Adipocytes

Todd Leff[1] *and James G. Granneman*[1,2]
[1] *Department of Pathology, Wayne State University School of Medicine, Detroit, MI, USA*
[2] *Department of Psychiatry, Wayne State University School of Medicine, Detroit, MI, USA*

1 Introduction – Adipocytes and the Regulation of Metabolism 3

2 Physiology of the Adipocyte 4
2.1 Absorption and Storage of Circulating Lipid 4
2.2 Mobilization and Release of Stored Lipid into the Circulation 5
2.3 Regulatory Mechanisms that Balance Lipid Storage and Utilization 6
2.4 Adipose Tissue as an Endocrine Organ 6

3 Developmental Origin of Adipocytes 9
3.1 Adipose Tissue Development and Plasticity 9
3.2 Differentiation of Adipocytes from Precursor Cells 11

4 Adipose Dysfunction and Metabolic Disease 13
4.1 Obesity – Too Much Fat 14
4.2 Lipodystrophy – Too Little Fat 15

5 The Adipocyte as a Therapeutic Target for Metabolic Disease 15
5.1 Antidiabetic PPARγ Ligands Act on Adipocytes 16

Bibliography 18
Books and Reviews 18
Primary Literature 19

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Keywords

Metabolism
The chemical processes that occur within a living organism in order to maintain life; the interconnected sequences of mostly enzyme-catalyzed chemical reactions by which a cell, tissue, organ, etc., sustains energy production, and synthesizes and breaks down complex molecules.

Lipids
Any of the large group of fats and fatlike compounds which occur in living organisms and are characteristically soluble in certain organic solvents but only sparingly soluble in water.

Fatty Acids
Long-chain aliphatic compounds that contain a carboxylic acid group.

Gene Expression
The process by which a gene's coded information is converted into the structures present and operating in the cell. Expressed genes include those that are transcribed into mRNA and then translated into protein and those that are transcribed into RNA but not translated into protein (e.g., transfer and ribosomal RNAs).

Transcription Factors
The DNA binding proteins that carry out the organic process whereby the DNA sequence in a a gene is copied into mRNA.

Nuclear Receptors
Transcription factors that are activated by the binding of specific ligans that are usually lipid soluble, such as fatty acids or steroids.

Adipose Tissue
Body tissue which is able to store high amounts of neutral fats.

Adipocytes
The cell type found in adipose tissue that is specialized for storage of neutral lipids.

The adipocyte, or fat cell, is a unique, highly specialized cell whose primary function is to provide a depot for storage of excess fat, derived mainly from the diet, which can be mobilized and distributed to other tissues as dictated by the energetic demands of the organism. Although long thought of as a simple passive storage depot for dietary lipid, it is now clear that adipose tissue plays a very active role in regulating energy balance and not only stores and secretes lipid as needed, but also produces hormonal signals that have multiple effects on energy metabolism in many tissue

and organs throughout the body. This new view of adipose tissue (and the adipocyte) as a key regulator of fuel metabolism and partitioning has dramatically increased the interest in the role that this tissue plays not only in normal physiology but also in the etiology and treatment of important metabolic diseases such as diabetes and obesity.

1 Introduction – Adipocytes and the Regulation of Metabolism

Although the physiology of adipose tissue and the molecular and cellular biology of the adipocyte have been subjects of research interest for many years, new work in this area has transformed our understanding of the adipocyte and its role in regulating human physiology. We now recognize that in addition to its long appreciated function as a lipid storage depot, the adipocyte plays a much more active role in regulating whole body fuel partitioning and metabolism. The growing interest in the physiology of adipose tissue and the role that it plays in metabolic regulation is fueled to a large degree by the realization that we are in the midst of a growing epidemic of the metabolic disorders of diabetes and obesity. In most developed countries, the incidence of both of these diseases has increased dramatically in the past few decades. Currently (as of 2003), there are 16 million diabetics in the United States, and the incidence of the disease has increased by 49% from 1990 to 2000, with projections indicating a 165% increase by 2050. There has also been a dramatic increase in the incidence of obesity, with over one-third of the population in the United States now classified as obese. Although the causal relationship between diabetes and obesity is not fully understood, a likely common link is the adipocyte.

In healthy individuals, excess fat is stored in adipocytes while only low amounts of triglyceride are maintained in nonadipocytes. It is thought that in obese individuals, the capacity for adipose tissue to accommodate excess lipid is exceeded, resulting in the abnormal accumulation of lipid in other tissues. This elevation in intracellular triglyceride content has been associated with physiological dysfunction (lipotoxicity) that contributes to the development of obesity-related type 2 diabetes. This pathological accumulation of lipid in nonadipose tissue may be the result of a physiological dysfunction of the adipocyte that is induced by the obese state. In this chapter, we will review the current state of knowledge about fat cell lipid metabolism, and how adipocytes function to balance lipid storage and mobilization to meet the energetic demands of the organism without exposing nonadipose tissues to deleterious fat accumulation.

In addition to its function as an energy-storage depot, we now understand that adipose tissue is also a bona fide endocrine organ, secreting hormones that regulate fat metabolism in other tissues throughout the body. The list of biologically active peptides known to be secreted by fat cells has grown significantly in recent years, and although the physiological function of most of these adipocyte-derived hormones (adipokines) is not fully understood, it is clear that they are important components of the physiological system

that controls lipid storage, distribution, and utilization throughout the body. Our current knowledge of the regulation of adipokine production and their downstream metabolic effects is reviewed below.

Our increased understanding of the active role that adipocytes play in regulating metabolism has stimulated a growing interest in adipose tissue as a therapeutic target for new agents to treat diabetes, obesity, and other metabolic diseases. The underlying assumption for this effort is that drugs acting on adipocyte lipid metabolism parameters or on hormone production pathways could have beneficial effects on metabolic abnormalities of diabetes and obesity. In support of this possibility are recent findings, which are reviewed below, demonstrating that an important family of antidiabetic drugs acts, at least in part, by modulating adipocyte physiology. A complete understanding of the medical significance of the adipocyte will require a comprehensive knowledge of the development of adipocytes and adipose tissue, of molecular biology and physiology of mature adipocyte, as well as the interaction of adipose tissue with the broader regulatory systems that control the whole body energy balance and fuel partitioning.

2 Physiology of the Adipocyte

Perhaps the most well-established function of adipose tissue in man is as a reservoir for the storage and mobilization of energy. Cells of the body utilize energy continuously, yet fuel is obtained from external sources intermittently. Thus, animals have a need to be able to efficiently store energy when food is available and access that energy during periods of fasting. Adipose tissue is highly specialized for the storage of lipid energy in the form of triglycerides and the mobilization of that energy in the form of free fatty acids. Of the energy substrates utilized by tissues, lipid is most efficiently stored and provides greater than twice the energy per gram than either carbohydrate (glycogen) or protein. The significance of this energy reserve is readily appreciated when considering that a man of normal body weight (70 kg, 10% body fat) can survive 40 days on energy reserves stored in adipose tissue.

2.1 Absorption and Storage of Circulating Lipid

Energy is stored in adipocytes in the form of triglycerides and is mainly derived from dietary fat, and *de novo* biogenesis from liver and within adipose tissue. Although human adipose tissue is capable of synthesizing lipid *de novo*, most triglycerides that accumulate in the tissue are derived directly or indirectly from dietary fat. Dietary fat is digested in the gut and repackaged as chylomicrons that are delivered to the systemic circulation via the lymphatics. The liver also plays a major role in the production and packaging of lipids in the form of various lipoproteins that are targeted for utilization in various target tissues, including adipose tissue.

A major mechanism for clearing chylomicrons and very low density lipoproteins (VLDL) from the circulation is through the action of lipoprotein lipase (LPL). As mentioned above, adipocytes express LPL, which is released and targeted to the capillary endothelium (Fig. 1). There, LPL acts on circulating chylomicrons and VLDL, hydrolyzing the neutral lipid core and releasing free fatty acids. The free fatty acids that are released can cross cell membranes and enter

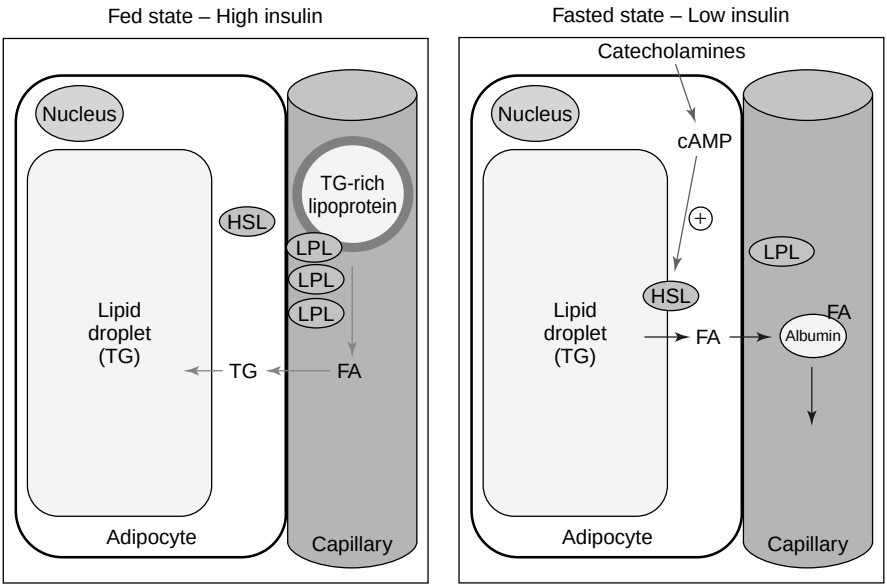

Fig. 1 Lipid storage and mobilization in adipocytes during fed and fasted states. See text for details. Flux of dietary lipid from circulating lipoprotein particles to the adipocyte lipid droplet is shown in green. Flux of stored lipid from adipocyte back into circulation shown in blue. TG, triglyceride; FA, fatty acid; HSL, hormone-sensitive lipase; LPL, lipoprotein lipase. (See color plate p. xxii).

adipocytes directly. Additionally, fatty acids can enter adipocytes via transport proteins, such as CD36 and FATP1. The flow of free fatty acids is governed by its concentration gradient. In the case of adipocytes, free fatty acid (FFA) that enter cells quickly encounter proteins that bind FFA, such as members of the fatty acid binding protein (FABP) family involved in cytoplasmic transport, or enzymes involved in metabolic conversion into triglyceride, such as long chain acyl-CoA synthase, thereby maintaining a significant concentration gradient.

2.2 Mobilization and Release of Stored Lipid into the Circulation

Triglycerides are typically stored within the adipocyte as a single lipid droplet. Recent work indicates that the lipid droplet is an extension of the endoplasmic reticulum surrounded by a single phospholipid layer. Closely associated with the lipid droplet are "coat proteins", especially perilipin isoforms and caveolin 2. Perilipin is the most abundant target for phosphorylation by cyclic AMP-dependent protein kinase (PK) in fat cells, and plays a key role in the retention and mobilization of energy from lipid droplets. The energy contained in the droplet is liberated when the triglyceride is hydrolyzed into free fatty acids and glycerol. The rate-limiting step in this process is the activity of hormone-sensitive lipase (HSL) at the surface of the lipid droplet. Although several protein kinases can influence lipolytic rate, the most significant of these is cyclic AMP-dependent protein kinase (PKA). Overall, the rate of lipolysis is

closely governed by the phosphorylation state of perilipin and HSL. Recent work indicates that phosphorylation of perilipin on multiple residues allows recruitment of phosphorylated HSL to the surface of the lipid droplet where triglyceride hydrolysis can take place. Phosphorylation of HSL has only a modest effect on the activity of the enzyme; rather, the dramatic increase in lipolysis produced by PKA-activation reflects the translocation and accessibility of the enzyme to its substrate within the cell. HSL appears to be physically associated with the cytosolic FABP4 (aP2). The significance of this association is not known, but could involve regulation of HSL activity or efflux of mobilized fatty acids. In this regard, mice lacking FABP4 show reduced rates of lipolysis.

Efflux of fatty acids from the fat cell and their transport across the capillary endothelium are thought to occur through passive diffusion across a concentration gradient, although this process has not been studied in detail. Long-chain fatty acids are essentially insoluble in aqueous solution and thus rely on carrier proteins for transport. In plasma, fatty acids are immediately bound with high affinity to serum albumin, which serves as a carrier of these substrates to the sites of oxidation. During fasting, mobilized fatty acids are quickly removed from the circulation, with a half-life of about two minutes, and provide an important source of energy for heart and skeletal muscle.

2.3 Regulatory Mechanisms that Balance Lipid Storage and Utilization

The pathways that promote FFA uptake and esterification to triglyceride are strongly regulated by hormones, particularly insulin. Insulin strongly promotes the synthesis of LPL whose expression coincides with the peak of triglyceride absorption. Insulin and chylomicrons also stimulate the secretion of acylation stimulation protein – a protein made by adipocytes that strongly promotes triglyceride synthesis. Insulin also promotes *de novo* fatty acid synthesis from glucose by stimulating glucose uptake and the expression of lipogenic enzymes.

Fatty acids, derived from adipose tissues, are an important energy source during fasting, exercise, and stress. As mentioned above, the overall rate of lipolysis is largely governed by the relative activation of PKA. The activity of PKA is governed by cAMP levels, which in turn are controlled by receptors coupled positively (beta adrenergic, glucagon, ACTH) or negatively (adenosine, niacin, alpha2, adrenergic) to adenylyl cyclase. In addition, insulin regulates adipocyte cAMP levels by activating phosphodiesterase PDE3b that degrades cAMP. The nature of the pathways controlling lipolysis depends on physiological circumstances. Lipolysis, in response to stress and exercise, is highly dependent on the activity of the sympathetic innervation of adipose tissue as well as the release of epinephrine from the adrenal gland. During fasting, lipolysis does not depend on neural activity, but rather appears to result from decline in insulin that provides tonic activation of PDE3b.

2.4 Adipose Tissue as an Endocrine Organ

As described above, the primary metabolic role of the adipocyte is to absorb and store excess lipid in the form of triglyceride, and to make it available to other tissues in the body as energy needs dictate, by measured release of fatty acids into the circulation. Proper functioning of

Tab. 1 Metabolically active proteins secreted by adipocytes.

Protein/hormone	*Physiological effects*
Leptin	Appetite, autonomic nervous activity
Adiponectin	Insulin sensitivity, fatty acid oxidation
Resistin	Insulin sensitivity
TNF-α	Insulin sensitivity, adipocyte differentiation, inflammation
ANG II	Lipogenesis, blood pressure
ASP	Lipogenesis
IGF	Lipogenesis, adipocyte differentiation
Interleukin-6	Lipolysis in adipocytes, inflammation
Kinins	Insulin sensitivity, tissue remodeling
PAI-1	Insulin sensitivity, blood clotting, atherosclerosis
TGF-β	Lipolysis, angiogenesis

Notes: ANG: angiotensin; ASP: acylation-stimulating protein; IGF: insulin-like growth factor; PAI-1: plasminogen activator inhibitor; TNF: tumor necrosis factor; TGF: transforming growth factor. Table adapted from Schling, P., Loffler, G. (2002) Cross talk between adipose tissue cells: impact on pathophysiology, *News Physiol. Sci.* **17**, 99–104.

this system requires communication between adipose tissue and essentially all the organ systems in the body. In addition to the hormonal and neural signals that regulate lipid uptake and storage, or induce lipolysis and release of fatty acids into the circulation during a fast (discussed above), there are signals that originate in adipose tissue that act to modify various physiological activities in tissues and organs throughout the body. For example, to insure that dietary intake is sufficient to maintain an adequate level of adiposity, there must be communication between adipose tissue and the centers in the brain that control appetite. Likewise, there is communication between adipose tissue and the organs, and the tissues that utilize fat for energy to insure that fatty acid delivered to nonadipose tissues, such as muscle and liver, are handled properly and do not accumulate to abnormal levels. One of the major advances in metabolic research in the last few years has been the discovery that adipocytes secrete hormones that act at specific sites in the body and have important effects on many aspects of energy metabolism. This new understanding of adipose tissue as an endocrine organ has dramatically changed our understanding of the significance of adipocytes in the regulation of metabolism. It is now believed that the adipose-derived hormones (referred to as adipokines) are important components of the integrated system of hormonal and neural signaling pathways that function to regulate the storage and use of metabolic energy. The following is a brief summary of the biology of four of the adipokines that have relatively clear effects on metabolism. Table 1 presents a more complete list of hormone like proteins produced by adipocytes.

The seminal contribution to the concept that adipose tissue produces hormones with important metabolic effects was made by Friedman and colleagues in 1994. These investigators identified the protein product of the *obese* (*ob*) gene that causes

severe obesity in mice when mutated. Leptin is secreted from adipose tissue and regulates body weight by acting directly in the CNS to inhibit feeding behavior. The control of leptin synthesis and secretion is still poorly understood. In general, leptin production and secretion are promoted under conditions of positive energy balance (fed state, high insulin) and suppressed by conditions of net energy deficit (e.g. fasting, catabolic hormone stimulation). As such, plasma leptin levels correlate strongly with total adipose tissue mass, and thereby provide an integrated, long-term signal indicating the status of lipid reserves. The actions of leptin are mediated through specific cell surface receptors, which are located in key central and peripheral target cells. Activation of leptin receptors in diverse brain regions signal a state of positive energy balance. Leptin-sensitive neural systems regulate the activity of the autonomic nervous system involved in energy storage and mobilization, feeding behavior, reproductive physiology, and sexual behavior. Leptin may also have direct effects on energy metabolism in peripheral tissues such as muscle, where it has been reported to cause an increase in fatty acid oxidation rates. Although leptin behaves as an antiobesity hormone in certain animal models, common human obesity does not appear to be due to abnormally low leptin levels.

Another recently identified adipocyte-secreted hormone that may play a role in both obesity and diabetes is adiponectin (also called ACRP30 or adipoQ). Originally identified as a secreted fat-specific protein whose expression was induced following adipogenesis, adiponectin levels were found to be reduced in obesity and increased by weight loss. In addition, the *adiponectin* gene maps to a region on chromosome 3 that is associated with diabetes and metabolic syndrome. Treatment of rodents with adiponectin was found to increase muscle fatty acid oxidation, reverse insulin resistance and improve hepatic insulin action. Together, these observations suggest that the physiological role of adiponectin may be to promote lipid oxidation in nonadipose tissues; in essence it may be a signal from fat indicating to the rest of the body that lipid energy is available and should be used.

In contrast to adiponectin, an adipokine that has recently been identified called *resistin* appears to have diabetes-promoting effects on metabolism. While adiponectin clearly promotes fatty acid oxidation and appears to have insulin-sensitizing effects throughout the body, resistin (also known as adipocyte secreted factor, ADSF or FIZZ3) was found to be over-expressed in rodent models of diet-induced obesity and to induce insulin resistance and glucose intolerance in normal mice. These data suggest that resistin acts in a converse manner to adiponectin, increasing insulin resistance and promoting the development of diabetes. However, this relationship between resistin and diabetes was not observed in all models of the disease and additional work will need to be carried out to fully clarify the role of resistin as another potential link between obesity and diabetes. Another potentially prodiabetic adipokine is the inflammatory cytokine tumor necrosis factor alpha (TNFα), which is secreted by adipocytes under some circumstances. TNFα production by adipocytes is elevated in obese rodents and humans and positively correlates with insulin resistance and in some studies inactivation of TNFα using antibody treatment improved insulin action. As with resistin, the combination of elevated expression in obesity

and insulin resistance promoting activity of these adipokines raises the possibility that it contributes to the functional link between obesity and diabetes. Although a great deal more work needs to be done in this area before we can fully appreciate the multiple roles that adipokine hormones play in the regulation of metabolism, it is clear that they are a crucial component of the physiological system that regulates energy balance and fuel partitioning.

3 Developmental Origin of Adipocytes

3.1 Adipose Tissue Development and Plasticity

The basic functional unit of adipose tissue is the adipocyte. Nevertheless, adipose tissue is complex and contains several cell types in addition to adipocytes, such as endothelial cells, interstitial cells, undifferentiated mesenchymal cells, pericytes, and "very small adipocytes". Indeed, adipocytes constitute less than 20% of the cells residing in typical adult fat tissue. Moreover, there are very important interactions among the various cell types that are critical to the proper functioning of the tissue. In view of the huge increase in obesity rates in the United States, and its negative impact on health, new attention has been focused on the development, maintenance, and plasticity of this important tissue.

Analysis of adipose tissue histogenesis and remodeling has relied mainly upon descriptive approaches to define cell phenotypes and deduce their transition to mature cells. In humans, adipose tissue appears as distinct lobules during the second trimester of fetal development. The specific timing of adipose tissue histogenesis and fat cell differentiation varies according to location in the body. Adipocytes within fat tissue are thought to derive initially from mesenchymal progenitors capable of differentiating into bone, muscle, as well as fat. Mesenchymal cells that are highly committed to the adipocyte lineage first appear closely associated with vessel formation, and there appears to be a close reciprocal association of developing fat cells with angiogenesis. This is not surprising since early committed cells express lipoprotein lipase that is targeted at the capillary lumen, and provides the mechanism for the transport of dietary triglyceride to fill the developing fat cells. As fat cells develop, triglycerides coalesce into small lipid droplets (nearly all of which are triglycerides) within the cytoplasm that eventually fuse to form a large single lipid droplet. The typical mature adipocyte is relatively large (30–50 micron diameter) and can reach a size of greater than 120 μm under certain conditions (Fig. 2).

Experimental investigations of adipose tissue have mainly utilized rodent models, although it is important to note that the ontogeny of adipose tissue varies widely among species, and even among fat depots within a given species. In rodent models, white adipose tissue generally appears late in embryonic development and continues to expand and differentiate during the neonatal period prior to weaning. Classic "flash" labeling experiments with an ^{3}H thymidine have shown that most proliferation of cells that are destined to become adipocytes occurs in the first postnatal week. Mitoses are mostly found in poorly differentiated mesenchymal cells that are closely associated with developing capillaries. The transition of cells from mesenchymal progenitors to mature cells can be deduced by evaluating the cellular distribution of ^{3}H label over time following

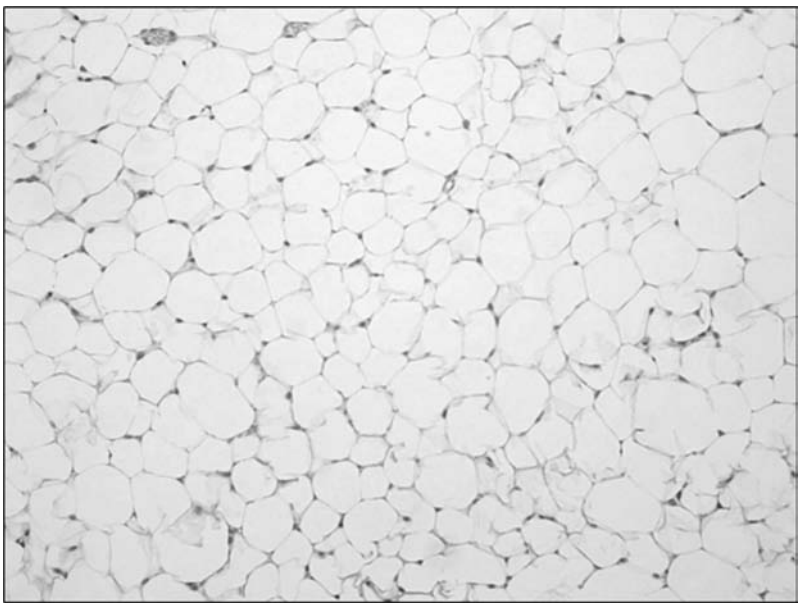

Fig. 2 Histological appearance of normal mouse adipose tissue. Shown is a 6-μm paraffin section stained with hematoxylin and eosin. Note the single large lipid droplet with numerous interstitial cells.

flash labeling. Over time, the percentage of labeled mesenchymal cells declines as the label appears in cells that have accumulated lipid. In contrast, labeling of endothelial cells remains relatively constant. These data suggest a dynamic process in which mature adipocytes are derived from committed mesenchymal progenitors that divide and develop into adipocytes. Interestingly, nearly 90% of the initial label is lost by five months, strongly indicating that cellular renewal occurs throughout life. In this regard, it is well established that cells can be isolated from human and rodent adipose tissue that readily differentiate into mature adipocytes *in vitro*. Indeed, pluripotent progenitors derived from adult adipose tissue may have numerous therapeutic applications. These observations indicate that adipose tissue contains a significant population of committed progenitors that are capable of contributing to tissue renewal and remodeling under appropriate conditions.

Under normal laboratory conditions, cellular proliferation in rat adipose tissue drops to very low levels after weaning. Nonetheless, a variety of physiological and pharmacological conditions reveal dynamic regulation of adipose tissue. For example, Hirsch and colleagues demonstrated in the 1970s that the obesity produced by high fat diets in rats involves both fat cell hyperplasia as well as hypertrophy. Fat cell renewal has also been observed after partial lipectomy, and elevated fat cell turnover has been observed in models of hypothalamic obesity. One of the best examples of physiological adipose tissue plasticity occurs in seasonal fat deposition of hibernators. Although the mechanisms involved in fat cell proliferation are largely unknown, adipose tissue itself is a rich source of growth factors and cytokines

that could trigger tissue expansion. As discussed below, certain pharmacological agents that exhibit antidiabetes properties in rodents and man target receptors that are enriched in fat tissue, and produce pronounced tissue remodeling that is likely related to the therapeutic actions of these agents.

3.2 Differentiation of Adipocytes from Precursor Cells

Together with adipocyte size, the number of adipocytes in the body is an important determinant of obesity and of multiple parameters of energy metabolism. The number of adipocytes present in an organism is determined to a large degree by the adipocyte differentiation process that generates mature adipocytes from fibroblast-like preadipocytes. Many of the molecular details of this process are now known, and the following section summarizes our current understanding of the molecular control of adipogenesis. It is important to note that our understanding of how adipocytes are generated from precursor cells is based primarily on cell culture models of adipogenesis such as the mouse 3T3L1 cell line. While these cell lines are very amenable to experimentation, they produce adipocytes that are strikingly different in some respects than native adipocytes found in adipose tissue *in vivo*. For example, fully differentiated 3T3L1 adipocytes are multilocular (contain multiple lipid droplets), while native adipocytes in white fat (the predominant type of adipose tissue in humans) display a unilocular distribution of lipid (compare Figs. 2 and 3). While we know that many of the characteristics of adipocyte differentiation in cultured cell lines are also important features of *in vivo* adipogenesis, it is important to bear in mind that some aspects of adipogenesis that have been learned from cell culture systems, as described below, may differ from the process as it occurs *in vivo*.

When cultured preadipocyte cultures are grown to confluence and cease cellular division (growth arrest), they can be induced to differentiate into adipocytes by treatment with an adipogenic hormonal cocktail containing insulin, dexamethasone, and an inducer of intracellular cAMP concentration. One of the first steps in the process of adipogenesis is the reentry of growth-arrested preadipocytes into the cell cycle and the completion of several rounds of clonal expansion. Multiple genes involved in the cell cycle control are required for this step to proceed, including the tumor suppressor retinoblastoma protein (Rb) and several cyclin-dependent kinases and their inhibitors (p18, p21, and p27). This and the subsequent steps of the program of adipogenesis are controlled, to a large degree, by a cascade of gene expression events regulated by a small set of transcription factors. Two families of transcription factors have emerged as the key determinants of this process: the three CCAAT/enhancer-binding proteins C/EBP α, β and δ, and the two-peroxisome proliferator-activated receptors gamma-1 and gamma-2 (PPARγ1 and PPARγ2).

One of the initial steps in the transcriptional cascade in response to adipogenic signals is the rapid induction of C/EBP β and δ expression. These transcription factors orchestrate cell cycle reentry by stimulating the expression of the CDK inhibitor p21, which acts to inhibit the Rb protein and relieve its block on cell cycle progression. C/EBP δ and β have also been shown to induce expression of the gene for the PPARγ transcription factor that plays

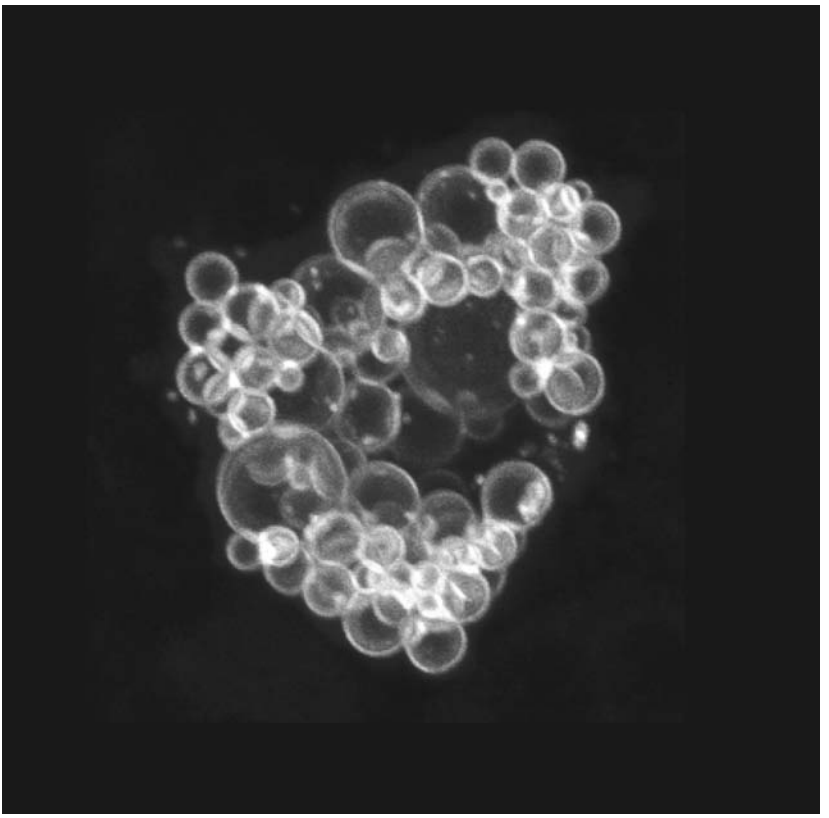

Fig. 3 Fluorescence micrograph of a cultured adipocyte expressing EGFP–perilipin fusion protein. A C3H 10T1/2 mesenchymal cell was differentiated in culture and transfected with an expression vector encoding perilipin fused to enhanced green fluorescent protein. Note the localization of the fluorescence to surface of the numerous lipid droplets (see color plate p. xxi).

a key role in the terminal differentiation of adipocytes (discussed in more detail below). The importance of C/EBP β and δ for adipogenesis was clearly demonstrated by loss-of-function and gain-of-function genetic studies in mice. Overexpression of either C/EBP β or δ in preadipocytes enhanced adipogenesis, while embryonic fibroblast cells derived from mice lacking either C/EBP β or δ had reduced levels of adipogenesis compared with wild type.

The induction of C/EBP β and δ is immediately followed by an increase in PPARγ and C/EBPα expression. PPARγ is a member of the nuclear hormone receptor family of ligand-activated transcription factors. It is absolutely required for adipocyte differentiation, as a genetic knockout of the PPARγ gene in mice prevents the development of all fat tissue. In addition to its crucial role in adipocyte differentiation PPARγ is the receptor for the thiazolidinedione (TZD) class of antidiabetic drugs, indicating that it is also important in metabolic regulation in adult organisms. In mice and humans there are two isoforms, PPARγ1 and PPARγ2, which are derived from the same gene by

alternative promoter usage and RNA splicing. While the expression of PPARγ2 is restricted almost exclusively to adipocytes, PPARγ1 has a broader pattern of expression although it is still most abundant in adipocytes. Although PPARγ2 is identical to PPARγ1 except that it contains an additional 28 amino acids on its N-terminus, it appears that the two proteins have distinct activities with regard to adipocyte differentiation. When the expression of the PPARγ2 isoform was blocked, adipogenesis was more strongly inhibited than when the PPARγ1 isoform was blocked. In addition, exogenous delivery of PPARγ2 into PPARγ deficient cells was able to completely restore the adipogenesis, whereas overexpression of PPARγ1 had little effect. It may be that PPARγ1, which is already expressed in preadipocytes, behaves as a priming factor (along with C/EBP β and δ) for the induction of PPARγ2 or for the generation of endogenous PPARγ ligands that play a role in later stages of adipogenesis.

As the program of differentiation proceeds, the expression of C/EBPα rises immediately after the increase in PPARγ2 expression. Like PPARγ, C/EBPα also plays an essential role in adipose development as targeted gene knockout in mice results in embryonic lethality and failure to develop normal adipose tissue. There has been an intense research effort to understand the relationship between these two transcription factors and the role they play in adipogenesis. Several studies have demonstrated that PPARγ2 and C/EBPα coregulate each other's expression. Mice with reduced PPARγ expression due to heterozygous gene knockout displayed a drastically reduced level of C/EBPα, and mice with disrupted C/EBPα expression showed a reduced level of PPARγ. Introduction of either PPARγ or C/EBPα into NIH3T3 cells is sufficient to convert these normally nonadipogenic cells from fibroblasts into adipocytes. However, it is unclear if either of the transcription factors, completely on its own, could induce adipogenesis. Taken together, most of the recent evidence supports the model that while both of the transcription factors work coordinately to carry out adipogenesis, PPARγ2 probably plays the primary role, while C/EBPα may act mostly by inducing and maintaining PPARγ2 expression. C/EBPα may also function to regulate the transcription of genes involved in the metabolic actions of insulin such as glucose transporter 4 (Glut 4). Clearly, PPARγ and C/EBPα are key transcription factors in adipogenesis, acting synergistically to generate fully differentiated, insulin-responsive adipocytes.

Although our understanding of adipocyte differentiation, as described above, is derived from work in cultured cell lines, it is likely that many of the pathways and key components also play an important role in generating adipocytes from precursor cells *in vivo* and in the remodeling of adipose tissue as it occurs during certain metabolic stress (see Sect. 3, above) or treatment with specific pharmacological agents (see Sect. 5, below).

4 Adipose Dysfunction and Metabolic Disease

Just as our appreciation of how active a role the adipose tissue plays in the metabolic regulation has grown in recent years, so too has the interest in the role that adipocytes play in the development (and treatment) of metabolic diseases such as obesity and diabetes. The investigation of the role of adipose tissue in these diseases has taken on added significance in recent years

because of the dramatic rise in the prevalence of these related disorders. Although the causal relationship between diabetes and obesity is not fully understood, a likely common link is the adipocyte. Given that adipose tissue serves to buffer excess lipid, and to confer insulin resistance that is associated with inappropriate accumulation of lipid in nonadipose cells, it is not surprising that the dysfunction in adipose tissue could profoundly affect diabetes susceptibility. The inability of adipose tissue to buffer circulating lipid levels and concurrent accumulation of lipid in muscle and liver may, in fact, be an early event in the development of diabetes. This potential link between adipose tissue function and diabetes raises the new and exciting possibilities for the development of therapeutic agents to treat diabetes. Drugs targeted to very specific metabolic or hormonal functions in the adipocyte could potentially have a profound beneficial effects on metabolism throughout the body. The following is a summary of the known defects in adipose function and their effects on energy metabolism and glucose homeostasis. We will also review what is known about the antidiabetic thiazolidinedione drugs that are thought to have their effects by modifying adipocyte physiology.

4.1 Obesity – Too Much Fat

A long-standing observation in the field of diabetes research is that obese individuals have an increased risk of developing insulin resistance and type 2 diabetes. The current epidemic of diabetes in western cultures began about thirty years ago and correlates precisely with the concurrent explosion of obesity. This has led to the assumption that there is a functional, or causative relationship between these two disorders; that the presence of obesity increases susceptibility to diabetes. Although the relation between these two metabolic disorders is unmistakable, it is not understood how obesity precipitates the multiple functional defects that precede overt diabetes. However, an emerging candidate is the acquired inability of adipocyte to protect important tissues and cells throughout the body from exposure to excessive lipids.

Normally, excess calories are stored in the body in the form of triglycerides, mainly in adipocytes of white fat tissue. However, nonadipocyte cells also store low levels of triglycerides to meet internal metabolic requirements. In these cells, the amount of stored triglyceride is maintained within a very narrow range. A fatty acid homeostatic system functions in healthy individuals to direct excess fat to adipocytes while maintaining normal amounts of triglyceride in nonadipocytes. Under conditions of chronic positive energy balance and ensuing obesity, however, the capacity for adipose tissue to buffer excess lipid energy may be exceeded, resulting in the pathological accumulation of lipid within key metabolic tissues.

The elevation of free fatty acid (FFA) in the blood is predictive of conversion from impaired glucose tolerance to diabetes, suggesting that FFA themselves could contribute to organ defects that precede type 2 diabetes, such as insulin resistance of skeletal muscle, and reduced pancreatic insulin secretion. More recently, it has been shown that intracellular accumulation of lipid within myocytes strongly predicts insulin resistance *independent of the magnitude of obesity*. Thus, it is not the accumulation of lipid in fat cells *per se* that is problematic, rather it is the redirection of lipid to other key cells that occurs once the capacity of the adipocyte to handle

lipid has been exceeded. This condition is perhaps most dramatically illustrated by patients with lipodystrophy, who are severely insulin-resistant and develop diabetes at an early age. These individuals lack certain adipose tissue depots and therefore have reduced capacity to buffer excess lipid (presented in more detail below). Viewed from this perspective, obesity leads to diabetes when abnormal amounts of triglyceride accumulate in nonadipose tissues such as skeletal muscle and pancreatic islets. Excess triglyceride deposition in nonadipocytes might generate a lipid environment that interferes with cellular physiology and gene expression in ways that contribute to insulin resistance and β-cell failure associated with type 2 diabetes.

4.2
Lipodystrophy – Too Little Fat

The model described above proposes that obesity promotes the development of type 2 diabetes because adipocytes from obese individuals have reduced capacity for further lipid storage, which leads to inappropriate fat accumulation in nonadipose tissue. As mentioned above, the effects of aberrant amounts of adipose tissue on physiology are best illustrated in the rare, but informative, lipodystrophy syndromes. These inherited and acquired syndromes are characterized by the partial or complete loss of adipose tissue, insulin resistance, and diabetes. Several mutations that cause inherited forms of lipodystrophy, such as partial familial lipodystrophy (PFLD), have been identified. One of the most common genetic defects in PFLD are mutations in the gene encoding the nuclear lamin a/c protein. Although the mechanism by which mutant lamin a/c protein causes lipodystrophy is not known at this time, it seems likely that the protein, which plays a fundamental role in the function of the nucleus, is required for normal adipocyte differentiation of maintenance. Patients carrying PFLD mutations exhibit an absence of adipose tissue in their extremities, elevated circulating lipids, abnormally high levels of intracellular triglyceride in muscle and liver, and diabetes. Interestingly, mutations causing PFLD have recently been identified in the PPARγ gene, underscoring the key role of this transcription factor in the development and maintenance of adipose tissue.

The conclusions derived from observations of human lipodystrophy syndromes are strongly supported by animal models of lipodystrophy. Several different genetic methodologies have been used to generate mice that have reduced amounts of adipose tissue. Like their human counterparts, these animals invariably exhibit elevated circulating lipids, abnormally high levels of intracellular triglyceride levels in multiple tissues including muscle and liver, severe insulin resistance, and diabetes. Interestingly, these metabolic defects can be improved by transplanting adipose tissue back into the fatless mice. These findings clearly illustrate the vital role that adipose tissue plays in protecting the organism from abnormal lipid accumulation and its deleterious effects on metabolism.

5
The Adipocyte as a Therapeutic Target for Metabolic Disease

As our understanding of the importance of adipose tissue in controlling metabolism has grown, so too has the realization that therapeutic agents designed to modify adipocyte physiology could provide new avenues for the treatment of diabetes, obesity, and other metabolic

diseases. As described above, our current understanding of the role of adipose in protecting other tissue from deleterious lipid accumulation suggests that drugs acting on this aspect of adipocyte function might have beneficial effects. Likewise, the observations (also described above) on adipocyte derived hormones and their metabolic activities, suggests that therapeutic agents modulating the production of key adipokines could also improve certain metabolic parameters. The view that drugs acting on adipocytes could be useful is reinforced by the relatively recent observation that a major class of antidiabetic drugs, the thiazolidinediones, have direct effects on adipocytes. Although this is an active area of research, a generally accepted model of TZD action has emerged in which the antidiabetic effect of these drugs is mediated by activation of the PPARγ transcription factor in adipocytes. The following is a review of our current understanding of how activation of PPARγ in adipocytes might improve insulin sensitivity and other metabolic parameters. This model can serve as a general example of how modification of adipocyte physiology could improve whole body metabolic status.

5.1 Antidiabetic PPARγ Ligands Act on Adipocytes

In patients with type 2 diabetes, treatment with TZD compounds results in a dramatic improvement in peripheral insulin sensitivity and a reduction in plasma glucose concentrations. The TZD compounds were first identified in the 1980s as antidiabetic agents in rodents, well before the discovery of the PPARγ receptor and in the absence of any knowledge of their mechanism of action. In the 1990s, it was discovered that TZDs could activate PPARγ and cause the differentiation of preadipocytes into adipocytes and it is now generally accepted that the antidiabetic activities of the TZDs are mediated by activation of PPARγ. Perhaps the clearest evidence that PPARγ activation mediates the antidiabetic effects of these drugs are the recently discovered synthetic ligands for PPARγ that have been selected exclusively for their ability to activate the receptor. These non-TZD PPARγ agonists show very similar antidiabetic activity to the TZDs.

Although the exact mechanism by which these drugs improve peripheral insulin sensitivity and reduce plasma glucose concentration is not fully understood, several general possibilities have emerged. First, TZDs may have a beneficial effect on metabolism by increasing the fat cell number and size, leading to greater lipid storage capacity and increased protection of nonadipose tissues from the deleterious effects of excess lipid accumulation. Another scenario is that, PPARγ agonists act on the mature adipocyte to alter the production of adipose-derived hormones or metabolic signals that function to improve metabolic parameters in other tissues and organs such as muscle, liver and pancreas. Finally, it is also possible that TZDs exert some of their metabolic effects through PPARγ present in nonadipose tissues such as skeletal muscle.

One feature of TZD treatment that could be central to the therapeutic actions of these compounds is a remodeling of adipose tissue. TZD treatment induces the appearance of clusters of small multilocular adipocytes and loss of large unilocular adipocytes in Zucker diabetic rats. This effect is also observed in mice treated with TZD compounds as seen in Fig. 4. It has been hypothesized that the newly

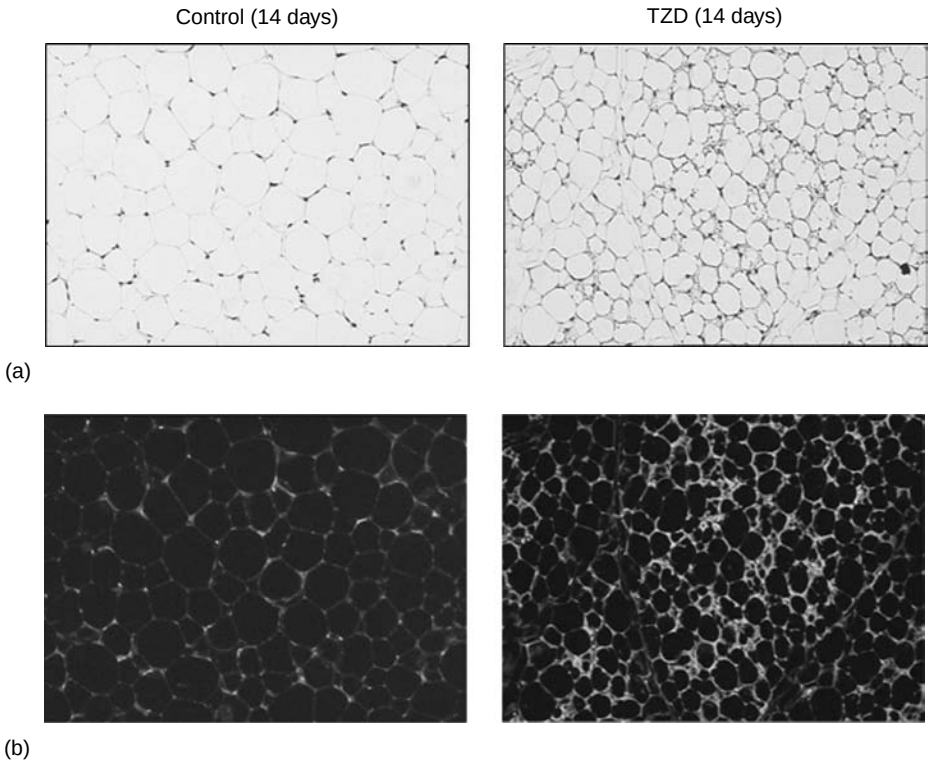

Fig. 4 Thiazolidinedione treatment remodels white adipose tissue. Mice were treated for 2 weeks with the TZD T174, and subcutaneous white fat was processed by (a) hematoxylin and eosin staining and (b) tissue autofluorescence. Note the appearance of numerous clusters of small multilocular adipocytes in TZD-treated tissue.

appearing small adipocytes might be more insulin sensitive and/or secrete lower levels of prodiabetes hormones and thereby contribute to the insulin-sensitizing effects of the drug. It is also possible that the increased number of smaller adipocytes, especially in the appropriate adipose beds, may improve the ability of adipose tissue to store excess lipid and reduce deleterious accumulation of triglyceride in muscle, liver, and pancreatic islets. In support of this possibility is the observation that many of the known target genes for PPARγ, whose expression would presumably be activated upon TZD treatment, are involved in lipogenesis. The increased rates of lipogenesis resulting from gene activation may increase the capacity of adipocytes to store lipid, thereby preventing triglyceride accumulation (and lipotoxicity) in nonadipose tissue. Whether the small adipocytes are derived from stem cell mitosis, recruitment of committed preadipocytes, or possibly by division of mature cells is not known. The loss of large fat cells was attributed to cellular apoptosis; however, the impact of TZD treatment on the fate and turnover of mature adipocytes has not been investigated directly.

Consistent with the observations of adipose tissue remodeling is the increased subcutaneous fat mass, and reduced

visceral fat mass seen in diabetic patients treated long-term with TZDs. Visceral fat is known to be more lipolytic in response to catecholamine stimulation than subcutaneous fat, and to efficiently deliver free fatty acids and other secreted factors to insulin-sensitive tissues such as liver and muscle, possibly causing an increase in insulin resistance. Although intrinsic metabolic differences between subcutaneous and visceral fat are not completely understood, current evidence suggests that subjects with increased visceral fat are at considerably higher risk for diabetes and cardiovascular complications than those with increased subcutaneous fat. These observations, plus the demonstration that PPARγ levels are higher in subcutaneous than in visceral fat, raise the possibility that PPARγ activation by TZDs is fat depot specific, and that differential activation of PPARγ in subcutaneous fat leads to a beneficial reproportioning of key metabolically active adipose beds.

The other possible mechanism by which activation of PPARγ in adipocytes could have effects on metabolism throughout the body, is by modulation of adipokine production. Adiponectin is an excellent candidate for a fat-derived hormone mediating the antidiabetic effects of PPARγ ligands. As described above, adiponectin has antidiabetic and antiobesity activity when introduced into rodents. Importantly, it has recently been demonstrated that levels of adiponectin are increased in patients treated with TZDs and that its expression in adipocytes is induced by PPARγ agonists. Another possibility is that the adipokine, resistin is negatively regulated by PPARγ. It has been reported that in some models, treatment with TZDs results in a reduction of resistin synthesis, which would have beneficial effects on insulin sensitivity. Finally, TZDs, could suppress expression of genes, such as TNFα and PAI-1, that might also contribute to systemic insulin resistance. Interestingly a mutual antagonism exists between TNFα and PPARγ; TNFα inhibits PPARγ expression in adipocytes whereas PPARγ activation by TZDs can partially overcome the diabetogenic effects of TNFα, potentially explaining at least some of the insulin-sensitizing activity of PPARγ ligands.

While there is still obviously a great deal to learn about which of these many possible mechanisms actually mediate the effects of PPARγ activation in adipose tissue, it is clear that an understanding of how these drugs affect adipose physiology will provide many clues into the complex pathways by which the adipocyte influences metabolism throughout the organism.

See also Diabetes Insipidus, Molecular Biology of.

Bibliography

Books and Reviews

Frayn, K.N., Karpe, F., Fielding, B.A., Macdonald, I.A., Coppack, S.W. (2003) Integrative physiology of human adipose tissue, *Int. J. Obes.* **27**, 875–888.

Friedman, J.M. (2002) The function of leptin in nutrition, weight, and physiology, *Nutr. Rev.* **60**, S1–14; discussion S68–84, 5–7.

Londos, C., Brasaemle, D.L., Schultz, C.J., Adler-Wailes, D.C., Levin, D.M., et al. (1999) On the control of lipolysis in adipocytes, *Ann. N. Y. Acad. Sci.* **892**, 155–168.

Picard, F., Auwerx, J. (2002) PPAR(gamma) and glucose homeostasis, *Annu. Rev. Nutr.* **22**, 167–197.

Renold, A.E., Cahill, J.F. (1965) Adipose Tissue, *Handbook of Physiology, A Critical, Comprehensive Presentation of Physiological Knowledge and Concepts*, Section 5, American Physiological Society, Washington, DC.

Rosen, E.D., Spiegelman, B.M. (2000) Molecular regulation of adipogenesis, *Annu. Rev. Cell. Dev. Biol.* **16**, 145–171.

Primary Literature

Agarwal, A.K., Garg, A. (2002) A novel heterozygous mutation in peroxisome proliferator-activated receptor-gamma gene in a patient with familial partial lipodystrophy, *J. Clin. Endocrinol. Metab.* **87**, 408–411.

Akazawa, S., Sun, F., Ito, M., Kawasaki, E., Eguchi, K. (2000) Efficacy of troglitazone on body fat distribution in type 2 diabetes, *Diabetes Care* **23**, 1067–1071.

Atanassova, P. (1998) Histochemical and ultrastructural criteria for early differentiation of human subcutaneous adipocytes, *Folia Med. (Plovdiv)* **40**, 24–27.

Atanassova, P.K. (2002) Cytochemical study of in situ differentiation of white adipose cells in rat, *Folia Med. (Plovdiv)* **44**, 32–36.

Bartness, T.J., Wade, G.N. (1985) Photoperiodic control of seasonal body weight cycles in hamsters, *Neurosci. Biobehav. Rev.* **9**, 599–612.

Berg, A.H., Combs, T.P., Du, X., Brownlee, M., Scherer, P.E. (2001) The adipocyte-secreted protein Acrp30 enhances hepatic insulin action, *Nat. Med.* **7**, 947–953.

Brodows, R.G., Campbell, R.G., Al-Aziz, A.J. (1976) Lack of central autonomic regulation of substrate during early fasting in man, *Metabolism* **25**, 803–807.

Brown, D.A. (2001) Lipid droplets: proteins floating on a pool of fat, *Curr. Biol.* **11**, R446–R449.

Cianflone, K., Xia, Z., Chen, L.Y. (2003) Critical review of acylation-stimulating protein physiology in humans and rodents, *Biochim. Biophys. Acta* **1609**, 127–143.

Cobb, J.E., Blanchard, S.G., Boswell, E.G., Brown, K.K., Charifson, P.S., et al. (1998) N-(2-Benzoylphenyl)-L-tyrosine PPARgamma agonists. 3. Structure-activity relationship and optimization of the N-aryl substituent, *J. Med. Chem.* **41**, 5055–5069.

Darlington, G.J., Ross, S.E., MacDougald, O.A. (1998) The role of C/EBP genes in adipocyte differentiation, *J. Biol. Chem.* **273**, 30057–30060.

de Souza, C.J., Eckhardt, M., Gagen, K., Dong, M., Chen, W., et al. (2001) Effects of pioglitazone on adipose tissue remodeling within the setting of obesity and insulin resistance, *Diabetes* **50**, 1863–1871.

De Ugarte, D.A., Ashjian, P.H., Elbarbary, A., Hedrick, M.H. (2003) Future of fat as raw material for tissue regeneration, *Ann. Plast. Surg.* **50**, 215–219.

Faust, I.M., Johnson, P.R., Hirsch, J. (1979) Adipose tissue regeneration in adult rats, *Proc. Soc. Exp. Biol. Med.* **161**, 111–114.

Faust, I.M., Johnson, P.R., Stern, J.S., Hirsch, J. (1978) Diet-induced adipocyte number increase in adult rats: a new model of obesity, *Am. J. Physiol.* **235**, E279–E286.

Freake, H.C., Moon, Y.K. (2003) Hormonal and nutritional regulation of lipogenic enzyme mRNA levels in rat primary white and brown adipocytes, *J. Nutr. Sci. Vitaminol. (Tokyo)* **49**, 40–46.

Friedman, J.M. (2002) The function of leptin in nutrition, weight, and physiology, *Nutr. Rev.* **60**, S1–S14; discussion S68–84, 5–7.

Fruebis, J., Tsao, T.S., Javorschi, S., Ebbets-Reed, D., Erickson, M.R., et al. (2001) Proteolytic cleavage product of 30-kDa adipocyte complement-related protein increases fatty acid oxidation in muscle and causes weight loss in mice, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2005–2010.

Gimble, J.M. (2003) Adipose tissue-derived therapeutics, *Expert Opin. Biol. Ther.* **3**, 705–713.

Haemmerle, G., Zimmermann, R., Hayn, M., Theussl, C., Waeg, G., et al. (2002) Hormone-sensitive lipase deficiency in mice causes diglyceride accumulation in adipose tissue, muscle, and testis, *J. Biol. Chem.* **277**, 4806–4815.

Hausberger, F.X. (1966) Pathological changes in adipose tissue of obese mice, *Anat. Rec.* **154**, 651–660.

Hausman, G.J., Kauffman, R.G. (1986) Mitotic activity in fetal and early postnatal porcine adipose tissue, *J. Anim. Sci.* **63**, 659–673.

Hausman, G.J., Thomas, G.B. (1984) Enzyme histochemical differentiation of white adipose tissue in the rat, *Am. J. Anat.* **169**, 315–326.

Hegele, R.A. (2001) Molecular basis of partial lipodystrophy and prospects for therapy, *Trends Mol. Med.* **7**, 121–126.

Hegele, R.A., Cao, H., Frankowski, C., Mathews, S.T., Leff, T. (2002) PPARG F388L, a transactivation-deficient mutant, in familial partial lipodystrophy, *Diabetes* **51**, 3586–3590.

Hellerstein, M.K. (1999) De novo lipogenesis in humans: metabolic and regulatory aspects, *Eur. J. Clin. Nutr.* **53**(Suppl. 1), S53–S65.

Hellerstrom, B.H.A.C. (1961) Cell renewal in the white and brown fat tissue of the rat, *Acta Pathol.* **51**, 347–353.

Hotamisligil, G.S., Arner, P., Caro, J.F., Atkinson, R.L., Spiegelman, B.M. (1995) Increased adipose tissue expression of tumor necrosis factor-alpha in human obesity and insulin resistance, *J. Clin. Invest.* **95**, 2409–2415.

Kelley, D.E., Goodpaster, B.H., Storlien, L. (2002) Muscle triglyceride and insulin resistance, *Annu. Rev. Nutr.* **22**, 325–346.

Londos, C., Brasaemle, D.L., Schultz, C.J., Adler-Wailes, D.C., Levin, D.M., et al. (1999) On the control of lipolysis in adipocytes, *Ann. N. Y. Acad. Sci.* **892**, 155–168.

Merkel, M., Eckel, R.H., Goldberg, I.J. (2002) Lipoprotein lipase: genetics, lipid uptake, and regulation, *J. Lipid Res.* **43**, 1997–2006.

Miller, W.H., Jr, Faust, I.M., Hirsch, J. (1984) Demonstration of de novo production of adipocytes in adult rats by biochemical and radioautographic techniques, *J. Lipid Res.* **25**, 336–347.

Minokoshi, Y., Kim, Y.B., Peroni, O.D., Fryer, L.G., Muller, C., et al. (2002) Leptin stimulates fatty-acid oxidation by activating AMP-activated protein kinase, *Nature* **415**, 339–343.

Mokdad, A.H., Bowman, B.A., Ford, E.S., Vinicor, F., Marks, J.S., Koplan, J.P. (2001) The continuing epidemics of obesity and diabetes in the United States, *JAMA* **286**, 1195–1200.

Moller, D.E. (2000) Potential role of TNF-alpha in the pathogenesis of insulin resistance and type 2 diabetes, *Trends Endocrinol. Metab.* **11**, 212–217.

Mori, Y., Murakawa, Y., Okada, K., Horikoshi, H., Yokoyama, J., et al. (1999) Effect of troglitazone on body fat distribution in type 2 diabetic patients, *Diabetes Care* **22**, 908–912.

Ochi, M., Furukawa, H., Yoshioka, H., Sawada, T., Kusunoki, T., Hattori, T. (1991) Adipocyte dynamics in hypothalamic obese mice during food deprivation and refeeding, *J. Nutr. Sci. Vitaminol. (Tokyo)* **37**, 479–491.

Ochi, M., Sawada, T., Hattori, T. (1987) Tritiated thymidine autoradiographic study on postnatal development of epididymal adipose tissue in the normal mouse, *Anat. Embryol. (Berl)* **177**, 139–145.

Okuno, A., Tamemoto, H., Tobe, K., Ueki, K., Mori, Y., et al. (1998) Troglitazone increases the number of small adipocytes without the change of white adipose tissue mass in obese Zucker rats, *J. Clin. Invest.* **101**, 1354–1361.

Paolisso, G., Howard, B.V. (1998) Role of non-esterified fatty acids in the pathogenesis of type 2 diabetes mellitus, *Diabet. Med.* **15**, 360–366.

Picard, F., Auwerx, J. (2002) PPAR(gamma) and glucose homeostasis, *Annu. Rev. Nutr.* **22**, 167–197.

Pilgrim, C. (1971) DNA synthesis and differentiation in developing white adipose tissue, *Dev. Biol.* **26**, 69–76.

Poissonnet, C.M., Burdi, A.R., Bookstein, F.L. (1983) Growth and development of human adipose tissue during early gestation, *Early Hum. Dev.* **8**, 1–11.

Poissonnet, C.M., Burdi, A.R., Garn, S.M. (1984) The chronology of adipose tissue appearance and distribution in the human fetus, *Early Hum. Dev.* **10**, 1–11.

Ranallo, R.F., Rhodes, E.C. (1998) Lipid metabolism during exercise, *Sports Med.* **26**, 29–42.

Reitman, M.L. (2002) Metabolic lessons from genetically lean mice, *Annu. Rev. Nutr.* **22**, 459–482.

Ren, D., Collingwood, T.N., Rebar, E.J., Wolffe, A.P., Camp, H.S. (2002) PPARgamma knockdown by engineered transcription factors: exogenous PPARgamma2 but not PPARgamma1 reactivates adipogenesis, *Genes Dev.* **16**, 27–32.

Rosen, E.D., Spiegelman, B.M. (2000) Molecular regulation of adipogenesis, *Annu. Rev. Cell. Dev. Biol.* **16**, 145–171.

Scheja, L., Makowski, L., Uysal, K.T., Wiesbrock, S.M., Shimshek, D.R., et al. (1999) Altered insulin secretion associated with reduced lipolytic efficiency in aP2-/- mice, *Diabetes* **48**, 1987–1994.

Schling, P., Loffler, G. (2002) Cross talk between adipose tissue cells: impact on pathophysiology, *News Physiol. Sci.* **17**, 99–104.

Scow, R.O., Blanchette-Mackie, E.J. (1991) Transport of fatty acids and monoacylglycerols in white and brown adipose tissues, *Brain Res. Bull.* **27**, 487–491.

Shen, W.J., Sridhar, K., Bernlohr, D.A., Kraemer, F.B. (1999) Interaction of rat hormone-sensitive lipase with adipocyte lipid-binding protein, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5528–5532.

Spiegelman, B.M. (1998) PPAR-gamma: adipogenic regulator and thiazolidinedione receptor, *Diabetes* **47**, 507–514.

Stahl, A., Gimeno, R.E., Tartaglia, L.A., Lodish, H.F. (2001) Fatty acid transport proteins: a current view of a growing family, *Trends Endocrinol. Metab.* **12**, 266–273.

Steppan, C.M., Lazar, M.A. (2002) Resistin and obesity-associated insulin resistance, *Trends Endocrinol. Metab.* **13**, 18–23.

Tanaka, T., Yoshida, N., Kishimoto, T., Akira, S. (1997) Defective adipocyte differentiation in mice lacking the C/EBPbeta and/or C/EBPdelta gene, *EMBO J.* **16**, 7432–7443.

Tauchi-Sato, K., Ozeki, S., Houjou, T., Taguchi, R., Fujimoto, T. (2002) The surface of lipid droplets is a phospholipid monolayer with a unique fatty acid composition, *J. Biol. Chem.* **277**, 44507–44512.

Tsao, T.S., Lodish, H.F., Fruebis, J. (2002) ACRP30, a new hormone controlling fat and glucose metabolism, *Eur. J. Pharmacol.* **440**, 213–221.

Unger, R.H. (2002) Lipotoxic diseases, *Annu. Rev. Med.* **53**, 319–336.

Wang, S.P., Laurin, N., Himms-Hagen, J., Rudnicki, M.A., Levy, E., et al. (2001) The adipose tissue phenotype of hormone-sensitive lipase deficiency in mice, *Obes. Res.* **9**, 119–128.

Way, J.M., Gorgun, C.Z., Tong, Q., Uysal, K.T., Brown, K.K., et al. (2001) Adipose tissue resistin expression is severely suppressed in obesity and stimulated by peroxisome proliferator-activated receptor gamma agonists, *J. Biol. Chem.* **276**, 25651–25653.

Yamauchi, T., Kamon, J., Waki, H., Murakami, K., Motojima, K., et al. (2001) The mechanisms by which both heterozygous peroxisome proliferator- activated receptor gamma (PPARgamma) deficiency and PPARgamma agonist improve insulin resistance, *J. Biol. Chem.* **276**, 41245–41254.

Yamauchi, T., Kamon, J., Waki, H., Terauchi, Y., Kubota, N., et al. (2001) The fat-derived hormone adiponectin reverses insulin resistance associated with both lipoatrophy and obesity, *Nat. Med.* **7**, 941–946.

Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L., Friedman, J.M. (1994) Positional cloning of the mouse obese gene and its human homologue, *Nature* **372**, 425–432.

Aggregation, Protein

Jeannine M. Yon
Université de Paris-sud, Orsay, France

1 Introduction 25

2 Protein Folding, Misfolding, and Aggregation 25
2.1 The New View of Protein Folding 25
2.2 Detection of Aggregates during the Refolding Process 27
2.2.1 Transient Aggregation 27
2.2.2 Irreversible Aggregation 29
2.3 Mechanisms of Protein Aggregation 30

3 Protein Folding in the Cellular Environment 34
3.1 Molecular Crowding in the Cells 34
3.2 The Role of Molecular Chaperones 34

4 Protein Aggregation in the Cellular Environment 40
4.1 The Formation of Inclusion Bodies 40
4.1.1 Occurrence of Inclusion Bodies 40
4.1.2 Characteristics of Inclusion Bodies 41
4.1.3 Strategies for Refolding Inclusion Body Proteins 43
4.2 The Formation of Amyloid Fibrils and its Pathological Consequences 45

Bibliography 49
Books and Reviews 49
References of Primary Literature 49

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Keywords

Aggregates
The association of nonnative protein molecules through intermolecular hydrophobic interactions.

Amyloid Fibrils
Ordered aggregates.

Inclusion Bodies
Insoluble, amorphous, disordered aggregates.

Molecular Chaperones
Proteins that assist protein folding within cells.

Protein Folding
The process by which polypeptide chains acquire their three-dimensional and functional structure.

Three dimensional–Domain Swapping
An aggregation mechanism in which one domain in a multidomain protein is swapped with the same domain of another molecule.

Protein misfolding and aggregation are frequent phenomena that occur under different conditions *in vivo* as well as *in vitro*. Aggregation is a serious problem affecting both the production of proteins in the biotechnology and pharmaceutical industries and human health. The aggregates are formed from nonnative proteins through intermolecular interactions that compete with intramolecular interactions. There is thus a kinetic competition between proper folding and misfolding, which can generate aggregates.

Recent evidence for transient association of intermediates during *in vitro* refolding has been obtained for several monomeric proteins. Irreversible and insoluble aggregates are formed in an off-pathway folding process; their formation is concentration dependent and could be prevented by using very small protein concentrations. These aggregates can dissociate and dissolve only in the presence of high concentrations of denaturant. The mechanisms involved in these aggregation processes will be discussed in light of the so-called *new view* of protein folding.

The environmental conditions within cells are markedly different from those used in *in vitro* refolding studies. In the production of recombinant proteins in foreign hosts, the formation of disordered aggregates, that is, inclusion bodies, is often observed. However, aggregation can also result in the formation of amyloid fibrils, which are ordered aggregates. These amyloid formations are at the origin of serious diseases.

1 Introduction

Protein misfolding and aggregation have been recognized for many years as common processes. Aggregation can occur under various conditions. The aggregation of which we are speaking is very different from the precipitation of a native protein at the isoelectric point or upon salting out, which can be reversed under appropriate conditions. In the precipitate, the protein remains in a native conformation. The aggregates, however, are formed from partially folded intermediates and result from intermolecular interactions, which compete with intramolecular interactions. Thermal denaturation of proteins is frequently accompanied by the formation of aggregates leading to the irreversibility of the process. As early as 1931, Wu, in a review on protein denaturation, distinguished between aggregation and precipitation. The aggregated species are not in equilibrium with the soluble species, complicating experimental approaches.

Aggregation has been reported to occur during the *in vitro* refolding of monomeric as well as oligomeric proteins, lowering the refolding yield. As mentioned above, the use of very low protein concentrations could prevent protein aggregation. However, during the folding of nascent polypeptide chains biosynthesized within prokaryotic and eukaryotic cells, aggregates can accumulate. The overexpression of genes in foreign hosts often result in aggregated nonnative proteins called *inclusion bodies*, which are disordered aggregates, leading to serious limitation in the production of recombinant proteins. It is a real problem needing a lot of effort to fully exploit the sequence information contained in the genome projects. Ordered aggregates resulting in amyloid fibrils lead to a number of serious human diseases such as Alzheimer's disease and the transmissible spongiform encephalopathies. The formation of amyloid aggregates has also been reported in *in vitro* experiments.

Experimental and theoretical studies together have provided significant insights into the mechanisms of protein folding, also allowing a better understanding of the aggregation processes.

The following different aspects of protein aggregation must be considered:

1. Theoretical and methodological aspects of protein folding, misfolding, and aggregation including the detection of aggregates and the mechanisms of aggregation processes.
2. Protein aggregation in the cellular environment including the folding into the cell, the role of molecular chaperones, and the formation of different aggregate morphologies, as well as the pathological consequences.

2 Protein Folding, Misfolding, and Aggregation

2.1 The New View of Protein Folding

The question of the mechanisms of protein folding has intrigued scientists for many decades. As early as the 1930s, attempts to refold denatured proteins were published, but significant progress began to be made when Anfinsen successfully refolded, denatured, and reduced ribonuclease into the fully active enzyme. In 1973, he stated the fundamental principle of protein folding referred to as the Anfinsen postulate: ''all the information necessary

to achieve the native conformation of a protein in a given environment is contained in its amino acid sequence.'' The thermodynamic control of protein folding was considered to be a corollary of the Anfinsen postulate, meaning that the native structure is at a minimum of the Gibbs free energy. This statement was discussed by Levinthal in a consideration of the short time required for the folding process *in vitro* as well as *in vivo*. It was concluded that a random search of the native conformation among all possible ones would require an astronomic time and is therefore unrealistic. Thus, it is clear that evolution has found an effective solution to this combinatorial problem. This is referred to as the Levinthal paradox and has dominated discussions for the last three decades.

In order to understand how the polypeptide chain could overcome the Levinthal paradox, different folding models were proposed and submitted to experimental tests. Kinetic studies were carried out to follow the folding pathway. A considerable number of experiments were performed to detect and characterize the folding intermediates. A stepwise sequential and hierarchical folding process in which several stretches of structure are formed and assembled at different levels following a unique route was supported by a majority of scientists for many years. According to this view, misfolded species could be formed from folding intermediates leading to the formation of aggregates in a kinetic competition with the correct folding.

Progressively, with the development of computers, theoretical studies have approached the folding problem, using simplified models to take into account the computational limitations in simulations of the folding from the random coil to the native structure. Different methods were developed using either lattice models or molecular dynamics simulations. In the lattice model, the polypeptide chain is represented as a string of beads on a two-dimensional square lattice or on a three-dimensional cubic lattice. The interactions between residues (the beads) provide the energy function for Monte Carlo simulations. In such simplified models, the essential features of proteins, that is, the heterogeneous character (hydrophobic or polar) of the interactions and the existence of long-range interactions, were included to explore the general characteristics of the possible folds. Lattice models were first applied to protein folding by Go and coworkers while simple exact models were initiated by Dill and his group, and have been used by several theoreticians. From the lattice simulations, insights into possible folding scenarios have been obtained, providing a basis for exploring the general characteristics of folding for real proteins. The exploration of such models supplies useful information that can be submitted to experimental tests.

The so-called ''new view'' has evolved during the past 10 years from both experiment and theory with the use of simplified models. It is illustrated by the metaphor of the folding funnel introduced in 1995 by Wolynes and coworkers. The model is represented in terms of an energy landscape and describes the thermodynamic and kinetic behavior of the transformation of an ensemble of unfolded molecules to a predominantly native state as illustrated in Fig. 1. According to this model, there are several micropathways, each individual polypeptide chain following its own route. Toward the bottom of the funnel, the number of protein conformations decreases as does the protein entropy. The steeper the slope, the faster the folding. As written by Wolynes et al., ''To fold, a protein navigates with remarkable ease

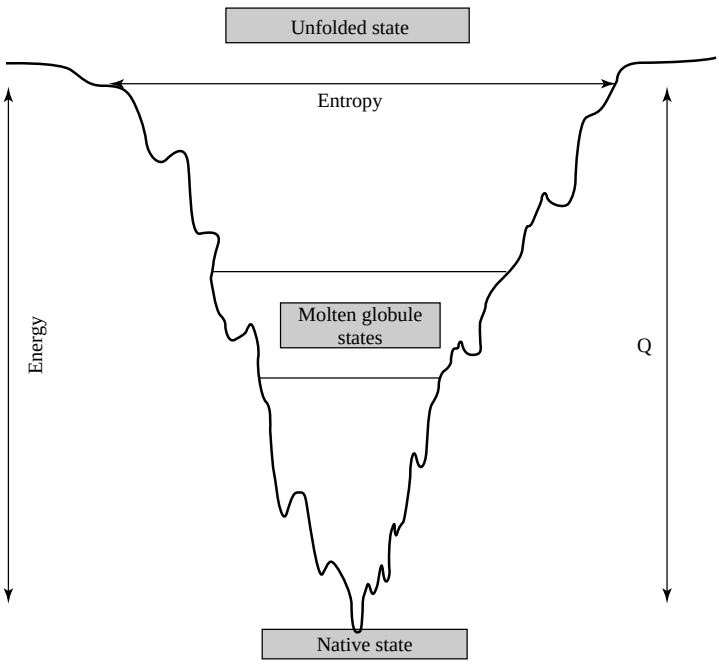

Fig. 1 Schematic representation of the folding funnel. Q is the number of native interactions.

through a complicated energy landscape.'' Thus, a wide variety of folding behaviors emerge from the energy landscape, depending on the energetic parameters and conditions. The folding rate could be slowed by ripples in the energy landscape corresponding to local minima populated by transiently stable intermediates. In a rugged energy landscape with kinetic traps formed by energy barriers, the folding will be even slower. When local energy barriers are high enough, protein molecules could be trapped and possibly aggregate.

The new view has progressively replaced the classical one of a unique sequential pathway and is now quite generally accepted. It is similar to the jigsaw puzzle model proposed in 1986 by Harrison and Durbin, suggesting the possibility of multiple folding routes to reach a unique solution. Many experimental results are consistent with this view. There is an increasing amount of evidence showing that the extended polypeptide chain folds through a heterogeneous population of partially folded intermediates in fluctuating equilibrium. Several alternative folding pathways have been observed for different proteins. From the convergence of theoretical and experimental studies, a unified view of the folding process has progressively emerged, also providing an explanation for the aggregation processes.

2.2 Detection of Aggregates during the Refolding Process

2.2.1 Transient Aggregation

Several observations indicate that transient aggregation could occur during *in vitro* protein refolding. Direct evidence for the

transient association of intermediates has been obtained from small angle X-ray scattering, in the case of apomyoglobin by Doniach and his group, and in the case of carbonic anhydrase by Semisotnov and Kuwajima, and by Silow et al. During the refolding of phosphoglycerate kinase, rapidly transient multimeric species (dimers, trimers, and tetramers) yielding to the native monomeric protein have been detected by Pecorari et al. These species are not in equilibrium, but are formed rapidly and disappear in the slow folding step. Unlike classical aggregates, their distribution does not depend on protein concentration, and they are produced at concentrations as low as 0.05 μM. The distribution of the oligomers is completely established at the end of the fast refolding step. To take into account all these observations, a model, which is formally similar to a reaction of copolymerization between two types of monomers, has been proposed. In this model, the refolding of the protein produces two types of intermediate conformers that can associate with the same or the other type. In the latter case, the association cannot be extended further (Fig. 2). Transient multimeric species have also been observed during the refolding of the isolated N-terminal domain under conditions in which neither the whole native protein nor the folded isolated N-domain associate. However, they cannot transform to the native form in the absence of the interactions with the complementary

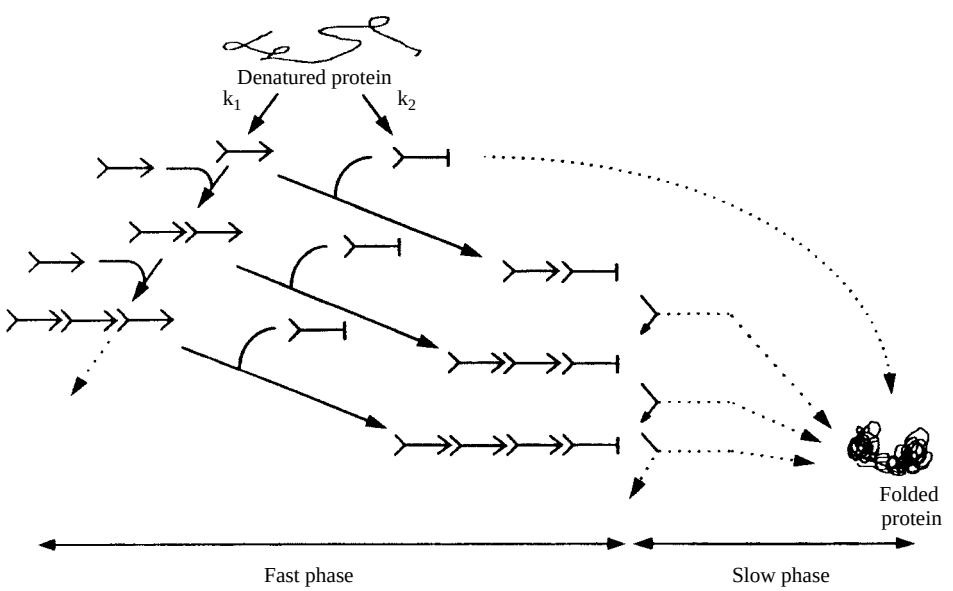

Fig. 2 Model proposed for the formation of transient multimeric species during the refolding of yeast phosphoglycerate kinase. Two types of conformers are produced in the early step of folding. One of these can be directionally extended by association with either the same conformer or another type of conformer. In this last case, the association cannot be further extended. The distribution of species results from a kinetic competition between two kinetic processes. (Reproduced from Pecorari, F., Minard, P., Desmadril, M.,Yon, J.M. (1996) Occurrence of transient multimeric species during the refolding of a monomeric protein, *J. Biol. Chem.* **271**, 5270–5276.)

domain indicating the importance of long-range interactions in directing the correct folding. Such species have not been observed with the C-terminal fragment. Thus, the occurrence of transient multimeric species arising from partially folded intermediates through hydrophobic interactions does not prevent the correct folding of a monomeric protein.

2.2.2 Irreversible Aggregation

Thermal unfolding of proteins is frequently accompanied by the formation of aggregates and therefore behaves as an irreversible process. It occurs at temperatures that vary widely according to the protein, since the temperature of optimum stability depends on the balance between hydrogen bonds and hydrophobic interactions. Generally, the products of thermal denaturation are not completely unfolded and retain some structured regions. At the end of the thermal transition, the addition of a denaturant such as urea or GdnHCl frequently induces further unfolding.

An apparent irreversibility at a critical concentration of denaturant has been observed during the refolding of monomeric as well as oligomeric proteins. It was reported for the first time by M.Goldberg and coworkers for the refolding of β-galactosidase, and for tryptophanase. It was also observed for a two-domain protein, horse muscle phosphoglycerate kinase by Yon and coworkers. In the latter study, when the enzyme activity was used as a conformational probe of the native structure, an irreversibility was observed for a critical concentration of denaturant equal to 0.7 M $\pm$ 0.1 M GdnHCl, a concentration very close to the end of the transition curve. Such irreversibility was found to be concentration dependent. For protein concentrations higher than 30 μM, restoration of enzyme activity was practically null. The formation of irreversible nonnative species was found to be temperature dependent; it was practically abolished at 4 °C, suggesting that aggregation occurs through hydrophobic interactions. The aggregation also depends on the time of exposure of the protein to the denaturant. When the unfolding–refolding process was observed using structural signals such as fluorescence or circular dichroism, it appeared completely reversible whatever the final denaturant concentration.

Another example is provided by rhodanese, a two-domain monomeric protein. During refolding at low denaturant concentration, an intermediate accumulates with partially structured domains and apolar surfaces exposed to the solvent, leading to the formation of aggregates. The aggregation can be prevented by refolding the protein in the presence of lauryl maltoside.

Most of the examples discussed above are related to multidomain proteins. Another degree of complexity appears in the folding of oligomeric proteins. It is generally accepted that the early steps of the process are practically identical to the folding of monomeric proteins. In the last step, subunit association and subsequent conformational readjustments yield the native and functional oligomeric protein. The correct recognition of subunit interfaces is required to achieve the process. The overall process of the folding of oligomeric proteins was extensively studied by Jaenicke and his coworkers for several enzymes and described in reviews. As with monomeric proteins, the formation of aggregates is concentration dependent. The kinetics of aggregation are complex and multiphasic, indicating that several rate-limiting reactions are involved in the process. In an attempt to characterize these aggregates, it was shown that noncovalent interactions occur between monomeric species with

partially restored secondary structures. The aggregates formed by either heat or pH denaturation can be disrupted in 6 M GdnHCl into monomeric unfolded species and then renatured under optimal conditions to yield an active enzyme. Only strong denaturants such as high concentrations of guanidine hydrochloride are efficient in this disruption process.

The presence of covalent cross-links such as disulfide bridges in a protein molecule can complicate the refolding of the denatured and reduced protein resulting in the formation of incorrect and intramolecular disulfide bridges leading to further aggregation. The first well-documented studies were performed by Anfinsen and his group on the refolding of reduced ribonuclease. The authors showed that the reoxidation of the enzyme produces a great number of species with incorrectly paired disulfide bonds. This scrambled ribonuclease is capable of regaining its native structure in a slow step, a process that is accelerated by the addition of a small quantity of reducing reagent such as β-mercaptoethanol yielding about 100% of active enzyme. The reshuffling of a protein's disulfide bonds takes place through a series of redox equilibria according to either an intramolecular or an intermolecular exchange. To prevent a wrong pairing of half-cystine and further aggregation, the addition of small amounts of reducing reagents or redox mixture is frequently used as investigated by Wetlaufer.

The detection and characterization of aggregates represent an important aspect of folding studies. The aggregation phenomenon can occur without precipitation. Indeed, the degree of association of protein intermediates during folding might be small, depending on the intermolecular interactions, and does not necessarily lead to a visible insolubility. The association state may be determined in several ways. The most common methods, available in any biochemistry laboratory, are gel permeation and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), used both with and without cross-linking. The detection of aggregates can also be monitored by other hydrodynamic methods such as analytical ultracentrifugation or classical light scattering. The latter method also gives information on the size of the aggregates. Quasi-elastic light scattering is a dynamic technique that can be used to determine macromolecule diffusion coefficients as a function of time, that is, to follow the kinetics of aggregation. Neutron scattering can also be used to detect protein aggregates, and mass spectrometry has become a useful tool as well.

2.3 Mechanisms of Protein Aggregation

A substantial body of information supports the idea that protein aggregation arises from partially folded intermediates through hydrophobic interactions. The formation of aggregates has often been considered as a trivial phenomenon, a nonspecific association of partially folded polypeptide chains to form a disordered precipitate. However, several analyses indicate that aggregation occurs by specific intramolecular associations involving the recognition of a sequence partner in another molecule rather than in the same molecule during the folding process. Analyses of the aggregation mechanisms of various proteins, such as bovine growth hormone and phosphoglycerate kinase, has permitted the identification of specific sites that are critical in the association.

An elegant demonstration of the specificity of aggregation was provided by King

and coworkers. The authors showed that during the *in vitro* refolding of a mixture of two proteins, tailspike endorhamnidase and coat protein from phage P22, no heterogeneous aggregates were formed. Tailspike endorhamnidase is a thermostable trimer whose folding intermediates are thermolabile and either undergo productive folding or form multimeric aggregates (Fig. 3). The P22 coat protein, which comprises the capsid shell of phage P22, yields either a correct fold or "off-pathway" aggregates upon refolding. Both proteins were intensively studied by King and coworkers who first denatured the two proteins in urea and then chose refolding conditions such that aggregation competes with correct folding. Folding and soluble aggregates of the two proteins were characterized either separately or mixed together. No heterogeneous aggregates were found, clearly indicating that only self-association of transient refolding molecules occurs in the formation of soluble multimers.

One mechanism that accounts for the formation of aggregates during refolding of multidomain proteins is domain swapping. This was first suggested by Monod and later proposed by Goldberg and colleagues to account for the formation of aggregates during the refolding of tryptophanase. The concept was foreshadowed by the results of Crestfield and coworkers in 1962. From their experiments based on chemical modification of bovine pancreatic ribonuclease, the authors proposed that the dimer is formed by exchanging the N-terminal fragments. The term *3D domain swapping* was introduced in 1994 by Bennett and coworkers to describe the structure of a diphtheria toxin dimer. The mechanism involves the replacement of one domain of a monomeric protein by the same domain of an identical neighboring

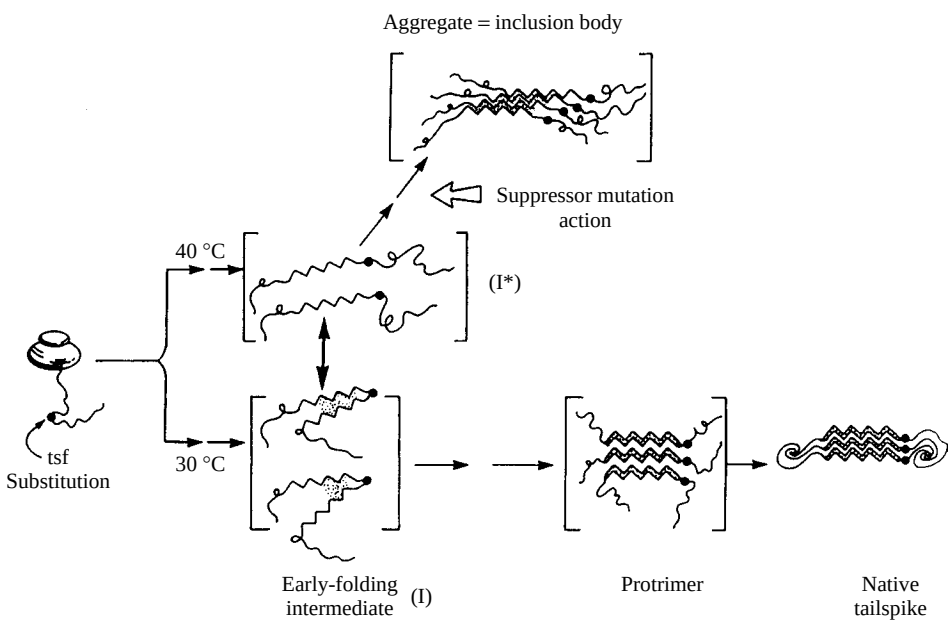

Fig. 3 The folding pathway of the P22 tailspike protein. (From Mitraki, A., King, J. (1992) *FEBS Lett.* **307**, 20–25; reproduced with permission.)

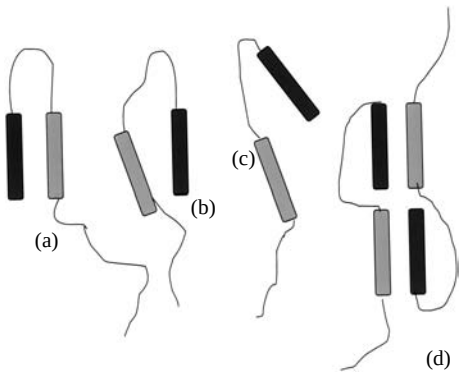

Fig. 4 Schematic representation of domain swapping. (a) monomeric protein, (b) and (c) partially unfolded monomers, (d) domain-swapped dimer.

molecule, thus resulting in an intertwined dimer or oligomer, as defined by Eisenberg and colleagues (Fig. 4). When the exchange is reciprocated, domain-swapped dimers are formed. However, if the exchange is not reciprocated but propagated along multiple polypeptide chains, higher order assemblies or aggregates may form. Domain-swapped oligomers are divided into two types, open and closed. The open oligomers are linear and have one closed interface (closed in the monomer) exposed to the solvent, whereas closed oligomers are cyclic and do not expose a closed interface. Eisenberg and coworkers have defined the structure of the monomer as the "closed monomer" and the conformation of the polypeptide chain in the domain-swapped oligomer as the "open monomer."

The ability of monomeric proteins to swap structural elements requires the presence of a hinge or linker region that permits the protein to attain the native fold with parts of two polypeptide chains. In fact, domain-swapped structures reveal regions of protein structure that are flexible. Bergdoll and coworkers have suggested that a proline in the linker region, by rigidifying the hinge region in intermediate states, might facilitate domain swapping. Baker and colleagues proposed that strain in a hairpin loop might predispose a protein to domain swapping. The possible role of 3D domain swapping in the evolution of oligomeric proteins has been discussed in several reviews. In the past years, the number of known domain-swapped proteins has increased and today about 40 such structures are solved. One common feature of these proteins is that all the swapped domains are from either the N-terminus or the C-terminus of the polypeptide chain. In this regard, an interesting example arises from the work of Eisenberg and his group on the dimerization of ribonuclease A. This protein forms two types of dimers upon concentration in mild acid. The minor dimer is formed by swapping of its N-terminal α-helix with that of an identical molecule. The major dimer results from the swapping of its C-terminal β-strand. RNase A was also reported to form trimers. On the basis of the structure of the N- and C-terminal swapped dimers, a model was proposed (Fig. 5). This indicates that two types of swapping can occur simultaneously in the same oligomer. Further biochemical studies have supported this model. A less abundant trimer in which only the C-terminal β-strand is swapped and exhibits a cyclic structure was also found. RNase represents the

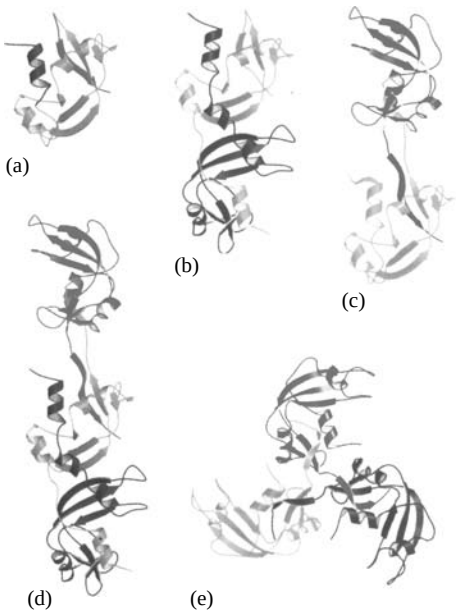

Fig. 5 Domain swapping in ribonuclease. Ribbon diagram of the structures of (a) the ribonuclease A monomer (2.0 Å), (b) the N-terminal swapped dimer (2.1 Å), (c) the C-terminal swapped dimer (1.75 Å), (d) the N- and C-terminal trimer model, and (e) the cyclic C-terminal swapped trimer (2.2 Å) (reproduced from Liu et al. Prot. Sci. **11**, 371, 2002 with permission).

first protein found to form both linear and cyclic domain-swapped oligomers. This protein also was described to form tetramers. Models based on the structures of dimers and trimers were proposed for these tetramers. Two linear models exhibit both types of swapping that occur in one molecule, and a cyclic tetramer shows the swapping of the C-terminal β-strand only. A trimeric domain-swapped barnase was obtained at low pH and high protein concentration. Crystallographic studies revealed a structure suggesting a probable folding intermediate. Domain swapping was described for the cell cycle regulatory protein p13suc1, a small protein of 113 amino acids.

Folding studies as well as molecular dynamics simulations have shown that domain swapping occurs in the unfolded state. Eisenberg and his colleagues have proposed a free energy diagram for the pathway of domain swapping. The free energy difference between the closed monomer and domain-swapped oligomer is small since they share the same structures except at the hinge loop, but the energy barrier can be reduced under certain conditions making domain swapping more favorable. Several molecular or environmental events may favor the formation of extended domain-swapped polymers. Genetic mutations introducing a deletion in the hinge loop can destabilize the monomeric form of a protein. The replacement of only one amino acid can also favor the polymerization of the mutated protein. Three-dimensional domain-swapped oligomers are expected to be increasingly favored as the protein concentration increases. Thus, a metabolic change that increases the concentration of a protein will favor aggregation. Charge effects, caused either by mutations or by pH change or salt concentration, can induce domain swapping; for example, in RNase A, a decrease in pH, by protonating the residues involved in hydrogen bonds and in salt bridges, lowers the energy barrier of the formation of the open monomer, hence inducing domain swapping.

There is great diversity of swapped domains, with different sizes and sequences. They can consist of entire tertiary domains or smaller structural elements made of several residues. No specific sequence motif seems to be involved among the swapped domains. Three-dimensional domain swapping has also been proposed as a mechanism for amyloid formation. This aspect will be discussed in Sect. 4.2.

As can be seen here, several mechanisms exist, which lead to the formation of aggregates. It is recognized that aggregation results from the association of incompletely or incorrectly folded intermediates through hydrophobic interactions. In the energy landscape of protein folding, the presence of local minima separated by an energy barrier allows the accumulation of intermediates. If the barrier is high enough, these intermediates cannot easily reach the native state, and kinetic competition thus favors the formation of aggregates.

3
Protein Folding in the Cellular Environment

3.1
Molecular Crowding in the Cells

The main rules of protein folding have been deduced from a considerable body of *in vitro* and *in silico* studies. It has been accepted that the same mechanisms are involved in *in vitro* refolding and in the folding of a nascent polypeptide chain in the cell. However, the intracellular environment differs markedly from that of the test tube where low protein concentrations are used. The interior of a cell is highly crowded with macromolecules. The concentration is so high that a significant proportion of the volume is occupied. As mentioned by Ellis, in general, 20 to 30% volume of the interior of the cells are occupied by macromolecules; for example, the concentration of total protein inside cells ranges from 200 to 300 g L^{-1}. The total concentration of proteins and RNA inside *Escherichia coli* ranges from 300 to 400 g L^{-1} depending on the growth phase. Polysaccharides also contribute to the crowding. It can be predicted practically that diffusion coefficients will be reduced by factors up to 10-fold due to crowding. Since the average time for a molecule to move a certain distance varies by D^{-2}, D being the diffusion coefficient, it will take 100 times longer to move this distance in the cell as would be necessary under low concentration conditions. Another prediction indicates that equilibrium constants for macromolecular associations may be increased by two to three orders of magnitude.

Molecular crowding inside cells also has consequences for protein folding, favoring the association of partly folded polypeptide chains into aggregates. This could explain why cells contain molecular chaperones, even though most denatured proteins refold spontaneously in the test tube.

3.2
The Role of Molecular Chaperones

The discovery of a ubiquitous class of proteins mediating the correct folding in cellular environment has led to a reconsideration of the mechanism of protein folding *in vivo*. Historically, the term *molecular chaperone* was introduced by Laskyard and coworkers in 1987 to describe the function of nucleoplasmin, which mediates the *in vitro* assembly of nucleosomes from separated histones and DNA. The concept was further extended by Ellis to define a class of proteins whose function is to ensure the correct folding and assembly of proteins through a transient association with the nascent polypeptide chain. Studies on heat-shock proteins have widely contributed to the development of this concept.

Today, more than 20 protein families have been identified as molecular chaperones. Molecular chaperones comprise

several highly conserved families of related proteins. They can be divided into two classes according to their size. Small chaperones are less than 200 kDa, whereas large chaperones are more than 800 kDa. During the past few years, a large amount of biochemical, biophysical, and low- and high-resolution structural data have provided mechanistic insights into the machinery of protein folding as assisted by molecular chaperones.

Molecular chaperones are involved in diverse cellular functions. The constitutive members of the heat-shock protein family (Hsp70) can stabilize nascent polypeptide chains during their elongation in ribosomes. The large cylindrical chaperonins GroEL in bacteria, mitochondria, and chloroplasts and the corresponding TriC in eukaryotes and archaebacteria provide a sequestered environment for productive folding. Several chaperones are stress-dependent; their expression is induced under conditions such as high temperatures, which provoke protein unfolding and aggregation. The members of the Hsp90 and Hsp100 families, as well as small Hsp, play a role in preventing protein aggregation under stress. Chaperone interactions are also important for the translocation of polypeptide chains into membranes.

Within cells, the nascent polypeptide chain is synthesized sequentially on the ribosome by a vectorial process. For many proteins, the rate of this process is slower than the rate of folding. Synthesis times range from 20 s for a 400 residue–polypeptide chain in *E. coli* at 37 °C to 10 times as long for such a chain in an eukaryotic cell. Many unfolded proteins refold completely in 20 s under the same conditions. Thus, there is the possibility for the elongating polypeptide either to misfold before completion or to be degraded by proteolytic enzymes. Chaperones prevent such unfavorable events by protecting the nascent chain. Hsp70 and its prokaryotic homolog DnaK recognize extended hydrophobic regions of the elongating polypeptides. These interactions are not specific. Hsp70 and DnaK interact with most unfolded polypeptide chains that expose hydrophobic residues. They do not recognize folded proteins. Binding and release of unfolded proteins from Hsp70 are ATP-dependent and require the presence of various cochaperones such as DnaJ and GrpE. The basic mechanism of Hsp70 (DnaK in *E. coli*) is represented in Fig. 6. In *E. coli*, DnaJ binds the nascent unfolded polypeptide, U; then the complex binds to the ATP-bound state of DnaK. ATP is hydrolyzed in the ternary complex

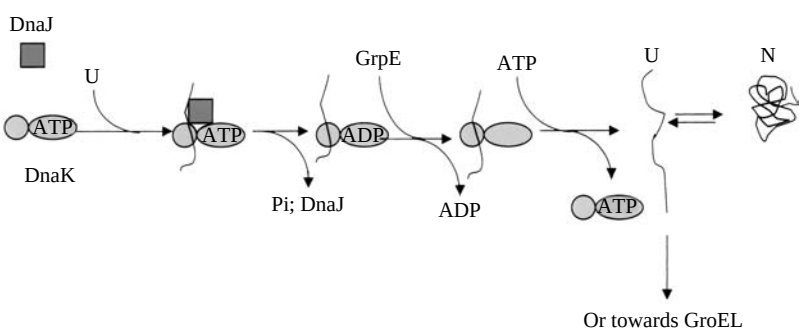

Fig. 6 Schematic representation of the basic mechanism of DnaK (see text).

allowing the release of DnaJ and P_i. In the following step, GrpE acts as an exchange factor to regenerate the ATP-bound state of DnaK. The unfolded polypeptide chain is released into the bulk solution. Thus, Hsp70 systems bind and release the polypeptide in an unfolded conformation. The unfolded protein has the possibility either to fold or to be transferred to the GroEL system, as illustrated in Fig. 6. Significant insights into this mechanism were obtained from structural data. The three-dimensional structures of Hsp70 and DnaJ as well as those of a complex DnaK–polypeptide and a complex of GrpE with the ATP binding domain of DnaK are known. DnaK and its homologs are composed of two domains, a C-terminal domain that binds ATP and an N-terminal domain that binds peptides. GrpE is a tight homodimer associated along two long helices. It binds DnaK–ATPase domain through its proximal monomer. DnaJ activates the ATP hydrolysis by DnaK. It was shown that a conformational change may occur upon ATP binding, opening the polypeptide binding cleft in the polypeptide binding domain of DnaK. The closed state may correspond to the ADP-bound conformation. The ADP-bound state of DnaK binds the peptide tightly. Peptide release requires the dissociation of ADP, which is mediated by GrpE. DnaK then rebinds ATP.

The GroEL–GroES system acts by a different mechanism in which the unfolded protein is sequestered. The chaperonins are large cylindrical protein complexes. The crystal structure of *E. coli* chaperonin GroEL was determined in 1994 and that of the asymmetric GroEL–GroES–(ADP)7 complex in 1997 by Sigler and his group. GroEL consists of two heptameric rings of 58-kDa subunits stacked back to back with a dyad symmetry and forming a porous cylinder (Fig. 7). Each subunit is organized in three structural domains. A large equatorial domain forms the foundation of the assembly and holds the rings together. It contains the nucleotide binding site. A large apical domain forms the end of the cylinder. The apical domain contains a number of hydrophobic

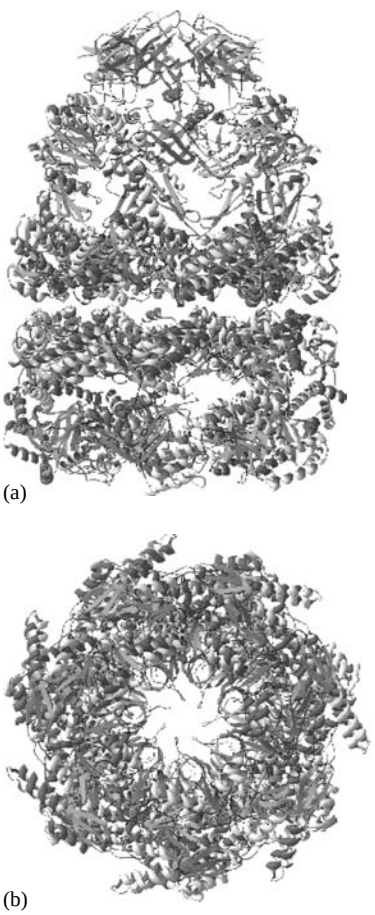

Fig. 7 Crystal structure of GroEL–GroES–(ADP)7 complex determined by Sigler et al. (a) view along the axis and (b) view from the top of the complex. (Reproduced from the PDB web site.)

residues exposed to the solvent. A small intermediate domain connects the two large domains. The intermediate segments have some flexibility allowing a hinge-like opening of the apical domains, which occurs upon nucleotide binding. These movements are large and have been visualized by three-dimensional reconstruction from cryoelectron microscopy by Sebil and her group.

GroES is a heptamer of 10 kDa subunits forming a flexible dome-shaped structure with an internal cavity large enough to accommodate proteins up to 70 kDa. Each subunit is folded into a single domain containing β-sheets and flexible loop regions. The loop regions are critical for the interactions between GroEL and GroES. It was deduced from electron microscopy studies that GroES binding to GroEL induces large movements in the apical GroEL domains. This provokes a significant increase in the volume of the central cavity in which protein folding proceeds. NMR coupled with the study of hydrogen-exchange techniques has indicated that small proteins are essentially unfolded in their GroEL-bound states. Mass spectroscopy has revealed the presence of fluctuating elements of secondary structure for several proteins. In a way, the GroEL–GroES system recognizes nonnative proteins.

The reaction cycle of the GroEL–GroES system is represented in Fig. 8. The nonnative protein binds to the apical domain of the upper ring of GroEL through hydrophobic interactions. Then, the equatorial domain of the same ring binds ATP, and GroES caps the upper ring, sequestering the protein inside the internal chamber in which the protein folding proceeds. The binding of GroES induces a conformational change in GroEL and ATP hydrolysis, which is a cooperative process that produces a conformational change in the lower ring, allowing it to bind a nonnative protein molecule. This promotes subsequent binding of ATP and GroES in the lower ring, and the dissociation of the upper complex, releasing the protein and ejecting GroES. If the protein has not reached the native state, it is subjected to a new cycle.

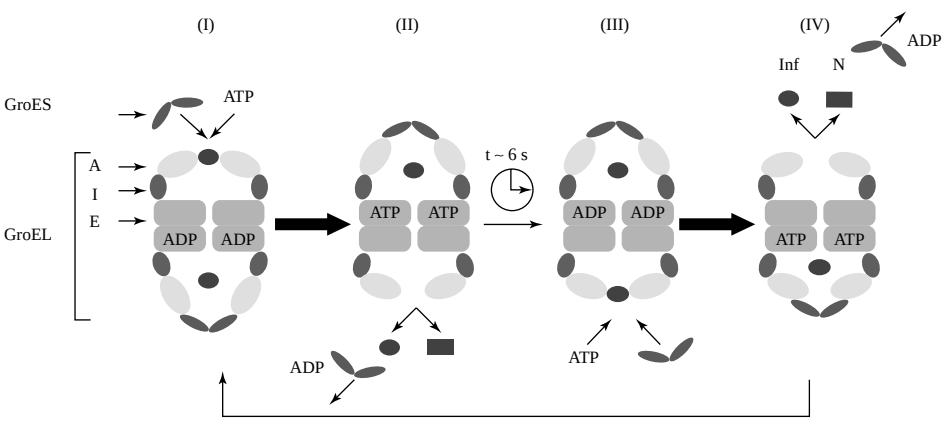

Fig. 8 The reaction cycle of GroEL–GroES. Inf is the unfolded protein, N the folded one, A is the apical domain, (in blue), I the intermediate domain (in red) and E the equatorial domain (in magenta). (Reproduced from Wang & Weissman (1999) *Nat. Struct. Biol.* **6**, 597, with permission.) (See color plate p. xxii).

The hydrolysis of ATP by GroEL is used only to induce conformational changes of the chaperone, which permits the release of the folded protein. The molecular chaperones, by their transient association through hydrophobic interactions with nascent, stress-destabilized, or translocated proteins, have a role in preventing improper folding and subsequent aggregation. They do not interact with folded proteins. They do not carry information capable of directing a protein to assume a structure different from that dictated by its amino acid sequence. Therefore, molecular chaperones assist the folding in the cells without violation of the Anfinsen postulate. They increase the yield but not the rate of folding reactions; in this respect they do not act as catalysts. Furthermore, the majority of newly synthesized polypeptide chains in both bacterial and eukaryotic cells fold spontaneously without the assistance of molecular chaperones.

Many proteins from prokaryotic and eukaryotic organisms are produced with an amino-terminal propeptide, which is removed by limited proteolysis during the activation process. Several of these propeptides consist of a long polypeptide chain; for example, there are 174 amino acids in the propeptide of pro-α-lytic protease, 91 in that of procarboxypeptidase Y, and 77 in that of prosubtilisin. Several studies have shown that the propeptide is required for proper folding of these proteins. The mature enzymes are not able to refold correctly. They seem to have kinetic stability only, whereas the proenzymes have thermodynamic stability. Since propeptides perform the function of mediating protein folding, they have been classified as intramolecular chaperones. However, this terminology is not appropriate since the nascent protein is the proenzyme, not the enzyme that has undergone proteolytic cleavage. Thus, it is not surprising that the proenzyme refolds spontaneously, whereas the mature protein does not. Indeed, the information is contained in the totality of the proenzyme sequence.

Two other classes of proteins play the role of helpers during protein folding *in vivo*: protein disulfide isomerases (PDIs) and peptidyl–prolyl cis – trans isomerases. Protein disulfide isomerase is an abundant component of the lumen of the endoplasmic reticulum in secretory cells. The enzyme was discovered independently in 1963 by two research groups: in rat and ox by Anfinsen and coworkers, and in chicken and pigeon pancreas by Straub and coworkers. Proteins destined to be secreted enter the endoplasmic reticulum in an unfolded state. In this environment, the folding process is associated with the formation of disulfide bonds, which is catalyzed by PDI through thiol–disulfide interchange. The first PDI cDNA was sequenced in 1985 by Edman et al. It displays sequence homologies implying a multidomain architecture. PDI consists of four structural domains arranged in the order a, b, b′, a′, with the b′ and a′ domains being connected by a linker region. Furthermore, it possesses an acidic C-terminal extension. The a and a′ domains contain the active site motif – W-C-G-H-C-. They display significant sequence identity to thioredoxin, a small cytoplasmic protein involved in several redox functions, and they have a similar active site sequence.

Recombinants of the a and b domains have been obtained and studied by high-resolution NMR. The a domain has the same overall fold as thioredoxin, an α/β fold with a central core made up of a five-stranded β-sheet surrounded by four helices. As in thioredoxin, the active site is located at the N-terminus of helix

α_2. Preliminary NMR data of the a′ domain confirm its structural similarity to the a domain. The b and b′ domains have significant sequence similarity to each other, but no similarity with the a domain. Nevertheless, NMR studies of the b domain have indicated a similar overall fold. From its sequence, it could be inferred that b′ also has the same fold. Neither b nor b′ contain the active site.

The folding pathway of disulfide-bound proteins involves isomerizations between a number of species containing disulfide bonds. *In vitro* experimental studies were performed using the isolated a and a′ domains, and the results were compared with those obtained with the holoenzyme. It was concluded that the activity of long length PDI is not simply the sum of the activities of the isolated a and a′ domains. Using a series of constructs including nearly every linear combination of domains, the contribution of each domain was investigated. It was determined that the thiol-disulfide chemistry requires only the a and a′ domains, and that simple isomerization requires one of these in a linear combination including b′, whereas complex isomerization involving large conformational changes requires all the PDI domains except the C-terminal extension. Thus, it appears that the b′ domain is the principal peptide binding site, but all domains contribute to the binding of larger polypeptide chains holding them in a partially unfolded conformation while the catalytic sites acts synergistically to perform the thiol-disulfide exchange. Since PDI has binding properties, it has been proposed that it acts as a molecular chaperone. However, as underlined by Freedman and coworkers, this property does not represent a chaperone activity and instead reflects its role as a catalyst to accelerate the formation of native disulfide bridges during protein folding.

Several gene products with similarity to PDI have been identified in higher eukaryotes. All are probably localized in the endoplasmic reticulum and have thiol-disulfide exchange activity.

In prokaryotes, the disulfide formation occurs in the periplasm and is catalyzed by a protein called DsbA, which exchanges its Cys30–Cys33 to a pair of thiols in the target protein, leaving DsbA in its reduced state. The crystal structure of oxidized DsbA displays a domain with a thioredoxin-like fold and another domain, which caps the thioredoxin-like active site C30-P31-H32-C33, located at the domain interface. Reoxidation of DsbA is catalyzed by a cytoplasmic membrane protein called DsbB, which contains four cysteine residues essential for catalysis. DsbB transfers the electrons from the reduced DsbA to membrane embedded quinones. The reduced quinones are then oxidized enzymatically either aerobically or anaerobically. Thus, DsbA is found in normal cells in its oxidized state. *E. coli* also has a complex reductive system including another periplasmic protein DsbC, which is a homodimer. The molecule consists of two thioredoxin-like domains with a CxxC motif, joined via hinged linker helices to an N-terminal dimerization domain. The hinge regions allow movement of the active site, and a broad hydrophobic cleft between the two domains may bind the polypeptide chain. Its function consists of reducing proteins with incorrect disulfide bonds. DsbC is maintained in its reduced form by a membrane protein called DsbD, which contains six essential cysteine residues. Then, the electrons are transferred to thioredoxin and ultimately to NADPH by thioredoxin reductase.

All these enzymes, which catalyze the pairing of cysteine residues in disulfide-bridged proteins, have functional domains pertaining to the thioredoxin superstructure.

Another type of enzyme, peptidyl–prolyl cis–trans isomerases, facilitates the folding of some proteins by catalyzing the cis–trans isomerization of X-Pro peptide bonds. Two classes of unrelated proteins demonstrate this activity, those that bind cyclosporin, which are known as *cyclophilins*, and those that bind FK506. The cellular function of these enzymes is important, since cyclosporin and FK506 are potent immunosuppressors that regulate T-cells activation. Both classes of peptidyl–prolyl isomerases are ubiquitous, and abundant in prokaryotes and eukaryotes. The sequences of several members of each family are known, and the three-dimensional structures of at least one member of each family have been elucidated by X-ray crystallography and multidimensional NMR. Their role is to accelerate the cis-trans isomerization of X-pro peptide bonds when this process is the rate-limiting step in protein folding. Although they do not present structural similarity, both exhibit a hydrophobic binding cleft favoring the rotamase activity by excluding water molecules.

4 Protein Aggregation in the Cellular Environment

4.1 The Formation of Inclusion Bodies

The overexpression of genes introduced in foreign hosts frequently results in aggregated nonnative proteins called *inclusion bodies*. In cells, inclusion bodies appear as unordered amorphous aggregates clearly separated from the rest of the cytoplasm; they form a highly refractive area when observed microscopically. A great variety of experimental studies indicates that the formation of inclusion bodies results from partially folded intermediates in the intracellular folding pathway and not from either totally unfolded or native proteins.

4.1.1 Occurrence of Inclusion Bodies

Inclusion bodies were first identified in the blood cells of patients with abnormal hemoglobins, the resulting pathology being anemia. Pathological point mutants of hemoglobin aggregate into inclusion bodies; this is the case for hemoglobin Köln (Val98Met on the β chain) and hemoglobin Sabine (Leu91Pro on the β chain). Similar deposits have been described in studies on the metabolism of abnormal proteins subjected to covalent modification in *E. coli*. The formation of aggregates also occurs when cells are subjected to heat shock.

The *in vivo* folding pathway of tailspike endorhamnosidase of Salmonella phage 22 is a well-documented system studied by J.King's group. Furthermore, it is one of the few systems in which the *in vivo* folding pathway has been compared with the *in vitro* refolding pathway. The protein is a trimer of 666 amino acids. The secondary structure is predominantly β-sheet. Newly synthesized polypeptide chains released from the ribosome generate an early partially folded intermediate. This intermediate further evolves into a species sufficiently structured for chain–chain recognition. In the following step, an incompletely folded trimer is formed upon close association with the latter species. The protrimer is then transformed into the native tailspike. A clear difference between the physicochemical properties of the intermediates and the native state has

allowed their identification. Figure 3 illustrates the folding pathway of the protein. The native protein is highly thermostable with a T_m of 88 °C; it is also resistant to detergents and proteases. During the *in vivo* folding process, the intermediates are sensitive to these factors, allowing their identification. At low temperature, almost 100% of the newly synthesized chains reach the native trimer conformation. When the temperature increases in the cells, the number of polypeptide chains achieving the native state decreases. At 39 °C, the maturation proceeds with 30% efficiency, while the remainder aggregates into inclusion bodies. It has been shown that the aggregation does not result from an intracellular denaturation of the native protein, but is generated from an early thermolabile intermediate. The aggregated chains cannot recover their proper folding by lowering the temperature. But when polypeptide chains that have been synthesized at high temperatures are shifted to low temperature early enough, they can refold correctly.

A set of mutations that alter protein folding without modifying the properties and stability of native P22 tailspike has been identified; they are referred to as temperature-sensitive folding (tsf) mutants. These mutations have been supposed to destabilize the already thermolabile intermediate and are located at more than 30 sites in the central region of the polypeptide chain. Starting from mutants kinetically blocked in their folding, a second set of mutants capable of correcting the folding defects was selected, and the sequences surrounding the suppressor mutations were identified. Only two substitution positions on the 666 amino acids of the polypeptide chain were sufficient to prevent inclusion body formation. Thus, single temperature mutations that affect the folding pathway but not the native conformation of a protein are efficient in preventing off-pathway and subsequent aggregation. A similar result has been found for heterodimeric luciferase. For recombinant proteins such as interferon-γ and interleukin 1β, as well as for P22 tailspike, amino acid substitutions that can decrease or increase the formation of inclusion bodies without alteration of the functional structure were found by Wetzel and coworkers.

The formation of inclusion bodies is frequently observed in the production of recombinant proteins. High levels of expression of these proteins result in the formation of inactive amorphous aggregates, and has been reported for proteins expressed in *E. coli* and also in several host cells, gram-negative as well as gram-positive bacteria, and eukaryotic cells such as *Saccharomyces cerevisiae*, insect cells, and even animal cells. The production of recombinant proteins, among them human insulin, interferon-γ, interleukin 1β, β-lactamase, prochymosin, tissue plasminogen activator, basic fibroblast growth hormone, and somatotropin, gives rise to inclusion bodies.

4.1.2 Characteristics of Inclusion Bodies

Inclusion bodies can form in the cytoplasm and in the periplasmic space of *E. coli*. Wild-type β-lactamase expressed in *E. coli* results in the formation of inclusion bodies in the periplasm, whereas the protein expressed without its signal sequence aggregates in the cytoplasm.

The characteristics of the aggregates depend on how the protein is expressed. Different sizes and morphologies have been observed. Generally, inclusion bodies appear as dense isomorphous aggregates of nonnative proteins separated from the rest of the cytoplasm, but not surrounded

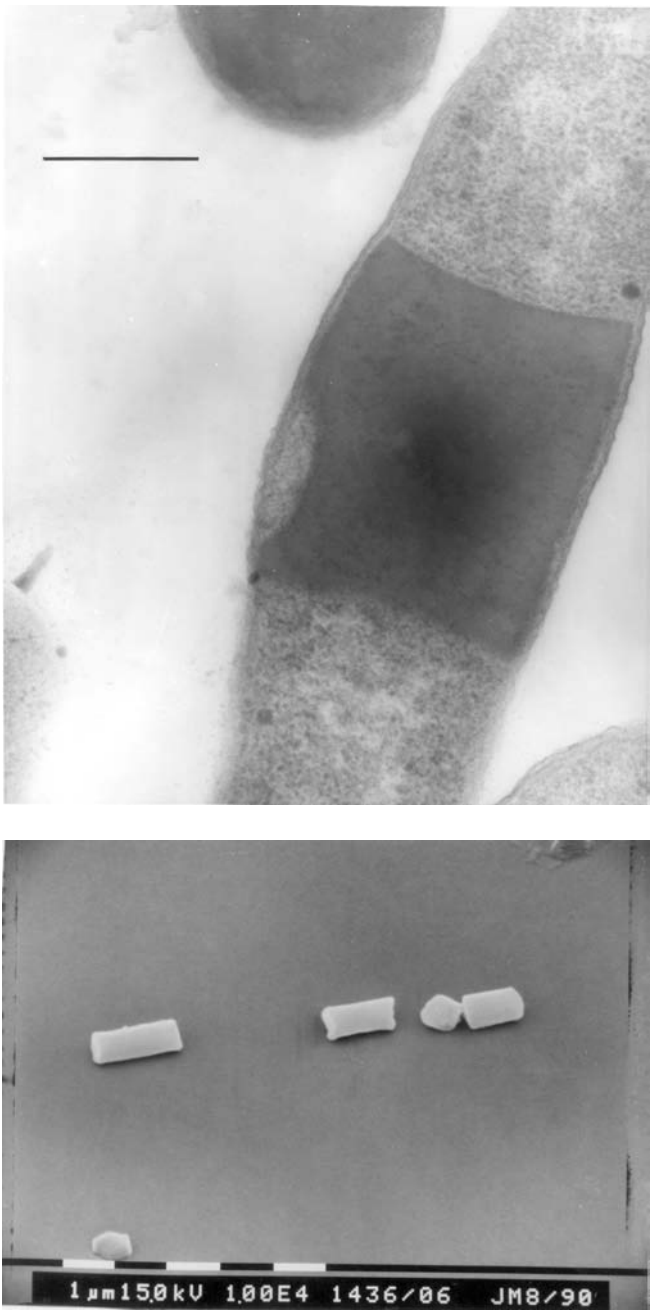

Fig. 9 Electron micrographs of (a) cytoplasmic β-lactamase inclusion bodies in *E. coli* RB791(pGB1) and (b) purified inclusion bodies from the same origin (courtesy of G.A. Bowden, A.M. Paredes & G. Georgiou).

by a membrane (Fig. 9). They look like refractile inclusions, which can be easily recognized by phase contrast microscopy when large enough. For prochymosin expressed in *E. coli*, the lack of birefringence indicates that inclusion bodies are not crystalline. The size distribution of inclusion bodies has been studied for prochymosin and interferon-γ, and Marston reported the mean size of particles to be 0.81 and 1.28 μm respectively, with a relatively high void fraction. The void volume was about 70% of the total volume for interferon-γ and 85% for prochymosin. Structural characterization studies using ATR-FTIR (attenuated total reflectance Fourier transformed infrared spectroscopy) have shown that the insoluble nature of inclusion bodies may be due to their increased levels of nonnative intramolecular β-sheet content.

Inclusion bodies consist mostly of the overexpressed recombinant protein, and can contain little contaminating molecules. Thus, they can be used as a source of relatively pure misfolded protein when refolding yields the active protein. However, some amorphous bodies incorporate other molecules, for example, inclusion bodies from *E. coli* cells overexpressing β-lactamase contain only between 35 and 95% intact β-lactamase. The rest consists of a variety of intracellular proteins, some lipids, and a small amount of nucleic acids. Homogeneous inclusion bodies were obtained by expressing β-lactamase without its leader peptide. Under these conditions, aggregation occurs within the cytoplasm. The extent of incorporation of other macromolecules in inclusion bodies depends upon the overexpressed protein.

The formation of inclusion bodies generally appears to be a disadvantage, since it requires the dissolving of the aggregates in denaturant and subsequent refolding of the protein. However, when the recovery of the active product can be obtained with a sufficient yield, certain advantages may accrue. Indeed, aggregation generally prevents proteolytic attack, except when the protein coaggregates with a protease. The formation of inclusion bodies is also an advantage for the production of proteins that are toxic for the host cells. Furthermore, these aggregates contain a great quantity of the overexpressed protein.

4.1.3 Strategies for Refolding Inclusion Body Proteins

The recovery of the active protein from inclusion bodies is crucial for industrial purposes. In structural proteomics today, efficient production of genetically engineered proteins is a prerequisite for exploiting the information contained in the genome sequences. The strategy to recover active proteins involves several steps of purification. The first step, the separation of the inclusion bodies from the cell, consists of cell lysis monitored either by high-pressure homogeneization, or by a combination of mechanical, chemical, and enzymatic techniques such as the use of EDTA and lysozyme. The lysates are then treated by low-speed centrifugation or filtration to remove the soluble fraction from the pellet containing inclusion bodies and cell debris. The most difficult task is to remove the contaminants; this is achieved by the washing steps, which commonly utilize EDTA and low concentrations of denaturants or detergents such as Triton X-100, deoxycholate, or octylglucoside. Using centrifugation in a sucrose gradient, it is generally possible to remove cell debris and membrane proteins. When the accumulation levels of aggregates are very high, inclusion bodies may be directly solubilized by treatment in a high concentration of denaturant, eliminating the need for gradient centrifugation. In this

case, the costs of production are considerably reduced.

A variety of techniques are available to solubilize purified inclusion bodies. The most commonly used solubilizing reagents are strong denaturants such as guanidine hydrochloride and urea. Generally, high denaturant concentrations are employed, 4 to 6 M for guanidine hydrochloride, and 5 to 10 M for urea to allow the disruption of noncovalent intermolecular interactions. Conditions may differ somewhat according to the denaturant and the protein. Lower denaturant concentrations have been used to solubilize cytokines from *E. coli* inclusion bodies. The purity of the solubilized protein was much higher at 1.5 to 2 M guanidinium chloride than at 4 to 6 M guanidinium chloride. At higher denaturant concentrations, contaminating proteins were also released from the particulate fractions.

Extremes of pH have also been used to solubilize inclusion bodies and for growth hormone, proinsulin, and some antifungal recombinant peptides. However, exposure to very low or very high pH may not be applicable to many proteins and may cause irreversible chemical modifications.

Detergents such as sodium dodecylsulfate (SDS) and n-cetyl trimethylammonium bromide (CTAB), have also been used to solubilize inclusion bodies. Extensive washing may then be needed to remove the solubilizing detergents. They also may be extracted from the refolding mixture by using cyclodextrins, linear dextrins, or cycloamylose. Recent developments include the use of high hydrostatic pressure (1–2 kbar) for solubilization and renaturation. For proteins with disulfide bonds, the addition of a reducing reagent such as dithiothreitol or β-mercaptoethanol is necessary to disrupt the incorrectly paired disulfide bonds. The concentrations generally used are 0.1 M for dithiothreitol and 0.1 to 0.3 M for β-mercaptoethanol.

When expression levels are very high, an *in situ* solubilization method can be used. It consists of adding the solubilizing reagent directly to the cells at the end of the fermentation process. The main disadvantage of this technique concerns the release of contaminants.

The last step is the recovery of the active protein. When inclusion bodies have been solubilized, the refolding is achieved by removal of the denaturant. This can be done by different techniques including dilution, dialysis, diafiltration, gel filtration, chromatography, or immobilization on a solid support. Dilution has been extensively used. It considerably reduces concentrations of both denaturant and protein. This procedure, however, cannot be applied to the commercial scale refolding of recombinant proteins, because large downstream processing volumes increase the cost of products. Although dialysis through semipermeable membranes has been used successfully to refold several proteins, it is not employed in large-scale processes. This is because it requires very long processing times, and there is the risk that during dialysis, the protein will remain too long at a critical concentration of denaturant and aggregate. The removal of the denaturant may be accomplished through gel filtration. However, here again, a possible aggregation could lead to flow restriction within the column. Dialfiltration through a semipermeable membrane allows the removal of denaturant and other small molecules and retains the protein. This procedure has been used for large-scale processing and was particularly efficient in the refolding of prorennin and interferon-β.

During the refolding process, the formation of incorrectly folded species and aggregates usually decreases the refolding yield. For disulfide-bridged proteins, the renaturation buffer must contain redox-shuffling mixture to allow the formation of correctly paired disulfide bridges. Stabilizing reagents may be added to improve the refolding yield. An efficient strategy is the addition of small molecules to suppress intermolecular interactions leading to aggregation. Sugar, alcohols, polyols (including sucrose, glycerol, polyethylene glycol, isopropanol), cyclodextrin, laurylmaltoside, sulfobetains, L-arginine, and low concentrations of denaturants and detergents, have been used to increase the refolding yield. L-arginine at a concentration ranging from 0.4 M to 0.8 M is the most widely used additive today.

Another important factor in the refolding process is the rate of removal of the denaturant. Since there is kinetic competition between the correct folding and the formation of aggregates from a folding intermediate, conditions that favor folding over the accumulation of aggregates must be found. To optimize this selection, Vilick and de Bernadez–Clark developed a strategy for achieving high protein refolding yields. They start from a model of refolding, develop the equations of refolding kinetics, characterize the rate-limiting step of the process, determine the influence of various environmental parameters, and finally optimize the system of equations in a scheme involving diafiltration to remove the denaturant. The approach was evaluated in the refolding of carbonic anhydrase from 8 M urea. The yield obtained after three diafiltration experiments was 69% whereas the model predicted a yield of 73%.

The properties of molecular chaperones have also been utilized to increase the refolding yield. Altamiro and coworkers have developed a system for refolding chromatography that utilizes GroEL, DsbA, and peptidyl–prolyl isomerase immobilized on an agarose gel. Kohler and coworkers have built a chaperone-assisted bioreactor; however, it could only be used for three cycles of refolding and needs to be improved. Another strategy consists of the co-overproduction of the DnaK–DnaJ or GroEL–GroES chaperones with the desired protein; this can greatly increase the soluble yield of aggregation-prone proteins. Fusion proteins have also been used to minimize aggregation.

The recovery of active proteins from inclusion bodies is a rather complex process. Although some general strategies have been developed, optimal conditions have to be determined for each protein. Recently, genetic strategies to improve recovery processes for recombinant proteins have been introduced. They consist of the introduction of combinatorial protein engineering to generate molecules highly specific to a particular ligand. Such methods, which allow efficient recovery of a recombinant protein, will be increasingly used in industrial scale bioprocesses as well.

4.2 The Formation of Amyloid Fibrils and its Pathological Consequences

The formation of amyloid fibrils plays a key role in the origin of several neurodegenerative pathologies, such as spongiform encephalopathies and Alzheimer's disease. Historically, the term amyloid was introduced to describe fibrillar protein deposits associated with diseases known as *amyloidoses* that involve the extracellular deposition of amyloid fibrils and plaques with the aspect of starch. For many of these diseases, the major fibrillar

protein component has been identified. In the 1970s, it was demonstrated that lysosomal proteins under acidic conditions could form amyloid fibrils. It was generally accepted at this time that proteolysis was the amyloidogenic determinant. Twenty years later, it was shown that purified transthyretin is converted into amyloid fibrils via an acid-induced conformational change *in vitro*, demonstrating that conformational changes alone were responsible for producing an intermediate generating amyloid structure. These aberrant protein self-assemblies are at the origin of more than hundred human amyloid diseases, some of them being lethal.

Twenty unrelated protein precursors are known to form amyloid fibrils, among them transthyretin, lysozyme, immunoglobulin light chain, β_2 microglobulin, Alzheimer Aβ1–40 and Aβ1–42 peptides, the mammalian prion protein, and the yeast prion-like proteins (Table 1). Since they are subjects of another chapter, prion proteins will not be discussed here. Although they have no homology in sequence and structure, all form amyloid fibrils with a similar overall structure, suggesting a common self-assembly

Tab. 1 Amyloidogenic proteins and the corresponding diseases.

Clinical syndrome	***Precursor protein***	***Fibril component***
Alzeimer's disease	APP	β-peptide 1–40 to 1–43
Primary systemic amyloidosis	Immunoglobulin light chain	Intact light chain or fragments
Secondary systemic amyloidosis	Serum amyloid A	Amyloid A (76-residue fragment)
Senile systemic amyloidosis	Transthyretin	Transthyretin or fragments
Familial amyloid polyneuropathy I	Transthyretin	Over 45 transthyretin variants
Hereditary cerebral amyloid angiopathy	Cystatin C	Cystatin C minus 10 residues
Hemodialysis-related amyloidosis	β_2-microglobulin Apolipoprotein A1	β_2-microglobulin Fragments of Apolipoprotein A1
Familial amyloid polyneuropathy III	Gelsosin	71-amino acid fragment of gelsosin
Finnish hereditory systemic amyloidosis	Islet amyloid polypeptide (IAPP)	Fragment of IAPP
Type II diabetes Medullary carcinoma of the thyroid	Calcitonin	Fragments of calcitonin
Spongiform encephalopathies	Prion	Prion or fragments thereof
Atrial amyloidosis	Atrial natriuretic factor (ANF)	ANF
Hereditary nonneuropathic systemic amyloidosis	Lysozyme	Lysozyme or fragments thereof
Injection-localized amyloidosis	Insulin	Insulin
Hereditary renal amyloidosis	Fibrinogen	Fibrinogen fragments
Parkinson disease	α-synuclein*	

Source: (According to Kelly, J.W. (1996) Alternative conformations of amyloidogenic proteins govern their behavior, *Curr. Opin. Struct. Biol.* **6**, 11–17); *From J.C. Rochet & P.T. Lansbury (2000) *Curr. Opin. Struct. Biol.* **10**, 60–68.

pathway. In all proteins known to form amyloid fibrils, there is a conversion of α- to β-structure. Amyloid fibrils are abnormal, insoluble, and generally protease-resistant structures. They were first recognized by their staining properties. The most commonly used method to detect amyloid is staining by Congo Red, which exhibits a green birefringence. Amyloid fibrils are generally 60 to 100 Å in diameter and of variable length. X-ray diffraction data on fibrils, solid-state NMR studies, cryoelectron microscopy, and infrared Fourier transform experiments have shown that amyloid fibrils are made of two or more β-sheet filaments wound around one another. They have a characteristic cross-β repeat structure, the individual β-strands being oriented perpendicular to the long axis of the fibril.

Recently, progress has been made in the knowledge of the mechanisms involved in the formation of amyloid fibrils. Oligomeric prefibrillar intermediates have been extensively characterized with respect to their structure and temporal evolution. A well-documented example is provided by the studies on transthyretin. The biological role of this protein is the transport of thyroxin by direct binding and the transport of retinol via the retinol binding protein. The wild-type protein is very stable at neutral pH. In certain individuals, however, it is converted into amyloid fibrils, and this is associated with the disease, senile systemic amyloidosis. Several variants are associated with familial polyneuropathies. *In vitro* biophysical studies have identified conditions leading to amyloid formation. The three-dimensional structure of the protein is known. The wild-type protein is a tetramer at pH ranging between 5 and 7; the tetramer dissociates into a monomer when the pH decreases. The dissociation is the rate-limiting step of the process. The monomer exhibits an altered tertiary structure, which aggregates in amyloid

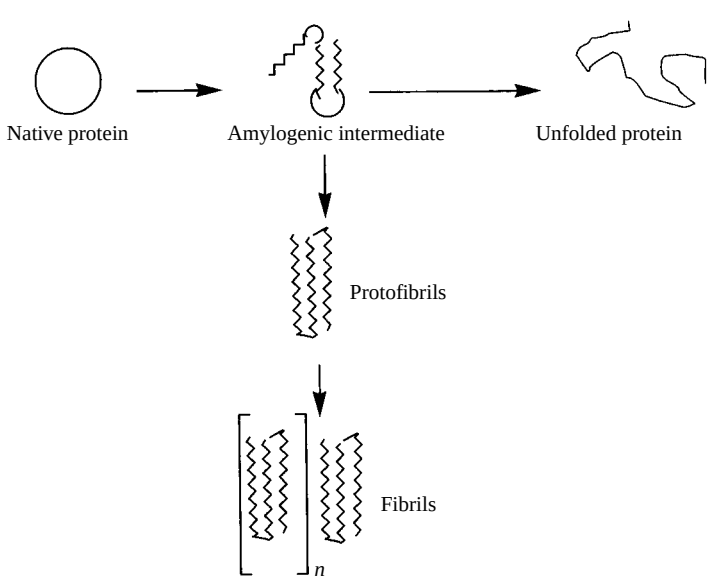

Fig. 10 Schematic representation of the formation of amyloid fibrils from a partially folded intermediate.

protofilaments, and then forms amyloid fibrils. This formation is at its maximum at pH 4.4. Using deuterium–proton exchange monitored by two-dimensional NMR spectroscopy on transthyretin at pH 5.75 and 4.5, Liu et al. have shown a selective destabilization of one half of the β-sandwich structure of the protein, increasing the mobility of this region. These studies have identified the residues that undergo increased conformational fluctuations under amyloidogenic conditions. The mutations in the pathological variants responsible for familial amyloid polyneuropathies are localized in this region. A strategy to delay the formation of amyloid fibrils proposed by Saccheti & Kelly was to develop molecules capable of stabilizing the tetramer.

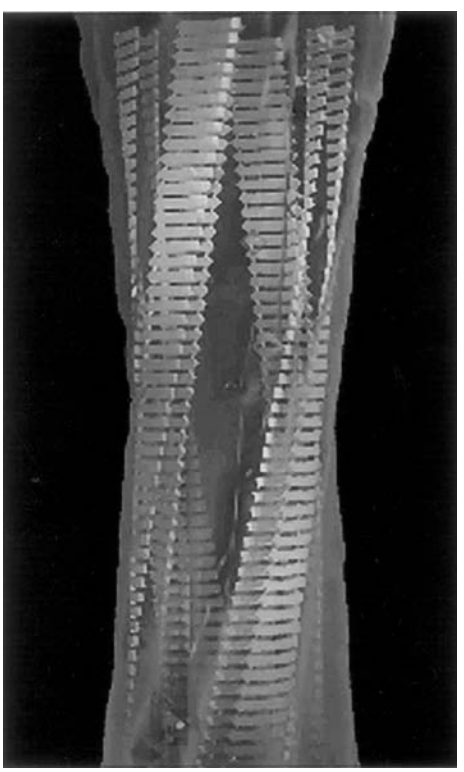

Two variants of human lysozyme, Ile56Thr and Asp67His have been reported to be amyloidogenic; they are responsible for fatal amyloidoses. Pepys and colleagues have determined the precise structures and properties of these mutants. The native fold of the two amyloidogenic variants, as resolved by X-ray crystallography, is similar to that of the wild-type protein. Both variants are enzymatically active, but have been shown to be unstable. The replacement of an aspartate by a histidine suppresses a hydrogen bond formed in the wild-type protein with a tyrosine in a neighboring β-strand. This rupture opens a large gap between two β-strands. In the other variant, the replacement of an isoleucine by a threonine suppresses a van der Waals contact with a neighboring helix. Consequently, changes in the interface between the α- and β-domains occur in both variants, destabilizing the molecule.

The mutations leading to amyloid fibril formation are observed to result in a decreased stability of the native state. In all cases, the formation of fibrils occurs from a partially structured molecule via nucleation-dependent oligomerization. It was observed for several proteins that fibrillation takes place only after a lag phase, which is abolished upon seeding. Nucleation is followed by the formation of protofibrils whose characteristics have been determined (Fig. 10). Atomic force microscopy and fluorescence correlation spectroscopy have been used to monitor transitions among the different types of assemblies.

Fig. 11 Molecular model of an amyloid fibril derived from cryoelectron microscopy analysis of fibrils grown from an SH3 domain by incubation of a solution containing the protein at low pH (reproduced from Dobson, C. (1999) *TIBS* **24**, 331, with permission).

Recent observations from Dobson and his group have shown that several proteins unrelated to amyloid diseases are able to aggregate *in vitro* into amyloid fibrils when exposed to mild denaturing conditions. These fibrils are indistinguishable from those found in pathological conditions. It was demonstrated for different proteins such as normal lysozyme, an SH_3 domain of a phosphatidyl inositol protein kinase (Fig. 11), an acyl phosphatase, and an α-helical protein, myoglobin suggesting a common mechanism for the formation of amyloid. These findings clearly indicate that amyloid formation is a general property of polypeptide chains rather than one restricted to definite sequences as occurs with chameleon sequences capable of adopting either a β- or an α-helicoidal structure depending on their environment. Furthermore, these aggregates exhibited an inherent toxicity when incubated with mouse fibroblasts. Several groups suggest that oligomeric intermediates rather than fibrils themselves are responsible for pathogenicity.

Significant progress has been made in understanding the mechanisms involved in the formation of amyloid fibrils. This is an important step in guiding research into the discovery of molecules with therapeutic efficiency.

See also Circular Dichroism in Protein Analysis.

Bibliography

Books and Reviews

Dobson, C.M., Šali, A., Karplus, M. (1998) Protein folding. A perspective from theory and experiment, *Angew. Chem. Int. Ed.* **37**, 868–893.

Ellis, R.J. (1991) Molecular chaperones, *Ann. Rev. Biochem.* **60**, 321–347.

Ellis, R.J., Hartl, F.U. (1999) Principles of protein folding in the cellular environment, *Curr. Opin. Struct. Biol.* **9**, 102–110.

Georgiou, G., de Bernadez-Clark, E. (Eds.) (1991) *Protein Refolding*, ACS Symposium Series 470, American Chemical Society, Washington, DC.

Ghélis, C., Yon, J.M. (1982) *Protein folding*, Academic Press, New York.

Jaenicke, R. (1987) Folding and association of proteins, *Prog. Biophys. Mol. Biol.* **49**, 117–237.

Kim, P.S., Baldwin, R.L. (1990) Intermediates in protein folding of small proteins, *Ann. Rev. Biochem.* **59**, 631–660.

Schlunegger, M.P., Bennett, M.J., Eisenberg, D. (1997) Oligomer formation by 3D domain swapping: a model for protein assembly and disassembly, *Adv. Protein Chem.* **50**, 61–122.

Wetlaufer, D.B., Ristow, S. (1973) Acquisition of the three-dimensional structure of proteins, *Ann. Rev. Biochem.* **42**, 135–158.

References of Primary Literature

Altamiro, M.M., Garcia, C., Possani, L.D., Fersht, A.R. (1999) Oxidative refolding chromatography: folding of the scorpion toxin Cn5, *Nat. Biotechnol.* **17**, 187–191.

Anfinsen, C.B. (1973) Principles that govern the folding of polypeptide chains, *Science* **181**, 223–230.

Anfinsen, C.B., Haber, E., Sela, M., White, F.H. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain, *Proc. Natl. Acad. Sci. USA* **47**, 1309–1314.

Baneyx, F. (1999) Recombinant protein expression in Escherichia coli, *Curr. Opin. Biotechnol.* **10**, 411–421.

Bennett, M.J., Choe, S., Eisenberg, D. (1994) Domain swapping: Entangling alliance between proteins, *Proc. Natl. Acad. Sci. USA* **91**, 3127–3131.

Bergdoll, M., Remy, M.H., Capron, C., Masson, J.M., Dumas, P. (1997) Proline-dependent oligomerization with arm exchange, *Structure* **5**, 391–401.

de Bernadez-Clark, E. (2001) Protein refolding for industrial processes, *Curr. Opin. Biotechnol.* **12**, 202–207.

de Bernadez-Clark, E., Schwarz, E., Rudolph, R. (1999) Inhibition of aggregation side-reactions during in vitro protein folding, *Methods Enzymol.* **309**, 217–236.

Booth, D.R., Suide, M., Belotti, V., Robinson, C.V., Hutchinson, W.L., Fraser, P.E., Hawkins, P.N., Dobson, C.M. Radford, S.E., Blake, C.C.F., Pepys, M.B. (1997) Instability, unfolding, and aggregation of human lysozyme variants underlying amyloid fibrillogenesis, *Nature* **385**, 787–793.

Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C.M., Stefani, M. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases, *Nature* **416**, 507–511.

Chiti, F., Wabster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., Dobson, C.M. (1999) Designing conditions for in vitro formation of amyloid protofilaments and fibrils, *Proc. Natl. Acad. Sci. USA* **96**, 3342–3344.

Crestfield, A.M., Stein, W.H., Moore, S. (1962) On the aggregation of bovine pancreatic ribonuclease, *Arch. Biochem. Biophys.* **1** (Suppl.), 217–222.

Dijkstra, K., Karvonen, P., Pirneskoski, A., Koivunen, P., Kivirikko, K.I., Darby, N.J., van Staaden, M., Sheek, R.M., Kemmink, J. (1999) Assignment of ^{1}H, ^{13}C and ^{15}N resonances of the a domain of protein disulfide isomerase, *J. Biomol. NMR* **14**, 194–196.

Dill, K.A. (1985) Theory for folding and stability of globular proteins, *Biochemistry* **24**, 1501–1509.

Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.H., Thomas, P., Chan, H.S. (1995) Principles of protein folding: a perspective from simple exact models, *Protein Sci.* **4**, 561–602.

van Duyne, G.D., Standaert, R.F., Karplus, P.A., Stuart, L., Shreiber, S.L., Clardy, J. (1993) Atomic structure of the human immunophilin FKBP12 complexes with FK506 and rapamycin, *J. Mol. Biol.* **229**, 105–124.

Edman, J.C., Ellis, L., Blacher, R.W., Roth, R.A., Rutter, W.J. (1985) Sequence of protein disulfide isomerase: implications of its relationship to thioredoxin, *Nature* **317**, 267–270.

Edwards, K.J., Ollis, D.L., Dixon, N.E. (1997) Crystal structure of cytoplasmic Escherichia coli peptidyl-prolyl isomerase: evidence for decrease mobility of loops upon complexation, *J. Mol. Biol.* **271**, 258–265.

Eliezer, D., Chiba, K., Tsurata, H., Doniach, S., Hodgson, K.O., Kihara, H. (1993) Evidence of an associative intermediate on the myoglobin refolding pathway, *Biophys. J.* **65**, 912–917.

Eliezer, D., Jennings, P.A., Wright, P.E.H. Doniach, S., Hodgson, K.O., Tsurata, H. (1995) The radius of gyration of an apomyoglobin folding intermediate, *Science* **270**, 487–488.

Ellis, R.J. (2001) Molecular crowding: an important but neglected aspect of the cellular environment, *Curr. Opin. Struct. Biol.* **11**, 114–119.

Ellis, R.J., Hartl, F.U. (1999) Principles of protein folding in the cellular environment, *Curr. Opin. Struct. Biol.* **9**, 102–110.

Fandrich, M., Flechter, M.A., Dobson, C.M. (2001) Amyloid fibrils from muscle myoglobin, *Nature* **410**, 163–166.

Fink, A.L. (1998) Protein aggregation, folding aggregates, inclusion bodies and amyloid, *Fold. Des.* **3**, R9–R23.

Freedman, R.B., Klappa, P., Ruddock, L.W. (2002) Protein disulfide isomerases exploit synergy between catalytic and specific binding domains, *EMBO Rep.* **31**, 136–140.

Georgiou, G., Valax, P. (1999) Isolating inclusion bodies from bacteria, *Methods Enzymol.* **309**, 48–58.

Goldberg, M.E. (1973) L'état natif est-il l'état fondamental? in: Sadron, C. (Ed.) *Dynamic Aspect of Conformational Changes in Biological Macromolecules*, Reidel, Dordrecht.

Guijarro, J.L., Sunde, M., Jones, J.A., Campbell, I.D., Dobson, C.M. (1998) Amyloid fibril by an SH3 domain, *Proc. Natl. Acad. Sci. USA* **95**, 4224–4228.

Harrison, S.C., Durbin, R. (1985) Is there a single pathway for the folding of a polypeptide chain? *Proc. Natl. Acad. Sci. USA* **82**, 4028–4030.

Horowitz, P. Criscimagna, N.L. (1986) Low concentrations of guanidinium chloride expose apolar surfaces and cause differential perturbation in catalytic intermediates of rhodanese, *J. Biol. Chem.* **261**, 15652–15658.

Jaenicke, R. (1991) Protein folding: local structures, domains, subunits and assemblies, *Biochemistry* **30**, 3147–3161.

Kadokura, H., Beckwith, J. (2002) Four cysteines of the membrane protein DsbB act in concert to oxidize its substrate DsbA, *EMBO J.* **21**, 2354–2363.

Karplus, M. Šali, A. (1996) Theoretical studies of protein folding and unfolding, *Curr. Opin. Struct. Biol.* **5**, 58–73.

Kelly, J.W. (1996) Alternative conformations of amyloidogenic proteins govern their behavior, *Curr. Opin. Struct. Biol.* **6**, 11–17.

Kelly, J.W. (1998) The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways, *Curr. Opin. Struct. Biol.* **8**, 101–106.

Kelly, J.W. (2002) Towards an understanding of amyloidogenesis, *Nat. Struct. Biol.* **9**, 323–325.

Kohler, R.J., Preuss, M., Miller, A.D. (2000) Design of a molecular chaperone-assisted protein folding bioreactor, *Biotechnol. Prog.* **16**, 671–675.

Kuhlman, B., O'Neill, J.M., Kim, D.E., Zhang K.Y., Baker, D. (2001) Conversion of a monomeric protein L to an obligate dimer by computational protein design, *Proc. Natl. Acad. Sci. USA* **98**, 10687–10691.

Langer, T., Lu, C., Flanagan, J., Hayer, A.K., Hartl, F.U. (1994) Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding, *Nature* **356**, 683–689.

Levinthal, C. (1968) Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44–45.

Liu, K. Cho, H.S., Lashuel, H.A., Kelly, J.W., Wemmer, D.E. (2000) A glimpse of a possible amyloidogenic intermediate of transthyretin, *Nat. Struct. Biol.* **7**, 754–757.

Liu, Y., Gotte, G. Libonati, M., Eisenberg, D. (2001) A domain-swapped ribonuclease A dimer with implication for amyloid formation, *Nat. Struct. Biol.* **8**, 282–284.

Liu, Y., Gotte, G. Libonati, M., Eisenberg, D. (2002) Structures of the two 3D domain-swapped ribonuclease A trimers, *Protein Sci.* **11**, 371–380.

London, J., Skrzynia C., Goldberg, M.E. (1974) Renaturation of Escherichia coli tryptophanase after exposure to 8 M urea. Evidence for the existence of nucleation centers, *Eur. J. Biochem.* **47**, 409–415.

Marston, F.A.O. (1986) The purification of eukaryotic polypeptides synthesized in Escherichia coli, *Biochem. J.* **240**, 1–12.

Martin, J.L., Bardwell, J.C., Kuriyan, J. (1993) Crystal structure of DsbA protein required for disulfide bond formation, *Nature* **365**, 464–468.

McCarthy, A.A., Haebel, P.W., Torronen, A., Rybin, V., Baker, E.N., Metcalf, P. (2000) Crystal structure of the protein disulfide bond isomerase, DsbC, from Escherichia coli, *Nat. Struct. Biol.* **7**, 196–199.

Mitraki, A., Betton, J.M., Desmadril, M., Yon, J.M. (1987) Quasi-irreversibility in the unfolding-refolding transition of phosphoglycerate kinase induced by guanidine hydrochloride, *Eur. J. Biochem.* **163**, 29–34.

Morozova-Roche, L.A., Zurdo, J., Spencer, A., Noppe, W., Receveur, V., Archer, D.B., Joniau, M., Dobson, C.M. (2000) Amyloid fibril formation and seeding by wild-type lysozyme and its disease-related mutational variants, *J. Struct. Biol.* **130**, 339–351.

Pecorari, F., Minard P., Desmadril, M., Yon, J.M. (1996) Occurrence of transient multimeric species during the refolding of a monomeric protein, *J. Biol. Chem.* **271**, 5270–5276.

Rochet, J.C., Lansbury, P.T. (2000) Amyloid fibrillogenesis: themes and variations, *Curr. Opin. Struct. Biol.* **10**, 60–68.

Roseman, A.M., Chen, S., White, H., Sebil, H.R. (1996) The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate binding domains in GroEL, *Cell* **87**, 241–251.

Saccheti, J.C., Kelly, J.W. (2002) Therapeutic strategies for human amyloid diseases, *Nat. Rev. Drug Disc.* **1**, 267–275.

Semisotnov, G.V., Kuwajima, K. (1996) Protein globularization during folding- a study by synchrotron small-angle X-ray scattering, *J. Mol. Biol.* **262**, 559–574.

Silow, M., Tan, Y.J., Fersht, A.N., Oliveberg, M. (1999) Formation of short-lived protein aggregates directly from the coil in two-state folding, *Biochemistry* **38**, 13006–13012.

Speed, M.A., Wang, D.I.C., King, J. (1996) Specific aggregation of partially folded polypeptide chains. The molecular basis of inclusion body composition, *Nat. Biotech.* **14**, 1283–1287.

Taketomi, H., Kano, F., Gô, N. (1988) The effect of amino acid substitutions on protein folding and unfolding studied by computer simulation, *Biopolymers* **27**, 527–560.

Wetzel, R., Perry, L.J., Veilleux, C. (1991) Mutations in human interferon γ affecting inclusion body formation by a general immunochemical screen, *Biotechnology* **9**, 731–737.

Wolynes, P.G., Onuchic, J.N., Thirumalai, D. (1995) Navigating the folding routes, *Science* **267**, 1618–1620.

Xu, Z., Horwich, A.L., Sigler, P.B. (1997) Crystal structure of the asymmetric GroEL-GroES-$(ADP)_7$ chaperonin complex, *Nature* **388**, 741–750.

Yon, J.M. (1996) The specificity of protein aggregation, *Nat. Biotech.* **14**, 1231.

Zhang, X., Beuron, F., Freemont, P.S. (2002) Machinery of protein folding and unfolding, *Curr. Opin. Struct. Biol.* **12**, 231–238.

Aging and Sex, DNA Repair in

Carol Bernstein and Harris Bernstein
University of Arizona, Tucson, AZ, USA

1 The DNA Damage Theory of Aging 56
1.1 Occurrence of DNA Damage and Pathways of DNA Repair 56
1.2 Consequences of Unrepaired DNA Damage 56
1.3 Life Span Extension by Genetic Alterations that Increase DNA Repair, Reduce Oxidative Damage, or Reduce Cell Suicide (Apoptosis) due to DNA Damage 57
1.4 Premature Aging Associated with Defects in DNA Repair or Increased Oxidant Status 63
1.5 Normal Aging in spite of Certain Defects in DNA Repair or Increases in Antioxidant Enzyme Production 64
1.6 Negative Correlation between Mitochondrial ROS Production and Life Span 65
1.7 Other Work Indicating the Central Role of DNA Damage and DNA Repair in Aging 67
1.8 Calorie Restriction and Aging 67
1.9 General Strategies for Coping with DNA Damage and Some Consequences 68
1.10 Potential Immortality of the Germ Line 69

2 DNA-repair Pathways and Their Relation to Aging 70
2.1 NER (Nucleotide Excision Repair) 70
2.2 BER (Base Excision Repair) 70
2.3 HRR (Homologous Recombinational Repair) 73
2.4 NHEJ (Nonhomologous End Joining) 75
2.5 MGMT (O^6-Methylguanine-DNA Methyltransferase) 76
2.6 Enzymes of DNA Repair Pathways and Enzymes Regulating DNA Damage–inducing ROS Contribute to Determination of Aging 77

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

3 The DNA Repair (and Complementation) Theory of Sex 77
3.1 Meiosis, an Adaptation for Repairing Germ Line DNA 77
3.2 Frequency of Sexual Reproduction 78
3.3 Costs of Sex 78
3.4 Benefits of Sex 79
3.5 Repair of DNA Damage of the Germ Line in Nonmeiotic Cell Divisions 80
3.6 Meiotic Recombination in Eukaryotes Probably Evolved from Recombination in Prokaryotes 81
3.7 Homologs of the Bacterial RecA Protein Have a Key Role in Eukaryotic Meiotic Recombinational Repair 81
3.8 The Adaptive Function of Recombination Appears to Be Removal of DNA Damage 82
3.9 Other Expectations of the DNA Repair and Complementation Theory of Sex 83

4 Vegetative Survival Strategies 84
4.1 Survival of Vegetative Cell Populations 84
4.2 Vegetative Survival Strategy for Mitochondria (and Chloroplasts) in the Germ Line 84
4.3 Dolly the Cloned Lamb, Cumulina the Cloned Mouse, and Low Success of Cloned Mammals 85

5 Three Levels of Sexual Communication Reflect (1) DNA Repair, (2) Complementation, and (3) Selection for Fitness 86
5.1 Sexual Communication in Bacteria, Primarily for DNA Repair 87
5.2 Sexual Communication in Fungi, Primarily for DNA Repair, but Also for Limited Complementation 88
5.3 Sexual Communication in Higher Eukaryotes, Primarily for Complementation while Repair Is an Automatic Concomitant of Sexual Reproduction 89
5.4 "Level 3" of Sexual Communication Is Used to Select for Fitness among Potential Mates 91

6 Overview 91

Bibliography 92
Books and Reviews 92
Primary Literature 92

Keywords

Aging
The progressive impairments of functions experienced by many organisms throughout their life span.

DNA Damage
A DNA alteration that has an abnormal structure, which cannot itself be replicated when the DNA is replicated, but which may be repaired.

DNA Repair
The process of removing damage from DNA and restoring the DNA structure.

Mutation
A change in the sequence of DNA base pairs, which may be replicated and thus inherited.

Sex
The process by which genetic material (usually DNA) from two separate parents is brought together in a common cytoplasm where recombination of the genetic material ordinarily occurs, followed by the passage of the recombined genome(s) to progeny.

Complementation
The masking of the expression of mutant genes by corresponding wild-type genes when two homologous chromosomes share a common cytoplasm.

A number of theories have been proposed to account for the biological phenomena of aging and sexual reproduction (sex). An emerging unified theory that accounts for a considerable amount of the data relating to both aging and sex is presented here.

Aging appears to be a consequence of DNA damage, while sexual reproduction (sex) appears to be an adaptation for coping with both DNA damage and mutation. DNA, the genetic material of most organisms, is composed of molecular subunits that are not endowed with any peculiar chemical stability. Thus, DNA is subject to a wide variety of chemical reactions that might be expected of any such molecule in a warm aqueous medium. DNA damages are known to occur very frequently, and organisms have evolved enzyme-mediated repair processes to cope with them. In any cell, however, some DNA damage may remain unrepaired despite repair processes. Aging appears to be due to the accumulation of unrepaired DNA damage in somatic cells, especially in nondividing cells such as those in mammalian brain and muscle.

On the other hand, the primary function of sex appears to be the repair of damages in germ cell DNA through efficient recombinational repair when chromosomes pair during the sexual process. This allows a relatively undamaged genome to initiate the next generation. In addition, in diploid organisms, sex allows chromosomes from genetically unrelated individuals (parents) to come together in a common cytoplasm (that of progeny). Since genetically unrelated parents ordinarily would not have common mutations, the chromosomes present in the progeny should complement each other, masking expression of any deleterious mutations that might be present.

Thus, aging and sex appear to be two sides of the same coin. Aging reflects the accumulation of DNA damage and sex reflects the removal of DNA damage, and in diploid organisms, the masking of mutations by complementation.

1 The DNA Damage Theory of Aging

1.1 Occurrence of DNA Damage and Pathways of DNA Repair

Except for certain viruses with an RNA genome, the genomes of most organisms are composed of DNA. If DNA damage is the cause of aging, then DNA damage is expected to occur frequently in multicellular organisms. Table 1 lists some important types of DNA damage caused by normal metabolic processes in mammals. These data suggest, for instance, that in the rat at least 95,000 DNA damages of various types occur, averaged over all cell types, per cell per day. The majority of these damages alter the structure of only a single DNA strand, so the redundant information in the complementary strand can usually be used to repair the damage. The damages shown in Table 1 are the newly occurring damages, most being rapidly repaired.

Five major DNA repair pathways known to be utilized by cells to repair the damages indicated in Table 1 are as follows:

- Nucleotide excision repair (NER) [with two subpathways, largely using the same enzymes: transcription coupled repair (TCR) and global genomic repair (GGR)]
- Base excision repair (BER)
- Nonhomologous end joining (NHEJ)
- Homologous recombinational repair (HRR)
- O^6-methylguanine-DNA methyltransferase (MGMT)

1.2 Consequences of Unrepaired DNA Damage

If accumulated DNA damages are the cause of aging, then repair processes would be less than 100% efficient; some types of unrepaired damages left each day would gradually build up in nondividing or slowly dividing cells. Most investigators examining the presence of DNA damages in tissues of young versus old mammals (usually rodents) have found an accumulation of damaged bases or single- or double-strand breaks with age. The tissues where accumulation of DNA damage has been shown include liver, kidney, heart, muscle, and brain.

A number of different types of DNA damage have been tested for their effects on transcription and DNA replication. It was found that transcription is blocked by UV-induced damages (mainly pyrimidine dimers) by adducts produced by derivatives of benzo[a]pyrene, *N*-acetoxy-2-fluorenylacetamide, or aflatoxin B_1 and also by the oxidized base, thymine glycol. UV-induced DNA damages and thymine glycol have also been shown to block DNA replication. These findings suggest that many types of DNA damage inhibit transcription and replication.

A reduction in the ability to transcribe mRNA should lead to a decline in the function of the cells. In fact, in mammalian

Tab. 1 Endogenous DNA damages in mammalian cells.

Type of damage	*Approximate average incidence (DNA damages/cell/day)*
Oxidative	500,000 (young mouse brain)
	2,000,000 (old mouse brain)
	86,000 (rats, all tissues)
	10,000 (humans, all tissues)
Depurinations	9,000 (humans and rats)
Single-strand break	7,200 (*in vitro*)
O^6-methylguanine	2,000 (*in vitro*)
Double-strand break	>40 (rats)[a]
	>3 (humans)[a]
DNA cross-link	>37 (rats)[a]
	>3 (humans)[a]
Glucose 6-phosphate adduct	3 (humans)

[a]These numbers were calculated from the values in the references by methods indicated in the literature.

brain, it has been shown that as single-strand DNA damages accumulate with age, mRNA synthesis and protein synthesis decline, neuron loss occurs, tissue function is reduced, and functional impairments directly related to the central processes of aging (e.g. cognitive dysfunction and decline in homeostatic regulation) occur. Similarly, it has been shown in muscle cells that as single-strand DNA damages accumulate, mRNA and protein synthesis decline, cellular structures deteriorate, cells die, and this is accompanied by a reduction in muscle strength and speed of contraction. Thus, for brain and muscle, accumulation of DNA damage is paralleled by declines in function, suggesting a direct cause-and-effect relationship between the accumulation of DNA damage and major features of aging. In other cells, including those of liver and lymphocytes, evidence for an increase in DNA damage paralleled by a decline in gene expression and cellular function has also been observed. In general, it appears that tissues composed of nondividing or slowly dividing cells accumulate DNA damage and experience functional declines with age.

1.3 Life Span Extension by Genetic Alterations that Increase DNA Repair, Reduce Oxidative Damage, or Reduce Cell Suicide (Apoptosis) due to DNA Damage

Table 2 lists alterations in genes controlling DNA repair, oxidant status, or apoptosis that result in increased life span. The increases in life span found with the genetic alterations in Table 2 are usually an increase in the maximum life span (not just the mean life span) by about 30 to 40%. Mean life span can be extended by reductions in tumorigenesis or acute and sporadic diseases, not generally regarded as a cause of aging. The organisms with increased maximum life span reported here showed longer spans of normal vigorous activity (not merely slowed metabolism, which can also extend life span). The cellular roles

Tab. 2 Life span extension: increased life span from alterations in genes controlling DNA repair, apoptosis, or oxidant status.

Organism	*Genetic alteration*	*Pathway*	*Aging phenotype*	*Fertility phenotype*	*Spontaneous cancer phenotype*	*Cellular ROS*	*Effect on*		
							Induced DNA damage	*Spont. mutation*	*Induced apoptosis*
Mouse	100-fold excess copies MGMT gene	O^6-meG DNA repair	Life span extension	n.t.	Reduced	n.t.	Reduced O^6-meG	No effect at *H-ras* locus	n.t.
Fruit fly	Excess SOD in neurons or all tissues or with catalase in all tissues	Removes ROS	Life span extension	No change	n.t.	Reduced occurrence of oxidized proteins	n.t.	n.t.	n.t.
Fruit fly	MsrA excess	Methionine sulfoxide reductase	Life span extension	Extended	n.t.	Repaired protein oxidation	n.t.	n.t.	n.t.
Mouse	$p66^{shc}$ defect	Blocks oxidant and apoptosis parts of p53 pathways	Life span extension	n.t.	n.t.	Reduced	Reduced oxidative damage	n.t.	Reduced
Human	Higher specific activity of PARP	BER	Life span extension	n.t.	n.t.	n.t.	n.t.	n.t.	n.t.

Notes: ROS: reactive oxygen species; Spont.: spontaneous; MGMT: O^6-methylguanine-DNA methyltransferase; n.t.: not tested; SOD: superoxide dismutase.

of these genetic alterations are described below.

1. *MGMT*. One frequent type of DNA damage (see Table 1) is O^6-methylguanine, caused by low levels of alkylating agents present in food, water, air, and tobacco smoke, and formed by normal processes in the body mediated by gastric bacteria and macrophages. O^6-methylguanine is specifically repaired by a DNA repair enzyme called *O^6-methylguanine-DNA methyltransferase* (MGMT). MGMT transfers the extra methyl group from guanine in DNA to a particular amino acid within itself and becomes "used up" after the transfer occurs. The *MGMT* gene codes for one of the five DNA repair mechanisms listed in Section 1.1. As indicated in Table 2, when 100 copies of the *MGMT* gene were inserted into the mouse genome, these mice (under the usual conditions of mouse maintenance) had their life span extended and died at a considerably slower rate than wild-type mice.

2. *SOD*. Another important type of metabolically caused DNA damage is oxidative damage, the most frequent damage identified (Table 1). An apparently unavoidable by-product of normal respiratory metabolism is the production of reactive oxygen species (ROS) from molecular oxygen, and ROS cause oxidative damage. ROS include free radicals (where the symbol $^\bullet$ indicates an unpaired electron): the superoxide radical ($O_2^{\bullet-}$) and the hydroxyl radical ($OH^\bullet$). Another oxygen respiration by-product is hydrogen peroxide (H_2O_2). H_2O_2, if not removed, it diffuses fairly easily through the cell, and when it encounters Fe^{2+} (the ferrous ion), it can undergo the Fenton reaction and produce $OH^\bullet$ and other ROS. ROS produce a number of lesions in DNA, including base lesions, sugar lesions (the deoxyribose sugar is in the backbone of DNA), DNA–protein cross-links, single-strand breaks, double-strand breaks, and abasic sites.

The major ROS produced by the cell is $O_2^{\bullet-}$, formed in the mitochondria (the energy-producing organelles of the cell). Superoxide dismutase (SOD) occurs in two forms, manganese SOD (MnSOD) and copper/zinc SOD (Cu/ZnSOD). Both forms of SOD convert $O_2^{\bullet-}$ to the less damaging H_2O_2, and then another enzyme, catalase, converts H_2O_2 to molecular oxygen and water. MnSOD occurs in the mitochondria and Cu/ZnSOD occurs in the cytoplasm. As shown in Table 2, a number of investigators have found that inserting genes producing higher than normal levels of superoxide dismutase into the fruit fly (*Drosophila melanogaster*) genome results in life span extension. Insertion of genes producing either MnSOD or Cu/ZnSOD caused life span extension, although the artificially inserted Cu/ZnSOD only produced life span extension when its expression was restricted to the motor neurons, or solely to the adult phase of the fruit fly life cycle.

Aging has been found to correlate with increased levels of oxidative products, such as protein carbonyls and 8-oxo-guanine in DNA, and fruit flies lacking either catalase or Cu/ZnSOD have a reduced life span. Further, selection of a population of fruit flies for increased life span correlates with strongly increased expression of both MnSOD and Cu/ZnSOD. Reverse selection of these long life span flies to a shorter life span resulted in reduced expression of Cu/ZnSOD.

3. *MsrA*. In addition to DNA damage, free radicals damage proteins, lipids, and carbohydrates. Most proteins have a short half-life (averaging about three days in mouse liver). Oxidatively damaged proteins and lipids are subject to both

degradation and some repair reactions. If cellular genes that code for enzymes involved in the replacement of damaged proteins are themselves damaged, then damaged proteins may not turn over as rapidly, and protein damages may become important as they accumulate with age. Table 2 shows that insertion of an extra gene encoding bovine methionine sulfoxide reductase (MsrA) in the fruit fly genome, which helps repair oxidatively damaged proteins, leads to life span extension. Consistent with this, MsrA, when defective in the mouse, results in early aging (Table 3).

4. *p66Shc*. The *p53* gene has a central role in response to DNA damage. The p53 protein is directly active in three forms of DNA repair (NER, BER, and HRR). When there is no externally induced DNA damage, p53 has a half-life of only 5 to 40 minutes since specific enzymes target p53 for degradation. Thus, p53 is kept at a low level when there is no DNA damage. However, upon exposure of a cell to DNA-damaging agents, p53 becomes metabolically stable and, in addition, more copies of it are produced in the cell. In the presence of various types of DNA damage, p53 undergoes modifications at some of the 18 different sites within the protein. Some of these modifications [phosphorylations, acetylations, poly(ADP-ribosyl)ations, or sumoylations (covalent attachments of small ubiquitin-like proteins) allow the p53 protein to act as a regulatory agent, activating numerous other genes, carrying out different responses to different kinds or levels of DNA damage. The p53 protein can regulate or act in at least four major types of responses to DNA damage (acting as a "master switch"), and which action or transactivation (regulating the induction of other genes) it performs depends on the level and type of DNA damage. p53 can (1) send the cell into cell cycle arrest (to allow extra time for repair of DNA damage); (2) act directly in DNA repair (see Fig. 1 for where p53 acts in NER); (3) cause the cell to switch into a cell suicide mode (apoptosis); or (4) cause the cell to produce higher levels of ROS (apparently as a preliminary to entering the cell suicide mode of apoptosis). When acting to increase the internal level of ROS and entry into apoptosis, p53 acts through another gene it controls, *p66Shc*.

When a mouse embryo is produced with both copies of its *p66Shc* gene inactive (a p66Shc "knockout"), mouse embryo fibroblast cells derived from it have intracellular levels of ROS reduced by about 40%. Consistent with this reduction in ROS, there is also greatly reduced oxidative damage accumulation in both nuclear and mitochondrial DNA of these cells. A similar reduction in nuclear and mitochondrial DNA damage is seen *in vivo* in the tissues of lung, spleen, liver, and skin in 3- and 24-month-old *p66Shc* knockout mice, although there is no reduction in the brain, where *p66Shc* is not normally expressed. Cells of these mice are inhibited from undergoing apoptosis after cellular oxidative damage (when challenged with externally applied H_2O_2). Knockout mice without *p66Shc* show life span extension without any notable increase in cancer or other pathological defects (Table 2). Mice with a type of overactive p53 (an increase in some p53 functions) and intact *p66Shc* show early aging (Table 3). On the other hand, removal of all p53 functions (some of which are protective in DNA repair) also results in early aging (Table 3).

5. *PARP*. DNA damages caused by alkylating agents (such as those that methylate guanine, discussed above),

Tab. 3 Early aging: decreased life span from alterations in genes controlling DNA repair or protein oxidation.

Organism	*Genetic alteration*	*Pathway*	*Aging phenotype*	*Fertility*	*Spontaneous cancer*	*Effect on*			
						Cellular ROS	*Induced DNA damage*	*Spont. mutation*	*Induced apoptosis*
Human	RECQ3 helicase and exonuclease defect (Werner syndrome)	HRR and NHEJ	Early aging	Reduced	Increased	n.t.	Increased	Increased	n.t.
Human	RECQ2 helicase defect (Bloom syndrome)	HRR and NHEJ	Early aging	Reduced	Increased	n.t.	Increased	Increased	n.t.
Human	RECQ4 helicase defect (Rothmund–Thomson syndrome)	DNA repair pathway, unknown type	Early aging	Reduced	Increased	n.t.	n.t.	Increased	n.t.
Human and Mouse	XPD helicase defective at certain sites (Trichothiodystrophy)	NER, also alters transcription initiation	Early aging	Reduced	No change	n.t.	n.t.	n.t.	Increased
Human	CSB defect at 2 helicase motifs or ATPase motif (Cockayne syndrome)	BER if defective at helicase motif V or VI, TCR if defective in ATPase function	Early aging	n.t.	No change	n.t.	Increased	n.t.	Increased
Mouse	Ku-80 (activator of Ku-70 helicase) defect	NHEJ	Early aging	n.t.	Increased	n.t.	n.t.	n.t.	
Mouse	Topoisomerase IIIβ defect	Unknown, but probably DNA repair, replication, or recombination	Early aging	n.t.	n.t.	n.t.	n.t.	n.t.	

(*continued overleaf*)

Tab. 3 (*continued*)

Organism	*Genetic alteration*	*Pathway*	*Aging phenotype*	*Fertility*	*Spontaneous cancer*	*Effect on*			
						Cellular ROS	*Induced DNA damage*	*Spont. mutation*	*Induced apoptosis*
Mouse	ERCC1 defect	NER and interstrand cross-link repair (HRR)	Early aging	Infertile	No change	n.t.	Increased	Increased	
Mouse	p53 overactivated	Increases some p53 functions	Early aging	n.t.	Reduced	n.t.	n.t.	n.t.	
Mouse	p53 defect	Blocks all p53 functions, including NER, BER and HRR	Early aging	Reduced	Increased	n.t.	Increased	Increased	
Mouse	MsrA Defect	Methionine sulfoxide reductase	Early aging	n.t.	n.t.	Increased protein oxidation	n.t.	n.t.	

Notes: ROS: reactive oxygen species; Spont.: spontaneous; n.t.: not tested.

ionizing radiation (which produces DNA single- and double-strand breaks and oxidative damages), and ROS result in rapid activation of an enzyme called *poly(ADP-ribose) polymerase*, or PARP. PARP, similar to p53 discussed above, has a role as a "master switch". PARP can (1) act directly in one form of DNA repair, BER, (2) control the function of many other proteins by catalyzing the addition of ADP-ribose branched polymers onto them (either activating or repressing their function), and (3) trigger apoptosis (cell suicide). In addition, PARP controls new transcription or activities of a number of genes affecting survival or apoptosis, including p53. It was found that centenarians (humans who have lived for more than 100 years) have a modified form of PARP, which is more efficiently activated than the PARP of noncentenarians (Table 2), thereby apparently causing life span extension. In addition, the maximal poly(ADP-ribosyl)ation capacity (efficiency of activation of PARP) in leukocytes of 13 mammalian species of different life span was measured. There was a strong correlation of PARP efficiency of activation with species-specific life span.

1.4 Premature Aging Associated with Defects in DNA Repair or Increased Oxidant Status

DNA damages are so frequent (Table 1) that total absence of DNA repair of a common damage is likely to be incompatible with life. If a DNA repair pathway lacks an essential enzyme, but the missing enzyme can be, at least, partially compensated for by a similar enzyme, then repair may be adequate to allow sufficient survival and growth to show premature aging. This will also occur if a particular DNA damage is preferentially repaired by one pathway, but another repair pathway, with less efficiency, also repairs that damage.

1. *Helicase*. There are at least 31 human enzymes that are helicases or contain helicase-motif domains. Helicases are enzymes that unwind and separate the strands of DNA, usually using the hydrolysis of ATP to provide the necessary energy. Some enzymes with multiple helicase-motif domains only act as ATPases, providing energy to DNA-related processes. Helicases or enzymes with helicase-motif domains participate in DNA repair, DNA replication, and DNA recombination. Usually, the helicase activity is specific for a particular DNA configuration. Some helicases involved in particular DNA-repair pathways may be partially replaceable, at least at a low level, by other helicases. That may be why five genes, which code for enzymes with helicase functions, or helicase motifs plus an ATPase function, and which are required in different DNA repair pathways, when genetically defective, cause syndromes characterized by early aging in humans (Table 3). These syndromes are Werner syndrome, Bloom syndrome, Rothmund–Thomson syndrome, Trichothiodystrophy and Cockayne syndrome (Table 3). Similarly, in the mouse, a defect in the *Ku-80* gene, which normally activates the Ku-70 helicase function, results in an early aging phenotype (Table 3). The different helicases listed in Table 3 have specificities for HRR, NHEJ, NER, TCR or BER, so that defects in each of these DNA repair pathways may allow accumulation of different types of DNA damage, each type being able to contribute to premature aging.

2. *Topoisomerase*. Topoisomerases interact with helicases in DNA repair, recombination, and replication. When a helicase

unwinds the two DNA strands of the double helix, this introduces supercoiling of the associated DNA. Topoisomerases introduce controlled breaks plus reattachments in DNA to relieve supercoiling. There are a number of topoisomerases in mouse and human cells. The different topoisomerases interact specifically with different helicases. However, some topoisomerases may be partially replaceable by another topoisomerase at a low level. In the mouse, a mutant lacking topoisomerase IIIβ develops to maturity but shows early aging (Table 3). Topoisomerase IIIβ interacts with human RecQ5β helicase and is thought to act in DNA repair, replication, or recombination (Table 3).

3. *ERCC1*. Excision Repair Cross Complementing 1 (*ERCC1*), when defective, is another gene whose absence or truncation causes an early aging phenotype in the mouse (Table 3). ERCC1 functions in both NER and interstrand cross-link repair (in a step prior to HRR). ERCC1 has homology with an endonuclease active in NER in yeast, and that yeast endonuclease can compensate for the loss of a topoisomerase or a helicase. Thus, ERCC1 may have some functional similarity to topoisomerase or helicase in DNA repair. Conversely, loss of ERCC1 may be partially compensated for by a helicase or topoisomerase, or by another endonuclease in mouse, so that a defect in ERCC1 is not lethal but causes early aging. ERCC1 primarily functions in NER as an endonuclease as illustrated in Fig. 1.

4. *p53*. Similar to helicase and topoisomerase, p53 occurs as one of a family of enzymes, p53, p63, and p73 (and both p73 and p63 have multiple isoforms), which share significant homology and have similar functions. In particular, p73 has a role in activating DNA repair enzymes and in carrying out apoptosis in the face of excess DNA damage (see below in Fig. 3). Thus, loss of p53 may, in part, be compensated for by functions of p73 and/or p63. In Section 1.3.4, we briefly discussed an overactive form of p53 that causes early aging. This mutant form of p53 has its effect in the presence of a wild-type p53 (a heterozygous situation) where it may increase some functions detrimental to the cell. Although it is not known which functions it increases, an increase of p66Shc under p53 control could reasonably be expected to cause early aging, since it would increase DNA damage through increases in ROS. In addition, a p53 knockout mouse, lacking all functions of p53, including its functionality in three DNA repair pathways (NER, BER, and HRR), is also viable but ages prematurely (Table 3).

5. *MsrA*. As discussed in Section 1.3.3, if cellular genes, which code for activity in the replacement of damaged proteins, are themselves damaged, then damaged proteins may not turn over as rapidly, and protein damages may become important as they accumulate with age. Added activity of MsrA in the fruit fly gave greater longevity. Defective MsrA in the mouse caused early aging (Table 3).

1.5 Normal Aging in spite of Certain Defects in DNA Repair or Increases in Antioxidant Enzyme Production

If a DNA repair pathway lacks an enzyme, but the missing enzyme is partially compensated for by a similar enzyme, then repair may be sufficient to allow survival, growth, and normal aging. While such fairly good compensation may allow normal aging, the repair of DNA damages would still be less than if the repair pathway were intact, and that could lead to

increased carcinogenesis. Such fairly good compensation for DNA repair enzyme defects may be the basis for normal aging, but increased carcinogenesis, shown by mice with the DNA repair mutations listed in Table 4. As shown in Fig. 1, XPA and XPC proteins each occur as one member of a pair of complexed proteins whose function is recognition of DNA damage to be repaired by an NER pathway. It is possible that the other complexes that can recognize DNA damage can compensate, at some level, when one of the recognition complexes is absent.

Further, in some instances, genetic alterations caused increased SOD production in fruit flies but did not affect aging. However, the inserted *SOD* gene may have been turned on under the control of promoters expressing in tissues where it may not have been useful, or at rather low levels.

1.6 Negative Correlation between Mitochondrial ROS Production and Life Span

If DNA damage is the major cause of aging, ROS are a major source of DNA damage, and mitochondria are a major source of ROS, then animals with mitochondria producing higher levels of ROS may have a shorter life span, other factors being equal. Comparisons were made between long-lived birds and short-lived mammals. Pigeons, with a maximum life span of 35 years were compared with rats (of similar body size) with a maximum life span of 4 years, and parakeets and canaries (maximum life spans of 21 and 24 years, respectively) were compared with mice (maximum life span of 3.5 years). Mitochondrial ROS production was lower in the longer-lived avian species. However, in addition, pigeons were shown to have higher levels of SOD in brain, heart, and kidney than the levels shown by rats, so there was also higher antioxidant enzymatic protection in the longer-lived pigeon. This indicates that antioxidant enzymes, which confer resistance to an externally added source of ROS may be of comparable importance to longevity as endogenous rates of ROS production. The values of one measured DNA-damaged base, 8-oxodeoxyguanine, were lower in canary brain and parakeet heart nuclear DNA than in the comparison tissues of the mouse, while in the other comparisons, the level of this one damaged base was not significantly different in nuclear DNA. In another experiment quoted by Herrero and Barja, other workers showed that nuclei of starlings (another long-lived bird, maximum life span of 20 years) have less DNA breaks and abasic sites after exposure to H_2O_2 than those of mice.

Mice heterozygous for a MnSOD defect ($MnSOD^{+/-}$ mice) have higher levels of oxidative damage to DNA, protein, and lipids in their mitochondria, but no increased damage to nuclear DNA or cytoplasmic proteins. The $MnSOD^{+/-}$ mice live as long as wild-type mice, showing that mitochondrial DNA damage (as distinct from nuclear DNA damage) may not be central to longevity. This may be because there are on the order of 1,000 mitochondria per cell, and mitochondria with excess damage may be replaced by replication of less damaged mitochondria.

In Section 1.3.2, we mentioned that the long-lived strains of fruit flies have higher levels of antioxidant defense enzymes. The long-lived flies also had mitochondria that had lower levels of ROS leakage. These less leaky mitochondria, when transferred to short-lived flies through maternal inheritance (only maternal mitochondria are passed on to and maintained in progeny

Tab. 4 Normal aging: alterations in genes controlling DNA repair or oxidant status but which do not alter life span.

Organism	*Genetic alteration*	*Pathway*	*Aging phenotype*	*Fertility*	*Spontaneous cancer*	*Effect On*			
						Cellular ROS	*Induced DNA damage*	*Spont. mutation*	*Induced apoptosis*
Mouse	XPA defect	NER	Unchanged[a]	Unchanged	Increased liver cancer	n.t.	n.t.	Increased in liver, no change in brain	n.t.
Mouse	XPC defect	GGR (NER)	Unchanged[a]	Unchanged	Unchanged (when not exposed to UV light)	n.t.	n.t.	Increased	n.t.
Fruit fly	Excess SOD	Removes ROS	Unchanged	Unchanged	n.t.	n.t.	n.t.	n.t.	n.t.

Notes: ROS: reactive oxygen species; Spont.: spontaneous; n.t.: not tested.
[a]Based on observations during the first half of life span.

flies), also transferred the ability to live about 25% longer.

1.7 Other Work Indicating the Central Role of DNA Damage and DNA Repair in Aging

Numerous studies have been performed in mammals on the correlation between the ability of cells to repair DNA and the life span of the species from which the cells were taken. The life spans of the species varied from 1.5 years for the shrew to 95 years for man. Almost all of the studies showed a positive correlation between DNA repair capacity and life span. Many experiments have been performed on the effect of adding antioxidants to the diets of organisms upon the organism's life span. Although the results of such experiments are not entirely consistent, certain antioxidants have been found to generally increase life span. Vitamin E, for example, has been found to increase the life span of rat, insects, rotifers, nematodes, and paramecium.

More than 50 studies have been performed to examine the possible experimental acceleration of aging by externally applied DNA-damaging agents. Overall, it has been found that sublethal doses of ionizing radiation or DNA-damaging chemicals in the diet shorten life span, but many specific aspects of normal aging are not accelerated. Several authors have noted that the distribution (over time and in different tissues) of DNA damages induced by external agents does not closely mimic that of natural damages. This difference could explain why the life-shortening effects induced by external agents do not closely conform to natural aging. In particular, natural damages probably accumulate gradually, so they would tend to build up in nondividing cells, while they would be diluted out in dividing cells. Exposure to an external agent over a brief period, on the other hand, could cause equally large numbers of damages to nondividing and rapidly dividing cells. The effect on rapidly dividing cells could be very large by interfering with DNA replication. In addition, if oxidative damages are important in normal aging, then brain cells, which have a high level of oxidative metabolism should have more damages than most other cell types. Externally applied damages would not be expected to produce this particular type of bias. Thus, the general finding that sublethal exposure to DNA-damaging agents shortens life span, while not uniformly accelerating the natural aging process, is consistent with the DNA damage theory of aging.

1.8 Calorie Restriction and Aging

Numerous studies have shown that calorie restriction in yeast, nematode worms, fruit flies, and rodents can give life span extension. The transcription of a large number of genes is altered with calorie restriction. In rat muscle, 34 of 800 genes tested were altered in their transcription under calorie restriction, with some of the largest changes being upregulation of Cu/ZnSOD and MnSOD. In fruit flies, where 14,028 genes were assessed, 2,188 showed significant up- or downregulation after calorie restriction. In particular, DNA repair genes were downregulated (which may reflect reduced DNA damage). Evaluation of upregulated genes was problematic, however, because expression measures are proportional to the total mRNA pool, and there was massive downregulation of a large number of genes related to cell growth, and

other targeted areas. Thus, genes whose absolute transcription abundance remains unchanged will appear to be upregulated.

Overall, calorie restriction results in slower accrual of oxidative damage. The steady state oxidative damage measured represents the equilibrium between oxidant generation, oxidant scavenging, repair, and protein and lipid turnover (DNA does not turn over, although it is repaired, with damaged bases and nearby bases being replaced). The consensus is that the primary reason for lower oxidative damage observed after calorie restriction is a reduction in the generation of ROS.

1.9 General Strategies for Coping with DNA Damage and Some Consequences

Different organisms, or even different tissues within the same organism, appear to use different strategies for dealing with DNA damage. The three major strategies are described below.

1. *Cell replacement strategy*. Bone marrow and hemopoietic cells of man, guinea pig, and mouse seem to maintain their population numbers by a cell-replacement strategy. For instance, mouse bone marrow cells have a turnover time of about 1 to 2 days, and there appear to be no significant differences in the erythrocyte production from marrow stem cell lines in old and young adult mice, suggesting that DNA damages do not accumulate in this cell population. However, rapidly dividing cells are vulnerable to accumulation of mutations, which arise by errors of replication. Hematopoietic stem cells accumulate mutations (not damages) as people age from birth to 96 years. The accumulation of mutations in replicating somatic cells is widely regarded as the cause of cancer.

2. *DNA repair in nondividing somatic cell populations*. The organs including brain, muscle, and liver consist, largely, of nondividing cells, and these cells carry out DNA repair as a strategy of coping with DNA damages. However, DNA repair is less than 100% efficient, and DNA damages accumulate with time. Brain, muscle, and liver are subject to some of the more conspicuous progressive declines in function characteristic of human aging. Investigations were reviewed in 1992 on the accumulation of DNA damage in mammalian muscle, brain, and liver. In that review, 4 studies on muscle, 9 on brain, and 14 on liver reported accumulation of DNA damage with age. In most of these investigations, the type of damage measured was single-strand breaks. In 1996, an increase in DNA single- and double-strand breaks in neurons of the rat cerebral cortex with age was found. Neurons in young 4-day-old rats had about 3,000 single-strand breaks, which increased to 7,400 in neurons of old rats more than 2 years of age. Double-strand breaks increased from about 156 in young rats to about 600 in old rats. It was suggested that gradual accumulation of DNA damage with age could be a primary reason for the breakdown of the metabolic machinery leading to the eventual senescence and death of the neuron. It was found in the same year that DNA adducts, a type of DNA damage, increase in rat brain with age. Some of these were identified as malondialdehyde adducts of dGMP. It was suggested that this accumulation of DNA damages may contribute to cerebral aging. A significant increase in single-strand breaks/alkali-labile sites with age in rat liver hepatocytes were reported in 1994. Thus, numerous studies

in mammals indicate that long-lived, nondividing, differentiated cells accumulate DNA damage with time. This damage may account for many of the progressive declines in functions that define aging.

3. *Cellular redundancy*. Like repair, cellular redundancy may be another strategy for coping with DNA damage in nondividing cells. The brain has an unusually high level of oxidative metabolism compared to other organs. Brain neurons are nondividing and there is evidence for DNA damage accumulation in the brain. There is a clear loss of neurons with age. Compared to young rats, old rats have about a 50% loss of neurons in many regions of their brains. Numerous studies in humans have also shown a loss of neurons with age. The brain appears to use a strategy of cellular redundancy to compensate for the loss of neurons with age. It has been estimated that the brain is twofold larger than necessary for short-term survival. Comparisons of different mammalian species indicate that the maximum life span in mammals is directly proportional to brain size. Thus, the brain appears to be protected from loss of neuronal function by cellular redundancy, and this type of redundancy may be significant in determining life span.

1.10 Potential Immortality of the Germ Line

Cells of the germ line are capable of avoiding aging. While multicellular organisms ordinarily age and die, their germ line is potentially immortal. The germ line is distinguished from other cell lines by periodic meiosis. HRR is especially promoted during meiosis, when the homologous chromosomes are closely paired along their length. Meiosis appears to be an adaptation for removing DNA damages through HRR, and the potential immortality of the germ line may be related to this special repair capability. This idea has been tested, using single-celled paramecia that can undergo either asexual or sexual reproduction. When they grow asexually, clones of *Paramecium tetraurelia* age (show reduced vigor) and then die. These paramecia have a macronucleus containing 800 to 1,500 copies of the genome that expresses cellular functions, and a micronucleus that contains the germ line DNA.

If the macronuclei of clonally young paramecia are injected into old paramecia, the old paramecia have their life span prolonged. In contrast, cytoplasmic transfer from young to old paramecia does not prolong the life span of the old paramecia. This suggests that the macronucleus, rather than the cytoplasm, determines clonal aging. Asexually growing clones of paramecia have been found to accumulate DNA damage in the macronucleus over successive generations of clonal growth. Upon sexual reproduction (conjugation) or self-fertilization (automixis), a new macronucleus develops from the micronucleus and the old macronucleus disintegrates. Both of these processes (conjugation and automixis) include meiosis, which involves pairing of homologous chromosomes and the opportunity for homologous recombinational repair (HRR) of DNA (see Section 2.3). It was found that at a few clonal generations after meiosis, the level of DNA damage in macronuclear DNA is low but then increases as the cells undergo clonal aging. Thus, accumulation of DNA damage in the macronucleus may account for clonal aging and DNA repair during meiosis (principally HRR) may account, in large part, for the potential immortality of the germ line.

2
DNA-repair Pathways and Their Relation to Aging

2.1
NER (Nucleotide Excision Repair)

NER repairs DNA with helix-distorting single-strand DNA damages, including some oxidative damages such as the cyclopurine 8,5′-(S)-cyclo-2′-deoxyadenosine. About 30 proteins are involved in NER. The key steps in the process (Fig. 1) are (1) recognition of a DNA defect; (2) recruitment of an initial repair complex; (3) preparation of the DNA for repair through the action of helicases; (4) incision of the damaged strand on each side of the damage with release of the damage in a single-strand fragment about 24 to 32 nucleotides long; (5) filling in of the gap by repair synthesis; and (6) ligation to form the final phosphodiester bond. The two subpathways of NER, TCR, and GGR, are initiated somewhat differently, but after initiation most enzymatic steps are the same.

Individuals with the inherited disease xeroderma pigmentosum (XP) are defective in NER and are sensitive to UV radiation. About 25% of such patients also have neurodegeneration attributed to neural oxidative damage. XP patients have defects in one of seven *XP genes*, *XPA* through *XPG*, with some of their roles indicated in Fig. 1. A number of specific mutations resulting in defects in XPD, a helicase, cause a milder disease than XP, called *trichothiodystrophy*, and these mutations cause early aging (Table 3). XPA or XPC, both involved in the recognition of DNA damage (Fig. 1), when defective, do not appear to affect aging (Table 4), perhaps because if either is defective, compensation can occur to some extent. However, a double-mutant mouse, defective with a trichothiodystrophy-causing XPD mutation plus an XPA mutation, ages more rapidly than a mouse with a single XPD defect.

ERCC1, an endonuclease, associates with XPF, and then this pair associates with the initial repair complex to carry out an incision on a strand carrying the DNA damage. Defects in ERCC1 cause early aging (Table 3). CSB is an enzyme with seven helicase motifs and an ATPase function. The ATPase function is required in TCR of NER. CSB interacts with the TFIIH complex [a transcription complex (transcription factor II H)] required in NER (Fig. 1), as well as with XPA, another NER protein (Fig. 1). Patients with alterations in CSB show early aging (Table 3). p53 affects the helicase step of NER and is also required for apoptosis when damage requiring NER is excessive (Fig. 1). Defects in p53 cause early aging (Table 3). Thus, at least four NER enzymes (XPD, CSB, p53, ERCC1), when defective, may allow helix-distorting DNA damages to accumulate more rapidly than normal, leading to premature aging. In addition, the p53-regulated pathway of apoptosis (Fig. 1), perhaps triggered by helix-distorting DNA damages, if defective in p66Shc, results in life span extension (Table 2).

2.2
BER (Base Excision Repair)

BER protects mammalian cells against single-base DNA damage by methylating agents, most oxidative damages, and a large number (about 9,000 per cell per day – see Table 1) of spontaneous depurinations. BER is mediated through at least two subpathways, one involving removal and replacement of a single nucleoside and the other involving a longer patch repair of 2 to 15 nucleotides. Repair

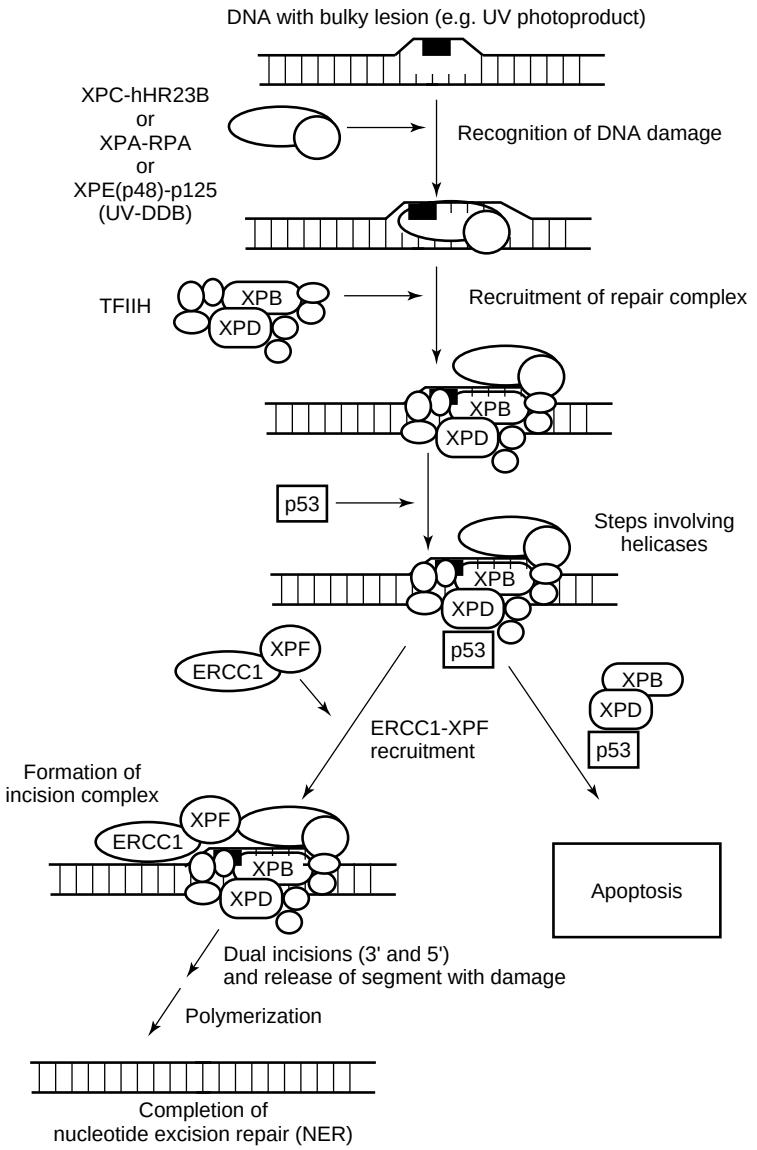

Fig. 1 Nucleotide excision repair (NER).

can be initiated by removal of a damaged base by a DNA glycosylase, which binds the altered deoxynucleoside in an extrahelical position and catalyzes cleavage of the base–sugar bond, or can occur at a site of spontaneous depurination [an apurinic (AP) site] (Fig. 2). The glycosylases that remove damaged bases are specific to the particular damage. Ref-1 (also called apurinic/apyrimidinic endonuclease or APE) then makes a 5′ nick in the DNA backbone, followed by PARP, 1 acting as a nick surveillance protein. PARP locates at the site of the single chain

nick. The nick then activates PARP, which catalyzes formation of poly(ADP-ribose) (polyAR) branched chain polymers, using the adenosine diphosphate ribose part of nicotinamide adenosine dinucleotide (NAD^+) with the release of nicotinamide (NAM) (Fig. 2). PolyAR is attached to numerous nearby proteins, including PARP itself and p53. Proteins modified by polyAR gain or lose functions. Fig. 2 indicates alteration in the transactivation role of p53 and its role in cell cycle arrest and apoptosis after it is modified by polyAR. Subsequently, a multiprotein complex (including XRCC1 and possibly DNA polymerase β and DNA ligase III) is recruited to the site by PARP, and repair patch synthesis and DNA ligation complete the process of BER (Fig. 2).

Individuals with Cockayne Syndrome show premature aging. Mutant cells with a defect in helicase motifs V or VI of Cockayne syndrome B (CSB) are defective in BER of one common oxidized base, 8-hydroxyguanine, but not defective in BER of two other oxidized bases, thymine glycol or 5-hydroxy-dCytosine. CSB is involved in the specific glycosylase step for BER of 8-hydroxyguanine (Fig. 2).

Centenarians (individuals who have lived for more than 100 years) appear to have an altered PARP with higher specific

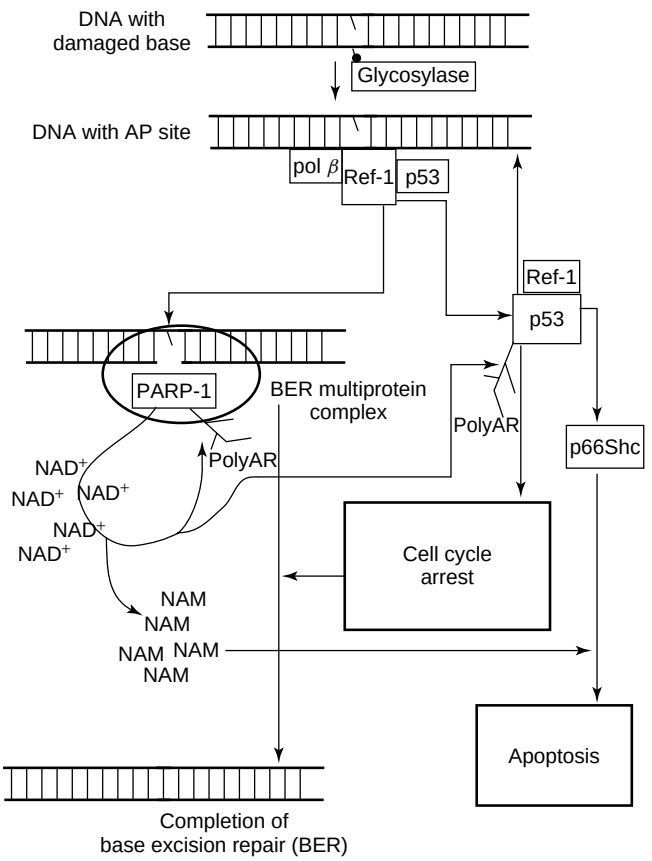

Fig. 2 Base excision repair (BER).

activity (Table 2). PARP has a central role in recruitment of the BER multiprotein complex (Fig. 2). p53 stimulates BER through interaction with ref-1 and DNA polymerase β (Fig. 2). Mice defective in p53 show early aging (Table 3).

p66Shc, when absent, causes life span extension of mice raised under protected laboratory conditions, and also longer survival of mice challenged by the oxidatively damaging compound paraquat. Mouse embryo fibroblast cells that are defective in p66Shc, when treated with H_2O_2, have much less apoptosis than wild-type cells treated with H_2O_2. This would be expected from the known role of p66Shc in the p53-controlled apoptosis pathway shown in Fig. 2. In addition, a defect in p66Shc causes reduced intracellular constitutive levels of ROS in lung, spleen, liver, and skin cells, where p66Shc is normally active. These reduced ROS levels are reflected in lower levels of oxidative DNA damage both in nuclear DNA and mitochondrial DNA of these tissues (both types of DNA damage normally repaired by BER).

Thus, two gene products involved in carrying out BER (CSB and p53), when defective, cause early aging. One gene product essential for BER, PARP, when more active, allows longer life span. One gene product, p66Shc, causes more DNA damages, needing BER when it is active. When p66Shc is absent, there are fewer DNA damages and there is life span extension.

2.3 HRR (Homologous Recombinational Repair)

Double-strand breaks, interstrand cross-links, and DNA damages blocking a replication fork can be repaired by one of two pathways, HRR or NHEJ. The selection of pathway depends on whether the cell has replicated its DNA (the S/G_2 phase of the cell cycle), so that sister chromatids are available (which favors HRR), or whether the cell is in G_0/G_1 (which favors NHEJ). The choice of HRR or NHEJ also depends on whether the cell is from embryonic tissue (which favors HRR) or from adult tissue (which, for certain damages, favors NHEJ). While HRR is especially promoted during meiosis (where there is ordinarily pairing and recombination between homologous chromosomes), HRR is also important in somatic cells. Defects in enzymes of HRR, including WRN (Werner syndrome), the RECQ3 helicase, and BLM (Bloom syndrome), the RECQ2 helicase, cause genetic instability and cancer (Table 3).

HRR is initiated through a multimeric complex, the BRCA1-associated genome surveillance complex (BASC). BASC includes the DNA repair proteins BRCA1, MSH2, MSH6, MLH1, ATM, BLM, and the RAD50-MRE11-NBS1 protein complex, some of which (BRCA1, ATM, and BLM) are indicated in Fig. 3. BASC is thought to act as a sensor for DNA damage, to which it binds. Depending on the type or amount of double-strand damage, BASC signals for (1) initiation of further steps of HRR, (2) interaction with the "master switch p53" to turn on cell cycle arrest (to allow more time for HRR), (3) apoptosis through the p53/p66Shc pathway, or (4) apoptosis through the p73 pathway (Fig. 3). As indicated in Fig. 3, BRCA1 and p53 have a type of feedback loop whereby BRCA1 may activate p53, and p53 may, in turn, inhibit BRCA1.

Further steps of HRR, shown in Fig. 3, involve (1) recruitment of a second homologous chromosome, (2) Rad51-dependent

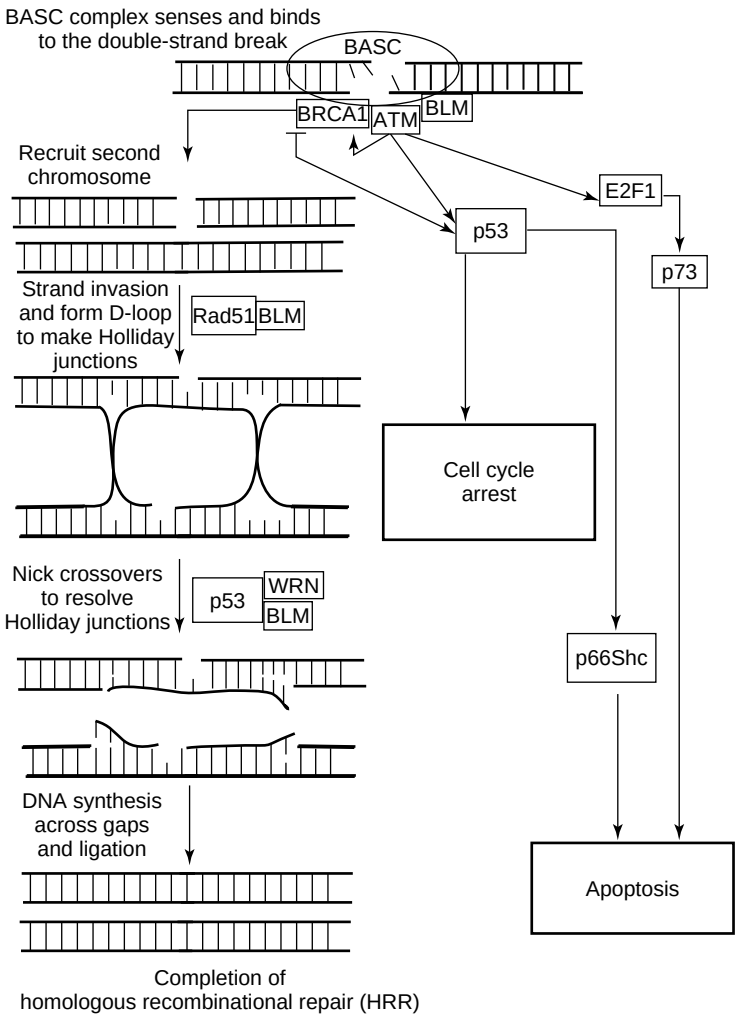

Fig. 3 Homologous recombinational repair (HRR).

strand invasion from the ends of the double-strand break plus formation of a "D-loop" in the invaded chromosome by strand migration facilitated by the helicase action of BLM, (3) nicking of the crossover strands to resolve the Holliday junctions, carried out by BLM and WRN helicases in association with p53 (which has a 3′ to 5′ exonuclease function and a strand transfer function that may act at this step), and (4) DNA synthesis across single-strand gaps and a final ligation step. An interstrand cross-link will be converted to a double-strand break, in part by action of ERCC1 (Table 3), and thus can also be repaired by this pathway.

Modified forms of HRR have been suggested for replicating somatic cells. In these models, the "D-loop" (the strand of DNA from the intact chromosome

that has been unwound and looped out to lend its information to cover the gap in the broken chromosome) is not cut to resolve the crossovers. These models, called the *synthesis-dependent strand annealing* or *migrating D-loop models*, were proposed to account for the prominence of mitotic gene conversion without reciprocal exchange observed in replicating cells. The standard model of HRR is shown in Fig. 3, however, since it applies to nonreplicating somatic cells, the major cells showing decreased function with age, and to germ line cells, as discussed below.

As indicated in Table 3, four genes that act directly in HRR (*BLM*, *WRN*, *p53*, and *ERCC1*), when defective, cause early aging. Oxidative damage to DNA includes double-strand breaks. Thus the gene product, *p66Shc*, which increases the level of intracellular ROS, may cause more DNA damage, needing HRR when it is active. When *p66Shc* is absent, there are fewer such DNA damages and there is life span extension.

2.4 NHEJ (Nonhomologous End Joining)

As noted in Section 2.3, double-strand breaks, interstrand cross-links, and DNA damages blocking a replication fork can be repaired by one of two pathways, HRR or NHEJ. HRR is an accurate repair pathway, obtaining information missing within a damaged chromosome from a homologous, undamaged chromosome. NHEJ, alone, however, is inaccurate at the nucleotide level because it involves end-joining reactions between single-stranded ends to form junctions containing regions of microhomology of 1 to 10 base pairs within 20 base pairs of the ends. However, coupled homologous and nonhomologous repair can occur (using short homologous regions on a nearby heterologous chromosome) to employ NHEJ in a more accurate process, with short regions of gene conversion from the undamaged chromosome. In addition, it was recently shown that NHEJ, when the number of double-strand breaks is small, serves to properly reconnect the two broken ends of a given chromosome. In the absence of NHEJ, more misrejoined chromosomes (chromosomal aberrations) are produced.

NHEJ of a double-strand break is illustrated in Fig. 4. Two Ku heterodimers (Ku consists of Ku70 complexed with Ku86) attach to two broken ends of a chromosome. Ku then recruits DNA-PK_{cs}, which becomes activated by the interaction with Ku and a DNA end. Activated DNA-PK_{cs} phosphorylates a wide range of DNA binding proteins and also phosphorylates Werner syndrome protein (WRN). WRN displaces DNA-PK_{cs} and, through interaction with Ku, activates its WRN exonuclease function, trimming back single strands on the chromosome ends. The complex of Mre11, Rad50, and NBS1 (MRN), plus BRCA1, may interact at this point with Ku to keep the end joining fairly accurate. In the absence of WRN, frequent gross chromosomal rearrangements occur. The Bloom syndrome protein, BLM, is also needed to keep the NHEJ fairly accurate. BLM may also assist with the alignment of two nearby ends for fairly accurate NHEJ (BLM is drawn with a dashed oval in Fig. 4 to indicate that its role in the NHEJ repair process is not yet clearly established). Pol μ then associates with small gaps in the aligned and stabilized joined ends, and pol μ, in association with Ku, recruits ligase IV and XRCC4 (X-ray repair cross complementing protein 4). Pol μ fills in

the gaps and ligase IV and XRCC4 complete the NHEJ repair of the double-strand break in a chromosome.

As indicated in Table 3, three genes that act directly in NHEJ [*Ku80* (also called *Ku86*), *WRN* and *BLM*], when defective, cause early aging. *Ku80* is only active in NHEJ, while *WRN* and *BLM* also have roles in HRR.

2.5 MGMT (O^6-Methylguanine-DNA Methyltransferase)

As indicated in Table 1, O^6-methylguanine is a frequent DNA damage. O^6-methylguanine is specifically repaired by a DNA repair enzyme called *O^6-methylguanine-DNA methyltransferase* (MGMT). As

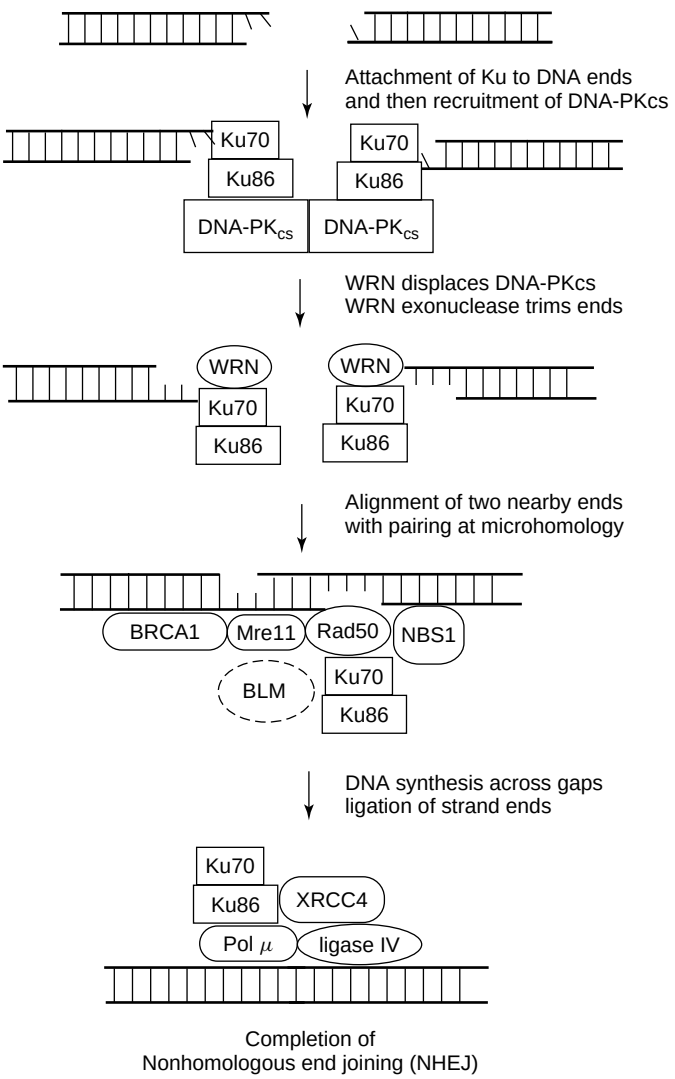

Fig. 4 Nonhomologous end joining (NHEJ).

indicated in Table 2, excess MGMT results in decreased aging (life span extension).

2.6
Enzymes of DNA Repair Pathways and Enzymes Regulating DNA Damage–inducing ROS Contribute to Determination of Aging

Eleven specific DNA repair enzymes distributed among NER, BER, HRR, NHEJ and MGMT were found to increase lifespan when elevated or decrease life span when defective (Tables 2 and 3). In addition, alterations of four enzymes causing reduced levels of DNA-damaging ROS (increases of Cu/ZnSOD, MnSOD or catalase, or loss of p66Shc) also increase life span. These findings strongly support the DNA damage theory of aging.

3
The DNA Repair (and Complementation) Theory of Sex

3.1
Meiosis, an Adaptation for Repairing Germ Line DNA

It is useful for further discussion to define sexual reproduction and to clarify its key components. Such a definition, to be general, should encompass sexual processes in all organisms, both eukaryotes and prokaryotes. *Sexual reproduction is the process by which genetic material (usually DNA) from two separate parents is brought together in a common cytoplasm where recombination of the genetic material ordinarily occurs, followed by the passage of the recombined genome(s) to progeny.* Thus, sexual reproduction has two basic elements: (1) recombination, in the sense of the exchange of genetic material between two homologous chromosomes, and (2) outcrossing, in the sense that the homologous chromosomes from two different individuals come together in the same cell.

In eukaryotes, the germ line is the cellular lineage that connects successive meioses. The germ line is a key feature of sexual reproduction. In obligate sexual organisms such as humans, the germ line is potentially immortal (see Section 1.9) in contrast with the somatic line. This feature of the germ line presumably is due to adaptations present in the germ line that are lacking in the somatic line. Meiosis is a process unique to the germ line, and it appears to be an adaptation specifically for promoting recombinational repair (HRR), since a major characteristic of meiosis is the close pairing of homologous chromosomes.

Some diploid organisms form gametes (such as sperm and egg cells) by meiosis, but then undergo self-fertilization or automixis (a process in which two haploid products of meiosis fuse to form a diploid zygote). Such processes are common in plants and invertebrates. Self-fertilization and automixis are not strictly sexual processes since they lack the outcrossing feature of sex. However, many of the organisms that undergo these asexual processes are also facultatively sexual. In the following discussion, we assume that meiosis serves essentially the same function in all organisms in which it occurs and that the germ line in facultative or obligate self-fertilizing and automictic eukaryotes is the cell lineage that connects successive meioses.

Tab. 5 Modes of reproduction among higher plant species.

Reproduction strategies	*Number*	*Percent*
Principally cross-fertilized	830	55
Self-incompatible, dichogamous, or dioecious		
Dichogamy: male and female parts mature at different times;		
Dioecious: unisexual (male or female) flowers on different plants		
Partially self-fertilized, partially cross-fertilized	105	7
Principally self-fertilized (autogamous)	229	15
Facultative apomicts (can be cross-fertilized)	16	1
Apomixis: vegetative methods of reproduction including propagation by runners or bulbs and agamospermy		
Agamospermy: embryos and seeds formed by asexual means		
Facultative apomicts, can be self- or cross-fertilized	3	0.2
Facultative apomicts, can be self-fertilized	2	0.1
Apomicts, not known if also self- or cross-fertilized	199	13
Obligate apomicts	121	8

3.2 Frequency of Sexual Reproduction

Sexual reproduction is a widespread strategy for reproduction. About 99.9% of the approximately one million known animal species are sexual. Among higher plants, the majority of species are sexual. As shown in Table 5, only 8% of higher plants are known to be obligate apomicts that reproduce only by vegetative means such as by runners, bulbs, or by asexual formation of seeds. Sex is also common among the simple eukaryotes, including fungi, algae, and protozoa. Sex is found among bacterial species as well, and is common in bacterial viruses and animal viruses.

3.3 Costs of Sex

Sex, while widespread, is very costly to the organism using it. For example, a sexual female lizard passes only 50% of her genes to a particular egg, while a comparable nonsexual (parthenogenetic) female lizard passes 100% of her genes to each egg. Thus, a sexual female is only half as efficient in propagating her genes as a nonsexual female (all other factors being equal). In addition, when two individuals must find each other to mate, there is a cost of searching out the other party.

Another cost of sex arises from the randomization of genetic information during meiosis. A parent organism, which has met the test of survival, has by definition, a well-adapted combination of genes. The process of meiosis, which includes recombination, generates untested new combinations of genes to be passed on to progeny. These new combinations, on average, should be less successful than the parental combinations of genes because random changes in successful genetic information are more likely to be deleterious than beneficial.

The noted biologist E.O. Wilson describes sex in humans as a "gratutiously consuming and risky activity". Reproductive organs are anatomically

complex in humans and these are vulnerable to such problems as ectopic pregnancy and venereal diseases. Further, courtship activities are costly. Even at the microscopic level in humans, genetic processes for determining sexual development are easily perturbed with one sex chromosome too few or too many, or a shift in the hormone balance of a fetus, causing abnormalities in physiology and behavior.

Clearly sex has large costs.

3.4 Benefits of Sex

Sex must have a large benefit to make up for its large costs. A major function of sex is to counteract two types of "noise" in the transmission of genetic information from parent to progeny: DNA damage and mutation.

As pointed out in Section 1.1, about 95,000 DNA damages occur, on average, per day per cell in the rat. Most of these are single-strand damages that can be removed by excision repair or other repair processes that only need the redundant information present on the opposite strand of DNA. However, excision repair of single-strand damages is not 100% efficient, and double-strand damages also occur at significant frequencies (Table 1). Such DNA damages remaining in germ cells (e.g. egg and sperm, in mammals) would cause the death of zygotes and loss of potential progeny. During the meiotic stage of the sexual process, however, HRR is strongly enhanced by the systematic homologous pairing of chromosomes, the major feature of meiosis. This enhanced HRR repair is one major benefit that can compensate for the costs of sex, since clearing the germ line of lingering damages can greatly increase viability of progeny.

The other type of noise in the transmission of genetic information from parent to progeny is mutation. Mutation is quite different from DNA damage. A mutation is a change in the DNA sequence rather than a change to a deformed DNA structure (DNA damage). Mutation and DNA damage have distinctly different consequences. Mutations can be replicated (when DNA replicates) and thus can be inherited, while a damage cannot be replicated. However, damages can be recognized by enzymes and repaired, whereas mutation cannot be recognized and thus cannot be repaired.

Genes carrying mutations often code for nonfunctional proteins. If one of a pair of chromosomes in a diploid cell carries a mutation in a given gene, and the second chromosome carries the homologous gene in a functional form, this second gene can usually provide an adequate level of gene expression for the organism to function normally. This masking of the expression of mutant genes by wild-type genes is called *complementation*. Complementation is available in the diploid phase of the life cycle of organisms. However, complementation is most beneficial when an organism undergoes outcrossing.

To see why mutation makes it beneficial to have outcrossing, consider a hypothetical population of diploid organisms that is strictly inbreeding (all of the organisms are self-fertilizing) and assume that the population has been long established. In such a population, the rate at which new mutations arise will be balanced by the rate at which they are lost from the population by natural selection. In this self-fertilizing population, each mutation present in an individual will have a one in four chance of being paired with

the same mutation (becoming homozygous recessive) in each progeny. If an average of one to several deleterious mutations are present in each individual, the cost of inbreeding in terms of defective progeny would be high. If a hypothetical outcrossing individual should arise in such an otherwise inbreeding population, any mutations in this outcrosser would very likely be complemented by the wild-type alleles from its mating partner and thus defective progeny would be avoided. Complementation would occur because the homologous chromosomes from the two parents are not likely to carry the same mutations. Thus, loss of progeny due to expression of mutations would be greatly reduced. This gives a strong immediate selective advantage to switching to outcrossing from inbreeding.

This advantage would not last indefinitely in our hypothetical example. Because of the ability of outcrossers to mask deleterious mutations by complementation, mutations would not be weeded out by natural selection as efficiently as in inbreeding individuals. Eventually, the mutations that build up in the population will cause as much lethality to progeny of the outcrosser as that in the original inbreeding population. However, if the outcrosser tried to switch back to inbreeding, there would be a great loss of progeny due to the larger number of mutations now present. In summary, there is a large immediate benefit in switching from inbreeding to outcrossing and a large immediate disadvantage in switching back. Therefore, mutation provides a selective pressure to maintain the outcrossing feature of sexual reproduction among diploids.

Overall, meiosis, with its promotion of recombinational repair, may be the only way to efficiently correct endogenous double-strand damages and leftover single-strand damages in the diploid cells that produce germ cells. On this view, the recombination aspect of sex is an adaptation for dealing with DNA damage, a major type of genetic noise. Furthermore, outcrossing allows the masking of mutations through complementation. The outcrossing aspect of sex deals with mutation, the second major type of genetic noise.

3.5 Repair of DNA Damage of the Germ Line in Nonmeiotic Cell Divisions

The germ line is characterized by periodic events of meiosis, but during the intervals between such events, cell divisions are ordinarily by mitosis. Cells of the germ line are presumably capable of the same types of repair processes that occur in somatic cells, discussed in Section 2. For instance, there is a relatively high level of PARP activity in premeiotic and meiotic spermatocytes, an indication of BER. Fraga and coworkers in 1990 measured the accumulation of one type of oxidatively damaged DNA base, 8-hydroxy-2′-deoxyguanosine, in various tissues of the rat. Although the average accumulation in the rat kidney was 80 residues per cell per day, there was no detectable accumulation in the testes. This lack of accumulation in the testes of normal individuals can be interpreted as a reflection of efficient BER in the germ line of normal individuals.

It is notable, in Table 2, that where life span is extended, fertility is either normal or extended as well, where it has been tested. In Table 3, where defects in DNA repair cause early aging, fertility is always reduced, where it has been tested.

3.6 Meiotic Recombination in Eukaryotes Probably Evolved from Recombination in Prokaryotes

Several workers have suggested that the processes of recombination in prokaryotes and eukaryotes share a common ancestry. Dougherty in 1955 may have been the first to conclude that the evolution of sexuality as it exists today was the result of a single phylogenetic sequence. He based this conclusion on the fact that recombination in bacteriophage and in bacteria, on the one hand, and meiotic recombination in eukaryotes, on the other hand, seemed to share fundamental similarities. Later Stahl in 1979 also concluded that despite numerous differences in detail, the "similarities in recombination in creatures as diverse as the phage and fungi are impressive."

In recent years, there has been considerable work on the biochemistry of recombination, with an emphasis on the RecA protein of the bacterium *Escherichia coli*. This protein catalyzes the key steps in recombination of homologous DNA pairing and strand exchange. Homologs of the *E. coli RecA* gene have been identified in over 60 bacterial species and in bacteriophage T4, suggesting that RecA-catalyzed recombination is very common in the prokaryotic world. Since about 1992, there has been much work indicating that RecA homologs play a key role in meiotic recombination in fungi and vertebrates. A RecA homolog in humans shows 30% amino acid sequence identity with the *E. coli* RecA protein. RecA homologs in yeast and humans form helical filaments with DNA, such as those formed by *E. coli* RecA, leading to the conclusion that the RecA protein has been conserved from bacteria to man. These findings suggest that eukaryotic meiotic recombination and recombination processes in extant bacteria are probably both derived from a common ancestor that existed before the divergence of prokaryotes and eukaryotes, at least 1.8 billion years ago.

3.7 Homologs of the Bacterial RecA Protein Have a Key Role in Eukaryotic Meiotic Recombinational Repair

Substantial evidence indicates that RecA homologs have a central role in recombinational repair during meiosis in yeast. The yeast *rad*51 and *dmc1* genes are homologs of the *recA* gene of *E. coli*. The Rad51 and Dmc1 proteins probably share redundant functions, since recombination is reduced by only a few fold in *dmc1* and *rad51* single mutants, but *dmc1* and *rad51* double mutants are profoundly defective in meiotic recombination. The Rad51 protein acts during both mitosis and meiosis, whereas Dmc1 protein acts only during meiosis. Both Rad51 and Dmc1 proteins function in repair of double-strand breaks. Sung in 1994 demonstrated that Rad51 protein, like *E. coli* RecA protein, catalyzes ATP-dependent homologous DNA pairing and strand exchange. The Dmc1 protein has an overall similarity to *E. coli* RecA protein in tertiary structure. These findings suggest that the yeast Rad51 and the bacterial RecA proteins have similar functions in recombinational repair. Morita and coworkers showed in 1993 that in the mouse a *recA* homolog is expressed at high level in the testes. Shinohara and coworkers in 1993 also showed that the mouse *recA* homolog is expressed at a high level in the testis and ovary, and suggested that the protein product of this gene is involved in meiotic recombination. The expression of a *recA* gene homolog has also been demonstrated in chicken testis and ovary and in human

testes. The human RecA homolog carries out the distinctive reactions of *E. coli* RecA protein, including DNA-dependent hydrolysis of ATP, renaturation of complementary strands, homologous pairing of a single strand with duplex DNA, and strand exchange. A pair of *recA* gene homologs of the lily, *lim15* and *rad51*, which are also homologs of the yeast *dmc1* and *rad51* genes, were found by Terasawa and coworkers in 1995. The *lim15* gene in lily is specifically expressed in meiotic prophase during microsporogenesis. Thus, in both animals and plants, homologs of bacterial *recA* appear to have an important role in meiotic recombination.

3.8 The Adaptive Function of Recombination Appears to Be Removal of DNA Damage

DNA repair is probably the principal adaptive function of the RecA protein and its homologs. RecA protein binding is largely limited to regions in the DNA containing suitable nucleation sites, especially single-strand gaps. A variety of DNA damages cause structural perturbations that provide favorable nucleation sites.

About 100 ATPs are hydrolyzed for every base pair of heteroduplex DNA generated by RecA-mediated strand exchange. In 1993, Cox reviewed evidence that the energy released by ATP hydrolysis is used specifically to allow the strand-exchange process to traverse damaged regions of DNA. He argued that this use of ATP is readily understood as an adaptation of repair of DNA damage. Thus, the evidence reviewed by Cox indicates that the adaptive function of RecA homologs, acting during meiosis, is DNA repair.

In humans, HRR is carried out by a group of interacting proteins in which the RecA homolog Rad51 has a central role. Other proteins involved in HRR in humans include the products of genes *BRCA1*, *BRCA2*, *ATM*, *ATR*, *FANCD2*, *HMLH1*, and *p53*.

These proteins appear to be responsible for HRR during meiosis and for HRR between sister DNA homologs in somatic cells. HRR is employed in removing a variety of DNA damages, particularly double-strand damages such as double-strand breaks and DNA cross-links. The proteins named just above, involved in HRR, also appear to use their DNA damage recognition capability to induce apoptosis (a form of programmed cell death) when the number of DNA damages in a cell is higher than those that can be repaired. In somatic cells, apoptosis is a protective mechanism for the whole organism since it eliminates cells that might otherwise survive despite having unrepaired DNA damages, and upon replication acquire mutations.

In fact, germ line mutations in genes *BRCA1*, *BRCA2*, *ATM*, *ATR*, *FANCD2*, *HMLH1*, and *p53* predispose individuals to cancer. Cancer predisposition results from an inadequate response to DNA damage in somatic cells leading to increased mutations, some of which cause progression to malignancy. Thus, the central function of these genes is to cope with DNA damages, either to repair the damages, or if their number in a cell is unmanageable, to induce apoptosis.

Males that are defective in *ATM* or *p53* are unable to respond appropriately to DNA damage during meiosis, and, as a likely result, they suffer from low sperm quantity and quality. The high effectiveness of DNA damage removal processes during normal meiosis is reflected in mice, where there is a one-third lower rate of mutation in germ cells compared to that in somatic cells.

Overall, the evidence reviewed in this section lends strong support to the idea that the primary general function of recombination enzymes, whether in somatic cells or during meiosis, is to remove DNA damage, either by repair or apoptosis.

3.9 Other Expectations of the DNA Repair and Complementation Theory of Sex

If DNA damage and its repair are important in maintaining sexual reproduction, then certain expectations follow. If recombination during meiosis and sexual reproduction reflect recombinational repair of germ line DNA, then there should be other evidence of avoidance of DNA damage in the germ line as well.

As indicated in Table 1, the largest known source of DNA damage is oxidative damage. Such damage occurs due to endogenous cellular metabolism. Presumably, to avoid DNA damage, germ line cells should have evolved ways to avoid high levels of metabolism. Eggs would seem, at first sight, to be poor candidates for avoiding metabolism. They are, in general, much larger than somatic cells of the organism (e.g. egg cells have about a 1,000-fold greater mass than somatic cells in humans). It requires considerable metabolism to produce large egg cells. However, much of the cytoplasmic material within an egg cell is generated by the activity of other cells. Some insects have nurse cells around each egg cell. These nurse cells are connected to the egg cell by cytoplasmic bridges and provide most of the ribosomes, mRNA, and proteins of the egg cell. The nurse cells themselves contain hundreds to thousands of copies of their genomes, presumably to protect the nurse cells themselves from losing function from the oxidative damage they may suffer while providing large amounts of metabolic products for the egg cell. Vertebrate egg cells are surrounded by follicle cells rather than nurse cells. The follicle cells do not have cytoplasmic bridges to the egg cell, but rather have small gap junctions connecting them to the egg. While these gap junctions are not large enough to transmit bulky macromolecules, they do transmit precursor molecules to the egg. In addition, for chickens, amphibians, and insects, the yolk proteins accumulated by the egg are made in liver or liver-type cells. These mechanisms allow eggs to be protected from oxidative damage while they store up material to sustain the zygote in its initial growth.

Sperm or pollen cells, in contrast to egg cells, are usually the smallest cells of an animal or plant. This allows a different strategy for effective protection against oxidative damage to their DNA. Because of their very small size, minimal metabolism would have been used in their formation. Thus, both egg and sperm production appears to have been adapted to circumvent the production of DNA damage in their especially important germ line DNA.

Another way to avoid production of sperm with damaged DNA is for spermatogenic cells with DNA damage to undergo cell cycle arrest to allow more time for meiotic repair, and then, if this fails, to undergo apoptosis. The p53 protein plays a key role in mediating cell cycle arrest in response to DNA damage. Low-level irradiation was found to activate a p53-dependent premeiotic delay, allowing time for increased DNA repair leading to increased motile spermatozoa. Higher levels of γ-irradiation induced p53-independent apoptosis during meiosis.

If complementation of mutations is important in maintaining the outcrossing

aspect of sex, this should be consistent with general biological observations as well. Indeed, both in animals and in plants, it is usually seen that when hybrids are formed from the crossing of two genetically distinct inbred lines, these hybrids are more vigorous than either of their two parental lines. This hybrid vigor is responsible for much of the crop improvement that has been achieved in modern agriculture. The opposite side of this observation is the fact that consanguineous marriages, in humans, result in an increased frequency of impaired offspring. Observations in other animals and in plants suggest that close inbreeding results in the production of less vigorous progeny. This inbreeding depression appears to be due largely to expression of deleterious recessive mutations and reflects in part the cumulative effect of numerous mildly deleterious mutations.

4
Vegetative Survival Strategies

4.1
Survival of Vegetative Cell Populations

Vegetatively growing populations of bacteria can be regarded as potentially immortal as long as nutrient resources are abundant. However, even in such populations there appears to be constant attrition due to DNA damage and deleterious mutation.

Flowering plants generally reproduce sexually, with gamete formation by a meiotic process followed by gamete fusion as the prelude to embryogenesis and seed formation. Plants, unlike most animals, are also able to generate complete new individuals of similar genetic constitution from vegetative parts. Meristematic buds or excised pieces of tissue can propagate in the appropriate environment. Apparently, plant vegetative cell lines can be maintained indefinitely under appropriate conditions. These lines probably maintain themselves by a strategy of replacement, where cells with lethal unrepaired DNA damage or expressed deleterious mutations die and are replaced by replication of nondefective cells. Nevertheless, in some plant tissues DNA damage may accumulate. For instance, in dry seeds, fragmentation of nuclear DNA occurs with time.

Most forest trees live for at least 100 years, many of them for more than 300 years, and a few survive for more than 1,000 years. Clonal tree species may occupy a location for several thousand years. Most of the tree is dead, and only a thin shell of dividing cells (cambium) occurs around the trunk and in the leaves. A tree actually represents a free-living clone of cells in which selective removal of cells with irreversible accumulated damage is constantly occurring. One would not expect to find old cells in a tree any more than one would find old cells in a growing culture of bacteria. The evidence discussed in this section suggests that some proliferating cell populations can cope with unrepaired DNA damage by a replacement strategy, which can be maintained indefinitely as long as nutrient resources are abundant and the level of unrepaired DNA damage is not excessive.

4.2
Vegetative Survival Strategy for Mitochondria (and Chloroplasts) in the Germ Line

An analog of cellular replacement at the molecular level ("molecular replacement") facilitates the purging of both DNA damage and deleterious mutations in cytoplasmic genomes (mitochondria and

chloroplasts) of germ cell lines. While nuclear genes occur as single copies per gamete, there are thousands of mitochondrial DNA (mtDNA) molecules in most cells, and several hundred thousand may occur in a mature oocyte. The many mtDNAs in oocytes appear to stem from a vastly smaller pool of mtDNA molecules that must have survived a process of replicative segregation in earlier cytokinetic divisions of the germ line lineage, since most heterogeneity of mtDNA is distributed among, rather than within, individuals. This implies that there are mtDNA population bottlenecks in germ lines. That is, the mtDNA, which is generally solely transmitted from one generation to the next through the oocyte, is of only one or a few genotypes. Avise concluded that the mtDNA molecules that survive and replicate to populate a mature oocyte probably have been scrupulously screened by natural selection for replicative capacity and functional competence in the germ cell lineages they inhabit. This strategy, at the molecular level, is equivalent to a vegetative cell replacement strategy for cell lineages that do not have a sexual cycle. While the mitochondria and chloroplasts within cells follow a vegetative replacement strategy, they, like vegetatively replicating bacteria and yeast, utilize homologous DNA recombinational repair to remove DNA damages due to oxidation or environmental stresses.

4.3 Dolly the Cloned Lamb, Cumulina the Cloned Mouse, and Low Success of Cloned Mammals

The evidence reviewed above indicates that germ cells, having undergone meiosis, are relatively free of DNA damage, and thus should be able to give rise to viable offsprings with high probability. At lower probability, a given somatic cell nucleus may also be sufficiently free of damage to be able to produce viable progeny if transferred to an enucleated egg cell.

Eight viable lambs were derived in 1997 – one (named Dolly) from donor cells of a mature sheep, four from donor embryo-derived cells, and three from donor fetal fibroblast cells. For Dolly, the lamb derived from a somatic cell of a mature sheep, they first made 277 fused couplets (enucleated oocytes fused to donor cells), using donor cells from their 3rd to 6th passage, cultured from the mammary gland of a six-year-old pregnant ewe. A morula or blastocyst was able to form from only 11.7% of the cultured fused couplets from the mature ewe. By comparison, when couplets were derived from embryo or fetal fibroblast cells, a morula/blastocyst was formed 27 to 39% of the time. Some of the morula/blastocysts implanted in recipient ewes formed fetuses detectable by ultrasound at 50 to 60 days. Subsequently, 62% of these fetuses were lost, a much greater proportion than the estimated 6% after natural (meiosis based) mating in sheep.

Similar observations were made with mice. Ten healthy mice were cloned, the first of which was named Cumulina, from donor nuclei of differentiated, nonreplicating granulosa cells of mature mice injected into enucleated recipient mouse oocytes. Cumulina and the other nine cloned healthy newborn mice were the only successful progeny produced from 800 injected oocytes, which had formed embryos and were transplanted into foster mother recipient mice. In multiple cloning experiments by these authors, the rate of successful implantation of embryos was 57

to 71%, formation of fetuses was 5 to 16%, and full-term development was 2 to 3%.

The low survival rate to full development in both sheep and mice could have three possible explanations. First, injuries introduced by the experimental manipulation of the embryos before implantation might be deleterious to further development. Even unreconstructed embryos experience some increased prenatal loss after manipulation or culture. Second, the differentiated donor nuclei, transplanted into the recipient oocytes, have to reprogram their developmental clock to zero, and errors in this process may have deleterious effects on fetal development. Third, the low survival of fetuses may be a consequence of the greater amount of DNA damage in donor somatic cells (than in meiotically produced gametes), which can give rise to deleterious mutations when replicating after transfer to an oocyte.

The gene expression profile of about 10,000 genes in the placentas and the livers of surviving cloned mice derived by nuclear transfer were obtained in 2002. The transferred nucleus came from either an embryonic stem (ES) cell or from a mature, differentiated cumulus cell. They compared those gene expression profiles (derived from ES nucleus or cumulus cell nucleus) with the gene expression profile in the same tissues (placenta, liver) of healthy mice derived from normal mating. These comparisons showed that several hundred (at least 4%) of the expressed genes had pronounced dysregulation in the cloned mice, probably accounting for the altered phenotype of cloned individuals (many cloned mice were obese, etc.). Their evidence pointed to difficulty in reprogramming the developmental clock back to zero, in the low percentage of surviving mice.

In humans, about 50 to 80% of all natural meiosis-based conceptions fail to result in live birth. Cytogenetic studies of spontaneous and induced abortions and on perinatal deaths indicate that many types of chromosome abnormalities are present in these failed conceptions. Chromosome abnormalities often derive from DNA damages.

Thus, DNA damage may be a serious problem for germ cells despite available mechanisms for avoiding and repairing such damage, and this problem may be considerably greater in nonmeiosis-derived conceptions.

5 Three Levels of Sexual Communication Reflect (1) DNA Repair, (2) Complementation, and (3) Selection for Fitness

Sexual communication occurs when signals are used to promote or modulate sexual interaction between individuals. Sexual communication is prevalent among organisms from bacteria to man. Sexual communication occurs at three levels. Level 1: the first level of sexual communication includes signals that increase the likelihood that two organisms will come together for sexual interaction. Level 2: the second level involves signals that modulate the sexual interactions to inhibit inbreeding or facilitate outbreeding. Level 3: the third level includes signals that further modulate the sexual interactions to promote selection among potential mating partners based on relative fitness. Evidence indicates that the selective advantages of the three levels of sexual communication are, respectively, (1) the repair of DNA damage, (2) the masking of mutation, and (3) the choice of a fit mating partner.

5.1 Sexual Communication in Bacteria, Primarily for DNA Repair

The simplest organisms in which sexual communication occurs are bacteria. Bacterial transformation is a form of sexual interaction in bacteria and has been shown in several bacterial species to involve sexual communication. Transformation involves the transfer of naked DNA from one member of a bacterial population to another through the surrounding medium. Transformation occurs naturally in a wide range of bacterial species. The bacteria that take up transforming DNA ordinarily incorporate this DNA into their genomes by recombination. Transformation is an evolved, rather than an incidental, trait since it results from a complex, energy-requiring, developmental process. For a bacterium to bind, take up, and recombine exogenous DNA into its genome, it must first enter a special physiological state referred to as *competence*. Transformation involves expression of genes required for competence and for recombination. In the completely sequenced genome of *Haemophilus influenzae*, transformation is promoted by at least 15 genes of the 1007 for which role assignments can be made. In another bacterium, *Bacillus subtilis*, about 40 genes necessary for competence have been identified.

Transformation in a number of bacterial species depends on the production of an extracellular factor (referred to as *competence factor*), which, upon release into the surrounding medium, induces the competent state in neighboring cells in the population. These extracellular factors are examples of pheromones. (Pheromones are molecules released by an organism into the external medium to influence other individuals of the same species.) These extracellular factors have been studied in *B. subtilis, Streptococcus pneumoniae,* and *B. cereus*. In *B. subtilis,* two different extracellular peptide pheromones are used for competence development. In *S. pneumoniae,* the competence pheromone is a 17 amino acid peptide.

The adaptive function of bacterial transformation has been studied in *B. subtilis* as a model system for understanding the adaptive function of bacterial transformation generally. The results indicated that transformation provides external DNA template for recombinational repair of DNA damage in the recipient. Transformation and recombinational repair in *B. subtilis* require the RecE protein, which is a homolog of the well-studied RecA protein of *E. coli*. DNA damages arising from such sources as UV irradiation, endogenous oxidative free radicals, and desiccation are a pervasive problem for bacteria. The RecE protein of *B. subtilis* probably catalyzes recombinational repair of these prevalent DNA damages. Thus, sexual communication via pheromones in bacteria is probably an evolved mechanism to promote sexual interaction for recombinational repair of prevalent DNA damages. This is "Level 1", the primary level of sexual communication.

Transformation in bacteria involves partial diploidy as an intermediate stage. Even though the initial advantage of this process, we think, was recombinational repair of DNA damage, diploidy (or partial diploidy) may also provide the benefit of masking deleterious recessive mutations through complementation. Thus, recombinational repair in bacteria may have set the stage, or may have been a precursor to, natural selection for diploidy as a distinct phase of the sexual cycle in eukaryotes. Further, recombinational variation is produced as a by-product

of recombinational repair in bacterial transformation as well as in other sexual processes. The infrequent beneficial new traits generated by recombination presumably promote evolutionary success, just as infrequent beneficial mutations do.

5.2 Sexual Communication in Fungi, Primarily for DNA Repair, but Also for Limited Complementation

1. *Sexual communication in fungi.* Among eukaryotic microorganisms, pheromones promote sexual interaction in many species. These simple eukaryotes use a variety of molecules, such as steroids, other lipids, peptides, and derivatives of organic acids, as well as large molecules such as glycoproteins as sex pheromones. Although these sex pheromones are usually transmitted through an aqueous medium, they may also be transmitted through the air as in the fungus *Mucos mucedo*.

2. *Sexual communication in yeast.* The sexual cycle and communication have been especially well described in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. Mating in these yeasts occurs between two cells of opposite mating type and is facilitated by the reciprocal action of pheromones. Cells of each mating type release pheromones that induce mating through promotion of specific modifications in cells of the opposite mating type. The two cells then conjugate, forming a diploid zygote that can undergo meiosis. This leads to sporulation and release of ascospores that can germinate to form haploid vegetative cells, thus completing the life cycle. The pheromones produced by these yeasts are short peptides.

The results of several studies bear on the adaptive advantage of the meiotic events promoted by these sex pheromones. Recombinational repair of DNA double-strand breaks occurs during meiosis in *S. cerevisiae* and *S. pombe*. The *rad51* and *dmc1* genes of *S. cerevisiae* are homologs of the *recA* gene of *E. coli*. A homolog of *E. coli* RecA also occurs in *S. pombe*.

It was found that treatment of *S. pombe* with H_2O_2, a DNA-damaging agent, caused increased mating. This stimulation of mating can be interpreted as an adaptation to promote recombinational repair of the introduced DNA damages. Thus, in *S. cerevisiae* and *S. pombe*, production of sex pheromones appears to be an adaptation for promoting recombinational repair of DNA, "Level 1" of sexual communication, as in bacteria.

In *S. pombe*, the mating pheromones M-factor (produced by the "Minus" mating type) and P-factor (produced by the "Plus" mating type) ensure sexual interaction of cells of opposite mating type. However, during vegetative growth, a cell of one mating type switches to the other mating type in about every two generations. Similarly, in *S. cerevisiae*, haploid descendants from spores of one mating type can change their mating type frequently so that after several cell divisions a diploid population develops from a single haploid spore. These observations indicate that inbreeding is avoided to some extent in *S. cerevisiae* and *S. pombe*, since mating with an individual of the opposite mating type is required, but that it is not strongly avoided since a cell of one mating type can give rise to a cell of the opposite mating type frequently, and then sibling cells can mate.

In *S. cerevisiae* and *S. pombe*, the diploid stage of the sexual cycle, formed by mating, probably expresses functions necessary for meiosis and sporulation. Recessive mutations in the genes encoding these functions may explain the observation that

inbreeding is avoided to some extent. We discussed above (Section 3.4) that the diploid stage of a life cycle serves the adaptive function of complementation, or the masking of deleterious recessive mutations. Optimal masking is achieved when the diploid zygote is formed from the union of haploid cells of two genetically unrelated individuals, so that the recessive mutations carried by each genome are most likely to be different. However, relatively few functions are probably expressed specifically in the diploid phase of *S. cerevisiae* and *S. pombe*, and this might explain why the masking of deleterious recessive mutations is only weakly promoted in these organisms.

In these yeasts, the pheromones operate at "Level 2" as well, to promote "outcrossing," in addition to acting at "Level 1" to promote mating between individuals.

3. *Sexual communication in neurospora.* *Neurospora crassa* is an ascomycete, like the yeasts *S. cerevisiae* and *S. pombe*, but its sexual phase is more elaborate than those of yeasts. *Neurospora crassa* encodes a gene *mei3*, which is homologous to *E. coli recA* and *S. cerevisiae dmc1* and *rad51*. A mutant defective in *mei3* has reduced repair of DNA damage and, when homozygous, is blocked in the zygotene stage of meiotic prophase. This implies that *N. crassa mei3*, like *S. cerevisiae dmc1* and *rad51*, is required for recombinational repair of DNA damage during meiosis.

Neurospora crassa has two stable mating types which produce sex pheromones. The sex pheromone of one mating type induces the formation of the female sexual structure, the differentiated protoperithecium, in the opposite mating type. Once a protoperithecium is formed, the vegetative conidiospores or mycelia of the opposite mating type can act as the male partner. Upon mating, a diploid nucleus is formed by fusion of male and female nuclei of opposite mating type. This diploid nucleus then undergoes meiosis, followed by mitosis, to produce eight ascospores. Each set of eight spores is enclosed within a sac, referred to as an *ascus*, and many asci are contained within the mature fruiting body, termed a *perithecium*. Genes promoting ascus and ascospore maturation, as well as perithecium development are expressed in the diploid stage. It was found that 74 of the 99 wild-collected *N. crassa* isolates from 26 populations carried one or more recessive mutations in genes that expressed in the diploid stage. They estimated that the number of genes expressed in the diploid stage is at least 435. Mutations in these genes, when homozygous in the diploid stage, cause formation of aborted asci, ascospores with maturation defects, or barren fruiting bodies with few sexual spores. The evolution of two distinct mating types, which communicate via pheromones, inhibits inbreeding in *N. crassa*, and this provides the advantage of masking recessive mutations in the numerous genes that express in the diploid stage.

Overall, in fungi, it appears that pheromones act at both the primary level of sexual communication to promote sexual interaction for meiotic recombinational repair and the secondary level to promote outcrossing, which allows masking of deleterious mutations.

5.3 Sexual Communication in Higher Eukaryotes, Primarily for Complementation while Repair Is an Automatic Concomitant of Sexual Reproduction

1. *Flowering plants.* Since the diploid stage is the most prominent stage in the life cycle of flowering plants, the advantages of complementation should be large. Thus,

the investment in sexual communication to promote cross-fertilization, as the most common mode of reproduction among flowering plants, is probably due to the prominence of the diploid stage and the benefit of masking deleterious mutations. Instances in which outcrossing has been abandoned in favor of self-fertilization or parthenogenesis may be explained by the need of some plants to survive in sparse populations in which the costs of sexual communication would be prohibitively high.

Sexual signals in plants promote mating, but the subsequent meiosis and recombinational repair happen in the succeeding generation when the progeny produce germ cells. Among those flowering plants that depend on cross-fertilization to reproduce, sexual communication acts at both the primary and the secondary level. However, in plants that can undergo self-fertilization, sexual communication is unnecessary when this option is used.

2. *Vertebrates.* In bacteria, fungi, and protozoa, sexual reproduction is generally facultative and favored only under certain conditions. In nearly all vertebrates, however, the sexual cycle is obligatory, and thus sexual communication is ordinarily necessary for reproduction.

In vertebrates, meiosis and concommitant recombinational repair ordinarily occur at a time separate from sexual signaling. However, sexual signaling to attract a mate is essential to continue the sexual cycle, which includes fertilization, production of progeny, and further meioses. Thus, sexual signals promote mating, but meiosis and recombinational repair occur subsequently in the succeeding generation, when progeny form germ cells.

Among vertebrates, sexual communication for promotion of outcrossing has been studied in many systems. In toads, advertisement vocalizations given by males apparently serve as cues by which females recognize their kin and thus avoid inbreeding.

In mouse, mating preference is strongly influenced by the major histocompatibility complex (MHC) genotype. A mouse can distinguish close relatives from more distantly related mice on the basis of MHC genotype through their sense of smell. Using this device, mice tend to outcross, avoiding mating with close relatives. Inbreeding of mice derived from wild populations has a significant detrimental effect on survivorship when the mice are reintroduced into a natural habitat. This effect is even more severe than that observed in laboratory studies of the population. In the other vertebrates, to be described below, the communication processes used to promote outcrossing have not been as well defined as in the mouse. Nevertheless, we infer that analogous processes also exist in these cases.

Outcrossing is promoted among birds. The great tit (*Parus major*) is a monogamous woodland bird. In natural populations, outcrossing is promoted by the dispersal of daughters, but not sons. Furthermore, among infrequent incestuous matings, nestling mortality was nearly double that of outbreeding pairs.

Among primates, avoidance of inbreeding occurs in rhesus monkeys, chimpanzees, and gorillas. Juvenile rhesus monkey males leave the troop that they were born into when they mature. This pattern is similar to that in many mammals in which male progeny leave the social group into which they are born and find a mate in another group. Chimpanzees and gorillas follow a less common pattern in which the female, rather than the males, leave their natal group and transfer to other groups.

Inbreeding ordinarily is avoided in humans. In humans, body odor preferences (determined by men and women scoring the odors of T-shirts previously worn by men or women) would serve to increase heterozygosity in the progeny. Prohibitions against marriage and sexual relations between close relatives are a near universal feature of human social behavior in many different cultures. This general prohibition is known as the incest taboo.

The basis for the incest taboo appears to be the avoidance of inbreeding because of the effect of the expression of deleterious mutations. There is a significantly higher incidence of major congenital malformations and postnatal mortality among the progeny of consanguineous marriages than among the progeny of nonconsanguinous marriages.

We have reviewed the evidence that "Level 2" of sexual communication, to avoid inbreeding and promote outcrossing, is common in vertebrates. Since sexual interaction in vertebrates is usually essential for reproduction, sexual communication at "Level 1" for sexual interaction is ubiquitous.

5.4 "Level 3" of Sexual Communication Is Used to Select for Fitness among Potential Mates

In addition to "Level 1" and "Level 2" functions of sexual communication, the first to promote homologous recombinational repair and the second to promote outcrossing, a third aspect of sexual communication involves courtship behavior. The adaptive function of courtship probably is selection of a mate with the characteristics that optimize reproductive success (selection for fitness). This leads to competition among individuals of the same gender. Such competition is the basis for sexual selection, regarded as having molded many features of animal behavior.

As an example, females of the guppy fish, *Poecilia reticulata*, respond more strongly to the mating displays of male guppies which have brighter orange spots. The brightness of the spots correlates positively with health and vigor and negatively with previous parasitic infection. Thus female guppies tend to select healthier mates that are more resistant to parasites.

Substantial evidence has been presented on human courtship behavior suggest that the basic biologic function of human courtship is to optimize reproductive success by the selection of a mate of the highest possible fitness.

6 Overview

The occurrence of high levels of endogenous DNA damage in mammals is now well established. DNA of mammalian somatic cells is the master informational molecule, and accumulated damages in this molecule probably cause the progressive irreversible deterioration of cell, tissue, and organ function that defines mammalian aging. When an organism forms progeny via germ cells, it is important that these cells be free of DNA damages, since such damages cause inviability. To facilitate efficient DNA repair, the redundant information available in the diploid cell can be used to replace damaged information through recombinational repair. Sexual reproduction appears to be an adaptation to promote pairing and exchange between homologous chromosomes for the purpose of efficient repair of the DNA, which is passed on to germ cells. In eukaryotes this occurs during meiosis,

while in bacteria and viruses it occurs during less complex, but similar, processes.

Mutations are another type of error in DNA in addition to damage. Unlike damages, mutations cannot be recognized by repair enzymes. Mutations, however, can be masked when information from two unrelated individuals (parents) is brought together through fertilization to form the progeny zygote. Thus, the outcrossing aspect of the sexual cycle in diploid organisms appears to be maintained by the advantage of masking mutations. Overall, aging appears to be a consequence of the accumulation of DNA damage, and sex appears to be an adaptation for the removal of damage through enhanced recombinational repair and the masking of mutations through outcrossing.

Sexual communication probably arose in bacteria as an adaptation to promote DNA transfer in order to allow recombinational repair of damages in DNA. Sexual communication is probably maintained in organisms with a distinct diploid stage of their life cycle by (1) the advantages of recombinational repair during meiosis, (2) outcrossing that facilitates complementation of recessive deleterious mutations, and (3) mate selection for fitness.

See also DNA Repair in Yeast; Female Reproductive System; Genetics, Molecular Basis of; Male Reproductive System.

Bibliography

Books and Reviews

Alexander, P. (1967) The role of DNA lesions in processes leading to aging in mice, *Symp. Soc. Exp. Biol.* **21**, 29–50.

Ames, B.N., Gold, L.S. (1991) Endogenous mutagens and the causes of aging and cancer, *Mutat. Res.* **250**, 3–16.

Arking, R. (1998) *Biology of Aging; Observations and Principles*, Sinauer Associates, Sunderland, MA.

Bernstein, C., Bernstein, H. (1991) *Aging, Sex and DNA Repair*, Academic Press, San Diego.

Bernstein, C., Bernstein, H., Payne, C. (1999) Cell Immortality: Maintenance of Cell Division Potential, in: Macieira-Coelho, A. (Ed.) *Progress in Molecular and Subcellular Biology Vol. 24; Cell Immortalization*, Springer-Verlag, Heidelberg, Germany, pp. 23–50.

Bernstein, H., Hopf, F.A., Michod, R.E. (1987) The molecular basis of the evolution of sex, *Adv. Genet.* **24**, 323–370.

Darwin, C. (1889) *The Effects of Cross and Self Fertilization in the Vegetable Kingdom*, D. Appleton & Co., New York.

Michod, R.E., Levin, B.R. (Eds.) (1988) *The Evolution of Sex: An Examination of Current Ideas*, Sinauer Associates, Sunderland, MA.

Primary Literature

Anson, R.M., Bohr, V.A. (2000) Mitochondria, oxidative DNA damage, and aging, *J. Am. Aging Assoc.* **23**, 199–218.

Arking, R., Buck, S., Novoseltev, V.N., Hwangbo, D.S., Lane, M. (2002) Genomic plasticity, energy allocations, and the extended longevity phenotypes of *Drosophila*, *Ageing Res. Rev.* **1**, 209–228.

Arking, R., Burde, V., Graves, K., Hari, R., Feldman, E., Zeevi, A., Soliman, S., Saraiya, A., Buck, S., Vettraino, J., Sathrasala, K., Wehr, N., Levine, R.L. (2000) Forward and reverse selection for longevity in *Drosophila* is characterized by alteration of antioxidant gene expression and oxidative damage patterns, *Exp. Gerontol.* **35**, 167–185.

Aufderhide, K.J. (1987) Clonal Aging in *Paramecium tetraurelia*. II Evidence of functional changes in the macronucleus with age, *Mech. Ageing Dev.* **37**, 265–279.

Avise, J.C. (1993) The evolutionary biology of aging, sexual reproduction, and DNA repair, *Evolution* **47**, 1293–1301.

Barrows, L.R., Magee, P.N. (1982) Nonenzymatic methylation of DNA by S-adenosylmethionine in vitro, *Carcinogenesis* **3**, 349–351.

Bernstein, C. (1998) Sex as a Response to Oxidative DNA Damage, in: Okefie, I.A., Halliwell, B. (Eds.) *Free Radicals and DNA: Techniques, Mechanisms and Applications*, OICA International, London, UK, pp. 99–129.

Bernstein, C., Bernstein, H. (1991) *Aging, Sex and DNA Repair*, Academic Press, San Diego.

Bernstein, C., Bernstein H. (1997) Sexual communication, *J. Theor. Biol.* **188**, 69–78.

Bernstein, C., Bernstein, H., Payne, C.M., Garewal, H. (2002) DNA repair/pro-apoptotic dual-role proteins in five major DNA repair pathways: fail-safe protection against carcinogenesis, *Mutat. Res.* **511**, 145–178.

Bernstein, C., Johns, V. (1989) Sexual reproduction as a response to H_2O_2, damage in *Schizosaccharomyces pombe*, *J. Bacteriol.* **171**, 1893–1897.

Bernstein, H., Byerly, H.C., Hopf, F.A., Michod, R.E. (1985) Genetic damage, mutation and the evolution of sex, *Science* **229**, 1277–1281.

Brooks, P.J., Wise, D.S., Berry, D.A., Kosmoski, J.V., Smerdon, M.J., Somers, R.L., Mackie, H., Spoonde, A.Y., Ackerman, E.J., Coleman, K., Tarone, R.E., Robbins, J.H. (2000) The oxidative DNA lesion 8,5′-(S)-cyclo-2′-deoxyadenosine is repaired by the nucleotide excision repair pathway and blocks gene expression in mammalian cells, *J. Biol. Chem.* **275**, 22355–22362.

Brosh, R.M. Jr., Bohr, V.A. (2002) Roles of the Werner syndrome protein in pathways required for maintenance of genetic stability, *Exp. Gerontol.* **37**, 491–506.

Bucala, R., Model, P., Cerami, A. (1984) Modification of DNA by reducing sugars: a possible mechanism for nucleic acid aging and age-related dysfunction in gene expression, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 105–109.

Bucala, R., Model, P., Russell, M., Cerami, A. (1985) Modification of DNA by glucose-6-phosphate induces DNA rearrangements in an *Escherichia coli* plasmid, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 8439–8442.

Buss, D.M. (1994) What do people desire in a mate? The evolution of human sexual strategies, *J. NIH Res.* **6**, 37–41.

Cai, Q., Tian, L, Wei, H. (1996) Age-dependent increase of indigenous DNA adducts in rat brain is associated with a lipid peroxidation product, *Exp. Gerontol.* **31**, 387–392.

Camerini-Otero, R.D., Hsieh, P. (1995) Homologous recombination proteins in prokaryotes and eukaryotes, *Annu. Rev. Genet.* **29**, 509–552.

Cathcart, R., Schwiers, E., Saul, R.L., Ames, B.N. (1984) Thymine glycol and thymidine glycol in human and rat urine: a possible assay for oxidative DNA damage, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5633–5637.

Cerutti, H., Johnson, A.M., Boynton, J.E., Gillham, N.W. (1995) Inhibition of chloroplast DNA recombination and repair by dominant negative mutants of *Escherichia coli* RecA, *Mol. Cell. Biol.* **15**, 3003–3011.

Chiarugi, A. (2002) Poly(ADP-ribose) polymerase: killer or conspirator? The 'suicide hypothesis' revisited, *Trends Pharmacol. Sci.* **23**, 122–129.

Cox, M.M. (1991) The recA protein as a recombinational repair system, *Mol. Microbiol.* **5**, 1295–1299.

Cox, M.M. (1993) Relating biochemistry to biology: how the recombinational repair function of RecA protein is manifested in its molecular properties, *Bioessays* **15**, 617–623.

Darwin, C. (1871) *The Descent of Man, and Selection in Relation to Sex*, Princeton, NJ, Reprinted 1981 by Princeton University Press.

de Boer, J., Andressoo, J.O., de Wit, J., Huijmans, J., Beems, R.B., van Steeg, H., Weeda, G., van der Horst, G.T., van Leeuwen, W., Themmen, A.P., Meradji, M., Hoeijmakers, J.H. (2002) Premature aging in mice deficient in DNA repair and transcription, *Science* **296**, 1276–1279.

Degler, C.N. (1991) *In Search of Human Nature*, Oxford University Press, New York.

Dizdaroglu, M. (1998) Mechanisms of Free Radical Damage to DNA, in: Okefie, I.A., Halliwell, B. (Eds.) *Free Radicals and DNA: Techniques, Mechanisms and Applications*, OICA International, London, UK, pp. 3–26.

Dougherty, E.C. (1955) Comparative evolution and the origin of sexuality, *Syst. Zool.* **4**, 145–190.

Driver, C., Tawadros, N. (2000) Cytoplasmic genomes that confer additional longevity in *Drosophila melanogaster*, *Biogerontology* **1**, 255–260.

Dudas, S.P., Arking, R. (1995) A coordinate upregulation of antioxidant gene activities is associated with the delayed onset of senescence in a long-lived strain of *Drosophila*, *J. Gerontol. Biol. Sci.* **50A**, B117–B127.

Fraga, C.G., Shigenaga, M.K., Park, J.-W., Degan, P., Ames, B.N. (1990) Oxidative damage to DNA during aging: 8-hydroxy-2′-deoxyguanosine in rat organ DNA and urine, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4533–4537.

Fryxell, P.A. (1957) Mode of reproduction in higher plants, *Botanical Rev.* **23**, 135–233.

Furuichi, Y. (2001) Premature aging and predisposition to cancers caused by mutations in RecQ family helicases, *Ann. N.Y. Acad. Sci.* **928**, 121–131.

Gaiddon, C., Lokshin, M., Ahn, J., Zhang, T., Prives, C. (2001) A subset of tumor-derived mutant forms of p53 down-regulate p63 and p73 through a direct interaction with the p53 core domain, *Mol. Cell. Biol.* **21**, 1874–1887.

Giese, H., Dolle, M.E., Hezel, A., van Steeg, H., Vijg, J. (1999) Accelerated accumulation of somatic mutations in mice deficient in the nucleotide excision repair gene XPA, *Oncogene* **18**, 1257–1260.

Gilley, D., Blackburn, E.H. (1994) Lack of telomere shortening during senescence in *Paramecium*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1955–1958.

Grammer, K. (1989) Human Courtship Behavior: Biological Basis and Cognitive Processing, in: Rasa, A.E., Vogel, C., Voland, E. (Eds.) *The Sociobiology of Sexual and Reproductive Strategies*, Chapman & Hall, New York, pp. 147–169.

Graymes, T.J., North, P.S., Brady, N., Hickson, I.D., Mufti, G.J., Rassool, F.V. (2002) Increased error-prone non homologous DNA end-joining – a proposed mechanism of chromosomal instability in Bloom's syndrome, *Oncogene* **21**, 2525–2533.

Grube, K., Burkle, A. (1992) Poly(ADP-ribose) polymerase activity in mononuclear leukocytes of 13 mammalian species correlates with species-specific life span, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2169–2173.

Gupta, R.C., Bazemore, L.R., Golub, E.I., Radding, C.M. (1997) Activities of human recombination protein Rad51, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 463–468.

Hamilton, M.L., Van Remmen, H., Drake, J.A., Yang, H., Guo, Z.M., Kewitt, K., Walter, C.A., Richardson, A. (2001) Does oxidative damage to DNA increase with age? *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10469–10474.

Hanawalt, C. (1994) Transcription-coupled repair and human disease, *Science* **266**, 1957, 1958.

Herrero, A., Barja, G. (1999) 8-oxo-deoxyguanosine levels in heart and brain mitochondrial and nuclear DNA of two mammals and three birds in relation to their different rates of aging, *Aging (Milano)* **11**, 294–300.

Higami, Y., Shimokawa, I., Okimoto, T., Ikeda, T. (1994) An age-related increase in the basal level of DNA damage and DNA vulnerability to oxygen radicals in the individual hepatocytes of male F344 rats, *Mutat. Res.* **316**, 59–67.

Holmes, G.E., Bernstein, C., Bernstein, H. (1992) Oxidative and other DNA damages as the basis of aging: a review, *Mutat. Res.* **275**, 305–315.

Holmes, G.E., Holmes, N.R. (1986) Accumulation of DNA damages in aging *Paramecium tetraurelia*, *Mol. Gen. Genet.* **204**, 108–114.

Houde, A.E., Torio, A.J. (1992) Effect of parasitic infection on male color pattern and female choice in guppies, *Behav. Ecol.* **3**, 346–351.

Huang, J., Dynan, W.S. (2002) Reconstitution of the mammalian DNA double-strand break end-joining reaction reveals a requirement for an Mre11/Rad50/NBS1-containing fraction, *Nucleic Acids Res.* **30**, 667–674.

Humphreys, D., Eggan, K., Akutsu, H., Friedman, A., Hochedlinger, K., Yanagimachi, R., Lander, E.S., Golub, T.R., Jaenisch, R. (2002) Abnormal gene expression in cloned mice derived from embryonic stem cell and cumulus cell nuclei, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12889–12894.

Imamura, O., Fujita, K., Shimamoto, A., Tanabe, H., Takeda, S., Furuichi, Y., Matsumoto, T. (2001) Bloom helicase is involved in DNA surveillance in early S phase in vertebrate cells, *Oncogene* **20**, 1143–1151.

Jimenez, J.A., Hughes, K.A., Alaks, G., Graham, L., Lacy, R.C. (1994) An experimental study of inbreeding depression in a natural habitat, *Science* **266**, 271–273.

Johnson, R.D., Jasin, M. (2000) Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells, *EMBO J.* **19**, 3398–3407.

Kaliraman, V., Mullen, J.R., Fricke, W.M., Bastin-Shanower, S.A., Brill, S.J. (2001) Functional overlap between Sgs1-Top3 and the Mms4-Mus81 endonuclease, *Genes Dev.* **15**, 2730–2740.

Kokoszka, J.E., Coskun, P., Esposito, L.A., Wallace, D.C. (2001) Increased mitochondrial oxidative stress in the Sod2 (+/−) mouse results

in the age-related decline of mitochondrial function culminating in increased apoptosis, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2278–2283.

Ku, H.H., Sohal, R.S. (1993) Comparison of mitochondrial pro-oxidant generation and anti-oxidant defenses between rat and pigeon: possible basis of variation in longevity and metabolic potential, *Mech. Ageing Dev.* **72**, 67–76.

Kuhne, M., Rothkamm, K., Lobrich, M. (2002) Physical and biological parameters affecting DNA double strand break misrejoining in mammalian cells, *Radiat. Prot. Dosim.* **99**, 129–132.

Kwan, K.Y., Wang, J.C. (2001) Mice lacking DNA topoisomerase IIIβ develop to maturity but show a reduced mean lifespan, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5717–5721.

Lehmann, A.R. (1998) Dual functions of DNA repair genes: molecular, cellular, and clinical implications, *BioEssays* **20**, 146–155.

Leslie, J.F., Raju, N.B. (1985) Recessive mutations from natural populations of *Neurospora crassa* that are expressed in the sexual diplophase, *Genetics* **111**, 759–777.

Li, B., Comai, L. (2002) Displacement of DNA-PKcs from DNA ends by the Werner syndrome protein, *Nucleic Acids Res.* **30**, 3653–3661.

Lindahl, T. (1977) DNA Repair Enzymes Acting on Spontaneous Lesions in DNA, in: Nichols, W.W., Murphy, D.G. (Eds.) *DNA Repair Processes*, Symposia Specialists Inc., Miami, FL, pp. 225–240.

Lindnor, N.M., Furuichi, Y., Kitao, S., Shimamoto, A., Arndt, C., Jalal, S. (2000) Rothmund-Thomson syndrome due to *RECQ4* helicase mutations: report and clinical and molecular comparisons with Bloom syndrome and Werner syndrome, *Am. J. Med. Genet.* **90**, 223–228.

Mahajan, K.N., McElhinny, S.A.N., Mitchell, B.S., Ramsden, D.A. (2002) Association of DNA polymerase μ (pol μ) with Ku and ligase IV: role for pol μ in end-joining double-strand break repair, *Mol. Cell Biol.* **22**, 5195–5202.

Mandavilli, B.S., Rao, K.S. (1996) Accumulation of DNA damage in aging neurons occurs through a mechanism other than apoptosis, *J. Neurochem.* **67**, 1559–1565.

Massie, H.R., Samis, H.V., Baird, M.B. (1972) The kinetics of degradation of DNA and RNA by H_2O_2, *Biochim. Biophys. Acta* **272**, 539–548.

Merry, B.J. (2002) Molecular mechanisms linking calorie restriction and longevity, *Int. J. Biochem. Cell Biol.* **34**, 1340–1354.

Michod, R.E. (1993) Genetic error, sex, and diploidy, *J. Heredity* **84**, 360–371.

Migliaccio, E., Giorgio, M., Mele, S., Pelicci, G., Reboldi, P., Pandolfi, P.P., Lanfrancone, L., Pelicci, P.G. (1999) The p66^{shc} adaptor protein controls oxidative stress response and life span in mammals, *Nature* **402**, 309–313.

Mohaghegh, P., Hickson, I.D. (2001) DNA helicase deficiencies associated with cancer predisposition and premature aging disorders, *Hum. Mol. Genet.* **10**, 741–746.

Morita, T., Yoshimura, Y., Yamamoto, A., Murata, K., Mori, M., Yamamoto, H., Matsushiro, A. (1993) A mouse homolog of the *Escherichia coli recA* and *Saccharomyces cerevisiae RAD51*genes, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6577–6580.

Moskovitz, J., Bar-Noy, S., Williams, W.S., Requena, J., Berlett, B.S., Stadtman, E.R. (2001) Methionine sulfoxide reductase (MsrA) is a regulator of antioxidant defense and lifespan in mammals, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12920–12925.

Muiras, M.L., Muller, M., Schachter, F., Burkle, A. (1998) Increased poly(ADP-ribose) polymerase activity in lymphoblastoid cell lines from centenarians, *J. Mol. Med.* **76**, 346–354.

Nakamura, J., Swenberg, J.A. (1999) Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues, *Cancer Res.* **59**, 2522–2526.

Nemoto, S., Finkel, T. (2002) Redox regulation of forkhead proteins through a p66shc-dependent signaling pathway, *Science* **295**, 2450–2452.

Ochem, A.E., Skooac, D., Costa, M., Rabilloud, T., Vuillard, L., Simoncsits, A., Giacca, M., Falaschi, A. (1997) Functional properties of the separate subunits of human DNA helicase II/Ku autoantigen, *J. Biol. Chem.* **272**, 29919–29926.

Orr, W.C., Sohal, R.S. (1994) Extension of lifespan by overexpression of superoxide dismutase and catalase in *Drosophila melanogaster*, *Science* **263**, 1128–1130.

Parkes, T.L., Elia, A.J., Dickinson, D., Hilliker, A.J., Phillips, J.P., Boulianne, G.L. (1998) Extension of *Drosophila* lifespan by overexpression of human *SOD1* in motor neurons, *Nat. Genet.* **19**, 171–174.

Pastink, A., Eeken, J.C., Lohman, P.H. (2001) Genomic integrity and the repair of double-strand DNA breaks, *Mutat. Res.* **480–481**, 37–50.

Pletcher, S.D., Macdonald, S.J., Marguerie, R., Certa, U., Stearns, S.C., Goldstein, D.B., Partridge, L. (2002) *Curr. Biol.* **12**, 712–723.

Qin, X., Zhang, S., Matsukuma, S., Zarkovic, M., Shimizu, S., Ishikawa, T., Nakatsuru, Y. (2000) Protection against malignant progression of spontaneously developing liver tumors in transgenic mice expressing O^6-methylguanine-DNA methyltransferase, *Jpn. J. Cancer Res.* **91**, 1085–1089.

Queille, S., Drougard, C., Sarasin, A., Daya-Grosjean, L. (2001) Effects of XPD mutations on ultraviolet-induced apoptosis in relation to skin cancer-proneness in repair-deficient syndromes, *J. Invest. Dermatol.* **117**, 1162–1170.

Richardson, C., Jasin, M. (2000) Coupled homologous and nonhomologous repair of a double-strand break preserves genomic integrity in mammalian cells, *Mol. Cell Biol.* **20**, 9068–9075.

Ruan, H., Tang, X.D., Chen, M.L., Joiner, M.L., Sun, G., Brot, N., Weissbach, H., Heinemann, S.H., Iverson, L., Wu, C.F., Hoshi, T. (2002) High-quality life extension by the enzyme peptide methionine sulfoxide reductase, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2748–2753.

Sancar, A. (1994) Mechanisms of DNA excision repair, *Science* **266**, 1954–1956.

Sands, A.T., Abuin, A., Sanchez, A., Conti, C.J., Bradley, A. (1995) High susceptibility to ultraviolet-induced carcinogenesis in mice lacking XPC, *Nature* **377**, 162–165.

Satoh, M.S., Jones, C.J., Wood, R.D., Lindahl, T. (1993) DNA excision-repair defect of xeroderma pigmentosum prevents removal of a class of oxygen free radical-induced base lesions, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6335–6339.

Schwartz, D., Goldfinger, N., Kam, Z., Rotter, V. (1999) p53 controls low DNA damage-dependent premeiotic checkpoint and facilitates DNA repair during spermatogenesis, *Cell Growth Differ.* **10**, 665–675.

Selzer, R.R., Nyaga, S., Tuo, J., May, A., Muftuoglu, M., Christiansen, M., Citterio, E., Brosh, R.M. Jr., Bohr, V.A. (2002) Differential requirement for the ATPase domain of the Cockayne syndrome group B gene in the processing of UV-induced DNA damage and 8-oxoguanine lesions in human cells, *Nucleic Acids Res.* **30**, 782–793.

Shen, J., Loeb, L.A. (2001) Unwinding the molecular basis of the Werner syndrome, *Mech. Ageing Dev.* **122**, 921–944.

Shapiro, R. (1981) Damage to DNA Caused by Hydrolysis, in: Seeberg, E., Kleppe, K. (Eds.) *Chromosome Damage and Repair*, Plenum Publishing, New York, pp. 3–18.

Shimamoto, A., Nishikawa, K., Kitao, S., Furuichi, Y. (2000) Human RecQ5β, a large isomer of RecQ5 DNA helicase, localizes in the nucleoplasm and interacts with topoisomerases 3α and 3β, *Nucleic Acids Res.* **28**, 1647–1655.

Shinohara, A., Ogawa, H., Matsuda, Y., Ushio, N., Ikeo, K., Ogawa, T. (1993) Cloning of human, mouse and fission yeast recombination genes homologous to RAD51 and recA, *Nat. Genet.* **4**, 239–243.

Sohal, R.S., Agarwal, A., Agarwal, S., Orr, W.C. (1995) Simultaneous overexpression of copper and zinc-containing superoxide dismutase and catalase retards age-related oxidative damage and increases metabolic potential in *Drosophila melanogaster*, *J. Biol. Chem.* **270**, 15671–15674.

Sreekumar, R., Unnikrishnan, J., Fu, A., Nygren, J., Short, K.R., Schimke, J., Barazzoni, R., Nair, K.S. (2002) Effects of caloric restriction on mitochondrial function and gene transcripts in rat muscle, *Am. J. Physiol. Endocrinol. Metab.* **283**, E38–E43.

Stahl, F.W. (1979) *Genetic Recombination: Thinking About it in Phage and Fungi*, Freeman & Co., San Francisco.

Sun, J., Folk, D., Bradley, T.J., Tower, J. (2002) Induced overexpression of mitochondrial Mn-superoxide dismutase extends the life span of adult *Drosophila melanogaster*, *Genetics* **161**, 661–672.

Sung, P. (1994) Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein, *Science* **265**, 1241–1243.

Terasawa, M., Shinohara, A., Hotta, Y., Ogawa, H., Ogawa, T. (1995) Localization of RecA-like recombination protein on chromosomes of the lily at various meiotic stages, *Genes Dev.* **9**, 925–934.

Trinei, M., Giorgio, M., Cicalese, A., Barozzi, S., Ventura, A., Migliaccio, E., Milia, E., Padura, I.M., Raker, V.A., Maccarana, M., Petronilli, V., Minucci, S., Bernardi, P., Lanfrancone, L., Pelicci, P.G. (2002)

A p53-p66Shc signalling pathway controls intracellular redox status, levels of oxidation-damaged DNA and oxidative stress-induced apoptosis, *Oncogene* **21**, 3872–3878.

Tuo, J., Muftuoglu, M., Chen, C., Jaruga, P., Selzer, R.R., Brosh, R.M. Jr., Rodriguez, H., Dizdaroglu, M., Bohr, V.A. (2001) The Cockayne syndrome group B gene product is involved in general genome base excision repair of 8-hydroxyguanine in DNA, *J. Biol. Chem.* **276**, 45772–45779.

Tyner, S.D., Venkatachalam, S., Choi, J., Jones, S., Ghebranious, N., Igelmann, H., Lu, X., Soron, G., Cooper, B., Brayton, C., Hee Park, S., Thompson, T., Karsenty, G., Bradley, A., Donehower, L.A. (2002) p53 mutant mice that display early ageing-associated phenotypes, *Nature* **415**, 45–53.

van Brabant, A.J., Ye, T., Sanz, M., German, J.L. Jr., Ellis, N.A., Holloman, W.K. (2000) Binding and melting of D-loops by the Bloom syndrome helicase, *Biochemistry* **39**, 14617–14625.

van Steeg, H., Mullenders, L.H.F., Vijg, J. (2000) Mutagenesis and carcinogenesis in nucleotide excision repair-deficient XPA knock out mice, *Mutat. Res.* **450**, 167–180.

Vikhanskaya, F., Marchini, S., Marabese, M., Galliera, E., Broggini, M. (2001) p73α overexpression is associated with resistance to treatment with DNA-damaging agents in a human ovarian cancer cell line, *Cancer Res.* **61**, 935–938.

Vogel, H., Lim, D.S., Karsenty, G., Finegold, M., Hasty, P. (1999) Deletion of Ku86 causes early onset of senescence in mice, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10770–10775.

Wakayama, T., Perry, A.C.F., Zucotti, M., Johnson, K.R., Yanagimachi, R. (1998) Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei, *Nature* **394**, 369–374.

Wang, X., Peterson, C.A., Zheng, H., Nairn, R.S., Legerski, R.J., Li, L. (2001) Involvement of nucleotide excision repair in a recombination-independent and error-prone pathway of DNA interstrand cross-link repair, *Mol. Cell. Biol.* **21**, 713–720.

Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S.J., Qin, J. (2002) BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures, *Genes Dev.* **14**, 927–939.

Wedekind, C., Furi, S. (1997) Body odor preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity? *Proc. R. Soc. London B Biol. Sci.* **264**, 1471–1479.

Weeda, G., Donker, I., de Wit, J., Morreau, H., Janssens, R., Vissers, C.J., Nigg, A., van Steeg, H., Bootsma, D., Hoeijmakers, J.H.J. (1997) Disruption of mouse *ERCC1* results in a novel repair syndrome with growth failure, nuclear abnormalities and senescence, *Curr. Biol.* **7**, 427–439.

Wijnhoven, S.W., Kool, H.J., van Oostrom, C.T., Beems, R.B., Mullenders, L.H., van Zeeland, A.A., van der Horst, G.T., Vrieling, H., van Steeg, H. (2000) The relationship between benzo[a]pyrene-induced mutagenesis and carcinogenesis in repair-deficient Cockayne syndrome group B mice, *Cancer Res.* **60**, 5681–5687.

Williams, M.D., Van Remmen, H., Conrad, C.C., Huang, T.T., Epstein, C.J., Richardson, A. (1998) Increased oxidative damage is correlated to altered mitochondrial function in heterozygous manganese superoxide dismutase knockout mice, *J. Biol. Chem.* **272**, 28510–28515.

Willis, J.H. (1999) Inbreeding load, average dominance and mutation rate for mildly deleterious alleles in *Nimulus guttatus*, *Genetics* **153**, 1885–1898.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., Campbell, K.H.S. (1997) Viable offspring derived from fetal and adult mammalian cells, *Nature* **385**, 810–813.

Wilson, E.O. (1978) *On Human Nature*, Harvard University Press: Cambridge, MA.

Wu, L., Davies, S.L., Levitt, N.C., Hickson, I.D. (2001) Potential role for the BLM helicase in recombinational repair via a conserved interaction with RAD51, *J. Biol. Chem.* **276**, 19375–19381.

Yang, Q., Zhang, R., Wang, X.W., Spillare, E.A., Linke, S.P., Subramanian, D., Griffith, J.D., Li, J.L., Hickson, I.D., Shen, J.C., Loeb, L.A., Mazur, S.J., Appella, E., Brosh R.M. Jr., Karmakar, P., Bohr, V., Harris, C.C. (2002) The processing of Holliday junctions by BLM and WRN helicases is regulated by p53, *J. Biol. Chem.* **277**, 31980–31987.

Yannone, S.M., Roy, S., Chan, D.W., Murphy, M.B., Huang, S., Campisi, J., Chen, D.J.

(2001) Werner syndrome protein is regulated and phosphorylated by DNA-dependent protein kinase, *J. Biol. Chem.* **276**, 38242–38248.

Zhong, Q., Chen, C.F., Chen, P.L., Lee, W.H. (2002) BRCA1 facilitates microhomology-mediated end joining of DNA double strand breaks, *J. Biol. Chem.* **277**, 28641–28647.

AIDS/HIV, Molecular and Cell Biology

Andrew M. L. Lever
University of Cambridge, Cambridge, UK

1 Origins of HIV 100

2 The Molecular Biology of HIV 102
2.1 Virus Structure 102
2.2 Life cycle 103
2.3 Additional Accessory Proteins 108
2.3.1 Virion Associated Proteins 108
2.3.2 Nonvirion Associated Proteins 108
2.4 Envelope Variants 109

3 Routes of Transmission 110

4 Natural History of Infection 110

5 Immune Response to HIV 112

6 Drug Therapy of HIV 114

7 Viral Escape 115

8 Vaccines 116

9 Summary 117

Acknowledgments 117

Bibliography 117
Books and Reviews 117
Primary Literature 118

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Keywords

HIV (Human Immunodeficiency Virus)
A retrovirus of the lentivirus family responsible for AIDS.

AIDS (Acquired Immune Deficiency Syndrome)
A clinical state of profound susceptibility to infection with low pathogenicity (opportunistic) infectious agent and unusual malignancies.

Retrovirus
A virus characterized by a diploid RNA genome converted to a DNA provirus by its reverse transcriptase enzyme and integrated into the target cell DNA.

Provirus
The integrated DNA form of the virus.

Antiretroviral Drugs
Pharmacological agents targeting virus-specific processes used in the treatment of AIDS.

This article describes the basic molecular biology of the viruses causing AIDS and the nature of the disease in causes, summarizing the immune response, antiviral therapy and varrine prospects.

1 Origins of HIV

HIV is a retrovirus of the lentivirus family. Closely related viruses have been discovered in many groups of African primates and sequence comparison would suggest that there has been more than one transspecies transmission of these viruses from monkeys into humans within the last 80 years. In the simian population, these viruses are transmitted by blood contact during biting and fighting. The butchering of monkeys for sale as bush meat is a highly plausible route of transmission of these viruses into humans through monkey blood contamination of cuts and scratches on human hands. Hypotheses surrounding contaminated polio vaccine have been comprehensively disproved.

There are two major divisions of HIV that have infected humans, HIV-1 and HIV-2. HIV-1 is phylogenetically closely related to SIV_{cpz}, a lentivirus of chimpanzees. HIV-2 is different in sequence from both of these and more closely related to the sooty mangabey lentivirus, SIV_{smm}. HIV-1 is divided into a number of different groups, the main (M) and the out (O) and new (N) groups within each of which are clades. The M group is estimated to have entered the human

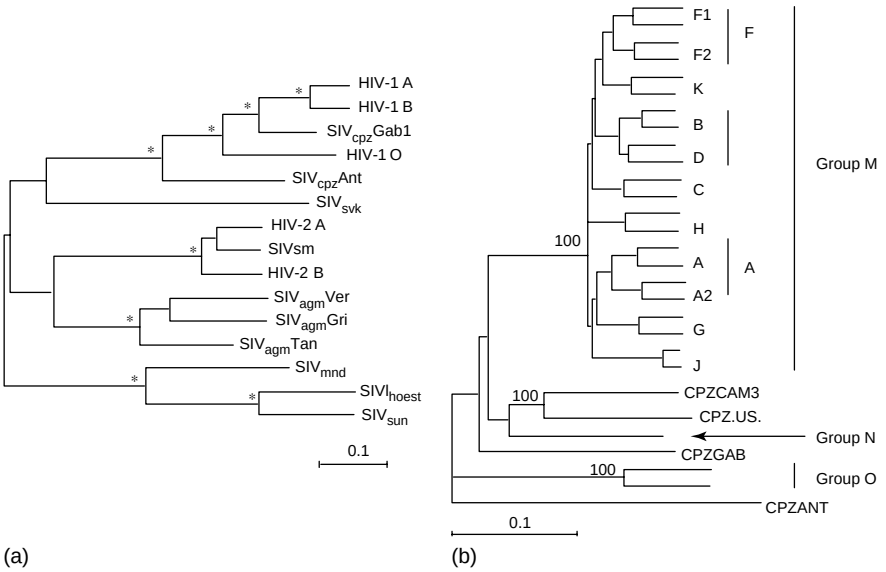

Fig. 1 (a) Relationships of HIV and SIV based on *pol* sequence comparison and (b) clades of HIV.

race in the 1930s. Within the M group, nine clades are recognized (Fig. 1a and b). In addition to clades, many recombinant or "mosaic" subtypes have also been identified. HIV-1 and HIV-2 have a very similar genetic structure except as detailed below. The sequence of the virus is, however, very variable indeed and two clades within the M group may differ in certain regions of the genome at the amino acid level by 20 to 40%. This is an astonishing level of variation and contributes significantly to the virus' success in establishing infection in the human population.

In their native hosts, simian lentiviruses appear to cause little disease. SIV from the African green monkey, SIV_{agm}, is asymptomatic in its natural host but produces profound immunodeficiency when introduced into rhesus macaques or cynomologus macaques. This may give us clues to the reason why HIV is so catastrophic in humans given the short time (<100 years) that it has been infecting humans and the consequent lack of host or pathogen adaptation.

The name lentivirus (lenti = slow) is derived from the archetypal member of this family, Maedi-Visna virus, which infects sheep and causes a very slow progressive disease involving pneumonic change and wasting, sometimes associated with central nervous system defects. In HIV, there may be symptoms associated with the initial viremia but thereafter, the speed of onset of clinical disease following infection is variable and infected individuals may remain asymptomatic for a number of years before the immunodeficiency becomes severe enough to cause illness.

Both HIV-1 and HIV-2 infection lead to a progressive decline in immune competence to the point where it becomes incapable of protecting the individual from otherwise very low pathogenicity infections, which are easily dealt with by a normal individual. This progressive loss of immune competence has given

the disease its name, *acquired immune deficiency syndrome* (*AIDS*). In general, the tempo of the disease is slower with HIV-2, possibly related to its replicating to a lower titer. This may also explain why it is less easily transmitted vertically from mother to child than HIV-1.

2 The Molecular Biology of HIV

Understanding the interactions of the virus with the immune system, its ability to cause disease, and potential therapies requires an understanding of the retroviral life cycle. The genetic structure of the virus is shown in Fig. 2. It has the conventional *gag, pol,* and *env* open reading frames of all retroviruses together with at least seven additional accessory and regulatory genes.

2.1 Virus Structure

HIV is a conventionally structured retrovirus (Fig. 3) with a C-type assembly mechanism. On the exterior, it has a lipid envelope derived from the surface plasma membrane of the cell from which it budded and in this are the envelope glycoproteins consisting of trimers of a transmembrane (TM) protein, each of which is noncovalently linked to an external surface (SU) glycoprotein. TM and SU are often referred to by their molecular weights as gp41 and gp120 respectively (in HIV-1).

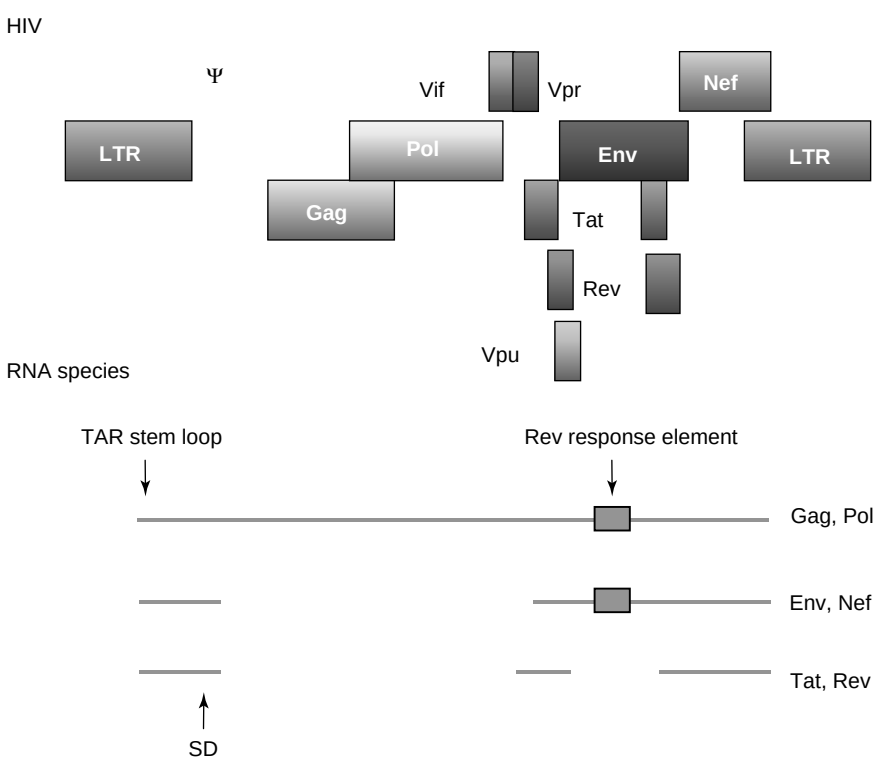

Fig. 2 Genetic structure of lentivirus DNA provirus and families of RNAs transcribed.

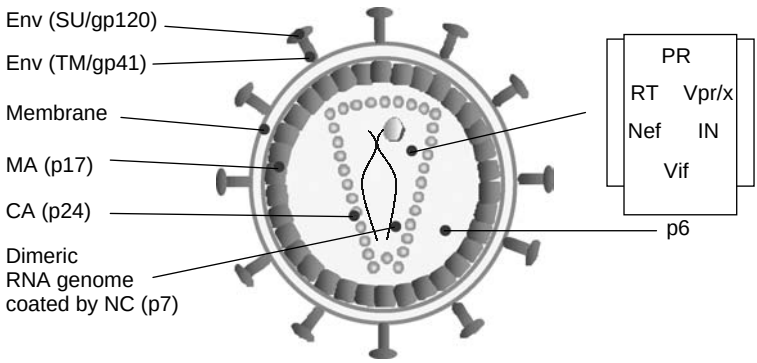

Fig. 3 Diagram of HIV 1 virus particle.

Within the mature virus particle, the inner surface of the envelope is lined by the matrix protein (MA), a cleavage derivative of the major core polyprotein, Gag. Within this is a pyramidal core structure made up of capsid (CA) protein subunits. This encloses the two copies of the RNA genome that are coated with the nucleocapsid (NC) protein. The two RNAs are tightly linked at a site known as the *dimer linkage site* (DLS). Each has a cell-derived $tRNA_{lys}$ annealed to a complementary sequence at the 5′ end of the genome, the primer binding site. Within the particle are a number of viral accessory proteins including Vif and Vpr. The virus also captures a number of cellular proteins including glycoproteins from the plasma membrane of the cell in its envelope and a number of cytoplasmic proteins including cyclophilin that binds specifically to the capsid protein in the core. The viral gene products of the *pol* gene are also incorporated into the virus. These include the protease (Pro) that cleaves the Gag and Pol precursor polyproteins to generate Gag products (p6, p7, p17, and p24) and Pol products including the reverse transcriptase enzyme (RT), which is responsible for generating a DNA copy from the RNA genome and the integrase enzyme (IN) that inserts the viral DNA into the chromosomes of the target cell.

2.2 Life cycle

Infection of a cell by HIV occurs following binding of the virus through its SU glycoproteins to the cell surface molecule CD4 and coreceptors (Fig. 4). The CD4 glycoprotein is found on a variety of cells, predominantly in those of the immune system. HIV tropism is broadly divided into viruses that can infect T lymphocytes (T-cell tropic) and those that infect cells of the macrophage/monocyte lineage (M tropic) (see Sect. 2.4).

Binding to the cell is followed by a conformational change in TM in which its hydrophobic terminus penetrates the target cell membrane and approximates it to the viral envelope lipids, fusing the two membranes together. Cellular lipid rafts appear to play a role in virus entry. The capsid and its contents enter the cell cytoplasm and the processes of disassembly and reverse transcription begin in which the viral RNA is copied into a DNA form. Reverse transcription is shown in Fig. 5. More detailed accounts can be found in the literature. Two important features of the

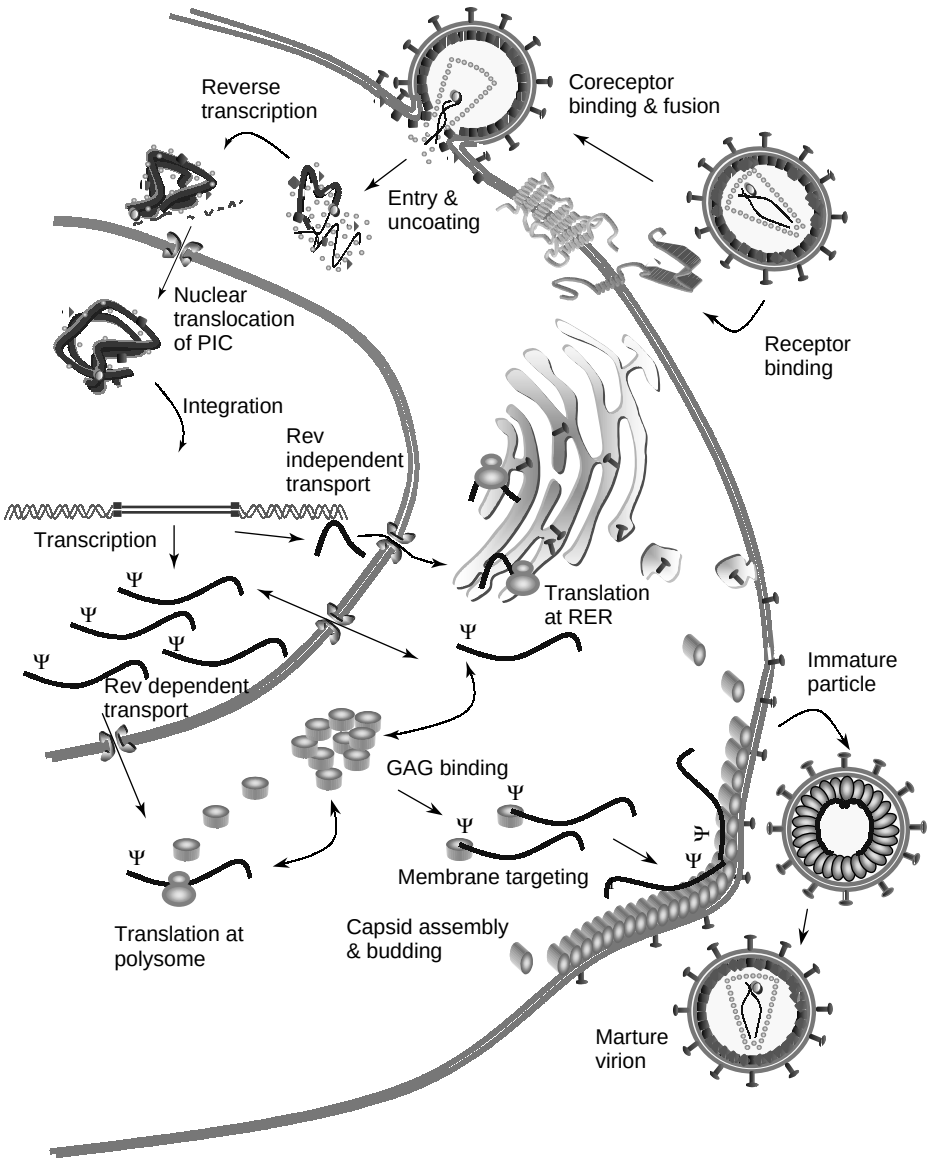

Fig. 4 Retroviral life cycle.

reverse transcription process are that both the RNA genomes are used as a template to produce a DNA copy in which template switching can occur such that the resultant DNA may be a combination of copies of segments from each of the RNA strands. Secondly, the resulting double-stranded DNA terminal sequences have both been templated by the 3′ long terminal repeat. It is during this reverse transcription step that the combination of template switching and nucleotide misincorporation, due

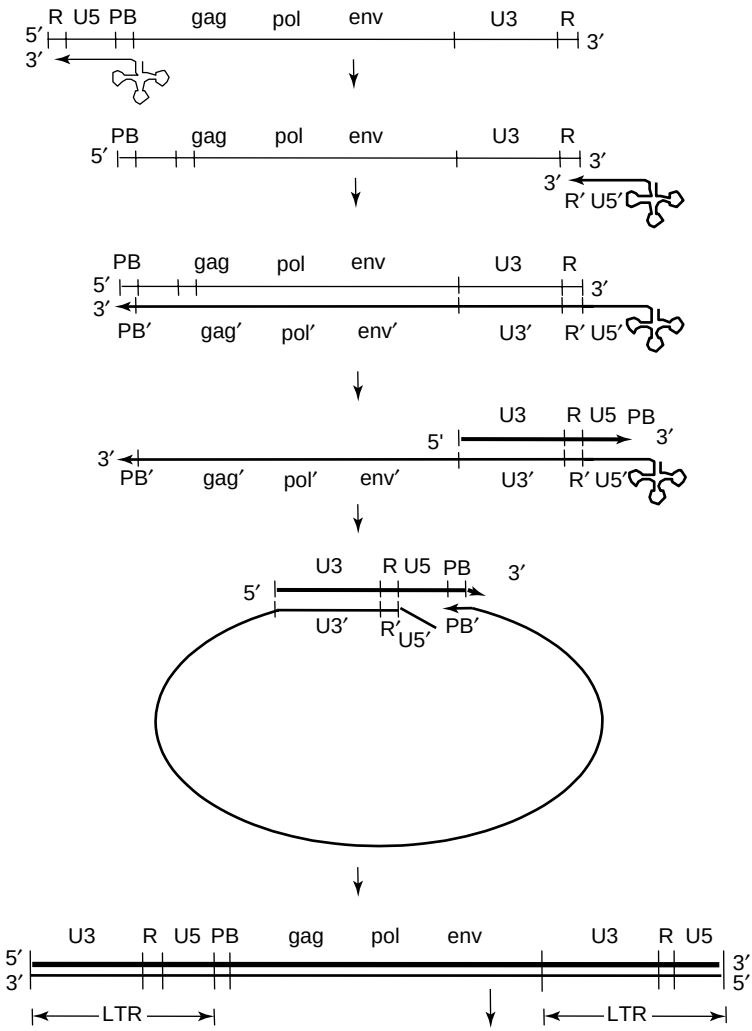

Fig. 5 Reverse transcription.

to the lack of RT proof-reading, produces the remarkable sequence diversity seen in HIV.

The DNA provirus is transported into the nucleus as a preintegration complex (PIC) using karyophilic signals found on proteins with the complex including the matrix protein, the integrase, and the Vpr accessory protein. There is some functional redundancy between these signals. In lentiviruses, the PIC can cross an intact nuclear membrane, meaning that these viruses have the unique capability among retroviruses of being able to integrate into cells that are not undergoing mitosis. The integration step involves the integrase enzyme cleaving the cellular DNA and making a 3′ to 5′ recess in the two ends of the proviral DNA. The chromosomal DNA undergoes a similar staggered

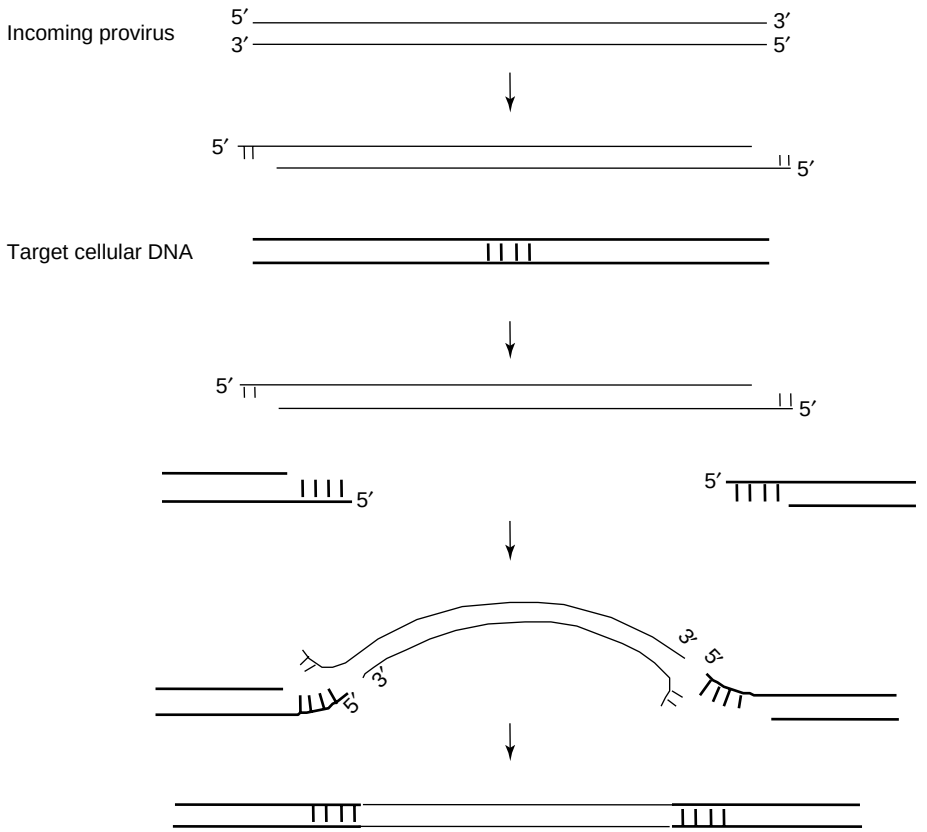

Fig. 6 Integration process.

cleavage and the viral and cellular DNA ends are linked. Cellular enzymes fill in the defect (Fig. 6). The resulting sequence shows a reduplication of the cellular nucleotides at either end of the provirus.

The integrated provirus may remain transcriptionally silent. Transcriptional activation depends in part on the degree of stimulation of the cell. In T cells, for example, mitotic stimulation is associated with activation of the retroviral promoter. Viral latency is a useful strategy for this pathogen to escape immune recognition and studies suggest that, at any one time, probably less than 10% of infected cells have a transcriptionally active virus.

The 5′ long terminal repeat of the virus contains a promoter element with a conventional TATA box. When activated, this is transcribed by RNA polymerase II to produce a genome length transcript. Multiple small abortive transcripts are seen early on but sufficient full-length RNAs are produced to initiate gene expression. Instability sequences are found in the full-length RNA, which default the RNA transcript to be spliced, and intronic sequences encompassing the Gag and Pol regions and the central part of envelope are spliced out to produce multiply spliced RNAs that are then exported into the cytoplasm and translated on free cytoplasmic

ribosomes. The earliest transcripts code for the regulatory proteins, Tat, Rev, and Nef (Fig. 2). Tat is a transcriptional activator that shuttles back into the nucleus and interacts with an RNA stem loop structure, the Tat Response Element (TAR) at the 5′ end of all viral transcripts (Fig. 2). Tat bound to TAR, recruits cyclin T and CDK9, which cause hyperphosphorylation of the C-terminal domain of the RNA polymerase II, massively enhancing its transcriptional efficiency. The sequence of transcription, export, and translation is replicated and amplified and more Tat protein is produced in what amounts to a positive feedback loop. The spliced RNA species encoding Rev is also translated and like the Tat protein, it shuttles back into the nucleus. It interacts with an RNA structure called the *Rev Responsive Element* (*RRE*) in the envelope intron (Fig. 2), enhancing export of both the full-length genomic RNA and the singly spliced RNA that encodes the envelope glycoprotein. Thus, the Tat protein causes an early amplification of transcription, and Rev subsequently leads to an early to late shift in the species of RNA, which are exported from the nucleus from those encoding regulatory factors to mRNAs for structural and enzymatic proteins. Unspliced full-length RNAs are translated on free cytoplasmic ribosomes to produce the Gag and Pol proteins. The singly spliced envelope encoding RNA binds to a ribosome and with the translation of its signal peptide, the complex is transported to the rough endoplasmic reticulum where envelope glycoproteins are translated and glycosylated using the conventional cellular glycoprotein secretory route.

The Gag and Pol proteins are encoded on the same transcript but in different reading frames. Translation of the polyprotein is interrupted by a frameshift sequence at the Gag/Pol overlap with an efficiency of about 5%. The ribosome shifts into the −1 frame, so that for approximately every 19 Gag proteins that are produced, 1 Gag/Pol fusion protein is synthesized. Gag and Gag/Pol proteins assemble at the plasma membrane in approximately the same stoichiometric ratio. At some stage, following transcription, the Gag protein has interacted with full-length genomic RNA transcripts and captured these through binding to a complex folded region of the RNA termed the *packaging signal* or Ψ region. This RNA is then transported into the assembling virion where it binds other Gag proteins through the zinc fingers of their nucleocapsid subfragments, the RNA acting to some extent as a scaffold for viral assembly. RNA–protein and protein–protein interactions contribute to the generation of a convex curved raft of protein underneath the cell membrane. Assembly continues with the viral bud bulging out of the cell and the Gag and Gag/Pol proteins accumulating to form a spherical viral core. This mode of assembly at the membrane defines lentiviruses as having a C-type morphology.

The viral core that forms is composed of Gag and Gag/Pol polyproteins, all aligned with the matrix region outermost and the NC region toward the center of the particle. At the extreme C-terminal of the Gag protein is a small region termed p6, which is intimately involved in the budding process. Mutations in p6 have been shown to lead to arrested budding in which stalked but otherwise intact particles remain attached to the plasma membrane of the cell. p6 has recently been shown to interact with a cellular protein, TSG101. In the normal cell, TSG101 is involved in the budding process of endosomes into multivesicular endosomes. It appears that the virus has

hijacked this budding mechanism for an external budding process and uses the same cellular factors to facilitate this. The complete array of proteins involved is not yet elucidated.

After or during the process of budding, the protease component of the Gag/Pol polyprotein becomes activated. This enzyme is inactive in the Gag/Pol monomer and functions as a homodimeric molecule. An as yet unknown process appears to approximate two protease monomers leading to autolytic cleavage of the first protease enzyme, following which cleavage of further Gag/Pol molecules will lead to release of further protease subunits and a cascade of proteolytic cleavage of Gag and Gag/Pol can occur. This cleavage is accompanied by a morphological change in which the particle changes from a spherical one containing a "doughnut" ring of structural proteins into a typical retrovirus particle with matrix protein underlying the envelope and an apparent space between this and the pyramidal core, which is composed of capsid proteins.

2.3 Additional Accessory Proteins

These can be divided into those that are or are not incorporated into the viral particle.

2.3.1 Virion Associated Proteins

As noted previously, the Vpr protein is involved in the preintegration complex targeting the nucleus of the newly infected cell. There is evidence that Vpr binds to the nucleoporin hCG1. Vpr also appears to have other functions. It may be a transcriptional transactivator in its own right through binding to p300/CBP coactivators. It has also been implicated in G2 phase cell cycle arrest. This is suggested to be the phase of the cell cycle at which Tat transactivation is most efficient. Vpr may also have a role in pathogenesis of AIDS since it has clearly been associated with apoptosis in a number of different cell types including neurons. It appears to gain entry to the virus by binding to the p6 region of Gag.

In HIV-2 and SIV, a gene duplication appears to have occurred to produce a second regulatory protein, Vpx, with homology to Vpr. There is evidence that in these viruses, Vpx may have subsumed the specific nuclear entry role of Vpr.

Vif is another small accessory protein that is incorporated into the virus particle. It has been clear for many years that Vif is an essential protein for viral replication in some cell lines but not others. The Vif phenotype depends on virion associated Vif since expression of Vif in a target cell does not appear to allow replication of an incoming Vif negative virus. Recent work suggests that Vif may be interacting with an inhibitory cellular protein that has sequence homology to proteins of the cytidine deaminase family and it may be that the actions of this nucleotide-modifying enzyme must be neutralized for successful reverse transcription and integration to occur.

2.3.2 Nonvirion Associated Proteins

Vpu is a small protein found only in HIV-1 and SIV_{cpz}. It has a number of roles that enhance the efficiency of budding of the virus. The envelope glycoprotein, gp120, which on the virion particle interacts with the CD4 protein, is synthesized and exported in the infected cell through the same pathway as the cell's own CD4. Complexing of these two molecules in the ER has been documented. Vpu appears to facilitate disaggregation of these two by binding to and degrading CD4, enhancing the export of the gp120 SU protein.

Vpu also appears to have additional envelope-independent effect, enhancing virus export.

The Nef protein is found throughout the primate lentiviruses. It is a multifunctional molecule. It has been documented as being capable of downmodulating cell surface expression of certain cellular proteins including CD4 and the MHC proteins. Recent evidence suggests that Nef may be involved in recruiting lymphocytes to HIV-infected macrophage/monocytes. Specific effects on migration inhibition of lymphocytes have been documented and these properties seem to be a powerful functional argument for Nef having important effects on virus spread and infectivity. This is further supported by the observation that Nef-deleted or Nef-mutated viruses have been shown to be associated with a severely attenuated pathogenicity. Recipients of HIV-infected blood from a single donor in Australia remained well for many years and the virus was identified as having a defective Nef protein. In SIV infection, Nef mutant viruses show an attenuated phenotype and reversion of the Nef mutation to wild type is associated with a regain of virulence.

2.4 Envelope Variants

The envelope gene of HIV-1 encodes a polyprotein comprising the two components of the external viral receptor ligand. At the N-terminus is the surface (SU) or gp120 protein. At the C-terminal, the TM is found. This latter sequence contains a hydrophobic stretch of amino acids, which comprise the membrane anchor and a second hydrophobic region involved in binding to the target cell membrane and fusing it with the virus envelope. The SU region is divided into five constant (C1-5) and five variable (V1-5) regions. The SU protein of HIV is responsible for binding to target cells. The V3 loop region of SU has attracted much attention because of its critical importance in receptor binding. The CD4 protein found on lymphocytes and cells of the monocyte/macrophage series was identified early in the AIDS epidemic as being a major receptor for the virus. At the same time, it was clear that binding to CD4 was necessary but not sufficient for infection. Observations that activated lymphocytes released substances that block infection by HIV and the identification of these as chemokines and their receptors as belonging to the family of 7-transmembrane spanning chemokine receptors revealed the existence of a family of coreceptor molecules, which are required for successful entry and infection. The two major families of these are the so-called *CC chemokine receptors* and the *CXC receptors*. Two of these are the principal HIV receptors. The CC chemokine receptor CCR5 responds to the cytokines MIP1α, MIP1β, and RANTES. The CXC chemokine receptor CXCR4 is bound by SDF1. There are a number of other receptors that have been identified as being involved in binding of HIV-1, HIV-2, and SIV. The identification of these receptors neatly mapped on the long-standing observation that there were at least two types of tropism identifiable in HIV. Some viruses appeared to be able to infect cells of the monocyte/macrophage series very efficiently but were very poor at infecting T lymphocytes. Others appear to be extremely efficient at infecting T lymphocyte cell lines *in vitro* but were very poor at infecting macrophages. These were broadly divided into "macrophage tropic" and "lymphocyte tropic" viruses (although this is a simplification). The chemokine receptor data link these tropisms with the

fact that the lymphocyte tropic viruses have a surface envelope protein that preferentially binds to the CXCR4 receptor and the macrophage tropic to the CCR5 receptor. These tropisms are commonly abbreviated by describing viruses as X4 or R5, respectively. More recently, a mutation has been identified, which affects about 1% of Caucasians who are homozygous for a stop codon causing a truncation of the CCR5 protein, and renders these individuals uninfectable by macrophage tropic virus. This explained another observation that there were some individuals who appeared to be repeatedly exposed to HIV through high-risk sexual practices and yet did not become infected.

3 Routes of Transmission

Infection can occur via cell-free virus or through contact between infected cells and those of the recipient. SIV in monkeys is probably primarily transferred from host to host by blood contact. HIV is also infectious by blood contact but the predominant mode of transmission is by the sexual route. HIV positive mothers can transmit the virus to their newborn child at birth or through breast feeding. The fact that cell-free virus is infectious means that HIV was transmitted very efficiently in clotting factor concentrates in the 1980s, resulting in the infection of large cohorts of hemophiliac patients. In the developed world, blood donations are screened for HIV as they are in many countries in the developing world. There is a very small but finite risk that a blood donor in the "window" period before seroconversion may transmit an infection through blood donation but this is in the order of less than one in a million. Blood transmission occurs quite efficiently in the case of intravenous drug users who share needles. Sexual transmission can occur from both male to female and female to male with an estimated risk of 1 : 200. Mucous membrane transmission probably begins with the infection of sessile cells of the macrophage/monocyte system in the mucous membrane tissues with subsequent replication and spread of virus to other cells of the immune system. Recent evidence implicates Langerhans cells as the first target. These may bind the virus using the DC-sign cell surface lectin, which has a high affinity for gp120. Whether replication occurs within these cells or whether surface-bound virus infects migrating T cells and macrophages is not established.

4 Natural History of Infection

The characteristic natural history of infection is illustrated in Fig. 7. Following early replication in lymph nodes and within a period of approximately four to eight weeks after exposure to virus, viremia becomes detectable in the form of ELISA positivity to the p24 capsid antigen and detection of viral RNA in the serum. This viremia rises rapidly to a peak that may be around five million copies of the viruses per milliliter of blood. At this time, the beginnings of an immune response against the virus become detectable, including special antibodies directed against the viral proteins. Subsequently, the virus level in the blood falls to a "set point" and this correlates most closely with detection of a rise in the cytotoxic T lymphocyte cell mediated immunity. At the time of this seroconversion, between a third and a

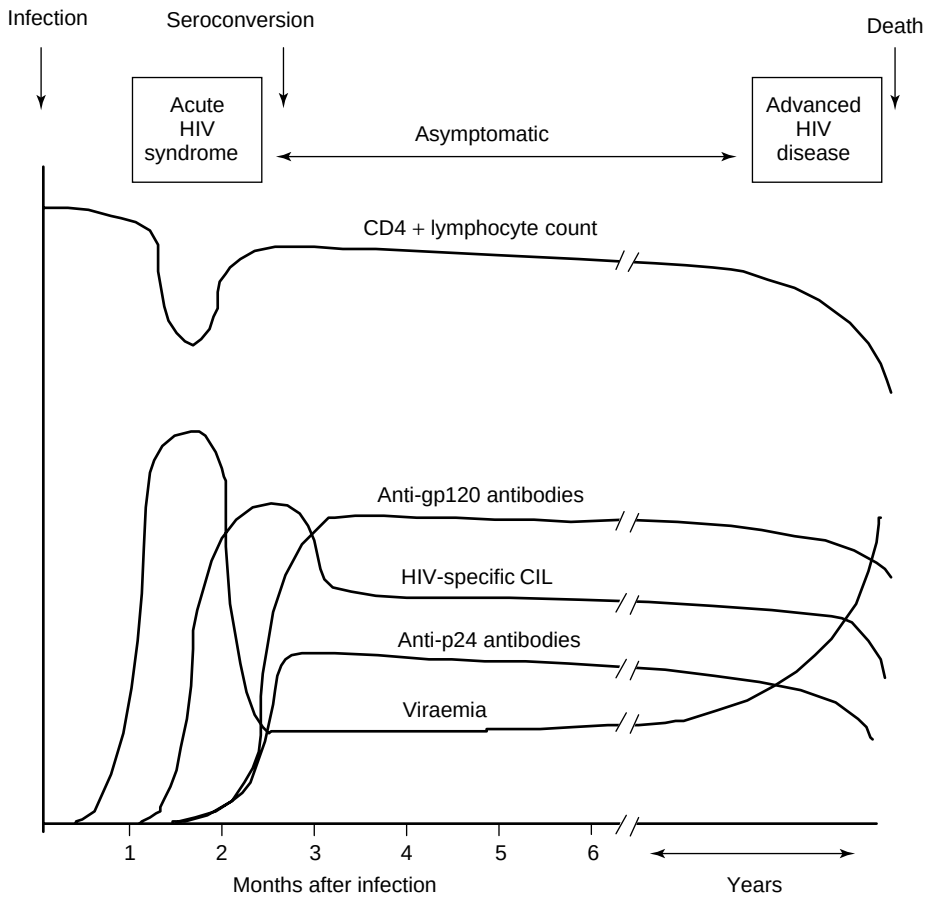

Fig. 7 Natural history of HIV infection.

half of infected patients will have symptoms resembling a glandular feverlike illness with lymphadenopathy, transient petechial rashes, arthropathy, evidence of peripheral neuropathy, and occasionally a brief meningoencephalitic illness. However, virtually without exception, these symptoms resolve. There then follows a period of apparent stability in which the viral level or "viral load" remains stable and there is ongoing evidence of an immune response. In some individuals, persistent enlargement of lymph nodes occurs associated with weight loss and fevers. This so-called *persistent generalized lymphadenopathy* appears to be associated with a worse prognosis in terms of the speed of onset of immunodeficiency. Despite the asymptomatic nature of infection in most individuals at this stage, measurements of the CD4 lymphocyte count show that during the initial viral peak, there is a corresponding trough in lymphocyte numbers, implying a significant amount of direct virus mediated lymphocyte destruction. This rises again as the virus load is brought under control but does not achieve the preinfection level and

from then on there is a very slow but progressive decline.

Calculations of the quantity of virus being produced during this time and the rate of lymphocyte destruction/regeneration have generated the remarkable figures that around 10^9–10^{10} viruses are produced and destroyed each day. This is probably an underestimate. The lymphocyte turnover is also around 10^9 per day, but while the viral load may be constant or may rise, the lymphocyte count slowly falls such that there is a net loss of around 10^6 lymphocytes per day. The fall in lymphocyte count correlates with a progressive decline in immunological competence and is a marker, if not the cause, of the eventual breakdown in immune competence. At this stage, the number of infected cells in the blood is very low, at around 1 in 10^5. This might suggest that virus mediated lymphocyte destruction is less important at this stage, although localized destruction in lymph nodes would be difficult to detect. Immune clearance of infected cells, however, may be the dominant force affecting lymphocyte numbers. Depending on the success of the immune response in controlling the virus, the length of time between seroconversion and immunodeficiency may be anything from 2 up to 12+ years. However, as yet, although there are examples of very slow progressors, there is no evidence of complete control of viral infection by the immune system once the virus has established itself. At the late stages of the disease, the predominant type of virus appears to change with the emergence of faster replicating T-cell tropic (X4) variants.

Symptomatic immunodeficiency becomes apparent when the lymphocyte count falls below 200×10^9 L^{-1}, at which point the individual is at very high risk of opportunistic infection and the development of unusual malignancies. The fall in lymphocyte count correlates with a loss of control over viral replication and a rise in detectable virus in the circulation in the absence of antiretroviral therapy. Death from an overwhelming infection or an uncontrolled malignancy usually occurs within two to three years of the onset of the severe immunodeficiency even with treatment of the individual infections and cancers.

As will be described later, the advent of antiretroviral therapy has transformed the natural history of disease in the western world. Pharmacological control of virus replication leads to a fall in viral load and a corresponding rise in the CD4 lymphocyte count. This leads to a reversal of the loss of immunocompetence, which, although it may not be complete, is enough to significantly reduce the risk of opportunistic infection.

5
Immune Response to HIV

Following the initial peak of virus in the circulation after infection, the viral load declines to a set point, which, on average, is around 30 000 copies per milliliter. If it is higher than this, the onset of AIDS is quicker and vice versa. It appears that qualitative and quantative parameters of the immune response determine this set point. A broad antibody response develops within the first three months of infection with specificity for some of the structural proteins such as Gag. Antibodies to Envelope and Polymerase proteins appear swiftly thereafter, and the regulatory and accessory proteins also trigger an antibody response.

Antibodies interfering with interaction between the envelope protein and the

cellular receptor may be neutralizing and the site at which they bind on the viral protein provides strong clues as to the important regions of SU involved in viral entry. In particular, the V3 loop is a major target for neutralizing antibody. A second region on gp120, which is involved in both CD4 and chemokine receptor binding, is another neutralization target. However, V3 loop neutralizing antibodies often appear to be more effective against laboratory strains of HIV than primary clinical isolates. It is possible that this relates to the different conformation of the envelope in the *in vitro* adapted strains in which the lack of immune selection pressure has allowed a more open structure that favors rapid cellular entry. *In vivo*, such a structure would predispose to the blockade by neutralizing antibodies and may be disadvantageous.

Other evidence suggesting that antibodies are of relatively low importance is the fact that some individuals in whom the illness appears to be held in check by the immune response do not have high levels of neutralizing antibody. In addition, passive protection using antibodies in animal models is only effective when they are present in very high titer. There is some evidence that high titers of neutralizing antibody may be associated with a lower "set point." The immense variability of the viral envelope facilitates immunological escape of the virus. This may occur either by amino acid variability or by modification of envelope glycosylation.

Cytotoxic T-cell responses can be detected before the antibody response and have long been considered one of the major immunological defenses against HIV. CD8+ T cells can kill virus infected cells by specific recognition of the HIV peptides presented on class I MHC proteins on the surface of the infected cell. *In vitro*, CTLs can eliminate infected cells even before new viruses have been produced. A second effect involves the release of β-chemokine mediators, which, as previously mentioned, are themselves able to bind to their natural ligand CCR5 and block the use of this protein by HIV for infection of the cell. The strong correlation between appearance of CD8+ T cells and the decline in viral replication in the early stages together with data on adoptive transfer in humans and CD8 depletion experiments in monkeys, which leads to a sustained increase in viral load, confirms the critical role played by this arm of the immune response in combating infection.

The signature of HIV infection is the decline in CD4 lymphocytes. The mechanisms of destruction of CD4 lymphocytes are probably many and varied as described earlier. Production of HIV with its gp120 within a cell that is already making CD4 may lead to protein aggregation within the cell and dysfunction leading to apoptosis. This can be observed *in vitro* as single cell killing. Expression of gp120 on the surface of the cell may induce fusion of that cell with a CD4+ cell. This latter phenomenon is also seen *in vitro* and leads to killing by syncitia formation. Passive binding of shed gp120 on to the CD4 proteins of immune competent cells may lead to them being a target for immunological clearance and the gp120 itself may have adverse effects. The cytotoxic T-cell response, clearing infected cells, is probably responsible for a very large proportion of the destruction of CD4+ infected cells. Thus, a number of direct and indirect mechanisms contribute to the loss of the CD4 cell population.

Individuals whose disease progresses very slowly (long-term nonprogressors – LTNPs) have clearly been shown to have powerful T-cell responses against HIV with the CD4 lymphoproliferative

response correlating well with successful suppression of the viral load. Whether these cells act independently or through enhancement of CD8+ CTL killing is not clear and it may be a combination of both. Apart from specific genetic variants such as the CCR5 truncation mutation described earlier, there does appear to be some genetic influences in that certain HLA alleles HLAB35 and CW4 are associated with more rapid disease progression, and certain other polymorphisms in chemokines and chemokine receptor genes may influence responses to therapeutic strategies and vaccination.

The immune response is capable of a measure of control of HIV replication in all patients, and in some, there is very significant delay in the onset of AIDS. Recent evidence suggests that not only does the ability to mount an immune response decline as disease progresses but the ability to regenerate this when antiretroviral treatment is started also declines the longer the initiation of treatment is delayed. Indeed, treatment during the early stages of infection at seroconversion has been correlated with a much better long-term CD4 lymphoproliferative response. Thus, early treatment might lead to the ability to maintain a strong CD4 cell response and this might provide the possibility of powerful long-term control of viral replication. Against this is the risk of generating drug resistance early in infection and the loss of utility of the limited range of chemotherapeutic agents when they may be more useful at a later stage.

6
Drug Therapy of HIV

Antiviral drugs are commonly targeted at viral enzymes. This is the case with HIV in which the first two types of highly successful pharmacotherapies have been directed to inhibit the actions of the viral reverse transcriptase gene and the protease gene. Reverse transcriptase is clearly a virus-specific process since the cell does not have any functional RNA-dependent DNA polymerase activity. The reverse transcriptase enzyme does not have proof-reading capability and while, as mentioned later, this may be an advantage to the virus in generating variants, it also is an Achilles' heel for the virus in that nucleoside analogs that mimic the bases that the enzyme is attempting to incorporate into the DNA chain can substitute for these and once inserted cannot be excised. Pyrimidine analogs have been particularly effective and they appear to act both by inhibiting the enzyme and by chain termination. They require phosphorylation within the cell to form triphosphates. Another class of drug that inhibits RT is the nonnucleoside reverse transcriptase inhibitors (NNRTIs), which bind to reverse transcriptase away from the active site but are not incorporated into the transcript. A third major class of antiretroviral drugs are those inhibiting the viral protease. These take advantage of the target peptide sequence of the aspartyl protease of HIV, being slightly different from cellular homologs of this enzyme. Predictably, however, the specificity of these drugs is somewhat lower than RT inhibitors and they are associated with a greater number of side effects, probably relating to effects on cellular proteases. Newer classes of drugs are being developed including those that may inhibit viral entry, viral integration, and viral export. Injection of a peptide, which interferes with the formation of the fusogenic envelope of the virus and blocks cell entry, has recently been introduced into clinical practice.

Structured treatment interruptions (STI) is being tested as a strategy to reduce drug toxicity. As yet, the effect of this on overall prognosis and immune regeneration is not fully established.

7
Viral Escape

The previously described infidelity of the reverse transcriptase enzyme is responsible for considerable sequence change in the virus. Undoubtedly, many of the incorporated nucleotides will cause lethal mutations and a nonviable integrated provirus will result. The huge numbers of viruses being produced everyday, however, will ensure that a large number of viable variants will be produced, which may have minor changes in their amino acid sequence. Thus, it is easy to see that a virus entering a macrophage may mutate slightly during its reverse transcription and integration process such that the progeny virions from that cell may have a slightly greater predisposition to infect lymphocytes than the parental version. The other major mechanism of sequence variation comes from recombination. When a cell is infected with two different viruses, the diploid nature of the genome means that it is possible for an RNA strand from two different integrated proviruses to be copackaged. During reverse transcription in the target cell, the reverse transcriptase enzyme is known to dissociate from its template and may reassociate with the partner template. Thus, exchange of large blocks of genetic information between the two strands is quite common. Indeed, it is estimated that the enzyme skips between strands at least four times during every replication cycle. In this way, mixing and matching of genomes can occur with ease and, in this case, large blocks of functionally intact genome will be changed. Recombination is, therefore, believed to be as significant if not more significant than RT errors in generating diversity. Some of the major worldwide clades have clearly been formed from recombination. Genotype E in Thailand, for example, is a recombinant between a clade A Gag and a clade E envelope sequence.

Variation allows the virus to introduce point mutations, which will render the binding of antiviral drugs much less efficient, and strategies to combat viral resistance have become an integral part of clinical management. This is particularly the case since antiviral therapy in the western world became more widespread. The chances of acquiring a virus, which has already generated some antiviral resistance, has increased greatly. For this reason, where possible, it is recommended that antiviral resistance testing is undertaken before initiating antiretroviral therapy so that the combination of drugs chosen is one in which the virus is sensitive to each component. As well as this, it is vital that combinations of three or more drugs are used to prevent emergence of resistance. Both of these strategies are designed to avoid presenting the virus with a single drug against which it can develop resistance.

The genetic variation also contributes very largely to the difficulty the immune system has in neutralizing this virus. There are clearly documented examples of immunological escape where viral proteins, which have cytotoxic T-cell epitopes, mutate and thus escape cytotoxic T-cell recognition. In terms of envelope variation, the virus has the additional weapon of variable glycosylation. HIV envelope is extremely heavily glycosylated compared to a number of its close viral relatives

and it is clear that glycosylation may significantly inhibit the function of neutralizing antibodies.

8
Vaccines

From the above, it is clear that the development of a vaccine against HIV is a monumental task compared to many other pathogens. The features that make it particularly challenging are as follows. Firstly, there is no evidence that complete sterilizing immunity occurs in the human race after infection. Thus, even the best immune response in the world does not appear to be capable of indefinitely checking virus replication. Secondly, the huge variability of the virus means that unlike, for example, polio or smallpox, vaccines containing a restricted number of variants will not successfully prevent infection and disease caused by HIV since the virus has such a vast repertoire of viable variants. Thirdly, the virus integrates into the genome of the cell. A number of the features of the life cycle of this and other retroviruses have been designed to specifically maintain cellular viability since death of the cell equates with death of the provirus. The action of many of the accessory genes of HIV in downmodulating cell surface proteins makes even a good immune response incapable of easily clearing infection. Over and above that, however, the virus has the capacity to become latent. Complete transcriptional arrest may occur (or a level of basal transcription insufficient to generate Tat), with transcriptional reactivation occurring under appropriate conditions such as triggering of the cell into mitosis. A transcriptionally inactive virus is completely invisible to the immune system and persistence of a small number of latently infected cells leaves the possibility of reactivation when the immune response wanes. Reactivation of other viruses such as *Varicella zoster* provides a comparable parallel. It is, thus, not surprising that progress toward a vaccine has been slow. To date, vaccines have been generated that are capable of preventing infection in SIV models. Most of these appear to be effective against homologous strains but with, as yet, relatively little evidence that effective cross-neutralization of heterologous strains occurs. Use of a live attenuated vaccine such as one deleted in *nef* and some of the other regulatory genes has been associated with protection against subsequent challenge. Attenuation, however, appears to be a finite phenomenon and mutation back to the wild-type sequence and full pathogenicity is a feature. Recent data using prime boost techniques in which carrier vaccine strains are used such as modified vaccinia Ankara have shown some promising evidence of good immune responses against HIV proteins. As yet, the duration of immune response using these techniques is not clear. A cohort of prostitutes in western Africa who were repeatedly exposed to HIV but remained uninfected originally, generated hope that a fully neutralizing immune response was possible with frequent low level viral challenge. The follow-up of these individuals showed that when sexual exposure to HIV ceases, the immunity wanes and the individuals become susceptible to infection again. This, again, is rather a concerning issue in terms of vaccine development since it would suggest that repeated boosts of vaccination will be required to maintain a level of immunity, which would otherwise be lost. There are strong parallels here with immunity to malaria.

9 Summary

The HIV epidemic is a global problem that has been controlled but not eliminated in the western world. The vast majority of infected individuals live in sub-Saharan Africa where access to treatment for the virus or its complications is extremely limited. Significant loss of life and disruption of family and society has occurred in this part of the world due to HIV. Recurrent suggestions that HIV may not be the causative agent of AIDS have seriously set back educational programs in a number of countries and are unfounded and irresponsible. Countries that have taken advice and help and approached the problem with a vigorous education campaign such as Uganda have reaped the rewards with the HIV positivity rate in the antenatal population halving over a period of less than 10 years. Until vaccines are developed and antiretroviral therapy becomes more widely available, these sort of approaches should be powerfully supported. In the western world, antiretroviral therapy has transformed the prognosis, and conceivably, a middle aged person acquiring HIV might now have a normal lifespan given the additive effects of the slow onset of clinical disease, the availability of antiretroviral drugs, and the combined time window these provide for the appearance of new therapies. While the death rate has fallen in the west, the incidence of HIV infection has continued to rise. This is partly because of increased longevity of infected individuals and also from increasing numbers of migrants from highly endemic areas who may seek access to western health resources. This trend is increasing and will provide an increasing burden on western healthcare services. There is still no reason why an uninfected person cannot be protected against HIV since transmission is preventable by barrier methods of contraception, vaginal virucidal agents, and so on. However, the key to the control of HIV in the near future in the developing and developed world is education.

Acknowledgments

I would like to acknowledge S. Griffin for contributions to the figures.

See also Genetic Engineering of Vaccines; Genetics, Molecular Basis of; RNA Virus Genome Packaging.

Bibliography

Books and Reviews

British HIV Association Guidelines for Antiretroviral Treatment of HIV Seropositive Individuals. BHIVA Guidelines Co-ordinating Committee (1997), *Lancet* **349**, 1086–1092.

Carpenter, C.C., Cooper, D.A., Fischl, M.A., Gatell, J.M., Gazzard, B.G., Hammer, S.M., Hirsch, M.S., Jacobsen, D.M., Katzenstein, D.A., Montaner, J.S., Richman, D.D., Saag, M.S., Schechter, M., Schooley, R.T., Thompson, M.A., Vella, S., Yeni, P.G., Volberding, P.A. (2000) Antiretroviral therapy in adults: updated recommendations of the international AIDS society-USA panel, *JAMA* **283**, 381–390.

Coffin, J.M., Hughes, S.H., Varmus, H.E. (Eds.) (1997) *Retroviruses*, Cold Spring Harbor Laboratory Press, New York.

Cohen, O.J., Fauci, A.S. (2001) Pathogenesis and Medical Aspects of HIV-1 Infection, in: Knipe, D.M., Howley, P.M. (Eds.) *Fields Virology*, 4th edition, Lippincott Williams & Wilkins, Philadelphia, pp. 2043–2094.

Fackler, O.T., Peterlin, B.M., Weis, K. (2001) Lessons from HIV: movement of macromolecules inside the cell, *Curr. Mol. Med.* **1**, 1–7.

Freed, E.O., Martin, M.A. (2001) HIVs and their Replication, in: Knipe, D.M., Howley, P.M. (Eds.) *Fields Virology*, 4th edition, Lippincott Williams & Wilkins, Philadelphia, pp. 1971–2042.

Jeang, K.-T. (Ed.) (2000) *Advances in Pharmacology: HIV-1: Molecular Biology and Pathogenesis – Viral Mechanisms (Vol 48)* & *Clinical Applications (Vol. 49)*, Academic Press, San Diego.

Persaud, D., Zhou, Y., Siliciano, J.M., Siliciano, R.F., (2003) Latency in human immunodeficiency virus type 1 infection: no easy answers, *J. Virol.* **77**, 1659–1665.

Weiss, R.A., Adler, M.W., Rowland-Jones, S.L. (Eds.) (2001) The changing face of HIV and AIDS, *Br. Med. Bull.* **58**(1).

Wyatt, R., Sodroski, J. (1998) The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens, *Science* **280**, 1884–1888.

Primary Literature

Addo, M.M., Altfeld, M., Rosenberg, E.S., Eldridge, R.L., Philips, M.N., Habeeb, K., Khatri, A., Brander, C., Robbins, G.K., Mazzara, G.P., Goulder, P.J., Walker, B.D. (2001) The HIV-1 regulatory proteins Tat and Rev are frequently targeted by cytotoxic T lymphocytes derived from HIV-1-infected individuals, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1781–1786.

Adjorlolo-Johnson, G., De Cock, K.M., Ekpini, E., Vetter, K.M., Sibailly, T., Brattegaard, K., Yavo, D., Doorly, R., Whitaker, J.P., Kestens, L., Ou, C.Y., George, J.R., Gayle, H.D. (1994) Prospective comparison of mother-to-child transmission of HIV-1 and HIV-2 in Abidjan, Ivory Coast, *JAMA* **272**, 462–466.

Amara, R.R., Villinger, F., Altman, J.D., Lydy, S.L., O'Neil, S.P., Staprans, S.I., Montefiori, D.C., Xu, Y., Herndon, J.G., Wyatt, L.S., Candido, M.A., Kozyr, N.L., Earl, P.L., Smith, J.M., Ma, H.L., Grimm, B.D., Hulsey, M.L., Miller, J., McClure, H.M., McNicholl, J.M., Moss, B., Robinson, H.L. (2001) Control of a mucosal challenge and prevention of AIDS by a multiprotein DNA/MVA vaccine, *Science* **292**, 69–74.

Andreasson, P.A., Dias, F., Naucler, A., Andersson, S., Biberfeld, G. (1993) A prospective study of vertical transmission of HIV-2 in Bissau, Guinea-Bissau, *AIDS* **7**, 989–993.

Asjo, B., Morfeldt-Manson, L., Albert, J., Biberfeld, G., Karlsson, A., Lidman, K., Fenyo, E.M. (1986) Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection, *Lancet* **2**, 660–662.

Baltimore, D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses, *Nature* **226**, 1209–1211.

Barouch, D.H., Kunstman, J., Kuroda, M.J., Schmitz, J.E., Santra, S., Peyerl, F.W., Krivulka, G.R., Beaudry, K., Lifton, M.A., Gorgone, D.A., Montefiori, D.C., Lewis, M.G., Wolinsky, S.M., Letvin, N.L. (2002) Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes, *Nature* **415**, 335–339.

Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., Montagnier, L. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS), *Science* **220**, 868–871.

Beaumont, T., van Nuenen, A., Broersen, S., Blattner, W.A., Lukashov, V.V., Schuitemaker, H. (2001) Reversal of human immunodeficiency virus type 1 IIIB to a neutralization-resistant phenotype in an accidentally infected laboratory worker with a progressive clinical course, *J. Virol.* **75**, 2246–2252.

Berry, N., Davis, C., Jenkins, A., Wood, D., Minor, P., Schild, G., Bottiger, M., Holmes, H., Almond, N. (2001) Vaccine safety. Analysis of oral polio vaccine CHAT stocks, *Nature* **410**, 1046–1047.

Borrow, P., Lewicki, H., Hahn, B.H., Shaw, G.M., Oldstone, M.B. (1994) Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection, *J. Virol.* **68**, 6103–6110.

Brodie, S.J., Patterson, B.K., Lewinsohn, D.A., Diem, K., Spach, D., Greenberg, P.D., Riddell, S.R., Corey, L. (2000) HIV-specific cytotoxic T lymphocytes traffic to lymph nodes and localize at sites of HIV replication and cell death, *J. Clin. Invest.* **105**, 1407–1417.

Bukrinsky, M.I., Haggerty, S., Dempsey, M.P., Sharova, N., Adzhubel, A., Spitz, L., Lewis, P., Goldfarb, D., Emerman, M., Stevenson, M. (1993) A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells, *Nature* **365**, 666–669.

Carrington, M., Nelson, G.W., Martin, M.P., Kissner, T., Vlahov, D., Goedert, J.J., Kaslow, R., Buchbinder, S., Hoots, K., O'Brien, S.J. (1999) HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage, *Science* **283**, 1748–1752.

Chackerian, B., Rudensey, L.M., Overbaugh, J. (1997) Specific N-linked and O-linked glycosylation modifications in the envelope V1 domain of simian immunodeficiency virus variants that evolve in the host alter recognition by neutralizing antibodies, *J. Virol.* **71**, 7719–7727.

Champagne, P., Ogg, G.S., King, A.S., Knabenhans, C., Ellefsen, K., Nobile, M., Appay, V., Rizzardi, G.P., Fleury, S., Lipp, M., Forster, R., Rowland-Jones, S., Sekaly, R.P., McMichael, A.J., Pantaleo, G. (2001) Skewed maturation of memory HIV-specific CD8 T lymphocytes, *Nature* **410**, 106–111.

Chen, M.Y., Maldarelli, F., Karczewski, M.K., Willey, R.L., Strebel, K. (1993) Human immunodeficiency virus type 1 Vpu protein induces degradation of CD4 in vitro: the cytoplasmic domain of CD4 contributes to Vpu sensitivity, *J. Virol.* **67**, 3877–3884.

Cheng-Mayer, C., Seto, D., Levy, J. (1988) Biologic features of HIV-1 that correlate with virulence in the host, *Science* **240**, 80–82.

Cho, M.W., Kim, Y.B., Lee, M.K., Gupta, K.C., Ross, W., Plishka, R., Buckler-White, A., Igarashi, T., Theodore, T., Byrum, R., Kemp, C., Montefiori, D.C., Martin, M.A. (2001) Polyvalent envelope glycoprotein vaccine elicits a broader neutralizing antibody response but is unable to provide sterilizing protection against heterologous Simian/human immunodeficiency virus infection in pigtailed macaques, *J. Virol.* **75**, 2224–2234.

Chun, R.F., Jeang, K.T. (1996) Requirements for RNA polymerase II carboxyl-terminal domain for activated transcription of human retroviruses human T-cell lymphotropic virus I and HIV-1, *J. Biol. Chem.* **271**, 27888–27894.

Cochrane, A.W., Perkins, A., Rosen, C.A. (1990) Identification of sequences important in the nucleolar localization of human immunodeficiency virus Rev: relevance of nucleolar localization to function, *J. Virol.* **64**, 881–885.

Dalgleish, A.G., Beverley, P.C., Clapham, P.R., Crawford, D.H., Greaves, M.F., Weiss, R.A. (1984) The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus, *Nature* **312**, 763–767.

Deacon, N.J., Tsykin, A., Solomon, A., Smith, K., Ludford-Menting, M., Hooker, D.J., McPhee, D.A., Greenway, A.L., Ellett, A., Chatfield, C. (1995) Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients, *Science* **270**, 988–991.

De Cock, K.M., Zadi, F., Adjorlolo, G., Diallo, M.O., Sassan-Morokro, M., Ekpini, E., Sibailly, T., Doorly, R., Batter, V., Brattegaard, K. (1994) Retrospective study of maternal HIV-1 and HIV-2 infections and child survival in Abidjan, Cote d'Ivoire, *BMJ* **308**, 441–443.

Dittmar, M.T., Simmons, G., Hibbitts, S., O'Hare, M., Louisirirotchanakul, S., Beddows, S., Weber, J., Clapham, P.R., Weiss, R.A. (1997) Langerhans cell tropism of human immunodeficiency virus type 1 subtype A through F isolates derived from different transmission groups, *J. Virol.* **71**, 8008–8013.

Eckert, D.M., Kim, P.S. (2001) Mechanisms of viral membrane fusion and its inhibition, *Annu. Rev. Biochem.* **70**, 777–810.

Felber, B.K., Drysdale, C.M., Pavlakis, G.N. (1990) Feedback regulation of human immunodeficiency virus type 1 expression by the Rev protein, *J. Virol.* **64**, 3734–3741.

Finzi, D., Hermankova, M., Pierson, T., Carruth, L.M., Buck, C., Chaisson, R.E., Quinn, T.C., Chadwick, K., Margolick, J., Brookmeyer, R., Gallant, J., Markowitz, M., Ho, D.D., Richman, D.D., Siliciano, R.F. (1997) Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy, *Science* **278**, 1295–1300.

Fischer, U., Huber, J., Boelens, W.C., Mattaj, I.W., Luhrmann, R. (1995) The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs, *Cell* **82**, 475–483.

Fouchier, R.A., Groenink, M., Kootstra, N.A., Tersmette, M., Huisman, H.G., Miedema, F., Schuitemaker, H. (1992) Phenotype-associated sequence variation in the third variable domain

of the human immunodeficiency virus type 1 gp120 molecule, *J. Virol.* **66**, 3183–3187.

Gallay, P., Hope, T., Chin, D., Trono, D. (1997) HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9825–9830.

Gallo, R.C., Sarin, P.S., Gelmann, E.P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V.S., Mann, D., Sidhu, G.D., Stahl, R.E., Zolla-Pazner, S., Leibowitch, J., Popovic, M. (1983) Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS), *Science* **220**, 865–867.

Garcia, J.V., Miller, A.D. (1991) Serine phosphorylation-independent downregulation of cell-surface CD4 by nef, *Nature* **350**, 508–511.

Geijtenbeek, T.B., Kwon, D.S., Torensma, R., van Vliet, S.J., van Duijnhoven, G.C., Middel, J., Cornelissen, I.L., Nottet, H.S., KewalRamani, V.N., Littman, D.R., Figdor, C.G., van Kooyk, Y. (2000) DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells, *Cell* **100**, 587–597.

Gonzalez, E., Dhanda, R., Bamshad, M., Mummidi, S., Geevarghese, R., Catano, G., Anderson, S.A., Walter, E.A., Stephan, K.T., Hammer, M.F., Mangano, A., Sen, L., Clark, R.A., Ahuja, S.S., Dolan, M.J., Ahuja, S.K. (2001) Global survey of genetic variation in *CCR5, RANTES*, and *MIP-1α*: Impact on the epidemiology of the HIV-1 pandemic, *PNAS* **98**, 5199–5204.

Gottlieb, G.S., Sow, P.S., Hawes, S.E., Ndoye, I., Redman, M., Coll-Seck, A.M., Faye-Niang, M.A., Diop, A., Kuypers, J.M., Critchlow, C.W., Respess, R., Mullins, J.I., Kiviat, N.B. (2002) Equal plasma viral loads predict a similar rate of CD4+ T cell decline in human immunodeficiency virus (HIV) type 1- and HIV-2-infected individuals from Senegal, West Africa, *J. Infect. Dis.* **185**, 905–914.

Gottlinger, H.G., Dorfman, T., Sodroski, J.G., Haseltine, W.A. (1991) Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3195–3199.

Gray, R.H., Wawer, M.J., Brookmeyer, R., Sewankambo, N.K., Serwadda, D., Wabwire-Mangen, F., Lutalo, T., Li, X., vanCott, T., Quinn, T.C. (2001) Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda, *Lancet* **357**, 1149–1153.

Greenberg, M.E., Iafrate, A.J., Skowronski, J. (1998) The SH3 domain-binding surface and an acidic motif in HIV-1 Nef regulate trafficking of class I MHC complexes, *EMBO J.* **17**, 2777–2789.

Gulick, R.M., Mellors, J.W., Havlir, D., Eron, J.J., Gonzalez, C., McMahon, D., Richman, D.D., Valentine, F.T., Jonas, L., Meibohm, A., Emini, E.A., Chodakewitz, J.A. (1997) Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy, *N. Engl. J. Med.* **337**, 734–739.

Guy, B., Kieny, M.P., Riviere, Y., Le Peuch, C., Dott, K., Girard, M., Montagnier, L., Lecocq, J.P. (1987) HIV F/3′ orf encodes a phosphorylated GTP-binding protein resembling an oncogene product, *Nature* **330**, 266–269.

Hammer, S.M., Squires, K.E., Hughes, M.D., Grimes, J.M., Demeter, L.M., Currier, J.S., Eron, J.J. Jr., Feinberg, J.E., Balfour, H.H. Jr., Deyton, L.R., Chodakewitz, J.A., Fischl, M.A., AIDS Clinical Trials Group 320 Study Team. (1997) A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less, *N. Engl. J. Med.* **337**, 725–733.

Harrer, T., Harrer, E., Kalams, S.A., Elbeik, T., Staprans, S.I., Feinberg, M.B., Cao, Y., Ho, D.D., Yilma, T., Caliendo, A.M., Johnson, R.P., Buchbinder, S.P., Walker, B.D. (1996) Strong cytotoxic T cell and weak neutralizing antibody responses in a subset of persons with stable nonprogressing HIV type 1 infection, *AIDS Res. Hum. Retroviruses* **12**, 585–592.

Heinzinger, N.K., Bukinsky, M.I., Haggerty, S.A., Ragland, A.M., Kewalramani, V., Lee, M.A., Gendelman, H.E., Ratner, L., Stevenson, M., Emerman, M. (1994) The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7311–7315.

Herrmann, C.H., Gold, M.O., Rice, A.P. (1996) Viral transactivators specifically target distinct cellular protein kinases that phosphorylate

the RNA polymerase II C-terminal domain, *Nucleic Acids Res.* **24**, 501–508.

Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M., Markowitz, M. (1995) Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection, *Nature* **373**, 123–126.

Huang, Y., Paxton, W.A., Wolinsky, S.M., Neumann, A.U., Zhang, L., He, T., Kang, S., Ceradini, D., Jin, Z., Yazdanbakhsh, K., Kunstman, K., Erickson, D., Dragon, E., Landau, N.R., Phair, J., Ho, D.D., Koup, R.A. (1996) The role of a mutant CCR5 allele in HIV-1 transmission and disease progression, *Nat. Med.* **2**, 1240–1243.

Isel, C., Karn, J. (1999) Direct evidence that HIV-1 Tat stimulates RNA polymerase II carboxyl-terminal domain hyperphosphorylation during transcriptional elongation, *J. Mol. Biol.* **290**, 929–941.

Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J., Varmus, H.E. (1988) Characterization of ribosomal frameshifting in HIV-1 gag-pol expression, *Nature* **331**, 280–283.

Jost, S., Bernard, M.C., Kaiser, L., Yerly, S., Hirschel, B., Samri, A., Autran, B., Goh, L.E., Perrin, L. (2002) A patient with HIV-1 superinfection, *N. Engl. J. Med.* **347**, 731–736.

Jowett, J.B., Planelles, V., Poon, B., Shah, N.P., Chen, M.L., Chen, I.S. (1995) The human immunodeficiency virus type 1 vpr gene arrests infected T cells in the G2 + M phase of the cell cycle, *J. Virol.* **69**, 6304–6313.

Kestler, H.W. III, Ringler, D.J., Mori, K., Panicali, D.L., Sehgal, P.K., Daniel, M.D., Desrosiers, R.C. (1991) Importance of the nef gene for maintenance of high virus loads and for development of AIDS, *Cell* **65**, 651–662.

Kino, T., Gragerov, A., Slobodskaya, O., Tsopanomichalou, M., Chrousos, G.P., Pavlakis, G.N. (2002) Human immunodeficiency virus type 1 (HIV-1) accessory protein Vpr induces transcription of the HIV-1 and glucocorticoid-responsive promoters by binding directly to p300/CBP coactivators, *J. Virol.* **76**, 9724–9734.

Kirchhoff, F., Greenough, T.C., Brettler, D.B., Sullivan, J.L., Desrosiers, R.C. (1995) Brief report: absence of intact nef sequences in a long-term survivor with nonprogressive HIV-1 infection, *N. Engl. J. Med.* **332**, 228–232.

Kjems, J., Calnan, B.J., Frankel, A.D., Sharp, P.A. (1992) Specific binding of a basic peptide from HIV-1 Rev, *EMBO J.* **11**, 1119–1129.

Koup, R.A., Safrit, J.T., Cao, Y., Andrews, C.A., McLeod, G., Borkowsky, W., Farthing, C., Ho, D.D. (1994) Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome, *J. Virol.* **68**, 4650–4655.

Le Gall, S., Erdtmann, L., Benichou, S., Berlioz-Torrent, C., Liu, L., Benarous, R., Heard, J.M., Schwartz, O. (1998) Nef interacts with the mu subunit of clathrin adaptor complexes and reveals a cryptic sorting signal in MHC I molecules, *Immunity* **8**, 483–495.

Le Rouzic, E., Mousnier, A., Rustum, C., Stutz, F., Hallberg, E., Dargemont, C., Benichou, S. (2002) Docking of HIV-1 Vpr to the nuclear envelope is mediated by the interaction with the nucleoporin hCG1, *J. Biol. Chem.* **277**, 45091–45098.

Lever, A., Gottlinger, H., Haseltine, W., Sodroski, J. (1989) Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions, *J. Virol.* **63**, 4085–4087.

Liu, R., Paxton, W.A., Choe, S., Ceradini, D., Martin, S.R., Horuk, R., MacDonald, M.E., Stuhlmann, H., Koup, R.A., Landau, N.R. (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection, *Cell* **86**, 367–377.

Lyall, E.G., Blott, M., de Ruiter, A., Hawkins, D., Mercy, D., Mitchla, Z., Newell, M.L., O'Shea, S., Smith, J.R., Sunderland, J., Webb, R., Taylor, G.P. (2001) Guidelines for the management of HIV infection in pregnant women and the prevention of mother-to-child transmission, *HIV Med.* **2**, 314–334.

Lyles, R.H., Munoz, A., Yamashita, T.E., Bazmi, H., Detels, R., Rinaldo, C.R., Margolick, J.B., Phair, J.P., Mellors, J.W., Multicenter AIDS Cohort Study. (2000) Natural history of human immunodeficiency virus type 1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men, *J. Infect. Dis.* **181**, 872–880.

Malim, M.H., Hauber, J., Le, S.Y., Maizel, J.V., Cullen, B.R. (1989) The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export

of unspliced viral mRNA, *Nature* **338**, 254–257.

Mariani, R., Kirchhoff, F., Greenough, T.C., Sullivan, J.L., Desrosiers, R.C., Skowronski, J. (1996) High frequency of defective nef alleles in a long-term survivor with nonprogressive human immunodeficiency virus type 1 infection, *J. Virol.* **70**, 7752–7764.

Mascola, J.R., Stiegler, G., VanCott, T.C., Katinger, H., Carpenter, C.B., Hanson, C.E., Beary, H., Hayes, D., Frankel, S.S., Birx, D.L., Lewis, M.G. (2000) Protection of macaques against vaginal transmission of a pathogenic HIV-1/SIV chimeric virus by passive infusion of neutralizing antibodies, *Nat. Med.* **6**, 207–210.

McCutchan, F.E., Artenstein, A.W., Sanders-Buell, E., Salminen, M.O., Carr, J.K., Mascola, J.R., Yu, X.F., Nelson, K.E., Khamboonruang, C., Schmitt, D., Kieny, M.P., McNeil, J.G., Burke, D.S. (1996) Diversity of the envelope glycoprotein among human immunodeficiency virus type 1 isolates of clade E from Asia and Africa, *J. Virol.* **70**, 3331–3338.

McKnight, A., Weiss, R.A., Shotton, C., Takeuchi, Y., Hoshino, H., Clapham, P.R. (1995) Change in tropism upon immune escape by human immunodeficiency virus, *J. Virol.* **69**, 3167–3170.

Mellors, J.W., Kingsley, L.A., Rinaldo, C.R. Jr., Todd, J.A., Hoo, B.S., Kokka, R.P., Gupta, P. (1995) Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion, *Ann. Int. Med.* **122**, 573–579.

Mellors, J.W., Rinaldo, C.R. Jr., Gupta, P., White, R.M., Todd, J.A., Kingsley, L.A. (1996) Prognosis in HIV-1 infection predicted by the quantity of virus in plasma, *Science* **272**, 1167–1170.

Migueles, S.A., Sabbaghian, M.S., Shupert, W.L., Bettinotti, M.P., Marincola, F.M., Martino, L., Hallahan, C.W., Selig, S.M., Schwartz, D., Sullivan, J., Connors, M. (2000) HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2709–2714.

Myers, E.L., Allen, J.F. (2002) Tsg101, an inactive homologue of ubiquitin ligase e2, interacts specifically with human immunodeficiency virus type 2 gag polyprotein and results in increased levels of ubiquitinated gag, *J. Virol.* **76**, 11226–11235.

Narayan, O., Cork, L.C. (1985) Lentiviral diseases of sheep and goats: chronic pneumonia leukoencephalomyelitis and arthritis, *Rev. Infect. Dis.* **7**, 89–98.

Ossendorp, F., Eggers, M., Neisig, A., Ruppert, T., Groettrup, M., Sijts, A., Mengede, E., Kloetzel, P.M., Neefjes, J., Koszinowski, U., Melief, C. (1996) A single residue exchange within a viral CTL epitope alters proteasome-mediated degradation resulting in lack of antigen presentation, *Immunity* **5**, 115–124.

Oxenius, A., Price, D.A., Easterbrook, P.J., O'Callaghan, C.A., Kelleher, A.D., Whelan, J.A., Sontag, G., Sewell, A.K., Phillips, R.E. (2000) Early highly active antiretroviral therapy for acute HIV-1 infection preserves immune function of CD8+ and CD4+ T lymphocytes, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3382–3387.

Palella, F.J. Jr., Delaney, K.M., Moorman, A.C., Loveless, M.O., Fuhrer, J., Satten, G.A., Aschman, D.J., Holmberg, S.D. (1998) Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators, *N. Engl. J. Med.* **338**, 853–860.

Parren, P.W., Moore, J.P., Burton, D.R., Sattentau, Q.J. (1999) The neutralizing antibody response to HIV-1: viral evasion and escape from humoral immunity, *AIDS* **13**, S137–S162.

Patel, C.A., Mukhtar, M., Pomerantz, R.J. (2000) Human immunodeficiency virus type 1 Vpr induces apoptosis in human neuronal cells, *J. Virol.* **74**, 9717–9726.

Phillips, R.E., Rowland-Jones, S., Nixon, D.F., Gotch, F.M., Edwards, J.P., Ogunlesi, A.O., Elvin, J.G., Rothbard, J.A., Bangham, C.R., Rizza, C.R. (1001) Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition, *Nature* **354**, 453–459.

Piguet, V., Chen, Y.L., Mangasarian, A., Foti, M., Carpentier, J.L., Trono D. (1998) Mechanism of Nef-induced CD4 endocytosis: Nef connects CD4 with the mu chain of adaptor complexes, *EMBO J.* **17**, 2472–2481.

Pircher, H., Moskophidis, D., Rohrer, U., Burki, K., Hengartner, H., Zinkernagel, R.M. (1990) Viral escape by selection of cytotoxic T cell-resistant virus variants in vivo, *Nature* **346**, 629–633.

Pornillos, O., Alam, S.L., Davis, D.R., Sundquist, W.I. (2002) Structure of the Tsg101 UEV domain in complex with the PTAP motif of the HIV-1 p6 protein, *Nat. Struct. Biol.* **9**, 812–817.

Quinn, T.C., Wawer, M.J., Sewankambo, N., Geberding, J.L. (1994) Incidence and prevalence of human immunodeficiency virus, hepatitis B virus, hepatitis C virus, and cytomegalovirus among health care personnel at risk for blood exposure: final report from a longitudinal study, *J. Infect. Dis.* **170**, 1410–1417.

Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K. (1985) Complete nucleotide sequence of the AIDS virus, HTLV-III, *Nature* **313**, 277–284.

Reitter, J.N., Means, R.E., Desrosiers, R.C. (1998) A role for carbohydrates in immune evasion in AIDS, *Nat. Med.* **4**, 679–684.

Rosenberg, E.S., Altfeld, M., Poon, S.H., Phillips, M.N., Wilkes, B.M., Eldridge, R.L., Robbins, G.K., D'Aquila, R.T., Goulder, P.J., Walker, B.D. (2002) Immune control of HIV-1 after early treatment of acute infection, *Nature* **407**, 523–526.

Rosenberg, E.S., Billingsley, J.M., Caliendo, A.M., Boswell, S.L., Sax, P.E., Kalams, S.A., Walker, B.D. (1997) Vigorous HIV-1-specific CD4+ T cell responses associated with control of viremia, *Science* **278**, 1447–1450.

Rowland-Jones, S., Sutton, J., Ariyoshi, K., Dong, T., Gotch, F., McAdam, S., Whitby, D., Sabally, S., Gallimore, A., Corrah, T. (1995) HIV-specific cytotoxic T-cells in HIV-exposed but uninfected Gambian women, *Nat. Med.* **1**, 59–64.

Schuitemaker, H., Kootstra, N.A., de Goede, R.E., de Wolf, F., Miedema, F., Tersmette, M. (1991) Monocytotropic human immunodeficiency virus type 1 (HIV-1) variants detectable in all stages of HIV-1 infection lack T-cell line tropism and syncytium-inducing ability in primary T-cell culture, *J. Virol.* **65**, 356–363.

Schwartz, D., Sharma, U., Busch, M., Weinhold, K., Matthews, T., Lieberman, J., Birx, D., Farzedagen, H., Margolick, J., Quinn T. (1994) Absence of recoverable infectious virus and unique immune responses in an asymptomatic HIV + long-term survivor, *AIDS Res. Hum. Retroviruses* **10**, 1703–1711.

Schwartz, O., Marechal, V., Le Gall, S., Lemonnier, F., Heard, J.M. (1996) Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein, *Nat. Med.* **2**, 338–342.

Sheehy, A.M., Gaddis, N.C., Choi, J.D., Malim, M.H. (2002) Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein, *Nature* **418**, 646–650.

Sodroski, J., Goh, W.C., Rosen, C., Dayton, A., Terwilliger, E., Haseltine, W. (1986) A second post-transcriptional trans-activator gene required for HTLV-III replication, *Nature* **321**, 412–417.

Sodroski, J., Rosen, C., Wong-Staal, F., Salahuddin, S.Z., Popovic, M., Arya, S., Gallo, R.C., Haseltine, W.A. (1985) Trans-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat, *Science* **227**, 171–173.

Spira, A.I., Marx, P.A., Patterson, B.K., Mahoney, J., Koup, R.A., Wolinsky, S.M., Ho, D.D. (1996) Cellular targets of infection and route of viral dissemination after an intravaginal inoculation of simian immunodeficiency virus into rhesus macaques, *J. Exp. Med.* **183**, 215–225.

Swingler, S., Mann, A., Jacque, J., Brichacek, B., Sasseville, V.G., Williams, K., Lackner, A.A., Janoff, E.N., Wang, R., Fisher, D., Stevenson, M. (1999) HIV-1 Nef mediates lymphocyte chemotaxis and activation by infected macrophages, *Nat. Med.* **5**, 997–1003.

Temin, H.M., Mizutani, S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus, *Nature* **226**, 1211–1213.

Tersmette, M., de Goede, R.E., Al, B.J., Winkel, I.N., Gruters, R.A., Cuypers, H.T., Huisman, H.G., Miedema, F. (1988) Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex, *J. Virol.* **62**, 2026–2032.

VerPlank, L., Bouamr, F., LaGrassa, T.J., Agresta, B., Kikonyogo, A., Leis, J., Carter, C.A. (2001) Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55(Gag), *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7724–7729.

Wagner, L., Yang, O.O., Garcia-Zepeda, E.A., Ge, Y., Kalams, S.A., Walker, B.D., Pasternack, M.S., Luster, A.D. (1998) Beta-chemokines are released from HIV-1-specific cytolytic T-cell granules complexed to proteoglycans, *Nature* **391**, 908–911.

Walker, B.D., Chakrabarti, S., Moss, B., Paradis, T.J., Flynn, T., Durno, A.G., Blumberg, R.S., Kaplan, J.C., Hirsch, M.S., Schooley, R.T. (1987) HIV-specific cytotoxic T lymphocytes in seropositive individuals, *Nature* **328**, 345–348.

Walker, B.D., Flexner, C., Paradis, T.J., Fuller, T.C., Hirsch, M.S., Schooley, R.T., Moss, B. (1988) HIV-1 reverse transcriptase is a target for cytotoxic T lymphocytes in infected individuals, *Science* **240**, 64–66.

Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H., Jones, K.A. (1998) A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA, *Cell* **92**, 451–462.

Wei, X., Ghosh, S.K., Taylor, M.E., Johnson, V.A., Emini, E.A., Deutsch, P., Lifson, J.D., Bonhoeffer, S., Nowak, M.A., Hahn, B.H., Saag, M.S., Shaw, G.M. (1995) Viral dynamics in human immunodeficiency virus type 1 infection, *Nature* **373**, 117–122.

Yang, O.O., Kalams, S.A., Rosenzweig, M., Trocha, A., Jones, N., Koziel, M., Walker, B.D., Johnson, R.P. (1996) Efficient lysis of human immunodeficiency virus type 1-infected cells by cytotoxic T lymphocytes, *J. Virol.* **70**, 5799–5806.

Yang, X., Herrmann, C.H., Rice, A.P. (1996) The human immunodeficiency virus Tat proteins specifically associate with TAK in vivo and require the carboxyl-terminal domain of RNA polymerase II for function, *J. Virol.* **70**, 4576–4584.

Yao, X.J., Gottlinger, H., Haseltine, W.A., Cohen, E.A. (1992) Envelope glycoprotein and CD4 independence of vpu-facilitated human immunodeficiency virus type 1 capsid export, *J. Virol.* **66**, 5119–5126.

Zaitseva, M., Blauvelt, A., Lee, S., Lapham, C.K., Klaus-Kovtun, V., Mostowski, H., Manischewitz, J., Golding, H. (1997) Expression and function of CCR5 and CXCR4 on human Langerhans cells and macrophages: implications for HIV primary infection, *Nat. Med.* **3**, 1369–1375.

Algae: *see* Chlamydomonas

Alternatively Spliced Genes

Jane Y. Wu[1,2], *Liya Yuan*[1], *and Necat Havlioglu*[1]
[1] *Washington University School of Medicine, St. Louis, MO, USA*
[2] *Departments of Pediatrics, Cell and Developmental Biology and Pharmacology; John F. Kennedy Vanderbilt University Medical Center, Nashville, TN USA*

1 Pre-mRNA Splicing and Splicing Machinery 127
1.1 Splicing Machinery: Spliceosome 127
1.2 Splicing Signals 128
1.3 Spliceosomal UsnRNP Biogenesis 137
1.4 Spliceosome Assembly 138
1.5 Biochemical Mechanisms of pre-mRNA Splicing 141

2 Alternative pre-mRNA Splicing 142
2.1 Alternative Splicing and its Role in Regulating Gene Activities and Generating Genetic Diversity 142
2.1.1 Different Patterns of Alternative Splicing 143
2.1.2 Alternative Splicing and Genetic Diversity 143
2.2 Mechanisms Underlying Alternative Splicing Regulation 144
2.2.1 Splicing Signals and Splicing Regulatory Elements 145
2.2.2 Trans-acting Splicing Regulators 148
2.3 Tissue-specific and Developmentally Regulated Alternative Splicing 151
2.4 Regulation of Alternative Splicing in Response to Extracellular Stimuli 152

3 Pre-mRNA Splicing and Human Diseases 152
3.1 Splicing Defects in Human Diseases 153
3.2 Molecular Mechanisms Underlying Splicing Defects Associated with Disease 157

4 Perspectives on Diagnosis and Treatment of Diseases Caused by pre-mRNA Splicing Defects 160
4.1 Diagnosis of Human Diseases Caused by Splicing Defects 160

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

4.2 Potential Therapeutic Approaches 161
4.2.1 Oligonucleotide-based Approaches: Antisense, RNAi, and Chimeric Molecules 161
4.2.2 Ribozymes 161
4.2.3 SMaRT 162
4.2.4 Chemical Compounds 162

5 Concluding Remarks 162

Acknowledgment 163

Bibliography 163
Books and Reviews 163
Primary Literature 166

Keywords

Pre-mRNA
Nascent transcripts that are precursors of mature messenger RNAs.

Exon
Gene regions that are present in mature messenger RNA transcripts.

Intron
Intervening sequences that are removed from pre-mRNA transcripts and not represented in mature mRNA species.

Pre-mRNA Splicing
The process of removing introns from pre-mRNAs and ligating exons to form mature mRNA transcripts.

UsnRNPs
Uridine-rich small nuclear ribonucleoprotein particles.

Spliceosome
The macromolecular machine in which pre-mRNA splicing reactions take place.

5′ and 3′ Splice Sites
Sequences that spliceosomes recognize at 5′ and 3′ ends of an intron. They are also named splice donor and splice acceptor sites respectively.

Alternative Splicing
The process in which cells selectively use different combinations of splice sites to generate two or more mRNA transcripts from a single pre-mRNA species.

Splicing Regulatory Elements
Specific sequences in introns or exons that are recognized by trans-acting factors to allow either enhancement or repression of splicing.

Splicing Regulators
Trans-acting factors that interact with splicing regulatory elements to either stimulate or inhibit splicing.

An important step of eukaryotic gene expression is pre-mRNA splicing, the process of removing intervening sequences (introns) from the nascent transcript (messenger RNA precursor, or pre-mRNA). The discovery of split genes in the viral genome and subsequent research in the field of pre-mRNA splicing have greatly advanced our understanding of mammalian gene regulation. Studies on pre-mRNA splicing have also facilitated sequence analyses of the human genome. With the completion of human genome sequencing, it is now further appreciated that pre-mRNA splicing and alternative splicing play critical roles in regulating gene expression and in enhancing genetic diversity.

Evolutionarily, the basic machinery for pre-mRNA splicing appears to be highly conserved among different species of metazoans. In *Saccharomyces cerevisiae*, although only a small percentage of genes undergo splicing, more than 100 genes have been identified that are either dedicated to or involved in pre-mRNA splicing. In mammals, pre-mRNA splicing is a crucial step for gene expression because the vast majority of mammalian transcription units contain one or more introns that must be accurately removed to form mature and functional messenger RNA (mRNA) species. In this chapter, we review the current knowledge about mammalian pre-mRNA splicing, with special emphasis on the aspects related to the pathogenesis or treatment of human diseases.

1 Pre-mRNA Splicing and Splicing Machinery

1.1 Splicing Machinery: Spliceosome

Pre-mRNA splicing is the first step of the major post-transcriptional processes in eukaryotic gene expression (Fig. 1). The biochemical reactions of pre-mRNA splicing occur in a macromolecular machine named the *spliceosome*. This large RNA–protein complex contains, in addition to the pre-mRNA substrate, several uridine-rich small nuclear ribonucleoprotein (UsnRNP) particles as well as a myriad of non-snRNP protein factors. The splicing machinery is similar in complexity and size to that of the protein synthesis machinery, the ribosome. In all cases studied so far, splicing reactions take place in the mature spliceosome inside the nucleus.

For processing the majority of introns (the major class, also called the U2-type

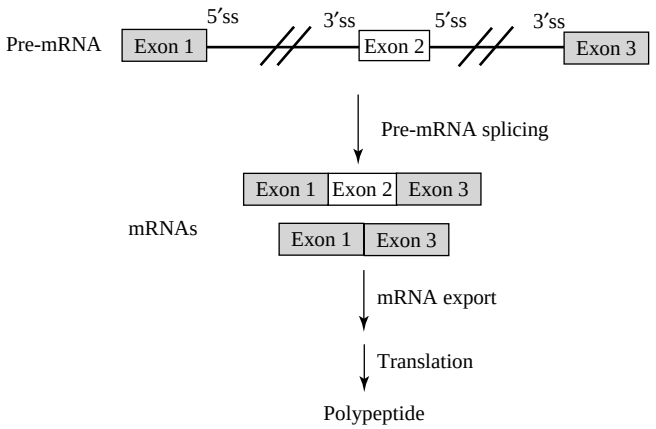

Fig. 1 The schematics for major posttranscriptional processes in eukaryotic gene expression. Different splice isoforms of mRNAs can be produced as a result of alternative selection of splice sites.

intron), the spliceosome contains U1, U2, U4/6, and U5snRNPs. The splicing of the minor class of introns (also called the U12-type) occurs in the spliceosome containing U11 and U12, in addition to U4atac, U6atac, and U5snRNPs (Figs. 2 and 3). This chapter will focus on the splicing of introns in the major class. This is because this class accounts for more than 99% of known introns. Most of our knowledge about pre-mRNA splicing has also come from studies of the major class of introns.

Comparison of genes coding for components of splicing machinery, including both snRNAs and protein factors, reveals a high degree of conservation through evolution. Amazing similarity is found among spliceosomal components from yeast, fruit fly, and human. A recent study using nanoscale microcapillary liquid chromatography tandem mass spectrometry to analyze partially purified human spliceosomes assembled on model splicing substrates revealed 145 distinct spliceosomal proteins. This suggests that the spliceosome is one of the most complex macromolecular machines for mammalian gene expression. Table 1 shows the currently known human proteins identified in functional spliceosomes assembled on model splicing substrates. The role of individual proteins in splicesome assembly and in splicing regulation will be discussed.

1.2 Splicing Signals

The sites for cleavage and ligation in splicing reactions are defined by conserved splicing signals in the pre-mRNA, including the 5′ splice site (5′ss), the branch site, the polypyrimidine tract, and the 3′ splice site (3′ss) consensus sequence. These cis-elements are important sites for RNA–RNA and protein–RNA interactions. Systematic analyses of both yeast and higher eukaryotic splicing signal sequences led to the classification of U2 and U12 types of introns based on the key UsnRNPs involved in the recognition of the branch sites in the corresponding introns (Fig. 2).

In general, the spliceosome for U2-type introns contains U1, U2, and U4/U6,

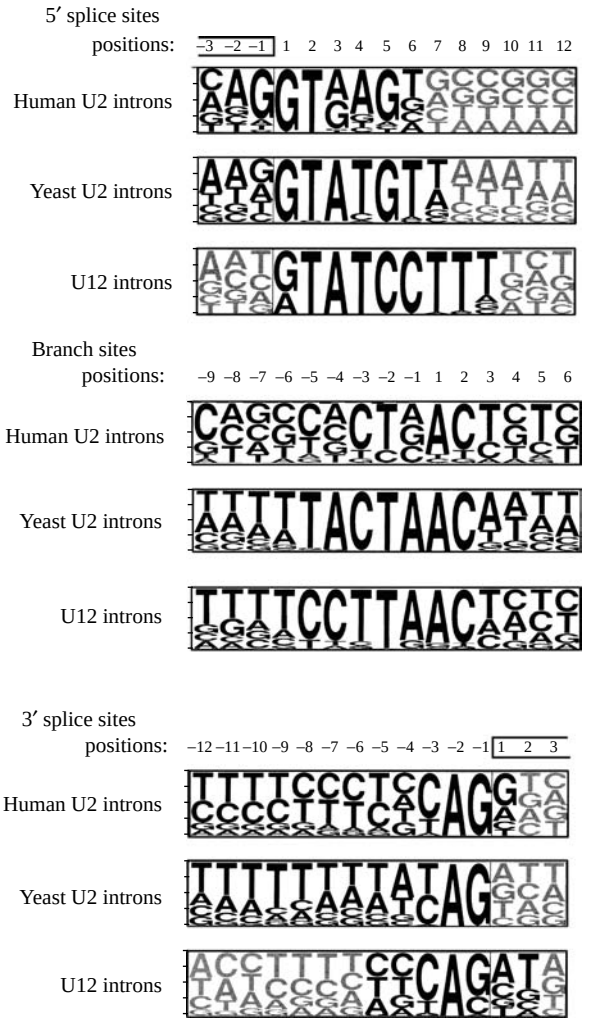

Fig. 2 The sequence features at the 5′ splice site, branch site, and 3′ splice site as derived from representative sets of introns of corresponding types. (Modified with permission from Burge, C.B., Tuschl, T.H., Sharp, P.A. (1999). in: The RNA World, Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds) *The RNA World*, Cold Spring Harbor Laboratory Press, New York, pp. 525–560). The frequency of each nucleotide (A, G, T, or C) at each sequence position is depicted by the heights of the corresponding letters in the diagram.

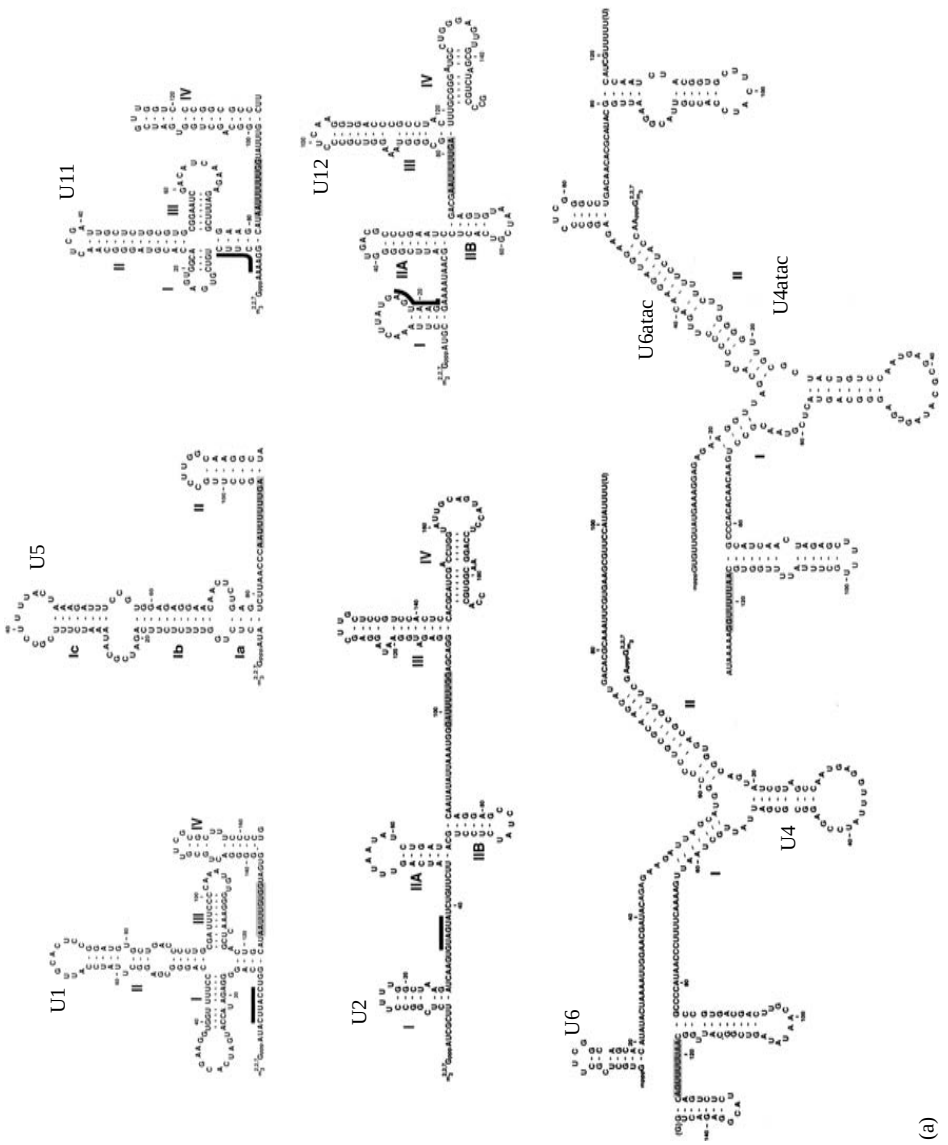

(a)

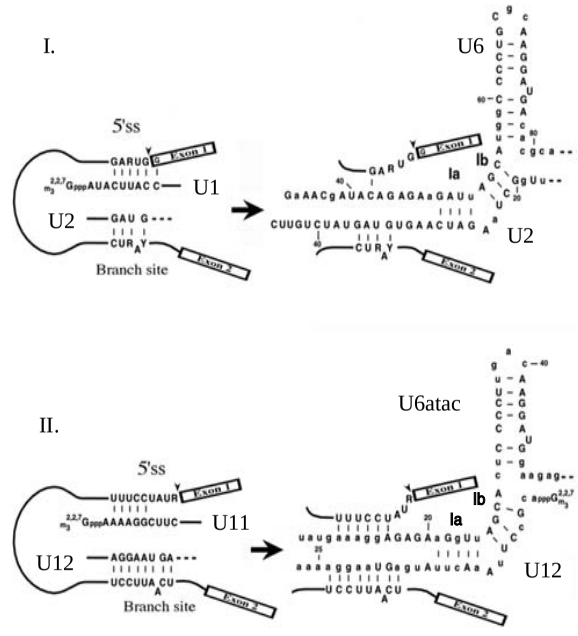

(b)

Fig. 3 Sequences of human spliceosomal UsnRNAs and their interactions with splice sites. Panel (a) Sequences and secondary structures of spliceosomal snRNAs: U1, U2, U4/6 (for the U2 type), and U11, U12, U4atac/U6atac snRNA (for the U12 type) as well as U5snRNAs (for both classes). The 5′ to 3′ orientation of U4 and U4atac is from right to left, and that of U1, U11, U2, U12, and U5snRNAs is from left to right. The RNA helices are indicated by Roman numerals, and nucleotide positions by Arabic numerals. The shaded nucleotides are binding sites for Sm proteins, and regions marked with black lines form base-pairing interactions with the pre-mRNA. (Modified with permission from Burge, C.B., Tuschl, T.H., Sharp, P.A. (1999). in: The RNA World, Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds) *The RNA World*, Cold Spring Harbor Laboratory Press, New York, pp. 525–560.) Panel (b) Base-pairing interactions in the U2 and U12 types of spliceosome. The exon-intron boundaries at 5′ splice sites are marked by arrowheads. During early stages of spliceosome assembly, U1 or U11snRNP interacts with the 5′ splice site, and U2 or U12snRNP with the branch sites. These interactions are displaced in mature spliceosome by U6 or U6atac associated with U2 and U12 respectively. In these drawings, the distances between interacting nucleotides do not reflect corresponding physical distances in the spliceosome. (Adapted from Tarn, W.Y., Steitz, J.A. (1996b) Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns, *Science* **273**(5283), 1824–1832.)

Tab. 1 Proteins detected in human spliceosomes.

Acce.#	*Protein name*	*Yeast homolog (SGD ORF)*	*Motifs*	*Cal. MWt*
A. snRNP proteins				
Sm/LSm core proteins				
P14678	Sm B/B'	SMB1 (YER029C)	Sm 1 and 2	24 610
P13641	Sm D1	SMD1 (YGR074W)	Sm 1 and 2	13 282
P43330	Sm D2	SMD2 (YLR275W)	Sm 1 and 2	13 527
P43331	Sm D3	SMD3 (YLR147C)	Sm 1 and 2	13 916
P08578	Sm E	SME1 (YOR159C)	Sm 1 and 2	10 804
Q15356	Sm F	SMX3 (YPR182W)	Sm 1 and 2	9725
Q15357	Sm G	SMX2 (YFL017W-A)	Sm 1 and 2	8496
Q9Y333	hLSm2	LSM2 (YBL026W)	Sm 1 and 2	10 835
Q9Y4Z1	hLSm3	LSM3 (YLR438C-A)	Sm 1 and 2	11 714
Q9Y4Z0	hLSm4	LSM4 (YER112W)	Sm 1 and 2	15 350
Q9Y4Y8	hLSm6	LSM6 (YDR378C)	Sm 1 and 2	9128
Q9UK45	hLSm7	LSM7 (YNL147W)	Sm 1 and 2	11 602
U1snRNP-specific proteins				
P08621	U1-70kD	SNP1 (YIL061C)	RS, 1 RRM	70 081
P09012	U1 A	MUD1 (YBR119W)	2 RRMs	31 280
P09234	U1 C	YHC1 (YLR298C)	ZF	17 394
AB034205	LUC7A	LUC7 (YDL087C)	RS-rich	51 466
U2snRNP-specific proteins				
AF054284	SAP155 (SF3b155)	HSH155 (YMR288W)	PP2A	1 45 815
Q13435	SAP145 (SF3b145)	CUS1 (YMR240C)	Pro-rich	97 657
AJ001443	SAP130 (SF3b130)	RSE1 (YML049C)		1 35 592
Q15459	SAP114 (SF3a120)	PRP21 (YJL203W)	SWAP, UBQ	88 886
Q15428	SAP62 (SF3a66)	PRP11 (YDL043C)	ZF	49 196
A55749	SAP61(SF3a60)	PRP9 (YDL030W)	ZF	58 849
Q15427	SAP49 (SF3b49)	HSH49 (YOR319W)	2 RRMs	44 386
P09661	U2 A′	LEA1 (YPL213W)	Leu-rich	28 444
P08579	U2 B″	MSL1 (YIR009W)	2 RRMs	25 486
Q9Y3B4	p14	SNU17 (YIR005W)	1 RRM	14 584
U5snRNP-specific proteins				
AB007510	U5-220kD (PRPC8)	PRP8 (YHR165C)		2 73 785
O75643	U5-200kD	BRR2 (YER172C)	DExD, Helicase	1 94 479
BC002360	U5-116kD	SNU114 (YKL173W)	G domain	1 09 478
BC001666	U5-102kD	PRP6 (YBR055C)	HAT	1 06 925
BC002366	U5-100kD	PRP28 (YDR243C)	DExD, Helicase, RS	95 583
BC000495	U5-52kD	SNU40 (YHR156C)	GYF	37 646
AF090988	U5-40kD		WD40s	39 299
O14834	U5-15kD	DIB1 (YPR082C)		16 786
U4/U6snRNP-specific proteins				
T50839	U4/U6-90kD (HPRP3)	PRP3 (YDR473C)	PWI	77 529
BC007424	U4/U6-60kD (HPRP4)	PRP4 (YPR178W)	WD40s	58 321

Tab. 1 (*continued*)

Acce.#	*Protein name*	*Yeast homolog (SGD ORF)*	*Motifs*	*Cal. MWt*
AL050369	U4/U6-61kD (PRPF31)	PRP31 (YGR091W)	NOP	55 424
AF036331	U4/U6-20kD	CPH1 (YDR155C)	Cyclophilin	19 207
P55769	U4/U6-15.5kD	SNU13 (YEL026W)	Putative RRM	14 174
U4/U6.U5 tri-snRNP specific proteins				
T00034	Tri-snRNP 110kD	SNU66 (YOR308C)	RS	90 255
AF353989	Tri-snRNP 65kD	SAD1 (YFR005C)	RS, UBQ, ZF	65 415
X76302	Tri-snRNP 27kD		RS	18 859
B. Non-snRNP spliceosomal proteins				
SR proteins				
Q08170	SRp75		2 RRMs, RS	56 792
Q05519	p54/SFRS11		1 RRM, RS	53 542
Q13247	SRp55		2 RRMs, RS	39 568
Q13243	SRp40		2 RRMs, RS	31 264
Q07955	ASF/SF2		2 RRMs, RS	27 613
Q16629	9G8		1 RRM, RS	27 367
Q01130	SC35		1 RRM, RS	25 575
Q13242	SRp30c		2 RRMs, RS	25 542
AF057159	hTra2		1 RRM, RS	21 935
P23152	SRp20		1 RRM, RS	19 330
Other spliceosomal proteins with known motifs				
NP_057417	SRm300		RS-rich	2 51 965
P26368	U2 AF65	MUD2 (YKL074C)	RS, 3 RRMs	53 501
Q01081	U2AF35		RS	27 872
Y08765	SF1	BBP (Ylr116w)	ZF, KH	68 632
AB018344	hPrp5	PRP5 (YBR237W)	DExD, Helicase, RS	1 17 461
O60231	hPrp2	PRP2 (YNR011C)	DExD, HELICc	1 19 172
P38919	IF4N	FAL1 (YDR021W)	DExD, HELICc	46 833
P17844	p68	DBP2 (YNL112W)	DExD, Helicase	69 148
Q92841	p72	DBP2 (YNL112W)	DExD, HELICc	72 371
T00333	fSAP164		DExD	1 63 986
Q08211	RHA	(YLR419W)	DExD, HELICc, DSRM	1 40 877
O00571	DDX3	DBP1 (YPL119C)	DExD, HELICc, SR	73 243
Q9UJV9	Abstrakt		DExD, HELICc	69 738
P42285	fSAP118	MTR4 (YJL050W)	DExD	1 17 790
BC002548	fSAP113		DExD	1 13 671
NP_055792	fSAP152 (Acinus)		1 RRM	1 51 887
Q14011	CIRP (cold inducible RNA-binding)		1 RRM, Gly-rich	18 648
BC007871	SPF45		1 RRM, Gly-rich	44 962
NP_008938	CF I-68kD (cleavage factor)		1 RRM	59 209

(*continued overleaf*)

Tab. 1 (*continued*)

Acce.#	*Protein name*	*Yeast homolog (SGD ORF)*	*Motifs*	*Cal. MWt*
AC004858	fSAP94		1 RRM, PWI	94 122
P52298	hCBP20	CBP20 (YPL178W)	1 RRM	18 001
BC003402	fSAP47	ECM2 (YBR065C)	1 RRM, ZF	46 896
BC006474	fSAPa		1 RRM, SWAP	Partial
AAB18823	TAT-SF1	CUS2 (Ynl286w)	2 RRMs	85 759
I55595	fSAP59 (transcription cofactor)		2 RRMs	58 657
U76705	IMP3		2 RRMs, 4 KH	63 720
NP_073605	OTT		3 RRMs	1 02 135
BC008875	PUF60		3 RRMs	58 171
Q15097	PAB2	PAB1 (YER165W)	4 RRMs, PABP	58 518
AF356524	SHARP (transcription cofactor)		4 RRMs	4 02 248
P41223	fSAP17	CWC14 (YCR063W)	ZF	16 844
O43670	ZNF207 (transcription factor)		ZF	50 751
AF044333	PLRG1	PRP46 (YPL151C)	WD40s	57 194
BC008719	hPrp19	PRP19 (YLL036C)	WD40s	55 181
BC006849	hTEX1	TEX1 (YNL253W)	WD40s	38 772
BC002876	fSAP57	PFS2 (YNL317W)	WD40s	57 544
BC003118	fSAP35		WD40s	34 849
AF083383	SPF38		WD40s	34 290
Other spliceosomal proteins with known motifs				
XM_087118	hFBP3		1 WW	24 297
AF049523	hFBP11	PRP40 (YKL012W)	2 WW, 1 FF	Partial
T08599	CA150		3 WW, 6 FF	1 23 960
XP_042023	CyP64	CPR3 (YML078W)	Cyclophilin, WD40	64 222
AF271652	PPIL3	CPR6 (YLR216C)*	Cyclophilin	18 155
S64705	PPIL2	CYP5 (YDR304C)*	Cyclophilin, Ubox	58 823
BC003048	PPIL1	CYP2 (YHR057C)*	Cyclophilin	18 237
Q9UNP9	PPIE	CPH1 (YDR155C)*	Cyclophilin	33 431
AY029347	hPrp4 kinase		TyrKc	1 16 973
Q9NYV4	CrkRS	CTK1 (YKL139W)	TyrKc	1 64 155
T12531	TIP39	(YLR424W)*	Gly-rich	96 820
BC001403	CF I-25kD (cleavage factor)		NUDIX	26 227
Q09161	hCBP80	CBP80 (YMR125W)	MIF4G	91 839
AB046824	fSAPb	CWC22 (YGR278W)	MIF4G	Partial
T02672	fSAPc		ERM	Partial
AF255443	hCrn	CLF1 (YLR117C)	HAT	99 201
BC007208	XAB2 (transcription cofactor)	SYF1 (YDR416w)	HAT	1 00 010

Tab. 1 (*continued*)

Acce.#	*Protein name*	*Yeast homolog (SGD ORF)*	*Motifs*	*Cal. MWt*
Q13573	SKIP	PRP45 (YAL032C)	SKIP	61 494
A53545	hHPR1	HPR1 (YDR138W)	DEATH	75 627
AF083385	SPF30**		TUDOR	26 711
NP_055095	SPF31		DnaJ	30 986
Q13123	RED		RED	65 630
Spliceosomal proteins without known motifs				
NP_001244	CDC5L	CEF1 (YMR213W)		92 250
AAK21005	ASR2			1 00 665
AF441770	hTHO2	THO2 (YNL139C)		1 69 573
I39463	fSAP79			78 536
AK027098	fSAP24			23 671
BC001621	SNP70			69 998
BC006350	fSAP71			70 521
P55081	MFAP1			51 855
AJ276706	WTAP			44 243
O15355	PP2Cγ	PTC3 (YBL056W)		59 272
AJ279080	fSAP105 (putative transcription factor)			1 04 804
NP_056311	fSAP121			1 21 193
BC004442	fSAP33	ISY1 (YJR050W)		32 992
AF161497	CCAP2	CWC15 (YDR163W)		26 610
Q9Y5B6	GCFC (putative transcription factor)			29 010
NP_073210	DGSI			52 567
NP_110517	NAP (nuclear protein inducing cell death)			65 173
NP_056299	fSAP29			28 722
NP_037374	fSAP23 (putative transcription factor)			22 774
BC004122	fSAP11			10 870
BC000216	fSAP18			18 419
AF081788	SPF27			26 131
C. Late-acting spliceosomal proteins				
Catalytic step II and late-acting proteins				
Q92620	hPrp16	PRP16 (YKR086W)	DExD, Helicase, RS	1 40 473
Q14562	hPrp22	PRP22 (YER013W)	DExD, Helicase, RS	1 39 315
O43143	hPrp43	PRP43 (YGL120C)	DExD, Helicase	92 829
BC010634	hSlu7	SLU7 (YDR088C)	ZF	68 343
AF038392	hPrp17	PRP17 (YDR364C)	WD40s	65 521
BC000794	hPrp18	PRP18 (YGR006W)		39 860

(*continued overleaf*)

Tab. 1 (*continued*)

Acce.#	*Protein name*	*Yeast homolog (SGD ORF)*	*Motifs*	*Cal. MWt*
Spliced mRNP/EJC proteins				
AF048977	SRm160		PWI, RS	93 519
Q13838	UAP56	SUB2 (YDL084W)	DExD, Helicase	48 991
PJC4525	RNPS1		1 RRM, Ser-rich	34 208
NM_005782	Aly	YRA1 (YDR381W)	1 RRM	26 861
Q9Y5S9	Y14		1 RRM	19 889
P50606	Magoh		Mago_nashi	17 164

Source: (Adapted with permission from Zhou, Z., Licklider, L.J., Gygi, S.P., Reed, R. (2002) Comprehensive proteomic analysis of the human spliceosome, *Nature* **419**(6903), 182–185).
Note: A. snRNP proteins. B. Non-snRNP spliceosomal proteins. C. Late-acting spliceosomal proteins. In each list, spliceosomal proteins are grouped by their structural motifs. Corresponding GenBank accession numbers are shown (Acce#), together with their yeast homologs (as SGD open reading frame, SGD ORF), sequence motifs, and calculated molecular weight (Cal. Mwt).
COLD: cold shock RNA-binding domain; Cyclophilin: cyclophilin type peptidyl–prolyl cis–trans isomerase; DEATH: domain found in proteins involved in cell death; DExD: DExD/H-like helicases superfamily; DSRM: double-stranded RNA-binding motif; ERM: ezrin/radixin/moesin family motif; fSAP: functional spliceosome-associated protein; FF: two conserved F residues; G domain: GTP-binding domain that contains a P-loop motif; GYF: contains conserved G-T-F residues; HAT: Half-A-TPR (tetratrico-peptide repeat); Gly-rich: peptide sequences rich in G, R, S residues; HELICc: helicase superfamily c-terminal domain; KH: hnRNP K homology RNA-binding domain; MIF4G: middle domain of eukaryotic initiation factor 4G; NOP: putative snoRNA binding domain; NUDIX: mutT-like domain; PABP: poly-adenylate binding protein, unique domain; PP2A: protein phosphatase 2A repeat; PWI: domain in splicing factors; RED: protein with extensive stretch of alternating R and E or D; RRM: RNA recognition motif; RS: arginine–serine-rich domains; SKIP: conserved domain found in chromatinic proteins; Sm: snRNP Sm proteins; SWAP: suppressor-of-white-apricot splicing regulator; TUDOR: a domain present in several RNA-binding proteins; TyrKc: tyrosine kinase, catalytic domain; UBQ: ubiquitin homologs; Ubox: modified RING finger domain; WD40s: WD40 repeats, structural repeats of the beta propeller domain; WW: domain with 2 conserved W residues, interacting with proline-rich polypeptides; ZF: Zinc finger domain.

whereas the U12-type of spliceosome involves U11, U12, and U4atac/U6atac, with both sharing U5snRNP. Although nucleotide sequences of U11, U12, and U4atac/U6atac snRNAs are different from their counterparts in the spliceosomes for U2-type introns, the predicted secondary structures of the corresponding snRNAs have striking similarity (Fig. 3a). In addition, the specific interactions between the spliceosomal snRNAs and pre-mRNA substrates also appear to be highly similar, as shown in Fig. 3(b). Both U2- and U12-types of introns can be found in the same genes. There is evidence suggesting that the two types of spliceosomes may interact and share some protein components in addition to the common U5snRNP. The evolutionary origins of these two types of spliceosomes remain largely speculative.

U12-class introns contain more conserved sequences at the 5′ss. The sequences around branch sites of the U12-class introns are also highly conserved with the TCCTTA<u>A</u>C (the underlined A as the branch site) consensus signal located approximately 10 to 20 nucleotides upstream

of the 3′ss. A typical U12-class intron lacks a polypyrimidine tract between the branch site and the 3′ss.

In the yeast, *S. cerevisiae*, only a small fraction of genes contain introns, and these are usually short introns (approximately 240 introns, averaging 270 nucleotides in length). Splicing signals in yeast introns are highly conserved and may contain sufficient information for defining splice junctions, especially considering the number of genes devoted to or associated with pre-mRNA splicing in the yeast genome.

In mammals, pre-mRNA transcripts are usually much longer and contain multiple introns of variable sizes. In humans, the average size of exons is 150 nucleotides, and that of introns is approximately 3500 nucleotides. Mammalian introns can be as large as 500 kbp. The basic splicing signals in mammalian pre-mRNAs are degenerate, especially in the case of U2-type of introns. The branch sites for U2-type introns are highly divergent. At both the 5′ss and 3′ss, only two nucleotides (/GT at the 5′ss and AG/ at 3′ss) are highly conserved. As a result, the nucleotide sequences surrounding the splice junctions, the 5′ss and 3′ss, usually contain only a limited amount of information. This is not sufficient for conferring the specificity required to achieve accurate splice site selection. In mammals, the recognition of not only exon–intron junction sequences but also the regulatory elements in intronic and exonic regions is important for defining splice junctions and maintaining splicing fidelity. In addition, multiple networks of interactions among the machineries for transcription, cap formation, splicing, and polyadenylation may also influence splice site selection. This high degree of degeneracy in the splicing signals in mammalian pre-mRNA transcripts provides the flexibility for alternative selection and pairing of different splice sites, a fundamental mechanism for regulating alternative splicing.

1.3 Spliceosomal UsnRNP Biogenesis

As essential subunits of the splicing machinery, spliceosomal UsnRNPs contain not only uridine-rich snRNAs but also a number of polypeptides. U1, U2, U4, and U5snRNAs are transcribed by RNA polymerase II as precursors containing additional 3′ nucleotides. After acquiring a monomethylated guanosine (m7G) cap structure, these pre-UsnRNAs are exported to the cytoplasm in a pathway dependent on the m7G cap, the cap-binding complex (CBC), RanGTP, and phosphorylated adaptor for RNA export. In the cytoplasm, pre-UsnRNAs interact with Sm proteins including B/B′, D3, D2, D1, E, F, and G (Table 1A) to form the snRNP core structure.

The Sm protein-binding sites in U1, U2, U4, and U5snRNAs are highly conserved, containing two stem-loop structures flanking PuAU4-6GPu sequence (Fig. 3a). A number of proteins interacting with Sm proteins have been identified, including the protein product of SMN (survival of motor neuron) gene, SMN-interacting protein/Gemin2, Gemin3, and Gemin4. Genetic defects in the SMN gene cause spinal muscular atrophy (SMA), possibly by interfering with UsnRNP Sm core assembly and therefore deficiency in UsnRNP biogenesis.

Following UsnRNP Sm core assembly, the m7G cap of their snRNA is converted to the 2,2,7-tri-methylated guanosine (m3G), and the 3′ extra nucleotides of pre-UsnRNAs are removed. These core UsnRNPs are imported into the nucleus to

be assembled into the spliceosome. This UsnRNP import process requires not only general nuclear import factors such as Importin-β but also specific factors such as Snurportin-1 that recognizes the m3G cap of the UsnRNP and interacts with general factors for snRNP import into the nucleus.

Before the association of UsnRNP particle-specific proteins, the UsnRNAs undergo internal modifications including pseudouridylation and 2′-O-methylation. Such modifications appear to be necessary for the assembly of a functional UsnRNP, as shown for the human U2snRNP. These posttranscriptional modifications are mediated by a small nucleolar (sno)RNA-guided mechanism through the action of snoRNPs. For example, U85 snoRNP directs both 2′-O-methylation and pseudouridylation of U5snRNA. Following these posttranscriptional internal modifications of snRNAs, the assembly of UsnRNPs is completed with the association with individual UsnRNP-specific proteins (Table 1A). Among these UsnRNP-specific proteins, U5-220 kDa (also named Prp8 or PRPC8) is a crucial spliceosomal protein that is most highly conserved through evolution. Prp8 interacts with sequences of all major splicing signals including 5′ss, branch site, and 3′ss. It has been proposed that Prp8 plays a critical role in catalysis by aligning 5′ss, and 3′ss at the catalytic center.

U6snRNP plays an important role at the catalytic center of the spliceosome, and its biogenesis has several unique features. The U6snRNA is transcribed by RNA polymerase III, and its cap structure is a γ-monomethyl group. The U6snRNP does not contain an Sm protein. Instead, U6snRNP contains seven Sm-like proteins (Lsm2, 3, 4, 5, 6, 7, and 8) interacting with the U-rich region at the 3′ end of the U6snRNA. The assembly of U6snRNP is believed to occur in the nucleus. The base-pairing interactions between U4snRNA and U6snRNA lead to the formation of the U4/U6snRNP, which associates with U5snRNP as the U4/U6.U5 tri-snRNP complex to join the spliceosome. The formation of U4/U6.U5 tri-snRNP complex and association of tri-snRNP with the spliceosome require PRPF31 (human homolog of yeast Prp31p) in addition to other proteins.

1.4 Spliceosome Assembly

A large number of studies show that mammalian spliceosomes are assembled on the pre-mRNA splicing substrate in an orderly fashion. A recent study, on the other hand, reported a penta-snRNP particle purified from the yeast extract. The finding of this ''preassembled'' penta-snRNP particle suggests the potential importance of concurrent multisite interactions during spliceosome assembly among spliceosomal components and different intronic as well as exonic regions. Many components of spliceosomes are conserved between yeast and human. In fact, much of our knowledge of spliceosome assembly is based on genetic studies in yeast and on biochemical experiments using both yeast and mammalian systems. The mammalian homologs of key spliceosomal components have been identified (Table 1). Systematic proteomic studies of spliceosomes assembled on model splicing substrates confirm the high degree of conservation between mammalian and yeast spliceosomal proteins (Table 1).

Spliceosome assembly is a highly dynamic process, with multiple RNA–RNA, RNA–protein, and protein–protein interactions. During spliceosomal assembly

and activation, there are a series of conformational rearrangements. This process involves not only changes in the RNA conformation but also remodeling of snRNPs and exchanges of protein factors. Studies over the past two decades have only provided a sketch of this multicomponent interactive process. At the catalytic center of the functionally mature spliceosome, the 5′ss and 3′ss must be juxtaposed precisely to ensure accurate cleavage and ligation.

The classical view of spliceosome assembly features stepwise interactions between the pre-mRNA substrate and different spliceosomal UsnRNPs. Immediately after transcription, nascent pre-mRNA transcripts interact with hnRNP (heterogeneous nuclear ribonucleoprotein) proteins (Table 2) to form *H complex*. Spliceosome assembly is initiated with the recognition of the 5′ splice site by U1snRNP in an ATP-independent manner. Efficient interaction between the U1snRNP and 5′ splice site requires U1snRNP-specific proteins (such as U-70 K) and possibly other proteins. Recognition of the branch site and 3′ splice site requires cooperative binding of SF1 (also named the *branch point binding protein*, BBP) to the branch site and binding of U2 auxiliary factor (U2AF65 and U2AF35 as a

Tab. 2 Human H complex proteins.

Acce. #	*Protein name*	*Motifs*	*Cal. MWt*
Q13151	hnRNP A0	2 RRMs	30 841
P09651	hnRNP A1**	2 RRM	38 715
P22626	hnRNP A2/B1	2 RRMs	37 430
P51991	hnRNP A3	2 RRMs	39 686
P07910	hnRNP C1/C2	1 RRM	33 299
Q14103	hnRNP D0	2 RRMs	38 434
P26599	hnRNP I/PTB**	4 RRMs	57 221
Q07244	hnRNP K	KH	50 976
P14866	hnRNP L	3 RRMs	60 187
O43390	hnRNP R	3 RRMs	70 943
AL031668	hnRNP RALY	1 RRM	32 214
P35637	FUS/hnRNP P2	1 RRM	53 426
B54857	NF-AT 90 k	2 DSRMs, ZF	73 339
AJ271745	NFAR-1	DSRM, ZF	76 033
A54857	NF-AT 45 k	DSRM, ZF	44 697
AF037448	GRY-RBP	3 RRMs	69 633
P43243	Matrin3		94 623
O43684	hBUB3	WD40s	37 155
Q15717	HuR	3 RRMs	36 062
Q92804	TAFII68	1 RRM, ZF	61 830
P16991	YB1**	COLD	35 924
P16989	DBPA	COLD	40 060
P08107	HSP70	HSP70	70 052
P11142	HSP71	HSP70	70 898
P11021	GRP78	HSP70	72 116

The H complex proteins that are involved in alternative splicing regulation are marked with "**."

heterodimer) to the intronic sequence at the 3′ splice site. Splicing reactions can occur in the absence of detectable U1snRNP. However, efficient splicing of most introns appears to require U1snRNP. The formation of such an early complex (*E complex*) is considered to be the commitment step that directs the nascent pre-mRNA transcript into the splicing pathway.

A distinct feature of the formation of mammalian splicing commitment complex is the involvement of a family of proteins named SR proteins that are not found in the genome of *S. cerevisiae*. SR proteins share common structural features, with one or two RNA recognition motifs (RRMs) of the RNP-consensus (RNP-cs) type at the amino-terminus and a carboxyl domain rich in arginine and serine residues (RS domain, Table 1). These proteins play important roles in recognizing exons and mediating interactions between splice sites. SR proteins interact with exonic sequences and recruit splicing factors for the formation of "cross-exon" complex, (i.e. exon definition). SR proteins also interact with 5′ss and facilitate the "cross-intron" recognition by mediating protein–protein interactions between U1snRNP and factors associated with the branch site 3′ss including U2snRNP, U2AF, and possibly other proteins.

The binding of U1snRNP and SR proteins promotes the interaction of the 17S U2snRNP with the branch site sequence. The stable association of U2snRNP with the branch site to form the *A complex* is ATP-dependent and requires the U2snRNP-specific proteins SF3a [SF3a60/SAP61 (Prp9), SF3a66/SAP62 (Prp11) and SF3a120/SAP114 (Prp21)]. This process also requires non-snRNP splicing factors including hPrp5 (a putative ATP-dependent RNA helicase). Upon the integration of the U2snRNP complex, a U2snRNP-protein, p14, contacts the branch adenosine residue and interacts with other U2snRNP proteins. The next ATP-dependent step is the association of U4/U6.U5 tri-snRNP complex to form the *B1 complex*, in which the interaction of the 5′ss with U1snRNA is destabilized by U5-100 kDa (hPrp28), a putative RNA helicase. The 5′ss sequence is then engaged in interactions with U6snRNA around the intronic region and with U5snRNA at the exonic region. After escorting the U6snRNP into the B1 complex, the U4snRNP is released to form the *B2 complex*, a transition that requires ATP and possibly U5-200 kDa (human homolog of Brr2 in yeast, another putative RNA helicase). The next ATP-dependent transition is the formation of the *C1 complex*, a process involving the activity of yet another possible RNA helicase hPrp2 (human homolog of Prp2/Ynr011c). This may directly lead to the formation of the catalytic site for the *first step of splicing*, the cleavage at the 5′ss with the formation of a lariat intermediate (see Fig. 4). U2, U6, and U5snRNPs are associated with the catalytically active form of splicing complex in which the first step of splicing occurs. The *second step of splicing* (cleavage at the 3′ss and ligation of exons) requires an additional set of protein factors including hPrp16, hPrp17, hPrp18, and hSlu7 (Table 1). The human homolog of the yeast Prp16, hPrp16, contains an RS domain in addition to DEXD-box helicase and ATPase domains. The transition to form the *C2 complex* is again ATP-dependent. The catalytic center for the second step is also formed by U6snRNA, U2snRNA, and/or associated proteins. The splicing products, ligated exons, and the lariat intron, are then released from the spliceosome by

another DEXD-box containing RNA helicase together with other factors. The release of the lariat intron and dissociation of U2, U5, and U6snRNPs from the *I complex* requires another putative RNA helicase, hPrp43. The released U6snRNP is reannealed with U4snRNP and associated with U5snRNP to form the U4/U6.U5 tri-snRNP complex to enter another cycle of spliceosome assembly. The lariat intron is debranched by the enzyme Dbr1. Many spliceosomal components are presumably recycled to form new spliceosomes.

It is clear that spliceosome assembly is a highly dynamic process involving multiple networks of RNA–RNA, protein–RNA, and protein–protein interactions at each step. Our understanding of this complex process, as well as its regulation, remains limited.

1.5 Biochemical Mechanisms of pre-mRNA Splicing

The complex and dynamic process of spliceosome assembly culminates in the formation of the catalytic core in which 5′ss and 3′ss are precisely juxtaposed. The biochemical reactions of pre-mRNA splicing involve two transesterification steps (Fig. 4). The first step is the cleavage at the 5′ss and the formation of the lariat intermediate with a 2′-OH displacing a 3′-OH group. The second step is the cleavage at the 3′ splice site with concomitant ligation of the 5′ and 3′ exons in which a 3′-OH displaces another 3′-OH group. Although it remains controversial, these two steps of splicing reactions may take place in two catalytic sites. In addition to the chemical differences between the

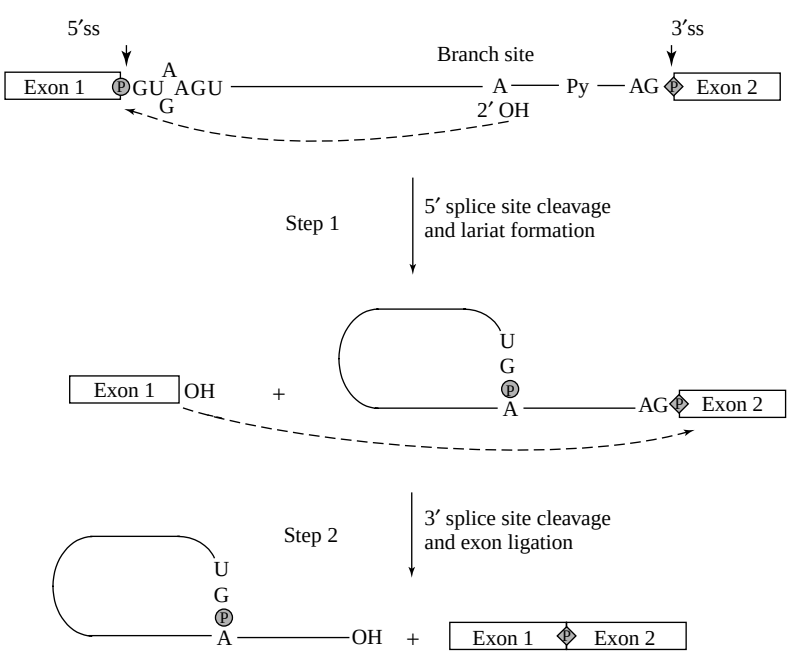

Fig. 4 The biochemical mechanisms of pre-mRNA splicing: two steps of transesterification. The phosphodiester linkages are indicated by the letter "p" inside a circle or a diamond.

two reaction steps, the stereoselectivity of the transesterification reactions and the difference in the metal-ion dependence support the existence of two reaction sites.

Accumulating evidence suggests that pre-mRNA splicing may be RNA-catalyzed with U6snRNA at the center for catalysis and protein factors including PRP8 as components of the catalytic core. U6snRNA plays a critical role in catalysis and in interacting with pre-mRNA and other spliceosomal snRNAs. In yeast, U6snRNA contributes to spliceosomal catalysis by coordinating catalytic metals. Precise catalytic mechanisms of mammalian pre-mRNA splicing remain to be elucidated.

2 Alternative pre-mRNA Splicing

2.1 Alternative Splicing and its Role in Regulating Gene Activities and Generating Genetic Diversity

The vast majority of vertebrate genes contains at least one intron. It is estimated that more than 50% of human genes undergo *alternative splicing*, the process that generates distinct splicing products from the same pre-mRNA transcript by using different splice sites (Fig. 5). Alternative splicing is an important mechanism for regulating gene activities in eukaryotic species. Alternative splicing events can

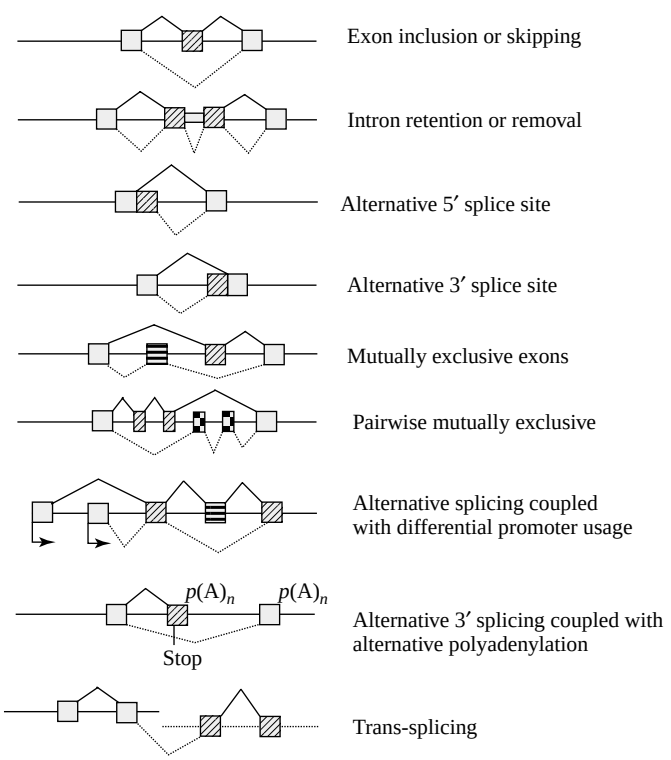

Fig. 5 Diverse patterns of alternative splicing. Alternatively spliced regions or exons are illustrated as shaded boxes.

occur in the protein-coding regions or in the regulatory regions of genes, including both 5′ and 3′ untranslated regions. Alternative splicing can affect peptide coding capacity or influence mRNA stability or translational control of the transcription products. Alternative splicing regulation can also be coupled with transcription, polyadenylation, RNA editing, or mRNA exporting processes. A recent example for the coupling of splicing with other processes of gene regulation is that steroid hormone receptors can simultaneously regulate transcription and alternative splicing by recruiting coregulators involved in both processes. In addition, alternative splicing can affect posttranslational modifications of protein products.

2.1.1 Different Patterns of Alternative Splicing

A number of distinct alternative splicing patterns have been reported (Fig. 5). Most common alternative splicing events include exon inclusion/skipping, intron removal/retention, and alternative selection of competing 5′ or 3′ splice sites. More complex patterns of alternative splicing include mutually exclusive or cassette types of exon inclusion. Alternative selection of terminal exons can be coupled with differential promoter usage or polyadenylation. Furthermore, recent studies have documented possible alternative trans-splicing of mammalian genes, although it may occur only at a low frequency in mammals.

2.1.2 Alternative Splicing and Genetic Diversity

Alternative splicing regulates gene activities involved in every aspect of cell survival and function. It is a major mechanism for generating the complexity of mammalian proteomes. Alternative splicing contributes to proteome expansion by a number of mechanisms, such as the usage of distinct translation start sites, in-frame nucleotide deletion or insertion, changes in peptide sequence, and alternative usage of different translation stop codons. Such changes in peptide sequence or length may lead to the formation of proteins with distinct properties, including biochemical/biophysical characteristics, subcellular localization (secreted versus membrane associated, membrane-tethered versus cytoplasmic, cytoplasmic versus nuclear), posttranslational modifications (glycosylation, phosphorylation, or lipid modification), or interactions with other cellular components.

Alternative splicing can be an excellent mechanism for generating functionally antagonistic products from the same genetic locus and for the fine-tuning of gene activities at the posttranscriptional level. For example, alternative splicing of a number of genes critical for cell death leads to the formation of both cell death–promoting and cell death–preventing splicing isoforms. These include genes encoding for death ligands, death receptors, Bcl-2 superfamily of death regulators, caspases, and other cell-death regulatory genes. Several human caspase genes utilize alternative splicing to produce protein products that either contain or lack their enzyme active sites, resulting in antagonistic activities in cell death.

The nervous system is a good example where alternative splicing is utilized to generate extreme functional diversity. A vast number of genes involved in neural development and function undergo complex alternative splicing. Some genes encoding neural receptors and axon guidance molecules can generate hundreds to

even thousands of different splicing isoforms. Many ion channels and neurotransmitter receptors have different alternative splicing isoforms with distinct electrophysiological properties. Alternative splicing regulation has a significant impact on the proper function of the nervous system, including learning, memory, and behavior development.

2.2 Mechanisms Underlying Alternative Splicing Regulation

The fundamental mechanisms underlying alternative splicing are the intricate interactions among trans-acting splicing factors and cis-elements in the pre-mRNA substrates, leading to selection of different splice sites. These cis-elements include splice sites and splicing regulatory sequences (splicing enhancers or silencers). The sizes of the exons or introns, as well as secondary structures of the pre-mRNA also influence splice site selection. The highly degenerate nature of mammalian splicing signals and the existence of large numbers of trans-acting factors allow versatile RNA–protein and RNA–RNA interactions during different stages of spliceosome assembly. Such interactions can be modulated by both exonic and intronic regulatory elements. The recognition of splice sites and cis-acting splicing regulatory sequences is mediated by a network of interactions between pre-mRNA and trans-acting factors including snRNPs and non-snRNP splicing factors. A number of splicing regulators, both positive and negative, play important roles in alternative splicing regulation (Table 3). The final outcome of alternative splicing of a given gene, including the production and the delicate balance of distinct isoforms,

Tab. 3 Pre-mRNA splicing regulators.

Protein name	*Motifs*	*RNA-binding sites*
A. Splicing activators		
SRp75	2 RRMs, RS	
SRp55	2 RRMs, RS	ESEs
SRp40	2 RRMs, RS	
ASF/SF2	2 RRMs, RS	
9G8	1 RRM, RS	
SC35	1 RRM, RS	
SRp30c	2 RRMs, RS	
hTra2	1 RRM, RS	
SRp20	1 RRM, RS	
SRm300	RS-rich	
SRm160	PWI, RS	
SF1 (BBP)	ZF, KH, CCHC,	ISE
KSRP	KH	
NOVA1	3 KH	(UCAUY)3
rSLM-2	KH, STAR	
CUG-BP	RRMs	ISE
ETR3	3 RRMs,	ISE
Halfpint (D.m.)	RRM (homolog of PUF60)	

Tab. 3 (*continued*)

Protein name	*Motifs*	*RNA-binding sites*
TIA-1	3 RRMs	U-rich
p72	DExD, Helicase, RGG	
YB1	Cold box	A/C-rich ESE
B. Splicing repressors		
hnRNP A1	2 RRMs, Gly-rich	ESSs or ISSs
hnRNP I/PTB	4 RRMs	U/C-rich
nPTB	4 RRMs	U/C-rich
SmPTB	4 RRMs	U/C-rich
Elav (D.m.)	3 RRMs	
Sxl (D.m.)	3 RRMs	
Mec8 (C.e.)	2 RRMs	
Fox1 (C.e.)	RRM	
PSI (D.m.)	4 KH	ISS
QKI-5	KH	ISS
SRp30c	2 RRMs, RS	
SRrp35	1 RRM, RS	
SRp38	1 RRM, RS	
SRrp40	1 RRM, RS	
SPF45	1 RRM, Gly-rich	
RSF1 (D.m.)	1 RRM, GRS	
SWAP (D.m.)	RS	
C. Bifunctional splicing regulators		
p54/SFRS11	1RRM, RS	
SRrP86	1RRM, RS	
p32		
HnRNP H	3RRMs, Gly	ESE or ISE ESS
Napor1	3RRMs	

Note: A. Splicing activators. B. Splicing repressors. C. Bifunctional splicing regulators.
STAR: signal transduction and activation of RNA.
For other abbreviations, see footnotes for Table 1.

is determined by combinatorial effects of multisite interactions among pre-mRNA, essential spliceosomal components, and regulatory factors.

Genetic and biochemical studies in *Drosophila* genes have tremendously advanced our understanding of alternative splicing regulatory mechanisms, although some mechanisms may not be conserved in corresponding mammalian systems. For example, *Drosophila* sex determination genes are under extensive regulation by alternative splicing, and this regulatory mechanism is not conserved in mammalian sex determination. In general, *Drosophila* genes contain more small introns than mammalian genes. Here, we focus on mammalian alternative splicing regulation.

2.2.1 Splicing Signals and Splicing Regulatory Elements

Unlike the yeast spliceosome, mammalian spliceosomes have the daunting task

of searching for small exons in the vast sea of introns. This process of "exon recognition" or "exon definition" is particularly remarkable because sequences at mammalian splice sites are so highly degenerate (see Fig. 2). A large number of sequence elements with similarity to authentic splice sites (pseudosplice sites) can be found in both intronic and exonic regions). Further complicating the issue, some of these elements act as *cryptic splice sites* that are only used by the splicing machinery when the authentic splice sites are altered by mutations. Therefore, specific recognition of authentic splice sites and correct pairing of corresponding 5′ and 3′ splice sites is a central issue for both constitutive splicing and alternative splicing regulation. The intrinsic sequence degeneracy of mammalian splice sites determines that alternative splicing is a rule rather than an exception during mammalian gene expression.

In addition to splice sites, sequence elements in both intronic and exonic regions modulate alternative splicing. Such regulatory elements can either enhance or suppress splicing, and are hence named exonic splicing enhancers (ESEs) and intronic splicing enhancers (ISEs), or exonic splicing silencers (ESSs) and intronic splicing silencers (ISSs).

A number of ESE motifs have been identified using biochemical systematic evolution of ligands by exponential enrichment (SELEX) or bioinformatical approaches. A/G-rich (also called purine-rich) and A/C-rich elements are among ESE motifs characterized by biochemical studies. Proteins containing SR domains play a major role in recognizing A/G-rich ESEs and recruiting other spliceosomal components (including snRNPs and other protein factors), thereby promoting the usage of neighboring splice sites. A cold-box protein, YB-1, has been shown to enhance splicing by interacting with an A/C-rich ESE. Another ESE in HIV-1 tev-specific exon interacts with hnNRP H and SR protein SC35 to enhance splicing. The SELEX method has been used to identify preferred binding sequences for individual SR proteins, and optimal binding sites for individual SR proteins are degenerate. ESE prediction programs based on SELEX and computational analysis of human genes have been developed (http://exon.cshl.org/ESE). These programs are useful in predicting alternative splicing patterns of natural pre-mRNA substrates in cells.

ISE elements have been studied in a number of genes, including c-src, β-tropomyosin, calcitonin/calcitonin gene-related peptide gene (CGRP), fibronectin, nonmuscle myosin heavy chain, cardiac troponin T (cTNT), FGFR-2, $\alpha 2$ subunit of glycine receptor, and other genes. Such ISEs may contain sequences similar to 5′ splice site, U-rich element adjacent to the regulated 5′ splice site, UGCAUG element, (UCAUY)3-containing sequences, or CUG-containing motif.

A number of ESSs have been characterized in different genes such as β-tropomyosin, CD44, and viral genes including HIV Tat, bovine papillomavirus type-1, and Rous sarcoma virus. They do not share any obvious sequence motifs. Their activities are often associated with interactions with proteins of the hnRNP family, including hnRNPA1, hnRNP H, and hnRNP F.

A variety of ISSs have been analyzed in alternatively spliced genes including hnRNP A1, fibroblast growth factor receptor 2 (FGFR2), caspase 2, $GABA_A R\gamma 2$ (γ-aminobutyric acid receptor typeA γ2subunit), NMDA R1 (*N*-methyl-D-aspartate receptor R1 subunit;), clathrin light chain B, and HIV tat. Many of

these ISSs contain extended polypyrimidine tracts. Some splicing repressor elements contain sequences similar to authentic splice sites, or decoy splice sites. It has been proposed that such decoy splice sites mediate nonproductive interactions to suppress the usage of upstream 5′ splice site. Again, several hnRNP proteins, including polypyrimidine tract binding protein (PTB, also known as hnRNP I), hnRNP A1, and hnRNP H play important roles in splicing repression by ISSs.

One of the common features of these cis-regulatory elements is that they are not simple sequence elements that act independently of each other. Sequence elements have been identified that can act to stimulate the splicing of one exon but repress another exon (e.g. in FGFR2 gene). Some regulatory elements contain both enhancer and silencer domains. The splicing of IgM exons M1 and M2 is regulated by a sequence containing juxtaposed splicing enhancer and silencer elements. Alternative splicing of protein 4.1R pre-mRNA generates multiple isoforms. This alternative splicing event is critical for red blood cell membrane biogenesis during erythroid differentiation and is regulated by multiple cis-elements and trans-acting regulators.

Splicing regulatory elements usually contain multiple binding sites for splicing regulators and function by recruiting other spliceosomal components to form RNP-like complexes. An evolutionarily conserved 100 bp intronic suppressor element in the caspase 2 (casp-2) gene, In100, specifically inhibits its exon 9 splicing. Alternative splicing of this 61 bp exon 9 leads to the formation of two functionally antagonistic products, casp-2L and casp-2S. Casp-2L product promotes cell death, whereas casp-2S prevents cell death. The In100 element contains a decoy 3′ splice site juxtaposed to a PTB-binding domain, both of which contribute to the full activity of In100 in inhibiting exon 9 inclusion. The upstream portion of In100 contains a sequence with features of an authentic 3′ splice site (including a branch site, a polypyrimidine tract, and AG dinucleotide). However, this site is not used under normal conditions. This sequence is only recognized as a 3′ss when the site is isolated with the downstream PTB-binding domain deleted. Biochemical and cell culture experiments show that this decoy 3′ acceptor site interacts nonproductively with the 5′ splice site of the alternative exon 9, thus repressing the efficient use of the 5′ splice site of exon 9 despite a high level of U1snRNP binding to this 5′ splice site. Downstream of the decoy 3′ splice site resides the second functional domain that interacts with PTB. The binding of PTB to CU-rich motifs within this downstream domain juxtaposed to the decoy 3′ acceptor site correlates well with the repressor activity of this domain. PTB can modulate recognition of the adjacent decoy 3′ acceptor site. In addition to factors interacting with an authentic 3′ splice site (such as U2AF or U2snRNP), PTB as well as other proteins interact with In100 and contribute to recognition of the In100 decoy 3′ splice site by the spliceosome as an intronic repressor element, rather than as an authentic 3′ splice site. The regulatory role of PTB in casp-2 alternative splicing and its mechanism of action appear to be distinct from other systems (Sect. 2.2.2). A recent survey of known human genes involved in cell death regulation suggests that In100-like intronic elements (i.e. 3′ splice site juxtaposed to PTB-binding domains) may represent a general intronic splicing repressor motif. Such intronic elements

may play a role in regulating alternative splicing of other cell-death genes.

Another example of the complexity of splicing regulatory elements is in the alternative splicing of exon 10 in the human *tau* gene. Alternative splicing of this exon is associated with the pathogenesis of dementias (Sect. 3.1). In this case, both exonic and intronic regulatory elements play important roles in controlling exon 10 inclusion. In addition, in the exonic region, both positive and negative elements are involved and form a multidomain composite regulatory element.

2.2.2 Trans-acting Splicing Regulators

The recognition and selection of splice sites are determined during spliceosome assembly, especially at early steps of spliceosomal formation. Splicing activators facilitate interaction of U1snRNP with 5′ splice site and of U2snRNP with 3′ splice site, whereas splicing repressors suppress the recognition of splice sites. A number of proteins involved in spliceosome assembly also play important roles in regulating alternative splicing (see Table 1).

Several families of splicing activators have been reported (Table 3). SR proteins are among the best characterized splicing activators. The interactions between SR proteins and ESEs play a critical role in exon recognition by the spliceosome. ESEs are present in both constitutively and alternatively spliced exons. By mediating protein–protein and RNA–protein interactions during early steps of spliceosome assembly, SR proteins coordinate the communication between 5′ and 3′ splice sites and promote interactions between exonic enhancers and splice sites. Enhancer complexes are usually multicomponent complexes, containing not only SR proteins but also other RS domain–containing proteins such as U2AF. Some RS domain–containing proteins, such as SRm300 and SRm160, act as coactivators for ESE function. The RS domain in SR proteins can be differentially phosphorylated, providing another level of regulation. The phosphorylation status of SR proteins regulates protein–protein interactions, intracellular distribution, and activities of SR proteins. Differential phosphorylation of SR proteins has been shown to play a role in regulating gene expression during development.

SF1, KSRP, NOVA-1, and rSLM-2 are RNA-binding proteins containing heterogeneous nuclear ribonucleoprotein K-type homology (KH) domain. They can enhance splicing by interacting with ISEs. SF1, a protein important for branch site recognition during spliceosome assembly, binds to GGGGCUG repeats in an ISE to activate the recognition of a 6-bp micro exon in cTNT gene. KSRP interacts with UGCAUG sequence and stimulates the neuronal-specific exon inclusion in c-src. NOVA-1 enhances the splicing of exon E3A in the $\alpha 2$ subunit of the glycine receptor gene (GlyR$\alpha 2$). The rat Sam68-like mammalian protein (rSLM-2) is a member of the STAR (signal transduction and activation of RNA) protein family. It can influence the splicing pattern of the CD44v5, human transformer-2beta, and tau minigenes in transfected cells.

Several members of CUG-BP and ETR-like factors (CELF) family activate the splicing of genes including cTNT, muscle-specific chloride channel, insulin receptor, and NMDA R1. Some of these CELF proteins activate splicing by interacting with ISEs.

TIA-1, a mammalian homolog of yeast NAM8 protein, interacts with the U-rich intronic sequence adjacent to the 5′ splice site of the K-SAM alternative exon in the FGFR2 gene. The activation of this splice

site is U1snRNP-dependent, and TIA-1 may function by facilitating U1-snRNP binding to the 5′ splice site.

Human Y-box binding protein (YB-1) was initially identified as a transcription factor interacting with single-stranded DNA in response to cold shock. YB-1 contains a five-stranded ß-barrel known as the "COLD" domain and accessory domains rich in basic amino acids and aromatic groups. This domain arrangement in YB-1 protein with specific interaction domain containing ß-sheets and a basic domain providing generic RNA binding is reminiscent of the modular structure of the SR proteins. YB-1 stimulates the splicing of the human CD44 alternative exon v4 by interacting with the A/C-rich element in the exonic splicing enhancer. Another protein, DEAD-box RNA helicase p72, has been reported to affect the splicing of alternative exons containing AC-rich exon enhancer elements. The mechanism of YB-1 or other proteins interacting with the A/C-rich type of ESEs in splicing activation remains to be elucidated.

HnRNP proteins hnRNPA1 A1 and PTB are among the best characterized splicing repressors. HnRNP A1 can interact with either ESSs or ISSs to prevent exon recognition by SR proteins or U2 assembly. A model has been proposed for interactions between hnRNP A1 and ESS based on studies on HIV Tat exon 3 splicing. In this model, the high-affinity binding of hnRNPA1 to ESS promotes nucleation of multiple A1 along the exon. The formation of this inhibition zone can be blocked by SF2/ASF, but not by another SR protein SC35, providing an explanation for differential antagonism between hnRNP A1 and different SR proteins. Two different models have been proposed for the function of hnRNPA1 in splicing silencing mediated by ISSs. In the case of HIV Tat intron 2 splicing, the interaction of hnRNP A1 with an alternative branch point sequence blocks the recognition of the branch site by U2snRNP. In the second model, the cooperative binding of hnRNP A1 to two intronic elements flanking the alternative exon inhibits exon inclusion by a looping-out mechanism during autoregulation of hnRNP A1 gene splicing.

PTB interacts with U/C-rich elements in ISSs in a number of genes including n-src, FGFR2, $GABA_AR\gamma 2$, tropomyosin, NMDA R1 exon 5, clathrin light chain B, caspase-2, and calcitonin/CGRP genes. Depletion of PTB using an RNA interference approach demonstrates that PTB is a negative regulator of exon definition in cultured cells. PTB-binding sites are frequently located in the intronic regions upstream of the regulated 3′ splice sites, although functionally active PTB-binding sites are also found in intronic elements downstream of the alternatively spliced exons (such as in c-src and caspase-2 genes). Mechanisms by which PTB represses splicing are not clear yet. One model for repression by PTB is via competition with U2AF binding to the polypyrimidine tract to block early spliceosome formation. This model is based on studies of alternative splicing of $GABA_AR\gamma 2$, NMDA R1 exon 5, and clathrin light chain B genes. In these genes, the high-affinity PTB binding to the long polypyrimidine tracts immediately upstream of the neural-specific exons represses the inclusion of these exons in nonneural tissues. A similar repression mechanism may be used in suppressing the inclusion of muscle-specific exons in rat α- or β-tropomyosin. PTB exists in a range of different tissues as an abundant splicing repressor. The repressor activity of PTB may be modulated by other regulatory proteins. Two tissue-specific PTB-related proteins have

been found, neuron-enriched (nPTB,) and smooth muscle–enriched (SmPTB). For example, nPTB can compete with PTB to promote inclusion of neuronal-specific N1 exon inclusion in c-src. The expression of SmPTB correlates with the smooth muscle–specific suppression of α-tropomyosin exon 3, which is included in nonsmooth muscle cells. The release of PTB suppression can be also achieved by cell-type specific CELF proteins such as ETR3 and Napor-1.

One exception to the general inhibitory activity of PTB has been reported where PTB stimulates the inclusion of an alternative 3′-terminal exon. In this case, the splicing regulation is coupled with alternative polyadenylation. In addition, PTB has also been implicated in translational control of viral transcripts. More comprehensive understanding of the biological roles of PTB in gene regulation requires further investigation.

Several proteins containing an RS domain and RRM-cs domains have also been reported to act as splicing repressors, including SRrp30c, SRrp35, SRp38, and SRrp40 (see Table 3). SR proteins are generally hyperphosphorylated by the kinase SRPK1. Dephosphorylated SRp38 is required for the splicing repression in mitotic cells. Another RS domain-containing protein, the *Drosophila* suppressor-of-white-apricot protein (SWAP) suppresses its own pre-mRNA splicing.

An RRM-containing spliceosomal protein, SPF45, represents a late-acting splicing regulator. In *Drosophila*, SPF45 blocks splicing at the second step by interacting with Sex-lethal (Sxl) protein, indicating that 3′ splice site recognition and splicing regulation can occur at the second catalytic step.

Proteins containing KH domains can also act as splicing repressors. For example, a nuclear isoform of quaking (qk) protein, QKI-5, regulates alternative splicing of myelin-associated glycoprotein gene by interacting with an intronic splicing repressor element. A *Drosophila* splicing repressor, P-element somatic inhibitor (PSI), is required for the soma-specific inhibition of splicing of P-element pre-mRNA both *in vitro* and *in vivo*.

Several splicing regulators can act as either positive or negative splicing regulators, depending on the sequence context of the pre-mRNA substrates. We classify these splicing regulators as bifunctional splicing regulators. Some SR proteins or SR-domain containing proteins can repress splicing of certain pre-mRNAs but activate splicing of other substrates. SRrp86 is an 86-kDa related to SR proteins. It can function as either an activator or a repressor by regulating the activity of other SR proteins. Other proteins can repress splicing by interacting with SR proteins. For example, an ASF/SF2-interacting protein, p32, was shown to act as a splicing regulator by inactivating the function of ASF/SF2.

HnRNP H and the related protein hnRNP F contain three RRMs of RNP-cs type. HnRNP H interacts with G-rich elements in either splicing enhancers or silencers. When binding to a splicing enhancer in the HIV env gene, hnRNP H acts as an activator by promoting the assembly of a complex containing SC35 and U1snRNP. When interacting with ESSs in the β-tropomyosin pre-mRNA or HIV Tat exon 2, or with a negative regulator of splicing (NRS) in the Rous Sarcoma Virus (RSV) genome, hnRNP H serves as a splicing repressor.

Napor1, a splice variant of ETR3, is expressed at a high level in the forebrain but at a low level in the cerebellum. Overexpression of Napor1 exerts opposite

effects on two different exons in NMDAR1, inhibiting exon 5 inclusion but stimulating exon 21 inclusion.

With more alternative splicing events characterized in detail, more cases of bifunctional splicing regulators may emerge. It is conceivable that these splicing regulators interact with different sequence elements and different spliceosomal components. When these interactions promote the productive recognition of splicing signals by the splicing machinery, such proteins act as splicing activators. On the other hand, when the same proteins compete with other splice activators or facilitate recognition of splicing silencers leading to nonproductive interactions between the spliceosome and splicing substrates, these proteins behave as splicing repressors.

2.3 Tissue-specific and Developmentally Regulated Alternative Splicing

Tissue-specific alternative splicing has been studied in mammalian systems for more than a decade. However, only a limited number of tissue-specific splicing regulators have been identified so far. These include Nova-1, Nova-2, nPTB, and SmPTB. Cell-type specific splicing regulators have also been found in *Drosophila* and *Caenorhabditis elegans*, such as neuronal-specific Elav (embryonic lethal and abnormal vision), ovary-specific Halfpint, and muscle-specific Mec8. From a large number of biochemical, molecular, and genetic studies, the emerging general theme is that no single factor dictates the tissue specificity of alternative splicing of any genes. Instead, multicomponent interactions among tissue-specific splicing regulators, pre-mRNA, and general splicing factors determine the specific alternative splicing pattern of a given gene in a tissue- or cell-type specific manner. On the other hand, genetic deletion of a single splicing regulator, even those highly tissue-specific ones, leads to defects in splicing of multiple target genes. For example, Nova-1 deletion in mice affects alternative splicing of $GABA_AR\gamma2$, $GlyR\alpha2$ and perhaps other unknown target genes. Similarly, the *Drosophila* neural-specific splicing regulator Elav has at least three known target genes, neuroglian (nrg), erectwing, and armadillo.

In addition to tissue-specific expression of splicing regulators, a number of other mechanisms modulate tissue-specific alternative splicing. Tissue-specific combination of different splicing regulators, relative concentrations of distinct splicing factors, and differential modification of splicing regulators all contribute to the tissue-specific alternative splicing of individual genes in a given tissue or cell type. For example, the level of the splicing repressor PTB in neural tissues is lower than in other tissues, providing an explanation for a more permissive environment for exon inclusion in a number of genes in neural tissues. Although most SR proteins are expressed in a wide range of tissues, their expression patterns vary in different tissues, both in different isoforms and at different levels. Many SR proteins have different isoforms either because of alternative splicing or posttranslational modifications. Kinases or phosphatases that regulate phosphorylation of different splicing factors can also be tissue specific. One example is the SR protein–specific kinase 2 (SRPK2), which is highly expressed in the brain and presumably capable of regulating activities of target SR proteins.

It is not clear yet how the extremely complex alternative splicing pattern of different genes in a given cell or a specific tissue is coordinated. In some cases, the alternative

splicing events of different genes appear to be coregulated by the same protein. For example, the splicing regulator PTB can differentially recognize neural and non-neural substrates. The expression pattern of PTB in different regions of the brain at different developmental stages supports a role for PTB to act as an alternative splicing coordinator for different splicing target genes.

2.4 Regulation of Alternative Splicing in Response to Extracellular Stimuli

Extracellular signals can induce changes in alternative splicing of different genes. For example, growth factors or hormones stimulate alternative splicing of intracellular responsive genes. Treatment with growth factors induces changes in alternative splicing of phosphotyrosine phosphatase PTP-1B gene. The alternative splicing pattern of protein kinase C beta is changed by insulin. Activation of calmodulin-dependent kinases (CaM kinases) stimulates changes in alternative splicing of BK potassium channels. In most cases, however, signal transduction pathways involved in splicing regulation have not been characterized.

A number of genes important for the functioning of the nervous system show activity-dependent changes in their splicing. Alternative splicing of syntaxin 3 changes in response to induction of long-term potentiation. NMDAR1 alternative splicing is modulated by both pH and Ca^{2+}. Stress hormones regulate alternative splicing of potassium channels. A cis-acting sequence named *calcium-responsive RNA element* (CaRRE) in the stress-inducible exon of BK potassium channels and in NMDAR1 exon 5 has been identified. This element mediates CaM kinase-dependent repression of the inclusion of the CaRRE-containing exons. The mechanisms by which the CaRRE causes exon skipping remain to be elucidated.

Drastic changes in the cell growth environment are expected to affect RNA metabolism, including alternative splicing. For example, ischemia in mice induces changes in alternative splicing of several genes examined. Accompanying such changes in alternative splicing are changes in the intracellular distribution of several splicing regulators. It is possible that these splicing regulators undergo changes in posttranslational modifications or at other levels induced by ischemia.

Chemical compounds, when administrated to animals or applied to cells in culture, can also cause changes in alternative splicings. Sodium butyrate has been tested in transgenic mice and shown to increase exon 7 inclusion of the SMN gene. Aclarubicin treatment induces changes in SMN gene alternative splicing in cultured cells. Treatment of A549 lung adenocarcinoma cells with cell-permeable ceramide, D-e-C(6) ceramide, downregulates the levels of Bcl-xL and caspase 9b splicing isoforms. The functional responses of nuclear splicing machinery to drug treatment opens the possibility of correcting aberrant or defective splicing using therapeutic agents (Sect. 4.2).

3 Pre-mRNA Splicing and Human Diseases

Aberrant pre-mRNA splicing has been implicated in the pathogenesis of a number of human diseases. Alterations, or dysregulation of either constitutive pre-mRNA splicing or alternative splicing can

cause disease phenotypes. Molecular genetic studies of human diseases have revealed a wide range of mutations that cause diseases because of their effects on pre-mRNA splicing. Investigating the molecular nature of these genetic defects and pathogenetic mechanisms will facilitate both diagnosis of such diseases and development of new therapeutic approaches.

3.1 Splicing Defects in Human Diseases

Genetic defects that cause aberrant pre-mRNA splicing can be found in every category of disease, from malignancies affecting a certain tissue or organ to diseases or syndromes involving multiple systems. Splicing mutations have been identified in genes involved in the pathogenesis of diseases in every system. Detailed information about these mutations can be found in databanks including HGMD (the Human Gene Mutation Database, http://www.uwcm.ac.uk/uwcm/mg/) as well as in a large number of publications. As shown in Table 4, we use a few examples to illustrate the diverse range of disease phenotypes and genes involved. The current understanding of these splicing defects will be summarized.

Malignancies are a major cause of mortality. Splicing mutations in genes critical for either cell proliferation or cell death result in malignancies in different systems. These include oncogenes, tumor suppressor genes, and genes involved in cell death. Only a few examples are included here to demonstrate the complexity of the involvement of defective/aberrant splicing in tumorigenesis. Cancer of the same tissue or organ can be caused by splicing mutations in different genes. For example, splicing defects in p53, p51, CD44, and members of epidermal growth factor receptor (EGFR) family lead to lung cancer, the most common cancer in humans. Breast cancer has been associated with splicing mutations in a number of genes, including HER-2/neu, p53, mdm2, and BRCA1 or BRCA2. Glioblastomas are also associated with splicing mutations in different genes. On the other hand, splicing defects in a single gene can lead to tumorigenesis in different tissues. For example, CD44 aberrant splicing has been associated with the development of a range of different tumors, such as breast cancer, prostate cancer, lung cancer, and Wilm's tumor. CD44 is a polymorphic family of cell surface glycoproteins important for cell adhesion and migration. Aberrant splicing of CD44 has been associated with the invasive behavior and metastasis of several types of tumors. Alternative CD44 splicing variants have been correlated with the poor prognosis of tumors. Another gene is Mdm2, whose alternative splicing defects have been implicated in oncogenesis in multiple tissues. Oncogenic splicing variants of Mdm2 that lack a domain important for its function in interacting with the tumor suppressor p53 have been found in breast cancer, glioblastoma, and rhabdomyosarcoma. Alterations in the balance of different FGFR2 splicing isoforms have been correlated with progression of prostate cancer. A significant fraction of neurofibromatosis type 1 (NF1) cases are associated with an aberrantly spliced NF1 gene. Production of tumor antigens as a result of alternative splicing has been associated with neuroblastoma and other tumors. The role of these tumor antigens in the development and metastasis of these tumors remains to be investigated. In pediatric acute myeloid leukemia (AML), splicing isoforms of gene fusion transcripts produced as a result of

Tab. 4 Examples of human diseases associated with aberrant or defective pre-mRNA splicing.

	Diseases	***Genes***	***Aberrant splicing***
Malignancies			
	Lung cancer	p53, p51, CD44, EGFR	Aberrant or defective splicing products
	Cancer in different tissues	p53, CD44	Defective or aberrant splicing isoforms
	Wilm's tumor	CD44	Aberrant splicing isoforms
	Breast cancer	HER2/neu	Aberrant splicing and imbalance of splicing isofors
	Breast and ovarian cancer	BRCA1, BRCA2	Exon skipping
	Breast cancer, rhabdomyosarcoma	Mdm2	Oncogenic variants lacking p53-binding domain
	Glioblastomas	Mdm2	Oncogenic variants lacking p53-binding domain
	Neurofibromatosis type 1	NF-1	Exon skipping
	Neuroblastoma	HUD, HUC, NNP-1, α-internexin	Production of tumor antigens
	AML	MLL-SEPTIN6	MLL-SEPTIN6 fusion-splicing variants
Cardiovascular diseases			
	Hypercholesterolemia	LDLR	Cryptic ss usage, exon skipping
	Hypertriglyceridemia	Hepatic lipase	Cryptic ss usage
	Marfan syndrome	Fibrillin-1	Cryptic ss usage, exon skipping
	Cardiomyopathy	cTNT	Aberrant splicing isoforms
	Hypertension	G-protein β3	Exon skipping
Metabolic diseases			
	Glycogen storage disease, type II	Lysosomal α-glucosidase	Cryptic ss usage
	Hereditary tyrosinemia, Type I	Fumarylacetoacetate hydrolase	Exon skipping
	Acute intermittent porphyria	Porphobilinogen deaminase	Exon skipping
	Ceruloplasmin deficiency	Ceruloplasmin	Cryptic ss usage
	Fabry's disease	Lysosomal α galactosidase A	Cryptic ss usage
	Tay–Sach's Disease	β-hexosaminidase	Intron retention, exon skipping
	Sandhof disease	β hexosaminidase β-subunit	Cryptic ss usage

Tab. 4 (*continued*)

	Diseases	*Genes*	*Aberrant splicing*
Neurodegenerative diseases			
	FTDP-17	Tau	Imbalance of splicing isoforms
	Alzheimer's disease	Presenilin-1	Exon skipping
	Alzheimer's disease	Presenilin-2	Exon skipping
	Ataxia telangiectasia	ATM	Cryptic ss usage
	Multiple sclerosis	CD45	Imbalance of splicing isoforms
	Spinal muscular atrophy	SMN1, SMN2	Exon skipping
	Retinitis pigmentosa	HPRP3, PRPF31, PRPC8	Unknown
Psychiatric disorders			
	Schizophrenia	$GABA_AR\gamma 2$, NCAM	Imbalance of splicing isoforms
	Schizophrenia	NMDAR1	Aberrant alternative splicing
	ADHD	Nicotinic acetylcholine receptor	Aberrant alternative splicing
Other syndromes or diseases			
	Cystic fibrosis	CFTR	Exon skipping
	IGHD II	GH-1	Imbalance of splicing isoforms
	Frasier syndrome	WT 1	Imbalance of splicing isoforms
	Epilepsy	AMPA receptor	Imbalance of splicing isoforms
	Menkes disease	MNK	Exon skipping
	Beta-thalassemia	β-globin	Cryptic ss usage
	Metachromatic leukodystrophy	Arylsulfatase A	Cryptic ss usage
	Myotonic dystrophy	DMPK	(CUG)n expansion:aberrant splicing

chromosomal translocation have been reported. Such aberrant gene fusion-splicing products may contribute to pathogenesis of tumors caused by chromosomal translocation. Finally, the development of drug resistance has been associated with changes in alternative splicing of members of multidrug resistance genes or genes involved in drug metabolism.

Cardiovascular diseases are among the leading causes of human death. Aberrant splicing of structural genes such as cTNT is associated with cardiomyopathy. Splicing defects resulting in the formation

of a defective receptor, such as the low-density lipoprotein receptor (LDLR), lead to familial hypercholesterolemia (FH). A single-nucleotide change altering the function of intracellular signal transduction molecules can cause diseases affecting multiple systems. For example, a single-nucleotide polymorphism leading to the formation of a G-protein beta3 subunit splice variant has been associated with hypertension. Aberrant or defective splicing can disrupt the function of genes encoding a wide range of proteins important for the function of the cardiovascular system, from cell surface receptors to intracellular signaling molecules. Both structural genes and regulatory genes can be affected by splicing mutations. These examples show that splicing mutations can affect development/formation of the cardiovascular system and regulation of cardiovascular system function.

A range of metabolic diseases is caused by splicing mutations, affecting the production of a single metabolic enzyme. For example, the glycogen storage diseases, Sandhof's disease, and Fabry's disease can be caused by the usage of a single cryptic splice site in the corresponding genes. Hereditary tyrosinemia (type I) and acute intermittent porphyria can develop because of improper exon skipping in the genes encoding for respective metabolic enzymes.

The critical role of splicing regulation for the normal function of the nervous system is demonstrated by the large number of neurodegenerative and psychiatric disorders associated with aberrant splicing. Splicing defects in *tau* and presenilin genes have been identified in different types of dementia, including frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17) and Alzheimer's disease. Splicing mutations that cause cryptic splice site usage in the ataxia telangiectasia (ATM) gene have been identified in ATM patients. Multiple sclerosis has been associated with imbalances of different splicing isoforms of CD45. A number of studies have shown that genetic mutations in ubiquitously expressed protein factors that affect spliceosome formation can lead to diseases with specific neurological manifestations, such as spinal muscular atrophy and retinitis pigmentosa. For example, mutations in SMN1 or SMN2 cause spinal muscular atrophy. Defects in genes essential for spliceosomal assembly, including HPRP3, PRPF31, and PRPC8, have been found in autosomal dominant retinitis pigmentosa. The molecular mechanisms underlying such neuronal-specific diseases caused by defects in the general splicing factors remain to be investigated.

Psychiatric disorders, including attention deficit/hyperactivity disorder (ADHD) and schizophrenia, have been associated with aberrant splicing. Imbalance of different splicing isoforms or aberrant splicing products of several genes were detected in the brain tissues of patients with schizophrenia. These genes include $GABA_AR\gamma 2$, NMDAR1, neuronal nicotinic acetylcholine receptors, and neural cell adhesion molecule (NCAM). Changes in splicing isoforms of astroglial AMPA receptors have also been reported in human temporal lobe epilepsy. Several other examples of diseases caused by splicing mutations or disregulation of alternative splicing are listed in Table 1, including cystic fibrosis, familial isolated growth hormone deficiency type II (IGHD II), Frasier's syndrome, Menkes disease, β-thalassemia, and metachromatic leukodystrophy. Again, a variety of aberrant splicing events in the corresponding genes cause

the formation of defective gene products and consequently the disease phenotypes.

3.2 Molecular Mechanisms Underlying Splicing Defects Associated with Disease

Genetic defects in pre-mRNA splicing that cause human diseases can be classified into two categories, cis-acting mutations that alter expression or function of single genes and trans-acting defects that affect the components of splicing machinery or regulators of alternative splicing. Both cis-acting and trans-acting splicing mutations can result in clinical manifestations, either primarily involving a single tissue/organ or affecting multiple systems. These mutations can cause diseases by their direct effects on single genes or by indirect mechanisms such as disrupting expression or regulation of multiple genes.

Cis-acting splicing mutations can cause diseases by either producing defective/aberrant transcripts or by simply altering the delicate balance of different naturally expressed splicing isoforms. Development of disease phenotypes can be the result of loss of function of the involved genes or gain-of-function toxicities associated with the aberrant gene products. A large number of studies have focused on relationships between cis-acting splicing mutations and specific defects in the formation or function of the affected genes. However, the relationship is much less understood between genetic defects in trans-acting splicing factors (such as spliceosomal components or splicing regulators) and specific disease phenotypes.

At least four types of molecular mechanisms have been described for cis-acting splicing mutations that cause human diseases (Fig. 6). These mutations often lead to the formation of defective protein products or the loss of function of the genes involved as a result of RNA or protein instability. Exon skipping is perhaps the most commonly reported mechanism for the production of defective gene products. Activation of cryptic splice sites has also been

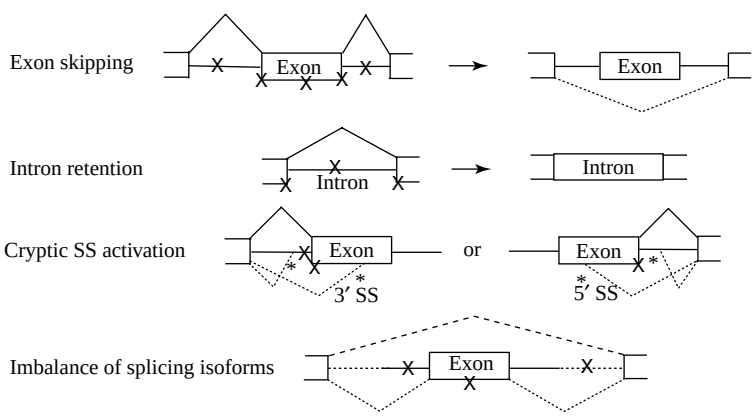

Fig. 6 Mechanisms for cis-acting splicing mutations. Positions of mutations are marked by "X," and they can be at the splice sites or within intronic or exonic sequences. The cryptic splice sites are marked by "*." Normal splicing events are illustrated by solid lines, and aberrant splicing events are depicted in dotted lines.

frequently associated with the production of aberrant or defective gene products. The failure to remove introns or intron retention is another mechanism for the formation of truncated protein products or the complete loss of gene function, because the inclusion of intronic sequences often introduces premature stop codons in gene transcripts. Finally, disturbance in the balances among naturally occurring splicing isoforms can lead to human disease. These mechanisms are not exclusive of each other. Multiple mutations acting by different mechanisms have been found in the same genes, leading to similar disease phenotypes. For example, both exon skipping and cryptic splice site activation contributing to the formation of defective LDLR lead to hypercholesterolemia. Similarly, exon skipping or cryptic splice site usage in Fibrillin-1 gene have been associated with Marfan syndrome.

Exon skipping can be caused by mutations at the splice sites or in splicing enhancers (either ESEs or ISEs). Translationally silent mutations can be functionally significant in splicing. For example, third codon changes that do not affect peptide sequences have often been overlooked as disease-causing mutations. Recently, such mutations have been examined at the splicing level. Such "silent" mutations may cause significant disruption in splicing, because defective function of ESEs leads to either improper exon skipping or imbalance of natural splicing isoforms.

Activation of cryptic splice sites in a large number of genes has been associated with human pathogenesis (see examples in Table 4). Single-nucleotide changes at the authentic splice sites result in inactivation of authentic splice sites. The consequences can be failures in exon inclusion or intron removal in some genes, or selection/activation of cryptic splice sites in other genes.

Intron retention has been examined in a number of disease genes. The average size of introns in human genes is approximately 3 kbp. The retention of even a small single intron may have catastrophic effects on the function of the genes affected because of the formation of defective or truncated peptides. Alternatively, intron retention can also cause a complete loss of expression of the mutated genes as a result of nonsense-mediated mRNA decay or instability of the gene products. Tay–Sach's disease can be caused by either intron retention or exon skipping in the splicing of the β-hexosaminidase gene.

Disturbance in the delicate balance of different natural splicing isoforms is now being recognized as a common mechanism for splicing diseases. It has been identified in neurodegenerative diseases, psychiatric disorders, and other diseases. These types of splicing defects can be caused by point mutations at splice junctions or in splicing regulatory elements (enhancers or silencers) located both in exon and intronic regions. FTDP-17, caused by mutations in the human *tau* gene, is a good example of such splicing defects. In addition to mutations that affect biophysical or biochemical function of Tau proteins, a large number of mutations that alter the ratio of natural splicing isoforms have been identified in FTDP-17 patients. The *tau* gene undergoes complex alternative splicing during the development of the nervous system. Six splicing isoforms are produced, three containing exon 10 and three lacking exon 10. These isoforms are named Tau4R and Tau3R respectively, because exon 10 encodes one of four microtubule-binding repeats. The balance among different Tau

isoforms is important for normal brain function, and disruption of the balance of Tau isoforms leads to the development of FTDP-17. Maintaining the appropriate balance of different Tau isoforms is controlled by complex interactions between the splicing machinery and various splicing regulatory elements in the *tau* gene. At the 5′ splice site of exon 10, a stem-loop type of secondary structure modulates the recognition of this 5′ splice site by U1snRNP. A number of intronic or exonic mutations destabilizing this stem-loop structure leads to an increase in Tau4R production. Other regulatory elements residing in exon 10, either splicing enhancers or silencers, may be disrupted by point mutations or small deletions, leading to the imbalance of different Tau isoforms. These mutations may increase (such as N279K) or decrease (such as L284L, del280K) the activity of the exonic splicing enhancer.

Other examples of diseases caused by disruption of the balance of different alternative splicing isoforms include Frasier syndrome, familial isolated growth hormone deficiency type II (IGHD II) and atypical cystic fibrosis. Such splicing defects have also been implicated in the pathogenesis of other diseases such as multiple sclerosis and schizophrenia. In Frasier syndrome, intronic mutations are found to change the ratio of +KTS to – KTS isoforms of Wilms tumor suppressor (WT1) gene products. The formation of these two splicing isoforms is the result of selection of two competing 5′ splice sites in exon 9. These two splice sites are separated by 9 nucleotides encoding KTS (lysine-threonine-serine). The use of the upstream and the downstream 5′ splice sites produces – KTS and +KTS isoforms, and the balance of these isoforms is important for kidney and gonad development. Mutations that decrease the use of the downstream +KTS 5′ splice site with an increase in the – KTS isoform have been associated with the majority of Frasier syndrome cases. These examples clearly show the significant role of splicing defects in human pathogenesis.

Spinal muscular atrophy (SMA) is an example that demonstrates the complexity of mechanisms underlying diseases caused by defects in trans-acting factors, such as proteins essential for spliceosomal snRNP biogenesis. SMA is a leading cause of infant mortality. Pathologically, it is characterized by degeneration of motor neurons in the anterior horn of the spinal cord leading to muscular atrophy. Genetic defects causing SMA include deletions or mutations in the survival of motor neuron genes (SMN). There are two copies of highly homologous SMN genes in humans, *SMN1* and *SMN2*. The *SMN1* gene is deleted or mutated in the majority of SMA patients. The SMN proteins are detected as distinctive speckles termed *gems* within the nucleus. SMN proteins contain an RNP1 motif and are critical for snRNP biogenesis and therefore, the formation of the spliceosome. Both SMN genes are expressed in a wide range of tissues. Although the predicted peptide sequences of *SMN1* and *SMN2* genes are identical, the gene products produced from the two genes in cells are different. A single translationally silent C to T nucleotide change at position +6 of exon 7 leads to the inefficient inclusion of exon 7 in the *SMN2* gene product, and thus the failure of *SMN2* to replace the function of *SMN1* and to provide protection against SMA. It is not clear why mutations in the *SMN* gene specifically cause motor neuron–specific disease. Although

the *SMN2* gene is not usually mutated in SMA patients, studies on stimulating *SMN2* exon 7 alternative splicing suggest a potential therapeutic approach based on the activation of alternative splicing of genes homologous to mutated genes (Sect. 4.2).

Myotonic dystrophy (DM), an autosomal dominant disease, is an example of trans-acting genetic mutations in which remarkable progress has been made. DM is a most common form of muscular dystrophy affecting both skeletal muscle and smooth muscles. Type I myotonic dystrophy (DM1) is caused by a CTG trinucleotide expansion in the 3′ untranslated region (3′UTR) of the DMPK (DM protein kinase) gene on chromosome 19q13.3. Type II myotonic dystrophy (DM2) is associated with a large CCTG repeat expansion in the intron 1 of the *ZNF9* gene. An "RNA gain-of-function" hypothesis has been proposed that these CTG or CCTG repeat expansions cause the formation of aberrant RNA transcripts containing large CUG or CCUG repeats. Such RNA transcripts containing long tracks of CUG/CCUG repeats may disrupt the normal function of certain RNA-binding proteins and induce secondary aberrant splicing of other genes. The RNA-binding proteins involved are likely to be the splicing regulators of CUG-BP family. The disruption of CUG-BP protein functions leads to aberrant splicing of genes including cardiac troponin T, insulin receptor, muscle-specific chloride channel, and tau. The aberrant splicing of these downstream genes can explain the cardiac phenotype, insulin resistance and myotonia found in DM patients. These findings demonstrate the complex roles of alternative splicing regulation in human pathogenesis.

Several trans-acting splicing defects have been reported in autosomal dominant retinitis pigmentosa (adRP). Recent genetic studies demonstrate that mutations in genes encoding general splicing factors such as HPRP3, PRPC8, or PRPF31 cause adRP. It remains unclear why these seemingly general defects in the splicing machinery cause such neuron-specific diseases. The underlying pathogenetic mechanisms await further investigation.

4 Perspectives on Diagnosis and Treatment of Diseases Caused by pre-mRNA Splicing Defects

4.1 Diagnosis of Human Diseases Caused by Splicing Defects

In the past three decades, significant advances have been made in the diagnosis of human genetic diseases, including diseases caused by splicing mutations. However, currently available molecular diagnostic tools remain limited for detecting splicing mutations.

Definitive diagnosis of splicing diseases should be based on the following criteria. First, the clinical manifestations should correlate with defects in a given gene. Second, corresponding splicing mutations are detected in genomic DNA samples of the patients. Third, aberrant or defective splicing products should be detected in the affected tissues or cells from patients. Detection of genomic mutations relies on DNA sequence analysis. In most cases, efforts in genomic DNA sequence analysis have been focusing on either exonic regions or near splice junctions. Detection of aberrant or defective splicing products

from patient samples can be technically challenging. The most frequently used methods include RT-PCR and analyses of defective protein products. In general, these methods are still limited to research studies. Systematic approaches to detecting disease-causing splicing mutations are yet to be developed for diagnostic applications in clinics.

4.2 Potential Therapeutic Approaches

Molecular studies and mechanistic characterization of splicing mutations that cause human disease have led to the development of a number of potential therapeutic approaches. They can be classified into four types, oligonucleotide-, ribozyme-, trans-splicing- and chemical compound-based approaches. These approaches are still at the stage of research and development.

4.2.1 Oligonucleotide-based Approaches: Antisense, RNAi, and Chimeric Molecules

Several groups have tested oligonucleotide-based methods for designing therapies to treat diseases caused by aberrant pre-mRNA splicing. These approaches include using modified antisense oligonucleotides or RNA interference (RNAi). Antisense oligonucleotides could be used to restore the proper function of mRNAs that are disrupted by splicing mutations. Antisense oligonucleotides designed to block cryptic sites in the mutated globin gene in β-thalassemia were reported to increase wild-type mRNA and decrease aberrant mRNA. A lentiviral vector-based system using a modified U7snRNA containing a sequence blocking the aberrant splice sites, reduced aberrant pre-mRNAs and increased levels of the correctly spliced β-globin mRNA and protein. The production of stress-induced aberrant AChE-R mRNA was reduced by antisense oligonucleotides. Another study using the antisense oligonucleotide approach reported the reversal of the aberrant splicing pattern caused by mutations associated with FTDP-17 in the human *tau* gene. On the other hand, antisense oligonucleotides have been used to increase the inclusion of exon 7 of SMN2 in order to develop therapeutic agents for spinal muscular atrophy. 2′-O-Methylated antisense oligoribonucleotides were used to modify the splicing pattern of the dystrophin pre-mRNA in the mdx mouse model of Duchenne muscular dystrophy. Recently, RNAi was used to selectively degrade alternatively spliced mRNA isoforms in *Drosophila* by treating cultured cells with dsRNA corresponding to an alternatively spliced exon.

One of most common splicing defects associated with human diseases is exon skipping. Because an RS domain can act as a splicing activator, small chimeric molecules containing a minimal synthetic RS domain covalently linked to an antisense moiety have been tested to target defective BRCA1 or SMN2 pre-mRNA transcripts and shown to restore splicing. These oligonucleotide-based approaches have shown feasibility in correcting splicing defects and treating the corresponding splicing diseases.

4.2.2 Ribozymes

Ribozymes are RNAs that catalyze biochemical reactions in cells, especially cleavage of other nucleic acids. Efforts have been made in developing derivatives from small naturally occurring RNAs, including the hammerhead, the hairpin, the tRNA processing ribonuclease P (RNase P), and group I and group II ribozymes as therapeutic agents. Ribozymes can be used to reduce aberrant or defective

splicing transcripts. In addition, a trans-splicing group I ribozyme was shown to convert mutant transcripts to normal mRNAs in the beta-globin and p53 genes.

4.2.3 SMaRT

Spliceosome-mediated RNA trans-splicing (SMaRT) was developed utilizing the endogenous trans-splicing activity in mammalian cells to correct aberrant splicing. SMaRT-mediated repair was reported to partially correct splicing defects in cystic fibrosis transmembrane conductance (CFTR) gene in cultured cells and bronchial xenografts.

4.2.4 Chemical Compounds

A number of chemical compounds have been shown to interact with RNA and/or RNA-binding proteins. Using chemical compounds to correct pre-mRNA splicing defects is being actively explored as a new therapeutic approach. Although the underlying mechanisms remain unclear, aclarubicin and sodium butyrate increase the inclusion of exon 7 in SMN2 transcripts in fibroblasts derived from spinal muscular atrophy patients or in transgenic mice, suggesting the therapeutic potential of chemical compounds in treating diseases associated with defective splicing.

Other chemical compounds are being tested that are capable of modifying splicing regulators, for example, kinases or phosphatases, that influence SR protein functions.

5 Concluding Remarks

After more than two decades of studies, a general picture of mammalian pre-mRNA splicing and alternative splicing regulation has begun to emerge. The basic components of the mammalian splicing machinery have been identified. The highly dynamic process of spliceosome assembly involves multiple networks of RNA–RNA, RNA–protein and protein–protein interactions. Recognition of splice sites and splicing regulatory sequences has been investigated in a number of genes, leading to identification and characterization of both cis-acting regulatory elements and trans-acting factors. We have begun to appreciate the contribution of alternative pre-mRNA splicing to creating genetic diversity, especially in mammals. A large number of splicing mutations that cause human diseases are being identified and characterized. We now have a glimpse of the complex picture of the involvement of aberrant or defective splicing in the pathogenesis of human diseases. Furthermore, efforts are being made to improve the diagnosis and treatment of diseases associated with pre-mRNA splicing defects.

Despite the significant progress, we are only at the beginning stage of understanding the molecular mechanisms controlling pre-mRNA splicing and alternative splicing regulation. A number of important questions remain to be addressed.

For the majority of genes, we have little knowledge about their complete expression profiles of different splicing isoforms in different cell types. We do not know how the splicing events of different genes are coordinated during development or in response to environmental changes. How a cell senses the environmental stimuli and responds by producing different splicing products remains largely unknown. Molecular pathways that transduce extracellular signals into the nuclear splicing machinery have not yet been delineated. Further development of new technologies such as

using microarray approaches to examine splicing isoforms under different physiological or pathological conditions at the genome level may be important for understanding the role of alternative pre-mRNA splicing in the biology of mammalian cells.

The basic components of mammalian splicing machinery are now characterized at the molecular level. However, the dynamic interactions among different components of the splicing machinery during spliceosome assembly and recognition of splicing regulatory elements are far from being understood. Little is known about the structural basis of such multicomponent interactions. Furthermore, the catalytic mechanisms of pre-mRNA splicing remain to be elucidated.

On the basis of studies of the relatively few model genes, a number of cis-acting elements and trans-acting splicing regulators have been identified. The mechanisms by which the splicing machinery specifically recognizes authentic versus decoy, or pseudosplice sites remain unclear. Some of the splicing regulatory factors are also important players for spliceosome assembly. Spliceosomal proteins also play a role in other processes of gene regulation. Further work needs to be carried out to understand the relationship between pre-mRNA splicing and other processes of gene expression and regulation, including transcription, RNA editing, RNA transport, translational control, and post-translational modification.

Studies of human diseases caused by splicing mutations or aberrant splicing have significantly advanced our understanding of human genetic diseases. Tremendous effort is still required to understand pathogenetic mechanisms underlying diseases caused by splicing defects, to develop efficient diagnostic tools and to design effective therapeutic approaches for these diseases.

Acknowledgment

We apologize to readers and our colleagues for having referred only to reviews instead of original research articles because of the space limit. The work in the authors' laboratory has been generously supported by Leukemia and Lymphoma Society, Muscular Dystrophy Association, Alzheimer's Association, the Society for Progressive Supranuclear Palsy, and NIH (AG17518, GM53945 and EY014576). We thank Drs Chris Burge, Reinhard Lührmann, Robin Reed, Phil Sharp, and Joan Steitz for sharing information and for giving permissions to modify figures from their previous publications. We are grateful to Drs Edward Benz, Doug Black, James Bruzik, Mariano Garcia-Blanco, Warren Gish, David Ornitz, Arnold Strauss, and members of Wu laboratory for their critical reading of the manuscript. We thank German G. Leparc, Joe Gee Seo, and Ying Zhang for the data bank searches and Mrs. Angie Hantak for outstanding assistance.

See also Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Adams, M.D., Rudner, D.Z., Rio, D.C. (1996) Biochemistry and regulation of pre-mRNA splicing, *Curr. Opin. Cell. Biol.* **8**(3), 331–339.

Bartel, F., Taubert, H., Harris, L.C. (2002) Alternative and aberrant splicing of MDM2 mRNA in human cancer, *Cancer Cell* **2**(1), 9–15.

Baserga, S.J., Yang, X.D., Steitz, J.A. (1993) The Diverse World of Small Ribonucleoproteins, in: Gesteland, R.F., Atkins, J.F. (Eds.) *The RNA World*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 359–381.

Berget, S.M. (1995) Exon recognition in vertebrate splicing, *J. Biol. Chem.* **2**, 2411–2414.

Black, D.L. (2000) Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology, *Cell* **103**(3), 367–370.

Black, D.L. (2003) Mechanisms of alternative pre-mRNA splicing, *Annu. Rev. Biochem.* **72**, 291–336.

Black, D.L., Grabowski, P.J. (2003) Alternative pre-mRNA splicing and neuronal function, *Prog. Mol. Subcell Biol.* **31**, 187–216.

Blencowe, B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases, *Trends Biochem. Sci.* **25**(3), 106–110.

Bilbao, D., Valcarcel, J. (2003) Getting to the heart of a splicing enhancer, *Nat. Struct. Biol.* **10**(1), 6–7.

Brett, D., Pospisil, H., Valcarcel, J., Reich, J., Bork, P. (2003) Alternative splicing and genome complexity, *Nat. Genet.* **30**(1), 29–30.

Brow, D.A. (2002) Allosteric cascade of spliceosome activation, *Annu. Rev. Genet.* **36**, 333–360.

Buchholz, T.A., Wazer, D.E. (2002) Molecular biology and genetics of breast cancer development: a clinical perspective, *Semin. Radiat. Oncol.* **12**(4), 285–295.

Burge, C.B., Tuschl, T.H., Sharp, P.A. (1999) in: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.) *The RNA World*, Cold Spring Harbor Laboratory Press, New York, pp. 525–560.

Caceres, J.F., Kornblihtt, A.R. (2002) Alternative splicing: multiple control mechanisms and involvement in human disease, *Trends Genet.* **18**, 186–193.

Cartegni, L., Chew, S.L., Krainer, A.R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing, *Nat. Rev. Genet.* **3**(4), 285–298.

Chabot, B., LeBel, C., Hutchison, S., Nasim, F.H., Simard, M.J. (2003) Heterogeneous nuclear ribonucleoprotein particle A/B proteins and the control of alternative splicing of the mammalian heterogeneous nuclear ribonucleoprotein particle A1 pre-mRNA, *Prog. Mol. Subcell Biol.* **31**, 59–88.

Cooper, T.A., Mattox, W. (1997) The regulation of splice-site selection, and its role in human disease, *Am. J. Hum. Genet.* **61**(2), 259–266.

Dredge, B.K., Polydorides, A., Darnell, R.B. (2001) The splice of life: alternative splicing and neurological disease, *Nat. Neurosci. Rev.* **2**, 43–50.

Faustino, N.A., Cooper, T.A. (2003) Pre-mRNA splicing and human disease, *Genes Dev.* **17**(4), 419–437.

Forch, P., Valcarcel, J. (2003) Splicing regulation in Drosophila sex determination, *Prog. Mol. Subcell Biol.* **31**, 127–151.

Fu, X.D. (1995) The superfamily of arginine/serine-rich splicing factors, *RNA* **1**(7), 663–680.

Goedert, M., Spillantini, M.G. (2001) Tau gene mutations and neurodegeneration, *Biochem. Soc. Symp.* **67**, 59–71.

Goldstrohm, A.C., Greenleaf, A.L., Garcia-Blanco, M.A. (2001) Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing, *Gene* **277**(1–2), 31–47.

Grabowski, P.J. (1998) Splicing regulation in neurons: tinkering with cell-specific control, *Cell* **92**(6), 709–712.

Grabowski, P.J., Black, D.L. (2001) Alternative RNA splicing in the nervous system, *Prog. Neurobiol.* **65**, 289–308.

Graveley, B.R. (2000) Sorting out the complexity of SR protein functions, *RNA* **6**(9), 1197–1211.

Hastings, M.L., Krainer, A.R. (2001) Pre-mRNA splicing in the new millennium, *Curr. Opin. Cell Biol.* **13**, 302–309.

Hertel, K.J., Lynch, K.W., Maniatis, T. (1997) Common themes in the function of transcription and splicing enhancers, *Curr. Opin. Cell Biol.* **9**(3), 350–357.

Holmila, R., Fouquet, C., Cadranel, J., Zalcman, G., Soussi, T. (2003) Splice mutations in the p53 gene: case report and review of the literature, *Hum. Mutat.* **21**(1), 101–102.

Konarska, M.M. (1998) Recognition of the 5′ splice site by the spliceosome, *Acta Biochim. Pol.* **45**(4), 869–881.

Krämer, A. (1996) The structure and function of proteins involved in mammalian pre-mRNA splicing, *Annu. Rev. Biochem.* **65**, 367–409.

Lopez, A.J. (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation, *Annu. Rev. Genet.* **32**, 279–305.

Maniatis, T., Reed, R. (2002) An extensive network of coupling among gene expression machines, *Nature* **416**(6880), 499–506.

Maniatis, T., Tasic, B. (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans, *Nature* **418**(6894), 236–243.

Manley, J.L. (2002) Nuclear coupling: RNA processing reaches back to transcription, *Nat. Struct. Biol.* **9**(11), 790–791.

Manley, J.L., Tacke, R. (1996) SR proteins and splicing control, *Genes Dev.* **10**(13), 1569–1579.

Maroney, P.A., Romfo, C.M., Nilsen, T.W. (2002) Functional recognition of 5′ splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly, *Mol. Cell* **6**, 317–328.

Nilsen, T.W. (2002) The spliceosome: no assembly required? *Mol. Cell* **9**(1), 8–9.

Nissim-Rafinia, M., Krem, B. (2002) Splicing regulation as a potential genetic modifier, *Trends Genet.* **18**, 123–127.

Moore, M.J. (2000) Intron recognition comes of AGe, *Nat. Struct. Biol.* **7**(1), 14–16.

Mount, S.M. (2000) Genomic sequence, splicing, and gene annotation, *Am. J. Hum. Genet.* **67**(4), 788–792.

Mount, S.M., Salz, H.K. (2000) Pre-messenger RNA Processing Factors in the *Drosophila* Genome, *J. Cell Biol.* **150**, F37–F43.

Paushkin, S., Gubitz, A.K., Massenet, S., Dreyfuss, G. (2002) The SMN complex, an assemblyosome of ribonucleoproteins, *Curr. Opin. Cell Biol.* **14**(3), 305–312.

Ponta, H., Wainwright, D., Herrlich, P. (1998) The CD44 protein family, *Int. J. Biochem. Cell Biol.* **30**(3), 299–305.

Proudfoot, N.J., Furger, A., Dye, M.J. (2002) Integrating mRNA processing with transcription, *Cell* **108**, 501–512.

Reed, R. (1996) Initial splice-site recognition and pairing during pre-mRNA splicing, *Curr. Opin. Genet. Dev.* **6**(2), 215–220.

Reed, R. (2000) Mechanisms of fidelity in pre-mRNA splicing, *Curr. Opin. Cell Biol.* **12**(3), 340–345.

Reed, R., Hurt, E. (2002) A conserved mRNA export machinery coupled to pre-mRNA splicing, *Cell* **108**, 523–531.

Roberts, G.C., Smith, C.W. (2002) Alternative splicing: combinatorial output from the genome, *Curr. Opin. Chem. Biol.* **6**(3), 375–383.

Rosonina, E., Blencowe, B.J. (2002) Gene expression: the close coupling of transcription and splicing, *Curr. Biol.* **12**(9), R319–R321.

Sharp, P.A. (1994) Split genes and RNA splicing, *Cell* **77**(6), 805–815.

Smith, C.W., Valcarcel, J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control, *Trends Biochem. Sci.* **25**(8), 381–388.

Soret, J., Tazi, J. (2003) Phosphorylation-dependent control of the pre-mRNA splicing machinery, *Prog. Mol. Subcell Biol.* **31**, 89–126.

Staley, J.P., Guthrie, C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things, *Cell* **92**(3), 315–326.

Steitz, J.A. (1984) Autoantibody probes for mammalian gene expression, *Harvey. Lect.* (1984–85) **80**, 39–47.

Stoilov, P., Meshorer, E., Gencheva, M., Glick, D., Soreq, H., Stamm, S. (2002) Defects in pre-mRNA processing as causes of and predisposition to diseases, *DNA Cell Biol.* **21**(11), 803–818.

Suzuki, T., Nishio, K., Tanabe, S. (2001) The MRP family and anticancer drug metabolism, *Curr. Drug Metab.* **2**(4), 367–377.

Tacke, R., Manley, J.L. (1999) Determinants of SR protein specificity, *Curr. Opin. Cell Biol.* **11**(3), 358–362.

Tarn, W.Y., Steitz, J.A. (1997) Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge, *Trends Biochem. Sci.* **22**(4), 132–137.

Tupler, R., Perini, G., Green, M.R. (2001) Expressing the human genome, *Nature* **409**(6822), 832–833.

Valadkhan, S., Manley, J.L. (2002) Intrinsic metal binding by a spliceosomal RNA, *Nat. Struct. Biol.* **9**(7), 498–499.

Valcarcel, J., Green, M.R. (1996) The SR protein family: pleiotropic functions in pre-mRNA splicing, *Trends Biochem. Sci.* **21**(8), 296–301.

Villa, T., Pleiss, J.A., Guthrie, C. (2002) Spliceosomal snRNAs: Mg(2+)-dependent chemistry at the catalytic core? *Cell* **109**(2), 149–152.

Wagner, E.J., Garcia-Blanco, M.A. (2001) Polypyrimidine tract binding protein antagonizes exon definition, *Mol. Cell Biol.* **21**(10), 3281–3288.

Wang, J., Manley, J.L. (1997) Regulation of pre-mRNA splicing in metazoa, *Curr. Opin. Genet. Dev.* **7**(2), 205–211.

Will, C.L., Luhrmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function, *Curr. Opin. Cell Biol.* **13**(3), 290–301.

Wu, J.Y., Tang, H., Havlioglu, N. (2003) Alternative pre-mRNA splicing and regulation of programmed cell death, *Prog. Mol. Subcell Biol.* **31**, 153–185.

Primary Literature

Abovich, N., Rosbash, M. (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals, *Cell* **89**, 403–412.

Andreasi, C., Jarecki, J., Zhou, J., Coovert, D.D., Monani, U.R., Chen, X., Whitney, M., Pollok, B., Zhang, M., Androphy, E., Burghes, A.H.M. (2001) Aclarubicin treatment restores SMN levels to cells derived from type I spinal muscular atrophy patients, *Hum. Mol. Genet.* **24**, 2841–2849.

Arinobu, Y., Atamas, S.P., Otsuka, T., Niiro, H., Yamaoka, K., Mitsuyasu, H., Niho, Y., Hamasaki, N., White, B., Izuhara, K. (1999) Antagonistic effects of an alternative splice variant of human IL-4, IL-4 delta2, on IL-4 activities in human monocytes B cells, *Cell Immunol.* **191**, 161–167.

Auboeuf, D., Honig, A., Berget, S.M., O'Malley, B.W. (2002) Coordinate regulation of transcription and splicing by steroid receptor coregulators, *Science* **298**(5592), 416–419.

Barnard, D.C., Li, J., Peng, R., Patton, J.G. (2002) Regulation of alternative splicing by SRrp86 through coactivation and repression of specific SR proteins, *RNA* **8**(4), 526–533.

Berget, S.M., Moore, C., Sharp, P.A. (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA, *Proc. Natl. Acad. Sci. U.S.A.* **74**(8), 3171–3175.

Berglund, J.A., Abovich, N., Rosbash, M. (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition, *Genes Dev.* **12**, 858–867.

Berglund, J.A., Chua, K., Abovich, N., Reed, R., Rosbash, M. (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC, *Cell* **89**(5), 781–787.

Biesiadecki, B.J., Elder, B.D., Yu, Z.B., Jin, J.P. (2002) Cardiac troponin T variants produced by aberrant splicing of multiple exons in animals with high instances of dilated cardiomyopathy, *J. Biol. Chem.* **277**(52), 50275–50285.

Bolduc, L., Labrecque, B., Cordeau, M., Blanchette, M., Chabot, B. (2001) Dimethyl sulfoxide affects the selection of splice sites, *J. Biol. Chem.* **276**(20), 17597–17602.

Boucher, L., Ouzounis, C.A., Enright, A.J., Blencowe, B.J. (2001) A genome-wide survey of RS domain proteins, *RNA* **7**(12), 1693–1701.

Brand, K., Dugi, K.A., Brunzell, J.D., Nevin, D.N., Santamarina-Fojo, S. (1996) A novel A→G mutation in intron I of the hepatic lipase gene leads to alternative splicing resulting in enzyme deficiency, *J. Lipid Res.* **37**(6), 1213–1223.

Bruzik, J.P., Van Doren, K., Hirsh, D., Steitz, J.A. (1988) Trans splicing involves a novel form of small nuclear ribonucleoprotein particles, *Nature* **335**(6190), 559–562.

Buckanovich, R.J., Darnell, R.B. (1997) The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo, *Mol. Cell Biol.* **17**(6), 3194–3201.

Buratti, E., Baralle, F.E. (2001) Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9, *J. Biol. Chem.* **276**, 36337–36343.

Buratti, E., Dork, T., Zuccato, E., Pagani, F., Romano, M., Baralle, F.E. (2001) Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping, *EMBO J.* **20**, 1774–1784.

Cáceres, J.F., Stamm, S., Helfman, D.M., Krainer, A.R. (1994) Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors, *Science* **265**, 1706–1709.

Cao, W., Jamison, S.F., Garcia-Blanco, M.A. (1997) Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro, *RNA* **3**(12), 1456–1467.

Caputi, M., Zahler, A.M. (2001) Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H′/F/2H9 family, *J. Biol. Chem.* **276**(47), 43850–43859.

Caputi, M., Zahler, A.M. (2002) SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D, *EMBO J.* **21**(4), 845–855.

Carlo, T., Sierra, R., Berget, S.M. (2000) A 5′ splice site-proximal enhancer binds SF1 and activates exon bridging of a microexon, *Mol. Cell Biol.* **20**(11), 3988–3995.

Carstens, R.P., McKeehan, W.L., Garcia-Blanco, M.A. (1998) An intronic sequence element mediates both activation and repression of rat fibroblast growth factor receptor 2 pre-mRNA splicing, *Mol. Cell Biol.* **18**(4), 2205–2217.

Carstens, R.P., Wagner, E.J., Garcia-Blanco, M.A. (2000) An intronic splicing silencer causes skipping of the IIIb exon of fibroblast growth factor receptor 2 through involvement of polypyrimidine tract binding protein, *Mol. Cell Biol.* **20**(19), 7388–7400.

Cartegni, L., Krainer, A.R. (2002) Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1, *Nat. Genet.* **30**(4), 377–384.

Cartegni, L., Krainer, A.R. (2003) Correction of disease-associated exon skipping by synthetic exon-specific activators, *Nat. Struct. Biol.* **10**(2), 120–125.

Celotto, A.M., Graveley, B.R. (2002) Exon-specific RNAi: a tool for dissecting the functional relevance of alternative splicing, *RNA* **8**(6), 718–724.

Chakarova, C.F., Hims, M.M., Bolz, H., Abu-Safieh, L., Patel, R.J., Papaioannou, M.G., Inglehearn, C.F., Keen, T.J., Willis, C., Moore, A.T., Rosenberg, T., Webster, A.R., Bird, A.C., Gal, A., Hunt, D., Vithana, E.N., Bhattacharya, S.S. (2002) Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa, *Hum. Mol. Genet.* **11**, 87–92.

Chalfant, C.E., Rathman, K., Pinkerman, R.L., Wood, R.E., Obeid, L.M., Ogretmen, B., Hannun, Y.A. (2002) De novo ceramide regulates the alternative splicing of caspase 9 and Bcl-x in A549 lung adenocarcinoma cells. Dependence on protein phosphatase-1, *J. Biol. Chem.* **277**(15), 12587–12595.

Chang, J.G., Hsieh-Li, H.M., Jong, Y.J., Wang, N.M., Tsai, C.H., Li, H. (2001) Treatment of spinal muscular atrophy by sodium butyrate, *PNAS* **98**, 9808–9813.

Charlet, B.N., Logan, P., Singh, G., Cooper, T.A. (2002a) Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing, *Mol. Cell* **9**(3), 649–658.

Charlet, B.N., Savkur, R.S., Singh, G., Philips, A.V., Grice, E.A., Cooper, T.A. (2002b) Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing, *Mol. Cell* **10**(1), 45–53.

Chen, C.D., Kobayashi, R., Helfman, D.M. (1999) Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat b-tropomyosin gene, *Genes Dev.* **13**, 593–606.

Chiba-Falek, O., Parad, R.B., Kerem, E., Kerem, B. (1999) Variable levels of normal RNA in different fetal organs carrying a cystic fibrosis transmembrane conductance regulator splicing mutation, *Am. J. Respir. Crit. Care Med.* **159**, 1998–2002.

Chou, M.Y., Rooke, N., Turck, C.W., Black, D.L. (1999) hnRNP H Is a Component of a Splicing Enhancer Complex That Activates a c-src Alternative Exon in Neuronal CellsMol, *Cell Biol.* **19**, 69–77.

Chow, L.T., Gelinas, R.E., Broker, T.R., Roberts, R.J. (1977) An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA, *Cell* **12**(1), 1–8.

Chui, D.H., Hardison, R., Riemer, C., et al. (1998) An electronic database of human hemoglobin variants on the world wide web, *Blood* **91**, 2643–2644.

Claes, K., Vandesompele, J., Poppe, B., Dahan, K., Coene, I., De Paepe, A., Messiaen, L. (2002) Pathological splice mutations outside the invariant AG/GT splice sites of BRCA1 exon 5 increase alternative transcript levels in the 5′ end of the BRCA1gene, *Oncogene* **21**(26), 4171–4175.

Cooper, S.R., Taylor, J.K., Miraglia, L.J., Dean, N.M. (1999) Pharmacology of antisense oligonucleotide inhibitors of protein expression, *Pharmacol. Ther.* **82**, 427–435.

Cote, J., Dupuis, S., Jiang, Z., Wu, J.Y. (2001a) Caspase-2 pre-mRNA alternative splicing: Identification of an intronic element containing a decoy 3′acceptor site, *Proc. Natl. Acad. Sci. U.S.A.* **98**(3), 938–943.

Cote, J., Dupuis, S., Wu, J.Y. (2001b) Polypyrimidine track-binding protein binding downstream of caspase-2 alternative exon 9 represses its inclusion, *J. Biol. Chem.* **276**(11), 8535–8543.

Coulter, L.R., Landree, M.A., Cooper, T.A. (1997) Identification of a new class of exonic splicing enhancers by in vivo selection, *Mol. Cell Biol.* **17**(4), 2143–2150.

Cowper, A.E., Caceres, J.F., Mayeda, A., Screaton, G.R. (2001) Serine-arginine (SR) protein-like factors that antagonize authentic SR proteins and regulate alternative splicing, *J. Biol. Chem.* **276**(52), 48908–48914.

Daoud, R., Da Penha Berzaghi, M., Siedler, F., Hubener, M., Stamm, S. (1999) Activity-dependent regulation of alternative splicing patterns in the rat brain, *Eur. J. Neurosci.* **11**(3), 788–802.

Daoud, R., Mies, G., Smialowska, A., Olah, L., Hossmann, K.A., Stamm, S. (2002) Ischemia induces a translocation of the splicing factor tra2-beta 1 and changes alternative splicing patterns in the brain, *J. Neurosci.* **22**(14), 5889–5899.

Das, S., Levinson, B., Whitney, S., Vulpe, C., Packman, S., Gitschier, J. (1994) Diverse mutations in patients with Menkes disease often lead to exon skipping, *Am. J. Hum. Genet.* **55**(5), 883–889.

Deguillien, M., Huang, S.C., Moriniere, M., Dreumont, N., Benz, E.J. Jr., Baklouti, F. (2001) Multiple cis elements regulate an alternative splicing event at 4.1R pre-mRNA during erythroid differentiation, *Blood* **98**(13), 3809–3816.

Del Gatto-Konczak, F., Bourgeois, C.F., Le Guiner, C., Kister, L., Gesnel, M.C., Stevenin, J., Breathnach, R. (2000) The RNA-binding protein TIA-1 is a novel mammalian splicing regulator acting through intron sequences adjacent to a 5′ splice site, *Mol. Cell Biol.* **20**(17), 6287–6299.

Eldridge, A.G., Li, Y., Sharp, P.A., Blencowe, B.J. (1999) The SRm160/300 splicing coactivator is required for exon-enhancer function, *Proc. Natl. Acad. Sci. U.S.A.* **96**(11), 6125–6130.

Fackenthal, J.D., Cartegni, L., Krainer, A.R., Olopade, O.I. (2002) BRCA2 T2722R is a deleterious allele that causes exon skipping, *Am. J. Hum. Genet.* **71**(3), 625–631.

Fairbrother, W.G., Yeh, R.F., Sharp, P.A., Burge, C.B. (2002) Predictive identification of exonic splicing enhancers in human genes, *Science* **297**(5583), 1007–1013.

Feltes, C.M., Kudo, A., Blaschuk, O., Byers, S.W. (2002) An alternatively spliced cadherin-11 enhances human breast cancer cell invasion, *Cancer Res.* **62**(22), 6688–6697.

Fischer, U., Liu, Q., Dreyfuss, G. (1997) The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis, *Cell* **90**(6), 1023–1029.

Fogel, B.L., McNally, M.T. (2000) A Cellular Protein, hnRNP H, Binds to the Negative Regulator of Splicing Element from Rous Sarcoma Virus, *J. Biol. Chem.* **275**, 32371–32378.

Forch, P., Puig, O., Kedersha, N., Martinez, C., Granneman, S., Seraphin, B., Anderson, P., Valcarcel, J. (2000) The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing, *Mol. Cell* **6**(5), 1089–1098.

Forch, P., Puig, O., Martinez, C., Seraphin, B., Valcarcel, J. (2002) The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5′ splice sites, *EMBO J.* **21**(24), 6882–6892.

Fouraux, M.A., Kolkman, M.J., Van der Heijden, A., De Jong, A.S., Van Venrooij, W.J., Pruijn, G.J. (2002) The human La (SS-B) autoantigen interacts with DDX15/hPrp43, a putative DEAH-box RNA helicase, *RNA* **8**(11), 1428–1443.

Fu, X.D., Maniatis, T. (1990) Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus, *Nature* **343**(6257), 437–441.

Fu, X.D., Maniatis, T. (1992) Isolation of a complementary DNA that encodes the mammalian splicing factor SC35, *Science* **256**(5056), 535–538.

Fujimaru, M., Tanaka, A., Choeh, K., Wakamatsu, N., Sakuraba, H., Isshiki, G. (1998) Two mutations remote from an exon/intron junction in the beta-hexosaminidase beta-subunit gene affect 3′-splice site selection and cause Sandhoff disease, *Hum. Genet.* **103**(4), 462–469.

Garcia-Blanco, M.A., Jamison, S.F., Sharp, P.A. (1989) Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns, *Genes Dev.* **3**(12A), 1874–1886.

Garcia-Blanco, M.A., Anderson, G.J., Beggs, J., Sharp, P.A. (1990) A mammalian protein of 220 kDa binds pre-mRNAs in the spliceosome: a potential homologue of the yeast PRP8 protein, *Proc. Natl. Acad. Sci. U.S.A.* **87**(8), 3082–3086.

Ge, H., Zuo, P., Manley, J.L. (1991) Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators, *Cell* **66**(2), 373–382.

Gebhardt, F., Zanker, K.S., Brandt, B. (1998) Differential expression of alternatively spliced c-erbB-2 mRNA in primary tumors, lymph node metastases, and bone marrow micrometastases from breast cancer patients, *Biochem. Biophys. Res. Commun.* **247**(2), 319–323.

Ghanem, M.A., Van Steenbrugge, G.J., Van Der Kwast, T.H., Sudaryo, M.K., Noordzij, M.A., Nijman, R.J. (2002) Expression and prognostic value Of CD44 isoforms in nephroblastoma (Wilms tumor), *J. Urol.* **168**(2), 681–686.

Gooding, C., Kemp, P., Smith, C.W. (2003) SmPTB: a novel polypyrimidine tract binding protein (PTB) paralog expressed in smooth muscle cells, *J. Biol. Chem.* in press.

Gozani, O., Patton, J.G., Reed, R. (1994) A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction, *EMBO J.* **13**(14), 3356–3367.

Graveley, B.R., Hertel, K.J., Maniatis, T. (1998) A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers, *EMBO J.* **17**(22), 6747–6756.

Graveley, B.R., Maniatis, T. (1998) Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing, *Mol. Cell* **1**(5), 765–771.

Groenen, P.J.T.A., Wansink, D.G., Coerwinkel, M., Van den broek, W., Jansen, G., Wieringa, B. (2000) Constitutive and regulated modes of splicing produce six major myotonic dystrophy protein kinase (DMPK) isoforms with distinct properties, *Hum. Mol. Genet.* **9**, 605–616.

Gui, J.F., Lane, W.S., Fu, X.D. (1994) A serine kinase regulates intracellular localization of splicing factors in the cell cycle, *Nature* **369**(6482), 678–682.

Grover, A., Houlden, H., Baker, M., Adamson, J., Lewis, J., Prihar, G., Pickering-Brown, S., Duff, K. (1999) Hutton M 5′ splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10, *J. Biol. Chem.* **274**, 15134–15143.

Guo, N., Kawamoto, S. (2000) An intronic downstream enhancer promotes 3′ splice site usage of a neural cell-specific exon, *J. Biol. Chem.* **275**(43), 33641–33649.

Hall, S.L., Padgett, R.A. (1996) Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns, *Science* **271**(5256), 1716–1718.

Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M., Luhrmann, R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method, *Proc. Natl. Acad. Sci. U.S.A.* **99**(26), 16719–16724.

Hasegawa, Y., Kawame, H., Ida, H., Ohashi, T., Eto, Y. (1994) Single exon mutation in arylsulfatase A gene has two effects: loss of enzyme activity and aberrant splicing, *Hum. Genet.* **93**(4), 415–420.

He, Y., Smith, S.K., Day, K.A., Clark, D.E., Licence, D.R., Charnock-Jones, D.S. (1999) Alternative splicing of vascular endothelial growth factor (VEGF)-R1 (FLT-1) pre-mRNA is important for the regulation of VEGF activity, *Mol. Endocrinol.* **13**, 537–545.

Hechtman, P., Boulay, B., De Braekeleer, M., Andermann, E., Melancon, S., Larochelle, J., Prevost, C., Kaplan, F. (1992) The intron 7 donor splice site transition: a second Tay-Sachs disease mutation in French Canada, *Hum. Genet.* **90**(4), 402–406.

Hibi, K., Trink, B., Patturajan, M., Westra, W.H., Caballero, O.L., Hill, D.E., Ratovitski, E.A., Jen, J., Sidransky, D. (2000) AIS is an oncogene amplified in squamous cell carcinoma, *Proc. Natl. Acad. Sci. U.S.A.* **97**(10), 5462–5467.

Hinek, A., Zhang, S., Smith, A.C., Callahan, J.W. (2000) Impaired elastic-fiber assembly by fibroblasts from patients with either Morquio B disease or infantile GM1-gangliosidosis is linked to deficiency in the 67-KD spliced variant of β galactosidase, *Am. J. Hum. Genet.* **67**, 23–36.

Honig, A., Auboeuf, D., Parker, M.M., O'Malley, B.W., Berget, S.M. (2002) Regulation of alternative splicing by the ATP-dependent DEAD-box RNA helicase p72, *Mol. Cell Biol.* **22**(16), 5698–5707.

Hou, V.C., Lersch, R., Gee, S.L., Ponthier, J.L., Lo, A.J., Wu, M., Turck, C.W., Koury, M., Krainer, A.R., Mayeda, A., Conboy, J.G. (2002) Decrease in hnRNP A/B expression during erythropoiesis mediates a pre-mRNA splicing switch, *EMBO J.* **21**(22), 6195–6204.

Huh, G.S., Hynes, R.O. (1994) Regulation of alternative pre-mRNA splicing by a novel repeated hexanucleotide element, *Genes Dev.* **8**(13), 1561–1574.

Hutton, M., Lendon, C.L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A., Hackett, J., Adamson, J., Lincoln, S., Dickson, D., Davies, P., Petersen, R.C., Stevens, M., de Graaff, E., Wauters, E., van Baren, J., Hillebrand, M., Joosse, M., Kwon, J.M., Nowotny, P., Heutink, P., et al. (1998)

Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17, *Nature* **393**, 702–705.

Ismaili, N., Sha, M., Gustafson, E.H., Konarska, M.M. (2001) The 100-kda U5 snRNP protein (hPrp28p) contacts the 5′ splice site through its ATPase site, *RNA* **7**(2), 182–193.

Jacquenet, S., Mereau, A., Bilodeau, P.S., Damier, L., Stoltzfus, C.M., Branlant, C. (2001) A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H, *J. Biol. Chem.* **276**(44), 40464–40475.

Jensen, K.B., Dredge, B.K., Stefani, G., Zhong, R., Buckanovich, R.J., Okano, H.J., Yang, Y.Y. (2000) Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability, *Neuron* **25**(2), 359–371.

Jiang, Z., Cote, J., Kwon, J.M., Goate, A.M., Wu, J.Y. (2000) Aberrant splicing of tau pre-mRNA caused by intronic mutations associated with the inherited dementia frontotemporal dementia with parkinsonism linked to chromosome 17, *Mol. Cell. Biol.* **20**, 4036–4048.

Jiang, Z., Tang, H., Havlioglu, N., Zhang, X., Stamm, S., Yan, R., Wu, J.Y. (2003) Mutations in tau gene exon10 associated with FTDP-17 alter the activity of an exonic splicing enhancer to interact with Tra2beta, *J. Biol. Chem.*

Jiang, Z.H., Zhang, W.J., Rao, Y., Wu, J.Y. (1998) Regulation of Ich-1 pre-mRNA alternative splicing and apoptosis by mammalian splicing factors, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9155–9160.

Jonghe, C.D., Cruts, M., Rogaeva, E.A., Tysoe, C., Singleton, A., Vanderstichele, H., Meschino, W., Dermant, B., Vanderhoeven, I., Backhovens, H., et al. (1999) Aberrant splicing in the presenilin-1 intron 4 mutation causes presenile Alzheimer's disease by increased Abeta42 secretion, *Hum. Mol. Genet.* **8**, 1529–1540.

Kalbfuss, B., Mabon, S.A., Misteli, T. (2001) Correction of alternative splicing of Tau in frontotemporal dementia and parkinsonism linked to chromosome 17, *J. Biol. Chem.* **276**, 42986–42993.

Kan, J.L., Green, M.R. (1999) Pre-mRNA splicing of IgM exons M1 and M2 is directed by a juxtaposed splicing enhancer and inhibitor, *Genes Dev.* **13**(4), 462–471.

Kanopka, A., Muhlemann, O., Akusjarvi, G. (1996) Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA, *Nature* **381**(6582), 535–538.

Kanopka, A., Muhlemann, O., Petersen-Mahrt, S., Estmer, C., Ohrmalm, C., Akusjarvi, G. (1998) Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins, *Nature* **393**(6681), 185–187.

Kennedy, C.F., Kramer, A., Berget, S.M. (1998) A role for SRp54 during intron bridging of small introns with pyrimidine tracts upstream of the branch point, *Mol. Cell. Biol.* **18**(9), 5425–5434.

Klamt, B., Koziell, A., Poulat, F., Wieacker, P., Scamber, P., Berta, P., Gessler, M. (1998) Frasier syndrome is caused by defective alternative splicing of WT1 leading to an altered ratio of WT1 +/− KTS splice isoforms, *Hum. Mol. Genet.* **7**, 709–714.

Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Luhrmann, R., Garcia-Blanco, M.A., Manley, J.L. (1994) Protein-protein interactions and 5′-splice-site recognition in mammalian mRNA precursors, *Nature* **368**(6467), 119–124.

Konforti, B.B., Koziolkiewicz, M.J., Konarska, M.M. (1993) Disruption of base pairing between the 5′ splice site and the 5′ end of U1 snRNA is required for spliceosome assembly, *Cell* **75**(5), 863–873.

Krainer, A.R., Conway, G.C., Kozak, D. (1990) The essential pre-mRNA splicing factor SF2 influences 5′ splice site selection by activating proximal sites, *Cell* **62**(1), 35–42.

Krainer, A.R., Maniatis, T., Ruskin, B., Green, M.R. (1984) Normal and mutant human beta-globin pre-mRNAs are faithfully and efficiently spliced in vitro, *Cell* **36**(4), 993–1005.

Kraus, A., Neff, F., Behn, M., Schuermann, M., Muenkel, K., Schlegel, J. (1999) Expression of alternatively spliced mdm2 transcripts correlates with stabilized wild-type p53 protein in human glioblastoma cells, *Int. J. Cancer* **80**(6), 930–934.

Labourier, E., Blanchette, M., Feiger, J.W., Adams, M.D., Rio, D.C. (2002) The KH-type RNA-binding protein PSI is required for Drosophila viability, male fertility, and cellular mRNA processing, *Genes Dev.* **16**(1), 72–84.

Labourier, E., Bourbon, H.M., Gallouzi, I.E., Fostier, M., Allemand, E., Tazi, J. (1999) Antagonism between RSF1 and SR proteins for both splice-site recognition in vitro and

Drosophila development, *Genes Dev.* **13**(6), 740–753.

Ladd, A.N., Charlet, N., Cooper, T.A. (2001) The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing, *Mol. Cell. Biol.* **21**(4), 1285–1296.

Laggerbauer, B., Achsel, T., Luhrmann, R. (1998) The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro, *Proc. Natl. Acad. Sci. U.S.A.* **95**(8), 4188–4192.

Lallena, M.J., Chalmers, K.J., Llamazares, S., Lamond, A.I., Valcarcel, J. (2002) Splicing regulation at the second catalytic step by Sex-lethal involves 3′ splice site recognition by SPF45, *Cell* **109**(3), 285–296.

Lander, E.S., et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Larriba, S., Bassa, L., Gimenez, J., Ramos, M.D., Segura, A., Nunes, V., Estivill, X., Casals, T. (1998) Testicular CFTR splice variants in patients with congenital absence of the vas deferens, *Hum. Mol. Genet.* **7**, 1739–1743.

Lavigueur, A., La Branche, H., Kornblihtt, A.R., Chabot, B. (1993) A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding, *Genes Dev.* **7**, 2405–2417.

Le Guiner, C., Lejeune, F., Galiana, D., Kister, L., Breathnach, R., Stevenin, J., Del Gatto-Konczak, F. (2001) TIA-1 and TIAR activate splicing of alternative exons with weak 5′ splice sites followed by a U-rich stretch on their own pre-mRNAs, *J. Biol. Chem.* **276**(44), 40638–40646.

Le Hir, H., Moore, M.J., Maquat, L.E. (2000) Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions, *Genes Dev.* **14**(9), 1098–1108.

Lim, S.R., Hertel, K.J. (2001) Modulation of survival motor neuron pre-mRNA splicing by inhibition of alternative 3′ splice site pairing, *J. Biol. Chem.* **276**(48), 45476–45483.

Lin, C.G., Bristol, L.A., Jin, L., Dykes-Hoberg, M., Crawford, T., Clawson, L., Rothstein, J.D. (1998) Aberrant RNA processing in a neurodegenerative disease: the cause for absent EAAT2, a glutamate transporter in amyotrophic lateral sclerosis, *Neuron* **20**, 589–602.

Lisbin, M.J., Qiu, J., White, K. (2001) The neuron-specific RNA-binding protein ELAV regulates neuroglian alternative splicing in neurons and binds directly to its pre-mRNA, *Genes Dev.* **15**(19), 2546–2561.

Liu, H.X., Chew, S.L., Cartegni, L., Zhang, M.Q., Krainer, A.R. (2000) Exonic splicing enhancer motif recognized by human SC35 under splicing conditions, *Mol. Cell. Biol.* **20**(3), 1063–1071.

Liu, X., Jiang, Q., Mansfield, S.G., Puttaraju, M., Zhang, Y., Zhou, W., Cohn, J.A., Garcia-Blanco, M.A., Mitchell, L.G., Engelhardt, J.F. (2002) Partial correction of endogenous DeltaF508 CFTR in human cystic fibrosis airway epithelia by spliceosome-mediated RNA trans-splicing, *Nat. Biotechnol.* **20**(1), 47–52.

Liu, W., Qian, C., Francke, U. (1997) Silent mutation induces exon skipping of fibrillin-1 gene in Marfan syndrome, *Nat. Genet.* **16**(4), 328–329.

Llewellyn, D.H., Scobie, G.A., Urquhart, A.J., Whatley, S.D., Roberts, A.G., Harrison, P.R., Elder, G.H. (1996) Acute intermittent porphyria caused by defective splicing of porphobilinogen deaminase RNA: a synonymous codon mutation at −22 bp from the 5′ splice site causes skipping of exon 3, *J. Med. Genet.* **33**(5), 437–438.

Lorson, C.L., Hahnene, E., Androphy, E.J., Wirth, B. (1999) A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy, *PNAS* **96**, 6307–6311.

Lou, H., Helfman, D.M., Gagel, R.F., Berget, S.M. (1999) Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3′-terminal exon, *Mol. Cell. Biol.* **19**(1), 78–85.

Lou, H., Yang, Y., Cote, G.J., Berget, S.M., Gagel, R.F. (1995) An intron enhancer containing a 5′ splice site sequence in the human calcitonin/calcitonin gene-related peptide gene, *Mol. Cell. Biol.* **15**(12), 7135–7142.

Lukas, J., Gao, D.Q., Keshmeshian, M., Wen, W.H., Tsao-Wei, D., Rosenberg, S., Press, M.F. (2001) Alternative and aberrant messenger RNA splicing of the mdm2 oncogene in invasive breast cancer, *Cancer. Res.* **61**(7), 3212–3219.

Lundquist, E.A., Herman, R.K., Rogalski, T.M., Mullen, G.P., Moerman, D.G., Shaw, J.E. (1996) The mec-8 gene of C. elegans encodes a protein with two RNA recognition motifs

and regulates alternative splicing of unc-52 transcripts, *Development* **122**(5), 1601–1610.

Lykke-Andersen, J., Shu, M.D., Steitz, J.A. (2001) Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1, *Science* **293**(5536), 1836–1839.

Lynch, K.W., Weiss, A. (2001) A CD45 polymorphism associated with multiple sclerosis disrupts an exonic splicing silencer, *J. Biol. Chem.* **276**, 24341–24347.

Makarova, O.V., Makarov, E.M., Liu, S., Vornlocher, H.P., Luhrmann, R. (2002a) Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing, *EMBO J.* **21**, 1148–1157.

Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., Luhrmann, R. (2002b) Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome, *Science* **298**, 2205–2208.

Mankodi, A., Takahashi, M.P., Jiang, H., Beck, C.L., Bowers, W.J., Moxley, R.T., Cannon, S.C., Thornton, C.A. (2002) Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy, *Mol. Cells* **10**(1), 35–44.

Mann, C.J., Honeyman, K., Cheng, A.J., Ly, T., Lloyd, F., Fletcher, S., Morgan, J.E., Partridge, T.A., Wilton, S.D. (2001) Antisense-induced exon skipping and synthesis of dystrophin in the mdx mouse, *Proc. Natl. Acad. Sci. U.S.A.* **98**(1), 42–47.

Markovtsov, V., Nikolic, J.M., Goldman, J.A., Turck, C.W., Chou, M.Y., Black, D.L. (2000) Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein, *Mol. Cell. Biol.* **20**(20), 7463–7479.

Maruyama, T., Miyake, Y., Yamamura, T., Tajima, S., Funahashi, T., Matsuzawa, Y., Yamamoto, A. (1998) A novel point mutation in a splice acceptor site of intron 1 of the human low density lipoprotein receptor gene which causes severe hypercholesterolemia: an unexpected absence of exon skipping, *Hum. Mutat.* **11**(6), 480–481.

McGrory, J., Cole, W.G. (1999) Alternative splicing of exon 37 of FBN1 deletes part of an 'eight-cysteine' domain resulting in the Marfan syndrome, *Clin. Genet.* **55**(2), 118–121.

Matsumura, T., Osaka, H., Sugiyama, N., Kawanishi, C., Maruyama, Y., Suzuki, K., Onishi, H., Yamada, Y., Morita, M., Aoki, M., Kosaka, K. (1998) Novel acceptor splice site mutation in the invariant AG of intron 6 of alpha-galactosidase A gene, causing Fabry disease, *Hum. Mutat.* **11**, 483.

Matter, N., Marx, M., Weg-Remers, S., Ponta, H., Herrlich, P., König, H. (2000) Heterogeneous ribonucleoprotein A1 is part of an exon-specific splice-silencing complex controlled by oncogenic signaling pathways, *J. Biol. Chem.* **275**, 35353–35360.

Mayeda, A., Krainer, A.R. (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2, *Cell* **68**(2), 365–375.

McCullough, A.J., Berget, S.M. (2000) An intronic splicing enhancer binds U1 snRNPs to enhance splicing and select 5′ splice sites, *Mol. Cell. Biol.* **20**(24), 9225–9235.

McGrory, J., Cole, W.G. (1999) Alternative splicing of exon 37 of FBN1 deletes part of an 'eight-cysteine' domain resulting in the Marfan syndrome, *Clin. Genet.* **55**(2), 118–121.

McKie, A.B., McHale, J.C., Keen, T.J., Tarttelin, E.E., Goliath, R., Van Lith-Verhoeven, J.J., Greenberg, J., Ramesar, R.S., Hoyng, C.B., Cremers, F.P., MacKey, D.A., Bhattacharya, S.S., Bird, A.C., Markham, A.F., Inglehearn, C.F. (2001) Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13), *Hum. Mol. Genet.* **10**, 1555–1562.

Mercatante, D., Kole, R. (2000) Modification of alternative splicing pathways as a potential approach to chemotherapy, *Pharmacol. Ther.* **85**, 237–243.

Merendino, L., Guth, S., Bilbao, D., Martinez, C., Valcarcel, J. (1999) Inhibition of msl-2 splicing by sex-lethal reveals interaction between U2AF35 and the 3′ splice site AG, *Nature* **402**(6763), 838–841.

Min, H., Turck, C.W., Nikolic, J.M., Black, D.L. (1997) A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer, *Genes Dev.* **11**(8), 1023–1036.

Min, H., Chan, R.C., Black, D.L. (1995) The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event, *Genes Dev.* **9**(21), 2659–2671.

Modafferi, E.F., Black, D.L. (1997) A complex intronic splicing enhancer from the c-src pre-mRNA activates inclusion of a heterologous exon, *Mol. Cell. Biol.* **17**(11), 6537–6545.

Moore, M.J., Sharp, P.A. (1993) Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing, *Nature* **365**(6444), 364–368.

Nabholtz, J.M., Reese, D.M., Lindsay, M.A., Riva, A. (2002) HER2-positive breast cancer: update on breast cancer international research group trials, *Clin. Breast Cancer* **3**(Suppl. 2), S75–S79.

Nelson, K.K., Green, M.R. (1990) Mechanism for cryptic splice site activation during pre-mRNA splicing, *Proc. Natl. Acad. Sci. U.S.A.* **87**(16), 6253–6257.

Nesic, D., Kramer, A. (2001) Domains in human splicing factors SF3a60 and SF3a66 required for binding to SF3a120, assembly of the 17S U2 snRNP, and prespliceosome formation, *Mol. Cell. Biol.* (19), 6406–6417.

Nissim-Rafinia, M., Krem, B. (2002) Splicing regulation as a potential genetic modifier, *Trends Genet.* **18**, 123–127.

Nicoll, M., Akerib, C.C., Meyer, B.J. (1997) X-chromosome-counting mechanisms that determine nematode sex, *Nature* **388**(6638), 200–204.

Nguyen, V.N., Mirejovsky, T., Melinova, L., Mandys, V. (2000) CD44 and its v6 spliced variant in lung carcinomas: relation to NCAM, CEA, EMA and UP1 and prognostic significance, *Neoplasma* **47**(6), 400–408.

Ono, Y., Ohno, M., Shimura, Y. (1994) Identification of a putative RNA helicase (HRH1), a human homolog of yeast Prp22, *Mol. Cell. Biol.* **14**(11), 7611–7620.

Ono, R., Taki, T., Taketani, T., Kawaguchi, H., Taniwaki, M., Okamura, T., Kawa, K., Hanada, R., Kobayashi, M., Hayashi, Y. (2002) SEPTIN6, a human homologue to mouse Septin6, is fused to MLL in infant acute myeloid leukemia with complex chromosomal abnormalities involving 11q23 and Xq24, *Cancer Res.* **62**(2), 333–337.

Oshika, Y., Nakamura, M., Tokunaga, T., Ohnishi, Y., Abe, Y., Tsuchida, T., Tomii, Y., Kijima, H., Yamazaki, H., Ozeki, Y., Tamaoki, N., Ueyama, Y. (2000) Ribozyme approach to downregulate vascular endothelial growth factor (VEGF) 189 expression in non-small cell lung cancer (NSCLC), *Eur. J. Cancer* **36**(18), 2390–2396.

Pagani, F., Buratti, E., Stuani, C., Bendix, R., Dork, T., Baralle, F.E. (2002) A new type of mutation causes a splicing defect in ATM, *Nat. Genet.* **30**, 426–429.

Perez-Tur, J., Froelich, S., Prihar, G., Crook, R., Baker, M., Duff, K., Wragg, M., Busfield, F., Lendon, C., Clark, R.F., et al. (1995) A mutation in Alzheimer's disease destroying a splice acceptor site in the presenilin-1 gene, *NeuroReport* **7**, 297–301.

Petersen-Mahrt, S.K., Estmer, C., Ohrmalm, C., Matthews, D.A., Russell, W.C., Akusjarvi, G. (1999) The splicing factor-associated protein, p32, regulates RNA splicing by inhibiting ASF/SF2 RNA binding and phosphorylation, *EMBO J.* **18**(4), 1014–1024.

Pinto, A.L., Steitz, J.A. (1989) The mammalian analogue of the yeast PRP8 splicing protein is present in the U4/5/6 small nuclear ribonucleoprotein particle and the spliceosome, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8742–8746.

Ploos Van Amstel, J.K., Bergman, A.J., Van Beurden, E.A., Roijers, J.F., Peelen, T., Van Den Berg, I.E., Poll-The, B.T., Kvittingen, E.A., Berger, R. (1996) Hereditary tyrosinemia type 1: novel missense, nonsense and splice consensus mutations in the human fumarylacetoacetate hydrolase gene; variability of the genotype-phenotype relationship, *Hum. Genet.* **97**(1), 51–59.

Poola, I., Speirs, V. (2001) Expression of alternatively spliced estrogen receptor alpha mRNAs is increased in breast cancer tissues, *J. Steroid Biochem. Mol. Biol.* **78**, 459–469.

Poltora, Z., Cohen, T., Neufeld, G. (2000) The VEGF splice variants: properties, receptors, and usage for the treatment of ischemic diseases, *Herz* **25**, 126–129.

Puig, O., Gottschalk, A., Fabrizio, P., Seraphin, B. (1999) Interaction of the U1 snRNP with nonconserved intronic sequences affects 5′ splice site selection, *Genes Dev.* **13**(5), 569–580.

Puttaraju, M., Dipasquale, J., Baker, C.C., Mitchell, L.G., Garcia-Blanco, M.A. (2001) Messenger RNA repair and restoration of protein function by spliceosome-mediated RNA trans-splicing, *Mol. Ther.* **4**(2), 105–114.

Query, C.C., Bentley, R.C., Keene, J.D. (1989) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein, *Cell* **57**(1), 89–101.

Rappsilber, J., Ajuh, P., Lamond, A.I., Mann, M. (2001) SPF30 is an essential human splicing factor required for assembly of the U4/U5/U6 tri-small nuclear ribonucleoprotein into the spliceosome, *J. Biol. Chem.* **276**(33), 31142–31150.

Rappsilber, J., Ryder, U., Lamond, A.I., Mann, M. (2002) Large-scale proteomic analysis of the human spliceosome, *Genome Res.* **12**(8), 1231–1245.

Richard, M.M., Erenberg, G., Triggs-Raine, B.L. (1995) An A-to-G mutation at the +3 position of intron 8 of the HEXA gene is associated with exon 8 skipping and Tay-Sachs disease, *Biochem. Mol. Med.* **55**(1), 74–76.

Rogan, P.K., Faux, B.M., Schneider, T.D. (1998) Information analysis of human splice site mutations, *Hum. Mutat.* **12**, 153–171.

Ruskin, B., Zamore, P.D., Green, M.R. (1988) A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly, *Cell* **52**(2), 207–219.

Reyes, J.L., Gustafson, E.H., Luo, H.R., Moore, M.J., Konarska, M.M. (1999) The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5′ splice site.å, *RNA* **5**(2), 167–179.

Sanford, J.R., Bruzik, J.P. (1999) Developmental regulation of SR protein phosphorylation and activity, *Genes Dev.* **13**(12), 1513–1518.

Sato, N., Imaizumi, K., Manabe, T., Tanigushi, M., Hitomi, J., Katayama, T., Takunari, Y., Morihara, T., Yasuda, Y., Takagi, T., Kudo, T., Tsudo, T., Itoyama, Y., Makifuchi, T., Fraser, P.E., St George-Hyslop, P., Tohyama, M. (2001) Increased production of beta amyloid and vulnerability to endoplasmic reticulum stress by an aberrant spliced form of presenilin 2, *J. Biol. Chem.* **276**, 2108–2114.

Savkur, R.S., Philips, A.V., Cooper, T.A. (2001) Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy, *Nat. Genet.* **29**(1), 40–47.

Seifert, G., Schroder, W., Hinterkeuser, S., Schumacher, T., Schramm, J., Steinhauser, C. (2002) Changes in flip/flop splicing of astroglial AMPA receptors in human temporal lobe epilepsy, *Epilepsia* **43**(Suppl. 5), 162–167.

Shin, C., Manley, J.L. (2002) The SR protein SRp38 represses splicing in M phase cells, *Cell* **111**(3), 407–417.

Siatecka, M., Reyes, J.L., Konarska, M.M. (1999) Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center, *Genes Dev.* **13**(15), 1983–1993.

Siebel, C.W., Kanaar, R., Rio, D.C. (1994) Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI, *Genes Dev.* **8**(14), 1713–1725.

Siffert, W., Rosskopf, D., Siffert, G., Busch, S., Moritz, A., Erbel, R., Sharma, A.M., Ritz, E., Wichmann, H.E., Jakobs, K.H. (1998) Horsthemke B association of a human G-protein beta3 subunit variant with hypertension, *Nat. Genet.* **18**(1), 45–48.

Sigalas, I., Calvert, A.H., Anderson, J.J., Neal, D.E., Lunec, J. (1996) Alternatively spliced mdm2 transcripts with loss of p53 binding domain sequences: transforming ability and frequent detection in human cancer, *Nat. Med.* **2**(8), 912–917.

Simard, M.J., Chabot, B. (2002) SRp30c is a repressor of 3′ splice site utilization, *Mol. Cell. Biol.* **22**(12), 4001–4010.

Singh, R., Valcarcel, J., Green, M.R. (1995) Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins, *Science* **268**(5214), 1173–1176.

Sontheimer, E.J., Steitz, J.A. (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome, *Science* **262**(5142), 1989–1996.

Sontheimer, E.J., Sun, S., Piccirilli, J.A. (1997) Metal ion catalysis during splicing of premessenger RNA, *Nature* **388**(6644), 801–805.

Spike, C.A., Davies, A.G., Shaw, J.E., Herman, R.K. (2002) MEC-8 regulates alternative splicing of unc-52 transcripts in C. elegans hypodermal cells, *Development* **129**(21), 4999–5008.

Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., Abelson, J. (2002) Composition and functional characterization of the yeast spliceosomal penta-snRNP, *Mol. Cell.* **9**(1), 31–44.

Sun, H., Chasin, L.A. (2000) Multiple splicing defects in an intronic false exon, *Mol. Cell. Biol.* **20**, 6414–6425.

Staley, J.P., Guthrie, C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things, *Cell* **92**(3), 315–326.

Staley, J.P., Guthrie, C. (1999) An RNA switch at the 5′ splice site requires ATP and the DEAD box protein Prp28p, *Mol. Cell.* **3**(1), 55–64.

Stamm, S., Zhu, J., Nakai, K., Stoilov, P., Stoss, O., Zhang, M.Q. (2000) An alternative-exon database and its statistical analysis, *DNA Cell. Biol.* **19**(12), 739–756.

Stark, J.M., Cooper, T.A., Roth, M.B. (1999) The relative strengths of SR protein-mediated associations of alternative and constitutive exons can influence alternative splicing, *J. Biol. Chem.* **274**(42), 29838–29842.

Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., Abelson, J. (2002) Composition and functional characterization of the yeast spliceosomal penta-snRNP, *Mol. Cell.* **9**(1), 31–44.

Stickeler, E., Fraser, S.D., Honig, A., Chen, A.L., Berget, S.M., Cooper, T.A. (2001) The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4, *EMBO J.* **20**(14), 3821–3830.

Stickeler, E., Kittrell, F., Medina, D., Berget, S.M. (1999) Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis, *Oncogene* **18**(24), 3574–3582.

Stoss, O., Olbrich, M., Hartmann, A.M., Konig, H., Memmott, J., Andreadis, A., Stamm, S. (2001) The STAR/GSG family protein rSLM-2 regulates the selection of alternative splice sites, *J. Biol. Chem.* **276**(12), 8665–8673.

Sun, H., Chasin, L.A. (2000) Multiple splicing defects in an intronic false exon, *Mol. Cell. Biol.* **20**(17), 6414–6425.

Tange, T.O., Damgaard, C.K., Guth, S., Valcarcel, J., Kjems, J. (2001) The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element, *EMBO J.* **20**(20), 5748–5758.

Tarn, W.Y., Steitz, J.A. (1996a) A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro, *Cell* **84**(5), 801–811.

Tarn, W.Y., Steitz, J.A. (1996b) Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns, *Science* **273**(5283), 1824–1832.

Teigelkamp, S., Mundt, C., Achsel, T., Will, C.L., Luhrmann, R. (1997) The human U5 snRNP-specific 100-kD protein is an RS domain-containing, putative RNA helicase with significant homology to the yeast splicing factor Prp28p, *RNA* **3**(11), 1313–1326.

Todd, R.D., Lobos, E.A., Sun, L.W., Neuman, R.J. (2003) Mutational analysis of the nicotinic acetylcholine receptor alpha 4 subunit gene in attention deficit/hyperactivity disorder: evidence for association of an intronic polymorphism with attention problems, *Mol. Psychiatry* **8**(1), 103–108.

Tysoe, C., Whittaker, J., Xuereb, J., Cairns, N.J., Cruts, M., Vanbroeckhoven, C., Wilcock, G., Rubinsztein, D.C. (1998) A presenilin-1 truncating mutation is present in two cases with autopsy-confirmed early-onset Alzheimer disease, *Am. J. Hum. Genet.* **62**, 70–76.

Vacek, M.M., Ma, H., Gemignani, F., Lacerra, G., Kafri, T., Kole, R. (2003) High-level expression of hemoglobin A in human thalassemic erythroid progenitor cells following lentiviral vector delivery of an antisense snRNA, *Blood* **101**(1), 104–111.

Valadkhan, S., Manley, J.L. (2001) Splicing-related catalysis by protein-free snRNAs, *Nature* **413**(6857), 701–707.

Valcarcel, J., Singh, R., Zamore, P.D., Green, M.R. (1993) The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA, *Nature* **362**(6416), 171–175.

Van Buskirk, C., Schupbach, T. (2002) Half pint regulates alternative splice site selection in Drosophila, *Dev. Cell.* **2**(3), 343–353.

Van Lith-Verhoeven, J.J., Van Der Velde-Visser, S.D., Sohocki, M.M., Deutman, A.F., Brink, H.M., Cremers, F.P., Hoyng, C.B. (2002) Clinical characterization, linkage analysis, and PRPC8 mutation analysis of a family with autosomal dominant retinitis pigmentosa type 13 (RP13), *Ophthalmic Genet.* **23**, 1–12.

Vithana, E., Al-Maghtheh, M., Bhattacharya, S.S., Inglehearn, C.F. (1998) RP11 is the second most common locus for dominant retinitis pigmentosa, *J. Med. Genet.* **35**, 174–175.

Vithana, E.N., Abu-Safieh, L., Allen, M.J., Carey, A., Papaioannou, M., Chakarova, C., Al-Maghtheh, M., Ebenezer, N.D., Willis, C., Moore, A.T., Bird, A.C., Hunt, D.M., Bhattacharya, S.S. (2001) A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11), *Mol. Cell.* **8**, 375–381.

Vorgerd, M., Burwinkel, B., Reichmann, H., Malin, J.P., Kilimann, M.W. (1998) Adult-onset glycogen disease type II: phenotypic and allelic heterogeneity in German patients, *Neurogenetics* **1**, 205–211.

Wagner, E.J., Garcia-Blanco, M.A. (2002) RNAi-mediated PTB depletion leads to enhanced exon definition, *Mol. Cells* **10**(4), 943–949.

Watanabe, T., Sullenger, B.A. (2000) Induction of wild-type p53 activity in human cancer cells by ribozymes that repair mutant p53 transcripts, *Proc. Natl. Acad. Sci. U.S.A.* **97**(15), 8490–8494.

Wang, H.Y., Lin, W., Dyck, J.A., Yeakley, J.M., Songyang, Z., Cantley, L.C., Fu, X.D. (1998) SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells, *J. Cell. Biol.* **140**(4), 737–750.

Wang, J., Takagaki, Y., Manley, J.L. (1996) Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability, *Genes Dev.* **10**(20), 2588–2599.

Watanabe, T., Sullenger, B.A. (2000) Induction of wild-type p53 activity in human cancer cells by ribozymes that repair mutant p53 transcripts, *Proc. Natl. Acad. Sci. U.S.A.* **97**(15), 8490–8494.

Weidenhammer, E.M., Ruiz-Noriega, M., Woolford, J.L. Jr. (1997) Prp31p promotes the association of the U4/U6 × U5 tri-snRNP with prespliceosomes to form spliceosomes in Saccharomyces cerevisiae, *Mol. Cell. Biol.* **17**, 3580–3588.

Weidenhammer, E.M., Singh, M., Ruiz-Noriega, M., Woolford, J.L. Jr. (1996) The PRP31 gene encodes a novel protein required for pre-mRNA splicing in Saccharomyces cerevisiae, *Nucleic Acids Res.* **24**, 164–170.

Will, C.L., Schneider, C., MacMillan, A.M., Katopodis, N.F., Neubauer, G., Wilm, M., Luhrmann, R., Query, C.C. (2001) A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site, *EMBO J.* **20**(16), 4536–4546.

Will, C.L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M., Luhrmann, R. (2002) Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein, *EMBO J.* **21**(18), 4978–4988.

Wimmel, A., Schilli, M., Kaiser, U., Havemann, K., Ramaswamy, A., Branscheid, D., Kogan, E., Schuermann, M. (1997) Preferential histiotypic expression of CD44-isoforms in human lung cancer, *Lung Cancer* **16**(2–3), 151–172.

Wu, J.Y., Maniatis, T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing, *Cell* **75**, 1061–1070.

Wu, J.A., Manley, J.L. (1991) Base pairing between U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA, *Nature* **352**(6338), 818–821.

Wu, J.I., Reed, R.B., Grabowski, P.J., Artzt, K. (2002) Function of quaking in myelination: regulation of alternative splicing, *Proc. Natl. Acad. Sci. U.S.A.* **99**(7), 4233–4238.

Wu, S., Romfo, C.M., Nilsen, T.W., Green, M.R. (1999) Functional recognition of the 3′ splice site AG by the splicing factor U2AF35, *Nature* **402**(6763), 832–835.

Xie, J., Black, D.L. (2001) A CaMK IV responsive RNA element mediates depolarization-induced alternative splicing of ion channels, *Nature* **410**(6831), 936–939.

Xie, J., McCobb, D.P. (1998) Control of alternative splicing of potassium channels by stress hormones, *Science* **280**(5362), 443–446.

Yang, L., Embree, L.J., Hickstein, D.D. (2000) TLS-ERG leukemia fusion protein inhibits RNA splicing mediated by serine-arginine proteins, *Mol. Cell. Biol.* **20**(10), 3345–3354.

Yang, L., Embree, L.J., Tsai, S., Hickstein, D.D. (1998) Oncoprotein TLS interacts with serine-arginine proteins involved in RNA splicing, *J. Biol. Chem.* **273**(43), 27761–27764.

Yang, F., Hanson, N.Q., Schwichtenberg, K., Tsai, M.Y. (2000) Variable number tandem repeat in exon/intron border of the cystathionine beta-synthase gene: a single nucleotide substitution in the second repeat prevents multiple alternate splicing, *Am. J. Med. Genet.* **95**, 385–390.

Yang, X.L., Miura, N., Kawarada, Y., Terada, K., Petrukhin, K., Gilliam, C., Sugiyama, T. (1997) Two forma of Wilson disease protein produced by alternative splicing are localised in distinct cellular compartments, *Biochem. J.* **326**, 897–902.

Yazaki, M., Yoshida, K., Nakamura, A., Furihata, K., Yonekawa, M., Okabe, T., Yamashita, N., Ohta, M., Ikeda, S. (1998) A novel splicing mutation in the ceruloplasmin gene rsponsible for hereditary ceruloplasmin deficiency with hemosiderosis, *J. Neurol. Sci.* **156**, 30–34.

Yeakley, J.M., Fan, J.B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M.S., Fu, X.D.

(2002) Profiling alternative splicing on fiber-optic arrays, *Nat. Biotechnol.* **20**(4), 353–358.

Yeakley, J.M., Tronchere, H., Olesen, J., Dyck, J.A., Wang, H.Y., Fu, X.D. (1999) Phosphorylation regulates in vivo interaction and molecular targeting of serine/arginine-rich pre-mRNA splicing factors, *J. Cell. Biol.* **145**(3), 447–455.

Yean, S.L., Wuenschell, G., Termini, J., Lin, R.J. (2000) Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome, *Nature* **408**(6814), 881–884.

Yu, L., Heere-Ress, E., Boucher, B., Defesche, J.C., Kastelein, J., Lavoie, M.A., Genest, J. (1999) Familial hypercholesterolemia. Acceptor splice site (G→C) mutation in intron 7 of the LDL-R gene: alternate RNA editing causes exon 8 skipping or a premature stop codon in exon 8, *Atherosclerosis* **146**, 125–131.

Zachar, Z., Chou, T.B., Kramer, J., Mims, I.P., Bingham, P.M. (1994) Analysis of autoregulation at the level of pre-mRNA splicing of the suppressor-of-white-apricot gene in Drosophila, *Genetics* **137**(1), 139–135.

Zahler, A.M., Lane, W.S., Stolk, J.A., Roth, M.B. (1992) SR proteins: a conserved family of pre-mRNA splicing factors, *Genes Dev.* **6**(5), 837–847.

Zahler, A.M., Neugebauer, K.M., Lane, W.S., Roth, M.B. (1993) Distinct functions of SR proteins in alternative pre-mRNA splicing, *Science* **260**(5105), 219–222.

Zamore, P.D., Patton, J.G., Green, M.R. (1992) Cloning and domain structure of the mammalian splicing factor U2AF, *Nature* **355**(6361), 609–614.

Zapp, M.L., Stern, S., Green, M.R. (1993) Small molecules that selectively block RNA binding of HIV-1 Rev protein inhibit Rev function and viral production, *Cell* **74**(6), 969–978.

Zhang, L., Liu, W., Grabowski, P.J. (1999) Coordinate repression of a trio of neuron-specific splicing events by the splicing regulator PTB, *RNA* **5**(1), 117–130.

Zhang, W., Liu, H., Han, K., Grabowski, P.J. (2002) Region-specific alternative splicing in the nervous system: implications for regulation by the RNA-binding protein NAPOR, *RNA* **8**(5), 671–685.

Zhang, W.J., Wu, J.Y. (1996) Functional properties of p54, a novel SR protein active in constitutive and alternative splicing, *Mol. Cell. Biol.* **16**(10), 5400–5408.

Zheng, Z.M., Huynen, M., Baker, C.C. (1998) A pyrimidine-rich exonic splicing suppressor binds multiple RNA splicing factors and inhibits spliceosome assembly, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14088–14093.

Zhou, Z., Licklider, L.J., Gygi, S.P., Reed, R. (2002) Comprehensive proteomic analysis of the human spliceosome, *Nature* **419**(6903), 182–185.

Zhu, J., Mayeda, A., Krainer, A.R. (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins, *Mol. Cells* **8**(6), 1351–1361.

Alzheimer's Disease

Jun Wang, Silva Hecimovic and Alison Goate
Department of Psychiatry, Washington University School of Medicine,
St. Louis, MO, USA

1 Pathology and Epidemiology of Alzheimer's Disease 182

2 The Genetics of Alzheimer's Disease 184
2.1 The Genetics of Familial Alzheimer's Disease 184
2.1.1 β-amyloid Precursor Protein Mutations 184
2.1.2 Presenilin Mutations 185
2.2 The Genetics of Sporadic Alzheimer's Disease 187
2.2.1 Apolipoprotein E 187
2.2.2 The Search for New Alzheimer's Disease Susceptibility Genes 188

3 Molecular and Cellular Mechanisms Underlying Alzheimer's Disease 189
3.1 β-amyloid Precursor Protein 189
3.2 Production of Aβ–Proteolytic Processing of APP 189
3.3 β-secretase – BACE1 190
3.4 γ-secretase – Presenilins and Cofactors 191
3.4.1 Presenilins 191
3.4.2 Cofactors 192
3.5 Neurofibrillary Tangles and Tau Protein 193
3.6 Apolipoprotein E 194

4 Transgenic Mouse Models of Alzheimer's Disease 195
4.1 APP Transgenic Mice 195
4.2 Presenilin 1/2 Transgenic Mice 196
4.3 Tau Transgenic Mice 197
4.4 Apolipoprotein E Transgenic Mice 198
4.5 Modulating and Enhancing the Phenotype: Transgenic Crosses 198

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

5 Potential Treatments 199
5.1 Secretase Inhibitors 200
5.2 Vaccine Approaches 201
5.3 Other Therapeutic Approaches 201

6 Summary 202

Bibliography 202
Books and Reviews 202
Primary Literature 202

Keywords

Association Study
Tests for differences in allele or genotype frequencies between unrelated, affected, and unaffected individuals.

Autosomal Dominant Trait
A phenotype that is expressed in a heterozygote, which maps to a chromosome other than the sex chromosomes.

β-secretase
Also known as BACE1; a transmembrane aspartyl protease involved in processing of APP-generating N-terminal end of Aβ peptides.

Candidate Gene Study
An association study performed on a particular gene(s) selected according to the biology of the disease and/or their location within chromosome regions found to be implicated in disease (by linkage or association study).

Full Genome Scan
A linkage or association analysis performed between the disease and polymorphic markers that are spaced throughout the genome.

γ-secretase
A multiprotein complex involved in intramembranous cleavage of APP-generating C-terminal end of Aβ peptides. It contains four proteins: the catalytic component (presenilins) and three cofactors (nicastrin, APH-1, and PEN-2).

Neurofibrillary Tangles (NFTs)
Abnormal intraneuronal deposits abundant in the brains of patients with AD and some other neurodegenerative diseases. The main constituent of NFTs is hyperphosphorylated tau protein.

Senile Plaques
Complex extracellular deposits abundant in the brains of AD patients. The main proteinaceous component of plaques is amyloid-β peptide.

Susceptibility Locus/Gene
A region of a chromosome or a specific gene that may influence the risk to a certain condition/disease.

Familial Alzheimer's Disease
Transmitted as an autosomal dominant trait; usually has an age of onset <60 years.

Sporadic Alzheimer's Disease
Shows modest familial clustering and probably results from the synergistic action of genetic and environmental factors. Usually has an age of onset >60 years.

Abbreviations

Aβ	β-amyloid peptide
AD	Alzheimer's disease
AICD	APP Intracellular domain
APLP	amyloid precursor-like protein
APOE	apolipoprotein E
APP	β-amyloid precursor protein
BBB	blood brain barrier
CAA	cerebral amyloid angiopathy
CNS	central nervous system
CTF	C-terminal fragment
FAD	familial Alzheimer's disease
FTDP-17	frontotemporal dementia with parkinsonism linked to chromosome 17
KPI	Kunitz-type protease inhibitor
MT	microtubule
NFT	neurofibrillary tangles
NSAID	nonsteroidal anti-inflammatory drugs
NTF	N-terminal fragment
PDGF	platelet-derived growth factor
PHF	paired helical filaments
PS	presenilin
TM	transmembrane

Alzheimer's disease (AD) is the most common cause of dementia in the elderly. It is characterized by a progressive loss of memory, reasoning, judgment, and orientation, by the presence of large numbers of extracellular senile plaques and intracellular neurofibrillary tangles, as well as substantial cell loss in the brain. Genetic studies in early onset families have identified mutations in three genes that cause AD, while genetic studies in sporadic AD have identified the apolipoprotein E4 allele as a risk factor for disease. *In vitro* cell biology and transgenic mouse studies implicate all of these genes in the biology of Aβ metabolism. A consistent effect of these genetic factors is an increase in Aβ deposition. Furthermore, cell biology studies have identified targets for possible treatment of AD. Transgenic models are now being used to test the validity of these targets and therapeutic approaches to the treatment of AD.

1
Pathology and Epidemiology of Alzheimer's Disease

The German psychiatrist Alois Alzheimer first described the clinical and pathological features of Alzheimer's disease (AD) in a 55-year-old female patient; he published the discovery as a case report in 1907. It was thought that the "presenile dementia" described by Alzheimer was distinct from "senile dementia" seen in older patients, but it is now widely accepted that the two conditions are the same disease, differing primarily in the age of onset.

Alzheimer's disease is the most common cause of dementia in the western world, and the fourth leading cause of deaths in the United States. The disease is confined mainly to aged individuals and is progressive. It affects 1% of people in the developed nations and is likely to become a major problem in developing countries as the proportion of elderly persons increases in these populations. The incidence of AD rises with increasing age: 1 to 5% of people aged 65 years are affected, while 10 to 20% of those over the age of 80 have AD.

The first clinical manifestation is usually a loss of short-term memory; this is followed, over the next 6 to 20 years, by a progressive loss of memory, reasoning, judgment, and orientation (dementia). Death occurs, on average, about 10 years after the onset of symptoms. Many epidemiological surveys have been carried out to identify risk factors. There is a clear correlation between AD and a positive family history in about one-third of cases; the risk of AD is also elevated in individuals with Down syndrome. Severe head trauma with loss of consciousness increases risk for AD, particularly in those individuals who also have an apolipoprotein E4 allele. In contrast, the use of nonsteroidal anti-inflammatory agents or cholesterol-lowering agents, such as the statins, may reduce risk for AD. Another factor that may potentially decrease risk for AD is the years of education; the more the years of education, the lower the risk of disease.

Alzheimer's disease is characterized neuropathologically by the presence of large numbers of neuritic plaques and neurofibrillary tangles (NFT) within the brain cortex (Fig. 1). In addition, there is massive

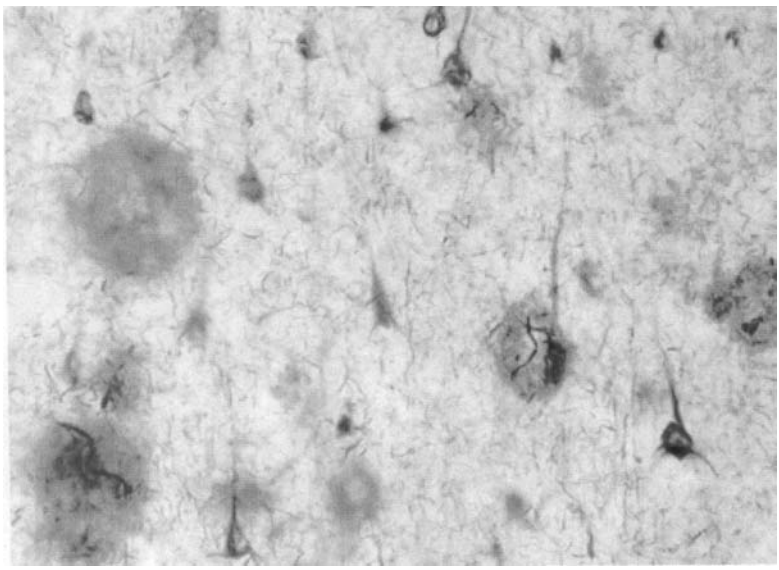

Fig. 1 Plaques and tangles in an AD brain. Shown is a brain section from an individual with mild AD. The section was double-stained with antibody PHF-1 specific to tangles and 10D5 specific to β-amyloid. Courtesy of Dr. Joel Price.

neuronal cell loss, and amyloid deposits that occur in the walls of the blood vessels of the meninges and cerebral cortex. Neither plaques nor tangles are specific to AD – they occur in intellectually normal, elderly people and in certain other diseases. However, they are more numerous and more widely distributed throughout the brains of Alzheimer's patients. Senile plaques are extracellular deposits, while tangles initially appear within the degenerating neurons. The major proteinaceous constituent of plaques is a 39 to 43 amino acid peptide called β-amyloid, or Aβ. The microtubule-associated protein, tau, is a major constituent of tangles. In the AD brain, there is extensive activation of astrocytes and microglia, the two primary mediators of inflammation in the central nervous system (CNS). A large number of inflammatory mediators and complement components are elevated in the AD brain, and colocalized with both plaques and tangles. The inflammatory response might be induced by increased Aβ levels or by plaque formation.

The deposition of Aβ has also been observed in association with several other pathological conditions. Individuals with Down syndrome, who live into their late thirties and beyond, develop a neuropathology indistinguishable from that of AD. Temporal studies in Down syndrome suggest that Aβ deposition is the earliest of the known pathological markers to appear, and that neurofibrillary tangles and neuronal cell loss occur decades after the first Aβ deposition. However, dementia correlates more closely with neurofibrillary tangles and neuron loss than with plaques. According to one hypothesis based on these observations, amyloid deposition precedes and induces tangle formation; tangles (and perhaps plaques) cause neuronal death, and the resultant neuronal damage destroys important

neurotransmission pathways in the brain, producing dementia.

2 The Genetics of Alzheimer's Disease

There are two forms of AD: familial Alzheimer's disease (FAD), in which the disease is transmitted as an autosomal dominant trait; and sporadic AD, which shows modest familial clustering and probably results from the synergistic action of genetic and environmental factors. Sporadic AD can have an early (<60 yrs) or late (>60 yrs) age of onset. FAD does not appear to be clinically or neuropathologically different from the more common sporadic form of AD, except that it generally has an earlier age of onset.

2.1 The Genetics of Familial Alzheimer's Disease

The existence of families in which AD segregates as a fully penetrant, autosomal, dominant trait presents the most striking evidence for the involvement of genetic factors in the etiology of AD. Mutations in three distinct genes, β-amyloid precursor protein (APP), *PS1* (presenilin 1) and *PS2* (presenilin 2) have been identified to cause FAD.

2.1.1 β-amyloid Precursor Protein Mutations

The main component of the AD-associated senile plaques, Aβ, is derived by the proteolytic processing of APP. The gene coding for APP is located on the long arm of chromosome 21. It contains 18 exons, three of which can be alternatively spliced to produce a variety of mRNA species.

The first mutation shown to cause FAD was a valine to isoleucine substitution at residue 717 of APP (Fig. 2). To date, 22 APP mutations have been reported, of which 17 are linked to early-onset AD (www.alzforum.org). Most mutations lie within the transmembrane (TM) domain of APP in the region that is involved in the generation of the C-terminal end of Aβ. All TM domain FAD-APP mutations result in an increased production of the longer, more amyloidogenic form of Aβ (Aβ42/43). The most severe mutation (T714I) causes a 11-fold increase in the Aβ42/Aβ40 ratio; it causes both a decrease in Aβ40 and an increase in Aβ42 production. Patients carrying the T714I FAD-APP mutation die in their 40s. FAD mutations in APP can differentially affect Aβ-peptide formation by either increasing Aβ42 or decreasing Aβ40, or both. The Swedish mutation (K670N, M671L) lies in the region outside of the TM domain in which cleavage by β-secretase generates the N-terminal end of the Aβ, resulting in an increase in both Aβ40 and Aβ42 levels. Five mutations are located within the Aβ sequence but outside the TM domain: Flemish (A692G), Dutch (E693Q), Arctic (E693G), Italian (E693K), and Iowa (D694N). Patients carrying these mutations deposit Aβ in senile plaques and/or in cerebral vascular walls (cerebral amyloid angiopathy – CAA) and present clinically with either AD or hemorrhagic strokes, or both. For example, Flemish A692G patients present with both AD and strokes, while most Dutch E693Q patients suffer from strokes followed by a progressive multi-infarct dementia (HCHWA-D, hereditary cerebral hemorrhage with amyloidosis Dutch type). Pathologically, Flemish A692G patients demonstrate a classical AD pathology although with a strong CAA component, while Dutch APP patients

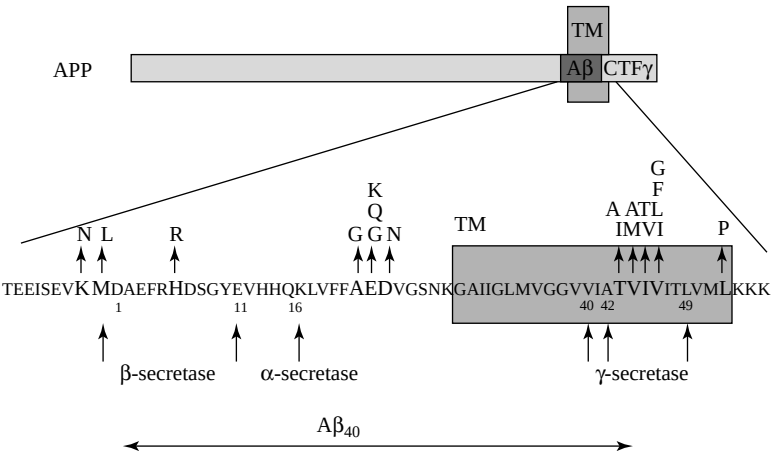

Fig. 2 FAD mutations in APP. Shown is a schematic representation of the APP protein showing the regions coding for Aβ peptide and C-terminal intracellular domain-CTFγ/AICD. The location of each FAD mutation is shown. Letters in bold are FAD-mutation sites, letters in gray are FAD substitutions. Major sites of APP processing, α-, β-, and γ-secretase sites, are also labeled. The amino acid numbering below the letters is according to the Aβ sequence.

have mainly vascular Aβ in the absence of neurofibrillary pathology. Recent analysis of pathogenic effect on Aβ levels of Arctic, Italian, Dutch, and Flemish APP mutations revealed that all, except the Flemish mutation, cause a decrease in both Aβ40 and Aβ42. In contrast, the Flemish mutation increases both Aβ40 and Aβ42 levels.

2.1.2 Presenilin Mutations

The discovery that mutations within the presenilin 1 *(PS1)* gene located on chromosome 14 (14q24.3) were responsible for most FAD cases opened up a whole new area of research into the molecular mechanisms of the pathogenesis of AD. Shortly after the discovery of *PS1,* DNA sequence database searches led to the identification of a homologous gene named presenilin 2 *(PS2)*, located on chromosome 1 (1q41.2). Mutations in *PS2* also result in FAD.

Pathogenic mutations in both of these genes have been identified; overall, mutations in *PS1, PS2,* and *APP* genes account for the majority of FAD cases. Although the proteins encoded by the two presenilin genes are 63% homologous at the amino acid level, intensive mutation and epidemiological analysis have revealed some interesting differences between the two proteins. The number of mutations identified in *PS1* is far greater than those in *PS2*. To date, more than 107 individual point mutations in *PS1* have been identified (Fig. 3, http://molgen-www.uia.ac.be/ADMutations) and found to be responsible for the majority of the early onset FAD. Most of these alterations are missense mutations that result in single amino acid changes in residues that are conserved between the PS1 and PS2 proteins. An intronic mutation (referred as ΔE9) results in the disruption of a splice acceptor site leading to the in-frame deletion of exon 9/10. This is the only *PS1*-FAD mutation that results in pathologically active protein that does not undergo proteolysis. Endoproteolytic

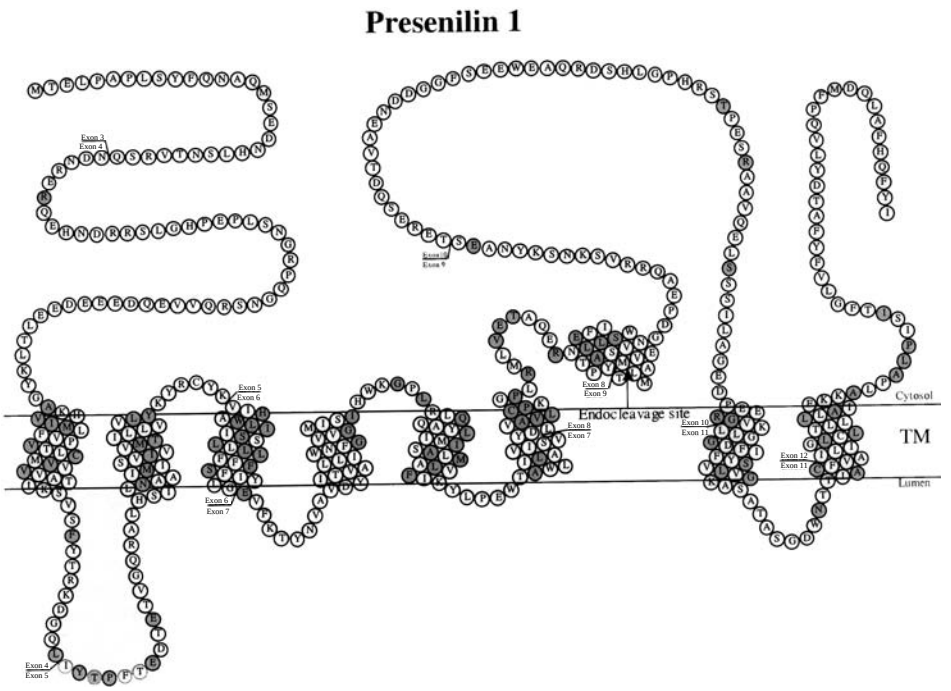

Fig. 3 FAD mutations in presenilin-1. Shown is a schematic representation of human presenilin 1 protein. The protein is predicted to have eight transmembrane domains. Arrow indicates the endocleavage site. Shaded residues are those associated with FAD mutations.

processing of the presenilins is an important aspect of the biochemistry of both PS1 and PS2 proteins, since it has been implicated in the formation of the functionally active complex (see Sect. 3). The most consistent effect mediated by the presenilin FAD mutations is increased production of Aβ42. This has been confirmed both in transfected cells and in transgenic mice. Although the mechanism by which FAD mutations in *PS1* and *PS2* cause AD is not fully understood, it is suggested that FAD mutations may alter γ-secretase activity toward cleavage of the more amyloidogenic peptide of Aβ, Aβ42/43. Although point mutations are distributed throughout the *PS1* gene, the majority of them are clustered within the TM domains and large hydrophilic loop between TM6 and TM7 (Fig. 3). The mean age of onset in all *PS1* families is about 45 years with a range of 28 to 62 years, suggesting that mutations in *PS1* result in a more aggressive form of FAD than APP mutations.

In contrast to *PS1*, only eight point mutations have been identified in *PS2* (http://molgen-www.uia.ac.be/ADMutations). The first *PS2* mutation to be identified was the N141I mutation that is responsible for disease in a large Volga German kindred. Whether the preponderance of *PS1* versus *PS2* mutations is due to the differing genomic environment of the two genes or the factors related to differing biological functions of the two molecules, remains unclear. Mutation analysis of the

two proteins has also led to the identification of a phenotypic difference. Overall, the mutations in *PS1* result in relatively early and constant age of onset while those in *PS2* lead to a somewhat later and more variable age of onset. The identification of a *PS1* N135I mutation, a position and substitution that corresponds to the *PS2* Volga German mutation, and the finding that the affected members of this pedigree have a relatively constant age of onset suggests that the observed variations in age of onset are dependent on the molecule itself as opposed to the specific position or nature of the amino acid substitution.

2.2 The Genetics of Sporadic Alzheimer's Disease

While the genetics of FAD is fairly well understood, our understanding of the more common sporadic AD remains much less complete. The majority of cases have onset >60 years. Some families have stronger clustering of AD cases than others, suggesting that the genetic component of the disease can be quite variable. The estimated cumulative risk to first-degree relatives of AD-affected probands approaches 50% by age 90 compared to a disease risk of 10 to 15% in the general population. To date, only the Apolipoprotein *E*4 (*APOE4*) allele has been linked to increased risk for sporadic AD. While it is clear that *APOE* is a major risk factor for AD, epidemiological studies estimate that 42 to 68% of AD cases do not have an *APOE4* allele, indicating that additional genetic and environmental factors are involved in this form of the disease. The balance of these risk factors most likely determines both the age of onset of the disease and the number of affected first-degree relatives.

2.2.1 Apolipoprotein E

Linkage analysis in late-onset AD pedigrees led to the observation that these families inherit a genetic risk factor located on chromosome 19. The defect involved in these particular pedigrees is a susceptibility defect as opposed to a causative defect such as those present in the gene coding for APP and the presenilins. The inheritance of the susceptibility gene defect alone is not sufficient to cause the disease, and one or more secondary events, either environmental or genetic, must accompany the inheritance of the primary defect to cause the disease. Association studies have shown that the chromosome 19 susceptibility gene (localized to 19q13.2) is within the *APOE* gene locus.

APOE protein is a lipid-transport molecule and the primary apolipoprotein observed in the CNS. It is encoded by a polymorphic gene that exists as three alleles designated *APOE2*, *APOE3*, and *APOE4*. These genetic variations result in amino acid substitutions (arginine or cysteine) at positions 112 and 158 of the protein. Thus, whereas the *APOE2* polypeptide has a cysteine at both of these positions, *APOE3* has a cysteine at 112 and an arginine at 158, and *APOE4* has arginine at both positions. In most Caucasian populations, the *APOE3* is the most common allele (frequency 0.78); *APOE4* (frequency 0.14) and *APOE2* (frequency 0.08) are considered variants.

When the association of the three common *APOE* alleles with AD was examined in late-onset AD pedigrees, it was found that the *APOE4* allele frequency was 52%, compared to 16% for age-matched controls. In addition, in sporadic AD cases, the *APOE4* allele frequency was 40%. The *APOE4* association with AD was rapidly and widely confirmed in patients with familial and sporadic AD. These multiple

confirmations established the *APOE4* allele as the most important genetic marker of risk for the disease identified so far, accounting for approximately 50% of the genetic component of AD. The frequency of *APOE4* allele among AD and control cases can vary in different populations, however, the association with AD is consistent.

Individuals inheriting two copies of the *APOE4* allele have an increased risk and earlier age of onset of AD than individuals with one *APOE4* allele. Likewise, individuals with one *APOE4* allele have a greater risk and earlier onset of disease than those with no *APOE4* alleles. These data demonstrate that the effect of *APOE4* on risk and age of onset is dose-related. In contrast to the effect of *APOE4* allele, the *APOE2* allele has a protective effect on AD, as evidenced by a decreased frequency of this allele in AD cases. The effect of the most commonly inherited allele, *APOE3*, on risk and age of onset falls between *APOE4* and *APOE2*. There also appears to be an interaction of *APOE* genotypes with the age of onset in FAD. When survival analysis was used in a large Colombian kindred carrying a presenilin mutation, the presence of an *APOE4* allele was observed to decrease age of onset, while presence of an *APOE2* allele increased age of onset. Although it is still not clear how an *APOE4* allele modifies risk for AD, several studies suggest that APOE may act via an Aβ-dependent mechanism.

2.2.2 The Search for New Alzheimer's Disease Susceptibility Genes

There are two genetic approaches to finding susceptibility loci: full genome scans and candidate gene studies. In candidate gene studies, genes are selected on the basis of the known biology of the disease or their location within chromosomal regions implicated in risk for late-onset AD, and are assessed individually in order to determine whether variants in each candidate are associated with disease. An alternative approach, a genome scan, does not select genes *a priori*, but instead tests for linkage or association between the disease and polymorphic markers that are spaced evenly through the genome. Association methods use unrelated case-control samples or siblings who are discordant for disease, and test for differences in allele frequencies between affected and unaffected individuals. In contrast, linkage studies use extended families or affected sibling pairs and look for regions of the genome in which there is an increased allele sharing between affected relatives.

To date, only five late-onset AD genome-wide screens have been reported. The most consistent findings among these studies are evidences for AD susceptibility loci on chromosomes 19 (*APOE*), 9, 10, and 12. Since FAD is closely associated with elevated Aβ42 levels, some investigators have hypothesized that loci, which modify plasma Aβ42, might also modify risk for AD. Using this approach, evidence for linkage was observed in the same region of chromosome 10 as was observed in the AD-linkage studies. Over the past few years, a large number of biologically interesting candidate genes as well as candidate genes that map within chromosome regions implicated in risk for late-onset AD have been evaluated in case-control populations. The genes that have received most attention are those encoding *APOE* on chromosome 19, α2-macroglobulin and the LDL-receptor related protein on chromosome 12 and insulin degrading enzyme and urokinase on chromosome 10. All of these genes have been closely linked to Aβ metabolism. However, with the exception of the *APOE* gene, whose association with

AD has been universally replicated, the relative significance of these associations is still in question.

3 Molecular and Cellular Mechanisms Underlying Alzheimer's Disease

3.1 β-amyloid Precursor Protein

β-Amyloid precursor protein (APP) is a large, type I transmembrane protein that is expressed in almost all cell types except erythrocytes. It consists of a large extracellular N-terminal region, a transmembrane domain, and a short intracellular C-terminal tail. The Aβ-peptide sequence begins close to the membrane on the extracellular part of APP and ends within the membrane-spanning domain. At least five different mRNAs arise as a result of alternative splicing of the primary transcript of the APP gene: APP770, 751, 714, 695, and 563. Transcript APP695 lacks the Kunitz-type protease inhibitor (KPI) domain and is the major form expressed in neurons, whereas APP751, which contains the KPI domain, is the major transcript in non-neuronal cell types. The APP transcripts appear to be developmentally regulated, and all cell types so far tested contain at least one of the transcripts.

In addition to differential RNA splicing, APP undergoes posttranslational modification, including phosphorylation, N- and O-linked glycosylation, sulfation and proteolytic processing. It is not clear how these different events modify APP metabolism or function. A detailed knowledge of post-translational events is undoubtedly crucial to understanding amyloid plaque formation and to elucidating the biological functions of APP and Aβ.

The normal function of APP is still unclear, but it has been implicated in neuroprotective action. APP is a member of a multigene family that contains at least two other homologs known as amyloid precursor-like protein 1 and 2 (APLP1 and APLP2). Both $APP^{-/-}$ and $APLP2^{-/-}$ mice are viable while APP/APLP2 double knockout mice are early postnatal lethal suggesting that APP and APLPs may share important physiological functions and the APLPs may compensate for the function of APP.

3.2 Production of Aβ – Proteolytic Processing of APP

It is clear that Aβ production does occur in normal cells, suggesting that excessive Aβ production in AD is due to enhanced or altered processing of APP by normal pathways that generate Aβ, rather than by an abnormal route specific to Alzheimer's cells. APP undergoes a series of proteolytic processing events (Fig. 4) by several different proteases called secretases. It is first cleaved by either α- or β-secretase. Two disintegrin metalloproteases, ADAM 10 and TACE (TNF-α converting enzyme), are involved in α-cleavage of APP. This cleavage produces a large soluble N-terminal fragment (NTF), APPsα and an 83-amino acid membrane-bound C-terminal fragment, C83 (CTFα). Alternatively, cleavage by the β-secretase results in an N-terminal fragment APPsβ and a 99-amino acid membrane-bound fragment, C99 (CTFβ). The β-secretase has recently been identified as a transmembrane aspartyl protease called BACE1, for β-site APP-cleaving enzyme. Both C83 and C99 are substrates for γ-secretase, a mysterious enzyme that carries out an unusual proteolysis in the middle of the transmembrane domain of

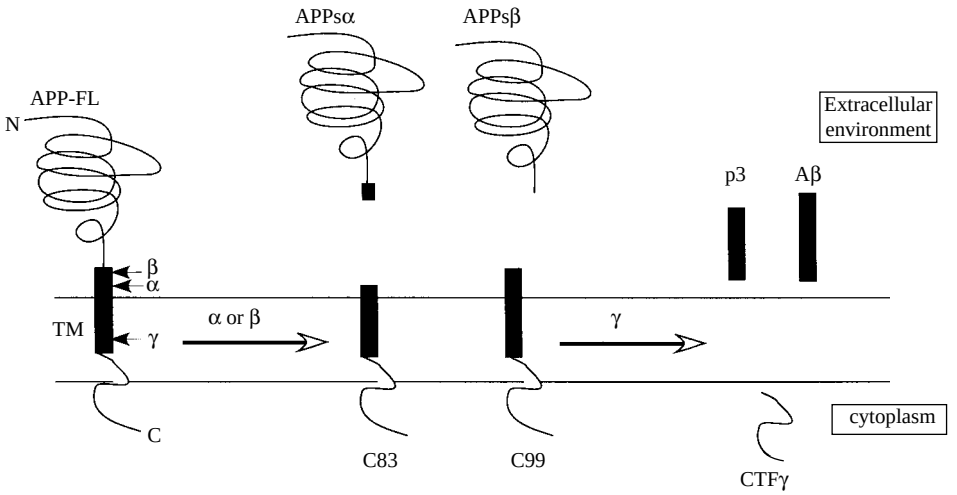

Fig. 4 Proteolytic processing of APP. APP is initially cleaved by either α- or β-secretase to generate membrane-bound C-terminal fragments, C83 or C99, which are subsequently cleaved by the γ-secretase (in the transmembrane domain) to produce p3 or Aβ.

APP, resulting in the release and secretion of the 4-kD Aβ from C99 and the 3-kD p3, a nonpathogenic peptide, from C83. γ-secretase cleavage is promiscuous and produces a spectrum of Aβ peptides varying from 39 to 43 residues in length. Over 90% of the Aβ in the brain is Aβ40 and the remaining 5 to 10% is the longer species Aβ42/43.

The C-terminal product of the γ-secretase cleavage is called CTFγ or AICD (from APP Intracellular Domain). Interestingly, CTFγ, purified from brain tissue, is not C59/57, which would be the predicted length of the C-terminal product of Aβ cleavage. Instead, sequence analysis reveals that it is C50/49, several amino acids shorter. The mechanism that accounts for such differential cleavages is currently under investigation. Several reports have suggested a role of CTFγ in APP signaling to the nucleus. Indeed, it was shown that CTFγ, when together with Fe65 protein, translocates to the nucleus and that this complex can activate transcription of a reporter gene *in vitro*.

3.3
β-secretase – BACE1

In late 1999, four groups almost simultaneously reported the discovery of BACE1, a novel aspartyl protease that exhibited all the known functional characteristics of the β-secretase. BACE1 is a type I transmembrane protein with an open reading frame of 501 amino acids. It is most closely related to the pepsin family of aspartyl proteases. BACE1 is expressed in most tissues and cell types at modest levels, but shows much higher expression levels in neurons. Structurally, it contains an N-terminal 21 amino acid signal peptide, a propeptide domain, a large lumenal domain followed by a transmembrane domain, and a short cytosolic C-terminal tail. BACE1 contains two aspartyl protease active site motifs in the lumenal domain: DTGS

(residues 93–96) and DSGT (residues 289–292). Both aspartic acid residues have been shown to be essential for enzymatic activity. BACE1 contains intrinsic proteolytic activity; as a purified recombinant, BACE1 can directly cleave APP *in vitro*. Overexpression of BACE1 in cell lines dramatically increases β-secretase activity and β-cleavage products: APPsβ and C99. In addition, mice deficient in BACE1 show a normal phenotype, but Aβ generation is abolished, further supporting that BACE1 is indeed a β-secretase.

BACE1 is initially synthesized in the ER as an immature precursor protein, proBACE1, which undergoes rapid maturation involving N-glycosylation as well as removal of the propeptide domain by a furin-like convertase in the Golgi apparatus. Unlike most zymogens, proBACE1 possesses β-secretase activity. The propeptide domain does not appear to significantly inhibit protease activity, but rather acts as a chaperone to assist proper folding of the protease domain. Phosphorylation of BACE1 at Ser498 appears to be important for the intracellular trafficking of the protein between early endosomes and late endosomes/TGN compartments.

A homologous aspartyl protease, BACE2 was shortly identified by searching the EST database with the BACE1 sequence. BACE1 and BACE2 share ~64% amino acid similarity. Together they define a novel family of transmembrane aspartyl proteases. Brain expression of BACE2 is very low and most evidence suggests that it does not play a major role in β-cleavage of APP.

3.4 γ-secretase – Presenilins and Cofactors

The enzyme that is clearly central to AD pathogenesis is the γ-secretase, which cleaves C99 and produces Aβ. Evidence suggests that it is an aspartyl protease with the unusual property of cleaving substrates within the transmembrane domain. Independent from its role in processing APP, γ-secretase also mediates the intramembranous cleavage of several other type I transmembrane proteins including Notch and ErbB-4. γ-secretase cleavage of Notch releases the Notch intracellular domain, which translocates to the nucleus and acts as a cofactor to modify transcription of target genes.

The identity of the γ-secretase is rather enigmatic. It appears to be a multiprotein, high molecular weight complex that consists of presenilins (PS) and three other cofactors: nicastrin, APH-1 and PEN-2.

3.4.1 Presenilins

Presenilins are a family of polytopic transmembrane proteins that are implicated in AD pathogenesis. PS homologs have been found in species as diverse as *Drosophila melanogaster*, *Caenorhabditis elegans* and *Arabidopsis*, but not in yeast. The function of PS appears to be conserved. Deficiency in sel-12, one of the nematode PS homolog, causes egg-laying defects. This phenotype can be rescued by expressing human PS1 in the mutant worms. In mammals, there are two PS genes, *PS1* and *PS2*, that share 65% identity. *PS1* is believed to account for the majority of PS functions, while *PS2* seems to play a complementary role.

Presenilins are found primarily within intracellular membranes including the ER and Golgi as well as the plasma membrane. Structurally, they are predicted to have eight transmembrane domains with both the amino and carboxyl termini oriented toward the cytoplasm (Fig. 3). A large cytoplasmic loop is postulated between TM6 and TM7. PS are synthesized as single

polypeptides that rapidly undergo endoproteolysis within the cytoplasmic loop, generating a 30-kD N-terminal fragment (NTF) and a 20-kD C-terminal fragment (CTF). The NTF and CTF form a stable complex, which appears to be the functional form of presenilins. PS endoproteolysis is tightly regulated such that overexpression of PS leads to the accumulation of a full-length protein, but does not result in an increase in the fragments. The NTF/CTF complex, but not the holoprotein, are components of a high molecular weight complex, estimated at ~250 kD or even larger. This high molecular weight complex is believed to be the γ-secretase complex.

Many lines of evidence indicate that *PS1* is the catalytic component of the γ-secretase. The two membrane-embedded aspartyl residues in *PS1* (D257 in TM6 and D385 in TM7) are essential for γ-secretase activity. The mutation of either residue abolishes γ-secretase processing of APP and Notch. The knockout of the PS genes blocks the Aβ production and Notch function. While in *PS1* single K/O cells, there is residual Aβ production, in PS1/2 double knockout cells, there is no detectable Aβ, demonstrating that there is no PS-independent γ-secretase activity. Inhibitors designed to be transition-state analogs of the γ-secretase bind to presenilin fragments. Recently, a family of signal peptide peptidases was discovered, which are apparently *bona fide* intramembranous aspartyl proteases, with similar active site motifs and similar topology to the presenilins. This finding further supports the hypothesis that presenilin is an aspartyl protease.

3.4.2 Cofactors

Nicastrin is a large type I transmembrane glycoprotein that was immunopurified from PS1 immunoprecipitates. It has no significant sequence homology to other functionally characterized proteins. Nicastrin is a 709 amino acid protein containing a putative signal peptide, a long N-terminal hydrophilic domain with numerous glycosylation sites, a TM domain and a short hydrophilic C-terminal tail of 20 residues. The mature protein migrates at ~150 kD owing to heavy glycosylation. There is evidence suggesting that the association of nicastrin with γ-secretase is tightly regulated via glycosylation. Nicastrin binds to presenilins and the active γ-secretase complex. Nicastrin deficiency in *Drosophila* abolishes Notch signaling and γ-secretase cleavage of APP. In addition, presenilin stability was compromised in the absence of nicastrin. All these data support that nicastrin is a component of the γ-secretase complex and an important regulator of the γ-secretase activity.

Genetic screens in *C. elegans* identified two proteins, APH-1 and PEN-2, which interact with worm homolog of *PS1* and nicastrin. Loss-of-function mutations of either protein produces phenotypes that resemble that of null mutants in nicastrin and presenilins. APH-1 is an integral protein containing seven predicted TM domains. There are two mammalian *APH-1* genes (*mAPH-1a* and *mAPH-1b*). *mAPH-1a* has at least two splice variants: *mAPH-1a*^{L} has seven exons and encodes a longer protein of 265 residues; mAPH-1a^{S} has six exons and encodes a shorter protein of 247 residues. All three variants are widely expressed, although their abundance seems to be coordinately regulated. To date, no functional difference among these variants has been documented. PEN-2 is a small protein of 101 amino acids containing two TM domains. Topology study suggests that both the N- and C-terminals of the protein face the lumen of the ER.

Human APH-1 and PEN-2 can partially rescue the *C. elegans* mutants, suggesting functional conservation among eukaryotes. The human genes must be present together to rescue the phenotypes and adding PS1 improves the rescue, indicating that these proteins act in concert. Consistent with the genetic studies, biochemical studies showed that reduction of APH-1 and PEN-2 expression in mammalian cells by RNA inhibition reduced γ-secretase activity.

While the exact functions of these cofactors remain to be elucidated, the data obtained so far suggest that all four proteins, presenilins, nicastrin, APH-1, and PEN-2, are necessary components of the γ-secretase. They appear to regulate one another for protein stability, proper maturation, and functional γ-secretase complex formation. It is also very likely that the least abundant component becomes the cellular-limiting factor that controls overall γ-secretase activity.

3.5 Neurofibrillary Tangles and Tau Protein

Neurofibrillary tangles are intracellular deposits that consist of aggregates of a fibrous substance, the paired helical filaments (PHF). These filaments are highly insoluble, even resistant to strong detergents such as SDS and guanidine hydrochloride. A major component of PHF is tau, a microtubule (MT)-associated phosphoprotein. Tau is abundantly expressed in the brain. At least six isoforms of tau are present in the brain, resulting from alternative mRNA splicing of a single gene on chromosome 17q.

Unlike APP, whose functions are not clear, tau has a biological role that is reasonably well understood. One function of tau is to promote microtubule assembly, a process controlled by phosphorylation of tau. In addition, it has been proposed that tau has a specific role in the generation of axon morphology. Increased phosphorylation of tau decreases its ability to promote tubulin polymerization to microtubules.

Tau in PHF is hyperphosphorylated, and is insoluble under conditions in which normal tau is soluble. It is believed that tangles are formed when tau becomes hyperphosphorylated and dissociates from microtubules and spontaneously forms insoluble PHF. Over 25 phosphorylation sites have been identified in PHF tau purified from AD brains. Most of the phosphorylation sites are located in the flanking region of the MT-binding domain. A variety of protein kinases, including glycogen synthase kinase-3β (GSK3β), cyclin-dependent kinase 5 (cdk5), mitogen-activated protein kinase (MAPK) and protein kinase A, have been found to be involved in hyperphosphorylation of tau. This suggests that multiple phosphorylation cascades may be activated in NFT-bearing neurons. However, it remains unclear whether these cascades are activated independently or whether one or more kinases are primarily responsible for initiating the cascade *in vivo*.

In addition to their presence in AD, filamentous tau inclusions accompanied by extensive gliosis and loss of neurons are the neuropathological hallmarks of other neurodegenerative diseases that are sometimes designated as tauopathies. There has been considerable debate in the AD literature concerning the primacy of Aβ versus NFT in the pathogenesis of AD. This was largely resolved by the identification of mutations in the tau gene in families with frontotemporal dementia with parkinsonism (FTDP-17). FTDP-17 is characterized clinically by the presence of

prominent behavioral changes early in the disease. The neuropathology associated with FTDP-17 is characterized by large numbers of NFT but no senile plaques. Mutations in the tau gene in FTDP-17 can alter splicing and produce a shift from the short tau isoform with three repeats to the longer isoform, which contains four repeats. Several other point mutations have been discovered in the coding region of tau that impact the functional integrity of the tau protein itself, P301L being the most common. So far, genetic analysis has not linked tau mutation directly to AD, but tau dysfunction may still have a significant role to play in the disease, if not in its initiation, then in its progression.

3.6 Apolipoprotein E

Apolipoprotein E is a 299 amino acid glycoprotein with a molecular weight of ~34 kD. Its N-terminal domain consists of four amphipathic α-helices and the C-terminal domain contains most of the lipid-binding activity. APOE is a major apolipoprotein that regulates cholesterol uptake and release. The CNS contains high levels of APOE with abundance only second to the liver. In the CNS, APOE is synthesized predominantly by astrocytes and is secreted into the CSF, where it is a major lipid carrier.

APOE appears to have multiple functions in the brain – any or all of which could potentially influence AD pathogenesis. The molecular mechanisms by which APOE protein is involved in AD pathogenesis are unclear because reports on the allele-specific effects of APOE are controversial. Nevertheless, numerous studies have demonstrated that APOE can affect AD pathogenesis either by influencing Aβ deposition or by a direct effect on neuronal survival. APOE is a major component of senile plaques and specifically binds to soluble Aβs. Allele-specific differences in Aβ binding, aggregation, and fibrillogenesis have been reported. Most histopathologic studies demonstrated a correlation between senile plaque density and *APOE4* allele dose in AD. Animal studies also show that APOE has isoform-specific effects on Aβ deposition and plaque formation (see Sect. 4.4).

There is also evidence that different alleles of *APOE* have different effects on neurite extension, neuronal survival, and the cell's response to oxidative stress. *In vitro* studies with primary neurons and neuronal cell cultures show that lipid-bound E3 promotes neurite extension, while E4 is neutral or inhibitory. These *in vitro* findings are consistent with reported defects in neuronal remodeling in AD brains carrying an E4 allele. APOE can have allele-specific differences on both neurotoxic or neuroprotective activities. High levels of APOE or some APOE synthetic peptides were shown to cause neurite degeneration or neuronal death in cultured cells, with the E4 having the most neurotoxic effects. On the other hand, APOE appears to play a role in protection against oxidative stress, a prominent phenomenon in AD pathogenesis. Apoe -/- mice exhibit increased markers of lipid peroxidation within the brain and appear to have increased susceptibility to neuronal damage. Different *APOE* alleles are shown to have isoform-specific neuroprotective effects (E2 > E3 > E4) on oxidative cytotoxicity caused by hydrogen peroxide, or Aβ peptides. A possible mechanism for the different antioxidant activities of the *APOE* alleles is their ability to detoxify 4-hydroxynonenal, a neurotoxic lipid-peroxidation product that is believed to play a key role in neuronal death in

AD. 4-hydroxynonenal covalently binds to APOE, with more binding to E2 and E3 than E4, probably because of the cysteine residues that are present in E2 and E3. E2 and E3 may therefore protect neurons by binding to, and thereby detoxifying, 4-hydroxynonenal.

4 Transgenic Mouse Models of Alzheimer's Disease

A major hindrance to basic research into the molecular pathogenesis of AD has been the absence of a mouse or rat model. As a result of the identification of mutations that cause AD in humans, this has begun to change. Although initial efforts to make transgenic models of AD failed, more recent efforts have recapitulated at least some of the key features of the disease. Many genetically altered mice have been designed to reproduce the neuropathology of AD, however, the majority of them share only some of the neuropathological and/or cognitive impairment characteristics of AD (http://www.alzforum.org). These include mice expressing human forms of the APP, the PS, APOE and, more recently, tau. Transgenic mice expressing human APP develop amyloid plaques, but neurodegeneration and neurofibrillary tangles are not observed. The Presenilin transgenic mice produced only increased Aβ42/43 levels and did not develop signs of AD pathology, even though the same mutations caused some of the earliest forms of inherited AD in humans. Strategies emphasizing tau resulted in increased phosphorylation of tau and tangle formation, although amyloid plaques were absent. Nevertheless, crossing transgenic animals expressing mutated tau and APP has produced a mouse that closely recapitulates the neuropathology of AD. This section provides a review of the various murine models and their role in understanding the pathogenesis of AD.

4.1 APP Transgenic Mice

The most extensively studied APP transgenic mouse lines are known as PDAPP and Tg2576. The PDAPP transgenic mouse expresses a human APP770 mini gene containing the V717F FAD-APP mutation, while Tg2576 expresses human APP695 that contains the Swedish FAD-APP mutation. In PDAPP, expression of the *APP* gene is under the control of the human platelet–derived growth factor (PDGF)-β chain neuronal promoter that targets expression preferentially to neurons in the cortex, hippocampus, hypothalamus, and cerebellum of the transgenic animals. This mouse was made on a mixed-strain background (C57BL/6, DBA/2, and Swiss-Webster), while Tg2576 was made on a single C57B6/SJL background. Expression of the *APP* gene in Tg2576 mice is under the control of the prion protein promoter resulting in a less-limited expression of APP both in the CNS and the periphery. The success of both models was due to the high level of APP expression achieved (>10-fold overexpression of human APP in PDAPP mouse and >6-fold higher expression of APP in Tg2576 compared to endogenous murine APP levels). Cortical and limbic amyloid deposits begin slightly earlier in the PDAPP mouse: by 3 months of age in homozygotes and at 6 to 9 months in heterozygotes, while in Tg2576 mouse they occur between 9 to 12 months of age, elevated Aβ production is observed as early as 3 months of age. Both models

develop β-amyloid deposits (in hippocampus and neocortex but limited in striatum and cerebellum), and the deposits are associated with dystrophic neurites, punctate immunoreactivity to hyperphosphorylated tau, astrocytosis, microgliosis, and vascular amyloidosis. The elevation of Aβ in the Tg2576 mouse correlates with the appearance of memory and learning deficits in the oldest group of transgenic mice as well. However, both APP transgenic models fail to show all the pathological features of human AD; no significant neuronal loss and no formation of NFTs were observed in either line.

Several other APP transgenic models were generated in order to obtain a complete or more comprehensive picture of the hallmarks of AD. The TgAPP22 transgenic mice overexpress a double mutant APP with Swedish (K670N, M671L) and London (V717I) FAD mutations in cis, while TgAPP23 only carries the Swedish mutation. In both models, the human APP751 isoform is used and its expression is under the control of the murine Thy-1 promoter. The TgAPP22 mice show twofold overexpression of the transgene over endogenous APP, and Aβ deposits were detected in the neocortex and hippocampus at 18 months of age, while the TgAPP23 has higher, sevenfold overexpression of human APP and typical Aβ plaques appear at the age of six months. Interestingly, these two murine AD models showed differences in plaque type; the majority of amyloid deposits in TgAPP22 mice brains are of the "diffuse" type, while in the TgAPP23 mice almost all extracellular amyloid deposits were fibrillar. Both substantial neurodegeneration and a reduction of neuron numbers were apparent in TgAPP23. In 14–18 months old TgAPP23 neuronal loss was 14% and reached 25% in mice with high Aβ plaque load. Additional pathological features included dystrophic neurites surrounding the plaques, hyperphosphorylated tau, but no NFTs developed.

Each of the models described above is remarkable in that the anatomical pattern of plaque formation parallels that seen in human AD. Furthermore, the morphology of amyloid plaques in aged APP-transgenic mice recapitulates amyloid pathology in human AD: the plaques span a continuum from diffuse Aβ deposits to compact core plaques with inflammation and neuritic dystrophy. However, none of these models reflects a complete picture of the neuropathology of AD.

4.2 Presenilin 1/2 Transgenic Mice

Several PS1 transgenic mice have been created to study presenilin biology *in vivo*. These mice differ in the promoter and the strains of mice used, but are similar in achieving high levels of protein production (1–3 fold over endogenous) in neuronal regions of the brain. Knockin transgenes were created expressing either the wild-type human PS1 or PS1 containing the FAD mutation A246E under the transcriptional control of the human Thy-1 promoter in PS1 null background. Both mice rescued the PS1 knockout mouse from embryonic lethality, confirming that PS1 mutation does not cause a loss of function of PS1 protein but has similar physiological properties as wild-type PS1. A second PS1 transgenic model expressing human mutant (M146L or M146V) and wild-type PS1 was generated under the control of the PDGF promoter. All three PS1 transgenics showed similar results: overexpressing mutant PS1 in the brains of transgenic mice lead to the elevation of

Aβ42/43, but not Aβ40. This effect was a direct result of the mutation and not overexpression of the human protein as the overexpression of the wild-type PS1 did not have any significant effect on Aβ levels. The physiological significance of the specific elevation of Aβ42/43 appears to be the observation that AD patients with PS1 mutations show plaques composed primarily of Aβ42/43. However, all three PS1 transgenic mice, in contrast to APP transgenics, revealed no Aβ deposition or other AD-associated pathology.

Recently, a PS1 mouse model carrying a PS1-P264L FAD mutation was developed using gene-targeting approach. The gene-targeted models are distinct from transgenic models because the mutant gene is expressed at normal levels, in the absence of the wild-type protein. $PS1^{P264L/P264L}$ mice had normal expression of PS1 mRNA, but levels of the N- and C-terminal protein fragments of PS1 were reduced while levels of the holoprotein were increased. $APP^{sw/sw}/PS1^{P264L/P264L}$ double gene-targeted mice had elevated levels of Aβ42, sufficient to cause Aβ deposition beginning at six months of age. This was the first animal model that exhibited Aβ deposition without overexpression of APP.

Three transgenic mice that express PS2-FAD mutation N141I under the control of PDGFβ chain promoter, chicken β-actin promoter or the neuron-specific enolase promoter were designed. The first two PS2 transgenic mice showed increased levels of Aβ42 in the brain of 12-month old mice. The third model utilizing neuron-specific enolase promoter did not show any difference in Aβ42 levels compared to the PS2 wild-type transgenic mice and both had no obvious differences in AD phenotypes (both had amyloid deposits).

4.3 Tau Transgenic Mice

One of the deficits of most AD transgenic models is the lack of tau pathology. APP and APP/PS transgenic mice, which exhibit extensive amyloid deposition also develop hyperphosphorylated tau around the amyloid deposits but it does not progress to resemble human NFTs. This suggests that some disruption of tau does occur in response to amyloid accumulation, but it does not progress to resemble human tauopathy in the mouse model. The idea that a species barrier prevents this progression has been proposed suggesting that primates and rodents differ in their response to injected amyloid not only in terms of neurodegeneration, but also in the accumulation of abnormal tau around the injected amyloid.

The most exciting development in the tau field is the creation of a transgenic mouse model with robust and reproducible tangle pathology. This transgenic mouse line, designated JNPL3, express human four repeat tau containing the most common FTDP-17 mutation, P301L, under the control of the mouse prion promoter. The mice develop motor and behavioral deficits that are associated with age- and gene-dosage-dependent development of congophilic NFT. Tau-immunoreactive pretangles are found in cortex, hippocampus, and basal ganglia, but at lower levels than commonly found in human disease. The tangles are associated with neurodegeneration, especially in the spinal cord where motor neurons were reduced by approximately 48%. This mouse has recently been used in breeding double and triple transgenic mouse models to produce a model system that more accurately recapitulates the hallmark pathology of AD (see further in the text). Recently, another

P301L tau transgenic model has been described that develops fivefold increase in the numbers of NFT soon after 18 days after the intracerebral injection of Aβ42 into the CA1 region of the hippocampus. These experiments indicate that an interaction of β-amyloid with the P301L tau mutation accelerates NFT formation.

4.4 Apolipoprotein E Transgenic Mice

A consistent consequence of carrying the *APOE4* allele is an increased number of amyloid plaques in the brain and more abundant amyloid deposition in the cerebral vasculature. The mechanism by which *APOE4* contributes to the development of neurodegeneration remains unknown, although the evidence suggests that this may be linked to the ability of APOE to interact with the amyloid β peptide and influence its concentration and structure. Recent studies with transgenic mice that overexpress the human APOE isoforms APOE2, APOE3, and APOE4 in a PDAPP hemizygous mouse (Apoe-/- background) show that isoforms that are known to increase the risk of AD enhance amyloid load and increase the neuritic pathology associated with fibrillar plaque development in aged animals (by 15 months of age APOE4 expressing mice having a 10-fold greater amyloid burden than APOE3 mice). The time course of Aβ deposition, Aβ levels, structure, and anatomic and subcellular distribution showed profound age-, species-, and isoform-dependent effect of APOE. It was also found that APOE not only impacts on the nature of aggregated Aβ, but also on the clearance of soluble Aβ from the brain across the blood–brain barrier (BBB). Aβ40 was rapidly cleared from the brain, while Aβ42 was cleared much less effectively, supporting the idea that Aβ42 production may favor amyloid deposition due to a reduced clearance across the BBB. However, Aβ clearance was not APOE isoform specific.

These recent findings are consistent with earlier work on Apoe null mice. When Tg2576 and PDAPP mice were crossed with Apoe null mice it was revealed that the absence of Apoe altered the quantity, character, and distribution of Aβ deposits in the transgenic animals, confirming that APOE affects the neuropathological phenotype in AD brains. The Aβ deposition is significantly reduced and is not thioflavine-S-positive.

4.5 Modulating and Enhancing the Phenotype: Transgenic Crosses

One of the great advantages of transgenic animals is that different mice can be mated together so that the effect of multiple transgenes on the disease phenotype can be observed. In particular, since most transgenic AD models do not exhibit all the pathological features of AD, different crosses have been made in order to design a model that will more closely describe the pathology, neurodegeneration, and memory loss seen in AD patients.

Double transgenic mice that carry *PS1-A246E* and *APPsw* FAD mutations show an elevated Aβ42/Aβ40 ratio in brain homogenates compared to the ratio observed in transgenic mice expressing *APPsw* alone or transgenic mice coexpressing wild-type human *PS1* and *APPsw*. In addition, double transgenic mice *APPswxPS1-A246E* developed numerous amyloid deposits much earlier (9 months of age) than age-matched mice expressing *APPsw* (18 months of age) and wild-type *huPS1* and *APPsw* alone. Interestingly, the majority of Aβ deposits in the double transgenic

mice were not immunoreactive to Aβ42 but instead were stained with antisera to Aβ40. Similarly, crossing the Tg2576 transgenic mice, which express mutant APPsw, with mice transgenic for *PS1-M146L* (PDPS1) revealed 41% increase in Aβ42/43 levels and formation of Aβ deposits in the cortex and hippocampus as early as 12 weeks of age, compared to development of Aβ deposits in Tg2576 mice of 9–12 months of age and approximately 1.5-fold elevation of Aβ42/43 from birth in PDPS1 line. The early Aβ deposits were found to be primarily composed of fibrillar Aβ and resembled compact amyloid plaques. As the mice aged, these fibrillar deposits did not increase substantially beyond the 12 months of age. Interestingly, the diffuse deposits appeared only until later, this is opposite to the general perception that in AD the compact deposits are formed by condensation of the diffuse material.

Studies on APPxPS1 double transgenics, therefore, revealed that marginal increases in Aβ42/43 levels, evoked by PS1 mutant transgene, can accelerate the deposition process by several months. As these mice have a severe Aβ pathology for an extended time relative to singly transgenic APP mice, other features of the disease such as tau abnormalities or major cell loss may become apparent as the animals age.

Double mutants produced from crossing JNPL3 transgenic mice expressing mutant P301L tau with Tg2576 mice expressing the APPsw mutation offered proof that Aβ influences the development of NFTs. The double mutant exhibited NFT pathology that was substantially enhanced in the limbic system and the olfactory cortex. These results suggest that APP or Aβ augments the formation of NFTs in the regions of the brain vulnerable to the formation of these lesions. Recently, a triple transgenic mouse was generated expressing APPsw, PS1-M146L, and P301L tau mutant protein that exhibited both plaques and NFT. However, other characteristics of AD such as neurodegeneration and memory loss have yet to be examined.

Although most of the transgenic models for AD do not completely recreate the disease phenotype, they have shown considerable utility, both for studying the disease mechanisms and for the preliminary testing of therapeutic agents, particularly those that are designed to modulate Aβ deposition.

5 Potential Treatments

The ultimate goal of AD research is to prevent disease and/or develop treatments for this devastating disease. Unfortunately, there is still no effective treatment available. Currently, the only FDA-approved therapy for AD is cholinesterase inhibitors, which can enhance cholinergic activity and temporarily improve cognitive function in some individuals in the early stages of disease. However, these drugs only treat the symptoms but have no impact on progression of the disease.

Recent advances in dissecting the molecular and cellular mechanisms that are involved in AD pathogenesis have provided substantial knowledge for determining novel targets for drug development. Several new therapeutic approaches targeted at distinct aspects of the disease are currently being pursued. Potential therapies include: anti-inflammatory agents, cholesterol-lowering drugs, antioxidants, hormonal therapy, and approaches that inhibit Aβ production or increase Aβ degradation and clearance. Strategies targeted directly at Aβ may be most effective

given the fact that Aβ deposition is the central and fundamental event that causes AD. Other approaches affecting downstream events in the neurodegenerative process are likely to be used as supplementary treatments.

5.1 Secretase Inhibitors

Two prime targets for drug development are β- and γ-secretases, the two enzymes that process APP to Aβ. Inhibitors to these two proteases are expected to have the greatest Aβ-lowering effect since inhibiting either enzyme can completely inhibit Aβ generation.

γ-secretase is considered to be central to AD pathogenesis because altering its activity invariantly favors Aβ42 production and causes AD. The major concern in developing a drug that targets the γ-secretase is its potential for side effects. It is clear that γ-secretase is important for several physiological functions through its role as the protease that mediates the intramembranous cleavage of a number of TM proteins, including Notch. One possible solution is to develop an inhibitor that selectively lowers Aβ (Aβ42 in particular) production without significantly affecting the cleavage of other γ-secretase substrates. The development of such inhibitors has been reported. Another possible approach is partial inhibition of γ-secretase activity, at inhibitor concentrations that reduce Aβ production without affecting Notch signaling. Indeed, it has been shown that compounds with no reported selectivity allow significant Aβ reduction without changing the Notch-related function. Some inhibitors, however, are reported to preferentially increase Aβ42 production when used at low concentrations.

Most evidence suggests that β-secretase (BACE1) may be a better therapeutic target than γ-secretase. BACE1 knockout mice do not generate Aβ, but are otherwise healthy and fertile without obvious deficits in the basal, neurological, and physiological functions. In addition, the X-ray structure of the BACE1 protease domain has been solved and should provide valuable knowledge for inhibitor design.

Although the knockout data are reassuring that the detrimental effects of inhibiting β-secretase may be minimal, concerns regarding the possible side effects remain. First, there are likely other β-secretase substrates since it is counterintuitive to assume that BACE1 has evolved just to generate Aβ. Indeed, one recent study suggests that β-secretase may be responsible for the cleavage and secretion of a Golgi resident sialyltransferase. It is very important to identify these substrates because they are valuable for predicting the possible side effects of BACE1 inhibitors. It will also be important to know the phenotype of BACE1/2 double knockout mice. Although BACE2 is not primarily involved in Aβ generation, it may have an important physiological role and inhibitors of BACE1 may also inhibit BACE2 because of the high degree of homology between these two proteins.

While inhibitors to both secretases may potentially be effective in AD treatment, the challenges facing drug development are significant. The inhibitors need to be highly selective for the target enzyme and target substrate to reduce possible side effects to a minimum. The inhibitors need to be able to efficiently cross the BBB. Potent peptide-based inhibitors to BACE1 have been developed. However, such large compounds are not viable drug candidates because they will not penetrate the BBB to a sufficient extent. Instead, the enzyme

inhibitors with therapeutic potential are preferably small organic molecules with high specificity. The X-ray structure of BACE1 suggests that its active site is more open and less hydrophobic than that of other aspartyl proteases. It may pose challenges for the development of small-molecule inhibitors.

5.2 Vaccine Approaches

Another novel approach for Alzheimer therapy is Aβ immunization. Many studies have shown that direct immunization with Aβ peptide or anti-Aβ antibodies can greatly reduce plaque formation, neuritic dystrophy, and other AD-like pathology in APP transgenic mice. One study also shows that such an immunization strategy is also effective in reversing behavioral deficits in the transgenic mouse model. It appears that antibodies to Aβ can enhance Aβ clearance, prevent Aβ fibril formation, disrupt Aβ fibrils, as well as block Aβ toxicity.

Despite the great effects in mice, there are serious concerns about the safety of this approach when used in humans. One concern is that autoimmunity may occur in humans, although it has not been reported in mice. Another major concern is the toxicity of Aβ42, which can cross the BBB and seed fibril formation in the brain, and therefore may actually promote plaque formation. In addition, Aβ42 may cause inflammation and neurotoxicity. A phase II clinical trial using Aβ42 vaccination was eventually terminated because of cerebral inflammation observed in several patients. The first autopsy result from this trial was just published. It shows exactly the two sides of this approach. Vaccination has such a powerful effect that the patient's cerebral cortex was almost cleared of Aβ deposition, but the side effect of cerebral inflammation was so severe that the patient developed meningoencephalitis, which led to death.

Safer therapeutic approaches including using nonamyloidogenic/nontoxic Aβ derivatives as an immunogen or passive immunization with Aβ antibodies are being developed. One study shows that passive immunization with $(Fab')_2$ fragments of Aβ-specific antibody can clear Aβ in a mouse model. This is promising because $(Fab')_2$ fragments do not interact with Fc receptors and so will not activate the cellular-immune response which has been the major cause for the severe side effects.

5.3 Other Therapeutic Approaches

Several FAD-approved nonsteroidal anti-inflammatory drugs (NSAIDs), including ibuprofen, indomethacin, and sulindac have been shown to selectively reduce Aβ42 levels in cultured cells. In addition, treatment of APP transgenic mice with ibuprofen reduces Aβ42 levels in the brain and suppresses plaque pathology. It appears that NSAIDs can subtly alter γ-secretase activity so that cleavage is shifted from Aβ42 to Aβ38.

Many studies also indicate that cholesterol influences Aβ metabolism and cholesterol-reducing drugs may have a beneficial effect on AD. Individuals taking cholesterol-lowering drugs such as statins show greatly reduced risk for developing AD, whereas individuals with elevated cholesterol are at a higher risk. Statins and other cholesterol-lowering drugs have been shown to reduce Aβ levels and Aβ deposition in both cell culture systems and animal models. The role of cholesterol in Aβ metabolism appears to be quite complex and the mechanisms are not clear.

Nevertheless, treatment with statins and other cholesterol-reducing drugs may have a significant clinical benefit in the prevention of AD.

6 Summary

We have seen a dramatic increase in our knowledge of the underlying molecular mechanisms of AD pathology during the last ten years. Understanding the effects of causative mutations in *APP* and the presenilins and the genetic risk factor, *APOE4* has led to the conclusion that changes in Aβ metabolism and deposition are central to the disease process. These studies suggest that similar mechanisms underlie both FAD and sporadic AD. Transgenic models of AD now provide a new hope for the rapid development in rational treatments for AD.

See also Brain Development; Medicinal Chemistry.

Bibliography

Books and Reviews

Duff, K. (2001) Transgenic mouse models of Alzheimer's disease: phenotype and mechanisms of pathogenesis, *Biochem. Soc. Symp.* **67**, 195–202.

Golde, T.E. (2003) Alzheimer disease therapy: Can the amyloid cascade be halted? *J. Clin. Invest.* **11**, 11–18.

Hardy, J., Selkoe, D.J. (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics, *Science* **297**, 353–356.

Myers, A., Goate, A.M. (2001) The genetics of late-onset Alzheimer's disease, *Curr. Opin. Neurol.* **14**, 433–440.

Richardson, J.A., Burns, D.K. (2002) Mouse models of Alzheimer's disease: a quest for plaques and tangles, *ILAR J.* **43**, 89–99.

Selkoe, D.J., Podlisny, M.B. (2002) Deciphering the genetic basis of Alzheimer's disease, *Annu. Rev. Genomics Hum. Genet.* **3**, 67–99.

Tandon, A., Fraser, P. (2002) The presenilins, *Genome Biol.* **3**, 3014.1–3014.9; reviews.

Vassar, R. (2002) β-Secretase (BACE) as a drug target for alzheimer's disease, *Adv. Drug Delivery Rev.* **54**, 1589–1602.

Primary Literature

Alzheimer, A. (1907) Über eine eigenartige Erkrankung der Hirnrinde, *Allg. Zeitsch. Psychiatrie Psychisch-gerichtliche Med.* **64**, 146–148.

Biernat, J., Gustke, N., Drewes, G., Mandelkow, E.M., Mandelkow, E. Phosphorylation of Ser262 strongly reduces binding of tau to microtubules: distinction between PHF-like immunoreactivity and microtubule binding, *Neuron* **11**, 153–163.

Blacker, D., Bertram, L., Saunders, A.J., Moscarillo, T.J., Albert, M.S., Wiener, H., Perry, R.T., Collins, J.S., Harrell, L.E., Go, R.C., Mahoney, A., Beaty, T., Fallin, M.D., Avramopoulos, D., Chase, G.A., Folstein, M.F., McInnis, M.G., Bassett, S.S., Doheny, K.J., Pugh, E.W., Tanzi, R.E. (2003) Results of a high-resolution genome screen of 437 Alzheimer's disease families, *Hum. Mol. Genet.* **12**, 23–32.

Borchelt, D.R., Ratovitski, T., van Lare, J., Lee, M.K., Gonzales, V., Jenkins, N.A., Copeland, N.G., Price, D.L., Sisodia, S.S. (1997) Accelerated amyloid deposition in the brains of transgenic mice coexpressing mutant presenilin 1 and amyloid precursor proteins, *Neuron* **19**, 939–945.

Borchelt, D.R., Thinakaran, G., Eckman, C.B., Lee, M.K., Davenport, F., Ratovitsky, T., Prada, C.M., Kim, G., Seekins, S., Yager, D., Slunt, H.H., Wang, R., Seeger, M., Levey, A.I., Gandy, S.E., Copeland, N.G., Jenkins, N.A., Price, D.L., Younkin, S.G., Sisodia, S.S. (1996) Familial Alzheimer's disease-linked presenilin 1 variants elevate Aβ1-42/1-40 ratio *in vitro* and *in vivo*, *Neuron* **17**, 1005–1013.

Braak, H., Braak, E. (1991) Neuropathological stageing of Alzheimer-related changes, *Acta Neuropathol. (Berl)* **82**, 239–259.

Brendza, R.P., Bales, K.R., Paul, S.M., Holtzman, D.M. (2002) Role of apoE/Aβ interactions in Alzheimer's disease: insights from transgenic mouse models, *Mol. Psychiatry* **7**, 132–135.

Cai, X.D., Golde, T.E., Younkin, S.G. (1993) Release of excess amyloid β protein from a mutant amyloid β protein precursor, *Science* **259**, 514–516.

Chung, H.M., Struhl, G. (2001) Nicastrin is required for Presenilin-mediated transmembrane cleavage in *Drosophila*, *Nat. Cell. Biol.* **3**, 1129–1132.

Citron, M., Oltersdorf, T., Haass, C., McConlogue, L., Hung, A.Y., Seubert, P., Vigo-Pelfrey, C., Lieberburg, I., Selkoe, D.J. (1992) Mutation of the β-amyloid precursor protein in familial Alzheimer's disease increases β-protein production, *Nature* **360**, 672–674.

Citron, M., Westaway, D., Xia, W., Carlson, G., Diehl, T., Levesque, G., Johnson-Wood, K., Lee, M., Seubert, P., Davis, A., Kholodenko, D., Motter, R., Sherrington, R., Perry, B., Yao, H., Strome, R., Lieberburg, I., Rommens, J., Kim, S., Schenk, D., Fraser, P., St. George Hyslop, P., Selkoe, D.J. (1997) Mutant presenilins of Alzheimer's disease increase production of 42-residue amyloid β-protein in both transfected cells and transgenic mice, *Nat. Med.* **3**, 67–72.

Corder, E.H., Saunders, A.M., Risch, N.J., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C. Jr., Rimmler, J.B., Locke, P.A., Conneally, P.M., Schmader, K.E., et al. (1994) Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease, *Nat. Genet.* **7**, 180–184.

Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L., Pericak-Vance, M.A. (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families, *Science* **261**, 921–923.

DeMattos, R.B., Bales, K.R., Cummins, D.J., Dodart, J.C., Paul, S.M., Holtzman, D.M. (2001) Peripheral anti-Aβ antibody alters CNS and plasma Aβ clearance and decreases brain Aβ burden in a mouse model of Alzheimer's disease, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8850–8855.

DeMattos, R.B., Bales, K.R., Cummins, D.J., Paul, S.M., Holtzman, D.M. (2002) Brain to plasma amyloid-β efflux: a measure of brain amyloid burden in a mouse model of Alzheimer's disease, *Science* **295**, 2264–2267.

De Strooper, B., Annaert, W., Cupers, P., Saftig, P., Craessaerts, K., Mumm, J.S., Schroeter, E.H., Schrijvers, V., Wolfe, M.S., Ray, W.J., Goate, A., Kopan, R.A. (1999) presenilin-1-dependent γ-secretase-like protease mediates release of Notch intracellular domain, *Nature* **398**, 518–522.

De Strooper, B., Saftig, P., Craessaerts, K., Vanderstichele, H., Guhde, G., Annaert, W., von Figura, K., Van Leuven, F. (1998) Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein, *Nature* **391**, 387–390.

Duff, K., Eckman, C., Zehr, C., Yu, X., Prada, C.M., Perez-Tur, J., Hutton, M., Buee, L., Harigaya, Y., Yager, D., Morgan, D., Gordon, M.N., Holcomb, L., Refolo, L., Zenk, B., Hardy, J., Younkin, S. (1996) Increased amyloid-β42(43) in brains of mice expressing mutant presenilin 1, *Nature* **383**, 710–713.

Edbauer, D., Winkler, E., Haass, C., Steiner, H. (2002) Presenilin and nicastrin regulate each other and determine amyloid β-peptide production via complex formation, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8666–8671.

Ertekin-Taner, N., Graff-Radford, N., Younkin, L.H., Eckman, C., Baker, M., Adamson, J., Ronald, J., Blangero, J., Hutton, M., Younkin, S.G. (2000) Linkage of plasma Aβ42 to a quantitative locus on chromosome 10 in late-onset Alzheimer's disease pedigrees, *Science* **290**, 2303–2304.

Esch, F.S., Keim, P.S., Beattie, E.C., Blacher, R.W., Culwell, A.R., Oltersdorf, T., McClure, D., Ward, P.J. (1990) Cleavage of amyloid β peptide during constitutive processing of its precursor, *Science* **248**, 1122–1124.

Estus, S., Golde, T.E., Younkin, S.G. (1992) Normal processing of the Alzheimer's disease amyloid β protein precursor generates potentially amyloidogenic carboxyl-terminal derivatives, *Ann. N. Y. Acad. Sci.* **674**, 138–148.

Fagan, A.M., Watson, M., Parsadanian, M., Bales, K.R., Paul, S.M., Holtzman, D.M. (2002) Human and murine ApoE markedly alters A β metabolism before and after plaque formation in a mouse model of Alzheimer's disease, *Neurobiol. Dis.* **9**, 305–318.

Farrer, L.A., Bowirrat, A., Friedland, R.P., Waraska, K., Korczyn, A.D., Baldwin, C.T. (2003) Identification of multiple loci for Alzheimer

disease in a consanguineous Israeli-Arab community, *Hum. Mol. Genet.* **12**, 415–422.

Flood, D.G., Reaume, A.G., Dorfman, K.S., Lin, Y.G., Lang, D.M., Trusko, S.P., Savage, M.J., Annaert, W.G., De Strooper, B., Siman, R., Scott, R.W. (2002) FAD mutant PS-1 gene-targeted mice: increased Aβ42 and Aβ deposition without APP overproduction, *Neurobiol. Aging* **23**, 335–348.

Francis, R., McGrath, G., Zhang, J., Ruddy, D.A., Sym, M., Apfeld, J., Nicoll, M., Maxwell, M., Hai, B., Ellis, M.C., Parks, A.L., Xu, W., Li, J., Gurney, M., Myers, R.L., Himes, C.S., Hiebsch, R., Ruble, C., Nye, J.S., Curtis, D. (2002) aph-1 and pen-2 are required for Notch pathway signaling, γ-secretase cleavage of βAPP, and presenilin protein accumulation, *Dev. Cell.* **3**, 85–97.

Games, D., Adams, D., Alessandrini, R., Barbour, R., Berthelette, P., Blackwell, C., Carr, T., Clemens, J., Donaldson, T., Gillespie, F. (1995) Alzheimer-type neuropathology in transgenic mice overexpressing V717F β-amyloid precursor protein, *Nature* **373**, 523–527.

Glenner, G.G., Wong, C.W. (1984) Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein, *Biochem. Biophys. Res. Commun.* **120**, 885–890.

Goate, A., Chartier-Harlin, M.C., Mullan, M., Brown, J., Crawford, F., Fidani, L., Giuffra, L., Haynes, A., Irving, N., James, L. (1991) Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease, *Nature* **349**, 704–706.

Goedert, M., Wischik, C.M., Crowther, R.A., Walker, J.E., Klug, A. (1988) Cloning and sequencing of the cDNA encoding a core protein of the paired helical filament of Alzheimer disease: identification as the microtubule-associated protein tau, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4051–4055.

Goldgaber, D., Lerman, M.I., McBride, O.W., Saffiotti, U., Gajdusek, D.C. (1987) Characterization and chromosomal localization of a cDNA encoding brain amyloid of Alzheimer's disease, *Science* **235**, 877–880.

Gómez-Isla, T., Price, J.L., McKeel, D.W., Morris, J.C., Growdon, J.H., Hyman, B.T. (1996) Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease, *J. Neurosci.* **16**, 4491–4500.

Gotz, J., Chen, F., van Dorpe, J., Nitsch, R.M. (2001) Formation of neurofibrillary tangles in P301l tau transgenic mice induced by Aβ42 fibrils, *Science* **293**, 1491–1495.

Goutte, C., Tsunozaki, M., Hale, V.A., Priess, J.R. (2002) APH-1 is a multipass membrane protein essential for the Notch signaling pathway in *Caenorhabditis elegans* embryos, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 775–779.

Grundke-Iqbal, I., Iqbal, K., Quinlan, M., Tung, Y.C., Zaidi, M.S., Wisniewski, H.M. (1986) Microtubule-associated protein tau. A component of Alzheimer paired helical filaments, *J. Biol. Chem.* **261**, 6084–6089.

Haass, C., Koo, E.H., Mellon, A., Hung, A.Y., Selkoe, D.J. (1992) Targeting of cell-surface β-amyloid precursor protein to lysosomes: alternative processing into amyloid-bearing fragments, *Nature* **357**, 500–503.

Haass, C., Schlossmacher, M.G., Hung, A.Y., Vigo-Pelfrey, C., Mellon, A., Ostaszewski, B.L., Lieberburg, I., Koo, E.H., Schenk, D., Teplow, D.B. (1992) Amyloid β-peptide is produced by cultured cells during normal metabolism, *Nature* **359**, 322–325.

Holcomb, L., Gordon, M.N., McGowan, E., Yu, X., Benkovic, S., Jantzen, P., Wright, K., Saad, I., Mueller, R., Morgan, D., Sanders, S., Zehr, C., O'Campo, K., Hardy, J., Prada, C.M., Eckman, C., Younkin, S., Hsiao, K., Duff, K. (1998) Accelerated Alzheimer-type phenotype in transgenic mice carrying both mutant amyloid precursor protein and presenilin 1 transgenes, *Nat. Med.* **4**, 97–100.

Holtzman, D.M. (2002) Aβ conformational change is central to Alzheimer's disease, *Neurobiol. Aging* **23**, 1085–1088.

Hong, C.S., Caromile, L., Nomata, Y., Mori, H., Bredesen, D.E., Koo, E.H. (1999) Contrasting role of presenilin-1 and presenilin-2 in neuronal differentiation *in vitro*, *J. Neurosci.* **19**, 637–643.

Hsiao, K., Chapman, P., Nilsen, S., Eckman, C., Harigaya, Y., Younkin, S., Yang, F., Cole, G. (1996) Correlative memory deficits, Aβ elevation, and amyloid plaques in transgenic mice, *Science* **274**, 99–102.

Hu, Y., Ye, Y., Fortini, M.E. (2002) Nicastrin is required for γ-secretase cleavage of the *Drosophila* Notch receptor, *Dev. Cell.* **2**, 69–78.

Hutton, M., Lendon, C.L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A.,

Hackett, J., Adamson, J., Lincoln, S., Dickson, D., Davies, P., Petersen, R.C., Stevens, M., de Graaff, E., Wauters, E., van Baren, J., Hillebrand, M., Joosse, M., Kwon, J.M., Nowotny, P., Heutink, P. (1998) Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17, *Nature* **393**, 702–705.

Iqbal, K., Zaidi, T., Thompson, C.H., Merz, P.A., Wisniewski, H.M. (1984) Alzheimer paired helical filaments: bulk isolation, solubility, and protein composition, *Acta Neuropathol. (Berl)* **62**, 167–177.

Kang, J., Lemaire, H.G., Unterbeck, A., Salbaum, J.M., Masters, C.L., Grzeschik, K.H., Multhaup, G., Beyreuther, K., Müller-Hill, B. (1987) The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor, *Nature* **325**, 733–736.

Leissring, M.A., Murphy, M.P., Mead, T.R., Akbari, Y., Sugarman, M.C., Jannatipour, M., Anliker, B., Muller, U., Saftig, P., De Strooper, B., Wolfe, M.S., Golde, T.E., LaFerla, F.M. (2002) A physiologic signaling role for the γ-secretase-derived intracellular fragment of APP, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4697–4702.

Levy, E., Carman, M.D., Fernandez-Madrid, I.J., Power, M.D., Lieberburg, I., van Duinen, S.G., Bots, G.T., Luyendijk, W., Frangione, B. (1990) Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type, *Science* **248**, 1124–1126.

Levy-Lahad, E., Wasco, W., Poorkaj, P., Romano, D.M., Oshima, J., Pettingell, W.H., Yu, C.E., Jondro, P.D., Schmidt, S.D., Wang, K. (1995) Candidate gene for the chromosome 1 familial Alzheimer's disease locus, *Science* **269**, 973–977.

Lewis, J., Dickson, D.W., Lin, W.L., Chisholm, L., Corral, A., Jones, G., Yen, S.H., Sahara, N., Skipper, L., Yager, D., Eckman, C., Hardy, J., Hutton, M., McGowan, E. (2001) Enhanced neurofibrillary degeneration in transgenic mice expressing mutant tau and APP, *Science* **293**, 1487–1491.

Lewis, J., McGowan, E., Rockwood, J., Melrose, H., Nacharaju, P., Van Slegtenhorst, M., Gwinn-Hardy, K., Paul Murphy, M., Baker, M., Yu, X., Duff, K., Hardy, J., Corral, A., Lin, W.L., Yen, S.H., Dickson, D.W., Davies, P., Hutton, M. (2000) Neurofibrillary tangles, amyotrophy and progressive motor disturbance in mice expressing mutant (P301L) tau protein, *Nat. Genet.* **25**, 402–405.

Lopez-Schier, H., St. Johnston, D. (2002) *Drosophila* nicastrin is essential for the intramembranous cleavage of notch, *Dev. Cell.* **2**, 79–89.

Masters, C.L., Simms, G., Weinman, N.A., Multhaup, G., McDonald, B.L., Beyreuther, K. (1985) Amyloid plaque core protein in Alzheimer disease and Down syndrome, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4245–4249.

Mayeux, R., Lee, J.H., Romas, S.N., Mayo, D., Santana, V., Williamson, J., Ciappa, A., Rondon, H.Z., Estevez, P., Lantigua, R., Medrano, M., Torres, M., Stern, Y., Tycko, B., Knowles, J.A. (2002) Chromosome-12 mapping of late-onset Alzheimer disease among Caribbean Hispanics, *Am. J. Hum. Genet.* **70**, 237–243.

Murrell, J., Farlow, M., Ghetti, B., Benson, M.D. (1991) A mutation in the amyloid precursor protein associated with hereditary Alzheimer's disease, *Science* **254**, 97–99.

Nicoll, J.A., Wilkinson, D., Holmes, C., Steart, P., Markham, H., Weller, R.O. (2003) Neuropathology of human Alzheimer disease after immunization with amyloid-β peptide: a case report, *Nat. Med.* **9**, 448–452.

Nilsberth, C., Westlind-Danielsson, A., Eckman, C.B., Condron, M.M., Axelman, K., Forsell, C., Stenh, C., Luthman, J., Teplow, D.B., Younkin, S.G., Naslund, J., Lannfelt, L. (2001) The 'Arctic' APP mutation (E693G) causes Alzheimer's disease by enhanced Aβ protofibril formation, *Nat. Neurosci.* **4**, 887–893.

Price, J.L., Davis, P.B., Morris, J.C., White, D.L. The distribution of tangles, plaques and related immunohistochemical markers in healthy aging and Alzheimer's disease, *Neurobiol. Aging* **12**, 295–312.

Rogaev, E.I., Sherrington, R., Rogaeva, E.A., Levesque, G., Ikeda, M., Liang, Y., Chi, H., Lin, C., Holman, K., Tsuda, T. (1995) Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene, *Nature* **376**, 775–778.

St. George-Hyslop, P.H., Tanzi, R.E., Polinsky, R.J., Haines, J.L., Nee, L., Watkins, P.C., Myers, R.H., Feldman, R.G., Pollen, D., Drachman, D. (1987) The genetic defect causing familial Alzheimer's disease maps on chromosome 21, *Science* **235**, 885–890.

Schenk, D., Barbour, R., Dunn, W., Gordon, G., Grajeda, H., Guido, T., Hu, K., Huang, J., Johnson-Wood, K., Khan, K., Kholodenko, D., Lee, M., Liao, Z., Lieberburg, I., Motter, R., Mutter, L., Soriano, F., Shopp, G., Vasquez, N., Vandevert, C., Walker, S., Wogulis, M., Yednock, T., Games, D., Seubert, P. (1999) Immunization with amyloid-β attenuates Alzheimer-disease-like pathology in the PDAPP mouse, *Nature* **400**, 173–177.

Scheuner, D., Eckman, C., Jensen, M., Song, X., Citron, M., Suzuki, N., Bird, T.D., Hardy, J., Hutton, M., Kukull, W., Larson, E., Levy-Lahad, E., Viitanen, M., Peskind, E., Poorkaj, P., Schellenberg, G., Tanzi, R., Wasco, W., Lannfelt, L., Selkoe, D., Younkin, S. (1996) Secreted amyloid β-protein similar to that in the senile plaques of Alzheimer's disease is increased *in vitro* by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease, *Nat. Med.* **2**, 864–870.

Seubert, P., Oltersdorf, T., Lee, M.G., Barbour, R., Blomquist, C., Davis, D.L., Bryant, K., Fritz, L.C., Galasko, D., Thal, L.J. (1993) Secretion of β-amyloid precursor protein cleaved at the amino terminus of the β-amyloid peptide, *Nature* **361**, 260–263.

Seubert, P., Vigo-Pelfrey, C., Esch, F., Lee, M., Dovey, H., Davis, D., Sinha, S., Schlossmacher, M., Whaley, J., Swindlehurst, C. (1992) Isolation and quantification of soluble Alzheimer's β-peptide from biological fluids, *Nature* **359**, 325–327.

Sherrington, R., Rogaev, E.I., Liang, Y., Rogaeva, E.A., Levesque, G., Ikeda, M., Chi, H., Lin, C., Li, G., Holman, K. (1995) Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease, *Nature* **375**, 754–760.

Shoji, M., Golde, T.E., Ghiso, J., Cheung, T.T., Estus, S., Shaffer, L.M., Cai, X.D., McKay, D.M., Tintner, R., Frangione, B. (1992) Production of the Alzheimer amyloid β protein by normal proteolytic processing, *Science* **258**, 126–129.

Sinha, S., Anderson, J.P., Barbour, R., Basi, G.S., Caccavello, R., Davis, D., Doan, M., Dovey, H.F., Frigon, N., Hong, J., Jacobson-Croak, K., Jewett, N., Keim, P., Knops, J., Lieberburg, I., Power, M., Tan, H., Tatsuno, G., Tung, J., Schenk, D., Seubert, P., Suomensaari, S.M., Wang, S., Walker, D., John, V. (1999) Purification and cloning of amyloid precursor protein β-secretase from human brain, *Nature* **402**, 537–540.

Skovronsky, D.M., Zhang, B., Kung, M.-P., Kung, H.F., Trojanowski, J.Q., Lee, V.M. (2000) *In vitro* detection of amyloid plaques in a mouse model of Alzheimer's disease, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7609–7614.

Stenh, C., Nilsberth, C., Hammarback, J., Engvall, B., Naslund, J., Lannfelt, L. (2002) The Arctic mutation interferes with processing of the amyloid precursor protein, *Neuro Report* **13**, 1857–1860.

Strittmatter, W.J., Saunders, A.M., Schmechel, D., Pericak-Vance, M., Enghild, J., Salvesen, G.S., Roses, A.D. (1993) Apolipoprotein E: high-avidity binding to β-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1977–1981.

Sturchler-Pierrat, C., Abramowski, D., Duke, M., Wiederhold, K.H., Mistl, C., Rothacher, S., Ledermann, B., Burki, K., Frey, P., Paganetti, P.A., Waridel, C., Calhoun, M.E., Jucker, M., Probst, A., Staufenbiel, M., Sommer, B. (1997) Two amyloid precursor protein transgenic mouse models with Alzheimer disease-like pathology, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13287–13292.

Sturchler-Pierrat, C., Staufenbiel, M. (2000) Pathogenic mechanisms of Alzheimer's disease analyzed in the APP23 transgenic mouse model, *Ann. N. Y. Acad. Sci.* **920**, 134–139.

Suzuki, N., Cheung, T.T., Cai, X.D., Odaka, A., Otvos, L., Eckman, C., Golde, T.E., Younkin, S.G. (1994) An increased percentage of long amyloid β protein secreted by familial amyloid β protein precursor (β APP717) mutants, *Science* **264**, 1336–1340.

Tanzi, R.E., Gusella, J.F., Watkins, P.C., Bruns, G.A., St. George-Hyslop, P., Van Keuren, M.L., Patterson, D., Pagan, S., Kurnit, D.M., Neve, R.L. (1987) Amyloid β protein gene: cDNA, mRNA distribution, and genetic linkage near the Alzheimer locus, *Science* **235**, 880–884.

Terry, R.D., Masliah, E., Salmon, D.P., Butters, N., DeTeresa, R., Hill, R., Hansen, L.A., Katzman, R. (1991) Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment, *Ann. Neurol.* **30**, 572–580.

Thinakaran, G., Borchelt, D.R., Lee, M.K., Slunt, H.H., Spitzer, L., Kim, G., Ratovitsky, T., Davenport, F., Nordstedt, C., Seeger, M., Hardy, J., Levey, A.I., Gandy, S.E., Jenkins, N.A., Copeland, N.G., Price, D.L., Sisodia, S.S. (1996) Endoproteolysis of presenilin 1 and accumulation of processed derivatives *in vitro*, *Neuron* **17**, 181–190.

Thinakaran, G., Harris, C.L., Ratovitski, T., Davenport, F., Slunt, H.H., Price, D.L., Borchelt, D.R., Sisodia, S.S. (1997) Evidence that levels of presenilins (PS1 and PS2) are coordinately regulated by competition for limiting cellular factors, *J. Biol. Chem.* **272**, 28415–28422.

Van Nostrand, W.E., Schmaier, A.H., Farrow, J.S., Cunningham, D.D. (1990) Protease nexin-II (amyloid β-protein precursor): a platelet α-granule protein, *Science* **248**, 745–748.

Vassar, R., Bennett, B.D., Babu-Khan, S., Kahn, S., Mendiaz, E.A., Denis, P., Teplow, D.B., Ross, S., Amarante, P., Loeloff, R., Luo, Y., Fisher, S., Fuller, J., Edenson, S., Lile, J., Jarosinski, M.A., Biere, A.L., Curran, E., Burgess, T., Louis, J.C., Collins, F., Treanor, J., Rogers, G., Citron, M. (1999) B-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BAC, *Science* **286**, 735–741.

Weggen, S., Eriksen, J.L., Das, P., Sagi, S.A., Wang, R., Pietrzik, C.U., Findlay, K.A., Smith, T.E., Murphy, M.P., Bulter, T., Kang, D.E., Marquez-Sterling, N., Golde, T.E., Koo, E.H. (2001) A subset of NSAIDs lower amyloidogenic Aβ42 independently of cyclooxygenase activity, *Nature* **414**, 212–216.

Weihofen, A., Binns, K., Lemberg, M.K., Ashman, K., Martoglio, B. (2002) Identification of signal peptide peptidase, a presenilin-type aspartic protease, *Science* **296**, 2215–2218.

Wolfe, M.S., Xia, W., Ostaszewski, B.L., Diehl, T.S., Kimberly, W.T., Selkoe, D.J. (1999) Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and γ-secretase activity, *Nature* **398**, 513–517.

Yu, G., Nishimura, M., Arawaka, S., Levitan, D., Zhang, L., Tandon, A., Song, Y.Q., Rogaeva, E., Chen, F., Kawarai, T., Supala, A., Levesque, L., Yu, H., Yang, D.S., Holmes, E., Milman, P., Liang, Y., Zhang, D.M., Xu, D.H., Sato, C., Rogaev, E., Smith, M., Janus, C., Zhang, Y., Aebersold, R., Farrer, L.S., Sorbi, S., Bruni, A., Fraser, P., St. George-Hyslop, P. (2000) Nicastrin modulates presenilin-mediated notch/glp-1 signal transduction and βAPP processing, *Nature* **407**, 48–54.

Animal Biotechnology and Modeling

Carl A. Pinkert, Michael H. Irwin, and Robert L. Howell
School of Medicine and Dentistry, University of Rochester, 601 Elmwood Ave., Rochester, NY 14642, USA

1 Introduction 211
1.1 Brief History and Definition of Transgenic Animals 211
1.2 Applications of Transgenic Animals 212
1.3 Societal and Ethical Considerations 214

2 Production of Transgenic Laboratory Animals 215
2.1 Choice of Animals 215
2.2 Gain-of-function, Loss-of-function and Conditional Modeling 216
2.3 Methodologies 216
2.3.1 DNA Microinjection 216
2.3.2 Retroviral Introduction of Transgenes 218
2.3.3 Embryonic Stem Cell Technology 219
2.3.4 Nuclear Transfer (Cloning) Methods 220
2.3.5 Other Methodologies 221
2.4 Other Laboratory Animal Models 222

3 Production of Transgenic Domestic Animals 223
3.1 Overview 223
3.2 Methodology 224
3.2.1 DNA Microinjection 224
3.2.2 Retroviral Infection 225
3.2.3 Nuclear Transfer 225
3.3 Traits Affecting Domestic Animal Productivity 227
3.4 Domestic Animals as Bioreactors 228
3.5 Examples from Domestic and Miniature Pigs 228

4 Analysis of Transgene Integration and Expression 231
4.1 Transgene Integration 231

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

4.2 Transgene-encoded mRNA Expression 232
4.3 Protein Expression 233

5 Conclusions and Future Directions 234
5.1 Gene Transfer Today 234
5.2 Future Directions 235

Bibliography 236
Books and Reviews 236
Primary Literature 236

Keywords

DNA Microinjection
A gene-transfer technique utilizing micromanipulation to inject DNA constructs (transgenes) directly into the pronuclei or nuclei of cells. The foreign DNA integrates into the host cell genome at a random location and usually in concatemers. DNA microinjection is the most commonly used gene-transfer technique for creating transgenic mammalian species.

Embryonic Stem (ES) Cell Transfer
The transfer of genetically manipulable pluripotent embryonic stem cells into a developing embryo; achieved by microinjection of stem cells into blastocysts or coculture with earlier stage ova. The resultant transgenic animal possesses a proportion of cells descended from the embryonic stem ES-cell lineage. Gene targeting resulting in both loss-of-function (knockout) and gain-of-function (knockin) is possible. In mammalian species, this gene-transfer technique has been successful only in producing germ line transmission of ES cell lineages in mice.

Gene Transfer
A set of techniques directed toward manipulating biological function via the introduction of foreign DNA sequences (genes) into living cells.

Nuclear Transfer (Cloning)
The nucleus of a cell with organism-specific DNA is provided by a donor cell and introduced into an enucleated oocyte (unfertilized ovum) or zygote, allowing for reprogramming fate of the developing embryo.

Transgenic Animal
An animal either integrating foreign DNA segments into its genome following gene transfer, or resulting from the molecular manipulation of endogenous genomic DNA.

Transgenic Founder
The first-generation transgenic organism directly produced as a result of gene transfer/germ plasm manipulation procedures. The term does not imply that the transferred gene (transgene) is heritable to future generations.

Transgenic Line
A direct familial lineage of organisms derived from one or more transgenic founders, characterized by the passing of the transgene(s) to successive generations as a stable genetic element. The line includes the founder and any subsequent offspring inheriting the specific germ line manipulation.

Transgenic animals and our ability to manipulate the genome of whole animals *in vivo* have influenced the life sciences in dramatic fashion. In less than 15 years, practical aspects of germ line manipulation of the genetic composition of animals have allowed researchers to address fundamental questions in fields ranging from medicine and biology to biotechnology and production agriculture. Studies emanating from basic efforts in gene regulation have led the way to *in vivo* gene-transfer studies focused on the expression of foreign genes and the ablation of endogenous genes. Our appreciation of basic biological development in animals and the potential ramifications of novel phenotypes resulting from genetic engineering efforts have been markedly influenced by pioneering efforts in transgenic animal technology.

1 Introduction

1.1 Brief History and Definition of Transgenic Animals

The ability to introduce functional genes into animals provides a very powerful tool for dissecting complex biological processes and systems. Hence, transgenic animals represent unique models that are custom-tailored to address specific biological questions. Furthermore, classical genetic monitoring cannot engineer a specific genetic trait in a directed fashion. For identification of interesting new models, the genetic screening and characterization of chance mutations remains a long and arduous task. Therefore, gene-transfer is also of great value, particularly for animal species in which long life cycles reduce the value of classical breeding practices for genetic modification.

Although the entire procedure for microinjection into living cells was described in the late 1960s, it took 15 years before transgenic animals were created. In 1980 and 1981, six pioneering laboratories reported successful gene transfer and the production of transgenic mice. Their success would not have been possible, however, were it not for the refinement of recombinant DNA technologies needed to develop viable protocols and to document results. In gene transfer, animals receiving new genes (foreign DNA sequences

integrated into their genome) are referred to as "transgenic," a term first coined by Gordon and Ruddle in 1981. As such, transgenic animals are recognized as specific species variants or strains, which result from the introduction and integration of new gene(s), or "transgenes," into the genome. More recently, the term *transgenic* has been extended to chimeric or *knockout* mice in which researchers have selectively removed or modified a gene or genes within the host genome.

For years, many studies were envisioned – if only an animal genome could be readily modified. The realization of the many technologies at hand today has opened numerous avenues for research. Production of transgenic mice has marked the convergence of earlier advances in the areas of recombinant DNA technology and the manipulation and culture of animal germ plasm (Fig. 1). Transgenic mice provide powerful models to explore the regulation of gene expression as well as the regulation of cellular and physiological processes. The use of transgenic animals in biomedical, agricultural, and biotechnological fields requires the ability to target gene incorporation and to control the timing and level of transgene expression. Experimental designs have taken advantage of the ability to direct specific (e.g. cell, tissue, organ specificity) as well as ubiquitous (whole-body) expression *in vivo*. Furthermore, from embryology to virology, transgenic technology provides unique animal models for studies in various disciplines that would otherwise be all but impossible to develop spontaneously.

1.2 Applications of Transgenic Animals

A number of methods exist for gene transfer in mammalian species, including DNA microinjection, embryonic stem (ES) cell transfer, retroviral infection, blastomere/embryo aggregation, teratocarcinoma cell transfer, electrofusion, nuclear transplantation, spermatogonial cell transfer, and perhaps spermatozoa-mediated transfer.

Transgenic technology has been extended to a variety of animal species in addition to the mouse, including rats, rabbits, swine, ruminants (sheep, goats, and cattle), poultry, and fish. Transgenic amphibians, insects, nematodes, and members of the plant kingdom have also been produced. With advances in the understanding of promoter/enhancer elements and transcription/regulatory proteins involved in the control of gene expression, the technology continues to evolve using different model systems (Table 1). In the systems explored to date, gene-transfer technology is a proven asset in science as a means of dissecting gene regulation and expression *in vivo*. As such, the primary questions that are addressed, concern the roles of individual genes in development or in particular developmental

Tab. 1 Applications of transgenic mice.

Transgenic mice have provided models in biomedical, veterinary, agricultural, and biotechnological disciplines in the study of gene expression and developmental biology, as well as for

- the fields of embryology, endocrinology, genetics, immunology, neurology, oncology, pathology, physiology, toxicology, and virology;
- the study of genetic bases of disease, and the design and testing of therapeutic strategies;
- gene therapy;
- disease resistance in humans and animals;
- drug and product efficacy testing/screening;
- novel product development through "molecular pharming"

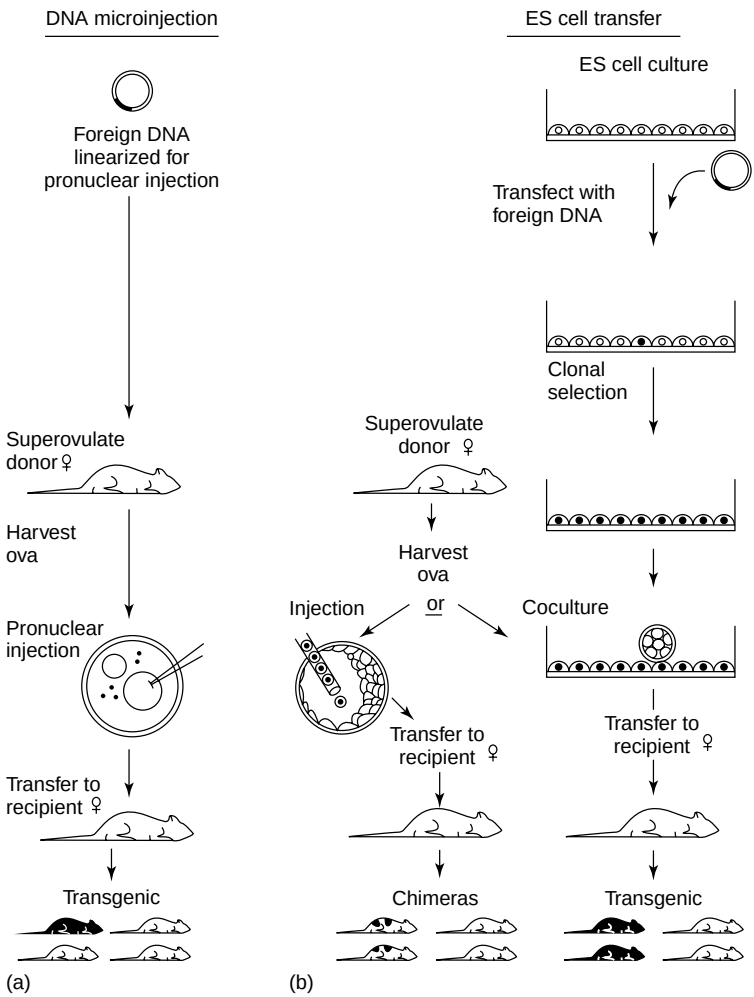

Fig. 1 (a) DNA microinjection and (b) embryonic stem (ES) cell transfer are the two most practiced methods for producing transgenic mice. Note that an *in vitro* culture step is not necessary for DNA microinjection; DNA is injected directly into the male pronucleus of fertilized one–cell ova (pronuclear zygotes). In general, if transgenic mice (represented by black mouse) are derived by DNA microinjection, all their cells will contain the new transgene(s). In (b), after clonal selection of transfected ES cells, one of two techniques is used for ES cell transfer. Ova are harvested between the eight-cell and blastocyst stages, ES cells are either injected directly into a host blastocyst (''injection'') or cocultured with eight-cell to morula stage ova, so that transfected ES cells are preferentially incorporated into the inner cell mass of the developing embryo (''coculture''). With blastocyst injection, transgenic offspring are termed *chimeric* because some of their cells are derived from the host blastocyst and some from transfected ES cells (denoted by white mice with black patches). Using coculture and tetraploid embryos, one can obtain founder mice derived completely from the transfected ES cells (denoted as all-black mice).

pathways. With this caveat, considerations include the ramifications of gene activity from intracellular to inter- and extracellular events within a given tissue or cell-type milieu.

Normally, gene function is influenced by cis-acting elements and trans-acting factors. For transferred genes, the cis- and trans-activators, in conjunction with the gene integration/insertion event within the host genome, influence regulation of both endogenous and transferred genes. Using genes that code for reporter proteins (e.g. oncogene or *lacZ*/βgal constructs), analysis of transgenic animals has revealed the importance of these factors in determining developmental timing, tissue distribution, and relative efficiency of gene expression. Transgenic animals have also proven quite useful in determining *in vivo* artifacts of other model systems or techniques.

There are a number of strategies in the development of transgenic mouse models, including systems designed to study dominant gene expression, homologous recombination/gene targeting and the use of ES cells, efficiency of transformation of eggs or cells, disruption of gene expression by antisense transgene constructs, gene ablation or knockout models, reporter genes, and marking genes for identification of developmental lineages.

1.3 Societal and Ethical Considerations

While gene-transfer technology continues to open exciting new biological frontiers, it also raises questions concerning regulation and commercialization. It is not within the scope of this presentation to fully address these issues; yet they do exist and will continue to influence the genetically engineered systems under development.

Contrary to the early prospects related to potential commercialization of this technology, there are societal challenges regarding perceived risks still ahead (Table 2). The risks at hand can be defined not only by scientific evidence but also in relation to public concern. Therefore, the central questions will revolve around the proper safeguards to employ and the development of coherent and unified regulatory guidelines. Can new animal reservoirs of fatal human diseases be created? Can more virulent pathogens be artificially created? What is the environmental impact of the "release" of genetically engineered animals? Do the advantages of a bioengineered product outweigh potential consequences of its use? These are but a few of the questions that we cannot ignore. Yet, the regulatory hurdles that exist today will challenge not only scientists and policy makers, but academicians, bioethicists, and legal scholars as well.

The strategies associated with the production of transgenic animals are quite variable, even between laboratories that utilize similar systems. However, most of the techniques for extensions to different systems are unique and timely

Tab. 2 Societal concerns regarding genetically engineered animals.

Environmental impact following "release" or commercialization of transgenic animals
Public perceptions of perceived risk
Ethical considerations of transgenic animal exploitation
Legislative oversight of technology
Safety of transgenic products, by-products, and foodstuffs
Patent aspects and product uniformity/economics of production

for new investigators. The overall efficiency of many procedures will vary, as will the cost–benefit ratios, however, one must not let the formidable challenges of gene-transfer experimentation outweigh the primary reason for instituting such studies – the desire to develop a biological model with specific genetic characteristics in a whole-animal environment.

2 Production of Transgenic Laboratory Animals

2.1 Choice of Animals

The relative importance of using particular strains or breeds of animals in gene-transfer experimentation will vary dramatically according to the species under consideration. Probably the most complex system is encountered in the production of transgenic mice, simply because so much work has been done with murine species. Here, well-documented differences in reproductive productivity, behavior, related husbandry requirements, and responses to various experimental procedures play major roles in determining the efficiency and the degree of effort associated with production of transgenic founder animals. A general discussion of these factors therefore serves as an appropriate starting point for understanding the many processes and procedures that must be evaluated and monitored when considering production of transgenic animals.

DNA microinjection, the most direct and reproducible method for producing transgenic animals, necessitates that several "pools" of animals be maintained for specific purposes. Most commonly, donor females are induced to superovulate by means of a regimen of injections of pregnant mare serum gonadotropin (PMSG, to stimulate follicular growth and development) followed by human chorionic gonadotropin (hCG, to induce ovulation of mature eggs). These donor females are then mated to fertile males, and large numbers of fertilized one-cell eggs ("zygotes") are obtained surgically. Alternatively, ova may be collected from donor females and then subjected to *in vitro* fertilization (IVF) to obtain zygotes. In either case, the DNA construct, in a buffer solution, is microinjected into the male pronuclei of the fertilized eggs. Ova that survive microinjection are surgically transferred to the reproductive tracts of hormonally synchronous female recipients, which carry the embryos to term. In mice, DNA from offspring is isolated from tissue samples (e.g. tail biopsies) obtained at weaning and analyzed by Southern blot analysis and/or methods based on the polymerase chain reaction (PCR) to determine which individuals carry the transgene.

Certainly, the relative costs of different strains of mice that are chosen to serve as gamete donors or recipients is a factor. This must be balanced with the background genetics of the various strains and the effects these may potentially exert on expression of the integrated, foreign DNA construct (the transgene) and the resulting phenotype. Strains used for donor females are chosen for (1) numbers of eggs ovulated in response to hormonal superovulation regimen, (2) the relative uniformity of this response within the strain, (3) the viability of superovulated eggs both before and after microinjection, (4) the ease with which fertilized eggs are visualized and recovered from the reproductive tract, and (5) the morphology and/or size of pronuclei (which affects the

ease and speed with which eggs can be microinjected).

The strain chosen for embryo transfer recipient use is evaluated for (1) parenting ability, (2) docile behavior (which allows for inspection of pups or cross-fostering after birth), (3) ease of embryo transfer procedures, and (4) the regularity with which females exhibit readily detectable vaginal plugs and undergo a pseudopregnant response when mated with vasectomized males.

Obviously, the genetic background of the strain used for vasectomized males (and also recipient females) is less important, since these animals do not contribute to the genetic composition of the transgenic offspring. Strains are favorably considered in which vasectomized males exhibit libido and are able to form visible copulatory plugs in the recipient females both frequently (two to three matings/week) and with regularity over a period of about 12 months.

Conversely, intact males are chosen for their (1) libido, (2) absence of nonmating aggressiveness when placed with families, (3) fertility and capacity to produce visible copulatory plugs, and, if serving as IVF donors, (4) sperm concentration and quality. Finally, it is desirable to maintain a pool of synchronous, pregnant females that can serve as foster mothers in the event that recipient females begin to cannibalize pups or fail to adequately nurse their offspring.

For most other nonmurine species, it will remain important to use donor females that respond well to hormonal synchronization and superovulation, and to have recipients that are able to carry fetuses to term and care for neonates appropriately. However, differences in methods used for hormonal synchronization, selection of proestrus females, and evaluation of other forms of reproductive behavior may differ significantly from those identified for mice.

2.2 Gain-of-function, Loss-of-function and Conditional Modeling

With advances in the characterization of factors that control gene expression, gene-transfer technology has become an important means of dissecting gene regulation and developmental pathways *in vivo*. Normally, endogenous gene function is regulated by a variety of molecular factors. Analysis of transgenic animals has illustrated molecular underpinnings and events associated with developmental timing, tissue distribution, and consequences of modifying gene expression. Additionally, transgenic animals have also proven quite useful in validating an array of *in vitro, in vivo*, and theoretical model systems. Today, there are multiple modeling capabilities that allow a researcher to develop a given hypothesis revolving around a gain-of-function experiment (e.g. DNA microinjection, retroviral-mediated transgenesis, or knockin gene targeting procedures are used), a loss-of-function model (e.g. knockout gene targeting or overexpression of cellular toxins may be envisioned) or conditional systems (e.g. use of traditional temporally or spatially regulated transgenes, or where Cre-lox or other *in vivo* recombinase systems are employed).

2.3 Methodologies

2.3.1 DNA Microinjection

The microinjection method, referred to earlier, involves the use of micromanipulators and an air- or oil-driven microinjection

apparatus to physically inject the DNA construct solution into the male pronuclei or nuclei of fertilized one- or two-cell ova. As such, the method is also sometimes referred to as *pronuclear microinjection*.

Virtually any cloned DNA construct can be used, albeit some caveats apply. Linearized (as opposed to closed, circular) DNA constructs are more readily integrated into the host genome. The presence of extraneous plasmid- or vector-related DNA sequences associated with the construct can adversely affect expression of the integrated construct/transgene. Also, greater care must be taken when handling larger DNA constructs, such as those derived from yeast artificial chromosomes (YACs), to avoid damage by shearing prior to and during microinjection.

With few exceptions, microinjected constructs integrate randomly throughout the host's genome, but usually only in single locations. This fact can be exploited to obtain functional linkage of independent transgenes by simultaneous coinjection of more than one DNA construct. In such a case, both constructs tend to integrate in the same randomly located site and may exhibit coexpression. Because integration is spatially random, transgenes may be integrated into either a somatic or a sex chromosome.

The integration process itself is poorly understood, but it apparently does not involve homologous recombination. During integration, single or multiple copies (1–200) of a transgene are incorporated into the genomic DNA, predominantly as head-to-tail concatemers. Regulatory elements in the host DNA near the site of integration, and the general availability of this region for transcription, appear to play a major role in affecting the level of transgene expression. This "positional effect" is presumed to explain why the levels of expression of the same transgene may vary dramatically between individual founder animals as well as their offspring (or "lines"). It is therefore prudent to examine transgene expression in offspring from at least three or four founder animals to determine what might be a result of position or incorporation site influences. Additionally, the presence of tissue- or cell-type-specific nuclear regulatory factors that act on host genes near the site of integration in cis or trans may restrict transgene expression to only a subpopulation of cells even though the transgene is present in virtually all cells of a founder.

The host DNA near the site of integration frequently undergoes various forms of sequence duplication, deletion, or rearrangement as a result of transgene incorporation. Such alterations, if sufficiently drastic, may disrupt the function of normally active host genes at the integration site, an activity that constitutes "insertional mutagenesis," wherein an aberrant phenotype may result. Such events cannot be purposefully designed, but they have led to the serendipitous discovery of previously unsuspected genes and gene functions.

Because gene transfer is accomplished at the one- to two-cell stage very early in embryonic development, the transgene is incorporated into essentially every cell that contributes to development of the embryo. Thus, transgenic animals (or "founders") produced by this method are usually considered to be "nonmosaic" or "nonchimeric," in the sense that the transgene is physically present in the nuclei of all cells of the body. The DNA or gene construct must be incorporated into the one-cell ovum (or a single viable blastomere) prior to DNA replication in order to appear in every descendent cell. If incorporation or integration occurs at

any later time point, then mosaicism can occur. Incorporation of the transgene into cells that will eventually contribute to development of germ cells (sperm or ova) is a common occurrence with this method, and it makes possible heritability of the transgene by offspring of founder animals. In such cases, the transgene has been said to have gone "germ-line." However, integration of the microinjected DNA construct into the host's genome occasionally may be inexplicably delayed. In such a case, if cells of the early embryo ("blastomeres") undergo mitosis before integration occurs, some but not all of the cells will contain the transgene, and the founder animal, although still considered to be transgenic, will be classified as a mosaic or chimera.

Advantages of the microinjection method are as follows: (1) the relatively high frequency of generating transgenic animals from viable microinjected embryos transferred to recipient females (20–30%), (2) the high probability of germ-line transmission of the transgene, (3) the relative lack of constraints on the size or type of DNA construct used, (4) the relative stability of the transgene as it is transmitted from generation to generation, and (5) the low frequency of mosaicism or double integrations (combined estimate of 10–30% of founders).

Disadvantages of this method include the following: (1) the random and potentially significant influence of the site of integration on transgene expression (positional effects), (2) the potential for undesired insertional mutagenesis, (3) the occasional production of mosaic founders, (4) the occasional lack of germ line incorporation, and (5) the time and expense required to obtain the skills necessary for micromanipulation and microinjection.

2.3.2 Retroviral Introduction of Transgenes

Transfer of foreign genes into animal genomes has also been accomplished using retroviruses. Although embryos can be infected with retroviruses up to midgestation, early eggs, usually at the 4- to 16-cell stages, are used for infection with one or more recombinant retroviruses containing a foreign gene. Immediately following the infection, the retrovirus produces a DNA copy of its RNA genome using the viral enzyme reverse transcriptase. Completion of this process requires the host cell to undergo the S phase of the cell cycle. Therefore, retroviruses effectively transduce only mitotically active cells. Incomplete infections, in which not all embryonic cells acquire the retrovirus, occur more frequently when using embryos after the 4-cell stage, with resultant chimeric embryos. Modifications to the retrovirus frequently consist of removal of structural genes, such as *gag*, *pol*, and *env*, which support viral particle formation. Additionally, most retroviruses and complementary lines are ecotropic in that they infect only rodents (e.g. rats, mice) and rodent cell lines rather than humans.

The DNA copy of the viral genome, or provirus, integrates randomly into the host cell genome, usually without deletions or rearrangements. However, as is the case for gene transfer by microinjection, because integration is not by way of homologous recombination, this method is not used effectively for site-directed mutagenesis.

Very high rates of gene transfer are achieved with the use of retroviruses. However, disadvantages include (1) a general limitation on the size of the foreign DNA insert that can be accommodated by individual viral strains, (2) biosafety

considerations, and potential for undesired genetic recombination that may alter the replicative characteristics of the retrovirus, (3) low copy number integration, (4) a high frequency of mosaicism (which necessitates subsequent outbreeding of founder animals to generate pure lines), and (5) possible interference by retroviral sequences on transgene expression.

2.3.3 Embryonic Stem Cell Technology

Gene transfer has been used to produce both random and targeted insertion or ablation of discrete DNA fragments into the mouse genome. For targeted insertions, where the integration of foreign genes is based on a recombinational gene insertion with a specific homology to cellular sequences (termed *homologous recombination*), the efficiency of DNA microinjection is extremely low. In contrast, the use of ES cell transfer into mouse embryos has been quite effective in allowing an investigator to preselect a specific genetic modification, via homologous recombination, at a precise chromosomal position. This preselection has led to the production of mice that (1) incorporate a novel foreign gene into their genome, (2) carry a modified endogenous gene, or (3) lack a specific endogenous gene following gene deletion or "knockout" procedures.

Technologies involving ES cells, and more recently primordial germ cells, have been used to produce a host of mouse models. Pluripotential ES cells are derived from early preimplantation embryos and maintained in culture for a sufficient period for one to perform various *in vitro* manipulations. The cells may be injected directly into the blastocoel of a host blastocyst or incubated in association with a zona-free morula. The host embryos are then transferred into intermediate hosts or surrogate females for continued development. Currently, the efficiency of chimeric mouse production results in about 30% of the live-born animals containing tissue derived from the injected stem cells. This ability to produce "chimeric" animals using ES cells has given researchers another tool in their armamentarium for the production of transgenic animals. In this set of techniques, the power of gene-transfer technology has been catapulted forward because such processes allow for targeted insertions into the genome. Such targeting is extremely important, particularly in areas of gene therapy and correction, wherein previous technologies allowed only for random integration events.

The genome of ES cells can be manipulated *in vitro* by introducing foreign genes or foreign DNA sequences by techniques including electroporation, microinjection, precipitation reactions, transfection, or retroviral insertion. The use of ES cells to produce transgenic mice faced a number of procedural obstacles before it became competitive with DNA microinjection as a standard technique in mouse modeling.

The first procedural obstacle in creating ES cell–derived transgenic animals was the extreme difficulty associated with the production and maintenance of ES cell lines. This problem lies in the general inability of embryos of most mammalian species, other than mice, to survive *in vitro*. A standard technique for producing murine ES cells involves culture of embryos *in vitro*, up to and beyond the point at which they would normally implant in the uterus. In mouse-embryo culture, this period is mimicked through coculture on fibroblast monolayers or by the inclusion of various growth factors, (leukemia inhibitory factor [LIF], ciliary neurotropic factor [CNTF], etc.) in the medium. In the case of coculture, fibroblasts serve as

a substrate for the attachment of developing embryos, enabling them to develop to a stage at which ES cells may be harvested. Additionally, although not completely understood, culture conditions and media/sera, as well as the handling of ES cells before and during transfer procedures, play critical roles in experimental efficiencies.

Following identification of true ES cells and subsequent gene-transfer procedures, it is necessary to accurately and efficiently identify ES cells that have integrated foreign DNA. In this regard, the use of marker-assisted selection schemes has greatly simplified mouse experimentation. Some genes were easily selectable *in vitro* and therefore made screening less difficult, but most were not. In these latter cases, PCR technology has been used effectively to screen for homologous recombination events. Mutated copies of normal endogenous gene sequences have been inserted in place of normal cellular sequences, and PCR techniques were used to screen for the novel mutants, with the capability of detecting even a single base-pair mutation. Hence, PCR techniques have proven to be valuable tools in selecting altered ES cell clones for use in transgenic animal production.

ES cell–chimeric animals may indeed harbor cells of both host embryo and ES cell derivation, extending the study of germ line transmission of desired transgenes to one or more additional generations, compared to most microinjection-derived transgenic animals. Concerns associated with the difficulty of line maintenance for DNA microinjection-derived transgenic animals are minimized, however, because the ES cells have the potential to give rise to multiple founder animals harboring the same transgene at the same chromosomal integration site. More recently, with the addition of coculture techniques involving tetraploid host embryos (eight-cell stage to morula), founders can be derived completely from the cocultured ES cells. Hence, the founders are no longer chimeras, having become "better engineered" models.

Still, for most modeling employed today, germ line transmission of transferred genes in ES cell derived mice is influenced to a great extent by the background genetics of host and donor strains of mice. Interestingly, germ line transmission is routinely problematic in some strains but not in others. The underlying biological mechanisms responsible for such differences are not known.

While putative ES cell lines have been identified for species other than the mouse, production of germ line ES cell derived/chimeric animals has proved difficult. If ES cells are identified that could give rise to germline–competent animals, it is probable that gene transfer into such cells would be successful – based on the wealth of available gene-transfer literature. Thus, the most obvious question to pose today is: Why has it been difficult to procure and utilize embryonic-derived pluripotential stem cells (ES cells) in mammalian species other than the mouse? Is it that present-day difficulties reflect an evolutionary mutation in the mouse genome, but not in other species, allowing immortalization of pluripotent cells? Or perhaps, might the idealized conditions just be beyond our immediate grasp?

2.3.4 Nuclear Transfer (Cloning) Methods

Nuclear transfer, or cloning, refers to the transfer of a nucleus from either an undifferentiated stem cell or a differentiated adult (somatic) cell to an enucleated oocyte, thereby producing a now-viable embryo. When this embryo is transferred into a

suitable female, it can develop to term and produce a live animal that is theoretically genetically identical to that from which the donor nucleus was obtained. If cloning can be done with differentiated cells from adult animals with desirable genotypic and phenotypic superiority, then many genetically identical copies can be produced as an important experimental or commercial resource.

The first cloned mice were identified in 1998 by the "Honolulu Method" of nuclear transfer using adult cells (cumulus cell nuclei). Furthermore, serial nuclear transfer resulted in cloned offspring from previously cloned animals, demonstrating conclusively that cells from cloned animals were not different from normal mice in their ability to create new clonal generations.

Embryonic stem cells, cumulus cells, cultured follicular epithelial cells, tail tip cells, fibroblast cells, neuronal cells, lymphocytes, fetal gonadal cells, Sertoli cells, as well as blastomeres up to the compacted morula stage have now been successfully used in nuclear transfer protocols. Yet surprisingly, at this time, cloning by nuclear transfer using adult donor cells has not been successful in nonhuman primates, suggestive of a comparative species-specific biological phenomenon; however, nuclear transfer was pioneered in domestic animals, and has now been demonstrated in a number of species to date.

In the Honolulu method, metaphase II oocytes are collected from the oviducts of superovulated donors and enucleated using a holding pipette to stabilize the oocyte and an enucleation pipette to aspirate the chromosome–spindle complex. A nucleus from a cumulus cell, with most of the cytoplasmic material removed, is then injected into the oocyte. Electrofusion has also been used in place of nuclear microinjection to reconstitute the embryo. Following activation, the embryos are cultured to the two-cell stage overnight and then transferred into oviducts of foster mothers.

The advantages of nuclear transfer over other techniques include the following:

- Most of the oocyte and embryo-handling procedures have already been worked out in the course of transgenic animal production.
- Because cumulus cells are relatively easily obtained and greater than 90% are in G_0/G_1 phase of the cell cycle, (no *in vitro* culturing required), cumulus cells are good nuclear donors for cloning in mice.
- Cloning of mice from ES cell nuclei combines the possibility of genetic manipulation of ES cells with cloning by nuclear transfer.

Yet, there are still a number of limitations to the current technologies. The principal concern, reported by a number of leading laboratories, is that cloning by nuclear transfer remains relatively inefficient. Well under 5% of ova and nuclear transfers with cumulus or somatic donor cells develop to term. Beyond the relative experimental efficiencies, potential reprogramming and developmental consequences of the various nuclear transfer related procedures are poorly characterized to date. Similarly, influences of epigenetic and mitochondrially encoded modifiers will undoubtedly influence potentially undesirable outcomes.

2.3.5 Other Methodologies

Today, a host of techniques have been applied to the genetic modification of mice. In the late 1960s and 1970s, blastomere or embryo aggregation and teratocarcinoma

cell transfer were used to transfer whole genomes into mouse ova. From the early 1980s, beyond DNA microinjection, retroviral infection, ES cell technologies, and nuclear transfer, we have also seen development of cytoplast and karyoplast fusions (including nuclear transfer), spermatozoa- and spermatogonial cell–mediated transfer, and particle bombardment and jet injection procedures. The later techniques expand upon earlier "building blocks" of genetic engineering technologies. As such, our capabilities and efficiencies will surely continue to evolve. Yet, there have been some conspicuous developments of general interest to the scientific community that may augment some of these existing methods.

Gain-of-function and loss-of-function modeling have, for the most part, concentrated on introducing specific mutations into the nuclear genome. From a gene ablation standpoint, creation of loss-of-function models will be facilitated by the emerging technology of RNA interference (RNAi). Short, interfering RNA (siRNA) exists in a double-stranded state and inhibits endogenous genes (and/or exogenous sequences as in viral genes) due to a complementary sequence homology. RNAi technology has potential applications including the inhibition of viral gene transcription and inhibition of endogenous genes coding for deleterious gene products. Indeed, germ line competent transgenic mouse and rat models using RNA interference have been shown to recapitulate null phenotypes.

In contrast to various methods to target modification of the nuclear genome, little attention was focused until recently, on the importance of mitochondrial genetics and the mitochondrial genome in mammalian development. This omission, in part, is related to the difficulty associated with *in vivo* mitochondrial transfer. Without an approach in hand, a significant technological hurdle remained in the identification of mitochondrial gene targets appropriate for engineering or modification. For a host of applications, the ability to manipulate the mitochondrial genome and to regulate mitochondrial gene function would provide an additional target in modifying mammalian development.

We and others have initiated studies revolving around mitochondrial transfer and techniques to produce animals harboring foreign mitochondrial genomes. The creation of transmitochondrial animals represents a new model system that will provide a greater understanding of mitochondrial dynamics, leading to the development of genetically engineered production animals as well as therapeutic strategies for human metabolic diseases affected by mitochondrial mutation or function.

2.4 Other Laboratory Animal Models

Transgenic animal protocols developed in mice have been modified to accommodate production of other transgenic species. In particular, the larger body sizes of rats and rabbits, their short estrous cycles and gestation lengths, their relatively rapid generation times, and their litter-bearing ability have helped make these animals preferred models for several areas of research employing gene transfer. As with mice, DNA microinjection has been the method used most frequently for introduction of foreign DNA. Differences between these species and mice in the size of ova, physical response to microinjection, requirements for *in vitro* ova culture, and numbers of fertilized eggs needed for embryo transfer (to maintain pregnancy), as well as differences in general husbandry

practices are well documented and used to optimize production of transgenic rats and rabbits.

In some cases, specific physiological characteristics unique to a species have made it a more appropriate or successful model for a particular disease; such was the case for lines of transgenic rats exhibiting hypertension-induced symptoms and pathology resulting from overexpression of specific transgenes (viz. a mouse *Ren*-2 renin transgene or a human angiotensinogen gene construct) that were not similarly expressed in transgenic mice.

Both transgenic rats and rabbits have provided models for the study of specific cancers: transgenic rats expressing an SV-40 T-antigen construct controlled by an albumin promoter develop hepatocellular carcinomas. Transgenic rabbits in which expression of a c-*myc* transgene construct is driven by a rabbit immunoglobulin enhancer exhibit malignant lymphocytic leukemia. Particular strains of rabbits have historically served as models for the study of lipid and lipoprotein metabolism. In this tradition, lines of transgenic rabbits bearing human apolipoprotein A-I antisense transgenes have been generated for the study of the biosynthesis of high density lipoproteins.

Finally, several studies have proven rabbits to be acceptable models for "gene pharming." In these experiments, foreign proteins are expressed in transgenic animals, preferentially in mammary tissue, and are secreted into milk, from which they can be readily purified. Rabbits are considered to be suitable models for such use for three reasons: lactating females produce relatively large volumes of milk daily, rabbit milk possesses a high protein concentration, and they tend to nurse their offspring infrequently. To date, transgenic rabbits have been separately produced, in which human proteins (e.g. α_1-antitrypsin, interleukin 2, growth hormone, tissue plasminogen activator) and bovine α-lactalbumin have been expressed in mammary tissue and harvested from milk.

3 Production of Transgenic Domestic Animals

3.1 Overview

The success of transgenic mouse experiments led a number of research groups to study the transfer of similar gene constructs into the germ line of domestic animal species. These efforts have focused on addressing one of the three research goals: improving the productivity traits of domestic food animal species, development of transgenic animals for use as bioreactors (i.e. producers of recoverable quantities of medically or biologically important proteins), and in developing tissues, organs, or reagents for transplantation-related endpoints. Toward these goals, a number of strategies have been employed to date. Such strategies run the gamut from molecular modeling (systems designed to study dominant gene expression, homologous recombination/gene targeting and the use of ES cells) to cell biology (efficiency of transformation of eggs or cells) to blocking or ablating gene function (disruption of gene expression by antisense or RNAi transgene constructs, gene ablation or knockout models). Since 1985, numerous studies have focused on transgenic farm animals created using growth-related gene constructs. Unfortunately, for the most part, deleterious consequences of various manipulations have failed to identify commercially useful animals because of an inability to coordinately regulate

gene expression – in deference to the ensuing cascade of endocrine events that unfolded. In contrast, efforts at using domestic farm animals as bioreactors or in transplantation-related efforts have continued to progress, and viable products are expected to petition for regulatory approval shortly. Whether they will gain market acceptance is still largely unknown.

3.2 Methodology

Today, DNA microinjection, retroviral transfection, nuclear transfer and passive transfer procedures (e.g. sperm-mediated transfer) have been used to successfully produce transgenic livestock. Genetic-engineering capabilities will continue to mature, in concert with advances in whole-animal and somatic-cell techniques (including liposome-mediated gene transfer, jet injection, and particle bombardment). Envisioned progress and expansion of our existing knowledge base will allow us to better maximize and engineer production traits in farm animals in a most effective fashion.

Along the same lines, procedures to augment gene-transfer efficiencies have a high priority, especially where long gestational intervals prove problematic. Protocols have been developed to permit removal and screening of individual blastomeres from preimplantation embryos maintained in culture prior to transfer to recipient females. Sensitive methods involving the use of conventional and real-time PCR allow analysis of DNA purified from individual blastomeres and subsequent identification of those embryos that bear genetic modifications of interest. Use of such methods offers the potential to greatly increase the efficiency associated with production and selection of transgenic farm animals and to thereby significantly reduce the associated costs.

3.2.1 DNA Microinjection

DNA microinjection was the first method used to successfully produce transgenic livestock. Although involved and at times quite tedious, the steps in the development of transgenic models are relatively straightforward. Once a specific gene sequence is cloned and characterized, sufficient quantities are isolated, purified, and tested in cell culture. If *in vitro* mRNA expression of the gene is identified, the appropriate fragment is linearized, purified, and readied for preliminary mammalian gene-transfer experiments. The DNA microinjection experiments are first performed in the mouse model. While the transgenic mouse model will not identify likely phenotypic expression patterns in domestic animals, we have not observed a single construct that would function in a pig when there was no evidence of transgene expression in mice. Therefore, preliminary *in vitro* and/or *in vivo* experimentation has been a crucial component of any gene-transfer experiment in domestic animals.

Following confirmation of transgene expression in mice, it then becomes cost-effective to initiate DNA microinjection experiments in other species. In mouse experiments, less than two months is required from the time the purified construct is ready for microinjection through weaning of founder pups. In contrast, for pig experiments, one month to a year is required for a sufficient number of DNA injections and recipient transfers to ensure the likelihood of success. Experimental efficiencies coupled with a long generational interval (i.e. 114 day gestation period, 21–28 day lactation, and onset of puberty between 6–9 months of age), reflect the efforts necessary to

identify and characterize transgenic pigs and illustrate the extended time lines associated with the production of any transgenic livestock model. In addition, the time frame from birth of pigs or cattle to the establishment of lines can be 1–2 years to 4–5 years respectively (while also dependent on the sex of founders). Hence, there is an obvious advantage to characterizing transgenic mouse models to expedite what will ultimately be a lengthy undertaking.

3.2.2 Retroviral Infection

Retroviruses have been used as vectors for transfer of foreign genes into domestic animal genomes, and are particularly exciting because of the promise of 100% transfer efficiency. Although embryos can be infected with retroviruses up to midgestation, early eggs, from oocytes to 16-cell stage ova, are used for infection with one or more recombinant retroviruses containing a foreign gene. Immediately following infection, the retrovirus produces a DNA copy of its RNA genome using the viral enzyme, reverse transcriptase. Completion of this process requires that the host cell undergoes the S-phase of the cell cycle. Therefore, retroviruses effectively transduce only mitotically active cells. Modifications to the retrovirus frequently consist of removal of structural genes, such as *gag, pol,* and *env*, which support viral particle formation. Additionally, most retroviruses and complementary lines are ecotropic in that they infect species-specific cell lines, limiting risk to humans in animal experimentation.

Advantages of retrovirus-mediated gene transfer include the technical ease and efficiency of existing methods. Very high rates of gene-transfer are achieved with the use of retroviruses, thus, minimizing the numbers of ova and animals needed for gene-transfer studies. Additionally, integration of viral genes does not appear to induce genetic rearrangements of the host genome (occasionally reported with other methods), with the exception of a possible short duplication at the site of integration.

However, of the disadvantages listed in Sect. 2, the first two limitations appear to raise the greatest concern at this time. From a size perspective, the maximum permissible size for efficient encapsidation or reverse transcription of each retroviral vector is approximately 10 kb. Because introns of foreign genes carried by retrovirus vectors are spliced out during replication, the insert does not have to contain introns, yet, intronic sequences can be critical for optimal gene expression and regulation. From a biosafety perspective, disabled retroviruses may become replication-competent via potential recombination events that may occur between or within retroviral vector, packaging cell lines, and the host genome. Biocontainment and safety considerations are still debated. Yet, work to date has illustrated and further characterized the utility of sensitive analytical methods to detect unexpected viral transmission and recombination risks. In addition to legislative and regulatory oversight, conscientious mechanisms now in place help ensure that proper methods and safeguards are incorporated into ongoing experimentation.

3.2.3 Nuclear Transfer

Cloning is generally associated with nuclear transfer whereby a nucleus provided by a donor cell is introduced into an enucleated oocyte (unfertilized ovum) or zygote allowing for reprogramming of the developing embryo. It should be noted that nuclear transfer, with nuclei obtained from either mammalian stem cells or differentiated adult cells, is an especially important

development in species beyond the mouse model. This is because a technological barrier was surpassed that allowed for specific *in vitro* manipulations that lead to targeted genetic modifications in first-generation founder animals. Prior to the mid-1990s, it was not possible to produce germ line–competent transgenics in mammalian species other than in mice, using any technique other than DNA microinjection (that only allowed for random and imprecise integration of transgenes in founder animals). Unfortunately, relative efficiencies for nuclear transfer experimentation still pales in comparison to conventional DNA microinjection. However, while nuclear transfer might be considered inefficient in its current form, major strides in enhancing experimental protocols within the next few years are envisioned, comparable perhaps to the early advances in DNA microinjection technology.

Nuclear transfer involves a technique in which nuclei of various cell origins are introduced as noted in Sect. 2 above (Fig. 2). The successful "cloning" of a sheep was reported in the mid-1990s and rekindled the imagination of researchers struggling with microinjection and related technologies. Yet, while nuclear transfer might be considered inefficient in its current form, major strides in enhancing experimental protocols have been envisioned, and with recent *in vitro* maturation of pluripotential cell lineages, the marriage of a number of existing techniques may prove fruitful in the near term.

The techniques for nuclear transfer in domestic animals mirror what is currently employed for mouse studies. Nuclei from quiescent (G_0) cells (or the intact

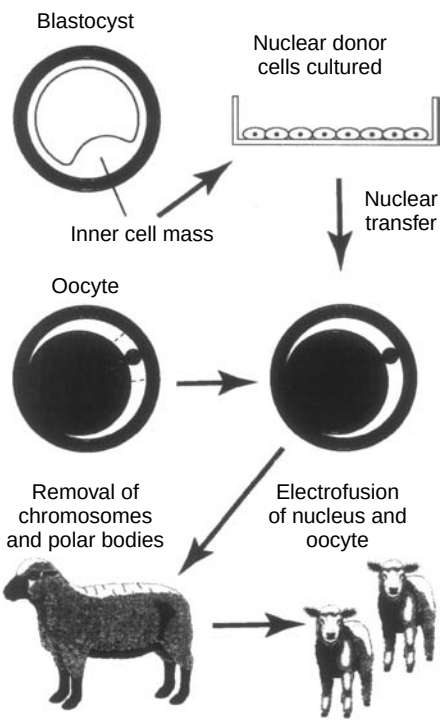

Fig. 2 Cloning sheep by nuclear transfer. In this example, inner cell mass cells from blastocyst stage embryos (or cells from somatic tissues) are obtained and propagated in culture. If so desired, the culture capability here, allows one to genetically modify cells prior to transfer. Selected cells are then used as nuclear donors for transfer into enucleated oocytes. In contrast to DNA microinjection, a fusion step is generally employed to fuse the transferred nuclei and enucleated oocytes. Here, electrofusion is used to fuse couplets (a donated nucleus plus enucleated oocyte) that are then transferred to recipients for the remainder of gestation. (Reprinted from Pinkert, C.A., (2000) *Reproduction in Farm Animals*, 7th ed., Hafez, E.S.E. and Hafez, B., Eds., Williams & Wilkins, Baltimore, Chap. 21, pp. 318–330. With permission.)

cells themselves) can be transferred by electrofusion, polyethylene glycol-induced fusion, or by direct injection into enucleated metaphase II oocytes. Following culture, the reconstituted embryos are transferred to the reproductive tracts of host animals. Since *in vitro* culture of reconstituted sheep embryos is relatively inefficient, they can instead be transferred to an intermediate (temporary) host. After *in vivo* incubation, they can be recovered and transferred to the final host to develop to term.

If cloning can be done with differentiated cells from adult animals with proven phenotype/genotype superiority, then many genetically characterized clones can be produced as an important experimental/commercial resource (particularly for the requisite "between batch" consistency needed for biomedical reagent production). However, there are still a number of procedural hurdles to reduce the costs associated with nuclear transfer – extending the overall utility of the methodology. With both sheep and cattle, there has been a very low success rate with a high percentage of fetal loss. With sheep, especially using adult donor cells, there is a significant loss of embryos, fetuses, and newborns and the process is very inefficient. Only 10–50% of nuclear-transferred embryos develop to the blastocyst stage, independent of *in vitro* or *in vivo* developmental concerns and again, there is less than a 5% yield in total live offspring (on the basis of embryos needed to generate founder animals). Lastly, the developmental consequences of nuclear transfer procedures are still controversial, with issues regarding birth weight of cloned animals, premature aging phenomena, and relative health and immunocompetence coming under significant scrutiny as various animal models entertain commercialization prospects.

3.3 Traits Affecting Domestic Animal Productivity

Interest in modifying traits that determine productivity of domestic animals was greatly stimulated by early experiments conducted by Ralph Brinster, Richard Palmiter, and their colleagues, in which body size and growth rates were dramatically affected in transgenic mice expressing growth hormone transgenes driven by a metallothionein (MT) enhancer/promoter. From that starting point, similar attempts followed in swine and sheep studies to enhance growth by introduction of various growth hormone (GH) gene constructs under control of a number of different regulatory promoters. Use of these constructs was intended to allow for tight regulation of individual transgene expression by dietary supplementation. However, although resulting phenotypes included altered fat composition, feed efficiency, rate of gain, and lean/fat body composition, they were accompanied by undesirable side effects (e.g. joint pathology, skeletal abnormalities, increased metabolic rate, gastric ulcers, infertility). Such problems were attributed to chronic overexpression or aberrant expression of the growth-related transgenes and could be mimicked, in several cases, in normal animals by long-term treatment with elevated doses of GH.

Subsequent efforts to genetically alter growth rates and patterns have included production of transgenic livestock harboring a vast array of genes – from growth-related transgenes to immune system modulators. Other productivity traits that are major targets for genetic engineering include altering the properties or proportions of caseins, lactose, or butterfat in milk of transgenic cattle and goats, more

efficient wool production, and enhanced resistance to viral and bacterial diseases (including development of "constitutive immunity" or germ line transmission of specific, rearranged antibody genes). Cumulatively, it has become apparent from these studies that greater knowledge of the biology of animal development will be required before lines of domestic animals with these desired characteristics can be genetically engineered.

3.4 Domestic Animals as Bioreactors

The second general area of interest has been the development of lines of transgenic domestic animals for use as bioreactors. One of the main targets of these so-called gene pharming efforts has involved attempts to direct expression of transgenes encoding biologically active human proteins. In such a strategy, the goal is to recover from serum or from the milk of lactating females large quantities of functional proteins that have therapeutic value. To date, expression of foreign genes encoding α_1-antitrypsin, tissue plasminogen activator, clotting factor IX, and protein C were successfully targeted to the mammary glands of goats, sheep, cattle, and/or swine.

Targeting expression of the cystic fibrosis transmembrane conductance regulator (CFTR) gene to the mammary epithelium, in which the protein becomes associated with the apical membranes of these cells, has resulted in apocrine secretion of the CFTR protein as a constituent of the milk fat globule membrane. The demonstrated utility of such a targeting strategy has led to proposals for largescale production and harvesting from milk of other classes of important membrane-associated proteins, such as hormone receptors, channel proteins, and transport proteins.

Similarly, lines of transgenic swine and mice have been created, which produce human hemoglobin or specific circulating immunoglobulins. The ultimate goal of these efforts is to harvest proteins, from the serum of transgenic animals, for use as important constituents of blood transfusion substitutes, or for use in diagnostic testing.

3.5 Examples from Domestic and Miniature Pigs

In contrast to gene transfer in mice, the efficiency associated with the production of transgenic livestock, including swine, is quite low. However, two advantages offered by swine over other domestic species include a favorable response to hormonal superovulation protocols (20–30 ova can be collected on average) and a significant uterine capacity, in that swine are a litter-bearing species. The highest transgenic success rate in domestic animals has been with outbred domestic pigs, while use of inbred strains was reported to have lower overall efficiency. Since the description in 1985 of the first transgenic livestock species, DNA microinjection has to had been the only successful method identified to produce germ line competent transgenic livestock prior to nuclear transfer.

An initial problem encountered during the creation of transgenic farm animal species concerned the visualization of the pronuclei or nuclei within the ova. In lipid-dense swine ova, for example, the cytoplasm is opaque and the nuclear structures are not discernible without some type of manipulation. This was a critical problem to overcome because transgenic pigs were not produced following injection

of DNA into the cytoplasm of one- and two-cell ova. Fortunately, it was found that centrifugation of pig ova resulted in stratification of the cytoplasm, rendering pronuclei and nuclei visible under differential interference contrast or Nomarski/Smith microscopy. Even though the situation was greatly improved, centrifugation failed to reveal the nuclear structures in 15–33% of fertilized eggs. However, when using conditions initially optimized for transgenic mouse production, DNA microinjection into pronuclei/nuclei resulted in nearly a 10% yield of transgenic offspring (vs. total live-born/viable offspring).

The proportion of transgenic swine that develop from microinjected ova, which varies from zero to about 12%, is generally much lower than that observed in mice. The survival of microinjected pig embryos is related to several factors including the developmental stage of ova injected, the duration of *in vitro* culture, synchrony of ova donors and recipients, the number of ova transferred, and donor age. Other factors that have been shown to influence the developmental efficiency of transgenic mouse production – including technician proficiency and embryo handling/transfer, pipette dimensions, DNA preparation, and the viability of microinjected ova – readily influence transgenic production efficiency for all other species.

If a foreign gene fragment is integrated into a domestic animal's genome, the gene copy number and orientation (head-to-head vs. head-to-tail arrays) vary considerably, as might be anticipated from the mouse studies. In addition, mosaic animals are sometimes produced, making analysis of transgene integration more difficult. Recently, the results of PCR analyses of founder animals were useful in detecting extreme mosaicism, where a given transgene is found in only a proportion of the animal cells. While the identification of transgenic founders (both mosaic and nonmosaic) can be enhanced using PCR analyses, confirmatory DNA slot-blot or Southern blot analyses are necessary to avoid extreme and costly breeding efforts to identify germ line transmission of the transgene.

In one set of experiments, approximately 70% of pigs harboring a growth hormone gene construct (as determined by Southern blot analysis) expressed the transgene-encoded growth hormone. Failure of the remaining 30% of transgenic swine to express the foreign gene was attributed to integration of the transgene into an inactive chromosomal locus or alteration of the transgene sequence during its integration. Of the pigs that did express the transgene, the level of expression appeared to vary greatly among individuals. While cis- and trans-acting factors are obviously important influences affecting transgene expression, the integration site of the transgene is equally critical. Transgene transcription rates are likely influenced by the level of activity present at the locus, or the site of integration, and the properties of enhancer sequences located in genes flanking the transgene.

While it is not unreasonable for an investigator to suggest that 40 transgenic mice can be produced in a week's worth of DNA microinjections, there are significant differences in undertaking pig experiments. In the pioneering collaboration of the University of Pennsylvania, the University of Washington, and the United States Department of Agriculture, which first reported on the production of transgenic pigs, more than four years went into amassing the experimental results that culminated in the production of 40 founder transgenic pigs. The 55% of founder animals that expressed transgene-encoded mRNA was

very much in line with results reported by other groups.

Currently, estimates for the cost of production for each transgenic pig run from $25 000 to $125 000 (U.S.$), with the broad range reflecting the low overall experimental efficiency. The costs are compounded (doubled or even tripled) if functional transgenes and pig survival are factored into the cost estimates. To address the efficiency of the gene-transfer experiments, a modified scheduling can be employed to decrease some of the fixed and variable costs, particularly in the use and scheduling of research animals. By overlapping days of experimentation rather than synchronizing one group of donor animals per week, actual experimental costs can be reduced by more than 25%. Superovulation can induce an average of 25–30 embryos per sow. While cost-effectiveness dictates the specific regimen, all pigs do not respond similarly, and only 60% respond within a specific window of opportunity for embryo collection. This low response rate may appear unacceptable, but it should be noted that the actual yield of suitable embryos is still higher than might be available using naturally synchronized females or other superovulatory schemes.

Surgical transfer of microinjected embryos usually follows the microinjection efforts. The mechanical insult to the embryos is severe, however, and only 15–25% of them will still be viable 5 days after transfer. Therefore, 30–50 microinjected embryos are routinely transferred per recipient sow, with the expectation that 50% of the recipient females will maintain pregnancy. While the number of embryos transferred may seem excessive, the basis is derived from classical studies that establish a requirement of four viable embryos at the time of implantation for a sow to initiate and maintain successful pregnancy.

For transgene expression studies targeting pigs, the use of outbred domestic pigs is the most practical way to produce and evaluate potential models. However, miniature or laboratory swine are now used with increasing frequency in biomedical research in which desirable, well-characterized background genetics make these animals more suitable for human modeling studies (e.g. xenotransplantation research). While miniature swine are commercially available, research in developing methodologies to explore or enhance reproductive performance has been limited. Reproductive efficiency in miniature swine is low compared to commercial swine and is characterized by a low ovulation rate, low birth weight, and small litter size. In addition, extensive management is required to maintain neonates. Average litter size is between 4 and 7 pigs at birth, with each breeding sow producing 12 to 18 pigs per year. Estrous cycles and gestation length are similar to standard commercial swine; however, sexual maturity in males and females occurs between 4 and 6 months of age in some breeds, which is sooner than the 6- to 9-month norm of commercial swine.

It is important to evaluate the potential breed differences and ultimate production efficiencies using a superovulatory protocol on groups of donor animals, to be able to assess and possibly enhance the reproductive performance prior to embarking on a gene-transfer program. Trials were conducted using Yucatan miniature pigs to ascertain the feasibility of a superovulation regimen, artificial insemination, and embryo transfer to standard commercial crossbred females. Yucatan semen was collected and used to successfully fertilize ova obtained by hormonal stimulation of donor females. As it happened, only 25% of the potential eggs were recoverable and

usable (fertilized one- and two-cell ova). However, fertilized ova were obtained with enhanced efficiency compared to spontaneous ovulations in both young and aged miniature pigs.

Breeding requirements for transgenic pigs are critical, and successful management calls for expertise in fields ranging from record keeping to animal husbandry. With a gestation length of 114 days, and an interval from birth to puberty of 6–12 months, the generational interval seems extraordinarily long, particularly when one is accustomed to the experimental progress observed in mouse studies. For xenotransplantation studies, the desire to breed animals to homozygosity for the transgene poses additional challenges. Breeding schemes have been reported for production efforts that, even under ideal circumstances, lead to a three- to five-year timetable (after founders have been produced and identified), assuming no lethality associated with transgene integration and insertional mutations. There are a number of additional questions that may arise in the breeding efforts. If transgenes interact, for example, their effects may not be cumulative; indeed, it might happen that such effects mask one another, perhaps in competition for trans-activators. Other issues involve stability of the genome as well as limiting the gene pool in intricate breeding schemes. Additionally, from a technical viewpoint, evaluation of homozygosity of low copy number or single-copy transgenic animals can be very difficult, with an additional breeding/generational cycle needed to ensure the homozygous status of a transgene in any particular animal.

In future experiments with domestic animals, the ability to utilize embryonic stem cells could represent a tremendous improvement over the microinjection protocols currently in use. The ability to "knockout" or replace endogenous genes has been an extremely valuable tool in mouse genetics and modeling. Yet, characterization and use of embryonic stem cells for gene transfer has not worked to date in any species other than the mouse. The need for such technology in domestic animal species is great, and the perseverance by the groups focused on this work is exemplary.

4 Analysis of Transgene Integration and Expression

4.1 Transgene Integration

When founder mice are three to four weeks of age, biopsies are readily performed to obtain representative tissues for analysis of transgene integration and expression. For integration analysis in mice, typically tail cuts (biopsies) are performed and the tissue is digested with proteinase K followed by DNA extraction using standard protocols. Preliminary determination of transgene integration by the PCR technique can be very useful when the target sequence, the transgene, possesses unique sequences (not endogenous to the genome of the animal). However, given the extreme sensitivity of PCR and the potential for mispriming and false positives, other more specific techniques should always be used to confirm the presence of the transgene in PCR-positive samples. These include DNA slot-blotting and Southern blot hybridization, the latter being not only more informative but also less likely to present false positive signals. Additionally, the transgene can be constructed to include a molecular "tag," a unique sequence

that is easily detected and has minimal similarity to any endogenous sequence. This strategy is especially useful when the transgene itself is similar or nearly identical to an endogenous sequence. In such cases, another strategy that may be helpful is the introduction of new restriction sites into the transgene, without perturbing function, so that restriction fragment length polymorphism (RFLP) analysis can be used to distinguish the transgene from its endogenous counterpart. Finally, in some instances phenotypic screening is possible if the expression of the transgene leads to an identifiable change in the appearance of the animal. Phenotypic screening by coinjection of a second transgene that gives rise to a specific phenotype can also be useful, since cointegration is much more likely than integration of the phenotypic marker gene alone. Later, however, the integration of the transgene of interest must be confirmed by other means.

Other concerns related to the analysis of transgene integration include determinations of copy number of the transgene per cell, orientation of tandemly arranged copies, and the presence of multiple integration sites. These questions can be addressed by Southern blot hybridization following digestion of the genomic DNA with appropriate restriction enzymes. In addition, the functionality of the transgene is an important consideration in the analysis of transgene integration. However, this aspect is usually dealt with in the analysis of transgene expression (detection of a functional mRNA transcript or protein product). Finally, the critical question of whether the transgene is integrated into the germ line of the founder animal and is heritable as a stable genetic element must be answered.

4.2 Transgene-encoded mRNA Expression

The analysis of expression of the transgene is, of course, absolutely essential in determining the utility of the transgenic animals produced. As with integration analysis, the presence or absence of similar or identical endogenous counterparts will determine, to a degree, the strategies that may be most useful. For transgenes that are unique (no endogenous counterpart) or contain some unique sequences, the strategies that can be used are more straightforward. The presence of a novel mRNA transcript or a unique protein (or enzyme activity) is more easily determined than in situations involving transcripts or protein products that are very similar to endogenous transcripts or proteins. As with integration analysis, molecular "tags" are also sometimes useful in that the transcripts will contain some unique identifying sequence that can be readily and unequivocally determined. Once again, care must be taken to ensure that any additions or deletions designed to aid in analysis do not interfere with expression.

Technically, the most critical step in analyzing transgene expression is the isolation of RNA. Great care must be taken to avoid contamination of RNA preparations with ribonucleases (enzymes that degrade RNA). Ribonucleases are present in significant amounts in a variety of locations from which they may gain access to samples of purified RNA (e.g. the skin surface, saliva aerosols from coughing or sneezing, and airborne dust). Once a high quality RNA preparation has been obtained from the appropriate tissue, the presence of mRNA transcripts of the transgene can be determined by RNA slot-blot hybridization. A more informative

technique, Northern blot hybridization, confirms not only the presence but also the size of the transcript of interest. The Northern blotting technique consists of electrophoresis into an agarose gel to separate the RNA molecules on the basis of size, transfer (blotting) of the RNA to a nylon or nitrocellulose membrane, and detection of the mRNA of interest by hybridizing it to a specific, radiolabeled probe, followed by autoradiography.

Additional techniques exist for determining the presence of relative levels of mRNA transcripts from transgenic animals. In the nuclease protection assay, a radiolabeled probe is allowed to hybridize to the RNA in solution, followed by nuclease digestion of unhybridized RNA. The sample is then resolved on a polyacrylamide gel, followed by autoradiography to detect the probe. This technique is very useful in determining the steady state levels of RNA in a given tissue and offers an inherently high signal-to-noise ratio. In the reverse transcription-PCR (RT-PCR) assay, the RNA of interest is transcribed into a cDNA molecule by means of a specific primer and the enzyme reverse transcriptase. A second primer is added, and a standard PCR amplification is performed.

The advantage of RT-PCR lies in its extreme sensitivity – theoretically, one mRNA molecule can be amplified to a quantity sufficient for visualization on an agarose gel after ethidium bromide staining. The drawbacks include the possibility of genomic DNA contamination, in addition to the concerns associated with standard PCR reactions (e.g. false positives, mispriming, contamination, etc.).

In the *in situ* hybridization technique, a labeled probe is hybridized to a target mRNA transcript in sections of tissue to permit the identification of individual cells containing the transcript. This technique is particularly useful in identifying subsets of positive cells within a given tissue in which the relative levels of specific transcript for the tissue as a whole might be too low to be detectable by other means.

4.3 Protein Expression

In addition to the foregoing assays for transcription of the transgene, various other techniques to determine the translation of the transcript into a protein product are often employed. Most of these depend on the availability of specific antibodies to the protein of interest. The immunoblotting technique, in which proteins are resolved on a polyacrylamide gel, transferred to a membrane, and detected by means of a labeled antibody, is useful in verifying the appropriate molecular weight of the protein product of interest. Although rarely used, the radioimmunoprecipitation assay can also serve in the determination of the size of the protein product and to allow for various other manipulations, including peptide mapping and isoelectric focusing, to verify proper modification of the protein. To identify the cells within a tissue that contain the protein product of the transgene, immunohistochemical "staining" of tissue sections with a labeled antibody is often employed.

Some experiments involving transgenic animal production focus on measuring the influence of various cis-regulating DNA sequences; the transcript and ultimate protein product are for the most part irrelevant. In these cases, "reporter genes" are often employed because they can simplify determination of expression levels by producing a transcript or protein that is easily and unequivocally detectable. Particularly useful in these applications are reporter genes that code for novel enzymes

in that the activities of the enzymes can be assayed, thus greatly increasing the sensitivity of detection.

In analyzing the expression of any transgene, it is always important to evaluate at least two separate lines of transgenic animals, to show that the expression patterns are consistent and reproducible. It is not uncommon for the site of integration within a given line of transgenic animals to have profound influence on transgene expression, independent of any transcriptional control sequences in the transgene itself.

5 Conclusions and Future Directions

5.1 Gene Transfer Today

The efforts and costs for gene-transfer experimentation in animal biology are challenging and limiting. Innovative technologies to enhance experimental gene-transfer efficiency in different species are desperately needed. Such enabling techniques would not only bring the cost of individual projects into a reasonable realm but would also increase the likelihood of breakthrough studies in many disciplines. As stated earlier, aside from mouse modeling, DNA microinjection into preimplantation embryos has been the only reproducible means for heritable gene transfer in the preponderance of mammalian species, other than multigene-targeted (nonspecific) blastomere aggregation and nuclear transplantation studies.

The outlined efficiencies for the production of transgenic animals represent what is considered to be current ''state-of-the-art'' technology. It is envisioned that procedures will be modified and enhanced as new research efforts are reported. Under the circumstances, there are a number of specific achievements that would significantly enhance experimental productivity. Enabling technologies would center on the following areas:

- development of alternate DNA delivery systems (e.g. embryonic stem cell transfer for stable gene transfer, liposome-mediated gene transfer, or perhaps targeted somatic-cell techniques);
- the identification of optimal experimental conditions for a given form of gene transfer and the identification of animal strains best suited to the individual technologies;
- complete animal genome mapping and identification of homologies to human genes;
- the establishment of routine and efficient embryo and germ plasm culture systems;
- the development of a means to reduce the number of animals and embryos required (e.g. use of PCR amplification of blastomere DNA or fluorescence-activated cell sorter (FACS) analyses and gating of transfected germ plasm);
- enhancement of relative experimental efficiencies to minimize the numbers of animals, gametes, and related reagents necessary to generate founder animals or experimental models.

Primary considerations for genetic manipulations and gene transfer are not limited to donor and recipient requirements, surgical manipulations, embryo handling, and gene-transfer skills, but also include expertise in tissue/cell culture and molecular biological techniques. For any animal experimentation, ethical considerations, regulatory requirements,

and experimental design are also important considerations that may limit research efforts. For most species, husbandry considerations are extensive in comparison to mouse requirements, running the gamut from animal and record-keeping management to female- (e.g. controlled estrus cycles/synchronization) and male-specific (e.g. sperm analysis/quantitation) requirements. For most nonrodent species, procurement of embryos and the trauma associated with most manipulations are significant factors. The refinement of nonsurgical techniques for embryo collection and transfer would bypass a huge financial and labor requirement for many large domestic animal species. More importantly, experimental animals would no longer be necessarily subjected to significant surgical manipulations. Perhaps recent advances, involving embryo aggregation and the derivation of transgenic mice completely from early passage ES cells, may shed new light on novel and efficient techniques.

5.2 Future Directions

For many species, there are technical drawbacks to various genetic-engineering technologies that are manifest in the lack of significant gene characterization and mapping efforts. With the complexity of the mammalian genome, this information is critical for the determination of appropriate genes to engineer and transfer, and in providing preliminary indications of the potential biological consequences of genetic-engineering experimentation. What are the appropriate genetic targets in different species and how will gene-targeting manipulations influence mammalian development? If genetic polymorphisms (indicative of related genes that code for different isoform proteins) or pseudogenes are present, targeting efforts can be long and arduous. For well-characterized genomes, the identification of specific genetic variations or deficiencies may provide only a glimpse of the effort necessary to develop in other species. Yet, it is hoped that with current efforts focused on the human and animal/species-specific genomes, accumulating data will prove quite informative and adaptable, inasmuch as conserved linkage groups are highly probable and would be exploitable.

While current experimental efficiencies for transgenic animal production are relatively poor, the mechanical aspects of producing transgenic animals should not overshadow our ability to create and evaluate biologically important animal models. Much has been learned about early development and effects of altering the growth cascade as well as other physiological processes in initial transgenic animal models. These models have had far-reaching effects and have redefined what is possible in the biological, biomedical, and agricultural sectors. While studies of transgene expression and overexpression in mammalian models may not always correlate exactly, nor reveal all the possible effects and consequences when extended to human biology, the utility of transgenic animal models to scientific discovery cannot be overestimated. The use of ES cell or related methodologies to provide efficient and targeted *in vivo* genetic manipulations offers the prospects of creating profoundly useful animal models for biomedical, biological, and agricultural applications. However, the road to such success will continue to be most challenging.

See also Gene Targeting.

Bibliography

Books and Reviews

Brinster, R.L. (1993) Stem cells and transgenic mice in the study of development, *Int. J. Dev. Biol.* **3**, 89–99.

Houdebine, L.M. (1997) *Transgenic Animals: Generation and Use*, Harwood Academic Publishers, Amsterdam.

Joyner, A.L. (2000) *Gene Targeting: A Practical Approach*, IRL Press, New York.

Murray, J.D., Anderson, G.B., Oberbauer, A.M., McGloughlin, M.M. (1999) *Transgenic Animals in Agriculture*, CABI Publishing, Cambridge.

Nagy, A., Gertsenstein, M., Vintersten, K., Behringer, R. (2003) *Manipulating the Mouse Embryo: A Laboratory Manual*, 3rd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Pinkert, C.A. (2002) *Transgenic Animal Technology: A Laboratory Handbook*, 2nd edition, Academic Press, San Diego.

Pinkert, C.A. (2003) Transgenic animal technology: alternatives in genotyping and phenotyping, *Comp. Med.* **53**, 116–129.

Pinkert, C.A., Trounce, I.A. (2002) Production of transmitochondrial mice, *Methods* **26**, 348–357.

Primary Literature

Baguisi, A., Behboodi, E., Melican, D.T., Pollock, J.S., Destrempes, M.M., Cammuso, C., Williams, J.L., Nims, S.D., Porter, C.A., Midura, P., Palacios, M.J., Ayres, S.L., Denniston, R.S., Hayes, M.L., Ziomek, C.A., Meade, H.M., Godke, R.A., Gavin, W.G., Overstrom, E.W., Echelard, Y. (1999) Production of goats by somatic cell nuclear transfer, *Nat. Biotechnol.* **17**, 456–461.

Behringer, R.R., Finegold, M.J., Cate, R.L. (1994) Mullerian-inhibiting substance function during mammalian sexual development, *Cell* **79**, 415–425.

Bondioli, K., Ramsoondar, J., Williams, B., Costa, C., Fodor, W. (2001) Cloned pigs generated from cultured skin fibroblasts derived from an H-transferase transgenic boar, *Mol. Reprod. Dev.* **60**, 189–195.

Bradley, A., Evans, M., Kaufman, M.H., Robertson, E. (1984) Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines, *Nature* **309**, 255–256.

Brem, G., Brenig, B., Goodman, H.M., Seldan, R.C., Graf, F., Kruff, B., Springman, K., Hondele, J., Meyer, J., Winnacker, E.L., Kraußlich, H. (1985) Production of transgenic mice, rabbits, and pigs by microinjection into pronuclei, *Zuchthygiene* **20**, 215–252.

Brinster, R.L., Allen, J.M., Behringer, R.R., Gelinas, R.E., Palmiter, R.D. (1988) Introns increase transcriptional efficiency in transgenic mice, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 836–840.

Brinster, R.L., Avarbock, M.R. (1994) Germline transmission of donor haplotype following spermatogonial transplantation, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11303–11307.

Brinster, R.L., Braun, R.E., Lo, D., Avarbock, M.R., Oram, F., Palmiter, R.D. (1989) Targeted correction of a major histocompatibility class II Eα gene by DNA microinjected into mouse eggs, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7087–7091.

Brinster, R.L., Chen, H.Y., Trumbauer, M.E., Senear, A.W., Warren, R., Palmiter, R.D. (1981) Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs, *Cell* **27**, 223–231.

Brinster, R.L., Ritchie, K.A., Hammer, R.E., O'Brien, R.L., Arp, B., Storb, U. (1983) Expression of a microinjected immunoglobulin gene in the spleen of transgenic mice, *Nature* **306**, 332–336.

Brinster, R.L., Sandgren, E.P., Behringer, R.R., Palmiter, R.D. (1989) No simple solution for making transgenic mice, *Cell* **59**, 239–241.

Campbell, K.H.S., Ritchie, W.A., McWhir, J., Wilmut, I. (1996) Sheep cloned by nuclear transfer from a cultured cell line, *Nature* **380**, 64–67.

Cappechi, M.R. (1989) Altering the genome by homologous recombination, *Science* **244**, 1288–1292.

Chan, A.W.S., Chong, K.Y., Martinovich, C., Simerly, C., Schatten, G. (2001) Transgenic monkeys produced by retroviral gene transfer into mature oocytes, *Science* **291**, 309–312.

Chan, A.W.S., Homan, E.J., Ballou, L.B., Burns, J.C., Bremel, R.D. (1998) Transgenic cattle produced by reverse-transcribed gene transfer in oocytes, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14028–14033.

Chen, H.Y., Trumbauer, M.E., Ebert, K.M., Palmiter, R.D., Brinster, R.L. (1986) Developmental changes in the response of mouse eggs to injected genes, in: Bogorad, L. (Ed.) *Molecular Developmental Biology*, Alan R. Liss, New York.

Chisari, F.V., Pinkert, C.A., Milich, D.R., Filippi, P., McLachlan, A., Palmiter, R.D., Brinster, R.L. (1985) A transgenic mouse model of the chronic hepatitis B surface antigen carrier state, *Science* **230**, 1157–1160.

Cibelli, J.B., Stice, S.L., Golueke, P.J., Kane, J.J., Jerry, J., Blackwell, C., de Leon, F.A.P., Robl, J.M. (1998) Cloned transgenic calves produced from nonquiescent fetal fibroblasts, *Science* **280**, 1256–1258.

Costantini, F., Lacy, E. (1981) Introduction of a rabbit β-globin gene into the mouse germ line, *Nature* **294**, 92–94.

Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S., Smithies, O. (1987) Targeted correction of a mutant *HPRT* gene in mouse embryonic stem cells, *Nature* **330**, 576–578.

Doetschman, T., Maeda, N., Smithies, O. (1988) Targeted mutation of the *Hprt* gene in mouse embryonic stem cells, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8583–8587.

Ebert, K.M., Selgrath, J.P., DiTullio, P., Denman, J., Smith, T.E., Memon, M.A., Schindler, J.E., Monastersky, G.M., Vitale, J.A., Gordon, K. (1991) Transgenic production of a variant of human tissue-type plasminogen activator in goat milk: generation of transgenic goats and analysis of expression, *Biotechnology* **9**, 835–838.

Engler, P., Haasch, D., Pinkert, C.A., Doglio, L., Glymour, M., Brinster, R., Storb, U. (1991) A strain-specific modifier on mouse chromosome 4 controls the methylation of independent transgene loci, *Cell* **65**, 939–947.

Fodor, W.L., Williams, B.L., Matis, L.A., Madri, J.A., Rollins, S.A., Knight, J.W., Velander, W., Squinto, S.P. (1994) Expression of a functional human complement inhibitor in a transgenic pig as a model for the prevention of xenogeneic hyperacute organ rejection, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11153–11157.

Furth, P.A., St. Onge, L., Boger, H., Gruss, P., Gossen, M., Kistner, A., Bujard, H., Hennighausen, L. (1994) Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9302–9306.

Golovan, S.P., Hayes, M.A., Phillips, J.P., Forsberg, C.W. (2001) Transgenic mouse models demonstrating reduced pollution potential in monogastric animals, *Nat. Biotechnol.* **19**, 429–433.

Gordon, J.W., Ruddle, F.H. (1981) Integration and stable germ line transmission of genes injected into mouse pronuclei, *Science* **214**, 1244–1246.

Gordon, J.W., Scangos, G.A., Plotkin, D.J., Barbosa, J.A., Ruddle, F.H. (1980) Genetic transformation of mouse embryos by microinjection of purified DNA, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7380–7384.

Gossler, A., Doetschman, T., Korn, R., Serfling, E., Kemler, R. (1986) Transgenesis by means of blastocyst-derived embryonic stem cell lines, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9065–9069.

Grosveld, F., van Assendelft, G.B., Greaves, D.R., Kollias, G. (1987) Position-independent, high level expression of the human β-globin gene in transgenic mice, *Cell* **51**, 975–985.

Hammer, R.E., Maika, S.D., Richardson, J.A., Tang, J.P., Taurog, J.D. (1990) Spontaneous inflammatory disease in transgenic rats expressing HLA-B27 and human β2-m: an animal model of HLA-B27-associated human disorders, *Cell* **63**, 1099–1112.

Hammer, R.E., Pursel, V.G., Rexroad, C.E. Jr., Wall, R.J., Bolt, D.J., Ebert, K.M., Palmiter, R.D., Brinster, R.L. (1985) Production of transgenic rabbits, sheep, and pigs by microinjection, *Nature* **315**, 680–683.

Harbers, K., Jahner, D., Jaenisch, R. (1981) Microinjection of cloned retroviral genomes into mouse zygotes; integration and expression in the animal, *Nature* **293**, 540–542.

Harris, A.W., Pinkert, C.A., Crawford, M., Langdon, W.Y., Brinster, R.L., Adams, J.M. (1988) The Eμ-myc transgenic mouse: a model for high incidence spontaneous lymphoma and leukemia of early B cells, *J. Exp. Med.* **167**, 353–371.

Hasty, P., Ramirez-Solis, R., Krumlauf, R., Bradley, A. (1991) Introduction of a subtle mutation into the Hox-2.6 locus in embryonic stem cells, *Nature* **350**, 243–246.

Hasuwa, H., Kaseda, K., Einarsdottir, T., Okabe, M. (2002) Small interfering RNA and gene silencing in transgenic mice and rats, *FEBS Lett.* **532**, 227–230.

Irwin, M.H., Johnson, L.W., Pinkert, C.A. (1999) Isolation and microinjection of somatic cell-derived mitochondria and germline heteroplasmy in transmitochondrial mice, *Transgenic Res.* **8**, 119–123.

Irwin, M.H., Moffatt, R.J., Pinkert, C.A. (1996) Identification of transgenic mice by PCR analysis of saliva, *Nat. Biotechnol.* **14**, 1146–1149.

Jakobovits, A., Moore, A.L., Green, L.L., Vergara, G.J., Maynard-Currie, C.E., Austin, H.A., Klapholz, S. (1993) Germ-line transmission and expression of a human-derived yeast artificial chromosome, *Nature* **362**, 255–258.

Kato, Y., Tani, T., Sotomaru, Y., Kurokawa, K., Kato, J.Y., Doguchi, H., Yasue, H., Tsunoda, Y. (1998) Eight calves cloned from somatic cells of a single adult, *Science* **282**, 2095–2098.

Kooyman, D.L., Pinkert, C.A. (1994) Transgenic mice expressing a chimaeric anti-*E. coli* immunoglobulin α heavy chain gene, *Transgenic Res.* **3**, 167–175.

Kubota, C., Yamakuchi, H., Todoroki, J., Mizoshita, K., Tabara, N., Barber, M., Yang, X.Z. (2000) Six cloned calves produced from adult fibroblast cells after long-term culture, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 990–995.

Lakso, M., Pichel, J.G., Gorman, J.R., Sauer, B., Okamoto, Y., Lee, E., Alt, F.W., Westphal, H. (1996) Efficient *in vivo* manipulation of mouse genomic sequences at the zygote stage, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5860–5865.

Lavitrano, M., Bacci, M.L., Forni, M., Lazzereschi, D., Di Stefano, C., Fioretti, D., Giancotti, P., Marfe, G., Pucci, L., Renzi, L., Wang, H., Stoppacciaro, A., Stassi, G., Sargiacomo, M., Sinibaldi, P., Turchi, V., Giovannoni, R., Della Casa, G., Seren, E., Rossi, G. (2002) Efficient production by sperm-mediated gene transfer of human decay accelerating factor (hDAF) transgenic pigs for xenotransplantation, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13090–13095.

Lavitrano, M., Camaioni, A., Fazio, V.M., Dolci, A., Farace, M.G., Spadafora, C. (1989) Sperm cells as vectors for introducing foreign DNA into eggs: genetic transformation of mice, *Cell* **57**, 717–723.

Li, Z.W., Stark, G., Gotz, J., Rulicke, T., Gschwind, M., Huber, G., Muller, U., Weissmann, C. (1996) Generation of mice with a 200-kb amyloid precursor protein gene deletion by Cre recombinase mediated site-specific recombination in embryonic stem cells, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6158–6162.

Martin, G.R., Silver, L.M., Fox, H.S., Joyner, A.L. (1987) Establishment of embryonic stem cell lines from preimplantation mouse embryos homozygous for lethal mutations in the t-complex, *Dev. Biol.* **121**, 20–28.

Matsui, Y., Zsebo, K., Hogan, B.L. (1992) Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture, *Cell* **70**, 841–847.

McGrath, J., Solter, D. (1983) Nuclear transplantation in the mouse embryo by microsurgery and cell fusion, *Science* **220**, 1300–1302.

McKnight, R.A., Shamay, A., Sankaran, L., Wall, R.J., Henninghausen, L. (1992) Matrix attachment regions can impart position-independent regulation of a tissue-specific gene in transgenic mice, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6943–6947.

Metzger, D., Clifford, J., Chiba, H., Chambon, P. (1995) Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6991–6995.

Meyers, E.N., Lewandoski, M., Martin, G.R. (1998) An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination, *Nat. Genet.* **18**, 136–141.

Nagano, M., Brinster, C.J., Orwig, K.E., Rye, B.-Y., Avarbock, M.R., Brinster, R.L. (2001) Transgenic mice produced by retroviral transduction of male germ-line stem cells, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13090–13095.

Nagy, A., Gocza, E., Diaz, E.M., Prideaux, V.R., Ivanyi, E., Markkula, M., Rossant, J. (1990) Embryonic stem cells alone are able to support fetal development in the mouse, *Development* **110**, 815–821.

Novak, A., Guo, C., Yang, W., Nagy, A., Lobe, C.G. (2000) Z/EG, a double reporter mouse line that expresses enhanced green fluorescent protein upon Cre-mediated excision, *Genesis* **28**, 147–155.

Onishi, A., Iwamoto, M., Akita, T., Mikawa, S., Takeda, K., Awata, T., Hanada, H., Perry, A.C.F. (2000) Pig cloning by microinjection of fetal fibroblast nuclei, *Science* **289**, 1188–1190.

Ono, Y., Shimozawa, N., Ito, M., Kono, T. (2001) Cloned mice from fetal fibroblast cells arrested at metaphase by a serial nuclear transfer, *Biol. Reprod.* **64**, 44–50.

Overbeek, P.A., Aguilar-Cordova, E., Hanten, G., Schaffner, D.L., Patel, P., Lebovitz, R.M., Lieberman, M.W. (1991) Coinjection strategy

for visual identification of transgenic mice, *Transgenic Res.* **1**, 31–37.

Overbeek, P.A., Lai, S.-P., Van Quill, K.R., Westphal, H. (1986) Tissue-specific expression in transgenic mice of a fused gene containing RSV terminal sequences, *Science* **231**, 1574–1577.

Palmiter, R.D., Brinster, R.L., Hammer, R.E., Trumbauer, M.E., Rosenfeld, M.G., Birnberg, N.C., Evans, R.M. (1982) Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes, *Nature* **300**, 611–615.

Palmiter, R.D., Norstedt, G., Gelinas, R.E., Hammer, R.E., Brinster, R.L. (1983) Metallothionein–human GH fusion genes stimulate growth of mice, *Science* **222**, 809–814.

Petitte, J.N., Clark, M.E., Liu, G., Verrinder Gibbins, A.M., Etches, R.J. (1990) The production of somatic and germline chimeras in the chick by transfer of early blastodermal cell, *Development* **108**, 185–190.

Piedrahita, J.A., Moore, K., Oetama, B., Lee, C.K., Scales, N., Ramsoondar, J., Bazer, F.W., Ott, T. (1998) Generation of transgenic porcine chimeras using primordial germ cell-derived colonies, *Biol. Reprod.* **58**, 1321–1329.

Pieper, F.R., deWit, I.C.M., Pronk, A.C.J., Kooiman, P.M., Strijker, R., Krimpenfort, P.K.A., Nuyens, J.H., deBoer, H.A. (1992) Efficient generation of functional transgenes by homologous recombination in murine zygotes, *Nucl. Acids Res.* **20**, 1259–1264.

Pinkert, C.A. (1990) A rapid procedure to evaluate foreign DNA transfer into mammals, *Biotechniques* **9**, 38–39.

Pinkert, C.A., Johnson, L.W., Irwin, M.H., Wong, S., Baetge, E.E., Wolfe, D.F., Simpkins, S.A., Owsley, W.F., Bartol, F.F. (2001) Optimization of superovulation and fertilization protocols in the production of transgenic swine, *Adv. Reprod.* **5**, 45–53.

Pinkert, C.A., Ornitz, D.M., Brinster, R.L., Palmiter, R.D. (1987) An albumin enhancer located 10 kb upstream functions along with its promoter to direct efficient, liver-specific expression in transgenic mice, *Genes Dev.* **1**, 268–276.

Polejaeva, I.A., Chen, S.H., Vaught, T.D., Page, R.L., Mullins, J., Ball, S., Dai, Y.F., Boone, J., Walker, S., Ayares, D.L., Colman, A., Campbell, K.H.S. (2000) Cloned pigs produced by nuclear transfer from adult somatic cells, *Nature* **407**, 86–90.

Prather, R.S., Barnes, F.L., Sims, M.M., Robl, J.M., Eyestone, W.H., First, N.L. (1990) Nuclear transplantation in the bovine embryo assessment of donor nuclei and recipient oocyte, *Biol. Reprod.* **37**, 859–866.

Pursel, V.G., Pinkert, C.A., Miller, K.F., Bolt, D.J., Campbell, R.G., Palmiter, R.D., Brinster, R.L., Hammer, R.E. (1989) Genetic engineering of livestock, *Science* **244**, 1281–1288.

Quaife, C.J., Pinkert, C.A., Ornitz, D.M., Palmiter, R.D., Brinster, R.L. (1987) Pancreatic neoplasia induced by ras expression in acinar cells of transgenic mice, *Cell* **48**, 1023–1034.

Richa, J., Lo, C.W. (1989) Introduction of human DNA into mouse eggs by injection of dissected chromosome fragments, *Nature* **245**, 175–177.

Rucker, E.B., Piedrahita, J.A. (1997) Cre-mediated recombination at the murine whey acidic protein (mWAP) locus, *Mol. Reprod. Dev.* **48**, 324–331.

Salter, D.W., Smith, E.J., Hughes, S.H., Wright, S.E., Fadly, A.M., Witter, R.L., Crittenden, L.B. (1986) Gene insertion into the chicken germ line by retroviruses, *Poultry Sci.* **65**, 1445–1458.

Sandgren, E.P., Quaife, C.J., Pinkert, C.A., Palmiter, R.D., Brinster, R.L. (1989) Oncogene-induced liver neoplasia in transgenic mice, *Oncogene* **4**, 715–724.

Schedl, A., Montoliu, L., Kelsey, G., Schütz, G. (1993) A yeast artificial chromosome covering the tyrosinase gene confers copy number-dependent expression in transgenic mice, *Nature* **362**, 258–261.

Schnieke, A.E., Kind, A.J., Ritchie, W.A., Mycock, K., Scott, A.R., Ritchie, M., Wilmut, I., Colman, A., Campbell, K.H.S. (1997) Human factor IX transgenic sheep produced by transfer of nuclei from transfected fetal fibroblasts, *Science* **278**, 2130–2133.

Simons, J.P., Wilmut, I., Clark, A.J., Archibald, A.L., Bishop, J.O., Lathe, R. (1988) Gene transfer into sheep, *Biotechnology* **6**, 179–183.

Storb, U., Pinkert, C., Arp, B., Engler, P., Gollahon, K., Manz, J., Brady, W., Brinster, R. (1986) Transgenic mice with μ and κ genes encoding antiphosphorylcholine antibodies, *J. Exp. Med.* **164**, 627–641.

Swanson, M.E., Martin, M.J., O'Donnell, J.K., Hoover, K., Lago, W., Huntress, V., Parsons, C.T., Pinkert, C.A., Pilder, S., Logan, J.S. (1992) Production of functional human hemoglobin in transgenic swine, *Biotechnology* **10**, 557–559.

Swift, G.H., Hammer, R.E., MacDonald, R.J., Brinster, R.L. (1984) Tissue-specific expression of the rat pancreatic elastase I gene in transgenic mice, *Cell* **38**, 639–646.

Takeda, K., Akagi, S., Takahashi, S., Onishi, A., Hanada, H., Pinkert, C.A. (2002) Mitochondrial activity in response to serum starvation in bovine (*Bos taurus*) cell culture, *Cloning Stem Cells* **4**, 223–230.

te Riele, H., Maandag, E.R., Berns, A. (1992) Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5128–5132.

Thomas, K.R., Folger, K.R., Capecchi, M.R. (1986) High frequency targeting of genes to specific sites in the mammalian genome, *Cell* **44**, 419–428.

Thompson, S., Clarke, A.R., Pow, A.M., Hooper, M.L., Melton, D.W. (1989) Germ line transmission and expression of a corrected *HPRT* gene produced by gene targeting in embryonic stem cells, *Cell* **56**, 313–321.

Townes, T.M., Lingrel, J.B., Chen, H.Y., Brinster, R.L., Palmiter, R.D. (1985) Erythroid-specific expression of human β-globin genes in transgenic *mice*, *EMBO J.* **4**, 1715–1723.

Vazquez, J.C., Nogues, C., Rucker, E.B., Piedrahita, J.A. (1998) Factors affecting the efficiency of introducing precise genetic changes in ES cells by homologous recombination: tag–and exchange versus the Cre–loxP system, *Transgenic Res.* **7**, 181–193.

Wagner, T.E., Hoppe, P.C., Jollick, J.D., Scholl, D.R., Hodinka, R.L., Gault, J.B. (1981) Microinjection of a rabbit β-globin gene into zygotes and its subsequent expression in adult mice and their offspring, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6376–6380.

Wagner, E.F., Stewart, T.A., Mintz, B. (1981) The human β-globin gene and a functional viral thymidine kinase gene in developing mice, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5016–5020.

Wakayama, T., Perry, A.C.F., Zuccotti, M., Johnson, K., Yanagimachi, R. (1998) Full–term development of mice from enucleated oocytes injected with cumulus cell nuclei, *Nature* **394**, 369–374.

Wall, R.J., Pursel, V.G., Hammer, R.E., Brinster, R.L. (1985) Development of porcine ova that were centrifuged to permit visualization of pronuclei and nuclei, *Biol. Reprod.* **32**, 645–651.

Wiles, M.V., Vauti, F., Otte, J., Füchtbauer, E.-M., Ruiz, P., Füchtbauer, A., Arnold, H.-H., Lehrach, H., Metz, T., von Melchner, H., Wurst, W. (2000) Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells, *Nat. Genet.* **24**, 13–14.

Wilkie, T.M., Brinster, R.L., Palmiter, R.D. (1986) Germline and somatic mosaicism in transgenic mice, *Dev. Biol.* **118**, 9–18.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., Campbell, K.H. (1997) Viable offspring derived from fetal and adult mammalian cells, *Nature* **385**, 810–813.

Winn, R.N., Norris, M.B., Brayer, K.J., Torres, C., Muller, S.L. (2000) Detection of mutations in transgenic fish carrying a bacteriophage lambda *cII* transgene target, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12655–12660.

Wood, S.A, Allen, N.D., Rossant, J., Auerbach, A., Nagy, A. (1993) Non-injection methods for the production of embryonic stem cell-embryo chimaeras, *Nature* **365**, 87–89.

Xu, X., Li, C., Garrett–Beal, L., Larson, D., Wynshaw-Boris, A., Deng, C.X. (2001) Direct removal in the mouse of a floxed neo gene from a three-loxP conditional knockout allele by two novel approaches, *Genesis* **30**, 1–6.

Yang, X.W., Model, P., Heintz, N. (1997) Homologous recombination-based modification in E. coli and germline transmission in transgenic mice of a bacterial artificial chromosome, *Nat. Biotechnol.* **15**, 859–865.

Yokoyama, T., Copeland, N.G., Jenkins, N.A., Montgomery, C.A., Elder, F.F.B., Overbeek, P.A. (1993) Reversal of left-right asymmetry: a situs inversus mutation, *Science* **260**, 679–682.

Annexins

Carl E. Creutz
University of Virginia, Charlottesville, VA, USA

1 **Diversity and Functions 242**

2 **Structure and Mechanism of Action 244**

3 **Regulation 245**

4 **Applications 246**

Bibliography 247
Books and Reviews 247
Primary Literature 247

Keywords

Exocytosis
Release of secretory product from a cell by fusion of an intracellular vesicle containing the product with the cell-surface membrane.

Membrane Fusion
Merger of two biological membranes to form a single membrane, as when the membrane of a secretory vesicle fuses with the surface membrane of the cell.

Phospholipid
Common constituent of biological membranes composed of two fatty acids esterified to a glycerol backbone with a head group containing a phosphate ester.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Phosphorylation
Covalent attachment of phosphate, as in the enzyme-catalyzed phosphorylation of serine or tyrosine amino acid side chains in a protein.

The annexins are a family of structurally related, calcium-dependent, phospholipid-binding proteins that have been postulated to mediate calcium-dependent activities at membrane surfaces such as membrane fusion, lipid metabolism and reorganization, and ion permeation. The basic annexin structure consists of 4 homologous 70–amino acid repeats and a unique N-terminal domain. These repeats do not contain sequences found in other intracellular calcium-binding proteins; therefore, the annexins represent a novel class of calcium-binding proteins. The "core" domains of all the annexins, composed of the four 70–amino acid repeats, are 40 to 60% identical in sequence. One annexin family member (annexin VI) has been formed as a result of gene duplication and consists of 8 of the 70–amino acid repeats. Another type of duplication has occurred with annexin II (calpactin), in which two 36-kDa molecules, the "heavy chains," each containing four repeats, bind to a dimer of a 10-kDa protein, the "light chains," to form a tetramer. In contrast to other lipid-binding proteins such as protein kinase C, or phospholipase A2, the annexins are unique in that most are bivalent. That is, they can attach to two membranes, rather than just one, and as a consequence, draw them together.

1 Diversity and Functions

Members of the annexin family of proteins, of which there are now 11 recognized mammalian variants, have been independently discovered in a number of contexts. As a consequence, a variety of names have been used for the different family members, such as *synexins, chromobindins, lipocortins, calcimedins, calphobindins,* and *calpactins*. However, a standard reference nomenclature has been adopted involving the term *annexin* followed by a Roman numeral (Fig. 1). Representative annexins have been isolated from animals, plants, and the slime mold *Dictyostelium,* attesting to the universality of the functions performed by annexins. Owing to the bivalent activity of the annexins, one of their postulated roles is the promotion of membrane fusion in exocytosis. After secretory vesicle membranes have been aggregated by an annexin, they undergo fusion if exposed to cis-unsaturated fatty acids. In a cell, such fusogenic fatty acids might be made available through activation of lipases. Alternatively, the annexins might function in concert with other proteins that promote membrane fusion and remodeling after the membranes have been brought into close apposition by the annexin. In addition to a possible role in exocytosis and membrane trafficking, some of the annexins have been suggested to be mediators of the anti-inflammatory effects of steroids, components of the submembranous cytoskeleton, inhibitors

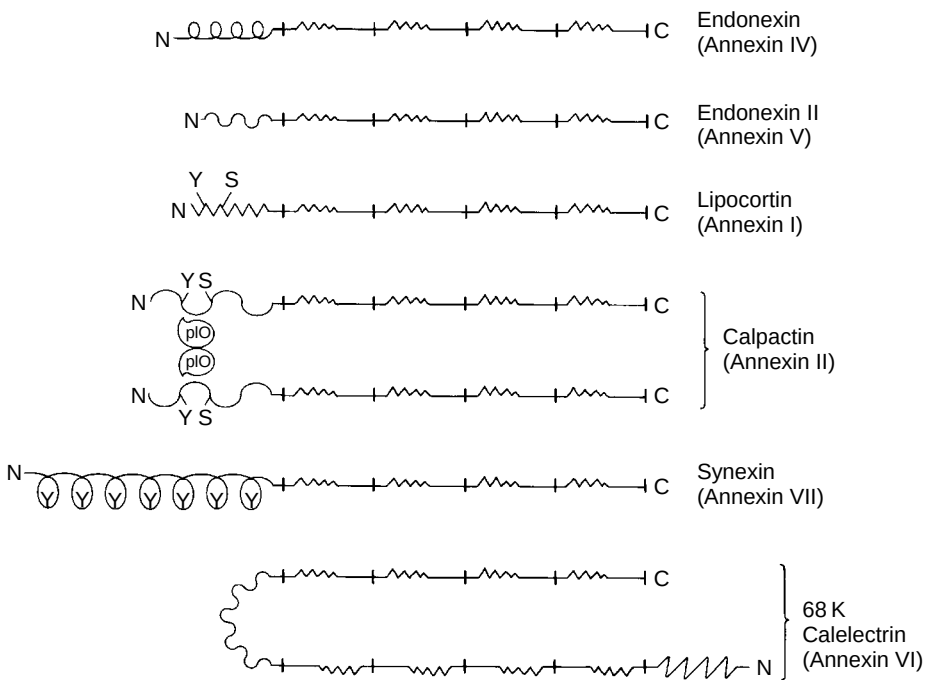

Fig. 1 Schematic illustration of the primary structures of six annexins. Each of the 4 (or 8) homologous domains contains the 17–amino acid "endonexin fold" sequence represented by a sawtooth line (sequence: KGhGTDExxLIpILApR: h, hydrophobic residue; p, polar residue; x, variable residue). The unique N-terminal structures are on the left (or right in the case of annexin VI); Y and S represent phosphorylation sites in the tails of annexins I and II. The annexin II tetramer is drawn showing the association of the N-termini of the heavy chains with the light-chain (p10) dimer. The *Y*s inside the loops in the tail of synexin represent a pro-β-helix. [Reproduced from Dedman and Smith (1990), with permission.]

of blood coagulation, transducers of signals generated by tyrosine kinases at the cell membrane, mediators of bone formation and remodeling, mediators of cell-matrix interactions, voltage-dependent ion channels, actin bundling proteins, regulators of calcium release channels in muscle, regulators of DNA polymerase activity, and extracellular binding sites for plasminogen. This diversity of hypotheses attests to the ubiquitous and abundant nature of these proteins. In most cases, the data supporting a role for the annexins in these various processes are derived from *in vitro* systems involving calcium and lipid membranes, and it has been difficult to relate these activities to true cellular functions. However, it seems likely that this family of proteins has radiated to perform a variety of cellular functions, since so many are coexpressed in single cells and their individual differences have been highly conserved during evolution. In that sense, the annexins may be similar to the other major class of calcium-binding proteins, the "EF-hand" family, which includes calmodulin and troponin C, and are responsible for a variety of nonoverlapping cellular phenomena.

2
Structure and Mechanism of Action

The annexins, in general, have proven to be relatively easy to isolate in high purity, and this advantage has led to several successful attempts to crystallize members of the family. The structure of annexin V reveals that each of the four 70–amino acid repeats forms a compact bundle of α-helices and these four bundles are arranged in a plane (Fig. 2). Viewed from the side, the molecule displays a convex face, thought to be the membrane-binding face, and a concave face, thought to face the cytoplasm. Each homologous domain forms one or two potential calcium binding sites, which are complex in that they involve convergent loops from different parts of the domain. In addition, the positions of sulfate ions in the loops (derived from the crystallization medium) suggest that the phosphate of a lipid head group may also sit in the binding pocket and coordinate the calcium ion. This geometry would explain the interdependence of calcium and lipid binding by the annexins. Since all the lipid binding sites of the annexin core domains face a single membrane, they may be able to integrate information about membrane-lipid composition. In general, all the annexins bind with higher affinity (i.e. at lower levels of calcium) to acidic phospholipids. However, there is some variation reported in the literature for different annexins with regard to order of preference for

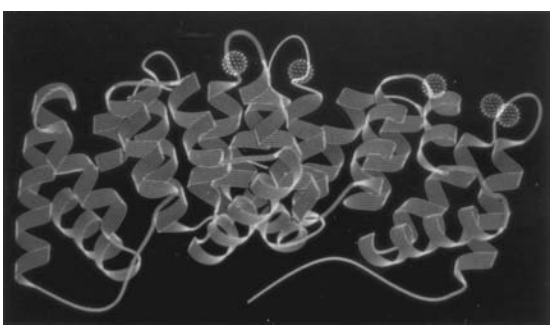

(a)

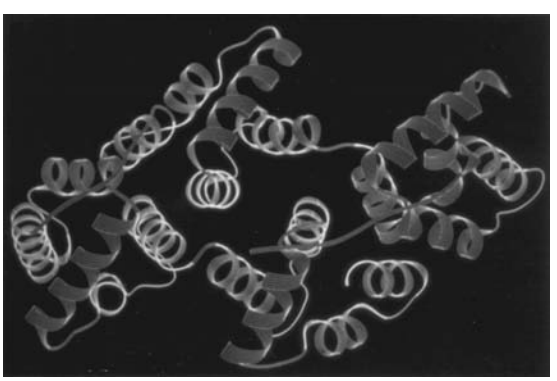

(b)

Fig. 2 Ribbon plots of the structure of annexin V as determined by X-ray diffraction. (a) "Side" view of the molecule, with the calcium binding sites and the membrane-binding face at the top. Calcium ions are represented by the spheres. The high-affinity sites are represented by the first, second, and fifth spheres from left to right; the third and fourth spheres indicate low-affinity ion binding sites that were identified by lanthanide binding. The extended N-terminus is seen at the bottom. (b) View of the "cytoplasmic" side of the molecule and the N-terminus. The calcium binding sites are on the opposite face. Note the potential channel structure in the center of the molecule. [From Creutz (1992), with permission. Original figure kindly provided by A. Burger and R. Huber.]

different acidic lipids. In addition, the four homologous repeats may have different specificities, since the repeats are only 40 to 50% identical in sequence.

The crystal structure of annexin V has also been interpreted in terms of the ion channel–forming properties of the protein. The molecule has a hydrophilic pore perpendicular to the face of the membrane (Fig. 2). Because the molecule has no hydrophobic external surfaces, it is assumed that it cannot actually enter the bilayer in this conformation. However, it has a calculated external electrical field similar to the field strength necessary to punch a hole through a membrane by electroporation. Therefore, it has been suggested that the molecule sits on the surface of the membrane and leads to a disordered state in the lipids immediately below it. In this disordered state, the membrane conducts ions that must also pass through the hydrophilic pore in the molecule. The molecule thus provides the ion selectivity of the overall transmembrane channel. A similar mechanism of membrane permeabilization has been suggested for the action of cholera toxin.

An alternative mechanism for the ability of annexins to promote ion permeation has been proposed that involves a significant reorganization of the molecule at low pH to form a transmembrane channel. However, a crystallographic model of this proposed structure has not yet been determined. It is also not known if the channel activity that annexins exhibit *in vitro* corresponds to a specific physiological channel in cells.

Since the calcium/lipid binding sites all appear to be on one side of the annexin molecule, it is not obvious how the annexins aggregate membranes. It seems likely that a self-association of annexin molecules attached to different membranes is required for membrane aggregation (Fig. 3). The concave faces of the molecules that face the cytoplasm might interlock during this self-association event. In the case of synexin acting *in vitro* on chromaffin granules (secretory vesicles of the adrenal medulla), binding to membranes occurs at low levels of calcium (less than 10 μmol), but membrane aggregation and fusion depend on higher levels of calcium (greater than 100 μmol). The calcium dependence of granule aggregation correlates exactly with the calcium dependence of synexin self-association in the absence of membranes, suggesting a mechanism for membrane aggregation whereby membrane-bound synexin molecules undergo self-association to bring the two membranes together. If this mechanism is correct, then the ''bottleneck'' in the system is the high level of calcium needed to promote annexin self-association. The annexin II (calpactin) tetramer promotes the aggregation and fatty acid–dependent fusion of chromaffin granules at the lowest level of calcium of any annexin (approximately 1 μmol). This may occur because the tetramer represents a permanently self-associated annexin (Fig. 1). The two heavy chains might be associated through the light-chain dimer in such a way that each heavy chain can bind a different membrane and thus pull them together. Therefore, the processes of overall membrane aggregation and fusion are catalyzed by the low levels of calcium needed simply to promote membrane binding.

3 Regulation

The N-terminal domains of the annexins are the major sites of sequence divergence in the family (Fig. 1). They may be the site of interaction with other proteins, as

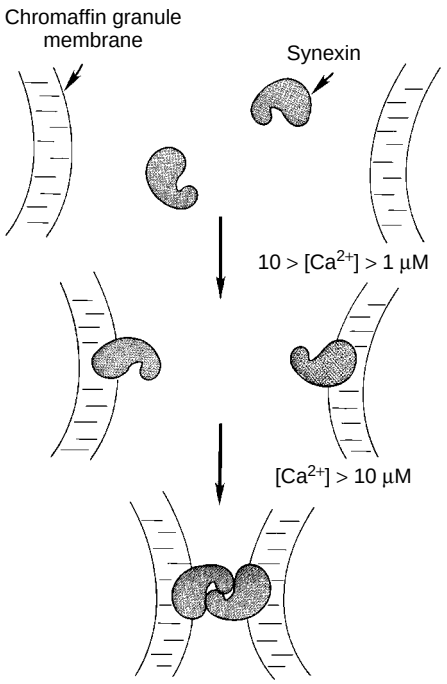

Fig. 3 Hypothesis for the mechanism of action of synexin (annexin VII) in promoting intermembrane contacts in a cell. In the resting cell (top), the Ca^{2+} concentration is less than 1 μM and synexin is freely soluble. In the stimulated cell (middle), as the calcium concentration rises to between 1 and 10 μM, synexin attaches to membranes. When the calcium concentration further increases above 10 μM (perhaps near the plasma membrane), synexin molecules on different membranes self-associate to bring the membranes together. [Reproduced from Creutz and Sterner (1983), with permission.]

in the case of the annexin II heavy chain, which associates with p10, and annexin XI, which associates with the small calcium-binding protein, calcyclin. The N-terminal domains are also potential regulatory sites for the rest of the molecule. For example, cleavage of the N-terminus of the isolated annexin II heavy chain is necessary for this protein to aggregate chromaffin granules at low levels of calcium. Furthermore, the calcium sensitivity of membrane aggregation by such cleaved molecules is strongly affected by the exact site of cleavage. In addition, annexins I and II are phosphorylated in the N-terminal domain by tyrosine- or serine-/threonine-specific kinases. Phosphorylation of the N-terminus of annexin I by protein kinase C *blocks* the ability of this protein to aggregate chromaffin granules, while slightly *enhancing* its ability to bind the granule membrane. These biochemical observations on the importance of the annexin N-terminal structure can be interpreted in terms of the crystal structure of annexin V. Since the N-terminus is located on the "cytoplasmic" side of the membrane-bound annexin, it may be in a position to participate in annexin–annexin self-association essential to membrane aggregation. This may explain why alterations of the tails of other annexins strongly affect membrane aggregation but not membrane binding. Such alterations, resulting from either proteolysis or phosphorylation, may provide additional mechanisms for cellular control of these calcium-dependent proteins.

4 Applications

Although the complete breadth and details of the biological roles of annexins are yet

to be determined, a number of biotechnological and medical applications are under development. Most of these applications relate to the signature ability of annexins to bind to acidic phospholipids in the presence of calcium. Members of this protein family have potential utility as anticoagulants because they can prevent the association of clotting factors with acidic lipids on the surface of activated platelets and endothelial cells. Annexins may also function as immunosuppressive agents by shielding phospholipids from attack by phospholipase A_2, thus preventing the release of arachidonic acid, which is the precursor for a number of inflammatory lipid mediators. Annexins are a staple of any research laboratory studying apoptosis *in vitro* since one hallmark of the cell death pathway is the exposure of phosphatidylserine on the cell surface, which can be marked by the binding of a fluorescent annexin. Labeled annexins have also been tested in whole animals for the ability to localize occult thromboses or apoptotic cells in whole-body imaging techniques, and annexins fused to clot-dissolving enzymes have been tested for the ability to direct these enzymes to thromboses.

See also Bioorganic Chemistry; Bioinorganic Chemistry; Calcium Biochemistry; Phospholipids.

Bibliography

Books and Reviews

Creutz, C.E. (1992) The annexins and exocytosis, *Science* **258**, 924–931.

Creutz, C.E. (2003) Reflections on twenty-five years of annexin research, in: Bandorowicz-Pikula, J. (Ed.) *Annexins: Biological Importance and Annexin-Related Pathologies*, Lindes Bioscience, Georgetown, Texas, pp. 1–20.

Crumpton, M.J., Dedman, J.R. (1990) Protein terminology tangle, *Nature* **345**, 212.

Dedman, J.R., Smith, V.L. (Eds.) (1990) *Stimulus–Response Coupling: The Role of Intracellular Calcium-Binding Proteins*, CRC Press, Boston.

Gerke, V., Moss, S.E. (2002) Annexins: from structure to function, *Physiol. Rev.* **82**, 331–371.

Moss, S.E. (Ed.) (1992) *The Annexins*, Portland Press, London.

Seaton, B.A. (Ed.) (1996) *Annexins: Molecular Structure to Cellular Function*, RG Landes Co., Austin.

Primary Literature

Avila-Sakar, A.J., Kretsinger, R.H., Creutz, C.E. (2000) Membrane-bound 3D structures reveal the intrinsic flexibility of annexin VI, *J. Struct. Biol.* **130**, 54–62.

Berendes, R., Voges, D., Demange, P., Huber, R., Burger, A. (1993) Structure-function analysis of the ion channel selectivity filter in human annexin V, *Science* **262**, 427–430.

Blankenberg, F.G., Katsikis, P.D., Tait, J.F., Davis, R.E., Naumovski, L., Ohtsuki, K., Kopiwada, S., Abrams, M.J., Darkes, M., Robbins, R.C., Maecker, H.T., Strauss, H.W. (1998) In vivo detection and imaging of phosphatidylserine expression during programmed cell death, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6349–6354.

Burns, A.L., Magendzo, K., Shirvan, A., Srivastava, M., Rojas, E., Alijani, M.R., Pollard, H.B. (1989) Calcium channel activity of purified human synexin and structure of the human synexin gene, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 3798–3802.

Creutz, C.E. (1981) cis-Unsaturated fatty acids induce the fusion of chromaffin granules aggregated by synexin, *J. Cell Biol.* **91**, 247–256.

Creutz, C.E., Pazoles, C.J., Pollard, H.B. (1978) Identification and purification of an adrenal medullary protein (synexin) that causes calcium-dependent aggregation of isolated chromaffin granules, *J. Biol. Chem.* **253**, 2858–2866.

Creutz, C.E., Sterner, D.C. (1983) Calcium dependence of the binding of synexin to isolated chromaffin granules, *Biochem. Biophys. Res. Commun.* **114**, 355–364.

Drust, D.S., Creutz, C.E. (1988) Aggregation of chromaffin granules by calpactin at micromolar levels of calcium, *Nature* **331**, 88–91.

Fava, R.A., Cohen, S. (1984) Isolation of a calcium-dependent 35-kilodalton substrate for the epidermal growth factor receptor/kinase from A-431 cells, *J. Biol. Chem.* **259**, 2636–2645.

Flower, R.J., Blackwell, G.J. (1979) Anti-inflammatory steroids induce biosynthesis of a phospholipase A2 inhibitor which prevents prostaglandin generation, *Nature* **278**, 456–459.

Funakoshi, T., Heimark, R.L., Hendrickson, L.E., McMullen, B.A., Fujikawa, K. (1987) Human placental anticoagulant protein: isolation and characterization, *Biochemistry* **26**, 5572–5578.

Geisow, M.J., Burgoyne, R.D. (1982) Calcium-dependent binding of cytosolic proteins by chromaffin granules from adrenal medulla, *J. Neurochem.* **38**, 1735–1741.

Geisow, M.J., Fritsche, U., Hexham, J.M., Dash, B., Johnson, T. (1986) A consensus amino-acid sequence repeat in Torpedo and mammalian Ca^{2+}-dependent membrane-binding proteins, *Nature* **320**, 636–638.

Genge, B.R., Wu, L.N., Wuthier, R.E. (1990) Differential fractionation of matrix vesicle proteins. Further characterization of the acidic phospholipid-dependent Ca2(+)-binding proteins, *J. Biol. Chem.* **265**, 4703–4710.

Gerke, V., Weber, K. (1984) Identity of p36K phosphorylated upon Rous sarcoma virus transformation with a protein purified from brush borders; calcium-dependent binding to non-erythroid spectrin and F-actin, *EMBO J.* **3**, 227–233.

Hirata, F., Schiffmann, E., Venkatasubramanian, K., Salomon, D., Axelrod, J. (1980) A phospholipase A2 inhibitory protein in rabbit neutrophils induced by glucocorticoids, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2533–2536.

Huber, R., Romisch, J., Paques, E.P. (1990) The crystal and molecular structure of human annexin V, an anticoagulant protein that binds to calcium and membranes, *EMBO J.* **9**, 3867–3874.

Huber, R., Schneider, M., Mayr, I., Romisch, J., Paques, E.P. (1990) The calcium binding sites in human annexin V by crystal structure analysis at 2.0 A resolution. Implications for membrane binding and calcium channel activity, *FEBS Lett.* **275**, 15–21.

Iwasaki, A., Suda, M., Nakao, H., Nagoya, T., Saino, Y., Arai, K., Mizoguchi, T., Sato, F., Yoshizaki, H., Hirata, M., Miyata, T., Shidara, Y., Murata, M., Maki, M. (1987) Structure and expression of cDNA for an inhibitor of blood coagulation isolated from human placenta: a new lipocortin-like protein, *J. Biochem. (Tokyo)* **102**, 1261–1273.

Kristensen, T., Saris, C.J., Hunter, T., Hicks, L.J., Noonan, D.J., Glenney, J.R., Tack, B.F. (1986) Primary structure of bovine calpactin I heavy chain (P36), a major cellular substrate for retroviral protein-tyrosine kinases: homology with the human phospholipase A2 inhibitor lipocortin, *Biochemistry* **25**, 4497–4503.

Langen, R., Isas, J.M., Hubbell, W.L., Haigler, H.T. (1998) A transmembrane form of annexin XII detected by site-directed spin labeling, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14 060–14 065.

Moore, P.B., Dedman, J.R. (1982) Calcium-dependent protein binding to phenothiazine columns, *J. Biol. Chem.* **257**, 9663–9667.

Pollard, H.B., Rojas, E. (1988) Ca^{2+}-activated synexin forms highly selective, voltage-gated Ca^{2+} channels in phosphatidylserine bilayer membranes, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2974–2978.

Radke, K., Gilmore, T., Martin, G.S. (1980) Transformation by Rous sarcoma virus: a cellular substrate for transformation-specific protein phosphorylation contains phosphotyrosine, *Cell* **21**, 821–828.

Reutelingsperger, C.P., Hornstra, G., Hemker, H.C. (1985) Isolation and partial purification of a novel anticoagulant from arteries of human umbilical cord, *Eur. J. Biochem.* **151**, 625–629.

Russo-Marie, F., Duval, D. (1982) Dexamethasone-induced inhibition of prostaglandin production does not result from a direct action on phospholipase activities but is mediated through a steroid-inducible factor, *Biochim. Biophys. Acta* **712**, 177–185.

Saris, C.J., Tack, B.F., Kristensen, T., Glenney, J.R., Hunter, T. (1986) The cDNA sequence for the protein-tyrosine kinase substrate p36 (calpactin I heavy chain) reveals a multidomain protein with internal repeats, *Cell* **46**, 201–212.

Stratton, J.R., Dewhurst, T.A., Kasina, S., Reno, J.M., Cerquiera, M.D., Baskin, D.G., Tait, J.F. (1995) Selective uptake of radiolabeled annexin V on acute porcine left atrial thrombi, *Circulation* **92**, 3113–3121.

Sudhof, T.C., Walker, J.H., Obrocki, J. (1982) Calelectrin self-aggregates and promotes membrane aggregation in the presence of calcium, *EMBO J.* **1**, 1167–1170.

Tanaka, K., Einaga, K., Tsuchiyama, H., Tait, J.F., Fujikawa, K. (1996) Preparation and characterization of a disulfide-linked bioconjugate of annexin V with the B-chain of urokinase: an improved fibrinolytic agent targeted to phospholipid-containing thrombi, *Biochemistry* **35**, 922–929.

Walker, J.H. (1982) Isolation from cholinergic synapses of a protein that binds to membranes in a calcium-dependent manner, *J. Neurochem.* **39**, 815–823.

Wallner, B.P., Mattaliano, R.J., Hession, C., Cate, R.L., Tizard, R., Sinclair, L.K., Foeller, C., Chow, E.P., Browning, J.L., Ramachandran, K.L., Pepinsky, R.B. (1986) Cloning and expression of human lipocortin, a phospholipase A2 inhibitor with potential anti-inflammatory activity, *Nature* **320**, 77–81.

Wang, W., Creutz, C.E. (1992) Regulation of the chromaffin granule aggregating activity of annexin I by phosphorylation, *Biochemistry* **31**, 9934–9939.

Anthology of Human Repetitive DNA

Vladimir V. Kapitonov, Adam Pavlicek, and Jerzy Jurka
Genetic Information Research Institute, 2081 Landings Drive, Mountain View, CA, USA

1 Introduction 255

2 Tandemly Arrayed Repeats 255
2.1 Microsatellites and Minisatellites 256
2.2 Satellites and Telomeric Repeats 257
2.2.1 Centromeric Satellites 258
2.2.2 Telomeric and Subtelomeric Repeats 260

3 Interspersed Repeats 261
3.1 Non-LTR Retrotransposons 263
3.1.1 L1 Non-LTR Retrotransposons 263
3.1.2 L1-dependent Nonautonomous Non-LTR Retrotransposons 265
3.1.3 CR1-like Non-LTR Retrotransposons 268
3.2 LTR Retrotransposons 269
3.3 DNA Transposons 279
3.3.1 hAT-like Transposons 280
3.3.2 Mariner/Tc1-like Transposons 282
3.3.3 piggyBac-like Transposons 282
3.3.4 MuDR-like Transposons 282

4 Repetitive Elements and Human Diseases 283

5 Potential Contribution of Repetitive DNA to Cellular Functions 295

6 Databases and Programs for Analysis of Repetitive DNA 298
6.1 Databases of Repetitive Elements 298
6.2 Programs for Detection and Analysis of Repetitive DNA 299

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Bibliography 301
Books and Reviews 301
Primary Literature 302
Seminal Early References 302
Recent Key Primary Literature 303

Keywords

AP Endonuclease
An enzyme that cleaves apyrimidinic or apurinic sites in DNA.

Autonomous Transposable Element
A mobile element encoding enzyme(s) necessary for its proliferation in the host genome. All other mobile elements are called nonautonomous.

Consensus Sequence
A DNA or protein sequence made of nucleotides or amino acid residues prevalent at each position in a set of multiply aligned homologous sequences.

cDNA
A DNA copy of an RNA molecule created during reverse transcription.

DNA Transposon
A transposable element that moves from one genomic site to another in the form of DNA only.

Family of Repetitive Elements
Multiple DNA copies derived from a number of related transposable elements called source, founder, or master genes. A repetitive family can be divided into subfamilies, each corresponding to a separate founder gene.

Interspersed Repeats
Randomly inserted copies of currently or historically active transposable elements.

Long Interspersed Repeat (LINE) Element
A non-LTR retrotransposon encoding reverse transcriptase and endonuclease catalyzing nuclear reverse transcription and integration of cDNA in the host genome.

Long Terminal Repeats (LTRs)
200 to 3000 bp long identical DNA sequences flanking a provirus sequence.

LTR Retrotransposons
A class of retrotransposons that are reverse transcribed in the cytoplasm. The resulting cDNA is integrated into the host genome, forming a provirus flanked by long terminal repeats.

Microsatellites
DNA sequences, typically shorter than 1000 bp, composed of tandemly repeated units 1 to 10 bp long (1–6 bp or 1–13 bp ranges are also used). Microsatellites are also known as short tandem repeats (STR) or simple sequence repeats (SSRs).

Minisatellites
DNA sequences composed of tandemly repeated units longer than in microsatellites. Minisatellites span from 1 kb to 100 kb and are also known as variable number of tandem repeats (VNTRs).

Non-LTR Retrotransposons
A class of retrotransposons that are transposed via target-primed reverse transcription. Autonomous non-LTR retrotransposons encode reverse transcriptase.

Processed Pseudogene (Retropseudogene)
A pseudogene derived from usually spliced and polyandenylated host mRNA by LINE-mediated reverse transcription.

Provirus
A DNA copy of a retrovirus inserted into the genome.

Pseudogene
A nonfunctional gene copy.

Recombination
The exchange of DNA or RNA molecules.

Repetitive Elements (Repeats)
DNA fragments present in multiple copies in the genome without clearly assigned biological function.

Ribonuclease H (RNase H)
An enzyme that degrades the RNA strand of RNA–DNA hybrid molecules.

Retrotransposition (Retroposition)
A process involving transcription, processing of mRNA, translation, reverse transcription of the transcribed mRNA, and integration of the resulting cDNA into the host genome.

Retrotransposon (Retroposon)
A transposable element that is reproduced via reverse transcription of an RNA intermediate. They can be split into two classes: LTR and non-LTR retrotransposons.

Retrovirus
An RNA virus, whose life cycle includes cytoplasmic reverse transcription of the RNA transcribed from its provirus integrated into the host genome.

Reverse Transcriptase
An RNA-dependent DNA polymerase.

Satellite DNA
An array of tandemly repeated units spanning up to 10^6 bp and present mainly in centromeric and paracentromeric heterochromatin. Satellite DNA usually forms a prominent "satellite band" separable from the rest of genomic DNA by density gradient centrifugation.

Short Interspersed Repeat (SINE) Element
A nonautonomous non-LTR retrotransposon derived mostly from host genes containing a pol III internal promoter. SINE elements are retrotransposed by reverse transcriptase encoded by LINE elements.

Superfamily of TEs
A group of related families of transposable elements.

Tandem Repeats
DNA units arranged in a head-to-tail manner.

Target-primed Reverse Transcription (TPRT)
A mechanism of integration of non-LTR retrotransposons into the genome.

Transposable Elements (TEs, Transposons, Mobile Elements)
DNA sequences that can move to different positions in the genome either by excision and reinsertion, or by retrotransposition.

Transposase
An enzyme encoded by a DNA transposon and involved in its insertion and excision from the host DNA.

Transposition
A relocation of a DNA sequence from one genomic site to another.

Human repetitive DNA includes tandemly arrayed and interspersed repeats. Tandemly arrayed microsatellites, minisatellites, telomeric repeats, and centromeric satellites constitute a significant portion of the genomic DNA, particularly of heterochromatin. Furthermore, around 45% of the genome is represented by so-called interspersed repeats that are mostly remnants of retrovirus-like LTR retrotransposons, non-LTR retrotransposons, and DNA transposons inserted in the genomic DNA over millions of years. Currently, the most active are L1 and Alu families of non-LTR retrotransposons, and they are known to cause genetic diseases through insertion into genes. Furthermore, repetitive sequences, particularly Alu elements, can stimulate illegitimate recombinations producing chromosomal instabilities leading to genetic disorders. Analysis of repetitive DNA is an important part of genome studies, and it is based on specialized databases and computer programs.

1 Introduction

The generally synonymous terms *repetitive elements*, *repetitive* (*reiterated*) *sequences*, and *repeats* describe a broad variety of DNA sequences with copies present in multiple locations throughout the genome. They can be divided into families of similar or sometimes identical elements. This definition also covers families of genes with known functions, such as RNA genes and other multigene families. However, the subject of this chapter is limited to repetitive elements that do not have a clearly assigned biological function.

Historically, the first categorization of repetitive DNA into "highly repetitive" and "middle repetitive" was based on DNA denaturation/renaturation kinetics. This categorization refers to the physical properties of DNA that cannot be translated satisfactorily into a classification of repeats based on cloning and detailed sequence studies. In general, highly repetitive DNA primarily includes the tandemly repeated sequences described in Sect. 2, whereas the middle-repetitive DNA includes repetitive elements interspersed with single-copy DNA (Sect. 3). However, the middle-repetitive DNA may contain sequences with a very high copy number such as the Alu family described in Sect. 3.1.2.2.

On the basis of chromosomal distribution, repetitive elements are divided into *tandem repeats*, consisting of direct repetitions of the same motif, and *interspersed repeats*. Interspersed repeats are mostly inactive copies of transposable elements (TEs). Together, these elements represent around 50% of the sequenced human genome. Some abundant interspersed repeats are relatively old and diverse and cannot be clearly distinguished from "unique" DNA. This is due to the acquisition of random mutations and sequence fragmentation over time. Therefore, it is believed that the real contribution of repetitive DNA to the human genome is underestimated, and that the cited proportion of 50% only reflects the current limits of detectability.

2 Tandemly Arrayed Repeats

Tandemly arrayed repeats are composed of multiple head-to-tail repetitions of simple

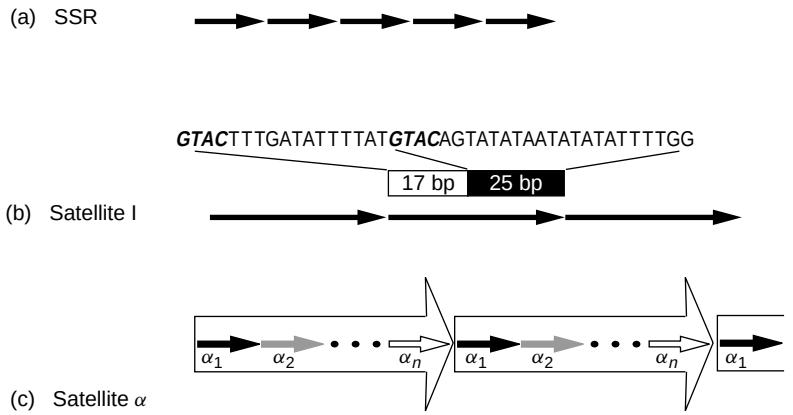

Fig. 1 Structure of human tandem repeats. (a) Schematic structure of simple sequence repeats. Arrows represent units ordered in head-to-tail fashion. (b) Structure of human classical satellite 1. Satellite 1 was originally described as a 17mer (A) and a 25mer (B) arranged in an alternating pattern A-B-A-B. This structure was inferred from the cutting of human DNA by the *RsaI* restrictase (the *RsaI* cutting sites are highlighted). Further sequence analyses revealed that satellite 1 is in fact composed of a single 42 bp-long basic unit. (c) Diagram of the hierarchical structure of human alpha satellites, showing the organization of a typical alpha satellite DNA region. Monomeric 171 bp units are represented by the small labeled arrows. The hierarchical structure of the alphoid DNA is shown by the large arrows.

sequence motifs (Fig. 1a). Human tandem repeats are composed of units ranging in size from 1 to ~1000 bp. On the basis of unit length and genomic distribution, they are somewhat arbitrarily divided into microsatellites, minisatellites, and satellites. Microsatellites are tandem arrays of short units, while minisatellites are composed of longer patterns. Micro- and minisatellites tend to be scattered all over chromosomal DNA, with some bias toward telomeric ends. The moderate degree of tandem repetition and high degree of dispersion throughout the chromosomes are the major features distinguishing micro- and minisatellites from satellites. Satellites are long, tandemly arrayed sequences located in well-defined chromosomal areas such as pericentromeric, subtelomeric, and telomeric regions. Satellites may consist of simple micro- or minisatellite-like units, but these are often organized into higher-order structures.

2.1 Microsatellites and Minisatellites

Originally, microsatellites were referred to as simple sequence repeats (SSRs) and minisatellites as variable number of tandem repeats (VNTR). Currently, the term simple sequence repeats is applied to both micro- and minisatellites. Micro- and minisatellite repeats are defined by three parameters: the *pattern* (or sequence of the unit), unit *length*, and the *number* of units.

The boundary separating micro- and minisatellites differs considerably in the scientific literature. Microsatellites are usually defined as repetition of 1–6 base pair (bp) long units, but the upper

boundary is often extended to 10–13 bp. Microsatellites are much more abundant than minisatellites.

The distribution of microsatellites in the human genome appears to be more uniform than that of minisatellites, with approximately one microsatellite for every 6 kb of DNA. Almost one-third of all potential microsatellites are simple runs of A. Such runs are often introduced as poly(A) tails of non-LTR retroposons, in particular, of Alu repeats (Sect. 3.1.2.2). Therefore, the abundance of $(A)_n$ may not be directly comparable to the abundance of other, spontaneously expanding, microsatellites. Other highly abundant microsatellites ranked in descending order are $(AC)_n$, $(AAAN)_n$, $(AAAAN)_n$, and $(AAN)_n$, including the complementary sequences. Many of these A-rich microsatellites are also spin-offs of poly(A) tails of non-LTR retroposons. Microsatellites composed of $(CAG/CTG)_n$ and $(CGG/CCG)_n$ trinucleotide repeats are less abundant, but nevertheless very important due to their connection with human diseases (Sect. 4).

The variation patterns of microsatellites and minisatellites appear to be very similar. In each case, the majority of variations are due to differences in the number of tandemly repeated units, not from mutations of the individual units. Despite the similarities between micro- and minisatellites, there are several different underlying mechanisms involved in their evolution. Indeed, the traditional division of tandem repeats into microsatellites and minisatellites is based on early hypotheses about mechanisms driving their expansion. For microsatellites, "replication slippage" during DNA replication was suggested, while minisatellite expansion was thought to result from unequal crossing-over. Nevertheless, our current knowledge suggests that processes driving SSR expansion are more complex and range from polymerase slippage and gap repair to DNA recombination and gene conversion. Another pathway of SSR expansion involves an initial comobilization of the microsatellite or minisatellite with transposition of another repeat, as in the case of poly(A) tails of non-LTR retrotransposons (Sect. 3.1).

Loci containing micro- or minisatellites produce a very large number of variant alleles, and the sequence length polymorphism makes micro- and minisatellites some of the most informative genomic markers. The most widely used microsatellites are the $(CA/GT)_n$ dinucleotide repeats. The variation at microsatellite loci has been used in genetic linkage analysis to establish the locations of disease genes as well as in forensics, human population genetics, and cancer genetics.

Many other kinds of simple repetitive sequences represent variable permutations of a limited number of motifs, sometimes interspersed with unique sequences. Such "composed" or "cryptically simple repeats" occur up to ten times more often in the genomic DNA than in any equivalent random sequence. They result from a systematic reshuffling of various DNA fragments differing in length and sequence composition. Therefore, they are of interest from the standpoint of genetic variability. These repeats have not yet been systematically classified or explored.

2.2 Satellites and Telomeric Repeats

Typically, satellites are arrays of 10^3–10^7 tandemly repeated units, or monomers, predominantly located in the well-defined chromosomal regions, such as centromeres and heterochromatin. Traditionally, the term *satellites* was restricted to centromere-specific repeats; however,

telomeric repeats can also be viewed as satellite-like sequences. Because of their distinct base composition, many satellites were originally discovered as a result of their separation from bulk DNA by centrifugation in buoyant density gradients. Satellites with GC content similar to that of the unique genomic DNA, which could not be separated by centrifugation, have been called *cryptic satellites*. The function of satellite DNA remains unclear. Satellites represent up to 20% of the genomic DNA, but they are underrepresented in the sequenced portion of the genome.

Table 1 lists representative human satellite sequences collected in Repbase Update (RU), the reference collection of repetitive sequences (Sect. 6.1). The most extensively studied centromere- and telomere-specific repeats are reviewed in the next two sections. In addition, there are several chromosome-specific tandem repeats, some of which are listed in Table 1.

2.2.1 Centromeric Satellites

Human centromeric DNA contains a variety of different satellite repeats. The major noncryptic satellite fractions were

Tab. 1 Human satellites and satellite-like sequences in repbase update.

Repbase	*Description*
Centromeric	
SAR	Satellite I
GAATG	Satellites II and III, both contain 5-ATTCC basic unit
HSATII	Satellite II (example, diverged copy)
ALR	Alpha satellite
ALR1	Alpha satellite subfamily 1
ALR2	Alpha satellite subfamily 2
BSR	Beta satellite
GSAT	Gamma satellite
GSATX	Gamma X satellite, X chromosome specific
CER	48-bp satellite
SN5	SN5 satellite
LSAU	Heterochromatic satellite
HSAT4	35-bp centromeric repeat
HSAT5	Centromeric repeat, 14-bp unit
GSATII	Centromere-associated repeat
D20S16	Centromere-associated repeat
Telomeric	
TAR1	Telomere-associated repeat sequence, TAR1 or SUBTEL_sat in RepeatMasker
REP522	Repetitive subtelomeric-like sequence
Others	
SATR1	Primate satellite-like sequence
SATR2	Primate satellite-like sequence
MSR1	37 bp minisatellite repeats, specific to chromosome 19, but detected in chromosome 19-derived segmental duplications located in other chromosomes
MER22	SSTI moderately repetitive family, strictly tandem repeat, appearing at 2 chromosomal locations only
ACRO1	Satellite/acromeric in all short arms of acromeric chromosomes
HSATI	Chromosome 22 satellite

defined as classical satellites I, II, III, and IV. Classical satellite I was found to be dominated by a simple component called *satellite 1*. Satellite 1 contains a 42-bp basic unit, originally described as a 17-mer (A) and a 25-mer (B), arranged in an alternating pattern: A-B-A-B (Fig. 1b). The major simple components of classical satellites II and III are named *satellites 2 and 3*, both containing the GAATG unit (sometimes described by the complementary ATTCC). Repetitive sequences from the satellite DNA fraction IV are nearly identical to those in fraction III.

Among other well-defined and extensively studied components of classical satellites II and III are the so-called *alpha satellites*, first discovered in African green monkeys and subsequently in other primates. Alpha satellite DNA is composed of a 171-bp unit (Fig. 1c). Interchromosomal sequence identities suggest that all human alpha satellites arose from two or three different ancestral monomers, which in turn diverged into 12 distinct types of alphoid monomers. They can differ by as much as 20 to 40% from one another by sequence identity.

Human chromosomes contain alpha satellites organized hierarchically into higher-order repeat arrays (Fig. 1c), which can differ between different chromosomal locations. On the basis of this organization, the alpha satellites from the human genome have been divided into five suprachromosomal families: the first family consists of alpha satellites located on chromosomes 1, 3, 5, 6, 7, 10, 12, 16, and 19; the second comprises chromosomes 2, 4, 8, 9, 13, 14, 15, 18, 20, 21, and 22; the third is located on chromosomes 1, 11, 17, and X; the fourth family is found on chromosomes 13, 14, 15, 21, 22, and Y; and, finally, the fifth family was detected on chromosomes 5, 6, 13, 14, 19, and 21.

Members of each suprachromosomal family share a common alpha satellite repeat unit. The size, conservation, and organization of these higher-order structures depend on the chromosome and on the physical distance from the centromere. Apart from the five classical suprachromosomal families, other higher-order structures have also been described. For example, the central region of the chromosome X centromere contains a specific alpha satellite array arranged into 2-kb long highly identical blocks.

The evolution of alpha satellites is complex, and is driven by intra- and interchromosomal recombination and gene conversion. The recombination occurs during chromosomal pairing and leads to an unequal crossing-over between chromosomes. As a result, stretches of alpha satellite DNA expand and contract at a very rapid rate. In addition, the alpha satellite regions of chromosomes within the same suprachromosomal families may undergo exchanges between nonhomologous chromosomes, thus increasing the mode and tempo at which these sequences evolve. However, the intrachromosomal recombination within alpha satellites appears to be the predominant evolutionary driving force. In general, the average intrafamily identity of alpha satellite monomers (~90%) is much higher than the interchromosomal identity (~75%). In addition, the average intrafamily identity within a species (99% in human and gorilla) is generally greater than that between species (90% between the human and gorilla satellite monomers). Humans and gorillas share the basic monomer as well as a higher-order repeat unit for suprachromosomal family 3, suggesting that this alpha satellite structure predates the human/great ape divergence. Thus, the alpha satellites within each species have evolved

in a "concerted" manner, giving rise to a significantly lower nucleotide divergence within a species than between species.

Alpha satellites have been found only within primate genomes, usually located around the centromere, and can bind key centromere proteins. Occasionally, however, ectopic centromeres are formed in noncentromeric segments of chromosomes, producing cytogenetic aberrations including visible karyotype disorders. These new centromeres were termed *neocentromeres*. However, not all neocentromeres detected in humans are derived from alpha satellites. About 60 reported neocentromeres completely lack alpha satellites but maintain their ability to bind main centromeric proteins. These neocentromeric sequences are promising material for the construction of artificial chromosomes. However, these results also suggest that centromeric repeats, including alpha satellites, may not have any essential role in the genome.

Table 1 lists other reported human centromeric satellites. The human *beta* (Sau3A) satellite family is defined by the 68-bp GC-rich monomer. This satellite is present in pericentromeric regions and in the short arms of human acrocentric chromosomes. *Gamma* satellite DNA was initially detected as a 220-bp repeat located on chromosome 8. A second subfamily of gamma satellite, *gamma X*, was identified in the centromeric region of the X chromosome. Both families share 220-bp units 62% identical to each other. The CER satellite family, characterized by the 48-bp unit, is present on chromosomes 13, 14, 15, 21, and Y.

Among poorly defined satellite families are the Sau (*Lsau*) and Sn5 families characterized by the 498- and 455-bp units, respectively. Many centromeric satellites listed in Table 1 were defined by sequence analyses. Classification of centromeric satellites continues, and additional families may be detected during the course of sequencing centromeric and heterochromatin regions.

2.2.2 Telomeric and Subtelomeric Repeats

In contrast to the highly variable monomers in centromeric satellites, the *TTAGGG* telomeric monomers are conserved in all vertebrates. They form tandem arrays 3 to 20 kb long and their conservation is indicative of the importance of telomeres for normal chromosomal maintenance and viability. Historically, telomeres have been found to be associated with the nuclear envelope, and they have been implicated in chromosomal positioning within the cell. Telomeres are implicated in aging, immortalization of cells, and chromosomal rearrangements.

The maintenance of telomeric repeats presents a unique challenge to the cell, since standard DNA replication would result in the progressive shortening of the telomere over time. The replication of human chromosomes is catalyzed by DNA polymerases. However, DNA polymerases require short template primers for the initiation of replication. Thus, the ends of chromosomes or telomeres would become gradually shorter over time through the exclusive use of this mode of replication. The length of telomeres is maintained by a special replication mechanism for the ends of chromosomes involving a telomere-specific terminal transferase. This enzyme, called *telomerase*, is a ribonucleoprotein (RNP), and it uses the RNA component as a template to initiate the replication of the telomeres. The enzymatic addition of bases is catalyzed by the protein component. Faithful replication of telomeres

protects chromosome ends from degradation, fusion, and recombination.

The length of telomeres appears to depend on the number of cell divisions with a general tendency toward shortening. This has been implicated in chromosomal aberrations, involving translocations in differentiated and abnormal cells. The length of telomeres has been used as a measure of the age of individual cells and is implicated indirectly in aging. It is still unclear whether changes in the telomere length have a direct impact on the aging process of a cell or whether they are the consequence of aging, possibly through a decrease in the telomerase fidelity.

The 100 to 300-kb long sequences adjacent to telomeric repeats are called *subtelomeric repeats*. They represent the boundary between the telomere and the rest of the chromosome. Most subtelomeric repeats contain motifs similar to TTAGGG, but there are exceptions such as *REP522*. Unlike telomeric repeats, most subtelomeric repeats are not substantially conserved in vertebrates. In fact, they are highly polymorphic in humans and even in different chromosomes. The distal ends of human chromosomes are characterized by an enhanced rate of recombination and aberrations in subtelomeric regions have been shown to be associated with several congenital abnormalities. The high divergence of subtelomeric repeats may prevent recombinations between nonhomologous chromosomes and increase their stability.

3 Interspersed Repeats

Usually, interspersed repeats are scattered remnants of transposable elements that were transpositionally active either in the human genome or genomes of our evolutionary ancestors. Active TEs have produced families and subfamilies of interspersed repeats of different age and size. There are over 600 families of human interspersed repeats represented in Repbase (Sect. 6.1). Most of them originated and became extinct during the last 30 to 150 million years but some of the oldest ones ceased to proliferate as long as 200 million years ago and are barely identifiable by sequence analysis. Detectable interspersed repeats represent ~45% of the human genome. Given the limits of detectability, it is very likely that the bulk of the remaining nonprotein coding DNA is also composed of very old, unrecognizable remnants of TEs.

The consensus sequence of a TE can be reconstructed by applying the so-called *simple majority rule* to its multiply aligned interspersed copies (Fig. 2). The average similarity of diverged repeats to their consensus sequence (y) is higher than the average similarity between individual repeats (x). The approximate relation between the two variables is given by the following simple formula:

$$y = \frac{1 + \sqrt{12x - 3}}{4}.$$

In many cases, consensus sequence is a quite accurate approximation of the original TE that generated a particular family of repeats. It can be further improved by correcting stop codons in open reading frames (ORFs) or restoring rapidly mutating CpG dinucleotides, which sometimes erroneously converge to CpA or TpG in the consensus sequence. At least one TE reconstructed this way, named *Sleeping Beauty*, has been shown to transpose *in vivo*.

Transposable elements harbored by the human genome can be divided into three

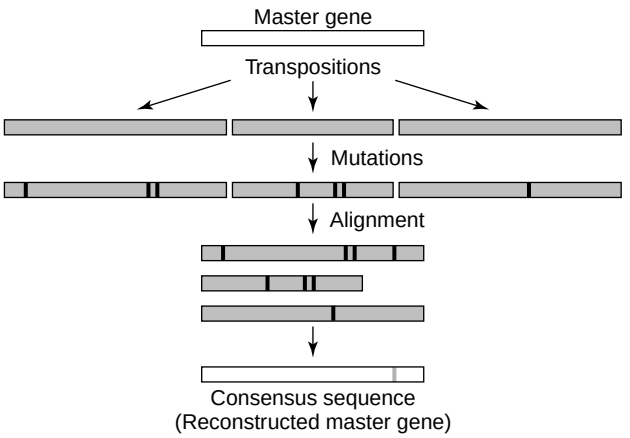

Fig. 2 Molecular paleontology of transposable elements. The top white rectangle represents an ancient active transposon that multiplied into interspersed repeats shown as gray rectangles subsequently marked by black vertical bars indicating mutations. Most of these mutations can be "undone" after multiple alignments and construction of a consensus sequence representing the reconstructed transposon (bottom white rectangle). The gray bar in the consensus sequence indicates an unresolved mutation.

groups. Two of these groups belong to a class of retroelements that include non-LTR retrotransposons and endogenous retrovirus-like elements also known as *long terminal repeat* (LTR) retrotransposons. Retroelements proliferate via reverse transcription of their RNA expressed in the host cell. An RNA copy of the transcribed retroelement is used as a template to produce its DNA copy, the so-called cDNA, which can be inserted in the genome. In non-LTR retrotransposons, a nick in the host DNA primes cDNA formation, whereas in retroviral-like elements, the DNA nicking and cDNA integration occur after reverse transcription. The nicking and cDNA formation are catalyzed by the endonuclease/integrase (IN) and reverse transcriptase (RT), respectively, encoded by the transposable element itself or by one of its autonomous relatives present in the genome. Unlike DNA transposons, active retroelements do not relocate from their original position.

The third group of TEs is represented by the *cut-and-paste* DNA transposons, called so because they are excised (cut) from the original genomic location and inserted (pasted) into a new site. Typically, a single enzyme called *transposase* mediates the process.

All groups of TEs are composed of autonomous and nonautonomous elements. An autonomous element encodes a complete set of enzymes characteristic of its family. Nonautonomous families use the protein machinery of their autonomous relatives. Therefore, the nonautonomous TEs can be viewed as parasites competing for enzymes encoded by the autonomous elements. Both groups also require additional cellular factors for amplification.

3.1 Non-LTR Retrotransposons

Detectable copies of non-LTR retrotransposons contribute ~34% of the human genome compared to the ~11% representing other detectable interspersed repeats. Non-LTR retroelements appear to be strictly intracellular parasites, inherited for millions of years rather than passed horizontally. For historical reasons, some autonomous non-LTR retrotransposons are referred to as long interspersed repeats (LINEs) due to the fact that they are several kilobases long. Analogously, their short nonautonomous associates are referred to as short interspersed repeats (SINEs). Autonomous non-LTR retrotransposons in eukaryotic species are currently divided into 13 superfamilies or clades: L1, CRE, NeSL, R4, R2, RTE, Tad1, R1, LOA, I, Ingi, Jockey, CR1, and Penelope. Only two of them, L1 and CR1 superfamilies are represented in the human genome. The human population continues to carry active elements of L1 non-LTR retrotransposons whereas CR1-like retroelements are extinct. L1-dependent Alu and SINE, VNTRs and ALU (SVA) nonautonomous retroelements also continue to proliferate in the human population.

3.1.1 L1 Non-LTR Retrotransposons

The typical L1 (or LINE1) autonomous retrotransposon is ~6 kb long, and contains two open reading frames referred to as ORF1 and ORF2, encoding the ~500-aa and ~1000-aa proteins, respectively (Fig. 3). The function of the L1-ORF1p protein is not well understood. Since it has a common leucine zipper (LZ) motif present in its sequence, it presumably binds RNA/DNA molecules. The L1-ORF2 protein includes the apurinic–apyrimidinic endonuclease (APE), reverse transcriptase (RT), and conserved cysteine-rich (C) domains. The endonuclease domain is involved in nicking the host DNA, followed by priming of an RNA-directed DNA synthesis that is catalyzed by the RT domain.

Out of several hundred thousand L1 copies present in the human genome, about 5000 are full-length elements and only around 100 contain intact open reading frames. In addition to ORF1 and ORF2, the full-length L1 elements contain 5′ and 3′ untranslated regions (UTRs), each several hundred bp long. The 5′ UTR includes a poorly characterized internal pol II transcription promoter. Unlike most human genes that contain pol II promoters upstream of the transcription start site, the L1 promoter is located downstream of the start site. Occasionally, the 5′ L1 UTR may also promote transcription that starts in the strand complementary to the 5′ UTR and extends to the genomic region flanking the L1 5′ end. The 3′ L1 UTR includes an AATAAA polyadenylation signal and a polyA tail.

In addition to the transcription, the L1 life cycle includes export of the transcribed

Fig. 3 Schematic structure of the L1 non-LTR retrotransposon. It consists of a 5′ untranslated region (5′ UTR) with pol II internal promoter, two open reading frames (ORF1 and ORF2), a 3′ untranslated region, a polyA signal (AATAAA), and a polyA tail. L1 elements are flanked by target site duplications (TSDs). The ORF1 protein includes a conserved leucine zipper motif (LZ). The ORF2 protein includes domains for the apurinic/apyrimidinic-like endonuclease (APE), RT, and a conserved cysteine-rich motif (C).

RNA from the nucleus to the cytoplasm, translation and formation of ribonucleoprotein particles, which are transported back into the nucleus, reverse transcription, and integration into the genome. L1 elements lack splice signals and therefore, L1 mRNA is bi-cistronic with two open reading frames separated by a short noncoding spacer. Translation of ORF1 is likely to be initiated at the 5′ UTR by ribosomal scanning. ORF2 translation may be initiated by an internal ribosomal entry site (IRES) or by the so-called *ribosomal shunting* or *jumping*. Also, it is unclear how the L1-encoded proteins enter the nucleus. Usually, nuclear export of proteins as large as those encoded by ORF1 and ORF2 is driven by specific amino acid sequences called *nuclear localization signals*. As there are no known nuclear localization signals encoded by either ORF1 or ORF2 proteins, the mechanism involved in their entry into the nucleus remains a mystery. It is even unclear if the ORF1 translation products ever enter the nuclear compartment.

The model of reverse transcription and integration of L1 into the genome invokes a coupled process known as *target-primed reverse transcription* (TPRT). According to this model, the APE domain of the ORF2 protein recognizes a 6-bp target site described by the TTAAAA consensus sequence, and nicks the second DNA strand between the bases complementary to TT and AAAA (Fig. 4). The free 3′-OH end of the nick is then used for priming reverse transcription of L1 RNA by the ORF2 reverse transcriptase, starting from the 3′ polyA tail. Eventually, a second nick is created on the first strand, preferably ~15 bp downstream of the first nicking site, leading to a staggered break and, after integration, to the formation of target site duplications (TSDs) flanking the integrated sequence. The typical size of TSDs generated by L1 retrotransposition is ~15 bp, but the size may vary and sometimes even short deletions of target sequences are observed. The final integration steps, including

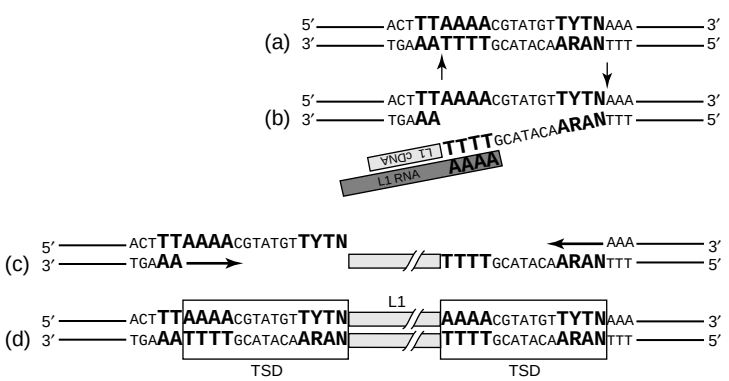

Fig. 4 The target-primed reverse transcription (TPRT) mechanism. (a) The L1 endonuclease cleaves the antiparallel DNA strand between 3′ AA and 3′ TTTT, and exposes a free 3′ hydroxyl group at the nick, serving as a primer for reverse transcription of L1 mRNA. (b) The second nicking occurs ~15 nucleotides from the first nicking site, preceded by a TYTN consensus sequence. (c) The remaining steps include elimination of L1 RNA, DNA synthesis, and repair, probably mediated by the host enzymes. (d) Integrated L1 retroelement and its flanking sites.

synthesis of the second strand, ligation, and elimination of the L1 RNA, are probably mediated by host enzymes.

The majority of L1 elements undergo 5′ truncations and inversions during retroposition and, as a result, they are represented mostly by 3′ terminal portions that are shorter than 1 kb in length. Possible reasons for the truncation may be digestion of the L1 RNA by cellular ribonucleases, or premature termination of the reverse transcription, or both.

Occasionally, transcription of L1 elements does not stop at the L1 polyadenylation signal but continues until another transcription termination signal encoded by the host DNA is encountered. As a result, host sequences adjacent to the 3′ L1 terminus and the new polyA sites may be transduced by the L1 to a new genomic location. Over 10% of the analyzed L1 insertions in the human genome contain 3′ transduced sequences that range from 30 bp to several kb in length. This represents an attractive mechanism that could be involved in exon shuffling in the host genome. However, no real exons have yet been detected in L1-transduced sequences. Therefore, it is possible that the L1-mediated 3′ transduction is suppressed in the vicinity of human genes. Some L1 retrotransposons can also be accidentally transcribed from upstream promoters leading to transduction of genomic DNA flanking the 5′ ends of L1 elements. Also, new L1 subfamilies can be generated by the addition of both 5′ and 3′ flanking DNA.

3.1.2 L1-dependent Nonautonomous Non-LTR Retrotransposons

There are two major groups of L1-dependent non-LTR retrotransposons present in the human genome: processed pseudogenes, and SINEs that include Alu and SVA elements. Elements from both groups are flanked by ~15-bp TSDs, have the 3′ polyA tails, and are inserted preferentially into the TTAAAA target sites. It has also been demonstrated experimentally that the propagation of processed pseudogenes and Alu elements is mediated by the L1-encoded RT. Therefore, both processed pseudogenes and Alu repeats are classified as nonautonomous non-LTR retroelements mobilized by the L1 retroposition machinery. The L1 reverse transcriptase, in a cis-acting manner, preferentially binds the same mRNA molecule it was translated from. Therefore, Alu and other nonautonomous retropseudogenes probably intercept the L1 RT while interacting with ribosomal components.

3.1.2.1 **Processed pseudogenes** Processed pseudogenes, or retropseudogenes, are products of occasional retrotransposition of spliced mRNA formed during regular transcription of the host genes (Fig. 5). Since transcription of these genes is usually driven by external pol II promoters, processed pseudogenes do not include the promoters. Therefore, it is extremely unlikely that they can be repeatedly transcribed and retrotransposed. Nevertheless, the human genome is estimated to contain at least 30 000 retropseudogenes formed from a variety of host genes. The most abundant retropseudogenes appear to be derived from highly expressed genes ranging from those coding for small nuclear, nucleolar, and cytoplasmic RNA, to those coding for ribosomal proteins.

3.1.2.2 **Alu repeats** Alu elements were first discovered using DNA renaturation–reassociation experiments that were performed during the early era of repeat studies. They are represented by more

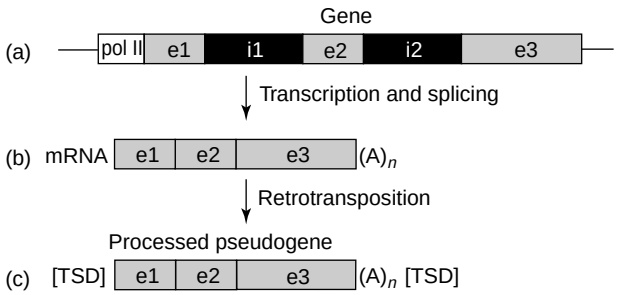

Fig. 5 Formation of processed pseudogenes (retropseudogenes). (a) The original gene that is transcribed into RNA, excluding pol II promoter. Exons e1–e3 and introns i1, i2 are indicated. (b) mRNA afer splicing and polyadenylation. (c) L1-retrotransposed, processed retropseudogene with ~15 bp TSDs.

than one million copies in the human genome, which makes them numerically the most abundant of all human interspersed repeats. The name "Alu" is derived from an *AluI* restriction site present in these elements.

Full-length Alu elements are ~300-bp long (Fig. 6) dimers composed of two homologous 120- and 148-bp left and right monomers respectively, derived from the *7SL RNA* gene ~100 million years (Myr) ago. The 7SL RNA molecule is an important component of the signal recognition particle (SRP) that binds to nascent signal sequences and transiently arrests translation. The *7SL RNA* gene contains the pol III internal promoter composed of the conserved ~15-bp A and B boxes. The right monomer contains an additional 31-bp 7SL-derived sequence

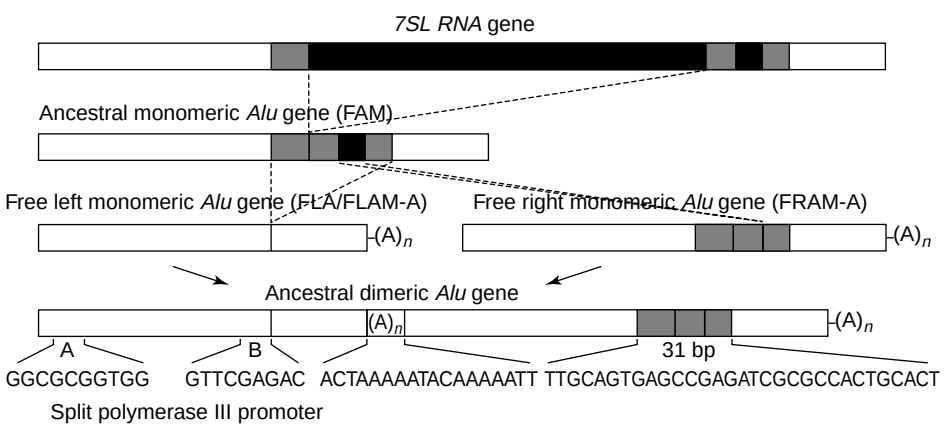

Fig. 6 A scheme for the origin and early evolution of human Alu master (source) genes. Dimeric Alu was derived from the *7SL RNA* gene after a two-step deletion process leading first to an ancestral family of Alu-like monomers (FAM) and then to other monomers, of which two, FLA/FLAM-A, and FRAM, merged into a precursor of dimeric Alu characteristic of primates.

relative to the left monomer, and the monomers are separated by the A-rich central region. The right monomer also contains a 3′-oligo-dA-rich tail (or polyA tail), usually 20 to 30 bp long. Free Alu monomers such as FLA, FLAM, and FRAM were very active during the early era of Alu evolution, and at least one of them, BC200, has been recruited as a regulated gene expressed in the brains of higher primates.

Probably around one hundred Alu elements out of one million appear to be currently participating in retrotransposition. It is unclear why Alu retroposition is confined to this limited number of "master" or "source" genes. One often invoked explanation is that most retroposed elements remain untranscribed. For example, pol III transcription depends on flanking sequences and most Alu elements may be inserted into genomic sites that do not complement their pol III promoters. Furthermore, most Alu sequences are heavily methylated at CpG positions that are also present in the promoter region. The methylation occurs at position 5 of cytosine residues. Deamination of 5-methyl cytosine produces thymine – a mutation that is likely to escape the repair mechanism. As a result, CpG dinucleotides mutate to TpG and CpA at a rate approximately ten times higher than that of the non-CpG dinucleotides. Other regions of rapid mutations include A-rich linker and tail regions often associated with microsatellite expansion. This rapid rate of mutations represents another potential way to eliminate or mitigate the ability of any particular Alu repeat to mobilize. Thus, the majority of Alu repeats appear to be fossil relics that integrate in the genome and accumulate mutations in a random manner, characteristic of pseudogenes.

The human Alu family has been divided into nine major subfamilies: AluJo, Jb, Sz, Sx, Sg, Sq, Sp, Sc, and Y, ordered approximately from the oldest to the youngest (Table 2). AluJo and Jb form a separate branch distinct from the AluS branch that includes all the other seven major subfamilies. Each Alu subfamily represents a set of elements derived from a small group of closely related master genes. As a result, members of the same family share, in principle, the same consensus sequence, which is different from

Tab. 2 Average chromosomal density and estimated age of major Alu subfamilies.

Alu subfamily[a]	*AluJo*	*AluJb*	*AluSz*	*AluSx*	*AluSg*	*AluSq*	*AluSp*	*AluSc*	*AluY*
Number[b]	216 465	130 297	192 932	155 182	88 697	106 344	61 465	47 071	146 176
Proportion[c]	18.91	11.3	16.9	13.6	7.8	9.3	5.4	4.1	12.8
Genomic density (%)	1.60	1.1	1.9	1.5	0.9	1.0	0.6	0.5	1.4
Age (Myr)[d]	81	81	48	37	31	44	37	35	19

[a]Alu subfamilies were classified by censor. Alu classification may slightly differ between different programs and parameters used for Alu detection.
[b]Numbers of Alu elements from major subfamilies present in the sequenced part of the human genome (2805 Mb).
[c]Proportions of elements from the corresponding subfamilies relative to all Alu sequences.
[d]Age (in million years), based on Kimura's distance.

consensus sequences for other subfamilies. AluJo and AluJb are the two oldest subfamilies that proliferated ~65 million years ago. AluY is the youngest among major AluS subfamilies as it was retrotransposed 5–30 million years ago. AluY gave rise to a number of still younger sub(sub)families of which the following 22 have been reported so far: AluYa1, Ya4, Ya5, Ya8, Yb3a1, Yb3a2, Yb8, Yb9, Ybc3a, Yc1, Yc2, Yd2, Yd3, Yd3a1, Yd8, Ye2, Ye5, Yf1, Yf2, Yg6, Yh9, Yi6. At least some of these sub(sub)families still contain elements actively proliferating in human populations, as indicated by their insertion polymorphism. Among the largest and best-studied active sub(sub)families are AluYa5 and Yb8. However, some of the smaller ones such as AluYb9, Yc2, Yg6 and Yi6 are also of growing interest in human population genetics studies.

Densities of recently retroposed Alu elements are about three times higher on chromosome Y than on chromosome X, and about two times higher than on autosomes. This indicates that Alu elements are primarily retroposed in paternal germlines. However, these original chromosomal proportions change very rapidly with time. The most striking long-term trend is the accumulation of Alu elements in GC-rich and gene-rich regions. At the same time L1 elements are predominant in AT-rich and gene-poor regions. Given that newly retrotransposed Alu and L1 elements are inserted into regions of similar GC content, the evolution of the human genome is characterized by massive subsequent redistributions of Alu repeats. The most likely factor involved in the redistributions may be related to the different patterns of postinsertion duplications and deletions.

3.1.3 CR1-like Non-LTR Retrotransposons

CR1 is one of the most abundant and widely distributed clades of non-LTR retrotransposons that are present in the genomes of birds, amphibians, fishes, reptiles, invertebrates, and mammals. Typically, autonomous CR1-like elements are several kb long and contain two open reading frames, ORF1 and ORF2, coding for multidomain proteins. The ORF1p may be involved in DNA/RNA binding, but other interesting functions are also likely (see below). The ORF2 protein includes the APE and RT domains, which are distantly similar to those identified in the L1 ORF2 protein. CR1-like elements do not generate target site duplications.

3.1.3.1 L2 and L3 retroelements

Two human families of long interspersed repeats called L2 and L3 were most likely amplified from active autonomous CR1-like retroelements (Fig. 7). Given the high sequence diversity, their retrotransposition probably occurred 200 to 300 million years ago, long before mammalian radiation. Together with the nonautonomous elements described below, L2 and L3 represent ~5% of the human genome. Most of the preserved L2 and L3 copies are 5′ truncated, which, in addition to their very old age, impedes reconstruction of their source genes. For example, the ~3300-bp L2 consensus sequence does not seem to contain the ORF1 protein that is present in other CR1-like retroelements. However, it is not clear whether this is due to incomplete reconstruction or because the active L2 source gene lacked such a protein. The ORF1 protein is present in the L3 consensus sequence, and it contains the esterase domain that is also present in the ORF1 proteins encoded by CR1-like elements in birds, reptiles, and fishes. Unlike L1, neither L2 nor L3 elements are flanked

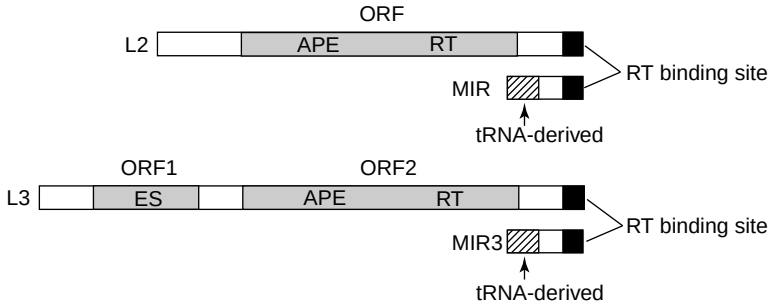

Fig. 7 L2 and L3 autonomous CR1-like retrotransposons and their nonautonomous companions MIR and MIR3. L2 and L3 encode enzymes containing endonuclease (APE) and reverse transcriptase (RT) domains. L3-ORF1 is homologous to esterase (ES).

by target site duplications. Furthermore, instead of the 3′ polyA tail, L2 and L3 carry the $(TGAA)_3$ and $(GATTCTAT)_2$ microsatellites at their 3′ termini respectively.

3.1.3.2 **Nonautonomous MIR and MIR3 elements** MIR and MIR3 are considered to be nonautonomous elements amplified by L2- and L3-encoded proteins, respectively. MIR is a classical 262-bp SINE element derived from a tRNA-like sequence and a 50 bp sequence fragment homologous to the 3′ terminus of the L2 element (Fig. 7). This terminus was probably recognized by the L2-encoded RT that led to amplification of MIRs. Approximately 4% of the human genome is made up of L2 and MIR copies.

MIR3 is viewed as a nonautonomous element that parasitized the L3 retroposition machinery (Fig. 7). Analogously to MIR and L2, MIR3 and L3 also share the ~50-bp 3′ termini, which are likely to serve as binding sites of the L3-encoded RT. In addition, the ~140-bp 5′ terminal portions of the MIR3 and MIR consensus sequences are 81% identical to each other. This portion is homologous to tRNA and probably includes the pol III internal promoter essential for proliferation of MIR3 and MIR retroelements. Approximately 0.5% of the human genome is derived from L3 and MIR3 elements.

3.2 LTR Retrotransposons

Human long terminal repeat retrotransposons are retrovirus-like elements retrotransposed in the genome and inherited by the host from generation to generation. It is believed that human LTR retrotransposons are descendants of exogenous retroviruses that repeatedly infected germlines of human ancestors and became endogenous. Endogenous retroviruses retain their ability to retrotranspose within cells for some time, but cannot be transmitted efficiently between cells. Their subsequent evolution reflects adaptation to an intracellular environment, simplification, and formation of nonautonomous copies.

Most human endogenous retroviruses (HERVs) are extinct and are represented only by mutated interspersed copies, particularly by LTRs. The interspersed copies of LTR retrotransposons constitute around 9% of the human genome. Although LTR retrotransposons are not as abundant as

the non-LTR retrotransposons, they are by far the most diverse and complex TEs harbored by our genome. Approximately 200 families of LTR retrotransposons have been identified in the human genome. A vast majority of them were identified and reconstructed using computer-assisted sequence analysis.

On the basis of the similarities to known animal exogenous retroviruses, all LTR retrotransposons present in the human genome can be divided into three classes (Table 3).

Class I non-LTR retrotransposons are similar to gamma retroviruses including mammalian type C retroviruses such as murine leukemia virus (MLV). All elements from this class are flanked by 4- and 5-bp TSDs that are generated upon their integration in the genome (Table 4).

Class II non-LTR retroelements are similar to beta retroviruses represented by mammalian type B and D retroviruses such as the mouse mammary tumor virus (MMTV), Mason-Pfizer monkey virus, and primate lentiviruses. The latter include human immunodeficiency virus (HIV). All Class II retroelements are characterized by 6-bp target site duplications (Table 5). They were probably active in hominoids including *Homo sapiens*.

Class III families represent the oldest LTR retrotransposons detected in the human genome. Some of them were active in the common ancestor of primates and other mammals. Class III elements amplified to a larger copy number than Class I and Class II elements together. Class III families are related to spumaviruses such as human foamy virus, and they are characterized by 5-bp target site duplications (Table 6).

Structural features of LTR retrotransposons

A typical retrovirus is inserted into the genome in the form of a provirus containing two copies of 200 to 3000-bp long LTRs, which are identical to each other at the time of insertion (Fig. 8). Most LTRs have conserved 5′-TG and CA-3′ termini, but in some families such as HERVK11 and HERVK11D, LTRs contain TA-3′ termini instead of CA-3′. The internal portion of a provirus is defined as a sequence flanked by LTRs. Usually

Tab. 3 Classification of human LTR retrotransposons.

Human LTR retrotransposons	*Percentage of the genome [%]*	*Target site duplications [bp]*	*Related exogenous retroviruses*	*Age [Myr]*
Class I	2.5	4–5	Gamma retroviruses (C type[a]): Murine leukemia virus	20–80
Class II	0.5	6	Beta retroviruses (B type[b], D type[c]): mouse mammary tumor virus; Mason–Pfizer monkey virus; lentiviruses	1–35
Class III	6	5	Spumaviruses: human foamy virus	20–150

Distinctive features seen in transmission electron micrographs:
[a]Central and spherical viral core, "C particles".
[b]Accentric and spherical viral core, "B particles".
[c]Cylindrical core, "D particles".

Tab. 4 Class I endogenous retroviruses.

Family name	*Names in repbase update*		*Alternative names*[a]	*Coding status*	*tRNA*[b]
	Internal portion	*LTR*			
HERV-I	HERVI	LTR10B, LTR10A, LTR10C, LTR10E	RTVL-I	Gag, PR, RT, RH, IN, ENV	I
HERVIK10D	HERVIP10D	LTR10D		Gag, PR, RT, RH, IN, ENV	K
HERVIP10F	HERVIP10F	LTR10F		Gag, PR, RT, RH, IN, ENV	I/P
HERVIP10FH	HERVIP10FH	LTR10F		Nonautonomous	I/P
HERV-E	HERVE	LTR2		Gag, PR, RT, RH, IN, ENV	E
Harlequin	HARLEQUIN	LTR2B, LTR2C		Nonautonomous	E
HERV3	HERV3	LTR4	BaEV, HERV-R, ERV3	Gag, PR, RT, RH, IN, ENV	R/L
HERV15	HERV15I	LTR15	RRHERV-I, HSRIRT	Gag, PR, RT, RH, IN, ENV	I
HERV17	HERV17	LTR17	HERV-W	Gag, PR, RT, RH, IN, ENV	R/W
HERVS71	HERVS71	LTR6A, LTR6B	S71	Gag, PR, RT, RH, IN, ENV	T
HERV9	HERV9	LTR12, LTR12B, LTR12C, LTR12D, LTR12E	HERV-pHE1, ERV9, PTR5, HRES-1/1	Gag, PR, RT, RH, IN, ENV	R
MER52AI	MER52AI	MER52A MER52B, MER52C, MER52D		Gag, PR, RT, RH, IN	P
HERVG25	HERVG25	LTR25		Gag, PR, RT, RH, IN, Env	G
HUERS-P3	HUERS-P3	LTR9	HuRRS-P, HERV-P	Gag, PR, RT, RH, IN, Env	P
HUERS-P3B	HUERS-P3B	LTR9B		Nonautonomous Env	P
HUERS-P2	HUERS-P2	LTR1, LTR1B, LTR1C, LTR1D		Nonautonomous Gag	P
HUERS-P1	HUERS-P1	LTR8, LTR8A		?	P
HERV-H	HERVH	LTR7A, LTR7B	RTVL-H, RTVL-H2, RGH	Gag, PR, RT, RH, IN, Env	H
HERV19	HERV19I	LTR19A		Nonautonomous	F
HERVFH19	HERVFH19I	LTR19B, LTR19C,	HERV-F	Gag, PR, RT, RH, IN, Env	F

(*continued overleaf*)

Tab. 4 (*continued*)

Family name	*Names in repbase update*		*Alternative names*[a]	*Coding status*	*tRNA*[b]
	Internal portion	*LTR*			
HERVH48	HERVH48I	MER48	HERV-Fb	Gag, PR, RT, RH, IN, Env	F
HERVFH21	HERVFH21I	LTR21A, LTR21B	HERV-XA34, HERV-F(XA34)	Gag, PR, RT, RH, IN, Env	F
HERV46	HERV46I	LTR46		Gag, PR, RT, RH, IN, Env	F
PABL_A	PABL_AI	PABL_A	HERV-Rb	Nonautonomous	R/L
PABL_B	PABL_BI	PABL_B		Gag, PR, RT, RH, IN, Env	R/L
LOR1	LOR1_I	LOR1		Nonautonomous	E
MER51	MER51I	MER51A-E MER4		Nonautonomous	E
MER110	MER110I	MER110, MER110A		?	?
HERV23	HERV23I	LTR23		Nonautonomous	?
HERV30	HERV30I	LTR30		Gag, PR, RT, RH, IN, Env	R/L
MER34B	MER34B_I	MER34B		Nonautonomous	?
HERV35	HERV35I	LTR35		Nonautonomous	?
HERV38	HERV38I	LTR38, LTR38B-C		Nonautonomous	M
HERV39	HERV39	LTR39		Nonautonomous	R
HERV43	HERV43I	LTR43		Nonautonomous (Env only)	H
HERV45	HERV45I	LTR45, LTR45, LTR45B-C		? PR, RT, RH, IN	?
HERV49	HERV49I	LTR49		Nonautonomous	?
HERV70	HERV70_I	LTR70		Nonautonomous	?
PRIMA4	PRIMA4_I	PRIMA4_LTR		Gag, PR, RT, RH, IN, Env	R/W
MER4I	MER4I	MER4		Nonautonomous	R/W
MER4BI	MER4BI	MER4B		Nonautonomous (+remnants of class III-derived gag and pol)	R/F
MER41	MER41I	MER41A-B, D-G		Nonautonomous	R/W
PRIMA41	PRIMA41	MER41C		Gag, PR, RT, RH, IN, Env	R/W?
MER57	MER57I	MER57A, MER57B	HERV-HS49C23	Nonautonomous	S

Notes: Internal sequences are not reconstructed yet for LTR24, LTR24B, LTR24C, LTR26, LTR26E, LTR27, LTR27B, LTR28, LTR29, LTR31, LTR34, LTR36, LTR37A-B, LTR44, LTR48, LTR48B, LTR51, LTR54, LTR54B, LTR56, LTR58, LTR59, LTR60, LTR61, LTR65, LTR68, LTR72, LTR72B, LTR73, LTR76, LTR77, MER4C, MER4D, MER4E, MER72, MER72B, MER90.

[a]Names given to partially sequenced retroviruses, and secondary names given to individual sequences of retroviruses whose consensus sequences were deposited in repbase update much earlier.

[b]tRNAa are designated by the corresponding one-letter amino acid code.

Tab. 5 HERVK-like endogenous retroviruses (class II).

Family	*Internal portion*	*LTRs*	*Alternative names*
HERVK10	HERVK	LTR5	HERV-K(HML-2)
HERVK3	HERVK3I	LTR3, LTR3B	HERV-K(HML-6)
HERVK11	HERVK11I	MER11A, MER11B, MER11C	HML-8
HERVK11D	HERVK11DI	MER11D	HML-7
HERVK9	HERVK9I	MER9	HML-3
HERVK22	HERVK22I	LTR22, LTR22A, LTR22B	HML-5
HERVK13	HERVK13I	LTR13	HML-4
HERVK14	HERVK14I	LTR14A, LTR14B	HML-1
HERVK(C4)	HERVKC4	LTR14	HML-10
HERVK14C	HERVK14CI	LTR14C	
HERVK(HML9)	HERVK(HML9)I	HERVK(HML9)_LTR	HML-9

Tab. 6 HERV-L-like endogenous retroviruses (class III).

Family	*Repbase update*		*ENV*
	Internal portion	*LTRs*	
HERV-L	HERVL	MLT2A1, MLT2A2	
ERVL	ERVL	MLT2B1, MLT2B2	
HERVL68	HERVL68I	MER68A, MER68B	
HERVL74	HERVL74	MER74, MER74A-C	
HERV16	HERV16	LTR16, LTR16A, LTR16A1, LTR16C-D	
HERV18[a]	HERV18	LTR18A, LTR18B	+
HERV32	HERV32	LTR32	
HERVL40	HERVL_40	LTR40A, LTR40B, LTR40C	
HERV47	HERV47	LTR47A, LTR47B	
HERVL66	HERVL66I	LTR66	+
HERV52	HERV52I	LTR52	
HERV57	HERV57I	LTR57	
HERVL70	MER70I	MER70A, MER70B	
THE1B	THE1BR	THE1B	
MSTA	MSTAR	MSTA	
MLT1	MLT1R	MLT1	
MLT1C	MLT1CR	MLT1C	
Unknown	Unknown	LTR33, LTR33A, LTR41, LTR42, LTR50, LTR53, LTR67, LTR69, LTR75, MER54, MER54A, MER54B, MER73, MER74B, MER88, MaLRs, MLT2-like, MLT1-like	

[a]This family is also known as HERV-S.

it is several kb long, and encodes a set of proteins necessary for the retroviral life cycle. These proteins are encoded by the *gag* (group specific antigen), *pro* (protease), *pol* (polymerase), and *env* (envelope) genes (Fig. 8). Also, some Class II and Class III retroviruses encode deoxyuridine triphosphatase (dUTPase).

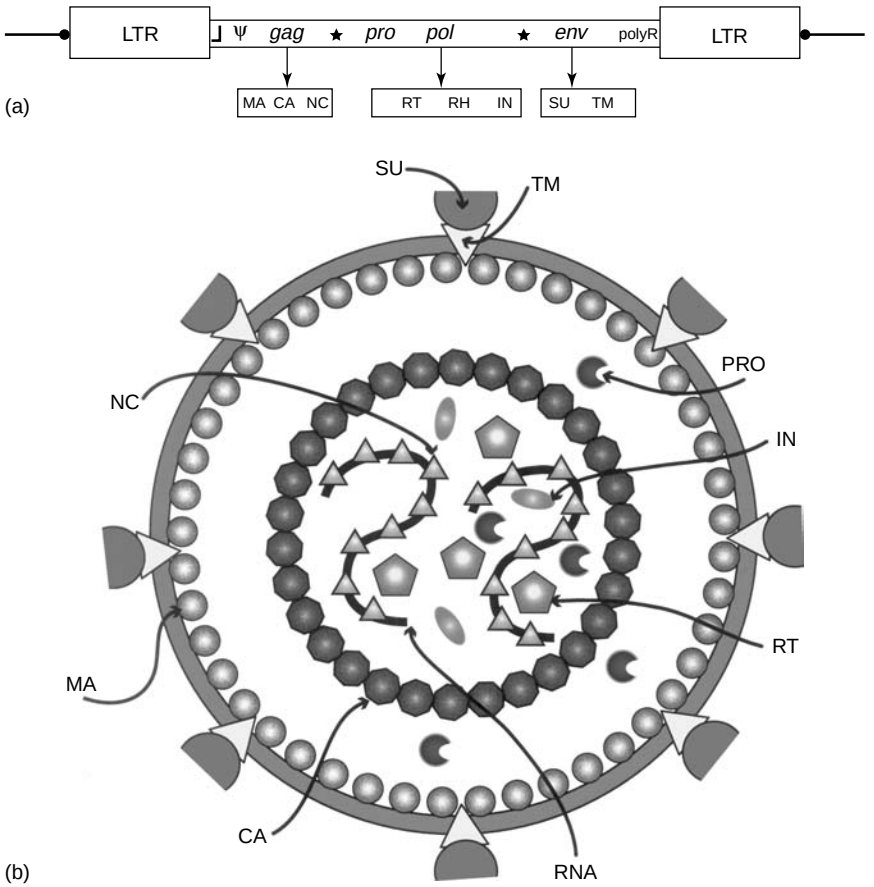

Fig. 8 The general structure of LTR retrotransposons. (a) Structural features of a provirus. Black circles indicate target site duplications; LTR, long terminal repeat; ⌟, primer binding site; Ψ, packaging signal; *gag*, capsid protein gene; *pro*, protease gene; *pol*, polymerase gene; *env*, envelope protein gene; ★, deoxyuridine triphosphatase (dUTPase), which lies between *gag* and *pro* in some Class III retroviruses, and between *pol* and *env* in Class II retroviruses. PPT stands for polypurine tract. The *pro* gene encodes the protease (PR), involved in the processing of the *gag*-, *pol*-, and *env*-encoded proteins. The *pol* gene encodes the reverse transcriptase (RT), ribonuclease H (RH), and integrase (IN). The *env* gene encodes the surface (SU) and transmembrane (TM) proteins. Typically, the *gag* gene encodes the matrix (MA), capsid (CA), and nuclear capsid (NC) structural proteins. (b) Schematic structure of a retroviral particle. The viral envelope (green) is made of a host-cell-produced lipid bilayer. It harbors multiple complexes formed by the *env*-encoded TM and SU components linked together by disulfide bonds. The retroviral capsid formed by the CA surrounds two RNA copies of the retrovirus, tRNAs, RT, IN, and PR. (See color plate p. xxiii).

Historically, LTR retrotransposons were viewed as endogenous retrovirus-like elements that did not encode the env protein. Currently, they are considered as nonautonomous retrovirus-derived elements, and the originally introduced distinction between retroviruses and LTR retrotransposons is becoming obsolete.

Gag is a multifunctional structural polyprotein separated into matrix, capsid, and nucleocapsid structural proteins by the *pro*-encoded protease (PR), which is related to cellular aspartic proteases.

The *pol* gene encodes a polyprotein precursor processed by the protease to reverse transcriptase and integrase. RT contains a ribonuclease H domain (RH), which digests RNA strands in RNA–DNA duplexes. RT encoded by LTR retrotransposons is capable of both RNA- and DNA-dependent DNA syntheses. It is the most conserved endoretroviral protein. The retroviral IN may be distantly related to $DD_{35}E$ transposases (see Sect. 3.3.2).

The *env*-encoded protein precursor is cleaved by the protease into the envelope, surface and transmembrane glycoproteins. It is thought that viral particles bind host cells through interactions between the surface glycoproteins and specific cell surface receptors (Fig. 8). Such interactions trigger fusion of the cell and virus membranes, followed by intrusion of the virus into the cell cytoplasm.

Transcription Transcription of the provirus starts at the pol II promoter present in the long terminal repeat (see LTR; Fig. 8a; Fig. 9). LTR is composed of three regions, called U3, R, and U5. The U3 region contains the pol II promoter. Additional signals affecting transcription, including enhancers, were also detected in the U3 region. The start of the LTR transcription corresponds to the U3-R boundary. Conversely, the R-U5 boundary is defined by the transcription stop-site. The transcription termination is mediated by the AATAAA or ATTAAA polyadenylation signals usually present in the U3 region. Therefore, the transcription is terminated in the second LTR only. In some retroviruses, including HIV, the polyadenylation signal is placed downstream of the U3-R boundary. However, the first LTR is still bypassed during transcription, presumably due to suppression of the transcription termination by a small distance between the transcription start-site and the polyadenylation signal.

Reverse transcription Reverse transcription of retroviral RNAs occurs in the cytoplasm of the infected host cell. It begins either when the viral particles enter the cytoplasm or when nuclear transcription of the already integrated provirus produces its functional RNA. The cytoplasmic localization of the retroviral reverse transcription is in sharp contrast to nucleus-based reverse transcription of non-LTR retrotransposons (Fig. 4). The retroviral reverse transcription is believed to follow a multistep process outlined schematically in Fig. 9 and described below.

1. The reverse transcription starts with the minus-strand DNA synthesis primed by the 3′ end of a tRNA molecule annealed to the primer binding site (PBS). Usually, the PBS is complementary to the ~18-bp 3′ end of the virus-specific tRNA that starts just a few nucleotides downstream of the U5 region.
2. Minus-strand DNA synthesis proceeds to the 5′ end of the RNA, producing a 100 to 150 bp DNA molecule called *minus-strand strong-stop DNA*, which is released from the original RNA–DNA duplex by the RNaseH mediated digestion of the complementary RNA.
3. Following the digestion, the minus-strand strong-stop DNA performs the first strand transfer, annealing to the R-region at the 3′ end of the plus-strand RNA.
4. After the strand transfer, minus-strand DNA synthesis continues, and it is

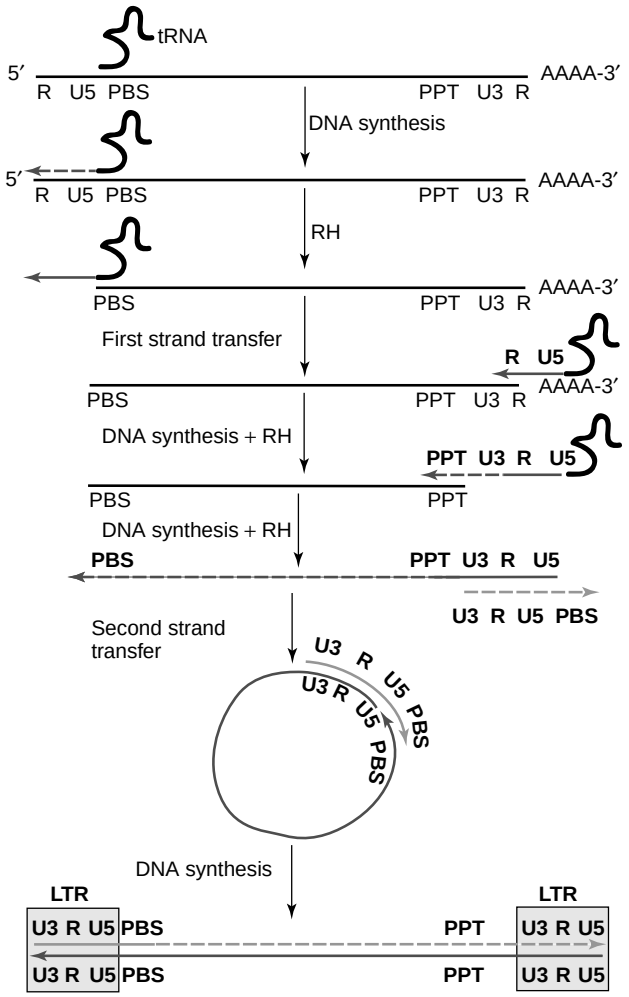

Fig. 9 Reverse transcription of the retroviral RNA. Dotted lines mark newly synthesized DNAs; green, plus-strand DNA; red, minus-strand DNA; RH, RNaseH digestion of RNAs. (See color plate p. xxiv).

accompanied by digestion of the RNA template, except for a short RNaseH-resistant polypurine tract (PPT).

5. The 3′ end of the PPT primes plus-strand DNA synthesis, which continues until the PBS-like portion of the tRNA molecule is reverse transcribed. As a result, the so-called *plus-strand strong-stop DNA* is generated.
6. RNaseH digests the PBS-annealed tRNA molecule.
7. The second strand transfer is performed by annealing the complementary PBS segments present at the termini of the plus-strand and minus-strand DNAs.
8. The plus and minus DNA strands serve as templates for each other.

The replication is completed when RT synthesizes the entire plus and minus DNA strands. The final product is a linear duplex DNA.

Life cycle of LTR retrotransposons The life cycle of an LTR retrotransposon starts with the transcription of its provirus copy integrated in the genome. The resulting mRNA is transported from the nucleus to the cytoplasm, where it is translated. Usually, reverse transcription of the proviral mRNA occurs in the viral particles in the cytoplasm. The resulting cDNA is then transported back to the nucleus, together with the retroviral integrase where the integrase inserts the cDNA into the host genome, producing a new provirus flanked by identical LTRs and short TSDs.

Subsequent homologous recombination between the LTRs may result in elimination of the provirus with the exception of a solo LTR (Fig. 10). The number of endogenous retroviruses that retain their internal portions flanked by both LTRs is about ten times smaller than the number of solo LTRs. Therefore, recombination between the flanking LTRs is a significant factor in elimination of LTR retrotransposons. Gene conversion between LTRs can also occur at a very low rate. Therefore, the average divergence between proviral LTRs does not adequately reflect the age of the corresponding family of LTR retrotransposons.

RNA recombination of LTR retrotransposons LTR retrotransposons copackage two copies of their RNAs, which can recombine with each other, in the same viral particle. Two RNAs representing essentially distinct LTR retrotransposons can also be copackaged together if they share similar ψ packaging signals. Retroviral recombination is usually mediated by the RNA template switch that occurs during minus-strand DNA synthesis, when RT is paused by a break in the reverse transcribed RNA. The RT pausing can also be caused by a specific secondary structure of the RNA template and by internal repeats.

Characteristics of human LTR retrotransposons The most abundant and ancient group of LTR retrotransposons preserved in the human genome are class III retrotransposons (Table 6). For example, the HERVL16 and HERVL33 families are more than 200 million years old. Despite their high abundance, they have remained undetected for a long time due to their extensive modification by mutations. Most

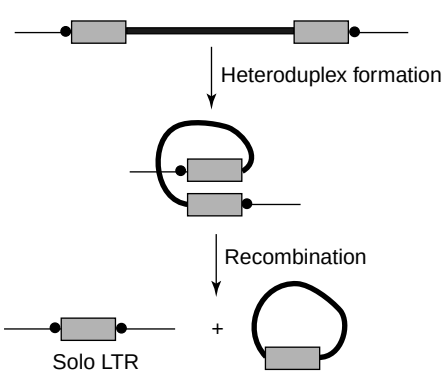

Fig. 10 Formation of a solo LTR through the illegitimate recombination between proviral LTRs.

Class III families of retroelements reconstructed so far are nonautonomous, including HERV-L, which was discovered as the first representative of class III elements. For example, the HERV-L family does not contain *env*, HERVL66 misses *gag* and protease, and HERVL68 does not encode any proteins at all. The most abundant Class III families are MaLRs (Mammalian apparent LTR retrotransposons) and THE1 elements, which probably did not encode any functional proteins at the time of their activity. MaLRs (THE1) were recently classified as members of Class III based on the observation that some of their noncoding sequences were derived from HERV-L-like elements. HERV18 was the first Class III element found to contain *env*. Many different families of LTR retrotransposons are still represented by their LTRs only. As shown in Tables 4 and 6, these families were reliably assigned to their proper classes based on similarities to LTRs from classified retroelements.

Class I LTR retrotransposons (Table 4) include dozens of nonautonomous families that are members of the same group, also called the MER4 or MER4I group after the MER4 element that was discovered over a decade ago and later identified as an LTR flanking the internal portion of a retrovirus-like element (MER4I). Other Class I families of LTR retrotransposons were discovered and classified on the basis of their sequence similarity to MER4I. All these families are characterized by 4-bp target site duplications. Analogously to MaLR and THE1 elements, the originally discovered MER4I-like retrotransposons, do not code for any proteins. Furthermore, the human genome also harbors autonomous Class I families whose members were ancestral to nonautonomous MER4I-like elements. For example, the internal portion of the PRIMA4 autonomous retrovirus is similar to the entire MER4I sequence; even their LTRs are similar. Analogously, the autonomous PRIMA41 retrovirus is an ancestor of the MER41 family of MER4I-like nonautonomous elements. Class I elements were involved in vigorous multiple recombinations between RNA copies of different retroviruses. As a result, many Class I families are chimeras composed of parts that came from retroviruses dissimilar to each other, for example, *Harlequin*.

Class III and Class I elements were probably engaged in extensive complementation, meaning that various families mutually assisted each other by sharing necessary enzymatic and packaging activities. For instance, the RNA of nonautonomous proviruses can be copackaged into functional viral cores produced by autonomous LTR retroelements and exogenous retroviruses. These viral cores can be released from the host cell and can infect other cells. Given the low fidelity of reverse transcription ($\sim 10^{-5}$ per nucleotide per cDNA molecule), the original nonautonomous retrovirus can be dramatically transformed after multiple rounds of reverse transcription and intercellular migrations. As a result of such transformation, the similarity between the original provirus and the transformed retrotransposon can wane rather fast.

Human Class II elements are represented by 11 families (Table 5) with ages ranging from $\sim$35 million to less than 5 million years. The oldest Class II family is HERVK22. On the other hand, some HERVK10 elements are polymorphic in the human population, and contain nearly intact ORFs coding for the standard retroviral proteins. In addition to the standard proteins, some HERVK elements encode dUTPase and K-Rev. The latter is a 105-aa

protein, also called cORF, which is a functional homolog of the HIV Rev protein. K-Rev mediates nuclear export of the unspliced HERVK RNA by binding to both the host Crm1 nuclear export factor and to a cis-acting retroviral RNA target.

Mechanisms of retroviral evolution are poorly understood since very few intermediate forms of exogenous retroviruses are available for analysis. Therefore, LTR retroelements deposited in mammalian genomes provide us with a unique opportunity to study retrovirus-like elements, which became extinct millions of years ago.

3.3 DNA Transposons

All DNA transposons detected in the human genome represent the so-called *cut-and-paste* category since they use excision from host DNA (cut) followed by reinsertion (paste) as two basic steps in their life cycle. Both steps are mediated by the transposon-encoded enzyme called *transposase*, which recognizes terminal inverted repeats (TIRs; Fig. 11). The transposon insertion is accompanied by target site duplication.

The cut-and-paste process is thought to be associated with the active replication process. As shown in Fig. 12, the excision is more likely in already replicated DNA segments, whereas the insertion occurs preferentially in nonreplicated segments. On the basis of this model, the predicted multiplication rate of cut-and-paste DNA transposons is 1.5 per replication cycle; but the real rate is expected to vary between 1 and 1.5.

To date, no active DNA transposons have been found in the human genome. All

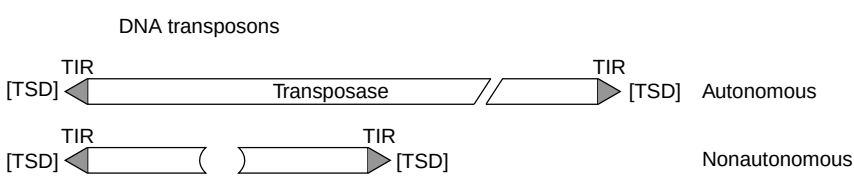

Fig. 11 Schematic structure of autonomous and nonautonomous DNA transposons.

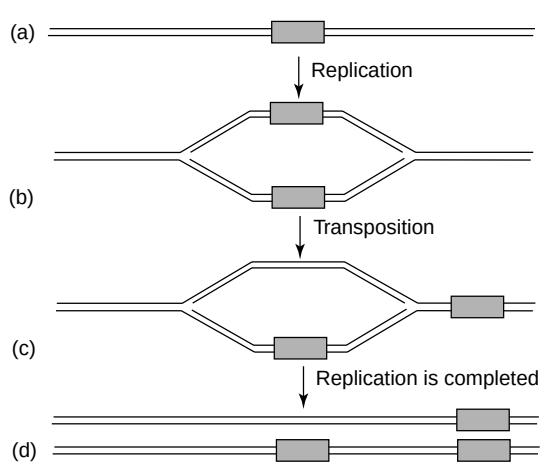

Fig. 12 Relationship between DNA transposition and host replication.

basic properties of human DNA transposons were deduced from sequence analysis of human repetitive DNA, followed by reconstruction of DNA and protein sequences encoded by once-active elements. On the basis of their structural commonalties, eukaryotic DNA transposons can be divided into eight different superfamilies, of which six are represented in the human genome (Table 7). DNA transposons that belong to the same superfamily, even those from different vertebrates, plants, and insects, are characterized by common structural hallmarks such as similar TIRs, transposases, and the same characteristic sizes of TSDs. Additional superfamily characteristics include conserved termini of TIRs, usually 2 to 3 bp long. Like other classes of TEs, DNA transposons are represented by autonomous and nonautonomous elements.

The human genome harbors families derived from over 29 autonomous transposons. Of these, over 25 belong to the hAT and *mariner/Tc1* superfamilies (Table 8). However, only 13 of them have been reconstructed and characterized in more detail. Some of the most abundant families of nonautonomous hAT elements, such as MER5A and Cheshire_A, are represented by more than 10^4 copies each. Overall, the transposon-derived interspersed repeats account for ~5% of the human genomic DNA.

3.3.1 hAT-like Transposons

The superfamily of hAT-like elements was named after the initials of *hobo, Activator* and *Tam* transposons identified in the fruit fly and maize. They are flanked by ~15-bp terminal inverted repeats that include characteristic 5′-CA and TG-3′ termini, and generate 8-bp target site duplications upon integration in the genome. Autonomous hAT-like transposons are ~3000-bp long and encode ~500-aa transposase similar to transposases encoded by all known hAT-like transposons detected in different eukaryotic species. Analogously, the ORFs in the human hAT-like transposons are also expected to encode the transposases.

The most abundant hAT elements are the nonautonomous deletion products of the corresponding autonomous transposons. They retain TIRs, but the

Tab. 7 Superfamilies of cut-and-paste DNA transposons represented in humans.

Superfamily	*Related bacterial transposases*	*Termini*	*TSDs*
Mariner/Tc1	IS630	CA-TG	2 (TA)
hAT	–	CA-TG	8
piggyBac	–	CCC-GGG	4 (TTAA)
MuDr[a]	IS256	GGG-CCC	9–10
Harbinger[b]	IS5	GGG-CCC	3
P[c]	–	CA-TG	7–8

[a]Only nonautonomous *Ricksha* has *MuDr* hallmarks: 9-bp TSDs, 70-bp TIR, 5′-GGG and CCC′- 3′ termini.
[b]*Harbinger*- and
[c]*P* – like transposases are present as single-copy genes conserved in mammals including humans.

Tab. 8 Families of DNA transposons identified in the human genome.

Superfamily	***Autonomous elements***	***Nonautonomous elements***	%
hAT	Blackjack	MER81, MER94	0.06
	Charlie1	Charlie1A, Charlie1B	0.21
	Charlie2	Charlie2A, Charlie2B	0.08
	Charlie3	MER1A, MER1B	0.11
	Charlie4	MER80	0.05
	Charlie5	MER3, MER33	0.40
	Charlie6		$0.04 \cdot 10^{-1}$
	Charlie7		0.04
	Charlie8	Charlie8A, MER102	0.08
	Charlie9	MER112	0.03
	Charlie10	MER5A, MER5B, MER5C	0.42
	Cheshire	Cheshire_A, Cheshire_B (MER58, MER58A-C)	0.31
	Zaphod	Zaphod2 (Fordprefect)	0.03
	Unknown	MER20, MER20B, MER30, MER30B, MER45, MER45A-C, MER45R, MER63A-D, MER69A-C, MER91A-C, MER96, MER96B, MER97A-C, MER99, MER106, MER106B, MER107, MER117, MER119, ORSL	0.54
			2.4
Mariner/Tc1	Golem (Tigger3)	Golem_A, Golem_B, Golem_C (Tigger3)	0.14
	HSMAR1	MADE1	0.04
	HSMAR2		0.04
	HSTC2	MER104, MER104A-C, Kanga1, Kanga2	0.06
	Tigger1		0.32
	Tigger2	MER28 (Tigger2a)	0.15
	Tigger5	TiggerA-B	0.01
	Tigger6	Tigger6A-B	0.03
	Tigger7	MER44A-D	0.09
	Zombi (Tigger4)	Zombi_A-C	0.03
	Unknown	MARNA, Tigger8, Tigger9, MER2, MER2B, MER6, MER6A-C, MER8, MER53, MER82	1.95
			2.8
piggyBac	Looper	MER75, MER85	0.02
MuDR	Unknown	Ricksha, Ricksha_0	0.03
	Merlin1_HS		$0.01 \cdot 10^{-3}$
Unclassified	Unknown	MER81, MER94, MER121, MER122, MER103, MER105, MER113, MER116	0.13

transposase is either partially or completely deleted. Typically, the number of nonautonomous copies representing a particular hAT family is at least an order of magnitude higher than the copy number of the corresponding autonomous elements. Sometimes, nonautonomous elements differ from their autonomous companions by possessing additional internal fragments (for example, MER5A-B and *Charlie10*). It is possible that these additional fragments represent genomic DNA

captured by the nonautonomous transposons (analogously to MuDR transposons in the *Arabidopsis thaliana* genome).

The most abundant human hAT families are *Charlie10, Charlie5, Cheshire* and their nonautonomous companions represented by ~60 000, ~50 000 and ~40 000 copies respectively (see Table 8). As indicated above, there are no active DNA transposons in the human genome. The youngest hAT elements are *Charlie3* repeats that are ~8% diverged from their family consensus sequence. Given the standard neutral mutation rate in the human genome, this corresponds to ~35 million years. Most of the human hAT-like families are much older and they might have been active in a common ancestor of primates and rodents.

3.3.2 Mariner/Tc1-like Transposons

The Mariner/Tc1 superfamily reflects the names of the two founding DNA transposons *mariner* and *Tc1*, first discovered in the *Drosophila melanogaster* and *Caenorhabditis elegans* genomes. Members of this superfamily are characterized by 2-bp target site duplications (usually TA), by ~28-bp TIRs, and by the 5′-CA...TG-3′ termini. However, some elements may have different termini; for example, HSMAR1 and HSMAR2 are characterized by the 5′-TT...AA-3′ and 5′-CG...CG-3′, respectively. Mariner/Tc1 transposases are distantly related to a broad set of bacterial transposases similar to that encoded by IS630 insertion element, and are characterized by the conserved $DD_{35}E$ motif.

3.3.3 piggyBac-like Transposons

The piggyBac superfamily includes DNA transposons with TTAA TSDs, the ~15-bp terminal inverted repeats, the 5′-CCC and GGG-3′ termini, and the unique piggyBac transposase. The piggyBac transposon was first discovered as an active mobile element in the cabbage looper, *Trichoplusia ni*. The human transposon called *Looper* was the first piggyBac-like element identified outside cabbage looper. It was reconstructed from copies transposed ~100 million years ago. Since then, piggyBac-like elements were identified in nonmammalian vertebrates, insects, and fishes. In addition to Looper, the human genome harbors MER75 and MER85 transposons that were active 10 to 50 million years ago. Only ~2000 copies of the piggyBac-like elements are present in the human genome.

3.3.4 MuDR-like Transposons

MuDR/Mu transposons, also known as *Mutators*, were first discovered in maize. They are characterized by 9 to 10-bp TSDs and by ~70 to 400-bp TIRs, typically ending with 5′-GGG and CCC-3′ conserved termini. However, some families of MuDR transposons in the *A. thaliana* genome can transpose without any discernible TIRs. Different autonomous MuDR transposons encode similar MuDR-like transposases, which share common conserved motifs with bacterial IS256-like transposases. Currently, only two families of MuDR-like transposons are known to be harbored by the human genome. One of them is represented by the nonautonomous transposons Ricksha and Ricksha_0, and characterized by 9 to 10-bp TSDs, ~70-bp TIRs, and 5′-GGG and CCC-3′ termini. Ricksha and Ricksha_0 were active 40 to 100 million years ago and are present in only ~1000 copies in the human genome. Ricksha is a composite transposon that differs from Ricksha_0 by having a ~700-bp insertion of the HERVL endogenous retrovirus including its long terminal repeat (MLT2B). Presumably, the HERVL

provirus was inserted into an active Ricksha_0 copy, which amplified into the Ricksha subfamily. A second family of human MuDR-like transposons was reported recently as the Merlin1_HS autonomous transposon, which encodes a transposase that is the most similar to the IS1016 bacterial transposase. Since the IS1016 and IS256 transposases are distantly related, Merlin1_HS can be considered to be a member of the MuDR superfamily, quite different from standard MuDR transposons. Merlin1_HS is characterized by the 8-bp TSDs and 21-bp TIRs with 5′-GGA and TCC-3′ termini. The human genome harbors fewer than 100 copies of Merlin1_HS. They are ~20% divergent from the consensus sequence, which indicates their activity some 100 million years ago.

4
Repetitive Elements and Human Diseases

The potential pathogenic impact of repetitive elements can manifest itself through a broad spectrum of pathways. The first category of disorders is linked to the expression of active TEs, particularly to the production of proteins. Such proteins can interfere with cellular processes and, most importantly, they can provoke host immune reactions against cells containing TE antigens. The second group of pathogenic effects linked to repetitive DNA includes mutations and genetic rearrangements such as satellite expansion, repeat insertions, deletions, illegitimate recombination, or nucleolytic activity of proteins encoded by TEs. These mutations can occur both in germline or somatic cells.

The pathogenic potential of TE-derived protein products is far from being understood. Relevant research was done mostly on human endogenous retroviruses (HERVs; Sect. 3.1.2). HERVs have been proposed as causative agents of autoimmune diseases such as rheumatoid arthritis, systemic lupus erytrematosus, insulin-dependent diabetes mellitus, and multiple sclerosis. Other potentially HERV-linked disorders include schizophrenia and several types of cancer. In many such cases, retroviral mRNAs or even entire particles were detected in malignant tissues. Retroviral antigens such as the *gag* gene product may indeed trigger immune reaction against affected tissues leading to autoimmune abnormalities. However, it is still unclear whether expression of retroviral mRNAs and proteins causes or merely correlates with the pathological processes. Increased expression of TEs is typical of many cellular alterations including cell transformations. This indicates that rather than inducing the pathogenesis, the expression of TEs may simply reflect relaxed transcription control in malignant cells.

Repeat-induced DNA rearrangements represent a wide range of mutation defects related to the amplification of repetitive elements and genetic recombination. Genetic diseases can be caused by both simple and interspersed repeats. The unstable expansions of simple sequence repeats, namely, trinucleotides, are involved in a number of genetic diseases. The fragile X syndromes (known as FRAXA and FRAXE) involve CGG and GGC triplet expansions. Myotonic dystrophy can be caused by the CTG/CAG repeat expansion. Other examples of diseases associated with moderate expansion of CAG repeats include neural diseases such as Kennedy's disease, Huntington's disease, spinocerebellar ataxia 1 (SCA1) and dentatorubropallidoluysian atrophy (DRPLA). The disease-related triplet expansions are much more frequent than

contractions, and occur primarily in GC-rich genes.

Centromeric and telomeric repeats can also trigger chromosomal instabilities in the human genome. Recombinations between both centromeric and telomeric satellites were reported in cases of chromosome rearrangements, chromosome fragility, or jumping chromosomal translocations.

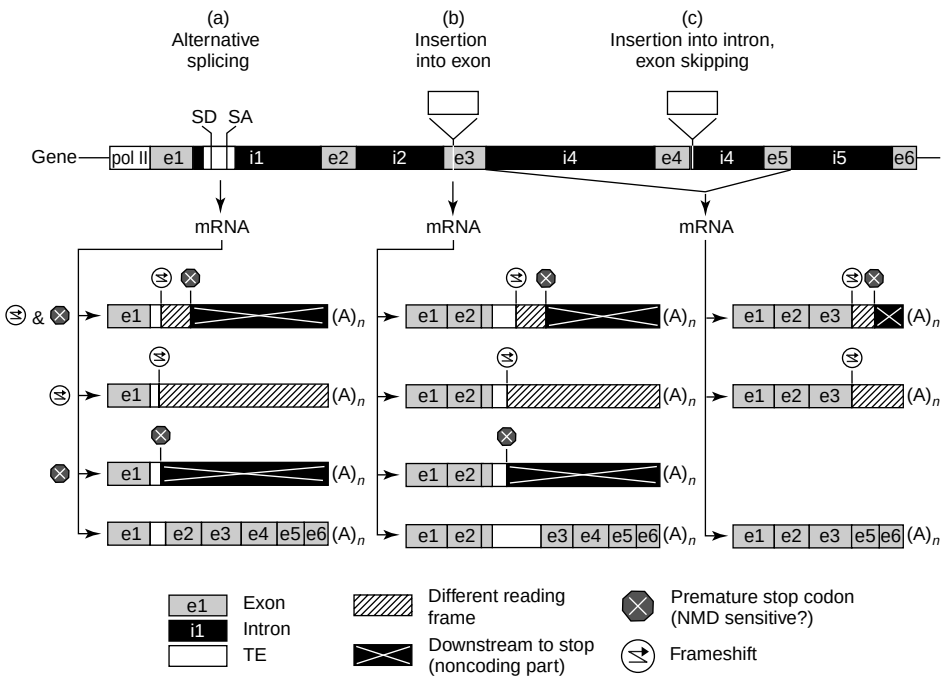

Fig. 13 Transposable elements and human disease. Three different mechanisms of TE-mediated disruptions affecting the gene functionality: (a) intronic repeats that contain splice donor (SD) and splice acceptor (SA) sites, can provide a new exon. (b) Insertion of transposable element into exon. (c) Insertion of transposable element into intron, disrupting proper splicing can lead to exon skipping. All three disruptions lead to similar changes on the mRNA level. Insertion or deletion in coding sequences may change the reading frame, if the indel length is not a multiple of three. Long frameshifts are likely to induce premature stop codon(s) into mRNA (case 1). mRNAs with premature stop codon (i.e. located more than 50–55 bp upstream of the last exon–exon junction) are likely to be destroyed by the nonsense-mediated RNA decay (NMD), and the corresponding protein is not produced. In some cases, particularly for short frameshifts, the random stop codon is found close enough to the last exon–exon boundary and the protein can be synthesized, albeit with an aberrant C-terminal domain (case 2). Note that mRNAs completely lacking stop codons are also destroyed by the cellular RNA control, although by a different pathway than the nonsense-mediated RNA decay. Insertions containing stop codon without change in the reading frame (case 3). Again, if the stop is more than 50 to 55 bp 5′ from the last exon–exon junction, the RNA is likely to be NMD-sensitive and the protein is not produced. Finally, if neither frameshifts nor new premature stop codons are induced, insertions lead to creation of a protein with a new repeat derived domain. Exon-skippings induce deletions of some domains. These new proteins may or may not be functional.

Interspersed repeats can affect alternative splicing and cause inclusion of intronic repeats into cellular mRNAs. This can alter or destroy the mRNA or protein functionality and be manifested as a disease phenotype. Notably, Alu repeats contain cryptic splice sites in the antisense orientation that, after acquiring specific mutations, can give rise to new Alu-derived exons (Fig. 13a). The majority of such Alu cassettes are alternatively spliced and produce premature termination (stop) codons. The affected mRNAs are probably destroyed by cellular mRNA quality control machinery, particularly by the nonsense-mediated RNA decay (NMD) pathway, which degrades mRNAs with premature stop codons at the time of protein translation. However, if the new exon becomes constitutively spliced in, the NMD control prevents efficient translation of the affected mRNA, and thus causes a deleterious phenotype at the protein level. So far, there are two known cases of disease-related Alu splicing into cellular mRNAs. One, reported in a case of ornithine-delta aminotransferase deficiency, was caused by alternative Alu splicing, creating a premature stop codon. In a second instance, a splice-mediated insertion of an Alu sequence in the precursor of alpha 3 type IV collagen mRNA caused autosomal recessive Alport syndrome in the affected patient. The affected proteins were not expressed in either case, despite the fact that the insertion preserved the open reading frame in the second case.

Insertion into genomic DNA is an integral part of the amplification of interspersed repeats. An integration into coding sequences is likely to cause an insertional inactivation of the targeted gene, which subsequently manifests itself as a genetic abnormality. To date, there are 41 published repeat insertions associated with human diseases (Table 9). All the insertions are either L1 elements (14 cases), L1-dependent Alu (23 cases), or SVA elements (4 cases). The Alu and L1 insertions come nearly exclusively from the two active subfamilies AluY and L1-Ta respectively (Table 9). No recent insertions of endogenous retroviruses or DNA transposons have been reported.

From Table 9 it is clear that some genes are overrepresented among insertion targets, particularly coagulation factors VIII and XI and the dystrophin gene. Furthermore, 42% (19 out of 41) of all documented insertions are on chromosome X, which represents only 5% of the human genome. The bias is particularly strong for L1 insertions 11/14 (79%), and less so for L1-dependent nonautonomous Alu and SVA elements, which represent 26% (6/23) and 25% (1/4) of all documented insertions respectively. Chromosome X is present in only one functional copy in both males and females, due to female imprinting (inactivation) of one X copy. Therefore, the X-linked genes are more vulnerable to mutations. The excess of detected insertions on the X chromosome may reflect the detection bias for dominant phenotypes rather than any preferential insertion patterns on this chromosome.

The majority of insertions listed in Table 9 occurred in exons and the remaining intronic insertions appear to have disrupted splicing or other essential signals located within introns. The deleterious effect of intraexonic insertions is mainly due to the incorporation of premature stop codons into the reading frame and mRNA degradation by NMD (Fig. 13b). Other defects can be caused by intron inclusion or exon skipping (Fig. 13a, c), leading to either truncation of the protein or induction of premature stop codons and NMD sensitivity.

Tab. 9 Insertions of interspersed repeats and human diseases.

Repeat	*Locus*	*Chrom*	*Gene*	*Sym bol*	*Phenotype*
Alu/Ya5	–	–	Human homolog of moloney murine virus leukemia (MMLV) retrovirus insertion locus	MLVI-2	Associated with leukemia
1x Alu/Yb8 1x Alu/Ya5	NM_000141	10q26.13	Fibroblast growth factor receptor 2 isoform 1	FGFR2	Apert syndrome
Alu/Y	NM_000062	11q12.1	Serine/cysteine proteinase inhibitor	C1NH	C1 inhibitor deficiency
Alu/Ya5	NM_000190	11q23.3	Porphobilinogen deaminase (hydroxymethylbilane synthase)	PBGD/HMBS	Acute intermittent porphyria
Alu/Y	NM_000059	13q13.1	Breast cancer 2	BRCA2	Breast cancer
Alu/Ya5	NM_000267	17q11.2	Neurofibromin 1	NF1	Neurofibromatosis
Alu/Ya5	NM_000789	17q23.3	Angiotensin I converting enzyme isoform 1	ACE	Protection against myocardial infarction and age-related macular geneneration?
Alu/n.a.	NM_000092	2q36.3	Alpha 4 type IV collagen precursor	COL4A3	Alport syndrome
Alu/Ya4	NM_000388	3q21.1	Calcium-sensing receptor	CASR	Familial hypocalciuric hypercalcemia
Alu/Yb8	NM_000055	3q26.1	Butyrylcholinesterase	BCHE	Acholinesterasemia
Alu/n.a.	NM_001119	4p16.3	Adducin 1, transcript variant 1	ADD1	Huntington disease
Alu/Yb1	NM_006264	4q21.3	Protein tyrosine phosphatase, nonreceptor type 13	PTPN13 (APO-1/CD95)	Autoimmune lymphoproliferative syndrome
Alu/Yb8	NM_000038	5q22.2	Adenomatosis polyposis coli	APC	Hereditary desmoid disease
Alu/Ya5	NM_000503	8q13.3	Eyes absent homolog 1 (Drosophila)	EYA1	Branchio-oto-renal syndrome
Alu/Ya5	NM_000167	Xp21.2	Glycerol kinase	GK	Glycerol kinase deficiency

Alu/Ya5	NM_000206	Xq13.1	Interleukin 2 receptor, gamma	IL2RG	X-linked severe combined immunodeficiency
Alu/Y	NM_000061	Xq22.1	Bruton agammaglobulinemia tyrosine kinase	BTK	X-linked agammaglobulinemia
Alu/Ya5	NM_000133	Xq27.1	Coagulation factor IX	F9	Hemophilia B
Alu/Ya3a1	NM_000133	Xq27.1	Coagulation factor IX	F9	Hemophilia B
Alu/n.a.	NM_000133	Xq27.1	Coagulation factor IX	F9	Hemophilia B
Alu/Yb8	NM_000132	Xq28	Coagulation factor VIII	F8	Hemophilia A
Alu/Yb8	AF074480	6p22.3	CMP-N-acetylneuraminic acid (CMP-Neu5Ac) hydroxylase	CMAH	Inactivated in humans, fixed
L1/Ta	NM_000518	11p15.4	b-Globin	HBB	Beta thalassemia
L1/Ta	NM_000038	5q22.2	Adenomatosis polyposis coli	APC	Colon cancer
L1/Ta	NM_006731	9q31.2	Fukutin	FCMD	Fukuyama-type congenital muscular dystrophy
L1/Ta	NM_006915	Xp11.3	Retinitis pigmentosa 2 gene	RP2	X-linked retinitis pigmentosa 2
2x L1/Ta	NM_000397	Xp11.4-Xp21.1	Cytochrome b-245, beta polypeptide	CYBB	Chronic granulomatous disease
4x L1/Ta	NM_000109	Xp21.1–2	Dystrophin	DMD	3x muscular dystrophy, 1x X-linked dilated cardiomyopathy
L1/Ta	NM_000133	Xq27.1	Coagulation factor IX	F9	Hemophilia B
2x L1/Ta 1x L1/pre-Ta	NM_000132	Xq28	Coagulation factor VIII	F8	Hemophilia A
SVA	NM_003126	1q23.1	Alpha spectrin	SPTA1	Hereditary elliptocytosis
SVA	NM_006731	9q31.2	Fukutin	FCMD	Fukuyama-type congenital muscular dystrophy
SVA	NM_000061	Xq22.1	Bruton agammaglobulinemia tyrosine kinase	BTK	X-linked agammaglobulinemia
SVA	NM_000527	19p13.2	Low density lipoprotein receptor	LDLR	Familial hypercholesterolemia

Source: The table is based on the updated online database of TE insertions into the human genome (http://www.med.upenn.edu/genetics/labs/kazazian/human.html).

Table 9 compiles both *de novo* insertions, such as in the case of L1-induced Fukuyama muscular dystrophy, and inherited insertions causing familial disorders, such as an Alu insertion into the adenomatosis polyposis coli (APC) gene, causing hereditary desmoid disease. The table also contains one somatic L1 insertion that inactivates APC, indicating that the L1 retrotransposition can take place in both germline and somatic tissues. Most phenotypes caused by insertional inactivations are negatively selected in the population and subsequently eliminated. Nevertheless, in cases of nonessential genes with mild defects, the insertions may eventually become fixed in the population. For example, an Alu insertion seems to have inactivated the CMP-*N*-acetylneuraminic acid hydroxylase (CMAH) in the human lineage after the evolutionary separation of human and chimp, since all primates except humans contain a functional copy of CMAH. In rare cases, insertions may even produce a positive phenotype. For instance, AluYa5 insertion into intron 16 of angiogenesin-converting enzyme (ACE) seems to provide protection against myocardial infarction and age-related macular degeneration, since genotypes without the insertion have a higher incidence of the aforementioned diseases in comparison to individuals having Alu insertion in both alleles.

Table 10 shows a list of the genetic defects attributed to recombination between interspersed repeats. Preliminary indications for linkage between Alu polymorphisms and genetic diseases, based solely on polymorphism studies, are not included. While for some genes only a single case of recombination between interspersed repeats is known, other loci including LDLR, C1NH, alpha-globin, APC, HEXB or MLL represent hotspots for recombinations. Many independent cases of such recombination have been reported. As in the case of insertions, there is a detection bias for recombinations in X-linked genes. Twelve out of 54 (22%) deletion/duplication loci are located on chromosome X.

In mammalian cells, two major recombination pathways exist: *homologous recombination* and nonhomologous end joining, also known as *nonhomologous recombination*. In the great majority of genes listed in Table 10, homologous recombination between TEs was detected, but nonhomologous recombination between one or more TEs was also found in many genes.

The majority of known rearrangements were reported for direct repeats located within or near one gene. While some rearrangements listed in Table 10 are inherited (i.e. passed through germline cells), others occur primarily during mitosis in somatic tissues. This includes many leukemia- or other cancer-related recombinations. Figure 14 shows a model of mitotic intrachromosomal recombination between direct repeats. Unequal sister chromatid exchange produces one duplicated and one deleted copy. Reciprocal intrachromatid exchange, on the other hand, results in one deletion and one circular episomal DNA, which is frequently lost. The mechanism is the same as for the creation of solo LTRs (see Fig. 10). Mispairing and crossing-over between repeats, leading to deletion, have been reported in 54 genes, compared to duplications in only five.

Internal indels can result in altered lengths of the protein, if the reading frame is preserved, or in premature stop codons within the reading frame. Typical outcomes of internal exon/gene deletions and duplications are shown in Fig. 14. RNAs with premature stop codons

Tab. 10 Recombination of interspersed repeats and human diseases.

Repeat	*Mutation*	*Locus*	*Chrom*	*Gene*	*Symbol*	*Phenotype*
Alu	Deletion	NM_000511	19q13.33	Alpha(1,2)fucosyl-transferase	FUT2	ABO-Bombay phenotype
Alu	Deletion	NM_000033	Xq28	ATP-binding cassette, subfamily D, member 1	ABCD1	Adrenoleukodystrophy
Alu	Deletion	NM_000061	Xq22.1	Bruton's tyrosine kinase	BTK	Agammaglobulinemia, X-linked
Alu	Deletion	NM_000273	Xp22.22	Albinism ocular type 1 gene	OA1	Albinism ocular type 1
L1	Deletion	NM_001847	Xq23	Type IV alpha 6 collagen	COL4A6	Alport syndrome and diffuse leiomyomatosis
Alu [2]	Deletion	NM_000021	14q24.2	Presenilin-1	PSEN1	Alzheimer's disease, early onset
Alu	Deletion	NM_000062	11q12.1	Complement component 1 inhibitor precursor	C1NH/SERPING1	Angioedema, hereditary
Alu	Deletion	NM_003190	6p21.32	Tapasin	TAPBP	Bare lymphocyte syndrome, type I
Alu/L1 [3]	Translocation	t(9;11)(p24;q23)	–	–		Bipolar affective disorder (manic depression)
Alu	Deletion	NM_007295	17q21.31	Breast cancer 1, early onset	BRCA1	Breast and/or ovarian cancer
Alu [4]	Deletion	NM_000059	13q13.1	Breast cancer 2	BRCA2	Breast cancer
Alu	Deletion	NM_005228	7p11.2	Epidermal growth factor receptor	EGFR	Cancer, glioblastomas
Alu	Deletion	NM_000321	13q14.2	Retinoblastoma	RB1	Cancer predisposition

(*continued overleaf*)

Tab. 10 (*continued*)

Repeat	*Mutation*	*Locus*	*Chrom*	*Gene*	*Symbol*	*Phenotype*
Alu	Deletion	NM_000038	5q22.2	Adenomatous polyposis coli	APC	Colorectal cancer, adenomatous polyposis
Alu	Deletion	NM_000251	2p21	Mismatch repair protein	MSH2	Colorectal cancer, nonpolyposis
Alu	Deletion	NM_000249	3p22.3	MutL homolog 1	MLH1	Colorectal cancer, nonpolyposis
Alu	Translocation	t(1;22)(q23;q11)	–	–		
Alu	Deletion	NM_000208	19p13.2	Insulin-receptor	INSR	Diabetes, insulin-resistant
Alu	Duplication	NM_000302	1p36.22	Lysyl hydroxylase 1/procollagen-lysine, 2-oxoglutarate 5-dioxygenase	PLOD	Ehlers Danlos syndrome
Alu	Translocation	t(11;22)	–	–		Ewing's sarcoma
Alu	Deletion	NM_005243	22q12.2	Ewing sarcoma breakpoint region 1	EWSR1	Protection against Ewing's sarcoma
Alu	Deletion	NM_000169	Xq22.1	Alpha-galactosidase A	GLA	Fabry disease
Alu	Deletion	NM_000135	16q24.3	Fanconi anemia, complementation group A	FANCA	Fanconi anemia
Alu	Deletion	NM_000419	17q21.31	Integrin alpha 2b precursor, platelet glycoprotein Iib, CD41B	ITGA2B	Glanzmann's thrombasthenia
L1	Deletion	NM_000293	16q12.1	Phosphorylase kinase, beta subunit	PHKB	Glycogen storage disease
Alu	Deletion	NM_000152	17q25.3	Alpha-glucosidase	GAA	Glycogen storage disease type II
Alu	Deletion	NM_002133	22q12.3	Heme oxygenase-1	HMOX1	Heme oxygenase-1 deficiency
Alu	Deletion	NM_000132	Xq28	Coagulation factor VIII	F8	Hemophilia A

Alu	Deletion	NM_004402	1p36.32	Caspase-activated DNase, beta polypeptide	CAD	Hepatoma
L1	Deletion	NM_000249	3p22.3	MutL homolog 1	MLH1	Hereditary nonpolyposis colorectal cancer
Alu	Transposition /Deletion	NM_000202	Xq28	Iduronate-2-sulfatase	IDS	Hunter disease
Alu	Deletion /Duplication	NM_000527	19p13.2	Low density lipoprotein receptor	LDLR	Hypercholesterolemia, familial
Alu	Deletion	NM_000064	19p13.3	Alpha chain of complement component 3	C3	Hyperlipidemia, familial combined
Alu	Deletion	NM_000384	2p24.1	Apolipoprotein B precursor	APOB	Hypobetalipoproteinemia
Alu	Deletion /Duplication	NM_000194	Xq26.2	Hypoxanthine phosphoribosyl-transferase	HPRT	Lesch–Nyhan syndrome
Alu	Deletion /Duplication /Translocation	Numerous translocations [1]	–	Several, including MLL, MOZ and CPB	–	Leukemia, various acute/chronic forms
Alu	Deletion	NM_000546	17p13.1	Tumor protein p53	TP53	Li-Fraumeni syndrome

(*continued overleaf*)

Tab. 10 (*continued*)

Repeat	*Mutation*	*Locus*	*Chrom*	*Gene*	*Symbol*	*Phenotype*
Alu	Deletion	NM_003982	14q11.2	Solute carrier family 7	SLC7A7	Lysinuric protein intolerance
HERV15 [5]	Duplication	AZFa locus	Yq11.21	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide|*Drosophila* fat facets related	DBY|USP9Y	Male infertility
Alu	Deletion	NM_002608	22q13.1	Platelet-derived growth factor beta	PDGFB	Meningioma
Alu	Translocation	t(1;19)(q21.3;q13.2)	–	–	PAFAH1B3-CLK2 fusion	Mental retardation, ataxia and atrophy of the brain
Alu [6]	Deletion	NM_000512	16q24.3	N-acetylgalactosamine-6-sulfatase precursor	GALNS	Mucopolysaccharidosis type IVA
Alu/L1 [3], Alu	Deletion	NM_000109	Xp21.1–2	Dystrophin	DMD	Muscular dystrophy
Alu	Deletion	NM_000533	Xq22.2	Proteolipid protein gene	PLP1	Pelizaeus–Merzbacher disease
Alu	Deletion	NM_001171	16p13.11	ATP-binding cassette, subfamily C, member 6	ABCC6	Pseudoxanthoma elasticum
Alu	Deletion	NM_000330	Xp22.13	Retinoschisis factor	RS1	Retinoschisis
Alu	Deletion	NM_000521	5q13.3	Hexosaminidase B, beta polypeptide	HEXB	Sandhoff disease
Alu	Deletion	NM_000022	20q13.12	Adenosine deaminase	ADA	Severe combined imminodeficiency (SCID)
Alu	Deletion	del(5)(q12.3)	–	–		Small-cell lung cancers

Alu	Translocation	t(X;18)	–	–		Synovial sarcoma
Alu	Deletion	NM_005502	9q31.1	ATP-binding cassette transporter, subfamily A member 1	ABCA1	Tangier disease
Alu	Deletion	NM_000520	15q24.1	Hexosaminidase A, alpha polypeptide	HEXA	Tay Sachs disease
Alu	Deletion	del(16)(p)	–	Alpha-globin gene cluster	–	Thalassemia, alpha
Alu	Deletion	del(11)(p)	–	Beta-globin gene cluster	–	Thalassemia, delta, gamma
Alu	Deletion	NM_000488	1q25.1	Serine (or cysteine) proteinase inhibitor, clade C (antithrombin)	SERPINC1	Thrombophilia
Alu	Deletion	NM_000548	16p13.3	Tuberous sclerosis 2	TSC2	Tuberous sclerosis
Alu	Deletion	NM_025237	17q21.31	Sclerosteosis	SOST	van Buchem disease
Alu [2]	Deletion	NM_000552	12p13.31	von Willebrand factor A	VWF	von Willebrand disease
Alu	Deletion	NM_000377	Xp11.23	Wiskott-Aldrich syndrome protein	WAS	Wiskott-Aldrich syndrome
Alu	Translocation	DXYS5 locus	–	–		XX maleness

Notes: The table represents a significant update of previously published reviews. Each row corresponds to one or several rearrangements affecting the same gene. [1]: There are dozens of cases of Alu involvement in recombination leading to various forms of both acute and chronic leukemias, and related carcinomas. The disorders known to be, at least partially, caused by Alu recombinations are: t(8;16)(p11;p13), t(11;18)(q21;q21), t(2;8), t(14;19), t(9;11), t(14;18), t(11;19)(q23;p13.3), t(10;16)(q22;p13), t(4;11), Philadelphia translocation t(9;22) and variants, 7q22-q34 inversion, and dup(11)(q23). [2]: Nonhomologous Alu recombination. [3]: Nonhomologous recombination between Alu and L1 elements. [4]: Nonhomologous recombination induced by a polyA tail of an Alu element. [5]: Homologous recombination between two HERV15 elements produced duplication of a 780-kb segment on chromosome Y harboring two genes DBY and USP9Y. [6]: Specific deletion of a single Alu monomer in an intronic Alu element, probably by nonhomologous recombination.

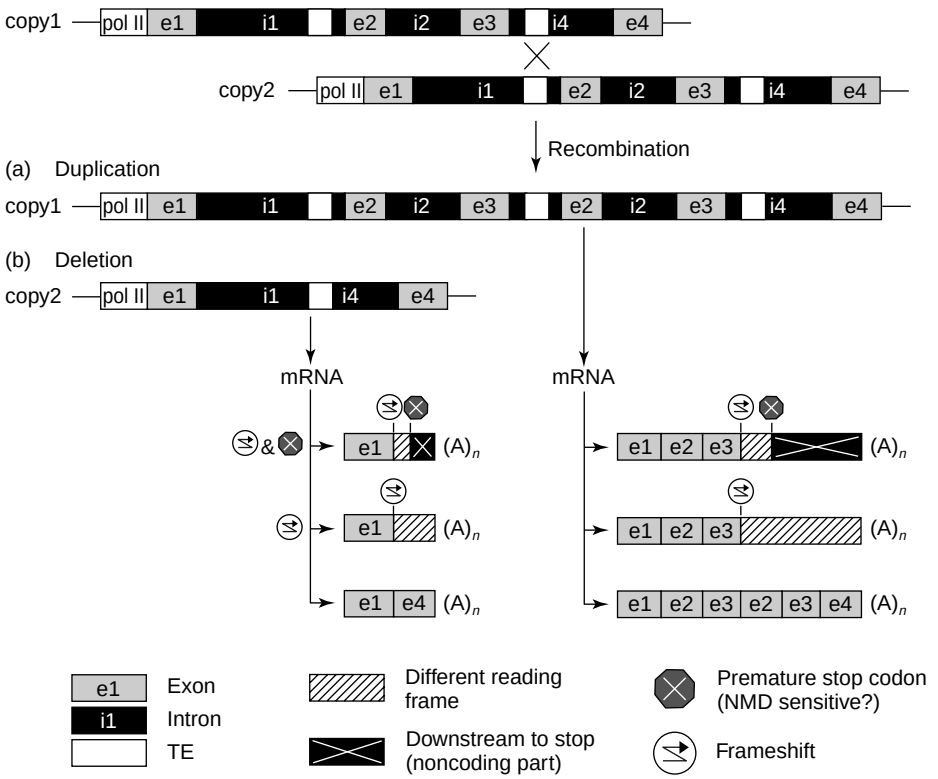

Fig. 14 Recombinations between repetitive elements and human disease. Figure shows classical recombination between two direct intragenic repeats, producing (a) exons duplication; and (b) exon deletion. The effects are similar to those described in Fig. 13.

are likely to be NMD-sensitive, and the resulting phenotype will be manifested by lack of the corresponding protein.

The prevalence of Alu-mediated rearrangements is overwhelming: 55 out of the 61 genes in Table 10 are affected by Alu-induced recombinations. Only three genes with L1 homologous recombinations have been reported so far, and only one case is known for endogenous retroviruses (Table 10). One deletion and one translocation were induced by nonhomologous Alu-L1 recombinations. The excess of detected Alu recombinations is probably related to the Alu genomic distribution and structure. A typical Alu is about 300 bp long, and this size seems to be sufficient for efficient homologous recombination in mammalian cells, where the minimal length requirement is around 150 bp. Alus are by far the most frequent repeats found in human genes. L1 and endogenous retroviruses are underrepresented in GC- and gene-rich regions and are often truncated or rearranged. The rearrangements disrupt the long sequence similarity necessary for homologous recombination. In summary, Alu elements frequently provide closely spaced, highly identical copies prone to recombination in GC-rich and

gene-rich regions, and thus represent the main risk for human genetic disorders caused by recombination between repetitive elements.

Alu sequences contain several regions with elevated occurrence of recombination breakpoints. These include promoters and the AT-rich linker between the left and right Alu monomer. Moreover, Alu elements harbor a 26-bp motif (positions 22–47 at the Alu consensus), possibly associated with increased frequency of recombination. The region contains the pentanucleotide motif core CCAGC, which is also part of chi, and an 8-bp sequence known to stimulate recBC-mediated recombination in *Escherichia coli*.

Alu repeats have also been implicated in the etiology of some recurrent somatic chromosome rearrangements detected in neoplastic tissues. The large-scale rearrangements lead to visible karyotype changes such as Philadelphia translocation, which is associated with chronic myelogenous leukemia. Chromosomal aberrations found in cancers result in loss of heterozygosity and in cancer progression. Apart from somatic translocation, nonlethal cytogenetically balanced chromosomal translocations in germline cells can lead to constitutional congenital disorders. An Alu-mediated translocation, t(9;11)(p23;q23), was shown to cause a familial form of the bipolar affective disorder (manic depression). Also, a patient with translocation t(1;19)(q21.3;q13.2) induced by nonhomologous Alu-L1 recombination was diagnosed with brain atrophy, ataxia, and mental retardation.

Some rearrangements may protect against deleterious changes. For example, the reciprocal translocation t(11;22)(q24;q12) fusing the Ewing's sarcoma (EWS) gene on 22q.12 with the *FLI1* gene on 11q.24, is associated with about 80% of primitive neuroectodermal tumors, including Ewing's sarcoma. However, a 2.4 kb Alu-mediated deletion in ESW intron 6, located near the translocation hotspot, seems to stabilize the region, since the carriers of this allele have a tenfold decreased incidence of EWS compared to populations without the 2.4 kb deletion.

In summary, repetitive elements contribute to human diseases through a variety of mechanisms. While the pathogenetic effects of repeat-encoded mRNAs and proteins are still debatable, repeat insertions and recombinations are significant contributors to genetic variability and abnormalities. The overall frequency of such events is unknown. Typical PCR-based detection of human mutations fails to amplify many large-scale variations, and the resulting defects are likely to be underrepresented in mutation databases. Also, germline recombination producing large rearrangements can terminate embryonic development in early stages, thus providing no opportunity for prenatal diagnostics.

The impact of retrotransposon insertions is small compared with the effect of deletions. Insertions correspond to about 0.1% of all human diseases. Alu-mediated recombinations are estimated to contribute to only 0.3% of human diseases. However, since large rearrangements (about 7.4% of all human mutations) often lack precise determination of breakpoints, it is possible that the impact of Alu recombination is higher.

5
Potential Contribution of Repetitive DNA to Cellular Functions

Satellites and TEs are thought to be involved in a broad variety of processes

directly related to the host cell. They can affect host replication, imprinting, gene silencing, transcriptional and posttranscriptional regulation, alternative splicing, and evolution of genes and proteins. Below are critically reviewed examples of such involvements.

It is believed that alpha satellites are functional elements in human centromeres, and they bind key proteins involved in host replication (see Sect. 2.2.1). However, the complete lack of alpha satellites in *de novo*-created (functional) neocentromeres indicates that they may represent well-adapted selfish DNA.

It has been proposed that L1s may be involved in inactivating chromosome X in female somatic cells. This view is based on the observed overrepresentation of L1 elements on human chromosome X, particularly near the X inactivation center. However, the lack of L1 clusters in the mouse X inactivation center indicates that L1 accumulation near the inactivation start may not be functionally related to the imprinting of the X chromosome.

As mentioned in Sect. 3, TEs contain different transcription regulation sites including promoters, enhancers, and polyadenylation signals. Accordingly, insertions of TEs into regulatory regions may alter the gene expression, and thus they can be negatively selected. For example, the majority of L1 and HERV elements present in introns are in the opposite orientation to the transcription direction of the corresponding genes. The most likely explanation is that polyA signals of these retrotransposons prematurely terminate the transcription if inserted in the direct orientation. Nevertheless, at least some TE-derived polyA signals may be recruited by cellular genes. Pol II transcription signals encoded by endogenous retroviruses and DNA transposons may serve as mobile "ready-to-use" blocks involved in the evolution of host genes. In several cases, TE-derived promoters and enhancers are incorporated into human genes. However, repetitive elements are found in many genomic regions, including promoters. Therefore, repeats present in these regions may be tolerated rather than being functional.

Potentially important regulation of genes by TEs may also occur at the RNA level. Antisense transcripts from promoter regions of TE inserted into introns can inhibit gene expression. For instance, a HERV-K element inserted into intron 9 of complement C4 gene in the orientation opposite to the gene transcription produces an antisense mRNA modulating the gene expression. Transposable elements are frequently found in untranslated regions of many mRNAs. Therefore, these mRNAs can also be regulated by transcripts derived from other similar copies of genomic TEs. Expressed SINE sequences can be acquired by the host as novel RNA genes. Indeed, *BC200*, a small RNA gene expressed in the primate brain, is the best-studied example of an Alu element recruited by the genome.

Protein-coding TEs can also be recruited as host protein-coding genes. Table 11 lists examples of TE-derived genes present in the human genome. Well-known examples of such genes are V(D)J recombinase and syncytin. V(D)J recombinase, which catalyzes rearrangements of immunoglobulin and T-cell receptor genes in lymphocytes, shares enzymatic characteristics with mariner/Tc1-like DNA transposases. Syncytin, a cell-fusion protein probably participating in placental morphogenesis and antiviral defense, was derived from the *env* gene encoded by the HERV17 (HERV-W) retrovirus. While the V(D)J

Tab. 11 Known human protein-coding genes derived from transposable elements.

GenBank ID	*Gene name*	*Related transposon family*	*Type*
pID 7959287	*KIAA1513*	Tc1/Pogo/Tc2	DNA transposon
pID 7662294	*KIAA0766*	hAT	DNA transposon
pID 7513096	*JRKL*	Tc1/Pogo	DNA transposon
pID 7513011	*KIAA0543*	hAT/Tip100/Tip100	DNA transposon
pID 7243087	*KIAA1353*	hAT	DNA transposon
pID 7243087	*Buster2*	hAT	DNA transposon
pID 7021900		hAT/Charlie3	DNA transposon
pID 6581097	*Buster3*	hAT	DNA transposon
pID 6581095	*Buster1*	hAT	DNA transposon
pID 6453533		piggyBac/MER85	DNA transposon
pID 4758872	*DAP4,pP52*^{rIPK}	hAT/Tip100/Zaphod	DNA transposon
pID 4504807	*Jerky*	Tc1/Pogo/Tigger4/Zombi	DNA transposon
EST 6986275	*Sancho*	Tc1/Pogo	DNA transposon
pID 4263748		hAT	DNA transposon
pID 4160548	*Tramp*	hAT	DNA transposon
pID 3413884	*KIAA0461*	Tc1/Pogo/Tc2	DNA transposon
pID 3327088	*KIAA0637*	hAT	DNA transposon
pID 29863	*CENP-B*	Tc1/Pogo	DNA transposon
pID 2231380		Tc1/Mariner/Hsmar1	DNA transposon
pID 131827	*V(D)J recombinase (RAG1/2)*	Tc1/Mariner	DNA transposon
pID 10439762		hAT/Tip100	DNA transposon
pID 10439744		hAT/Tip100	DNA transposon
pID 10439678		hAT	DNA transposon
pID 10047247	*KIAA1586*	hAT/Tip100	DNA transposon
BAC 4309921		piggyBac/MER85	DNA transposon
BAC 3522927		hAT/Tam3	DNA transposon
NP_776172		Harbinger	DNA transposon
NT_006413		P	DNA transposon
pID 4773880	*Syncytin*	HERV-W	LTR retrotransposon
pID 1196425	*HERV-3 env*	HERV-R	LTR retrotransposon
NP_055883	*PEG10*	Gypsy	LTR retrotransposon
AL117190	*Rtl1*	Gypsy	LTR retrotransposon
XP_291322		Gypsy	LTR retrotransposon
NP_689907		Gypsy	LTR retrotransposon
XP_293405		Gypsy	LTR retrotransposon
NP_065820		Gypsy	LTR retrotransposon
NP_036449		Gypsy	LTR retrotransposon
Z75407		Gypsy	LTR retrotransposon

recombinase is an ancient gene, possibly derived from the DNA transposase ~400 million years ago, syncytin is less than 5 million years old. Some host proteins listed in Table 11, like syncytin and Jerky, are entirely derived from proteins encoded originally by TEs. The remaining protein-coding genes were formed by the fusion of the preexisting host proteins with transposon-encoded proteins. Such fusions might have been initiated by alternative splicing.

DNA transposons are most abundant among recruited genes, despite their relatively low abundance in the genome compared to other TEs (Table 11). Rtl1 is the only clear example of RT recruited in the host gene. Genes derived from DNA transposase are probably involved in specific DNA binding and cleavage (e.g. the V(D)J recombinase). Surprisingly, eight genes are derived from Gypsy-like LTR retrotransposons, whose remnants are otherwise not detectable in the human genome. Two genes are derived from *env* genes encoded by HERV-W and HERV-R elements. Proteins from *env*-derived genes may prevent retroviral infections by saturating the cellular receptors for exogenous retroviruses.

The extent of the contribution of TEs to the host protein-coding capacity is still not well known. The number of well-supported examples of TE-derived host proteins is relatively small (Table 11). On the other hand, analyses of human mRNA sequences disclosed several hundreds of mRNAs with one or several exons derived from TEs. However, some of these TE-containing sequences can represent common contaminations in EST libraries. In addition, many of the transcripts contain short reading frames, indicating that they do not code for functional proteins. Indeed, it should be noted that the majority of cellular polymerase II transcripts do not code for any proteins and never reach the cytoplasm. Cellular mRNA control mechanisms such as the nonsense-mediated RNA decay protect cells from potentially pathogenic truncated or aberrant proteins. Indeed, products of TE-containing transcripts are seldom detected at the protein level.

In conclusion, while there are a few well-documented examples of novel functional genes derived from TEs, the extent of this phenomenon is unknown. Many proteins deposited in public databases are hypothetical ORFs predicted by statistical methods based on computational analysis of genomic DNA and/or cDNA sequences. These methods are less than 90% accurate and, as a result, they may include TEs in erroneously predicted proteins. Analogously, databases of sites involved in transcription regulation may include regions whose "regulatory" effects are shown *in vitro* only.

6
Databases and Programs for Analysis of Repetitive DNA

About half of the human genome is composed of various repetitive elements. Detection of repetitive elements is a basic step in many biologically important analyses, including, but not limited to, sequence assembly during genome sequencing, genome annotation, similarity searches, and gene and coding sequence prediction. Therefore, specialized tools and databases have been developed for the detection and masking of repetitive elements.

6.1
Databases of Repetitive Elements

The first database of human repetitive elements, Repbase, was published in 1992. Currently it contains over 620 families and subfamilies of human repeats. Repbase became a regularly updated general database of eukaryotic repetitive elements, renamed as Repbase Update (RU). RU is accompanied by the monthly online journal Repbase Reports, which publishes newly discovered eukaryotic repeats.

Currently, RU is released in the EMBL and Fasta formats. Only the EMBL format contains a detailed description and annotation of every RU consensus sequence. In addition, a version of RU manually tailored for RepeatMasker is available. RU contains several sections dedicated to separate eukaryotic species. Human TEs are listed in the section "humrep.ref." Most of the abundant processed pseudogenes can be obtained from the "pseudo.ref" library. The "simple.ref" library contains the most common simple repeat motifs.

Apart from Repbase Update, there are other useful online resources for studies of human repeats. Table 12 lists selected databases as well as related resources devoted to human repetitive elements.

6.2 Programs for Detection and Analysis of Repetitive DNA

There are two basic approaches for detection of repetitive elements in biological sequences. The first is based on the detection of similarity with previously known repeats. The second *de novo* approach is based on a

Tab. 12 Databases of human repeats, and related sources.

Database	*URL description*
Repbase Update	http://www.girinst.org/Repbase_Update.html *De facto* the only specialized database of eukaryotic repetitive elements. To date, Repbase Update contains consensus sequences and annotation for more than 4000 different eukaryotic repetitive elements. The content can be searched through sequence similarity search on the Censor web server, or through the text search using keywords.
HERVd	http://herv.img.cas.cz/ The Database of Human Endogenous Retroviruses (HERVs) maps all known endogenous retroviruses in the human genome. Retroviruses can be searched using several criteria such as family, length, chromosome etc., or by fasta similarity search. Several characteristics including the GC content, CpGs are calculated.
Human TE Insertion	http://www.med.upenn.edu/genetics/labs/kazazian/human.html The database lists known insertions of repetitive elements into the human genome associated with human genetic diseases.
Exapted retrogenes	http://www-ifi.uni-muenster.de/exapted-retrogenes/tables3.html Examples of vertebrate regulatory elements, parts of coding regions or genes generated by retroelements. Vertebrate genes generated by retrosequences.
UCSC GoldenPath	http://genome.ucsc.edu/ The UCSC human genome assembly contains extensive annotation files including tables with maps of repetitive elements in the human genome.
Ensembl	http://www.ensembl.org/ The ensembl genome assembly contains extensive annotation files including tables with maps of repetitive elements in the human genome.

search for characteristic properties of particular repeats. Table 13 lists some of the most popular programs developed for the detection of repetitive elements.

Similarity-based methods typically compare a library of known repetitive sequences against genomic DNA using general programs for sequence similarity search such as blast, wublast or cross_match. The repetitive library can be any custom set of repeats; however, specialized programs nearly exclusively rely on consensus sequences from Repbase. *De novo* detection methods do not require prior knowledge of repeats and the guiding principle of this approach is to search for the presence of multiple copies in the genome.

Standard detection programs for interspersed repeats use similarity-based detection of repeats. Censor and MaskerAid require wublast, while RepeatMasker uses the Smith-Waterman algorithm implemented in cross_match. For large quantities of data or for quick searches, it is better to use Censor or

Tab. 13 Programs for detection of repetitive sequences in biological sequences.

Program (package)	***URL description***
Censor	http://www.girinst.org/Censor_Server.html General package for similarity-based identification of interspersed and tandem repeats in both nucleic and amino acid sequences. Requires wublast and Repbase Update repeat libraries. Both online and local versions are available.
RepeatMasker	http://ftp.genome.washington.edu/cgi-bin/RepeatMasker General package for similarity-based identification of interspersed and tandem repeats in nucleic acid sequences. Requires cross_match and Repbase Update repeat libraries. Both online and local versions are available.
MakerAid	http://sapiens.wustl.edu/maskeraid/ Program performs a similarity-based detection of interspersed and tandem repeats in nucleic acid sequences. Can be used with RepeatMasker for fast repeat identification. Requires wublast and external repeat libraries. Local versions only.
Tandem repeat finder	http://tandem.biomath.mssm.edu/trf/trf.html An excellent program to for *de novo* detection of tandem repeats in DNA sequences. Both online and local versions available.
Satellites	http://bioweb.pasteur.fr/seqanal/interfaces/satellites.html Program to for *de novo* detection of tandem repeats in DNA sequences. Online version only.
Sputnik	http://espressosoftware.com:8080/esd/pages/sputnik.jsp Scans for 2–5 long tandem repeats using *de novo* search.
EMBOSS	http://www.hgmp.mrc.ac.uk/Software/EMBOSS/index.html A general open source package for analysis of biological sequences. Two included programs etandem and equicktandem detect tandem repeats using *de novo* search. Local versions only.

MaskerAid, since they are much faster than cross_match-based RepeatMasker. RepeatMasker and MaskerAid are restricted to nucleic acid sequences only. Censor, on the other hand, can perform virtually any similarity search implemented in the wublast package, including translated searches, and can even detect repetitive elements in protein sequences. All these programs produce maps of repetitive elements in query sequences, with the exact position of each repeat. These three programs also "mask" sequences, that is, they replace detected repeats by special characters, usually by "N" for nucleic and "X" for protein sequences. Such masking is essential for many database searches or gene predictions. Moreover, all these programs detect the most frequent simple repeats.

Unlike in the case of interspersed repeats, *de novo* detection is the primary approach in the analysis of simple repeats. There are several packages for the detection of microsatellites with short unit sizes, but for minisatellites with large units there are only a few suitable programs. In such situations, Tandem Repeat Finder can be applied.

In summary, for routine detection of known repeats it is best to use Censor, RepeatMasker, or MaskerAid with appropriate repeat libraries (Repbase Update libraries). These programs also detect the majority of simple sequence repeats, particularly the most abundant microsatellites. Specialized analysis of simple sequence repeats may involve specialized *de novo* detection programs. The majority of programs listed in Table 13 can be installed only on Linux/Unix-related systems. Users of other operating systems may use several programs that have online submission forms, as specified in Table 13.

See also Genetics, Molecular Basis of; Genomic Sequencing (Core Article).

Bibliography

Books and Reviews

Batzer, M.A., Deininger, P.L. (2002) Alu repeats and human genomic diversity, *Nat. Rev. Genet.* **3**, 370–379.

Berg, D.E., Howe, M.M. (Eds.) (1989) *Mobile DNA*, American Society for Microbiology, Washington.

Coffin, J.M., Hughes, S.H., Varmus, H.E. (Eds.) (1997) *Retroviruses*, Cold Spring Harbor Laboratory Press, New York.

Cooper, D.N. (2003) *Encyclopedia of The Human Genome*, Nature Publishing Group, London.

Jurka, J. (1998) Repeats in genomic DNA: mining and meaning, *Curr. Opin. Struct. Biol.* **8**, 333–337.

Lower, R., Lower, J., Kurth, R. (1996) The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5177–5184.

Maraia, R.J. (Ed.) (1995) *The Impact of Short Interspersed Elements (SINEs) on the Host Genome*, Springer, New York.

Ostertag, E.M., Kazazian, H.H. Jr. (2001) Biology of mammalian L1 retrotransposons, *Annu. Rev. Genet.* **35**, 501–538.

Smit, A.F. (1996) The origin of interspersed repeats in the human genome, *Curr. Opin. Genet. Dev.* **6**, 743–748.

Weiner, A.M., Deininger, P.L., Efstratiadis, A. (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information, *Annu. Rev. Biochem.* **55**, 631–661.

Wilkinson, D., Mager, D., Leong, J. (1994) Endogenous Human Retroviruses, in: Levy, J. (Ed.) *The Retroviridae*, Plenum Press, New York, pp. 465–535.

Primary Literature

Seminal Early References

Britten, R.J., Davidson, E.H. (1971) Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty, *Q. Rev. Biol.* **46**, 111–137.

Britten, R.J., Baron, W.F., Stout, D.B., Davidson, E.H. (1988) Sources and evolution of human Alu repeated sequences, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4770–4774.

Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F., Kazazian, H.H. Jr. (1991) Isolation of an active human transposable element, *Science* **254**, 1805–1808.

Corneo, G., Ginelli, E., Polli, E. (1967) A satellite DNA isolated from human tissues, *J. Mol. Biol.* **23**, 619–622.

Esnault, C., Maestre, J., Heidmann, T. (2000) Human LINE retrotransposons generate processed pseudogenes, *Nat. Genet.* **24**, 363–367.

Feng, Q., Moran, J.V., Kazazian, H.H. Jr., Boeke, J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition, *Cell* **87**, 905–916.

Hattori, M., Kuhara, S., Takenaka, O., Sakaki, Y. (1986) L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein, *Nature* **321**, 625–628.

Jeffreys, A.J., Wilson, V., Thein, S.L. (1985) Hypervariable 'minisatellite' regions in human DNA, *Nature* **314**, 67–73.

Jelinek, W.R., Toomey, T.P., Leinwand, L., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L., Schmid, C.W. (1980) Ubiquitous, interspersed repeated sequences in mammalian genomes, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1398–1402.

Jurka, J. (1997) Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1872–1877.

Jurka, J., Smith, T. (1988) A fundamental division in the Alu family of repeated sequences, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4775–4778.

Martin, F., Olivares, M., Lopez, M.C., Alonso, C. (1996) Do non-long terminal repeat retrotransposons have nuclease activity? *Trends Biochem. Sci.* **8**, 283–285.

Moos, M., Galwitz, D. (1983) Structure of two human beta-actin-related processed genes one of which is located next to simple repetitive sequence, *EMBO J.* **2**, 757–761.

Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., Kazazian, H.H. Jr. (1996) High frequency retrotransposition in cultured mammalian cells, *Cell* **87**, 917–927.

Ono, M., Yasunaga, T., Miyata, T., Ushikubo, H. (1986) Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome, *J. Virol.* **60**, 589–598.

Prosser, J., Frommer, M., Paul, C., Vincent, P.C. (1986) Sequence relationships of three human satellite DNAs, *J. Mol. Biol.* **187**, 145–155.

Quentin, Y. (1988) The Alu family developed through successive waves of fixation closely connected with primate lineage history, *J. Mol. Evol.* **27**, 194–202.

Repaske, R., O'Neill, R.R., Steele, P.E., Martin, M.A. (1983) Characterization and partial nucleotide sequence of endogenous type C retrovirus segments in human chromosomal DNA, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 678–682.

Schmid, C.W., Deininger, P.L. (1975) Sequence organization of the human genome, *Cell* **6**, 345–358.

Schmid, C.W., Jelinek, W.R. (1982) The Alu family of dispersed repetitive sequences, *Science* **216**, 1065–1070.

Shafit-Zagardo, B., Maio, J.J., Brown, F.L. (1982) KpnI families of long, interspersed repetitive DNAs in human and other primate genomes, *Nucleic Acids Res.* **10**, 3175–3193.

Smit, A.F., Riggs, A.D. (1996) Tiggers and DNA transposon fossils in the human genome, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1443–1448.

Soriano, P., Meunier-Rotival, M., Bernardi, G. (1983) The distribution of interspersed repeats is nonuniform and conserved in the mouse and human genomes, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1816–1820.

Ullu, E., Tschudi, C. (1984) Alu sequences are processed 7SL RNA genes, *Nature* **312**, 171–172.

Van Arsdell, S.W., Dension, R.A., Bernstein, L.B., Weiner, A.M., Manser, T., Gesteland, R.F. (1981) Direct repeats flank three small nuclear RNA pseudogenes in the human genome, *Cell* **26**, 11–17.

Vanin, E.F. (1985) Processed pseudogenes: characteristics and evolution, *Annu. Rev. Genet.* **19**, 253–272.

Weber, J.L., May, P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction, *Am. J. Hum. Genet.* **44**, 388–396.

Weiner, A.M. (1980) An abundant cytoplasmic 7S RNA is complementary to the dominant interspersed middle repetitive DNA sequence family in the human genome, *Cell* **22**, 209–218.

Recent Key Primary Literature

Amor, D.J., Choo, K.H. (2002) Neocentromeres: role in human disease, evolution, and centromere study, *Am. J. Hum. Genet.* **71**, 695–714.

Andersson, M.-L., Lindeskog, M., Medstrand, P., Westley, B., May, F., Blomberg, J. (1999) Diversity of human endogenous retrovirus class II-like sequences, *J. Gen. Virol.* **80**, 255–260.

Antonarakis, S.E. (1998) Recommendations for a nomenclature system for human gene mutations, *Hum. Mutat.* **11**, 1–3.

Batzer, M.A., Deininger, P.L., Hellmann-Blumberg, U., Jurka, J., Labuda, D., Rubin, C.M., Schmid, C.W., Zietkiewicz, E., Zuckerkandl, E. (1996) Standardized nomenclature for Alu repeats, *J. Mol. Evol.* **42**, 3–6.

Bedell, J.A., Korf, I., Gish, W. (2000) MaskerAid: a performance enhancement to repeat masker, *Bioinformatics* **16**, 1040–1041.

Berkhout, B., Jebbink, M., Zsiros, J. (1999) Identification of an active reverse transcriptase enzyme encoded by a human endogenous HERV-K retrovirus, *J. Virol.* **73**, 2365–2375.

Brosius, J. (1999) RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements, *Gene* **238**, 115–134.

Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., Kazazian, H.H. Jr. (2003) Hot L1s account for the bulk of retrotransposition in the human population, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5280–5285.

Chou, H.H., Hayakawa, T., Diaz, S., Krings, M., Indriati, E., Leakey, M., Paabo, S., Satta, Y., Takahata, N., Varki, A. (2002) Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11736–11741.

Chureau, C., Prissette, M., Bourdet, A., Barbe, V., Cattolico, L., Jones, L., Eggen, A., Avner, P., Duret, L. (2002) Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine, *Genome Res.* **12**, 894–908.

Cordonnier, A., Casella, J.F., Heidmann, T. (1995) Isolation of novel human endogenous retrovirus-like elements with foamy virus-related pol sequence, *J. Virol.* **69**, 5890–5897.

Deininger, P.L., Batzer, M.A. (1999) Alu repeats and human disease, *Mol. Genet. Metab.* **67**, 183–193.

Deininger, P.L., Batzer, M.A., Hutchison III, C.A., Edgell, M.H. (1992) Master genes in mammalian repetitive DNA amplification, *Trends Genet.* **8**, 307–311.

Dewannieux, M., Esnault, C., Heidmann, T. (2003) LINE-mediated retrotransposition of marked Alu sequences, *Nat. Genet.* **35**, 41–48.

Duncan, C., Biro, P.A., Choudary, P.V., Elder, J.T., Wang, R.R., Forget, B.G., de Riel, J.K., Weissman, S.M. (1979) RNA polymerase III transcriptional units are interspersed among human non-alpha-globin genes, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5095–5099.

Goncalves, I., Duret, L., Mouchiroud, D. (2000) Nature and structure of human genes that generate retropseudogenes, *Genome Res.* **10**, 672–678.

Grimaldi, G., Skowronski, J., Singer, M.F. (1984) Defining the beginning and the end of KpnI family segments, *EMBO J.* **3**, 1753–1759.

Hohjoh, H., Singer, M.F. (1997) Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon, *EMBO J.* **16**, 6034–6043.

Jeffreys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neil, D.L., Armour, J.A. (1994) Complex gene conversion events in germline mutation at human minisatellites, *Nat. Genet.* **6**, 136–145.

Jurka, J., Klonowski, P., Dagman, V., Pelton, P. (1996) CENSOR – a program for identification and elimination of repetitive elements from DNA sequences, *Comput. Chem.* **20**, 119–121.

Jurka, J., Kapitonov, V.V. (1999) Sectorial mutagenesis by transposable elements, *Genetica* **107**, 239–248.

Jurka, J. (2000) Repbase update: a database and an electronic journal of repetitive elements, *Trends Genet.* **16**, 418–420.

Jurka, J., Krnjajic, M., Kapitonov, V.V., Stenger, J.E., Kokhanyy, O. (2002) Active Alu elements are passed primarily through

paternal germlines, *Theor. Popul. Biol.* **61**, 519–530.

Kapitonov, V., Jurka, J. (1996) The age of Alu subfamilies, *J. Mol. Evol.* **42**, 59–65.

Kapitonov, V.V., Jurka, J. (2003) The esterase and PHD domains in CR1-like non-LTR retrotransposon, *Mol. Biol. Evol.* **20**, 38–46.

Kazazian, H.H. Jr. (1998) Mobile elements and disease, *Curr. Opin. Genet. Dev.* **8**, 343–350.

Kazazian, H.H. Jr., Goodier, J.L. (2002) LINE drive retrotransposition and genome instability, *Cell* **110**, 277–280.

Kolosha, V.O., Martin, S.L. (2003) High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1), *J. Biol. Chem.* **278**, 8112–8117.

Kolomietz, E., Meyn, M.S., Pandita, A., Squire, J.A. (2002) The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors, *Genes Chromosomes Cancer* **35**, 97–112.

Korenberg, J.R., Rykowski, M.C. (1988) Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands, *Cell* **53**, 391–400.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Lee, C., Wevrick, R., Fisher, R.B., Ferguson-Smith, M.A., Lin, C.C. (1997) Human centromeric DNAs, *Hum. Genet.* **100**, 291–304.

Lehrman, M.A., Schneider, W.J., Sudhof, T.C., Brown, M.S., Goldstein, J.L., Russel, D.W. (1985) Mutation on LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains, *Science* **227**, 140–146.

Lerman, M.I., Thayer, R.E., Singer, M.F. (1983) KpnI family of long interspersed repeated DNA sequences in primates: polymorphism of family members and evidence for transcription, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3966–3970.

Li, T.H., Schmid, C.W. (2001) Differential stress induction of individual Alu loci: implications for transcription and retrotransposition, *Gene* **276**, 135–141.

Lower, R. (1999) The pathogenic potential of endogenous retroviruses: facts and fantasies, *Trends Microbiol.* **7**, 350–356.

Lyon, M.F. (1998) X-chromosome inactivation: a repeat hypothesis, *Cytogenet. Cell. Genet.* **80**, 133–137.

Mager, D.L. (1999) Human endogenous retroviruses and pathogenicity: genomic considerations, *Trends Microbiol.* **7**, 431–432.

Martignetti, J.A., Brosius, J. (1993) BC200 RNA: a neural RNA polymerase III product encoded by a monomeric Alu element, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11563–11567.

Medstrand, P., Mager, D.L. (1998) Human-specific integrations of the HERV-K endogenous retrovirus family, *J. Virol.* **72**, 9782–9787.

Mi, S., Lee, X., Li, X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y., Edouard, P., Howes, S., Keith, J.C. Jr., McCoy, J.M. (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**, 785–789.

Moran, J.V., DeBerardinis, R.J., Kazazian, H.H. Jr. (1999) Exon shuffling by L1 retrotransposition, *Science* **283**, 1530–1534.

Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T.D., Taccioli, G.E., Batzer, M.A., Moran, J.V. (2002) DNA repair mediated by endonuclease-independent LINE-1 retrotransposition, *Nat. Genet.* **31**, 159–165.

Paces, J., Pavlicek, A., Paces, V. (2002) HERVd: database of human endogenous retroviruses, *Nucleic Acids Res.* **30**, 205–206.

Okada, N., Hamada, M., Ogiwara, I., Ohshima, K. (1997) SINEs and LINEs share common 3′ sequences: a review, *Gene* **205**, 229–243.

Prak, E.T., Kazazian, H.H. Jr. (2000) Mobile elements and the human genome, *Nat. Rev. Genet.* **1**, 134–144.

Schlotterer, C., Tautz, D. (1992) Slippage synthesis of simple sequence DNA, *Nucleic Acids Res.* **20**, 211–215.

Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K., Willard, H.F. (2001) Genomic and genetic definition of a functional human centromere, *Science* **294**, 109–115.

Scott, A.F., Schmeckpeper, B.J., Abdelrazik, M., Comey, C.T., O'Hara, B., Rossiter, J.P., Cooley, T., Heath, P., Smith, K.D., Margolet, L. (1987) Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence, *Genomics* **1**, 113–125.

Shafit-Zagardo, B., Brown, F.L., Zavodny, P.J., Maio, J.J. (1983) Transcription of the KpnI families of long interspersed DNAs in human cells, *Nature* **304**, 277–280.

Singer, M.F. (1982) SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes, *Cell* **28**, 433–434.

Smit, A.F. (1993) Identification of a new, abundant superfamily of mammalian LTR-transposons, *Nucleic Acids Res.* **21**, 1863–1872.

Smit, A.F., Toth, G., Riggs, A.D., Jurka, J. (1995) Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences, *J. Mol. Biol.* **246**, 401–417.

Smit, A.F., Riggs, A.D. (1995) MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation, *Nucleic Acids Res.* **23**, 98–102.

Smit, A.F. (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes, *Curr. Opin. Genet. Dev.* **9**, 657–663.

Speek, M. (2001) Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes, *Mol. Cell. Biol.* **21**, 1973–1985.

Swergold, G.D. (1990) Identification, characterization, and cell specificity of a human LINE-1 promoter, *Mol. Cell. Biol.* **10**, 6718–6729.

Toth, G., Gaspari, Z., Jurka, J. (2000) Microsatellites in different eukaryotic genomes: survey and analysis, *Genome Res.* **10**, 967–981.

Voliva, C.F., Jahn, C.L., Comer, M.B., Hutchison, C.A., Edgell, M.H. (1983) The L1Md long interspersed repeat family in the mouse: almost all examples are truncated at one end, *Nucleic Acids Res.* **11**, 8847–8859.

Wallace, M.R., Andersen, L.B., Saulino, A.M., Gregory, P.E., Glover, T.W., Collins, F.S. (1991) A de novo Alu insertion results in neurofibromatosis type 1, *Nature* **353**, 864–866.

Wellinger, R.J., Sen, D. (1997) The DNA structures at the ends of eukaryotic chromosomes, *Eur. J. Cancer* **33**, 735–749.

Anthrax (*Bacillus anthracis*), Molecular Biology of

Kathryn B. Garber
Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA

1 Anthrax 309

2 The Virulence Plasmids 311

3 Molecular Identification of *B. anthracis* 312
3.1 Molecular Methods to Distinguish *B. anthracis* from Other Bacteria 312
3.1.1 Post-October 2001 Research 315
3.2 How do You Tell One Strain from Another? 317

4 Regulation of Virulence 319
4.1 AtxA 319
4.2 PagR 320
4.3 AbrB 321
4.4 AcpA 322
4.5 Other Potential Virulence Factors 322

5 Regulation of Sporulation/Germination 323

6 Conclusions 323

Bibliography 324
Books and Reviews 324
Primary Literature 324

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Keywords

Spore
A dormant, protective form that certain bacteria can take to help them survive adverse conditions.

Phagosome
Cellular compartments formed when cells engulf large particles, such as bacteria, from their environment.

Endocytosis
The process by which cells take up material from their environment. This involves membrane invagination and the formation of an internal membranous compartment called an *endosome*.

Transposon
A mobile genetic element that is able to insert a copy of itself into a chromosome or plasmid.

Virulence Factor
A factor that helps a pathogen cause disease.

Regulon
A set of genes whose expression is regulated in a coordinated fashion.

Abbreviations

PA	Protective antigen
EF	Edema factor
LF	Lethal factor
AFLP	Amplified fragment length polymorphism
MLVA	Multiple-locus variable-number tandem repeat analysis
LR REP PCR	Long-range repetitive element polymorphism-PCR
VNTR	Variable-number tandem repeat

Bacillus anthracis, the causative agent of anthrax, has been a subject of research for over 150 years, beginning with the work of Rayer and Davaine who found "small filiform bodies" in the blood of anthrax-infected sheep and, from this finding, coined the term *bacteria*. Several years later, Koch proved conclusively that *B. anthracis* was the causative agent of anthrax, providing the first demonstration that a disease was

caused by a specific bacterium. He later developed the first live attenuated bacterial vaccine with a heat-attenuated strain of *B. anthracis*. Despite its illustrious beginnings as a subject of research, an understanding of the molecular biology of *B. anthracis* has lagged behind that of some other bacteria. *Bacillus anthracis* has recently come back into focus as a target of research with the goal of being able to readily detect this bacterium in environmental samples and to understand its pathogenesis and evolution better.

1 Anthrax

Bacillus anthracis is a gram-positive, spore-forming bacterium that normally resides in the soil, but when taken up into the body, causes anthrax. This facultative aerobe grows on many types of culture media and its colonies appear mucoid and are not generally hemolytic. Upon microscopic examination, chains of encapsulated, non-motile, rod-shaped bacteria are seen. For infection to occur, spores of *B. anthracis* may be taken up through one of three routes: inhalation, ingestion, or via an abrasion of the skin. From there, the spores are engulfed by macrophages and carried to regional lymph nodes. Spore germination in macrophages appears to be required for the establishment of disease, and the spores require this intracellular environment to stimulate germination. It has been shown that, at least in culture, the newly vegetative bacilli escape the phagosome to reach the cytoplasm of macrophages, where they multiply and are eventually released into the extracellular environment. Macrophages, therefore, end up playing a major role in the pathogenesis of *B. anthracis* because they provide an appropriate environment for spore germination and they also provide a transportation route that gives the bacteria access to the blood stream where high levels of bacteremia can ensue. Once this stage is reached, a shock-like syndrome and death follow; although this severe disease course is usually associated with inhalation or gastrointestinal anthrax, a cutaneous anthrax infection is generally contained prior to the achievement of bacteremia.

The majority of *B. anthracis* strains carry two large virulence plasmids, called pXO1 and pXO2. Both plasmids are required for full virulence. pXO2 carries the *cap* region, an operon encoding a poly-D-glutamic acid capsule that helps the bacteria evade the immune system by allowing them to resist phagocytosis. The other plasmid, pXO1, carries *pagA, lef,* and *cya*, the genes encoding the three components of the anthrax toxin: protective antigen (PA), lethal factor (LF) and edema factor (EF) respectively. The symptoms of anthrax are largely mediated by this toxin. In fact, an anthrax infection can be fatal even after the blood has been cleared of all *B. anthracis* owing to the toxin that has already been released into the body.

The toxin proteins work in binary combinations to produce toxic effects; PA is the carrier molecule that delivers the enzymatic portions of the toxin, EF and LF, to the cytosol of cells. There, EF, an adenylate cyclase, increases cAMP levels, which are presumed to play a role in the development of the

pulmonary edema and tissue swelling that are seen with anthrax infections. LF, on the other hand, appears to alter the production of cytokines by macrophages and to induce macrophage lysis. It is a zinc-dependent metalloprotease that cleaves MAP kinase kinases, leading to apoptosis of the cells through inhibition of an antiapoptotic signaling pathway that involves p38 MAP kinase. LF gets its ominous name from the fact that coinjection of PA and LF, even in the absence of any *B. anthracis*, is sufficient to kill an animal.

Briefly, the intoxication process, as we currently understand it, occurs as follows (Fig. 1): PA binds to a ubiquitous receptor, named *anthrax toxin receptor* (ATR), on mammalian cells. A 20-kD amino-terminal piece of PA is then cleaved off the protein by a furin-like protease, leaving the remaining 63-kD carboxy-terminal portion of PA (PA63) bound to the receptor. This proteolysis, also called *nicking*, facilitates the organization of PA63 into a seven-membered ring-shaped structure, reveals the binding sites for EF and LF, and stimulates endocytosis of the PA-LF and PA-EF complexes. The toxin complexes are trafficked to endosomal compartments, wherein low pH triggers the PA63 ring to insert into the endosomal membranes as a protein channel. Through a mechanism that involves unfolding of EF and LF, the PA63 channel delivers its ligands to the cytosol of the intoxicated cells, where they exert their enzymatic effects.

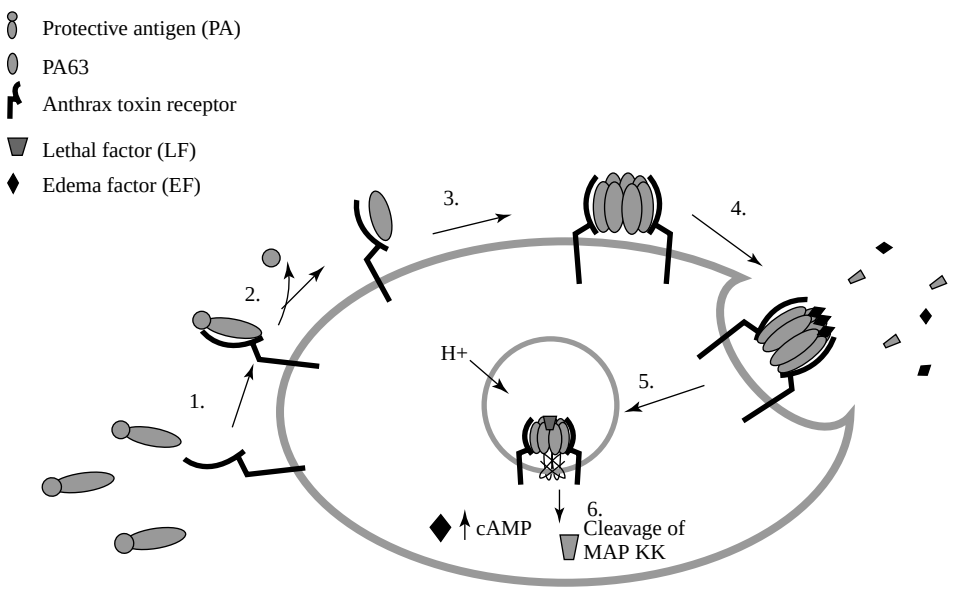

Fig. 1 Cellular intoxication cycle for anthrax. Step 1. PA binds to the anthrax toxin receptor. Step 2. PA is nicked, removing a 20-kD amino-terminal portion of the protein and leaving PA63 bound to the receptor. Step 3. Seven PA63 molecules come together to form a ring-shaped heptamer. Step 4. The PA63 heptamer binds EF or LF. Step 5. The toxin complex is endocytosed. Step 6. In response to a drop in pH in the endosomal compartment, the PA63 heptamer forms a membrane channel that mediates the translocation of EF and LF to the cytosol of cells, where they exert their enzymatic effects.

2
The Virulence Plasmids

A gapped, linear sequence of approximately 93 kbp has been generated for pXO2, but a detailed analysis of the sequence has not yet been published. Comparison of the pXO2 sequences from various *B. anthracis* isolates identified several variable pXO2 markers, leading to the suggestion that this plasmid is the source of much of the variability in the *B. anthracis* genome. More work needs to be done before the functions of the gene products encoded by this plasmid are discovered.

The 181-kbp pXO1 plasmid has been fully sequenced. It has 143 predicted open reading frames, most encoding proteins of unknown function. There are also several pXO1-associated phenotypes that have not been assigned to a particular gene. Among these phenotypes are different nutritional requirements, altered sensitivity to bacteriophages, and escape of *B. anthracis* from cultured macrophages.

The most striking feature of this plasmid is a 44-kbp region that has been designated a pathogenicity island (Fig. 2). These islands are large regions of DNA that confer virulence phenotypes and are believed to have been acquired from unrelated organisms. The pathogenicity island of pXO1 contains 31 predicted open reading frames, including all three toxin genes and their regulatory elements. It is flanked by two inverted and nearly identical copies of an insertion sequence element that serve as footprints for previous transposition of DNA. Also, within the island are genes encoding transposases and integrases – enzymes that are involved in the movement of DNA segments – which, along with the insertion sequences, suggest that there may be plasticity to this region. In fact, comparative restriction analysis of this region in different *B. anthracis* isolates indicates that the pathogenicity island is inverted in some strains.

The origin of this pathogenicity island is unclear, if, in fact, it did come from another species. One possibility is that it came from a member of the *Bacillus cereus* phylogenetic group, to which *B. anthracis* belongs, as some of the insertion sequence elements in the pathogenicity island are very similar to those isolated from other members of this group. Further supporting this hypothesis is the fact that the GC content of the pathogenicity island does not differ significantly from that of the rest of

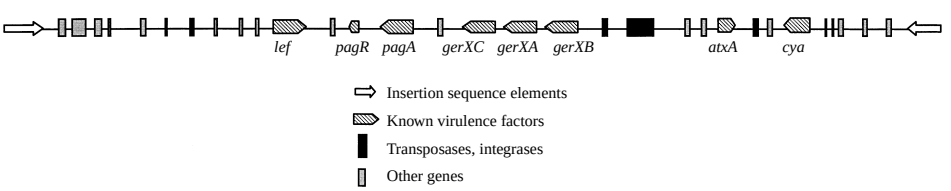

Fig. 2 The pathogenicity island of *B. anthracis*. Depicted is the basic structure of the pathogenicity island that is located on the pXO1 plasmid. It encodes, among other things, a major virulence gene regulator (*atxA*), the anthrax toxin proteins (*pagA*, *lef*, *cya*), and an operon involved in sporulation (*gerXA*, *gerXB*, and *gerXC*). Adapted from Okinaka, R.T., Cloud, K., Hampton, O., Hoffmaster, A.R., Hill, K.K., Keim, P., Koehler, T.M., Lamke, G., Kumano, S., Mahillon, J., Manter, D., Martinez, Y., Ricke, D., Svensson, R., Jackson, P.J. (1999) Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes, *J. Bacteriol.* **181**, 6509–6515.

pXO1 or the *B. anthracis* genome, nor is there a difference in codon usage between them, so the pathogenicity island must have evolved in a species very similar to *B. anthracis*.

In fact, other members of the *B. cereus* group possess many sequences related to the pXO1 coding sequences. A recent hybridization study found that DNA from three members of this group (two *B. cereus* strains and a *Bacillus thuringiensis* strain) each hybridized with over half of the probes that were used to represent the pXO1 open reading frame complement. This conservation of pXO1 genes is independent of evolutionary relatedness in terms of the chromosomal sequences, further supporting the idea that these sequences may shuffle between related species.

Lest the reader fear that the deadly anthrax toxin genes are jumping around between bacteria, I will point out that approximately 20% of the open reading frames tested in this study, including those encoding the anthrax toxin, were not found in any of the related species that were tested, and are, as far as we know, unique to anthrax. Further, the results of a phylogenetic analysis of *pag* sequences from several *B. anthracis* strains largely agreed with those from a similar analysis of chromosomal sequences. This suggests that the *pag* gene did not evolve separately from the chromosome, so this gene probably did not undergo horizontal transfer during the recent evolution of *B. anthracis*.

3 Molecular Identification of *B. anthracis*

Standard microbiological techniques can be used to distinguish *B. anthracis* from other bacilli. Distinguishing characteristics include lack of motility and hemolysis, a typical colony morphology (including a flat, tacky appearance and gray-white color), and susceptibility to lysis by gamma phage. Although these techniques can be used to identify *B. anthracis* in most cases, the threat of bioterrorism makes the development of rapid molecular surveillance methods, for use in environmental samples, a priority. Microbiological techniques rely upon isolation and culture of bacteria, and they generally take at least 24 h to complete. This fact can limit their successful use in certain situations. It would be desirable to develop faster diagnostic methods so that anthrax outbreaks can be identified and managed as rapidly as possible. A second limitation to standard microbiological techniques is the fact that anthrax victims who are treated with antibiotics before blood cultures are taken often have blood cultures that are negative for *B. anthracis*. For accurate diagnosis of these patients, molecular diagnostic techniques must be perfected.

As mentioned above, *B. anthracis* is genetically closely related to *B. cereus* and *B. thuringiensis* and, in fact, some researchers have proposed on the basis of multilocus enzyme electrophoresis as well as 16S and 23S rRNA sequence analysis, that these organisms be classified as subtypes of the same species. The relatedness of these bacteria has complicated efforts to develop *B. anthracis*-specific tests because the detection of related species as false-positives can be a problem.

3.1 Molecular Methods to Distinguish *B. anthracis* from Other Bacteria

The presence of virulence plasmids, including pXO1 and pXO2 in *B. anthracis*,

is the primary characteristic used to distinguish *B. anthracis* from related bacilli. However, it has been demonstrated experimentally that plasmids can be transferred between *B. cereus, B. thuringiensis,* and *B. anthracis.* Further, there are naturally occurring isolates of plasmidless *B. anthracis,* so a method to distinguish chromosomal DNA is important. Because the genomes of *B. anthracis* and *B. cereus* are estimated to share 98% DNA sequence homology, identification of unique diagnostic markers will be important for the definitive identification of the related strains. In the following section, several genetic sequences that have been purported to be unique to *B. anthracis* will be discussed, as will their usefulness in the identification of this species.

Ba813 was the first locus that was proposed as a unique marker for *B. anthracis.* A PCR assay for the detection of this chromosomal DNA fragment was proposed as a reliable method to detect this species. Subsequent studies found the Ba813 marker in other bacilli, so this marker alone does not appear to be sufficient to distinguish anthrax from environmental samples. The relationship of the Ba813-carrying, nonanthrax *Bacillus* strains to verifiable *B. anthracis* is not yet clear. Genetically, these strains are closely related to each other but distinct from *B. anthracis.*

Other assays for specific chromosomal regions have been developed, including one for variability in *rpoB,* which is almost exclusively found in *B. anthracis* strains. Interestingly, the one exception to this was a nonanthrax *Bacillus* strain that also carried Ba813.

A variable region in the *B. anthracis* chromosome has been identified using arbitrarily primed PCR. Only one complete open reading frame was present in this region, and it was dubbed *vrrA,* for variable region with repetitive sequence. PCR of this region amplified a fragment of distinct size from *B. cereus, Bacillus mycoides,* and *B. anthracis,* allowing these closely related species to be distinguished from each other. Although most of this variation was located in *vrrA,* many of the single nucleotide substitutions that differed between *B. anthracis* and the other bacilli were third position synonymous mutations that would not change the encoded amino acid sequence. The role of the *vrrA*-encoded protein is unknown, but the gene most closely resembles *shp2,* a gene that encodes a sheath protein in microfilarial parasites. A repetitive sequence, QYPQ, in the parasite protein is also predicted in the *vrrA*-encoded protein, and this sequence has been implicated in the covalent cross-linking of multiple protein subunits.

Rather than looking at one locus at a time, a more effective method of distinguishing *B. anthracis* from related species is to use techniques that look at multiple loci simultaneously. One of the methods that has been used is amplified fragment length polymorphism, or AFLP. In this method, restriction fragments of DNA from the test sample are ligated to linker fragments that are then used to prime an amplification reaction (Fig. 3). The labeled primers used in this reaction can be made more or less specific, so that a manageable number of fragments are produced in the final output. These amplified fragments can then be screened for length polymorphisms using gel electrophoresis, with the idea that there will be a species-specific pattern of restriction polymorphisms.

AFLP has the advantage that a large number of loci are examined at once and that no prior knowledge of the

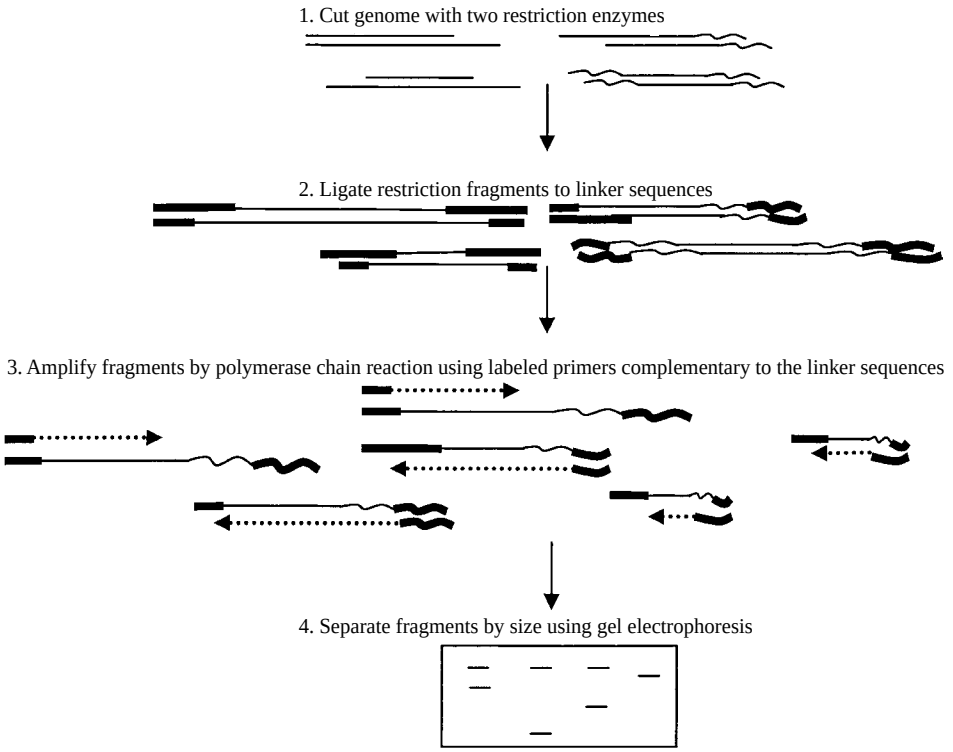

Fig. 3 The basic method for AFLP analysis. In this method, DNA from a sample is cut into pieces using restriction enzymes. The size of these fragments should differ between samples, based on the location of restriction enzyme sites in the DNA sequences. Rather than amplifying the DNA restriction fragments using sequences within them, linkers of known DNA sequence are attached to the ends of these fragments so that they may be amplified using primers complementary to the linker fragments. This allows one to amplify DNA even when the sequences in the target organism are unknown. The amplified fragments can then be separated by size so that the patterns of fragments between samples can be compared.

genetic makeup of the target organism is necessary. The AFLP profile of a sampled organism can be compared to that of known profiles for identification. These comparisons can be logistically difficult, especially when a large number of fragments are examined at once. To make this easier, AFLP profiles for a number of organisms are being archived and algorithms developed to automate these comparisons. This method does show promise for use in the identification of *B. anthracis*, because the AFLP profiles generated for a variety of *B. anthracis* strains show significant differences from closely related *Bacillus* species.

Long-range repetitive element-polymorphism PCR (LR REP-PCR) is another method that uses length polymorphisms to distinguish between bacterial species. This technique uses highly conserved repetitive extragenic palindromic (REP) sequences in DNA as target sequences in a PCR reaction. DNA sequences that lie between two adjacent REP sequences are amplified in these reactions (Fig. 4).

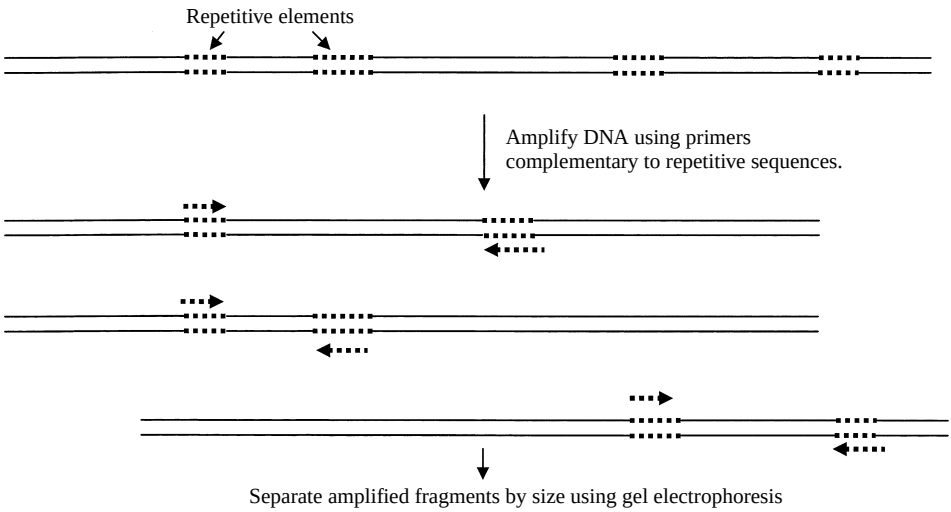

Fig. 4 Long-range repetitive element polymorphism-PCR (LR REP-PCR). Certain sequences occur many times within a bacterial genome, and these are called repetitive elements. The number and location of these REPs will vary between bacterial species. One can use primers complementary to the REPs to amplify fragments of DNA that will vary in size depending on the spacing of the REPs in the genome. As with AFLP, these fragments can be separated by size using gel electrophoresis and the patterns of fragments can be compared between samples.

PCR fragments of different size will be amplified in different strains, depending on the spacing of REPs in each. This technique clearly distinguishes between *B. anthracis* and related *Bacillus* species.

Another method that is proving useful for the identification of *B. anthracis* is multiple-locus variable-number tandem repeat analysis (MLVA). Variable-number tandem repeat (VNTR) loci are short, repeated sequences of DNA that vary in copy number, thereby creating a length polymorphism that can be detected by PCR amplification of fragments in the region. Because of their greater diversity, these markers have more power to discriminate between species of bacteria than other molecular typing methods. Most of the MLVA primers developed for *B. anthracis* do not amplify sequences from related *Bacillus* species, so this method appears to be useful in distinguishing *B. anthracis* from closely related species.

3.1.1 Post-October 2001 Research

Research efforts aimed at understanding and detecting anthrax have been ramped up since the intentional release of anthrax in 2001. This research has covered everything from epidemiologic surveillance, to methods of sample collection during an investigation, and to rapid methods for the detection of anthrax during a potential outbreak. The following section provides a discussion of the technologies that were found to be useful in the detection of anthrax.

The sequence of the 16S rRNA gene is often used to identify unknown bacteria. Owing to the similarities between the 16S rRNA sequences of *B. anthracis* and its relatives, it was thought that this

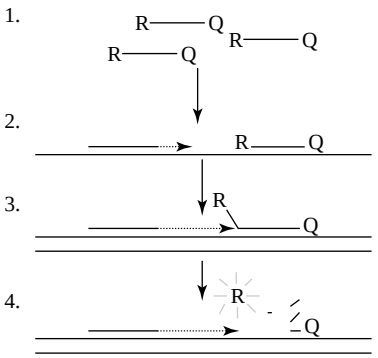

Fig. 5 5′ nuclease real-time PCR assay. 1. A probe complementary to the target sequence is labeled with a fluorescent reporter (R) and a quencher (Q). At the beginning of the reaction, the close proximity of these molecules allows the quencher to act on the reporter and little fluorescent signal can be detected. 2. As the target sequence is amplified, the probe binds to it. 3. The Taq DNA polymerase in the reaction uses its 5′ to 3′ nuclease activity to degrade the probe. 4. As the probe is degraded, the reporter and quencher separate and the quencher no longer acts on the reporter, leaving it free to fluoresce. This signal is used as a measure of the amount of probe that has been degraded, which is directly related to the amount of target sequence in the reaction.

analysis would not be useful for the identification of *B. anthracis*. However, this conclusion was based on a limited number of 16S rRNA sequences. During the investigation of the outbreak, the 16S rRNA sequences from a number of *B. anthracis, B. cereus*, and *B. thuringiensis* strains were determined, and it was found that, indeed, it was possible to distinguish between the species. All bacillus strains in this study, including 86 *B. anthracis* strains, possessed 16S rRNA types that were unique to their species. In Genbank, there are, however, 3 16S rRNA sequences from *B. anthracis* that fall into a type that was exclusively found in *B. cereus* in this study. The explanation for this discrepancy is unclear at this point, and it may be due to a methodological or an actual biological difference.

Real-time PCR has also proven effective for the rapid identification of anthrax. The method that has been used is a 5′ nuclease assay, which exhibits increased sensitivity compared to a traditional PCR method (Fig. 5). In addition to primers, 5′ nuclease assays use a fluorescently labeled probe that is complementary to the target sequence. This probe is labeled with a reporter fluorophore as well as a quencher for this reporter. When the probe is added to the reaction, the quencher and reporter are in close proximity and little fluorescent signal is detectable. However, as the target sequence is amplified and the probe binds to it, the 5′ to 3′ exonuclease activity of the Taq polymerase in the reaction degrades the probe, releasing the reporter fluorophore from its quencher and leading to an increase in the fluorescent signal. In the case of *B. anthracis*, three primer and probe sets were used; one each to detect pXO1, pXO2, and the *B. anthracis* chromosome. All fully virulent *B. anthracis* strains tested were positive for all three targets, yielding 100% sensitivity for virulent bacteria. Further, none of the non-*Bacillus. anthracis* strains, including *B. cereus* and *B. thuringiensis*, tested positive for all three targets, giving 100% specificity, as well. The lower limit of detection with this assay is estimated to be 1 pg of DNA or approximately 5 to 10 spores. *B. anthracis* strains lacking one or the other virulence plasmid

will obviously not test positive in all three reactions, and additional methods will need to be used to confirm their identity.

Some anthrax detection methods look for actual bacteria rather than for DNA. One such method is a *fluorescent antibody assay* that has also been validated for rapid identification of *B. anthracis*. In this method, two fluorescently labeled monoclonal antibodies, one against the cell wall polysaccharide antigen and one against the capsule antigen, are incubated with bacterial samples. The cells are then washed to remove unbound antibody and scored under a fluorescent microscope. Although other bacillus strains express a similar polysaccharide antigen, these bacteria lack the poly-D-glutamic acid capsule and will not score positive with both antibodies. The combined assay was 99% specific, with 227 of 230 *B. anthracis* isolates scoring positive, and 100% specific, with none of the 56 non-*Bacillus anthracis* strains scoring positive. It also could provide results within 3 to 6 h. Several clinical samples that were positive for *B. anthracis* by the real-time PCR assay were not detected by this fluorescent assay, so the latter assay may not be as sensitive. However, the antibody assay requires less lab space and sophisticated equipment and, therefore, may be useful in an outbreak situation, especially as a rapid confirmatory assay.

A method that may be useful for the detection of exposure to *B. anthracis* in clinical samples is an enzyme-linked immunosorbent assay (ELISA) that detects antibodies to PA. Only individuals who have been exposed to or vaccinated for anthrax will show positive results in this assay. In this ELISA, recombinant PA is bound to a microtiter plate. Serum from the test subject is then added to the plate and incubated to allow anti-PA antibodies from the serum to bind to the plate. Following extensive washes, an antihuman IgG antibody labeled with a reporter molecule is used to measure the amount of anti-PA antibody bound to the plate. The specificity of this assay can be enhanced through competitive inhibition, where test sera are incubated in the presence or absence of excess recombinant PA in solution before analysis in the standard ELISA. If the antibodies detected in the ELISA are truly specific for PA (and not just sticking to something in the assay nonspecifically), they should be removed from the assay by this competitive inhibition, and little antibody will bind to the microtiter plate. This assay will be used as part of a panel of tests for the confirmation of clinical human anthrax. This test should not be used as the sole confirmation of anthrax because it is known that, at least in nonhuman primates, early antibiotic treatment of an anthrax infection can abrogate the development of anti-PA antibodies.

3.2 How do You Tell One Strain from Another?

Not only is it important to be able to distinguish *B. anthracis* from related *Bacillus* species, it has also become important to distinguish between different strains of *B. anthracis*. This is important for tracking the history of an environmental outbreak and in the investigation of acts of terrorism that use *B. anthracis*. It has been difficult to develop methods to distinguish between *B. anthracis* strains because this is one of the most monomorphic species known – which means that it has very little genetic variation. Of 1221-AFLP fragments examined in 79 *B. anthracis* isolates in one study, 97% were monomorphic – no sequence variation was found between the

isolates. Looking at a specific gene important for virulence, Price et al. found only 5 single base pair changes in *pagA* between 26 *B. anthracis* strains, using complete DNA sequence data for the gene. Further, LR REP-PCR analysis of 105 *B. anthracis* strains of diverse geographical origins allowed the strains to be classified into only five groups. A few explanations for the monomorphic nature of *B. anthracis* have been proposed. It may be that anthrax has a slow rate of evolution and divergence, although this is thought to be unlikely. Other possible explanations include a very recent origin for this species, giving it little time to evolve and diverge, or a population bottleneck that reduced its genetic diversity in the recent past.

So far, the best screening method for distinguishing *B. anthracis* strains has been MLVA. In one study, MLVA of 426 *B. anthracis* isolates using 8 VNTR loci identified 89 distinct genotypes that could be studied in terms of their genetic relatedness. MLVA was used for molecular subtyping of *B. anthracis* in the 2001 anthrax outbreak. All 135 isolates from the outbreak had identical MLVA characteristics. Rather than being useful for distinguishing this strain from others, MLVA characterization turned out to be useful for the confirmation that each case was caused by the same *B. anthracis* strain and for separating these results from environmental contamination. During this investigation, it became clear that MLVA was not sufficient to determine the origin of the attack strain. On the basis of MLVA analysis, the strain of *B. anthracis* isolated in the attacks appeared to be identical to the Ames strain that is used by many research labs, so this technique did not provide the resolution necessary for the genetic investigation.

As the *B. anthracis* genome sequence was nearing completion at this time, research turned to comparing the DNA sequence of the *B. anthracis* strain isolated from a patient who contracted anthrax in Florida with that of the Porton isolate of the Ames strain, which was used to generate the standard *B. anthracis* genome sequence. Even at this level of resolution, only four differences were found between the chromosomal sequences of the Florida and Porton isolates. As for plasmid sequences, the Porton isolate has been cured of its plasmids, so the pXO1 and pXO2 sequences of the attack isolate were compared with those of the Sterne (pXO1) and Pasteur (pXO2) strains. A total of 38 single nucleotide polymorphisms, 8 VNTRs, 3 insertion/deletion polymorphisms, and 2 inversions were identified. One of the inversions was found on pXO1 and was centered on the pathogenicity island. In fact, the Florida isolate contained a mixture of both orientations of this plasmid segment, confirming the fact that this DNA segment can invert, as had been suggested previously.

The polymorphisms identified in this sequence comparison were then used in an attempt to distinguish five additional Ames isolates that previously had been indistinguishable. When plasmid sequences were compared, one laboratory isolate and a strain isolated from an infected goat did have a small number of plasmid sequence differences that allowed them to be classified separately. The polymorphisms identified in this study also enabled researchers to readily distinguish two non-Ames *B. anthracis* strains from the Ames strains. However, the remaining three Ames lab isolates and the Florida strain could not be distinguished from each other, on the basis of either the chromosomal or plasmid sequences.

Owing to the monomorphic nature of the *B. anthracis* genetic composition, direct sequence comparisons between isolates are the most effective method for distinguishing related strains. Even with this complete sequence comparison, it is not possible to distinguish between all isolates, as indicated by the difficulty in distinguishing between various isolates of the Ames strain.

4 Regulation of Virulence

As mentioned above, the major virulence factors for *B. anthracis* are a poly-D-glutamic acid capsule and the anthrax toxin. Synthesis of the capsule requires four genes: *capA, capB, capC,* and *dep (capD)*. The toxin, composed of protective antigen, edema factor, and lethal factor, is encoded by *pagA, cya,* and *lef* respectively. The genes encoding these factors are expressed in a regulated fashion, through the action of two major regulators, AtxA and AcpA, but additional regulatory factors are also involved. A figure summarizing the virulence regulation pathways in *B. anthracis,* as we currently understand them, is shown in Fig. 6.

4.1 AtxA

The toxin genes and *atxA* are expressed by vegetative cells shortly after spore germination in cultured macrophages. Expression of *atxA* at this time is important because it is essential for the release of the bacteria from macrophages. The *atxA* gene is located on pXO1 between the *cya* and *pagA* genes. In *atxA*$^+$ strains, the toxin genes are coexpressed, and this expression is enhanced by elevated CO_2 levels and temperature, with induction at

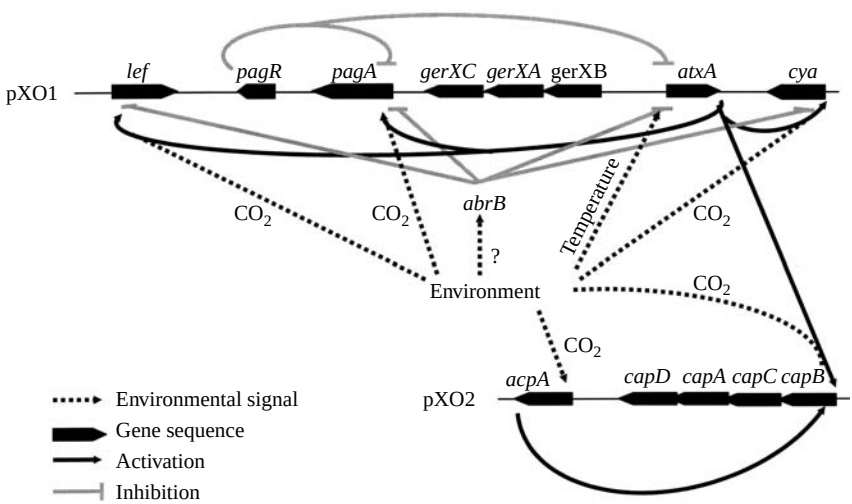

Fig. 6 Regulation of plasmid-encoded virulence gene expression. This figure is an overview of regulatory pathways known to be involved in *B. anthracis* virulence gene expression. Additional factors involved in the control of virulence gene expression will certainly be identified in the future. Although environmental signals are known to be crucial for regulation of virulence gene expression, the mechanism for this signaling is not known, nor have all of the factors involved in environmental sensing probably been identified.

5% CO_2 and 37 °C. *AtxA*$^-$ strains show significantly reduced production of all three toxin proteins in culture and are avirulent in a mouse model.

AtxA is an activator of toxin synthesis that acts at the transcriptional level. Although *pagA*, *lef*, and *cya* are transcribed as part of an AtxA-dependent regulon, there are no obvious sequence similarities upstream of the virulence genes that would suggest an AtxA binding site, nor does the AtxA protein sequence suggest any characteristics of DNA-binding proteins, so the mechanism it uses to regulate transcription of these genes is not clear.

The temperature requirement for toxin synthesis is at the level of the *atxA* transcript, as steady state levels of *atxA* mRNA are significantly increased in cells grown at 37 °C compared to cells grown at 28 °C. Higher levels of AtxA do not directly translate into increased levels of toxin production, however. Strains that carry *atxA* on a multicopy plasmid have high levels of AtxA, but they do not produce more, and some strains even produce significantly less, protective antigen than do matched strains carrying *atxA* in a single copy on pXO1. Further, although CO_2-enhanced toxin expression is AtxA-dependent, CO_2 does not affect AtxA levels. Although altering AtxA levels does not directly lead to an alteration of toxin levels, it can have a regulatory effect. It has been reported that some *B. anthracis* strains carrying a multicopy plasmid encoding *atxA* produce toxin in a CO_2-independent fashion in contrast to the CO_2-dependence seen in the parental strain.

The toxin genes are not the only *B. anthracis* genes that are AtxA-regulated. The presence of pXO1 also increases capsule synthesis. AtxA appears to be necessary and sufficient for this stimulation, as indicated by enhanced transcription of a *capB-lacZ* fusion in an AtxA-dependent manner. To identify additional AtxA-dependent gene expression, Hoffmaster and Koehler used a *lacZ*-containing transposon as a marker for genes that were induced by CO_2 in an AtxA-dependent manner. In this scheme, *lacZ,* which is a gene whose expression can be easily monitored, will be placed randomly under the control of the regulatory sequences of the gene into which the tranposon inserts. *lacZ* expression will therefore mark genes that are expressed under the conditions of interest – in this case, the presence of CO_2 and AtxA. In this experiment, there were 10 independent transposon insertions with AtxA-dependent expression that did not map to the toxin genes. Two-dimensional electrophoresis of protein extracts from *atxA*+ versus *atxA*−cells also indicates that synthesis of nontoxin proteins is influenced by AtxA. Some of these AtxA-regulated genes may play a role in the growth of *B. anthracis*, because *atxA*-null mutations cause *B. anthracis* to grow poorly on certain minimal media, and in sporulation, because an *atxA*-null mutant sporulates more efficiently than the parent strain when grown in rich media. Finally, *atxA*, but not the toxin genes, is required for release of *B. anthracis* from macrophages, so an *atxA*-regulated gene is likely to be involved in this process.

4.2 PagR

One of the CO_2-enhanced, *atxA*-dependent loci on pXO1 is a repressor of the *pagA* operon that serves to limit *pagA* expression under conditions optimal for PA synthesis. It was named *pagR* for protective antigen repressor. *pagR* is located downstream of

pagA and is cotranscribed with that gene. Transcription from the *pagA* promoter results in a monocistronic transcript containing *pagA* alone, and in a bicistronic transcript that encodes both *pagA* and *pagR*. Expression of *pagR*, as with *pagA*, is *atxA*-dependent and CO_2-enhanced. PagR has sequence similarity to transcription regulators, and it has been found that the level of the *pagA* transcript, and PA protein, is higher in a *pagR* mutant than in its parent strain, indicating that PagR normally represses *pagA* expression. Interestingly, PagR also represses *atxA* expression; the *pagR* mutant strain produces approximately twofold more *atxA* mRNA than its parent strain. It is not believed that the effects of *pagR* on *pagA* expression occur through AtxA because AtxA levels are not limiting for *pagA* expression.

Before *pagR* was discovered, a mutant *B. anthracis* strain was found that expressed the toxin genes in the absence of CO_2. An unidentified virulence regulator was therefore proposed. This strain was postulated to have a mutation in a negative regulator of toxin gene expression, whose repressive effect is alleviated in the presence of CO_2, thereby allowing AtxA to stimulate toxin expression. CO_2 has the opposite effect on *pagR*-regulated toxin expression – it would be predicted to decrease toxin gene expression because it stimulates *pagR* expression. *pagR*, therefore, cannot be the sole CO_2-dependent regulator of toxin gene expression because CO_2 has the overall effect of increasing toxin gene expression, rather than limiting it. Does CO_2 turn off the proposed repressor while at the same time increasing repression by PagR so that toxin gene expression does not get out of control? Unfortunately, several things about the CO_2 regulatory system are unknown, including the effects of PagR on expression of *lef* and *cya*, the effect of CO_2 on toxin expression in a *pagR* mutant, and the identity of the gene mutated in the strain that produces toxin in the absence of CO_2. As such, more research needs to be done before the regulatory pathways that limit anthrax toxin synthesis are fully understood.

4.3 AbrB

Another regulator of toxin expression is *abrB*, an ortholog to *abrB* from *Bacillus subtilis*. The *B. subtilis* gene is a transition state regulator, meaning it mediates the transition of cells from one state to another, such as from quiescence to active growth. The *B. anthracis* ortholog is located on the chromosome and appears to work similarly. An *abrB* null mutant shows increased expression of toxin gene–lacZ reporter fusions, with the largest effect seen for *pagA::lacZ*. Slightly elevated expression of an *atxA::lacZ* reporter was also found in the *abrB* null mutant, although, in the reciprocal experiment, *abrB* expression was found to be independent of *atxA*. Thus, similar to *pagR*, *abrB* appears to be a negative regulator of toxin gene and *atxA* expression. As with *pagR*, however, the *abrB*-mediated regulation of toxin gene expression is not believed to occur solely through *abrB*'s effects on *atxA* expression because AtxA levels in cells are not limiting for toxin gene expression.

In *B. subtilis*, *abrB* is regulated by Spo0A. The situation is similar in *B. anthracis*. Normally, low levels of AbrB are found in early log phase. These levels increase to a maximum level in log phase then drop back down to a low level as the cells enter stationary phase. In contrast, AbrB in an *spo0A* mutant persists at high levels in stationary phase, suggesting that *abrB* expression is normally negatively

regulated by *spo0A*. In conjunction with these findings, the pattern of toxin gene expression is affected by *abrB*. In a parent strain, toxin expression increases gradually throughout log phase, reaching a maximum in late log phase. The derivative *abrB* mutant strain, on the other hand, shows a sharp increase in toxin gene expression in early log phase, and maximum expression is reached in mid-log phase before a decline in expression is seen in late log phase. Although it has not yet been tested, it is likely that signaling pathways containing AbrB play a role in environmental sensing in *B. anthracis,* as they do in *B. subtilis*.

4.4 AcpA

Another major virulence regulator in *B. anthracis* is AcpA, a protein with significant similarity to AtxA. AcpA is encoded by a gene on pXO2 and is a regulator of capsule production that works at a transcriptional level. Northern blots indicate that the genes for capsule production, *capA* to *capD*, are transcribed as one message and thus constitute an operon. Strains with mutations in *acpA* are significantly reduced in capsule production and in virulence. AcpA is involved in the bicarbonate-mediated regulation of capsule production. Unlike AtxA, whose production requires a temperature of 37 °C, *acpA* transcription is regulated by bicarbonate.

Green et al. reported that the requirement for CO_2 in capsule production increases in the absence of pXO1. And, as mentioned above, the presence of pXO1 increases capsule synthesis in an AtxA-dependent manner. This led Uchida et al. to speculate that AtxA may work with AcpA to induce capsule production. Indeed, they found that an *acpA*-null mutant is complemented, not only by *acpA* but also by *atxA* in a CO_2-dependent fashion. This regulation is at the level of transcription, and the same sequences upstream of *cap* are required for both AtxA- and AcpA-mediated activation of transcription. Reciprocal experiments suggest that AcpA cannot complement an *atxA* defective strain. No obvious similarities between the sequences upstream of *pagA* and *cap* have been identified, so an AtxA target sequence is yet to be identified.

4.5 Other Potential Virulence Factors

Homologs of virulence genes from closely related species have been identified in *B. anthracis,* but their roles in virulence have not been fully studied. These include extracellular proteases, hemolysins, enterotoxins, and phospholipases. In *B. cereus* and *B. thuringiensis,* the expression of some of these nonspecific virulence factors is regulated by PlcR, a pleiotropic activator. Strains of *B. cereus* and *B. thuringiensis* with *plcR* deletions have attenuated virulence in a mouse model, indicating that the proteins turned on by this regulator do, in fact, play a role in virulence. In contrast, wild-type, fully virulent *B. anthracis* has a nonsense mutation in *plcR* and encodes a truncated, nonfunctional protein. This mutation silences a regulon; if a functional *plcR* is expressed in *B. anthracis,* a number of genes are activated. The addition of a functional *plcR* does not, however, affect the virulence of the organism, but it does inhibit sporulation of *B. anthracis* in an *atxA*-dependent manner. A strain that carries a functional *plcR* but no functional *atxA* sporulates normally. It has been postulated, therefore, that, after the acquisition of pXO1, there may have been counterselection for the

plcR regulon in *B. anthracis* because it puts the bacteria at a selective disadvantage due to its effect on sporulation.

5 Regulation of Sporulation/Germination

Appropriate sporulation and germination responses are crucial for the survival of *B. anthracis*. Spores can survive for years and can withstand harsh environmental conditions. The spores must be able to sense and respond to the environment, such that germination occurs under the appropriate conditions. In the case of *B. anthracis*, spores must germinate once they are taken up by a host so that the bacteria can replicate and the disease can be established. A tricistronic operon on pXO1 is responsible for spore germination. The genes in this operon, *gerXB*, *gerXA*, and *gerXC*, are located between *pagA* and *atxA* on the plasmid. The *ger* genes have significant similarity to the germination genes found in other bacilli, such as the *gerB* operon of *B. subtilis* and the *gerI* operon of *B. cereus*. In culture, expression of this operon is barely detectable during exponential growth and increases rapidly during stationary phase so that maximal expression is seen 3 h after the initiation of sporulation. Expression is limited to the developing forespore. In macrophages, spores of a *gerX* null mutant do not germinate as readily as do wild-type spores, and this is also true in mice infected subcutaneously with the *gerX* null strain. In the animals, there is a significant delay in the onset of germination with the *gerX* mutant, and this is associated with a fivefold decrease in virulence. Once the gerX null spores have germinated, they are able to multiply and invade the host. It has been speculated that the rate of germination may affect the rate of the infection and that this may play a role in the reduced virulence of this strain.

The *gerX* operon is not the only regulator of sporulation and germination in *B. anthracis*; plasmidless strains of *B. anthracis* germinate and grow in animal models. Inosine and L-alanine are two of the major germinants that induce plasmidless strains to germinate. At physiological levels, these germinants work only in combination with other compounds. For alanine, an aromatic amino acid (either L-histidine, L-tyrosine, or L-tryptophan) is also required for germination, and this has been termed the *aromatic-enhanced alanine* response (AEA). For inosine, either a nonaromatic amino acid or an aromatic amino acid are required as cogerminants, and these exhibit different germination kinetics. These responses have been called AAID for amino acid – and inosine-dependent responses. A *gerX* homolog was found on the chromosome of *B. anthracis* and called *gerS*. Characterization of a *gerS* mutant indicated that it is involved in both the AAID and AEA germination responses. Analogous to the situation in even more well-characterized germination pathways in other bacilli, it has been proposed that *gerS* encodes an aromatic responsive element that works in conjunction with other sensors to detect germinants with multiple components. Homologs of *gerI* and *gerL* from the *B. cereus* germination operon have been identified in the *B. anthracis* genome, and they may be components of this pathway.

6 Conclusions

Great strides have been made recently in the development of methods for the

detection of *B. anthracis*. These methods will facilitate the investigation of anthrax outbreaks in the future. Although research yields ever-increasing knowledge of the molecular biology of *B. anthracis*, much more work remains to be done to achieve a full understanding of this organism's life cycle, how it evolved into such a deadly pathogen, and how its virulence is controlled. As has been outlined in this review, *B. anthracis* shows evidence of gene functions that are encoded by unknown genes and of genes that encode proteins of unknown function. Further research on *B. anthracis*, which should be facilitated by the completed genome sequence, will hopefully make it possible to connect additional genes with their functions and result in a deeper understanding of the life cycle and pathogenesis of *B. anthracis*.

See also Autoantibodies and Autoimmunity; Genetic Engineering of Vaccines; Genetics, Molecular Basis of; Infectious Disease Testing by LCR

Bibliography

Books and Reviews

Baillie, L., Read, T.D. (2001) Bacillus anthracis, a bug with attitude! *Curr. Opin. Microbiol.* **4**, 78–81.

Mourez, M., Lacy, D.B., Cunningham, K., Legmann, R., Sellman, B.R., Mogridge, J., Collier, R.J. (2002) 2001: a year of major advances in anthrax toxin research, *Trends Microbiol.* **10**, 287–293.

Thorne, C.B. (1993) *Bacillus anthracis, Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*, American Society of Microbiology, Washington, DC, pp. 113–124.

Bioterrorism-related anthrax, *Emerg. Infect. Dis.* **8**(10), 1013–1182 (2002). This is a whole issue devoted to the anthrax outbreak of 2001.

Primary Literature

Anderson, G.L., Simchock J.M., Wilson, K.H. (1996) Identification of a region of genetic variability among *Bacillus anthracis* strains and related species, *J. Bacteriol.* **178**, 377–384.

Ash, C., Collins, M.D. (1992) Comparative analysis of 23S ribosomal RNA gene sequences of *Bacillus anthracis* and emetic *Bacillus cereus* determined by PCR-direct sequencing, *FEMS Microbiol. Lett.* **94**, 75–80.

Ash, C., Farrow, J.A.E., Dorsch, M., Stackebrandt, E., Collins, M.D. (1991) Comparative analysis of *Bacillus anthracis, Bacillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S rRNA, *Int. J. Syst. Bacteriol.* **41**, 343–346.

Battisti, L., Green, B.D., Thorne, C.B. (1985) Mating system for transfer of plasmids among *Bacillus anthracis, Bacillus cereus*, and *Bacillus thuringiensis*, *J. Bacteriol.* **162**, 543–550.

Beauregard, K.E., Collier, R.J., Swanson, J.A. (2000) Proteolytic activation of receptor-bound anthrax protective antigen on macrophages promotes its internalization, *Cell. Microbiol.* **2**, 251–258.

Bradley, K.A., Mogridge, J., Mourez, M., Collier, R.J., Young, J.A. (2001) Identification of the cellular receptor for anthrax toxin, *Nature* **414**, 225–229.

Brumlik, M.J., Szymajda, U., Zakowska, D., Liang, X., Redkar, R.J., Patra, G., Del Vecchio, V.G. (2001) Use of long-range repetitive element polymorphism-PCR to differentiate *Bacillus anthracis* strains, *Appl. Environ. Microbiol.* **67**, 3021–3028.

Carter, K.C. (1988) The Koch-Pasteur dispute on establishing the cause of anthrax, *Bull. Hist. Med.* **62**, 42–57.

Christopher, G.W., Cieslak, T.J., Pavlin, J.A., Eitzen, E.M. (1997) Biological warfare. A historical perspective, *JAMA* **278**, 412–417.

Dai, Z., Koehler, T.M. (1997) Regulation of anthrax toxin activator gene (*atxA*) expression in *Bacillus anthracis*: temperature, not CO_2/bicarbonate, affects AtxA synthesis, *Infect. Immun.* **65**, 2576–2582.

Dai, Z., Sirard, J.-C., Mock, M., Koehler, T.M. (1995) The *atxA* gene product activates transcription of the anthrax toxin genes and is essential for virulence, *Mol. Microbiol.* **16**, 1171–1181.

Dixon, T.C., Fadl, A.A., Koehler, T.M., Swanson, J.A., Hanna, P.C. (2000) Early *Bacillus anthracis*-macrophage interactions: intracellular survival and escape, *Cell. Microbiol.* **2**, 453–463.

Duesbery, N.S., Webb, C.P., Leppla, S.H., Gordon, V.M., Klimpel, K.R., Copeland, T.D., Ahn, N.G., Oskarsson, M.K., Fukasawa, K., Paull, K.D., Vande Woude, G.F. (1998) Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor, *Science* **280**, 734–737.

Friedlander, A.M. (1986) Macrophages are sensitive to anthrax lethal toxin through an acid-dependent process, *J. Biol. Chem.* **261**, 7123–7126.

Gordon, V.M., Leppla, S.H., Hewlett, E.L. (1988) Inhibitors of receptor-mediated endocytosis block the entry of *Bacillus anthracis* adenylate cyclase toxin but not that of *Bordetella pertussis* adenylate cyclase toxin, *Infect. Immun.* **56**, 1066–1069.

Green, B.D., Battisti, L., Koehler, T.M., Thorne, C.B., Ivins, B.E. (1985) Demonstration of a capsule plasmid in *Bacillus anthracis*, *Infect. Immun.* **49**, 291–297.

Guidi-Rontani, C., Levy, M., Ohayon, H., Mock, M. (2001) Fate of germinated *Bacillus anthracis* spores in primary murine macrophages, *Mol. Microbiol.* **42**, 931–938.

Guidi-Rontani, C., Pereira, Y., Ruffie, S., Sirard, J.-C., Weber-Levy, M., Mock, M. (1999) Identification and characterization of a germination operon on the virulence plasmid pXO1 of *Bacillus anthracis*, *Mol. Microbiol.* **33**, 407–414.

Guidi-Rontani, C., Weber-Levy, M., Labruyère, E., Mock, M. (1999) Germination of *Bacillus anthracis* spores within alveolar macrophages, *Mol. Microbiol.* **31**, 9–17.

Guignot, J., Mock, M., Fouet, A. (1997) AtxA activates the transcription of genes harbored by both *Bacillus anthracis* virulence plasmids, *FEMS Microbiol. Lett.* **147**, 203–207.

Hanna, P.C. (1998) Anthrax pathogenesis and host response, *Curr. Top. Microbiol. Immunol.* **225**, 13–35.

Helgason, E., Økstad, O.A., Caugant, D.A., Johansen, H.A., Fouet, A., Mock, M., Hegna, I., Kolstø, A.-B. (2000) Bacillus anthracis, Bacillus cereus, and Bacillus thuringiensis – one species on the basis of genetic evidence, *Appl. Environ. Microbiol.* **66**, 2627–2630.

Hoffmaster, A.R., Koehler, T.M. (1997) The anthrax toxin activator gene *atxA* is associated with CO_2-enhanced non-toxin gene expression in *Bacillus anthracis*, *Infect. Immun.* **65**, 3091–3099.

Hoffmaster, A.R., Koehler, T.M. (1999) Autogenous regulation of the *Bacillus anthracis pag* operon, *J. Bacteriol.* **181**, 4485–4492.

Ireland, J.A.W., Hanna, P.C. (2002) Amino acid- and purine ribonucleoside-induced germination of *Bacillus anthracis* ΔSterne endospores: *gerS* mediates responses to aromatic ring structures, *J. Bacteriol.* **184**, 1296–1303.

Jackson, P.J., Hill, K.K., Laker, M.T., Ticknor, L.O., Keim, P. (1999) Genetic comparison of *Bacillus anthracis* and its close relatives using amplified fragment length polymorphism and polymerase chain reaction analysis, *J. Appl. Microbiol.* **87**, 263–269.

Jackson, P.J., Walthers, E.A., Kalif, A.S., Richmond, K.L., Adair, D.M., Hill, K.K., Kuske, C.R., Andersen, G.L., Wilson, K.H., Hugh-Jones, M.E., Keim, P. (1997) Characterization of the variable-number tandem repeats in *vrrA* from different *Bacillus anthracis* isolates, *Appl. Environ. Microbiol.* **63**, 1400–1405.

Jernigan, J.A., Stephens, D.S., Ashford, D.A., Omenaca, C., Topiel, M.S., Galbraith, M., Tapper, M., Fisk, T.L., Zaki, S., Popovic, T., Meyer, R.F., Quinn, C.P., Harper, S.A., Fridkin, S.K., Sejvar, J.J., Shepard, C.W., McConnell, M., Guarner, J., Shieh, W.-J., Malecki, J.M., Gerberding, J.L., Hughes, J.M., Perkins, B.A., Members of the Anthrax Bioterrorism Investigation Team (2001) Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States, *Emerg. Infect. Dis.* **7**, 933–944.

Keim, P., Kalif, A., Schupp, J., Hill, K., Travis, S.E., Richmond, K., Adair, D.M., Hugh-Jones, M., Kuske, C.R., Jackson, P. (1997) Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers, *J. Bacteriol.* **179**, 818–824.

Keim, P., Klevytska, A.M., Price, L.B., Schupp, J.M., Zinser, G., Smith, K.L., Hugh-Jones, M.E., Okinaka, R., Hill, K.K., Jackson, P.J.

(1999) Molecular diversity in *Bacillus anthracis*, *J. Appl. Microbiol.* **87**, 215–217.

Keim, P., Price, L.B., Klevytska, A.M., Smith, K.L., Schupp, J.M., Okinaka, R., Jackson, P.J., Hugh-Jones, M.E. (2000) Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*, *J. Bacteriol.* **182**, 2928–2936.

Klimpel, K.R., Molloy, S.S., Thomas, G., Leppla, S.H. (1992) Anthrax toxin protective antigen is activated by a cell surface protease with the sequence specificity and catalytic properties of furin, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10277–10281.

Koehler, T.M., Dai, Z., Kaufman-Yarbray, M. (1994) Regulation of the *Bacillus anthracis* protective antigen gene: CO_2 and a *trans*-acting element activate transcription from one of two promoters, *J. Bacteriol.* **176**, 586–595.

Lee, C.A. (1996) Pathogenicity islands and the evolution of bacterial pathogens, *Infect. Agents Dis.* **5**, 1–7.

Leppla, S.H. (1982) Anthrax toxin edema factor: a bacterial adenylate cyclase that increases cyclic AMP concentrations of eukaryotic cells, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3162–3166.

Makino, S.-I., Uchida, I., Terakado, N., Sasakawa, C., Yoshikawa, M. (1989) Molecular characterization and protein analysis of the *cap* region, which is essential for encapsulation in *Bacillus anthracis*, *J. Bacteriol.* **171**, 722–730.

Mignot, T., Mock, M., Robichon, D., Landier, A., Lereclus, D., Fouet, A. (2001) The incompatibility between the PlcR- and AtxA-controlled regulons may have selected a nonsense mutation in *Bacillus anthracis*, *Mol. Microbiol.* **42**, 1189–1198.

Milne, J.C., Furlong, D., Hanna, P.C., Wall, J.S., Collier, R.J. (1994) Anthrax protective antigen forms oligomers during intoxication of mammalian cells, *J. Biol. Chem.* **269**, 20607–20612.

Okinaka, R.T., Cloud, K., Hampton, O., Hoffmaster, A.R., Hill, K.K., Keim, P., Koehler, T.M., Lamke, G., Kumano, S., Mahillon, J., Manter, D., Martinez, Y., Ricke, D., Svensson, R., Jackson, P.J. (1999) Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes, *J. Bacteriol.* **181**, 6509–6515.

Pannucci, J., Okinaka, R.T., Sabin, R., Kuske, C.R. (2002) *Bacillus anthracis* pXO1 plasmid sequence conservation among closely related bacterial species, *J. Bacteriol.* **184**, 134–141.

Park, J.M., Greten, F.R., Li, Z.-W., Karin, M. (2002) Macrophages apoptosis by anthrax lethal factor through p38 MAP kinase inhibition, *Science* **297**, 2048–2051.

Patra, G., Sylvestre, P., Ramisse, V., Thérasse, J., Guesdon, J.-L. (1996) Isolation of a specific chromosomic DNA sequence of *Bacillus anthracis* and its possible use in diagnosis, *FEMS Immunol. Med. Microbiol.* **15**, 223–231.

Price, L.B., Hugh-Jones, M., Jackson, P.J., Keim, P. (1999) Genetic diversity in the protective antigen gene of *Bacillus anthracis*, *J. Bacteriol.* **181**, 2358–2362.

Qi, Y., Patra, G., Liang, X., Williams, L.E., Rose, S., Redkar, R.J., Del Vecchio, V.G. (2001) Utilization of the *rpoB* gene as a specific chromosomal marker for real-time PCR detection of *Bacillus anthracis*, *Appl. Environ. Microbiol.* **67**, 3720–3727.

Ramisse, V., Patra, G., Vaissaire, J., Mock, M. (1999) The Ba813 chromosomal DNA sequence effectively traces the whole *Bacillus anthracis* community, *J. Appl. Microbiol.* **87**, 224–228.

Read, T.D., Salzberg, S.L., Pop, M., Shumway, M., Umayam, L., Jiang, L., Holtzapple, E., Busch, J.D., Smith, K.L., Schupp, J.M., Solomon, D., Keim, P., Fraser, C.M. (2002) Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*, *Science* **296**, 2028–2033.

Ruhfel, R.E., Robillard, N.J., Thorne, C.B. (1984) Interspecies transduction of plasmids among *Bacillus anthracis*, *B. cereus*, and *B. thuringiensis*, *J. Bacteriol.* **157**, 708–711.

Saile, E., Koehler, T.M. (2002) Control of anthrax toxin gene expression by the transition state regulator *abrB*, *J. Bacteriol.* **184**, 370–380.

Schmitt, W. (1982) Robert Koch; man and work, *Zentralbl. Chir.* **107**, 251–259.

Schupp, J.M., Klevytska, A.M., Zinser, G., Price, L.B., Keim, P. (2000) *vrrB*, a hypervariable open reading frame in *Bacillus anthracis*, *J. Bacteriol.* **182**, 3989–3997.

Singh, Y., Chaudhary, V.K., Leppla, S.H. (1989) A deleted variant of *Bacillus anthracis* protective antigen is non-toxic and blocks anthrax toxin action in vivo, *J. Biol. Chem.* **264**, 19103–19107.

Sirard, J.C., Mock, M., Fouet, A. (1994) The three *Bacillus anthracis* toxin genes are coordinately regulated by bicarbonate and temperature, *J. Bacteriol.* **176**, 5188–5192.

Turnbull, P.C.B. (1999) Definitive identification of *Bacillus anthracis* – a review, *J. Appl. Microbiol.* **87**, 237–240.

Uchida, I., Hornung, J.M., Thorne, C.B., Klimpel, K.R., Leppla, S.H. (1993) Cloning and characterization of a gene whose product is a *trans*-activator of anthrax toxin synthesis, *J. Bacteriol.* **175**, 5329–5338.

Uchida, I., Makino, S.-I., Sekizaki, T., Terakado, N. (1997) Cross-talk to the genes for *Bacillus anthracis* capsule synthesis by *atxA*, the gene encoding the *trans*-activator of anthrax toxin synthesis, *Mol. Microbiol.* **23**, 1229–1240.

Vietri, N.J., Marrero, R., Hoover, T.A., Welkos, S.L. (1995) Identification and characterization of a *trans*-activator involved in the regulation of encapsulation by *Bacillus anthracis*, *Gene* **152**, 1–9.

Vitale, G., Pellizzari, R., Recchi, C., Napolitani, G., Mock, M., Montecucco, C. (1998) Anthrax lethal factor cleaves the N-terminus of MAPKKs and induces tyrosine/threonine phosphorylation of MAPKs in cultured macrophages, *Biochem. Biophys. Res. Commun.* **248**, 706–711.

Watson, A., Keir, D. (1994) Information on which to base assessments of risk from environments contaminated with anthrax spores, *Epidemiol. Infect.* **113**, 479–490.

Wesche, J., Elliott, J.L., Falnes, P., Olsnes, S., Collier, R.J. (1998) Characterization of membrane translocation by anthrax protective antigen, *Biochemistry* **37**, 15737–15746.

Antibody Molecules, Genetic Engineering of

Manuel L. Penichet and Sherie L. Morrison
Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, CA, USA

1 Antibody Structure and Engineering 331
1.1 The Basic Structure of Antibodies 331
1.2 Classes and Subclasses of Antibodies 332

2 From Mouse to Human Antibodies 334
2.1 Murine Monoclonal Antibodies 334
2.2 Chimeric Antibodies 336
2.3 Humanized Antibodies 337
2.4 Human Monoclonal Antibodies in Mice 337

3 *In Vitro* Antibody Production by Phage Libraries 338

4 Further Genetic Modifications of Antibodies 340
4.1 Engineering Antibody Fragments: Monovalent, Bivalent, and Multivalent scFvs 340
4.2 Bispecific Antibodies 342
4.3 Polymers of Monomeric Antibodies 343
4.4 Antibody Fusion Proteins 344

5 Expression Systems 346

6 Conclusion 346

Bibliography 347
Books and Reviews 347
Primary Literature 347

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Keywords

Antibody/Antigen
Antibodies are proteins known as *immunoglobulins* (Igs), which are produced by the immune system in response to the presence of a foreign substance. Antigens are molecules (usually proteins) on the surface of cells, viruses, fungi, bacteria, and some nonliving substances such as toxins, chemicals, drugs, and foreign particles, which are recognized and specifically bound by antibodies. The immune system recognizes immunogenic antigens (also known as *immunogens*), and produces antibodies that destroy or neutralize substances or organisms containing antigens.

Antibody-dependent Cellular Cytotoxicity (ADCC)
Cell-killing reaction in which the Fc receptor bearing killer cells recognizes target cells via specific antibodies.

Bacteriophage
Bacteriophages or phages are viruses that infect bacteria. Bacteriophage λ is a temperate phage, which can grow lytically lysing the bacteria and forming a clear plaque on a lawn of bacteria. Infection with filamentous phages such as M13 is not lethal and the host bacteria do not lyse. Instead, their rate of growth slows and they form turbid plaques on the bacterial lawn.

Complement
Group of serum proteins participating in the lysis of foreign cells and pathogens [a process known as *complement-dependent cytotoxicity* (CDC)]; they also play an important role in phagocytosis.

Constant Region
Portion of the antibody molecule exhibiting little variation and determining the isotype (class or subclass) of the antibody.

Fab
The Fab fragment is a monovalent antigen-binding fragment of an antibody that consists of one light chain and part of one heavy chain (the variable region and the first constant region domain). It can be obtained by digestion of intact antibody with papain or by genetic engineering techniques.

Fc
The Fc is a nonantigen binding fragment of an antibody that consists of the carboxy-terminal portion of both heavy chains. It can be obtained by papain digestion of an intact antibody. Two Fabs and one Fc fragment comprise a complete IgG antibody.

Fv/scFv
The Fv is a monovalent antigen-binding fragment of an antibody composed of the variable regions from the heavy and light chains. scFv fragments are composed of the

variable domains of heavy and light chains (V_L and V_H) joined by a synthetic flexible linker peptide. Thus, the scFv provides a fully functional antigen binding domain that is encoded in a single gene and expressed as a single polypeptide.

Hybridoma
Cell derived by a fusion between a normal cell, usually a lymphocyte, and a tumor cell, usually a myeloma cell.

Variable Region
Variable portion of the antibody molecule that is responsible for antigen binding.

Antibodies have long been appreciated for their exquisite specificity. With the development of the hybridoma technology, it was possible to produce rodent (mouse or rat) monoclonal antibodies that are the product of a single clone of antibody-producing cells and have only one antigen-binding specificity. Advances in genetic engineering and expression systems have been applied to overcome problems of immunogenicity of rodent-produced antibodies and to improve their ability to trigger human immune effector mechanisms. The production of chimeric, humanized, and totally human antibodies as well as antibodies with novel structures and functional properties has resulted in improved monoclonal antibodies. As a consequence, recombinant antibody-based therapies are now used to treat a variety of diverse conditions that include infectious diseases, inflammatory disorders, and cancer. This article summarizes and compares different strategies for developing recombinant antibodies and their derivatives.

1 Antibody Structure and Engineering

1.1 The Basic Structure of Antibodies

Antibodies are molecules with multiple properties that make them a critical component of the immune system. These properties include the ability to recognize a vast array of different molecules known as *antigens* and to interact with and activate the host effector systems.

The basic structure of all antibodies, also known as *immunoglobulins* (Igs), is a unit consisting of two identical light polypeptide chains and two identical heavy polypeptide chains linked together by disulfide bonds (Fig. 1). Heavy and light chains are encoded by separate genes and are organized into discrete globular domains separated by short peptide segments. The amino-terminus end of both heavy and light chains is the antigen binding site and consists of one domain characterized by sequence variability (variable region or V) in both the heavy and light chains, called the V_H and V_L regions respectively. The rest of the molecule has a relatively constant (C) structure. The constant region of the light chain is termed

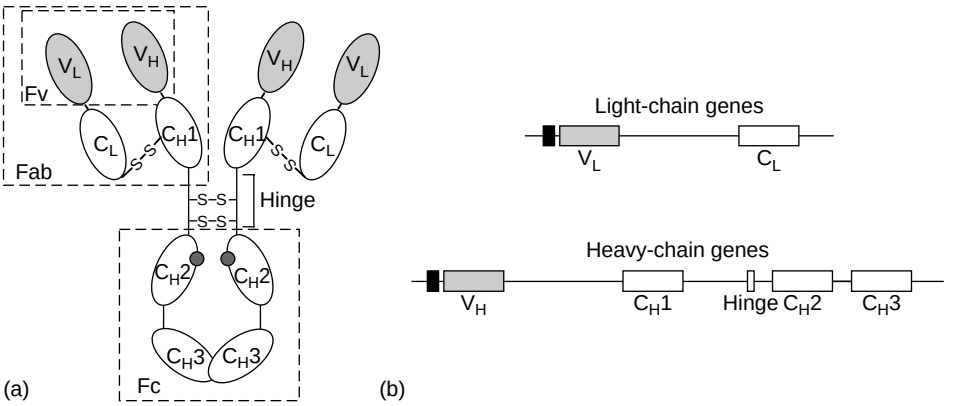

Fig. 1 (a) Diagram of an immunoglobulin G (IgG) molecule (the most abundant antibody in serum) and the active fragments that can be derived from it. The antibody molecule is divided into discrete functional domains: two domains constitute the light chain (V_L and C_L), while four domains make up the heavy chain (V_H, C_H1, C_H2, and C_H3). The variable region domains make the antibody binding site and are designated as the Fv region. The effector functions of the antibody are properties of the constant region domains. The carbohydrate units (black circles) present within the C_H2 domains contribute to the functional properties of the antibody. The hinge region provides flexibility to the antibody molecule, facilitating antigen binding and some effector functions. The enzyme papain cleaves the antibody into two Fab fragments containing the antigen binding sites and an Fc fragment responsible for the effector functions. (b) Genes that encode the heavy and light chains. In the genes, each domain is encoded by a discrete exon (indicated by boxes) separated by intervening sequences (introns) indicated by the line; the intervening sequences are present in the primary transcript but are removed from the mature mRNA by splicing. Both heavy and light chains contain hydrophobic leader sequences (indicated by the black exon) necessary for their secretion. This leader sequence is present in the newly synthesized heavy and light chains but is cleaved from them after they enter the endoplasmic reticulum and, therefore, is not present in the mature chains.

the C_L region. The constant region of the heavy chain is further divided into three structural domains stabilized by intrachain disulfide bonds: C_H1, C_H2, C_H3 (Fig. 1). The C_H3 domain of the heavy chain represents its carboxy-terminus. The domain structure of the antibodies is very important for genetic engineering because it facilitates protein engineering, allowing the exchange between molecules of functional domains carrying antigen-binding activities (Fabs or Fvs) or effector functions (Fc).

The hinge region, a segment of heavy chain between the C_H1 and C_H2 domains, provides flexibility in the molecule. Papain digestion of IgG yields two Fab fragments and one Fc fragment. The Fab region binds antigen, while the Fc region mediates effector functions such as complement activation, antibody-dependent cellular cytotoxicity (ADCC), and placental transmission. All antibodies are glycoproteins, and the carbohydrate present in the constant region has been shown to be essential for many of its effector functions.

1.2 Classes and Subclasses of Antibodies

The constant region of the heavy chain determines the class or isotype of the

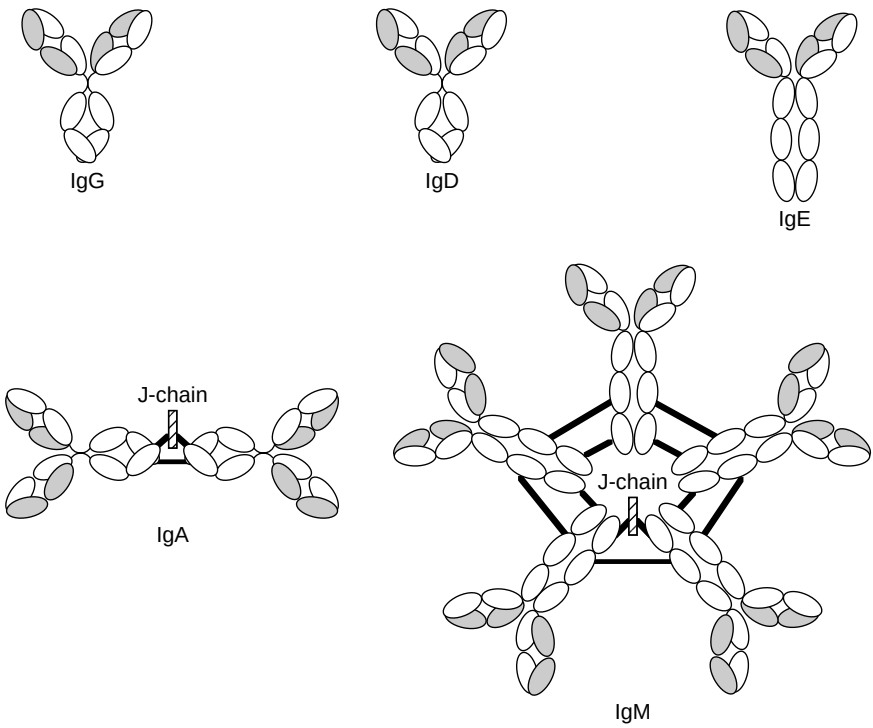

Fig. 2 The five classes of antibodies: IgG, IgD, and IgE are monomeric antibodies; IgA and IgM are polymeric antibodies.

antibody. Figure 2 shows the five different classes of antibodies, which differ in the constant region of their respective heavy chains. The different heavy-chain isotypes that define these classes of antibodies are designated by lower-case Greek letters: γ for IgG, δ for IgD, ε for IgE, α for IgA, and μ for IgM. IgM and IgE heavy chains contain an extra C_H domain (C_H4) and lack the hinge region found in IgG, IgD, and IgA. However, the absence of a hinge region does not imply that IgM and IgE lack flexibility; electron micrographs of IgM molecules binding to ligands have shown that the Fab arms can bend, relative to the Fc fragment. There are also two different light-chain isotypes, which are designated by the lower-case Greek letters κ and λ. Light chains of both isotypes are found associated with all the heavy-chain isotypes.

In humans, there is only one class of IgD, IgE, and IgM with monomeric structures having a molecular weight (MW) of around 190 kDa. However, human IgA antibodies can be further subdivided in two subclasses (IgA1 and IgA2) with a molecular weight of around 160 kDa for each monomer, and human IgG antibodies can be subdivided into four subclasses: IgG1, IgG2, IgG4 (all three with a molecular weight of around 150 kDa) and IgG3 with a molecular weight of around 165 kDa. The higher molecular weight exhibited by human IgG3 is due to the presence of an extended hinge region, which provides extraordinary flexibility. In fact, IgG3 is the most flexible human IgG.

Although all antibody molecules are constructed from the basic unit of two heavy and two light chains (H_2L_2), both IgA and IgM form polymers (Fig. 2). IgA forms a dimeric structure in which two H_2L_2 units are joined by a J-chain [$(H_2L_2)_2J$], and IgM forms either a pentameric structure with five H_2L_2 units joined by a J-chain [$(H_2L_2)_5J$] or a hexameric structure [$(H_2L_2)_6$], which does not have the J-chain (not shown). IgM and IgA heavy-chain constant regions contain a "tailpiece" of 18 amino acids that contains a cysteine residue essential for polymerization. The J-chain is a 15 kDa polypeptide produced by B-lymphocytes and plasma cells (the same cells that produce the antibodies) that promotes polymerization by linking to the cysteine of the tailpiece (Fig. 2). Differences in the isotype of the heavy chain determine the number of carbohydrate units and the ability to engage in various effector functions such as complement activation, ADCC, and placental transmission. Differences in the isotype of the light chain do not appear to significantly influence the structure or the effector functions of the antibody molecule.

2 From Mouse to Human Antibodies

2.1 Murine Monoclonal Antibodies

During the normal immune response, a wide variety of antibodies is produced. These antibodies, known as *polyclonal antibodies* (i.e. they are the product of many different antibody-producing cells), include antibodies with different variable regions as well as antibodies with the same variable regions associated with different constant regions. Rarely do different individuals mount an identical immune response. This heterogeneity in the immune response plus ethical and safety concerns has made it difficult to use polyclonal antibodies for many applications.

A significant breakthrough was made when it became possible to produce stable cell lines that synthesize a single homogeneous antibody (Fig. 3). By fusing a normal B-cell from the splenocytes of an immunized animal (initially a mouse or a rat) with a myeloma cell, it is possible to generate a "hybridoma," which possesses the immortality of the myeloma cell and secretes the antibody characteristic of the normal B-cell. Antibodies produced by hybridomas are monoclonal (i.e. they are the product of a single antibody-producing cell) and therefore have a single variable region associated with only one constant region. The immortality of the hybridoma ensures the continued availability of a well-characterized antibody. Once a hybridoma cell line is developed, it can be grown *in vitro* or *in vivo* for large-scale production of the monoclonal antibody, or it can be used to clone the variable regions of the monoclonal antibody for genetic engineering purposes.

Owing to their high affinity and exquisite specificity, murine monoclonal antibodies seemed to be the ideal "magic bullets" for diagnosis or therapy of multiple diseases including cancer. However, the progression of "magic bullets" from dream to reality has been slow because mouse (murine) monoclonal antibodies are not the ideal agents to be administered into a human. Murine monoclonal antibodies compared to human antibodies require more frequent dosing to maintain a therapeutic level of monoclonal antibodies because of a shorter circulating half-life in humans than human antibodies. In addition, the

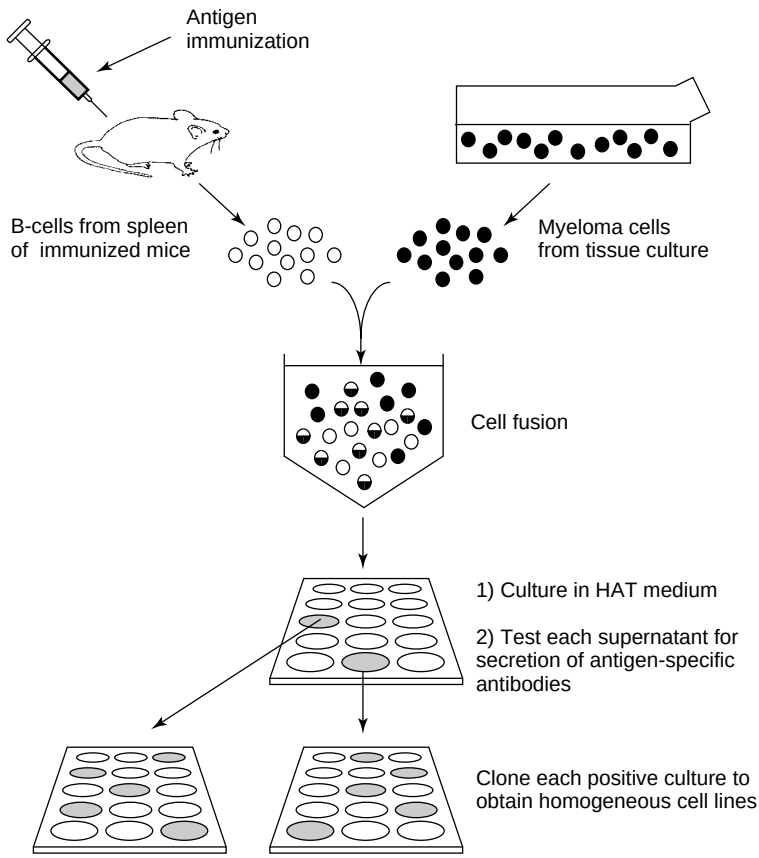

Fig. 3 Production of monoclonal antibodies. Mice are immunized with the antigen of interest and spleen cells are isolated and then fused with myeloma cells. HAT (hypoxanthine, aminopterin, and thymidine) medium is used to separate unfused myeloma cells from fused hybridoma cells. It takes advantage of the fact that normal mammalian cells can synthesize nucleotides by both a *de novo* and a salvage pathway, while the myeloma cells used have a defect in the salvage pathway. When the *de novo* pathway is blocked by aminopterin, cells must then utilize the enzymes of the salvage pathway. Thus, in the presence of aminopterin, unfused myeloma cells will die. Normal spleen cells do not grow. Only cells that have acquired the enzymes of the salvage pathway from the normal cells and the capacity for continuous growth from the myeloma cell will survive. Supernatants of hybridomas are screened for the presence of the antibody with the desired specificity, and positive cell cultures are subcloned to obtain a homogeneous cell line.

human immune system recognizes the mouse protein as foreign, generating a human antimouse antibody (HAMA) response, which results in an even more rapid clearance of the murine antibody (rendering the therapeutic useless) and, in some cases, in a severe allergic reaction. Moreover, murine constant regions can be ineffective in interacting with the human immune effector system. These

problems would be overcome by producing human monoclonal antibodies using human B-cells from immunized human donors. However, human monoclonal antibodies are difficult to produce using the hybridoma technology originally designed for the production of murine antibodies: the cell lines are unstable and frequently produce antibodies of the IgM isotype, which have low affinity for the antigen and are difficult to purify and handle. In addition, there are ethical and safety problems in obtaining humans immunized with certain antigens.

2.2 Chimeric Antibodies

To overcome the problems associated with the administration of murine monoclonal antibodies to humans, protein engineering has been used to convert murine monoclonal antibodies to mouse/human chimeric antibodies by genetically fusing the mouse variable regions to the human constant regions (Fig. 4), a procedure that is facilitated by the structure of antibodies.

Initially, variable region genes were obtained from genomic or cDNA libraries produced using DNA or mRNA from antibody-producing cells. More recently, cloning of genes encoding for the antibody variable region has been greatly facilitated by polymerase chain reaction (PCR)-based procedures. The variable region consists of hypervariable complementarity determining regions CDRs that are responsible for antibody specificity supported by relatively conserved framework regions. A limited number of different hydrophobic leader sequences are also found associated with the different variable regions. Therefore, it is possible to design sets of oligonucleotide primers on the basis of either the framework or the leader regions that will bind to virtually all mouse variable regions. A limited number of primers are required for all constant regions. Using upstream consensus primers for framework or leader regions in light- and heavy-chain variable regions and downstream primers for the constant regions, PCR can be used to amplify cDNAs generated by reverse-transcription (RT-PCR) directly from hybridoma mRNA.

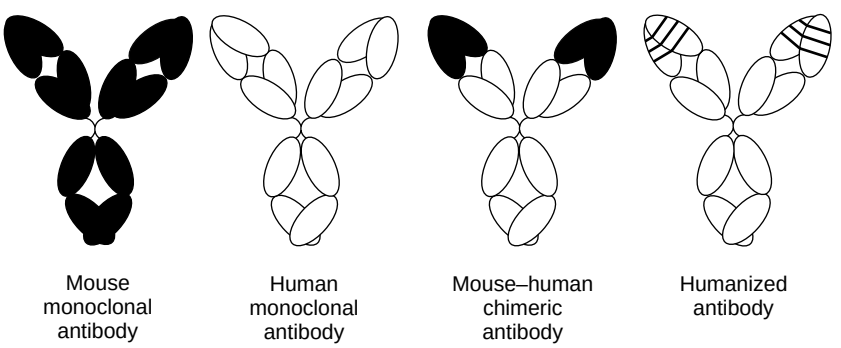

Fig. 4 Schematic representation of a murine monoclonal antibody, a human monoclonal antibody, a mouse–human chimeric antibody, and a CDR-grafted or humanized antibody. Chimeric antibodies have murine derived variable regions and binding specificities joined to human constant regions with their corresponding effector functions. Humanized antibodies are composed mostly of human sequences except for the areas in contact with the antigen (CDR regions), which are derived from mouse sequences.

This product is cloned and sequenced, and then used for the construction of chimeric antibodies or for further genetic modifications, as described below.

For the most part, chimeric antibodies retain their target specificity and show reduced HAMA responses. An example of a successful mouse–human chimeric antibody approved for clinical use is Rituximab (Rituxan/Mabthera), which targets the CD20 antigen and is now widely used to treat non-Hodgkin's lymphoma.

2.3 Humanized Antibodies

Although mouse–human chimeric antibodies are less immunogenic than mouse antibodies, in some cases they can still elicit a significant human antichimeric antibody (HACA) response. One approach to overcome this problem is to further manipulate the antibody variable region encoding for the antigen binding site resulting in humanized antibodies (Fig. 4). Each variable domain consists of a β-barrel with seven antiparallel β-strands connected by loops. Among the loops are the CDR regions. It is feasible to move the CDRs and their associated specificity from one scaffolding β-barrel to another, thereby creating "CDR-grafted" or "humanized" antibodies. However, it is rarely sufficient to move only the CDRs from a murine antibody onto a completely human framework because the resulting antibody frequently has reduced or no binding activity. In these cases, additional mouse residues near the CDRs are incorporated until binding is restored.

An example of a successful humanized monoclonal antibody that has been approved for clinical use is Trastuzumab (Herceptin), which has demonstrated significant antitumor activity in patients affected with breast cancer overexpressing the tumor-associated antigen HER2/*neu*.

2.4 Human Monoclonal Antibodies in Mice

Recently, mice have been produced that make antigen-specific antibodies that are totally human. To accomplish this goal, transgenic mice carrying portions of the human IgH and Igκ loci [in germ line configuration using megabase-sized YACs (yeast artificial chromosome)] were obtained that included the majority of the variable region repertoire, the genes for Cμ, Cδ and Cγ1, Cγ2, or Cγ4, as well as the *cis* elements required for their function. The IgH and Igκ transgenic was then bred into a genetic background deficient in the production of murine immunoglobulin. Therefore, the resulting mouse model, named *XenoMouse*, has elements of the human heavy- and light-chain loci in a murine context in which endogenous mouse heavy and light chains were disrupted.

The large and complex human variable region repertoire encoded on the immunoglobulin transgenes in XenoMouse strains support the development of large peripheral B-cell compartments and the generation of a diverse primary immune repertoire similar to that of adult humans. The human genes are compatible with mouse enzymes mediating class switching from IgM to IgG as well as somatic hypermutation and affinity maturation. Importantly, the immune system of the XenoMouse recognizes human antigens as foreign, with a concomitant strong human humoral immune response. The use of XenoMouse mice in conjunction with well-established hybridoma procedures (see Fig. 3) reproducibly results in human IgG monoclonal antibodies with

high affinity for human antigens and suitable for repeated administration to humans. To date, these engineered mice can produce only some of the human isotypes, but further genetic modification of these mice promises to expand their potential.

3 *In Vitro* Antibody Production by Phage Libraries

The production of monoclonal antibodies frequently requires several injections of antigen, which can take weeks to months. Hybridoma construction and screening require additional time. In addition, the immune response is often biased toward certain "immunodominant" epitopes of the antigen, making it difficult or impossible to produce monoclonal antibodies with the specificity desired for a particular use. Moreover, the production of antibodies against antigens conserved among species may be difficult or impossible. Phage display, a technology that allows the expression of immunoglobulin genes in bacteriophages (viruses that infect bacteria) without the need for developing hybridomas, provides an alternative approach that overcomes many of these problems.

In phage display, the heavy and light V-gene obtained from the spleen of an immunized mouse or from the peripheral blood of a naive or immunized donor are expressed as Fab or single-chain Fv (scFv) on the surface of filamentous phages (f1, M13, and fd). Phage libraries can be generated using variable antibody gene repertoires from any species including humans, or even synthetic sequences. Figure 5 shows the development of scFv libraries using a human donor.

Both scFv and Fab fragments can be expressed on the surface of filamentous bacteriophages (f1, M13, and fd) as either single or multiple copies of the antibody of interest depending on the phage protein used for fusion. Expression of multiple copies facilitates the identification and isolation of low-affinity antibodies. When functional antibody V domains are displayed on the surface of filamentous phages, the resulting phages bind specifically to antigen, and rare phages can be isolated on the basis of their ability to bind antigen. Multiple rounds of enrichment consisting of binding to immobilized antigen, expanding the bound phage, and further enriching by again binding to immobilized antigen can yield specific phages even if the desired specificity was present on less than 1 in 10^6 of the original phages in the library. The selected variable regions will generally have affinities similar to monoclonal antibodies and can be expressed as antibody fragments in the bacterium *Escherichia coli* or they can be used to produce complete antibodies and can be expressed in mammalian hosts. In addition, specific variable regions can be mutagenized and phages that express variable regions with increased affinity can be selected. The bacteriophage expression systems are designed to allow the genes encoding for heavy and light chains to undergo random combinations, which are tested for their ability to bind the desired antigen.

The decision of whether to produce scFv or Fab libraries depends partly on the intended use. The single gene format of the scFv is an advantage for the construction of fusion proteins such as "immunotoxins" (antibodies or antibody fragments fused to a toxin) or for targeted gene therapy approaches where the *scFv* gene is fused to a viral envelope protein gene. The use of scFv also appears to

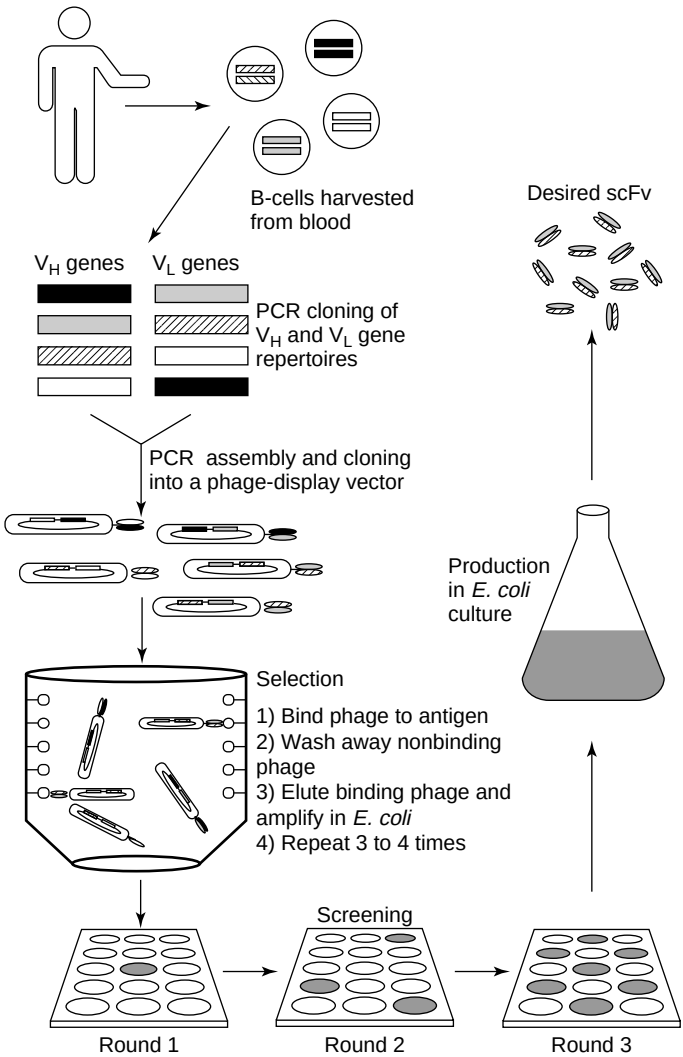

Fig. 5 *In vitro* human antibody production using phage libraries (phage-display technology). Peripheral blood mononuclear cells are harvested from a nonimmunized or immunized human donor, and the heavy- and light-chain V-gene regions (V_L and V_H) are amplified by PCR and assembled as a single-chain Fv region (scFv). Note that the original heavy- and light-chain pairings become scrambled during scFv assembly. The scFv genes are cloned into filamentous bacteriophages, where the encoded scFvs are displayed in a functional form on the phage surface. Multiple rounds of selection with a solid-phase antigen allow the isolation of even rare phages from the original library. The selected scFv can be expressed in *E. coli* (The illustration was adapted from Powers D.B. and Marks J.D. Monovalent Phage Display of Fab and scFv fusions. In *Antibody fusion proteins*. John Wiley & Son, Inc., New York, 1999).

be a better option for "intracellular immunization," a procedure in which the gene encoding the antibody binding site is delivered intracellularly to achieve phenotypic knockout. However, the use of Fab fragments offers the advantage that heavy- and light-chain libraries can be produced independently and reassorted. The use of two vectors also facilitates the construction of hierarchical libraries in which a fixed heavy or light chain is paired with a library of partners. In theory, Fab libraries may be preferred where the final product will be a complete antibody, since, in some instances, removal of the scFv linker might alter the antigen-binding properties. However, scFvs have been successfully converted into complete antibodies.

Phage display has proved to be very effective in generating a variety of antibodies that are difficult to obtain using the hybridoma technology. For preexisting mouse monoclonal antibodies, phage display can be used to obtain antibodies that are entirely human in sequence, but which bind to the same part of the antigen (epitope) as the mouse monoclonal antibodies. Antibodies to targets previously inaccessible using immunization approaches (for example, self-antigens, ubiquitous compounds, or toxic compounds) have been isolated by selection of phage on antigen *in vitro*. Building on the concept of phage display, alternative display strategies such as ribosome display and cell-surface display have been recently developed.

4 Further Genetic Modifications of Antibodies

4.1 Engineering Antibody Fragments: Monovalent, Bivalent, and Multivalent scFvs

As explained in Sect. 3, scFv fragments consist of the variable domains of heavy and light chains (V_L and V_H) genetically fused by a flexile linker peptide. scFvs are much smaller than intact antibodies (Fig. 6) (25–27 kDa vs around 150 kDa for intact IgG), yet can retain the specificity and the affinity of the parental molecule. The small size of scFvs is a tremendous

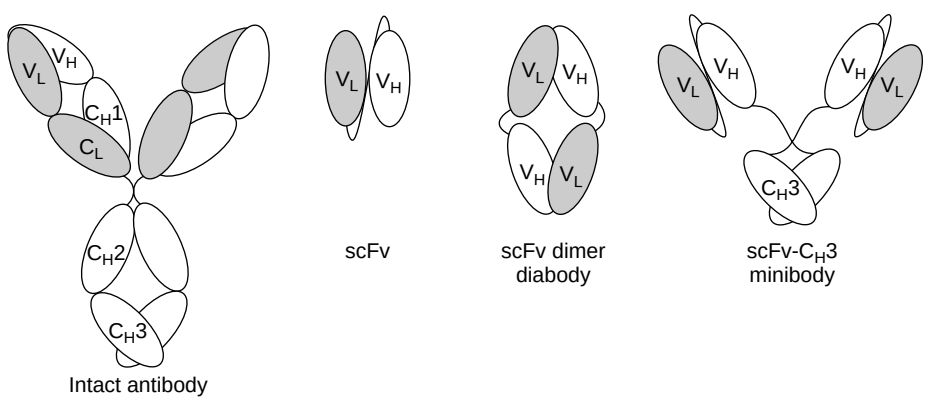

Fig. 6 Schematic representation showing the relative size and domain relationships between intact IgG (around 150 kDa) and engineered single-chain Fv (scFv, 25–27 kDa), scFv dimer or diabody (55 kDa), and minibody (scFv-C_H3, 80 kDa) fragments.

advantage in certain applications such as tumor targeting for detection and/or therapeutic purposes. In fact, scFv fragments display a combination of rapid, high-level tumor targeting with concomitant clearance from normal tissues and circulation that have made radiolabeled scFvs important tools for detection and treatment of cancer metastasis in both preclinical models and patients. An example of a successful scFv in the clinic is MFE-23. This scFv is specific for carcinoembryonic antigen (CEA), a glycoprotein that is highly expressed in colorectal adenocarcinomas. MFE-23, expressed in bacteria, has been used in two clinical trials: a gamma camera imaging trial using ^{123}I-MFE-23 and a radioimmunoguided surgery trial using ^{125}I-MFE-23, in which tumor deposits are detected by a handheld probe during surgery. Both trials showed that MFE-23 is safe and effective in localizing tumor deposits in patients with cancer.

Despite their successful use in some tumor-targeting studies, a significant limitation of using scFvs for targeting *in vivo* is their monovalent binding to antigen. Intact antibodies have significant avidity as a result of the presence of two antigen binding sites. To address this problem, antibody engineers have used the monomeric scFv as a building block for larger engineered fragments. One approach consists of producing dimers of scFv by incorporating a carboxy-terminal cysteine residue so that a disulfide bridge forms, yielding $(scFv)_2$ fragments. Alternately, the single-chain concept has been extended by using an additional linker peptide to join the two scFv molecules in tandem. Another approach for the production of scFv dimers results from the observation that the use of a very short linker peptide to connect the antibody variable regions caused the formation of "cross-paired" dimers, in which the V_L of one molecule associates with the V_H of a second and the V_L of the second molecule associates with the V_H of the first (Fig. 6). These noncovalent dimers, also known as *diabodies*, are capable of bivalent binding to antigen. scFv dimers and diabodies have a molecular weight similar to that of the antibody Fab fragments (55–60 kDa) but contain two antigen binding sites. Diabodies show significant improvement in tumor targeting compared to monovalent scFv. scFv fragments can also be fused to an immunoglobulin C_H3 domain, resulting in a self-assembling bivalent "minibody" (Fig. 6).

To further increase the avidity of antibody fragments, several laboratories have generated fragments with an increased valence. One strategy to develop fragments with increased valence has been to extend the diabody approach by decreasing the length of the interdomain linker peptide, which may result in the formation of tribodies and tetrabodies. It is also possible to obtain larger bivalent or multivalent fragments by fusing scFvs to protein domains normally involved in protein association such as helix bundles or leucine zippers. An alternative approach for producing multivalent fragments has been the fusion of scFvs to the bacterial protein streptavidin. Since streptavidin is a tetramer composed of four noncovalently linked monomers, four scFvs assemble to form a tetrameric structure with four antigen binding sites. A practical example of this technology was the development of Rh-specific scFv fused to streptavidin. This fusion protein named scFv::strep was able to directly agglutinate antigen-positive red blood cells (a reaction that is impossible to achieve by using antigen-specific monomeric scFv or IgG), suggesting the potential use of scFv::strep as a blood-typing reagent.

4.2 Bispecific Antibodies

In contrast to normal antibodies, which are monospecific, bispecific antibodies contain binding sites with two different specificities. Antibodies with two different specificities have been prepared by chemical modification to combine univalent fragments of different pepsin-treated antibodies, by fusing two hybridomas secreting antibodies of different specificities yielding a quadroma and by joining two different single-chain antibodies (scFvs). However, chemical modification of antibodies is inefficient and can lead to side reactions that damage the combining site. In a quadroma, it is difficult to separate the desired bispecific antibodies from the mixed population of heavy and light chains produced by the two hybridomas, and scFvs lacking constant regions also lack the antibody effector functions, which may be critical in certain applications. To address these problems, novel bispecific antibodies have been produced in which an scFv of one specificity has been genetically fused after the hinge (hinge-scFv) or at the carboxy-terminus (C_H3-scFv) of an antibody with a different specificity. Both fusion proteins were expressed by gene transfection in the context of a murine variable region. Transfectomas secreted a homogeneous population of the recombinant antibody with the two different specificities, one at the amino-terminus (anti-dextran) and the other at the carboxy-terminus (anti-dansyl). The C_H3-scFv antibody, which maintains the constant region of human IgG3, has some of the associated effector functions such as long half-life and Fc receptor binding. As expected, hinge-scFv antibody, which lacks the C_H2 and C_H3 domains, has no known effector functions.

Bispecific antibodies provide potential tools for use in immunotherapy. They take advantage of the great specificity of variable regions for their antigens and can be envisioned as transporters of therapeutic drugs or molecules or even immune effector cells to the specific targets identified by one of their binding sites. In fact, bispecific antibodies have been shown to be beneficial in the recruitment of immune cells for the treatment of cancer. Depending on the effector/target interaction, bispecific antibodies enhance cytotoxicity and phagocytosis. Several of these antibodies are currently undergoing evaluation in phase I and II clinical trials. Among the most extensively studied bispecific antibodies are 2B1 and MDX-210. 2B1 produced from a hybrid hybridoma is specific for the tumor-associated antigens HER2/*neu* and FcγRIII (CD16, the Fc gamma receptor expressed by key cytotoxic effector cells such as natural killer cells, neutrophils, and activated mononuclear phagocytes), while MDX-210 is a chemically conjugated hetero F(ab)′2 fragment specific for HER2/*neu* and FcγRI (CD64, the Fc gamma receptor expressed by key cytotoxic effector cells such as monocytes, macrophages, and IFN-γ–activated granulocytes). Both antibodies showed therapeutic promise in late-stage cancer patients whose tumor was refractory to conventional therapy. Since conventionally prepared bispecific antibodies have already shown promise in clinical trials and results from preclinical studies of recombinant bispecific antibodies are encouraging, in the future recombinant bispecific antibodies will undoubtedly be used in multiple clinical trials.

4.3 Polymers of Monomeric Antibodies

Although IgM is a naturally occurring polymeric antibody (see Fig. 2), for several applications it is desirable to have other classes of antibodies in an IgM-like (polymeric) structure. For example, IgM does not bind the Fc receptors on phagocyte cells, whereas IgG effectively triggers effector functions mediated through gamma-specific Fc receptors. In addition, since in IgG both heavy and light variable regions are usually somatically mutated (because of isotype switching to IgG), it is expected that polymeric IgG may be of higher intrinsic and functional affinity than the currently available IgM, resulting in more sensitivity and/or specificity. In addition, many methods such as protein-A binding are available that facilitate the isolation of IgG.

Vectors have been developed for the construction and expression of human polymeric IgG. It was observed that the 18–amino acid carboxyl-terminal tailpiece from human μ heavy chain is sufficient for polymer assembly. This finding was exploited to produce IgM-like polymers of IgG by fusing the 18–amino acid carboxyl-terminal tailpiece from human μ to the carboxyl-terminal of γ constant regions (Fig. 7). Using this technique, IgM-like polymers of IgG1, IgG2, IgG3, and IgG4 have been produced. IgGs obtained by this approach possess up to 6 Fcs and 12 antigen-combining sites, greatly increasing the avidity of their interactions with other molecules. These polymeric antibodies possess the Fcγ receptor–binding properties of IgG. Not surprisingly, the complement activity of normally active IgG1 and IgG3 and somewhat less-active IgG2 antibodies is dramatically enhanced upon polymerization. An unexpected result is that IgG4, normally devoid of complement activity, when polymerized in the same fashion directs complement-mediated lysis of target cells almost as effectively as the other polymers. These experiments demonstrate that polymerization of monomeric antibodies such as IgG is an effective approach to obtain antibodies with broader and more powerful effector functions than their wild-type counterparts.

An alternative strategy to make polymers of IgG is to genetically fuse chicken

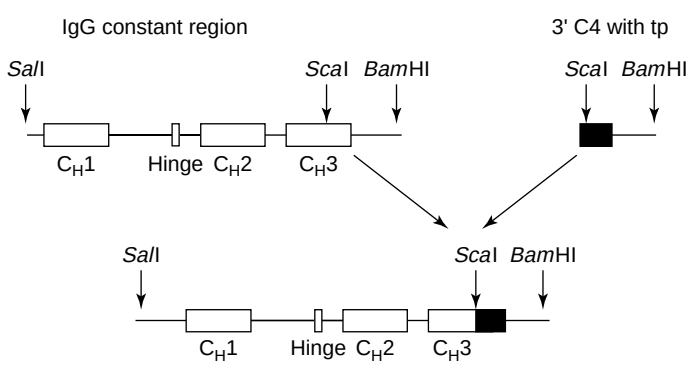

Fig. 7 Strategy for the construction of polymeric IgG. Using appropriate restriction sites, the μ tailpiece (μtp) of human IgM is genetically fused to the end of the heavy chain of human IgG.

avidin to the carboxy-terminus of the heavy chain of IgG. This approach is similar to that described in Sect. 4.1 (production of tetramers of scFv by fusing it with streptavidin) based on the fact that both streptavidin and avidin are tetramers of four noncovalently linked monomers. Since each antibody-avidin protein contains two molecules of avidin (one genetically fused at the carboxy-terminus of each heavy chain), two independent antibody fusion proteins bind to each other through their respective avidins forming a dimeric structure. One example of this approach is a recently developed human IgG3-avidin fusion protein specific for the transferrin receptor (TfR). The anti-TfR IgG3-avidin was able to function as a universal vector to deliver different biotinylated compounds into cancer cells overexpressing the TfR. Furthermore, it was unexpectedly discovered that anti-TfR IgG3-avidin, but not a recombinant anti-TfR IgG3 or a nonspecific IgG3-avidin, possesses a strong antiproliferative/proapoptotic activity against hematopoietic malignant cell lines. Studies confirmed that anti-TfR IgG3-avidin exists as a dimer, suggesting that cross-linking of the surface transferrin receptor may be responsible for the cytotoxic activity. These findings demonstrate that it is possible to transform an antibody specific for a growth factor receptor that does not exhibit inhibitory activity into a novel drug with significant intrinsic cytotoxic activity against selected cells by fusing it with avidin. The antitumor activity may be enhanced by delivering biotinylated therapeutics into cancer cells. Further development of this technology may lead to effective therapeutics for *in vivo* eradication of hematological malignancies and *ex vivo* purging of cancer cells in autologous transplantation.

4.4 Antibody Fusion Proteins

Fusion proteins with nonantibody molecules fused to antibodies can be produced using different approaches. Antibody fusion proteins that contain an intact antigen binding site should retain the ability to bind antigen, while the attached nonantibody partner should be able to exert its function. Such molecules, which have been called *immunoligands*, can be produced in several different ways (Fig. 8 a–f). When the nonantibody partner is fused to the end of the C_H3 domain (C_H3-ligand) (Fig. 8a), the antibody-combining specificity can be used to deliver an associated biological activity as well as antibody-related effector functions. An example is the anti-TfR IgG3-avidin fusion protein described in Sect. 4.3. Other examples are antibodies targeting cancer cells fused with interleukin-2 and GM-CSF. The goal of this approach to cancer therapy is to concentrate the cytokine in the tumor microenvironment and, by so doing, enhance the tumoricidal effect of the antibody and/or the host immune response against the tumor, while limiting severe toxic side effects associated with a high dose of cytokine administration. Such antibody–cytokine fusion proteins have shown significant antitumor activity in mice bearing tumors, leading to clinical trials. Immunoligands with the nonantibody partner fused immediately after the hinge (H-ligand) (Fig. 8b) or to the C_H1 domain (C_H1-ligand) (Fig. 8c) may be useful when the antibody-related effector functions are unnecessary or harmful. In addition, for many applications such as tumor targeting, the small size of the H-ligand and C_H1-ligand may be an advantage over the larger C_H3-ligand.

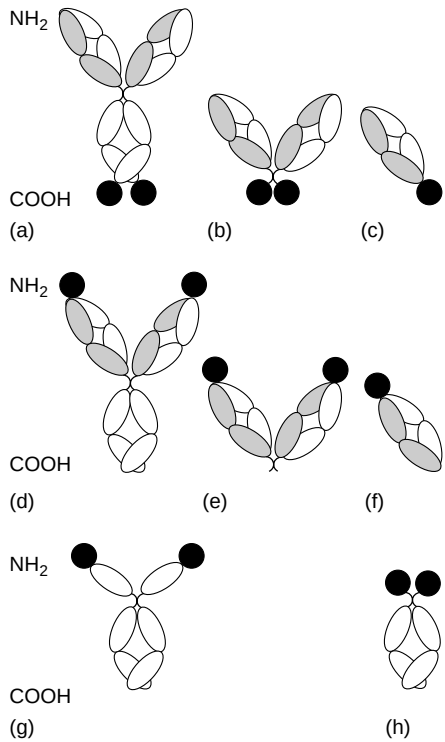

Fig. 8 Schematic representation of antibody fusion proteins. (a) to (c) represent different antibody fusion proteins in which the nonantibody partner was fused at the carboxy-terminus after the C_H3 domain (a), immediately after the hinge (b), or after the C_H1 domain (c). (d) to (f) represent antibody fusion proteins in which the nonantibody partner has been joined to the amino-terminus of the full-length heavy chain (d) or the truncated heavy chain (e and f). (g) and (h) represent two fusion proteins with the nonantibody partner fused to the amino-terminus of the C_H1 domain (g) or immediately before the hinge (h).

An alternative approach is to construct antibody fusion proteins with the ligand fused to the amino-terminus of the heavy chain (Fig. 8 d–f). This may be necessary for proteins that require N-terminal processing or folding for activity such as nerve growth factor (NGF), the costimulatory molecule B7.1, and interleukin-12 (IL-12). In fact, antibody-NGF, antibody-(B7.1), and antibody-(IL-12) fusion proteins containing the ligand fused to the amino-terminus of the antibody retain both the ability to bind antigen and the activity of the nonantibody partner.

Nonantibody sequences can also be used to replace the V_H domain or the V_H-C_H1 domains (Fig. 8 g and h). These molecules, which lack the ability to bind antigen, have been called *immunoadhesins* because they contain an adhesive molecule linked to the immunoglobulin Fc effector domains. In these proteins, the fused moiety acquires antibody-associated properties such as effector functions or improved pharmacokinetics. An example is the tumor necrosis factor (TNF) receptor IgG fusion protein, which binds to TNF (a mediator of inflammation) and neutralizes its activity. In fact, this molecule has been demonstrated to be efficacious for the treatment of rheumatoid arthritis.

Although Fig. 8 shows the nonantibody partner fused to the heavy chain, it should be appreciated that the nonantibody partner can also be fused to the light chain. It is also possible to construct antibody fusion proteins that combine more than one kind of nonantibody partner at the amino- or carboxy-termini of the heavy and/or light chains. In addition, a nonantibody

molecule can be fused to an scFv molecule as described in Sect. 4.1. Moreover, two antibody Fc fragments can be genetically fused resulting in a molecule with novel biological properties. An example of this approach is the protein GE2, which is the product of the fusion between the Fc fragments of human IgG and human IgE (Fcγ–Fcε fusion protein). GE2 does not have an Fab and, as a consequence, is unable to target an antigen. However, GE2 was able to form complexes with both FcγRII and FcεRI, resulting in an inhibition of mast cell and basophil activation that results in the blocking of the anaphylaxis in transgenic mice expressing human FcεRIα. This approach has therapeutic potential in IgE- and FcεRI-mediated diseases such as allergic asthma, allergic rhinitis, chronic urticaria, angioedema, and anaphylaxis.

5 Expression Systems

A large variety of expression systems have been used for the production of genetically engineered antibodies and antibody fragments. These expression systems include bacteria, yeast, plants, baculovirus, and mammalian cells.

Antibody fragments are commonly expressed in bacteria and yeast. The bacterium *E. coli* is a frequently used expression system owing to its rapid growth and easy genetic manipulation. However, proteins expressed in *E. coli* are frequently insoluble and/or inactive, and refolding may be required to obtain functional fragments. Secretion of fragments into the bacterial periplasm or culture supernatant provides an alternative means to obtain the desired functional fragments without the need for refolding. Although the results are highly antibody-dependent, there are many examples in which the bacterial expression system is successful.

Complete functional antibodies have been most successfully expressed in mammalian cells, as these cells possess the mechanisms required for correct immunoglobulin assembly, posttranslational modification (glycosylation), and secretion. Posttranslational modifications can influence the biologic properties and effector functions, important considerations especially when the antibody is to be used for therapy. Examples of mammalian cells that have been successfully used to express properly assembled and glycosylated antibodies and antibody fusion proteins are the mouse myeloma cell lines P3X63Ag8.653, Sp2/0-Ag14, and NS0/1. These three myeloma cell lines have lost the ability to produce endogenous H and L chains and are derived from the parent myeloma. Antibodies produced in nonlymphoid cell lines such as Chinese hamster ovary (CHO), HeLa, C6, and PC12 are also properly assembled and glycosylated. Owing to the slower growth of mammalian cells, mammalian expression requires a longer time frame and higher costs than bacterial or yeast expression, but it is preferred when complete functional antibodies with proper glycosylation and disulfide bonds are required. Other expression systems that have been extensively used to produce complete functional antibodies include insect cells and plants. There is no "universal" expression system – each system has its advantages and disadvantages.

6 Conclusion

Rapid progress has been made in producing genetically engineered

antibodies. The ability to express foreign DNA in a variety of host cells has made it possible to produce chimeric, humanized, and human antibodies as well as antibodies with novel structures and functional properties in quantities sufficiently large for many applications including clinical therapy. The available experience suggests that antibody-based therapies can be successfully developed for use in clinical situations in which no alternative effective therapy is available. However, continued progress in the development of antibody-based therapies will require extensive research to further define the mechanism of antibody action and on how to optimally use the novel proteins with unique functional properties.

See also Genetics, Molecular Basis of; Medicinal Chemistry.

Bibliography

Books and Reviews

Carter, P. (2001) Improving the efficacy of antibody-based cancer therapies, *Nat. Rev. Cancer* **1**, 118–29.

Chamow, S.M., Ashkenazi, A. (Eds.) (1999) *Antibody fusion proteins*, John Wiley & Son, New York.

Helguera, G.F., Morrison, S.L., Penichet, M.L. (2002) Antibody-cytokine fusion proteins: harnessing the combined power of cytokines and antibodies to cancer therapy, *Clin. Immunol.* **105**, 233–246.

Janeway, C.A., Travers, P., Walport, M., Shlomchik, M. (Eds.) (2001) The Generation of Lymphocyte Antigen Receptors, *Immunobiology: The Immune System in Health and Disease*, 5th edition, Garland Publisher, New York, pp. 123–154.

Kriangkum, J., Xu, B., Nagata, L.P., Fulton, R.E., Suresh, M.R. (2001) Bispecific and bifunctional single chain recombinant antibodies, *Biomol. Eng.* **18**, 31–40.

Rader, C., Barbas, C.F. (1997) Phage display of combinatorial antibody libraries, *Curr. Opin. Biotechnol.* **8**, 503–508.

Sensel, M.G., Coloma, M.J., Harvill, E.T., Shin, S.U., Smith, R.I., Morrison, S.L. (1997) Engineering novel antibody molecules, *Chem. Immunol.* **65**, 129–158.

Verma, R., Boleti, E., George, A.J. (1998) Antibody engineering: comparison of bacterial, yeast, insect and mammalian expression systems, *J. Immunol. Methods* **216**, 165–181.

Wu, A.M., Yazaki, P.J. (2000) Designer genes: recombinant antibody fragments for biological imaging, *Q. J. Nucl. Med.* **44**, 268–283.

Yoo, E.M., Chintalacharuvu, K.R., Penichet, M.L., Morrison, S.L. (2002) Myeloma expression system, *J. Immunol. Methods* **261**(1–2), 1–20.

Primary Literature

Abramowicz, D., Crusiaux, A., Goldman, M. (1992) Anaphylactic shock after retreatment with OKT3 monoclonal antibody, *N. Engl. J. Med.* **327**, 736.

Atwell, J.L., Breheney, K.A., Lawrence, L.J., McCoy, A.J., Kortt, A.A., Hudson, P.J. (1999) scFv multimers of the anti-neuraminidase antibody NC10: length of the linker between VH and VL domains dictates precisely the transition between diabodies and triabodies, *Protein Eng.* **12**, 597–604.

Bajorin, D.F., Chapman, P.B., Wong, G.Y., Cody, B.V., Cordon-Cardo, C., Dantes, L., Templeton, M.A., Scheinberg, D., Oettgen, H.F., Houghton, A.N. (1992) Treatment with high dose mouse monoclonal (anti-GD3) antibody R24 in patients with metastatic melanoma, *Melanoma Res.* **2**, 355–362.

Barbas, C.F., Bain, J.D., Hoekstra, D.M., Lerner, R.A. (1992) Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4457–4461.

Biocca, S., Pierandrei-Amaldi, P., Campioni, N., Cattaneo, A. (1994) Intracellular immunization with cytosolic recombinant antibodies, *Biotechnology (NY)* **12**, 396–399.

Boulianne, G.L., Hozumi, N., Shulman, M.J. (1984) Production of functional chimaeric mouse/human antibody, *Nature* **312**, 643–646.

Challita-Eid, P.M., Penichet, M.L., Shin, S.-U., Mosammaparast, N., Poles, T.M., Mahmood, K., Slamon, D.L., Morrison, S.L., Rosenblatt, J.D. (1997) A B7.1-Ab fusion protein retains antibody specificity and ability to activate via the T cell costimulatory pathway, *J. Immunol* **160**, 3419–3426.

Chester, K.A., Bhatia, J., Boxer, G., Cooke, S.P., Flynn, A.A., Huhalov, A., Mayer, A., Pedley, R.B., Robson, L., Sharma, S.K., Spencer, D.I., Regent, R.H. (2000) Clinical applications of phage-derived sFvs and sFv fusion proteins, *Dis. Markers* **16**, 53–62.

Chester, K.A., Mayer, A., Bhatia, J., Robson, L., Spencer, D.I., Cooke, S.P., Flynn, A.A., Sharma, S.K., Boxer, G., Pedley, R.B., Begent, R.H. (2000) Recombinant anti-carcinoembryonic antigen antibodies for targeting cancer, *Cancer Chemother. Pharmacol.* **46**, S8–S12.

Coloma, M.J., Hastings, A., Wims, L.A., Morrison, S.L. (1992) Novel vectors for the expression of antibody molecules using variable regions generated by polymerase chain reaction, *J. Immunol. Methods* **152**, 89–104.

Coloma, M.J., Morrison, S.L. (1997) Design and production of novel tetravalent bispecific antibodies [see comments], *Nat. Biotechnol.* **15**, 159–163.

Dangl, J.L., Wensel, T.G., Morrison, S.L., Stryer, L., Herzenberg, L.A., Oi, V.T. (1988) Segmental flexibility and complement fixation of genetically engineered chimeric human, rabbit and mouse antibodies, *EMBO J.* **7**, 1989–1994.

Davis, A.C., Roux, K.H., Pursey, J., Shulman, M.J. (1989) Intermolecular disulfide bonding in IgM: effects of replacing cysteine residues in the mu heavy chain, *EMBO J.* **8**, 2519–2526.

Davis, A.C., Roux, K.H., Shulman, M.J. (1988) On the structure of polymeric IgM., *Eur. J. Immunol.* **18**, 1001–1008.

Davis, A.C., Shulman, M.J. (1989) IgM–molecular requirements for its assembly and function, *Immunol. Today* **10**, 127, 128.

Dubel, S., Breitling, F., Fuchs, P., Zewe, M., Gotter, S., Welschof, M., Moldenhauer, G., Little, M. (1994) Isolation of IgG antibody Fv-DNA from various mouse and rat hybridoma cell lines using the polymerase chain reaction with a simple set of primers, *J. Immunol. Methods* **175**, 89–95.

Dubel, S., Breitling, F., Kontermann, R., Schmidt, T., Skerra, A., Little, M. (1995) Bifunctional and multimeric complexes of streptavidin fused to single chain antibodies (scFv), *J. Immunol. Methods* **178**, 201–209.

Ernst, M., Meier, D., Sonneborn, H.H. (1999) From IgG monoclonals to IgM-like molecules, *Hum. Antibodies* **9**, 165–170.

Figini, M., Marks, J.D., Winter, G., Griffiths, A.D. (1994) In vitro assembly of repertoires of antibody chains on the surface of phage by renaturation, *J. Mol. Biol.* **239**, 68–78.

George, A.J., Titus, J.A., Jost, C.R., Kurucz, I., Perez, P., Andrew, S.M., Nicholls, P.J., Huston, J.S., Segal, D.M. (1994) Redirection of T cell-mediated cytotoxicity by a recombinant single- chain Fv molecule, *J. Immunol.* **152**, 1802–1811.

Glennie, M.J., Brennand, D.M., Bryden, F., McBride, H.M., Stirpe, F., Worth, A.T., Stevenson, G.T. (1988) Bispecific F(ab' gamma)2 antibody for the delivery of saporin in the treatment of lymphoma, *J. Immunol.* **141**, 3662–3670.

Glennie, M.J., McBride, H.M., Worth, A.T., Stevenson, G.T. (1987) Preparation and performance of bispecific F(ab' gamma)2 antibody containing thioether-linked Fab' gamma fragments, *J. Immunol.* **139**, 2367–2375.

Goldenberg, M.M. (1999) Trastuzumab, a recombinant DNA-derived humanized monoclonal antibody, a novel agent for the treatment of metastatic breast cancer, *Clin. Ther.* **21**, 309–318.

Green, N.M. (1990) Avidin and streptavidin, *Methods Enzymol.* **184**, 51–67.

Green, L.L. (1999) Antibody engineering via genetic engineering of the mouse: XenoMouse strains are a vehicle for the facile generation of therapeutic human monoclonal antibodies, *J. Immunol. Methods* **231**, 11–23.

Griffiths, A.D., Malmqvist, M., Marks, J.D., Bye, J.M., Embleton, M.J., McCafferty, J., Baier, M., Holliger, K.P., Gorick, B.D., Hughes-Jones, N.C., Hoogboom, H., Winter, H. (1993) Human anti-self antibodies with high specificity from phage display libraries, *EMBO J.* **12**, 725–734.

Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H.R., Plückthun, A. (1998) Ribosome display efficiently selects and evolves high-affinity antibodies *in vitro* from immune libraries, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14130–14135.

He, M., Taussig, M.J. (1997) Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for *in vitro* display and evolution of antibody combining sites, *Nucleic Acids Res.* **25**, 5132–5134.

Higuchi, K., Araki, T., Matsuzaki, O., Sato, A., Kanno, K., Kitaguchi, N., Ito, H. (1997) Cell display library for gene cloning of variable regions of human antibodies to hepatitis B surface antigen, *J. Immunol. Methods* **202**, 193–204.

Holliger, P., Prospero, T., Winter, G. (1993) "Diabodies": small bivalent and bispecific antibody fragments, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6444–6448.

Hu, S., Shively, L., Raubitschek, A., Sherman, M., Williams, L.E., Wong, J.Y., Shively, J.E., Wu, A.M. (1996) Minibody: a novel engineered anti-carcinoembryonic antigen antibody fragment (single-chain Fv-CH3) which exhibits rapid, high-level targeting of xenografts, *Cancer Res.* **56**, 3055–3061.

Hudson, P.J., Kortt, A.A. (1999) High avidity scFv multimers; diabodies and triabodies, *J. Immunol. Methods* **231**, 177–189.

Iliades, P., Kortt, A.A., Hudson, P.J. (1997) Triabodies: single chain Fv fragments without a linker form trivalent trimers, *FEBS Lett.* **409**, 437–441.

Jaffers, G.J., Fuller, T.C., Cosimi, A.B., Russell, P.S., Winn, H.J., Colvin, R.B. (1986) Monoclonal antibody therapy. Anti-idiotypic and non-anti-idiotypic antibodies to OKT3 arising despite intense immunosuppression, *Transplantation* **41**, 572–578.

James, N.D., Atherton, P.J., Jones, J., Howie, A.J., Tchekmedyian, S., Curnow, R.T. (2001) A phase II study of the bispecific antibody MDX-H210 (anti-HER2 × CD64) with GM-CSF in HER2+ advanced prostate cancer, *Br. J. Cancer* **85**, 152–156.

Jespers, L.S., Roberts, A., Mahler, S.M., Winter, G., Hoogenboom, H.R. (1994) Guiding the selection of human antibodies from phage display repertoires to a single epitope of an antigen, *Biotechnology (NY)* **12**, 899–903.

Kohler, G., Milstein, C. (1975) Continuous cultures of fused cells secreting antibody of predefined specificity, *Nature* **256**, 495–497.

Kontsekova, E., Kolcunova, A., Kontsek, P. (1992) Quadroma-secreted bi(interferon alpha 2–peroxidase) specific antibody suitable for one-step immunoassay, *Hybridoma* **11**, 461–468.

Kortt, A.A., Lah, M., Oddie, G.W., Gruen, C.L., Burns, J.E., Pearce, L.A., Atwell, J.L., McCoy, A.J., Howlett, G.J., Metzger, D.W., Webster, R.G., Hudson, P.J. (1997) Single-chain Fv fragments of anti-neuraminidase antibody NC10 containing five- and ten-residue linkers form dimers and with zero- residue linker a trimer, *Protein Eng.* **10**, 423–433.

Kostelny, S.A., Cole, M.S., Tso, J.Y. (1992) Formation of a bispecific antibody by the use of leucine zippers, *J. Immunol.* **148**, 1547–1553.

Kuus-Reichel, K., Grauer, L.S., Karavodin, L.M., Knott, C., Krusemeier, M., Kay, N.E. (1994) Will immunogenicity limit the use, efficacy, and future development of therapeutic monoclonal antibodies? *Clin. Diagn. Lab. Immunol.* **1**, 365–372.

Leget, G.A., Czuczman, M.S. (1998) Use of rituximab, the new FDA-approved antibody, *Curr. Opin. Oncol.* **10**, 548–551.

Lewis, L.D., Cole, B.F., Wallace, P.K., Fisher, J.L., Waugh, M., Guyre, P.M., Fanger, M.W., Curnow, R.T., Kaufman, P.A., Ernstoff, M.S. (2001) Pharmacokinetic-pharmacodynamic relationships of the bispecific antibody MDX-H210 when administered in combination with interferon gamma: a multiple-dose phase-I study in patients with advanced cancer which overexpresses HER-2/neu, *J. Immunol. Methods* **248**, 149–165.

Liu, S.J., Sher, Y.P., Ting, C.C., Liao, K.W., Yu, C.P., Tao, M.H. (1998) Treatment of B-cell lymphoma with chimeric IgG and single-chain Fv antibody-interleukin-2 fusion proteins, *Blood* **92**, 2103–2112.

Lloyd, F. Jr., Goldrosen, M. (1991) The production of a bispecific anti-CEA, anti-hapten (4-amino-phthalate) hybrid-hybridoma, *J. Natl. Med. Assoc.* **83**, 901–904.

Mallender, W.D., Voss, E.W. Jr. (1994) Construction, expression, and activity of a bivalent bispecific single-chain antibody, *J. Biol. Chem.* **269**, 199–206.

Marasco, W.A. (1997) Intrabodies: turning the humoral immune system outside in for

intracellular immunization, *Gene Ther.* **4**, 11–15.

Marks, J.D., Hoogenboom, H.R., Bonnert, T.P., McCafferty, J., Griffiths, A.D., Winter, G. (1991) By-passing immunization. Human antibodies from V-gene libraries displayed on phage, *J. Mol. Biol.* **222**, 581–597.

McGrath, J.P., Cao, X., Schutz, A., Lynch, P., Ebendal, T., Coloma, M.J., Morrison, S.L., Putney, S.D. (1997) Bifunctional fusion between nerve growth factor and a transferrin receptor antibody, *J. Neurosci. Res.* **47**, 123–133.

Milstein, C., Cuello, A.C. (1983) Hybrid hybridomas and their use in immunohistochemistry, *Nature* **305**, 537–540.

Moreland, L.W., Baumgartner, S.W., Schiff, M.H., Tindall, E.A., Fleischmann, R.M., Weaver, A.L., Ettlinger, R.E., Cohen, S., Koopman, W.J., Mohler, K., Widmer, M.B., Blosch, C.M. (1997) Treatment of rheumatoid arthritis with a recombinant human tumor necrosis factor receptor (p75)-Fc fusion protein [see comments], *N. Engl. J. Med.* **337**, 141–147.

Moreland, L.W., Schiff, M.H., Baumgartner, S.W., Tindall, E.A., Fleischmann, R.M., Bulpitt, K.J., Weaver, A.L., Keystone, E.C., Furst, D.E., Mease, P.J., Ruderman, E.M., Horwitz, D.A., Arkfeld, D.G., Garrison, L., Burge, D.J., Blosch, C.M., Lange, M.L., McDonnell, N.D., Weinblatt, M.E. (1999) Etanercept therapy in rheumatoid arthritis. A randomized, controlled trial, *Ann. Intern. Med.* **130**, 478–486.

Morrison, S.L., Johnson, M.J., Herzenberg, L.A., Oi, V.T. (1984) Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6851–6855.

Ng, P.P., Dela Cruz, J.S., Sorour, D.N., Stinebaugh, J.M., Shin, S.U., Shin, D.S., Morrison, S.L., Penichet, M.L. (2002) An anti-transferrin receptor-avidin fusion protein exhibits both strong proapoptotic activity and the ability to deliver various molecules into cancer cells, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10706–10711.

Orlandi, R., Gussow, D.H., Jones, P.T., Winter, G. (1989) Cloning immunoglobulin variable domains for expression by the polymerase chain reaction, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 3833–3837.

Pack, P., Plückthun, A. (1992) Miniantibodies: use of amphipathic helices to produce functional, flexibly linked dimeric FV fragments with high avidity in *Escherichia coli*, *Biochemistry* **31**, 1579–1584.

Parmley, S.F., Smith, G.P. (1988) Antibody-selectable filamentous fd phage vectors: affinity purification of target genes, *Gene* **73**, 305–318.

Peng, L.S., Penichet, M.L., Morrison, S.L. (1999) A single-chain IL-12 IgG3 antibody fusion protein retains antibody specificity and IL-12 bioactivity and demonstrates antitumor activity, *J. Immunol.* **163**, 250–258.

Penichet, M.L., Dela Cruz, J.S., Shin, S.U., Morrison, S.L. (2001) A recombinant IgG3-(IL-2) fusion protein for the treatment of human HER2/neu expressing tumors, *Hum. Antibodies* **10**, 43–49.

Penichet, M.L., Harvill, E.T., Morrison, S.L. (1998) An IgG3-IL-2 fusion protein recognizing a murine B cell lymphoma exhibits effective tumor imaging and antitumor activity, *J. Interferon Cytokine Res.* **18**, 597–607.

Plückthun, A., Pack, P. (1997) New protein engineering approaches to multivalent and bispecific antibody fragments, *Immunotechnology* **3**, 83–105.

Raso, V., Griffin, T. (1981) Hybrid antibodies with dual specificity for the delivery of ricin to immunoglobulin-bearing target cells, *Cancer Res.* **41**, 2073–2078.

Russell, S.J., Hawkins, R.E., Winter, G. (1993) Retroviral vectors displaying functional antibody fragments, *Nucleic Acids Res.* **21**, 1081–1085.

Shin, S.U., Wu, D., Ramanathan, R., Pardridge, W.M., Morrison, S.L. (1997) Functional and pharmacokinetic properties of antibody-avidin fusion proteins, *J. Immunol.* **158**, 4797–4804.

Shusta, E.V., Kieke, M.C., Parke, E., Kranz, D.M., Wittrup, K.D. (1999) Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency, *J. Mol. Biol.* **292**, 949–956.

Smith, G.P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface, *Science* **228**, 1315–1317.

Smith, R.I., Coloma, M.J., Morrison, S.L. (1995) Addition of a mu-tailpiece to IgG results in polymeric antibodies with enhanced

effector functions including complement-mediated cytolysis by IgG4, *J. Immunol.* **154**, 2226–2236.

Smith, R.I., Morrison, S.L. (1994) Recombinant polymeric IgG: an approach to engineering more potent antibodies, *Biotechnology (NY)* **12**, 683–688.

Somia, N.V., Zoppe, M., Verma, I.M. (1995) Generation of targeted retroviral vectors by using single-chain variable fragment: an approach to *in vivo* gene delivery, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7570–7574.

Tada, H., Kurokawa, T., Seita, T., Watanabe, T., Iwasa, S. (1994) Expression and characterization of a chimeric bispecific antibody against fibrin and against urokinase-type plasminogen activator, *J. Biotechnol.* **33**, 157–174.

Valone, F.H., Kaufman, P.A., Guyre, P.M., Lewis, L.D., Memoli, V., Deo, Y., Graziano, R., Fisher, J.L., Meyer, L., Mrozek-Orlowski, M., Wardwell, K., Guyre, V., Morley, T.L., Arvizu, C., Fanger, M.W. (1995) Phase Ia/Ib trial of bispecific antibody MDX-210 in patients with advanced breast or ovarian cancer that overexpresses the proto-oncogene HER-2/*neu*, *J. Clin. Oncol.* **13**, 2281–2292.

van Dijk, M.A., van de Winkel, J.G. (2001) Human antibodies as next generation therapeutics, *Curr. Opin. Chem. Biol.* **5**, 368–374.

Vaughan, T.J., Williams, A.J., Pritchard, K., Osbourn, J.K., Pope, A.R., Earnshaw, J.C., McCafferty, J., Hodits, R.A., Wilton, J., Johnson, K.S. (1996) Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library, *Nat. Biotechnol.* **14**, 309–314.

Verhoeyen, M., Milstein, C., Winter, G. (1988) Reshaping human antibodies: grafting an antilysozyme activity, *Science* **239**, 1534–1536.

Weiner, L.M., Clark, J.I., Ring, D.B., Alpaugh, R.K. (1995) Clinical development of 2B1, a bispecific murine monoclonal antibody targeting c-erbB-2 and Fc gamma RIII, *J. Hematother.* **4**, 453–456.

Weinstein, J.N., Eger, R.R., Covell, D.G., Black, C.D., Mulshine, J., Carrasquillo, J.A., Larson, S.M., Keenan, A.M. (1987) The pharmacology of monoclonal antibodies, *Ann. N.Y. Acad. Sci.* **507**, 199–210.

Zhu, D., Kepley, C.L., Zhang, M., Zhang, K., Saxon, A. (2002) A novel human immunoglobulin Fc gamma Fc epsilon bifunctional fusion protein inhibits Fc epsilon RI-mediated degranulation, *Nat. Med.* **8**, 518–521.

Antigen Presenting Cells (APCs)

Harald Kropshofer and Anne B. Vogt
Roche Center for Medical Genomics, F. Hoffman-La Roche Ltd, Basel, Switzerland

1 Antigen-specific Immunity 357

2 Antigen Presenting Cells (APCs) 358
2.1 The Discovery of Langerhans 358
2.2 Professional APCs 358
2.3 Nonprofessional APCs 359

3 Antigens Presented by APCs 360
3.1 Antigen Types 360
3.2 Biochemical Nature of Antigens 361
3.3 Origin of Antigens 362

4 Antigen Entry Sites 362
4.1 APCs in the Skin 362
4.2 APCs in the Mucosa 363
4.3 APCs in the Gut 363
4.4 Antigen Uptake in the Respiratory Tract 364

5 Antigen Presenting Molecules 364
5.1 MHC Class I Molecules 364
5.2 MHC Class II Molecules 366
5.3 CD1 Molecules 366

6 Antigen Processing by APCs 367
6.1 MHC Class I Processing Pathways 368
6.1.1 Generation of Peptide Ligands 368
6.1.2 TAP 368
6.1.3 The MHC Class I Loading Complex 369

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

6.2 MHC Class II Processing Pathways 370
6.2.1 Endosomal Generation of Peptides 370
6.2.2 Invariant Chain and CLIP 370
6.2.3 HLA-DM 370
6.2.4 Tetraspan Network 371

7 Activation of T Cells by APCs 371
7.1 Cell Adhesion Molecules 372
7.2 Costimulatory Molecules 373
7.3 The Immunological Synapse 374

8 Macrophages 375
8.1 Evolutionary and Cellular Origin 375
8.2 Antigen Recognition Receptors 376
8.3 Macrophage Activation 378
8.4 Immunological and Nonimmunological Effector Functions 379
8.4.1 Microbial Killing 379
8.4.2 Tissue Remodeling 380
8.5 Deactivated and Alternatively Activated Macrophages 380

9 B lymphocytes 381
9.1 Antigen-induced B-Cell Activation 381
9.2 Antigen Presentation Outside the Follicular Center 382
9.2.1 B-cell Epitope versus T-cell Epitope 382
9.2.2 CD40–CD40L Engagement 384
9.3 Antigen Presentation in the Germinal Center 384
9.3.1 B Cell/Follicular DC Contacts 385
9.3.2 HLA-DO 385

10 Dendritic Cells (DCs) 386
10.1 DC Subsets 387
10.1.1 Myeloid Versus Lymphoid DCs 387
10.1.2 Human DC Subtypes 388
10.2 Antigen Uptake Mechanisms 389
10.2.1 Macropinocytosis Versus Phagocytosis 390
10.2.2 Receptors for Endocytosis 390
10.2.3 The Adhesion Receptor DC-SIGN 391
10.3 Maturation of DCs 391
10.3.1 Maturation Stimuli 391
10.3.2 The Danger Hypothesis 392
10.4 Antigen Processing and Presentation 394
10.4.1 Classical and Cross-presentation via MHC Class I 394
10.4.2 MHC Class II–restricted Antigen Presentation 394

10.5 T Cell Priming and Polarization 395
10.5.1 Priming of $CD8^+$ and $CD4^+$ T Cells 395
10.5.2 T Helper Cell Polarization 396
10.6 Immune Evasion and Plasticity of DCs 397
10.7 DCs in Immunotherapy 398

Bibliography 399
Books and Reviews 399
Primary Literature 400

Keywords

Antigen
Any piece of a substance that can induce a specific immune response or reacts to a specific antibody or T cell.

Adhesion Molecules
Proteins, such as integrins, selectins, members of the Ig superfamily, or CD44, that mediate the binding of one cell to other cells or to the extracellular matrix.

Adjuvant
A substance that enhances the immune response to antigens and is therefore used in immunizations.

Activation
A process by which the immune system (e.g. T cells, B cells) is switched on to deal with an infection.

Costimulation
Secondary signals provided to T cells by costimulatory molecules (e.g. CD80, CD86) on antigen presenting cells. Regulated expression of costimulatory molecules allows T-cell activation in inflammation but not under healthy conditions.

CpG
Denotes a deoxynucleotide motif in bacterial DNA that differs in sequence and methylation state from mammalian DNA. It binds to and activates certain subsets of dendritic cells.

Danger
A term popularized by Matzinger, that covers the various signals from damaged tissues or from microbial products that trigger activation of dendritic cells.

dsRNA
Denotes double-stranded RNA, known as the genome of retro-viruses.

Endotoxin
Lipopolysaccharides (LPS) that are an integral part of the outer cell envelope membrane of gram-negative bacteria. They evoke toxic and pyrogenic effects and cause sepsis.

Epitope
Specific amino acid sequence of foreign or self-antigens that an antibody (B-cell epitope) or a T-cell receptor (T-cell epitope) recognizes, binds to, and reacts against.

Germinal Center
Sites in secondary lymphoid organs where activated B cells undergo proliferation, selection, maturation, and apoptosis.

HLA
Acronym for human leukocyte antigen, denoting the human MHC molecules.

Humoral Immunity
Immune response involving B cells that produce antibodies.

Immunoglobulin
A protein produced by a B cell that binds to a specific antigen leading to attack by the immune system.

LPS
Lipopolysaccharide (cf. endotoxin).

Lymphatic System
A network of vessels that is separate from the blood circulation. It includes the lymph nodes that are the command centers of the immune system.

MHC
Major histocompatibility complex – a gene cluster encoding MHC molecules, originally described as antigens responsible for transplant rejection. MHC molecules bind antigenic peptides in order to activate T cells in an antigen-specific manner.

Peptide
Short fragment of a protein. On binding to MHC molecules, it can activate T cells and thereby becomes antigenic.

Processing
Set of activities of an antigen presenting cell encompassing proteolysis of an antigen, generation of antigen peptides, peptide loading onto MHC molecules, and transport of MHC-peptide complexes to the cell surface.

T_{H1}
Helper T cells that produce high levels of IFN-γ and promote activation of macrophages.

T_{H2}
Helper T cells that produce IL-4, IL-5, and IL-I0 and promote eosinophil, B-cell, and mast-cell functions.

Tolerance
The failure of the immune system to respond to antigens.

Vaccination
Administering an antigen to generate long-term antibody or cellular immune responses or both.

Antigen presenting cells play a key role in the immune system as they are the only cells of the body that are able to trigger antigen-specific immune responses upon infection with foreign invaders. Three differentially specialized types of cells are in charge of this vital task: dendritic cells, macrophages, and to a certain extent, B lymphocytes. We are beginning to understand the molecular mechanisms of how these cells take up, process, and present antigens and how this translates into T-cell activation or silencing. This knowledge is fundamental for both our current concepts about how the immune system discriminates between self and nonself and for the development of novel vaccination strategies against diseases still threatening mankind.

1 Antigen-specific Immunity

Our environment is rich in microbes and parasites, a multitude of them being pathogenic to humans and/or to other mammals or vertebrates. The necessity of means to combat microbial invaders is evident and is accomplished through the presence of a functional immune system. One type of weapon was discovered by Behring and Kitasato at the end of the nineteenth century in the serum of rabbits immunized with diphtheria bacilli: the cell-free serum was capable of destroying diphtheria toxins, it was specific for diphtheria and remained effective in the organism of other animals – the breakthrough in immunizing against diphtheria. Behring and his colleagues named the serum components of this monumental discovery *antitoxins*. Later on they were named *antibodies*, and the agents able to induce the generation of antibodies became known as *antigens*.

An immune response can be antigen-specific, as in Behring's immunization experiments with diphtheria and tetanus toxins, but also antigen-nonspecific (Fig. 1). Nonspecific immunity is based on specialized cells, such as natural killer (NK) cells,

Innate immunity	Acquired immunity
Eosinophils	B lymphocytes
Neutrophils	Macrophages
Macrophages	Dendritic cells
NK cells	T lymphocytes

Fig. 1 Cells of the innate immune system versus cells of the acquired immune system. Eosinophils, neutrophils, macrophages, and natural killer (NK) cells confer innate immunity. B lymphocytes, macrophages, and dendritic cells are professional antigen presenting cells, often referred to as "APCs." They are the key players in facilitating acquired immunity.

eosinophils and neutrophils, and phagocytic cells, such as microglial cells in the brain or Kupffer cells in the liver. None of these elements require specific antigen recognition. These cells belong to the innate immune system. However, most pathogens cannot be eliminated by innate immunity alone but only through the involvement of elements of the adaptive or acquired immune system, which is antigen-specific. This system relies on antigen presenting cells that do not simply destroy antigens but expose their antigens on the surface for recognition by T cells. In particular, macrophages link the innate with the acquired immune system being both phagocytes and antigen presenters. Antigen exposure to effector T lymphocytes leads to production of antigen-specific antibodies, to cytolytic elimination of infected cells in an antigen-specific manner and/or to pathogen killing by phagocytosis. Thus, antigen presenting cells govern the initiation of both humoral and cellular immunity.

2 Antigen Presenting Cells (APCs)

2.1 The Discovery of Langerhans

In the second half of the nineteenth century, the histologist Langerhans was the first to detect a cell that immunologists nowadays call an *antigen presenting cell*. He used a gold impregnation method to stain skin sections and saw irregularly shaped cells with long protrusions reminiscent of dendrites of cutaneous nervous system cells. These cells formed a network of dendrites within the spinal layer of the epidermis and were called *Langerhans cells* (LCs). At the ultrastructural level, LCs can be identified by unique organelles in their cytoplasma, termed *Birbeck granules*. Birbeck granules are unique to LCs and allow us to distinguish them from macrophages and other dendritic cells. Almost a century later, LCs were rediscovered by immunologists and shown to function as APCs. Today, it is obvious that LCs are representatives of the dendritic cell family, specialized in trapping antigens that have penetrated the epidermis of the skin. Hence, they function as the sentinels of the antigen-specific skin immune response.

2.2 Professional APCs

The term APC refers to not only LCs but to all types of leukocytes that are responsible for taking up antigens at the entry sites of the body, transporting them to lymphoid organs where they recruit antigen-specific T lymphocytes (short: T cells). APCs thus act at the very beginning of an adaptive immune response. In order to mount such a response against an antigen that enters the body for the first time, APCs must be able to activate so-called *naive* T cells that

have not been stimulated before and "see" the respective antigen for the first time.

APCs that are able to prime naive T cells are termed *professional APCs* and include three sets of cells (Fig. 2): mononuclear phagocytes or macrophages, dendritic cells (DCs), and B lymphocytes (B cells). These three classes of APCs can be distinguished from nonprofessional APCs in that they express a specific combination of surface proteins that are critical for priming naive T cells. These surface proteins are (1) the gene products of the major histocompatibility complex (MHC): MHC class I and class II molecules bind peptides derived from antigenic proteins and present them to either cytotoxic T cells, characterized by the coreceptor CD8, or helper T cells expressing the coreceptor CD4, respectively; (2) the costimulatory molecules CD80 (B7.1) and CD86 (B7.2), which promote growth and differentiation of T cells upon engaging APCs; and (3) adhesion molecules such as ICAM-1 or LFA-3 that facilitate the primary contact between T cells and APCs.

DCs, although being highly diverse in origin and shape, appear to be the most potent professional APCs: DCs localize to all putative antigen entry sites of the skin, mucosa, and airways, they express high amounts of both classes of MHC molecules and, as members of the phagocyte family, they have an enormously high capacity to pino- and phagocytose antigenic material. Owing to their superior impact on the functionality of the acquired immune system, numerous attempts are ongoing to exploit DCs in immunotherapy of cancer and other diseases. Interestingly, about 100 years ago, the writer George Bernard Shaw stated in his famous book The Doctor's Dilemma: "There is only one genuinely scientific treatment for all diseases and that is to stimulate the phagocytes."

2.3 Nonprofessional APCs

One major prerequisite of bone marrow–derived professional APCs is their

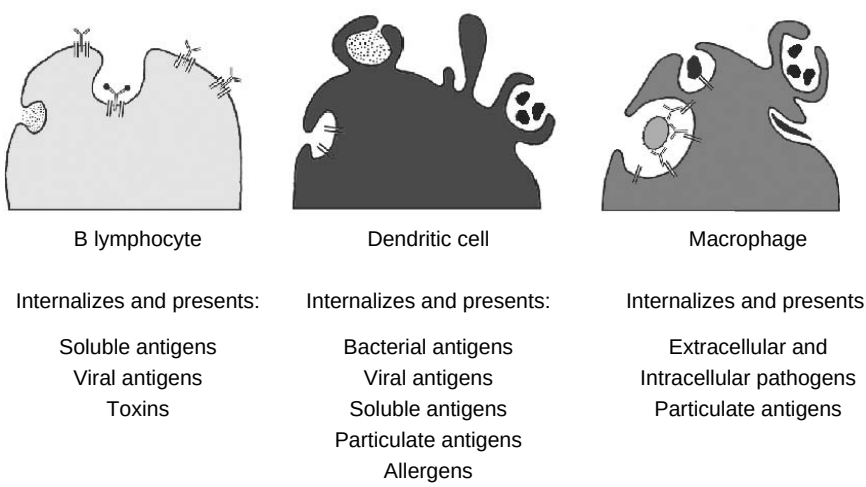

Fig. 2 The three types of antigen presenting cells (APCs). They have overlapping but distinct repertoires of antigens that they internalize, process and present to T cells.

Tab. 1 Nonprofessional APCs of the human immune system.

Cell type	*Constitutive MHC class I*	*Constitutive MHC class II*	*Costimulators CD80/CD86*
Thymic cortical epithelial cells	+	+	−
Thymic medullary epithelial cells	+	+	−
Keratinocytes	+	−	−
Myoblasts	+	−	−
T lymphocytes	+	−	−

concerted expression of MHC class II and costimulatory molecules CD80 and CD86. Apart from a very few exceptions, for example, vascular endothelial cells, only B cells, macrophages, and DCs are able to express reasonable amounts of CD80 and CD86 that are essential for activation of naive T cells. However, a number of other cell types can express MHC class II molecules (Table 1).

Thymic cortical epithelial cells (cTECs) that are critical for positive selection of $CD4^+$ thymocytes, and thymic epithelial cells of the medulla, which are responsible for deletion of self-reactive $CD4^+$ thymocytes, express MHC class II molecules in a constitutive fashion. The importance of MHC class II expression on cTECs is demonstrated by MHC class II–negative mutant mice that lack mature $CD4^+$ T cells. Following cytokine stimulation *in vitro* or during inflammatory states *in vivo*, MHC class II molecules are induced on a variety of epithelial and endothelial cells. IFN-γ induces MHC class II molecules in keratinocytes of the skin and even in myoblasts of muscles. Likewise, activated T lymphocytes in humans can become MHC class II positive.

It has been shown that cells becoming MHC class II–positive in the presence of proinflammatory cytokines, such as IFN-γ, are able to present peptides to $CD4^+$ T cells and to activate them, provided that the T cells do not need costimulation. Neo-MHC class II–expressing cells that have the potential to present antigen, are termed *nonprofessional* APCs.

3 Antigens Presented by APCs

3.1 Antigen Types

According to the initial definition brought up by the discoveries of Behring and Ehrlich, "antigens" are substances that cause the production of antibodies. However, this historic definition only covers humoral immune responses. Nowadays, molecules that are recognized by the antigen receptor of a B lymphocyte (B-cell receptor) and/or the antigen receptor of a T lymphocyte (T-cell receptor) are termed *antigens*. Antigens that initiate an immune response are named *immunogens*, whereas others may induce tolerance and are named *tolerogens*. Another subset of antigens that have a low molecular weight (<4000 Da) and are immunogenic only if they are covalently bound to a carrier protein, is termed *haptens*. Examples of haptens are substances, such as poison

ivy or penicillins, that bind covalently to proteins before or after take-up by APCs and then can cause allergy.

3.2 Biochemical Nature of Antigens

APCs are able to successfully present a large variety of organic substances (mainly consisting of carbon, hydrogen, oxygen, and nitrogen atoms); however, not a single inorganic substance is known to be able to trigger an antigen-specific immune response without the involvement of proteins. This is also true for large crystals, such as renal calculi. This clear difference in the immunogenicity indicates that the adaptive immune system evolved to circumvent the danger originating from organic and not inorganic material.

The most potent antigens are proteins and polypeptides. This is also true for polypeptides that consist of D-amino acids instead of naturally occurring L-amino acids. The reason for the strong immunostimulatory capacity is that proteins are easily recognized by the B-cell receptor and by (soluble) antibodies in the blood as well. In addition, proteins give rise to peptides during processing by APCs and peptides are the substances of choice to activate cellular immune responses through recognition by the T-cell receptor of T lymphocytes (Table 2).

Polysaccharides and glycans are by far weaker antigens than proteins. They can activate B cells but not T cells because APCs cannot present polysaccharides to T cells. Nevertheless, *Saccharomyces glycan* stimulates macrophages to secrete IL-1 and TNF-α, thereby supporting antitumor immunity in mice. Polysaccharides from bacteria, such as multiple branched glucose polymers (dextranes), induce strong antibody responses in humans and mice, however, not in rabbits or guinea pigs.

Most lipids are too small to act as antigens that induce antibody production. Similar to other low molecular weight substances, a few lipids can function as haptens, such as cardiolipin. Cardiolipin is a phospholipid derived from the mitochondrial membrane. It is released from cells upon bacterial infection with *Treponema pallidum*, binds to carrier proteins and, in this context, initiates formation of anticardiolipin antibodies (Table 2).

However, DCs such as LCs express CD1 molecules that bind and present certain lipids or glycolipids in order to induce T-lymphocyte activation. Nonglycosylated and glycosylated mycolates,

Tab. 2 Antigens presented by APCs via MHC and CD1 molecules, versus haptens.

Restriction	*Antigen*	*Hapten*
MHC class I/MHC class II	Proteins	Polysaccharides
	Polypeptides	DNA
	Peptides	Penicillin
	Glycopeptides	Cardiolipin
CD1	Mycolates	Organic substances
	Diacylglycerols	
	Sphingolipids	
	Polyisoprenoids	

diacylglycerols, sphingolipids, and polyisoprenoids derived from endogenous and microbial sources have been described (Table 2).

Similar to pure lipids, only DNA or RNA is widely unable to trigger generation of antibodies. Moreover, APCs obviously lack specialized receptors for DNA or RNA necessary for activating T lymphocytes. Nucleoproteins, however, are suitable to generate anti-DNA antibodies. Thus, DNA can serve as a hapten. Anti-DNA antibodies are actually found in the serum of patients suffering from the autoimmune disease systemic lupus erythematosus (SLE). Some of these antibodies react with denatured DNA, others with intact double-strand DNA.

3.3 Origin of Antigens

Under normal circumstances, the immune system relying on APCs reacts only against foreign antigens, but not against self-antigens. In cases in which a foreign antigen is a homolog of a self-antigen, the extent of the immune response decreases with increasing similarity between both antigens. To take this rule into account, transplantation immunologists differentiate between the donor and the recipient of a transplanted organ: antigens derived from a recipient of the same species are called *alloantigens*. When the antigen is derived from another species, it is named *xenoantigen*. In contrast to that, *autoantigens* are derived from the same organism and may cause autoimmune diseases.

4 Antigen Entry Sites

Vertebrates offer microbes a large interface to enter the body. Consequently, the lymphatic system is organized in such a way that lymphoid tissues are located in close proximity to all putative entry sites. Afferent lymphatic vessels drain fluid and APCs from the skin and from other peripheral tissues to the lymph node. The spleen filters antigen mainly from the blood stream. The gut-associated lymphoid tissues (GALT), which include the tonsils, adenoids, appendix, and the Peyer's patches (PP) in the small intestine, collect antigen from the epithelial surfaces of the gastrointestinal tract.

Likewise, the bronchial-associated lymphoid tissues (BALT) and the mucosal-associated lymphoid tissues (MALT) protect the respiratory epithelia and other mucosa respectively. Each of these lymphatic tissues traps foreign antigens through the activity of migratory APCs that present antigenic peptides to T lymphocytes, thereby initiating antigen-specific immunity. Importantly, even in the absence of foreign antigen, lymphoid tissues provide signals to lymphocytes via presentation of self-antigens. This is important for the survival of naive and silent T lymphocytes, thereby contributing to the maintenance of T-cell homeostasis.

4.1 APCs in the Skin

Antigen that has penetrated the epidermis is primarily picked up by LCs. LCs are derived from monocytes that leave the blood stream and migrate into the epidermis where they adopt the shape of typical DCs. Antigen-fed LCs leave the skin and migrate as so-called *veiled cells* via the afferent lymph vessels to the draining lymph nodes. There, they accumulate as interdigitating dendritic cells (IDCs) in the T cell–rich areas.

Under pathological conditions, for example, in skin suffering from the autoimmune disease psoriasis, not only LCs but also keratinocytes express high amounts of MHC class II molecules and become capable of functioning as APCs. In other squamous epithelia, DCs different from LCs do occur. This is true of the oral and nasal mucosa, the trachea and bronchus, and the esophagus and the tonsils. In the epithelia lining the lung, the stomach, and the gut, DCs have not been found, as yet.

4.2 APCs in the Mucosa

Although tight junctions and the mucous layer of mucosal surfaces widely exclude microbes and large particulate antigen(s), antigenic proteins are known to surpass this barrier through endo- and transcytosis. Granulocytes, as representatives of the innate immune system, and macrophages functioning as APCs take up these antigens in the lamina propria. Furthermore, in the gut and bronchus, an antigen transfer system exists that transports antigens from the lumen to the GALT and BALT respectively. A prerequisite of this system is the presence of a specialized type of epithelial cell, the so-called *M cell* (see below).

4.3 APCs in the Gut

The uptake of bacteria, proteins, and abiotic substances (e.g. latex) by the Peyer's patches and by lymphoid follicles in other parts of the gut, for example, the appendix, is well known. The GALT is separated from the gut lumen by a single layer of epithelium, containing M cells (Fig. 3). In ultrathin sections, an M cell is seen as a rim of apical cytoplasm that bridges the space between two adjacent epithelial cells. An M cell forms a kind of an umbrella above a space surrounded by epithelial cells and filled with all sorts of APCs including DCs, macrophages, and B cells. Thus, antigen from the gut lumen is transcytosed through the epithelial monolayer and

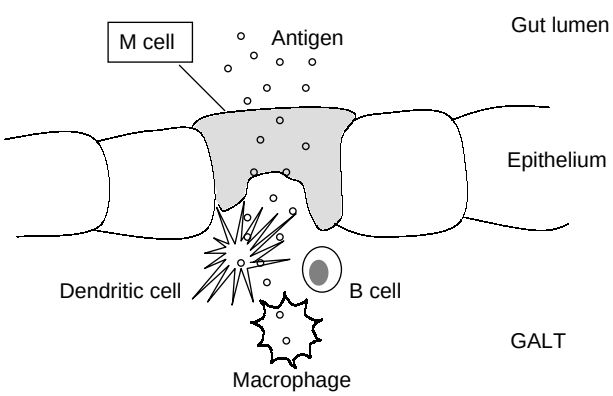

Fig. 3 The gut-associated lymphoid tissue (GALT). Antigens from the gut lumen traverse the single layer of epithelia via M cell–mediated transcytosis whereupon they get in contact with APCs (dendritic cells, macrophages, B cells) of the GALT. APCs are very similarly organized in the lymphoid tissues of the bronchus and Waldeyer's ring.

reaches APCs, which take up and process the antigen and present it to T lymphocytes in the PP or in mesenteric lymph nodes.

4.4 Antigen Uptake in the Respiratory Tract

In the upper respiratory tract, antigen presenting LCs occur in the transitional zones between the keratinizing epithelium of the skin and the mucosal surface of the nasal cavity and the nasopharyngeal region, for example, the lips. Moreover, APCs fulfill their sentinel function in and around the lymphoid tissue of Waldeyer's ring. The epithelium above the ring of Waldeyer contains, just as BALT and GALT do, antigen transporting M cells. Furthermore, DCs are found in the epithelium of the trachea. The rest of the mucosal surface of the upper respiratory tract serves to exclude rather than take up antigen.

Antigen exclusion is also dominant in the lower respiratory tract, in particular, the lung. Antigen clearance mainly depends on mucociliary activity of the epithelia and/or alveolar macrophages. Macrophages in the alveoli act as both phagocytic cells of the innate defense system and as APCs giving rise to acquired immunity. Some of them simply phagocytose and digest foreign antigens involving the lung; others stimulate a local immune response. For this purpose, alveolar macrophages loaded with antigen cross the lining of the alveoli and migrate into lymph nodes located in the lung tissue. Thus, similar to DCs, alveolar macrophages form a lung-specific antigen handling system.

5 Antigen Presenting Molecules

Both types of the classical molecules of the MHC, class I and class II, are peptide receptors and encoded by genes that display the highest degree of polymorphism in the genome. Within the MHC molecules, the polymorphism is mainly restricted to the regions that constitute the antigenic peptide-binding cleft. The advantage of the multitude of MHC alleles encoding for a multitude of distinct peptide-binding specificities is that there will be hardly any foreign peptide that cannot be bound by at least a few individuals of a population. In contrast, CD1 molecules are MHC class I–like receptors for mycobacterial lipids or glycolipids. They are not encoded in the MHC and are monomorphic. Their physiological role is, as yet, ill defined.

5.1 MHC Class I Molecules

MHC class I molecules are glycosylated transmembrane proteins that consist of two subunits: the larger α-chain encoded in the MHC genetic locus is polymorphic and forms the peptide-binding site. It associates with the smaller subunit, β_2-microglobulin (β_{2m}), which is nonpolymorphic and encoded outside the MHC. β_{2m} and the membrane-proximal α_3-domain of the α-chain fold as immunoglobulin domains, whereas the membrane-proximal (membrane-distal) α_1- and α_2-domain constitute the peptide-binding cleft (Fig. 4).

Quite in contrast to other peptide receptors described in the hormone system, the binding cleft of MHC molecules is highly promiscuous so that a large variety of antigenic peptides can be presented. In the case of MHC class I molecules, this peculiarity is accomplished through three structural criteria shared by peptide ligands: (1) the peptide length is limited to 8 to 10 amino acids, with the majority

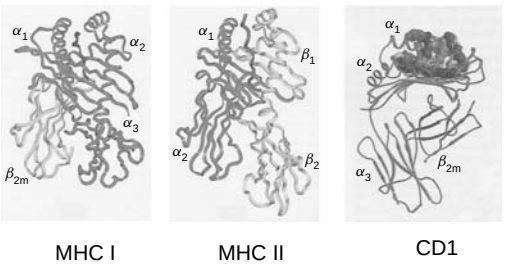

Fig. 4 X-ray structure of the MHC class I molecule HLA-Aw68, the MHC class II molecule HLA-DR1, and the CD1b molecule, which are key surface molecules of APCs. Adapted from Guo, H. C. et al. (1992) *Nature* **26**, 300–301; Murthy, V. L., Stern, L. J. (1997) *Structure* **5**, 1385–1396; Gadola, S. D., Zaccai, N. R., Harlos, K., Shepherd, D., Castro-Palomino, J. C., Ritter, G., Schmidt, R. R., Jones, E. Y. Cerundola, V. (2002) Structure of human CD1b with bound ligands at 2.3 Å, a maze for alkyl chains, *Nat. Immunol.* **3**, 721–726.

of peptides being 9-mers; (2i) 1 to 2 side chains of the peptide, called *anchors*, fit into corresponding specificity pockets in the binding cleft. Both the position and identity of these anchor residues vary, depending on the particular MHC class I molecule (allele-specific anchor motifs); and (3) the amino- and carboxyterminni of peptides are fixed in the binding cleft through a hydrogen-bonding network.

To illustrate this in more detail, the sequences of self- and foreign peptides restricted by the human MHC class I molecule HLA-A2 are compared (Table 3): 70% of these peptides are 9-mers with a leucine or methionine at the anchor position P2 and a valine at the anchor position P9. Owing to the fact that the sequences at the nonanchor positions are widely irrelevant for binding, more than

Tab. 3 Peptide anchor motif of antigenic peptides binding to the MHC class I molecules HLA-A1 and HLA-A2.

Antigen					***Sequence***						***Allele***
		P2							**P9**	**P10**	
Tyrosinase (369–377)	Y	**M**	N	G	T	M	S	Q	**V**		
EBV LMP2 (426–434)	C	**L**	G	G	L	L	T	M	**V**		
HIV RT(476–484)	I	**L**	K	E	P	V	H	G	**V**		
HTLV-1 Tax(11–19)	I	**L**	F	G	Y	P	V	Y	**V**		HLA-A2
Hepatitis B sAg(335–343)	W	**L**	S	L	L	V	P	F	**V**		
Tyrosinase (1–9)	M	**M**	N	G	T	M	S	Q	**V**		
pmel 17/gp100	I	**L**	D	G	T	A	T	L	R	**L**	
Influenza B NP(85–94)	K	**L**	G	E	F	Y	N	Q	M	**M**	
Influenza MP(59–68)	I	**L**	G	F	V	F	T	L	T	**V**	
HPV11 E7 (4–12)	R	**L**	V	T	L	K	D	I	**V**		
					Anchor motif						
		L							V		
		M							L		
									M		
					Anchor motif						
–			**P3**						**P9**		
			D						Y		HLA-A1
			E								

500 different peptides have been estimated to constitute the self-peptide repertoire of HLA-A2 expressed on the surface of a B cell; however, none of these peptides binds to HLA-A1 and vice versa. The reason is that HLA-A1 and HLA-A2 differ strongly in the anchor motif of their peptide ligands.

Today, more than 220 distinct HLA-A and more than 450 distinct HLA-B type alleles are known. Not all of the alleles have distinct anchor motifs, however, the high diversity of allelic MHC products observed in the human population allows a high diversity of self- and foreign peptides to be bound. This strategy reduces the risk that some pathogens may escape recognition by the immune system due to holes in the binding capacity of MHC molecules.

5.2 MHC Class II Molecules

The crystallographic structure of MHC class II molecules shows that they are folded very much like MHC class I molecules, although class II dimers consist of a noncovalent complex of an α- and a β-chain, which are both encoded within the MHC. Again, highly polymorphic regions are located in the peptide-binding cleft, consisting of 8β-strands and 2α-helical segments that serve to sandwich peptide ligands (Fig. 4).

The only and decisive structural difference is that the class II binding cleft is open at both ends, whereas the class I cleft is closed. A critical consequence is that naturally processed peptides binding to class II MHC molecules are at least 12 to 14 amino acids long and can be much longer. The majority of class II–associated peptides are 15- to 17-mers. The binding forces of the class II peptide-binding cleft are distributed to 3 to 4 specificity pockets, most often localized at relative positions P1, P4, P6, and P9 (Table 4). Moreover, about a dozen hydrogen bonds contacting the whole backbone of a typical 15-mer stabilize the class II–peptide interaction. In further contrast to class I molecules, the specificity pockets of the class II groove are more permissive, often allowing 3 to 5 structurally similar amino acid side chains to fit into a particular pocket (Table 4).

The advantage of this peculiarity is that class II molecules encoded by the same MHC class II allele are able to accommodate a large repertoire of different peptide sequences. This may be viewed as an adaptation to pathogens that constantly change their immunodominant antigens by high mutation rates. The disadvantage is that MHC class II allele-specific peptide ligand motifs are rather complex and class II–restricted immunodominant epitopes are difficult to predict.

5.3 CD1 Molecules

CD1 molecules are nonpolymorphic transmembrane glycoproteins encoded by genes outside the MHC. The overall structure of CD1 is similar to that of MHC class I molecules in that CD1 molecules form heterodimers of ~45-kDa heavy chains associated with β_{2m}. However, unlike MHC class I molecules, CD1 is not retained within the endoplasmic reticulum but is targeted at endocytic compartments where it binds its ligands. CD1 molecules are expressed only on dendritic cells, monocytes, and some thymocytes. Like classical MHC class I and class II molecules, CD1 molecules are recognized by T cells. However, unlike classical MHC molecules, group I CD1 molecules (CD1a, CD1b, CD1c), which are found in humans but not in mice, present mycobacterial lipids, mycolic acids, lipoarabinomannan,

Tab. 4 Self-peptides eluted from the class II molecule HLA-DR4 of human monocyte-derived dendritic cells, stimulated with TNF-α.

Antigen	***Sequence***																			
							P1			**P4**		**P6**								
HLA-B60	D	T	Q	F	V	R	**F**	D	S	**D**	A	**A**	S	Q	R	M				
HLA-B60	D	T	Q	F	V	R	**F**	D	S	**D**	A	**A**	S	Q	R					
HLA-B60		T	Q	F	V	R	**F**	D	S	**D**	A	**A**	S	Q	R					
HLA-B60			Q	F	V	R	**F**	D	S	**D**	A	**A**	S	Q	R					
HLA-B60				F	V	R	**F**	D	S	**D**	A	**A**	S	Q	R					
HLA-B60				F	V	R	**F**	D	S	**D**	A	**A**	S	Q	R	M	E	P		
β_{2m}				Y	L	L	**Y**	Y	T	**E**	F	**T**	P	T	E	K	D	E		
β_{2m}					L	L	**Y**	Y	T	**E**	F	**T**	P	T	E	K	D	E		
Mannose receptor			F	E	N	K	**W**	Y	A	**D**	C	**T**	S	A	G	R	S	D	G	
Calreticulin			D	N	P	E	**Y**	S	P	**D**	P	**S**	I	Y	A	Y	D	N		
Transferrin receptor			T	G	Q	F	**L**	Y	Q	**D**	S	**N**	W	A	S	K	Y			
Apolipoprotein d		V	L	N	Q	E	**L**	R	A	**D**	G	**T**	V	N	Q	I	E	G		
Apolipoprotein d			L	N	Q	E	**L**	R	A	**D**	G	**T**	V	N	Q	I	E	G		
Apolipoprotein d					Q	E	**L**	R	A	**D**	G	**T**	V	N	Q	I	E	G		
IL-10 receptor			D	K	L	S	**V**	I	A	**E**	D	**S**	E	S	G	K	Q	N	P	G
Rab-7			F	P	E	P	**I**	K	L	**D**	K	**N**	D	R	A	K	A	S	A	
Rab-7			F	P	E	P	**I**	K	L	**D**	K	**N**	D	R	A	K	A	S		
Rab-7			F	P	E	P	**I**	K	L	**D**	K	**N**	D	R	A	K	A			
										Anchor motif										
							F			D		A								
							Y			E		T								
							W					S								
							I					N								
							L													
							V													

phosphatidylinositol mannosides, and hexosyl-1-phosphoisoprenoids to cytotoxic T cells. These ligands are derived either from internalized mycobacteria or from the uptake of lipoarabinomannans by the mannose receptor expressed on the surface of monocytes and dendritic cells. No naturally processed bacterial antigens have been identified in the context of the group II molecule CD1d molecule, which is expressed in both humans and mice. CD1d molecules present bacterial α-galactosylceramide to NK T cells. The relationship between the peptide- and lipid-binding capacities of CD1 is not clear, as yet. Structural studies show that CD1 molecules bear a deep and hydrophobic ligand-binding cleft in which glycolipids bind – it is open whether peptides associate to the same site (Fig. 4).

Compared to MHC class I and class II molecules, CD1 proteins appear to represent a separate lineage of antigen-presenting molecules specialized in presenting microbial lipids, glycolipids, and a subset of peptide antigens to T cells.

6 Antigen Processing by APCs

Virus-infected cells can be destroyed by $CD8^+$ cytotoxic T cells provided that the

respective cells present viral peptide antigens bound to MHC class I molecules on their cell surface. Likewise, exogenous foreign antigens that are pino- or phagocytosed by APCs at sites of inflammation lead to activation of $CD4^+$ helper T cells when APCs present antigen-derived peptides in the context of MHC class II molecules. The concept that foreign antigenic peptides stimulate T cells only in the context of MHC molecules has been introduced by Zinkernagl and Doherty as *MHC restriction*. The conversion of protein antigens derived from the extracellular space or the cytosol into peptides and the conservation of this antigenic information through loading onto MHC class II or class I molecules is accomplished by pathways of antigen processing.

The different fates of vesicular and cytosolic antigens are mainly due to the segregated pathways of biosynthesis and assembly of class I and class II molecules. Both pathways, however, have evolved as adaptations of basic cellular functions, such as endocytosis, transmembrane transport, and protein degradation, which are not exclusively used for antigen presentation.

6.1 MHC Class I Processing Pathways

The peptide-receptive binding cleft of class I molecules is localized on the luminal side of the ER. Class I–restricted antigenic peptides, however, are derived from cytosolic or nucleic protein antigens. Therefore, accessory molecules are necessary for antigen processing. IFN-γ increases the expression of these accessory molecules and MHC class I molecules (Fig. 5).

6.1.1 Generation of Peptide Ligands

The majority of antigenic peptides presented by MHC class I molecules appears to be generated by the multicatalytic proteinase complex, named the proteasome. The proteasome is localized in the cytosol and nucleus and consists of 28 subunits that form a central channel where unfolded polypeptides are cleaved into short peptides (Fig. 5). In the presence of IFN-γ, the catalytic activity of the proteasome changes due to the replacement of 3 constitutive subunits for the new subunits LMP2, LMP7, and MECL-1. The proteasome containing these new subunits is called the immunoproteasome and preferably generates peptides with a C-terminal hydrophobic or basic amino acid. Peptides with these structural features are superior to other peptides with regard to binding to MHC class I molecules.

Apart from the proteasome, other proteases and peptidases are involved in processing of antigens in the MHC class I pathway: cytosolic calpains and aminopeptidases are sometimes required, the actual involvement being strongly influenced by the type of antigen to be cleaved. Moreover, the ER-resident aminopeptidase ERAP-1 has recently been described to be mainly responsible for N-terminal trimming of precursor peptides, thereby giving rise to typical 8-mer, 9-mer, and 10-mer epitopes.

6.1.2 TAP

The transporter associated with antigen processing (TAP) belongs to the ABC family of heterodimeric transmembrane transporters, which are fueled by ATP hydrolysis. TAP is localized in the membrane of the ER and shuttles peptides of variable length ($n = 6$ to 30 amino acids) from the cytosol into the ER lumen (Fig. 5). This is a prerequisite

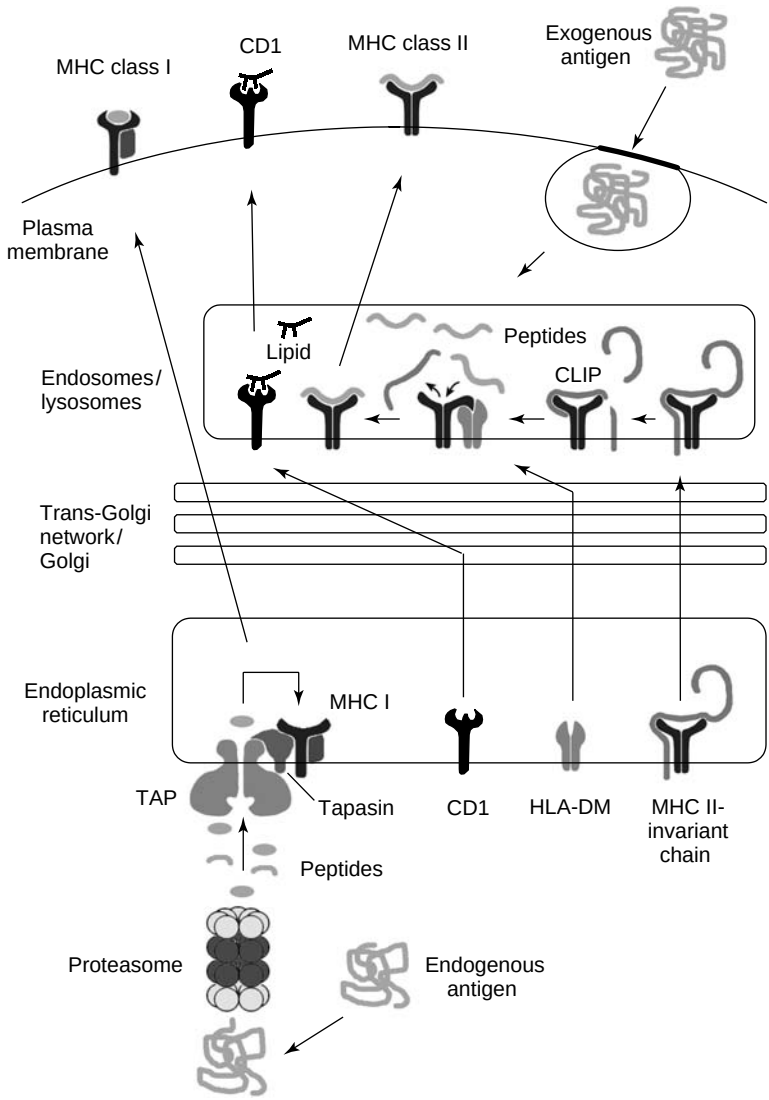

Fig. 5 Antigen processing pathways leading to the generation of MHC class I-peptide, MHC class II–peptide, and CD1-lipid complexes.

for the regular supply of MHC class I molecules with peptide antigen. Moreover, TAP is an essential constituent of a multimolecular peptide–loading complex in the ER that regulates loading of MHC class I molecules with cognate antigenic peptides.

6.1.3 The MHC Class I Loading Complex

Newly synthesized MHC class I heavy chains (α) are stabilized by binding to the chaperone calnexin until the heavy chain forms dimers with β_2-microglobulin (β_{2m}). After dissociation of calnexin, the chaperone calreticulin allows α:β_{2m}

complexes to enter a multimolecular loading complex, which is responsible for the formation of cognate class I-peptide complexes. Constituents of this loading complex are α:β_{2m}, calreticulin, TAP, the protein-disulfide-isomerase ERp57, and the TAP-associated chaperone tapasin (Fig. 5). During peptide loading, ERp57 gives allowance for the regulated opening and closing of a disulfide bond in the binding cleft domain of the class I heavy chain. Tapasin bridges TAP and α:β_{2m}, and retains peptide-free α:β_{2m} dimers in the ER until a stably binding peptide occupies the groove. Consequently, preferentially long-lived α:β_{2m}-peptide complexes leave the ER and make it to the cell surface. Thus, tapasin is thought to function as a peptide editor, similar to HLA-DM in the MHC class II processing pathway (see below).

6.2 MHC Class II Processing Pathways

MHC class II-associated peptides originate from two sources: endocytosed exogenous antigens or endogenous self-antigens, the majority of them being membrane proteins derived from endosomes or the nucleus (Table 4).

6.2.1 Endosomal Generation of Peptides

Professional APCs are equipped to internalize whole pathogens, for example, bacteria or viruses, small vesicles, multimolecular complexes, proteins or peptides via phagocytosis, pinocytosis or in a receptor-mediated fashion. In all these cases, endocytosed antigens end up in endosomal compartments, known for their low pH and the high redox potential (Fig. 5). These conditions favor denaturation of protein antigens and proteolytic generation of peptides via hydrolases, amino- and carboxypeptidases, and endopeptidases. Endosomal/lysosomal compartments of APCs, which are enriched in MHC class II molecules, as determined by electron microscopy or immunohistochemically, have been termed *MIICs* (MHC class II compartments).

6.2.2 Invariant Chain and CLIP

The invariant chain (Ii) is a chaperone dedicated primarily to MHC class II molecules. In DCs, Ii also binds to CD1 molecules. In general, Ii trimers bind to newly synthesized class II $\alpha\beta$ dimers, allowing them to obtain their native conformation. Since Ii binding blocks the sole peptide-binding cleft, Ii-associated class II molecules cannot bind antigens. Owing to a targeting signal in the cytosolic tail of Ii, Ii class II complexes are sorted from the trans-Golgi-network directly to endosomal/lysosomal MIICs (Fig. 5).

In MIICs, Ii is proteolytically degraded whereas $\alpha\beta$ dimers remain intact. The proteases mainly responsible for Ii degradation are cathepsin S in DCs and B cells and cathepsin L in cortical epithelial cells. The terminal Ii fragment that still occupies the peptide-binding cleft of class II molecules is termed *CLIP* (class II MHC-associated Ii peptide). Under steady state conditions, CLIP is part of the self-peptide repertoire of class II molecules. Since it dissociates from the majority of class II allelic proteins rather slowly, CLIP has to be actively removed from the class II binding cleft.

6.2.3 HLA-DM

HLA-DM (termed H2-M in mice) is a nonclassical MHC class II protein, displaying 20 to 25% sequence homology to both classical class I and class II molecules. X-ray analysis revealed that HLA-DM lacks

a peptide-binding groove and, thus, is unable to function as an antigen-presenting molecule. It bears a targeting signal for MIICs where it accumulates and transiently engages in complexes with classical class II MHC molecules. HLA-DM has three important functions (Fig. 5): (1) it catalytically removes the Ii-derived precursor peptide CLIP, thereby promoting peptide loading; (2) it remains bound to empty class II molecules and prevents their unfolding as a molecular chaperone; and (3) it functions as a peptide editor, as it removes peptides that bind with low kinetic stability to class II dimers. Consequently, preferentially high-stability class II peptide complexes are displayed on the surface of APCs, as long as these cells bear sufficient HLA-DM. This principle allows APCs, once loaded with cognate antigen, to activate $CD4^+$ T cells for prolonged periods of time, even a couple of days after encounter of antigenic material.

6.2.4 Tetraspan Network

Professional or nonprofessional APCs and even T cells express proteins, such as CD9, CD37, CD53, CD63, CD81, CD82, and CD151, belonging to a family termed *tetraspan proteins* or *tetraspanins*. The nomenclature relates to the fact that all of them traverse the membrane four times. It is well established that they are able to form homo- and heterodimers and thus form two-dimensional networks in the membrane. In APCs, tetraspanins form clusters with various integrins and MHC molecules. In MIICs, CD82 and CD63 associate with HLA-DM and classical class II molecules. On the cell surface, CD81, CD9, and CD53 form microdomains together with MHC class II and class I molecules. The function of tetraspan microdomains may be to increase the local density of MHC class II peptide complexes in loading compartments and on the surface, so that the avidity of an APC–T cell encounter is increased, thereby raising the efficacy of T-cell activation.

7 Activation of T Cells by APCs

After leaving the thymus, mature T cells recirculate between blood and peripheral lymphoid tissue until they recognize an MHC-peptide complex on the surface of an APC. T cells that have not yet been activated via encounter of an antigen are termed *naive T cells*. In contrast, T cells that have already contacted MHC molecules carrying foreign peptide antigens start to proliferate and differentiate into cells capable of contributing to the removal of the antigen. They are termed *effector T cells*.

Effector T cells can be subdivided into three classes, according to the type of pathogen that may skew the efficiency of peptide entry into different processing pathways. Peptides from pathogens that multiply in the cytosol, such as influenza virus, vaccinia, and *Listeria monocytogenes*, are presented by MHC class I molecules and activate $CD8^+$ cytotoxic T cells that kill infected target cells. Pathogens that accumulate in endosomal vesicles of macrophages, such as mycobacteria or *Leishmania*, facilitate the differentiation of T_{H1} helper cells that express Fas ligands enabling them to kill infected macrophages. Extracellular antigen derived from viruses tend to stimulate the production of T_{H2} helper cells that initiate a humoral immune response by activating naive B cells to produce neutralizing IgM antibodies.

The activation of naive T cells following engagement with MHC–peptide complexes in APCs constitutes a primary

immune response. In the case of such a primary response, a subset of long-lived T cells is generated that gives an accelerated response upon seeing the same antigen for the second time. These T cells are called *memory T cells*. Memory T cells differ in several instances from naive T cells, but like naive T cells they require activation by professional APCs in order to become effector T cells.

Finally, it has been shown that T cells need to contact APCs continuously in order to stay alive. This is accomplished by the recognition of MHC-self-peptide or MHC-foreign peptide complexes on the surface of APCs.

7.1 Cell Adhesion Molecules

The migration of naive T cells through lymph nodes and their initial contacts with APCs depend on interactions that are not antigen-specific, hence are independent from MHC-peptide complexes. The initial cell–cell contact between APCs and T cells is controlled by an array of adhesion molecules on the surface of T cells that recognize a complementary array of adhesion molecules on the surface of an APC. The main classes of adhesion molecules are the selectins, the integrins, members of the Ig superfamily, and mucin-like molecules.

Selectins are particularly important for the T cell homing into particular tissues. For example, L-selectins on the T-cell surface binds to addressins, such as CD34 or MAdCAM-1, on vascular endothelial cells.

To guide naive T cells to DCs, chemokines play a role: the chemokine MIP-3β directs naive T cells and mature DCs into lymphoid tissues. Moreover, DCs in lymphoid tissues express the chemokine SLC (secondary lymphoid tissue chemokine) that binds to the chemokine receptor CCR7 expressed on naive T cells. This event obviously reinforces the strength of successive T cell–DC interactions mediated via integrins and proteins of the Ig superfamily.

Integrins are heterodimeric cell surface proteins, consisting of a large α-subunit and a smaller β-subunit. There are several subfamilies of integrins broadly defined by their common β-chains. An important integrin expressed mainly by T cells, but also by macrophages, is LFA-1 (lymphocyte function-associated antigen-1), an $\alpha_L\beta_2$ integrin. LFA-1 mediates migration of both naive and effector T cells out of the blood and the initial contact with DCs. Another integrin VLA-4 (very late activation antigen-4), which is particularly strongly expressed by activated effector T cells, and the Ig family member CD2 function as substitutes for LFA-1. The majority of cell surface adhesion molecules being involved in T cell–APC interactions are members of the Ig superfamily. Three very similar intercellular adhesion molecules (ICAM-1, ICAM-2, ICAM-3) bind to the integrin LFA-1. ICAM-1 (CD54) and ICAM-2 (CD102) are expressed on APCs and on epithelia, whereas ICAM-3 (CD50) is only expressed on T cells and B cells. In addition to binding to LFA-1, ICAM-3 has a high affinity for the lectin DC-SIGN (DC-specific ICAM-3-grabbing nonintegrin; CD209), which is only found on DCs. The ICAM-3/DC-SIGN interaction appears to be exclusive to the contact between naive T cells and DCs, whereas CD2 binding to LFA-3 synergizes with LFA-1 binding to ICAM-1 and ICAM-2 on all types of APCs.

Importantly, it turned out that the strength of the LFA-1/ICAM interaction of a naive T cell and an APC depends on the type of MHC–peptide complexes

being presented by the APC; in those cases in which the T cell recognizes a cognate MHC-peptide ligand, a conformational change in LFA-1 increases its affinity for ICAM-1 or ICAM-2 of the APC. This phenomenon allows that the T cell–APC contact can persist for several days, so that a naive T cell can proliferate and become an effector T cell. In the majority of encounters, the T cell will not recognize an appropriate peptide antigen and separates from APCs, keeping on migrating through the lymph node. Thus, the transient nature of binding of naive T cells to APCs via adhesive interactions is crucial in the sampling of large numbers of MHC molecules and APCs.

7.2 Costimulatory Molecules

MHC–peptide complexes expressed on APCs are recognized by ligation of a corresponding T-cell receptor on the T-cell side. This interaction can be viewed as a first signal and it is critical for the antigen-specificity of a cellular immune response. However, in order to trigger clonal expansion and differentiation of a naive T cell, a second or costimulatory signal delivered by the same APC is required.

The best-studied costimulatory molecules solely expressed by APCs are B7.1 (CD80) and B7.2 (CD86). Both are homodimeric members of the Ig superfamily. Both B7 molecules share a common receptor on T cells, CD28. Ligation of CD80 or CD86 by CD28 costimulates the clonal expansion of naive T cells. Lack of B7 molecules abrogates T-cell proliferation.

Other costimulatory molecules on APCs are the TNF family member CD40 and 4-1BBL, which bind to CD40 ligand and 4-1BB (CD137) on the T-cell side, respectively. Both pairs of molecules are induced upon B7/CD28-mediated activation and function in sustaining the development of a full T-cell response. Notably, not only does the T cell receive a signal via the CD40/CD40L engagement but the APC also receives a signal. The best-characterized case is the conditioning of developing DCs via the engagement of activated $CD4^+$ helper T cells, so that the DC gains the capacity to fully activate $CD8^+$ cytotoxic T cells (“license-to-kill model”; cf Sect. 10.5.1.).

Activated T cells express an additional receptor for B7 molecules, denoted as CTLA-4, a homolog of CD28. However, CTLA-4 has a manyfold higher affinity to B7 than CD28 and transmits an inhibitory signal to T cells. Thus, CTLA-4 upregulation on the surface of T cells serves to downregulate proliferation and cytokine secretion of activated T cells.

A further CD28-like costimulator is ICOS, which recognizes LICOS produced on activated B cells, monocytes, and DCs. ICOS is poorly investigated so far – it is only known to induce IL-10 secretion of activated T cells.

The physiological relevance of the involvement of costimulatory molecules is to prevent destructive immune responses in the absence of signals mediated by pathogens, for example, responses against self-tissues. This is accomplished through the limitation that APCs express costimulatory B7 molecules only upon receiving danger signals from bacterial or viral products or equivalent proinflammatory signals. Consequently, naive T cells become effector T cells only when they receive signal 1 and signal 2 from the same APC. When signal 2 is lacking, the T cell is rendered anergic. Anergy means that the T cell becomes refractory to signals, even when signal 1 has been sent by professional APCs.

7.3 The Immunological Synapse

High-resolution confocal microscopy revealed that the key molecules involved in the formation of heterotypic junctions between APCs and T cells are reorganized in a highly characteristic manner. This specialized junction that results from cytoskeleton-driven clustering of MHC-peptide complexes, T-cell receptor molecules, adhesion molecules, and costimulators has been termed the *immunological synapse*. It is the close apposition of two membranes and the peculiar shape of the T cell that is reminiscent of neurological synapses (Fig. 6).

The immunological synapse consists of two concentric rings. In the center, proteins of the so-called central supramolecular activation cluster (c-SMAC), such as the T-cell receptor (TCR), CD2, CD3, the costimulator CD28, the protein kinases lck, fyn, and PKC-θ cocluster on the T-cell surface. Complementarily, MHC-peptide complexes and the costimulator CD80 have been described to cluster in the c-SMAC zone of APCs. The c-SMAC is surrounded by a second zone, the peripheral supramolecular activation cluster (p-SMAC). On T cells, the p-SMAC contains the integrin LFA-1 and the cytoskeletal protein talin. Accordingly, the adhesion molecule ICAM-1, known to be a ligand of LFA-1, coalesces in the p-SMAC of APCs.

The p-SMAC provides adhesive anchoring of the T cell to the APC, whereas the c-SMAC represents a protected zone for sustained signaling via the TCR. Immunological synapses are maintained for at least 1 h. Outside the p-SMAC is a third area containing proteins that are excluded from the synapse, such as the mucin CD43 or the phosphatase CD45. Exclusion of both types of proteins from the T cell–APC contact area is essential, as both molecules are very large and would therefore prevent

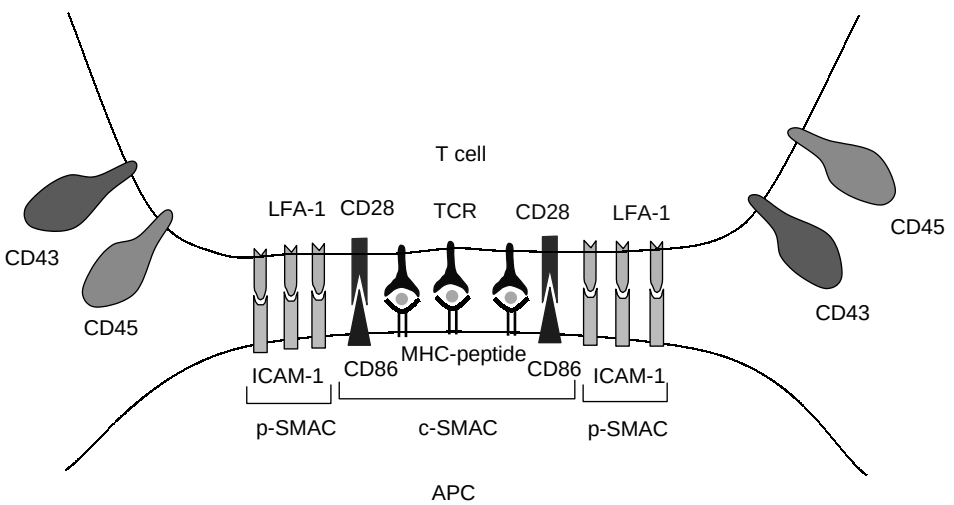

Fig. 6 The immunological synapse between a T cell and an APC. In the central zone (c-SMAC), MHC-peptide complexes, costimulators, for example, CD86, TCRs, and CD28 accumulate. In the surrounding ring zone, termed p-SMAC, the adhesion molecules ICAM-1 and LFA-1 are found. Excluded are large molecules, for example, CD43 and CD45.

small molecules, such as the TCR and MHC molecules, from interacting with each other.

Exploratory adhesive interactions of T cells and APCs precede formation of the synapse. The adhesion molecule ICAM-3 has been reported to initiate antigen-independent scanning of the APC surface by T cells followed by early signaling events and rearrangements of the T cellular cytoskeleton. The synapse itself appears to be a highly dynamic structure. At the beginning, cognate TCR ligands localize to an outermost ring of the nascent synapse, whereas LFA-1 and ICAM-1 constitute the c-SMAC. After 5 to 10 min, ICAM-1 moves to the p-SMAC and MHC-peptide clusters concentrate at the heart of the synapse. Finally, the coreceptor CD4 progressively migrates out of the c-SMAC in the course of synapse maturation. Strikingly, synapses between naive $CD4^+$ or $CD8^+$ T cells and DCs can form in the absence of antigen or MHC molecules. This finding underscores the fact that the synapse is a flexible rather than a static entity. In conclusion, the immunological synapse can be viewed as a platform that favors interactions between costimulatory molecules, such as CD28, thereby facilitating priming of naive T cells.

8 Macrophages

Macrophages belong to the first line of innate defense to protect the vertebrate body from invasive microorganisms. When a microorganism crosses an epithelial barrier and replicates in a host tissue, it is rapidly recognized by mononuclear phagocytes, or macrophages. Macrophages differentiate from monocytes that leave the blood circulation to migrate into tissues throughout the body. Together with neutrophils, macrophages are key players of the innate immune response because they are equipped to recognize, ingest, and destroy many pathogens without the aid of an adaptive immune response. It is noteworthy that macrophages endocytose the equivalent of their entire cell volume in about 30 min and, on the other hand, secrete more distinct products than even hepatocytes, which secrete numerous serum proteins.

As macrophages react very rapidly on encountering an infecting microorganism, and since the immune system of invertebrates relies entirely on macrophage-driven innate responses, in the early 1960s, Elie Metchnikoff believed that the same would be true for vertebrates. In the meantime, we know that macrophages are of superior importance in the nonadaptive branch of our immune system, but play a less dominant role in mounting an adaptive immune response, since they express comparably low amounts of MHC and costimulatory molecules. Nevertheless, they do cross talk extensively with B and T cells and may be very important for targeted effector T-cell functions in the context of infected cells.

8.1 Evolutionary and Cellular Origin

Macrophages in culture behave very similar to amoeba feeding on microorganisms. From this and other similarities, it was concluded that the amoeba is one of the earliest or the earliest form of a phagocytic cell or macrophage. The evolutionary pathway from the ancient amoeba to the modern macrophage is unknown. However, the urgent need for an amoeba to distinguish between other amoeba that

should be left untouched and nutrient microorganisms that should be engulfed and degraded, is obvious. Hence, any surface receptor that allowed the amoeba to specifically recognize food would have been a great evolutionary progress. Such a receptor could have been the ancestor of a pattern-recognition receptor (PRRs). Such PRRs have originally been proposed by Janeway. They could serve in distinguishing bacteria or viruses, which carry regularly organized membrane or coat proteins named *pathogen-associated molecular patterns* (PAMPs) from self-proteins, which are not organized in such patterns. In the meantime, equivalents of PRRs have been found: the family of Toll-like receptors (TLRs), which recognize various types of viral or bacterial substances, such as lipopolysaccharide (LPS) or dsRNA. And indeed, TLRs are expressed on macrophages and on DCs.

All vertebrates and many invertebrates have a population of phagocytic cells that patrol their bodies. Macrophages are found in higher vertebrates in especially large numbers in connective tissue, in association with the gastrointestinal tract, in the lung, where they are found in both the interstitium and the alveoli, along blood vessels in the liver (termed *Kupffer cells*), throughout the spleen, where they remove senescent blood cells, and in the thymus, where they aid in thymocyte education.

Beyond that, macrophages are also found at several strategically important locations in the lymph nodes. They are particularly enriched in the marginal sinus zones where the afferent lymph enters the lymphoid tissue, and in the medulla, where the efferent lymph collects before having access to the blood. Here, they obviously prevent antigens from entering the blood stream, a critical event, as otherwise sepsis would occur.

Irrespective of their tissue localization, they differentiate from bone marrow–derived monocytes, which are considerably smaller and not phagocytically active.

For targeting macrophages into sites of infection, macrophages bear low levels of chemotactic receptors, such as the f-Met-Leu-Phe or 7-transmembrane-α-helical receptor. It binds to N-formylated peptides produced by bacteria, and rarely by mitochondria.

8.2 Antigen Recognition Receptors

Macrophages express several receptors on their surface, which are able to recognize a large variety of bacterial or viral constituents (Table 5). The peculiarity of some of these surface receptors is that they can bind pathogen surfaces directly. These receptors are also known as pattern-recognition molecules, as discussed above. Receptors of the innate immune system mediate different functions. Several of the

Tab. 5 Antigen recognition receptors on macrophages.

Receptor	*Ligand(s)*
Mannose receptor (CD206)	Terminal mannose or fucose of glycoprotiens or lipids
Scavenger receptor (CD36)	Anionic polymers on microbial surfaces, apoptotic bodies
f-Met-Leu-Phe receptor	N-formylated peptides
TLR2	Peptidoglycans Lipopeptides
TLR3	dsRNA
TLR4/CD14	Endotoxin (LPS)
TLR4/CD91	Hsps
TLR5	Bacterial flagellin
TLR9	CpG oligonucleotides

known receptors are phagocytic receptors that stimulate ingestion of pathogens. One of these receptors is the mannose receptor (MR). The MR is also expressed on DCs, but not on monocytes or neutrophils. The MR is a C-type lectin that binds specific arrays of terminal mannose or fucose residues of glycoproteins and glycolipids, as they are found on the surface of many bacteria and some viruses, including the human immunodeficiency virus (HIV).

Mammalian equivalents contain terminal sialic acid or N-acetylgalactosamine, which is why the MR recognizes microbes and not host cells. Bacterial carbohydrates are also bound by the glucan receptor. Another receptor is the scavenger receptor (CD36). Originally, this receptor was found to bind to aberrant low-density lipoproteins (LDLs) that are unable to be taken up via the LDL receptor. Today it is known that CD36 is involved in removal of old red blood cells and in recognition and removal of pathogens.

Moreover, macrophages express several types of Fcγ receptors suitable for binding the Fc portion of IgG antibodies. During a humoral response, soluble IgG molecules coat the microbes and thereby promote phagocytosis by macrophages. Host proteins such as IgG that promote phagocytosis of microorganisms, such as bacteria, are called *opsonins*. Further opsonins are fragments of the complement factor C3 and plasma proteins, such as fibronectin, fibrinogen, C-reactive protein, or the mannose-binding lectin. They coat microbes early in the course of an infection before specific antibodies are available and bind to receptors, such as the integrins $\alpha_v\beta_3$ and Mac-1 (CD11b/CD18).

A third class of receptors that binds microbial components induces effector molecules that mediate initiation of an innate immune response. The best-defined activation pathway of this type is triggered through the family of Toll-like receptors. The name for these receptors reflects the fact that their extracellular regions are homologous to the protein Toll of the fruit fly Drosophila. In this invertebrate, Toll triggers the production of antifungal peptides in response to fungal infection. In mammals, a Toll-family protein called *Toll-like receptor 4*, or TLR-4, signals the presence of the gram-negative cell-wall component LPS, also named *endotoxin*. The polysaccharide moieties of LPS vary strongly between bacterial strains and are major antigens in inducing an adaptive immune response. The lipid moiety, by contrast, is conserved and is a good example of a molecular pattern recognized by the innate immune system.

The LPS recognition system of macrophages consists of 3 components: (1) a plasma protein, termed *LPS–binding protein* (LBP); (2) CD14, which binds LPS bound to LBP; and (3) TLR-4, the signal-transducing receptor. Circulating LPS is captured by LBP, whereupon the LPS-LBP protein binds to CD14. Upon dissociation of LBP, LPS-CD14 engages TLR-4 leading to activation of the macrophage. This is accomplished via a signal transduction pathway sharing structural homology with the type-I IL-1 receptor pathway. Both IL-1 and LPS stimulate activation of the transcription factor NF-κB, leading to cytokine secretion and activation of inflammation.

Another mammalian Toll-like receptor, TLR-2, signals the presence of another set of bacterial components that mainly include proteoglycans of gram-positive bacteria. TLR-2 and TLR-4 induce similar but distinct signals. Macrophages are also able to sense the intracellular presence of bacterial DNA. Bacterial DNA contains unmethylated cytidine–guanidine sequences, the so-called CpG nucleotides.

Unmethylated CpG nucleotides, organized in palindromic stretches are rather rare in mammalian DNA. If enough of these CpG sequences accumulate intracellularly, macrophages start secreting IFN-γ and IL-12. According to recent results, this response is mediated via TLR-9.

Two other receptors of the TLR family, which are functionally investigated are TLR-5 and TLR-3. TLR-5 was found to bind the bacterial motor protein flagellin, whereas TLR-3 is discussed to bind viral dsRNA. The functions of other proteins of the TLR family, which comprises 10 members, are currently being explored.

8.3 Macrophage Activation

Unactivated macrophages stand out for their high rate of phagocytosis and chemotactic movements and they proliferate. This is sufficient to cope with low numbers of pathogens in an antigen-independent way. However, in order to function as an APC and to gain high antimicrobial effectiveness, macrophages have to be activated. Such an activated macrophage can damage a broad spectrum of microbes and even certain tumor cells, but also healthy self-tissue. Therefore, macrophage activation must be tightly regulated. This regulation is achieved through the requirement of two coinciding signals. Signal 1 is provided by IFN-γ, signal 2 can be provided by different means, for example, LPS, engagement of CD40L, or cognate TCR (Fig. 7).

Activated $CD4^+$ T_{H1} cells can deliver both signals very efficiently: they secrete high doses of IFN-γ and express CD40L. Activated $CD8^+$ CTLs also secrete IFN-γ; however, recognition of MHC-antigenic peptide complexes by macrophages is not always sufficient to attain full activation of macrophages – the presence of low amounts of LPS or membrane TNF-α can compensate for the lack of CD40L on $CD8^+$ T cells.

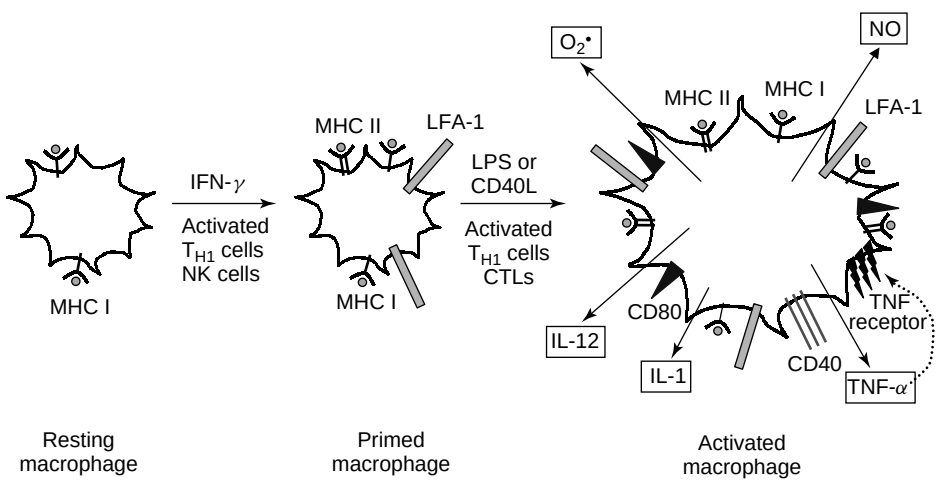

Fig. 7 The two-step model of macrophage activation. In the first step, resting macrophages are primed by IFN-γ secreted by activated T_{H1} or NK cells. Fully activated macrophages are generated in a second step upon encounter of CD40L or LPS. Activated macrophages express MHC class II molecules, costimulators, for example, CD80, CD40, and the TNF receptor. They secrete the cytokines IL-12, IL-1, and TNF-α and the radicals NO and superoxide ($O_2^{\bullet}$).

Another source of IFN-γ, which is rapidly available at the onset of an infection, is NK cells, thus supporting T_{H1}-mediated macrophage activation. T_{H2} cells counteract macrophage activation in that they secrete IL-10 but not IFN-γ. Macrophage activation by T_{H1} cells expressing CD40L and secreting IFN-γ is central to the host response to pathogens that proliferate in macrophage phagosomes, such as *Mycobacteria tuberculosis*, *Mycobacteria leprae* or *Leishmania* species. The same is true for vaccinia virus. Activated macrophages can clear internal microbes by overcoming the block in fusing phagosomes with lysosomes, so that lysis and degradation in phagolysosomes can occur. More importantly, activated macrophages produce the bactericidal nitrogen metabolite NO (see below).

By the late 1960s, the basis of acquired cellular immunity to facultative and obligate intracellular parasites was ascribed to activated macrophages. E. Metchnikoff wrote "The acquisition of immunity against microorganisms is therefore due not only to the change from negative to positive chemotaxis but also to the perfection of the phagocytic and digestive powers of the leukocytes – a general superactivity and adaptation of the phagocytic reaction is produced."

8.4 Immunological and Nonimmunological Effector Functions

8.4.1 Microbial Killing

T_{H1} cells activate infected macrophages through their TCR engaging MHC class II–antigenic complexes on the macrophage surface and the focal secretion of IFN-γ. The consequence is a series of events that converts the macrophage into a potent antimicrobial effector cell. An important intracellular event, which can be nicely observed by modern microscopy technologies, is the induced fusion of phagosomes with lysosomes, thereby exposing intracellular or recently ingested microbes to microbicidal lysosomal hydrolases leading to their destruction.

Other changes render macrophages more potent APCs and thereby help amplify an adaptive immune response: surface MHC class II, CD80, CD86, LFA-1, CD40, and TNF receptor are upregulated, thereby increasing the efficacy of macrophages in presenting antigen to resting $CD4^+$ T cells and the responsiveness to CD40L and TNF-α.

TNF-α synergizes with IFN-γ in inducing the enzyme inducible NO synthetase (iNOS), which produces the reactive nitrogen metabolite NO. NO together with oxygen radicals are secreted by activated macrophages leading to cell damage in the close neighborhood. Together with secreted proteases, NO and O_2 radicals are even able to attack large extracellular pathogens such as parasitic worms, which cannot be ingested, or intracellular pathogens that resist or survive phagolysosome fusion. The price to pay is that host tissue is destroyed as well by radicals and other toxic mediators.

Fully activated macrophages also secrete various types of cytokines, for example, TNF-α, IL-1, IL-6, and IL-12. IL-1 increases the access of effector T cell, TNF-α is an autocrine stimulus and increases the vascular permeability, IL-6 favors lymphocyte activation and IL-12 favors differentiation of $CD4^+$ T cells into T_{H1} cells and activates NK cells.

Besides priming macrophages through the focal secretion of IFN-γ, T_{H1} cells are very important in the recruitment of macrophages to sites of infection. This

is accomplished in three ways: (1) T_{H1} cells secrete the hematopoietic growth factors IL-3 and GM-CSF (granulocytes-macrophage-colony-stimulating factor), which stimulate generation of monocytes and neutrophils in the bone marrow; (2) T_{H1} cells at sites of infection secrete TNF-α and TNF-β which promote diapedesis of monocytes in endothelia; (3) T_{H1} cells in inflamed tissues secrete the chemokine MCP-1 (macrophage chemotactic protein), which attracts macrophages.

8.4.2 Tissue Remodeling

Macrophages, in particular activated ones, can secrete a number of proteins not directly related to their functions in innate or acquired immune responses but leading to local changes in the architecture of the tissue where macrophages are residing. Macrophages are the principal source of angiogenic factors, for example, vascular-endothelial growth factor (VEGF), factors that stimulate fibroblast proliferation, for example, platelet-derived growth factor (PDGF), and factors that regulate connective tissue biosynthesis, for example, transforming growth factor-β (TGF-β). Beyond that, unlike neutrophils, macrophages secrete proteases belonging to the matrix-metalloproteinases that degrade extracellular matrix proteins. At the same time, they activate fibroblasts to synthesize new matrix proteins. In settings of prolonged activation, macrophages even mediate tissue fibrosis.

8.5 Deactivated and Alternatively Activated Macrophages

In order to minimize local tissue damage and energy consumption, macrophage activation needs to be downregulated at a certain point of time. One way of achieving this is by T_{H1} cells regulating the half-life of their mRNA encoding IFN-γ. Activation of T_{H1} cells through engaging MHC class II–peptide complexes, CD80/CD86, and CD40 on the macrophage surface induces a new protein that promotes cytokine mRNA degradation, including IFN-γ mRNA. Another regulatory mechanism is the production of deactivating glucocorticoids and cytokines, such as TGF-β, IL-4, IL-10, and IL-13, especially by T_{H2} cells. Thus, the induction of differentiation of T_{H2} cells is a critical pathway for controlling the effector functions of activated macrophages.

More recently, however, IL-4 and glucocorticoids were found to induce increased expression of the MR and to enhance the capacity for endocytosis and antigen presentation of macrophages. This argues against mere deactivation and gave rise to the concept of alternative activation of macrophages. Alveolar and placental macrophages are typical examples. They express certain molecules selectively, such as the chemokine AMAC-1 (alternative macrophage activation-associated chemokine-1). AMAC-1 is related to MIP-1α, induced by classical macrophage activators, such as LPS, and inhibited by IL-4. Conversely, AMAC-1 is specifically induced by IL-4 and IL-10 and inhibited by IFN-γ. AMAC-1 is supposed to facilitate a downregulatory T_{H2} circuit in inflammatory reactions. Moreover, tumors secreting IL-10 and TGF-β may systemically and locally induce alternative macrophage activation. These macrophages may facilitate immune escape mechanisms, observed in malignant tumors, by secretion of growth and angiogenic factors supporting enhanced vascularization and nutrition of the tumor. More insight into the role of alternatively activated APCs is required.

9 B lymphocytes

B lymphocytes (short: B cells) can also serve as APCs, in particular, in the phase after a T-cell response has been initiated. In contrast to macrophages, B cells are known for their constitutive expression of MHC class II molecules. Furthermore, only B cells and some medullary thymic epithelial cells, but not other APCs, express the nonclassical MHC class II allele HLA-DO which is thought to regulate antigen processing in a B cell–specific manner. Similar to macrophages, B cells need an activation signal to express costimulatory molecules, thus, avoiding interaction with self-reactive T cells.

It has been known for a while that B cells are the key players in a humoral immune response, as they are the only cells of the immune system which can produce and secrete antibodies recognizing antigenic molecules. We will focus here on the capacity of B cells to function as APCs. Interestingly, it is still open as to how important B cells are in priming naive T cells in natural immune responses. This relates to the fact that B cells are uniquely adapted to bind and internalize mainly soluble molecules through their B-cell receptor (BCR), which is surface immunoglobulin (cf Fig. 2). However, during a bacterial or viral infection, particulate antigen is abundant but soluble antigen is rare. Therefore, it is not very likely that the limited number of B cells will detect soluble antigen that is only present at very low concentrations and that such an event is needed to prime a helper T cell. It is by far more likely that the main task of B cells is to present peptide antigens in the context of MHC class II molecules in order to receive a signal from activated T cells leading to their differentiation into antibody-secreting plasma cells.

9.1 Antigen-induced B-Cell Activation

The activation of B cells in an antigen-specific manner is initiated by the binding of antigen to the BCR. The BCR serves two key roles in B-cell activation: (1) antigen-mediated clustering of the BCR–antigen complexes initiates a signaling cascade; (2) BCR mediates endocytosis of BCR–antigen complexes into MIIC compartments where MHC class II molecules capture antigenic peptides to be presented to $CD4^+$ helper T cells.

Delivery of signals by the BCR starts upon clustering of at least two BCR complexes. *In vivo*, clustering is thought to occur by multivalent antigen. In resting B cells as well as in B cells that have undergone isotype switching, membrane molecules are associated with two other molecules, termed Igα and Igβ. Both of them are disulphide-linked heterodimers. The cytoplasmic domains of Igα and Igβ contain immunoreceptor tyrosine-based activation motifs (ITAMs), which are tyrosine enriched. Cross-linking of IgM or IgD molecules brings several ITAMs into closer proximity, thereby triggering subsequent signaling events. Initiation of signaling is accomplished via src family protein tyrosine kinases, such as Lyn, Fyn, or Blk, which are associated with the BCR and trans-phosphorylate ITAMs. Downstream signaling events ultimately activate transcription factors such as Fos, JunB, or NF-κB that induce the expression of genes whose products facilitate B-cell activation.

B-cell activation requires, in addition to antigen, second signals, which may be provided by complement proteins; for example, C3d binds to the type 2 complement–receptor CR2 on B cells (Fig. 8). Since C3d binds covalently to microbial antigens in the course of complement activation, antigen-C3d complexes can bridge the BCR and CR2. Binding of C3d to the B-cell complement receptor recruits CD19 and the tetraspanin CD81 into the complex. CR2-CD19-CD81 is often called the B cell coreceptor complex. Phosphorylation of the cytoplasmic tail of CD19 by the BCR-associated kinases leads to augmentation of the humoral immune response by 100- to 1000-fold. In summary, a proteolytic fragment of complement, which is induced by microbes, provides the second signal for B-cell activation. This is reminiscent of T-cell activation depending not only on signal 1 (TCR-MHC-peptide interaction) but also on costimulation (CD80/CD86-CD28 interaction).

9.2 Antigen Presentation Outside the Follicular Center

Within 1 to 2 days after antigen administration, B cells recognize antigen in the follicles of peripheral lymphoid organs, are activated, and begin to migrate out of the follicles toward the T-cell zones. The initial encounters between antigen-stimulated B and T cells occur at the interphase of the follicles and the T-cell zones (Fig. 9). In striking contrast to other APCs, B cells do not migrate into tissues where pathogens enter the body, but the antigens have to be transported into lymph nodes or to the spleen via the lymphatic vessels or blood stream to meet B cells.

9.2.1 B-cell Epitope versus T-cell Epitope

The three-dimensional molecular structure of an antigen that is recognized by the BCR of a B cell is termed the *B-cell epitope*, which is a conformational epitope.

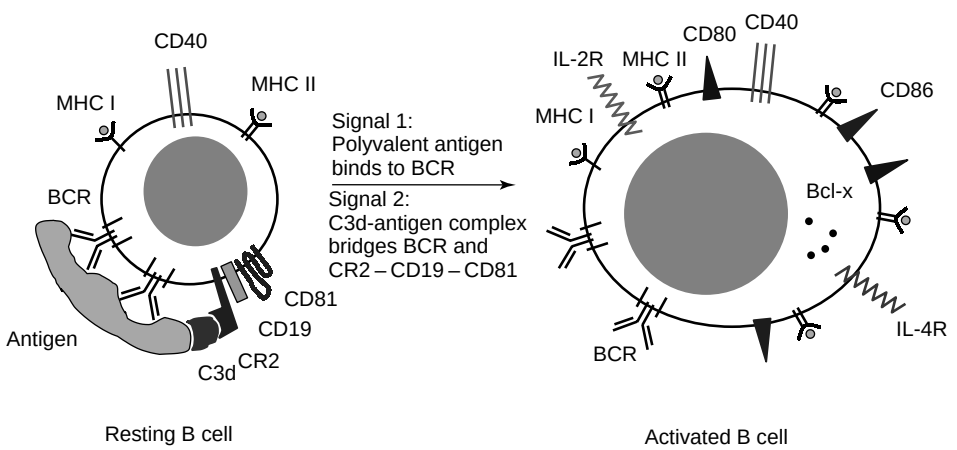

Fig. 8 Activation of B cells: resting B cells bind polyvalent antigen via the B-cell receptor (BCR) thereby generating signal1. The complement factor C3d bridges the antigen with another signaling complex consisting of the complement receptor CR2, CD19, and CD81, thereby generating signal 2. Activated B cells express costimulatory molecules, for example, CD80 and CD86, the IL-2 receptor (IL-2R), and the IL-4 receptor (IL-4R).

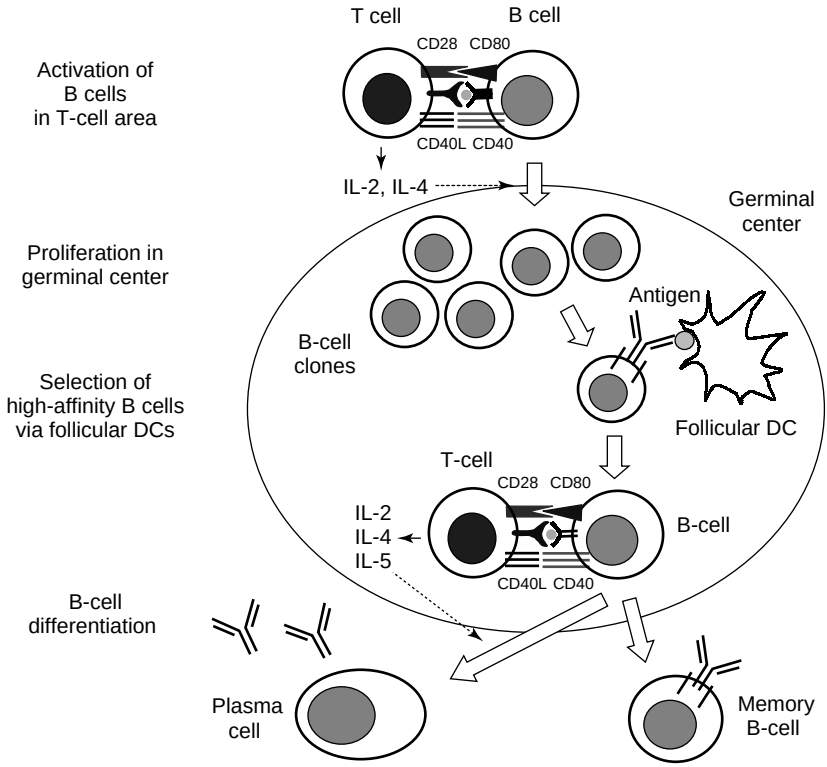

Fig. 9 T cell–B cell interactions outside and inside germinal centers: B cells encountering activated T cells in the T-cell areas of a lymph node are stimulated via cytokines to proliferate and form the germinal center. Binding to follicular DCs, which present native antigen, leads to selection of high-affinity B cells. The B cells that have survived reencounter activated T cells thereby receiving a differentiation signal. This gives rise to antibody-secreting plasma cells and memory B cells.

The peptide that is generated by proteolytic digestion of the antigen, bound and presented by MHC class II molecules, and recognized by a $CD4^+$ T cell is termed "*T-cell epitope*." It is a linear epitope.

In the majority of cases, B- and T-cell epitopes are distinct entities with different functions. This can be nicely illustrated with hapten–protein conjugates: hapten-specific B cells bind the hapten via the BCR, which mediates efficient internalization. This means that the hapten contains the B cell epitope. The hapten, however, cannot be bound by MHC class II molecules. This is why a carrier protein is necessary: the T-cell epitope to be presented to $CD4^+$ T cells is a peptide derived from the carrier protein. The requirement for MHC-associated presentation of the T-cell epitope for T-cell activation accounts for the MHC restriction of B cell–T cell interactions.

Which mechanisms prevent B cells that do not possess an antigen-specific BCR but are localized close to an antigen-specific B cell, from so-called *bystander activation*? There are several characteristics of the B cell–T cell interactions that counteract

activation of bystander B cells: (1) the first activation signal is dependent on an antigen-specific BCR; (2) binding of an antigen, endocytosis, processing and loading onto MHC class II molecules occurs at 10^4 to 10^6-fold lower antigen dose in B cells expressing an antigen-specific BCR than in other cells; (3) only B cells that remain in prolonged and tight contact with a T cell have access to cytokines secreted by T cells in a polarized fashion upon activation. This is accomplished by formation of the immunological synapse (cf. Sect. 7.3.).

9.2.2 CD40–CD40L Engagement

CD40 is a member of the TNF receptor family and constitutively expressed on the surface of B cells. Its ligand, CD40L, is only found on the surface of T cells upon seeing cognate antigenic peptide in the context of MHC class II molecules and costimulatory molecules CD80 and/or CD86. When these activated helper T cells bind antigen-presenting B cells, CD40–CD40L interactions lead to CD40 oligomerization. This oligomerization initiates enzyme cascades finally leading to the activation and nuclear translocation of transcription factors, including NF-κB and AP-1. This pathway has already been discussed in the context of T cell–mediated macrophage activation; hence, it is a general mechanism for the stimulation of target cells by helper T cells.

The importance of the CD40–CD40L interaction in humoral immunity is underscored by the observations that CD40 or CD40L gene knockout mice and humans with mutations in the CD40L gene exhibit profound defects in affinity maturation of immunoglobulin and memory B-cell generation. Interestingly, a transforming protein of Epstein–Barr Virus (EBV), which infects human B cells, utilizes the same signaling pathway as CD40, with the consequence that EBV-transformed B cells proliferate rather rapidly, leading to lymphomas.

Three helper T-cell derived cytokines, IL-2, IL-4, and IL-5, contribute to proliferation of activated B cells (Fig. 9). Besides CD40-mediated signals, it is again cytokines, mainly IL-2, IL-4, and IL-6, that stimulate antibody synthesis, the production of secreted Ig and switching from IgM to the IgG, IgA, or IgE isotype. There is a clear preference of some cytokines to induce secretion of particular Ig isotypes, for example, IFN-γ promotes secretion of IgG2a, IL-4 favors IgG1, and IgE ad IL-5/TGF-β favor IgA production.

Ig-secreting B cells are found in the red pulp of the spleen, the medulla of the lymph nodes, and the bone marrow. Many of the antibody-secreting B cells change into plasma cells, which are specialized in producing and secreting soluble antibody.

9.3 Antigen Presentation in the Germinal Center

Within 4 to 7 days after antigen exposure, some of the activated B cells migrate back toward the center of the follicle and begin to proliferate rapidly, thereby forming the germinal center. The B cells in the germinal center are named *centroblasts* and *centrocytes*. They have a doubling time of 6 to 12 h, so that 1 B cell gives rise to a clone of about 5000 centroblasts in 5 days. It was estimated that on average, three B-cell clones colonize each follicle after a single immunization. In particular, CD40–CD40L interactions and presentation of cognate peptide antigens on MHC class II molecules are required to maintain proliferation of centroblasts and to allow further differentiation.

9.3.1 B Cell/Follicular DC Contacts

Some of the mutations generated during somatic mutations in immunoglobulin genes are likely to be useful because antibodies with higher affinity are generated. However, the majority of the mutations may lead to a decline or even loss of antigen binding or to recognition of self-antigen. Therefore, the next step in the process of affinity maturation is the selection of the useful, high-affinity B cells that do not react with self.

To this end, follicular dendritic cells (FDCs) are required to engage hypermutated B cells. FDCs have long cytoplasmic processes that form a network around which germinal centers are formed (Fig. 9). FDCs do not express MHC class II molecules, and hence, are not related to classical DCs, but they express several complement receptors, Fc receptors, and CD40L. These receptors bind and display intact antigen that are decorated with Ig or complement factors. Signals generated by antigen binding – mediated by FDCs – to surface Ig on the B cell block a default cell death pathway by inducing antiapoptotic protein of the Bcl family.

With increasing numbers of antigen-specific antibodies being available, more and more of the antigen is captured, and thus, the concentration of available antigen in the germinal center decreases. Therefore, these B cells that are still able to capture antigen have to have a high affinity for the antigen – and only they will survive. The survival of the respective B cells involves uptake and processing of antigen hold on FDCs, and its presentation to T cells via MHC class II molecules (Fig. 9). This event will lead to T-cell activation – the CD40–CD40L interaction conveys long-term survival to the respective centrocyte. The net result of this selection process is a population of B cells producing antibodies with significantly higher affinities for antigen than the antibodies produced by the same clones of B cells earlier in the immune response.

Some antigen-activated B cells acquire the ability to survive for long periods of time. They are termed *memory B cells* capable of mounting rapid responses to subsequent infections. It is possible that memory cells are continually generated and maintained by antigenic stimulation, facilitated by the help of FDCs. Memory B cells typically bear high-affinity Ig. The production of large quantities of high-affinity antibodies is strongly accelerated after secondary exposures to antigen. This can be attributed to the activation of B cells in germinal centers and the rapid formation of immune complexes presented by FDCs and processed by memory B cells in order to contact cognate T helper cells.

9.3.2 HLA-DO

HLA-DO is an MHC class II molecule that is unable to bind peptide but serves as a regulator of the peptide editor and chaperone HLA-DM. HLA-DO is expressed in B cells, but not in other APCs. In further contrast to the classical MHC class II molecules and to HLA-DM, HLA-DO expression is initiated only after B-cell development is complete, that is, when B cells start becoming capable of interacting successfully with T cells.

HLA-DO is tightly bound to HLA-DM and is rapidly degraded in the absence of HLA-DM. HLA-DM–HLA-DO complexes are poorly active in removal of the CLIP peptide, especially at moderate pH found in early and late endosomes, suggesting that HLA-DO is a negative regulator of antigen presentation. However, at the

same time, HLA-DO acts as a cochaperone of HLA-DM, which is particularly relevant at the low pH found in lysosomal MIICs of B cells. Thus, by limiting the catalytic function of HLA-DM in all but the most acidic MIICs, HLA-DO appears to selectively promote loading and presentation of antigenic peptides derived from proteins internalized through the BCR. The rationale is that very low pH, a high redox potential, and a high abundance of proteases, typical characteristics of lysosomal MIICs, are required for efficient processing of antigens tightly bound to BCRs.

Strikingly, the capacity of B cells to present antigen to helper T cells is controlled via differential expression of HLA-DO in quiescent B cells as compared to activated B cells in germinal centers. In peripheral blood B cells, 50 to 100% of HLA-DM is associated to HLA-DO. This leads to inefficient CLIP removal, and hence to high surface levels of MHC class II-CLIP complexes and to a reduced capacity to process self- or foreign antigen. The conclusion is that due to the comparably high abundance of HLA-DO, peripheral B cells are reduced in their potential to present foreign antigen. The role of class II-CLIP complexes, as typical representatives of self, in peripheral tolerance induction with regard to self-reactive T cells has to be elucidated.

In germinal centers, however, HLA-DO is markedly downregulated so that CLIP is efficiently released from MHC class II molecules and antigenic peptides derived from internalized BCR–antigen complexes have access to the binding cleft of class II molecules. As 20 to 30% of HLA-DM is still associated to HLA-DO, peptide receptive "empty" class II molecules are chaperoned by HLA-DM–HLA-DO complexes until BCR-mediated uptake of foreign antigen occurs. In summary, regulated HLA-DO expression allows optimization of antigen processing and presentation to helper T cells in germinal centers during the development of humoral immune responses.

10 Dendritic Cells (DCs)

Dendritic cells (DCs) represent a highly heterogeneous population of APCs, at the same time being the most potent APCs. They are bone marrow–derived and reside in most peripheral tissues as sentinels of the adaptive immune system. As DCs, similar to cells of the innate immune system, express various types of TLRs that mediate recognition of microbial danger motifs, DCs are also most critical in bridging innate and adaptive immunity.

DCs are specialized for the uptake, transport, processing, and presentation of antigens to T cells. Any encounter with microbial products or tissue damage initiates the migration of DCs out of peripheral tissues to lymph nodes and their differentiation into the mature state. Mature DCs express strongly upregulated levels of MHC and costimulatory molecules so that, in lymph nodes, DCs can trigger an immune response by any T cell with a receptor that is specific for MHC-foreign peptide complexes on the DC surface. Since subsets of DCs are constantly carrying self-antigens into lymph nodes without receiving a maturation stimulus, DCs are thought to play a major role in the maintenance of peripheral tolerance against self as well.

10.1 DC Subsets

Phagocytic cells carrying numerous dendrites and expressing considerable amounts of MHC and costimulatory molecules in their unactivated state are found in distinct microenvironments of our body, for example, in peripheral tissues, in lymph nodes, and in the thymus. In addition, DCs do not only interact with T cells, but also with B cells and with NK cells. These aspects render it most likely that the various and often opposing functions ascribed to DCs can only be performed by different sets of DCs. According to the functional plasticity model, only a single DC lineage (hematopoietic precursor, DC precursor, immature DC, mature DC) exists and local environmental factors cause functional diversity (Fig. 10). According to the specialized lineage model, however, specialized DC subtypes are the product of entirely separate developmental lineages, with DC precursors being already functionally committed. It appears that in reality, we are facing a complicated mixture of both models.

10.1.1 Myeloid Versus Lymphoid DCs

Distinct DC subtypes were initially more evident among mouse DCs than among human DCs because of the availability of

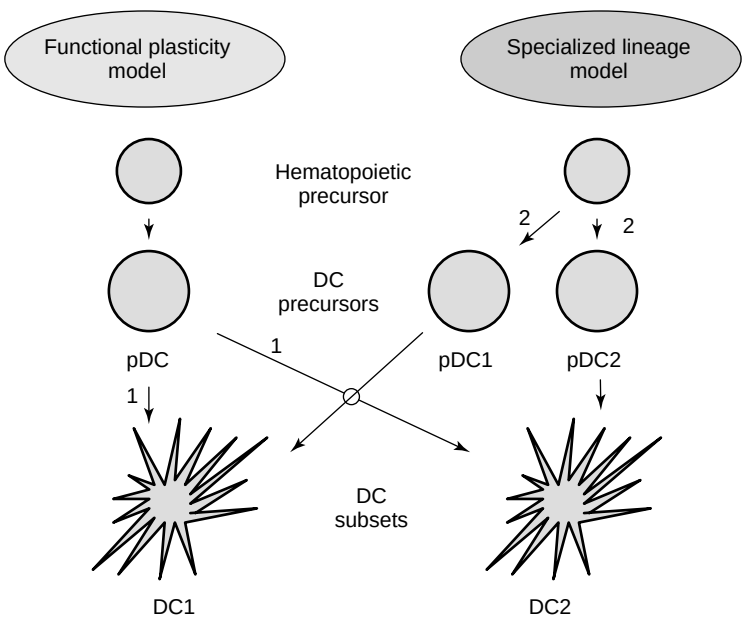

Fig. 10 Generation of functionally distinct DC subsets. According to the functional plasticity model, all DCs belong to a single hematopoietic lineage, the subtypes DC1 and DC2 being generated by local environmental factors in the periphery (1). The specialized lineage model proposes that DC subsets originate from early divergences in the development, triggered by the microenvironment of the bone marrow (2), giving rise to distinct precursors, pDC1 and pDC2. Subsequently, pDC1 and pDC2 function as the precursors of the DC1 and DC2 subsets respectively.

different murine lymphoid tissues and the expression of distinct markers not present on human DCs. Hence, there was evidence that the $CD8^+$ DCs of the thymus shared early steps of development with T cells. They were thought to derive from lymphoid-restricted precursors and were named *lymphoid DCs*. All other DCs being $CD8^-$ were thought to originate from myeloid-restricted precursors; they were called *myeloid DCs*. In the meantime, when myeloid- and lymphoid-restricted precursors were isolated from bone marrow, it became evident that each precursor could produce all the mature splenic and thymic DC subtypes. Besides CD8 and CD4, the integrin α_M chain (CD11b), the lectin DEC205 (CD205) and the integrin α_X chain (CD11c) are used as surface markers to distinguish murine DC subtypes. Currently, the terms *myeloid* and *plasmacytoid* (Table 6) are used to distinguish DC lineages in both mouse and man.

10.1.2 Human DC Subtypes

In contrast to the situation in mice, only blood and bone marrow can be used as an *ex vivo* source of human DCs. Human blood DCs are highly heterogeneous in their marker expression, with many markers reflecting differences in the maturation or activation state rather than separate lineages. A comparison to mouse subtypes is difficult, since human DCs lack expression of CD8. Only human thymic DCs make a strong case for subset segregation, similar to that in mouse: most human thymic DCs are $CD11c^+$ $CD11b^-$ and lack myeloid markers, and so resemble murine thymic $CD8^+$ DCs.

As yet, most insight into DC subsets has come from studies of their development in culture from precursors. Essentially three distinct precursor types have been used: $CD34^+$ bone marrow–derived precursors, blood monocytes, and plasmacytoid cells (Table 7).

$CD34^+$ precursor pathway The earliest precursors used are $CD34^+$ and derived from bone marrow or umbilical-cord blood. Liquid culture with granulocyte-macrophage colony-stimulating factor (GM-CSF) and TNF-α leads to two

Tab. 6 Tissue distribution of murine DC subtypes.

Tissue	*Percentage of total DCs*[a]			
	Myeloid DCs		*Plasmacytoid DCs*	
	$CD4^+$ $CD8^-$ $CD205^-$ $CD11b^+$	*$CD4^-$ $CD8^-$ $CD11b^+$*	*$CD4^-$ $CD8^+$ $CD205^+$ $CD11b^-$ lymphoid*	*$CD4^-$ $CD8^+$ $CD205^+$ $CD11b^+$ Langerhans DCs*
Spleen	56	20	23	<1
Thymus	–	–	60	–
Mesenteric lymph nodes	4	63	19	<4
Skin-draining lymph nodes	4	37	17	33

[a] Adapted from Shortman, K., Lin, Y.J. (2002) *Nature Reviews* **2**, 151–161.

Tab. 7 Different subsets of human dendritic cells (DCs).

Name	*Monocyte-derived DC*	*Langerhans DC*	*IL-3 receptor$^+$ DC*
Origin	Blood	Bone marrow cord blood	Blood
Precursor	CLA$^-$/CD14$^+$	CD34$^+$/CLA$^+$	pDC2
Lineage	Myeloid	Plasmacytoid	Plasmacytoid
Phenotype (immature):			
MHC class II	++	++	−
CD207 (Langerin)	−	+++	−
CD11c	+	+	−
IL-3 receptor	−	−	++
CD4	+	+	+++

apparently separate pathways of DC development, which can be distinguished by the expression of a skin-homing factor, known as *cutaneous lymphocyte-associated antigen* (CLA).

One pathway (via CD34$^+$/CLA$^+$ intermediates), depending on TGF-β, leads to DCs resembling Langerhans cells: they express the Langerhans cell–associated antigens Langerin and E-cadherin, CD11c, CD1a, and have Birbeck granules, characteristic structures known from electron microscopy studies of Langerhans epidermal DCs (Table 7).

The second pathway (via CD34$^+$/CLA$^-$, CD14$^+$ intermediates, which resemble blood monocytes) produces DCs reminiscent of intestinal DCs (Table 7). They lack Langerin and E-cadherin, but express the tetraspanin CD9, the coagulation factor XIIIa and CD68.

CD14$^+$ monocyte pathway Peripheral blood monocytes are the most commonly used precursors for generating human DCs in culture. CD14$^+$ monocytes differentiate into DCs (termed DC1) in the presence of GM-CSF and IL-4. In the presence of M-CSF, macrophages are generated. The monocyte-to-DC differentiation can also be accomplished by seeding monocytes onto a layer of endothelial cells over a collagen matrix, mimicking the entry of monocytes into tissues. Most of the monocytes that have migrated into the collagen matrix remain there and become macrophages. A subset of them, however, migrate back through the endothelial barrier and become DCs (Table 7).

Plasmacytoid cell pathway Human plasmacytoid cells, termed DC2, were found by their plasma cell–like morphology and their unique surface phenotype: they are strongly positive for the IL-3 receptor, but negative for CD11c and for myeloid mineage markers (Table 7). Strikingly, they produce mRNA for germ-line Igκ and pre-T-cell receptor α. Plasmacytoid cells are found in blood and many lymphoid tissues. They respond to viral and microbial stimuli by producing IFN-α and IFN-β.

10.2 Antigen Uptake Mechanisms

Tissue DCs capture pathogens, infected cells, dead cells, or their derived products, and also soluble material to use for antigen processing and presentation. Importantly,

binding and uptake of certain pathogen-derived components are in many cases linked to DC maturation. During DC maturation, they downregulate their endocytic capacity. The consequence is that antigen uptake is no longer possible during migration to and in the lymph nodes, but only in peripheral organs. Down-modulation of endocytosis is achieved in at least two ways: (1) a decrease in surface expression of several antigen-binding receptors; (2) inactivation of the GTPases cdc42 and rac1 widely abolishes macropinocytosis and phagocytosis.

10.2.1 Macropinocytosis Versus Phagocytosis

Macropinocytosis is a cytoskeleton-dependent type of fluid-phase endocytosis that certain growth factors can induce in macrophages. In immature DCs, however, macropinocytosis is a constitutive phenomenon. It allows DCs to sample large amounts of extracellular fluid in short periods of time. This is the nonspecific way of DCs to search for antigenic material. Antigens that have been captured by macropinocytosis have access to MIICs and hence can be loaded onto MHC class II or CD1 molecules.

Phagocytosis is initiated by the engagement of specific surface receptors; hence, it is to a certain extent antigen-specific. Immature DCs phagocytose whole bacteria, irrespective of whether they are gram-positive or gram-negative, but also yeast cells and hyphae. As part of the GALT, interstitial DCs penetrate tight junctions so that their dendrites have access to bacteria in the gut lumen.

Immature DCs also internalize cell debris from cells that have undergone cell death by apoptosis or necrosis. Human monocyte-derived DCs take up apoptotic and necrotic bodies derived from B or T cells, virus-infected apoptotic monocytes, or tumor cells, including melanoma cells or various carcinoma cells. Moreover, DCs have been shown to transport apoptotic intestinal epithelial cells to T-cell areas of mesenteric lymph nodes in rats. Phagocytosis of apoptotic bodies mainly occurs through a complex including the scavenger receptor CD36 and the integrins $\alpha_V\beta_5$ and $\alpha_V\beta_3$. The receptors needed for phagocytosis of necrotic cells are less well defined.

10.2.2 Receptors for Endocytosis

Receptors for the F_c portion of immunoglobulin belong to the prominent molecules on the DC surface: human monocyte-derived DCs express mainly FcγRII and FcαR, LCs express FcγRI and FcεRI, whereas blood DCs are positive for FcγRII and FcγRI. Recently, a novel member of the Ig superfamily, named Ig-like transcript (ILT)-3 was found to facilitate antigen processing in DCs, but also in monocytes and macrophages.

In contrast to macrophages, immature DCs express only the complement receptors CR3 and CR4, but not CR1 or CR2. Similar to macrophages, immature monocyte-derived DCs, blood DCs, and interstitial DCs of the dermis display high levels of the MR. The MR not only allows DCs to endocytose a large variety of bacterial and yeast antigens, but also desialylated immunoglobulin. Strikingly, LCs do not express the MR, but Langerin (CD207), which is also a lectin with mannose specificity. Unlike the MR, CD207 induces the formation of Birbeck granules. Finally, DCs are also very potent in binding and taking up Hsps, such as hsp70 and gp96. Recently, CD91, known as the α2−macroglobulin receptor, and CD40 were described to mediate endocytosis of Hsps of the Hsp70 and Hsp90

protein family. These findings are intriguing, as endogenous as well as foreign Hsps are discussed to mediate maturation of DCs. In particular, the activatory capacity of necrotic cells has been ascribed to Hsps

10.2.3 The Adhesion Receptor DC-SIGN

DC-SIGN (DC-specific ICAM-3-Grabbing Nonintegrin), also termed CD209, is a mannose-binding C-type lectin that is strongly and selectively expressed on DCs, irrespective of whether they are immature or mature.

DC-SIGN recognizes with high affinity the adhesion molecule ICAM-2 on vascular and lymphoid endothelia. ICAM-2 serves as a rolling counterreceptor, thereby enabling DC-SIGN$^+$ DCs to tether to and roll along endothelia. This is a prerequisite for trans-endothelial migration into peripheral tissues and secondary lymphoid organs. Thus, one function of DC-SIGN is to function as a DC-specific rolling receptor.

Upon reaching T cell–rich areas in secondary lymphoid organs, DC-SIGN mediates initiation of T cell–dependent immune responses: it enables transient DC–T-cell interactions through binding to ICAM-3, which is expressed on the surface of naive T cells. In contrast, DC-SIGN does not bind to the principal ligand of LFA-1, ICAM-1.

Finally, DC-SIGN functions as an HIV-1 trans-receptor important in the dissemination of HIV-1. HIV-1 is captured on DCs present in the periphery by DC-SIGN, which binds to the HIV-1 coat protein gp120. Thus, HIV-1 is transported by DCs that migrate into lymphoid tissues where DC-SIGN-associated HIV-1 efficiently infects target CD4$^+$ T cells.

10.3 Maturation of DCs

The various subtypes of human DCs described above vary considerably with regard to the expression of marker proteins and their behavior in the immature state. However, as soon as they are subjected to inflammatory stimuli, they share a common program, termed maturation, which transforms immature into mature DCs that are highly efficient in priming and activating T cells. They are by far more potent than activated macrophages or B cells in initiating a cellular immune response. This is mainly due to the fact that key proteins, for example, MHC molecules, costimulators, and accessory molecules, are much more abundant on mature DCs than on stimulated macrophages or B cells.

10.3.1 Maturation Stimuli

Immature DCs respond to two types of maturation stimuli: (1) direct recognition of pathogens via TLRs and other PPRs; and (2) inflammatory cytokines and other exogenous or endogenous mediators that are secreted during an infection or in other acute alarm situations (Fig. 11). In detail, DCs can recognize bacterial peptidoglycans, lipopeptides, and mycoplasma lipoproteins via TLR2, LPS via TLR4, viral dsRNA via TLR3 and certain DNA oligonucleotides via TLR9. It should be emphasized that TLR2 is not present on the cell surface, but specifically recruited to yeast-containing phagosomes. Likewise, TLR4 is not compulsory for internalization of LPS-carrying bacteria. This suggests that TLRs are involved in discrimination of microbes and not necessarily in their uptake. Apart from that, DCs have receptors in order to sense infections and

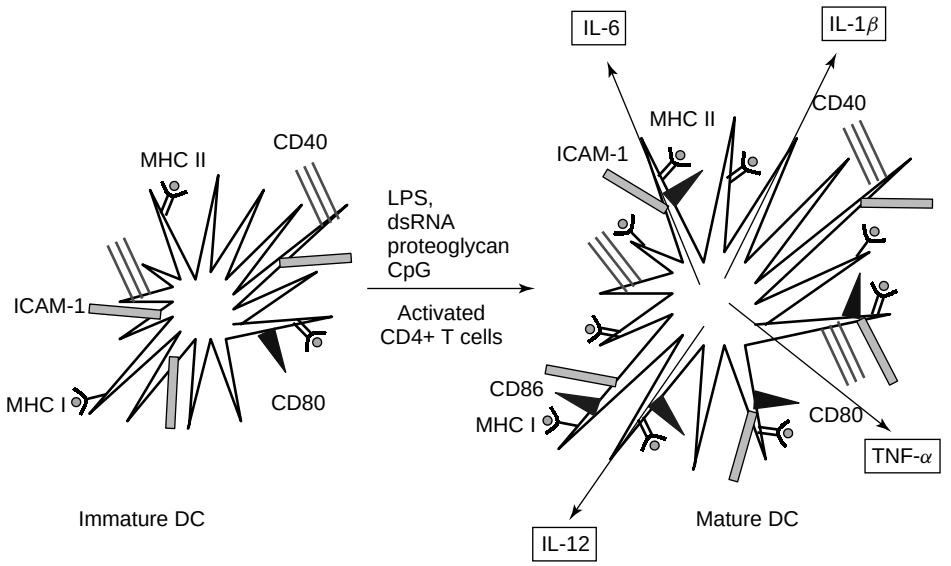

Fig. 11 Maturation of dendritic cells (DCs). Several microbial signals, including LPS, dsRNA, proteoglycan, and CpG, or the encounter of activated T helper cells induces the mature phenotype. Mature DCs strongly upregulate MHC, costimulatory and adhesion molecules and secrete the proinflammatory cytokines IL-12, IL-6, IL-1β, and TNF-α.

danger indirectly, via binding of TNF-α, IL-1β, or PGE_2, whose secretion is triggered by pathogens.

Moreover, whole viruses or bacteria decorated with immunoglobulins induce DC maturation via binding to FcR, including FcγRI, FcγRIII, and FcγR. Even more importantly, activated $CD4^+$ T cells provide a strong stimulus via CD40L that triggers CD40 of DCs. This signal appears to be essential for conditioning DCs in such a way that they are able to fully activate certain types of $CD8^+$ CTLs.

There is accumulating evidence that cell death gives rise to signals that render DC mature, as well. In particular, necrosis induces DC maturation, however, the mediators of this activation are still unknown. Apoptotic cells, at least after heat stress, are also capable of stimulating immature DCs. Nucleotides, such as ATP and UTP, and Hsps, including gp96, Hsp90, and Hsc70 are liberated extensively in the course of necrosis and have been proposed to function as danger signals being able to trigger DC differentiation. However, contaminations with LPS have been implicated in this context, so that it is doubtful, as yet, whether Hsps per se or Hsp-LPS complexes have induced DC maturation. Moreover, Hsp60 and Hsc70 have been described to activate macrophages through a CD14/TLR4 complex, but receptors on DCs are not characterized, yet.

10.3.2 The Danger Hypothesis

The model of self–nonself discrimination foresees that the baseline state of the immune system is "OFF." To initiate a cellular immune response, the key players, namely, APCs and T cells have to be activated. T cells get activated by the encounter of mature APCs. Such mature

APCs provide signal 1 via MHC-peptide complexes engaging TCRs. Since APCs appear to be unable to distinguish self-peptide from foreign peptide and since self-reactive T cells do exist in our body, the safe guard of the self–nonself discrimination model is signal 2. Signal 2 is only provided via costimulators of APCs that have recognized patterns of non-self molecules found only on evolutionary distinct organisms, such as bacteria or viruses. This is thought to be accomplished via PRRs, for example, TLRs. However, even Janeway, who pioneered the idea of pattern recognition serving for costimulation, pointed out that PRRs cannot explain autoimmunity or immune responses to tumors or transplants. Moreover, it is still open how the immune system deals with a changing self. The self–nonself discrimination model, for example, fails to give us a clue why vertebrates do not attack themselves at puberty and why females do not reject their own newly lactating breasts when they begin to produce milk proteins that were not part of self until that time.

The danger model pioneered by Matzinger gives an answer to these questions. This model is based on the hypothesis that the ultimate controlling signals are endogenous, not exogenous. They are the alarm signals that emanate from stressed or injured tissues (Fig. 12). Hence, cells of normal bodily tissue, when distressed, send signals that are called *danger signals*, which serve to activate local APCs. Healthy cells, in fact, may send calming signals to local APCs, whereas cells that are damaged or die by necrosis should favor maturation of APCs. In contrast, cells that die by apoptosis, a process of normal programmed cell death, should merely send signals that may favor phagocytosis but should not induce the expression of costimulatory molecules. This view is consistent with findings where necrotic cells do induce maturation of DCs, but not apoptotic cells. This correlates with the liberation of Hsps from necrotic but not apoptotic cells. Strikingly, apoptotic cells that have undergone heat stress were also able to activate DCs. As yet, Hsps are the only candidate for endogenous mediators of danger.

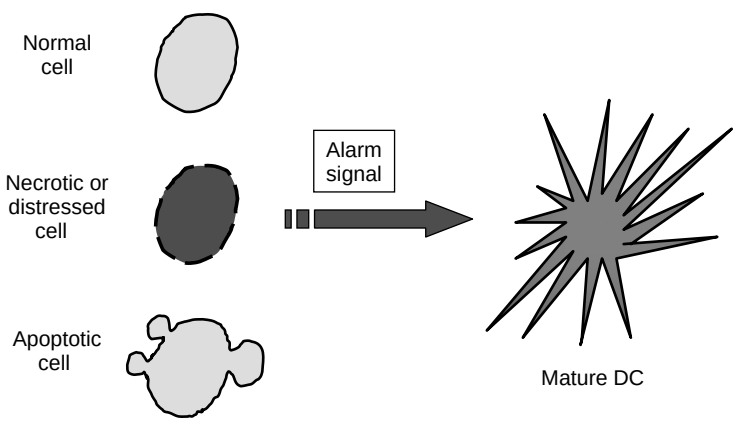

Fig. 12 The danger model DCs receive an alarm signal from injured, distressed, or necrotic cells leading to their maturation. This is not the case with normal cells or in cells dying by apoptosis.

10.4 Antigen Processing and Presentation

10.4.1 Classical and Cross-presentation via MHC Class I

Like other APCs and other nucleated cells, DCs present self- or virus-derived endogenous antigens to $CD8^+$ CTLs in an MHC class I–restricted manner. In contrast to other cells, DCs express low levels of the immunoproteasome already in the immature state and upregulate it upon maturation. Moreover, the abundance of MHC class I–peptide complexes increases during maturation. However, in contrast to the MHC class II pathway, mature DCs still synthesize and replace surface MHC class I molecules in their mature state. This finding raises the question, whether DCs that have captured a class I–restricted antigen in peripheral tissues, are the same DCs that will present this antigen to $CD8^+$ T cells in secondary lymphoid organs. Alternatively, DCs may capture antigens in the lymph node. In this case, DCs should be able to take up exogenous antigen and transfer it onto MHC class I molecules. This nonclassical mechanism had already been proposed by Bevan and colleagues: they showed that priming of CTL responses *in vivo* can occur after presentation of exogenous antigens by class I MHC molecules. These phenomena were referred to as *cross-presentation* of antigen and *cross-priming* of $CD8^+$ T cells.

In the meantime, cross-presentation has been shown to be a major pathway utilized by DCs in the context of various sources of exogenous antigens. Physiologically, phagocytosis appears to be a major route for antigen uptake and cross-presentation, as FcR-mediated uptake of immune complexes or opsonized dead cells promotes cross-priming. Furthermore, phagocytosis of bacteria, bacteria-derived antigens, and apoptotic cells results in cross-presentation of bacterial antigens and the same has been observed for viral and tumor antigens. Likewise, peptides bound to Hsps, including gp96, Hsp90, Hsp60, and Hsc70, are cross-presented by mature DCs.

Two pathways for cross-presentation were reported: the first foresees peptide loading in the endocytic system, and hence is TAP- and proteasome-independent. Recycling MHC class I molecules are, indeed, found in endosomes of immature DCs, where low pH is supposed to favor exchange of prebound peptide for exogenous peptide. The second pathway is TAP-dependent and was reported in DCs in several experimental settings. This pathway foresees the existence of a membrane transport system linking the lumen of endosomes and the cytosol. Other cells, for example, thymic epithelia cells or liver sinusoidal epithelial cells, are also able to present soluble exogenous antigen in a TAP-dependent fashion, albeit less efficiently than DCs.

As an alternative to explain cross-presentation, one could envisage that ER compartments fuse with nascent endosomal membranes before endosomes separate from the plasma membrane, as most recently shown for macrophages. This would allow loading of exogenous antigens onto MHC class I molecules in the ER without the need for a specialized transporter in the endosomal membrane. Physiologically, this mechanism could serve to feed the plasma membrane with phospholipids from the ER to compensate the loss of membranes due to endocytosis.

10.4.2 MHC Class II–restricted Antigen Presentation

The cell surface display of MHC class II–peptide complexes is tightly regulated

during the life cycle of DCs. An immature DC exposes around 10^6 MHC class II molecules on the surface. This is already more than activated B cells or macrophages can ever express, however, 5 to 10 times less than in mature DCs. Several reasons account for that: the efficiency of major proteases involved in antigen degradation in immature DCs, for example, cathepsins B and S, is low. This may be due to comparably high levels of endogenous protease inhibitors such as cystatin C and the Ii splice variant Iip41. Another aspect is that not all endosomes or phagosomes may have access to MIICs, where MHC class II molecules are loaded with peptide. Finally, MHC class II–peptide complexes have a short half-life ($\leq$10 h) in immature DCs, since they are constantly internalized and transported to recycling endosomes, where peptide exchange occurs, or is degraded in lysosomes. Nevertheless, it should be pointed out that immature DCs are able to present self- and foreign antigen peptides to $CD4^+$ T cells. According to current concepts, this process serves mainly to tolerize T cells against self-antigens.

Inflammatory signals activate a program leading to profound changes in antigen processing for class II molecules. MHC class II biosynthesis is transiently upregulated, the activity of processing proteases increases accompanied by their relocalization into MIICs and recycling of class II molecules ceases widely. Following peptide loading in MIICs, subsets of class II molecules associate with the costimulator CD86 and the tetraspanins CD82 and CD63, thereby becoming a part of multimeric clusters that are transported to the cell surface. At the cell surface, class II peptide complexes remain organized in a multimerized fashion, mainly associated to tetraspanins. The tetraspan microdomains, which are strongly upregulated in mature DCs, not only concentrate MHC-peptides complexes but also enrich MHC class II molecules loaded with particular peptide. This increases the local avidity to T cells and, hence, increases the efficacy of T-cell activation. This is further supported by the fact that the mature DCs no longer endocytose class II molecules so that the latter become rather long-lived. After 10 to 20 h of maturation, no further class II molecules are synthesized and the activity of HLA-DM is strongly downregulated so that peptide loading ceases. Thus, mature DCs that have reached the T-cell areas of lymph nodes mainly present those peptide antigens to $CD4^+$ T cells that they have captured before or during induction of maturation in peripheral tissues.

10.5 T Cell Priming and Polarization

10.5.1 Priming of $CD8^+$ and $CD4^+$ T Cells

Several lines of evidence support the concept that DCs are the principal APCs for T-cell priming. This is based on *in vitro* observations, where DCs were compared to other APCs for priming alloreactive, naive T cells or the expansion of naive precursors from polyclonal populations. Furthermore, injection of antigen-loaded DCs induces potent $CD4^+$ and $CD8^+$ T-cell primary responses. By immunofluorescence on lymph node sections, interactions of DCs and T cells could be directly visualized. Beyond that, DCs break peripheral tolerance against transplantation antigens and tumor antigens. Moreover, the exclusive expression of MHC class II molecules in DCs was sufficient for the establishment of central tolerance through Vβ-specific deletion of T cells by a superantigen presented by DCs in the thymus medulla.

The outstanding T-cell priming capacity of mature DCs is mainly due to the enhanced surface levels of costimulatory CD80 and CD86 and adhesion molecules, but also connected to a rapid relocalization of antigen-bearing DCs to the T-cell zones of secondary lymphoid organs. The *in vivo* relevance of DC maturation in T-cell priming is particularly evident in the case of $CD8^+$ T cells. Although there are examples of direct priming of CTLs by DCs without the involvement of $CD4^+$ T cells, numerous cases have been described, where $CD8^+$ naive T cells are dependent on $CD4^+$ T-cell help. It was found that CD40 ligation on DCs by CD40L on $CD4^+$ T cells was necessary and sufficient to confer to DCs the ability to prime CTLs. At the same time, DCs are rendered mature (Fig. 13).

This mechanism allows a temporal dissociation of interactions between DCs, $CD4^+$ helper T cells, and $CD8^+$ CTLs: once licensed by $CD4^+$ T cells, mature DCs are competent to prime $CD8^+$ T cells. This model, also known as the *license-to-kill model* bypasses the requirement of a tricellular interaction of DCs, activated helper T cells and naive CTLs in lymph nodes, which is rather unlikely to occur. Apart from helper T cells, pathogen-derived products, such as LPS, viruses or apoptotic bodies carrying CD40L have been shown to deliver licensing signals to DCs.

10.5.2 T Helper Cell Polarization

In the most recent literature, the issue of T-helper cell polarization by DCs has received considerable attention. Various pathways of regulation are operative. Immature DCs *per se* favor T_{H2} polarization, but they can be reinstructed by the characteristics of the antigen or its accompanying adjuvant to induce a T_{H1}, T_{H2}, or an undifferentiated T_{H0} response (Table 8).

Microbe-derived pattern molecules, such as LPS or CpG oligonucleotides, drive T_{H1} polarization by induction of IL-12, IL-6, and IL-1β secretion through binding to TLRs. However, microbial products often need additional host-derived microenvironmental instruction to

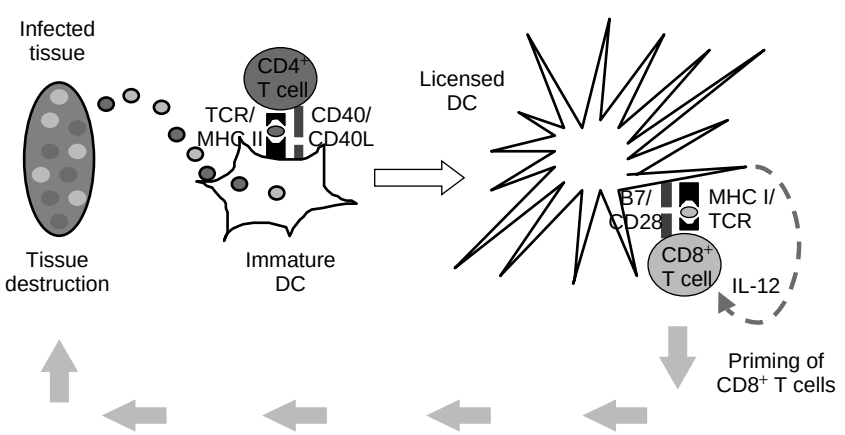

Fig. 13 The license-to-kill model of DCs. Immature DCs that encounter activated $CD4^+$ Thelper cells receive a signal through CD40/CD40L engagement. DCs that have received this signal are able to render cytotoxic $CD8^+$ T cells highly potent in killing infected target tissues. Thus, DCs conditioned by Thelper cells bear the license to kill.

Tab. 8 T helper cell polarization mediated by DCs.

	T_{H1} *bias*	T_{H2} *bias*
Antigen dose	High	Low
Type of antigen	*Candida albicans yeast* Toxoplasma antigen	*Candida albicans hyphae* Allergens
Adjuvants	CpG oligonucleotides LPS	PGE_2 Cholera toxin
MHC-TcR interaction	High affinity Short interaction	Low affinity Sustained interaction
Stimulator/responser ratio	High	Low
Costimulators	ICAM-1 CD40	OX40-L ICOS-L
DC state	Mature	Immature
DC type	$CD1a^+$ (myeloid)	$CD1a^-$ (myeloid)
DC localization		Mucosal DC
Cytokine status	High IL-12 Low IL-10	Low IL-12 High IL-10 High IL-6

induce stable IL-12 production by DCs, for example, high doses of antigen, IFN-γ, or IL-4. Interestingly, monocyte-derived DCs prime strong T_{H1} responses early after onset of maturation, whereas at later time points, the same cells prime T_{H2} or T_{H0} responses. Likewise, when DCs stimulate naive T cells at low stimulator/responder ratios, T_{H2} polarization is favored despite the presence of IL-12. Recently, it was proposed that different developmental lineages of DCs could induce either T_{H1} or T_{H2} responses, hence the terminology DC1 or DC2 (Fig. 10) was introduced. Monocyte-derived or myeloid DCs were thought to prime T cells towards the T_{H1} phenotype, whereas plasmacytoid DCs were believed to exclusively prime towards the T_{H2} phenotype. However, this is not the case, as evidenced by at least two recent observations: CD1a-negative subpopulations of monocyte-derived DCs do not secrete IL-12, even after stimulation with LPS, IFN-γ, or CD40L. Likewise, plasmacytoid DCs can be very potent in priming naive T cells towards T_{H1} differentiation, especially after viral infection.

In conclusion, T helper cell priming by DCs is very flexible, allowing the simultaneous generation of T_{H1}, T_{H2}, and T_{H0} responses.

10.6 Immune Evasion and Plasticity of DCs

In general, DCs, after having met viruses or other pathogens, will mature and secrete IL-12, thereby skewing T_{H1} differentiation. However, infection of immature human DCs with herpes simplex virus type 1 (HSV-1) or vaccinia virus has been shown to inhibit DC maturation. Likewise, the bacterial metabolite n-butyrate, which is found in high concentrations in the gastrointestinal tract, can interfere with LPS-induced IL-12 release, but not of IL-10. IL-10 released during infection with *Bordetella pertussis* not only promotes the differentiation of monocytes into macrophages but also the differentiation of T regulatory cells, which suppress protective T_{H1}

responses (Fig. 14). In response to viruses and *Streptococcus pneumoniae*, DCs can activate B cells via secretion of proinflammatory type 1 interferons, thereby promoting immunoglobulin isotype switching. Most recently, it has been observed that there are reciprocal activatory interactions between NK cells and DCs. The cross talk between both cell types appears to be relevant in innate antitumor responses.

10.7 DCs in Immunotherapy

The superior role of DCs in triggering antigen-specific T-cell responses has provided the rationale for developing new vaccination strategies based on the injection of DCs loaded with antigen *in vitro*. In most immunotherapy protocols, autologous $CD34^+$ precursor-derived DCs or monocyte-derived DCs from healthy volunteers or patients are being used.

In vaccination against tumors, the ultimate goal is to prime $CD4^+$ and $CD8^+$ T cells with antitumor reactivity. To achieve this, efficient and stable loading of MHC molecules of DCs with cognate tumor antigens is necessary. Two strategies are employed to provide antigens: an endogenous and an exogenous way. Endogenous antigen expression was obtained by using bulk RNA prepared from tumors or retroviral, adenoviral, or poxviral vectors to transduce DCs with genes coding for the respective tumor antigens. Another

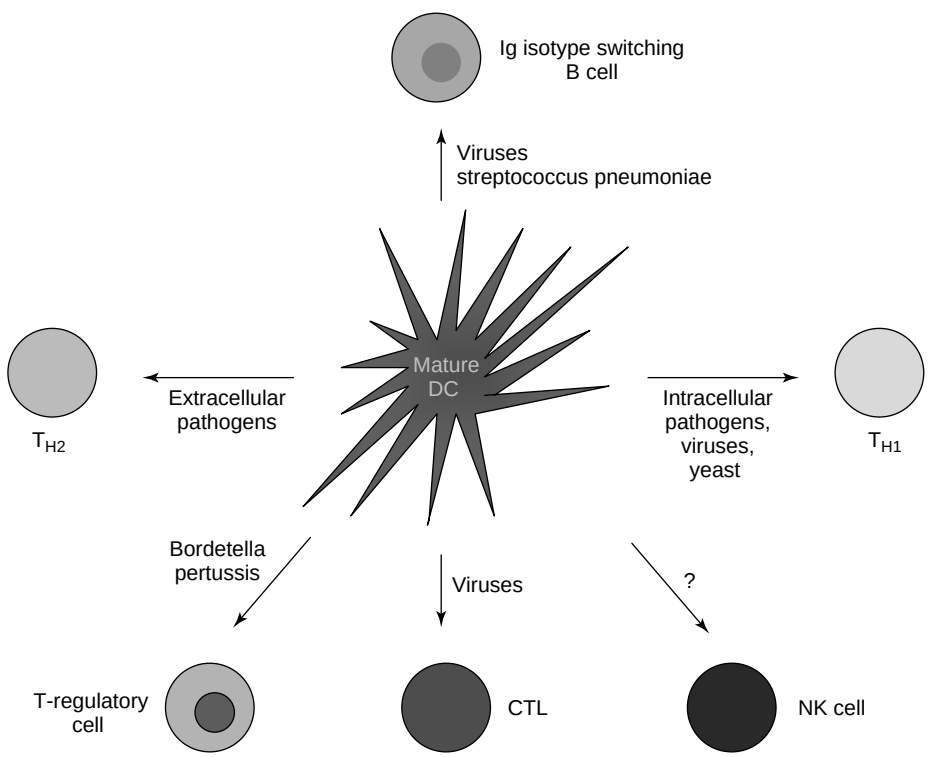

Fig. 14 The plasticity of DCs. Mature DCs in the draining lymph node are capable of delivering different signals according to the microbes they encounter in the periphery.

approach to obtain endogenous antigen synthesis in DCs has been to fuse DCs to tumor cells. The immunogenicity of such vaccines was confirmed in several experimental systems, although the mechanisms leading to the clinical success of allogeneic approaches remained open.

The alternative strategy foresees to provide tumor antigens exogenously. To this end, tumor cell lysates and defined tumor antigens, applied as a protein or a peptide, were coincubated with autologous DCs so that MHC class II–restricted peptide antigens become cross-presented. An alternative to this latter strategy is to prepare exosomes from the respective tumor cells. Exosomes are the internal vesicles of multivesicular endosomes. They were shown to contain tumor rejection antigens and to be able to transfer them to DCs, so that tumor-specific CTL responses are generated in mice. Although several clinical trials – where DC-based vaccination was employed in the context of malignant melanoma, lymphoma, myeloma, prostate, and renal cancer – were successful to a certain extent, there are still several potential caveats that have to be taken into account. Some tumor antigens are not tumor-specific but tissue-specific and hence widely expressed in normal tissues. In case of vaccination with melanocyte/melanoma antigens, this leads to vitiligo due to autoimmunity at the vaccination site. Conversely, the immunoproteasome of DCs may process tumor antigens in a way that is different from processing by the normal proteasome being active in tumor cells. Moreover, feeding DCs with whole tumor cell lysates is risky since T cells that are not directed against tumor antigens may be primed as well. Therefore, precise knowledge of MHC class I–and MHC class II–associated tumor-specific peptide antigens is necessary. Unfortunately, only a few class II–restricted helper epitopes are known so far. Therefore, further effort is needed to obtain a more comprehensive set of tumor-specific peptide epitopes in order to optimize DC-based vaccination against major malignancies.

See also Dendritic Cells.

Bibliography

Books and Reviews

Banchereau, J., Steinman, R.M. (1998) Dendritic cells and the control of immunity, *Nature* **392**, 245–252.

Beck, S., Trowsdale, J. (1999) Sequence organisation of the class II region of the human MHC, *Immunol. Reviews* **167**, 201–210.

Fearon, D.T., Carter, R.H. (1995) The CD19/CR2/TAPA-1 complex of B lymphocytes: linking natural to acquired immunity, *Ann. Rev. Immunol.* **13**, 127–149.

Fehder, W.P., Ho, W.Z., Douglas, S.D. (2000) Macrophages, in: Fink, G. (Ed.) *Encyclopedia of Stress*, Academic Press, San Diego pp. 382–398.

Greenfield, E.A., Nguyen, K.A., Kuchroo, V.K. (1998) CD28/B7 co-stimulation: a review, *Crit. Rev. Immunol.* **18**, 389–418.

Humphreys, R.E., Pierce, S.K. (Eds.) (1994) *Antigen Processing and Presentation*, Academic Press, New York.

Kronenberg, M., Brossay, L., Kurepa, Z., Forman, J. (1999) Conserved lipid and peptide presentation functions of nonclassical class I molecules, *Immunol. Today* **20**, 515–521.

Kropshofer, H., Hammerling, G.J., Vogt, A.B. (1999) The impact of the non-classical MHC proteins HLA-DM and HLA-DO on loading of MHC class II molecules, *Immunol. Reviews* **172**, 267–278.

Lotze, M., Thompson, A. (Eds.) (1998) *Dendritic cells: Biology and Clinical Application*, Academic Press, New York.

Ma, X., Trinchieri, G. (2001) Regulation of interleukin-12 production in antigen presenting cells, *Adv. Immunol.* **79**, 55–92.

Matzinger, P. (2001) The danger model in its historical context, *Scand. J. Immunol.* **54**, 4–9.
Parker, D.C. (1993) T cell-dependent B cell activation, *Ann. Rev. Immunol.* **11**, 331–360.
Paulnock, D.M. (1992) Macrophage activation by T cells, *Curr. Opin. Immunol.* **4**, 344–349.
Srivastava, P.K., Menoret, A., Basu, S., Binder, R.J., McQuade, K.L. (1998) Heat shock proteinscome of age: primitive functions acquire new roles in an adaptive world, *Immunity* **8**, 657–665.
Stoy, N. (2001) Macrophage biology and pathobiology in the evolution of immune responses: a functional analysis, *Pathobiology* **69**, 179–211.
Vogt, A.B., Spindeldreher, S., Kropshofer, H. (2002) Clustering of MHC peptide complexes prior to their engagement in the immunological synapse: lipid raft and tetraspan microdomains, *Immunolog. Reviews* **189**, 136–151.
Yewdell, J.W., Norbury, C.C., Bennink, J.R. (1999) Mechanisms of exogenous antigen presentation by MHC class I molecules in vitro and in vivo: implications for generating $CD8^+$ T cell responses to infectious agents, tumors, transplants, and vaccines, *Adv. Immunol.* **73**, 1–77.
Zinkernagel, R.M. (2000) Localization, dose and time of antigens determine immune reactivity, *Semin. Immunol.* **12**, 215–219.

Primary Literature

Aderem, A., Ulevitch, R.J. (1998) Toll-like receptors in the induction of the innate immune response, *Nature* **406**, 782–787.
Bachmann, M.F., McKall-Faienza, K., Schmidts, R., Bouchard, D., Beach, J., Speiser, D.E., Mak, T.W., Ohashi, P.S. (1997) Distinct roles for LFA-1 and CD28 during activation of naïve T cells: adhesion versus co-stimulation, *Immunity* **7**, 549–557.
Basu, S., Binder, R.J., Ramalingam, T., Srivastava, P.K. (2001) CD91 is a common receptor for heat shock proteins gp96, hsp90, hsp70, and calreticulin, *Immunity* **14**, 303–313.
Cella, M., Engering, A., Pinet, V., Pieters, J., Lanzavecchia, A. (1997) Inflammatory stimuli induce accumulation of MHC class II complexes on dendritic cells, *Nature* **388**, 782–787.
Cella, M., Facchetti, F., Lanzavecchia, A., Colonna, M. (2000) Plasmacytoid dendritic cells activated by influenza virus and CD40L drive a potent TH1 polarization, *Nat. Immunol.* **1**, 305–310.
Chen, X., Laur, O., Kambayashi, T., Li, S., Bray, R.A., Weber, D.A., Karlsson, L., Jensen, P.E. (2002) Regulated expression of human histocompatibility leucocyte antigen (HLA)-DO during antigen-dependent and antigen-independent phases of B cell development, *J. Exp. Med.* **195**, 1053–1062.
Gadola, S.D., Zaccai, N.R., Harlos, K., Shepherd, D., Castro-Palomino, J.C., Ritter, G., Schmidt, R.R., Jones, E.Y., Cerundola, V. (2002) Structure of human CD1b with bound ligands at 2.3 Å, a maze for alkyl chains, *Nat. Immunol.* **3**, 721–726.
Garside, P., Ingulli, E., Merica, R.R., Johnson, J.G., Noelle, R.J., Jenkins, M.K. (1998) Visualization of specific B and T lymphocyte interactions in the lymph node, *Science* **281**, 96–99.
Goerdt, S., Politz, O., Schledzewski, K., Birk, R., Gratchev, A., Guillot, P., Hakiy, N., Klemke, C.D., Dippel, E., Kodelja, V., Orfanos, C.E. (1999) Alternative versus classical activation of macrophages, *Pathobiology* **67**, 222–226.
Gorak, P.M., Engwerda, C.R., Kaye, P.M. (1998) Dendritic cells but not macrophages, produce IL-12 immediately following *Leishmania donovani* infection, *Eur. J. Immunol.* **28**, 687–695.
Grakoui, A. Bromley, S.K., Sumen, C., Davis, M.M., Shaw, A.S., Allen, P.M., Dustin, M.L. (1999) The immunological synapse: a molecular machine controlling T cell activation, *Science* **285**, 221–227.
Gunzer, M., Schafer, A., Borgmann, S., Grabbe, S., Zanker, K.S., Brocker, E.B., Kampgen, E., Friedl, P. (2000) Antigen presentation in extracellular matrix: interactions of T cells with dendritic cells are dynamic, short-lived, and sequential, *Immunity* **13**, 323–332.
Humphrey, J.H., Grennan, D., Sundaram, V. (1984) The origin of follicular dendritic cells in the mouse and the mechanism of trapping of immune complexes on them, *Eur. J. Immunol.* **14**, 1859–1867.
Kropshofer, H., Arndt, S.O., Moldenhauer, G., Hammerling, G.J., Vogt, A.B. (1997) HLA-DM acts as a molecular chaperone and rescues empty HLA-DR molecules at lysososmal pH, *Immunity* **6**, 293–302.
Kropshofer, H., Spindeldreher, S., Roehn, T.A., Platania, N., Grygar, C., Daniel, N., Woelpl, A.,

Langen, H., Horejsi, V., Vogt, A.B. (2002) Tetraspan microdomains distinct from lipid rafts enrich select peptide-MHC class II complexes, *Nat. Immunol.* **3**, 61–68.

Kupfer, H., Monks, C.R., Kupfer, A. (1994) Small splenic B cells that bind to antigen-specific T helper (Th) cells and face the site of cytokine production in the Th cells selectively proliferate: immunofluorescence microscopic studies of Th-B antigen-presenting cell interactions, *J. Exp. Med.* **179**, 1507–1515.

Langenkamp, A., Messi, M., Lanzavecchia, A., Sallusto, F. (2000) Kinetics of dendritic cell activation: impact on priming of TH1, TH2 and nonpolarized T cells, *Nat. Immunol.* **1**, 311–316.

Lohoff, M., Koch, A., Rollinghoff, M. (1992) Two signals are involved in polyclonal B cell stimulation by T helper type 2 cells: a role for LFA-1 molecules and interleukin 4, *Eur. J. Immunol.* **22**, 599–602.

Ludewig, B., Odermatt, B., Landmann, S., Hengartner, H., Zinkernagel, R.M. (1998) Dendritic cells induce autoimmune diabetes and maintain disease via de novo formation of local lymphoid tissue, *J. Exp. Med.* **188**, 1493–1501.

Mitchell, D.A., Nair, S.K., Gilboa, E. (1998) Dendritic cell/macrophage precursors capture endogenous antigen for MHC class I presentation by dendritic cells, *Eur. J. Immunol.* **28**, 1923–1933.

Munoz-Fernandez, M.A., Fernandez, M.A., Fresno, M. (1992) Synergism between tumor necrosis factor-α and interferon-γ on macrophage activation for the killing of intracellular *Trypanosoma* crusi through a nitric oxide-dependent mechanism, *Eur. J. Immunol.* **22**, 301–307.

Nestle, F.O., Alijagic, S., Gilliet, M., Sun, Y., Grabbe, S., Dummer, R., Burg, G., Schadendorf, D. (1998) Vaccination of melanoma patients with peptide- or tumor lysate-pulsed dendritic cells, *Nat. Med.* **4**, 328–332.

Nestle, F.O., Banchereau, J., Hart, D. (2001) Dendritic cells: on the move from bench to bedside, *Nat. Med.* **7**, 761–765.

O'Rourke, A.M., Mescher, M.F. (1992) Cytotoxic T lymphocyte activation involves a cascade of signaling and adhesion events, *Nature* **358**, 253–255.

Ridge, J.P., Di Rosa, F., Matzinger, P. (1998) A conditioned dendritic cell can be a temporal bridge between a $CD4^+$ T-helper and a T-killer cell, *Nature* **393**, 474–478.

Roncarolo, M.G., Levings, M.K., Traversari, C. (2001) Differentiation of T regulatory T cells by immature dendritic cells, *J. Exp. Med.* **153**, 5–9.

Sauter, B., Albert, M.L., Francisco, L., Larsson, M., Somersan, S., Bhardwaj, N. (2000) Consequences of cell-death: exposure to necrotic tumor cells, but not primary tissue cells or apoptotic cells, induces the maturation of immunostimulatory dendritic cells, *J. Exp. Med.* **191**, 423–434.

Sallusto, F., Lanzavecchia, A. (1994) Efficient presentation of soluble antigen by cultured human dendritic cells is maintained by granulocyte/ macrophage colony-stimulating factor plus interleukin 4 and down-regulated by tumor necrosis factor alpha, *J. Exp. Med.* **179**, 1109–1118.

Shi, G.P., Villadangos, J.A., Dranoff, G., Small, C., Gu, L., Haley, K.J., Riese, R., Ploegh, H., Chapman, H.A. (1999) Cathepsin S required for normal MHC class II peptide loading and germinal center development, *Immunity* **10**, 197–206.

Stout, R., Bottomly, K. (1989) Antigen-specific activation of effector macrophages by interferon-γ-producing (TH1) T-cell clones: failure of IL-4 producing (TH2) T-cell clones to activate effector functions in macrophages, *J. Immunol.* **142**, 760–768.

Sugita, M., Grant, E.P., van Donselaar, E., Hsu, V.W., Rogers, R.A., Peters, P.J., Brenner, M.B. (2000) Separate pathways for antigen presentation by CD1 molecules, *Immunity* **11**, 743–752.

Underhill, D.M., Bassetti, M., Rudensky, A., Aderem, A. (1999) Dynamic interactions of macrophages with T cells during antigen presentation, *J. Exp. Med.* **190**, 1909–1914.

Antisense Oligonucleotides as Potential Drugs

Joachim W. Engels[1] *and Fritz Eckstein*[2]
[1] *Goethe-Universitat Institut für Organische Chemie und Chemische Biologie (OCCB), Frankfurt am Main, Germany*
[2] *Max-Planck-Insitut für Experimentelle Medizin, Göttingen, Germany*

1 Introduction 405

2 Chemical Modification of Oligonucleotides 406
2.1 Unmodified Oligonucleotides 406
2.2 Phosphorothioates 406
2.3 Alkylphosphonates 406
2.4 2′-*O*-Modification 408
2.5 Sterically Locked Nucleic Acids 408
2.6 N3′-P5′-Phosphoramidates 409
2.7 Morpholino Oligonucleotides 409
2.8 Peptide Nucleic Acids (PNA) 409
2.9 Oligonucleotide Conjugates 409
2.10 5′-end Conjugates 409
2.11 3′-end Conjugates 409

3 Mechanism of Action 410

4 Positioning 410

5 Delivery 411

6 Application 411

7 RNA Interference 412

8 Ribozymes 413

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

9 Conclusion 413

Bibliography 413
Books and Reviews 413
Primary Literature 414

Keywords

Anneal (hybridize)
To form doubled-stranded oligo/polynucleotides (duplex) from two complementary single-stranded polynucleotides by heating and cooling.

Antisense
A strand of DNA that has the same sequence as mRNA (coding strand).

Ribozyme
Nucleic acid polymer (RNA) that catalyzes a chemical reaction, mostly transesterification.

RNase H
A ribonuclease that recognizes and cleaves internally the RNA strand of a DNA–RNA hybrid.

RNAi
RNA interference, a natural process whereby small double-stranded (21-mers) RNAs induce gene silencing.

Phosphorothioate
Oligonucleotides in which one of the nonbinding oxygens on phosphorus is replaced by sulfur.

Backbone
Refers to the sugar–phosphate linkage in DNA or RNA that can be replaced by isosteric linkages.

Antisense technology is a novel method for drug discovery and drug development. Antisense oligodeoxynucleotides (ODN) work at the genetic level by inhibiting the translation of disease-causing proteins.

Proteins are the key players in the etiology of diseases. Especially for cancer and viral infections, the antisense technology shows promise. The underlying mechanism of this approach is to sequence-specifically interfere in the genetic flow at the mRNA level. This molecular recognition is believed to rely on Watson–Crick base pairing

between the mRNA and the antisense oligonucleotide. Different modes of action are to be considered. The ODNs can either simply function as a steric block to interfere with the ribosomal translation or induce the action of particular RNases to destroy the mRNA in the hybrid formed with the ODN. In addition to this pharmacodynamic consideration, pharmacokinetic processes are important for optimal drug activity. Key points are the mode of delivery of the ODNs to the patient, to assure that the target is reached in the cells or organs of concern. The synthesis of the oligonucleotides has to be straightforward, economic, and reliable. The ODNs should be fairly metabolically stable, possess high affinity to the target, be nontoxic, and preferentially orally available. The first generation of ODNs are the phosphorothioate analogues that have been extensively tested. However, further developments of analogues for improvement are actively pursued.

This article will feature the design of oligos, their synthesis, modifications, and a discussion of the mechanisms of action. The ODN application will be illustrated with some selected examples, with the emphasis on those that are in clinical trials.

1 Introduction

The classical treatment of a disease is the inhibition of function of the protein that is presumed to be the causative agent of the disease, mostly with a small molecular weight drug. Even though this concept is very successful, problems arise when the drug is not very specific for the protein in question, for example, when the wild type and the mutated protein are not sufficiently different in their affinity for the drug. Thus, the idea to interfere earlier in the process of disease development by preventing the synthesis of the unwanted protein at the first place is attractive. It is based on the idea that the wild type and the mutated protein are sufficiently different in their nucleotide sequence that it should be possible to sequence-specifically inhibit either transcription or translation for the latter but not the former. The antisense methodology follows the route of translation interference. This approach has been applied beyond gene functional analysis to application in the clinic where a considerable number of trials are underway.

In the original concept, the antisense method for the sequence-specific inhibition of gene expression was quite simply steric blockage of translation by binding of the oligonucleotide (ODN) to the mRNA target. However, it is now generally accepted that the mode of action is more complicated. Even though some of the chemically modified antisense oligonucleotides (AS-ODN) seem to act by this mechanism, the most commonly used AS-ODN act by degradation of the RNA target by RNase H in the hybrid formed upon annealing of the AS-ODN. As it was recognized early that the unmodified ODNs are unsuitable for most *in vitro* and certainly for *in vivo* applications, chemical modification has played a major part in developing this methodology. The phosphorothioate analogues have played a crucial role in this development.

More recently, RNA interference has been described for the inhibition of gene

expression, which will be discussed later in a separate section.

The antisense oligodeoxynucleotide approach has been reviewed recently. A very thoughtful review of the different types of inhibition of gene expression with oligonucleotides has also appeared recently.

2 Chemical Modification of Oligonucleotides

2.1 Unmodified Oligonucleotides

Natural oligodeoxynucleotides are widely used as tools in molecular biology; however, they have serious limitations in their application as therapeutics for the sequence-specific inhibition of gene expression. Such oligonucleotides with a natural phosphodiester internucleoside linkage are degraded in serum within minutes to a few hours, mainly by the action of fast-cleaving 3′-exonucleases that are accompanied by slower-cleaving endonucleases. In certain tissues, a significant 5′-exonuclease activity has also been observed. However, oligonucleotides can be protected against nucleolytic cleavage by packaging into liposomes or nanoparticles that, at the same time, serve as carriers for *in vivo* application of these compounds as discussed below.

2.2 Phosphorothioates

Because of the nuclease sensitivity of ODNs, the major aim from the beginning was to strengthen the stability against such degradation by chemical modification. The first such modification was the introduction of phosphorothioate groups as internucleotidic linkages, which has been quite successful so that most antisense oligonucleotides (AS-ODNs) in use are the phosphorothioate analogues. This modification indeed enhances the stability of the ODNs against enzymatic degradation and at the same time is compatible with activation of RNase H. Thus, these analogues look ideal but unfortunately they have some drawbacks. Thus, the thermal stability of the complex formed with the target RNA is not as stable as that formed with an unmodified oligonucleotide. Moreover, for reasons that are not well understood, the phosphorothioate ODNs can have a much higher affinity to certain proteins such as those responsible for polyanion binding. This can result in undesired side effects often associated with cytotoxicity. Thus, reducing the number of phosphorothioates in an AS-ODN is advisable and will be discussed below.

As replacement of one of the nonbridging oxygens of the internucleotidic phosphate results in a chiral phosphorus, the chemical synthesis of the ODN results in a mixture of diastereomers unless efforts are undertaken to render the synthesis stereospecific, which is possible but not routinely done. There does not seem any particular advantage for either isomer in terms of the combination of thermal stability and RNase H activation over the mixture.

In summary, the advantages of the phosphorothioate modification outweighs the disadvantages at least up to now, making them a first generation of AS-ODNs.

2.3 Alkylphosphonates

Rather early, it was assumed that the negative charge of the phosphate backbone of the ODNs might interfere with

Structures for antisense application

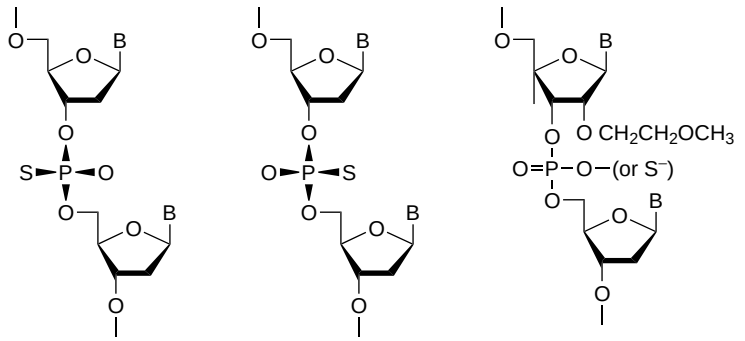

Phosphorothioates

Sp-diastereomer Rp-diastereomer 2′-*O*-Methoxyethyl

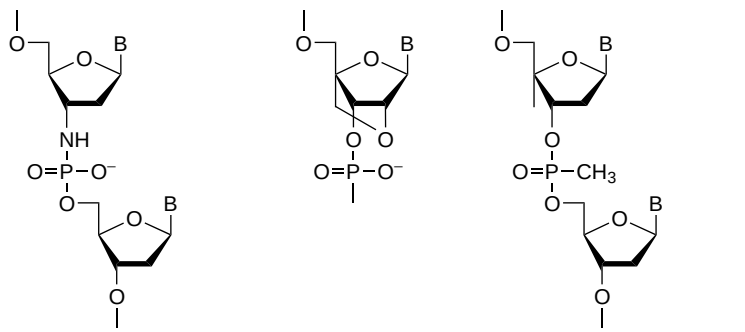

N3′ → P5′ Phosphoramidates LNA Alkyl phosphonates (mixture of isomers)

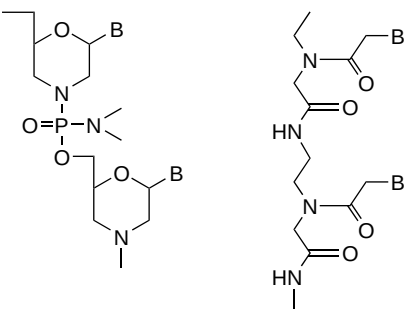

Morpholino PNA

cellular uptake, and thus constructing ODNs with a neutral backbone was attempted. Representatives for such analogues are the methylphosphonates in which the negatively charged phosphate oxygen is replaced by a neutral methyl group. These analogues have been used very early for antisense inhibition of gene expression. They are available through automated oligonucleotide synthesis using methylphosphonamidites as synthons. The methylphosphonate linkage is highly

stable to degradation by nucleases. As in the phosphorothioates, the methylphosphonates exist as two diastereomers of the Rp or the Sp configuration. Uniformly methylphosphonate-modified oligonucleotides do not form duplexes with RNA to induce RNase H cleavage. However, chimeric oligonucleotides containing both methylphosphonate and a window of at least five to seven phosphodiester linkages retain RNase H activity. This class of ODNs has not been very useful for the inhibition of gene expression. Besides the lack of RNase H activity, it also turned out that, contrary to expectation, cellular uptake is more sluggish than that of ODNs with a negative charge.

2.4 2′-*O*-Modification

In order to limit the number of phosphorothioates in an AS-ODN, alternative modifications were sought to enhance enzymatic stability without the inherent side effects of the phosphorothioates. It was recognized early that alkylation of the 2′ position of the ribose indeed increases stability. Thus, 2′-*O*-methyl modification was used, followed by the 2′-*O*-methoxyethyl derivatives. These analogues adopt an RNA-like (A-form) conformation, assuring high binding affinities for the RNA target and thus being superior in this respect to the phosphorothioates. As such, 2′ modification does not support RNase H activity (see below). The general policy is to locate them at the flanks of an AS-ODN, leaving a gap of approximately six to eight phosphorothioate nucleotides in the center to take advantage of RNase H activity. Thus, the second-generation oligos of the ISIS company have 2′-methoxy-ethoxy moiety termed *MOE-Oligos*. Often, particularly for animal model application, even the flanks contain the phosphorothioates to further improve stability. Apparently, in conjunction with a 2′alkyl modification, the cytotoxic side effects of phosphorothioates are not severe.

2.5 Sterically Locked Nucleic Acids

Nucleic acid analogues with conformationally restricted sugar–phosphate backbones have been attractive as it was envisaged that they might form more-stable duplexes for entropy reasons. Thus, the DNA analogue bicyclo-[3.2.1]-DNA has a rigid phosphodiester backbone, emulating a B-DNA-type conformation, to which the nucleobases are attached via a flexible open-chain linker. Although bicyclo-[3.2.1]-DNA forms less-stable duplexes with complementary DNA than natural DNA, base-mismatch discrimination is slightly enhanced compared to pure DNA duplexes. Importantly, bicyclo-[3.2.1]-DNA oligomers are resistant to 3′-exonuclease degradation. None of these analogues has been tested for inhibition of gene expression yet.

A further development along these lines are the locked nucleic acids (LNA), which represent a totally new and very promising class of compounds. In these analogues, the 2′ oxygen is linked by a methylene bridge to the 4′ carbon, forming a methylene-linked bicyclic ribofuranosyl nucleoside. These analogues, like the 2′-*O*-alkylated oligoribonucleotides, adopt an RNA-like (A-form) conformation and thus bind very well to target RNA. Probably not surprisingly, they are not substrates for nucleases. Oligonucleotides consisting purely of LNA units do not activate RNase H, but LNA/DNA chimeras do. Thus, these analogues combine many favorable

features for application. Although some cell culture data have been published, further studies should demonstrate the advantages of this class of compounds in a wider setting, in particular, in animal model studies.

2.6
N3′-P5′-Phosphoramidates

Oligonucleotides based on this structural element also show remarkable resistance to nucleases but do not activate RNase H. Extensive studies for the inhibition of gene expression are still awaited.

2.7
Morpholino Oligonucleotides

This class of compounds again is RNase H–irresponsive but, of course, is very stable to nucleases. It has found acceptance mainly in developmental studies in the zebra fish and Xenopus, both systems in which the compounds can be administered (injected) to the embryo. For other *in vivo* studies, uptake of the compound seems to be limited, as discussed below.

2.8
Peptide Nucleic Acids (PNA)

Peptide nucleic acids (PNA) deviate from all other analogues in that the linkage between the monomeric units is a peptide bond and not a phosphodiester bond. They show excellent affinity to RNA targets and thus should be ideally suited for interference with gene expression, even though they cannot rely on RNase H activity to destruct the target. Some cell culture data have been published, but further studies are desired to fully appreciate the potential of this class of compounds.

2.9
Oligonucleotide Conjugates

Certain properties of ODNs can be influenced by the covalent attachment of nonnucleosidic moieties to either the 3′ or 5′ ends. This type of derivatization allows modulation of nuclease stability, cellular uptake, and organ distribution of ODNs. In addition, the internucleoside linkages, the C5 position of pyrimidine bases, and the 2′ position of ribose are also positions for derivatisation, but these will not be considered here.

2.10
5′-end Conjugates

Conjugation of molecules to the 5′ end of ODNs is straightforward by coupling a phosphoramidite or H-phosphonate derivative of the desired molecule to the 5′-hydroxy group of the oligomer, following chain elongation by solid-phase synthesis. A broad range of phosphoramidite derivatives of ligands, such as fluorescein, biotin, cholesterol, cholic acids, dinitrophenyl, acridine, and psoralen derivatives, are commercially available. Alternatively, the 5′-terminal hydroxy group of the oligonucleotide is reacted with an aminoalkyl linker phosphoramidite, which, after deprotection, results in a free aminoalkyl function. This amino function can then be reacted in solution with suitably activated conjugate molecule derivatives, such as active esters, isothiocyanates, or iodoacetamides.

2.11
3′-end Conjugates

The conjugation of molecules to the 3′ end of oligonucleotides is achieved by using appropriately functionalized solid supports

that, in addition, harbor a hydroxyl function from which the oligonucleotide chain is extended during solid-phase synthesis. After oligonucleotide synthesis is complete, the oligonucleotide conjugate is cleaved from the solid support and deprotected by ammonia treatment. Similarly to the 5′-conjugation, appropriate 3′-amino-modifier solid supports are commercially available, which allow coupling of suitably activated derivatives of the molecule to be conjugated with the 3′-amino alkyl group in a postsynthetic solution-phase step.

3 Mechanism of Action

For most of the AS-ODNs, activation of RNase H is the mechanism of action to interfere with translation. This enzyme recognizes a duplex between DNA and RNA, resulting in the cleavage of the RNA strand of this duplex. The essential role of this enzyme in inhibiting gene expression is based on the observation that the target RNA is degraded, which would not be expected if a simple blockage of translation, as originally conceived, was the mechanism of inhibition of translation. The most convincing evidence that the enzyme RNase H is responsible for this degradation relies on the fact that AS-ODNs that are incompatible with the action of this enzyme do not result in degradation of the mRNA. This enzymatic step, of course, makes AS-ODN turnover a key step in the process requiring, in principle, substoichiometric amounts of the oligonucleotide for complete destruction of the RNA.

Activation of RNase H demands a DNA-like or B-conformation of the ODN. This requirement is obviously best met by oligodeoxynucleotides as they exist in this conformation. However, DNA–RNA complexes are not as stable as RNA–RNA complexes, which makes ODN with an RNA-like or A-conformation more desirable for hybridization with RNA. This criterion is met by the 2′-alkylated ODNs and the LNAs, which indeed show high thermal stability. Thus, there is a dilemma in that one conformation is most suitable for one of the desired properties that is, unfortunately, not optimal for the other. The phosphorothioates apparently represent a good compromise for these two qualities. As mentioned above, the alkylated analogues do not activate RNase as the 2′ group interferes with recognition by this enzyme or prevents substrate cleavage.

Neither the 2′ alkylated analogues, the alkylphosphonates, the morpholino, the N3′-P5′-phosphoramidates, nor the PNAs activate RNase H, but they do show inhibition of gene expression. The simplest explanation is that they follow the original concept in physically blocking translation by preventing the ribosome either to bind at the entry site or to move along the mRNA. Alternatively, the oligo may replace one strand in a duplex. Whether this is a sufficient explanation or whether other factors are involved remains to be seen.

4 Positioning

An important factor in obtaining efficient inhibition of gene expression is the positioning of the AS-ODN on the mRNA. The secondary structure of the latter will have regions that are more suitable than others for the annealing process. Although a number of computational methods exist to predict such regions, many researchers prefer experimental approaches to identify such positions. The method based on the

annealing of a randomized oligodeoxynucleotide preferably to the native RNA and subsequent cleavage by endogenous RNase H is the nearest to the intracellular situation and yet is relatively simple to perform. In any event, in most cases a number of AS-ODNs will be synthesized to select the most active. Depending on the laboratory, this number can be up to twenty.

5 Delivery

One of the problems for the antisense strategy is the intracellular availability of the negatively charged oligonucleotides with a molecular weight of several thousand daltons. ODNs are usually not taken up efficiently in cell culture. Electroporation is one way of delivering the ODNs to the cell, but this method is not of general applicability. The most commonly used strategy for cell culture experiments is the complexation of the oligonucleotides with synthetic lipsosomes with a positive charge on the surface. A large variety of these compounds are commercially available. These carriers by themselves are usually toxic to the cell because of the positive charges, but once covered by the ODN this cytotoxicity is in most cases abolished. Transfection efficiencies can widely differ, but this can be monitored by cotransfection with a fluorescent ODN. In most cases, the ODNs are trapped to a considerable degree in the endosome so that only a fraction is located in the cytoplasm or in the nucleus, places where they should be operative. Most interesting is the observation that uptake of ODNs *in vivo*, that is, in animal models or even man, does not require such carriers. The delivery there is simply achieved, for example, by subcutaneous or intravenous infusion. This difference for *in vitro* and *in vivo* cellular uptake signifies a considerable difference between the two systems, which is not understood at present. Whatever the cause of it, it should caution for extrapolating from the *in vitro* results to *in vivo* relevance. Independent of the method of delivery, considerable efforts are spent on directing the ODNs specifically to the desired cell or organ type.

Uptake of the ODNs with a neutral backbone such as the morpholinos or the PNAs cannot rely on cationic carriers because of lack of binding. They require cell scraping, injection, or, as shown for PNA, conjugation with cationic peptides to facilitate cellular uptake. At present, this presents a certain barrier for wide application, in particular, for medical purposes.

6 Application

The literature on the application of ODNs for the inhibition of gene expression is vast.

We have summarized in Table 1 those examples in which the technique has advanced to clinical-phase trials, although this collection is probably not complete. Not surprisingly, considering the enormous expense involved, these investigations are almost exclusively conducted by companies. The only notable exception is the study by the laboratory of A. Gewirtz on CML for *ex vivo* treatment.

Not all the examples listed can be discussed here, so some comments must suffice. Common to all ODNs used in the clinical trials is one aspect of their chemical nature in that they all are phosphorothioate-containing ODNs. The first approved antisense drug was developed by ISIS, Fomivirsen or Vitravene, a trade

Tab. 1 Antisense drug pipeline.

Target	*Lead indication*	*Company*	*Product*	*Phase*
CMV	CMV-retinitis	ISIS	Vitravene	Drug
PKC-α	Cancer-NSCLC	ISIS/Lilly	Affinitac	III
C-RAF	Cancer-ovarian	ISIS	ISIS 5132	II
H-ras	Cancer-pancreatic	ISIS	ISIS 2503	II
ICAM-1	Crohn's disease	ISIS	ISIS 2302/Alicaforsen	III
ICAM-1	Tropical psoriasis	ISIS	ISIS 2302/Alicaforsen	II
ICAM-1	Ulcerative colitis	ISIS	ISIS 2302/Alicaforsen	II
HCV	Hepatitis C	ISIS	ISIS 14803	II
TNF-α	Crohn's disease	ISIS	ISIS 104838	II
TNF-α	Topical psoriasis	ISIS	ISIS 104838	II
c-myc	Cancer, restenosis	AVI-Biopharm	Resten-NG	III
bcl	Cancer	Genta /Aventis	Genasense	III
HIV	AIDS	Enzo Biochem	HGTV43	II
Adenosine A1	Respiratory disease	EpiGenesis	EPI2010	II
PKA	Cancer	Hybridon	Gem 231	II
Rib. reductase	Cancer	Lorus Terapeutics	GTI2040	II
TGF-β2	Cancer	Antisense Pharma	AP 12009	I/II
VEGFR1	Cancer	Ribozyme Pharmaceuticals	Angiozyme	II
HCV	Hepatitis C	Ribozyme Pharmaceuticals	Heptazyme	I
HER2	Cancer	Ribozyme Pharmaceuticals	Herzyme	I

name of Novartis, certainly a fantastic achievement. However, the number of patients with CMV-retinitis has grown lesser and lesser because of the successful treatment of AIDS at least in the Western hemisphere. ICAM-1 from the same company, as a medication for Crohn's disease, has not been approved yet but is pursued further under the trade name Alicaforsen. Very far advanced is the Genta anti-*BCL2* oligonucleotide Genasense for the treatment of malignant melanoma in conjunction with chemotherapy. Genasense is in phase III clinical trial, developed together with Aventis.

Affinitac is a potent inhibitor of PKCα in non–small cell lung cancer, which shows almost double survival times (16 instead of 8 months) in combination with carboplatin/taxol. The main objective is to obtain a good synergistic profile by the combination with chemotherapeutic agents, as is the goal with Genasense, with taxol in this case. Affinitac is also tested in a single-agent (phase II) trial against non-Hodgkin's lymphoma. Here, a codevelopment together with Eli Lilly is under way.

7 RNA Interference

Inhibition of gene expression by RNA interference is a rapidly growing area of research and has already found considerable application for gene functional analysis. It differs from the antisense method fundamentally. It is a natural process in plants undergoing transgene- or virus-induced posttranscriptional gene silencing (PTGS). Here, short double-stranded oligoribonucleotides are processed from long dsRNA for the inhibition of expression of the foreign gene.

This process has now been extended to many organisms, mainly in cell culture but also in a few animal model studies for gene function analysis. It requires double-stranded oligoribonucleotides 21 nucleotides in length (called *siRNAs*) of which one strand is completely complementary to the mRNA target. Upon binding of this strand to the mRNA in a RNA induced silencing complex (RISC) complex, the target is cleaved opposite the oligoribonucleotide. The breakthrough in the field came by the observation that the oligonucleotides can be added exogenously, thus facilitating wide application. They can be chemically synthesized or obtained by *in vitro* transcription to be transfected with the aid of a carrier as detailed earlier. Alternatively, vectors can be constructed for stable transfection where the sequence of the oligonucleotides is positioned, most often behind the U6 promoter, for intracellular transcription. The inhibition by the siRNAs is not very dependent on the secondary structure of the mRNA. Usually three constructs are sufficient to identify the most active oligonucleotide pair, thus much fewer than that required for the AS-ODNs as described earlier. Also, in general, the inhibition achieved is very satisfactory.

8
Ribozymes

Although not normally considered in the context of the antisense methodology, ribozymes fall into this category as the basis for their interaction with RNA is based on the sense–antisense principle. The fundamental difference between the classical antisense method and that of RNA interference is that ribozymes have the inherent catalytic power to cleave the target RNA rather than to have to rely on cellular proteins for this step. Ribozymes will not be discussed here in depth, but two recent reviews give an excellent account of the present understanding of the various types of ribozymes. Some clinical trials are ongoing that are predominantly directed against Hepatitis C (Heptazyme), the VEGF receptor (Angiozyme), and the human epidermal growth factor HER2 (Herzyme).

9
Conclusion

This review demonstrates that the antisense methodology is not as simple as originally anticipated. As research progresses, more and more insight is gained into the complexity of responses induced by the insertion of an oligonucleotide in a cell. Given these problems, it is a great achievement that quite a few ODNs are now in clinical trials. It provides encouragement to pursue this line of development further. With the knowledge available from past experience, it is expected that the pace of the development of oligonucleotide-based drugs will accelerate.

See also Bioorganic Chemistry.

Bibliography

Books and Reviews

Agrawal, S., Kandimalla, E. (2001) Antisense and/or immunostimulatory oligonucleotide therapeutics, *Curr. Cancer Drug Targets* **1**, 197–209.

Bennet, C.F. (1998) Use of Cationic Lipid Complexes for Antisense Oligonucleotide Delivery, in: Stein, C.A., Krieg, A.M. (Eds.)

Applied Antisense Oligonucleotide Technology, Wiley-Liss, pp. 129–145.

DeRose, V.J. (2002) Two decades of RNA catalysis, *Chem. Biol.* **9**, 961–969.

Eckstein, F. (2000) Phosphorothioate oligodeoxynucleotides: What is their origin and what is unique about them? *Antisense Nucleic Acid Drug Dev.* **10**, 117–121.

Lebedeva, I., Stein, C.A. (2001) Antisense oligonucleotides: promise and reality, *Annu. Rev. Pharmacol. Toxicol.* **41**, 403–419.

McManus, M.T., Sharp, P.A. (2002) Gene silencing in mammals by small interfering RNAs, *Nat. Rev. Genet.* **3**, 737–747.

Miller, P.S. (1991) Oligonucleoside methylphosphonates as antisense reagents, *BioTechnology* **9**, 358–362.

Nielsen, P.E. (1999) Peptide nucleic acids as therapeutic agents, *Curr. Opin. Struct. Biol.* **9**, 353–357.

Opalinska, J.B., Gewirtz, A.M. (2002) Nucleic acid therapeutics: basic principles and recent applications, *Nat. Rev. Drug Discovery* **1**, 503–514.

Primary Literature

Braasch, D.A., Corey, D.R. (2001) Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA, *Chem. Biol.* **8**, 1–7.

Doudna, J.A., Cech, T.R. (2002) The chemical repertoire of natural ribozymes, *Nature* **418**, 222–228.

Engels, J., Jaeger, A. (1982) A new synthesis of nucleoside methylphosphonates, *Angew. Chem.* **94**, 931.

Epple, C., Leumann, C. (1998) Bicyclo(3.2.1)-DNA, a new DNA analog with a rigid backbone and flexibly linked bases: pairing properties with complementary DNA, *Chem. Biol.* **5**, 209–216.

Gryaznov, S.M. (1999) Oligonucleotide N3′-P5′phosphoramidates as potential therapeutic agents, *Biochim. Biophys. Acta* **1489**, 131–140.

Grillone, L.R., Lanz, R. (2001) Fomivirsen, *Drugs Today* **37**, 245–255.

Heasman, J. (2002) Morpholino oligos: making sense of antisense? *Dev. Biol.* **243**, 209–214.

Jensen, B., Wacheck, V., Heere-Ress, E., Schlagbauer-Wadl, H., Hoeller, C., Lucas, T., Hoermann, M., Hollenstein, U., Wolff, K., Pehamberger, H. (2000) Chemosensitisation of malignant melanoma by *BCL2* antisense therapy, *The Lancet* **356**, 1728–1733.

Luger, S.M., O'Brien, S.G.O., Ratajczak, J., Ratajczak, M.Z., Mick, R., Stadtmauer, E.A., Nowell, P.C., Goldman, J.M., Gewirtz, A.M. (2002) Oligodeoxynucleotide-mediated inhibition of c-*myb* gene expression in autografted bone marrow: a pilot study, *Blood* **99**, 1150–1158.

Manoharan, M. (2002) Oligonucleotide conjugates s potential antisense drugs with improved uptake, biodistribution, targeted delivery, and mechanism of action, *Antisense Nucleic Acid Drug Dev.* **12**, 103–128.

Scherr, M., Rossi, J.J. (1998) Rapid determination and quantitation of the accessibility to native RNAs by antisense oligodeoxynucleotides in murine cell extracts, *Nucleic Acids Res.* **26**, 5079–5085.

Yuen, A., Halsey, J., Fisher, G., Advani, R., Moore, M., Saleh, M., Ritch, P., Harker, G., Ahmed, F., Jone, C., Polikoff, P., Keiser, W., Kwoh, J., Holmlund, J., Dorr, A., Sikic, B. (2001) Phase I/II Trial of ISI 3521, an Antisense Inhibitor of PKC-Alpha, with Carboplatin and Paclitaxel in Non-Small Cell Lung Cancer (May 12, 2001), *American Society of Clinical Oncology's 37th Annual Meeting*, San Francisco, CA.

Antitumor Agents: Taxol and Taxanes – Production by Yew Cell Culture

Arthur G. Fett-Neto,[1] *Hideki Aoyagi,*[2] *Hideo Tanaka*[2], *and Frank DiCosmo*[3]
[1] *Universidade Federal Do Rio Grande Do Sul, Porto Alegre, Brazil*
[2] *University of Tsukuba, Tsukuba, Ibaraki, Japan*
[3] *University of Toronto, Toronto, ON, Canada*

1 Plant Material Selection and Preparation 417
1.1 Type and Source of Explant 417
1.2 Sterilization Procedure 418
1.3 Explant Size and Final Preparation 418

2 Establishment of Callus and Cell Cultures 418
2.1 General Procedure 418
2.2 Medium Optimization 419

3 Analyses of Taxol™ and Taxane Production 422
3.1 Extraction Procedure 422
3.2 Separation, Quantification and Identification 423

4 Taxol™ Biosynthesis 424

5 General Characteristics of Taxol™ Yield and Growth in *Taxus* Cell Culture 426

6 Techniques to Improve Taxol™ Yield in Cell Culture 429
6.1 Precursor and Sugar Feeding 429
6.2 Cell Line Selection 430
6.3 Light and Temperature 430
6.4 Elicitation and Metabolic Inhibitors 431
6.5 Gas Composition 432
6.6 Protoplast Cultures 432
6.7 Culture System 435

7 Scale-up of Taxus Cell Cultures 435

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

8 **Future Perspectives** 435

Bibliography 437
Books and Reviews 437
Primary Literature 437

Keywords

Callus
A mass of disorganized plant tissue, generally with low degree of cell differentiation, usually cultured on solid media.

Cell Suspension
Cells derived from callus or explant tissue and cultivated in liquid media, usually proliferating cell aggregates or single cells.

Explant
The portion of the intact plant used to initiate callus cultures.

Taxane
A compound related to taxol that can be described as a diterpene composed of three pentadecene rings.

Taxol™ (general name Paclitaxel)
A diterpene amide with antineoplastic properties.

Taxus
A genus of gymnosperm trees belonging to the family Taxaceae and commonly known as yew.

Taxol™ is an anticancer drug increasingly used for the approved treatment of various cancer types, including ovarian, breast, lung, head and neck, bladder and cervix, melanomas, and AIDS-related Karposi's sarcoma. The supply of this compound is limited because of expanding demand and the fact that the only commercial source is the biomass of *Taxus* species. Yew trees grow relatively slowly, contain low and variable amounts of Taxol™, and in many cases are located in protected, environmentally sensitive areas. The bulk of current Taxol™ production derives from semisynthesis starting from 10-deacetyl baccatin III, a precursor obtained from yew needles. Cell culture of yew species is a viable alternate source of Taxol™ and related taxanes, which can be used for production and/or semisynthesis of the drug. Cell cultures are readily renewable and yield relatively homogeneous, easily manipulated systems for biosynthetic studies on Taxol™, amenable to large-scale commercial production.

1 Plant Material Selection and Preparation

1.1 Type and Source of Explant

The procedures described here were established primarily from results obtained with *Taxus cuspidata* (Japanese yew), although some studies were performed using *Taxus canadensis* and *Taxus brevifolia*. In general lines, variations of these procedures have been successfully applied to a number of other *Taxus* species. In the recent literature of taxane production by yew cell culture, preference has been given to *T. chinensis*, *T. media*, *T. cuspidata*, and *T. canadensis*.

The response of yew tissue to *in vitro* culture is dependent on the type of explant used to initiate the culture, on the physiological status of the donor plant, and on its genetic makeup. Explants that contain meristematic or parenchymal tissue are preferable. These are the types of tissue most likely to display cell proliferation in response to the growth regulators in the culture medium, and this tendency is attributed to these tissues' low degree of differentiation, which seems to be associated with a certain developmental flexibility in the presence of the appropriate stimuli. Relatively more differentiated tissues, such as sclerenchyma or xylem, are less responsive to exogenous plant growth regulators, possibly because of the genetic commitment to a specialized developmental pathway.

Young stem segments (flexible, but bearing dark green needles) are ideal for rapid initiation of callus to be used in developing cell cultures. Seed contents (megagametophyte and embryo) are better suited for establishing long-term maintenance of callus cultures of *Taxus brevifolia*. Needles are also a suitable explant type for callus initiation, although initial growth rates tend to decrease more rapidly than those of young stem segments. Arils (fleshy basal covering of the seed) display different responses to *in vitro* culture depending on their stage of development. Green arils grow relatively fast for the first two weeks, but from then on, the growth seems to stop. Calli derived from green arils are relatively hard and not friable. Mature red arils, not surprisingly, fail to grow in culture, possibly as a result of the advanced degree of cell lysis at this later stage of their development (Table 1).

The physiological condition of the donor plant must be stable and favorable to growth. Outdoor-grown plants display higher contamination rates and tissue pigmentation when cultures are started in the winter months in Canada (November to March) relative to those started in other seasons. Genetic differences between plants of the same species are also factors to be considered; it may be necessary to test different donor plants to develop *in vitro* cultures. Donor plants of the same species, although grown under similar environmental conditions, may display significant differences in the

Tab. 1 Effect of different explant sources of *T. cuspidata* cultivated in B5 medium (containing 1 mg L^{-1} 2,4-D) on callus formation and growth.

Explant	*Growth*[a]	*Friability*
Green aril	+	No
Red aril	–	–
Seed content	++	Yes
Young stem[b]	+++	Yes
Needle[b]	++	Yes

[a] +, directly proportional to growth; –, poor growth.
[b] Also tested with *T. Canadensis*.

amount of growth *in vitro* as a result of genetic variability. It may also be beneficial to begin *in vitro* culture using plants shown to produce high yields of Taxol™. High taxane yields *in vitro*, however, are not always warranted by this phenotype in explant donor plants.

1.2 Sterilization Procedure

Sterilization is accomplished using standard plant tissue culture techniques. The tissue is washed with distilled water to remove dust and solid particles, immersed in 70% ethanol for one minute, immersed with a few drops of Tween 20 (surfactant agent) in NaClO (1.5% v/v) for 30 minutes, and rinsed three times with sterile distilled and deionized water. During the sterilization period, the tissue must be kept fully immersed and under agitation. Sterilization times may have to be reduced for more sensitive and tender tissues. Our experience has yielded typical contamination rates using outdoor-grown donor plants that are less than 3% of the total number of explants used. In winter, however, the incidence of contamination can be 10 times the regular rate. In certain cases, a particular plant may be extensively infested with endophytes that are not destroyed by surface-sterilization procedures; in this situation, contamination can be very high and a change of donor plant is advisable.

1.3 Explant Size and Final Preparation

Following sterilization, the exposed cut portions of the tissue that were damaged by the ethanol and NaClO solutions are excised, and the healthy tissue is appropriately trimmed and placed in the growth medium. Excision of explants should be made smoothly, with a relatively new blade, to minimize the production of phenolics in the cut areas. In the case of stem segments, approximately 1-cm long segments give good results. Stem segments can be cultured at an approximate angle of 45° to the agar surface, with one of the transversely sectioned ends immersed in the medium and the other partly exposed to the air of the culture vessel. Alternatively, longitudinally cut stem segments can be cultured with their cut portions facing the surface of the medium. The contents of longitudinally halved seeds can be placed on the medium after separation from the seed coat with a spatula. Isolated embryos can be used instead, in order to avoid calli derived from the haploid megagametophyte tissue; haploid tissues, on the other hand, may be useful for mutant generation and screening. Hypocotyl segments or other portions of aseptically germinated seedlings are also good sources of explants. Yew tissues seem to be very sensitive to heat, relative to tissues of other woody plants. All tools used in explant preparation and cell transfer should be sterilized in ethanol, flamed briefly to remove ethanol residue, and cooled to room temperature before touching the plant material.

2 Establishment of Callus and Cell Cultures

2.1 General Procedure

Cultures of yew are relatively difficult to establish because growth rates are slow and the explants produce and release phenolics; in turn, the oxidation of the phenolics leads to darkening, growth

inhibition, and eventual loss of viability of the tissue. The inclusion in the medium of phenolic-binding compounds, such as polyvinylpolypyrrolidone (PVP), improves the growth of the explants. The inclusion of activated charcoal in the media can also reduce darkening, although it may also diminish explant growth as a result of binding of medium components. In addition, transfers to fresh medium (or rotation of the explants to new positions in the culture vessel containing portions of unused medium) at approximately 15-day intervals help prevent growth inhibition due to phenolic oxidation.

More frequent transfers may be required in the early period of cultivation or when the donor plant is more prone to darkening upon *in vitro* culture. After several subcultures and growth stabilization of the cultures, the need for phenolic-binding adjuvants may decrease or even disappear.

Following an initial screening and evaluation of some standard medium formulations, the basic medium B5bPVP was established for callus initiation in yew; it consists of B5 medium salts and organic supplements, 3.0% (w/v) sucrose, 1.5% (w/v) insoluble PVP, 0.8% (w/v) agar, and 1 mg L^{-1} of 2,4-dichlorophenoxyacetic acid (2,4-D). The media are prepared with distilled and deionized water, plant cell culture or reagent-grade chemicals, adjusted to pH 5.8 with NaOH and HCl (after the addition of PVP and prior to the addition of agar) and autoclaved (1.1 kg cm^{-2} at 121 °C for 20 min).

Preferably, the glassware used for plant cell culture should be washed with neutral detergent, abundantly rinsed with tap water, and then rinsed with distilled water. The cultures can be grown in plastic petri dishes (9 cm in diameter) containing 30 mL of medium (3–4 explants per plate) and sealed with parafilm™ or air permeable tape, or in 30 mL glass vials with polypropylene or aluminum foil lids containing 10 mL of medium each (1 explant per vial). The cultures are incubated at 25 °C in dark. Various other media formulations based on MS, WPM, SH and B5 salts and organics supplemented with auxins and cytokinins have been effectively used for several *Taxus* species.

2.2 Medium Optimization

Additional improvements on the basic medium formulation for yew tissue culture can be accomplished by investigating several medium components (e.g. plant growth regulator balance, salt and organic supplement composition) that promote the growth of related species, or seem likely to have a stimulatory effect on the putative biosynthetic pathway to Taxol™. The effect of the medium components when included in various amounts in the basic formulation, or when replacing some of its original constituents, can be analyzed relative to growth and Taxol™ yield in developing yew calli. Medium factors with a positive effect on growth and Taxol™ yield are combined in new formulations to investigate the possibility of synergistic or antagonistic interactions.

A great deal of work is entailed in the high number of treatments and respective replicates needed for accurate statistical analysis, but these comprehensive experiments allow the identification of limiting factors for growth and Taxol™ production. The overall analysis of the results also shows the relationship between growth and Taxol™ production. From such a series of experiments, two improved media formulations were devised: B5bPhe (B5bPVP with 0.1 mM of phenylalanine)

and B5Phe8/0GA (B5bPhe salts and organic supplements with 8.0 mg L^{-1} of 2,4-D and 0.5 mg L^{-1} of gibberellic acid). After autoclaving, gibberellic acid (GA3) is added to the culture medium by filtration through 0.22-μm sterile filters. All other medium components are added prior to autoclaving. The overall analysis of the results showed a trend toward an inverse relationship between Taxol™ production and growth in the developing callus cultures.

Cell suspension cultures can be initiated from approximately two-month-old callus cultures developed from stem segments. Depending on the availability of callus biomass for inoculum and on incubator space, cell cultures can be grown in 125-, 250-, 500-, and 1000-mL Erlenmeyer flasks containing one-fifth of their total volume of liquid medium. Basic incubation conditions are 22–25 °C in dark on gyratory shakers at 110 rpm. Experiments done in callus and some additional optimization for cell suspension culture medium yielded the following basic formulations: B5 medium salts, threefold the regular amount of B5 organic supplements, 3.0% sucrose (w/v), 1.5% soluble PVP (MW $\leq 40 \times 10^3$), and 1 mg L^{-1} of 2,4-D (B5C1) or the same composition of B5C1, but with a modified balance of plant growth regulators (i.e. 4 mg L^{-1} of 2,4-D and 1 mg L^{-1} of kinetin (B5C2)).

Medium preparation, pH adjustment, and sterilization procedure are the same as in callus culture medium. Insoluble PVP is replaced with soluble PVP in liquid cell cultures to facilitate visual and microscopic observation of the cells and cell harvesting. Inoculum weights are preferably high (e.g. one-third of the volume of the liquid medium) to initially establish cell suspension cultures. After 1 to 2 weeks in culture on a gyratory shaker at 110 rpm, the cells released from the calli are collected and transferred into fresh liquid medium (1 : 1 or 2 : 1 proportion of old to fresh medium). The calli with attendant stem debris can be discarded or replated in solid medium to regenerate more callus tissue. During the first months of culture, the media can be replaced at 2-week intervals following 10 min of centrifugation at $600 \times g$ at 20 °C under sterile conditions to pellet the cells; in this way, the maximum amount of cell biomass is saved until the density is great enough to start suspension culture propagation in new flasks. Once suspensions have been stabilized (generally after several cycles of propagation), inoculum amounts can be reduced to one-sixth (w/v) of the liquid medium for propagation of new cell suspension flasks. Transfers to fresh medium in stock cultures can be done every 20 days. Viability checks of newly developing cell lines can be performed via dye exclusion of cells in the presence of Evans Blue (cells staining blue under the microscope are considered not viable) or other cytochemical method, and/or by analyzing the occurrence of growth in microcalli rescued from the liquid culture and plated in solid B5bPVP. Microscopic examination of cell viability and culture sterility should be performed regularly, particularly during the first few months of culture in liquid medium.

Cell cultures can also be immobilized after the first months of culture to facilitate medium replacement and to aid in the establishment of recalcitrant cell lines; this approach has been particularly useful in our efforts to establish *T. cuspidata* cell lines (Fig. 1). Immobilization is performed by spontaneous adhesion to 5 cm × 5 cm, 10- to 15-mm thick glass fiber mats with a uniform diameter of 10 μm and fiber density of approximately 61 mg cm^{-2}.

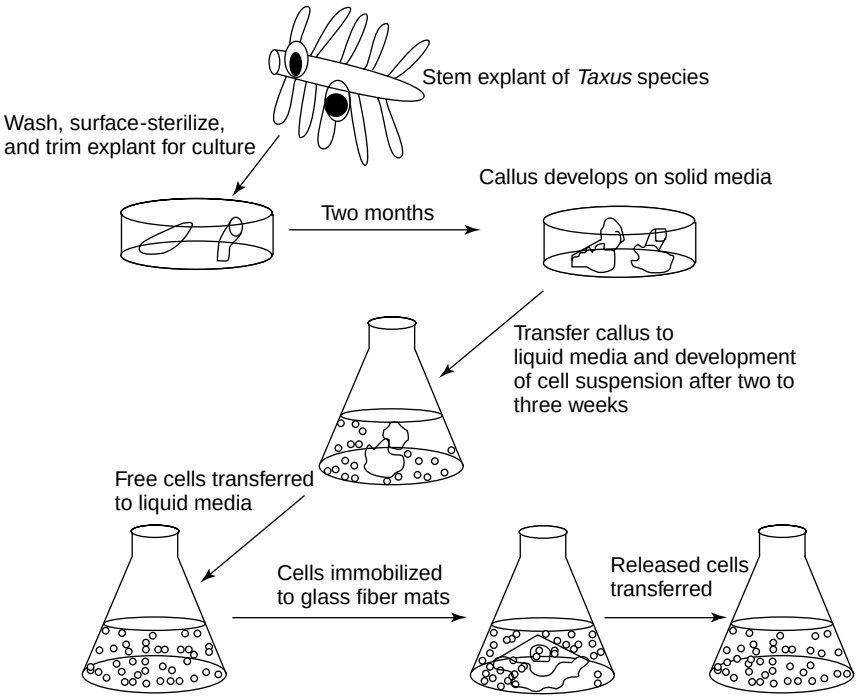

Fig. 1 Establishment of immobilized cell cultures and cell suspension cultures of *T. cuspidata*.

Within the first days of contact with the substrate, the cells are spontaneously immobilized. The cells grow on top of and within the mats, forming several layers of cells; after the cells have saturated the immobilization substrate, they begin to be released into the liquid medium. These cells can be used to establish new suspension cultures or new immobilized cultures. Alternatively, cell cultures can be propagated by carefully removing external cell layers with a sterile spatula, or by cutting the mats in two or four portions and placing these on fresh medium flasks containing a new glass fiber mat (Fig. 1).

Two-year-old cell suspensions of *T. cuspidata* have various degrees of cell aggregation. Moreover, the shape of the cells is variable, from round to considerably long, cylinder-like cells. In general, most cell suspension cultures display a relatively high number of cell aggregates of various sizes, with relatively fewer cells occurring individually. Cell aggregates may constitute a problem in large-scale cell cultures, which can be more susceptible to shear stress and may create difficulties in the mixing of cells, gases, and medium components. It is possible to select for smaller aggregate sizes by allowing larger aggregates to settle for a short period of time and then transferring the smaller aggregates and single cells from the upper portion of the flask into a new culture flask. The selection for the single-cell or small-aggregate type of suspension, however, may have implications for Taxol™ productivity. Aggregated cell suspensions tend to be more productive than suspensions

of the single-cell type, possibly because of their relatively higher degree of differentiation.

In our experience with cell lines of *T. cuspidata*, the stabilization of cell lines is a relatively lengthy process, normally requiring 6 to 12 months for completion. This is not only the case for yew species, but also for many woody species, as a result of their relatively slow growth and often recalcitrant behavior *in vitro*.

3 Analyses of Taxol™ and Taxane Production

3.1 Extraction Procedure

At least 0.5 g of fresh callus tissue or 1 g of fresh cells from suspensions is collected in glass test tubes (previously weighed and identified), frozen (−20 °C), and homogenized with a glass rod in 1 to 2 mL of hexane, followed by 1 h of sonication at room temperature. After 12 h in hexane at 25 °C, the samples are centrifuged at 2500 × *g* for 20 min to pellet the cells, and the supernatants are discarded. The pellets are then extracted in 1 to 2 mL of methanol/dichloromethane (1 : 1) with glass rod rehomogenization, followed by sonication for 2 h at room temperature. After centrifugation at 2500 × *g* for 30 min, the supernatants containing the taxanes are collected; the pellets are used for determining the extracted dry weights at 60 °C until constant weight.

The crude extracts are dried under vacuum or nitrogen at 25 °C, redissolved in 1 mL of dichloromethane, partitioned with 1 mL of water, and centrifuged at 2500 × *g* for 20 min. The dichloromethane fraction is collected, dried under vacuum or nitrogen at 25 °C, redissolved in an exact volume of high performance liquid chromatography (HPLC) grade methanol (100–500 μL), and filtered through 0.45 μm filters for HPLC analysis (Fig. 2).

Alternatively, the tissue can be extracted with 100% methanol. The hexane step can be avoided for cleaner extracts (e.g. cell suspension extracts) or if the column for liquid chromatography to be used in extract analysis is equipped with a precolumn, which prevents irreversible binding of impurities to the column. The Taxol™ extraction step (methanol/dichloromethane, 1 : 1) can be repeated two or three times with the pellet of the first Taxol™ extraction to achieve more complete Taxol™ recoveries. In experiments involving several treatment comparisons and large number of samples with respective replicates, however, a single methanol/dichloromethane extraction step is more practical; this procedure allows for Taxol™ recovery of 80 to 85%. Media from cell cultures can be extracted for Taxol™ with dichloromethane (1 : 1 v/v ratio of medium to solvent). For small volumes, the mixture contained in glass tubes is vigorously agitated on a tube mixer for 2 to 5 min, followed by sonication for an hour at room temperature. The tubes are briefly mixed again and centrifuged for 30 min at 2000 × *g* at 20 °C. For large volumes, the mixture is agitated overnight in flask shakers and partitioned for an hour in separation funnels. The upper aqueous layer is discarded and the organic fraction is dried under vacuum at 25 °C. The preparation for HPLC analysis is similar to the one just described, although the filtering of media extracts tends to be more difficult than in the case of plant material extracts. This difficulty in filtering media extracts prior to HPLC analysis is probably a result of the presence of relatively large amounts of soluble

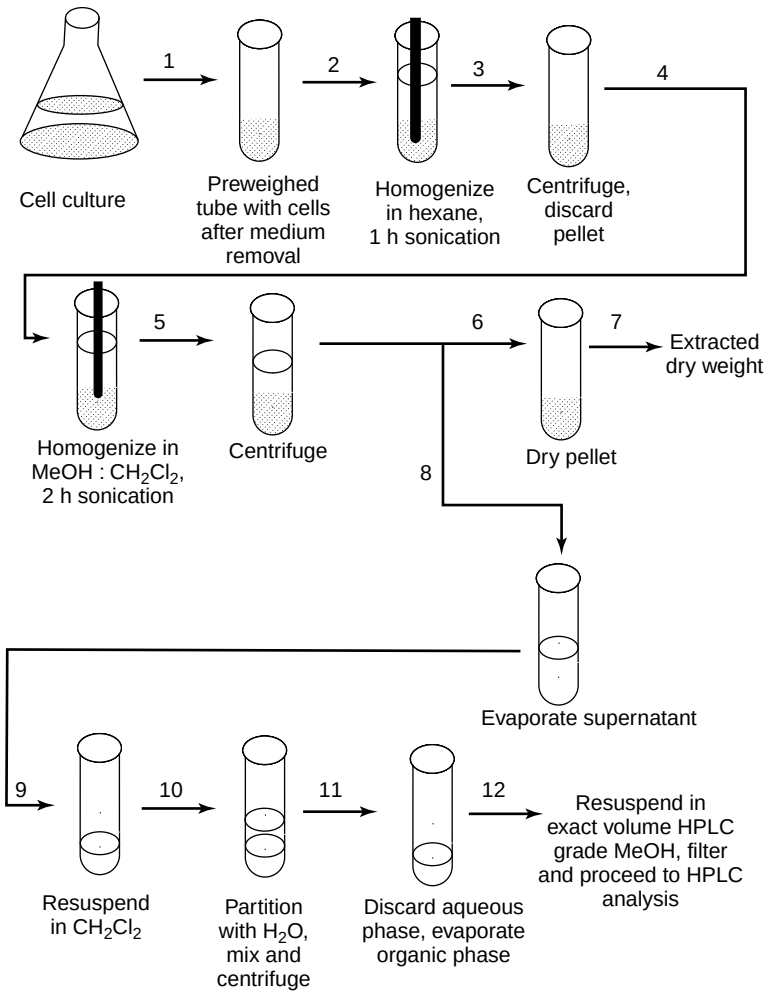

Fig. 2 Extraction of Taxol™ and taxanes from plant material.

PVP in the media. It may be necessary to use larger filters (keeping the pore size at 0.45 μm) and larger volumes of HPLC grade methanol to filter and resuspend extracts of large volumes of medium. Once filtered, these extracts can be further concentrated prior to HPLC analysis. Other alternatives to facilitate medium sample preparation are possible: for example, the removal of PVP by pore size exclusion prior to medium extraction or the substitution of the last filtration step by centrifugation prior to HPLC analysis.

3.2 Separation, Quantification and Identification

The taxane analyses are commonly done by HPLC. There are various chromatographic methods to carry out such analyses. One of these methods is described below.

The samples are passed through a reversed phase microbore column (e.g. Hewlett-Packard ODS Hypersil, 5 μm, 100 mm × 2.1 mm), using a mobile phase containing methanol, water, and acetonitrile; the linear gradient elution profile starts with 20 : 67 : 13 (solvent A), ends with 20 : 27 : 53 (solvent B) within 50 min, and includes a 5-min wash in both solvents to reestablish the initial condition. The flow rate is 0.2 mL min^{-1}, and chromatograms are plotted at 227 nm, using an HPLC system equipped with a photodiode array detector (e.g. Hewlett-Packard 1090A HPLC system). Typical injection volumes are 10 μL, and solvents are all HPLC grade, filtered through 0.45 μm filters prior to use.

Microbore columns have the advantages of high sensitivity, low sample volume, and low solvent consumption. For cell suspension extracts analyzed with the foregoing solvent system, however, a better resolution of early eluting taxanes, such as baccatin III and 10-deacetyl-baccatin III, can be achieved using longer reversed phase analytical columns; in this case, the flow rate is increased to 1 mL min^{-1}, whereas all other chromatographic and detection parameters are not changed.

Quantification is done by the method of external standards, that is, a standard curve is generated with authentic standard (at least five different concentrations covering the concentration range of the extract samples) and run under the same chromatographic conditions as the samples. Taxol™ and taxane identification is done by retention time, UV spectra and, cochromatography with standard. Peak purity checks can be performed as additional evidence for positive identification if the HPLC equipment has a diode array detector.

Further proof of identification can be obtained by collecting the corresponding HPLC peak and performing nuclear magnetic resonance (NMR) analyses. Over a series of injections, a concentrated sample is run through the HPLC system, the corresponding Taxol™ peak in the sample extract is collected and pooled, until approximately 50 μg of Taxol™ has been collected. The collected material is dried under vacuum at 25 °C and resuspended in 150 μL of HPLC grade methanol; then a small portion of it is injected again for purity evaluation and final quantification. If a single peak with the same retention time and UV spectrum as Taxol™ is present in the amount required, the sample is dried under vacuum as before, and freeze-dried for 24 to 48 hours to remove solvent residue. Taxol™ identification, for example, may be obtained by 1H NMR analysis in deuterated chloroform (≈50 μg of HPLC-purified sample Taxol™ in 0.75 mL of deuterated chloroform with 0.05% tetramethylsilane as internal reference) at 500 MHz and 25 °C on a Varian UNITY 500 spectrometer equipped with an inverse detection 5-mm probe. Taxol™ and related taxane standards may be requested for nonhuman research use from the Drug Synthesis and Chemistry Branch, Developmental Therapeutics Program, Division of Cancer Treatment of the National Cancer Institute (U.S.A.). Several taxanes are also available from chemical supply companies.

4
Taxol™ Biosynthesis

Taxol™ is a highly substituted and oxygenated cyclic diterpenoid with a taxane ring system. Its chemical uniqueness among taxanes is due to the presence

of an oxetane ring system, substitution pattern, and the ester side chains. Studies on Taxol™ biosynthesis have been conducted particularly by the groups of R. Croteau and H. Floss in the United States.

Considerable advances have been made both on the biochemical and molecular aspects of taxane biosynthesis. As diterpenes, the taxane skeleton has been shown to be derived from isopentenyl diphosphate (IPP), synthesized via the 1-deoxy-D-xylulose-5-phosphate (DXPS) pathway, on the basis of ^{13}C labeled glucose studies. This plastidic pathway functions in parallel with the classic mevalonate dependent route, which is now believed to be responsible for the assembly of IPP for the synthesis of sesquiterpenes, triterpenes, and phytosterols. However, the mevalonate pathway may be at least partly involved in taxane assembly. Virtually, all productive cultures of yew species are grown in dark (see below), indicating that the DXPS pathway is likely operating in proplastids or poorly differentiated chloroplasts, a scenario that is reinforced by the well established suppression of chloroplast differentiation and photosynthetic pathways by high sucrose levels used in culture media. The first committed step in the assembly of Taxol™ and related taxanes is the cyclization of the universal diterpenoid precursor geranylgeranyl diphosphate to taxa-4-(5), 11(12)-diene, a slow, but apparently not rate-limiting reaction, catalyzed by a diterpene cyclase (Fig. 3).

Cloning of the cDNA of taxadiene synthase in *T. brevifolia* by a homology-based PCR strategy revealed a 79-KDa peptide size and an N-terminal plastid transit peptide, in agreement with the plastid location of the diterpene biosynthesis. Heterologous expression of taxadiene synthase in bacteria followed by kinetic studies showed that, unlike several terpene synthases in plants, taxadiene synthase produces essentially one product and not multiple taxadiene isomers. The second specific step in taxane assembly involves the cytochrome P450-dependent hydroxylation at the C-5 position of the taxane skeleton, with allylic rearrangement, yielding taxa-4(20),11(12)-diene-5α-ol. This intermediate is acetylated at the hydroxyl in C-5 by an acetyl transferase to yield taxa-4(20),11(12)-dien-5α-yl acetate (Fig. 3). This is likely the third specific step in Taxol™ assembly and is responsible for generating the 4(20)-en-5α-acetoxy functional grouping from which the oxetane D-ring is established. A P450-dependent monooxygenase catalyzes the hydroxylation of the later molecule, resulting in taxa-4(20),11(12)-dien-5α-acetoxy-10β-ol. An additional cytochrome P450 cDNA clone encoding a product capable of hydroxylating taxadien-5α-ol at the C-13 position has also been identified.

The exact order of hydroxylation steps and the sequence of intermediates in the Taxol™ biosynthetic pathway are not clearly established, and it may be possible that various interconnected routes to this compound exist. The same applies to acylation reactions and the exact timing of epoxidation. After oxygenation, various hydroxylated positions in the taxane skeleton undergo esterification.

These acylation reactions appear to span early and later steps in the pathway. Whereas acetylation of the C-5 hydroxyl group occurs early in biosynthesis, acetylation at the C-10 hydroxyl and benzoylation at the C-2 hydroxyl take place later. Clones of cDNAs encoding these three taxane *o*-acyltransferases have been identified. The oxetane ring could arise by epoxidation of the 4(20)-double bond of the α-acetoxy group, followed by migration of

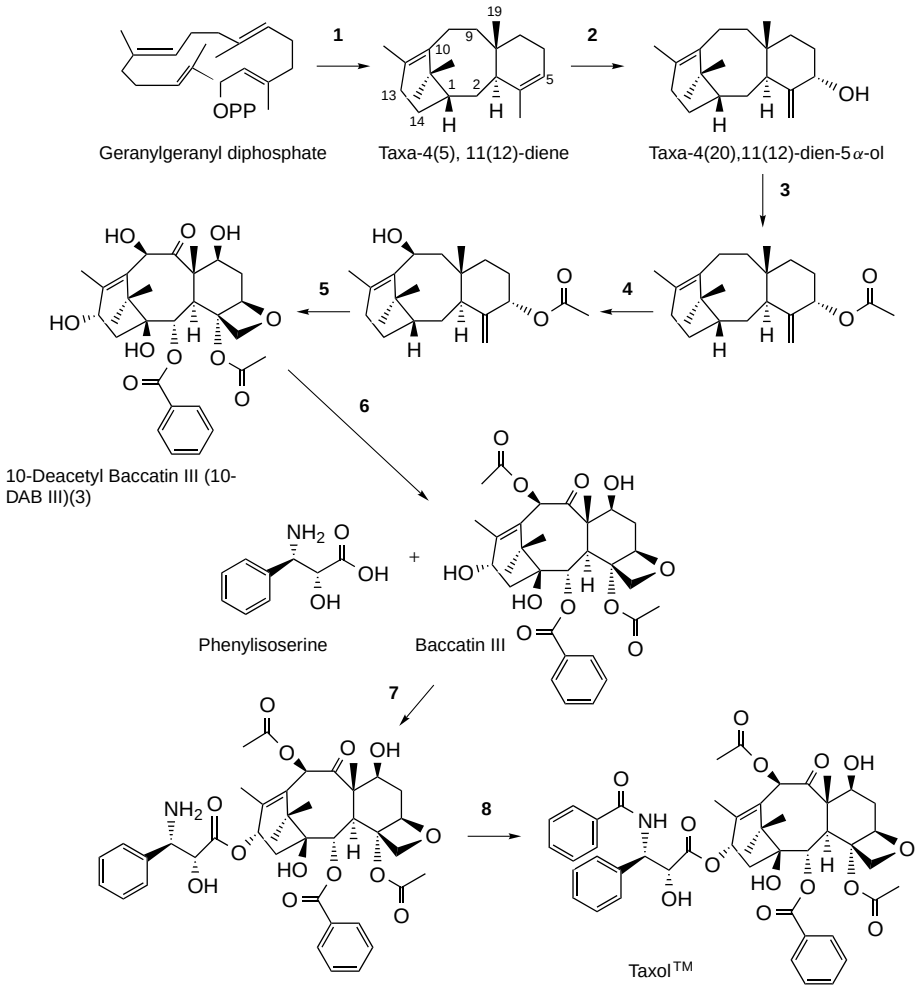

Fig. 3 A scheme of Taxol™ biosynthesis. **1**. Taxadiene synthase, **2**. Cytochrome P450 taxane-13α-hydroxylase, **3**. Taxa-4(20), 11(12)-dien-5α-*o*-acetyl transferase, **4**. Cytochrome P450 taxane-10β-hydroxylase, **5**. Multiple steps, **6**. 10-deacetyl baccatin III acetyl transferase, **7**. C-13 side chain addition, **8**. *N*-benzoylation.

the α-acetoxy group from C-5 to C-4, with oxirane-to-oxetane ring expansion. The attachment of the C-13 side chain to the taxane ring seems to involve initial esterification with phenylisoserine, followed by *N*-benzoylation. Phenylisoserine would be generated by an aminomutase reaction of phenylalanine to β-phenylalanine, followed by α-hydroxylation (Fig. 3).

5 General Characteristics of Taxol™ Yield and Growth in *Taxus* Cell Culture

This section focuses on the detailed analysis of a set of cell lines of *T. cuspidata* developed in the early and mid 90s at the University of Toronto to characterize the relationships and general features

of taxane production and cell growth. Early experiments with callus and cell cultures of *T. cuspidata* have shown that growth and Taxol™ yield are generally inversely proportional. The reason for this phenomenon, also observed with the production of other secondary metabolites by a number of plant species, is not well understood. Possible explanations include the lack of expression of key enzyme and/or transcriptional activator genes in cells with low differentiation, diversion of substrates for secondary metabolism into growth processes, unregulated catabolism, and lack of end product transport systems and storage sites.

The analysis of Taxol™ yields in different *in vitro* cultures of *T. cuspidata* supports the observation that faster growth rates and/or lower degrees of differentiation are normally associated with lower Taxol™ amounts. Stem-derived explants of *T. cuspidata* after two months of culture in B5bPVP displayed an average eightfold increase in fresh weight and Taxol™ yields of $0.020 \pm 0.005\%$ of the extracted dry weight (equivalent to the yield of stems of the intact plant). Slow-growing, six-month-old immobilized cell cultures of *T. cuspidata* grown in B5C2 had variable amounts of Taxol™, up to $0.012 \pm 0.007\%$ of the dry weight (equivalent on average to the bark of the intact plant), whereas six-month-old immobilized cell cultures visually selected for faster growth yielded $0.0014 \pm 0.0003\%$ of Taxol™ on a dry weight basis. A two-year-old cell suspension line of *T. cuspidata* grown in B5C2 (dry weight doubling time of 17 days) yielded $0.0014 \pm 0.0005\%$ of Taxol™ (dry weight basis) when cultured as 300-mL suspensions in 1 L Erlenmeyer flasks. These yields may be altered by precursor feeding and various other strategies (see further sections).

Studies on the kinetics of growth and Taxol™ accumulation in cell suspensions of *T. cuspidata* (cell line FCL1F) showed a relatively slow growth rate and confirmed the relationship between growth and Taxol™ yield. The doubling time was approximately 17 days for fresh weight and 21 days for dry weight. Although relatively long in contrast to data from herbaceous species, these doubling times are comparable to or even shorter than those of other woody species that grow slowly (Table 2). The maximum specific growth rates were $0.040\ g^{-1}$ (fresh weight) and $0.034\ g^{-1}$ (dry weight). The accumulation of Taxol™ in the cells was not growth linked. The highest taxol accumulation values occurred in cells in stationary phase, and Taxol™ amounts in the cells decreased with the increase in cell biomass (Table 3).

Other taxanes in yew callus and cell cultures have been detected by HPLC, such as baccatin III, 10-deacetyl-baccatin III, 1-dehydroxybaccatin V, 7-epitaxol, and cephalomannine in both types of culture of *T. cuspidata*. Baccatin III yields in two-month-old callus of this species grown in B5bPVP were $0.043 \pm 0.016\%$ of the extracted dry weight. A two-year-old cell line suspension of *T. cuspidata* (FCL1) has been estimated to yield $0.018 \pm 0.010\%$ of baccatin III (extracted dry weight basis).

Tab. 2 Growth parameters of *T. cuspidata* cell suspensions in comparison to some other plant species (shake-flask experiments).

Species	*Doubling time [days]*	*Specific growth rate [day^{-1}]*
Taxus cuspidata	20.6	0.034
Brucea javanica	32.0	
Callitris drummondi	40.0	
Catharanthus roseus	3.0	0.32

Tab. 3 Growth and Taxol® accumulation in cells and in medium of suspensions of *T. cuspidata*[a].

Time [days]	Fresh weight [g mL^{-1}]	Taxol® accumulated[b]	
		In cells [μg g^{-1} EDW]	In medium [μg mL^{-1}]
0	0.15	4.0	0.013
6	0.20	nd	0.356
12	0.24	2.0	0.060
18	0.32	nd	bd
25	0.35	0.4	bd
32	0.39	1.2	bd
38	0.28	2.8	0.100
44	0.27	3.2	0.050

[a]Values are the mean of two to three determinations; inoculi were 1/6 w/v obtained from a single stationary phase donor suspension.
[b]EDW, estimated dry weight; nd, not determined; bd, below limit of detection.

Taxanes such as 10-deacetyl-baccatin III and baccatin III can be used to produce Taxol™ through semisynthesis. These taxanes can also be used to produce Taxotere™, a semisynthetic Taxol™ analog with higher antineoplastic properties and higher water solubility than Taxol™; Taxotere™ was first developed by P. Potier and associates and Rhône-Poulenc Rorer in France.

Media collected from six-month-old immobilized cell suspensions of *T. cuspidata* and extracted with dichloromethane may contain trace amounts of Taxol™ and baccatin III as indicated by HPLC analysis. In differentiated physiologically active cells, Taxol™ may be stored inside the cells, possibly in the vacuole or other membranous compartment. In a detailed time course study using the FCL1F cell suspension line of *T. cuspidata*, the analysis of the volumetric total production of Taxol™ at stationary phase indicated that 66% of the total Taxol™ was in the medium. In contrast, at a given time, Taxol™ accumulation in the medium on a per-milliliter basis was generally a small portion of that contained per gram of extracted dry weight basis in the cells (Table 3). The occurrence of a larger extracellular component of the total volumetric Taxol™ at stationary phase (productive phase) may be the result of excretion by the cells that reach highest amounts of Taxol™ accumulation in this phase. Studies on the production of various secondary products by plant cell cultures show that the proportion of excreted to endogenously accumulated product can be highly variable depending on species and cell line.

Taxol™ accumulation in the medium of cell suspensions of *T. cuspidata* can vary considerably with the time of the growth cycle. In the kinetic study just mentioned, the accumulation of Taxol™ in the medium had two peaks (Table 3). The first peak of Taxol™ in the medium (≈0.4 μg mL^{-1}) occurred in the early portion of the exponential growth phase, soon after inoculation in fresh medium. This early release of Taxol™ in the medium could be a response to the transfer into fresh medium ("transfer effect") and/or to the increase of the internal pH of the cells due to the absorption of ammonium (exchanged with H^-), which takes place in this portion of the growth cycle. Following the first peak, Taxol™ was degraded, possibly by enzymes released into the medium. A second peak of Taxol™ in the medium (0.1 μg mL^{-1}) occurred at stationary phase and may have resulted from excretion by the cells that reach their highest level of Taxol™ at this stage and/or from a small degree of cell lysis, common at later stages of the growth cycle. Baccatin III accumulation in the medium peaked after approximately nine days of culture at 6.2 μg mL^{-1} in the

same cell line suspension. The occurrence of product excretion into the medium is useful for industrial application of plant cell culture technology, since it facilitates downstream processing (especially when using immobilized cells).

From the mid-1990s on, a large research effort from various groups brought Taxol™ and taxane productivity levels in cell cultures to progressively higher values, in some cases reaching industrially interesting amounts. In the next section, different proven strategies to promote taxane yield in cell cultures are described.

6 Techniques to Improve Taxol™ Yield in Cell Culture

6.1 Precursor and Sugar Feeding

The problem of overall low Taxol™ yields, particularly at higher growth rates, can be partly overcome by feeding Taxol™ precursors to the cell cultures. The appropriate combination of type, amount, and feeding time of precursors able to enter the cells can significantly increase the yield of secondary metabolites in plant cell cultures, provided the biosynthesis of the metabolite of interest is at least partly limited by poor activity of enzymes involved in precursor formation or appropriate supply of precursor to the biosynthetic reaction. In two-month-old callus cultures of *T. cuspidata*, phenylalanine supplementation at 0.1 mM in B5bPVP can double the Taxol™ yield in relation to regular B5bPVP without affecting growth significantly; similar results can be obtained with cell suspensions after a 25-day growth period in medium supplemented with the same amount of this amino acid. Phenylalanine acts as a precursor for the side chain of Taxol™ and the benzoyl moiety at C-2.

Other potential precursors for the Taxol™ side chain were investigated in feeding experiments to both callus and cell suspension cultures of *T. cuspidata*. Compounds that significantly enhanced Taxol™ yields in cultures were phenylalanine, benzoic acid, hippuric acid, serine, and glycine. Impacts on Taxol™ yields were concentration dependent. Taxol™ yields were increased by two to four times when the putative precursors were provided at concentrations of 0.2 mM for callus cultures and 0.05 mM for cell suspension cultures; all cultures were fed at the onset of the stationary phase of growth. The best results were obtained with phenylalanine or benzoic acid feeding.

In a related approach, replacement of 2,4-D by Indole 3 acetic acid–glycine or Indole 3 acetic acid–phenylalanine conjugates yielded a two and threefold increase in Taxol™ production by cell cultures of *T. wallichiana* (Himalayan yew), respectively. Results suggest that precursor feeding is a promising approach to increase Taxol™ yields in yew cell cultures without detrimental effects on biomass generation.

Kinetic studies of taxane production in cell cultures showed that carbohydrate availability is closely related to biomass increase in *Taxus* cell cultures. Feeding of sucrose to *Taxus chinensis* cultures initially kept at 20 gL^{-1} and fed-batch system increased growth and taxane production; intermittent feeding of various concentrations of sucrose or maltose improved Taxol™ production in the same species.

An important aspect to consider is the osmotic effect of sugars on taxane production. It has been shown that high concentration of sucrose (60 gL^{-1}), combinations of sucrose and mannitol, as well as nonsugar osmotic agents

(polyethyleneglycol – PEG 4,000) were capable of improving Taxol™ yield in *T. chinensis* cell cultures. Cells cultivated in PEG-supplemented medium yielded a threefold increase of Taxol™ over the control. The authors have suggested that osmotic stress may improve secondary metabolism by directly modulating enzymes or by reducing growth rates and making more carbon available for secondary metabolism.

6.2 Cell Line Selection

Screening different cell lines for high yields of Taxol™ and/or taxanes may lead to the establishment and propagation of elite cell lines that can be used for large-scale production purposes. The technique has been successfully applied, for instance, to the production of ajmalicine by *C. roseus*, shikonin by *Lithospermum erythrorhizon*, and berberine by *Coptis japonica*. Although laborious and time-consuming, the approach may turn out to be advantageous for long-term potential industrial applications of yew cell cultures as a source of Taxol™ and taxanes.

The presence of a visual feature associated with high productivity is highly desirable for selecting high yielding cell lines (e.g. colored product, such as anthocyanin). In the case of *T. cuspidata*, additional studies are required to identify a reliable visual feature associated with taxane productivity. When such a facilitating screening parameter is not present at all, however, as it often happens, the use of standard phytochemical extraction and separation procedures is required or, if the product of interest can be made antigenic, immunoassays can be used in the screening procedure. Antitaxol and several other antitaxane antibodies are commercially available.

6.3 Light and Temperature

These physical parameters can be used to manipulate taxane production. The vast majority of protocols for yew cell cultures uses dark incubation. In a study on the effects of light on heterotrophic *T. cuspidata* cell suspension cultures and Taxol™ yield, it was established that white light of low intensity caused significant reduction in Taxol™ yield and growth; moreover, light-grown cells retained more intracellular Taxol™ and had much lower amounts of the taxane in the medium fraction (only 25% of Taxol™ was released to the medium versus 65% in dark-grown suspensions), presumably contributing to negative feedback of biosynthetic enzymes. Diversion of the phenylalanine pool from taxane biosynthesis toward synthesis of phenolic compounds in light-grown cells is another possible reason for light inhibition of taxol production. The effects of light have been shown to be the result of a direct effect of light on cell physiology and not photochemical degradation of medium components or taxanes, as judged from HPLC analysis of light-exposed Taxol™ solutions and growth performance in light irradiated versus control medium.

Temperature is generally maintained at approximately 25 °C for yew cell culture studies. A detailed investigation on suspensions of *T. chinensis* established that the optimum temperature for cell growth was 24 °C, whereas for Taxol™ production this parameter was 29 °C. By applying a strategy of temperature shift (from 24 to 29 °C) after a high cell biomass was achieved, a sustained higher Taxol™

production was obtained; whereas a reduction in the yield of the main taxane in this species, taxuyunnanine C, was observed. These results suggested that different branches of a metabolic pathway might have different optimal operating temperature conditions, and the temperature shifting strategy must be tailored to the specific target compound.

6.4 Elicitation and Metabolic Inhibitors

Elicitation is the stimulation of secondary product formation by specific compounds or physical conditions often mimicking environmental stresses. It has been successfully applied to improve Taxol™ yields in cell cultures. Jasmonate, a lipid-derived hormone, known as a signal molecule for herbivore and pathogen defense responses in plants, and related compounds have been used to induce taxane overproduction in *T. media, T. baccata and T. brevifolia* cell suspensions. Methyl jasmonate (MJ) added at 100 μM in medium at the time of cell inoculation has promoted Taxol™ and baccatin III accumulation to various extents (e.g. Taxol™ from 0.120 to 0.606% dw in *T. media* and 0.002 to 0.229% dw in *T. baccata*), with only moderate decreases in cell yield. Optimal configuration of MJ stereoisomers has been investigated to maximize taxane induction and minimize cell growth inhibition. A crude fungal extract of *Fusarium oxysporum* caused apoptosis in cells of *T. chinensis* var. *mairei*, as detected by microscopic and DNA analyses; Taxol™ yields tripled in treated cells.

Abiotic elicitors such as the heavy metals silver, cobalt, cerium, and lanthanium have also been used to stimulate taxane production in cell cultures. Enhanced Taxol™ production and release in *T. chinensis* cell cultures was obtained by a combination of ultrasound (US) treatment, MJ exposure and *in situ* solvent extraction (two-phase culture, see below). Ultrasound applied for 2 min once or twice during a four-week culture period increased volumetric Taxol™ yield 1.5- to 1.8-fold; MJ at 60 to 120 μM resulted in 5-fold increase, and *in situ* solvent extraction, 7- to 9-fold. The most effective combined approach was the application of US or MJ, followed by *in situ* solvent extraction, affording 33- to 35 mgL^{-1} of Taxol™, a 17-fold increase over control levels.

Suspension cultures are generally heterogeneous in the cell cycle phase of individual cells, which contribute to variable and relatively low magnitude production responses upon elicitation. Stable and optimized Taxol™ yields have been achieved in *T. chinensis* cultures by synchronization induced with phosphate starvation and low temperature treatment. Phosphate starvation was induced by transferring 15d old cells to phosphate-free medium for 25 days, followed by a second transfer to phosphate-free medium for additional 25 days. Low temperature treatment involved two cycles of 72 h exposure to 4 °C, each followed by transfer to 24 °C for 24 h. Synchronized cultures, particularly by low temperature treatment, doubled Taxol™ yields in response to elicitation with MJ relative to asynchronous cultures elicited in the same way.

Metabolic inhibitors have been examined as a means to divert resources from competing branch pathways towards the main pathway leading to the compound of interest. Since phenylalanine acts as a precursor to the phenylisoserine side chain of Taxol™ and the benzoyl moiety at C-2, attempts were made to inhibit the phenylpropanoid pathway at the level of phenylalanine ammonia lyase (PAL), expecting an increase in phenylalanine availability

for Taxol™ biosynthesis. However, the use of three different PAL inhibitors had little or inhibitory effect on Taxol™ yield, suggesting that more specific steps of the phenylpropanoid pathway need to be evaluated in order to achieve practical results.

Inhibitors have also been used to study the contribution of the mevalonate and 1-deoxy-D-xylulose (DXP) pathway to Taxol™ and baccatin III accumulation in *T. baccata* cell suspensions. Both mevinolin (inhibitor of the mevalonate branch, acting on HMG-CoA reductase) and fosfomidomycin (inhibitor of the DXP branch, acting on DXP reducto- isomerase) caused decreases in Taxol™ accumulation, whereas baccatin III was only reduced by fosfomidomycin. The results indicate that both pathways may be used for Taxol™ assembly with a predominance of the non-mevalonate one.

6.5 Gas Composition

The gas phase of flasks and bioreactors is often a determinant of secondary product accumulation. The effects of gas phase composition on Taxol™ was examined in shake-flask suspensions of *T. cuspidata*. Low head space oxygen (10% v/v) promoted early production of Taxol™, whereas high CO_2 (10% v/v) had an inhibitory effect. The same authors examined the role of ethylene, a gaseous phytohormone involved in various stress responses, in Taxol™ production. Gas composition also affected sugar, calcium, and phosphate uptake, which may have influenced taxane metabolism. The best combination for Taxol™ production was 10% (v/v) oxygen, 0.5% carbon dioxide, and 5 ppm ethylene. However, ethylene may have a dual effect on Taxol™ yield. In fungal elicitor-induced cell suspension of *T. chinensis* and *T. yunnanensis*, ethylene biosynthesis inhibitors (e.g. Co and Ni salts) and the ethylene action inhibitor $AgNO_3$ enhanced Taxol™ yield (up to threefold by combining Co and Ag); the opposite was observed with the ethylene-releasing compound ethrel. Such positive and negative effects were observed when adjuvants were added before or simultaneously with the elicitor, becoming progressively negligible when added 24 h after elicitation. Control of ethylene action and production during fungal elicitation appears to favor Taxol™ accumulation.

6.6 Protoplast Cultures

In general, protoplasts have not been used for useful metabolite production but as an important material for cell fusion and genetic transformation. One of the major advantages of using protoplasts is that the products are released freely into the broth with the double consequences of increasing overall productivity and facilitating down stream processing. Furthermore, since many high molecular weight substances (including elicitors) that are otherwise not accessible to the cells (due to the presence of cell wall) can directly access the cell membrane in protoplasts, the composition and concentrations of the metabolites produced by protoplasts may vary from those produced by cells. However, protoplasts are very fragile and cannot be used for long-term production process, and active protoplasts easily regenerate their cell walls during cultivation. In order to solve these problems, various gels were used as immobilization carriers (providing stability to the protoplasts) combined with an inhibitor of cell wall synthesis. Thus, active protoplasts were maintained for a long time without

cell wall regeneration. Immobilizing protoplasts is akin to providing them with an artificial cell wall. This concept has merits in that various useful functions can be freely incorporated in protoplasts by using artificial cell walls with such functions. Production of Taxol™ using *Taxus* protoplasts with artificial cell walls was feasible.

The optimal condition for isolating large amounts of viable protoplasts from *T. cuspidata* cultured cells was investigated. With a combination of cellulase, pectin lyase, polygalacturonase, hemicellulase, potassium citric acid buffer solution, bovine serum albumin (BSA), and potassium dextran sulfate and degassing treatments were found to be effective in increasing the yield of protoplasts isolated from *T. cuspidata* cultured cells. Protoplasts yields ($3.2–6.4 \times 10^6$ number/g-fresh weight cells) were achieved with specific culture and cell phases. This condition and these factors are thought to be useful for isolation of various *Taxus* species protoplasts.

The localization of Taxol™ in cultured *T. cuspidata* cells was investigated using protoplast isolation reaction. The amount of Taxol™ released by protoplasts isolation reaction was considered to be located in the cell wall and/or between cell wall (CW) and cell membrane. The amount of paclitaxel in the protoplasts was considered to be located in the cytoplasm. About 30% and 35% of total Taxol™ was located in CW of suspension cells in the growth phase and in the stationary phase, respectively (Table 1). It was also found that about 30% and 43% of total Taxol™ was located in CW of the cells grown in solid culture at growth phase and at the stationary phase, respectively (Table 4).

Since some of the Taxol™ is located in CW and the accumulation of Taxol™ in the medium was low, it was assumed that the cell wall limits the excretion of Taxol™ and that the amount of Taxol™ in CW may affect production of the same compound. It should be noted that Taxol™ production by *Taxus* cells is controlled by feedback inhibition due to the accumulation of Taxol™ in CW. Therefore we considered that the removal of cell walls would enhance Taxol™ production and developed a novel production system for isolated protoplasts. Figure 4 shows the Taxol™

Tab. 4 Localization of Taxol™ in *T. cuspidata* cells (1.0-g fresh weight) in suspension and solid cultures.

Phase	*Taxol™ in CW[a] [μg]*	*Taxol™ in cytoplasm [μg]*	*Ratio[b] [%]*
Suspension culture			
Growth phase (15 days)	6.6 ± 0.8	15.0 ± 1.4	30.5 ± 1.9
Stationary phase (22 days)	11.7 ± 0.8	21.6 ± 1.6	35.1 ± 1.0
Solid culture			
Growth phase (20 days)	9.9 ± 0.9	23.0 ± 2.2	30.1 ± 1.3
Stationary phase (35 days)	22.8 ± 0.8	30.0 ± 1.8	43.2 ± 1.4

[a]CW; cell wall and/or between cell wall and cell membrane.
[b]Ratio; [Taxol™ in CW/(Taxol™ in CW + Taxol™ in cytoplasm)] × 100.
Numbers represent mean value ± S.E.

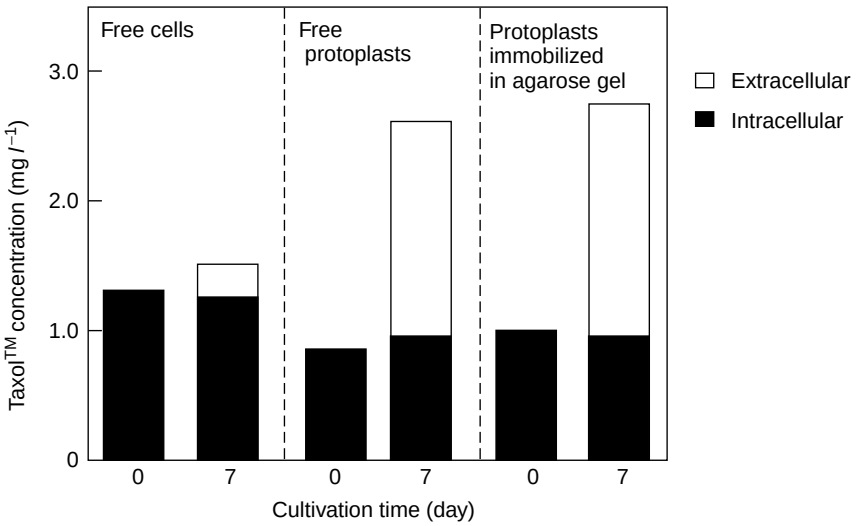

Fig. 4 Production of Taxol™ by *T. cuspidata* cells and free and immobilized protoplasts.

production by free cells and protoplasts in a static culture. Protoplasts produced a sixfold increase in Taxol™ in the medium compared to that of the free cells. In order to improve the stability of protoplasts, 3.0% agarose gel was used as the artificial cell wall material so that they could be cultivated by shaking culture without disruption. Agarose gel provided an artificial mechanical support to the protoplasts thus enabling them to withstand hydrodynamic stress. The protoplasts immobilized in agarose gel produced Taxol™ and released the compound into the medium as did protoplasts in static culture, yielding about six times more Taxol™ than free cells.

The system described here may be useful for the production of Taxol™ and other taxanes using different *Taxus* cell lines. Agarose was employed as an artificial cell wall and 2,6-dichlorobenzonitrile, for example, was used as inhibitor of cell wall synthesis. This system has merits in that various useful functions can be freely incorporated into protoplasts by using artificial cell walls with engineered functions. By using an appropriate artificial cell wall and an inhibitor of cell wall synthesis, it may be possible to further promote Taxol™ production. For example, Indole alkaloids production by *C. roseus* protoplasts was synergistically enhanced by using alginate gel as both elicitor and immobilization carrier, and $CaCl_2$ was used for both inhibitor of cell wall synthesis and signal transducer. Protoplasts with artificial cell walls have important potential applications.

The numerous approaches described above for enhancing taxane production in cell suspension cultures may be examined for Taxol™ production by protoplast cultures. Protoplast cultures provide a valuable tool for detailed study of biosynthetic pathways leading to efficient production of Taxol™. Furthermore, protoplast cultures represent a sustainable and relatively clean source of enzymes and genes involved in Taxol™ biosynthesis.

6.7 Culture System

Various culture systems have been tested for taxane production by cell cultures, both with free and immobilized cultures (glass fiber, alginate, etc.). Use of 50% conditioned medium instead of full replacement of medium has proven useful for the growth of *T. chinensis* and production of taxuyunnanine C. Two-phase culture systems involve the presence of a nonpolar phase to capture by partitioning the product of interest, thereby alleviating feedback inhibition on biosynthetic enzymes of cells in the aqueous phase. Product purification is also facilitated in this system. Accumulation sites for compounds of interest of the solid resin type (e.g. Amberlite XAD series) have promoted Taxol™ yield in *T. cuspidata* by 40 to 70%.

Organic solvents used in two-phase culture systems have also been used for taxane production in cell cultures; application of 10 to 11% (v/v) dibutylphthalate as the *in situ* solvent extraction has improved Taxol™ with minimal deleterious effects on cell growth and integrity.

Two-stage cultivation, involving a growth and a production medium, has been largely used for taxane production in cell cultures. The most frequent strategy involves addition of precursors and elicitors or the exposure to abiotic factors (such as temperature shifts) after a certain cell biomass has been attained.

Semicontinuous and perfusion cultures, which involve medium replacement throughout the growth cycle, have been used for Taxol™ production studies. In *T. canadensis*, a fourfold increase in accumulation was observed in semicontinuous cultivation with total cell recycle versus batch culture. Alginate immobilized cultures and free cells with a nylon-mesh separator of *T. cuspidata* were cultivated in continuous medium perfusion mode and resulted in an order of magnitude higher Taxol™ specific production (0.3 mg gdw^{-1} day^{-1} for 40 days) relative to batch cultivation; however, some degree of cell growth inhibition owing to a dilution effect was observed.

7 Scale-up of Taxus Cell Cultures

The scale-up of plant cell cultures remains a challenge in spite of considerable efforts to bring this technology to commercially viable use. In general, as culture systems move from flask to bioreactor and further up to large-scale cultivation, productivity and growth drop considerably. Problems during scale up include shear stress, control of gas composition, lack of adequate mixing of medium and cells, and oxygen supply. Yew cell cultures have been cultivated in bioreactors, in a large scale, to a significant degree of success compared to other species capable of accumulating industrially interesting chemicals.

The general strategy used for scale up cultivation of yew cell cultures has been to combine some of the above-described techniques and adapt their application to the reactor setting to improve yields. Some examples of bioreactor cultures of yew species are depicted in Table 5. Taxol™ from cell culture–based large-scale production has been commercialized since 2001 by Samyang Genex (Taejon, Korea).

8 Future Perspectives

Data to date shows that cultured cells of *Taxus* spp. represent a viable source

Tab. 5 Examples of bioreactor production of taxanes to industrially significant amounts.

Species	*Equipment*	*Taxane*	*Production [mg L^{-1}]*	*Productivity [mg L^{-1} × d]*	*Treatment*
Taxus media	Stirred reactor, 5 L capacity	Taxol™ Baccatin III	21.1 56.0	1.06 2.8	Methyl jasmonate, Precursor feeding, 2-stage media
T. cuspidata	Baloon (bubble reactor), 20 to 500 L	Total taxanes	74	2.74	Feeding or medium replacement
T. chinensis	Bubble column, 1 L	Taxuyunnanine C	229	13.5	Ethylene addition
	Airlift, 1 L	Taxuyunnanine C	336	20.2	Methyl jasmonate, ethylene addition and sucrose feeding
	Airlift, 1 L	Taxuyunnanine C	612	24.7	Repeated methyl jasmonate application and sucrose feeding

of Taxol™ and related taxanes, which can be used in the production and semisynthesis of this antineoplastic agent and its analogs. Scaled-up cell cultures and alternate modes of taxane production, such as protoplasts and perfusion culture systems need to be further developed by combining the knowledge that has been successfully applied to suspension cultures with the particular features of these culture methods.

The molecular characterization of taxane biosynthetic genes, the availability of their sequences, and protein products expressed in microorganisms for detailed kinetic studies open important perspectives for the establishment of engineered yew cell cultures transformed with genes controlling slow or limiting biosynthetic steps driven by strong or easily inducible promoter sequences. The knowledge of cis elements in promoter sequences of taxane biosynthetic genes may shed light on their modes of regulation and contribute to their efficient manipulation. Alternatively, antisense transformation or the use of RNA interference techniques may be useful for the suppression or inhibition of competing pathways, making more carbon, nitrogen, and energy flux available for relevant taxane assembly. In that regard, the recently observed preferential induction of C-13 oxygenated taxanes (more pharmaceutically relevant) by MJ in *T. media* cell cultures points to taxoid 14β-hydroxylase, a recently cloned cytochrome P450-dependent monooxygenase, as an interesting target for inhibition in cultures that produce C-14 oxygenated taxanes. As the molecular basis of taxane biosynthesis if further detailed, the identification of trans-activating factors (DNA binding proteins) controlling the coordinated expression of multiple biosynthetic genes

may allow for more efficient and rational control of the metabolic flux toward specific taxanes of interest. Genetically engineering cell cultures with these "master" regulatory genes may facilitate the control of the complex pathways leading to Taxol™ and related taxanes. Interestingly, recent progress has been made in the control of the pathways leading to monoterpene indole alkaloids in *C. roseus* by genetically modifying the expression of the ORCA family of transcription factors, which regulate the expression of biosynthetic genes involved in the indolic portion of the pathway and which are activated by jasmonate and related compounds. Since taxane biosynthesis is also strongly induced by jasmonate, it is possible that homologs of these transcription factors operate in yew species. Other potential targets for genetic modification in yew cell cultures include gene products responsible for taxane transport in and out of cells and between subcellular compartments.

The continued analysis and characterization of new taxanes produced in cell cultures may directly yield or provide leads to the development of pharmaceutically more efficient Taxol™ analogs. Moreover, new and important bioactivities may be uncovered from the vast chemical array of taxanes.

See also Medicinal Chemistry.

Bibliography

Books and Reviews

Jennewein, S., Croteau, R. (2001) Taxol: biosynthesis, molecular genetics and biotechnological applications, *Appl. Microbiol. Biotechnol.* **57**, 13–19.

Memelink, J., Verpoorte, R., Kinje, J.W. (2001) ORCAnization of jasmonate-responsive gene expression in alkaloid metabolism, *Trends Plant Sci.* **6**, 212–219.

Walker, K., Croteau, R. (2001) Taxol biosynthetic genes, *Phytochemistry* **58**, 1–7.

Zhong, J.J. (2002) Plant cell culture for production of paclitaxel and other taxanes, *J. Biosci. Bioeng.* **94**, 591–599.

Primary Literature

Aoyagi, H., DiCosmo, F., Tanaka, H. (2002) Efficient paclitaxel production using protoplasts isolated from cultured cells of *Taxus cuspidata*, *Planta Med.* **68**, 420–424.

Aoyagi, H., Sakamoto, Y., Asada, M., Tanaka, H. (1998) Indole alkaloids production by *Catharanthus roseus* protoplasts with artificial cell walls containing guluronic acid rich alginate gel, *J. Ferment. Bioeng.* **85**, 306–311.

Brincat, M.C., Gibson, D.M., Schuler, M.L. (2002) Alterations in taxol Production in plant cell culture via manipulation of the phenylalanine ammonia lyase pathway, *Biotechnol. Prog.* **18**, 1149–1156.

Choi, H.K., Kim, S.I., Son, J.S., Hong, S.S., Lee, H.S., Lee, H.J. (2000) Enhancement of paclitaxel production by temperature shift in suspension culture of *Taxus chinensis*, *Enzyme Microb. Technol.* **27**, 593–598.

Cusidó, R.M., Palazón, J., Bonfill, M., Navia-Osorio, A., Morales, C., Piñol, M.T. (2002) Improved paclitaxel and baccatin III production in suspension cultures of *Taxus media*, *Biotechnol. Prog.* **18**, 418–423.

Dong, H.D., Zhong, J.J. (2002) Enhanced taxane productivity in bioreactor cultivation of *Taxus chinensis* cells by combining elicitation, sucrose feeding and ethylene incorporation, *Enzyme Microb. Technol.* **31**, 116–121.

Eisenreich, W., Menhard, B., Hylands, P.J., Zenk, M.H. (1996) Studies on the biosynthesis of taxol: the taxane carbon skeleton is not of mevalonoid origin, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6431–6436.

Fett-Neto, A.G., DiCosmo, F., Reynolds, W.F., Sakata, K. (1992) Cell culture of *Taxus* as a source of taxol and related taxanes, *Biotechnology* **10**, 1572–1575.

Fett-Neto, A.G., Melanson, S.J., Sakata, K., DiCosmo, F. (1993) Improved growth and taxol yield in developing calli of *Taxus cuspidata* by media composition modification, *Biotechnology* **11**, 731–734.

Fett-Neto, A.G., Melanson, S.J, Nicholson, S.A., Pennington, J.J., DiCosmo, F. (1994b) Improved taxol yield by aromatic carboxylic acid and amino acid feeding to cell cultures of *Taxus cuspidata*, *Biotechnol. Bioeng.* **44**, 967–971.

Fett-Neto, A.G., Pennington, J.J., DiCosmo, F. (1995) Effect of white light on taxol and baccatin III accumulation in cell cultures of *Taxus cuspidata* Sieb and Zucc, *J. Plant Physiol.* **146**, 584–590.

Fett-Neto, A.G., Zhang, W.Y., DiCosmo, F. (1994) Kinetics of taxol production, growth, and nutrient uptake in cell suspensions of *Taxus cuspidata*, *Biotechnol. Bioeng.* **44**, 205–210.

Fleming, P.E., Knaggs, A.R., He, X.-G., Mocek, U., Floss, H.G. (1994) Biosynthesis of taxoids: mode of attachment of the taxol side chain, *J. Am. Chem. Soc.* **116**, 4137–4138.

Jha, S., Sanyal, D., Ghosh, B., Jha, T.B. (1998) Improved taxol yield in cell suspension culture of *Taxus wallichiana* (Himalayan yew), *Planta Med.* **64**, 270–272.

Ketchum, R.E.B., Rithner, C.D., Qiu, D., Kim, Y.S., Williams, R.M., Croteau, R.B. (2003) *Taxus* metabolomics: methyl jasmonate preferentially induces production of taxoids oxygenated at C-13 in *Taxus x media* cell cultures, *Phytochemistry* **62**, 901–909.

Kim, S.I., Choi, H.K., Kim, J.H., Lee, H.S., Hong, S.S. (2001) Effect of osmotic pressure on paclitaxel production in suspension cell cultures of *Taxus chinensis*, *Enzyme Microb. Technol.* **28**, 202–209.

Mirjalili, N.,, Linden, J.C. (1995) Gas phase composition effects on suspension cultures of *Taxus cuspidata*, *Biotechnol. Bioeng.* **48**, 123–132.

Palazón, J., Cusidó, R.M., Bonfill, M., Morales, C., Piñol, M.T. (2003) Inhibition of paclitaxel and baccatin III accumulation by mevinolin and fosmidomycin in suspension cultures of *Taxus baccata*, *J. Biotechnol.* **101**, 157–163.

Pan, Z.W., Wang, H.Q., Zhong, J.J. (2000) Scale-up study on suspension cultures of *Taxus chinensis* cells for production of taxane diterpene, *Enzyme Microb. Technol.* **27**, 714–723.

Phisalaphong, M., Linden, J.C. (1999) Kinetic studies of paclitaxel production by *Taxus canadensis* cultures in batch and semicontinuous with total cell recycle, *Biotechnol. Prog.* **15**, 1072–1077.

Seki, M., Ohzora, C., Takeda, M., Furusaki, S. (1997) Taxol (paclitaxel) production using free and immobilized cells of *Taxus cuspidata*, *Biotechnol. Bioeng.* **53**, 214–219.

Son, S.H., Choi, S.M., Lee, Y.H., Choi, K.B., Yun, S.R., Kim, J.K., Park, H.J., Kwon, O.W., Noh, E.W., Seon, J.H., Park, Y.G. (2000) Large – scale growth and taxane production in cell cultures of *Taxus cuspidata* (Japanese yew) using a novel bioreactor, *Plant Cell Rep.* **19**, 628–633.

Tanaka, H., Yamashita, T., Aoyagi, H., Yamamoto, Y., Fukunaga, Y. (1996) Efficient production of chitinase by *Wasabia japonica* protoplasts immobilized In double-layered gel fibers, *J. Ferment. Bioeng.* **81**, 394–399.

Wang, Z.Y., Zhong, J.J. (2002) Repeated elicitation enhances taxane production in suspension cultures of *Taxus chinensis* in bioreactors, *Biotechnol. Lett.* **24**, 445–448.

Wu, J., Lin, L. (2003) Enhancement of taxol production and release in *Taxus chinensis* cell cultures by ultrasound, methyl jasmonate and in situ solvent extraction, *Appl. Microbiol. Biotechnol.* **62**, 151–155.

Yu, L.J., Lan, W.Z., Qin, W.M., Xu, H.B. (2002) High stable production of taxol in elicited synchronous cultures of *Taxus chinensis*, *Process Biochem.* **38**, 207–210.

Yuan, Y.J., Lin, C., Hu, Z.D., Wu, J.C., Zeng, A.P. (2002) Fungal elicitor-induced cell apoptosis in suspension cultures of *Taxus chinensis* var. *mairei* for taxol production, *Process Biochem.* **38**, 193–198.

Yukimune, Y., Tabata, H., Higashi, Y., Hara, Y. (1996) Methyl jasmonate-induced overproduction of paclitaxel and baccatin III in *Taxus* cell suspension cultures, *Nat. Biotechnol.* **14**, 1129–1132.

Zhang, C.H., Wu, J.Y. (2003) Ethylene inhibitors enhance elicitor-induced paclitaxel production in suspension cultures of *Taxus* spp. cells, *Enzyme Microb. Technol.* **32**, 71–77.

Antitumor Steroids

Guy Leclercq
Institut J. Bordet, Université Libre de Bruxelles, Brussels, Belgium

1 Background 440

2 Steroid Hormone Receptors 441
2.1 Structure and Conformation of the Hormone Binding Domains 441
2.2 Conformational Changes of Ligand-binding Domains Induced by Agonists or Antagonists 442

3 Restrictive Conditions for Efficient Targeting of Antitumor Estrogens 445
3.1 Main Structural Requirements for Estradiol Binding 446
3.1.1 A-ring 446
3.1.2 B-ring 446
3.1.3 C-ring 446
3.1.4 D-ring 447
3.2 Nucleocytoplasmic Shuttle of Estrogen Receptors 447
3.3 Ligand-induced Regulation of Estrogen Binding Capacity 448
3.4 Ligand-induced Regulation of Estrogen Receptor Level 449

4 Antitumor Estrogens 449
4.1 Organometallic Estradiol Derivatives 450
4.1.1 A-ring Complexion 450
4.1.2 7α and 11β Derivatives 451
4.1.3 17α Derivatives 451
4.2 Radiolabeled Estradiol Derivatives 453
4.2.1 ^{123}I and ^{125}I Labeled Compounds 454
4.2.2 α-Emitting Compounds 455
4.3 Electrophilic 11β Derivative 456

Bibliography 457
Books and Reviews 457
Primary Literature 457

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Keywords

Steroid Hormones
Lipophilic hormones (lipid-soluble) which derive from cholesterol and act through association with their cognate receptors (specific nuclear receptors for each class of hormones, i.e. estrogens, androgens, progestins, gluco- and mineralocorticoids).

Nuclear Receptor
Large family of ligand-dependent transcription factors (peptides able to modulate the expression of genes).

Targeting
Modality to selectively deliver a compound to a given cell type or tissue to produce a biological effect only at his level.

1 Background

Steroid hormones once secreted are transported through the blood by proteins such as albumin or sex hormone binding globulin to be delivered to specific receptors located in various tissues, including tumors. Hence, steroid hormones appear as attractive vectors that might be used to achieve a selective targeting of anti-neoplastic drugs, at least in hormone-dependent cancer, which contains high amounts of receptors (i.e. breast, prostate etc.). While association of these drugs with receptors is obviously of major importance, their binding ability to transport proteins may also play a role (i.e. albumin localize at the tumor due to enhanced vascular permeability).

Such a concept is obviously not new. High doses of steroid hormones are indeed routinely used for the treatment of various endocrine disorders. Synthetic antagonists aimed to specifically block the action of their receptors or/and metabolic enzymes have also been in clinical practice since a long time. With specific regard to neoplasia, antiestrogens, antiprogestins, as well as aromatase inhibitors (enzymes that convert androgens to estrogens) are drugs of prominent importance for the treatment of breast cancer. Therapeutic effectiveness as well as mechanism of action of such compounds has been the subject of a large number of reviews and will therefore not be addressed here. Potential utility of steroidal derivatives aimed to selectively kill tumor cells (i.e. ''antitumor steroids'') will solely be analyzed.

In fact, during the last 30 years, several investigators prepared a large number of steroids bearing cytotoxic molecules (e.g. alkylating agents, nitrosoureas, *cis*-platinum complexes, intercalating agents, etc.). Unfortunately, such syntheses led almost always to drugs devoid of therapeutic interest primarily due to the fact that their cytotoxic groups were not linked in appropriate positions onto the steroids, leading to a drastic loss of the binding affinity of the latter for their cognate receptors with concomitant decrease of cellular uptake. Moreover, for the few compounds that overpass this limiting step, specific killing of tumor cells was rarely observed in cell culture

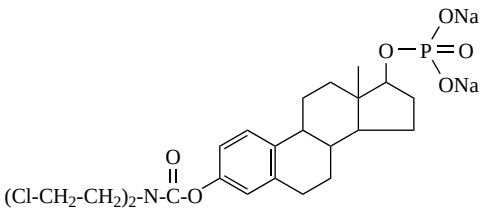

Fig. 1 Chemical formula of estramustine phosphate (Estracyt).

(receptor-positive vs receptor-negative cell lines) precluding their potential testing in clinical trials. Metabolistic instability of compounds also precludes their use. In fact, out of the hundreds of compounds synthesized, estramustine phosphate (Estracyt) (Fig. 1) is the only one having a place in the therapeutic arsenal (treatment of advanced prostatic cancer).

The growing number of experimentally determined crystal structures of ligand-binding domains of steroid hormone receptors provided a theoretical basis for the synthesis of a new generation of cytotoxic-linked steroids having a chance to associate with their cognate receptor. It is our purpose to focus on this topic. First, we will rapidly describe the structural determinants that govern the interactions of a ligand with its receptor, with special emphasis on estrogens and antiestrogens. Major advances in the synthesis of cytotoxic-linked estrogens in comparison with other classes of steroids justify this option. Next, we will move to the description of estrogen derivatives with potential antitumor activity.

2 Steroid Hormone Receptors

2.1 Structure and Conformation of the Hormone Binding Domains

Steroid hormone receptors are members of a superfamily of nuclear receptors, which regulate the expression of gene involved in cell proliferation and/or differentiation. All these receptors encode various domains with specific functions (A/B, C, D, E, and F; Fig. 2). The N-terminal A/B domain of each receptor contains a ligand-independent activation function (AF-1) required for the transcriptional activation of target genes. The adjacent C-domain contains two zinc fingers, which play an important role in DNA sequence-specific bindings (association with specific response elements). The D-domain is a hinge region, which separates ABC domains from the hormone binding domain (HBD; domain E). This binding domain contains a ligand-dependent activation function (AF-2) located in its carboxy-terminal part, a third additional AF-site located on the amino-terminal part of the HBD has been reported in the gluco and estrogen receptors (AF-2b). The C-terminal part of the receptors (end of domain E and F) is involved in the regulation of their half-life (proteolysis) and transcriptional activity.

Binding of a steroidal hormone to its receptor always provokes a conformational change in the latter that allows, ultimately, the transcription of target genes. Receptor molecules, in homo- and heterodimeric associations, exert their transcriptional effect through interaction with response elements located in the promoter region of these genes; alternatively, they may operate via their interaction with other classes of DNA-bound transcription factors (i.e. AP1,

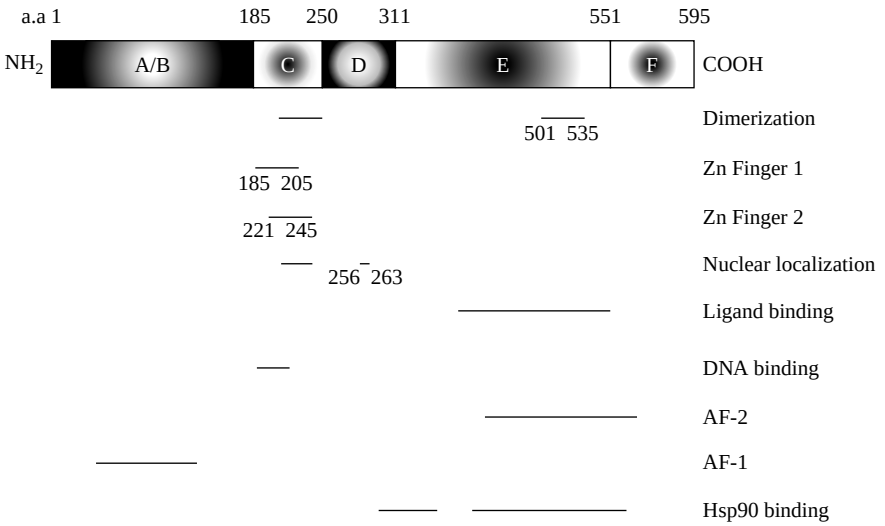

Fig. 2 ERα domains. ERα is composed of six molecular domains, which allow it to localize into the nucleus, dimerize, bind to ligand and DNA, and regulate the transcription through hormone-independent (AF-1) and hormone-dependent (AF-2) mechanisms.

Sp1 etc.). Transcription requires nuclear proteins (coactivators) that are recruited by the DNA-bound receptors. Coactivators vary among tissues eventually giving rise to tissue-specific responses to a given hormonal stimulation. Antagonists repress transcription by a reverse process (blockade of association with response elements and/or recruitment of corepressors).

AF1 expression is regulated by growth factors via classical signal transduction cascades while AF2 (and apparently AF-2b) is entirely dependent on the association of the receptor with the ligand. Under most conditions, the ligand induces a functional synergism between AF1 and AF2 due to major reorganization in the tertiary structure of the receptor. HBD contains 12α-helixes (H1–H12) folded into a 3-layered antiparallel α-helical sandwich. The central core of the HBD layer contains 3α-helixes (H5/6, H9, and H10) sandwiched between 2 additional layers of helixes composed of H1–4, H7, H8, and H11; it is flanked by H12, which contains the core region of AF2. Agonists elicit a receptor conformation of the HBD so that H12 is aligned over the hormone binding cavity; the formation of a salt bridge between H4 and H12 stabilizes the reposition of the latter. This main conformational change results in the emergence of a specific binding site (a groove) for a consensus α-helical motif (LxxLL, "NR" box) found in coactivators. Antagonists sterically interfere with H12 positioning thereby preventing recruitment of coactivators and favoring binding of corepressors.

2.2 Conformational Changes of Ligand-binding Domains Induced by Agonists or Antagonists

HBD structure of each steroidal hormone receptor has been established by amino acid labeling with reactive ligands,

site-directed mutageneses, hydrophobic cluster analyses, motif-based searches for homology with other proteins of known structures, and crystallographic studies. All HBDs are hydrophobic pockets, except at subsites corresponding to polar centers of cognate hormones; virtually, amino acids of these binding pockets surround the steroids. HBDs are somehow flexible in order to accommodate various ligands. Enlargement of a subsite results from the conformational reorganization of the whole HBD upon ligand binding: occupation of one subsite may modify the flexibility at a second subsite indicating that they respond to the ligand as a whole entity. Subsites are flexible because HBDs are composed of polypeptide segments rich in amino acid residues with mobile side chains (e.g. lysine, methionine, etc.).

Interaction of a steroidal hormone (or any synthetic ligand) with its cognate receptor depends upon several parameters (size, shape, polarity of atoms within the steroidal rings, etc.). Polar functions located at the extremities of the steroidal core (pos. 3 et 17) are of prominent importance for the onset of a biological function, as illustrated hereunder by the binding of estradiol (E_2) and the partial antiestrogen hydroxytamoxifen (OH-Tam) to the estrogen receptor (α-isoform).

Most important amino acids involved in the binding of E_2 to the ER-HBD are Glu353 and Arg394 on the one hand, His524 on the other hand: Glu353 and Arg394 interact with the C_3 phenolic group of the hormone (H-bridge, participation of an H_2O molecule) while His524 interacts with 17β oxygen (Fig. 3). As stated above, location of E_2 within the HBD leads to a displacement of H12 favoring the access of coactivators to their binding site; virtually H12 surrounds E_2 and shields the pocket where it is locked from the environment ("closed conformation").

OH-Tam (Fig. 4) is a triphenylethylenic derivative of prominent importance for the treatment of ER-positive breast cancer; aminoalkyl side chain of the compound is responsible for its antiestrogenic (and cytostatic/cytotoxic) activity. Phenolic group of the drug (pos. 4) interacts with Glu353 and Arg394, as the C_3 phenolic group of E_2.

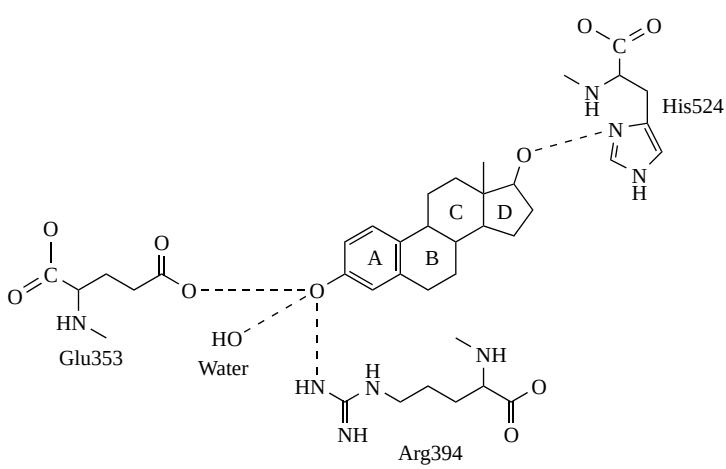

Fig. 3 Docking of E_2 in the ERα-HBD. The phenolic A-ring of E_2 is located in the HBD by Glu353 and Arg394 and tethered by His524.

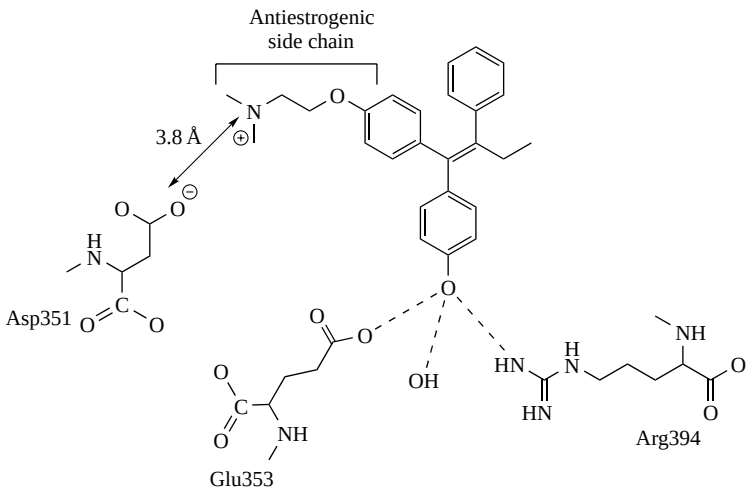

Fig. 4 Docking of OH-Tam (partial antiestrogen) in the ERα-HBD. The phenolic ring of OH-Tam is located in the HBD by Glu353 and Arg394, in analogy with the A-ring of E_2 (see Fig. 3). The nitrogen atom of the amino alkyl side chain of OH-Tam weakly interacts with Asp351. This interaction seems to be the key to the antagonistic activity of the compound.

However, in the hormone binding pocket, binding of OH-Tam to ER promotes a conformational change differing from that driven by E_2 inasmuch as the orientation of H12 appears essentially different. In the OH-Tam–ER complex, H12 is oriented in such a way that it maintains the pocket wide open, allowing the specific interaction of the nitrogen atom of the side chain of OH-Tam with a residue of the binding pocket (Asp351). In this "open conformation", which is stabilized by the side chain of the antiestrogen, H12 occupies a part of a coactivator binding site, which normally becomes accessible upon E_2 exposure. As a consequence, OH-Tam-associated H12 reorientation abrogates the recruitment of coactivators perhaps in favor of corepressors (AF-2 silencing). Hence, interaction between the side chain of OH-Tam and Asp351 seems to be the key to the antagonistic activity of the compound. Note that this open conformation of the HBD allows the expression of other transcriptional activating sites (AF-1 and to some extent AF-2b) explaining the partial estrogenicity of OH-Tam.

OH-Tam belongs to an important class of clinically relevant compounds, recognized now as *selective estrogen receptor modulators* (SERMs). These compounds act as antagonists or agonists, depending on the cellular type promoters and coregulators (ER isoform is also of importance). A second class of antagonists, (ICI 182,780, ICI 164,384 and RU 58,668; Fig. 5) are capable of completely blocking the transcriptional activity of ER via the inhibition of both AF1 and AF2 action. One may speculate that these "pure antiestrogens" could play an important role as a second-line therapy against advanced breast cancer in patients who develop resistance to tamoxifen treatment. They correspond to 7α or 11β derivatives of E_2 in which specific reactive groups are linked to the steroidal moiety

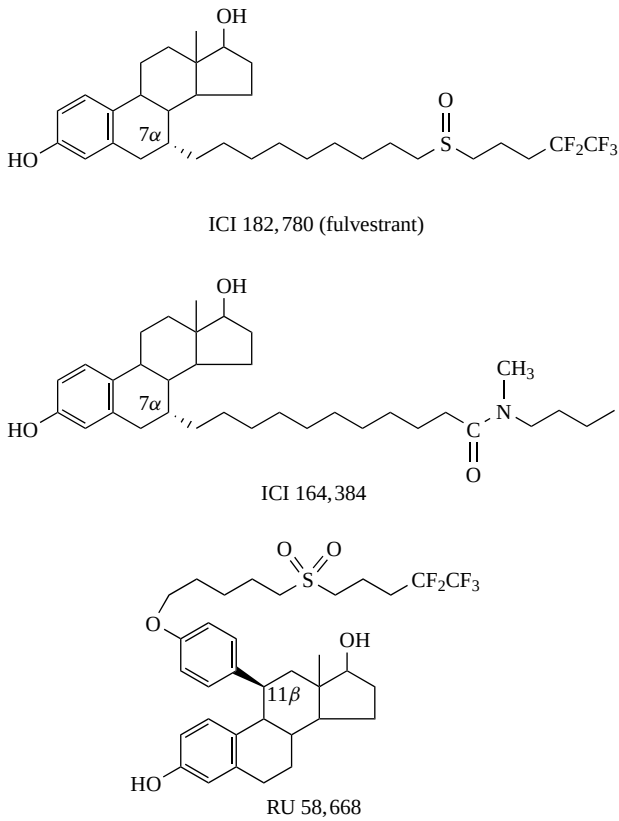

Fig. 5 Chemical formulas of pure antiestrogens.

through a long alkyl side chain intended to protrude out of the HBD when bound to the receptor (as the side chain of OH-Tam and related SERMs). The resulting conformational state of the receptor is, however, slightly different from that elicited by all investigated SERMs, although their long alkyl side chain similarly abrogate the alignment of H-12 onto the steroidal core.

3 Restrictive Conditions for Efficient Targeting of Antitumor Estrogens

To utilize steroid hormones as effective delivery vector of cytotoxic drugs to the tumor cells, several conditions should be satisfied. Concerning the association of the compounds with their receptors, we stress the need for the following four conditions:

1. It is important to link the cytotoxic agent in such a way as to not interfere with the recognition of the steroidal vector.
2. The cytotoxic moiety should be shipped to its target. Since receptors focalize their action at the genomic level, drugs to be linked to the steroid are restricted to those producing damage at the DNA level or associated nuclear proteins.

3. The compounds should not produce a rapid loss of binding ability of their steroidal vector.
4. The concentration of receptors should be sufficiently high to produce an effective cellular uptake of the cytotoxic moiety. This concentration should be maintained until this cytotoxic moiety has reached its target, potential degradation of the receptor (downregulation) induced by the steroidal vector should be limited.

The following paragraphs analyze how such conditions could be satisfied for (anti)estrogenic conjugates.

3.1 Main Structural Requirements for Estradiol Binding

The main and unique characteristic of E_2 is its lack of binding affinity for steroid hormone receptors other than its own receptor. The phenolic group in C_3, not recorded in the other steroid hormone, generates this big difference. Binding data of E_2 and derivatives led to the following conclusions.

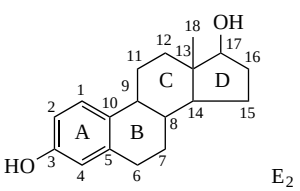

3.1.1 A-ring

C_1 is intolerant of both hydrophobic and polar substituents, suggesting close contacts with the HBD. The hydroxy group in C_3 donates a proton to Glu353 and Arg394 (see above), and this bond is donated *syn* to the C_2–C_3 bond. C_4 may be in close proximity to a cysteine residue, based on inactivation assays of the receptor by 4-mercuri-E_2. The A-ring electron cloud may be engaged in a weak polar interaction with a slightly positively charged receptor residue, situated on the β-face of the steroid.

3.1.2 B-ring

C_6 seems to be intolerant for both polar and nonpolar substituents, once again suggesting close interaction of this area to the receptor essential spatial volume. Large aliphatic chains at $C_{7\alpha}$ are well tolerated (i.e. pure antiestrogens ICI 164,384 and ICI 182,780; Fig. 5), they localize within a subsite of the HBD accepting 11β nonpolar substituents (rotation of the E_2 core around a virtual symmetry axis that runs roughly through C_3 and C_{17}). Positions in $C_{7\beta}$ and $C_{8\alpha}$ are unexplored and $C_{9\alpha}$ has low steric tolerance.

3.1.3 C-ring

A preformed hydrophobic subsite of the HBD has a high steric tolerance for 11β residues. It can accommodate aromatic substituents bearing aliphatic side chains (pure antiestrogen RU 58,668; Fig. 5) while still retaining high affinity; it is intolerant of polar groups. The covalent binding of tamoxifen aziridine has suggested an accessible receptor nucleophile in $C_{11\beta}$ vicinity; such a hypothesis was rejected since 11β-methyl aziridine estradiol does not covalently label ER. Moreover, 11β-chloromethyl and bromoacetyl analogs also fail to covalently label the receptor. The 11β-chloromethyl substituent of E_2 produces, however, a *quasi*-irreversible binding of the hormone with the receptor. Interaction of the chloride atom with a putative

Zn^{++} ion located within the receptor has been proposed to explain this extremely high binding affinity; hydrophobic interaction with a tryptophane or phenylalanine has been advocated as an alternative hypothesis. The 12β-site accommodates small nonpolar groups as methyl; otherwise, C_{12} has been scarcely investigated.

3.1.4 D-ring

C_{14} can accept small, nonpolar groups at the α- or β-sites without appreciable loss of binding affinity. C_{15} is intolerant of small hydrophobic substituents (methyl) or replacement with oxygen. $C_{16\alpha}$ accommodates large nonpolar groups (e.g. iodine), but $C_{16\beta}$ is sterically less accessible, and even fluorine substitution decreases the binding affinity substantially. ER accepts small nonpolar groups in $C_{17\alpha}$ or groups in which the extended bulk should be directed in an endo manner, for example, *Z*−CH=CHI binds more effectively than the corresponding *E*-isomer. Spatial tolerances of this area are precise, because even an ethyl group at this position results in significant loss of binding affinity; this effect may be secondary to alteration of rotation of the 17β-hydroxyl group. The 17β-hydroxyl group acts as a H-bond acceptor from His524. There may be a lysine and two cysteine residues (one positioned on each face of the steroid) in the vicinity of C_{15}-C_{16}-C_{17}, based on the reactivity of E_2 derivatives bearing covalently labeling groups. The 18-methyl group is not a significant contributor to binding, indicating no productive contact with the receptor at this site.

Data summarized here define localizations for grafting of a cytotoxic compound onto estradiol.

1. Positions 7α and 11β appear especially appropriate for large substituents without provoking a dramatic loss of binding affinity. In fact, this has been proved by linkage of a porphyrin in position 11β through a side chain. Linkage on 7α or 11β would localize the substituent within the same region of the receptor: the steroid may indeed rotate around a virtual symmetry axis between C_3 and C_{17}. Hence, 7α and 11β seem quite equivalent for grafting a side chain bearing a cytotoxic compound. Flexibility of such chains does not appear of prime importance as revealed by the study of mobile $[(CH_2)_n]$ or inflexible $[(CF_2)_n]$ spacers.
2. Substitution in 17α may also be recommended, provided that a short spacer is introduced between E_2 and the cytotoxic compound to maintain an effective binding affinity. Ethynyl and vinyl groups are especially accurate since they give an effective separation without excessive conformational flexibility. Directing these substituents away from the β-face of the steroid (side of the 17 OH) is required to limit steric interference with the binding pocket.

3.2 Nucleocytoplasmic Shuttle of Estrogen Receptors

Nuclear receptors are "translocating binding proteins," which continuously move between the cytoplasm and the cell nucleus (shuttling mechanism). The nuclear localization, however, always dominates especially under agonist stimulation (a dynamic event resulting from the continuous active transport of the receptor into the nucleus, slightly counterbalanced by its diffusion into the cytoplasm). Receptor detection in the cytosol of extracts

from unstimulated cells may derive from its nuclear release at the time of homogenization.

Conveyance of ER into the cell nucleus is governed by specific amino acid sequences and nuclear localization signals (NLSs) situated between its DNA and hormone binding domains (HBDs) (hinge region). The existence of nuclear export signals (NESs) has also been reported; both NESs and NLSs play an important role in nuclear export of the steroid receptors. Nuclear uptake of receptors requires ATP, while nuclear export apparently occurs in the absence of energy production.

Investigations performed on ER-transfected cells revealed an additional subnuclear movement of the receptor under ligand stimulations. E_2 changes the receptor pattern from a diffuse nucleoplasmic aspect to a hyper-speckled distribution, suggesting that the ligand-dependent organization of the receptors within the nuclear matrix involves more complex events than a simple recognition of DNA-binding sites and coregulators. Interestingly, such a hyper-speckled distribution also occurs with the partial antiestrogen OH-Tam, while a substantial effect is solely detected with the pure antagonist ICI 182,780. In this context, it should be stressed that several other investigations localize the receptor within the cytoplasm under exposure to pure antiestrogens. Hence, the cellular localization of the receptor under pure antiestrogens is still not established, although the tendency for a major distribution within the cytoplasm dominates (when downregulation does not occur, see Sect. 3.4).

Concepts outlined here suggest that linkage of cytotoxic residues aimed at altering DNA and associated proteins should be restricted to estrogens (and SERM), pure antiestrogen having only a chance to strongly associate to the nuclear matrix.

3.3 Ligand-induced Regulation of Estrogen Binding Capacity

Shuttling of ER between the cytoplasm and the nucleus influences its ligand-binding ability. The nuclear localization of a receptor is associated with a loss of ligand-binding ability, which is restored by its return into the cytoplasm (dephosphorylation/phosphorylation cycle). After several cycles, ER is degraded by the proteasomal pathway, most probably into the cytoplasm (half-life 3 ~ 4 h); ubiquitination of the receptor favors its shipment to the proteasome. Inhibitors of protein synthesis (cycloheximide, puromycin) interfere with this mechanism leading to a stabilization of the receptor within the cell nucleus in a form devoid of strong ligand-binding ability.

As stated above (Sect. 2.2), E_2 is stably incorporated within the HBD, decreasing the chance of exchange with unbound ligands (closing of the binding pocket). Such a property fixes the majority of the receptor molecules into a conformation appropriate for a specific anchorage within the cell nucleus (hyper-specked pattern) to accomplish functions imposed by the hormone. A progressive decrease of estrogen binding capacity subsequently occurs. As expected, this decrease is closely related to the binding affinity of the estrogen for the receptor, which is a major determinant of its nuclear stabilization. One may anticipate that this receptor form, unable to bind the hormone is subjected to a return into the cytoplasm to be recycled or degraded.

Partial antiestrogens (OH-Tam and related SERMs), which maintain the HBD in an open conformation, accumulate the receptor within the nucleus in which it progressively loses its ability to bind estradiol. Blockade of the proteasomal degradation of ER by these antagonists associated with maintenance of its synthesis (see Sect. 3.4), explains this accumulation process.

3.4 Ligand-induced Regulation of Estrogen Receptor Level

Ligands strongly influence the turnover rate of ER, by regulating its production and ubiquitin-dependent proteolysis. Estrogens block ER synthesis (decrease of ER mRNA level) while antiestrogens do not. Maintenance of ER synthesis under antiestrogen treatment may therefore upregulate the receptor when its proteasomal degradation is weak, as observed with OH-Tam and related SERMs. Estrogens and pure antiestrogens induce a strong and rapid ubiquitination of the receptor, leading to its rapid proteolysis; SERMs are largely less effective in this regard. As a consequence of these regulatory processes, estrogens, and pure antiestrogens downregulate the receptor while SERMs may lead to its accumulation.

The ability of cycloheximide and puromycin to stabilize the activated receptor within the cell nucleus suggests the induction of a protein favoring its return into the cytoplasm where it would be degraded. This concept of cytoplasmic degradation is supported by the observation that cycloheximide and puromycin fail to block the ER elimination when it is provoked by pure antiestrogens, which fail to stabilize the receptor within the nucleus.

In fact, only 10% of ER molecules should be saturated by a ligand to produce the proteasomal degradation of the whole population. Moreover, receptors of which the HBD has been blocked by a covalent binding (i.e. tamoxifen aziridine) are similarly eliminated. The mechanism underlying this amplification process is not known. Secretion of unidentified hydrophobic compounds by early-stimulated cells may play a role. Such compounds, the existence of which has been detected in conditioned media from ligand-treated cells, may act by themselves or in synergy with undetectable concentration of ligands.

4 Antitumor Estrogens

ER content is extremely variable among breast cancer. Median concentrations vary among studies reaching a maximum of 100 fmoles mg^{-1} proteins, which correspond roughly to 10 000 receptors per cell. Only a limited number of cytotoxic drugs would be able to exert significant antitumor effect with such a low concentration of vectors. The high reactivity of alkylating agents (1000 DNA interstrand cross-links is lethal) led to their selection for the synthesis of antitumor estrogens. Numerous compounds were synthesized in the 1970s and 1980s. However, as stated above, most of them failed to display a sufficient binding affinity for ER, or any unambiguous specificity of action, leading to a progressive disinterest to this approach. Other cytotoxic agents such as vinca alkaloids and intercalating agents (like ellipticin and daumomycin) also failed to show any potential therapeutic interest. Note that

this remark also holds for nonsteroidal cytotoxic derivatives (i.e. diphenyl and triphenylethylene (anti)estrogens).

In the following years, accumulation of data concerning the receptor's structure, binding requirement, and mode of action (previous sections) dictated new guidelines for syntheses. Research axes of the last decade are overviewed in the following paragraphs.

4.1 Organometallic Estradiol Derivatives

The advance in organometallic chemistry associated with the discovery of the antitumor activity of cisplatin led to the analysis of the potential use of metal-conjugated estrogens. Since cisplatin is unsuitable for breast cancer treatment, the search concentrates mainly on non-platinum metal complexes such as cyclopentadienyl cycles of Ru, Ti, Fe, Co, and so on, which appear attractive in cancer chemotherapy. Nonetheless, cisplatin was also linked with the hope of extending its therapeutic use. While DNA is the expected target of such drugs, some of them were found to irreversibly associate with the receptor, a behavior that may produce unusual biological properties.

Unfortunately, endocrine and therapeutic activity of most of the metal-conjugated estrogens has not yet been explored. Investigations mainly concern their binding to the receptor. The effect of ER-positive and ER-negative cell lines has been assessed in only a few cases.

4.1.1 A-ring Complexion

An important feature displayed by organometallic clusters is the possibility of their complexion on either the α- or β-side of the steroid A-ring (Fig. 6). Studies conducted with chromium and ruthenium complexes reveal that a π-organometallic adduct on the β-face prevents the binding of E_2 to the receptor while its addition on the α-face maintains some binding affinity. Of these compounds, the neutral chromium complex displayed the highest binding affinity (~30% of E_2). Although the cationic species Cp^*Ru^+ failed to seriously alter the binding of the hormone, as demonstrated

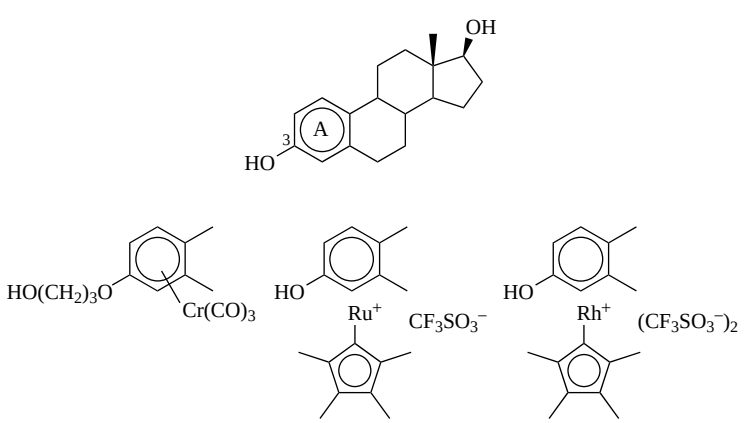

Fig. 6 Organometallic E_2 derivatives – chemical formulas of A-ring substituted derivatives described in Sect. 4.1.1. Binding affinity for ER of the organometallic abduct is higher on the α face of A-ring those on its β face.

by α-[Cp*Ru(17β-estradiol)][CF_3SO_3], an increase in the positive charge of the organic moeity (i.e. Rh^{++}) appears detrimental. This effect may be attributed to an increase in acidity; the complexion of the phenolic A-ring increases its acidity. In a basic medium, the complexed phenolic rings are transformed into the corresponding dienonylic rings with the concomitant loss of 3-OH function required for strong association with the receptor.

4.1.2 7α and 11β Derivatives

Grafting of organometallic clusters in these strategic regions is expected to be relatively well tolerated (Sect. 3.1). Hence, various compounds have been synthesized.

Nonradioactive Re systems are analogous to corresponding nonradioactive systems containing ^{99m}Tc, ^{186}Re or ^{188}Re. For this reason, they may be viewed as prototypes for the production of drugs to be used in imaging and radiotherapy. Attempts to produce 7α derivatives in which the rhenium moeity is remote from E_2 by a long side chain led to the identification of a few compounds with high binding affinity (~15% of E_2; Fig. 7). The polarity of these complexes was modified to improve biodistribution without loss of binding affinity by introducing (poly)ether linkages into the 7α side chains. Attachment of a rhenium complex in position 11β (Fig. 7) was also relatively well tolerated (binding affinity ~1/10 of E_2). Tested on a reported gene, this compound displayed an agonist activity almost as strong as E_2.

Linkage of a cobalt cluster in 11β (i.e. 11β-[(ethynyl) $Co_2(CO)_6$]estradiol; Fig. 7) strongly increases the stability of the association of the parent-free ligand with the receptor. This behavior is reminiscent of the high performance of 11β-chloromethylestradiol in ER binding (almost irreversible binding in buffered solution). Tested *in vitro* in MVLN cells (MCF-7 cells stably transfected with an estrogen responsive luciferase gene), this cobalt conjugate displays a strong estrogenicity (as effective as E_2); no cytotoxicity was recorded.

4.1.3 17α Derivatives

Various organometallic moeities (rhenium, cobalt, ruthenium, tungsten, etc.) fail to strongly affect the binding affinity of E_2 for ER when linked in its 17α position at the end of a rigid ethynyl spacer (Fig. 8). In contrast, the affinity largely decreases when the spacer is shortened to a simple sp^3 carbon atom confirming that the ER-HBD possesses a subsite able to accommodate rigid narrow substituents. Affinity also decreases when the organometallic moeity is moved toward the D-ring of the steroid, even on a rigid spacer. Moreover, neutral species (ferrocenyl) are relatively well tolerated while cationic species (ruthenium) decrease the binding affinity indicating that the HBD does not accept a positive change in the 17α position of E_2. Hence, bulky organometallic moiety may be tolerated if they localize outside a zone of steric/electronic constraint.

Coordination of selected transition-metal moeities to 17α-alkynylestradiol may induce an irreversible (covalent) binding with the receptor (Fig. 8). Nucleophilic sulfur residues within or in the close vicinity of the HBD are good candidates for the establishment of such a covalent bond. The key step in this chemical reaction is the loss of the 17β-OH functionality for the generation of a carbenium ion; this electron-deficient center stabilized by the organometallic fragment

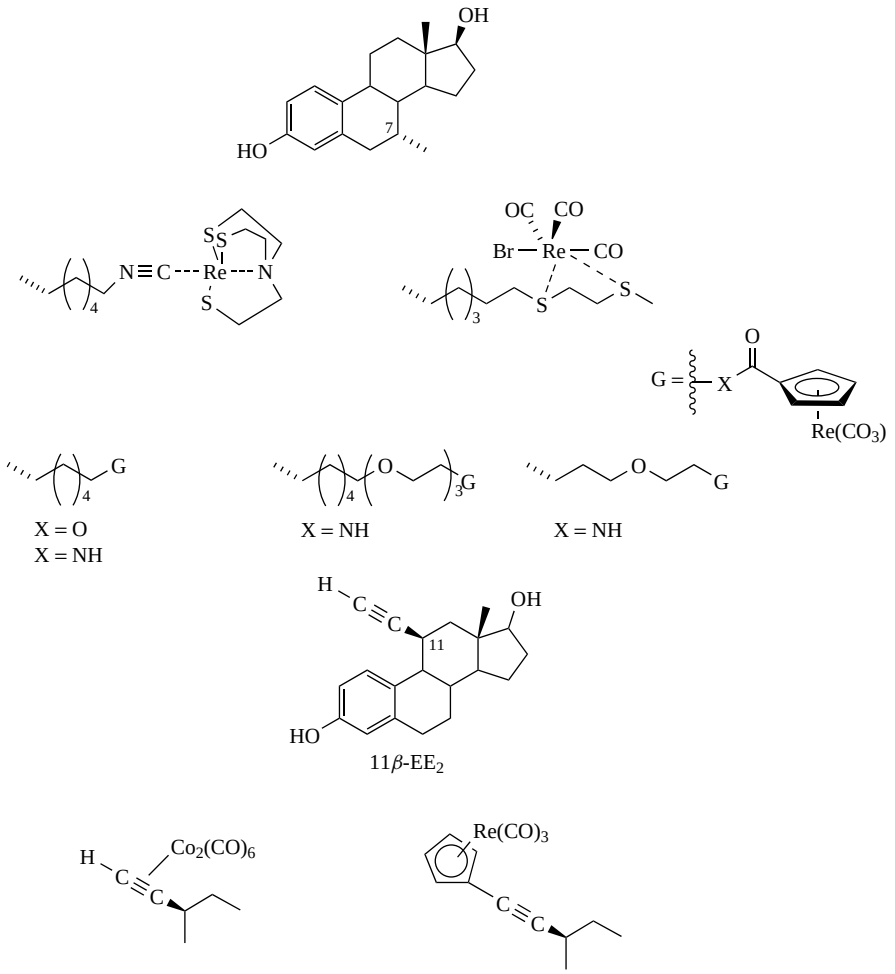

Fig. 7 Organometallic E_2 derivatives – chemical formulas of 7α and 11β derivatives described in Sect. 4.1.2.

promotes an interaction with a vicinal nucleophilic residue. Among E_2 transition-metal complexes examined, dicobalt clusters [$Co_2(CO)_6$] were found to be the most active affinity markers. Of note, FeCo-$(CO)_6$(17α-ethynyl-17β-dehydroxyestradiol) in which the heterobimetallic fragment is sterically similar to the $Co_2(CO)_6$ homolog, prevents the formation of a carbenium-ion-like species confirming the requirement at the 17-position of an OH group and an organometallic unit capable of dissipating the positive charge of the transient carbocation.

Potential influence of covalent binding of these dicobalt clusters on the ER level and transcriptional activity is unknown, and therefore, to be investigated. Potential therapeutic activity should also be assessed. Note in this regard that alkyne $Co_2(CO)_6$ itself shows no cytotoxic effect.

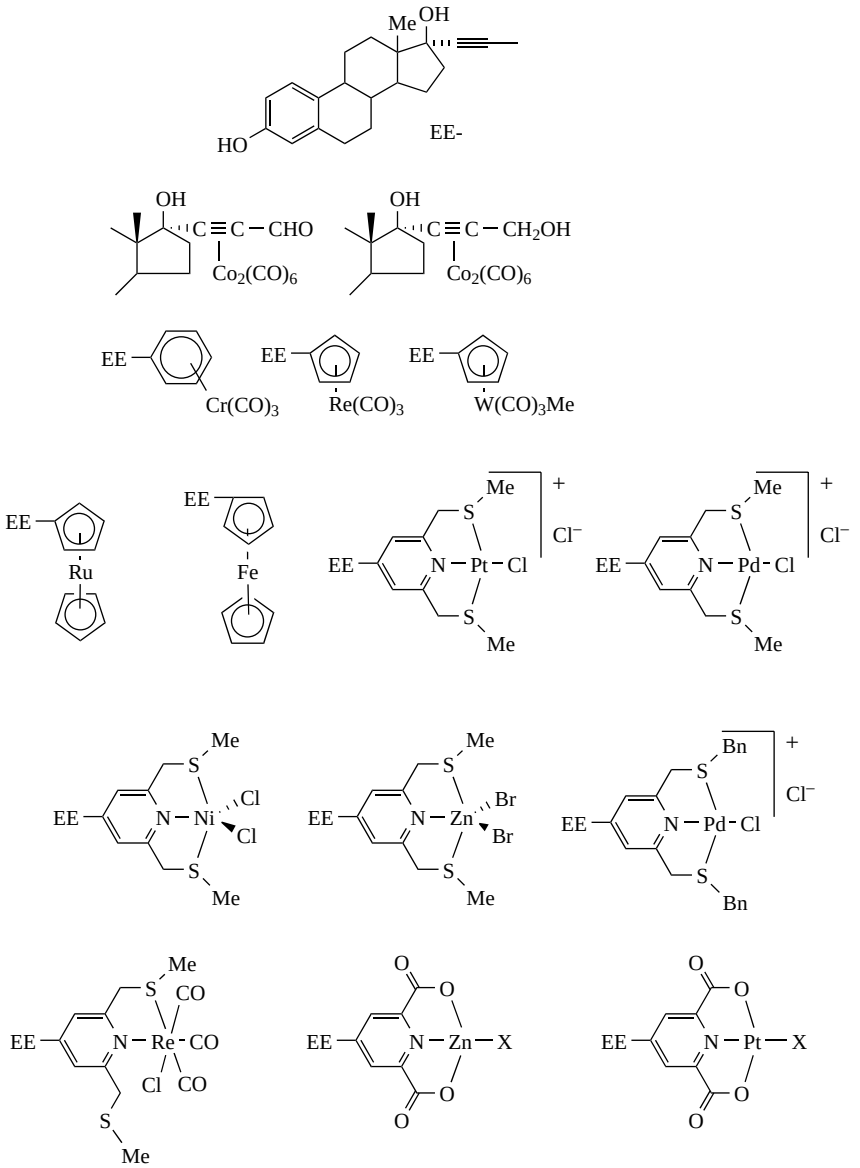

Fig. 8 Organometallic E_2 derivatives – chemical formulas of 17α ethynyl E_2 (EE) derivatives described in Sect. 4.1.3.

4.2 Radiolabeled Estradiol Derivatives

The presence of ER-positive and -negative cells in all tumors led to the concept of using receptor-directed radiochemicals of which the range of radiation is larger than one cell diameter. Compounds of this type would be of great interest since their action would not restrict solely to

ER-positive cells (as for antiestrogens) but also to ER-negative cells located in their neighborhood.

4.2.1 ^{123}I and ^{125}I Labeled Compounds

Several [^{125}I] labeled E_2 derivatives have been synthesized for *in vivo* imaging of ER-positive tumors. Utility of this approach in cancer diagnostic has, however, never been demonstrated. Nevertheless, biodistribution as well as imaging investigations performed on both animals and humans suggest some uptake selectivity, especially for 17α-iodovinyl derivatives. Stereochemistry of such compounds appears of prime importance for ER binding; *Z*-isomers bind more effectively than corresponding *E*-isomers. These data confirm the concept that the HBD possesses a hydrophobic subsite able to accept lipophilic moieties linked in position 17α of E_2 (see Sect. 3.1.4). In an effort to utilize this concept, a series of 17α-substituted E_2 derivatives were synthesized, four of them bearing iodine-123 (Fig. 9). The binding of these four compounds depends on the nature of their substituents with an alkene linkage being greatly preferred over alkynyl and alkyl. Preference for the *Z*-alkene geometry confirms the conclusion from studies on parent iodovinyls. Additional biodistribution studies performed with the *Z*-alkene suggested some uptake selectivity (uterus/ovary) and metabolic instability (accumulation of free iodine within the thyroid).

16α-[^{125}I]iodo-estradiol has been shown to produce a receptor-mediated cytotoxicity on MCF-7 breast cancer cells. The decrease of survival rate provoked

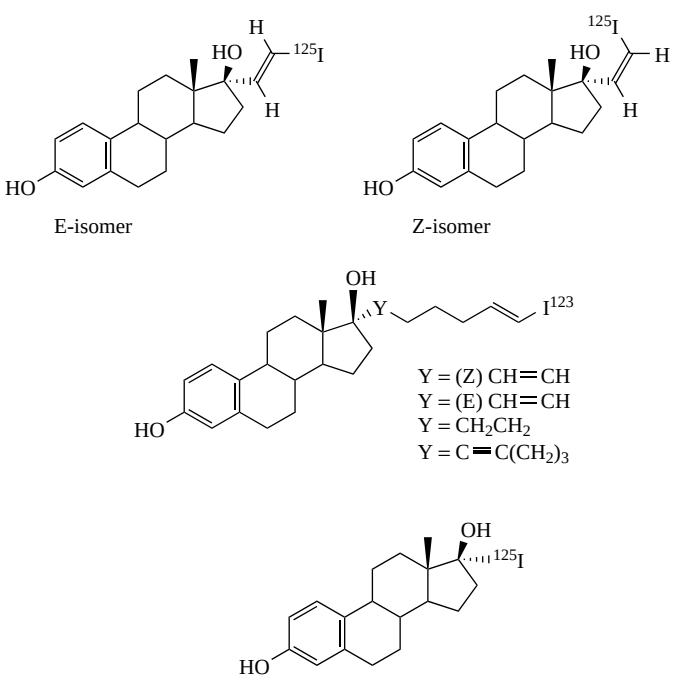

Fig. 9 Radiolabeled E_2 derivatives. Chemical formulas of iodinated (^{123}I and ^{125}I) derivatives described in Sect. 4.2.1.

by this radioligand was associated with chromosomal aberrations; of note, an excess of diethylstilbestrol, used as potential competitor, prevented the decrease of survival rate as well as chromosomal aberrations. Despite the promising aspect of these data, estimated residence times of ^{125}I-labeled estrogens in ER-positive cells (ER half-life: less than 3 h) appear be too low for an efficient cell killing. In this regard, the Auger electron-emitting nucleotides such as iodine-123 that release their energy within a short period of time would be more powerful to selectively kill ER-positive cells (i.e. half-life of ^{123}I = 4.4 h vs ^{125}I = 60 days). Testing of 17α-[^{123}I]iodovinyl-11β methoxy-estradiol on Chinese hamster ovary cells expressing or not high level of ER provided credence to this concept.

4.2.2 α-Emitting Compounds

Auger electron emitter agents would only affect ER-positive cells due to their weak distance of irradiation potency. α-particle emitting radionucleotides such as astatine-211 are attractive candidates to palliate this gap and affect neighboring ER-negative cells (range of radiation in tissue is 10 to 100 µm); astatine-211 can substitute iodine in many syntheses making the production of ^{211}At-labeled E_2 easily feasible. In fact, [^{211}At]astatinovinylestradiol (as well as its 11β methoxylated derivative) has been established. One may hope that such compounds produce a cytotoxic effect as their ^{125}I-labeled analogs.

The possibility to generate α-emitting isotopes by boron capture therapy is another approach under exploration. Boron-10, when irradiated with low energy thermal neutrons, yields α-particles and ^{7}Li-nuclei. The success of this therapy is dependant on the quantity of boron-10 and thermal neutrons to deliver to the cancer cells to sustain an α-lethal radiation. Carboranes (dicarba-closo-dodecaboranes) are a class of carbon-containing polyhedral-boron clusters having remarkable thermal stability and exceptional hydrophobic character; they have been utilized to incorporate a large number of boron atoms into tumor cells. Remarkably, hydrophobic interactions between carboranes and ER may occur when linked to a phenol. This finding led to the design of several compounds sharing high binding affinity for the receptor as well as estrogenic and/or antiestrogenic activity on an ER-dependent reporter gene (i.e. analogs of E_2, OH-Tam or RU 39,411; Fig. 10). One may speculate that such drugs may play a role in cancer therapy since they combine suitable endocrine and radiotoxic properties.

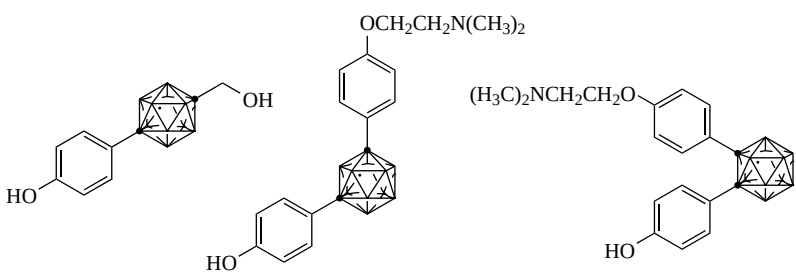

Fig. 10 Chemical formulas of phenolic carboranes with structural and ER binding properties to (a) E_2, (b) OH-Tam, and (c) RU 39,411 (Sect. 4.2.2).

4.3
Electrophilic 11β Derivative

Estrogenic and antiestrogenic activity of a few 11β-ethyl, butyl, and decyl derivatives of E_2 bearing various electrophilic functionalities (i.e. bromide, bromoacetamido, (methylsulfonyl)oxy, and (*p*-tolylsulfonyl)oxy) were determined with MVLN cells (ER-dependent luciferase expression). Ethyl derivatives were mainly estrogenic, whereas butyl and decyl essentially antiestrogenic (most compounds share a significant binding affinity for ER).

Long-term treatment of MVLN cells with OH-Tam progressively abolishes the ability of cells to express luciferase in response to estrogens. This irreversible effect seems to result from an epigenetic mechanism, such as methylation or chromatin remodeling rather than gene mutation. It progressively affects the whole-cell population and occurs by a first-order process, with

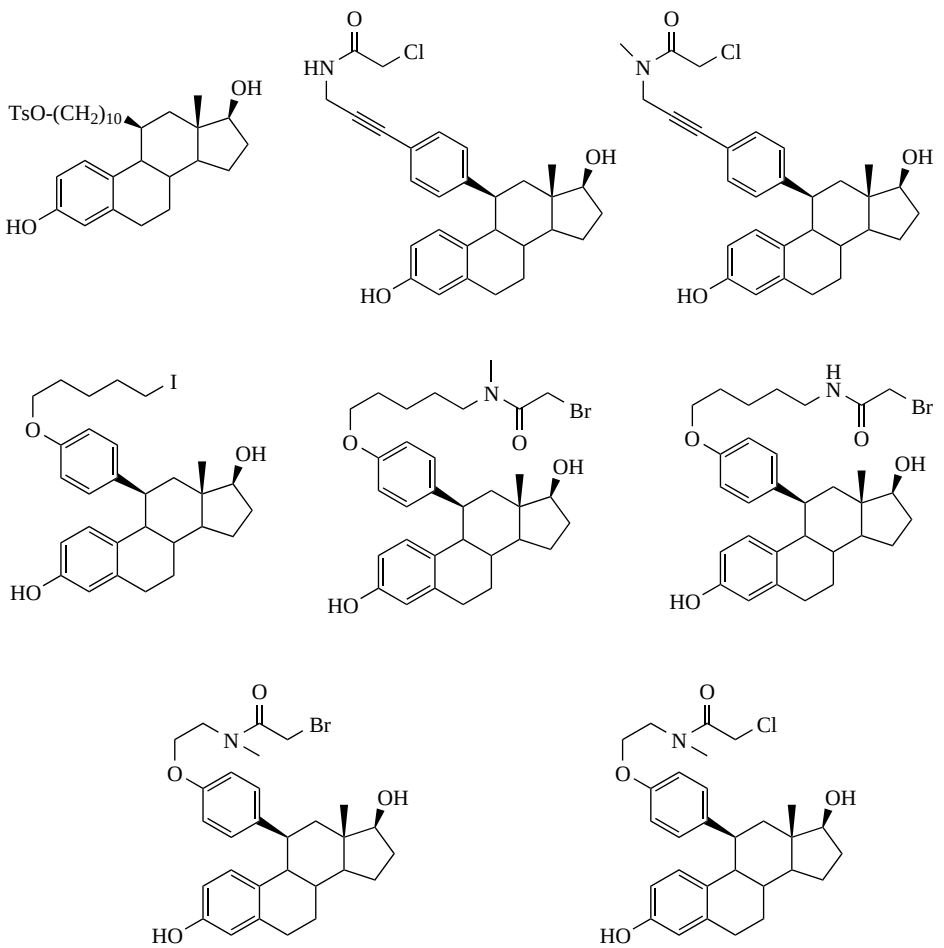

Fig. 11 Electrophilic derivatives of E_2 described in Sect. 4.3.

a half-inactivation time of ~10 days. 11β electrophile conjugates described here, similarly suppresses the expression of the luciferase gene; antiestrogenic electrophiles were more potent than the estrogenic ones but less efficient than OH-Tam. Of note, these antiestrogenic electrophiles decrease cell proliferation with potency less than OH-Tam.

These effects could not be ascribed to an irreversible binding of the compounds to ER since 11β-[(tosyloxy)decyl]E_2 (Fig. 11) solely shared such a property. A cysteine residue does not seem to be involved in the covalent attachment of this compound since it was not prevented by a thiol-specific alkylating reagent (MMTS). In fact, length and mobility of the *N*-decyl side chain precluded the identification of the anchorage site. To overcome this drawback, 11β aryl analogs aimed to restrict the localization of the electrophilic carbon to the β-side of the steroid were synthesized (Fig. 11). Both cysteinyl (probably C381 and/or C530) and non-cysteinyl residues located on the β-side and remote from C11 of the steroid (distance > "seven bonds") appear to be potential electrophile covalent attachment sites for such E_2 derivatives.

As for most other estrogens bearing reactive groups, no data were reported upon their potential use in antitumor therapy.

See also Bioorganic Chemistry; Medicinal Chemistry.

Bibliography

Books and Reviews

Blickenstaff, R.T. (1992) *Antitumor Steroids*, Academic Press, New York.

von Angerer, E. (1995) The estrogen receptor as a target for rational drug design, *Molecular Biology Intelligence Unit*, Springer-Verlag, Heidelberg, Germany.

Primary Literature

Aliau, S., Delettre, G., Mattras, H., El Garrouj, D., Nique, F., Teutsch, G., Borgna, J.L. (2000) *J. Med. Chem.* **43**, 613–628.

Aliau, S., Mattras, H., Richard, E., Bonnafous, J.-C., Borgna, J.-L. (2002) *Biochemistry* **41**, 7979–7988.

Anstead, G.M., Carlson, K.E., Katzenellenbogen, J.A. (1997) *Steroids* **62**, 268–303.

Brzozowski, A.M., Pike, A.C.W., Dauter, Z., Hubbard, R.W., Bonn, T., Angström, O., Öhman, L., Greene, G.L., Gustafsson, J.Ä., Carlquist, M. (1997) *Nature* **389**, 753–758.

Cummins, C.H. (1993) *Steroids* **58**, 245–259.

DeFranco, D.B. (2002) *Mol. Endocrinol.* **16**, 1449–1455.

DeSombre, E.R., Shafii, B., Hanson, R.N., Kuivanen, P.C., Hughes, A. (1992) *Cancer Res.* **52**, 5752–5758.

Endo, Y., Iijima, T., Yamakoshi, Y., Yamaguchi, M., Fukasawa, H., Shudo, K. (1999) *J. Med. Chem.* **42**, 1501–1504.

Endo, Y., Yoshimi, T., Iijima, T., Yamakoshi, Y. (1999) *Bioorg. Med. Chem. Lett.* **9**, 3387–3392.

El Hamouri, H., Vessières, A., Vichard, D., Top, S., Gruselle, M., Jaouen, G. (1992) *J. Med. Chem.* **35**, 3130–3135.

El Khissiin, A., Journé, F., Laïos, I., Seo, H.-S., Leclercq, G. (2000) *Steroids* **65**, 903–913.

Hart, L.L., Davie, J.R. (2002) *Biochem. Cell Biol.* **80**, 335–341.

Jaouen, G., Vessières, A. (1993) *Acc. Chem. Res.* **26**, 361–369.

Jackson, A., Davis, J., Pither, R.J., Rodger, A., Hannon, M.J. (2001) *Inorg. Chem.* **40**, 3964–3973.

Kabalka, G.W., Shoup, T.M., Daniel, G.B., Goodman, M.M. (2000) *Nucl. Med. Biol.* **27**, 279–287.

Lobaccaro, C., Pons, J.-F., Duchesne, M.-J., Auzou, G., Pons, M., Nique, F., Teutsch, G., Borgna, J.-L. (1997) *J. Med. Chem.* **40**, 2217–2227.

Morel, P., Top, S., Vessières, A., Stéphan, E., Laïos, I., Leclercq, G., Jaouen, G. (2001) *C. R. Acad. Sci., Paris, Chem.* **4**, 201–205.

Osella, D. (2002) *Second Workshop on Pharmaco-Bio-Metallics*; Communication in Sienna, Italy.
Osella, D., Cavigiolio, G., Vincenti, M., Vessières, A., Laïos, I., Leclercq, G., Napolitano, E., Fiaschi, R., Jaouen, G. (2000) *J. Organomet. Chem.* **596**, 242–247.
Reid, G., Denger, S., Kos, M., Gannon, F. (2002) *Cell. Mol. Life Sci.* **59**, 1–11.
Shiau, A.K., Barstad, D., Loria, P.M., Chang, L., Kushner, P.J., Agard, D.A., Greene, G.L. (1998) *Cell* **95**, 927–937.
Skaddan, M.B., Wüst, F.R., Katzenellenbogen, J.A. (1999) *J. Org. Chem.* **64**, 8108–8121.
Top, S., El Hafa, H., Vessières, A., Huché, M., Vaissermann, J., Jaouen, G. (2002) *Chem. Eur. J.* **8**, 5241–5249.
Tsai, M.J., O'Malley, B.W. (1994) *Annu. Rev. Biochem.* **63**, 451–486.

Autoantibodies and Autoimmunity

Kenneth Michael Pollard
The Scripps Research Institute, La Jolla, CA, USA

1 **Autoantibodies and Autoimmunity 460**

2 **Autoantibodies as Diagnostic Markers 464**

3 **Autoantibodies as Molecular and Cellular Probes 466**

4 **Autoantibodies in Experimental Models of Autoimmunity 472**

5 **Perspectives 478**

Bibliography 478
Books and Reviews 478
Primary Literature 479

Keywords

Antibody
A protein product of B cells that combines with a specific molecular target called an antigen.

Antigen
A substance that interacts with an antibody.

Epitope
The region of an antigen that is directly recognized by a specific antibody.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Immunogen
A substance that elicits an antibody response.

Indirect Immunofluorescence (IIF)
A technique whereby an antibody is overlaid onto an antigen containing cellular substrate and the antigen–antibody formed complex is detected by a fluorescently labeled anti-antibody.

Autoimmunity is an immunological reaction against constituents of the organism that are normally tolerated by the immune system of that organism. Autoimmune reactions can be either cell- or antibody-mediated. Autoantibodies are therefore antibodies that recognize normally tolerated cell and tissue constituents (or autoantigens). The antigenic specificity of an autoantibody can be a useful aid in clinical diagnosis. Autoantibodies are either cell (or tissue) specific, as found in organ-specific autoimmune diseases such as autoimmune thyroiditis, or non–organ-specific and reactive with ubiquitous intracellular antigens, as found in multisystem autoimmune diseases such as the systemic rheumatic diseases. The latter group includes autoantibodies that recognize components of macromolecular complexes of nucleic acids and/or proteins such as small nuclear ribonucleoprotein (snRNP) particles, nucleosomal and subnucleosomal structures, and tRNA synthetases, which are intrinsic components of all cell types present in an organism. Autoantibodies also recognize components of subcellular structures, including mitochondria, ribosomes, Golgi apparatus, nuclear membrane, and substructures within the nucleus and nucleolus. The ability of autoantibodies to recognize components of the cellular machinery of replication, transcription, RNA processing, RNA translation, and protein processing has made them important reagents for isolating cDNA clones that code for proteins involved in these cellular processes and for probing the relationship between molecular and cellular structure and function. The evolutionarily conserved nature of many autoantigens allows the use of autoantibodies to identify their target antigens in diverse species, ranging in some cases from humans to lower eukaryotes such as yeast. Autoantibodies have been used to inhibit the biological function of autoantigens and/or to recognize autoantigens in a defined functional state.

1
Autoantibodies and Autoimmunity

An autoimmune response is an attack by the immune system on the host itself. In healthy individuals, the immune system is "tolerant" of its host ("self") but attacks foreign ("nonself") constituents such as bacteria and viruses. The ability to distinguish self from nonself is considered to be the determining factor in whether the immune system responds to a suspected challenge. Although it may appear obvious, there is actually considerable debate over

what constitutes "self" and "nonself," and what cellular/molecular mechanisms are involved. Possible discriminators between "self" and "nonself" include recognition of infection or identification of danger signals. The outcome of the debate on self/nonself discrimination notwithstanding, autoimmunity represents an obvious disruption of the mechanism by which the immune system regulates its activities. Importantly, the responsible effector mechanisms appear to be no different from those used to combat exogenous infective reagents, and include soluble products such as antibodies (humoral immunity) as well as direct cell-to-cell contact resulting in specific cell lysis (cell-mediated immunity). No single mechanism has been described that can account for the diversity of autoimmune responses or the production of autoantibodies. Figure 1 outlines the common features of hypothetical models of autoantibody elicitation. Most models, particularly those relating to autoimmune disease in animals, include a genetic predisposition. Breeding experiments between inbred strains of mice have shown that the genetic control of autoantibody production is complex, involving multiple genes. Although most of the required genetic elements remain to be characterized, it appears that both acceleration and suppression of autoimmune responses are under genetic control. The most frequently observed genetic requirement involves the major histocompatibility complex (MHC) class II genes, which encode proteins responsible for the presentation of processed antigen to $CD4^+$ T cells via the T-cell receptor.

The most perplexing and challenging aspect of autoimmunity and autoantibody elicitation is the identification of the events involved in the initiation of the response. Although these early events are poorly understood for most autoimmune diseases, it is thought that an exogenous trigger can provide the first step in the initiation of some autoimmune responses. The best evidence for this comes from drug- and

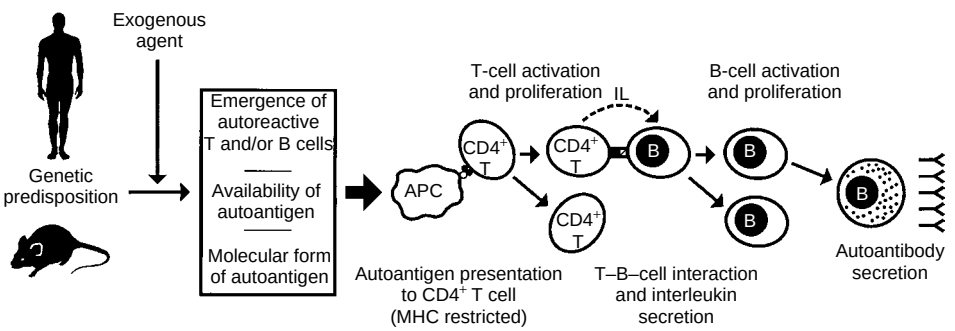

Fig. 1 Hypothetical pathway of autoantibody elicitation in human disease and experimental animal models. This model combines features from the most commonly accepted postulated mechanisms for autoantibody production. Genetically predisposed individuals may be triggered to begin the response by an exogenous agent such as exposure to a drug, chemical toxin, or other environmental influence. The events that follow (listed in large box) are poorly understood but must involve the emergence of autoreactive lymphoid cells and the presence of autoantigen in a molecular form reactive with autoreactive cells. Once the presentation of autoantigen has activated autoreactive lymphoid cells, the production of autoantibody proceeds essentially as it would for a nonautoimmune antibody response.

chemical-induced autoimmunity, which has been described in both human disease and animal models of autoimmunity. However, even in exogenously induced autoimmunity, many of the events between the administration of a chemical or a drug and the appearance of autoantibodies remain to be unveiled. Induction of autoantibodies by exogenous agents can take from several weeks to many months. Drug-induced systemic autoimmunity in humans can take prolonged periods of time to develop and can be provoked by a large number of chemically unrelated drugs. The autoantibody response, however, appears quite restricted, targeting histones and histone–DNA complexes, the components of chromatin. Complexes of drug and autoantigen are not the immunogens responsible for the autoantibody response, since the drug is not required for autoantibody interaction with the autoantigen. Withdrawal of the drug often leads to cessation of clinical symptoms, clearly implicating the participation of the drug in some mechanism inciting the autoimmune response, although the autoantibody may persist for months in the absence of the drug. In several animal models, exposure to chemicals, particularly inorganic forms of heavy metals such as mercury, silver, or gold, can lead to autoantibody expression within weeks. In these murine models, the autoantibody response is again restricted, but here the predominant targets are nonchromatin components of the nucleolus. The development of restricted autoantibody specificities in humans given many different drugs or in mice given heavy metals suggests that it is not the parent molecule that is important but rather the metabolic products of these compounds that lead on the one hand to antichromatin autoantibodies and on the other to antinucleolar antibodies. In human drug-induced autoimmunity, a common pathway of oxidative metabolism via the ubiquitous neutrophil has been suggested as a means of producing reactive drug metabolites that may perturb immune regulation sufficiently to produce autoimmune disease. Another mechanism that has been proposed is disruption by drug metabolites of positive selection of T cells during their development in the thymus. This mechanism has been shown to result in mature $CD4^+$ T cells that are able to respond to self-antigen leading to T-cell proliferation as well as autoantibody production by B cells.

In Fig. 1, the large boxed area highlights several concepts that form pivotal points in many hypothetical postulates of autoantibody elicitation but about which little is known. How do B- and T cells, with receptors for autoantigen, emerge from and escape the regulatory mechanisms that normally keep them in check, then make their way to the secondary lymphoid tissues? Studies involving transgenic mice possessing neoautoantigens suggest that possible mechanisms include avoidance of apoptotic elimination, escape from tolerance induction, and reversal of an anergic state. Molecular identification of autoantigens, their presence in macromolecular complexes, the occurrence of autoantibodies in different components of the same complex, and the appearance of somatic mutations in the variable regions of autoantibodies have suggested that it is the autoantigen that drives the autoimmune response. It remains unclear how autoantigens, particularly intracellular autoantigens, are made available to autoreactive lymphoid cells, and what molecular forms of these complex macromolecular structures interact with autoreactive lymphoid cells. One mechanism that has been proposed as a means by which autoantigens

might be made available to the immune system is apoptotic cell death. The impetus for this hypothesis is the finding that many autoantigens undergo proteolytic cleavage during apoptotic cell death and that apoptotic bodies (debris from dying cells) contain multiple autoantigens. Processing and presentation of such material by antigen-presenting cells (APC) has been suggested as a means of providing antigen to autoreactive T cells. However, uptake of apoptotic cellular material does not lead to the activation of APCs, which is necessary if APCs are to activate T cells. Inability of apoptotic material to activate APCs may stem from the observation that apoptosis is a descriptor for programmed cell death (PCD), which is a physiological process. This contrasts sharply with necrotic cell death, which is a nonphysiological process that produces cellular material that activates APCs. Also of note is that necrotic cell death induced by mercury leads to proteolytic cleavage of the autoantigen fibrillarin. Immunization with the N-terminal fragment of such cleavage leads to autoantibodies against fibrillarin that possess some of the characteristics of the antifibrillarin response elicited by mercury alone. In contrast, the antibody response elicited by immunization with full-length fibrillarin does not mimic the mercury-induced response, suggesting that processing and presentation of fragmented autoantigens may allow loss of self/nonself discrimination. Examination of the molecular forms of autoantigens during and after cell death and their roles in activating both APC and T cells will be fruitful areas of future research.

Roles in autoantibody production have been argued for pathways that either are or are not dependent on the presence of T cells. A T cell–dependent response is shown in Fig. 1 with an APC supplying processed antigen to $CD4^+$ T cells. An essential element in any model of autoantibody elicitation is the emergence of antibody-secreting B cells, which recognize material derived from the host. The antibody secreted by a B cell is directed against a single region (or epitope) on an antigen. An autoantibody response can target a number of epitopes on any one antigen, clearly showing that multiple autoreactive B-cell clones are activated during an autoimmune response. In the systemic autoimmune diseases, many autoantigens are complexes of nucleic acid and/or protein, and an autoimmune response may target several of the components of a complex. It is unknown whether the autoantibody responses to the components of a complex arise simultaneously, sequentially, independently, or through interrelated mechanisms.

In only a few diseases have autoantibodies been shown to be the causative agents of pathogenesis (e.g. antiacetylcholine receptor autoantibodies in myasthenia gravis, antithyroid stimulating hormone receptor autoantibodies in Graves' disease). It is noteworthy not only that these diseases are organ specific but also that their autoantigens are extracellular or on the surface of cell membranes and are therefore easily targeted by the immune system. In some individuals, the largest organ, the skin, can suffer insult from several blistering conditions now known to be autoimmune diseases characterized by autoantibodies against products of keratinocytes. The autoantigens involved are cell adhesion molecules that are important in maintaining the integrity of the skin by cell–cell contact between the various cell layers in the epidermis and at the dermal–epidermal junction. In

the non–organ-specific autoimmune disease systemic lupus erythematosus (SLE), anti–double-stranded DNA (dsDNA) autoantibodies have been shown to participate in pathogenic events by way of complexing with their cognate antigen to cause immune complex–mediated inflammation. These examples show that in both the organ-specific and systemic autoimmune diseases, *in vivo* disposition of autoantibody in tissues and organs has clinical significance inasmuch as it indicates sites of inflammation, which may contribute to the pathological process. Moreover, detection of autoantibody deposits in the organ-specific autoimmune diseases has particular significance because some organ-specific autoantibodies have been found to be the direct mediators of pathological lesions. In most autoimmune diseases, however, it has not been determined whether autoantibodies cause or contribute to disease or are merely a secondary consequence of the underlying clinical condition.

2 Autoantibodies as Diagnostic Markers

The diseases associated with autoantibodies can be divided into two broad groups: the organ-specific autoimmune diseases, in which autoantibodies have the ability to react with autoantigens from a particular organ or tissue, and the multisystem autoimmune diseases, in which autoantibodies react with common cellular components that appear to bear little relevance to the underlying clinical picture. In both cases, particular autoantibody specificities can serve as diagnostic markers (Table 1).

In the multisystem autoimmune diseases, there are several features of the relationship between autoantibody specificity and diagnostic significance that bear consideration. Autoantigens in these diseases are components of macromolecular structures such as the nucleosome of chromatin and the small nuclear ribonucleoprotein (snRNP) particles of the spliceosome, among others. Autoantibodies to different components of the same macromolecular complex can be diagnostic for different clinical disorders. Thus, the core proteins B, B′, D, and E, which are components of the U1, U2, and U4–U6 snRNPs and are antigenic targets in the anti-Smith antigen (Sm) response in SLE, are different from the U1 snRNP–specific proteins of 70 kDa, A and C, which are targets of the anti-nRNP response in mixed connective tissue disease (MCTD; see Table 1). It has also been observed that certain autoantibody responses are consistently associated with one another. The anti-Sm response, which is diagnostic of SLE, is commonly associated with the anti-nRNP response, but the anti-nRNP response can occur without the anti-Sm response, in which case it can be diagnostic of MCTD. These two observations suggest that the snRNP complexes responsible for the autoantibody response against the spliceosome in MCTD may differ from the snRNP complexes that produce the antispliceosome response in SLE. Other autoantibody responses demonstrate similar associations and restrictions. The anti-SS-A/Ro response (see Table 1) frequently occurs alone in SLE, but the anti-SS-B/La response in Sjögren's syndrome is almost always associated with the anti-SS-A/Ro response. Similarly, the antichromatin response occurs alone in drug-induced lupus but is usually associated with the anti-dsDNA response in idiopathic SLE.

Tab. 1 Examples of clinical diagnostic specificity of autoantibodies.

Autoantibody specificity[a]	***Molecular specificity***	***Clinical association***
Organ-specific autoimmune diseases		
Antiacetylcholine receptor*	Acetylcholine receptor	Myasthenia gravis
Anti-TSH receptor*	TSH receptor	Graves' disease
Antithyroglobulin*	Thyroglobulin	Chronic thyroiditis
Antithyroid peroxidase*	Thyroid peroxidase	Chronic thyroiditis
Antimitochondria*	Pyruvate dehydrogenase complex	Primary biliary cirrhosis
Antikeratinoctye*	Desmoplakin I homologue	Bullous pemphigoid
Antikeratinoctye*	Desmoglien	Pemphigus foliaceus
Multisystem autoimmune diseases		
Anti–double-stranded DNA*	B-form of DNA	SLE
Anti-Sm*	B, B′, D, and E proteins of U1, U2, U4–U6 snRNP	SLE
Anti-nRNP	70 kDa, A and C proteins of U1-snRNP	MCTD, SLE
Anti-SS-A/Ro	60 and 52 kDa proteins associated with hY1-Y5 RNP complex	SS, neonatal lupus, SLE
Anti-SS-B/La	47 kDa phosphoprotein complexed with RNA polymerase III transcripts	SS, neonatal lupus, SLE
Anti-Jo-1*	Histidyl tRNA synthetase	Polymyositis
Antifibrillarin*	34 kDa protein of box C/D containing snoRNP (U3, U8, etc.)	Scleroderma
Anti-RNA polymerase 1*	Subunits of RNA polymerase 1 complex	Scleroderma
Anti-DNA topoisomerase 1 (anti-Scl-70)*	100 kDa DNA topoisomerase I	Scleroderma
Anti-centromere*	Centromeric proteins CENP-A, B, C	CREST (limited Scleroderma)
Canca	Serine proteinase (proteinase 3)	Wegener's vasculitis

[a] Disease-specific diagnostic marker antibodies indicated by an asterisk.
Notes: SLE: systemic lupus erythematosus; MCTD: mixed connective tissue disease; SS: Sjögren's syndrome; cANCA: cytoplasmic antineutrophil cytoplasmic antibody; TSH: thyroid-stimulating hormone; CREST: calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, telangiectasia.

Autoantibody specificities may occur at different frequencies in a variety of diseases, and the resultant profile consisting of distinct groups of autoantibodies in different diseases can have diagnostic use. In some cases, the grouping of autoantibody specificities, such as the preponderance of antinucleolar autoantibodies in scleroderma (Table 1), provides provocative but as yet little-understood relationships with clinical diagnosis. Unlike SLE, where a single patient may have multiple autoantibody specificities to a number of unrelated nuclear autoantigens (e.g. DNA, Sm, SS-A/Ro), scleroderma patients infrequently

have multiple autoantibody specificities to nucleolar autoantigens that are unrelated at the macromolecular level (i.e. not part of the same macromolecular complex).

3 Autoantibodies as Molecular and Cellular Probes

Autoantibodies can be used for the detection of their cognate antigens using immunoprecipitation, immunoblotting, enzyme-linked immunosorbent assay (ELISA), and a variety of microscopy techniques including immunoelectron microscopy. The most visually impressive demonstration of the usefulness of autoantibodies as biological probes is the indirect immunofluorescence (IIF) test. In this technique (Fig. 2), a cell or tissue source containing the autoantigen of interest is permeabilized, to allow entry of the antibody into the cell, and fixed, to ensure that the target antigen is not leached away during the procedure. Although some procedures are inappropriate for particular antigens, workable means of cell permeabilization and fixation have been developed. The cell substrate is incubated with the autoantibody to allow interaction with the antigen, and any excess is washed away. The location of the antigen/autoantibody complex within the cell is revealed by addition of an anti-antibody tagged with a fluorochrome. Fluorescence microscopy is then used to view the cells to determine the location of the antigen-/autoantibody-/fluorochrome-tagged anti-antibody complex. Using this technique, investigators are identifying an increasing number of autoantibody specificities that recognize cellular substructures and domains (Table 2, Figs. 3 and 4). The nucleus can be identified by a variety of autoantibodies such as those against chromatin and DNA or, as shown in Fig. 3(a), autoantibodies to the nuclear lamina, which underlies the nuclear envelope and produces a ringlike fluorescence around the nucleus. The nucleolus and its subdomains can also be identified by a variety of autoantibody specificities (Table 2). Autoantibodies against the 34 kDa protein fibrillarin, a component of the C/D box containing small nucleolar ribonucleoprotein (snoRNP) particles, label the nucleolus in a distinctive "clumpy" pattern (Fig. 3e). The list of autoantibodies that are able to distinguish subnuclear domains and compartments, some considerably smaller than the nucleolus, continues to grow. One example is the Cajal body, a small subnuclear structure described using light microscopy by the Spanish cytologist Santiago Ramon y Cajal in 1903 and subsequently named after him. This nuclear domain can now be easily identified using human autoimmune sera that react with p80 coilin (Fig. 3d), a protein highly enriched in the Cajal body. Using other autoantibodies, it has been found that Cajal bodies contain snRNP particles and fibrillarin (previously thought to be restricted to the nucleolus and prenucleolar bodies). Knowledge of the functional associations of these coiled-body constituents suggests that the Cajal body may play a role in RNA processing and/or in the accumulation of components involved in RNA processing.

Many features of subcellular structures such as size, shape, and distribution can be studied by IIF during the cell cycle, viral infection, mitogenesis, or any cellular response that may result in changes in the distribution of an antigen or a subcellular structure. As shown in Fig. 3(a) (arrowheads), antinuclear lamin autoantibodies can be used to reveal re-formation of the lamina during telophase. Autoantibodies

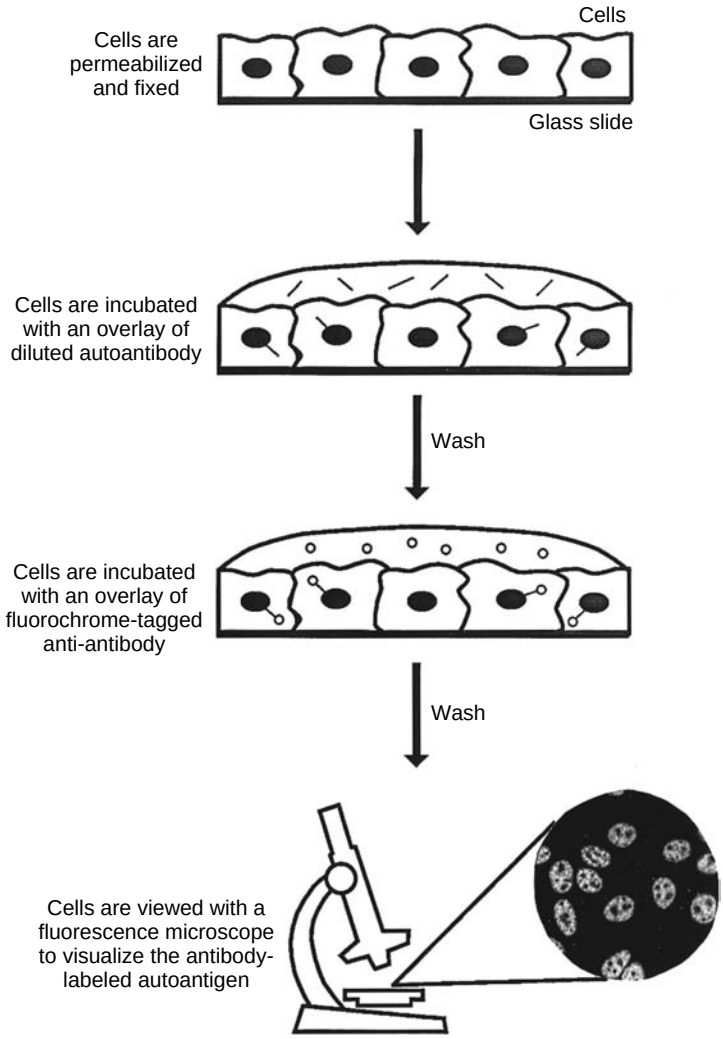

Fig. 2 Diagrammatic representation of the steps involved in the indirect immunofluorescence (IIF) test: see text for explanation.

have identified unexpected protein distributions such as the localization of the nucleolar protein fibrillarin to the outer surface of the chromosomes during cell division (Fig. 3e, arrowheads). The localization of some autoantigens during the cell cycle has aided in their identification. Detection of proliferating cell nuclear antigen (PCNA) in S-phase cells (Fig. 3c) suggested its involvement in DNA synthesis, while the distribution of speckles along the metaphase plate produced by other autoantibodies (Fig. 3f, arrowheads) was an important clue in their identification as autoantibodies to centromeric proteins A, B, and C.

The IIF test has also proved useful in the identification of autoantibodies

Tab. 2 Examples of subcellular structures and domains recognized by autoantibodies.

Autoantibody	*Molecular specificity*	*Subcellular structure*
Nuclear components		
Antichromatin	Nucleosomal and subnucleosomal complexes of histones and DNA	Chromatin
Anti-nuclear pore	210 kDA glycoprotein (gp210)	Nuclear pore
Antilamin	Nuclear lamins A, B, C	Nuclear lamina
Anticentromere	Centromere proteins (CENP) A, B, C, F	Centromere
Anti-p80 coilin	p80-coilin (80 kDa protein)	Coiled body
Anti-PIKA	p23–25 kDa proteins	Polymorphic interphase karyosomal association (PIKA)
Anti-NuMA	238 kDa protein	Mitotic spindle apparatus
Nucleolar components		
Antifibrillarin	34 kDa fibrillarin	Dense fibrillar component of nucleolus
Anti-RNA polymerase 1	RNA polymerase 1	Fibrillar center of nucleolus
Anti-Pm-Scl	75 and 100 kDa proteins of Pm-Scl complex	Granular component of nucleolus
Anti-NOR 90	90 kDa doublet of (human) upstream binding factor (hUBF)	Nucleolar organizer region (NOR)
Cytosolic components		
Antimitochondria	Pyruvate dehydrogenase complex	Mitochondria
Antiribosome	Ribosomal P-proteins (P_0, P_1, P_2)	Ribosomes
Anti-Golgi	95 and 160 kDa golgins	Golgi apparatus
Antiendosome	180 kDa protein	Early endosomes
Antimicrosomal	Cytochrome P450 superfamily	Microsomes
cANCA	Serine proteinase (proteinase 3)	Lysosomes
Antimidbody	38 kDa protein	Midbody
Anticentrosome/centriole	Pericentrin (48 kDa)	Centrosome/centriole

Notes: NuMA: nuclear mitotic apparatus; Pm-Scl: polymyositis-scleroderma; cANCA: cytoplasmic antineutrophil cytoplasmic antibody.

that react with subcellular structures other than the nucleus. Figure 4 shows the IIF patterns produced by some of these autoantibodies. Prior knowledge of subcellular organelles and their relative cellular distribution was instrumental in identifying the structures recognized by these and other autoantibodies. In turn, autoantibodies, by virtue of their reactivity with individual autoantigens, have allowed cell and molecular biologists an insight into the molecular constituents of these subcellular organelles.

Comparative studies using human autoantibodies, and nonhuman autoantibodies raised by immunization against specific cellular proteins, can be useful in determining the cellular distribution of the specific protein. This is achieved by using antihuman antibodies labeled with one chromophore and antibodies specific for the nonhuman antibody that

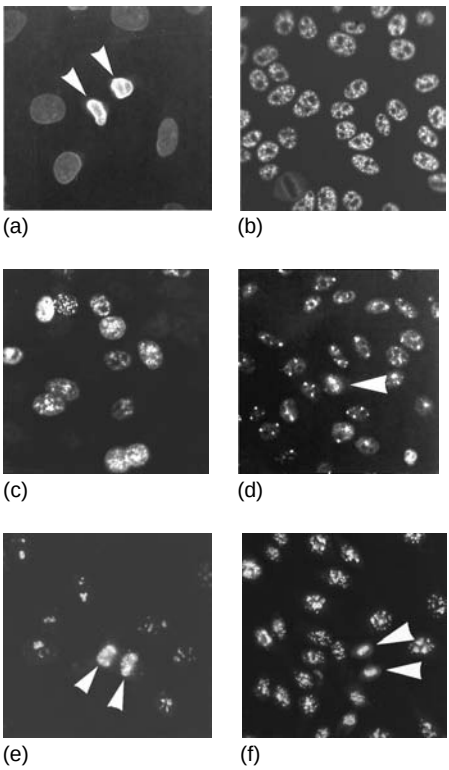

Fig. 3 Immunofluorescence patterns produced by autoantibodies recognizing structural and functional domains within the cell nucleus (magnification, 350×). (a) Antinuclear lamin B1 antibodies identify the periphery of the nucleus; arrowheads show the re-formation of the nuclear envelope during late telophase. (b) Anti-Sm antibodies localize the U1, U2, and U4–U6 snRNP particles as a speckled nuclear pattern. (c) Anti-PCNA antibodies recognize the auxiliary protein of DNA polymerase delta during active DNA synthesis, producing different fluorescence patterns as cells progress through mitosis. (d) Anti-p80 coilin antibodies highlight subnuclear domains known as Cajal bodies, which disappear during metaphase (arrowhead). (e) Antifibrillarin antibodies target the nucleolus and produce a characteristic clumpy pattern in interphase cells, decorating the chromosomes from late metaphase until cell division (arrowhead). (f) Antibodies to centromeric proteins A, B, and C produce a discrete speckling of the interphase nucleus and identify the centromeric region of the dividing chromosomes during cell division (arrowheads).

are labeled with a different chromophore and by comparing the fluorescence patterns. Immunolocalization of the non-snRNP spliceosome component SC-35 was achieved in this way by comparison of anti-SC-35 antisera with the IIF pattern of autoimmune anti-Sm sera, which recognize protein components of the spliceosome.

One feature of autoantibodies that distinguishes them from antibodies raised by specific immunization, and underscores their uniqueness, is their ability to recognize their target antigen not only from the host but also from a variety of species. The extent of this species cross-reactivity is dependent on the evolutionary conservation of the autoantigen and is related to the conservation of protein sequence. One example is the snoRNP protein fibrillarin. Using autoantibodies in a variety of immunological techniques, including IIF, this protein can be recognized from species as diverse as humans and the unicellular yeast *Saccharomyces cerevisiae*. cDNA cloning of fibrillarin has confirmed the expected high degree of conservation of the protein sequence.

The reactivity of autoantibodies with conserved sequence and conformational protein elements has made them useful reagents in the cloning of cDNAs of expressed proteins from cDNA libraries from a variety of species. However, because of their reactivity with the human

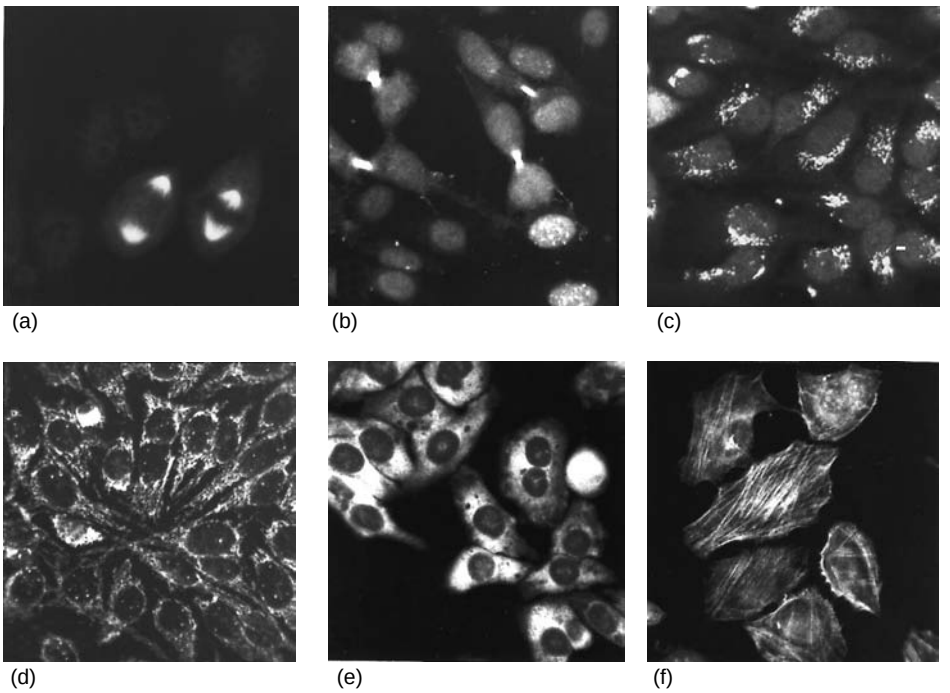

Fig. 4 Immunofluorescence patterns produced by autoantibodies recognizing intracellular structures other than the nucleus. (a) Antimitotic spindle apparatus antibodies identify spindle poles and spindle fibers during cell division. (b) Antimidbody antibodies react with the bridgelike midbody that connects daughter cells following chromosome segregation but before cell separation. (c) Anti-Golgi complex antibodies decorate the Golgi apparatus, which in most cells is shown as an accumulation of fluorescence in a discrete cytoplasmic region. (d) Antimitochondrial antibodies demonstrate the presence of mitochondria throughout the cytoplasm; the discrete nuclear dots represent an additional autoantibody specificity in this serum unrelated to mitochondria. (e) Antiribosome antibodies produce a diffuse cytoplasmic staining pattern that spares the nucleus but may show some weak nucleolar fluorescence. (f) Anticytoskeletal antibodies react with a variety of cytoskeletal components; in this case, the antibody reacts with nonmuscle myosin. (magnification: a, 700×; b–f, 350×).

protein they have found most use in the cloning and characterization of the primary structure of numerous human cellular proteins. This diversity of targets that can be exploited by this approach is clearly illustrated in Tables 1 and 2.

Elucidation of the structure of the autoantigens that are the targets of autoantibodies from systemic autoimmune diseases has revealed that many are functional macromolecular complexes involved in nucleic acid and protein synthesis and processing (Table 3). A distinguishing feature of many of these complexes of nucleic acids and/or proteins is that autoantibodies do not recognize all the components of the complex. An extreme, but useful, example is the ribosome, which in eukaryotes may contain more than 70 proteins. However, only the P-proteins (P_0, P_1, and P_2), S10, and L12 are recognized by autoantibodies. Nonetheless,

Tab. 3 Examples of the function of nuclear autoantigens and the effects of autoantibody on antigen function.

Autoantigen	***Function***	***Autoantibody effect***[a]
	Known function	
Sm/nRNP (U1, 2, 4–6 snRNP)	Pre-mRNA splicing	Inhibition of pre-mRNA splicing
PCNA (DNA polymerase delta auxiliary protein)	DNA replication	Inhibition of DNA replication and repair
RNA polymerase I	Transcription of rRNA	Inhibition of rRNA transcription
tRNA polymerase	Aminoacylation of tRNA	Inhibition of charging of tRNA
Ribosomal RNP	mRNA translation	Inhibition of protein synthesis
Centromere/kinetochore	Microtubule-based chromosome movement during mitosis	Inhibition of centromere formation and function
	Probable function	
Fibrillarin (Box C/D snoRNP)	Processing and methylation of pre-rRNA	Blocks translocation of fibrillarin during the cell cycle, thereby influencing the ultrastructure of the nucleolus.
NOR-90	Nucleolar transcription factor	Not tested

[a] Inhibition of function has been demonstrated *in vitro* or following injection of autoantibody into living cells.

the use of autoantibodies that identify specific components of such complexes has aided in identifying other subunits of these complexes, with profound consequences. Thus, the initial identification of anti-Sm and anti-nRNP autoantibodies in SLE led to the observation that they recognize some of the protein components of the snRNP particles, fueling subsequent studies that showed the snRNPs as components of the spliceosome complex that functions in pre-mRNA splicing.

As the functional associations of autoantigens have become known, attempts to uncover the role of the autoantigen itself have revealed that autoantibodies can directly inhibit the function of their target autoantigen (Table 3). Although it remains to be determined, it seems likely that such inhibition reflects the involvement of a conserved protein sequence or structure in functional activity. An increasing number of autoantibodies, many of unknown molecular specificity, recognize their autoantigen only in a particular functional state or phase of the cell cycle. Of the several examples known, the best characterized is PCNA, which is the auxiliary protein of DNA polymerase delta and is recognized by autoantibodies only during mitosis, even though PCNA is present throughout the cell cycle. When a population of cells at different stages of the cell cycle is used as substrate in the IIF test, anti-PCNA autoantibodies produce varying degrees of fluorescence intensity, being negative for G_0 cells and highly positive for S-phase cells (Fig. 3c). These

intriguing features of some autoantibodies have added new dimensions to the biological usefulness of these proteins and have suggested that functionally active macromolecular complexes may play a role in the elicitation of the autoantibody response.

The presence of multiple autoantibody specificities in the blood of individual human patients with autoimmune diseases poses a limitation on their use in studies involving a single autoantigen. Only infrequently are patients found whose autoantibody response is so restricted that they express autoreactivity to a single autoantigen or autoantigenic complex; such autoantibodies are termed *monospecific*. For some autoantibody specificities, this condition has been overcome by the production of hybridomas secreting a monoclonal autoantibody. Some hybridomas have been produced by fusion of B cells from human patients, but most have come from fusion of lymphoid cells from animal models of autoimmunity, particularly inbred murine strains. Monoclonal antibody specificities include reactivity against the nucleic acids DNA and RNA, subunits of chromatin, protein components of snRNP particles, fibrillarin, and immunoglobulin.

4 Autoantibodies in Experimental Models of Autoimmunity

Research into the mechanisms of autoimmunity and the antigenic specificity, and possible pathogenic role, of autoantibodies has been significantly advanced by the availability of animal models. Four different types of models have been used (Table 4). Specific antigen immunization models are produced by direct injection of purified antigen into animals to elicit autoantibody. Direct immunization has proven most useful when the autoantigen is extracellular or is on the cell membrane. In such examples, the elicited autoantibody response can produce pathological consequences such as the myasthenia gravis–like disease produced in rodents following immunization with purified acetylcholine receptor. The animals used in this type of model are most often healthy, normal individuals with fully functional immune systems and are able to downregulate the autoimmune response produced by the immunization of autoantigen. As a result, direct immunization models often produce transient autoimmune responses and the animals return to a healthy state.

Comparison of the autoantigenic reactivities of antibodies raised by immunization, especially to intracellular autoantigens, has revealed distinct differences in comparison to autoantibodies found in human autoimmune disease. Direct immunization requires a purified antigen, which means subjecting the antigen to rigorous biophysical, biochemical, and sometimes immunological separation techniques. The resulting preparation may therefore be partially or totally denatured and no longer in association with other cellular components that constitute its *in vivo* molecular form. Even if the native *in vivo* macromolecular complex can be purified, direct immunization experiments are "best guess" attempts to mimic the natural autoimmunization process because the molecular structure of the putative autoimmunogen that contains the autoantigen of interest is unknown. A further complication is the use of adjuvants to boost the immune response. As a result, direct immunization produces antibodies that, although they react with the autoantigen, usually recognize a denatured form rather than

Tab. 4 Examples of animal models of autoantibody production.

Model	***Animal***	***Human disease***	***Autoantibody specificity***
Spontaneous			
NZB	Mouse	Hemolytic anemia	Erythrocyte
(NZB × NZW) F1	Mouse	SLE	Chromatin, DNA
MRL-*lpr*/*lpr*	Mouse	SLE	Chromatin, DNA, Sm, ribosome
MRL-+/+	Mouse	SLE	Chromatin, DNA, Sm
BXSB	Mouse	SLE	Chromatin, DNA
Obese strain (white leghorn chicken)	Bird	Thyroiditis	Thyroglobulin
Induced by exogenous agents			
Chronic GVHD	Mouse	SLE	DNA, chromatin, snRNP, ribosome
Mercuric chloride	Mouse (H-2^s)	Scleroderma[a] (immune-complex nephritis)	Fibrillarin (Box C/D snoRNP)
Mercuric chloride	Mouse (non-H-2 restricted)	SLE	Chromatin
Mercuric chloride	Rat ($RT1^n$)	Immune-complex nephritis	GBM
Pristane	Mouse	SLE	DNA, Sm, RNP, Su
Direct immunization (antigen)			
EAT (thyroglobulin)	Rabbit, mouse (H-$2^{k,s,\text{ or } q}$)	Thyroiditis	Thyroglobulin
GMB nephritis (GBM)	Sheep, mouse	Immune-complex nephritis	GBM
EMG (acetylcholine receptor)	Lewis rat	Myasthenia gravis	Acetylcholine receptor
Gene mutation			
C1q knockout	Mouse (MRL-+/+)	SLE	DNA, rheumatoid factor
Dnase1 knockout	Mouse	SLE	Chromatin, DNA
SAP knockout	Mouse	SLE	Chromatin, DNA
c-mer knockout	Mouse	SLE	Chromatin
IFN-γ transgenic	Mouse	SLE	DNA, histones

[a]Autoantibody specificity is specific for scleroderma, but a scleroderma-like disease has not been described in mice treated with mercuric chloride.
Notes: GVHD: graft-versus-host disease; GMB: glomerular basement membrane; EAT: experimental autoimmune thyroiditis; EMG: experimental myasthenia gravis; C1q: component of serum complement; Dnase1: deoxyribonuclease 1; SAP: serum amyloid P component; c-mer: tyrosine kinase.

the autoantigen in its native state; only rarely do they react with the autoantigen when it is in association with other cellular subunits that make up its *in vivo* molecular form. Although direct-immunization antibodies can recognize conformational epitopes, they do not appear to recognize conserved epitopes and therefore cannot exhibit the same lack of species specificity that allows, for example, antifibrillarin autoantibodies to recognize fibrillarin in all species that contain this protein. Lack of reaction against conserved epitopes means that direct immunization antibodies are less efficient at inhibiting the functional activity of their target autoantigen than autoantibodies from patients with autoimmune diseases. Animal models of other types, described next, can produce autoantibodies with reactivities that are extremely difficult to differentiate from those of human patients. As a result, such models more closely approximate their human counterparts.

The second type of model also involves the manipulation of normal, nonautoimmune animals to produce an autoimmune response. In these cases, the triggering event is the introduction of exogenous material into the animal, which, unlike the case of direct immunization, may appear to bear little relationship to the ensuing autoimmune response. An excellent example of this type of model is the autoimmunity induced by heavy metals. Administration of mercury by several different routes and in several different forms, most notably subcutaneous injection of mercuric chloride, produces in mice an autoantibody response that targets the nucleolus. The principal autoantigen involved is the 34 kDa protein fibrillarin (Fig. 5), a protein component of the box C/D snoRNP particles. Mercury induces this autoantibody response in a restricted number of histocompatibility genotypes, most commonly H-2^s. Although offspring of crosses between the autoimmune-sensitive H-2^s strains and the autoimmune-resistant strains such as C57BL/6 (H-2^b) or DBA/2 (H-2^d) are sensitive to antifibrillarin induction following $HgCl_2$ treatment, the response does not appear to be solely due to the product of a dominant *H-2* gene but involves multiple loci as well. This is supported by back-crossing of hybrids onto the autoimmune-sensitive H-2^s background, where the $HgCl_2$-induced antifibrillarin response is even less frequent, even though 50% of the mice would be expected to be homozygous for H-2^s. Although antifibrillarin autoantibodies are a marker for human scleroderma, mercury does not appear to produce a scleroderma-like disease in mice; the importance of the model lies instead in the similarity of this toxin-induced murine autoantibody's response to the spontaneous antifibrillarin autoantibody in human scleroderma.

Another example of the exogenous factor-type model is murine graft-versus-host disease (GVHD). In this model, the offspring of the mating of two inbred nonautoimmune mouse strains are injected (grafted) with lymphocytes from one of the parental strains. The injected lymphocytes recognize genetic differences in the host strain that are inherited from the other parental strain and are stimulated to mount a variety of immune responses against the host animal; hence the name "graft versus host." Unlike the case of direct immunization models, the autoimmunity produced in this type of model can lead to severe pathological consequences including lethal immune-complex disease. The immunological sequelae that occur during a GVHD response depend on the murine inbred strains used. Injection of

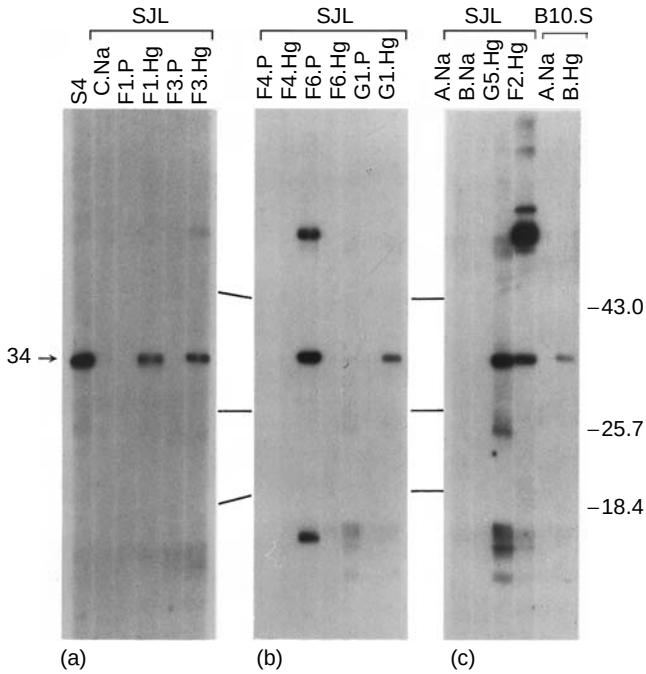

Fig. 5 Antifibrillarin autoantibodies in H-2^s murine strains (SJL and B10.S) after treatment with mercuric chloride. Autoantibodies were detected by immunoblotting using mouse liver nucleoli resolved by sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS-PAGE). (a) Prototype human antifibrillarin serum (S4), serum from a NaCl-treated SJL mouse (C.Na), and serum from mice F1 and F3 before (P) and after (Hg) mercuric chloride treatment. (b) Immunoblotting using sera from three additional SJL mice before and after mercuric chloride treatment. (c) Immunoblotting using sera from two SJL mice treated with NaCl (A.Na; B.Na); two SJL mice treated with mercuric chloride (G5.Hg; F2,Hg); and two B10.S mice, one treated with NaCl (A.Na) and the other treated with mercuric chloride (B.Hg). (Reproduced from Hultman et al., *Clin. Exp. Immunol.* 78 (1989), 470–477 with permission from the publishers.)

DBA/2 lymphocytes into a cross between the DBA/2 and C57B16 mice produces a chronic GVHD, which results in an autoimmune response similar to human SLE including the presence of autoantibodies against chromatin and DNA. Injection of lymphocytes from the A/J strain into Balb/c X A/J hybrids also produces a chronic GVHD in which autoantibodies to snRNP particles including the U3 snoRNP are found. The relationship of different autoantibody specificities to the use of different strains of inbred mice in the GVHD model again highlights the influence of genotype on autoimmunity and autoantibodies.

The third type of model does not require any manipulation of the animal at all; the disease develops spontaneously. The best described of these are the murine strains

BXSB, (NZB × NZW) F1, NZM, MRL-+/+, and MRL-*lpr*/*lpr*, which develop forms of SLE that serve as excellent models of the autoantibody specificities and pathology of the human disease. While the variety of autoantibodies developed by these different strains continues to be investigated, the common autoantibody response, like human SLE, is against chromatin and its subcomponents including DNA. In the (NZB × NZW) F1 strain, autoimmune disease and autoantibodies occur earlier and more frequently in female mice, a finding that has been found to be associated with the presence of female sex hormones. Because of this and other features, the (NZB × NZW) F1 strain is considered the best animal model of human SLE. As noted above, it is the genetic makeup of these inbred strains that has significant potential to address the genesis of the autoimmune response. A much studied aspect of several of the spontaneous models of systemic autoimmune disease is the presence of single-gene defects that accelerate or exacerbate autoimmunity in these already susceptible mouse strains. In the MRL substrains, the *lpr* phenotype is responsible for massive lymphoproliferation of $CD4^-$, $CD8^-$, and $B200^+$ T cells and an accelerated occurrence of autoimmune phenomena compared to the MRL-+/+. Recent studies have indicated that the *lpr* defect is due to a mutation in the *fas* gene that leads to defective expression of *fas* on T- and B cells, which allows them to escape apoptotic elimination and reach the peripheral circulation. Breeding experiments to impart the *lpr* gene to nonautoimmune genetic backgrounds have shown that the *fas* defect does influence the development of autoimmunity and the expression of autoantibodies. A dominant role for *fas* in the initiation of autoimmunity and autoantibodies is questionable, however, because the MRL-+/+, which does not have the *fas* defect, does develop a autoantibody profile and immunopathological disease that is similar to the MRL-*lpr*/*lpr*, albeit at a much later age. Other genes that appear to play a role in acceleration of autoimmunity include *gld*, the ligand for *fas*, and *Yaa*, a sex-linked gene that produces a defect in B cells and is the accelerator gene of autoimmunity in the male BXSB mouse. Exposure of lupus-prone strains to exogenous agents known to elicit autoimmunity in normal mice can result in accelerated appearance of disease features including autoantibodies. In some cases, the exogenous agent accelerates the appearance of idiopathic disease, while in others the elicited disease has features of xenobiotic-induced disease. Thus, mercury exposure accelerates idiopathic disease in BXSB mice including antichromatin autoantibodies of the IgG2a subclass, while pristane injection into (NZB × NZW) F1 mice elicits anti-Sm/RNP and Su autoantibodies that are not part of the idiopathic disease of the (NZB × NZW) F1 but are found in pristane-induced autoimmunity. These observations suggest that not only idiopathic and induced autoimmunity may arise through different mechanisms but also exogenous triggers can influence disease expression.

The fourth type of model involves genetic manipulation in which a gene is deleted (''knockout'') or added (''transgenic'') in order to influence the expression of autoimmunity. Both types of genetic modification can be used to study the influence of single genes on the animal models described above. Perhaps not unexpectedly, many gene deletions have little

or no effect on the expression of autoimmunity and autoantibodies. Such negative effects need to be interpreted carefully as they may indicate a genetically redundant process rather than an unimportant gene. Other gene deletions have been reported to influence differing aspects of autoimmunity in a gene-specific manner, although the extent of the effect may vary between experimental models. Some gene deletions exhibit highly consistent responses. Thus, deletion of the gene for the pleiotropic cytokine interferon-γ (IFN-γ) abrogates autoantibody production and immunopathology in mercury-induced autoimmunity of B10.S mice and spontaneous autoimmunity in MRL-*lpr*/*lpr* mice. The significance of IFN-γ in systemic autoimmunity has been demonstrated in nonautoimmune-prone mice made transgenic for IFN-γ expression in the epidermis. The increased expression of IFN-γ leads to a lupus-like disease characterized by the production of autoantibodies and immune complex–mediated tissue damage. Further evidence for the importance of IFN-γ has come from an examination of gene expression in the Nba2 locus of chromosome 1 of the mouse. Nba2 is a genetic interval identified as a locus of genetic susceptibility for lupus in the NZB strain. The offspring of Nba2 interval–specific congenic C57BL/6 mice mated with NZW mice develop autoimmunity similar to the SLE-prone (NZB × NZW) F1 mouse. Examination of gene expression by DNA array revealed a relationship between increased expression of interferon inducible gene (*Ifi*) 202 and features of systemic autoimmunity. Importantly, the gene for Ifi202 lies within the Nba2 interval. Confirmation that increased expression of Ifi202 occurs in other models of systemic lupus would significantly enhance its stature as a lupus susceptibility gene. However, as susceptibility for SLE maps to multiple genetic loci, it is highly likely that additional genes contribute to full disease expression in the (NZB × NZW) F1 mouse.

A number of other gene deletions are associated with expression of autoimmunity and autoantibodies. Some of these such as deficiency of C1q, a component of the complement system, have particular relevance as complement deficiencies in humans and can lead to development of systemic lupus. Significantly, lack of C1q is not sufficient for the development of murine lupus; this gene deletion must occur on genetic backgrounds carrying additional susceptibility genes for autoimmunity to occur. It must also be noted that although many knockout and transgenic models exhibit features of autoimmunity, they may also exhibit other features that are not consistent with the known spectrum of clinical and immunological facets of autoimmune diseases. More telling is the finding that many of the genetic mutations that lead to autoimmunity in mice are not necessary for the development of human systemic autoimmune disease. As described above, mutation in the *fas* gene contributes significantly to the severity of murine SLE. However, mutations in the *fas* gene are not associated with human SLE but rather with autoimmune lymphoproliferative syndrome (ALPS). ALPS is characterized by lymphoproliferation of double-negative T cells and autoantibodies against DNA and cardiolipin, features found in mice with a *fas* mutation but without other lupus susceptibility genes. Similarly, Dnase1-deficient mice develop a lupus-like disease with antichromatin autoantibodies, but deficiency of Dnase1 is not common in human systemic lupus. A nonsense mutation in exon 2 of the *DNASE1* gene has been reported in two apparently unrelated young Japanese

patients but not in SLE patients in Spain or the United States.

Although care must be taken in their interpretation, identification and characterization of genes that are associated with autoimmunity and autoantibody production constitute fertile ground for the molecular biologist. Elucidation of the roles of the many genes that appear to contribute to the development of autoimmunity will help define the critical molecular events in the disease process. The murine strains described above have proven valuable model systems for studies on a variety of facets of autoimmunity, and will play significant roles in future genetic studies. It will be important to focus attention not only on the genetic loci that impart susceptibility to autoimmunity but also on those that may allow an individual to resist the development of autoimmune phenomena.

5 Perspectives

Initially used as aids in clinical diagnosis, autoantibodies have become increasingly useful "reporter" molecules in the identification of structure–function relationships. New autoantigens continue to be discovered, while many described autoantigens remain to be characterized both structurally and functionally. Autoantibodies will figure prominently in these characterization studies. As the molecular structures of the interaction between autoantigen and autoantibody become known, it should be possible to design peptide configurations capable of perturbing the functional activity of numerous cellular processes.

Understanding the influence of genes and their products not only on susceptibility but also on resistance to autoimmunity and autoantibody expression is in its infancy. However, the tools to mature this field (inbred animal models of spontaneous and induced autoimmunity, and the molecular techniques of transgenics and gene knockout) are already available. They await the complex but potentially fruitful identification and functional analysis of candidate genes.

See also Antibody Molecules, Genetic Engineering of.

Bibliography

Books and Reviews

Earnshaw, W.C., Rattner, J.B. (1991) The Use of Autoantibodies in the Study of Nuclear and Chromosomal Organization, in: Hamkalo, B.A., Elgin, S.C.R. (Eds.) *Methods in Cell Biology*, Vol. 35, Academic Press, New York.

Kono, D.H., Theofilopoulos, A.N. (2001) Genetics of murine models of systemic lupus erythematosus, in: Wallace, D.J., Hahn, B.H. (Eds.) *Dubois' Systemic Lupus Erythematosus*, 6th edition, Williams & Wilkins, Baltimore.

Matzinger, P. (2002) The danger model: a renewed sense of self, *Science* **296**, 301–305.

Medzhitov, R., Janeway, C.A. Jr. (2002) Decoding the patterns of self and nonself by the innate immune system, *Science* **296**, 298–300.

Peter, J.B., Shoenfeld, Y. (Eds.) (1996) *Autoantibodies*, Elsevier, Amsterdam.

Pollard, K.M., Hultman, P. (1997) Effects of mercury on the immune system, *Met. Ions Biol. Syst.* **34**, 421–440.

Rose, N.R., Mackay, I.R. (Eds.) (1998) *The Autoimmune Diseases*, 3rd edition, Academic Press, New York.

Rubin, R.L., Pollard, K.M. (1997) Chemical-Induced Autoimmunity, in: Hertzenberg, L.A., Weir, D.M., Hertzenberg, L.A., Blackwell, C. (Eds.) *Handbook of Experimental Immunology*, 5th edition, Blackwell Scientific Publications, London.

Tan, E.M. (1989) Antinuclear antibodies: diagnostic markers for autoimmune diseases

and probes for cell biology, *Adv. Immunol.* **44**, 93–151.

Primary Literature

Balomenos, D., Rumold, R., Theofilopoulos, A.N. (1998) Interferon-gamma is required for lupus-like disease and lymphoaccumulation in MRL-lpr mice, *J. Clin. Invest.* **101**, 364–371.

Bickerstaff, M.C., Botto, M., Hutchinson, W.L., Herbert, J., Tennent, G.A., Bybee, A. et al. (1999) Serum amyloid P component controls chromatin degradation and prevents antinuclear autoimmunity, *Nat. Med.* **5**, 694–697.

Chatenoud, L., Salomon, B., Bluestone, J.A. (2001) Suppressor T cells – they're back and critical for regulation of autoimmunity! *Immunol. Rev.* **182**, 149–163.

Gall, J.G. (2000) Cajal bodies: the first 100 years, *Annu. Rev. Cell. Dev. Biol.* **16**, 273–300.

Kono, D.H., Theofilopoulos, A.N. (2000) Genetics of systemic autoimmunity in mouse models of lupus, *Int. Rev. Immunol.* **19**, 367–387.

Kono, D.H., Balomenos, D., Pearson, D.L., Park, M.S., Hildebrandt, B., Hultman, P., Pollard, K.M. (1998) The prototypic Th2 autoimmunity induced by mercury is dependent on IFN-gamma and not Th1/Th2 imbalance, *J. Immunol.* **161**, 234–240.

Kono, D.H., Park, M.S., Szydlik, A., Haraldsson, K.M., Kuan, J.D., Pearson, D.L., Hultman, P., Pollard, K.M. (2001) Resistance to xenobiotic-induced autoimmunity maps to chromosome 1, *J. Immunol.* **167**, 2396–2403.

Matzinger, P. (2002) The danger model: a renewed sense of self, *Science* **296**, 301–305.

Medzhitov, R., Janeway, C.A. Jr. (2002) Decoding the patterns of self and nonself by the innate immune system, *Science* **296**, 298–300.

Mitchell, D.A., Pickering, M.C., Warren, J., Fossati-Jimack, L., Cortes-Hernandez, J., Cook, H.T., Botto, M., Walport, M.J. (2002) C1q deficiency and autoimmunity: the effects of genetic background on disease expression, *J. Immunol.* **168**, 2538–2543.

Napirei, M., Karsunky, H., Zevnik, B., Stephan, H., Mannherz, H.G., Moroy, T. (2000) Features of systemic lupus erythematosus in Dnase1-deficient mice, *Nat. Genet.* **25**, 177–181.

Pollard, K.M. (2002) Cell death, autoantigen cleavage, and autoimmunity, *Arthritis Rheum.* **46**, 1699–1702.

Pollard, K.M., Hultman, P. (1997) Effects of mercury on the immune system, *Met. Ions Biol. Syst.* **34**, 421–440.

Pollard, K.M., Pearson, D.L., Bluthner, M., Tan, E.M. (2000) Proteolytic cleavage of a self-antigen following xenobiotic-induced cell death produces a fragment with novel immunogenic properties, *J. Immunol.* **165**, 2263–2270.

Pollard, K.M., Pearson, D.L., Hultman, P., Deane, T.N., Lindh, U., Kono, D.H. (2001) Xenobiotic acceleration of idiopathic systemic autoimmunity in lupus-prone bxsb mice, *Environ. Health Perspect.* **109**, 27–33.

Pollard, K.M., Pearson, D.L., Hultman, P., Hildebrandt, B., Kono, D.H. (1999) Lupus-prone mice as models to study xenobiotic-induced acceleration of systemic autoimmunity, *Environ. Health Perspect.* **107**(Suppl. 5), 729–735.

Rozzo, S.J., Allard, J.D., Choubey, D., Vyse, T.J., Izui, S., Peltz, G., Kotzin, B.L. (2001) Evidence for an interferon-inducible gene, Ifi202, in the susceptibility to systemic lupus, *Immunity* **15**, 435–443.

Rubin, R.L., Kretz-Rommel, A. (2001) A nondeletional mechanism for central T-cell tolerance, *Crit. Rev. Immunol.* **21**, 29–40.

Salomon, B., Bluestone, J.A. (2001) Complexities of CD28/B7: CTLA-4 costimulatory pathways in autoimmunity and transplantation, *Annu. Rev. Immunol.* **19**, 225–252.

Satoh, M., Richards, H.B., Shaheen, V.M., Yoshida, H., Shaw, M., Naim, J.O., Wooley, P.H., Reeves, W.H. (2000) Widespread susceptibility among inbred mouse strains to the induction of lupus autoantibodies by pristane, *Clin. Exp. Immunol.* **121**, 399–405.

Seery, J.P., Carroll, J.M., Cattell, V., Watt, F.M. (1997) Antinuclear autoantibodies and lupus nephritis in transgenic mice expressing interferon gamma in the epidermis, *J. Exp. Med.* **186**, 1451–1459.

Steinman, R.M., Nussenzweig, M.C. (2002) Avoiding horror autotoxicus: the importance of dendritic cells in peripheral T cell tolerance, *Proc. Natl. Acad. Sci. USA* **99**, 351–358.

von Muhlen, C.A., Tan, E.M. (1995) Autoantibodies in the diagnosis of systemic rheumatic diseases, *Semin. Arthritis Rheum.* **24**, 323–358.

Vyse, T.J., Kotzin, B.L. (1998) Genetic susceptibility to systemic lupus erythematosus, *Annu. Rev. Immunol.* **16**, 261–292.

Autoimmunity in Scleroderma

Yoshinao Muro[1] *and Kenji Sugimoto*[2]
[1] *Nagoya Graduate University School of Medicine, Nagoya, Japan*
[2] *Osaka Prefecture University, Sakai, Osaka, Japan*

1 Scleroderma and Autoantibodies 483
1.1 Methods of Autoantibody Detection 483
1.2 Clinical Subsets of Scleroderma and Autoantibodies 483
1.3 Detection of Scleroderma Autoantibodies and Epitope Analysis 486

2 Centromere Antigens 486
2.1 Autoimmunity against Centromere Antigens 486
2.2 CENP-B and B-cell Epitopes 488
2.3 CENP-C and B-cell Epitopes 490
2.4 CENP-A and B-cell Epitopes 491
2.5 Centromere-associated Autoantigens 491

3 Topoisomerase I 492
3.1 Autoimmunity against Topoisomerase I 492
3.2 B-cell Epitope 492
3.3 T-cell Response 493

4 Scleroderma-associated Nucleolar Autoantigens 494
4.1 Fibrillarin 494
4.2 PM/Scl Antigen 494

5 Molecular Mimicry and Scleroderma Autoantigens 495

6 Apoptosis and Scleroderma Autoantigens 495

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Bibliography 497
Books and Reviews 497
Primary Literature 497
Recent Key Primary Literature 499

Keywords

Anticentromere Antibodies
A centromere is the essential domain for proper segregation of chromosomes to daughter cells at mitosis and meiosis. Antibodies to the centromere are found in about 20% of patients with scleroderma, almost all of whom have limited scleroderma, formerly called *CREST* (calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia) syndrome.

Antitopoisomerase I Antibodies
Anti-DNA topoisomerase I antibodies, formerly termed *anti-Scl-70* antibodies, recognize a nuclear enzyme that catalyzes the relaxation of supercoiled DNA. The presence of antitopoisomerase I antibodies is highly specific to scleroderma sera. Antitopoisomerase I antibodies identify a scleroderma subset with a high frequency of diffuse skin thickening and pulmonary interstitial fibrosis.

Epitope
Antigenic determinants recognized by T and/or B cells are called *immunogenic epitopes*. Recombinant techniques of antigens can yield information crucial to uncovering T- and B-cell epitopes. Epitopes in autoimmunity can be expanded to other epitopes on the same protein or other proteins in the same macromolecular complexes, a phenomenon termed *epitope spreading*.

Recombinant Protein
Large amounts of antigen can be expressed by introducing its cloned gene into *Escherichia coli* or insect cells. Fusion proteins consisting of the amino-terminal peptides encoded by a portion of the *E. coli* β-galactosidase or a glutathione S-transferase peptide linked to eukaryotic proteins have been used often.

Scleroderma
Scleroderma is a multisystem disorder of unknown etiology characterized by abnormal deposition of collagen in the skin and internal organs, and microvascular obliteration. Circulating autoantibodies are one of the features of scleroderma.

Scleroderma is a multisystem connective tissue disease of unknown etiology characterized by abnormal deposition of collagen in the skin and internal organs, and microvascular obliteration. Several autoantibodies against cellular components are striking features of scleroderma and some are associated with certain clinical subsets. This article mainly reviews the autoimmune phenomenon against centromere antigens and topoisomerase-I, which are commonly found in the limited skin sclerosis type and the proximal type respectively. For the production of autoantibodies, molecular mimicry hypothesis and implications of apoptosis are also discussed.

1 Scleroderma and Autoantibodies

1.1 Methods of Autoantibody Detection

Sera from human autoimmune patients contain autoantibodies that react with many important cellular antigens. Detection of antinuclear antibody (ANA) is widely accepted as an important test in the diagnosis of many connective tissue diseases such as systemic lupus erythematosus, scleroderma, Sjögren's syndrome, and polymyositis. ANA testing by indirect immunofluorescence (IIF) is commonly used as a broad screen for the detection of a variety of autoimmune diseases. Initially, ANA tests by IIF methods were performed on mouse kidney or rat liver tissue sections, but lately many laboratories use HEp-2 human larynx epithelioma cancer cell line as substrate. Human cultured cells are better for the detection of human and proliferating nuclear antigens. In contrast to a few specific fluorescent patterns for particular antigens, the main patterns in IIF studies are not specific for any autoantibodies or disease entity. Characterization of the specific antibodies using double immunodiffusion assay, enzyme-linked immunosorbent assay (ELISA), or immunoblotting with appropriate reference sera is necessary. Immunoblotting with an extract from cultured cells is a very sensitive assay and permits detection of autoantibodies against soluble and insoluble antigens. This method has proved useful in a number of studies to characterize many autoantigens. Since expression cloning was first applied to isolation of autoantigens in 1985, cloned autoantigens themselves have been used for diagnostic autoantibody determination, molecular analysis of autoimmune epitopes, and molecular investigation of autoimmune processes. More recently, many clinical laboratories use commercial ELISA kits with cloned recombinant autoantigens, which are often expressed by *E. coli*. Although there might be the lack of proper folding or protein modification of some recombinant proteins obtained from prokaryotic sources, ELISA is highly sensitive, quantitative, and suitable for simultaneous analysis of a large number of serum samples, providing substantial cost and time savings.

1.2 Clinical Subsets of Scleroderma and Autoantibodies

Connective tissue disease patients have various autoantibodies against a variety of cellular components required for universal

cell functions such as cell division, DNA replication, transcription, and translation. Although the primary event that triggers production of autoantibodies is not known, more than 95% of scleroderma patients produce autoantibodies to nuclear components. Scleroderma (systemic sclerosis) is a disease characterized by collagen deposition and subsequent fibrosis in skin and internal organs. The extent of fibrosis shows considerable variation in severity, which classifies scleroderma into two distinct subtypes: the limited cutaneous form and diffuse form. In the former, skin thickening is relatively restricted to the fingers, hands, and/or face but without extension to other internal organs. In the latter, skin changes are widespread and often rapidly progressive, involving the trunk and distal and proximal portions of the extremities with internal organ involvement. In addition, some patients have overlapping features with other connective tissue diseases.

ANAs in patients with scleroderma can be mainly classified into antinuclear and antinucleolar antibodies by immunofluorescence staining patterns of tissue culture cells as substrate (Fig. 1). The former is further subdivided into several distinct nuclear staining patterns: discrete speckles (almost equal to the centromere), coarse speckles, diffusely grainy, and homogeneous. The latter is divided into speckled, homogeneous, and clumpy. With these distinctive autoantibodies, a number of targeted autoantigen molecules have been identified: DNA topoisomerase I (topo I, formerly described as Scl-70), three centromere proteins (CENP-A, CENP-B, and CENP-C), PM/Scl (polymyositis/scleroderma)-associated antigens, RNA polymerases, and fibrillarin (U3 RNP-associated 34 kDa peptide).

Among autoantibodies against these antigens, two antibodies have been well correlated with subtypes of scleroderma:

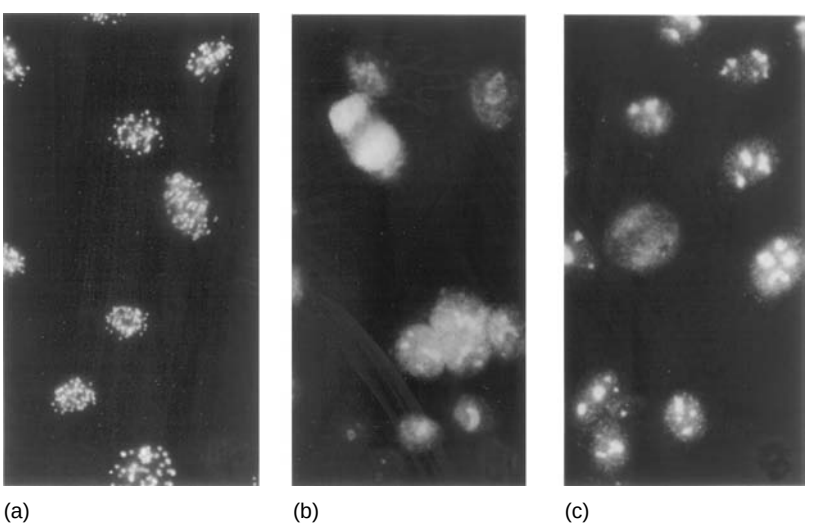

Fig. 1 Indirect immunofluorescence pictures of HEp-2 cells with scleroderma sera. (a) Anticentromere antibody. (b) Antitopoisomerase I antibody. (c) Antifibrillarin antibody.

anti-topo I and anticentromere antibodies (ACA). The former antibody has been frequently found in the patients with diffuse scleroderma, while the latter has been strongly associated with limited scleroderma. Furthermore, Fritzler et al. reported that no anti-topo I was found in more than 127 patients with connective tissue diseases other than scleroderma, while ACA was found in only less than 2% of 113 patients with connective tissue diseases other than scleroderma and Raynaud's disease. In addition, the coexistence of these two antibodies in the same patient is very rare. Therefore, both antibodies show striking disease associations and are found almost exclusively in patients with scleroderma.

Nevertheless, ACA and anti-topo I–positive patients constitute only a half or two-thirds of scleroderma patients (Table 1). The remaining one-third or another half are patients without ACA or anti-topo I. Although the latter percentage may decrease when the detection system of these autoantibodies is improved, other autoantibodies are also present in patients with distinct subsets of scleroderma, and sometimes even coexist with ACA or anti-topo I antibody. For example, some individuals with diffuse scleroderma have anti-RNA polymerase (RNA polymerase I, II, III) antibody or antifibrillarin antibody, and others in overlap with polymyositis have anti-PM/Scl antibody.

The characterization of these scleroderma subset-specific autoantibodies will be helpful in the diagnosis and prognosis of the patients with scleroderma as well as in the study on the mechanism of

Tab. 1 Human autoantigens associated with scleroderma.

Antigenic complex	*Autoantibody frequency in scleroderma [%]*	*Autoantigen*				
		Antigenic polypeptide	*Number of residues*	*B-cell major epitope*[a]	*B-cell minor epitope*[b]	*T-cell epitope*
Topoisomerase I	~50	Topoisomerase I	765	74–139 484–560 658–765		209–386 363–563
Centromere	~30	CENP-A	140	3–17 25–38		?
		CENP-B	599	1–73 538–599	438–492	?
		CENP-C	943	74–244 820–943	396–552	?
		HP1$^{HS\alpha}$	191	16–43 106–191		?
PM/Scl	3	PM/Scl-100	885[c]	166–245	672–866	?
		PM/Scl-75	355	218–355		?
U3 RNP	~10	Fibrillarin	321	1–85	149–229 276–321	?

[a,b]Major and minor epitopes are defined here as the reactive region with >50% and <50% of the corresponding antibodies, respectively.
[c]The isoform with 860 amino acid residues is also reported.

the immunization process in scleroderma autoimmunity.

1.3 Detection of Scleroderma Autoantibodies and Epitope Analysis

Detection of the scleroderma autoantibodies depends on the assay systems. For example, Kumar et al. reported that the anti-topo I sera that were positive on the immunoblot were only 71 to 86% positive in the immunodiffusion test. Similar results were obtained by Hildebrandt; anti-topo I IgG were 88% positive by ELISA with the purified topo I from calf thymus, in contrast to being 100% positive by immunoblotting. In addition, Maul et al. reported that a few anti-topo I–negative sera were positive by double diffusion with the recombinant protein blotted on the membrane. Sensitivity of ACA detection was also compared by Hildebrandt et al. among ELISA, immunoblotting, and IIF. IgG-ACA was detected to be 85% by ELISA with cloned C-terminal CENP-B antigen, 95% by immunoblotting with HeLa chromosomal proteins, and 100% by IIF with HEp-2 cells. These results indicate that among the above detection systems, immunoblotting seems to display a comparable sensitivity for the detection of these autoantibodies, even if these antigens have some conformational epitope (described below).

Epitope mapping studies have been performed more extensively with B-cell than with T-cell studies because autoantibodies (B cells) are easy to handle. Various approaches have been used for the studies; for example, deletion-type immunoblotting experiments with recombinant fragments of the autoantigen and ELISA with synthesized peptides covering the whole sequence of the autoantigen. The former method somehow detects conformational epitopes, consisting of amino acid residues from distant regions in the sequence of a protein that are spatially juxtaposed upon, while the latter method is suitable for the detection of linear epitopes in the primary structure of a protein. Since the epitope regions are renatured to some extent during protein transfer and while blocking the membrane, the immunoblot analysis can reveal the linear and conformationally reversible epitopes.

The major autoantigens associated with scleroderma are shown in Table 1: topo I, CENP-A, CENP-B, CENP-C, 75 kDa and 100 kDa antigens of PM/Scl, and fibrillarin. Epitope analyses of the epitope structure for these autoantigens at the molecular level would provide a significant insight into the interactions between the antigen and the autoantibodies. What are the trigger components leading to autoimmune response? (pathogens? or the antigens themselves?) Are there some characteristic features of structures targeted by autoantibodies? How are autoantibodies related to the disease process? We focus on the epitope structure of scleroderma autoantigens, mainly describing the analysis of centromere antigens.

2 Centromere Antigens

2.1 Autoimmunity against Centromere Antigens

The attachment site of spindle microtubules on mitotic chromosomes is called the *centromere*. The centromere has a

central role in the pairing and disjunction of daughter chromosomes in mitosis and meiosis. The human centromere at the metaphase chromosome has a trilaminar disk-shaped structure that was revealed on electromicroscopic observation, which is called the *kinetochore*.

Moroi et al. first reported ACA in the sera of patients with scleroderma in 1980. ACA is recognized as a "disease marker" antibody of limited cutaneous scleroderma, which was formerly called *CREST syndrome*. In addition, ACA are indispensable probes for the analysis of the structure of the centromere/kinetochore.

The IIF pattern by ACA is discrete speckled nucleoplasmic staining in interphase cells and strong centromeric dots decorated in the metaphase plate (Fig. 1a). Three major target autoantigens, *cen*tromere *p*rotein (CENP)-A, B, and C, have been identified with ACA-positive sera. Earnshaw et al. first isolated cDNA encoding CENP-B, the 80 kDa human centromere autoantigen, by expression cloning. Next, Saitoh et al. isolated cDNA clones that encode the human CENP-C, 140 kDa human centromere autoantigen, with immunoscreening. As for CENP-A, the 17 kDa histone-like centromere antigen, Sullivan et al. cloned its bovine cDNA using PCR with degenerated oligonucleotides derived from the amino acid sequences previously reported by Palmer et al. and isolated its human homologue. These three antigens exist in the centromere during cell cycles (Fig. 2). In contrast, the 330 kDa CENP-F antigen localizes in the centromere from metaphase

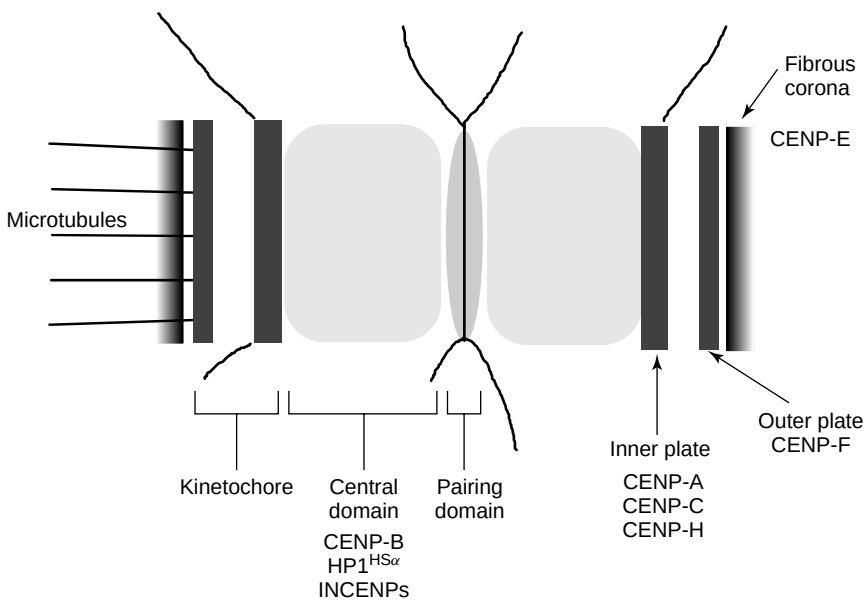

Fig. 2 Centromere domain and centromere antigens. Elements of centromeric region are depicted mainly for CENPs and autoantigens. Centromere proteins are divided into two groups: one is the constitutive proteins, for example, CENP-A, -B, -C, and -H, and another is the chromosome passenger proteins, for example, CENP-E and -F. Some proteins shown here have not been reported as autoantigens, for example, CENP-H and INCENPs.

to anaphase. Rattner et al. reported that autoantibodies to CENP-F were found mainly in patients who suffered from malignant tumors but not systemic autoimmune disease and rarely found them in "classical" ACA-positive sera, which detect CENP-A, B, and C. Its cDNA was also isolated by immunoscreening with human autoantibodies, while a cDNA that encoded human CENP-E was isolated using experimentally induced monoclonal antibodies by immunization of fractionated chromosomal proteins. CENP-H cDNA was incidentally obtained by differential display. cDNA clones of CENP-D and -G, which were also shown to be recognized by ACA, have not been obtained. Although several proteins that are not named "CENP" are known to exist in the centromere region – centromere-chromatin proteins and centromere-associated proteins – they are rarely targeted by "classical" ACA.

2.2 CENP-B and B-cell Epitopes

We start with CENP-B epitopes since its gene was cloned at first and has been most extensively studied. CENP-B has 599 amino acid residues and an apparent molecular mass of 80 kDa, larger than the predicted molecular weight of 65 kDa. This antigen was used for the demonstration of the direct interaction between the constitutive centromeric protein and centromeric satellite DNA (alphoid DNA). Alphoid DNA is a long, tandemly repeated DNA family based on a 171 bp fundamental monomer repeat unit. The alphoid DNA family in humans is composed of many subclasses, which share varying degrees of sequence homology (60–90%) and have been localized to the primary constriction (centromere) of specific chromosomes. Masumoto et al. found that the CENP-B antigen directly interacts with an alphoid DNA and they named the 17 bp motif required for the specific binding "CENP-B box." The DNA binding domain of CENP-B is located within the N-terminal 125–amino acid region containing 4 potential α-helices. On the contrary, the C-terminus of the protein from 541 to 599 amino acids forms a homodimer by itself (Fig. 3). Thus, CENP-B was presumed to contribute to the formation of the heterochromatin structure in the centromere region, binding to alphoid

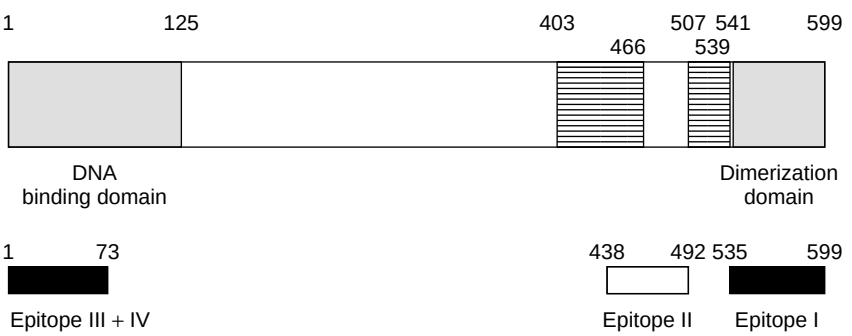

Fig. 3 Functional domains and autoepitopes of CENP-B. The top rectangle represents the full-length human CENP-B antigen. Two horizontal lined boxes show highly acidic regions. Major autoepitopes, which are recognized by more than 50% of ACA, are shown as black bars. Minor epitope, which is recognized by less than 50% of ACA, is shown as a white bar.

DNA at the N-terminus, and forming a homodimer at the C-terminus.

Epitope mapping of CENP-B were conducted to deduce the important antigenic regions for autoimmunity reactions against human centromere. Bacterial recombinant proteins expressed from a set of truncated CENP-B cDNAs were used for immunoblotting analysis. From the C-terminus to the N-terminus, four representative epitopes have been shown, epitopes I to IV. Epitopes III and IV locate in the N-terminal 73 residues corresponding to the N-terminal half of the DNA binding domain. They are overlapping since the N-terminal boundary of epitope III has not yet been precisely determined. Interestingly, the inhibitory effect of the patients' sera containing antibodies against epitopes III and/or IV on the DNA binding activity was shown *in vitro*, while no inhibitory effect was observed with ACA-positive sera without reactivities to these epitopes. Epitope III plus IV (1–73) reacts with ~75% of the ACA-positive sera. In contrast, almost all ACA sera reacted with the region spanning amino acids 535 to 599, the epitope I, which corresponds to the region required for its homodimerization. In order to restrict a minimal region of epitope I, further studies for epitope mapping were performed but no shorter reactive region in epitope I was identified. The epitope I, a major epitope of CENP-B, is used as a substrate of ELISA for ACA for clinical use.

CENP-B is a highly conserved protein through species. The similarity of the amino acid sequences of CENP-B between mouse and human are 96%. Amino acid sequences of the epitopes I, III, and IV are identical between the two species. In contrast, the epitope II (438–492), which is only reactive to the minority of ACA, is called a "human specific" epitope because the antibody to this region did not react to mouse centromere. Actually, this epitope has lower homology to amino acid sequences for CENP-B between human and mouse. This fact supported an idea that the autoantibodies were derived from stimulation of the autoantigen itself, consistent with the "antigen driven theory." In addition, the antigenic pleural sites in a single molecule also supported this theory. For ordinary antibodies' and antigens' interaction, 15 to 24 residues of the antigen participate in antibody binding. As for the epitopes I and II, we could not limit the region below the 30 to 40 amino acid length. These results supported the idea that some conformational structures in the epitopes were required for the interaction between the antibodies and antigens.

In order to confirm the importance of protein conformation for antigenicity, we tried a DNA immunoprecipitation method for quantitative assays of ACA. A ^{32}P radio-labeled 59 bp oligonucleotide probe, including the 17 bp motif of the CENP-B box, bound with cellular purified CENP-B proteins was immunoprecipitated and the radioactivity was used for the estimation of anti-CENP-B/DNA complex–antibody activity. Even though no ACAs reacted to only DNA molecules, a few sera without reactivities to CENP-B in immunoblotting methods did immunoprecipitate the CENP-B/DNA complex. Such a discrepancy of results between the DNA immunoprecipitation method and the immunoblotting as well as that relatively long stretches of amino acids were required for epitope mapping led us to an idea that the conformational structure was critical for the epitope targeted by some autoantibodies.

2.3 CENP-C and B-cell Epitopes

CENP-C is an essential component of the inner plate of the trilaminar structure of the kinetocore. It is required for maintaining proper kinetochore size and is implicated in the metaphase-to-anaphase transition in mitotic cells by stabilizing microtubule attachments. It also has a DNA binding domain and a region that contributes to the formation of the homodimer as described above for CENP-B. CENP-C has 943 amino acids and an apparent molecular mass of 140 kDa in SDS-PAGE, quite larger than the molecular weight derived from the sequence. Autoepitopes of CENP-C were mainly restricted to three regions that correspond to functional domains (Fig. 4). Epitope 1, the first major epitope, resides at the N-terminus (74–244 a.a.) that corresponds to the presumptive instability domain reported by *in vivo* analysis. Epitope 7 located at the central part (396–552 a.a.) contains the DNA binding domain, which was recently shown to bind the alphoid DNA selectively – the CENP-C region between amino acids 410 and 537. Together with CENP-B, CENP-C associates with the centromeric alphoid DNA *in vivo* and defines a highly structured organization within the human centromere. Epitope 9, the other major epitope, locates in the C-terminal region that contributes to the formation of the homodimer. This epitope spanning a 124 amino acid stretch from 476 to 599 is the second major epitope. Autoimmunity to this region was quite unique.

By immunoblotting, the reactivities between monomer and homodimer formed by chemical cross-linking were compared. The reactivities as well as the frequencies of the antibodies against the dimer form were obviously higher in both IgG and IgM responses. Interestingly, GroEL protein, which is a heat-shock chaperon protein of *E. coli*, was copurified with the C-terminal region. When purified antibodies against the dimer in a liquid phase containing GroEL were applied to various assays to detect ACA, they did not react to the recombinants but very weakly recognized cellular CENP-C by immunoblotting. Furthermore, they did show centromere staining in IIF analysis. These results suggest that these purified antibodies recognize predominantly native forms of centromere structures. Therefore, there are assumed

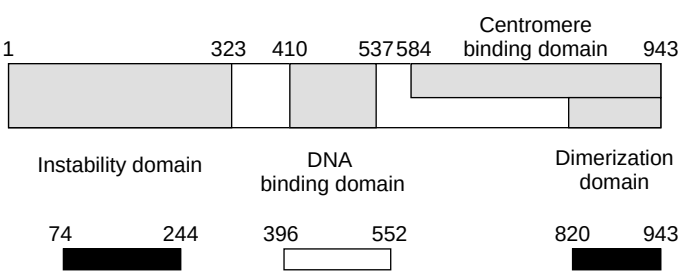

Fig. 4 Functional domains and autoepitopes of CENP-C. The top rectangle represents the full-length human CENP-C antigen. Major autoepitope, which is recognized by more than 50% of ACA, is shown as a black bar. Minor epitope, which is recognized by less than 50% of ACA, is shown as a white bar. For the epitope at the C-terminus, the dimer form is recognized by around 50% of ACA (see text).

to be three different types of epitopes on the C-terminus, each specific to a linear sequence, a three-dimensional structure newly formed by a dimer complex, and the conformational structure formed by a macromolecular complex in the presence of GroEL.

2.4 CENP-A and B-cell Epitopes

CENP-A is also a component of the inner plate of the trilaminar structure of the kinetocore. It has 140 amino acids and a molecular weight of 17 kDa (Fig. 5). Because of the presence of the region homologous to histone H3 in the C-terminal half, it is considered to function as a centromere-specific core histone. This region shares 62% identity with histone H3 and is called the *histone fold* domain. Although the N-terminal region from 1 to 48 residues is unique for CENP-A, its localization to the centromere is dependent on the histone fold domain. Nevertheless, the two-third part of the C-terminal peptide of CENP-A showed no reactivity to ACA.

The basic amino acids in the N-terminus may be involved in the protein–protein interaction with an acidic protein(s) or the DNA binding activities. Brendel et al. suggested that autoantigens generally contain more long charge runs or clusters of charged groups, which may be exposed on their surface and be strongly immunogenic, and consist of extended surface structures that are charge-, proline-, or glycine-rich. Two peptides composed of amino acid residues 3 to 17 (named peptide A) and residues 25 to 38 (peptide B) were synthesized and examined for the reactivity to ACA. Peptide A contains seven arginine (R), one lysine, and three proline (P), and two repeats of PRRRS. Peptide B has a few charged residues (two arginine and one histidine) but contains two dipeptide Pro-Ser or one Ser-Ser, leaving each of the three residues (PSXXXPSXXXSS) and two repeats of Gly-Pro-Ser. Interestingly, nearly 90% of ACA showed reactivity to both peptides. These two, epitopes 1 and 2, are the major autoepitopes. Peptide A was further used by Valdivia et al. to produce experimental ACA in rabbits and its strong antigenicity was demonstrated. Additional epitope(s) is present in the region between 39 to 52 amino acid residues, but very minor populations of ACA have reactivities.

2.5 Centromere-associated Autoantigens

Heterochromatin protein 1 (HP1) was first identified in *Drosophila* as a nonhistone chromosomal protein that functions in the establishment and maintenance of higher-order chromatin structures playing a role in chromosome segregation and gene silencing. Three related proteins homologous to *Drosophila* HP1

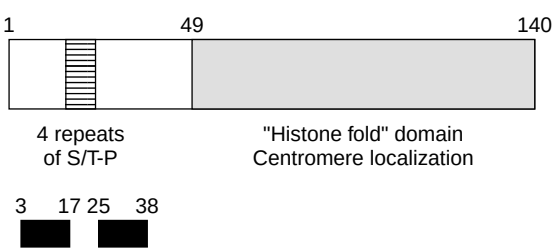

Fig. 5 Functional domains and autoepitopes of CENP-A. The top rectangle represents the full-length human CENP-A antigen. Major autoepitopes, which are recognized by more than 80% of ACA, are shown as black bars.

have been found in humans. They are closely related and contain two functionally important and conserved domains: N-terminal chromo and C-terminal chromo shadow domains. Three human homologues $HP1^{HS\alpha}$, $HP1^{HS\beta}$, and $HP1^{HS\gamma}$ are targeted by "antichromo antibodies" in a subset of ACAs. Interestingly, antichromo antibodies are always associated with ACAs. Furuta et al. reported that 20% of ACA-positive sera also had antichromo antibodies that targeted p23 ($HP1^{HS\gamma}$) and p25 (a mixture of $HP1^{HS\alpha}$ and $HP1^{HS\beta}$).

$HP1^{HS\alpha}$, which has 191 amino acids, was used for epitope mapping with immunoblotting. Regions 1 (16–43) and 2 (106–191) were found to be important for autoepitopes and in the two conserved regions: chromodomain (13–78) and chromo shadow domain (114–179). Since $HP1^{HS\alpha}$ localizes in the inner centromeric region of mitotic chromosomes and antichromo antibodies are exclusively found in ACAs, ACA is likely to recognize maculomolecular antigenic complexes. Interestingly, some ACA-positive sera were found to develop antichromo antibodies after a long period. This phenomenon implies "epitope spreading," which means that the epitope is spreading to another molecule or to domains of the same molecule.

Another autoantigen targeted by a minor subset of ACA has been characterized by ACA-positive patients. This antigen, p27, has 199 amino acids with a predicted molecular weight of 22 kDa and was nomenclatured "SSSCA1" (*S*jögren *s*yndrome/*sc*leroderma *a*utoantigen 1). Many charged residues (17 acidic and 12 basic) are present in the N-terminal region (7–83 a.a.) and two short Pro-rich stretches are present in the remaining region (113–122 and 138–145 a.a.). So far, it is not known whether p27 is a structural component such as a centromere-associated protein; however, the exclusive association between antibodies against p27 and the centromere is suggestive of a biological interaction between these antigenic molecules. Interestingly, Zhou et al. showed that p27 as well as CENP-B, fibrillarin, and RNA polymerase II was overexpressed in fibroblasts from scleroderma patients.

3 Topoisomerase I

3.1 Autoimmunity against Topoisomerase I

In the eukaryotic cell, the topological status of DNA is modulated by DNA topoisomerases. Topo I causes a single-strand break, which can allow the two sections of DNA helix on either side of the nick to rotate freely relative to each other. This enzyme plays important roles in chromatin organization, DNA replication, recombination, and transcription. It localizes in the nucleoplasm, nucleolus, and nucleolar organizing regions, giving IIF staining patterns of speckled nucleolar and densely speckled nucleoplasmic decoration (Fig. 1b). Racial differences in the prevalence and autoepitopes of anti-topo I antibodies are noted. Immunoglobulin isotypes are IgG, IgA, and IgM, and their frequencies are IgG >= IgA ≫ IgM. Isotype switch with persistence of IgG expression in longitudinal studies and the presence of all four IgG anti-topo I subclasses suggest chronic B-cell stimulation by antigen-driven mechanism.

3.2 B-cell Epitope

Topo I, a 105 kDa polypeptide, has 4 major domains: the N-terminal domain (1–200

a.a.), the globular core (200–635 a.a.), a small linker domain (636–712 a.a.), and the C-terminal domain (713–765 a.a.). The globular core and the C-terminal domains are responsible for the catalytic activity. The active site of human topo I is a tyrosine residue at position 723.

A number of studies have shown that anti-topo I antibodies recognize multiple epitopes on the entire molecule of topo I, although the ethnic backgrounds and the sites of recombinant topo I fragments used were various. Major epitopes, which are recognized by more than 80% of anti-topo I–positive patients' sera, have been determined in the central part of the molecule – the globular core domain. All these defined epitopes contain the region 484 to 560 a.a. Interestingly, residues 440 to 614 share significant structural similarity with the bacteriophage family of DNA integrases.

The N-terminal domain is highly charged and contains several nuclear-targeting signals. This domain is also antigenic and reactive to more than 70% of autoantibodies against topo I: for example, 1 to 229 a.a., recognized by 74% sera; 1 to 139 a.a., recognized by 74% sera; 74 to 248 a.a., recognized by 86% sera.

The third epitope is located at the C-terminal domain. Sixty to 70% sera reacted to the region 657(8) to 765 a.a. Since the antigenicity of truncated peptide (658–700 a.a.) was reduced to ~40% of anti-topo I antibodies, the C-terminal half of this region is likely to contribute the antigenicity of topo I. Interestingly, the eight residues (717–724 a.a.), including the active site tyrosine723, are also similar to the DNA integrase. In fact, most anti-topo I sera inhibit the relaxation activity of this enzyme *in vitro*.

3.3 T-cell Response

Autoreactive T-cell responses to topo I have been extensively studied by the group of Dr Wright. Kuwana et al. showed by peripheral blood mononuclear cell (PBMC) cultures with recombinant topo I and pokeweed mitogen that HLA-DR-restricted collaboration between topo I–specific T- and B cells is essential for the antibody production in scleroderma patients with anti-topo I antibodies. T- and B-cell epitopes are located in close proximity on the topo I molecule. Autoreactive T cells are of the CD4+ helper phenotype, recognize majorly fragments F5 and F6 (209–386 a.a. and 363–563 a.a. respectively), and express highly restricted T-cell receptor β-chain variable gene fragments ($V_\beta 20.1$). Surprisingly, these features are almost common in scleroderma patients and healthy individuals. That is, the topo I–specific T-cell responses are detected in healthy individuals with the appropriate HLA-DR alleles. T-cell clones from healthy individuals can also provide help to HLA-DR-matched patients' B cells in producing anti-topo I antibodies through a CD40-dependent mechanism.

The ability to promote anti-topo I autoantibody production is strictly correlated with IL-2 and IL-6 expression by the T-cell clones. Most topo I–specific T-cell clones express both Th1 and Th2 cytokines, and the patterns of cytokine expression are heterogeneous, supporting the concepts that the human T-cell cytokine profile is not controlled by a simple binary switch between two sets of genes, each cytokine gene expression is regulated independently, and synergy between T cells with complementary cytokine production plays a role in regulating B-cell function. According to kinetic studies, it is strongly suggested that

topo I–reactive T cells having a strong helper activity are selectively expanded in patients with anti-topo I antibody. Also interestingly, for autoreactive T-cell responses, antigen forms (native full-length vs fragmented) and cells of antigen presentation (dendritic cell vs antigen-presenting cell in PBMC = monocytes and B cells) are critical. The fact that fragmented recombinant protein strongly induces T-cell responses suggests cryptic epitopes (see Chapter 6) recognized by autoreactive T cells.

4 Scleroderma-associated Nucleolar Autoantigens

4.1 Fibrillarin

Fibrillarin is a 34 kDa protein that derives its name from its localization to both the fibrillar center and dense fibrillar component of the nucleolus. It is associated with U3 and other small nucleolar RNAs (snoRNAs) required for rRNA processing. Kasturi et al. showed, using an immunoblot with recombinant fragments, that antifibrillarin antibodies reacted primarily with epitopes present in the N- (1–80 a.a.) and C- (276–321 a.a.) domains of fibrillarin in addition to a minor epitope at an internal region (149–229 a.a.).

Autoantibodies against fibrillarin can also be induced in H-2^s mice by treatment with mercuric chloride ($HgCl_2$). Immunoprecipitation studies with truncated yeast fibrillarin shows similar recognition of a conserved conformational determinant of antifibrillarin responses in murine $HgCl_2$–induced autoimmunity and human scleroderma. However, the target of the autoantibodies is not a fibrillarin–mercury complex even though a metal–protein interaction does occur. Proteolytic cleavage of fibrillarin following xenobiotic-induced cell death can lead to the generation of novel protein fragments that obtain more appropriate immunogenic activity. These studies by Pollard et al. suggest that cryptic peptides (see Chapter 6) would serve as antigenic determinants for autoreactive T cells.

4.2 PM/Scl Antigen

A nuclear/nucleolar particle termed *PM/Scl* has been shown to immunoprecipitate a complex of 11 to 16 polypeptides ranging from 20 to 110 kDa. The PM/Scl complex is the human homologue of the yeast exosome, which is an RNA-processing complex. In immunoblotting studies, two major antigenic polypeptides have been shown – a 100 kDa antigen (PM/Scl-100) recognized by most anti-PM/Scl sera and a 75 kDa antigen (PM/Scl-75) recognized by around half the sera. According to epitope mapping of PM/Scl-75 by Alderuccio et al., the major epitope was estimated to be present within the C-terminal 160 amino acid residues. Of eight anti-PM/Scl sera, however, only two sera react with the cellular 75 kDa antigen as well as the C-terminal region. Interestingly, three additional anti-PM/Scl sera displayed a weaker reactivity with the entire PM/Scl-75 recombinant protein, suggesting the existence of another epitope in the N-terminal half. By mapping studies of PM/Scl-100 with truncated recombinant proteins, Bluthner and Bautz mapped two autoepitopes on the recombinant PM/Scl-100: one antigenic domain between amino acids 153 to 324 recognized

by all anti-PM/Scl–positive sera, another domain between amino acids 492 to 703 recognized by the minority of sera. Ge et al. identified three antigenic regions; 156 to 312, 507 to 749, and 750 to 882. Using synthetic peptides, Bluthner et al. demonstrated that a cluster of epitopes is located between amino acids 166 and 245 and that the second cluster is located between 672 and 866. Among them, a major epitope is located between 231 and 245, which is suggested to form a local α-helical secondary structure. Recently, anti-PM/Scl antibodies are shown to recognize also other components in human exosome.

5 Molecular Mimicry and Scleroderma Autoantigens

Production of some autoantibodies has been considered to be caused by cross-reactivity between viral or bacterial protein and autoantigen, which were named *molecular mimicry*. The concept of molecular mimicry was first tested in an experimental animal model of autoimmune encephalomyelitis in 1985, and since that time, substantial evidence has accumulated such that it has become the favored mechanism for causing some autoimmune diseases. If the first reaction was initiated by a foreign stimulus, the major epitope would have a similar antigenicity to the true antigen as a reminiscence of the primary immunization. However, recent epitope-mapping studies indicate that none other than the antigen molecule itself is a trigger molecule as for autoimmunity in systemic autoimmune diseases. Maul et al. reported that the sequence spanning 741 to 746 in the C-terminal region of topo I matches a sequence present in mammalian retroviral $p30^{gag}$ core nucleoprotein. However, antibodies reacting to the sequence are shown to be rare in anti-topo I autoantibodies, and topo I fragments containing this sequence were rarely recognized by peripheral blood T cells from anti-topo I–positive patients.

Because the two autoepitopes of CENP-A have been restricted to small peptide regions, its possible molecular mimicry was explored. A variety of autoantibodies has been described in hepatitis B and C for ages and we found that major epitopes of CENP-A, peptides A and B, are homologous to the core antigens of hepatitis B virus (HBV) and hepatitis C virus (HCV). Especially, the HCV core antigen contains a sequence similar to peptide B and the PRRRS sequence that repeats in peptide A. However, the corresponding regions of viruses were not generally recognized by HBV- or HCV-infected sera according to the previous epitope-mapping studies of hepatitis patients. Detailed mapping studies by Mahler et al. showed that the CENP-A major epitope motif (GPXRX) was present in Epstein–Barr nuclear antigen 1. However, the gradual development of ACA in a patient from an antihistone response was not mediated by the molecular mimicry mechanism resulting from an anti-Epstein–Barr virus response.

6 Apoptosis and Scleroderma Autoantigens

A lot of evidence suggests the involvement of biochemical events activated during apoptosis (programmed cell death), such as caspase activation in the generation of particular autoantigens targeted in systemic autoimmune diseases. Selective substrate phosphorylation/association to splicing factors has been seen to be

a crucial event for production of some autoantibodies. During apoptosis, for example, the lupus autoantigens cluster and become concentrated in the surface blebs of apoptotic cells where several of these molecules are specifically cleaved by proteases of the interleukin 1β–converting enzyme family. Lymphocytes are educated not to react against immunodominant epitopes of autoantigens and consequently autoantibodies are not produced. However, the fragmentation of specific autoantigens as a result of apoptosis can generate cryptic epitopes, which may break tolerance to autoantigens.

Casciola-Rosen et al. demonstrated that scleroderma autoantigens (topo I, the large subunit of RNA polymerase II, NOR-90) are fragmented in the presence of reactive oxygen species (ROS) and specific metals, such as copper and iron. They propose that the autoantibody production in scleroderma is induced and driven by cycles of ischemia/reperfusion generating ROS. The observation that scleroderma autoantigens are enriched in nucleoli should be noted since they colocalize with sites of high-affinity metal binding. Fibrillarin can also be fragmented and can acquire high immunogenicity during apoptosis induced by administration of mercury (see Chapter 4.1).

Although a number of autoantigens have been shown to be cleaved by caspases during apoptosis, susceptibility to cleavage by caspases is not considered to be a general feature of autoantigens due to the reasons that (1) caspases during all forms of apoptosis produce identical fragments of autoantigens and (2) several autoantigens (e.g. CENP-B, fibrillarin) cannot be cleaved by caspase. Casciola-Rosen et al. found that several autoantigens are exclusively or preferentially cleaved by Granzyme B, which is a serine protease found in the cytoplasmic granules of cytotoxic T lymphocytes and natural killer cells, and induces apoptosis in target cells during granule exocytosis–induced cytotoxicity. The function of Granzyme B is achieved partly by catalyzing the cleavage and activation of several caspases, as well as through caspase-independent pathways. Also, for scleroderma autoantigens, cleavage by Granzyme B is strongly predictive as shown in Table 2. B-cell epitopes of these antigens are believed not to be hidden in the intact molecule because

Tab. 2 Scleroderma autoantigens cleaved by granzyme B.

Autoantigens	*Cleavage site*
Topoisomerase I	IEAD^{15}-F^{a}
Fibrillarin	VGPD^{184}-G^{a}
PM/Scl-100^{b}	VEQD^{252}-M
CENP-B^{b}	VDSD^{457}-E
CENP-C	ITQD^{452}-E^{c}, VGHD^{607}-E^{c}, LESD^{621}-E^{c}, LSPD^{775}-T^{c}, LDND^{804}-E^{c}, VGQD^{888}-I^{c}
PM/Scl-75	LEND^{188}-Q^{c}, LEPD^{321}-K^{c}
p27 (SSSCA1)	LDSD^{79}-V^{c}

[a]Cleaved sites are experimentally defined by Casciola-Rosen et al.
[b]Cleavage by Granzyme B was confirmed by Casciola-Rosen et al.
[c]Possible cleaved sites according to the consensus sequences.
As for CENP-A and HP1$^{HS\alpha,\beta,\gamma}$, no consensus sequences are found.

(1) most antibodies do not preferentially recognize cleaved forms, (2) most autoepitopes are the functional sites of antigens, and (3) cleavage sites are not associated with the major epitope region (Tables 1 and 2), although we cannot speculate on conformational change induced by cryptic peptides. Although the preferential autoantibodies against Granzyme B-cleaved CENP-C in scleroderma patients with ischemic digital loss were recently shown, the theory that Granzyme B-generated fragments serve as a source of T-cell antigenic determinants is more relevant as shown in the study of topo I (see Chapter 3.3).

Investigations of early-involvement lesions in scleroderma suggest that apoptosis of endothelial cells, possibly caused by cytotoxic antiendothelial cell antibodies, may occur before the infiltration of lymphocytes. Although a direct role for the antibodies inducing apoptosis *in vivo* remains to be elucidated, the recent report by Lunardi et al. showed that antibodies in scleroderma patients, reacting with the human cytomegalovirus late protein UL94, cause apoptosis of endothelial cells through specific interaction with the cell-surface integrin–NAG-2 protein complex. Other than protein cleavage described above, recent studies showed that, in the case of infection of herpes simplex virus type 1, viral immediate-early protein ICP0 induced degradation of CENP-A and CENP-C by a proteasome-dependent manner. This is one of the cases in which viral infection can induce an altered autoantigen. In order to clarify the pathogenesis of scleroderma, further studies especially to reveal mechanisms of exposure of a cryptic epitope and its reactive T-cell activation are needed.

See also Antigen Presenting Cells (APCs).

Bibliography

Books and Reviews

Black, C.M., Denton, C.P. (1998) Scleroderma and Related Disorders in Adults and Children, in: Maddison, P.J., Isenberg, D.A., Woo, P., Glass, D.N. (Eds.) *Oxford Textbook of Rheumatology*, Oxford University Press, Oxford, pp. 1217–1247.

Harvey, G.R., McHugh, N.J. (1999) Serologic abnormalities in systemic sclerosis, *Curr. Opin. Rheumatol.* **11**, 495–502.

Pollard, K.M. (2002) Cell death, autoantigen cleavage, and autoimmunity, *Arthritis Rheum.* **46**, 1699–1702.

Rosen, A., Casciola-Rosen, L. (1999) Autoantigens as substrates for apoptotic proteases: implications for the pathogenesis of systemic autoimmune disease, *Cell Death Differ.* **6**, 6–12.

Rothfield, N.F. (1996) Autoantibodies to Scleroderma-associated Antigens, in: Venrooij, W.J., Maini, R.N. (Eds.) *Manual of Biological Markers of Disease*, Kluwer, Dordrecht, C5.1, pp. 1–7.

Verheijen, R. (1992) B-cell epitopes of scleroderma-specific autoantigens, *Mol. Biol. Rep.* **16**, 183–189.

von Muhlen, C.A., Tan, E.M. (1995) Autoantibodies in the diagnosis of systemic rheumatic diseases, *Semin. Arthritis Rheum.* **24**, 323–358.

Wucherpfennig, K.W. (2001) Mechanisms for the induction of autoimmunity by infectious agents, *J. Clin. Invest.* **108**, 1097–1104.

Primary Literature

Alderuccio, F., Chan, E.K.L., Tan, E.M. (1991) Molecular characterization of an autoantigen of PM-Scl in the polymyositis/scleroderma overlap syndrome: a unique and complete human cDNA encoding an apparent 75-kD acidic protein of the nucleolar complex, *J. Exp. Med.* **173**, 941–952.

Bernstein, R.M., Steigerwald, J.C., Tan, E.M. (1982) Association of antinuclear and antinucleolar antibodies in progressive

systemic sclerosis, *Clin. Exp. Immunol.* **48**, 43–51.

Bluthner, M., Bautz, F.A. (1992) Cloning and characterization of the cDNA coding for a polymyositis-scleroderma overlap syndrome-related nucleolar 100-kD protein, *J. Exp. Med.* **176**, 973–980.

Brendel, V., Dohlman, J., Blaisdell, B.E., Karlin, S. (1991) Very long charge runs in systemic lupus erythematosus-associated autoantigens, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1536–1540.

D'Arpa, P., Machlin, P.S., Ratrie, H., Rothfield, N., Cleveland, D.W., Earnshaw, W.C. (1988) cDNA cloning of human DNA topoisomerase I: Catalytic activity of a 67.7-kDa carboxy-terminal fragment, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2543–2547.

Earnshaw, W.C., Bordwell, B., Marino, C., Rothfield, N. (1986) Three human chromosomal autoantigens are recognized by sera from patients with anti-centromere antibodies, *J. Clin. Invest.* **77**, 426–430.

Earnshaw, W.C., Machlin, P.S., Bordwell, B.J., Rothfield, N.F., Cleveland, D.W. (1987) Analysis of anticentromere autoantibodies using cloned autoantigen CENP-B, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4979–4983.

Earnshaw, W.C., Sullivan, K.F., Machlin, P.S., Cooke, C.A., Kaiser, D.A., Pollard, T.D., Rothfield, N.F., Cleveland, D.W. (1987) Molecular cloning of cDNA for CENP-B, the major human centromere autoantigen, *J. Cell Biol.* **104**, 817–829.

Fritzler, M.J., Kinsella, T.D., Garbutt, E. (1980) The CREST syndrome: a distinct serologic entity with anticentromere antibodies, *Am. J. Med.* **69**, 520–526.

Fujinami, R.S., Oldstone, M.B. (1985) Amino acid homology between the encephalitogenic site of myelin basic protein and virus: mechanism for autoimmunity, *Science* **230**, 1043–1045.

Ge, Q., Wu, Y., James, J.A., Targoff, I.N. (1996) Epitope analysis of the major reactive region of the 100-kd protein of PM-Scl autoantigen, *Arthritis Rheum.* **39**, 1588–1595.

Hildebrandt, S., Weiner, E., Senecal, J.-L., Noell, S., Daniels, L., Earnshaw, W.C., Rothfield, N.F. (1990) The IgG, IgM, and IgA isotypes of anti-topoisomerase I and anticentromere autoantibodies, *Arthritis Rheum.* **33**, 724–727.

Kumar, V., Kowalewski, C., Koelle, M., Qutaishat, S., Chorzelski, T.P., Beutner, E.H., Jarzabek-Chorzelska, M., Kolacinska, Z., Jablonska, S. (1988) Scl-70 antigen stability and its effect on antibody detection in scleroderma, *J. Rheumatol.* **15**, 1499–1502.

Masumoto, H., Masukata, H., Muro, Y., Nozaki, N., Okazaki, T. (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite, *J. Cell. Biol.* **109**, 1963–1973.

Masumoto, H., Sugimoto, K., Okazaki, T. (1989) Alphoid satellite DNA is tightly associated with centromere antigens in human chromosomes throughout the cell cycle, *Exp. Cell Res.* **181**, 181–196.

Maul, G.G., Jimenez, S.A., Riggs, E., Ziemnicka-Kotula, D. (1989) Determination of an epitope of the diffuse systemic sclerosis marker antigen DNA topoisomerase I: Sequence similarity with retroviral p30gag protein suggests a possible cause for autoimmunity in systemic sclerosis, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8492–8496.

Meesters, T.M., Hoet, M., van den Hoogen, F.H., Verheijen, R., Habets, W.J., van Venrooij, W.J. (1992) Analysis of an immunodominant epitope of topoisomerase I in patients with systemic sclerosis, *Mol. Biol. Rep.* **16**, 117–123.

Moroi, Y., Peebles, C., Fritzler, M.J., Steigerwald, J., Tan, E.M. (1980), Antibody to centromere (kinetochore) in scleroderma sera, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1627–1631.

Muro, Y., Masumoto, H., Okazaki, T., Ohashi, M. (1991) Purification of a human centromere antigen (CENP-B) and application of DNA immunoprecipitation to quantitative assay for anti-CENP-B antibodies, *J. Invest. Dermatol.* **97**, 378–380.

Muro, Y., Masumoto, H., Yoda, K., Nozaki, N., Ohashi, M., Okazaki, T. (1992) Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box, *J. Cell Biol.* **116**, 585–596.

Muro, Y., Sugimoto, K., Okazaki, T., Ohashi, M. (1990) The heterogeneity of anticentromere antibodies in immunoblotting analysis, *J. Rheumatol.* **17**, 1042–1047.

Reimer, G., Scheer, U., Peters, J.-M., Tan, E.M. (1986) Immunolocalization and partial characterization of a nucleolar autoantigen (PM-Scl) associated with polymyositis/scleroderma

overlap syndromes, *J. Immunol.* **137**, 3802–3808.

Reimer, G., Steen, V.D., Penning, C.A., Medsger, T.A., Tan, E.M. (1988) Correlates between autoantibodies to nucleolar antigens and clinical features in patients with systemic sclerosis (scleroderma), *Arthritis Rheum.* **31**, 525–532.

Saitoh, H., Tomkiel, J., Cooke, C.A., Ratrie, H. III, Maurer, M., Rothfield, N.F., Earnshaw, W.C. (1992) CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochore plate, *Cell* **70**, 115–125.

Saunders, W.S., Chue, C., Goebl, M., Craig, C., Clark, R.F., Powers, J.A., Eissenberg, J.C., Elgin, S.C.R., Rothfield, N.F., Earnshaw, W.C. (1993) Molecular cloning of a human homologue of *Drosophila* heterochromatin protein HP1 using anti-centromere autoantibodies with anti-chromo specificity, *J. Cell Sci.* **104**, 573–582.

Steen, V.D., Powell, D.L., Medsger, T.A. Jr. (1988) Clinical correlations and prognosis based on serum autoantibodies in patients with systemic sclerosis, *Arthritis Rheum.* **31**, 196–203.

Sugimoto, K., Migita, H., Hagishita, Y., Yata, H., Himeno, M. (1992) An antigenic determinant on human centromere protein B (CENP-B) available for production of human-specific anticentromere antibodies in mouse, *Cell Struct. Funct.* **17**, 129–138.

Sugimoto, K., Muro, Y., Himeno, M. (1992) Anti-helix-loop-helix domain antibodies: discovery of autoantibodies that inhibit DNA binding activity of human centromere protein B (CENP-B), *J. Biochem.* **111**, 478–483.

Sullivan, K.F., Glass, C.A. (1991) CENP-B is a highly conserved mammalian centromere protein with homology to the helix-loop-helix family of proteins, *Chromosoma* **100**, 360–370.

Verheijen, R., van den Hoogen, F., Beijer, R., Richter, A., Penner, E., Habets, W.J., van Venrooij, W.J. (1990) A recombinant topoisomerase I used for autoantibody detection in sera from patients with systemic sclerosis, *Clin. Exp. Immunol.* **80**, 38–43.

Recent Key Primary Literature

Bluthner, M., Mahler, M., Muller, D.B., Dunzl, H., Bautz, F.A. (2000) Identification of an α-helical epitope region on the PM/Scl-100 autoantigen with structural homology to a region on the heterochromatin p25β autoantigen using immobilized overlapping synthetic peptides, *J. Mol. Med.* **78**, 47–54.

Bordron, A., Dueymes, M., Levy, Y., Jamin, C., Leroy, J.P., Piette, J.C., Shoenfeld, Y., Youinou, P.Y. (1998) The binding of some human antiendothelial cell antibodies induces endothelial cell apoptosis, *J. Clin. Invest.* **101**, 2029–2035.

Brouwer, R., Vree Egberts, W.T.M., Hengstman, G.J., Raijmakers, R., van Engelen, B.G., Seelig, H.P., Renz, M., Mierau, R., Genth, E., Pruijn, G.J., van Venrooij, W.J. (2002) Autoantibodies directed to novel components of the PM/Scl complex, the human exosome, *Arthritis Res.* **4**, 134–138.

Casciola-Rosen, L., Andrade, F., Ulanet, D., Wong, W.B., Rosen, A. (1999) Cleavage by granzyme B is strongly predictive of autoantigen status: implications for initiation of autoimmunity, *J. Exp. Med.* **190**, 815–826.

Casciola-Rosen, L., Wigley, F., Rosen, A. (1997) Scleroderma autoantigens are uniquely fragmented by metal-catalyzed oxidation reactions: implications for pathogenesis, *J. Exp. Med.* **185**, 71–79.

Casiano, C.A., Martin, S.J., Green, D.R., Tan, E.M. (1996) Selective cleavage of nuclear autoantigens during CD95 (Fas/APO-1)-mediated T cell apoptosis, *J. Exp. Med.* **184**, 765–770.

Everett, R.D., Earnshaw, W.C., Findlay, J., Lomonte, P. (1999) Specific destruction of kinetochore protein CENP-C and disruption of cell division by herpes simplex virus immediate-early protein Vmw110, *EMBO J.* **18**, 1526–1538.

Everett, R.D., Earnshaw, W.C., Pluta, A.F., Sternsdorf, T., Ainsztein, A.M., Carmena, M., Ruchaud, S., Hsu, W.L., Orr, A. (1999) A dynamic connection between centromeres and ND10 proteins, *J. Cell Sci.* **112**, 3443–3454.

Furuta, K., Hildebrandt, B., Matsuoka, S., Kiyosawa, K., Reimer, G., Luderschmidt, C., Chan, E.K., Tan, E.M. (1998) Immunological characterization of heterochromatin protein p25beta autoantibodies and relationship with centromere autoantibodies and pulmonary fibrosis in systemic scleroderma, *J. Mol. Med.* **76**, 54–60.

Hayashi, Y., Muro, Y., Kuriyama, K., Tomita, Y., Sugimoto, K. (2000) Differences in specificities of anti-centromere sera for the monomeric and dimeric C-terminal peptides of human

centromere protein C, *Int. Immunol.* **12**, 1431–1437.

Henry, P.A., Atamas, S.P., Yurovsky, V.V., Luzina, I., Wigley, F.M., White, B. (2000) Diversity and plasticity of the anti-DNA topoisomerase I autoantibody response in scleroderma, *Arthritis Rheum.* **43**, 2733–2742.

Iwai, T., Muro, Y., Sugimoto, K., Matsumoto, Y., Ohashi, M. (1996) Clinical features of anti-chromo antibodies associated with anti-centromere antibodies, *Clin. Exp. Immunol.* **105**, 285–290.

Kasturi, K.N., Hatakeyama, A., Spiera, H., Bona, C.A. (1995) Antifibrillarin autoantibodies present in systemic sclerosis and other connective tissue diseases interact with similar epitopes, *J. Exp. Med.* **181**, 1027–1036.

Kato, T., Yamamoto, K., Takeuchi, H., Okubo, M., Hara, E., Nakada, S., Oda, K., Ito, K., Nishioka, K. (1993) Identification of a universal B cell epitope on DNA topoisomerase I, an autoantigen associated with scleroderma, *Arthritis Rheum.* **36**, 1580–1587.

Kuwana, M., Feghali, C.A., Medsger, T.A. Jr., Wright, T.M. (2001) Autoreactive T cells to topoisomerase I in monozygotic twins discordant for systemic sclerosis, *Arthritis Rheum.* **44**, 1654–1659.

Kuwana, M., Kaburaki, J., Arnett, F.C., Howard, R.F., Medsger, T.A. Jr., Wright, T.M. (1999) Influence of ethnic background on clinical and serologic features in patients with systemic sclerosis and anti-DNA topoisomerase I antibody, *Arthritis Rheum.* **42**, 465–474.

Kuwana, M., Kaburaki, J., Medsger, T.A. Jr., Wright, T.M. (1999) An immunodominant epitope on DNA topoisomerase I is conformational in nature: heterogeneity in its recognition by systemic sclerosis sera, *Arthritis Rheum.* **42**, 1179–1188.

Kuwana, M., Kaburaki, J., Mimori, T., Kawakami, Y., Tojo, T. (2000) Longitudinal analysis of autoantibody response to topoisomerase I in systemic sclerosis, *Arthritis Rheum.* **43**, 1074–1084.

Kuwana, M., Kaburaki, J., Mimori, T., Tojo, T., Homma, M. (1993) Autoantigenic epitopes on DNA topoisomerase I. clinical and immunogenetic associations in systemic sclerosis, *Arthritis Rheum.* **36**, 1406–1413.

Kuwana, M., Medsger, T.A. Jr., Wright, T.M. (1995) T and B cell collaboration is essential for the autoantibody response to DNA topoisomerase I in systemic sclerosis, *J. Immunol.* **155**, 2703–2714.

Kuwana, M., Medsger, T.A. Jr., Wright, T.M. (1995) T cell proliferative response induced by DNA topoisomerase I in patients with systemic sclerosis and healthy donors, *J. Clin. Invest.* **96**, 586–596.

Kuwana, M., Medsger, T.A. Jr., Wright, T.M. (1997) Highly restricted TCR-alpha beta usage by autoreactive human T cell clones specific for DNA topoisomerase I: recognition of an immunodominant epitope, *J. Immunol.* **158**, 485–491.

Kuwana, M., Medsger, T.A. Jr., Wright, T.M. (2000) Analysis of soluble and cell surface factors regulating anti-DNA topoisomerase I autoantibody production demonstrates synergy between Th1 and Th2 autoreactive T cells, *J. Immunol.* **164**, 6138–6146.

Liao, H., Winkfein, R.J., Mack, G., Rattner, J.B., Yen, T.J. (1995) CENP-F is a protein of the nuclear matrix that assembles onto kinetochores at late G2 and is rapidly degraded after mitosis, *J. Cell Biol.* **130**, 507–518.

Lomonte, P., Everett, R.D. (1999) Herpes simplex virus type 1 immediate-early protein Vmw110 inhibits progression of cells through mitosis and from G(1) into S phase of the cell cycle, *J. Virol.* **73**, 9456–9467.

Lomonte, P., Sullivan, K.F., Everett, R.D. (2001) Degradation of nucleosome-associated centromeric histone H3-like protein CENP-A induced by herpes simplex virus type 1 protein ICP0, *J. Biol. Chem.* **276**, 5829–5835.

Lunardi, C., Bason, C., Navone, R., Millo, E., Damonte, G., Corrocher, R., Puccetti, A. (2000) Systemic sclerosis immunoglobulin G autoantibodies bind the human cytomegalovirus late protein UL94 and induce apoptosis in human endothelial cells, *Nat. Med.* **6**, 1183–1186.

Mahler, M., Mierau, R., Bluthner, M. (2000) Fine-specificity of the anti-CENP-A B-cell autoimmune response, *J. Mol. Med.* **78**, 460–467.

Mahler, M., Mierau, R., Genth, E., Bluthner, M. (2002) Development of a CENP-A/CENP-B-specific immune response in a patient with systemic sclerosis, *Arthritis Rheum.* **46**, 1866–1872.

McHugh, N.J. (1995) Frequency of autoantibodies to a major epitope on the carboxyl terminal fragment of CENP-B in patients with autoimmune disease, *Br. J. Rheumatol.* **34**, 407–412.

Muro, Y., Azuma, N., Onouchi, H., Kunimatsu, M., Tomita, Y., Sasaki, M., Sugimoto, K. (2000) Autoepitopes on autoantigen CENP-A are restricted to the N-terminal region, which has no homology with histone H3, *Clin. Exp. Immunol.* **120**, 218–223.

Muro, Y., Yamada, T., Himeno, M., Sugimoto, K. (1998) cDNA cloning of a novel autoantigen targeted by a minor subset of anticentromere antibodies, *Clin. Exp. Immunol.* **111**, 372–376.

Muro, Y., Yamada, T., Iwai, T., Sugimoto, K. (1996) Epitope analysis of chromo antigen and clinical features in a subset of patients with anti-centromere antibodies, *Mol. Biol. Rep.* **23**, 147–151.

Oriss, T.B., Hu, P.Q., Wright, T.M. (2001) Distinct autoreactive T cell responses to native and fragmented DNA topoisomerase I: influence of APC type and IL-2, *J. Immunol.* **166**, 5456–5463.

Parkinson, J., Everett, R.D. (2000) Alpha herpes virus proteins related to herpes simplex virus type 1 ICP0 affect cellular structures and proteins, *J. Virol.* **74**, 10006–10017.

Pollard, K.M., Lee, D.K., Casiano, C.A., Bluthner, M., Johnston, M.M., Tan, E.M. (1997) The autoimmunity-inducing xenobiotic mercury interacts with the autoantigen fibrillarin and modifies its molecular and antigenic properties, *J. Immunol.* **158**, 3521–3528.

Pollard, K.M., Pearson, D.L., Bluthner, M., Tan, E.M. (2000) Proteolytic cleavage of a self-antigen following xenobiotic-induced cell death produces a fragment with novel immunogenic properties, *J. Immunol.* **165**, 2263–2270.

Rattner, J.B., Rees, J., Arnett, F.C., Reveille, J.D., Goldstein, R., Fritzler, M.J. (1996) The centromere kinesin-like protein, CENP-E. an autoantigen in systemic sclerosis, *Arthritis Rheum.* **39**, 1355–1361.

Rattner, J.B., Rees, J., Whitehead, C.M., Casiano, C.A., Tan, E.M., Humbel, R.L., Conrad, K., Fritzler, M.J. (1997) High frequency of neoplasia in patients with autoantibodies to centromere protein CENP-F, *Clin. Invest. Med.* **20**, 308–319.

Redinbo, M.R., Stewart, L., Kuhn, P., Champoux, J.J., Hol, W.G. (1998) Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA, *Science* **279**, 1504–1513.

Reveille, J.D., Fischbach, M., McNearney, T., Friedman, A.W., Aguilar, M.B., Lisse, J., Fritzler, M.J., Ahn, C., Arnett, F.C. (2001) Systemic sclerosis in 3 US ethnic groups: a comparison of clinical, sociodemographic, serologic, and immunogenetic determinants, *Semin. Arthritis Rheum.* **30**, 332–346.

Schachna, L., Wigley, F.M., Morris, S., Gelber, A.C., Rosen, A., Casciola-Rosen, L. (2002) Recognition of Granzyme B-generated autoantigen fragments in scleroderma patients with ischemic digital loss, *Arthritis Rheum.* **46**, 1873–1884.

Sgonc, R., Gruschwitz, M.S., Boeck, G., Sepp, N., Gruber, J., Wick, G. (2000) Endothelial cell apoptosis in systemic sclerosis is induced by antibody-dependent cell-mediated cytotoxicity via CD95, *Arthritis Rheum.* **43**, 2550–2562.

Sgonc, R., Gruschwitz, M.S., Dietrich, H., Recheis, H., Gershwin, M.E., Wick, G. (1996) Endothelial cell apoptosis is a primary pathogenetic event underlying skin lesions in avian and human scleroderma, *J. Clin. Invest.* **98**, 785–792.

Soriano, E., Whyte, J., McHugh, N.J. (1994) Frequency and clinical association of anti-chromo antibodies in connective tissue disease, *Ann. Rheumatic Dis.* **53**, 666–670.

Sugimoto, K., Fukuda, R., Himeno, M. (2000) Centromere/kinetochore localization of human centromere protein A (CENP-A) exogenously expressed as a fusion to green fluorescent protein, *Cell Struct. Funct.* **25**, 253–261.

Sugimoto, K., Kuriyama, K., Himeno, M., Muro, Y. (1998) Epitope mapping of human centromere autoantigen CENP-C; heterogeneity of anti-CENP-C response in rheumatic diseases, *J. Rheumatol.* **25**, 474–481.

Sugimoto, K., Kuriyama, K., Shibata, A., Himeno, M. (1997) Characterization of internal DNA-binding and C-terminal dimerization domains of human centromere/kinetochore autoantigen CENP-C *in vitro*: role of DNA-binding and self-associating activities in kinetochore organization, *Chromosome Res.* **5**, 132–141.

Sullivan, K.F., Hechenberger, M., Masri, K. (1994) Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere, *J. Cell Biol.* **127**, 581–592.

Takeuchi, K., Turley, S.J., Tan, E.M., Pollard, K.M. (1995) Analysis of the autoantibody response to fibrillarin in human disease and murine models of autoimmunity, *J. Immunol.* **154**, 961–971.

Valdivia, M.M., Figueroa, J., Iglesias, C., Ortiz, M. (1998) A novel centromere monospecific serum to a human autoepitope on the histone H3-like protein CENP-A, *FEBS Lett.* **422**, 5–9.

Vazquez-Abad, D., Monteon, V., Senecal, J.L., Walsh, S., Rothfield, N. (1997) Analysis of IgG subclasses of human antitopoisomerase I autoantibodies suggests chronic B cell stimulation, *Clin. Immunol. Immunopathol.* **84**, 65–72.

Vazquez-Abad, D., Russell, C.A., Cusick, S.M., Earnshaw, W.C., Rothfield, N.F. (1995) Longitudinal study of anticentromere and antitopoisomerase-I isotypes, *Clin. Immunol. Immunopathol.* **74**, 257–270.

Vazquez-Abad, D., Tian, L., Zanetti, M., Rothfield, N.F. (1997) A cross-reactive idiotype in scleroderma, *Clin. Exp. Immunol.* **108**, 420–427.

Zhou, X., Tan, F.K., Xiong, M., Milewicz, D.M., Feghali, C.A., Fritzler, M.J., Reveille, J.D., Arnett, F.C. (2001) Systemic sclerosis (scleroderma): specific autoantigen genes are selectively overexpressed in scleroderma fibroblasts, *J. Immunol.* **167**, 7126–7133.

Bacillus anthracis (Anthrax), Molecular Biology of: *see* Anthrax (*Bacillus anthracis*), Molecular Biology of

Bacterial Cell Culture Methods

Andrew S. Ball
Department of Biological Sciences, University of Essex, Colchester, UK

1 Cultivation of Bacteria 505
1.1 Introduction 505
1.2 Cultures 505
1.3 Culture Incubations 506
1.4 Types of Media Used for Growing Selected Bacteria 506

2 Methods of Bacterial Isolation 507
2.1 Sterilization 507
2.2 Purification of Bacteria from Contaminated Cultures 507
2.3 Isolation of Bacteria from Natural Sources 508
2.4 Use of Selective Media 508
2.5 Methods for Anaerobic Cultures 509

3 Microscopic Examination of Bacteria 510
3.1 Introduction 510
3.2 The Examination of Living Unstained Bacteria 510
3.3 The Measurement of Bacteria 510
3.4 Methods of Staining Bacteria 510

4 Estimation of Population Sizes of Bacteria 512
4.1 Bacterial Biomass Measurement 512
4.2 Turbidometric Estimations 512
4.3 Theory of Bacterial Growth 513
4.4 Counting Methods 513

5 Characterization of Bacteria Using Biochemical and Cultural Tests 516
5.1 Introduction 516
5.1.1 Minimal Growth Requirements and Substrate Utilization 517
5.1.2 Detection of Metabolic End Products 517

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

5.1.3 Biochemical Examination of Anaerobic Bacteria 520
5.1.4 Rapid Testing Methods 520
5.2 Cultural Characteristics 521

Bibliography 522
Appendix 1 Sources of Media and Supplies 522

Keywords

Bacterial Isolation
This is he separation from a sample of one or more bacterial species. The basis of the separation is ability of these isolates to grown under selective conditions. This selectivity can be physical, chemical or biological in nature.

Bacterial Culture
A particular strain or kind of bacteria growing in a laboratory medium. It is usually assumed that the bacteria is growing in pure culture, without contamination.

Microscopy
The observation of bacteria requires the use of microscopy. A number of methods can be used to enhance the appearance of bacteria under the microscope. Simple or complex stains are widely used.

Population Estimation
Determining the numbers of bacteria present in a sample is often required. A number of techniques can be used including direct microscopic counts, dilution plate counts, absorbance and most probable numbers. Each can give useful estimates of population increases and hence growth rates.

Biochemical Characterization
Microscopic observations rarely allow bacterial identification to the genus level. The metabolic activities and other characteristics of the bacteria are also required. Biochemical characterization of the bacteria assits in the identification of the organism.

This chapter provides essential information to help the researches in the successful culturing of a wide range of bacteria. It includes guidance on the isolation, culturing, microscopic examination, characterization and identification of bacteria.

1 Cultivation of Bacteria

1.1 Introduction

Much of microbiological work depends upon the maintenance of pure cultures of microorganisms. Aseptic technique is largely a matter of commonsense, but the following points should be borne in mind:

- Clean air contains many bacterial and fungal spores, carried on dust particles or in water droplets; any surface exposed to the air quickly becomes contaminated. Instruments that can be sterilized by heating in a Bunsen flame, for example, inoculating loops, can be left exposed, but must be flamed thoroughly before use. Items that cannot be treated in this way, for example, pipettes should be sterilized in wrappings or containers.
- Microbial contamination in the laboratory is most often due to currents of unsterile air. The chief merit of inoculation cabinets therefore lies in the protection they give from draughts. This protection can be supplemented by keeping all windows and doors shut and by cutting down personal movement within the laboratory.

Prior to identification it is often necessary to obtain pure cultures of bacteria growing on plates containing a suitable nutrient medium. Bacteria display many variations of the major nutritional requirements and so the utilizable sources of energy also vary. However, the media for the cultivation of bacteria must contain (1) water; (2) some source of energy, usually as a carbon compound; (3) other essential elements such as N, S, P, K, Na, Mg, Ca, Fe and Cl in balanced proportions. Some bacteria also require vitamins. In preparing the media, chemicals of defined composition are preferred to those of undefined composition. However, since the nutritional patterns of all bacteria are not sufficiently well known, it is often necessary to use substances of unknown composition, for example, peptone. In addition, media are often solidified with complex substances like agar. Most chemicals should be dissolved in distilled or demineralized water. Tap water may contain toxic elements, particularly copper.

Most bacteria can only grow within a restricted pH range, and are usually intolerant of acid conditions. The reaction of the medium must be adjusted so that the final pH after sterilization is between 6.8 and 7.2, unless a different reaction is needed for some special purpose. If a medium contains substances that may become more acid during sterilization, it is usually adjusted to a pH value about 0.2 greater than that required, that is, 7.0 to 7.4. Buffers are often required to facilitate the growth. This is particularly true of media composed of simple compounds or media in which acid-producing bacteria are cultivated.

1.2 Cultures

Liquid media (broths) are distributed in flasks or tubes plugged with nonabsorbent cotton wool or plastic caps. Liquid medium can be solidified with agar (1.0–2.0%). Agar is an extract of seaweed and is a sulfate ester of a linear galactan, though contaminated with other organic impurities. Generally, agar can be regarded as inert. Agar gels liquefy between 96 and

100 °C, but do not set until their temperature has dropped to 42 °C.

Occasionally, it may be necessary to exclude all organic material (including agar) from the media. In this case, silica gel can be used to solidify the medium. $Na_2SiO_3 \cdot 9H_2O$ (13.4 g) is dissolved in 100 mL distilled water. The pH must be adjusted to 10.0 using cation-exchange resin AG50 WX-8 (Bio-Rad Laboratories) under constant mixing. After 30 min, the pH of the resin must be readjusted to 10.0. The resin must then be removed by filtration using Whatman No. 1 filter paper. After the addition of appropriate nutrients for bacterial growth, sterilization is achieved by filtration through a sterile membrane filter (0.45 μm). The pH of the solution is then adjusted to 7.0 using phosphoric acid (5 M). Finally the medium is poured and allowed to gel (40 min at room temperature).

Solid media can be used in a variety of ways:

Slope cultures. Tubes or small bottles containing 6 mL of medium are heated to melt the agar. This is allowed to set in a sloping position, thus increasing the surface area of the medium for the growth of bacteria.

Deep stab cultures. Tubes or bottles containing 10 to 15 mL of solid medium are heated and the molten medium is allowed to cool to 45 °C. While still liquid, the medium is inoculated, mixed well by rotating the tube between the hands, and then allowed to set.

Plate cultures. About 15 to 20 mL of solid medium are needed for each 9-cm petri dish. The medium is melted and cooled to 45 °C and poured into the dish. Hot media must never be poured since water may condense inside the lid. The medium can be inoculated in several ways:

1. The inoculum is first placed in the petri dish and then the medium is poured over. To ensure even distribution of the inoculum, the plate is moved gently with a rotating movement. This method is used when all the bacteria in a measured inoculum are required to grow into separate colonies, for example, a dilution-plate count.
2. Alternatively, the agar can be inoculated by the poured plate or the streak plate methods, which are described below.

1.3 Culture Incubations

Cultures should be incubated at the optimum temperature for growth or for the specific activity being studied. Pathogens and commensal species grow best at body temperature, that is, 37 °C. Soil organisms and plant pathogens are incubated at 20 to 25 °C.

1.4 Types of Media Used for Growing Selected Bacteria

Bacteriological media are frequently classified as complex or defined according to their chemical composition. A complex medium is one that contains such ill-defined substances as yeast or beef extract, while a defined medium can have 10 to 20 different chemicals present, but the identity of each is known. All of the most widely used media are available from a range of supplies, some of which are listed in Appendix 1.

2 Methods of Bacterial Isolation

2.1 Sterilization

Media are usually sterilized after distribution into bottles, tubes, or flasks. Sterile media may be transferred aseptically into previously sterilized containers. Table 1 lists the methods commonly used for the sterilization of media.

2.2 Purification of Bacteria from Contaminated Cultures

Progressive dilution is the basis of isolation techniques. This can be applied as follows:

Poured plate method

- Melt and cool to 45 °C three tubes of an agar medium. Inoculate one tube with one loopful of the suspension and mix well by rotating it between the hands.
- Flame the loop thoroughly and transfer one loopful of the mixture from the first tube to the second. Mix again.
- Repeat for the third tube. Pour the contents of all the tubes into the sterile petri dishes and allow the contents to set.
- The second or third plate should show well-separated colonies after incubation. Distinct colonies can then be picked off and examined. If the suspension is very heavy, initial dilution in sterile saline may have to be made.

The streak plate method. This is a quicker, though less reliable method.

- Prepare a plate of solid medium. Dry the plates for 30 min before use by placing them open with the open sides facing downwards in a 45 °C incubator.
- Mark the plate on the underside into four sectors. With a sterile loop, place a small drop of suspension on the agar at one side of a marked sector near the edge of the dish. Draw this out in a single broad tangential streak.
- Flame the loop, cool it by touching the agar at the end of the dish away from the inoculum and draw a set of

Tab. 1 Common methods of sterilization.

Treatment	*Comments*
Autoclave	Most media can be sterilized by treatment in an autoclave, usually a 15–20 min treatment at a pressure of 1 kg cm^{-2}. This raises the steam temperature to 121 °C
Steaming	Media that cannot be autoclaved (e.g. sugar media) may be sterilized by intermittent steaming. Samples are heated over boiling water in a steamer (85–95 °C) for 15 min on each of 3 successive days. Between treatments the material must be left at a suitable temperature for the growth of endospores
Filter sterilization	It is necessary to sterilize some ingredients separately before use. Heat-labile ingredients (some antibiotics, sugars, and vitamins) are sterilized by filtration through a membrane filter (pore size 0.2–0.45 μm), which excludes bacteria
Dry heat	Dry glassware such as empty flasks, e.g. pipettes cannot be satisfactorily autoclaved and must be sterilized in a hot air oven for a minimum of 2 h at a temperature of 160 °C. All material should be allowed to cool to room temperature before removal from the oven

tangential streaks cutting across the first streak and covering one half of the plate.

- Turn the plate through 90° and continue the process, flaming and cooling the loop between each set of streaks.

After incubation, well-separated colonies should be found along the streak marks. Before deciding that a culture is pure by either method, colonies should be picked off, grown and then reseparated, until all the colonies are the same. Staining must be used as a check on the purity of the final isolations.

2.3 Isolation of Bacteria from Natural Sources

The object of viable counting techniques is to estimate the number of all live organisms in a given sample of material. To do this, a medium satisfying the nutritional requirements of as many bacteria in the sample as possible is required. Other types of viable count and all isolation techniques embody the reverse principles. Here it is necessary to pick out and encourage one type of organism and to prevent or repress the development of the other types. In natural habitats, the organisms that one is isolating may be present only in small numbers. The first step in such cases is thus to obtain an enrichment culture by one or more of three methods:

1. using selective media,
2. using selective conditions of incubation,
3. selective pretreatment of the material.

Several generations of subcultures on liquid, or solid media may be necessary, but the final step will consist of plating out the organisms.

2.4 Use of Selective Media

A completely selective medium allowing the growth of only a single species is not attainable in practice, but media that will discourage all but the required species can be obtained.

Some selective media are also differential. Certain organisms, when grown on them, exhibit distinctive biochemical or morphological characters, which enable them to be recognized easily. This may be very important from a medical point of view, where the range of species concerned is small. Where a large range of species is involved, for example, in soil, differential media cannot be used to separate taxonomic groups, but are useful to distinguish groups concerned with some biochemical process, for example, cellulose decomposition.

Selectivity may be achieved in three ways,

1. by adding something to the medium to discourage the growth of species that are not required,
2. by altering the pH of the medium,
3. by omission of some ingredient required by most bacteria, but not the ones required.

Most bacteria are facultative aerobes, that is, they can grow under aerobic or anaerobic conditions, the latter ability being dependent upon the presence of some substance that can be utilized as a hydrogen acceptor by the species concerned. Some bacteria are obligate aerobes, unable to use anything but oxygen as a hydrogen acceptor. Others are obligate anaerobes, which cannot use oxygen as a hydrogen acceptor. A few bacteria are intermediate, growing best

in low oxygen tensions. These are called microaerophilic bacteria.

Anaerobic forms occur in several families of bacteria. They include several pathogens, for example, *Clostridium tetani*, certain autotrophic bacteria, for example, the purple sulfur bacteria, and some other soil or water bacteria, such as *Desulfovibrio* and *Clostridium*.

During growth, aerobic bacteria tend to utilize all available oxygen and so reduce the medium. Thus, in mixed cultures the oxidation–reduction potential (Eh) of the medium may become low enough to allow anaerobes to develop. This property has also been utilized to grow pure cultures of anaerobic bacteria. Thus, double petri dishes have been cemented together, so that the aerobe in one half uses the oxygen, allowing the anaerobe to develop in the other half.

Pure cultures of anaerobes are usually grown making use of other methods. Steps must be taken to ensure that access of oxygen to cultures is prevented or reducing substances must be added to the medium. Media to be used for the cultivation of anaerobes must be heated at 100 °C for 10 min to drive off the dissolved oxygen and cooled quickly, immediately before use.

2.5 Methods for Anaerobic Cultures

Stab and shake cultures. Many anaerobes can grow in deep stab or shake cultures in glucose agar. This method is particularly useful for microaerophilic species. A seal of liquid paraffin or Vaseline is sometimes advocated to help maintain anaerobic conditions, for example, the Hugh & Leifson fermentation test.

Anaerobe jar culture. If surface cultures are required, the plates or slopes should be placed in an anaerobe jar. This is a glass or plastic jar with a lid that can be clamped down firmly to form an effective seal. Usually a Gas-Pak method is used in which a sachet (to which water has been added) is placed in the jar. The chemical reaction uses oxygen and generates both hydrogen and carbon dioxide. A strip of methylene blue paper acts as an indicator.

Robertson's cooked meat broth. 1 kg of lean meat is minced and simmered with 1 L of water plus 1.5 mL of 1 M NaOH for 20 min, strained and washed very thoroughly with distilled water and partially dried. It is distributed into tubes to form a layer at least 1-cm deep. Nutrient broth is then added to a depth of about 8 cm and the medium is sterilized. The sterilized tissue contains reducing substances that are effective in maintaining anaerobic conditions at the bottom of the tube. The reducing activity of the meat is shown by the pink hematin at the bottom of the tube.

Thioglycollate medium. This is prepared by adding 0.1% (up to 0.4%) thioglycollic acid to the nutrient broth before adjusting the pH. 1% glucose must also be added. Although the medium may be solidified with agar, it is more usual to use a semisolid medium, that is, 0.5% agar. The increased viscosity of the medium prevents the distribution of oxygen (dissolved at the exposed surface), by convection currents. Methylene blue is added to act as an indicator of reducing conditions. A blue-green layer at the surface shows the depth to which oxygen has diffused into the medium. Inocula should be introduced carefully by means of a fine pipette, to the bottom of the tube.

3 Microscopic Examination of Bacteria

3.1 Introduction

Since most bacteria are not more than 2 to 4 μm in length and 0.3 to 1.0 μm in width, any examination of bacteria involves the use of high magnifications. To attain these magnifications use is made of oil immersion lenses with focal lengths of 2 mm or less. Care must be taken when using them as the working distance between the lens and object is very small. Always locate bacteria on a slide with the ×10 and ×40 lenses before preparing to examine them under oil.

3.2 The Examination of Living Unstained Bacteria

The most useful and frequently adopted method of examining live bacteria is the "hanging drop" preparation.

- A small piece of plasticine should be rolled out until its length equals the circumference of the cover slip.
- The plasticine is placed on the slide in the form of a ring or a square.
- With a wire loop, a small drop of the bacterial suspension is placed on a cover slip laid on the bench.
- The slide is then inverted, the plasticine pressed gently on the cover slip and the slide quickly turned over again, leaving the suspension as a "hanging drop."
- The slide should then be examined under the ×10 objective, and with the edge of the drop focused. Swing the ×40 objective lens into position and adjust the microscope. The method is used to detect motility, but this should be distinguished from the Brownian movement that is exhibited by any small particles suspended in a liquid. Examination of these preparations under oil is difficult since currents of liquid are set up during focusing.

Phase contrast microscopy is the easiest method for examining hanging drops or even ordinary cover slip preparations. Bacteria appear as dark objectives against a pale background. The method, utilizing special objectives and condensers, depends upon the change of phase of the white light as it passes through objects with different refractive indices.

3.3 The Measurement of Bacteria

The measurement of bacteria is achieved by using a stage micrometer, together with a micrometer eyepiece. On the stage micrometer, a scale graduated in hundredths of a millimeter is engraved. In the micrometer, the eyepiece is a second graduated scale, which may be focused with a movable eyepiece lens.

- The micrometer eyepiece is inserted into the microscope and the stage micrometer is focused.
- The number of divisions of the eyepiece scale corresponding to a definite number of stage micrometer scale divisions is determined.
- The stage micrometer is removed, the object to be measured is focused, and the number of divisions of the eyepiece scale, which cover the object, is noted.

3.4 Methods of Staining Bacteria

The preparation of films for staining

- A drop of the bacterial suspension should be transferred to a clean grease-free slide with a sterile loop. The drop

should be small, it should be spread evenly, and allowed to air-dry.

- Because of their small size, bacteria dry without great distortion and so the only fixation required is to pass the slide quickly two or three times through the Bunsen flame at this stage.

General staining principles. Dyes may be of the acidic or basic type, the former owing their colored properties to an anion, the latter to a cation. Basic dyes have greatest affinity for the nuclei of cells because of the acidic nature of the nuclear material. Acid dyes have greatest affinity for the cytoplasm.

The chemicals used in the preparation of staining solutions are frequently variable in composition. Variations of this type and those of technique make it essential to use proper controls. Smears of known positive and negative organisms should always be prepared.

Differential stains. There are a number of more complex staining procedures designed to demonstrate different parts of the bacterial cell, and different types of bacteria.

Gram's method. Some bacteria when treated with para-rosaniline dyes and iodine retain the stain when subsequently treated with decolorizing agents such as alcohol or acetone. Other bacteria lose the stain. This property is used to distinguish two types of bacteria, gram-positive and gram-negative forms. Gram-negative forms, which are those that lose the stain on decolorization, can be made visible by using a suitable counterstain. It is known that the stain plus the iodine react with the RNA fraction of the cytoplasm and form a colored, alcohol-soluble substance. Gram-positive forms have cell walls (~95% peptidoglycan) that prevent the "leaching out" of the stain from the cytoplasm with alcohol while the gram-negative ones (~5% peptidoglycan) do not. A useful form of the gram's stain is Hucker's modification.

- The smears are stained with ammonium oxalate crystal violet solution for 1 min. Ammonium oxylate crystal violet is prepared using ammonium oxylate 0.8 g, crystal violet 2.0 g, ethyl alcohol 20 mL and distilled water 80 mL.
- The slides are washed in tap water for 2 to 3 s and then immersed for a further minute in Lugol's iodine solution. Lugol's iodine is prepared by adding iodine 1.0 g, potassium iodide 2.0 g, distilled water 300.0 mL.
- The slides are washed again and blotted dry.
- This is followed by treatment with 95% *ethanol* for 30 s. The slide should be gently rocked during this step.
- After further quick washing with water and blotting, the slides are counterstained with *safranin solution* for 10 to 30 s. Safranin solution is prepared by adding safranin, (2.5% solution in 95% ethanol), 10 mL and distilled water 100 mL.
- The preparation is washed, blotted, and examined under the oil immersion lens. Gram-positive organisms will be stained purple and gram-negative organisms will be stained red. Endospores remain unstained. Organisms with known reactions should be used as controls on the same slide.

Acid-fast staining (Modified Ziehl–Neilsen method). Certain bacteria, for example, *Mycobacterium*, cannot be stained by ordinary aniline dyes except by using strong solutions with phenol and possibly heat. Once stained, these bacteria resist decolorization by acids and are called *acid-fast bacteria*. This stain is used

primarily to detect *Mycobacterium tuberculosis*, but other species of *Mycobacterium* are acid-fast to some extent and can be destained by following the acid treatment with 95% alcohol.

- The dried film is stained with Ziehl's carbol fuchsin for 5 min, applying enough heat to give steam. Do not let the slide dry out. Ziehl's carbol fuchsin is prepared by dissolving basic fuchsin (0.3 g) in 10 mL of ethanol. This is then added to phenol (heat melted crystals), 5 mL dissolved in 95 mL of distilled water. The solution must be mixed, allowed to stand for a few days, and filtered prior to use.
- The slide is washed.
- After removing the excess water, the slide is treated with a decolorizing solution (hydrochloric acid, 3.0 mL; ethyl alcohol, (95% v/v) 97 mL).
- The film is immediately washed with water and the decolorization repeated until the film appears faintly pink.
- The organisms are counterstained with 0.3% aqueous methylene blue for 20 to 30 s, washed and blotted dry. Acid-fast organisms retain the red stain while others are stained blue. Endospores of *Bacillus* are also acid-fast (see below).

Spore staining (Conklin's method). Spores are not so readily stained by Gram's stain as is the rest of the cell. Thus, frequently they are observed as clear areas in the center of otherwise stained cells. This is because the spore is rather more acid-fast than the rest of the cell, that is, it does not stain easily, but once stained does not release the stain readily. Perhaps the easiest method of staining spores is by Conklin's method.

- Smears are prepared and fixed in the normal way. The slide is stained with 5% aqueous malachite green for 5 min, heating the slide gently over a low flame until it steams. Aqueous malachite green is produced by adding 5 g of malachite green to 100 mL distilled water.
- After washing the slide for 30 s, the smear is counterstained with 5% aqueous mercurochrome or 0.25% aqueous safranin (10 mL 2.5% safranin solution in 95% ethanol added to 100 mL of distilled water) for 30 s.
- The slide is then washed and blotted dry. Under oil, the spores are stained green and the rest of the cell is stained red.

4 Estimation of Population Sizes of Bacteria

4.1 Bacterial Biomass Measurement

Biomass is the dry weight of bacterial substance per mL of a suspension and can be determined directly provided that the suspension contains no foreign matter.

A sample of culture is treated with formalin (to give a final concentration of 1% v/v) and centrifuged in a weighed tube. The deposit is washed with 0.85% w/v saline +1% formalin and then in 0.05% saline and finally in distilled water. The deposit is dried to constant weight at 105°, placed in a desiccator over phosphorus pentoxide and allowed to cool, and then weighed in a balance (previously dried atmosphere with silica gel).

4.2 Turbidometric Estimations

Where large numbers of determinations are to be made, for example, in making growth curves, turbidometric estimations can be made. These are meaningless,

however, unless translated into terms of cell concentration (viable or total counts), but the relationship between turbidity and cell concentration is different for each different species. Turbidity readings obtained for a series of dilutions of a known number of organisms should be plotted as a calibration graph.

4.3 Theory of Bacterial Growth

Logarithmic growth. The growth of a bacterial culture undergoing binary fission is expressed by the equation,

$$x = x_o 2^n$$

where $x =$ the final number of organisms, x_o the initial number and n is the number of generations.

If n generations are produced in time t, then

$$g = \frac{t}{n} \text{ and } n = \frac{t}{g}$$

where $g =$ the generation or doubling time. Thus

$$x = x_o 2^{t/g}$$

and

$$\ln x = \ln x_o + \ln 2 \frac{t}{g}$$

This is usually written in the form

$$\ln x = \ln x_o + \mu t$$

where $\mu = \ln 2/g = 0.6931/g$.

μ represents the number of e-fold increases in cell numbers that have occurred in time t.

If the equation

$$\ln x = \ln x_o + \mu t$$

is differentiated, it becomes the familiar growth equation

$$\frac{dx}{dt} = \mu x$$

which may be stated as the rate of growth of the population is proportional to the initial number of organisms present.

4.4 Counting Methods

Counts may be used to determine the number of organisms in a sample and these counts can be converted to estimates of biomass if suitable conversion factors are known.

Total counts. Total counts include the various methods of estimating the total number of bacteria per unit volume or weight of a sample, without distinguishing between viable and nonviable organisms.

Counting chambers. The number of bacteria per mL of a suspension can be determined by means of a counting chamber. The best form of a counting chamber is the Helber cell, but a hemocytometer is still useful. Suspensions of pathogenic or motile bacteria should be inactivated before they are counted, by heating for 1 h at 56 to 60 °C.

- Prepare a dilution that will give an average count of about 20 bacteria per small square, if necessary, examining a drop of undiluted suspension to decide the dilution. The following method is suitable unless extreme accuracy is needed. Measure out 0.5 mL of suspension into a small tube and add an equal volume of diluent (formol saline; 0.5% formalin in 0.9% sodium chloride solution is a suitable diluent) to make a twofold dilution. Mix well by drawing up and down several times in the pipette

and then transfer a measured volume to a second tube. Wash the pipette out with a diluent and add an equal volume of diluent to this second tube. Repeat making further 2-fold dilutions to give a series of 4-fold, 8-fold, 16-fold dilutions of the original suspension to the required degree.

- A few drops of methylene blue may be added to the diluent to make the bacteria more easily seen.
- Place a No. 1 coverslip on the slide and then exactly fill the chamber (not the ditches) with the suspension using a fine pipette.
- Find the grid under the low power, cutting down the light intensity if necessary, and use the ×40 objective for counting. Oil should not be used since it causes sagging of the coverslip and a consequent change in volume of the liquid.
- Count the number of organisms per small square that cover all planes of focus, in 20 squares taken diagonally across the grid. Repeat with a second mount and if time allows, with a third. At least 1000 cells should be counted. The more organisms that are counted, the more accurate is the estimate.
- If cells lie on the lines of the grid, only those on two sides of the square should be counted. All the component cells in small clumps of completely separated cells should be counted as single individuals.
- From the number of organisms per small square, the number per mL of the suspension is calculated. The volume of liquid lying over each square is given on the slide. In a Thoma hemocytometer, the depth of the chamber is 0.01 cm and the area of each small square is 0.000025 cm^2, hence, the number of bacteria per milliliter of the suspension is the average number per small square $\times 4 \times 10^6 \times$ the dilution factor.
- In the Helber cell, the depth of the chamber is 0.002 cm, and the area of each small square is 0.000025 cm^2. Consequently, the number of cells per milliliter is the number per small square $\times 2 \times 10^7 \times$ the dilution factor.

The direct smear method. A known volume of bacteria in suspension is spread over a defined area of a microscope slide, stained, and counted. A variety of differential stains can be used, for example, methylene blue.

- Transfer 0.01 mL of the suspension to a 1 cm^2 area of a microscope slide.
- Allow the slide to dry at room temperature.
- Stain in a damp chamber for 5 min with the following solution (1.3 mL, 0.5 M carbonate–bicarbonate buffer, pH 9.6; 6.0 mL, 0.001 M phosphate buffered saline, pH 7.1; 5.7 mL, 0.85% physiological saline, 5.3 mg crystalline Fluorescein isothiocyanate (FITC)).
- Wash gently but thoroughly in water and blot dry. Observe with a fluorescence microscope.

Viable counts. These methods aim at estimating the number of bacteria capable of growth and multiplication in the material. Aseptic methods must be used throughout.

Making dilutions. In all these methods, the preparation of dilutions is the first step. A weighed amount or measured volume of the material being examined is suspended in a known volume of sterile diluent, mixed well by shaking and further dilutions are made from this.

The diluent used is most important. Sometimes the natural milieu that is,

seawater or river water is used, but more often sterile quarter-strength Ringer's solution is recommended (NaCl, 9 g; KCl, 0.42 g; $CaCl_2$, 0.48 g; NaH_2CO_3, 0.2 g; water, 4 L). A 0.1% peptone solution may also be used as a diluent.

- The required number of tubes containing 9 mL diluent are prepared with a 1-mL sterile blow-out pipette, 1 mL of the sample is added to the first tube.
- With a fresh pipette tip, the liquid is mixed and 1 mL of this 1/10 dilution transferred to a second 9 mL of diluent.
- Repeat until the sample is diluted sufficiently. The number of dilutions needed will depend on the material concerned, only the two or three highest being used.

The dilution-plate method and the spread plate method. Tenfold dilutions are prepared as described above.

- With the pipette used for mixing the dilution, 1 mL of this dilution is placed in each of three or more sterile petri dishes; then after making the next dilutions, 1 mL of this is treated in the same manner and so on until a range of dilutions is plated (usually 3).
- After mixing the dilutions, 15 or 20 mL of an agar medium, previously melted and cooled to 45 °C, is added to each plate and an even distribution is ensured.
- When set, the plates are incubated at the required temperature. After incubation, dilutions where the plates show more than 300 or fewer than 30 colonies are discarded. Errors due to overcrowding or due to sampling make these counts unreliable. Thus only one set of plates will actually be counted if tenfold dilutions are used.
- The inoculum may, as an alternative, be spread over the surface of prepared plates of agar. Usually an inoculum of 0.1 mL is then used and the agar is dried to remove surface water. The inoculum is spread with a sterile glass spreader.

The membrane filter method. This method is used for water samples in which only a few organisms are present. It depends upon the concentration of the inoculum rather than the dilution.

- Sterile cellulose acetate membranes (pore size ~0.45 μm) are placed in a sterilized filter holder.
- A known volume of the fluid is passed through the filter.
- This filter is then laid on a thin layer of the medium in a petri dish and incubated. Bacteria held in the filter develop into colonies that can be counted. Difficulty is often experienced since the even distribution of the colonies cannot be ensured. By using suitable media, particular groups of organisms can be selected.

The "Most Probable Number" method. As well as being economical in time and materials, this method can be used where the plate count cannot, for example, with materials that contain species that form "spreading" colonies on agar or where it is desired to determine numbers of a particular group that is sparsely represented in material with a mixed bacterial flora.

A series of tenfold dilutions are prepared, so that when 1 mL samples of the highest dilution are inoculated into a suitable medium, very few if any bacteria are likely to grow. Thus, if a suspension contains 5×10^9 bacteria per mL, 1 mL of a 10^{-9} dilution would contain 5 organisms, and most or all tubes of medium would

show growth. In a 10^{-12} dilution, all 5 tubes of medium would probably show no growth as there would be only 1 organism per 200 mL.

- Transfer 3 (4 or 5 if possible) replicate 1 mL samples of each suitable dilution to tubes of liquid or plates of solid medium and incubate. A greater range of dilutions (usually 5 or 6) is required in this method, unless the result can be predicted accurately.
- After incubation, record the presence or absence of growth in the recovery cultures.
- Statistical tables, for example, McGrady tables have been compiled, enabling the most probable number of organisms per milliliter of the suspension to be determined. The following example of a McGrady table has been computed for a dilution factor of 10, and for 5 replicate tubes inoculated from each dilution of culture (Table 2).

The entries in the table are the number of turbid tubes observed after inoculation of the same volume from any three successive tenfold dilutions. The stated MPN is the estimated mean viable count per inoculum taken from the most concentrated of the three dilutions being considered.

5 Characterization of Bacteria Using Biochemical and Cultural Tests

5.1 Introduction

Because bacteria are small, morphological characters have never proved satisfactory

Tab. 2 Values of the Most Probable Number (MPN) for 5 tubes inoculated from each of 3 successive 10-fold dilutions.

Significant number Numbers of replicate tubes showing turbidity at 3 successive dilutions			*MPN*	*Significant number Numbers of replicate tubes showing turbidity at 3 successive dilutions*			*MPN*
0	1	0	0.18	5	0	0	2.3
1	0	0	0.20	5	0	1	3.1
1	1	0	0.40	5	1	0	3.3
2	0	0	0.45	5	1	1	4.6
2	0	1	0.68	5	2	0	4.9
2	1	0	0.68	5	2	1	7.0
2	2	0	0.93	5	2	2	9.5
3	0	0	0.78	5	3	0	11.0
3	0	1	1.1	5	3	1	11.0
3	1	0	1.1	5	3	2	14.0
3	2	0	1.4	5	4	0	13.0
4	0	0	1.3	5	4	1	17.0
4	0	1	1.7	5	4	2	22.0
4	1	0	1.7	5	4	3	28.0
4	1	1	2.1	5	5	0	24.0
4	2	0	2.2	5	5	1	35.0
4	2	1	2.6	5	5	2	54.0
4	3	0	2.7	5	5	3	92.0
–	–	–	–	5	5	4	160.0

for the delimitation of species, although in certain cases, it is possible to assign the organisms to families and sometimes to genera. Other characters have been used therefore, to enable more precise classifications of organisms to be made. Perhaps the most widely used tests, however, have been those designed to demonstrate the biochemical and physiological characteristics of microorganisms. The biochemical methods used fall into four groups:

1. Mineral growth requirements and substrate utilization.
2. Detection of metabolic end products.
3. Identification of characteristic compounds, for example, enzymes, gums, and toxins.
4. Resistance and sensitivity to chemicals, for example, antibiotics.

5.1.1 Minimal Growth Requirements and Substrate Utilization

Nutritional studies. It is often useful to know something of the nutrition of a particular bacterium, from the point of view of both taxonomic and cultural studies. It is important to obtain information on the relation of bacteria to nitrogen sources and growth factors. Growth of bacteria with an inorganic nitrogen source can be detected in the Ayer, Rupp, and Johnson's medium: $NH_4H_2PO_4$, 1.0 g; KCl, 0.2 g; $MgSO_4.7H_2O$, 0.2 g; glucose, 10.0 g; water, 1 L. The pH should be adjusted to about 7.0. Several serial subcultures should be attempted in this medium before recording growth as positive since the inoculum may carry over organic nitrogen and growth factors. Further nutritional studies can be made by replacing sources of nitrogen or growth factors in this medium with others. If growth is obtained in the basal medium, all subsequent tests can be carried out using this medium with the particular substances under investigation substituted, for example, in fermentation tests. If no growth is recorded, peptone media must be used as a basis for growth.

Effects of environment. The relation of organisms to free oxygen and pH also fall into this group of tests. These relationships can be elucidated by methods outlined earlier.

Utilization of organic acids. Some bacteria are able to utilize salts of certain organic acids as sole carbon sources. Pathogenic species, which are nutritionally exacting, are less likely to be able to do this, while many of the less exacting soil and water forms can. Thus, while the soil coliforms can utilize citrate, intestinal coliforms cannot, for example, *Escherichia coli*, a commensal in the animal intestine.

Koser's citrate medium is a mineral medium with 0.2% anhydrous sodium citrate as the sole carbon source. Bromothymol blue is usually added to assist in detecting slight growth, by the alkaline reaction produced. Other organic acids can be detected in a similar way, but a lower concentration, for example, 0.1% should be used for the more toxic acids, that is, formic and oxalic acids. The mineral medium most frequently used is KH_2PO_4, 1.0 g; $MgSO_4{\cdot}7H_2O$, 0.2 g; $Na(NH_4)HPO_4{\cdot}4H_2O$, 1.5 g; distilled water, 1 L.

5.1.2 Detection of Metabolic End Products

Indole. Indole is produced by some bacteria from the amino acid tryptophan. A tube of peptone water is inoculated and samples of the culture are tested

at intervals with Ehrlich's reagent (2% para-dimethylaminobenzaldehyde in 95% alcohol). 1 mL is added to the culture and concentrated HCl is added drop by drop until a red zone appears between the alcohol and the peptone solution. No more than 0.5 mL acid is needed. If the colored pigment is soluble in amyl alcohol or chloroform it can be considered to be indole.

Hydrogen sulfide. Hydrogen sulfide may be produced from sulfur-containing amino acids, for example, hydrogen cystine present in peptone. A test for hydrogen sulfide production can be combined with the above test. When the peptone water is inoculated, a filter paper strip impregnated with a saturated lead acetate solution is inserted between the plug and the tube, taking care to keep it dry. If hydrogen sulfide is produced, the paper will become blackened.

Some bacteria also reduce inorganic sulfur compounds. A chemically defined medium containing a sulfur compound, for example, sulfate, or thiosulfate, can be used instead of peptone water. Only organisms that can utilize inorganic nitrogen can be tested in this way.

A number of special media containing iron salts and a source of sulfate can also be used, for example, Kligler's iron agar. Hydrogen sulfide production is recognized by blackening of the medium.

Acetyl methyl carbinol – The Vosges–Proskauer (VP) test. Acetoin or acetylmethylcarbinol can be produced from glucose via pyruvic acid. Two enzymes are necessary for this to occur, that is, pyruvate enzyme and acetolactic decarboxylase. Acetoin may be reduced to 2.3 butanediol and so give a negative result. Sensitive tests should therefore be used to detect acetoin.

The normal method of detection is to oxidize acetoin to diacetyl, which reacts with guanidine residues in the medium to give a cherry red color. The medium used usually contains 0.5% glucose, 0.5% peptone and 0.5% K_2HPO_4 in distilled water. To prevent reducing conditions from developing, fumarate is sometimes added as a hydrogen acceptor or else glucose is replaced by pyruvate.

- After 3 days of incubation, a 2 mL sample of the culture fluid is tested by adding about 5 mL of 40% KOH and a trace of creatine, shaking vigorously and allowing it to stand for up to 60 min. The culture is positive if a cherry red color develops.

Methyl red test. In the examination of coliform bacteria, the VP Test is carried out at the same time as the methyl red test. The latter determines the hydrogen ion concentration produced in a definite time and in a specific medium. The glucose–phosphate medium described above must be used and not a modification of it.

- The culture is tested after 3 days at 30 °C.
- To one part of the culture is added a 0.4% solution of methyl red (1 drop per mL of culture). A magenta red color is positive, a yellow color is negative.

Reduction of nitrates. Certain organisms are capable of reducing nitrate to nitrite and even to nitrogen or ammonia.

- The organism under test is grown in peptone water containing 1% potassium nitrate.
- After incubation, the culture fluid is divided into three portions.
- One portion is tested by adding a few drops of sulfanilic acid reagent followed by a few drops of α-napthylamine reagent (Griess' reagents I and II). A red color indicates the presence

of nitrite. The production of free nitrogen may be detected by allowing it to collect in a Durham tube.

- A second portion of the fluid may be tested for ammonia with *Nessler's reagent*, but since many organisms produce ammonia from peptone, the result must be interpreted with care. [A sample of uninoculated medium should also be tested since some peptones give a positive reaction (i.e. an orange/brown precipitate) with Nessler's reagent].
- If no nitrite has been detected, the third portion of the culture should be tested for residual nitrate for it is possible that all the nitrate will have been converted first to nitrite and then to nitrogen or ammonia. To do this, add a small quantity of zinc dust. This will reduce the residual nitrate chemically to nitrite and this can then be tested for as shown above. The interpretation is as follows.
- *If nitrite is then +, it follows that the bacterium did not reduce nitrate to nitrite; rather it was the Zn dust that reduced it.* If the test on the third portion is negative for nitrite, it follows that all the nitrate had been reduced to another product via nitrite by bacteria and thus the bacterium will be recorded as a positive nitrate reducing organism.

Fermentation tests – Hugh and Leifson's method. Certain carbohydrates that is, sugars, alcohols, and glucosides are attacked and produce quantities of acid and/or gas. It is also important to know which acid and which gas are produced, that is, 2 keto-gluconic acid, propionic acid, lactic acid and so on. Generally speaking, simple tests for detecting the nature of the acids are lacking and so results are most frequently recorded as no acid, acid or acid plus gas. One of the most convenient tests for this purpose is that of Hugh and Leifson.

- A semisolid agar medium is tubed to a depth of 4 cm, two tubes being required for each culture and for each sugar tested. The formula for this medium is as follows:

Peptone	2.0 g
Sodium chloride	5.0 g
Dipotassium hydrogen phosphate	0.3 g
Bromothymol blue (1% aqueous solution)	3.0 mL
Carbohydrate	10.0 g
Water	1.0 L
pH	7.1

  The bromothymol blue is dissolved in water and 3.0 mL of a 1% solution added to each 1000 mL of medium. Alcoholic solutions of indicators should not be used as acid may be produced from the alcohol. For critical studies, the carbohydrate should be sterilized separately and added to the otherwise complete sterile medium.
- After inoculation, one of the pairs of tubes is covered by a layer of sterile melted Vaseline to a depth of 2 cm.
- Several types of reaction may be observed. *Fermentative organisms,* which phosphorylate glucose and then split it into 2 triose molecules before forming acid will produce an acid reaction in both tubes (or the sealed tube only). *Oxidative organisms,* which convert the aldehyde group to a carboxyl group in glucose to form gluconic acid will produce an acid reaction in the open tube only, leaving the sealed tube unchanged. The acid

reaction produced by oxidative organisms is apparent first at the surface and extends gradually downward into the medium. Organisms that oxidize glucose but do not ferment it have never been observed to ferment any other carbohydrates and so the sealed tube can be omitted when testing the other sugars.

Production of aesculetin from aesculin. Hydrolysis of aesculin is shown by the production of blackening in the following medium. Peptone, 10.0 g; aesculin, 1.0 g; ferric citrate, 0.5 g; water, 1 L; agar, 1.2 g.

Levan production. Many bacteria, for example, plant pathogens, are characterized by an ability to produce levan polysaccharides from sucrose. Such compounds may also be produced by soil bacteria and help in maintaining soil crumb structure. This may be detected on a plate of nutrient agar containing 5% sucrose, poured and a channel cut in the agar. This channel is filled with the same medium with 0.1% aniline blue in it. Polysaccharide producing organisms attract the dye toward them while others remain colorless. The inoculum is usually added in two ways; as streaks to the left of the channel and as individual spaced colonies to the right of the channel.

5.1.3 Biochemical Examination of Anaerobic Bacteria

For tests requiring plate cultures, anaerobic jars should be used for incubation, but many of the tests can be carried out without this apparatus, merely by slightly modifying the media.

1. The media may be rendered semisolid by adding 0.5% agar.
2. Fermentative metabolism can be demonstrated by sealing cultures with Vaseline.

5.1.4 Rapid Testing Methods

Analytical Profile Index (API) test strips. API test strips, produced by Bio Mérieux SA, France, may be used to obtain test results quickly. These consist of a series of miniature cupules on a molded plastic strip, each of which contains a sterile dehydrated medium in powder form. Addition of water containing a bacterial suspension simultaneously rehydrates and inoculates the medium. A rapid reaction is obtained because of the small volume of medium and the large inoculum used. Reactions are evident from color changes of indicator chemicals either with or without addition of further reagents. These changes occur after incubation of the strip in a small, humidified, plastic chamber. A range of strips is available containing between 20 and 50 tubes with cupules. These can be used to identify specific groups of organisms, for example, members of the Enterobacteriaceae, staphylococci, anaerobes, lactic acid bacteria, streptococci, and clinical yeasts.

Enterotube II system. The basic philosophy of the Enterotube II system is the same as the API system. The system provides ease, speed, and low cost in the identification of gram-negative Enterobacteriaceae. The system consists of a single tube containing 12 compartments, each containing a different agar-solidified culture medium. Compartments that require aerobic conditions have small openings that allow in air. Those requiring anaerobic conditions have a layer of paraffin wax. There is a self-enclosed inoculating

needle. This needle can touch an isolated bacterial colony and is then drawn through the 12 compartments, inoculating the test media. Fifteen standard tests are performed. After 18 to 24 h of incubation the color changes that occur in each of the compartments are recorded. Interpretation of the results is determined using a five-digit code obtained from the results. A differential chart, supplied by the manufacturer, or a computer-based program, (ENCISE) Enterobacteriaceae numerical coding and identification system for Enterotube can be used.

5.2 Cultural Characteristics

In addition to the morphological and biochemical characters, it is essential to record the appearance of bacterial colonies under defined conditions, for example, on specific media as agar streak plates, gelatin stabs, and in broth culture. The shape and structure of colonies can be described as shown in the figure below:

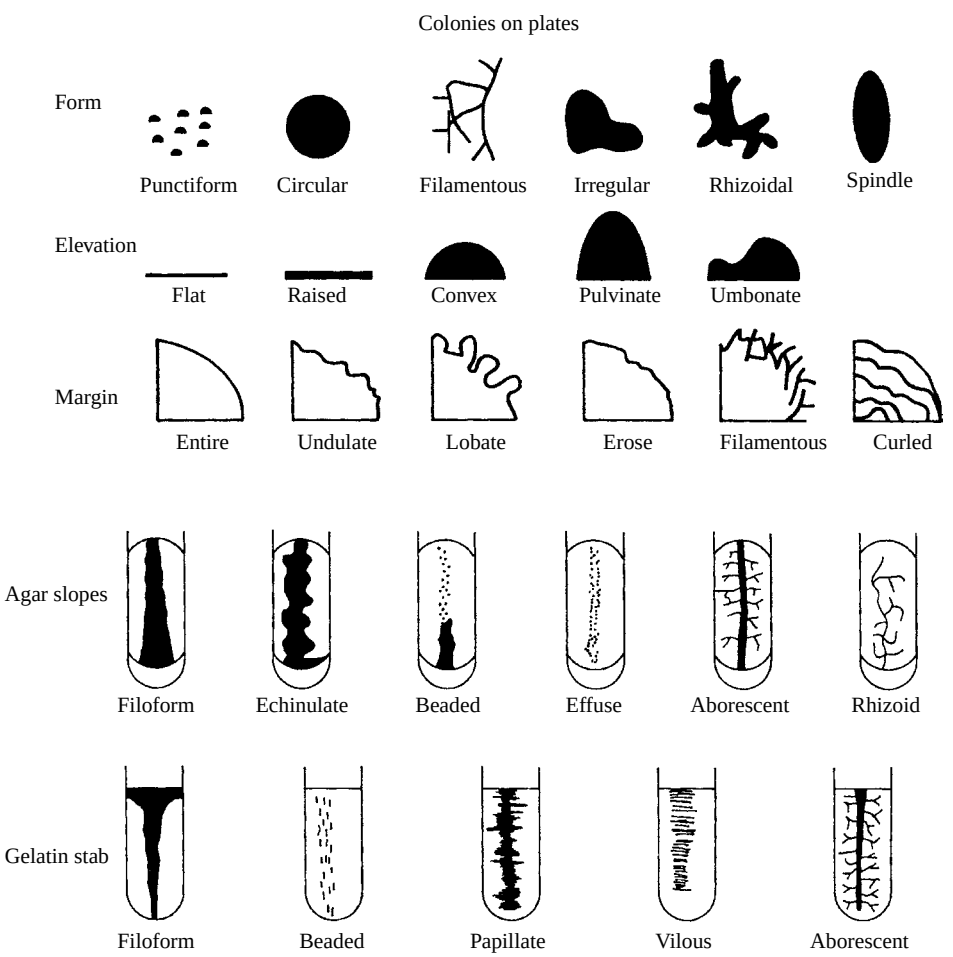

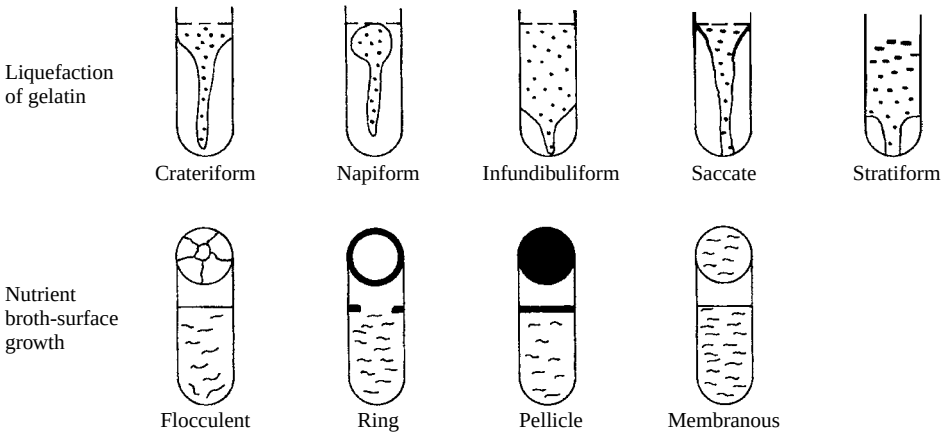

See also Bacterial Growth and Division; Cell Growth in Microgravity.

Bibliography

Atlas, R.M. and Parks, L.C. (1983) *Handbook of Microbiological Media*, CRC Press, Boca Raton, FL.

Ball, A.S. (1997) *Bacterial Cell Culture: Essential Data*, Wiley and Sons, Chichester.

Balows, A. et al. (1992) The Prokaryotes: *A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications*, 2nd edition, Springer-Verlag, New York.

Block, S.S. (Ed.) (1991) *Disinfection, Sterilization and Preservation*, 4th edition, Lea and Febiger, Philadelphia.

Boatman, E.S. et al. (1987) *Biosciences* **37**, 384.

Gerhardt, P. et al. (Eds.) (1981) *Manual of Methods for General Microbiology*, American Society for General Microbiology, Washington, DC.

Gerhardt, P. et al. (Eds.) (1984) *Manual of Methods for General Bacteriology*, American Society for General Microbiology, Washington, DC.

Holt, J.G. et al. (Eds.) (1994) *Bergey's Manual of Determinative Bacteriology*, 9th edition, Williams and Wilkins, Baltimore, MD.

Isenberg, H.D. (Ed.) (1992) *Clinical Microbiology Procedures Handbook*, American Society for Microbiology, Washington, DC.

Issac, S., Jennnings, D. (1995) *Microbial Culture*, Introduction to Biotechniques, BIOS Scientific Publishers, Oxford.

Koch, A.L. (1981) *Manual of Methods for General Bacteriology*, American Society for Microbiology. Washington, DC.

Perkins, J.J. (1983) *Principles and Methods of Sterilization in Health Sciences*, 2nd edition, Charles C. Thomas, Springfield, IL.

Priest, F., Austin, B. (1993) *Modern Bacterial Taxonomy*, 2nd edition, Chapman & Hall, London.

Rawlings, D.J. (1992) *Light Microscopy*, BIOS Scientific Publishers, Oxford.

Appendix 1 Sources of Media and Supplies

Anachem Ltd
Anachem House
20 Charles Street
Luton
Bedfordshire LU2 OEB
UK

BDH Laboratory Suppliers
Merck Ltd
Hunter Boulevard
Magna Park
Lutterworth
Leicestershire LE17 4XN
UK

Bio-Rad Laboratories
Main Office
2000 Alfred Noble Drive,
Hercules, CA 94647

Bio-Rad Laboratories Ltd
Bio-Rad House
Marylands Avenue
Hemel Hempstead
Hertfordshire HP2 7TD
UK

Difco Laboratories
PO Box 331058
Detroit, MI 48232
USA

Oxoid USA Inc.
9017 Red Branch Road
Columbia
MD 21045
USA

Scott Laboratories
Fiskville
RI 02823
USA

Bacterial Growth and Division

Stephen Cooper
University of Michigan Medical School, Ann Arbor, MI, USA

1 Growth of Bacterial Cultures 528
1.1 The Classical Growth Curve 528
1.2 Bacterial Growth at Different Growth Rates 529
1.3 The Shift-up 529
1.4 Classical Growth Pattern Is Merely a Series of Shift-ups and Shift-downs 530

2 The Bacterial Division Cycle 530
2.1 Cell Age and the Age Distribution during Balanced Growth 530
2.2 Aggregation Theory and the Control of Synthesis during the Division Cycle 531

3 Synthesis of Cell Components during the Division Cycle 532
3.1 Cytoplasm Synthesis during the Division Cycle 532
3.1.1 Exponential Cytoplasm Synthesis during the Division Cycle 532
3.1.2 Variation in the Rate of Cytoplasm Increase 533
3.2 DNA Synthesis during the Division Cycle 534
3.2.1 Pattern of DNA Synthesis during the Division Cycle 534
3.2.2 Regulation of Initiation of DNA Replication 536
3.2.3 The Specific Origin of Replication 536
3.2.4 Protein Factors Controlling DNA Replication 537
3.2.5 Chromosome Organization and Bidirectional Replication 537
3.2.6 DnaA Protein Is Involved in Initiation of DNA Replication 537
3.3 Surface Synthesis during the Division Cycle 538
3.3.1 Peptidoglycan Synthesis during the Division Cycle 538
3.3.2 The Invagination Process 541
3.3.3 FtsZ Protein Is Involved in Septum Formation and Cell Division 541
3.3.4 Localization of the Division Site 541

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

4 Plasmid Replication during the Division Cycle 542
4.1 High-copy Plasmids 542
4.2 Low-copy plasmids 542
4.3 Minichromosome Replication during the Division Cycle 542

5 Segregation of Cell Components at Division 542
5.1 Segregation of Cytoplasm 542
5.2 Segregation of DNA 543
5.3 Segregation of the Cell Wall 544

6 Are There Events during the Division Cycle? 544

7 Do Checkpoints Control the Bacterial Cell Cycle? 545

8 Temporal Variation of the Bacterial Division Cycle 546

9 Alternative Bacterial Division Cycles 546
9.1 *Caulobacter crescentus* Division Cycle 546
9.2 *Bacillus subtilis* Division Cycle 546
9.3 Streptococcal Growth during the Division Cycle 547

10 Mass Increase Is the Driving Force of the Division Cycle 547

11 The Bacterial Growth Law during the Division Cycle 548

12 Experimental Analysis of the Bacterial Division Cycle 549

13 Genetic Analysis of the Bacterial Cell Cycle 549

14 Die Einheit der Zellbiologie 550

Bibliography 551
Books and Reviews 551

Keywords

C-period
The time for a round of DNA replication, from initiation at the origin to completion at the terminus. In *Escherichia coli*, the C-period, for cells growing between 20 and 60 min doubling times, is a constant of approximately 40 min.

D-period
The time between termination of DNA replication and cell division. In *Escherichia coli*, the D-period, for cells growing between 20 and 60 min doubling times, is a constant of approximately 20 min.

Multiple Forks
The situation that occurs when initiation of DNA replication occurs on a molecule that has not finished its previous round of replication; there are forks upon forked material.

Cytoplasm
The collective name for the numerous components of the cell that are not associated with either the cell wall or the genome. It is composed of the small molecules such as ions, metabolites, and cofactors, macromolecules such as soluble enzymes, tRNAs, mRNAs, and higher complexes such as ribosomes.

Peptidoglycan
The stress-bearing, presumably shape-maintaining, layer of the bacterial cell wall. It is composed of glycan chains cross-linked with amino acids.

Cell Membrane
The lipid-containing structures that lie adjacent to the peptidoglycan layer. Some cells have a single membrane layer, while others have two different membranes, an outer and an inner membrane.

Segregation
The distribution of cell material from the dividing mother cell into the two daughter cells. Different cell components may have different segregation patterns.

Origin
A sequence on a chromosome or a plasmid where normal replication of DNA initiates.

Terminus
A point or region on a chromosome or plasmid where a round of replication terminates.

Bidirectional Replication
Replication of DNA with two replication forks proceeding in opposite directions away from the origin of replication.

GC-skew
The excess (or deficit) of G residues over C residues within a single strand of DNA. While G = C in a double helix, an individual strand may have an excess of G within a region or an excess of C residues within a region.

FtsZ Ring
The ring of synthesis of the pole region during pole formation.

A dividing bacterial cell must, on average, have precisely twice as much of everything found in a newborn cell. How does a bacterial cell ensure that all cell components are duplicated between divisions, and how does a cell ensure that it does not divide prior to duplicating all of its components? These questions have been investigated by studying biosynthesis of cell components during the division cycle. Cell cytoplasm increases uniformly and exponentially during the division cycle. There does not appear to be any cell-cycle-specific syntheses of cytoplasmic components during the division cycle of a well-studied bacterium such as *E. coli*. DNA replication is initiated at a specific time in the division cycle when the cell mass per chromosome origin reaches a particular value. Once initiated, DNA replicates at a constant rate such that the time between initiation and termination of replication is relatively independent of the growth rate. The time between termination of replication and cell division is also relatively constant. High-copy plasmids replicate throughout the division cycle and low-copy plasmids replicate at a particular time during the division cycle according to rules similar to that for initiation of chromosome replication. The cell surface grows in response to the increase in cell mass so that the turgor pressure inside the cell, and the cell density, are constant during the division cycle. At division, the components of the cell are segregated to the two daughter cells. The cytoplasm segregates randomly, the DNA segregates stochastically but nonrandomly, and the peptidoglycan segregates in a manner consistent with the fixed location of the synthesized peptidoglycan. The key "events" during the division cycle are the initiation and termination of DNA replication and the initiation and termination of pole formation. There do not appear to be other cell-cycle-specific events or syntheses during the division cycle that are related to the regulation of the division cycle.

1 Growth of Bacterial Cultures

1.1 The Classical Growth Curve

For almost a century (since Buchanan, 1918), the classic growth curve has been used as a description of the pattern of bacterial growth. An overgrown culture is diluted into fresh medium and at first there is a "lag" phase, experimentally defined as a period before viable cell number begins to increase. Then cell numbers begin to increase and the culture enters the "log" phase. The log phase is the period of exponential increase in cell number. Although the cells during this period of steady state cell number increase are growing exponentially, the cells are generally referred to as being in "log" phase because of the usual plotting of cell number on semilogarithmic coordinates. During the log phase, plotting cell number increase on semilogarithmic paper yields a straight line that can be used to determine the doubling time of a culture. After the cells grow to higher concentrations, cell growth slows as the cells enter the stationary phase. Cell growth then ceases. Finally, if a culture is studied long enough, cells begin to lose viability or the ability to form colonies, and the culture enters the death phase.

During these phases, there are changes in cell sizes. The initial cells in the overgrown culture are small, cell size increases during the lag phase, cells are largest during the log phase, and then cell size decreases as cells enter the stationary phase. This classical view of the way in which cells grow is really a special case of the phenomenon of changes in cell size with changes in growth rate.

1.2 Bacterial Growth at Different Growth Rates

It is possible to grow a given bacterium in cultures at a wide range of growth rates. For example, it is possible to grow *E. coli* with a 300-min interdivision time with serine as the sole source of carbon, and a little faster than 20-min interdivision times in various complex broths available for bacterial culture. By varying the medium composition, it appears that any possible growth rate is allowable, so that there is a continuous range of allowable growth rates between the extremely slow and the extremely fast.

When bacteria are grown at a given temperature in different media at a number of different growth rates, a regular pattern of cell composition and cell size is observed. The faster a cell grows (richer medium) the larger the cell and the more the DNA, RNA, protein, and other components per cell (Fig. 1). When cell components are accurately measured, it is found that the material per cell increases exponentially as a function of the reciprocal of the doubling time (i.e. growth rate, as a shorter doubling time indicates a faster growth rate).

1.3 The Shift-up

During a shift-up from a slow growth medium (e.g. minimal medium, in which cells are small) to rapid growth in a rich medium (e.g. nutrient broth, in which cells are larger), there is a regular pattern of change in cell size and composition. Immediately after the shift-up, there is a rapid change in the rate of RNA and protein synthesis to the new rate of increase. There is a slower change in the rate of DNA synthesis. There is a delay before there is any change in the rate of cell number increase. Because of

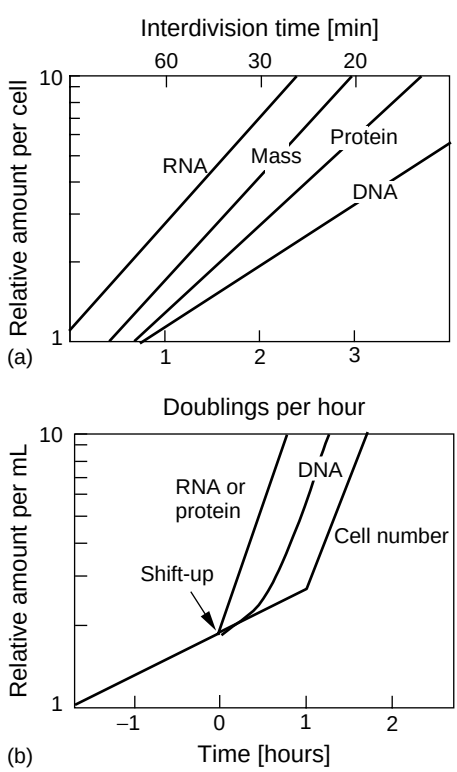

Fig. 1 (a) Composition of bacteria as a function of growth rate. (b) Synthesis of cell components during a shift-up from minimal medium to faster growth in richer medium.

the dissociation of mass increase and cell number increase, the average cell size increases during this first part of the shift-up. After a period of time, there is a sharp change in the rate of cell increase from the old rate to the new rate. This occurs 60 min after the shift-up (Fig. 1). Because there is a lag in the change in the rate of cell division, while cytoplasm synthesis increases to the new rate almost immediately, during the shift-up period there is an increase in cell size producing the larger-sized cells in the faster growth medium.

1.4 Classical Growth Pattern Is Merely a Series of Shift-ups and Shift-downs

The classical growth pattern for a culture can now be seen as a series of shift-ups and shift-downs. When an overgrown culture (nongrowing, and thus of essentially zero growth rate) containing small cells is diluted into fresh medium, there is a "shift-up" in growth rate. The mass increases immediately with no cell division occurring. The "lag" phase is thus one of a lag in cell number increase, but there is essentially no lag when considering mass increase. After the normal cell size is achieved for that medium, both the mass and the cell number increase in parallel. The rate of cell increase is determined by the medium supporting the more rapid growth. There is a constant larger-size cell during exponential growth. The reverse occurs as growth begins to slow down and the cells approach the stationary phase. The rate of mass synthesis decreases first, followed later by a cessation in the increase in cell number. Because division continues after mass increase ceases, the average cell size decreases in an overgrown culture.

2 The Bacterial Division Cycle

2.1 Cell Age and the Age Distribution during Balanced Growth

By convention, a newborn cell has an age of 0.0 and grows during the division cycle to divide at age 1.0. The cell ages during the division cycle are thus numbers between 0.0 and 1.0 indicating where, in the division cycle, a cell is located at a particular instant. A cell half-way between birth and division is age 0.5. Cell age, by definition, increases linearly during the division cycle. However, cells at the same absolute time after birth are only approximately the same cell-cycle age because of the variability in the absolute time required for the division cycle of an individual cell.

In a growing culture, the distribution of cell ages is not uniform. There are twice as many newborn cells as dividing cells. The age distribution is given by the formula $N = 2^{1-x}$, where x is the cell age between 0.0 and 1.0. This formula indicates that there are twice as many newborn cells ($x = 0.0$) as dividing cells ($x = 1.0$). If all cells grew with exactly the same interdivision time, this formula would give the age distribution exactly. However, because of cycle variability there is a smoothing of the function and the actual distribution is an approximation of the theoretical distribution. Because of the age distribution, the properties of an exponentially growing culture are independent of time. This means that all patterns of synthesis within the division cycle give an exponentially increasing amount of material in a growing culture. This is because the age distribution is constant during steady state, exponential growth. Thus, whether something is

made linearly, exponentially, or at an instant during the division cycle, in an exponentially growing culture the amounts of all cell components increase exponentially and in parallel. This is referred to as balanced growth. The observed balanced growth is the result of the age distribution during growth being invariant.

Given that between the birth of a cell and its subsequent division, all of the cell components double, we can ask, ''At what rate is material synthesized during the division cycle of an individual cell, and how does the cell ensure that there is a precise doubling of cell components by the time the cell divides?'' Let us first analyze which components must be dealt with in terms of the division cycle.

2.2 Aggregation Theory and the Control of Synthesis during the Division Cycle

In economics, the aggregation problem is how to combine various sectors of an economy. Should the figures for the production of capital machinery be combined with those for the production of consumer goods? Is paper produced for boxes in the same economic category as stationery? It is difficult to treat each item in an economy individually; some aggregation is necessary in order to understand the whole system.

Now, consider the aggregation problem for the analysis of the bacterial division cycle. How should one aggregate the different cell components in order to achieve an understanding of the biochemistry of growth and division? Is there a unique pattern of synthesis during the division cycle for each enzyme, or are there a limited number of patterns with different enzymes or molecules synthesized according to any one of these patterns? Are there ways of grouping proteins or RNA molecules so that one can consider classes of molecules rather than individual molecular species? Should we consider the cell membrane as a different category from that of peptidoglycan? There are approximately a thousand proteins in the growing cell, and if each protein had a unique cell-cycle synthetic pattern, or if there were only a few enzymes exhibiting any particular pattern, we would have an insuperable task describing the biosynthesis of the cell during the division cycle.

Fortunately, one need consider only three categories of molecules, each of which is synthesized with a unique pattern. The growth pattern of the cell is the sum of these three biosynthetic patterns. The first category is the cytoplasm, which is the entire accumulation of proteins, RNA molecules, ribosomes, small molecules, water, and ions that make up the bulk of the bacterial cell. It is the material enclosed within the cell surface that is not the genome. The second category is the genome, the one-dimensional linear DNA structure. The third category is the cell surface, which encloses the cytoplasm and the genome. The surface is composed of peptidoglycan, membranes, and membrane-associated proteins and polysaccharides. Everything in the cell fits into one of the three categories, and each category has a different pattern of synthesis during the division cycle. These three patterns are simple to understand as they can be derived from our current knowledge of the principles involved in the biosynthesis of the cytoplasm, the genome, and the cell surface.

The division cycle of bacteria in general may be best described by discussing the division cycle of the bacterium about which most is known, *E. coli*, and then describing

other bacteria and their patterns of growth and division with respect to this archetypal pattern. The division cycle of *E. coli* will be described, indicating how the various components of the cell are made and distributed into two new daughter cells.

3 Synthesis of Cell Components during the Division Cycle

3.1 Cytoplasm Synthesis during the Division Cycle

3.1.1 Exponential Cytoplasm Synthesis during the Division Cycle

Consider a unit volume of cytoplasm. It contains enzymes required for the breakdown of nutrients, enzymes for the production of energy for biosynthesis, enzymes for the biosynthesis of low molecular weight precursors of macromolecules, and enzymes for the synthesis of macromolecules. Each unit of cytoplasm produces a small amount of new cytoplasm over a small interval of time. If the new cytoplasm is indistinguishable from the old, and if the new cytoplasm acts to synthesize more cytoplasm immediately, then the pattern of cytoplasm synthesis is exponential. This is because after each interval of synthesis, the rate of synthesis increases owing to the new increment of material added to the cytoplasm. This exponential pattern is illustrated in Fig. 2 with both the rate of synthesis and the amount of material increasing exponentially during the division cycle.

The pattern for cytoplasm as a whole is reflected in the synthesis of the individual cytoplasmic components. There are no changes in the specific rate of synthesis or pattern of cytoplasm synthesis that are related to any particular cell-cycle event. (Actually, there is a minor *caveat* to this generalization, which is treated below.) No particular molecule of the cytoplasm is made differently from any other molecule of the cytoplasm during the division cycle. In addition, there is no variation, during the division cycle, in the relative concentration of any component of the cytoplasm. All parts of the cytoplasm accumulate exponentially during the division cycle, and this pattern is unchanged even when the cell is dividing. At the instant of division, the combined rate of cytoplasm synthesis in the two new daughter cells is precisely equal to the synthetic rate in the dividing mother cell.

One minor exception to the rule that specific proteins are made exponentially during the division cycle regards the effect of momentary sequestration of DNA upon replication. Specific regions of DNA are sequestered in the membrane at different times during the cell cycle at the time of their replication. During this sequestration period, the DNA is not available for transcription. Thus, for all genes there

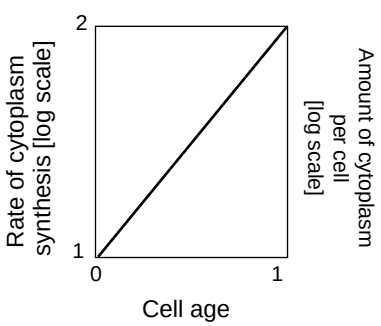

Fig. 2 Cytoplasm synthesis during the division cycle. Both the rate of cytoplasm synthesis and the pattern of accumulation of cytoplasm are exponential during the division cycle.

is a time during the division cycle at which gene expression is prohibited. This momentary cessation of transcription should not obscure the more fundamental point that there does not appear to be any specific control of gene expression for any gene during the division cycle that is related to the regulation of that gene during the division cycle. One should imagine this cessation of gene expression as merely a necessary accommodation of transcription to the needs of the cell for DNA replication. It could be imagined that if the cell were able to replicate DNA without any cessation of transcription, there would be no reason to expect that the cell would alter specific gene expression at any point during the division cycle.

The evidence for a smooth, continuous, and exponential increase in cytoplasm comes from experiments that use the membrane-elution method, colloquially referred to as the *baby machine* (see below). The pattern of DNA synthesis during the division cycle, which was subsequently shown to be correct by flow-cytometric and other analyses, was actually discovered using the membrane-elution method. Thus, the results from a method capable of producing an accurate analysis of biosynthesis during the division cycle have confirmed the exponential synthesis of cytoplasm.

In addition to this experimental support, an evolutionary argument can be made for an invariant rate of accumulation of cytoplasm during the division cycle. If the synthetic rate of an enzyme changed abruptly during the division cycle, there would be a relative excess or deficiency of that enzyme at some point during the division cycle; this would be an inefficient use of resources. The optimal allocation of resources is for each component to be at a constant concentration during the division cycle. Cells would not evolve controls that make them less efficient in cell production. The ideal pattern of cytoplasm synthesis is an invariant cytoplasm composition during the division cycle.

A third argument supporting exponential cytoplasm synthesis is that only this pattern is explained in known biochemical terms. The cycle-independent exponential synthesis of cytoplasm can be derived from our current understanding of the biochemistry of macromolecule synthesis. Enzymes make RNA and protein, which make ribosomes, which then lead to more protein synthesis. More proteins mean more RNA polymerases, more catabolic and anabolic enzymes, and the continuously increasing ability of the cell to make more and more cytoplasm. In contrast, if newly synthesized material were *not* active or available for synthesis when made, but instead was recruited for biosynthesis only at the instant of division or at some specific cell-cycle age, one would have linear cytoplasm synthesis during the division cycle. There are no known mechanisms enabling the cell to distinguish between newly synthesized cytoplasm and old cytoplasm, or which lead to cell-cycle-specific variation in the synthetic rate of any particular molecule during the division cycle. We therefore conclude that in theory as well as in practice, cytoplasm increases uneventfully and exponentially during the division cycle.

3.1.2 Variation in the Rate of Cytoplasm Increase

The continuous variation in allowable growth rates may be understood by partitioning the cytoplasm into those molecules involved in cytoplasm synthesis and that are absolutely necessary for growth (ribosomes, RNA polymerases, etc.) and those molecules that are involved in dispensable

syntheses (e.g. enzymes involved in amino acid biosynthesis). In a minimal medium, in which many dispensable enzymes are being synthesized, the ribosomes may be considered to be involved in the synthesis of ribosomal protein and a large number of dispensable proteins. Since only a fraction of the ribosomes are making ribosomal protein at any moment, the time for a doubling in the ribosomal content is relatively long. As one increases the number of nutrients in the medium, thus relieving the cell of having to make the dispensable enzymes, a larger fraction of ribosomes are involved in ribosomal protein synthesis. Thus, in rich medium the time for ribosome number to double is shorter than in poor medium. The variation in growth rates is thus a reflection of the continuous ability of the cell to partition the ribosomes (or the RNA polymerases) between making ribosomal components or nonribosomal components.

Considering the time required for a ribosome to make another ribosome (i.e. using 15 amino acids incorporated in a protein chain in a second), one can calculate that it would take eight minutes for one ribosome to synthesize a single ribosome. If a cell were in some way supplied with all proteins other than ribosomes, one could consider that the fastest rate of cell doubling would be eight minutes, the time for ribosomes to double their number.

3.2 DNA Synthesis during the Division Cycle

One of the most important generalizations regarding the regulation of linear macromolecule synthesis is that the rate of synthesis is regulated at the point of initiation of polymerization. This principle is exhibited most clearly in DNA replication, where the process of initiation of replication can be clearly dissociated from the continued replication of DNA that has already initiated replication. If one inhibits the initiation of replication, then the rounds of DNA replication in progress will continue until their normal termination.

3.2.1 Pattern of DNA Synthesis during the Division Cycle

The DNA synthesis pattern during the bacterial division cycle (specifically for *E. coli* and related bacteria) is composed of one or more periods of constant rates of DNA accumulation (Fig. 3). In cells with a 60 min interdivision time, there is a constant rate of DNA synthesis for the first 40 min, with one bidirectional pair of replication forks, followed by a zero rate of synthesis during the last 20 min. In cells with a 30 min interdivision time, there is a constant rate of DNA synthesis for the first 10 min, which falls to a constant rate that is two-thirds of the initial rate for the last 20 min. In cells with a 20 min interdivision time, there is a constant rate of DNA synthesis throughout the division cycle. Other growth rates have similar particular patterns of DNA synthesis.

The constant rates of DNA synthesis can be understood in terms of the molecular aspects of DNA synthesis. DNA is synthesized by the movement of a replication point along the parental double helix, leaving two double helices in its wake. The rate of movement of a growing point appears to be invariant and independent of its location in the genome. This means that the rate of DNA synthesis is proportional to the extant number of replication points. The number of replication points is constant for any period of the division cycle. Thus, the rate of DNA synthesis during any period of the division cycle is also constant. The

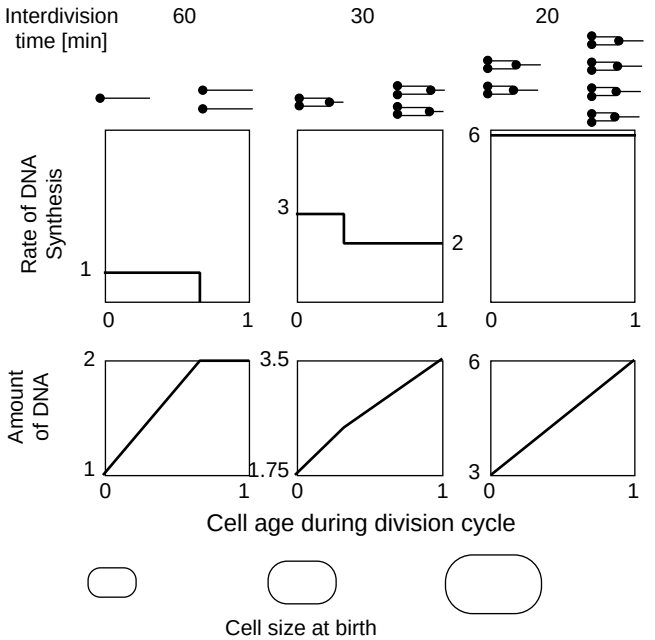

Fig. 3 DNA synthesis during the division cycle. Three different patterns are illustrated for cells growing with 20, 30, and 60 min doubling times. For each growth rate, the time from initiation to termination of DNA replication is 40 min. The time between termination of replication and cell division is 20 min. The proposed chromosome patterns at the start and finish of the division cycle are illustrated above the graphs. In addition to the rate of DNA synthesis during the division cycle, the pattern of accumulation of DNA during the division cycle is presented. Accumulation of DNA is composed of periods of linear synthesis. The rates during these periods are proportional to the existing number of growing points. The graph of the rate of synthesis is the differential of the accumulation plot. A representation of the cell size expected for cells at the start of the division cycle is presented below each synthetic pattern. The size of the newborn cells is in the ratio of 1 : 2 : 4 in the three cultures illustrated here. The number of new initiations at the start of each division cycle is also in the ratio of 1 : 2 : 4.

pattern of constant rates of synthesis is derived from, and is consistent with, our understanding of the biochemistry of DNA synthesis.

Embedded in the patterns of Fig. 3 are the two major rules for the determination of the pattern of replication of DNA during the bacterial division cycle. The first rule is that once DNA replication is initiated at the origin of replication, the time for a complete round of replication – the C-period – is relatively invariant (40 min in *E. coli* at 37 °C). The second rule is that the time between termination of replication and cell division – the D-period – is also relatively invariant.

The constant C- and D-periods are most clearly evident in the 60 min cells in Fig. 3.

At other growth rates, the patterns of synthesis appear more complicated, but this is due to the possibility of initiation occurring before completion of previous rounds of replication and the possibility of initiation of replication occurring in one division cycle and replication of the chromosome being completed in a subsequent division cycle.

3.2.2 Regulation of Initiation of DNA Replication

Why is replication initiated at these particular times during the division cycle? How is the initiation process regulated? Rounds of replication are initiated in a bacterial cell when the amount of cell mass or something proportional to the cell mass (the "initiator") is present at a certain amount per origin. Initiation of new rounds of replication occurs once per division cycle when the mass (or "initiator") per origin reaches a fixed value. In the 60 min cell, initiation occurs at one origin with a cell of unit size (size 1.0). In the 30 min cells, there are two origins in the newborn cell and the size of the newborn cell is twice that of a 60 min cell. And the 20 min cell has four origins initiated in the newborn cell and the cell size is four times the newborn 60 min cell. Thus, not only does the DNA per cell increase with growth rate but the average cell size increases with growth rate as well. In the cells shown in Fig. 3, the relative sizes of the newborn cells, as well as the average cell in the culture, are in the ratio of 1 : 2 : 4. This increase in cell size with growth rate is consistent with the exponential increase in cell size as a function of the inverse of the doubling time.

The actual nature of the initiator is not known, but a number of candidate molecules have been identified. Among these, the DnaA protein appears as a good candidate for the regulator of DNA initiation.

Alternative models of initiation have been proposed, such as the sudden accumulation of an inhibitor of initiation that is diluted out by cell growth. This model does not accommodate the constant synthesis of cell-cytoplasmic components. In all cases, it is indistinguishable from the initiator accumulation model, and there is no evidence for such an inhibitor being synthesized during the division cycle.

3.2.3 The Specific Origin of Replication

A specific sequence of DNA in the bacterial genome is the origin of replication (oriC). The required sequence (for *E. coli*) is 245 bp. Within this sequence are 5 highly conserved sequence blocks of 15 to 20 bp. These conserved regions are interspersed with sequences that are random. Regions of random sequence but of fixed length separate the conserved regions. Thus, the length of separation of conserved sequences is more important than the specific sequence between the conserved sequences. Four sequences of 9 bp are specific recognition sites for the binding of dnaA, the specific initiator protein. In addition, there are three 13 bp regions that are sites for opening the duplex DNA for initiation. Finally, there are 14 GATC sequences located in the origin (oriC) region and the neighboring regions. These 14 four bp sequences are expected to occur at a much lower frequency if they were not a necessary part of the *E. coli* origin.

The ultimate proof that this sequence is the origin of replication is the production of minichromosomes, small replicating circles of DNA that have the bacterial origin of replication. An *E. coli* cell can have many copies of the minichromosome

without any apparent impairment of cell growth.

3.2.4 Protein Factors Controlling DNA Replication

Initiation of replication of DNA is a function of numerous proteins. The key regulatory protein appears to be the product of the *dnaA* gene. The dnaA protein binds to the 9 bp "dnaA boxes" in the oriC region. After binding to the origin, the protein promotes melting of the DNA (i.e. separation of the DNA strands in a local region) and helps in loading additional proteins such as the dnaB protein. After this initial opening of the DNA strands, replication fork movement is allowed to proceed bidirectionally from the origin.

Initiation of replication is distinct from the continued replication of DNA that has been initiated. Once initiated, DNA replication can continue even though the accumulation of cell cytoplasm is inhibited.

3.2.5 Chromosome Organization and Bidirectional Replication

The phenomenon of bidirectional replication has two effects on the organization of the chromosome. One effect is the variation in the relative abundance of G residues versus C residues in one of the strands. While $G = C$ in the double helix, such an equality need not exist in a given strand or in a particular region of a strand. As a general rule, it is found that in the two halves of the bidirectionally replicated bacterial genome, one half of the chromosome (e.g. that from noon past three down to six on a clock representation of the genome) has $G > C$ and the other half (noon, past nine, down to six) has $G < C$. If one calculates the $G - C/G + C$ for a running window of nucleotides, and accumulates the value of this "GC-skew," there is a valley that corresponds to the origin and a peak that corresponds to the terminus. As long as one reads the strand in the 5′ to 3′ direction, the location of the peak and the valley is independent of the strand being used to define the origin and the terminus.

The bidirectional organization of the genome also affects the direction of gene transcription. As a general rule (with numerous exceptions), the direction of gene transcription is in the same direction as the movement of the replication fork. The rationale for this phenomenon is that this correspondence of direction of replication and transcription reduces the "head-on collision" of replication and transcription. The fact that there are many exceptions means that the replication system can accommodate running into transcriptional units without a problem.

3.2.6 DnaA Protein Is Involved in Initiation of DNA Replication

The DnaA protein, identified by mutants affecting the initiation of DNA replication and not the progression of rounds of replication already in process, appears to be the ultimate regulatory protein with regard to the initiation of DNA synthesis. The initiation process involves the cooperative binding of approximately 20 DnaA protein molecules to the origin of replication. This complex then leads to the open complex in which a region of AT-rich 13-mer repeats denatures so that single-stranded DNA is exposed. Other proteins such as the DnaB helicase are now introduced into the forks of the melted DNA where it expands the region of single-stranded DNA. Other molecules then join the initiation complex with the final loading of DNA polymerase III. When this occurs, the replication of DNA progresses

around the genome to the opposite end of the genome where replication terminates.

3.3 Surface Synthesis during the Division Cycle

3.3.1 Peptidoglycan Synthesis during the Division Cycle

The cell surface is made to perfectly enclose, without excess or deficit, the cytoplasm synthesized by the cell. Cell cytoplasm increases continuously during the division cycle (described above), and therefore the cell surface is made continuously. How is the bacterial cell surface made and duplicated during the division cycle? What is the rate and topography of cell-surface synthesis during the division cycle?

Consider an imaginary cell in which the cytoplasm is enclosed in a cell-surface tube that is open at each end; the cytoplasm remains within the bounds of the tube. The cytoplasm in the newborn cell is thus encased in the cylinder of cell surface that is made up of membrane and peptidoglycan. As the cytoplasm increases exponentially (see above), the tube length increases to exactly enclose the newly synthesized cytoplasm. In this imaginary case, the cell surface increases exponentially in the same manner as the cytoplasm. When the cell cytoplasm doubles, the tube divides into two new cells and the cycle repeats. In this imaginary cell, the cytoplasm increases exponentially, the internal volume of the cell increases exponentially, the surface area increases exponentially, and the density of the cell, that is, the total weight of the cell per cell volume, is constant during the division cycle.

However, a real rod-shaped cell does have ends, and therefore the pattern of cell-surface synthesis during the division cycle is not simply exponential. If the cell surface were synthesized exponentially, the cell volume cannot increase exponentially. Since the cytoplasm increases exponentially, there would have to be a change in cell density. (In theory, it may be that major changes in cell shape could allow a pure exponential increase in cell mass and cell surface, but such changes are clearly not observed during the cell-division cycle.) Cell density, however, is constant during the division cycle. A proposal for cell-surface synthesis that allows an exponential increase in cell volume, and therefore a constant cell density, is presented in Fig. 4. Before invagination, the cell grows only in the cylindrical sidewall. After invagination, the cell grows in the pole area and the sidewall. Any volume required by the increase in cell cytoplasm that is not accommodated by pole growth is accommodated by an increase in sidewall. The cell is considered a pressure vessel, and when the pressure in the cell increases, there is a corresponding increase in cell surface area. The pole is preferentially synthesized and any residual pressure due to new cell cytoplasm that is not accommodated by pole growth is relieved by an increase in cylindrical sidewall area.

The resulting pattern of synthesis is approximately exponential. The formula describing surface synthesis during the division cycle is a complex one, including terms for the shape of the newborn cell, the cell age at which invagination starts, the pattern of pole synthesis, and the age of the cell. One consequence of this model of surface synthesis is that at no time is surface synthesis exponential, since the rate of surface synthesis is not proportional to the amount of surface present over any period of time.

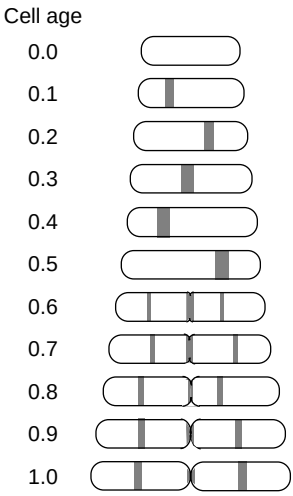

Fig. 4 Cell-surface synthesis during the division cycle. Before invagination, the cell grows only by cylinder extension. Each cell is drawn to scale, with the volume of the cells increasing exponentially during the division cycle. The shaded regions of the cell indicate the amount and location of wall growth (whether in pole or sidewall) each tenth of a division cycle. The width of the shaded area is drawn to scale. Cell-surface growth actually occurs throughout the sidewall and not in a narrow contiguous zone. The variable locations of wall growth in this figure have no specific meaning, but are meant to show synthesis occurring all over the sidewall during the division cycle. Before invagination, the ratio of the surface increase to volume increase is constant. When pole synthesis starts, at age 0.5 in this example, there is an increase in the ratio of the surface increase to the volume increase. Any volume not accommodated by pole growth is accommodated by cylinder growth. At the start of pole growth, there is a reduction in the rate of surface growth in the cylinder. As the pole continues to grow, there is a decrease in the volume accommodated by the pole and an increase in the rate of growth in the sidewall.

As with cytoplasm and DNA synthesis, we can derive the cell-cycle pattern of surface synthesis from our understanding of the molecular aspects of peptidoglycan synthesis. The peptidoglycan sacculus of the cell is composed of glycan strands encompassing the cell, perpendicular to the long axis, as shown in Fig. 5. These strands are cross-linked by peptide chains. Because the cross-linking in the load-bearing layer is not complete, one can have new strands in place, below the taut load-bearing layer, prior to the cutting of the bonds linking old strands. As shown in Fig. 5, the load-bearing layer is stretched by the turgor pressure in the cell. An infinitesimal increase in cytoplasm leads to an infinitesimal increase in the turgor pressure of the cell. This increase in turgor pressure places a stress on all of the peptidoglycan bonds, thus reducing the energy of activation for bond hydrolysis. The result is that there is an increase in the cutting of stressed bonds between the glycan chains. When a series of cuts is made, allowing the insertion of a single new strand into the load-bearing layer, there is an infinitesimal increase in cell volume. This increase in volume leads to a reduction of the stress on the remaining bonds. By a continuous repetitive series of cytoplasm increases, surface stresses,

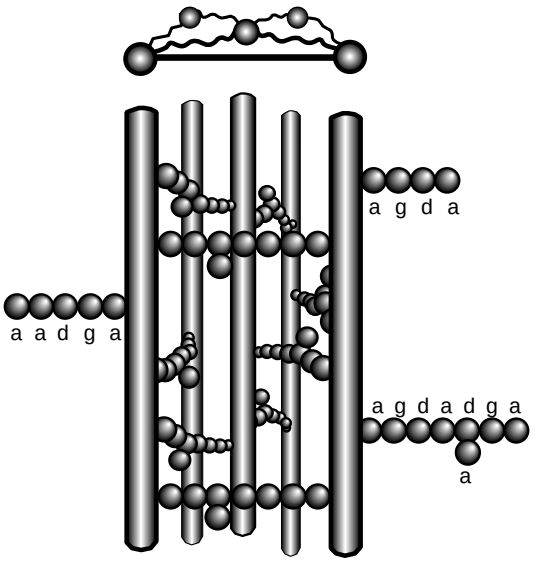

Fig. 5 Three-dimensional representation of peptidoglycan structure. This is an idealized representation of peptidoglycan structure as seen from the outer surface. The thick bars represent glycan chains at the outside of the peptidoglycan layer. The thinner bars represent glycan chains below the outer layer. The stretched chains of circles represent amino acids cross-linking the glycan chains. The chains below the stretched surface of the cell rise to the outer layer when the taut layers of the peptidoglycan are hydrolyzed. The figure shows a cross-sectional view through the glycan chains illustrating the taut outer layer and the more loosely inserted inner material. The letters indicate the amino acids composing the cross-links; a = alanine, d = diaminopimelic acid, and g = glutamic acid, with italicized letters present in the D-configuration.

enzymatic cuts, and volume increases, one gets an increased cell volume that precisely accommodates the increase in cell cytoplasm.

An inside-to-outside mode of peptidoglycan growth, similar to that for *Bacillus subtilis*, has been proposed for *E. coli*. This proposal is based on the observed recycling of murein, the calculated amount of peptidoglycan per cell, and the existence of trimeric and tetrameric fragments that are consistent with a multilayered peptidoglycan structure. The insertion of new peptidoglycan strands in an unstressed configuration prior to their movement into the load-bearing layer of the peptidoglycan can explain all of these observations. The recycling of peptidoglycan may be a strain-specific result, as there is no major release or recycling of peptidoglycan in *Salmonella typhimurium* or in *E. coli* B/r. At this time, the inside-to-outside mode of surface growth cannot be excluded.

Although this discussion of the rate of surface synthesis during the division cycle has dealt primarily with peptidoglycan, it applies equally to membranes and other surface-associated elements. The cell membrane grows in response to the increase in peptidoglycan surface and coats the peptidoglycan without stretching or buckling. The area of the membrane increases in the same way in which bacterial peptidoglycan increases.

One must make distinctions among cell components with regard to their location in the cell, rather than with regard to their chemical properties. Cell proteins can be divided into two categories – those associated with the cytoplasm and those associated with the surface. Proteins associated with the membranes are included in the surface category of synthesis during the division cycle. If a bacterial cell had a histone-like protein associated with DNA, there would be a third category of protein-synthetic patterns, proteins synthesized

during the division cycle with constant rates of synthesis. Proteins are synthesized during the division cycle with a pattern that is presumably consistent with their final location or category. Proteins are not a monolithic group nor are they divided into a large number of groups. The three-category system proposed here allows us to conceptualize the possible patterns of protein synthesis observed during the division cycle.

3.3.2 The Invagination Process

The division of the bacterial cell into two daughter cells occurs by the formation of two new poles in the middle of the dividing cell. As there are very few DNA-less cells produced during balanced growth, the division process must be coordinated with the replication process so that division does not occur until two genomes are present. The invagination process that forms two poles in the middle of the cell involves a complex process whereby there is an inward growth of the peptidoglycan and the associated membrane of the cell surface.

One of the earliest events related to the invagination process is the localization of the FtsZ protein to the future site of pole formation. The FtsZ protein is localized to the division site in a pattern called the *FtsZ ring*. The FtsZ protein accumulates at the mid-cell site where division will occur before any visible invagination. During the invagination process, the FtsZ protein is located at the leading edge of the invagination process. Thus, the FtsZ ring continues to define a ring of decreasing size as invagination proceeds.

Other proteins necessary for invagination and division are apparently localized at the center of the cell because of their association with FtsZ.

3.3.3 FtsZ Protein Is Involved in Septum Formation and Cell Division

Recent results indicate that the tubulin-like FtsZ protein plays a central role in cytokinesis or cell division as a major component of a contractile cytoskeleton. An indication that the FtsZ protein is the key initiator of septation is that mutants of this protein that form filaments show no indication of any constriction. Other septation mutants presumed to be later in the pathway of septum formation do show some indication of septation. The FtsZ protein resembles the eukaryotic tubulin molecule and this similarity has led to speculations that the FtsZ protein has cytoskeletal properties that allow it to form the new pole in the middle of the cell.

3.3.4 Localization of the Division Site

The actual mechanism by which the division site is localized at mid-cell position is unknown. One model of the division-site localization proposes that the site exists prior to any observation of FtsZ localization. An alternative model proposes that the division site is determined in some way by the reaction of the cell surface to the presence of DNA; the absence of DNA in the center of the cell would thus lead to invagination-site localization.

An understanding of site location is strongly influenced by the existence of minicells. If invagination occurs, upon occasion, near a pole rather than in the center of a cell, a very small DNA-less "minicell" is produced that is essentially two poles without an intervening sidewall. The genetics of minicell formation is complicated, but the main idea is that the products of the min genes have a negative function and normally inhibit abnormal invagination sites. Thus, inactivation of any of a number of min genes leads to minicell

formation owing to the inappropriate location of invagination sites.

4 Plasmid Replication during the Division Cycle

4.1 High-copy Plasmids

If a cell has a large number (20–100) of plasmids per newborn cell, then one can consider that these plasmids replicate during the division cycle in proportion to those present (i.e. exponentially) and then assort themselves randomly into the newborn cells. Because the plasmid number is high, there is little chance of one of the daughter cells ending up with no plasmids and with all of the plasmids segregating to the other daughter cell. This pattern appears to be the case for naturally occurring, high-copy number plasmids.

One possible exception is the artificial plasmid made by cloning the origin (oricC) region of the bacterial chromosome. This appears to have a random assortment mechanism, but it is synthesized at a precise time during the division cycle that coincides with the normal time at which the chromosome initiates replication.

4.2 Low-copy plasmids

Low-copy plasmids are usually larger than high-copy plasmids and are present on the order of 1 to 2 per cell. Examples of such plasmids are the F-factor and the P1 plasmid. These replicate in a precise way at a particular time during the division cycle. The rules of their replication are similar to that regulating chromosome replication, with initiation of plasmid replication occurring when a fixed amount of cell mass is present per plasmid origin.

4.3 Minichromosome Replication during the Division Cycle

Minichromosomes replicate at the same time as normal chromosome initiation. The behavior of minichromosomes is paradoxical in that in a cell with numerous additional origins of replication there would be a competition for the "initiator" of replication. Such a competition would lead to abnormal cell sizes or some alteration in cell growth pattern. No such alteration is observed.

5 Segregation of Cell Components at Division

Producing the right amount of material so that the cell can divide is not enough for a successful division cycle. At division, the material in the mother cell must also be properly apportioned to the newborn cells. In a cell that divides to produce two equivalent daughter cells, the segregation problem is to ensure that both cells have the same amount of each portion of the cell material. Let us see how the three categories of material in the mother cell are segregated properly to the new daughter cells.

5.1 Segregation of Cytoplasm

The cytoplasmic components of the cell appear to segregate randomly at division. There does not appear to be any compartmentalization or hindrance to segregation of cytoplasm (Fig. 6).

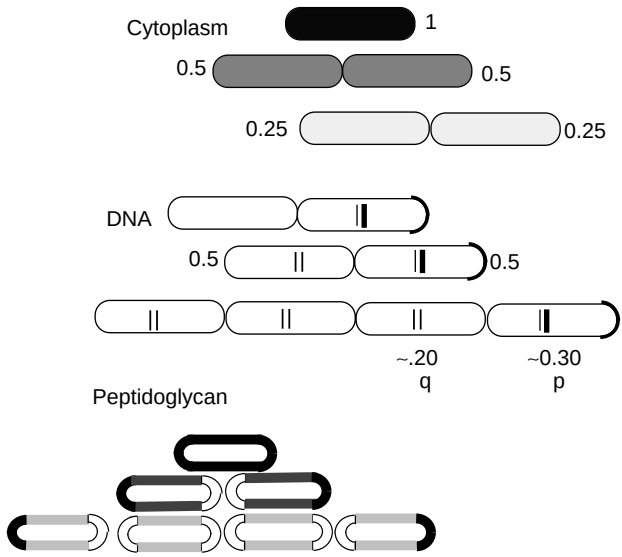

Fig. 6 Segregation of cytoplasm, surface, and DNA at division. The cytoplasm of the cell is segregated randomly at division. Each daughter cell shares equally in the parental cytoplasm (ribosomes, soluble proteins, ions, etc.), which can be considered randomly distributed throughout the dividing cell. The DNA is segregated nonrandomly, but each daughter cell gets a complete and equal complement of DNA from the dividing cell. The nonrandom segregation is at the level of a single strand. Since each DNA duplex is made up of an ''older'' strand and a new complement made on the older strand, and each cell is composed of an ''older'' pole and a pole just made at the previous division, one can see whether strands of one age go preferentially toward one or the other age pole. Experimental analysis indicates that the DNA strands segregate as though the older DNA strand goes preferentially toward the pole that it went to at the previous division. This may be phrased as the newer strand goes toward the newer pole. The degree of nonrandom segregation is not perfect, but is stochastic, and is approximately 60/40 rather than 50/50. The surface of the cell segregates to the daughter cells at division such that the poles are conserved and the sidewall segregates with no apparent conservation of sidewall peptidoglycan. It is not clear whether insertion of new wall material over the side is perfectly uniform or whether there is some nonuniform pattern of insertion.

5.2 Segregation of DNA

DNA segregates so that each daughter cell receives a complete genome from the dividing mother cell (Fig. 6). As the DNA nucleoid is relatively compact within a dividing cell, if the nucleoids were apportioned completely randomly, one might expect that at some divisions one of the daughter cells would not receive any DNA and the other daughter cell would

receive two genomes. That this is a rare event supports the notion that segregation of the DNA is a regulated and deterministic process that leads to very few nonnucleated cells in a culture.

In addition to the segregation of individual nucleoids, the separate strands of the DNA have rules that describe their segregation and which are compatible with the deterministic equipartition of DNA to the two daughter cells. As all rod-shaped cells are asymmetric in that one of the poles is the new pole (produced at the last division) and one pole is necessarily older, one can see how older DNA strands and newer DNA strands (the newer strands having just been made on an older, preexisting strand) segregate with regard to pole age. The general rule is that the older strands segregate slightly preferentially to the older poles and the younger strands segregate preferentially to the newer or younger poles. This phenomenon is consistent with other data that indicate some attachment of the DNA to the cell surface.

5.3
Segregation of the Cell Wall

The peptidoglycan is a large macromolecule that grows by intercalation of new material between old material. Since all the peptidoglycan is covalently linked in one macromolecule, there is no ability for the peptidoglycan components to move with respect to each other. The segregation of cell wall material is therefore determined by the location of the cell wall material. Once in place, the wall segregates into the newborn cells so that the old poles go to each cell, and the sidewall is segregated to the newborn cells (Fig. 6).

6
Are There Events during the Division Cycle?

There are only four discrete events that occur during the division cycle of a growing rod-shaped bacterial cell. They are (1) the initiation of DNA replication, (2) the termination of DNA replication, (3) the initiation of pole formation, and (4) the completion of pole formation. Other aspects of cell growth are too continuous to be considered cell-cycle-specific events. For example, the addition of new nucleotides at position 43,567 on the chain of DNA may be considered a unique event, and one that occurs at a particular time in the division cycle, but because of cell-cycle variability, the time of this synthetic occurrence can never be precisely obtained. The time of insertion of a particular nucleotide pair has no meaning for the cell cycle; only the initiation and termination of DNA replication are important. With regard to the division cycle, the DNA is an amorphous material with no biochemical specificity along the strand.

Similar considerations apply to the cell surface and the cytoplasm. No aspect of peptidoglycan strand insertion is unique in time during the division cycle. A new strand is inserted at random sites in response to the stretching of the cell surface. This may occur in one cell at position 0.376 of the cell length; at the same time, in another cell, the new strand may be inserted at position 0.549. For an analysis of the cell cycle, we should consider pole and sidewall growth as uneventful extensions of the cell surface. Regarding the cytoplasm, one may consider the increase in ribosome content from 37,411 to 37,412 an event; but such an individual event is not measurable and is without meaning in the cell cycle. Thus, cytoplasm synthesis,

DNA synthesis, and cell-surface synthesis are uneventful during the division cycle.

In contrast to our understanding of the principles of cytoplasm, cell-surface, and DNA synthesis, the biochemical mechanisms of the four events occurring during the division cycle – the initiation and termination of DNA synthesis and the initiation and cessation of pole formation – are still unknown. While a great deal is known about the biochemistry of initiation of DNA replication, the mechanism by which this initiation is regulated is less well understood. In addition, very little is known of the biochemical principles involved in either the termination of DNA replication or the start and cessation of pole formation. A cell-surface structure, the periseptal annulus, has been proposed as a possible first step in the formation of a new pole. If the periseptal annulus is the start of pole formation, the important question is whether there are definable steps between the formation of the periseptal annulus and the start of invagination.

7
Do Checkpoints Control the Bacterial Cell Cycle?

In eukaryotic cell-cycle studies, the concept of "checkpoints" looms large. Checkpoints are the postulated functions of a cell that determine that one function is properly completed before a subsequent function occurs. Thus, in eukaryotic cells it is postulated that a checkpoint function determines that DNA has replicated completely and properly before allowing mitosis to occur. The checkpoint is thus a valuable function that prevents abnormal functions from occurring.

Studies of this type of phenomenon in bacteria can provide a corrective lens for understanding or more rigorously defining checkpoints. In bacteria, if DNA synthesis is inhibited, cells with only one genome (i.e. cells that have not completed replication of a total chromosome) do not divide. Even when DNA is inhibited or damaged, a cell with two genomes will divide, while a cell with one genome will not. Thus, in bacteria, DNA damage *per se* does not activate a checkpoint to prevent division. It is believed that the presence of a single genome in the center of a cell prevents division by its central presence, not the activation of a function that checks for completion of DNA replication.

The bacterial experience thus suggests that a distinction should be made between the existence of a function external to the tested function (i.e. some activity that checks on whether DNA replication has been completed) and the actual function itself acting as a brake on further activity. The failure to complete DNA replication is the inhibitor of division, and there is no checking by some "checkpoint" activity.

A similar analysis may be applied to the ubiquitous postulation of restriction points in eukaryotic cells. Restriction points are points or events postulated to exist in the G1-phase of the division cycle. Passage of a cell through the restriction point is a requirement for entering the S-phase and eventual division. The restriction point was defined by the effect of different starvation regimens on DNA replication and the resumption of DNA replication following release from starvation.

It is interesting that a bacterial restriction point was postulated to exist before the proposal of the eukaryotic restriction point, although the term *restriction point* was not used. Subsequent analysis of the bacterial restriction point indicated that the proposed control point was an experimental artifact related to leakage of

synthesis under starvation conditions. The eukaryotic restriction point is subject to the same critique.

8
Temporal Variation of the Bacterial Division Cycle

Division cycles in a culture are not precisely timed. There is a great deal of variability in division cycles, with a culture having a 60-min doubling time probably having cells with interdivision times between 40 and 100 min. The variability of the division cycle can be visualized as due to the variability of the separate components that make up the division cycle. Thus, for two identical newborn cells one can have the subsequent division occurring at different times because of (1) variability in mass synthesis in the two cells leading to variation in the time of initiation of DNA replication, (2) variation in the rate of DNA replication in the two cells, (3) variation in the time between termination and division, (4) and variation in the equality of division (which leads to variation in the next division cycle).

9
Alternative Bacterial Division Cycles

9.1
Caulobacter crescentus Division Cycle

At first glance, the division cycle of the stalked bacterium *Caulobacter crescentus* appears quite different from that of *E. coli*. At division, two different-sized and different looking cells are produced (division is not in the middle of the mother cell and one cell is stalked and the other is flagellated), the DNA patterns during the division cycle differ in each of the daughter cells, and there are cell-cycle-specific syntheses of flagellum-related materials. The most important difference is that the poles of the daughter cells are different in that one cell has a flagellum and the other has a stalk. Because of the production of two different appearing cells, the growth and division of *Caulobacter* has been described as a model of cell differentiation.

The division cycle of *Caulobacter* can be shown to correspond to that of the standard *E. coli* model by considering that the different DNA patterns are due to the different sizes of the two daughter cells arising by division not occurring in the middle of the dividing mother cell, and that the cell-cycle-specific pattern of expression of flagellum-related genes is related to the completion of poles after the act of cell division. Flagellum synthesis is thus considered a part of surface synthesis. The cycle-specific pattern observed in cells is related more to the window of time in which poles are made rather than synthesis at a particular time during the division cycle (Fig. 7).

9.2
Bacillus subtilis Division Cycle

The division cycle of the gram-positive bacterium *Bacillus subtilis* is relatively similar to that of *E. coli*. The major difference between the organisms lies in the fact that the *B. subtilis* cell wall is much thicker (giving rise to its gram-positive character), and this requires an inside-to-outside pattern of peptidoglycan growth. The cell wall is made adjacent to the cytoplasmic membrane, and this wall layer moves out as subsequent layers are made between it and the cytoplasmic membrane. This leads to the multilayered *B. subtilis* peptidoglycan.

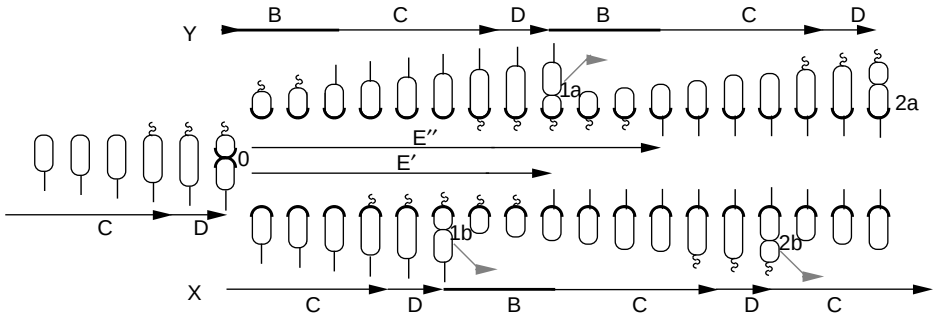

Fig. 7 The *Caulobacter* cell cycle. *Caulobacter* cells divide unequally, producing a smaller swarmer cell that is motile and a larger stalked cell that is generally attached to a surface by the stalk. As drawn here, the development of *Caulobacter* can be followed by concentrating on the two new poles formed by the first division. The development of the poles in both cells is a succession from bald, to flagellated, to stalked. Because the swarmer cell is smaller the development is slightly delayed, suggesting a size control for both the initiation of DNA replication and the completion of pole development. As drawn here, the cell-cycle-specific synthesis of flagella protein (flagellin) is a result of pole development following cell separation. The ultimate end of pole development is a stalk, after which no further changes in the pole occur. Thus, with regard to the general model of cytoplasm synthesis proposed here, the *Caulobacter* cytoplasm (excluding surface) does not exhibit cell-cycle-specific patterns of gene expression.

9.3
Streptococcal Growth during the Division Cycle

The growth of spherical bacteria such as Streptococci reveals a different pattern of surface growth. There is no apparent intercalation of new peptidoglycan material into the preexisting older peptidoglycan. Rather, the old peptidoglycan is conserved and new peptidoglycan grows at the center of the cell to make a completely new cell pole.

10
Mass Increase Is the Driving Force of the Division Cycle

The driving force of the division cycle is cytoplasm synthesis. Some part of the energy used by the cell to make cytoplasm drives the biosynthesis of the cell surface by causing stresses along the cell surface and this stress leads to the breaking of peptidoglycan bonds. The insertion of new peptidoglycan then leads to the increase in cell size. Cytoplasm increase also regulates DNA synthesis. Cell size at birth is greater at faster growth rates (Fig. 3) because the initiation of new rounds of DNA replication occurs when the cell has a unit (or critical) amount of cytoplasm per origin of DNA. Cell size at initiation is proportional to the number of replication points in the cell, and that is why cells are larger at faster growth rates; at faster growth rates there are more origins per cell at initiation. The cell "titrates" the amount of cytoplasm or some specific molecule that is a constant fraction of the cytoplasm. Thus, DNA synthesis is initiated at all available origins when the amount of cytoplasm per origin reaches a particular value.

The cell cytoplasm increases as fast as it can, given the external nutrient conditions. The synthesis of DNA and cell

surface cannot outpace or lag behind cytoplasm synthesis. DNA synthesis cannot go faster because it is waiting to initiate new rounds of DNA replication in response to cytoplasm synthesis nor can it go slower because the cytoplasm would increase without initiating new rounds of DNA replication. Cell surface is made to just enclose the newly synthesized cytoplasm. The regulation of surface and DNA synthesis by cytoplasm thus explains why there is no dissociation of these syntheses during the division cycle and during the growth of a culture. Furthermore, the failure to complete DNA replication (so that a cell has only one genome) prevents cell division; the cell thus ensures that there will be enough material (genome, cytoplasm, and cell surface) to allow the production and survival of two new daughter cells.

11 The Bacterial Growth Law during the Division Cycle

It has been speculated that there is a "bacterial growth law" that can be discovered by sensitive methods of analysis. Does the cell grow linearly, bilinearly, exponentially, or perhaps follow some other, yet undiscovered, growth law? If there were a general law of cell growth, independent of the biosynthetic patterns of the three categories that comprise the cell, then the individual categories of biosynthesis would have to accommodate themselves to this overall growth law.

There is no simple mathematical growth law regulating or describing bacterial growth during the division cycle. Bacterial growth during the division cycle is the simple weighted sum of the biosynthetic processes that are described by the three categories proposed here. As presented in Table 1, it can be seen that the components of the cytoplasm of the cell are 80% of the total cell weight. This means that the growth of the cell is approximately exponential. A slight deviation from exponential growth is due to the contribution of cell surface and DNA synthesis. Cell growth, then, is the result of a large number of individual reactions, regulated by local conditions, and not conforming to any overriding growth law.

Tab. 1 Composition of *Escherichia coli*, a typical gram-negative, rod-shaped bacterium.

Cell component	*Molecules per cell*	*Percent dry weight*
Cytoplasm		
Protein	2,350,000	55.0
RNA	255,480	20.5
Polyamines	6,700,000	1.1
Glycogen	4,300	3.5
Metabolites, ions, cofactors		2.5
Surface		
Peptidoglycan	1	2.5
Lipid	22,000,000	9.1
Lipopolysaccharide	1,430,000	2.4
Genome		
DNA	2.1	3.1

If one wished to state the growth law, the growth law is simply the sum of the individual synthetic patterns of the three categories of the cell material.

12
Experimental Analysis of the Bacterial Division Cycle

How can one study the pattern of synthesis of cell components during the division cycle? At first thought, it appears that one could study the division cycle by preparing a synchronized bacterial culture – that is, a culture composed of cells all of the same age, for example, newborn cells – and taking measurements of the cell components at different times during the division cycle. It turns out that this method has rarely, if ever, led to any concrete results. This is presumably due to the introduction of perturbations and artifacts when cells are synchronized, as no naturally synchronized populations are available.

The methods that have given results are nonperturbing, nonsynchrony methods. One is the membrane-elution method, or "baby machine," which analyzes the division cycle of unperturbed cells. In this method, a label is added to a growing culture, the labeled cells are filtered onto a membrane, the cells bind as though bound at one end or pole, and when medium is pumped through the inverted membrane, newborn cells are eluted from the bound cells. Since the newborn cells arise from the bound cells as a function of the original cell age, the first newborn cells arise from the oldest cells at the time of labeling. With time, the newborn cells arise from cells labeled at younger and younger cell ages. By measuring the amount of label per cell during elution, one can determine the pattern of synthesis during the division cycle. This has been done successfully for DNA, plasmids, protein, RNA, peptidoglycan, and cell membrane.

Another method for cell-cycle analysis is flow cytometry, where the amount of material per cell in a cell population is determined for individual cells by a complex system of liquid handling and laser-excitation engineering. The quantitative distribution of cell material can give the relative rate of synthesis during the division cycle. This method has been successful for DNA replication during the division cycle and has confirmed the initial results obtained with the membrane-elution method. It is not clear whether the method is sensitive enough for bacteria to enable the discovery of a particular pattern of DNA replication.

13
Genetic Analysis of the Bacterial Cell Cycle

Complementing the physiological and biochemical analyses of the bacterial division cycle has been a large amount of work on isolating and analysing mutants that affect the division cycle. Because a successful division requires that the cell make all of its required material, a mutation affecting the synthesis of any required material is, in some trivial sense, a mutation affecting the division cycle. Thus, a mutant unable to make a particular amino acid, such as leucine, that is required for growth would stop progressing through the division cycle. However, this would be a trivial relationship to the division cycle because the leucine is required throughout the division cycle and is not related to a specific control mechanism.

More interesting are those mutants that affect the major events of the division cycle

such as the start of DNA synthesis and the start of septation. A large number of mutations affecting these processes have been isolated, and their study has revealed that a number of enzymes and proteins are required for cell-surface synthesis and DNA synthesis.

14
Die Einheit der Zellbiologie

Are the interrelationships of cell-cycle regulation, proposed here for *E. coli*, applicable to other organisms? Other regulatory mechanisms have been proposed in other cell types. For example, a circular regulatory system was proposed for *Caulobacter* and a stochastic mechanism as well as a rate-controlled model for animal cells. Is the mechanism of cell-cycle regulation subject to historical accident? Or is there a unity of mechanisms regulating the division cycle of all types of cells? I conjecture that the *E. coli* model of cell-cycle regulation is generally applicable to other organisms, and that the optimal design of the cell cycle is exhibited by the *E. coli* cell cycle.

In the second decade of this century, Kluyver and his colleagues explicitly proposed the principle of the Unity of Biochemistry, or as originally proposed, Die Einheit der Biochemie (the oneness of biochemistry). Although the original proposal dealt primarily with the nature of oxidation–reduction reactions, the basic principle has evolved to encompass even more unifying principles. The nature of the gene, the genetic code, the mechanism of protein synthesis, the use of enzymes for biochemical changes, the patterns of enzyme regulation, and all of the fundamental aspects of biochemistry appear similar throughout the different organisms on earth. While there are exceptions and unique aspects to many organisms, the essential principle that there is a unified biochemistry has been supported.

Here we must now add, with regard to the cell cycle, the Unity of Cell Biology – *Die Einheit der Zellbiologie* – that is analogous to the Unity of Biochemistry. Phenomena suggesting the existence of other modes of cell-cycle regulation in animal cells have been shown to be consistent with the model for the bacterial division cycle proposed here. For example, the constancy of C- and D-periods is mimicked by the constancy of their analogues in animal cells, the S and G2 periods. The variation in cell size with growth rate and the rules regarding size determination also appear to hold for all organisms. The rules of the cell-division cycle – that is, the logic and design principles of the division cycle – are the same for *Escherichia* and escargot, *Salmonella* and salmon. Although there are many detailed differences between eukaryotic and prokaryotic division cycles, at the deepest level the principle of the Unity of Cell Biology proposes that the division cycle is ultimately regulated by the continuous accumulation of some molecule that is titrated against a fixed amount of another cell component. From this viewpoint, cell-cycle-specific events identified in eukaryotic cells are the result or symptom of a deeper regulatory principle that is produced in a cell-cycle-independent manner. The question for future analysis is whether there is some deeper rule that ensures this common pattern in all cells. I conjecture that this is indeed the case and hope that a search for the underlying meaning of size control in all cells is now the new goal of cell-cycle research.

See also Bacterial Cell Culture Methods; Bacterial Pathogenesis, Molecular Basis of.

Bibliography

Books and Reviews

Bramhill, D. (1997) Bacterial cell division, *Annu. Rev. Cell. Dev. Biol.* **13**, 395–424.

Cooper, S. (1984) The Continuum Model as a Unified Description of the Division Cycle of Eukaryotes and Prokaryotes, in: Nurse, P., Streiblova, E. (Eds.) *The Microbial Cell Cycle*, CRC Press, Boca Raton, Florida, pp. 7–18.

Cooper, S. (1991) Synthesis of the cell surface during the division cycle of rod-shaped, gram-negative bacteria, *Microbiol. Rev.* **55**, 649–674.

Cooper, S. (1991) *Bacterial Growth and Division: Biochemistry and Regulation of Prokaryotic and Eukaryotic Division Cycles*, Academic Press, San Diego.

Cooper, S. (1996) Segregation of Cell Surface Structures, in: Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M., Umbarger, H.E. (Eds.) *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edition, American Society for Microbiology, Washington, DC, pp. 1652–1661.

D'Ari, R. (2001) Cycle-regulated genes and cell cycle regulation, *BioEssays* **23**, 563–565.

Draper, G.S., Gober, J.W. (2002) Bacterial chromosome segregation, *Annu. Rev. Microbiol.* **56**, 567–597.

Gordon, G.S., Wright, A. (2000) DNA segregation in bacteria, *Annu. Rev. Microbiol.* **54**, 681–708.

Helmstetter, C.E. (1996) Timing of Synthetic Activities in the Cell Cycle, in: Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M., Umbarger, H.E. (Eds.) *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edition, American Society for Microbiology, Washington, DC, pp. 1627–1639.

Helmstetter, C.E. (1996) Timing of Synthetic Activities in the Cell Cycle, in: Neidhardt, F.C., Curtiss, R. III, Ingraham, J.L., Low, K.B., Magasanik, B., Reznikoff, W.S., Riley, M., Schaechter, M., Umbarger, H.E. (Eds.) *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington, DC, pp. 1627–1639.

Levin, P.A., Grossman, A.D. (1998) Cell cycle: the bacterial approach to coordination, *Curr. Biol.* **8**, R28–R31.

Lutkenhaus, J., Addinall, S.G. (1997) Bacterial cell division and the Z ring, *Annu. Rev. Biochem.* **66**, 93–116.

Lutkenhaus, J., Mukherjee, A. (1996) Cell Division, in: Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M., Umbarger, H.E. (Eds.) *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edition, American Society for Microbiology, Washington, DC, pp. 1615–1626.

Messer, W., Weigel, C. (1996) Initiation of Chromosome Replication, in: Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M., Umbarger, H.E. (Eds.) *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edition, American Society for Microbiology, Washington, DC, pp. 1579–1601.

Newton, A., Ohta, N. (1990) Regulation of the cell division cycle and differentiation in bacteria, *Annu. Rev. Microbiol.* **44**, 689–719.

Zyskind, J.W., Smith D.W. (1992) DNA replication, the bacterial cell cycle, and cell growth, *Cell* **69**, 5–8.

Bacterial Pathogenesis, Molecular Basis of

Melanie Filiatrault and Luciano Passador
University of Rochester Medical Center, Rochester, NY, USA

1 Introduction 555

2 Concepts and Principles of Pathogenesis 555

3 What Makes a Pathogen a Pathogen? 557

4 Regulation of Virulence 559

5 The Pathogenic Cycle 561
5.1 Transmission of Pathogens to the Host and Host Susceptibility 562
5.2 Entry into the Host 562
5.3 Adherence to Host Cells 563
5.4 Evasion of the Host Immune Response 566
5.5 Invasion and Intracellular Survival 567
5.6 Dissemination 569

6 Resistance to Antimicrobials 569

7 Closing Remarks 571

Bibliography 571
Books and Reviews 571
Primary Literature 572

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Keywords

Virulence
The relative ability of an organism to cause disease.

Pathogen
Organisms capable of causing disease.

Adhesin
A molecule used by a bacterial cell to adhere to a surface.

Colonization
The ability of a pathogen to adhere to host cells or surfaces and to multiply within the host.

Quorum sensing
A regulatory mechanism by which the expression of various genes is coordinated to population density via the interaction of a receptor protein and a small organic signaling molecule.

Antimicrobial Resistance
The ability of a bacterial pathogen to become refractory to the actions of a given antimicrobial compound.

Pathogenic Cycle
The common steps that pathogens need to undertake to establish themselves within a host and survive.

The incidence of infectious disease due to bacteria continues to be a leading cause of morbidity and mortality worldwide despite advances in public health measures, modern medicine, and the availability of myriad antimicrobial drugs. Those organisms that are able to cause disease are referred to as *pathogens*. These organisms have evolved many mechanisms to subvert the immune defenses of the human host and usurp usual host cell processes to allow their own proliferation and persistence during infections. In addition to uncovering the inner working of both bacterial and host cells, the elucidation of these processes provides the potential for the development of new therapeutic approaches.

1 Introduction

Despite tremendous advances in medicine and public health measures, infectious diseases caused by microbial agents are still the leading cause of morbidity and mortality worldwide. This is due not only to the emergence of new infectious agents but also to the reemergence of diseases once thought to be under control. Furthermore, increasing numbers of organisms resistant to numerous antimicrobials are being isolated and identified. Organisms capable of causing disease are referred to as pathogens and may be generally divided into three main categories as follows: (1) viruses, (2) bacteria, and (3) fungi and parasites. This chapter will concern itself with bacterial pathogens, and, in the ensuing discussion, we will deal with the strategies used by various bacteria to promote disease. For obvious reasons, most of the discussions will be based on a human serving as the host. However, it should be noted that bacterial pathogens exist for almost every animal and plant life form and that the mechanisms and strategies outlined for survival and causation of disease in the human are also generally applicable to those hosts as well. Bacterial pathogenesis is a large and complex topic and it would be impossible to discuss every detail in a short review such as this. Indeed, there are myriad books that deal solely with this topic, and the reader is directed to these and other publications listed at the end of this chapter for further details. Rather, this chapter is intended to serve as a general overview to provide the basic concepts of this complex field to the reader.

2 Concepts and Principles of Pathogenesis

A concept important to any discussion of pathogenesis is that of virulence, which may be defined as the relative ability of an organism to cause disease. As we shall see, the virulence of a pathogen is intimately tied to a large and complex group of factors (virulence factors, see Table 1) that it can generate and the relative fitness of a given host. The ability of bacteria to be pathogenic is directly associated with the ability of these organisms to produce various cell-associated or extracellular virulence factors. These components are often involved in processes such as adherence to host cells and in the evasion of host defense mechanisms. The presence of these factors allows the bacterium to gain an advantage over the host and results in various pathologies.

To cause disease, some organisms use a combination of tactics, while a few rely on a single approach. Pathogenesis may be caused by factors produced after the bacteria have colonized the host or by factors that are produced exogenous to the host. The latter case, in which the host must contact preformed toxins produced by the bacteria, is primarily limited to certain forms of food poisoning. Interestingly, even though a large number of pathogenic bacteria exist, recent evidence demonstrates that many of these pathogens utilize strikingly similar

Tab. 1 Representative virulence factors/mechanisms and their functions.

Virulence factor/mechanism	*Organism(s)*	*Virulence function*
Motility	*Pseudomonas aeruginosa*	Enhanced ability of motile organisms to disseminate to other tissues.
Altered lipopolysaccharide side chains	*Salmonella typhimurium*	Inhibition of cell lysis by inhibiting the formation of the activated complement complex.
Proteases	*Pseudomonas aeruginosa* *Streptococcus pyogenes*	Destruction of tissue for dissemination; provision of nutrients. Destruction of complement, antibodies, and other host cell defense proteins.
Antigenic variation	*Neisseria gonorrhoeae*	Variation of surface proteins to evade host antibody response.
Exotoxins	A–B types: *Corynebacterium diphtheriae* *Pseudomonas aeruginosa* *Clostridium tetani* *Vibrio cholerae* Membrane disrupting: *Listeria monocytogenes* *Clostridium perfringens*	Killing of host immune response cells (phagocytes, PMNs, etc). Destruction of tissue to enhance nutrient availability and possible dissemination to other sites.
Pili or fimbriae	*Pseudomonas aeruginosa* *Escherichia coli* *Vibrio cholerae*	Adhesion to host cells and mucosal tissue.
Nonpilus adhesins	*Pseudomonas aeruginosa* *Streptococcus pyogenes* *Vibrio cholerae*	Adhesion to cells and mucosal tissue.
Secreted (siderophores) and surface proteins that bind iron	*Escherichia coli* *Pseudomonas aeruginosa* *Yersinia pestis*	Acquisition of iron from host sources.
Capsules	*Streptococcus pneumoniae*	Inhibit phagocytosis. Decrease activation of complement.
Survival within phagocytes	*Salmonella typhimurium*	Evasion of host immune response.

strategies and mechanisms to cause infection and disease. While not applicable in every case, almost all pathogens must successfully complete a series of common steps in order to flourish. First, a pathogen must gain entry into the host. Second, the pathogen must be able to establish itself (colonize) and adhere to host cell tissues. Third, the pathogen must be able to survive within the host. Fourth, the pathogen must be able to cause cell or tissue damage. Lastly, the pathogen must be able to disseminate from the host in order to infect a new host.

The process may appear to be simple, but one must remember that pathogens face considerable challenges in attempting to complete the steps outlined above. Almost immediately these organisms face the task of outcompeting the normal

microbial flora that may be present at a given infection site. There is competition for nutrients and adherence sites, and many microbes considered part of the normal flora are able to defend their territory via the production of various proteases, toxins, and other deleterious factors. If a bacterium can survive this challenge, it must then face the formidable challenge of trying to evade the human immune system.

Clearly, in order to survive, a pathogen must be able to respond and adapt to its changing environment. Pathogens must adapt to changes induced during transmission to a new host or to changes that occur in the new host as a result of the pathogens' presence. The fact that many successful pathogens have been identified is a testament to the adaptability and diversity of these organisms. The concept of adaptability raises the point that we must recognize that pathogenic bacteria do not cause disease with malicious intent as is so often implicitly conveyed in any such discussion. Rather, they cause disease as a consequence of the evolutionary stresses that select for those organisms that can survive in an environment that is most often very different from the environment in which they are usually found. As will become evident, the balance between host and pathogen is quite delicate and a small shift in this balance can give one or the other the upper hand.

3 What Makes a Pathogen a Pathogen?

Bacteria cover and colonize most of the exterior surface of the body including mucosal surfaces. While the majority of these organisms usually serve a beneficial role to the host, or at least do not appear to be detrimental, a small percentage of bacteria are pathogenic. This demonstrates that the association of pathogen and host does not always lead to disease. In fact, some pathogens such as *Streptococcus pneumoniae* are able to colonize the host in large numbers; yet the host does not exhibit signs of streptococcal pneumonia. The number of organisms needed to be present to prove sufficient to cause disease is termed an *infectious dose*. The specific infectious dose can vary between organisms and even for the same organism in different hosts. Some pathogens are considered "overt" in that they can cause disease even in immunocompetent individuals. Other pathogens are considered "opportunistic," causing infection only when some predisposing factor is present in the host. A classic example of an opportunist is *Pseudomonas aeruginosa*, which infects individuals who have been immunocompromised in some way.

The finding that a small percentage of organisms can actually be pathogenic leads us to the question of what makes a given organism a pathogen. The rise of molecular biology and advances in genetics allowed us to gain an appreciation for the complexity of the answer to this question. Clearly, the susceptibility of the host plays a large role in the success of a pathogen, and there is evidence (discussed further in this chapter) that specific genetic traits of the host may predispose an individual to a given infection. Just as clear is the fact that the genetic makeup of the bacterium itself plays an equally important role. Closely related bacteria can demonstrate dramatically different pathogenic potential. This is evident with the identification of specific strains of organisms that are pathogenic, while other strains of the same organism are not.

In the genetic sense, a bacterium can become pathogenic by one of a number of events. The organism could harbor genes unique to itself. A good example of such genes is the "pathogenicity islands" (PAIs) that have been identified in numerous species of bacterial pathogens. PAIs are chromosomal segments of DNA that encode genes required for the expression and production of virulence factors and that are absent from nonpathogenic related bacteria. The PAIs can range in size from 10 to 200 kb and thus can encode few or many genes important for virulence. Many of the identified PAIs are located between loci that encode transfer RNA, sites that are frequently used by specific bacteriophages as integration sites. This finding suggests that the acquisition of a PAI by a given organism can be accomplished simply by the infection of the bacterium with a phage harboring the DNA as part of its genome. This horizontal gene transfer concept is further supported by the finding that PAIs are usually flanked by direct repeat sequences that indicate that the region is prone to recombination events such as those necessary for the integration of a foreign fragment of DNA. Further evidence for the concept that the PAI is acquired from a different organism is that they often exhibit G + C contents that are significantly different from that of the organism in which they are found. Implicit in this finding is that transfer of a PAI between organisms could be accomplished simply by infection with a bacteriophage that is able to infect both organisms.

It should be noted that PAIs are not the only virulence elements that can be acquired by horizontal transfer. For example, many of the genes required for the virulence of *Shigella spp.* are maintained on a large plasmid carried by the pathogenic strain. *Shigella* is also able to produce a number of cytotoxins, the genes for which are carried by a bacteriophage that has integrated into the chromosome. This bacteriophage has also been shown to be transferred to the enterohemorrhagic strains of *Escherichia coli*, which are now able to produce the cytotoxins. Thus, the transfer of genes encoding virulence factors within the same specie or between species has been documented and provides a mechanism for the acquisition of virulence capabilities by organisms that were previously avirulent.

A second mechanism by which a bacterium can evolve to pathogen status is via the loss of a function required for suppression of virulence factor expression. Suppressors, as the name implies, function to maintain the expression of a given gene(s) quiescent. One of the first examples of such a mechanism was the loss of the OmpT surface protease of *Shigella*. This protease suppresses virulence since its expression interferes with the expression of the VirG protein, which is required for intercellular spread. The regions of DNA corresponding to the genes that have been lost are known as *black holes* to contrast them with the PAIs whose acquisition leads to increased virulence.

A bacterial species may become more virulent by the differential regulation of the same group of genes. That is, the expression of the same functional group of genes may be regulated by different environmental controls. An excellent example of such differential regulation occurs again in the genus *Shigella*. Both *S. enterica* and *S. flexneri* harbor very similar genes encoding functions required for cellular invasion. However, the *S. enterica* genes respond to oxygen tension, while the expression of the *S. flexneri* cohort of genes is modulated by temperature.

Finally, a bacterium can increase its virulence simply by harboring mutations in a gene required for pathogenic success. There are numerous examples of such a mechanism in action, but one of the best is the *fimH* gene of *E. coli*, which encodes a protein required for adherence to host tissue. It has been shown that various single mutations in the sequence of this gene can give rise to variants of the adhesin, which increase the binding of the *E. coli* to tissue, thereby raising its virulence potential.

Clearly, the mechanisms described in the preceding text have implications for the evolution of pathogenic bacteria. A given organism may be the subject of one or more of the mechanisms and may thereby evolve into a more potent, efficient, and ever-changing pathogen. In addition, the ability of vectors such as bacteriophages and plasmids to harbor virulence genes and their potential as vectors for horizontal transfer of virulence traits suggests that bacteria are constantly evolving during passage through their hosts.

4 Regulation of Virulence

As mentioned in the previous section, pathogens produce an array of products collectively termed *virulence factors*, which enable them to cause disease. However, pathogens do not express these factors constitutively. Rather, expression of most virulence factors is tightly controlled. Given that a pathogen must be able to sense the condition of its host, it is not surprising that many of the signals that initiate expression of virulence factors are environmental. Diverse bacterial pathogens can recognize an array of signals including pH, temperature, growth phase, osmolarity, oxygen tension, and availability of iron and carbon source as cues for virulence factor expression. By way of example, the Anr protein is responsible for the modulation of gene expression due to oxygen availability in *P. aeruginosa*. There are a number of regulatory factors that are responsive to the presence or absence of iron. The most recognized of these is the Fur protein that binds to specific DNA elements only in the presence of iron, resulting in the repression of the target gene(s). The ability to sense such signals is critical to the organisms' ability to adapt to the microenvironment in which it finds itself.

The sensing of a specific signal results in the triggering of a cascade of events that lead to a regulation of the genes responsible for the production of a given virulence factor(s). The signals are converted to a specific response via the action of molecules deemed regulators. In the simplest sense, the regulators alter the expression of a gene so that it is either increased or decreased in response to the signals that are sensed. Some of these regulatory molecules may respond to more than one signal, and, in this way, the expression of various virulence factors is coordinated and made interdependent to allow the required response of the organism. While the mechanism(s) used to regulate gene expression are numerous, they can be broadly categorized into several common motifs. In addition, a pathogen can use one or more mechanisms to simultaneously respond to the signals that they sense. Furthermore, one or more mechanisms may act to regulate the production of a single virulence factor, allowing the pathogen to monitor and respond to multiple environmental cues to finely tune the expression of the given factor. Although many diverse organisms

use a variety of regulators, many of these regulators can be grouped into families on the basis of their function. Thus, although homologies at the sequence level (both nucleotide and protein) may occur, functionally the products are the same. The number of regulators identified has exploded with the introduction of molecular techniques and their application to studies of bacterial pathogenesis. While all of them cannot be discussed here, it is important to acquaint the reader with some examples and hence three major regulatory mechanisms are briefly described in the following text.

One of the simplest ways to regulate the expression of a specific gene is to alter the recognition of the promoter elements recognized by RNA polymerase (the enzyme responsible for synthesis of RNA from DNA). Sigma factors, synthesized by most bacteria regardless of pathogenic potential, are proteins whose function is to interact with RNA polymerase and thereby alter the expression of a given gene(s). Many organisms express various sigma factors, which are responsible for coordinating gene expression to various specific environmental cues. Not all organisms will produce the same sigma factors, but the role of sigma factors as regulators of gene expression has been well documented in many bacteria. These factors may be produced during specific periods of growth of the organism or in response to specific stimuli. One excellent example is the sigma factor RpoS (also known as σ^{32}). This sigma factor has been shown to be important for the virulence of a number of organisms including *E. coli*, *Salmonella typhimurium*, and *P. aeruginosa*. RpoS is expressed during the stationary phase of growth when the organism is under various stresses such as nutrient deprivation and oxygen tension. Other sigma factors, which have a demonstrated role in the expression of virulence genes, are RpoE, which responds to various stresses and is important for the virulence of *S. typhimurium*; and RpoH, which is involved in the heat shock response of many organisms. The production of the exopolysaccharide alginate, responsible for the mucoid phenotype of *P. aeruginosa* in infections of individuals with cystic fibrosis, is regulated by two sigma factors RpoN and AlgU.

One of the most common types of regulator is that which belongs to the two-component systems. These regulatory systems consist of two proteins – a sensor protein, which is embedded in the membrane of the bacterial cell wall and which acts to sense specific signals; and a response regulator, which is then the mediator of the regulatory response. The sensor molecules usually contain a histidine kinase function that allows the molecule to phosphorylate itself when a specific signal is sensed. The phosphate is then transferred to the response regulator, which alters its conformation and leads to the formation of an "active" form of the molecule. The response regulator may then either interact with the RNA polymerase enzyme or bind to the promoter region to result in stimulation or repression of the gene of interest. Two-component systems are involved in the sensing of and response to signals such as iron availability, carbon and nitrogen levels, and phosphate levels. The expression of many different virulence factors is known to be regulated by two-component systems. Among these are the *Bordetella pertussis* toxin, fimbriae and filamentous hemagglutinin Fha (controlled by the BvgA/BvgS system), the formation of pili and toxin in *Vibrio cholerae* (the ToxR/ToxS system), and

alginate production in *P. aeruginosa* (the AlgZ/AlgR system).

In addition to the two-component systems, a number of other regulatory factor families have been implicated in virulence factor regulation. The families are usually based on the homology or conservation of distinct motifs (e.g. the ability to bind a specific DNA element) within the members of the group. Two of the best-characterized are the AraC and LysR regulator families. The AraC regulators are a group of proteins that exhibit homologies to the AraC protein, which regulates arabinose utilization in *E. coli*. These proteins bind to specific DNA sequences upstream of target genes to regulate their expression. A member of the AraC family, the VirF protein of *Yersinia spp.* is involved in regulating the expression of the Yop proteins, which are involved in the virulence of this organism. It should be noted that the expression of VirF itself is regulated by temperature and that the Yop proteins are only expressed at 37 °C, demonstrating the networking of different regulatory mechanisms to produce expression of the correct product at the correct time during infection. The LysR family includes SpvR, a DNA-binding protein that regulates the expression of a number of genes required for the survival of *Salmonella spp.* in the mouse.

Over the past decade, a strong connection has been made between the coordination of population density and the regulation of virulence genes in a mechanism known as *quorum sensing*. This coordination is achieved by the interaction of small organic signal molecules, which can traverse the bacterial cell wall, with specific receptor proteins that can act as regulators of transcription. In gram-negative organisms, the signals appear to be primarily homoserine lactones, while gram-positive organisms appear to use peptides. It is proposed that as cell population density increases, the concentration of the appropriate signal molecule increases both external and internal to the bacterial cell. Within the pathogen, the number of regulator/signal complexes is increased until, upon reaching a threshold concentration, gene expression is coordinated. The discovery of this mechanism in pathogens is relevant to the observation that many pathogens must reach a critical density within the host prior to causation of disease. It is postulated that quorum sensing is used by pathogens to allow production of virulence factors at the appropriate time during infection and that the pathogen can unleash an arsenal of factors to overwhelm the host. As an extension of this postulate, it may be proposed that a pathogen does not express unwanted factors too early, a situation that, in addition to being wasteful, may also give premature warning of the pathogen's presence to the immune system. The field of quorum sensing is intensively studied, and the list of pathogens that use such a mechanism as well as the virulence factors regulated by this mechanism is rapidly expanding.

5 The Pathogenic Cycle

In the previous section, there was a discussion of the steps that most pathogens undertake in their progression to causing disease. In the sections that follow, these steps will be described to allow the reader to gain some appreciation of the complexity of the process as well as to gain some insight into the hurdles that pathogens face within the human host.

5.1 Transmission of Pathogens to the Host and Host Susceptibility

The information regarding the effective transmission of a pathogen to its host is somewhat limited. Some pathogens such as *B. pertussis*, the causative agent of whooping cough, can be transmitted via aerosols created by sneezing. Others, including *Treponema pallidum* and *Neisseria gonorrhoeae*, which cause syphilis and gonorrhea respectively, are transmitted via sexual contact. Some organisms rely on transmission via animal or arthropod vectors. *Borrelia burgdorferi*, the cause of Lyme disease is transmitted by the deer tick, which can bite humans and in effect inoculate them with the organism. Still others, such as *V. cholerae*, are internalized by ingestion of contaminated matter or via the fecal–oral route.

Within the host, any available surface is a potential target for bacterial attack. However, not all pathogens can interact with all surfaces or, for that matter, infect all hosts. Some pathogens may be able to interact with a wide variety of cells and tissues, whereas others are restricted to specific cell types. This demonstrates that some microorganisms exhibit preferences for specific tissues and hosts. For example, *N. gonorrhoeae* can only infect humans and more specifically only certain mucosal surfaces. In contrast, *P. aeruginosa* can cause a wide variety of infections affecting various tissues in a number of hosts.

There is increasing evidence that some factors may predispose a given host to infection with a given pathogen and that the severity of the disease may actually be modulated by these factors. A prime example is the NRAMP1 protein, which was initially identified in mice but which has also recently been identified in humans. This protein is located on the surface of macrophages, cells that are part of the immune system and that are used to engulf bacteria and destroy them, where it functions as a transporter of specific metals such as iron and manganese into the mammalian cell. Mutation of this protein leads to increased susceptibility to infection with organisms such as *Mycobacterium spp*. A single amino acid mutation (glycine at position 169 changed to aspartate) is highly associated with this increased susceptibility. It is proposed that the loss of proper function of this protein results in increased levels of such metals in the extracellular spaces. Since these metals are essential for the basic growth requirement of all cells including bacteria, it is proposed that the increased availability of the metals allows the pathogens to thrive in an environment in which these metals would normally be limited in supply. Furthermore, since the metals are used by bacteria as integral components of enzymes required to defend against host defenses, the availability of the metals in NRAMP1 mutant hosts would allow the organisms to resist those defenses and survive.

5.2 Entry into the Host

The primary barrier to infection is the epithelium, which is considered part of the human hosts' innate immunity. Bacteria may breach this protective layer in many ways. One approach is to enter through breaks in the skin caused by trauma, including cuts, burns, and insect bites. *Pseudomonas aeruginosa* is an opportunistic pathogen that can cause horrific infections of individuals who suffer burn trauma. Arthropod vectors can transmit *Yersinia pestis*, the causative agent of the

bubonic plague, via direct inoculation in the form of a bite. The increased use of indwelling medical devices such as catheters also provides portals of entry into the host and may actually serve as recurrent sources of infection. Additional routes of entry include mucosal surfaces, which are involved in nutrient uptake or gas exchange. Examples of such surfaces are the gastrointestinal and respiratory tracts. These mucosal surfaces are in constant contact with the external environment and hence are under risk of being breached by bacteria. These tissues serve as a thin barrier between the host and the external environment and provide a portal of entry for many bacteria that are capable of invading a host through intact epithelia. The breach of the skin barrier is important because it allows not only the normal flora on the surface to enter but also invites invasion by pathogens. Indeed, organisms that are relatively innocuous on the surface of the skin may actually present problems if allowed to penetrate the skin barrier and gain access to underlying tissue or blood vessels.

Some organisms do not directly enter the host nor require a living vector system to cause disease. Although they can also cause wound infections, members of the genus *Clostridium* can elaborate toxins in foodstuffs that can result in severe disease upon ingestion. One of the most common examples of disease without colonization is botulism. *Clostridium botulinum* is able to produce botulinum toxin that essentially acts as a neurotoxin. Upon ingestion of tainted food, the toxin is absorbed from the intestines into the circulation and binds to susceptible neurons. Part of the toxin (i.e. the light chain) is internalized in the cytoplasm of the neurons where it begins to exert its proteolytic activity and disrupt cellular pathways required for the release of chemical messenger molecules. As a result, the muscles do not receive the proper signals and essentially enter a state of flaccid paralysis. As the paralysis progresses further, there is risk of fatal respiratory and cardiac failure.

5.3 Adherence to Host Cells

Once inside a host, bacterial pathogens can adhere either directly to a host cell surface or to a component of the extracellular matrix, which functions as a foundation for the host cells. Some pathogens remain on the surface of a cell and multiply at that location, while others make use of their ability to invade the tissue in search of a more suitable location. Nonetheless, even invasive organisms must initially attach to a host cell. This is a key step in the interaction between the pathogen and the host. Implicit in this interaction is that the organism must possess a molecule that can allow it to adhere to a host and that the host must possess a receptor for the bacterial adhesin. The surfaces of attachment can be the skin, the mucosal membranes, and some of the deeper tissues. The host has evolved mechanisms such as coughing, sneezing, secretion of saliva, and mucus flow in an attempt to wash away the bacteria. However, some pathogens are able to strongly attach themselves and prevent their eradication from the host. These organisms utilize specific components, termed *adhesins*, to adhere themselves to receptor sites on the host cell surfaces.

The types of adhesins used by bacterial pathogens are varied. They may be composed of either proteins or carbohydrates. The protein adhesins are generally separated into two groups known as fimbrial (pili) and afimbrial. Those adhesins

that consist of carbohydrate are generally components of the bacterial cell membrane, wall, or an exopolysaccharide coating known as the capsule. Many organisms are believed to use a two-step process in which an initial loose binding is mediated by adhesins such as pili followed by a tighter binding owing to the action of other adhesins. This demonstrates that it is common for a single pathogen to synthesize and utilize more than one adhesin.

The function of an adhesin implies the presence of a receptor for that adhesin. The variety of molecules that can be used as receptors include proteins of the host cell membrane, glycolipids, glycoproteins, and proteins that make up the extracellular matrix such as fibronectin and collagen. In some cases, the pathogen can even inject its own receptor into the host, have it located to the surface, and then use it to bind to the host cell. This appears to be the mechanism used by enteropathogenic *E. coli*.

The best-understood adherence mechanism is the use of protein structures known as fimbriae or pili. These rodlike structures are composed of protein subunits termed *pilin*, which are arrayed and extend outward from the surface of the pathogen. In some organisms, the pili are distributed over the entire surface, while in others the pili are localized to a preferential location on the surface. In some cases, the tip of the pilus mediates the actual adherence. This is accomplished by a very specific binding of the pilus to a molecule on the host cell surface. The host cell receptors are most often carbohydrate residues that are part of glycoprotein or glycolipid molecules. The specificity of the binding to these host receptors is significant in determining where, within the body, the organism will eventually bind, since some receptors may be tissue-specific. In some cases, the specificity of binding is determined by a special tip structure that consists of several proteins different from pilin. In others, it appears that pilin itself is involved.

The synthesis and assembly of pili is a complex process that requires proteins other that those involved in the actual pilus structure itself. Usually, the tip of the pilus is assembled and secreted into the periplasm, where specialized proteins known as chaperones hold the pilus proteins in the proper conformation for the addition of further pilin subunits and the eventual transport of the growing pilus to the outer membrane. Pilin subunits are added to the base of the pilus molecule until the proper length is reached. The signals involved in determining the proper length of the pilus are not understood. It is known that a periplasmic protein is involved in signaling for the end of pilus synthesis, but the mechanism by which it conveys this information is still a mystery. There are numerous families of pili that use diverse mechanisms of biogenesis and that have distinct structural characteristics.

Attachment is especially important in areas such as the mouth, the intestinal tract, and the bladder, which are washed by fluids. In these areas, only organisms capable of adhering will be capable of persisting. One of the most obvious of these sites is the urinary tract, which is subject to washing by urine. A number of the adhesins of uropathogenic *E. coli* strains have been identified and studied. Several pilus types have been identified. The type 1 pili appear to be involved in the colonization of the vaginal tract; however, the contribution of these pili in bladder infections appears to be minimal. The P pili, synthesized by the products of the *pap* (*p*yelonephritis-*a*ssociated *p*ili) genes, are extremely important in strains that

cause kidney infections (pyelonephritis). Although P pili of many antigenic types exist, they all appear to have the same type of receptor (globobiose).

As mentioned above, surface molecules that are important in adherence but which do not form a piluslike structure are often called the *nonpilus adhesins* or *afimbrial adhesins*. In addition to the fimbrial adhesins, the uropathogenic *E. coli* strains also carry adhesins that are not pili. These adhesins include the afimbrial adhesins Dr, AFAI, and AFAIII. It is proposed that these molecules are responsible for the tighter binding of the bacterium to the host cell surface subsequent to the initial binding by the pili. Very little is known about the structure and functions of these types of molecules, but it appears that some of them may recognize host cell surface proteins rather than carbohydrates. Many pathogens maintain several alternative means of adhering to cells that may be used under specific environmental conditions or to allow attachment to specific cell types. Members of the genus *Yersinia*, including *Y. enterocolitica*, *Y. pestis*, and *Y. pseudotuberculosis*, produce a number of adhesins that have also been implicated in the invasion of host cells. These adhesins (Inv, Ail, PsaA, and YadA) are extremely interesting in that their expression is often temperature-dependent and, in most cases, they do not appear to use the same receptors. Both the Inv and YadA proteins are expressed at 37 °C and both are capable of using β1-integrins as their receptors. However, YadA can also bind to laminin and fibronectin. In addition, the gene for Inv is chromosomally located, whereas the gene for YadA resides on a large plasmid carried by the organism. In contrast, Ail can be expressed at both 30 °C and 37 °C (albeit expression is higher at 37 °C) and although its receptor has not been identified, it does not appear to be the β1-integrins. PsaA is interesting because it is expressed in both *Y. pseudotuberculosis* and *Y. pestis* but not in *Y. enterocolitica*. Moreover, in contrast with all the other adhesins, it appears to be a pilus. It is expressed only at 37 °C and only when the pH is below 7.

Bordetella pertussis, the causative agent of whooping cough, provides an interesting case for studies of the specificity of adherence. *Bordetella pertussis* exhibits an extraordinary preference for ciliated cells of the respiratory tract. These organisms produce a number of adhesins of which the two best studied are the filamentous hemagglutinin (Fha) and pertussis toxin (Ptx). Both adhesins lack the characteristic ordered structure of a pilus. The receptors for Fha are the galactose residues of sulfated glycolipids. Part of the preference for the ciliated respiratory cells may be due to the high content of a type of sulfated glycolipid known as sulfatide in the membranes of these cells. Mutants in Fha are still able to colonize the host, indicating that other adhesins are present.

The Ptx molecule is capable of functioning both as a toxin and an adhesin. Ptx is composed of six subunits termed S1 to S5, all of which are present in single copies except for S4, which is present in two copies. During an infection, some of the toxin molecules are secreted into the surrounding environment, but the rest remain attached to the bacterial cell surface. These attached Ptx molecules can now function as adhesins. The S2 and S3 subunits are responsible for binding to receptors on the host cell. It is proposed that these molecules act as adhesins when they are left on the bacterial cell surface.

Even though both these molecules have a high degree of homology at the amino acid level, they appear to bind two different receptors. S2 specifically binds to a glycolipid called *lactosylceramide*, which is found primarily on ciliated respiratory cells. S3 binds to a ganglioside found primarily on phagocytic cells. It should be noted that Fha can also bind CR3 on polymorphonuclear monocytes. *Bordetella pertussis* also produces pili and a surface protein known as pertactin.

Some organisms use polysaccharide-based adhesins to attach to host tissue. These adhesins are exemplified by the teichoic acids (components of the gram-positive bacterial cell wall), lipopolysaccharides or lipooligosaccharides (found only in gram-negative organisms), or capsules (found on both gram-positive and negative organisms). As examples, both *Staphylococcus* and *Streptococcus spp.* utilize their teichoic acids as adhesins; *Haemophilius influenzae* makes use of lipooligosaccharides to attach to respiratory epithelial cells; and adhesion of *Mycobacterium spp.* is promoted by the glucan and mannan polysaccharides found in their capsule.

A little-discussed facet of adherence is the role that the host cell itself may play in aiding the attachment of the pathogen. It is known that interactions with pathogens activate various host cell–signaling systems. Recent data suggest that pathogens may be able to connect to these signaling mechanisms to encourage the host cell to synthesize receptors to facilitate bacterial adherence. Recent data suggest that various enteropathogenic *E. coli* may secrete proteins that stimulate the expression of specific receptors on host cell surfaces. The *E. coli* then utilizes these receptors for adherence. *Streptococcus pneumoniae* has also been shown to stimulate host cells to express a receptor for its adherence.

5.4 Evasion of the Host Immune Response

After attachment, the bacteria must survive in their new environment. Toward this end, the bacteria must compete for nutrients not only with the host but also with the indigenous microbial flora. In addition, the organism must be able to evade the components of the host immune response. Antibodies and complement are the major components of the humoral response and when these proteins encounter a pathogenic organism or its soluble components, the antibodies bind to form an antibody–antigen complex. In the case of whole organisms, a series of complement proteins bind the complex and result in the lysis and death of the pathogen. Some pathogens (e.g. *S. typhimurium*) exhibit serum resistance in that they are able to inhibit the formation of the complement complex and thereby avoid cell lysis. In the case of soluble factors, the formation of the complex makes it easy to phagocytize and remove pathogens by specific cells. However, the formation of the immune complexes themselves may cause pathology as seen in complex-mediated hypersensitivity in which the host immune system is turned against the host itself. An excellent example is the case of post-*Streptococcal* glomerulonephritis. The classic view of this disease postulates that lysis products of *Streptococcus pyogenes* that result from the host inflammatory response attempting to clear a throat infection eventually enter the bloodstream and initiate an antibody response. The high levels of antibodies produced by the host then interact with antigens present in the bloodstream, resulting in the formation of antibody–antigen complexes, which may then accumulate in the kidney. The complexes themselves may

elicit an inflammatory response resulting in damage to kidney tissue and eventually impairment of normal kidney function. The same organism is also responsible for rheumatic heart disease in which antibodies directed against a number of *S. pyogenes* products include those against M protein, a cell surface protein of *S. pyogenes* that shares epitopes with proteins on the surface of cardiac tissue. Hence the formation of antibodies against M protein results in the production of a set of antibodies that may also recognize the host tissue, leading to an immune response against the host.

Alternatively, the interaction of pathogens with antibodies may make them more susceptible to phagocytosis. The process of coating cells with antibodies termed *opsonization* leads to their recognition by phagocytic cells and other cells of the immune system. This interaction may also lead to hypersensitivity reactions since many of these cells will release molecules that result in the development of an inflammatory response. Some pathogens directly attack the immune response (e.g. *P. aeruginosa*) by producing proteolytic enzymes that can degrade certain classes of antibodies and complement components. Others evade the immune response by cloaking themselves in exopolysaccharide capsules (e.g. *Klebsiella pneumoniae*, *H. influenzae*), which are believed to impair the ability of antibodies to bind and to prevent phagocytosis, or by periodically changing their surface antigens to render existing antibodies ineffective (antigenic variation). *Neisseria gonorrhoeae*, the causative organism of gonorrhea, provides a good example for antigenic variation. This organism is capable of using a variety of methods in order to vary the amount and composition of the pili on its surface. In this way, it presents an ever-changing surface to the host response, resulting in a decreased ability of the host to clear the organism. Still other organisms utilize a process of molecular mimicry in which they present a surface marker that the host recognizes as "self" and thus does not respond with a normal immune response. Good examples of this mimicry are the hyaluronic acid capsule of *S. pyogenes* or the sialic acid capsule of *N. meningitidis*. The importance of evading the immune response is illustrated by opportunistic pathogens that are limited to immunocompromised hosts but cannot survive in a host whose immune system is intact.

5.5 Invasion and Intracellular Survival

In order to persist and perpetuate the infection, some pathogens need to gain access to deeper regions of the host through a process generally known as invasion. The ability of pathogens to invade normally phagocytic cells such as the macrophages, which function to engulf pathogens, is not very surprising. However, pathogens have had to evolve more elaborate mechanisms in order to invade nonphagocytic cells. The fact that some invasive pathogens can invade an extremely diverse population of cells while others are restricted to specific cell types implies that host cell receptors are important in invasion. Integrins that play a role in host cell interactions including attachment, phagocytosis, and the binding of proteins present in the extracellular matrix have been identified as receptors. *Yersinia pseudotuberculosis*, *Mycobacterium tuberculosis*, and *Legionella pneumophila* have all been shown to utilize members of the integrin family of proteins as receptors.

In actuality, the concept of invasion can be subcategorized into two types, extracellular and intracellular. In extracellular invasion, a pathogen breaches the barriers of various tissues in order to disseminate within the host but in a manner in which it does not actually enter the host cells. Some organisms (e.g. *P. aeruginosa* and *Staphylococcus aureus*) produce and secrete enzymes that degrade host cell molecules such as elastin, an extracellular protein, and hyaluronic acid, the cement that holds cells together. Destruction of these molecules allows the organism to move from one location to another more suitable location where they can survive and proliferate.

Many pathogens, both gram-positive and gram-negative, have the ability to invade intracellularly. In fact, some pathogens such as *Chlamydia spp.* and *Rickettsia spp.* are obligate intracellular pathogens and absolutely require a host cell for growth. Other pathogens are classified as facultatively intracellular and invade host cells to proliferate or to spread to other tissues. The advantage of intracellular invasion is several-fold. First, it sequesters the organism from the potentially lethal effects of the host immune system. Second, it places the organism in an environment that is nutrient-rich. Lastly, it positions the organism in a location where competition from other organisms is almost nonexistent. However, invasion requires a set of complex biological mechanisms that must allow the organism a means to not only enter the cells but also survive, multiply, and eventually disseminate from the target tissue. Intracellular invasion can also be subdivided into two types. There are those pathogens that take advantage of the ability of highly differentiated professional phagocytes, which normally ingest bacteria. Other pathogens are able to invade cells that do not have the ability to ingest bacteria. In this type of invasion, the process is considered directed. That is, the bacterium directs the production of various factors known as invasins or internalins that allow the organism to invade. One example of such a pathogen is *Y. pseudotuberculosis*. This organism utilizes invasins including the Inv and YadA products to associate with the β1-integrin on the host cell surface. The association with multiple receptors leads to the bacterium becoming literally enveloped by the host cell and thus internalized. The process has sometimes been referred to as a *zippering* on the basis of the obvious structural analogies as the host cell surrounds the bacterium. In contrast, *Shigella spp.*, which are also intestinal pathogens, utilize a different approach involving the temperature-sensitive induction of Mxi–Spa and Ipa products. Immediately following ingestion, the cells begin to synthesize the Mxi, Spa, and Ipa products owing to the temperature of the host. The Mxi–Spa products assemble in the membranes of the organism and form a translocation apparatus that is used to move the Ipa proteins out of the cell once contact with the host cell has been made. The Ipa proteins form complexes that interact with the host cell surface and alter the signaling pathways of the host cell. The altered signaling results in rearrangements of the host cell cytoskeleton and leads to the formation of membrane projections giving the host cell a "ruffled" appearance. Eventually, the projections of the membrane encircle the *Shigella* and the organism is taken into the cytoplasm of the host cell. It should be noted that some pathogens make use of both extracellular and intracellular mechanisms of invasion in order to proliferate.

Once inside a cell, there are a number of possible outcomes for the pathogen. The

majority of the cells that enter a phagocytic host cell are destroyed. However, some pathogens have evolved mechanisms that allow them to survive within this environment. For those pathogens that do not invade professional phagocytes, the task of survival is somewhat facilitated in that their host cells do not maintain an arsenal of host defense mechanisms as do the phagocytes. Survival within host cells is accomplished by a number of strategies. Upon ingestion of the pathogen, the host cell stimulates fusion of the vacuole harboring the pathogen (endosome) with vacuoles that contain degradative enzymes (lysosomes). Some organisms such as *Mycobacterium spp.* are able to inhibit the fusion of the endosome to the lysosome, thereby eluding the toxic enzymes. Other pathogens (e.g. *Rickettsia* and *Listeria monocytogenes*) are able to dissolve the endosomal membrane and escape directly into the cytoplasm of the cell. Still others (e.g. *L. pneumophila*) utilize outer-membrane proteins designated Mip that interact with specific receptors on the host cell surface to stimulate engulfment via pathways that expose the pathogen to fewer killing mechanisms. As an added benefit, *L. pneumophila* is also able to inhibit the fusion of the endosome to the lysosome.

5.6 Dissemination

Many pathogens remain localized to the initial site of attachment, while others are capable of dissemination within the host. The ability to disseminate allows the pathogen to move to environments of greater nutrient availability as well as to increase the probability of locating a preferred niche (see preceding text). Dissemination may involve movement through cells or between cells in the extracellular matrix. Many pathogens are capable of eroding tissue through the secretion of extracellular products. *Pseudomonas aeruginosa* and *Staph. aureus* may secrete a large number of extracellular products including toxins, which can have an exfoliative effect, and collagenases and elastases, which degrade major connective tissue components. *Listeria monocytogenes* can take intracellular invasion one step further to aid in dissemination. This organism makes use of various virulence factors, including a membrane-disrupting toxin known as Listeriolysin O, and phospholipases to escape from the endosome into the cytoplasm. Once in the cytoplasm, it begins to disrupt normal cellular processes, resulting in the deposition of actin on the bacterial cell surface. These "actin tails" are further extended and the organism uses these tails to move about. The reorganization of the host cell cytoskeleton also permits the formation of "philopodia" or extensions of the infected host cell. By moving through these philopodia, the pathogen can gain access to neighboring cells and invade them. In this manner, *Listeria monocytogenes* is not required to exit the host cells and remains protected from the host immune defenses.

6 Resistance to Antimicrobials

The increasing resistance of many bacterial pathogens to a wider range of antimicrobial agents is an intriguing yet alarming observation. Clearly, the increased use of antibiotics by the medical profession has contributed to the problem by placing pressure on bacterial populations to select individual cells that can survive in the presence of these agents. As a result,

the concept of many agents as "wonder drugs" is quickly fading and is being replaced with the reality that we may, if we have not done so already, be selecting for "super organisms" resistant to all available drugs. By way of example, strains of methicillin-resistant *Staph. aureus* have been identified, and for those unfortunate individuals who are infected by such strains, there are no viable antimicrobial options available. Aside from the purely scientific standpoint, an understanding of how pathogens become resistant to antimicrobials is quickly becoming an area of intense research owing to the global threat that such super organisms would pose.

In general, there are several mechanisms that would lead to the acquisition of resistance to an antimicrobial.

Alteration of the target. An organism could evolve to select for those individuals who have altered the target of the antimicrobial so as to render the target insensitive to the drug. This is seen in a number of cases including the transpeptidases, which are enzymes involved in bacterial cell wall synthesis and are normally inhibited by penicillins by virtue of the fact that they are capable of binding the drug. Single amino acid changes within the protein structure of these molecules can render them insensitive to the drug altogether or merely reduce their binding affinity for the drug so as to decrease their sensitivity. Other antimicrobials target the ribosomes, molecules essential for protein synthesis by the cell. Single mutations within subunits of this molecule can also render it insensitive to various drugs such as spectinomycin.

Overproduction of the target. While not a common mechanism, bacterial pathogens could overproduce a target and effectively increase the amount of drug required to be effective.

Modification of the antimicrobial. This is a common mechanism and is easily exemplified by the bacterial β-lactamases. These enzymes hydrolyze the lactam ring component of various β-lactam antibiotics (e.g. penicillin) to inactivate the drug. A variation on this mechanism is the addition of an acetyl moiety onto the structure of chloramphenicol, thereby rendering it inactive.

Interfering with the accessibility of a target to an antimicrobial. A number of potential mechanisms would fall into this category. Given that the membrane is the primary barrier between the bacterium and its external environment, alteration of the permeability of the membrane would affect the ability of certain lipophilic antimicrobials to enter the cell. This alteration in permeability could be achieved by altering the composition of the lipid bilayer. Another mechanism makes use of specific transporter systems that many bacteria maintain to specifically export antibiotics out of the bacterial cell. A wide variety of these so-called drug efflux pumps have been described for many organisms (e.g. *P. aeruginosa*). It should be noted that other more general efflux systems have been described for some organisms. The *mtr* locus of *N. gonorrhoeae* is such an example.

Production of new enzymes to bypass the targeted site. Some bacteria are able to synthesize new enzymes that can carry out a specific function but that have a significantly decreased affinity for a specific antibiotic.

An interesting mechanism for drug resistance may also be available to those pathogens that are capable of growing

in multicellular, sessile, and encapsulated communities known as biofilms. Evidence indicates that bacteria in the biofilm mode of growth exhibit increased resistance to antimicrobial agents. It is not clear whether the resistance is due to a combination of the various mechanisms described earlier, the fact that the biofilm in enclosed in a matrix of extracellular material, or specific physiological changes occurring within the organisms that elicit as-yet undiscovered resistance mechanisms.

As we have seen in previous sections of this chapter, the facility with which bacteria can exchange genetic information allows for the acquisition of many traits by sometimes-unrelated organisms. Such genetic-transfer mechanisms are also important for the bacteria to acquire antimicrobial resistance.

7 Closing Remarks

As has been briefly outlined above, bacterial pathogenesis involves a large, complex interplay of regulatory networks and virulence factors to allow pathogens to inhabit their hosts and cause disease. This interplay between host and bacteria has resulted in many pathogens becoming specific for certain host species as they evolve to proficiently colonize and invade the host. The transfer of various virulence genes between pathogenic species increases the fitness of pathogen populations and allows them to survive in an increasing number of hosts and specialized niches. Pathogens adapt to changes in their environment by being able to sense changes in their hosts and respond by producing virulence factors that aid in their survival. As one might imagine, these adaptations require an intricate balance of the expression of certain genes. For some of the virulence factors, details regarding the mechanisms and genetics have been elucidated. In many cases, the genetic regulation is poorly understood. While we have made great progress in characterizing the molecular basis of pathogenesis, we cannot define the full complement of factors required for pathogenicity. The advent of new technologies such as DNA microarrays combined with the sequencing of ever-increasing numbers of bacterial genomes is allowing for the comparison of gene complements and genetic expression between organisms. Many common products and motifs have been discovered between pathogenic organisms and these observations are increasing the pace with which the molecular mechanisms of pathogenesis can be and are being deciphered. Nonetheless, studies of the array of products and strategies that bacterial pathogens use to survive within their hosts are always revealing novel and surprising findings that illustrate the complexity that is bacterial pathogenesis.

See also Bacterial Growth and Division; *E. Coli* Genome; Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Daffe, M., Etienne, G. (1999) The capsule of *Mycobacterium tuberculosis* and its implications for pathogenicity, *Tuberc. Lung Dis.* **79**, 153–169.

Finlay, B.B., Falkow, S. (1997) Common themes in microbial pathogenicity revisited, *Microbiol. Mol. Biol. Rev.* **61**, 136–169.

Groisman, E.A. (Ed.) (2001) *Principles of Bacterial Pathogenesis*, Academic Press, San Diego.

Groisman, E.A., Ochman, H. (1996) Pathogenicity islands: bacterial evolution in quantum leaps, *Cell* **87**, 791–794.

Miller, M.B., Bassler, B.L. (2001) Quorum sensing in bacteria, *Annu. Rev. Microbiol.* **55**, 165–199.

McClane, B.A., Mietzner, T.A., Dowling, J.N., Phillips, B.A. (Eds.) (1999) *Microbial Pathogenesis*, Fence Creek Publishing, Madison.

Poole, K., Srikumar, R. (2001) Multidrug efflux in *Pseudomonas aeruginosa:* components, mechanisms and clinical significance, *Curr. Top. Med. Chem.* **1**, 59–71.

Whitehead, N.A., Barnard, A.M., Slater, H., Simpson, N.J., Salmond, G.P. (2001) Quorum-sensing in gram-negative bacteria, *FEMS Microbiol. Rev.* **25**, 365–404.

Primary Literature

Blum, G., Ott, M., Lischewksi, A., Ritter, A., Imrich, H., Tschape, H., Hacker, J. (1994) Excision of large DNA regions from termed pathogenicity islands from tRNA-specific loci I the chromosome of an *Escherichia coli* wild-type pathogen, *Infect. Immun.* **62**, 606–614.

Caldwell, A.L., Gulig, P.A. (1991) The *Salmonella typhimurium* virulence plasmid encodes a positive regulator of a plasmid-encoded virulence gene, *J. Bacteriol.* **173**, 7176–7185.

Cellier, M., Govoni, G., Vidal, S., Kwan, T., Groulx, N., Liu, J., Sanchez, F., Skamene, E., Schurr, E., Gros, P. (1994) Human natural resistance-associated macrophage protein: cDNA cloning, chromosomal mapping, genomic organization, and tissue-specific expression, *J. Exp. Med.* **180**, 1741–1752.

Deretic, V., Dikshit, R., Konyecsni, W.M., Chakrabarty, A.M., Misra, T.K. (1989) The *algR* gene, which regulates mucoidy in *Pseudomonas aeruginosa*, belongs to a class of environmentally responsive genes, *J. Bacteriol.* **171**, 1278–1283.

DiRita, V.J. (1992) Co-ordinate expression of virulence genes by ToxR in *Vibrio cholerae*, *Mol. Microbiol.* **6**, 451–458.

Fang, F.C., Libby, S.J., Buchmeier, N.A., Loewen, P.C., Switala, J., Harwood, J., Guiney, D.G. (1992) The alternative sigma factor *katF* (*rpoS*) regulates *Salmonella* virulence, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 11 978–11 982.

Hoe, N.P., Minion, F.C., Goguen, J.D. (1992) Temperature sensing in *Yersinia pestis*: regulation of *yopE* transcription by *lcrF*, *J. Bacteriol.* **174**, 4275–4286.

Hromockyj, A.E., Maurelli, A.T. (1989) Identification of *Shigella* invasion genes by isolation of temperature-regulated *inv::lacZ* operon fusions, *Infect. Immun.* **57**, 2963–2970.

Jackson, M.P., Neill, R.J., O'Brien, A.D., Holmes, R.K., Newland, J.W. (1987) Nucleotide sequence analysis and comparison of the structural genes for Shiga-like toxin-I and Shiga-like toxin-II encoded by bacteriophages from *Escherichia coli*, *FEMS Microbiol. Lett.* **44**, 109–114.

Kenny, B., DeVinney, R., Stein, M., Reinscheid, D.J., Frey, E.A., Finlay, B.B. (1997) Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells, *Cell* **91**, 511–520.

Lee, C.A., Jones, B.D., Falkow, S. (1992) Identification of a Salmonella typhimurium invasin by selection for hyperinvasive mutants, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1847–1851.

Maurelli, A.T., Fernandez, R.E., Bloch, C.A., Rode, C.K., Fasano, A. (1998) "Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella spp.* and enteroinvasive *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3943–3948.

Nakata, N., Tobe, T., Fukuda, I., Suzuki, T., Komatsu, K., Yoshikawa, M., Sasakawa, C. (1993) The absence of a surface protease, OmpT, determines the intercellular spreading ability of *Shigella*: the relationship between the *ompT* and *kcpA* loci, *Mol. Microbiol.* **9**, 459–468.

Pan, W., Spratt, B.G. (1994) Regulation of the permeability of the gonococcal cell envelope by the *mtr* system, *Mol. Microbiol.* **11**, 769–775.

Rajakamur, K. (1997) Use of a novel approach, termed island probing, identifies the *Shigella flexneri she* pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins, *Infect. Immun.* **65**, 4606–4614.

Schaffer, S., Hantke, K., Braun, V. (1985) Nucleotide sequence of the iron regulatory gene *fur*, *Mol. Gen. Genet.* **200**, 110–113.

Skurnik, M., Toivanen, P. (1992) LcrF is the temperature-regulated activator of the *yadA* gene of *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*, *J. Bacteriol.* **174**, 2047–2051.

Sokurenko, E.V., Chesnokova, V., Dykhuizen, D.E., Ofek, I., Wu, X.R., Krogfelt, K.A.,

Struve, C., Schembri, M.A., Hasty, D.L. (1998) Pathogenic adaptation of *Escherichia coli* by natural variation of FimH adhesin, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8922–8926.

Suh, S.J., Silo-Suh, L., Woods, D.E., Hassett, D.J., West, S.E., Ohman, D.E. (1999) Effect of *rpoS* mutation on the stress response and expression of virulence factors in *Pseudomonas aeruginosa*, *J. Bacteriol.* **181**, 3890–3897.

Uhl, M.A., Miller, J.F. (1996) Integration of multiple domains in a two-component sensor protein: the *Bordetella pertussis* BvgAS phosphorelay, *EMBO J.* **15**, 1028–1036.

Zimmermann, A., Reimmann, C., Galimand, M., Haas, D. (1991) Anaerobic growth and cyanide synthesis of *Pseudomonas aeruginosa* depend on *anr*, a regulatory gene homologous with *fnr* of *Escherichia coli*, *Mol. Microbiol.* **5**, 1483–1490.

Bacteriorhodopsin, Molecular Biology of

Janos K. Lanyi
University of California, Irvine, CA, USA

1 Introduction 577

2 Overall Structure of the Protein 577

3 Chromophore and the Photocycle 578
3.1 The K State 579
3.2 The L State 579
3.3 The M States and Proton Release to the Extracellular Side 579
3.4 The N State and Proton Uptake from the Cytoplasmic Side 581
3.5 The O State 581

4 Properties of Single-site Mutants 582

5 Pump Modes, Energetics 582

6 A Family of Retinal Proteins 583

Bibliography 583
Books and Reviews 583
Primary Literature 583

Keywords

Photoisomerization
Rotation of the $C_{13}-C_{14}$ double bond of the retinal from trans to cis, the only directly light-driven reaction of bacteriorhodopsin.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Photocycle
Reaction sequence initiated by photoisomerization of the retinal of bacteriorhodopsin.

Photointermediates
States of bacteriorhodopsin that arise transiently in the photocycle, identifiable through their distinct spectral properties and from the structure of the protein and the retinal.

Retinal Schiff Base
Covalent bond between the C_{15} atom of the retinal and the side-chain amino nitrogen (NZ) of Lys216, with the nitrogen normally in the protonated state.

Schiff Base Counterion
Complex of ionic side chains and bound water with a net negative charge that compensates the positive charge of the protonated Schiff base.

Transmembrane Helix
An α-helical polypeptide segment that traverses the lipid bilayer, a typical structural motif in membrane proteins.

Ion Pump
Membrane protein that utilizes a chemical or photochemical reaction to drive the energetically uphill translocation of an ion across the membrane.

Chemiosmotic Coupling
General biological principle in which pumps generate a transmembrane proton gradient, and the protons moving down the gradient through specific membrane proteins provide energy for a variety of metabolically useful reactions.

Bacteriorhodopsin is a light-driven pump that generates an electrochemical gradient for protons across the membrane. It is unique among such transporters, widespread in biology, in being a small, seven-transmembrane helical protein. Its remarkably simple transport mechanism is driven by the thermal reisomerization of the photoisomerized 13-*cis* retinal to all-trans. The ease of performing static and time-resolved spectroscopy of many kinds, as well as the propensity of the protein for crystallization, have generated a large amount of valuable information about the transformations of the retinal and the protein during the transport cycle. The data describe how the motions of the retinal cause structural changes that result in proton transfers between donor and acceptor groups inside the protein and in the release and uptake of a proton at the opposite membrane surfaces. As a result, today, bacteriorhodopsin is the best-understood ion pump.

1 Introduction

Bacteriorhodopsin is a small (26 kDa) integral membrane protein, the prototype of seven-helical G-protein-linked receptors. It is found in extended two-dimensional hexagonal arrays in the cytoplasmic membrane of halobacteria, the "purple membrane" patches. Upon illumination, this retinal-containing protein transports protons against a transmembrane gradient. Photoisomerization of the retinal from all-trans to 13-cis,15-anti sets off a sequence of thermal reactions, the "photocycle," in which changes of the retinal cause a cycle of structural changes in the protein, and these result in vectorial proton transfers between donor and acceptor groups. Together, the internal transfers add up to full translocation of a proton from the cytoplasmic to the extracellular side of the protein and thus generate a transmembrane electrochemical gradient for protons. This light-driven gradient is utilized for chemiosmotic coupling, in the same way as in other membranes in which it is generated by redox reactions, to the synthesis of ATP, the uptake of K^+ and nutrients (amino acids), and the transport of Na^+ out of the cells.

2 Overall Structure of the Protein

Bacteriorhodopsin forms trimers, which assemble in the two-dimensional hexagonal lattice of the purple membrane. This specialized membrane contains only bacteriorhodopsin and specific lipids (about 10 lipids/protein), and its regular crystalline lattice made it possible to determine the structure of the protein by cryoelectron microscopy of single sheets, ultimately to 3 Å resolution. Three-dimensional crystals can be grown from detergent-solubilized bacteriorhodopsin in cubic lipid phase, and from X-ray diffraction the structure of the protein has been described at increasingly better resolutions, most recently to 1.47 Å. From these crystallographic studies, the protein is known to consist of seven transmembrane helices with short interhelical loops and short N and C termini. Three of the helices, B, C, and D, are normal to the plane of the membrane and four, A, E, F, and G, are inclined at various but small angles to the perpendicular. The single retinal per protein molecule is bound to the ε-amino group of Lys216 forming a protonated Schiff base near the middle of helix G, while its polyene chain lies at a small angle from the membrane plane (Fig. 1).

The Schiff base divides the seven-helical bundle into extracellular and cytoplasmic halves. The trajectory of the transported proton is through the two "half-channels" formed by these regions. Identification of the residues and bound water that participate in these half-channels has been the objective of much work in the last few decades. Their locations and how they interact are known from crystallography, infrared spectroscopy, and solid-state nuclear magnetic resonance (NMR). The extracellular half-channel contains numerous polar and hydrogen-bonding residues and a three-dimensional network of tightly bound water. Mutational studies have shown that the polar side chains play roles in the release of protons to the extracellular surface. The anionic Asp85 and Asp212 are hydrogen-bonded to the centrally located water402 that receives a hydrogen bond from the protonated retinal Schiff base. The Asp85/Asp212/water402 complex is the counterion to the charge of the Schiff base, but in the photocycle it is Asp85

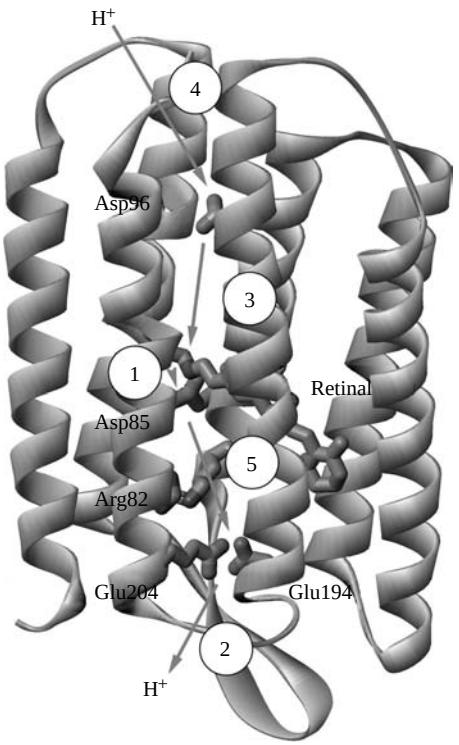

Fig. 1 Three-dimensional structure of bacteriorhodopsin showing the seven transmembrane helices, the retinal, and functionally important residues as marked. The arrows with numbers represent proton transfer events in their sequence. (Reproduced with permission from Luecke et al. (1999) *Science* **286**, 255–260.)

that becomes the acceptor of the Schiff base proton. The pathway for the transported proton is a chain that leads from Asp85 via Arg82, Glu204, and Glu194, with seven intervening water molecules, to the extracellular surface. In contrast, the cytoplasmic half-channel contains hydrophobic residues between the retinal and Asp96, the proton donor in the reprotonation of the Schiff base. The side chain of Thr46 forms a hydrogen bond with the carboxyl group of Asp96, raising its pK_a high enough to keep it protonated. This region is also connected to the retinal, but through a chain of covalent bonds and hydrogen bonds, with the participation of part of helix G that forms a π–bulge near Lys216, two water molecules, and Trp182 that contacts the retinal at the 13-methyl group. The participants of the extracellular and cytoplasmic chains are the candidates for linking the events at the retinal Schiff base first to proton release and then to uptake at the two membrane surfaces.

3 Chromophore and the Photocycle

The chromophore with all-trans, 15-anti retinal has a broad absorption band with a maximum at 568 nm. This is considerably red-shifted from that of retinal (380 nm) or a retinal analogue with a protonated Schiff base (440 nm). The "opsin shift" originates from the only partial compensation of the positively charged retinal Schiff base by the counterion complex, and the distortion of the polyene chain of the retinal in its binding site. The all-trans chromophore exists in thermal equilibrium with the 13-cis,15-syn configuration, which absorbs at 555 nm. Sustained illumination converts the retinal to 100% all-trans, 15-anti ("light-adaptation"), and it is the photoreaction of this isomer that is normally active in transport.

The photochemical cycle of the all-trans chromophore is described by the intermediate states J, K, L, M, N, and O and their substates and the sequence of their interconversions (Fig. 2). Each intermediate is characterized by a distinct absorption maximum in the visible and numerous vibrational bands from the retinal in the infrared and Raman, as well as from the

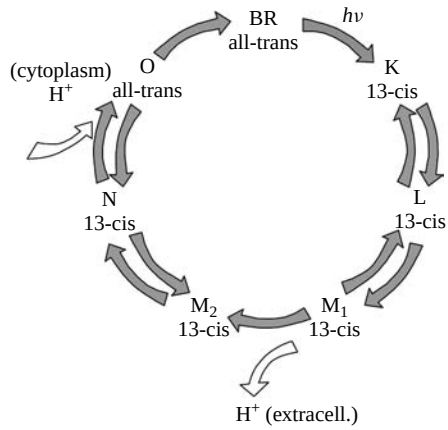

Fig. 2 Photocycle scheme for bacteriorhodopsin. The intermediate states, described in the text, are shown with the isomeric state of the retinal indicated. In all states other than BR and O, the retinal is 13-cis. Proton release is to the extracellular side of the membrane; uptake is from the cytoplasmic side.

protein in the infrared. They correspond to stages of the progress of the photocycle.

3.1 The K State

In the K intermediate that arises on the picosecond timescale, the retinal assumes a twisted 13-cis,15-anti configuration, as indicated by the observation of high-amplitude hydrogen out-of-plane bands, and confirmed by the X-ray diffraction structure of the retinal. The latter had shown that the bond angle at C_{13} is anomalously wide, attributable to the fact that the retinal binding site resists deformation on the timescale in K. This keeps the retinal linearly extended in spite of rotation of the $C_{13}{=}C_{14}$ bond that would otherwise result in a bent polyene chain. Counterrotations of the $C_{14}{-}C_{15}$ and $C_{15}{=}N$ bonds ensure that the direction of the N−H bond remains roughly in the extracellular direction. However, the N−H−O angle that the Schiff base forms with water 402 is now more sharply angled, making its hydrogen bond weaker. This is in accordance with an anomalously low Schiff base band frequency and loss of its deuterium isotope effect, as well as a shift of the O−H stretch bands of a water molecule (assigned to be water 402) to higher frequencies.

3.2 The L State

K converts in about 1 μs into the L state that absorbs at about 540 nm. Fourier-transform infrared (FTIR) spectra indicate that in L, changes begin to appear in the protein also, in both extracellular and cytoplasmic regions. The Schiff base band increases in frequency and regains its deuterium isotope effect, indicating that it forms a stronger hydrogen bond than in K. The X-ray diffraction structure of L indicates that this is because rotations of the $C_{13}{=}C_{14}$ and $C_{15}{=}N$ bonds return the N−H−O geometry toward its initial angle. This means that further relaxation of the retinal by proton transfer from the Schiff base to water 402 and thence to Asp85 is made possible.

3.3 The M States and Proton Release to the Extracellular Side

The M states, with deprotonated Schiff base, have strongly blue-shifted absorption

maximum, near 410 nm. The kinetics of this conversion, on a timescale of tens of microseconds as measured in the visible and the infrared, is multiphasic. Analysis of the kinetics suggests that L is in equilibrium with an early M state, M_1, and the mixture of L and M_1 decays together to form M_2 and then the late M state, M_2' (Fig. 2). At pH > 6, the latter occurs in a unidirectional reaction, and thus L disappears as M_2' is formed. The reason for this appears to be that at the higher pH, a proton dissociates from a site in the extracellular region that interacts with Asp85. The proton is then released to the extracellular surface. The site is likely to consist of bound water, and its function depends on Arg82 and the carboxyl groups of Glu194 and Glu204.

The anomalous titration properties of Asp85 reveal how proton release is coupled to protonation of Asp85. The appearance of two apparent pK_as for Asp85 could be modeled by interaction between the proton affinities of the aspartate and the proton release site. They interact in such a way that either may be protonated but not both. This seems to occur also during the photocycle. When Asp85 becomes protonated by the Schiff base, the pK_a of the proton release group is lowered and it dissociates. When the proton is released to the bulk at a pH higher than the pK_a of the proton release site, the pK_a of Asp85 is in turn driven higher, and the deprotonation of the Schiff base becomes complete. The identity of the proton release is still uncertain. Although Arg82, Glu194, and Glu204 are all involved, the released proton is most probably a proton delocalized in the water network rather than one that originates from these amino acid residues.

Solid-state NMR spectra show that at this time a guanidinium group becomes strongly asymmetrical, suggesting relocation of the side chain of an arginine, very probably Arg82. In the crystallographic structures of the M_2 and M_2' states, the side chain of Arg82 moves away from the region of the now-protonated Asp85 and approaches the region of Glu194/Glu204. This shuttling of Arg82 is therefore the means by which Asp85 interacts with the proton release site and the direct cause of the release of the proton to the surface in M_2'. In M_1, the Schiff base remains initially more or less in the same position as before its reprotonation, but in the M_1 to M_2 reaction it rotates to face the cytoplasmic side. Thus, after deprotonation of the Schiff base there is a drastic change in the geometry of the retinal. Before deprotonation, it faces in the extracellular direction, toward water 402 and Asp85, its proton acceptor, but after its deprotonation it turns to the cytoplasmic direction, toward its eventual proton donor, Asp96. This "reorientation switch" as well as the modulation of the pK_a of Asp85 prevent reprotonation of the Schiff base by reversal of its deprotonation and confer directionality on the pump.

In the X-ray diffraction structures of the intermediate states, the retinal continues to relax as it passes through the M states, and in M_2' it reaches the configuration bent at C_{13}, as expected for the 13-cis isomer. This thrusts the 13-methyl group against the indole ring of Trp182, which moves toward the cytoplasmic side. The movement of Trp182 and rearrangements of the side chains between helices F and G initiate the outward tilt of helix F and an "opening" of the cytoplasmic region. As a result, in M_2, a cluster of three water molecules accumulates near Asp96. One of these, water 504, separates the otherwise hydrogen-bonded Asp96 and Thr46. Another, water 503, connects water 504 with water 502, located near, although

not within hydrogen-bonding distance, the Schiff base. The water cluster in M_2 may be expected to (1) lower the pK_a of Asp96 so as to make it a proton donor and (2) constitute the beginnings of a hydrogen-bonded chain that would, in the next photocycle step, conduct the proton to the Schiff base.

3.4
The N State and Proton Uptake from the Cytoplasmic Side

Reprotonation of the Schiff base by Asp96 produces the N intermediate, which absorbs near 560 nm but with a lower extinction than the initial BR state, as expected for its 13-*cis* retinal. A large-scale protein conformation change in N is evident from electron diffraction, measured either at various times after flash illumination and freezing or at an ambient temperature in a photostationary state of a mutant with long N lifetime. The extensive displacement of the cytoplasmic end of helices F and G in N is confirmed by distance measurements with pairs of spin labels, and with maps of Hg attached to engineered cysteines. The changed accessibilities of spin labels to aqueous quenching agents, and engineered cysteine residues to hydrophilic reagents, further define the conformational changes of the helices and the interhelical loops.

The effects of osmotic agents, hydrostatic pressure, and humidity on the $M_2' \rightarrow$ N reaction and on the protein conformation change, and kinetic analysis of the in-plane cooperativity in the purple membrane lattice, suggest that the rationale of the helical tilt is to increase the hydration of the cytoplasmic region and thereby decrease the pK_a of Asp96. Although widely assumed, the hydrogen-bonded chain of water molecules between Asp96 and the Schiff base has not been directly observed. The conformation with such a chain of water might be a transient state that leads to N, one that does not accumulate in measurable quantities. Recovery of the tilt of helix F is during decay of the N state to the O intermediate, as it is absent in the crystal structure of a mutant that should have an O-like structure. Recovery of the initial state of the cytoplasmic region will reestablish the initial high pK_a of Asp96, and it will have been reprotonated from the cytoplasmic surface.

3.5
The O State

Reisomerization of the retinal to all-trans is made possible by the lowered barrier to bond rotations in the polyene chain upon protonation of the Schiff base. It occurs in the N to O transition, coupled to the reprotonation of Asp96. Residues that contact the chain near the 9-methyl and 13-methyl groups, such as Trp182 and Leu93, facilitate the reisomerization, through steric interaction, that transmits displacements of the protein to the retinal and *vice versa*. The O state has a strongly red-shifted maximum in the visible, at least partly because Asp85 is still protonated, and thus the main component of the counterion to the protonated Schiff base is lacking. Large-amplitude hydrogen out-of-plane vibrations indicate that, as in the K state, the retinal chain is twisted. These features disappear in the final O $\rightarrow$ BR reaction, which appears to be limited by the rate of proton transfer from Asp85 to the still unprotonated proton release site. As expected from the recovery of the low initial pK_a of Asp85, this reaction is unidirectional under all conditions and ensures not only the full repopulation of the initial state but also the

functioning of the proton pump against large transmembrane proton gradients.

4
Properties of Single-site Mutants

Of the 248 amino acid residues in bacteriorhodopsin, only a handful plays an important role in the transport. Their replacement with less functional (e.g. nonprotonable, wrong polarity, or size-mismatched) residues will produce a less effective pump, but only at a saturating light intensity, through increasing the overall turnover time. Only one mutation, that of Lys216, has been reported to eliminate transport altogether by eliminating the covalent binding of retinal. However, even such mutants show transport if reconstituted with a retinal analogue in which the Schiff base is protonated.

The important roles played by Asp85 and Asp96 in bacteriorhodopsin are illustrated, nevertheless, by the altered phenotypes of recombinant proteins in which they are replaced by nonprotonatable residues like asparagine.

When Asp85 is replaced, the purple chromophore becomes blue, similar to the color produced upon protonation of this anionic residue at low pH (<3). The pK_a of the retinal Schiff base is shifted from >13 in the wild type to about 8. The mutation removes the negative charge of Asp85 that is the primary counterion to the positively charged Schiff base and thus the origin of the stability of the protonated Schiff base. The mutation also removes the proton acceptor, and the photocycle of this mutant produces only small amounts of the M intermediate, through loss of the Schiff base proton in the cytoplasmic direction. The D85N mutant shows proton transport at a pH at which the Schiff base is unprotonated, through a series of photocycle steps that occur out of their normal sequence.

Replacement of Asp96 with asparagine greatly stabilizes the M intermediate, as the internal proton donor to the Schiff base is removed. The Schiff base is now reprotonated directly from the cytoplasmic surface, and its rate becomes slow and pH dependent. Mutation of many residues at the cytoplasmic surface near Asp96 does not hinder reprotonation of the Schiff base, but it decouples the reisomerization of the retinal to all-trans from reprotonation of Asp96. In these mutants, the lifetime of N becomes very slow and pH independent.

Replacement of residues implicated in proton release, Arg82, Glu194, and Glu204, with glutamine eliminates proton release during the rise of the M state. Instead, the proton is released to the extracellular surface in the last step of the photocycle, the O to BR reaction. Most likely, with the proton release site rendered nonfunctional, this proton is directly from Asp85 that deprotonates in the last photocycle step in any case.

5
Pump Modes, Energetics

The release and uptake of protons at the extracellular and cytoplasmic surfaces respectively, brought about by the intraprotein proton transfer reactions of the photocycle, can be measured by absorption changes of pH indicator dyes either in the bulk or covalently linked to a surface residue. The pump functions well both at pH > 6, at which proton release to the extracellular surface precedes proton uptake from the cytoplasmic side, and at pH < 6 (the pK_a for proton release), at which the

sequence is reversed. The light-driven proton pump is highly effective: the quantum yield for producing the K intermediate is about 0.6, and the protons transported is close to one/photocycle. From calorimetry, it appears that nearly 50% of the energy gain in the K intermediate is conserved in the form of a transmembrane electrochemical difference for protons. On the other hand, feedback mechanisms, so far not well defined, appear to limit the pump at very high proton gradients, perhaps by shunt reactions.

6
A Family of Retinal Proteins

Bacteriorhodopsin is one of three kinds of similar retinal proteins in halobacterial membranes. Their functions are all based on the photoisomerization of all-trans retinal to 13-cis,15-anti and the protein reactions that accompany its thermal reisomerization. Halorhodopsin is an inward-directed light-driven chloride ion pump. It lacks homologues for Asp85 and Asp96 (containing threonine and alanine at these locations respectively), and the retinal Schiff base does not deprotonate during the photocycle. Sensory rhodopsins I and II are receptors for phototactic behavior. A profound similarity in the mechanisms of these proteins with different functions is indicated by the fact that, for the most part, their activities are interconvertible with minimal perturbations. Thus, the D85T mutant of bacteriorhodopsin binds chloride, exhibits a photocycle similar to that of halorhodopsin, and transports chloride from the extracellular to the cytoplasmic direction. Halorhodopsin, in turn, transports protons when the weak acid, azide, is added. The azide binds near the Schiff base and functions as a proton acceptor like Asp85 in bacteriorhodopsin, and another azide molecule is the proton donor on the cytoplasmic side, as in the D96N mutant of bacteriorhodopsin. Sensory rhodopsin I transports protons such as bacteriorhodopsin when the transducing protein that is normally tightly bound to it is genetically deleted.

Bibliography

Books and Reviews

Ebrey, T.G. (1993) Light Energy Transduction in Bacteriorhodopsin, in: Jackson, M. (Ed.) *Thermodynamics of Membranes, Receptors and Channels*, CRC Press, New York, pp. 353–387.

Haupts, U., Tittor, J., Oesterhelt, D. (1999) Closing in on bacteriorhodopsin: progress in understanding the molecule, *Annu. Rev. Biophys. Biomol. Struct.* **28**, 367–399.

Lanyi, J.K., Váró, G. (1995) The photocycles of bacteriorhodopsin, *Isr. J. Chem.* **35**, 365–386.

Lanyi, J.K. (2000) Molecular mechanism of ion transport in bacteriorhodopsin: Insights from crystallographic, spectroscopic and mutational studies, *J. Phys. Chem. B* **104**, 11441–11448.

Lanyi, J.K. (Ed.) (2000) Bacteriorhodopsin, *Biochim. Biophys. Acta* **1460**, 1–239.

Luecke, H., Lanyi, J.K. (2002) Structural clues to the mechanism of ion pumping in bacteriorhodopsin, *Adv. Protein Chem.* **63**, 111–130.

Maeda, A. (1995) Application of FTIR spectroscopy to the structural study on the function of bacteriorhodopsin, *Isr. J. Chem.* **35**, 387–400.

Mathies, R.A., Lin, S.W., Ames, J.B., Pollard, W.T. (1991) From femtoseconds to biology: mechanism of bacteriorhodopsin's light-driven proton pump, *Annu. Rev. Biophys. Biophys. Chem.* **20**, 491–518.

Primary Literature

Balashov, S.P., Imasheva, E.S., Govindjee, R., Ebrey, T.G. (1996) Titration of aspartate-85 in bacteriorhodopsin: what it says

about chromophore isomerization and proton release, *Biophys. J.* **70**, 473–481.
Blaurock, A.E., Stoeckenius, W. (1971) Structure of the purple membrane, *Nature* **233**, 152–155.
Bogomolni, R.A., Baker, R.A., Lozier, R.H., Stoeckenius, W. (1976) Light-driven proton translocations in *Halobacterium halobium*, *Biochim. Biophys. Acta* **440**, 68–88.
Grigorieff, N., Ceska, T.A., Downing, K.H., Baldwin, J.M., Henderson, R. (1996) Electron-crystallographic refinement of the structure of bacteriorhodopsin, *J. Mol. Biol.* **259**, 393–421.
Luecke, H., Schobert, B., Richter, H.-T., Cartailler, J.-P., Lanyi, J.K. (1999) Structure of bacteriorhodopsin at 1.55 Angstrom resolution, *J. Mol. Biol.* **291**, 899–911.
Luecke, H., Schobert, B., Richter, H.-T., Cartailler, J.-P., Lanyi, J.K. (1999) Structural changes in the M photointermediate of bacteriorhodopsin at 2 Angstrom resolution, *Science* **286**, 255–260.
Lozier, R.H., Niederberger, W. (1977) The photochemical cycle of bacteriorhodopsin, *Fed. Proc.* **36**, 1805–1809.
Sasaki, J., Brown, L.S., Chon, Y.-S., Kandori, H., Maeda, A., Needleman, R., Lanyi, J.K. (1995) Conversion of bacteriorhodopsin into a chloride ion pump, *Science* **269**, 73–75.
Subramaniam, S., Gerstein, M., Oesterhelt, D., Henderson, R. (1993) Electron diffraction analysis of structural changes in the photocycle of bacteriorhodopsin, *EMBO J.* **12**, 1–8.

Behavior Genes

Tanya Button and Peter McGuffin
MRC Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, De Crespigny Park, London, UK

1 Complex Phenotypes 587

2 Quantitative Genetics 588
2.1 Family Studies 589
2.2 Twin Studies 589
2.3 Adoption Studies 591
2.4 Partitioning of the Phenotypic Variance 593
2.5 Path Analysis 596
2.6 Structural Equation Models 597

3 Molecular Genetics 597
3.1 Animal Studies 597
3.2 Human Studies 599
3.2.1 Linkage Analysis 599
3.2.2 Association Studies 601

4 Applications 603
4.1 General Cognitive Ability 603
4.2 Schizophrenia 605
4.3 Mood Disorders 607

5 Future Directions 609

Bibliography 611
Books and Reviews 611
Primary Literature 611

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Keywords

Quantitative Trait Loci
Loci that contribute to the variance of a continuously distributed characteristic.

Concordance
The proportion of a sample of pairs of relatives (e.g. twins) who both show a particular trait or disease. The most useful way of calculating concordance is ''probandwise,'' that is, a series of twins is ascertained via affected index cases or probands. The probandwise concordance is then simply the number of affected cotwins divided by the total number of cotwins.

Relative Risk
Strictly, relative risk is the proportion of individuals exposed to a risk factor for a disorder who are affected divided by the proportion of nonexposed individuals who are affected. In genetic studies, the term is sometimes used slightly more loosely to refer to the proportion of relatives of affected individuals of a particular type (e.g. siblings) who are affected, divided by the proportion of the general population who are affected.

Etiology
The causes of a particular disorder.

Variable Number Tandem Repeat (VNTR)
A small motif of DNA that is repeated tandemly a varying number of times to produce alleles of different sizes.

Single Nucleotide Polymorphism (SNP)
A nucleotide substitution occurring within DNA.

Proband
An individual identified as having a characteristic or disorder of interest through whom other family members are ascertained.

Phenotype
An observable characteristic of an individual.

Behavior genetics is the study of the inheritance of normal psychological traits, such as personality and cognitive ability, as well as abnormal traits and illnesses, such as mood disorders and psychoses. The questions addressed by behavior genetics include the following: Why do individuals differ? Why aren't children of the same parents alike? To what extent are genes and environmental factors accountable for the differences and similarities in human behavioral traits? What are the salient

genes and environmental factors? How do these genes and environmental factors act, coact, and interact to produce the phenotype? To what extent is the covariation of phenotypes the result of the same genes and environmental factors acting on both? Does the extent to which genes affect a phenotype change over the life span and are different genes important at different stages of development? Behavioral traits and disorders, with some rare exceptions, show complex patterns of inheritance involving gene–environment interplay and therefore quantitative genetic methods assume a particularly important role alongside molecular techniques.

1 Complex Phenotypes

Behavioral and psychiatric phenotypes usually exhibit complex, or multifactorial, inheritance, resulting from a combination of multiple genes and environmental contributors, and most normal behavioral traits are continuously distributed in the population. For readers used to dealing with classic Mendelian traits, these features may at first be puzzling. However, if we start by considering a single gene with two alternative forms or alleles, *A* and *a*, as we see in Fig. 1, we have three genotypes *aa, Aa,* and *AA,* whose phenotypic values on some particular scale are x_1, x_2, and x_3 respectively. Depending on the value of x_2, we can have three possible situations. When x_2 is equal to x_3, we have classical dominance of *A* with respect to *a*, whereas if x_2 is equal to x_1, we have recessivity.

If the value of x_2 is exactly midway between x_1 and x_3, we have a purely additive gene effect. Suppose the two alleles have frequencies in the population of p and q (equal to $1 - p$) and assuming there is no migration, mutation, or selection against a genotype, the genotype would be distributed in the population as follows:

Genotypes: *aa Aa AA*

Frequency: $p^2 2pqq^2$

that is, in *Hardy–Weinberg equilibrium.* If, for the sake of simplicity, we take $p = q = 0.5$, and suppose that we have variations about the mean value in each genotype due to environmental factors, we might see three phenotypic distributions as in Fig. 2(a), occurring in the proportions 1 : 2 : 1. For a trait controlled by two loci with each having two alleles of equal frequency and additive effect, we would see, as in Fig. 2(b), five phenotypic distributions occurring in the proportions 1 : 4 : 6 : 4 : 1. In general then, the relative proportions of phenotypes controlled by N loci will be given by $(p + q)^{2^N}$. As the size of N increases, the overall phenotypic distribution more closely approximates a normal distribution, which is the limiting case when N becomes very large. It is thought that most continuously

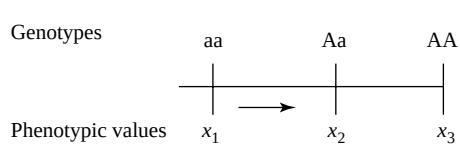

Fig. 1 Genotypes and corresponding phenotypic values.

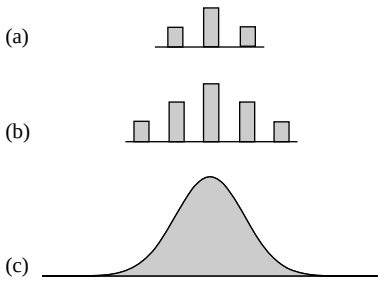

Fig. 2 Distribution of traits influenced by additive genetic effects. (a) A single biallelic gene gives rise to three phenotypes; (b) two biallelic genes give rise to five different phenotypes; (c) numerous additive genetic and environmental factors give rise to continuously distributed phenotypes.

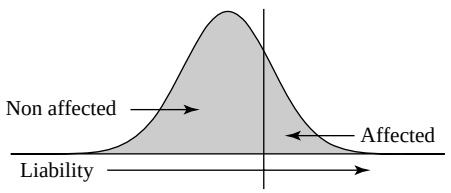

Fig. 3 A liability threshold model in which it is assumed that multiple genes and environmental factors contribute to the underlying liability to a produce a normally distributed liability towards a disorder. A person who falls below the predefined threshold is not affected by the disorder whilst all those who fall above the threshold are affected.

distributed hereditary traits are due to *polygenic* mechanisms of this kind, with each individual gene still being transmitted according to Mendel's laws. The multiple genes that contribute to the continuous variation of a multifactorial phenotype are referred to as quantitative trait loci (QTL).

Some behavioral traits such as psychiatric disorders display discontinuous characteristics, for example, when an individual is categorized as unaffected or affected by schizophrenia, or depression. Although some comparatively rare disorders such as Huntington's disease and early-onset varieties of familial Alzheimer's disease show Mendelian segregation, most psychiatric disorders do not conform to Mendelian patterns of inheritance. For such phenotypes, we can consider that there is an underlying continuum of liability to the disorder. It is usually assumed that liability has an approximately normal distribution resulting from the combined effects of multiple QTLs and environmental effects. Individuals who fall below a predefined threshold are unaffected whilst those who at some point exceed it are affected. This is referred to as the liability threshold model (see Fig. 3).

2 Quantitative Genetics

Phenotypic variance results from both genetic and environmental influences, and one of the concerns of quantitative genetics is the partitioning of the phenotypic variance into its environmental and genetic components. This is discussed in more detail in Sect. 2.4. The phenotypic covariance between family members results from both their shared genes and shared environments. By taking advantage of the different genetic and environmental relationships between family members, we are able to quantify the extent to which genes and environment influence a phenotype. To do this, we employ family, twin, and adoption studies.

2.1 Family Studies

Family studies are used to establish the extent to which behavioral phenotypes cluster in families. For example, a familial influence would result in relatives correlating positively for a continuous trait or, when studying a disorder, relatives of probands (individuals affected by the disorder, also referred to as an index case) having a higher prevalence than the relatives of controls (individuals unaffected by a disorder, often matched to the proband for characteristics such as age, sex, and ethnicity). We can also compare correlations and concordances between family members with differing degrees of genetic relatedness to the proband. The main sources of genetic correlation or covariance between relatives are called additive variance and dominance variance. Essentially, additive variance reflects the extent to which the QTLs contributing to a trait show simple additive effects and dominance variance reflects the extent to which heterozygotes deviate from the mid values between homozygotes (Fig. 1). The genetic covariance between pairs of relatives is given by the simple expression $\beta V_A + \alpha V_D$, where β is the (average) proportion of genes held in common, α is the probability that at any given locus the pair of relatives share two alleles that are identical by descent (IBD) from common ancestors, V_A is the additive genetic variance, and V_D is the nonadditive genetic, or dominance variance. In addition to genetic covariance, family members share the home environment in which they are raised and this contributes a proportion of variance V_C. The sources of covariance between different family members can be seen in Table 1.

One of the major limitations associated with family studies is the inability to distinguish between genetic and environmental effects. So-called competition effects between family members may reduce resemblance between relatives and make a trait appear less genetically influenced than it actually is. More commonly shared environment contributes positively to resemblance between family members and therefore familial clustering of a phenotype is not sufficient evidence of genetic causation.

2.2 Twin Studies

Identical, or monozygotic (MZ) twins are the product of a single fertilized egg

Tab. 1 The sources of covariance between pairs of relatives.

Relationship to proband	*Genetic relationship*	*Phenotypic covariance*
MZ twins	$V_A + V_D$	$V_A + V_D + V_C$
DZ twins	$0.5V_A + 0.25V_D$	$0.5V_A + 0.25V_D + V_C$
Full sibs	$0.5V_A + 0.25V_D$	$0.5V_A + 0.25V_D + V_C$
Parent or offspring	$0.5V_A$	$0.5V_A$
Half sibs raised together	$0.25V_A$	$0.25V_A + V_C$
Raised apart		$0.25V_A$
Uncles/aunts/nephews/nieces	$0.25V_A$	$0.25V_A$
Cousins	$0.125V_A$	$0.125V_A$
Grandparents	$0.25V_A$	$0.25V_A$

and consequently are genetically identical. Fraternal or dizygotic (DZ) twins, on the other hand, result from two separate fertilized eggs and so share on average only 50% of their genes, the same as full siblings. Hence, if genes contribute toward predisposition for a phenotype, MZ twins will tend to be more similar than DZ twins. If MZ twin pair correlations or concordances significantly exceed those observed for DZ twins, we can conclude a genetic influence for the phenotype, with environmental variables accounting for all the difference between MZ twins. As full sibs share, on average, the same number of genes as DZ twins, it would theoretically be possible to substitute them into such analyses. However, DZ twins are a better comparison since they are the same age and are likely to have shared more similar environments and major life events. One of the requirements of twin studies is that correct determination of zygosity is essential, and misclassification can result in meaningless results. For example, if genetic effects are present and MZ pairs are incorrectly assigned as DZ, the DZ correlations will be artificially inflated, decreasing the difference between MZ and DZ correlations and resulting in an underestimate of heritability, and an overestimate of the common environment effect. In practice, zygosity assignment is frequently based on questionnaires. These typically ask such questions as whether the twins are often mistaken one for the other by friends or if they are as alike as "two peas in a pod." While genotyping at several highly polymorphic DNA markers is a more accurate method of zygosity testing, it is also far more costly and may be uneconomical for very large studies where the approximately 95% accuracy of questionnaires may be acceptable.

A potential limitation is that twins have atypical obstetric and perinatal histories and suffer from birth complications and low birth weight more frequently than nontwins do, and consequently, conclusions drawn from twin studies may not be generalizable to nontwins. For example, on average, the early cognitive ability of twins is slightly lower than that of singletons and twins probably have a slightly elevated risk of mental retardation. However, in practice, the long-term impact of this for most behaviors is limited, and by approximately five years of age, biological, behavioral, and emotional differences between twins and nontwins would have diminished.

The classic twin approach depends on the "equal environments assumption", that is, MZ twins share their environments to roughly the same extent as DZ twins. However, some studies suggest that MZ twins may share more similar postnatal environments, such as having more friends in common, and they may also be perceived and treated more similarly than DZ twins. The extent to which this may invalidate the results of twin studies has been tested in a number of ways. One approach is to attempt to measure aspects of environmental sharing and examine whether they are associated with the trait being studied. For example, a study of depressive disorder in adult same-sex twins found that MZ twins had on average shared a bedroom for longer as children and were in more frequent contact as adults than DZ twins, but these indicators of environmental sharing did not predict concordance for depression. Another approach is to look at the effects of mistaken zygosity on parental treatment of their twins. Up to 20% of parents think that their MZ twins are DZ or vice versa. The prediction might be that MZ twins thought to be DZ by their parents would be treated less similarly than MZ

twins where the parents were correct in their assumption about zygosity. DZ twins thought to be MZ would be treated more similarly than correctly assigned DZ twins. In fact, studies of traits such as personality, intelligence quotient (IQ), and depression symptoms in childhood have found that parental misclassification has negligible effect on twin resemblance. Finally, studies of MZ twins who have been reared apart make it possible to eliminate the effects of shared environment altogether, and comparisons of the MZ twins reared together with those reared apart for traits such as personality reveal that correlations are similar for both types. Thus, even if MZ twins do share a greater environment than DZ twin pairs, it does not significantly influence their similarity for behavioral phenotypes. Therefore, the available evidence suggests that the equal environment assumption is a reasonable conjecture. The potential effects of prenatal environment also need to be considered in twin studies as MZ twins often share a chorion, whereas DZ twins never do. This may also influence MZ twin similarity, although there is little evidence to support this.

Twin studies also assume random mating of the parents, in which proband parents do not correlate more for the observed phenotype than would be expected by chance. However, assortative mating where the parents correlate highly for the trait has been detected for some phenotypes, such as IQ and antisocial behavior. This may increase the proportion of predisposing genes shared by DZ twins above the expected 50%; thus DZ twins would be phenotypically more similar than if random mating was occurring. There is no effect on the genetic relatedness between MZ twins as they already share 100% of their genes. Consequently, the difference between MZ and DZ twin correlations will be reduced, resulting in underestimation of the heritability.

Despite their limitations, twin studies remain the most used tool in disentangling the genetic and environmental influences on behavior. Figure 4 summarizes twin study results for a number of disorders and traits. There are clear MZ–DZ differences for some phenotypes such as schizophrenia, manic depressive disorder, unipolar depression, and cognitive ability as measured by IQ tests, suggesting genetic effects. In the case of bulimic behavior and childhood conduct disorder, there is evidence of familiality (there are positive correlations) but little suggestion of genetic effects (modest differences in the MZ and DZ correlations).

However, the classic twin study is not the end of the story. There are a number of possible directions in which twin studies can be elaborated upon. Extending twin studies to include other family members is one useful method that has already been applied and which has potential for future research. Furthermore, the offspring of MZ twin pairs are of particular interest. The son of an MZ father shares 50% of his genes with both his father and his uncle, thus comparison of the correlation between uncles and nephews with the father and son pairs may inform us more about the genetic basis of the phenotype, whilst eliminating any confounding effects of common environment or G × E correlation. Similarly, we could compare both sets of offspring from such a family set-up as these are genetically similar to half sibs. See Fig. 5.

2.3 Adoption Studies

The effects of genes and environment on development of behavioral traits can also

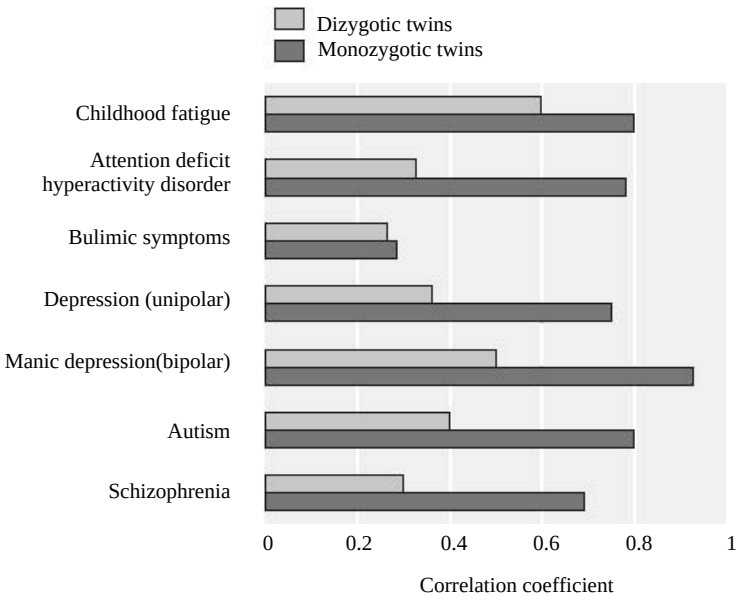

Fig. 4 Monozygotic and dizygotic twin correlations for various disorders. (Taken from McGuffin, P., Martin, N. (1999) Science, medicine, and the future. Behaviour and genes, *BMJ* **319**, 37–40.)

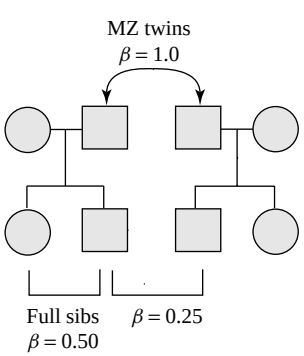

Fig. 5 The genetic relationships (β) between the offspring of monozygotic twins.

be differentiated using adoption studies, although for practical reasons, these have been used less extensively than twin studies. There are three main variations on the adoption design. Firstly, the adoptee study design compares the frequency of a disorder or trait in the adopted-away offspring of parents who have the trait and adopted-away offspring of parents who do not. A second type is the adoptee's family study. This involves taking affected adoptees as the index cases. The prevalence of the trait in biological relatives can be compared with that in adoptive relatives or to that in the relatives of control adoptees. A higher rate of disorder in those biologically related to the index cases indicates a genetic effect. One can also compare the prevalence of the disorder in the biological parents of adopted-away probands with that in the

biological parents of nonadoptees as a test of whether the environment or rearing has any effect. A third approach is the cross-fostering study, which compares the rates of disorder in adoptees with no biological risk but raised by an affected adoptive parent with the disorder in adoptees with biological but not environmental risk. This is the most complicated and practically difficult design but it allows the possibility of testing for the effects of genes, environment of rearing, and the combination of the two. An example was a study of antisocial behavior by Mednick and colleagues who found an increased risk of having a criminal record in male adoptees with either a criminal adoptive father or a criminal biological father but an even higher rate in those who had both a biological and an adoptive father who had a criminal record. However, the disorder where adoption designs have been used most extensively is schizophrenia. Some of the main results from schizophrenia adoption studies are summarized in Table 2. They show a consistent pattern favoring a genetic contribution but in contrast with the study of antisocial behavior just mentioned, there is little evidence of a family environmental effect.

A general criticism of adoption studies is that adoptees may not be representative of the population as a whole. For example, the circumstances surrounding adoption may be particularly stressful placing adoptees at increased risk of behavioral and psychiatric disorders. Parents who give up their children for adoption may do so because of illness and therefore will tend to have increased levels of psychopathology. Finally, adoptive parents are likely to be unrepresentative in the opposite direction as adoption agencies will tend to screen out many "negative" attributes such as preexisting psychopathology when placing a child.

A related complication with adoption studies is selective placement. Some agencies attempt to place adoptees with families that resemble their biological parents in terms of social or ethnic background. This could result in genetic relatives, who have been independently adopted, sharing a correlated environment. For example, a tendency toward selective placement has been suggested in some studies of IQ, although there is little evidence for selective placement in many other traits. Another fundamental limitation of adoption studies is that adoption is becoming rarer as a result of increasing contraception and abortion, along with the removal of the stigma of being a single parent.

2.4 Partitioning of the Phenotypic Variance

As described above, observed phenotype (V_P) can be partitioned into genetic (V_G) and environmental (V_E) components, such that:

$$V_P = V_G + V_E \tag{1}$$

Furthermore, the genetic variance can be partitioned into additive genetic effects (V_A) and dominance deviation effects (V_D) that result from interactions of the alleles found at a single locus. We may also encounter epistasis, the interaction or multiplicative effect of genes at different loci. In practice, this is not easy to distinguish from dominance in human studies. However, one example of a disorder where the pattern of inheritance would be in keeping with epistasis is schizophrenia, where the concordance rate (i.e. the proportion) of relatives and of index cases who are also affected is around

Tab. 2 Examples of adoption studies of schizophrenia, not age-corrected. (Taken from McGuffin, P., Owen, M. J., O'Donovan, M. C., Thapar, A., Gottesman, I. I. (1994) *Seminars in Psychiatric Genetics*, Gaskell, London.)

Study	*Type of study*	*Diagnosis*	*Genetic relatives of a schizophrenic*	*Not genetically related to schizophrenic*
Heston (1966)	Adoptee	Schizophrenia	10.6% of 47 adoptees who had a schizophrenic mother	0% of 50 control adoptees
Rosenthal et al. (1968)	Adoptee	Schizophrenia spectrum disorder	18.8% of 69 children of schizophrenics raised by normals	10.1% of 79 control adoptees
Wender et al. (1974)	Cross-fostering	Schizophrenia spectrum disorder	18.8% of 69 children raised by normals	10.7% of 28 children of controls raised by future schizophrenics
Kety (1983) Kety et al. (1994)	Adoptee's family: national sample (47 chronic schizophrenic adoptees)	Chronic and latent (DSM-II) schizophrenics	15.8% of 279 biological relatives of adopted-away schizophrenics	18% of 228 adoptive relatives of schizophrenics and relatives of control adoptees
Kendler et al. (1984, 1994)	Reassessment of Kety's data (31 adoptees with spectrum disorder)	DSM-III schizophrenia, schizotypal personality disorder, RDC schizoaffective disorder, mainly schizophrenic	14.4% of 209 biological relatives of adopted-away schizophrenics (23.5% in first-degree relatives, 9.9% in second-degree relatives)	3% of 299 adoptive relatives of schizophrenics and relatives of control adoptees
Tienari (2000)	Adoptee	Schizophrenic/paranoid psychosis	6.7% of 164 adoptees who had schizophrenic biological parents	2.0% of 197 control adoptees

50% for MZ twins, 10 to 12% for DZ twins or full siblings, but only 2 to 3% for second-degree relatives such as uncles/aunts or grandparents. One explanation of why concordance declines more steeply than the proportion of shared genes (see Fig. 6) would be an interactive effect of multiple loci. For example, Risch has suggested on mathematical grounds that just two or three loci with epistatic effects would be sufficient to explain the inheritance of schizophrenia. However, recent molecular studies appear to indicate that a larger number of loci is involved.

As we also discussed earlier, part of the environmental variance is common or shared environment (V_C), which is defined as any environment that is common to a pair of relatives and makes them more similar. In contrast, individual-specific, or nonshared environmental variance results from any environmental effects that produce differences between relatives (V_E).

Phenotypic variance may also be affected by gene–environment interaction and gene–environment correlation. Gene–environment interaction occurs when specific genotypes confer susceptibility to specific environments. For instance, it has been demonstrated that adoptees who faced adversity during their upbringing were at increased risk for criminality if at least one biological parent had antisocial behavior. More recently, Caspi and his colleagues found more specific evidence of gene–environment interaction. Young men who had experienced early parental maltreatment had an increased risk of antisocial behavior if they had inherited the X chromosomal gene that results in low activity of an enzyme involved in brain signaling, monoamine oxidase A (MAOA). Men who had the alternative, high-activity form of the MAOA gene and early maltreatment showed no significant increase in antisocial behavior when compared with men who had not been maltreated. Gene–environment correlation may result from the covariation genotype and environmental factors. Three types of gene–environment (G-E) correlation have been described: passive, active, and evocative. Passive G-E correlation arises because children inherit both genetic risk factors and environment factors from their parents. For example, a musically talented parent could transmit both predisposing genes and a predisposing environment to their offspring. Active correlation occurs because people search out or modify environments to suit their genetic predisposition. An example of this would be a person of high musical ability seeking out environments where they could make best use of their talents.

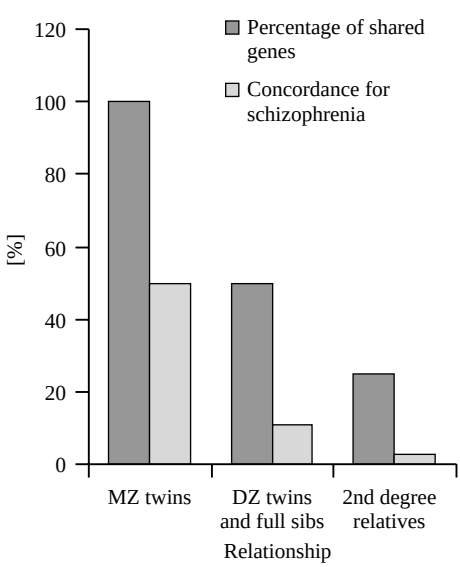

Fig. 6 The genetic concordances between family members and corresponding concordances for schizophrenia. The concordances decline more steeply than the percentage of shared genes.

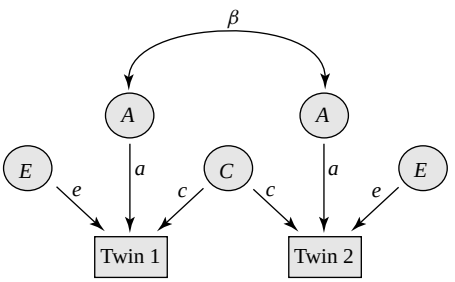

Fig. 7 An ACE path diagram. Path analysis can be used to estimate the contribution of the path coefficients *a*, *c*, and *e*, to the covariance or correlation between twin pairs. β, genetic correlation between twin pairs, 1.0 for monozygotic twin pairs and 0.5 for dizygotic twin pairs; *a*, path coefficient for additive genetic effects; *c*, path coefficient for common environmental effects; *e*, path coefficient for nonshared environmental effects; *A*, additive genetic variance; *C*, common environment variance; *E*, nonshared environment variance.

Finally, evocative correlation results from others' reactions to a person based at least partially on that person's genotype. For example, a musically gifted child's abilities may be recognized by her teachers and parents, who ensure that she has the tuition and equipment to fulfil her promise.

Therefore, the phenotypic variance is the consequence of additive genetic variance, nonadditive genetic variance, common environmental variance, and nonshared environmental variance, as well as the effect of gene–environment interactions and correlations. This can be expressed in Eq. (2):

$$V_P = V_A + V_D + V_C + V_E + V_{G\times E} + 2\mathrm{Cov}_{GE} \qquad (2)$$

where $V_{G\times E}$ is the variance due to gene–environment interaction and $2\mathrm{Cov}_{GE}$ is the covariance due to gene–environment correlation.

Although genetic and environmental variance cannot be directly measured or observed, it may be possible to estimate them indirectly from the observed phenotypic variances and covariances.

The extent to which genes are accountable for the phenotypic variance in the population, or the proportion of variance in the liability for a phenotype is referred to as the heritability. Strict sense or narrow heritability refers only to the additive genetic effects whereas broad heritability (or degree of genetic determination) is the proportion of the phenotypic variance accounted for by total genetic variance V_G, including both additive and nonadditive effects.

$$\text{Broad heritability} = \frac{V_G}{V_P} \qquad (3)$$

$$\text{Narrow heritability} = \frac{V_A}{V_P} \qquad (4)$$

Similarly, it may be of interest to estimate the extent to which the variance of a phenotype is influenced by shared environment, V_C/V_P, and by environment that is not shared, V_E/V_P. A useful methodology for such tasks is provided by path analysis.

2.5 Path Analysis

Figure 7 illustrates a simple path model for a trait where it is assumed that the variance can be explained by additive genetic effects (A), shared environmental effects (C), and nonshared environment (E). The observed correlation between a series of twins (twin 1 and twin 2) can be calculated for the trait under investigation. The expected value of the correlation between the twins in terms of path coefficients can be obtained by tracing the connecting paths, providing the equation:

$$r(\text{Twin 1})(\text{Twin 2}) = (a\times\beta\times a) + (c\times c) = \beta a^2 + c^2 \qquad (5)$$

where the correlation between genotypes $\beta = 1.0$ for MZ twins and 0.5 for DZ twins.

Therefore, if we calculate the correlations based on observed measures for MZ and DZ twins, rMZ and rDZ, we have a set of simultaneous equations that can be easily solved for a^2 and c^2, giving:

$$a^2 = 2(r\mathrm{MZ} - r\mathrm{DZ}) \quad (6)$$

$$c^2 = 2r\mathrm{DZ} - r\mathrm{MZ} \quad (7)$$

Although path analysis is most commonly carried out with twin data, the approach is perfectly general and can be applied to other types of relative pairs. It can also incorporate other types of observations such as direct environmental measures.

2.6 Structural Equation Models

Structural equation modeling enables formal testing of the extent to which the genetic and environmental effects described above contribute to the variance of the phenotype, or the variance of liability of categorical phenotypes. This is typically performed using computer programs such as LISREL or Mx, which use an iterative process to test which estimates for each of the parameters best explain both the observed data and the expected variance and covariance values. These expected values are calculated by the program using information regarding the causes of covariation, for example, that the covariance between MZ twins results from genetic and common environmental effects, and that between DZ twins is due to half the genetic effects and all the common environment. The best fitting model is the one in which there is maximum agreement between the expected and observed covariances. This is achieved by maximizing a likelihood function or minimizing the goodness-of-fit χ^2.

A principle of parsimony is adopted whereby there is an attempt to explain the observed data with as few parameters as possible. Thus nested models, in which parameters are dropped, are tested and compared with the full model to see if the remaining parameters can account for all the similarity between the twins without significantly worsening the fit of the model.

Model fitting need not be restricted to explaining the cause of phenotypic variance and covariance. It can also be used, for example, to identify the existence of qualitative and quantitative sex differences, identify the causes of phenotypic comorbidity, assess sibling interaction effects, and tease out gene–environment interaction effects.

3 Molecular Genetics

3.1 Animal Studies

Animal models are useful in the identification of genes with a role in behavior since it is possible to modify and manipulate both the genome and environment of laboratory animals in a way that is not possible in human samples. Other mammals have a high level of genetic homology with humans, and also display certain behaviors that are analogs of human traits. For example, the mouse genome shares approximately 85% homology with humans and mice show behaviors analogous to such human behaviors as anxiety, aggression, and hyperactivity.

Animal studies use inbreeding and selection studies to identify the presence of

a genetic influence on behavior. Inbreeding, the mating together of close relatives such as siblings, when repeated over many generations, creates animals that are virtually identical. Consequently, differences within strains are largely or entirely due nongenetic effects, whilst differences between strains kept in similar environments are due to genetic effects. Inbreeding has been used to demonstrate a genetic basis to numerous mouse behaviors such as fearfulness, held to be a model of human anxiety, and ability to solve problems such as mazes, which provides analogies for components of human intelligence. Currently, there are over 100 inbred strains of mice commercially available. Mouse models are, of course, limited when it comes to more complex human behaviors and cannot shed direct light on behaviors involving speech and language. Another limitation is that inbred strains may not represent the behavioral repertoire of naturally segregating genetic populations. Another practical problem is that inbreeding depression may result in reduced infertility, interfering with breeding of laboratory strains.

The ability to select for a trait is only possible when there is an underlying genetic contribution. Thus, if we are able to selectively breed for a trait, we can infer a genetic component. For example, selection studies have been carried out with regard to so-called open field activity in rodents, where the rodents are introduced to an enclosure that is open at the top, often in the form of a cage with beams of light transmitted near the floor to record activity. A rat or mouse that displays low levels of activity and frequently "freezes" and defecates in an open field is thought to display the equivalent of high anxiety, whilst one that is highly active in an open field and shows less frequent defecation is thought to have low anxiety. By selectively mating together animals that show high activity, and mating together those that display low activity it has been possible to develop behaviorally distinct lines, the best known of which are the "Maudsley reactive" and "Maudsley nonreactive" rats.

There are a number of methods used to identify genes using animal models. These include studies of transgenic, knockout, and recombinant inbred strains. In transgenic mice, genes from another species, such as humans, are inserted into the mouse genome, and the effects of the gene can be observed. Genes such as the Dopamine D1 receptor gene has been implicated in anxiety and hyperactivity using transgenic mice.

Knockout mice result from eliminating activity of genes thought to be influential and observing the effect this has on the animal's behavior. This technique has been used to show that the monoamine oxidase A gene and the nitric oxide synthase gene both play a role in mouse aggression as knockout mice for each of these displayed higher rates of aggression compared to the wild-type mice.

By mating together a single male and female from two different inbred strains and then carrying out brother sister matings for 20 or more generations, it is possible to develop a recombinant inbred (RI) strain. Like any form of inbred strain, RIs are essentially identical and each contains only two alleles for each locus, one from each of the original inbred strains from which they were derived. It is possible to identify which of the progenitor strains each of the alleles at a locus originate from. When two alleles at different loci originate from the same progenitor strain, they are said to be concordant, and when they are from different progenitors, they are said to be discordant. The greater

the physical distance between two alleles, the greater the chance of discordance. RI strains are ideal for studying linkage and have a number of advantages. As they are genetically identical, all people working on an RI strain are essentially using the same animal and thus each strain needs to be genotyped only once. Moreover, the intensive inbreeding reduces the influence of environmental effects on a phenotype, making it possible to study phenotypes of low and moderate heritability and complex traits.

3.2 Human Studies

Quantitative genetic approaches to behavior can tell us whether genes are involved in a particular trait and provide us with estimates of how much genes contribute. However, such statistical approaches are not ends in themselves and molecular genetic approaches are the next stage allowing us to move on to discover what genes are involved and how they act in concert with the environment.

Broadly, there are two approaches to identifying the genes involved in complex traits. These are candidate gene studies and positional cloning. Candidate genes are those that encode for proteins where there is a prior hypothesis that they might be involved in a trait or disorder. Positional cloning starts out with no prior hypothesis except that there is a genetic contribution to the trait. The locations of genes are then inferred by linkage and association (Sect. 3.2.1) and then this positional information is refined until the genes are pinpointed and identified. Since at least half of all genes are expressed in the brain, the potential number of behavioral candidate genes is overwhelming and has to be narrowed down in some way. The most common approach with regard to psychiatric disorders has been to draw up a list of candidates based on informed guesswork with useful leads being provided by the mode of action of drugs. For example, all of the standard drug treatments used in schizophrenia, the "typical" antipsychotics such as Chlorpromazine and Haloperidol, are known to block a class of receptors for the neurotransmitter dopamine, called D2 receptors. The newer "atypical" antipsychotic drugs (e.g. Clozapine and Olanzapine) block both dopamine and serotonin receptors in the brain, most importantly, a type of receptor designated 5HT2A. Consequently, researchers have attempted to discover polymorphisms (variants) in the 5HT2A receptor gene and in D2 receptor genes, and compare their frequency in cases affected by schizophrenia and controls who do not have the disorder. The results overall suggest a small but significant effect of the variant in the 5HT2A receptor gene and a dopamine receptor gene DRD3, a member of the D2-like family.

Such investigations are examples of allelic association studies, to which we will return to discuss in more detail in Sect. 3.2.2. A complementary approach and one that is the starting point for positional cloning is to carry out genetic linkage studies. Linkage studies provide information about the location of genes involved in traits or disorders where the causation is unknown. Researchers can then move from location to gene identification and then to studies of structure and sequence and finally to studies of the gene products themselves.

3.2.1 Linkage Analysis

Loci that are on different chromosomes or that are far apart on the same chromosome obey Mendel's law of independent

assortment, that is, they are transmitted together from one generation to the next, or cosegregate, at a frequency no higher than chance. Loci that are close together tend to depart from chance assortment and, within certain limits, the frequency of recombination, or crossing-over between loci on paired chromosomes is proportional to the physical distance between them. In gene mapping studies, the genome is measured in units called centimorgans (cM) with 1 cM corresponding to a pair of loci showing crossing-over, or recombination, one meiosis in a hundred. The whole genome, averaged across the sexes (there is more crossing-over in female than in male meioses) is about 3500 cM long, with 1 cM roughly corresponding to just under a mega base of DNA (or a million base pairs). Linkage can be detected over relatively large distances, 10 cM or more, so that it is possible to mount a whole genome search using just a few hundred evenly spaced markers. The disadvantage of linkage is that it is capable of detecting only comparatively large effects. Linkage is therefore straightforward in rare, simple Mendelian neuropsychiatric disorders such as Huntingdon's disease or some forms of early-onset familial Alzheimer's disease that show autosomal dominant inheritance. However, in the case of common, complex traits, such as schizophrenia or affective disorders, linkage studies present more of a challenge.

Parametric linkage analysis While in studies of plants or animals under laboratory conditions it may be possible to arrange crosses on a large scale and simply count the numbers of recombinants and nonrecombinants, studies of humans are more complicated. The standard statistical approach to detecting linkage and estimating recombination is to calculate LOD (log of the odds) scores. An LOD score is the common log of the ratio of the probability that a recombination fraction has a certain value of less than one half to the probability that the recombination fraction equals one half. LOD scores are calculated for a range of values of the recombination fraction from 0 to 0.5 and the value that gives the maximum LOD is taken as the most likely value of the recombination fraction. By convention, an LOD of three or more (i.e. odds on linkage of 1000 to 1) are accepted as sufficient support for linkage and an LOD of – 2 (i.e. odds of 100 to 1 against) is accepted as excluding the presence of linkage. Unfortunately, LOD scores were originally intended for simple traits. Therefore, one of the difficulties that besets attempts to detect linkage in psychiatric disorders and other common diseases is that the mode of transmission is unknown. That is, the disease does not usually follow a simple Mendelian pattern of segregation in families.

The LOD score approach can be extended to complex traits by incorporating penetrance parameters. Penetrance is defined as the probability that a particular phenotype will be manifested given a certain genotype. In Mendelian traits, penetrances are always either 0 or 1. For example, if we consider a trait where there are two alleles, A_1 and A_2, and where A_2 is the "disease" allele, there will be three possible genotypes A_1A_1, A_1A_2, and A_2A_2. If A_2 is recessive, the three genotypes will show penetrances of 0, 0, and 1 respectively, whereas if A_2 is dominant, the penetrances will be respectively 0, 1, and 1. Complex traits that depart from Mendelian patterns can be modeled using a more general single major locus (SML) approach where penetrances

can take values between 0 and 1. However, in order for linkage studies to be carried out accurately, the penetrance parameters together with the disease gene frequency need to be specified correctly. Otherwise, at best, the estimate of recombination will be incorrect, and at worst, linkage will not be detected even when it is present.

Other problems include possible genetic heterogeneity in common disorders and relative imprecision in defining who is affected or unaffected within families. This is not a difficulty when there are clear-cut cases and clear-cut healthy family members, but common diseases often also show milder or uncertain cases within the same families. The most straightforward way of dealing with these problems has been to concentrate on extended families with many affected members and to make an informed guess at the mode of transmission. It is also assumed that even if there is heterogeneity in the disorder as a whole, there is homogeneity at least within the families. Such simplifying assumptions have worked well with some common diseases, with the most notable success amongst psychiatric disorders being early-onset familial Alzheimer's disease. However, in other disorders such as schizophrenia, concentration on large multiply affected families has produced results that are more difficult to interpret, hence there has been a move toward using nonparametric or ''model-free'' methods of analysis.

Nonparametric linkage analysis The most common approach concentrates on affected sibling pairs. Affected sib pairs may be more representative of a common disorder than subjects found in multiplex pedigrees and, in general, finding families containing affected sib pairs is easier than finding extending pedigrees. However, the real advantage of focusing on sib pairs who are both affected by the disorder is that the statistical analysis is relatively straightforward and robust. Detection of linkage depends on the fact that at any given locus, the probabilities of a pair of siblings inheriting 0, 1, or 2 alleles that are identical by descent are respectively 1/4, 1/2, and 1/4. Greater allele sharing at a marker than would be expected by chance in pairs of siblings who are both affected by the disease suggests that the marker locus is close to a gene conferring susceptibility to the disorder. The unfortunate disadvantage of sib pair methods is that although they are simple and robust, there is a comparative lack of statistical power. Both practical experience and simulation studies suggest that unless the sample sizes are extremely large, linkage is difficult to detect with genes that confer a relative risk of a disorder of much less than about three or account for less than about 10% of a variation in liability to the disorder. Genes with effects as large as that may be rare in common diseases and so, for example, a genome scan for linkage in schizophrenia in almost 200 families effectively excluded any gene having a relative risk of three from most of the genome. The need to detect genes having only small effects has brought about a renewed interest in applying allelic association studies in a more systematic fashion than just focusing on potential candidate genes.

3.2.2 Association Studies

Allelic association is the phenomenon whereby a particular allele at a marker locus is found together with a trait or disorder more than would be expected by chance. In contrast with linkage, the same allele is found across the population rather

than just within particular families. The simplest and most commonly used study design is to compare allelic frequency between affected cases and controls. It has long been known that association can detect those genes that account for as little as 1% of the variance in a trait. However, association only occurs either if the marker itself contributes to the trait or disorder or if a gene contributing to the trait or disorder is so close to the marker locus that there is linkage disequilibrium (LD). LD is the phenomenon whereby alleles at adjacent loci are found in a combination that is undisturbed over many generations of potential crossing-over.

One of the difficulties in interpreting association studies is that recent admixtures of populations can bring about stratification. This occurs when two ethnic groups have different frequencies of a disease and the trait that is being studied and also have different frequencies of marker alleles. Consequently, in mixed populations from these groups, there may be a spurious association unless cases and controls are carefully matched for ethnicity. To overcome the need to carry out case control matching, which may be difficult in some highly admixed populations where there has been much immigration, one can use family based methods with "internal controls." For example, in the transmission disequilibrium test (TDT), trios consisting of an affected offspring and both parents are studied. At least one parent needs to be heterozygous at the marker locus. It is then possible to test the frequency of affected offspring to whom a particular allele is transmitted, with the frequency of those who do not receive that allele. That is, the untransmitted marker alleles are effectively the "controls" while the transmitted alleles are the "case alleles." A significant distortion of the chance expectation of transmission of alleles is easily tested by a simple statistic called the *McNemar* χ^2 where a significant result indicates that LD is present. The principles of TDT are shown in Fig. 8.

The drawback of allelic association studies is that LD can only be detected over very short distances and hence many thousands, not hundreds, of markers would be needed to carry out a whole genome scan. There is currently considerable debate over the approximate number of markers that would be required to carry out a genome scan using LD. One controversial suggestion based on simulations suggests that this might be as many as 500 000 markers. However, some empirical studies have shown that in moderately outbred populations, LD can be detected over distances up to 1 cM. Although LD is not evenly distributed across the genome, this would mean that a few thousand markers would probably be sufficient for an initial screen of the genome. Even so, until recently, a genome scan using this number of markers was not feasible, but now the availability of very dense marker maps based on single nucleotide polymorphisms (SNPs)

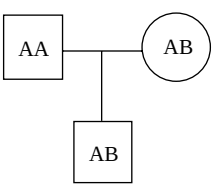

Allele	Father	Mother
Transmitted	A	B
Non transmitted	A	A

Fig. 8 Transmission/Disequilibrium Test (TDT).

means that it is possible to construct an SNP map across regions of interest (e.g. a region implicated by linkage studies) and narrow it down using case control studies to detect LD. Databases can then be searched for genes within the narrowed region of interest and SNPs selected within such genes to explore them further in case control comparisons. This process, which in effect is a convergence of the positional cloning and candidate gene approaches, is now referred to as the *positional candidate approach* and has been successful, as we will discuss below, in implicating several genes in the etiology of schizophrenia.

4 Applications

Quantitative approaches have now been used extensively in a wide range of behavioral traits and disorders and increasingly, molecular genetic approaches are being applied. Here, we have selected three major examples, one a normal trait and the other two, disorders (or rather groups of disorders). These are general cognitive ability or intelligence, mood disorders and schizophrenia.

4.1 General Cognitive Ability

Intelligence in its various guises is one of the most extensively studied aspects of human behavior. There is now a broad consensus among psychologists that although various types of intelligence can be described and tests devised to measure them, performance on such tests is strongly correlated with a common factor called *general cognitive ability* or "g." Thus, it is has been shown that tasks such as copying a design with wooden blocks (visuo–spatial ability), vocabulary tests (verbal ability), remembering figures (memory), and comparing lists of objects (processing speed) all correlate with "g." Commonly used tests provide measures of general intelligence quotient (IQ) as well as specific abilities including the Wechsler Adult Intelligence Scale (WAIS) and the Wechsler Intelligence Scale for Children (WISC).

Quantitative genetics and cognitive ability

Family studies of IQ and "g" demonstrate moderate to high correlations of around 0.4 to 0.5 for first-degree relatives. More distant relatives such as cousins typically show correlations in the order of 0.15. Although such a pattern suggests the involvement of genes, it is not on its own sufficient proof. In fact, nearly all of the twin studies performed in childhood suggest that the familial correlation in IQ is partly the result of genes and partly the result of shared environment. For example, identical twins reared together show very high correlations for g, on average around 0.85. This is somewhat lower, at around 0.7, in identical twins reared apart. Similarly, DZ twins reared together show an average correlation for g of about 0.6, but this falls to an average 0.3 in DZ twins reared apart. It can be shown using the sort of structural equation models that have been discussed earlier that this pattern of correlations corresponds to a heritability of around 50% with about 25% of the variance explained by shared environment and the remainder explained by nonshared environment.

Adoption studies in childhood are almost completely consistent with the twin results in that adoptive siblings show a correlation for g of around 0.25 compared with a correlation of 0.5 in biological siblings

reared together. The pattern, however, is rather different in adults and, surprisingly, it is only comparatively recently that much attention has been focused on this. It turns out that biologically unrelated individuals who are raised as siblings have an average correlation of zero for general intelligence when measured as adults. Twin studies similarly show that the heritability of IQ steadily increases in adolescents compared with children and in adults compared with adolescents. At the same time, the effect of shared environment decreases. Very recent studies of twins in late life show that the nondemented elderly have the highest heritability for general intelligence of all age groups at around 60% with no evidence of a shared environmental effect.

In summary, quantitative genetic studies have recently converged on a consensus that general intelligence is at least moderately heritable and perhaps highly heritable, particularly in late adulthood. Much interest has now begun to shift towards the goal of actually being able to locate and identify genes involved in cognitive ability.

Molecular studies: searching for "IQ-QTLs"

It is broadly accepted from the pattern of inheritance of general cognitive ability that it is likely to be a highly polygenic trait. That is, there is likely to be a contribution from many genes each of which on its own has only a very small effect. If this assumption is correct, then linkage studies will be very difficult and therefore almost all the attention has focused on allelic association. A pioneering study was by Plomin and colleagues who in 1995 published results on 100 polymorphisms selected on the basis that they were at or near genes that were expressed in the brain. A comparison of allele distributions was made in three groups of subjects showing high, middle, and low scores on IQ tests and positive results were followed up on a replication sample. Although there were some promising preliminary findings including a locus on chromosome 6p and an association with a mitochondrial polymorphism, these failed to replicate on an independent sample.

The same group then embarked on an ambitious project attempting to search the whole genome, chromosome by chromosome, for linkage disequilibrium using a highly polymorphic set of microsatellite markers. Plomin and colleagues also adopted an innovative approach using DNA pooling in which the DNA from very high IQ subjects was combined in one pool and this was compared with the DNA from a pool of subjects with average IQ. The aim here was to achieve a rapid way of testing many markers. Thus, the initial study searching for genes having an effect on IQ used around 2000 markers on 200 unrelated subjects with high IQ and 100 with average IQ. This would have entailed 600 000 individual genotypings, but with DNA pooling of the high-IQ and average-IQ groups, the initial genotyping was reduced to a more manageable 4000. Positive results were then subsequently followed up by individual genotyping. To date, there are no definite positive associations that have been discovered with general cognitive ability that have replicated on all samples. However, there have been some intriguing, suggestive findings, for example, of an association with a variant in the insulin-like growth factor receptor 2 gene (*IGF2R*). While these results may be worth pursuing further, it should be emphasized that *IGF2R* is not, as some popular press reports suggested, the "gene for" IQ. In fact, the *IGF2R* association reported by the group of Plomin et al. appears at best to account

for about 2% of the variance in general cognitive ability.

4.2 Schizophrenia

Quantitative genetics of schizophrenia As already noted, there is consistent evidence that schizophrenia is more common in the relatives of schizophrenics than in the general population. The lifetime risk of schizophrenia is surprisingly constant in all of those countries for which figures are available at around 1% of the population. This compares with around 10% in the siblings and offspring of schizophrenics. It is at first sight perhaps curious that the risk in parents of schizophrenics is much lower at around 5%, but this is probably because of a "censoring" effect. Schizophrenia is a disorder with a markedly adverse effect on the ability of a sufferer to find a mate and procreate. This means that the parents of schizophrenics, who are themselves schizophrenic, will almost all have their onset *after* having had their children. Consequently, the period of risk for developing schizophrenia is reduced and individuals who have both had children and developed schizophrenia are likely to consist of comparatively late-onset cases.

Familiality does not, of course, on its own prove that schizophrenia has a partly genetic basis. Indeed, schizophrenia was at one time among the hotly contested topics in the "nature versus nurture" debate. This came to its height in the 1960s when "antipsychiatrists," such as the followers of R. D. Laing, criticized the very concept of schizophrenia, taking the somewhat romantic view that the manifestations of the disorder, hallucinations, delusions, and a distorted view of reality, could be seen as a "sane response to an insane world." Subsequently, adoption and twin studies have convinced all but the most obdurate skeptics that schizophrenia cannot be explained purely as a reaction to an adverse environment. As Seymour Kety, one of the leaders of an influential group of Danish–American adoption studies commented, "If schizophrenia is a myth then it is a myth with a strong genetic basis." We have already reviewed the adoption study data (Table 2). The twin data are also completely compatible with important genetic influences, and the five most recent twin studies (four from Europe and one from Japan) show concordance rates in MZ twins of 41 to 65% compared with 0 to 28% in DZ twins. A meta-analysis of these data suggested a heritability in excess of 80%.

Molecular genetics: finding the susceptibility genes Given that there is consistent evidence of a strong genetic effect, it might be expected that the business of locating and identifying genes should prove to be relatively straightforward. However, schizophrenia appears to be a polygenic disorder where single-gene forms are rare or perhaps nonexistent and where each of the contributory genes, on its own, has a very small effect. Indeed, attempts to map genes involved in schizophrenia date back to almost 50 years and there were several attempts to discover schizophrenia-associated genes in the "premolecular" era using what are now called classical genetic markers such as blood groups and human leukocyte antigen (HLA) types. More recently, attention has turned to polymorphisms in functional candidate genes and, as we have already mentioned, there is evidence of a small but consistent effect resulting from variants in the serotonin *5HT2A* receptor gene and dopamine

DRD3 receptor gene. There has also been renewed interest in another gene involved in dopamine neurotransmission, catechol-O-methyltransferase (*COMT*). *COMT* inactivates catecholamines, including dopamine, by methylating their m-hydroxy group. A functional *COMT* gene polymorphism has been described that influences enzyme activities and the high-activity allele (*val*-108) has been found to be associated with schizophrenia in some studies. This is of particular interest because schizophrenics, compared with controls, on average show reduced metabolism in the prefrontal lobe of the brain and do less well on cognitive tests that are designed to test prefrontal function. Otherwise healthy subjects who have the high-activity allele fair less well in such tests and have less efficient prefrontal physiology than those who have only the low-activity (*met*-108) allele. Unfortunately, the *val/met* polymorphism association with schizophrenia has not proved consistent across studies. The largest single study to focus on *COMT* suggests that there is a weak but highly significant association with schizophrenia, but this results from a set of variants in a haplotype in a different part of the gene.

Perhaps of even greater interest, several new positional candidates, genes that are involved in neurochemical pathways and/or neurodevelopmental processes that are of relevance to schizophrenia, have been implicated in linkage regions. One of these, neuregulin-1 (*NRG1*), maps to chromosome 8p, a region of the genome implicated in schizophrenia in a number of studies including one based on Icelandic families. The Icelandic researchers performed an association analysis across the region that gave significant evidence of an effect of a variant in *NRG1*. The *NRG1* association with schizophrenia was subsequently replicated in Scottish patients and controls. Mice in which one copy of *NRG1* or of the gene for its receptor *ErbB4* have been "knocked out" show some behaviors that occur in schizophrenic patients, including a response to simple stimuli called prepulse inhibition. These changes can be partially reversed in the mice with the antipsychotic drug Clozapine. The mutant mice also have fewer functional receptors for the excitatory amino acid glutamine of the NMDA type than do wild-type mice.

A completely novel gene encoding for a protein called G72 also appears to have a role in NMDA receptor function. The *G72* gene maps to chromosome 13q to a region that again has been implicated in several linkage studies. In a study involving Canadian patients, which was then replicated in cases from Russia, the region was again narrowed down by linkage disequilibrium mapping. Potential genes in the narrowed region were then identified by searching computer databases. After experimental annotation using RT-PCR, two genes were identified, one of which, *G72*, was found to be expressed in the human brain. It is also of interest that given the quintessentially human and language-based character of schizophrenic symptoms, *G72* appears to be found only in primates. Furthermore, the longest transcript of the gene in the chimpanzee brain is only about half that of its human counterpart, perhaps suggesting rapid evolutionary change. Using a way of identifying interacting proteins called the yeast 2 hybrid method, an interaction with the enzyme D-amino acid oxidase (DAAO) was found. DAAO is also expressed in the brain and oxidizes D-serine, which in turn has an activating effect at an NMDA receptor.

A third linkage region, this time on chromosome 6p originally identified by

a large study of Irish families, contains a gene called dystrobrevin-binding protein-1 or dysbindin, which again contains variants that appear to be more common in schizophrenics than in controls. Dysbindin is widely expressed in mouse brain, and by binding to dystrobrevin, appears to play a role in synapse formation and maintenance and possibly in NMDA receptor clustering.

A region on chromosome 22q has been implicated not just by linkage studies but also by studies of individuals in whom there is a micro deletion of the region. Depending on the size of the deletion, those who carry it present with an overlapping set of syndromes of which the central one is called velo-cardio-facial syndrome (VCFS). About a quarter to a third of individuals who have VCFS also show schizophrenic-like symptoms. Interestingly, the VCFS region contains the *COMT* gene, but some studies of schizophrenia, particularly the early-onset type, have suggested an association with variants in another gene, proline dehydrogenase (*PRODH2*). There is again some evidence that *PRODH2* is involved in the modulation of glutamine transmission in the brain at NMDA receptors.

The fact that all of these genes have an effect on glutamate transmission seems unlikely to be coincidental. Glutamate receptors are of two broad types – those such as the NMDA receptor that are ion channels ("ionotropic") and those that are coupled to a class of intracellular messengers called g-proteins ("metabotropic"). G-protein signaling is in turn influenced by a class of proteins called regulators of g-protein (RGS). One such protein, RGS4, has been shown to have altered expression in schizophrenic brains compared with controls. Subsequently, a haplotype, a set of variants in the *RGS4* gene, has been found to show differences between patients and controls. *RGS4* is also relevant to other g-protein coupled receptors including dopamine receptors belonging to the D2 family.

4.3 Mood Disorders

The major mood disorders belong to two main groups, bipolar affective disorders and unipolar depression. Bipolar disorder is so called because there are both, episodes of mania (characterized by elation, over-activity, disinhibition, and often reckless behavior) and episodes of depression (characterized by low mood often accompanied by sleep disturbance, appetite and weight loss, and suicidal thoughts). Unipolar disorder, also called depressive disorder, consists of episodes of depression alone. Depressive disorder is very common. Illness that is sufficiently severe to warrant referral to a psychiatrist has a lifetime risk in the general population, of around 6 to 8%. Less severe forms are even more common. Affected women outnumber men by about 2 : 1. By contrast, there is no sex difference in the frequency of bipolar disorder, which is a less common condition affecting around 0.5 to 1% of the population per lifetime.

Quantitative genetics of mood disorder

Both forms of disorder show a marked tendency to aggregate in families. The majority of studies show a stronger familial effect with bipolar disorder, although some more recent studies, particularly those focusing on severe depression, find that unipolar disorder is almost as familial. A major and consistent difference in all studies is that index cases with unipolar disorder have an excess only of unipolar depression among their relatives, whereas

index cases with bipolar disorder have an excess of relatives both with bipolar disorder and with unipolar depression.

In contrast with schizophrenia, the adoption data on affective disorders are fairly sparse, but the most carefully conducted studies suggest that the familial clustering of these disorders largely results from genetic effects. This is supported by twin research. In recent studies of unipolar depression, there is consistently higher concordance in monozygotic than in dizygotic twins. The effect appears to be stronger in more severe or recurrent depression and model fitting studies aimed at estimating heritability suggest that this is around 70% in studies where the index cases have been ascertained via hospitals or clinics. If the depressed index cases are ascertained via population surveys, perhaps consisting of milder disorder, the heritability is typically around 30 to 40%.

The twin study model fitting results also suggest that whether the sample is obtained clinically or is population-based, the environmental effects are all of the nonshared kind. There have been several studies attempting to examine the interplay between genetic and environmental risk factors in unipolar depression. There is a great deal of evidence supporting the common sense hypothesis that onsets of depression are frequently associated with unpleasant events. However, there is no evidence that depression that is associated with adversity is any less familial than depression that arises "out of the blue." There is also considerable evidence from studies of families and twins that life events are not completely randomly distributed in the population. Some individuals are more prone to life events than others and people who score high on personality traits designed to measure dimensions such as extraversion or novelty-seeking seem to have more eventful lives than those who score low on such trait measures. These personality traits have been found to show both positive correlations in siblings and, in twin studies, to be heritable. Therefore, it is perhaps not too surprising that at least some forms of life events also appear to be familial. For some events, the familial effect is purely environmental. For example, the death of a parent affects both members of a pair of siblings or twins. Nevertheless, several studies indicate that some events reflect the influence of genetic factors rather than shared environment.

Molecular genetics: searching for "blue genes" Bipolar disorder has been the subject of quite a number of linkage studies including whole genome scans. The data are less extensive than those for schizophrenia and recent meta-analyses would suggest that they are much less clear-cut. Despite this, regions of interest have been identified on chromosomes 13q and 22q, which overlap those that have been implicated in schizophrenia. Although the conventional view is that schizophrenia and bipolar disorder are largely different genetically, families have frequently been described containing sufferers from both conditions and occasionally identical twins or triplets have been described who have nonidentical psychoses. A recent twin study using multivariate structural equation modeling looked at symptom patterns rather than conventional diagnoses, that is, subjects were classified as having a bipolar syndrome, schizophrenia, or schizoaffective disorder, an intermediate form. Instead of assigning an exclusive diagnosis, subjects were classified according to the symptom pattern as having one, two, or all three disorders. The analyses suggested a large

overlap in the genes contributing to all three disorders as well as the existence of specific genes contributing to the liability of either schizophrenia or bipolar disorder. The findings are therefore consistent with linkage study findings pointing to an overlap in the genomic regions that contain schizophrenic and bipolar genes. As we write, no studies have yet been published on bipolar disorder and *G72* or *DAAO*, the genes on chromosome 13q discussed earlier in relation to schizophrenia. Nor have there been any published studies on PR0DH2 and bipolar disorder on the chromosome 22q genes implicated in schizophrenia. However, *COMT* variants have been implicated in bipolar disorder, particularly inpatients who show rapid cycling, that is, those having four or more episodes per year.

Molecular genetic studies are much less advanced in unipolar disorder, although several large-scale linkage and association studies are currently under way. To date, the most provocative results have been based on the serotonin transporter gene. One of the ways in which serotonin is cleared from the synapse is by reuptake into the presynaptic neuron. This is the site of action of the antidepressant drugs called selective serotonin reuptake inhibitors of which fluoxetine (Prozac) is probably the best-known example. The serotonin transporter gene contains a common variable number tandem repeat (VNTR) polymorphism within one of its introns, as well as a common variant in the promoter region that has been shown to effect levels of gene expression. Both variants have been implicated in association studies of mood disorders and the promoter polymorphism may also be associated with the personality trait of neuroticism, which may in turn be associated with vulnerability to depression.

5 Future Directions

Behavioral genetics is now moving into an era of behavioral genomics, offering an improved understanding of the neurobiology of disease. Not only is the complete human genome sequence known but there are also complete genome sequences or draft sequences available on over 60 species. Comparing genomes across species and noting differences in those genes that have been conserved will show the major variations between species and offer an understanding of the basis for neurobiological and hence behavioral evolution. Knowing the structure and function of all human genes has been compared with the discovery of a "periodic table of life." It paves the way for a series of paradigm shifts where the emphasis moves from the structure of the genome to functional genomics and to proteomics, the study of proteins, at a functional level. A striking example of how one complex aspect of human behavior, circadian rhythms, can be dissected and its basis understood at a molecular level is given by the recent discovery of new "clock genes," simply by the analysis of the draft sequence of the human genome that was discovered when searching for genes with a high similarity to known clock genes originally discovered in mice or fruit flies.

Even before the publication of the draft sequence of the human genome, discoveries arrived at by positional cloning in Alzheimer's disease (AD) have begun to provide important new understandings of the neuropathological mechanism. Although Alzheimer's disease has a characteristic neuropathology that was described at the beginning of the twentieth century, the problem of discovering its pathogenesis did not appear to have a tractable solution until the discovery of the role

of mutations in the amyloid precursor protein and presenilin. Interestingly, although presenilin1 on chromosome 14 was identified by positional cloning, presenilin2 was discovered essentially by its homology with presenilin1 and subsequently was demonstrated to have an etiological role by linkage analysis. As with the discovery of new clock genes, we can expect the discovery of families of genes involved in abnormal and normal behaviors to become more commonplace.

Identification of genes and gene products involved in psychiatric disorders will have important implications for pharmacotherapy in two ways. The first of these is in directing drug discovery. At present, the drugs used in the treatment of psychiatric or other central nervous system diseases have their actions at a limited number of target sites that include cell surface receptors, nuclear receptors, ion channels, and enzymes. It is likely that detecting genes involved in the pathogenesis of psychiatric disorders will identify new targets, some of which fall within these categories and some of which have not thus far been included as target sites for drug action. In addition to targeting of treatments, advances in genomics will allow the tailoring of treatments. There is already some evidence that the response to a typical antipsychotic such as Clozapine is influenced by an individual genotype at the serotonin 5HT2A receptor. It is also well established that the rate at which most psychoactive drugs are metabolized is influenced by genetic factors, in particular, genes that code for a family of enzymes called the cytochrome P450 system. This may be relevant to the development of side effects as well as to treatment response. This whole area of pharmacogenomics is still comparatively novel. It holds enormous promise for individual tailoring of treatments that will be a major advance over the current trial and error approach that the clinician is now saddled with.

There has been a degree of public concern that the current pace of advance will tend to "geneticize" common diseases and encourage deterministic attitudes. In particular, worries have been expressed that insurance companies may wish to force DNA testing on individuals thought to be at high genetic risk of disorder. While prediction with a high degree of accuracy is already possible for rare early-onset dementia such as Huntington's disease or the single-gene forms of AD, this is not possible for complex disorders. For example, the apolipoprotein €4 allele, despite the confirmed association with risk of late-onset AD in the general population, is of limited value as a predictor at an individual level. The situation is likely to prove even more complicated with disorders such as schizophrenia. At best, DNA-based tests may be used to modify the predicted risk in individuals who are already at high risk because of having a schizophrenic close relative. However, it is unlikely that risk prediction will ever be better than about 50% accurate, since genetically identical individuals, MZ twins, are discordant for schizophrenia 50% of the time. This means that DNA-based population screening for complex psychiatric disorders (including AD of late onset) will never become a reality, but the option of offering screening to high-risk individuals probably will, with the caveat that such testing will refine risk prediction but not give clear yes/no answers. Given that most common diseases, and not just psychiatric ones, will depend upon the combined action of multiple common gene variations together with environmental risk factors, batteries of genetic tests will be of limited usefulness to insurance companies and are not likely to be widely employed.

Finally, the other potentially worrying aspect of psychiatric disorders becoming "geneticized" is, it has been suggested, an increase in stigma. Of course, just the opposite could be the case, so that increasing knowledge about the causation of disorders may serve to demystify them and therefore make them, in the public eyes, something that is more tangible and acceptable. Part of postgenomic psychiatry's impact on disorders such as schizophrenia, bipolar affective disorder, and depression might therefore be to legitimize them as "real" diseases rather than, as is all too often the case, their being seen as phenomena that result from personal failing or weakness. AD might be taken as a good example of how public perceptions clearly have changed. "Becoming senile" in old age was once seen by many as somehow morally reprehensible, whereas it is now acceptable for the families of a past President of the United States or a famous novelist such as Iris Murdoch to "come out" and admit that they suffer from AD. In the authors' present view, this could turn out to be a general effect. Therefore, rather than an increasing stigma, it is quite possible that the ultimate effect of genetic research on the public image of psychiatric diseases will be wholly positive.

See also Brain Development; Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Falconer, D.S., Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*, 4th edition, Longman, Harlow.

Faraone, S.V., Tsuang, M.T.T., Tsuang, D.W. (2002) *Genetics of Mental Disorders: What Practitioners and Students Need to Know*, Guilford Press, New York.

Mackintosh, N.J. (1998) *IQ and Human Intelligence*, Oxford University Press, Oxford.

McGuffin, P., Owen, M.J., O'Donovan, M.C., Thapar, A., Gottesman, I.I. (1994) *Seminars in Psychiatric Genetics*, Gaskell, London.

McGuffin, P., Owen, M.J., Gottesman, I.I. (2002) *Psychiatric Genetics and Genomics*, Oxford University Press, Oxford.

Neale, M.C., Boker, S.M., Xie, G., Maes, H.H. (1999) *Mx: Statistical Modelling*, 5th edition, Box 126 MCV, Department of Psychiatry, Richmond, VA.

Neale, M.C., Cardon, L.R. (1992) *Methodology for Genetic Studies of Twins and Families*, Kluwer Academic Publishers, Dordrecht.

Plomin, R, Defries, J.C., McClearn, G.E., McGuffin, P. (2000) *Behavioral Genetics*, W.H. Freeman, New York.

Plomin, R., Owen, M.J., McGuffin, P. (1994) The genetic basis of complex human behaviors, *Science* **264**, 1733–1739.

Sham, P.C. (1998) *Statistics in Human Genetics*, Edward Arnold, London.

Primary Literature

Allison, D.B., Heo, M., Kaplan, N., Martin, E.R. (1999) Sibling-based tests of linkage and association for quantitative traits, *Am. J. Hum. Genet.* **64**, 1754–1763.

Allison, D.B., Neale, M.C., Zannolli, R., Schork, N.J., Amos, C.I., Blangero, J. (1999) Testing the robustness of the likelihood-ratio test in a variance- component quantitative-trait loci-mapping procedure, *Am. J. Hum. Genet.* **65**, 531–544.

Almasy, L., Blangero, J. (1998) Multipoint quantitative-trait linkage analysis in general pedigrees, *Am. J. Hum. Genet.* **62**, 1198–1211.

Amos, C.I. (1994) Robust variance-components approach for assessing genetic linkage in pedigrees, *Am. J. Hum. Genet.* **54**, 535–543.

Bellivier, F., Henry, C., Szoke, A., Schurhoff, F., Nosten-Bertrand, M., Feingold, J., Launay, J.M., Leboyer, M., Laplanche, J.L. (1998) Serotonin transporter gene polymorphisms in patients with unipolar or bipolar depression, *Neurosci. Lett.* **255**, 143–146.

Bellivier, F., Leboyer, M., Courtet, P., Buresi, C., Beaufils, B., Samolyk, D., Allilaire, J.F., Feingold, J., Mallet, J., Malafosse, A. (1998) Association between the tryptophan hydroxylase

gene and manic- depressive illness, *Arch. Gen. Psychiatry* **55**, 33–37.

Bouchard, T.J. Jr., Lykken, D.T., McGue, M., Segal, N.L., Tellegen, A. (1990) Sources of human psychological differences: the Minnesota study of twins reared apart, *Science* **250**, 223–228.

Cadoret, R.J., O'Gorman, T.W., Heywood, E., Troughton, E. (1985) Genetic and environmental factors in major depression, *J. Affect Disord.* **9**, 155–164.

Carter, C.O. (1969) Genetics of common disorders, *Br. Med. Bull.* **25**, 52–57.

Chorney, M.J., Chorney, K., Seese, N., Owen, M., Daniels, J., McGuffin, P., Thompson, L.A., Detterman, D.K., Benbow, C.P., Lubinski, D., Eley, T., Plomin, R. (1998) A quantitative trait locus associated with cognitive ability in children, *Psychol. Sci.* **9**, 159–166.

Curtis, D. (1997) Use of siblings as controls in case-control association studies, *Ann. Hum. Genet.* **61**, 319–333.

Daniels, J., Holmans, P., Williams, N., Turic, D., McGuffin, P., Plomin, R., Owen, M.J. (1998) A simple method for analyzing microsatellite allele image patterns generated from DNA pools and its application to allelic association studies, *Am. J. Hum. Genet.* **62**, 1189–1197.

DeFries, J.C., Fulker, D.W. (1985) Multiple regression analysis of twin data, *Behav. Genet.* **15**, 467–473.

DeFries, J.C., Fulker, D.W., LaBuda, M.C. (1987) Evidence for a genetic aetiology in reading disability of twins, *Nature* **329**, 537–539.

Devlin, B., Roeder, K. (1999) Genomic control for association studies, *Biometrics* **55**, 997–1004.

Dolan, C.V., Boomsma, D.I., Neale, M.C. (1999) A simulation study of the effects of assignment of prior identity-by-descent probabilities to unselected sib pairs, in covariance-structure modeling of a quantitative-trait locus, *Am. J. Hum. Genet.* **64**, 268–280.

Eaves, L.J., Neale, M.C., Maes, H. (1996) Multivariate multipoint linkage analysis of quantitative trait loci, *Behav. Genet.* **26**, 519–525.

Eaves, L.J., Silberg, J.L., Meyer, J.M., Maes, H.H., Simonoff, E., Pickles, A., Rutter, M., Neale, M.C., Reynolds, C.A., Erikson, M.T., Heath, A.C., Loeber, R., Truett, K.R., Hewitt, J.K. (1997) Genetics and developmental psychopathology: 2. The main effects of genes and environment on behavioral problems in the Virginia twin study of adolescent behavioral development, *J. Child Psychol. Psychiatry* **38**, 965–980.

Egan, M.F., Goldberg, T.E., Kolachana, B.S., Callicott, J.H., Mazzanti, C.M., Straub, R.E., Goldman, D., Weinberger, D.R. (2001) Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6917–6922.

Elston, R.C., Stewart, J. (1971) A general model for the genetic analysis of pedigree data, *Hum. Hered.* **21**, 523–542.

Falconer, D.S. (1965) The inheritance of liability to certain diseases, estimated from the incidence among relatives, *Ann. Hum. Genet.* **29**, 51–76.

Fisher, R.A. (1918) the correlation between relatives on the supposition of Mendelian Inheritance, *Trans. R. Soc. Edinburgh* **52**, 399–433.

Fisher, P.J., Turic, D., Williams, N.M., McGuffin, P., Asherson, P., Ball, D., Craig, I., Eley, T., Hill, L., Chorney, K., Chorney, M.J., Benbow, C.P., Lubinski, D., Plomin, R., Owen, M.J. (1999) DNA pooling identifies QTLs on chromosome 4 for general cognitive ability in children, *Hum. Mol. Genet.* **8**, 915–922.

Foley, D.L., Neale, M.C., Kendler, K.S. (2001) Genetic and environmental risk factors for depression assessed by subject-rated symptom check list versus structured clinical interview, *Psychol. Med.* **31**, 1413–1423.

Fulker, D.W., Cherny, S.S. (1996) An improved multipoint sib-pair analysis of quantitative traits, *Behav. Genet.* **26**, 527–532.

Fulker, D.W., Cherny, S.S., Sham, P.C., Hewitt, J.K. (1999) Combined linkage and association sib-pair analysis for quantitative traits, *Am. J. Hum. Genet.* **64**, 259–267.

Gottesman, I.I. (1963) Genetic aspects of intelligent behavior, in: Ellis, N. (Ed.) *The Handbook of Mental Deficiency: Psychological Theory and Research*, McGraw-Hill, New York, pp. 346–357.

Gottesman, I.I. (1991) *Schizophrenia Genesis: The Origins of Madness*, Freeman, New York.

Gottesman, I.I., Shields, J. (1967) A polygenic theory of schizophrenia, *Proc. Natl. Acad. Sci. U.S.A.* **58**, 199–205.

Greenberg, D.A., Abreu, P., Hodge, S.E. (1998) The power to detect linkage in complex disease by means of simple LOD-score analyses, *Am. J. Hum. Genet.* **63**, 870–879.

Henderson, N.D. (1970) Genetic influences on the behavior of mice can be obscured by laboratory rearing, *J. Comp. Physiol. Psychol.* **72**, 505–511.

Heston, L.L. (1966) Psychiatric disorders in foster home reared children of schizophrenic mothers, *Br. J. Psychiatry* **112**, 819–825.

Ingraham, L.J., Kety, S.S. (2000) Adoption studies of schizophrenia, *Am. J. Med. Genet.* **97**, 18–22.

Jacquet, H., Raux, G., Thibaut, F., Hecketsweiler, B., Houy, E., Demilly, C., Haouzir, S., Allio, G., Fouldrin, G., Drouin, V., Bou, J., Petit, M., Campion, D., Frebourg, T. (2002) PRODH mutations and hyperprolinemia in a subset of schizophrenic patients, *Hum. Mol. Genet.* **11**, 2243–2249.

Kendler, K.S., Gardner, C.O. Jr. (1998) Twin studies of adult psychiatric and substance dependence disorders: are they biased by differences in the environmental experiences of monozygotic and dizygotic twins in childhood and adolescence?, *Psychol. Med.* **28**, 625–633.

Kendler, K.S., Gruenberg, A.M. (1984) An independent analysis of the Danish adoption study of schizophrenia. VI. The relationship between psychiatric disorders as defined by DSM- III in the relatives and adoptees, *Arch. Gen. Psychiatry* **41**, 555–564.

Kendler, K.S., Gruenberg, A.M., Kinney, D.K. (1994) Independent diagnoses of adoptees and relatives as defined by DSM-III in the provincial and national samples of the Danish adoption study of schizophrenia, *Arch. Gen. Psychiatry* **51**, 456–468.

Kendler, K.S., Heath, A.C., Martin, N.G., Eaves, L.J. (1987) Symptoms of anxiety and symptoms of depression. Same genes, different environments?, *Arch. Gen. Psychiatry* **44**, 451–457.

Kendler, K.S., Neale, M.C., Kessler, R.C., Heath, A.C., Eaves, L.J. (1992) A population-based twin study of major depression in women. The impact of varying definitions of illness, *Arch. Gen. Psychiatry* **49**, 257–266.

Kety, S.S. (1983) Mental illness in the biological and adoptive relatives of schizophrenic adoptees: findings relevant to genetic and environmental factors in etiology, *Am. J. Psychiatry* **140**, 720–727.

Kety, S.S., Wender, P.H., Jacobsen, B., Ingraham, L.J., Jansson, L., Faber, B., Kinney, D.K. (1994) Mental illness in the biological and adoptive relatives of schizophrenic adoptees. Replication of the Copenhagen study in the rest of Denmark, *Arch. Gen. Psychiatry* **51**, 442–455.

Knapp, M. (1999) The transmission/disequilibrium test and parental-genotype reconstruction: the reconstruction-combined transmission/ disequilibrium test, *Am. J. Hum. Genet.* **64**, 861–870.

Kruglyak, L., Daly, M.J., Reeve-Daly, M.P., Lander, E.S. (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach, *Am. J. Hum. Genet.* **58**, 1347–1363.

Lander, E.S., Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results, *Nat. Genet.* **11**, 241–247.

Martin, N., Boomsma, D., Machin, G. (1997) A twin-pronged attack on complex traits, *Nat. Genet.* **17**, 387–392.

McGrath, M.J., Campbell, K.M., Veldman, M.B., Burton, F.H. (1999) Anxiety in a transgenic mouse model of cortical-limbic neuropotentiated compulsive behavior, *Behav. Pharmacol.* **10**, 435–443.

McGuffin, P., Asherson, P., Owen, M., Farmer, A. (1994) The strength of the genetic effect. Is there room for an environmental influence in the aetiology of schizophrenia?, *Br. J. Psychiatry* **164**, 593–599.

McGuffin, P., Katz, R., Watkins, S., Rutherford, J. (1996) A hospital-based twin register of the heritability of DSM-IV unipolar depression, *Arch. Gen. Psychiatry* **53**, 129–136.

McGuffin, P., Martin, N. (1999) Science, medicine, and the future. Behaviour and genes, *BMJ* **319**, 37–40.

McGuffin, P., Tandon, K., Corsico, A. (2003) Linkage and association studies of schizophrenia, *Curr. Psychiatry Rep.* **5**, 121–127.

Mednick, S.A., Gabrielli, W.F. Jr., Hutchings, B. (1984) Genetic influences in criminal convictions: evidence from an adoption cohort, *Science* **224**, 891–894.

Mendlewicz, J., Rainer, J.D. (1977) Adoption study supporting genetic transmission in manic – depressive illness, *Nature* **268**, 327–329.

Moises, H.W., Zoega, T., Gottesman, I.I. (2002) The glial growth factors deficiency and synaptic destabilization hypothesis of schizophrenia, *BMC Psychiatry* **2**, 8.

Morton, N.E. (1955) Sequential test of the detection of linkage, *Am. J. Hum. Genet.* **7**, 277–318.

Morton, N.E., MacLean, C.J. (1974) Analysis of family resemblance. 3. Complex segregation of quantitative traits, *Am. J. Hum. Genet.* **26**, 489–503.

Neale, M.C. (1997) *Mx: Statistical Modeling*, Department of Psychiatry, Medical College of Virginia, Richmond, VA.

Ohara, K., Nagai, M., Tsukamoto, T., Tani, K., Suzuki, Y., Ohara, K. (1998) Functional polymorphism in the serotonin transporter promoter at the SLC6A4 locus and mood disorders, *Biol. Psychiatry* **44**, 550–554.

Ott, J. (1974) Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies, *Am. J. Hum. Genet.* **26**, 588–597.

Pedersen, N.L., Plomin, R., Nesselroade, J.R., McClearn, G.E. (1992) A quantitative genetic analysis of cognitive abilities during the second half of the life span, *Psychol. Sci.* **3**, 346–353.

Penrose, L.S. (1935) The detection of autosomal linkage in data which consist of pairs of brothers and sisters of unspecified parentage, *Ann. Eugenics* **6**, 133–138.

Plomin, R., Fulker, D.W., Corley, R., DeFries, J.C. (2001) Nature, nurture, and cognitive development from 1–16 years: A parent-offspring adoption study, *Psychol. Sci.* **31**, 497–509.

Plomin, R., Hill, L., Craig, I.W., McGuffin, P., Purcell, S., Sham, P., Lubinski, D., Thompson, L.A., Fisher, P.J., Turic, D., Owen, M.J. (2001) A genome-wide scan of 1842 DNA markers for allelic associations with general cognitive ability: a five-stage design using DNA pooling and extreme selected groups, *Behav. Genet.* **31**, 497–509.

Pritchard, J.K., Rosenberg, N.A. (1999) Use of unlinked genetic markers to detect population stratification in association studies, *Am. J. Hum. Genet.* **65**, 220–228.

Pritchard, J.K., Stephens, M., Rosenberg, N.A., Donnelly, P. (2000) Association mapping in structured populations, *Am. J. Hum. Genet.* **67**, 170–181.

Purcell, S., Cherny, S.S., Hewitt, J.K., Sham, P.C. (2001) Optimal sibship selection for genotyping in quantitative trait locus linkage analysis, *Hum. Hered.* **52**, 1–13.

Rao, D.C., Keats, B.J., Morton, N.E., Yee, S., Lew, R. (1978) Variability of human linkage data, *Am. J. Hum. Genet.* **30**, 516–529.

Risch, N. (1990) Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs, *Am. J. Hum. Genet.* **46**, 242–253.

Risch, N., Zhang, H. (1995) Extreme discordant sib pairs for mapping quantitative trait loci in humans, *Science* **268**, 1584–1589.

Rosenthal, D., Wender, P.H., Kety, S.S., Schulsinger, F., Welner, J., Ostergaard, L. (1968) Schizophrenics' offspring reared in adoptive homes, in: Rosenthal, D., Kety, S.S. (Eds) *The Transmission of Schizophrenia*, Pergamon, Oxford, pp. 377–392.

Schork, N.J. (1993) Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations, *Am. J. Hum. Genet.* **53**, 1306–1319.

Sham, P.C., Cherny, S.S., Purcell, S., Hewitt, J.K. (2000) Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data, *Am. J. Hum. Genet.* **66**, 1616–1630.

Sham, P.C., Curtis, D. (1995) An extended transmission/disequilibrium test (TDT) for multi-allele marker loci, *Ann. Hum. Genet.* **59**, 323–336.

Sham, P.C., Lin, M.W., Zhao, J.H., Curtis, D. (2000) Power comparison of parametric and nonparametric linkage tests in small pedigrees, *Am. J. Hum. Genet.* **66**, 1661–1668.

Sham, P.C., Zhao, J.H., Cherny, S.S., Hewitt, J.K. (2000) Variance-components QTL linkage analysis of selected and non-normal samples: conditioning on trait values, *Genet. Epidemiol.* **19**(Suppl. 1), S22–S28.

Shifman, S., Bronstein, M., Sternfeld, M., Pisante-Shalom, A., Lev-Lehman, E., Weizman, A., Reznik, I., Spivak, B., Grisaru, N., Karp, L., Schiffer, R., Kotler, M., Strous, R.D., Swartz-Vanetik, M., Knobler, H.Y., Shinar, E., Beckmann, J.S., Yakir, B., Risch, N., Zak, N.B., Darvasi, A. (2002) A highly significant association between a COMT haplotype and schizophrenia, *Am. J. Hum. Genet.* **71**, 1296–1302.

Spielman, R.S., McGinnis, R.E., Ewens, W.J. (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM), *Am. J. Hum. Genet.* **52**, 506–516.

Stefansson, H., Sigurdsson, E., Steinthorsdottir, V., Bjornsdottir, S., Sigmundsson, T., Ghosh, S., Brynjolfsson, J., Gunnarsdottir, S., Ivarsson, O., Chou, T.T., Hjaltason, O., Birgisdottir, B., Jonsson, H., Gudnadottir, V.G., Gudmundsdottir, E., Bjornsson, A., Ingvarsson, B., Ingason, A., Sigfusson, S., Hardardottir, H., Harvey, R.P., Lai, D., Zhou, M., Brunner, D., Mutel, V., Gonzalo, A., Lemke, G., Sainz, J., Johannesson, G., Andresson, T., Gudbjartsson, D., Manolescu, A., Frigge, M.L., Gurney, M.E., Kong, A., Gulcher, J.R., Petursson, H., Stefansson, K. (2002) Neuregulin 1 and susceptibility to schizophrenia, *Am. J. Hum. Genet.* **71**, 877–892.

Straub, R.E., Jiang, Y., MacLean, C.J., Ma, Y., Webb, B.T., Myakishev, M.V., Harris-Kerr, C., Wormley, B., Sadek, H., Kadambi, B., Cesare, A.J., Gibberman, A., Wang, X., O'Neill, F.A., Walsh, D., Kendler, K.S. (2002) Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia, *Am. J. Hum. Genet.* **71**, 348.

Suarez, B.K., Rice, J., Reich, T. (1978) The generalized sib pair IBD distribution: its use in the detection of linkage, *Ann. Hum. Genet.* **42**, 87–94.

Terwilliger, J.D., Ott, J. (1992) A haplotype-based 'haplotype relative risk' approach to detecting allelic associations, *Hum. Hered.* **42**, 337–346.

Thapar, A., McGuffin, P. (1997) Anxiety and depressive symptoms in childhood–a genetic study of comorbidity, *J. Child Psychol. Psychiatry* **38**, 651–656.

Tienari, P., Wynne, L.C., Moring, J., Laksy, K., Nieminen, P., Sorri, A., Lahti, I., Wahlberg, K.E., Naarala, M., Kurki-Suonio, K., Saarento, O., Koistinen, P., Tarvainen, T., Hakko, H., Miettunen, J. (2000) Finnish adoptive family study: sample selection and adoptee DSM-III-R diagnoses, *Acta Psychiatr. Scand.* **101**, 433–443.

Wender, P.H., Rosenthal, D., Kety, S.S., Schulsinger, F., Welner, J. (1974) Crossfostering. A research strategy for clarifying the role of genetic and experiential factors in the etiology of schizophrenia, *Arch. Gen. Psychiatry* **30**, 121–128.

Whittaker, J.C., Lewis, C.M. (1999) Power comparisons of the transmission/disequilibrium test and sib- transmission/disequilibrium-test statistics, *Am. J. Hum. Genet.* **65**, 578–580.

Zhao, H. (2001) Family-based association studies, *Statistical Methods Med. Res.* **9**, 563–587.

Zhao, J.H., Curtis, D., Sham, P.C. (2000) Model-free analysis and permutation tests for allelic associations, *Hum. Hered.* **50**, 133–139.

Zhao, H., Zhang, S., Merikangas, K.R., Trixler, M., Wildenauer, D.B., Sun, F., Kidd, K.K. (2000) Transmission/disequilibrium tests using multiple tightly linked markers, *Am. J. Hum. Genet.* **67**, 936–946.

Biodiversity (see Genetic Diversity in Microorganisms): *see* Genetic Variation and Molecular Evolution

Biogenesis, Structure and Function of Lysosomes

James Fred Dice
Department of Physiology, Tufts University School of Medicine,
136 Harrison Avenue, Boston, MA, USA

1 Lysosome Structure 620

2 Biogenesis of Mammalian Lysosomes 621
2.1 Proteins in the Lumen 621
2.2 Proteins in the Membrane 623

3 Biogenesis of the Yeast Vacuole 624
3.1 Proteins in the Lumen 624
3.2 Proteins in the Membrane 624
3.3 Vacuolar Inheritance 625

4 The Importance of Phosphoinositides in Lysosomal Biogenesis 625

5 Lysosomal Pathways of Proteolysis 626
5.1 Endocytosis 626
5.2 Exocytosis 627
5.3 Macroautophagy 627
5.4 Microautophagy 628
5.5 Vacuolar Import and Degradation Pathway 628
5.6 Chaperone-mediated Autophagy 629

6 Lysosomal Pathways of RNA Degradation 629

7 Lysosomal Storage Diseases 629

9 Conclusions 630

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Bibliography 630
Books and Reviews 630
Primary Literature 631

Keywords

Lysosome
An acidic organelle containing hydrolytic enzymes.

Vacuole
A large lysosome-like organelle found in yeast and plants.

Mannose 6-Phosphate Receptor
One of two proteins in mammalian cells that recognize lysosomal lumen proteins tagged with mannose 6-phosphate.

Phosphoinositides
Lipids that act as second messengers and regulate vesicular traffic.

Lysosomal Storage Disease
Defects in lysosomal enzyme activity or targeting that result in the lysosomal accumulation of undigested materials.

Abbreviations

ALP	alkaline phosphatase
AP	adaptor protein
Aph	autophagosome
API	aminopeptidase I
Apg	autophagy
ARF	ADP-ribosylation factor
Cma	chaperone-mediated autophagy
COP	coatomer protein
CPY	carboxypeptidase Y
Cvt	cytoplasm-to-vacuole targeting
ER	endoplasmic reticulum
FBPase	fructose-1,6-bisphosphatase
FYVE	Fab1, YGLO23, Vps27, EEA1
GGA	Golgi-localizing, gamma adaptin ear homology domain, ARF-binding
Hsp70	heat shock protein of 70 kDa
Hsc70	the constitutively expressed hsp70

IGF	insulin-like growth factor
LAMP	lysosome-associated membrane protein
LIMP	lysosomal integral membrane protein
M6P	mannose 6-phosphate
M6PR	mannose 6-phosphate receptor
MVB	multivesicular body
NSF	*N*-ethylmaleimide sensitive factor
PI	phosphotidylinositol
PI3P	phosphotidylinositol-3-phosphate
Rab	ras-related GTPase
t-SNARE	target membrane soluble *N*-ethylmaleimide sensitive attachment receptor
V-ATPase	Vacuolar proton-pumping ATPase
Vam3p	a t-SNARE in yeast
Vid	Vacuolar import and degradation
Vps	vacuolar protein sorting
v-SNARE	vesicle membrane soluble *N*-ethylmaleimide sensitive attachment receptor

Lysosomes are organelles that are surrounded by a single membrane and contain many hydrolases that are most active at an acidic pH. The pH within the lysosomal lumen is usually maintained between 5.0 and 5.5 by a multisubunit, proton-pumping ATPase in the lysosomal membrane of mammalian cells and in the vacuolar membrane of yeast. Other organelles having some of the properties of lysosomes include late endosomes and multivesicular bodies. However, these organelles have protein and lipid constituents at least partially distinct from lysosomes.

Enzymes in the lysosomal lumen are targeted to that location by mannose 6-phosphate (M6P) carbohydrate modifications in the case of mammalian cell lysosomes or by linear peptide sequences for the yeast vacuole. Integral membrane proteins within the lysosomal and vacuolar membranes are targeted to that location due to critical tyrosine or dileucine peptide motifs in cytosolic regions of the proteins. These motifs interact with vesicle coat protein complexes that concentrate the proteins into vesicles destined for fusion with lysosomes. Peripheral lysosomal membrane proteins generally interact strongly with one or more integral membrane proteins and/or lipids.

Lysosomes account for 1 to 15% of cell volume and of cell protein in mammalian cells and 30 to 90% of cell volume in fungal and plant cells. This variability depends on cell type and physiological status. The morphological appearance of lysosomes also varies from vesicular to a more complex tubular lattice. Lysosomes are responsible for degrading both extracellular and intracellular proteins as well as other macromolecules.

The role of lysosomes in overall intracellular protein degradation depends on the cell type and nutritional conditions. Lysosomes are responsible for most protein degradation in liver and kidney, and certain lysosomal pathways of proteolysis are activated in response to nutritional deprivation. On the other hand, lysosomes play a

minor role in protein degradation in skeletal muscle and lymphocytes, where the ubiquitin-proteasome proteolytic pathway predominates. Lysosomal pathways of proteolysis operate to some extent in all eukaryotic cells except in mature red blood cells that do not contain organelles, and there are at least six different pathways by which proteins can be delivered to lysosomes for digestion.

1 Lysosome Structure

Lysosomes are usually spherical organelles surrounded by a single membrane. There are 50 to 1,000 lysosomes per mammalian cell, but a single large or multilobed lysosome called the *vacuole* in fungi and plants. Lysosomes can be quite heterogeneous in appearance and are difficult to identify on the basis of purely morphological criteria. In fact, spherical lysosomes in many cell types are able to fuse into a tubular network.

The lumen or matrix of the lysosome contains approximately 50 hydrolases most of which are optimally active at acidic pH. The concentration of some of these hydrolases is estimated to be in the range of 100 to 200 mg mL^{-1}. The lysosomal lumen also contains a molecular chaperone of the heat shock protein of 70-kDa (hsp70) family. The vacuolar lumen of yeast and plant cells also serves as storage sites for amino acids, polyphosphates, and other compounds.

The lysosomal membrane is typically a single bilayer. It contains phospholipids, including sphingomyelin, relatively high concentrations of cholesterol, and lysobisphosphatidic acid. This lipid is also found in late endosomes and multivesicular bodies. The lysosomal membrane also contains many different proteins. The *v*acuolar proton-pumping ATPase (V-ATPase) is composed of five different integral membrane proteins and eight different peripheral membrane proteins. Some of each class are present in multiple copies per functional complex. The V-ATPase subunits are not restricted to lysosomal membranes since they can also be found in transport vesicles, endosomes, Golgi, secretory vesicles, and the plasma membrane.

Lysosomal membranes also contain several transporters. Several different sugar transporters have been described that can transport monosaccharides produced in mammalian lysosomes by digestion of glycoconjugates. There is also a polymannose transporter in the lysosomal membrane that transfers oligosaccharides from the cytosol to the lysosome for digestion. Most amino acids exit lysosomes after proteolysis through general or specific amino acid transporters. A cystine transporter, cystinosin, is an integral lysosomal membrane protein that when mutated leads to cystinosis.

Lysosome-associated membrane protein (LAMP) type 2a (LAMP2a) is a receptor for protein substrates of chaperone-mediated autophagy (Cma; see Sect. 5.6). LAMP2a multimerizes in the lysosomal membrane and may form the protein translocon as well as act as the substrate receptor. Another isoform of LAMP2 formed by alternative splicing, LAMP2b, is required for docking or fusion of autophagic vacuoles and lysosomes, a critical step in macroautophagy (see Sect. 5.3), so

LAMP2a and LAMP2b are required for the operation of two different pathways of lysosomal proteolysis. A role for the third splicing isoform, LAMP2c, is not known.

The LAMP2s, like LAMP1 and lysosomal integral membrane proteins (LIMP)s, are integral proteins with most of the protein facing the lysosomal lumen. This portion of the protein is highly glycosylated, and this glycosylation protects the proteins from rapid digestion by the battery of lysosomal proteases. These glycoproteins may also protect the inner leaflet of the lysosomal membrane from digestion by lumenal lipases.

A recent report of detergent-resistant lipid raft microdomains in the lysosomal membrane suggests an additional level of complexity to lysosomal membrane structure. The physiological significance of such microdomains remains to be determined.

Certain specialized cells also contain lysosome-related organelles that are thought to be made by pathways similar to those for lysosomal biogenesis. They contain not only lysosomal proteins but also additional proteins. Such lysosome-related organelles include melanosomes within melanocytes and retinal pigment epithelium and lytic granules within cytotoxic T-lymphocytes and natural killer cells. The acrosome of sperm is also a lysosome-related organelle.

2 Biogenesis of Mammalian Lysosomes

Both lumenal and membrane lysosomal proteins contain specific targeting signals that are required for their correct localization in lysosomes. Receptors or vesicle coat proteins recognize these targeting signals and play critical roles in trafficking the proteins to lysosomes.

2.1 Proteins in the Lumen

Most lysosomal proteins that are soluble within the lysosomal lumen travel through the protein secretory pathway. That is, they are synthesized by polysomes that are associated with the endoplasmic reticulum (ER). They typically have a cleavable membrane-spanning signal sequence that is required for their translocation into the ER lumen. In the ER, the signal sequence is cleaved, and the resulting processed protein is glycosylated. The protein is packaged into coatomer protein (COP) II-coated vesicles and transferred to the Golgi where it is recognized by a phosphotransferase enzyme that transfers *N*-acetyl-glucosamine-1-phosphate to one or more mannose residues on the lysosomal enzyme. A glucosaminidase in the Golgi removes the glucosamine to generate the M6P.

The phosphotransferase recognizes the sugar group and the features of the polypeptide to identify the protein as belonging to lysosomes. The features of the protein that are recognized are not linear but appear to result from protein folding. Multiple lysine residues in a particular chemical environment and the distance from the carbohydrate group appear to be important determinants of phosphotransferase recognition (Fig. 1).

Transfer of the lysosomal protein through the Golgi may be by COPI-coated vesicles. Alternatively, the Golgi may be dynamic with enzyme distribution determined by forward membrane flow and retrograde protein retrieval, a process termed *cisternal maturation*.

All forms of vesicular transport require specific *v*esicle membrane *s*oluble *N*-ethylmaleimide sensitive *a*ttachment *re*ceptor (v-SNARE) and *t*arget membrane

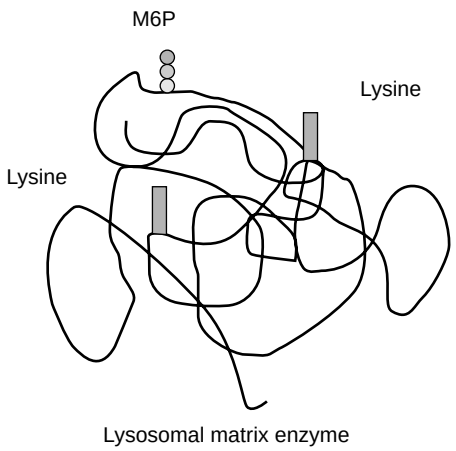

Fig. 1 Proposed recognition elements for the phosphotransferase on a lysosomal matrix protein. Both mannose sugars and lysines within a certain distance from the carbohydrate are recognized by the phosphotransferase enzyme. The lysines may also be in a particular chemical environment created by neighboring amino acid side chains.

*s*oluble *N*-ethylmaleimide sensitive *a*ttachment *re*ceptor (t-SNARE) proteins along with their regulatory proteins, ADP ribosylation factor (ARF) and ras-related GTPases (Rabs). The Rabs are especially important in ensuring specificity of vesicle docking. In order for vesicle fusion to occur, the v-SNARE and t-SNARE first have to be pried apart. This is accomplished by *N*-ethylmaleimide sensitive factor (NSF), a protein complex that is an ATPase.

The M6P is recognized by a M6P receptor (M6PR). There are two types of M6PR in most cells, and both are integral membrane proteins associated with late endosomes, Golgi, and the plasma membrane. One is 215 kDa and binds proteins marked with M6P independently of divalent cations. The other is 46 kDa, and its binding to lysosomal proteins is highly dependent on the presence of divalent cations. These different M6PRs may recognize different lysosomal enzymes and/or be differentially regulated. In the trans-Golgi network, M6PR and lysosomal enzymes are packaged into differently coated vesicles. The M6PR travels to late endosomes but then recycles to Golgi. M6PRs thereby avoid entering lysosomes.

A small amount of the M6PRs reside on the plasma membrane. Any M6P-tagged proteins that may be secreted instead of being delivered to lysosomes can bind to this cell surface M6PR and be delivered to lysosomes by this route. Interestingly, the cell surface cation-independent M6PR also acts as a receptor for insulin-like growth factor II (IGF-II). Binding of IGF-II to the cation-independent M6PR may be a mechanism for delivery of IGF-II to lysosomes. Alternatively, some aspects of IGF-II growth signaling may occur through the cation-independent M6PR.

There are indications that not all lysosomal matrix proteins are targeted to lysosomes by M6P. For example, in I-cell disease, there is a deficiency of the phosphotransferase enzyme, so no M6P can be formed on lysosomal enzymes. As expected, these cells accumulate degradative substrates within lysosomes owing to the deficiency of lysosomal hydrolases. However, lysosomes from some cell types from patients with I-cell disease have almost normal levels of many lysosomal enzymes. Such cells must have M6P-independent lysosomal targeting pathways.

2.2 Proteins in the Membrane

Lysosomal membrane proteins are also made from polysomes associated with the ER. They insert into the ER membrane because of a hydrophobic stretch of 22 to 26 amino acids within the protein sequence. These proteins contain critical amino acids in their cytosolic regions that direct their targeting to lysosomes (Fig. 2). The sequences are remarkably simple and consist of a tyrosine and a bulky hydrophobic amino acid that are present within the last four amino acids of the carboxy terminus. Alternatively, a leucine, typically in the form of a dileucine motif, can also target membrane proteins to lysosomes.

Heterotetrameric adaptor protein (AP) complexes, AP-1, AP-2, AP-3, and AP-4, will all bind tyrosine- or dileucine-based motifs. AP-1 is responsible for recycling of the M6PRs from endosomes to the Golgi. AP-2 is involved in clathrin-mediated internalization from the plasma membrane and is important for the internalization of secreted lysosomal enzymes. AP-3 is important for the intracellular biogenesis of lysosomes. AP-4 is localized to the Golgi and may also be involved in lysosome biogenesis.

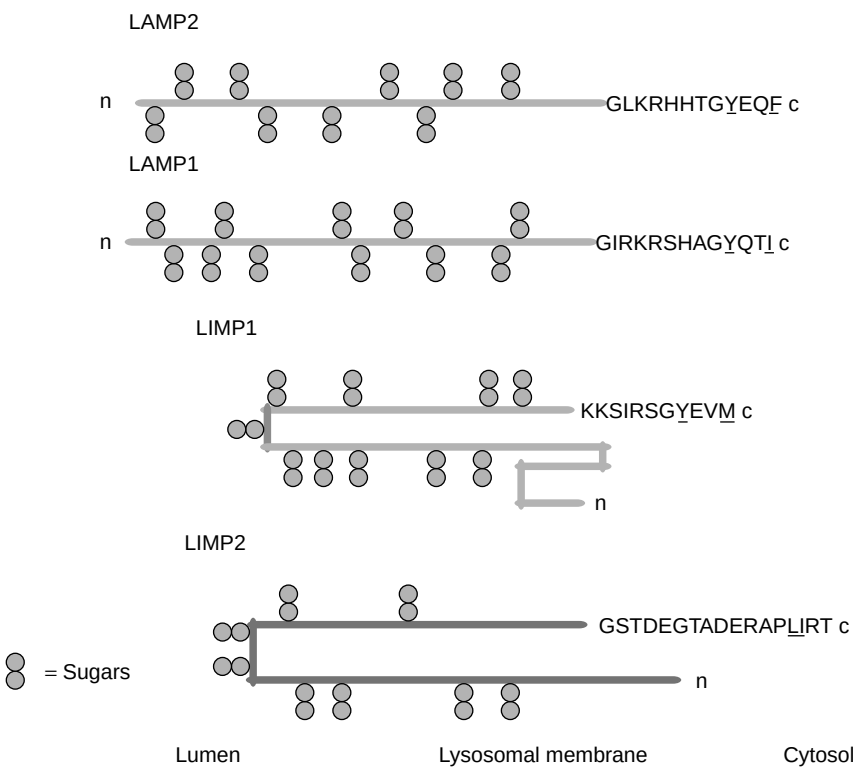

Fig. 2 Diagrammatic structures of integral lysosomal membrane proteins. n = amino terminus; c = carboxyl terminus. LAMPs and LIMPs are defined in the text. The amino acid sequences of the cytosolic tails are shown. Amino acids experimentally shown to be important in targeting the protein to the lysosomal membrane are underlined.

COPI coats also recognize both tyrosine-based and dileucine motifs. An additional family of proteins called Golgi-localizing, gamma adaptin ear homology domain, ARF-binding (GGAs) form vesicular coats and bind to dileucine motifs in the M6PRs. The GGAs are thought to participate in lysosomal biogenesis and/or trafficking of the M6PRs.

Rabs and SNAREs are also required for this vesicular transport. Rab7 is most important for the targeting of lysosomal enzymes to lysosomes while Rab9 is important for targeting the M6PRs from endosomes to the Golgi.

3 Biogenesis of the Yeast Vacuole

We have a more complete understanding of protein targeting to the yeast vacuole compared to the biogenesis of mammalian lysosomes in part because of the power of yeast genetics. The initial stages of protein synthesis and transport through the ER and Golgi, glycosylation of lysosomal enzymes, packaging the enzymes into vesicles, and docking and fusion of those vesicles are similar to processes described for mammalian cells.

3.1 Proteins in the Lumen

Targeting of vacuolar lumen proteins does not require M6P or any other glycosylation signal. The vacuolar targeting of carboxypeptidase Y (CPY) has been studied in detail. More than 50 vacuolar protein sorting (*Vps*) gene product mutations result in CPY missorting and abnormal vacuolar morphology. The proper sorting of CPY requires a linear peptide sequence in the amino-terminal propeptide of the protein. A CPY receptor, Vps10p, binds to this region of CPY in the Golgi and the complex trafficks to a prevacuole compartment called a *multivesicular body* (MVB) prior to entering the vacuole. The Vps10p recycles to the Golgi. Various t-SNAREs and v-SNAREs are required for vesicle fusion with the MVBs or for fusion of MVBs with the vacuole (Fig. 3).

Two vacuolar proteins, aminopeptidase I (API) and α-mannosidase, enter the vacuole independent of any *VPS* genes. They require gene products from the cytoplasm-to-vacuole targeting (Cvt) pathway, which has considerable overlap with genes required for macroautophagy (Fig. 3). Both pathways utilize double-membrane vesicles to initially sequester material. The Cvt pathway selectively sequesters API or α-mannosidase multimers while macroautophagy isolates cytoplasm including API and α-mannosidase. Macroautophagy is stimulated by carbon or nitrogen starvation and is discussed in Sect. 5.3.

3.2 Proteins in the Membrane

Another vacuolar protein sorting pathway for vacuolar membrane proteins bypasses the MVBs and targets Golgi-derived vesicles directly to the vacuole. Two proteins that are targeted to the vacuole as membrane proteins, alkaline phosphatase (ALP) and the t-SNARE, Vam3p, follow this pathway to the vacuole (Fig. 3). The yeast homolog of AP-3 coats the vesicles containing these vacuolar proteins, and the AP-3 appears to be recruited by tyrosine or dileucine peptide sequences in the cytoplasmic tails of ALP and Vam3p. Many *VPS* genes are common for the CPY and ALP pathways, but mutants in specific *VPS* genes selectively disrupt one pathway or the other.

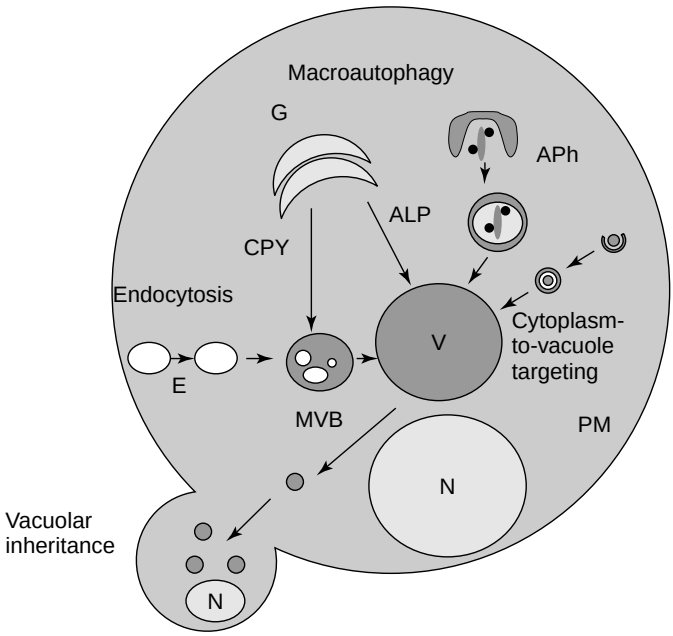

Fig. 3 Pathways for assembly of the vacuole in yeast. G = Golgi; APh = autophagosome; ALP = alkaline phosphatase; CPY = carboxypeptidase Y; E = endosome; MVB = multivesicular body; V = vacuole; PM = plasma membrane; N = nucleus. Pathways of vacuolar assembly and vacuolar inheritance are described in the text.

3.3
Vacuolar Inheritance

Vacuolar inheritance by daughter cells is initiated by a protrusion of the vacuole into the bud region. This protrusion vesiculates, and the vesicles move into the bud (Fig. 3). The vesicles eventually fuse to form the vacuole for the daughter cell.

4
The Importance of Phosphoinositides in Lysosomal Biogenesis

The lysosomal biogenesis pathways in mammalian cells and the vacuolar biogenesis pathways in yeast require phosphotidylinositol (PI) lipids. Phosphorylation of PI to phosphotidylinositol-3-phosphate (PI3P) can be inhibited by compounds such as wortmannin and 3-methyladenine. These inhibitors block lysosomal biogenesis and disrupt several different membrane trafficking processes. For example, both macroautophagy and endocytosis are inhibited by these compounds. PI3P acts at least in part by binding to proteins that are required for vesicular traffic. Some of these proteins contain Fab1, YGLO23, Vps27, EEA1 (FYVE) domains that bind PI3P. Different FYVE proteins are involved in different steps of vesicular traffic, docking, and fusion both in mammalian and in yeast cells.

5
Lysosomal Pathways of Proteolysis

There are several different pathways for intracellular protein degradation. The major pathway in the cytosol and nucleus is the ubiquitin/proteasome system. Other proteolytic systems in the cytosol include calpains and caspases. Most cells also exhibit a lysosomal component of proteolysis, and six different pathways of lysosomal proteolysis have been described in yeast (Fig. 4) and mammalian cells (Fig. 5).

5.1
Endocytosis

Extracellular proteins such as hormones, growth factors, and serum proteins can be internalized by receptor-mediated endocytosis, absorptive endocytosis, or fluid-phase endocytosis by both yeast (Fig. 4) and mammalian cells (Fig. 5). Proteins that are internalized by fluid-phase endocytosis do not interact with the plasma membrane. Proteins that adsorb to plasma membrane proteins or lipids or that bind to plasma membrane receptors can be internalized at rates hundreds or thousands of times more efficiently than by fluid-phase endocytosis alone.

Endosomes are usually coated by clathrin and AP-2 adaptor molecules. The clathrin is uncoated by the constitutively expressed hsp70 (hsc70), and the internalized material travels through early endosomes, late endosomes, and then to the vacuole (Fig. 4) or to lysosomes (Fig. 5) where they are degraded. In some cells, the degradation begins in early or late endosomes but is often completed within lysosomes.

The endocytic pathway can also deliver plasma membrane proteins to lysosomes for degradation (Figs 4 and 5). For example, the insulin receptor along with insulin is degraded within lysosomes. Other receptors may be spared lysosomal proteolysis because they are recycled to the plasma membrane. A growing number of plasma membrane proteins in yeast and mammalian cells are delivered to lysosomes for degradation only after a cytosolic domain has been tagged by ubiquitin.

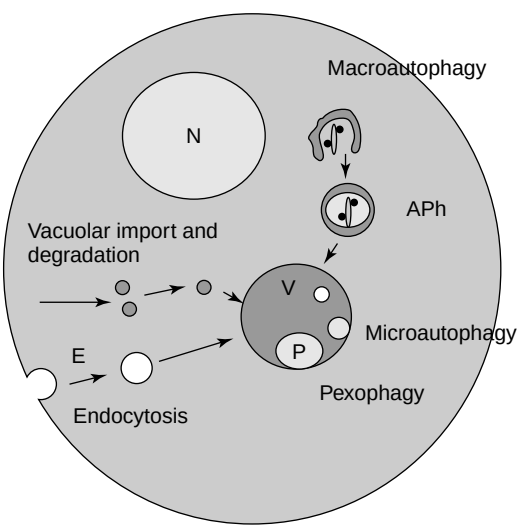

Fig. 4 Vacuolar pathways of protein degradation in yeast. N = nucleus; V = vacuole; APh = autophagosome; E = endosome; P = peroxisome. Four different pathways for vacuolar degradation of proteins, as listed in the figure and described in the text, have been described in yeast. Pexophagy refers to the microautophagic uptake of peroxisomes by the yeast vacuole.

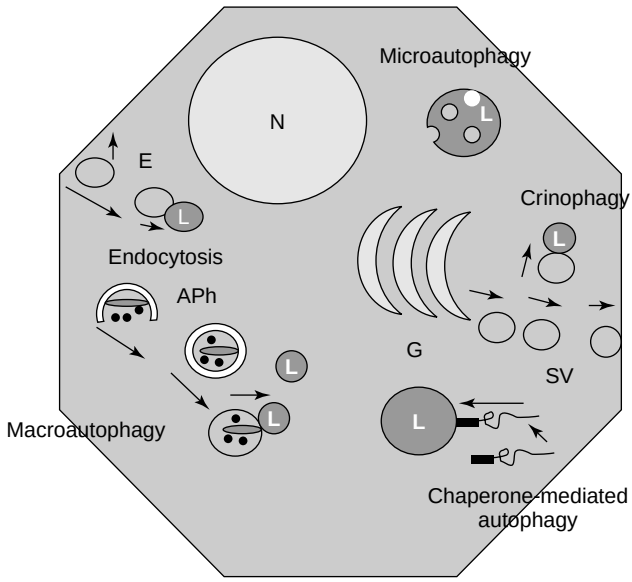

Fig. 5 Lysosomal pathways of protein degradation in mammalian cells. N = nucleus; L = lysosome; E = endosome; APh = autophagosome; G = Golgi; SV = secretory vesicle. Five different lysosomal degradation pathways in mammalian cells are listed in the figure and described in the text.

5.2 Exocytosis

Proteins can be secreted from mammalian cells by constitutive vesicular pathways or by regulated mechanisms in which the protein is stored in secretory vesicles (Fig. 5). When there is reduced physiological need for the secreted proteins, the secretory vesicles can fuse with lysosomes rather than the plasma membrane in a process called *crinophagy*. For example, when blood glucose levels are low, insulin secretion by the pancreatic islet β-cells is reduced. Under such conditions, the secretory vesicles containing insulin fuse with lysosomes and the packaged insulin is degraded. A percentage of serum albumin made by the liver and secreted constitutively is also degraded in this way. This degradation of secretory proteins has not been described in yeast.

5.3 Macroautophagy

Macroautophagy is a degradative pathway that is found in yeast (Fig. 4) and in mammalian cells (Fig. 5). Regions of the cytoplasm are sequestered in double-membrane vesicles called *autophagosomes* (APhs). The membranes surrounding APhs are lipid-rich and contain primarily a single protein. APhs often contain recognizable material such as mitochondria, peroxisomes, ribosomes, or glycogen. The APh acidifies, and then fuses with lysosomes after which the contents of the APh are digested.

Macroautophagy is induced in yeast because of nitrogen or carbon starvation and

in liver, kidney, and other tissues during fasting. The cause for the increased macroautophagy in mammalian tissues is a reduction in circulating amino acids combined with reduced insulin and elevated glucagon. Certain cancer cells are able to grow at least in part because they contain mutations that suppress macroautophagy.

Yeast mutants that fail to carry out macroautophagy under starvation conditions have been isolated and placed into approximately 20 complementation groups. These macro*a*uto*p*hagy (Apg) mutants result in the inhibition of different steps of macroautophagy including formation of APhs, growth of the APhs, docking and fusion with lysosomes, and digestion of the APh contents. The genes required for APh formation include a protein conjugation system reminiscent of ubiquitination. Apg12 is covalently linked to Apg5 through the actions of Apg10 and Apg7. Apg7 is a homolog of the ubiquitin activating protein, E1, and Apg10 is a homolog of a ubiquitin-conjugating protein, E2. Many of the yeast *APG* genes have homologs in mammals, so macroautophagic processes appear to be highly conserved.

Many *apg* mutants are also defective in the Cvt pathway used for delivery of API and α-mannosidase to the yeast vacuole. Other genes are unique for the Cvt pathway or for macroautophagy. For example, the Cvt19 protein acts as a receptor for API and α-mannosidase, and this protein is not required for macroautophagy.

5.4 Microautophagy

Microautophagy is the invagination of vacuolar membrane such that cytosol, organelles, or pieces of organelles are internalized within vesicles in the vacuole (Figs 4 and 5). Vacuolar lipases cause the break down of the internalized membrane. Microautophagy can also operate in modes that are selective for particular organelles such as peroxisomes or the nucleus. Microautophagy of peroxisomes (pexophagy) is stimulated when yeast is switched from methanol to glucose as an energy source (Fig. 4). A form of microautophagy called *piecemeal autophagy* of the nucleus is a constitutive process that is accelerated during starvation. It probably accounts for the degradation of nuclear preribosomal particles under conditions in which ribosome numbers in the cytosol decline. Microautophagy and macroautophagy have partially overlapping genetic requirements, but they also have distinct elements.

Microautophagy has been reproduced using isolated yeast vacuoles. The invagination of the vacuolar membrane requires ATP, GTP, and cytosol. In addition, a membrane potential across the vacuolar membrane is required.

5.5 Vacuolar Import and Degradation Pathway

Proteins such as fructose-1,6-bisphosphatase (FBPase) are synthesized in yeast when gluconeogenesis is required. These proteins are rapidly degraded within vacuoles in response to addition of glucose by a process called *vacuolar import and degradation* (Vid). FBPase may also be degraded by the ubiquitin/proteasome pathway. FBPase trafficking to the vacuole occurs in two steps: (1) import of the protein into an intermediate vesicle, the Vid vesicle, and (2) delivery of the Vid vesicle to the vacuole. The Vid vesicles are unique and contain one prominent membrane protein. Import of the protein into Vid vesicles requires hsc70 while delivery of the Vid vesicles to

the vacuole requires distinct t-SNAREs, v-SNAREs, and other membrane trafficking regulatory molecules.

Whether or not this pathway exists in mammalian cells is not yet known. It shares some mechanistic similarities with chaperone-mediated autophagy. Nevertheless, transport into vesicles prior to lysosome/vacuole delivery of the proteins remains a major difference in these two protein-degradation pathways.

5.6 Chaperone-mediated Autophagy

This lysosomal pathway of proteolysis is activated in liver and other mammalian tissues by prolonged starvation (Fig. 5). Macroautophagy is activated early in starvation but is subsequently suppressed when Cma is activated. In cultured fibroblasts and other cultured cells, macroautophagy is stimulated when cells reach confluence. Withdrawal of serum growth factors increases intracellular protein degradation due to Cma. However, other cell types in culture may stimulate macroautophagy and Cma simultaneously in response to withdrawal of growth factors.

Approximately 30% of cytosolic proteins are substrates for Cma. These proteins contain peptide-targeting sequences biochemically related to KFERQ. Examples of substrates of Cma include ribonuclease A, glyceraldehydes-3-phosphate dehydrogenase, cytosolic aspartate aminotransferase, aldolase B, and certain annexin isoforms, but not others. The KFERQ sequences are recognized by a complex of molecular chaperones including hsc70. The substrate protein-hsc70 molecular chaperone complex then binds to LAMP2a. The substrate protein is unfolded at the lysosomal surface presumably by the molecular chaperone complex prior to the protein's entry into the lysosome in a molecule-by-molecule process. The substrate protein's entry into lysosomes requires an intralysosomal form of hsc70 presumably to pull the substrate protein into the lysosome although the intralysosomal hsc70 may play other roles in Cma as well. The substrate protein is rapidly degraded in the lysosomal lumen.

6 Lysosomal Pathways of RNA Degradation

Lysosomes can also degrade RNA. Lysosomes are able to take up RNA, probably by macroautophagy since the RNA uptake by rat liver lysosomes increases in the absence of amino acids. It is interesting that glucagon also stimulates macroautophagy in the liver, but in this case protein, but little RNA, is delivered to lysosomes for degradation.

7 Lysosomal Storage Diseases

Mutations within genes encoding individual lysosomal enzymes can reduce their targeting to lysosomes and/or their activities within lysosomes. Such mutations result in the accumulation of materials within lysosomes, and identifying this stored material is often critical for identifying the missing enzyme activity (Table 1). For example, Tay-Sachs disease is associated with the lysosomal accumulation of particular gangliosides due to the absence of β-hexosaminidase. A more complicated lysosomal storage disease results from mutations in lysosomal cathepsin A. This peptidase cleaves a major protein of the

Tab. 1 Examples of lysosomal storage diseases.

Disease	*Enzyme deficiency*	*Stored material*
Tay-Sachs	β-Hexosaminidase	Gangliosides
α-Mannosidosis	α-Mannosidase	α-Mannose oligosaccharides
Sialidosis	Sialidase	Sialyl oligosaccharides
Farber	Ceramidase	Ceramide
Gaucher	Glucocerebrosidase	Glucoceramide
Niemann–Pick	Sphingomyelinase	Sphingomyelin
Hunter	Iduronate sulfatase	Heparin sulfate and dermatan sulfate
Galactosialidosis	Cathepsin A	Galactyl and sialyl oligosaccharides
I-cell	6-Phosphotransferase	Multiple substances

lysosomal membrane, LAMP2a. In addition, independent of its peptidase activity, cathepsin A protects two lysosomal glycosidases, β-D-galactosidase and *N*-acetyl-α-neuraminidase, from rapid intralysosomal destruction. In addition, cathepsin A is required for the proper lysosomal targeting of *N*-acetyl-α-neuraminidase. A defect in cathepsin A results in a combined deficiency of lysosomal activities of these two glycosidases and leads to the disease galactosialidosis.

9 Conclusions

Lysosomes are dynamic, multifunctional organelles. Lysosomal enzymatic content can be regulated under certain circumstances, but most control over the digestion of macromolecules is maintained by the sequestration of hydrolytic enzymes into membrane-bound organelles. Control of substrate delivery to the lysosomes appears to be the rate-limited process in these proteolytic pathways. Much less is known about delivery of nucleic acids and lipids for lysosomal degradation, but these topics are worthy of study.

See also Developmental Cell Biology; Electron Microscopy in Cell Biology.

Bibliography

Books and Reviews

Bryant, N.J., Stevens, T.H. (1998) Vacuolar biogenesis in *Saccharomyces cerevisiae*: protein transport pathways to the yeast vacuole, *Microbiol. Mol. Biol. Rev.* **62**, 230–247.

Conibear, E., Stevens, T.H. (1995) Vacuolar biogenesis in yeast: sorting out the sorting proteins, *Cell.* **83**, 513–516.

Cuervo, A.M., Dice, J.F. (1998) Lysosomes, a meeting point of proteins, chaperones and proteases, *J. Mol. Med.* **76**, 6–12.

de Camilli, P., Emr, S.D., McPherson, P.S., Novick, P. (1996) Phosphoinositides as regulators of membrane traffic, *Science* **271**, 1533–1539.

Dice, J.F. (2000) *Lysosomal Pathways of Protein Degradation*, Landes Bioscience, Austin, TX, p. 106.

Dunn, W. (1994) Autophagy and related mechanisms of lysosome-mediated protein degradation, *Trends Cell. Biol.* **4**, 139–143.

Ghosh, P., Dahms, N.M., Kornfeld, S. (2003) Mannose 6-phosphate receptors: new twists in the tale, *Nat. Rev. Mol. Cell. Biol.* **4**, 202–212.

Kornfeld, S., Mellman, I. (1989) Biogenesis of lysosomes, *Annu. Rev. Cell Biol.* **5**, 483–525.

Klionsky, D.J., Emr, S.D. (2000) Autophagy as a regulated pathway of cellular degradation, *Science* **290**, 1717–1721.

Mullins, C., Bonifacino, J.S. (2001) The molecular machinery for lysosome biogenesis, *Bio Essays* **23**, 333–343.

Primary Literature

Abeliovich, H., Dunn, W.A. Jr., Kim, J., Klionsky, D.J. (2000) Dissection of autophagosome biogenesis into distinct nucleation and expansion steps, *J. Cell Biol.* **151**, 1025–1033.

Agarraberes, F.A., Dice, J.F. (2001) A complex of molecular chaperones at the lysosomal membrane is required for protein translocation, *J. Cell Sci.* **114**, 2491–2499.

Agarraberes, F.A., Terlecky, S.R., Dice, J.F. (1997) An intralysosomal hsp70 is required for a selective pathway of lysosomal protein degradation, *J. Cell Biol.* **137**, 825–834.

Ahlberg, J., Glaumann, H. (1985) Uptake-microautophagy- and degradation of exogenous proteins by isolated rat liver lysosomes. Effects of pH, ATP, and inhibitors of proteolysis, *Exp. Mol. Pathol.* **42**, 78–88.

Ahlberg, J., Marzella, L., Glaumann, H. (1982) Uptake and degradation of proteins by isolated rat liver lysosomes. Suggestion of a microautophagic pathway of proteolysis, *Lab. Invest.* **47**, 523–532.

Barrett, A.J. (1992) Cellular proteolysis: an overview, *Ann. N. Y. Acad. Sci.* **67**, 1–15.

Blott, E.J. (2002) Secretory lysosomes, *Nat. Rev. Cell. Mol. Biol.* **3**, 122–131.

Brodsky, J.L., Goeckeler, J., Schekman, R. (1995) Sec63p and BiP are required for both co- and post-translational protein translocation into yeast microsomes, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9643–9646.

Burgess, L., Kelly, R. (1987) Constitutive and regulated secretion of proteins, *Annu. Rev. Cell Biol.* **3**, 243–293.

Burgouyne, R., Morgan, A. (1993) Regulated exocytosis, *Biochem. J.* **293**, 305–316.

Catlett, N.L., Weisman, L.S. (2000) Divide and multiply: organelle partitioning in yeast, *Curr. Opin. Cell Biol.* **12**, 509–516.

Chen, W., Sun, Y., Welch, C., Gorelik, A., Leventhal, A.R., Tabas, I., Tall, A.R. (2001) Preferential ATP-binding cassette transporter A1-mediated cholesterol efflux from the endosomes/lysosomes, *J. Biol. Chem.* **276**, 43564–43569.

Cherqui, S., Sevin, C., Hamard, G., Kalatzis, V., Sich, M., Pequignot, M.O., Gogat, K., Abitbol, M., Broyer, M., Gubler, M.C., Antignac, C. (2002) Intralysosomal cystine accumulation in mice lacking cystinosin, the protein defective in cystinosis, *Mol. Cell Biol.* **22**, 7622–7632.

Chiang, H.L., Terlecky, S.R., Plant, C.P., Dice, J.F. (1989) A role for a 70-kilodalton heat shock protein in lysosomal degradation of intracellular proteins, *Science* **246**, 382–385.

Chirico, W.J., Waters, M.G., Blobel, G. (1988) 70 K heat shock related proteins stimulate protein translocation into microsomes, *Nature* **332**, 805–810.

Chou, H.F., Passage, M., Jonas, A.J. (1997) ATP stimulates lysosomal sulfate transport at neutral pH: evidence for phosphorylation of the lysosomal sulfate carrier, *Biochem. J.* **327**, 781–786.

Clague, M.J. (1998) Molecular aspects of the endocytic pathway, *Biochem. J.* **336**, 271–282.

Conradt, B., Haas, A., Wickner, W. (1994) Determination of four biochemically distinct, sequential stages during vacuolar inheritance in vitro, *J. Cell Biol.* **126**, 99–110.

Creutz, C. (1992) The annexins and exocytosis, *Science* **258**, 924–931.

Cuervo, A.M., Dice, J.F. (1996) A receptor for the selective uptake and degradation of proteins by lysosomes, *Science* **273**, 501–503.

Cuervo, A.M., Dice, J.F. (2000) Regulation of lamp2a levels in the lysosomal membrane, *Traffic* **1**, 570–583.

Cuervo, A.M., Dice, J.F. (2000) Unique properties of lamp2a compared to other lamp2 isoforms, *J. Cell Sci.* **113**, 4441–4450.

Cuervo, A.M., Dice, J.F., Knecht, E. (1997) A population of rat liver lysosomes responsible for the selective uptake and degradation of cytosolic proteins, *J. Biol. Chem.* **272**, 5606–5615.

Cuervo, A.M., Knecht, E., Terlecky, S.R., Dice, J.F. (1995) Activation of a selective pathway of lysosomal proteolysis in rat liver by prolonged starvation, *Am. J. Physiol.* **269**, C1200–C1208.

Cuozzo, J.W., Tao, K., Cygler, M., Mort, J.S., Sahagian, G.G. (1998) Lysine-based structure responsible for selective mannose phosphorylation of cathepsin D and cathepsin L defines a common structural motif for lysosomal enzyme targeting, *J. Biol. Chem.* **273**, 21067–21076.

Deshaies, R., Koch, B., Werner-Washburne, M., Graig, E., Schekman, R. (1988) 70 kD stress

protein homologues facilitate translocation of secretory and mitochondrial precursor polypeptides, *Nature* **332**, 800–805.

Desnick, R.J., Schuchman, E.H. (2003) Enzyme replacement and enhancement therapies: lessons from lysosomal disorders, *Nat. Rev. Genet.* **3**, 954–966.

Dice, J.F. (1990) Peptide sequences that target cytosolic proteins for lysosomal proteolysis, *Trends Biochem. Sci.* **15**, 305–309.

Doray, B., Bruns, K., Ghosh, P., Kornfeld, S. (2002) Interaction of the cation-dependent mannose 6-phosphate receptor with GGA proteins, *J. Biol. Chem.* **277**, 18477–18482.

Eskelinen, E.-L., Illert, A.L., Tanaka, Y., Schwarzmann, G., Blanz, J., von Figura, K., Saftig, P. (2002) Role of lamp2 in lysosome biogenesis and autophagy, *Mol. Biol. Cell.* **13**, 3355–3368.

Franch, H.A., Sooparb, S., Du, J., Brown, N.S. (2001) A mechanism regulating proteolysis of specific proteins during renal tubular growth, *J. Biol. Chem.* **276**, 19126–19131.

Frydman, J., Hohfeld, J. (1997) Chaperones get in touch: the Hip-Hop connection, *Trends Biochem. Sci.* **22**, 87–92.

Gagescu, R., Gruenberg, J., Smythe, E. (2000) Membrane dynamics in endocytosis: structure-function relationship, *Traffic* **1**, 84–88.

Garver, W.S., Heidenreich, R.A. (2002) The Niemann-Pick C proteins and trafficking of cholesterol through the late endosomal/lysosomal system, *Curr. Mol. Med.* **2**, 485–505.

Havelaar, A.C., Mancini, G.M., Beerens, C.E., Souren, R.M., Verheijen, F.W. (1998) Purification of the lysosomal sialic acid transporter. Functional characteristics of a monocarboxylate transporter, *J. Biol. Chem.* **273**, 34568–34574.

Hayes, S., Dice, J.F. (1996) Roles of molecular chaperones in protein degradation, *J. Cell Biol.* **132**, 255–258.

Heydrick, S.J., Lardeux, B.R., Mortimore, G.E. (1991) Uptake and degradation of cytoplasmic RNA by hepatic lysosomes. Quantitative relationship to RNA turnover, *J. Biol. Chem.* **266**, 8790–8796.

Hicke, L. (1999) Getting' down with ubiquitin: turning off cell-surface receptors, transporters, and channels, *Trends Cell Biol.* **9**, 107–112.

Kim, J., Klionsky, D.J. (2000) Autophagy, cytoplasm-to-vacuole targeting pathway, and pexophagy in yeast and mammalian cells, *Annu. Rev. Biochem.* **69**, 303–342.

Knecht, E., Hernandez-Yago, J., Grisolia, S. (1984) Regulation of lysosomal autophagy in transformed and nontransformed mouse fibroblasts under several growth conditions, *Exp. Cell Res.* **154**, 224–232.

Kundra, R., Kornfeld, S. (1999) Asparagine-linked oligosaccharides protect lamp-1 and lamp-2 from intracellular proteolysis, *J. Biol. Chem.* **274**, 31039–31046.

Lecker, S.H., Solomon, V., Price, S.R., Kwon, Y.T., Mitch, W.E., Goldberg, A.L. (1999) Ubiquitin conjugation by the N-end rule pathway and mRNAs for its components increase in muscles of diabetic rats, *J. Clin. Invest.* **104**, 1411–1420.

Lusa, S., Tanhuanpaa, K., Ezra, T., Somerharju, P. (1998) Direct observation of lipoprotein cholesterol ester degradation in lysosomes, *Biochem. J.* **332**, 451–457.

Luzio, J.P., Rous, B.A., Bright, N.A., Pryor, P.R., Mullock, B.M., Piper, R.C. (2000) Lysosome-endosome fusion and lysosome biogenesis, *J. Cell Sci.* **113**, 1515–1524.

Mach, L. (2002) Biosynthesis of lysosomal proteinases in health and disease, *Biol. Chem.* **383**, 751–756.

Matlack, K.E., Misselwitz, B., Plath, K., Rapoport, T.A. (1999) BiP acts as a molecular ratchet during posttranslational transport of prepro-alpha factor across the ER membrane, *Cell* **97**, 553–564.

Mizushima, N., Kuma, A., Kobayashi, Y., Yamamoto, A., Matsubae, M., Takao, T., Natsume, T., Ohsumi, Y., Yoshimori, T. (2003) Mouse apg16L, a novel WD-repeat protein, targets to the autophagic isolation membrane with the Apg12-Apg5 conjugate, *J. Cell Sci.* **116**, 1679–1688.

Mizushima, N., Noda, T., Yoshimori, T., Tanaka, Y., Ishii, T., George, M., Klionsky, D.J., Ohsumi, M., Ohsumi, Y. (1998) A protein conjugation system essential for autophagy, *Nature* **395**, 395–398.

Mizushima, N., Sugita, H., Yoshimori, T., Ohsumi, Y. (1998) A new protein conjugation system in human. The counterpart of the yeast Apg12p conjugation system essential for autophagy, *J. Biol. Chem.* **273**, 33889–33892.

Mortimore, G.E., Lardeux, B.R., Heydrick, S.J. (1989) Mechanism and control of protein and RNA degradation in the rat hepatocyte: two modes of autophagic sequestration, *Revis Biol. Cell.* **20**, 79–96.

Muller, O., Sattler, T., Flotenmeyer, M., Schwarz, H., Plattner, H., Mayer, A. (2000) Autophagic tubes: vacuolar invaginations involved in lateral membrane sorting and inverse vesicle budding, *J. Cell Biol.* **151**, 519–528.

Nebenfuhr, A. (2002) Vesicle traffic in the endomembrane system: a tale of COPs, Rabs, and SNAREs, *Curr. Opin. Plant Biol.* **5**, 507–512.

Nishi, T., Forgac, M. (2002) The vacuolar (H+)-ATPases – Nature's most versatile proton pumps, *Nat. Rev. Mol. Cell Biol.* **3**, 94–103.

Noda, T., Suzuki, K., Ohsumi, Y. (2002) Yeast autophagosomes: de novo formation of a membrane structure, *Trends Cell Biol.* **12**, 231–235.

Ohsumi, Y. (2001) Molecular dissection of autophagy: two ubiquitin-like systems, *Nat. Rev. Mol. Cell Biol.* **2**, 211–216.

Peters, C., von Figure, K. (1994) Biogenesis of lysosomal membranes, *FEBS Lett.* **346**, 108–114.

Roberts, P., Moshitch-Moshkovitz, S., Kvam, E., O'Toole, E., Winey, M, Goldfarb, D.S. (2003) Piecemeal microautophagy of the nucleus in *Saccharomyces cerevisiae*, *Mol. Biol. Cell.* **14**, 129–141.

Rock, K., Gramm, C., Rothstein, M., Clark, K., Stein, R., Dick, L., Hwang, D., Goldberg, A. (1994) Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules, *Cell* **78**, 761–771.

Rotin, D., Staub, O., Haguenauer-Tsapis, R. (2000) Ubiquitination and endocytosis of plasma membrane proteins: role of nedd4/rsp5p family of ubiquitin-protein ligases, *J. Membr. Biol.* **176**, 1–17.

Rouille, Y., Rohn, W., Hoflack, B. (2000) Targeting of lysosomal proteins, *Cell Dev. Biol.* **11**, 165–171.

Sagne, C., Agulhon, C., Ravassard, P., Darmon, M., Hamon, M., El Mestikawy, S., Gasnier, B., Giros, B. (2001) Identification and characterization of a lysosomal transporter for small neutral amino acids, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7206–7211.

Saint-Pol, A., Codogno, P., Moore, S.E. (1999) Cytosol-to-lysosome transport of free polymannose-type oligosaccharides. Kinetic and specificity studies using rat liver lysosomes. *J. Biol. Chem.* **274**, 13457–13455.

Salvador, N., Aguado, C., Knecht, E. (2000) Import of a cytosolic protein into lysosomes by chaperone-mediated autophagy depends on its folding state, *J. Biol. Chem.* **275**, 27447–27456.

Sattler, T., Mayer, A. (2000) Cell-free reconstitution of microautophagic vacuole invagination and vesicle formation, *J. Cell Biol.* **151**, 529–538.

Schatz, G., Dobberstein, B. (1996) Common principles of protein translocation across membranes, *Science* **271**, 1519–1526.

Schmid, S. (1997) Clathrin-coated vesicle formation and protein sorting: an integrated process, *Annu. Rev. Biochem.* **66**, 511–548.

Schnell, A., Swenne, I., Borg, L. (1988) A quantitative estimation of crinophagy in the mouse pancreatic ß-cell, *Cell Tissue Res.* **252**, 9–15.

Shintani, T., Hyang, W.P., Stromhaug, P.E., Klionsky, D.J. (2002) Mechanism of cargo selection in the cytoplasm to vacuole targeting pathway, *Dev. Cell.* **3**, 825–837.

Swanson, J., Burke, E., Silverstein, S.C. (1987) Tubular lysosomes accompany stimulated pinocytosis in macrophages, *J. Cell Biol.* **104**, 1217–1222.

Taute, A., Watzig, K., Simons, B., Lohaus, C., Meyer, H., Hasilik, A. (2002) Presence of detergent-resistant microdomains in lysosomal membranes, *Biochem. Biophys. Res. Commun.* **298**, 5–9.

Teter, S.A., Eggerton, K.P., Scott, S.V., Kim, J., Fischer, A.M., Klionsky, D.J. (2001) Degradation of lipid vesicles in the yeast vacuole requires function of Cvt17, a putative lipase, *J. Biol. Chem.* **276**, 2083–2087.

Tisdale, M.J. (2002) Biochemical mechanisms of cellular catabolism, *Curr. Opin. Clin. Nutr. Metab. Care* **5**, 401–405.

Walkley, S. (1998) Cellular pathology of lysosomal storage disorders, *Brain Pathol.* **8**, 175–193.

Weisz, O.A. (2003) Organelle acidification and disease, *Traffic* **4**, 57–64.

Wenger, D.A., Coppola, S., Liu, S.L. (2003) Insights into the diagnosis and treatment of lysosomal storage diseases, *Arch. Neurol.* **60**, 322–328.

Zhu, Y., Doray, B., Poussu, A., Lehto, V.P., Kornfeld, S. (2001) Binding of GGA2 to the lysosomal enzyme sorting motif of the mannose 6-phosphate receptor, *Science* **292**, 1716–1718.

Bioinformatics: *see* Informatics

Bioinorganic Chemistry

R. Bruce Martin
University of Virginia, Charlottesville, VA, USA

1 Essentiality 637

2 Abundances 639

3 Ionic Radii 639

4 Ca^{2+} Versus Mg^{2+} 639

5 Stability Sequences 641

6 Stability Ruler 642

7 Metal-ion Hydrolysis 643

8 Hard and Soft Acids and Bases 644
8.1 Definitions 644
8.2 Substitution Reactions 644
8.3 Metal-ion Scales 645
8.4 Ligand Scale 646
8.5 Lead 647
8.6 Conclusions 648

9 Nonaqueous Environments 648

10 Other Factors 649

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

11 pH-dependent Stabilities 649

12 Cooperativity 651

13 Metal Ion Binding Characteristics 651
13.1 Amino Acids 651
13.2 Peptides 651
13.3 Proteins 652
13.4 Nucleosides 652
13.5 Nucleic Acids 653

14 Dioxygen 653

15 Metal Ion Complexation Kinetics 654
15.1 Substitution Reactions 654
15.2 Ligand Exchange 655

Bibliography 656
Books and Reviews 656
Primary Literature 657

Keywords

Chelate (from Greek claw)
Multiple bonding of two or more atoms of a single molecule to a metal ion, often to form five- or six-membered rings.

Ligand
A molecule or ion with a donor atom possessing a lone pair of electrons that interacts with a metal ion.

$pH = -\log(H^+)$
The negative decade logarithm of the hydrogen ion activity or concentration. Almost all chemists and biochemists calibrate a pH meter with buffers that yield a pH scale based closely on hydrogen ion activity. Many coordination chemists calibrate a pH meter using known concentrations of hydrogen ion. In the region of 0.1 to 0.2 ionic strength, the concentration scale yields acidity constant logarithms (pK_a values) about 0.12 log unit lower than those on the more common activity-based scale. Stability constants of metal ions with ligands are unaffected by the choice of pH scale.

Bioinorganic chemistry focuses on the roles played by noncarbon elements in life processes. Yet, bioinorganic chemistry is inseparable from the general chemistry of life. Carbon itself cycles among the many bioorganic compounds and inorganic carbon dioxide and carbonates. More than 80% of all the carbon in the earth's crust occurs as $CaCO_3$. About 30% of all enzymes contain metal ion cofactors. The most common metal ion, zinc, appears in over 100 enzymes; iron and copper, in a substantial number; manganese, cobalt, and molybdenum, in a few cases. Selenium appears in the enzyme glutathione peroxidase. Metal ions stabilize nucleic acid polymers, which bear a negative charge on each residue. Though blood is loaded with an array of organic molecules, its main constituent is NaCl and that of intracellular fluids is KCl. Even among the vitamins, a word coined from amines essential to life, the action of vitamin B_{12} depends upon a cobalt ion. Metal ions and nonmetals other than carbon are intimately and inseparably involved in life processes.

1 Essentiality

Twenty-one elements are essential to humans. An essential nutrient is one for which a deficiency results in an impairment of function that is relieved only by administration of that substance. Vitamins by definition and some minerals are essential. The significance of essentiality may be illustrated by burlesquing an old adage.

> "For want of a nail the shoe is lost, for want of a shoe the horse is lost, for want of a horse the rider is lost."
> George Herbert, *Jacula Prudentum* (Outlandish Proverbs), 1640

> "For want of a nutrient the enzyme is lost, for want of an enzyme the function is lost, for want of a function the life is lost."
> Bruce Martin, *Summa Veritatis* (Lofty Truth), 1989

Four essential elements H, O, C, and N comprise more than 99 atom % and about 96 wt % of the human body. These 4 and 14 other essential elements occur among the first 30 elements (through zinc) of the periodic table. Three heavier trace elements Se, Mo, and I are also essential in humans. For 17 tabulated essential elements, Table 1 shows the predominant elemental form at pH 7, typical adult concentrations in the blood plasma or serum, the approximate amount found in a 70 kg adult, and a recommended adult daily allowance. In addition to the four basic elements, the essential elements include two alkali metal ions, two alkaline earth metal ions, seven transition metals (the most common, iron, contributes to less than 0.01% of body weight), phosphorus, sulfur, selenium, and three halogens. Table 1 shows that most of the remaining 4% of body weight consists of 2 elements found in bone, Ca and P. Many of the elements do not exist predominantly in their pH 7 forms in the serum as they are combined with other components. For example, Fe^{3+} does not precipitate as the hydroxide but is retained by tightly chelating ligands. There is little free iodide; it occurs as part of the thyroid hormones. For sulfur, Table 1 lists the total serum concentration, most of which appears in proteins; there is only about 1 mM nonprotein sulfur. Sulfur is not

Tab. 1 Essential elements in humans.[a]

Element	*pH 7 form*	*Serum conc.*	*Human amt.*	*Daily allowance*
Na	Na^+	140 mM	70 g	1–2 g
K	K^+	4 mM	130 g	2–5 g
Mg	Mg^{2+}	0.8 mM	22 g	0.3 g
Ca	Ca^{2+}	2.4 mM	1,100 g	0.8 g
Mn	Mn^{2+}	10 nM	12 mg	3 mg
Fe	$Fe(OH)_3 \downarrow$	17 μM	4 g	10–20 mg
Co	Co^{2+}	2 nM	1 mg	3 μg vit. B12
Cu	Cu^{2+}	17 μM	80 mg	3 mg
Zn	Zn^{2+}	14 μM	2.3 g	15 mg
Cr	$Cr(OH)_2{}^+$	3 nM	6 mg	0.1 mg
Mo	$MoO_4{}^{2-}$	6 nM	5 mg	0.2 mg
Cl	Cl^-	104 mM	80 g	2–4 g
P	$HPO_4{}^{2-}$	1.1 mM	600 g	1 g
S	$SO_4{}^{2-}$	24 mM	120 g	0.7 g Met[b]
Se	$HSeO_3{}^-$	1 μM	5 mg	0.1 mg
F	F^-	2 μM	2.5 g	2 mg
I	I^-	0.4 μM	30 mg	0.15 mg

[a]In addition to H, C, N, and O.
[b]Essential amino acid methionine.

Tab. 2 Body fluids.

mM concentrations	*Extracellular*	*Intracellular*
Na^+	140	10
K^+	4	140
Ca^{2+}	2	0.0001
Mg^{2+}	1	1
Cl^-	104	4
$HCO_3{}^-$	27	10
$HPO_4{}^{2-}$ & $H_2PO_4{}^-$	1	10
$SO_4{}^{2-}$	1	7
pH	7.4	6.6

important as an inorganic element but only as part of the amino acids cysteine and essential methionine. An additional four elements, B, Si, V, and Ni, not included in Table 1, are essential for other organisms.

Table 2 shows the concentrations of several ions in the blood plasma and in a typical intracellular fluid. While the concentration of Mg^{2+} is similar on both sides of the cell membrane, the concentration of K^+ is greater on the inside and those of Na^+ and Ca^{2+} are greater on the outside of a cell. A pump, driven by the protein Ca-ATPase, maintains the 20 000-fold concentration difference for Ca^{2+}. These concentration differences illustrate how compartmentalization in biology leads

to differences in concentration from a surrounding milieu.

Essentiality is not the only criterion for inclusion of elements in a survey of bioinorganic chemistry. Organisms may accumulate elements without their being essential. Some elements such as arsenic and antimony have been used therapeutically, and bismuth still is. Others such as Al and the heavy metals Cd, Hg, Tl, and Pb are toxic and prevalent, and their interactions with life processes are of concern.

2 Abundances

In the natural environment, most essential elements are relatively abundant and most toxic elements are relatively rare. Abundant crustal elements that have not become essential, such as Si, Al, and Ti, are not easily accessible. Life evidently evolved by using elements that are not only abundant but also accessible and suitable. Some relatively common elements are not used in living organisms because similar roles are played by more common elements; Cl is 10^3 times more abundant than Br, Ca is 100 times more abundant than Sr, and Na and Mg are both 100 times more abundant than Li. Boron, the 11th-ranking element in oceans, is essential for some plants but not animals. Essential Mo, I, and Se do not even occur in the upper 50th percentile in crustal atom abundance, but Mo ranks 20th in abundance in the oceans and is the most abundant transition metal ion in seawater.

Of the most toxic elements – arsenic (As), cadmium (Cd), mercury (Hg), thallium (Tl), and lead (Pb) – only the last occurs in the upper 50th percentile in the crust, and As and Cd occur similarly in the oceans. Thus, the most toxic elements are much rarer on the whole than the essential elements. However, both in the crust and in the oceans essential Mo and toxic As appear in comparable amounts as do essential Se and toxic Cd.

3 Ionic Radii

Like size usually plays a greater role than identical charge in promoting metal ion substitutions in biology. Table 3 presents effective ionic radii for many metal ions. For any metal ion, the radius increases with coordination number, since the greater number of bonds weakens the strength of any one bond. The radius of the most common coordination number is underlined in Table 3. The alkali and alkaline earth metal ions exhibit variable coordination numbers without strong directionality in bonding. Because they are of similar size, Ca^{2+} and Na^{+} of differing charges often interchange in biological systems. Ba^{2+} is poisonous because of its antagonism to K^{+} of similar radius and not to Ca^{2+} of identical charge but smaller radius. Ba^{2+} is a muscle poison for which one antidote is intravenous infusion of K^{+}.

4 Ca^{2+} Versus Mg^{2+}

To many, there are insufficient fundamental differences between the pair of alkaline earth metal ions Ca^{2+} and Mg^{2+} to explain their widely differing roles in biology. There are, however, ample ways to account for the behavior differences between Ca^{2+} and Mg^{2+} observed in biochemical systems.

Tab. 3 Effective ionic radii in picometers.

Ion	*Coordination number*[a]					
	4	*5*	*6*	*7*	*8*	*9*
Be^{2+}	27		45			
Al^{3+}	39	48	54			
Cr^{3+}			62			
Ga^{3+}	47	55	62			
Fe^{3+}	49	58	64[b]		78	
Ni^{2+}	55	63	69			
Mg^{2+}	57	66	72		89	
Cu^{2+}	57	65	73			
Co^{2+}	58	67	74		90	
Zn^{2+}	60	68	74		90	
Li^{+}	59		76		92	
Fe^{2+}	63		78[c]		92	
Mn^{2+}	66	75	83	90	96	
Lu^{3+}			86		98	103
Gd^{3+}			94	100	105	111
Cd^{2+}	78	87	95	103	110	
Ca^{2+}			100	106	112	118
La^{3+}			103	110	116	122
Na^{+}	99	100	102	112	118	124
Sr^{2+}			118	121	126	131
Pb^{2+}	98		119	123	129	135
Ba^{2+}			135	138	142	147
K^{+}	137		138	146	151	155
Tl^{+}			150		159	
Rb^{+}			152	156	161	163
Cs^{+}			167		174	178

[a] Radius of the most common coordination number is underscored.
[b] High spin; low-spin value is 55.
[c] High spin; low-spin value is 61.

1. Ca^{2+} favors a higher-fold coordination number than Mg^{2+}. The favored coordination number for Mg^{2+} is sixfold octahedral. Ca^{2+} exhibits a greater diversity of coordination numbers, with eight- and sixfold coordination being the most common with small ligands. Since coordination in both metal ions is mainly ionic, the greater coordination numbers for Ca^{2+} derive ultimately from its larger size.
2. Ca^{2+} is substantially larger than Mg^{2+}. In the same coordination number, the effective ionic radii of Ca^{2+} are at least 23 pm greater than those of Mg^{2+} (Table 1.) If the most favored coordination numbers are considered, Ca^{2+} (eightfold) possesses a 40 pm greater effective ionic radius than Mg^{2+} (sixfold).
3. Ca^{2+} to ligand donor atom bond distances vary more than those of Mg^{2+}. Bond distances to oxygen donor atoms typically range from only 20 to 21 pm for Mg^{2+} and a broader 23 to 26 pm for Ca^{2+}. Thus, Mg^{2+} prefers to form

well-defined six-coordinate octahedral complexes with precise bond lengths and Ca^{2+} forms looser complexes of higher and more variable coordination number, without directionality and with variable bond lengths.

4. Though both ions favor coordination at oxygen donor ligands, Mg^{2+} binds more strongly than Ca^{2+} to small ligands with nitrogen donors. With respect to the pair of related tridentate ligands oxydiacetate and iminodiacetate, Ca^{2+} binds more strongly to the former and Mg^{2+} to the latter N-donor ligand. The stability constant for Mg^{2+} binding is greater for iminodiacetate than for oxydiacetate, while for Ca^{2+} the reverse is true. Mg^{2+} binds to four nitrogen donors in chlorophyll, while Ca^{2+} does not.
5. Ca^{2+} is more comfortable than Mg^{2+} in binding to large, multidentate, anionic ligands. Although Mg^{2+} binds more strongly than Ca^{2+} to iminodiacetate, Ca^{2+} binds more strongly to nitrilotriacetate, EGTA, and EDTA. Ca^{2+} binds to EGTA a striking $10^{5.6}$ times more strongly than does Mg^{2+}. In all three ligands, there are steric requirements that the larger Ca^{2+} fulfills better than Mg^{2+}. Mg^{2+} binds slightly more strongly to nucleoside phosphates,, but within cells, there is insufficient free Ca^{2+} for binding. Within cells, Mg^{2+} associates with phosphates and Ca^{2+} binds tightly to proteins.
   The equilibrium picture shows Mg^{2+} forming more restrictive complexes, usually hexacoordinate with precise bond lengths, favoring smaller chelate groupings, and more readily accepting nitrogen donors. In contrast, Ca^{2+} forms less well-defined complexes with eight-, six-, and other fold coordination with variable bond lengths and favors larger multidentate anionic ligands. These equilibrium stability and structural differences account for differentiation between Ca^{2+} and Mg^{2+} in biological systems.
6. Ca^{2+} undergoes substitution at least 10^3 times faster than Mg^{2+}. The slower rate for Mg^{2+} probably results from a more limited capability to aid substitution by reducing or enlarging its coordination sphere. Thus, from both static and dynamic viewpoints, Mg^{2+} forms the more prescribed and tighter complexes.
7. Insolubilities of their common phosphate salts provide only a weak differentiation between Mg^{2+} and Ca^{2+}. Inorganic Mg^{2+} phosphates are up to 10 times more soluble than Ca^{2+} salts of similar structure. Both metal ions form soluble salts with organic phosphates. The main difference between Mg^{2+} and Ca^{2+} is that, owing to its small size, Mg^{2+} fails to form a structure like the most insoluble of the Ca^{2+} phosphates, hydroxyapatite, $Ca_5(PO_4)_3OH$.

5 Stability Sequences

From the many studies on stability constants, one finds the order of metal ion stabilities to be dependent on the ligand. Increasing metal ion stabilities follow the orders

Glycine:
$Ca, Mg \ll Mn < Fe, Cd, Pb < Co, Zn < Ni \ll CH_3Hg^+, Cu \ll Hg$
1,2-Diaminoethane:
$Mg \ll Mn \ll Fe < Pb, Cd, Co, Zn \ll Ni < CH_3Hg^+ \ll Cu \lll Hg$

The ion Ca^{2+} does not form stable amine complexes. Except for methyl mercury,

CH_3Hg^+, all the metal ions carry two positive charges. Owing to the strongly chelating bidentate ligands, glycine (gly) and 1,2-diaminoethane (en), CH_3Hg^+ with only a single strong binding site is at a competitive disadvantage in the above series. To a unidentate ligand, CH_3Hg^+ binds more strongly than all of the above metal ions except Hg^{2+}.

In the above series, each inequality sign stands for an approximate 10-fold increase in stability constant. The two series gly and en are similar, their major difference being a stability constant span from Mg^{2+} to Hg^{2+} of 10^9 for glycine and 10^{14} for en. Generally, the increment between metal ions increases on passing from O < N < S donor atoms. The presence of a sulfur donor promotes Cd^{2+} and Pb^{2+} to higher positions than that in the above series. The order of increasing sulfur-binding strengths is $Zn^{2+} < Cd^{2+} < Pb^{2+} < CH_3Hg^+ < Hg^{2+}$. Sulfhydryl group interactions are the main mode of toxicity of the heavy-metal ions. These general bonding features are expressed in the concept of the stability ruler.

6 Stability Ruler

For a single ligand, the Irving–Williams stability sequence of dipositive metal ions is Mg < Mn < Fe < Co < Ni < Cu > Zn, invariant of ligand. The uniformly progressive part of this sequence from Mg^{2+} to Cu^{2+} defines a *stability ruler* to which more variable metal ions may be compared. The stability ruler appears across the top of Table 4. For several ligand donor sets, Table 4 shows the relative binding strengths of the variable dipositive metal ions Zn^{2+}, Cd^{2+}, Pb^{2+}, and Hg^{2+}. Their placement in Table 4 corresponds to their stability constants compared with those of the metal ions that define the ruler. The entries for glycine and en agree with the series in the previous section. With most of the metal ions in Table 4, histidine will be tridentate. The increment between metal ions and hence the length of the stability ruler increases with the substitutions O < N < S. The ruler length appears in the second to last column of Table 4 in log*K* units for binding to each ligand. The longer the ruler, the more discriminating

Tab. 4 Stability ruler.[a]

Ligand	*Donors*	*Mg*	<	*Mn*	<	*Fe*	<	*Co*	<	*Ni*	<	*Cu*	*Length*[b]	*Hg*[c]
Hydroxide	OH^-				Cd						Zn	Pb	3.7	1.2
Acetate	O^-									Zn	Cd	Pb	1.3	1.2
Imidazole	=N							Zn	Cd				4.0	1.1
Ammonia	NH_3						Pb		Zn	Cd			4.0	1.1
Oxalate	O^-,O^-				Cd			Zn	Pb				2.1	
Glycine	N,O^-					Cd	Pb	Zn					6.1	0.4
Histidine	N,=N					Cd	Pb	Zn					8.2	0.2
$NH_2(CH)_2NH_2$	N,N					Pb	Cd	Zn					10.1	0.4
$NH_2(CH)_2S^-$	N,S^-									Zn	Cd	Pb	12(est.)	0.2

[a]All dipositive cations.
[b]Ruler length in difference between log stability constants of Cu^{2+} and Mg^{2+} complexes.
[c]Relative magnitude that the Mg^{2+} to Cu^{2+} length must be extended to reach the Hg^{2+} value.

the ligand in selecting among metal ions. Thus, ligands with oxygen donor atoms are least discriminating among metal ions, nitrogen donors intermediate, and sulfur donors most discriminating.

Not only do Co^{2+} and Zn^{2+} display nearly identical radii for the same coordination number in Table 3 but they also exhibit similar stabilities for all ligands in Table 4, except for those involving sulfur and hydroxide. The similarities allow the facile and useful substitution of Co^{2+} for Zn^{2+} in many enzymes. When a sulfhydryl group is present, as in 2-mercaptoethylamine, Zn^{2+} binding strengthens, equaling that for Ni^{2+}. Table 4 also shows the relative strengthening of Cd^{2+} and Pb^{2+} binding with the sulfhydryl donor in 2-mercaptoethylamine.

To most unidentate ligands, Cd^{2+} binds more strongly than Zn^{2+}. Probably because of an unfavorable ring bite size for the relatively large Cd^{2+} (Table 3), Zn^{2+} chelates more strongly to ligands containing O and N donors. Table 2 shows that upon introduction of an S donor atom into a chelate, Cd^{2+} becomes the stronger metal ion binder.

For all donor sets in Table 4, Hg^{2+} binds so strongly that it is off the end of the ruler scale. The number under Hg in the last column of Table 4 refers to the relative distance by which the length of the whole log stability-constant scale from Mg^{2+} to Cu^{2+} must be extended to reach the value for Hg^{2+}. A telling contrast appears between the length of the extension for most bidentate compared to unidentate ligands. The scale extension for Hg^{2+} amounts to 0.2 to 0.4 log units for bidentate ligands and to 1.1 to 1.2 log units for the 3 unidentate ligands at the end of Table 4. The difference arises because Hg^{2+} prefers linear two-coordination and binds the second donor atom in small chelate rings much more weakly than the first donor atom.

We may generalize the results and conclusions of the stability ruler by noting that alkali metal and alkaline earth metal ions, lanthanides, and Al^{3+} prefer oxygen donors; transition metal ions, oxygen and nitrogen donors; and the heavy-metal ions, nitrogen and sulfur donors.

7 Metal-ion Hydrolysis

The higher the charge density or charge-to-radius ratio, the more likely it is for a metal ion to undergo hydrolysis in aqueous solutions to form hydroxo complexes. Hydroxo complexes may abruptly form polynuclear complexes and precipitate even in solutions more acidic than the pK_a for first hydroxo complex formation. The first five small metal ions in Table 3 hydrolyze even in acidic solutions and form precipitates. In six coordination, the charge-to-radius ratio for the first five metal ions is greater than 0.044, while for all the other metal ions in Table 3, the ratio is less than 0.035. The first five metal ions cannot occur to an appreciable extent in the bloodstream (pH 7.4) as the free aqueous ion and must be complexed in some way. Covalence may also promote complex formation and hydrolysis in acidic solutions, as is the case for Hg^{2+}.

Hydroxo complex formation follows the same stability order as other ligands with the strongest binding at the right of the stability ruler in Table 4, where the hydroxide ion appears as the first entry. The stability ruler in Table 4 shows that Zn^{2+} and Pb^{2+} exhibit relatively strong tendencies to form hydroxo complexes and precipitates.

8
Hard and Soft Acids and Bases

8.1
Definitions

More than 45 years ago, metal ions were divided into two groups: a majority labeled class (a) for which the anion-binding strength in aqueous solutions is greatest for F^- and generally follows the order $F > Cl > Br > I$, and a minority class (b) in which F^- binding is weaker than at least one of the heavier anions and generally follows the order $F < Cl < Br < I$ (in the gas phase, all metal ions follow the first order). The same trends occur in other columns in the periodic table; for example, class (a) metal ions tend to favor binding to oxygen and class (b) metal ions to sulfur ligands. Subsequently, Pearson extended the application and changed the terminology so that class (b) metal ions became soft and class (a) metal ions were subdivided into hard and borderline groups. The greater popularity of the altered nomenclature is partly because it is easier to say harder or softer than more (a) character or more (b) character. Thus, it is now said that fluoride and chloride are hard, bromide is borderline, and iodide is soft.

The dictum by Pearson that hard acids prefer to interact with hard bases and soft acids with soft bases has become well known. In biochemistry, there are only four main types of ligand donor atoms, of which O and aliphatic N are classified as hard, aromatic N as borderline, and S as soft bases. All alkali, alkaline earth, and lanthanide ions are classified as hard, as are the first five metal ions in Table 3 and Mn^{2+}. Borderline metal ions include the last four of the stability ruler (Table 4) and Zn^{2+} and Pb^{2+}. Soft metal ions include Ag^+, Cd^{2+}, CH_3Hg^+, and Hg^{2+}.

Despite the popularity of the hard–soft designation, there was a lapse of a quarter century before Pearson proposed a quantitative scale. Pearson defined softness as the reciprocal of hardness and presented quantitative absolute hardness values for metal ions and ligands (in which the hardness of an atom and an anion such as F and F^- are identical). A few years later, Pearson disavowed his quantitative hardness scales for at least some cations and anions. In any case, the hardness values for metal ions do not correlate at all with the stability constants of complexes, which are correlated best by the electron affinity of the metal ion. Though often spoken of by chemists as if it is the determining factor in complex stability, the contribution of hardness or softness may pale when compared to the intrinsic stability, which often dominates the binding strength.

8.2
Substitution Reactions

Comparisons of hardness and softness, like the stability order of halide complexes mentioned in the first paragraph, are equivalent to describing the extent of substitution (not exchange) reactions. We quantitatively evaluate the free energy change in the form of log stability-constant differences of substitution reactions. If M and N are two different metal ions and X and Y two different ligands, we may compare the hardness of a metal ion by considering the substitution on the metal ion of one ligand by another

$$MX + Y \rightleftharpoons X + MY$$

for which the free energy change is proportional to $\log K_{MY} - \log K_{MX}$, where K_{MY} and K_{MX} are the stability constants for complexes MY and MX respectively.

An advantage of such a comparison is that it greatly reduces the overlooked effect of a reduction in coordination number from the aqueous ion for several soft metal ions in augmenting stability constants by powers of ten. Similarly, we may compare the hardness of a ligand by considering the substitution on the ligand of one metal ion for another

$$MX + N \rightleftharpoons M + NX$$

for which the free energy change is proportional to $\log K_{NX} - \log K_{MX}$. By considering such substitution reactions, one gains consistency in application of the principle of hard and soft acids and bases. The scales presented are practical scales dependent only upon experimental stability constants determined in aqueous solutions, and they do not involve any other quantities such as electron affinities and other parameters from gas-phase reactions, heats of hydration, or heats of reaction. Nor do the scales presented assume a model or rely on derived parameters with arbitrary scaling. The practical scales probably represent what most investigators imply when they make comparisons using the terms *hardness* and *softness*.

8.3 Metal-ion Scales

Of two scales involving oxygen and nitrogen donors hydroxide–ammonia and acetate–ammonia, only the latter is described here. As an oxygen donor ligand, hydroxide may not be typical (see below), and for applications acetate may offer a more representative oxygen donor ligand than hydroxide. Acetate–ammonia stability-constant log differences for 22 metal ions and the proton are ordered in the following acetate–ammonia (O–N) scale, with the differences in parentheses. The scale spans 8.0 log units and each of the 8 greater than signs indicates a factor of near 10. Pearson's hard metal ions appear in normal type, borderline in bold, and soft in italics.

Sc(2.8) > La(1.6), Ce(1.5), Y(1.3), Lu(1.2) > *Tl(I)*(0.8), Ca(0.7), Li(0.6), **Pb**(0.6), Mg(0.3) > Fe(III)(−0.4), Mn(−0.4), In(−0.5) > **Co**(−1.0), *Cd*(−1.1), **Zn**(−1.2) > **Ni**(−2.0), **Cu**(−2.3) > *Ag*(−2.9), *Tl(III)*(−2.9) > CH_3Hg^+(−4.1), H^+ (−4.5) > *Hg*(−5.2).

O–N scale

Metal ions favoring acetate appear at the beginning of the O–N scale and those favoring ammonia at the end. Of the first 13 metal ions, 11 are hard, but in this group there is 1 soft (Tl^+) and 1 borderline (Pb^{2+}) metal ion. Next, a group of 4 borderline metal ions also contains 1 soft metal ion (Cd^{2+}). The scale ends with four soft metal ions but also with the hard proton.

Unfortunately, it is impossible to develop a comparable hardness scale with unidentate sulfur donor ligands as very few reliable experimentally determined stability constants are available. Polymerization has gone unrecognized in many of the determinations. Many of the constants involve chelate rings that introduce the additional variable of chelate ring bite size with metal ions of varying radii. Ag^+ and Hg^{2+} prefer linear coordination making formation of five- and even six-membered chelate rings highly strained. For this reason, unidentate ligands are the focus of this article.

By relaxing the requirement for unidentate ligands, we may compare the tendency of some metal ions to bind

to oxygen or sulfhydryl donor atoms. Stability constants have been tabulated for binding to substituted iminodiacetates, $R{-}N(CH_2COO^-)_2$. For $R = -CH_2COO^-$ (nitrilotriacetate) and $R = -CH_2CH_2S^-$, the difference in stability-constant logs for the former minus the latter tridentate ligand is given in parentheses, in sequence, after each dipositive metal ion.

Ca(1.5), Sr(1.4), Ba(1.3), Mg(1.1)
≫ Mn(−1.9), **Ni**(−2.2), **Fe**(−2.9)
≫**Zn**(−5.3), **Pb**(−5.6), *Cd*(−6.9),
Hg(<−8).

O–S scale

Similar differences and sequences are obtained with $R = -CH_2CH_2OH$ as the oxygen donor, with the addition of poorly chelating Hg^{2+}, still with the greatest negative value. Metal ions in the O–S scale fall distinctly into three groups: alkaline earths with differences from +1.5 to +1.1, transition metal ions from −1.9 to −2.9, and filled d-orbital ions from −5.3 to <−8. The spans within the first two groups are narrower than the gaps between the three groups. For oxygen–sulfhydryl binding, the alkaline earths may be designated as hard, the three transition metal ions as borderline, and the four filled d-shell metal ions as soft. The recognized avidity of Zn^{2+} for sulfhydryl donors is confirmed. Though assigned as hard by Pearson, by this grouping Mn^{2+} is emphatically borderline. This borderline designation for Mn^{2+} is also more consistent with other scales.

8.4 Ligand Scale

We now create a ligand scale, corresponding to the second substitution reaction of Sect. 8.2, by considering the differences between stability-constant logarithms for soft and hard metal ions for several ligands. Hg^{2+} stands as the preeminent soft metal ion. Sc^{3+} is the hardest metal ion in several scales, but results are limited. More ligands may be included if results for the smaller and heavier lanthanides are considered for the hard metal ion. For 16 ligands, the difference between the stability-constant logarithms for Hg^{2+} minus a lanthanide in parentheses is as follows:

F^-(−3), H_2O(0), acetate(2),
HCO_3^-(4), CO_3^{2-}(5),
OH^-(5) > **pyridine**(6),
benzimidazole(6), **imidazole**(7),
Cl^-(7), NH_3(8), *SCN^-*(9),
Br^-(9) > *I^-*(13), *CN^-*(13),
$S_2O_3^{2-}$ (14).

Ligand scale

Ligands assigned by Pearson as hard appear in normal type, borderline in bold, and soft in italics. For Pearson, aliphatic amines are hard and aromatic amines are borderline. Since imidazole binds metal ions through a borderline pyridine-type nitrogen, imidazole is also classed as borderline in the ligand scale. Use of CH_3Hg^+ in place of Hg^{2+} yields a virtually identical ordering with a lesser span of 13 log units.

The hardest ligands appear at the beginning of the ligand scale, which spans 17 log units. Cyanide and thiocyanate are ambivalent ligands; virtually all metal ions bind to CN^- through the carbon atom. Even upon binding the first CN^-; Cu^+, Ag^+, and Hg^{2+} undergo a reduction to two-coordination. A reduction to four-coordination occurs upon binding of the second CN^- to Ni^{2+} and Zn^{2+}. A reduction in coordination number strengthens binding.

The above order of increasing softness of the amines is also the order of increasing basicity. The influence of greater basicity in yielding increasing softness is also illustrated by two other comparisons in the ligand scale: $H_2O(0) < OH^-(5)$ and $HCO_3^-(4) < CO_3^{2-}(5)$. These sequences point a general feature of metal ion stabilities: greater basicity yields greater metal ion stability, which in turn provides greater discrimination among metal ions, which in turn gives rise to softness. Often, however, other factors overwhelm the influence of basicity: among the halides, fluoride is both the strongest base and the hardest.

In the ligand scale, the relative softness of the presumably hard ammonia(8), comparable to borderline bromide(9) and soft thiocyanate(9), emerged unexpectedly. The above ligand scale serves as a quantitative indication of the relative hardness of ligands. The 2 inequality signs divide the 16 members of the ligand scale into 3 groups on the basis of the magnitude of the differences. The first 6 ligands are assigned as hard by Pearson; of the next closely spaced 7 ligands, 2 are hard, 4 borderline, and 1 soft. The ligand scale ends with 3 soft ligands. On the basis of the quantitative results of the ligand scale, previously hard ammonia and chloride and soft thiocyanate are more appropriately switched into the borderline group. These switches more consistently place all amines, aliphatic and aromatic, in the same borderline group. The recommended designations lead to a consistent sequence of 6 hard, 7 borderline, and 3 soft ligands. Thus, all oxygen donor ligands are designated hard and all nitrogen donors borderline. Within the borderline group, aliphatic amines are softer than aromatic amines.

8.5 Lead

Lead (Pb^{2+}) presents a case that demonstrates limitations in the hard–soft concept. The first stability-constant logarithms for lead and the halide ions, all at 25 °C and 1.0 M ionic strength, appear in parentheses: F(1.44) > Cl(0.90) < Br(1.10) < I(1.26). The increasing trend for the last three halides describes class (b) behavior, while the greatest value for F^- characterizes a class (a) metal ion. Pb^{2+} interacts relatively strongly with both oxygen (hard) and sulfur (soft) donor ligands, and relatively weakly with nitrogen donor ligands, as illustrated in the stability ruler. Despite the presence of a free sulfhydryl group on the proteins, Pb^{2+} binding occurs exclusively at the Ca^{2+} sites (composed solely of O donors) of oncomodulin and chick vitamin D–induced intestinal calcium-binding protein. Pb^{2+} also combines with the components of nucleic acids. In many examples, Pb^{2+} eschews borderline (is antiborderline) and opts for either hard or soft behavior, making the simple hard–soft concept ineffective for use with this metal ion. In contrast, compared to other metal ions, Ni^{2+} exhibits a tendency to prefer nitrogen over oxygen or sulfur donors (except when two or more sulfur donors promote a diamagnetic complex of reduced coordination number). The position of the proton swings from one end to another of the several metal ion scales. Much more than Pb^{2+}, the proton displays strong antiborderline behavior appearing at either the hard or the soft ends of the scales. Thus, biology offers metal ion environments inconsistent with and unpredictable under

the hard and soft classification scheme. In another example, in the iron–sulfur protein rubredoxin, sulfur exclusively (soft) coordinates to both Fe^{2+} (borderline) and Fe^{3+} (hard).

8.6 Conclusions

Note again that the numerical values associated with hardness–softness in this article are based wholly on experimentally determined log stability-constant values determined in aqueous solutions at room temperature. The values are not absolute; addition of more metal ions might extend the scales in either direction. Small differences between metal ions should not be over-interpreted. The relative difference scales are linear in log stability constant, stretching from very hard at one end to very soft at the other. As hardness decreases, softness correspondingly increases, and vice versa. These practical scales are not comparable to that of Pearson, where absolute hardness values are derived from gas-phase parameters, and softness is the reciprocal of hardness.

In practice, the principle of hard and soft acids and bases is often contorted in such a way as to provide nonfalsifiable explanations for almost any observation. The limitations of the principle need greater exposure. The very hard Sc^{3+} and the very soft CH_3Hg^+ both bind strongly and nearly equally to the hard ligands hydroxide on one hand and acetate on the other (Table 1). Though both metal ions are considered as very soft, Hg^{2+} is among the strongest binders to hydroxide and acetate, and Ag^+ among the weakest. Many other such anomalies may be found by considering differences of stability-constant logs. It is only when these differences are compared that one finds some quantitative justification for the principle of hard and soft acids and bases. The principle of hard and soft acids and bases finds its most consistent application not in direct stability-constant comparisons but rather in the free energy or log stability-constant differences of the substitution reactions of Sect. 8.2.

9 Nonaqueous Environments

Nonaqueous solvents that lower the dielectric constant increase stability constants between oppositely charged ions, decrease the constants between identically charged ions, and leave relatively unaffected stabilities with one neutral reactant species. For example, on passing from water to a mixed solvent system containing 70% by weight dioxane, the dielectric constant drops from 79 to 18. This solvent change increases the pK_a of acetic acid by 3.6 log units and decreases the pK_a of anilinium ion by 1.0 log unit. The much larger change in the former case is due to the reaction between oppositely charged proton and acetate becoming much more favored in the lower dielectric medium. In contrast, the latter reaction is merely transfer of a proton from water to amine without the creation or destruction of any charges.

Binding sites in proteins display lower dielectric constants than the surrounding water. An equivalent solution dielectric constant applicable to binding sites in metalloproteins is calibrated with metal ion stabilities in nonaqueous solvent mixtures. For the zinc ion active sites in bovine carbonic anhydrase and carboxypeptidase A, the equivalent solution dielectric constants are estimated as 35 and 70 respectively. The former enzyme reacts with a smaller

substrate and possesses a much less open binding site.

10 Other Factors

Numerous other factors affect metal ion affinities including size and number of chelate rings, presence of other ligands, noncovalent interactions among ligands, ligand field strength (at most 10% of total binding energy), and spin state (important in heme proteins).

11 pH-dependent Stabilities

If at a given pH a ligand occurs in a protonated form, competition develops between metal ion and proton for a basic ligand site. The concentration of basic ligand available to the metal ion drops in the presence of protonated ligand species. Neutral solutions contain ligands with protonated amine, phenolic, and sulfhydryl groups. In these and other cases, the tabulated stability constants overstate the effective binding strength, and its magnitude needs to be reduced to reflect competition from the proton.

The tabulated stability constant K_s refers to the following equilibrium:

$$\mathrm{M + L \longrightarrow ML} \quad K_s = \frac{[\mathrm{ML}]}{([\mathrm{M}][\mathrm{L}])}$$

With a protonated ligand, however, the relevant formation reaction becomes $\mathrm{M + HL \rightarrow ML + H^+}$. There is the additional acid–base equilibrium of the ligand.

$$\mathrm{HL \longrightarrow H^+ + L} \quad K_a = \frac{(\mathrm{H^+})[\mathrm{L}]}{[\mathrm{HL}]}$$

The fraction of the ligand in the basic form available for metal ion binding is given by

$$\alpha = \frac{[\mathrm{L}]}{([\mathrm{HL}] + [\mathrm{L}])} = \frac{K_a}{[(\mathrm{H^+}) + K_a]} \text{ so that } 0 < \alpha < 1.$$

The conditional, pH-dependent stability constant is given by $K_c = \alpha K_s$, or

$$\log K_c = \log K_s - \mathrm{p}K_a - \log[(\mathrm{H^+}) + K_a]$$

For $\mathrm{pH} \gg \mathrm{p}K_a$, as for carboxylates in neutral solutions, the basic form of the unbound ligand predominates and $\log K_c = \log K_s$. For $\mathrm{pH} \ll \mathrm{p}K_a$, as for aliphatic amines in neutral solutions, the unbound ligand is predominantly protonated and $\log K_c = \log K_s - \mathrm{p}K_a + \mathrm{pH}$, and the conditional constant shows a pH dependence. When the pH is within 1.5 log units of $\mathrm{p}K_a$, the complete equation should be used. By allowing for withdrawal of deprotonated ligand from solution by protonation, the value of the conditional stability constant becomes less than the standard stability constant, $K_c \leq K_s$. (The conditional constant may exceed the tabulated constant for cases in which a subsequent deprotonation occurs from a complex.)

In Table 5, conditional stability constants at pH 7.0 (represented as K_7) are compared with the tabulated stability constants (K_s) for three metal ions and four unidentate ligands. The four ligands represent carboxylate, imidazole, sulfhydryl, and aliphatic amine donors in biological systems. For acetate, $\mathrm{pH} = 7 \gg \mathrm{p}K_a$, $\alpha = 1$, virtually all ligand is in its basic, carboxylate form, and the conventional and conditional stability constants are equal. For imidazole, 44% of the ligand is in the basic form. For glutathione and ammonia, only 1.2 and 0.47% respectively of the ligands are in the basic form.

Tab. 5 Conventional and conditional (pH 7.0) stability constants.

Ligand	pK_a	Ni^{2+}		Zn^{2+}		Cd^{2+}	
		$\log K_s$	$\log K_7$	$\log K_s$	$\log K_7$	$\log K_s$	$\log K_7$
Acetate (O^-)	4.7	0.7	0.7	0.9	0.9	1.3	1.3
Imidazole (=N)	7.11	3.0	2.6	2.52	2.2	2.80	2.4
Glutathione (RS^-)	8.92	4.0	2.1	5.0	3.1	6.16	4.2
Ammonia (NH_3)	9.33	2.79	0.5	2.31	0.0	2.65	0.3
Glycine	9.68	5.80	3.1	5.03	2.3	4.28	1.6
Histidine	9.15 6.10	8.7	6.5	6.6	4.4	5.4	3.2
1,2-Diaminoethane	10.0 7.2	7.3	3.9	5.7	2.3	5.4	2.0

Glutathione is the tripeptide γ-L-glutamyl-L-cysteinylglycine, and it serves as a superb model for unhindered sulfhydryl groups in proteins.

The results in Table 5 show that for all three metal ions, the ligand stabilities from the conventional constants follow the increasing sequence acetate < ammonia < imidazole < glutathione. This sequence differs from the stability order in neutral solutions because the sequence applies only when all ligands are predominantly in their basic form, which is true only for acetate. For conditional constants at pH 7.0, Table 5 shows that for Ni^{2+} the increasing stability order is ammonia < acetate < glutathione < imidazole, while for Zn^{2+} and Cd^{2+} the order is ammonia < acetate < imidazole < glutathione. For all three metal ions in neutral solutions, the most basic ammonia has become a weaker ligand than the least basic acetate. The stability order from conditional constants suggests that in neutral solutions, it is apt for Ni^{2+} to be found at imidazole sites and Zn^{2+} and Cd^{2+} at sulfhydryl groups. Of course, chelation by adjacent donors may increase stabilities and alter preferred binding sites. Yet, the examples in Table 5 serve as models for the analogous groups in proteins and prove that binding strengths and stability orders are pH-dependent.

In the nucleosides of the 6-oxopurines, guanosine and inosine, in neutral solutions, the much more basic N1 site is protonated while the N7 site is difficult to protonate. Metal ions coordinate in acid solutions at N7, and as the pH increases, they compete better with the proton and crossover to N1.

The principles developed in this section extend to ligands with several basic sites. For diamines, EDTA, and catecholates in which metal ions displace two protons in neutral solutions, the fraction of unbound, doubly deprotonated ligand becomes $\alpha = K_{a1}K_{a2}/[(H^+)^2 + (H^+)K_{a1} + K_{a1}K_{a2}]$, where $pK_{a1} < pK_{a2}$. Table 5 also concludes with two bidentate amine ligands and histidine. For 1,2-diaminoethane, the conditional stability constants at pH 7.0 are 3.4 log units less than tabulated stability constants.

Ignored in this section and Table 5 is the formation of metal ion hydroxo complexes and precipitates. Formulation of conditional stability constants that incorporate hydroxo complex formation is outside the scope of this section.

12
Cooperativity

Life abounds in cooperative interactions and would not exist, as we know it, without them. For example, the important enzyme that maintains the 20 000-fold Ca^{2+} gradient across cell membranes binds 2 Ca^{2+} or 2 protons so cooperatively that it is not possible to resolve reliably the successive stability constants. What is not appreciated, and what makes meaningless all calculations of the origin of life based on random events, is the cooperation that occurs in even the simplest systems. For statistical reasons at least, and in the case of charged ligands for electrostatic reasons as well, we expect the successive stability constants for ligand binding to a metal ion to decrease as the number of ligands increases. This decrease is generally observed. Yet, there are simple systems in which the second ligand is bound more strongly than the first.

Both the ammonia and imidazole complexes of Ag^{+} exhibit an inversion of the usual stability order: for NH_3, $\log K_1 = 3.20$ and $\log K_2 = 3.83$, while for imidazole, $\log K_1 = 3.1$ and $\log K_2 = 3.8$. Since there is a statistical factor of 4 or 0.60 log units favoring K_1 over K_2, the observed inversion is actually 1.23 to 1.3 log units or a factor of about 20. This result means that once the first ammonia molecule is bound, the second is bound significantly more strongly. Binding of two ligands to Hg^{2+} and four to Zn^{2+} also exhibits cooperative behavior. In these cases, the explanation lies in a decrease in coordination number upon ligand binding, resulting in shorter and stronger bonds to the remaining ligands.

In hemoglobin, the successive stability constants for dioxygen binding to the four iron-containing heme groups are cooperative with $K_1 < K_2 < K_3 \ll K_4$. In this case, the inverse order is caused by interactions among the α and β protein chains and is of enormous physiological significance. The inverse order means that once the first dioxygen binds, the second, third, and especially the fourth are bound even more avidly. As a result, wholly deoxygenated and wholly oxygenated forms predominate. They do so to such an extent that there are only small amounts of species with one, two, and three dioxygen molecules, and the individual stability constants are resolved with difficulty though the overall product $K_1K_2K_3K_4$ is well known. In an oxygen-rich environment such as the lung, hemoglobin becomes wholly oxygenated, while in an oxygen-poor tissue, the entire oxygen load undergoes efficient release.

13
Metal Ion Binding Characteristics

13.1
Amino Acids

Amino acids bind transition metal ions more avidly with increasing pH since the amino group suffers less competition from the proton. The potentially tridentate amino acids histidine and cysteine are especially strong transition metal ion binders. Alkali and alkaline earth metal ions, Al^{3+}, and lanthanides bind only weakly to amino acids.

13.2
Peptides

Owing to loss of basic carboxylate and amino groups, peptides usually bind metal ions more weakly than amino acids unless the metal ion deprotonates the peptide

nitrogen. The oxygen of a peptide bond is only a weak metal ion binder, and the peptide nitrogen is not a proton or metal ion binding site unless deprotonated. For amide deprotonations, $pK_a \approx 15$. A deprotonated peptide nitrogen serves as a strong metal ion binder, but unless there is an anchor for the metal ion, amide deprotonation does not take place until the solution is very basic. The Cu^{2+}–biuret reaction is an amide deprotonation in very basic solutions.

The tripeptide Gly-Gly-His serves as a strong quadridentate ligand with an amino nitrogen (an anchor), two deprotonated peptide nitrogens, and an imidazole nitrogen (a second anchor) providing a planar array of four donor atoms. Binding in this mode occurs best in neutral and even acidic solutions with Cu^{2+}, Pd^{2+}, and Ni^{2+}, which undergoes a transition from a blue, octahedral, high-spin complex to a yellow, planar, low-spin complex upon cooperative deprotonations of the peptide nitrogens. In blood, Cu^{2+} interacts in this binding mode with the amino terminus of human serum albumin where histidine is the third amino acid residue.

For neutral amides, the weakly basic oxygen ($pK_a \approx -1$) is the proton and metal ion binding site. The major determinant of peptide oxygen basicity is the basicity of the amino group making up the bond. Among the 20 amino acids in proteins, the amino group basicity of proline is almost 1 log unit greater than that of the other amino acids. As a result, the peptide oxygen from the amino acid linked to a proline nitrogen should be relatively basic compared to other peptide oxygens not linked likewise. Proline-linked peptide oxygens appear disproportionately as calcium ion binding sites in calcium proteins. Proline itself never furnishes the peptide oxygen that coordinates a metal ion. How much a relatively greater basicity, rather than appearance of proline in loop regions, contributes to proline-linked peptide oxygen occurrence as a metal binder remains uncertain.

13.3 Proteins

Metal ions interact in numerous ways with proteins, from the weak fairly nonspecific interactions of many metal ions with proteins such as serum albumin to the highly specific protein sites made for exclusive binding of a single kind of metal ion. For example, the iron-transporting protein of the plasma, transferrin, binds two Fe^{3+} under blood plasma conditions with conditional dissociation constants of 10^{-22} M. Such a minuscule value is necessary because the solubility of goethite, FeO(OH), only allows 10^{-21} M Fe^{3+} at pH 7.4. About 30% of all enzymes contain metal ions, and in most of these, only one kind of metal ion is bound almost exclusively. Individual cases are the subject of other articles in this volume.

13.4 Nucleosides

Let us designate the nucleic bases with their usual alphabetical symbols, A, C, G, T, and U. The order of decreasing basicity is given by

$$H^+: T3 > U3 > G1 \gg C3 > A1 > G7 > A7$$

In contrast, for a heavy-metal ion in neutral solutions, the order of decreasing stability is given by

$$\text{Heavy-metal ion: } G7 > A7 > C3 > A1 > G1 > U3 > T3$$

Comparison of the two series reveals a marked promotion of the G7 and A7 sites on passing from the first to the second series. Moreover, in the DNA double helix, the N7 sites are not involved in interstrand hydrogen bonding and are consequently relatively exposed. Thus, the antitumor *cis*-diaminodichloroplatinum(II) binds mainly at the guanosine N7 site of DNA. The sugar moiety binds metal ions very weakly and insignificantly.

Though proposed many times as metal ion binding sites, primary amino ($-NH_2$) groups located at C4 in cytidine, C6 in adenosine, and C2 in guanosine are neither proton nor metal ion binding sites in neutral solutions. These amino groups are not basic, the flat $-NH_2$ group is nearly coplanar with the ring, and the amino nitrogen–to-carbon bond lengths are about 6-pm shorter in the nucleic bases than in aniline. This combination of properties indicates appreciable double-bond character and extensive delocalization of electron density into the rings. Consistent with their appreciable *positive* charge density, the primary amino groups serve as hydrogen bond donors in specific base pairing. Moreover, in strongly acidifying solvents, even the dication of cytosine and the trications of adenine and guanine have still not undergone amino group protonation as the ring nitrogens and O2 in cytosine and O6 in guanine protonate before the amino group. Only upon its deprotonation in strongly basic solutions does the primary amino group ($pK_a \approx 16$) coordinate a few strongly binding metal ions such as CH_3Hg^+.

13.5 Nucleic Acids

With the introduction of more than one basic phosphate group, the di- and tri-nucleotides ($pK_a = 6.5$) become strong metal ion binders. Virtually all reactions of ATP require a Mg^{2+} cofactor, and ATP occurs as a Mg^{2+} complex in cells. Nucleic acid chemists were slow to recognize the role of ambient metal ions in stabilizing polymeric nucleic acid structures. With a negative charge on the phosphate of each residue, only a random-coil form exists at low salt concentrations. The phosphate groups in the polymers are not basic ($pK_a \approx 1$) and bind metal ions very weakly. Required for structures such as the double helix are nonspecifically bound alkali metal ions and Mg^{2+} serving as counterions to offset the negative charges on the phosphates. Mg^{2+} is also important in stabilizing various RNA structures.

14 Dioxygen

It is impossible to overstate the effects that increases in the dioxygen content of the atmosphere have wrought for life on Earth. On primitive Earth, the atmosphere was free of dioxygen, and beginning about 3 500 million years ago, anaerobic prokaryotes thrived. About 2 700 million years ago, prokaryotes began emitting dioxygen, and the atmospheric content began increasing. Soluble Fe^{2+} was oxidized to Fe^{3+}, which formed a very insoluble goethite, $Fe(OH)_3$, that allows only 10^{-20} M free Fe^{3+} at pH 7.0. (For a typical cell volume of 10^{-12} L, this implies only one Fe^{3+} ion in 10^8 cells!) Thus, this important and formerly freely available element became scarce, and it was necessary for nature to find new ways to handle it. The ever-increasing dioxygen concentration became an environmental catastrophe for an anaerobe unable to find a niche or evolve.

However, by allowing a greater diversity of energy-producing reactions, the increasing dioxygen concentration promoted the rise of eukaryotes about 1 500 million years ago and of the first multicellular organisms about 650 million years ago.

Prokaryotes do not form true multicellular structures. Later, sunlight interacting with dioxygen made enough ozone in the stratosphere to provide an umbrella for ultraviolet radiation, so that about 410 million years ago, plants were able to invade the land. Thus, though life appeared at just about the first 1 000 million years of the earth's history, it took over 3 000 million years from the origin of life on earth through the production of a dioxygen-containing atmosphere and ozone shield to extensive land plants. Dioxygen is the pivotal element the absence or presence of which determines the forms of life on earth. It took more than 2 000 million years after the first appearance of dioxygen in the atmosphere to yield an environment in which multicellular organisms arose and became the basis for life as we know it today.

Dioxygen, so necessary for the life forms on earth today, is also toxic and contributes to cell aging. With two unpaired spins, O_2 reacts readily with radicals and some metal ion complexes. Subsequent reactions generate other reactive molecules and radicals: hydrogen peroxide, H_2O_2; hydroperoxides, ROOH; superoxide, O_2^-; and the hydroxyl radical, HO, which is highly damaging. The last two species initiate deleterious radical chain lipid peroxidations that contribute to aging. Therefore, pivotal dioxygen provides life with both the source of its versatility via oxidation processes and the seeds of its demise by radical reactions.

Almost all organisms possess an enzyme, superoxide dismutase, to rid cells of readily generated and reactive superoxide, O_2^-, by converting it to less-toxic hydrogen peroxide, $2\ O_2^- + 2H^+ \rightarrow H_2O_2 + O_2$. Superoxide dismutase is much more widespread than catalase, which catalyzes the decomposition of hydrogen peroxide.

15 Metal Ion Complexation Kinetics

Belying a common understanding of the word *formation* that might imply an addition reaction, metal ion complex formation in solution is almost always a *substitution* (occasionally with a reduction in coordination number). In aqueous systems, bound water undergoes substitution either in the reactant complex or as an intermediate in the route from reactants to products.

15.1 Substitution Reactions

The greater complexity of kinetics over thermodynamics is shown in the simple-appearing ligand replacement reaction with *n* fixed ligand donor atoms, L, bound to metal ion M (charges not specified).

$$L_nMX + Y \longrightarrow L_nMY + X$$

While thermodynamics deals with the species above, kinetics also includes additional species involved in the mechanism of the reaction. Substitution of unidentate ligand X by unidentate ligand Y may proceed by a reduction of coordination number (dissociation) to give the intermediate L_nM, by an increase in coordination number (association) to yield the intermediate $L_nM(X)Y$, or by a process in which the incoming and outgoing ligands interchange in a more or less concerted fashion. From the wholly dissociative mechanism

through the varying degrees of the interchange route to the wholly associative mechanism, there is a continuous spectrum of possibilities. Moreover, in aqueous systems, virtually all such substitution reactions occur via an aquo intermediate $L_nM(OH_2)$. Thus, for the most general case in which X and Y are not water, the question just posed as to the coordination number of the intermediates needs to be answered twice: for both formation and reaction of the aquo intermediate.

Detailed answers about the mechanism of substitution reactions in metal ion complexes have been sought for many years. Some general conclusions now seem reliable, but universal agreement on all details remains to be reached. *Two statements sum up many of the results*: (1) since room is available for approach of an axial ligand, most substitutions in linear, trigonal, and planar complexes proceed by the associative mechanism with an increase in coordination number and (2) since the environment about the metal ion is more crowded, most substitutions in octahedral complexes proceed in the interchange-dissociative region of the continuum.

15.2 Ligand Exchange

Even if they do not appear as reactants or products, the intermediacy of aquo complexes in most substitution reactions, whatever the detailed mechanism of substitution in any particular case, makes the properties of the aquo complexes of primary importance.

The characteristic rate constant for exchange of inner-sphere water (both X and Y are H_2O in the above reaction) gives insight into the differences among metal ions. Increasing water exchange rates on wholly aquo metal ions follow the order

$$Cr^{3+}, Ru^{3+} \ll Pt^{2+} \ll Ru^{2+} < Co^{3+} < Al^{3+} \ll V^{2+}, Fe^{3+} < VO^{2+}, Ga^{3+}, V^{3+}, Pd^{2+}, Be^{2+} < Ni^{2+} < Mg^{2+} < Co^{2+}, Fe^{2+} < Mn^{2+}, Zn^{2+}, Sc^{3+} < Ca^{2+}, Cd^{2+}, Gd^{3+} < \text{alkali metal ions}, Cr^{2+}, Cu^{2+}, Hg^{2+}, Pb^{2+}$$

with each inequality sign indicating an approximate 10-fold increase in rate. Water exchanges slowly with the first 6 aqua cations through Al^{3+}, rapidly with the next 14 through Sc^{3+}, and virtually instantaneously with the last 8 entries. The series spans a phenomenal 15 powers of 10 from a lifetime at 25 °C of about 5 days for the mean time of a water molecule on aqueous Cr^{3+} to 10^{-9} s for the metal ions at the end of the list. Indeed, exchange lifetimes are so short for the cations in the last line that some values are not known with certainty: some of the lifetimes may be 10^{-10} s or even shorter.

The above series is useful in several ways. The contrast between the presence of Zn^{2+} in numerous mammalian enzymes and of Ni^{2+} in only a few plant enzymes has proved puzzling. Ni^{2+} differs by neither size nor most complex stabilities from Zn^{2+}. The above exchange rate series explains the difference since Zn^{2+} undergoes ligand exchange about 10^3 times faster than Ni^{2+}, a significant factor for a metal ion bound at enzyme active sites.

In four cases, reduction of the metal ion substantially increases the rate. An approximate 10^4-fold rate increase occurs for the $d^5 \rightarrow d^6$ reduction in both Ru^{3+} and Fe^{3+}. An even greater 10^7-fold rate increase takes place upon the reduction of $Co^{3+} \rightarrow Co^{2+}$. Finally, the 10^{15} span of the entire series is gained upon reduction of $Cr^{3+} \rightarrow Cr^{2+}$.

Though these lifetimes refer to water exchange in the wholly aquated metal ions, they also reflect the relative rates of exchange of other ligands. When bound to a metal ion, other ligands, including hydroxide ion, increase modestly the release rate of bound water, typically by about 10^2. As a poorer leaving group, hydroxide ion itself exchanges much more slowly than water. Chelated ligands exchange more slowly, but again, the relative order prevails. In the case of chelates, binding occurs first by one donor group followed by chelate ring closure. In some cases, such as five-membered rings, the chelate ring closes rapidly, and in others, especially with larger rings, ring closure may become the slow step. Selective broadening of lines in nuclear magnetic resonance spectroscopy that have been used to indicate the binding site of a paramagnetic metal ion at stoichiometric concentrations often fails. This occurs because chelate ring closure is slow compared to rapid passage of the metal ion to other ligands present in large excess, which bind only in a unidentate mode.

We conclude with an application that incorporates many of the principles mentioned in this chapter. Surprisingly, the above exchange rate series demonstrates that some of the strongest-binding metal ions undergo the most rapid ligand exchange. Examples from the high-rate end of the series include Pb^{2+}, Hg^{2+}, Cu^{2+}, and Cd^{2+}, all strong ligand binders. A significant feature of the toxicology of Hg^{2+} and CH_3Hg^+ is the rapid exchange of ligands in and out of coordination to the metal ion. For CH_3Hg^+ binding to the sulfhydryl group of glutathione (present in blood plasma at 4 mM), the stability-constant logarithm is an extraordinary $\log K = 15.9$. For such strong binding, the half-life for exchange by a dissociative mechanism is about 10 days. Yet, the average lifetime for the CH_3HgSR complex in the presence of excess glutathione is less than 0.01 s. This comparison shows that exchange does not occur by a dissociative mechanism. The ability of linear two-coordinated Hg^{2+} and CH_3Hg^+ to associate weakly with additional donor atoms accounts for rapid metal ion exchange among donor atoms. The last two ligands are bound much more weakly than the first two and exhibit longer Hg-to–donor atom bond lengths. Excess ligand participates in nucleophilic attack at an uncoordinated site on the metal ion with rearrangement in the coordination geometry and release of a formerly bound ligand. This addition–elimination mechanism accounts for the very rapid exchange in Hg^{2+} and CH_3Hg^+ complexes.

See also Bioorganic Chemistry; Calcium Biochemistry.

Bibliography

Books and Reviews

Berthon, G. (Ed.) (1995) *Handbook of Metal-Ligand Interactions in Biological Fluids: Bioinorganic Chemistry, Vols. 1 and 2 and Bioinorganic Medicine, Vols. 1 and 2*, Marcel Dekker, New York.

Bertini, I., Sigel, A., Sigel, H. (2001) *Handbook on Metalloproteins*, Marcel Dekker, New York.

da Silva, J.J.R.F., Williams, R.J.P. (2001) *The Biological Chemistry of the Elements*, 2nd edition, Clarendon Press, Oxford.

Holm, R.H., Solomon, E.I. (Eds.) (1996) Bioinorganic Enzymology, *Chem. Rev.* **96**, 2237–3042. Thematic issue with 24 articles.

Kaim, W., Schwederski, B. (2000) *Bioinorganic Chemistry: Inorganic Elements in the Chemistry of Life*, Wiley, Chichester.

Klotz, I.M., Kurtz, D.M. Jr. (Eds.) (1994) Metal dioxygen complexes, *Chem. Rev.* **94**, 567–856. Thematic issue with 12 articles.

Lippard, S.J., Berg, J.M. (1994) *Principles of Bioinorganic Chemistry*, University Science Books, Mill Valley, CA.

Martin, R.B., Mariam, Y.H. (1979) Interactions between metal ions and nucleic bases, nucleosides, and nucleotides in solution, *Metal Ions Biol. Syst.* **8**, 57–124.

Ochiai, E. (1987) *General Principles of Biochemistry of the Elements*, Plenum Press, New York.

Seiler, H.G., Sigel, H. (1988) *Handbook on Toxicity of Inorganic Compounds*, Marcel Dekker, New York.

Sigel, H. (Ed.) (1973–2003) *Metal Ions in Biological Systems, Vols. 1–41*, Marcel Dekker, New York.

Sigel, H., Martin, R.B. (1982) Coordinating properties of the amide bond, *Chem. Rev.* **82**, 385–426.

Primary Literature

Martin, R.B. (2001) Peptide bond characteristics, *Metal Ions Biol. Syst.* **38**, 1–23.

Martin, R.B. (2002) Practical hardness scales for metal ion complexes, *Inorg. Chim. Acta* **339**, 27–33.

Biological Regulation by Protein Phosphorylation

Clay W. Scott[1], *Ratan V. Bhat*[2], *and Gaochao Tian*[1]
[1] *Lead Discovery Department, AstraZeneca Pharmaceuticals LP, Wilmington, DE, USA*
[2] *Department of Bioscience, AstraZeneca R&D Södertälje, Huddinge, Sweden*

1 Importance of Protein Phosphorylation in Biological Regulation 660

2 Classification and Properties of Protein Kinases 661

3 Classification and Properties of Phosphoprotein Phosphatases 665

4 Phosphorylation Cascades and Multisite Phosphorylations 667

5 Genetic and Pharmacological Modulation of Protein Phosphorylation 668

Bibliography 671
Books and Reviews 671
Primary Literature 671

Keywords

Protein Kinase
An enzyme that catalyzes the transfer of a phosphate group from a nucleoside triphosphate (usually ATP) to an amino acid side chain of a substrate protein.

Phosphoprotein Phosphatase
An enzyme that removes phosphate groups from the amino acid side chains of phosphorylated proteins.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Second Messenger
An intracellular molecule, generated by the cell in response to an extracellular signal (first messenger), which triggers a biochemical cascade leading to a change in the function or phenotype of the cell.

The biological activity of many proteins is modulated by the phosphorylation of specific sites within the protein. Protein kinases catalyze the transfer of a phosphate group to the protein, while protein phosphatases remove the phosphate group. This reversible modification is utilized by the cell to dynamically regulate proteins that are involved in almost all cellular functions. Protein phosphorylation plays a particularly prominent role in the transduction of extracellular signals. A hormone, neurotransmitter, or growth factor binds to its cell surface receptor and may directly or indirectly activate a protein kinase (or phosphatase) inside the cell, which leads to changes in the phosphorylation states of key regulatory proteins. This cascade of events ultimately results in a functional or phenotypic cellular response (e.g. contraction, release of neurotransmitter, changes in metabolism, or altered gene expression).

1
Importance of Protein Phosphorylation in Biological Regulation

Protein phosphorylation is one of the most common posttranslational modifications to occur in eukaryotic cells. It has been estimated that ~30% of proteins encoded by the human genome are phosphorylated at one or more sites. Phosphorylation and dephosphorylation is a fundamental molecular switch used by cells to control the activity of a variety of cellular proteins, which in turn regulates a vast array of cellular functions. The types of proteins known to undergo reversible phosphorylation include metabolic enzymes, cytoskeletal proteins, transcription factors, ion channels, and cell surface receptors. Protein phosphorylation is particularly prominent in signal-transduction processes. A cell can rapidly respond to a biological signal by phosphorylating key intracellular molecules thereby initiating a cascade of events resulting in a change in a physiological property such as cell motility, release of neurotransmitters, modulation of ion fluxes or alteration in gene expression. Because of the prominent role of this reversible and tightly coupled protein-modifying process, dysregulated protein phosphorylation can contribute to the pathogenesis of human disease.

The phosphorylation state of a particular protein is controlled by specific protein kinases and phosphoprotein phosphatases (Fig. 1). Protein kinases are enzymes that transfer the γ-phosphate from a nucleoside triphosphate (usually ATP) to an amino acid side chain of the substrate protein. Structural studies have shown that, in the case of enzymes, phosphorylation can occur within the active site and directly affect substrate binding. Phosphorylation

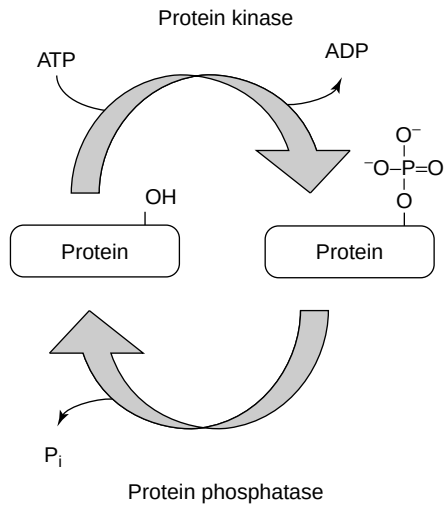

Fig. 1 Protein kinases catalyze the transfer of a phosphate group onto a specific amino acid side chain of a substrate protein. ATP usually serves as the phosphate donor, and serine, threonine, and tyrosine residues serve as phosphate acceptors. The reversible phosphorylation of a protein causes a change in its biological activity. Phosphoprotein phosphatases remove the phosphate group allowing the protein to return to its former functional state. The relative activities of the protein kinase and protein phosphatase generally determine the phosphorylation state of the protein. (Modified from Scott and Patel, Encyclopedia of Human Biology (1991) 6; 201–211; reproduced by permission of Academic Press.)

can also occur at sites distal to the catalytic site and regulate enzyme function by inducing long-range conformational changes. Phosphorylation has been shown to affect the biological properties of proteins by altering their intracellular location, enhancing their susceptibility to proteolysis, and modulating their ability to interact with other proteins. Phosphoprotein phosphatases remove the phosphate group, thereby allowing the protein to return to its previous functional state. Some proteins become active when dephosphorylated and are inactivated by phosphorylation. An example is glycogen synthase, an enzyme that catalyzes the conversion of glucose to glycogen. This enzyme is active in the dephosphorylated form. When phosphorylated by glycogen synthase kinase 3 (GSK3), glycogen synthase becomes inactive thereby limiting the synthesis of glycogen. In general, both protein kinases and phosphoprotein phosphatases are under stringent regulatory control. The activation state of the relevant kinase and phosphatase will usually determine when a protein becomes phosphorylated and how long it remains phosphorylated.

2 Classification and Properties of Protein Kinases

Protein kinases share a highly conserved catalytic domain called the *kinase domain* (also known as the *catalytic core*). The high conservation of this domain has helped identify protein kinase genes from the various large-scale genome sequencing efforts of human, yeast (*Saccharomyces cerevisiae*), fly (*Drosophila melanogaster*), and worm (*Caenorhabditis elegans*). Over 500 protein kinase genes have been identified in the human genome, representing about 2% of the human genes. The kinase domain is the third most abundant domain encoded in the human genome, with the immunoglobulin and zinc finger domains being the most prevalent.

More than 170 crystal structures of protein kinases have been resolved, providing an insight into the structure and function of these enzymes. The kinase domain contains two major structural domains named N- and C-terminal domains (Fig. 2). These two domains are bridged through a short linker peptide, around which they rotate in

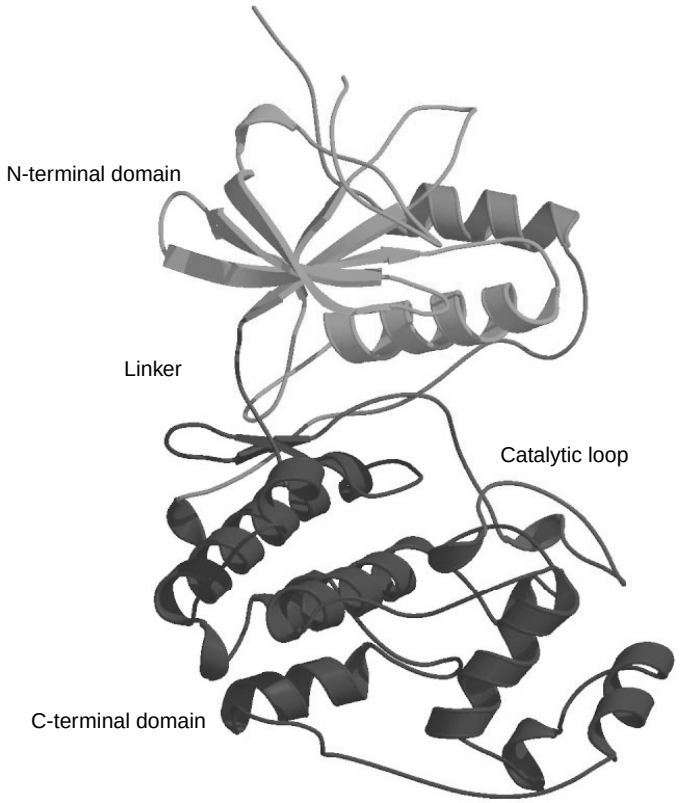

Fig. 2 Ribbon diagram of the kinase domain of ERK2 (p42 MAPK), a protein serine–threonine kinase. The four prototypic subdomains are labeled and color-coded.

response to protein substrate and/or ATP-binding. Also located in the C-terminal domain is a short peptide strand called the *catalytic loop*, which contains an invariant aspartate that is critical for catalysis. The ATP binds in a deep cleft between the N- and C-terminal domains that contain a part of the linker region and the catalytic loop. The protein substrate binding site is composed of shallow surface depressions, whose structural features vary among different protein kinases and thereby influence substrate specificity. The shallow surface depressions do not provide adequate interaction sites for high affinity binding of small molecules. Thus, the vast majority of small-molecule inhibitors identified so far bind to the ATP binding site.

Protein kinases recognize phosphorylation sites within particular amino acid sequences called *consensus sequences* or *recognition motifs*. Studies using peptide substrates containing altered amino acid sequences have revealed the importance of primary sequence in distinguishing phosphorylation sites. For example, cAMP-dependent protein kinase will phosphorylate a serine or threonine residue in the sequence -Arg-Arg-*X*-Ser/Thr-*X*-, where *X* represents any amino acid. Eliminating

or displacing either arginine residue can dramatically alter the kinetics of phosphorylation. Although amino acid sequence is an essential determinant for substrate recognition, higher orders of structure can also affect phosphorylation. Conformational states that mask or expose phosphorylation sites can alter the phosphorylation of the substrate protein. Many protein kinases can provide additional substrate specificity by engaging substrate with domains outside the catalytic core, such as SH2 (src homology 2) domains.

Protein kinases are classified according to the amino acid residue they phosphorylate. Greater than 98% of phosphorylation events in eukaryotic cells occur on serine and threonine residues (by protein–serine/threonine kinases); the remainder occur on tyrosine residues (by protein tyrosine kinases). A few protein kinases appear to phosphorylate tyrosine as well as serine/threonine residues and are classified as dual specificity protein kinases. Tyrosine kinases (TK) are found only in metazoans, whereas serine/threonine kinases are conserved throughout the eukaryotes. The lipid kinase PI3K possesses protein–serine/threonine kinase activities, in addition to phosphorylating lipid phosphoinositides. Histidine–aspartate kinases have been identified in bacteria, yeast, and plants but are not known to exist in mammals.

By aligning the kinase domain amino acid sequences for all the protein kinases, one can develop a sequence-based hierarchy of groups, families, and subfamilies (Fig. 3). Serine/threonine kinases are divided into 5 groups and further divided into 89 families. Tyrosine kinases, which represent a small fraction (~15% in humans) of the kinase superfamily, constitute a single group (TK) of 30 families. This phylogenetic classification of protein kinases correlates closely with general themes of protein kinase function and regulation. For example, the TK group is composed of tyrosine kinases only, which are widely recognized for their roles in controlling cell growth and differentiation. The group of Ca^{2+}/calmodulin-dependent protein kinases (CAMK) tend to phosphorylate serine/threonine residues located near basic residues. Many protein kinases in this group are activated by Ca^{2+}/calmodulin binding to a small domain located C-terminal to the catalytic domain.

Tyrosine kinases were identified first as the products of viral oncogenes. These viral protein tyrosine kinases, which we now know are structural variants of normal cytoplasmic protein tyrosine kinases, have unregulated tyrosine kinase activity and thereby induce the malignant transformation of cells. Their normal cellular homologs are under stringent regulatory control and play critical roles in cell growth and metabolism. Another important example of cytosolic nonreceptor tyrosine kinases is the Janus (JAK) kinase family. The JAKs are involved in signal transduction of the cytokine receptor superfamily. JAKs bind to cytoplasmic domains on cytokine receptors and become activated when the receptor binds its ligand. The activated JAK phosphorylates the receptor and intracellular proteins such as STATs. Once phosphorylated, STATs translocate to the nucleus where they regulate gene expression.

Some tyrosine kinases are cell surface receptors that contain an intrinsic tyrosine–protein kinase within their cytoplasmic domains. These receptor tyrosine kinases constitute a significant portion of the TK group. Of the 90 protein tyrosine kinases found in humans, 58 are

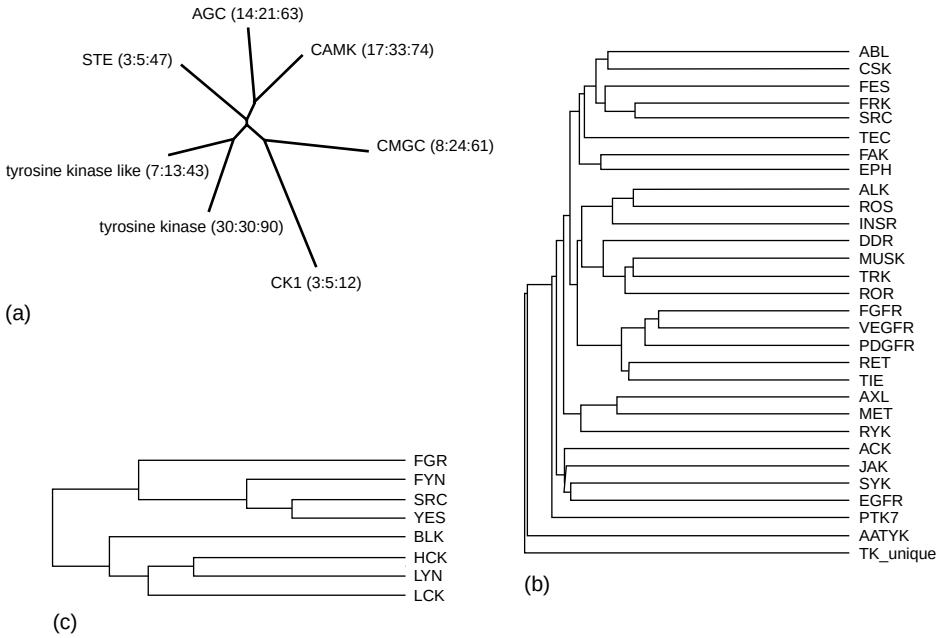

Fig. 3 Dendrogram or phylogenetic tree of protein kinases. (a) A circular tree representing the evolutionary divergence of the eukaryotic protein kinase groups based on the homology within the kinase domains. Each group can be further divided into families and in some instances into subfamilies. For example, the AGC group includes the cyclic nucleotide-regulated protein kinase family, which can be divided into the cAMP-dependent and cGMP-dependent protein kinase subfamilies. The numbers in parentheses in (a) represent the number of families, subfamilies, and human kinases in each protein kinase group. (b) The tyrosine kinase group can be divided into 30 families, as depicted in the flat tree. (c) The src family of tyrosine kinases contains 8 human tyrosine kinases and their evolutionary divergence is depicted. A comprehensive classification scheme for each protein kinase group is available at www.kinase.com.

the receptor type. Examples of receptor tyrosine kinases include the EGF receptor and the PDGF receptor. These receptors undergo dimerization following ligand binding. Receptor dimerization leads to activation of the receptor tyrosine kinase and autophosphorylation of the receptor on multiple tyrosine residues. The phosphorylated tyrosine and surrounding amino acids serve as binding sites for secondary signal transducing proteins containing SH2 domains, for example, Src, phospholipase C-γ, phosphatidylinositol 3′ kinase, and *ras* GTPase-activating protein (GAP). Specificity of binding is generated by the sequence context surrounding the individual phosphotyrosines. These newly bound proteins link the receptor with a variety of different signal-transduction pathways either directly due to the activation of their catalytic activity or indirectly by serving as an adaptor protein to couple other signal-transduction molecules.

Eph receptors, the largest family of receptor tyrosine kinases, are unique in that their ligands, ephrinAs and ephrinBs, are cell surface proteins. The interaction of ephrins with Eph receptors on opposing

cells induces receptor clustering and phosphorylation, and subsequent recruitment of intracellular signaling molecules. The binding of ephrin to the Eph receptor also initiates a functional response in the ligand-expressing cell, although the biochemical mechanism for this effect is not yet clear. Such reciprocal signal-transduction between Eph receptors and their ligands regulate cell–cell repulsion and adhesion mechanisms involved in the formation and maintenance of organ systems.

Several serine/threonine-receptor kinases exist. These include lectin receptor kinases, transforming growth factor β (TGF-β) receptors, and activin. TGF-β receptors (types I and II) control the development and homeostasis of most tissues in metazoan organisms. To exert their signal, type II and type I receptors act in sequence. TGF-β first binds to the type II receptor, which contains intrinsic kinase activity. The type I receptor is then recruited and phosphorylated in its intracellular domain by the type II receptor, leading to activation of its kinase activity. Similar to JAK-STAT signaling, activated type I TGF-β receptor recruits and phosphorylates Smad proteins, which then translocate into the nucleus where they regulate gene transcription.

Mechanistically, many protein kinases are regulated through autoinhibition. This involves a pseudosubstrate domain, found either in the primary sequence of the kinase itself or on a separate regulatory protein that interacts with the active site. Activation is achieved when the autoinhibitory domain is displaced from the active site. Protein serine/threonine kinases activated by second messengers, intracellular molecules that are generated in response to extracellular signals (Fig. 4), are kinases regulated by autoinhibition. For these enzymes, binding of the second messenger induces a conformational change that releases the inhibitory domain. These protein kinases are classified according to the second messenger that stimulates their activity. These groups include protein kinases activated by cyclic nucleotides (cAMP- and cGMP-dependent protein kinases), calcium plus calmodulin (including Ca^{2+}/calmodulin-dependent protein kinase II and myosin light-chain kinase), and diacylglycerol (protein kinase C). In the protein tyrosine kinase Src, activation arises from phosphorylation-induced conformational changes, resulting in displacing part of the C-terminal tail that otherwise occupies the active site and prevents substrate from binding to enzyme.

Activation of many protein kinases, but not all, require phosphorylation of either a conserved threonine or tyrosine residue in a loop structure called the activation loop (Fig. 2). These protein kinases have low basal activities as a result of the unphosphorylated form possessing low affinity for ATP and incorrectly positioned catalytic residues. Phosphorylation induces a structural change in the activation loop, which results in remodeling of the active site leading to optimized substrate binding and catalysis.

3 Classification and Properties of Phosphoprotein Phosphatases

Traditionally, phosphoprotein phosphatases were categorized by their selectivity for phosphoserine/phosphothreonine or phosphotyrosine residues and further separated on the basis of sensitivities to endogenous inhibitors, requirements for cations, and *in vitro* substrate specificity. More recent gene sequencing efforts

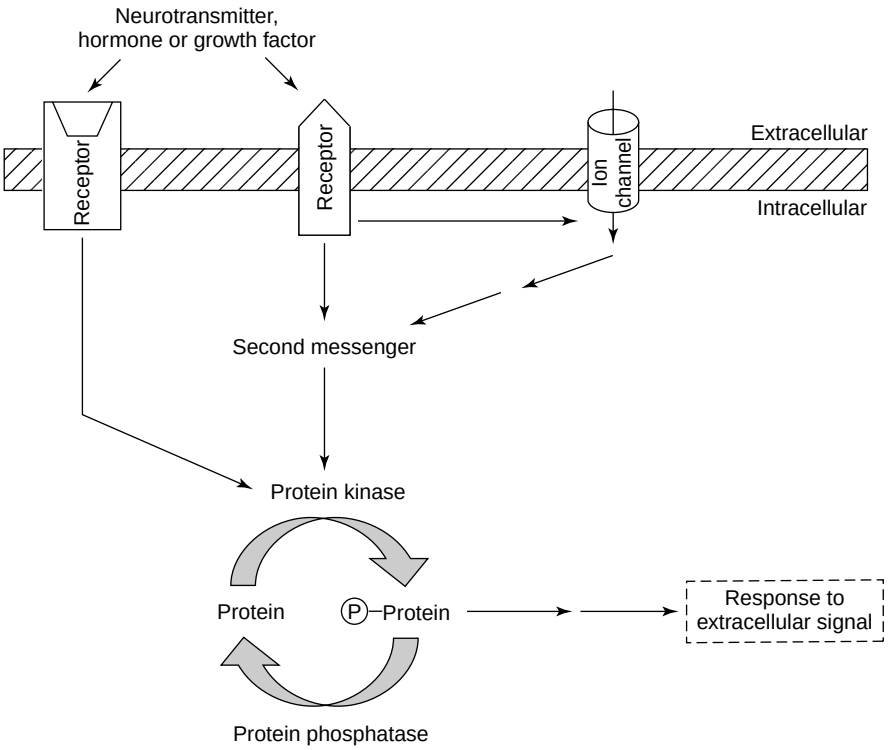

Fig. 4 Protein phosphorylation is a common mechanism used to transduce extracellular stimuli into intracellular responses. Extracellular molecules bind to their cell surface receptors and activate protein kinases or phosphatases. Some kinases and phosphatases are intrinsic components of receptors and are directly activated by ligand binding. Others are activated, either directly or indirectly, by changes in the levels of second messengers such as cAMP, cGMP, diacylglycerol, and Ca^{2+}. The activity of some kinases and phosphatases are regulated by other kinases and phosphatases as part of a phosphorylation cascade. (Modified from Scott and Patel, Encyclopedia of Human Biology (1991) 6; 201–211; reproduced by permission of Academic Press.)

have revealed three distinct superfamilies of phosphatases; two families of protein serine/threonine phosphatases (PPP and PPM families) and one protein tyrosine phosphatase family (PTP). Within the PTP family is a subgroup of enzymes that dephosphorylate both tyrosine and serine/threonine residues: this group is called the *dual-specificity phosphatases*.

Thirty-two protein-serine/threonine phosphatase genes have been identified in the human genome sequence. The enzymes encoded by these genes are responsible for dephosphorylating proteins that are acted upon by over 500 serine/threonine kinases. Thus, it is very likely that most serine/threonine phosphatases will have multiple substrates. In most cases, the catalytic subunits of these phosphatases are capable of binding to numerous regulatory and targeting subunits. These auxiliary subunits are expressed in cell-type specific patterns and can be located within different regions of the cell. They regulate

the biological activity and subcellular location of the phosphatase and thereby limit access to potential substrates. Thus, one serine/threonine phosphatase can be utilized within the cell to regulate multiple cellular events.

The PPP family of serine/threonine phosphatases has the most members and is responsible for more than 90% of the serine/threonine phosphatase activity in mammalian cells. This family includes some of the earliest and most extensively studied enzymes including PP1, PP2A, and PP2B (also called calcineurin). The principal members of the PPM family include PP2C and pyruvate dehydrogenase phosphatase. Although these enzymes are unrelated to the PPP family in terms of protein sequence, they share structural features within their active sites and have a common catalytic mechanism.

Compared with the serine/threonine phosphatases, the tyrosine phosphatases have different structural features within their catalytic domains and use a different molecular mechanism for catalysis. The tyrosine phosphatases share a common amino acid sequence within their active site termed the *PTP signature motif*. Although the PTPs are responsible for a very small fraction of the overall cellular phosphatase activity, data from the human genome project predict over one hundred human enzymes. These phosphatases can be divided into three subfamilies: tyrosine-specific, dual-specific, and low-molecular weight phosphatases.

Some of the tyrosine-specific phosphatases have transmembrane receptor-like structures with an intracellular catalytic domain, suggesting the existence of receptor-linked proteins that utilize the dephosphorylation of tyrosine residues for signal transduction. Other tyrosine-specific phosphatases are intracellular enzymes that contain SH2 domains allowing them to interact with either receptor or cytosolic protein tyrosine kinases or localize to different subcellular regions where they function. The dual-specificity phosphatases include the Cdc25 and mitogen-activated protein kinase (MAPK) phosphatase families. The Cdc25 phosphatases are highly specific for adjacent phosphothreonine and phosphotyrosine residues on cyclin-dependent kinases. The regulated dephosphorylation of these sites plays a crucial role in eukaryotic cell cycle progression. The MAPK phosphatases regulate signal transduction pathways that are activated by mitogens and stress and which control cell proliferation, differentiation, and death (apoptosis). Much has been learned about the structure of PTPases, but many questions on their function and physiological substrates remain. This is especially true for the low molecular weight tyrosine phosphatases, a family of enzymes whose biological function(s) remain undefined.

4 Phosphorylation Cascades and Multisite Phosphorylations

Protein kinase cascades represent a common underlying theme in many signal-transduction pathways. A classical example is the regulation of glycogenolysis in skeletal muscle. Epinephrine binds to the β-adrenergic receptor, causing activation of adenylate cyclase, an increase in cAMP production, and an activation of cAMP-dependent protein kinase. This enzyme phosphorylates and activates phosphorylase kinase, which in turn phosphorylates and activates glycogen phosphorylase. In addition, cAMP-dependent protein kinase

phosphorylates and inactivates phosphoprotein phosphatase-1. Because a single protein kinase molecule can phosphorylate many substrate molecules, each step in this cascade amplifies the signaling pathway. In addition, some protein kinases in these cascades can phosphorylate many different substrate proteins, thereby modulating multiple biochemical pathways, which lead to an amplification of a signal. This is most prominent in the MAPK cascade. In response to an extracellular stimulus, several stress-activated protein kinases also converge onto key protein kinases, resulting in the amplification of a cell stress signal. Thus, protein kinase cascades contain sequential steps to amplify an extracellular signal as well as divergent steps to coordinate different cellular processes.

Many proteins appear to be phosphorylated at multiple sites with varying consequences on their biological activity. Examples of multisite phosphorylation have revealed that (1) different protein kinases can phosphorylate a common site on a protein; (2) different protein kinases can phosphorylate distinct sites on a protein yet produce similar changes in biological activity; (3) different protein kinases can phosphorylate distinct sites on a protein resulting in opposite changes in biological activity; and (4) phosphorylation of one site by a protein kinase can generate a consensus sequence for a second protein kinase. Proteins that are phosphorylated at multiple sites often function at critical points in a biochemical cascade. Thus, multisite phosphorylation represents one approach by which different biological signals, acting through different transduction pathways, can converge to regulate the function of a single protein. This convergence would allow for coordinated regulation of a single cellular process. Reversible phosphorylation of proteins represents a unifying and dynamic mechanism by which different signaling systems can be integrated to produce the appropriate biological response.

5 Genetic and Pharmacological Modulation of Protein Phosphorylation

In 1979, the following four criteria for establishing that an enzyme undergoes physiologically significant phosphorylation were defined:

1. The enzyme is phosphorylated at stoichiometric levels at a significant rate by a protein kinase and is dephosphorylated by a protein phosphatase.
2. There is a correlation between the degree of phosphorylation of an enzyme and changes in its functional properties.
3. *In vivo* studies or studies with intact cells show that the enzyme can be phosphorylated and dephosphorylated with subsequent changes in its functional properties.
4. The cellular levels of protein kinase and phosphatase effectors correlate with the extent of phosphorylation of the enzyme.

These criteria can be extended to all other classes of phosphoproteins such as receptors, ion channels, and proteins that form the cytoskeleton and proteins that regulate transcription and translation. Of these four criteria, the third element is often difficult to demonstrate. Selective activators or inhibitors of a specific protein kinase or phosphatase are often used

in an attempt to modify the phosphorylation state of a particular protein in intact cells and to test for changes in its biological activity. By incubating cells with a molecule that selectively activates or inhibits a particular kinase or phosphatase, one can study the downstream events that develop as a consequence of the pharmacological manipulation. The number of pharmacological agents that target protein kinases and phosphatases have increased significantly in recent years, and have been used to provide a wealth of information linking a particular enzyme with a particular biological response. Agents that modify protein phosphorylation can be classified into a number of different categories including activators or inhibitors, and natural or synthetic molecules. The synthetic molecules can be further separated into small molecules, peptides/proteins, and nucleic acids (e.g. antisense oligonucleotides and small interfering RNAs to regulate expression of an enzyme). Regardless of the classification scheme, these agents have great value when trying to identify the physiological substrates for a particular kinase or phosphatase, and to dissect the role of a particular phosphoprotein in cell function.

Several synthetic analogs of cAMP have been developed and shown to activate cAMP-dependent protein kinase in a manner analogous to cAMP. Some of these compounds are membrane-permeable and therefore can activate the enzyme when applied to intact cells. Because these compounds effectively mimic the actions of extracellular signals on the cAMP signal-transduction pathway, these cAMP analogs are useful to determine if a cellular response to a particular hormone, growth factor or neurotransmitter is mediated through the cAMP pathway.

Phorbol esters are a class of small organic molecules first identified in croton oil. These compounds mimic the actions of diacylglycerol and activate members of the protein kinase C family. Both diacylglycerol and inositol trisphosphate are generated by the hydrolysis of inositol phospholipids, which occurs in response to a number of different extracellular signals. Diacylglycerol activates the protein kinase C signal-transduction pathway while inositol trisphosphate stimulates the release of calcium from intracellular stores. Because the phorbol esters are potent and selective activators of protein kinase C in intact cells, they have become invaluable tools for investigating the involvement of protein kinase C isoenzymes in various biochemical pathways and to dissociate responses mediated via the diacylglycerol signal-transduction pathway from the inositol trisphosphate pathway. However, caution must be used when interpreting the results from such experiments, for recent studies have revealed additional targets for diacylglycerol and phorbol esters that can regulate cell function independent of protein kinase C.

Several natural toxins have been identified that exert their biological effects through direct and potent inhibition of the phosphatases PP1 and PP2A. These toxins are found in biologically diverse organisms. Okadaic acid, the first inhibitor identified within this class, is produced by several species of marine plankton. This plankton can accumulate in shellfish and, when ingested by humans, cause an illness termed diarrhetic seafood poisoning. Cantharidin is another interesting example, for it has a rich history of medicinal and recreational use prior to understanding its mechanism of action. Cantharidin is produced by the blister beetle as a defense mechanism against predators. The

dried bodies of these beetles have been used in China for over 2000 years as a traditional medicine for several indications including cancer and piles. Hippocrates also described its use for the treatment of dropsy. It has been exploited (incorrectly) as an aphrodisiac named Spanish Fly because it is an irritant of the urogenital tract, and causes pelvic engorgement in women. More recently, dermatologists have used it as a topical agent for the treatment of warts. Because cantharidin has a relatively simple molecular structure, a long history of apparent medicinal value and a newly discovered molecular mechanism of action, drug hunters have begun to synthesize derivatives of cantharidin in search of new medicines with greater specificity for a subset of serine/threonine phosphatases, thereby reducing side effects that occur with nonselective inhibitors.

In recent years, the development of genetically modified organisms has provided important tools for research. These organisms have been particularly helpful in evaluating whether phosphorylation by a particular kinase affects a protein's function *in vivo*. New molecular and genetic techniques have been developed to create animals expressing kinases that are driven by cell-specific or developmentally regulated promoters. Knockout mice have been generated carrying targeted mutations that prevent the expression of a functional enzyme. Animals derived from such work have greatly enhanced our understanding of protein phosphorylation pathways involved in pathological conditions such as diabetes, Alzheimer's disease, and cancer.

Modulators of protein phosphorylation are useful not only to dissect and understand molecular mechanisms but, as noted above for cantharidin, may also have therapeutic value. For example, FK506 and Cyclosporin A are drugs that produce their pharmacological effects by inhibiting a serine/threonine phosphatase in T-lymphocytes and preventing the dephosphorylation of a key cellular protein. They are used therapeutically as immunosuppressive drugs and are prescribed to organ-transplant patients. FK506 and Cyclosporin A bind to their respective intracellular binding proteins, FKBP12 and cyclophilin A. Each of these complexes can bind to PP2B and block its activation. Normally, T-cell activation causes an increase in intracellular calcium, which stimulates PP2B. Active PP2B dephosphorylates the transcription factor NF-AT thereby facilitating its translocation to the nucleus to initiate gene transcription. FK506 and Cyclosporin A block the dephosphorylation of NF-AT, preventing transcription of specific genes in T-lymphocytes that induce transplant rejection.

Current drugs for cancer therapy include cytotoxic agents that target basic cellular processes such as DNA replication or microtubule dynamics. Drug discovery strategies aimed at inhibiting specific protein kinases have led to the discovery of compounds that, in theory, should be effective and more tolerable anticancer therapies. Several dozen compounds are now in clinical trials for a wide range of cancers. Two protein kinase inhibitors have demonstrated remarkable clinical value and have been approved by the FDA: Gleevec, a Bcr-Abl kinase inhibitor for the treatment of chronic myelogenous leukemia, and Iressa, an epidermal growth factor receptor kinase inhibitor for the treatment of advanced non-small-cell lung cancer.

The future holds much promise for our increased understanding of the structure and function of protein kinases and phosphatases. A combination of gene mutation studies, X-ray crystallography and

NMR structural structures are helping to reveal the molecular basis of enzymatic activity and substrate specificity. In conjunction with molecular modeling studies, this structural information can aid in the rational design of more potent and selective enzyme inhibitors. These compounds are helping to define further the role of particular kinases and phosphatases in signal transduction and in the evaluation of their suitability as targets for therapeutic intervention in disease. The next few years represent an exciting time in the field of protein phosphorylation in terms of seeing how understanding fundamental biology is enabling the development of new medicines for the treatment of human diseases.

See also Bioorganic Chemistry.

Bibliography

Books and Reviews

Cohen, P. (2002) The origins of protein phosphorylation, *Nat. Cell Biol.* **4**, E127–E130.

Hunter, T. (1995) Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signalling, *Cell* **80**, 225–236.

Johnson, L.N., Lewis, R.J. (2001) Structural basis for control by phosphorylation, *Chem. Rev.* **101**, 2209–2242.

Kennelly, P. (2001) Protein phosphatases – a phylogenetic perspective, *Chem. Rev.* **101**, 2291–2312.

Krebs, E.G., Beavo, J.A. (1979) Phosphorylation-dephosphorylation of enzymes, *Annu. Rev. Biochem.* **48**, 923–959.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S. (2002) The protein kinase complement of the human genome, *Science* **298**, 1912–1934.

McCluskey, A., Sim, A.T.R., Sakoff, J.A. (2002) Serine-threonine protein phosphatase inhibitors: development of potential therapeutic strategies, *J. Med. Chem.* **45**, 1151–1175.

Scapin, G. (2002) Structural biology in drug design: selective protein kinase inhibitors, *Drug Discovery Today* **7**, 610–611.

Primary Literature

Ablooglu, A.J., Frankel, M., Rusinova, E., Ross, J.B., Kohanski, R.A. (2001) Multiple activation loop conformations and their regulatory properties in the insulin receptor's kinase domain, *J. Biol. Chem.* **276**, 46933–46940.

Aggen, J.B., Humphrey, J.M., Gauss, C.M., Huang, H.B., Nairn, A.C., Chamberlin, A.R. (1999) The design, synthesis, and biological evaluation of analogues of the serine-threonine protein phosphatase 1 and 2A selective inhibitor microcystin LA: rational modifications imparting PP1 selectivity, *Bioorg. Med. Chem.* **7**, 543–564.

Anderson, D., Koch, C.A., Grey, L., Ellis, C., Moran, M.F., Pawson, T. (1990) Binding of SH2 domains of phospholipase C gamma 1, GAP, and Src to activated growth factor receptors, *Science* **250**, 979–982.

Barford, D., Flint, A.J., Tonks N.K. (1994) Crystal structure of human protein tyrosine phosphatase 1B, *Science* **263**, 1397–1404.

Bax, B., Carter, P.S., Lewis, C., Guy, A.R., Bridges, A., Tanner, R., Pettman, G., Mannix, C., Culbert, A.A., Brown, M.J., Smith, D.G., Reith, A.D. (2001) The structure of phosphorylated GSK-3beta complexed with a peptide, FRATtide, that inhibits beta-catenin phosphorylation, *Structure* **9**, 1143–1152.

Brambilla, R., Schnapp, A., Casagranda, F., Labrador, J.P., Bergemann, A.D., Flanagan, J.G., Pasquale, E.B., Klein, R. (1995) Membrane-bound LERK2 ligand can signal through three different Eph-related receptor tyrosine kinases, *EMBO J.* **14**, 3116–3126.

Brown, N.R., Noble, M.E., Lawrie, A.M., Morris, M.C., Tunnah, P., Divita, G., Johnson, L.N., Endicott, J.A. (1999) Effects of phosphorylation of threonine 160 on cyclin-dependent kinase 2 structure and activity, *J. Biol. Chem.* **274**, 8746–8756.

Bruckner, K., Pasquale, E.B., Klein, R. (1997) Tyrosine phosphorylation of transmembrane ligands for Eph receptors, *Science* **275**, 1640–1643.

Cheifetz, S., Weatherbee, J.A., Tsang, M.L., Anderson, J.K., Mole, J.E., Lucas, R., Massague, J. (1987) The transforming growth

factor-beta system, a complex pattern of cross-reactive ligands and receptors, *Cell* **48**, 409–415.

Collett, M.S., Erikson, R.L. (1978) Protein kinase activity associated with the avian sarcoma virus src gene product, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2021–2024.

Davis, S., Gale, N.W., Aldrich, T.H., Maisonpierre, P.C., Lhotak, V., Pawson, T., Goldfarb, M., Yancopoulos, G.D. (1994) Ligands for EPH-related receptor tyrosine kinases that require membrane attachment or clustering for activity, *Science* **266**, 816–819.

Dobrowsky, R.T., Kamibayashi, C., Mumby, M.C., Hannun, Y.A. (1993) Ceramide activates heterotrimeric protein phosphatase 2A, *J. Biol. Chem.* **268**, 15523–15530.

Dor, Y., Djonov, V., Abramovitch, R., Itin, A., Fishman, G.I., Carmeliet, P., Goelman, G., Keshet, E. (2002) Conditional switching of VEGF provides new insights into adult neovascularization and pro-angiogenic therapy, *EMBO J.* **21**, 1939–1947.

Egloff, M.P., Cohen, P.T., Reinemer, P., Barford, D. (1995) Crystal structure of the catalytic subunit of human protein phosphatase 1 and its complex with tungstate, *J. Mol. Biol.* **254**, 942–959.

Favelyukis, S., Till, J.H., Hubbard, S.R., Miller, W.T. (2001) Structure and autoregulation of the insulin-like growth factor 1 receptor kinase, *Nat. Struct. Biol.* **8**, 1058–1063.

Fienberg, A.A., Hiroi, N., Mermelstein, P.G., Song, W., Snyder, G.L., Nishi, A., Cheramy, A., O'Callaghan, J.P., Miller, D.B., Cole, D.G., Corbett, R., Haile, C.N., Cooper, D.C., Onn, S.P., Grace, A.A., Ouimet, C.C., White, F.J., Hyman, S.E., Surmeier, D.J., Girault, J., Nestler, E.J., Greengard, P. (1998) DARPP-32: regulator of the efficacy of dopaminergic neurotransmission, *Science* **281**, 838–842.

Fischer, E.H., Krebs, E.G. (1955) Conversion of phosphorylase b to phosphorylase a in muscle extracts, *J. Biol. Chem.* **216**, 121–132.

Gille, H., Kowalski, J., Li, B., LeCouter, J., Moffat, B., Zioncheck, T.F., Pelletier, N., Ferrara, N. (2001) Analysis of biological effects and signaling properties of Flt-1 (VEGFR-1) and KDR (VEGFR-2). A reassessment using novel receptor-specific vascular endothelial growth factor mutants, *J. Biol. Chem.* **276**, 3222–3230.

Goldberg, J., Huang, H.B., Kwon, Y.G., Greengard, P., Nairn, A.C., Kuriyan, J. (1995) Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1, *Nature* **376**, 745–753.

Griffith, J.P., Kim, J.L., Kim, E.E., Sintchak, M.D., Thomson, J.A., Fitzgibbon, M.J., Fleming, M.A., Caron, P.R., Hsiao, K., Navia, M.A. (1995) X-ray structure of calcineurin inhibited by the immunophilin-immunosuppressant FKBP12-FK506 complex, *Cell* **82**, 507–522.

Hanks, S.K., Hunter, T. (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification, *FASEB J.* **9**, 576–596.

Hemmings, H.C., Jr., Greengard, P., Tung, H.Y., Cohen, P. (1984) DARPP-32, a dopamine-regulated neuronal phosphoprotein, is a potent inhibitor of protein phosphatase-1, *Nature* **310**, 503–505.

Holash, J., Maisonpierre, P.C., Compton, D., Boland, P., Alexander, C.R., Zagzag, D., Yancopoulos, G.D., Wiegand, S.J. (1999) Vessel cooption, regression, and growth in tumors mediated by angiopoietins and VEGF, *Science* **284**, 1994–1998.

Holash, J., Wiegand, S.J., Yancopoulos, G.D. (1999) New model of tumor angiogenesis: dynamic balance between vessel regression and growth mediated by angiopoietins and VEGF, *Oncogene* **18**, 5356–5362.

Holland, S.J., Gale, N.W., Mbamalu, G., Yancopoulos, G.D., Henkemeyer, M., Pawson, T. (1996) Bidirectional signalling through the EPH-family receptor Nuk and its transmembrane ligands, *Nature* **383**, 722–725.

Hunter, T., Sefton, B.M. (1980) Transforming gene product of Rous sarcoma virus phosphorylates tyrosine, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1311–1315.

Ingebritsen, T.S., Cohen, P. (1983) The protein phosphatases involved in cellular regulation. 1. Classification and substrate specificities, *Eur. J. Biochem.* **132**, 255–261.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921. [erratum appears in Nature 2001 Aug 2; 412(6846):565].

Jiang, Y., Chen, C., Li, Z., Guo, W., Gegner, J.A., Lin, S., Han, J. (1996) Characterization of the structure and function of a new mitogen-activated protein kinase (p38beta), *J. Biol. Chem.* **271**, 17920–17926.

Johnson, D.F., Moorhead, G., Caudwell, F.B., Cohen, P., Chen, Y.H., Chen, M.X., Cohen, P.T. (1996) Identification of protein-phosphatase-1-binding domains on the glycogen and myofibrillar targetting subunits, *Eur. J. Biochem.* **239**, 317–325.

Kissinger, C.R., Parge, H.E., Knighton, D.R., Lewis, C.T., Pelletier, L.A., Tempczyk, A., Kalish, V.J., Tucker, K.D., Showalter, R.E., Moomaw, E.W. (1995) Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex, *Nature* **378**, 641–644.

Klauck, T.M., Faux, M.C., Labudda, K., Langeberg, L.K., Jaken, S., Scott, J.D. (1996) Coordination of three signaling enzymes by AKAP79, a mammalian scaffold protein, *Science* **271**, 1589–1592.

Knighton, D.R., Zheng, J.H., Ten Eyck, L.F., Ashford, V.A., Xuong, N.H., Taylor, S.S., Sowadski, J.M. (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase, *Science* **253**, 407–414.

Krebs, E.G., Fischer, E.H. (1955) The phosphorylase b to a converting enzyme of rabbit skeletal muscle, *Biochim. Biophys. Acta* **20**, 150–157.

Li, M., Guo, H., Damuni, Z. (1995) Purification and characterization of two potent heat-stable protein inhibitors of protein phosphatase 2A from bovine kidney, *Biochemistry* **34**, 1988–1996.

Li, M., West, J.W., Lai, Y., Scheuer, T., Catterall, W.A. (1992) Functional modulation of brain sodium channels by cAMP-dependent phosphorylation, *Neuron* **8**, 1151–1159.

Lin, P., Buxton, J.A., Acheson, A., Radziejewski, C., Maisonpierre, P.C., Yancopoulos, G.D., Channon, K.M., Hale, L.P., Dewhirst, M.W., George, S.E., Peters, K.G. (1998) Antiangiogenic gene therapy targeting the endothelium-specific receptor tyrosine kinase Tie2, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8829–8834.

Liu, J., Farmer, J.D., Jr., Lane, W.S., Friedman, J., Weissman, I., Schreiber, S.L. (1991) Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes, *Cell* **66**, 807–815.

Millward, T., Cron, P., Hemmings, B.A. (1995) Molecular cloning and characterization of a conserved nuclear serine(threonine) protein kinase, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5022–5026.

Moasser, M.M., Basso, A., Averbuch, S.D., Rosen, N. (2001) The tyrosine kinase inhibitor ZD1839 ("Iressa") inhibits HER2-driven signaling and suppresses the growth of HER2-overexpressing tumor cells, *Cancer Res.* **61**, 7184–7188.

Pallas, D.C., Shahrik, L.K., Martin, B.L., Jaspers, S., Miller, T.B., Brautigan, D.L., Roberts, T.M. (1990) Polyoma small and middle T antigens and SV40 small t antigen form stable complexes with protein phosphatase 2A, *Cell* **60**, 167–176.

Pautsch, A., Zoephel, A., Ahorn, H., Spevak, W., Hauptmann, R., Nar, H. (2001) Crystal structure of bisphosphorylated IGF-1 receptor kinase: insight into domain movements upon kinase activation, *Structure* **9**, 955–965.

Petit, A.M., Rak, J., Hung, M.C., Rockwell, P., Goldstein, N., Fendly, B., Kerbel, R.S. (1997) Neutralizing antibodies against epidermal growth factor and ErbB-2/neu receptor tyrosine kinases down-regulate vascular endothelial growth factor production by tumor cells in vitro and in vivo: angiogenic implications for signal transduction therapy of solid tumors, *Am. J. Pathol.* **151**, 1523–1530.

Rall, T.W., Sutherland, E.W., Wosilait, W.D. (1956) The relationship of epinephrine and glucagon to liver phosphorylation, *J. Biol. Chem.* **218**, 483–495.

Ray, L.B., Sturgill, T.W. (1987) Rapid stimulation by insulin of a serine/threonine kinase in 3T3-L1 adipocytes that phosphorylates microtubule-associated protein 2 in vitro, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1502–1506.

Rudrabhatla, P., Rajasekharan, R. (2003) Mutational analysis of stress-responsive peanut dual specificity protein kinase. Identification of tyrosine residues involved in regulation of protein kinase activity, *J. Biol. Chem.* **278**, 17328–17335.

Sadowski, I., Stone, J.C., Pawson, T. (1986) A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps, *Mol. Cell. Biol.* **6**, 4396–4408.

Schindler, T., Sicheri, F., Pico, A., Gazit, A., Levitzki, A., Kuriyan, J. (1999) Crystal structure of Hck in complex with a Src family-selective tyrosine kinase inhibitor, *Mol. Cell.* **3**, 639–648.

Sicheri, F., Moarefi, I., Kuriyan, J. (1997) Crystal structure of the Src family tyrosine kinase Hck, *Nature* **385**, 602–609.

Stewart, A.A., Ingebritsen, T.S., Manalan, A., Klee, C.B., Cohen, P. (1982) Discovery of a Ca^{2+}- and calmodulin-dependent protein phosphatase: probable identity with calcineurin (CaM-BP80), *FEBS Lett.* **137**, 80–84.

Till, J.H., Ablooglu, A.J., Frankel, M., Bishop, S.M., Kohanski, R.A., Hubbard, S.R. (2001) Crystallographic and solution studies of an activation loop mutant of the insulin receptor tyrosine kinase: insights into kinase mechanism, *J. Biol. Chem.* **276**, 10049–10055.

Till, J.H., Annan, R.S., Carr, S.A., Miller, W.T. (1994) Use of synthetic peptide libraries and phosphopeptide-selective mass spectrometry to probe protein kinase substrate specificity, *J. Biol. Chem.* **269**, 7423–7428.

Tonks, N.K., Charbonneau, H., Diltz, C.D., Fischer, E.H., Walsh, K.A. (1988) Demonstration that the leukocyte common antigen CD45 is a protein tyrosine phosphatase, *Biochemistry* **27**, 8695–8701.

Tonks, N.K., Diltz, C.D., Fischer, E.H. (1988) Purification of the major protein-tyrosine-phosphatases of human placenta, *J. Biol. Chem.* **263**, 6722–6730.

Turowski, P., Favre, B., Campbell, K.S., Lamb, N.J., Hemmings, B.A. (1997) Modulation of the enzymatic properties of protein phosphatase 2A catalytic subunit by the recombinant 65-kDa regulatory subunit PR65alpha, *Eur. J. Biochem.* **248**, 200–208.

Waltenberger, J., Claesson-Welsh, L., Siegbahn, A., Shibuya, M., Heldin, C.H. (1994) Different signal transduction properties of KDR and Flt1, two receptors for vascular endothelial growth factor, *J. Biol. Chem.* **269**, 26988–26995.

Walter, G., Ruediger, R., Slaughter, C., Mumby, M. (1990) Association of protein phosphatase 2A with polyoma virus medium tumor antigen, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2521–2525.

Wilks, A.F., Harpur, A.G., Kurban, R.R., Ralph, S.J., Zurcher, G., Ziemiecki, A. (1991) Two novel protein-tyrosine kinases, each with a second phosphotransferase-related catalytic domain, define a new class of protein kinase, *Mol. Cell. Biol.* **11**, 2057–2065.

Xie, X., Gu, Y., Fox, T., Coll, J.T., Fleming, M.A., Markland, W., Caron, P.R., Wilson, K.P., Su, M.S. (1998) Crystal structure of JNK3: a kinase implicated in neuronal apoptosis, *Structure* **6**, 983–991.

Xu, W., Harrison, S.C., Eck, M.J. (1997) Three-dimensional structure of the tyrosine kinase c-Src, *Nature* **385**, 595–602.

Zhao, Y., Boguslawski, G., Zitomer, R.S., DePaoli-Roach, A.A. (1997) Saccharomyces cerevisiae homologs of mammalian B and B′ subunits of protein phosphatase 2A direct the enzyme to distinct cellular functions, *J. Biol. Chem.* **272**, 8256–8262.

Zwijsen, A., van Grunsven, L.A., Bosman, E.A., Collart, C., Nelles, L., Umans, L., Van de, P.T., Wuytens, G., Huylebroeck, D., Verschueren, K. (2001) Transforming growth factor beta signalling in vitro and in vivo: activin ligand-receptor interaction, Smad5 in vasculogenesis, and repression of target genes by the deltaEF1/ZEB-related SIP1 in the vertebrate embryo, *Mol. Cell. Endocrinol.* **180**, 13–24.

Bioluminescence and Chemiluminescence: *see* Chemiluminescence and Bioluminescence, Analysis by

Bioorganic Chemistry

Simon H. Friedman
University of Missouri, Kansas City, MO, USA

1 Introduction 2

2 General Principles 3
2.1 Covalent 3
2.2 Noncovalent Interactions 6
2.2.1 Electrostatic 8
2.2.2 Hydrogen Bonding 9
2.2.3 Hydrophobic Effect 10
2.2.4 van der Waals' Interactions 10
2.2.5 Conformational/Configurational Aspects 10

3 Nucleotide Interactions 11
3.1 Oligonucleotides 11
3.2 Small Molecules 13

4 Protein Studies 16
4.1 Enzyme Inhibition 16
4.2 Enzyme Catalysis 18
4.3 Transition-state Analogs and Catalytic Antibodies 20
4.4 Solid-phase Peptide Synthesis 22

5 Combinatorial Chemistry 23

6 Molecular Modeling 27

7 Summary 30

Bibliography 30
Books and Reviews 30
Primary Literature 30

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Keywords

Solid-phase Oligonucleotide Synthesis
The primary technique used for chemical synthesis of oligonucleotides via sequential addition of monomers to a growing, resin-bound chain.

Phosphorothioate-containing DNA
A nuclease-resistant form of DNA in which one of the two unesterified phosphate oxygens is replaced with a sulfur atom.

Enzyme Inhibitors
Molecules that are able to prevent enzyme catalytic action.

Catalytic Antibodies
Antibodies that have been generated to catalyze specific chemical reactions.

Transition-state Analogs
Molecules that closely resemble the shape and charge of a reaction's transition state (more accurately known as the activated complex).

Combinatorial Chemistry
Method of synthesis of large numbers of molecules by the reaction of sets of monomer units at multiple positions.

Bioorganic chemistry is the study of biological systems, utilizing the tools of the organic chemist, usually including a synthetic component.

1 Introduction

Bioorganic chemistry is the study of biological systems utilizing the tools of the organic chemist, usually including a synthetic component. Bioorganic chemists pursue a wide range of activities including the study of biomechanisms, enzyme models, biosynthesis, biomimetic synthesis, molecular recognition, enzymology, peptide chemistry, nucleic acid chemistry, immunology, and the design and synthesis of therapeutic agents. Within the past decade, the classical distinctions among scientific disciplines have gradually faded so that frequently one finds the synthetic organic chemist making compounds of biological interest and, conversely, the biochemist using synthetic techniques to aid in answering biological questions. In this article, we discuss some highlights of bioorganic chemistry ranging from the more organic to the more biochemical. Areas of greatest interest to molecular biologists are emphasized.

Before delving into specific topics that relate bioorganic chemistry to molecular biology, we will explore in a more general way the physical and chemical principles that govern these topics and, indeed, govern all systems that one can observe in

nature. An understanding of these principles will allow a deeper understanding of all of biochemistry. The principles will be divided into two subareas: covalent and noncovalent. The covalent section will explore some of the fundamental organic reactions that one sees in much of biochemistry. The noncovalent section will examine the forces that govern the association of molecules.

2 General Principles

2.1 Covalent

There are literally hundreds of specific chemical reactions that are found in biological systems, and a study of the breadth of these is obviously outside the scope of this review. This huge number of reactions can be ordered by a variety of reaction types. One of the most prevalently observed reactions in biochemistry is hydrolysis and its reverse reaction, which is considered a type of condensation. These two general classes of reaction account for such processes as DNA/RNA polymerization and degradation, protein synthesis and degradation, muscle action, regulation via phosphorylation, and so on. Despite the variety of the reactants and products in these examples, they all have features that make them chemically similar.

To define terms, let us examine the hydrolysis of ethyl acetate, which is the ethanolic ester of acetic acid (Fig. 1). In the reaction, the electropositive carbonyl carbon is attacked by a water molecule. A tetrahedral intermediate is formed that can either revert to the starting materials or form products, that is, acetic acid and ethanol. In this reaction, ethanol is the leaving group and water is the attacking group or nucleophile. The reverse reaction, in which ethanol and acetic acid form ethyl acetate, is called a *condensation* because in the process of reaction, a molecule of water is released.

Ethyl acetate is an ester, a subclass of the compounds known as acyl derivatives, that is, compounds that are formed when carboxylic acids condense with another species. Figure 2 shows some of the variety of acyl derivatives found in nature and in the laboratory. The amide is the condensation product of an amine with a carboxylic acid, in the same way that an ester is the condensation product of an alcohol with a carboxylic acid. The thioester is the product of condensation of the thioether coenzyme A with butyric acid. Acetyl chloride is the condensation product of acetic acid and hydrochloric acid. As with the ester ethyl acetate, each of these compounds can undergo hydrolysis to form the free acid and another product. The ease with which this reaction can occur depends crucially on the species that has condensed with the carboxylic acid. For example, in aqueous solution at pH 7, it will take weeks for an amide to show signs of decomposition by hydrolysis, whereas an acid chloride will

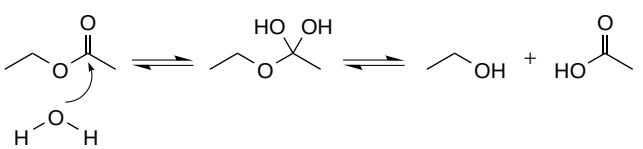

Fig. 1 Hydrolysis of ethyl acetate.

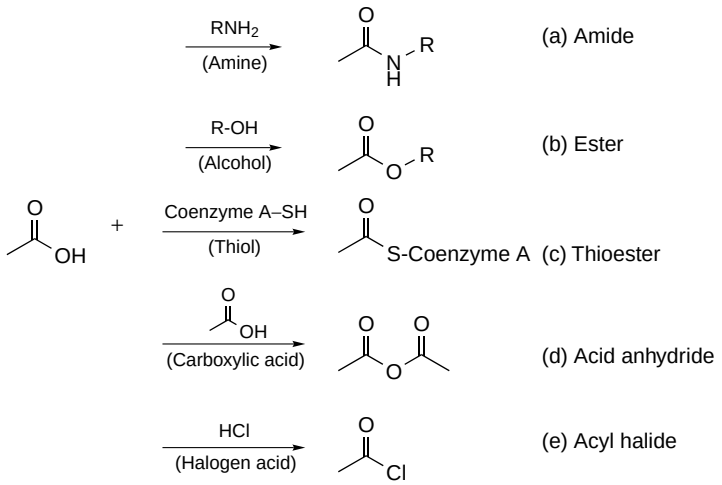

Fig. 2 Acyl derivative ranked from greatest hydrolytic stability (highest) to lowest hydrolytic stability (lowest).

be quantitatively hydrolyzed in a matter of minutes. This ease of hydrolysis, or stability, of the compound has profound biochemical ramifications.

It is fairly straightforward to estimate the relative stability of acyl derivatives. This stability is roughly proportional to the pK_a of the conjugate acid of the group attached to the carbonyl carbon (the group that will be replaced by water); the stronger the conjugate acid, the less stable towards hydrolysis the acyl derivative will be. For example, in the case of the amide, the group attached is RNH. The conjugate acid of this is RNH_2 or simply an amine. The pK_a of this is very high, on the order of 38. (Note, this is the pK_a of the *amine*, not the pK_a of the corresponding ammonium ion.) On the other end of the scale, if we examine acetyl chloride, the group attached is Cl. The conjugate acid of this is HCl or hydrochloric acid. The pK_a of this is on the order of −7. The other acyl derivatives, esters, thioesters, and acid anhydrides, all having attached groups whose conjugate acid pK_a values are intermediate to these two extremes, are found to be of intermediate stability (Fig. 2).

What is the reason behind this relationship? One way of looking at it is as follows: if the pK_a of the conjugate acid of the attached group is high, it means that it is not very acidic. In other words, the acidic proton is bound tightly. This tight binding indicates that the attached group is capable of donating electrons to produce effective bond formation, which in turn reduces the amount of free proton in solution. When the attached group is attached to the carbonyl carbon instead, this effectiveness of bond formation is preserved and is manifested in the bond stability.

For example, consider acetic anhydride from Fig. 2. This is the condensation product of two molecules of acetic acid. It is an effective reagent for adding acetyl groups to nucleophiles, such as amines. The conjugate acid of the attached group would simply be acetic acid, which with a pK_a of 4.5 satisfies the criterion of being a good leaving group for the reasons stated above. When the amine attacks the

carbonyl carbon to form the tetrahedral intermediate, either the amine or the acetic acid can be expelled to form the starting materials. Most of the time, the acid will be expelled because it is less able to contribute its electrons to make an effective bond.

In the same way that carboxylic acids can condense with alcohols to produce esters, so can other acids such as phosphoric acid and sulfuric acid. In the biological realm, condensation products of phosphoric acid are of greatest interest. It is useful to think of the chemistry of phosphoric acid condensation products as analogous to carboxylic acid condensation products (Fig. 3). A major difference between the two is that phosphoric acid has three acidic OH groups whereas carboxylic acids have only one. The result of this is that phosphoric acid can form multiple condensations. The most important of these in the molecular biology realm is the phosphodiester linkage that forms part of the backbone of DNA. And as with carboxylic acids, phosphoric acids can also form anhydrides. In the same way that carboxylic acid anhydrides are effective acylating agents because of the quality of the leaving group, phosphoric acid anhydrides are effective phosphorylating agents. The most well-known biologically important phosphorylating agents are the nucleotide triphosphates such as adenosine 5′-triphosphate (ATP).

The formation of the phosphodiester linkage between two nucleotides during DNA biosynthesis demonstrates some of the similarities and differences between phosphorylation and acylation chemistry (Fig. 4). The nucleophile in this case is the 3′-hydroxyl of the elongating nucleotide strand. It attacks the alpha phosphate of the next nucleotide triphosphate, forming a pentacoordinate intermediate. (It is pentacoordinate because the phosphate initially has four groups attached.) The intermediate can then either revert to starting materials or proceed to the diester product by ejecting the pyrophosphate group (the two attached phosphates). As with the acylation of amines with acetic anhydride, the reaction proceeds toward products because of the superiority of the phosphate as a leaving group compared to the attacking alcohol. In this reaction, the pyrophosphate is the leaving group. Nucleotide triophosphates are frequently used in nature to

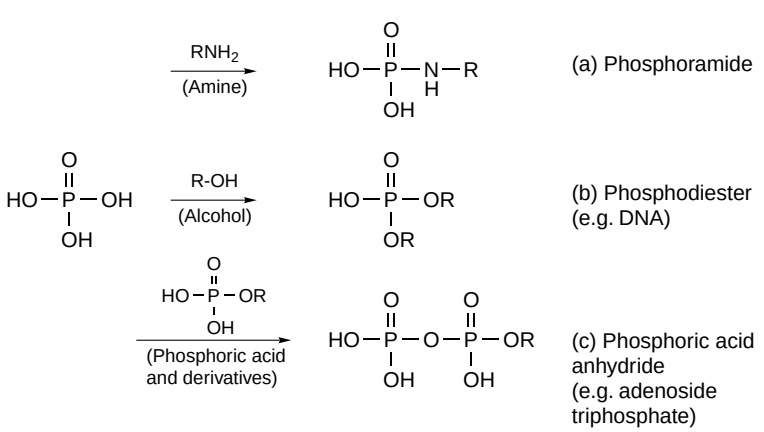

Fig. 3 Phosphoric acid condensation products.

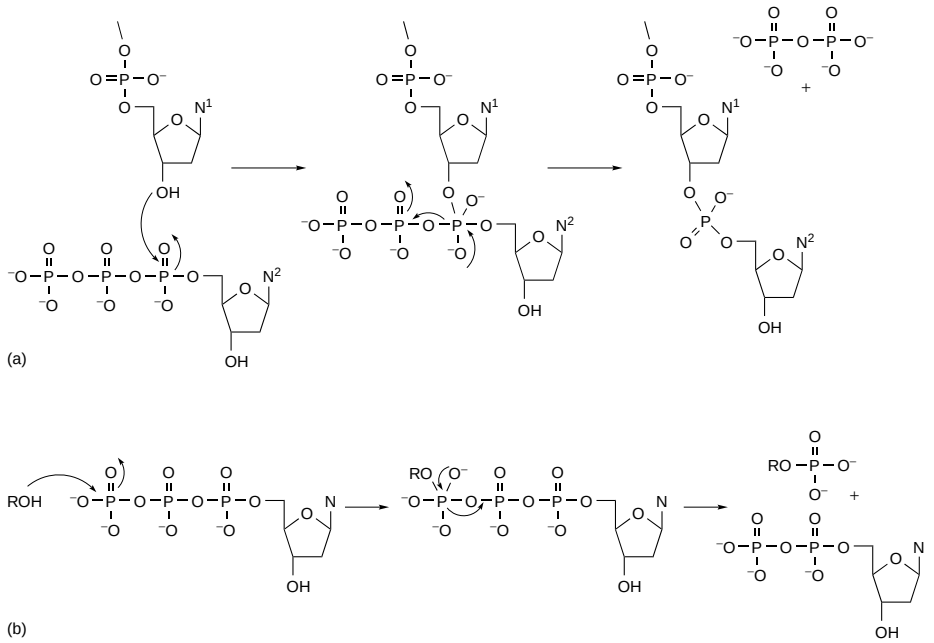

Fig. 4 (a) Polynucleotide biosynthesis. (b) Nucleophile phosphorylation with ATP.

transfer phosphate groups to nucleophiles like the hydroxyl groups of glucose during glucose metabolism. In this case, the nucleophile will add to the gamma or terminal phosphate, and the leaving group will be adenosine diphosphate, ADP, not pyrophosphate.

2.2 Noncovalent Interactions

We now turn our attention to the realm of noncovalent interactions, those forces that cause molecules to associate with and recognize each other. These interactions are of importance in all of biochemistry, from the specific hydrogen bonding patterns in the base pairs of DNA that allow for information to be coded to the enzymes' ability to distinguish between two ligands that can differ by a single methyl group. In addition to aiding in the understanding of binding processes, these forces are of key importance in the realm of catalysis as well. This concept will be explored at greater length later.

The hypothetical binding process of a protein "P" and a ligand "L," forming a complex PL can be described by a simple equilibrium (Fig. 5). The equilibrium association constant $K_a = [PL]/[P][L]$ is related to the standard free energy of binding (ΔG°_{bind}) through the expression $\Delta G^\circ_{bind} = -RT \ln K_a$ (where R is the universal gas constant and T is the temperature). This ΔG°_{bind} in turn is the sum of many factors such as coulombic attractions, hydrogen bonding, hydrophobic effects, van der Waals' forces, conformational effects, and translational/rotational entropy. A good rule of thumb is that for every 1.4 kcal mol^{-1} increase in ΔG°_{bind}, there is a tenfold increase in binding affinity. ΔG°_{bind} being a free energy, which

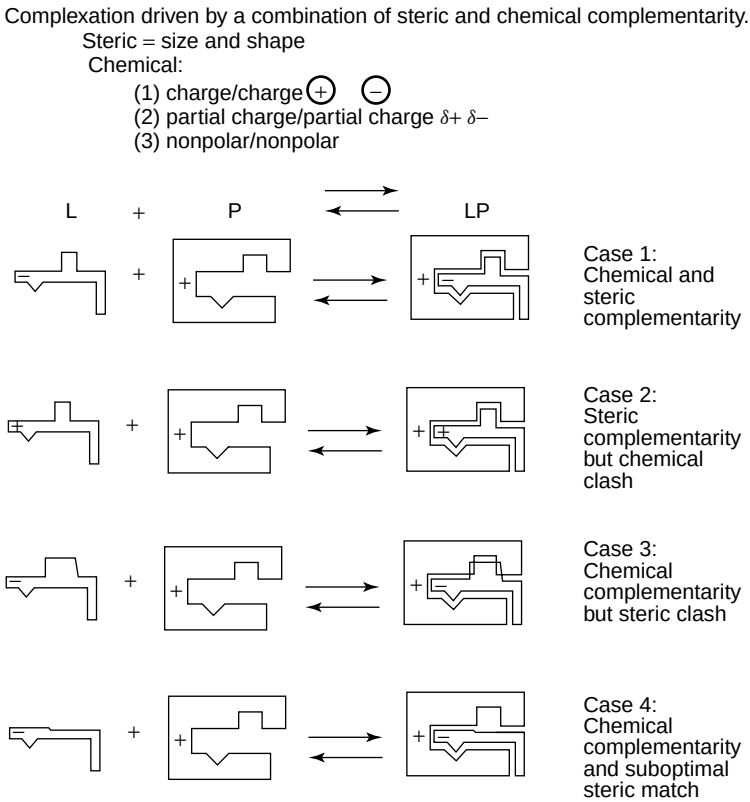

Fig. 5 Schematic representation of protein (P) ligand (L) association. Case 1 represents the optimal situation with simultaneous steric and chemical complementarity.

reflects the overall stability of the system, is a combination of two terms, enthalpy (ΔH°) and entropy (ΔS°) following the equation $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$. An overall negative $\Delta G^\circ_{\text{bind}}$ is indicative of a favorable process, so the trend is that reactions that have a decrease in enthalpy and an increase in entropy will tend to occur spontaneously, although either term can overcome the other. Enthalpy refers to the change in attractions and repulsions that occur during the reaction that are the result of forces (such as coulombic forces and intramolecular forces such as bond stretching.) The entropic term refers to the change in the "randomness" in the system or the number of configurations the system can adopt.

For the most efficient binding of the ligand, it should have both steric and chemical complementarity to the target site. Steric complementarity refers to the shape of the ligand versus the shape of the binding site. Chemical complementarity refers to the appropriate matching of surfaces on the ligand with the binding site. On average, surface elements that contact each other should have opposite charges on the ligand and the site to allow for effective

electrostatic interaction, or both should be nonpolar (uncharged) to allow for effective hydrophobic interaction (Fig. 5).

2.2.1 Electrostatic

Electrostatic interactions refer to charge–charge interactions. An example of this is the close contact of a positively charged amino group and a negatively charged carboxyl group (this arrangement is also known as a salt bridge). Coulomb's law describes the energy that is released when two charges are brought together (Fig. 6). From the form of the equation, we can see that the energy is directly proportional to the charges on the species and inversely proportional to the distance separating them. In addition, the dielectric constant indicates the ease with which the electrostatic field is propagated through the medium (a vacuum has a dielectric constant of 1, whereas the value for water is 79). Misuse of this equation can lead to bizarre conclusions. For example, if one calculates the energy released upon formation of a salt bridge using the distance of van der Waals' contact and the dielectric constant of a vacuum, the result is an energy on the order of 60 kcal mol^{-1}. In

$$\text{Electrostatic interaction energy} = \frac{(\text{Charge 1}) \times (\text{charge 2})}{\text{Distance} \times \text{dielectric constant}}$$

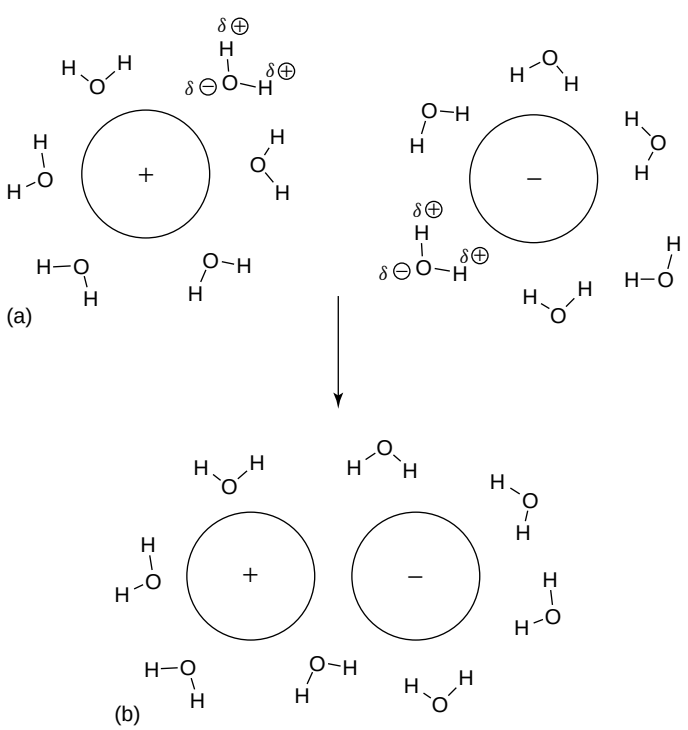

Fig. 6 (a) Electrostatic interaction energy. (b) Schematic representation of the salt bridge–forming process showing the desolvation of the charged species and its interaction with water dipoles.

reality, the contribution of a salt bridge to the stability of a complex is on the order of 0 to 3 kcal mol^{-1}. The source of this difference lies in the presence of water. In all associative processes, it is imperative not only to examine the energy being released by bringing two groups together but also the energy required to remove the species from contact with the solvent (i.e. water). In the case of the salt bridge, this requires taking into account the desolvation of the amino group and carboxyl group as the salt bridge forms. This desolvation is an energetically expensive process because although water has no net charge, it still has a strong attraction to charged groups because of its dipole moment (i.e. separation of charge within the molecule) (Fig. 6). In addition to this effect, water decreases all electrostatic interactions through a general shielding effect.

An example of salt bridges in action may be found in the structure of the restriction endonuclease ECO RV bound to an oligomeric double-stranded DNA. Along the contact surface of the nuclease and the polynucleotide, there are seven positively charged groups (arginines, lysines, and histidines) within salt bridge–forming range of the anionic phosphate backbone of DNA, while no negatively charged groups (aspartic and glutamic acids) are in similar positions. These nonspecific salt bridges ensure that the nuclease will bind to DNA in general, while other specific hydrogen bonding interactions give it specificity for its unique cleavage site.

2.2.2 Hydrogen Bonding

Hydrogen bonding is also an electrostatic interaction, although the participants are not necessarily formally charged. Examples of hydrogen bonds are shown in Fig. 7. The defining characteristics are (1) a hydrogen atom attached to an electronegative atom such as nitrogen or oxygen (this is the hydrogen bond donor) and (2) a lone electron pair on a nitrogen or oxygen (this is the hydrogen bond acceptor). The hydrogen has a partial positive charge due to the electron-withdrawing nature of the atom to which it is attached, and this interacts with the negatively charged lone pair on the hydrogen bond acceptor in an essentially coulombic attraction. Hydrogen bonds also contribute on the order of 0 to 3 kcal mol^{-1} to molecular associations, although there can be variation in this number.

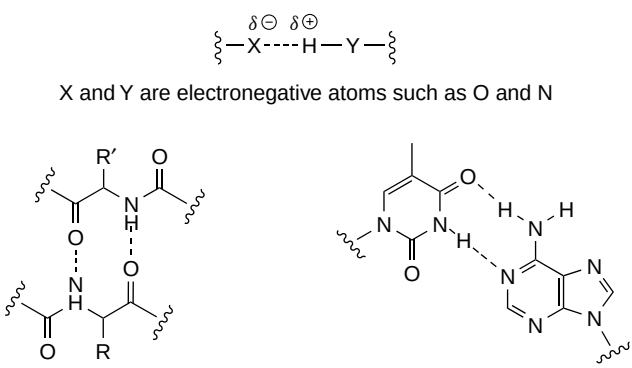

Fig. 7 Hydrogen bond formation in (left) protein beta sheets and (rights) nucleotide base pairing.

2.2.3 Hydrophobic Effect

The hydrophobic effect refers to the tendency of nonpolar species to associate in water and thereby limit the extent of their water contact. Macroscopically, this is manifested in the separation of oil from water. Microscopically, this is manifested by the tendency for side chains of nonpolar amino acids to cluster in the interior of proteins, away from water contact. In a variety of studies, it has been demonstrated that there is a linear dependence on the amount of nonpolar surface a molecule has and its tendency to prefer nonpolar environments. Although this is a complex and by no means a fully understood phenomenon, it appears to be entropically driven. A nonpolar surface cannot hydrogen bond effectively with water molecules. The result is that there is an ordering of water molecules around nonpolar species, with each water orienting so as to maximize the number of hydrogen bonds. The result is that these waters are limited in their range of motion, resulting in a decrease in entropy and therefore an (free) energetically unfavorable situation. The more nonpolar molecules cluster, the less (non–hydrogen bonding) surface they expose to water and the less of this water ordering takes place.

2.2.4 van der Waals' Interactions

In addition to the strong interactions of charges found in salt bridges and hydrogen bonds, there are weaker forces that allow principally uncharged species to also have attractive (and repulsive) interactions with other species. This is so because although these nonpolar species have no formal charge or dipole, they can transiently form a dipole due to random fluctuations of electrons within them. These transient dipoles can interact with other transient (and nontransient) species. In general, because of the random generation of these dipoles, van der Waals' forces are rather weak forces, acting over short distances. A classical example of van der Waals' attraction is between the stacked bases of DNA.

2.2.5 Conformational/Configurational Aspects

In addition to all of the specific attractive and repulsive forces described above, there are several other important factors that influence molecular associations. First of all, there is conformational free energy. If a molecule has to adopt a highly strained conformation in order to bind productively to its host, its affinity for the host will be decreased relative to a similar molecule that can make the same interactions without adopting the strained conformation. Furthermore, if a molecule can adopt many conformations (i.e. a so-called conformationally flexible molecule) and only a few of these can productively bind, then its affinity too will be decreased relative to a rigid, conformationally restricted molecule that can make similar interactions.

Finally, there is the rotational/translational entropy loss that accompanies binding. This is the free-energetic cost of preventing a molecule from moving freely in solution. This is a fixed cost in the sense that it will always have to be paid when two molecules associate. Before two molecules associate, they each have three rotational and three translational (movement through space) degrees of freedom, for a total of 12. These degrees of freedom are related to the system's entropy, the greater the number of degrees of freedom, the greater the randomness (and entropy). After the complex forms, there is essentially a single species having three degrees of rotational and three degrees of translational freedom. The overall entropy is decreased,

which is unfavorable from the standpoint of free energy. It is difficult to determine exactly the free-energetic cost of losing rotational/translational entropy upon complexation, but it has been estimated through a variety of techniques to be on the order of 7 to 11 kcal mol^{-1}.

3 Nucleotide Interactions

A wide variety of bioorganic investigations deal with molecules that interact with DNA. The purposes behind these investigations are manifold, ranging from the development of therapeutics to the understanding of the forces required for molecular recognition. The following section is divided into two parts. The first deals with synthetic oligonucleotides interacting with DNA, the second with small molecules that interact with DNA.

3.1 Oligonucleotides

The revolution in recombinant DNA technology relies critically upon the ability to synthesize a variety of oligonucleotides quickly and cheaply. The development of solid-phase oligonucleotide synthesis has made this process relatively simple. In addition to the many uses of these oligonucleotides in recombinant DNA manipulation (ranging from use as polymerase chain reaction (PCR) primers to the generation of site-directed mutants), they have potential as therapeutics in the field of antisense therapy, gene therapy, and RNA interference. Paralleling that use, in solid-phase peptide synthesis (SPPS), a common synthetic strategy that is used in oligonucleotide synthesis is depicted in Fig. 8.

There are many variations on this chemistry, not only for the production of natural DNA but also for the production of unnatural polynucleotides that have alternative backbone and nucleic acid structures. An example of the latter is phosphorothioate-containing DNA, in which one of the two unesterified phosphate oxygens is replaced with a sulfur. This may be achieved by replacing the hydrogen peroxide or iodine in the phosphite oxidation step shown in Fig. 8 with a solution of elemental sulfur. Phosphorothioate DNA has been found to have greater resistance to nucleases (enzymes that cleave the diester backbone of DNA), which makes it more suitable for antisense therapy than naturally occurring DNA.

Other approaches to artificial DNA have been tried that completely bypass the use of phosphate derivatives. Once again, the rationale behind this is to make a selective agent that will bind to a specific gene target but will be resistant to nucleases. P.E. Nielsen and coworkers have recently developed a group of DNA analogs in which the bases are connected to a poly*amide* backbone (Fig. 9). The result is an uncharged species that is capable of forming specific base-pairing hydrogen bonds with target genes. In addition to having resistance to nucleases, these peptide nucleic acids, or PNAs, form complexes with their target nucleic acid sequences that appear to be even tighter than those found in natural DNA duplexes. A probable cause of this phenomenon is the lack of negative charge on the PNA compared with the normal hybridization partner. We previously described the energetic advantage of coulombic attraction through, for example, the formation of salt bridges between negatively and positively charged species. The flip side to

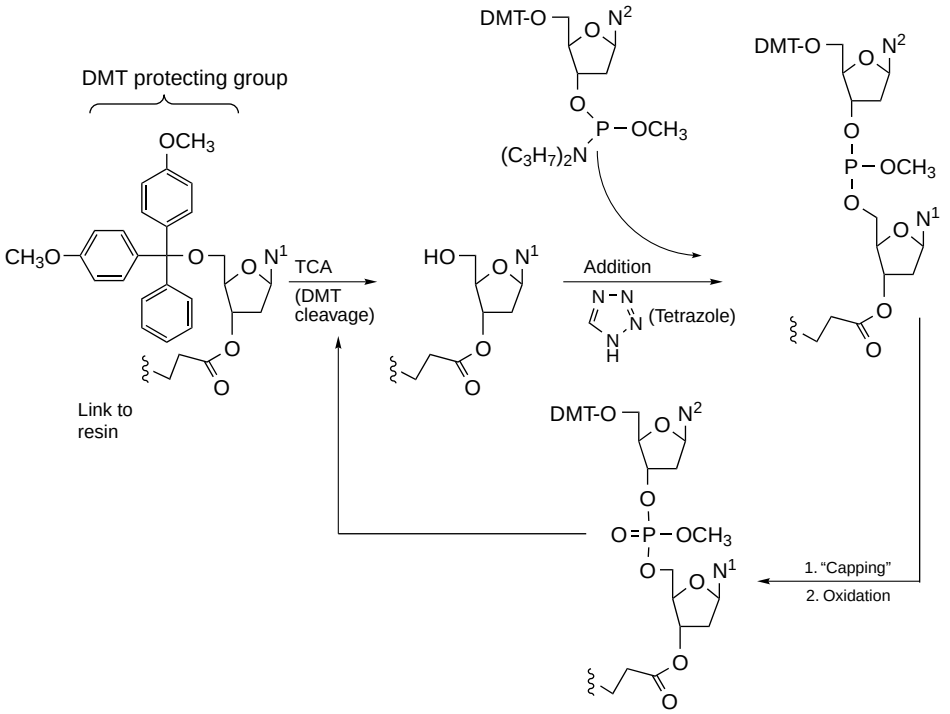

Fig. 8 Oligonucleotide synthesis begins with the tethering of the first monomer to an insoluble resin via an ester linkage to its 3′ hydroxyl group. To prevent esterification of the 5′ hydroxyl instead, it is blocked with a dimethoxy trityl group (DMT). After the ester linkage to resin has been formed, the DMT group is cleaved with acid. The next monomer (also 5′ hydroxyl DMT protected) contains a phosphoramidite group attached to its 3′ hydroxyl. The displacement of the phosphoramidite amine by the deprotected 5′ hydroxyl is catalyzed by the addition of tetrazole. "Capping" or acetylation of the 5′ hydroxyl group of any monomer that did not react with the phosphoramidite prevents errors from compounding. The capping prevents the unreacted, and therefore incorrect, oligomer from undergoing further chemistry. After the coupling between monomers has taken place, the phosphoramidite is oxidized to form the phosphate triester using either hydrogen peroxide or iodine. The DMT group is cleaved, thereby releasing the 5′ hydroxyl that can then participate in the next round of monomer addition. When the oligomer is complete, a final step of deprotection removes the final DMT group, any protecting groups on the nucleotide bases, and converts the phosphate triester into the naturally occurring diester.

this is that there is a destabilizing effect produced when like-charged species are brought together. This is precisely the situation with double-stranded DNA. The distance between phosphate groups on opposite strands can be as little as 8 Å, thus producing a significant electrostatic repulsion. Obviously, there are enough other interactions in dsDNA such as the hydrogen bonding between the base pairs and van der Waals' interactions between stacked base pairs to overcome this destabilizing influence. In the case of the PNA, these positive interactions are maintained, while the destabilizing electrostatic repulsion is eliminated. Thus, the stability of

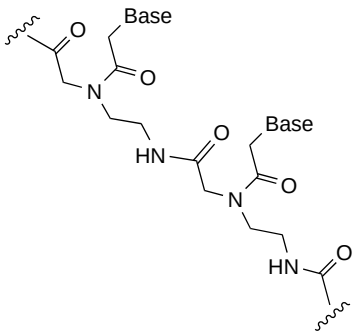

Fig. 9 PNA: peptide nucleic acid, a synthetic polynucleotide with an amide backbone.

the hybridized PNA is greater relative to its DNA analog.

In addition to alterations in the backbone, numerous bioorganic investigations have examined the effect of novel bases incorporated into DNA. (Many of these studies would be impossible without the prior development of efficient solid-phase oligonucleotide synthesis.) For example, Kool and coworkers have incorporated steric analogs of the nucleoside bases that have the same shape as the normal bases but have no hydrogen bonding capability. Difluorotoluene (Fig. 10) closely mimics the shape of thymine but is nonpolar and does not closely mimic its charge properties. Despite this, it can efficiently and specifically direct the incorporation of adenine opposite to it by DNA polymerase when it is used in oligonucleotides. This kind of result helps tease apart the relative importance of charge and shape in molecular recognition processes.

3.2 Small Molecules

There are literally hundreds of small molecules that have been found to have interesting or useful effects that depend upon their binding to DNA and form the basis for bioorganic investigations. The uses of these compounds include (1) antiviral and antineoplastic (anticancer) chemotherapies (e.g. bleomycin), (2) nucleotide stains (e.g. ethidium bromide), and (3) biochemical probes for nucleic acid structure (e.g. phenanthroline complexes). The major interaction modes are binding to the minor and major grooves, intercalation between stacked base pairs, as well as a host of potential

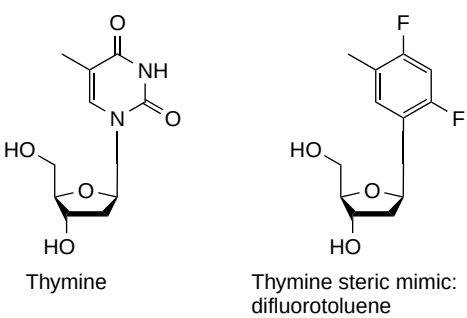

Fig. 10 Comparison of thymine and its steric analog difluorotoluene.

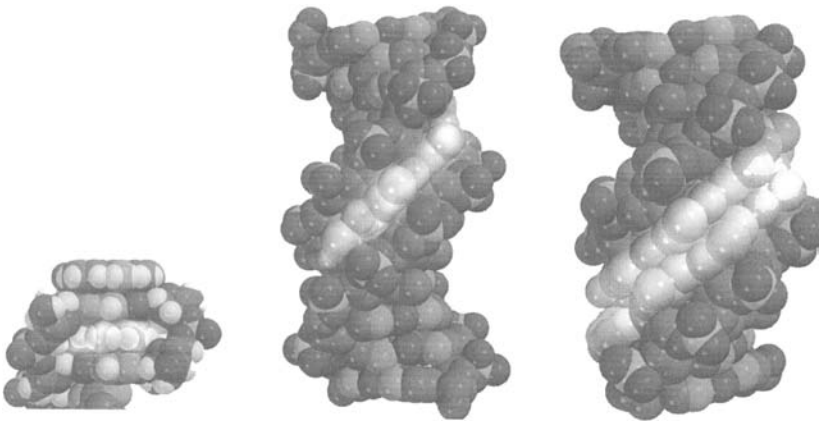

Intercalation to DNA by an acridine derivative (highlighted above, chemical structure shown below). Planer structure stacks between planer base-pairs.

Binding to minor groove by distamycin (highlighted above, chemical structure shown below). Exclusive hydrogen bond donation to AT rich regions which are exclusively hydrogen bond accepting.

Binding to minor groove by side by side polyamide (highlighted above). Sequence specificity tuned by use of both pyrrole and imidazole, the latter which is able to accept a hydrogen bond from G.

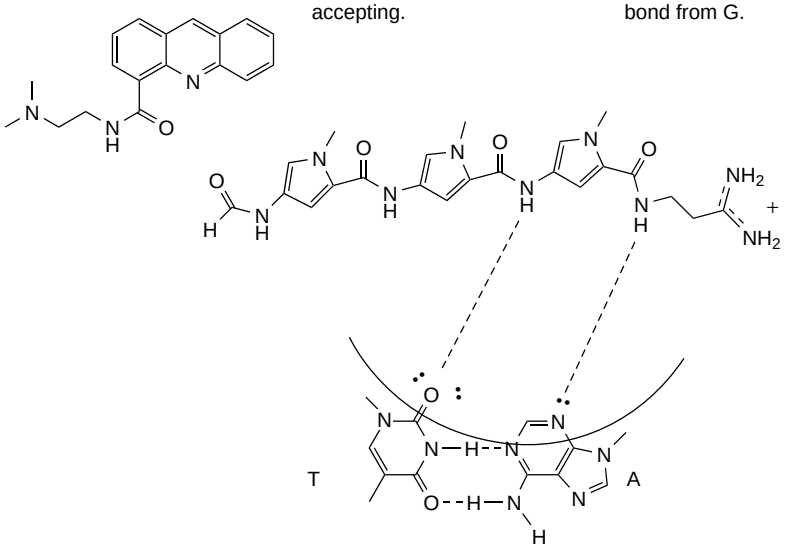

Fig. 11 Major sites of small-molecule interaction with DNA. (left) intercalation by acridine derivative (middle) minor groove binding by distamycin (right) minor groove binding by side-by-side polyamide oligomers.

covalent modifications (including bond cleavage and alkylation) of the nucleotides, sugars, and phosphate backbone (Fig. 11). A well-studied DNA-binding small molecule is the natural product distamycin (Fig. 11), which has been shown to bind noncovalently and selectively to the minor groove of AT-rich portions of DNA.

In addition to having antiviral properties, which are probably linked to its ability to inhibit the enzyme DNA polymerase, distamycin has found use as a template for designing molecules with altered sequence specificity and sequence-specific cleavage.

The binding of distamycin to the minor groove of double-stranded DNA involves the combination of several of the previously described intermolecular forces. These include (1) conformational energy, (2) van der Waals' interactions, (3) hydrogen bonding, and (4) electrostatic interactions. The minor groove of DNA in the AT-rich regions to which distamycin binds is quite narrow, roughly the width of a single carbon atom. The distamycin molecule is composed of principally sp^2-hybridized (and approximately planer) atoms. Furthermore, uncomplexed distamycin exhibits a slight twist. This twist, brought about by subtle conformational forces within the molecule, is almost exactly the same as the twist observed in the spiraling minor groove of the double helix. Thus, there are none of the conformational energy penalties upon binding as described previously. Furthermore, because the groove width is very close to being the width of distamycin, there is an opportunity for extensive van der Waals' interactions between the faces of distamycin and the walls of the minor groove. Specific hydrogen bonds also play an important role in the binding. In the AT-rich regions of the minor groove to which distamycin binds, there are exclusively hydrogen bond *acceptors* in the form of lone pairs extending from nitrogens and oxygens of the exposed edge of the nucleotide bases. Distamycin in its natural conformation has exclusively hydrogen bond *donors*, in the form of amide protons, directed toward the floor of the minor groove. Finally, distamycin has a positively charged amidinium group at one of its ends. This contributes to the binding through electrostatic interaction with both the negatively charged phosphate backbone and the lone pairs on the minor groove floor.

The work of Dervan and others has extended the sequences that can be specifically targeted by minor groove–binding molecules such as distamycin. So-called hairpin polyamides that contain both the pyrrole rings of distamycin as well as the imidazole rings are able to bind to sequences of DNA by specifically recognizing and differentiating GC base pairs, CG base pairs, and AT/TA base pairs. The basis for this recognition is a side-by-side binding motif, where two planer oligomers (joined by a linker moiety) simultaneously bind to the minor groove (Fig. 12). It should be noted that this binding could not have been easily predicted from examining crystal structures of DNA, as the minor groove in most of these is far too narrow to accommodate two such molecules. DNA is a dynamic molecule and has been shown to be able to distort sufficiently to allow both molecules to bind. Binding specificity is rooted in the arrangement of pyrrole and imidazole rings. As noted above, the floor of the minor groove in AT regions of DNA is exclusively hydrogen bond accepting, exposing lone pairs of electrons with partial minus charges. Regions of the side-by-side hairpin polyamide that have a pyrrole across from a pyrrole bind to these AT regions, with hydrogen bonds being donated from the amide NHs that are adjacent to the pyrrole rings. Regions with GC or CG base pairs are recognized by an imidazole/pyrrole pair or pyrrole/imidazole pair respectively. This is because only G has the ability to donate a hydrogen bond (from an exocyclic NH) and only imidazole (with the lone pair on the unprotonated

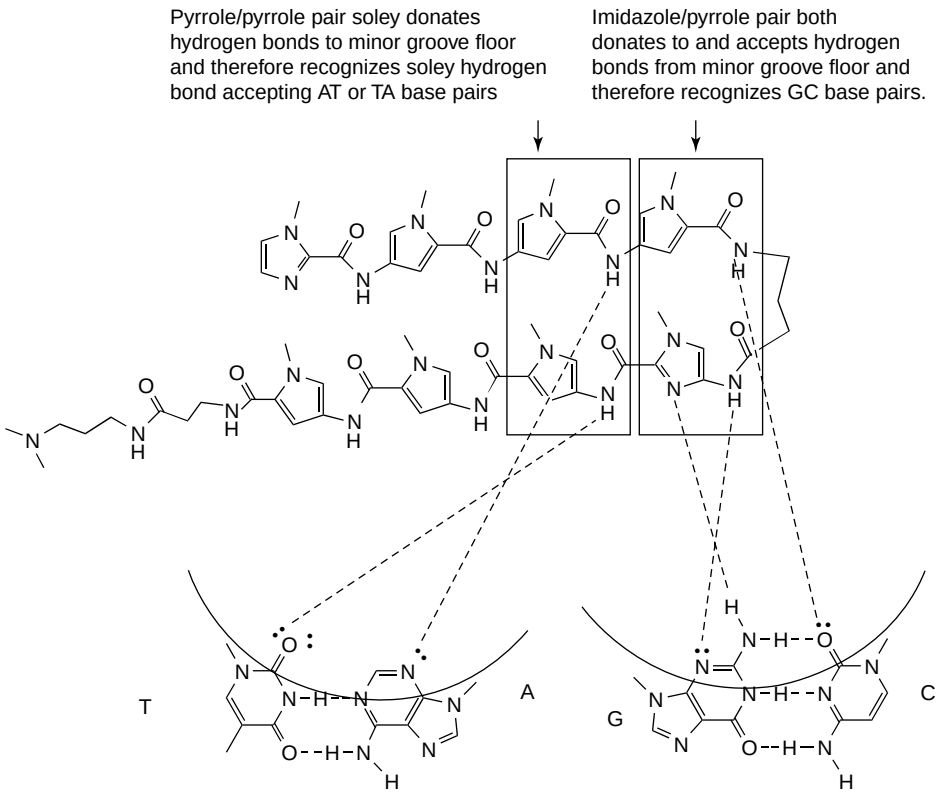

Fig. 12 Details of the mechanism of differentiation of AT and GC base pairs by "side-by-side" hairpin polyamides.

ring nitrogen) has the ability to accept a hydrogen bond. Oligomers built using these "pairing rules" of oligomer sequence matched with DNA sequence have shown remarkable affinity and specificity for sequences of DNA and may have application as in diagnostics and therapeutics.

4 Protein Studies

4.1 Enzyme Inhibition

The study of enzyme inhibition is a major focus of bioorganic chemistry for two major reasons. Firstly, enzyme inhibitors are useful tools for understanding enzyme structure and mechanism of action. Secondly, enzyme inhibition has proven to be a successful strategy for the design of therapeutics. Examples of drugs that act by inhibiting enzymes include methotrexate, Deprenyl, and Enalapril. Inhibitors can block enzyme function by preventing access of the natural substrate to the active site or by stopping the subsequent reaction from taking place by, for example, chemically blocking a key catalytic group on the enzyme.

There are two broad classes of inhibitors, covalent and noncovalent, the

distinction being whether or not a covalent chemical bond forms between enzyme and inhibitor. These classifications may be further divided into categories on the basis of kinetic and mechanistic criteria. Major subclasses include transition-state analogs, multisubstrate analogs, affinity labels, and mechanism-based inhibitors.

Noncovalent enzyme inhibitors rely on having two important features: (1) enough structural similarity to the actual enzyme's substrate to allow recognition and tight binding to the enzyme that is to be inhibited and (2) differences in the chemical structure that prevents the normal chemistry from taking place. An example of this type of compound is the drug Enalapril. It is the prodrug of Enalaprilat, which is an inhibitor of the protease angiotensin-converting enzyme (ACE). This enzyme cleaves the 10-residue peptide angiotensin I into the 8-residue peptide angiotensin II. These peptides are part of a pathway that controls blood pressure, and it has been found that preventing the conversion of angiotensin I to angiotensin II can be effective in reducing certain types of blood pressure (Fig. 13). This figure also shows schematically the interactions between enzyme and substrate that have been deduced to be crucial for binding affinity. Of note are a hydrogen bond to a carbonyl group, a salt bridge from the carboxyl terminus to a cationic site on the protein, and a putative hydrophobic pocket that binds the phenyl ring of the phenylalanine residue that is adjacent to the scissile (or cleavable) peptide bond. Also indicated schematically is an enzyme-bound zinc ion that is crucial for catalysis. It is shown to be interacting with the carbonyl of the peptide bond that is to be cleaved. It is believed that the zinc polarizes the carbonyl group, pulling electron density from the oxygen and therefore, indirectly, from the carbonyl

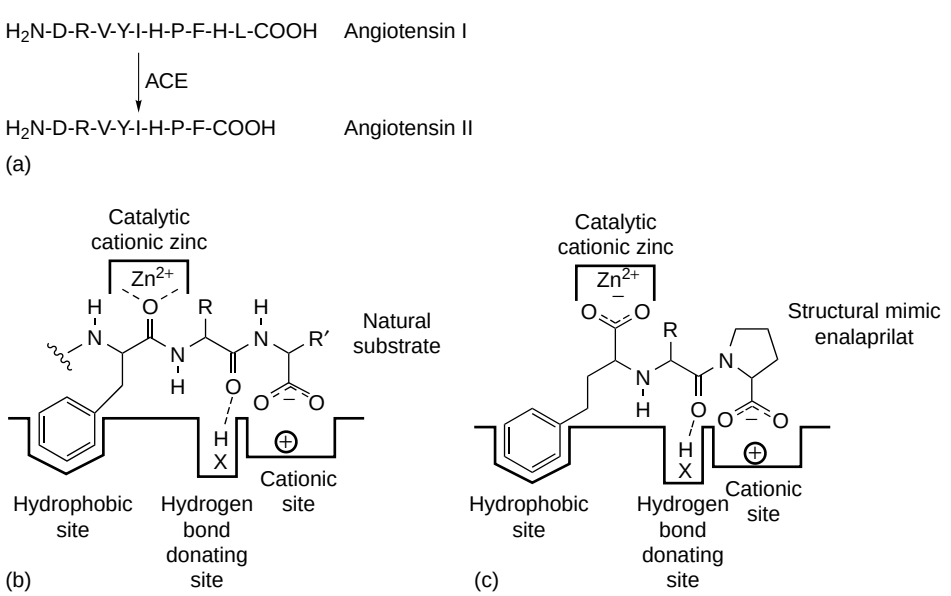

Fig. 13 (a) The reaction catalyzed by angiotensin-converting enzyme (ACE). (b) Schematic representation of the interaction of ACE with both substrate and (c) Enalaprilat, an inhibitor.

carbon. This makes the carbonyl carbon more susceptible to nucleophilic attack and reaction. Enalaprilat is able to make similar interactions with ACE. Figure 13 shows that the terminal salt bridge, phenyl ring, and carbonyl hydrogen bond are all maintained in positions similar to that of the natural substrate. A major difference is that the original scissile amide bond has been replaced with an unreactive amine. In order to maintain the interaction with the catalytic zinc, a carboxyl group is positioned in a similar orientation as the original carbonyl. This salt bridge is crucial for tight binding. Enalapril, the prodrug to this compound, is identical except that this carboxylate group is esterified, thereby making salt bridge formation impossible. It is also a far less potent inhibitor, evidently due to the lack of half of the salt bridge.

Covalent enzyme inhibitors also must have some structural similarity to the natural substrate in order to provide initial binding and specificity. An example of a covalent enzyme inhibitor is aspirin, or acetylsalicylic acid. It is a nonsteroidal anti-inflammatory agent. One of the ways by which it exerts this effect is by covalently modifying a critical amino group on the enzyme cyclooxygenase. Cyclooxygenase is an enzyme in the synthetic pathway of prostaglandins, which are key parts of the inflammatory response. The actual modification reaction is an acetylation.

4.2 Enzyme Catalysis

The previous paragraphs describe some of the fundamental organic chemistry found in biological systems. We now wish to demonstrate by example how some of these principles are applied in enzymes to achieve catalysis. One of the ways by which bioorganic chemistry examines this issue is through the design of small-molecule mimics of enzyme action. Enzymes are able to accelerate reactions in a variety of ways. One of these strategies, transition-state stabilization, will be discussed later. Another strategy is nucleophilic catalysis, where a reactive covalent bond is formed between enzyme and substrate that is then easily cleaved by another reactant. Frequently, enzymes take advantage of their ability to provide specific electrostatic interactions that can stabilize and promote the movements of charge within a molecule that will encourage its reaction.

For example, let us consider the reaction of ribonuclease A with its substrate, a single strand of RNA (Fig. 14). The reaction that is catalyzed is the cleavage of the phosphodiester bond to form a 3′-phosphorylated species and a 5′-free hydroxyl species. RNA is fairly stable at neutral pH, although the presence of the 2′-hydroxyl makes it less stable than DNA under similar conditions. The reason this is so is because the adjacent 2′-hydroxyl is a potential nucleophile that is in an ideal position to attack the phosphorus atom. Even though alcohols are not considered "good" nucleophiles, the proximity to the reactive center increases its reactivity. Ribonuclease A takes advantage of this adjacent nucleophile and improves upon its nucleophilicity by deprotonating it with the imidazole base of an active site histidine. In a sense, the partial minus charge on the unprotonated imidazole is transferred to the 2′-OH through the transfer of a proton. The anion is a better nucleophile (having greater negative charge, it will be attracted to the slightly positive phosphorus atom). The result of the 2′-OH attack is the formation of a pentacoordinate

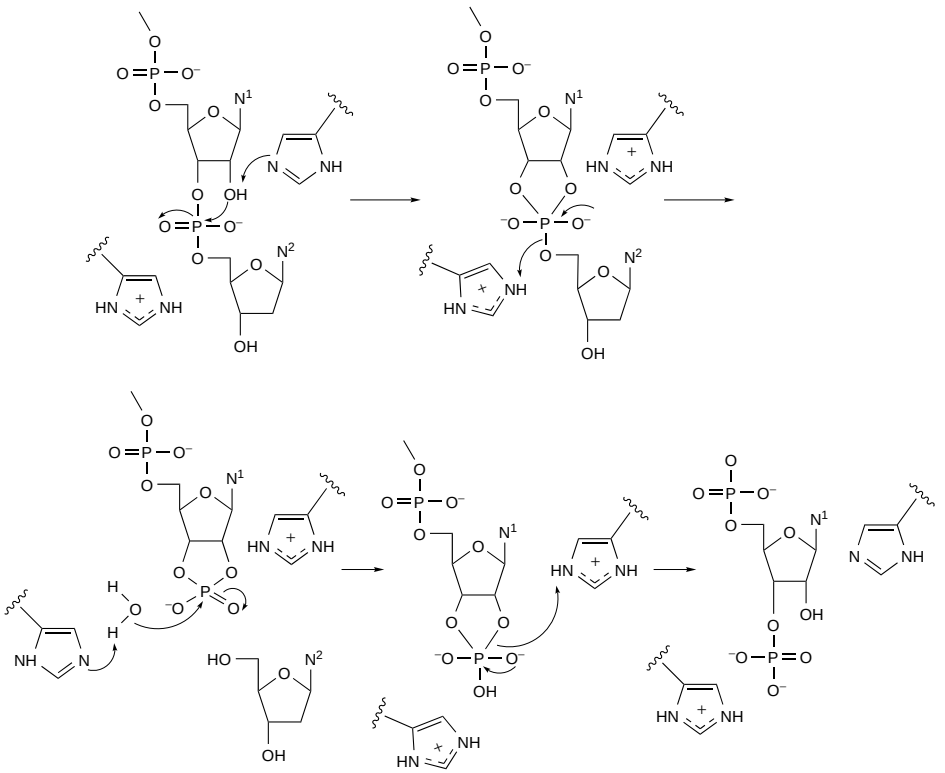

Fig. 14 The mechanism of catalysis of phosphodiester cleavage in single-stranded RNA by RNase A, accomplished by coordinated acid–base chemistry of two histidines.

intermediate. In essentially the reverse process, another active site histidine (which is protonated) protonates the 5′-OH of the other nucleotide. In doing so, it allows the flow of electrons in the phosphorus oxygen bond to migrate to the (now) positively charged 5′-OH–histidine complex. The result of this electron movement is the release of one of the products, the free 5′-OH nucleotide, and the formation of the cyclic phosphate form of the 3′-OH phosphate. In the same way that histidine A deprotonated the 2′-OH for attack originally, in the final step, histidine B deprotonates a water, essentially producing a nucleophilic hydroxide anion. This then attacks the cyclic phosphate intermediate. Histidine A can now act as an acid and protonate the 2′-OH. This again will allow the electrons in that P–O bond to migrate toward the positive charge. The result is the final product, the 3′-phosphate nucleotide. This reaction sequence demonstrates how the reactivity of specific groups can be significantly altered by specific electrostatic interactions. Furthermore, enzymes can both protonate and deprotonate virtually simultaneously, so that if the reaction calls for it, electrons can simultaneously be "pushed" and "pulled."

A rich area of bioorganic chemistry is in the development of small molecules that can in some way mimic the actions of

such enzymes by arranging key catalytic groups on an appropriate scaffold. A classic example of this is the development of small molecules that mimic the enzymatic action of ribonuclease (described above) by Breslow and coworkers. Using cyclodextran as a scaffold, they covalently attached two imidazole rings positioned to mimic the key catalytic side chains of the two histidines involved in catalysis by ribonuclease. This molecule showed activity in hydrolyzing a model phosphodiester in a manner analogous to ribonuclease (Fig. 15).

4.3 Transition-state Analogs and Catalytic Antibodies

Transition-state analogs are molecules that closely resemble a reaction's transition state (i.e. the highest energy species in the reaction pathway, more accurately known as an activated complex) in shape and charge. Enzymes are observed to bind tightly not to the substrates and the products of the reaction that they catalyze but rather to the transition state, the highest energy species that forms during

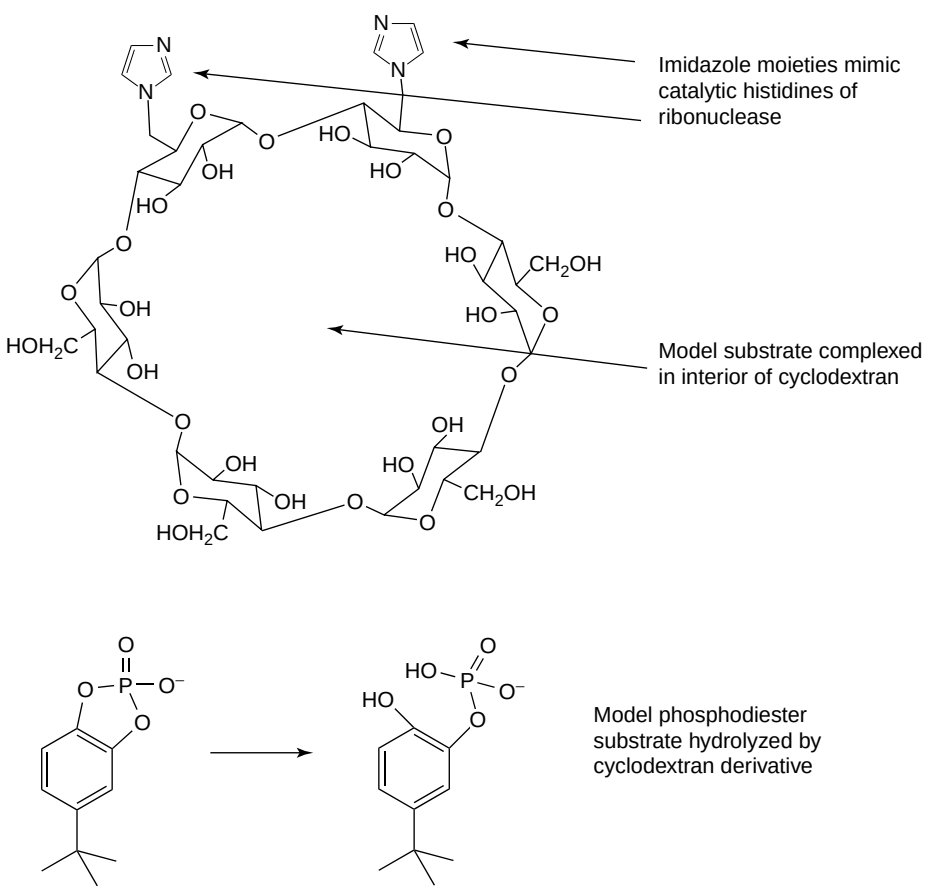

Fig. 15 Small-molecule mimic of RNase A, incorporating cyclodextran as a substrate binding site, and two imidazole moieties for catalyzing hydrolysis of the indicated model substrate.

the reaction. Enzymes have evolved to use this strategy for catalysis. By forming an active site that is complementary in shape and charge to the transition state, enzymes can stabilize or lower the energy of this species. The effect of this is to accelerate the reaction because the activation barrier is more easily overcome. This complementarity in charge and shape to the transition state is readily demonstrated by the effectiveness of transition-state analogs as enzyme inhibitors. A classic example of a transition-state analog is a phosphonate-containing molecule that mimics the transition state of the hydrolysis of the corresponding acyl derivative (Fig. 16). When an ester is hydrolyzed, the central carbonyl group changes from being planer (sp^2 hybridization of the carbonyl carbon) to being tetrahedral (sp^3 hybridization). The phosphonate-containing analog is able to reproduce the shape of the transition state as well as the partially negatively charged oxygens. In addition, it is chemically stable, whereas the actual transition state exists only fleetingly. In this example, an enzyme that catalyzed the hydrolysis of this ester would also bind tightly to the phosphonate analog. Transition-state analogs are a successful strategy for enzyme inhibition, one that forms the basis of several drugs.

In addition to generating effective enzyme inhibitors, transition-state analogs can be used to generate antibodies that are capable of catalyzing chemical reactions (so-called "catalytic antibodies"). The theory underlying this approach is that antibodies that are elicited against a transition-state analog, and therefore have the ability to bind them, will be chemically and sterically complementary to the transition state and will therefore be potentially capable of catalyzing the reaction. The general strategy behind the production of a catalytic antibody is to (1) design and synthesize a molecule whose shape and charge closely resemble that of the transition state of the reaction one wishes to catalyze, (2) tether this molecule to a larger molecule, (3) elicit an immune response to this conjugate, and (4) screen the resultant monoclonal antibodies for

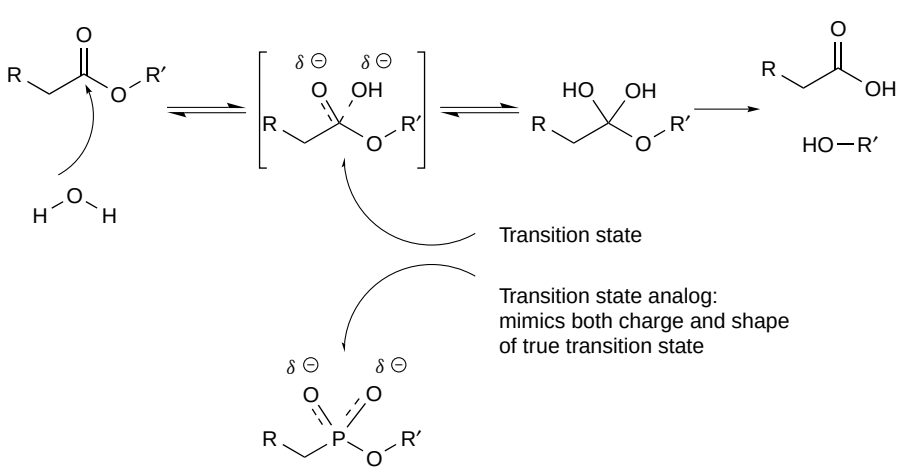

Fig. 16 Comparison of transition-state structure and transition-state analog structure for an ester hydrolysis.

catalytic activity of the type desired. This approach has proven successful in a range of systems.

4.4 Solid-phase Peptide Synthesis

Solid-phase peptide synthesis (SPPS) is a technique that allows the semi- or fully automatic synthesis of relatively long peptides without the need for purification of reaction intermediates. The basic strategy of SPPS is outlined in Fig. 17. The first amino acid of the peptide is tethered to a derivatized resin via its carboxyl group. The amino terminus of this first amino acid is freed from its protecting group by an appropriate cleaving system and then treated with the activated carboxyl of the next amino acid to be added to the chain. The amino group of this newly added amino acid is also protected to prevent condensation with the activated carboxyl. After the new peptide bond is formed, the new terminal amino group is deprotected and the cycle of condensation followed by deprotection can continue. In addition to protection of the alpha amino group of the newly added amino acid, the chemically reactive side chains of the peptide (such as the amino group of lysine and the carboxyl group of aspartic acid) are also protected with groups that will not be cleaved until after the peptide synthesis is complete. A major advantage of SPPS is that after each of these steps, the reaction by-products can be flushed from the container holding the resin. The peptide that is attached to the insoluble resin will remain in the container. The alternative to SPPS is solution-phase peptide synthesis, which, as with most organic synthesis, requires the isolation and purification of reaction intermediates, usually after each condensation reaction. Thus, making a peptide of even modest length becomes a major undertaking. In addition to removing a large amount of the purification from the process, SPPS lends itself very well to automation. A modern automatic solid-phase peptide synthesizer can make a peptide of 20 residues in length in about a day. The equivalent synthesis by solution methods would typically be on the order of months. Since each intermediate is not isolated and purified, the key to making SPPS a viable technique was the

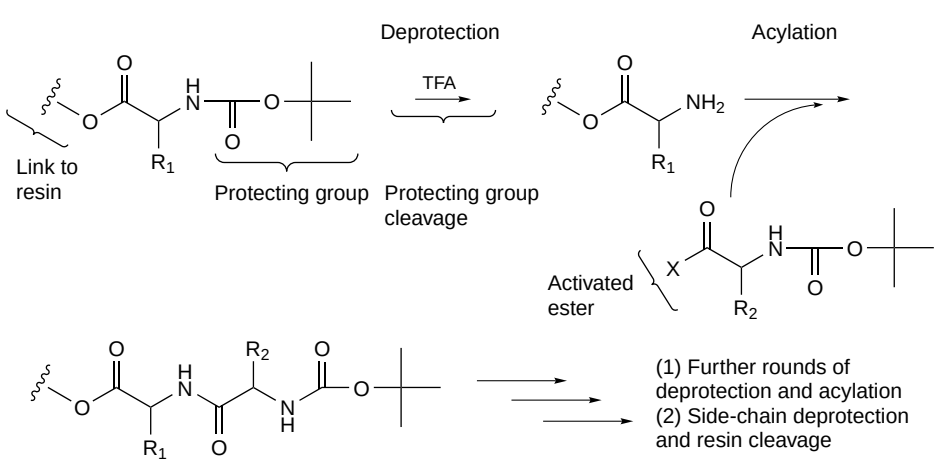

Fig. 17 Strategy of solid-phase peptide synthesis using BOC chemistry.

development of very efficient coupling reactions. If the coupling efficiency is 95%, a high yield under normal circumstances, in the process of making a 20-mer, 0.95^{20} or 36% of the resultant peptide will be of the sequence specified. The remainder will be a mixture of similar but incorrect peptides. Modern solid-phase peptide synthesizers are able to attain efficiencies on the order of 99%, which allows good purities to be attained.

5 Combinatorial Chemistry

A central area of bioorganic chemistry is the identification of small-molecule ligands for macromolecules. Such molecules can be used as biochemical probes as well as potential lead therapeutics. There are several paths that can lead to high-affinity ligands for macromolecules. Molecules that mimic the structures of known substrates can act as lead compounds. Three-dimensional structures of macromolecules determined by X-ray or NMR can be used as a guide to "rationally" design molecules that bind to identified binding sites. If efficient screening methods exist, large collections of compounds can be individually tested until a "hit" is found, which can then be optimized through the synthesis of derivatives. An additional method, that of combinatorial chemistry, has proven itself to be useful for identifying ligands for a range of systems. It has been made possible by the revolution of solid-phase organic synthesis, which allows the rapid synthesis of collections of molecules containing arbitrarily large numbers of species.

An example of this is shown in Fig. 18. It depicts the synthesis of a simple tripeptide library using 10 standard L amino acids. The peptide is constructed on an insoluble resin in a stepwise fashion. If each position during the synthesis is allowed to vary and be any one of 10 amino acids (we will discuss how this is done shortly), there will be a mixture of $10 \times 10 \times 10 = 1000$ possible tripeptides present at the end of the synthesis. The challenge then is to determine which, if any, of the members of this library bind to the target of interest. There are many approaches to this process, known as *deconvolution* of a library. Two of the most widely used approaches have been positional scanning and iterative deconvolution. In both these approaches, mixtures of compounds (sublibraries of the larger library) are tested and the results are used to identify an individual hit or hits.

In the process of positional scanning, sublibraries are constructed by using a specific amino acid at one given position and mixtures of amino acids at the remaining positions (Fig. 18). This is done multiple times, moving the specific amino acid along the length of the peptide. Therefore, at the end of the process for the tripeptide library described above, one will have three collections, each of which contain 10 sublibraries. As shown in the figure, each of the 10 sublibraries contains a mixture of products. This is a mixture of peptides that have a random amino acid at two positions and a specific amino acid at the remaining position. When all of these sublibraries are assayed, if there is a single "winner" to be found in the total library (for the purpose of argument Gly-Pro-Asp), then the subelements that contain that sequence in high concentration will show activity. The sequence of the "hit" can therefore be read directly from the activity results. It can then be resynthesized as an individual species and confirmed as an active molecule. In practice, there can

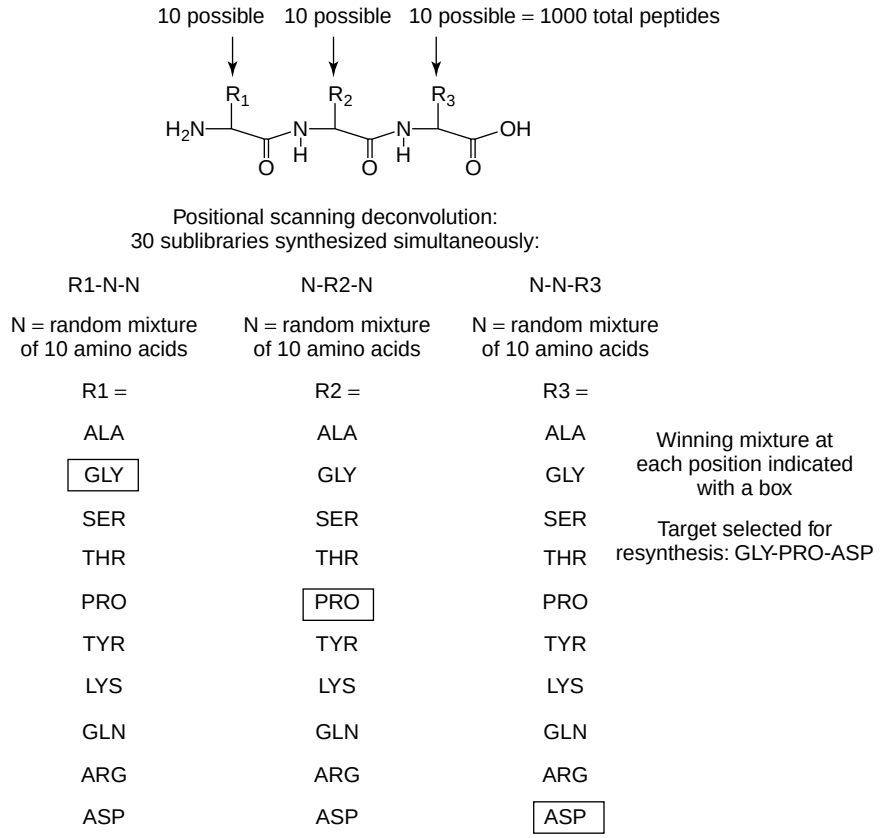

Fig. 18 Deconvolution of a three-position peptide library using the positional scanning approach.

be more than one clear winner at each position in the oligomer. If this is the case, then multiple resyntheses are warranted. For example, if there are two active sublibraries at the first position, two at the second, and three at the third, then $2 \times 2 \times 3$ or 12 new individual molecules should be synthesized as individual species to assess for activity.

A second approach for deconvolution is the method of iterative deconvolution (Fig. 19). This also uses the technique of having multiple random positions and individual fixed positions to identify individual active species from large libraries. Whereas positional scanning allows one to simultaneously synthesize and assay all sublibraries at once, iterative deconvolution requires a series of alternating synthesis and assay steps. Using the example of the tripeptide library above, deconvolution would begin by making a collection of 10 sublibraries in which the first position is fixed and the second two positions have random amino acids. Therefore, one would have a sublibrary of 10 elements. Each of these 10 elements would be assayed and the most active element would identify the specific amino acid at that position. A new collection of

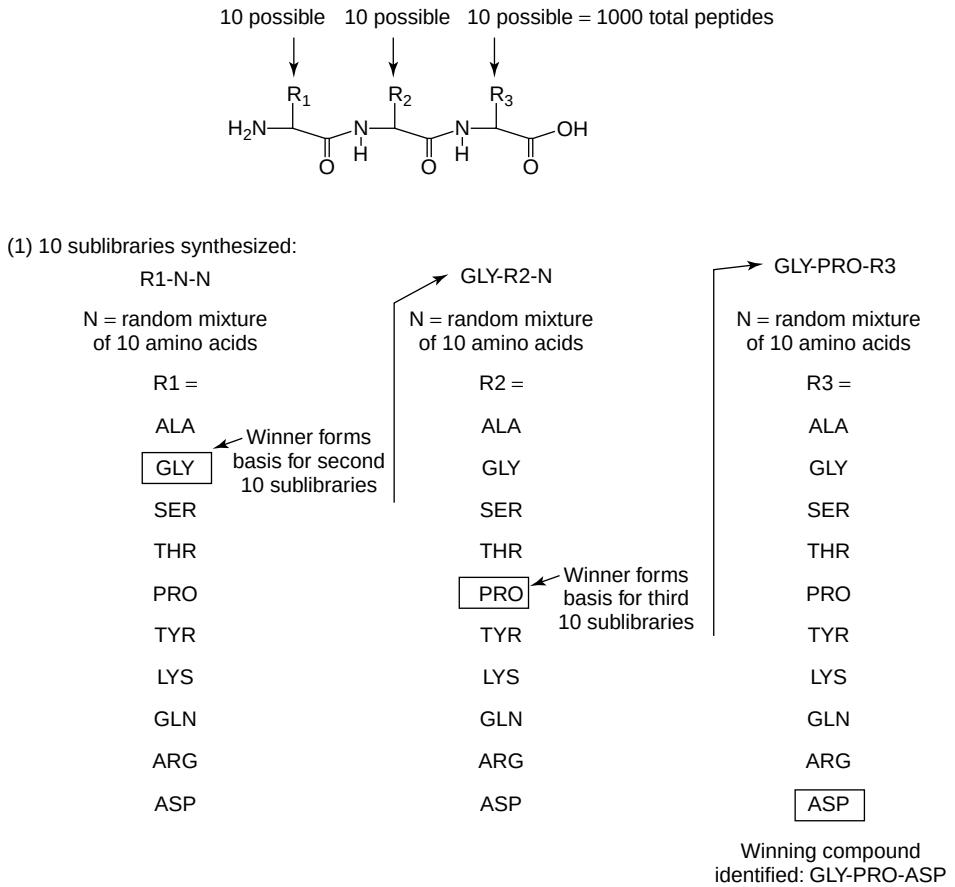

Fig. 19 Deconvolution of the same library from Fig. 18 using the method of iterative deconvolution.

10 sublibraries would then be synthesized, with the specific selected amino acid from the first round in the first position, 1 of the 10 possible amino acids in the second position, and a random amino acid selected from the third position. (See Fig. 19). This process would continue until a single high-activity species was identified from the final sublibrary.

Both these methods have been extensively used to determine the sequence of active species from large combinatorially generated mixtures. There is a range of other deconvolution strategies that have been described and contrasted. Both iterative deconvolution and positional scanning require that some positions in a sequence be randomized. There are two possible ways of doing this randomization. The most conceptually straightforward approach is to simply use a mixture of amino acids when coupling to a given position. One of the disadvantages of this method is that each amino acid in a mixture will have a different coupling rate and efficiency, which depends not only on itself but on the amino acid it is coupling to. The result of this is that some amino acids, if they are particularly reactive, can be overrepresented at the position. A way to avoid

this situation is to use a 1 : 1 ratio of incoming amino acid to resin-bound amine. This then requires longer coupling times to allow the slowest amino acids an opportunity to react. Some of the advantages of solid-phase synthesis are thus lost, as a large excess cannot be employed to drive the reaction.

An alternative to the use of mixtures of amino acids during coupling is the use of so-called split and mix synthesis. This method will work *only* with the method of iterative deconvolution. Using an example of a dipeptide library with three possible amino acids, a typical split and mix synthesis would begin by separating the resin to be used into three equal portions (Fig. 20). To each of these portions, a single specific amino acid would be attached (in the example these are G, P, and T). Because an individual amino acid is being used, it can be used in excess to drive the reaction to completion. (This is not possible when using a mixture of amino acids, without running the risk of more reactive amino acids overpopulating the product mixture.) When these reactions are complete, these three separate portions of resin are combined and mixed thoroughly, and again separated into three separate portions. Again, to each of these portions, a single specific amino acid is attached. These three portions are then again mixed thoroughly and separated into random mixtures. The resultant beads have fully random amino acids at each position and can then be specifically modified with a specific amino acid for use in an iterative deconvolution strategy. The difference between this approach and that of using mixtures of amino acids is the manner in which the random positions were generated. In the case of split and mix synthesis, the random positions are generated by using individual amino acids reacting with, and *then* mixing the resultant beads. The method of amino acid mixtures generates random positions by using mixtures of amino acids reacting at each position.

Both these methods for generating random positions can be effective and have strengths and weaknesses. Split and mix synthesis allows one to avoid the differential reactivity problem that can exist with random mixtures by allowing large excesses to react. Random mixtures on the other hand are easier to manipulate. In addition, for logistical reasons, only they can

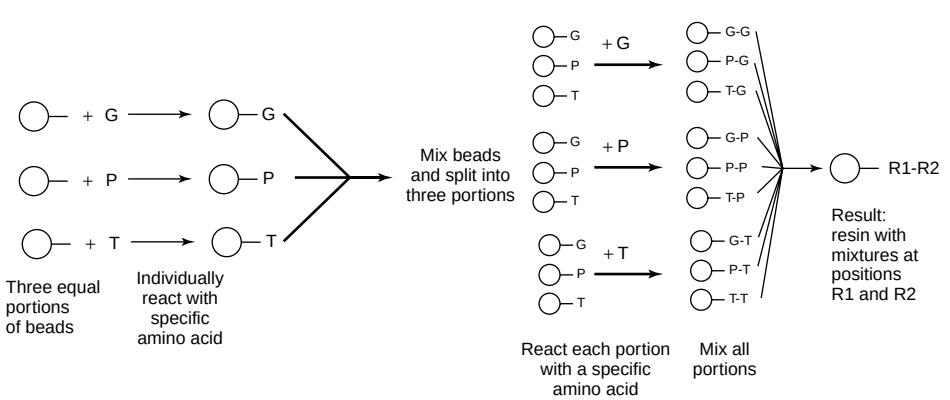

Fig. 20 Using the split and mix approach to generate resin with a bound mixture of all possible dipeptides using the amino acids G, P, and T.

be used to generate randomized positions in positional scanning libraries. (This is because in positional scanning libraries, there are random positions that occur *after* the specific defined positions.) Splitting and mixing would therefore lose the information contained in the defined position. In iterative deconvolution, however, the terminal position is always the defined position, so this issue is not relevant.

6 Molecular Modeling

A very useful tool in the study of bioorganic chemistry is molecular modeling, the simulation of molecules and their interactions. This single designation covers a vast region of study including methods of molecular mechanics, a range of approaches for conformational sampling, semiempirical methods, *ab initio* methods as well as the analysis of complex interactions (which may include all of the previously named techniques). The overall aim of these methods is to simulate properties of molecules, from the small scale (e.g. dipole moments) to the larger scale (e.g. thermal stability of a protein). All of these tools have limitations and, by definition, approximate the behavior of physical systems. At the minimum, it is suggested that a given computational technique be validated with a related system for which experimental evidence can be used to support the technique's application. If no such external validation is available, then the results of a simulation should be used cautiously.

Molecular mechanics is the computationally least-intensive method of estimating the energy of a molecule and the dependency of this energy on conformation. The basis of molecular mechanics is the *force field*, a mathematical representation of a molecule's energy based on an arithmetic sum of different terms. These terms assess the contributions made by bond stretching, bond angles, dihedral rotations, and other movements to the overall energy of the molecule. Figure 21 shows the equation for the AMBER force field. Because the forces that control these movements are simulated using very simple physical models (a stretching spring to represent a stretching bond, for example) the overall energy can be calculated very rapidly. The challenge is *parameterization*, the determination of the factors

$$\text{Energy} = \underbrace{\Sigma K_r (r - r_{eq})^2}_{\text{Bond stretching}} + \underbrace{\Sigma K_\theta (\theta - \theta_{eq})^2}_{\text{Bond angle}} + \underbrace{\Sigma (V_n/2)[(1 + \cos(n\varnothing - \gamma)]}_{\text{Bond dihedral angle}} + \underbrace{\Sigma (A_{ij}/R_{ij}^{12} - B_{ij}/R_{ij}^{6})}_{\text{Van der Waal interactions}} + \underbrace{\Sigma q_i q_j / eR_{ij}}_{\text{Electrostatic interactions}}$$

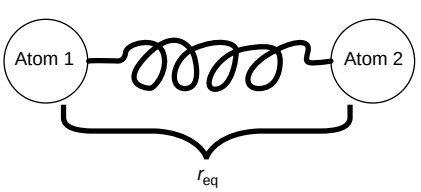

Fig. 21 Sample force field (AMBER). Simple physical model of energy dependence on atomic position: for example, bond stretching modeled as a spring stretch with energy required to displace atoms beyond their equilibrium separation.

that determine the changes in energy due to deviation from the ideal bond length, angle, and so on. The parameters of a force field are determined in multiple ways, including experimental techniques (e.g. ideal bond lengths from crystal structures, bond stretch force constants from spectroscopy) and high-level quantum mechanical calculations where no experimental values are available. There are many force fields available, including MMFF94, CHARM, AMBER, OPLS, MM2/MM3. Each is constructed and parameterized using different approaches and therefore each gives different results for different systems. Again, before overinterpreting the results given by a force field in analyzing a new system, it is important to assess the accuracy of the given force field in predicting experimentally known values in a related system.

Once a molecule is "drawn up" using a modeling package such as Sybyl, Spartan, or Hyper-Chem, it may have nonoptimal bond angles, distances, and so on. Energy minimization can then be used to "relax" the molecule into a local energy minimum. Energy minimization takes a multidimensional derivative of the energy with respect to a given atomic motion and determines the forces on a given atom. These forces sum to push the atoms into their local lowest energy position. This is a local and not a global energy minimum, because it is the nature of energy minimization that the molecule is always going "down-hill" energetically, and therefore it will be unable to traverse energetic barriers that may allow access to a lower energy conformation.

Why would a bioorganic chemist use molecular modeling? Often, it will be to understand what a molecule looks like, to determine, for example, how it may interact with a target macromolecule. For this purpose, it is of greater use to identify the global energy minimum structure, that is, the conformation of the molecule that will have the greatest population. This problem can be approached using a variety of tools, including systematic conformational searching (SCS), Monte Carlo searching, and molecular dynamics. Since a majority of the movement in a molecule is due to rotations around bonds, SCS looks for the global energy minimum by generating all possible conformers by incrementally advancing the dihedral angles of the molecule. This approach will definitely explore all possible conformations (within the resolution of the dihedral angle increment), but the number of conformations necessary to examine increases exponentially with each additional rotatable bond. Eventually, this can become computationally unwieldy.

To address this problem, Monte Carlo searches also rotate bonds in an attempt to traverse high-energy barriers but do so in a random manner that is biased away from conformers that are too high in energy. There is a trade-off at play here: The Monte Carlo search will not be as thorough as a systematic search (except at the highest number of cycles) but will avoid the very highest energy regions of confomer space, which the SCS will be forced to examine.

An additional method of searching for the global energy minimum of a molecule is simulated annealing, a tool of molecular dynamics. In molecular dynamics, the same force field used to model the static energy of a given configuration of atoms is used to determine its dynamic nature. Momentum is imparted to each atom (the amount being dependent on the "temperature" at which the dynamic simulation is run) and the resulting motions of atoms simulated using the force field. In simulated annealing, a high temperature is simulated (i.e. a large amount of momentum applied to the

atoms in the molecule), which results in portions of the molecule traversing energy barriers that could not normally be traversed. The result of this is that the conformations of the molecule are explored. "Cooling" of the molecule in the simulation then allows the molecule to settle into a local energy minimum. With enough rounds of heating and cooling, it is then possible to find the global energy minimum.

Molecular mechanics is an effective tool for predicting conformational energy; however, it cannot describe the electronic nature of a molecule, such as the energy of a transition state associated with the breaking of a bond, or overall reaction energies. For this and other applications where a detailed understanding of the electronic nature of a molecule is needed, quantum mechanical methods are required. The list of approaches is vast, including semiempirical methods (e.g. AM1, PM3), Hartree Fock methods, and correlated methods including Moeller–Plesset and density functional methods. The common element in all of these approaches is that they estimate solutions to the Schrödinger equation. As such, they are able to address questions involving movement of electrons (such as reaction energy), which molecular mechanics is unable to. This fundamental assessment of electron movement and energy in a molecule comes at a cost, which is the increased computational resources demanded of these methods.

A common aim in bioorganic chemistry is the design of small molecules to target macromolecules. Modeling of the structure of the small molecule can be performed using the techniques described above. Of further interest is modeling the potential for interaction of the small molecule with the target of interest. This examination can be done "by hand" in a qualitative fashion using interactive molecular graphics (e.g. programs such as MIDAS, Chimera, Rasmol). In addition, a range of docking programs exists that allows for rapid generation of potential complexes and their assessment via scoring functions. The algorithms used by these programs are fast enough that databases of $>100{,}000$ compounds can be virtually screened for potential interaction with a target macromolecule in a reasonable amount of time ($\sim\leq 1$ week). Docking programs include Dock, Autodock, and FlexX.

The methods described above for modeling a molecule's structure (molecular mechanics, semiempirical, and *ab initio* methods) are in general *in vacuo* methods. In other words, the result is a predicted structure or energy based on the molecule and its conformations while completely desolvated. For some molecules, in fact for most molecules that are of interest to bioorganic chemistry, this eliminates a very important factor from the calculation – the effect of solvent water. The presence of water diminishes electrostatic interactions by screening full and positive charges. In addition, it strengthens nonpolar–nonpolar interactions through the hydrophobic effect. For a true and accurate simulation of molecules in an aqueous environment, solvation energy corrections have to be applied. There are multiple approaches to doing this. One is to incorporate a shell of explicit water molecules in the simulation. This can be computationally very demanding. More efficient methods include so-called continuum methods, which do not include individual water molecules but model the solvent as a continuum. These include the SM (Solvation Model) methods of Cramer and Truhlar.

7
Summary

We have briefly attempted to highlight the principles of bioorganic chemistry that are of general importance and of special interest to the molecular biologist. In addition, we have sought to illustrate these with specific examples. Bioorganic chemistry makes connections with a very broad range of fields, ranging from physical chemistry to biochemistry. Obviously, we have just scratched the surface of these topics in this vital and rapidly expanding discipline.

See also Bioinorganic Chemistry.

Bibliography

Books and Reviews

Benkovic, S.J. (1992) Catalytic antibodies, *Annu. Rev. Biochem.* **61**(29), 29–54.

Bertozzi, C.R., Kiessling, L.L. (2001) Chemical glycobiology, *Science* **291**(5512), 2357–2364.

Breslow, R. (1998) Biomimetic chemistry: a frontier at the chemistry/biology interface, *Chem. Biol.* **5**(2), R27, R28.

Breslow, R., Xu, R. (1993) Recognition and catalysis in nucleic acid chemistry, *Proc. Natl. Acad. Sci. U.S.A.* **90**(4), 1201–1217.

Dolle, R. (2001) Comprehensive survey of combinatorial library synthesis: 2001, *J. Comb. Chem.* **3**(6), 477–517.

Dugas, H. (1996) Bioorganic chemistry: A Chemical Approach to Enzyme Action, 3rd edition. *Springer Advanced Texts in Chemistry* (Cantor, R.C., Ed.), Springer-Verlag, New York.

Fersht, A. (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W.H. Freeman, New York.

Fields, G.B. (1997) Solid Phase Peptide Synthesis, *Methods in Enzymology*, Vol. 289, Academic Press, San Diego.

Hecht, S.M. (1996) *Bioorganic Chemistry: Nucleic Acids. Topics in Bioorganic and Biological Chemistry*, Oxford University Press, New York.

Jung, G. (1999) *Combinatorial Chemistry: Synthesis, Analysis, Screening*, Wiley-VCH, Weinheim, Cambridge.

Muscate, A., Levinson, C.L., Kenyon, G.L. (1993) *Enzyme Inhibitors, in Encyclopedia of Chemical Technology*, John Wiley & Sons, New York, Vol. 9.

Nielsen, P.E., Egholm, M. (1999) An introduction to peptide nucleic acid, *Curr. Issues Mol. Biol.* **1**(1–2), 89–104.

Primary Literature

Anderson, J.C., Magliery, T.J., Schultz, P.G. (2002) Exploring the limits of codon and anticodon size, *Chem. Biol.* **9**(2), 237–244.

Barsky, D., Kool, E.T., Colvin, M.E. (1999) Interaction and solvation energies of nonpolar DNA base analogues and their role in polymerase insertion fidelity, *J. Biomol. Struct. Dyn.* **16**(6), 1119–1134.

Bertozzi, C.R., Kiessling, L.L. (2001) Chemical glycobiology, *Science* **291**(5512), 2357–2364.

Bishop, A.C., Ubersax, J.A., Petsch, D.T., Matheos, D.P., Gray, N.S., Blethrow, J., Shimizu, E., Tsien, J.Z., Schultz, P.G., Rose, M.D., Wood, J.L., Morgan, D.O., Shokat, K.M. (2000) A chemical switch for inhibitor-sensitive alleles of any protein kinase, *Nature* **407**(6802), 395–401.

Boger, D.L., Schmitt, H.W., Fink, B.E., Hedrick, M.P. (2001) Parallel synthesis and evaluation of 132 (+)-1,2,9,9a-tetrahydrocyclopropa[c] benz[e]indol-4-one (CBI) analogues of CC-1065 and the duocarmycins defining the contribution of the DNA-binding domain, *J. Org. Chem.* **66**(20), 6654–6661.

Braasch, D.A., Corey, D.R. (2002) Novel antisense and peptide nucleic acid strategies for controlling gene expression, *Biochemistry* **41**(14), 5403–5410.

Bruner, S.D., Norman, D.P., Verdine, G.L. (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA, *Nature* **403**(6772), 859–866.

Chen, Z., Monia, B.P., Corey, D.R. (2002) Telomerase inhibition, telomere shortening, and decreased cell proliferation by cell permeable 2′-o-methoxyethyl oligonucleotides, *J. Med. Chem.* **45**(25), 5423–5425.

Chin, J.W., Schepartz, A. (2001) Concerted evolution of structure and function in a

miniature protein, *J. Am. Chem. Soc.* **123**(12), 2929, 2930.

Christensen, L., Fitzpatrick, R., Gildea, B., Petersen, K.H., Hansen, H.F., Koch, T., Egholm, M., Buchardt, O., Nielsen, P.E., Coull, J., Berg, R. et al. (1995) Solid-phase synthesis of peptide nucleic acids, *J. Pept. Sci.* **1**(3), 175–183.

Conway, K.A., Rochet, J.C., Bieganski, R.M., Lansbury, P.T. Jr. (2001) Kinetic stabilization of the alpha-synuclein protofibril by a dopamine-alpha-synuclein adduct, *Science* **294**(5545), 1346–1349.

Cramer, C.J., Truhlar, D.G. (1999) Implicit solvation models: equilibria, structure, spectra and dynamics, *Chem. Rev.* **99**(8), 2161–2200.

Daura, X., Gademann, K., Schafer, H., Jaun, B., Seebach, D., van Gunsteren, W.F. (2001) The beta-peptide hairpin in solution: conformational study of a beta-hexapeptide in methanol by NMR spectroscopy and MD simulation, *J. Am. Chem. Soc.* **123**(10), 2393–2404.

Dickinson, L.A., Gulizia, R.J., Trauger, J.W., Baird, E.E., Mosier, D.E., Gottesfeld, J.M., Dervan, P.B. (1998) Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands, *Proc. Natl. Acad. Sci. U.S.A.* **95**(22), 12890–12895.

Eriksson, M., Nielsen, P.E. (1996) Solution structure of a peptide nucleic acid-DNA duplex, *Nat. Struct. Biol.* **3**(5), 410–413.

Feng, S., Chen, J.K., Yu, H., Simon, J.A., Schreiber, S.L. (1994) Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions, *Science* **266**(5188), 1241–1247.

Feng, S., Kapoor, T.M., Shirai, F., Combs, A.P., Schreiber, S.L. (1996) Molecular basis for the binding of SH3 ligands with non-peptide elements identified by combinatorial synthesis, *Chem. Biol.* **3**(8), 661–670.

Fisk, J.D., Gellman, S.H. (2001) A parallel beta-sheet model system that folds in water, *J. Am. Chem. Soc.* **123**(2), 343, 344.

Frackenpohl, J., Arvidsson, P.I., Schreiber, J.V., Seebach, D. (2001) The outstanding biological stability of beta- and gamma-peptides toward proteolytic enzymes: an in vitro investigation with fifteen peptidases, *Chembiochem* **2**(6), 445–455.

Freidinger, R.M. (1989) Non-peptide ligands for peptide receptors, *Trends Pharmacol. Sci.* **10**(7), 270–274.

Gabriel, G.J., Iverson, B.L. (2002) Aromatic oligomers that form hetero duplexes in aqueous solution, *J. Am. Chem. Soc.* **124**(51), 15174, 15175.

Gestwicki, J.E., Strong, L.E., Kiessling, L.L. (2000) Tuning chemotactic responses with synthetic multivalent ligands, *Chem. Biol.* **7**(8), 583–591.

Horlacher, J., Hottiger, M., Podust, V.N., Hubscher, U., Benner, S.A. (1995) Recognition by viral and cellular DNA polymerases of nucleosides bearing bases with nonstandard hydrogen bonding patterns, *Proc. Natl. Acad. Sci. U.S.A.* **92**(14), 6329–6333.

Houghten, R.A., Pinilla, C., Appel, J.R., Blondelle, S.E., Dooley, C.T., Eichler, J., Nefzi, A., Ostresh, J.M. (1999) Mixture-based synthetic combinatorial libraries, *J. Med. Chem.* **42**(19), 3743–3778.

Huizenga, D.E., Szostak, J.W. (1995) A DNA aptamer that binds adenosine and ATP, *Biochemistry* **34**(2), 656–665.

Hung, D.T., Jamison, T.F., Schreiber, S.L. (1996) Understanding and controlling the cell cycle with natural products, *Chem. Biol.* **3**(8), 623–639.

Keefe, A.D., Szostak, J.W. (2001) Functional proteins from a random-sequence library, *Nature* **410**(6829), 715–718.

Kemp, D.S., Boyd, J.G., Muendel, C.C. (1991) The helical s constant for alanine in water derived from template-nucleated helices, *Nature* **352**(6334), 451–454.

Kiessling, L.L., Pohl, N.L. (1996) Strength in numbers: non-natural polyvalent carbohydrate derivatives, *Chem. Biol.* **3**(2), 71–77.

Konings, D.A.M., Wyatt, J.R., Ecker, D.J., Freier, S.M. (1997) Strategies for rapid deconvolution of combinatorial libraries: comparative evaluation using a model system, *J. Med. Chem.* **40**, 4386–4395.

Kostrewa, D., Winkler, F.K. (1995) Mg^{2+} binding to the active site of EcoRV endonuclease: a crystallographic study of complexes with substrate and product DNA at 2 Å resolution, *Biochemistry* **34**(2), 683–696.

Li, T., Janda, K.D., Lerner, R.A. (1996) Cationic cyclopropanation by antibody catalysis, *Nature* **379**(6563), 326, 327.

Liu, B., Han, Y., Corey, D.R., Kodadek, T. (2002) Toward synthetic transcription activators: recruitment of transcription factors to DNA by a PNA-peptide chimera, *J. Am. Chem. Soc.* **124**(9), 1838, 1839.

Liu, L., Rozenman, M., Breslow, R. (2002) Hydrophobic effects on rates and substrate selectivities in polymeric transaminase mimics, *J. Am. Chem. Soc.* **124**(43), 12660, 12661.

MacBeath, G., Kast, P., Hilvert, D. (1998) Redesigning enzyme topology by directed evolution, *Science* **279**(5358), 1958–1961.

Maly, D.J., Choong, I.C., Ellman, J.A. (2000) Combinatorial target-guided ligand assembly: identification of potent subtype-selective c-Src inhibitors, *Proc. Natl. Acad. Sci. U.S.A.* **97**(6), 2419–2424.

Mammen, M., Shakhnovich, E.I., Whitesides, G.M. (1998) Using a convenient, quantitative model for torsional entropy to establish qualitative trends for molecular processes that restrict conformational freedom, *J. Org. Chem.* **63**, 3168–3175.

Mapp, A.K., Ansari, A.Z., Ptashne, M., Dervan, P.B. (2000) Activation of gene expression by small molecule transcription factors, *Proc. Natl. Acad. Sci. U.S.A.* **97**(8), 3930–3935.

Matray, T.J., Kool, E.T. (1999) A specific partner for abasic damage in DNA, *Nature* **399**(6737), 704–708.

McDonnell, K.A., Imperiali, B. (2002) Oligomeric beta(beta)(alpha) miniprotein motifs: pivotal role of single hinge residue in determining the oligomeric state, *J. Am. Chem. Soc.* **124**(3), 428–433.

Morales, J.C., Kool, E.T. (1998) Efficient replication between non-hydrogen-bonded nucleoside shape analogs, *Nat. Struct. Biol.* **5**(11), 950–954.

Moran, S., Ren, R.X.F., Rumney, S., Kool, E.T. (1997) Difluorotoluene, a nonpolar isostere for thymine, codes specifically and efficiently for adenine in DNA replication, *J. Am. Chem. Soc.* **119**, 2056, 2057.

Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C., Schultz, P.G. (1989) A general method for site-specific incorporation of unnatural amino acids into proteins, *Science* **244**(4901), 182–188.

Norman, D.P., Bruner, S.D., Verdine, G.L. (2001) Coupling of substrate recognition and catalysis by a human base-excision DNA repair protein, *J. Am. Chem. Soc.* **123**(2), 359, 360.

Novotny, J., Bruccoleri, R.E., Saul, F.A. (1989) On the attribution of binding energy in antigen-body complexes McPC 603, D1.3, and HyHEL-5, *Biochemistry* **28**(11), 4735–4749.

Orner, B.P., Ernst, J.T., Hamilton, A.D. (2001) Toward proteomimetics: terphenyl derivatives as structural and functional mimics of extended regions of an alpha-helix, *J. Am. Chem. Soc.* **123**(22), 5382, 5383.

Ottesen, J.J., Imperiali, B. (2001) Design of a discretely folded mini-protein motif with predominantly beta-structure, *Nat. Struct. Biol.* **8**(6), 535–539.

Peterson, R.T., Link, B.A., Dowling, J.E., Schreiber, S.L. (2000) Small molecule developmental screens reveal the logic and timing of vertebrate development, *Proc. Natl. Acad. Sci. U.S.A.* **97**(24), 12965–12969.

Saxon, E., Bertozzi, C.R. (2000) Cell surface engineering by a modified Staudinger reaction, *Science* **287**(5460), 2007–2010.

Sefler, A.M., Kozlowski, M.C., Guo, T., Bartlett, P.A. (1997) Design, synthesis, and evaluation of a depsipeptide mimic of tendamistat, *J. Org. Chem.* **62**(1), 93–102.

Stewart, J.D., Benkovic, S.J. (1995) Transition-state stabilization as a measure of the efficiency of antibody catalysis, *Nature* **375**(6530), 388–391.

Swann, S.L., Bergh, J., Farach-Carson, M.C., Ocasio, C.A., Koh, J.T. (2002) Structure-based design of selective agonists for a rickets-associated mutant of the vitamin d receptor, *J. Am. Chem. Soc.* **124**(46), 13795–13805.

Thomas, C.J., Chizhov, A.O., Leitheiser, C.J., Rishel, M.J., Konishi, K., Tao, Z.F., Hecht, S.M. (2002) Solid-phase synthesis of bleomycin A(5) and three monosaccharide analogues: exploring the role of the carbohydrate moiety in RNA cleavage, *J. Am. Chem. Soc.* **124**(44), 12926, 12927.

Trauger, J.W., Baird, E.E., Dervan, P.B. (1996) Recognition of DNA by designed ligands at subnanomolar concentrations, *Nature* **382**(6591), 559–561.

Wang, L., Brock, A., Herberich, B., Schultz, P.G. (2001) Expanding the genetic code of Escherichia coli, *Science* **292**(5516), 498–500.

Wang, L., Schultz, P.G. (2002) Expanding the genetic code, *Chem. Commun. (Camb)* **JAN 7**(1), 1–11.

White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E., Dervan, P.B. (1998) Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands, *Nature* **391**(6666), 468–471.

Wilson, C., Szostak, J.W. (1995) In vitro evolution of a self-alkylating ribozyme, *Nature* **374**(6525), 777–782.

Wilson, D.S., Szostak, J.W. (1999) In vitro selection of functional nucleic acids, *Annu. Rev. Biochem.* **68**, 611–647.

Woods, C.R., Ishii, T., Boger, D.L. (2002) Synthesis and DNA binding properties of iminodiacetic acid-linked polyamides: characterization of cooperative extended 2 : 1 side-by-side parallel binding, *J. Am. Chem. Soc.* **124**(36), 10676–10682.

Ye, X.S., Wong, C.H. (2000) Anomeric reactivity-based one-pot oligosaccharide synthesis: a rapid route to oligosaccharide libraries, *J. Org. Chem.* **65**(8), 2410–2431.

Yu, C., Henry, A.A., Romesberg, F.E., Schultz, P.G. (2002) Polymerase recognition of unnatural base pairs, *Angew. Chem., Int. Ed. Engl.* **41**(20), 3841–3844.

Zimmer, C., Wahnert, U. (1986) Nonintercalating DNA-binding ligands: specificity of the interaction and their use as tools in biophysical, biochemical and biological investigations of the genetic material, *Prog. Biophys. Mol. Biol.* **47**(1), 31–112.

Bioprocess Engineering

Kimberly L. Ogden
Department of Chemical and Environmental Engineering, University of Arizona, Tucson, AZ 85721

1 **Introduction 36**

2 **Enzyme Engineering 39**

3 **Whole-cell Bioreactors 40**

4 **Transgenic Animals and Plants 43**

5 **Transport and Adhesion of Cells 44**

6 **Bioseparations 45**

7 **Concluding Remarks 48**

Bibliography 48
Books and Reviews 48
Primary Literature 48

Keywords

Bioprocessing
The engineering component of the commercial exploitation of biological materials, living organisms, and their activities.

Enzyme
High molecular weight proteins that lower the activation energy of a specific reaction resulting in faster reaction rates.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Genomics
A set of experimental and computational tools that allow genetic information to be read and related to the structure and function of an organism.

Hybridoma
A cell line produced by fusing a myeloma (cancer) cell with a lymphocyte; used to produce antibodies.

Recombinant DNA
A general term for laboratory manipulation in which DNA molecules or fragments from various sources are severed, combined enzymatically, and reinserted into host organisms to make a protein product.

Transgenic
A term describing an animal or a plant created by introducing new DNA sequences into the germ line via addition to the egg or the seed.

Bioprocessing is the engineering component of the commercial exploitation of biological materials, living organisms, and their activities. It is a diverse field requiring the collaboration of scientists, physicians, and engineers working in the areas of agriculture, biology, biochemistry, chemical engineering, electrical engineering, environmental engineering, genetics, material science, medicine, water, and wastewater treatment. Current biotechnology products include pharmaceuticals, fine chemicals, agricultural and food products, and biomaterials. In addition, bioprocessing is being evaluated for the production of nano-bio products, bioelectronics, and commodity chemicals. The use of more biologically based products is growing; global genomic initiatives and miniaturization assure industrial biotechnological growth. The desire to move toward more ''green factories'' that generate less environmental waste also contributes to industrial expansion. The current biotechnology industry is estimated to generate between $35 and $50 billion in sales and continues to be self-sustaining in terms of research and development.

1 Introduction

The roots of bioprocess engineering are in the food, tobacco, tea, leather, and other ancient industries. Of particular interest are the industries that use fermentation, not only in beer brewing and wine making but also in the production of antibiotics, citric acid, various amino acids, and other biomolecules. The art of yeast and bacterial fermentation is therefore well established. Throughout the 1990s, cell culture technology grew significantly as growth hormones, interferons, monoclonal antibodies, and artificial tissue

products entered clinical trials. Current frontiers include stem cell research for tissue, organ, and bone replacements; integrated fermentation and downstream process development; and production of protein from transgenic plants and animals.

Another important precursor of present-day bioprocessing is the fractionation of human plasma proteins for the production of its components: mainly, serum albumin, immunoglobulins, fibrinogen, and various clotting factors. Use of purified proteins is a far more efficient utilization of human plasma – an expensive national resource – than total plasma transfusion. The industry originated during World War II and is based on fractional precipitation of plasma component proteins by the addition of ethanol under controlled conditions of temperature, pH, and ionic strength. The ethanol process developed by Cohn and his numerous associates is used worldwide with remarkably few changes since its introduction.

Modern purification of therapeutic proteins that are produced through recombinant DNA technology is not based on ethanol precipitation; rather, it relies mainly on chromatography. Nevertheless, protein purification benefited enormously from the older plasma fractionation industry, which developed such essential ancillary equipment and processes as centrifuges, filters, freeze-dryers, sterile operation, pyrogen testing, and quality control. Plasma fractionation still handles much larger volumes of proteins and at a lower cost than any other form of therapeutic protein processing.

The use of radioisotopes and electron microscopes, also around World War II, contributed significantly to advances in bioprocessing. The use of these techniques to track isotopes in the body and to visualize cell components when combined with silicon transistors in the 1950s lead to the advanced diagnostic and analytical tools of today including nuclear magnetic resonance machines, gas and liquid chromatographs, ultrasound, laser and magnetic imaging techniques, and X-ray diffraction tools. These advances in instrumentation allowed for a switch from "trial-and-error" medicine and products to physiology-based biotechnology.

The current bioprocessing industry relies on several complementary technologies. The backbone of the industry is the use of recombinant proteins, obtained by inducing a host organism to express foreign proteins. To this end, a gene regulating the production of this protein is inserted into a plasmid, an autonomous, self-replicating piece of DNA that is inserted into a living cell. The host organism can be a bacterium, yeast, plant, mammalian, or insect cell, as well as an intact plant or higher animal. The first recombinant plasmid was patented by Cohen and Boyer in 1973. The plasmid is constructed to cause a cell to manufacture a large quantity of foreign or product protein (up to 50 g of protein per liter of cell medium). The excess protein is, in many cases, expressed in a denatured form as a densely aggregated particle. This step facilitates the purification of the protein, but it then requires solubilization and refolding in the biologically active form – an often daunting problem. Obtaining the active protein usually dictates the choice of plasmid and host cell.

The formulation of the original gene (i.e. the sequence of the nucleotides forming the DNA) is greatly facilitated by the polymerase chain reaction (PCR) method, first developed in the mid-1980s by the Cetus Corporation and commercialized by Perkin-Elmer. PCR is an enzymatic

process that is carried out in discrete cycles of amplification, each cycle doubling the amount of available DNA. Thus, minute quantities of a rare target sequence can be amplified to produce adequate quantities with excellent faithfulness. PCR is used to amplify DNA for protein products as well as for detection of trace quantities of contaminants. Specific sequences of DNA can also be chemically synthesized using solid-phase procedures.

A second important technology is the production of monoclonal antibodies. Sensitized antibody-producing spleen cells are hybridized with stable cancer cell lines; the so-formed hybrid has the longevity of typical cancer cells in tissue culture and the specificity of the original splenocyte. The reproducible character of the monoclonal antibodies has permitted the development of new diagnostic procedures of exquisite specificity, new protein separation specificity, and new therapeutics such as Herceptin, an antibody used to block the overexpressed Her2 receptor in breast cancer patients. While recombinant proteins can be expressed in all types of host cells like bacteria and yeast, the production of hybridoma cells requires more stringent and far more expensive mammalian tissue culture procedures.

A third major branch of biotechnology is the direct chemical synthesis of peptides using the solid phase methodology first developed by Bruce Merrifield, a Nobel laureate. The first amino acid of a desired peptide sequence is coupled to a latex particle, and subsequent amino acids are coupled to it, step by step, in highly automated instruments. The direct synthesis yields an easier approach for the synthesis of families of peptides differing in only a single amino acid, a point mutation. If the future holds more personalized pharmaceuticals through the field of pharmacogenomics (study of polymorphisms in drug-metabolizing enzymes and the resulting differences in drug effects), then the direct synthesis method may be the production method of choice.

A fourth major branch of biotechnology is genomics. Now that the human genome is sequenced, a huge amount of controversy surrounds the use of the information. Genes for diseases are rapidly being discovered, and medicines and therapies are based on knowledge that is more complete compared to the trial and error methods of the past. The genome sequence in combination with new techniques in combinatorial chemistry, a greater understanding of receptors, and improved data management tools using the fastest computers is opening up an entirely new type of bioprocessing industry. However, there is a huge amount of resistance from a variety of religious organizations fearing the loss of human individuality, and many other government agencies and individuals worldwide that are scared of transgenic plants and animals after a couple of incidents. Laws were passed or are under consideration in many countries to limit the use of genetics to protect human tissue use and track genetically modified plants from farm to consumer.

Detailed descriptions of all these and other current processes are beyond the scope of this brief review. Other articles within this encyclopedia discuss some in detail. Therefore, the key features of only five aspects of bioprocessing are presented: enzyme engineering, whole-cell reactors, transgenic plants and animals, transport and adhesion of cells, and bioseparations or downstream processing.

2
Enzyme Engineering

Enzymes are an alternative to viable cells for synthesis of biological products. The applications vary from alternative fuel synthesis to environmental remediation to biosensors to food processing to textiles. Enzymes themselves are also bioprocessing products. With the increased knowledge of recombinant DNA technology, rare enzymes are now made easily in large quantities and are obtained from organisms grown in unusual environments like hot springs or the ocean. One increasing area of enzyme use is in the separation of chirally pure compounds. Typically, one enantiomer in a mixture is often not useful while the other has therapeutic value. This unique enzyme function leads to the development of processes that use both chemical and enzymatic synthesis for the production of new pharmaceuticals.

The most common reactor configuration for enzyme synthesis of a product is a packed or fluidized immobilized enzyme reactor. This system has the advantage of being a continuous process able to handle high throughputs of process streams. Enzymes are reasonably protected in this environment. Cellulase is an example of an enzyme that is used in an immobilized configuration for the conversion of biomass feedstock in ethanol production for alternative fuels. In addition, these systems show promise for environmental remediation to selectively remove heavy metals or transform hazardous organic substances into nontoxic compounds.

Enzyme kinetics for simple enzyme-catalyzed reactions are often referred to as saturation or Michaelis–Menten kinetics. The reaction mechanism describing this process is

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E} + \mathrm{P}$$

where E is the enzyme, S is the substrate, P is the product, ES is the enzyme–substrate complex, and the k values are the reaction rate constants. The scheme involves a reversible step for formation and dissociation of the enzyme–substrate complex. It assumes that the second reaction, product generation, is irreversible, which is a good assumption when product accumulation is negligible at the beginning of the reaction. To develop the rate expression for enzyme-catalyzed reaction, either a rapid equilibrium approach or a quasi-steady state approach is used. Either way, the resulting rate expression is

$$v = \frac{\mathrm{d}P}{\mathrm{d}t} = \frac{V_m \mathrm{S}}{K_m + \mathrm{S}} \tag{1}$$

where v is the reaction rate, t is time, V_m is the maximum forward reaction rate, and K_m is the Michaelis–Menten constant. This rate expression is commonly used to describe enzyme-catalyzed reactions. Many enzymatic reactions are described by Michaelis–Menten kinetics. The rate constants are usually determined from initial rate experiments.

Enzymes are effected by pH and temperature. Typically, an enzyme is active only over a certain pH (1 to 2 pH units) and temperature range (5 to 10°). The pH and temperature effects are incorporated into the enzyme kinetic expression when needed. Enzymes are also inhibited by many heavy metals or organic materials. The inhibitors bind to the active site of the enzyme directly (competitive inhibition), to another site on the enzyme, thus reducing the enzyme affinity for the substrate (noncompetitive inhibition), or to

the enzyme–substrate complex (uncompetitive inhibition). All these cases result in reduction of the enzyme-catalyzed reaction rate. In addition, the substrate itself or the product formed may be inhibitory, also resulting in a decrease in reaction rate.

When immobilized reactors are utilized, the reaction may be diffusion limited. A concentration gradient causes the substrate molecule to move, and this movement may be slower than the enzyme-catalyzed reaction rate. The diffusion rate is a function of temperature, pressure, and the composition of the system. If a porous support is used, the following equation is used to describe the enzyme kinetics:

$$D_e\left(\frac{d^2S}{dr^2}+\frac{2\,dS}{r\,dr}\right)=\frac{V_mS}{K_m+S} \qquad (2)$$

where D_e is the effective diffusivity of the substrate within the porous support matrix, and r is the radius of the porous support. This expression assumes the following:

- The system is at steady state.
- Flux is by diffusion only, not convection.
- Diffusion is governed by Fick's law.
- The enzyme-catalyzed reaction follows Michaelis–Menten kinetics.
- The substrate does not adsorb to the support.
- The support matrix has a constant void volume.
- The enzyme is evenly distributed within the support matrix.

Enzyme technology is exploited in the area of biosensors. Enzyme-coated electrodes provide a highly selective and sensitive method for determining the amount of a given substrate. Examples of electrodes include sensors for fermentation products and intermediates (e.g. amino acids, lactic acid, penicillin, alcohols). Current research involves coupling enzyme reactions to improve sensitivity by eliminating interfering signals, as well as improving enzyme stability by using enzymes isolated from organisms that live in extreme environments like hot springs or oceans. The use of these specific electrodes facilitates the control of bioprocesses, resulting in higher productivity. Biosensors are also playing an increased role in medical applications like the detection of glucose in blood and urine and detection of scarce enzymes such as tissue plasminogen activator. Future use will also involve detection of biological and chemical contaminants in the environment and in water supplies as security tightens around the world.

For everyday use, the food and detergent industries purchase the largest amounts of enzymes. Proteases that hybridize proteins into smaller peptide units constitute the largest part of the enzyme market. They are used in cheese making, baking, meat tenderizing, brewing, tanning, and detergents for the hydrolysis of protein stains. Other enzyme uses include medicine (penicillinase removes penicillin from an allergic individual and trypsin is used as an anti-inflammatory agent); pulp and paper manufacture (lignases biopulp wood); and the textile and leather industries. The issue becomes one of how to process or cheaply make large quantities of the enzymes. This leads to the next area of the bioprocessing industry – live cell bioreactors.

3 Whole-cell Bioreactors

Living cells can be viewed as small biochemical reactors of great complexity. They are utilized in most aspects of biotechnology, and most biotechnology

processes begin with a bioreactor. Despite the advances in enzyme technology, it is important to keep in mind that enzymes come from cells; thus, whole-cell bioreactor technology will be important for enzyme production as well as for production of vaccines, insecticides, antibodies, hormones, immunobiological regulators, and artificial tissue.

The roots of fermentation technology are traced back to many centuries. The earliest bioreactors or fermentors were used in the brewing and wine-making industries. During the last 50 years, fungal fermentations have been used to make antibiotics, beginning with penicillin. Recombinant DNA technology, monoclonal antibody production, and genomic research continue to push the whole-cell reactor technology forward. The first major products made by recombinant organisms were insulin (Eli Lilly), tissue plasminogen activator (Genetech), and erythropoietin (Amgen). Today the major question that remains is how to choose the best host–vector system. The most common hosts for commercial production are the bacterial strain, *Escherichia coli*, and the Chinese hamster ovary (CHO) cells. A comparison of the advantages and disadvantages of using these two types of cell systems is given in Table 1. Overall, the choice of the host–vector system is very complicated and the "ideal" system does not as yet exist. The characteristics of the protein product and cost are the critical values. For example, if a bulk enzyme for detergents is being produced, some impurities are okay, and bacterial

Tab. 1 Comparison of bacterial and mammalian cell host–vector systems for product synthesis.

E. coli cultures		*CHO cell cultures*	
Advantages	***Disadvantages***	***Advantages***	***Disadvantages***
High growth rate	Formation of inclusion bodies	Availability of host and vector systems	Slow growth
Wide range of host backgrounds	Misfolded protein products	Excretion of product	Expensive, complex media
Wide range of vectors and promoters	No ability to glycosylate	Ability to produce authentic proteins (correct folding, glycosylation, and posttranslational processing)	Low protein expression levels
High expression levels and cell densities	Proteolytic activity	Low proteolytic activity	Shear sensitive cells
Low-cost media	Usually low levels of product excretion		Vectors are derived from virus and there is fear of reversion to pathogenic form.
	No posttranslational processing		Must assure removal of nucleic acids from hybridoma lines
	Must assure removal of endotoxins during purification		Most strains are mortal so dead cells must be removed.

or yeast cultures are more often used. Large protein molecules and antibodies are made only in mammalian cultures and product purity is paramount if the product is an injected biopharmaceutical. Compounds such as insulin or tissue plasminogen activator are made in either system – bacterial or mammalian.

In addition to *E. coli* and CHO cells, other types of bacterial cells, yeast, plant cells, and insect cells have been studied for synthesis of bioproducts. Gram-positive bacteria such as *Bacillus subtilis* excrete proteins better but make more proteases that rapidly degrade the protein products. The yeast strain, *Saccharomyces cerevisiae*, is used in the food industry and the genetics are reasonably well understood; however, it tends to express only low levels of a foreign protein and hyperglycosylates the product. Plants themselves offer the advantage of diversity. As much as 25% of today's pharmaceuticals (primarily nonprotein products) are extracted from plants. Plant cell cultures do allow for more control than using an intact plant; however, genetic knowledge is less than what is known for bacterial and animal cells, and product expression levels are low. Only a few plant cell systems are used commercially in Japan and Germany. The most well known product made in plant cell culture is the anticancer agent, paclitaxel or Taxol. The insect cell–baculovirus system is used primarily as a research tool and for small-scale studies (100 L). The advantages of this system are that it does allow for high expression of foreign protein and offers potential safety advantages, since virus that affect insect cells do not affect humans. However, the insect system does not quite mimic the mammalian cell system in that the protein product may have slight structural modifications, which are useful when making vaccines but not necessarily so for complex products.

Optimization of recombinant protein production is a continuing objective of industrial and academic research. Engineering and science work together to reach this objective. For instance, a common problem is genetic instability (segregational, structural, host cell gene mutation, and/or growth rate dominated), since the overproduction of foreign proteins is always detrimental to cell growth and survival. Novel reactor strategies developed to alleviate segregational instability include selective recycle reactors in which productive plasmid-bearing cells are selectively recycled to a fermentor through flocculation and size separation, while nonproductive rapidly growing cells and/or dead cells are removed from the reactor. Thus, a productive, continuous reactor can exist "in theory" for infinite time. Cells have been genetically manipulated to alleviate this problem as well, by inserting antibiotic resistance genes into the plasmids and supplying antibiotics to the medium. Research continues to find the "ideal" host–vector system, which rapidly grows and expresses high levels of foreign protein that are excreted. The ideal system also does not produce proteases, and folds and glycosylates the protein product as required.

The newest area of bioreactor research and development is the growth of tissues in biofilm or scaffold reactors for regeneration of tissues. Great strides continue to be made. For example, it is now possible for a patient's cells to be shipped to a company, cultured in a scaffold bioreactor and subsequently reimplanted. Regenerative and rejuvenating therapies may become common in the near future. Human genes, proteins, antibodies, and cells are used in combination to replace, repair, and restore

tissue that is damaged by disease, injures, or old age. Tissue cell reactors for these purposes require precise control compared to those used for the production of food or pharmaceutical products, but the principles of operation build upon the knowledge learned from existing whole-cell bioreactors. A more controversial area of research is the use of stem cells for medicinal purposes. The ethical issues surrounding this type of work may slow its development but these areas will be an important part of the next few decades of research.

Various models ranging from unstructured, nonsegregated to structured, segregated are used to describe cell growth. The most common kinetic expression is the Monod equation:

$$\mu = \frac{\mu_{\max} S}{K_s + S} \tag{3}$$

where μ is the specific growth rate, $\mu_{\max}$ is the maximum specific growth rate, K_s is the saturation constant, and S is the substrate. This growth expression is inserted into batch, continuous, biofilm, scaffold, or immobilized reactor equations to predict the rate of utilization of substrate, cell concentration, and growth-associated product concentration. The semi-empirical expression assumes that a single chemical species, *S*, is limiting, while changes in other nutrient concentrations have no effect; and that a single enzyme system with Michaelis–Menten kinetics is responsible for the uptake of *S*. Although this premise is seldom true, the Monod equation is used routinely to describe bioreactor behavior of everything from well-defined recombinant bacterial or mammalian systems to wastewater treatment systems. Growth rates are affected by temperature, pH, and media composition.

Chemically structured models provide a more general approach with greater predictive power by relating cell growth and product production not to just one substrate, but to nitrogen, carbon, and oxygen uptake, as well as including expressions that relate important kinetic interactions among cellular subcomponents such as RNA, DNA, lipids, and proteins. These more sophisticated models predict growth rates better and, thus, reactor behavior, but consist of between 4 and 40 equations. Segregated models typically differentiate between productive and nonproductive cells within one reactor system. The very complex models are not typically used by industry but are used in some research laboratories.

4 Transgenic Animals and Plants

A technology that has made great strides in the past decade is the use of transgenic animals and plants for production of proteins. The living animal or plant becomes the "bioreactor". The use of transgenic animals is less developed than that of plants. Animals have the advantage of performing complex posttranslational processing steps that cannot be done in animal cell culture. How it works is that new genetic information is inserted into the embryo of the animal and the nontoxic protein is expressed by the mature animal – typically in the milk. High concentrations of complex proteins are achieved, and the processes can be made cost effective. Sheep, goats, and pigs are the primary species being evaluated for this task. Harvesting the milk and downstream processing technologies are areas of current study. In addition, there are some serious concerns about the use of

animals to produce proteins. The primary concern is safety. Besides the possibility of having a negative health effect on the animal, there is fear regarding mutation of animal viruses that jump species and become a serious health issue for humans. The fear is real since some viruses such as AIDS and SARS that still lack effective treatment are believed to have originated in animals. Screening of animals for viruses is an extremely expensive undertaking and may limit the cost effectiveness of their use for production.

Transgenic plant technology holds great promise and products are currently in Phase I clinical trials. Large companies including Monsanto, Dow Chemical, and Sigma-Aldrich are teaming with smaller companies such as Prodigen, Inc. and Epicyte Pharmaceutical to manufacture antibodies and proteins in transgenic plants. Several crops including corn, tobacco, potatoes, and soybeans have been studied for production of antibodies and proteins. Products tend to degrade within the tobacco leaf and potato; protein separation from soybeans is more challenging; but corn results are very promising. The product, especially antibodies, concentrates in the corn kernels, and is easily separated. The demand is for "green" engineering processes. The desire is to eliminate environmental hazardous waste generation altogether. Thus, a green plant that works as a bioreactor is an extremely attractive, environmentally friendly alternative. The systems are also safe in terms of virus transmission, since plant viruses do not survive in animals. One concern is the ability to keep the plants isolated from other crops. The pharm plant products, as they are being called, cannot be allowed into the food chain. So currently these plants are not grown in any food producing areas. The tremendous advantage of using pharm plants to make antibodies and other products is that the estimated production costs are an order of magnitude less than the cost of production using mammalian cells. This savings will push the development of transgenic plant products forward.

5
Transport and Adhesion of Cells

Cellular and viral transport and adhesion are important in immobilized reactors, the development of new drugs, tissue repair, drug delivery, *in situ* bioremediation efforts, and security issues related to bioterrorism. Understanding how cells are able to recognize and adhere specifically to insoluble extracellular tissue matrix proteins and to other cells is crucial to many physiological processes. Cells of all types – blood, bacterial, tumor, and endothelial – respond to environmental stimuli through receptors found on their cells surfaces. For example, a cell will have a receptor on its surface for epidermal growth factor; binding and internalization of the growth factor then stimulates DNA synthesis. Cell receptors are responsible for cell adhesion to surfaces such as tissue, scaffold reactors, and for self-agglutination. White blood cells adhere to ligands on vessel walls, or mammalian cells adhere to surfaces coated with ligands through integrins. Cell receptors are also responsible for chemotaxis – the migration of cells in response to a specific chemical gradient. Chemotaxis is of particular importance in the development and repair of inflammations.

Understanding the mechanisms of intra- and intercellular transport is the key to the development of site-directed or target-specific drugs. For instance,

monoclonal antibodies may be used to deliver drugs to cancer cells only through specific ligand–receptor binding. Gene therapy relies on a better understanding of developing vectors to deliver genes and proteins to specific targets, to deliver vaccines, for example. Animal models are improving and the future may hold more gene therapy for humans, although this is another area of controversial research and application.

Understanding cellular transport also aids in environmental remediation. Once microorganisms have been injected into the environment, they must move from the point of injection to the contamination site. Cell receptors known as *pili* or *fimbriae* are responsible for some cell adhesion to the soil, minerals, and other contaminants in soil. Polysaccharide production is also responsible for cell adhesion. If pilin and polysaccharide synthesis and binding mechanisms are understood, the efficiency of using biodegradation to remediate soil and water can be determined through process engineering analysis and compared to other existing technologies such as chemical oxidation and incineration.

A growing area of concern is the understanding of pathogen transport through water systems. Once a bacterial or viral pathogen enters a water supply, it is important to understand how its transport can be limited. Efficient, cost-effective alternatives for pathogen death and removal are currently being investigated, as well as the fundamental principles influencing pathogen transport such as the effect of minerals, organics, and temperature. Since the environment in which we live is complex and changing, how a pathogen behaves in a particular environment is still not well understood.

6 Bioseparations

Although a large quantity of work has been done on fermentation and other upstream processing, the main cost of making a biotechnological product is in the downstream processing or separation steps. Improved methods are always needed, since most of the separation equipment employed are simple scale-ups of analytical and micro-preparative techniques pioneered in life sciences laboratories.

After a product is made by a cell or a plant, it needs to be separated from the rest of the by-products. The standard industrial separation techniques for separating cells from spent media entail either centrifugation or cross-flow filtration. If the cell product is an intracellular product, the cells must be lysed through homogenization, osmotic shock, or bead milling. Then the cell debris is removed by centrifugation or filtration. Bacterial products tend not to secrete protein products and require this step, whereas mammalian cells typically do secrete. The expensive part is further purification of the product. For soluble protein products, all of the cellular protein is usually precipitated as a first step by changing the temperature or ionic strength or by adding a solvent such as ethanol. The purity and efficiency of the process is governed by thermodynamics – in particular, by the chemical potential of the precipitant and the solution. A number of technologies exist for further product separations, depending on the purity of the product required. These include affinity, ion exchange, reversed phase, size exclusion and other forms of chromatography; electrophoresis; liquid–liquid extraction; membranes; and distillation. There are process design heuristics for purification/separation,

which include steps such as separating the most plentiful impurities first; performing the most difficult and costly separations last; and sequencing processes to exploit different separation driving forces such as size, charge, and differences in boiling points.

Chromatography is defined as the separation of substances by the selective binding of components to porous, solid, sorptive media. Various types of chromatographic separation are employed including adsorption, liquid–liquid partition, ion exchange, gel filtration or molecular sieving, affinity, hydrophobic and high performance liquid chromatography or HPLC. Typically, a combination of chromatographic processes is used to purify proteins. For example, an ion exchange column separates negatively charged products from positively charged and neutral products, whereupon the desired products are isolated by means of an affinity column. Affinity chromatography is particularly versatile, often displaying exquisite selectivity. Various affinity systems are used such as antigen–antibody, lectin–carbohydrate moiety, enzyme–substrate, immobilized metal ion-histidine-containing protein, receptor–hormone, and so on. In addition, affinity chromatography may be used for environmental remediation using ligands (chelators or macromolecules) that specifically bind metal ions.

Since chromatography involves the adsorption of species to be purified (i.e. a protein) to a support for separation, the process is often described by either Freundlich or Langmuir isotherms. An empirical formula that describes the adsorption of many antibiotics, hormones, and steroids is the Freundlich isotherm:

$$q = K\gamma^n \tag{4}$$

where q is the amount of solute (product or contaminant) adsorbed per amount of adsorbent, γ is the solute concentration in solution, K is similar to an equilibrium constant, and n is an empirical constant. The values of K and n are determined experimentally. If the adsorption is favorable, then $n < 1$; if it is unfavorable, $n > 1$. The Langmuir isotherm is not empirical; instead it has a theoretical basis:

$$q = \frac{q_0\gamma}{K + \gamma} \tag{5}$$

The values of q_0 and K are determined experimentally. This equation assumes that there are a specific number of active sites; the amount of sites for a given adsorbent is reflected in the value of q_0. At low solute concentrations, $q = q_0$. The Langmuir equation also assumes that an equilibrium is reached between adsorption and desorption to the active sites. Other assumptions are that adsorption takes place monomolecularly and is localized. The only disadvantage of chromatographic processes is that large columns are sometimes required that result in large pressure drops over the columns, requiring large auxiliary equipment like pumps.

Electrophoresis is used for the separation of charged molecules based on differences in both size and charge in an electric field. The separation is accomplished by balancing the drag force of a charged particle with the electrostatic forces when the particle is moving with a constant velocity. The net charge, q, on a protein determines its velocity, V_t, and thus its separation:

$$V_t = \frac{qE}{3\pi\mu D_P} \tag{6}$$

where E is the electric field intensity, D_P is the particle diameter, and μ is the viscosity. The pH of the solution greatly affects the

separation. The major drawback of traditional electrophoresis is lack of scalability. Thus, electrophoresis has not made significant inroads into protein downstream processing, although two research-scale instruments are commercially available. Both are based on isoelectric focusing in free solution and do not need gels or other supporting matrices. Preparative isoelectric focusing does not require expensive buffers and is more easily scaled up.

Liquid–liquid extraction originally used immiscible liquids such as an organic and an aqueous phase. For example, an aqueous DNA solution is extracted with phenol and ether to remove lipids and other cell debris. Extraction is used to separate inhibitory end products such as alcohols from fermentation broths, or to recover antibiotics. However, organic solvents are not compatible with many proteins. Instead, protein separation is accomplished by means of aqueous two-phase extraction. Typical aqueous phases contain water-soluble but mutually incompatible polymers such as polyethylene glycol and dextran. The latter polymer, or any other bottom phase polymer, such as cellulose derivatives, may be substituted by highly concentrated salt solutions.

Partitioning of the solute or protein is described using a partition coefficient, K:

$$K = \frac{x}{y} \tag{7}$$

where x is the concentration of protein in the lighter liquid phase and y is the concentration of the same protein in the heavier liquid phase. The value of K is governed by thermodynamics; the logarithm of K is proportional to the difference in the chemical potentials of the standard states of heavy and light liquids. Factors that affect the partition coefficient include temperature, ionic strength, and pH.

Membranes are used to concentrate products and to clarify solutions; they are used to separate proteins based on size. Membranes are used to purify water through reverse osmosis. Filtration is generally governed by Darcy's law, which relates the velocity of the liquid through porous media, v, to the pressure drop ΔP:

$$v = \frac{k\Delta P}{\mu l} \tag{8}$$

where k is the permeability constant, which is a function of the filtrate, μ is the viscosity of the liquid, and l is the bed or membrane thickness. Darcy's law holds for low Reynolds number flow. The major drawback of membranes is biofouling. Layers of proteins build up on the surface and clog the membranes. Cross-flow or tangential-flow filtration decreases fouling, but does not eliminate it.

Other separation processes include crystallization – typically the last step in producing highly purified products like antibiotics, and drying to remove any solvents. The focus for the future continues to be combining techniques such as chromatography with membranes or chelators with liquid–liquid extraction. Furthermore, clever methods to combine fermentation and separation are always being investigated, since optimization of both will lead to the most cost-effective process.

Finally, an essential part of all downstream processing is quality control – the assessment of sterility, absence of genetic materials such as viruses or nucleic acids, and absence of pyrogens, as well as the purity of the final product such as the desired protein or antibody. Various chromatographic, mass spectrometric, and electrophoretic processes are routinely used for quality assurance. These techniques, when combined with

immunochemical reagents, allow for product and contaminant identification. The advent of multidimensional nuclear magnetic resonance, DNA hybridization, and Fourier-transform ion cyclotron resonance in tandem with large advances in computer technology significantly impact the ability to perform bioanalytical analysis. These techniques, while typically used more during the research and development stage, open the door for better methodology for product purity and safety.

7 Concluding Remarks

Currently, biopharmaceuticals constitutes only about 8% of the worldwide drug market, and is expected to grow to reach about 15% of the drug market by 2006. However, biopharmaceuticals is only one part of the bioprocessing market. Bioprocessing involves the collaboration of scientists and engineers from industry, government agencies, and universities. Systems biology is the wave of the future to evaluate all multivariable components simultaneously. Thrusts for expansion include developing combined reactor–separator equipment, pharm plant systems, human therapeutics, better and more vaccines and antibiotics, commodity chemicals, bioremediation, proteomics, instrumentation and software, ''biochips'', and gene and rejuvenative therapies. Legislation addressing the use of stem cells, gene therapy, transgenic animals, environmental remediation and emissions, and genetically engineered foods will have a tremendous impact on the future of the bioprocessing industry. Advances in current bioprocess techniques will be of essence if the cost of production is to be reduced. The competition is international and there are currently many company mergers. All are trying to be at the forefront of bioprocessing technology.

See also Gene Targeting; Genetic Engineering of Vaccines; DNA Replication and Transcription.

Bibliography

Books and Reviews

Flickinger, M.C., Drew, S.W. (Eds.) (1999) *Encyclopedia of Bioprocess Technology: Fermentation, Biocatalysis and Bioseparation*, John Wiley & Sons, New York.

Ladish, M.R. (2001) *Bioseparations Engineering: Principles, Practice and Economic*, Wiley-Interscience, New York.

Primary Literature

Adams, M.W.W., Kelly, R.M. (1995) Enzymes from microorganisms in extreme environments, *C&EN* **73**, 32–42.

Bailey, J.E. (1991) Towards a science of metabolic engineering, *Science* **252**, 1668–1675.

Bailey, J.E. (1998) Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities, *Biotechnol. Prog.* **14**, 8–22.

Belfort, G., Davis, R.H., Zydney, A.J. (1994) The behavior of suspensions and macromolecular solutions in crossflow microfiltration, *J. Membrane Sci.* **96**, 1–8.

Bergmeyer, H.U. (1974) Determination of Enzyme Activities, in: Bergmeyer de, H.U. (Ed.) *Methods of Enzymatic Analysis*, Vol. 1, Verlag Chemie, Weihmeim, pp. 121–130.

Borman, S. (2000) Proteomics: taking over where genomics leaves off, *C&EN* **79**, 31–37.

Bull, J.B., Breese, K. (1978) Interaction of alcohols with proteins, *Biopolymers* **17**, 2121–2130.

Bylund, F., Guillard, F., Enfors, S.O., Tragardh, C., Larsson, G. (1999) Scale down of recombinant protein production: a comparative study of scaling performance, *Bioprocess Eng.* **20**, 327–389.

Chalmers, J.J. (2000) Animal Cell Culture: Effects of Agitation and Aeration on Cell Adaptation, in: Speir, R., Griffiths, J.B., Seragg, A.H. (Eds.) *Encyclopedia of Cell Technology*, Wiley, New York.

Characklis, W.G., Marshall, K.C. (1990) Biofilms: A Basis for an Interdisciplinary Approach, in: Characklis, W.G., Marshall, K.C. (Eds.) *Biofilms*, John Wiley and Sons, New York, pp. 4–16.

Cregg, J.M., Higgins, D.R. (1995) Production of foreign proteins in the yeast pichia pastoris, *Cand. J. Bot.* **73**, 5891–5897.

Datar, R.V., Cartwright, T., Rosen, C.G. (1993) Process economics of animal cell and bacterial fermentations: a case study analysis of tissue plasminogen activator, *Bio/Technology* **11**, 349–357.

Datar, R.V., Rosen, C.G. (1993) Cell and Cell Debris Removal: Centrifugation and Crossflow Filtration, in: Rehm, H.J., Reed, G. (Eds.) *Biotechnology*, Vol. 3, *Bioprocessing*, VCH Publishers, Weinheim, pp. 486–498.

De Gooijer, C.D., Bakker, W.A.M., Beeftink, H.H., Tramper, J. (1996) Bioreactors in series: an overview of design procedures and practical applications, *Enzyme Microb. Technol.* **18**, 202–219.

Doran, P.M. (1999) Design of mixing systems for plant cell suspensions in stirred reactors, *Biotechnol. Prog.* **15**, 319–355.

Evangelista, R.L., Kusnadi, A.R., Howard, J.A., Nikolov, Z.L. (1998) Process and economic evaluation of the extraction and purification of recombinant β-glucuronidase from transgenic corn, *Biotechnol. Prog.* **14**, 607–614.

Fernandez, J.M., Hoeffler, J.P. (1999) *Gene Expression Systems*, Academic Press, New York.

Flores, H.E. (1992) Plant roots as chemical factories, *Chem. Ind.* **(May) 18**, 374–377.

Freshney, R.I. (1987) *Culture of Animal Cells: A Manual of Basic Technique*, 2nd edition, Alan R. Liss, Inc., New York.

Georgiou, G. (1988) Optimizing the production of recombinant proteins in microorganisms, *AIChE J.* **34**, 1233–1243.

Gundlach, H., Muller, M.J., Kutchan, T.M., Zenk, M.H. (1992) Jamonic acid is a signal transducer in elicitor-induced plant cell cultures, *Proc. Natl. Acad. Sci.* **89**, 2389–2393.

Hamel, J.J., Hunter, B., Sikdar, S.K. (1990) *Downstream Processing and Bioseparations: Recovery and Purification of Biological Products*, American Chemical Society, Washington, DC.

Hammer, M.J., Hammer, M.J. Jr. (2001) *Water and Wastewater Technology*, 4th edition, Prentice Hall, NJ.

Hileman, B. (2003) Clashes over Agbiotech, *C&EN* **81**, 25–33.

Ho, C.S., Wang, D.I.C. (1991) *Animal Cell Bioreactors*, Butterworth-Heinemann Press, Stoneham, MA.

Hsiao, T.Y., Bacani, F.T., Carvalho, E.B., Curtis, W.R. (1999) Development of a low capital investment reactor system: application for plant cell suspension culture, *Biotechnol. Prog.* **15**, 114–122.

Hu, W.-S., Peshwa, M.V. (1993) Mammalian cells for pharmaceutical manufacturing, *ASM News* **59**(2), 65–68.

Imanaka, T., Aiba, S. (1981) A perspective on the application of genetic engineering: stability of recombinant plasmid, *Ann. N. Y. Acad. Sci.* **369**, 1–14.

Jegalian, K. (1999) The gene factory, *Technol. Rev.* **March/April**, 64–68.

Kargi, F. (1987) Plant cell bioreactors, present states and future trends, *Biotechnol. Prog.* **3**, 1–11.

Kargi, F., Moo-Young, M. (1985) Transport Phenomena in Bioprocesses, in: Moo-Young, M. (Ed.) *Comprehensive Biotechnology*, Vol. 2, Pergamon Press, Oxford, UK, pp. 5–55.

Katchalski-Katzir, E. (1993) Immobilized enzymes-learning from past successes and failures, *Trends Biotechnol.* **11**, 471–478.

Kelley, B.D., Hatton, T.A. (1993) Protein Purification by Liquid-Liquid Extraction, in: Stephanopoulus, G. (Ed.) *Biotechnology*, Vol. 3, *Bioprocessing*, VCH Publishers, Weinheim, pp. 593–599.

Ketchum, R.E.B., Gibson, D.M., Croteau, R.B., Shule, M.L. (1999) The kinetics of taxoid accumulation in cell suspension cultures of Taxus following elicitation with methyl jasmonate, *Biotechnol. Bioeng.* **62**, 97–105.

Kuchner, K., Arnold, F.H. (1997) Directed evolution of enzyme catalysts, *Trends Biotechnol.* **15**, 523–530.

Kundu, A., Barnthouse, K.A., Cramer, S.M. (1997) Selective displacement chromatography of proteins, *Biotechnol. Bioeng.* **56**, 119–129.

Kunkel, J.P., Jan, D.C.H., Butler, M., Jamieson, J.C. (2000) Comparisons of the glycosylation of a monoclonal antibody produced under nominally identical cell culture conditions in

two different bioreactors, *Biotechnol. Prog.* **16**, 462–470.

Lilie, H., Schwartz, E., Rudolph, R. (1998) Advances in refolding proteins produced in E. coli, *Curr. Opin. Biotechnol.* **9**, 497–502.

MacLoughlin, P.F., Malone, D.M., Murtagh, J.T., Kieran, P.M. (1998) The effect of turbulent jet flows on plant cell suspension cultures, *Biotechnol. Bioeng.* **58**, 595–604.

Marino, M., Angelo, C., Ippolito, A., Cassani, G., Fassina, G. (1997) Effect of bench-scale culture conditions on murine IgG heterogeneity, *Biotechnol. Bioeng.* **54**, 7–25.

Moo-Young, M. (1988) *Bioreactor Immobilized Enzymes and Cells: Fundamentals and Applications*, Elsevier Science Publishing, Inc., New York.

Mueller, R.F., Characklis, W.G., Jones, W.L., Sears, J.T. (1992) Characterization of initial events in bacterial surface colonization by two Psuedomonas species using image analysis, *Biotechnol. Bioeng.* **25**, 1522–1529.

Naughton, G.K. (1999) Skin: the first tissue engineered products, the advanced tissue sciences story, *Sci. Am.* **280**, 84–85.

Ogden, K.L., Davis, R.H. (1991) Plasmid maintenance and protein overproduction in selective recycle reactors, *Biotechnol. Bioeng.* **37**, 325–333.

Privalov, P.L. (1989) Thermodynamic problems of protein structure, *Annu. Rev. Biophys. Chem.* **18**, 47–75.

Roberts, S.C., Shuler, M.L. (1997) Large scale plant cell culture, *Curr. Opin. Biotechnol.* **8**, 154–159.

Sauer, T., Robinson, C.W., Glick, B.R. (1989) Disruption of native and recombinant Escherichia coli in a high-pressure homogenizer, *Biotechnol. Bioeng.* **33**, 1330–1336.

Scholl, M.A., Harvey, R.W. (1992) Laboratory investigations on the role of sediment surface and groundwater chemistry in transport of bacteria through a contaminated sandy aquifer, *Environ. Sci. Technol.* **26**, 1410–1417.

Sellek, G.A., Chaudhuri, J.B. (1999) Biocatalysis in organic media using enzymes from extremophiles, *Enzyme Microb. Technol.* **25**, 471–482.

Seo, J.-H., Bailey, J.E. (1985) Effects of recombinant plasmid content on growth properties and cloned gene product formation in Escherichia coli, *Biotechnol. Bioeng.* **27**, 1668–1674.

Shiloach, J., Kennedy, R.M. (2000) Expanded Bed Adsorption Process for Protein Capture, in: Ahuja, S. (Ed.) *Handbook of Separations*, Academic Press, San Diego, CA, pp. 431–442.

Shuler, M.L. (1985) On the use of chemically structured models for bioreactors, *Chem. Eng. Commun.* **36**, 161–185.

Shuler, M.L., Wood, H.A., Granados, R.R., Hammer, D.A. (1995) *Baculovirus Expression Systems and Biopesticides*, Wiley-Liss, New York.

Stephanopoulos, G., Nielsen, J., Aristidou, A. (1998) *Metabolic Engineering: Principles and Methodologies*, Academic Press, CA.

Sukla, A.A., Cramer, S.M. (2000) Bioseparations by Displacement Chromatography, in: Ahuja, S. (Ed.) *Handbook of Separations*, Academic Press, San Diego, CA, pp. 379–386.

Van Loosdrecht, M.C.M., Norde, W., Lyklema, J., Sehnder, A.J.B. (1990) Hydrophobic and electrostatic parameters in bacterial adhesion, *Aquat. Sci.* **52**, 103–114.

Velander, W.J., Lubon, H., Drohan, W.N. (1997) Transgenic livestock as drug factories, *Sci. Am.* **276**, 70–74.

Biotransformations of Drugs and Chemicals

Paul R. Ortiz de Montellano
Department of Pharmaceutical Chemistry, University of California, San Francisco, CA, USA

1 General Purpose of Xenobiotic and Drug Metabolism 52

2 Classification and Properties of Enzymes of Xenobiotic Metabolism 53

3 Phase I Enzymes 54
3.1 Cytochrome P450 and Flavin Monooxygenases 54
3.2 Monoamine Oxidase 57
3.3 Alcohol and Aldehyde Dehydrogenases 57
3.4 Esterases and Amidases 59
3.5 Epoxide Hydrolases 59

4 Phase II Enzymes 60
4.1 Glucuronyl Transferases 60
4.2 Sulfotransferases 63
4.3 *N*-Acyl Transferases 64
4.4 Methyl Transferases 65
4.5 Glutathione Transferases 66

5 Xenobiotic Metabolism and Toxicity 68

6 Therapeutic Aspects of Drug Metabolism 69

7 Species Differences and Extrapolation to Humans 69

8 High-throughput Methods in Drug Metabolism Studies 70

Bibliography 70
Books and Reviews 70
Primary Literature 70

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Keywords

Electrophilic
A compound, group, or function that is electron deficient and therefore preferentially reacts with electron-rich moieties.

Endobiotic
A compound that is endogenous to the organism.

LogP
The log of the partition coefficient for a compound between an organic solvent, often octanol, and water.

Nucleophilic
A compound, group, or function that is electron rich and therefore preferentially reacts with electron-deficient moieties.

Prodrug
A structure without a desired pharmacological property that is converted by metabolism to a product that has the desired property.

Xenobiotic
A compound foreign to the organism.

Drug and xenobiotic metabolism converts the lipophilic compounds that can be absorbed into the body into polar, excretable products, thus preventing their accumulation, but also sometimes forms more toxic products.

1 General Purpose of Xenobiotic and Drug Metabolism

Life is based on the organization of biological systems into compartments segregated by means of lipid bilayer membranes. This applies to subcellular organelles such as the nucleus and mitochondria, to the cells themselves, and through the assembly of cells into surfaces, to organs, and larger entities. The membrane barriers prevent the passage of charged or highly polar molecules, with the exception of small molecules that diffuse through pores in the membrane barriers, molecules for which active transport systems exist, or molecules taken up during pinocytosis. Lipophilic molecules can cross the membrane barriers relatively unimpeded unless they are so lipophilic that they are effectively trapped within the membrane. In contrast, mammals and other complex organisms only efficiently excrete polar or ionic compounds because lipophilic compounds are readily reabsorbed from excretory compartments such as the kidney. Thus, the properties that allow a xenobiotic to be taken up by a biological system are the

very same properties that militate against its excretion.

The primary purpose of xenobiotic metabolism is to convert lipophilic compounds to polar materials that can be eliminated. For this reason, the highest concentrations of enzymes of drug metabolism are found at sites that are portals of entry into the organism, such as the lungs, intestine, and liver of mammals. It is instructive in this context to consider the example of compounds, typified by manmade polyhalogenated aromatic hydrocarbons, that are poorly metabolized by the enzymes evolved for the metabolism of natural products. Compounds such as the highly halogenated biphenyls accumulate in adipose tissues and a major fraction remains there for essentially the lifetime of the individual. This accumulation of lipophilic, poorly metabolized compounds in fatty tissues is the basis for the well-known concentration of xenobiotics in natural food chains.

2 Classification and Properties of Enzymes of Xenobiotic Metabolism

The enzymes of drug metabolism predominantly catalyze oxidation/reduction, hydrolysis, and conjugation reactions, although there is some ambiguity in assigning enzymes to the drug-metabolizing system. For organizational purposes, the redox and hydrolytic enzymes are considered to be responsible for early stages of xenobiotic metabolism and are known as Phase I enzymes. The conjugative enzymes are usually responsible for the later transformations in a metabolic sequence and are known as Phase II enzymes. This classification is artificial because Phase II conjugation reactions need not be preceded by Phase I reactions, although it is true that Phase I redox or hydrolytic reactions usually occur before the Phase II conjugations that make compounds highly polar. Phase I xenobiotic metabolism is largely catalyzed by cytochrome P450 and flavoprotein monooxygenases, monoamine oxidases, alcohol and aldehyde dehydrogenases, esterases and amidases, and epoxide hydrolases. The enzymes involved in Phase II metabolism are primarily the glucuronyl transferases, sulfotransferases, *N*-acyl transferases, methylases, and glutathione transferases. Despite their involvement in drug metabolism, some members of both the Phase I and Phase II classes of enzymes are primarily or exclusively involved in the processing of endogenous substrates. Furthermore, many drugs resemble endogenous substances and are subject to metabolism by the more specialized enzymes involved in processing of endobiotics. For example, fatty acid β-oxidation, a pathway for the catabolism of endogenous fatty acids, also degrades xenobiotic alkyl chains that terminate in carboxyl groups.

Some generalizations can be made concerning the properties of enzymes that are primarily devoted to xenobiotic metabolism. In contrast to enzymes that process endobiotics, the enzymes of drug metabolism have broad substrate specificities and are stereoselective rather than stereospecific. There are not many classes of drug-metabolizing enzymes, but each class encompasses multiple enzyme forms. This combination of loose substrate specificity and enzyme multiplicity provides the flexibility required to metabolize the diversity of natural and manmade substances to which organisms are routinely exposed. The ability to deal with specific xenobiotics is enhanced by the

fact that the concentrations of some of the enzymes, most notably the cytochrome P450 monooxygenases (see Sect. 3.1), are transiently increased by exposure to specific substrates or inhibitors. This selective induction of enzyme forms that are required to deal with specific xenobiotics enhances the flexibility of the system in its response to xenobiotic exposure.

3 Phase I Enzymes

3.1 Cytochrome P450 and Flavin Monooxygenases

The mammalian cytochrome P450 monooxygenases are membrane-bound hemoproteins that are found in highest concentrations in the liver, kidney, nasal tissues, and lungs but are present in smaller amounts in most other organs. The name cytochrome P450 derives from the distinctive absorption maximum at approximately 450 nm of the ferrous–carbon monoxide complex of these enzymes. The genes for more than 1000 cytochrome P450 enzymes have been identified in prokaryotic and eukaryotic species. Approximately 60 such enzymes are found in the human genome. A nomenclature system has been developed that is based on division of the superfamily of cytochrome P450 enzymes according to their degree of sequence identity. With few exceptions, the members of a cytochrome P450 family share >40% amino acid identity and the members of a subfamily >55% identity. In this nomenclature, cytochrome P450 enzymes are designated as, for example, CYP4A1, in which the number after the prefix CYP identifies the family, the subsequent letter the subfamily, and the final number the individual enzyme within the subfamily. A similar nomenclature has been developed for other families of drug-metabolizing enzymes, including the glucuronyl transferases and sulfotransferases. Citations prior to 1987 employ nonsystematic nomenclatures.

The tissue levels of the different cytochrome P450 enzymes in an individual are controlled by genetic and environmental factors. Genetic variability in the expression of the various cytochrome P450 enzymes is well established and gives rise to populations with distinct metabolic patterns. Important differences exist in the ability of individuals to metabolize drugs owing to the existence of populations with high and low activities of, for example, the CYP2D6 form of cytochrome P450 responsible for the metabolism of debrisoquine and a variety of other drugs. In addition to genetic factors, the levels of cytochrome P450 enzymes are modulated by diet, age, disease, gender, and hormonal status. The ability of drugs and xenobiotics to alter the profile of cytochrome P450 enzymes in an individual by elevating the concentrations of specific forms is particularly relevant in this context. The increase in the levels of certain enzymes by a xenobiotic, termed *enzyme induction*, can be mediated by several mechanisms, among which the most important are an increase in transcription of the message for the enzyme, stabilization of the RNA transcript, and stabilization of the enzyme itself.

The mechanism of induction of the cytochrome P450 enzyme CYP1A1 involves binding of the inducer to a soluble receptor, termed the *aryl hydrocarbon receptor* (or AHR), that translocates to the nucleus and enhances the expression of the appropriate gene. Comparable receptor-mediated mechanisms have more recently been identified for the induction of other classes of cytochrome P450 enzymes. The classical

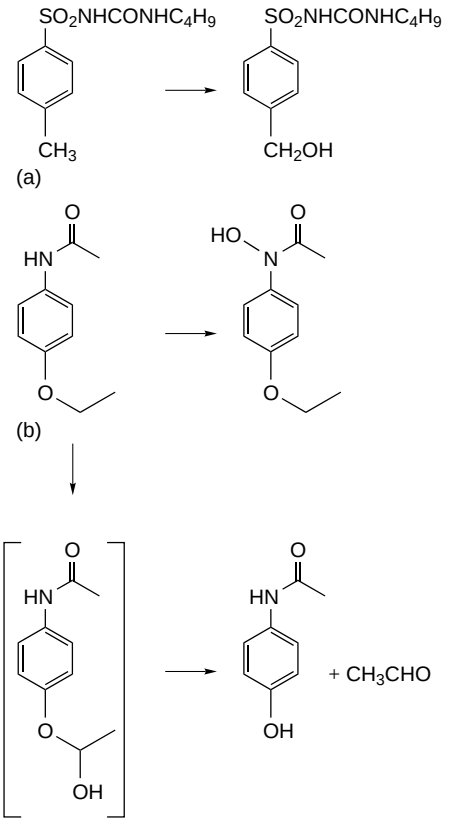

Fig. 1 Examples of cytochrome P450–catalyzed hydroxylation reactions. Hydroxylation of the methyl group in tolbutamide (a) and the amide nitrogen in phenacetin (b) yields, respectively, the corresponding alcohol and *N*-hydroxyamide metabolites. Alternative hydroxylation of the ethyl ether function in phenacetin produces a hemiacetal that decomposes to the phenol and acetaldehyde, a classic example of a cytochrome P450–catalyzed *O*-dealkylation.

cytochrome P450–inducing agents are phenobarbital, the polycyclic aromatic hydrocarbons, clofibrate, and agents such as dexamethasone, although compounds as simple as ethanol are known to elevate the concentrations of specific cytochrome P450 enzymes. In a general sense, the cytochrome P450 enzymes that are induced are those that bind, and usually metabolize, the inducing drug or xenobiotic. The induction response thus permits the organism to increase its ability to clear offending xenobiotics.

Cytochrome P450 monooxygenases catalyze the insertion of one atom of molecular oxygen into their substrates with concomitant reduction of the other oxygen atom to a molecule of water. The electrons required for this reductive activation of molecular oxygen are provided by reduced pyridine nucleotides. A flavoprotein known as cytochrome P450 reductase mediates the transfer of electrons from NADPH to the cytochrome P450 enzymes of interest in mammalian drug metabolism. The mitochondrial cytochrome P450 enzymes are primarily involved in the metabolism of endogenous sterols and, like most bacterial cytochrome P450 enzymes, obtain their electrons from NADH via a two-protein system consisting of a flavoprotein reductase and an iron sulfur protein. In contrast to the diversity of cytochrome P450 enzymes, there is relatively little diversity in the reductase components.

One of the three general oxidative reactions supported by cytochrome P450 is insertion of the catalytically activated oxygen atom into a C–H or N–H bond to give the corresponding hydroxylated product (Fig. 1). In principle, the oxygen can be inserted into the bond between a hydrogen and any other heavy atom. In the absence of enzyme-imposed steric or substrate orientation factors, the reactivity of a C–H bond is inversely correlated with the strength of the bond, so that tertiary alcohols are more easily formed than primary alcohols. One consequence of this is that

hydrocarbon chains are most commonly hydroxylated at the next-to-the-last ($\omega - 1$) position, although the CYP4 family of cytochrome P450 enzymes has specifically evolved to promote the thermodynamically more difficult terminal (ω) hydroxylation of fatty acids. Oxidation by cytochrome P450 is particularly critical for hydrocarbons and other nonpolar structures that do not have a functional group that can be used for further metabolism. This is clearly illustrated, as already mentioned, by the fact that polyhalogenated hydrocarbons (e.g. hexachlorinated biphenyls) that are resistant to cytochrome P450–catalyzed oxidation are very poorly excreted and accumulate in adipose tissue.

The diversity of the products formed by the cytochrome P450 system is greatly enhanced by the decomposition or rearrangement of unstable oxidation products to stable metabolites. The best example of this is hydroxylation adjacent to a heteroatom, which is normally followed by elimination of the heteroatom and conversion of the hydroxyl group to a carbonyl function. Alkyl ethers are thus *O*-dealkylated (Fig. 1), alkyl amines *N*-dealkylated, alkyl thioethers *S*-dealkylated, and alkyl halides dehalogenated.

The second general cytochrome P450 reaction is transfer of the activated oxygen to the π-bond of an olefin, an aromatic ring, or another unsaturated substructure to give the epoxide or a product formally obtained by rearrangement of such an epoxide. In general, electron-rich π-bonds are more readily oxidized than electron-deficient π-bonds. The epoxides formed in this reaction are relatively reactive and potentially toxic, but their toxicity is attenuated by the presence of epoxide hydrolases (see Sect. 3.6) and glutathione transferases (see Sect. 4.5) that convert the epoxides to less-reactive vicinal diols and glutathione adducts, respectively. The oxidation of aromatic rings yields particularly unstable epoxides that are subject to the so-called *NIH rearrangement* to give phenolic products (Fig. 2). Indeed, it is unclear in some instances whether the epoxide is an actual intermediate or a species closer to a transition state in the conversion of the aromatic ring to a phenol.

The third general cytochrome P450 reaction is addition of the activated oxygen to the electron pair of a heteroatom. A trisubstituted nitrogen is thus converted to an N-oxide and a disubstituted sulfur to a sulfoxide (Fig. 3). Whereas hydroxylation and epoxidation reactions are primarily catalyzed by cytochrome P450, the oxidation of heteroatoms is also catalyzed by flavin-containing monooxygenases.

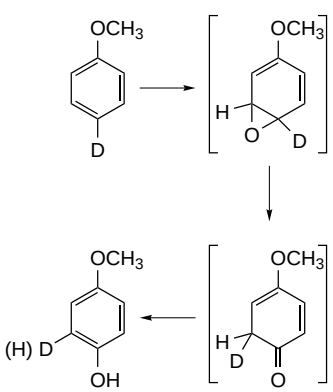

Fig. 2 The cytochrome P450–catalyzed epoxidation of aromatic rings is commonly followed by the "NIH Shift", a nonenzymically catalyzed rearrangement that converts the epoxide to the corresponding phenol.

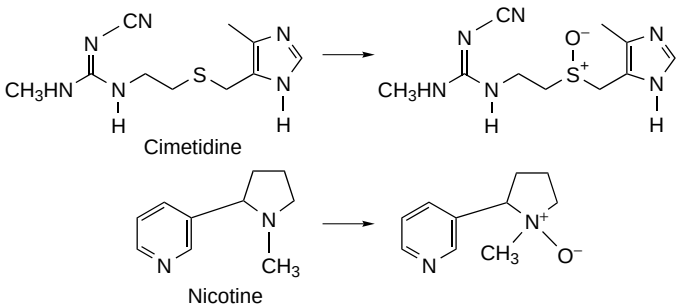

Fig. 3 Cytochrome P450 and the flavin-containing monooxygenase (FMO) catalyze the oxidation of nitrogen and sulfur, as illustrated by the sulfoxidation of cimetidine and the N-oxidation of nicotine.

The flavin-containing monooxygenases, or FMOs, are generally limited to the NADPH- and oxygen-dependent oxidation of nitrogen and sulfur atoms and therefore have a more restricted metabolic scope than cytochrome P450 enzymes. Nevertheless they can play an important role in the metabolism of individual drugs. Both the cytochrome P450 and flavin-containing monooxygenases preferentially oxidize electron-rich heteroatoms.

Cytochrome P450 enzymes can catalyze oxidative reactions other than the three general reactions described above, including reactions such as the dehydrogenation of alkyl groups, deformylation of aldehydes, and cleavage of carbon–carbon bonds. Although relatively rare, these reactions have been more frequently observed in the past few years. Under conditions of low oxygen tension, cytochrome P450 enzymes also catalyze reductive reactions, including the reduction of nitro and azo groups to amines and the dehalogenation of haloalkanes.

3.2 Monoamine Oxidase

Monoamine oxidase is an oxygen-dependent mitochondrial flavoprotein that oxidizes endogenous neurotransmitter amines and related xenobiotics to the corresponding imines. The imines are subsequently hydrolyzed in the aqueous medium to the carbonyl derivatives and either ammonia or an amine, as shown below for 2-phenylethylamine.

$$\mathrm{PhCH_2CH_2NH_2} \longrightarrow \mathrm{PhCH_2CH{=}NH}$$
$$\longrightarrow \mathrm{PhCH_2CHO} + \mathrm{NH_3}$$

Primary amines are particularly good substrates for the enzyme, but secondary amines in which one of the substituents is a methyl or another small alkyl group are readily oxidized. The oxidation of trisubstituted amines often results in inactivation of the enzyme, a phenomenon exploited in the development of clinically useful inhibitors of the monoamine oxidases.

3.3 Alcohol and Aldehyde Dehydrogenases

Alcohol dehydrogenases are members of a family of cytosolic enzymes that catalyze the NAD-dependent oxidation of alcohols to aldehydes or ketones as well as the reverse reaction, the NADH-dependent reduction of carbonyl compounds to alcohols (Fig. 4). At least 16 different alcohol dehydrogenase isoforms have been identified in human liver. Less well characterized

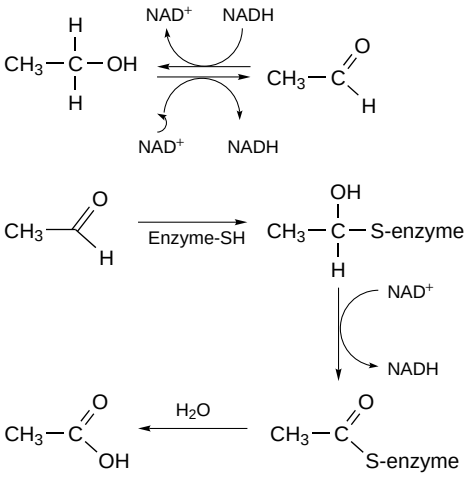

Fig. 4 Alcohol dehydrogenases catalyze the reversible oxidation of alcohols to aldehydes and ketones, as illustrated by the oxidation of ethanol to acetaldehyde. The oxidation of aldehydes to carboxylic acids, such as that of acetaldehyde to acetic acid, is catalyzed by aldehyde dehydrogenases. The aldehyde reacts with a thiol in the aldehyde dehydrogenase to give an enzyme-bound hemithioacetal intermediate that is then oxidized to an enzyme-bound thioester and finally released by hydrolysis of the thioester bond.

aldehyde and ketone reductases that employ NADPH as the reducing cofactor are also involved in the reduction of carbonyl compounds to alcohols. The direction of the reaction catalyzed by alcohol dehydrogenases is dictated by the redox balance of the pyridine nucleotide pool and by whether an alcohol or carbonyl compound is the substrate, although a thermodynamic equilibrium mixture of the alcohol and carbonyl compound would be obtained were it not for the fact that the substrate and product are removed from the solution by alternative metabolic reactions and the dynamic nature of the circulatory and excretory systems.

Aldehydes are primarily oxidized by NAD^+ -dependent aldehyde dehydrogenases (Fig. 4). A molybdenum-containing enzyme related to xanthine oxidase may also contribute to aldehyde oxidation, but its physiological importance in the general metabolism of xenobiotic aldehydes is unclear. Unlike the alcohol dehydrogenases, the aldehyde dehydrogenases do not simply transfer a hydride from the aldehyde substrate to the pyridine nucleotide cofactor. The first step of the reaction is addition of the aldehyde to an active site cysteine thiol to give a hemithioacetal that is then dehydrogenated by a hydride transfer to an enzyme-bound thioester. Hydrolysis of the thioester releases the acid metabolite and regenerates the enzyme (Fig. 4). Because of the complexity of this mechanism, aldehyde dehydrogenases catalyze the oxidation of aldehydes to acids but not the reverse reaction. In effect, carboxylic acids and xenobiotics with functionalities such as amides and esters in which the carbon has a comparable oxidation state are rarely metabolically reduced.

Alcohol and aldehyde dehydrogenases are the primary enzymes involved in the metabolism of ethanol, the former catalyzing the oxidation of ethanol to acetaldehyde and the latter the conversion of acetaldehyde to acetic acid (Fig. 4). The oxidation of ethanol to acetaldehyde can also be mediated by catalase and cytochrome P450, but under normal conditions alcohol dehydrogenases are the principal catalysts for this transformation. Genetic polymorphisms in the levels of both the alcohol and aldehyde dehydrogenases are responsible for the lower tolerance of certain populations to the pharmacological effects of ethanol. Inhibition of acetaldehyde

dehydrogenase by drugs such as disulfiram [$Et_2NC(S)SSC(S)NEt_2$] or by metabolites of other drugs that also give stable adducts with the catalytic thiol group causes hypersensitivity to the noxious effects of ethanol. This property of disulfiram has been exploited in the treatment of chronic alcoholism.

3.4 Esterases and Amidases

The hydrolysis of esters and amides, a very general metabolic pathway, is catalyzed by a diverse group of carboxylesterases. These enzymes differ in their mechanisms and substrate specificities but commonly catalyze ester and amide, as well as thioester, hydrolysis. Many of the carboxylesterases are serine hydrolases in which an active site serine catalyzes the hydrolytic reaction, but enzymes of other types are also involved. Large concentrations of these enzymes are found in the liver and kidney, but they can be found in most tissues. As a rule, ester hydrolysis occurs more rapidly than amide hydrolysis, in accord with the relative susceptibilities of the two functionalities toward chemical hydrolysis.

The carboxylesterases hydrolyze esters to the carboxylic acid and alcohol, and amides to the corresponding carboxylic acid and amine.

$$RCO_2R' \longrightarrow RCO_2H + R'OH$$

$$RCONR'R'' \longrightarrow RCO_2H + R'NHR''$$

Ester and amide hydrolysis greatly increases the polarity of the substrate, as it replaces the neutral ester or amide function with a highly ionized (at pH 7) carboxylic acid group. The amine group released by amide hydrolysis is also likely to be protonated at physiological pH, and the alcohol, albeit not ionized, is likely to be more polar due to the presence of the hydrogen-bonding hydroxyl group. Furthermore, the functionalities unmasked by ester or amide hydrolysis are susceptible to glucuronidation, sulfation, and/or amino acid conjugation (See Sect. 4).

Esterification is often used to convert drugs to prodrugs from which the parent drugs are released metabolically at rates, or in tissue locations, that improve the pharmacokinetic properties and therapeutic activities of the drugs. For example, fluphenazine is a short-lived antipsychotic agent that must be administered several times a day to be effective. Esterification with a short-chain fatty acid, however, yields a derivative that can be administered in a single, relatively large dose that is deposited in adipose tissue. Hydrolytic release of therapeutic doses of fluphenazine from the ester occurs over a period of two to four weeks (Fig. 5).

3.5 Epoxide Hydrolases

The hydrolysis of epoxides is catalyzed by epoxide hydrolases located in both the

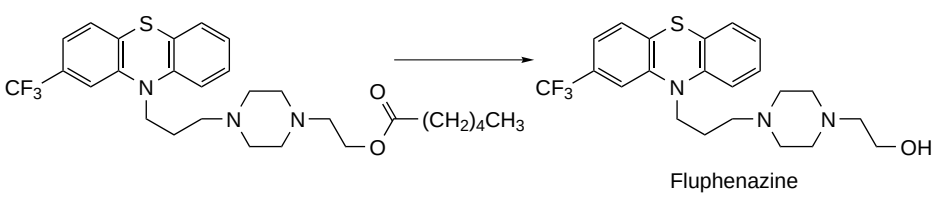

Fig. 5 Fatty acid ester derivatives of fluphenazine are lipophilic prodrugs from which fluphenazine, the active principle, is released by enzymatic hydrolysis.

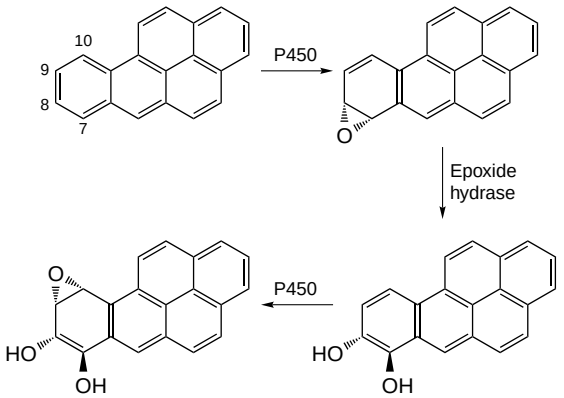

Fig. 6 The epoxidation of benzo[a]pyrene by cytochrome P450 is followed by enzymatic hydrolysis of the epoxide and a second cytochrome P450–catalyzed epoxidation reaction. The resulting highly reactive diol epoxide alkylates DNA and this reaction is at least partially responsible for the carcinogenic properties of this polycyclic aromatic hydrocarbon.

endoplasmic reticulum and the cytosol. The membrane-bound and soluble enzymes have somewhat different substrate specificities, but both formally catalyze backside addition of water to the epoxide to give a trans-diol product. The actual mechanism of the reaction, however, is addition of a carboxylic acid group of the enzyme to the epoxide followed by internal hydrolysis of the resulting protein-bound ester. In general, increasing the number of substituents on the epoxide decreases the rate of enzymatic hydrolysis by the membrane-bound enzyme but has less effect on hydrolysis of the epoxide by the soluble enzyme. Although epoxide groups are usually not found in drugs or xenobiotics because of their chemical reactivity, the ability to detoxify epoxides is very important because epoxides are formed *in situ* by the cytochrome P450–catalyzed oxidation of olefins and aromatic rings (Sect. 3.1). Epoxides are also detoxified by the glutathione transferase–mediated addition of glutathione (Sect. 4.5). The redundancy represented by the abilities of both epoxide hydrolases and glutathione transferases to detoxify epoxides reflects the importance of doing so to the well-being of the host.

Epoxide hydrolysis is normally a detoxifying reaction because it converts chemically reactive epoxides into nonreactive diols that can be glucuronidated or sulfated. Nevertheless, the potential deleterious interplay between cytochrome P450 enzymes and epoxide hydrolases is well illustrated by the bioactivation of carcinogenic polycyclic aromatic hydrocarbons. In the case of benzo[a]pyrene, one of the reaction sequences thought to play an important role in carcinogenesis involves oxidation of the 7,8-double bond of the hydrocarbon by cytochrome P450, hydrolysis of the epoxide to the *trans*-7,8-diol by epoxide hydrolase, and epoxidation of the 9,10-double bond by cytochrome P450 to give the 9,10-epoxy-7,8-diol (Fig. 6). This highly reactive diol epoxide alkylates DNA, causing a lesion that engenders a carcinogenic response.

4 Phase II Enzymes

4.1 Glucuronyl Transferases

Glucuronidation is among the most common and quantitatively most

important pathways of drug and xenobiotic metabolism. The reaction is catalyzed by a family of membrane-bound enzymes known as UDP-glucuronosyltransferases (or glucuronyl transferases). The number of enzymes in this family is not known, but evidence exists for at least a dozen rat enzymes and a corresponding number of human enzymes. UDP-glucuronosyltransferases transfer the glucuronic acid moiety from the endogenous donor uridine-5′-diphospho-α-D-glucuronic acid (UDPGA) to an acceptor function on the drug or xenobiotic (Fig. 7). Glucuronidation greatly increases the polarity and water solubility of most xenobiotics due to the high ratio of oxygen to carbon atoms in glucuronic acid and the fact that its carboxyl group ($pK_a = 3.5$) is highly ionized at physiological pH. This increase in ionic character and polarity greatly accelerates excretion of the xenobiotic. Glucuronidation usually occurs only once and it is the last step in a metabolic sequence because the glucuronide metabolite is usually too polar to interact efficiently with the membrane-bound enzymes of drug metabolism and is readily excreted. A few lipophilic compounds are known, however, that are glucuronidated twice. Bilirubin produced endogenously by the catabolism of heme is one of these compounds.

Glucuronidation involves displacement of the UDP phosphate group by a nucleophile in the xenobiotic. The stereochemistry of the glucuronic acid carbon involved in the reaction is inverted, in accord with a backside displacement mechanism (Fig. 7). The glucuronic acid moiety can be transferred by UDP-glucuronosyltransferases to most nucleophilic groups if they can be deprotonated at physiological pH to accommodate

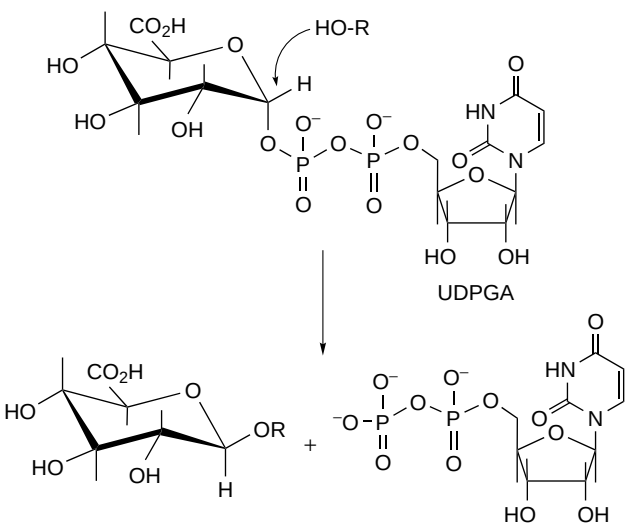

Fig. 7 Glucuronidation by glucuronosyl transferases involves transfer of the glucuronic acid moiety from uridine-5′-diphospho-α-D-glucuronic acid (UDPGA) to the substrate, illustrated here by an alcohol (ROH).

the glucuronic acid moiety. Hydroxyl, carboxyl, amino, and sulfhydryl functions such as those found in phenols, alcohols, enols, carboxylic acids, hydroxylamines, primary amines, and mercaptans are thus commonly glucuronidated. Acetaminophen, clofibrate, and 1-naphthylamine are examples of drugs and xenobiotics that are extensively glucuronidated (Fig. 8).

Trisubstituted amines, nitrogen heterocycles, and a few other nitrogen-containing functionalities are unique in that they can be glucuronidated in the absence of deprotonation to give quaternary, charged products. An example of such a transformation is the glucuronidation of ketotifen (Fig. 8). Carbon glucuronidation is possible with compounds such as phenylbutazone that have highly acidic carbon atoms.

Glucuronyl transferase activity depends on a variety of factors, including age, diet, genetic factors, hormonal status, and cigarette smoking. The increase in glucuronidation observed during the neonatal period is very important because bilirubin, the catabolite of heme, is not readily excreted unless it is first glucuronidated. The low levels of glucuronosyltransferase activity at birth can result in bilirubin accumulation, jaundice, and nervous system damage unless measures (usually phototherapy) are taken to eliminate bilirubin by other mechanisms until the appropriate glucuronosyltransferase increases to the appropriate functional level.

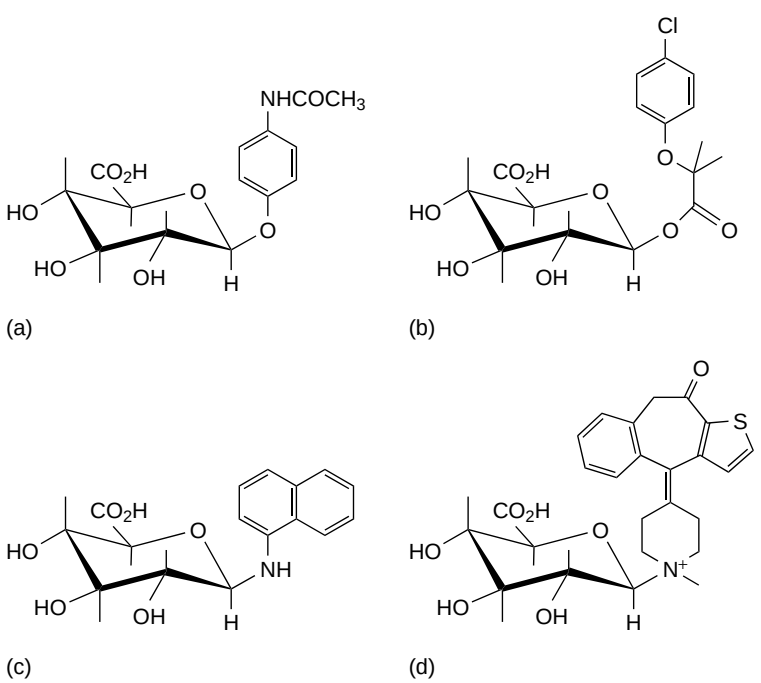

Fig. 8 Examples of glucuronides in which the functionality that is glucuronidated is (a) the phenol group of acetaminophen, (b) the carboxylic acid of clofibrate, (c) the amino group of 1-naphthylamine, and (d) the trisubstituted amine group of ketotifen.

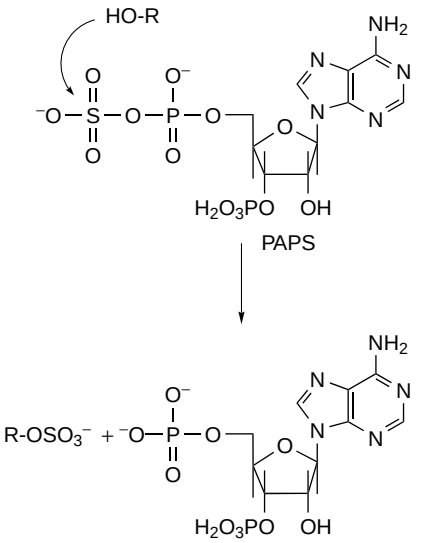

Fig. 9 Sulfation catalyzed by sulfotransferases involves transfer of the sulfate group of 3-phosphoadenosine-5′-phosphosulfate (PAPS) to the substrate, represented in the figure by an alcohol (ROH).

4.2 Sulfotransferases

Sulfotransferases are soluble enzymes that transfer the sulfate group from 3-phosphoadenosine-5′-phosphosulfate to a nucleophile on the drug or xenobiotic (Fig. 9). The pK_a value (~ -7) of the sulfate function is such that the resulting sulfate conjugates are completely ionized, and therefore more readily eliminated, under all physiological conditions. Sulfation of drugs and xenobiotics is almost as common as glucuronidation. The physiological sulfate pool appears to be limited, however, so that the proportional importance of sulfation often decreases as the dose of the drug or xenobiotic increases.

The range of functionalities that undergo sulfation is similar to, but more limited than, the range that undergoes glucuronidation. Hydroxyl groups, particularly phenolic groups such as that in terbutalin (Fig. 10), and amino functions are the most commonly sulfated. The sulfation of carboxylic acids, trialkylamines, and sulfhydryl groups, unlike glucuronidation, is not observed. In the case of sulfhydryl groups, limited evidence suggests that

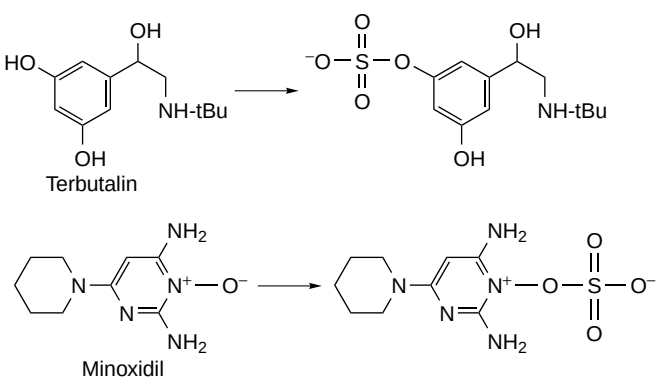

Fig. 10 Sulfation of the phenol function of terbutalin increases its polarity and excretability. Sulfation of the N-oxide group of minoxidil is unusual and produces a metabolite that is responsible for the vasodilatory properties of the drug (actually a prodrug).

sulfate derivatives may be formed but react rapidly with glutathione or other protein sulfhydryls to give disulfide products. Sulfation accelerates xenobiotic excretion but, in a few instances, converts prodrugs to their active form. The best example of this is conversion of the vasodilatory prodrug minoxidil to its physiologically active sulfate metabolite (Fig. 10).

4.3 *N*-Acyl Transferases

N-Acyl transferases are involved in three pathways of xenobiotic metabolism: (1) conjugation of the carboxyl group of xenobiotics to the α-amino group of various amino acids to give amino acid conjugates, (2) acetylation of the amino functions of xenobiotics, and (3) acetylation of the cysteine conjugates produced by the glutathione transferase pathway (Sect. 4.5). In each of these acetylation reactions, the carboxyl group is first activated by enzymatic conversion to the corresponding CoA ester. This reaction is catalyzed by ATP-dependent acid:CoA ligases that are also involved in the formation of medium-length fatty acid CoA esters. For the formation of amino acid conjugates, the carboxyl group of the xenobiotic is activated, whereas acetyl CoA provides the activated carboxyl function for *N*-acetylation of the amino functions in xenobiotics or the α-amino groups of cysteine conjugates (Sect. 4.5). Transfer of the activated acyl moiety from the acyl CoA to the amino group of the acceptor is catalyzed by *N*-acyl transferases.

The most common amino acids involved in the formation of amino acid conjugates in mammals are glycine, glutamine, and taurine, although conjugates with arginine, asparagine, histidine, lysine, and serine are also formed in a species-dependent manner. The *N*-acyl transferases and the enzymes that synthesize the acyl CoA esters belong to enzyme families located primarily in the matrix of liver and kidney mitochondria. The range of xenobiotics susceptible to amino acid conjugation is restricted to compounds with a carboxylic acid function. The reaction is catalyzed by acyl-CoA:amino acid *N*-acyltransferases. The increase in polarity achieved by this transformation is modest because the reaction replaces one carboxylic acid function with another and therefore does not greatly alter the ionization state of the xenobiotic. The conversion of benzoic acid to hippuric acid by conjugation with glycine is a case in point.

$$\mathrm{PhCO_2H} \longrightarrow \mathrm{PhCOSCoA}$$
$$\longrightarrow \mathrm{PhCONHCH_2CO_2H}$$

The acetyl CoA–dependent acetylation of xenobiotic amino groups is catalyzed by cytosolic *N*-acetyl transferases. In man, one form of the enzyme predominates in the liver and gut and another in extrahepatic tissues. Arylamines and hydrazines are the most commonly *N*-acetylated functions, although the reaction also occurs with sulfonamides, aliphatic amines, and related functions. Typical examples of this process are the *N*-acetylation of isoniazid and sulfamethazine (Fig. 11). Because acetylation decreases rather than increases the polarity of the xenobiotic, it can occur at any stage in the metabolic process.

The ability to *N*-acetylate xenobiotics is polymorphically distributed in the human population and has therapeutic and toxicological consequences. For a given dose of a drug, slow acetylators accumulate higher concentrations of the parent drug than fast acetylators. This can have clinical and pathological consequences if the therapeutic index of the drug is fairly

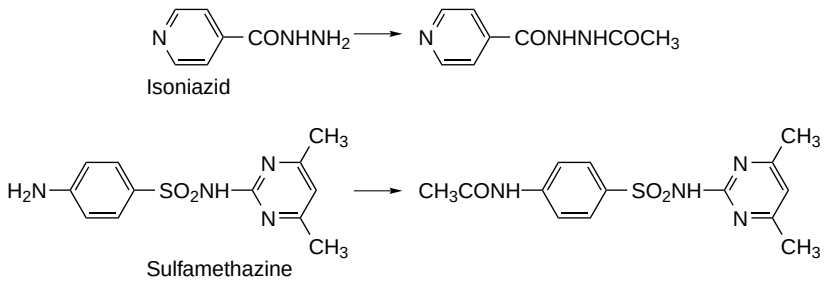

Fig. 11 *N*-Acetylation of a hydrazine (isoniazid) and an aryl amine (sulfamethazine), two typical reactions catalyzed by *N*-acetyl transferases.

narrow. For example, slow acetylators are more susceptible to the peripheral neuropathy associated with isoniazid or the lupus erythematosus side effect of hydralazine. The variability of the *N*-acetylation phenotype in humans is due, at least in part, to point mutations in the *N*-acetyltransferase genes.

4.4 Methyl Transferases

N-, *O*-, and *S*-methylations of xenobiotics are catalyzed by *S*-adenosyl-*L*-methionine-dependent methyltransferases, including enzymes such as catechol *O*-methyltransferase and indolethylamine *N*-methyltransferase, whose primary substrates are endogenous neurotransmitters and their metabolites (Fig. 12). The substrate specificities of these enzymes are broad enough that they often also methylate drugs that have the required functionality and structurally resemble the endogenous substrates. Methylation of the meta-hydroxyl group in a catechol substrate such as dopamine (Fig. 12) is usually favored over methylation of the para-hydroxyl, although both products are formed. One consequence of the fact that the primary function of the *N*- and *O*-methyltransferases is the metabolism of endogenous substrates is that methylation of hydroxyl and amine groups primarily occurs with xenobiotics that structurally resemble endogenous neurotransmitters. A second consequence is that the polarity of the substrate is decreased rather than

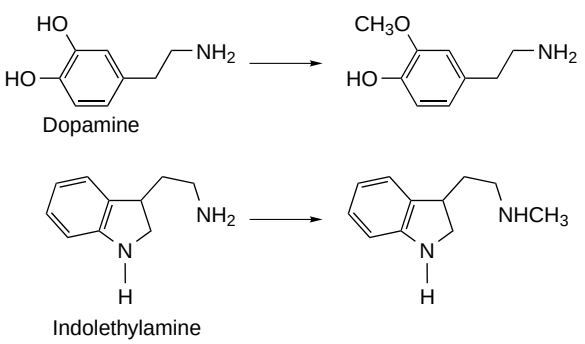

Fig. 12 Methylation of the catechol group of dopamine and the primary amine group of indolethylamine.

increased by methylation. Methylation and *N*-acetylation are therefore the two general pathways that decrease substrate polarity.

4.5 Glutathione Transferases

Glutathione is a tripeptide in which the carboxyl and amino groups of cysteine are linked respectively to a glycine and the γ-carboxyl group of glutamic acid (Fig. 13). Incorporation of cysteine into this tripeptide protects it from utilization in protein synthesis and makes possible the accumulation of high (5–10 mM) concentrations of the cysteine sulfhydryl in tissues. These high concentrations are desirable because glutathione serves as the major line of defense against electrophilic, chemically reactive species, including xenobiotics, xenobiotic metabolites, and reactive oxygen derivatives. Highly reactive electrophilic metabolites can be trapped by direct, uncatalyzed reaction with the glutathione sulfhydryl group, but the addition of glutathione to moderately or weakly electrophilic compounds is catalyzed by a family of enzymes known as glutathione transferases. The glutathione transferases have broad, overlapping substrate specificities and are primarily, but not exclusively, located in the cytosol.

Glutathione reacts chemically or with the catalytic assistance of a glutathione transferase with essentially all electrophilic functionalities, including alkyl and acyl halides (Fig. 13), epoxides (Fig. 14), electrophilically activated aromatic ring systems (Fig. 14), and α, β-unsaturated carbonyl compounds (Fig. 15). The products of the reactions are the thioethers or thioesters in which the cysteine sulfhydryl has added to the electrophilic center, usually with concomitant displacement of a leaving group from the same center. The glutathione transferases promote these reactions, in part, by simply binding the lipophilic, electrophilic substrates in close proximity to the enzyme-bound glutathione.

Glutathione conjugates are usually more polar and/or highly ionized than the parent xenobiotic due to the hydrophilic nature of the tripeptide and the presence of ionizable amino and carboxyl functions in it. Glutathione conjugates are actively transported out of cells, but the conjugates undergo further transformation before they are excreted. Glutamic acid

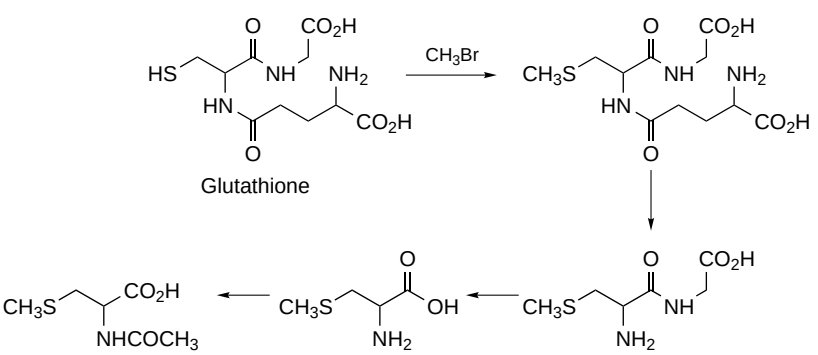

Fig. 13 The sequence of steps involved in conjugation of the tripeptide glutathione with an electrophilic substrate, here methyl bromide, and subsequent processing of the glutathione conjugate to the *N*-acetylcysteine conjugate that is actually excreted.

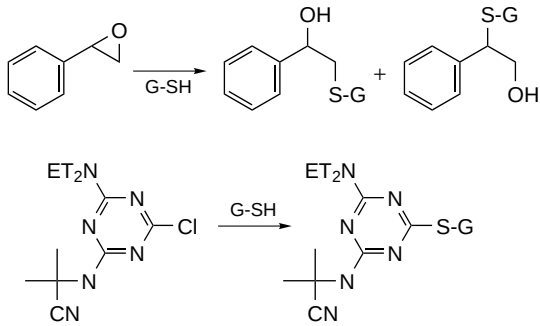

Fig. 14 Addition of glutathione to the epoxide of styrene oxide and the activated aromatic ring of a herbicide. GSH stands for glutathione.

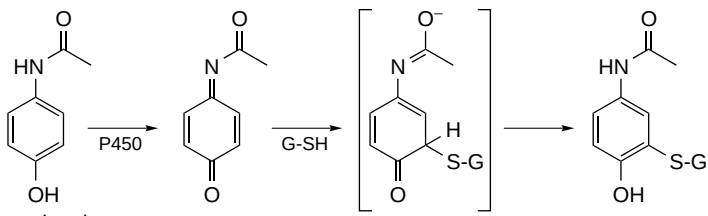

Acetaminophen

Fig. 15 Acetaminophen is oxidized by cytochrome P450 to a chemically reactive iminoquinone structure that is trapped by glutathione to give a glutathione conjugate. In the absence of glutathione, the iminoquinone alkylates multiple liver proteins.

and glycine are first removed from the conjugate by the ordered, sequential action of a γ-glutamyl transpeptidase and a peptidase (Fig. 13). High concentrations of these enzymes are found in the kidney. The cysteine conjugate obtained by removal of the two protective amino acids is then *N*-acetylated by the action of an acetyl CoA–dependent *N*-acetyltransferase (See Sect. 4.3). *N*-Acetylation converts the zwitterionic amino acid derivative into a simple carboxylic acid and thus prevents reabsorption of the conjugate from the kidney by amino acid transporters. The *N*-acetyl cysteine conjugates, known as mercapturic acids, are generally excreted in the urine.

The cysteine conjugates prior to *N*-acetylation, or regeneration by hydrolysis of the *N*-acetyl group of the mercapturic acid derivatives, are subject to an alternative metabolic pathway initiated by a β-lyase. This pyridoxal phosphate-dependent enzyme catalyzes removal of the proton adjacent to the amino group with concomitant elimination of the thioether or thioester group. The products of the reaction are the sulfhydryl derivatives of the original xenobiotic and pyruvic acid (CH_3COCO_2H) (Fig. 16). The free sulfhydryl group is then methylated by an *S*-adenosyl-*L*-methione-dependent *S*-methyl transferase (See Sect. 4.4). The thiomethyl product is relatively nonpolar and can be oxidized by a cytochrome P450 or flavin-containing monooxygenase to the corresponding sulfoxide.

The high concentrations of glutathione present in most tissues provide a reserve

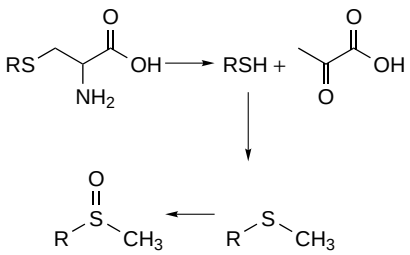

Fig. 16 The cysteine β-lyase pathway for the metabolism of cysteine conjugates derived by normal proteolytic processing of glutathione conjugates. The thiol derivative released from the cysteine conjugate by the action of cysteine β-lyase can be methylated by an *S*-methyltransferase and subsequently oxidized to the sulfoxide by a monooxygenase.

for the detoxification of reactive, electrophilic compounds. This reserve is nevertheless limited and exposure to large amounts of electrophilic species can result in its temporary depletion. Depletion of the glutathione reservoir in the face of continued exposure to electrophilic products can result in life-threatening cell and organ damage. The relationship between glutathione depletion and cell damage has been most extensively studied with acetaminophen, a small fraction of which is oxidized by cytochrome P450 to a highly reactive iminoquinone derivative that is normally detoxified by glutathione (Fig. 15). The protection provided by glutathione is such that normal doses of this over-the-counter analgesic agent are nontoxic. Ingestion of large doses of acetaminophen, however, results in the formation of sufficiently large amounts of the reactive metabolite that the glutathione reservoir is depleted. The iminoquinone metabolite is then free to react with cellular constituents, causing hepatic necrosis and, if the damage is sufficiently extensive, death. The same toxic consequences are observed but at a lower dose of acetaminophen if the glutathione concentration is first depleted by administration of a second agent or if the proportion of the reactive metabolite is increased by induction of the appropriate cytochrome P450 enzyme.

5 Xenobiotic Metabolism and Toxicity

The low reaction control implicit in the broad specificity of drug-metabolizing enzymes and the unpredictable diversity of xenobiotic structures make drug metabolism a major contributor to the toxicity and carcinogenicity of xenobiotics. For example, the cytochrome P450–catalyzed oxidation of aflatoxin and polycyclic aromatic hydrocarbons to epoxides (Fig. 6) that bind covalently to DNA is directly responsible for the carcinogenic properties of these substances. The analogous covalent binding of reactive metabolites to proteins is responsible for the toxic properties of many xenobiotics. Thus, the oxidation of allyl alcohol by alcohol dehydrogenases produces acrylaldehyde ($CH_2{=}CHCH_2OH \rightarrow CH_2{=}CH{-}CH{=}O$), a reactive electrophilic agent that alkylates proteins and causes tissue damage. As noted in Sect. 4.4, a major role of the glutathione system is to prevent tissue damage by trapping metabolically produced reactive species. The relationship between the glutathione system and toxicity has been most clearly defined in the case of acetaminophen. Thus, the metabolism of lipophilic compounds is essential for their elimination but not infrequently produces species of higher toxicity than the parent structure.

6
Therapeutic Aspects of Drug Metabolism

The magnitude and duration of drug action is controlled by the extent and rate of absorption of the drug, the extent and rate at which it is delivered to the target site, the extent and rate of its metabolism, and the rate at which it is cleared from the body. Factors that influence the role of metabolism in determining drug action include exposure to other drugs or xenobiotics, genetic makeup, gender, diet, age, and alterations in physiological status. Concomitant exposure to a second xenobiotic can increase the metabolism of a compound by enzyme induction or decrease its metabolism by enzyme inhibition. For example, phenobarbital increases the metabolism of phenytoin when both are used in the treatment of epilepsy by inducing cytochrome P450 isoforms that oxidize phenytoin.

The genetic makeup of the individual is important because the levels of individual drug-metabolizing enzymes in the human population are genetically determined. Gender and physiological status are important because hormonal factors alter, among other parameters, the levels and types of drug-metabolizing enzymes. The profile of drug-metabolizing enzymes is thus determined by a combination of heredity and environment and varies from individual to individual.

7
Species Differences and Extrapolation to Humans

In view of the sensitivity of drug-metabolizing systems to genetic and environmental factors, it is not surprising that there are major differences in the metabolism of xenobiotics by different species. The species-dependence of drug metabolism is a major concern in the extrapolation to humans of metabolic, pharmacokinetic, and toxicological data obtained with animals. Interspecies variability involves xenobiotic absorption, distribution, and excretion as well as metabolism. Unfortunately, no single animal is a reliable mimic of drug metabolism in humans, although on the average the rhesus monkey is probably one of the better mimics. For example, glucuronidation is a negligible metabolic pathway for oxaprozin in rats, but accounts for roughly half of the total metabolic products in rhesus monkeys and man (Fig. 17). It is not uncommon, however, for other species to predict the metabolism of specific xenobiotics via individual pathways

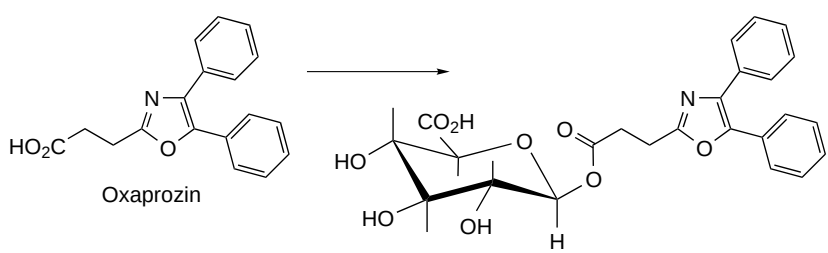

Fig. 17 Glucuronidation of oxaprozin is a major pathway of metabolism in humans and rhesus monkeys but not in rats.

better than the rhesus monkey. The extrapolation and scaling of drug metabolism data from animals to humans, despite its practical importance, is therefore still a difficult and imprecise science.

8
High-throughput Methods in Drug Metabolism Studies

The acceleration of lead candidate discovery by combinatorial synthesis and high-throughput screening methods, combined with the realization that metabolism and pharmacokinetics are critical determinants of a candidate's success, has led to efforts to develop methods that allow a large number of drug candidates to be screened for metabolic properties at an early stage of the drug discovery process. The properties of greatest interest are absorption, metabolism, and the inhibition and induction of individual cytochrome P450 enzymes. The approaches being pursued include computational methods for the prediction of probable metabolic profiles, *in vitro* studies of metabolic stability, reporter constructs in which the activation of receptors responsible for the induction of specific cytochrome P450 enzymes is readily visualized, DNA and protein microarray analysis of changes in metabolic enzyme levels, and high-throughput mass spectrometric approaches to determining actual metabolite profiles.

See also Bioorganic Chemistry; Drug Bioavailability, Distribution and Clearance Prediction.

Bibliography

Books and Reviews

Gibson, G.G., Skett, P. (2001) Introduction to Drug Metabolism, 3rd edition, Nelson Thornes, Cheltenham, UK.

Ioannides, C. (Ed.) (2002) *Enzyme Systems that Metabolise Drugs and Other Xenobiotics*, John Wiley & Sons, New York.

Mulder, G.J. (Ed.) (1990) *Conjugation Reactions in Drug Metabolism. An Integrated Approach*, Taylor & Francis, London, New York.

Ortiz de Montellano, P.R. (Ed.) (1995) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 2nd edition, Plenum Press, New York.

Ortiz de Montellano, P.R. (Ed.) (2004) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd edition, Kluwer Plenum Press, New York, in press.

Phillips, I.R., Shephard, E.A. (Eds.) (1998) *Cytochrome P450 Protocols*, Humana Press, Totowa, NJ.

Testa, B. (1995) *Biochemistry of Redox Reactions; Metabolism of Drugs and Other Xenobiotics*, Academic Press, New York.

Testa, B., Mayer, J.M. (2003) *Hydrolysis in Drug and Prodrug Metabolism*, Wiley-VCH, Zurich.

Woolf, T.F. (Ed.) (1999) *Handbook of Drug Metabolism*, Marcel Dekker, New York.

Primary Literature

Armstrong, R.N. (1999) Kinetic and chemical mechanism of epoxide hydrolase, *Drug Metab. Rev.* **31**, 71–86.

Borhan, B., Jones, A.D., Pinot, F., Grant, D.F., Kurth, M.J., Hammock, B.D. (1995) Mechanism of soluble epoxide hydrolase. Formation of an α-hydroxy ester-enzyme intermediate through Asp-333, *J. Biol. Chem.* **270**, 26923–26930.

Benedetti, M.S., Tipton, K.F. (1998) Monoamine oxidases and related amine oxidases as phase I enzymes in the metabolism of xenobiotics, *J. Neural Transm. Suppl.* **52**, 149–171.

Bu, H.Z., Knuth, K., Magis, L., Teitelbaum, P. (2001) High-throughput cytochrome P450 (CYP) inhibition screening via cassette probe-dosing strategy: III. Validation of a direct injection/on-line guard cartridge extraction-tandem mass spectrometry method for

CYP2C19 inhibition evaluation, *J. Pharm. Biochem. Anal.* **25**, 437–442.

Cashman, J.R. (2000) Flavin monooxygenase: substrate specificity and role in drug metabolism, *Curr. Drug Metab.* **1**, 181–191.

Dashwood, R.H., Arbogast, D.N., Fong, A.T., Pereira, C., Hendricks, J.D., Bailey, G.S. (1989) Quantitative interrelationships between aflatoxin B1 carcinogen dose, indole-3-carbinol anticarcinogen dose, target organ DNA adduction and final tumor response, *Carcinogenesis* **10**, 175–181.

Denison, M.S., Nagy, S.R. (2003) Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals, *Annu. Rev. Pharmacol. Toxicol.* **43**, 309–334.

Denissenko, M.F., Pao, A., Tang, M., Pfeifer, G.P. (1996) Preferential formation of Benzo[a]pyrene adducts at lung cancer mutational hotspots in P53, *Science* **274**, 430–432.

Dreyfuss, J., Ross, J.J. Jr., Shaw, J.M., Miller, I., Schreiber, E.C. (1976) Release and elimination of ^{14}C-fluphenazine enanthate and decanoate esters administered in sesame oil to dogs, *J. Pharm. Sci.* **65**, 502–507.

Fadominska-Pandya, A., Czernik, P.J., Little, J.M., Battaglia, E., Mackenzie, P.I. (1999) Structural and functional studies of UDP-glucuronosyltransferases, *Drug. Metab. Rev.* **31**, 817–899.

Faigle, J.W., Dieterle, W. (1977) The biotransformation of phenylbutazone (Butazolidin), *J. Int. Med. Res.* **5**(Suppl. 2), 2–14.

Fretland, A.J., Omiecinski, C.J. (2000) Epoxide hydrolases: biochemistry and molecular biology, *Chem. Biol. Interact.* **129**, 41–59.

Gerhold, D., Lu, M., Xu, J., Austin, C., Caskey, C.T., Rushmore, T. (2001) Monitoring expression of genes involved in drug metabolism and toxicology using DNA microarrays, *Physiol. Genomics* **5**, 161–170.

Gordon, E.R., Sommerer, U., Goresky, C.A. (1983) The hepatic microsomal formation of bilirubin diglucuronide, *J. Biol. Chem.* **258**, 15028–15036.

Hempel, J., Perozich, J., Chapman, T., Rose, J., Boesch, J.S., Liu, Z.J., Lindahl, R., Wang, B.C. (1999) Aldehyde dehydrogenase catalytic mechanism. A proposal, *Adv. Exp. Med. Biol.* **463**, 53–59.

Hirshey, S.J., Dooley, T.P., Reardon, I.M., Heinrikson, R.L., Falany, C.N. (1992) Sequence analysis, *in vitro* translation, and expression of the cDNA for rat liver minoxidil sulfotransferase, *Mol. Pharmacol.* **42**, 257–264.

http://drnelson.utmem.edu/human.P450.table.html.

Idle, J.R., Mahgoub, A., Lancaster, R., Smith, R.L. (1978) Hypotensive response to debrisoquine and hydroxylation phenotype, *Life Sci.* **22**, 889–904.

Janssen, F.W., Kirkman, S.K., Knowles, J.A., Ruelius, H.W. (1978) Disposition of 4,5-diphenyl-2-oxazolepropionic acid (oxaprozin) in beagle dogs and rhesus monkeys, *Drug Metab. Dispos.* **6**, 465–475.

Jerina, D.M., Daly, J.W. (1974) Arene oxides: a new aspect of drug metabolism, *Science* **185**, 573–582.

Jornvall, H., Hoog, J.O., Persson, B., Pares, X. (2000) Pharmacogenetics of the alcohol dehydrogenase system, *Pharmacology* **61**, 184–191.

Knecht, K.T., Mason, R.P. (1988) *In vivo* radical trapping and biliary secretion of radical adducts of carbon tetrachloride-derived free radical metabolites, *Drug Metab. Dispos.* **16**, 813–817.

Lam, J.P., Mays, D.C., Lipsky, J.J. (1997) Inhibition of recombinant human mitochondrial and cytosolic aldehyde dehydrogenases by two candidates for the active metabolites of disulfiram, *Biochemistry* **36**, 13748–13754.

Le Bigot, J.F., Cresteil, T., Kiechel, J.R., Beaune, P. (1983) Metabolism of ketotifen by human liver microsomes. *In vitro* characterization of a tertiary amine glucuronidation, *Drug Metab. Dispos.* **11**, 585–589.

Lin, J. (1995) Species similarities and differences in pharmacokinetics, *Drug Metab. Dipos.* **23**, 1008–1021.

Lin, J.H., Levy, G. (1986) Effect of prevention of inorganic sulfate depletion on the pharmacokinetics of acetaminophen in rats, *J. Pharmacol. Exp. Ther.* **239**, 94–98.

Mackenzie, P., Owens, I., Burchell, B., Bock, K., Bairoch, A., Bélanger, A., Fournel-Gigleux, S., Green, M., Hum, D., Iyanagi, T., Lancet, D., Louisot, P., Magdalou, J., Chowdhury, J., Ritter, J., Schachter, H., Tephly, T., Tipton, K., Nebert, D. (1997) The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence, *Pharmacogenetics* **7**, 255–269.

McDonagh, A.F., Palma, L.A., Lightner, D.A. (1980) Blue light and bilirubin excretion, *Science* **208**, 145–151.

Mühlebach, S., Bickel, M.H. (1981) Pharmacokinetics in rats of 2,4,5,2′,4′,5′-hexachlorobiphenyl, an unmetabolizable lipophilic model compound, *Xenobiotica* **11**, 249–257.

Nagata, K., Yamazoe, Y. (2000) Pharmacogenetics of sulfotransferase, *Annu. Rev. Pharmacol. Toxicol.* **40**, 159–176.

Nagata, K., Yamazoe, Y. (2002) Genetic polymorphism of human cytochrome P450 involved in drug metabolism, *Drug Metab. Pharmacokinet.* **17**, 167–189.

Nelson, S.D. (1990) Molecular mechanisms of the hepatotoxicity caused by acetaminophen, *Semin. Liv. Dis.* **10**, 267–278.

Nelson, D.R., Kamataki, T., Waxman, D.J., Guengerich, F.P., Estabrook, R.W., Feyereisen, R., Gonzalez, F.J., Coon, M.J., Gunsalus, I.C., Gotoh, O., Okuda, K., Nebert, D.W. (1993) The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names or enzymes, and nomenclature, *DNA Cell Biol.* **12**, 1–51.

Nordling, E., Persson, B., Jornvall, H. (2002) Differential multiplicity of MDR alcohol dehydrogenases: enzyme genes in the human genome versus those in organisms initially studied, *Cell. Mol. Life Sci.* **59**, 1070–1075.

Ohno, Y., Ormstad, K., Ross, D., Orrenius, S. (1984) Mechanism of allyl alcohol toxicity and the protective effects of low-molecular weight thiols, studied with isolated rat hepatocytes, *Toxicol. Appl. Pharmacol.* **78**, 169–179.

Okita, R.T., Okita, J.R. (2001) Cytochrome P450 4A fatty acid omega hydroxylases, *Curr. Drug Metab.* **2**, 265–281.

Paine, M.J.I., Scrutton, N.S., Munro, A.W., Roberts, G.C.K., Wolf, C.R. (2004) Electron transfer partners of cytochrome P450, ibid reference 5, in press.

Patsalos, P.N., Perucca, E. (2003) Clinically important drug interactions in epilepsy: general features and interactions between antiepileptic drugs, *Lancet Neurol.* **2**, 347–356.

Radel, M., Goldman, D. (2001) Pharmacogenetics of alcohol response and alcoholism: the interplay of genes and environmental factors in thresholds for alcoholism, *Drug Metab. Dispos.* **29**, 489–494.

Raucy, J., Warfe, L., Yueh, M.F., Allen, S.W. (2002) A cell-based reporter gene assay for determining induction of CYP3A4 in a high-volume system, *J. Pharmacol. Exp. Ther.* **303**, 412–423.

Rettie, A., Rettenmeier, A.W., Howald, W.N., Baillie, T.A. (1987) Cytochrome P-450-catalyzed formation of Δ^4-VPA, a toxic metabolite of valproic acid, *Science* **235**, 890–893, 28.

Roberts, S.A. (2001) High-throughput screening approaches for investigating drug metabolism and pharmacokinetics, *Xenobiotica* **31**, 557–589.

Salinas, A.E., Wong, M.G. (1999) Glutathione S-transferases – a review, *Curr. Med. Chem.* **6**, 2790309.

Stevens, J., Jakoby, W.B. (1983) Cysteine conjugate β-lyase, *Mol. Pharmacol.* **23**, 761–765.

Smith, C.V., Jones, D.P., Guenthner, T.M., Lash, L.H., Lauterburg, B.H. (1996) Compartmentation of glutathione: Implications for the study of toxicity and disease, *Toxicol. Appl. Pharmacol.* **140**, 1–12.

Umehara, K., Kudo, S., Hirao, Y., Morita, S., Ohtani, T., Uchida, M., Miyamoto, G. (2000) *In vitro* characterization of the oxidative cleavage of the octyl side chain of olanexidine, a novel antimicrobial agent, in dog liver microsomes, *Drug Metab. Dispos.* **28**, 1417–1424.

Vaz, A.D.N., Pernecky, S.J., Raner, G.M., Coon, M.J. (1996) Peroxo-iron and oxenoid-iron species as alternative oxygenating agents in cytochrome P450-catalyzed reactions: switching by threonine-302 to alanine mutagenesis of cytochrome P450 2B4, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4644–4648.

Waxman, D.J., Chang, T.K.H. (1995) Hormonal regulation of liver cytochrome P450 enzymes, ibid reference 4, 391–417.

Weber, W.W., Hein, D.W. (1985) N-Acetylation pharmacogenetics, *Pharmacol. Rev.* **37**, 25–79.

Whitlock, J.P. Jr., Denison, M.S. (1995) Induction of cytochrome P450 enzymes that metabolize xenobiotics, ibid reference 4, 367–390.

Williams, R.T. (1959) *Detoxification Mechanisms*, Chapman & Hall, London.

Williams, S.N., Dunham, E., Bradfield, C.A. (2003) Induction of cytochrome P450 enzymes, ibid reference 5, in press.

Xu, W., Bak, S., Decker, A., Paquette, S.M., Feyereisen, R., Galbraith, D.W. (2001) Microarray-based analysis of gene expression in very large

gene families: the cytochrome P450 gene superfamily of *Arabidopsis thaliana*, *Gene* **272**, 61–74.

Yin, H., Anders, M., Jones, J. (1996) Metabolism of 1,2-dichloro-1-fluoroethane and 1-fluoro-1,2,2-trichloroethane – electronic factors govern the regioselectivity of cytochrome P450-dependent oxidation, *Chem. Res. Toxicol.* **9**, 50–57.

Yoshida, A., Rzhetsky, A., Hsu, L.C., Chang, C. (1998) Human aldehyde dehydrogenase gene family, *Eur. J. Biochem.* **251**, 549–557.

Zamora, I., Arzelius, L., Cruciani, G. (2003) Predicting drug metabolism: A site of metabolism prediction tool applied to the cytochrome P450 2C9, *J. Med. Chem.* **46**, 2313–2324.

Zimmerman, H.J., Maddrey, W.C. (1995) Acetaminophen (paracetamol) hepatotoxicity with regular intake of alcohol: analysis of instances of therapeutic misadventure, *Hepatology* **22**, 767–773.

Body Expression Map of Human Genome

Shinichi Morishita[1] *and Asao Fujiyama*[2]
[1] *Department of Computational Biology, Graduate School of Frontier Sciences, University of Tokyo*
[2] *Biosciences Information Research, Research Information Research Division, National Institute of Informatics*

1 Overview 77
1.1 Genes and Genome 77
1.2 Body-Map Data 79

2 Integration of Millions of EST Alignments on the Human Genome 79
2.1 Association of Gene Expression Patterns with EST Groups on the Human Genome 79
2.2 Identification of Alternative Splice Forms 80
2.3 Data Cleansing and SNP Classification 80

3 Computational Methods 80
3.1 Query into Database of EST Alignments 81
3.2 Efficient Computational Alignment of ESTs with the Human Genome 81
3.3 Resolution of EST Orientations by Alignments 82

4 Use of Human Genome for Observing Gene Expression Patterns 82
4.1 Identification of Less-frequent Subsequences 82
4.2 Selection of Oligomers, and Genome Markers 83

Bibliography 83
Books and Reviews 83
Primary Literature 83
Human Genome Sequencing 83
BodyMap 83
Human DNA Sequence Databases 84
Alignment Algorithms 84

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Alignment with Large-scale Genomic Sequences 84
Oligonucleotide Probe Design 84

Keywords

Gene
A part of genome, which is coded for proteins or templates of many functional RNAs.

Genome
A set of genes, which are required for a particular organism.

Gene Expression
A process in which DNA of a gene is transcribed into RNA, and then translated into protein.

Alignment Algorithm
A computer algorithm used to compare multiple nucleotide or amino acid sequences.

SNP Classification
A classification that is necessary because current single nucleotide polymorphism (SNP) data tend to have errors.

EST Orientation
EST sequences can be aligned onto genomic DNA according to which strand of the genomic DNA is transcribed.

Oligomer Design
Oligo-nucleotide whose sequence is unique within the particular genome is useful for various purposes, for example, designing of DNA-chip.

One of the major objectives of the Human Genome Project (HGP) was to list the entire set of genes and their functions together with control mechanisms written in the human genome. Thus, genome-wide mRNA/cDNA analyses were conducted at about the same time as the major research topic of the Human Genome Project. The first approach was to compile a complete catalog of genes written in the genome of particular organisms, such as human or mouse, through a random collection of expressed cDNA sequences (expressed-sequence tags, ESTs). The second approach has more focus on the status of expressed genes through the acquisition of expression profiles of mRNA/cDNA (body mapping) isolated from a particular cell group, tissue, or organ. Much of the EST approach, such as the gene cataloguing project,

is qualitative. In contrast, body-mapping projects are designed to clarify the differences of cells or tissues in terms of gene expression profiles; thus, they have to be quantitative at every step during the experiment. Integration of millions of EST alignments on the human genome needs browsers that facilitate efficient searching and browsing of an enormous quantity of EST alignments. To achieve this goal, it is essential to meet specific computational requirements, such as the acceleration of sensitive-but-slow (dynamic programming) alignment algorithms, and the resolution of EST orientations. Genome browsers such as Ensembl, UCSC, NCBI, and GRL support these functions.

1 Overview

Nearly two years after the publication of the historic paper on the human draft genome sequence and initial analyses, sequencing of most of the euchromatin region of the human genome is near completion, ushering in the age of genome science (announcement of the completion of human genome was made in April 2003). The complete sequence of the human genome or any genome will provide researchers and the society with valuable information concerning the positions and primary structures of genes, genetic markers, repetitive sequences, functional elements, and lines of evidence of chromosomal rearrangement during human evolution. Many of these aspects are, at present, incompletely understood, and await intensive studies in the future by scientists from various disciplines. Structural analyses of other functionally important elements, such as centromeric and telomeric regions that are buried in heterochromatin must wait for a longer period. In addition, the sequences and position data will be updated for many years.

However, the sequence we will see in a database of a human genome is a sort of patchwork of several genomes because of the technical considerations relating to the sequencing and preparation of DNA materials. Thus, the sequence in the public databases (GenBank, DDBJ, and EMBL) should be regarded as a reference set for our genome. It is very important, when we look at the sequence data, that we keep in mind that the human genome does not reflect the heterogeneity of human population. Because of this, variation studies, such as identification of Single nucleotide polymorphisms (SNPs) and typing of haplotype patterns using various human subpopulations will be important to identify the functional differences linked with specific genomic variation. In other words, functional studies of personal differences will be the basis of the medical and pharmaceutical studies in the future.

In this chapter, a set of databases and tools that we believe will be useful to such studies will be presented. Statistical and mapping data presented in the later section of this chapter are calculated using human genome assembly data available in January 2003.

1.1 Genes and Genome

One of the major objectives of the Human Genome Project (HGP) was to list

the entire set of genes and their functions together with control mechanisms written in the human genome. However, thanks to the progress of recent molecular biology/genetics over the past decade, we now know that genes are not mere stretches of nucleotides but are dynamically controlled, depending on, for example, the developmental stages, and the conditions of the outer environments of cells or organisms. Thus, genome-wide mRNA/cDNA analyses were conducted at about the same time as the major research topic of the Human Genome Project.

Several different approaches have been tested and are still in use for these purposes. The first approach was to compile a complete catalog of genes written in the genome of particular organisms such as human or mouse through a random collection of expressed cDNA sequences (expressed-sequence tags, ESTs) from various tissues. Once as many of these as possible have been collected, computational methods involving clustering were used to compile each of the cDNA fragments into groups of mRNA sequences, based on their sequence similarities. In other words, the final goal of this kind of approach is to make a list of entire genes and their genomic localizations linked with annotated databases. Recent development of a high-throughput computational approach is discussed in the later part of this chapter.

The second approach has more focus on the status of expressed genes through the acquisition of expression profiles of mRNA/cDNA (body mapping) isolated from a particular cell group, tissue, or organ. Human body mapping was first started using quantitatively prepared cDNA libraries constructed from mRNA isolated from anatomically purified cells or tissues through microscopic dissection, cell culture, and other cell purification technologies (Fig. 1). For the experimental details of body mapping, refer to the previous edition of this encyclopedia. An alternative approach to determine the expression profiles is SAGE and DNA microarrays that will be discussed in other chapters.

Expression profiling is the only practical approach to quantify the extent and to characterize the specificities of gene expression in a particular cell at a specific time, through the identification of gene signature, elements of gene expression control, and the network of gene

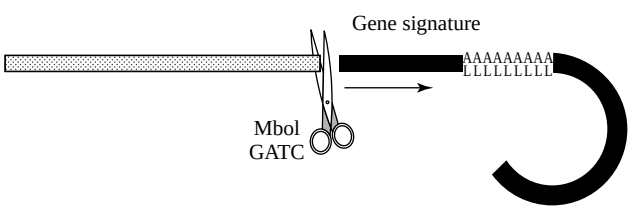

Fig. 1 Libraries and data collection: All libraries used, harbor only the 3′-terminal cDNA fragments made by digesting vector-primed cDNA molecules with *Mbo*I (GATC) in order to facilitate the 3′-EST collection and to minimize the difference in cloning efficiencies among transcripts of different sizes. They were not subjected to amplification, prescreening, or normalization. Accordingly, the clonal recurrence roughly represents transcript abundance in this library.

functions. Altogether, combined with genomic sequences, we will understand the exact mechanisms and signals of gene expression control written in genomic DNA in the near future through experimental data obtained from such studies.

1.2 Body-Map Data

Much of the EST approach, such as the gene cataloguing project, is qualitative. In contrast, body-mapping projects are designed to clarify the differences of cells or tissues in terms of gene expression profiles; thus, they have to be quantitative at every step during the experiment. In general, the following points have to be considered very carefully when carrying out an expression-profiling experiment or interpreting the experimental data obtained from such projects. First, it is important to know the origin and quality of the tissue or cells from which cDNA libraries are prepared. This is especially important when comparing data obtained from different mRNA sources such as tissues, organs, individuals, or species. Secondly, the quality and complexity of the cDNA library have to be carefully examined. Thirdly, any sampling procedure potentially has biases that will severely affect the composition of the final dataset. Without these precautions and evaluation of starting materials and experimental steps, the data will lead to improper conclusions. Statistical assessment of the data as well as the linearity of measuring technology are also important, especially when using fluorescently labeled material because both body mapping and DNA chip technology have experimental characteristics that users have to be aware of when interpreting the final data.

2 Integration of Millions of EST Alignments on the Human Genome

Before the human draft genome became available, ESTs were classified primarily on the basis of sequence similarities. UniGene, for example, employs this method to produce clusters of ESTs. In contrast, the advent of the draft human genome sequence makes it possible to align millions of ESTs with the genomic sequence. Integration of these alignments is helpful in identifying groups of ESTs that are coded at the same locus, in gathering information on alternatively spliced transcripts and their representatives (see Fig. 2), and in associating gene expression patterns with specific loci. It is noteworthy that these two approaches are likely to output different clusters for ESTs that have low-level sequence similarities. This is because the traditional sequence-similarity approach tends to overlook a typical cluster of alignments that share a small number of exons at the same positions.

2.1 Association of Gene Expression Patterns with EST Groups on the Human Genome

Correlation of expression patterns with loci of gene clusters is a crucial step in functional analysis. In an effort to achieve this task, one can attempt to align ESTs associated with gene expression patterns in various tissues with the human genome. For instance, the expression levels of BodyMap representative genes in 30 distinct human tissues are available. BodyMap sequences are hundreds of base pairs in length and are therefore sufficiently long to locate their positions in the human genome. This makes it possible

Fig. 2 Identification of alternatively spliced transcripts and their representatives: there are millions of redundant EST alignments, because it is possible for several distinct ESTs to map to the same locus. In this case, alignments that share common exons on the same strand should be placed in the same group. In the figure, all of the EST sequences are aligned in the 5′ to 3′ direction and displayed from left to right. Each thick line represents the alignment of one EST, in which the narrow yellow boxes are exons, orange boxes are protein coding regions, and the brown boxes are introns. (See color plate p. xxi.)

to annotate about 10 500 EST loci with BodyMap gene expressions.

2.2 Identification of Alternative Splice Forms

Integration of millions of EST alignments enables us to collect alternatively spliced transcripts and their representatives. Figure 2 illustrates a group of alignments in which the second line from the bottom represents the EST alignment of the RefSeq sequence (TAC1, NM_003182) that contains seven exons, while some alternatively spliced transcripts do not use the fourth and sixth exons. Since the protein-coding region (CDS) in the second-from-bottom alignment starts with the second exon and ends with the last (seventh) exon, the bottom alignment, which skips the sixth exon, actually encodes a different protein than the second-from-bottom alignment. Similarly, the third- and fourth-from-bottom alignments do encode distinct proteins.

2.3 Data Cleansing and SNP Classification

Single nucleotide polymorphisms, DNA sequence variations among individuals, have been collected and are stored in databases such as dbSNP (http://www.ncbi.nlm.nih.gov/SNP) and Japanese SNP (http://snp.ims.u-tokyo.ac.jp/). The entry of each SNP involves two sequences that are located before and after the SNP nucleotide, for the identification of the SNP nucleotide, thereby demanding the alignments of the two sequences with the human genome specify the unique location of the SNP nucleotide. Thus, selection of those sequences that map with at least 99% identity to a unique location in the draft genome is effective in eliminating the incorrect SNPs. Furthermore, integration of millions of EST alignments is helpful in classifying the aligned SNPs as regulatory, coding, or noncoding, according to their locations.

3 Computational Methods

Integration of millions of EST alignments on the human genome needs browsers that facilitate efficient searching and browsing of an enormous quantity of EST alignments. To achieve this goal, it is essential to meet specific computational requirements, such as the acceleration of sensitive-but-slow (dynamic programming) alignment algorithms, and the resolution of EST orientations.

3.1
Query into Database of EST Alignments

Since millions of EST alignments with the huge human genome constitute a very large database, various ways to access alignments are indispensable to obtain valuable information. For instance, in a positional cloning project for hunting a human disease gene, if some linkage data suggest that the gene of interest lies between two sequence-tagged site markers, it is informative to enumerate all the known genes and the single nucleotide polymorphisms in the interval between the two markers. It would then follow the precise analysis of alternatively spliced transcripts or individual SNPs that fall within the coding region of a gene. Furthermore, association of a BodyMap sequence with the coding region provides the expression levels of 30 distinct human tissues as supplementary information. These tasks are comfortably facilitated by devising query and search functions in the database that accepts the STS marker name, the GenBank accession number, the RefSeq symbol name, the BodyMap GS number, or the raw sequence of nucleotides of interest. It is also helpful to provide a graphical interface that is capable of browsing genes in one chromosome or in one BAC contig, the exon/intron structure of alternatively spliced transcripts, and the exact positions of SNPs by zooming in and out of the map smoothly and seamlessly. Genome browsers such as Ensembl, UCSC, NCBI, and GRL partly support these functions.

3.2
Efficient Computational Alignment of ESTs with the Human Genome

Here, we briefly mention the key ideas relating to efficient and complex computational methods to align four million ESTs to a newly revised draft genome in just one day. Figure 3 depicts an efficient way of processing millions of ESTs in a reasonable amount of time while retaining sensitivity. The algorithm consists of three key steps. The first step is to build an index lookup table (hash table) of the positions of all the nucleotide sequences of length N (called N-mers) in the DNA sequence. The index aids in locating a single N-mer, say 8mers, instantly, and this represents a crucial step in accelerating the overall performance of mapping millions of ESTs in the genome. The second step is to approximate the start and end positions of each EST alignment by mapping, for instance, 12mers at each end of an EST, which is illustrated in the upper part of the figure. The algorithm scans the EST from

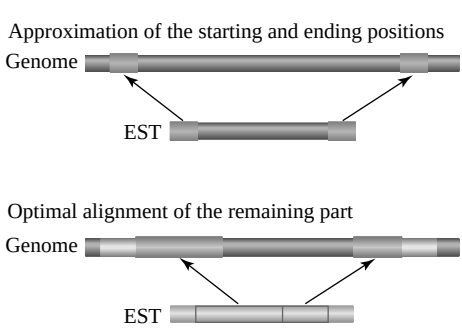

Fig. 3 Accelerated dynamic programming of EST alignments using an index lookup.

the 5′-end until it finds one 12mers that fully matches to the genome; some 12mers are not immediately evident because of errors in the EST sequences. Similarly, the algorithm scans the EST from the 3′-end. The third step is to align the intervening sequence, that is, the sequence between the 12mers at the start and end. The lower part of the figure illustrates a special case, in which our algorithm fractionates an EST into two exons and aligns the exons to the DNA sequence. This method is generally capable of processing ESTs that have more than two exons or intronless ESTs. In situations in which there are multiple potential candidates at the start and end positions, the intervening sequences are investigated with every combination of ends. At the intron boundaries between two exons, it is necessary to move the windows of the two exons so that the exon/intron junctions obey the so-called "GT. . .AG" rule.

3.3
Resolution of EST Orientations by Alignments

Individual ESTs are aligned against both the plus and minus strands of the genome. However, even if an EST is aligned to one strand, the EST might actually be read from the other strand. To resolve EST orientation, considering the combinations of EST alignments on genome is effective. The standard test would be to check whether the EST contained poly-A stretches or polyadenylation signals. The existence of a poly-T subsequence at the beginning of an EST indicates that the sequence should be reversed and then complemented to achieve the correct orientation. However, this check does not always work, since the 3′-end sequences are lacking for many ESTs. Another useful rule of thumb is to look at the introns of each EST alignment and to infer the orientation by utilizing the "GT. . .AG" rule. For example, if the intron boundary is "CT. . .AC," the alignment needs to be reversed and complemented. However, these two rules are not able to identify the frequently encountered, intronless EST alignments that have neither poly-A stretches nor polyadenylation signals. In this case, it is necessary to see whether the alignment overlaps another EST alignment whose strand orientation has been already confirmed, thereby allowing one to assign the ambiguous EST alignment to the confirmed strand.

4
Use of Human Genome for Observing Gene Expression Patterns

Sequences such as cDNAs, mRNAs, ESTs, and partial fragments of human genome have been the primary sources of designing primers and oligomers for observing gene expression patterns before the elucidation of human genome. Because the human genome involves all the sequence information, full utilization of the human genome may yield novel methods useful in designing primers and oligomers.

4.1
Identification of Less-frequent Subsequences

Traditional methods of primer design make it difficult to select primers that hybridize at only one position. Although RepBase encompasses well-known repetitive sequences, such as Alu, LTR, and LINE, its coverage of less-frequent repetitive sequences is incomplete. Fortunately,

the use of entire genomic sequences appears to offer an effective solution to these problems, because subsequences that appear at extremely low frequencies in the genome would be good primers. Thus, the number of occurrences (frequency) of a subsequence of length N (an N-mer) would serve as a criterion to assess subsequence uniqueness. Subsequently, our primary interest is to list N-mers that appear only once in the genome, that is, unique N-mers. The number of unique N-mers in human genome expands as N increases, but the number converges once N exceeds 18. There are about 1.7 billion unique 18mers, indicating that unique 18mers make up approximately half of all the 3 billion 18mers, and efficiently cover the entire human genome.

4.2 Selection of Oligomers, and Genome Markers

As mentioned above, to evaluate the specificity of primers about 20-nucleotides long, the first step is to determine how frequently they occur. However, for longer oligomers or genome markers of about 50 nucleotides, determination of their frequencies is not useful because the frequencies are usually 1. A more suitable procedure would be to consider the mismatch tolerance of an oligomer, that is, the minimum number of mismatches that allow a given oligomer to match a subsequence other than the target sequence anywhere in the genome. Although calculating the exact value of mismatch tolerance is computationally costly and impractical, it becomes feasible to check if an oligomer meets the relaxed constraint that its mismatch tolerance is no less than a given threshold.

See also DNA Libraries; Genetics, Molecular Basis of; Genomic DNA Libraries, Construction and Applications; Gel Electrophoresis, 2D-difference.

Bibliography

Books and Reviews

Ewens, W.J., Grant, G.R. (2001) *Statistical Methods in Bioinformatics*, Springer Verlag.

Gusfield, D. (1997) *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*, Cambridge University Press.

Mount, D.W. (2001) *Bioinformatics: Sequence and Genome Analysis*, Cold Spring Harbor Laboratory.

Pevzner, P.A. (2000) *Computational Molecular Biology: An Algorithmic Approach*, The MIT Press.

Waterman, M.S. (1995) *Introduction to Computational Biology: Maps, Sequences and Genomes*, CRC Press.

Primary Literature

Human Genome Sequencing

Boguski, M.S., Schuler, G.D. (1995) Establishing a human transcript map, *Nat. Genet.* **10**, 369–371.

Collins, F.S., Green, E.D., Guttmacher, A.E., Guyer, M.S. (2003) A vision for the future of genomics research, *Nature* **422**, 835–847.

Lander, E.S., et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

BodyMap

Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y., Matsubara, K. (1992) Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression, *Nat. Genet.* **2**, 173–179.

Okubo, K., Matsubara, K. (1997) Body expression map of the human genome, *Encycl. Mol. Cell Biol. Mol. Med.* 203–206.

Sese, J., Nikaidou, H., Kawamoto, S., Minesaki, Y., Morishita, S., Okubo, K. (2001) BodyMap incorporated PCR-based expression profiling data and a gene ranking system, *Nucleic Acids Res.* **29**, 156–158.

Human DNA Sequence Databases

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L. (2004) GenBank: update, *Nucleic Acids Res.* **32**, D23–D26.

Birney, E. et al. (2004) Ensembl 2004, *Nucleic Acids Res.* **32**, D468–D470.

Gabriel, S.B. et al. (2002) The structure of haplotypes blocks in the human genome, *Science* **296**, 2225–2229.

Galperin, M.Y. (2004) The molecular biology database collection: 2004 update, *Nucleic Acids Res.* **32**, D3–D22.

Honkura, T., Ogasawara, J., Yamada, T., Morishita, S. (2002) The gene resource locator: gene locus maps for transcriptome analysis, *Nucleic Acids Res.* **30**, 221–225.

Jurka, J. (2000) RepBase update: a database and an electronic journal of repetitive elements, *Trends Genet.* **9**, 418–420.

Karolchik, D. et al. (2004) The UCSC table browser data retrieval tool, *Nucleic Acids Res.* **32**, D497–D501.

Kulikova, T. et al. (2004) The EMBL nucleotide sequence database, *Nucleic Acids Res.* **32**, D27–D30.

Miyazaki, S. et al. (2004) DDBJ in the stream of various biological data, *Nucleic Acids Res.* **32**, D31–D34.

Pruitt, K.D., Maglott, D.R. (2001) RefSeq and LocusLink: NCBI gene-centered resources, *Nucleic Acids Res.* **29**, 137–140.

Sherry, S.T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation, *Nucleic Acids Res.* **29**, 308–311.

The International SNP Map Working Group. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms, *Nature* **409**, 928–933.

Thorisson, G.A., Stein, L.D. (2003) The SNP consortium website: past, present and future, *Nucleic Acids Res.* **31**(1), 124–127.

Wheeler, D.L. et al. (2004) Database resources of the national center for biotechnology information: update, *Nucleic Acids Res.* **32**, D35–D40.

Alignment Algorithms

Altschul, S.F., Gis, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool, *J. Mol. Biol.* **215**, 403–410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* **25**(17), 3389–3402.

Needleman, S.B., Wunsch, C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins, *J. Mol. Biol.* **48**, 443–453.

Pearson, W.R., Lipman, D.J. (1988) Improved tools for biological sequence comparison, *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.

Smith, T.F., Waterman, M.S. (1981) Identification of common molecular subsequences, *J. Mol. Biol.* **147**, 195–197.

Alignment with Large-scale Genomic Sequences

Florea, L., Hartzell, G., Zhang, Z., Rubin, G.M., Miller, W. (1998) A computer program for aligning a cDNA sequence with a genomic sequence, *Genome Res.* **8**(9), 967–974.

Kent, W.J. (2002) BLAT–the BLAST-like alignment tool, *Genome Res.* **12**(4), 656–664.

Ma, B., Tromp, J., Li, M. (2002) PatternHunter: faster and more sensitive homology search, *Bioinformatics* **18**(3), 440–445.

Ogasawara, J., Morishita, S. (2003) Fast and sensitive algorithm for aligning ESTs to human genome, *J. Bioinf. Comp. Biol.* **1**(2), 363–386.

Schwartz, S. et al. (2000) PipMaker–a web server for aligning two genomic DNA sequences, *Genome Res.* **10**(4), 577–586.

Schwartz, S. et al. (2003) Human mouse alignments with BLASTZ, *Genome Res.* **13**(1), 103–107.

Zhang, Z., Schwartz, S., Wagner, L., Miller, W. (2000) A greedy algorithm for aligning DNA sequences, *J. Comput. Biol.* **7**(1–2), 203–214.

Oligonucleotide Probe Design

Kaderali, L., Schliep, A. (2002) Selecting signature oligonucleotides to identify organisms using DNA arrays, *Bioinformatics* **18**(10), 1340–1349.

Li, F., Stormo, G. (2001) Selection of optimal DNA oligos for gene expression arrays, *Bioinformatics* **17**(11), 1067–1076.

Rahmann, S. (2002) Rapid large-scale oligonucleotide selection for microarrays, *Proc. of IEEE Computer Society Bioinformatics Conference*, pp. 54–63.

Rouillard, J.-M., Herbert, C.J., Zuker, M. (2002) OligoArray: genome-scale oligonucleotide design for microarrays, *Bioinformatics* **18**(3), 486–487.

Sung, W.-K., Lee, W.-H. (2003) Fast and accurate probe selection algorithm for large genomes, *Proc. of IEEE Computer Society Bioinformatics Conference*, pp. 65–74.

Yamada, T., Morishita, S. Computing highly specific and noise-tolerant oligomers efficiently, *J. Bioinf. Comp. Biol.* (in press).

Zheng, J., Close, T.J., Jiang, T., Lonardi, S. (2003) Efficient selection of unique and popular oligos for large EST databases, *Proc. of the 14th Annual Symposium on Combinatorial Pattern Matching*, pp. 384–401.

Brain Development

Takayoshi Inoue
National Institute of Neuroscience, 4-1-1, Ogawahigashi, Kodaira, Tokyo 187-8502, Japan

1 Neural Induction – The First Step Involved in Forming Neural Tissue 89

2 Early Patterning of the Neural Tissue Along the A–P and D–V Axes 91
2.1 D–V Patterning in the Neural Plate/Tube 91
2.2 A–P Patterning in the Neural Plate/Tube 98
2.2.1 Hindbrain Specification 99
2.2.2 Midbrain Specification 103
2.2.3 Forebrain Specification 104

3 Cellular Aspects to Generate Distinct Types of Cells and/or Cellular Organization (e.g. Nuclei, Layers, and so on) in Patterned Tissue 107

4 Molecular Aspects of Generating Distinct Types of Cells and/or Cellular Organization (e.g. Nuclei, Layers, and so on) in Patterned Tissue 109
4.1 Generation of Neurons 109
4.1.1 Cell Intrinsic Mechanisms 109
4.1.2 Cell Nonautonomous Mechanisms 109
4.2 Nuclei Formation 112
4.3 Layer Formation 112

5 Cellular Aspects to Wire up the Nerve Cells into Functional Circuits 113

6 Molecular Aspects to Wire up the Nerve Cells into Functional Circuits 114
6.1 Contact-dependent (Short Range) Axon Guidance Mechanisms 116
6.1.1 Ig Superfamily Molecules 116
6.1.2 Integrins 117
6.1.3 Cadherins 117

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

6.2 Contact-independent (Long-range) Axon Guidance Mechanisms 118
6.3 Plasticity of the Neuronal Circuit-potentiation of Synaptic Transmission 119

7 How can Molecular Functions in Complex Tissues be Studied? 120
7.1 Loss-of-function Studies 120
7.2 Gain-of-function Studies 121

Acknowledgments 123

Bibliography 123
Books and Reviews 123
Primary Literature 123

Keywords

Neural induction
The first step involved in forming neural tissue.

Early patterning
Steps involved in dividing neural tissue along the A–P and D–V axes.

Differentiation
Generation of distinct types of cells and/or cellular organization (e.g. nuclei, layers, and so on) in the patterned tissue.

Wiring
Formation of cell processes and connections into functional neuronal circuits.

The human brain is an intricate organ with over a hundred billion cells. Brain activity is precisely coordinated to generate electrical and hormonal signals that control basic body functions, movement, awareness, and behaviors. Intriguingly, the incredible complexity of the brain is formed from a simple cellular sheet called the *neural plate*. A progressive series of precisely controlled morphogenetic movements, in combination with cell proliferation, differentiation, and specification of regional character, generates functionally distinct brain structures. In this article, I summarize some of the crucial events happening at the cellular and molecular levels during neural development to help understand the complex processes implicated in vertebrate brain organization.

1
Neural Induction – The First Step Involved in Forming Neural Tissue

One of the most important events that occur during early vertebrate embryogenesis is *gastrulation*, by which all the three germ layers; *ectoderm, mesoderm*, and *endoderm* are first established. Neural tissue is formed in the ectodermal layer overlying the mesodermal tissue. In 1924, Spemann and Mangold demonstrated that during early gastrulation a distinct population of cells around the dorsal lip in amphibian embryos had the ability to induce an additional head and trunk if grafted into the opposite site of the dorsal lip (i.e. presumptive epidermis) of a host embryo (Fig. 1a). Notable was the fact that the cells in the newly formed head and/or trunk structures were totally derived from the host embryo, indicating that the grafted tissue had an instructive role in organizing the head and trunk tissues. The cells that harbored the ability to induce the secondary axis in the embryo were named *organizer*. The organizer tissue in their experiment had the ability to pattern the mesoderm, including the *notochord* (Fig. 1a). Similar organizer regions have been shown to have a role in the formation and patterning of the axial mesoderm in various organisms, and these regions indeed induce additional neural tissues/axes in developing embryos. Collectively, such processes are termed *neural induction*. While Spemann and colleagues proposed that distinct tissues appear to induce either anterior or posterior characters in the neural tissue of amphibian embryos, Nieuwkoop later suggested a two-step model. In this case, an early activating signal induces neural tissue with an anterior character (i.e. forebrain), and a second posteriorizing or transforming signal that converts neural tissues into a more posterior character further produces regional differences along the axis (i.e. hindbrain and spinal cord). Both classical models that explain frog neural induction involve a degree of neural patterning as very early events. In mice, three distinct tissues, the early gastrula organizer (EGO), anterior visceral endoderm (AVE), and anterior epiblasts, have been implicated in playing a critical role in the induction of anterior neural structures during early gastrulation stages (Fig. 1c). EGO is a cell population in the early primitive-streak stage of the embryo, which displays the cellular properties typical of Spemann's frog organizer. The transplanted EGO/node region indeed has the ability to induce the secondary axis with posterior neural characters, but the AVE as well as the anterior epiblast are further required to induce complete sets of anterior neural structures in mouse embryos, indicating that multiple signaling mechanisms should be involved in neural induction events to simultaneously generate regional differences along the Anterior–Posterior (A–P) axis. I will come back to this point later (Sect. 2,on A–P patterning).

Basic molecular mechanisms involved in neural induction appear to be conserved among vertebrates and invertebrates, although tissue organization of the three germ layers varies from species to species. Key molecules determining neural and nonneural tissues are a signaling factor, Decapentaplegic (Dpp)/bone morphogenetic protein (BMP), and its antagonist secreted from the axial mesoderm, Shorted gastrulation (Sog)/Chordin. BMP signaling participates in the process that converts cells into surface ectoderm (skin), and cells in which the Dpp/BMP signal is antagonized by Sog/Chordin inhibitors are fated to form neural tissue (Fig. 1a). It is now known, however, that vertebrate neural

tissue can be generated by the serial inductive actions of surrounding tissues prior to gastrulation without the organizer and/or axial mesoderm signals. For example, even surgical ablations of the notochord as well as genetic manipulations to mutate genes expressed in the organizer, that harbors a role in the formation of the notochord, result in neural tissue formation. Recently, using early chicken embryos as a good

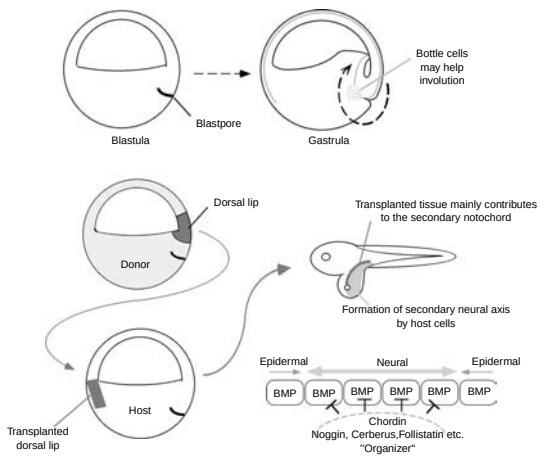

(a) Amphibian

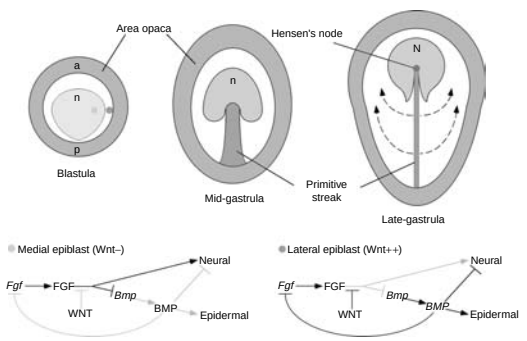

(b) Chicken

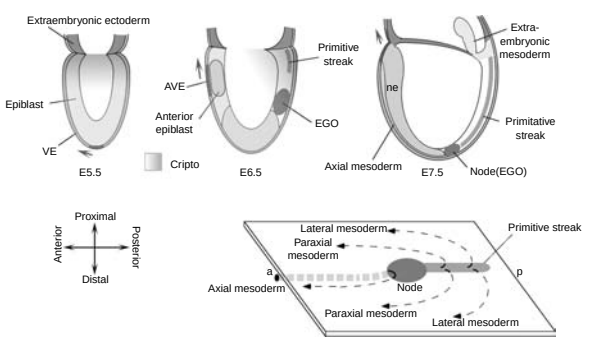

(c) Mouse

model system, signaling machineries involved in neural induction were examined, and a new genetic cascade has been proposed (Fig. 1b). According to the hypothesis, a balance of Wnt (see Fig. 2b) and FGF (fibroblast growth factor) signaling pathways is the most critical aspect of neural induction at earlier stages (Fig. 1b). The BMP signaling that finally determines the neural tissue is well-incorporated in this model, yet it remains elusive to what degree this hypothesis can explain the intricate machineries required to establish neural tissue among various vertebrate species.

2 Early Patterning of the Neural Tissue Along the A–P and D–V Axes

Neural induction events determine the definitive area of future neural tissue within the single-layered ectoderm, termed *neuroectoderm*. The neuroectoderm consists of tall epithelial cells and is recognizable as a sheet or plate. All complex structures of the brain emerge from this single-layered cellular sheet termed the *neural plate* (Fig. 3a). Importantly, the neural plate has been exposed to different environments along the A–P and dorsal–ventral (D–V) axes in the embryo. This allows the neural plate to further adopt distinct features in the central nervous system (CNS). I next outline early patterning events in the neural plate along the A–P and D–V axes.

2.1 D–V Patterning in the Neural Plate/Tube

The neural plate rolls up, detaches from the nonneural ectoderm, and finally zips along the dorsal midline of the embryo to form a tubelike structure, called the *neural tube* (Fig. 3a). The D–V axis in the

Fig. 1 Gastrulation and neural induction in various vertebrate species. (a) Gastrulation and neural induction in an amphibian embryo. During gastrulation, bottle cells are observed around the blastpore, allowing the involution of cells (a dotted arrow). Amphibian neural induction events appear to be very dependent upon BMP (bone morphogenetic protein) signaling inhibition, in which BMP antagonists such as Chordin, Noggin, Cerberus, Follistatin, and so on secreted from the organizer play crucial roles. Regarding BMP signaling machineries, see Fig. 2(a) as BMP is a member of the TGF-β superfamily. (b) Gastrulation and neural induction in chicken embryo. During gastrulation, mesodermal cells detach from the epiblast layer and emigrate anterior-laterally (dotted arrows), forming the primitive streak. In the latest model, chicken neural tissue can be formed by FGF and Wnt signals earlier than gastrulation: At the blastula stage, only lateral epiblast (green) expresses Wnt3A/8C, demarcating medial epiblast as the prospective neural tissue (light pink region). The Wnt receptor, Frizzled8, is ubiquitously expressed by all epiblasts at this stage (regarding the Wnt signaling pathway, see Fig. 2(b)). a, anterior; N, neural tissue; n, future neural tissue, p, posterior. (c) Gastrulation, neural induction, and early anterior–posterior axis formation during mouse development. Blue arrows indicate the crucial movement of cells in the endodermal layer, determining anterior characters of the neuroectoderm. Expression of the secreted molecule, Cripto, is dynamically regulated at earlier stages, and this is thought to be critical to initiate A–P patterning events, including the endodermal cell movement as well as primitive-streak formation (purple region). During gastrulation, mesodermal cells detach from the epiblast layer (light green region) and emigrate just like those in the chicken embryo (dotted arrows). a, anterior; AVE, anterior visceral endoderm (dark blue region); E, embryonic day; EGO, early gastrula organizer (red region); ne, neuroectoderm (pink region); p, posterior; VE, visceral endoderm (light blue region). The light gray shading indicates Cripto expression, and the axial mesoderm is colored by orange. (See color plate p. xxii).

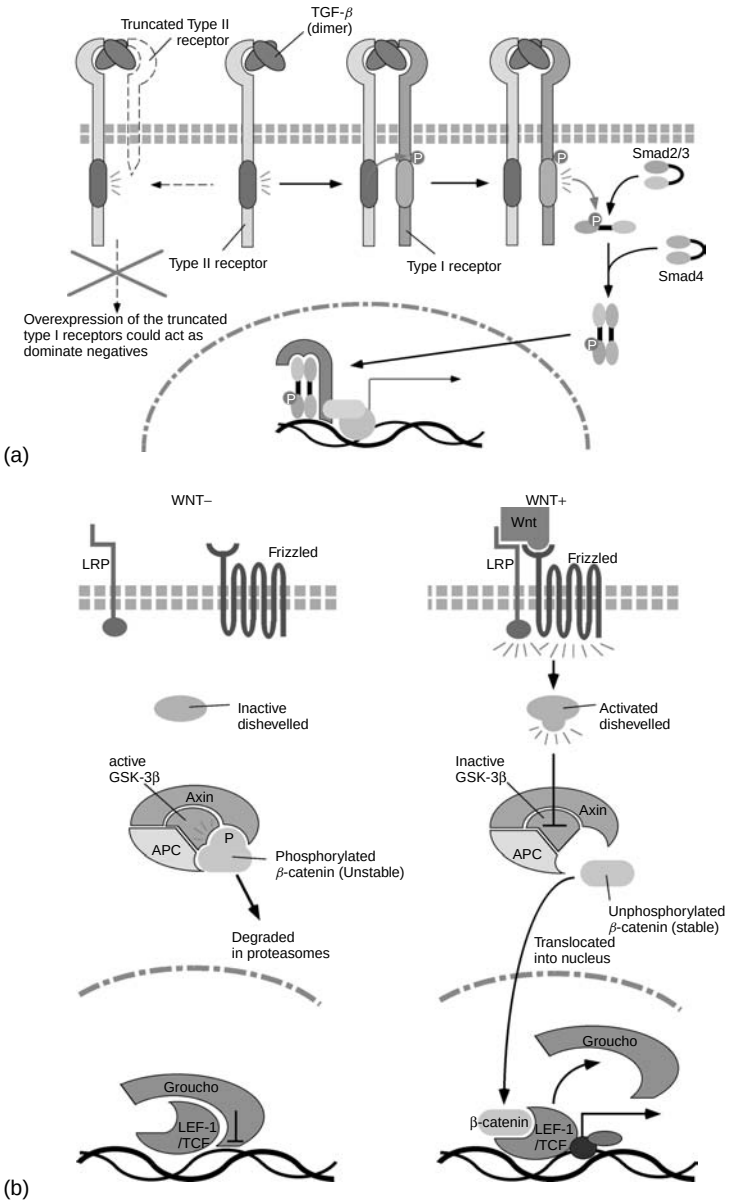

Fig. 2 (a) TGF-β signaling machineries. (b) The canonical Wnt signaling pathway.

neural tube therefore corresponds to the mediolateral axis in the early neural plate: both lateral ends of the neural plate, named the *neural ridge*, become the dorsalmost portion of the neural tube, while the midline of the neural plate, which overlies the axial mesoderm notochord, becomes the ventralmost portion of the neural

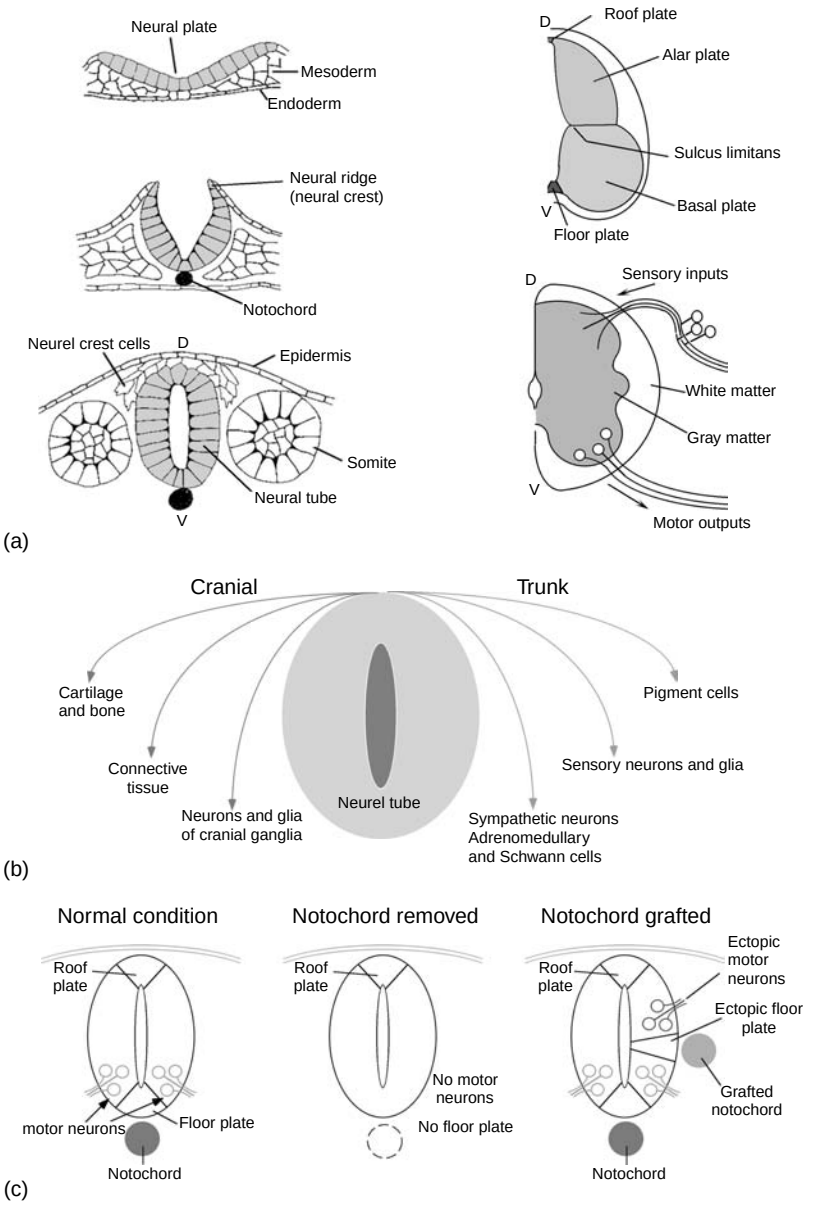

Fig. 3 Neurulation and D–V patterning of the neural tube. (a) Neurulation and D–V patterning of the neural tube during vertebrate development. D, dorsal; V, ventral. (b) Neural crest cells can differentiate into various cell types. Note that cranial neural crest cells are different from trunk neural crest cells in generating a distinct set of cell types *in vivo*. (c) The role of the notochord in the neural tube D–V patterning.

tube (Fig. 3a). During the morphogenetic processes to form the neural tube, termed *neurulation*, a group of cells delaminates from the neural ridge/dorsal midline (*neural crest*) of the neural tube and migrates to generate the *neural crest cells*. These cells differentiate into components of the peripheral nervous system (PNS) as well as other types of cells including bones (Fig. 3b). The neural crest cells can produce such a variety of cell types that they have often been designated as the fourth germ layer. The timing of neural crest cell emigration varies among species. For example, mammalian cranial neural crest cells emigrate before the neural tube is closed, while chicken neural crest cells first migrate just after the neural tube closure. The neurulation step also varies from species to species. For instance, in fish embryo, neuroectoderm (the neural keel) first segregates out from the ectoderm layer as a group of cells, and later a cavity is generated in the middle of the segregated tissue to form the neural tube. Another unique process, observed in the posterior part of mouse and chicken embryos, is termed *secondary neurulation*. In this process, neural tissues are segregated from a mixture of cells in the tail bud and this forms the remainder of the neural structures in the spinal cord. The dorsalmost cells in the neural tube differentiate to constitute the *roof plate*, while the ventralmost cells form the *floor plate* with distinct features and functions (Fig. 3a). The *sulcus limitans* is an anatomically distinguishable boundary between the dorsal and ventral neural tube, and the dorsal half of the neural tube subdivided by this landmark is termed the *alar plate*, while the ventral half is called the *basal plate* (Fig. 3a). In the mature spinal cord, sensory inputs from the peripherals (i.e. dorsal root ganglia) innervate the interneurons in the alar plate (Fig. 3a). Then, the outputs from interneurons either innervate the brain or directly rule the motor neurons in the basal plate, generating the motor outputs to the peripherals (i.e. muscles) (Fig. 3a). This organization along the D–V axis in the spinal cord is also represented in the rest of the CNS.

D–V differences in the neural tube are patterned by the surrounding tissues, among which the notochord and epidermal ectoderm (i.e. surface ectoderm) seem to play pivotal roles. For example, if the notochord is eliminated at early developmental stages, no ventral identities are known to emerge (Fig. 3c). Furthermore, if the notochord is grafted into the medial region of the early chicken spinal cord, it has been demonstrated that components in the basal plate, such as the floor plate and motor neurons, are ectopically induced (Fig. 3c). The floor plate has also been shown to have inducing activity, and a component of the signaling molecules is the secreted molecule Sonic hedgehog (Shh). Both the notochord and floor plate cells express this molecule during early neural development, and experimental data support a working hypothesis whereby several types of motor neurons as well as interneurons are produced in response to the Shh gradient along the D–V axis. According to the hypothesis, the cells exposed to the highest dose of Shh become the floor plate cells, while those exposed to lower doses constitute the distinct progenitor domains in the ventricular zone. A gradient of Shh signaling is also important to eventually generate several types of motor neurons and interneurons defined by a differential expression of transcription factors and functions (Fig. 4a, b). *In vitro* experiments have indeed confirmed that different concentrations of Shh set strict thresholds to

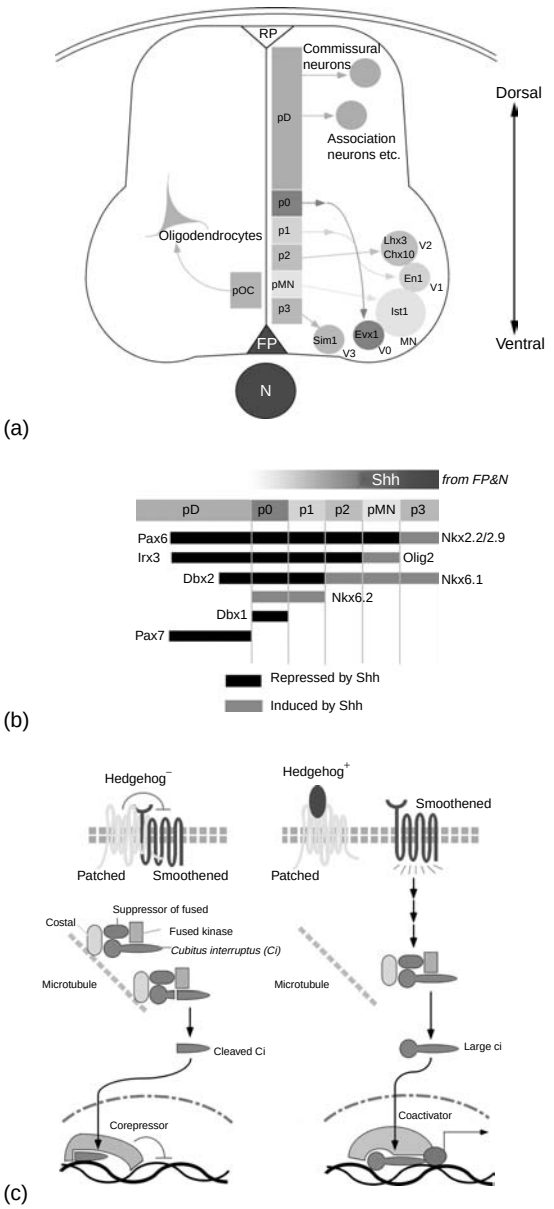

Fig. 4 Shh dependent D–V patterning in the neural tube. (a) Various cell types are generated along the D–V axis of the neural tube. The floor plate cells (FP) and the underlying axial mesoderm, notochord cells (N) produce the secreted molecule sonic hedgehog (Shh). MN, motor neurons; p, progenitor domains; RP, roof plate, (b) Ventral sources of Shh (FP and N) set a Shh gradient, inducing and/or repressing distinct sets of transcription factors within the progenitor domains. This helps generate various types of neurons defined by the functions and gene expression profiles. (c) The hedgehog signaling machinery revealed in *Drosophila*. In vertebrates, there are three hedgehogs (Sonic, Indian, and Desert), and Gli transcription factors are the vertebrate homologs of Ci (*cubitus interruptus*).

induce each cell type in the spinal cord. This raises the question of how a single molecular gradient can serve to define distinct progenitor domains along the D–V axis. The known Shh receptor components are the seven-transmembrane proteins, Patched and Smoothened, and interactions between the ligand Shh and the receptor molecules are shown to finally activate the Gli transcription factors in vertebrates (refer Fig. 4c). As the degree of receptor activation appears to control the

degree of Gli activation, it is assumed that such differential concentrations of the Gli molecule in cell nuclei affect the binding possibility of the target genes, allowing the distinct sets of downstream transcription factors to be switched on/off.

In the dorsal half of the spinal cord, distinct neuronal cell types are generated at characteristic times and positions. The identity and pattern of generation of these dorsal neurons are shown to be dependent initially on BMP-mediated signals that derive from the epidermal ectoderm and induce the dorsal midline cells called the *roof plate*. Roof plate cells are further revealed to provide a secondary source of TGF (transforming growth factor) beta-related signals that are required for the generation of distinct classes of dorsal interneurons. Additionally, WNT-secreted molecules are suggested to play a crucial role in constituting the dorsal cell types. For instance, when the *wnt* genes expressed at the dorsal midline of the neural tube (Wnt1 and Wnt3a) are eliminated by gene targeting in the mouse embryo, expansion of the neural crest as well as dorsal CNS progenitors is abolished. This indicates that local WNT signaling is important for the patterning and generation of various types of cells in the dorsal neural tube. Furthermore, neural crest induction has been shown to involve interactions between the neural plate (neuroectoderm) and epidermal (surface) ectoderm: If an isolated neuroectoderm and neural plate are combined *in vitro*, neural crest cells are induced at the interface between these tissues. Recently, Wnt6 localized in the early chicken surface ectoderm has been demonstrated to play a role in initiating neural crest induction, yet the detailed molecular machinery including the relationship between Wnt6 and BMP signaling has not been determined.

The alar/basal boundary, which coincides with the morphologically identifiable sulcus limitans, has been proposed to play a critical role in maintaining the D–V pattern in the neural plate/tube during development. Cadherin cell adhesion molecules could be responsible for the process, because one of the cadherin subclasses, F-cadherin, is found to delineate the alar/basal boundary cells, whereas other subclasses of cadherins demarcate a distinct cellular population in the frog neural plate/tube. Cadherin is a transmembrane protein and is known to confer rigid adhesiveness to cells in a Ca^{2+} dependent manner (Fig. 5a). More than 20 subclasses have been cloned thus far, and notable is the fact that each subclass has a

Fig. 5 Roles of cadherin cell adhesion molecules in embryogenesis (a) Cadherin is a transmembrane protein that confers rigid adhesiveness to cells in a Ca^{2+} dependent manner. Several molecules are known to interact with cadherin at the cytoplasmic domain, and they are important to tether cadherins to the actin cytoskeltons. Cadherins might be involved in the Wnt signaling pathway, as beta-catenin required for cadherin-mediated cell adhesion, is also a crucial element of Wnt signaling (Fig. 2b). (b) Cells with a particular subclass of cadherin can only adhere to cells with the same subclass of cadherin: when cells with different cadherin subclasses (red or blue) are mixed, they sort out from each other in *in vitro* aggregation assays, although some pairs of cadherins show heterotypic bindings. (c) During chicken neurulation, the expression pattern of cadherins is dynamically regulated. Note that the changes of the cadherin expression patterns are tightly linked to morphogenetic events, and it has been shown that overexpression of cadherins perturbs the processes of neurulation as well as neural crest emigrations. (See color plate p. xxiv).

specific binding property: cells with a particular subclass of cadherin can generally adhere to those with the same subclass of cadherin, although some pairs of cadherins show heterotypic bindings. Owing to this characteristic, if cells with different cadherin subclasses are mixed, they sort out from each other in *in vitro* aggregation assays (Fig. 5b). The restricted expression pattern of F-cadherin at the *sulcus limitans*

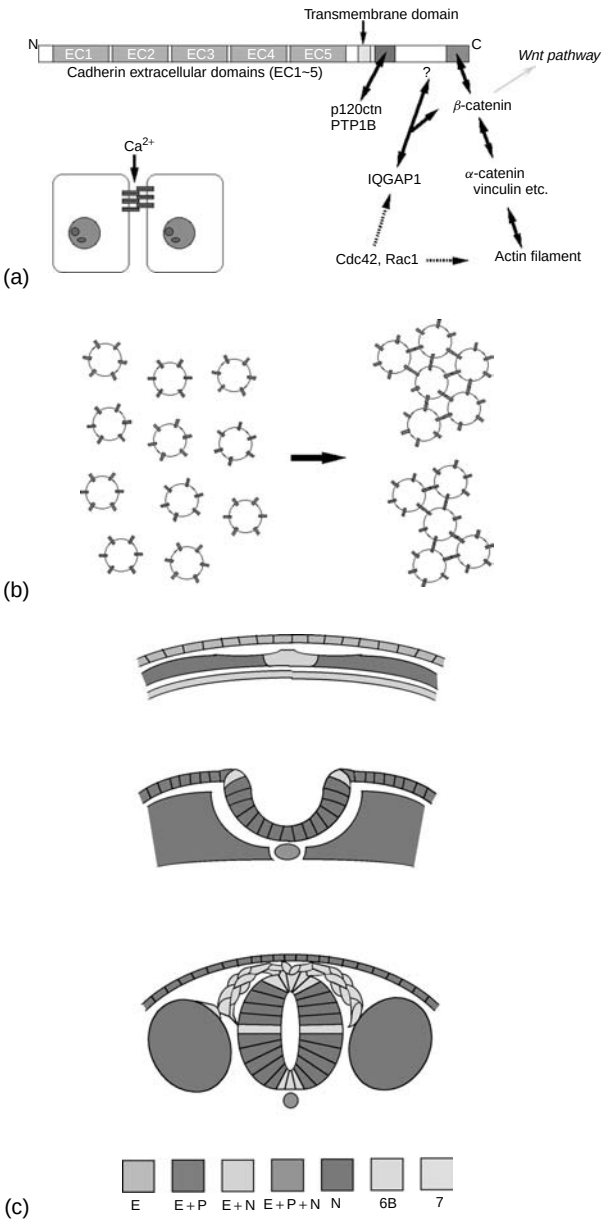

may therefore be required to prevent the intermingling of cells between the alar and basal plates. Overexpression of F-cadherin in the frog neural tube indeed resulted in the abnormal intermixing of cells in the neural tube. Chicken cadherin-7, most homologous to frog F-cadherin, further delineates the sulcus limitans, suggesting a conserved role for cadherins in maintaining this important boundary between the alar and basal plates during evolution (Fig. 5c).

Curiously, in the process of neurulation, the expression of cadherin subclasses is dynamically regulated. In chicken embryos, for instance, the neuroectoderm loses E-cadherin expression and begins to express N-cadherin (Fig. 5c). Cells at the neural ridge then lose N-cadherin expression and begin to express cadherin-6B, while the emigrating neural crest cells switch off cadherin-6B expression and acquire cadherin-7 expression (Fig. 5c). These serial switches of cadherin subclasses during development are crucial, as ectopic expression of N-cadherin in the frog embryo causes abnormal segregation between the ectoderm and neural tube. Overexpression of N-cadherin or cadherin-7 in the chicken neural tube is further revealed to result in the deterioration of neural crest cell emigration, indicating that the spatiotemporal expression pattern of cadherin is important for normal development. The repressive transcription factor, Slug/Snail, is known to be involved in epithelial–mesenchymal transition processes by downregulating E-cadherin expressions *in vitro*. This type of regulation might control emigrations of neural crest cells *in vivo*, as Slug/Snail expression delineates the neural fold during neural crest emigrations. Yet, cadherin genes are huge (~200 kbp) with large introns, and the transcriptional machineries required for the spatiotemporary restricted expression of cadherins has been poorly investigated.

2.2 A–P Patterning in the Neural Plate/Tube

It is known that tissues surrounding the future neuroectoderm at the gastrulation stage help establish A–P patterning in neural tissue. As I already mentioned earlier, Nieuwkoop proposed a two-step model in which an early activating signal has a role in inducing neural tissue with an anterior character (i.e. forebrain). Subsequently, a second posteriorizing or transforming signal converts some of the posterior neural tissues into an even more posterior character (i.e. hindbrain and spinal cord). Supporting this model, distinct tissues have been shown to induce different axial characters. For instance, in mice, it has been demonstrated that the combined action of the AVE and the node-derived axial mesoderm (prechordal plate) was important to represent anterior neural characters (Fig. 1c). Various signaling molecules required to establish the anterior character have indeed been identified. For example, secreted molecules such as Noggin and Chordin have been shown to induce anterior characters as the gene targeting mice have severe defects only in the anterior structures. Another secreted molecule, Cripto with a cysteine rich and EGF-like motif has further been shown to be a crucial morphogen organizing the mouse A–P axis (refer to Fig. 1c and the legend). Regarding the posteriorizing signals, retinoids, Wnt, and FGF signaling have been implicated as the candidates. For instance, explants from the frog anterior neural plate were shown to acquire the posterior identity when treated with FGF. Additional mechanisms must be involved in the posteriorization, as a simple block of

FGF was not enough to abolish the posterior patterning. Interactions among those secreted molecules could generate the regional differences of signaling events in the downstream, eventually turning on differential sets of transcription factors to represent the morphological and functional variations along the A–P axis in the neural plate/tube.

Following the neural induction and/or neurulation events, several bulges and sulci become apparent in an orderly manner in the anterior part of the neural plate/tube, generating the anatomically distinguishable segmented structures along the A–P axis. This step is termed *segmentation* and Fig. 6(a) shows how the vertebrate anterior neural plate/tube can be subdivided into major brain regions such as the *forebrain, midbrain*, and *hindbrain*, all of which have separable characteristics compared with the posterior CNS region, namely, the *spinal cord*. Within these major brain regions, further bulges and *sulci* are generated, and these minimal brain units, anatomically distinguishable during development, are defined as *neuromeres* (Fig. 6c, 7c). Curiously, some of the neuromeres, which constitute the cell lineage restricted compartment units, are considered to play a major role in maintaining embryonic brain organization, because the boundaries prevent cells, once patterned, from random intermingling during development. Neuromeres may further provide a basic framework for complex neuronal circuits since the neuromere boundaries often coincide with the territories between functional brain subdivisions and with initial axonal tracts in the embryonic brain. Next, I summarize genetic mechanisms involved in A–P patterning within the (1) hindbrain (rhombencephalon), (2) midbrain (mesencephalon), and (3) forebrain (prosencephalon).

2.2.1 Hindbrain Specification

The hindbrain region later generates the pons, cerebellum, and medulla. While the cerebellum is thought to play a central role in the coordination of complex movements, the medulla is considered as the critical center for involuntary activities such as respiration. The pons is also an important turning point for various nerve tracts. In the developing vertebrate hindbrain, seven to eight segmental units are morphologically recognizable. These units are called *rhombomeres* and play a crucial role in hindbrain patterning. For instance, distinct types of cranial motor neurons are differentiated at later stages along rhombomere organization (Fig. 7c). The neural crest migration pattern is also determined by this segmental organization, eventually affecting the metameric craniofacial structures represented in the branchial arches (Fig. 7c). What molecular mechanisms then regulate hindbrain segmentation?

The Hox transcription factors are one of the gene families that play a central role in hindbrain segmentation. *Hox* genes are originally identified as homologs of the genes involved in the segmentation of *Drosophila* embryos: The A–P axis of *Drosophila* embryos is first realized as opposing maternal mRNA gradients along the A–P axis of the egg, which are translated into transcription factors immediately after fertilization (Fig. 7a). In response to the combinatorial concentration of these transcription factors, differential sets of genes termed gap genes are then turned on along the A–P axis (Fig. 7a). Downstream of these gap gene transcription factors, pair-rule genes are activated (Fig. 7a), and finally, the segment polarity genes and the homeotic selector genes are respectively expressed to determine the polarity and identity of

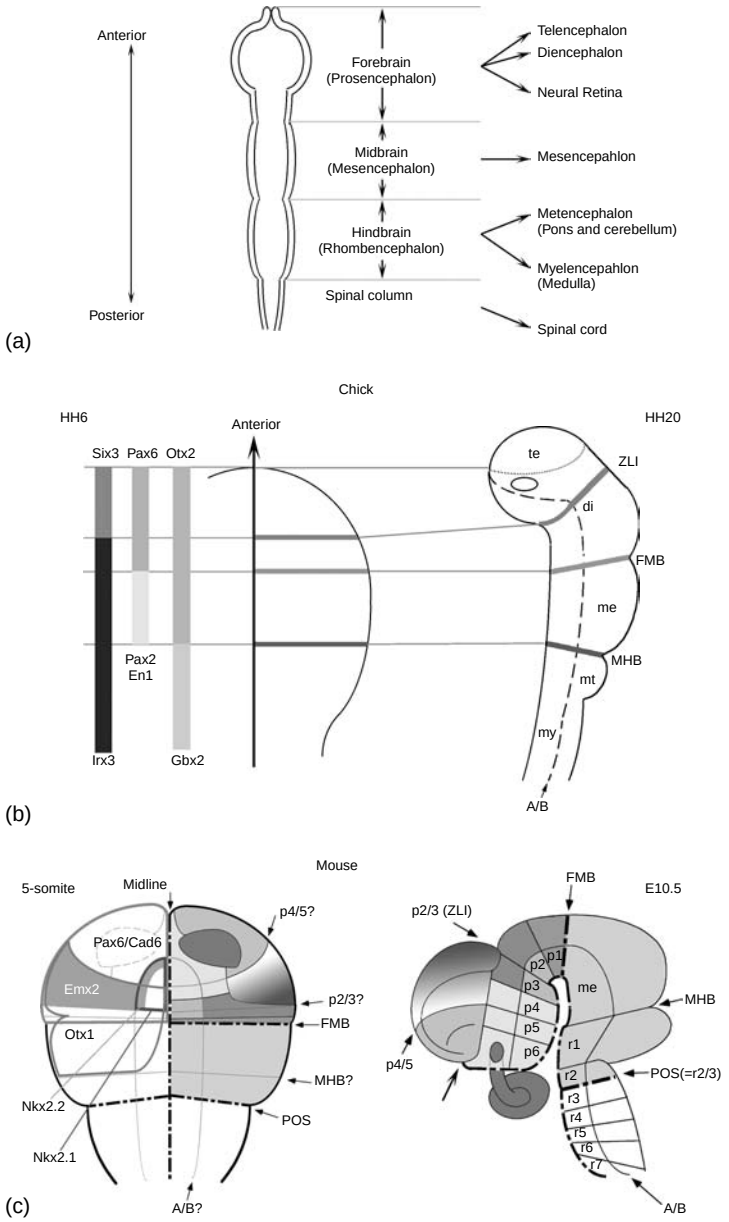

each segmental unit (Fig. 7a). Interestingly, the homeotic selector genes, containing a unique DNA-binding motif with 60 amino acids called "*homeodomain*," make a cluster in *Drosophila* chromosome 3 such that their arrangement is colinear to the expression domain along the A–P axis of the organism (Fig. 7b). Lewis proposed a model in which the combinatorial expression of this class of genes

plays an essential role in conferring regional identities. Supporting the model, many anterior homeotic transformations are found in *Drosophila* embryos lacking these homeodomain-containing genes.

Intriguing is the fact that clusters similar to the *Drosophila* homeodomain-containing gene complex (*antenapedia*/bithorax complex) are found in the genome of various animals including humans, and they are named *Hox* gene clusters as the genome sequences encoding the homeodomain are called *homeobox* (Fig. 7b). Notably, the anterior boundaries of *Hox* gene expressions coincide with the rhombomere boundaries, implicating them in the important role of hindbrain segmentation (Fig. 7c). The functional analyses by gene targeting in mice (see Sect. 7 for this methodology) indeed indicate that regional identities in the hindbrain are controlled by the differential expression of these *Hox* genes along the A–P axis with gene knockout phenotypes just being similar to the *Drosophila* homeotic mutant phenotypes.

Hox genes have been revealed to be regulated by early posteriorizing signals such as Wnt, FGF, and retinoic acid (RA), all of which are secreted by surrounding tissues including somites, highlighting the pivotal role of *Hox* genes in linking the molecular events that occur between neural induction and hindbrain segmentation. For example, Krumlauf and colleagues have found several RA responsible elements in the *Hox* gene regulatory regions, where RA and its receptor complex can directly interact to control the *Hox* expression, and RA treatment has been shown to affect the Hox expression *in vivo*. Itasaki et al. have further shown that anteriorly grafted somites have an activity that shifts the expression boundary of Hox anteriorly. This posteriorizing activity of the somites can be separated with a simple treatment by RA, indicating that several factors could cooperatively regulate the Hox gene expression in the hindbrain.

One of the critical events that occur during hindbrain segmentation is compartmentalization. If a cell is labeled by vital dyes (DiI) and traced for a few days, sibling cell migration and/or proliferation is restricted within each rhombomere

Fig. 6 A–P patterning of the neural tube/plate. (a) Anterior part of the neural/plate becomes segmented during development, generating brain regions with distinct structure and functions, (b) Early patterning of the chicken neural plate by the differential expression of transcription factors. Gene expression boundaries eventually form the anatomically distinguishable territories in the brain. A/B, alar/basal boundary; di, diencephalon; FMB, forebrain/midbrain boundary; HH, Hamburger Hamilton stages; me, mesencephalon; MHB, midbrain/hindbrain boundary; mt, metencephalon; my, myelencephalon; te, telencephalon; ZLI, zona limitans interthalamica. (c) A fate-mapping result of the mouse neural plate. The area painted by a given color on the right side of neural plate at the 5-somite stage (left) produces the brain region of the same color at E10.5 (right). Faint lines indicate prospective boundaries of brain regions. Cell lineage restricted compartment boundaries are indicated by bold dashed lines. Note that several gene expression boundaries demarcated by different colors on the left side of the neural plate coincide well with prospective boundaries of brain regions: blue line; Pax6/cadherin-6 expression boundary, red line; Otx1 expression boundary, orange line; Nkx2.2 expression boundary, purple line; Nkx2.1 expression boundary, and green area; Emx2 mRNA positive region in the neural plate. A/B; alar/basal boundary, FMB; forebrain/midbrain boundary, me; mesencephalon, MHB; midbrain/hindbrain boundary, p; prosomeres, POS; preotic sulcus, r; rhombomeres, ZLI; zona limitans interthalamica. An arrow indicates the anterior most part of the brain at the embryonic day (E) 10.5. (See color plate p. xxvi).

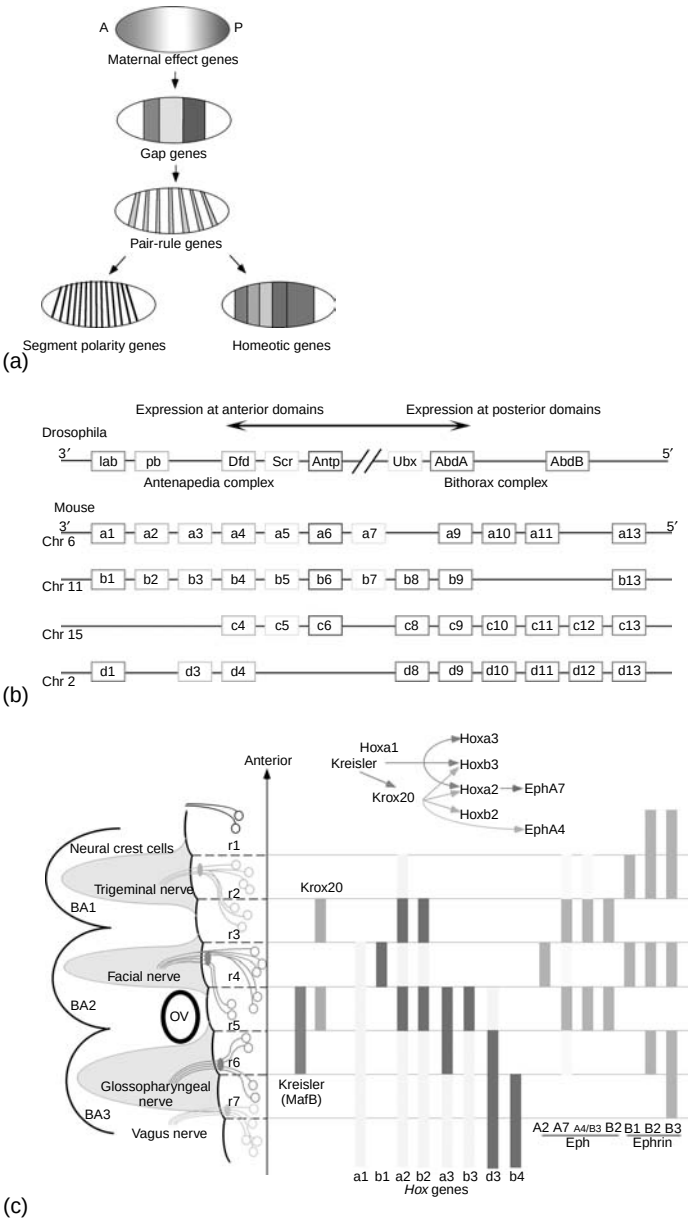

unit in the chicken embryo. Then, what molecular machinery is involved in the hindbrain compartmentalization? First of all, differential affinities among compartmental units have been suggested. The expression of several transcription factors such as *Hox* genes, Krox20, and Krisler/MafB are known to delineate rhombomere units (Fig. 7c). Downstream of these transcription factors, several cell

surface molecules are implicated in restricting cell intermingling, as dissociated cells from even-numbered rhombomeres and those from odd-numbered rhombomeres segregate from each other if mixed in suspensions. Eph receptor tyrosine kinases and their ligands, ephrins, are good candidates in conferring such differential affinity to cells: Even-numbered rhombomeres express ephrins, while odd-numbered rhombomeres express Ephs (Fig. 7c). This complementary pattern of Eph/ephrin expression is important because bidirectional signaling between cells with receptors and those with ligands leads to mutual repulsion at the interface. As a result, cell intermixing between rhombomeres is strictly regulated and rhombomere boundaries are maintained.

The cadherin cell adhesion molecules are another set of candidates involved in cell lineage restrictions between rhombomeres. As I mentioned in Sect. 2.1, cadherins confer subclass specific adhesiveness to cells. Within the mouse hindbrain, cadherin-6 is transiently expressed in rhombomere-6 under regulation of the transcription factor Hoxa-1. Additionally, R-cadherin expression has been reported in odd-numbered rhombomeres. The sorting activity between cells from even-numbered rhombomeres and those from odd-numbered rhombomeres has been shown to be Ca^{2+} dependent. Cadherins might thus serve a role in restricting cell movement during mouse hindbrain development. Cadherins are linked to the actin filaments via catenins and this provides rigid forces for cells to bind each other with polarity (Fig. 5a). Eph-ephrin signaling activates Rho GTPases and could collapse actin cytoskeletal elements to alter cell movements. Cadherins and eph-ephrin signaling may therefore act cooperatively to regulate the status of actin cytoskeltons at the rhombomere boundary region.

2.2.2 Midbrain Specification

Although no segmental organization has been identified along the A–P axis in the midbrain during development, this region contains many types of neurons with critical functions. For example, the ventral compartment produces dopaminergic neurons of the substantia nigra to control the motor functions, while the tectum (or the superior colliculus) region forms the precise positioning of neurons along the A–P axis to process visual information. Because no *Hox* genes are expressed in the brain anterior to rhombomere2, additional mechanisms are required to pattern this region.

At the early neural plate stages of chicken embryos, the expression borders of several transcription factors are known to demarcate the future midbrain territory. For instance, the anterior limit of the future midbrain (forebrain/midbrain boundary) has been shown to correspond to the

Fig. 7 A–P patterning of the developing hindbrain. (a) The genetic cascades involved in *Drosophila* A–P patterning, (b) Genomic organization and expression profiles of mouse *Hox* gene clusters are reminiscent of those of *Drosophila* homeotic genes in *antenapedia* and the bithorax complex, (c) The metameric units in the hindbrain (r, rhombomeres) delineated by the differential expression of transcription factors and Eph/ephrins are important to produce distinct sets of neurons along the A–P axis. Dark and light shading represent high and low levels of gene expression in individual rhombomeres respectively. Note that the migration pattern of neural crest cells as well as the formation of the branchial arches (BA) are also affected by the hindbrain organization.

posterior expression border of transcription factor Pax6 (Fig. 6b). The anterior expression border of another paired domain containing transcription factor Pax2 as well as a homeodomain-containing transcription factor Engrailed1 (En1) complementarily delineates the Pax6 negative domain in the neural plate (Fig. 6b). It has also been reported that the posterior expression border of the Otx2 transcription factor as well as the anterior border of Gbx2 coincides with the posterior limit of the future midbrain (Fig. 6b). Repressive interactions among such transcription factors have been suggested as being required for the establishment and maintenance of the mutually exclusive expression of these transcription factors at the future forebrain/midbrain boundary (FMB) and midbrain/hindbrain boundary (MHB). Ectopic Gbx2 expression in the midbrain using the *in vivo* electroporation (EP) method (see Sect. 7), for instance, downregulates Otx expression only within the cells expressing ectopic Gbx2. While the FMB is a compartment boundary at the mouse neural plate stages, the midbrain and hindbrain cells are able to freely intermix. Hence, the mutually repressive interactions between Otx2 and Gbx2 are very important to precisely define the MHB during early development.

The Otx2/Gbx2 expression border then generates the anatomically identifiable boundary between the midbrain and hindbrain called the *isthmus*, which harbors the local organizing activity for midbrain A–P patterning. For instance, if this region is grafted into the forebrain region, the entire structure of the posterior midbrain would be induced with its mirror image in the anterior midbrain and the topological innervations of the retinal neurons would be bifurcated (Fig. 8a). Wnt1 and Fgf8 secreted proteins appear to be the candidate signals for midbrain patterning. Wnt1 mutant mice lack the midbrain, and Fgf8 expressing cells or beads can mimic organizer activity (Fig. 8a). In addition, Wnt1 and Fgf8 signaling pathways form a regulatory loop that controls homeobox-containing transcription factors, such as En1/2, to establish and maintain the MHB organizer as well as the entire midbrain structure (Fig. 8a).

2.2.3 Forebrain Specification

The forebrain (prosencephalon), which is the most anterior portion of the CNS, gives rise to a variety of tissues during development including the eyes, telencephalon, and diencephalon. Within this area, mammals have acquired a six-layered architecture of cells named the neocortex, and the region has evolved, especially in human beings, for total cognitive association as well as linguistic communication. In the developing forebrain, segmental structures specifically termed *prosomeres* are known to appear (Fig. 6c). In chicken and mouse embryos, at least six units (p1–p6) are anatomically recognizable, and using the dye-labeling method, some of their boundaries have been shown to stand as limits of cell lineage restrictions just like the rhombomeres in the hindbrain. Several axon tracts seem to coincide with the prosomere boundaries, suggesting that prosomeres might also provide the basic framework for forebrain structures. Then, how are such prosomeres formed and maintained during development?

In the early chicken neural plate stage, mutual repression between the transcription factors, Irx3 and Six3, was found to first set a boundary within the future forebrain, which later coincided with the p2/3 boundary specifically called the *zona limitans interthalamica* (ZLI) (Fig. 6b). A recent study using the organ culture system has

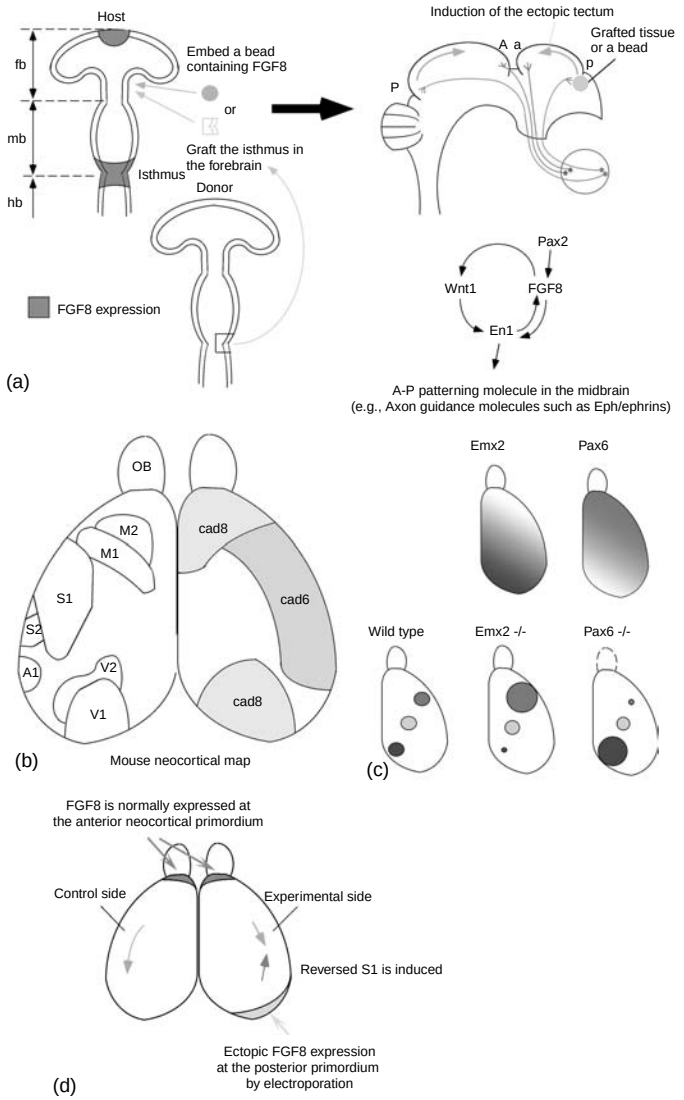

Fig. 8 A–P patterning of the midbrain and mammalian neocortex
(a) The role of the isthmic organizer in midbrain patterning. Fgf8 is a component of the organizer. See details in the text. fb, forebrain; hb, hindbrain; mb, midbrain, (b) Dorsal view of the mouse postnatal neocortex. The mouse neocortex can be subdivided into several functional areas, which are delineated by differential expression of type II cadherins. A, auditory cortex; cad, cadherin; M, motor cortex; OB, olfactory bulb; S, somatosensory cortex; V, visual cortex,
(c) Countergradient of Pax6 and Emx2 in the neocortical primordium is important to bring out differences among neocortical areas later in the development, (d) Fgf8 expressed at the anterior neocortical primordium is a crucial organizer to pattern the mouse neocortical areas.

shown that the mouse neural plate at the 2- to 3-somite stage already possesses different competencies for gene expression between the anterior and posterior regions adjoining the Six3/Irx3 gene expression boundary. This could depend partly on the fact that the underlying axial mesoderm is subdivided into the prechordal plate and notochord just beneath the future ZLI at this stage. In mouse embryos, it had been difficult to confirm if similar units exist because the embryonic development proceeds *in utero*. I performed a cell labeling experiment at the early neural plate stages using the whole embryo culture system that allows recapitulating the development *in vitro*. According to the results, there are cell populations around the 3-somite stage destined to be only forebrain cells, and the posterior limit of such cells appears to fall on the future ZLI. At the ZLI, in later developmental stages, signaling molecules such as Shh are expressed, implicating them in an organizing role in patterning the prosomeres.

Immediately after the differential competency between the region anterior to the future ZLI and that posterior to the future ZLI, the FMB is set as the Pax6 gene expression boundary, which can also be defined as a compartment boundary at the mouse 5-somite stage (Fig. 6c). In the chicken system, the FMB is revealed to be molecularly established by the 11-somite stage as the result of the mutually repulsive interactions between Pax6 and Pax2/En1 (Fig. 6b).

It is still unclear when and how the other prosomere boundaries are established during early development. In the later stages of mouse development, however, the boundary between the future cerebral cortex (ctx) and lateral ganglionic eminence (lge) has been shown to be a compartment boundary by E10.5. This boundary is called the *zona limitans intertelencephalica* and thought to correspond to the p4/5 limit. According to the fate-mapping analyses in the mouse neural plate, the future lge was indeed demonstrated to be located more anteriorly to the future ctx.

For over several decades, it has been a big problem to determine which region is the anterior most and how the floor/roof plate as well as the alar/basal boundary ends in the forebrain. This is simply because forebrain development could only be traced by morphological landmarks. However, recent gene expression analyses provide a variety of landmarks in the brain, and such molecular markers strengthen the prosomeric model. In the prosomeric model, the roof, alar, and basal plates concentrically cross the midline of the anterior most neural plate, disposing the eyes and telencephalon within the alar plate (Fig. 6b). Recent fate-mapping analyses in chicken and mouse well support the model (Fig. 6c).

Then, what molecular mechanism are involved in maintaining the prosomere organization at later stages of forebrain development? Several transcription factors are again implicated as their expression delineates later prosomeres and/or the boundaries. For example, *Pax6, Gsh2*, and *Dlx* genes demarcate restricted brain regions along the D–V axis of the mouse E12.5 telencephalon, which almost corresponds to the A–P axis of the neural plate. Interestingly, it has been demonstrated that cells from distinct regions in the telencephalon segregate from each other in a Ca^{2+} dependent manner if the cells are dissociated and mixed together. A cell surface carbohydrate epitope, CD15 (Lewis(x)), is expressed within the cortex and suggested to be the candidate molecule that mediates Ca^{2+} dependent segregation *in vivo*. In addition, several cadherins

are known to delineate prosomeres in the mouse brain, and dissociated cells have been demonstrated to sort each other out *in vitro*, dependent on the cadherin subclass expressed. In the E11.5 mouse telencephalon, cadherin-6 demarcates the lge (p5), while R-cadherin complementary delineates the ctx (p4). As the overexpression of cadherins at the ctx/lge boundary using the *in vivo* EP method (see Sect. 7 for this methodology) results in the sorting of cells with ectopic cadherin into the territory expressing the same cadherin at the boundary, this indicates that the differential expression of cadherins plays a role in maintaining the compartment boundary. Notably, R-cadherin expression in the ctx has been shown to be downstream of Pax6. Transcription factors might therefore regulate cell adhesion molecules in order to maintain their expression pattern by sorting mechanisms in the telencephalon.

Most recently, an important genetic cascade that patterns the rostro-caudal axis within the mouse cerebral cortex at later stages of development has been identified. As summarized in Fig. 8(c), the expression level of transcription factors Pax6 and Emx2 was found to make the countergradients along the mediolateral axis in the neocortex primordium at E12.5, and if one of the gradients was deprived by means of genetic mutation, regional identities were revealed to be shifted accordingly. Upstream of Pax6/Emx2 countergradients, Fgf8 expressed at the anterior neural ridge has been suggested as playing a pivotal role. Indeed, overexpression of Fgf8 in the posterior cerebral neocortex caused the duplication of the cortical area with its mirror image, indicating the organizer activity of FGF8 in the cerebral cortex primordium (Fig. 8d). As the mammalian neocortex has been thought to evolve via multiple duplication of identities, local changes in organizer activities such as ectopic sources of FGF8 expression might actually contribute to emerge the diversity of brain regions.

3 Cellular Aspects to Generate Distinct Types of Cells and/or Cellular Organization (e.g. Nuclei, Layers, and so on) in Patterned Tissue

While the neural tissue at earlier stages has just a single-cell layer organization with the epithelial character, differentiated tissues contain multiple types of cells with complex cellular organizations. There are three major classes of cells in the developing and adult CNS: *nerve, glial,* and *ependymal cells*.

Nerve cells, also called *neurons,* play a central role in constituting functional neuronal networks of the nervous system. Most nerve cells have large cell bodies called *soma*, which are accompanied by relatively large cell nuclei as well as multiple processes called either *axons* or *dendrites*: the former are defined by their ability to gather the action potentials and send electric signals to other nerve cells, and the latter are the sites for receiving these electric signals from axons. One nerve cell normally has one axon and multiple dendrites. In an *in vitro* situation, the processes of nerve cells are also called neurites, and the longest neurite has been shown to acquire axon characters. In the CNS, axon terminals of nerve cells make connections with other nerve cells and the contact sites between nerve cells are called *synapses*. Nerve cells also innervate muscles and the contact site is specifically called the *neuromuscular junction* (NMJ).

Nerve cells are considered to be all postmitotic and never generate siblings. In the process of nerve cell production, progenitor cells are known to get into the elevator movement, in which the cell body of the progenitor cells migrates from the ventricular to the marginal zone in a cell cycle dependent manner. During the movement, the asymmetrical divisions appear to produce neuroblasts and eventually generate various types of nerve cells.

Nerve cells can be distributed in different patterns in the CNS as the postmitotic nerve cells usually migrate toward their destinations after final cell division. For example, in the spinal cord, motor neurons born in the ventricular zone migrate radially and make clusters called motor columns in the ventral horn. Within the spinal motor column, neurons are further segregated into groups by means of their birthdays and targets: While the earlier generated (i.e. older) motor neurons constitute the lateral group in the column and innervate the limb muscles, the younger ones home in on the medial column and innervate the axial muscles. In the mammalian cerebral neocortex, nerve cells born in the ventricular zone travel radially as well as tangentially and settle in the mantle zone depending on their birthdays, forming the characteristic layer organization. The migration pattern of neurons born in the subventricular zone of the telencephalon is another interesting example, in which the postmitotic neurons adhere to each other to make a chain migratory pathway into the olfactory bulb. In the mature brain, neurons with distinct functions often gather and are anatomically distinguishable in the gray matter as the group of cells called the *nucleus*. At the dorsal surface of the CNS, neurons with similar functions tend to be assembled in the laminated structure called the *cortex*. The neurons with same functions may further exist in a sparse manner, and such a distribution pattern is termed the *reticular formation*, which plays a crucial role in modulating both sensory and motor inputs.

Glial cells, or simply *glia*, are supporting cells in the nervous system, and there are normally between 10 and 50 times more glial cells than neurons. Various types of glial cells with distinct functions are generated in both the CNS and PNS. For example, the oligodendrocyte attaches to the nerve axon and surrounds it to make the myelin sheath, enhancing the speed of electrical signal transmission along the axon. The Schwann cells in the PNS have a role similar to the oligodendrocyte. The astrocyte is thought to relay nutrients from blood vessels to neurons, and the microglia plays a role in maintaining and protecting the cellular organization of the CNS.

Finally, the ependymal cells are a unique population in the CNS with stem cell characteristics. They are homed in the ventricular zone and are shown to produce nerve cells as well as glial cells (astrocytes). The ependymal cells are sometimes called *radial glial cells* as they extend long radial fibers from the ventricular to the pial surface, which are thought to support the radial migration of nerve cells during development. In the adult brain, the ependymal cells continue to produce siblings and this might help generate novel neuronal circuits during the process of learning and/or memory formation. Several researchers have now begun to apply this unique character of the ependymal cells to regenerate a variety of cells, including nerve cells in the CNS.

4 Molecular Aspects of Generating Distinct Types of Cells and/or Cellular Organization (e.g. Nuclei, Layers, and so on) in Patterned Tissue

4.1 Generation of Neurons

There are mainly two molecular mechanisms in generating a variety of neurons. One is the cell intrinsic mechanism (cell autonomous) and the other is the cell nonautonomous mechanism. These two mechanisms might closely correlate with each other. In the vertebrate nervous system, the latter mechanism plays major roles, yet the former mechanism is indispensable. In Sects. 4.1.1 and 4.1.2, I summarize those molecular mechanisms involved in the production of multiple cell types in the nervous system.

4.1.1 Cell Intrinsic Mechanisms

Cell intrinsic mechanisms play important roles especially in animals where cell lineage is strictly regulated during development. Cell intrinsic differences could be mainly realized by asymmetric cell divisions in which particular sets of determinants might be differentially divided into two daughter cells. For example, in the process of *Drosophila* neurogenesis, *prospero* mRNAs as well as the protein products are distributed asymmetrically in the progenitor cells (Fig. 9a). This asymmetric distribution of Prospero is crucial in producing distinct sets of neurons, because *prospero* encodes a transcription factor and the molecule is transported into the nucleus after asymmetric cell division, allowing distinct sets of genes to be switched on/off in each daughter cell (Fig. 9a). In order to localize *prospero* mRNAs as well as proteins asymmetrically in the progenitor cell, the RNA-binding protein, Staufen, and a Staufen binding protein, Miranda, which can also bind to the Prospero protein, appear to play a central role and Inscutable regulates the polarization of these proteins (Fig. 9a). Inscutable protein can also control the spindle orientation in the cell, suggesting that this protein is a main determinant of cell intrinsic differences between apical and basal compartments in a cell. Curiously, Inscutable can further interact with the Bazooka protein, another important determinant of apical–basal polarity formation in a cell. Bazooka homologs have been identified in the nematode *Caenorhabditis elegans* (Par-3) as well as in vertebrates (ASIP; atypical PKC isotype specific interacting protein), indicating a conserved mechanism for intracellular asymmetry beyond species. Notable is the fact that the *Drosophila* glial/neuronal lineage can also be initiated by the asymmetric distribution of *glial cell missing* (*gcm*) mRNA (Fig. 9a). As *gcm* encodes a transcription factor required for glial fates, the asymmetric segregation between the two daughter cells may play an essential role in the bifurcation of glioblast and neuroblast lineages (Fig. 9a).

4.1.2 Cell Nonautonomous Mechanisms

It would be very important to control the relative ratio of various cell types during neural development. In this critical process, Notch/Delta mediated cell–cell communications (cell nonautonomous machinery) appear to play a pivotal role. The Notch/Delta signaling pathway was first reported to be essential for the neurogenesis of the *Drosophila* embryo, where cells that happen to express a higher amount of ligands, Delta, inhibit lateral cells expressing receptors, Notch, to further differentiate into neurons. This process is specifically

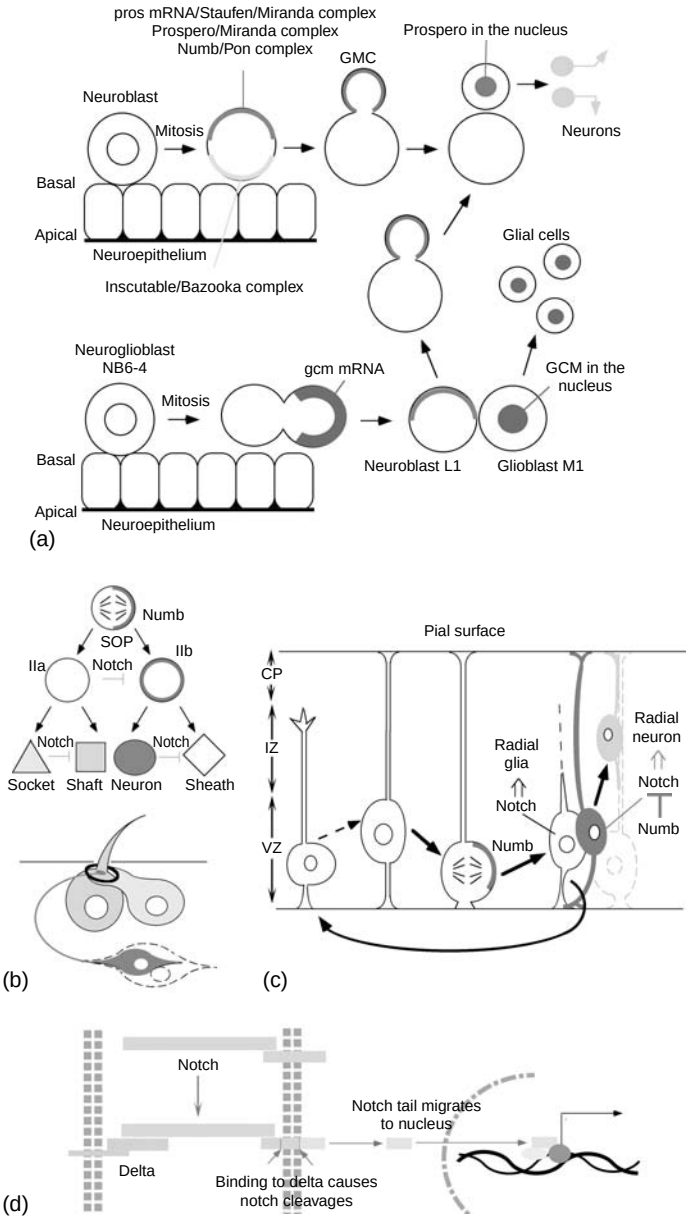

called *lateral inhibition*. In the process of Notch signaling, the cytoplasmic region is cleaved and transported into the cell nuclei to switch on the downstream target genes (Fig. 9d). Signaling may be modified by Numb (an intrinsic determinant localized at the cytoplasmic side of cell membrane and/or Fringe (a member of glycosyltransferase). Although the initiation of this process is stochastic, a

proper number of neurons can be generated in the neuroectoderm, otherwise all cells become neurons (the so-called *neurogenic phenotype*) after neural induction. At later stages of *Drosophila* neural development, Notch/Delta signaling is utilized to control the cell type specification in the series of cell divisions. For example, after the sensory precursor cells are selected from the epithelium, each of them should produce four distinct cell types: a neuron, glia (sheath), socket, and shaft (Fig. 9b). As shown in Fig. 9(b), Notch signaling, together with Numb, seems to suppress the glia as well as shaft cells to become neuron and shaft cells. In vertebrate nervous systems, the Notch/Delta signaling pathway appears to be involved in determining primary neuron production in a regulated way; prospective primary neurons expressing the ligand Delta inhibit lateral cells expressing the receptor Notch to further differentiate into neurons, resulting in the stripe domains of neuron production along the D–V axis of the neural plate/tube. In the mammalian cerebral cortex, similar mechanisms help generate cortical granule cells (Fig. 9c). Using time lapse imaging, it has been demonstrated that a radial glial cell (ependymal cell) in the cortical ventricular zone divides to produce one neuron and a cell maintaining the radial glial cell characters (Fig. 9c). In this process, the cells that get into the neuronal lineage have been shown to inherit the Numb protein from the radial glial cells/stem cells via asymmetric cell divisions, repressing the receptor Notch activity to finally be determined as neuronal (Fig. 9c). PNS gangliogenesis is also shown to be dependent on this signaling mechanism, controlling the neuron/glia ratio in the dorsal root ganglia.

As I described earlier, early patterning events partition the vertebrate neural plate/tube along the A–P and D–V axes via a series of complex actions of secreted molecules such as BMP, Shh, WNT, FGF, RA, and so on. This directly helps in generating various sets of neurons, as a distinct address in the neural plate/tube *per se* restricts the definitive number of transcription factors able to be turned on. Besides the Shh dependent neuronal generation machinery along the D–V axis of the spinal cord, which I have already described in detail earlier (see Sect. 2.1; Fig. 4), *Hox* genes have been shown to play a role in generating distinct sets of neurons along the A–P axis. For instance, it has been demonstrated that inactivation of members of the *Hox-c* and *Hox-d* gene clusters expressed in the spinal motor neurons leads to alterations in the motor innervation of specific muscles in the limb. Additionally, in the developing

Fig. 9 Mechanisms involved in neurogenesis during development (a) Cell intrinsic machineries play crucial roles in *Drosophila* neurogenesis as well as gliogenesis. GMC, ganglion mother cell; Pon, partner of numb; pros, Prospero. (b) Roles of Numb and Notch signaling in the *Drosophila* sensory organ precursors (SOPs). (c) Neurogenesis in the mouse cerebral cortex. A radial glial cell (a stem cell) divided asymmetrically within the ventricular zone (VZ) produce a radial glial cell (Numb negative) and a radial neuron (Numb positive). The latter will migrate radially along the long process of radial glial cells, differentiate into a neuron in the intermediate zone (IZ), and finally settle within the cortical plate (CP). (d) Notch/Delta signaling machineries. Notch with 36 EGF-like repeats can bind to Delta with 9 EGF-like repeats expressed by a neighboring cell, resulting in the release of Notch cytoplasmic domain into the nucleus. The cytoplasmic tail can directly bind to DNA and transactivate target genes.

hindbrain, *Hox-a* and *Hox-b* genes have been shown to regulate the identity as well as the axonal projection pattern of cranial motor neurons. Recently, a possible role has been suggested for the secreted molecules FGFs, Gdf11, and Retinoids in regulating the *Hox-c* expression in the motor neurons along the A–P axis. It would be an intriguing subject in the near future to see how the signals that regulate A–P differences can interact with those that define the D–V axis (Shh) and/or neuron/glia ratio (Notch/Delta) to generate distinct types of neurons in a timely and spatially restricted manner.

4.2 Nuclei Formation

To arrange cells into groups after cell divisions, coordinated processes must be required. First of all, the timing of cell division has a role in sorting the final localization of neurons. For example, in the spinal cord, motor neurons generated at earlier stages are shown to occupy the distal domain. Recently, Jessell and colleagues have reported that cadherin cell adhesion molecules demarcate distinct populations of neurons in the chicken spinal motor column. Using the *in vivo* EP technique (see Sect. 7), they showed that neurons with the ectopic cadherin subclass could localize in the subdivisions of the motor column (pools) expressing the ectopic cadherin, implicating the cadherins in the important role of sorting different subclass of neurons into a pool within the spinal motor column during development.

In the thalamus, there are many distinct subdivisions in the neurons, which innervate specific telencephalic region. For example, neurons in the medial geniculate nucleus send axons to just the auditory cortex, while those in the lateral geniculate nucleus innervate only the visual cortex. As is the case with the spinal motor neurons, neurons produced at earlier stages usually settle in the lateral subdivisions within a thalamic nucleus. Recently, various types of molecular markers such as transcription factors (LIM homeodomain transcription factors, Gbx2, Dlx1, Nkx2.2, Pax6, and so on) and cell adhesion molecule cadherins were found to demarcate specific subdivisions. As cadherins have subclass specific cell adhesiveness, differential expression of cadherins may contribute to the segregation of specific cells into the thalamus subdivisions, just like the case with the motor column segregations in the spinal cord. Curiously, it has been reported that secreted molecules such as the Shh expressed at the ZLI might regulate the differential expression of transcription factors in the thalamus, implicating a genetic cascade in the regulation of nucleus formation, yet the detailed machineries have been poorly studied.

4.3 Layer Formation

In the mammalian cerebral neocortex there are six major layers (layers I to VI), which can be distinguished by differences in the morphology and density of the neurons, with the thickness of each layer varying from area to area. Neurons in each cortical layer are generated in the ventricular zone of the dorsal telencephalon and travel toward the cortical plate located in a remote position from their birthplace using the radial glial fibers as the migrating substrates, while some interneurons are known to be generated in the ventricular zone of the ventral telencephalon and migrate tangentially. Cortical layers are formed in an orderly manner from the

inside to the outside, being the very reverse of tree ring formation, in that, newly generated neurons stop their migration at the marginal zone so as to get over the predominated old neurons in the cortical plate. A series of mutant mice exist that have apparent phenotypes in layer formation; for example, the *reeler* mutation is known to affect the order of layers in that newly generated neurons cannot get over the old ones in the cortical layer and so stay inside the prepopulated cell layer, similar to the formation of a tree ring, and ultimately resulting in the reversed order of the cortical layers. The gene responsible for this phenotype encodes a large extracellular protein named *Reelin*, which is secreted by the Cajal–Retzius (CR) cells in the preplate. CR cells are thought to originate from the telencephalic ventricular zone and are the first to reach the cerebral marginal zone to form the preplate. The role of CR cells in layer formation is significant because the secreted Reelin protein appears to promote neuronal cell migration from the ventricular to the marginal zone in order to get over the prepopulated layer of cells. The Reelin receptor is thought to be a molecular complex. A component is the LDL receptor family, namely, the very-low-density lipoprotein receptor (VLDLR) and the apolipoprotein E receptor type 2 (ApoER2) expressed by the migrating cortical neurons. The cytoplasmic region of these receptors can interact with an adapter protein Disabled (Dab-1), and Reelin binding to the receptor complex has been shown to induce tyrosine phosphorylation of Dab-1. Double knockout mice for both receptor genes as well as Dab-1 mutant mice (*Scrambler* or *Yotari* mice) resulted in the reeler phenotype, indicating that these genes constitute the main complex to regulate the neuronal migration and form the normal cortical layer organization in a Reelin dependent manner. Importantly, alpha3beta1 integrin, expressed by migrating neurons was also demonstrated to be able to bind to Reelin and regulate the interaction between radial glial fibers and migrating neurons *in vitro* as well as *in vivo*. Gene knockout of this integrin resulted in reduction of the Dab-1 protein level, suggesting that integrin mediated cell adhesion might play a role in the radial glial fiber guided neuronal migration. The Reelin-Dab1 regulated cell migration machinery has further been shown to be required for the layer formation in other tissues including the cerebellum and hippocampus. In reeler mutant mice, for instance, the Purkinje cell layer was totally disorganized, and the cell arrangement in the dentate gyrus of the hippocampus was also perturbed, although cell type differentiation *per se* was not affected. This implicates Reelin-Dab1 regulated cell migration in the common role of layer formation during development and evolution.

5
Cellular Aspects to Wire up the Nerve Cells into Functional Circuits

Once the nervous system is patterned and unique sets of neurons are generated, they must be connected to each other to constitute functional neuronal circuits. In this process, neurons should precisely send axons into remote target positions. Although Ramon y Cajal had already suggested possible mechanisms explaining axon guidance in the early 1880s, experimental supports had been relatively poor for several decades after him. In the 1950s, Sperry rotated an eye in an adult newt by 180°, and made the eye–tectum connections regenerate. He then demonstrated that the axons

from the retinal neurons at the ectopic position could finally reach the original position, indicating that the axon pathfinding events are totally dependent on various environmental cues (*Chemoaffinity hypothesis*). Landmesser and colleagues also found that if the spinal cord at the limb level is rotated by 180° along the A–P axis in chicken embryos, the motor axons originating from the ectopic position could finally reach the normal muscle target.

To sense such various environmental cues, the tip of the axon is specialized into a structure called *growth cone* (Fig. 10a). During axon guidance, growth cones appear to contact either supporting cells or previously established nerve tracts from other neurons. Those cells harboring the activity to guide growth cones are named *guidepost/stepping-stone cells* and *pioneer neurons* respectively (Fig. 10a). Once a growth cone arrives at a target region, it stops and forms a stable junction with the target neuron, the maturated form of which is termed synapse. The neural circuits at this point are, however, premature and sometimes contain wrong connections. For instance, while only a single climbing fiber innervates each Purkinje neuron in the adult mouse cerebellum, multiple innervations were found at perinatal stages. Such wrong connections are eliminated in an activity dependent manner. In the mammalian visual cortex, inputs from the right eye compete with those from the left eye to make the ocular dominance column. In this process, inputs from both eyes have been shown to occupy the same area at the initial step of the development but eventually lose the connection to form the ocular dominance column. Elimination of inputs from one eye during the initial stages resulted in the loss of the ocular dominance column, indicating that this process was also activity dependent. Even in the adult brain, neuronal circuits are always modified over time mainly by changing the synaptic junctions in number and strength.

6 Molecular Aspects to Wire up the Nerve Cells into Functional Circuits

There are mainly two mechanisms to explain axon guidance; one is a contact dependent and the other is a contact-independent machinery. In the former, interaction between the transmembrane and/or membrane-bound receptor in the growth cone and its ligands exists on the neighboring cell membrane or extracellular matrix (ECM) and plays a major role, while secreted molecules from remote positions and their receptors on the membrane of growth cones are involved in the latter. The following are examples of these two mechanisms:

Fig. 10 Mechanisms involved in axon guidance. (a) A schematic drawing explaining how a growth cone can be guided to the target area. For details, see text, (b) The retino-tectum connection system in chicken. Visual information is precisely represented in the tectum because of the topographic connection pattern. In order to establish such a connection pattern, the graded expression of the Eph receptor in the retina and that of its ligand ephrin in the tectum appear to play an important role, (c) A commissural interneuron in the spinal cord can express different sets of molecules along the axon in response to various environmental cues. (d) Secreted molecule Slit mediated axon guidance in the *Drosophila* embryo is strictly regulated by the posttranscriptional modification of the Slit receptor Robo. The midline glia is the source of the chemorepellant, Slit.

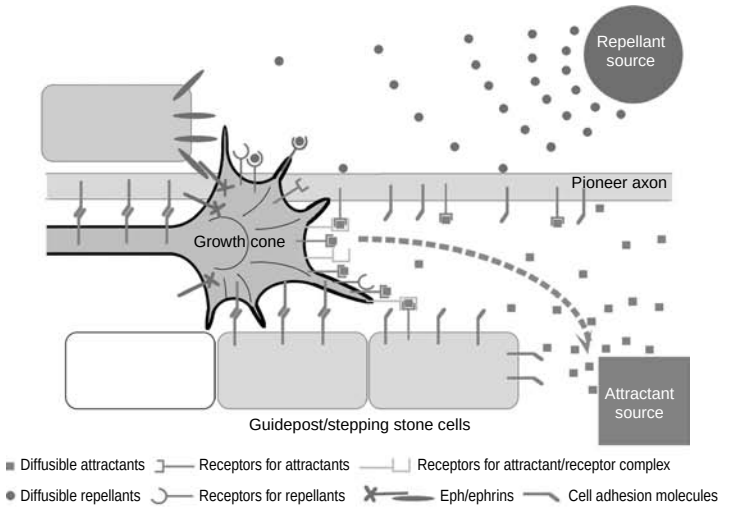

(a)

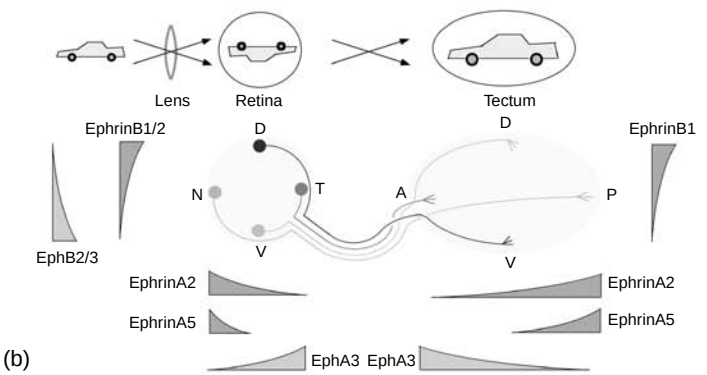

(b)

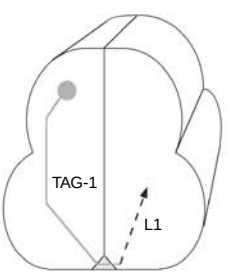

(c) Floor plate

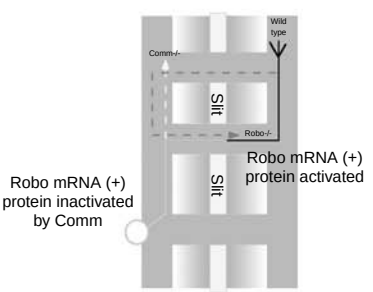

(d) Midline glia

6.1 Contact-dependent (Short Range) Axon Guidance Mechanisms

In chicken embryos, the retinal ganglion cells in the eye send axons toward a midbrain region, the tectum. Noticeable is the fact that the connection pattern between the retina and the tectum is totally topographic; nasal ganglion cells send axons to the posterior tectum and the temporal ones to the anterior tectum. Furthermore, the dorsal ganglion cells send axons to the ventral tectum and the ventral ones to the dorsal tectum, enabling them to simply relay the two-dimensional information caught by an eye to the brain (Fig. 10b). To precisely establish such a connection pattern, Eph tyrosine kinase receptor, and its ligand ephrin have been shown to play a crucial role: EphA3 expression makes a gradient along the nasal–temporal axis in the retina (higher in the temporal; see Fig. 10b) while the ligand ephrinA2/5 makes a gradient along the A–P axis in the tectum (higher in the posterior; see Fig. 10b). The interactions between Eph and ephrin cause the collapse of growth cones, steering the axon into the area with less interactions between the receptor and the ligand. The same mechanism seems to be utilized in the thalamocortical projections in mice, as EphrinA2 and A5 expression make a gradient in the thalamic lateral geniculate nucleus, and EphrinA2/A5 double mutant mice have defects in this connection system. Regarding the upstream factors required to regulate the EphA3 gradient in the chicken retina along the nasal-temporal axis, transcription factors such as CBF-1/2 are known to make a similar gradient to EphA3. Ectopic expression of CBF-1/2 in the retina has been shown to perturb the topographic retina–tectum projection patterns, implicating these molecules in the important role of axon guidance. However, there is no evidence that these transcription factors directly regulate EphA3 expression and there could be additional uncharacterized mechanisms. In the tectum, FGF8 and Wnt1 expression at the MHB is known to be required to set up an expression gradient of transcription factors Engrailed1/2 along the A–P axis, controlling the gradual expression of ephrinA2 and A5 (Fig. 8a).

Other important molecules steering the growth cone in a contact-dependent manner are cell adhesion molecules. Cell adhesion molecules could further be a component to stabilize the synaptic junctions immediately after the axon guidance processes. There are three major types of cell adhesion molecules: (1) immunoglobulin (Ig) superfamily molecules, (2) integrins, and (3) cadherins. The following is a brief summary of the roles of these classes of adhesion molecules in axon guidance:

6.1.1 Ig Superfamily Molecules

Ig superfamily molecules are simply defined by containing the Ig domain folded by a disulfide link. Numbers of Ig superfamily molecules with multiple Ig domains are expressed in the CNS. The smallest Ig superfamily molecule expressed in the CNS is Thy-1, which contains just one Ig domain. Thy-1 is a major fraction of the surface protein on long axons, stabilizing the membrane to prevent neurite sprouting during the guidance process. N-CAM (Neural cell adhesion molecule) has five Ig domains and two fibronectin type III (FNIII) repeats, and its expression is found everywhere in the CNS, providing a good neural marker during development. *In vitro* assay showed that homophilic interaction between axons and cell substrates expressing N-CAM is crucial to promote axon outgrowth. Interestingly,

it is known that sugar groups could modify the N-CAM extracellular domain (polysialylated form; PSA-N-CAM). This modification lowers the binding affinity, providing repulsive cues for axon guidance. TAG-1/axonin-1 and L1 with six Ig domains show a curious expression pattern during the axon guidance process: The commissural neurons of the vertebrate spinal cord first send axons toward the floor plate and the axons cross the midline at the floor plate. After crossing the midline, the axons immediately change direction by 90° to ascend the spinal cord (Fig. 10c). While TAG-1 expression is detected during the initial axon guidance process and is downregulated after the turning, L1 expression is only switched on after the turning (Fig. 10c). This indicates that localization and/or expression of guidance molecules even along the same axon could be tightly regulated, allowing growth cones to sense various environmental cues over time and space. Interestingly, TAG-1 is known to heterotypically interact with Nr-CAM (Ng-CAM related Ig superfamily adhesion molecule), which is expressed by the floor plate cells. Such molecular interactions occurring at a restricted territory *per se* might have a role in changing the growth cone behaviors.

6.1.2 Integrins

Integrins are the cell surface receptors and their substrates are the ECM molecules such as laminin, fibronectin, and so on. Integrins consist of two subunits, namely, α- and β-chains. Distinct genes encode α- and β-chains, and differential combinations of α/β-chains produce various affinities to ECM ligands/substrates. In the CNS, multiple types of receptors and their substrates are expressed, and preferential interactions between integrins and molecules in the ECM are indeed shown to play essential roles in guiding neurites *in vitro*. The cytoplasmic domain of integrins is known to interact with cytoskeletal elements, such as vinculin, talin, and α-actinin, each of which is found to localize at the growth cones. Importantly, integrins have been shown to play a major role in directed cell migration by remodeling the cytoskeltons, and migration of mouse cortical neurons has been demonstrated to be dependent of integrins. Mechanisms that regulate cellular migrations would be utilized in the directed steering of axons.

6.1.3 Cadherins

Cadherins are transmembrane proteins that confer subclass specific adhesiveness to cells: cells that express one particular subclass can generally contact with those that express the same subclass in a Ca^{2+} dependent manner. Takeichi and colleagues first demonstrated that functional neuronal circuits in the chicken nervous system were subdivided by differential expressions of N- and R-cadherin. In the mouse thalamocortical connection system, expressions of type II cadherins have also been shown to delineate the specific division in the thalamus as well as the neocortical area. Furthermore, cadherin-6 expression was found to demarcate entire parts of the neuronal circuit such as the auditory circuit. Redies and colleagues have revealed that such differential expression patterns along the neuronal circuit play a crucial role in finally targeting the axons into a specific area by means of the EP mediated cadherin overexpression experiment. In the *Drosophila* embryo, it has also been shown that *D*N-cadherin plays a role in the segregation of axon tracts in a fascicle, otherwise the neural circuits were totally disorganized.

6.2 Contact-independent (Long-range) Axon Guidance Mechanisms

Many secreted molecules and their ligands play a central role in steering axon guidance. For example, interneurons in the vertebrate spinal cord send the commissural axons toward the ventralmost midline cells, the floor plate. Growth cones then cross the midline and project anteriorly as shown in Fig. 10(c). In the first step of these processes, *Netrin* secreted from the floor plate has been shown to have an activity that attract the growth cone expressing the receptor DCC (Deleted in Colorectal Cancer), an Ig superfamily molecule with four Ig domains and six FNIII repeats. Netrin was originally identified as the gene responsible for the uncoordinated cell migration phenotype (unc-6) in *C. elegans*, and importantly, it is known that Netrin can function as a repellant if the DCC receptor interacts with another receptor UNC-5. The trochlear motor neurons are located in the ventral region of the hindbrain and send axons dorsally. Netrin has indeed been shown to regulate this axon guidance step in a repulsive manner, yet the receptor components in this process are not confirmed to be UNC-5 dependent. Recently, the *Drosophila* Netrin receptor Frazzled (*Drosophila* DCC homolog) has been demonstrated to guide axons by controlling Netrin distribution. In this model, Netrin captured and relocated by the receptor Frazzled along the pioneer axon, could further act as a contact-dependent attractive guidance cue for growth cones expressing unknown Netrin receptors other than Frazzled, implicating complex molecular machineries involved in the Netrin-mediated axon guidance (Fig. 10a).

The secreted molecule Slit can act as a repellant and the receptor is an Ig superfamily member Roundabout (Robo) with five Ig domains and three FNIII repeats. In *Drosophila*, Slit is expressed at the ventral midline cells and this prevents axons that had once crossed the midline from recrossing it (Fig. 10d). Therefore, growth cones in the Slit receptor Robo mutants cross the midline many times (Fig. 10d). Notable is the fact that growth cones after crossing the midline become sensitive to Slit repulsive cues in the wild type, although the Robo mRNA is expressed by almost all neurons before crossing the midline (Fig. 10d). It is revealed that a transmembrane protein, Commissureless (Comm), transiently inactivates the Robo protein until growth cones cross the midline, otherwise axons can never cross the midline because of the repulsive activity of Slit, resulting in the commisureless phenotype in the mutant (Fig. 10d). As the Robo protein is found to be sequestered in the internal vesicles of growth cones before crossing the midline, Comm might regulate the localization of the Robo protein during axon guidance. Robo has further been demonstrated to be able to interact with DCC, the Netrin receptor. This interaction has been shown to be essential not only to silence the Netrin attractive signals but also to potentiate the Slit repulsive signals after crossing the midline.

Semaphorins are another type of secreted molecule, and both neuropilins and plexins are the receptor components of semaphorins. In the mouse, neuropilin-1 (NP-1) is a known receptor component of SemaphorinD (SemaD; the human homolog is named SemaphorinIII, while the chicken homolog is termed CollapsinI) and is expressed by the neurons in the

dorsal root ganglia (DRG). *In vitro* experiments have revealed that SemaD expressing cells indeed repel or collapse growth cones from DRG neurons, whereas growth cones from DRG neurons isolated from NP-1 gene targeting mice cannot be affected by SemaD expressing cells. *In vivo*, efferent axons from the NP-1 mutant DRG were found to be defasciculated and/or harbor excess spreading, suggesting an important role for NP-1/SemaD in constituting the guidance pathway. The NP-1/SemaD interactions also appear to regulate the initial patterning of projections from neocortical neurons: cortical neurons expressing NP-1 sense the SemaD signal from the marginal zone, sending efferent axons to avoid the SemaD signals.

Receiving various contact-dependent and contact-independent signals and/or cues, migrating growth cones sum up these stimulations and determine the final direction to go. Such decisions are mainly made within the cytoplasm of growth cones, and small signaling molecules play major roles. For example, VAB-8, a cytoplasmic protein with kinesin motor similarity, is normally required for the posterior axon guidance in *C. elegans*, and ectopic expression of VAB-8 has been shown to be sufficient to totally reverse the direction of axon outgrowth. The ratios of activity of a cytoplasmic tyrosine kinase, ABL, its substrate ENA, and the tyrosine phosphatases, which antagonize ENA, have also been demonstrated to be balanced by various Robo mediated guidance signals and finally set a direction to go. Activity and/or localization of a second messenger such as cyclic-AMP and cyclic-GMP have further been revealed to affect the final direction to go in *Xenopus* spinal axons and the cortical efferents respectively. Regarding the turning response of growth cones, local activation of the Rho family of GTPases, which is able to rearrange the cytoskeltal organization, has been implicated. For instance, Eph/ephrin interactions could modulate Rho activity, yet the detailed machinery that link distinct GTPases to other guidance cues is not well understood.

It should be apparent that the molecular mechanisms involved in axon guidance are used in controlling the direction of cellular migrations. For instance, the Eph-ephrin repulsive signals together with the cadherin-mediated cell adhesion machinery play a crucial role in generating cell lineage restriction between neuromeres (see above). Neural crest cells also require Eph/ephrin signaling, when they migrate from the dorsal neural tube and travel toward their final locations. Furthermore, the rostral migration stream from the ganglionic eminence to the olfactory bulb in the forebrain as well as the cytokine induced leukocytes migration has been shown to be dependent on Slit signaling. Such cellular machineries are so important that they might repeatedly be utilized during development and evolution.

6.3 Plasticity of the Neuronal Circuit-potentiation of Synaptic Transmission

Here, I just outline the molecular mechanisms involved in the potentiation of synaptic transmission, because detailed mechanisms have been well summarized in a chapter of the previous series written by Dr Eric Kandel and references therein.

As I described earlier, it has been reported that the morphological as well as physiological features of synaptic junctions can vary over time in the mature brain. For instance, if an excitory presynaptic neuron is artificially stimulated at

high frequency, the postsynaptic membrane potential is known to increase and the situation can last from minutes to several hours. This persistent increase in the synaptic membrane potential is called *posttetanic potentiation* and in the *in vivo* context, the process is called long-term facilitation (LTF) in *Aplysia*, while in vertebrates it is named long-term potentiation (LTP). Morphologically, many buds and/or branches termed *spines* are formed along the postsynaptic dendrites. In addition, the area of synaptic contact has been demonstrated to become broader during LTP. Molecularly, it is revealed in the excitory synapses that components of the postsynaptic ion channels are modified to facilitate the Ca^{2+} permeability through the postsynaptic membrane. Increased amounts of Ca^{2+} then activate the Ca^{2+} dependent kinase and further modify ion channels to increase the Ca^{2+} permeability. Continuous Ca^{2+} accumulations finally activate the PKA pathways at the postsynapse, resulting in turning on a set of gene transcriptions, such as cell adhesion molecules, to strengthen the synaptic junction. From the postsynaptic membrane, signaling molecules such as nitric oxide (NO) and neurotrophic factors, such as BDNF (brain-derived neurotrophic factor), are thought to be released and affect transmitter release as well as survival rates of the presynaptic neuron terminals. Neurotrophic factors further play an important role in modulating components of ion channels. Recently, the role of those genes that are implicated in LTP formation have been eliminated by the gene targeting method in mice (see Sect. 7), and these mice more or less showed aberrant phenotype in the activity dependent generation of brain functional units during development as well as during learning and memory processes at adult stages, indicating LTP might be a crucial element of these intricate processes of the brain.

7 How can Molecular Functions in Complex Tissues be Studied?

Many methods have been established to determine molecular functions *in vivo*. There are basically two methods to examine molecular functions: *loss-of-function* and *gain-of-function* studies. In the former, a gene and/or its protein product of interest is eliminated in the relevant tissue and/or entire organism, and researchers analyze the abnormalities to understand its normal functions. In the latter, a gene and/or its protein product is ectopically expressed in the tissue where the gene is not expressed, and researchers examine the acquired phenotypes of the tissue to know its original functions. Here, I briefly outline emerging methodologies to clarify the complex molecular functions in the brain.

7.1 Loss-of-function Studies

Applying materials that block generation of a gene product and/or the function of the gene product would be the simplest way in this criterion. For example, several drugs/antagonists are known to block functions of specific sets of ion channels expressed by neurons, and these reagents have been playing pivotal roles in elucidating synaptic functions in the nervous system. Designing *dominant negative molecules* could be another option to inhibit the normal functions of endogenous molecules. For instance, several types of receptors such as Notch, BMP receptors, FGF receptors, and so on are

known to be activated through dimer formations, and if truncated forms are expressed, normal receptor functions are effectively perturbed (Fig. 2a). In the *C. elegans* system, double-stranded RNA sequences, which are simply designed from the coding sequences of a gene of interest, have been demonstrated to specifically interfere with mRNA transcription and/or translation of the gene. This mechanism termed *RNAi* is now shown to be applicable not only to invertebrates but also to vertebrates including the mouse. Applying the *antisense oligonucelotides* for a given gene sequence, has also been revealed to block mRNA translation. Most recently, it has been demonstrated that the *morpholino oligonucleotides* serve as the best tool to stably block gene functions because the morpholino reagents are not degraded, thereby inhibiting mRNA translation for a long time.

Generating an organism that completely lacks a gene of interest is a prevailing method, and the manipulated organism could yield more reproducible and reliable results. In the mouse, embryonic stem cells (ES) that can differentiate into all cell types in an organism have been established, and it is now a routine process to replace a gene of interest with a drug selection gene in the ES cells via homologous recombination. Selected ES cells can directly be injected in a host embryo, and chimeric mice containing cells derived from the gene manipulated ES cells are easily obtained. If ES cell–derived cells contribute to the germ line, mice homozygous for the mutated locus can be finally produced by means of serial mating processes (Fig. 11). This procedure, called *gene targeting* or *gene knockout*, has greatly contributed to the understanding of the unveiled functions of genes expressed in the mouse brain. Major problems in loss-of-function studies are that redundant functions of related genes sometimes mask the mutant phenotypes and that the abnormalities at earlier developmental stages and/or lethality prevent the mutant phenotypes from being examined at later stages. The former problem might simply be overcome by generating double or more related gene knockouts, while the latter could partly be solved by using the conditional gene knockout system that allows the gene of interest to be inactivated in a spatiotemporally restricted manner.

7.2 Gain-of-function Studies

To express a gene of interest ectopically, ubiquitous promoters such as heat shock, beta actin, cytomegalovirus (CMV) promoters, and so on must be linked to the gene, and the construct should be delivered into tissues or organisms. If a tissue or cell type specific overexpression is required, the researchers could use proper types of promoters. For example, the promoter of neuron-specific enolase is widely used to drive the neuron-specific expression of exogenous genes. Synthesized mRNA with a 5′ cap can also be injected into a fertilized egg or the blastomere of a developing embryo, and strong expression from the injected mRNA has been reported in the zebrafish and frog systems. In the mouse, the constructs for overexpression can be easily microinjected into the pronuclei of fertilized eggs, and mice lines with the exogenous gene will be obtained after oviduct transfer of the injected eggs.

Recently, the EP mediated gene transfer method into living embryos and/or tissues has emerged a powerful tool. Using controlled square pulses instead of authentic pulses with exponential decay,

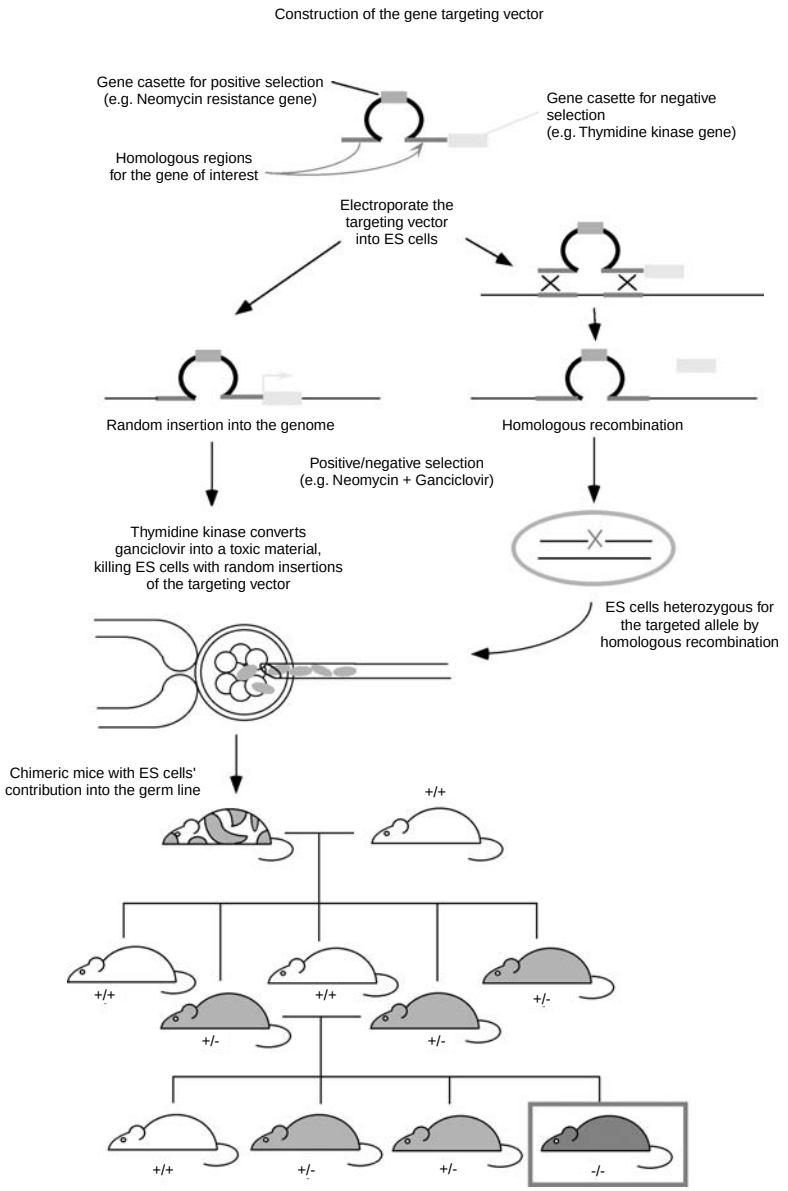

Fig. 11 How to generate gene targeting (knockout) mice.

it is now possible to drive gene expressions in any tissue at any developmental stage. Compared to the virus mediated gene transfer method, EP mediated gene transfer is less toxic to the relevant tissue as well as for researchers. Furthermore, EP is superior in its capability to control the area of gene transfer because DNA constructs with a negative charge will only migrate to the tissue facing the

positive electrode, directed by the electric field generated during EP. EP may further be combined with loss-of-function studies, as EP can target any charged molecules. Indeed, RNAi as well as morpholino oligo-mediated gene attenuation in a restricted area using EP has already been reported. Hence, EP will offer a relatively high-throughput means of assaying or evaluating gene function and regulation, which will be important in light of the flood of information coming from the genome projects.

Acknowledgments

I would like to thank Dr. Robb Krumlauf for valuable suggestions in preparing the manuscript at the Stowers Institute for Medical Research, Kansas City, MO, and an anonymous reviewer for critical reading of the manuscript. I also thank the members of the Krumlauf lab for helpful discussions. Takayoshi Inoue is supported by the Human Frontier Science Program Long-term fellowship (LT00293/2000-M).

See also Behavior Genes; Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Arendt, D., Nuler-Jung, K. (1999) Comparison of early nerve cord development in insects and vertebrate, *Development* **126**, 2309–2325.

Bayer, S.A., Altman, J. (1991) *Neocortical Development*, Raven Press, New York.

De Robertis, E.M., Sasai, Y. (1996) A common plan for dorsoventral patterning in bilateria, *Nature* **380**, 37–40.

Inoue, T., Krumlauf, R. (2001) An impulse to the brain: using in vivo electroporation, *Nat. Neurosci.* **4 Suppl.**, 1156–1158.

Lee, S.K., Pfaff, S.L. (2001) Transcriptional networks regulating neuronal identity in the developing spinal cord, *Nat. Neurosci.* **4 Suppl.**, 1183–1191.

Lumsden, A., Krumlauf, R. (1996) Patterning the vertebrate neuraxis, *Science* **274**, 1109–1115.

Matsuzaki, F. (2000) Asymmetric division of Drosophila neural stem cells: a basis for neural diversity, *Curr. Opin. Neurobiol.* **10**, 38–44.

O'Leary, D.D.M., Koester, S. (1993) Development of projection neuron types, axon pathways, and patterned connection of the mammalian cortex, *Neuron* **10**, 991–1006.

Puelles, L., Rubenstein, J.L. (1993) Expression patterns of homeobox and other putative regulatory genes in the embryonic mouse forebrain suggest a neuromeric organization, *Trends Neurosci.* **16**, 472–479.

Redies, C. (2000) Cadherins in the central nervous system, *Prog. Neurobiol.* **61**, 611–648.

Spemann, H., Mangold, H. (1924) Der Induktion von Embryonalanlagen durch Implantation artfremder Organisatoren, *Arch. mikrosk. Anat. und Entw.mechan.; Wilh. Roux Arch. Entw. Mech. Organ.* **100**, 599–638.

Takeichi, M. (1995) Morphogenetic roles of classic cadherins, *Curr. Opin. Cell Biol.* **7**, 619–627.

Tam, P.P., Behringer, R.R. (1997) Mouse gastrulation: the formation of a mammalian body plan, *Mech. Dev.* **68**, 3–25.

Yu, T.W., Bargmann, C.I. (2001) Dynamic regulation of axon guidance, *Nat. Neurosci.* **4 Suppl.**, 1169–1176.

Primary Literature

Akiyama-Oda, Y., Hosoya, T., Hotta, Y. (1999) Asymmetric cell division of thoracic neuroblast 6-4 bifurcate glial and neuronal lineage in *Drosophila*, *Development* **126**, 1967–1974.

Altman, J., Bayer, S.A. (1984) The development of the rat spinal cord, *Adv. Anat. Embryol. Cell Biol.* **85**, 1–166.

Amaya, E., Musci, T.J., Kirschner, M.W. (1991) Expression of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in *Xenopus* embryos, *Cell* **66**, 257–270.

Ang, S.L., Rossant, J. (1994) *HNF-3b* is essential for node and notochord formation in mouse development, *Cell* **78**, 561–574.

Anton, E.S., Kreidberg, J.A., Rakic, P. (1999) Distinct functions of alpha3 and alpha(v) integrin receptors in neuronal migration and laminar organization of the cerebral cortex, *Neuron* **22**, 277–289.

Bachiller, D., Klingensmith, J., Kemp, C., Belo, J.A., Anderson, R.M., May, S.R., McMahon, J.A., McMahon, A.P., Harland, R.M., Rossant, J., De Robertis, E.M. (2000) The organizer factors Chordin and Noggin are required for mouse forebrain development, *Nature* **403**, 658–661.

Baker, C.V., Bronner-Fraser, M. (1997) The origin of the neural crest. Part I: embryonic induction, *Mech. Dev.* **69**, 3–11.

Bashaw, G.J., Kidd, T., Murray, D., Pawson, T., Goodman, C.S. (2000) Repulsive axon guidance: Abelson and Enabled play opposing roles downstreams of the roundabout receptor, *Cell* **101**, 703–715.

Bayer, S.A., Altman, J. (1990) Development of layer I and the subplate in the rat neocortex, *Exp. Neurol.* **107**, 48–62.

Bell, E., Wingate, R., Lumsden, A. (1999) Homeotic transformation of rhombomere identity after localized *Hoxb1* misexpression, *Science* **284**, 21682171.

Bishop, K.M., Goudreau, G., O'Leary, D.M. (2000) Regulation of area identity in the mammalian neocortex by Emx2 and Pax6, *Science* **288**, 344–349.

Blair, S.S. (2000) Notch signaling: fringe really is a glycosyltransferase, *Curr. Biol.* **10**, R608–R612.

Briscoe, J., Pierani, A., Jessell, T.M., Ericson, J. (2000) A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube, *Cell* **101**, 435–445.

Broccoli, V., Boncinelli, E., Wurst, W. (1999) The caudal limit of Otx2 expression positions the isthmic organizer, *Nature* **401**, 164–168.

Brose, K., Bland, K.S., Wang, K.H., Arnott, D., Henzel, W., Goodman, C.S., Tessier-Lavigne, M., Kidd, T. (1999) Slit proteins bind Robo receptors and have an evolutionarily conserved role in repulsive acon guidance, *Cell* **96**, 795–806.

Bulfone, A., Puelles, L., Porteus, M.H., Frohman, M.A., Martin, G.R. (1993) Spatially restricted expression of Dlx-1, Dlx-2 (Tes-1), Gbx-2, and Wnt-3 in the embryonic day 12.5 mouse forebrain defines potential transverse and longitudinal segmental boundaries, *J. Neurosci.* **13**, 3155–3172.

Calegari, F., Haubensak, W., Yang, D., Huttner, W.B., Buchholz, F. (2002) Tissue-specific RNA interference in postimplantation mouse embryos with endoribonuclease-prepared short interfering RNA, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14236–14240.

Cano, A., Perez-Moreno, M.A., Rodrigo, I., Locascio, A., Blanco, M.J., del Barrio, M.G., Portillo, F., Nieto, M.A. (2000) The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression, *Nat. Cell Biol.* **2**, 76–83.

Carpenter, E.M., Goddard, J.M., Chisaka, O., Manley, N.R., Capecchi, M.R. (1993) Loss of Hox-A1(Hox-1.6) function results in the reorganization of the murine hindbrain, *Development* **118**, 1063–1075.

Castellani, V., Rougon, G. (2002) Control of semaphorin signaling, *Curr. Opin. Neurobiol.* **12**, 532–541.

Caviness, V.S.J. (1982) Neocortical histogenesis in normal and reeler mice: a developmental study based upon ^{3}H-Thymidine autoradiography, *Dev. Brain Res.* **4**, 293–302.

Chitnis, A., Henrique, D., Lewis, J., Ish-Horowicz, D., Kintner, C. (1995) Primary neurogenesis in *Xenopus* embryos regulated by a homologue of the *Drosophila* neurogenic gene Delta, *Nature* **375**, 761–766.

Cobos, I., Shimamura, K., Rubenstein, J.L., Martinez, S., Puelles, L. (2001) Fate map of the avian anterior forebrain at the four-somite stage, based on the analysis of quail-chick chimeras, *Dev. Biol.* **239**, 46–67.

Colamarino, S.A., Tessier-Lavigne, M. (1995) The axonal chemoattractant netrin-1 is also a chemorepellent for trochlear motor axons, *Cell* **81**, 621–629.

Corey, D.R., Adrams, J.M. (2001) Morpholino antisense oligonucleotides: tools for investigating vertebrate development, *Genome Biol.* **2**, reviews1015.1–3.

Crossley, P.H., Martinez, S., Martin, G.R. (1996) Midbrain development induced by *FGF8* in the chick embryo, *Nature* **380**, 66–68.

Cypher, C., Letourneau, P.C. (1991) Identification of cytoskeltal, focal adhesion, and cell adhesion proteins in growth cone particles isolated from developing chick brain, *J. Neurosci. Res.* **30**, 259–265.

D'Arcangelo, G., Hamayouni, R., Keshvara, L., Rice, D.S., Sheldon, M., Curran, T. (1999) Reelin is a ligand for lipoprotein receptors, *Neuron* **24**, 471–479.

D'Arcangelo, G., Miao, G.G., Chen, S.C., Soares, H.D., Morgan, J.I. (1995) A protein related to extracellular matrix proteins deleted in the mouse mutant reeler, *Nature* **374**, 719–723.

Davidson, B.P., Kinder, S.J., Steiner, K., Schoenwolf, G.C., Tam, P.P. (1999) Impact of node ablation on the morphogenesis of the body axis and the lateral asymmetry of the mouse embryo during early organogenesis, *Dev. Biol.* **211**, 11–26.

Davidson, B.P., Tam, P.P. (2000) The node of the mouse embryo, *Curr. Biol.* **10**, R617–R619.

de la Cruz, C.C., Der-Avakian, A., Spyropoulos, D.D., Tieu, D.D., Carpenter, E.M. (1999) Targeted disruption of Hoxd9 and Hoxd10 alters locomotor behavior, vertebral identity, and peripheral nervous system development, *Dev. Biol.* **216**, 595–610.

Detrick, R.J., Dickey, D., Kintner, C.R. (1990) The effect of N-cadherin misexpression on morphogenesis in *Xenopus* embryos, *Neuron* **4**, 493–506.

Ding, J., Yang, L., Yan, Y.T., Chen, A., Desai, N., Wynshaw-Boris, A., Shen, M.M. (1998) Cripto is required for correct orientation of the anterior-posterior axis in the mouse embryo, *Nature* **395**, 702–707.

Dodd, J., Morton, S.B., Karagogeos, D., Yamamoto, M., Jessell, T.M. (1988) Spatial regulation of axonal glycoprotein expression on subsets of embryonic spinal neurons, *Neuron* **1**, 105–116.

Doherty, P., Walsh, F.S. (1991) The contrasting roles of N-CAM and N-cadherin as neurite outgrowth-promoting molecules, *J. Cell Sci. Suppl.* **15**, 13–21.

Driever, W., Nuslein-Volhard, C. (1988) A gradient of bicoid protein in *Drosophila* embryos, *Cell* **54**, 83–93.

Driever, W., Nuslein-Volhard, C. (1989) The bicoid protein is a positive regulator of hunchback transcription in the early *Drosophila* embryo, *Nature* **337**, 138–143.

Dulabon, L., Olson, E.C., Taglienti, M.G., Eienhuth, S., McGrath, B., Walsh, C.A., Kreidberg, J.A., Anton, E.S. (2000) Reelin binds alpha3beta1 integrin and inhibits neuronal migration, *Neuron* **27**, 33–44.

Ericson, J., Rashbass, P., Schedl, A., Brenner-Morton, S., Kawakami, A., van Heyningen, Y., Jessell, T.M., Briscoe, J. (1997) Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling, *Cell* **90**, 169–180.

Espeseth, A., Marnellos, G., Kintner, C. (1998) The role of F-cadherin in localizing cells during neural tube formation in *Xenopus* embryos, *Development* **125**, 301–312.

Feldheim, D.A., Vanderhaeghen, P., Hansen, M.J., Frisen, J., Lu, Q., Barbacid, M., Flanagan, J.G. (1998) Topographic guidance labels in a sensory projection to the forebrain, *Neuron* **21**, 1303–1313.

Figdor, M.C., Stern, C.D. (1993) Segmental organisation of embryonic diencephalon, *Nature* **363**, 630–634.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*, *Nature* **391**, 806–811.

Fishell, G., Mason, C.A., Hatten, M.E. (1993) Dispersion of neural progenitors within the germinal zones of the forebrain, *Nature* **362**, 636–638.

Fraser, S., Keynes, R., Lumsden, A. (1990) Segmentation in the chick embryo hindbrain is defined by cell lineage restrictions, *Nature* **344**, 431–435.

Fujimori, T., Miyatani, S., Takeichi, M. (1990) Ectopic expression of N-cadherin perturbs histogenesis in *Xenopus* embryos, *Development* **110**, 97–104.

Fukuchi-Shimogori, T., Grove, E.A. (2001) Neocortex patterning by the secreted signaling molecule FGF8, *Science* **294**, 1071–1074.

Garcia-Castro, M.I., Marcelle, C., Bronner-Fraser, M. (2002) Ectodermal Wnt function as a neural crest inducer, *Science* **297**, 848–851.

Gavalas, A., Davenne, M., Lumsden, A., Chambon, P., Rijli, F. (1997) Role of *Hoxa-2* in axon pathfinding and rostral hindbrain patterning, *Development* **124**, 3693–3702.

Goffinet, A.M. (1997) Unscrambling a disabled brain, *Nature* **389**, 668,669.

Gotz, M., Wizenmann, A., Lumsden, A., Price, J. (1996) Selective adhesion of cells from different telencephalic regions, *Neuron* **16**, 551–564.

Gould, A., Itasaki, N., Krumlauf, R. (1998) Initiation of rhombomeric *Hoxb4* expression requires induction by somites and a retinoid pathway, *Neuron* **21**, 39–51.

Greenwald, I., Rubin, G.M. (1992) Making a difference: the role of cell-cell interactions in establishing separate identities for equivalent cells, *Cell* **68**, 271–281.

Guo, M., Jan, L.Y., Jan, Y.N. (1996) Control of daughter cell fates during asymmetric division: interaction of Numb and notch, *Neuron* **17**, 27–41.

Hatta, K., Takeichi, M. (1986) Expression of N-cadherin adhesion molecules associated with early morphogenetic events in chick development, *Nature* **320**, 447–449.

Hedgecock, E.M., Culotti, J.G., Hall, D.H. (1990) The unc-5, unc-6, and unc-40 genes guide circumferential migrations of pioneer axons and mesodermal cells on the epidermis in *C. elegans*, *Neuron* **4**, 61–85.

Heikkila, R., Schwab, G., Wickstrom, E., Loke, S.L., Pluznik, D.H., Watt, R., Neckers, L.M. (1987) A c-myc antisense oligodeoxynucleotide inhibits entry into S phase but not progress from G0 to G1, *Nature* **328**, 445–449.

Hemmati-Brivanlou, A., Melton, D.A. (1992) A truncated activin receptor inhibits mesoderm induction and formation of axial structures in *Xenopus* embryos, *Nature* **359**, 586–587.

Hiesberger, T., Trommsdorff, M., Howell, B.W., Goffinet, A., Mumby, M.C., Cooper, J.A., Herz, J. (1999) Direct binding of Reelin to VLDL receptor and ApoE receptor 2 induces tyrosine phosphorylation of disabled-a and modulates tau phosphorylation, *Neuron* **24**, 481–489.

Hiramoto, M., Hiromi, Y., Giniger, E., Hotta, Y. (2000) The *Drosophila* Netrin receptor frazzled guides axons by controlling Netrin distribution, *Nature* **406**, 886–889.

Hirata, J., Nakagoshi, H., Nabeshima, Y., Matsuzaki, F. (1995) Asymmetric segregation of the homeoprotein Prospero during *Drosophila* development, *Nature* **377**, 627–630.

Hirotsune, S., Takahara, T., Sasaki, N., Hirose, K., Yoshiki, A., Ohashi, T., Kusakabe, M., Murakami, Y., Muramatsu, M., Watanabe, S., et al. (1995) The reeler gene encodes a protein with an EGF-like motif expressed by pioneer neurons, *Nat. Genet.* **10**, 77–83.

Hong, K., Hinck, L., Nishiyama, M., Poo, M.M., Tessier-Lavigne, M., Stein, E. (1999) A ligand-gated association between cytoplasmic domains of UNC5 and DCC family receptors converts netrin-induced growth cone attraction to repulsion, *Cell* **97**, 927–941.

Horwitz, A., Duggan, K., Greggs, R., Decker, C., Buck, C. (1985) The cell substrate attachment (CSAT) antigen has properties of a receptor for laminin and fibronectin, *J. Cell Biol.* **101**, 2134–2144.

Ikeya, M., Lee, S.M., Johnson, J.E., McMahon, A.P., Takada, S. (1997) Wnt signalling required for expansion of neural crest and CNS progenitors, *Nature* **389**, 966–970.

Ingham, P.W., Martinez-Arias, A. (1986) The correct activation of Antennapedia and bithorax complex genes requires the fushi tarazu gene, *Nature* **324**, 592–597.

Inoue, T., Chisaka, O., Matsunami, H., Takeichi, M. (1997) Cadherin-6 expression transiently delineates specific rhombomeres, other neural tube subdivisions, and neural crest subpopulations in mouse embryos, *Dev. Biol.* **183**, 183–194.

Inoue, T., Nakamura, S., Osumi, N. (2000) Fate mapping of the mouse prosencephalic neural plate, *Dev. Biol.* **219**, 373–383.

Inoue, T., Tanaka, T., Suzuki, S.C., Takeichi, M. (1998) Cadherin-6 in the developing mouse brain: expression along restricted connection systems and synaptic localization suggest a potential role in neuronal circuitry, *Dev. Dyn.* **211**, 338–351.

Inoue, T., Tanaka, T., Takeichi, M., Chisaka, O., Nakamura, S., Osumi, N. (2001) Role of cadherins in maintaining the compartment boundary between the cortex and striatum during development, *Development* **128**, 561–569.

Itasaki, N., Ichijo, H., Hama, C., Matsuno, T., Nakamura, H. (1991) Establishment of rostrocaudal polarity in tectal primordium: engrailed expression and subsequent tectal polarity, *Development* **113**, 1133–1144.

Itasaki, N., Sharpe, J., Morrison, A., Krumlauf, R. (1996) Reprogramming *Hox* expression in the vertebrate hindbrain: influence of paraxial mesoderm and rhombomere transposition, *Neuron* **16**, 487–500.

Iwai, Y., Usui, T., Hirano, S., Steward, R., Takeichi, M., Uemura, T. (1997) Axon patterning requires *D*N-cadherin, a novel neuronal adhesion receptor, in the *Drosophila* embryonic CNS, *Neuron* **19**, 77–89.

Jack, T., Regulski, M., McGinnis, W. (1988) Pair-rule segmentation genes regulate the expression of the homeotic selector gene, deformed, *Genes Dev.* **2**, 592–597.

Jacques, T.S., Relvas, J.B., Nishimura, S., Pytela, R., Edwards, G.M., Streuli, C.H., ffrench-Constant, C. (1998) Neural precursor cell chain migration and division are regulated through different beta1 integrins, *Development* **125**, 3167–3177.

Jessell, T.M., Bovolenta, P., Placzek, M., Tessier-Lavigne, M., Dodd, J. (1989) Polarity and patterning in the neural tube: the origin and function of the floor plate, *Ciba Found. Symp.* **144**, 255–276.

Johansson, C.B., Momma, S., Clarke, D.L., Risling, M., Lendahl, U., Frisen, J. (1999) Identification of a neural stem cell in the adult mammalian central nervous system, *Cell* **96**, 25–34.

Jungbluth, S., Bell, E., Lumsden, A. (1999) Specification of distinct motor neuron identities by the singular activities of individual Hox genes, *Development* **126**, 2751–2758.

Jungbluth, S., Larsen, C., Wizenmann, A., Lumsden, A. (2001) Cell mixing between the embryonic midbrain and hindbrain, *Curr. Biol.* **11**, 204–207.

Katahira, T., Sato, T., Sugiyama, S., Okafuji, T., Araki, I., Funahashi, J., Nakamura, H. (2000) Interaction between Otx2 and Gbx2 defines the organizing center for the optic tectum, *Mech. Dev.* **91**, 43–52.

Kaufman, T.C., Seeger, M.A., Olsen, G. (1990) Molecular and genetic organization of the Antennapedia gene complex of *Drosophila melanogaster*, *Adv. Genet.* **27**, 309–362.

Keino-Masu, K., Masu, M., Hinck, L., Leonardo, E.D., Chan, S.S., Culotti, J.G., Tessier-Lavigne, M. (1996) Deleted in colorectal cancer (DCC) encodes a netrin receptor, *Cell* **87**, 175–185.

Keleman, K., Rajagopalan, S., Cleppien, D., Teis, D., Paiha, K., Huber, L.A., Technau, G.M., Dickson, B.J. (2002) Comm sorts robo to control axon guidance at the *Drosophila* midline, *Cell* **110**, 415–427.

Kennedy, T.E., Serafini, T., de la Torre, J.R., Tessier-Lavigne, M. (1994) Netrins are diffusible chemotropic factors for commissural axons in the embryonic spinal cord, *Cell* **78**, 425–435.

Kidd, T., Bland, K.S., Goodman, C.S. (1999) Slit is the midline repellent for the robo receptor in *Drosophila*, *Cell* **96**, 785–794.

Kitamura, K., Miura, H., Yanazawa, M., Miyashita, T., Kato, K. (1997) Expression patterns of Brx1 (Rieg gene), sonic hedgehog, Nkx2.2, Dlx1 and Arx during zona limitans intrathalamica and embryonic ventral lateral geniculate nuclear formation, *Mech. Dev.* **67**, 83–96.

Kitsukawa, T., Shimizu, M., Sanbo, M., Hirata, T., Taniguchi, M., Bekku, Y., Yagi, T., Fujisawa, H. (1997) Neuropilin-Semaphorin III/D-mediated chemorepulsive signals play a crucial role in peripheral nerve projection in mice, *Neuron* **19**, 995–1005.

Klingensmith, J., Ang, S.L., Bachiller, D., Rossant, J. (1999) Neural induction and patterning in the mouse in the absence of the node and its derivatives, *Dev. Biol.* **216**, 535–549.

Knoblich, J.A., Jan, L.Y., Jan, Y.N. (1995) Asymmetric segregation of Numb and Prospero during cell division, *Nature* **377**, 624–627.

Kobayashi, D., Kobayashi, M., Matsumoto, K., Ogura, T., Nakafuku, M., Shimamura, K. (2002) Early subdivisions in the neural plate define distinct competence for inductive signals, *Development* **129**, 83–93.

Krull, C.E., Lansford, R., Gale, N.W., Marcelle, C., Collazo, A., Yancopoulos, G.D., Fraser, S.E., Bronner-Fraser, M. (1997) Interactions of Eph-related receptors and ligands confer rostrocaudal pattern to trunk neural crest migration, *Curr. Biol.* **7**, 571–580.

Krumlauf, R. (1993) Hox genes and pattern formation in the branchial region of the vertebrate head, *Trends Genet.* **9**, 106–112.

Kuhn, T.B., Schmidt, M.F., Kater, S.B. (1995) Laminin and fibronectin guideposts signal sustained but opposite effects to passing growth cones, *Neuron* **14**, 275–285.

Lamb, T.M., Harland, R.M. (1995) Fibroblast growth factor is a direct neural inducer, which combined with noggin generates anterior-posterior neural pattern, *Development* **121**, 3627–3636.

Lance-Jones, C., Landmesser, L. (1980) Motor neuron projection patterns in chick hindlimb following partial reversals of the spinal cord, *J. Physiol.* **302**, 581–602.

Lavdas, A.A., Grigoriou, M., Pachnis, Y., Parnavelas, J.G. (1999) The medial ganglionic eminence gives rise to a population of early neurons in the developing cerebral cortex, *J. Neurosci.* **19**, 7881–7888.

Lewis, E.B. (1978) A gene complex controlling segmentation in *Drosophila*, *Nature* **276**, 565–570.

Liem, K.F. Jr., Tremml, G., Jessell, T.M. (1997) A role for the roof plate and its resident TGFbeta-related proteins in neuronal patterning in the dorsal spinal cord, *Cell* **91**, 127–138.

Liu, J.P., Laufer, E., Jessell, T.M. (2001) Assigning the positional identity of spinal motor neurons. Rostrocaudal patterning of Hox-c expression by FGFs, Gdf11, and retinoids, *Neuron* **32**, 997–1012.

Lois, C., Gracia-Verdugo, J.M., Alvarez-Buylla, A. (1996) Chain migration of neuronal precursors, *Science* **271**, 978–981.

Lufkin, T., Dierich, A., LeMeur, M., Mark, M., Chambon, P. (1991) Disruption of the Hox 1.6 homeobox gene results in defects in a region corresponding to its rostral domain of expression, *Cell* **66**, 1105–1119.

Lumsden, A. (1990) The cellular basis of segmentation in the developing hindbrain, *Trends Neurosci.* **13**, 329–335.

Lyckman, A.W., Jhaveri, S., Feldheim, D.A., Vanderhaeghen, P., Flanagan, J.G., Sur, M. (2001) Enhanced plasticity of retinothlamic projections in an ephrin-A2/A5 double mutant, *J. Neurosci.* **21**, 7684–7690.

Marshall, H., Nonchev, S., Sham, M.H., Muchamore, I., Lumsden, A., Krumlauf, R. (1992) Retinoic acid alters hindbrain Hox code and induces transformation of rhombomeres 2/3 into a 4/5 identity, *Nature* **360**, 737–741.

Marshall, H., Studer, M., Popperl, H., Aparicio, S., Kuroiwa, A., Brenner, S., Krumlauf, R. (1994) A conserved retinoic acid response element required for early expression of the homeobox gene *Hoxb-1*, *Nature* **370**, 567–571.

Matsunaga, E., Araki, I., Nakamura, H. (2000) Pax6 defines the di-mesencephalic boundary by repressing En1 and Pax2, *Development* **127**, 2357–2365.

Matsunami, H., Takeichi, M. (1995) Fetal brain subdivisions defined by R- and E-cadherin expressions: evidence for the role of cadherin activity in region-specific cell-cell adhesion, *Dev. Biol.* **172**, 466–478.

McGinnis, W., Krumlauf, R. (1992) Homeobox genes and axial patterning, *Cell* **68**, 283–302.

McKerracher, L., Chamoux, M., Arregui, C.O. (1996) Role of laminin and integrin interactions in growth cone guidance, *Mol. Neurobiol.* **12**, 95–116.

McMahon, A., Bradley, A. (1990) The Wnt-1 (int-1) proto-oncogene is required for development of a large region of mouse brain, *Cell* **62**, 1073–1085.

Mellitzer, G., Hallonet, M., Chen, L., Ang, S.L. (2002) Spatial and temporal 'knock down' of gene expression by electroporation of double-stranded RNA and morpholinos into early postimplantation mouse embryos, *Mech. Dev.* **118**, 57–63.

Mellitzer, G., Xu, Q., Wilkinson, D. (1999) Eph receptors and ephrins restrict cell intermingling and communication, *Nature* **400**, 77–81.

Millet, S., Campbell, K., Epstein, D.J., Losos, K., Harris, E., Joyner, A.L. (1999) A role for Gbx2 in repression of Otx2 and positioning the mid/hindbrain organizer, *Nature* **401**, 161–164.

Miyata, T., Kawaguchi, A., Okano, H., Ogawa, M. (2001) Asymmetric inheritance of radial glial fibers by cortical neurons, *Neuron* **31**, 727–741.

Momose, T., Tonegawa, A., Takeuchi, J., Ogawa, H., Umesono, K., Yasuda, K. (1999) Efficient targeting of gene expression in chick embryos by microelectroporation, *Dev. Growth Differ.* **41**, 335–344.

Morris, R.J., Tiveron, M.C., Xue, G.P. (1992) The relation of the expression and function of the neuronal glycoprotein Thy-1 to axonal growth, *Biochem. Soc. Trans.* **20**, 401–405.

Muramatsu, T., Mizutani, Y., Ohmori, Y., Okumura, J.-I. (1997) Comparison of three non-viral transfection methods for foreign gene expression in early chicken embryos *in ovo*, *Biochem. Biophys. Res. Commun.* **230**, 376–380.

Myat, A., Henry, P., McCabe, V., Flintoft, L., Rotin, D., Tear, G. (2002) Drosophila Nedd4, a ubiquitin ligase, is recruited by commissureless to control cell surface levels of the roundabout receptor, *Neuron* **35**, 447–459.

Nakagawa, Y., O'Leary, D.D. (2001) Combinatorial expression patterns of LIM-homeodomain and other regulatory genes parcellate developing thalamus, *J. Neurosci.* **21**, 2711–2725.

Nakagawa, S., Takeichi, M. (1995) Neural crest cell-cell adhesion controlled by sequential and subpopulation-specific expression of novel cadherin, *Development* **121**, 1321–1332.

Nakagawa, S., Takeichi, M. (1998) Neural crest emigration from the neural tube depends on regulated cadherin expression, *Development* **125**, 2963–2971.

Nieuwkoop, P. (1952) Activation and organisation of the central nervous system in amphibians, *J. Exp. Zool.* **120**, 1–108.

Nose, A., Nagafuchi, A., Takeichi, M. (1988) Expressed recombinant cadherins mediate cell sorting in model systems, *Cell* **54**, 993–1001.

Nothias, F., Vernier, P., von Boxberg, Y., Mirman, S., Vincent, J.D. (1997) Modulation of NCAM polysialylation is associated with morphofunctional modifications in the hypothalamo-neurohypophysial system during lactation, *Eur. J. Neurosci.* **9**, 1552–15565.

Ogawa, M., Miyata, T., Nakajima, K., Yagyu, K., Seike, M., Ikenaka, K., Yamamoto, H., Mikoshiba, K. (1995) The reeler gene-associated antigen on Cajal-Retzius neurons is a crucial molecule for laminar organization of cortical neurons, *Neuron* **14**, 899–912.

Palmiter, R.D., Brinster, R.L. (1986) Germ-line transformation of mice, *Annu. Rev. Genet.* **20**, 465–499.

Panin, V.M., Papayannopoulos, V., Wilson, R., Irvine, K.D. (1997) Fringe modulates Notch-ligand interactions, *Nature* **387**, 908–912.

Peifer, M., Bejsovec, A. (1992) Knowing your neighbors: cell interactions determine intrasegmental patterning in *Drosophila*, *Trends Genet.* **8**, 243–249.

Polleux, F., Giger, R.J., Ginty, D.D., Kolodkin, A.L., Ghosh, A. (1998) Patterning of cortical efferent projections by Semaphorin-Neuropilin interactions, *Science* **282**, 1904–1906.

Polleux, F., Morrow, T., Ghosh, A. (2000) Semaphorin 3A is a chemoattractant for cortical apical dendrites, *Nature* **404**, 567–573.

Price, S.R., De Marco Gracia, N.V., Ranscht, B., Jessell, T.M. (2002) Regulation of motor neuron pool sorting by differential expression of type II cadherins, *Cell* **109**, 205–216.

Rakic, P. (1974) Neurons in rhesus monkey visual cortex: systematic relation between time origin and eventual disposition, *Science* **183**, 425–427.

Rebay, I., Fehon, R.G., Artavanis-Tsakonas, S. (1993) Specific truncations of *Drosophila* Notch define dominant activated and dominant negative forms of the receptor, *Cell* **74**, 319–329.

Redies, C., Engelhart, K., Takeichi, M. (1993) Differential expression of N- and R-cadherin in functional neuronal systems and other structures of the developing chicken brain, *J. Comp. Neurol.* **333**, 398–416.

Redies, C., Takeichi, M. (1996) Cadherins in the developing central nervous system: an adhesive code for segmental and functional subdivisions, *Dev. Biol.* **180**, 413–423.

Rhinn, M., Dierich, A., LeMeur, M., Ang, S.L. (1999) Cell autonomous and non-cell autonomous functions of Otx2 in patterning the rostral brain., *Development* **126**, 4295–4304.

Rietze, R.L., Valcanis, H., Brooker, G.F., Thomas, T., Voss, A.K., Bartlett, P.F. (2001) Purification of a pluripotent neural stem cell from the adult mouse brain, *Nature* **412**, 736–739.

Roelink, H., Augsburger, A., Heemskerk, J., Korzh, V., Norlin, S., Ruiz i Altaba, A., Tanabe, Y., Placzek, M., Edlund, T., Jessell, T.M., Dodd, J. (1994) Floor plate and motor neuron induction by *vhh-1*, a vertebrate homolog of *hedgehog* expressed by the notochord, *Cell* **76**, 761–775.

Roelink, H., Porter, J.A., Chiang, C., Tanabe, Y., Chang, D.T., Beachy, P.A., Jessell, T.M. (1995) Floor plate and motor neuron induction by different concentrations of the amino-terminal cleavage product of Sonic hedgehog autoproteolysis, *Cell* **81**, 445–455.

Shen, Q., Zhong, W., Jan, Y.N., Temple, S. (2002) Asymmetric Numb distribution is critical for asymmetric cell division of mouse cerebral cortical stem cells and neuroblasts, *Development* **129**, 4843–4853.

Shimamura, K., Hartigan, D.J., Martinez, S., Puelles, L., Rubenstein, J.L.R. (1995) Longitudinal organization of the anterior neural plate neural tube, *Development* **121**, 3923–3933.

Shimamura, K., Rubenstein, J.L.R. (1997) Inductive interactions direct early regionalization of the mouse forebrain, *Development* **124**, 2709–2718.

Smith, A., Robinson, V., Patel, K., Wilkinson, D.G. (1997) The EphA4 and EphB1 receptor tyrosine kinases and ephrin-B2 ligand regulate targeted migration of branchial neural crest cells, *Curr. Biol.* **7**, 561–570.

Sobue, K., Kanda, K. (1989) Alpha-actinins, calspectin (brain spectrin or fodrin), and actin participate in adhesion and movement of growth cones, *Neuron* **3**, 311–319.

Song, H., Ming, G., He, Z., Lehmann, M., McKerracher, L., Tessier-Lavigne, M., Poo, M. (1998) Conversion of neuronal growth cone responses from repulsion to attraction by cyclic nucleotides, *Science* **281**, 1515–1518.

Stanojevic, D., Small, S., Levine, M. (1991) Regulators of a segmentation stripe by

overlapping activators and repressors in the *Drosophila* embryo, *Science* **254**, 1385–1387.

Stein, E., Tessier-Lavigne, M. (2001) Hierarchical organization of guidance receptors: silencing of netrin attraction by slit through a Robo/DCC complex, *Science* **291**, 1928–1938.

Stoeckli, E.T., Sonderegger, P., Pollerberg, G.E., Landmesser, L.T. (1997) Interference with axonin-1 and NrCAM interactions unmasks a floor-plate activity inhibitory for commissural axons, *Neuron* **18**, 209–221.

Stoykova, A., Gotz, M., Gruss, P., Price, J. (1997) Pax6-dependent regulation of adhesive patterning, R-cadherin expression and boundary formation in developing forebrain, *Development* **124**, 3765–3777.

Streit, A., Berliner, A.J., Papanayotou, C., Sirulnik, A., Stern, C.D. (2000) Initiation of neural induction by FGF signalling before gastrulation, *Nature* **406**, 74–78.

Struhl, G., Struhl, K., Macdonald, P.M. (1989) The gradient morphogen bicoid is a concentration-dependent transcriptional activator, *Cell* **57**, 1259–1273.

Studer, M., Lumsden, A., Ariza-McNaughton, L., Bradley, A., Krumlauf, R. (1996) Altered segmental identity and abnormal migration of motor neurons in mice lacking *Hoxb-1*, *Nature* **384**, 630–635.

Suzuki, S.C., Inoue, T., Kimura, Y., Tanaka, T. (1997) Neural circuits are subdivided by differential expression of type-II classic cadherins in neonatal mouse brains, *Mol. Cell. Neurosci.* **9**, 433–447.

Tam, P.P., Steiner, K.A., Zhou, S.X., Quinlan, G.A. (1997) Lineage and functional analyses of the mouse organizer, *Cold Spring Harb. Symp. Quant. Biol.* **62**, 135–144.

Tam, P.P., Steiner, K.A. (1999) Anterior patterning by synergistic activity of the early gastrula organizer and the anterior germ layer tissues of the mouse embryo, *Development* **126**, 5171–5179.

Tamamaki, N., Fujimori, K.E., Takauji, R. (1997) Origin and route of tangentially migrating neurons in the developing neocortical intermediate zone, *J. Neurosci.* **17**, 8313–8323.

Tamkun, J.W., DeSimone, D.W., Fonda, D., Patel, R.S., Buck, C., Horwitz, A.F., Hynes, R.O. (1986) Structure of integrin, a glycoprotein involved in the transmembrane linkage between fibronectin and actin, *Cell* **46**, 271–282.

Tan, S.S., Morriss-Kay, G. (1985) The development and distribution of the cranial neural crest in the rat embryo, *Cell Tissue Res.* **240**, 403–416.

Thomas, K.R., Capecchi, M.R. (1987) Site-directed mutagenesis by gene targetting in mouse embryo-derived stem cells, *Cell* **51**, 503–512.

Tiret, L., Le Mouellic, H., Maury, M., Brulet, P. (1998) Increased apoptosis of motoneurons and altered somatotopic maps in the branchial spinal cord of *Hoxc8*-deficient mice, *Development* **125**, 279–291.

Toresson, H., Potter, S.S., Campbell, K. (2000) Genetic control of dorsal-ventral identity in the telencephalon: opposing roles for Pax6 and Gsh2, *Development* **127**, 4361–4371.

Treubert-Zimmermann, U., Heyers, D., Redies, C. (2002) Targeting axons to specific fiber tracts in vivo by altering cadherin expression, *J. Neurosci.* **22**, 7617–7626.

Uemura, T., Shepherd, S., Ackerman, L., Jan, L.Y., Jan, Y.N. (1989) numb, a gene required in determination of cell fate during sensory organ formation in *Drosophila* embryos, *Cell* **58**, 349–360.

van Straaten, H.W., Hekking, J.W., Wiertz-Hoessels, E.J., Thors, F., Drukker, J. (1988) Effect of the notochord on the differentiation of a floor plate area in the neural tube of the chick embryo, *Anat. Embryol.* **177**, 317–324.

Wahl, S., Barth, H., Ciossek, T., Aktories, K., Mueller, B.K. (2000) Ephrin-A5 induces collapse of growth cones by activating Rho and Rho kinase, *J. Cell Biol.* **149**, 263–270.

Wakamatsu, Y., Maynard, T.M., Weston, J.A. (2000) Fate determination of neural crest cells by NOTCH-mediated lateral inhibition and asymmetrical cell division during gangliogenesis, *Development* **127**, 2811–2821.

Wang, H.U., Anderson, D.J. (1997) Eph family transmembrane ligands can mediate repulsive guidance of trunk neural crest migration and motor axon outgrowth, *Neuron* **18**, 383–396.

Wargelius, A., Ellingsen, S., Fjose, A. (1999) Double-stranded RNA induces specific developmental defects in zebrafish embryos, *Biochem. Biophys. Res. Commun.* **263**, 156–161.

Wianny, F., Zernicka-Goetz, M. (2000) Specific interference with gene function by double-stranded RNA in early mouse development, *Nat. Cell Biol.* **2**, 70–75.

Wightman, B., Clark, S.G., Taskar, A.M., Forrester, W.C., Maricg, A.V., Bargmann, C.I.,

Garriga, G. (1996) The *C. elegans* gene vab-8 guides posteriorly directed axon outgrowth and cell migration, *Development* **122**, 671–682.

Wilson, S., Rydstrom, A., Trimborn, T., Willert, K., Nusse, R., Jessell, T.M., Edlund, T. (2001) The status of Wnt signalling regulates neural and epidermal fates in the chick embryo, *Nature* **411**, 325–330.

Wizenmann, A., Lumsden, A. (1997) Segregation of rhombomeres by differential chemoaffinity, *Mol. Cell. Neurosci.* **9**, 448–459.

Wolf, F.W., Hung, M.S., Wightman, B., Way, J., Garriga, G. (1998) vab-8 is a key regulator of posteriorly directed migrations in *C. elegans* and encodes a novel protein with kinesin motor similarity, *Neuron* **20**, 655–666.

Wong, K., Park, H.T., Wu, J.Y., Rao, Y. (2002) Slit proteins: molecular guidance cues for cells ranging from neurons to leukocytes, *Curr. Opin. Genet. Dev.* **12**, 583–591.

Xu, Q., Mellitzer, G., Robinson, V., Wilkinson, D. (1999) *In vivo* cell sorting in complementary segmental domains mediated by *Eph* receptors and *ephrins*, *Nature* **399**, 267–271.

Yamada, T., Placzek, M., Tanaka, H., Dodd, J., Jessell, T.M. (1991) Control of cell pattern in the developing nervous system: polarizing activity of the floor plate and notochord, *Cell* **64**, 635–647.

Ye, W., Bouchard, M., Stone, D., Liu, X., Vella, F., Lee, J., Nakamura, H., Ang, S.L., Busslinger, M., Rosenthal, A. (2001) Distinct regulators control the expression of the mid-hindbrain organizer signal FGF8, *Nat. Neurosci.* **4**, 1175–1181.

Yuasa, J., Hirano, S., Yamagata, M., Noda, M. (1996) Visual projection map specified by topographic expression of transcription factors in the retina, *Nature* **382**, 632–635.

Calcium Biochemistry

Joachim Krebs
Institute of Biochemistry, Swiss Federal Institute of Technology, ETH-Hoenggerberg, Zuerich, Switzerland

1 Introduction 135

2 Ca^{2+} Ligation 138

3 Calcium in the Extracellular Space 138
3.1 The Calcium Binding EGF-domain 139
3.2 The Extracellular Calcium Sensor 140

4 Signal Transduction Principles 140
4.1 General Properties 140
4.2 Calcium Oscillations in the Cell 141
4.3 Calcium and the Cell Cycle 142
4.4 Calcium and Apoptosis 142

5 Intracellular Calcium Binding Proteins 144
5.1 The EF-hand Protein Family 144
5.1.1 Calmodulin and its Targets 146
5.1.2 Neuronal Calcium Sensors 152
5.1.3 S100 Proteins 152
5.1.4 The Penta EF-hand Family 153
5.2 Non-EF-hand Ca^{2+} Binding Proteins 155
5.2.1 Annexins 155
5.2.2 Gelsolin 156
5.2.3 C_2-domains 156

6 Systems Controlling Intracellular Ca^{2+} Concentration: Structural and Functional Properties 156
6.1 Calcium Transport Systems of the Plasma Membrane 157

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

6.1.1 The Calcium Channel 157
6.1.2 The Ca^{2+} Pump 158
6.1.3 Na^{+}/Ca^{2+} Exchanger 161
6.2 Ca^{2+} Transport Systems of the Reticulum 161
6.2.1 The IP_3 Receptor 161
6.2.2 The Ca^{2+}-release Channel 162
6.2.3 The Ca^{2+} Pump 163
6.3 The Mitochondrial Calcium Transporting Systems 165
6.4 Calcium in the Nucleus 165

7 Role of Calcium in Specialized Tissues 166
7.1 Calcium and Fertilization 166
7.2 Calcium in Calcifying Tissues 166
7.3 The Role of Calcium in Plants 167
7.3.1 Touch Genes 167
7.3.2 Calmodulin-like Domain Protein Kinase (CDPK) 167

8 Calcium and Disease 168

9 Conclusions 168

Bibliography 169
Books and Reviews 169
Primary Literature 169

Keywords

Calcium
The fifth most abundant element of the human body; it has a stabilizing function in shells, bones, and teeth (bound as hydroxyapatite), and in proteins in the extracellular fluid. Inside cells, calcium functions as a second messenger by interacting with specific Ca^{2+} binding proteins (e.g. calmodulin).

Calmodulin
A highly conserved, intracellular Ca^{2+} binding sensor protein involved in triggering Ca^{2+} signals. It contains four helix-loop-helix or EF-hand Ca^{2+} binding domains.

Channel
Integral membrane proteins responsible for transport across membranes; opening and closing (i.e. gating) can be voltage-, ligand-, or store-operated.

EF-hand Proteins
The term coined by R.H. Kretsinger to describe the helix-loop-helix calcium binding domains of specific proteins. The highly conserved motif (first described on the basis of

the crystal structure of parvalbumin containing six helices, A–F) in which certain amino acids are invariant consists of two helices enclosing the Ca^{2+} binding loop. As a model, the forefinger and the thumb of the right hand can resemble the two helices (e.g. E and F of the second Ca^{2+} binding domain of parvalbumin), and the bent midfinger, the enclosed loop, hence the EF-hand.

Exchanger
A transmembrane protein using the downhill gradient of one ion as an energy source to transport another ion against its gradient across the membrane (e.g. Na^{+}/Ca^{2+}).

Pump
A transmembrane protein using ATP or another nucleotide triphosphate as an energy source to transport an ion against its gradient across the membrane.

Second Messenger
Intracellular small molecules (e.g. cyclic nucleotides or inositol polyphosphates) or ions such as Ca^{2+}, indispensable for the transduction of signals converting extracellular stimuli (e.g. hormones = primary messengers) into intracellular responses.

Calcium is of pivotal importance for many biological processes. It may have a rather static, structure-stabilizing role, or it may participate as one of the second messengers of the cell in signal transduction pathways, fulfilling a more dynamic function. This is made possible by some specific properties of the Ca^{2+} ion (e.g. high dehydration rate, great flexibility in coordinating ligands, largely irregular geometry of the coordination sphere). The control of calcium homeostasis is of central importance to the organism, involving an exchange of the mineral between the skeleton (as the major calcium reservoir), the intestine, and the kidney (as the organs of calcium absorption or reuptake), from the extracellular fluid (ECF) and intracellular calcium, respectively. This highly integrated process consists of a number of hormonally controlled feedback loops and an elaborate system of channels, exchanger, and pumps to control Ca^{2+} fluxes into and out of cells. This article describes the different roles of calcium in the regulation of biological functions and the proteins involved in these processes.

1 Introduction

Calcium is one of the most common elements on earth, and it is the fifth most abundant element of the human body. Next to its central role in cellular functions as one of the second messengers, calcium is a major constituent of the skeleton. It has a stabilizing function in shells, bones, and teeth and therefore is an old component of organisms, the record

being 2×10^9 years (blue–green algae). In minerals as well as in solution, calcium occurs predominantly in a complex form, mostly as calcium phosphate (e.g. hydroxyapatite [$Ca_{10}(PO_4)_6(OH)_2$]), which makes up 60% of the weight of human skeleton, that is, the skeleton of a man contains 1.0 to 1.3 kg of calcium, that of a woman 0.8 to 0.9 kg, comprising 99% of the total calcium of a human body. Compared to this amount, the calcium found in the extracellular fluid and intracellularly in the cytosol or in other intracellular compartments is almost negligible. In the ECF or in the lumen of intracellular reticular systems, the calcium concentration is millimolar (2–5 mM, of which about 50% is unbound, i.e. corresponds to free Ca^{2+} concentration), whereas the cytosolic free Ca^{2+} concentration of a resting cell is about 100 to 300 nM. This results in a steep concentration gradient of ionized Ca^{2+} across cellular membranes, which is regulated by a variety of channels, pumps, and other transporting systems controlling the fluxes of Ca^{2+} into and out of the cell and between the various intracellular compartments. On the other hand, calcium homeostasis of the ECF is maintained through a highly integrated and complex endocrine system. This involves the interplay between a receptor sensing the level of Ca^{2+} in ECF (see below) and two antagonistic polypeptide hormones, parathyroid hormone (PTH) and calcitonin (CT), and a Vitamin D metabolite, 1,25 $(OH)_2D_3$ (see Fig. 1). They measure the level of calcium in the ECF and regulate its flow into and out of the ECF by acting on target cells of the intestine, kidney, and bone. Normally, PTH prevents calcium of the ECF from falling below a threshold level, whereas CT prevents abnormal increases of serum calcium. PTH is also responsible for the formation of 1,25 $(OH)_2D_3$, which acts on specific receptors in the intestine to promote absorption of Ca^{2+}.

At the end of the nineteenth century, the English physiologist Sidney Ringer discovered that calcium plays an essential role in regulating the contraction of the heart, but for decades, calcium was considered to be a specific factor merely important for muscle contraction. Evidently, the general implication of Ringer's observation came ahead of its time. It needed the formulation of the "second messenger" concept, which was developed by Earl Sutherland on the basis of the discovery of cyclic AMP as an intracellular regulatory constituent. It was supported by the work of Ebashi et al. who showed how Ca^{2+} works in skeletal muscle, and identified troponin C as the first member of the EF-hand type calcium binding proteins. However, it was not until the discovery of calmodulin by Cheung and Kakiuchi in the sixties of the twentieth century to realize the pivotal role of Ca^{2+} as a general regulator of cellular functions. This is made possible due to the tightly controlled steep concentration gradient of ionized Ca^{2+} across cellular membranes. If a cell becomes activated because of an external signal, this often results in up to a 100-fold rise in the intracellular free Ca^{2+} concentration owing to the uptake of extracellular Ca^{2+} and/or the release of Ca^{2+} from intracellular stores. These changes in the free Ca^{2+} concentration can cause significant oscillations of Ca^{2+} in the cytosol providing the possibility of signal transduction for a number of different cellular activities (e.g. muscle contraction; glycogen metabolism; fertilization; cell growth, division, and apoptosis; excitation–contraction and stimulus-secretion coupling; mineralization). Many of these functions are accomplished through the interaction of Ca^{2+} with specific proteins, resulting in specific modulations of

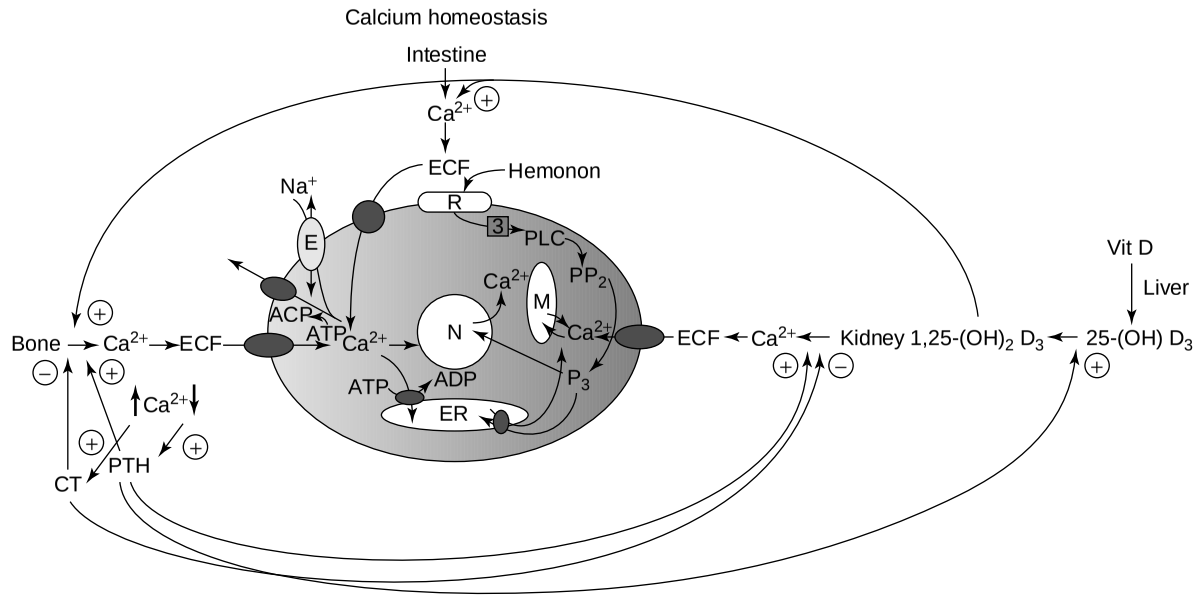

Fig. 1 Schematic view of the different systems participating in calcium homeostasis. The figure integrates the different organs, hormones, and the cellular systems involved in the regulation of calcium. CT, calcitonin; PTH, parathyroid hormone; ECF, extracellular fluid; ER, endoplasmic reticulum; M, mitochondrion; N, nucleus; R, receptor; C, channel; E, exchanger; P, pump; G, G protein; PLC, phospholipase C; PIP2, phosphatidylinositol-4,5-diphosphate; IP3, inositol-1,4,5-triphosphate.

protein–protein interactions owing to conformational changes of the Ca^{2+} receptors.

2
Ca^{2+} Ligation

Ca^{2+} ligation usually occurs via carboxylates (mono- or bidentate) or neutral oxygen donors since the calcium ion overwhelmingly prefers oxygen as donor groups of ligands. Owing to the great flexibility of calcium in coordination (coordination numbers usually 6–8, but up to 12 are possible) and a largely irregular geometry, both in bond length and in bond angles, calcium has a superior binding property to proteins as compared to Mg^{2+}, which requires a fixed geometry of an octahedron with six coordinating oxygen atoms and a fixed ionic bond distance due to the smaller ionic radius (0.64 Å) as compared to 0.97 Å of calcium. This difference in flexibility of complexation geometry permits calcium a greater versatility in coordinating ligands, leading to a higher exchange rate. This is reflected in a three-orders-of-magnitude difference in dehydration rate between the two ions, which makes Ca^{2+} much more suitable as a signal-transducing factor, especially in the presence of a high excess of Mg^{2+}, the concentration of which is in the mM range on both sides of the cellular membrane.

3
Calcium in the Extracellular Space

It was stated in the introduction that calcium in the ECF is tightly controlled to maintain its concentration in a range of 2 to 5 mM. One of the important functions of Ca^{2+} in the ECF is to stabilize the structure of proteins and to mediate cell–cell or cell–matrix interactions. As a consequence, there are a number of important extracellular processes that require calcium, such as blood clotting, complement activation, or interaction of cell-surface receptors with their ligands (e.g. Notch signaling). One important difference in comparing the Ca^{2+} binding sites of extracellular and intracellular proteins (e.g. EF-hand proteins, see below) is the spatial arrangement of the ligating groups. As it will be discussed in detail below, the intracellular Ca^{2+} binding proteins show a sequential arrangement of their ligating residues in contrast to the extracellular proteins in which these ligands are usually located at distant positions in the amino acid sequence (e.g. α-lactalbumin). Therefore, those proteins have a preformed cavity of Ca^{2+} binding sites with a relative high degree of rigidity with the consequence that the on-rate for Ca^{2+} binding is fairly slow. Since, on the other hand, the off-rate is relatively fast, the affinity of Ca^{2+} for the extracellular proteins is usually low. However, owing to the relative high concentration of Ca^{2+} in the ECF, these proteins occur in their Ca^{2+} bound form and are therefore protected against proteolytic cleavage.

Another important function of Ca^{2+} in the extracellular space is in mediating cell–cell adhesion. Essential for these cellular contacts are proteins called *adherins*. These transmembrane glycoproteins require calcium for their cell-binding activity. One of the best-studied molecules is uvomorulin or E-cadherin. The extracellular part of this protein is largely composed of three repeating domains, each containing two putative Ca^{2+} binding sites different from the later-described EF-hand binding

sites, but rich in aspartic acid and asparagine residues. Recently, it was shown that replacing a single aspartic acid residue by either lysine or alanine in one of these sites renders the protein more susceptible to protein degradation due to the loss of Ca^{2+} binding and abolishes the adhesive function of the molecule.

Bone-forming cells, osteoblasts, synthesize and secrete a number of noncollagenous Ca^{2+} binding proteins such as osteocalcin, osteopontin, and osteonectin that bind to bone minerals such as hydroxyapatite in a calcium-dependent manner. Osteocalcin belongs to the class of Ca^{2+} binding proteins rich in the unusual amino acid γ-carboxyl glutamic acid mediating the Ca^{2+} binding property, whereas osteopontin, a glycoprotein, is rich in sequences of aspartic or glutamic acid residues responsible for Ca^{2+} binding. In addition, both proteins contain arginine-glycine-aspartic acid sequence domains known to be important in mediating binding to cell-surface receptors. Osteonectin, on the other hand, is an extracellular Ca^{2+} binding protein that can contain at least one EF-hand type of high-affinity Ca^{2+} binding sites that are normally typical for *intra*cellular Ca^{2+} binding proteins. One interesting abnormality of this EF-hand is the fact that it is stabilized by an S–S bridge that normally does not occur in EF-hand type proteins (see below).

Similar to the intracellular Ca^{2+} binding proteins that are composed of EF-hand domains (see below), most extracellular proteins are assembled from a limited number of domain structures or modules. This is well documented for the calcium binding domains of a number of extracellular proteins like the γ-carboxyl glutamic acid–rich Gla-domain in the vitamin K–dependent clotting factors, the cadherin module, or the epidermal growth factor (EGF)-like domain. The latter is one of the best-studied examples and should be described in more detail.

3.1 The Calcium Binding EGF-domain

The epidermal growth factor or EGF-like domain is one of the most common modules identified in extracellular proteins. It plays a general role in cell–cell adhesion, blood coagulation, and receptor–ligand interaction. EGF modules contain about 40 to 45 amino acids including 6 cysteine residues that normally build S–S disulfide bond bridges. A distinct subset of EGF-domains containing a calcium binding domain (cbEGF) has been identified and its structure studied in detail. A consensus sequence associated with calcium binding has been identified as D/N-x-D/N-E/Q-y-D/N-y-Y/F, where x symbolizes a variable amino acid, y a variable sequence of amino acids, and D/N, E/Q, and Y/F represent the usual one-letter codes for aspartic acid/asparagine, glutamic acid/glutamine, or tyrosine/phenylalanine respectively.

CbEGF domains have been identified in a variety of proteins such as different coagulation factors, the low-density lipoprotein receptor (LDLR), the cell-surface signaling receptor Notch or the extracellular matrix protein fibrillin-1. In most of these proteins, the cbEGF domains exist as multiple tandem repeats. In a number of proteins, mutations within cbEGF domains have been linked to genetic diseases affecting blood clotting owing to factor deficiency, familial hypercholesterolaemia owing to mutations in the LDLR, inherited forms of cerebrovascular disorder such as CADASIL owing to mutations or deletions in the Notch signaling receptor, or the

Marfan syndrome owing to mutations disrupting the pair-wise interaction of cbEGF domains in fibrillin-1.

3.2 The Extracellular Calcium Sensor

As stated before, calcium in the extracellular fluid is tightly controlled because of a strict regulation by different hormones keeping the level of calcium concentration within a narrow window (see Fig. 1). Receptors of the different hormones, such as PTH, CT and Vitamin D_3, involved in this regulation have been identified, but the molecular basis for sensing the extracellular calcium level was poorly understood. On the basis of several observations, however, it was postulated that a cell-surface Ca^{2+}-sensing receptor might exist with functional similarities to the G-protein–coupled receptors of the Ca^{2+}-mobilizing hormones. In 1993, Brown et al. reported the successful cloning and characterization of a 500 kDa protein from parathyroid glands with a large, extracellular N-terminal region and a smaller, cytosolic C-terminal domain that could be activated by mM Ca^{2+}, corresponding to the concentration of extracellular Ca^{2+}. Since the predicted amino acid sequence of the protein contained seven putative transmembrane domains, it was suggested that this calcium receptor (CaR) functionally belongs to the well-characterized group of G protein-coupled receptors. The CaR is a glycoprotein that shares some homologies with glutamate receptors. The extracellular domain contains low-affinity Ca^{2+} binding sites characterized by clusters of negatively charged amino acid residues, typical for extracellular, Ca^{2+}-sensing proteins. CaR can be found in most cells throughout the body, but especially in cells of those organs that mainly control the level of calcium in the extracellular fluid, that is, in bone-forming and -resorbing cells, in parathyroid glands, and in cells of kidney and intestine. Long before the discovery of CaR, clinical observations of inheritable hypo- and hypercalcemic syndromes have been described that could now be linked to specific mutations of CaR, most of them in the extracellular calcium-sensing domain.

4 Signal Transduction Principles

4.1 General Properties

Cell-surface receptors can recognize extracellular signals ("primary messengers") and multiply them into a cascade of intracellular events using a limited number of intracellular signal transducers, so-called "second messengers." The most common intracellular signal multiplication system makes use of the phosphorylation/dephosphorylation of proteins or enzymes by activating a bunch of different kinases/phosphatases. To date, five classes of intracellular messengers are known:

1. cyclic nucleotides [cyclic adenosine 3′–5′ monophosphate (cyclic AMP); cyclic guanosine 3′–5′ monophosphate (cyclic GMP); cyclic ADP-ribose];
2. derivatives of phosphatidylinositol [inositolpolyphosphates, e.g. inositol-1,4,5-triphosphate (IP3), inositol-1,3,4,5-tetraphosphate (IP4), and diacylglycerol (DAG), stemming from the same precursor phosphatidylinositol-4,5- diphosphate (PIP2);
3. free Ca^{2+} ions;
4. gases such as nitric oxide (NO), probably also carbon monoxide (CO);
5. nicotinic acid adenine dinucleotide phosphate (NAADP).

The primary event in all signal-transducing pathways is the reception of an external signal by a specific receptor in the cellular membrane, activating a chain of reactions that will finally result in an intracellular response, that is, the cellular or plasma membrane is the main information barrier.

Ca^{2+} performs its second messenger function owing to changes in the distribution at the two sides of the membrane, which modulates its messenger function, in contrast to the metabolic synthesis and degradation of the other second messenger molecules. If the intracellular Ca^{2+} rises owing to the opening of Ca^{2+} channels upon receiving an extracellular signal, the ion will bind to specific proteins. This event will result in a conformational change enabling the triggering device to multiply the incoming signal. The occurrence of such Ca^{2+} binding regulatory proteins was first described in muscle tissues (i.e. troponin C) as mentioned before. The concept was later corroborated because of the detection of a variety of homologous proteins, as described in detail in Sect. 5.

The different second messengers are central components of intracellular control mechanisms. They are connected in their action through a complex network of feedback mechanisms. In addition, the interaction of polypeptide growth factors with their receptors is connected with signal transduction pathways. Thus, one of the earliest responses upon binding of growth factors to their receptors is a rapid increase in cytosolic free Ca^{2+} owing to receptor-activated, membrane-bound phospholipase C. This enzyme specifically cleaves phosphatidylinositol-4,5-diphosphate (PIP2) to release IP_3 and DAG, the former releasing Ca^{2+} from intracellular stores, whereas DAG activates the Ca^{2+}- and phospholipid-dependent protein kinase C.

4.2 Calcium Oscillations in the Cell

In 1986, Cobbold et al. made a striking observation: by monitoring the cytosolic calcium level of individual hepatocytes, they noticed that upon addition of vasopressin, a hormone known to mobilize intracellular calcium, the intracellular Ca^{2+} did not rise to a sustained level, but instead increased and decreased repetitively with a certain frequency (calcium spikes). They further observed that it was the frequency and not the amplitude of spiking that increased proportional to the rise in concentration of the added hormone, that is, an extracellular *analog* signal was converted into an intracellular *digital* signal. This seminal observation was soon confirmed in other laboratories on a variety of different cells indicating that calcium "spikes" and, in addition, calcium "waves," the spatial counterpart of calcium spiking, are fundamentally important processes in signal transduction pathways. Two models describing these phenomena have been recently developed:

1. The IP_3–Ca^{2+} crosscoupling (ICC) model by Meyer & Stryer
2. The calcium-induced calcium-release (CICR) model by Berridge

Pivotal to both systems are positive feedback, cooperativity, deactivation, and reactivation of the process. The key element of the model described by Meyer & Stryer is the mutual reinforcement of the IP_3-induced Ca^{2+} release and the Ca^{2+}-stimulated IP_3 formation leading to a crosscoupling of two oscillating second messenger systems. By contrast, the

key assumption of the model proposed by Berridge et al. is the existence of two independent cytosolic Ca^{2+} stores: one that influences the cytosolic Ca^{2+} level owing to the IP_3-induced Ca^{2+} release, and the other that contains the Ca^{2+}-induced Ca^{2+} release channel. In this model, receptor activation leads to the formation of IP_3, which releases Ca^{2+} from IP_3-sensitive Ca^{2+} stores for inducing a constant efflux of calcium. The latter triggers Ca^{2+} from Ca^{2+}-sensitive stores, leading to an autocatalytic process in which the spiking characteristics depend on the frequency of the depletion and refilling of the Ca^{2+}-sensitive Ca^{2+} stores. Owing to the distribution of the storage systems over much of the cellular volume, calcium spiking leads to calcium waves (propagation rate 10–100 $\mu m\ s^{-1}$) as observed for many cells. Since the diffusion rate of IP_3 is 100 times faster than that of Ca^{2+}, calcium stores have to be much closer in space for the CICR model as compared to the ICC model to generate efficient calcium waves across the cell.

4.3 Calcium and the Cell Cycle

Calcium is one of the control elements of the cell division cycle (cdc) during proliferation of cells. Usually, there are three control points in a cell cycle: START, mitosis ENTRY, and mitosis EXIT. At all three points, a transient rise in the intracellular Ca^{2+} can be observed. The probable mediator of these Ca^{2+}-dependent cell cycle controls is calmodulin, a multifunctional intracellular calcium binding protein, as described later. It has been shown that overexpression of calmodulin leads to a reduction of the G1-phase in the cell cycle, whereas reduced calmodulin synthesis leads to arrest in G1 and mitosis EXIT. Of special interest is the observation that CaM-kinase activity is required for entry into mitosis. The relevant target for CaM kinase is a tyrosine phosphatase, known as cdc25 in yeast. This enzyme is activated because of phosphorylation by CaM kinase that enables it to dephosphorylate $pp34^{cdc2}$, a kinase involved in nuclear envelope breakdown, one of the key events during the cell cycle.

The essential role of calmodulin during cell cycle progression was demonstrated by Davis et al. for yeast. Evidence was provided that calmodulin was required for chromosome segregation and the polarized growth of bud formation. Interestingly, Davis identified two calmodulin targets that are important for chromosomal segregation, the spindle pole body protein SPC110 and the nonconventional myosin Myo2P. A mammalian homolog of SPC110 has recently been identified as a calmodulin target, supporting the earlier findings that calmodulin is associated with centrosomes during mitosis.

4.4 Calcium and Apoptosis

Apoptosis, or programmed cell death (PCD), is a process during which cells shrink and dissociate from their surrounding neighbors, their organelles retain in size, and in the nucleus chromatin forms dense aggregates on the nuclear membrane. Eventually, chromatin undergoes fragmentation, and the formation of membrane-enclosed structures termed *apoptotic bodies* that are extruded into the extracellular environment can be observed. On the other hand, necrotic cells swell, their mitochondria enlarge, the plasma membranes disrupture, but the nuclear changes are marginal.

PCD is an important mechanism during development that has emerged in multicellular organisms to remove unnecessary, damaged, or aged cells. Therefore, abnormal resistance toward apoptosis may lead to autoimmune diseases or cancer, whereas uncontrolled enhancement of apoptotic processes could favor chronic pathologies such as neurodegenerative diseases (e.g. dementia of the Alzheimer type) or immune deficiencies such as AIDS.

A number of highly conserved proteins, called *death proteins*, are involved in the complex apoptotic pathways by either causing or preventing cell death. This was mainly due to detailed studies of the nematode *Caenorhabditis elegans*. During development, 131 of the total 1090 cells of *C. elegans* are predestined to die. This apoptotic process is controlled by the coordinated expression of the death proteins ced-3, ced-4, and ced-9, some of which belong to the highly conserved family of cysteine proteases, called *caspases*, others to oncogenes such as bcl-2.

Work in a number of different labs has demonstrated that alterations in intracellular Ca^{2+} homeostasis are commonly involved in promoting apoptosis. One aspect of the function of bcl-2 involves preventing these alterations. So it has been demonstrated that sustained levels of Ca^{2+} in mitochondria can be significantly lowered by the expression of bcl-2 and that bcl-2 can block the depletion of endoplasmic reticular Ca^{2+} stores. In addition, it has been shown that IP_3-dependent Ca^{2+} release is essential for triggering apoptosis. One of the consequences may lead to the activation of Ca^{2+}-dependent endonucleases that are responsible for the cleavage of DNA linker regions between nucleosomes, leading to the frequently observed DNA laddering during apoptotic processes.

Another interesting link between Ca^{2+} and apoptosis was recently reported by the discovery of a new member of a Ca^{2+} binding protein of the EF-hand type, called ALG-2, which was linked to apoptosis (see below). It was observed that T-cell hybridomas depleted of ALG-2 were protected against PCD induced by a variety of stimuli including dexamethasone and Fas/CD95 triggering. Evidence was provided that members of the ICE protease family were activated upon stimulation by either Fas/APO-1 antibodies or dexamethasone in ALG-2 depleted cells, but progression of cell death was impaired indicating that ALG-2 is necessary for the apoptotic function of ICE/Ced-3 proteases. However, later it was reported that mice lacking the *ALG-2* gene not only developed normally but it was also observed that in T cells obtained from those mice, apoptotic processes could be induced indicating that ALG-2 is not essential for apoptosis.

By using a similar functional selection strategy of cloning to identify genes linked to PCD, a *d*eath *a*ssociated *p*rotein (DAP) was identified as a kinase, therefore called DAP-kinase. This protein turned out to be a Ca^{2+}-calmodulin-dependent kinase of 160 kDa. It is specific for phosphorylating serine/threonine residues, and contains a number of other domains such as ankyrin repeats, P-loops, cytoskeleton binding domains, and death domains. It was demonstrated that the enzyme phosphorylated myosin light chains in a Ca^{2+}-calmodulin-dependent manner and is associated with the cytoskeleton of the cells. The latter observation could be of special interest since changes in the actin microfilament of the cytoskeleton organization preceded the nuclear condensation and segmentation in response to interferon-γ-stimulation or DAP-kinase overexpression.

5
Intracellular Calcium Binding Proteins

5.1
The EF-hand Protein Family

As pointed out before, the concentration of the intracellular Ca^{2+} in eucaryotic cells is closely regulated to remain below 5×10^{-7} M in a resting cell, whereas outside cells [Ca^{2+}] is 10^{-3} M, resulting in a steep concentration gradient across the plasma membrane. The intracellular level of Ca^{2+} can transiently increase owing to a response to extracellular signals. In order to function as a second messenger, the Ca^{2+} signal is transduced by a variety of Ca^{2+} binding proteins. In contrast to the extracellular, low-affinity Ca^{2+} binding proteins, the intracellular proteins bind Ca^{2+} with high affinity owing to a sequential arrangement of the amino acids ligating calcium in a loop flanked by two helical segments. This common helix-loop-helix Ca^{2+} binding motif, also known as the EF hand as coined by Kretsinger, is an important entity of the intracellular Ca^{2+} receptor proteins for triggering cellular responses. The helix-loop-helix motif can be present in several copies in these proteins. The number of the latter is steadily increasing, (to date more than 600 have been described, but for most of them a precise function is not known yet).

In 1973, Kretsinger and Nockolds reported the crystal structure of parvalbumin. The protein consists of two homologous calcium binding domains each of which contains two α-helices perpendicular to each other enclosing a Ca^{2+} binding loop, the well-known "EF-hand". Parvalbumin thus became the prototype of numerous Ca^{2+} binding proteins with similar structural properties, such as calmodulin, troponin C, recoverin, S100 proteins, and others. To date, more than 50 structures of EF-hand proteins are known. All EF-hand domains show a pentagonal-bipyramidal coordination of the Ca^{2+} ion in the loop flanked by the two approximately perpendicular helices. The residues forming the ligands to Ca^{2+} are highly conserved within a contiguous sequence of 12 residues spanning the loop and the beginning of the second helix. Loop residues in positions 1, 3, 5, and 12 contribute monodentate (1, 3, 5) or bidentate (12) Ca^{2+} ligands through side-chain oxygens and residue 7 through its backbone carbonyl oxygen. Therefore, an invariant glycine residue is in position 6 to permit the sharp bend necessary to ligate Ca^{2+} through the oxygen of a side chain (5) and a backbone carbonyl (7). Residue 9 provides an additional ligand, either directly through an oxygen of its side chain or indirectly via a water molecule. In position 1 of the loop, usually an aspartate is located, whereas in position 12, glutamic acid is invariant. The latter has two reasons:

1. both oxygens of the side-chain carboxylate ligate to calcium and
2. glutamic acid has α-helix-inducing propensity.

The Ca^{2+} binding domains usually occur in pairs stabilized by hydrogen-bond bridges between the central residues of adjacent loops that form a minimal antiparallel β-sheet. Owing to these pair-forming Ca^{2+} binding domains, the Ca^{2+} affinity to these sites and the cooperativity of binding is enhanced. Calmodulin and troponin C contain four Ca^{2+} binding sites. Both proteins display Ca^{2+} binding characteristics compatible with a "pair-of-pairs" model of EF-hands, that is, the two globular domains at the N- and C-terminal

site each contain a pair of EF-hands, and the domains are connected by a long central helix providing a dumbbell-shaped appearance as revealed by crystal-structure determination. The N- and C-terminal domains bind calcium cooperatively, the C-terminal domain with higher affinity. Recent multinuclear magnetic resonance experiments of the Ca^{2+}-bound form of calmodulin corroborated the general validity of the dumbbell-shaped structure in solution, but indicated the high degree of flexibility of the central part of the helix that is also supported by a high B-factor (temperature-factor) in that region in the crystal structure. This flexibility in the center of the molecule is an important property for the interaction with targets. Upon binding of calcium and subsequent conformational change of the protein, the extended dumbbell-shaped form of calmodulin in the absence of a target is transformed into a more compact form upon interacting with a target. By bending the central helix and bringing the two domains close together, thereby engulfing the peptide representing the CaM binding domain is documented by numerous examples, some of which will be discussed below in more detail (see Fig. 2).

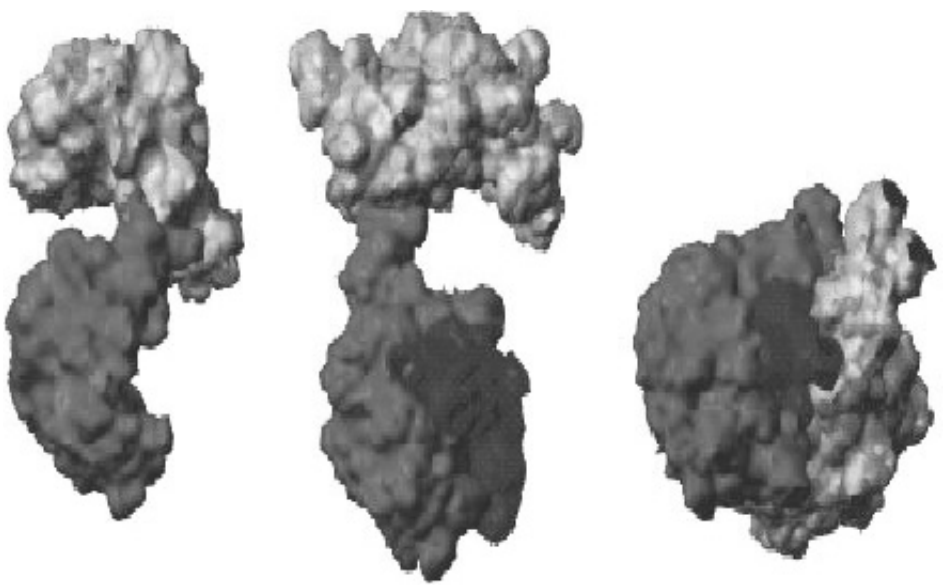

(a) CaM (b) CaM/C20W (c) CaM/M13

Fig. 2 Surface representation of calcium-bound calmodulin (CaM) and its different binding modes to peptides. The N-terminal half of CaM is in orange, the C-terminal half is in red, and the peptides are in blue. The orientation of the C-terminal half of CaM is the same in all cases. (a) Crystal structure of calcium-bound CaM. (b) Solution structure of the CaM/C20W complex showing that the peptide C20W, corresponding to the N-terminal portion of the CaM binding domain of the plasma membrane Ca^{2+} pump, but lacking the C-terminal hydrophobic anchor residue, binds only to the C-terminal half of CaM. (c) The solution structure of CaM/M13 shows a compact globular complex involving both C- and N-terminal halves of CaM in binding the peptide M13 corresponding to the CaM binding domain of MLCK. (Reprinted with permission from Elshorst et al., 1999, *Biochemistry*, **38**, 12320–12332).

5.1.1 Calmodulin and its Targets

5.1.1.1 Structural principles of Calmodulin-target interactions

As pointed out before, calmodulin (CaM) undergoes a profound conformational change upon binding of Ca^{2+} to the different EF-hand loops of the protein. As a consequence, deep hydrophobic clefts are built that are a prerequisite for the interaction with targets. In the absence of a target, the N-terminal and the C-terminal halves of CaM behave as independent structural units in solution that are linked by a flexible tether. The latter is part of the seven-turn central helix observed in the crystal structure of the protein.

CaM binding sequences of different targets reveal similar properties (see Table 1). They are composed of short peptides with an average length of 20 to 30 amino acids that have the propensity to form a basic amphiphilic α-helix. Several three-dimensional structures of complexes between calmodulin and target peptides representing binding domains of different CaM-dependent kinases, such as skeletal muscle and smooth muscle myosin light chain kinase (MLCK), CaM-dependent protein kinase II, and CaM-dependent protein kinase kinase, have been solved by either NMR or X-ray crystallography. These structures share some common structural features:

1. Calmodulin, which usually has an extended, dumbbell-type shape, is collapsed to a globular structure wrapped around the target peptide
2. The peptide is located in a hydrophobic channel passing through the center of the globular molecule
3. Two hydrophobic residues of the binding peptide spaced by a defined number of residues are essential for the interaction with CaM (see Table 1).
4. The peptides interact with CaM in an antiparallel manner, that is, the N-terminal half of CaM binds the C-terminal part of the peptide and *vice versa*. Structural examples of some CaM-target interactions are presented in Fig. 2 showing the structures of calmodulin in its Ca^{2+}-bound form and how it interacts with the CaM binding domains of the plasma membrane calcium pump and of MLCK.

An exception to rule (4) was recently reported by Ikura et al. who solved the structure of the complex of CaM bound to the peptide of the binding domain of CaMKK by NMR. In this complex, the peptide comprises unusual structural properties, that is, only the N-terminal

Tab. 1 Aligned sequences of CaM binding Domains of CaM-dependent Protein Kinases.

skMLCK (M13)	KRR	W	KKNFIAVS	A	ANR	F	KKI
smMLCK (R20)	RRK	W	QKTGHAVR	A	IGR	L	SSM
CAMKI	KSK	W	KQAFNATA	V	VRH	M	RKL
CAMKII	RRK	L	KGAILTTM	L	ATRN		FSA
CAMKIV	RRK	L	KAAVKAVV	A	SSR	L	GSA
CAMKK	IPS	W	TTVILVKS		MLRK		RS F G

Boxed hydrophobic residues represent the critical anchoring amino acids of the different CaM binding domains.

half forms an α-helix, the C-terminal half is a hairpin loop folding back on itself. In addition, the peptide is oriented in a parallel fashion relative to the two halves of CaM.

Interestingly, recent experiments with small-angle X-ray scattering and neutron scattering by the group of Trewhella provided evidence that also when CaM is complexed to an intact enzyme, that is, MLCK, CaM undergoes a conformational collapse identical to that observed with the peptide corresponding to its binding domain.

The first crystal structure of a CaM-dependent enzyme, CaM-dependent kinase I (CaMKI), obtained in the absence of CaM, that is, in the autoinhibited state, supports the view that the C-terminal regulatory domain forms a helix-loop-helix segment that interferes with the two domains of the catalytic core, that is, with the binding site of peptide substrates and with the ATPbinding site. An interesting feature of this structure is the striking observation that the N-terminal part of the CaM binding domain is accessible for an initial interaction with CaM. This part comprises the loop region of the regulatory domain including the conserved tryptophan that provides the hydrophobic anchor for the C-terminal half of CaM. Furthermore, it was recently postulated that the interaction between the C-terminal half of CaM and the CaM binding domain of a CaM-dependent enzyme is necessary, and sometimes already sufficient, to release the autoinhibited state of the enzyme. This view was based on the three-dimensional NMR structure of the complex between CaM and the N-terminal part of the binding domain of the plasma membrane calcium pump (Fig. 2b), a view that is supported by recent structural studies on MLCK indicating that initial binding of CaM to MLCK occurs already at substoichiometric Ca^{2+} concentrations, that is, at about 2 moles of Ca^{2+}/mol of CaM.

Calmodulin is a very versatile modulator protein interacting with a variety of different enzymes such as adenylate cyclases, cyclic nucleotide phosphatases, NO synthases, protein kinases, and others. One of the best-studied families of CaM-dependent enzymes is the CaM-dependent protein kinases, which will be outlined in more detail.

5.1.1.2 **Calmodulin-dependent Kinases**

Calmodulin-dependent protein kinases can be divided into two categories

1. Kinases dedicated to single substrates, called *monosubstrate kinases*. Examples of this class are the myosin light chain kinase (MLCK), the phosphorylase kinase, and CaM-dependent kinase III or elongation factor 2 kinase. MLCK will be described in more detail.
2. Kinases dedicated to different substrates, called *multisubstrate kinases*. Examples of this class are CaM kinase I, II, and IV, and the CaM kinase kinase, activating CaMKI and IV.

5.1.1.2.1 **Monosubstrate CaM Kinases**

Myosin Light Chain Kinase (MLCK) The most extensively studied calmodulin-dependent enzyme is the myosin light chain kinase (MLCK). This enzyme catalyzes the phosphorylation of a specific serine residue in the N-terminal portion of the myosin II regulatory light chains. Two types of MLCK exist that differ genetically, biochemically, and physiologically and that are expressed in two different types of muscles: the MLCK of skeletal muscles (skMLCK) and of smooth or nonmuscle cells (smMLCK). The two different MLCK

enzymes are encoded by different genes. SkMLCK corresponds to a molecular mass of 65 kDa, whereas the expression pattern and genomic organization of sm/nm MLCK is more complex. This is due to alternative splicing that results in different isoforms with significantly different molecular weights, ranging between 150 and 200 kDa owing to different lengths of their amino termini.

To date, the regulatory myosin light chain is the only established substrate of MLCK. In smooth muscles, the Ca^{2+}-dependent trigger of muscle contraction is directly mediated by the phosphorylation of the light chain by MLCK thus removing the inhibition of the myosin ATPase. In skeletal muscle, however, the actomyosin ATPase is regulated by binding of Ca^{2+} to troponin C, the regulatory calcium binding protein in the troponin complex. A possible function of MLCK in skeletal muscle is to modulate the rate of enhancement and extent of isometric contraction of the muscle.

The substrate specificity of MLCK is very strict. Substrate recognition is basically dependent on the appropriate location of at least four basic residues over a more extended region N-terminal of the phosphorylated serine residue. Calmodulin-dependent enzymes are activated by disrupting an autoinhibitory segment that interacts with the catalytic core of the enzyme and thus deinhibiting the enzyme. This concept was first developed and verified for the cAMP-dependent protein kinase (PKA) that could be blocked by an endogenous inhibitor (PKI) competing with the substrate about the substrate binding site, thus serving as a pseudosubstrate. MLCK is the CaM-dependent enzyme that has been studied very thoroughly by Kemp et al. to provide evidence for the existence of a pseudosubstrate to serve as an autoinhibitory domain in MLCK. This view gained further support by the finding that controlled proteolysis at the C-terminal end of MLCK generated a constitutively active fragment independent of calmodulin activation, that is, the autoinhibitory domain was removed.

5.1.1.2.2 **Multisubstrate CaM Kinases** Four different enzymes of calmodulin-dependent protein kinases fall into the class of multifunctional or multisubstrate kinases, CaMKI, CaMKII, CaMKIV, and CaMKK, the activating kinase of CaMKI and IV. This class can be subdivided into CaMKII on one side and CaMKI, CaMKIV, and CaMKK on the other for two reasons: (1) CaMKI, CaMKIV, and CaMKK act as monomeric proteins, whereas CaMKII is a multimeric enzyme composed of up to 12 subunits; (2) CaMKI and CaMKIV are activated by another calmodulin-dependent protein kinase, that is, by CaMKK. Whether CaMKK itself is also activated by an upstream acting kinase is not clear at the moment, but there are indications that this indeed might be the case. In contrast, CaMKII is activated by autophosphorylation. Therefore, the latter is discussed separately from CaMKI, CaMKIV, and CaMKK.

CaMKII Calmodulin-dependent protein kinase II (CaMKII) represents a family of closely related protein kinases with broad substrate specificity. This class of protein kinases is involved in cellular metabolism, neurotransmitter release, cytoskeletal structure, gene expression, cell cycle control, and, more recently, in long-term potentiation and depression (LTP and LTD), functions that are thought to be involved in memory and learning.

Structural Organization The holoenzyme is an oligomeric complex of 6 to 12 subunits with a total molecular mass ranging from 300 to 700 kDa, which can be either a homomultimer or a heteromultimer. According to the amino acid sequences, the different isoforms are closely related to each other and reveal the segmental organization typical for CaM-dependent protein kinases, consisting of catalytic, regulatory, variable, and association domains on the same polypeptide chain.

The overall structure of the catalytic domain of CaMKII is likely to be bilobal, as expected for all CaM-dependent protein kinases, on the basis of the crystal structure of CaMKI. Catalysis takes place within a cleft between the two lobes orienting the consensus sequence of the substrate in such a way that the basic residue(s) in proximity to the serine/threonine residue that should be phosphorylated could interact with the conserved polar residues located on the surface of the large lobe. The small lobe contains a lysine residue that is essential for ATP binding, that is, if mutated to a methionine renders the kinase inactive.

Regulation of the Catalytic Activity Since CaMKII has a broad range of substrates, the consensus sequence for substrate determination is less strict than that described for MLCK before. Basically, an arginine at the P−3 position, the phosphorylated serine/threonine being P(0), is the only essential requirement for a multifunctional kinase such as CaMKII. In addition, hydrophobic residues are preferred at P−5 and P+1. Downstream of the catalytic domain, the autoinhibitory, the CaM binding, and the association domains are located, interspersed with variable domains.

As discussed in detail before, CaM activates its targets by wrapping around the binding domain thereby releasing the enzyme from its autoinhibitory state. In the case of CaMKII, the three-dimensional structure of the complex between CaM and a peptide representing the CaM binding domain of CaMKII has been determined by X-ray crystallography. This complex also shows the general feature observed for similar CaM–peptide complexes, that is, CaM is collapsed to a globular structure engulfing the helical peptide that is fixed by numerous hydrophobic and electrostatic interactions. An important difference between the interactions of CaM with CaMKII and MLCK, respectively, is the different affinity of CaM for the two enzymes, that is, the affinity for CaMKII is much lower (K_{act} 20–100 nM) than that for MLCK (K_{act} 1 nM). However, the affinity for CaMKII can be substantially enhanced (>1000-fold) by autophosphorylation of Thr286 resulting in a significant lowering of the CaM off-rate from the complex, a phenomenon termed *calmodulin trapping* that might have some important implications for the function of the enzyme as discussed below.

CaMKII and Neuronal Plasticity Cyclic cellular depolarization or hormonal stimuli can trigger a series of calcium pulses inside cells. In neurons, these Ca^{2+} oscillations can be used to modulate synaptic activity and strength with subsequently increased release or resynthesis of neurotransmitters or other cellular activities. The coordinator of at least some of these activities appears to be CaMKII; however, the question is, "Is CaMKII able to decode Ca^{2+} oscillations, and if yes, how?"

The hippocampus has been used to study processes in which high-frequency

stimulation leads to long-term potentiation (LTP) and low-frequency stimulation to long-term depression (LTD). These processes have been implied as important models for studying spatial learning and memory storage. It was also reported that both processes are induced by an increase in intracellular Ca^{2+} in the postsynaptic neuron concommitant with a rise in CaMKII activity and phosphorylation of synapsin I.

An important link has been made between LTP and some forms of learning, and its dependence on CaMKII by gene-targeted disruption of CaMKIIα, by the group of Tonegawa. These studies provided evidence that those mice that developed normally were not only impaired in spatial learning but also that induction of LTP was blocked in hippocampal slices from those mice.

Recently, De Koninck and Schulman provided direct evidence that CaMKII indeed can decode the frequency of Ca^{2+} pulses. Owing to the complex activation pattern of CaMKII by CaM, autophosphorylation, and subsequent CaM trapping, the enzyme can become autonomous. Autophosphorylation is an intersubunit reaction between proximate subunits in which CaM not only activates the "kinase" subunit but also presents the "substrate" subunit for phosphorylation. Subsequently, the kinase is transformed into the "trapped" state, that is, a cooperative, positive feedback loop resulting in a short molecular "memory" that could enable the enzyme to respond to digital and cyclic activation associated with Ca^{2+} oscillations. In simulation calculations, it was predicted that repetitive Ca^{2+} pulses lead to recruitment of CaM, autophosphorylation and trapping of CaM, establishing a threshold frequency at which the activity of the enzyme is sustained. These predictions were exactly confirmed by the experiments of De Koninck and Schulman who could demonstrate that, independent of the Ca^{2+}-CaM pulse duration, the autonomous activation of CaMKII increased steeply as a function of frequency. On the other hand, once a threshold value was achieved, it was possible to maintain the response level with signals of substantially lower frequency.

CaMKI and IV In contrast to CaMKII, which is a multimeric protein, CaMKI and CaMKIV are monomeric enzymes. CaMKI is, like CaMKII, ubiquitously expressed, whereas the tissue distribution of CaMKIV is restricted to nervous tissues, the thymus, particularly T lymphocytes, the bone marrow, keratinocytes, testis, and ovary. Both enzymes are activated not only by autophosphorylation but also by another CaM-dependent protein kinase, the CaM kinase kinase (CaMKK).

The Crystal Structure of CaMKI In 1996, the group of Kuriyan reported the crystal structure of CaMKI in the autoinhibited state, that is, in the absence of CaM. This is the first three-dimensional structure of a calmodulin-dependent enzyme determined at high resolution, as mentioned before. The crystal structure of CaMKI provided evidence of significant contacts between the autoinhibitory sequence and the catalytic core of the kinase, thereby supporting the pseudosubstrate model for activation of calmodulin-dependent protein kinases. Another important result was the finding that Trp303 corresponding to the first anchoring residue of the CaM binding domain (see Table 1) lies on the surface of the protein. This arrangement makes it very likely that CaM could first bind to the exposed Trp303 and subsequently release the β-strands of the ATP

binding domain from their interaction with the regulatory segment comprising the core of the CaM binding domain. Such an arrangement could be of general relevance for the activation mechanism of CaM-dependent enzymes. The structure further suggests that binding of CaM and subsequent reorientation of the ATP binding domain would open up the substrate binding domain that would permit Thr177 to be accessible for the activating CaMKK.

Regulation of Activity The regulation of the activity of CaMKI and CaMKIV is a rather complex process and involves three major steps:

1. Binding of Ca^{2+}/CaM to the enzyme
2. Phosphorylation of either Thr177 (CaMKI) or Thr196 (CaMKIV) located in the activation loop of the kinases by CaMKK
3. Autophosphorylation of serine residue(s) at the N-terminus of the enzyme

CaMKII on one hand and CaMKI and CaMKIV on the other are activated by different mechanisms. As discussed before, CaMKII is a multisubunit, homo- or hetero-oligomer enzyme complex that is activated by autophosphorylation between two neighbouring subunits. On the other hand, CaMKI and CaMKIV do not seem to multimerize, and both enzymes are activated by CaMKK; whether both are activated by the same kinase kinase or by different enzymes is not entirely clear. In CaMKIV, the critical residue phosphorylated by CaMKK is Thr196, which is located in the activation loop of the kinase and is homologous to Thr177 in CaMKI. Accessibility of the two threonine residues is brought about by the binding of calmodulin, that is, binding of calmodulin not only releases autoinhibition of I or IV but also enables the enzymes to be phosphorylated at this critical residue. This phosphorylation results in a 10- to 20-fold activation in the case of CaMKIV. Furthermore, it was reported for CaMKIV that Ser11 and Ser12 could be autophosphorylated in a slow process as indicated before, but CaMKK could considerably enhance this process by phosphorylating Thr196.

CaMKIV and Ca^{2+}-dependent Gene Expression One of the consequences of elevated calcium in the cell, especially in the nucleus, is the induction of gene expression. Since transcription factors such as CREB, CREMτ, ATF-1, SRF, and ETS-1 are among the best substrates for CaMKIV, and CaMKIV has been localized to the nucleus, the enzyme could have direct access to transcription factors to regulate their function in a Ca^{2+}-dependent manner. Thus, it has been shown in different cell lines that CaMKIV is involved in the regulation of expression of immediate early genes (IEGs) either through CREB or through SRF.

Originally, CREB was identified as a transcription factor induced in a cAMP-dependent process, but later studies demonstrated that several kinases activating CREB phosphorylated the transcription factor at the same serine residue, Ser133. In contrast to CaMKIV, CaMKII phosphorylates next to Ser133 also Ser142, which has an inhibitory effect on gene expression.

CaM-dependent Protein Kinase Kinase (CaMKK) As indicated before, CaMKI and IV are activated by calmodulin-dependent protein kinase(s), the CaMKK. Two isoforms of CaMKK are known to date, α and β which are encoded by different genes. Both isoforms are organized

similar to other CaM-dependent kinases with catalytic, autoinhibitory, and CaM binding domains. In addition, CaMKK contains an unusual arginine–proline rich insert within the N-terminal part of its catalytic domain that might be important for the recognition of CaMKI and IV as substrates since deletion of these inserts abolishes its activation potential toward CaMKI and IV.

CaMKK can be found both in the cytoplasm and in the nucleus; the latter is important for the activation of CaMKIV that is mainly responsible for the Ca^{2+}-dependent gene expression. In this context, it is interesting to note that the distribution of βCaMKK in the brain is much more similar to CaMKIV, and that of αCaMKK follows the distribution of CaMKI, suggesting that βCaMKK could be the actual activator of CaMKIV, whereas αCaMKK would be that of CaMKI. Nevertheless, both isoforms of CaMKK can activate CaMKI and IV with the same efficiency, but whether there exist more specific isoforms and whether there are additional kinases further upstream like in the MAP-kinase cascade is not known to date.

5.1.2 Neuronal Calcium Sensors

A group of EF-hand proteins that has recently received attention is that of the neuronal calcium sensors (NCS). They are divided into five subfamilies. Two of them are expressed in retinal photoreceptors, called recoverins and guanylate cyclase-activating proteins (GCAPs), the three others frequenins, visinin-like proteins, and the Kv channel–interacting proteins. Recoverins and GCAPs have antagonistic roles in phototransduction: recoverin inhibits rhodopsin kinase and GCAPs activate guanylate cyclase. The other three NCS families are supposed to regulate the release of neurotransmitters, the biosynthesis of polyphosphoinositides, the metabolism of cyclic nucleotides, or the activity of Kv channels. Most NCSs are N-terminally myristoylated, favoring interaction with membranes (or target proteins) in a Ca^{2+}-dependent manner. Thus, the Ca^{2+}-dependent myristoyl switch could be a means to compartmentalize signaling cascades in neurons or to transduce Ca^{2+} signals to the membranes.

5.1.3 S100 Proteins

The multigenic family of the so-called S100 proteins (named because they are soluble at 100% ammonium sulfate solution) is growing steadily, and possible functions have been implicated in intracellular and extracellular regulatory activities. To this family belong proteins that are involved in cell cycle control (e.g. calcyclin) or can display neurite growth factor activities such as the S100β dimer. S100B was the first member of the EF-hand family that is secreted to perform its extracellular function as a growth factor. The latter function is exclusively connected with the β-isoform and only so if it can form a dimer. Mutation of one of the cysteine residues necessary for the dimer formation leads to the loss of this function. Recently, it has been suggested that the S100β dimer carries out its growth factor function through the RAGE receptor (receptor for advanced glycation end products).

In general, members of the S100 protein family have low molecular weights (M_r between 9 and 14 kDa). They usually function as homodimers, but some also as heterodimers. Of special note is the finding reported by Heizmann et al. that in human S100A1–S100A13 are clustered on chromosome 1 (1q21) within 5 Mb, but a common regulation of expression has not been observed.

5.1.4 The Penta EF-hand Family

The EF-hand motifs in Ca^{2+} binding proteins are usually tandemly repeated. Owing to the recent determination of the X-ray structure of the small subunits of the calcium-dependent protease calpain, it was recognized that also uneven numbers of EF-hand domains may exist in proteins. By comparing the primary structures of a number of different calcium binding proteins, Maki et al. identified several proteins that, like calpain, contained five EF-hand domains, and therefore introduced the name "penta EF-hand family" or PEF family for this subgroup of EF-hand-containing proteins. Members of this family include calpain, sorcin, grancalcin, peflin, and ALG-2; the latter was originally identified as "**a**poptosis-**l**inked **g**ene". Characteristic to all the members of this family is a two–amino acid insertion into the loop of the fifth EF-hand domain, and therefore, at least under physiological conditions, may not bind calcium anymore. Instead, the fifth domain may serve as a dimerization domain, as originally observed in the calpain structure and later confirmed by solving the structure of ALG-2 and grancalcin.

Proteins of the PEF family have been quite conserved during evolution; homologs have been found in invertebrates, plants, fungi, and even in protists. On the basis of comparison of the gene structures, PEF proteins can be classified into two groups: group I contains ALG-2, peflin, and their homologs, and group II is composed of calpain, sorcin, and grancalcin.

As indicated before, calpain is a calcium-dependent, intracellular protease, of which two major isozymes exist, calpain I and II, which differ in their Ca^{2+} dependence; that is, calpain I or μ-calpain is activated by micromolar calcium concentrations, whereas calpain II or m-calpain needs millimolar calcium to be activated *in vitro*. Both isoforms exist as heterodimers consisting of a large subunit (80 kDa), containing the catalytic site, and of a small regulatory subunit (30 kDa). Recently, Suzuki, Bode et al. determined the structure of the full-length, heterodimeric human calpain II in its Ca^{2+}-free form as shown in Fig. 3: the two subdomains IIa and IIb of the large subunit comprising the catalytic site are rotated against each other, disrupting the active and the substrate binding sites, which would explain the inactivity of calpain in the absence of calcium. Surprisingly, the group of Davies who determined the structure of calpain I in the Ca^{2+}-bound form suggested that the cooperative binding of Ca^{2+} to two non-EF-hand-type Ca^{2+} binding sites derived from two different peptides aligned the active site cleft and converts the protein into an active enzyme, and Ca^{2+} binding to the EF-hand-type binding sites would contribute to the Ca^{2+} sensitivity of the enzymes, as documented by mutational analysis. In addition, both isoforms of calpain are maintained in an inactive form because of the interaction with an endogenous inhibitor, calpastatin, until external signals activate the protease.

Increasing evidence indicates that calpain is involved in the regulation of basic cellular processes such as cell proliferation, differentiation, and apoptosis owing to cytoskeletal remodeling, which, under pathological conditions, could contribute to tissue damage in heart and brain ischemias as well as neurodegeneration in Alzheimer's disease.

ALG-2 was originally identified as a gene linked to apoptosis. It is a 22 kDa highly conserved and ubiquitously expressed protein that forms dimers through the fifth EF-hand, as documented by the crystal structure. This could provide a new

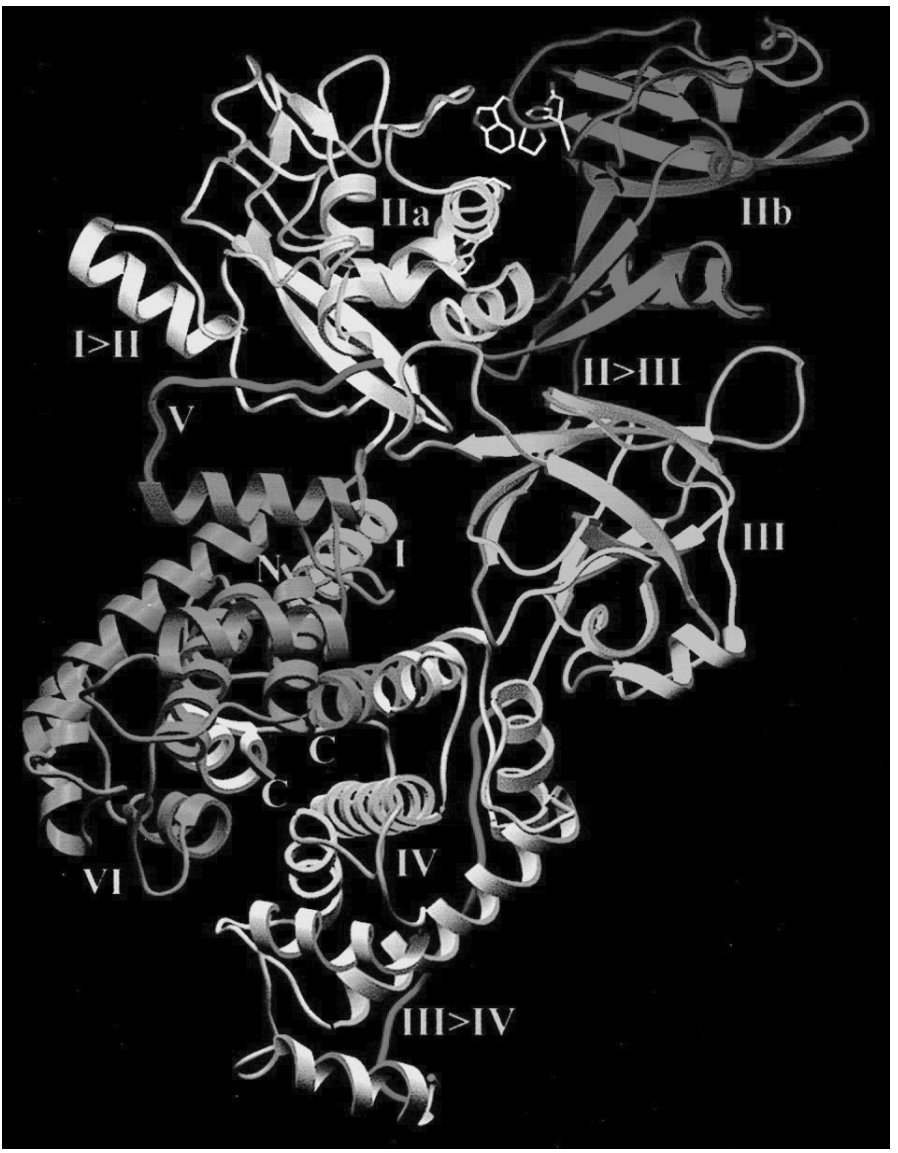

Fig. 3 Ribbon structure of human m-calpain in the absence of calcium. The 80 kDa large subunit starts in the molecular center, folds into the surface of the dIIa subdomain, forms the catalytic domain dII, builds domain dIII, runs down, and forms the right-side calmodulin-like domain dIV. The 30 kDa small subunit becomes visible before forming the left-side calmodulin domain dVI. (Reprinted with permission from Strobl et al., 2000, *Proc. Natl. Acad. Sci. USA*, **97**, 588–592.)

interface for the interaction with possible targets of which a number of different proteins such as AIP or Alix have been identified, which in Xenopus is involved in the maturation process of oocytes, peflin, and different annexins. All these targets interact with ALG-2 in a calcium-dependent manner indicating that ALG-2, like calmodulin, may have calcium sensing, that is, signal transduction–regulating properties. Therefore, the involvement in apoptosis, which gave the protein its name, would be one but not the exclusive signal transduction pathway. In addition, as mentioned before, it has been shown that ALG-2 is not essential for apoptotic responses since by using T cells from ALG-2-deficient mice, apoptotic stimuli could be induced by either stimulating T-cell receptors (TCR), Fas/CD95, or by glucocorticoids.

Recent experiments carried out in our lab indicated that ALG-2 may not only be involved in apoptotic processes but may also play a role during cell proliferation. A significant nuclear concentration of ALG-2 was found in cells prior to cell division, in addition to a significant increase in ALG-2 expression in highly proliferative cells obtained from cancerous tissues, especially evident for those from metastatic tissues. ALG-2 may thus be an important modulator involved in the cellular decision between cell proliferation and cell death.

5.2 Non-EF-hand Ca^{2+} Binding Proteins

5.2.1 Annexins

Next to the family of the EF-hand-containing Ca^{2+} binding proteins, another intracellular Ca^{2+} binding protein family became known in recent years. These soluble, amphipathic proteins bind to membranes containing negatively charged phospholipids in a Ca^{2+}-dependent manner and are therefore called *annexins*. They are widespread in the animal and plant kingdom and have been claimed to be involved in a variety of cellular functions such as interaction with the cytoskeleton, membrane fusion, anticoagulation, signal transduction, or phospholipase inhibition. The primary structure of the annexins known to date contain 4 or 8 conserved repeat units of about 75 amino acids in length, which are separated by intervening sequences of variable length. For some of these proteins, the crystal structure is known. The most detailed studied protein, annexin V, is almost entirely α-helical. It consists of five α-helices bundled into a right-handed super-helix. On the basis of this structure, it was proposed that annexin V functions as a calcium channel, and some experiments using a reconstituted system seem to support a voltage-gated mechanism. In contrast to the EF-hand-containing Ca^{2+} binding proteins, the ligands coordinating calcium in the annexins are not adjacent in sequence. Three calcium binding sites seem to exist in annexins, two invariably in repeats II and IV, one in either repeat I or III. As indicated for the EF-hand proteins, also in the annexins, calcium is hepta-coordinated with ligands organized in a pentagonal bipyramid. The ligating oxygens mainly stem from peptide carbonyls or water molecules.

An interesting feature of annexin I and II is their phosphorylation by either the EGF receptor kinase or the *src*-encoded tyrosine kinase respectively suggesting their involvement as coupling factors between growth factor receptors and their cellular targets. Furthermore, annexin II has been found to form stable heterotetrameric complexes with p11, a member of the S100

family, thereby increasing its Ca^{2+} sensitivity by 3 orders of magnitude. However, formation of such a complex interferes with the phosphorylation of annexin II by different kinases.

5.2.2 Gelsolin

Gelsolin is a multifunctional protein binding actin in a Ca^{2+}-dependent manner. It can also nucleate actin polymerization by binding two actin monomers. Alternative transcription initiation and selective RNA processing produces two isoforms from the same gene; one isoform is located intracellular and the other extracellular. The intracellular gelsolin is involved in cell motility–regulating actin function, whereas extracellular gelsolin can act as an actin-scavenging system to prevent the polymerization of actin released after cell death.

Gelsolin consists of 6 repeat units (G1 to G6) containing 120 to 150 amino acids. The units are organized in 2 clusters of similar architecture and are connected by a flexible linker of about 50 amino acids that may be cleaved by caspase-3. One repeat unit, G1, has been crystallized in a complex with actin. It is organized in a central four-stranded β-sheet motif faced on one side by a four-turn α-helix parallel to the strands and on the other side by a shorter α-helix perpendicular to the strands. The longer helix also serves as the binding partner to a cleft between two actin subdomains. The G1-actin-complexed domain comprises two Ca^{2+} binding sites.

The regulation of gelsolin activity by Ca^{2+} has recently been proposed. This was made possible by comparing the Ca^{2+}-free structure of nonactive gelsolin with the active, Ca^{2+}-bound gelsolin complexed to actin. In the absence of Ca^{2+}, the six repeat units of gelsolin provide a very compact globular structure, thus blocking the actin-binding helices of the appropriate subdomains. By binding of Ca^{2+} to gelsolin, the N- and C-terminal halves of the protein, that is, G1 to G3 and G4 to G6, become separated, thereby unmasking the actin binding sites, and hence permit binding to actin strands.

5.2.3 C_2-domains

Synaptotagmin, a transmembrane protein of synaptic vesicles, is believed to act as the major Ca^{2+} sensor of exocytosis and neurotransmitter release. The cytoplasmic region of synaptotagmin contains two C_2 domains (C_2A and C_2B) that are Ca^{2+} binding domains of non-EF-hand character. These domains have been first described in protein kinase C by Nishizuka, and have been suggested to be responsible for binding to membranes in response to Ca^{2+}. The C_2 domain consists of a β-sandwich of two four-stranded β-sheets. The β-strands are connected by loops, some of which bind clusters of Ca^{2+} ions, primarily through oxygens of aspartate side chains. Sequence homology and structural similarity among C_2 domains of different proteins are high for the β-strands, which probably constitute scaffolds to support the Ca^{2+} binding loops. A recent detailed mutational analysis of synaptotagmin in *Drosophila* seems to indicate that only the C_2B domain represents the necessary Ca^{2+} sensing domain required for synaptic vesicle fusion.

6 Systems Controlling Intracellular Ca^{2+} Concentration: Structural and Functional Properties

Owing to its function as an intracellular second messenger, the ionized Ca^{2+} has

to be kept in a narrow concentration range in the resting cell (i.e. 100–300 nM). On the other hand, as indicated before, the free Ca^{2+} concentration in the extracellular milieu is 2 to 5 mM. This results in a very steep concentration gradient across the membrane, and thus small changes in the free Ca^{2+} concentration can influence the different signal transduction pathways. Therefore, a number of different transmembrane Ca^{2+} transporting systems participate in controlling the free Ca^{2+} concentration in the cell. Most of these systems are either located in the plasma membrane (Ca^{2+} channel, ATP-dependent transporting system, Na^+/Ca^{2+} exchanger), in the sarco (endo)plasmic reticular system (ATPase, Ca^{2+}-release channel), or in mitochondria (an electrophoretic uptake system, a Na^+-dependent Ca^{2+} exchanger).

6.1 Calcium Transport Systems of the Plasma Membrane

6.1.1 The Calcium Channel

Transient changes in the intracellular free Ca^{2+} can be due to calcium deriving either from intracellular stores (see below) or from the extracellular fluid by passing through specifically regulated (i.e. gated) channels in the plasma membrane down their electrochemical gradient. Electrophysiological and pharmacological properties have been used to characterize different subtypes of the calcium channels. To date, six different types of channels have been identified: L, N, T, P, Q, and R. On the basis of their electrophysiological properties, using the well-established patch-clamp technique, it can be shown that these channels differ by their opening kinetics and their conductance:

1. L-type: *long-lasting* inward current and strong conductance (7–8 pS at 0.1 M Ca^{2+}) antagonist: dihydropyridines
2. T-type: *transient* inward current and a conductance of 5 to 10 pS at 0.1 M Ca^{2+}
3. N-type: originally defined as *neither* L-*nor* T-type channel (mainly found in *neurons*) with a characteristic activation by strong depolarization and intermediate conductance (11–15 pS at 0.1 M Ca^{2+}). Insensitive to dihydropyridines, but blocked by ω-conotoxin-GVIA
4. P-type: high-voltage activated channel, mainly found in cerebellar *Purkinje* cells, blocked by ω-Agatoxin-IVA but not by ω-conotoxin-GVIA

Recently, two other types of channels have been identified in Purkinje cells, Q-type and R-type channels. Q-type channels are less sensitive to ω-Agatoxin-IVA, whereas R-type channels are insensitive to ω-Agatoxin-IVA and ω-conotoxin-GVIA but sensitive to Ni^{2+}. All these channels are characterized by their difference in voltage dependence and their response to pharmaca. A different type of classification depending on the mechanism by which the transition between the "open" and "closed" conformations are regulated distinguishes between

1. voltage-operated channels (VOC) in which the gating depends on voltage as described above
2. receptor-operated channels (ROC) in which gating depends on ligand binding
3. store-operated channels (SOC) in which activation depends on the depletion of Ca^{2+} stores of the endoplasmic reticulum (ER) through a mechanism known as capacitative calcium entry (CCE).

Structurally, all VOC-type channels are complexes of four or five subunits ($\alpha 1$, $\alpha 2$, β, γ, and δ) forming large macromolecular complexes. The best-characterized examples are the **L**-type channels. The representative from skeletal muscle is composed of two high molecular weight components, $\alpha 1$ and $\alpha 2$, plus three smaller subunits β, γ, and δ. The channel protein represents a multigene family. To date, 10 distinct genes for the $\alpha 1$-subunit and three different ones for the β-subunit have been identified in mammals. The $\alpha 1$-subunit reveals some properties characteristic of voltage-gated channels. It is composed of four homologous repeated units containing six transmembrane regions and contains the conduction pore, the voltage sensor, and sites for gating and for channel regulation, for example, by protein kinases, by toxins, and by drugs. Dihydropyridines (e.g. nifedipine), phenylalkylamines (e.g. verapamil), and benzothiazepines (e.g. diltiazem) bind to three allosterically coupled binding sites on the $\alpha 1$-subunit. A physiologically important means of regulating these channels is provided by cAMP-dependent (PKA) phosphorylation increasing the opening probability of the channels. In reconstituted systems, it was shown that the $\alpha 1$- and β-subunits could be phosphorylated by PKA in a stoichiometric manner.

Three different kinds of **r**eceptor-**o**perated Ca^{2+} **c**hannels (ROC) have been identified. They are activated by L-glutamate and are named after the different agonists binding to these receptors as kainate (KA), α-amino-3-hydroxy-5-methyl-4-isoxazole proprionate (AMPA), or as N-methyl-D-aspartate (NMDA) receptors. They are located in the postsynaptic membrane. Most channels formed by KA and AMPA receptors are primarily permeable to Na^+ or K^+, but some of the AMPA receptors are also permeable to Ca^{2+}, whereas the NMDA receptors are permeable to Na^+ and Ca^{2+}. Several genes have been identified characterizing the different types of glutamate receptors. The mass of the different receptors ranges from 95 to 165 kDa.

In neuroendocrine cells, activation of an inward Ca^{2+} current occurs through channels known as store-operated Ca^{2+} channels (SOC). These channels have not been characterized in detail on the protein level, but they are homologous to the transient receptor potential (*trp* or *trp*-like) gene products of *Drosophila*. The mechanism by which these intracellular stores, probably derived from the endoplasmic reticulum, communicate their Ca^{2+} level to the plasma membrane in the form of a channel-opening signal is still under debate. The process, termed *capacitative calcium entry* (CCE) mediates smaller Ca^{2+} influx than the voltage-gated channels. Two hypotheses for the mechanism of CCE are discussed: one favors the liberation of a small chemical factor inducing the opening of the CCE channel, the other proposes some form of physical interaction between the ER and the plasma membrane CCE to trigger the opening.

6.1.2 The Ca^{2+} Pump

The Ca^{2+} pump of plasma membranes (PMCA) is important for the fine-tuning of the intracellular free Ca^{2+} concentration. It belongs to the P-type ion-motive ATPases like its counterpart from the endo (sarco)plasmic reticulum, the SERCA pump; that is, it forms an aspartylphosphate intermediate to transport Ca^{2+} against its concentration gradient across the plasma membrane on the expense of ATP. It is a protein of low abundance (e.g. 0.1% of the total membrane

proteins of the human erythrocytes), but it is ubiquitous in the plasma membranes of all cells. Calmodulin (CaM) activates the PMCA by direct interaction, thereby lowering the K_M for Ca^{2+} by one order of magnitude and increasing the V_{Max} two- to threefold; that is, as a result, the Ca^{2+} affinity and its transport rate are increased. This calcium-dependent affinity of the Ca^{2+} pump to CaM was also exploited to isolate the enzyme in pure form from detergent-solubilized erythrocyte membranes by applying calmodulin affinity chromatography. Next to the stimulation by CaM, the plasma membrane Ca^{2+} pump can also be activated by alternative treatments, that is, by acidic phospholipids, fatty acids, phosphorylation by different protein kinases (PKA or PKC, see below), oligomerization, or controlled proteolytic treatment. The latter procedure helped identify a number of functional domains of the enzyme (see Fig. 4). A significant difference in the properties between PMCA and the SERCA pump concerns the Ca^{2+}/ATP stoichiometry, which is 1 for PMCA but 2 for SERCA. In this respect, it is of interest that most polar residues that have been predicted (and confirmed through the structure) as participating in Ca^{2+} binding and transport through the

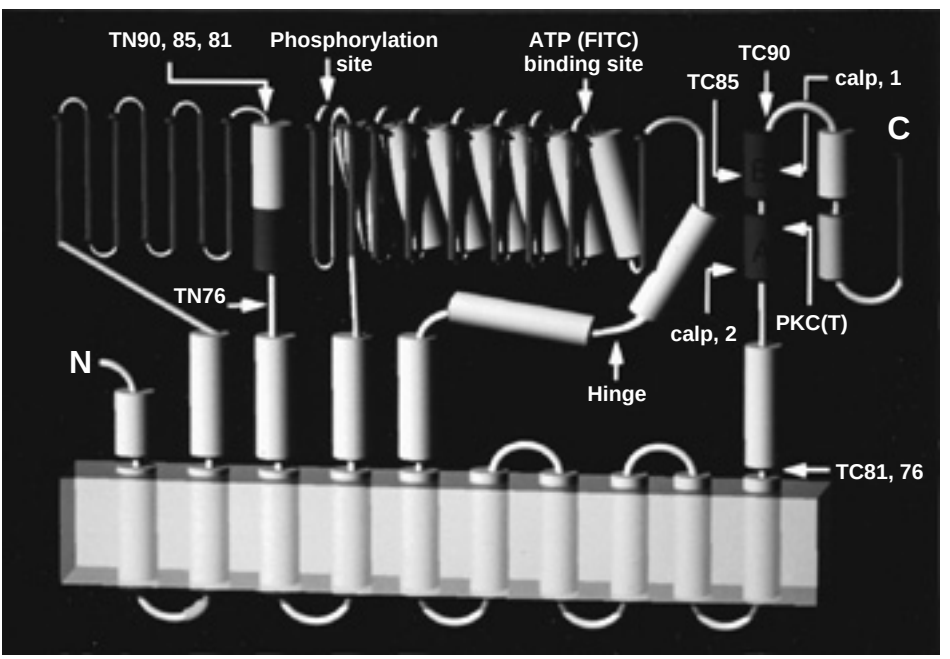

Fig. 4 Model of the topography of the plasma membrane Ca pump, illustrating the arrangement of the different functional domains assigned according to the primary structure of the enzyme and on the basis of secondary structure predictions. (A similar model of the SERCA pump was recently validated by solving the crystal structure, see text). The figure also includes different proteolytic cleavage sites (TN, TC), producing a number of enzymatically active fragments of different size (not discussed in the text); calp, calpain cleavage sites; PKC(T), threonine phosphorylatable by protein kinase C. (Reprinted with permission from Carafoli, 1992, *J. Biol. Chem.* **267**, 2115–2118.)

membrane within the SERCA pump (see below) have also been conserved in the PMCA pump. Consistent with the idea that the two pumps differ in the stoichiometry of Ca^{2+}/ATP was the finding that mutating Met882 of PMCA (M6) that corresponds to Thr799 (M6) in SERCA, and is known to be part of Ca^{2+} site II, had no influence on the activity of the PMCA pump. However, replacing Ala854 in PMCA (homologous to Glu771 in SERCA in M5) by a charged residue made the PMCA pump more similar to the SERCA pump with respect to the Ca^{2+}/ATP stoichiometry.

As typical for the P-type ion pumps, the Ca^{2+} pump of plasma membranes consists of a single polypeptide (ca. 1200 amino acids ≅135 kDa, depending on the isoform) spanning the membrane an even number of times (see Fig. 4); that is, the N- and the C-terminus are located on the same side of the membrane. This model of the PMCA pump gained validity owing to the recently solved structures of the Ca^{2+}-bound and the Ca^{2+}-free forms of the SERCA pump by Toyoshima et al., as will be discussed below. As demonstrated by this model, most of the pump mass protrudes into the cytosol with very short loops connecting the proposed transmembrane domains on the extracellular side. The cytosolic part of the enzyme can be divided into three different units.

1. Transduction or actuator (A) domain: It comprises the part of the pump between transmembrane domain 2 and 3, and, according to the structure of the SERCA pump, it should also contain the N-terminal sequence upstream of transmembrane domain 1. It is proposed to couple ATP hydrolysis to ion transport as in other ion pumps. An interesting part of this domain is the phospholipid binding domain that is unique to the plasma membrane Ca^{2+} pump as compared to other ion pumps, and that might play a role in specific tissue expression of isoforms since it was identified as one of the two alternative splicing zones of this protein (see below).

2. Catalytic domain: This unit spans the cytosolic part between the fourth and the fifth transmembrane domain (see Fig. 4) and, in analogy to the SERCA pump, contains the phosphorylation (P) and the nucleotide binding (N) domains. The P-domain contains the aspartyl phosphate site, the N-domain the ATP binding site, and the so-called ''hinge'' region, a highly conserved amino acid sequence among ion-pumping ATPases, which is involved in bringing the phosphorylation site near to the bound ATP. The catalytic unit further contains a receptor for the ''autoinhibitory'' calmodulin binding domain, a property that seems to be typical of CaM-dependent enzymes, and which differs PMCA from SERCA. This receptor is composed of two contact sites as identified by cross-linking experiments, one located between the P- and the N-domain, the other N-terminal to the phospholipid binding domain, thereby ''bridging'' the transduction and the catalytic domain and thus limiting the access of the substrate Ca^{2+} to the enzyme.

3. Regulatory domain: The sequence protruding into the cytosol after the last transmembrane domain until the C-terminus comprises several sites important for the regulation of the Ca^{2+} pump: the CaM binding domain and consensus sequences of two protein kinases, PKA and PKC (see Fig. 4). The PKA site is located C-terminal to the CaM binding domain, whereas the

PKC site is located within the CaM binding domain, and phosphorylation of this site can be prevented by the presence of CaM.

To date, four different genes of the PMCA pump have been identified in mammalian species. Additional variability of this multigene family is produced by alternative splicing within the phospholipid and the calmodulin binding domains. Splicing could give rise to differences in tissue-specific expression as well as to differences in affinity to calmodulin, for example, as documented for PMCA2, which has the highest affinity for calmodulin of all the isoforms studied to date. Four genes have been localized on human chromosomes: PMCA1 on chromosome 12, PMCA2 on chromosome 3, PMCA3 on chromosome X, and PMCA 4 on chromosome 1. PMCA1 and 4 are ubiquitously distributed, whereas PMCA2 and 3 are much more restricted in their tissue distribution, that is, the latter two are mainly found in the brain.

6.1.3 Na^+/Ca^{2+} Exchanger

Next to the Ca^{2+} pump, there also exists a Na^+-dependent Ca^{2+} exchanger in the plasma membranes of many cells, especially in excitable tissues. This Na^+/Ca^{2+} exchanger has a lower affinity for Ca^{2+} compared to the Ca^{2+} pump, but a much higher transport capacity. It is an electrogenic system; that is, it transports three Na^+ for one Ca^{2+}, the direction of the transport depending on the ionic transmembrane gradients or the transmembrane electrical potential, that is, the exchanger can operate fully reversible. Three homologous exchanger gene products (*NCX1, 2, 3*) have been identified; they are differently expressed in the heart, the brain, and the kidney. Hydropathy analysis has predicted 9 to 11 transmembrane domains, depending on the model, with the consequence that N- and C-terminus are located on opposite sides of the membrane. The size of the protein is between 110 and 120 kDa. Expression of the different gene products is controlled by tissue-specific promoters; some can be modulated by Ca^{2+}, for example, expression of *NCX2* in neurons is under the control of calcineurin, the CaM-dependent phosphatase.

6.2 Ca^{2+} Transport Systems of the Reticulum

Next to the Ca^{2+} transporting systems located in the plasma membrane, there are also intracellular organelles involved in controlling the free Ca^{2+} ion concentration of the cell. The reticular systems have best been studied in skeletal, smooth, and cardiac muscle cells, that is, the sarcoplasmic reticulum, but also the related endoplasmic reticulum of other cells gained importance owing to the knowledge that the Ca^{2+} content of these stores can be released into the cytosol by inositol-1,4,5-phosphate (IP_3). This latter second messenger links plasma membrane receptor activation response to extracellular hormonal stimuli to Ca^{2+} mobilization from intracellular stores.

6.2.1 The IP_3 Receptor

IP_3 is produced by receptor-activated phospholipase C–dependent hydrolysis of phosphatidyl inositol diphosphate (PIP_2). Subsequently, IP_3 binds to a specific receptor located in the membranes of the ER, thereby releasing calcium into the cytosol through the Ca^{2+}-ion channel of the receptor. The IP_3 receptor, which is highly concentrated in the Purkinje cells of the cerebellum, has been identified

and purified to a single protein band corresponding to 260 kDa. However, the molecular weight of the native receptor is about 1 million Da, indicating that the receptor is a homotetramer.

IP_3 receptors can be phosphorylated by different kinases (PKA, PKC, CaM kinase II) in a stoichiometric manner, resulting in a reduced potency of IP_3 in releasing Ca^{2+} from the ER. Thus, regulation of the IP_3-induced Ca^{2+} release by phosphorylation of the receptor is a means of "communication" between different second messenger systems.

Cloning and sequencing of cDNA of the IP_3 receptor was first accomplished by Mikoshiba et al. IP_3 receptors from different species revealed a high degree of conservation and homology. The sequence further demonstrates a high degree of similarity to the ryanodine-binding Ca^{2+}-release channel of the SR with which it also shares functional similarity. Like the latter, the main protein mass protrudes into the cytosol with both the N- and the C-terminus being on the cytosolic side, resulting in an even number of transmembrane domains. The IP_3 binding site lies within the first 400 amino acids from the N-terminus, whereas the putative Ca^{2+} channel is located within the transmembrane regions of the C-terminal end, indicating a large spatial distance for coupling the IP_3 binding to the Ca^{2+} release through a conformational change.

6.2.2 The Ca^{2+}-release Channel

The sarcoplasmic reticulum (SR) of striated muscles is an important feature involved in the regulation of excitation–contraction coupling of muscles. It is composed basically of two elements: (1) the longitudinal tubules surrounding the myofibrils and (2) the terminal cisternae. The latter are in contact with the transverse tubular system (T-tubules), a periodic inflection from the plasma membrane (the sarcolemma) forming a junctional gap that is crossed by periodic "feetlike" structures. These structures, originally described as the ryanodine receptor (RyR), have been identified as the calcium-release channel of the SR, forming tetramers as functional units. The amino acid sequence of the monomer, consisting of more than 5000 amino acids, has been deduced from the cDNA by the group of Numa. An unusual feature of these proteins is the existence of only four potential transmembrane domains right at the C-terminal end indicating that more than 90% of the protein protrudes into the cytosol. This, on the other hand, would provide an attractive morphological explanation for the existence of these "feetlike" structures spanning the 150-Å gap between the SR and the T-tubules and, furthermore, would provide the possibility for a physical interaction between the Ca^{2+}-release channel of the SR and the voltage-dependent Ca^{2+} channel concentrated in the T-tubules and identified as dihydropyridine receptors or L-type Ca^{2+} channels (see above). These features would provide a rational explanation for a direct triggering of Ca^{2+} release by these Ca^{2+} channels located in the T-tubules.

Three distinct RyRs have been identified. RyR1 is predominantly expressed in skeletal muscles, RyR2 in heart and brain, and RyR3 in some regions of the brain (e.g. hippocampus, diencephalon). Even if RyR was originally described as the Ca^{2+}-release channel of the sarcoplasmic reticulum, it is now well known that RyR is also widely distributed in nonmuscle cells. In these cells, cyclic ADP-ribose appears to be the natural second messenger for inducing Ca^{2+} release from the channel.

6.2.3 The Ca^{2+} Pump

The principle protein component of the sarcoplasmic reticulum (SR) is the Ca^{2+} pump (SERCA) transporting Ca^{2+} from the cytosol into the lumen of the reticulum against the concentration gradient across the membrane. This protein can represent as much as 90% of the membrane protein (in SR of skeletal muscles), but even in the SR of heart cells, it can still make up to 50%. Similar to the plasma membrane Ca^{2+} pump, the protein consists of a single polypeptide chain of about 100 kDa. The predicted secondary structure and membrane topography as suggested by MacLennan, Green et al. was impressively confirmed by the recently solved structures of the Ca^{2+}-bound and Ca^{2+}-free forms (the latter in the presence of the SERCA-specific inhibitor thapsigargin) by Toyoshima et al. (see Fig. 5). The properties of the SR Ca^{2+} pump can be summarized as follows:

1. The hydrophobic portion is made up of 10 transmembrane helices as predicted.
2. Transmembrane helix M5 continues into the cytosol (60 Å total length) coupling the membrane portion to the cytosolic domains.
3. The major protein mass protrudes into the cytosol divided into the actuator (A), phosphorylation (P), and the nucleotide binding (N) domains, the latter being inserted within the P-domain (see Fig. 5). The P- and N-domains, connecting transmembrane helices 4 and 5, provide the catalytic center of the enzyme, with the conserved aspartyl residue (D351) being phosphorylated during the reaction cycle.
4. Two Ca^{2+} binding sites could be identified within the transmembrane part of the SERCA pump comprising polar residues of helices M4, M5, M6, and M8. This is a necessary prerequisite for the stoichiometry of two calcium ions being pumped/ATP during one reaction cycle, and confirms strongly those residues participating in the transport of calcium as identified by mutational experiments.
5. Large-scale movements of the A-, P-, and N-domains occur during the reaction cycle as suggested by comparing the Ca^{2+}-free with the Ca^{2+}-bound structure. These domain movements are coupled to changes in tilt and position of helices M1 to M6.
6. One of the significant differences between the two Ca^{2+} pumps of plasma and reticular membranes concerns the C-terminal part. In contrast to the plasma membrane Ca^{2+} pump, the SR protein does *not* directly interact with calmodulin; that is, it lacks the corresponding regulatory domain, but some isoforms can be regulated by a CaM kinase through phosphorylation of phospholamban (PLN) (see below).
7. The SR Ca^{2+} pump of cardiac or smooth muscles, that is, of slow twitch muscles, is regulated by a highly hydrophobic, phosphorylatable protein called *phospholamban* (PLN), composed of an N- and a C-terminal helix connected by a short, flexible hinge region. PLN interacts with the SERCA pump within the transmembrane region as well as with a small loop of the P-domain. PLN can be phosphorylated by different kinases, that is, by PKA and by a CaM-dependent kinase, thereby leading to a significant stimulation (up to fivefold) of the SR Ca^{2+} pump. Thus, the unphosphorylated form of phospholamban can be viewed as an endogenous inhibitor of the SR Ca^{2+} pump similar to the CaM binding domain of the Ca^{2+} pump of plasma

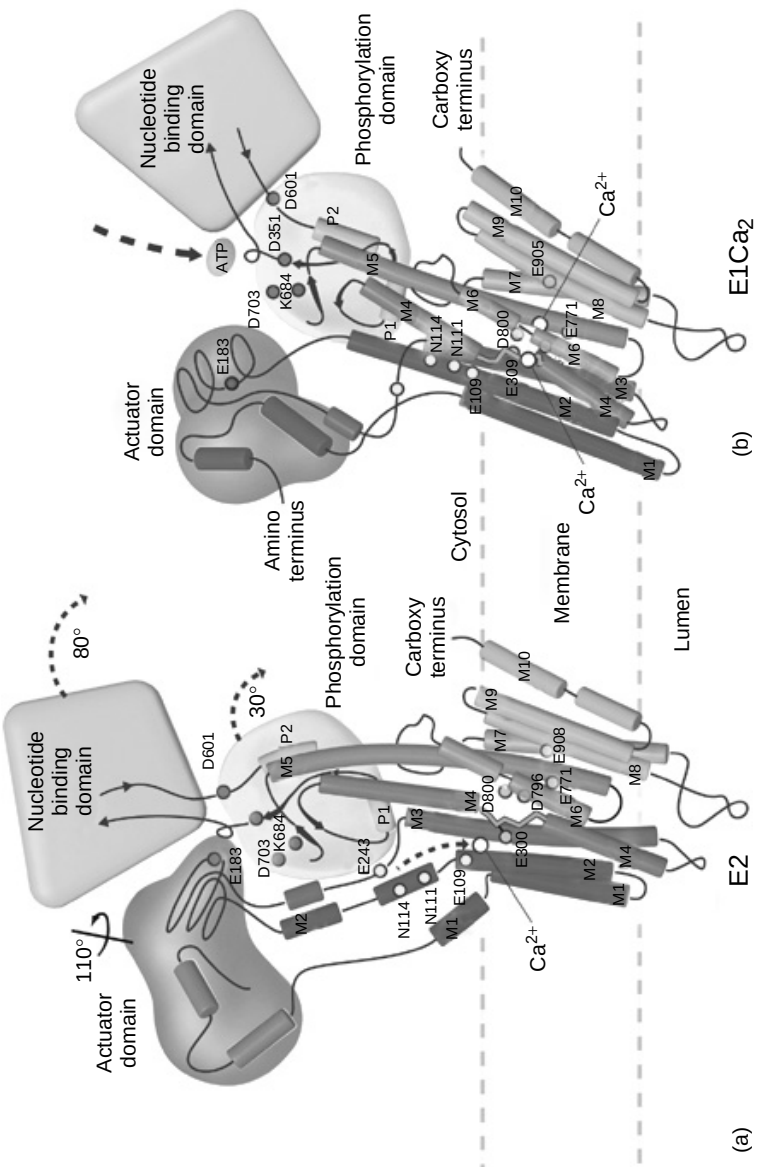

Fig. 5 Structures of the SERCA pump in (a) the Ca^{2+}-free (thapsigargin-bound) form and (b) the Ca^{2+}-bound form. The three cytosolic domains are indicated as A (actuator), P (phosphorylation), and N (nucleotide binding) domains. The P-domain (with N inserted) connects transmembrane helices M4 and M5. Within the P-domain, Asp351, which is phosphorylated during the reaction cycle, is also indicated. A number of residues important for the catalytic center or the Ca^{2+} binding sites within the membrane are shown (Reprinted with permission from Green & MacLennan, 2002, *Nature* **418**, 598–599).

membranes, which is an autoinhibitory domain.

Three different genes have been identified to date to encode the SERCA pump. SERCA1 is mainly found in the fast-twitch skeletal muscle, whereas SERCA2 is the major form of cardiac, smooth, and nonmuscle tissues. SERCA3 has been cloned from a kidney library and is prominently present in platelets. For SERCA1 and 2, several spliced isoforms also have been described.

6.3 The Mitochondrial Calcium Transporting Systems

Like the reticular systems, mitochondria also provide intracellular calcium-transporting systems sequestering Ca^{2+} by means of energy-dependent processes. The mitochondrial calcium transporters, that is, a Ca^{2+} uniporter and a Na^{+}/Ca^{2+} exchanger, were thought, for a long time, to play a minor role in the constant regulation of the cytosolic Ca^{2+} concentration since the mitochondrial Ca^{2+} uptake rate is about 10-fold slower than that of the SR, and the K_M value for the transporter is 10-fold higher than that of the reticular systems. Using Ca^{2+}-specific indicators targeted to different organelles, it became clear that mitochondria were very important for sensing Ca^{2+} changes in "microdomains" of the cell, rather than in the bulk cytosol; that is, increase of cytosolic Ca^{2+} induced by IP_3 was paralleled by a rapid increase of mitochondrial Ca^{2+}. It soon became clear that the activation of mitochondrial Ca^{2+} uptake was the result of the proximity of the mitochondria and the ER, and the release of large amounts of Ca^{2+} by the latter could create microdomains of local high Ca^{2+} concentrations, sufficient to activate the low-affinity mitochondrial Ca^{2+} uniporter. In spite of concentrated studies, however, the molecular nature of this mitochondrial Ca^{2+} transporting system could not be identified to date.

6.4 Calcium in the Nucleus

The nucleus is separated from the cytosol by the nuclear envelope consisting of two concentric membranes, of which the outer membrane is continuous with the ER. The nuclear envelope is not a completely closed system but is interrupted by the nuclear pores, large multicomponent protein complexes that regulate the transport into and out of the nucleus. Whether Ca^{2+} can freely diffuse through these pores like most small molecules or whether it is gated is still controversial, but it is clear that there are important Ca^{2+}-dependent processes occurring in the nucleus, including Ca^{2+}-dependent gene expression. A key factor of Ca^{2+}-dependent gene expression (including a number of Ca^{2+} transporting systems) is the transcription factor CREB, originally identified as the protein recognizing cAMP-responsive elements. Thus, CREB can be activated by phosphorylation of a serine residue, Ser133, either by cAMP-dependent PKA or by a CaM-dependent protein kinase, CaMKII or IV. In this respect, a recent finding is of particular interest, showing that the mode of entry of Ca^{2+} into cells, that is, via the voltage-dependent calcium L-channel or through other means, specifies the response to Ca^{2+}-dependent gene transcription.

Another pathway of Ca^{2+}-dependent gene transcription recently attracted much attention, that is, the transcription factor NFAT that has originally been described as being important in T cells and that is under

the control of the CaM-dependent phosphatase calcineurin. Similar findings have recently been reported for neurons, where calcineurin regulates the gene expression of some of the Ca^{2+} transporters. This is especially true for the cerebellum, where calcineurin mediates the rapid transcriptional downregulation of specific isoforms of the PMCA pumps and of the NCXs.

Ca^{2+} can also regulate transcription directly through the recently identified transcription factor DREAM (**d**ownstream **r**egulatory **e**lement **a**ntagonist **m**odulator), which is an EF-hand-type protein. Ca^{2+}-free DREAM binds to the regulatory element DRE, thereby silencing the gene that is released owing to the binding of Ca^{2+} to DREAM.

7 Role of Calcium in Specialized Tissues

7.1 Calcium and Fertilization

Interaction of sperm with the egg upon fertilization results in a sharp increase in Ca^{2+} concentration, thereby depolarizing the plasma membrane. Subsequently, phospholipase C is activated and phosphoinositide lipid turnover is induced leading to IP_3 generation, a prerequisite to release Ca^{2+} from intracellular stores. Also, cyclic ADP-ribose/RyR channels and NAADP-sensitive Ca^{2+} stores contribute to Ca^{2+} mobilization. Several mechanisms are discussed to be responsible for the elevation of Ca^{2+} during fertilization; the most likely candidate is of proteinaceous nature. Nevertheless, irrespective of the mechanism, the immediate consequence of Ca^{2+} rise is the exocytosis of cortical granules responsible for the elevation of the fertilization envelope to prevent polyspermy.

7.2 Calcium in Calcifying Tissues

Bone is formed by osteoblasts of mesenchymal origin and resorbed by osteoclasts of hematopoietic stem cell origin. In this highly integrated process, it is important to maintain a balance between the formation and resorption of bone. If one part dominates over the other, leading either to osteopetrosis, that is, increase of bone formation, or to osteoporosis, that is, increase of bone resorption, one of the consequences is the imbalance of the calcium homeostasis. Therefore, it is important for a variety of cells to sense changes in the extracellular Ca^{2+} levels to be able to modify their functions accordingly. The existence of such calcium sensors has first been demonstrated for the parafollicular cell of the thyroid gland (C-cell) and for the parathyroid cell, which through the secretion of parathyroid hormone regulate the level of serum calcium, that is, responding to changes in PTH secretion inversely related to the ambient ionized calcium concentration. These cells can recognize small changes in the extracellular Ca^{2+} concentration through a "Ca^{2+} receptor," presumably a cell-surface receptor protein, which in the case of cytotrophoblast cells of the human placenta has been identified as a protein of 500 kDa. In the case of osteoclasts that can locally be exposed up to 30 mM Ca^{2+} during bone resorption, increase in extracellular Ca^{2+} concentration leads to an increase in intracellular Ca^{2+} resulting mainly from influx of extracellular Ca^{2+} through voltage-insensitive channels. This in turn leads to a modification of the cytoskeletal arrangement of these highly polarized cells, to a reduction of specific adhesion structures responsible for the attachment to bone, and, finally, to an

inhibition of bone resorption accompanied by a reduction in the secretion of resorptive enzymes.

7.3 The Role of Calcium in Plants

Most Ca^{2+} dependent signal transduction pathways as described for animal cells are also true for plant cells. So it is not surprising that most systems described before controlling calcium homeostasis found their counterparts in plants also. Nevertheless, there are some peculiar Ca^{2+}-dependent processes or properties, only found in plants, which should be described in more detail.

7.3.1 Touch Genes

Plants show some specific reactions that seem to be calcium-dependent; that is, they sense and actively respond to many environmental stimuli such as rain, wind, gravity, changes in osmotic conditions, and touch. The latter is demonstrated very impressively by the Venus fly trap (*Dionaea muscipula*) using a sensitive touch response to catch preys, or by the *Mimosa pudica* closing the leaflets upon touch stimuli. It was discovered by Braam and Davies that the plant *Arabidopsis* demonstrated a rapid and strong induction of gene expression following touch stimulation. Several of these genes revealed either virtual identity or close relationship to calmodulin, indicating that these gene products may play a central role in Ca^{2+}-dependent signal transduction pathways involved in touch response. The response is relatively fast; for example, all of the responding genes showed up to 100-fold increased expression after 10 min, but maximal expression varied for the different genes. Furthermore, these genes not only demonstrated a time-dependent but also a dose-dependent response to touch stimuli; that is, the level of transcripts of the touch genes was proportional to the strength of the stimuli. These findings indicated that there is an immediate, but transient, requirement for Ca^{2+} binding proteins following mechanical stimulation, resulting in growth inhibition. It is possible that these proteins act as Ca^{2+}-dependent modulators of target enzymes, but other explanations are also possible such as involvement in the regulation of cellular Ca^{2+} homeostasis or changing microtubular arrangements. The latter is particularly attractive since this activity could influence the orientation of the axes of cell division and elongation as a consequence of mechanical stimulation.

7.3.2 Calmodulin-like Domain Protein Kinase (CDPK)

In plants, a new class of calcium-dependent kinases has been described. These enzymes require micromolar Ca^{2+} concentrations for activity, but they do not seem to need the presence of calcium effector molecules such as calmodulin, phosphatidylserine, or diacylglycerol. It could be demonstrated that they bind Ca^{2+} directly through a regulatory domain on the same polypeptide chain containing the catalytic domain. Cloning of the corresponding cDNA encoding a calmodulin-like domain protein kinase (CDPK) revealed characteristic regions of serine/threonine kinases. In addition, the regulatory region contains a domain with a high degree of homology to calmodulin including four Ca^{2+} binding sites of the EF-hand-type family. Thus, the CDPK family can be distinguished from other calcium-regulated protein kinases by fusion of a regulatory, calcium binding, calmodulin-like domain to a protein kinase catalytic domain. Investigation of the substrate specificity of this

kinase revealed an overlapping with CaM kinase II and protein kinase C. Together with the ubiquitous distribution of CDPK in plants, this observation suggests a multifunctional kinase property for this enzyme, but specific endogenous substrates still have to be identified to characterize the role of CDPK in calcium signal transduction pathways in plants.

8 Calcium and Disease

Sustained raise of cellular Ca^{2+} into the micromolar range leads to deleterious effects upon Ca^{2+} signaling, which could eventually lead to cell death. Mutations of the intracellular Ca^{2+}-dependent protease, calpain, the activity of which normally is tightly controlled – like disruption of the gene responsible for a muscle isoform – cause muscular dystrophy type 2A, whereas mutations of another isoform lacking the C-terminal Ca^{2+} binding domain favor the onset of type 2 diabetes.

As indicated before, bone formation and bone resorption is a highly integrated process that, in cases of misbalance, leads to severe diseases (e.g. osteopetrosis, osteoporosis, Paget's disease) and thereby influences calcium homeostasis. The reason for these malfunctions can be manifold and will not be discussed here, but recently two examples became known in which disturbance of calcium homeostasis could be traced to mutations of single proteins involved in Ca^{2+} regulation.

Malignant hyperthermia is manifested in humans and swine as an acute hyperthermic reaction accompanied by skeletal muscle contracture and high fever, indicating a lack of regulation of Ca^{2+} within muscle cells. This could be due to either an enhanced or chronic release of Ca^{2+} into the cell or due to a defective reuptake of Ca^{2+}. Detailed studies by MacLennan and his laboratory implicated a defect in the Ca^{2+} release mechanism that could be traced to a number of mutations within the gene of the ryanodine-sensitive Ca^{2+}-release channel in the sarcoplasmic reticulum of skeletal muscle. To date, more than 20 different mutations linked to this defect have been described in the *RyR1* gene.

Defects in the genes of the two brain-specific isoforms of the plasma membrane Ca^{2+} pump, PMCA2 and 3, have also been described. Of special note is PMCA2 that is abundant in the outer hair cells of the organ of Corti. Recent findings in mice with hearing defects localized genetic mutations within PMCA2, which could be used as a model for investigating hearing defects in humans.

Vitamin D–resistant rickets is an autosomal recessive disorder resulting from an inability to regulate Ca^{2+}, which leads to bone demineralization, loss of hair, and increased levels of the active vitamin D metabolite 1,25 $(OH)_2D_3$. As shown by Hughes et al., the reason for these defects lies in a number of inherited mutations of the vitamin D receptor, a member of the steroid/thyroid family of hormone-sensitive transcription factors. The mutations identified so far fall either into the DNA binding or into the hormone binding domain.

9 Conclusions

In summary, several general points can be made:

1. Calcium, as one of the oldest components of organisms, plays a central

role in biological systems, controlling a myriad of key cell processes. It can fulfill a static function, in stabilizing structures, or a dynamic function, participating in signal transduction pathways as a second messenger.

2. Calcium homeostasis in an organism is carefully controlled, involving a variety of systems in the skeleton, in the ECF, and inside cells. Depending on its function, calcium can be complexed in different forms: by hydroxyapatite in the skeleton; by acidic, low-affinity proteins in the ECF; by the high-affinity EF-hand proteins inside cells.
3. Extracellular and intracellular concentrations of calcium differ by several orders of magnitude. Therefore, cells are exposed to a steep Ca^{2+} gradient across the membranes, which makes it possible for even small changes in membrane permeability to lead to substantial changes in intracellular free Ca^{2+} concentration. Signals can be converted from an extracellular analog to an intracellular digital form. The control of cellular calcium is maintained by an elaborate system of channels, exchangers, and pumps located in the plasma membrane and in intracellular membranes.
4. The EF-hand proteins play a pivotal role in permitting Ca^{2+} to function as a second messenger. These proteins bind Ca^{2+} with high affinity, selectivity, and cooperativity, thereby permitting interaction with targets. More than 600 EF-hand proteins have been identified to date, which fulfill the different tasks of calcium-dependent mechanisms (e.g. glycogen metabolism, muscle contraction, excitation–secretion coupling, cell cycle control, gene expression, mineralization).
5. Solving the structure of several key components of calcium-controlled pathways helped understand their mechanism of action.

See also Bioinorganic Chemistry; Cellular Interactions.

Bibliography

Books and Reviews

Carafoli, E., Klee, C. (1999) *Calcium as a Cellular Regulator*, Oxford University Press, New York.

Carafoli, E., Krebs, J. (2000) *Calcium Homeostasis*, Topics in Biological Inorganic Chemistry, Vol. 3, Springer, Berlin.

Carafoli, E., Santella, L., Branca, D., Brini, M. (2001) *Generation, control, and processing of cellular calcium signals*, *Crit. Rev. Biochem. Mol. Biol.* **36**, 107–260.

Krebs, J. (1991) Calcium, Biochemistry, in: Dulbecco, R. (Ed.) *Encyclopedia of Human Biology*, Vol. 2, Academic Press, San Diego, pp. 89–99.

Krebs, J., Guerini, D. (1996) The Calcium Pump of Plasma Membranes, in: Lee, A.G. (Ed.) *Biomembranes*, Vol. 5, Jai Press, Greenwich, pp. 101–131.

Krebs, J. (1998) The role of calcium in apoptosis, *BioMetals* **11**, 375–382.

Pochet, R. (2000) *Calcium. The Molecular Basis of Calcium Action in Biology and Medicine*, Kluwer Academic Publishers, Dordrecht.

Primary Literature

Babu, Y.S., Sack, J.S., Greenhough, T.J., Bugg, C.E., Means, A.R., Cook, W.J. (1985) Three-dimensional structure of calmodulin, *Nature* **315**, 37–40.

Brown, E.M., Gamba, G., Riccardi, D., Lombardi, M., Butters, R., Kifor, O., Sun, A., Hediger, M.A., Lytton, J., Hebert, S.C. (1993) Cloning and characterization of an extracellular Ca(2+)-sensing receptor from bovine parathyroid, *Nature* **366**, 575–580.

Elshorst, B., Hennig, M., Forsterling, H., Diener, A., Maurer, M., Schulte, P., Schwalbe, H., Griesinger, C., Krebs, J., Schmid, H., Vorherr, T., Carafoli, E. (1999) NMR solution structure of a complex of

calmodulin with a binding peptide of the Ca^{2+} pump, *Biochemistry* **38**, 12320–12332.

Furuichi, T., Yoshikawa, S., Miyawaki, A., Wada, K., Maeda, N., Mikoshiba, K. (1989) Primary structure and functional expression of the inositol 1,4,5-trisphosphate-binding protein P400, *Nature* **342**, 32–38.

Goldberg, J., Nairn, A.C., Kuriyan, J. (1996) Structural basis for the autoinhibition of calcium/calmodulin-dependent protein kinase I, *Cell* **84**, 875–887.

Herzberg, O., James, M.N. (1985) Structure of the calcium regulatory muscle protein troponin-C at 2.8 A resolution, *Nature* **313**, 653–659.

Ikura, M., Clore, G.M., Gronenborn, A.M., Zhu, G., Klee, C.B., Bax, A. (1992) Solution structure of a calmodulin-target peptide complex by multidimensional NMR, *Science* **256**, 632–638.

James, P., Maeda, M., Fischer, R., Verma, A.K., Krebs, J., Penniston, J.T., Carafoli, E. (1988) Identification and primary structure of a calmodulin binding domain of the Ca^{2+} pump of human erythrocytes, *J. Biol. Chem.* **263**, 2905–2910.

Kretsinger, R.H. (1975) Hypothesis: Calcium Modulated Proteins Contain EF-Hands, in: Carafoli, E., Clementi, F., Drabikowski, W., Margreth, A. (Eds.) *Calcium Transport in Contraction and Secretions*, Elsevier, Amsterdam, pp. 469–478.

Kretsinger, R.H., Nockolds, C.E. (1973) Carp muscle calcium-binding protein. II. Structure determination and general description, *J. Biol. Chem.* **248**, 3313–3326.

MacLennan, D.H., Brandle, C.J., Korczak, B., Green, N.M. (1985) Amino-acid sequence of a Ca^{+} Mg^{2+}-dependent ATPase from rabbit muscle sarcoplasmic reticulum, deduced from its complementary DNA sequence, *Nature* **316**, 696–700.

Meador, W.E., Means, A.R., Quiocho, F.A. (1992) Target enzyme recognition by calmodulin: 2.4 Å structure of a calmodulin-peptide complex, *Science* **257**, 1251–1255.

Moldoveanu, T., Hosfield, C.M., Lim, D., Elce, J.S., Jia, Z., Davies, P.L. (2002) A Ca(2+) switch aligns the active site of calpain, *Cell* **108**, 649–660.

Strobl, S., Fernandez-Catalan, C., Braun, M., Huber, R., Masumoto, H., Nakagawa, K., Irie, A., Sorimachi, H., Bourenkow, G., Bartunik, H., Suzuki, K., Bode, W. (2000) The crystal structure of calcium-free human m-calpain suggests an electrostatic switch mechanism for activation by calcium, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 588–592.

Toyoshima, C., Nakasako, M., Nomura, H., Ogawa, H. (2000) Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution, *Nature* **405**, 647–655.

Toyoshima, C., Nomura, H. (2002) Structural changes in the calcium pump accompanying the dissociation of calcium, *Nature* **418**, 605–611.

Verma, A.K., Filoteo, A.G., Stanford, D.R., Wieben, E.D., Penniston, J.T., Strehler, E.E., Fischer, R., Heim, R., Vogel, G., Mathews, S., Strehler-Page, M.-A., James, P., Vorherr, T., Krebs, J., Carafoli, E. (1988) Complete primary structure of a human plasma membrane Ca^{2+} pump, *J. Biol. Chem.* **263**, 14152–14159.

Cancer Chemotherapy, Theoretical Foundations of

Nicholas B. La Thangue
Division of Biochemistry and Molecular Biology, Davidson Building, University of Glasgow, Glasgow, G12 8QQ, UK

1 **Cancer Therapy 173**

2 **The New Era of Mechanism-based Drug Design 173**

3 **Gleevec and CML 174**

4 **The ErbB Family 174**

5 **Manipulating p53 Tumor Suppressor Activity 176**

6 **Regulating the Cancer Cell Cycle: E2F and Cdks 177**

7 **Targeting Chromatin Control in Cancer 179**

8 **Regulating Hsp90 180**

9 **Blocking Angiogenesis Through Inhibiting HIF1 Activity 182**

10 **Regulating Protein Turnover 184**

11 **Conclusions and Perspectives 184**

Acknowledgment 186

Bibliography 186
Books and Reviews 186
Primary Literature 186

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Keywords

Cancer
One of the most common human diseases.

Drug Design
The process of developing a new medicine.

Therapy
The administration to human subjects of a drug

p53
A tumour suppressor protein seen to be frequently mutated in cancer cells.

Cell Cycle
The process through which cells grow and divide.

Angiogenesis
The formation of new blood vessels.

Hsp90
A ubiquitously expressed protein that acts to protect cells from thermal stress.

Erb B
The epidermal growth factor receptor.

E2F
A key protein transcription regulating protein involved in cell cycle control.

The massive research efforts that have been applied to understanding the workings of the cancer cell have provided a wealth of detail on key mechanisms that become abnormal during tumorigenesis. Translating and exploiting this large body of cancer research information in the design and implementation of new drugs should yield better medicines for the cancer patient than those that are currently available. It is an implicit assumption in this objective that mechanism-based agents that target discrete abnormalities in tumor cells will provide improved treatments over conventional therapies. In this respect, we are beginning to witness emerging clinical efficacy in a new generation of cancer drugs designed to target key mechanisms in tumor cells.

1 Cancer Therapy

Cancer remains largely a clinically unmet disease. Statistics tell us that cancer affects one person in three, and in the Western world, it is the cause of about a quarter of all mortality. Worldwide, cancer is the second largest cause of death, and the World Health Organization estimates that by 2020, there will be 20 million new cancer patients each year. Cancer is not a single disease, as there are over 200 different types of diseases that fall into this category. However, four cancers dominate, namely, lung, breast, colon, and prostate, which account for over half of all new cases.

Most current cancer therapies employ one of three approaches, namely, surgery, radiotherapy, or chemotherapy. Surgery and radiotherapy are frequently used to treat localized primary tumors, but have limited application in disseminated disease in which the tumor is not localized, usually as a result of metastasis. Needless to say, disseminated disease is the cause of most cancer deaths. Current cancer chemotherapy is dominated by treatments that use cytotoxic agents or hormone-based therapy, for example, in breast and prostate cancer (Table 1).

Tab. 1 Examples of conventional cancer chemotherapeutic agents.

Drug	*Category*
Methotrexate	Antimetabolite
Taxol	Spindle modulator
Cyclophosphamide	Alkylating agent
Cisplatin	Platinum-DNA complexes
Doxorubicin	DNA intercalating/topoisomerase inhibitor

Furthermore, while advances have been made in all three approaches to treatment, the impact on mortality rates has been modest. As an example, lung cancer, the leading cause of cancer death worldwide, has a five-year survival rate of only 5%, which hardly differs from the survival that prevailed 30 years ago (www.cancerresearchuk.org). However, there are some notable improvements. Cures are achievable in certain childhood leukemias and testicular cancer, although of the six most common cancers, only breast cancer has a survival rate greater than 50%. Overall, conventional chemotherapy is relatively ineffective in many cancers.

This situation reflects the mechanism of action of most current cancer therapy regimes, which are dominated by cytotoxic chemotherapies (Table 1). In general, these chemotherapies target mechanisms employed by most dividing cells, rather than specific genetic abnormalities associated with tumor cells. Because of this, their normal healthy counterparts are affected almost as severely as the tumor cells, leading to the highly debilitating side effects seen with many therapies, such as myelosuppression, hair loss, and gastrointestinal toxicity. This problem, combined with the poor quality of life and diminishing effectiveness in the patient (usually because of the emergence of drug resistance), means that many late stage cancer treatments provide minimal survival advantage.

2 The New Era of Mechanism-based Drug Design

Because of the limited therapeutic value of many current treatments, major efforts are being applied toward identifying new cancer therapies that target

specific mechanisms in tumor cells, in the hope that this approach will yield drugs with increased specificity for cancer. Here, we review progress and provide a snapshot of this rapidly advancing field. Selected examples of mechanism-based drug design programs, which illustrate the guiding principles behind the new era of cancer drug development, are discussed. The success stories are clear; a situation we anticipate will be repeated in the continuing efforts to deliver tailored therapies into the cancer clinic.

3
Gleevec and CML

Gleevec (ST1571) provides a paradigm for a successfully developed, rationally designed, and mechanism-based therapy for the treatment of a specific cancer. In chronic myelogenous leukemia (CML), the consistent chromosomal abnormality, the Philadelphia (Ph) chromosome, is a translocation event between chromosome 9 and 22 that results in the bcr-*abl* fusion product derived from the juxtaposition of c-*abl* oncogene and breakpoint cluster region (bcr). bcr-*abl* is present in over 95% of CML patients and in about 5% of acute lymphoblastic leukemia (ALL) patients in whom it functions as a constitutively active tyrosine kinase that is essential for the oncogenic activity of the fusion protein (Fig. 1).

Originally, Gleevec was identified by high-throughput screening for compounds with kinase inhibitory activity, and was subsequently optimized using rational drug design procedures for activity against Abl tyrosine kinase in CML cells. A successful Phase I study indicated remarkable single agent activity during CML blast crisis (and Ph-positive ALL), although the responses did not appear to be long term. Nevertheless, the success in Phase I prompted further Phase II clinical analysis, followed by FDA approval.

The appearance of resistance to Gleevec was a surprising clinical outcome: the majority of CML patients that relapse after initial response have reactivated bcr-*abl* kinase. Molecular analysis indicated that in many cases resistant cells carry point mutations in the *abl* kinase gene that rendered the kinase less sensitive to STI571, for example, at a site that would be predicted to contact STI571 from the crystal structure. In another category of patients, relapse was correlated with amplification at the bcr-*abl* translocation. These results suggest that while rationally designed drugs like Gleevec may exhibit promising clinical efficacy, the appearance of resistance could be a significant problem.

Gleevec also inhibits the c-kit tyrosine kinase where activating mutations are found in gastrointestinal stromal tumors (GIST), a tumor frequently refractory to chemotherapy. While a range of other tumors express c-kit, the encouraging results of the Gleevec Phase I study on GIST suggest that its clinical application may be wider than initially anticipated, with the proviso that clinical resistance may also appear in other indications.

4
The ErbB Family

The epidermal growth factor (EGF) family of tyrosine kinase receptors (the ErbB receptors) play a crucial role in regulating proliferation. In tumor cells, a variety of mutagenic events give rise to increased ErbB activity, including gene amplification and mutations that alter protein stability.

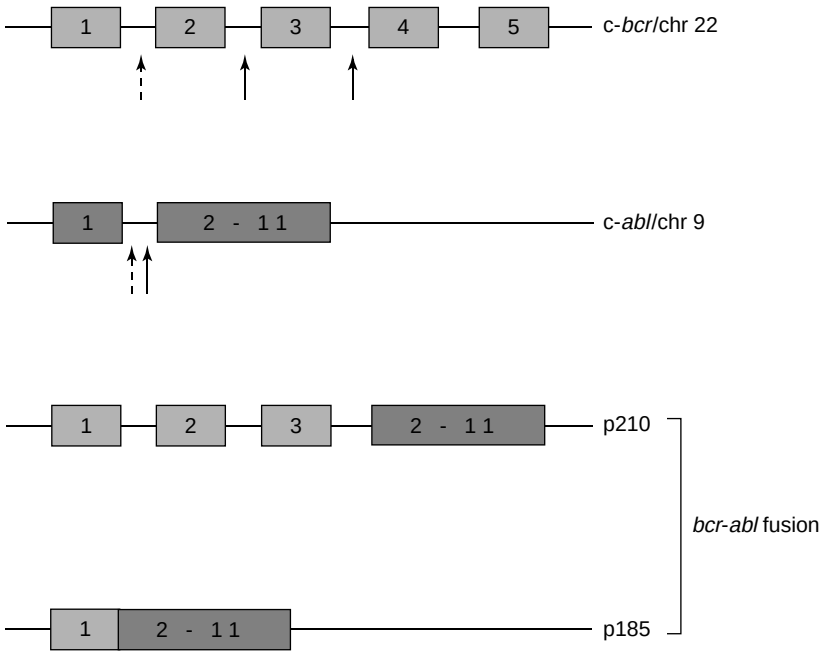

Fig. 1 Chromosomal breakpoints in chronic myelogenous leukemia (CML) and acute lymphoblastic leukemia (ALL). In CML and ALL, a breakpoint occurs between the first exon of c-*abl* on chromosome 9, translocating most of c-*abl* to c-*bcr* on chromosome 22. In CML, the translocations in c-*bcr* occur after exon 2 or 3, to yield the p210 fusion protein, and in Ph-positive ALL, they occur after exon 1 to create p185. The solid arrows indicate CML breakpoints and broken arrows indicate ALL breakpoints.

Both humanized monoclonal antibodies and small molecules have been developed against ErbB1 (also known as EGF receptor and HER1) and ErbB2 (also known as HER2 and Neu). Herceptin (trastuzumab) has received regulatory approval as a single agent for treatment of metastatic breast cancers expressing ErbB2, whereas other humanized antibodies against ErbB1 such as Cetuximab remain in clinical development.

Iressa (ZD1839) is an optimized small molecule ATP competitive antagonist of the ErbB1 tyrosine kinase, originally identified by high-throughput screening and subsequently shown to be active against breast cancer cells. Phase II clinical trials with Iressa demonstrated activity in nonsmall cell lung cancer (NSCLC), squamous cell carcinomas of the head and neck, and hormone refractory prostate cancer. Recently, Iressa was evaluated in a randomized Phase II study for activity in advanced NSCLC, where it was found that it did not provide improvement in survival when added to standard platinum-based chemotherapy.

A significant question regarding the clinical utility of mechanism-based drugs like Iressa relates to how they will be used in the clinical setting. A relevant administration schedule and, as mechanism-based drugs may be cytostatic, identifying an appropriate combination agent and patient population are likely to be critical determinants in reaching optimal clinical efficacy.

5 Manipulating p53 Tumor Suppressor Activity

As one of the most frequently mutated genes in human cancer, the p53 tumor suppressor protein has attracted great interest from drug discoverers. In normal cells, p53 acts as a stress-responsive transcription factor, becoming activated in response to, for example, ionizing radiation, ultraviolet light, and many cancer chemotherapeutic drugs. Once activated, p53 targets genes involved in limiting cellular proliferation, causing either cell cycle arrest or apoptosis (Fig. 2). Its frequent inactivation in cancer has led to many attempts to identify ways of reinstating p53 activity, or approaches that exploit the absence of p53 in tumor cells. However, though it is an intrinsically difficult task for a small molecule to reinstate wild-type activity in a mutant protein, the extensive number of distinct mutations (over 1700 different mutations have been reported) that can occur in the p53 gene provides an additional challenge to this approach.

Despite these significant obstacles, compounds have been identified that reinstate certain wild-type functions on mutant p53 protein. These compounds appear to stabilize p53 in a form that allows some properties of the p53 response to be retained. Relating to this general strategy is the identification of p53-binding peptides, which similarly stabilize mutant p53, providing further support for the idea that the ultimate goal of generating compounds that reinstate p53 activity may be achievable.

In another approach that exploits p53 activity, attempts have been made to generate compounds that block p53 activity. A potential therapeutic application for such inhibitors would be in reducing the level of p53-dependent toxicity in normal healthy tissues that occurs, for example, upon cancer therapy. Pifithrin α, which reduces p53 activity, was identified by high-throughput screening in a cell-based p53 screen. The

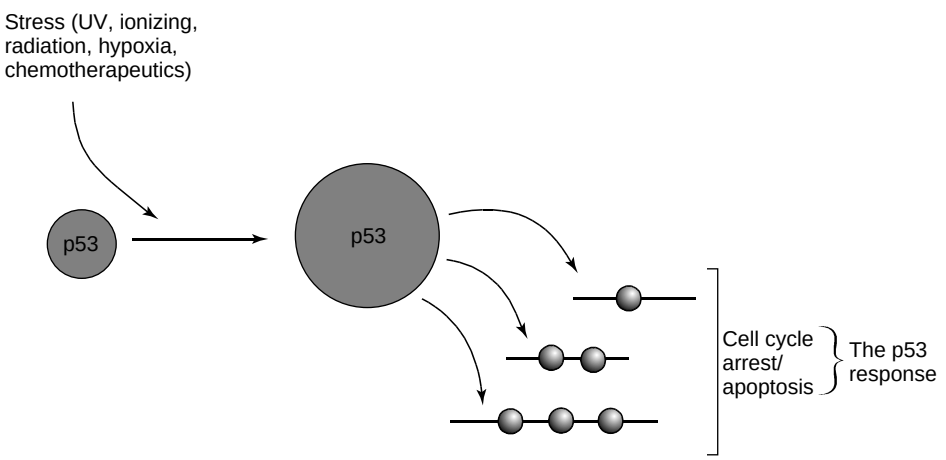

Fig. 2 The p53 response. In normal cells, p53 can be activated by different types of stress, and it then acts to transcriptionally induce target genes that cause cell cycle arrest and apoptosis. In tumor cells, the activity of p53 is lost through mutation in the p53 gene, or in upstream regulators that control p53 activity.

compound is active in mice where it reduces the side effects, such as hair loss and myelosuppression, of the widely used cancer cytotoxic agent doxorubicin.

Although the anticipated clinical utility of p53 inhibitor drugs would be in combination treatments to enhance the efficacy of a conventional cytotoxic, a shortcoming of this approach could lie in the potential mutagenic activity resulting from p53 inhibition. Developing a p53 inhibitor with a sufficiently short half-life may overcome such a concern.

6 Regulating the Cancer Cell Cycle: E2F and Cdks

Abnormal cellular growth and division is a hallmark of tumor cells. It is acquired through a multistep process involving the gradual transformation of a normal cell to a tumor cell, and it involves mutation of critical pathways that act to restrain proliferation, together with the activation of those that promote proliferation.

The retinoblastoma tumor suppressor protein (pRb) provides a perfect example of how a crucial point of growth control can become abnormal in tumor cells, and thereby contribute to tumor cell growth. An important function of pRb is in regulating cell cycle progression through G1 into S phase, particularly passage through the restriction point, where it influences the activity of the E2F family of transcription factors (Fig. 3). E2F coordinates the timely expression of a large body of genes required for cell cycle progression (including DNA synthesis, cyclins, and regulatory proteins), and it is the progressive phosphorylation of pRb that relieves its physical interaction with E2F, thereby enabling E2F to activate target genes (Fig. 3).

Tumor cells employ diverse mechanisms to overcome the tumor suppressor activity of pRb. The Rb gene is mutated in about 25% of human tumors, crippling pRb so that it is not able to bind to E2F. Alternatively, aberrant levels of upstream regulators of pRb, like cyclin D-dependent kinase, occur which leads to the constitutive phosphorylation of pRb and release of E2F. Because of the high frequency of abnormal pRb control, it is believed that deregulation of the pRb/E2F pathway is essential in achieving the transformed state.

In this respect, E2F is an obvious target for therapeutic intervention, and several approaches have been made toward its validation as a cancer target. That E2F plays a crucial role in the cell cycle is supported by E2F knockout cells, in which inactivation of E2F-1, -2 and -3 causes cell cycle arrest. Furthermore, administering short oligonucleotides containing the E2F DNA binding site to sequestor E2F activity away from cellular target genes, the so-called *decoy* approach, prevents cellular proliferation, and short peptides that block E2F function by inhibiting dimerization (with its essential partner protein DP-1) or its interaction with DNA cause apoptosis in tumor cells. Furthermore, a small molecule that is believed to act by modulating the E2F pathway has entered Phase I clinical studies.

Other regulatory factors involved in control of the pRb/E2F pathway, including cyclin-dependent kinases (Cdk) have also attracted considerable attention as a drug target. Though most members of the Cdk family control cell cycle transitions such as cyclin D/Cdk4 and cyclin E/Cdk2 (Fig. 3), other members have been assigned much more specific roles, like cyclinT1/Cdk9, which regulates targets including RNA

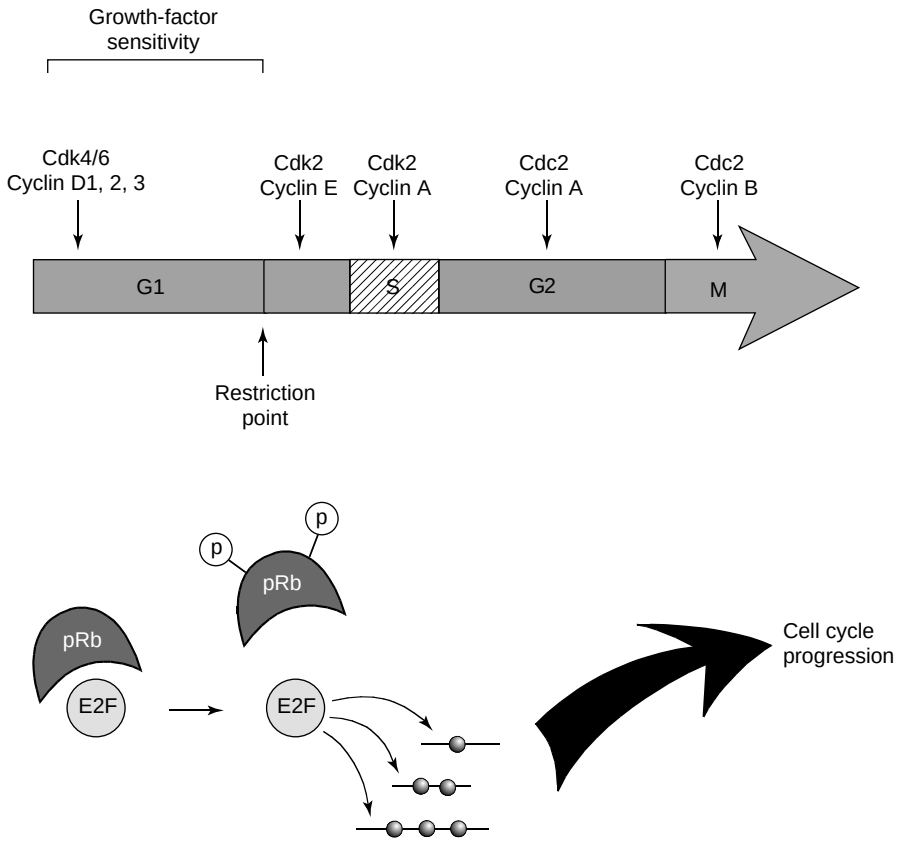

Fig. 3 Regulation of the cell cycle. The four phases of the cell cycle is indicated, together with the nature and the timing of the cyclin-Cdk complexes involved in cell cycle progression. The Restriction point, positioned in G1(R), is the point at which cells become committed to cell cycle progression and no longer require growth factors. The regulation of E2F activity, which controls the early cell cycle, is indicated underneath. In normal cells, the pRb binds to E2F to inactivate E2F activity. As pRb becomes phosphorylated (indicated by p) by Cdk complexes, E2F is released leading to the activation of target genes required for cell cycle progression. Normal control of pRb activity is frequently lost in human tumor cells, either through mutation in pRb or altered levels of its upstream regulators, including Cdk complexes.

polymerase II to influence the expression of a diverse set of genes.

Compounds that act on multiple Cdk enzymes, together with those that are more selective, have been taken into clinical development. UCN-01 is an example of a nonspecific kinase inhibitor active against a variety of kinases including Cdks, protein kinase C, and checkpoint 1 kinase. In contrast, flavopiridol and R-roscovitine exhibit greater specificity for Cdk targets. Flavopiridol inhibits Cdk1, 2, 4, and 7 (including Cdk2 at nM concentrations) by competing with the ATP binding site, and has reached Phase II clinical trials as a monotherapy for cancer. Roscovitine possesses moderate specificity for Cdk2, but also inhibits other kinases like Cdk7

and Cdk2 at lower potency. In a similar fashion to flavopiridol, roscovitine acts as a competitive inhibitor and induces apoptosis in tumor cells. Roscovitine (also known as CYC 202) has completed a Phase I clinical trial and appears to be well tolerated.

Advances are continuing to be made in improving the selectivity of Cdk inhibitors. For example, a number of oxindole-substituted Cdk inhibitors have greater than tenfold selectivity for Cdk2 against a panel of other Cdks. In this respect, it is hoped that the clinical development of selectively acting Cdk inhibitors will identify molecules with greater efficacy than some of the broad-spectrum agents currently in clinical development.

7 Targeting Chromatin Control in Cancer

Recent advances have highlighted a new and exciting area for therapeutic intervention, namely, the interplay between the cell cycle and chromatin control. Chromatin is the DNA-proteinaceous material in which chromosomal DNA resides in the nucleus. The majority of chromatin proteins is composed of histones that assemble into nucleosomes, and thereby assist in DNA packaging and transcriptional control. The histone tail is subject to a variety of enzymatic modifications, including phosphorylation, acetylation, and methylation, and many of the critical enzymes responsible for these modifications have recently come to light. Histone deacetylases (HDACs) are responsible for removing acetyl groups from lysine residues in histones and many proteins involved in the cell cycle (Fig. 4).

From the perspective of cancer therapy, HDACs have gained recognition as an important target, which, in part, reflects the identification of proteins other than histones that are subject to acetylation control. The activity of cell cycle regulators including E2F, p53, and pRb is influenced by acetylation, and pRb controls E2F activity through the recruitment of chromatin-modifying enzymes, such as HDACs (Fig. 4). Several oncogenic proteins have altered recruitment of HDACs, which leads to aberrant gene transcription. This is exemplified by the fusion protein PML-RARα, which occurs through a chromosomal translocation in acute promyelocytic leukemia (APL). The PML-RARα fusion protein recruits HDAC and represses transcription, causing a block to differentiation and promoting the oncogenic phenotype in APL. Perhaps, not surprisingly, compounds that inhibit HDAC activity cause potent cell cycle effects and frequently induce apoptosis; encouraging results of HDAC inhibitor clinical trials have begun to validate HDAC enzymes as drug targets.

SAHA (suberoylanilide hydroxamic acid) is a small molecule HDAC inhibitor that has reached Phase II clinical trials for the treatment of solid and hematological malignancy, and a range of other HDAC inhibitors are gaining clinical acceptance, for example, PXD101 is a highly potent HDAC inhibitor that blocks proliferation of tumor cells and has entered clinical trials. In addition, a variety of studies have demonstrated the potential for synergy using combinations of HDAC inhibitors with several mechanistically distinct antitumor agents, such as SAHA, which together with radiotherapy, produces an additive effect in human prostate cancer spheroids. While HDAC inhibitors have yet to be completely validated, nevertheless, they do represent

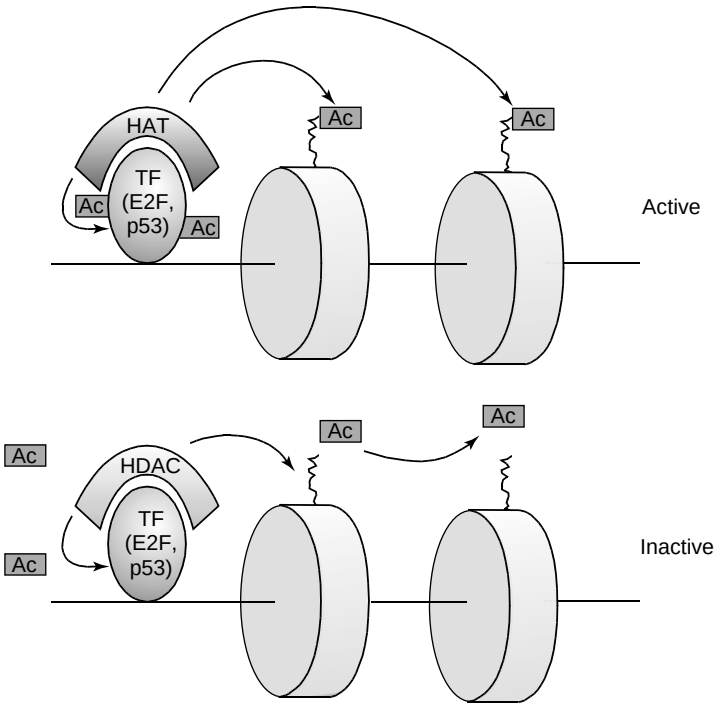

Fig. 4 Acetylation control in the cell cycle. Acetylation (Ac) mediated by acetyltransferases (HATS) can target histone tails in the form of nucleosomes (yellow) or transcription factors (TF) involved in cell cycle control, like E2F and p53. In most cases so far studied, acetylation appears to activate transcription. In contrast, deacetylation mediated by histone deacetylases (HDAC) causes transcriptional inactivity by targeting histones in nucleosomes, leading to a more transcriptionally inert state, together with dampening the activity of transcription factors. (See color plate p. xxv.)

a promising new approach to cancer treatment.

8 Regulating Hsp90

Heat shock protein 90 (Hsp90) is a molecular chaperone that regulates the activity of a variety of cellular proteins, enabling protein folding and preventing denaturation, and perhaps aggregation, of misfolded proteins. Hsp90 is gaining recognition as a cancer target in part because it is overexpressed in a variety of tumor cells and associates with oncoproteins including the tyrosine kinase v-Src, the serine/threonine kinase Raf1, HIF1 (discussed later), and mutated p53. A small molecule inhibitor of Hsp90, geldanamycin, blocks the interaction of Hsp90 with client proteins, including v-Src and Raf1, by interacting with the hydrophobic ATP-binding pocket in the N-terminal region and inhibiting the Hsp90 ATPase activity.

Geldanamycin has been used as a molecular probe to explore the regulation

and composition of Hsp90. Two distinct sets of Hsp90 chaperone are endowed with opposing functions, reflecting the nucleotide occupying the binding pocket. When ATP is present, Hsp90 assembles with the cochaperone proteins p23 and $p50^{Cdc37}$, resulting in the stabilization of client proteins. In contrast, in the presence of ADP, Hsp90 assembles with a different set of proteins, including Hsp70 and $p60^{Hop}$, to promote client protein ubiquitination and degradation via the proteasome (Fig. 5). Binding of geldanamycin to the nucleotide pocket of Hsp90 locks Hsp90 in the form that favors client protein degradation. The mechanism through which Hsp90 causes the degradation of client proteins remains unclear, although E3 ligases that associate with Hsp90 and ubiquitinate some client proteins have been identified. Thus, Hsp90 complexes have opposing functions, and geldanamycin promotes formation of the degrading Hsp90 chaperone complex.

An interesting property of Hsp90 inhibitors observed in preclinical studies is the induction of cytostasis or apoptosis, depending on the cell-type. This may reflect the nature of the Hsp90 chaperone complex and the variation in target client proteins present in cells. Some evidence suggests that the integrity of pRb may dictate sensitivity to Hsp90 inhibition, as cells lacking pRb appear to be quite sensitive to inhibitor-induced apoptosis.

In preclinical studies, Hsp90 inhibitors have shown promising activity. While geldanamycin exhibits levels of hepato-toxicity

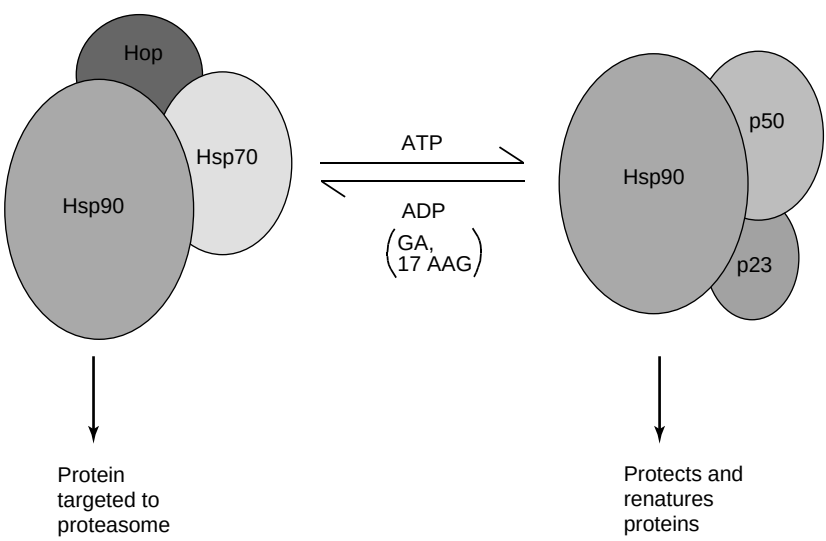

Fig. 5 Regulation of Hsp90 activity. Hsp90 can form different protein chaperone complexes with opposing activities. When ATP is bound to the nucleotide-binding pocket, assembly with p23 and $p50^{Cdc37}$ occurs resulting in a chaperone complex that stabilizes and protects client proteins, whereas ADP in the pocket causes Hsp90 to assemble with Hsp70 and $p60^{Hop}$, resulting in client protein degradation via the proteasome. With geldanamycin (GA) or its derivative 17-allylaminogeldanamycin (17AAG), the chaperone assembles with Hsp70 and $p60^{Hop}$ to favor client protein degradation.

that are too high for clinical application, the derivative 17-allylaminogeldanamycin (17AAG) exhibits improved efficacy, and has now reached phase I clinical trials.

In the clinical setting, if cytostasis were to be the primary outcome of treating a patient with an Hsp90 inhibitor, then sustained chronic administration of the drug may not be possible because of the associated toxicity profile. Conversely, if Hsp90 inhibition causes apoptosis upon short administration, then efficacy with limited toxicity may be achievable. As suggested by Neckers, this situation may be solved eventually by profiling the patient's tumor for the nature and expression level of Hsp90 client proteins. Combined with further information on the appropriate clinical schedule for Hsp90 inhibitors, it may be possible to identify and treat those tumors that will preferentially enter apoptosis upon inhibition of Hsp90 activity. This general principle could be important for other new agents, such as HDAC inhibitors.

9
Blocking Angiogenesis Through Inhibiting HIF1 Activity

Hypoxia-inducible factor 1 (HIF1) plays an important role in regulating angiogenesis by controlling the expression of genes linking vascular oxygen supply to metabolic demand. Tumor progression is associated with higher levels of vascularization that results from the increased synthesis of proangiogenesis factors and decreased synthesis of antiangiogenic factors, and angiogenesis progresses hand-in-hand with the adaptation of tumor cells to growth in low oxygen levels, by increasing glucose transport and glycolysis. HIF1 controls the activity of a variety of genes involved in these processes; for example, the gene for vascular endothelial growth factor (VEGF), which is required for tumor angiogenesis, is regulated by HIF1. Similarly, target genes include glucose transporters GLUT1 and 3, together with enzymes involved in glycolysis like aldolase A and C, enolase 1, and hexokinase 1 and 3.

In cells, HIF1 exists as a heterodimer composed of HIF1α and HIF1β; HIF1β is constitutively expressed, whereas HIF1α is maintained at low levels in most cells under normoxic conditions. While HIF1α is regulated through a multistep process, a key event involves the hydroxylation of specific residues in an oxygen-dependent manner (Fig. 6). The modification of two prolyl residues (P402 and P564) mediates the binding of the von Hippel-Lindau (VHL) protein, which is the targeting component of an E3 ubiquitin ligase that confers degradation on HIF1 through the proteasomal pathway. In hypoxic conditions, prolyl hydroxylation becomes rate limiting, allowing HIF1α to escape degradation to favor increased levels of HIF1α. Further, Lando et al. identified a second hydroxylation-dependent control process occurring in the C-terminal transcriptional activation domain, where an asparagine residue (N803) is hydroxylated by an asparaginyl hydroxylase, which blocks the interaction with the p300 co-activator, again downregulating HIF activity (Fig. 6). Collectively, HIF1α activity is regulated by a series of oxygen-dependent posttranslational modifications mediated by three prolyl hydroxylases and one or more asparaginyl hydroxylases.

Increased activity of the HIF pathway is seen in many tumor types, resulting from the intratumoral hypoxia together with genetic alterations that affect HIF1 control. The impact of genetic alterations on the

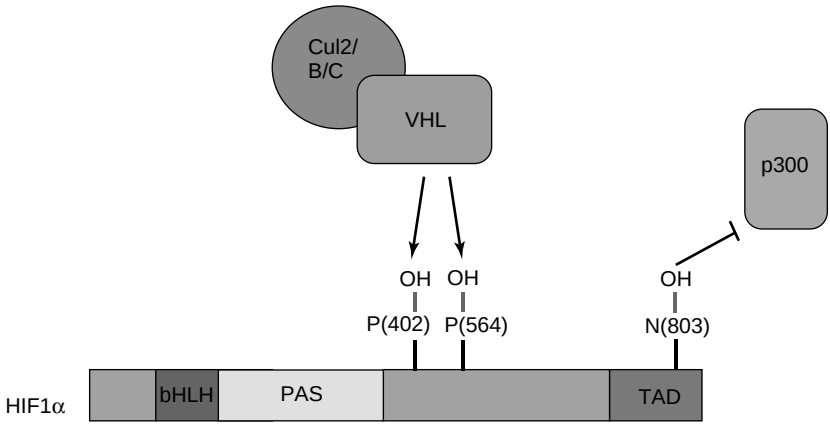

Fig. 6 Regulation of the HIFIα protein. The domains in HIFIα including the basic helix–loop–helix (bHLH) and PAS domains, required for dimerization and DNA binding and the C-terminal trans-activation domain are indicated. Hydroxylation of proline (P) residues at 402 and 564 is required for binding of the von Hippel-Lindau (VHL) tumor suppressor protein, which recruits an E3 ligase (Cul2/B/C complex) that degrades HIFIα. Hydroxylation of the asparagines (N) residue at 803 in the TAD prevents the interaction with p300 and downregulates HIF activity.

activity of HIF1 is influenced by the signal transduction pathway active in the tumor, and may involve pathways regulated by insulin, epidermal growth factor, Ras, or Src. In this respect, VHL is an important component of HIF1 regulation, and inactivation of VHL tumor suppressor activity in VHL hereditary cancer syndrome is characterized by a highly restricted cancer predisposition involving renal cell carcinoma, hemanglioblastoma, and pheochromocytoma. In many other tumors, HIF1α is commonly overexpressed, where its levels sometimes correlate with increased vascular density such as in brain tumors. In other cases, HIF1α overexpression is a marker for aggressive disease, such as cervical cancer. However, it should be borne in mind that HIF1α is not a universal marker for clinical outcome. In ovarian cancer, the overexpression of HIF1α with p53 is correlated with apoptosis, contrasting with the coexpression of HIF1α with mutant p53.

HIF1α is gaining increased acceptance as a potential therapeutic target. The antitumor effect of some chemotherapeutic agents may be derived from effects on the HIF1 pathway. For example, rapamycin inhibits hypoxia-induced HIF1 expression, most likely by blocking the serine/threonine FRAP kinase required for translation of HIF1 mRNA. Agents that target other signaling molecules in the PI3K-AKT-FRAP pathway would similarly be expected to alter HIF1 activity together with many other effects independent of HIF1.

To date, mechanism-based therapeutics that directly target HIF1 are not available. However, a recent Phase III trial supports the therapeutic value of agents that modulate angiogenesis in cancer therapy. The therapeutic antibody, Avastin™, designed to inhibit VEGF activity, which plays a critical role in tumor angiogenesis, provided a 50% increase in patient survival chance compared to patients

receiving chemotherapy alone. While cancer therapies that inhibit angiogenesis remain in their infancy, clinical studies like the Avastin trial emphasize future potential.

10 Regulating Protein Turnover

Increasing evidence suggests that drugs that influence protein degradation may prove to be beneficial in cancer treatment. As we have discussed, protein degradation plays a crucial role in regulating cell viability, not only in the control of damaged proteins but also in controlling key regulatory processes. The turnover of many cell cycle regulatory proteins is required for normal cell cycle control, for example, the Cdk inhibitor p27 whose degradation is enhanced in certain tumor cells. Similarly, signaling to the NFκB transcription factor is dependent upon protein turnover of its specific inhibitor IκB, allowing NF-κB to activate a variety of genes including cytokines (like tumor necrosis factor) and antiapoptotic proteins.

The proteasome plays a central role in regulating protein turnover. The 26S proteasome is a large multiprotein complex composed of a 20S catalytic complex together with two 19S regulatory subunits that degrade proteins targeted by ubiquitination through the addition of ubiquitin to lysine residues. Ubiquitinated proteins are broken down to short peptides by the central subunit rings of the proteasome, which contain multiple active sites recognizing threonine as the catalytic residue (Fig. 7).

It has become increasingly clear that compounds that block proteasome function possess potential as anticancer agents. The first compounds identified exhibited inappropriate pharmacokinetic properties but ultimately led to the generation of compounds with improved properties, including PS-341 (known as Velcade and developed by Millenium). Velcade has activity across a wide range of tumor cells but appears to act preferentially on tumor cells compared to normal cells. Myeloma cells undergo apoptosis when treated with Velcade, in contrast, for example, to normal bone marrow stromal cells, which exhibit a much reduced sensitivity profile.

Other promising properties of proteasome inhibitors relate to the potential utility in resistant disease. Velcade will arrest the growth of myeloma cells that are resistant to conventional chemotherapies, and overexpression of Bcl2 does not affect the sensitivity of myeloma to Velcade. Velcade was the first proteasome inhibitor to enter clinical trials in hematological malignancy and solid tumors. While the drug is rapidly cleared from plasma (over 90% clearance within 15 min of administration), early indications of efficacy were observed in Phase I trials. Efficacy was subsequently confirmed in a Phase II trial on patients with refractory multiple myeloma. Velcade has successfully completed Phase III clinical trials in multiple myeloma, and is the first proteasome inhibitor to be approved (by the US FDA), for multiple myeloma.

11 Conclusions and Perspectives

We are beginning to see subtle hints that exploiting mechanism-based approaches will deliver improved cancer medicines over many existing chemotherapeutics.

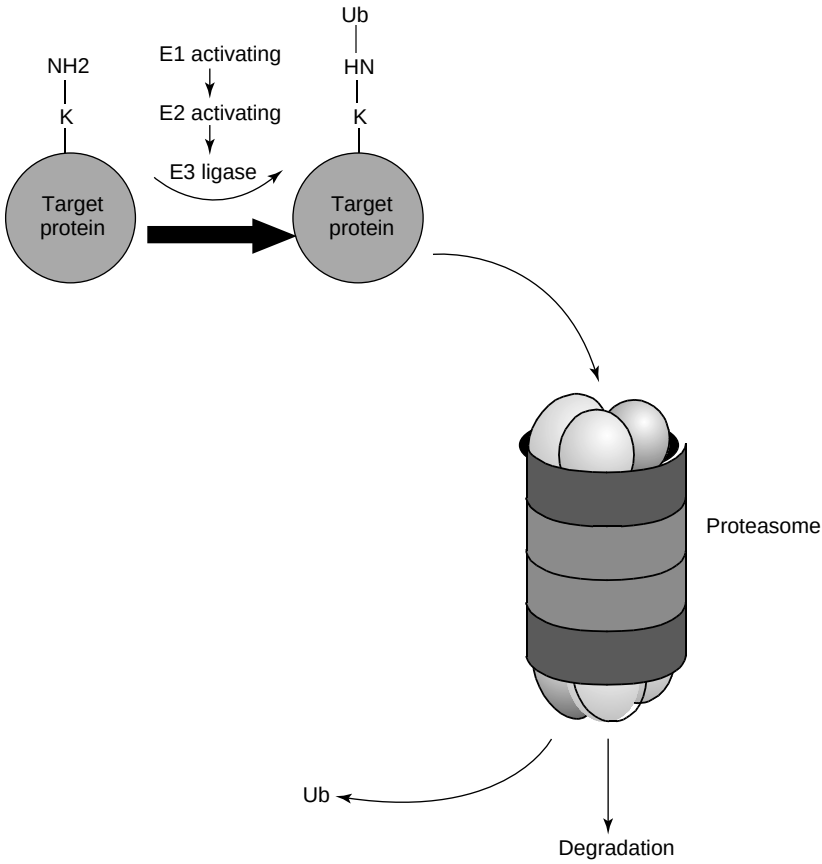

Fig. 7 The proteasome in protein degradation. Target proteins for degradation are ubiquitinated (Ub) at lysine (K) residues through a process involving a ubiquitin-activating enzyme (E1), the transfer of ubiquitin to a conjugating enzyme (E2), which is then transferred to a target protein by an E3 protein ligase. Polyubiquitinated proteins are then recognized and degraded by the 26S proteasome.

Both Gleevec and Valcade directed against the bcr-*abl* tyrosine kinase and the proteasome respectively have shown clinical efficacy. While we await further clinical results with Iressa, HDAC, Cdk, and Hsp90 inhibitor programs, we are left with the compelling view that improved therapies will also arise from these routes of enquiry.

Perhaps, it will be of greater significance to understand the clinical practice with these new cancer agents. Should these compounds be progressed as single agents or, which seems more likely, be applied as combination therapies perhaps with existing cytotoxic agents? It remains a significant hurdle to address this question, and to further provide scientific rationale in considering the array of possible combination options that could be studied. An interrelated question concerns the scheduling regime for administering the drug, particularly in combination therapies, which will reflect knowledge of the critical targets

through which the inhibitor regulates cancer cell growth and the expression of these targets in the particular cancer type. In this respect, we have already discussed the different properties of Hsp90 inhibitors in apoptosis and cell cycle arrest, and how these properties could impact on the clinical regimen and perhaps require patient stratification for maximum clinical benefit.

The possibility of resistance to mechanism-based agents remains to be determined. Tumor cells are adept at acquiring drug resistance phenotypes through diverse mechanisms, and we have already seen clinical resistance to Gleevec, but we await further information before this becomes clinically exemplified for other new agents. Of course, if drug resistance does appear as a clinical hurdle, then it may be possible to modulate its progression by altering drug concentration and exposure through a combination therapy approach.

It is clear that many questions remain to be answered. Nevertheless, we are witnessing an exciting era in cancer drug discovery in which newer drugs are beginning to replace conventional cytotoxic approaches. Our views and anticipations, which are shared by many others, reflect the gathering momentum that translating research knowledge about the cancer cell will yield improved and more efficacious drugs for treating the cancer patient.

Acknowledgment

We thank Marie Caldwell for help in preparing the manuscript. Work in our laboratory was supported by the MRC, CRUK, LRF, AICR and EC.

See also Bioorganic Chemistry; Drug Bioavailability, Distribution and Clearance Prediction; Medicinal Chemistry.

Bibliography

Books and Reviews

Adams, J., Palombella, V.J., Elliott, P.J. (2000) Proteasome inhibition: a new strategy in cancer treatment, *Invest. New Drugs* **18**, 109–121.

Brooks, G., La Thangue, N.B. (1999) The cell cycle and drug discovery: the promise and the hope, *Drug Discovery Today*, **4**, 455.

Demonacos, C., La Thangue, N.B. (2002) p53 and drug discovery, *PCCR* **5**, 375–382.

Faderl, S., Talpaz, M., Estrov, Z., Kantarjian, H.M. (1999) Chronic myelogenous leukaemia: biology and therapy, *Ann. Intern. Med.* **131**, 207–219.

Hanahan, D., Weinberg, R. (2000) The hallmarks of cancer, *Cell* **100**, 57–70.

McLaughlin, F., Finn, P., La Thangue, N.B. (2003) The cell cycle, chromatin and cancer: mechanism-based therapeutics come of age, *Drug Discovery Today* **8**, 793–802.

Neckers, L. (2002) Hsp90 inhibitors as novel cancer chemotherapeutic agents, *Trends Mol. Med.* **8**(4 Suppl.), S55–S61.

Semenza, G.L. (2001) Hypoxia-inducible factor 1: oxygen homeostasis and disease pathophysiology, *Trends Mol. Med.* **7**, 345–350.

Simon, M.A. (2000) Receptor tyrosine kinases: specific outcomes from general signals, *Cell* **103**, 13–15.

Workman, P., Kaye, S. (2002) Translating basic cancer research into new cancer therapeutics, *Trends Mol. Med.* **8**(4 Suppl.), S1–S9.

Primary Literature

Agnew, E.B., Wilson, R.H., Grem, J.L. et al. (2001) Measurement of the novel anti-tumour agent 17-(allylamino)-17-demethoxygeldanamycin in human plasma by high-performance liquid chromatography, *J. Chromatogr., B.: Biomed. Sci. Appl.* **755**, 237–243.

Akiyama, T., Yoshida, T., Tsujita, T., et al. (1997) G1 phase accumulation induced by UCN-01 is associated with dephosphorylation of Rb and CDK2 proteins as well as induction of CDK inhibitor p21/Cip1/WAF1/Sdl1 in p53-mutated human epidermoid carcinoma A431 cells, *Cancer Res.* **576**, 1495–1501.

An, W.G., Schulte, T.W., Neckers, L.M. (2000) The heat-shock protein 90 antagonist geldanamycin alters chaperone association with p210bcr-abl and v-src proteins before their degradation by the proteasome, *Cell Growth Differ.* **11**, 355–360.

Anderson, P.G.A. (2003) A Phase 2 study of bortezomib in relapsed, refractory myeloma, *N. Engl. J. Med.* **348**, 2609–2617.

Bandara, L.R., Girling, R., LaThangue, N.B. (1997) Apoptosis induced in mammalian cells by small peptides that functionally antagonise the Rb-regulated E2F transcription factor, *Nat. Biotech.* **15**, 896–901.

Baselga, J. (2000) Continuous administration of ZD1839 (Iressa), a novel oral epidermal growth factor receptor tyrosine kinase inhibitor, in patients with five selected-tumour types: evidence of activity and good tolerability, *Proc. Am. Soc. Oncol.* **19**, 177a.

Baselga, J. (2001) The EGFR as a target for anticancer therapy-focus on cetuximab, *Eur. J. Cancer* **37**, S16–S22.

Birner, P., Schindl, M., Obermair, A., et al. (2000) Over-expression of hypoxia-inducible factor 1α is a marker for an unfavourable prognosis in early-stage invasive cervical cancer, *Cancer Res.* **60**, 4693–4696.

Birner, P., Schindl, M., Obermair, A., Breitenecker, G., Oberhuber, G. (2001a) Expression of hypoxia-inducible factor 1α in epithelial ovarian tumours: its impact on prognosis and on response to chemotherapy, *Clin. Cancer Res.* **7**, 1661–1668.

Birner, P., Gatterbauer, B., Oberhuber, G., et al. (2001b) Expression of hypoxia-inducible factor 1α in oligodendrogliomas: its impact on prognosis and on neoangiogenesis, *Cancer* **92**, 165–171.

Chan, H.-M., La Thangue, N.B. (2001) p300/CBP proteins: HATS for transcriptional bridges and scaffolds, *J. Cell Sci.* **114**, 2363–2373.

Chiarle, R., Budel, L.M., Skolnik, J., et al. (2000) Increased proteasome degradation of cyclin-dependent kinase inhibitor p27 is associated with a decreased overall survival in mantle cell lymphoma, *Blood* **95**, 619–626.

Ciardiello, F., Tortora, G. (2001) A novel approach in the treatment of cancer: targeting the epidermal growth factor receptor, *Clin. Cancer Res.* **7**, 2958–2970.

Connell, P., Ballinger, C.A., Jiang, J., et al. (2001) The co-chaperone CHIP regulates protein triage decisions mediated by heat-shock proteins, *Nat. Cell Biol.* **3**, 93–96.

Deininger, M.W., Goldman, J.M., Lydon, N., Melo, J.V. (1997) The tyrosine kinase inhibitor CGP57148B selectively inhibits the growth of BCR-ABL-positive cells, *Blood* **90**, 3691–3698.

Deininger, M.W., Goldman, J.M., Melo, J.V. (2000) The molecular biology of chronic myeloid leukaemia, *Blood* **96**, 3343–3356.

Druker, B.J., Lydon, N.B. (2000) Lessons learned from the development of an abl tyrosine inhibitor for chronic myelogenous leukaemia, *J. Clin. Invest.* **105**, 3–7.

Dyson, N. (1998) The regulation of E2F by pRB-family proteins, *Genes Dev.* **12**, 2245–2262.

Feinman, R., Gangurde, P., Miller, S., et al. (2001) Proteasome inhibitor PS-341 inhibits constitutive NF-KB activation and bypass the anti-apoptotic Bcl2 signal in human multiple myeloman cells, *Blood* **98**, 640a.

Foster, B.A. (1999) Pharmacological rescue of mutant p53 conformation and function, *Science* **286**, 2507–2510.

Friedler, A., Hansson, L.O., Veprintsev, D.B., et al. (2002) A peptide that binds and stabilizes p53 core domain: chaperone strategy for rescue of oncogenic mutants, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 937–942.

Gorre, M.E., Mohammed, M., Ellwood, K., et al. (2001) Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification, *Science* **293**, 876–880.

Hainaut, P. (2002) Tumour-specific mutations in p53: the acid test, *Nat. Med.* **8**, 21–23.

Hideshima, T., Richardson, P., Chauhan, D., et al. (2001) The proteasome inhibitor PS-341 inhibits growth, induces apoptosis, and overcomes drug resistance in human multiple myeloma cells, *Cancer Res.* **61**, 3071–3076.

Hochhaus, A., Kreil, S., Corbin, A., et al. (2001) Roots of clinical resistance to ATI-571 cancer therapy, *Science* **293**, 2136.

Hockel, M., Vaupel, P. (2001) Tumour hypoxia: definitions and current clinical, biologic, and molecular aspects, *J. Natl. Cancer Inst.* **93**, 266–276.

Hollstein, M., Hergennahn, M., Yang, Q. et al. (1999) New approaches to understanding p53

gene tumour mutation spectra, *Mutat. Res.* **431**, 199–209.

Honma, T., Yoshizumi, T., Hashimoto, N., et al. (2001) A novel approach for the development of selective cdk4 inhibitors: library design based on location of cdk4 specific amino acid residues, *J. Med. Chem.* **44**, 4628–4640.

Hostein, I., Robertson, D., DiStefano, F., Workman, P., Clarke, P.A. (2001) Inhibition of signal transduction by the Hsp90 inhibitor 17-allylamino-17-demethoxygeldanamycin results in cytostasis and apoptosis, *Cancer Res.* **61**, 4003–4009.

Ivan, M., Kondo, K., Yang, H., et al. (2001) HIFα targeted for VHL-mediated destruction by proline hydroxylation: implications for O_2 sensing, *Science* **292**, 464–468.

Jaakkola, P., Mole, D.R., Tian, Y.M., et al. (2001) Targeting of HIF-α to the von Hippel-Lindau ubiquitylation complex by O_2-regulated prolyl hydroxylation, *Science* **292**, 468–472.

Kantarjian, H. (2000) Phase II study of ST1571, a tyrosine kinase inhibitor, in patients with resistant or refractory Philadelphia chromosome positive chronic myeloid leukaemia, *Blood* **96**, 470a.

Ko, L., Prives, C. (1996) p53: puzzle and paradigm, *Genes Dev.* **10**, 1054–1072.

Komarov, P.G., Komarova, E.A., Kondratov, R.V., et al. (1999) A chemical inhibitor of p53 that protects mice from the side effects of cancer therapy, *Science* **285**, 1733–1737.

Kouzarides, T. (1999) Histone acetylases in cell proliferation, *Curr. Opin. Genet. Dev.* **9**, 40–48.

Lando, D., Peet, D.J., Whelan, D.A., Gorman, J.J., Whitelaw, M.L. (2002) Asparagine hydroxylation of the HIF transactivation domain: a hypoxic switch, *Science* **295**, 858–861.

Laughner, E., Taghavi, P., Chiles, K., Mahon, P.C., Semenza, G.L. (2001) HER2 (neu) signalling increases the rate of hypoxia-inducible factor 1α (HIF1α) synthesis: novel mechanism for HIF-1 mediated vascular endothelial growth factor expression, *Mol. Cell. Biol.* **21**, 3995–4004.

Laurence, V., Faivre, S., Vera, K., et al. (2002) Preliminary results of an ongoing phase 1 and pharmacokinetic study of CYC202, a novel oral cyclin-dependent kinases inhibitor, in patients with advanced malignancies, *Eur. J. Cancer* **38**, S49.

Li, Y., Sun, X., Lamout, J.T., Pardee, A.B., Li, C.J. (2003) Selective killing of cancer cells by beta-lapachone; direct checkpoint activation as a strategy against cancer, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2674–2678.

Lin, R.J., Nagy, L., Inoue, S., Shao, W., Miller, W.H. Jr., Evans, R.M. (1998) Role of the histone deacetylase complex in acute promyelocytic leukaemia, *Nature* **391**, 811–814.

Manegold, C. (2003) Gefitinib (Iressa, ZD 1839) for non-small cell lung cancer (NSCLC): recent results and further strategies, *Adv. Exp. Med. Biol.* **532**, 247–252.

Mann, M.J., Whittemore, A.D., Donaldson, M.D., et al. (2001) Ex-vivo gene therapy of human vascular bypass grafts with E2F decoy: the PREVENT single centre, randomised, controlled trial, *Lancet* **354**, 1493.

Marks, P.A., Rifkind, R.A., Richon, V.M., Breslow, R. (2001) Inhibitors of histone deacetylases are potentially effective anticancer agents, *Clin. Cancer Res.* **7**, 759–760.

McCarthy, M. (2003) Anti-angiogenesis drug promising for metastatic colorectal cancer, *Lancet* **361**, 1959.

McClue, S.J., Blake, D., Clarke, R., et al. (2002) *In vitro* and *in vivo* anti-tumour properties of the cyclin-dependent kinase inhibitor CYC202 (R-Roscovitine), *Int. J. Cancer* **102**, 463–468.

Montigiani, S., Müller, R., Kontermann, R.E. (2003) Inhibition of cell proliferation and induction of apoptosis by novel tetravalent peptides inhibiting DNA binding of E2F, *Oncogene* **22**, 4943–4952.

Moulder, S.L., Yakes, F.M., Muthuswamy, S.K., Bianco, R., Simpson, J.F., Arteaga, C.L. (2001) Epidermal growth factor receptor (HER1) tyrosine kinase inhibitor ZD1839 (Iressa) inhibits HER2/*neu* (erbB2)-over-expressing breast cancer cells *in vitro* and *in vivo*, *Cancer Res.* **61**, 8887–8895.

Nguyen, V.T., Kiss, T., Michels, A.A., Bensaude, O. et al. (2001) The 7SK small nuclear RNA inhibits the CDK9/cyclinT1 kinase to control transcription, *Nature* **414**, 322–325.

Nimmanapalli, R., O'Bryan, E., Bhalla, K. (2001) Geldanamycin and its analogue 17-allylamino-17-demethoxygeldanamycin lowers Bcr-Abl levels and induces apoptosis and differentiation of Bcr-Abl-positive human leukaemic blasts, *Cancer Res.* **61**, 1799–1804.

Nix, D., Pien, C., Newman, R., et al. (2001) Clinical development of a proteasome inhibitor, PS-341, for the treatment of cancer, *Proc. Am. Soc. Clin. Oncol.* **20**, 86a.

Papandreou, C., Daliani, D., Millikan, R.E., et al. (2001) Phase I study of intravenous (I.V.) proteasome inhibitor PS-341 in patients (Pts) with advanced malignancies, *Proc. Am. Soc. Clin. Oncol.* **20**, 86a.

Plumb, J.A., Finn, P.W., Williams, R.J., et al. (2003) Inhibition of human tumour cell growth by the novel histone deacetylase inhibitor PXD101, *Mol. Cancer Ther.* **2**, 721–728.

Scheibel, T., Buchner, J. (1998) The Hsp90 complex – a super-chaperone machine as a novel drug target, *Biochem. Pharmacol.* **56**, 675–682.

Seidman, A.D., Fornier, M.N., Esteva, F.J., et al. (2001) Weekly trastuzumab and paclitaxel therapy for metastatic breast cancer with analysis of efficacy by HER2 immunophenotype and gene amplification, *J. Clin. Oncol.* **19**, 2587–2595.

Selivanova, G., Ryabchenko, L., Jansson, E., Iotsova, V., Wiman, K.G. (1998) Reactivation of mutant p53: a new strategy for cancer therapy, *Semin. Cancer Biol.* **8**, 369–378.

Sgouros, G., Yang, W.-H., Richon, V.M., Kelly, W.K., Rifkind, R.A., Marks, P.A. (2002) Synergistic interaction of suberoylanillde hydroxamic acid (SAHA) and radiation in human prostate tumour spheroids, *Proc. Am. Soc. Clin. Oncol.* **2**, 105.

Sherr, C.J., Roberts, J.M. (1999) CDK inhibitors: positive and negative regulators of cyclin-dependent regulators of G1-phase progression, *Genes Dev.* **13**, 1501–1512.

Slamon, D.J., Leyland-Jones, B., Shak, S., et al. (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that over-expresses HER2, *N. Engl. J. Med.* **344**, 783–792.

Stancato, L.F., Silverstein, A.M., Owens-Grillo, J.K., Chow, Y.H., Jove, R., Pratt, W.B. (1997). The hsp90-binding antibiotic geldanamycin decreases Raf levels and epidermal growth factor signaling without disrupting formation of signaling complexes or reducing the specific enzymatic activity of Raf kinase, *J. Biol. Chem.* **272**, 4013–4020.

Stinchcombe, T.E., Mitchell, B.S., Depcik-Smith, N., et al. (2000) PS-341 is active in multiple myeloma: preliminary report of a phase I trial of the proteasome inhibitor PS-341 in patients with hematologic malignancies, *Blood* **96**, 516a.

Strahl, B.D., Allis, C.D. (2000) The language of covalent histone modifications, *Nature* **403**, 41–45.

Srethapakdi, M., Liu, F., Tavorath, R., Rosen, N. (2000) Inhibition of Hsp90 function by ansamycins causes retinoblastoma gene product-dependent G1 arrest, *Cancer Res.* **60**, 3940–3946.

Stebbins, C.E., Russo, A.A., Schneider, C., Rosen, N., Hartl, F.U., Pavletich, N.P. (1997) Crystal structure of an Hsp90-geldanamycin complex: targeting of a protein chaperone by an antitumour agent, *Cell* **89**, 239–250.

Thomas, J.P., Tutsch, K.D., Cleary, J.F., et al. (2002) Phase I clinical and pharmacokinetic trial of the cyclin-dependent kinase inhibitor flavopiridol, *Cancer Chemother. Pharmacol.* **50**, 465–472.

Van Oosterom, A.T., Judson, I., Verweij, J., et al. (2001) STI571, an active drug in metastatic gastrointestinal stromal tumours (GIST), an EORTC Phase 1 study, *Proc. Am. Soc. Clin. Oncol.* **20**, 1a.

Xu, Y., Lindquist, S. (1997) Heat-shock protein hsp90 governs the activity of pp60v-src kinase, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7074–7078.

Wang, C.Y., Mayo, M.W., Korneluk, R.G., Goeddel, D.V., Baldwin, A.S. Jr. (1998) NF-KB anti apoptosis: induction of TRAF1 and TRAF2 and c-IAP1 and c-IAP2 to suppress caspase 8 activation, *Science* **281**, 1680–1683.

Wu, L., Timmers, C., Maiti, B., et al. (2001) The E2F1-3 transcription factors are essential for cellular proliferation, *Nature* **414**, 457–462.

Zagzag, D., Zhong, H., Scalzitti, J.M., Laughner, E., Simons, J.W., Semenza, G.L. (2000) Expression of hypoxia inducible factor 1α in brain tumours: association with angiogenesis, invasion, and progression, *Cancer* **88**, 2606–2618.

Zhong, H., Chiles, K., Feldser, D., et al. (1999) Over-expression of hypoxia-inducible factor 1α in common human cancers and their metastases, *Cancer Res.* **59**, 5830–5835.

Zhong, H., De Marzo, A.M., Laughner, E., et al. (2000) Modulation of HIF1α expression by the epidermal growth factor phosphatidylinositol 3 kinase/PTEN/AKT/FRAP pathway in human prostate cancer cells: implications for tumour angiogenesis and therapeutics, *Cancer Res.* **60**, 1541–1545.

Cancer of the Prostate: Molecular Genetics

Rajvir Dahiya
University of California San Francisco and Veterans Affairs Medical Center
San Francisco, CA, USA

1 **Hereditary Factors 192**

2 **Genetic Alterations in Precursor Lesions 194**

3 **Genetic Alterations in Prostate Cancer 195**

4 **Tumor Suppressor Genes 196**

5 **Oncogenes 198**

6 **DNA Repair Genes 200**

7 **Telomerase 201**

8 **Growth Factors 202**

9 **Metastasis Suppressor Genes 203**

10 **DNA Methylation 206**

11 **Androgen Receptor 208**

12 **Conclusions 209**

Bibliography 210
Primary Literature 210

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Adenocarcinoma of the prostate is one of the most common cancers in men worldwide and is the second leading cause of cancer-specific death in the United States. Over the past 15 years, significant advances have been made in the treatment of prostate cancer, largely owing to earlier diagnosis and the evolution of surgical techniques and improvements in radiation therapy. Nevertheless, tremendous uncertainty and controversy exist regarding the appropriate clinical management of prostate cancer in many cases. Similarly, despite the magnitude of the problem, understanding of the pathogenesis, biology, and natural history of human prostate cancer remains incomplete and unclear. Progress has been made in the past decade in elucidating the molecular and genetic changes involved in prostate cancer development and progression. A better understanding of the underlying disease mechanisms will allow rational treatment strategies and improved outcomes. The classic model of tumor development describes a multistep process in which a series of genetic alterations leads to aberrant, uncontrolled cell growth. This is best characterized in the case of colon cancer, for which specific genetic changes have been correlated with clinical and pathologic disease progression. These genetic alterations can be both hereditary (germ-line mutation) and acquired (somatic mutation). The paradigm is less established in prostate cancer (Fig. 1). We discuss the current understanding of the molecular genetics and biology of prostate cancer development, including genetic predispositions, early changes associated with prostatic intraepithelial neoplasia (PIN), and the traditional classes of genes involved in promoting and repressing cancer development (protooncogenes/oncogenes and tumor suppressor genes, respectively). In addition, we describe recent studies on the role of genetic instability, telomerases, growth factors, cell-adhesion molecules, DNA ethylation, and the androgen receptor (AR) in prostate cancer.

1 Hereditary Factors

Multiple etiologies have been proposed to contribute to the development of prostate cancer. Environmental factors are likely play a major role in most cases; however, inherited genetic factors are clearly important in some men. Epidemiologic analysis of almost 3000 men in the Utah Cancer Registry demonstrated familial clustering of prostate cancer patients, even greater than that for both breast and colon cancer, for which a hereditary component has already been demonstrated. The risks of developing the disease are dependent on both the number of affected first-degree relatives, and early age of disease onset. Using segregation analysis in 691 families, a model of a rare autosomal dominant susceptibility gene with an allele frequency of 0.006 and penetrance of 89% at 85 years of age was proposed. Overall, this gene may account for approximately 9% of all prostate cancer cases but an increased proportion in those patients diagnosed at earlier ages. The criteria for hereditary prostate cancer include a cluster of at least three first-degree relatives with disease, at least two relatives with

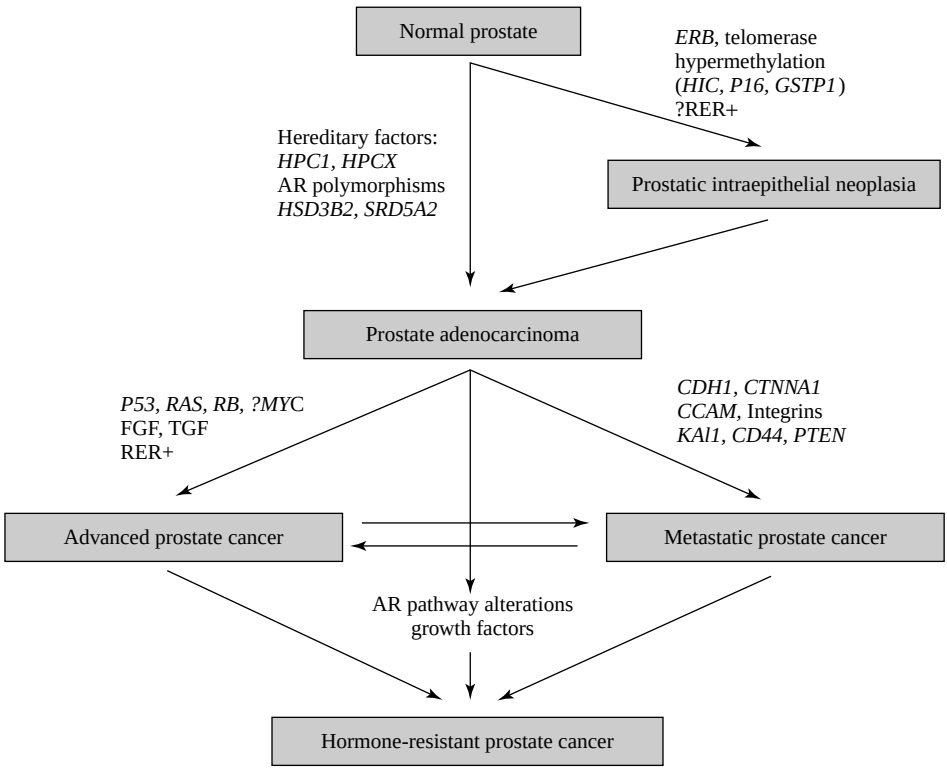

Fig. 1 Schematic illustrating the potential factors in the development of prostatic intraepithelial neoplasia and prostate cancer and the subsequent progression of prostate cancer. AR = androgen receptor; FGF = fibroblast growth factor; TGF = transforming growth factor; RER = replication error.

prostate cancer diagnosed under the age of 55 years, or prostate cancer in each of three generations in the paternal or maternal lineage.

Genome-wide analysis in 66 high-risk prostate cancer families suggested evidence of linkage to a locus on the long arm of chromosome 1 (1q24-25). This hereditary prostate cancer locus (HPC1) demonstrated linkage in an additional 25 North American and Swedish families studied. The gene has not yet been cloned, but candidate genes in the region include *SKI, ABL2, TRK,* and *LAMC2*. Other reports, however, have shown no evidence of linkage and no significant loss of heterozygosity (LOH) at the *HPC1* locus. Narod and colleagues suggested a recessive or X-linked gene in hereditary prostate cancer. A second prostate cancer susceptibility gene (*HPCX*) has recently been identified, accounting for approximately 16% of hereditary prostate cancer cases. A total of 360 prostate cancer families in North America, Finland, and Sweden were analyzed, and linkage to chromosome Xq27-28 was observed. This finding was confirmed in a study of another 153 families as well as in a study of 104 German prostate cancer families.

RNASEL gene (2′,5′-oligoisoadenylate-synthetase dependent), identified by a positional cloning/candidate gene method, maps to the HPC-predisposition locus at 1q24-q25 (HPC1) and was recently shown to harbor truncating mutations in two families with linkage to HPC1. A nonsense mutation and a mutation in an initiation codon of RNASEL segregate independently in two HPC1-linked families. Microdissected tumors with a germ-line mutation showed loss of heterozygosity and loss of RNase L protein, and that RNASEL activity was reduced in lymphoblasts from heterozyogous individuals compared with family members who were homozygous with respect to the wild-type allele. Thus, germ-line mutations in RNASEL may be of diagnostic value, and the 2-5A pathway might provide opportunities for developing therapies for those with prostate cancer. Thus, there is strong evidence of genetic factors in the potential for development of prostate cancer. The *HPC1* gene may play a role in up to a third of hereditary prostate cancer cases. Further work should identify the *HPC1* and *HPCX* gene products and their functions, as well as other potential prostate cancer susceptibility genes.

2 Genetic Alterations in Precursor Lesions

Histopathologic features of PIN have provided evidence that it represents a precursor lesion of prostate cancer. Changes noted in both high-grade PIN and early invasive prostate cancer include disruption of the basal cell layer, alterations in markers of secretory differentiation, nuclear and nucleolar enlargement, and increased cell proliferation. Identification of similar molecular changes in PIN and prostate cancer further strengthens this hypothesis. Sequential magnification of the abnormalities in PIN is seen in the progression to localized cancer, metastatic cancer, and hormone-refractory cancer. Allelic loss is common in both PIN and prostate cancer (Table 1). Qian and colleagues used centromere-specific fluorescence *in situ* hybridization (FISH) probes against chromosomes 7, 8, 10, and Y, to demonstrate chromosomal anomalies in 50% of PIN and carcinoma foci. Foci of metastatic cancer demonstrated increased chromosomal anomalies. Other alterations of chromosome 8 are common in both PIN and prostate cancer, with gain of chromosome 8q being the most frequent. Using polymerase chain reaction (PCR) and a novel microdissection technique, allelic imbalance of chromosome 8p12-21 in 64% of PIN foci and 91% of cancer foci has been described; no LOH at chromosome 8p12-21 was found in benign tissue. Additional LOH at chromosomes 8p, 10q, and 16q has been identified in 29% of PIN and 42% of primary tumors. Taken together, PIN is a precursor of prostate cancer and exhibits abnormalities in biomarkers,

Tab. 1 Genetic alterations in prostate cancer.

	Gains	*Losses*
Prostalic intraepithelial neoplasia	7, 8q, 10	8p, 10q, 16q, Y
Prostate adenocarcinoma	7, 8q, 10, X	2q, 5q, 6q, 7q, 8p, 9p, 10q, 13q, 15q, 16q, 16p, 17p, 17q, 18q, 20q, 22q, Y

which likely represent the early genetic changes in prostate cancer progression.

3
Genetic Alterations in Prostate Cancer

Evolution in techniques has allowed more refined analysis of somatic genetic changes in prostate cancer cells. Methods that have been used include cytogenetic analysis, FISH, and comparative genomic hybridization (CGH). More recently, molecular analysis has focused on specific oncogenes and tumor suppressor genes (Table 2), growth factors and their receptors, and the AR. Whereas cytogenetic analysis has yielded important information regarding hematologic malignancies, interpretation of solid tumor karyotypes is more difficult. In addition, prostate cancer is characterized by low mitotic rates, tumor heterogeneity, significant stromal elements, and poor morphology of metaphase spreads. Nevertheless, the most common changes observed are nonrandom loss of the Y chromosome and gain of chromosome 7. Others have observed deletions of chromosomes 7q, 8p, and 10q, as well as structural aberrations of chromosomes 8p22, 10q24, and 1. These chromosomal abnormalities were observed more frequently in poorly differentiated, locally extensive tumors with poor clinical outcomes. Identification of double minutes and homogeneously staining regions, features associated with oncogene amplification, has been rare in prostate cancer specimens. FISH and CGH data have confirmed cytogenetic findings and enabled detection of additional areas of change in prostate cancer cells. Aneuploidy occurs in 66 to 100% of prostate cancers. The most common alterations include chromosome 8 (23%) and chromosome 7 (20%), with involvement of chromosomes 10, 12, X, and Y occurring less frequently. Aneuploid tumors demonstrated more frequent disease progression, most notably with chromosomes 8 and Y, and may have use as prognostic markers for locally advanced prostate cancer. Chromosome 7 trisomy was noted more often in tumors of higher stage and Gleason grade. Others have postulated that S-phase fraction, like DNA ploidy, correlates with both pathologic features and clinical outcomes. However, the importance of both DNA ploidy and S-phase fraction over standard clinical and pathologic information remains unclear. Comparative genomic hybridization allows genome-wide survey for regions of amplification or loss by hybridizing DNA from normal and tumor tissues with normal metaphase

Tab. 2 Genes potentially involved in prostate carcinogenesis.

Tumor suppressor genes
N33
MX11
P53
RB1
DCC
Oncogenes
MYC
ERBB2
RAS
Metastatic genes
CDH1 (E-cadherin)
CTNNA 1 (α-catenin)
CCAM
KA11
CD44
PTEN
Androgen cascade
Androgen receptor
HSD3B2 (3β-hydroxysterold dehydrogenase)
SRD5A2 (5α-reductase)

chromosomes. Alterations as detected by CGH are present in up to 70% of prostate cancer cases, with losses five times more common than gains. Areas of losses involve chromosomes 2q, 5q, 10q, 13q, 15q, 16q, 16p, 17p, 17q, 18q, 9p, 20q, and Y, whereas gains involve chromosomes 11p, 1q, 3q, 9q, 20, 23, and 2p. New alterations revealed by CGH include loss at chromosomes 6q, 9p, and 22q and amplifications at chromosome 8q24. FISH and molecular methods of restriction fragment length polymorphism and PCR have generally confirmed these regions of change detected by CGH. The more frequent finding of DNA loss suggests that inactivation of tumor suppressor genes in these areas may play an important role in prostate cancer development. More recent studies have specifically examined alterations in Y chromosome number. It was found that prostate cancer cell lines were able to consistently induce an increase in the number of copies of the Y chromosome in three murine tumor cell lines *in vivo*. FISH demonstrated that the majority of cells had two to three copies of the Y chromosome and that the amplification was specific, without other chromosomes being involved. However, significant changes were not found in Y chromosome number. Using touch preparations from prostate tumors, 42 specimens were analyzed by whole Y chromosome paint FISH. Only a single case of Y chromosome gain, and no evidence of loss of the entire Y chromosome, was found.

4
Tumor Suppressor Genes

Cancer development relies on alterations in normal homeostatic mechanisms exhibiting precise control on cellular proliferation. The genes involved in negative regulation of cell growth can be disrupted, thus allowing tumorigenesis. These tumor suppressor genes are recessive and typically require inactivation of both alleles. As mentioned, sites of chromosomal deletion from LOH studies may suggest the presence and location of tumor suppressor genes. Loss of part, or all, of the short arm of chromosome 8 (8p) is the most common genetic defect in prostate cancer and has been confirmed by several studies. Loss of heterozygosity studies in other cancers such as lung, colon, and breast also demonstrate frequent chromosome 8p loss, further suggesting the presence of a potential tumor suppressor gene that may be fundamental in tumor development. Smaller regions of deletion, chromosomes 8p11-12 and 8p22, have been reported. Only recently has a candidate tumor suppressor gene been identified from the chromosome 8p22 region. This *N*33 gene, expressed in most tissues, is downregulated in colon cancer cell lines; however, studies in prostate cancer cell lines have not revealed mutations in *N*33 or evidence of downregulation. No other candidate tumor suppressor gene has been reported in these regions. Another frequent site of loss, initially described by classic cytogenetic analysis, is the terminal region of chromosome 10q (10q24-25). A potential tumor suppressor gene, *MXI1*, has been mapped to this chromosomal region. The protein acts as a negative regulator of the *MYC* oncogene, and its inactivation may potentially result in increased activity of the MYC protein. Mutations of the *MXI1* gene have been found to be rare, and it has been demonstrated that the allelic losses of chromosomes 10q24-25 occur near the chromosome 10q23-24 boundary and not at the *MXI1* locus. 44, 45 P53 is one of the

most widely studied tumor proteins, and abnormalities of the *P53* tumor suppressor gene are the most common genetic alterations associated with human malignancy. P53 plays a key role in the regulation of the cell cycle, ensuring DNA integrity by negatively affecting cell growth after DNA damage (Fig. 2). Overall, the incidence of LOH and mutations at the *P53* locus (chromosome 17p13.1) in prostate cancer is less than in other malignancies, ranging between 10 and 30%. Evidence suggests that changes in P53 are important in the subset of patients with advanced disease and thus may be a late event. Studies of cell lines derived from metastatic deposits demonstrated frequent *P53* mutations, and transfection of wild-type P53 resulted in suppression of growth. Ability to detect P53 by immunohistochemistry often accompanies mutations in the gene caused by increases in the protein half-life, thus allowing study of paraffin-embedded specimens. Nuclear accumulation of P53 correlated with metastatic disease or androgen independence. In addition, P53-positive tumors were found to have a greater frequency of *P53* mutations. These mutations are primarily localized to exon 7. An association between P53 expression and poor clinical outcome has been reported. Moreover, alterations in P53 may affect and therefore predict response to radiation therapy. Several regulatory genes, including *P53*, *BCL2*, *CMYC*, and *RAS*, have been implicated in the apoptotic process. It is postulated that defects in these genes may limit the ability of cells to induce apoptosis; thus, partial resistance to agents that act through programmed cell death may be imparted with P53 alterations. Several studies demonstrate an attenuated response to ionizing radiation in cells harboring *P53* mutations. Similarly, *P53* mutations adversely predict overall survival in patients with locally advanced disease treated with androgen deprivation after radiotherapy. Impaired apoptosis, after either radiotherapy or androgen ablation, may be important in tumors with *P53* mutations. Variations in studies of P53 may result from many factors. First, commercially available antibodies recognizing P53 differ in sensitivity and specificity. Second, quantification of P53 staining is subject to interobserver variability. Third,

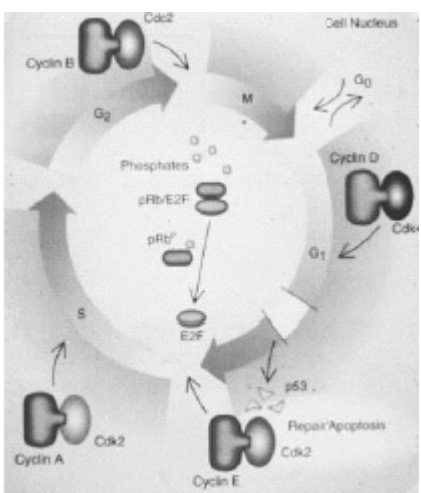

Fig. 2 Overview of the role of P53, RB, and related proteins in the cell cycle and regulation of apoptosis.

ability to detect *P53* mutations is dependent on the technique used, such as temperature or denaturing gradient gel electrophoresis or single-stranded conformation polymorphisms. In addition, heterogeneity in the distribution of *P53* alterations can result in variations, depending on which regions of the tumors are examined. The retinoblastoma tumor suppressor gene (*RB1*) is associated with the development of retinoblastoma in the immature retina as well as tumors of the breast and lung. Because of the frequent LOH of chromosome 13q (30 to 35%) and the location of *RB1* at chromosome 13q14, *RB1* has been postulated to be a tumor suppressor gene involved in prostate tumorigenesis. The phosphoprotein likely interacts with transcription factors, such as the E2F family, and negatively regulates cell growth. Although decreased or absent RB protein expression was found in a third of tumors with LOH, few mutations of the *RB1* gene have been identified. The DU145 cell line possesses a mutated *RB1* gene, and reintroduction of wild-type *RB1* is able to reduce tumorigenicity. Using single-stranded conformation polymorphism analysis of ribonucleic acid (RNA), Kubota and colleagues revealed that 4 of 25 primary cancers (16%) had *RB* alterations. Interestingly, LOH at *RB* has been found not only in cancer specimens but also in benign prostatic hyperplasia (BPH) samples. More recent data do not demonstrate prognostic significance of altered RB protein expression. The true importance of *RB1* and exact nature of LOH at 13q remain to be determined. Another region at chromosome 13, 13q21, has also been found frequently deleted in various human cancers including prostate cancer. Frequent genomic deletion and loss of expression as well as cell growth suppression indicate that KLF5 is a reasonable candidate for the tumor suppressor gene at 13q21 in prostate cancer.

In prostate cancer, chromosome 18q locus LOH has been reported in up to 45% of primary tumors. The deleted locus in colorectal carcinoma (*DCC*) gene is a candidate tumor suppressor gene located in this region and possesses homology with neural cell-adhesion molecules. Although mutations of the gene have not been reported, decreased expression of the *DCC* gene, as well as allelic loss, suggest a role in prostate cancers. It has been suggested that alternative tumor suppressor genes may be located at chromosome 18q21.

Pentaerythritol tetranitrate (PTEN) (the phosphatase and tensin homolog deleted on chromosome-10), a dual specifity phosphatase, is a tumor suppressor gene whose inactivation has been associated with many different types of cancer including prostate cancer. Combined loss of PTEN and p27 expression is associated with tumor cell proliferation by Ki-67 and increased risk of recurrent disease in localized prostate cancer.

5 Oncogenes

The second major class of genes implicated in tumor development is the so-called oncogenes. Initially recognized in cells transformed by retroviruses, oncogenes are derived from normal cellular genes, or proto-oncogenes, that contribute to tumorigenesis once activated by mutation or increased expression. The typical functions of these proto-oncogenes in normal cells include control of differentiation and proliferation. The critical role of oncogenes has been well documented

in many malignancies. Unlike other cancers, though, prostate cancer initiation and progression do not predominantly rely on oncogenes. As discussed above, genetic losses are much more common than DNA amplification. Nevertheless, region 8q, specifically 8q24, is the only region to demonstrate significant gain or amplification (5 to 16%). *MYC*, located at chromosome 8q24, is a member of the *MYC* proto-oncogene family. *MYC* encodes a nuclear DNA-binding phosphoprotein involved in transcriptional regulation and is frequently amplified in breast, lung, and cervical cancers. Amplification, rearrangement, and overexpression of *MYC* has been documented in the LNCaP cell line but not in primary tumor specimens. Elevated *MYC* expression has been shown in prostate cancer, compared to BPH, by Northern blot analysis. Others have suggested that higher *MYC* levels, seen with higher grades and lymph node metastases, correlate with biologic aggressiveness. Reporter gene and mobility shift assays demonstrated that c-Myc could repress *TMEFF2* gene (a transmembrane protein that can inhibit prostate cancer cell growth) expression through its cognate site. In light of the role of *TMEFF2* in inhibiting cell growth, its suppression may contribute to the oncogenic properties of c-Myc. Conflicting data from *in situ* hybridization do not confirm any role of *MYC* in prostate cancer, and additional studies are required. Another oncogene, *ERBB2* or *HER2/NEU* located at chromosome 17p21, has been studied in prostate cancer. The proteins in the *ERBB* family are transmembrane receptors with tyrosine kinase activity, including the *ERBB1* gene encoding the epidermal growth factor (EGF) receptor. *ERBB2* is implicated in cancers of the breast and ovaries and is typically overexpressed as a result of gene amplification or transcriptional and posttranscriptional processing. The results do not clearly show increased ERBB2 expression in prostate cancer; some have found higher protein expression in prostate cancer, whereas others demonstrate overexpression in BPH. Similarly, the prognostic value of *ERBB2* amplification and P185*NEU* protein expression is controversial. The finding of P185*NEU* in PIN and carcinoma by immunohistochemistry suggests that *ERB* oncogenes may be an early event in prostate transformation and provide further evidence for PIN as a pre-neoplastic state. Fourteen molecular studies do agree that *ERBB2* amplification is not the mechanism of activation in prostate cancer. As with P53 studies, technical factors, such as method of antigen retrieval and type of antibody used, may significantly influence the results of immunohistochemical staining. It has recently been demonstrated that FISH is more sensitive in the detection of *HER2/NEU* abnormalities, and further studies should investigate the use of this method. Recent studies suggest that ErbB1/ErbB2 RTKs play an important role in the biology of androgen-independent prostate cancer and provide a rationale for clinical evaluation of inhibitors targeted to this pathway. Analysis for Her-2 may improve prognostic algorithms for clinically relevant endpoints other than biochemical relapse.

The *RAS* oncogene family is involved in a wide variety of human cancers. The three proteins (H-RAS, K-RAS, and N-RAS) are guanosine triphosphate binding proteins essential for signal transduction. Point mutations within codons 12, 13, and 61 are the most common activating changes found that lead to uncontrolled cell growth. Immunohistochemistry has suggested high levels of RAS expression in prostate cancer; overexpression correlates

with progression and metastatic disease rather than with tumor initiation and has limited value in diagnosis and staging. *RAS* mutations, detected by molecular techniques, are rare in Western populations (<5%) and almost absent in localized prostate cancer. In Japan, however, mutations in the *RAS* gene have been found in up to 25% of cases. These data suggest that different mechanisms of prostate cancer initiation and progression may exist within different ethnic or geographic groups. Recent studies showed that stable expression of Ras effector-loop mutants that activate the Ras/MAP kinase pathway is sufficient to reduce the androgen requirement of LNCaP prostate cancer cells for growth, prostate-specific antigen (PSA) expression, and tumorigenicity; and attenuation of Ras signaling restores androgen sensitivity to hormone-refractory C4-2 prostate cancer cells.

6
DNA Repair Genes

The classic paradigm of tumor development involves oncogenes and tumor suppressor genes. More recent research describes a third class of genes that lead to tumorigenesis, those involved in DNA repair. Maintenance of genomic integrity is essential to prevent the acquisition of mutations; thus, genetic instability allows accumulation of changes, which may lead to tumor formation. It has been estimated that mutation in up to 100 genes can result in genomic instability. Efforts have focused primarily on microsatellite instability, originally described in hereditary nonpolyposis colorectal cancer (HNPCC). Mutations, either inherited or acquired, in DNA mismatch repair genes lead to the replication error phenotype (RER). Several of these genes have been identified in humans (*MSH2, MLH1, PMS1, PMS2*) and are homologous to the same genes in yeast. Defects are manifested by variations in the lengths of noncoding mono-, di-, tri-, and tetranucleotide repeats, the so-called microsatellites. Whereas microsatellite instability is found in HNPCC as well as sporadic colorectal, endometrial, stomach, and pancreatic cancers, the frequency in prostate cancer ranges from 5 to 70%. Two studies reported high frequency of alterations at specific microsatellite loci, although microsatellite instability at multiple loci is rare in prostate cancer. We found frequent microsatellite instability (45%) in prostate cancer using normal and tumor cells of 40 microdissected prostate cancer specimens. It is essential to use pure epithelial cells isolated by microdissection from 5-μ sections of prostate tissue. Variations in results may be the product of differences in PCR primer and tissue selection as well as definitions of RER. Recent reports have not found the RER phenotype in prostate cancer, with the more common finding being focal instability and the lack of widespread microsatellite instability. Evidence of microsatellite instability in 19% of cancers has been found, with no difference between clinically apparent tumors and incidentally discovered cancers at transurethral prostatectomy. In an analysis of 10 loci in 60 prostate cancers, 113 RER was found in both areas of cancer and PIN but rarely in areas of glandular hyperplasia. Racial differences have been noted in the frequency of microsatellite instability. Another report suggested that the frequency of RER is much higher in the United States when compared with Japan. The DU145 cell line demonstrates microsatellite instability and mutations in *MLH1* and *PMS2*. Overall, RER is not

predominant in most prostate cancers but may be associated with high-grade and advanced tumors, although this remains controversial.

7 Telomerase

Telomeres are repetitive, noncoding DNA sequences found at the ends of all eukaryotic chromosomes that play a role in protection and replication of the chromosomes, as well as prevent loss of genetic coding regions. In humans, hundreds of TTAGGG tandem repeats compose the telomeres. Normal cells lose between 50 and 200 of these nucleotides at each mitosis and, after 50 to 100 cell doublings, a critical amount of sequence is lost and the cells senesce. Thus, telomeres and their shortening serve as a mitotic clock. Telomerase, a ribonucleoprotein enzyme, is able to add telomeric sequences to the ends of newly replicating DNA and can compensate for telomere shortening. Germ-line cells and fetal cells typically have high telomerase activity, whereas normal somatic cells have undetectable telomerase activity. It is postulated that increased telomerase activity allows unlimited cell division and, therefore, is associated with cellular immortalization, an important phenotypic feature in the progression of normal cells to malignant ones. Normal prostate tissue does not exhibit telomerase activity. In an assay of benign and malignant tissue for telomerase activity using a sensitive PCR technique, a large number (84%) of cancers were strongly positive, as well as all lymph nodes with metastases. All human prostate cancer cell lines tested (DU145, LNCaP, PC3, PPC1, TSU) demonstrated telomerase activity. In contrast, no normal tissue from the cancer specimens or BPH samples from patients without evidence of cancer demonstrated telomerase activity. Benign prostatic hyperplasia tissue adjacent to areas of cancer demonstrated weak telomerase activity in 12% of cases. Paradoxically, telomere length from cancer tissue was significantly and consistently shorter than that detected in either adjacent normal or BPH tissue. Other studies have confirmed the high incidence of telomerase activity in prostate cancer, and some data suggest a correlation with Gleason grade. Interestingly, telomerase activity has been found frequently in adjacent areas of PIN (74%). These findings lead to several conclusions. First, telomerase activity in a majority of PIN and cancer further supports the hypothesis than PIN is a precursor of cancer and harbors early genetic changes. Second, telomerase activity is nearly ubiquitous in prostate cancer, and expression occurs early in prostate carcinogenesis. Third, the finding of telomerase activity in associated areas that appear histologically normal suggests that telomerase activity may be useful as a sensitive diagnostic marker for malignant tissue prior to visible, microscopic changes. The relationships between telomerase activity and expression levels of human telomerase RNA (hTR) and human telomerase reverse transcriptase (hTERT) mRNA in benign and malignant alterations of the human prostate gland has been investigated. Remarkably, one-third of all cancer and BPH tissue samples with hTR and hTERT expression lack telomerase activity. Quantitative analyses contradict the assumption that a certain threshold level of hTR or hTERT mRNA is required for telomerase activation, thus indicating that telomerase regulation in prostate cancer occurs more likely on a posttranscriptional level. Nevertheless, the observation

that hTR and hTERT mRNA levels are significantly ($p < 0.002$) correlated suggests some common mechanisms in the upregulation of hTR and hTERT expression. Because *in situ* hybridization revealed strong hTERT expression not only in all cells of the tumor glands but also in high-grade PIN foci, this upregulation seems to occur early in prostate carcinogenesis.

8
Growth Factors

Growth factors in the prostate have been studied in two settings: the production of factors within the tissue and the effects of exogenous factors on prostate cells in culture. It should be noted that virtually all studies relating to growth factor expression have been performed in normal adult prostates, which are either growth quiescent or only slowly growing. However, one report used prostates from young rats (6 to 8 weeks) to characterize expression of heparin-binding growth factors. Most studies in human tissue have used BPH specimens from older men, in whom growth is exceedingly slow. Thus, the function of growth factors in the adult prostate, as determined by current data, is likely more homeostatic than stimulatory. Further description of growth factor expression in the growing prostate is required to clarify these issues and inconsistencies. Polypeptide growth factors and their receptors play a role in both normal prostate growth and prostate cancer development. These growth factors act in either an autocrine or paracrine fashion and play a role in stimulating or inhibiting proliferation, differentiation, development, chemotaxis, and tissue repair. They are divided into four major families: EGF, fibroblast growth factor (FGF), insulin-like growth factor (IGF), and transforming growth factor-α (TGF-α). The EGF family includes EGF, TGF-α, amphiregulin, heparin-binding EGF, and *cripto*. Both EGF and TGF-α have been implicated in prostate cancer and act on the EGF receptor in a paracine manner. Autocrine stimulation of the EGF receptor may also be important in the regulation of prostate epithelial proliferation and differentiation, although this remains controversial. Prostate cancer cell proliferation is strongly reduced by the EGF receptor tyrosine kinase inhibitor ZD1839 *in vitro* on human cell lines and primary cultures. TGF-α whose overexpression in transgenic mice results in hyperplasia and dysplasia of the anterior prostate, has been found in tumor cells but not in normal prostatic cells.

The FGF family has at least 9 members; only FGF1, 2, 3, 7, and 8 have been found in prostate tissue. Moreover, their role in prostate cancer is currently unclear. Increased expression of both FGF7 and FGF8 has been reported in adenocarcinoma and was associated with hormone-refractory disease and higher grade, respectively. Interestingly, the stroma produces FGF7 exclusively and is the site of most FGF messenger RNA (mRNA) expression. This further confirms the vital role of the prostatic stoma and fibroblasts in affecting epithelial cell behavior and carcinogenesis. In the presence of an active AR, the combined production of FGF2 and FGF-BP may play an important role in the progression of prostate cancer through the selection of AR-clones expressing high levels of Bcl-2. By using the Transgenic adenocarcinoma mouse prostate (TRAMP) model, the expression of FGFR1iiib in intraductal vasculature and the expression of FGF-2 protein were found to be concomitant with the emergence of

PIN. These observations implicate specific changes in the FGF axis with the initiation and progression of prostate cancer, and underscore the utility of animal models to identify specific molecular changes in early disease. Fibroblast growth factor, also known as keratinocyte growth factor, is expressed only by epithelial cells, and studies in the developing mouse seminal vesicle have shown it to be an important paracrine mediator of androgen action. Similar effects may be important in prostate cancer and in epithelial cell proliferation and morphology.

The IGF family, comprised of IGF-I, IGF-II, and relaxin, is involved in cell growth, cell division, and apoptosis. Insulin-like growth factor-II is an autocrine and paracrine regulator of cell proliferation in the adult prostate. In androgen-insensitive prostate cancer, the IGF-I receptor is overexpressed, and IGF-II autocrine stimulation results in uncontrolled proliferation. Typically, the IGF-II gene demonstrates imprinting, in which one allele is inactivated and silent. Bi-allelic expression in cases of cancer has been reported and it has also been suggested that loss of imprinting may be involved in abnormal growth in prostatic tissue. The Physician's Health Study prospectively examined a matched population of 152 men for IGF-I. The study concluded that higher plasma IGF-I levels were associated with higher rates of prostate cancer; men in the highest quartile were 4.3 times more likely to develop prostate cancer. These findings were independent of PSA level, stage, and grade. The TGF-β family has three isoforms in humans and regulates cell cycle kinetics by acting on RB, MYC, and cyclin-dependent kinase inhibitor activity. Transforming growth factor-β, detected in normal and neoplastic prostate tissue, is an inhibitor of normal prostate epithelial cell growth, and its activity is modulated by other growth factors, androgens, and the extracellular matrix. Conversely, TGF-β can also stimulate cell growth and contribute to tumor aggressiveness. Immunohistochemistry revealed increasing levels of TGF-β1 in tumors with metastatic and nodal involvement. Plasma TGF-β1 levels may have use as a tumor marker or predictor of clinical prognosis. The TGF-β effects are mediated through the heteromeric complex of the type I and II receptors. Thus, loss of function of either receptor results in resistance to TGF-β inhibitory signals. Decreased levels of type I and II receptors are correlated with higher grade and advanced disease. Activation of TGF-β signaling in human prostate cancer cells suppresses tumorigenicity via deregulation of cell cycle progression and induction of caspase-1 mediated apoptosis.

Other growth factors implicated include hepatocyte growth factor, platelet-derived growth factor, nerve growth factor, and interleukin-6. Further investigation is required to elucidate their exact roles in prostate carcinogenesis.

9 Metastasis Suppressor Genes

The invasive nature of prostate cancer is dependent on the loss of usual cell–cell and cell–matrix interactions. Several candidate genes involved in this process have been identified in prostate cancer that typically act as tumor suppressor genes. Ultimately, a metastatic phenotype may be acquired after inactivation of one or many of these genes. As discussed above, loss at chromosome 16q, as detected by all methods, is frequent in prostate cancer and led to the search for a tumor suppressor

gene. E-cadherin, located at chromosome 16q22.1, has been studied thoroughly. The protein is a calcium-dependent, epithelial cell–cell adhesion molecule and plays a vital role in the invasive potential of cancer cells (Fig. 3). Decreased E-cadherin expression is found in prostate cancer and is associated with higher grade and worse clinical outcome. In addition, metastases express less E-cadherin than primary tumors. Blocking E-cadherin function using antibodies or antisense transcripts increases invasive behavior. Conversely, reexpression of E-cadherin in epithelial cells can induce a reversion to a noninvasive phenotype (Fig. 4). Newer techniques in immunohistochemistry may reduce the inconsistencies from formalin fixation and autolytic artifacts that characterized prior studies. An alterative approach has been described for fixing prostatectomy specimens that yields samples amenable to E-cadherin immunohistochemistry as well as extraction of longer DNA fragments.

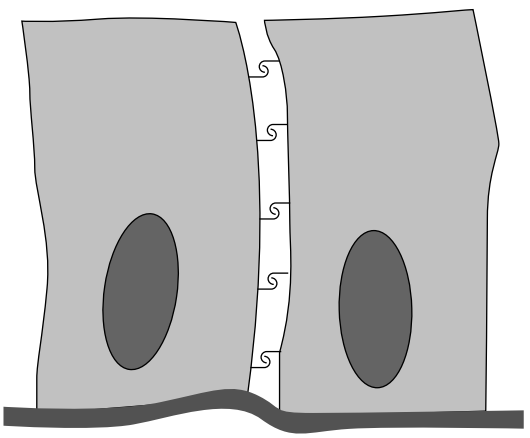

Fig. 3 Illustration depicting the role of E-cadherin. The surface molecule interacts with other E-cadherin molecules, contributing to epithelial cell adhesion. The cytoplasmic domain, not shown, normally interacts with intracellular catenin proteins.

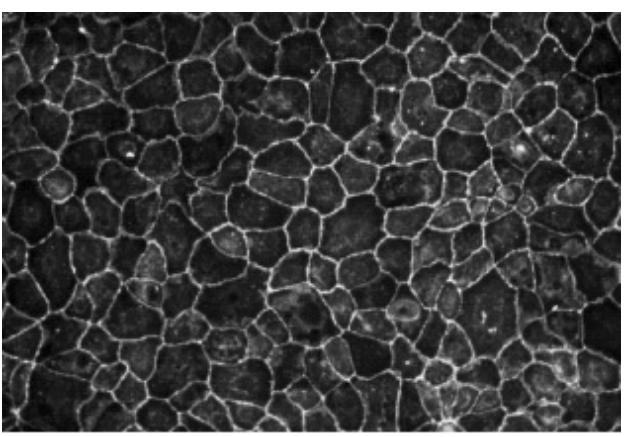

Fig. 4 Immunocytochemistry after staining using an antibody to E-cadherin. The immortalized, but nontumorigenic, prostate epithelial cells demonstrate strong staining for E-cadherin in the areas of cell–cell adhesion.

Mutations in the E-cadherin gene (*CDH1*) are rare, suggesting other mechanisms of regulation and inactivation. Hypermethylation of the 5′-CpG islands of *CDH1* may silence the gene, and this has been found in E-cadherin–negative cell lines. After treatment with demethylating agents, E-cadherin RNA and protein levels were partially restored. E-cadherin function is also dependent on sex steroids, which increase expression, and interaction with cytoskeletal and other adhesion molecules such as the catenins. The catenins are a family of three (α-, β-, γ-) proteins involved in intercellular adhesion and signal transduction. They associate with the cytoplasmic domain of E-cadherin to allow coupling to the microfilament cytoskeleton. Because of frequent chromosome 5q losses and mapping of α-catenin (*CTNNA1)* to chromosome 5q21-q22, they are of particular interest. The PC-3 cell line expresses E-cadherin; however, α-catenin is absent owing to homozygous deletion of the gene. Reexpression of α-catenin by either transfection with α-catenin complementary DNA or transfer of chromosome 5 was associated with a return of cell–cell contact and decreased tumorigenicity in mice. Thus, both E-cadherin and α-catenin function is required in reducing invasive potential. Like E-cadherin, loss of α-catenin expression, as determined by immunohistochemistry, predicts invasiveness and poor prognosis. Further study is needed of β- and γ-catenin. Only 5β-catenin mutations in 104 specimens have been found. In addition, the adenomatous polyposis gene product (APC) interacts with β-catenin. APC is also located at chromosome 5q21, and inactivation may lead to altered cell–cell function. Other epithelial cell-adhesion molecules may act as repressors of growth and invasion. One that has been evaluated is CCAM, a member of the immunoglobulin gene superfamily. CCAM expression in human prostate cancer cells reduces anchorage-independent growth and tumor invasion *in vitro* and *in vivo* respectively. Antisense CCAM expression in less tumorigenic cell lines was able to increase tumorigenicity. Studies in human tissue reveal absence of CCAM in prostate cancer and decreased staining in BPH and high-grade PIN. Members of the integrin family mediate interactions between cells and extracellular matrix proteins, affecting cell migration, proliferation, and invasion. These heterodimeric cell-surface receptors are composed of noncovalently associated α- and β- subunits. Overall, downregulation of the various integrin subunits occurs in prostate cancer development. These include α2, α4, α5, αv, and β4. Conversely, increased expression of integrin α6 is associated with more invasive activity. An increased invasive phenotype in prostate cancer is associated with $\alpha3\beta1$, $\alpha4\beta1$, and $\alpha6\beta1$. More recently, it has been found that $\alpha v\beta3$ integrin expression in prostate cancer generates a migratory phenotype, which is modulated by a focal adhesion kinase signaling pathway. Other integrins, $\alpha6\beta4$ and $\alpha2\beta1$, attach the cells to the basement membrane and bone matrix, respectively. The *KAI1* metastasis suppressor gene is found on chromosome 11p11.2. It belongs to a family of membrane glycoproteins and is thought to participate in cell–cell interaction and cell migration. KAI1 is expressed in normal human prostate epithelial cells but is absent in cell lines derived from metastases, such as DU145, PC-3, and LNCaP. Reintroduction of the *KAI1* gene into the Dunning rat AT6.1 cell line suppressed the metastatic potential. Downregulation of KAI1 is found in the progression of prostate cancer, with decreased protein expression in 70% of

primary cancer specimens and in almost all patients with lymph node metastases or hormone-refractory disease. A recent study showed that KAI1 expression could be a predictor of stage A prostate cancer progression. Allelic loss or mutations do not appear to be the mechanism of KAI1 decrease. EWI2/PGRL, a KAI1 associated protein, was recently found to be important for KAI1-mediated suppression of cancer cell migration.

Located near *KAI1*, CD44 is a transmembrane protein associated with cell–cell and cell–extracellular matrix interactions. Conflicting data exist regarding CD44's function in prostate cancer. Loss of expression of CD44 standard isoform (CD44s) correlated with poor prognosis in several studies, whereas others demonstrate increased CD44 expression associated with high-grade and metastatic potential. Neutralizing antibodies to CD44 inhibited cell proliferation and basement membrane invasion. In another report, transfection-enhanced expression suppressed metastatic ability in the AT3.1 cell line without affecting *in vivo* growth or tumorigenicity. Recent study indicates the important role of CD44 methylation in the progression and metastasis of prostate cancer. Pentaerythritol tetranitrate is a tyrosine phosphatase, mapping to chromosome 10q23 that negatively regulates cell migration and cell survival. In addition, its homology to a protein that interacts with actin filaments at focal adhesions suggests that PTEN may regulate cell invasion and metastasis. Loss of *PTEN* may occur in 13 to 37% of primary prostate cancers. Increased frequency of mutations and LOH is found in cases with metastases. Recent immunohistochemistry confirms these observations. Total absence of PTEN expression correlated significantly with Gleason score and advanced pathologic stage. Pentaerythritol tetranitrate, along with CD44 and KAI1, has potential as a marker to evaluate the metastatic ability of prostate cancer, and further studies are warranted.

10 DNA Methylation

Abnormal patterns of DNA methylation often characterize human cancer cells. These include hypermethylation, redistribution of methylation, and demethylation of normally methylated regions. Although hypomethylation has been reported in prostate cancer, its significance is unclear. More recently, hypermethylation of particular areas has been noted and characterized. The most important sites of hypermethylation are the CpG islands, with a high density of C-G dinucleotide sequences, found in the 5′ regions of genes. Methylation of CpG-rich islands may be associated with transcriptional inactivation of the associated gene. This has been demonstrated for the CDKN2 (*P16*) gene at chromosome 9p21. The product, a cyclin-dependent kinase inhibitory protein, controls passage through the G1 phase of the cell cycle. Inactivation by homozygous deletion, found in 20% of prostate specimens studied, may induce progression through the cell cycle in tumors. Analysis of the CpG-rich promoter region in exon 1 of *P16* in the prostate cancer cell lines PC-3, PPC1, and TSU revealed dense methylation in these areas. In addition, this hypermethylation correlated with lack of mRNA expression as determined by reverse transcriptase PCR. *In vitro* treatment with the demethylation agent 5-aza-2′-deoxycytidine induced reexpression of CDKN2 transcripts. Promoter methylation has been found in 13

and 8% of primary and metastatic tumor samples, respectively, whereas no normal prostate tissues were methylated at the CDKN2 gene promoter. Gu and colleagues were unable to demonstrate intragenic alterations in methylation pattern in early-stage cancers, suggesting that *P16* methylation may be involved in disease progression rather than initiation. The π-class glutathione-S-transferase (*GSTP1*) gene at 11q13 is important in the pathway, preventing damage from electrophilic metabolites of carcinogens and reactive oxygen species. Decreased GSTP1 expression was found to accompany human prostate carcinogenesis. Analysis of a CpG island in the promoter region of the *GSTP1* gene revealed hypermethylation in all prostate cancer specimens examined but no aberrant methylation in 12 nonmalignant tissues, including BPH. Further immunohistochemistry in 91 cases of cancer was unable to detect GSTP1 expression in almost all samples (97%). These findings were confirmed *in vitro*, with protein expression limited to prostatic cell lines with *GSTP1* alleles hypomethylated at the promoter sequences. Additional work has substantiated that CpG-island hypermethylation is abnormal in both alleles in most prostate cancers; in normal tissue, the entire CpG island is unmethylated with extensive methylation outside the *GSTP1* gene. *GSTP1* hypermethylation and associated reduction in GSTP1 expression have also been demonstrated in areas of PIN. The high frequency of *GSTP1* CpG hypermethylation may serve as a biomarker in prostate cancer detection and staging. Recent study on quantitation of GSTP1 hypermethylation can accurately detect the presence of cancer even in small, limited tissue samples. It represents a promising diagnostic marker that could be used as an adjunct to tissue biopsy as part of prostate cancer screening.

Another locus of hypermethylation in the prostate has been found to be methylated in cancer (*HIC1*) gene on chromosome 17p13.3. The methylation of NotI restriction sites at D17S5 is a common occurrence in prostate cancer and is thus postulated to be an important step in carcinogenesis. However, nonmalignant prostate, but not seminal vesicle tissues also demonstrated *HIC1* hypermethylation, suggesting that the locus, and its tissue-specific methylation, may be important in the organ's susceptibility to uncontrolled growth. The CpG island in the *CD44* transcriptional regulatory region is also often hypermethylated. This leads to transcriptional inactivation of the gene, as with *GSTP1*, and a resultant decrease in CD44 expression. CD44-negative cell lines, such as LNCaP, demonstrate hypermethylation of the promoter, whereas CD44-expressing cells do not exhibit this methylation. Examination of primary prostate cancer specimens confirms the increased presence of CpG hypermethylation when compared to normal tissues. Aberrant methylation has also been found at the endothelin B (*ETB*) receptor gene. Endothelin B is not expressed in human prostate cancer cell lines and shows decreased binding to endothelin-1 in prostate cancer tissues. The low ETB expression in prostate cancer was correlated with methylation of the CpG-island sequences in the *ETB* gene. Normal tissues were unmethylated at the *ETB* transcriptional regulatory region, whereas treatment of cancer cell lines with a demethylating agent induced *ETB* mRNA expression. Thus, CpG-island methylation and transcriptional silencing appear to be important in many genes involved in prostate cancer.

11
Androgen Receptor

The AR mediates the actions of androgens in both the normal and abnormal prostate cell. This, combined with the necessity of androgens in prostate carcinogenesis and the inevitable development of androgen insensitivity with continued androgen deprivation, has focused much effort into examining the role of AR in early and late prostate cancer. Inherited susceptibility to prostate cancer may be transmitted by AR polymorphisms. Two microsatellite repeats, CAG and GGN, are located within exon 1 of the *AR* gene. Whereas expansion of the CAG repeats from 40 to 62 is associated with spinal and bulbar muscular atrophy (Kennedy's syndrome), less than 19 CAG repeats are associated with a 1.52 relative risk for prostate cancer ($p = .04$). In addition, men with shorter CAG microsatellites were at higher risk for metastatic disease. A 3% decrease in prostate cancer risk with each additional CAG repeat has been noted, and various studies have found similar results. However, variations in the lengths have not been consistently associated with altered AR function. Although patients with Kennedy's syndrome have been reported to possess reduced androgen binding capacity, other investigators have found normal hormone binding, transactivation, and transrepression potential in these men. The frequency of CAG repeat lengths differs among the ethnic groups. Shorter repeats are more common in black men, who have a higher incidence of prostate cancer. Conversely, more CAG repeats are found in populations at lower risk such as Asians and Whites. This polymorphism could, in part, explain the ethnic differences in prostate cancer incidence. Length of the GGN repeat has been inconsistently associated with potential for prostate cancer development. Some have found any deviation from a repeat length of 16 to impart a higher risk (odds ratio 1.18), whereas others report lengths less than 16 to be a risk factor. No significant differences were found between cancer cases and controls with respect to GGN repeat length. The observed gain of chromosome Xq in hormone-refractory prostate cancer suggested involvement of the AR, located at chromosome Xq12. Amplification of AR was detected in 30% of recurrent cancers, but none of the primary tumors, in patients receiving hormonal therapy. *AR* gene amplification leads to increased mRNA levels, and high AR expression is seen in androgen-independent tumors. Almost all primary and metastatic lesions express AR, however, prior to endocrine treatment. Thus, mechanisms other than loss of AR are involved in progression to a metastatic or androgen-independent state. Indeed, prior androgen ablation may select for *AR* gene amplification. Significantly longer median survival has been found in patients with *AR* amplification than those without it. *AR* mutations are relatively infrequent, ranging from 0 to 41%. This further suggests that the development of prostate cancer occurs in the presence of normal AR expression. Moreover, the AR pathway likely remains important in advanced and hormone-refractory cancers. Often, mutations are found in metastatic but not primary tumors. *AR* mutations have been found in half of the hormone-independent, metastatic lesions, all of which were clustered in the steroid-binding domain (exons 4 to 8). Changes in this area may lead to loss of repression of AR transactivation in the absence of androgen or increased affinity to steroids such as estrogen and progesterone. Over 50 *AR* mutations have been identified and

can occur in the amino-terminal domain (exon 1), DNA-binding domain (exons 2 to 3), and hinge region (exon 4). Mutations in the AR and overexpression may provide the cells with a means of continued growth despite low androgen conditions, a sensitization of the AR pathway. Other means of sensitizing the pathway include activation of AR in a ligand-independent manner by growth factors and cytokines or amplification of coactivators. Finally, some prostate cancer cells have bypassed the AR pathway altogether. A recent study demonstrated the important roles of AR-associated protein 55 (dARA55) coregulator in the AR-mediated growth of prostate cancer and thus provided evidence that AR function could be suppressed without mutation or change in AR itself.

Overall, the AR plays an important role in prostate cancer recurrence. *AR* amplification may lead to continued stimulation with low, residual androgens, whereas *AR* mutations can alter receptor specificity. New reports document an important role of AR coactivators in bypassing traditional methods of androgen ablation. A better understanding of the mechanisms of AR-dependent and AR-independent growth will improve current and target future methods of androgen manipulation in prostate cancer. Polymorphisms in other genes of the androgen-signaling cascade may contribute to the racial and ethnic differences in prostate cancer incidence and biologic behavior. 3β-hydroxysteroid dehydrogenase (type 2) (*HSD3B2)* is involved in the catabolism of dihydrotestosterone. Twenty-five alleles of the (TG)n(TA)n(CA)n dinucleotide repeat in the *HSD3B2* gene have been identified, with the 289-bp form being the most common. The shorter 275-bp allele was more common in Blacks, and longer alleles (281-bp, 302 to 340 bp) were more common in European-Americans and Asians. A similar polymorphism exists in the 3′ untranslated region of the 5α-reductase type 2 gene (*SRD5A2*), whose product converts testosterone to dihydrotestosterone. The most common allele is the (TA)0 form, without TA repeats, found in a majority of men, whereas the two other alleles are less frequent ((TA)9, (TA)18). Blacks possessed the (TA)18 allele in 18% of cases studied; this allele was not found in Asian or White populations. In addition to reporting the epidemologic finding of increased risk of cancer in Black and Hispanic men with the A49T missense substitution in *SRD5A2*, one study found high enzymatic activity of the A49T variant in an *in vitro*, mammalian system. It is thought that the *HSD3B2* and *SRD5A2* polymorphisms result in increased prostatic levels of dihydrotestosterone and therefore increased risks of prostate cancer. Additional functional and *in vivo* studies are necessary to substantiate this hypothesis.

12 Conclusions

A wide variety of genetic alterations are potentially involved in prostate carcinogenesis. An improved understanding of the molecular basis will improve screening of patients at high risk for tumor, allow more accurate assessment of prognosis prior to and after initial therapy, and direct research toward novel and rational treatment strategies.

See also Genetics, Molecular Basis of; Medicinal Chemistry.

Bibliography

Primary Literature

Alexandrow, M.G., Moses, H.L. (1995) Transforming growth factor α and cell cycle regulation, *Cancer Res.* **55**, 1452–1457.

Allen, M.V., Smith, G.J., Juliano, R., et al. (1998) Downregulation of the beta4 integrin subunit in prostatic carcinoma and prostatic intraepithelial neoplasia, *Hum. Pathol.* **29**, 311–318.

Avery, A., Standberg, M.D. (1992) Chromosomal abnormalities and related events in prostate cancer, *Hum. Pathol.* **23**, 368–380.

Bakin, R.E., Gioeli, D., Bissonette, E.A., Weber, M.J. (2003) Attenuation of Ras signaling restores androgen sensitivity to hormone-refractory C4-2 prostate cancer cells, *Cancer Res.* **63**(8), 1975–1980.

Bakin, R.E., Gioeli, D., Sikes, R.A., Bissonette, E.A., Weber, M.J. (2003) Constitutive activation of the ras/mitogen-activated protein kinase signaling pathway promotes androgen hypersensitivity in LNCaP prostate cancer cells, *Cancer Res.* **63**(8), 1981–1989.

Bedford, M.T., van Helden, P.D. (1987) Hypomethylation of DNA in pathological conditions of the human prostate, *Cancer Res.* **47**, 5274–5276.

Behrens, J., Mareel, M.M., VanRoy, F.M., Birchmeier, W. (1989) Dissecting tumor cell invasion: epithelial cells acquire invasive properties after loss of uvomorulin-mediated cell-cell adhesion, *J. Cell Biol.* **108**, 2435–2447.

Blackburn, E.H. (1991) Structure and function of telomeres, *Nature* **350**, 569–573.

Bochum, S., Paiss, T., Vogel, W., Herkommer, K., Hautmann, R., Haeussler, J. (2002) Confirmation of the prostate cancer susceptibility locus HPCX in a set of 104 German prostate cancer families, *Prostate* **52**(1), 12–19.

Bookstein, R., MacGrogan, D., Hilsenbeck, S.G., et al. (1993) p53 is mutated in a subset of advanced-stage prostate cancers, *Cancer Res.* **53**, 3369–3373.

Bookstein, R., Shew, J.-Y., Chen, P.L., et al. (1990) Suppression of tumorigenicity of human prostate carcinoma cells by replacing a mutated RB gene, *Science* **247**, 712–715.

Bostwick, D.G., Pacelli, A., Lopez-Beltran, A. (1996) Molecular biology of prostate intraepithelial neoplasia, *Prostate* **29**, 117–134.

Bova, G.S., Carter, B.S., Bussemakers, M.J., et al. (1993) Homozygous deletion and frequent allelic loss of chromosome 8p22 loci in human prostate cancer, *Cancer Res.* **53**, 3869–3873.

Boyer, J., Umar, A., Risinger, J., et al. (1995) Microsatellite instability, mismatch repair deficiency, and genetic defects in human cancer cell lines, *Cancer Res.* **55**, 6063–6070.

Bronner, C.E., Baker, S.M., Morrison, P.T., et al. (1994) Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer, *Nature* **368**, 258–261.

Brooks, J.D., Bova, G.S., Isaacs, W.B. (1995) Allelic loss of the retinoblastoma gene in primary human prostatic adenocarcinomas, *Prostate* **26**, 35–39.

Brooks, J.D., Bova, G.S., Ewing, C.M., et al. (1996) An uncertain role for p53 gene alterations in human prostate cancers, *Cancer Res.* **56**, 3814–3822.

Brooks, J.D., Weinstein, M., Lin, X., et al. (1998) CG island methylation changes near the GSTP1 gene in prostatic intraepithelial neoplasia, *Cancer Epidemiol Biomarkers Prev* **7**, 531–536.

Brothman, A.R., Peehl, D.M., Patel, A.M., McNeal, J.E. (1990) Frequency and pattern of karyotypic abnormalities in human prostate cancer, *Cancer Res.* **50**, 3795–3803.

Brown, J.A., Alcaraz, A., Takahashi, S., et al. (1994) Chromosomal aneusomies detected by fluorescent in situ hybridization analysis in clinically localized prostate carcinoma, *J. Urol.* **152**, 1157–1162.

Buchman, E.C., Nayak, R.N., Bushman, W. (1995) Immunohistochemical staining of ras p21: staining in benign and malignant prostate tissue, *J. Urol.* **153**, 233–237.

Buttyan, R., Sawczuk, I.S., Benson, M.C., et al. (1987) Enhanced expression of the c-myc protooncogene in high-grade human prostate cancers, *Prostate* **11**, 327–337.

Cairns, P., Okami, K., Halachmi, S., et al. (1997) Frequent inactivation of PTEN/MMAC1 in primary prostate cancer, *Cancer Res.* **57**, 4997–5000.

Cairns, P., Polascik, T.J., Eby, Y., et al. (1995) Frequency of homozygous deletion at p16/CDKN2 in primary human tumors, *Nat. Genet.* **11**, 210–212.

Cannon, L., Bishop, D.T., Skolnick, M., et al. (1982) Genetic epidemiology of prostate

carcinoma in the Utah Mormon genealogy, *Cancer Surv.* **1**, 48–69.

Carpten, J., Nupponen, N., Isaacs, S., Sood, R., Robbins, C., Xu, J., Faruque, M., Moses, T., Ewing, C., Gillanders, E., Hu, P., Bujnovszky, P., Makalowska, I., Baffoe-Bonnie, A., Faith, D., Smith, J., Stephan, D., Wiley, K., Brownstein, M., Gildea, D., Kelly, B., Jenkins, R., Hostetter, G., Matikainen, M., Schleutker, J., Klinger, K., Connors, T., Xiang, Y., Wang, Z., De Marzo, A., Papadopoulos, N., Kallioniemi, O.P., Burk, R., Meyers, D., Gronberg, H., Meltzer, P., Silverman, R., Bailey-Wilson, J., Walsh, P., Isaacs, W., Trent, J. (2002) Germline mutations in the ribonuclease L gene in families showing linkage with HPC1, *Nat. Genet.* **30**(2), 181–184.

Carruba, G., Miceli, D., D'Amico, D., et al. (1995) Sex steroids up-regulate E-cadherin expression in hormone-responsive LNCaP human prostate cancer cells, *Biochem. Biophys. Res. Commun.* **212**, 624–631.

Carter, B.S., Beaty, T.H., Steinberg, G.D., et al. (1992) Mendelian inheritance of familial prostate cancer, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 3367–3371.

Carter, B.S., Bova, G.S., Beaty, T.H., et al. (1993) Hereditary prostate cancer: epidemiologic and clinical features, *J. Urol.* **150**, 797–802.

Carter, B.S., Epstein, J.I., Isaacs, W.B. (1990) Ras gene mutations in human prostate carcinoma, *Cancer Res.* **50**, 6830–6832.

Castagnaro, M., Yandell, D.W., Dockhorn-Dworniczak, B., et al. (1993) Androgen receptor gene mutations and p53 gene analysis in advanced prostate cancer, *Verh. Dtsch. Ges. Pathol.* **77**, 119–123.

Chan, J.M., Stampfer, M.J., Giovannucci, E., et al. (1998) Plasma insulinlike growth factor-I and prostate cancer risk: a prospective study, *Science* **279**, 563–566.

Chen, C., Bhalala, H.V., Vessella, R.L., Dong, J.T. (2003) KLF5 is frequently deleted and down-regulated but rarely mutated in prostate cancer, *Prostate* **55**(2), 81–88.

Chen, W.C., Obrink, B. (1991) Cell-cell contacts mediated by E-cadherin (uvomorulin) restrict invasive behavior of L-cells, *J. Cell Biol.* **114**, 319–327.

Cheng, K.D., Loeb, L.A. (1993) Genomic instability and tumor progression: mechanistic considerations, *Adv. Cancer Res.* **60**, 121–156.

Cheng, L., Nagabhushan, M., Pretlow, T.P., et al. (1996) Expression of E-cadherin in primary and metastatic prostate cancer, *Am. J. Pathol.* **148**, 1375–1380.

Cher, M.L., Bova, G.S., Moore, D.H., et al. (1996) Genetic alterations in untreated metastases and androgen-independent prostate cancer detected by comparative genomic hybridization and allelotyping, *Cancer Res.* **56**, 3091–3102.

Cher, M.L., MacGrogan, D., Bookstein, R., et al. (1994) Comparative genomic hybridization, allelic imbalance, and fluorescence in situ hybridization on chromosome 8 in prostate cancer, *Genes Chromosomes Cancer* **11**, 153–162.

Ching, K.Z., Ramsey, E., Pettigrew, N., et al. (1993) Expression of mRNA for epidermal growth factor, transforming growth factor-alpha and their receptor in human prostate tissue and cell lines, *Mol. Cell. Biochem.* **126**, 151–158.

Cress, A.E., Rabinovitz, I., Zhu, W.G., Nagle, R.B. (1995) The alpha-6-beta-1 and alpha-6-beta-4 integrins in human prostate cancer progression, *Cancer Metastasis Rev.* **14**, 219–228.

Crundwell, M.C., Morton, D.G., Arkell, D.G., Phillips, S.M. (1999) Genetic instability in incidentally discovered and advanced prostate cancer, *BJU Int.* **84**, 123–127.

Cunningham, J.M., Shan, A., Wick, M.J., et al. (1996) Allelic imbalance and microsatellite instability in prostatic adenocarcinoma, *Cancer Res.* **56**, 4475–4482.

Dahiya, R., Deng, G., Chen, K.M., et al. (1996) P53 tumor-suppressor gene mutations are mainly localised on exon 7 in human primary and metastatic prostate cancer, *Br. J. Cancer* **74**, 264–268.

Dahiya, R., Lee, C., McCarville, J., et al. (1997) High frequency of genetic instability of microsatellites in prostatic adenocarcinoma, *Int. J. Cancer* **72**, 762–767.

Davis, D.L., Russell, D.W. (1993) Unusual length polymorphism in human steroid 5 alpha-reductase type 2 gene (SRD5A2), *Hum. Mol. Genet.* **2**, 820.

DeLange, T. (1994) Activation of telomerase in a human tumor, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2882–2885.

Deocampo, N.D., Huang, H., Tindall, D.J. (2003) The role of PTEN in the progression and survival of prostate cancer, *Minerva Endocrinol.* **28**(2), 145–153.

Devgan, S.A., Henderson, B.E., Yu, M.C., et al. (1997) Genetic variation of 3 beta-hydroxysteroid dehydrogenase type II in three racial/ethnic groups: implications for prostate cancer risk, *Prostate* **33**, 9–12.

Dong, J.T., Isaacs, W.B., Isaacs, J.T. (1997) Molecular advances in prostate cancer, *Curr. Opin. Oncol.* **9**, 101–107.

Dong, J.T., Lamb, P.W., Rinker-Schaeffer, C.W., et al. (1995) KAI1, a metastasis suppressor gene for prostate cancer on human chromosome 11p11.2, *Science* **268**, 884–886.

Dong, J.T., Suzuki, H., Pin, S.S., et al. (1996) Down-regulation of the KAI1 metastasis suppressor gene during progression of human prostatic cancer infrequently involved gene mutation or allelic loss, *Cancer Res.* **56**, 4387–4390.

Eagle, L.R., Yin, X., Brothman, A.R., et al. (1995) Mutation of the MXI1 gene in prostate cancer, *Nat. Genet.* **9**, 249–255.

Eastham, J.A., Truong, L.D., Rogers, E., et al. (1995) Transforming growth factor-beta-1: comparative immunohistochemical localization in human primary and metastatic prostate cancer, *Lab. Invest.* **73**, 628–635.

Edwards, A., Hammond, H.A., Jin, L., et al. (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups, *Genomics* **12**, 241–253.

Egawa, S., Uchida, T., Suyama, K., et al. (1995) Genomic instability of microsatellite repeats in prostate cancer: relationship to clinicopathological variables, *Cancer Res.* **55**, 2418–2421.

Emmert-Buck, M.R., Vocke, C.D., Pozzatti, R.O., et al. (1995) Allelic loss on chromosome 8p12–21 in microdissected prostatic intraepithelial neoplasia, *Cancer Res.* **55**, 2959–2962.

Ewing, C.M., Ru, N., Morton, R.A., et al. (1995) Chromosome 5 suppresses tumorigenicity of PC3 prostate cancer cells: correlation with re-expression of alpha-catenin and restoration of E-cadherin function, *Cancer Res.* **55**, 4813–4817.

Fearon, E.R., Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis, *Cell* **16**, 759–767.

Feilotter, H.E., Nagai, M.A., Boag, A.H., et al. (1998) Analysis of PTEN and the 10q23 region in primary prostate carcinomas, *Oncogene* **16**, 1743–1748.

Fleming, W.H., Hamel, A., MacDonald, R., et al. (1986) Expression of the c-myc protooncogene in human prostatic carcinoma and benign prostatic hyperplasia, *Cancer Res.* **46**, 1535–1538.

Fossa, A., Lilleby, W., Fossa, S.D., Gaudernack, G., Torlakovic, G., Berner, A. (2002) Independent prognostic significance of HER-2 oncoprotein expression in pN0 prostate cancer undergoing curative radiotherapy, *Int. J. Cancer* **99**(1), 100–105.

Fournier, G., Latil, A., Amet, Y., et al. (1994) Gene amplifications in advanced-stage human prostate cancer, *Urol. Res.* **22**, 343–347.

Fowler, J.E., Lau, J.T., Ghosh, L., et al. (1988) Epidermal growth factor and prostatic carcinoma: an immunohistochemical study, *J. Urol.* **139**, 857–861.

Fox, S.B., Persad, R.A., Coleman, N. (1994) Prognostic value of c-erbB-2 and epidermal growth factor receptor in stage A1 (T1a) prostatic adenocarcinoma, *Br. J. Urol.* **74**, 214.

Frixen, U.H., Behrens, J., Sachs, M., et al. (1991) E-cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells, *J. Cell Biol.* **113**, 173–185.

Fukumoto, M., Shevrin, D.H., Roninson, I.B. (1988) Analysis of gene amplification in human tumor cell line, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6846–6850.

Gao, X., Honn, K.V., Grignon, D., et al. (1993) Frequent loss of expression and loss of heterozygosity of the putative tumor suppressor gene DCC in prostatic carcinomas, *Cancer Res.* **53**, 2723–2727.

Gao, A.C., Lou, W., Dong, J.T., Isaacs, J.T. (1997) CD44 is a metastasis suppressor gene for prostatic cancer located on human chromosome 11p13, *Cancer Res.* **57**, 846–849.

Gao, X., Wu, N., Grignon, D., et al. (1995) High frequency of mutator phenotype in human prostate cancer., *Oncogene* **9**, 2999–3003.

Gery, S., Koeffler, H.P. (2003) Repression of the TMEFF2 promoter by c-Myc, *J. Mol. Biol.* **328**(5), 977–983.

Giovannucci, E., Stampfer, M.J., Krithivas, K., et al. (1997) The CAG repeat within the androgen receptor gene and its relationship to prostate cancer, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3320–3323.

Giri, D.K., Wadhwa, S.N., Upadhaya, S.N., Talwar, G.P. (1993) Expression of NEU/HER-2 oncoprotein (p185neu) in prostate tumors: an immunohistochemical study, *Prostate* **23**, 329–336.

Giroldi, L.A., Bringuier, P.P., Schalken, J.A. (1994) Defective E-cadherin function in urological cancers: clinical implications and molecular mechanisms, *Invasion Metastasis* **14**, 71–81.

Gottlieb, B., Lehvaslaiho, H., Beitel, L.H., et al. (1998) The androgen receptor gene mutations database, *Nucleic Acids Res.* **26**, 234–238.

Graff, J.R., Herman, J.G., Lapidus, R.G., et al. (1995) E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas, *Cancer Res.* **55**, 5195–5199.

Gray, I.C., Phillips, S.M., Lee, S.J., et al. (1995) Loss of the chromosomal region 10q23-24 in prostate cancer, *Cancer Res.* **55**, 4800–4803.

Grignon, D.J., Caplan, R., Sarkar, F.H., et al. (1997) p53 status and prognosis of locally advanced prostatic adenocarcinoma: a study based on RTOG 8610, *J. Natl. Cancer Inst.* **89**, 158–165.

Gu, K., Mes Masson, A.M., Gauthier, J., Saad, F. (1996) Overexpression of her-2/neu in human prostate cancer and benign hyperplasia, *Cancer Lett.* **99**, 185–189.

Gu, K., Mes-Masson, A.M., Gauthier, J., Saad, F. (1998) Analysis of the p16 tumor suppressor gene in early-stage prostate cancer, *Mol. Carcinog.* **21**, 164–170.

Gumerlock, P.H., Poonamallee, U.R., Meyers, F.J., de Vere White, R.W. (1991) Activated ras alleles in human carcinoma of the prostate are rare, *Cancer Res.* **51**, 1632–1637.

Halvorsen, O.J., Haukaas, S.A., Akslen, L.A. (2003) Combined Loss of PTEN and p27 expression is associated with tumor cell proliferation by Ki-67 and increased risk of recurrent disease in localized prostate cancer, *Clin. Cancer Res.* **9**(4), 1474–1479.

Hamdy, S.M., Aprikian, A.G., Begin, L.R., et al. (1994) ras p21 overexpression is a late event in prostate cancer, *Int. J. Oncol.* **4**, 627–631.

Harden, S.V., Guo, Z., Epstein, J.I., Sidransky, D. (2003) Quantitative GSTP1 methylation clearly distinguishes benign prostatic tissue and limited prostate adenocarcinoma, *J. Urol.* **169**(3), 1138–1142.

Hardy, D.O., Scher, H.I., Bogenreider, T., et al. (1996) Androgen receptor CAG repeat lengths in prostate cancer: correlation with age of onset, *J. Clin. Endocrinol. Metab.* **81**(12), 4400–4405.

Heidenberg, H.B., Sesterhenn, I.A., Gaddipati, J.P., et al. (1995) Alteration of the tumor suppressor gene p53 in a high fraction of hormone refractory prostate cancer, *J. Urol.* **154**, 414–421.

Hsieh, J.T., Luo, W., Song, W., et al. (1995) Tumor suppressive role of an androgen-regulated epithelial cell adhesion molecule (CCAM) in prostate carcinoma cell revealed by sense and antisense approaches, *Cancer Res.* **55**, 190–197.

Huss, W.J., Barrios, R.J., Foster, B.A., Greenberg, N.M. (2003) Differential expression of specific FGF ligand and receptor isoforms during angiogenesis associated with prostate cancer progression, *Prostate* **54**(1), 8–16.

Irvine, R.A., Yu, M.C., Ross, R.K., Coetzee, G.A. (1995) The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer, *Cancer Res.* **55**, 1937–1940.

Isaacs, W.B., Bova, G.S., Morton, R.A., et al. (1994) Molecular biology of prostate cancer, *Semin. Oncol.* **21**, 514–521.

Ittmann, M.M., Wieczorek, R. (1996) Alterations of the retinoblastoma gene in clinically localized, stage B prostate adenocarcinomas, *Hum. Pathol.* **27**, 28–34.

Ivanovic, V., Melman, A., David-Joseph, B., et al. (1995) Elevated plasma levels of TGF-α1 in patients with invasive prostate cancer, *Nat. Med.* **1**, 282–284.

Jarrard, D.F., Bova, G.S., Ewing, C.M., et al. (1997) Deletional, mutational, and methylation analyses of CDKN2 (p16/MTS1) in primary metastatic prostate cancer, *Genes Chromosomes Cancer* **19**, 90–96.

Jarrard, D.F., Bussemakers, M.J., Bova, G.S., Isaacs, W.B. (1995) Regional loss of imprinting of the insulin-like growth factor II gene occurs in human prostate tissues, *Clin. Cancer Res.* **1**, 1471–1478.

Jenster, G. (1999) The role of the androgen receptor in the development and progression of prostate cancer, *Semin. Oncol.* **26**, 407–421.

Joos, S., Bergerheim, U.S.R., Pan, Y., et al. (1995) Mapping of chromosomal gains and losses in prostate cancer by comparative genomic hybridization, *Genes Chromosomes Cancer* **14**, 267–276.

Kamradt, J., Drosse, C., Kalkbrenner, S., Rohde, V., Lensch, R., Lehmann, J., Fixemer, T., Bonkhoff, H., Stoeckle, M., Wullich, B. (2003) Telomerase activity and telomerase subunit gene expression levels are not related in prostate cancer: a real-time quantification and

in situ hybridization study, *Lab. Invest.* **83**(5), 623–633.

Kantoff, P.W., Febbo, P.G., Giovannucci, E., et al. (1997) A polymorphism of the 5 alpha-reductase gene and its association with prostate cancer: a case-control analysis, *Cancer Epidemiol. Biomarkers Prev.* **6**, 189–192.

Kim, I.Y., Ahn, H.J., Zelner, D.J., et al. (1996) Loss of expression of transforming growth factor beta type I and type II receptors correlates with tumor grade in human prostate cancer tissues, *Clin. Cancer Res.* **2**, 1255–1261.

Kim, I.Y., Kim, J.H., Zelner, D.J., et al. (1996) Transforming growth factor-beta-1 is a mediator of androgen-regulated growth arrest in an androgen-responsive prostatic cancer cell line, LNCaP, *Endocrinology* **137**, 991–999.

Kim, I.Y., Zelner, D.J., Sensibar, J.A., et al. (1996) Modulation of sensitivity to transforming growth factor beta-1 (TGF-beta-1) and the level of type II TGF-beta receptor in LNCaP cells by dihydrotestosterone, *Exp. Cell Res.* **222**, 103–110.

Kito, H., Suzuki, H., Ichikawa, T., Sekita, N., Kamiya, N., Akakura, K., Igarashi, T., Nakayama, T., Watanabe, M., Harigaya, K., Ito, H. (2001) Hypermethylation of the CD44 gene is associated with progression and metastasis of human prostate cancer, *Prostate* **49**(2), 110–115.

Kleinerman, D.I., Troncoso, P., Lin, S.H., et al. (1995) Consistent expression of an epithelial cell adhesion molecule (C-CAM) during human prostate development and loss of expression in prostate cancer: implication as a tumor suppressor, *Cancer Res.* **5**, 1215–1220.

Koivisto, P., Kononen, J., Palmberg, C., et al. (1997) Androgen receptor gene amplification: a possible molecular mechanism for androgen deprivation therapy failure in prostate cancer, *Cancer Res.* **57**, 314–319.

Konishi, N., Hiasa, Y., Tsuzuki, T., et al. (1996) Detection of Rb, p16/cdkn2 and p15 (Ink4b) gene alterations with immunohistochemical studied in human prostate carcinomas, *Int. J. Oncol.* **8**, 107–112.

Konishi, N., Hiasa, Y., Tsuzuki, T., et al. (1997) Comparison of ras activation in prostate carcinoma in Japanese and American men, *Prostate* **30**, 53–7.

Kubota, Y., Fujinami, K., Uemura, H., et al. (1995) Retinoblastoma gene mutations in primary human prostate cancer, *Prostate* **27**, 314–320.

Kuhn, E.J., Kurnot, R.A., Sesterhenn, I.A., et al. (1993) Expression of the c–erbB-2 (HER-2/neu) oncoprotein in human prostatic carcinoma, *J. Urol.* **150**, 1427–1433.

Kunimi, K., Amano, T., Uchibayashi, T. (1996) Point mutation of the p53 gene is an infrequent event in untreated prostate cancer, *Cancer Detect. Prev.* **20**, 218–222.

Kunimi, K., Bergerheim, U., Larsson, I.L., et al. (1991) Allelotyping of human prostatic adenocarcinoma, *Genomics* **11**, 530–536.

Kyprianou, N. (1999) Activation of TGF-beta signalling in human prostate cancer cells suppresses tumorigenicity via deregulation of cell cycle progression and induction of caspase-1 mediated apoptosis: significance in prostate tumorigenesis, *Prostate Cancer Prostat. Dis.* **2**(S3), S18.

Lacombe, L., Orlow, I., Reuter, V.E., et al. (1996) Microsatellite instability and deletion analysis of chromosome 10 in human cancer, *Int. J. Cancer* **69**, 110–113.

Lange, E.M., Chen, H., Brierley, K., et al. (1999) Linkage analysis of 153 prostate cancer families over a 30-cM region containing the putative susceptibility locus HPCX, *Clin. Cancer Res.* **5**, 4013–4020.

LaSpada, A.R., Wilson, E.M., Lubahn, D.B., et al. (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy, *Nature* **352**, 77–79.

Latil, A., Baron, J.-C., Cussenot, O., et al. (1994) Genetic alterations in localized prostate cancer identification of a common region of deletion on chromosome arm 18q, *Genes Chromosomes Cancer* **11**, 119–125.

Latil, A., Baron, J.-C., Cussenot, O., et al. (1994) Oncogene amplifications in early-stage human prostate carcinomas, *Int. J. Cancer* **59**, 637–638.

Latil, A., Cussenot, O., Fournier, G., Lidereau, R. (1997) Infrequent allelic imbalance at the major susceptibility HPC1 locus in sporadic prostate tumors, *Int. J. Cancer* **71**, 1118.

Latil, A., Lidereau, R. (1998) Genetic aspects of prostate cancer, *Virchows Arch.* **432**, 389–406.

Leach, F.S., Nicolaides, N.C., Papadopoulos, N., et al. (1993) Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer, *Cell* **75**, 1215–1225.

Lee, J.M., Bernstein, A. (1993) p53 mutations increase resistance to ionizing radiation, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5742–5746.

Lee, W.H., Isaacs, W.B., Bova, G.S., Nelson, W.G. (1997) CG island methylation changes near the GSTP1 gene in prostatic carcinoma cells detected using the polymerase chain reaction: a new prostate cancer biomarker, *Cancer Epidemiol. Biomarkers Prev.* **6**, 443–450.

Lee, W.H., Morton, R.A., Epstein, J.I., et al. (1994) Cytidine methylation of regulatory sequences near the pi-class glutathione Stransferase gene accompanies human prostatic carcinogenesis, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11733–11737.

Leung, H.Y., Dickson, C., Robson, C.N., Neal, D.E. (1996) Over-expression of fibroblast growth factor-8 in human prostate cancer, *Oncogene* **12**, 1833–1835.

Leung, H.Y., Mehta, P., Gray, L.B., et al. (1997) Keratinocyte growth factor expression in hormone insensitive prostate cancer, *Oncogene* **15**, 1115–1120.

Li, J., Yen, C., Liaw, D., et al. (1997) PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer, *Science* **275**, 1943–1947.

Limon, J., Lundgren, R., Elfving, P., et al. (1989) Double minutes in two primary adenocarcinomas of prostate, *Cancer Genet. Cytogenet.* **39**, 191–194.

Lin, Y., Uemura, H., Fujinami, K., et al. (1997) Telomerase activity in primary prostate cancer, *J. Urol.* **157**, 1161–1165.

Loeb, L.A. (1994) Microsatellite instability: marker of a mutator phenotype in cancer, *Cancer Res.* **54**, 5059–5063.

Lokeshwar, B.L., Lokeshwar, V.B., Block, N.L. (1995) Expression of CD44 in prostate cancer cells: association with cell proliferation and invasive potential, *Anticancer Res.* **15**, 1191–1198.

Lou, W., Krill, D., Dhir, R., et al. (1999) Methylation of the CD44 metastasis suppressor gene in human prostate cancer, *Cancer Res.* **59**, 2329–2331.

Lowe, S.W., Schmitt, E.M., Smith, S.W., et al. (1993) p53 is required for radiation-induced apoptosis in mouse thymocytes, *Nature* **362**, 847–849.

Lundgren, R., Manhad, N., Heim, S., et al. (1992) Cytogenetic analysis of 57 primary prostatic adenocarcinomas, *Genes Chromosomes Cancer* **4**, 16–24.

MacGrogan, D., Levy, A., Bova, G.S., et al. (1996) Structure and methylation-associated silencing of a gene within a homozygously deleted region of human chromosome band 8p22, *Genomics* **35**, 55–65.

Macoska, J.A., Trybus, T.M., Benson, P.D., et al. (1995) Evidence for three tumor suppressor gene loci on chromosome 8p in human prostate cancer, *Cancer Res.* **55**, 5390–5395.

Macoska, J.A., Trybus, T.M., Sakr, W.A., et al. (1994) Fluorescence in situ hybridization analysis of 8p allelic loss and chromosome 8 instability in human prostate cancer, *Cancer Res.* **54**, 3824–3830.

Makridakis, N.M., Ross, R.K., Pike, M.C., et al. (1999) Association of mis-sense substitution in *SRD5A2* gene with prostate cancer in African-American and Hispanic men in Los Angeles, USA, *Lancet* **354**, 975–978.

Mansson, P.E., Adams, P., Kan, M., McKeehan, W.L. (1989) Heparinbinding growth factor gene expression and receptor characteristics in normal rat prostate and two transplantable rate prostate tumors, *Cancer Res.* **49**, 2485–2494.

Matusik, R.J., Fleming, W.H., Hamel, A., et al. (1987) Expression of the c-myc proto-oncogene in prostatic tissue, *Prog. Clin. Biol. Res.* **239**, 91–112.

McIndoe, R.A., Stanford, J.L., Gibbs, M., et al. (1997) Linkage analysis of 49 high-risk families does not support a common familial prostate cancer-susceptibility gene at 1q24-25, *Am. J. Hum. Genet.* **61**, 347–353.

McMenamin, M.E., Soung, P., Perera, S., et al. (1999) Loss of PTEN expression in paraffin-embedded primary prostate cancer correlates with high Gleason score and advanced stage, *Cancer Res.* **59**, 4291–4296.

McPherson, J.D., Morton, R.A., Ewing, C.M., et al. (1994) Assignment of the human α-catenin gene (CTNNA1) to chromosome 5q21-q22, *Genomics* **19**, 188–190.

Mellinghoff, I.K., Tran, C., Sawyers, C.L. (2002) Growth inhibitory effects of the dual ErbB1/ErbB2 tyrosine kinase inhibitor PKI-166 on human prostate cancer xenografts, *Cancer Res.* **62**(18), 5254–5259.

Mellon, K., Thompson, S., Charlton, R.G., et al. (1992) p53, c-erbB-2, and the epidermal growth factor receptor in the benign and malignant prostate, *J. Urol.* **147**, 496–499.

Micale, M.A., Mohamed, A., Sakr, W., et al. (1992) Cytogenetics of primary prostatic adenocarcinoma. Clonality and chromosome

instability, *Cancer Genet. Cytogenet.* **61**, 165–173.

Miet, S.M., Neyra, M., Jaques, R., et al. (1999) RER(+) phenotype in prostate intra-epithelial neoplasia associated with human prostate-carcinoma development, *Int. J. Cancer* **82**, 635–639.

Millar, D.S., Ow, K.K., Paul, C.L., et al. (1999) Detailed methylation analysis of the glutathione S-transferase pi (GSTP1) gene in prostate cancer, *Oncogene* **18**, 1313–1324.

Monroe, K.R., Yu, M.C., Kolonel, M.N., et al. (1995) Evidence of an X linked or recessive genetic component to prostate cancer risk, *Nat. Med.* **1**, 827–829.

Morin, G.B. (1989) The human telomere transferase is a ribonucleoprotein that synthesizes TTAGGG repeats, *Cell* **59**, 521–529.

Morton, D.M., Barrack, E.R. (1995) Modulation of transforming growth factor beta 1 effects on prostate cancer cell proliferation by growth factors and extracellular matrix, *Cancer Res.* **55**, 2596–2602.

Morton, R.A., Ewing, C.M., Watkins, J.J., Isaacs, W.B. (1995) The E-cadherin cell-cell adhesion pathway in urologic malignancies, *World J. Urol.* **13**, 364–368.

Morton, R.A., Watkins, J.J., Bova, G.S., et al. (1996) Hypermethylation of chromosome 17P locus D17S5 in human prostate tissue, *J. Urol.* **156**, 512–516.

Moul, J.W., Friedrichs, P.A., Lance, R.S., et al. (1992) Infrequent Ras oncogene mutations in human prostate cancer, *Prostate* **20**, 327–338.

Multani, A.S., Ozen, M., Agrawal, A., et al. (1999) Amplification of the Y chromsome in three murine tumor cell lines transformed *in vivo* by different human prostate cancers, *In Vitro Cell Dev. Biol. Anim.* **35**, 236–239.

Nag, A., Smith, R.G. (1989) Amplification, rearrangement, and elevated expression of c-myc in the human prostatic carcinoma cell line LNCaP, *Prostate* **15**, 115–122.

Nagle, R.B., Hao, J.S., Knox, J.D., et al. (1995) Expression of hemidesmosomal and extracellular matrix proteins by normal and malignant human prostate tissue, *Am. J. Pathol.* **146**, 1498–1507.

Narod, S.A., Dupont, A., Cusan, L., et al. (1995) The impact of family history on early detection of prostate cancer, *Nat. Med.* **1**, 99–101.

Natali, P.G., Nicotra, M.R., Bigotti, A., et al. (1990) Expression of the p185 encoded by HER2 oncogene in normal and transformed human tissues, *Int. J. Cancer* **45**, 457–461.

Nelson, J.B., Lee, W.H., Nguyen, S.H., et al. (1997) Methylation of the 5′ CpG island of the endothelin B receptor gene is common in human prostate cancer, *Cancer Res.* **57**, 35–37.

Noordzij, M.A., vanSteenbrugge, G.-J., Verkaik, N.S., et al. (1997) The prognostic value of CD44 isoforms in prostate cancer patients treated by radical prostatectomy, *Clin. Cancer Res.* **3**, 805–815.

O'Connor, P.M., Jackman, J., Jondle, D., et al. (1993) Role of the p53 tumor suppressor gene in cell cycle arrest and radiosensitivity of Burkitt's lymphoma cell lines, *Cancer Res.* **53**, 4776–4780.

Paradis, V., Dargere, D., Laurendeau, I., et al. (1999) Expression of the RNA component of human telomerase (hTR) in prostate cancer, prostatic intraepithelial neoplasia, and normal prostate tissue., *J. Pathol.* **189**, 213–218.

Parsons, R., Li, G.M., Longley, M.J., et al. (1993) Hypermutability and mismatch repair deficiency in RER+ tumor cells, *Cell* **75**, 1227–1236.

Platz, E.A., Giovannucci, E., Dahl, D.M., et al. (1998) The androgen receptor gene GGN microsatellite and prostate cancer risk, *Cancer Epidemiol. Biomarkers Prev.* **7**, 379–384.

Prendergast, N.J., Atkins, M.R., Schatte, E.C., et al. (1996) p53 immunohistochemical and genetic alterations are associated at high incidence with post-irradiated locally persistent prostate carcinoma, *J. Urol.* **155**, 1685–1692.

Prendergast, N.J., Walther, P.J. (1995) Genetic alterations in prostate adenocarcinoma, *Surg. Oncol. Clin. North Am.* **4**, 241–255.

Qian, J., Bostwick, D.G., Takahashi, S., et al. (1995) Chromosomal anomalies in prostatic intraepithelial neoplasia and carcinoma detected by fluorescence in situ hybridization, *Cancer Res.* **55**, 5408–5414.

Rabinovitz, I., Nagle, R.B., Cress, A.E. (1996) Integrin alpha 6 expression in human prostate carcinoma cells is associated with a migratory and invasive phenotype *in vitro* and *in vivo*, *Clin. Exp. Metastasis* **13**, 481–491.

Rahman, M.M., Miyamoto, H., Lardy, H., Chang, C. (2003) Inactivation of androgen receptor coregulator ARA55 inhibits androgen receptor activity and agonist effect of antiandrogens in prostate cancer cells, *Proc. Natl. Acad. Sci. U.S.A.* **100**(9), 5124–5129.

Reichardt, J.K., Makridakis, N., Henderson, B.E., et al. (1995) Genetic variability of the human SRD5A2 gene: implications for prostate cancer risk, *Cancer Res.* **55**, 3973–3975.

Richmond, P.J., Karayiannakis, A.J., Nagafuchi, A., et al. (1997) Aberrant Ecadherin and α-catenin expression in prostate cancer: correlation with patient survival, *Cancer Res.* **57**, 3189–3193.

Rokman, A., Ikonen, T., Seppala, E.H., Nupponen, N., Autio, V., Mononen, N., Bailey-Wilson, J., Trent, J., Carpten, J., Matikainen, M.P., Koivisto, P.A., Tammela, T.L., Kallioniemi, O.P., Schleutker, J. (2002) Germline alterations of the RNASEL gene, a candidate HPC1 gene at 1q25, in patients and families with prostate cancer, *Am. J. Hum. Genet.* **70**(5), 1299–1304.

Rosini, P., Bonaccorsi, L., Baldi, E., Chiasserini, C., Forti, G., De Chiara, G., Lucibello, M., Mongiat, M., Iozzo, R.V., Garaci, E., Cozzolino, F., Torcia, M.G. (2002) Androgen receptor expression induces FGF2, FGF-binding protein production, and FGF2 release in prostate carcinoma cells: role of FGF2 in growth, survival, and androgen receptor down-modulation, *Prostate* **53**(4), 310–321.

Ross, J.S., Sheehan, C., Hayner-Buchan, A.M., et al. (1997) HER-2/neu gene amplification status in prostate cancer by fluorescence in situ hybridization, *Hum. Pathol.* **28**, 827–833.

Ruijter, E., van de Kaa, C., Miller, G., et al. (1999) Molecular genetics and epidemiology of prostate cancer, *Endocr. Rev.* **20**, 22–45.

Ruijter, E.T., Miller, G.J., Aalders, T.W., et al., Biomed-II MPC Study Group. (1997) Rapid microwave stimulated fixation of entire prostatectomy specimens, *J. Pathol.* **183**, 369–375.

Sakr, W.A., Macoska, J.A., Benson, P., et al. (1994) Allelic loss in locally metastatic, multisampled prostate cancer, *Cancer Res.* **54**, 3273–3277.

Santourlidis, S., Florl, A., Ackermann, R., et al. (1999) High frequency of alterations in DNA methylation in adenocarcinoma of the prostate, *Prostate* **39**, 166–174.

Schaid, D.J., McDonnell, S.K., Blute, M.L., Thibodeau, S.N. (1998) Evidence of autosomal dominant inheritance of prostate cancer, *Am. J. Hum. Genet.* **62**, 1425–1438.

Schalken, J.A., Smit, F., Bringuier, P.P., et al. (1995) The role of E-cadherin mutations in bladder and prostate cancer: is E-cadherin a classical tumor suppressor gene? *Urol. Res.* **23**, P57.

Serrano, M., Hannon, G.J., Beach, D. (1993) A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4, *Nature* **366**, 704–707.

Smith, J.R., Freije, D., Carpten, J.D., et al. (1996) Major susceptibility locus for prostate cancer on chromosome 1 suggested by genome-wide search, *Science* **274**, 1371–1374.

Smith, J.R., Freije, D., Carpten, J.D., Gronberg, H., Xu, J., Isaacs, S.D., Brownstein, M.J., Bova, G.S., Guo, H., Bujnovszky, P., Nusskern, D.R., Damber, J.E., Bergh, A., Emanuelsson, M., Kallioniemi, O.P., Walker-Daniels, J., Bailey-Wilson, J.E., Beaty, T.H., Meyers, D.A., Walsh, P.C., Collins, F.S., Trent, J.M., Isaacs, W.B. (1996) Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search, *Science* **274**(5291), 1371–1374.

Sommerfeld, H.-J., Meeker, A.K., Piatyszek, M.A., et al. (1996) Telomerase activity: a prevalent marker of malignant human prostate tissue, *Cancer Res.* **56**, 218–222.

Stanford, J.L., Just, J.J., Gibbs, M., et al. (1997) Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk, *Cancer Res.* **57**, 1194–1198.

Stapleton, A.M.F., Timme, T.L., Gousse, E.A., et al. (1997) Primary human prostate cancer cells harboring p53 mutations are clonally expanded in metastases, *Clin. Cancer Res.* **3**, 1389–1397.

Steiner, M.S. (1995) Review of peptide growth factors in benign prostatic hyperplasia and urological malignancy, *J. Urol.* **153**, 1085–1096.

Suzuki, H., Freije, D., Nusskern, D.R., et al. (1998) Interfocal heterogeneity of PTEN/MMAC1 gene alterations in multiple metastatic prostate cancer tissues, *Cancer Res.* **58**, 204–209.

Suzuki, H., Komiya, A., Aida, S., et al. (1995) Microsatellite instability and other molecular abnormalities in human prostate cancer, *Jpn. J. Cancer Res.* **86**, 956–961.

Suzuki, H., Komiya, A., Aida, S., et al. (1995) Microsatellite instability and other molecular abnormalities in human prostate cancer, *Jpn. J. Cancer Res.* **86**, 956–961.

Suzuki, H., Komiya, A., Emi, M., et al. (1996) Three distinct commonly deleted regions of

chromosome arm 16q in human primary and metastatic prostate cancers, *Genes Chromosomes Cancer* **17**, 225–233.

Takahashi, S., Alcaraz, A., Brown, J.A., et al. (1996) Aneusomies of chromosomes 8 and Y detected by fluorescence in situ hybridization are prognostic markers for pathological stage C (pT3N0M0) prostate carcinoma, *Clin. Cancer Res.* **2**, 137–145.

Takahashi, H., Furusato, M., Allsbrook, W.C. Jr., et al. (1995) Prevalence of androgen receptor gene mutations in latent prostatic carcinomas from Japanese men, *Cancer Res.* **55**, 1621–1624.

Takahashi, S., Qian, J., Brown, J.A., et al. (1994) Potential markers of prostate cancer aggressiveness detected by fluorescence in situ hybridization in needle biopsies, *Cancer Res.* **54**, 3574–3579.

Takeichi, M. (1991) Cadherin cell adhesion receptors as a morphogenetic regulator, *Science* **251**, 1451–1455.

Taplin, M.-E., Bubley, G.J., Shuster, T.D., et al. (1995) Mutation of the androgen-receptor gene in metastatic androgen-independent prostate cancer, *N. Engl. J. Med.* **21**, 1393–1398.

Taylor, S.M. (1993) 5-Aza-2′-deoxycytidine: cell differentiation and DNA methylation, *Leukemia* **7**, 3–8.

Terrell, R.B., Willie, A.H., Cheville, J.C., et al. (1995) Microsatellite instability in adenocarcinoma of the prostate, *Am. J. Pathol.* **147**, 799–805.

Thompson, T.C. (1990) Growth factors and oncogenes in prostate cancer, *Cancer Cells* **2**, 345–354.

Tozawa, K., Akita, H., Kawai, N., Okamura, T., Sasaki, S., Hayashi, Y., Kohri, K. (2001) KAI1 expression can be a predictor of stage A prostate cancer progression, *Prostate Cancer Prostat. Dis.* **4**(3), 150–153.

Trapman, J., Sleddens, H.F., van der Weiden, M.M., et al. (1994) Loss of heterozygosity of chromosome 8 microsatellite loci implicates a candidate tumor suppressor gene between loci D8S87 and D8S133 in human prostate cancer, *Cancer Res.* **54**, 6061–6064.

Tricoli, J.V. (1999) Y chromosome enumeration in touch preparations from 42 prostate tumors by interphase fluorescence in situ hybridization analysis, *Cancer Genet. Cytogenet.* **111**, 1–6.

Turkeri, L.N., Sakr, W.A., Wykes, S.M., et al. (1994) Comparative analysis of epidermal growth factor receptor gene expression and protein product in benign, premalignant, and malignant prostate tissue, *Prostate* **25**, 199–205.

Uchida, T., Wada, C., Wang, C., et al. (1995) Microsatellite instability in prostate cancer, *Oncogene* **10**, 1019–1022.

Ueda, T., Komiya, A., Emi, M., et al. (1997) Allelic losses on 18q21 are associated with progression and metastasis in human prostate cancer, *Genes Chromosomes Cancer* **20**, 140–147.

Umbas, R., Isaacs, W.B., Bringuier, P.P., et al. (1994) Decreased E-cadherin expression is associated with poor prognosis in patients with prostate cancer, *Cancer Res.* **54**, 3929–3933.

Umbas, R., Isaacs, W.B., Bringuier, P.P., et al. (1997) Relation between aberrant b-catenin expression and loss of E-cadherin function in prostate cancer, *Int. J. Cancer* **74**, 374–377.

VanDenBerg, C., Guan, X.-Y., VonHoff, D., et al. (1995) DNA sequence amplification in human prostate cancer identified by chromosome microdissection: potential prognostic implication, *Clin Cancer Res.* **1**, 11–18.

Vesalainen, S., Lipponen, P. (1995) Expression of retinoblastoma gene (Rb) protein in T1,2M0 prostatic adenocarcinoma, *J. Cancer Res. Clin. Oncol.* **121**, 429–433.

Vicentini, C., Festuccia, C., Gravina, G.L., Angelucci, A., Marronaro, A., Bologna, M. (2003) Prostate cancer cell proliferation is strongly reduced by the epidermal growth factor receptor tyrosine kinase inhibitor ZD1839 *in vitro* on human cell lines and primary cultures, *J. Cancer Res. Clin. Oncol.* **129**(3), 165–174.

Viola, M.V., Fromovitz, F., Oravez, S., et al. (1986) Expression of ras oncogene p21 in prostate cancer, *N. Engl. J. Med.* **314**, 133–137.

Visakorpi, T., Hyytinen, E.R., Koivisto, P., et al. (1995) *In vivo* amplification of the androgen receptor gene and progression of human prostate cancer, *Nat. Genet.* **9**, 401–406.

Visakorpi, T., Kallioniemi, A.H., Syvänen, A.-C., et al. (1995) Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization, *Cancer Res.* **55**, 342–347.

Vleminckx, K., Vakaet, L. Jr., Mareel, M., et al. (1991) Genetic manipulation of E-cadherin expression by epithelial tumor cells reveals an invasion suppressor role, *Cell* **66**, 107–119.

Voeller, H.J., Truica, C.I., Gelmann, E.P. (1998) α-catenin mutations in human prostate cancer, *Cancer Res.* **58**, 2520–2523.

Wang, S.I., Parsons, R., Ittmann, M. (1998) Homozygous deletions of the PTEN tumor suppressor gene in a subset of prostate adenocarcinomas, *Clin. Cancer Res.* **4**, 811–815.

Watanabe, M., Imai, H., Kato, H., et al. (1996) Microsatellite instability in latent prostate cancers, *Int. J. Cancer* **69**, 394–397.

Wertz, I.E., Deitch, A.D., Gumerlock, P.H., et al. (1996) Correlation of genetic and immunodetection of TP53 mutations in malignant and benign prostate tissues, *Hum. Pathol.* **27**, 573–580.

Whittemore, A.S. (1994) Prostate cancer, *Cancer Surv.* **19**, 309–322.

Williams, R.H., Stapleton, A.M.F., Yang, G., et al. (1996) Reduced levels of transforming growth factor beta receptor type II in human prostate cancer: an immunohistochemical study, *Clin. Cancer Res.* **2**, 635–640.

Xu, J., Meyers, D., Freije, D., et al. (1998) Evidence for a prostate cancer susceptibility locus on the X chromosome, *Nat. Genet.* **20**, 175–179.

Yan, G., Fukabori, Y., Nikolaropoulos, S., et al. (1992) Heparin-binding keratinocyte growth factor is a candidate stromal to epithelial andromedin, *Mol. Endocrinol.* **6**, 2123–2128.

Yeh, S., Lin, H.K., Kang, H.Y., et al. (1999) From HER2/neu signal cascade to androgen receptor and its coactivators: a novel pathway by induction of androgen target genes through MAP kinase in prostate cancer cells, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5458–5463.

Zhang, W., Kapusta, L.R., Slingerland, J.M., Klotz, L.H. (1998) Telomerase activity in prostate cancer, prostatic intraepithelial neoplasia, and benign prostatic epithelium, *Cancer Res.* **58**, 619–621.

Zhang, X.A., Lane, W.S., Charrin, S., Rubinstein, E., Liu, L. (2003) EWI2/PGRL associates with the metastasis suppressor KAI1/CD82 and inhibits the migration of prostate cancer cells, *Cancer Res.* **63**(10), 2665–2674.

Zhang, X.H., Sakamoto, H., Takenaka, I. (1996) Accumulation of p53 and expression of CD44 in human prostatic cancer and benign prostatic hyperplasia: an immunohistochemical study, *Br. J. Urol.* **77**, 441–444.

Zheng, D.Q., Woodard, A.S., Fornaro, M., et al. (1999) Prostatic carcinoma cell migration via alpha(v)beta3 integrin is modulated by a focal adhesion kinase pathway, *Cancer Res.* **59**, 1655–1664.

Cancer Stem Cells

Michael F. Clarke[1] *and Andrew T. Hass*[2]
[1] *Department of Hematology and Oncology, The University of Michigan, Ann Arbor, Michigan, USA*
[2] *Cellular and Molecular Biology Program, The University of Michigan, Ann Arbor, Michigan, USA*

1 Introduction 223

2 Evidence for Cancer Stem Cells 223
2.1 Viewing a Tumor as an Abnormal Organ 223
2.2 Recent Evidence for AML and Breast Cancer Stem Cells 224

3 Properties of Normal Stem Cells 226
3.1 The Hematopoietic Stem Cell and its Properties 226
3.2 The Hierarchy of the Hematopoietic System 227
3.3 Self-renewal: The Defining Property of Stem Cells 227

4 Genetic Regulation of Self-renewal in Normal Stem Cells and Cancer Cells 228
4.1 Tissue and Tumor Growth: A Balancing Act 228
4.2 Multiple Oncogenes Implicated in Self-renewal 228
4.3 Possibility of a Common Self-renewal Pathway among Multiple Cell Types 230

5 Target Cells for Malignant Transformation 230
5.1 Stem Cells as Targets for Malignant Transformation 230
5.2 Progenitor Cells as Targets for Malignant Transformation 231
5.3 Targets for Malignant Transformation in Hematopoietic Malignancies 232

6 Implications of Cancer Stem Cells for the Diagnosis and Treatment of Cancer 233
6.1 Cancer Stem Cells Would Explain Several Conundrums of Cancer Biology 233

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

6.2 Tumor Relapse Might Result from a Failure to Target Cancer Stem Cells 233
6.3 Cancer Stem Cells Might Be More Resistant to Chemotherapy 234
6.4 Identification of Cancer Stem Cells Might Lead to Better Drug Selection 234

7 Future Implications of Cancer Stem Cells 235

Bibliography 235
Books and Reviews 235
Primary Literature 236

Keywords

Cancer Stem Cell
A cell within a tumor that possesses the ability to self-renew and is responsible for maintaining the growth of the tumor; this cell population is the only cell population within a tumor that has the ability to form a new tumor and can produce progeny of multiple phenotypes.

Self-renewal
The process by which, upon cell division, at least one of the daughter cells retains the multipotent capacity of the parent cell.

Tumorigenic Cell
A tumor cell capable of self-renewal and of forming a new tumor.

Researchers have made stunning advances over the last fifty years in characterizing the oncogenic mutations that lead to cancer. However, we are just beginning to understand the complex consequences of these mutations and how they result in neoplastic transformation at the cellular level. Most cancers arise in tissues that contain a stem cell population, such as gut, breast, prostate, lung, and bone marrow. Stem cells are defined by three critical properties: the ability to divide and give rise to a new stem cell (self-renewal), the ability to give rise to the differentiated cells of an organ, and the genetic constraints on expansion of the cell population. It is known that cancer results from the accumulation of mutations in a single target cell, sometimes over a period of many years. Since stem cells are the only long-lived cells in most tissues, this suggests that they are the cells in which early mutations leading to cancer accumulate. Additionally, stem cells are the only normal cells that share cancer cells' intrinsic ability to self-renew. Taken together, these two facts suggest that, in some cases, stem cells may be the target cells for neoplastic transformation.

Presently, therapeutic targets are selected on the basis of the assumption that all of the cancer cells within a particular tumor are equivalent in that they are all capable of driving tumorigenesis and metastasis. New data suggests that, in at least some cancers, a small, phenotypically distinct subset of the cancer cells (tumorigenic cells, or cancer stem cells) has the exclusive ability to form tumors. The rest of the cells, which form the bulk of the tumor, are unable to self-renew or sustain tumorigenesis. Tumorigenic and nontumorigenic cancer cells can be identified on the basis of cell surface-marker expression, utilizing the same tools used in normal stem cell biology. This suggests that therapies could be made more effective by targeting the tumorigenic cell population. The ability to identify tumorigenic cancer cells should therefore allow the identification of new diagnostic markers and therapeutic targets, better evaluation of candidate therapeutics, and more accurate determination of therapeutic efficacy.

1 Introduction

Most tissues in which common cancers arise contain a subpopulation of proliferating cells that are responsible for replenishing the short-lived mature cells as they undergo apoptosis or are otherwise damaged. Cell maturation is arranged in a hierarchy in which this rare subpopulation of stem cells give rise to progenitor cells and eventually mature cells; the stem cell pool replenishes itself through a process called *self-renewal*.

Recent discoveries indicated that it is likely that the tumors contain a minority "cancer stem cell" population with indefinite proliferative potential that drives the growth and metastasis of tumors. Furthermore, emerging evidence suggests that similar molecular mechanisms regulate self-renewal in normal stem cells and their malignant counterparts. If this is true, a more complete understanding of the cellular biology of the tissues in which cancers arise, and specifically of the stem cells residing in those tissues, could provide new insights into cancer biology. Certain aspects of stem cell biology apply to cancer. First, both normal stem cells and cancer stem cells undergo self-renewal – a property that separates them from the bulk of normal tissue cells and the bulk of tumor cells, respectively. Additionally, it is quite likely that mutations that lead to cancer accumulate in normal stem cells since they are the only long-lived cells in many tissues.

2 Evidence for Cancer Stem Cells

2.1 Viewing a Tumor as an Abnormal Organ

Cancers consist of phenotypically heterogeneous populations of cancer cells. These phenotypically distinct cell populations could arise, in part, from sequential mutations occurring because of either genetic instability and/or environmental factors. Another possible source of this heterogeneity becomes apparent if one views a tumor as an aberrant organ containing a tumorigenic (stem cell) population that

drives tumor growth. These tumorigenic cells would have acquired oncogenic mutations that result in deregulated self-renewal, but their phenotypically diverse progeny would still be differentiated and lack the ability to self-renew. Several lines of evidence suggest that this model accounts for some of the cellular heterogeneity seen in tumors, although genetic instability and environmental factors undoubtedly contribute to the variability in phenotypes of cancer cells in a particular tumor. It is well documented that many types of cancer contain heterogeneous populations of cells, which variably express differentiation markers that reflect the tissue or origin, as well as cancer cells that have an immature morphology. Examples of this observation include the variable expression of milk proteins by some breast cancers and the variable expression of myeloid and/or lymphoid markers in chronic myelogeous leukemia (CML) and acute myelogenous leukemia (AML). By contrast, in some tumors, specifically in mature germ cell teratomas, only a minority of the cancer cells express immature cell markers such as α-fetoprotein. Because the terminally differentiated cells that form the teeth and hair in germ cell teratomas are unlikely to be able to proliferate and form new tumors, these data are consistent with the notion that the minority population of α-fetoprotein expressing cancer cells might have the exclusive ability to form new tumors consisting of more tumorigenic cells as well as the phenotypically diverse populations of abnormally differentiated cells lacking the ability to self-renew. If this proved to be true, these cells could be considered to be cancer stem cells.

The evidence suggesting that cancer is a disease of aberrant differentiation, proliferation, and self-renewal makes it plausible to view tumors as abnormal organs; in this context, the principles of stem cell biology can be applied to better understand the biology of these diseases. It has been shown in both hematopoietic malignancies and solid cancers that only a subset of cancer cells are clonogenic when placed in tissue culture or injected into immunodeficient mice. For example, only 0.01 to 1% of the mouse myeloma cells formed colonies in *in vitro* colony-forming assays. Similarly, only 1 to 4% of leukemic cells formed spleen colonies when transplanted into mice. This trend was demonstrated in solid cancers as well, with only 0.02 to 0.1% of ovarian cancer or lung cancer cells forming colonies in soft agar. Since only a minority of normal bone marrow cells were clonogenic, the clonogenic cancer cells were initially described as cancer stem cells; this implied that only a distinct population of cancer cells were able to proliferate extensively in these assays. However, an alternative explanation was that all cancer cells have the intrinsic ability to proliferate extensively, but only a small minority of cells do so in a particular assay.

2.2 Recent Evidence for AML and Breast Cancer Stem Cells

Proving that a phenotypically distinct population of cancer cells is responsible for perpetuating the disease would require isolating different populations of cancer cells and demonstrating that one or more groups were enriched for the ability to initiate the disease while other populations lacked this ability. This was first shown in AML when it was shown that a leukemia tumor-initiating subpopulation of cells could be prospectively identified and purified from multiple patients' bone marrows. In most cases of AML,

the minority population of $CD34^{+}CD38^{-}$ cells was the only group of cells capable of establishing human AML in the bone marrow of Nonobese Diabetic Severe Combined Immuno-deficiency (NOD/SCID) mice. Within this population are also $Thy1^{+}$ $CD34^{+}CD38^{-}$ normal hematopoietic stem cells (HSCs). Since leukemia-initiating cells lack the Thy1 expression characteristic of normal HSCs, it is likely that while the early mutations occurred in the HSC, the final transforming mutations occurred in early downstream progenitors lacking Thy1 expression. An alternative explanation is that Thy1 expression could have been lost as a consequence of neoplastic transformation.

Tumorigenic and nontumorigenic subsets of cancer cells have recently been isolated from human breast cancer tumors, providing the first conclusive evidence for cancer stem cells in solid tumors. Using a model for human breast cancer similar to that used for AML, cells were isolated, sorted by surface marker expression using flow cytometry, and grown in immunocompromised mice, as shown in Fig. 1. A minority population of breast cancer cells was identified in these tumors, which was exclusive in its ability to form new tumors. Tumorigenic cells could be distinguished from nontumorigenic cancer cells based upon surface-marker expression that was consistent from patient to patient. In eight out of nine patients, tumorigenic cells could be prospectively identified and isolated as $CD44^{+}CD24^{-/low}Lineage^{-}$ cells. In a subset of these tumors, Epithelial Specific Antigen (ESA) provided further enrichment of the tumorigenic population. Limiting dilution assays

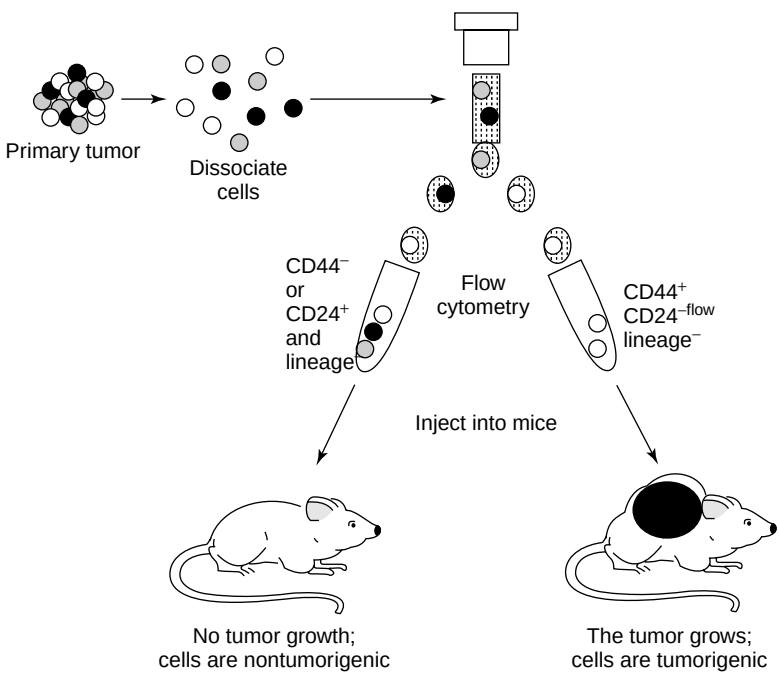

Fig. 1 Identification of breast cancer tumorigenic cells.

demonstrated that as few as one hundred $CD44^{+}CD24^{-/low}Lineage^{-}$ cells were able to form tumors, while tens of thousands of cancer cells from outside this population failed to form tumors in NOD/SCID mice, as shown in Fig. 1. These tumorigenic cells could be serially passaged, and each time cells within this population generated new tumors containing additional $CD44^{+}CD24^{-/low}Lineage^{-}$ tumorigenic cells. Additionally, phenotypically mixed populations of nontumorigenic cancer cells were produced that composed the bulk of the tumor. Importantly, their phenotypic distribution resembled that of the original tumor. These data demonstrate the presence of a hierarchy of cells within a breast cancer tumor in which only a fraction of the cells have the ability to proliferate extensively, while other cells have only a limited proliferative potential, suggesting that the tumorigenic cells can both self-renew as well as differentiate. The phenotype of the tumorigenic breast cancer cells may be similar to normal breast stem or progenitor cells, since early multipotent epithelial progenitor cells have been reported to express both ESA and CD44. While this would be consistent with the similarities observed between normal HSCs and AML tumor-initiating cells, definitive proof will depend upon the characterization of the differentiation hierarchy of breast tissue.

The $CD44^{+}CD24^{-/low}Lineage^{-}$ tumorigenic breast cancer cells and the $CD34^{+}CD38^{-}$ leukemia-initiating cells share with normal stem cells the ability to proliferate extensively, as well as to give rise to cells of diverse phenotypes with reduced developmental or proliferative potential. The extensive proliferative potential of tumorigenic cancer cells contrasts with the bulk of the cancer cells, which lack the ability to form detectable tumors and form smaller colonies *in vitro*. Thus, both tumorigenic breast cancer cells and leukemia-initiating cells from most tumors appear to exhibit the properties of cancer stem cells. However, before these cells can definitively be called cancer stem cells, new assays will be necessary to demonstrate that a single transplanted cell gives rise to all of the diverse populations of cancer cells within a tumor.

Cancer stem cells share many key properties with normal stem cells, including the ability to self-renew. In addition, normal stem cells are likely to be the cells in which oncogenic mutations accumulate and in some cases may be the target cells for neoplastic transformation. Appreciating the implications of the discovery of cancer stem cells for the diagnosis and treatment of cancer consequently depends upon understanding the key properties of stem cells, the genetic regulation of these properties, and how cells can be targeted for neoplastic transformation.

3 Properties of Normal Stem Cells

3.1 The Hematopoietic Stem Cell and its Properties

The accessibility of the hematopoietic system to study has made the HSC the most studied and best understood somatic stem cell population; consequently, it serves as a model for stem cells from other tissues. Hematopoiesis is a tightly regulated process in which a pool of HSCs, through a series of divisions and immature progenitors, eventually gives rise to the mature lymphohematopoietic system consisting of the formed blood

elements, for example, red blood cells, platelets, granulocytes, macrophages, and B- and T-lymphocytes. These cells are responsible for oxygenation, prevention of bleeding, infection prevention, and immunity, respectively. The HSC pool has three fundamental properties in the adult. First, self-renewal allows the maintenance of the stem cell pool. Self-renewal is not analagous to proliferation; self-renewal is a cell division in which one or both the daughter cells remain undifferentiated and retain the ability to give rise to another stem cell as well as to differentiated progeny. Proliferation does not require either daughter cell to be a stem cell, nor does it retain the ability to give rise to multiple lineages. Second, they must differentiate in order to maintain a genetically regulated pool of mature cells under homeostatis, and to produce increased numbers of a particular lineage in response to environmental stresses, including bleeding or infection. Third, the total number of HSCs must be under strict genetic regulation. Since only stem cells comprise a rare subset of cells within a tissue, they must be isolated in order to study their biological, molecular, and biochemical properties. Although it is thought that they give rise to most tissues, stem cells have been rigorously identified and purified only from a handful. The stem cells that give rise to the lymphohematopoietic system, called *hematopoietic stem cells*, were one of the first stem cell populations characterized and have been isolated from both mice and humans. The utility of stem cell biology in the clinic has already been demonstrated in cancer therapy with their extensive use for bone marrow transplantation to regenerate the hematolymphoid system following myeloablative protocols.

3.2 The Hierarchy of the Hematopoietic System

The multipotent cells of the hematopoietic system constitute only 0.05% of mouse bone marrow cells and are heterogeneous in their ability to self-renew. The lymphohematopoietic system consists of a hierarchy in which long-term HSCs give rise to short-term HSCs, which in turn give rise to multipotent progenitors; only long-term and short-term HSCs have detectable self-renewal potential. Cells become more mitotically active as they progress through the differentiation hierarchy, but lose the ability to self-renew. Only long-term HSCs can replenish the full hematopoietic system for the lifetime of an animal; short-term HSCs and multipotent progenitors reconstitute lethally irradiated mice for less than eight weeks.

3.3 Self-renewal: The Defining Property of Stem Cells

While the functional and phenotypic properties of mouse and human HSCs have been extensively studied, our understanding of the fundamental property defining stem cells, self-renewal, is minimal. HSCs differentiate when exposed to combinations of growth factors that can induce extensive proliferation in long-term cultures, complicating *in vitro* studies of self-renewal. Despite recent progress in identifying culture conditions that maintain HSC activity in culture for a limited period of time, it has proven exceedingly difficult to identify tissue culture conditions that allow a significant and prolonged expansion of progenitors while maintaining a transplantable HSC population.

4
Genetic Regulation of Self-renewal in Normal Stem Cells and Cancer Cells

4.1
Tissue and Tumor Growth: A Balancing Act

The maintenance and composition of a tissue or tumor is determined by the balance of proliferation and cell death. It is therefore logical that many of the mutations that are known to drive tumor expansion and progression regulate either cell proliferation or survival. For example, the prevention of apoptosis by enforced expression of the oncogene *BCL-2* promotes the development of lymphoma as well as results in an increased number of HSCs *in vivo*, suggesting that cell death plays a role in the regulation of HSC homeostasis. Inhibition of intrinsic cell death pathways has been further implicated in tumorigenesis by the observation that the progression to experimental AML in a mouse model requires at least three, and likely four independent events. Proto-oncogenes that drive proliferation of tumor cells, such as *c-myb* and *c-myc*, are also essential for HSC development and proliferation.

4.2
Multiple Oncogenes Implicated in Self-renewal

Self-renewal is critical for the maintenance of both normal and cancer stem cell populations. In a normal tissue, stem cell numbers are under strict genetic regulation, resulting in the maintenance of a stable stem cell pool. Cancer stem cells, however, have escaped this homeostatic regulation. Not only is the number of stem cells constantly expanding but they are also proliferating without the normal constraints on their expansion. This results in the relentless growth of the tumor. Since cancer cells and normal stem cells share the ability to self-renew, it is not surprising that a number of genes classically associated with cancer have recently been implicated in normal stem cell development.

In combination with other growth factors, Sonic Hedgehog (Shh) signaling was implicated in the regulation of self-renewal when it was found that cell populations highly enriched for human HSCs exhibited increased self-renewal in response to Shh stimulation *in vitro*. Mice that are deficient in *tal-1/SCL*, which is involved in some cases of human acute leukemia, lack embryonic hematopoiesis. This suggests that *tal-1/SCL* is required either to initiate hematopoiesis for the maintenance of the earliest definitive blood cells, or for the decision to form blood cells downstream of embryonic HSCs. Members of the Hox family have likewise been implicated in human leukemia. Enforced expression of *HoxB4* can affect stem cell functions. One of the major targets of the p53 tumor suppressor gene, $p21^{cip1}$, is necessary for HSCs to repopulate lethally irradiated recipients. Failure at serial transfer could result from exhaustion of the stem cell pool, loss of telomeres, or loss of transplantability. The *bmi-1* gene, which cooperates with *c-myc* in mice to induce lymphoma, is required for the maintenance of both adult HSCs and leukemia cells. Taken together, this body of evidence suggests that the same pathways involved in stem cell fate decisions and the ability of stem cells to self-renew are also involved in malignant transformation.

The Wnt/β-catenin and Notch pathways, which have been implicated in cancer in both mice and humans, may have critical functions in the self-renewal of both normal and cancer stem cells. The Notch

family of receptors was first identified in *Drosophila* and has been implicated in both development and differentiation. In *Caenorhabditis elegans*, Notch plays a role self-renewal of germ cells. In higher animals, Notch plays a critical role in the fate determination of neural crest stem cells. Specifically, transient Notch activation induces an irreversible switch from neurogenesis to gliogenesis. In cultured HSCs, Notch activation using either of the Notch ligands (Jaged-1 or Delta) transiently increased the number of primitive progenitors; this result proved true in *in vivo* studies as well. This suggests that Notch activation promotes either the maintenance of progenitor cell multipotentiality or the maintenance of self-renewal in HSCs. The Notch pathway was implicated in cancer when it was determined that the mouse oncogene *int-3* was a truncated Notch-4. However, the role for Notch in *de novo* human cancer is complex and incompletely understood. It is known that some members of the Notch signaling family are expressed in cancers of epithelial origin, and activation of Notch by chromosomal translocation is involved in some cases of leukemia. Microarray analysis has demonstrated overexpression by tumor cells. Specifically, overexpression of Notch-1 leads to growth arrest of a small cell lung cancer cell line, while inhibition of Notch-1 signals can induce leukemia cell lines to undergo apoptosis. Elegant work by Miele and colleagues showed that activation of Notch-1 signaling maintains the neoplastic phenotype in ras-transformed human cells. They also found that in *de novo* cancers, cells with an activating ras-mutation also demonstrated increased expression of Notch-1 and Notch-4. Additionally, a truncated Notch-4 mRNA is expressed by some breast cancer cell lines.

In addition to the Notch pathway, Wnt/β-catenin signaling has also been implicated as a major player in the malignant transformation and self-renewal of normal stem cells. The Wnt pathway was first implicated in cancer when it was discovered that some murine mammary tumors were the result of mouse mammary tumor virus proviral insertion into the *Wnt-1* gene, causing deregulated expression of the locus. Subsequently, it has been shown that Wnt proteins play a central role in pattern formation during embryogenesis as well as in some adult tissues. The Wnts are a large family of highly hydrophobic secreted proteins that signal through their cognate receptors, members of the frizzled and low-density lipoprotein receptor-related protein families. Activation of these receptors results in the cytoplasmic and nuclear stabilization of β-catenin by blocking its degradation. In the absence of receptor activation, β-catenin is marked for degradation by ubiquitination by a complex consisting of the adenomatous *polyposis coli* (APC), axin, and glycogen synthase-3β proteins. Activation of Wnt/β-catenin signaling by Wnt proteins in HSCs *in vitro* or by expression of a constitutively active β-catenin in HSCs expands the pool of early progenitor cells and enriches normal transplantable HSCs in tissue culture and *in vivo*. Inhibition of Wnt/β-catenin by ectopic expression of axin (which promotes β-catenin degradation) leads to the inhibition of stem cell proliferation both *in vitro* and *in vivo*. Additional studies implicate the Wnt/β-catenin pathway in the maintenance of stem cell or progenitor cell self-renewal in other tissues as well. Specifically, the level of β-catenin within a particular keratinocyte directly correlates with its proliferative capacity. Additionally,

enforced expression of an activated β-catenin increased the ability of epidermal stem cells to self-renew and decreased their ability to differentiate. In the intestinal epithelium, it has been observed that β-catenin levels are highest in the bottom of the crypts (where the epithelial stem cells are thought to reside), and decrease in a gradient as cells differentiate and move up and out of the crypt. Knockout mice for TCF-4, one of the transcription factors that is activated when bound to β-catenin, soon exhaust their undifferentiated crypt epithelial progenitor cells; this further suggests that Wnt signaling is involved in the self-renewal of epithelial stem cells in the intestine.

Activation of the Wnt/β-catenin can take place by one of several mechanisms. In colon cancer, inactivation of the β-catenin protein degradation pathway, most frequently by mutation of APC, is common. Alternatively, elevated expression of *Wnt* genes in some epithelial cancers suggests that activation of β-catenin might sometimes be secondary to ligand activation. There is now evidence that constitutive activation of the Wnt/β-catenin pathway may confer upon cancer cells some of the key properties of cancer stem cells. Inhibition of β-catenin/TCF-4 signaling in a colon cancer cell line induced the expression of the cell cycle inhibitor $p21^{cip-1}$ and induced the cells to stop proliferating and to acquire a more differentiated phenotype. Additionally, enforced expression of the proto-oncogene *c-myc*, which is activated by β-catenin through TCF-4, inhibited the expression of $p21^{cip-1}$ and allowed the colon cancer cells to proliferate when β-catenin/TCF-4 signaling was blocked. This further enforced the notion that *c-myc* is a key *Wnt* signaling target gene in the regulation of cell proliferation and differentiation.

4.3 Possibility of a Common Self-renewal Pathway among Multiple Cell Types

The implication that genes such as *Notch*, *Wnt*, *c-myc*, and *Shh* all play roles in the regulation of not only HSC self-renewal, but also in the self-renewal of stem cells from multiple tissues, suggests that there may be a common set of self-renewal pathways utilized both in normal stem cells and possibly in cancer stem cells. It is clear that identifying the molecular mechanisms by which these pathways produce their biological effects, as well as determining whether these pathways interact to regulate the self-renewal of normal stem cells and cancer stem cells, could have profound importance in our understanding of cancer biology and our ability to effectively treat the disease.

5 Target Cells for Malignant Transformation

5.1 Stem Cells as Targets for Malignant Transformation

It has been established that oncogenic mutations often target signaling pathways that regulate proliferation and self-renewal. This fact raises an interesting question: are stem cells, highly proliferative progenitor cells, or both, the target cells of neoplastic transformation? Theoretically, stem cells could be transformed by acquiring mutations that deregulate proliferation, while highly proliferative progenitors could be transformed by mutations conveying the ability to self-renew.

Several lines of evidence suggest that stem cells may play several key roles in the evolution of a cancer. First, multiple

mutations are necessary for a cell to become cancerous. Since both progenitor cells and mature cells have a very limited lifespan, it is unlikely that all of the mutations could occur during the life of these relatively short-lived cells. This evidence suggests that, in many cases, mutations accumulate in a stem cell. Second, the regulation of stem cell expansion and self-renewal is under strict genetic regulation by multiple genes. Unregulated expansion of the stem cell pool would therefore be analogous cancer. Third, most cancers arise in tissues that contain stem cells that have the intrinsic ability to self-renew. Cancer cells must overcome the tight genetic regulation restricting both self-renewal and proliferation. Since cancer cells must possess the ability to self-renew, it follows that stem cells (which could be transformed by altering only proliferative pathways) may more easily undergo malignant transformation than progenitor cells that lack this fundamental property and must therefore activate both proliferative and self-renewal pathways to become malignant. In the hematopoietic system, the only cells with self-renewal capability are HSCs and mature lymphocytes. The common blood cancers, acute leukemias, and lymphomas, may arise from the HSCs or lymphocytes, respectively. This may take place through the constitutive activation of mitogenic pathways associated with the proliferation of normal cells. Finally, while many cancers contain cells displaying aberrant expression of differentiation markers from their tissue of origin, germ cell cancers are the only cancers to produce differentiated cells from multiple tissues. For this reason, marker expression is routinely used to determine the tissue of origin in metastatic disease. If a differentiated cell were to be the origin of a cancer, it would be necessary for the cell to dedifferentiate exclusively along its parent lineage to produce expression of markers from multiple cell types within the tissue. However, if a somatic stem cell or early multipotent progenitor were transformed, the cell would already be genetically programmed to give rise to cells expressing markers of multiple lineages, but from only one tissue type, as is usually observed clinically. Since most identified oncogenic mutations are involved in the regulation of proliferation and self-renewal and not lineage commitment, the latter hypothesis seems more likely.

5.2 Progenitor Cells as Targets for Malignant Transformation

Although evidence supports the hypothesis that stem cells are frequently the target of mutations that lead to malignant transformation, it is likely that progenitor cells, possibly arising from mutated stem cells, may be transformed by subsequent genetic events that confer immortality and/or self-renewal potential to these normally non-self-renewing cells. In CML patients, the BCR-ABL mutation is present in both normal HSCs and leukemic stem cells. In otherwise normal hematopoietic cells, the BCR-ABL mRNA is expressed solely by the progenitor cells. BCR-ABL expression targeted to myeloid progenitor cells by the hMRP-8 promoter in a transgenic mouse model of CML resulted in CML-like disease in a subset of the mice. Furthermore, when hMRP8p210$^{BCR/ABL}$ mice were crossed with hMRP8bcl-2 mice (which possessed myeloid progenitors resistant to apoptosis), a portion of the mutant mice developed a disease resembling AML. Although expression of transforming genes was targeted to early progenitor cells and not the HSCs, the

morphology of the leukemia cells and clinical course resembled human CML and AML in the hMRP8p210$^{BCR/ABL}$ mice and the hMRP8p210$^{BCR/ABL}$/hMRP8bcl-2 mice, respectively.

In a mouse model of high grade glioblastoma, enforced expression of the epidermal growth factor receptor (EGF-R) in enriched populations of either Ink4a/Arf null neuronal stem cells or Ink4a/Arf null astrocytes led to malignant glioblastomas when the cells were injected orthotopically into mice. Notably, in the majority of cases the transformed astrocytes appeared to acquire an immature phenotype in the brains of the mice, suggesting that there was "dedifferentiation." However, such dedifferentiation has never been conclusively demonstrated. There are two other possible explanations for these results. First, since it is unlikely that the cell culture achieved 100% purity, it is possible that the astrocyte tissue culture cells could have contained a population of neuronal stem cells that were transformed; these stem cells could then have been responsible for generating the tumors. In this case, the immature appearance of the cells could reflect a deregulated expansion of the stem cell pool or early progenitor pool, the result of which would be the overt glioblastoma that was observed. Second, it is possible that the tissue culture conditions, in conjunction with the enforcement of EGF-R expression, could have caused the dedifferentiation of the astrocytes *in vitro*, and that unless the astrocytes were grown in tissue culture first they could not later give rise to glioblastomas in an animal.

The human and mouse observations support the notion that oncogenic mutations accumulate in the stem cells. Expression of the mutated gene by progenitors downstream of the stem cells, or an additional mutation in a short-lived progenitor cell, can lead to neoplastic transformation. These observations have implications for targeted therapies.

It is possible that only a minority of the mice, whose progenitors express BCR-ABL, develop leukemia because the progenitors must acquire an additional mutation that causes deregulated self-renewal. There are two lines of evidence that support this hypothesis. First, expression of *bmi-1* is necessary for the self-renewal of adult HSCs; blast cells of patients with AML express large amounts of this protein. While expression of *HoxA9* and *Meis1* induces transplantable AML in normal mice, expression of these genes in the absence of Bmi-1 does not. This suggests that, in at least some leukemic blasts, a mutation activating *bmi-1* must take place in order to confer self-renewal capability onto the blast cells. Second, deregulated β-catenin signaling occurs in many *de novo* human cancers and causes cancer in transgenic mouse models. Because expression of a constitutively active β-catenin can promote the self-renewal of normal HSCs and stem cells from other tissues, it is possible that its activation promotes self-renewal of cancer cells in these tissues as well. On the basis of these data, it is clear that achieving an understanding of the molecular regulation of self-renewal in both normal stem cells and cancer cells could lead to more effective therapies for cancer.

5.3 Targets for Malignant Transformation in Hematopoietic Malignancies

The systems in which the target cells for transformation are best understood are hematopoietic malignancies; this is because the developmental hierarchy of the

blood is well understood. When this knowledge is combined with analysis of oncogenic mutations, it can be inferred from which cell type along the differentiation lineage the cancer arose. One of the most frequent mutations in AML is the t(8 : 21) mutation, which results in the expression of a chimeric AML-ETO transcript in the leukemic cells. HSCs isolated from patients in clinical remission were found to express chimeric AML-1-ETO transcript up to 90% of the time. When analyzed using *in vitro* assays, these HSCs gave rise to normal progeny, thus demonstrating that the mutation was present in phenotypically normal stem cells. In these patients, the Thy1$^-$ subset of CD34$^+$CD38$^-$ Lin$^-$ cells gave rise to leukemic colonies *in vitro*. These cells could represent HSCs that lost Thy1 expression, or downstream multipotent progenitors that have gained the ability to self-renewal. Collectively, these observations support the idea that mutations accumulate in stem cells, and either subsequent mutations in stem cell progeny or selective activation in downstream progenitors results in overt leukemia.

6 Implications of Cancer Stem Cells for the Diagnosis and Treatment of Cancer

6.1 Cancer Stem Cells Would Explain Several Conundrums of Cancer Biology

Although the NOD/SCID mouse model provides compelling evidence in support of the cancer stem cell model, the ultimate confirmation of this hypothesis will require proof in humans. If this model can be confirmed in humans, however, it will have profound implications for the diagnosis and treatment of cancer. Several conundrums of cancer biology would be explained if it were conclusively demonstrated that, in most tumors, only a small population of cancer cells has the ability to self-renew while other populations of cancer cells have only a limited ability to proliferate. For example, for many years it has been recognized that disseminated cytokeratin-positive breast cancer cells can be detected in the bone marrow of patients that never relapse. One possible explanation of this observation is that the cancer cells lie dormant until some unknown event triggers them to renew proliferation. An alternative explanation is that the disseminated cancer cells in this group of patients arose from the spread of nontumorigenic cells, and only when cancer stem cells disseminate and subsequently self-renew, will metastatic tumors form. Thus, the development of diagnostic reagents that allow cancer stem cells to be identified may have prognostic significance for patients with breast cancer.

6.2 Tumor Relapse Might Result from a Failure to Target Cancer Stem Cells

A frustrating clinical observation is that in most solid cancers such as breast cancers, chemotherapy can frequently shrink tumors, but in most patients the tumors rapidly recur and there is only a small increase in patient survival. Most of the present cancer therapeutic agents have been developed for their ability to shrink a tumor; essentially, all of the phenotypically diverse cancer cells are treated as if they possess the ability to significantly contribute to the growth and maintenance of the tumor, as shown in Fig.2a. However, the failure of oncogenic mutations to confer unto all breast cancer

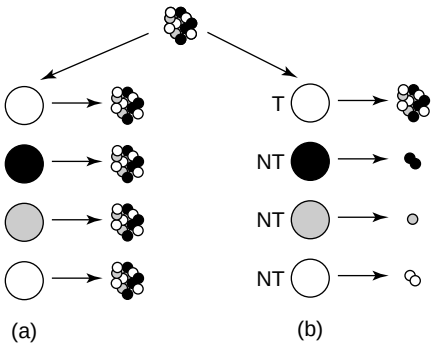

Fig. 2 (a) Therapies are currently evaluated assuming all cancer cells are equivalently tumorigenic. (b) The cancer stem cell model says only a distinct subpopulation drives tumorigenesis. Symbols: T = tumorigenic, NT = nontumorigenic.

and leukemia cells the ability to form new tumors or proliferate extensively underscores the fact that the effects of oncogenic mutations are not equivalent in all cancer cells, as shown in B in Fig 2. If only a minority of the cancer cells are tumorigenic and therefore responsible for driving tumor growth and metastasis, then tumor shrinkage must primarily reflect the elimination of the bulk population of nontumorigenic cells. If a substantial number of tumor stem cells were spared by the therapy, then the tumors would regenerate from these cells. The clinical observation that many patients treated with chemotherapy initially realize tumor shrinkage but then experience relapse at sites of prior disease, supports this model.

6.3 Cancer Stem Cells Might Be More Resistant to Chemotherapy

The similarities between normal stem cells and AML tumor-initiating cells suggest that they might share additional clinically significant parallels. For instance, compared to their differentiated progeny, normal HSCs express high levels of genes that make them more resistant to cytotoxic agents, including both antiapoptotic members of the *BCL-2* family and members of the ABC transporter family that pump many drugs out of the cell. If the same is true for cancer stem cells, then these cells may be significantly more resistant to cytotoxic agents than their nontumorigenic progeny. In support of this notion, while the chemotherapeutic agent cytosine arabinoside very efficiently killed leukemic blast cells isolated from many patients, it selectively spared the leukemia-initiating cells. This observation suggests that the effect of a particular therapeutic agent on the cancer stem cell population must be taken into account when evaluating its therapeutic efficacy.

6.4 Identification of Cancer Stem Cells Might Lead to Better Drug Selection

The fact that therapeutic agents are currently selected on the basis of their ability to rapidly shrink tumors suggests that agents that selectively target the cancer stem cells could be overlooked in screens to identify potential therapeutic agents. Initially, such agents would be expected to produce only a moderate decrease in tumor growth. However, elimination of the cancer stem cell population would eventually halt the spread of the tumor, possibly leaving a residual, benign mass. Perhaps, the best clinical evidence of this model occurs in patients with teratocarcinoma.

Platinum-based chemotherapy is curative in the majority of these patients; however, many patients are left with residual masses. The immature cancer cells have been eliminated by the chemotherapy, leaving only differentiated cancer cells, in a mature teratoma. Patients with mature teratomas rarely develop metastases; most are cured. This demonstrates that the elimination of the presumed stem cell population is sufficient for curing this solid cancer.

7 Future Implications of Cancer Stem Cells

The ability to prospectively identify cancer stem cells should have exciting implications for the development of new diagnostic tools and therapeutic agents. Currently, all of the cancer cells within a tumor are treated equivalently in their presumed ability to drive tumor growth, invade, and metastasize. The ability to identify these crucial cells will allow efforts to develop novel diagnostics and therapies focused on the cells responsible for the maintenance of the malignancy – the cancer stem cells. For example, immune therapy efforts to identify genes and proteins expressed by cancer cells presently use either whole tumors or all of the phenotypically diverse cells within a tumor. Since the cancer stem cells represent only a minority of the cancer cells in most tumors, it is nearly impossible to identify diagnostic markers or therapies that target these cells without their purification. However, directing expression analysis to enriched populations of cancer stem cells would allow the identification of novel diagnostic markers and therapeutic targets that can be exploited to more effectively diagnose and treat patients. This principle is illustrated by the observation that the BCR/ABL oncogene mRNA appears not to be expressed by HSCs that carry the mutation in their DNA; therefore, efforts to eradicate the cancer by targeting the BCR/ABL protein might fail to target the cancer stem cells.

As illustrated above for AML-initiating cells, the ability to prospectively identify cancer stem cells should also improve our ability to evaluate the curative potential of new therapeutic agents. Although cancer cell lines are useful for the evaluation of particular biochemical pathways, they have proven somewhat unreliable when used to predict the clinical efficacy of a particular therapeutic agent in patients. These cell lines are relatively homogeneous, and even tumors grown using these cell lines are homogeneous. This contrasts sharply with the phenotype of real tumors. However, tumors that are grown in the immunodeficient mouse model of human cancer appear to more closely recapitulate the phenotypic diversity of patients' original tumors, including the generation of both tumorigenic and nontumorigenic cells. This model might therefore more effectively predict the potential usefulness of a particular therapeutic agent. New agents could be tested for their ability to eliminate the tumorigenic (cancer stem cell) component of tumors from multiple patients, allowing the agents with the greatest curative potential to proceed to human clinical trials.

See also Developmental Cell Biology; Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Cotta, C., Swindle, C., Weissman, I.L., Klug, C.A. (2001) in: Klug, C., Jordan, C. (Eds.) *Retroviral*

Transduction of FACS-Purified Hematopoietic Stem Cells, Humana Press, Totowa, NJ.
Reya, T., Morrison, S.J., Clarke, M.F., Weissman, I.L. (2001) Stem cells, cancer, and cancer stem cells, *Nature* **414**, 105–111.

Primary Literature

Akashi, K., Weissman, I.L. (2001) *Developmental Biology of Hematopoiesis*, Oxford University Press, New York.
Al-Hajj, M., Wicha, M., Morrison, S.J., Clarke, M.F. (2003) Prospective identification and isolation of breast cancer tumorigenic cells, *Proc. Natl. Acad. Sci. U.S.A.* **100**(7), 3983–3988.
Antonchuk, J.S.G., Humphries, R.K. (2002) HOXB4-induced expansion of adult hematopoietic stem cells ex vivo, *Cell* **109**(1), 39–45.
Artavanis-Tsakonas, S., Rand, M.D., Lake, R.J. (1999) Notch signaling: cell fate control and signal integration in development, *Science* **284**(5415), 770–776.
Aubele, M., Werner, M. (1999) Heterogeneity in breast cancer and the problem of relevance of findings, *Anal. Cell Pathol.* **19**(2), 53–58.
Austin, T.W., Solar, G.P., Ziegler, F.C., Liem, L., Matthews, W. (1997) A role for the Wnt gene family in hematopoiesis: expansion of multilineage progenitor cells, *Blood* **89**(10), 3624–3635.
Bachoo, R.M., Maher, E.A., Ligon, K.L., Sharpless, N.E., Chan, S.S., You, M.J., Tang, Y., DeFrances, J., Stover, E., Weissleder, R., Rowitch, D.H., Louis, D.N., DePinho, R.A. (2002) Epidermal growth factor receptor and Ink4a/Arf: convergent mechanisms governing terminal differentiation and transformation along the neural stem cell to astrocyte axis, *Cancer Cell* **1**(3), 269–277.
Baum, C., Uchida, N., Peault, B., Weissman, I.L., (1994) *Bone Marrow Transplantation*, Blackwell Scientific Publications, Boston.
Baum, C.M., Weissman, I.L., Tsukamoto, A.S., Buckle, A.M., Peault, B. (1992) Isolation of a candidate human hematopoietic stem-cell population, *Proc. Natl. Acad. Sci. U.S.A.* **89**(7), 2804–2808.
Bedi, A., Zehnbauer, B.A., Collector, M.I., Barber, J.P., Zicha, M.S., Sharkis, S.J., Jones, R.J. (1993) BCR-ABL gene rearrangement and expression of primitive hematopoietic progenitors in chronic myeloid leukemia, *Blood* **81**(11), 2898–2902.
Bergsagel, D.E., Valeriote, F.A. (1968) Growth characteristics of a mouse plasma cell tumor, *Cancer Res.* **28**(11), 2187–2196.
Berry, L., Westlund, B., Schedl, T. (1997) Germline tumor formation caused by activation of glp-1, a Caenorhabditis elegans member of the Notch family of receptors, *Development* **124**, 925–936.
Bhardwaj, G., Murdoch, B., Wu, D., Baker, D.P., Williams, K.P., Chadwick, K., Ling, L.E., Karanu, F.N., Bhatia, M. (2001) Sonic hedgehog induces the proliferation of primitive human hematopoietic cells via BMP regulation, *Nat. Immunol.* **2**(2), 172–180.
Bhatia, M., Wang, J.C., Kapp, U., Bonnet, D., Dick, J.E. (1997) Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice, *Proc. Natl. Acad. Sci. U.S.A.* **94**(10), 5320–5325.
Bonnet, D., Dick, J. (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell, *Nat. Med.* **3**, 730–737.
Braun, S., Pantel, K. (1998) Prognostic significance of micrometastatic bone marrow involvement, *Breast Cancer Res. Treat.* **52**(1–3), 201–216.
Braun, S., Pantel, K. (1999) Micrometastatic bone marrow involvement: detection and prognostic significance, *Med. Oncol.* **16**(3), 154–165.
Brown, J.M. (1997) NCI's anticancer drug screening program may not be selecting for clinically active compounds, *Oncol. Res.* **9**(5), 213–215.
Bruce, W.R., Gaag, H. (1963) A quantitative assay for the number of murine lymphoma cells capable of proliferation in vivo, *Nature* **199**, 79–80.
Buske, C., Feuring-Buske, M., Abramovich, C., Spiekermann, K., Eaves, C.J., Coulombel, L., Sauvageau, G., Hogge, D.E., Humphries, R.K. (2002) Deregulated expression of HOXB4 enhances the primitive growth activity of human hematopoietic cells, *Blood* **100**(3), 862–868.
Cadigan, K.M., Nusse, R. (1997) Wnt signaling: a common theme in animal development, *Genes Dev.* **11**(24), 3286–3305.
Capobianco, A.J., Zagouras, P., Blaumueller, C.M., Artavanis-Tsakonas, S., Bishop, J.M. (1997) Neoplastic transformation by truncated alleles of human NOTCH1/TAN1 and NOTCH2, *Mol. Cell. Biol.* **17**(11), 6265–6273.

Chan, E.F., Gat, U., McNiff, J.M., Fuchs, E. (1999) A common human skin tumour is caused by activating mutations in beta-catenin, *Nat. Genet.* **21**(4), 410–413.

Cheng, T., Rodrigues, N., Shen, H., Yang, Y.-G., Dombkowski, D., Sykes, M., Scadden, D.T. (2000) Hematopoietic stem cell quiescence maintained by p21cip1/waf1, *Science* **287**, 1804–1808.

Christensen, J.L., Weissman, I.L. (2001) Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells, *Proc. Natl. Acad. Sci. U.S.A.* **98**(25), 14541–14546.

Clarke, M.F., Kukowska-Latallo, J.F., Westin, E., Smith, M., Prochownik, E.V. (1988) Constitutive expression of a c-myb cDNA blocks friend murine erythroleukemia cell differentiation, *Mol. Cell. Biol.* **8**(2), 884–892.

Danish, R., el-Awar, O., Weber, B.L., Langmore, J., Turka, L.A., Ryan, J.J., Clarke, M.F. (1992) c-myb effects on kinetic events during MEL cell differentiation, *Oncogene* **7**(5), 901–907.

Datta, Y.H., Adams, P.T., Drobyski, W.R., Ethier, S.P., Terry, V.H., Roth, M.S. (1994) Sensitive detection of occult breast cancer by the reverse-transcriptase polymerase chain reaction, *J. Clin. Oncol.* **12**(3), 475–482.

DiStefano, A., Tashima, C.K., Yap, H.Y., Hortobagyi, G.N. (1979) Bone marrow metastases without cortical bone involvement in breast cancer patients, *Cancer* **44**(1), 196–198.

Domen, J., Cheshier, S.H., Weissman, I.L. (2000) The role of apoptosis in the regulation of hematopoietic stem cells: overexpression of BCL-2 increases both their number and repopulation potential, *J. Exp. Med.* **191**(2), 253–264.

Domen, J., Gandy, K.L., Weissman, I.L. (1998) Systemic overexpression of BCL-2 in the hematopoietic system protects transgenic mice from the consequences of lethal irradiation, *Blood* **91**(7), 2272–2282.

Domen, J., Weissman, I.L. (2000) Hematopoietic stem cells need two signals to prevent apoptosis; BCL-2 can provide one of these, Kitl/c-Kit signaling the other, *J. Exp. Med.* **192**(12), 1707–1718.

Ellisen, L.W., Bird, J., West, D.C., Soreng, A.L., Reynolds, T.C., Smith, S.D., Sklar, J. (1991) TAN-1, the human homolog of the drosophila notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms, *Cell* **66**(4), 649–661.

Fearon, E.R., Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis, *Cell* **61**(5), 759–767.

Feuerhake, F., Sigg, W., Hofter, E.A., Dimpfl, T., Welsch, U. (2000) Immunohistochemical analysis of Bcl-2 and Bax expression in relation to cell turnover and epithelial differentiation markers in the non-lactating human mammary gland epithelium, *Cell Tissue Res.* **299**(1), 47–58.

Fidler, I.J., Hart, I.R. (1982) Biological diversity in metastatic neoplasms: origins and implications, *Science* **217**(4564), 998–1003.

Fidler, I.J., Kripke, M.L. (1977) Metastasis results from preexisting variant cells within a malignant tumor, *Science* **197**(4306), 893–895.

Gallahan, D., Callahan, R. (1997) The mouse mammary tumor associated gene INT3 is a unique member of the NOTCH gene family (NOTCH4), *Oncogene* **14**(16), 1883–1890.

Gat, U., DasGupta, R., Degenstein, L., Fuchs, E. (1998) De Novo hair follicle morphogenesis and hair tumors in mice expressing a truncated beta-catenin in skin, *Cell* **95**(5), 605–614.

George, A.A., Franklin, J., Kerkof, K., Shah, A.J., Price, M., Tsark, E., Bockstoce, D., Yao, D., Hart, N., Carcich, S., Parkman, R., Crooks, G.M., Weinberg, K. (2001) Detection of leukemic cells in the CD34(+)CD38(−) bone marrow progenitor population in children with acute lymphoblastic leukemia, *Blood* **97**(12), 3925–3930.

Golub, T.R. (2001) Genome-wide views of cancer, *N. Engl. J. Med.* **344**(8), 601–602.

Gudjonsson, T., Villadsen, R., Nielsen, H.L., Ronnov-Jessen, L., Bissell, M.J., Petersen, O.W. (2002) Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties, *Genes Dev.* **16**(6), 693–706.

Guzman, M.L., Swiderski, C.F., Howard, D.S., Grimes, B.A., Rossi, R.M., Szilvassy, S.J., Jordan, C.T. (2002) Preferential induction of apoptosis for primary human leukemic stem cells, *Proc. Natl. Acad. Sci. U.S.A.* **99**(25), 16220–16225.

Hamburger, A.W., Salmon, S.E. (1977) Primary bioassay of human tumor stem cells, *Science* **197**(4302), 461–463.

Hanahan, D., Weinberg, R.A. (2000) The hallmarks of cancer, *Cell* **100**(1), 57–70.

Harrison, D.E., Lerner, C.P. (1991) Most primitive hematopoietic stem cells are stimulated to cycle rapidly after treatment with 5-fluorouracil, *Blood* **78**(5), 1237–1240.

Hedgepeth, C.M., Deardorff, M.A., Rankin, K., Klein, P.S. (1999) Regulation of glycogen synthase kinase 3beta and downstream Wnt signaling by axin, *Mol. Cell. Biol.* **19**(10), 7147–7157.

Henrique, D., Hirsinger, E., Adam, J., Le Roux, I., Pourquie, O., Ish-Horowicz, D., Lewis, J. (1997) Maintenance of neuroepithelial progenitor cells by Delta-Notch signalling in the embryonic chick retina, *Curr. Biol.* **7**(9), 661–670.

Heppner, G.H. (1984) Tumor heterogeneity, *Cancer Res.* **44**(6), 2259–2265.

Hoffman, R.M. (1999) Orthotopic metastatic mouse models for anticancer drug discovery and evaluation: a bridge to the clinic, *Invest. New Drugs* **17**(4), 343–359.

Imatani, A., Callahan, R. (2000) Identification of a novel NOTCH-4/INT-3 RNA species encoding an activated gene product in certain human tumor cell lines, *Oncogene* **19**(2), 223–231.

Ingle, J.N., Tormey, D.C., Bull, J.M., Simon, R.M. (1977) Bone marrow involvement in breast cancer: effect on response and tolerance to combination chemotherapy, *Cancer* **39**(1), 104–111.

Jaiswal, S., Traver, D., Miyamoto, T., Akashi, K., Lagasse, E., Weissman, I.L. (2002) Expression of BCR/ABL and BCL-2 in myeloid progenitors leads to myeloid leukemias, *Proc. Natl. Acad. Sci. U.S.A.* **100**(18), 10002–10007.

Janni, W., Gastroph, S., Hepp, F., Kentenich, C., Rjosk, D., Schindlbeck, C., Dimpfl, T., Sommer, H., Braun, S. (2000) Prognostic significance of an increased number of micrometastatic tumor cells in the bone marrow of patients with first recurrence of breast carcinoma, *Cancer* **88**(10), 2252–2259.

Jehn, B.M., Bielke, W., Pear, W.S., Osborne, B.A. (1999) Cutting edge: protective effects of notch-1 on TCR-induced apoptosis, *J. Immunol.* **162**(2), 635–638.

Kirikoshi, H., Inoue, S., Sekihara, H., Katoh, M. (2001) Expression of WNT10A in human cancer, *Int. J. Oncol.* **19**(5), 997–1001.

Knudson, A.G. Jr., Strong, L.C., Anderson, D.E. (1973) Heredity and cancer in man, *Prog. Med. Genet.* **9**, 113–158.

Korinek, V., Barker, N., Moerer, P., van Donselaar, E., Huls, G., Peters, P.J., Clevers, H. (1998) Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4, *Nat. Genet.* **19**(4), 379–383.

Kowenz-Leutz, E., Twamley, G., Ansieau, S., Leutz, A. (1994) Novel mechanism of C/EBP beta (NF-M) transcriptional control: activation through derepression, *Genes Dev.* **8**, 2781–2791.

Kyoizumi, S., Baum, C.M., Kaneshima, H., McCune, J.M., Yee, E.J., Namikawa, R. (1992) Implantation and maintenance of functional human bone marrow in SCID-hu mice, *Blood* **79**(7), 1704–1711.

Lagasse, E., Weissman, I.L. (1994) bcl-2 inhibits apoptosis of neutrophils but not their engulfment by macrophages, *J. Exp. Med.* **179**(3), 1047–1052.

Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M., Dick, J. (1994) A cell initiating human acute myeloid leukaemia after transplantation into SCID mice, *Nature* **17**(367), 645–648.

Leethanakul, C., Patel, V., Gillespie, J., Pallente, M., Ensley, J.F., Koontongkaew, S., Liotta, L.A., Emmert-Buck, M., Gutkind, J.S., (2000) Distinct pattern of expression of differentiation and growth-related genes in squamous cell carcinomas of the head and neck revealed by the use of laser capture microdissection and cDNA arrays, *Oncogene* **19**(28), 3220–3224. Lippman, M.E. (2000) High-dose chemotherapy plus autologous bone marrow transplantation for metastatic breast cancer, *N. Engl. J. Med.* **342**(15), 1119–1120.

Liu, A., True, L., LaTray, L., Nelson, P., Ellis, W., Vessella, R., Lange, P., Hood, L., van den Engh, G. (1997) Cell-cell interaction in prostate generegulation and cytodifferentiation, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10705–10710.

Liu, Y., Dehni, G., Purcell, K., Sokolow, J., Carcangiu, M., Artavanis, T., Stifani, S. (1996) Epithelial expression and chromosomal location of human TLE genes: implications for notch signaling and neoplasia, *Genomics* **31**(1), 58–64.

Manegold, C., Krempien, B., Kaufmann, M., Schwechheimer, K., Schettler, G.(1988) The value of bone marrow examination for tumor staging in breast cancer, *J. Cancer Res. Clin. Oncol.* **114**(4), 425–428.

Miller, C.L., Eaves, C.J. (1997) Expansion in vitro of adult murine hematopoietic stem cells with transplantable lympho-myeloid reconstituting ability, *Proc. Natl. Acad. Sci. U.S.A.* **94**(25), 13648–13653.

Miyamoto, T., Weissman, I.L., Akashi, K. (2000) AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation, *Proc. Natl. Acad. Sci. U.S.A.* **97**(13), 7521–7526.

Morrison, S., Hemmati, H., Wandycz, A., Weissman, I. (1995) The purification and characterization of fetal liver hematopoietic stem cells, *Proc. Natl. Acad. Sci. U.S.A.* **92**(22), 10302–10306.

Morrison, S., Perez, S., Qiao, Z., Verdi, J., Hicks, C., Weinmaster, G., Anderson, D. (2000) Transient notch activation initiates an irreversible switch from neurogenesis to gliogenesis by neural crest stem cells, *Cell* **101**, 499–510.

Morrison, S.J., Qian, D., Jerebek, L., Thiel, B.A., Park, I.-K., Ford, P.S., Kiel, M.J., Schork, N.J., Weissman, I.L., Clarke, M.F. (2002) A genetic determinant that specifically regulates the frequency of hematopoietic stem cells, *J. Immunol.* **168**(2), 635–642.

Morrison, S.J., Shah, N.M., Anderson, D.J. (1997) Regulatory mechanisms in stem cell biology, *Cell* **88**(3), 287–298.

Morrison, S.J., Uchida, N., Weissman, I.L. (1995) The biology of hematopoietic stem cells, *Annu. Rev. Cell Dev. Biol* **11**, 35–71.

Morrison, S.J., Weissman, I.L. (1994) The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype, *Immunity* **1**(8), 661–673.

Mucenski, M.L., McLain, K., Kier, A.B., Swerdlow, S.H., Schreiner, C.M., Miller, T.A., Pietryga, D.W., Scott, W.J.J., Potter, S.S. (1991) A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis, *Cell* **65**(4), 677–689.

Muller-Sieburg, C.E., Cho, R.H., Sieburg, H.B., Kupriyanov, S., Riblet, R. (2000) Genetic control of hematopoietic stem cell frequency in mice is mostly cell autonomous, *Blood* **95**(7), 2446–2448.

Negrin, R.S.,, Atkinson, K., Leemhuis, T., Hanania, E., Juttner, C., Tierney, K., Hu, W.W., Johnston, L.J., Shizurn, J.A., Stockerl-Goldstein, K.E., Blume, K.G., Weissman, I.L., Bower, S., Baynes, R., Dansey, R., Karanes, C., Peters, W., Klein, J. (2000) Transplantation of highly purified CD34+Thy-1+ hematopoietic stem cells in patients with metastatic breast cancer, *Biol. Blood. Marrow Transplant.* **6**(3), 262–271.

Negrin, R.S., Weissman, I.L. (1992) Hematopoietic stem cells in normal and malignant states, *Marrow Transplant.* **2**, 23–26.

Nowell, P.C. (1986) Mechanisms of tumor progression, *Cancer Res.* **46**(5), 2203–2207.

Nusse, R. (1992) The Wnt gene family in tumorigenesis and in normal development, *J. Steroid Biochem. Mol. Biol.* **43**((1–3)), 9–12.

Nusse, R., Brown, A., Papkoff, J., Scambler, P., Schackleford, G., McMahon, A., Moon, R., Varmus, H. (1991) A new nomenclature for int-1 and related genes: the Wnt gene family, *Cell* **64**(2), 231.

Orkin, S.H., Morrison, S.J. (2002) Stem-cell competition, *Nature* **418**(6893), 25–27.

Osawa, M., Hanada, K., Hamada, H., Nakauchi, H. (1996) Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell, *Science* **273**(5272), 242–245.

Ozols, R.F., Willson, J.K., Weltz, M.D., Grotzinger, K.R., Myers, C.E., Young,, R.C. (1980) Inhibition of human ovarian cancer colony formation by adriamycin and its major metabolites, *Cancer Res.* **40**(11), 4109–4112.

Park, C.H., Bergsagel, D.E., McCulloch, E.A. (1971) Mouse myeloma tumor stem cells: a primary cell culture assay, *J. Natl. Cancer Inst.* **46**(2), 411–422.

Park, I.K., Clarke, M.F. (2003) Bmi-1 and senescence regulation in stem cells, *JCI* (submitted).

Park, I.K., Qian, D., Kiel, M.J., Becker, M.W., Pihalja, M., Weissman, I.L., Morrison, S.J., Clarke, M.F. (2003) Bmi-1 is required for the maintenance of self-renewing adult hematopoietic stem cells, *Nature* **423**(693), 302–305.

Peters, R., Leyvraz, S., Perey, L. (1998) Apoptotic regulation in primitive hematopoietic precursors, *Blood* **92**(6), 2041–2052.

Phillips, R.L., Reinhart, A.J., Van Zant, G. (1992) Genetic control of murine hematopoietic stem cell pool sizes and cycling kinetics, *Proc. Natl. Acad. Sci. U.S.A.* **89**(23), 11607–11611.

Platt, F.M., Cebra-Thomas, J.A., Baum, C.M., Davie, J.M., McKearn, J.P. (1992) Monoclonal antibodies specific for novel murine cell surface markers define subpopulations of

germinal center cells, *Cell. Immunol.* **143**(2), 449–466.

Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F., Orkin, S., (1996) The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages, *Cell* **86**(1), 47–57.

Prochowinik, E., Kukowska, J. (1986) Deregulated expression of c-myc by murine erythroleukaemia cells prevents differentiation, *Nature* **322**(6082), 848–850.

Reya, T., Morrison, S.J., Clarke, M.F., Weissman, I.L. (2001) Stem cells, cancer, and cancer stem cells, *Nature* **414**(6859), 105–111.

Reya, T., O'Riordan, M., Okamura, R., Devaney, E., Willert, K., Nusse, R., Grosschedl, R. (2000) Wnt signaling regulates B lymphocyte proliferation through a LEF-1 dependent mechanism, *Immunity* **13**(1), 15–24.

Rhoades, K.L., Hetherington, C.J., Harakawa, M., Yergeau, D.A., Zhou, L., Liu, L.Q., L.M.T., T.D.G. and Z.D.E. (2000) Analysis of the role of AML1-ETO in leukemogenesis, using an inducible transgene mouse model, *Blood* **96**(6), 2108–2115.

Saitoh, T., Mine, T., Katoh, M. Frequent Upregulation of WNT5A mRNA in primary gastric cancer, *Int. J. Mol. Med.* **9**(5), 515–519.

Saitoh, T., Mine, T., Katoh, M. (2002). Upregulation of WNT8B mRNA in human gastric cancer, *Int. J. Oncol.* **20**(2), 343–348.

Salsbury, A.J. (1975) The significance of the circulating cancer cell, *Cancer Treat. Rev.* **2**(1), 55–72.

Schultz, L.B., Weber, B.L. (1999) Recent advances in breast cancer biology, *Curr. Opin. Oncol.* **11**(6), 429–434.

Shelly, L.L., Fuchs, C., Miele, L. (1999) Notch-1 inhibits apoptosis in murine erythroleukemia cells and is necessary for differentiation induced by hybrid polar compounds, *J. Cell. Biochem.* **73**(2), 164–175.

Shivdasani, R., Mayer, E., Orkin, S. (1995) Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL, *Nature* **373**(6513), 432–434.

Southam, C., Brunschwig, A. (1961) Quantitative studies of autotransplanation of human cancer, *Cancer* **14**, 971–978.

Spangrude, G.J., Heimfeld, S., Weissman, I.L. (1988) Purification and characterization of mouse hematopoietic stem cells, *Science* **241**(4861), 58–62.

Spink, K.E., Polakis, P., Weis, W.I. (2000) Structural basis of the axin-adenomatous polyposis coli interaction, *EMBO J.* **19**(10), 2270–2279.

Stadtmauer, E.A., Tsai, D.E., Sickles, C.J., Mick, R., Luger, S.M., Porter, D.L., Mangan, P.A., Schuchter, L.M., Schuster, S.J., Loh, E.Y., Magee, D.A., Sachs, R.A., Wall, M.E., Moore, J., Buzby, G.P., Zaleta, E., Kamoun, M., Silberstein, L.E. (1999) Stem cell transplantation for metastatic breast cancer: analysis of tumor contamination, *Med. Oncol.* **16**(4), 279–288.

Stingl, J., Eaves, C., Kuusk, U., Emerman, J. (1998) Phenotypic and functional characterization in vitro of a multipotent epithelial cell present in the normal adult human breast, *Differentiation* **63**, 201–213.

Taipale, J., Beachy, P.A. (2001) The hedgehog and Wnt signalling pathways in cancer, *Nature* **411**(6835), 349–354.

Terskikh, A.V., Easterday, M.C., Li, L., Hood, L., Kornblum, H.I., Geschwind, D.H., Weissman, I.L. (2001) From hematopoiesis to neuropoiesis: evidence of overlapping genetic programs, *Proc. Natl. Acad. Sci. U.S.A.* **98**(14), 7934–7639.

Traver, D., Akashi, K., Weissman, I.L., Lagasse, E. (1998) Mice defective in two apoptosis pathways in the myeloid lineage develop acute myeloblastic leukemia, *Immunity* **9**(1), 47–57.

Tsukamoto, A.S., Grosschedl, R., Guzman, R.C., Parslow, T., Varmus, H.E. (1988) Expression of the int-1 gene in transgenic mice is associated with mammary gland hyperplasia and adenocarcinomas in male and female mice, *Cell* **55**(4), 619–625.

Uchida, N., Buck, D., He, D., Reitsma, M., Masek, M., Phan, T., Tsukamoto, A., Gage, F., Weissman, I. (2000) Direct isolation of human central nervous system stem cells, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14720–14725.

Uchida, N., Weissman, I.L. (1992) Searching for hematopoietic stem cells: evidence that Thy-1.1lo Lin- Sca-1+ cells are the only stem cells in C57bL/Ka-Thy1.1 bone marrow, *J. Exp. Med.* **175**, 175–184.

van de Wetering, M., Sancho, E., Verweij, C., de Lau, W., Oving, I., Hurlstone, A., van der Horn, K., Batlle, E., Coudreuse, D., Haramis, A.P., Tjon-Pon-Fong, M., Moerer, P., van den Born, M., Soete, G., Pals, S., Eilers, M., Medema, R., Clevers, H. (2002) The beta-catenin/TCF-4 complex imposes a crypt

progenitor phenotype on colorectal cancer cells, *Cell* **111**(2), 241–250.

van der Lugt, N.M., Domen, J., Linders, K., van Roon, M., Robanus-Maandag, E., te Riele, H., van der Valk, M., Deschamps, J., Sofroniew, M., van Lohuizen, M. (1994) Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene, *Genes Dev.* **8**(7), 757–769.

van Lohuizen, M., Frasch, M., Wientjens, E., Berns, A. (1991) Sequence similarity between the mammalian bmi-1 proto-oncogene and the drosophila regulatory genes Psc and Su(z)2, *Nature* **353**(6342), 353–355.

VanDenBerg, D.J., Sharma, A.K., Bruno, E., Hoffman, R. (1998) Role of members of the Wnt gene family in human hematopoiesis, *Blood* **92**, 3189–3202.

Varnum-Finney, B., Xu, L., Brashem-Stein, C., Nourigat, C., Flowers, D., Bakkour, S., Pear, W.S., Bernstein, I.D. (2000) Pluripotent, cytokine-dependent, hematopoietic stem cells are immortalized by constitutive notch1 signaling, *Nat. Med.* **6**(11), 1278–1281.

Wechsler-Reya, R., Scott, M.P. (2001) The developmental biology of brain tumors, *Annu. Rev. Neurosci.* **24**, 385–428.

Weeraratna, A.T., Jiang, Y., Hostetter, G., Rosenblatt, K., Duray, P., Bittner, M., Trent, J.M. (2002) Wnt5 signaling directly affects cell motility and invasion of metastatic melanoma, *Cancer Cell* **1**(3), 279–288.

Weisenthal, L., Lippman, M.E. (1985) Clonogenic and nonclonogenic in vitro chemosensitivity assays, *Cancer Treat. Rep.* **69**, 615–632.

Weissman, I.L. (2000) Translating stem and progenitor cell biology to the clinic: barriers and opportunities, *Science* **287**(5457), 1442–1446.

Weizen, S., Rizzo, P., Braid, M., Vaishnav, R., Jonkheer, S., Zlobin, A., Osborne, B., Gottipati, S., Aster, J., Hahn, W., Rudolf, M., Siziopikou, K., Kast W., Miele, L. (2002) Activation of notch-1 signaling maintains the neoplasitc phenotype in human Ras-transformed cells, *Nat. Med.* **8**, 979–986.

Williams, S.D., Birch, R., Einhorn, L.H., Irwin, L., Greco F.A., X Loehrer P.J., (1987) Treatment of disseminated germ-cell tumors with cisplatin, bleomycin, and either vinblastine or etoposide, *N. Engl. J. Med.* **316**(23), 1435–1440.

Wodinsky, I., Swiniarski, J., Kensler, C.J. (1967) Spleen colony studies of leukemia L1210.I. Growth kinetics of lymphocytic L1210 cells in vivo as determined by spleen colony assay, *Cancer Chemother. Rep.* **51**, 415–421.

Wu, C., Zeng, Q., Blumer, K.J., Muslin, A.J. (2000) RGS proteins inhibit Xwnt-8 signaling in xenopus embryonic development, *Development* **127**(13), 2773–2784.

Zagouras, P., Stifani, S., Blaumueller, C., Carcangiu, M., Artavanis-Tsakonas, S. (1995) Alterations in notch signaling in neoplastic lesions of the human cervix, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6414.

Zhou, S., Schuetz, J.D., Bunting, K.D., Colapietro, A.M., Sampath, J., Morris, J.J., Lagutina, I., Grosveld, G.C., Osawa, M., Nakauchi, H., Sorrentino, B.P. (2001) The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype, *Nat. Med.* **7**(9), 1028–1034.

Zhu, A.J., Watt, F.M. (1999) beta-catenin signalling modulates proliferative potential of human epidermal keratinocytes independently of intercellular adhesion, *Development* **126**(10), 2285–2298.

Carbohydrate Analysis

Martin F. Chaplin
South Bank University, London, UK

1 Outlines of Methodology 246
1.1 Introduction 246
1.2 Monosaccharides 249
1.3 Oligosaccharides 250
1.4 Polysaccharides 250
1.5 Proteoglycans 250
1.6 Glycoproteins 251
1.7 Glycolipids 253

2 Key Methods 253
2.1 The Use of Enzymes in Analysis 253
2.2 Chromatography 255
2.3 Determination of Size and Conformation 259
2.3.1 Light Scattering 259
2.3.2 Size-exclusion Chromatography 260
2.3.3 Field Flow Fractionation 260
2.3.4 Mass Spectrometry (MS) 261
2.3.5 NMR Spectrometry 263
2.3.6 Molecular Modeling 266
2.4 Linkage Analysis 266
2.4.1 Glycan Detection 266
2.4.2 Glycan Release and the Use of Enzymes 266
2.4.3 Methylation Analysis 266
2.4.4 Mass Spectrometry 268
2.4.5 NMR Spectroscopy 268
2.5 Use of Lectins 271

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Bibliography 272
Books and Reviews 272
Primary Literature 272

Keywords

Anomeric Structure
The steric configuration around the hemiacetal (reducing) carbon atom in carbohydrates.

Antennary Structure
The number and composition of the nonreducing oligosaccharide chains in glycoprotein glycans.

Capillary Electrophoresis
Chromatographic method that utilizes the interaction of eluant and analyte with the capillary wall to separate molecules in an electric field.

COSY
Correlated spectroscopy. A two-dimensional NMR technique that determines the through-bond connectivity of atoms.

Endoglycosidase
Enzyme that hydrolyzes internal glycosidic links in carbohydrate chains.

Exoglycosidase
Enzyme that hydrolyzes nonreducing terminal glycosidic links in carbohydrate chains.

FAB-MS
Fast atom bombardment mass spectrometry.

FFF
Field flow fractionation is a chromatographic technique that separates on the basis of molecular size.

GC
Gas–liquid chromatography (also abbreviated GLC). A method for separating volatile molecules.

Glycan
Oligosaccharide or polysaccharide chain present in carbohydrate-containing molecules.

HPAEC-PAD
High-performance anion-exchange chromatography with pulsed amperometric detection; an HPLC system developed for polyalcohols.

HPLC
High-performance liquid chromatography; a method for separating molecules in solution, often conducted at relatively high pressures.

ICUMSA
The International Commission for Uniform Methods of Sugar Analysis, c/o British Sugar plc, Technical Centre, Colney, Norwich, UK. Publishes standardized methods of food carbohydrate analysis at regular intervals.

Lectins
Proteins or glycoproteins of nonimmune origin that bind carbohydrates.

MALDI-TOF
Matrix-assisted laser-desorption ionization mass spectrometry.

MALLS
Multiangle laser light scattering is a technique for determining the molecular weight of polymers.

NMR
Nuclear magnetic resonance; a spectroscopic process for examining nuclear environments.

NOESY
Nuclear Overhauser effect spectroscopy. A two-dimensional NMR technique that determines the through-space relationship of atoms.

O-Glycoside
Compounds in which a sugar is linked to another molecule through its reducing group to form an acetal.

ROESY
Rotational nuclear Overhauser effect spectroscopy. A variation in the two-dimensional NMR NOESY technique that determines the through-space relationship of atoms. It is especially useful for cases in which NOESY signals are weak in spite of atoms being close.

SEC
Size-exclusion chromatography is a technique that separates on the basis of molecular size.

TLC
Thin-layer chromatography; a chromatographic system based on partition.

Carbohydrates form a complex family of molecules that necessitate a variety of analytical approaches. They are capable of forming a large number of distinct structures using relatively few units. In addition, the more complex carbohydrates are rarely found to exist as single structural entities; polymorphism is common, and often a given molecule is found as an assortment of closely related structures. These mixtures of complex structures give rise to analytical problems that are much more difficult to solve than those concerning protein analysis. Typical protocols involve isolation and purification, determination of the constituent carbohydrates, partial hydrolysis of polysaccharides or glycosaminoglycans or cleavage of the intact glycans from the protein or lipid moieties, and isolation and purification of these fragments. Chromatographic purification techniques are followed by mass spectrometry and/or NMR spectroscopy. Absolute molecular weights of larger molecules are determined by light scattering, but uncovering their conformation often requires the assistance of molecular dynamics. However difficult to obtain, a solid understanding of the structure and conformation of carbohydrates is required a our proper understanding of their biological roles.

1
Outlines of Methodology

1.1
Introduction

Carbohydrates, which form the most diverse of all the major groups of biological compounds, possess a wide variety of structural types with very different chemical and physical characteristics. They are formed from basic structures that commonly have a backbone of five or six carbon atoms and possess several alcohol groups and a hemiacetal or, more rarely, a hemiketal group. Less commonly found are carbohydrate units containing a nine-carbon backbone, whilst those with seven and eight carbons are only rarely found. In spite of the extraordinarily large number of isomers possible, only relatively few basic units (*monosaccharides*) are found as the basic building blocks of more complex molecules (Fig. 1). These units link in a number of ways between the monosaccharide units utilizing different hydroxyl groups and forming simple or complex structures. Different monosaccharides may be joined together in an ordered or apparently random manner to form a mainly homogeneous or heterogeneous family of structural entities, termed *glycans*. Again, only a limited subgroup of the possible isomers is generally found. For example, more than 50 different D-glucose dimers and more than 1000 trimers are possible, although only 5 dimers and 2 trimers are commonly encountered. These carbohydrate moieties may also be linked to proteins or lipids in glycoproteins, proteoglycans,

and glycolipids. As these glycans are not directly genetically coded, but their organization depends on the sequential action of enzymes, all such carbohydrates are commonly found to be somewhat heterogeneous. This may be a microheterogeneity as evident in glycoprotein glycans in which most of the structure remains constant, but a variety of structures exist due to missing residues at the molecule's periphery. Alternatively, there may be sequence and linkage heterogeneity, as found in heteropolysaccharides, or molecular weight heterogeneity, as is common in homopolysaccharides.

The principal classes of carbohydrates are listed in Table 1, together with the most commonly encountered analytical methods. Since glycoproteins form the class of carbohydrates containing material that is of major importance to cell and molecular biologists, this article concentrates on methods for their analysis. The analytical problems surrounding the carbohydrates are clearly orders of magnitude more difficult than that of protein analysis; indeed it is apparent that for many polysaccharides no two molecules are likely to be identical and the concept of "structure" must be investigated and interpreted differently from that for proteins with more emphasis on the statistical relevance. There is a further level of complexity if the conformation of the carbohydrates in solution is sought as, in contrast to protein structures, these are generally not fixed but fluctuate widely and may change irreversibly during isolation or with apparently minor changes in the temperature, solvent, pH, ionic strength, concentration, or co-solutes.

Tab. 1 The important analytical methods for carbohydrates.

Carbohydrate	*Usual analytical method(s)*
Monosaccharides and disaccharides	
In syrups	ICUMSA empirical methods
In food	Enzymatic methods
In glycoconjugates and polysaccharides	Colorimetric assays
Hydrolyzed carbohydrate	HPLC, HPAEC-PAD, GC, GC-MS
Oligosaccharides	HPLC, HPAEC-PAD, GC-MS
Neutral polysaccharides	Separation and extraction techniques, component analysis, methylation analysis, partial acid and enzymatic hydrolysis, mass spectrometry, and ^{13}C and ^{1}H NMR spectroscopy
Proteoglycans	Separation and extraction techniques, component analysis, β-elimination, and enzymatic hydrolysis followed by HPAEC-PAD, and ^{13}C and ^{1}H NMR spectroscopy
Glycoproteins and glycolipids	Cleavage from protein/lipid, lectin binding, component analysis, methylation analysis, enzymatic hydrolysis, mass-spectrometric techniques (fast atom bombardment, laser desorption, and electrospray), ^{13}C and ^{1}H NMR, and two-dimensional NMR.

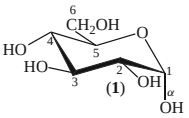
(1)

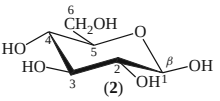
(2)

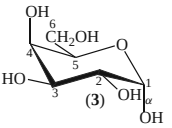
(3)

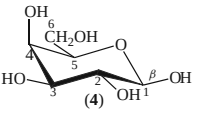
(4)

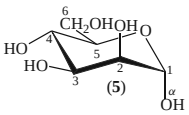
(5)

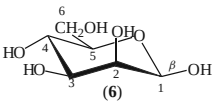
(6)

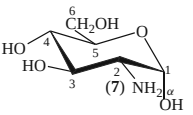
(7)

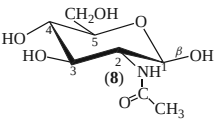
(8)

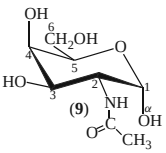
(9)

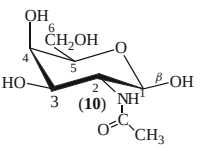
(10)

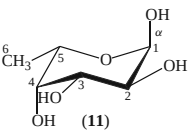
(11)

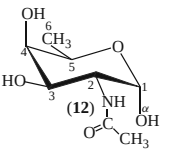
(12)

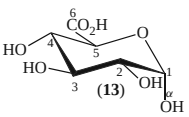
(13)

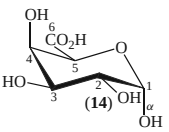
(14)

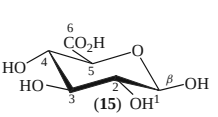
(15)

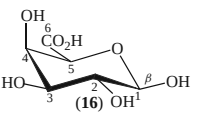
(16)

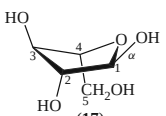
(17)

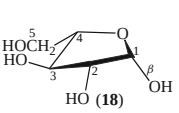
(18)

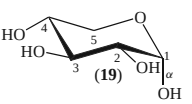
(19)

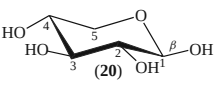
(20)

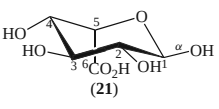
(21)

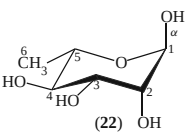
(22)

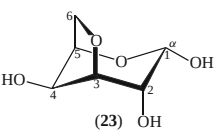
(23)

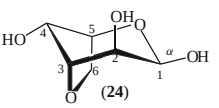
(24)

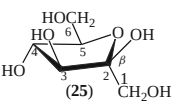
(25)

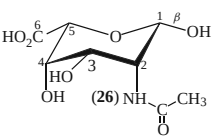
(26)

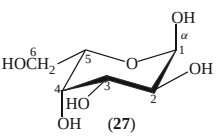
(27)

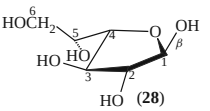
(28)

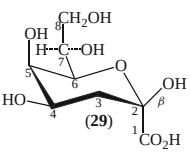
(29)

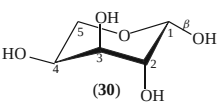
(30)

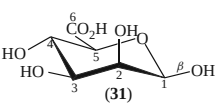
(31)

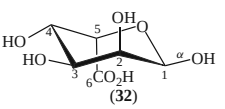
(32)

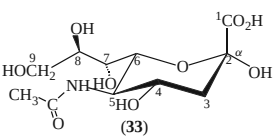
(33)

Fortunately, although the analytical task remains formidable, the full range of monosaccharide ring size and linkage diversity is not found, with most glycans being constructed out of a very limited subset of possible molecular building blocks. However, the analysis of polysaccharides and glycans remains challenging and must be tackled in a systematic manner. There is no single method applicable to this analysis. Even for simple carbohydrates like glucose, a number of methods are used depending on the purpose of the analysis. Complex carbohydrates may consist of several monosaccharide moieties combined in differing ways and proportions and covalently attached to proteins or other molecules. In these cases, it is necessary to determine not only the relative amounts of the carbohydrates present but also their absolute configuration (D or L), their ring size (pyranose or furanose), the anomeric configuration of the linkages (α or β), the linkage positions and sequence, and the nature and location of any inorganic groups or other nonsugar components. Structures are sometimes determined or confirmed by the use of specific lectin binding. Detailed structural analysis of the carbohydrate moieties in glycoconjugates enables the investigation and determination of biosynthetic pathways and biological processes and may offer evidence for the mechanism of metabolic disorders and clinical behavior. As a final refinement, the molecular conformation(s) may have to be determined in order to fully understand the biological relevance of the structure.

1.2 Monosaccharides

There are well-established standardized methods, validated by the International Commission for Uniform Methods of Sugar Analysis (ICUMSA), for analyzing individual low molecular weight carbohydrates in foodstuffs. These make use of nonstoichiometric colorimetric assays for reducing groups or physical methods such

Fig. 1 The commonest monosaccharides present in glycans, with their common abbreviations in brackets. (1) α-D-glucose (Glc); (2) β-D-glucose (Glc); (3) α-D-galactose (Gal); (4) β-D-galactose (Gal); (5) α-D-mannose (Man); (6) β-D-mannose (Man); (7) α-D-glucosamine (GlcN); (8) *N*-acetyl-β-D-glucosamine (GlcNAc); (9) *N*-acetyl-α-D-galactosamine (GalNAc); (10) *N*-acetyl-β-D-galactosamine (GalNAc); (11) α-L-fucose (Fuc); (12) *N*-acetyl-α-D-fucosamine (FucNAc); (13) α-D-glucuronic acid (GlcA); (14) α-D-galacturonic acid (GalA); (15) β-D-glucuronic acid (GlcA); (16) β-D-galacturonic acid (GalA); (17) α-L-arabinose (Ara); (18) β-L-arabinose (Ara); (19) α-D-xylose (Xyl); (20) β-D-xylose (Xyl); (21) α-L-iduronic acid (IdoA); (22) α-L-rhamnose (Rha); (23) 3,6-anhydro-α-D-galactose; (24) 3,6-anhydro-α-L-galactose; (25) β-D-fructose (Fru); (26) *N*-acetyl-β-L-talosaminuronic acid; (27) α-L-galactose; (28) β-D-galactofuranose (Gal*f*); (29) 3-deoxy-β-D-*manno*-oct-2-ulopyrasonic acid (Kdo); (30) β-L-arabinose (Ara); (31) β-D-mannuronic acid (ManA); (32) α-L-guluronic acid (GulA); (33) *N*-acetyl-α-neuraminic acid (NeuAc or sialic acid, SA). Compounds 17, 18, 25, and 28 are furanoses, the rest being pyranoses. Compounds 1–6, 25, and 27–28 are hexoses, 7 an aminohexose, 8–10 *N*-acetamidohexoses, 11, 12, and 22 deoxyhexoses, 13–16, 21, and 31, 32 hexuronic acids, and 17–20 and 30 pentoses. Compounds 2, 4–6, 8, 10–11, 14–15, 17, 20, 22, and 33 are relatively common, whereas compounds 7, 12, 16, 18, 23–24, 26 – 28, and 29–32 are rarely found, the remainder occurring occasionally. Other monosaccharide residues (not shown) are found very rarely, with microorganisms in particular producing a wide variety of rare monosaccharide units. Compounds are also found as derivatives with some or all of sulfate, phosphate, or acetyl groups or as their methyl ethers.

as optical rotation and specific gravity. Enzymatic methods of analysis are becoming more commonly used owing to their specificity and stoichiometry. Monosaccharides derived from more complex carbohydrates by dilute acid hydrolysis may be identified and quantified using gas–liquid chromatography (GC), high-performance liquid chromatography (HPLC), or high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) by comparison of their elution positions and detector response, respectively, with known standards (Sect. 2.2).

1.3 Oligosaccharides

Oligosaccharides containing between 2 and about 30 sugar residues may be analyzed by means of HPLC or, more commonly, HPAEC-PAD (Sect. 2.2). Known carbohydrate oligomers are usually quantified by comparison with standards. The structures of unknown oligosaccharides may be determined by means of the complementary techniques of mass spectrometry (MS, Sect. 2.3.4) and nuclear magnetic resonance (NMR, Sect. 2.3.5).

1.4 Polysaccharides

Polysaccharides are high molecular weight polymers (>5 kDa) of monosaccharides. They may be linear or branched and may contain just one (homopolysaccharides) or several (heteropolysaccharides) types of monosaccharide. Most polysaccharides have very high molecular weight (≫100 kDa) and possess repeating units of relatively uniform structure. It is this structural unit, together with the molecular weight and/or molecular weight distribution, which is generally sought, rather than an accurate picture of a complete molecule.

The type of polysaccharide may be concluded from the component analysis obtained after acid hydrolysis. The glycosidic linkage positions may be determined by means of methylation analysis (Sect. 2.4.3) or NMR spectroscopy (Sect.2.4.5), and the anomeric configuration of these links may be found by means of their susceptibility to hydrolysis using various specific glycosidases (Sect.2.4.2).

1.5 Proteoglycans

Proteoglycans consist of a number of highly charged polyanionic glycosaminoglycan chains attached to a central protein core via serine or threonine residues. These glycans are distinguished from those occurring in glycoproteins in that they are high molecular weight and essentially unbranched chains of repeating disaccharide units containing carboxyl groups and/or substituted with sulfate ester groups. Analysis of proteoglycans involves the determination of the protein's structure and the positions and local configurations at the linkage regions in addition to the structural determination of the attached glycans.

Glycosaminoglycan chain fragments may be cleaved from the proteoglycans by β-elimination with alkali or enzymic hydrolysis (Sect. 2.4.2) and purified by high-performance anion-exchange chromatography (HPAEC). They may then be analyzed by, for example, ^{13}C-NMR spectroscopy. Although the class of the glycosaminoglycans can be relatively easily deduced on the basis of a component analysis and chemical and enzymatic degradation, the discovery of their

detailed fine structure is much more arduous.

1.6 Glycoproteins

Generally, the most important carbohydrates to cell and molecular biologists are the glycoproteins, where carbohydrate often confers the specificity inherent in many molecular and cellular interactions. These compounds possess diverse structures in which carbohydrate moieties, known as glycans, are covalently attached to proteins. A number of different carbohydrates may be attached in different ways, with specific but varying anomeric configurations, linkage positions, and locations within the sequence. All these parameters must be determined if a complete analysis of the glycoprotein is to be achieved.

Usually the glycans are of relatively low molecular weight, containing between 1 and 20 monosaccharides of between 1 and 5 different types. They often involve complex, highly branched structures (Fig. 2). The same type of monosaccharide may be connected in different ways even within a single glycan group. It is frequently found that several different types of glycan are attached to one protein molecule. Since incomplete glycan biosynthesis or subsequent partial hydrolysis commonly results in glycan heterogeneity, and since such differences in structure often do not significantly influence the chromatographic behavior of the intact glycoprotein, a number of different, but related, glycans may be present and coupled to the same amino acid residue, even in "purified" glycoprotein preparations. The analysis of such structures is a serious undertaking for which there are ground rules but no generally applicable, single analytical technique. Methods may vary between laboratories even for similar glycoproteins. There are, however, stages through which any complete analysis must pass.

1. *Isolation and purification.* This involves a number of chromatographic steps such as fast protein liquid chromatography (FPLC), HPLC, gel exclusion chromatography, and lectin affinity chromatography (Sect. 2.5).

2. *Determination of the constituent carbohydrates.* Monosaccharides may be liberated from the glycoproteins by mild acid hydrolysis (e.g. 2 M HCl, 4 h, 100 °C). Once liberated, and after any necessary derivatization, they are separated and determined in one step by either GC or HPLC. This component analysis usually indicates the type of glycans, hence the likely glycan–protein linkages involved.

3. *Cleavage of the intact glycans from the protein moieties.* The different families of glycoprotein glycans require different methods for their cleavage (Fig. 2). As an alternative to the use of enzymes, automated chemical cleavage using hydrazinolysis releases either O-linked or O- and N-linked intact oligosaccharides (Fig. 2). The glycosylation sites may be determined from glycoprotein peptide mapping. The cleaved glycans are often labeled with a radioactive tracer (e.g. by having their reducing ends reduced with sodium borotritide) to make it easier to follow them through their purification.

4. *Glycan isolation and purification.* A combination of size-exclusion HPLC and HPAEC with pulsed amperometric detection (HPAEC-PAD) is commonly used for the purification of glycans. Lectin affinity chromatography is a powerful technique used in their separation.

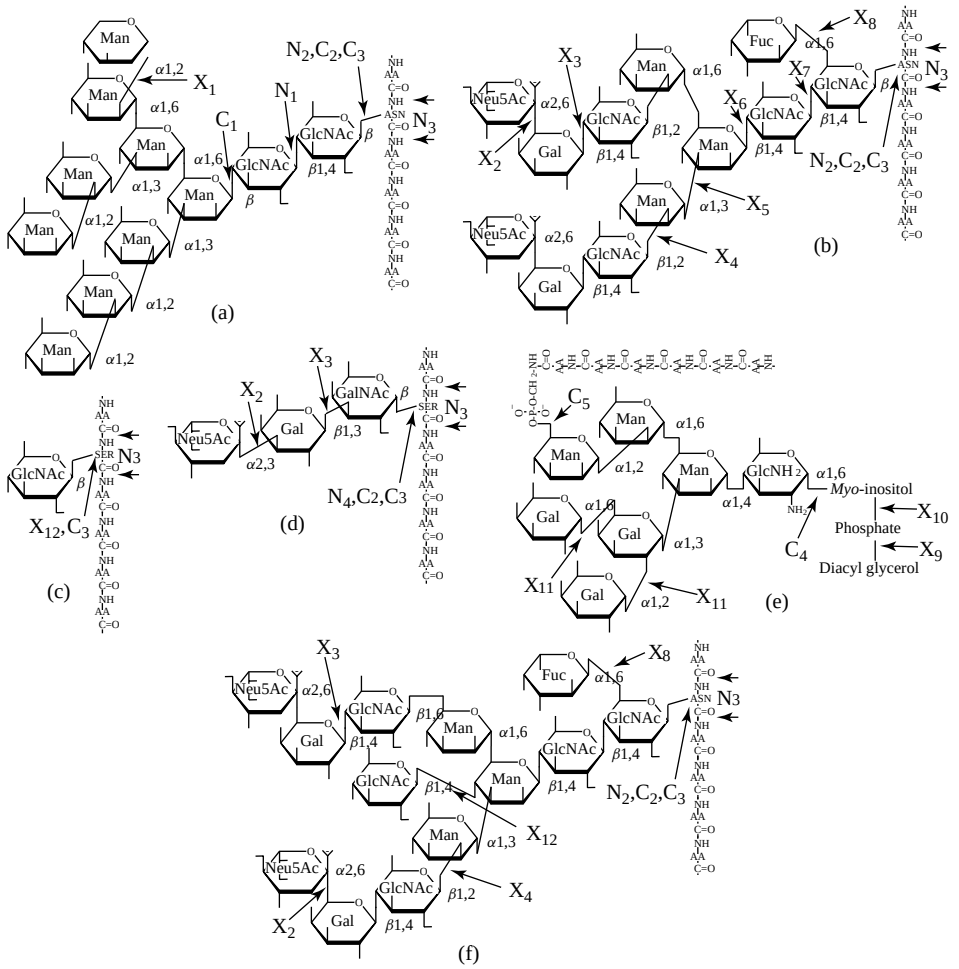

Fig. 2 Examples of the major classes of glycoprotein glycans, showing some methods for their cleavage: (a) *N*-linked high-mannose triantennary, (b) *N*-linked complex biantennary, (c) *O*-linked *N*-acetylglucosamine, (d) *O*-linked complex, (e) glycosylphosphatidylinositol anchor, (f) *N*-linked bisected biantennary. The cleavages may be achieved by the exoglycosidases: X_1, α-mannosidase; X_2, neuraminidase; X_3, β-galactosidase after X_2; X_4, β-*N*-acetylglucosaminidase after X_3; X_5, α-mannosidase after X_4; X_6, β-mannosidase after X_5; X_7, *N*-acetylglucosaminidase after X_6; X_8, α-fucosidase; X_9, phospholipase C; X_{10}, GPI-phospholipase D; X_{11}, α-galactosidase; X_{12}, β-*N*-acetylglucosaminidase; the endoglycosidases: N_1, endoglycosidase H; N_2, peptide *N*-glycosidase F; N_3, pronase; N_4, *O*-glycosidase after X_2; and the chemical methods: C_1, trifluoroacetolysis; C_2, hydrazinolysis; C_3, alkaline sodium borohydride; C_4, nitrous acid; C_5, cold hydrofluoric acid.

5. *Structural studies on the intact glycan.* Much analytical information may be obtained by the use of NMR techniques and mass spectrometry (MS) of the intact glycopeptides. Although it is a powerful tool for determining the presence, sequence, and arrangement of carbohydrates in a glycan chain of up to about 40 residues, mass spectrometry alone is generally not able to distinguish the anomeric type or the linkage positions. NMR, on the other hand, provides a complementary tool that less easily determines sequences but is able to provide detailed information on the anomeric type of the linkages present and the linkage positions. NMR is also often the only experimental technique that can provide the three-dimensional structure. Lectin binding is a powerful, yet simple, technique for discriminating between glycan types.

6. *Complete glycan analysis.* To be sure of any conclusions drawn from spectroscopic studies on an intact glycan, it is necessary to know the complete component composition of all such glycans. In addition, it is usual to confirm the sequence of carbohydrates and their anomeric linkages through the use of specific endo- and exoglycosidases (see Fig. 2 for examples). The linkage positions of the glycan components are still often determined or confirmed by means of methylation analysis.

1.7 Glycolipids

Glycolipids form a diverse group of biological molecules that contain glycans covalently bound to a wide variety of lipid molecules. The intact molecules are generally separated by HPLC. The purified glycolipid molecules may be structurally analyzed by a combination of NMR and MS or by other techniques similar to those applicable to glycoprotein glycans.

2 Key Methods

2.1 The Use of Enzymes in Analysis

Enzymatic methods of analysis, which use the specificity of enzymes to selectively pick out substances from mixtures, are ideally placed for the analysis of complex biological mixtures. There are two ways in which enzymes are generally used:

1. Direct determination of substrates by conversion to products that may be quantified.
2. Structural determination by the ability of enzymes to specifically cleave carbohydrate moieties in di-, oligo-, and polysaccharides and glycoconjugates.

A most important quantitative method is that for glucose. This is not only because glucose is a common carbohydrate analyte but also because it is often hydrolytically produced from other carbohydrates, in their assays. Analysis is by the completely specific conversion of the glucose to 6-phosphogluconolactone with the concomitant conversion of the coenzyme $NADP^+$ to its reduced form NADPH, which absorbs light of 340 nm wavelength. If care is taken to ensure the purity of the enzyme(s), such assays are generally very reliable methods, exhibiting good sensitivity, accuracy, linearity, and precision. All currently accepted assays make use of endpoint determinations; quantification being determined on the basis of a change in color or ultraviolet absorption before and

Tab. 2 Sources and specificity of some important enzymes used in carbohydrate analysis.

Enzyme commission (EC) number	*Common name*	*Usual source*	*Useful specificity*[a]
Oxidoreductases			*Oxidation of*
1.1.1.49	Glucose-6-phosphate dehydrogenase	*Saccharomyces cerevisiae*	Glucose 6-phosphate
1.1.3.4	Glucose oxidase	*Aspergillus niger*	Glucose
1.1.3.9	Galactose oxidase	*Dactylium dendroides*	Terminal galactose residues in saccharides and glycoconjugates
Kinases			*Phosphorylation of*
2.7.1.1	Hexokinase	*S. cerevisiae*	Hexoses
Hydrolases			*Hydrolysis of*
3.2.1.1	α-Amylase	*Bacillus amyloliquefaciens*	α-1,4-Glucans
3.2.1.3	Amyloglucosidase	*A. niger*	Internal α-linked glucose residues
3.2.1.4	Cellulase	*Trichoderma viride*	β-1,4-Glucans
3.2.1.14	Chitinase	*Streptomyces sp.*	β-1,4-*N*-Acetylglucosamine links in chitin
3.2.1.17	Lysozyme	Hen egg white	GlcNAc(β1 $\rightarrow$ 4)-*N*-acetylmuramic acids in peptidoglycans
3.2.1.18	Neuraminidase	*Clostridium perfringens*	Terminal α2-linked *N*-acetyl/*N*-glycolyl neuraminic acids
3.2.1.18	Neuraminidase	*Vibrio cholerae*	Terminal α2-linked *N*-acetyl/*O*-acetyl neuraminic acids
3.2.1.20	α-Glucosidase	*S. cerevisiae*	Terminal α-linked glucose in di- and trisaccharides
3.2.1.21	β-Glucosidase	Sweet almonds	β-Linked hexose in disaccharides
3.2.1.22	α-Galactosidase	Green coffee beans	Terminal α-linked galactose in glycans
3.2.1.23	β-Galactosidase	*Escherichia coli*	β-Linked galactose in disaccharides
3.2.1.23	β-Galactosidase	Bovine testes	Terminal β-linked galactose in glycans
3.2.1.24	α-Mannosidase	Jack bean	Terminal α-linked mannose in glycans
3.2.1.25	β-Mannosidase	*A. niger*	Terminal β-1,4-linked mannose in glycans
3.2.1.26	Invertase	*S. cerevisiae*	Terminal β-linked fructose in oligosaccharides
3.2.1.30	*N*-Acetyl-β-glucosaminidase	*Diplococcus pneumoniae*	Terminal β-linked GlcNAc in glycans
3.2.1.31	β-Glucuronidase	*E. coli*	Terminal β-linked glucuronic acid in polysaccharides
3.2.1.35	Hyaluronidase	Ovine testes	GlcNAc(β1 $\rightarrow$ 4)-glucuronic acids in hyaluronates

Tab. 2 (*continued*)

Enzyme commission (EC) number	*Common name*	*Usual source*	*Useful specificity*[a]
3.2.1.41	Pullulanase	*A. aerogenes*	Internal Glc(α1 $\rightarrow$ 6)-Glc in predominantly α1 $\rightarrow$ 4 glucans
3.2.1.45	Ceramide glycanase	Leech	Glucosyl β-1,4-linkages in neutral glycosylceramides
3.2.1.51	α-Fucosidase	Beef kidney	Terminal α-linked fucose in glycans
3.2.1.96	Endoglycosidase D	*D. pneumoniae*	Chitobiose in high-mannose (Man_5) glycoproteins
3.2.1.96	Endoglycosidase F	*Flavobacterium meningosepticum*	Chitobiose in high-mannose, hybrid, and biantennary glycoproteins
3.2.1.96	Endoglycosidase H	*Streptomyces plicatus*	Chitobiose in high-mannose and hybrid glycoproteins
3.2.1.97	*O*-Glycosidase	*D. pneumoniae*	Linkage of Gal(β1 $\rightarrow$ 3)GalNAc to Ser/Thr in glycoproteins
3.2.1.103	Endo-β-galactosidase	*Bacteroides fragilis*	Internal galactose β1,4-linked to GlcNAc in polylactosamines
3.2.2.18	*N*-Glycosidase F	*F. meningosepticum*	Linkage of *N*-glycans to Asn in glycoproteins
Lyases			*Cleavage of*
4.2.2.4	Chondroitinase	*Proteus vulgaris*	Chondroitin sulfate/dermatan sulfate from proteoglycans
Isomerases			*Isomerization of*
5.3.1.9	Phosphoglucose isomerase	*S. cerevisiae*	Glucose-6-phosphate

[a] For abbreviations, see Fig. 1.

after the material has totally reacted. To produce the change in absorption or to pull the reaction in the required direction, several coupled enzymatic reactions are often required.

Enzymes are also used in *biosensors*—analytical devices that convert biological responses directly to electrical signals. A number of such devices are commercially available. The most widespread is the glucose biosensor, which utilizes the reaction of glucose oxidase at the electrode surface. The other major use of enzymes is in the sequence and linkage analysis of polysaccharides and glycoconjugates (Table 2 and Fig. 2).

2.2 Chromatography

Given the wide variety of carbohydrates, it is not surprising that there exists a wide diversity of chromatographic methods for the separation of carbohydrates (Table 3). They generally deliver quantitative data where standards are available. Positive confirmation of the identity of chromatographic peaks (e.g. by MS) may be needed when the analysis involves complex mixtures or if standards are not available. HPLC and HPAEC are the most common methods of chromatographic analysis, with GC still used sometimes for

Tab. 3 HPLC, HPAEC, and GC chromatographic methods for carbohydrate analysis. There are many commercially available columns that will perform similar functions. Choice of appropriate column for a given separation is often due as much to personal experience, cost, and availability of materials as choice of separation chemistry.

Column type HPLC	*Mobile phase*	*Mechanism*	*Analysis*
Quaternary ammonium bonded to PS-DVB, OH^- form	Aqueous	Anion exchange	Monosaccharides, sialic acids, uronic acids, glycans
Anion exchange (aminopropylsilane-bonded silica, OH^- form)	Acetonitrile/water	Hydrophobic and polar	Mono-, di-, and oligosaccharides
Cation exchange (PS-DVB sulfonate, H^+ form)	Acetonitrile/water or aqueous	Ion-moderated partition and ligand exchange	Glycoprotein- and proteoglycan-derived monosaccharides
Cation exchange (PS-DVB sulfonate, Ca^{2+}, Pb^{2+}, Ag^+ forms)	Hot water	Ion-moderated partition, size exclusion, ligand exchange	Alditols, mono-, di-, and oligosaccharides
C_{18}-bonded reversed-phase silica and graphitized carbon	Acetonitrile/water	Hydrophobic interactions	Derivatized carbohydrates and oligosaccharides
Gel-permeation resin or derivatized silica	Aqueous	Size exclusion	Oligosaccharides, glycans, glycoconjugates
Quaternary ammonium pellicular resin	Usually alkaline aqueous	Anion exchange	Mono-, di-, and oligosaccharides and glycans
Column type GC[a]	***Liquid phase***	***Derivatization***	***Analysis***
Dimethyl	Nonpolar	Trimethylsilylated	Mono- and disaccharides
Trifluoropropyl/methyl	Slightly polar	Trifluoroacetylated	Methyl glycosides from glycoproteins
Cyanopropyl/phenyl/methyl	Polar	Peracetylated	Hydrolyzed and reduced methylated glycans

PS-DVB, divinylbenzene/polystyrene resin.
[a]Column type (silicone).

monosaccharides. Capillary electrophoresis (CE) is a powerful, if less used, technique used for separating and classifying glycans and glycoconjugates.

In GC, separation is due to the extractive distillation of the volatile derivatives. The normal protocol uses highly efficient, fused-silica, wall-coated open tubular (WCOT) columns containing a silicone stationary liquid phase. Detection is by means of a flame ionization detector or by mass spectrometry. For low molecular

weight sugars, the choice between HPLC and GC is not often clear-cut, but in general, HPLC has been preferred for the separation and analysis of oligosaccharide mixtures, while GC has been used for more complex mixtures of monosaccharides and disaccharides.

A wide variety of columns has been used for the HPLC of carbohydrates. Sulfonated cross-linked styrene-divinylbenzene columns containing metal-loaded cation exchangers (H^+, Ca^{2+}, Pb^{2+}, or Ag^+) used at moderately high temperatures (~85 °C) or amino-bonded silica columns operating at around ambient temperature are both popular. The cation exchangers act by ion-moderated partition, ligand exchange, and size exclusion, elution being roughly in order of decreasing molecular weight. Amino-bonded silica columns, which use an acetonitrile–water mobile phase, separate by hydrophobic and polar interactions and partition between the acetonitrile-rich mobile phase and the water-enriched stationary phase. Underivatized carbohydrate molecules generally do not absorb light of useful wavelengths, so mass detectors, changes in refractive index (RI), or post-column chromophore-producing reactions are used for detection. Columns are short (generally 5–25 cm) and contain 5 μm, or 3 μm usually spherical particles. Silica-based packing materials have a high rigidity and do not swell in any solvent. Their major weakness is their limited stability in aqueous mobile phases at high pH. Polymeric-based packing materials are more compressible, but they are compatible with all mobile phases including the entire aqueous pH range. Both the inorganic- and the polymeric-based materials offer a wide range of separation modes including reversed-phase, ion-exchange, hydrophilic-interaction, hydrophobic-interaction, and size-exclusion chromatography, which is discussed in Sect. 2.3.2.

UV detectors are not commonly used for carbohydrate analysis, though it is possible to monitor sugars directly at wavelengths below 200 nm. However, at such short wavelengths, low sensitivity and background interference levels often make their use impractical. Refractive index (RI) detectors work on the principle of differential refractometry and have, until recently, been the mainstay of HPLC carbohydrate detection. However, RI detectors have low sensitivity, are temperature-dependent, and cannot be used with a gradient, as they are extremely sensitive to changes in mobile-phase concentration. Pulsed amperometric detectors offer much more versatility and sensitivity and are commonly used in tandem with anion-exchange chromatography.

High-performance liquid chromatography, on strongly basic anion-exchange resins (HPAEC), is generally the method of choice for the analytical separation of disaccharide, oligosaccharide, and glycan mixtures. Coupled with pulsed amperometric detection (PAD), HPAEC-PAD offers high resolution at picomole sensitivity. Polyhydroxyl compounds, such as the carbohydrates, are negatively charged at high pH and are separated by anion exchange with the degree of ionization, composition, size, linkage position, and anomeric configuration all affecting retention. Retention times can be manipulated by altering the strength of the sodium hydroxide eluant or by addition of the more strongly eluting acetate anion (Fig. 3).

Pulsed amperometric detection is a very sensitive technique (~1 ng glucose) that makes use of the electrochemical oxidation, at a gold electrode poised at positive potential, of a tiny proportion of the alkaline anionic carbohydrates. The

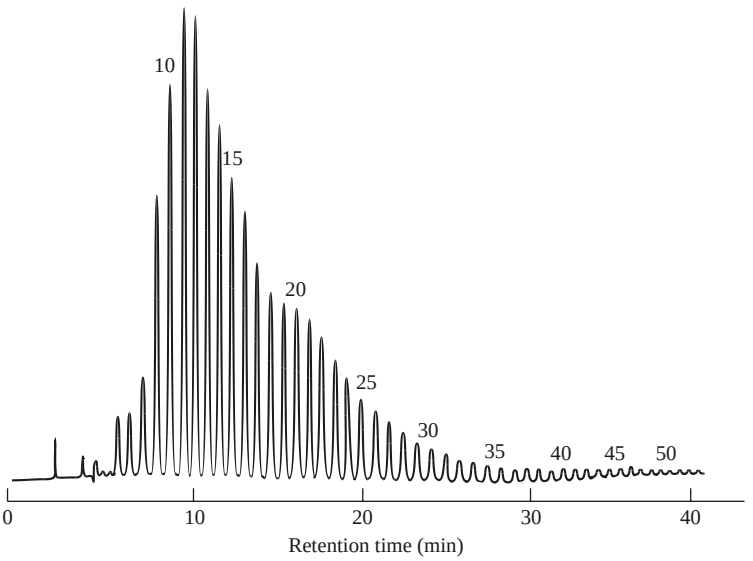

Fig. 3 HPAEC elution profile of the components of debranched wheat amylopectin on a HPIC-AS6 (250 × 4 mm i.d.) column using a gradient of 200 mM sodium acetate at the start, 250 mM at 2 min, 300 mM at 10 min, and 400 mM at 40 min in 150 mM sodium hydroxide solution (1 mL min^{-1}; ambient temperature) and PAD detection. The number on each peak indicates the number of glucose residues. Redrawn from Koizumi et al. (1991) with permission from Elsevier Science.

electrode surface is kept clean by using short pulses of higher positive and negative potentials, thus removing unwanted oxidation products. The advantage of detection with PAD for carbohydrates is that no pre- or post-column derivatization is necessary. Sample preparation is also simplified because only oxidizable analytes will be detected by PAD, and the sensitivity for carbohydrates is orders of magnitude greater than that for the usual contaminant species. HPAEC-PAD has been used, for example, to estimate the chain length of different amylopectins after enzymatic digestion by isoamylase (Fig. 3).

Capillary electrophoresis (CE) is a relatively new technique that provides rapid, highly efficient separations for carbohydrates and possesses several advantages over HPLC owing to its ease of operation, low cost per analysis, and the use of nontoxic chemicals. It provides great resolution by high-voltage (10–20 kV) electrophoresis of mixtures within a polymer-coated fused-silica capillary tube (generally 50 or 70 μm i.d., 375 μm o.d., and between 30 to 100 cm in length). The separation process involves electroendosmotic flow owing to interactions between the charged capillary surface and oppositely charged ions in the buffer. Charged species are separated on the basis of their electrophoretic mobility (charge:mass ratio) and the electroosmotic flow (migration of positive buffer ions to the cathode under the influence of the applied field). Separation can be manipulated by changing either buffer composition or pH. The technique is also applicable to neutral polysaccharides, which can be charged under alkaline

conditions or through complexation with borate or similar ions. Sensitive detection techniques such as pulsed amperometric detection and fluorescence detectors are required owing to the tiny amounts of material separated. Like HPLC, it is suited to the analysis of both polar and nonpolar carbohydrates and it is replacing many of the HPLC techniques. CE offers fast and efficient separation, relatively affordable and durable capillary columns, requires small sample volumes, and has low reagent consumption. It is also a flexible technique available in a number of modes including zone, isoelectric, and micellar electrokinetic. However, it cannot be used for preparative-scale applications.

2.3 Determination of Size and Conformation

Historically, the molecular weight (M) of pure macromolecules, such as glycoproteins, has been determined by ultracentrifugation, using the relationship shown in Eq. (1)

$$\frac{s}{D} = \frac{M(1 - \bar{v}\rho)}{\mathrm{R}T} \tag{1}$$

where both the sedimentation coefficient (s) and the diffusion coefficient (D) are determined under the same conditions. T is the temperature (K), R is the gas constant, $\bar{v}$ is the partial specific volume (the effective volume per unit mass) of the solute, and ρ is the solution density. Today, this methodology has been largely replaced by size-exclusion chromatography, field flow fractionation, and laser light scattering methodology.

2.3.1 Light Scattering

It has been known for many years that light is scattered from molecules in solution in a way that depends on the concentration and molecular size of the molecules present. Recently, the use of multiangle laser light scattering (MALLS) has enabled the accurate nondestructive determination of the molecular weight, molecular weight distribution, and radius of gyration of many high molecular weight carbohydrates without the need for known calibrating standards.

Light scattering from molecules in solution may be described by Eq. (2)

$$\frac{Kc}{R_\theta} = \frac{1}{M_\mathrm{w}}\left(1 + \frac{16\pi^2}{3\lambda^2}\langle r_\mathrm{g}^2\rangle \sin^2(\theta/2)\right) + 2A_2c \tag{2}$$

where

K is an optical parameter given in Eq. (3)

$$K = \frac{4\pi^2 n^2 \left(\frac{\mathrm{d}n}{\mathrm{d}c}\right)^2}{\lambda^2 N_\mathrm{A}} \tag{3}$$

n is the refractive index

$\frac{\mathrm{d}n}{\mathrm{d}c}$ is the RI increment of the solute.

c is the sample concentration, calculated from the differential RI response

$$\mathrm{RI} = K_\mathrm{RI}c\left(\frac{\mathrm{d}n}{\mathrm{d}c}\right) \tag{4}$$

where K_RI is the instrumental calibration constant,

R_θ is the excess intensity of scattered light at angle theta (θ). M_w is the weight average molar mass (molecular weight, Eq. 5)

$$M_\mathrm{w} = \frac{\sum_i n_i M_i^2}{\sum_i n_i M_i} \tag{5}$$

where M_i is the mass of the ith component and n_i is the number of component molecules with that mass.

λ is the wavelength of the scattered light in vacuum.

A_2 is the second virial coefficient, a measure of the solute–solvent interaction that may be important for charged molecules at low ionic strength or where the solution is moderately concentrated.

N_A is Avogadro's number

$\langle r_g^2 \rangle$ is the z-average mean square of the radius of gyration (R_G, which reflects the average displacement of each residue from the molecular centre. Eq. 6)

$$R_G^2 = < r_g^2 > = \frac{\sum_i n_i M_i^2 \langle r_g^2 \rangle_i}{\sum_i n_i M_i^2} \qquad (6)$$

Use of Eq. (2) normally involves extrapolation of theta to zero angle. The molecular weight and radius of gyration are determined by plotting $\frac{Kc}{R_\theta}$ against $\sin^2(\theta/2)$ (Zimm plot) to yield an often almost linear curve whose intercept gives M_w and whose slope at low angles gives the radius of gyration and hence the hydrodynamic volume.

When combined with the separating power of high-performance size-exclusion chromatography (HPSEC), light scattering forms an important tool for the polysaccharide analyst. A typical SEC-MALLS system includes an RI detector, placed after the MALLS detector, for determining the concentration of the eluting polysaccharides.

2.3.2 Size-exclusion Chromatography

Size-exclusion chromatography (SEC) has proved to be a useful technique for separating carbohydrates on the basis of their molecular size. Molecules are eluted in order of decreasing molecular size from a stationary phase containing a wide range of different size pores, allowing greater access and hence slower elution for smaller molecules. SEC has difficulty in determining the absolute molecular size of polysaccharides in which there is a requirement for similar carbohydrates, of known molecular weight, for calibration. In combination with MALLS detection, however, exact values for polysaccharide molecular weight polydispersity can be routinely determined.

The drawback of this technique is that separation of very large molecules is relatively poor due to the difficulty in manufacturing the required high-performance, high-porosity resins. Also, there can sometimes be losses or poor chromatography due to nonspecific absorption. For these reasons, SEC is to some extent being replaced by field flow fractionation.

2.3.3 Field Flow Fractionation

Field flow fractionation (FFF) is a technique for separating polymers on the basis of their size. In combination with MALLS, it is capable of determining molecular weight distributions of complex polysaccharides and delivering far superior resolution to that obtained by SEC-MALLS for high molecular weight materials (>100 000 Da). The separatory principle of the method is the reduction in the diffusion coefficient with increase in molecular size (more exactly, their hydrodynamic size). Molecules to be separated are carried through a thin rectangular channel with at least one of the broader surfaces permeable to small molecules and aqueous solvent (Fig. 4). An introduced cross flow forces the molecules against this permeable surface. Larger molecules with low diffusion constants are left in the slow-moving carrier stream at this surface, whereas molecules with greater diffusion coefficients are able to leave the surface and be carried along more quickly by the fast stream of aqueous solvent further out. Thus, molecules will leave the system in order of their diffusion

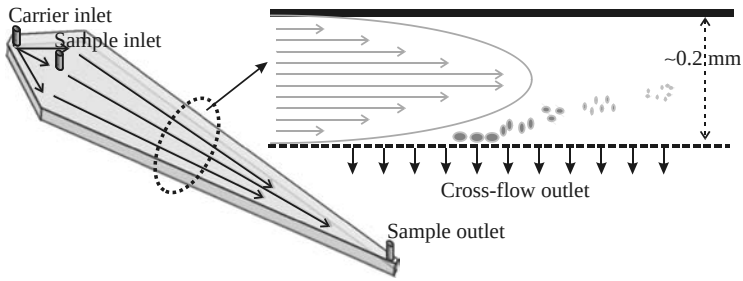

Fig. 4 Outline of asymmetric flow field flow chromatography. Sample (about 100 µL containing 100 µg) is injected and concentrated onto the crossflow membrane. Eluant then carries it along the surface of the crossflow membrane to the sample outlet. Eluant but not sample leaves through the crossflow outlet at a rate determined by the flow rate and the chamber's asymmetry, thus maintaining the sample close to this surface. Molecules possessing greater diffusivity are able to enter the faster flowing stream above the surface and are carried out of the chamber first.

coefficients with those that have higher values (i.e. smaller molecules) exiting first and the largest molecules exiting last. The concentrations of the chromatographed material on exit are usually determined from the RI changes and their absolute molecular weights are determined using MALLS. As the separation only depends on the diffusion coefficient and there is effectively no absorption, this method is capable of giving the absolute hydrodynamic size of molecules plus useful information concerning the molecules' conformations and structures (e.g. the degree of chain branching).

2.3.4 Mass Spectrometry (MS)

Mass spectrometry (MS) is a sensitive and powerful technique for the analysis of glycans. It involves the production of ionized carbohydrate moieties that are in the gas phase. The charged ions move in a magnetic field along curved paths, which are determined by the imposed magnetic field and by the ratio of their mass to their charge (m/z). Usually, singly charged cations are produced most abundantly and may be detected with masses up to several kilodaltons with an accuracy approaching 1 ppm. The key stage of the technique is the ionization step. Sugars are not normally volatile and must be derivatized or otherwise persuaded to enter the gas phase.

In electrospray ionization mass spectrometry (ES-MS), a strong electric field is applied between the end of a capillary tube containing the sample solution and a counter electrode, to disperse the emerging sample into an aerosol of highly charged droplets. As the droplets evaporate, assisted by the flow of warm gas, multiply charged ions emerge free from solvent (Fig. 5). This rapid and sensitive method produces intact molecular ions of even very large molecules and allows the determination of the molecular weight of glycoproteins with an accuracy of 0.01% using as little as 1 pmol of sample. Thus, glycosylation heterogeneity can be accurately quantified. As an example, the ES-MS of the carboxy-terminal half of hen ovotransferrin gives 3 major peaks centered on molecular masses of 40 220, 40 056, and 39 854. This indicates microheterogeneity

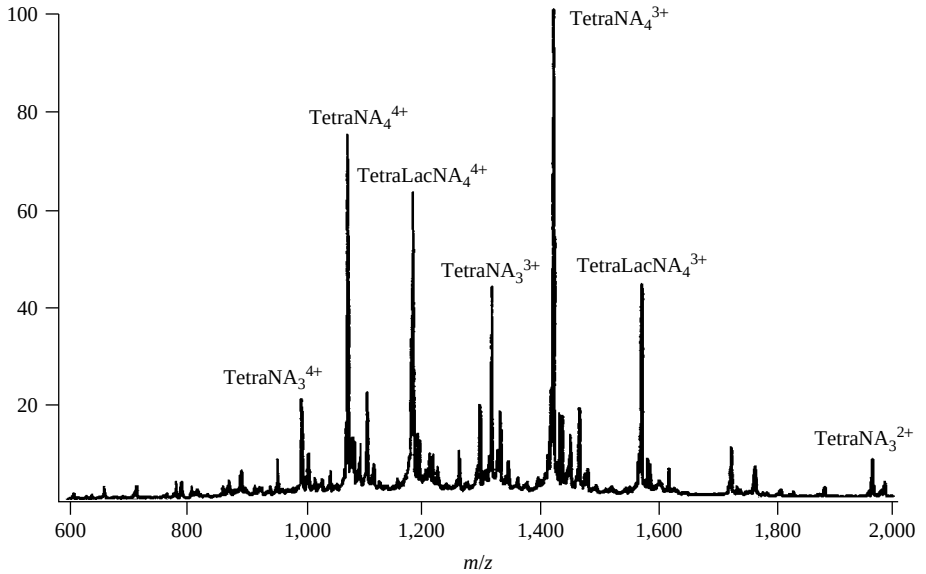

Fig. 5 ES-MS of oxidized, reduced ($NaBD_4$), and methylated glycan from recombinant erythropoietin. ES-MS produces multiply charged ions. This example is centered on triply positively charged ions. Periodate oxidation, which oxidizes *cis*-diols only, of the glycans was used to resolve isomeric structures and discover linkage positions. For example, the most abundant peak ($TetraNA_4$, m/z 1422.4, mass 4267.2 Da) is derived from a tetraantennary glycan containing four terminal neuraminic acid residues and a fucosyl residue (4633.5 Da). The other major species, Tetra NA_3 and $TetraLacNA_4$, are derived from similar glycans but with an *N*-acetylneuraminic acid residue loss and an *N*-acetyllactosamine residue gain respectively. Redrawn from Linsley et al. (1994) with permission from Elsevier Science.

due to the absence of first a hexose and then an *N*-acetylhexosamine residue. ES-MS is ideally suited for direct interfacing to HPLC and capillary electrophoresis. It is far less able to handle buffers and detergents than MALDI-MS, however.

Matrix-assisted laser-desorption ionization mass spectrometry (MALDI-MS) gives results similar to those obtained from ES-MS but can extend the range of detectable molecular ions. In this method, the ionization of dry samples is achieved through the use of a short pulse of intense UV-laser light in the presence of a 100-fold excess of a UV-absorbing organic molecule (known as the matrix, e.g. 2,5-dihydroxybenzoic acid). Even very large molecules can be vaporized and ionized as MALDI is capable of determining molecular weights, in exceptional cases, of up to m/z 300 kDa using time-of-flight (TOF) detection. It requires only femtomoles to picomoles of material, no sample derivation, and tolerates more buffer and detergent than other MS methods. Although able to handle proteins, it has not had similar success, so far, with natural polysaccharides in the upper part of this range (>25 kDa) and seems to work best with oligosaccharides below 5 kDa. This may be partly due to the polydispersity of such samples and partly due to the inherently poor resolution of MALDI-TOF MS.

Mass spectrometry must generally be combined with other analytical methods of analysis as it cannot distinguish between

isomeric sugar residues and cannot easily provide the linkage and stereochemical information of NMR.

2.3.5 NMR Spectrometry

The nuclei of many atoms (e.g. ^{1}H, ^{13}C, ^{15}N, ^{31}P) act as spinning magnets that can orient themselves with or against an external magnetic field. The strength of this imposed magnetic field determines both the frequency at which the nuclei rotate (at high field strengths, the ^{1}H resonance frequencies are 400–600 MHz) and the resolution of the NMR spectra. The very small difference in energy between the orientations of the spin of a nucleus is affected by local variations in the magnetic field due to neighboring nuclei and can be used to differentiate between the nuclei and to provide information on neighboring nuclei and their relative orientation. The differences in the magnetic environment of the nuclei are reported in terms of the relative shift, in parts per million, of the applied magnetic field due to the local molecular magnetic field relative to a known structure. For protons, shifts are usually reported downfield (i.e. higher ppm) from the methyl protons of sodium 3-trimethylsilyl-1-propanesulfonate.

^{1}H-NMR spectroscopy is an extremely powerful and elegant technique for the determination of the structure of glycans, although it is between three and six orders of magnitude less sensitive than mass spectrometry. It can provide information about the number and variety of the component carbohydrate residues, on how they are joined together, and can indicate the presence or absence of noncarbohydrate groups. Taken together with component and methylation analysis, ^{1}H-NMR spectra often allow the unambiguous assignment of the complete molecular structures of purified glycans and the composition of less-purified mixtures of glycans.

The key features of carbohydrate NMR are the downfield shifts owing to the spatial proximity of oxygen atoms. Proton NMR carbohydrate spectra contain a mainly unresolved heap of signals between 3.5 and 4 ppm with relatively few resolved outlying peaks (Fig. 6, top). These outliers are termed *structural reporter groups* and contain generally well-resolved anomeric signals (H-1) with greater shifts (4.4–5.4 ppm). As the anomeric carbon (C-1) is bonded to two oxygen atoms, these anomeric ^{13}C and ^{1}H atoms are easily distinguishable, providing important inroads into the structural determination. Their chemical shifts and coupling constants indicate the sugar residues involved, together with the types and anomeric character (i.e. α or β) of the glycosidic linkages and their ring size (pyranose or furanose).

Other reporter groups also provide useful well-resolved resonances. The methyl protons and carbon (^{13}C) atoms in deoxysugars are upfield (i.e. at lower ppm) and the carboxylate carbon (^{13}C) atoms in uronic acids are downfield. Thus, the position of fucose residues may be confirmed by the shifts of their H-5 (4.0–4.8 ppm) and methyl hydrogen (~1.2 ppm) atoms. Signals due to mannose H-2 atoms (4.1–4.2 ppm) help confirm the type of antennary structure indicated by the anomeric proton shifts. The shifts of both the axial (~1.7 ppm) and equatorial (~2.7 ppm) H-3 atoms in sialic acids are characteristic for the type and configuration of their glycosidic links and may be indicative of their position in the chain. The shifts of the hydrogen atoms in acetyl groups (2.0–2.1 ppm) are very useful in indicating the structure local to *N*-acetylhexosamine and sialic acid residues.

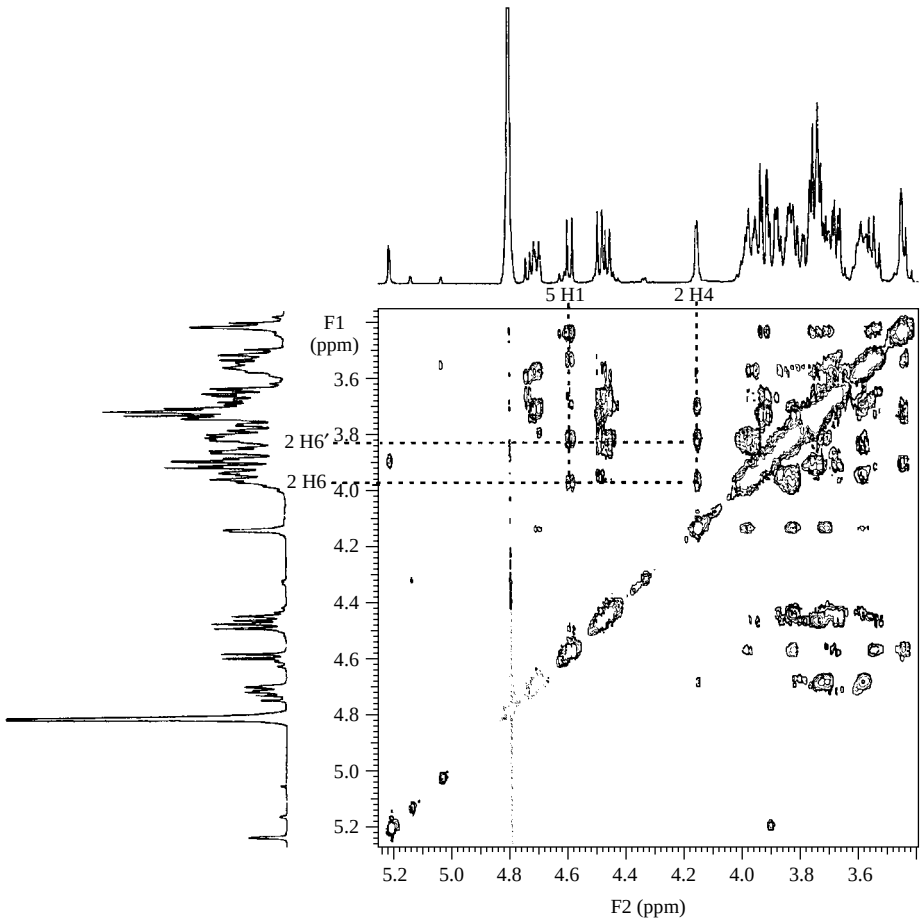

Fig. 6 Two-dimensional ROESY spectrum of a pentasaccharide derived by the action of a rat serum β-(1 $\rightarrow$ 6)-D-*N*-acetylglucosaminyltransferase. A ROESY spectrum gives through-space information concerning the closeness of protons. The diagonal consists of the 1D spectrum also shown to the top and side. The off-diagonal peaks show close atoms (<0.5 nm). The anomeric proton on the newly inserted residue (5H1, GlcNAc) is shown to lie close to the two protons on the 6-position of the second residue (2 H6 and 2 H6′, Gal) that in turn lies close to the 4-proton on this Gal residue (2 H4). Thus, the linkage of this GlcNAc residue to the 6-position on the second galactose residue was confirmed. Redrawn from Maaheimo et al. (1997) with permission from Elsevier Science.

^{13}C-NMR requires substantially larger samples than ^{1}H-NMR both because of the intrinsically lower sensitivity (1.6%) of ^{13}C nuclei relative to ^{1}H and because of the low natural abundance of ^{13}C relative to ^{12}C (1.1%). Advantage is taken of this lower abundance as it also simplifies the spectra inasmuch as ^{13}C–^{13}C coupling is virtually nonexistent. It has also proved popular for the determination of naturally acetyl-derivatized sugar residues.

Polysaccharides, glycoproteins, and large glycans are usually run in D_2O ($^{2}H_2O$). Working in D_2O eliminates the

less-well-defined exchangeable hydroxyl protons and the vast excess (110 M) of H_2O protons. Intramolecular hydrogen bonding may be investigated by dissolving the carbohydrate in per-deuterated dimethyl sulfoxide (DMSO) when the hydroxyl (and amine) protons do not exchange with solvent and give useful information. In some cases, hydrogen bonding may be determined from proton exchange in mixed H_2O, HOD, and D_2O solvents where solvent exposure may be quantified from the differential shifts in proton and carbon spectra.

Although 1D proton NMR spectra can be used as fingerprints to enable structural determination where databased structures exist, they are generally too complex to assign fully even for relatively simple oligosaccharides. ^{13}C spectra are simpler but are also usually too complex to use without the availability of fully assigned NMR spectra from suitable known structures for guidance. The structures of unknown carbohydrates generally demand a variety of through-bond (COSY), through-space (NOESY), and heteronuclear $^{13}C-^{1}H$ 2D NMR techniques using high-field spectrometers (600 MHz) for their elucidation.

Complete assignment of ^{1}H NMR and ^{13}C NMR resonances usually starts with the well-resolved anomeric H-1 or C-1 peak using 1D NMR. The proton resonances can then be assigned sequentially using through-bond techniques. The anomeric conformation may be confirmed from the size of the anomeric $^{1}J_{C-H}$ coupling constant, which is typically 170 Hz for axial hydrogen and 160 Hz if equatorial. Structural conformations may be given by the $^{3}J_{H-H}$ coupling constants, which obey a Karplus relationship (Eq. 7).

$$^{3}J = A\cos^2(\theta) - B\cos(\theta) + C \qquad (7)$$

where A, B and C are fitted parameters derived from a number of known trainer molecules. The cosine nature of Karplus curves means that it gives one to four possible angles every 360° for a given ^{3}J (Fig. 7). Simple molecular modeling can usually make a choice between these, as often only one conformation is structurally reasonable. For example, the

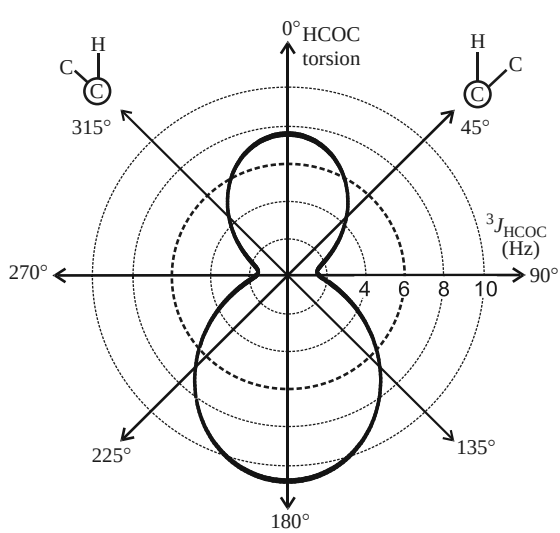

Fig. 7 The Karplus relationship $^{3}J = A\cos^2(\theta) - B\cos(\theta) + C$. The coupling constants involving the glycosidic linking torsions ($^{3}J_{COCH}$)A, B, and C are 7.6, 1.7 and 1.6 Hz respectively. It can be noted that for any experimental value of ^{3}J, there may be up to four possible torsional angles (e.g. a $^{3}J_{COCH}$ of 6 Hz may be derived from a torsion of 28°, 131°, 229°, or 332°).

anomeric configuration of glucopyranose and galactopyranose are easily observed, from the $^3J_{H1-H2}$ coupling constant, as the H-1 and H-2 protons of the beta anomers are trans-diaxial (~180°) giving large couplings, whereas those in alpha anomers are gauche (~60°) with small couplings.

2.3.6 Molecular Modeling

Molecular modeling techniques treat structures as mechanical entities in which bond lengths, angles, and torsions are treated as springs requiring forces to move them from their preferred values. Together with a consideration of atomic charges and van der Waals interactions, the structure of the molecule is optimized, giving preferred conformations. Various molecular dynamics and random walk molecular mechanics (RAMM) methods exist to ensure that the molecule does not settle into biologically irrelevant local potential energy minima.

The additional information available from such molecular modeling is particularly useful for carbohydrates, when compared to proteins, as carbohydrates generally do not form crystal structures (if they form them at all) with similar conformations to their conformation in freely dissolved solution. The information is also useful for clarifying NMR structural analyses where the coupling constants may not be measurable and where there are mixed conformational populations present.

2.4 Linkage Analysis

2.4.1 Glycan Detection

There are several methods available for detecting the presence of glycans attached to proteins (glycoproteins) and lipids (glycolipids). One way is to oxidize all the pairs of adjacent hydroxyl groups (i.e. *cis*-diols) in the carbohydrate to aldehyde groups, using periodate, followed by covalent coupling to hydrazide derivatives of biotin or the steroid digoxigenin. The respective labeled glycoconjugates may be detected by an enzyme immunoassay using streptavidin or a digoxigenin-specific antibody conjugated to alkaline phosphatase. This procedure may be applied *in situ* for the detection of nanogram quantities of glycoproteins, previously separated by sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS-PAGE) and blotted onto nitrocellulose membranes. More specific oxidation methods, such as the use of galactose oxidase or prior specific enzyme hydrolysis (Sect. 2.4.2), may allow the detection of glycan subclasses.

2.4.2 Glycan Release and the Use of Enzymes

The carbohydrate structures in glycoproteins are normally investigated after they are released from the protein and chromatographically separated from each other. There are a number of methods by which these glycans may be released; some utilize specific enzymes (Table 2), whilst others use chemical methods (Fig. 2).

2.4.3 Methylation Analysis

Methylation analysis is a powerful and widely used method for the determination of the glycosidic linkage positions in the structural analysis of glycoprotein, as well as glycolipids, glycans, and polysaccharides. It involves replacing all the free hydroxyl groups with methyl ether groups (Fig. 8). The methylated polysaccharide is then hydrolyzed, converted to alditols, acetylated, and analyzed by GC or GC-MS.

The carbohydrate is initially dissolved in dry DMSO. In the presence of a very strong base, such as sodium hydride or potassium *tert*-butoxide, and the absence

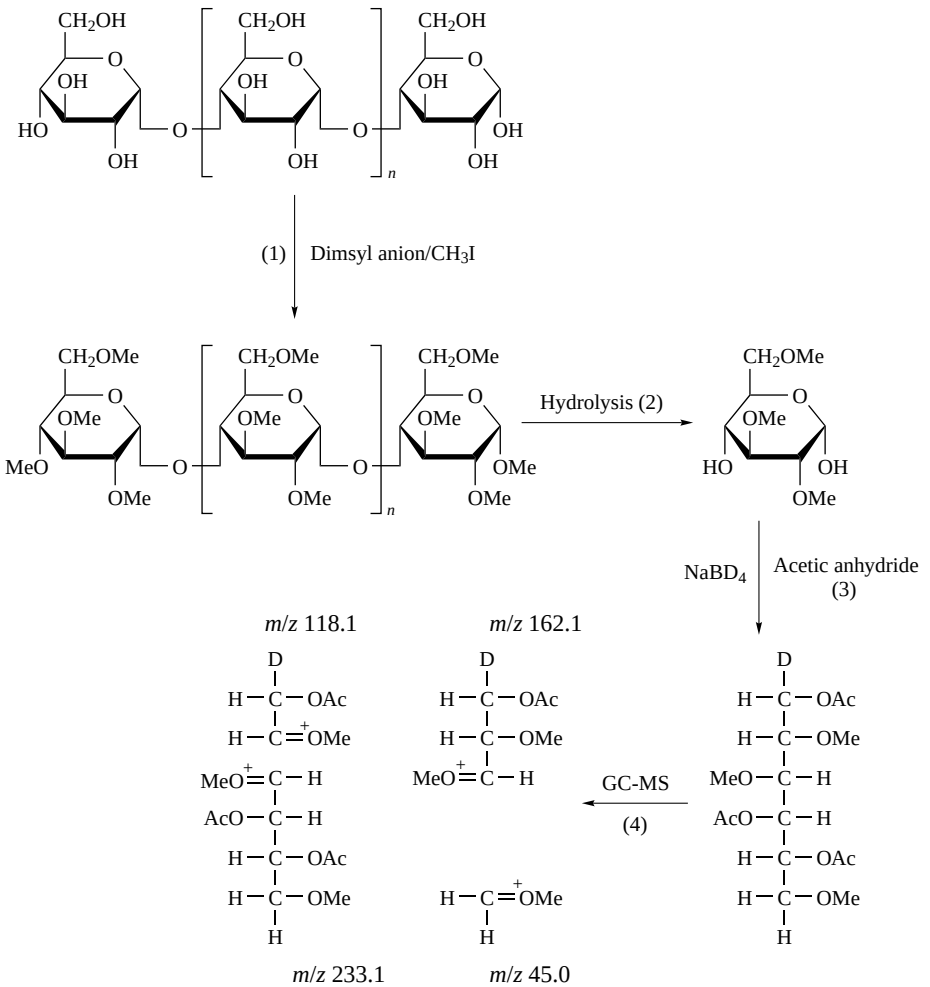

Fig. 8 Methylation analysis. In this technique all the free hydroxyl groups in the intact glycan are methylated (1). The fully methylated glycan is then hydrolyzed (2), reduced, and acetylated (3). The linkage positions may be deduced from the positions of the methyl and acetyl groups in the resultant partially methylated alditol acetates, as determined using GC-MS. The ring hydrogen atoms are omitted for the sake of clarity. The breakdown of such compounds in EI-MS occurs primarily between neighboring methyl ethers. Secondary fragmentation then occurs by the elimination of the neutral molecules acetic acid (CH_3COOH, −60 Da), methanol (CH_3OH, −32 Da), ketene (CH_2CO, −42 Da), and formaldehyde (CH_2O, −30 Da) from these primary fragments. Thus, the *m/z* 162 fragment gives rise to an *m/z* 102 peak. The deuterium atoms in these structures distinguish the C-1 atoms from the C-6 atoms.

of water, the DMSO is converted into the dimsyl anion, which is then able to convert all the free hydroxyl, carboxyl, and amino groups to their anions. Methylation of the polyanion is then carried out using methyl iodide. At the end of the reaction, the methylated product is hydrolyzed with acid to its component partially methylated

residues. These have free hydroxyl groups in the positions where they were linked to each other and methyl ether groups where the original carbohydrate had free hydroxyl groups. The products are further derivatized to render them volatile and are analyzed qualitatively and quantitatively by GC coupled to electron impact ionization spectroscopy (EI-MS). This MS produces very energetic ions that fragment readily, nearly always resulting in the loss of an intact molecular ion. It is the breakdown pattern rather than the molecular ion that enables us to distinguish between the methylated carbohydrate residues (Fig. 8). The positions of the acetyl groups indicate where the unmethylated free hydroxyl groups were in the partially methylated hydrolyzed sugars, thus showing where the residue was substituted. However, it gives no information on either the sequence of these residues or their anomeric conformation.

Naturally methylated or acetylated glycans may be analyzed by methylation, as just described, but using deuterated methyl iodide or acetic anhydride to distinguish between the natural and chemically derivatized sites by their different masses.

2.4.4 Mass Spectrometry

Mass spectrometry is capable of providing sequence information in addition to its previously described use, determining molecular weight. In FAB-MS, as little as 1 pmol of glycan sample is bombarded with an accelerated beam of neutral atoms (e.g. Xe, Cs) in a vacuum. This causes desorption of the sample from the liquid surface, forming a characteristic molecular ion plus a wealth of structurally diagnostic fragment ions derived from it. The structure of a sample is determined given the known fragmentation patterns of related model structures. Improved spectra can usually be obtained by the use of fully methylated or acetylated carbohydrates. As fragmentation predominantly occurs between glycosidic links, the breakdown pattern may give an abundance of information concerning the glycan's sequence. As an example, the biantennary glycan (Fig. 2b) when fully methylated gives the following m/z peaks:

2492	$(\text{Neu5Ac} \rightarrow \text{Gal} \rightarrow \text{GlcNAc} \rightarrow \text{Man})_2\text{Man} \rightarrow \text{GlcNAc})^+$
1029	$(\text{Neu5Ac} \rightarrow \text{Gal} \rightarrow \text{GlcNAc} \rightarrow \text{Man})^+$
825	$(\text{Neu5Ac} \rightarrow \text{Gal} \rightarrow \text{GlcNAc})^+$
376	$(\text{Neu5Ac})^+$

Although linkage information is produced, the MS is not able, in this case, to differentiate the order of the hexoses or *N*-acetylhexosamines, nor is it able to determine the linkage positions.

Tandem mass spectrometry (MS/MS) utilizes two or more mass-selection devices in series so that the known ion from one analyzer may be selected and induced to give daughter fragments that are determined using a second analyzer. It can provide complete sequence, branching, and linkage information. Even pyranose and furanose ring systems may be distinguished owing to differences in their cleavage patterns. Tandem mass spectrometry can be used with FAB, ESI, or MALDI and the use of post source decay (PSD) in MALDI-TOF or collision-induced dissociation (CID) in ESI has proved particularly useful as the extent and distribution pattern of fragments is highly reproducible.

2.4.5 NMR Spectroscopy

The power of NMR is most useful when it is used to determine the sequence, linkage positions, and conformation of carbohydrates. Linkage and sequence

information is commonly obtained using the off-diagonal peaks apparent in two-dimensional NMR (Fig. 6). Spectra can be obtained in a number of ways, by such techniques as COSY and NOESY after a full assignment of the ^{1}H and ^{13}C spectra. In particular, COSY techniques indicate neighboring connected nuclei, whereas NOESY techniques give indication as to the closeness of groups through space, the effects of both being due to perturbations caused to each other's atom's local magnetic field. Linkages can be determined from the $^{3}J_{CH}$ scalar coupling through and across glycosidic bonds and from interglycosidic ^{1}H NOE effects using NOESY, which gives those protons that are in close proximity (less than about 0.5 nm).

By spreading out the NMR spectra over two or more dimensions, all the spectral peaks may be unambiguously assigned to atoms in the glycan, fully confirming its structure. Such spectra also allow the determination of the positions of noncarbohydrate substituents such as sulfate and phosphate groups.

Sequences may be determined from through-space connectivities. However, with increasing size above about 20 glycosyl residues, it becomes increasingly difficult to delineate branch locations unambiguously. NOE effects are also difficult to interpret in the presence of significant internal molecular motion, owing to positional averaging. Although NMR is difficult to apply to many polysaccharides, the presence of repeating units can be identified from the NMR spectra of oligomers and then quantified in the polymers.

Linkage positions may also be discovered by determination of the position of the free hydroxyl groups, giving information similar to that obtained from methylation analysis (Sect. 2.4.3), but nondestructively. ^{1}H NMR spectra in H_2O (~10% D_2O) and D_2O are compared to identify the position(s) of the hydroxyl groups.

The biological roles of oligosaccharides and polysaccharides are related to their conformations. This mainly depends on the trans-glycosidic torsional angles connecting the carbohydrate units (Fig. 9).

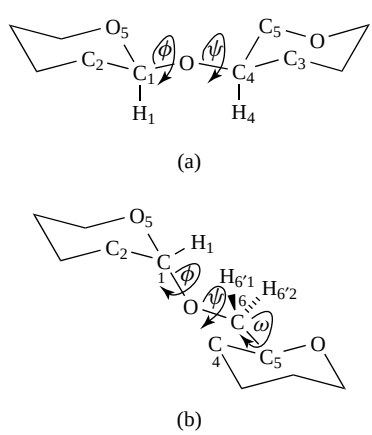

Fig. 9 The torsions and possible coupling constants involved in the anomeric linkages in oligosaccharides and polysaccharides. (a) Torsions ϕ (phi) and ψ (psi) are defined as the clockwise torsions $O_rC_aO_aC_n$, and $C_aO_aC_nC_{n-1}$ respectively, where *r*, *a*, and *n* represent ring, anomeric, and structural number, respectively. The relevant and useful coupling constants are $^{3}J_{C_1OC_4C_3}$ giving ψ and $^{3}J_{C_2C_1OC_4}$ giving ϕ and $^{3}J_{C_1OC_4H_4}$ and $^{3}J_{H_1C_1OC_4}$. (b) Linkages to the terminal carbon atom may involve an exocyclic link; the additional linking torsion ω (omega) is defined as the clockwise torsion $O_aC_nC_{n-1}C_{n-2}$. The relevant coupling constants are $^{3}J_{C_1OC_6C_5}$ giving ψ and $^{3}J_{H_1C_1OC_6}$, $^{3}J_{C_1OC_6H_{6'1}}$, $^{3}J_{C_1OC_6H_{6'2}}$, $^{3}J_{C_2C_1OC_6}$, $^{3}J_{H_{6'1}C_6C_5C_4}$, and $^{3}J_{H_{6'2}C_6C_5C_4}$.

Tab. 4 The specificity of some lectins used for characterizing carbohydrates.

Lectin	*Systematic name*	*Abbreviation*	*Useful specificity*[a]	*Example (a–f designate parts of Fig. 2)*
Orange peel mushroom	*Aleuria aurantia*	AAA	Terminal fucose ($\alpha 1 \rightarrow 6$) GlcNAc	(b), (f)
Love lies bleeding	*Amaranthus caudatus*	ACA	Terminal/internal Gal($\beta 1 \rightarrow$ 3)GalNAc($\alpha 1 \rightarrow$)	(d)
Jack bean (concanavalin A)	*Canavalia ensiformis*	ConA	Terminal/internal mannose/glucose	(a), (b), (f)
Horse gram	*Dolichos biflorus*	DBA	Terminal GalNAc($\alpha 1 \rightarrow$)	Blood type A
Thorn apple	*Datura stramonium*	DSA	Terminal Gal($\beta 1 \rightarrow 4$) GlcNAc	Neuraminidase-treated (b) or (f)
			Terminal GlcNAc($\beta 1 \rightarrow 3$) Ser/Thr	(c)
Common snowdrop	*Galanthus nivalis*	GNA	Terminal Man($\alpha 1 \rightarrow 2/3/6$) Man	(a)
A west African legume	*Griffonia simplicifolia*	GS IA_4, BS IA_4	Terminal *N*-acetyl-α-galactosamine	Blood type A[b]
		GS IB_4	Terminal α-galactose	(e), blood type B[c]
Roman snail	*Helix pomatia*	HPA	Terminal N acetyl-α-galactosamine	Blood type A
Lentil seed	*Lens culinaris*	LCA	Terminal/internal α-mannose/α-glucose	(a), (b), (f)
Winged pea	*Lotus tetragonolobus*	LTA	Terminal fucose($\alpha 1 \rightarrow 6$) GlcNAc	(b), (f)
Maackia tree	*Maackia amurensis*	MAA	Terminal sialic acid ($\alpha 2 \rightarrow 3$) Gal	(d)
Red kidney bean	*Phaseolus vulgaris*	PHA	Terminal Gal($\beta 1 \rightarrow 4$) GlcNAc	Neuraminidase-treated (b)
Peanut	*Arachis hypogaea*	PNA	Terminal Gal($\beta 1 \rightarrow 3$) GalNAc	Neuraminidase-treated (d)
Castor bean	*Ricinus communis*	RCA_1	Terminal Gal($\beta 1 \rightarrow 4$) GlcNAc($\beta 1 \rightarrow$)	Neuraminidase-treated (b)
		RCA_2	Terminal Gal($\beta 1 \rightarrow 3$) GalNAc($\alpha 1 \rightarrow 3$) Ser/Thr	Neuraminidase-treated (d)
Elderberry	*Sambucus nigra*	SNA	Terminal sialic acid($\alpha 2 \rightarrow 6$) Gal/GalNAc	(b), (f)

Tab. 4 (*continued*)

Lectin	*Systematic name*	*Abbreviation*	*Useful specificity*[a]	*Example (a–f designate parts of Fig. 2)*
Potato	*Solanum tuberosum*	STA	GlcNAc(β1 → 4) GlcNAc(β1 → 4) GlcNAc(β1 → 4)	
Gorse seed	*Ulex europaeus*	UEA-I	Terminal fucose(α1 →)	(b), blood type H (O)[d]
Wheat germ	*Triticum vulgaris*	WGA	Internal GlcNAc(β1 → 4) Man(β1 → 4)	(f)

[a] For abbreviations, see Fig. 1.

[b] Fuc ↓ α1, 2
GalNAc(α1 → 3)Gal(β1 → 3 or 4)GlcNAc(β1 → R)

[c] Fuc ↓ α1, 2
Gal(α1 → 3)Gal(β1 → 3 or 4)GlcNAc(β1 → R)

[d] Fuc ↓ α1, 2
Gal(β1 → 3 or 4)GlcNAc(β1 → R)

Conformational analysis requires independent data for each independent torsion angle (ϕ, ψ, and sometimes ω) and, where mixtures of conformations exist, the relative population occupancy of each. These may be available from interresidue proton NOE (e.g. between the anomeric proton and the proton on the linkage carbon, H–C_1–O–C_n–H) and the $^3J_{COCH}$, $^3J_{COCC}$, $^3J_{CCCH}$, and $^2J_{COC}$ coupling constants, but the cosine nature of the Karplus equation for ^3J coupling means modeling may be required. Whilst the conformation of rigid oligosaccharides can be directly derived from NOESY data, the upper and lower bounds of the possible distances from NOESY often enclose a large number of ϕ and ψ combinations. As multiple conformations are often in fast equilibrium on the NMR timescale and NMR data is effectively time-averaged, unrealistic putative torsional angles may arise and must be discriminated using molecular modeling (Sect. 2.3.6).

Although NMR is an extremely powerful tool, it has not yet been proved possible to use it on intact glycoproteins because of the complexity of the multitude of overlapping peaks in the spectra. A further limitation of NMR is that it requires samples of at least about 25 nmol (i.e. milligram amounts).

2.5 Use of Lectins

Lectins are widespread and serve in many biologically significant activities, such as agglutinating cells and precipitating glycoproteins and polysaccharides. Because of this, many lectins are extremely toxic. The carbohydrate binding sites are large and able to distinguish complex carbohydrate structures (Table 4).

For this reason, lectins are used in blood banks to identify blood groups. In some cases, lectin binding is able to discriminate between structures that are otherwise difficult to distinguish; examples include α-2-6- and α-2-3-linked terminal sialic acids.

Lectins may be immobilized on Sepharose columns and used to separate, purify, and type glycoconjugates by a process of affinity chromatography. Submicrogram amounts of glycoconjugates may be differentiated by means of their specific binding to lectins that have been covalently labeled with biotin or digoxigenin. A range of lectins may be used; those bound are detected by enzyme immunoassay using streptavidin or a digoxigenin-specific antibody, respectively, conjugated to alkaline phosphatase in a similar manner to that described in Sect. 2.4.1.

See also Bioorganic Chemistry.

Bibliography

Books and Reviews

Chaplin, M.F., Kennedy, J.F. (Eds.) (1994) *Carbohydrate Analysis: A Practical Approach*, 2nd edition, Oxford University Press, London, New York (also on CD-ROM).

El Rassi, Z. (2002) *Carbohydrate Analysis by Modern Chromatography and Electrophoresis*, 2nd edition, Elsevier Health Sciences, Amsterdam.

Montreuuil, J., Vliegenthart, J.F.G., Schachter, H. (1995) Glycoproteins, *New Comprehensive Biochemistry*, Vol. 29a, Elsevier, Amsterdam.

Montreuuil, J., Vliegenthart, J.F.G., Schachter, H. (1997) Glycoproteins II, *New Comprehensive Biochemistry*, Vol. 29b, Elsevier, Amsterdam.

Singh, R.S., Tiwory, A.K., Kennedy, J.F. (1999) Lectins: sources, activities and applications, *Crit. Rev. Biotechnol.* **19**, 145–178.

White, C.A. (Ed.) (1996) *Advances in Carbohydrate Analysis*, JAI Press, London.

Primary Literature

Arentoft, A.M., Michaelson, S., Sorenson, H. (1993) Determination of oligosaccharides by capillary zone electrophoresis, *J. Chromatogr., A.* **652**, 517–524.

Bello-Pérez, L.A., Roger, P., Baud, B., Colonna, P. (1998) Macromolecular features of starches determined by aqueous high-performance size exclusion chromatography, *J. Cereal Sci.* **27**, 267–278.

Bendiak, B. (1999) Nuclear magnetic resonance spectroscopy of peracetylated oligosaccharides having ^{13}C-labeled carboyl groups in lieu of permethylation analysis for establishing linkage substitutions of sugars, *Carbohydr. Res.* **315**, 206–221.

Benincasa, M.-A., Cartoni, G., Fratte, C.D. (2002) Flow field-flow fractionation and characterization of ionic and neutral polysaccharides of vegetable and microbial origin, *J. Chromatogr., A* **967**, 219–234.

Beri, R.G., Walker, J., Reese, E.T., Rollings, J.E. (1993) Characterization of chitosans via coupled size-exclusion chromatography and multiple-angle laser light-scattering technique, *Carbohydr. Res.* **238**, 11–26.

Chaplin, M.F. (1982) A rapid and sensitive method for the analysis of carbohydrate components in glycoproteins using gas-liquid chromatography, *Anal. Biochem.* **123**, 336–341.

Charlwood, J., Bryant, D., Skehel, J.M., Camilleri, P. (2001) Analysis of N-linked oligosaccharides: progress towards the characterisation of glycoprotein-linked carbohydrates, *Biomol. Eng.* **18**, 229–240.

Churms, S.C. (1996) Recent progress in carbohydrate separation by high-performance liquid chromatography based on size exclusion, *J. Chromatogr., A* **720**, 151–166.

Cooper, C.A., Harrison, M.J., Wilkins, M.R., Packer, N.H. (2001) GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources, *Nucleic Acids Res.* **29**, 332–335.

D'Ambra, A.J., Baugher, J.E., Concannon, P.E., Pon, R.A., Michon, F. (1997) Direct and indirect methods for molar-mass analysis of fragments of the capsular polysaccharide of

Haemophilus influenzae type b, *Anal. Biochem.* **250**, 228–236.

Davies, M.J., Hounsell, E.F. (1996) Carbohydrate chromatography: towards yoctomole sensitivity, *Biomed. Chromatogr.* **10**, 285–289.

De Beer, T., van Zuylen, C.W.E.M., Leeflang, B.R., Hård, K., Boelans, R., Kaptein, R., Kamerling, J.P., Vliegenthart, J.F.G. (1996) NMR studies of the free alpha-subunit of human chorionic gonadotropin; structural influences of *N*-glycosylation and the beta-subunit on the conformation of the alpha subunit, *Eur. J. Biochem.* **241**, 229–242.

Dell, A., Reason, A.J., Khoo, K.-H., Panico, M., McDowell, R.A., Morris, H.R. (1994) Mass spectrometry of carbohydrate-containing biopolymers, *Methods Enzymol.* **230**, 108–132.

Dische, Z. (1949) Spectrophotometric method for the determination of free pentose and pentose in nucleotides, *J. Biol. Chem.* **181**, 379–392.

Dubois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A., Smith, F. (1956) Colorimetric method for determination of sugars and related substances, *Anal. Chem.* **28**, 350–356.

Dwek, R.A. (1995) Glycobiology: towards understanding of the function of sugars, *Biochem. Soc. Trans.* **23**, 1–25.

El Rassi, Z., Mechref, Y. (1996) Recent advances in capillary electrophoresis of carbohydrates, *Electrophoresis*, **17**, 275–301.

Guerardel, Y., Kol, O., Maes, E., Lefebvre, T., Boilly, B., Davril, M., Strecker, G. (2000) *O*-glycan variability of egg-jelly mucins from *Xenopus laevis*: characterization of four phenotypes that differ by the terminal glycosylation of their mucins, *Biochem. J.* **352**, 449–463.

Guttman, A., Ulfelder, K.W. (1997) Exoglycosidase matrix-mediated sequencing of a complex glycan pool by capillary electrophoresis, *J. Chromatogr., A* **781**, 547–554.

Hakomori, S. (1964) A rapid permethylation of glycolipids, and polysaccharide catalyzed by methylsulfinyl carbanion in dimethyl sulfoxide, *J. Biochem.* **55**, 205–208.

Hashimoto, S., Hashimoto, M., Yanai, M., Kumasaka, K., Kawano, K., Senou, A., Araki, H., Tsukamoto, Y., Takahashi, N. (2000) *N*-glycan structures derived from monoclonal immunoglobulin G cryoglobulins, *J. Clin. Biochem. Nutr.* **28**, 81–89.

Huckerby, T.N., Lauder, R.M., Nieduszynski, I.A. (1998) Structure determination for octasaccharides derived from the carbohydrate-protein linkage region of chondroitin sulphate chains in the proteoglycan aggrecan from bovine articular cartilage, *Eur. J. Biochem.* **358**, 669–676.

Jay, A. (1996) The methylation reaction in carbohydrate analysis, *J. Carbohydr. Chem.* **15**, 897–923.

Karamanos, N.K., Vanky, P., Tzanakakis, G.N., Tsegenidis, T., Hjerpe, A. (1997) Ion-pair high performance liquid chromatography for determining disaccharide composition in heparin and heparan sulphate, *J. Chromatogr., A* **765**, 169–179.

Knobloch, J.E., Shaklee, P.N. (1997) Absolute molecular weight distribution of low-molecular weight heparins by size-exclusion chromatography with multiangle laser light scattering detection, *Anal. Biochem.* **245**, 231–241.

Koizumi, K., Fukuda, M., Hizukuri, S. (1991) Estimation of the distributions of chain length of amylopectins by high-performance liquid chromatography with pulsed amperometric detection, *J. Chromatogr.*. **585**, 233–238.

Kormelink, F.J.M., Hoffmann, R.A., Gruppen, H., Voragen, A.G.J., Kamerling, J.P., Vliegenthart, J.F.G. (1993) Characterisation by ^{1}H NMR spectroscopy of oligosaccharides derived from alkali-extractable wheat-flour arabinoxylan by digestion with endo-(1 → 4)-β-D-xylanase III from *Aspergillus awamori*, *Carbohydr. Res.* **249**, 369–382.

Kragh-Hansen, U., Donaldson, D., Jensen, P.H. (2001) The glycan structure of albumin Redhill, a glycosylated variant of human serum albumin, *Biochim. Biophys. Acta-Protein Struct. Mol. Enzyme* **1550**, 20–26.

Küster, B., Wheeler, S.F., Hunter, A.P., Dwek, R.A., Harvey, D.J. (1997) Sequencing of N-linked oligosaccharides directly from protein gels: in-gel deglycosylation followed by matrix-assisted laser desorption/ionization mass spectrometry and normal-phase high-performance liquid chromatography, *Anal. Biochem.* **250**, 82–101.

Lamari, F., Militsopoulou, M., Gioldassi, X., Karamanos, N.K. (2001) Capillary electrophoresis: a superior miniaturized tool for analysis of the mono-, di-, and oligosaccharide constituents of glycan

moieties in proteoglycans, *Fresenius' J. Anal. Chem.* **371**, 157–167.

Lee, Y.C. (1996) Carbohydrate analyses with high-performance anion-exchange chromatography, *J. Chromatogr.* **720**, 137–149.

Leeflang, B.R., Faber, E.J., Erbel, P., Vliegenthart, J.F.G. (2000) Structure elucidation of glycoprotein glycans and polysaccharides by NMR spectrometry, *J. Biotechnol.* **77**, 115–122.

Levery, S.B., Toledo, M.S., Suzuki, E., Salyan, M.E.K., Hakomori, S., Straus, A.H., Takahashi, H.K. (1996) Structural characterization of a new galactofuranose-containing glycolipid antigen of *Paracoccidioides brasiliensis*, *Biochem. Biophys. Res. Commun.* **222**, 639–645.

Linsley, K.B., Chan, S.-Y., Chan, S., Reinhold, B.R., Lisi, P.J., Reinhold, V.N. (1994) Application of electrospray mass spectrometry to erythropoietin *N*- and *O*-linked glycans, *Anal. Biochem.* **219**, 207–217.

Maaheimo, H., Räbinä, J., Renkonen, O. (1997) ^{1}H and ^{13}C NMR analysis of the pentasaccharide Galβ(1 → 4)GlcAcβ (1 → 3)-[GlcNAcβ (1 → 6)Galβ (1 → 4)GlcNAc synthesized by the mid-chain β-(1 → 6)-D-*N*-acetylglucosaminyltransferase of rat serum, *Carbohydr. Res.* **297**, 145–151.

Mattu, T.S., Royle, L., Langridge, J., Wormald, M.R., Van den Steen, P.E., Van Damme, J., Opdenakker, G., Harvey, D.J., Dwek, R.A., Rudd, P.M. (2000) *O*-glycan analysis of natural human neutrophil gelatinase B using a combination of normal phase HPLC and online tandem mass spectrometry: implications for the domain organization of the enzyme, *Biochemistry-USA* **39**, 15695–15704.

Packer, N.H., Lawson, M.A., Jardine, D.R., Sanchez, J.C., Gooley, A.A. (1998) Analysis of glycoproteins separated by two-dimensional gel electrophoresis, *Electrophoresis* **19**, 981–988.

Parekh, R.B., Patel, T.P. (1992) Comparing the glycosylation patterns of recombinant glycoproteins, *Trends Biotechnol.* **10**, 276–280.

Perreault, H., Hronowski, X.L., Koul, O., Street, J., McCluer, R.H., Costello, C.E. (1997) High sensitivity mass spectral characterization of glycosphingolipids from bovine erythrocytes, mouse kidney and fetal calf brain, *Int. J. Mass Spectrum. Ion Processes* **169**, 351–370.

Picton, L., Bataille, I., Muller, G. (2000) Analysis of a complex polysaccharide (gum arabic) by multi-angle laser light scattering coupled online to size exclusion chromatography and flow field flow fractionation, *Carbohydr. Polym.* **42**, 23–31.

Rouse, J.C., Strang, A.-M., Yu, W., Vath, J.E. (1998) Isomeric differentiation of asparagine-linked oligosaccharides by matrix-assisted laser desorption-ionization postsource decay time-of-flight mass spectrometry, *Anal. Biochem.* **256**, 22–46.

Saaki, T., Yamada, H., Matsumura, K., Shimizu, T., Kobata, A., Endo, T. (1998) Detection of *O*-mannosyl glycans in rabbit skeletal muscle α-dystroglycan, *Biochim. Biophys. Acta* **1425**, 599–606.

Sheng, S., Cherniak, R., van Halbeek, H. (1998) A 1H NMR spectroscopic approach to the unambiguous determination of glycosyl linkage positions in oligosaccharides, *Anal. Biochem.* **256**, 63–66.

Shibata, N., Onozawa, M., Tadano, N., Hinosawa, Y., Suzuki, A., Ikuta, K., Kobayashi, H., Suzuki, S., Okawa, Y. (1996) Structure and antigenicity of the mannans of *Candida famata* and *Candida saitoana*: comparative study with the mannan of *Candida guilliermondii*, *Arch. Biochem. Biophys.* **336**, 49–58.

Smith, F., Van Cleve, J.W. (1955) Reduction of the products of periodate oxidation of carbohydrates. I Hydrogenation with Raney nickel of the dialdehydes from the methyl glycopyranosides, *J. Am. Chem. Soc.* **77**, 3091–3096.

Somogyi, M. (1952) Determination of blood sugar, *J. Biochem.* **195**, 19–23.

Stahl, B., Linos, A., Karas, M., Hillenkamp, F., Steup, M. (1997) Analysis of fructans from higher plants by matrix-assisted laser desorption/ionization mass spectrometry, *Anal. Biochem.* **246**, 195–204.

Striegel, A.M., Plattner, R.D., Willett, J.L. (1999) Dilute solution behaviour of dendrimers and polysaccharides: SEC, ESI-MS and computer modelling, *Anal. Chem.* **71**, 978–986.

Suortti, T. (1997) Coupled size-exclusion chromatography-anion-exchange chromatography in the analysis of poly- and oligosaccharides, *J. Chromatogr., A* **763**, 331–335.

Suzuki, S., Honda, S. (1998) A tabulated review of capillary electrophoresis of carbohydrates, *Electrophoresis* **19**, 2539–2560.

Thornton, D.J., Khan, N., Mehotra, R., Howard, M., Veerman, E., Packer, N.H., Sheehan, J.K. (1999) Salivary mucin MG1 is comprised almost entirely of different glycosylated forms of the MUC5B gene product, *Glycobiology* **9**, 293–302.

Tseng, K., Lindsay, L.L., Penn, S., Hedrick, J.L., Lebrilla, C.B. (1997) Characterization of neutral oligosaccharide-alditols from *Xenopus laevis* egg jelly coats by matrix-assisted laser desorption Fourier transform mass-spectrometry, *Anal. Biochem.* **250**, 18–28.

Van Rooijen, J.J.M., Voskamp, A.F., Kamerling, J.P., Vliegenthart, J.F.G. (1999) Glycosylation sites and site-specific glycosylation of Human Tamm-Horsfall glycoprotein, *Glycobiology* **9**, 21–30.

Viëtor, R.J., Hoffmann, R.A., Angelino, S.A.G.F., Voragen, A.G.J., Kamerling, J.P., Vliegenthart, J.F.G. (1994) Structures of small oligomers liberated from barley arabinoxylans by endoxylanase from *Aspergillus awamori*, *Carbohydr. Res.* **254**, 245–255.

Viseux, N., de Hoffmann, E., Domo, B. (1998) Structural assignment of permethylated oligosaccharide subunits using sequential tandem mass spectrometry, *Anal. Chem.* **70**, 4951–4959.

Vives, R.R., Pye, D.A., Salmivirta, M., Hopwood, J.J., Lindahl, U., Gallagher, J.T. (1999) Sequence analysis of heparan sulphate and heparin oligosaccharides, *Biochem. J.* **339**, 767–773.

Weiskopf, A.S., Vouros, P., Harvey, D.J. (1998) Electrospray ionization-ion trap mass spectrometry for structural analysis of complex N-linked glycoprotein oligosaccharides, *Anal. Chem.* **70**, 4441–4447.

Wittgren, B., Wahlund, K.G. (1997) Fast molecular mass and size characterization of polysaccharides using asymmetrical flow field-flow fractionation-multiangle light scattering, *J. Chromatogr., A* **760**, 205–218.

Wyatt, P.J. (1993) Light scattering and the absolute characterization of macromolecules, *Anal. Chim. Acta* **272**, 1–40.

Yang, Y., Orlando, R. (1996) Identifying the glycosylation sites and site-specific carbohydrate heterogeneity of glycoproteins by matrix assisted laser desorption/ionization mass spectrometry, *Rapid Commun. Mass Spectrom.* **10**, 932–936.

Carbohydrate Antigens

Denong Wang
Columbia University, New York, NY, USA

1 Classifications 279
1.1 Microbial Antigens 280
1.2 Alloantigens and Xenoantigens 280
1.3 T-dependent and T-independent Antigens 280

2 Diversity of Carbohydrates 281
2.1 Composition, Sequences, and Linkages 281
2.2 Conformational Diversity 284
2.3 Biological Complexity 287

3 Sugar Chains as Antigenic Determinants 288
3.1 Sizes of Antigenic Determinants 288
3.2 Types of Carbohydrate Epitopes: Terminal and Internal Structures 289

4 Carbohydrate-protein Interactions 291
4.1 Lectins 291
4.2 Carbohydrate-based Self/nonself Discrimination 292

5 Immune Responses to Microbial Polysaccharides 293
5.1 T-independency of Anti-polysaccharide Responses 293
5.2 Complexity of Anticarbohydrate Responses 294
5.3 Delayed Maturation of Anticarbohydrate Responses 295

6 Future Prospects 297

Acknowledgment 298

Bibliography 298
Books and Reviews 298
Primary Literature 299

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Keywords

Monosaccharide
The most basic carbohydrate molecule that cannot be divided into simpler sugars by hydrolysis.

Oligosaccharide
A linear or branched chain composed of a small number of monosaccharides (2 to 10 or 20) that can be broken down into simpler sugars via hydrolysis.

Polysaccharide
A polymer composed of a large number of monosaccharides, at least more than 10 to 20, and usually many hundreds or thousands of monosaccharide units arranged in repeated structures.

Glycoconjugate
A molecule in which an oligosaccharide is covalently linked to a protein or a lipid, forming either a glycoprotein or a glycolipid, respectively.

Glycan
Equivalent to polysaccharide but usually refers to the sugar chains of glycoconjugates.

Glycosaminoglycan
A heteropolysaccharide chain usually composed of a hexosamine and glucuronic acid as seen in proteoglycan.

Glycoprotein
A molecule in which an oligosaccharide is covalently linked to a protein.

Proteoglycan
A subclass of glycoproteins with large heteroglycan chains usually composed of a hexosamine and glucuronic acid.

Glycolipid
A molecule in which an oligosaccharide is covalently linked to a lipid.

Glycosphingolipid
A subclass of glycolipid in which an oligosaccharide is covalently linked to the lipid sphinganine.

Lectin
A carbohydrate binding protein or glycoprotein of nonimmune origin that binds noncovalently to carbohydrates with specificity and selectivity.

Carbohydrate antigens are polysaccharides and glycocojugates of multiple structural configurations that are able to evoke carbohydrate-specific antibodies and react with these antibodies. Although recognition of carbohydrates as antigens began with the study of microbial polysaccharides, the scope of investigation of carbohydrate antigens has since been substantially extended. Carbohydrates are unsurpassed in generating structural diversity and are prominent in surface display. They contain valuable biological information and are characteristically recognizable by either soluble or membrane-bound cellular proteins. In the higher eukaryotic species, expression of cellular glycoconjugates, and especially their complex carbohydrate structures, is frequently cell type–specific or tissue-specific. In microbes, many sugar chains, including those displayed on the surface of a microbial cell and those secreted outside it, have been recognized as "signatures" of specific pathogens. Exploring the biological information content of sugar chains is the current focus of postgenomic research. Carbohydrate antigen, as a classical immunological topic, has gained renewed interest in the new millennium.

1 Classifications

Carbohydrate antigens exist in multiple structural configurations, including polysaccharides, glycoproteins, glycolipids, glycosphingolipids, glycosaminoglycans, and proteoglycans. Given their origins, carbohydrate antigens were traditionally classified as microbial antigens and alloantigens. Microbes produce the former; the latter are expressed by mammalian cells. Carbohydrates may be "complete-antigens," such as microbial polysaccharides, in that they are able to induce reactive antibodies in certain species. They may also be components or antigenic determinants of glycoconjugates. Such carbohydrate moieties, or oligosaccharides, are termed *haptens* in immunology. They determine the specificity of an immune response but are not being able to evoke an immune response in their free, unconjugated forms. Consideration of glycoconjugates as carbohydrate antigens makes it necessary to further classify carbohydrate antigens into two large immunological categories, the T-dependent (TD) and T-independent (TI) forms of carbohydrate antigens.

Therefore, different methods have been used to classify carbohydrate antigens. Their relationship is summarized in Table 1. We must emphasize that this table is an attempt to illustrate how researchers of different scientific disciplines "see" a

Tab. 1 Classification of carbohydrate antigens.

	TI-1	*TI-2*	*TD*
Microbial antigens			
Polysaccharide	No	Yes	No
Lipopolysaccharide	Yes	No	No
Glycoprotein	No	No	Yes
Glycolipid	No	Yes	No
Allo- & xenoantigens			
Glycoproteins	No	No	Yes
Glycolipids	No	Yes	No
Glycosaminoglycans	No	Yes	No
Proteoglycans	No	Yes	Yes (?)

Notes: TI: T-independent; TD: T-dependent.

common object and work together to better understand the object – carbohydrates. When a specific molecule is involved, experiments must be conducted to learn its immunological properties. Diversity of carbohydrate antigens and the complexity of host immune systems have made it difficult to summarize a general scheme for the antibody responses induced by carbohydrate antigens. Seeing exceptions regarding the TD and TI characteristics of an antigen from a given class may not be surprising.

1.1
Microbial Antigens

Microbial carbohydrate antigens include polysaccharides, glycolipids, and glycoproteins. Polysaccharides may be cell wall components of microorganisms, such as lipopolysaccharides (LPS) in gram-negative bacteria, or capsules in gram-positive bacteria, known as *capsular polysaccharides*. They may also be secreted as exopolysaccharides, forming slime-like amorphous materials to trap microbial cells in a viscous matrix. These polysaccharides are all located on the cell surface and are, therefore, of importance in the recognition and immune response of a higher organism to microbial infections. Glycoproteins were recently recognized in a number of bacterial species. By contrast, they were found essentially in almost all membrane-bound or secretory proteins of viruses and many parasites.

1.2
Alloantigens and Xenoantigens

Eukaryotic cells also express carbohydrate antigens, notably alloantigens of the human blood group systems and xenoantigens expressing α-Gal epitopes. The former is an allogenic marker of human red blood cells and is responsible for the rejection of an allogenic blood transfusion. The latter is a cross-species barrier in organ transplantation and is known as the main cause of human immune rejection of a donor pig's kidney. These antigens are present as glycoconjugates of different molecular configurations such as glycoproteins and glycolipids. The carbohydrate chains attached to either proteins or lipids serve as antigenic determinants. Some carbohydrate structures are composed of more than two different sugar residues and have further chemical modifications at specific positions. Such structures are called *complex carbohydrates*. They are systematically assembled by sets of enzyme systems via complex biosynthesis processes. The expression of blood group substances A, B, H, and Lewis (Le) appears to be developmentally and differentiationally regulated. Atypical expression of such substances may lead to the formation of tumor-associated antigens.

1.3
T-dependent and T-independent Antigens

Protein antigens fail to elicit antibody responses in athymic mice and *nude* (−/−) mice. Polysaccharides and other macromolecules with repetitive antigenic determinants can induce unimpaired antibody responses in these mice. These observations establish two broad categories of antigens, T-dependent (TD) and T-independent (TI). TI antigens are further divided into type I (TI-I) and type II (TI-II), based on the ability to elicit antibodies in the CBA/N mouse strain, with an X-chromosome linked immunodeficiency (*xid*). Lipopolysaccharides (LPS)

of gram-negative bacteria, which induce antibodies in such strains, are the TI-I antigens; capsular polysaccharides of gram-positive bacteria and exopolysaccharides that induce no response in *xid* strains are classified as TI-II antigens.

Glycoproteins are TD antigens; other forms of glycoconjugates, including glycolipids, Glycosphingolipids, glycosaminoglycans, and proteoglycans, are generally T-independent in nature. The carbohydrate chains of glycosaminoglycans and proteoglycans are composed of repeats of disaccharide units and are relatively large. Their antigenic reactivities may vary significantly among molecules and depend on the structure of a specific preparation and the animal species used for immunization. Glycolipids are a unique category of antigens since their hydrophobic lipid tails may integrate into cellular membranes of distinct cell types while the hydrophilic sugar chains are displayed on the surfaces. There are many documented examples to prove that molecules with such structural configurations are potent antigens and vaccines.

2 Diversity of Carbohydrates

2.1 Composition, Sequences, and Linkages

Carbohydrates are unique in their structural diversity. The structure of a sugar chain is determined by the composition and sequence of its sugar residues, as well as by their glycosidic linkages. This is different from nucleic acids and proteins that are synthesized linearly in cells with a single type of covalent bond to bridge their monomers. Carbohydrates can be linked together in multiple ways and can thereby generate different structures with identical residue compositions and sequences.

Monosaccharides are the basic building blocks of carbohydrate structures. They can be linked by means of either α- or β-anomeric linkages at each joining point, whereby the free hydroxyl on given monosaccharide links an adjacent monosaccharide (Fig. 1). Unlike proteins that are connected solely by a peptide bond, carbohydrates utilize many possible

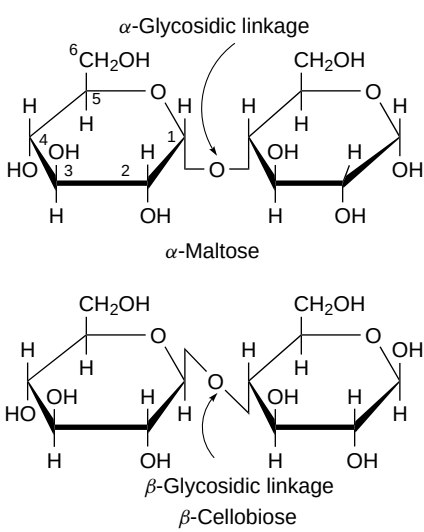

Fig. 1 α- or β-anomeric linkages of carbohydrates. The diagram illustrates the two kinds of linkages that may exist between at least two monosaccharides or sugar residues. These linkages may be either α- or β-anomeric where the free hydroxyl on one monosaccharide connects to an adjacent monosaccharide.

glycosidic linkages so as to extensively diversify their structures. Two amino acid residues, such as two alanines, can produce only one possible dipeptide; however, two molecules of glucose have the potential to generate 11 different disaccharides. A trimer of any of the 9 common sugar residues of the human body theoretically can give rise to 119 736 different structural isomers. This is strikingly in contrast to the maximal construction of 8000 tripeptides using 20 different amino acid residues. Theoretically, carbohydrate structures can have unlimited variation. This potential makes them well suited to provide various recognition signals.

Oligosaccharides and polysaccharides are composed of monosaccharides. A chain of 2 to 10 monosaccharide residues is called an *oligosaccharide*. A polymer of monosaccharides that is longer than the oligosaccharide chains of 10 to 20 sugar units in length is termed a polysaccharide. Usually, an oligosaccharide means also a structurally defined carbohydrate molecule. Polysaccharides are, however, large polymers of unspecified length with repeated structures. Oligosaccharides may be simply the shorter stretches or fragments of a polysaccharide. They could also be the sugar components of glycocojugates of diverse sources. The oligosaccharides attached to the glycocojugates of mammalian cells frequently consist of multiple distinct monosaccharide residues linked in complex structures and associated with different biological molecules. For example, the sugar chain structure of human blood group A substance differs in composition from that of group B or group H by only one monosaccharide residue (Fig. 2). These sugar structures may link to either a protein or a lipid molecule, forming a glycoprotein and a glycolipid, respectively.

Polysaccharide chains composed of a single type of sugar residue are termed *homopolysaccharides*. Those formed by more than one type of monosaccharides are named *heteropolysaccharides*. In a homopolysaccharide, the glycosidic bonds linking the monosaccharides are the only determining factor of their structural diversity. A well-documented example is

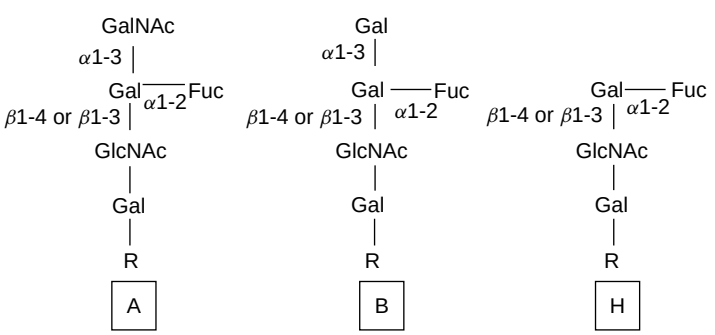

Fig. 2 Structures of human blood group A, B, and H. This diagram illustrates three different human blood group antigens A, B, and H. The addition of $\alpha(1\rightarrow2)$linked fucosyl residue to these oligosaccharides provides relatively rigid terminal branches. Within the A, B, or H blood group, the R group may be either lipid or protein. This kind of terminal branch is also found in other blood group substances such as the lewis (Le) sugar series, which are a family of fucosylated glycans (see also Fig. 5).

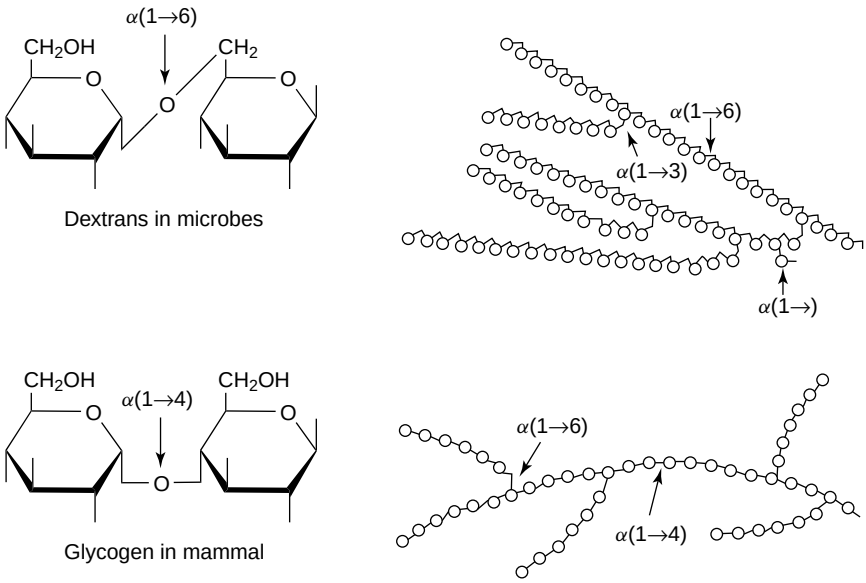

Fig. 3 Schematics of microbe's dextran and mammalian glycogen: linkages make the difference. The only source of structural diversity in homopolysaccharides is the glycosidic bonds linking the monoaccharides. This figure illustrates two commonly found polysaccharides – one in microbes and another in mammals. The polysaccharide with $\alpha(1\rightarrow6)$linkage is dextran, which is found in many bacteria including *Lactobacillaceae*. The polysaccharide with the $\alpha(1\rightarrow4)$ is glycogen and it is found in mammals as one of the energy storage biological molecules.

a class of polysaccharide called *dextrans* (Fig. 3). These macromolecules are composed entirely of glucose residues and are produced by numerous bacteria of *Lactobacillaceae*. However, dextran molecules derived from different strains differ significantly in their glycosidic linkage compositions. Some dextran preparations are predominantly or solely $\alpha(1\rightarrow6)$linked, forming molecules with dominantly linear chain structures; others are composed of multiple glycosidic linkages, including $\alpha(1\rightarrow6)$-, $\alpha(1\rightarrow3)$-, $\alpha(1\rightarrow2)$- and others, generating heavily branched molecules. These bacterial products and mammalian glycogen differ from each other only in linkage. Mammalian enzymes cannot digest dextrans as well as most microbial polysaccharides, and thus they may remain *in vivo* for a long period of time after an immunization. Linkage largely determines sensitivity or accessibility of polysaccharides to enzymes and, frequently, their antigenic reactivities. Branches occur in a polysaccharide when two or more C-atoms of one residue are involved in glycosidic linkage with other residues. Polysaccharides with more branches present more terminal structures.

Most heteropolysaccharides are composed of two or three different sugar residues and termed as *diheteroglycans* and *triheteroglycans*, respectively. The heteropolymers that consist of more than six distinct monosaccharides are rarely seen in living organisms. On the basis of the organization of sugar residues in heteropolysaccharides, they are further

classified as simple heteropolysaccharides and complex heteropolysaccharides. In the former, each polysaccharide chain is a homopolysaccharide, though two or more different homopolymers are present in a large heteropolysaccharide. In a complex heteropolysaccharide, the polymer contains two or more different types of sugar residues in a single branch.

2.2 Conformational Diversity

Conformational flexibility is an intrinsic property of carbohydrate molecules. An oligosaccharide in solution may exist in many different conformations with the lower energy forms predominating. Significant influence of the composition of glycosidic linkages on the structural characteristics of oligosaccharides and polysaccharides had been suggested for some time. In the 1960s, it was pointed out: "It is important to remember that each glucose residue is free to rotate around the $\alpha(1\rightarrow6)$ glycosidic bond so that the schematic representation in Fig. 3 for isomaltose is not rigid. It may well be that the molecules prefer certain conformations in solution, conformations in which the planes of the ring bear a definite orientation to one another (see: Kabat, E.A., Structural Concepts in Immunology and Immunochemistry, Second Edition, Holt, Rinehart and Winston, New York, 1976, Page 21)." In reacting with protein binding sites, however, certain conformations of an oligosaccharide may be selected and stabilized. Such phenomenon was later termed *induced fit*.

Recent technological advances have made it possible to "see" whether an induced fit indeed occurs when an oligosaccharide intereacts with its receptor. This can be achieved by comparing the solution conformers of oligosaccharides with those in carbohydrate–protein complexes. Nuclear magnetic resonance (NMR), fluorescence energy transfer (FET), optical rotation (OR), or Raman optical activity, can be applied for detecting the molecular flexibility in solution. X-ray crystallography gives high-resolution determinations of molecular coordinates in the "frozen" state. These comparative studies demonstrated that (1) $\alpha(1\rightarrow6)$ glycosidic bond in an oligosaccharide is more flexible than other linkages; (2) the linkages attaching terminal residues to main chain structures are more flexible; (3) those in internal chains are relatively stable; and (4) sugar rings are generally rigid in conformation, but, induced ring distortion can be seen.

The $\alpha(1\rightarrow6)$ glycosidic bond is superior to other glycosidic bonds in conformational flexibility since three torsion angles are required to define the conformation of this glycosidic bond (Fig. 4). Using NMR, the solution conformation of melibiose [Gal$\alpha(1\rightarrow6)$Glc] was compared with that in its complex with ricin B-chain. The disaccharide exists predominantly as two conformers, resulting from rotation about the $\alpha(1\rightarrow6)$linkage. Only one of the two conformers binds to the lectin. Enzymes may also recognize only certain conformations of their carbohydrate substrates. GlcNAc transferase V can use Man $\alpha(1\rightarrow6)$ Man for extending N-linked oligosaccharides into higher branched structures. To see which conformer is suitable for the enzyme, two rigid bicyclic analogs of Man $\alpha(1\rightarrow6)$ Man were generated by fixing the $\alpha(1\rightarrow6)$linkage in either a gauche–gauche (gg) or gauche–trans (gt) conformation. It was shown that only the gg conformer was used as substrate by the enzyme. Thus, this enzymatic selectivity depends on the conformational property of carbohydrate molecules.

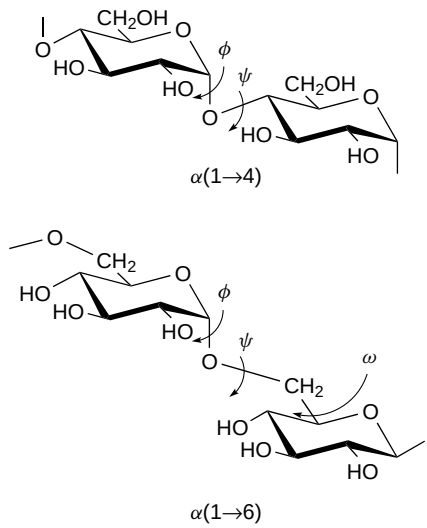

Fig. 4 Torsion angles of $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$linkages. The glycosidic bond with the greatest conformational flexibility is the $\alpha(1\rightarrow6)$ glycosidic bond. Its flexibility is attributable to the fact that three torsion angles are required to define the conformation of this glycosidic bond.

The type of linkage also determines the chain conformation or the 3D structure of the intact polysaccharide. In polysaccharides, all component monosaccharides exist in the ring configuration, except for the terminal-reducing residue, in which there is equilibrium between the ring and its open-chain form. The geometry of the individual sugar rings in a polysaccharide is essentially rigid. However, adjacent monosaccharides are potentially rotatable around their glycosidic bonds. The range of rotation varies for different glycosidic linkages and is limited by steric hindrance between adjacent rings and by hydrogen-bonding patterns between groups in adjacent residues. Depending on sugar chain linkage types, homopolysaccharides may exist in four different conformations: the extended ribbon (Type A), the flexible helix (Type B), the crumpled ribbon (Type C), and the flexible coil (Type D). The fourth type has the most flexibility, as seen in $\alpha(1\rightarrow6)$ dextran. Differing from other glycosidic bonds, three torsion (rotational) angles, ϕ, ψ and ω are required to define the conformation of a $\alpha(1\rightarrow6)$linkage. Irregularities in composition and the existence of large branched structures can inhibit regular conformational patterns.

Oligosaccharides with terminal branches are relatively rigid in conformation. Such structures are frequently seen in alloantigens and serve as the key carbohydrate structures for immunorecognition. As shown in Fig. 2, addition of $\alpha(1\rightarrow2)$linked fucosyl residue in the linear sugar chains generates the terminal branch-type structures of human blood group A, B, and H. Terminal branches are seen also in other blood group substrances, such as the Lewis (Le) sugar series, which are a family of fucosylated glycans (Fig. 5). Sialyl Le^x is a ligand for E-selectin. Its solution conformation was studied using proton NMR to assign each proton and advanced NOE (Nuclear Overhauser Effect) measurements to determine inter-proton distances. Unlike other oligosaccharides with multiple conformers in solution, Le^x is predominantly in a single rigid conformation. The conformation of sulphated Le^a approximates also that of the unsulfated analog (Le^a). The enhanced binding of sulfated

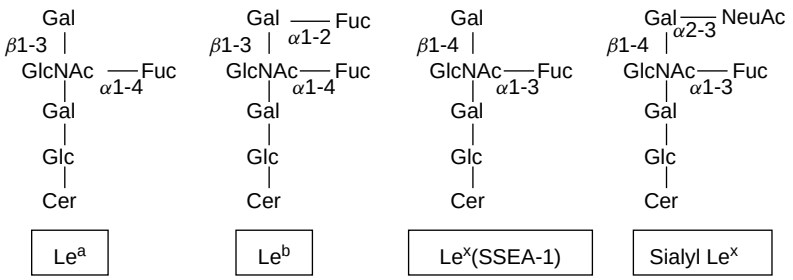

(a) Lewis system

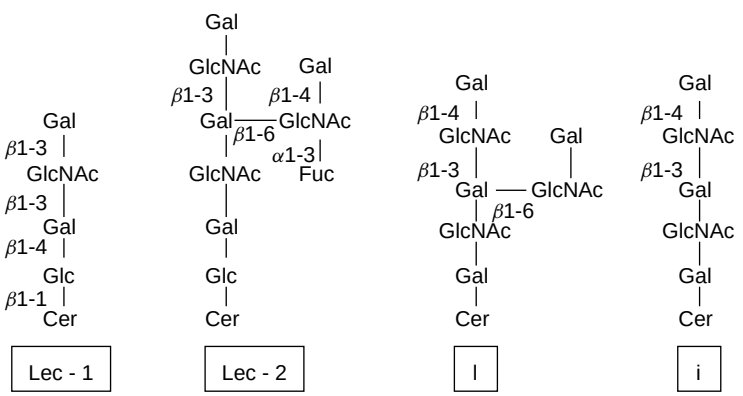

(b) I and i antigens

Fig. 5 Schematics of blood group substances I, i, and Le series. This diagram illustrates the rigid conformation of various terminal chain possessing oligosaccharides. Some sugar chains represented here have had $\alpha(1\rightarrow2)$linked fucosyl residues added to create the terminal branches of the human blood group substances. (a) Lewis (Le) sugar series. (b) I and i sugar chains.

Le^a to E-selectin was thus attributed to favorable electrostatic interactions between the charged sulfate group and the selectin molecules. Close packing between fucose and galactose residues might be a major cause of this conformational rigidity. Such conformational rigidity is also seen in other blood group substances.

Understanding the conformational flexibility or rigidity of oligosaccharides is important for the rational design of drugs, vaccines, and other biological reagents using carbohydrate molecules. Recent recognition of a conformational epitope of a sugar chain as an effective vaccination against meningococcal infections highlights this concept. Meningococcal meningitis is a severe childhood disease that often results in significant disability or death. Two major etiological agents of meningitis are the group B *meningococci* and capsular type K1 *Escherichia coli*. The virulence of these organisms is attributable to structural mimicry between their common $\alpha(2-8)$-polysialic acid capsular polysaccharide and human tissue antigens. This property allows the bacteria to evade immune

surveillance. Accordingly, such structural mimicry poses a challenge to the development of an effective vaccine to protect against this infection. NMR and other experiments found, however, that the capsular polysaccharide of the bacteria can adopt a unique "antigenic conformation" of helical structures, which are pathogen-specific and are not cross-reactive with the smaller oligomers of sialic acid expressed by human tissue. This led the way to design and develop chemically modified sugar chain vaccines, whereby the pathogen-specific polymer helical structures were better preserved. In an immunization using one of the compounds, N-propionylated group B meningococcal polysaccharide (NPrGBMP), the major population of antibodies elicited showed no significant cross-reactivity with $\alpha(2-8)$-polysialic acid, but preserved all bactericidal activity *in vitro* and protectivity *in vivo*.

2.3 Biological Complexity

By reviewing the chemical characteristics of carbohydrates, it is clearly seen that carbohydrate chains can have almost unlimited variation. The actual variation in carbohydrate structure in a given organism, is, however, restricted by the complexity of biosynthesis of carbohydrates. Sugar chains are not primary gene products. The biosynthesis of a carbohydrate chain is catalyzed step by step by a cluster of specific enzymes, called *glycosyltransferases*. For example, a cluster of genes directs the biosynthesis of the group B *meningococci* capsular polysaccharide (GBMP) that we discussed above. Each enzyme is responsible for a specific chemical reaction for the synthesis of the molecule. Glycosyltransferases are protein products of genes. For every new carbohydrate structure, a gene must be provided. It is, thus, believed that any given organism has a limited number of variations in carbohydrates. Carbohydrates are, however, the most abundant organic substances and are produced by virtually all the living organisms. They may employ various mechanisms and pathways for sugar chain synthesis. Thus, the overall repertoires of diversity and complexity of naturally occurring carbohydrate molecules can be extraordinarily large.

Bacterial polysaccharides are typically much more complex than those produced by plants or animals. They are often partially composed of unusual sugar residues that are rarely seen in higher eukaryotic species. Although only nine monosaccharides are commonly seen in mammals, over a hundred different monosaccharides are found in bacteria. Some of these sugars, such as ketodeoxyoctonic acid, were previously thought to be limited in their occurrence to LPS. Other interesting examples include the identification of hexosaminuronic acids in a wide range of bacterial polymers, aminohexuronic acid in *E. coli K7*, *Streptococcus pneumoniae* type 12 F and *Achromobacter georgipolitanum*; amphipathic exopolysaccharide "emulsan" from *Acinetobacter calcoaceticus*; and the presence of several *o*-methyl sugars, including an *o*-methyl-6-deoxyhexose, in the exopolysaccharide sheath. In addition to carbohydrates, the microbial polysaccharides contain various ester-linked substituents and pyruvate ketals, which are frequently of immunological significance.

Glycoproteins are present in both mammals and microbes. The prokaryotic microorganisms synthesize the N-linked and the O-linked oligosaccharides on proteins, just as mammals and other eukaryotes do. Their glycosylation machineries have,

however, significant differences. Prokaryotes lack internal membrane-bound organelles, such as a nucleus, endoplasmic reticulum (ER), and Golgi apparatus. They process oligosaccharides on the outside of the cell membrane in the so-called extracellular space. The sugar chains decorating the glycoproteins of microbes of various strains or species are quite diverse in their structures. Many of them are selectively or specifically expressed by certain strains and/or species. There are, however, instances where bacteria-produced oligosaccharides share similarity with eukaryotic sugar chain structures. As discussed above, a documented example is the structural similarity between the $\alpha(2-8)$-polysialic acid capsular polysaccharide of human pathogens and the sialic acid structures of human brain tissues. By mimicking host carbohydrates, the pathogen evades detection by the immune system, which recognizes it as part of "self." This has been recognized as a strategy by which a number of pathogens evade host immune recognition.

Microheterogeneity in carbohydrates is another documented phenomenon that can be attributed to the characteristics of sugar chain biosynthesis. Sharply contrasted with the template-dependent mechanism of nucleic acid and protein synthesis, there is no rigorous quality-control mechanism in the biosynthesis of sugar chains. A polysaccharide may thus not only have a main sugar chain composed of a type of repetitive structure but also contain a few different minor structural moieties as well. Immunologically, the former may form a dominant antigenic determinant of a polysaccharide. The latter may generate a minor antigenic determinant of the antigen.

3 Sugar Chains as Antigenic Determinants

As discussed above, microbial carbohydrate antigens of multiple structural configurations have been recognized. Polysaccharides and lipopolisaccharides are widely produced by many bacterial species. Glycoproteins are expressed and displayed on the surface of virions in almost all known viral species infecting mammals. Viruses take advantage of the cellular machineries of carbohydrate synthesis to produce sugar chains and glycoproteins. Parasites of mammals are higher eukaryote species and have their own enzymatic systems for the biosynthesis of carbohydrate molecules and protein glycosylation, generating diverse sugar structures. Presence of glycoproteins in some bacterial species has been also recently recognized. The oligosaccharide chains that define the antigenic specificities of these microbial antigens are their antigenic determinants.

3.1 Sizes of Antigenic Determinants

An antigenic determinant is a part of an antigen that is bound by and is complementary to the combining site of a specific antibody. It is the key element for host recognition and immune response. The antigenic determinants of a carbohydrate antigen are generally its oligosaccharide chains in certain size, shape, and conformation. Given this widely accepted definition, it is possible to estimate the size of a carbohydrate-based antigenic determinant by characterizing the specific interaction between carbohydrate ligand and its antibody-combining sites. The exact sizes and shapes of an antigenic determinant can be, however, visualized by

X-ray crystallographic characterization of its 3D structure.

A model system, $\alpha(1\rightarrow6)$ dextran and antidextran antibodies, was applied to address the question regarding sizes of a carbohydrate-based antigenic determinant. Native dextran N279 is a near-linear macromolecule of glucose with predominant $\alpha(1\rightarrow6)$linkage (90%). The relative structural simplicity of this carbohydrate antigen and the availability of oligosaccharides of $\alpha(1\rightarrow6)$linked glucose made it suitable for such studies (See Fig. 6). Sizes of antibody-combining sites were measured by competitively displacing the dextran from dextran–antibody complexes by various isomaltose oligosaccharides. The combining site is considered to be complementary in size to the smallest oligosaccharide giving maximum competition. Such an oligosaccharide is defined as the epitope or antigenic determinant. It may vary in size from a lower limit of one to two glucoses to an upper limit of six to seven.

3.2
Types of Carbohydrate Epitopes: Terminal and Internal Structures

A carbohydrate antigen may display both the terminal sugar moieties and internal chain structures on its solvent accessible surface. This is owing to the hydrophilic property of carbohydrates, making them strikingly different from proteins. In aqueous solution, proteins tend to fold to bring their hydrophobic side chains together, forming an oily core with polar side chains exposed. Surface moieties of a protein antigen may serve as antigenic determinants interacting with B-cell Ig-receptors;

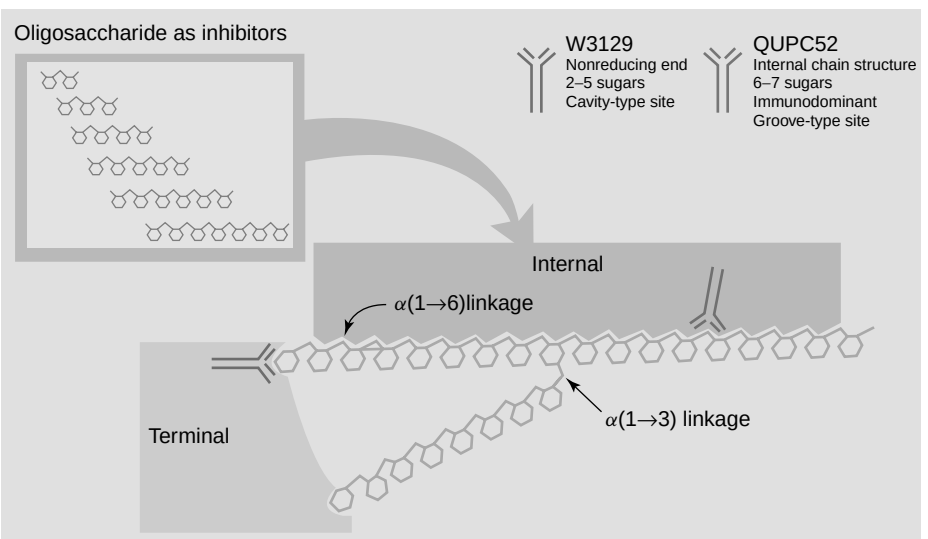

Fig. 6 Isomaltose oligosaccharides as "rulers" to measure the sizes of antigenic determinants. The diagram illustrates the two different specificities of anti-$\alpha(1\rightarrow6)$ dextran antibodies. Antibodies bound to polysaccharide-$\alpha(1\rightarrow6)$ dextran can be either cavity-type for terminal epitopes or groove-type for internal epitopes. The size of the antibody binding site can be determined by using oliogsaccharides of different number of sugar residues as competitive inhibitors that displace the dextran–antibody complexes.

interior residues are generally not accessible to such interactions. Carbohydrates are built up by monosaccharides whereby the enriched hydroxy groups readily interact with water molecules by hydrogen bonding. Their glycosidic linkages are more flexible than the peptide bonds in proteins and protein-like folding patterns are not seen in polysaccharides. Thus, not only are the terminals of the carbohydrate chains accessible for molecular recognition but residues in the internal chain are also exposed in solvent and are frequently reactive.

This important concept was first established in an immunochemical study focusing on the fine specificities of antidextran antibodies. Immunochemical mapping of the combining sites of two monoclonal myeloma proteins specific for $\alpha(1\rightarrow6)$ dextran, W3129 and QUPC52, established experimentally the existence of combining sites, which are specific for either the internal linear chain or the terminal structures of carbohydrate molecules (See Fig. 6 for a schematic view of their binding specificities). W3129 had a site saturated by isomaltopentaose, whereas QUPC52 accommodated isomaltohexaose. The binding constants were $1 \times 10^5\ M^{-1}$ for the former and $5 \times 10^3\ M^{-1}$ for the latter. Further analyses showed the terminal nonreducing glucose to contribute 50 to 60% of the total binding energy with W3129 but less than 5% with QUPC52. The two antibodies also differ in their ability to precipitate a synthetic linear dextran with about 200 glucoses. QUPC52 but not W3129 forms precipitins in saline. These observations were interpreted to indicate that the combining sites of QUPC52 must be a "groove" into which internal chain epitopes of $\alpha(1\rightarrow6)$linked glucose could fit; in contrast, the nonreducing ends of dextran may be held by the cavity-type site of W3129. Computer model building studies on two cavity-type mAb, W3129 and 16.4.12E, and one groove-type mAb 19.1.2 support the immunochemical distinction of these two basic types of combining sites. Recognition through either terminal moieties or internal structures of a carbohydrate molecule may also occur in other protein–carbohydrate interactions.

X-ray crystallography visualizes the epitopes bound in their combining sites. Such analyses have been performed on several carbohydrate–antibody complexes. The presence of terminal, branched terminal, and internal chain epitopes of polysaccharide antigens, as well as their corresponding antibody-combining sites, have been confirmed. The crystal structure of the murine Fab S-20-4 from a protective anticholera, Ab specific for the lipopolysaccharide Ag of the Ogawa serotype was determined in its unliganded form and in complex with synthetic fragments of the Ogawa O-specific polysaccharide (O-SP). The upstream terminal O-SP monosaccharide was shown to be the primary antigenic determinant to accommodate the antibody binding pocket. This antibody-combining site is immunochemically similar to a cavity-type anti-$\alpha(1\rightarrow6)$ dextran, W3129, in that a single terminal sugar residue accounts for 90% of the maximal binding energy. A *Salmonella* trisaccharide epitope bound by mAb Se 155-4 was also extensively studied. The Fab of this mAb was successfully cocrystallized with a branched trisaccharide αD-Galp$(1\rightarrow2)$[α-D-Abep$(1\rightarrow3)$]-α-D-Manp. The trisaccharide fills a hydrophobic pocket, 8-Å deep by 7-Å wide, a size close to previous estimates of the minimum for anti-$\alpha(1\rightarrow6)$ dextran sites.

Unlike the above examples, whereby the terminal or branched terminal epitopes

was seen in crystallized carbohydrate antibody complexes, the crystal structure of a monoclonal Fab (YsT9.1.) and a *Brucella* A cell wall polysaccharide showed binding between a groove-type site and an internal chain epitope of microbial polysaccharide. It discriminates the *Brucella abortus* A antigen from the nearly identical *Brucella melitensis* M antigen. By forming a groove-type binding site, lined with tyrosine residues, this antibody accommodates the rodlike A antigen but excludes the kinked structure of the M antigen. This is consistent with immunochemical data that the binding site of this groove-type anti-Brucella A polysaccharide antibody can be optimally filled by a pentamer or hexamer of $\alpha(1\rightarrow 2)$linked A polysaccharide. The size of this epitope is similar to that which accommodates to the groove-type anti-$\alpha(1\rightarrow 6)$ dextran antibodies, but much larger than the cavity-type anti-$\alpha(1\rightarrow 6)$ dextrans.

4 Carbohydrate-protein Interactions

The biological significance of the sugar chain's structural diversity depends on whether cellular machineries can recognize their signals. In fact, at least two large classes of proteins, lectins and antibodies, are present in living organisms and serve as specialized "decoders" to decipher the information content of sugar chains. Lectins and antibodies represent two distinct models of protein–carbohydrate interactions. The former is widely seen in the events of cell–cell interactions and cell signaling, but is also found in the innate immune systems as antimicrobe reactors. The latter is a specialized system of immune recognition and anti-infection response. Our main interest here is with the carbohydrate binding characteristics and their functional correlations.

4.1 Lectins

Lectins are a large class of proteins or glycoproteins that can specifically or selectively bind carbohydrates. They are, however, neither enzymes nor antibodies. A wide range of living organisms, from microbes to mammals, produces lectins. Apparently, they are conserved in evolution. According to species origin, they are classified into at least four categories, which include microbial lectins, plant lectins, invertebrate lectins, and vertebrate lectins. In the last decade, most attention has focused on vertebrate lectins. On the basis of the similarities in sequence homology and activity, vertebrate lectins are grouped into five subfamilies, including C-type (Calcium-dependent), P-type, S-type, I-type, and pentraxins.

The carbohydrate binding activity of a lectin is commonly described in terms of the monosaccharide specificity, as suggested first by Professor Y. C. Lee. Experimentally, a type of inhibition assay, such as agglutination inhibition, is performed to identify the monosaccharide that most effectively inhibits the reaction. Lately, disaccharides have been considered to be a better way to define a lectin's specificity. Most lectins are able to cross-react with a panel of sugar chains with a common terminal sugar residue.

X-ray crystallographic studies identified some common characteristics in the binding pockets of some lectins and anticarbohydrate antibodies. Aromatic amino acid residues appear frequently in carbohydrate binding sites of various fine specificities. Such residues are large and participate in

a wide variety of van der Waals and electrostatic interactions. The majority of crystal structures so far resolved are associated with a network of H-bonds, frequently with water molecules, in their combining sites. Such bound water molecules play important roles at the contacting interface between carbohydrate ligands and receptors. In addition, polyamphiphilic surfaces were identified in the interfaces of a number of carbohydrate ligand and binding sites. The presence of relatively rigid structures is required to produce amphiphilic surfaces in solution. Carbohydrate molecules with branched termini, such as many blood group substances, are favorable for the generation of such contact surfaces.

4.2 Carbohydrate-based Self/nonself Discrimination

In the higher eukaryotic species, carbohydrate molecules serve as recognition signals for the cell–cell communication that occurs in a living organism. Carbohydrates also play critical roles in mediating the interactions that occur between hosts and microbes. This raises an important question on how molecular and cellular mechanisms allow self and nonself discrimination on the basis of carbohydrate structures. As discussed above, carbohydrates are unique in structural diversity. The microbe-produced sugar chains are structurally different from those expressed by the host organisms and thereby recognized by the host immune system as "foreign" and elicit an antibody response. This antibody or B-lymphocyte-mediated machinery is specialized for immune defense. It is, however, no longer useful for cell–cell communication in a host since B-cell clones that produce self-reactive antibodies are generally prohibited in the development of an individual's immune system. The lectin-mediated recognition machinery plays important roles in cell–cell communication.

Some characteristic differences between the lectin-mediated and antibody-mediated model of carbohydrate-recognition have been identified. First, lectins seem to have lower binding affinity to carbohydrates than antibodies do. Second, most lectins recognize the terminal epitopes or terminal branches of carbohydrate antigens; antibodies are able to bind either the terminal or internal epitopes. Third, lectins recognize a specific pattern, such as clusters of sugar structures. Their selectivity in carbohydrate binding is heavily dependent on the topological distribution of carbohydrate ligands. By contrast, antibody binding to carbohydrates is generally less dependent on the topological distribution of sugar epitopes but strongly relies on its binding affinity. The so-called cluster effect is seen in many lectin–carbohydrate interactions in cellular systems. In fact, the genetic and molecular machineries for constructing high affinity antibody-combining sites are not present in lectin systems. Apparently, the two carbohydrate-recognition systems have evolved to be better suited for their respective cellular functions.

In invertebrates and many lower vertebrate species, there are no acquired immune systems, such as those including B-cells and T-cells as found in mammals. They depend on the innate immune systems to combat potential microbial pathogens. In fact, both invertebrates and vertebrates have a self/nonself pattern recognition innate immunity system, which is based on the detection of carbohydrates that decorate the surface of microbial pathogen. This machinery is

responsible for the initial recognition step in the first line of internal defense, which triggers the initiation of various pathogen-killing or inhibition machineries. These include agglutination, endocytosis by phagocytic cells, and the activation of protease cascades, resulting in clotting, melanization, or complement-mediated killing of the pathogen. A fucose binding lectin, called *fucolectin*, was demonstrated to be an immune-recognition molecule in invertebranes and vertebrates, such as the horseshoe crab (*Tachypleus tridentatus*) and the Japanese eel (*Anguilla japonica*).

5 Immune Responses to Microbial Polysaccharides

5.1 T-independency of Anti-polysaccharide Responses

Microbial polysaccharides are T-independent antigens. As illustrated in Fig. 7, a polysaccharide activates B-cells directly by cross-linking their membrane-bound immunoglobulins, the B-cell antigen receptors (BCRs). Cross-linking BCRs is sufficient to induce B-cell proliferation,

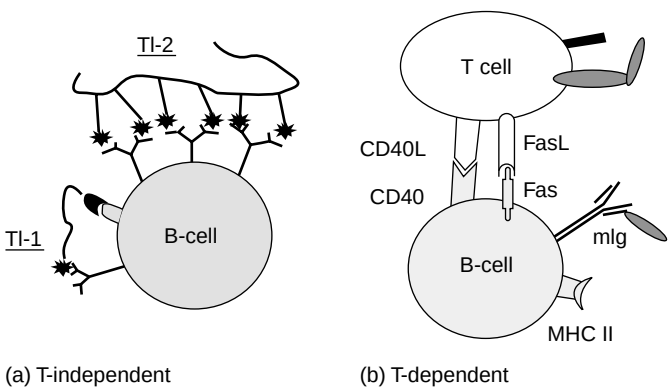

Fig. 7 Schematics of the TI- and TD-models of B-cell activations. The diagram demonstrates two distinct models of B-cell activation, the T-independent (TI) and the T-dependent (TD) models. (a) Both TI-I antigen (LPS) and TI-II antigen (polysaccharide) are able to cross-link the BCR (Ig) of B-cells and induce B-cell proliferation in the absence of T-cell help. TI-1 antigens (LPS) are polyclonal B-cell activators; TI-II antigens activate only the B-cells that express antibodies specific for given TI-II antigens. (b) TD protein activates B-cells in an antigen-specific and MHC-restricted manner. Unlike the TI antigens, protein antigens, in general, do not contain repeating structures and thereby are unable to cross-link the BCR (Ig) of B-cells on their own. Instead, specialized antigen-presenting cells, such as dendritic cells, macrophages, or B-cells themselves, take up protein antigens, process them in cells, and present the protein-derived peptides by their MHC class II molecule. Such peptide-bound MHC class II molecules are able to bind and activate CD4+ T-helper cells (Th) that express the appropriate T-cell receptor (TCR). Subsequent to the first initiation, additional molecular interactions, such as CD40 and CD40 ligand, and CD28/CTLA4-Ig with B7-1/B7-2, take place and activate further the interaction between T-B cells.

but not sufficient to induce class switching and antibody production. Secondary signaling is required for induction of class switching and for the terminal differentiation of B-cells into antibody-secreting cells (ASCs). This is significantly different from the B-cell activation by a T-dependent protein antigen.

B-cell activation via TD antigens is initiated by antigen-specific, MHC-restricted, T-cell–B-cell interactions. As a consequence of the first initiation, additional molecular interactions, such as CD40 and CD40 ligand, and CD28/CTLA4-Ig with B7-1/B7-2, can proceed. These molecular events further activate the interaction of T-cells and B-cells. The CD40–CD40 ligand interaction is particularly important in the TD antibody response. Blocking this molecular interaction also blocks the TD antibody response.

B-cells stimulated by TI and TD antigens may undergo different pathways of cellular activation and differentiation, that is, germinal center reactions and antibody-secreting cell responses. TD antigens induce both germinal centers (GC)s and ASCs. Many critical molecular and cellular events take place in GCs: clonal selection and differentiation of B-memory cells, IgH-class switching, somatic hypermutation and affinity maturation, and the production of precursors of ASCs. By contrast, most TI antigens induce an ASC response without GC induction. Thus, the typical TI-antigen response fails to induce memory cells, somatic mutation, and class switching (IgM to IgG). As a consequence, antibodies to polysaccharides and other T-independent antigens are generally in germ-line gene configurations without somatic mutations in their V regions. These antibodies were thought to be natural antibodies of relatively low affinity and are frequently isotype and idiotype restricted.

5.2 Complexity of Anticarbohydrate Responses

Given the structural diversity of carbohydrate antigens and the presence of families of cellular receptors with carbohydrate binding activities (lectins), the complexity of anticarbohydrate immune responses *in vivo* must be emphasized. This point was highlighted by a series of immunological studies on two microbial polysaccharides, $\alpha(1\rightarrow6)$ dextran (N279) and $\alpha(1\rightarrow3)\alpha(1\rightarrow6)$ dextran (B1355S).

Dextran B1355S and N279 are produced by different strains of *Leuconostoc mesenteroides*. Structurally, they differ only in their composition of glycosidic linkages. N279 is the simplest dextran with $\alpha(1\rightarrow6)$linked internal linear chains predominating; B1355S is the prototype of dextrans having significant proportions of non-$\alpha(1\rightarrow6)$linkages. Therefore, the tertiary structure of B1355S differs strikingly from that of N279. The predominant antigenic determinant of B1355S is the linear backbone of alternating $\alpha(1\rightarrow3)\alpha(1\rightarrow6)$linkages. $\alpha(1\rightarrow6)$ dextran has type D conformation, as a flexible coil in solution. However, by insertion of $\alpha(1\rightarrow3)$linkages into a $\alpha(1\rightarrow6)$linked linear chain, this structure may be completely altered. The backbone of alternating $\alpha(1\rightarrow3)\alpha(1\rightarrow6)$linkages is much more rigid than the $\alpha(1\rightarrow6)$ glycosidic internal linear chain.

When the two glucose-composed polysaccharides were used to immunize mice, distinct patterns of B-cell responses were induced. These include the repertoires of their specific antibodies and the pathways of splenic B-cell activation. $\alpha(1\rightarrow3)\alpha(1\rightarrow6)$ dextran elicits a highly restricted VHJ558/λ1 dominant response, while the $\alpha(1\rightarrow6)$ dextran responses are highly diverse. Moreover, unlike $\alpha(1\rightarrow3)$

$\alpha(1\rightarrow6)$ dextran, which induces only an ASC response, primary immunization with $\alpha(1\rightarrow6)$ dextran elicits both an ASC response as well as the GC reaction. In addition, $\alpha(1\rightarrow6)$ dextran but not $\alpha(1\rightarrow3)\alpha(1\rightarrow6)$ dextran is able to elicit an antigen-specific IgA response in a strain of mice that lack functional T-cells. This strain was generated by knocking out the genes encoding T-cell receptor β and δ, resulting in a complete deletion of T-cells in these animals.

Molecular and cellular mechanisms underlying the above observations are largely unknown. Polysaccharides of distinct physicochemical properties, showing characteristic patterns in their cellular compartment localization have been, however, recognized for some years. In the spleen, macrophages located in various cellular compartments were shown to retain polysaccharides selectively. Acidic polysaccharides were predominantly localized in red pulp macrophages; neutral polysaccharides were detectable exclusively or very intensely in macrophages of the marginal zone of the white pulp. Memory B-cells to T-dependent and T-independent antigens were found in the marginal zone. Follicular localization of some polysaccharides was shown to be complement dependent. Lipopolysaccharides (LPS) may interact with host cells in a number of diverse ways, as B-cell polyclonal activators or as T-independent antigen performing specific stimulations. The former is believed to be associated in their interactions with macrophages. Lipoarabinomannan of mycobacterial pathogens exhibits a wide spectrum of immunoregulatory functions, such as inhibition of interferon-mediated activation of murine macrophages, the scavenging of potential cytotoxic oxygen-free radicals, inhibition of protein kinase C activity, and evocation of a large array of cytokines associated with macrophages. This mycobacterial lipoglycan mediates the production of macrophage-derived cytokines, which, in turn, may evoke many of the clinical manifestations of tuberculosis and leprosy.

The above two dextrans are neutral polysaccharides but show differences in their cellular compartment localization. Dextran B1355S [$\alpha(1\rightarrow3)\alpha(1\rightarrow6)$ dextran] was detectable in splenic marginal zones of the white pulp, but not in GCs. Dextran N279 [$\alpha(1\rightarrow6)$ dextran] persisted, however, in both splenic germinal centers and marginal zone. Therefore, their corresponding B-cell clones were incubated *in vivo* in different microenvironments and were interacting with different types of cells, such as macrophages, follicular dendritic cells, and so on. Expression of lectin-like receptors by some of these cells has been recently documented. For example, mannose binding lectin (MBL) is expressed by some dendritic cells, which can bind dextrans. Involvement of distinct lineages or subtypes of B-cells is another important factor that may influence the pathway of B-cell activation in an immune response. Förster and Rajewsky demonstrated that in BALB/c mice, the dominant $\lambda1$ response to B1355S $\alpha(1\rightarrow3)\alpha(1\rightarrow6)$ dextran is derived, at least in part, from cells of the B-1 lineage. A broad repertoire of anti-$\alpha(1\rightarrow6)$ dextrans, associated with multiple V_H and V_L genes and solely with κ L chains, suggests, perhaps, that different lineages or subsets of B-cells are involved in anti-$\alpha(1\rightarrow6)$ dextran responses.

5.3 Delayed Maturation of Anticarbohydrate Responses

The delayed maturation of antibody response is a common characteristic of TI

antigens. The time required to develop full responsiveness to TI stimulations differs significantly among antigenic systems studied. In mice, antibody responses to TI-I antigens (LPS and other) and protein antigens reach adult levels within 1 to 2 weeks; antibodies to TI-II antigens can be detected only at 2 to 3 weeks. For pneumococcal polysaccharides SSS-III and dextran B1355S, full development of the antibody response is not reached until 4 weeks; with polysaccharides levan and $\alpha(1\rightarrow6)$ dextran, 7 and 13 weeks respectively are needed. In humans, children younger than 18 months fail to respond to microbial polysaccharides or produce antibodies at levels too low to be protective. Such poor responsiveness generally lasts until 5 years of age. Thus, there is a period when maternally derived protective antibodies have declined; yet the age-related development of immunity to bacterial infection remains immature. Pathogens causing severe problems during this high-risk period have long posed the need for developing efficient vaccines.

The underlying mechanism for the age delay in antipolysaccharide responses is still poorly understood. It has been illustrated that two distinct B-cell lineages/subtypes – B-1 and conventional B-cell – have different kinetics of B-cell development and functional maturation. It is interesting to see whether the delayed maturation of antibody responses to TI antigens coincides with B-1 or B-2 cell maturation. B-1 cell lineages (B-1a and B-1b) develop earlier than do conventional B-cells. By 4 weeks after birth, the mouse B-1 cell population approaches adult size and is able to mount a normal level of antibody response. Conventional B-cells reach maximal levels at 12 to 14 weeks of age. By comparing the kinetics of B-cell development and the timing of maturation of anti-TI antigen responses, it is clear that the delayed ontogeny of antibody response to some, but not all TI antigens, may be attributed to the time required for B-1 cell maturation.

However, an immunization strategy was successfully developed to improve childhood vaccination. Several important molecular events elicited by T-dependent antigens are lacking in the polysaccharide induced antibody responses, including memory cell induction, affinity maturation, and Ig H-chain class switching to IgGs. These events may occur in GCs of the peripheral lymphoid organs during T-dependent immune responses. Thus, efforts have been made to convert carbohydrate antigens into T-dependent antigens. Studies with T-dependent forms of $\alpha(1\rightarrow6)$ dextran, made by coupling isomaltose oligosaccharides to protein carriers, experimentally support this strategy. This conversion resulted in a significant shift in the age at which the maximum anti-$\alpha(1\rightarrow6)$ dextran was obtained. With native dextran (TI-II), mouse antibody responses reach adult levels in about 12 to 13 weeks; with the T-dependent form, the peak is at 3 to 4 weeks. V region sequence data and combining site mapping established that the T-dependent repertoire of anti-$\alpha(1\rightarrow6)$ dextran differs from the T-independent repertoire, indicating that distinct B-cell populations were elicited by the T-dependent conjugates.

Robbins and Schneerson (1990) introduced the conjugate strategy for human vaccination. By coupling purified capsular polysaccharides of *Haemophilus influenzae* B with tetanus toxoid, the first polysaccharide–protein conjugate against this bacterium was made. This conjugate vaccine has been applied in over 40 countries and is included in the Extended Program on Immunization of the World Health

Organization (W.H.O). Instead of large capsular polysaccharides, its oligosaccharide was conjugated with different protein carriers for vaccinations. Age-related shifting of antibody responses and Ig class switching to the protective IgG isotypes were observed with these vaccines, resulting in better protection of high-risk populations from *H. influenzae* B. Since these conjugate vaccines have become available, meningitis and other systemic *H. influenzae* B infections have virtually disappeared. The principles underlying the approaches to *H. influenzae* B vaccines have been extended to microbial polysaccharides of other bacteria, viruses, parasites, and so on.

6 Future Prospects

Recognition of carbohydrates as antigens began with the study of microbial polysaccharides. In 1917, Dochez and Avery found that when *Pneumococci* were grown in fluid media, there was a substance in the culture fluid that precipitated specifically with antisera to the same *Pneumococcus*. Heidelberger and Avery showed that this substance was a polysaccharide and not a protein, as previously thought. In the past few decades, a large amount of valuable information regarding the structural and immunological properties of various classes of carbohydrate antigens has been accumulated. In the new millennium, this field faces serious challenges as well as remarkable new opportunities.

The genome-sequencing project has reached a conclusion that only about 30 000 genes in the human genome must account for the complexity of the human organism. This finding emphasizes the importance of posttranslational protein modifications in modulating the biological activities of proteins and their cellular functions. Glycosylation is one of the most important protein posttranslational modifications, which introduces sugar chains of various structures to a newly synthesized protein molecule at given positions. Unraveling the biological roles of carbohydrate diversity attracts scientists from many scientific disciplines. An interesting analysis that appeared recently is about the possible contribution of carbohydrate chains to the neural network in the brain (Dr. Yoshitaka Nagai, Mitsubishi Kagaku Institute of Life Sciences). It is estimated that 10^4 synaptic contacts are present for a single neuron. The total number of neurons in a human brain is as large as 1.4×10^{10}. The information content of such a network system in the brain is, thus, estimated to be on the order of more than 10^{14}. This is far beyond the information carried by the human genome at the DNA sequence level, which is around the order of 3×10^9 to 10^{10}. Recent investigations using gene-targeting technology were able to generate animals with selective genetic defect. For example, a single glycosyltransferase can be targeted and destroyed. Several groups of scientists investigated such genetically engineered animals by all available means and from different scientific aspects. They systematically investigated not only the tissue expression of sugar chains but also the learning ability and behavior of an animal. Evidences obtained begin to suggest a link between specific glycosyltransferase and certain functional disorders in the brain. Perhaps, such comprehensive investigation reflects the characteristics of a new scientific discipline – glycomics or systemic glycobiology.

The worldwide epidemic spread of HIV infections and potential bioterrorism attacks have posed serious challenges to the entire scientific community. High-priority infectious agents that are current risks to our national security include multiple microbial pathogens. These agents are listed as category A, B, and C pathogens by the Center for Disease Control (CDC) of the United States. The genome-sequencing projects have uncovered the whole genome sequences of a number of human pathogens and are rapidly reaching completion of genome-sequencing for others. In the postgenomic era, it is important for us to learn to take advantage of the enriched genome sequence information, novel research concepts, and high throughput postgenomic technologies to benefit carbohydrate research. Professor Peng George Wang's group of Wayne State University has illustrated one of the excellent examples of such an effort. This group transferred genes of the Galactose biosynthetic cycle into an *E. coli* strain and was able to use this strain to produce Galactoside on a large-scale. Another example is, perhaps, the recent development of carbohydrate microarrays. In the year 2002, carbohydrate microarrays of different types were reported in scientific journals. In principle, these technologies are able to display a large repertoire of carbohydrates on a surface and thereby allow detection of a broad range of carbohydrate–protein interactions in a single assay. A platform of carbohydrate microarrays that was developed by the Columbia University Genome Center was designed to take advantage of the state-of-the-art technology of cDNA micoarray and to display different classes of carbohydrate antigens, including polysaccharide and glycoconjugates of different structural configurations. This technology has achieved the sensitivity to recognize the profiles of human anticarbohydrate antibodies with as little as a few microliters of serum specimen and reached the chip capacity to include the antigenic preparations of most common pathogens (~20 000 microspots per biochip). Substantial efforts must be made, however, to extend the repertoires of antigenic diversity and complexity of carbohydrate microarrays and to further improve the methods for chip production, microarray scanning, and data processing. Considerable bioinformatic efforts are also critically needed to enable the clinical application of diagnostic carbohydrate microarrays.

Acknowledgment

I am very grateful to Dr Shaoyi Liu, Dr Jose Morales, Ms Ellen Gluzman, and Mr Brian Trummer for their excellent technical assistance and many helpful discussions in preparing this manuscript and proofreading of this manuscript. This work was supported by grants from the National Institute Health and Compass Pacific, Ltd, to D. Wang.

See also Bioorganic Chemistry.

Bibliography

Books and Reviews

Basten, A., Howard, J.G. (1973) Thymus independence, in: Davies, A.J.S. (Ed.) *Contemporary Topics in Immunobiology*, Vol. 2, Plenum, New York, pp. 265–300.

Brooks, S.A., Dwek, M.V., Schumacher, U. (2002) *Functional & Molecular Glycobiology*, BIOS Scientific Publishers, Oxford.

Durie, F.H., Foy, T.M., Masters, S.R., Laman, J.D., Noelle, R.J. (1994) The role of CD40 in the regulation of humoral and cell-mediated immunity, *Immunol. Today* **15**(9), 406–411.

Feizi, T. (2000) Carbohydrate-mediated recognition systems in innate immunity, *Immunol. Rev.* **173**, 79–88.

Hakomori, S. (1985) Aberrant glycosylation in cancer cell membranes as focused on glycolipids: overview and perspectives, *Cancer Res.* **45**(6), 2405–2414.

Herzenberg, L.A., Herzenberg, L.A. (1989) Toward a layered immune system, *Cell* **59**, 953–954.

Laman, J.D., Claassen, E. (1996) T-cell-independent and T-cell-dependent humoral immunity, in: Snapper, C.M. (Ed.) *Cytokine Regulation of Humoral Immunity*, John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore, pp. 24–72.

Loris, R. (2002) Principles of structures of animal and plant lectins, *Biochim. Biophys. Acta* **1572**(2–3), 198–208.

Mond, J.J., Lees, A., Snapper, C.M. (1995) T cell-independent antigens type 2, *Annu. Rev. Immunol.* **13**, 655–692.

Wang, D. (2000) T-independent IgA responses to microbial polysaccharides, in: Wu, A.M. (Ed.) *Molecular Immunology of Complex Carbohydrates 2*, Kluwer Academic/Plenum, New York, Boston, Dordrecht, London, Moscow.

Wang, D., Kabat, E.A. (1996) Carbohydrate Antigens (Polysaccharides), in: Regenmortal, M.H.V.V. (Ed.) *Structure of Antigens*, Vol. 3, CRC Press, Boca Raton, New York, London, Tokyo, pp. 247–276.

Wang, D., Kabat, E.A. (1998) Antibodies, specificity, in: Roitt, D.A. (Ed.) *Encyclopedia of Immunology*, Academic Press, London, pp. 148–154.

Primary Literature

Adachi, M., Hayami, M., et al. (1988) Expression of Le^Y antigen in human immunodeficiency virus-infected human T cell lines and in peripheral lymphocytes of patients with acquired immune deficiency syndrome (AIDS) and AIDS-related complex (ARC), *J. Exp. Med.* **167**(2), 323–331.

Bundle, D.R., Baumann, H., Brisson, J.R., Gagne, S.M., Zdanov, A., Cygler, M. (1994) Solution structure of a trisaccharide-antibody complex: comparison of NMR measurements with a crystal structure, *Biochemistry* **33**(17), 5183–5192.

Chen, X., Zhang, J., Kowal, P., Liu, Z., Andreana, P.R., Lu, Y., Wang, P.G. (2001) Transferring a biosynthetic cycle into a productive Escherichia coli strain: large-scale synthesis of galactosides, *J. Am. Chem. Soc.* **123**, 8866–8867.

Cisar, J., Kabat, E.A., Dörner, M.M., Liao, J. (1975) Binding properties of immunoglobulin combining sites specific for terminal or nonterminal antigenic determinants in dextran, *J. Exp. Med.* **142**, 435–459.

Cygler, M., Rose, D.R., Bundle, D.R. (1991) Recognition of a cell-surface oligosaccharide of pathogenic Salmonella by an antibody Fab fragment, *Science* **253**(5018), 442–445.

Feinberg, H., Mitchell, D.A., Drickamer, K., Weis, W.I. (2001) Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR, *Science* **294**(5549), 2163–2166.

Feizi, T. (1982) The antigens Ii, SSEA-1 and ABH are in interrelated system of carbohydrate differentiation antigens expressed on glycosphingolipids and glycoproteins, *Adv. Exp. Med. Biol.* **152**, 167–177.

Finne, J., Leinonen, M., Makela, P.H. (1983) Antigenic similarities between brain components and bacteria causing meningitis. Implications for vaccine development and pathogenesis, *Lancet* **2**(8346), 355–357.

Howard, J. (1987) T cell-independent responses to polysaccharides: their nature and delayed ontogeny, in: Bell, R., Torrigiani, G. (Eds.) *Proceedings of a meeting organized by the World Health Organization*, Geneva, October 9–11 1986, *Towards Better Carbohydrate Vaccines*, John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore, pp. 221–229.

Humphrey, J.H. (1981) Tolerogenic or immunogenic activity of hapten-conjugated polysaccharides correlated with cellular localization, *Eur. J. Immunol.* **11**, 212–220.

Humphrey, J.H., Grennan, D. (1981) Different macrophage populations distinguished by means of fluorescent polysaccharides. Recognition and properties of marginal-zone macrophages, *Eur. J. Immunol.* **11**, 221–228.

Kohler, G., Milstein, C. (1975) Continuous cultures of fused cells secreting antibody

of predefined specificity, *Nature* **256**(5517), 495–497.

Leteux, C., Chai, W., et al. (2000) The cysteine-rich domain of macrophage mannose receptor is a multispecific lectin that recognizes chondroitin sulfates A and B and sulfated oligosaccharides of blood group Lewis[a] and Lewis[x] types in addition to the sulfated N-glycans of lutropin, *J. Exp. Med.* **191**, 1117–1126.

Mandrell, R.E., McLaughlin, R., Aba Kwaik, Y., Lesse, A., Yamasaki, R., Gibson, B., Spinola, S.M., Apicella, M.A. (1992) Lipooligosaccharides (LOS) of some haemophilus species mimic human glycosphingolipids, and some LOS are sialylated, *Infect. Immun.* **60**(4), 1322–1328.

Muller-Loennies, S., MacKenzie, C.R., Patenaude, S.I., Evans, S.V., Kosma, P., Brade, H., Brade, L., Narang, S. (2000) Characterization of high affinity monoclonal antibodies specific for chlamydial lipopolysaccharide, *Glycobiology* **10**(2), 121–130.

Parkhill, J., Wren, B.W., et al. (2001) Genome sequence of Yersinia pestis, the causative agent of plague. *Nature* **413**, 523–527.

Patenaude, S.I., MacKenzie, C.R., Bilous, D., To, R.J., Ryan, S.E., Young, N.M., Evans, S.V. (1998) Production, crystallization and diffraction to atomic resolution of an antibody Fv specific for the blood-group A oligosaccharide antigen, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54**(Pt 6 Pt 2), 1456–1459.

Patenaude, S.I., Vijay, S.M., Yang, Q.L., Jennings, H.J., Evans, S.V. (1998) Crystallization and preliminary X-ray diffraction analysis of antigen-binding fragments which are specific for antigenic conformations of sialic acid homopolymers, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54**(Pt 5), 1005–1007.

Rosati, F., Capone, A., Giovampaola, C.D., Brettoni, C., Focarelli, R. (2000) Sperm-egg interaction at fertilization: glycans as recognition signals, *Int. J. Dev. Biol.* **44**(6), 609–618.

Rose, D.R., Przybylska, M., To, R.J., Kayden, C.S., Oomen, R.P., Vorberg, E., Young, N.M., Bundle, D.R. (1993) Crystal structure to 2.45 A resolution of a monoclonal Fab specific for the Brucella A cell wall polysaccharide antigen, *Protein Sci.* **2**(7), 1106–1113.

Schneerson, R., Barrera, O., Sutton, A., Robbins, J.B. (1980) Preparation, characterization, and immunogenicity of Haemophilus influenzae type b polysaccharide-protein conjugates, *J. Exp. Med.* **152**(2), 361–376.

Sharon, J., Kabat, E.A., Morrison, S.L. (1982) Immunochemical characterization of binding sites of hybridoma antibodies specific for $\alpha(1 \rightarrow 6)$ linked dextran, *Mol. Immunol.* **19**(3), 375–388.

Thomas, R., Patenaude, S.I., MacKenzie, C.R., To, R., Hirama, T., Young, N.M., Evans, S.V. (2002) Structure of an anti-blood group A Fv and improvement of its binding affinity without loss of specificity, *J. Biol. Chem.* **277**(3), 2059–2064.

Villeneuve, S., Souchon, H., Riottot, M.M., Mazie, J.C., Lei, P., Glaudemans, C.P., Kovac, P., Fournier, J.M., Alzari, P.M. (2000) Crystal structure of an anti-carbohydrate antibody directed against Vibrio cholerae O1 in complex with antigen: molecular basis for serotype specificity, *Proc. Natl. Acad. Sci. U.S.A.* **97**(15), 8433–8438.

Wang, D., Chen, H., Liao, J., Akolkar, P.N., Sikder, S.K., Gruezo, F., Kabat, E.A. (1990) Two families of monoclonal antibodies to α(1,6)dextran, V_H19.1.2 and V_H9.14.7, show distinct patterns of J_κ and J_H minigene usage and amino acid substitutions in CDR3, *J. Immunol.* **145**, 3002–3010.

Wang, D., Liao, J., Mitra, D., Akolkar, P.N., Gruezo, F., Kabat, E.A. (1991) The repertoire of antibodies to a single antigenic determinant, *Mol. Immunol.* **28**(12), 1387–1397.

Wang, D., Liu, S., Trummer, B.J., Deng, C., Wang, A. (2002) Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells, *Nat. Biotechnol.* **20**(3), 275–281.

Wang, D., Wells, S.M., Stall, A.M., Kabat, E.A. (1994) Reaction of germinal centers in the T-cell-independent response to the bacterial polysaccharide $\alpha(1 \rightarrow 6)$ dextran, *Proc. Natl. Acad. Sci. U.S.A.* **91**(7), 2502–2506.

Weissman, I.L., Gutman, G.A., Friedberg, S.H., Jerabek, L. (1976) Lymphoid tissue architecture. III. Germinal centers, T cells, and thymus-dependent vs thymus-independent antigens, *Adv. Exp. Med. Biol.* **66**, 229–237.

Wood, C., Kabat, E.A. (1981) Immunochemical studies of conjugates of isomaltosyl oligosaccharides to lipid. I. Antigenicity of the glycolipids and the production of specific antibodies in rabbits, *J. Exp. Med.* **154**(2), 432–449.

Young, A.C., Valadon, P., Casadevall, A., Scharff, M.D., Sacchettini, J.C. (1997) The three-dimensional structures of a polysaccharide binding antibody to cryptococcus neoformans and its complex with a peptide from a phage display library: implications for the identification of peptide mimotopes, *J. Mol. Biol.* **274**(4), 622–634.

Zdanov, A., Li, Y., Bundle, D.R., Deng, S.J., MacKenzie, C.R., Narang, S.A., Young, N.M., Cygler, M. (1994) Structure of a single-chain antibody variable domain (Fv) fragment complexed with a carbohydrate antigen at 1.7-A resolution, *Proc. Natl. Acad. Sci. U.S.A.* **91**(14), 6423–6427.

Zopf, D.A., Tsai, C.M., Goinsburg, V. (1978) Carbohydrate antigens: coupling of oligosaccharidephenethylamine derivatives to edestin by diazotization and characterization of antibody specificity by radioimmunoassay, *Methods Enzymol.* **50**, 163–169.

Cardiovascular Diseases, Genetics and Gene Therapy: *see* Gene Therapy and Cardiovascular Diseases

Cell Growth in Microgravity

Alamelu Sundaresan[1], *Diana Risin*[2], *and Neal R. Pellis*[2]
[1] *Division of Space and Life Sciences, Universities Space Research Association (USRA), Houston, Texas, USA* [2] *NASA, Lyndon B. Johnson Space Center, Houston, Texas, USA*

1 Introduction 304

2 Unique Cell Culture Conditions Offered by Microgravity 305
2.1 The Effects of Microgravity on Cellular Organelles and Cellular Function 305

3 Modeling Microgravity 306

4 Cellular Response to Microgravity 310
4.1 Signal Transduction in Microgravity 311

5 The Experience in Microgravity 314

6 Conclusions 316

Acknowledgments 317

Bibliography 317
Books and Reviews 317
Primary Literature 317

Keywords

Microgravity
Gravitational or acceleration forces less than10^{-2} g.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Microgravity Analog Culture System (MACS)
Refers to the NASA series of bioreactors, including the rotating-wall vessel (RWV), slow turning lateral vessel (STLV), and the high aspect ratio vessel (HARV).

Modeled Microgravity
Analog paradigms of microgravity such as drop tower, suborbital rockets, MACS, parabolic flight, and so on.

Three-dimensional Cell Growth
Growth and assembly of cells into definite constructs with a third dimension.

T-cell Activation
The process by which T-cell receptor triggering leads to cytokine secretion and cell proliferation.

Microgravity affects cells and tissues either through direct response elements within the cell or by influencing the environment in which the cell lives. Cells respond to decreased gravity and the reordering of forces by changes in gene expression and cellular function. This review will address the unique attributes of microgravity that influence cell growth and function, progressing from ground-based models of microgravity, the cellular response to microgravity, and experimental results from cell experiments in microgravity.

1 Introduction

Microgravity is a unique environment in which cells, tissues, and organisms undergo novel adaptations. The adaptational responses provide insight into basic cellular mechanisms and, furthermore, provide opportunities in applied cell science. Investigations in microgravity and ground-based model systems have opened new vistas in tissue engineering, disease modeling, drug testing, vaccine production, and space cell biology. In microgravity, cells no longer sediment to surfaces, but rather freely associate with each other, forming large assemblies that can replicate in three dimensions. No doubt, the diminution of gravity from the cell culture setting allows the remaining forces to be more apparent. The rapid assembly and three-dimensional growth afford insight into the nature of physical forces that affect life at the cellular level and other new prospects in tissue engineering. Microgravity tissue morphogenesis produces functional three-dimensional constructs that are useful experimental models and may ultimately be amenable to transplantation. The reordering of forces in microgravity is reflected in the profound shape change, gene expression, differentiation, and metabolic alterations identified in space experiments. Despite the dramatic

response, cells not only survive, but also thrive to give a new perspective to biological systems.

2 Unique Cell Culture Conditions Offered by Microgravity

As life evolved on Earth, a multiplicity of physical and chemical factors invoked adaptations and participated in the complicated selection process. For many factors, there are clear examples of the role of physical forces in determining the pathways in evolution. A notable exception is gravity. The force of gravity has been constant for the 4.8 billion years of life evolutionary processes on Earth. Therefore, there is little or no anticipated genetic memory for terrestrial life to respond to gravitational force changes. As we transition terrestrial life to low gravity environments and study the adaptive processes in the cell, we will increase our understanding of gravity's role in shaping life on Earth.

The species on Earth evolved to perform various functions and processes in a gravitational force field of $9.8\ \mathrm{m\ s^{-2}}$. Some gravity-dependent processes, such as balance and the ability of higher organisms to proceed along a surface, are obvious. Others, such as the directional growth of plants (gravitropism), are less obvious but are demonstrated through experimentation. For most species (microbial, plant or animal), only a small proportion of the functions influenced by gravity are known. During the past 40 years, the ability to achieve Earth orbit and experience long-term microgravity has created a need to understand the role of gravity in human, animal, and plant systems in preparation for exploration class missions into the solar system.

2.1 The Effects of Microgravity on Cellular Organelles and Cellular Function

Many biological stresses are induced by spaceflight and are visible in many areas of human physiology. The most dramatic response to spaceflight is the significant loss of bone and muscle mass and function, most likely caused by skeletal unloading and reduced activity. In addition, cephalad fluid shifting, neurovestibular disturbances, and a general malaise termed *space sickness* are experienced by astronauts soon after arrival in microgravity. In many instances, the basis of the negative influence of microgravity on human systems may be investigated using cell-based systems. This communication will offer a selected review of the past work at the cellular level and invoke models for testing in forthcoming opportunities on the International Space Station.

A major challenge in studying the adaptation of terrestrial life to space conditions resides in convincingly delineating the responses induced by microgravity from the radiation and stress experienced in spaceflight. Various *in vitro* parameters of immune function are suppressed in humans during and after returning from space missions. Exposure to cosmic radiation, along with physical and psychological stresses occurs concurrently with the microgravity of spaceflight. Understanding the effect of each of these factors in the physiological changes observed during spaceflight is a formidable challenge for space research programs (e.g. for immune function, which is affected independently by all three factors: microgravity, stress, and radiation).

It is unlikely that single cells can "sense" gravity through biophysical changes within the cell. Three theoretical approaches

address this problem and are reviewed by Cogoli and Cogoli–Greuter: (1) a direct effect: the direct interaction of gravity with one or more cellular organelles of a density different from that of the cytoplasm generates a pressure on neighboring structures (e.g. the cytoskeleton) and consequently a signal that is transduced into a biological event; (2) a nonequilibrium thermodynamic effect: the interaction of gravity with a few organelles is not sufficient to trigger one event, but a series of small changes is amplified to generate an important effect; and (3) an indirect effect: alterations in gravity induce changes in the microenvironment of the cell and the cell responds to the new environmental conditions.

Studies by Pollard suggest that in cells smaller than 10 μm in diameter, Brownian motion caused by impacts from molecules in the culture fluid are sufficient to counteract gravity-induced sedimentation. How much sedimentation can occur in cells larger than 10 μm is not known. Furthermore, this approach assumes that cells act like a bag of water containing solutes and undissolved particles (starch granules, oil vacuoles, and organelles). This theory does not agree with the observations that the aqueous phase of the cell is heterogenous, containing solution properties, colloidal regions, suspended and attached organelles, and contractile cytoskeletal structures. Studies by Nace suggest that organelles of the cell may impart significant torque on the cytoskeleton using the force of gravity. It is not known what effect the microgravity-related loss of this torque has on cellular physiology. Nevertheless, the possibility of a direct gravity sensor within mammalian cells exists although it is not known as yet.

The likelihood that gravity, a small force, would have an effect on cell function is questionable. The nonequilibrium thermodynamic effect, or the bifurcation theory, provides the reasoning to allow such small changes to exert a significant effect on biochemical processes of the cell. This theory, outlined by Mesland, argues that the biochemical processes of a cell are described in terms of nonlinear, nonequilibrium thermodynamics with chaotic outcomes. Therefore, the direction in which these reactions proceed is extremely sensitive to environmental conditions. This sensitivity may allow the seemingly small effect of gravity changes to have dramatic effects on cellular processes.

Microgravity can possibly alter the biophysical microenvironment around cells, and in turn, affect the shape, metabolism, and function of the cell. For example, the lack of sedimentation and thermal convection in microgravity leaves only simple diffusion and possibly surface tension–driven phenomena to move waste products away from the cell and bring nutrients toward the cell. Depending on the metabolic rate of the cell, diffusion and surface tension may not be sufficient to fulfill the metabolic requirements of the cell. These and other microgravity-induced alterations in the physical and chemical microenvironment alter cellular function.

3 Modeling Microgravity

Modeling microgravity for investigation of cell biological phenomena is a challenge that the space agency and scientists address in many familiar venues: parabolic flight, suborbital rocketry, drop towers, and neutral buoyancy. The former three strategies provide only short duration analog conditions, while the latter permits extended observations. A novel model emerged from early attempts to keep

cells suspended without invoking excessive hydrodynamic shear while providing abundant mass transfer. The rotating-wall vessel (RWV) bioreactors developed at NASA's Johnson Space Center (Fig. 1) model some aspects of microgravity and allow the extensive analysis of microgravity-related biological phenomena at the cellular level. We now refer to these culture vessels as the microgravity analog culture system (MACS). MACS is based on a previous clinostat design and consists of a zero headspace cylindrical culture vessel that rotates at a slow speed around a horizontal axis. These conditions compensate the gravity vector by rotation and create a solid body fluid rotation that can suspend cells or small particles in a sustained free fall condition, modeling conditions of microgravity for cell cultures.

This system provides us with the ability to assess function at a cellular level, without the potential confounding influences of psychoneuroendocrine factors. Furthermore, the cells are analyzed in real time while in analog microgravity in addition to a postflight or postanalog. Initially, the MACS demonstrated suppression of lymphocyte locomotion through collagen in ground-based experiments. Subsequent flight experiments conducted onboard Shuttle missions STS-54 and STS-56 confirmed suppression of lymphocyte locomotion suggested by MACS. This was

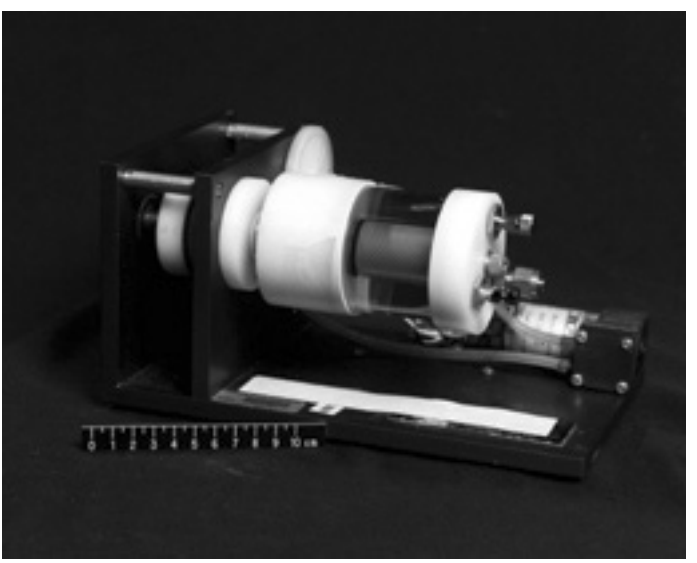

Fig. 1 The NASA rotating bioreactor allows for solid body fluid rotation. The growth medium completely fills the cylindrical vessel. Rotation of the vessel along its horizontal axis suspends the cells without stirring, randomizes orientation to the *g* vector, and renders cells to continuous fall at terminal velocity through the culture fluid. The cells have less interaction with inert surfaces and experience greater freedom for three-dimensional association. Several derivatives of the rotating bioreactor are available commercially (Synthecon, Inc.). The advantages lie in the low shear environment and the potential space analog condition.

a first in a series of experiments that demonstrated shared characteristics between orbital microgravity and rotational culture. The MACS provides for a unique opportunity to investigate the cellular responses as candidates for flight experiment opportunities. As with many models in science, a proportion of results are shared with the results from actual conditions in microgravity.

Hind-limb suspension is a common ground-based microgravity model used for experiments with rodents. Antiorthostatic suspension of rodents provides a 30° head-down tilt, unloading the hind limbs and weight bearing the front limbs. This position provides the cephalad-fluid shift observed in astronauts during spaceflight and mimics the arm-priority locomotion of astronauts. Studies in this model demonstrated organ-specific immune changes. Lymphocytes from lymph nodes and the peripheral blood of suspended animals had a reduced response to phytohemagglutinin (PHA) stimulation, whereas spleen cells of suspended animals displayed enhanced response. Other suspension experiments, as well as spaceflight studies, support this organ-specific response. In contrast, recent antiorthostatic suspension experiments reported enhanced responses in animals. Antiorthostatically suspended mice had enhanced resistance to primary infection and an enhanced protective immunological memory to *Listeria monocytogenes*. Mechanisms for this enhancement are not known, but the results indicate that both suspension and spaceflight cause complex immune alterations rather than a simple shut down of the immune system. Indeed, some effects are related to the altered orientation to gravity, while others may be the result of stress induced by the suspension model.

The effect of psychoneuroendocrine changes on immune-response studies is minimized by performing experiments on isolated cell populations in controlled culture conditions. Ground-based models of microgravity enable refining of hypotheses in advance of flight. The response of cells to microgravity is complex and difficult to identify if the cell is responding to the change in gravity or to a culture condition that is created by microgravity. In the former resides the posit that cells possess a "gravity sensor." No one has as yet described a gravity sensor in mammalian cells, despite numerous studies demonstrating that mammalian cells respond to changes in gravity. In contrast, cells in the roots of plants have geotropic organelles that identify the direction of the gravity vector and promote root propagation that follows gravity.

Microgravity affects fluid systems such that there are no density-driven phase separations, no particle sedimentation, and no gravity-driven convection. Absence of these three conditions in cell culture results in a cell culture environment that is stressful or even hostile to many types of cells. In microgravity, and without facilitation, cell cultures rapidly become diffusion-limited, with regard to nutrient assimilation and waste-product dissipation. Therefore, the cellular response in microgravity may be related to these adverse culture conditions and not directly to microgravity.

Einstein described gravity in terms of a curvature of space–time in his theory of general relativity. In relation to biological systems, gravity is simply an acceleration force (9.8 m s^{-2} on Earth, G). Astronauts in orbit typically experience acceleration forces of 10^{-2} to 10^{-4} G. The force of gravity at the distance that most orbits take place in is practically the same as it is at

sea level. Microgravity is experienced by the astronauts because they are in free-fall around the planet. This is similar to the microgravity experienced during free-fall achieved in microgravity analogs conducted on Earth, for example, in drop towers (Fig. 2), parabolic aircraft (Fig. 3), and sounding rockets (Fig. 4).

To achieve the same level of microgravity by distance alone that is achieved by free-fall in orbit, the astronaut would have to be 17 times as distant from the Earth as

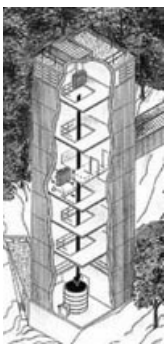

Fig. 2 Drop Tower. Drop towers offer a way to conduct 5- to 60-s microgravity analog experiments in free-fall. This analog is better suited for fluid characterization of the culture environment and rapid cellular events such as shape change and signal transduction.

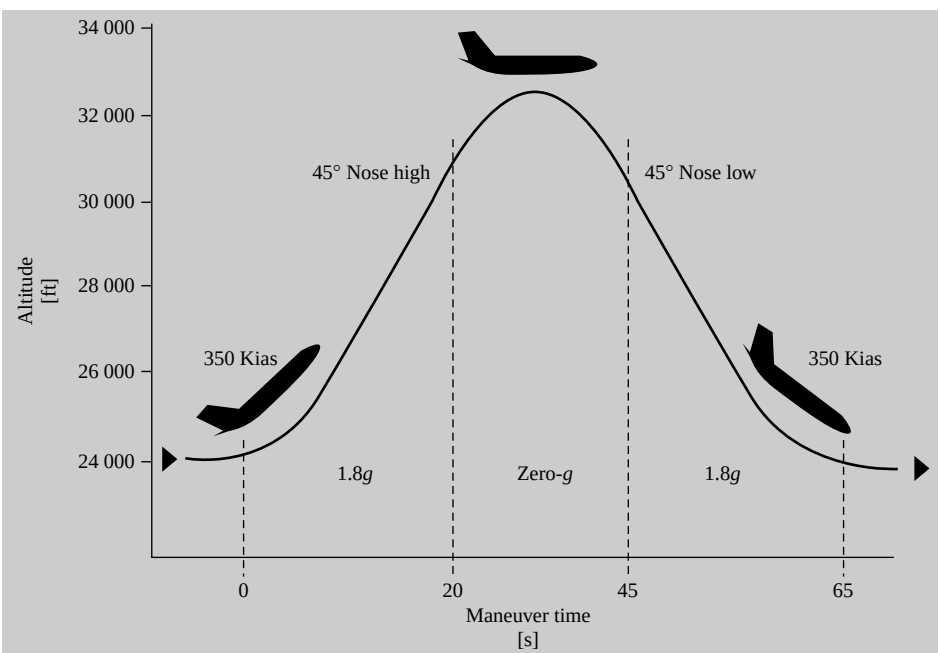

Fig. 3 Parabolic Flight. Parabolic flight, where a specially modified jet alternates steep climbs with long drops (much like a roller coaster), and generates 20 to 30 s of analog microgravity. A typical mission lasts 2 to 3 h and consists of 30 to 40 parabolas. Multiple parabolas are flown, allowing repeat experiments to be performed under the same conditions. Flight maneuvers can be modified to provide any *g*-force level less than 1.

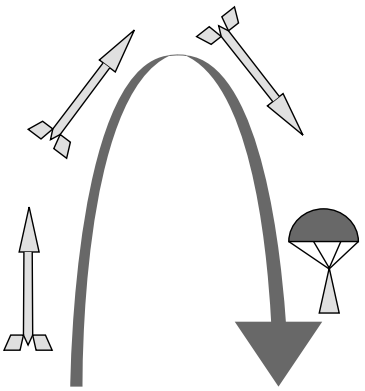

Fig. 4 Suborbital Flight. Suborbital rockets release a research payload at high altitude and offer 4 to 12 min of free-fall. Investigators can study cell movement, signal transduction, and characterization of fluid dynamics in the culture environment.

is the moon. Cell cultures can experience microgravity in space (orbit) or in models for simulating microgravity on Earth: clinostatic culture, parabolic flight (aircraft or sounding rocket), and drop towers. Drop towers only provide very short durations of microgravity (3 to 4 s) and are of limited use for biological experiments. Parabolic flight can provide somewhat longer durations of microgravity (aircraft: 15–35 s, sounding rocket: 7–15 min). Cell cultures can be maintained in clinostatic rotation indefinitely depending on the culture system used. Thus, there are analog systems that can be used to refine hypotheses for spaceflight. Each of these systems shares some characteristics with microgravity. It is essential to understand both the advantages and limitations of model systems. The main advantage for cell science is that many experiments can be modeled for extended periods on the ground using one or more of these analogs.

4
Cellular Response to Microgravity

More than 120 experiments have been performed in space during the past 15 years that demonstrate that microgravity induces changes in single cells. The opportunity to conduct experiments using cells in microgravity and the development of ground-based models provide *in vitro* models to study adaptation to the conditions in microgravity. Cellular responses to microgravity are varied, depending on cell type and species of origin. Lymphocytes from humans, monkeys, and rodents are affected by gravitational unloading. The list spans from the plant kingdom all the way across the animal phyla. The different parameters affecting cells seem to depend on their size, dynamics, shape, circadian rhythms, function, location, and so on. Mammalian cells are especially susceptible to microgravity-related perturbations. In the immune system, there is extensive evidence that microgravity suppresses the immune function *in vitro* and possibly *in vivo*. The outcome of the suppressed functions in long duration space travel is unknown. Nevertheless, it is of concern that widely accepted parameters of immune function are compromised in microgravity. The underlying mechanisms of dysfunction in microgravity are the subjects of flight and ground-based experiments.

4.1
Signal Transduction in Microgravity

An important phase of the microgravity-induced perturbations in transmembrane signaling is selective and not random. Most cells adapt to the new signal motifs and are able to survive. Several studies, beginning with Spacelab D-1 in 1985, SLS-1 in 1991, IML-1 in 1992, IML-2 in 1994, STS-55, -56, and many sounding rocket experiments from the mid-1980s to the mid-1990s, attest to the ability of cells to adapt to change induced by microgravity environments. The flight experiments, and ground-based studies using microgravity analog culture, suggest that single cells are sensitive to changes in gravity, as reflected by proliferation, gene expression, locomotion, and receptor-mediated signal transduction and differentiation.

Mammalian immune performance is critically tied to the function of thymus-dependent (T) lymphocytes. A hallmark of immune performance is the response of T-lymphocytes to activation through cell-surface receptors. The first *in vitro* T-cell activation experiments were performed aboard Spacelab 1 in 1983 by Cogoli et al. (Table 1). Results showed a dramatic suppression of T-cell activation in response to the widely used polyclonal activator, Concanavalin A. Spacelab D1 in 1985 and Spacelab SLS1 in 1991 continued the investigation of T-lymphocytes. The experiments confirmed the inhibition of T-cell activation in microgravity and raised the possibility of direct and indirect effects of gravity unloading on the T cells. The first signal required for T-cell activation is the presentation of antigen by the antigen-presenting cell, which activates a G-protein and phospholipase C. In the SLS-1 flight, it was found that production of the cytokine interleukin 1 (IL-1) expression was impaired in the monocytes. Interleukin-1 is a second signal that is necessary for T-cell activation. The subsequent signal to achieve full T-cell activation is the synthesis of the cytokine Interleukin 2 (IL-2) and the expression of its receptor (IL-2R). The first signal of antigen presentation is modeled *in vitro* using Concanavalin A (a plant lectin) that binds to saccharide groups on the T-cell antigen receptor. An interesting observation from this study was that the binding of Concanavalin A to the cell membrane and subsequent events, such as capping and patching of membrane proteins, were not altered in microgravity. SLS-1 experiments showed that monocytes present in the preparations flown did not release Interleukin-1 (IL-1). Therefore, it was hypothesized that the reason for the inability of T cells to activate in microgravity was the absence of IL-1 production by monocytes and hence, absence of the accessory signal. However, the production of interferon-gamma (IFN-γ) was slightly but significantly restored. The receptor for IL-2 was secreted in adequate amounts in lymphocyte supernatants in microgravity. Importantly, the absence of activation was not due to a lack of cell–cell contact.

It is known that spaceflight conditions and microgravity *per se* alters immune function. Indeed, the altered immune function may be the summation of the individual influence of stress (psychological and physical), cosmic radiation, and microgravity. Thus, the space environment may affect immune performance either directly, because of microgravity-induced changes or indirectly, through the stress related to performance and confinement. Therefore, interpretation of spaceflight studies on immunity must delineate the direct and indirect effects of microgravity on cells from the other factors associated with spaceflight.

Lymphocyte locomotion through interstitium is a critical stage in the immune response to antigenic challenge. Initially, the MACS was used to demonstrate the suppression of lymphocyte locomotion through collagen in modeled microgravity. Locomotion can be investigated *in vitro* using gelled type I collagen as a test matrix. In analog culture, lymphocytes completely lose their capacity to randomly locomote in gelled collagen.

Experiments conducted in microgravity onboard Shuttle missions STS-54 and STS-56 demonstrated similar suppression of lymphocyte locomotion and support the relationship between orbital microgravity and analog culture.

In a study by Cooper and Pellis, the MACS was used to further characterize the suppression of polyclonal activation of T cells with PHA (phytohemagglutinin). Results suggest that accessory cell costimulation in the form of cytokines is intact in the MACS. Furthermore, the calcium-signaling pathway may be intact during PHA stimulation in the MACS. Therefore, signal transduction mechanisms between T-cell receptor engagement and downstream protein kinase C (PKC) activation are sensitive to modeled microgravity. The suppression of lymphocytic PHA response is observed in both the lymphocytes of astronauts after spaceflight and in *in vitro* experiments in spaceflight and modeled microgravity. Although some studies suggest there may be a microgravity-induced PKC defect, T cells were fully activated to proliferate in modeled microgravity by direct activation of PKC using phorbol myristate acetate (PMA) and ionomycin. These pharmacological agents bypass mechanisms at the cell surface and impart their effect inside the cell: PMA by mimicking 1, 2-diacylglycerol (DAG) and activating the serine/threonine kinase PKC; ionomycin by increasing the cytoplasmic-free calcium ion concentrations. The ability to activate T cells in the RWV with PMA and ionomycin implies that all the cellular mechanisms involved in T-cell activation downstream of PKC activation and calcium ion flux are intact in modeled microgravity. Some flight studies indicate that PKC may be unaffected or activated by microgravity. It is still unclear as to how PKC is affected by microgravity and this requires further investigation. These results explain why PHA responsiveness could be partially restored by preincubating PHA-stimulated cultures at normal gravity before exposing them to modeled microgravity. Cells that achieved PKC activation and calcium flux continued on to proliferate even when they were exposed to modeled microgravity. Since PKC activation and calcium flux occur quickly after receptor triggering, one might expect full activation to be restored by advance exposure to PHA cultures. Although these events do occur within minutes, sustained receptor triggering of up to 48 h is required for optimal lectin activation. Once PHA primed cultures were exposed to modeled microgravity, the signal from the receptor was no longer sustained owing to the microgravity-induced lesion somewhere upstream of PKC activation that was observed with the phenomenon of suppression of lymphocyte locomotion in microgravity and modeled microgravity. In addition to polyclonal and oligoclonal models, antigen-specific models of T-cell activation are also suppressed in microgravity and its cell culture analog.

Currently, the results strongly suggest that signaling pathways upstream of PKC are affected in microgravity. An emerging hypothesis is that the microgravity-induced lesion in T-cell activation and lymphocyte locomotion resides between the cell surface and PKC. There is some

reflection of effect on PKC *per se*. Microgravity culture differentially alters the translocation of individual PKC isoforms in monocytes and T cells providing a partial explanation of adverse effects on T-cell activation. A significant variation in the intracellular distribution and quantity of PKC according to *g* levels was also observed in monocytes and peripheral blood lymphocytes aboard a Biorack experiment. These experiments demonstrate the sensitivity of PKC expression to gravitational force. There is no clear demonstration that the PKC changes are a direct consequence of variations in gravity or an indirect outcome of upstream changes as the cell adapts to the new physical environment. Commensurately, a marked decrease in the expression of the calcium-independent PKC isoforms, PKC-δ and PKC-ε, was described in human lymphocytes exposed to modeled microgravity culture. Functional association in microgravity may be determined by specific blocking of PKC isoforms expression followed by assessment of T-cell activation and locomotion both in microgravity and cell culture analog.

Microgravity experiments have been performed on several different mammalian cells. Flight experiments included HeLa cervical carcinoma cells, A431 human epidermoid carcinoma cells, and various immune cells, including primary isolated peripheral blood leukocyte populations and the transformed cell lines of Jurkat leukemic T cells and monocytic THP-1 cells. While HeLa cells proliferate normally in microgravity during orbital flight, other studies revealed responsiveness to hypergravity. At 10, 35, and 70 G in a centrifuge, HeLa cell proliferation and c-*myc* expression were significantly increased. Subsequently, HeLa displayed increased inositol 1, 4, 5-triphosphate (IP_3) levels and decreased cAMP levels. Studies with A431 cells examined the inducible expression of the transcription factor c-*fos*. In both clinostat and sounding rocket studies, the expression of c-*fos* was suppressed when induced by epidermal growth factor (EGF) or phorbol ester but not when induced by a calcium ionophore, A23187, or by forskolin, an activator of protein kinase A (PKA). Conversely, the EGF induced expression of c-*fos* in hypergravity (10 G) was slightly increased. The studies with A431 cells demonstrate that alterations in gravity have specific impacts on signal transduction mechanisms. Experiments with Friend murine erythroleukemic virus transformed cells, WI38 human embryonic lung cells, hamster kidney cells, and L8 rat myoblast cells, demonstrate normal proliferation and growth in microgravity. Thus, despite changes induced by microgravity, many cells adapt and proceed with many functions (i.e. proliferation still intact). In cultured liver cells, formation of multidimensional tissue-like spheroids from primary human liver cells was observed in modeled microgravity. These spheroids were composed of hepatocytes and biliary epithelial cells that arranged as bile duct–like structures along newly formed vascular sprouts.

Several studies addressed the ability of proteins and other macromolecules to interact with each other in microgravity. The binding of (1) EGF to receptors on rat osteosarcoma cells; (2) EGF to and the clustering of EGF receptors on A431 cells; (3) lectin to rhizobia; (4) alpha-fetoprotein to immobilized monoclonal antibodies; and (5) Concanavalin A to lymphocytes were all normal in microgravity. Thus, normal protein–protein interactions occur in microgravity.

5
The Experience in Microgravity

Microgravity and randomized gravity facilitate spatial colocalization and three-dimensional assembly of cells into large aggregates. It is believed that diminution or nullification of gravity results in other forces displaying dominance. More than 25 types of cells have been flown in space. For 20 years, flight experiments have shown that single cells respond to space conditions with changes in cellular morphology and function. Cytoskeletal reorganization occurs early after achieving orbit. Human dermal fibroblasts displayed a significant increase in collagen synthesis aboard Space lab. This might facilitate larger or more differentiated organoids by cell aggregation. Space flight decreases the amounts of osteocalcin and pre-pro-alpha 2 chain of type 1 precollagen mRNA in long bone and calvarial periosteum of rats. Hence, cell types adversely affected in microgravity (such as muscle, bone, and cartilage) may yield inferior types of tissue morphogenesis. Space bioreactors were used for a 137-day cell culture on Mir to generate three-dimensional cartilaginous constructs from chondrocytes, which were preaggregated on resorbable polyglycolic acid scaffolds. In parallel control experiments, the constructs were maintained in analog culture for the same period. In both flight and control cultures, chondrogenesis occurred, yielding many of the histochemical characteristics of native collagen. Flight constructs were mechanically inferior to constructs grown on earth, largely due to a disproportionate loss of one of the matrix components, glycosaminoglycan. Cartilaginous constructs obtained in the modeled microgravity

Tab. 1 Partial list of cells and tissues cultivated in microgravity in space and on the ground.

Cell/tissue	*Environment*	*Observation*
Lymphocytes	Several shuttle missions, Maser sounding rockets	Inhibited locomotion, Impaired immunocompetence, impaired mitogenicity, changes in cytokine production, altered cellular signaling
Chondrocytes	RWV in space (MIR/STS-79 for 4 months)	In space culturing of pre-assembled 3-D aggregates results in constructs which are mechanically inferior to similar aggregates grown in RWV on the ground. Effect might mimic microgravity-induced loss of cartilage
PC12 cells	Six weeks serial passaging of cells in culture bags on MIR Space Station (STS-86)	Establishes feasibility of long term, serial passaging of cells in space, formation of large aggregates with epitheloid morphology
MIP 101 colon carcinoma cells	EDU-1 on STS-70	Increased proliferation, enhanced carcinoembryonic antigen (CEA) production
Osteoblasts	Four days on STS-56 in Materials Dispersion Apparatus minilabs (MDA)	Inhibition of growth, reduction in serum growth activation, changes in microfilament structure

Tab. 1 (*continued*)

Cell/tissue	*Environment*	*Observation*
Myoblasts	Ten days on STS-45 in space Tissue Loss Flight Module "A" (TLMA)	Several permanent phenotypic alterations, including failure to fuse into myotubes
Neonatal rat heart cells	HARV	3-D organization, synchronous beating, isolated heart cells reestablish structural & molecular phenotypes of heart tissue
Rodent skeletal muscle satellite cells	HARV/STLV	Enhance proliferation, 3-D organization, attenuation differentiation
MIP 101 colon carcinoma cells	STLV	Increased proliferation, enhanced carcinoembryonic antigen (CEA) production
Human Ovarian Tumor Cells	STLV	Capability of growth out of the body, organization into differentiated tissue-like constructs, enhanced oncogene expression
Diverse human tumor cell	HARV	Organization of tissue-specific epitheloid structures, enhanced production of cell adhesion molecules
Human prostatic cancer cells	STLV, HARV	Enhanced differentiation, reduced proliferation in 3-D aggregates. Upregulation of growth factors and basement membrane proteins in co-cultures mimic physiological growth conditions of prostate epithelial cells with stromal cells, altered responses to sex hormones and growth factors
PC12 cells	STLV, HARV	Formation of large aggregates exhibiting neuroendocrine differentiation, altered cellular signaling mechanisms
Normal human kidney cells	STLV	Re-expression of tissue-specific morphology (microvilli) and differentiation markers
Bovine cartilage	STLV	Macroscopic large 3-D constructs with enhanced tissue-specific differentiation (ECM proteins e.g., chondroitin sulfate, and cytoskeletal proteins [e.g. vimentin])
Murine osteoblasts	Clinostat	Decrease of differentiated phenotype (reduction of alkaline phosphatase and osteocalcin)
Lymphoid tissue HIV	HARV	Repopulation of human tonsil fragments with exogenously added T and B lymphocytes; capable of infection by HIV
Liver	HARV	Expansion of tissue from microscopic fragments, angiogenesis

Source: Unsworth, B.R., Lelkes, P.I. (1998) Growing tissues in microgravity. *Nat. Med.* **4**(8): 904–907. (Adapted with copyright permission of the authors and Nature Publishing Company, New York, NY [http://www.nature.com/]).

environment of the MACS were more similar in composition and mechanical strength to natural cartilage. This experiment demonstrated the microgravity effect on cellular functions and processes in the synthesis, secretion, and assembly of extracellular matrix proteins. It is known that microgravity exerts deleterious effects on existing cartilage and bone. This investigation demonstrated the effect that microgravity may have on remodeling of tissues such as cartilage. Earth-grown constructs in the analog culture were superior to the flight effects of 1*g*. Subsequent experiments should delineate the effects of *g* forces on the process of tissue morphogenesis.

Space cell culture may not offer advantages in tissue morphogenesis for all cell types. Experiments with osteoblasts (i.e. bone-forming cells) were performed on STS-56 to assess the gene expression changes that occurred in microgravity. The results were that prostaglandin synthesis did occur during space flight and there were significant changes in cellular morphology and the acting cytoskeleton. Additionally, there was reduced growth of osteoblasts after four days in microgravity. Following this, a series of experiments on the European Space Agency Biorack (STS-76, STS-81, and STS-84) were performed on osteoblasts using ground and in-flight 1*g* controls. The expression of immediate early genes such as c-*fos*, and *cox-2* was altered significantly in microgravity. However, osteoblast fibronectin mRNA, protein synthesis, and matrix remained unchanged after exposure to microgravity. This suggests the possibility of altered signaling in the osteoblast, since the expression of these genes is critical for gene transcription and cell activation. The aggregation and secretion of bioactive products in MIP101 colon carcinoma cells were enhanced in microgravity. The extent of aggregation, proliferation, and differentiation of PC 12 pheochromocytoma (neoplastic adrenomedullary) cells in space was also enhanced. In addition, the rate of glucose consumption was five times higher than in ground cultures.

Primary cultures of human cortical renal epithelial cells were flown on the space shuttle to test the hypothesis that the adaptation of cells to conditions in space is reflected in the changes in gene expression. Cells in microgravity and ground-based controls were grown for six days and fixed. RNA was extracted, and automated gene array analysis of the expression of 10 000 genes was performed. A select group of 1632 genes was regulated in microgravity. The magnitude of the response signifies that the conditions in spaceflight offer a multitude of perturbations to the cells and some of the motifs are coincident with known response suites. These were consistent with shear stress response genes and heat shock proteins. In other instances, there are gene expression changes that may be reflective of novel suites that may be related to microgravity, culture conditions, and changes in other forces such as convection. Finally, some are most likely related to the promotion of tissue morphogenesis. These included the Wilm's tumor zinc finger protein and the vitamin D receptor expressed in space cultures of differentiated renal cortical cells.

6 Conclusions

Microgravity thus affects cells in different ways. In bone and muscle systems, microgravity exposure results in inhibition

of cell function, while in others there is an encouragement of three-dimensional cell aggregate formation and secretion of bioactive products. Cellular mechanotransduction properties change, as does transmembrane signaling. These changes are more often than not selective. Across phyla from plant cells where subcellular responses, cell division, and growth and development are altered to bacterial virulence, which is enhanced in the case of *Salmonella typhimurium*, microgravity has multiple effects. In mammalian cells, cytoskeletal distortion is probably one of the major methods by which a cell initiates response to microgravity. In plant cells and in microbes with cell walls, the mechanism may be substantially different. The whole organism, cell or tissue could possibly act as a gravity sensor as opposed to having individual gravity sensing subcellular organelles. The concept of *tensegrity* offers a focus on mechanical energy at specific points and allows testing of hypotheses of how changes in gravity translate into a cellular response. Gene expression changes are also profound indicating the magnitude of gravity in the adaptational responses in cells in space. Genomic and proteomic analysis of different cells on the ISS will greatly enhance the existing knowledge. Utilization of simple, defined, and well-studied model systems such as *Escherichia coli*, *Caenorhabditis elegans*, *Drosophila melanogaster*, pisceans, and amphibians will also provide insight into microgravity effects on many fundamental processes as they transcend the cell to tissues, organs, systems, and organisms. Microgravity as an experimental tool has gained impetus, and with the establishment of the ISS and reliable ground-based analogs, microgravity will provide new inroads into cell biology.

Acknowledgments

The authors thank Ms. Mildred D. Young for her invaluable help in the preparation of this manuscript. We also thank Dr. Steven R. Gonda and Mr. Charles M. Lundquist for reviewing the manuscript.

See also Bacterial Growth and Division; Cellular Interactions.

Bibliography

Books and Reviews

Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection*, John Murray, London.

Einstein, A. (1961) *Relativity: The Special and the General Theory*, (Lawson, R.W. trans), Three Rivers Press, New York.

Montgomery, P.O., Cook, J.E., Reynolds, R.C., Paul, J.S., Hayflick, L., Stock, D., Schulz, W.W., Kimzey, S., Thirolf, R.G., Rogers, T., Campbell, D., Murrell, J. (1977) The Response of Single Human Cells to Zero Gravity, in: Johnson, R.S., Dietlein, L. (Eds.) *Biomedical Results of Skylab*, NASA SP 377.

Nicogossian, A.E., Pool, S.L., Uri, J.J. (1993) Historical Perspectives, in: Nicogossian, A.E., Huntoon, C.L., Pool, S.L. (Eds.) *Space Physiology and Medicine*, Lea & Febiger, Philadelphia, pp. 3–49.

Pollard, E.C. (1971) Physical Determinants of Receptor Mechanisms, in: Gordon, A., Cohen, M. (Eds.) *Gravity and the Organism*, University of Chicago Press, Chicago, pp. 25–34.

Primary Literature

Armstrong, J.W., Nelson, K.A., Simske, S.J., Luttgers, M.W., Iandolo, J.J., Chapes, S.K. (1993) Skeletal unloading causes organ-specific changes in immune cell responses, *J. Appl. Physiol.* **75**, 2734–2739.

Backup, P., Westerlind, K., Harris, S., Spelsberg, T., Kline, B., Turner, R. (1994) Space

flight results in reduced mRNA levels for tissue specific protein in the musculoskeletal system, *Am. J. Physiol.* **266**, E567–E573.

Bechler, B., Hunzinger, E., Müller, O., Cogoli, A. (1993) Culture of hybridoma and Friend leukemia virus transformed cells in microgravity. Spacelab IML-1 mission, *Biol. Cell* **79**, 45–50.

Bonting, S.L. (1992) Utilization of biosensors and chemical sensors for space applications, *Biosens. Bioelectron.* **7**(8), 535–548.

Chapes, S.K., Mastro, A.M., Sonnenfeld, G., Berry, W.D. (1993) Antiorthostatic suspension as a model for the effect of spaceflight on the immune system, *J. Leukoc. Biol.* **54**, 227–235.

Claasen, D.E., Spooner, B.S. (1994) Impact of altered gravity on aspects of cell biology, *Int. Rev. Cytol.* **156**, 301–373.

Cogoli, M., Bechler, B., Cogoli, A., Arena, N., Barni, S., Pippia, P., Sechi, G., Valora, N., Monti, R. (1990) Lymphocytes on sounding rockets, *COSPAR Adv. Space Res.* **12**, 141–144.

Cogoli, A., Gmünder, F.K. (1991) Gravity effects on single cells: techniques, findings, and theory, *Adv. Space Biol. Med.* **1**, 183–248.

Cogoli, A., Gmünder, F.K. (1991) Gravity effects on single cells: techniques, findings, and theory, *Adv. Space Biol. Med.* **1**, 183–248.

Cogoli, A. (1993) The effect of hypogravity and hypergravity on cells of the immune system (1993), *J. Leukoc. Biol.* **54**, 259–268.

Cogoli, A. (1997) Signal transduction of T lymphocytes in microgravity, *Gravit. Space Biol. Bull.* **10**(2), 5–16.

Cogoli, A., Cogoli-Greuter, M. (1997) Activation and proliferation of lymphocytes and other mammalian cells in microgravity, *Adv. Space Biol. Med.* **6**, 33–79.

Cogoli, A., Cogoli-Greuter, M. (1997) Activation and proliferation of lymphocytes and other mammalian cells in microgravity, *Adv. Space Biol. Med.* **6**(2), 33–79.

Cogoli-Greuter, M., Pippia, P., Sciola, L., Cogoli, A. (1994) Lymphocytes on sounding rocket flights, *J. Gravit. Physiol.* **1**, 90–91.

Cooper, D., Pellis, N.R. (1998) Suppressed PHA activation of T lymphocytes in simulated microgravity is restored by direct activation of protein kinase C, *J. Leukoc. Biol.* **63**, 550–562.

Croxdale, J., Cook, M., Tibbitts, T., Brown, C., Wheeler, R. (1997) Structure of potato tubers formed during spaceflight, *J. Exp. Bot.* **48**, 2037–2043.

De Groot, R.P., Rijken, P.J., Boonstra, J., Verkley, A.J., de Laat, S.W., Kruijer, W. (1991) Epidermal growth factor induced expression of *c-fos* is influenced by altered gravity conditions, *Aviat. Space Environ. Med.* **62**, 37–40.

De Groot, R.P., Rijken, P.J., den Hertog, J., Boonstra, J., Verkleij, A.J., de Laat, S.W., Kruijer, W. (1991) Nuclear responses to protein kinase C signal transduction are sensitive to gravity changes, *Exp. Cell Res.* **197**, 87–90.

Freed, L.E., Langer, R., Martin, I., Pellis, N.R., Vunjak-Novakovic, G. (1997) Tissue engineering of cartilage in space, *Proc. Natl. Acad. Sci.* **94**, 13885–13890.

Freed, L.E., Hollander, A.P., Martin, I., Barry, J.R., Langer, R., Vunjak-Novakovic, G. (1998) Chondrogenesis in a cell-polymer bioreactor system, *Exp. Cell Res.* **240**, 58–65.

Grigoriev, A.I., Egorov, A.D. (1991) The effects of prolonged spaceflights on the human body, *Adv. Space Biol. Med.* **1**, 1–35.

Grigoriev, A.I., Egorov, A.D. (1992) General mechanisms of the effect of weightlessness on the human body, *Adv. Space Biol. Med.* **2**, 1–42.

Grigoriev, A.I., Egorov, A.D. (1992) Physiological aspects of adaptation of main human body systems during and after spaceflights, *Adv. Space Biol. Med.* **2**, 43–82.

Hammond, T.G., Bennes, E., O'Reilly, K.C., Wolf, D.A., Linnehan, R.M., Taher, A., Kaysen, J.H., Allen, P.L., Goodwin, T.J. (2000) Mechanical culture conditions effect gene expression: gravity induced changes on the space shuttle, *Physiol. Genomics* **3**, 163–173.

Hashemi, B.B., Penkala, J.E., Vens, C., Huls, H., Cubbage, M., Sams, C.F. (1999) T cell activation responses are differentially regulated during clinorotation and in spaceflight, *FASEB J.* **13**, 2071–2082.

Hatton, J.P., Gaubert, F., Cazenave, J.P., Schmitt, D. (2002) Microgravity modifies protein Kinase C isoform translocation in the human monocytic cell line U937 and human peripheral blood T cells, *J. Cell. Biochem.* **87**, 39–50.

Henry, R.L., Green, P.D., Wong, P.P., Guikema, J.A. (1990) Binding of isolated plant lectin by rhizobia during episodes of reduced gravity obtained by parabolic flight, *Plant Physiol.* **92**, 262–264.

Hughes-Fulford, M., Lewis, M.L. (1996) Effects of microgravity on osteoblast growth activation, *Exp. Cell Res.* **224**, 103–109.

Hughes-Fulford, M., Gilbertson, V. (1999) Osteoblast fibronectin mRNA, protein synthesis, and matrix are unchanged after exposure to microgravity, *FASEB J.* **13**(4), S121–S127.

Ingber, D.E. (1997) Tensegrity: the architectural basis of cellular mechanotransduction, *Annu. Rev. Physiol.* **59**, 575–599.

Ingber, D. (1999) How cells (might) sense microgravity, *FASEB J.* **13**, S3–15.

Jessup, J.M., Goodwin, T.J., Spaulding, G. (1993) Prospects for use of microgravity-based bioreactors to study three-dimensional host-tumor interactions in human neoplasia, *J. Cell. Biochem.* **51**, 290–300.

Jessup, J.M., Goodwin, T.J., Garcia, R., Pellis, N.R. (1996) STS-70: first flight of EDU-1, *In Vitro Dev. Biol. Anim.* **32**, 13A.

Khaoustov, V.I., Darlington, G.J., Soriano, H.E., Krishnan, B., Risin, D., Pellis, N.R., Yoffe, B. (1999) Induction of three-dimensional assembly of human liver cells by simulated microgravity, *In Vitro Dev. Biol. Anim.* **35**, 501–509.

Kulesh, D.A., Anderson, L.H., Wilson, B., Otis, E.J., Elgin, D.M., Barker, M.J., Mehm, W.J., Kearney, J.P. (1994) Space shuttle flight (STS-45) of L8 myoblast cells result in the isolation of a nonfusing cell line variant, *J. Cell. Biochem.* **55**, 530–544.

Kumei, Y., Nakajima, T., Sato, A., Kamata, N., Enomoto, S. (1989) Reduction of g-1 phase duration and enhancement of c-myc gene expression in HeLa cells at hypergravity, *J. Cell Sci.* **93**, 221–226.

Kumei, Y., Whitson, P.A., Sato, A., Cintron, N.M. (1991) Hypergravity signal transduction in HeLa cells with concomitant phosphorylation of proteins immunoprecipitated with anti-microtubule-associated protein antibodies, *Exp. Cell Res.* **192**, 492–496.

Lelkes, P.I., Galvan, D.L., Hayman, G.T., Goodwin, T.J., Chatman, D.Y., Cherian, S., Garcia, R.M.G., Unsworth, B. (1998) Simulated microgravity conditions enhance differentiation of cultured PC12 cells towards the neuroendocrine phenotype, *In Vitro Dev. Biol. Anim.* **34**, 316–324.

Limouse, M., Manié, S., Konstantinova, I.V., Ferrua, B., Schaffar, L. (1991) Inhibition of phorbol ester-induced cell activation in microgravity, *Exp. Cell Res.* **197**, 82–86.

Lorenzi, G., Gmünder, F.K., Cogoli, A. (1993) Cultivation of hamster kidney cells in dynamic cell culture system in space (Spacelab IML-1 Mission), *Micrograv. Sci. Technol.* **6**, 34–38.

Meehan, R., Whitson, P., Sams, C. (1993) The role of psychoneuroendocrine factors on spaceflight-induced immunological alterations, *J. Leukoc. Biol.* **54**, 236–244.

Merkys, A., Darginaviciene, J. (1997) Plant gravitropic response, *Adv. Space Biol. Med.* **6**, 213–230.

Mesland, D.A. (1987) Biorack experiments in spacelab D-1 and IML-1: further developments in gravitational biology, Proceedings of the 3rd European Symposium on Life Sciences Research in Space, ESA SP-271, pp. 307–312.

Mesland, D.A.M. (1992) Mechanisms of gravity effects on cells: are there gravity-sensitive windows, *Adv. Space Biol. Med.* **2**, 211–228.

Miller, E.S., Sonnenfeld, G. (1993) Influence of suspension on the expression of protective immunological memory to murine *Listeria monocytogenes* infection, *J. Leukoc. Biol.* **54**, 578–583.

Miller, E.S., Sonnenfeld, G. (1994) Influence of antiorthostatic suspension on resistance to murine *Listeria monocytogenes* infection, *J. Leukoc. Biol.* **55**, 371–378.

Montgomery, P.O.B., Cook, J.E., Reynolds, R.C., Paul, J.S., Hayflick, L., Stock, D., Schulz, W.W., Kimsey, S., Thirolf, R.G., Rogers, T., Campell, D. (1978) The response of single human cells to zero gravity, *In Vitro* **14**, 165–173.

Nace, G.W. (1983) Gravity and homoeostasis of the cell, *COSPAR Adv. Space Res.* **3**, 159–168.

Nash, P.V., Bour, B.A., Mastro, A.M. (1991) Effect of hindlimb suspension simulation of microgravity on *in vitro* immunological responses, *Exp. Cell Res.* **195**, 353–360.

Nash, P.V., Mastro, A.M. (1992) Variable lymphocyte responses in rats after space flight, *Exp. Cell Res.* **202**, 125–131.

Nash, P.V., Mastro, A.M. (1993) Activation of primary lymphocytes requires prolonged lectin stimulation, *J. Leukoc. Biol.* **53**, 73–78.

Pellis, N.R., Goodwin, T.J., Risin, D., McIntyre, B.W., Pizzini, R.P., Cooper, D., Baker, T.L., Spaulding, G.F. (1997) Changes in gravity

inhibit lymphocyte locomotion through type I collagen, *In Vitro Cell. Dev. Biol. Anim.* **33**, 398–405.

Pippia, P., Sciola, L., Cogoli-Greuter, M., Meloni, M.A., Spano, A., Cogoli, A. (1996) Activation signals of T lymphocytes in microgravity, *J. Biotechnol.* **47**, 215–222.

Pollard, E.C. (1965) Theoretical studies on living systems in the absence of mechanical stress, *J. Theor. Biol.* **8**, 112–123.

Rijken, P.J., de Groot, R.P., Kruijer, W., de Laat, S.W., Verkleij, A.J., Boonstra, J. (1992) Identification of specific gravity sensitive signal transduction pathways in human A431 carcinoma cells, *COSPAR Adv. Space Res.* **12**, 145–152.

Rijken, P.J., de Groot, R., van Belzen, N., de Laat, S.W., Boonstra, J., Verkleij, A. (1993) Inhibition of EGF-induced signal transduction by microgravity is independent of EGF receptor redistribution in plasma membrane of human A431 cells, *Exp. Cell Res.* **204**, 373–377.

Rijken, P.J., Boonstra, J., Verkleij, A.J., de Laat, S.W. (1994) Effects of gravity on the cellular response to epidermal growth factor, *Adv. Space Biol. Med.* **4**, 159–188.

Schatz, A., Linke-Hommes, A. (1989) Gravity and the membrane-solution interface: theoretical investigations, *COSPAR Adv. Space Res.* **9**, 61–64.

Schatz, A., Reitstetter, R., Briegleb, W., Linke-Hommes, A. (1992) Gravity effects on biological systems, *COSPAR Adv. Space Res.* **12**, 51–53.

Schmitt, D.A., Malouvier, A., Holy, X., Marie, P.J., Caissard, J.C., Zérath, E. (1991) Receptor-ligand binding on osteoblasts in microgravity obtained by parabolic flight, *Physiologist* **34**, S64–S65.

Schmitt, D.A., Ohlmann, P., Gachet, C., Cazenave, J.P. (1993) Platelet shape change and protein phosphorylation induced by ADP and thrombin are not sensitive to short periods of microgravity, *J. Cell Sci.* **104**, 805–810.

Schmitt, D.A., Hatton, J.P., Emond, C., Chaput, D., Paris, H., Levade, T., Cazenave, J.P., Schaffar, L. (1996) The distribution of protein kinase C in human leukocytes is altered in microgravity, *FASEB J.* **10**, 1627–1634.

Schwarz, R.P., Goodwin, T.J., Wolf, D.A. (1992) Cell culture for three-dimensional modeling in rotating-wall vessels: an application in microgravity, *J. Tissue Cult. Methods* **14**, 51–58.

Seitzer, U., Bodo, M., Muller, P.K., Acil, Y., Batge, B. (1995) Microgravity and hypergravity effects on collagen biosynthesis on human dermal fibroblasts, *Cell. Tissue Res.* **282**, 513–517.

Silver, I.L. (1976) The dynamics of a discrete geotropic sensor subject to rotation-induced gravity compensation, *J. Theor. Biol.* **61**, 353–362.

Spooner, B.S., Guikema, J.A., Barnes, G. (1990) Binding of alpha fetoprotein by immobilized monoclonal antibodies during episodes of zero-gravity obtained by parabolic flight, *Aviat. Space Environ. Med.* **61**, 725–728.

Sundaresan, A., Risin, D., Pellis, N.R. (2002) Loss of signal transduction and inhibition of lymphocyte locomotion in a ground-based model of microgravity, *In Vitro Cell. Dev. Biol. Anim.* **38**, 118–122.

Sundaresan, A., Yamauchi, K., Kulkarni, A.D., Pellis, N.R. (2002) Micro gravity and modeled micro gravity effects on lymphocyte signal transduction: comparisons between human and mouse lymphocyte signaling, *Jpn. J. Aerospace Res.* 2256–2261.

Sundaresan, A., Risin, D., Pellis, N.R. (2003) Modeled Microgravity-Induced Protein Kinase C Isoform Expression in Human Lymphocytes. *Manuscript in submission.*

Taylor, G.R., Janney, R.P. (1992) *In vivo* testing confirms a blunting of the human cell-mediated immune mechanism during space flight, *J. Leukoc. Biol.* **51**, 129–132.

Taylor, G.R. (1993) Immune changes during short-term missions, *J. Leukoc. Biol.* **54**, 202–208.

Truneh, A., Albert, F., Golstein, P., Schmitt-Verhulst, A.M. (1985) Early steps of lymphocyte activation bypassed by synergy between calcium ionophores and phorbol ester, *Nature* **313**, 318–320.

Tsao, Y.D., Goodwin, T.J., Wolf, D.A., Spaulding, G.F. (1992) Responses of gravity level variations on the NASA/JSC bioreactor system, *Physiologist* **35**, S49–S50.

Unsworth, B.R., Lelkes, P. (1998) Growing tissues in microgravity, *Nat. Med.* **4**, 901–907.

Vunjak-Novakovic, G., Searby, N., De Luis, J., Freed, L.E. (2002) Microgravity studies of cells and tissues, *Ann. N. Y. Acad. Sci.* **974**, 504–517.

Wilson, J.W., Ramamurthy, R., Porwollik, S., McClelland, M., Hammond, T., Allen, P.,

Ott, M.C., Pierson, D.L., Nickerson, C.A. (2002) Microassay analysis identifies Salmonella genes belonging to the low-shear modeled microgravity regulon, *Proc. Natl. Acad. Sci. USA* **99**(21), 13807–13812.

Woods, K.M., Chapes, S.K. (1994) Abrogation of TNF-mediated cytotoxicity by space flight involved protein kinase C, *Exp. Cell Res.* **211**, 171–174.

Yamauchi, K., Sundaresan, A., Hales, N.W., Yamamoto, Y., Pellis, N.R., Kulkarni, A.D. (2002) Nutritional countermeasure to obviate immune dysfunction in microgravity, *Jpn. J. Aerospace Res.* 2262–2267.

Cell Junctions, Structure, Function, and Regulation

Susan E. LaFlamme[1] *and Peter A. Vincent*[2]
[1] *Center for Cell Biology and Cancer Research, Albany Medical College, Albany, NY, USA*
[2] *Center for Cardiovascular Sciences, Albany Medical College, Albany, NY, USA*

1 Cell–matrix Adhesion 326
1.1 Overview 326
1.2 Extracellular Matrix 326
1.2.1 Collagens 327
1.2.2 Laminins 329
1.2.3 Other Extracellular Matrix Proteins 330
1.3 Integrins 331
1.3.1 The Integrin Family 331
1.3.2 Integrin Ligands 332
1.3.3 Integrin Signaling 333
1.3.4 Integrin-associated Proteins 336
1.4 Focal Adhesions 341
1.4.1 Components of Focal Adhesion 341
1.4.2 Regulation of Focal Adhesion Assembly 342
1.5 Hemidesmosomes 343
1.5.1 Assembly of Hemidesmosomes 344
1.5.2 Signaling by $\alpha_6\beta_4$ 345

2 Cell–Cell Junctions 345
2.1 Overview 345
2.2 Cadherin Superfamily 347
2.3 Type I/II Cadherins and Adherens Junctions 349
2.3.1 Extracellular Domain: Cis- and Trans-cadherin Dimerization 350
2.3.2 Cytoplasmic Domain: Lateral Clustering and Cytoskeletal Attachment 352
2.3.3 Regulation of Cadherin–Catenin Complex 354
2.3.4 Other Components of Adherens Junctions: Nectin and Afadin 356

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

2.4 Desmosomal Cadherins and Desmosomes 356
2.4.1 Extracellular Domain of Desmoglein and Desmocollin 357
2.4.2 Cytoplasmic Tail, Plaque Proteins, and Attachment to the Cytoskeleton 358
2.4.3 Functions of Desmosomes 361
2.5 Tight Junctions 361
2.5.1 Claudins and Other TJ Transmembrane Proteins 362
2.5.2 Plaque Proteins of Tight Junctions 365
2.5.3 Regulation of Tight Junction Assembly/Disassembly 366

3 Summary and Future Directions 368

Bibliography 370
Books and Reviews 370
Primary Literature 371

Keywords

Adherens Junction
Cadherin-mediated cell–cell junction that attaches to the actin cytoskeleton thereby providing strength to the junction and allowing the cells to act as a single structure.

Cadherins
A family of single pass transmembrane proteins possessing five homologous domains in the extracellular region. Cadherins on adjacent cells bind in a homophilic, calcium-dependent manner to form a cell–cell junction.

Desmosome
A ''spot weld''-like cell–cell junction characterized by a dense plaque on the cytoplasmic face of the plasma membrane into which intermediate filaments attach. Mediated by members of the cadherin family, this specialized junction has tremendous cell–cell adhesion strength and is found in tissues that require strong cell–cell attachment such as the heart and skin.

Extracellular Matrix
An insoluble meshwork of cell-secreted macromolecules (including polysaccharides, such as glycosaminoglycans and large proteins, such as collagen and fibronectin) that holds cells together in tissues, and tissues together in organs, and also provides an adhesive substrate for cells that regulates cell behavior.

Focal Adhesion
A specialized cell-ECM junction where there is tight association between the cell and an ECM protein, such as fibronectin. Adhesion is mediated by the direct binding of

integrins to the ECM protein. Adhesion is strengthened by the transmembrane association of integrins with intracellular proteins that link the adhesion site to actin stress fibers within the cell's cytoplasm.

Hemidesmosome
A specialized cell-ECM junction that mediates very strong adhesion between the basal layer of stratified epithelia and the underlying basement membrane. The $\alpha_6\beta_4$ integrin functions at hemidesmosomes as a transmembrane link between the basement membrane and cytoplasmic proteins that link hemidesmosomes to the intermediate filament cytoskeleton.

Integrins
Heterodimeric transmembrane proteins containing one α- and one β-subunit. Integrins are generally known for their ability to bind ECM proteins; however, some integrins bind soluble ligands and/or counter receptors on other cells. Although integrins are known as adhesion receptors, they also function as signaling receptors that regulate cell behavior.

Tight Junction
A cell–cell junction that acts as a barrier to the passage of solutes between epithelial and endothelial cells and serves as a fence to prevent the movement of proteins within the plasma membrane from passing between the apical and basal surfaces of the cell layer.

A critical component in the development and homeostasis of higher eukaryotic organisms is the adhesion of cells to each other and to large insoluble proteins that form the extracellular matrix (ECM). These adhesions are referred to as cell–cell and cell–matrix junctions respectively. Specific families of cell surface transmembrane receptors connect these junctions with the cell's cytoskeleton. Three different cell–cell junctions promote the stable adhesion between cells: tight junctions formed by the claudin family, and adherens junctions and desmosomes both mediated by members of the cadherin family. Members of the integrin family function at cell–matrix junctions. A subset of integrins also mediate cell–cell adhesion events that are important for the immune response. In addition to providing structural linkages important for tissue architecture, cell–matrix and cell–cell adhesions also activate signaling pathways that act coordinately with signals from receptors for growth factors, chemokines, cytokines and morphogens to regulate the cell adhesion, migration, proliferation, survival, and differentiation. In addition to adhesion receptors themselves, large multiprotein complexes are assembled at the cytoplasmic face of these junctions by protein interactions initiated at the intracellular domains of each type of adhesion receptor. Biochemical and cell culture experiments have provided information on how these protein interactions and complexes connect adhesion junctions to the cytoskeleton and the cell's signaling apparatus. Many components of these adhesion complexes have been identified.

Important insights have also been gleaned from gene knockout studies, which have identified critical roles for many components of these adhesion complexes during development. Mutations in protein components of these adhesion complexes have also been identified in the etiology of a variety of human diseases.

1 Cell–matrix Adhesion

1.1 Overview

The adhesion of cells to extracellular matrix (ECM) proteins is essential for embryonic development, tissue morphogenesis, and homeostasis of all vertebrate organisms. Integrins comprise the major class of cell surface proteins that bind to ECM proteins. Integrins contribute to tissue structure by forming transmembrane links between ECM proteins and the cell's cytoskeleton at sites referred to as cell–matrix junctions. The best-characterized cell–matrix junctions are focal adhesions and hemidesmosomes. At focal adhesions, integrins connect the ECM with the actin microfilament cytoskeleton, and at hemidesmosomes integrins connect the ECM with the intermediate filament cytoskeleton. Integrins are also signaling receptors, providing cells with information to evaluate the composition and physical state of their environment. Integrin signaling contributes to the regulation of cell migration, cell proliferation, survival and differentiation. Important insights into the physiological roles of extracellular matrix proteins and their integrin receptors has been gained from cell culture models, gene knockout approaches in mice, and the analysis of human hereditary disorders involving genes encoding cell adhesion proteins.

1.2 Extracellular Matrix

The extracellular matrix (ECM) is comprised of cell-secreted macromolecules that form an insoluble meshwork outside cells. The ECM holds cells together in tissues, and tissues together in organs. Large glycoproteins such as collagens, laminins, and fibronectin are important components of the ECM. They contribute to the physical characteristics of the ECM and function as adhesive ligands for cells. Also important to the structure of the ECM are the long, unbranched, charged polysaccharides referred to as glycosaminoglycans (GAGs) and proteins containing GAG chains called proteoglycans. Because of their extended and charged nature, GAGs easily form hydrated gels that are important for tissues to resist compressive forces.

ECM is the major component in connective tissue, such as skin, bone, and cartilage. The ECM is secreted and organized by cells that reside within the connective tissue: fibroblasts in the connective tissue in skin, chondrocytes in cartilage, and osteoblasts in bone. Other cells such as epithelial cells, endothelial cells, and nerve cells secrete ECM proteins that become assembled into a basal lamina or basement membrane (Fig. 1). Unlike the ECM of connective tissue, which surrounds the cells that secrete it, the basement membrane is found only on the basal cell surface of endothelial and epithelial cells. Basement membranes

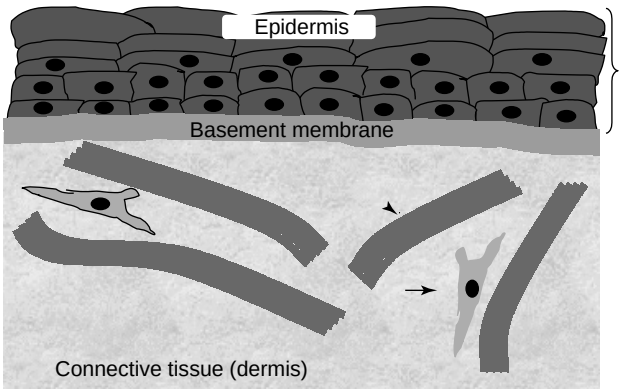

(a)

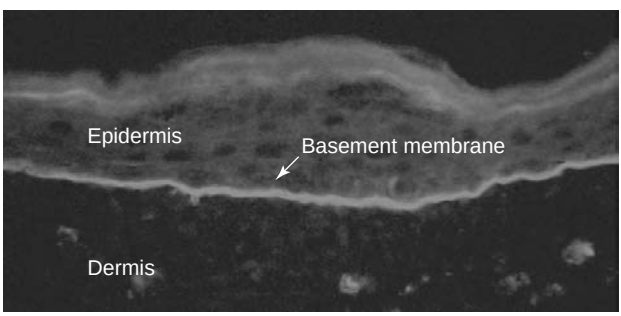

(b)

Fig. 1 Extracellular matrices of the skin. (a) A schematic representation of a cross-section of skin. Keratinocytes of the epidermis are shown in red. The basement membrane is shown in green. The connective tissue of the underlying dermis is shown in tan. The black arrows and arrowheads indicate dermal fibroblasts and collagen fibrils respectively. (b) A cross-section of mouse skin stained with fluorescent antibodies that recognize laminin-5. The epidermis consisting of keratinocytes, the basement membrane, and the underlying connective tissue of the dermis are labeled. Bright staining for laminin-5 is seen in the basement membrane. Some background autofluorescence is seen in the upper layers of epidermis. The immunofluorescence micrograph was kindly provided by Dr. Michael DiPersio (see color plate p. xxviii).

separate these cells from connective tissue and other cell types (Fig. 1) and also function as permeability barriers to the passage of macromolecules. This is particularly important in the kidney glomerulus, where the basement membrane regulates the passage of molecules from the blood into the urine. Additionally, the adhesion of cells to basement membranes regulates cell behavior including cell survival, proliferation and differentiation.

1.2.1 Collagens

Collagens comprise a large family of related proteins (Table 1) that are important components of all ECMs. All collagens

Tab. 1 Collagens. (Adapted from Prockop, D.J., Kivirikko, K.I. (1995) *Annu. Rev. Biochem.* **64**, 403–434.)

Collagens[a]	*Subunit composition*[b]	*Tissue distribution*
Fibril-forming (fibrillar)		
Type I	$[\alpha_1(I)_2\alpha_2(I)$	Most connective tissues including bone, skin, tendons, ligaments, cornea, internal organs
Type II	$[\alpha_1(II)]_3$	Cartilage, invertebral disc, notochord, vitreous humor of the eye
Type III	$[\alpha_1(III)]_3$	Skin, blood vessels, internal organs
Type V	$[\alpha_1(V)]_2\alpha_2(V)$ and $\alpha_1(V)\alpha_2(V)\alpha_3(V)$	Similar to Type I
Type XI	$\alpha_1(IX)\alpha_2(IX)\alpha_3(IX)$	Similar to Type II
Fibril-associated		
Type IX	$\alpha_1(IX)\alpha_2(IX)\alpha_3(IX)$	Cartilage
Type XII	$[\alpha_1(IV)]2\alpha_2(IV)$	Tendons, ligaments, some other tissues
Network forming		
Type IV	$[\alpha_1(IV)]2\alpha_2(IV)$	Basement membranes
Anchoring filaments		
Type VII	$[\alpha_1(VII)]3$	Beneath stratified squamous epithelia
Transmembrane		
Type XVII (BP180)	$[\alpha_1(XCII)]3$	Hemidesmosomes present in skin

[a]Only a subset of the known collagens are presented.
[b]Collagen subunits are also referred to as collagen chains.

contain a triple helical domain containing three collagen subunits referred to as α-chains or α-subunits. The triple helical structure is dictated by a stretch of glycine-X-Y amino acid repeats in each α-chain, where X and Y are frequently proline and 4-hydroxy-proline respectively. Depending upon the type of collagen, the triple helical domain can form from three identical or different α-chains. Collagens also differ in structure and function. Fibrillar collagens, which are major components of connective tissue, have long stretches of triple helical domains that promote the formation of fibrils and larger fibers. Fibril-associated collagens mediate the interactions of collagen fibrils with each other and with other matrix components. Network-forming collagens form extended lattices that are believed to have important scaffolding functions in basement membranes. The triple helical domains of fibril-associated and network-forming collagens are interrupted with nonhelical domains, making these collagen proteins more flexible. Other collagens form anchoring fibrils that connect basement membranes of stratified epithelia to the underlying connective tissue.

Mutations have been identified in collagen genes that have been linked to human genetic disorders. Mutations in genes that encode type I fibrillar collagen are responsible for the genetic disorder

osteogenesis imperfecta. In its most severe form, this disorder is characterized by brittle bones, abnormal teeth, thin skin, and weak tendons, which reflects the wide tissue distribution of type I collagen. Mutations in genes that encode type II collagen, which is found mostly in cartilage, are responsible for disorders referred to as chondrodysplasias. Affected individuals can have bone and joint deformities. Mutations in genes that encode type III collagen affect the integrity of blood vessels and skin, reflecting the tissue distribution of type III collagen. Null mutations in the homologous genes in mice result in similar phenotypes.

1.2.2 **Laminins**

Laminins comprise a family of high molecular weight heterotrimers that are essential components of basement membranes (Table 2). Each laminin contains one α-subunit, one β-subunit, and one γ-subunit. To date there are five known α-subunits, three known β-subunits, and three known γ-subunits that form 12 different laminin isoforms, each containing a distinct combination of α-, β- and γ-subunits that assemble in a cross-like structure with one coil-coiled domain containing portions of all three subunits. Most information about the tissue distribution of individual laminin subunits is known (Table 3). However, the expression and function of laminin isoforms in specific tissues has also been established. For example, laminin-1 ($\alpha_1\beta_1\gamma_1$) expression is critical for early vertebrate development. Laminin-2 ($\alpha_2\beta_1\gamma_1$) and laminin-4 ($\alpha_2\beta_2\gamma_1$) play important roles in the development and maintenance of muscles and nerves. Mutations in the gene for the laminin α_2 subunit, which is present in both laminin-2 and laminin-4, cause some forms of congenital muscular dystrophy in humans, in which affected individuals show both muscular and neurological defects. Laminin-5 ($\alpha_3\beta_3\gamma_2$) is present in basement membranes of stratified epithelia. Epithelial cell adhesion to laminin-5 is required for epithelial tissues to resist physical stress. Laminin-5 is also discussed in Sect. 1.5, because of its central role in cell–matrix junctions known as hemidesmosomes. Laminins also play a critical role in the kidney where laminin expression is differentially regulated for kidney development and function. Laminin-1 ($\alpha_1\beta_1\gamma_1$), laminin-8 ($\alpha_4\beta_1\gamma_1$) and laminin-10 ($\alpha_5\beta_1\gamma_1$) are expressed early in the formation of the glomerular basement membrane. These laminins are later replaced by laminin-9 ($\alpha_4\beta_2\gamma_1$) and finally by laminin-11 ($\alpha_5\beta_2\gamma_1$). Mice with null mutations in the gene encoding the laminin α_5 subunit lack one or both kidneys. When one kidney is present in α_5 laminin-null mice, this kidney is smaller than normal and shows defects in glomerulogenesis and in the glomerular basement membrane. As indicated above, the basement membrane of kidney glomeruli performs an important filtration function, and β_2

Tab. 2 Laminins.

Laminins	*Chain composition*
Laminin 1	$\alpha_1\beta_1\gamma_1$
Laminin 2	$\alpha_2\beta_1\gamma_1$
Laminin 3	$\alpha_1\beta_2\gamma_1$
Laminin 4	$\alpha_2\beta_2\gamma_1$
Laminin 5	$\alpha_3\beta_3\gamma_2$
Laminin 6	$\alpha_3\beta_1\gamma_1$
Laminin 7	$\alpha_3\beta_2\gamma_1$
Laminin 8	$\alpha_4\beta_1\gamma_1$
Laminin 9	$\alpha_4\beta_2\gamma_1$
Laminin 10	$\alpha_5\beta_1\gamma_1$
Laminin 11	$\alpha_5\beta_2\gamma_1$
Laminin 12	$\alpha_2\beta_1\gamma_3$

Tab. 3 Tissue distribution of laminin chains. (Adapted from Patarroyo et al. (2002) *Seminars in Cancer Biol.*, **12**, 197–207).

Laminin chain	*Tissue distribution*
α_1	Early embryo, neuroretina, kidney proximal tubes, salivary, and mammary glands
α_2	Skeletal and cardiac muscle, peripheral nerve, a few capillaries, trophoblast, brain, and other tissues
α_3	Skin and most other epithelia
α_4	Fetal skeletal muscle, cardiac and visceral smooth muscle, nerve, blood vessel endothelia, bone marrow, and other tissues
α_5	Diverse epithelia, kidney, blood vessels, bone marrow, developing muscle and nerve, synaptic BM
β_1	Most tissues
β_2	Neuromuscular junction, blood vessels, kidney glomerulus
β_3	Skin and most other epithelia
γ_1	Most tissues
γ_2	Skin and most other epithelia
γ_3	Kidney, lung, reproductive tract, nerve, and brain

subunit containing laminins play a critical role in this filtration function. Mice with a null mutation in the laminin β_2 gene maintain high levels of expression of the β_1 laminin chain; nonetheless, these mice die after birth because of massive proteinuria, because of the failure of the glomerular basement membrane to function as a filtration barrier. Thus, laminins play isoform-specific roles in tissue differentiation and function.

1.2.3 Other Extracellular Matrix Proteins

In addition to collagens and laminins, there are other ECM proteins that play important roles in regulating cell behavior, including fibronectin and fibrinogen/fibrin. Fibronectin is a large disulfide-linked dimer that is present in a soluble form in blood. Fibronectin is also secreted by fibroblasts and other cells into the ECM where it forms fibrils associated with the cell surface. Like laminin-1, fibronectin is essential for early vertebrate development. Characterization of fibronectin-null embryos indicated a requirement for fibronectin in the development of mesodermal structures, blood vessels, and the heart. Fibronectin also plays an important role in wound healing in the adult. During cutaneous wound healing, a provisional matrix is formed from plasma proteins released from injured vessels. These proteins include fibronectin and fibrinogen. Fibrinogen is proteolytically processed at the site of blood vessel injury to form a fibrin clot. This provisional matrix supports the formation of new blood vessels, the proliferation of fibroblasts and the reepithelialization of the wound. The provisional matrix is eventually replaced by the normal ECM of the skin.

In summary, extracellular matrices are essential to tissue and organ structure and development not only because of their scaffolding function but also because they provide important regulatory information that guides cell behavior and differentiation. As discussed in the next sections, much of this information is relayed by integrin receptors that

provide the physical link at cell–matrix junctions.

1.3 Integrins

1.3.1 The Integrin Family

Integrins comprise a large family of cell surface proteins. They are heterodimeric transmembrane receptors for ECM proteins (Fig. 2). Each integrin contains one α-subunit and one β-subunit. Both α- and β-subunits pass through the plasma membrane, and each subunit contains a relatively large extracellular segment with a globular ligand-binding domain. Most α- and β-subunits also contain relatively small intracellular domains (cytoplasmic domains or tails) consisting of from 13 to 70 amino acids. The exception is the β_4 subunit cytoplasmic domain, which contains 1000 amino acids.

Integrins are highly evolutionarily conserved. Integrin homologs have been identified in all metazoans from flies and nematodes to humans. In mammals, there are currently 8 known β-subunits and 18 α-subunits that heterodimerize in different combinations to generate 24 distinct integrin heterodimers (Table 4). Several integrin subunits can be modified by alternative splicing, further adding to the complexity of the integrin family. The specific combination of α- and β-subunits determines ligand-binding specificity. Some integrins are specific for a single component of the ECM, such as the $\alpha_5\beta_1$ integrin that binds only to fibronectin. Most integrin heterodimers can bind to more than one extracellular matrix component. The $\alpha_v\beta_3$ integrin is one of the more promiscuous integrins; it binds to vitronectin, fibronectin, and fibrinogen, among other ECM components.

Most cells express several different integrin heterodimers allowing them to interact with multiple components of the ECM, and a cell's integrin repertoire can be regulated both developmentally and in response to changes in the physiological environment. Although most integrin heterodimers are expressed by a variety of cell types, some integrins are expressed in a cell-type specific manner. The $\alpha_{IIb}\beta_3$ integrin is expressed only in platelets. The $\alpha_7\beta_1$ integrin is expressed exclusively in neuronal cells and integrins containing the β_2 subunit are expressed only by immune cells also referred to as leukocytes.

The identification of human hereditary diseases involving mutations in genes encoding integrins, and the analysis of null mutations in integrin genes in mice have underscored the importance of integrin receptors in physiological and developmental processes. To date there are four inherited disorders involving integrins. Leukocyte adhesion deficiency-1 (LAD1) is caused by mutations in the integrin β_2 subunit where affected individuals show

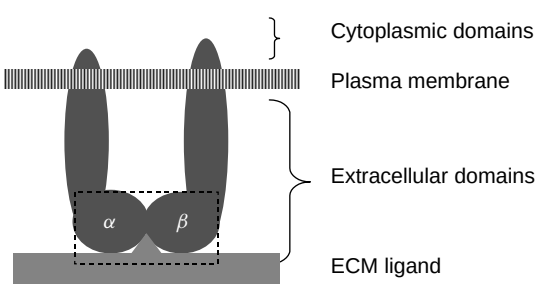

Fig. 2 Integrin heterodimer. Shown is a schematic representation of an integrin heterodimer. The typical small cytoplasmic domains and large extracellular domains of the α- and β-subunits are indicated by brackets. The N-terminal globular domains of the α- and β-subunits that function in ligand binding are indicated by a dashed box.

Tab. 4 Integrins and their ligands. (Adapted from Table 1 in van der Flier, A., Sonnenberg, A. (2001) Function and interactions of integrins, Cell Tissue Res. **305**, 285–298).

Integrin	*Ligand[a] (ECM, soluble, counter receptors)*
$\alpha_1\beta_1$	Co, Ln[2]
$\alpha_2\beta_1$	Co, Ln
$\alpha_3\beta_1$	Ln
$\alpha_4\beta_1$	Fn, VCAM
$\alpha_4\beta_7$	Fn, VCAM, MadCAM
$\alpha_5\beta_1$	Fn
$\alpha_6\beta_1$	Ln
$\alpha_6\beta_4$	Ln
$\alpha_7\beta_1$	Ln
$\alpha_8\beta_1$	Fn, Tn, VCAM
$\alpha_9\beta_1$	Tn, Co, Ln, VCAM
$\alpha_{10}\beta_1$	Co
$\alpha_{11}\beta_1$	Co
$\alpha_v\beta_1$	Fn, Vn
$\alpha_v\beta_3$	Vn, Fn, vWF, Op, Tn
$\alpha_v\beta_5$	Vn
$\alpha_v\beta_6$	Fn, Tn
$\alpha_v\beta_8$	Co, Ln, Fn
$\alpha_{IIb}\beta_3$	Vn, Fn, vWF, Fg, Fg
$\alpha_L\beta_2$	ICAM 1–5
$\alpha_M\beta_2$	Fg, iC3b, ICAM
$\alpha_x\beta_2$	Fg, iC3b
$\alpha_D\beta_2$	ICAM, VAM
$\alpha_E\beta_7$	E-cadherin

[a]ECM, soluble, and counter receptors ligands are shown in red, blue, and green, respectively.

severe deficiencies in their immune responses. Glanzmann's thrombasthenia, a bleeding disorder, is caused by mutations that prevent the proper expression of the $\alpha_{IIb}\beta_3$ platelet integrin and some mild forms of muscular dystrophy are associated with mutations in the α_7 subunit, which affect the function of the $\alpha_7\beta_1$ laminin binding integrin. Additionally, mutations that inhibit $\alpha_6\beta_4$ function result in a skin blistering disorder (see Sect. 1.5). These phenotypes are recapitulated in mice carrying the corresponding homozygous null mutations in genes encoding the β_2, α_{IIb}, β_3, α_7, α_6, and β_4 integrin subunits.

The generation of targeted deletions in integrin genes in mice has probably provided the most insight into the developmental role of individual integrin heterodimers. Homozygous null mutations in the gene encoding the integrin β_1 subunit gene result in early embryonic lethality with defects in implantation and gastrulation. Targeted deletion of the integrins α_5 and α_4 subunit genes also result in embryonic lethality. The α_5-null embryos have defects in vascular and mesoderm development, whereas α_4-null embryos show defects in heart development. Deletion of the integrin α_8 subunit gene results in defects in kidney development and deletion of the integrin α_9 subunit results in defects in the development of the lymphatics and thoracic duct. The α_9-null mice die several days after birth from respiratory failure due to fluid accumulation. Mice with null mutations in the integrin α_3 subunit gene show abnormalities in kidney and lung development and also have a skin blistering phenotype. Interestingly, null mutations in some integrin genes do not produce obvious developmental phenotypes. However, in some instances, phenotypes are observed in adult animals when challenged or stressed in some way. For example, mice with null mutations in the integrin β_6 subunit gene are viable and fertile but show increased skin and lung inflammation in response to environmental triggers.

1.3.2 Integrin Ligands

In addition to ECM ligands, some integrins have soluble ligands, and other integrins bind to cell surface receptors (counterreceptors) on neighboring cells

(Table 4). For example, the platelet integrin $\alpha_{IIb}\beta_3$ binds to soluble fibrinogen leading to platelet aggregation. β_2 integrins bind to counterreceptors on the surface of other cells. For example, the $\alpha_L\beta_2$ integrin expressed on T-lymphocytes binds to its counterreceptor ICAM-1 expressed on antigen-presenting cells, leading to the formation of the adhesion complex known as the *immune synapse*, which is required for full and sustained T-lymphocyte activation. The $\alpha_L\beta_2$ integrin is also expressed on neutrophils and monocytes/macrophages and can bind to ICAM-1 expressed on the endothelial cells that line blood vessels. Similarly, the $\alpha_4\beta_1$ and $\alpha_4\beta_7$ integrins expressed on leukocytes can bind to their counterreceptor(s), VCAM-1 and MadCAM-1 ($\alpha_4\beta_7$ only) on endothelial cells. The binding of leukocyte integrins to their counterreceptors on endothelial cells is important to enable leukocytes to exit the circulation in order to fight infection in tissues in the case of neutrophils and monocyte/macrophages, or to enter lymphoid tissues in the case of T-lymphocytes.

1.3.3 Integrin Signaling

The most intriguing aspect of integrin biology is the ability of integrins to function as bidirectional signaling receptors. Ligand binding to integrins induces transmembrane conformational changes that transduce signals to integrin intracellular domains, resulting in the linkage of integrins to cytoskeletal networks and signaling pathways that regulate cell behavior. The process by which ligand binding to integrins activates intracellular signals is referred to as integrin "outside-in" signaling.

Integrin "outside-in" signaling provides cells with important information about their ECM environment that allows them to make decisions regarding proliferation, survival, migration, and differentiation in response to growth factors, cytokines, chemokines, and morphogens. The formation of integrin-ECM adhesions regulates these processes by activating cascades of biochemical signals that regulate the assembly of cytoskeletal networks and signaling complexes (Fig. 3). The formation of signaling complexes is regulated by protein–protein interactions. For example, Src homology 2 (SH2) domains bind to phosphorylated tyrosine residues in proteins. Thus, the activation of tyrosine kinase signaling generates binding sites for SH2 domains, providing a mechanism to recruit proteins to signaling complexes. Other protein interactions are mediated by Src homology 3 (SH3) domains; these domains bind to proline-rich regions in other proteins. There are still other conserved domains that mediate protein–protein interactions; these will be described as they become relevant to the discussion.

An early signaling event that follows integrin engagement in cell adhesion is the activation of the cytoplasmic tyrosine kinase, focal adhesion kinase (FAK). Integrins trigger the autophosphorylation of FAK at tyrosine 397, which promotes the association of Src family of cytoplasmic tyrosine kinases and other signaling proteins such as phosphoinositide (PI) 3-kinase and phospholipase Cγ via their SH2 domains. Src kinases phosphorylate FAK's kinase domain and also tyrosine 925, increasing FAK's kinase activity and promoting the binding of SH2 domain containing signaling proteins at tyrosine 925, including Grb2, and Grb7. Activated FAK/Src kinase complexes trigger a cascade of protein tyrosine phosphorylations, including the phosphorylation of the adaptor proteins paxillin and p130Cas, which in turn promotes the formation of protein complexes

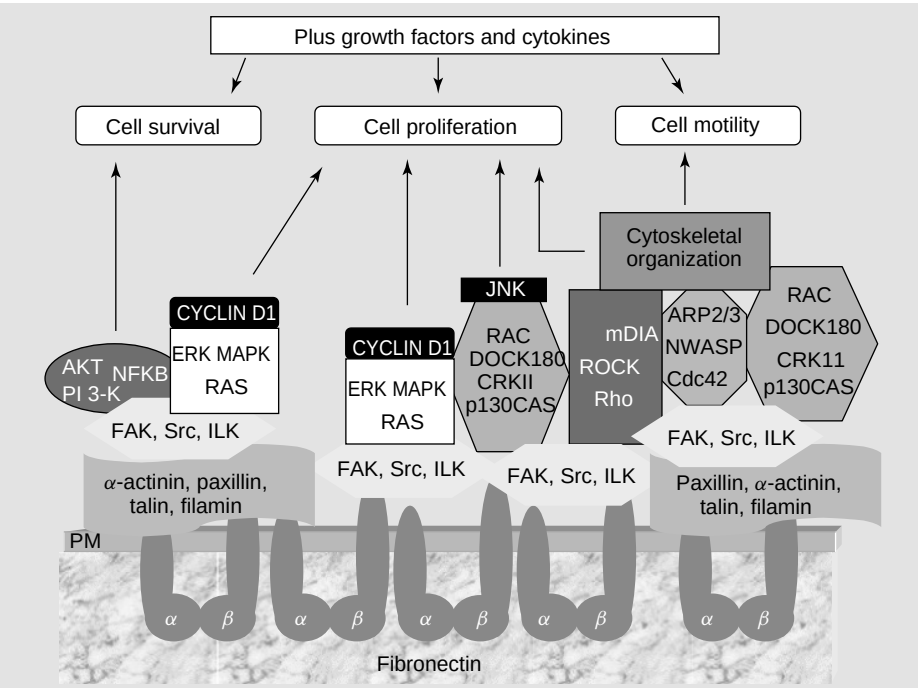

Fig. 3 Integrin signaling. Pathways regulated by integrin signaling that promote cell motility, proliferation, and survival are shown in the diagram. Integrins signal together with growth factor and cytokine receptors to regulate cell motility, survival, and proliferation, although the mechanisms of cooperation between these different types of cell surface receptors are not fully defined. Most of the cytoskeletal, adaptor and signaling proteins are described in the text. In addition to these are the NFKB transcription factor, the kinase ROCK, which is activated by Rho and in turn activates actinomyosin contractility, and mDIA, which is also activated by Rho to promote actin polymerization. Also shown is Cdc42, which activates NWASP, which in turn activates the Arp2/3 complex to nucleate actin polymerization.

that trigger many downstream signals. The small GTP-binding protein Ras gets activated downstream of FAK/Src kinases and, in turn, Ras triggers the activation of ERK MAP kinase and PI 3-kinase. However, FAK can also activate PI 3-kinase directly. In turn, PI 3-kinase plays a role in the activation of several other signaling proteins. It generates PIP3 phospholipids, which provide bindings sites at the plasma membrane for PH domains, which are conserved domains present on many different types of signaling proteins. PI 3-kinase signaling plays a role in activating the kinase AKT and the Rho family of GTP-binding proteins, Cdc42, Rac and Rho. AKT signaling promotes cell survival and proliferation, whereas the Rho family GTPases regulate the actin cytoskeleton. The Rho family is particularly important in adhesion signaling. Cdc42 and Rac promote actin polymerization, resulting in the formation of filopodia and lamellipodia, respectively. The formation of these structures drives cell spreading and migration. Rho promotes the formation of

stress fibers and focal adhesions to regulate cell–matrix adhesion and contractility.

FAK and paxillin are important regulators of cell spreading and cell migration. FAK/Src complexes trigger the phosphorylation of p130Cas. Phosphorylated p130Cas recruits additional signaling proteins including the adaptor protein CrkII, and Dock180, which is a guanine exchange factor (GEF) that activates Rac. Integrin-mediated adhesion is sufficient to activate Rac, and is also required for the membrane targeting of activated Rac and its association with downstream effectors. Thus, even when Rac is activated by signals from other cell surface receptors, such as growth factor receptors, integrin-mediated adhesion is still required for Rac to activate downstream signaling pathways. Activated FAK/Src complexes also trigger the tyrosine phosphorylation of paxillin, which can also lead to the recruitment of CrkII and potential activation of Rac by similar pathways. However, current evidence suggests that paxillin regulates migration by binding of paxillin-kinase linker (PKL). PKL is an Arf GTPase activating protein (GAP) that regulates vesicular trafficking. In addition, PKL binds to the Cdc42/Rac GEF, PAK interacting exchange factor (PIX), another GEF for Rac. PKL also binds to the serine/threonine kinase PAK, which is activated by Rac to promote cell migration. Thus, paxillin-PKL association may provide the appropriately localized scaffold for the activation of signaling complexes that are required to direct cell migration.

In addition, stable integrin-mediated adhesion cooperates with other cell surface receptors such as growth factor receptors to regulate cell behavior, such as proliferation, survival, and differentiation. Numerous studies have demonstrated an important role for integrins in the regulation of cell proliferation. Most normal cells require integrin-mediated adhesion to specific components of the ECM for cell proliferation; activation of growth factor receptors is not sufficient. Integrins promote cell proliferation by regulating the activities of the G1 cyclin-dependent kinases (Cdk4/6 and Cdk2). Integrin-mediated adhesion is necessary for the transcription of cyclin D1, which is required for the activation of the cyclin-dependent kinase(s) Cdk4/6. The cyclin D1 promoter is regulated by transcription factors that are phosphorylated and activated by ERK. Integrin-mediated adhesion is required for integrin- and growth factor-activated ERK to be translocated to the nucleus and for the sustained ERK activity that is necessary to induce transcription from the cyclin D1 promoter. Integrin signaling is also required for the activation of cyclin E-Cdk2. Integrin-mediated adhesion promotes cyclin E-Cdk2 activity by contributing to the downregulation of the Cdk inhibitors, p21 and p27.

In some cell types, the ability of cell–matrix adhesion to promote cell proliferation is dependent upon the integrin heterodimer mediating adhesion. The $\alpha_5\beta_1$-mediated endothelial cell adhesion to fibronectin promotes proliferation in response to growth factors, whereas $\alpha_2\beta_1$-mediated endothelial adhesion to laminin1 does not. Interestingly, this regulation is at the level of cyclin D1 translation and depends on Rac signaling, which occurs efficiently when cells are adhered to fibronectin, but not to laminin1. Thus, integrin signaling not only promotes the transcription of cyclin D1, but can also regulate the translation of cyclin D1 by modulating the activation of Rac signaling.

"Inside-out" integrin signaling provides cells with a mechanism to rapidly regulate integrin function. In some cell types, the ability of integrins to bind

their ligands is tightly regulated. In these cells, intracellular signals trigger conformational changes in integrins that regulate the ability of their extracellular domains to bind to their ligands. This is referred to as integrin "inside-out" signaling and is most prevalent in circulating blood cells, including platelets and leukocytes, such as neutrophils, monocytes and T-lymphocytes. The best-studied examples of inside-out signaling are the regulation of $\alpha_{IIb}\beta_3$ integrin on platelets and β_2 integrins and leukocytes. As mentioned above, the $\alpha_{IIb}\beta_3$ integrin promotes platelet aggregation by binding to soluble fibrinogen. Although platelet aggregation is important in preventing excessive bleeding when blood vessels are injured, inappropriate platelet aggregation can lead to thrombosis. Thus, the $\alpha_{IIb}\beta_3$ integrin is inactive on resting platelets, and when platelets become activated at sites of blood vessel injury, intracellular signaling events activate $\alpha_{IIb}\beta_3$. The $\alpha_{IIb}\beta_3$ integrin is then able to bind to soluble fibrinogen with high affinity and promote the formation of platelet aggregates or plugs to stop bleeding.

Similarly, the activity of β_2 integrins (Table 4) is also tightly regulated to prevent the inappropriate activation of an immune response. The activation of β_2 integrins involves conformational changes that increase the affinity of ligand binding; however, changes in the avidity with which β_2 integrins interact with their ligands is also important. Signals that increase the avidity of β_2 integrin-ligand interactions promote the diffusion and clustering of β_2 integrins on the cell surface. In the case of inactivated T-lymphocytes, β_2 integrins are tethered to the actin cytoskeleton in an inactive state. Signals that activate T-lymphocytes release β_2 integrins from the cytoskeleton. β_2 integrins are then free to diffuse on the plasma membrane, cluster and form new attachments to the cytoskeleton at adhesion sites.

Both inside-out and outside-in integrin signaling are mediated by conformational changes in the integrin heterodimer. Interestingly, the characterization of conformation-dependent antibodies to integrins has indicated that the conformations of activated and ligand-bound integrins are similar. Recent electron microscopy, X-ray crystallography, NMR, and mutational studies have provided important insight into the conformational regulation of integrin activity. These studies suggest that integrins in the low affinity state (inactive) have a bent, closed conformation with their globular ligand binding heads bent toward the plasma membrane and their α- and β-subunits closely associated (Fig. 4). In contrast, integrins in the high affinity state (fully active) have straightened and opened conformations with their α- and β-subunits separated and their globular domains extended away from the plasma membrane for ligand binding (Fig. 4). Integrins in the intermediate affinity state have their ligand-binding domains extended from the plasma membrane; however, their conformations are not fully opened. Current evidence suggests that these integrin conformations are in dynamic equilibrium and that conditions that inactivate or activate integrins shift this equilibrium.

1.3.4 Integrin-associated Proteins

The ability of integrins to mediate adhesion and to activate signals depends on their interaction with other cellular proteins (Table 5). Most integrins associate with the cytoskeleton and signaling complexes by protein interactions mediated by their β-subunit cytoplasmic domain (β-tail). In fact, clustering isolated β_1 and β_3

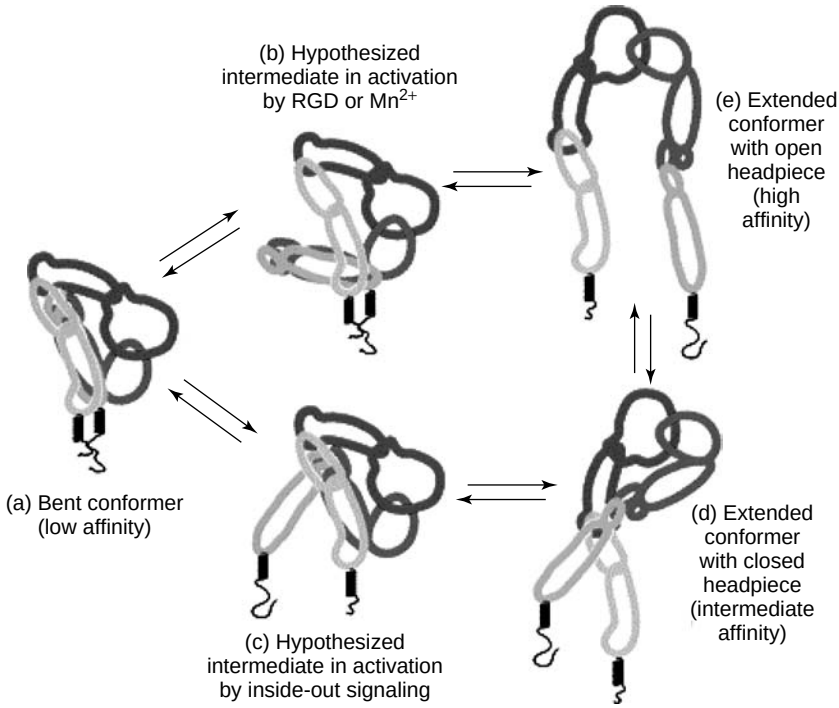

Fig. 4 Conformational changes in integrin structure associated with integrin activation. Results from X-Ray crystallography, electron microscopy and mutational studies suggest that integrins with different ligand-binding affinities have different conformations. Integrins with a bent structure are in the low affinity conformation (a), whereas Sho integrins with an extended structure with an open headpiece (consisting of the globular ligand-binding domains) are in the high affinity conformation (e). Integrins in intermediate affinity conformations have extended structures, but with closed headpieces (d). Other intermediate affinity conformations have been hypothesized (b and c). (Reprinted with permission from Takagi, J., Petre, B.M., Walz, T., Springer, T.A. (2002) Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling, *Cell* **110**, 599–611).

tails is sufficient to activate tyrosine kinase and Rac signaling.

The eight integrin β-tails have varying degrees of amino acid homology (Fig. 5) with the β_1, β_2, and β_3 tails showing the most similarity. Not surprisingly, several proteins have been shown to bind to more than one of these β-tails, whereas other protein interactions are β-tail specific. The β_8 tail has no homology with other integrin β-tails and to date no proteins have been identified that bind to the β_8 tail. The β_4 tail shows no similarity with other β tails except that it contains NPXY and NXXY motifs that are present in other β-tails. Protein interactions with the β_4 tail will be discussed in Sect. 1.5.

The β_1, β_2 and β_3 cytoplasmic tails share considerable amino acid sequence homology and the cytoskeletal linker protein talin has been shown to bind to all three tails. Since talin binds to actin and the actin binding protein vinculin, integrin-talin interactions are likely to regulate many

Tab. 5 Integrin binding proteins.

Integrin binding proteins	*β-tail*	*α-tail*	*Other regions*
Talin	$\beta_1, \beta_2, \beta_3$		
Filamin	$\beta_1, \beta_2, \beta_3, \beta_7$		
α-actinin	β_1, β_2		
Skelemin	β_1, β_3		
FAK	$\beta_1, \beta_2, \beta_3$		
ICAP-1	[a] β_1		
Paxillin[a]	β_1, β_3	α_4, α_9	
Rack1	$\beta_1, \beta_2, \beta_3$		
Cytohesin	[a] β_2		
β-endonexin	[a] β_2		
ILK	β_1, β_3		
SHC	β_3		
TAP-20	[a] β_5		
CD98	β_1, β_3		
Calreticulin		α_5	
CIB		α_{IIb}	
Nischarin		α_5	
CD47(IAP) (tetraspanins)			$\alpha_v\beta_3, \alpha_{IIb}\beta_3, \alpha_2\beta_1$
CD9 (tetraspanins)			Multiple integrins
CD81 (tetraspanins)			$\alpha_3\beta_1, \alpha_6\beta_1$
CD151 (tetraspanins)			$\alpha_3\beta_1, \alpha_6\beta_1, \alpha_6\beta_4, \alpha_7\beta_1$
Caveolin			$\alpha_1\beta_1, \alpha_5\beta_1, \alpha_v\beta_3$

[a] Paxillin has been shown to bind to both α- and β-tails.

processes requiring integrin-cytoskeletal interactions. Talin also binds to FAK, linking integrins to several signaling pathways (see Sect. 1.3.3). The cytoskeletal proteins, filamin and α-actinin, provide additional linkages between integrins and the actin cytoskeleton. The actin binding protein α-actinin can bind to the β_1 and β_2 tails, whereas filamin binds to the β_1, β_2, and β_7 tails. Filamin also links integrins to signaling pathways that regulate the actin cytoskeleton, including RalA, Rho family GTPases, their activators, as well as PAK. Thus, filamin–integrin interactions may be important in directing the actin assembly to sites of integrin-mediated adhesion. Skelemin has recently been shown to bind to the β_1 and β_3 cytoplasmic tails, and current evidence suggests that it may function in regulating integrin-mediated cell spreading. Skelemin is a muscle-specific protein that binds to myosin; however, evidence is accumulating suggesting that a skelemin-related protein is expressed in nonmuscle cells, suggesting that skelemin-β tail interactions may regulate integrin function in several cell types.

Integrin-β cytoplasmic domains have also been shown to bind to signaling proteins. The membrane-proximal regions of the β_1, β_2, and β_3 cytoplasmic domains bind to FAK, and the same region of the β_1 and β_3 tails can also bind to paxillin. The functional significance of these interactions is not yet known; however, mutations in the FAK binding region in the β-tail do not inhibit the activation

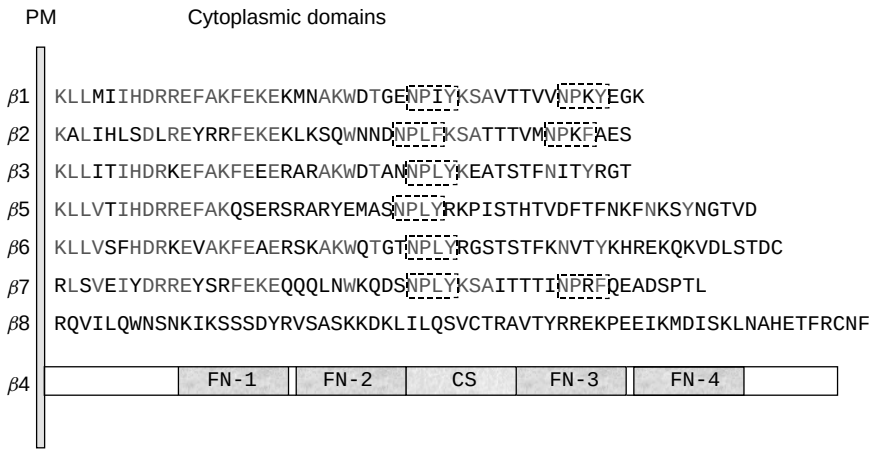

Fig. 5 Comparison of integrin β-subunit cytoplasmic domains. The amino acid sequences of the $\beta_1-\beta_8$ cytoplasmic domains are shown, with the exception of β_4. Amino acids that are conserved in at least three β-cytoplasmic domains are shown in red. The β_2 and β_7 cytoplasmic domains contain NPXF motifs in place of NPXY motifs found in other β-cytoplasmic domains. These phenylalanine residues (F) are shown in green. The conserved NPXY/F motifs are indicated by dashed boxes. The β_4 cytoplasmic domain contains 1000 amino acids with 4 regions homologous to conserved domains identified first in fibronectin, referred to as fibronectin type III repeats. These regions are labeled FN-1 to FN-4 in the β_4 cytoplasmic domain. The first pair of FN repeats is separated by a region referred to as the connecting segment (CS) (see color plate p. xxiii).

of FAK by integrins. Interestingly, phosphorylated FAK has been shown to bind to the C-terminal region of the β_5 tail. This interaction is important for vascular endothelial growth factor (VEGF) to promote migration mediated by the $\alpha_v\beta_5$ integrin. Integrin-linked kinase (ILK) has also been shown to bind to the β_1 and β_3 tails. In addition to its serine/threonine kinase activity, ILK also functions as an adaptor protein that localizes to focal adhesions. ILK plays a role in the regulation of cell adhesion, cell proliferation and fibronectin matrix assembly; however, there is no evidence that ILK-β tail interactions are required for ILK function in these processes.

β tail–specific protein interactions have been identified for the β_1, β_2, β_3, and β_5 cytoplasmic domains. Many of these interactions regulate integrin-ligand interactions. Cytohesins bind specifically to the β_2 cytoplasmic domain to upregulate β_2 integrin-dependent adhesion. β_3-endonexin binds specifically to the β_3 tail and increases the ability of $\alpha_{IIb}\beta_3$ to bind to its ligand. Although talin is not a β-tail-specific binding protein, it is interesting to note that the binding of a fragment of talin to the β_3 tail can also activate $\alpha_{IIb}\beta_3$. Thus, β_3-endonexin and talin may both regulate the activation of $\alpha_{IIb}\beta_3$. TAP-20 binds specifically to the β_5 tail and appears to play a role in promoting $\alpha_v\beta_5$-dependent migration by decreasing $\alpha_v\beta_5$-mediated adhesion. Interestingly, TAP-20 shares amino acid homology with β_3-endonexin, although their functions may be distinct since β_3-endonexin enhances $\alpha_{IIb}\beta_3$-ligand interactions. Integrin-associated protein-1 (ICAP-1) binds specifically to the β_1 cytoplasmic

domain; there is some evidence that this interaction promotes cell migration.

Recently, phosphotyrosine-binding (PTB) domains have been shown to be involved in some protein interactions with integrin β-tails. The PTB domain represents another conserved protein–protein interaction domain. PTB domains bind to NPXY and NPXF motifs found in many proteins, including most integrin β-tails. Recent studies have indicated that the region of talin that binds to integrins has a three-dimensional structure similar to a PTB domain. Talin's PTB domain has been shown to bind to the membrane-proximal NPXY motif in the integrin β-tail. Interestingly, this β-tail motif is required for most aspects of integrin function. Some PTB domains bind NPXY motifs when the tyrosine residue is phosphorylated. However, talin-integrin β-tail interactions do not require tyrosine phosphorylation. ICAP-1 also interacts with the integrin β_1 cytoplasmic domain via a PTP domain, but requires the C-terminal NPXY motif of the β_1 tail. Recently, PTB domains from several other proteins have been shown to bind to integrin β-tails, although the functional significance of these interactions is not yet known. Nonetheless, since NPXY/F motifs are highly conserved among the different β-tails, as well as across species, it is likely that β-tail–PTB domain interactions will play central roles in integrin function. The binding of some PTP domain-containing proteins to individual β-tails may require the tyrosine phosphorylation of the NPXY, motifs further increasing the diversity of protein interactions with β-tails and the mechanisms that regulate them.

Several proteins have been identified that bind to α-subunit cytoplasmic domains (Table 5). The functional significance of the interaction of paxillin with the α_4 tail has been established. This association modulates cell migration mediated by integrins containing α_4 tails. Additionally, serine phosphorylation of α_4 tails has been shown to inhibit the binding of paxillin. Thus, paxillin-α_4 tail association and α_4 integrin-dependent cell migration are likely to be regulated by signaling pathways that modulate the phosphorylation of the α_4 tail. Paxillin has also been shown to bind to the α_9 tail to regulate $\alpha_9\beta_1$-dependent cell migration. The ability of paxillin to regulate migration by this mechanism is likely to be important for $\alpha_4\beta_1, \alpha_4\beta_7$, and $\alpha_9\beta_1$ function during development and $\alpha_4\beta_1$ and $\alpha_4\beta_7$ function during the immune response.

Integrins also associate with other transmembrane proteins, including caveolin, tetraspanins (TM4 proteins) and CD47 (also known as integrin-associated protein, IAP). Caveolin binds to the α-subunit of several integrin heterodimers, including $\alpha_5\beta_1$ and links these integrins to Fyn, an Src family kinase that can couple integrins to the Ras/ERK MAP kinase pathway, which is important for cell proliferation. Several members of the TM4 family have been shown to bind to integrin α-subunit extracellular domains. The strongest association appears to be between CD151 and $\alpha_3\beta_1$. The cytoplasmic domain of CD151 recruits the signaling proteins phosphatidylinositol 4-kinase and protein kinase C to CD151-integrin complexes. These signaling proteins regulate the actin cytoskeleton and cell spreading. Interestingly, mutations that inhibit the association of CD151 and $\alpha_3\beta_1$ also inhibit cell spreading, indicating that this association is functionally significant. CD47 is also an important regulator of integrin function. CD47 associates with β_1, β_2,

and β_3 integrins in cholesterol-rich membrane domains. CD47-integrin association enhances the interaction of integrins with their ligands, as well as integrin function in adhesion and migration. The mechanisms involve signaling through heterotrimeric Giα proteins.

1.4 Focal Adhesions

1.4.1 Components of Focal Adhesion

One of the most studied cell–matrix junctions is the focal adhesion, sometimes referred to as the focal contact. Focal adhesions contain clustered integrins that mediate very tight adhesion with their ECM ligands and also associate with the ends of actin stress fibers. More than 50 proteins have been shown to localize to focal adhesions. Not surprisingly, many of the proteins that can bind to integrin-β cytoplasmic domains, as well as proteins that function in integrin signaling pathways are concentrated in focal adhesions (Figs. 3 and 6). As discussed above, several of these proteins are tyrosine-phosphorylated; thus, focal adhesions can be visualized by using antibodies that recognize phosphotryrosine residues (Fig. 7). In general, focal adhesion proteins contain multiple protein–protein interaction domains enabling them to bind to several proteins simultaneously. Indeed, many focal adhesion proteins function in multiprotein complexes linking integrins, the actin cytoskeleton and multiple signaling cascades. For example, FAK binds to paxillin, talin, p130Cas, Src family kinases, PI 3-kinase, Grb2, and Grb7. Paxillin binding partners include vinculin, FAK, ILK, the tyrosine phosphatase PTP-PEST, PKL, and CrkII. Long lists of binding partners can be made for most focal adhesion proteins. On the basis of biochemical studies that have characterized protein–protein

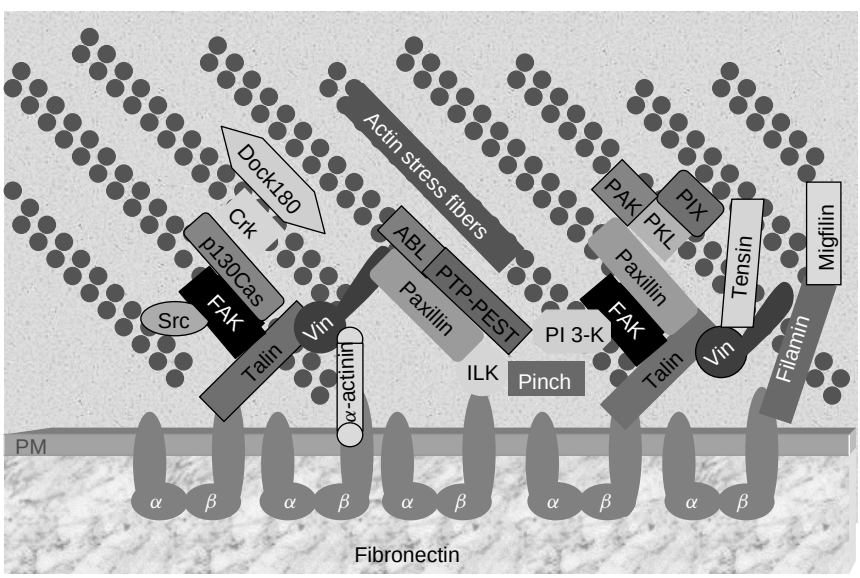

Fig. 6 Components of focal adhesions. Shown is a schematic of a focal adhesion. Only a subset of proteins and protein interactions present at focal adhesion is shown. Vinculin is abbreviated as vin.

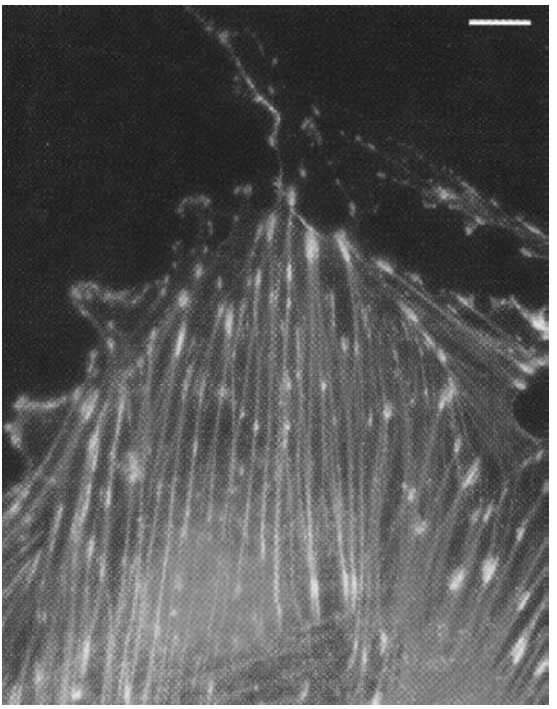

Fig. 7 Tyrosine phosphorylated proteins at focal adhesions. Shown is a rat embryo fibroblast stained for actin in stress fibers (green) and with antibodies that recognize phosphotyrosine (red). Phosphotyrosine staining is seen in focal adhesions at the ends of stress fibers. Regions of overlapping red and green staining are yellow. (Reprinted with permission from Burridge, K., Chrzanowska-Wodnicka, M., Zong, C. (1997) Focal adhesion assembly, *Trends Cell Biol.* **7**, 342–347.) (see color plate p. xxvii).

interactions involving integrins and other focal adhesion proteins, complex linkage maps have been proposed connecting integrins to the activation of various signaling pathways (Figs. 3 and 6). Surprisingly, new focal adhesion proteins are still being identified every year. The most recently identified focal adhesion protein is the Migfilin, which is a filamin-binding protein that regulates cell spreading.

1.4.2 Regulation of Focal Adhesion Assembly

Focal adhesions are formed by most cell types when cultured on two-dimensional substrates. The assembly of focal adhesions has been studied in spreading and migrating cells. In these instances, the activation of Rac or Cdc42, either by integrin or growth factor signaling, promotes the formation of filopodia and lamellipodia and new integrin-mediated cell–matrix junctions. These nascent adhesions, referred to as focal complexes, contain integrins, as well as a subset of focal adhesions proteins. Following the activation of Rho and actinomyosin contractility, focal complexes mature into focal adhesions with their associated stress fibers. The formation of focal adhesion has also been studied in quiescent cells, where the activation Rho and downstream myosin II-dependent contractility by extracellular factors induces focal adhesion formation. Focal adhesion formation can also be promoted when external force is applied to integrin-mediated adhesions. When this force is increased, focal adhesions become larger, whereas a reduction in the applied force causes disassembly of focal adhesions. The stimulation of focal adhesion formation by external force also requires Rho signaling,

but appears to be independent of myosin II. Importantly, the stimulation of focal adhesion formation by external forces leads to activation of integrin signaling with physiological consequences. For example, increases in blood flow result in enhanced hemodynamic shear stress placed on integrin-mediated adhesions on the basal surface of endothelial cells, resulting in the remodeling of adhesion sites and the activation of integrin signaling. The increased integrin signaling and subsequent changes in gene expression are thought to play an important role in the development of atherosclerotic plaques. Thus, integrin-mediated adhesions are important sites at which cells sense and respond to changes in their physical environment, a process referred to as mechanotransduction.

1.5 Hemidesmosomes

The basal layer of stratified epithelia mediate specialized adhesions to their underlying basement membrane. These specialized adhesions are referred to as hemidesmosomes. $\alpha_6\beta_4$ integrins are the only integrins present at hemidesmosomes, and function as critical transmembrane links between their ligand laminin 5 in the basement membrane and the keratin filament cytoskeleton within the cell. The cytoplasmic domain of the β_4 subunit plays an important role in this linkage (Fig. 8), by binding to the cytoskeletal linker proteins plectin and BP230, which, in turn, can bind directly to intermediate filaments. BP180

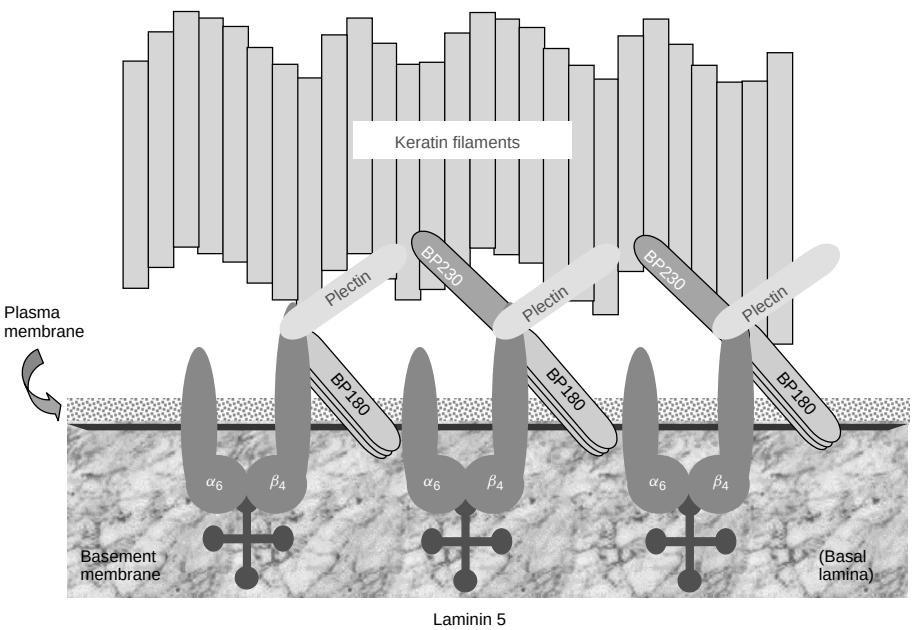

Fig. 8 Hemidesmosomes. Hemidesmosomes are sites of strong adhesion between cells and their underlying basement membrane. The $\alpha_6\beta_4$ integrin binds to its ligand laminin-5 in the basement membrane. The transmembrane protein BP180 has a collagenous extracellular domain that forms a trimer and associates with the basement membrane on the outside of the cell. The cytoplasmic domain of BP180 binds to the β_4 tail and plectin on the inside of the cell. Both plectin and BP230 function in linking $\alpha_6\beta_4$ to keratin filaments within the cell.

is another transmembrane protein that functions at hemidesmosomes. BP180 has a collagenous extracellular domain; its intracellular domain binds to the β_4 cytoplasmic domain.

The physiological importance of hemidesmosomal components is clear from studies of inherited human skin blistering disorders. In humans, loss of $\alpha_6\beta_4$ expression or laminin 5 expression causes a form of epidermolysis bullosa (EB). Patients with this disorder have very fragile skin, which is prone to extensive blistering due to a lack of normal hemidesmosomes. The adhesion of other epithelia is fragile as well, affecting organ function such as the intestines. Most patients suffering from this disorder survive only several months after birth at most. Mutations in the BP180 gene cause a less severe form of EB. Another form of EB results from mutations in the plectin gene; this form of EB is associated with muscular dystrophy indicating the importance of plectin function in other tissues.

Mice with null mutations in genes encoding hemidesmosomal components have similar phenotypes to those described for the corresponding human hereditary disorder. These null mice have been useful animal models for these diseases, providing important insights into the functions of hemidesmosome components. For example, the skin of plectin-null mice showed a reduced number of hemidesmosomes, and the hemidesmosomes present had fewer associated keratin filaments and exhibited reduced mechanical stability. Thus, plectin is not strictly required for the association of hemidesmosomes with keratin filaments. This is not true for BP230. Targeted deletion of BP230 in mice also results in a skin blistering phenotype. The number of hemidesmosomes present in the skin of BP230-null mice was similar to wild-type animals; however, the hemidesmosomes in BP230-null mice were not associated with keratin filaments. Thus, BP230 plays an essential role in connecting $\alpha_6\beta_4$ to keratin filaments at hemidesmosomes *in vivo*.

1.5.1 Assembly of Hemidesmosomes

Biochemical and molecular genetic approaches have been used to characterize protein interactions with the β_4 cytoplasmic domain and among the components of the hemidesmosome. Using these approaches, the β_4 cytoplasmic domain has been shown to bind directly to plectin, BP230 and the cytoplasmic domain of BP180. Other studies have shown that BP180 binds to plectin and BP230. The β_4 cytoplasmic domain has also been shown to associate with itself; the C-terminal region of the β_4 cytoplasmic domain binds to a site more toward the N-terminal that overlaps the binding site for plectin. This β_4 tail-β_4-tail interaction could potentially occur either intramolecularly or intermolecularly, and it has been suggested that the binding of plectin to the β_4 tail disrupts this interaction. The importance of individual interactions between the various hemidesmosomal components has been examined by testing the effect of expressing recombinant forms of these proteins, containing mutations that inhibit specific protein–protein interactions, on the formation of hemidesmosome-like structures on the basal cell surface of cultured epithelial cells. Emerging evidence from these types of studies suggests that the assembly of hemidesmosomes occurs by a stepwise process. An early step is the binding of plectin to the β_4 cytoplasmic domain. The next step is the recruitment of BP180, which requires the interaction of BP180 with both β_4 and plectin, as well as the association of plectin with the β_4 cytoplasmic

domain. Finally, BP230 is recruited. The recruitment of BP230 requires $\alpha_6\beta_4$ adhesions containing both BP180 and plectin. Thus, plectin appears to function in the assembly of hemidesmosomes by facilitating the interaction of BP230 and BP180 with the β_4 cytoplasmic domain and also by stabilizing the association of keratin filaments with hemidesmosomes.

1.5.2 Signaling by $\alpha_6\beta_4$

Like other integrins, $\alpha_6\beta_4$ can function as a signaling receptor. $\alpha_6\beta_4$ signaling impacts the survival, proliferation, and migration of epithelial cells. $\alpha_6\beta_4$ activates ERK MAP kinase that plays a critical role in regulating cell proliferation and survival. The β_4 cytoplasmic domain becomes phosphorylated on tyrosine residues when $\alpha_6\beta_4$ binds to laminin 5. These phosphorylations are important in linking $\alpha_6\beta_4$ engagement to the activation of intracellular signals. Although $\alpha_6\beta_4$ can signal in epithelial cells that do not form hemidesmosomes, it is not yet known whether $\alpha_6\beta_4$ integrins can activate signals when associated with intermediate filaments at hemidesmosomes. However, the activation of the epidermal growth factor (EGF) receptor can trigger the phosphorylation of the β_4 cytoplasmic domain resulting in the disassembly of hemidesmosomes, which is likely to be important for cell migration and proliferation. The $\alpha_6\beta_4$ integrin has been found to play a role in promoting cell migration and invasion of some carcinomas by activating PI 3-kinase, and the small GTP-binding proteins Rac and Rho. In cell culture models, $\alpha_6\beta_4$ integrin can promote cell migration on laminin 1 by associating with the microfilament cytoskeleton and stabilizing the adhesion of lamellipodia and filopodia to laminin1, suggesting that $\alpha_6\beta_4$ can function in both intermediate filament-associated and actin-microfilament-associated cell–matrix adhesions. The $\alpha_6\beta_4$ integrin is also expressed by the endothelial cells and neuronal cells. The adhesion complexes formed by $\alpha_6\beta_4$ in these cell types have not been fully characterized. However, it is known that $\alpha_6\beta_4$ associates with the intermediate filament cytoskeleton in endothelial cells and is likely to do so in neuronal cells as well.

In summary, members of the integrin family of cell adhesion receptors play important roles in most physiological processes. Integrins regulate these processes by providing a transmembrane connection between the various integrin ligands and components of the cell's cytoskeletal networks, and also functioning as signaling receptors in cooperation with other cell surface receptors.

2 Cell–Cell Junctions

2.1 Overview

Cell–cell junctions serve to bind cells together allowing for these cells to act as a tissue and mediate physiological functions within an organism. For example, cell–cell junctions in endothelial cells, which line the walls of all blood vessels, allow endothelial cells to regulate the movement of fluid, macromolecules, and immune cells from the vascular space to the extravascular space of tissues. Cell junctions are separated into groups according to their functions, and these groups include adhesive junctions, occlusive junctions and junctions that allow direct communication between cells. These junctions have been extensively studied in epithelial cells and are shown in Fig. 9. Adhesive junctions include desmosomes and adherens

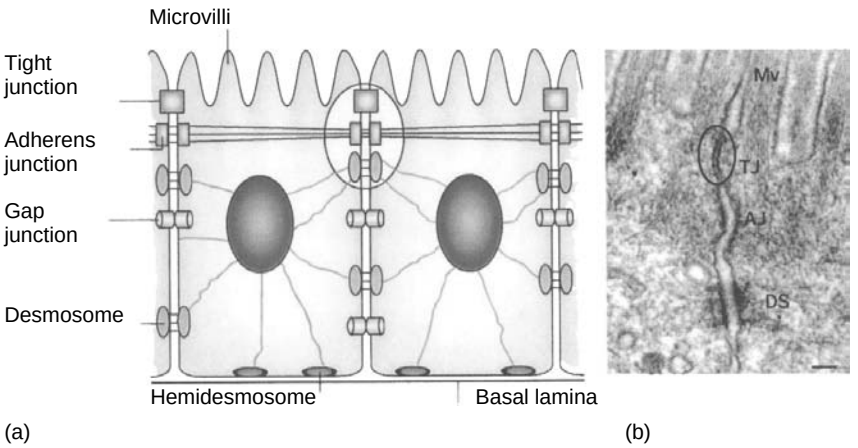

Fig. 9 Junctional complexes in epithelial cell monolayer. (a) Schematic showing the cellular location of tight junctions, adherens junctions, desmosomes, and gap junctions in an epithelial cell layer. (b) Electron micrograph of tight junctions (TJ), adherens junctions (AJ), and desmosomes (DS). Note the electron-dense regions associated with the plasma membrane in each junction. This is due to the localization of plaque proteins in these areas. (Reprinted with permission from Tsukita, S., Furuse, M., Itoh, M. (2001) Multifunctional strands in tight junctions, *Nat. Rev. Mol. Cell Biol.* **2**, 285–293).

junctions (AJs). Desmosomes are found primarily in tissues that experience mechanical stress, such as epithelial cells of the skin and myocardial cells of the heart. These junctions are connected to intermediate filaments and are thought to have mainly a structural function. AJs are connected to the actin cytoskeleton and also serve a structural function. Indeed, in certain epithelial cells such as those in the intestine, the AJ acts together with the actin cytoskeleton to allow contraction of the epithelial sheet. Occlusive junctions, also termed *tight junctions* (TJ), form restrictive barriers that regulate the paracellular flux of molecules across epithelial and endothelial monolayers, and act as a fence to prevent the movement of membrane components from the apical to the basal surface of a cell layer. Gap junctions allow for communication between cells. These junctions are formed by proteins called *connexins*. Connexins form a hexameric multimer that create a pore in the plasma membrane of the cell. The hexameric pore binds to a similar pore on an adjacent cell, thereby forming a continuous pore between cells that allows for the passage of ions and small molecules between neighboring cells. Gap junctions provide electrical and chemical coupling between cells whereas AJs, desmosomes, and TJs provide mechanical coupling between cells. The following discussion will focus on the junctions that provide mechanical coupling. A number of recent reviews provide further information on gap junctions and their role in cell–cell communication.

A common theme exists with respect to the structure of cell–cell junctions. All contain transmembrane proteins whose extracellular domains mediate binding to similar proteins on neighboring cells. The transmemebrane proteins of desmosomes and AJs are members of the cadherin

superfamily. Claudins, occludins, and junctional adhesion molecules (JAM) are all transmemebrane proteins that participate in the formation of TJs. The ultrastructures of TJs, AJs and desmosomes all display a dense plaque region adjacent to the plasma membrane when observed through an electron microscope. This region appears as a dense plaque due to the large number of adaptor proteins that form a complex that supports the function of the transmembrane proteins (Fig. 9b). The cytoplasmic domains of the transmembrane proteins bind to linker proteins that connect the transmembrane proteins to the cytoskeleton. In addition, both the cytoplasmic domain of the transmembrane proteins and the linker proteins can bind to and localize signaling proteins, such as the Src family of tyrosine kinases, to the junctional complex. Thus, as is the case in focal adhesions mediated by the integrins, cell–cell junctions too are now considered sites that mediate inside-out or outside-in signaling.

2.2 Cadherin Superfamily

In the early 1980s, two proteins were being investigated for their role in mediating cell–cell adhesion: uvomorulin and L-CAM. Upon sequencing of these proteins it was found that they shared a large number of amino acids and a number of structural features such as the requirement of calcium for cell–cell adhesion. These proteins became collectively known as E-cadherin. At this time, two other transmembrane proteins having a similar structure and amino acid sequence to E-cadherin were also identified: N-cadherin and P-cadherin. Most of what is known about cadherins has been learned from studies of E-, N-, and P-cadherin. To date, over 50 different cadherins have been identified in both vertebrates and invertebrates. The discovery of novel cadherin-like proteins, including the recent identification of the protocadherins, has lead to the formation of the cadherin superfamily containing cadherins and cadherin-related proteins.

All members of the cadherin superfamily possess cadherin homology repeats, referred to as EC domains, in their extracellular region. Cadherin-related proteins show a high degree of variability in the number of cadherin repeats in their extracellular domain with reported numbers ranging from 5 to 34. In addition, extracellular regions of cadherin-related proteins can contain structural motifs, such as laminin-A G repeats or EGF repeats, found in other proteins. A high degree of variability also exists with respect to the membrane-spanning region of cadherin-related proteins. Although most cadherin-related proteins contain a single spanning transmembrane region, some members contain seven membrane-spanning regions, while others are anchored to the plasma membrane by glysyl phosphotidylinositol (GPI) linkages. The cadherin-related proteins also show a diverse array of cytoplasmic tails with unique binding domains to a wide variety of molecules. The following discussion will not include cadherin-related proteins due to the variability of their structures and functions.

The other members of the cadherin superfamily have a number of identical structural features. These include five cadherin repeats in the extracellular domain, a single spanning transmembrane region, and a highly conserved cytoplasmic tail. In addition, these cadherins are synthesized as precursor molecules containing a signal peptide and a propeptide

region that are both removed by proteolytic cleavage.

Cadherins have been divided into five families on the basis of the domain structure and genomic organization of the cadherin proteins (Table 6). The five families are (1) the classical/type I cadherins; (2) the atypical/type II cadherins, the two families of desmosomal cadherins; (3) desmocollins; (4) desmogleins; and (5) a family containing only one member, BS-cadherin. Sequence similarities in the EC1 domain allow for classification of the different cadherins into families. Type I cadherins all have the triple amino acid sequence H-A-V in the first cadherin repeat (EC1). This sequence has been termed the cell adhesion recognition (CAR) sequence (Fig. 10), and as will be discussed, is required for cell–cell adhesion of type I cadherins. The other cadherin family members also have a central alanine in their CAR sequence, but the surrounding residues are different for each family. The HAV sequence has been found to interact with the N-terminal tryptophan (Trp2) residue also found in the EC-1 domain (Fig. 10). The amino acid sequence surrounding the Trp2 residue is also highly conserved within the type I cadherin family. If one looks at a sequence comparison surrounding the N-terminal

Tab. 6 Major cadherin subfamilies. (Adapted from Nollet, F., Kools, P., van Roy, F. (2000) Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members, *J. Mol. Biol.* **299**, 551–572).

Subfamily	*Protein name*	*Synonyms*
Classical/type I	Cadherin 1	E-cadherin, uvomorulin
	Cadherin 2	N-cadherin
	Cadherin 3	P-cadherin
	Cadherin 4	R-cadherin
	Cadherin 13	T- or H-cadherin
	Cadherin 15	M-cadherin
	EP-cadherin	C-cadherin (*Xenopus*)
Atypical/type II	Cadherin 5	VE-cadherin
	Cadherin 6	K-cadherin
	Cadherin 7	
	Cadherin 8	
	Cadherin 9	T1-cadherin
	Cadherin 10	T2-cadherin
	Cadherin 11	OB-cadherin
	Cadherin 12	Br-cadherin, N-cadherin 2
	Cadherin 18	Cadherin 14
	Cadherin 19	
	Cadherin 20	
Desmocollin	Desmocollin 1	
	Desmocollin 2	
	Desmocollin 3	
Desmoglein	Desmoglein 1	
	Desmoglein 2	
	Desmoglein 3	
BS Cadherin		

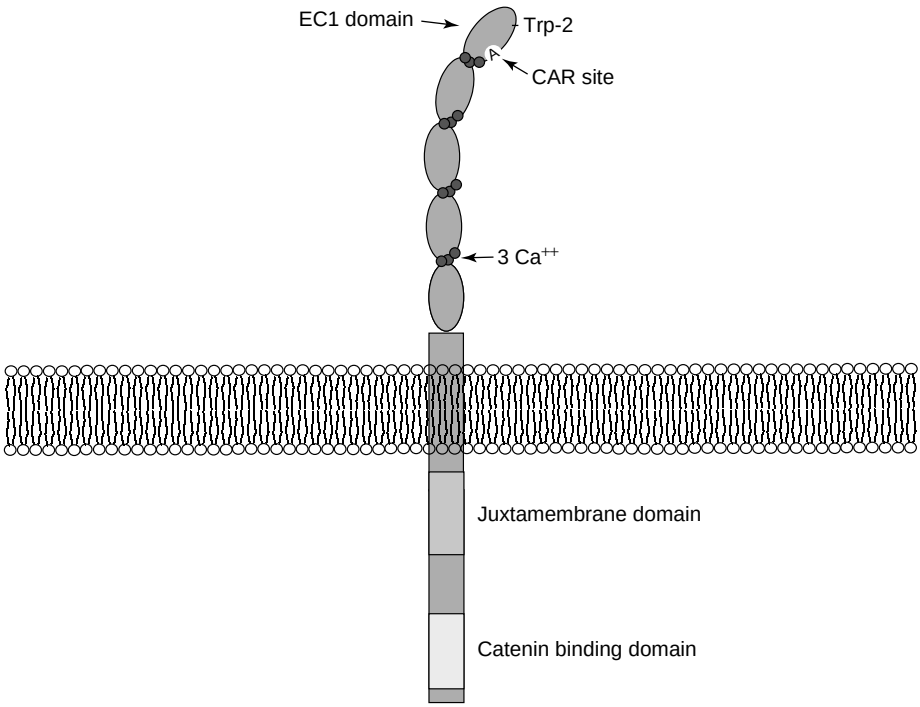

Fig. 10 Cadherin monomer. Schematic showing the important sites that are implicated in the regulation of cadherin-mediated adhesion, including the first cadherin repeat (EC1 domain), the N-terminal tryptophan (Trp2), cell adhesion recognition site (CAR site), calcium (Ca++) binding sites, the juxtamembrane region, and catenin binding domain.

Trp2 residue in the type II cadherins desmocollin and desmoglein, one also finds a high degree of sequence similarity within each type or family of cadherin. In addition to sequence similarities, genomic organization is also used to classify cadherins. For example, the precursor region and the cytoplasmic tail of atypical/type II cadherins are encoded by a single exon, whereas these regions in other cadherins contain one or more introns. The type II cadherins also contain two extra introns in the sequence coding the extracellular region that is not found in other cadherin families. On the basis of these structural and genomic similarities members of the cadherin family have been divided into families as shown in Table 6.

2.3 Type I/II Cadherins and Adherens Junctions

The binding of cadherins and the formation of stable AJs is a multistep process. The formation of cell–cell junctions by classical cadherins has been extensively studied in epithelial cells. In epithelial cells, cadherins are diffusely spread on the cell surface prior to cell–cell contact. Cells first approach one another through the formation of filopodia. As the filopodia from adjacent cells come into contact, cadherins form small clusters at the filopodia tip and recruit the proteins that promote cortical actin assembly. These clusters are brought into larger plaque-like structures as the cortical actin remodels. The binding

and formation of a mature cell–cell junction requires the following steps involving cadherins: (1) the binding of one cadherin to a second cadherin on the same cell, referred to as cis-dimer formation; (2) the binding of a cadherin on one cell to a cadherin on an adjacent cell, referred to as trans-dimer formation; (3) the formation of lateral complexes or clusters; and (4) the attachment of these complexes to the cytoskeleton to enhance the strength of the cell–cell attachment. The first two steps are mediated by the extracellular domain, whereas the last two steps occur through domains on the cytoplasmic tail.

2.3.1 Extracellular Domain: Cis- and Trans-cadherin Dimerization

The extracellular domain mediates cadherin–cadherin binding. The extracellular domain consists of five cadherin domains that form a rigid, rodlike structure through the binding of three calcium ions to the intervening sequence located at the base of each domain as shown in Fig. 10. Lateral cis-dimerization is the binding of two adjacent cadherins within the plasma membrane of one cell and is mediated by the interaction of the EC-1 and EC-2 regions of two cadherin molecules (Fig. 11). This process is thought to require the Trp2 and the hydrophobic pocket created by the HAV region on the EC-1domain. Indeed, crystallographic data has shown two possible scenarios for how these two peptide regions interact to establish the formation of a cadherin cis-dimer. In the first model, the Trp2 on one cadherin inserts into the hydrophobic pocket created by the HAV region on an adjacent cadherin. In the second model, the Trp2 inserts into the hydrophobic pocket on the same EC-1 domain resulting in a change in the conformation of the surrounding amino acids, which allows interaction with the EC-1 domain on an adjacent cadherin.

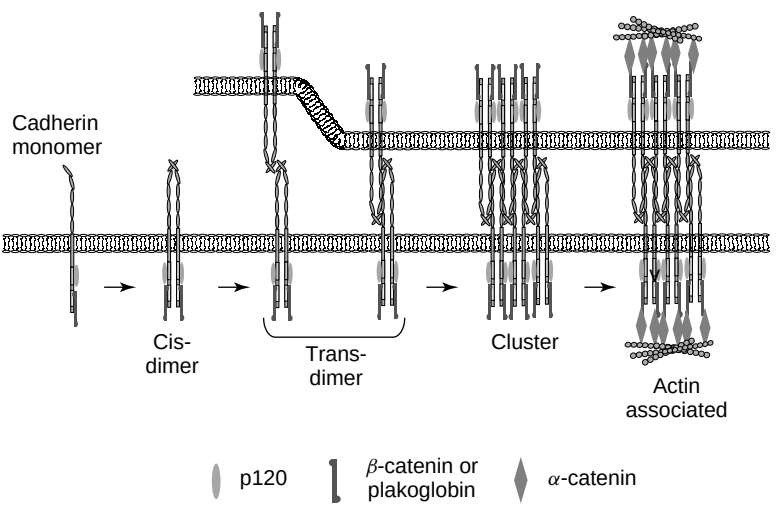

Fig. 11 Formation of cadherin-mediated cell–cell junction. Cadherin monomers first form cis-dimers through the interaction of the first two cadherin repeats. This is followed by formation of trans-dimers through the interaction of cadherin cis-dimers on adjacent cells. The trans-dimers then cluster, increasing the strength of cell–cell adhesion. The adhesion strength is increased further by the attachment of the cadherin–catenin complex to the actin cytoskeleton. (See text for details).

Interestingly, a recent crystallographic study using C-cadherin has suggested a third model in which cis-binding may actually form lateral multimers through interaction of EC-1/EC-2 domains on adjacent cadherins. Although the mechanism of cis-interaction of cadherins remains controversial, crystallographic, mutagenesis, and *in vitro* binding studies all support a requirement for cis-dimerization in cadherin-mediated cell–cell adhesion.

The structural basis involved in mediating homophilic binding of cadherins on adjacent cells is also fraught with controversy. As in cis-dimerization, trans-dimerization, which is the binding of two cadherins on different cells, has also been shown to require the EC-1 domain. In one model, the EC-1 domains of two cadherins on opposing cells bind to form a zipper-like structure as shown in Fig. 11. This model is supported by studies that have shown that the HAV motif and the amino acids surrounding this peptide mediate the specificity of cadherin adhesion. However, a trans-dimer with this structure seems too long to be accommodated in the intercellular space observed by electron microscopy. In a second model, the EC-1 domain of one cadherin interacts with the EC-4 or EC-5 domain of the cadherin on an adjacent cell. This model is consistent with the distance between plasma membranes observed by electron microscopy and is also supported by studies showing that antibodies to the EC-4 and EC-5 regions disrupt cadherin-mediated binding. In addition, direct force measurements would suggest that a greater overlap of the cadherins shown in Fig. 11 is required for maximum adhesive strength, and that interaction mediated only by the EC-1 domain would not be strong enough to maintain cell–cell adhesion. Nonetheless, both models support the formation of a zipper-like structure by the trans-interaction of cadherins on opposing cells.

Understanding the mechanisms of cadherin-mediated adhesion has important implications for therapeutic interventions in processes such as tumor angiogenesis. Monoclonal antibodies to the EC1 region of VE-cadherin have been shown to prevent branching morphogenesis of endothelial cells in culture and to inhibit angiogenesis in a mouse cornea model. Unfortunately, antibodies to the EC1 domain also cause a decrease in barrier function of confluent endothelial monolayers, particularly in the microcirculation of the lung. However, monoclonal antibodies to the EC4 domain inhibit angiogenesis, without altering the barrier function of endothelia with stable, well-structured AJs, such as those found in the lung. These results suggest that antibodies to the EC4 domain may be used to prevent angiogenesis and thus inhibit the growth of many tumors without having the adverse side effect of changing vascular permeability. These results further underscore the importance of understanding structure–function relationships that govern cadherin-mediated cell–cell adhesion.

Using cell culture aggregation assays, it was demonstrated that the EC1 domain and the HAV region play essential roles in determining the specificity of homophilic binding of classical cadherins. In early studies, E-cadherin and P-cadherin were expressed as recombinant proteins in fibroblasts that normally do not express these cadherins and do not demonstrate calcium-dependent cell–cell adhesion. These studies demonstrated that when cells expressing E-cadherin were mixed with cells expressing P-cadherin, the resulting cell aggregates would contain either E-cadherin cells or P-cadherin

cells but not a mixture of the two. Mutational studies demonstrated that the EC1 domain and the amino acids surrounding the HAV peptide play an important role in homophilic trans-cadherin adhesion. Further studies using domain-swapping approaches, antibodies or peptides to the EC1 domain confirmed these findings. Interestingly, recent studies have suggested that cadherins may not be restricted to homophilic binding, but may also interact with each other in a heterophilic manner. This is particularly true of the type II cadherins. The members of this family of cadherins have been shown to bind in a heterophilic manner to other members of the same family. For example, cadherin 8 can bind to cadherin 11 allowing for the formation of cell aggregates with strong cell–cell adhesion. One should remember that type II cadherins do not contain the HAV sequence and do not bind at all to type I cadherins. Recent studies have found that heterophilic binding may also occur within the type I cadherin family; however, the studies have suggested that these heterophilic interactions are weaker than homophilic interactions. As the studies demonstrating heterophilic binding of cadherins have been performed only in cell culture systems, the physiological relevance of heterophilic binding in tissue and organ function is unknown.

2.3.2 Cytoplasmic Domain: Lateral Clustering and Cytoskeletal Attachment

Changes in the strength of cadherin-mediated adhesion are necessary for the dynamic cellular movements found during processes such as angiogenesis and tissue morphogenesis. Environmental signals, such as growth factors, act to modulate the strength of cadherin-mediated adhesion by changing the association of cadherin with the actin cytoskeleton and by disrupting cadherin clustering. Both of these processes are regulated through the cadherin cytoplasmic tail. The cytoplasmic tail of cadherin contains two important domains, the juxtamembrane domain (JMD) and the C-terminal catenin binding domain (CBD), both of which bind to the catenins to form a cadherin–catenin complex. The catenins are a family of related proteins that share a central domain called the armadillo-repeat domain (ARM domain). This domain contains 10 to 13 armadillo repeats, so named because of their sequence homology to the armadillo gene product first identified in *Drosophila*. The ARM repeats have been shown to have a tertiary structure that results in the formation of a groove that contains a number of positively charged residues. Proteins that bind to the ARM region do not possess a single consensus sequence, but do possess many negatively charged amino acids that bind within the positively charged groove created by the ARM domain.

The extracellular domain of cadherin can support cell–cell adhesion independent of the cytoplasmic tail, but the strength of adhesion is considerably weaker than that supported by the full-length molecule. The binding mediated by the extracellular domain of cadherins can be increased by clustering cadherin extracellular domains on the cell surface. This was shown using chimeric cadherin molecules containing an intracellular regulatable oligomerization motif connected to the extracellular and transmembrane domain of cadherins. Under normal conditions, the JMD of the cytoplasmic tail is believed to regulate cadherin clustering. Deletion mutants of cadherin that contain the JMD, but not the CBD, were found to cluster and mediate strong adhesion demonstrating the importance of the JMD in regulating cadherin function.

The interaction of the JMD with p120 catenin has been implicated in regulating lateral clustering and the strength of cadherin-mediated cell–cell adhesion. p120 is a member of the p120 family, which includes p120, p0071, δ-catenin, and AVRCF. These proteins have greater than 45% identity in the ARM domains and bind to cadherins via the interaction of the ARM domain with a highly conserved region in the JMD. The most extensively studied member of this family to date is p120. Initial studies presented conflicting results concerning the effect of p120 on cadherin function with one study showing that the p120 cadherin interaction was required for lateral clustering and enhanced the strength of cell adhesion, while a second study presented contrasting data suggesting that the JMD and its interaction with p120 had a negative effect on cell–cell adhesion. This discrepancy was attributed to differences in the phosphorylation state of p120 in the different cell types that were used for these studies. Indeed, phosphorylation of p120 on serine and threonine residues in the N-terminus results in a decrease in cell–cell adhesion, whereas inhibition of phosphorylation of these residues using staurosporin, or deletion of the N-terminus results in an increase in cadherin-mediated cell–cell adhesion. In addition to containing serine/threonine phosphorylation sites, the N-terminus of p120 also contains a number of tyrosine phosphorylation sites. Although the phosphorylation of p120 has been shown to alter cadherin function, the molecular mechanism responsible for the change in cadherin function with phosphorylation of p120 has not been reported.

The interaction of p120 with the JMD has been implicated in mediating a number of functions in addition to cadherin clustering. For example, the juxtamembrane region of VE-cadherin has been implicated in excluding N-cadherin from cell–cell junctions in cells expressing these two cadherins, which may be why N-cadherin is not observed at cell–cell junctions in endothelial cells. The interaction of JMD with p120 has also been shown to protect E-cadherin from degradation. The colon tumor cell line SW48 possesses a defect in the p120 gene that prevents expression of p120. These cells show minimal expression of E-cadherin and have poor cell–cell adhesion. Interestingly, restoring the levels of p120 increased both the E-cadherin levels and E-cadherin-mediated cell–cell adhesion. We have found similar results in endothelial cells, where a decrease in the level of p120 results in a decrease in the level of VE-cadherin and a decrease in endothelial monolayer integrity. Thus, interaction of p120 with the JMD region of at least some cadherins is required for the expression at the plasma membrane.

The C-terminus of the cytoplasmic tail contains the CBD and mediates attachment to the actin cytoskeleton through binding to β-catenin, plakoglobin (also known as γ-catenin), and α-catenin. Plakoglobin and β-catenin share a similar structure containing 12 ARM repeats found in the central ARM domain that is flanked on either side by N-terminal and C-terminal domains. The binding of β-catenin and plakoglobin to cadherins is mediated through the ARM domains, and these two proteins bind to the CBD in a mutually exclusive fashion. Attachment to the cytoskeleton is then achieved by the binding of α-catenin to the N-terminus and the first ARM repeat of plakoglobin or β-catenin and to actin. In addition to binding directly to the actin cytoskeleton, α-catenin can also mediate attachment to microfilaments by binding to actin-associated

proteins, such as α-actinin, vinculin, or ZO-1.

Disruption of the binding of β-catenin/plakoglobin to either cadherin or α-catenin regulates changes in the strength of cell–cell adhesion and has been suggested as playing a physiological role in cell functions requiring dynamic changes in cell–cell adhesion, such as the migration of leukocytes through the endothelium lining the blood vessel wall. In addition, disruption of cadherin attachment to the actin cytoskeleton has been implicated in disease states such as cancer. Tumor cells deficient in α-catenin, but expressing E-cadherin and β-catenin, have a high rate of proliferation and display a highly invasive phenotype. Expression of α-catenin in these cells will restore the attachment of cadherins to the actin cytoskeleton and reverse the proliferative and invasive phenotype. Attachment to the actin cytoskeleton also is important for the contraction of epithelial cell sheets, a process that is required for the formation of various structures during development. Indeed, mutations in α-catenin result in trophoblast malformation and inhibition of development at the blastula stage due to loss of intercellular adhesion.

In addition to performing structural and regulatory roles at the AJ, catenins have also been implicated in transcriptional regulation of gene expression. A number of catenins including p120 family members, plakoglobin, and β-catenin have been found in the nucleus. Although the role of many of these catenins in the nucleus is not fully understood, transcriptional regulation by β-catenin has been shown to be part of the TCF/LEF signaling pathway. The majority of the β-catenin in a cell is found in the AJ and any free β-catenin is usually rapidly degraded through a complex series of reactions that lead to the ubiquitin-proteasome pathway. However, activation of the Wnt signaling pathway results in the stabilization of cytoplasmic β-catenin that then translocates into the nucleus. Once in the nucleus, β-catenin complexes with LEF/TCF transcription factors, and activates the LEF/TCF target genes that encode proteins involved in cell proliferation, apoptosis, and remodeling of the extracellular matrix. Increases in β-catenin nuclear signaling, owing to a loss of AJs coupled with activation of the Wnt pathway, have been implicated in cancer progression. Although a similar pathway has not been elucidated for other catenins, many have been found to bind to transcription factors, suggesting that they may also participate in transcriptional activation and further support the association between the AJ and regulation of gene expression.

2.3.3 Regulation of Cadherin–Catenin Complex

The phosphorylation of serine, threonine, or tyrosine residues on β-catenin, plakoglobin, and the CBD of cadherin has also been implicated in regulating the association of cadherins with the actin cytoskeleton. A number of studies have found that receptor tyrosine phosphatases such as PTPμ are part of the cadherin–catenin complex, possibly binding directly to cadherins. Activation of Src kinases, interaction with receptor tyrosine kinases, such as Erb-2 and EGF receptors, or the addition of phosphatase inhibitors, all result in an increase in the phosphorylation of β-catenin or plakoglobin. Numerous studies have found that increases in the phosphorylation of β-catenin, plakoglobin and cadherin are associated with decreases in cell–cell adhesion and a disruption of cadherin β-catenin α-catenin complex. However, these studies

did not examine the effect of phosphorylation of specific amino acid residues on the disassembly of the cadherin–catenin complex. More recent investigations have begun to examine the role of individual amino acids of β-catenin. Mutation of tyrosine 654 in β-catenin was found to prevent the dissociation of β-catenin E-cadherin complexes triggered by exposure to the phosphatase inhibitor sodium orthovanadate, thereby demonstrating a cause-and-effect relationship between phosphorylation of this residue and cadherin binding. Similarly, phosphorylation of β-catenin on Tyr142 was found to decrease its association with α-catenin. Interestingly, this tyrosine residue is phosphorylated by Fer or Fyn tyrosine kinases that can both bind to p120. Thus, p120 may regulate β-catenin α-catenin association by recruiting these kinases to cadherin complexes. The further study of recombinant cadherins and catenins with mutations that target individual tyrosine, serine, and theosine residues will provide important insights into how phosphorylation regulates cell–cell adhesion at the AJ.

Rho, Rac, and Cdc42 are members of the Rho GTPase family that regulate changes in the actin cytoskeleton and play a major role in cell processes that require actin reorganization. Given the close association of cadherin function with attachment to the actin cytoskeleton, it is not surprising that the small GTPases Rac, Rho, and Cdc42 have all been implicated in regulating cadherin-mediated adhesion. Use of dominant negative or constitutively active forms of these small GTPases has shown that each can influence the state of cadherin-mediated adhesion, with Rho and Rac being required for stable localization of cadherins at cell–cell contacts in epithelial cells. Recent investigations have also begun to elucidate the role of Rho GTPases in signaling pathways initiated by cadherin binding. These studies have found that Rac is activated upon engagement of cadherins and is recruited to nascent cadherin-mediated adhesion sites. Rac is eventually lost from AJs as they mature, indicating that although Rac is activated upon contact formation, it is not continually activated by cadherins. Rac may increase the strength of cadherin-mediated cell–cell adhesion via its influence on IQGAP. IQGAP is localized to the cell–cell junction by binding to β-catenin at the same site as α-catenin, suggesting that binding of IQGAP prevents attachment to the actin cytoskeleton. Activation of Rac disrupts the binding of IQGAP to β-catenin resulting in an increase in α-catenin β-catenin association and an increase in cadherin attachment to the actin cytoskeleton. In addition, Rac activation also results in the localization of the Arp 2/3 complex at cadherin-mediated cell–cell junctions. The Arp 2/3 complex initiates actin assembly and may be required for actin polymerization that results in the expansion of cell–cell contacts.

The function of Rho has been more elusive. Studies have reported that Rho is activated upon cadherin binding while others have found inhibition of activity with cadherin binding. This discrepancy may be due to the different cell types used in these studies, as the effect of small GTPases on cadherin-mediated junctions is dependent on cellular context. For example, VE-cadherin does not require Rac or Rho activity to localize to the cell–cell junction in endothelial cells. However, when VE-cadherin is expressed in CHO cells, localization of VE-cadherin to cell–cell junctions is Rho-dependent. This study underscores the difference between endothelial cell and epithelial cell junctions and shows the importance

of choosing the appropriate cell type for investigating specific cell functions involving the regulation of cadherin-mediated adhesion.

2.3.4 Other Components of Adherens Junctions: Nectin and Afadin

In addition to the cadherin–catenin complex, the AJ of many cell types including epithelial cells, neurons and sertoli cells contain a nectin/afadin complex that also anchors to the actin cytoskeleton. Nectins are a family of transmembrane proteins that are highly homologous to human poliovirus receptor. There are 4 isoforms of nectin with nectin 1, 2, and 3 having two to three splice variants. The extracellular domain of nectin contains three Ig-like domains. There is a single pass transmembrane region coupled to a carboxyterminal cytoplasmic tail that, in most nectins, binds to the PDZ domain of afadin. Similar to cadherins, nectins form cis-dimers in the lateral plane of the membrane prior to forming trans-dimers. Cis-dimerization is regulated by the second Ig-like domain in the extracellular region, while the first Ig-like domain is responsible for trans-dimer formation. In contrast to cadherins, trans-dimerization in nectins is calcium independent; however, nectins can form both homo- and hetero-trans dimers.

Afadin is a cytoplasmic protein that binds to the C-terminus of nectin. There are two isoforms of afadin – "s" and "l." L-afadin contains a number of domains, including a Ras association domain, forkhead association domain, DIL domain, PDZ domain, three proline-rich regions, and an actin binding domain at the C-terminus. The "s" isoform contains these same domains; however, the C-terminus is deleted removing the third proline-rich domain and the actin binding region. L-afadin can bind to the side of actin filaments, but not to the ends, thereby linking nectin to the actin cytoskeleton. Afadin has been found to play an important role in development, as knocking out this gene results in embryonic lethalality at day 10.5. These embryos show defects at gastrulation including disorganization of the ectoderm, impaired migration of the mesoderm, and a loss of somites. Interestingly, in afadin−/− embryos the E-cadherin localization was disrupted in the ectodermal cell layer, but was localized to cell–cell junctions in the endodermal layer. This suggests that the nectina fadin complex may be required for the formation of cadherin-mediated junctions in certain cell types. This is supported by studies showing that the formation of nectin-mediated junctions will recruit E-cadherin, and the formation of cadherin-mediated junctions will recruit nectins. The recruitment of E-cadherin by nectin or nectin by E-cadherin requires afadin and α-catenin. Although afadin and α-catenin do bind to one another, the interaction is weak, leading investigators to believe that an intermediate protein may be required for the interaction of E-cadherin and nectin in the AJ.

2.4 Desmosomal Cadherins and Desmosomes

Desmosomes are spot weldlike structures that provide strong cell–cell adhesion to tissues that require adhesive strength to maintain organ function. When observed by electron microscopy, desmosomes have a very distinctive ultrastructure consisting of two electron-dense plaques surrounding a lighter central core (Fig. 12a). The central core region contains the plasma membranes and the extracellular domains of the desmosomal cadherins from each of the two connected cells. Each electron-dense band is actually a tripartite structure

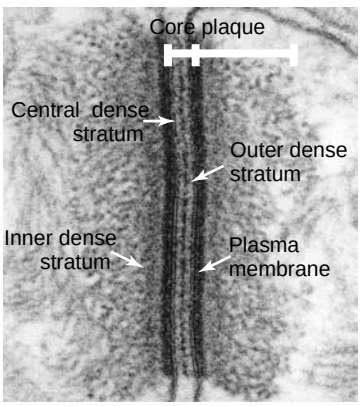

(a)

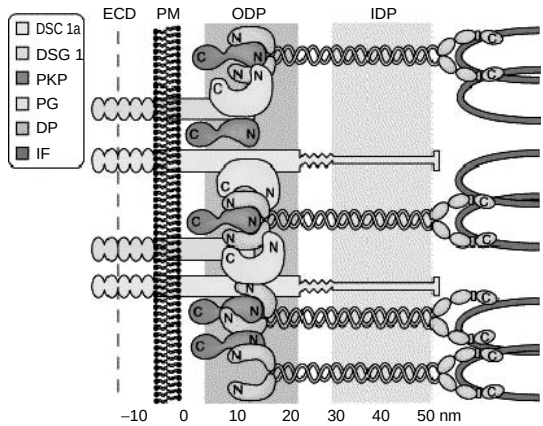

(b)

Fig. 12 Ultrastructure of desmosomal junction. (a) Electron micrograph showing the ultrastructure of a desmosome. The core region is the intercellular space between the plasma membranes of adjacent cells. The plaque regions contain proteins whose interactions serve to anchor the desmosome to intermediate filaments. The electron micrograph was kindly provided by Andrew Kowalczyk. (b) Schematic of desmosomal components. The cytoplasmic tail of desmocollins (Dsc) and desmoglein (Dsg) bind to plakoglobin (PG), plakophilin (Pkp), and the amino-terminus of desmoplakin (Dsp) in the outer dense plaque (ODP). The C-terminus of Dsp binds to the intermediate filaments. (Reprinted with permission from Leung, C.L., Green, K.J., Liem, R.K.H. (2002) Plakins: a family of versatile cytolinker proteins, *Trends Cell Biol.* **12**, 37–45).

with an outer dense plaque located next to the plasma membrane, a translucent middle region, and an inner dense plaque bordering the cytoplasm. Intermediate filaments converge on the desmosome, forming loops that pass into the inner dense plaque. The dense plaques are sites containing a high density of proteins that serve to attach the desmosomal cadherins to the intermediate filaments.

Desmosomes contain cadherins from two families – the desmocollins (Dsc) and the desmogleins (Dsg), with Dsc having three isoforms and Dsg having four isoforms. Dsg2 and Dsc2 are the most widely distributed of the desmosomal cadherins found in simple epithelia, the basal layer of epidermis, and in myocardium, all of which are tissues that require cell adhesive strength for a highly resilient junction. Indeed, in the myocardium these cadherins are localized to the intercalated discs that connect adjacent cardiac myocytes to one another. Dsg3 and Dsc3 are localized to the basal and suprabasal layers of epidermis and stratifying epithelia. Dsg1 and Dsc1 have the most limited distribution and are found in the highly differentiated upper layers of the epidermis. The importance of desmosomes to the maintenance of skin integrity is demonstrated by *Pemphigus foliaceus* and *Pemphigus vulgaris*, two autoimmune blistering diseases that result from the generation of antibodies to Dsg1 or to Dsg3 respectively.

2.4.1 Extracellular Domain of Desmoglein and Desmocollin

The structural organization of the extracellular domain of desmosomal cadherins

is similar to that of type I/II cadherins, although major differences exist in the sequences of the five cadherin repeats. The EC1 domain has the well-conserved tryptophan as its second amino acid and the surrounding AA sequence (10 mer) for each family is well conserved. The desmosomal cadherins also posses a CAR site with a central alanine. The surrounding AA differ from those in classical cadherins, but again are conserved within the desmogleins (Dsg1 R-A-L, Dsg2 Y-A-L, Dsg3 R-A-L) and the desmocollins (Dsc1a Y-A-T, Dsc2a FAT, Dsc3a Y-A-S). The CAR site for Dsg/Dsc has been shown to play a role in the trans-adhesion of the desmosomal cadherins; however, as with the type I and II cadherins, there is confusion on whether the desmosomal cadherins bind in a strict homophilic or heterophilic manner.

To determine the manner by which these two cadherin families mediate cell–cell adhesion, Dsg or Dsc were expressed in fibroblasts that do not express endogenous cadherins. These early studies demonstrated that the expression of a single desmosomal cadherin would not support cell–cell adhesion, but the expression of certain Dsg/Dsc pairs resulted in the formation of cell aggregates. These studies also demonstrated that in addition to requiring specific Dsg/Dsc pairs, specific cytoplasmic proteins must also be expressed in order to achieve strong cell–cell adhesion. Interestingly, the addition of CAR peptides homologous to either Dsg or Dsc of the Dsg/Dsc pairs was able to prevent cell–cell adhesion suggesting that Dsg/Dsc bound in a heterophilic manner. Coimmunoprecipitation of Dsc and Dsg further supported the concept that these cadherins bind in a heterophilic manner. However, more recent studies have indicated that CAR peptides homologous to both the Dsg and the Dsc are required in order to inhibit epithelial morphogenesis in tissues expressing Dsg2 and Dsc2, or Dsg3 and Dsc3, suggesting that homophilic interactions are involved in the *in vivo* function of desmosome cadherins during epithelial morphogenesis. Thus, current evidence suggests that a combination of heterophilic and homophilic interactions may participate in cell–cell adhesion at desmosomes.

2.4.2 Cytoplasmic Tail, Plaque Proteins, and Attachment to the Cytoskeleton

The cytoplasmic tails of Dsg and Dsc exhibit distinct structural features. The cytoplasmic tail of Dsc is much shorter than that of Dsg and has two forms as a result of alternative splicing. The "a" form has a CBD that binds to plakoglobin but not to β-catenin. The CBD is replaced in the "b" form of the cytoplasmic tail with an insertion of 11 amino acids not found in the "a" form resulting in the inability of the "b" form to bind to plakoglobin. Although Dsc-b does not bind plakoglobin, this desmosomal cadherin can still link to the cytoskeleton, as it can bind to desmoplakin (DP) and plakophilin-3, two proteins that also link to intermediate filaments. As in Dsc, the cytoplasmic domain of Dsg contains a CBD that binds to plakoglobin but not to β-catenin. This is followed by a proline-rich linking region, a repeated unit domain (RUD), and a terminal glycine rich domain (DTD). The cytoplasmic tail of Dsg does not bind directly to DP. Interestingly, the cytoplasmic regions of Dsc and Dsg behave differently when expressed as recombinant proteins lacking the extracellular domain. Indeed, deletion mutants containing only the cytoplasmic domain of Dsc-a can recruit DP and intermediate filaments to the plasma

membrane, whereas the Dsg cytoplasmic tail acts as a dominant negative that prevents the formation of desmosomes. The dominant negative activity of the Dsg cytoplasmic tail can be reversed by mutating the CBD. The mechanism of this response is not known but the indications are that the interaction of Dsg/Dsc with intermediate filaments is more complex than just direct structural connections through plaque proteins.

As described previously, plakoglobin is a member of the armadillo family of proteins, which links type I and type II cadherins to α-catenin, thereby supporting attachment to the actin cytoskeleton. In desmosomes, plakoglobin binds to the CBD of Dsc and Dsg. Interestingly, the binding site for Dsc/Dsg on plakoglobin overlaps with the binding site for α-catenin (the first ARM repeat); thus, the binding of plakoglobin prevents the association of α-catenin with the desmosomal cadherins. Instead of binding to α-catenin, plakoglobin binds directly to DP, and appears to play an important role in linking the desmosomal cadherins to DP and the intermediate filament cytoskeleton. The C-terminus of plakoglobin has also been found to play a role in regulating desmosome structure, as mutation of this region results in desmosome fusion and the formation of desmosomes that cover a large area of the plasma membrane. The importance of plakoglobin to the function of desmosomes *in vivo* has been demonstrated by gene ablation experiments in mice. Plakoglobin-null mice exhibit early embryonic lethality due to mechanical defects in the heart that render the organ unable to withstand the hemodynamic forces generated by cardiac muscle contraction. At the ultrastructural level, there is a loss of intercalated disc integrity, and the normal organization of desmosomal structures is disrupted. Interestingly, the plakoglobin-null phenotype is not fully penetrant, and survivors exhibit skin defects (fragility and blistering) even though desmosomes in most tissues appear morphologically normal.

In addition to plakoglobin, a second family of armadillo proteins is found in desmosomes. The plakophilins (PKP) are ARM proteins belonging to the p120 subfamily. In humans, three distinct plakophilin isoforms arise from three different genes. The first human genetic disorder to be attributed to a desmosomal gene was a mutation in plakophilin-1. Plakophilin-1 mutations lead to a skin fragility and dysplasia syndrome, resulting in epidermal blistering. The plakophilins exhibit a tissue- and differentiation-specific expression pattern and, depending on the tissue, the plakophilins can localize to both desmosomes and to the nucleus. In desmosomes, PKP-1 is found in the upper, most differentiated layers of stratified epithelia, and PKP-2 is found in the basal cell layers of stratified epithelia. PKP-2 is also found in some single-layered epithelia. Recently, plakophilin 3 has been implicated as playing a central role in desmosomal structure owing to its multiple binding partners and its widespread distribution in different epithelial cells. PKP-3 has been found to bind to all three Dsgs and to Dsc3a and 3b, as well as to Dsc1a and Dsc2a. In addition, PKP-3 was found to bind to plakoglobin, DP, and epithelial keratin-18. Interestingly, this study was unable to find PKP-3 in the nucleus, suggesting that it may play a more structural role than PKP-1 and -2. In contrast, plakophilin-1 and -2 exhibit strong nuclear localization, and plakophilin-2 has been shown to interact with the RNA polymerase III. Furthermore, PKP-2 has recently been found to participate in transcriptional regulation in

conjunction with β-catenin. These findings suggest that plakophilin-1 and -2, much like β-catenin and p120, may play dual roles in the cell: a structural role at the desmosome and a transcriptional role in the nucleus.

Desmoplakin (DP) is the protein that plays the central role in linking the desmosomal cadherin complex to intermediate filament networks. There are two closely related isoforms of DP resulting from alternative splicing, DP1-322 kD and DP2-259 kD. Both forms are found in all tissues except in the heart where only DP1 is located. DP is a dumbbell-shaped molecule that possesses three domains, an N-terminal globular domain, a coiled-coiled rod domain, and a carboxyterminal globular domain, and bind to intermediate filaments. DP is part of a large family of cytoskeletal linking proteins termed the plakin family of cytolinkers. The founding members of this family include DP, the hemidesmosomal protein BPAG-1 (BP230) and plectin, a cytoskeletal cross-linking protein with a wide tissue distribution. This family of proteins is defined by an amino-terminal plakin domain, which appears to contain protein interaction motifs that target individual plakin family members to specific subcellular destinations. In the case of DP, the N-terminal region of this protein has been shown to bind to plakoglobin, plakophilin, and Dsg/Dsc. The coiled-coiled rod domain of DP is believed to be responsible for dimerization (possibly multimerization). The length of this region is variable; DP-1 contains 599 AAs more than DP-2. The last region contains three globular plakin domains that are responsible for binding to intermediate filaments.

As shown in Fig. 12(b), DP plays a central role in the desmosome as it is the primary protein that connects to the intermediate filaments. Gene knockout studies have reinforced the importance of DP to desmosome function. Initial gene knockout studies found that loss of DP was lethal at day E5.5 to 6.5 due to a loss of DP in extra-embryonic tissue surrounding the embryo. The function of the extra-embryonic tissue was rescued using an approach that allows the extra-embryonic tissue to have a wild-type genotype, while the developing embryonic tissue contains the DP knockout. This approach allowed development until embryonic day 10. As expected, defects were found in the heart and the skin, organs known to require strong cell–cell adhesion to maintain tissue integrity. The neuroepithelium, found in the head folds and in the lining of the spinal cord, was also affected by the loss of DP. Prior to neural tube closure there is normally an increase in desmosomes followed by a loss of these junctions once the neural tube is formed. In DP−/− embryos, the neural tube was collapsed and the neuroepithelium was disorganized. Interestingly, the loss of DP did not affect differentiation of the neuroepithelium but prevented the stabilization and/or completion of neural tube formation.

The DP knockout animal also displayed fewer capillaries than wild-type mice. Interestingly, endothelial cells do not contain desmosomes. Instead, plakoglobin and DP bind to VE-cadherin (cadherin-5) forming a structure called the *complexus adherens junction* or *syndesmosome*. This junction has been observed in the microcirculation *in vivo*, but does not appear to be present in large-vessel endothelium. In DP−/− embryos, only the capillaries appeared to be disrupted, while the large vessels appeared to develop normally. In addition, the DP−/− phenotype resembles the VE-cadherin phenotype (−/−); both

exhibit lethality at embryonic day 9.5 due impaired remodeling and maturation of blood vessels. *In vitro* studies also support the roles of plakoglobin and DP in attaching VE-cadherin to intermediate filaments, as expression of these three components of the complexus AJ will induce the formation of cell–cell junctions in fibroblasts. One should note that, to date, VE-cadherin is the only cadherin from the type I and type II families to form attachments to intermediate filaments and the actin cytoskeleton. Thus, VE-cadherin is a very unique member of the cadherin superfamily.

2.4.3 Functions of Desmosomes

The previous discussion highlights the importance of desmosomes in the maintenance of tissue integrity. However, recent studies suggest that desmosomes may mediate other functions in addition to providing strong cell–cell adhesion. For example, desmosomal adhesion has been implicated in morphogenic cell positioning in the mammary gland. *In vivo*, the luminal epithelial cells of the mammary gland possess desmosomal adhesions mediated by Dsg2/Dsc2, while the myoepithelial cells surrounding the luminal cells have desmosomes containing Dsg3/Dsc3 in addition to Dsg2/Dsc2. Isolating these cells from tissue and allowing them to recombine in culture results in the formation of aggregates in which the Dsg3/Dsc3 myoepithelial cells surround the luminal epithelial cells. The addition of CAR peptides homologous to Dsg3/Dsc3 inhibited the correct association of these cells, suggesting that Dsg3/Dsc3-mediated adhesion is required for the sorting of mammary epithelial cells into appropriate cell layers.

In addition to regulating morphogenesis, desmosomes may also play a role in regulating differentiation. Recent studies demonstrated that the ratio of Dsg1 to Dsg3 expression varies in different epithelial tissues and that this ratio is important in regulating the differentiation of these tissues. For example, Dsg1 is expressed in all layers of the stratified epithelial layers of the skin, whereas Dsg3 is expressed only in the suprabasal layer. In contrast, Dsg1 and Dsg3 are expressed in all cell layers of the stratified epithelia of the oral mucosa. Using a transgenic approach, Dsg3 was expressed in all epithelial cell layers of the mouse skin producing a Dsg1 : Dsg3 ratio similar to that in the oral mucosa. Interestingly, the stratum corneum, the outermost layer of the skin of these transgenic mice, now had a histological appearance similar to that of the oral mucosa. Furthermore, the skin of these mice had excessive transepidermal water loss that resulted in death due to dehydration, indicating that the Dsg3/Dsg1 ratio is also important in the regulation of barrier function.

2.5 Tight Junctions

Tight junctions form a meshwork of interconnected strands grouped in a narrow belt at the apical region of epithelial or endothelial cell–cell adhesion. These structures have been studied primarily in epithelial cells whose monolayers form vital cell barriers that tightly regulate the movement of fluid and molecules between different tissue compartments. In addition to regulating paracellular transport, TJs separate the apical plasma membrane from the basal plasma membrane, and thus act as a fence to maintain the different lipid and protein compositions of these two cell surfaces. At the ultrastructural level, the plasma membranes of two

adjacent cells come into such close proximity that the outer leaflets of the adjacent plasma membranes appear to be fused. The anastomosing filaments of tTJs are best observed using freeze fracture as shown in Fig. 13(a). Freeze fracture is a platinum carbon replica of a fractured membrane surface that provides a three-dimensional relief of the junction when viewed by scanning electron microscopy. This technique creates an image of either strands of beads on the protoplasmic face or grooves on the exoplasmic face.

Early hypotheses speculated that the tight junction may be formed by specialized lipid arrangements between the two plasma membranes of adjacent cells; however, more recent studies have shown that transmembrane proteins are responsible for the ultrastructure of the tight junction (Fig. 13c). In a manner similar to AJs and desmosomes, the transmembrane proteins of TJs attach to the cytoskeleton through linker proteins and contain scaffolding proteins that allow TJs to participate in a number of signal transduction pathways.

2.5.1 Claudins and Other TJ Transmembrane Proteins

Initial studies using transepithelial electrical resistance (TER) to assess the barrier function of an epithelial monolayer suggested a relationship between the number of tight junction strands and TER. However, this concept was challenged by studies using Madin-Darby canine kidney (MDCK) epithelial cells, a cell line that is used extensively for *in vitro* studies of the epithelial barrier as they form functionally intact TJs. Interestingly, two strains of MDCK cells had been identified that had vastly different TERs (MDCK I $\sim$3000 Ωcm^2 and MDCK II $\sim$100 Ωcm^2), but possessed the same number of tight junction strands with the same morphology. This suggested that in addition to the number of tight junction strands, the

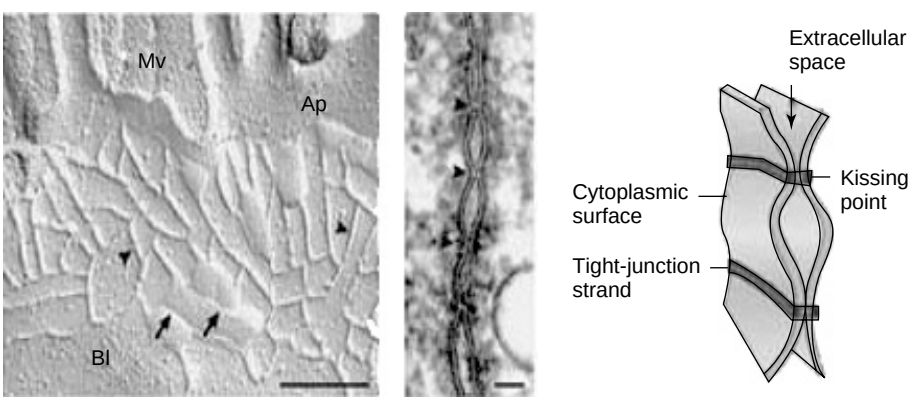

(a) (b) (c)

Fig. 13 Ultrastructure of tight junctions. (a) Electron microscopic image showing freeze-fracture replica of the tight junctions in intestinal epithelial cells. The interconnected strands appear as fibers on the P-face (arrow heads) and as grooves on the E-face (arrows). (b) Electron micrograph showing the close association of the plasma membranes in tight junction strands. Note the dense plaque region associated with the junction. (c) Schematic showing the "kissing point" produced by protein interactions on adjacent cells; this interaction produces the close association of membranes shown by arrow heads in (b).

protein composition of the tight junction strand may control the selectivity of the epithelial barrier. In 1998, two novel integral membrane proteins, claudin 1 and claudin 2, were identified from TJ enriched fractions. When expressed in fibroblasts, these proteins were able to produce TJ strands that were morphologically similar to those of epithelial cells. Since that time, over 20 claudin isoforms have been identified. Claudins are 20 to 27 kD proteins with 4 transmembrane domains that form two extracellular loops and whose N- and C-termini are localized on the cytoplasmic side of the plasma membrane. The C-terminal tail is long and is divergent between isoforms suggesting that anchorage to the cytoskeleton or scaffolding to signaling molecules may also be important in regulating claudin function.

Claudins were found to play an important role in determining the TER in MDCK1 and MDCK2 cells. Expression of claudin 1 in MDCK cells was found to increase the TER 4 to 5-fold over wild-type MDCK. Expression of claudin 2 in highly resistant MDCK1 cells, which normally express claudins 1 and 4, resulted in a decrease in TER and tight junction morphology similar to low-resistive MDCK2 that typically express claudin 2 in addition to claudin 1 and 4. This study demonstrated that strands containing different claudin isoforms are responsible for the variable barrier properties of different tight junction strands. Claudins have also been found to regulate ion selectivity. Indeed, expression of claudin 4 in MDCK will increase TER by decreasing Na^+ permeability without changing the permeability to Cl^-, the flux of mannitol, or the rank order of permeabilities of alkali metal cations. These studies all demonstrate that claudins are the proteins in the tight junction strand that are responsible for regulating the selective permeability of epithelial and endothelial cells.

How do different claudin isoforms create barriers with different selectivity? Although the mechanisms involved are not fully understood, current evidence supports the hypothesis that the specific structural characteristics of different claudin isoforms play a role in regulating the selectivity and barrier properties of TJs. Studies using expression of different claudin isoforms both singularly as well as in pairs have found that claudins can form mixed strands containing more than one isoform in each strand. In addition, claudins have been shown to bind in both a homophilic and a heterophilic manner. Heterotypic interactions only occur with specific pairs of claudin isoforms. For example, cells expressing claudin 1 or claudin 3 form tight junction strands when placed in mixed culture, whereas cells expressing claudin 2 will not form tight junction strands when mixed with either claudin 1-or claudin 3-expressing cells. It has been proposed that mixed strands containing claudin 1 and 3 would form a very restrictive barrier as these two isoforms bind to each other in a homotypic and heterotypic manner as shown (Fig. 14b). In contrast, a strand containing claudin 1 and claudin 2 would form a less restrictive barrier. In this strand, claudin 1 would bind to claudin 1 and claudin 2 would bind to claudin 2 in a homotypic manner, creating a barrier. However, at certain points in the strand, claudin 1 would oppose claudin 2 forming a pore that would allow the transport of molecules through the strand (Fig. 14b). As over 20 claudin isoforms have been identified, a number of different pairings could exist, allowing for different barrier properties in specific tissues.

Since the discovery of claudin, a number of inherited human diseases have been

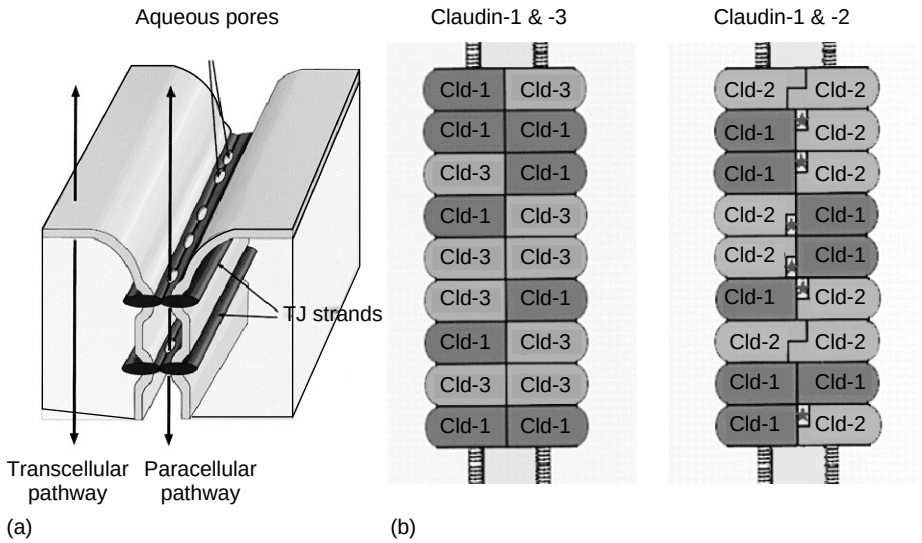

Fig. 14 Schematic of proposed model for paracellular transport through tight junction. A(a) The claudins are believed to form a strand containing aqueous pores that acts as a selective barrier to the flux of molecules between cells termed paracellular transport. (Bb) Schematic showing how claudins may interact to form a selective barrier. (See text for details). (Reprinted with permission from Tsukita, S., Furuse, M. (2000) Pores in the wall: Claudins constitute tight junction strands containing aqueous pores, *J. Cell Biol.* **149**, 13–16; Tsukita S., Furuse, M. 2000, *J. Cell Biol.*, **149**, 13–15).

identified for which mutations in specific claudin isoforms were demonstrated to be the cause of the underlying pathophysiology of the disease. Indeed, mutations in claudin 16 (paracellin-1) were identified in patients with renal hypomagnesemia, a rare disease characterized by a massive urinary loss of magnesium. Claudin 16 is expressed in the thick ascending limb in the kidney, a primary site of magnesium uptake. The current thought is that claudin 16 creates a channel that is selective for magnesium. This channel would be absent in patients with claudin 16 mutations; thus, magnesium would remain in the tubule and be lost in the urine. Mutations in claudin 14 have been shown to be responsible for nonsyndromic recessive deafness. Claudin 14 is located in the sensory epithelium in the organ of Corti, although its exact function in the inner ear is not completely understood. Knockout animals also provide evidence that claudins are responsible for the formation of highly selective and tissue-specific channels in tight junction strands. Claudin 5 is a major tight junction protein in endothelial cells with the greatest expression being found in the brain vasculature. Claudin 5-null mice were found to have normal vascular development and showed no change in the vascular barrier as assessed by the lack of edema during development. However, the transport of small molecules (<800 Da) was selectively increased in endothelial cells of the blood-brain barrier resulting in death 10 h after birth. Interestingly, mice lacking claudin 1 die shortly after birth due to dehydration that is the result of water loss across the epidermal

barrier of the skin. Importantly, removal of one claudin isoform does not result in the complete loss of tight junction strands in the affected tissue, thereby demonstrating that it is the combined function of different claudin isoforms that creates the selectivity of the tight junction.

Two other transmembrane components have been found to play a major role in the formation and regulation of TJs, occludin and junctional adhesion molecule (JAM). Occludin was the first protein to be isolated from tight junction enriched fractions. Immunofluorescent and immunoelectron microscopic analysis showed that this protein was localized to the tight junction strand. Like claudins, occludins contain four transmembrane regions with two extracellular loops, and their amino and carboxyterminal tails are located in the cytoplasm. Initial studies found that overexpression of occludin in MDCK cells would increase the number of tight junction strands and elevate TER. However, by generating occludin-deficient embryonic stem cells and occludin-knockout mice, it was shown that functional tight junction strands could be formed in the absence of occludin. Histological examination of certain tissues did show some abnormalities such as chronic inflammation and hyperplasia of the gastric epithelium, suggesting that occludin plays a regulatory role rather than establishing the barrier itself. This is supported by studies that have correlated changes in occludin phosphorylation with changes in barrier function. JAM is a single membrane-spanning protein with two IgG-like domains in its extracellular region that mediate homotypic binding. JAM is not integrated into the tight junction strand; however, studies have shown that JAM is involved in tight junction formation and barrier function possibly through an interaction with occludin.

2.5.2 Plaque Proteins of Tight Junctions

Like desmosomes and AJs, TJs also have a cytoplasmic plaque region containing proteins that serve both structural and signaling roles (Table 7). The plaque region of the tight junction is made up of a large number of proteins containing PDZ domains, named from the first three proteins first identified within this domain [PSD-95-postsynaptic density protein 95; Dlg-discs large proteinan; Zonula (ZO) occludins-1]. The PDZ domain mediates interaction with other proteins, including other PDZ proteins or transmembrane proteins containing the S/T-X-V in their C-terminus. As seen in Table 7, the PDZ containing proteins found in the tight junction can be divided into two subsets based on whether the protein contains a guanylate kinase homology region. Proteins containing the membrane-associated

Tab. 7 Tight junction plaque proteins.

Type	*Protein name*
PDZ (MAGUK)	ZO-1
	ZO-2
	ZO-3
	MAGI/BAP1
PDS (non-MAGUK)	PAR-3 (ASIP)
	PAR-6
	AF-6
Non-PDZ	Cingulin
	Symplekin
	ZONAB
	ASH1
	4.1R
Cytoskeletal	Actin
	Spectrin
Signaling	RAB13, RAB3B
	Atypical protein kinase C (ζ, λ)
	ZAK
Other	Sec6/8
	7H6
	BG9.1
	19B1

guanylate kinase (MAGUK) also contain three to five PDZ domains and an SH3 domain. Of these proteins, the most closely studied group is the zonula occluding proteins (ZO-1, ZO-2, ZO-3). The zonula occluding proteins all bind to the claudins via their PDZ domains and can attach to actin filaments, thereby anchoring the tight junction to the actin cytoskeleton. ZO-1 and ZO-2 have also been found in AJs, presumably due to their ability to bind to α-catenin; however, their presence in the AJ may only occur in the absence of TJs. In addition, these proteins serve a scaffolding, role localizing transcription factors and signaling proteins to the tight junction plaque. For example, ZONAB (ZO-1-associated nucleic acid binding protein) is a transcription factor that acts as a repressor of Erb-2 promoter. ZONAB function is regulated, in part, by the sequestration of ZONAB at TJs through its binding to ZO-1.

Attachment to the actin cytoskeleton has been implicated as a possible control point for the regulation of tight junction function. A number of proteins located in the tight junction plaque can directly link claudin to the actin cytoskeleton as described for ZO proteins. Interestingly, occludin has been found to directly link to actin filaments. The tight junction may also serve as a site regulating cytoskeletal dynamics. Indeed, myosin II is localized to the tight junction plaque through its association with cingulin. The structure of cingulin contains a globular N-terminus linked to a coiled–coiled domain that allows for the formation of dimers. The N-terminus mediates binding to ZO-1, ZO-2, JAM, and actin. The C-terminus has been shown to bind nonmuscle myosin II. Other proteins that act to bundle actin filaments, such as cortactin, are also localized to the tight junction plaque. Although it is clear that the tight junction is associated with the actin cytoskeleton, the functional significance of this attachment is not clear. Studies have attempted to disrupt actin interaction through the expression of mutant proteins; however, these studies are hampered by the endogenous levels of these proteins and/or by the ability of other proteins to compensate for the lost protein function. A major area for future research will be to determine how the interaction of the plaque proteins with the actin cytoskeleton regulates *de novo* assembly, reorganization, and acute regulation of the paracellular permeability of the tight junction.

2.5.3 Regulation of Tight Junction Assembly/Disassembly

The close association between tight junction function and the actin cytoskeleton implicates the small GTPases (Rho, Rac, and Cdc42) in the regulation of the barrier function of tight junctions. A number of studies have used the expression of dominant negative or constitutively active forms of these proteins and found that the expression of either form decreases TER, increases paracellular soluble flux, and decreases fence function. These findings suggest that the activity of Rho GTPases needs to be tightly regulated, as either increased or decreased activity will disrupt tight junction function. Alternatively, localized activation of GTPase activity may be an important factor in regulating changes in TJ function. Local changes in activity could be controlled by guanine nuclear exchange factors (GEFs) associated with the tight junction plaque proteins. GEF-H1/Lfc has recently been found to localize in the TJ and overexpression of wild-type GEF-H1/Lfc increased

paracellular permeability to 4-kDa dextran, whereas depletion of this GEF by siRNA decreased the permeability. Interestingly, changes in the level of GEF-H1/Lfc did not change TER or the permeability of higher molecular weight molecules. In addition, no change in the morphology of the tight junction was noted, suggesting that local activation of Rho produces only subtle changes in tight junction activity. Identification of other GEFs and GAPs associated with the tight junction along with identification of signaling pathways that activate these molecules will be important for elucidating the role of small GTPases in the regulation of tight junctions.

As is the case with AJ proteins, the plaque proteins found in tight junctions are also phosphorylated on serine, threonine, and tyrosine. Occludin, ZO-1 and ZO-2 have all been found to be phosphorylated on all three residues under various experimental conditions; however, no clear association between phosphorylation of these proteins and TJ assembly and/or function has been firmly established. A number of kinases have been identified as residing in the tight junction complex including occludin-associated kinase, ZO-1-associated kinase, and several PKC isoforms. Of the PKC isoforms localized to the TJs, only the atypical PKCs (aPKCs) have been shown to play a definitive role in junction assembly. The aPKCs have been shown to form a complex with PAR3 (also known as ASIP, a PKC-specific interacting protein), PAR6, and two PDZ proteins that are found in the tight junction complex. A number of studies support a role for aPKC/PAR3/PAR6 in the formation of the tight junction. For example, expression of kinase-deficient aPKC was found to prevent the reforming of tight junction in the Ca^{++} switch assay. In addition, the expression of wild-type PAR3 has also been shown to increase the rate of tight junction formation. Phosphorylation of PAR on Ser827 by aPKC was shown to decrease aPKC binding to PAR3 and increase junction formation. PAR6 was found to complex with PAR3 and aPKC and to link this complex with Cdc42, which, as explained in the following section, may play an important role in tight junction formation.

The formation of AJs has been demonstrated to trigger the formation of TJs. Although the exact order of arrival for the components of AJs and TJs is not known, the following model was proposed. Initial contact between two epithelial cells has been shown to occur when the tip of a filopodia from one cell contacts an adjacent cell. This allows E-cadherin and nectin to form microclusters at this relatively small point of cell–cell contact. The cadherin in this cluster begins to recruit α-catenin, while the nectins recruit afadin, resulting in the association of this microcluster with the actin cytoskeleton. The junction expands as actin polymerization is stimulated through activation of Rac and possibly through recruitment of Arp 2/3 by E-cadherin and nectin transdimerization. Nectin has also been shown to activate Cdc42, which, in conjunction with Rac, will also stimulate actin polymerization, thereby increasing junction stability. In addition, Cdc42 also activates aPKC resulting in the phosphorylation of PAR3. PAR3 can associate with JAM, which is recruited to the apical side of the AJ. JAM and PAR3 and possibly ZO-1, which can associate with α-catenin, may then serve as an assembly site for the tight junction, which matures into a distinct junction complex as shown in Fig. 15. The knockdown of specific

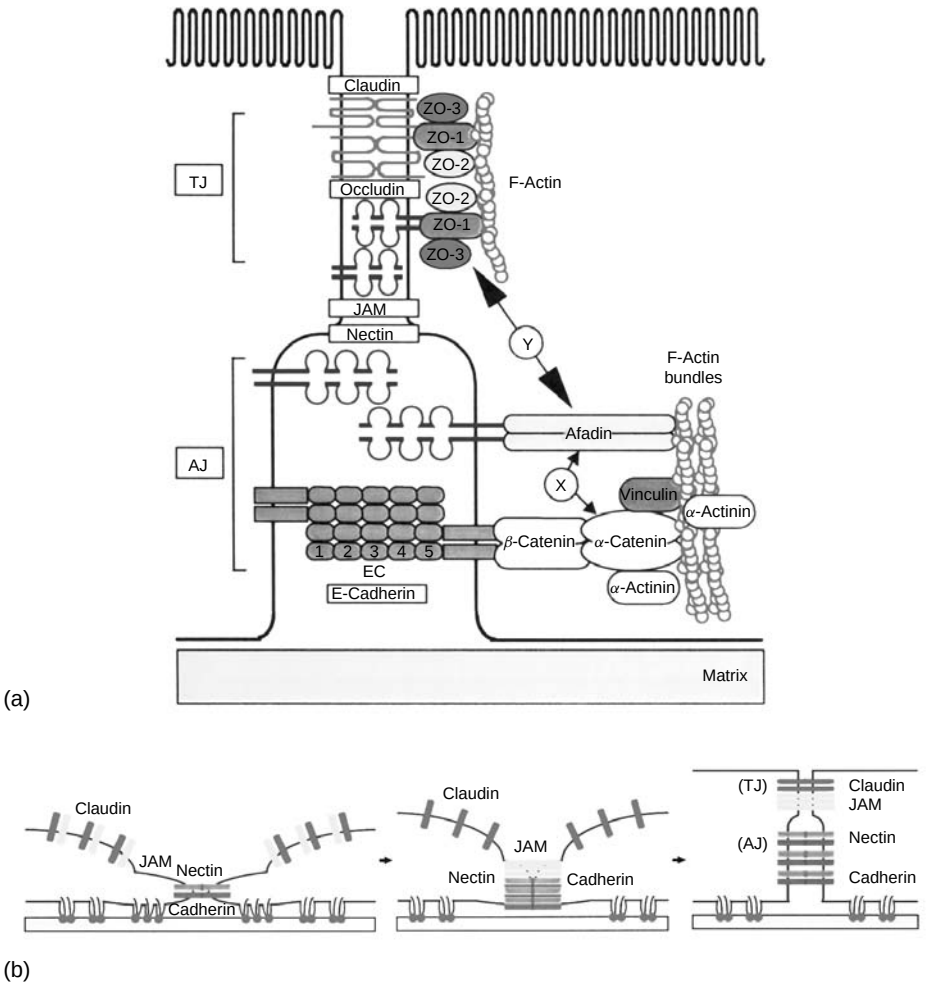

Fig. 15 Model for the formation of adherens and TJs. (a) Multiple proteins are localized to mature adherens and TJs. Some proteins such as ZO-1 can be found in both types of junctions and have been proposed to serve as intermediates allowing AJs to initiate tight junction formation. (b) Filopodia from one cell contact an adjacent cell allowing cadherins and nectins to cluster. As the cadherin/nectin junction expands JAM is incorporated, resulting in the localization of tight junction proteins and the initiation of tight junction formation. The tight junction matures into a separate structure. (Reprinted with permission from Takai, Y., Nakanishi, H. (2003) Nectin and afadin: novel organizers of intercellular junctions, *J. Cell Sci.* **116**, 17–27).

proteins in this proposed model using techniques such as siRNA (small interfering RNAs) will be critical in determining the importance of the AJ to tight junction assembly.

3 Summary and Future Directions

The interaction of cells with the ECM and withone another plays an essential

role in most if not all physiological processes. Much is already known about the mechanisms regulating cell–matrix and cell–cell adhesion. Important components of cell junctions and their signaling pathways have been identified and characterized. Current evidence suggests that there are many different junctional-binding proteins providing alternative cytoskeletal linkages and signaling pathways to mediate cell function. A challenge for the future will be to determine the significance of individual linkages and pathways in diverse physiological contexts. Additionally, most of what we know about cell junctions has been gained from two-dimensional (2D) cell culture models focusing on specific protein–protein interactions. Although these studies have provided important insights into the mechanisms regulating the functions of specific proteins, a future task will be to determine the role of these mechanisms in complex biological systems. Gene knockout studies have provided an important starting point. However, cell adhesion and signaling proteins have multiple interaction partners. Thus, more information will be gained from future "knockin" studies of these proteins with mutations that inhibit individual aspects of protein function.

Additionally, the view of cell–matrix and cell–cell adhesions as sites of mechanotransduction is growing in momentum. Multiple examples are available to demonstrate the activation of integrin signaling by changes in hemodynamic shear stress on endothelial cells. Additionally, it has been known for many years that the response of cells to their ECM environment not only depends on specific integrin–ligand integrations but also on the physical state of these ECM ligands. The rigidity of the ECM can dramatically affect cell behavior. Changes in ECM rigidity have been modeled experimentally using three-dimensional collagen matrices. Rigid collagen matrices that are attached to a culture provide resistance to cell-generated contractile forces at cell–matrix junctions, whereas nonrigid collagen matrices that float in culture medium do not. Interestingly, mammary epithelial cells differentiate into aveolar-like structures containing basement membranes, and secrete milk proteins when cultured in nonrigid three-dimensional collagen matrices. However, mammary cells fail to differentiate and continue proliferating when cultured in rigid three-dimensional matrices. In contrast, the differentiation of fibroblasts into myofibroblasts requires the increase in ECM rigidity that accompanies wound healing. In culture, fibroblasts differentiate into myofibroblasts when cultured in rigid, but not in nonrigid, three-dimensional collagen matrices. Clearly, identifying the molecular mechanisms that govern mechanotransduction at cell junctions is an important area for future investigation.

Finally, in most instances, integrins and cadherins regulate cell behavior by collaborating with other cell surface receptors. For example, the interaction of VE-cadherin with VEGF-receptor 2 modulates VEGF-induced signaling, perhaps by clustering the VEGF receptor with phosphatases in the AJ. In addition, evidence suggests that E-cadherin may cocluster and activate the EGF receptor resulting in activation of EGF signaling independent of the EGF ligand. Similar observations have been made in studies examining the collaboration between integrin and growth factor receptor signaling. Thus, the molecular mechanisms governing the cross-talk between cell junction proteins and other signaling receptors will be a continuing area of intense investigation.

See also Electron Microscopy in Cell Biology.

Bibliography

Books and Reviews

Anastasiadis, P.Z., Reynolds, A.B. (2000) The p120 catenin family: complex roles in adhesion, signaling and cancer, *J. Cell Sci.* **113**, 1319–1334.

Angst, B.D., Marcozzi, C., Magee, A.J. (2001) The cadherin superfamily: diversity in form and function, *J. Cell Sci.* **114**, 629–641.

Blaschuk, O.W., Rowlands, T.M. (2002) Plasma membrane components of adherens junctions, *Mol. Membr. Biol.* **19**, 75–80.

Brakebusch, C., Fassler, R. (2003) The integrin-actin connection, an external love affair, *EMBO J.* **22**, 2324–2333.

Brown, E.J., Frazier, W.A. (2001) Integrin associated protein (CD47) and its ligands, *Trends Cell Biol.* **11**, 130–135.

Burridge, K., Chrzanowska-Wodnicka, M., Zong, C. (1997) Focal adhesion assembly, *Trends Cell Biol.* **7**, 342–347.

Cereijido, J., Anderson, J. (Eds.) (2001) *Tight Junctions*, CRC Press, Boca Raton.

Colognato, H., Yurchenco, P.D. (2000) Form and function: the laminin family of heterotrimers, *Dev. Dyn.* **218**, 213–234.

Garrod, D.R., Merritt, A.J., Nie, Z. (2002) Desmosomal cadherins, *Curr. Opin. Cell Biol.* **14**, 537–545.

Geiger, B., Bershadsky, A. (2002) Exploring the neighborhood: adhesion-coupled cell mechanosensors, *Cell* **110**, 139–142.

Grinnell, F. (2003) Fibroblast biology in three-dimensional collagen matrices, *Trends Cell Biol.* **13**, 264–269.

Hynes, R.O. (1996) Targeted mutations in cell adhesion genes: what have we learned from them? *Dev. Biol.* **180**, 402–412.

Hynes, R.O. (2002) Integrins: bidirectional, allosteric signaling machines, *Cell* **110**, 673–687.

Leung, C.L., Green, K.J., Liem, R.K.H. (2002) Plakins: a family of versatile cytolinker proteins, *Trends Cell Biol.* **12**, 37–45.

Liu, S., Calderwood, D.A., Ginsberg, M.H. (2000) Integrin cytoplasmic domain-binding proteins, *J. Cell Sci.* **113**, 3563–3571.

Matter, K., Balda, M.S. (2003) Signalling to and from tight junctions, *Nat. Rev. Mol. Cell Biol.* **4**, 225–236.

Mercurio, A.M., Rabinovitz, I., Shaw, L.M. (2001) The $\alpha_6\beta_4$ integrin and epithelial cell migration, *Curr. Opin. Cell Biol.* **13**, 541–545.

Miranti, C.K., Brugge, J.S. (2002) Sensing the environment: a historical perspective on integrin signal transduction, *Nat. Cell Biol.* **4**, 83–90.

Nievers, M.G., Schaapveld, R.Q., Sonnenberg, A. (1999) Biology and function of hemidesmosomes, *Matrix Biol.* **18**, 5–17.

Nollet, F., Kools, P., van Roy, F. (2000) Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members, *J. Mol. Biol.* **299**, 551–572.

Prockop, D.J., Kivirikko, K.I. (1995) Collagens: molecular biology, diseases, and potentials for therapy, *Annu. Rev. Biochem.* **64**, 403–434.

Schlaepfer, D.D., Hauck, C.R., Sieg, D.J. (1999) Signaling through focal adhesion kinase, *Prog. Biophys. Mol. Biol.* **71**, 435–478.

Sheppard, D. (2000) In vivo functions of integrins: lessons from null mutations in mice, *Matrix Biol.* **19**, 203–209.

Takai, Y., Hakanishi, H. (2003) Nectin and afadin: novel organizers of intercellular junctions, *J. Cell Sci.* **116**, 17–27.

Tomasek, J.J., Gabbiani, G., Hinz, B., Chaponnier, C., Brown, R.A. (2002) Myofibroblasts and mechano-regulation of connective tissue remodeling, *Nat. Rev. Mol. Cell Biol.* **3**, 349–363.

Tsukita, S., Furuse, M. (2000) Pores in the wall: Claudins constitute tight junction strands containing aqueous pores, *J. Cell Biol.* **149**, 13–16.

Tsukita, S., Furuse, M., Itoh, M. (2001) Multifunctional strands in tight junctions, *Nat. Rev. Mol. Cell Biol.* **2**, 285–293.

van der Flier, A., Sonnenberg, A. (2001) Function and interactions of integrins, *Cell Tissue Res.* **305**, 285–298.

Wu, C., Dedhar, S. (2001) Integrin-linked kinase (ILK) and its interactors: a new paradigm for the coupling of extracellular matrix to actin cytoskeleton and signaling complexes, *J. Cell Biol.* **155**, 505–510.

Yamada, K.M., Geiger, B. (1997) Molecular interactions in cell adhesion complexes, *Curr. Opin. Cell Biol.* **9**, 76–85.

Yap, A.S., Kovacs, E.M. (2003) Direct cadherin-activated cell signaling: a view from the plasma membrane, *J. Cell Biol.* **160**, 11–16.

Zhurinsky, J., Shtutman, M., Ben-Ze'ev, A. (2000) Plakoglobin and β-catenin: protein interactions, regulation and biological roles, *J. Cell Sci.* **113**, 3127–3139.

Primary Literature

Akiyama, S.K., Yamada, S.S., Yamada, K.M., LaFlamme, S.E. (1994) Transmembrane signal transduction by integrin cytoplasmic domains expressed in single-subunit chimeras, *J. Biol. Chem.* **269**, 15961–15964.

Andra, K., Lassmann, H., Bittner, R., Shorny, S., Fassler, R., Propst, F., Wiche, G. (1997) Targeted inactivation of plectin reveals essential function in maintaining the integrity of skin, muscle, and heartácytoarchitecture, *Genes Dev.* **11**, 3143–3156.

Aono, S., Nakagawa, S., Reynolds, A.B., Takeichi, M. (1999) $p120^{ctn}$ acts as an inhibitory regulator of cadherin function in colon carcinoma cells, *J. Cell Biol.* **145**, 551–562.

Aplin, A.E., Stewart, S.A., Assoian, R.K., Juliano, R.L. (2001) Integrin-mediated adhesion regulates ERK nuclear translocation and phosphorylation of Elk-1, *J. Cell Biol.* **153**, 273–282.

Berrier, A.L., Martinez, R., Bokoch, G.M., LaFlamme, S.E. (2002) The integrin β tail is required and sufficient to regulate adhesion signaling to Rac1, *J. Cell Sci.* **115**, 4285–4291.

Boggon, T.J., Murray, J., Chappuis-Flament, S., Wong, E., Gumbiner, B.M., Shapiro, L. (2002) C-cadherin ectodomain structure and implications for cell adhesion mechanisms, *Science* **296**, 1308–1313.

Bonne, S., Gilbert, B., Hatzfeld, M., Chen, X., Green, K.J., van Roy, F. (2003) Defining desmosomal plakophilin-3 interactions, *J. Cell Biol.* **161**, 403–416.

Bottazzi, M.E., Zhu, X., Bohmer, R.M., Assoian, R.K. (1999) Regulation of $p21_{cpi1}$ expression by growth factors and the extracellular matrix reveals a role for transient ERK activity in G1 phase, *J. Cell Biol.* **146**, 1255–1264.

Braga, V.M.M., Del Maschio, A., Machesky, L., Dejana, E. (1999) Regulation of cadherin function by Rho and Rac: Modulation of junction maturation and cellular context, *Mol. Biol. Cell* **10**, 9–22.

Calderwood, D.A., Fujioka, Y., de Pereda, J.M., Garcia-Alvarez, B., Nakamoto, T., Margolis, B., McGlade, C.J., Liddington, R.C., Ginsberg, M.H. (2003) Integrin β cytoplasmic domain interactions with phosphotyrosine-binding domains: a structural prototype for diversity in integrin signaling, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2272–2277.

Calderwood, D.A., Yan, B., de Pereda, J.M., Alvarez, B.G., Fujioka, Y., Liddington, R.C., Ginsberg, M.H. (2002) The phosphotyrosine binding-like domain of talin activates integrins, *J. Biol. Chem.* **277**, 21749–21758.

Cary, L.A., Han, D.C., Polte, T.R., Hanks, S.K., Guan, J.L. (1998) Identification of p130Cas as a mediator of focal adhesion kinase-promoted cell migration, *J. Cell Biol.* **140**, 211–221.

Chang, D.D., Hoang, B.Q., Liu, J., Springer, T.A. (2002) Molecular basis for interaction between Icap1αPTB domain and β1 interin, *J. Biol. Chem.* **277**, 8140–8145.

Chang, D.D., Wong, C., Smith, H., Liu, J. (1997) ICAP-1, a novel β_1 integrin cytoplasmic domain-associated protein, binds to a conserved and functionally important NPXY sequence motif of β_1 integrin, *J. Cell Biol.* **138**, 1149–1157.

Chitaev, N.A., Troyanovsky, S.M. (1997) Direct Ca2+-dependent heterophilic interaction between desmosomal cadherins, desmoglein and desmocollin, contributes to cell-cell adhesion, *J. Cell Biol.* **138**, 193–201.

Colegio, O.R., Itallie, C.V., Rahner, C., Anderson, J.M. (2003) Claudin extracellular domains determine paracellular charge selectivity and resistance but not tight junction fibril architecture, *Am. J. Physiol. Cell. Physiol.* **284**, C1346–C1354.

Cukierman, E., Pankov, R., Stevens, D.R., Yamada, K.M. (2001) Taking cell-matrix adhesions to the third dimension, *Science* **294**, 1708–1712.

Degani, S., Balzac, F., Brancaccio, M., Guazzone, S., Retta, S.F., Silengo, L., Eva, A., Tarone, G. (2002) The integrin cytoplasmic domain-associated protein ICAP-1 binds and regulates Rho family GTPases during cell spreading, *J. Cell Biol.* **156**, 377–388.

Dowling, J., Yu, Q.-C., Fuchs, E. (1996) β_4 integrin is required for hemidesmosome formation, cell adhesion and cell survival, *J. Cell Biol.* **134**, 559–572.

Elias, P.M., Matsuyoshi, N., Wu, H., Lin, C., Wang, Z.H., Brown, B.E., Stanley, J.R. (2001) Desmoglein isoform distribution affects stratum corneum structure and function, *J. Cell Biol.* **153**, 243–249.

Eliceiri, B.P., Puente, X.S., Hood, J.D., Stupack, D.G., Schlaepfer, D.D., Huang, X.Z., Sheppard, D., Cheresh, D.A. (2002) Src-medicated coupling of focal adhesion kinase to integrin $\alpha_v\beta_5$ in vascular endothelial growth factor signaling, *J. Cell Biol.* **157**, 149–159.

Fukuhara, A., Irie, K., Nakanishi, H., Takekuni, K., Kawakatsu, T., Ideda, W., Yamada, A., Katata, T., Honda, T., Sato, T., Shimizu, K., Ozaki, H., Horiuchi, H., Kita, T., Takai, Y. (2002) Involvement of nectin in the localization of junctional adhesion molecule at tight junctions, *Oncogene* **21**, 7642–7655.

Furuse, M., Furuse, K., Sasaki, H., Tsukita, S. (2001) Conversion of zonulae occludentes from tight to leaky strand type by introducing claudin-2 into Madin-Darby canine kidney I cells, *J. Cell Biol.* **153**, 263–272.

Gallicano, G.I., Bauer, C., Fuchs, E. (2001) Rescuing desmoplakin function in extra-embryonic ectoderm reveals the importance of this protein in embryonic heart, neuroepithelium, skin and vasculature, *Development* **128**, 929–941.

Green, J.M., Zhelesnyak, A., Chung, J., Lindberg, F.P., Sarfati, M., Frazier, W.A., Brown, E.J. (1999) Role of cholesterol in formation and function of a signaling complex involving $\alpha_v\beta_3$ integrin-associated protein (CD47), and heterotrimeric G proteins, *J. Cell Biol.* **146**, 673–682.

Guan, J.L., Trevithick, J.E., Hynes, R.O. (1991) Fibronectin/integrin interaction induces tyrosine phosphorylation of a 120-kDa protein, *Cell Regul.* **2**, 951–964.

Guo, L., Degenstein, L., Dowling, J., Yu, Q.-C., Wollman, R., Perman, B., Fuchs, E. (1995) Gene targeting of BPAG1: Abnormalities in mechanical strength and cell migration in stratified epithelia and neurologic degeneration, *Cell* **81**, 233–243.

Hagel, M., George, E.L., Kim, A., Tamimi, R., Opitz, S.L., Turner, C.E., Imamoto, A., Thomas, S.M. (2002) The adaptor protein paxillin is essential for normal development in the mouse and is a critical transducer for fibronectin signaling, *Mol. Cell. Biol.* **22**, 901–915.

Han, J., Liu, S., Rose, D.M., Schlaepfer, D.D., McDonald, H., Ginsberg, M.H. (2001) Phosphorylation of the integrin α_4 cytoplasmic domain regulates paxillin binding, *J. Biol. Chem.* **276**, 40903–40909.

Honda, T., Shimizu, K., Kawakatsu, T., Yasumi, M., Shingai, T., Fukuhara, A., Ozaki-Kuroda, K., Irie, K., Nakanishi, H., Takai, Y. (2003) Antagonistic and agonistic effects of an extracellular fragment of nectin on formation of E-cadherin-based cell-cell interaction, *Genes Cells* **8**, 51–63.

Ilic, D., Furuta, Y., Kanazawa, S., Takeda, N., Sobue, K., Nakatsuji, N., Nomura, S., Fujimoto, J., Okada, M., Yamamoto, T. (1995) Reduced cell motility and enhanced focal adhesion contact formation in cells from FAK-deficient mice, *Nature* **377**, 539–544.

Ireton, R.C., Davis, M.A., van Hengel, J., Mariner, D.J., Barnes, K., Thoreson, M.A., Anastasiadis, P.Z., Matrisian, L., Bundy, L.M., Sealy, L., Gilbert, B., van Roy, R., Reynolds, A.B. (2002) A novel role for p120 catenin in E-cadherin function, *J. Cell Biol.* **159**, 465–476.

Kazarov, A.R., Yang, X., Stipp, C.S., Sehgal, B., Hemler, M.E. (2002) An extracellular site on tetraspanin CD151 determines α_3 and α_6 integrin-dependent cellular morphology, *J. Cell Biol.* **158**, 1299–1309.

Koster, J., Geerts, D., Favre, B., Borradori, L., Sonnenberg, A. (2003) Analysis of the interactions between BP180, BP230, plectin and the integrin $\alpha_6\beta_4$ important for hemidesmosome assembly, *J. Cell Sci.* **116**, 387–399.

LaFlamme, S.E., Aiyama, S.K., Yamada, K.M. (1992) Regulation of fibronectin receptor distribution, *J. Cell Biol.* **117**, 437–447.

Lammerding, J., Kazarov, A.R., Huang, H., Lee, R.T., Hemler, M.E. (2003) Tetraspanin CD151 regulates $\alpha_6\beta_1$ integrin adhesion strengthening, *PNAS* **100**, 7616–7621.

Li, R., Mitra, N., Gratkowski, H., Vilaire, G., Litvinov, R., Nagasami, C., Weisel, J.W., Lear, J.D., DeGrado, W.F., Bennett, J.S. (2003) Activation of integrin $\alpha_{IIb}\beta_3$ by modulation of transmembrane helix associations, *Science* **300**, 795–798.

Liu, D., Aguiree-Ghiso, J., Estrada, Y., Ossowski, L. (2002) EGFR is a transducer of the urokinase receptor initiated signal that is required for in vivo growth of a human carcinoma, *Cancer Cell* **1**, 455–457.

Liu, S., Slepak, M., Ginsberg, M.H. (2001) Binding of paxillin to α_9 integrin cytoplasmic domain inhibits cell spreading, *J. Biol. Chem.* **276**, 37086–37092.

Liu, S., Thomas, S.M., Woodside, D.G., Rose, D.M., Kiosses, W.B., Plaff, M., Ginsberg, M.H. (1999) Binding of paxillin to α_4 integrins modifies integrin-dependent biological responses, *Nature* **402**, 676–681.

Mainiero, F., Murgia, C., Wary, K.K., Curatola, A.M., Pepe, A., Blumemberg, M., Westwick, J.K., Der, C.J., Giancotti, F.G. (1997) The coupling of $\alpha_6\beta_4$ integrin to Ras-MAP kinase pathways mediated by Shc controls keratinocyte proliferation, *EMBO J.* **16**, 2365–2375.

Marcozzi, C., Burdett, I.D.J., Buston, R.S., Magee, A.I. (1998) Coexpression of both types of desmosomal cadherin and plakoglobin confers strong intercellular adhesion, *J. Cell Sci.* **111**, 495–509.

Mariotti, A., Kedeshian, P.A., Dans, M., Curatola, A.M., Gagnoux-Palacios, L., Giancotti, F.G. (2001) EGF-R signaling through Fyn kinase disrupts the function of integrin $\alpha_6\beta_4$ at hemidesmosomes: role in epithelial cell migration and carcinoma invasion, *J. Cell Biol.* **155**, 447–458.

Merritt, A.J., Berika, M.Y., Zhai, W., Kirk, S.E., Ji, B., Hardman, M.J., Garrod, D.R. (2002) Suprabasal desmoglein 3 expression in the epidermis of transgenic mice results in hyperproliferation and abnormal differentiation, *Mol. Cell. Biol.* **22**, 5846–5858.

Mettouchi, A., Klein, S., Guo, W., Lopez-Lago, M., Lemichez, E., Westwick, J.K., Giancotti, F.G. (2001) Integrin-specific activation of Rac controls progression through the G(1) phase of the cell cycle, *Mol. Cells* **8**, 115–127.

Miner, J.H., Li, G. (2003) Defective glomerulogenesis in the absence of laminin α_5 demonstrates a developmental role for the kidney glomerular basement membrane, *Dev. Biol.* **217**, 278–289.

Moro, L., Dolce, L., Cabodi, S., Bergatto, E., Erba, E.B., Smeriglio, M., Turco, E., Retta, S.F., Giuffrida, M.G., Venturino, M., Godovac-Zimmermann, J., Conti, A., Schaefer, E., Beguinot, L., Tacchetti, C., Gaggini, P., Silengo, L., Tarone, G., Defilippi, P. (2002) Integrin-induced epidermal growth factor (EGF) receptor activation requires c-Src and p130Cas and leads to phosphorylation of specific EGKF receptor tyrosines, *J. Biol. Chem.* **277**, 9405–9414.

Murgia, C., Blaikie, P., Kim, N., Dans, M., Petrie, H.T., Giancotti, F.G. (1998) Cell cycle and adhesion defects in mice carrying a targeted deletion of the integrin β_4 cytoplasmic domain, *EMBO J.* **17**, 3940–3951.

Niessen, C.M., Gumbiner, B.M. (2002) Cadherin-mediated cell sorting not determined by binding or adhesion specificity, *J. Cell Biol.* **156**, 389–399.

Noakes, P.G., Miner, J.H., Gautam, M., Cunningham, J.M., Sanes, J.R., Merlie, J.P. (1995) The renal glomerulus of mice lacking s-laminin/laminin β_2: nephrosis despite molecular compensation by laminin β_1, *Nat. Genet.* **10**, 400–406.

O'Connor, K.L., Shaw, L.M., Mercurio, A.M. (1998) Release of cAMP gating by the $\alpha_6\beta_4$ integrin stimulates lamellae formation and the chemotactic migration of invasive carcinoma cells, *J. Cell Biol.* **143**, 1749–1760.

Ohkubo, T., Ozawa, M. (1999) $p120^{ctn}$ binds to the membrane-proximal region of the E-cadherin cytoplasmic domain and is involved in modulation of adhesion activity, *J. Biol. Chem.* **274**, 21409–21415.

Piedra, J., Martinez, D., Castano, J., Miravet, S., Dunach, M., Garcia de Herreros, A. (2001) Regulation of β-catenin structure and activity by tyrosine phosphorylation, *J. Biol. Chem.* **276**, 20436–20443.

Piedra, J., Miravet, S., Castano, J., Palmer, H.G., Heisterkamp, N., Garcia de Herreros, A., Dunach, M. (2003) p120 catenin-associated Fer and Fyn tyrosine kinases regulate β-catenin Tyr-142 phosphorylation and b-catenin-a-catenin interaction, *Mol. Cell Biol.* **23**, 2287–2297.

Rabinovitz, I., Toker, A., Mercurio, A.M. (1999) Protein kinase C-dependent mobilization of the $\alpha_6\beta_4$ integrin from hemidesmosomes and its association with actin-rich cell protrusions drive the chemotactic migration of carcinoma cells, *J. Cell Biol.* **146**, 1147–1160.

Reddy, K.B., Gascard, P., Price, M.G., Negrescu, E.V., Fox, J.E.B. (1998) Identification of an interaction between the M-band protein skelemin and β-integrin subunits, *J. Biol. Chem.* **273**, 35030–35047.

Roura, S., Miravet, S., Piedra, J., Garcia de Herreros, A., Dunach, M. (1999) Regulation of E-cadherin/catenin association by tyrosine

phosphorylation, *J. Biol. Chem.* **274**, 36734–36740.

Runswick, S.K., Hare, M.J., Jones, L., Streuli, C.H., Garrod, D.R. (2001) Desmosomal adhesion regulates epithelial morphogenesis and cell positioning, *Nat. Cell Biol.* **3**, 823–830.

Schaapveld, R.Q., Borradori, L., Geerts, D., van Leusden, M.R., Kuikman, I., Nievers, M.G., Niessen, C.M., Steenbergen, R.D., Snijders, P.J., Sonnenberg, A. (1998) Hemidesmosome formation is initiated by the β_4 integrin subunit, requires complex formation of β_4 and HD1/plectin, and involves a direct interaction between β_4 and the bullous pemphigoid antigen 180, *J. Cell Biol.* **142**, 271–284.

Schwartz, M.A., Assoian, R.K. (2001) Integrins and cell proliferation: regulation of cyclin-dependent kinases via cytoplasmic signaling pathways, *J. Cell Sci.* **114**, 2553–2560.

Shapiro, L., Fannon, A.M., Kwong, P.D., Thompson, A., Lehmann, M.S., Grubel, G., Legrand, J.-F., Als-Nielsen, J., Colman, D.R., Hendrickson, W.A. (1995) Structural basis of cell-cell adhesion by cadherins, *Nature* **374**, 327–337.

Shaw, L.M., Rabinovitz, I., Wang, H.H., Toker, A., Mercurio, A.M. (1997) Activation of phosphoinositide 3-OH kinase by the $\alpha_6\beta_4$ integrin promotes carcinoma invasion, *Cell* **91**, 949–960.

Shimoyama, Y., Tsujimoto, G., Kitajima, M., Natori, M. (2000) Identification of three human type-II classic cadherins and frequent heterophilic interactions between different subclasses of type-II classic cadherins, *Biochem. J.* **349**, 159–167.

Shyy, Y., Chien, S. (2002) Role of integrins in endothelial mechanosensing of shear stress, *Circ. Res.* **91**, 769–775.

Smith, F.J.D., Eady, R.A.J., Leigh, I.M., McMillan, J.R., Rugg, E.L., Kelsell, D.P., Bryant, S.P., Spurr, N.K., Geddes, J.F., Kirtschig, G., de Bono, A.G., Owaribe, K., Wiche, G., Pulkkinen, L., Uitto, J., McLean, W.H.I., Lane, E.B. (1996) Plectin deficiency results in muscular dystrophy with epidermolysis bullosa, *Nat. Genet.* **13**, 450–457.

Sterk, L.M., Geuijen, C.A., van den Berg, J.G., Claessen, N., Weening, J.J., Sonnenberg, A. (2002) Association of the tetraspanin CD151 with the laminin-binding integrins $\alpha_3\beta_1$, $\alpha_6\beta_1$, $\alpha_6\beta_4$ and $\alpha_7\beta_1$ in cells in culture and in vivo, *J. Cell Sci.* **115**, 1161–1173.

Tachibana, K., Nakanishi, H., Mandai, K., Ozaki, K., Ikeda, W., Yamamoto, Y., Nagafuchi, A., Tsukita, S., Takai, Y. (2000) Two cell adhesion molecules, nectin and cadherin, interact through their cytoplasmic domain-associated proteins, *J. Cell Biol.* **150**, 1161–1175.

Tahiliani, P.D., Singh, L., Auer, K.L., LaFlamme, S.E. (1997) The role of conserved amino acid motifs within the integrin β_3 cytoplasmic domain in triggering focal adhesion kinase phosphorylation, *J. Biol. Chem.* **272**, 7892–7898.

Takagi, J., Petre, B.M., Walz, T., Springer, T.A. (2002) Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling, *Cell* **110**, 599–611.

Tselepis, C., Chidgey, M., North, A., Garrod, D. (1998) Desmosomal adhesion inhibits invasive behavior, *Proc. Natl. Acad. Sci.* **95**, 8064–8069.

Tu, Y., Wu, S., Shi, X., Chen, K., Wu, C. (2003) Migfilin and Mig-2 link focal adhesions to filamin and the actin cytoskeleton and function in cell shape modulation, *Cell* **113**, 37–47.

Turner, C.E., West, K.A., Brown, M.C. (2001) Paxillin-ARF GAP signaling and the cytoskeleton, *Curr. Opin. Cell Biol.* **13**, 593–599.

van der Neut, R., Krimpenfort, P., Calafat, J., Niessen, C.M., Sonneberg, A. (1996) Epithelial detachment due to absence of hemidesmosomes in integrin $\beta4$ null mice, *Nat. Genet.* **13**, 366–369.

Van Itallie, C., Rahner, C., Anderson, J.M. (2001) Regulated expression of claudin-4 decreases paracellular conductance through a selective decrease in sodium permeability, *J. Clin. Invest.* **107**, 1319–1327.

van Kooyk, Y., Figdor, C.G. (2000) Avidity regulation of integrins: the driving force in leukocyte adhesion, *Curr. Opin. Cell Biol.* **12**, 542–547.

Wang, F., Weaver, V.M., Petersen, O.W., Larabell, C.A., Dedhar, S., Briand, P., Lupu, R., Bissell, M.J. (1998) Reciprocal interactions between β_1-integrin and epidermal growth factor receptor in three-dimensional basement membrane breast cultures: a different perspective in epithelial biology, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14821–14826.

Wary, K.K., Mainiero, F., Isakoff, S.J., Marcantonio, E.E., Giancotti, F.G. (1996) The adaptor protein Shc couples a class of integrins to

the control of cell cycle progression, *Cell* **87**, 733–743.

Wary, K.K., Mariotti, A., Zurzolo, C., Giancotti, F.G. (1998) A requirement for caveolin-1 and associated kinase Fyn in integrin signaling and anchorage-dependent cell growth, *Cell* **94**, 625–634.

Weaver, V.M., Petersen, O.W., Wang, F., Larabell, C.A., Briand, P., Damsky, C., Bissell, M.J. (1997) Reversion of the malignant phenotype of human breast cells in three-dimensional culture and in vivo by integrin blocking antibodies, *J. Cell Biol.* **137**, 231–245.

Welsh, C.F., Roovers, K., Villanueva, J., Liu, Y., Schwartz, M.A., Assoian, R.K. (2001) Timing of cyclin D1 expression within G1 phase is controlled by Rho, *Nat. Cell Biol.* **3**, 950–957.

West, K.A., Zhang, H., Brown, M.C., Nikolopoulos, S.N., Riedy, M.C., Horwitz, A.F., Turner, C.E. (2001) The LD4 motif of paxillin regulates cell spreading and motility through an interaction with paxillin kinase linker (PKL), *J. Cell Biol.* **154**, 161–176.

Yap, A.S., Brieher, W.M., Pruschy, M., Gumbiner, B.M. (1997) Lateral clustering of the adhesive ectodomain: a fundamental determinant of cadherin function, *Curr. Biol.* **7**, 308–315.

Yap, A.S., Niessen, C.M., Gumbiner, B.M. (1998) The juxtamembrane region of the cadherin cytoplasmic tail supports lateral clustering, adhesive strengthening, and interaction with $p120^{ctn}$, *J. Cell Biol.* **141**, 779–789.

Young, B.A., Taooka, Y., Liu, S., Askins, K.J., Yokosaki, Y., Thomas, S.M., Sheppard, D. (2001) The cytoplasmic domain of the integrin α_9 subunit requires the adaptor protein paxillin to inhibit cell spreading but promotes cell migration in a paxillin-independent manner, *Mol. Biol. Cell* **12**, 3214–3225.

Zhang, X.A., Hemler, M.E. (1999) Interaction of the integrin β_1 cytoplasmic domain with ICAP-1 protein, *J. Biol. Chem.* **274**, 11–19.

Zhao, J., Pestelll, R., Guan, J.L. (2001) Transcriptional activation of cyclin D1 promoter by FAK contributes to cell cycle progression, *Mol. Biol. Cell* **12**, 4066–4077.

Cell Nucleus Biogenesis, Structure and Function

Dean A. Jackson
Department of Biomolecular Sciences, Manchester, UK

1 Introduction 379

2 Nuclear Function 380
2.1 Gene Expression 381
2.1.1 Activating Gene Expression 381
2.1.2 Specific Transcription Factors 382
2.1.3 Posttranscriptional Events – RNA Processing 383
2.2 DNA Replication 384
2.2.1 DNA Repair 386
2.3 The Cell Division Cycle 388
2.3.1 Mitosis 389
2.3.2 Meiosis 391

3 Nuclear Structure 392
3.1 Nuclear Pore Structure 392
3.1.1 Pore Function 393
3.2 The Nuclear Lamina 393
3.2.1 The Lamin Genes 395
3.2.2 The Lamin Proteins and Filament Assembly 395
3.2.3 Lamin Function 395
3.2.4 Lamin-associated Proteins 396
3.2.5 Lamin Mutants 396
3.3 Biogenesis of the Nuclear Envelope During Cell Proliferation 397
3.4 Nuclear Transport and Karyopherins 398
3.4.1 The RanGTP/GDP Cycle 398
3.4.2 RNA Export Pathways 399
3.4.3 Nuclear Transport – Facilitated Diffusion 400

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

3.5 Nuclear Architecture and Nuclear Compartments 400
3.5.1 Nuclear Compartments 400
3.5.2 The Nucleolus 401
3.5.3 Nuclear Bodies and Domains 402
3.5.4 Nuclear Speckles 403
3.5.5 Active Sites – Synthetic Factories 404
3.6 Chromosome Territories 408
3.6.1 Chromosome Structure 410
3.6.2 Functional Implications of Chromosome Structure and Location 410
3.7 The Principles of Global Nuclear Structure 411
3.7.1 Organizing the Cytoplasm 411
3.7.2 The Nuclear Matrix – Fact or Fiction 413
3.7.3 Proteins of the Nuclear Matrix 413
3.7.4 Functions at the Matrix 415

4 Integrating Nuclear Structure and Function 415
4.1 Chromatin Structure and Function 416
4.1.1 Chromatin Function 416
4.1.2 DNA Packaging and Chromatin 416
4.1.3 Euchromatin and Heterochromatin 418
4.1.4 Histone Modifications 418
4.1.5 ATP-dependent Chromatin Remodeling Machines 419
4.2 Higher-order Chromatin Folding 420
4.2.1 Chromatin Domains and Loops 421

5 Overview 423

Bibliography 423
General Reference and Books 423
Research Articles are Reviews 424

Keywords

Chromosome
The structural unit of genetic material.

Nuclear Architecture
Nuclear structures that influence nuclear function.

Nuclear Matrix
Protein network throughout the nucleus.

Gene Expression
Genes are decoded into RNA, which is translated into protein.

DNA Replication
DNA is copied by enzymes into 2 identical molecules.

DNA Repair
DNA damage is corrected to maintain genetic integrity.

Cell Cycle
Sequence of events that yields 2 daughter cells from a single mother cell.

The nucleus is an organelle that is found in eukaryotes but not in prokaryotes. Nuclei contain the genetic information that defines the appearance and behaviour of an organism. The genetic material - DNA - is packaged as a DNA-protein complex, called chromatin, into units of manageable size, called chromosomes. In multicellular eukaryotes, different types of cells are defined by the different genes that they express. Patterns of gene expression arise during cell differentiation and reflect a cells position in the organism during development. Gene expression is regulated at many levels in each cell. The major regulatory step occurs during transcription, when the gene is copied by transcription into messenger RNA. Messenger RNAs transfer genetic information to the cytoplasm and provide the template for protein synthesis during translation. Higher eukaryotes contain very large numbers of cells that are derived from a single precursor - a fertilised egg. The genetic material of this cell is derived from the egg and sperm of the female and male parents. During development, cells must perform many cycles of DNA duplication and cell proliferation. The genetic information must be faithfully copied during DNA replication and any damaged corrected by DNA repair, so that the same genetic information is passed to all cells of the organism. This chapter describes these fundamental biological processes and explains how they are regulated within the nucleus of eukaryotic cells.

1 Introduction

Eukaryotes are defined by the presence of a nucleus in their cells. In most eukaryotes, the nucleus is the most prominent organelle. It is the cellular compartment where the vast majority of genetic information is held (mitochondria and chloroplasts contain very small DNA molecules). The genetic code is held within the DNA. Each human cell has roughly 6×10^9 bp DNA. This is far too much to be held in a single molecule within the cell, as in bacteria, so the DNA is divided between a number of

chromosomes. Somatic human cells are diploid and a vast majority have the same genetic information, which is held within the 46 chromosomes – 22 pairs of autosomes called chromosomes 1–22 and 1 pair of sex chromosomes, XX in females or XY in males. This DNA is folded as chromatin to form chromosomes and it is this DNA–protein complex that serves as the genetic substrate that supports the major functions performed by DNA. The major functions are as follows:

1. DNA synthesis – called *DNA replication* – ensures that the same genetic complement is passed from the fertilized egg to all subsequent cell descendants.
2. RNA synthesis – called *RNA transcription* – ensures that genes are expressed in the appropriate cells at the required times throughout the development of an organism. Different patterns of gene expression in different types of cells and tissues are determined by patterns of differentiation.
3. DNA repair ensures that the integrity of genetic information is maintained so that the same DNA sequence is found in all cells of the body. Defects in DNA repair can lead to a variety of diseases through sporadic mutations, and are particularly important during the development of cancer.

Different aspects of chromatin function and nuclear structure are fundamental to the development of multicellular organisms. The nucleus itself plays a vital role in this development by ensuring that functions performed on chromatin are isolated from major cytoplasmic activities such as protein synthesis and energy metabolism. This separation inevitably demands that a critical step of the gene expression process involves the transfer of information from each gene to the cytoplasm where it can be decoded to generate the corresponding protein. The genetic intermediary involved is called *messenger RNA* (mRNA) and the process of protein synthesis is called *protein translation*. The fact that eukaryotic cells are divided into two major compartments by the nuclear membrane has a number of implications. One disadvantage is the need to develop complex systems to regulate compartmentalization and control traffic between the two compartments. The major advantage is that separating protein synthesis from the compartment in which DNA function is performed allows quality controls to be installed to ensure that protein synthesis is only performed on intact and authentic mRNAs. In bacteria, transcription, translation, protein folding, and incorporation of the nascent folding polypeptide into the functional sites are coupled, so that in principle the nascent protein might still be attached – through the ribosome, mRNA and RNA polymerase – to its gene. In eukaryotes, much more complex gene structures, and in particular, the fact that most genes contain intervening sequences that do not hold protein coding information, demands that the events of mRNA and protein synthesis are uncoupled.

2 Nuclear Function

The basic molecular mechanisms that dictate most aspects of nuclear function are known in detail and it is only appropriate to cover the fundamental principles here.

2.1 Gene Expression

Eukaryotic gene expression follows the general principles developed during the evolution of prokaryotes. Put simply, gene expression is dictated by DNA sequences within gene promoters that determine how the RNA synthetic machinery, RNA polymerase, is positioned on the gene. In multicellular eukaryotes, highly complex patterns of gene expression have evolved and correspondingly complex mechanisms of gene regulation are seen. Nevertheless, the same basic principle underlies the activation of RNA synthesis.

In contrast to prokaryotes, which have a single RNA polymerase, three different RNA polymerase complexes perform RNA synthesis in eukaryotic cells. In most mammalian cells, RNA polymerase II (pol II) is the major activity, transcribing all protein-coding genes to generate patterns of gene expression that determine cell type; humans have about 250 distinct cell types. Synthesis is performed by an ~4 MDa holoenzyme containing the pol II core enzyme and other activities required during RNA synthesis and processing. RNA polymerase I (pol I) is dedicated to the synthesis of the repeated ribosomal RNA (rRNA) genes, within specialized nuclear sites – nucleoli – and RNA polymerase III (pol III), a minor nucleoplasmic activity, transcribes transfer RNA (tRNA) and 5S rRNA genes. Small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA) genes encode structural RNAs needed for RNA processing; some are transcribed by pol II and others by pol III.

A proliferating mammalian cell supports a continuous rate of RNA synthesis of $\sim 2 \times 10^8$ nucleotides min^{-1}. This will be roughly 40% pre-rRNA, 60% pre-mRNA (hnRNA) and only 2 to 3% 4 to 5S RNA (5S and pre-tRNA). Synthetic rates are 3 to 4 kbp min^{-1} for pol I and 1.5 to 2 kbp min^{-1} for pol II/III respectively and average transcript lengths of ~13.5, ~15 and ~0.1 kb for pol I, pol II, and pol III. These values correspond to ~20 000, ~60 000 and ~3000 engaged pol I, II, and III complexes per cell. It is also important to consider the density of pol complexes on individual genes. A chromatin spreading technique called *Miller spreads* can be used to show that the active rRNA genes of mammalian cells each support 100 to 120 pol I complexes. In sharp contrast, pol II transcription units generally support very few widely dispersed active complexes. Even when studying adenovirus at the time of the highest viral transcription, the active genomes have only four (on average) associated transcription complexes.

Most tissues of mammalian origin express at least 10 000 genes. In rat liver, for example, there are ~10 species present at ~10 000 copies per cell, 500 at ~200 copies per cell and ~10 000 at ~10 copies per cell (33) – ~300 000 mRNAs per cell, in all. Note that even after accounting for mRNA turnover, it is clear that the majority of genes generate so few transcripts that they must be transcriptionally inert for most of the time. In addition, it is interesting that many pol II complexes seem to be involved in synthesis that generates RNA molecules that turnover in the nucleus. Many of these are intergenic transcripts that have been speculated to play a role in maintaining an active chromatin status across extended gene domains.

2.1.1 Activating Gene Expression

The critical feature of gene expression in eukaryotes is that the transcription machinery is recruited to genes through its interaction with transcription factors and

adaptor proteins that interact with these factors. Genetic elements within the gene promoters define where the transcription will begin. General transcription factors play a fundamental role in this process. For most eukaryotic promoters, the process is activated by association of the factor TFIID with DNA. TFIID is a large multiprotein complex, one subunit of which, the TATA sequence binding protein (TBP), can recognize and associate with the canonical promoter element TATA. TFIID bound to the promoter can then activate the sequential recruitment of other factors as follows: TFIIB; TFIIE with TFIIF and RNA polymerase; TFIIH. Once assembly of this preinitiation complex is complete, the TFIIH complex directs phosphorylation of a domain of the RNA polymerase complex called the *C-terminal domain* (CTD). The CTD is conserved in eukaryotic RNA polymerase II proteins. The mammalian enzyme has 52 repeats of the amino acid (consensus) sequence YSPTSPS, in which the serine (S) and threonine (T) residues can be phosphorylated. Phosphorylation of the CTD correlates with release of the synthetic complex from the promoter and the initiation of RNA synthesis. During synthesis, the DNA template is used to generate a corresponding RNA molecule, which is polymerized from its 5′ toward its 3′ end. The 3′ end of the transcript is generated by a process that cuts the nascent RNA and adds a poly(A) tail. This process recognizes a set of sequence elements in the RNA that interact with the poly(A) synthesis machinery. Once the nascent RNA is processed in this way, the engaged polymerase complex is destabilized so that it will soon dissociate from the template – this termination step appears not to be directed by a specific sequence in DNA and is thought to be more probabilistic, so that the polymerase "drops off" the DNA at some point within a few kbp at the end of the gene.

2.1.2 Specific Transcription Factors

The general transcription factors provide only part of the mechanism that regulates gene expression in higher eukaryotes. In multicellular organisms, it is clear that complex mechanisms of regulation are required to ensure that particular genes are expressed in the desired cells and at the appropriate times of development. This function is performed by specific transcription factors. As the name implies, specific transcription factors function to activate transcription from a single gene or a group of genes with similar expression characteristics. As a general rule, genes that are expressed in a specific cell lineage will only be expressed in cells with the appropriate constellation of transcription factors. Certain specific transcription factors will play a dominant role in maintaining the active chromatin status across target genes. Others will activate or enhance levels of gene expression beyond the levels that are provided by the general transcription factors alone.

It is self-evident that this class of transcription factors must achieve specificity by virtue of their interaction with specific DNA elements within target genes. For most factors, the DNA target will lie within 1 to 2 turns of the double helix. These DNA elements can be located close to the promoter itself, within promoter proximal activating sequences, but, in many cases, these are located in transcriptional enhancer elements that can be as much as 100 kbp for the gene promoter. This surprising observation reflects the fact that the chromatin itself is flexible and able to fold, so that the remote sequences and associated factors can contribute to the efficiency with which gene expression is

activated at the corresponding promoter. An excellent example of this is provided by enhancer elements within the mammalian β-globin locus control region (LCR), which are located about 50 kbp from the gene promoters in this locus that are activated at different times of development.

Particular features of the transcription factor proteins generate binding specificity. Many classes of transcription factor have been described, but the majority of factors fall into three main groups with (1) helix–turn–helix motifs; (2) zinc fingers; (3) leucine zippers.

The homeodomain of vertebrate hox genes is a classical example of the helix–turn-helix transcription factor motif. The homeodomain is composed of 60 amino acids that fold as three helices. Helices 2 and 3 form the helix–turn–helix motif and helix 3 interacts with the major groove of DNA.

Though variants have been described, the characteristic feature of the zinc finger motif is a zinc atom associated with two cysteine and two histidine residues. The human Gli protein is a good example that is involved in bone development and also amplified in some human tumors. The factor contains five zinc fingers, of which four interacts with each of the major grooves of DNA where they wrap around the DNA for a full helical turn. Extensive contacts between the amino acids of these structures and the bases and phosphate residues in DNA together generate specificity of factor binding.

The human oncogenes *Fos* and *Jun* are members of the family of transcription factors that bind to AP-1 sites within promoters. *Fos* and *Jun* are able to form heterodimers by virtue of the precise spacing of hydrophobic leucine residues in the domains of the two monomers that provide a surface for interaction – the leucine zipper. The alpha helices within the monomers "zip" together to create a coiled coil and a proximal basic region that is generated in the dimer to provide a DNA-binding domain. This basic region has two lysine/arginine-rich regions separated by an invariant asparagine that together generate an interaction with the major groove of the DNA.

2.1.3 Posttranscriptional Events – RNA Processing

In eukaryotes, the primary transcripts of protein coding genes must be processed before the mature messenger RNAs (mRNAs) are able to pass to the cytoplasm. The processing steps include the following:

1. Addition of a 5′ cap structure – this occurs cotranscriptionally and serves a role in RNA structure and stability. Formation of the RNA cap involves the addition of 7-methyl guanylic acid to the 5′ triphosphate of the nascent transcript, in a reaction that involves the two triphosphate moieties. During this three-step reaction, the γ-phosphate of the RNA is first removed by RNA triphosphatase, GTP-RNA guanyltransferase then adds GMP to the new terminal and finally the methyl group is added by RNA-guanine-7-methyltransferase. The cap may be further modified by the addition of 2–O-methyl groups to the first or first and second ribose residues of the transcript.
2. The protein coding sequences of many eukaryotic genes are interspersed with noncoding introns, which must be removed in order to generate mature mRNA. This process, called *RNA splicing*, begins at the time of transcription and is completed posttranscriptionally, at the transcription site. Splicing is

catalyzed by a number of small nuclear RNA-protein complexes (snRNPs). The RNAs involved are called U1, U2, U4, U5, and U6 (U3 is involved in rRNA processing). U1 snRNP binds to the 5′ splice site and U2 snRNP the intronic branch acceptor site. A U4/U6.U5 complex then assembles so that U6 snRNP replaces U1 at the 5′ splice site. U2 and U6 RNAs then pair to form the catalytic site and U5 pairs with the end of the 5'exon. Splicing proceeds through two transesterification reactions that form first a lariat at the internal acceptor site and finally the spliced product. Many proteins – such as the SR proteins (see 3.5.4) – are involved in this splicing reaction and one consequence of the splicing process is the formation of a mature RNA–protein complex that is required for mRNA export to the cytoplasm.

3. The 3′ end of the mRNA is generated by a processing event that adds a poly(A) tail (typically of 150–200 residues) to the nascent transcript. This serves a structural role and is important for RNA stability. Two elements within RNA determine the location of the poly(A) tail. The polyadenylation signal, AAUAAA, binds the cleavage poly(A) specificity factor (CPSF) protein and a downstream element, UGUGUG, binds the cleavage stimulatory factor (CstF). These two proteins interact with and position the cleavage factor proteins (CFI and CFII) and poly(A) polymerase (PAP), which together generate cleaved RNA with a poly(A) tail.

2.2 DNA Replication

The vast majority of cells in multicellular eukaryotes contain the same genetic information that is descended from an original diploid cell – the fertilized egg. It is self-evident, that if cells are to maintain their genetic integrity the proliferation process must be tightly controlled so that cell division can only proceed once the genetic material – DNA – of the mother cell is fully duplicated. This process of DNA replication is clearly central to the activities of cell proliferation and is the target for a wide variety of cellular controls and checkpoints (quality controls).

Prokaryotes such as the bacterium *Escherichia coli* have a circular genome and initiate DNA synthesis at a single site, the replication origin, *OriC*. The molecular details of this process are known. Quite clearly, the size and complexity of eukaryotic genomes demand that many more initiation events occur. In human cells, origins of DNA replication are about 150 kbp apart, which means that an average chromosome will have ~1000. The molecular mechanisms that define replication origins involve the association of a protein complex with DNA. This complex includes a multi-subunit origin recognition complex (ORC1–6), and an MCM 2 to 7 complex that restricts initiation to one event/origin in each cell cycle, and proteins that are targets for the cell cycle machinery that activates the replication process. Once an origin is activated the synthetic machinery is recruited to the origin and replication ensues.

For most eukaryotes, it is not clear how the origins of DNA replication are defined. The yeast *Saccharomyces cerevisiae* is an exception. In this organism, replication origins can be defined as autonomous replicating sequence (ARS) elements that are about 200 bp in length and have consensus sequence motifs – called *the A and B elements* – that act as DNA unwinding

elements and serve as target sites for the recruitment of ORC. No equivalent sequences have been defined in multicellular eukaryotes and it is believed that origins are defined by a complex combination of genetic and epigenetic features.

Duplication of the DNA double helix proceeds by a semiconservative mechanism. The DNA duplex is first opened to separate the two strands, each of which then serves as a template to generate two identical daughter DNA molecules. The synthetic enzymes responsible for the synthetic process are called *DNA polymerases* (Table 1). Many other proteins are involved in the replication process (Fig. 1). The critical first step of the replication process, DNA denaturation is driven by an enzyme called *DNA helicase*. Helicases unwind the DNA template. The newly formed single-stranded regions are stabilized by a DNA-binding protein called *replication factor-A* (*RF-A*). The DNA replication then begins. However, DNA polymerases are unable to initiate the synthetic process-initiation requires that a DNA dependent RNA polymerase called *a primase*, first generates short (usually about 5 bp) RNA primers at each replication origin. DNA polymerase-α then proceeds to synthesize DNA by extending the RNA primer. DNA polymerase-α is not especially processive (i.e. efficient at copying long stretches of DNA) and soon after initiation the synthetic process is assumed by DNA polymerases-δ and/or -ε.

The antiparallel structure of DNA adds a certain complexity to the replication process. Because of the reaction mechanisms that drive chain elongation, the polymerization process can take place only in one direction – nucleic acids only grow from the 5′ to 3′ ends. Replication from a single origin almost always occurs in two directions, so that each origin will have four associated polymerase complexes, two at each growing replication fork (Fig. 1 shows the proteins at one of the two forks). At each fork, replication of one

Tab. 1 Eukaryotic DNA polymerases.

Name	*Family*	*Amino acids*	*Major function*
Pol α	B	1462	Initiates replication
Pol β	X	335	Base excision repair
Pol γ	A	1239	Mitochondrial DNA replication
Pol δ	B	1107	Nuclear DNA replication
Pol ε	B	3000	Nuclear DNA replication
Pol ζ	B	3130	Mutagenic synthesis?
Pol η	Y	713	Translesion synthesis
Pol θ	A	2724	Cross-link repair?
Pol ι	Y	715	Specialized base excision repair?
Pol κ	Y	870	Unknown
Pol λ	X	575	Base excision repair, meiotic roles
Pol μ	X	495	ds DNA break repair
Pol σ	X	543	Sister chromatid cohesion
REV1	Y	1251	Mutagenic translesion DNA synthesis

Note: See Bebenek, K., Kunkel, T.A. (2002) Family growth: the eukaryotic DNA polymerase revolution, *Cell Mol. Life Sci.* **59**, 54–57 for further details.

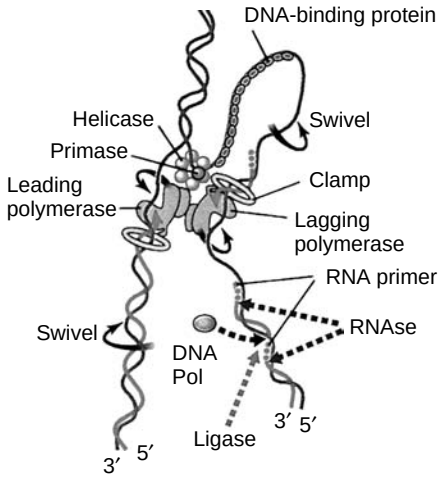

Fig. 1 Replication proteins of a single replication fork. Two replication complexes are required to replicate each replication fork. The activity of the two complexes is coordinated in the cell – they are thought to be associated as shown here. The requirement for leading and lagging strand synthesis is described in the text. In mammalian cells, the DNA clamp, which holds the polymerase complex on DNA is a trimer of PCNA (proliferating cell nuclear antigen) and the primer that is needed to activate DNA synthesis is a component of DNA polymerase-α. DNA synthesis is performed by the combined activities of DNA polymerases-α, -δ, and -ε. The swivel that is required to release superhelical stress that develops during elongation is provided by topoisomerases I and II. (Reproduced from von Hippel, P.H., Jing, D.H. (2000) Structural biology – Bit players in the trombone orchestra, *Science* **287**, 2435–2436 and adapted from Yuzhakov, A., Kelman, Z., O'Donnell, M. (1999) Trading places on DNA – A three-point switch underlies primer handoff from primase to the replicative DNA polymerase, *Cell* **96**, 153–163. With permission of American Association for the Advancement of Science and Cell Press.)

strand can occur continuously in the 5′ to 3′ direction. This is called the *leading strand synthesis*. On the other strand, called the *lagging strand*, replication occurs only when a single-stranded patch of about 200 bp has formed. Synthesis then proceeds to fill the patch, so that the replication products on the lagging strand are all initially short, in mammalian cells typically 100 to 500 bp in length. These short fragments, called *Okazaki fragments*, are then ligated together – by an enzyme called *DNA ligase* – in a process that replaces the RNA primers with DNA and joins the gaps between adjacent Okazaki fragments. It is believed that the leading and lagging strand enzyme complexes are associated so that their activities can be readily coordinated (Fig. 1) and it has even been suggested that the complexes associated with the two forks arising from each origin are held together.

2.2.1 DNA Repair

The maintenance of genetic integrity demands that the DNA is duplicated faithfully from one cell generation to the next and that any damage that might arise in the DNA can be recognized and corrected. In mammalian cells, our best estimates suggest that the sequence of DNA changes at a rate of 1 bp/10^9 bp replicated during each division cycle. In fact, it is important that this level of accuracy is maintained, as most mutations in protein coding genes are likely to be deleterious. Repair is clearly extremely important. Xeroderma pigmentosum (NER), Blooms syndrome (ligase) and ataxia telangiectasia (ATM) are examples of diseases in which affected individuals have defective DNA repair pathways. These defects correlate with a massively increased risk of cancer.

In proliferating cells, a major potential source of DNA mutation is base misincorporation during the course of replication.

Errors are not uncommon, but very effective proofreading and postreplicative mismatch repair activities ensure that the genetic information encoded in the parental DNA strand is preserved during DNA replication.

In addition to this, numerous sources of DNA damage have the potential to generate lesions that will ultimately give rise to mutations in DNA (Table 2). Classes of DNA damage include the following:

1. Chemical or oxidative damage that might lead to base modification or base excision. For example, spontaneous hydrolysis can cause deamination (e.g. turning cytosine to uracil) and depurination (e.g. removing a guanine base).
2. Ultraviolet light induces a photochemical cyclization between adjacent thymine/thymine or thymine/cytosine residues.
3. Ultraviolet light and other types of radiation can induce single- and double-strand breaks in the DNA backbone.
4. Some chemicals can interact with the DNA to form bulky adducts or intercalate into the DNA. Both processes are potentially mutagenic.

Together these insults ensure that the DNA within each mammalian cell will experience many thousands of lesions every day. To cope with this, cells employ a variety of repair mechanisms (Table 2). In each case, the efficiency of the process relies on the fact that the structure of DNA provides a template for repair. The general mechanism of DNA repair involves three basic steps. The first involves the recognition of damage and excision of the damage by DNA repair nucleases. A DNA polymerase can then bind to the free 3′-OH end of the cut DNA and fill the gap using the complementary DNA strand as template. Finally, the nick that remains in the DNA backbone following incorporation of the new base(s) is sealed by the DNA ligase.

Depurination is the most common form of a potentially mutagenic lesion

Tab. 2 DNA repair mechanisms and damage types.

Mechanism	*Damage*	*Repair proteins involved*
Exonuclease of DNA polymerase	Inaccuracies during DNA replication (proofreading)	DNA pol δ/ε, ligase
Base-mismatch repair	Small adducts	MSH6/3-MSH2 (dimer)
	Insertions/deletions	PMS1, MLH1, PCNA
	Oxidative damage	DNA polδ, ligase
Nucleotide excision repair	Bulky adducts	XPA – XPG, RPA, TFIIH
	UV cross-links	ERCC1, DNApolδ/ε
		PCNA, ligase
Base excision repair	Abasic sites	N-glycosidases
	Deamination	AP endonuclease
	Oxidative damage	DNApolβ, DNA ligase
	Alkylations	
Homologous recombination	Strand breaks	RAD51, other RAD proteins RPA, DNApol, ligase
		p53, BRCA1, BRCA2
DNA end-joining	Strand breaks	Ku70, XRCC5, DNAPK p53, ATM

in DNA. In this case, the DNA with a missing base is recognized by the enzyme AP endonuclease, which then cuts the DNA phosphodiester backbone at the 5′ side of the altered site. The sugar and phosphate are then removed by a phosphodiesterase and DNA sequence restored by DNA polymerase-β and DNA ligase using the complementary DNA strand as template. Base excision repair provides an alternative repair pathway that involves a family of DNA glycosylases. These enzymes recognize altered bases in DNA and catalyze their removal so that the natural DNA sequence can be restored.

Nucleotide excision repair provides a quite different repair pathway that is capable of repairing many types of DNA that distort the structure of DNA. Typical example of damage that is removed by this pathway include the thymine dimers and 6 to 4 photoproducts (C-T dimers) caused by ultraviolet irradiation (sunlight) and "bulky adducts" such as those created by the covalent reaction of DNA with hydrocarbons like the carcinogen benzopyrene. In this case, a large multienzyme complex scans DNA for damage and on recognition uses nucleases to remove a DNA patch of about 30 bp. Helicase activity (XPD) is required to remove the DNA strand containing the damage and this is then repaired by DNA polymerase and ligase using the complementary DNA strand as template. The process also involves the transcription factor TFIIH. This is recruited to the repair site and concomitantly leads to a downregulation of RNA synthesis, which prevents the cell from attempting transcription of damaged DNA.

The importance of preserving the genome intact cannot be overstated as many classes of lesion can be mutagenic and will ultimately lead to cancer. Robust pathways are in place that minimize the likelihood of damage leading to cancer. For example, both ionizing and ultraviolet radiation induce the activity of protein kinases such as the mammalian ATM protein, which alters the phosphorylation status of the protein p53; p53 has been termed *the guardian of the genome*. This enhances the stability of p53 by reducing its interaction with the inhibitor of p53 function called MDM2 and stimulates the transcription factor activities of p53, increasing the synthesis of p21, an inhibitor of cell cycle progression, and the proapoptotic protein BAX. If DNA damage is sufficiently severe, the cell is committed to die. The combined activities of ATM, ATR, chk1, and chk2 also serve to ensure that DNA synthesis is downregulated, to reduce the possibility that replication might proceed through damaged regions of the genome and so fix mutations into DNA.

2.3 The Cell Division Cycle

Multicellular organisms must continuously replace cells that are damaged or have otherwise fulfilled their natural purpose. Cell replacement operates through a proliferative "cell cycle" during which mitogenic cues (growth factors) from the local environment activate a complex series of events that will ultimately lead each parental or mother cell to divide into two genetically identical daughters. The cell cycle should be viewed as a continuous process, though for descriptive simplicity it is usually broken down into four phases – G1, S, G2, and mitosis. Mitosis is by far the most visibly distinct phase, which can be watched in living cells using a simple optical microscope. Mitosis might be viewed as the culmination of the

cycle as it incorporates the dramatic structural changes that are required to generate the daughter cells. Mitosis can only occur if the chromosomes within the mother cell are first duplicated. This is the most critical requirement of the proliferation process and eukaryotes have developed a number of quality control mechanisms called *checkpoints* to ensure that growing cells can only attempt mitosis once the process of DNA duplication is complete. Other mechanisms ensure that the initiation of replication is tightly regulated so that chromosomes are replicated once, and only once in each cycle. The process of DNA synthesis takes place in the S-phase (for synthesis phase) of the cell cycle. G1 and G2 are gap phases that separate the distinct S- and M-phases. During these phases – and also during S-phase – the cell grows and so continuously assimilates and accumulates the components that are required to generate two new cells at the end of mitosis. It goes without saying that if proliferating cells are to maintain the required structural characteristic, each cell that forms during mitosis must double in mass before the next mitosis takes place.

For mammalian cells, a typical cell cycle will be completed in about 24 h. Breaking down as: G1 6 to 10 h; S 8 to 10 h; G2 3 to 5 h; M 1 h. Variations in cell cycle occur predominantly during G1. The major cues for proliferation are driven by families of cell cycle proteins called *cyclins* and cyclin dependent kinases (cdks or cdcs). The critical step that commits a cell to proliferation occurs late in G1, about 2 h prior to the onset of S-phase. This position of the cell cycle is referred to as "start" and once a cell passes this point it is locked into a cycle of events that will inevitably lead to the formation of the daughter cells. A critical part of this transition involves the phosphorylation of the tumor suppressor protein Rb (retinoblastoma protein) by a cyclinD/cdk4/6 and cyclinE/cdk2 complexes (Fig. 2). This Rb modification alters the association of Rb with a family of transcription factors – E2F1–6 and DP1–2 – that function to activate the expression of other genes involved in the cell cycle progression and DNA replication, such as cdc2 and DNA polymerase 2. The activity of the cyclin/cdk complexes is driven by mitogenic cues and restrained by the tumor suppressor proteins $p21^{cip/kip}$ and $p16^{INK4a}$.

Cells with protracted cell cycle times have extended pre-start periods. In addition to this, cells with proliferative potential can "rest" in a G0 phase of the cycle. This state is usually perceived as an extension of the G1 phase where cells are able to respond to growth signals once the appropriate signaling cues are provided. Once cells in this quiescent state are activated by the requisite mitogens, it will usually take 15 to 20 h before the onset of S-phase. It is also important to remember that most cells have limited proliferative potential so that after growing and dividing for some time (perhaps 50-cell doubling would be a typical duration) they assume a nonproliferative state. Now cells will be destined to perform the specific function that their pattern of gene expression dictates. Such cells become terminally differentiated and are no longer capable of accessing the cell cycle. A minority of cells in multicellular organisms maintain a long-term undifferentiated status. These special cells, called *stem cells*, have a privileged status that allows them to evade differentiation and senescence.

2.3.1 Mitosis

Once a cell has performed DNA replication and has acquired the mass needed

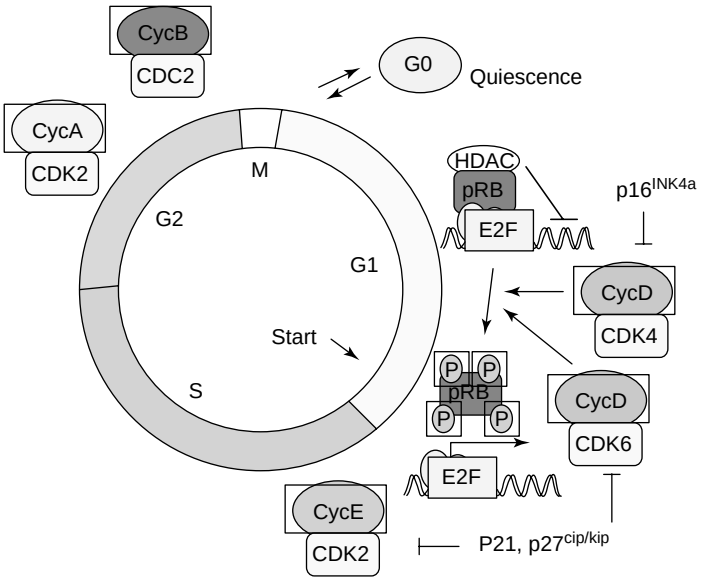

Fig. 2 The cell cycle. In proliferating mammalian cells, the cell cycle covers roughly 24 h. This growth cycle is depicted by the central ring, in which the amount of DNA in the cell is depicted (relative intensity of blue shading). The notable events of this cycle are the S-phase where DNA is replicated and the M-phase where chromosome segregation and daughter cell formation take place. These are separated by two gap phases – G1 and G2. Progression through the cell cycle is dictated by mitogenic signals from growth factors that interact with the cell cycle machinery, the critical components of which are shown here. Cyclin/CDK/C complexes drive the cycle. Kinase activities within the CDK proteins are responsible for regulating the downstream effects and the activity of the kinases is regulated by the cell cycle dependent expression of the cyclin proteins. The cell cycle machinery has many targets. For simplicity, the figure shows one major target, the cell cycle regulator Rb. Rb is phosphorylated by the cell cycle kinases and this activity releases Rb from complexes that suppress the transcription of critical proteins involved in proliferation, particularly DNA synthesis. Once Rb phosphorylation has activated this process, a cell is committed to pass through the cycle and on to mitosis. The activity of the cell cycle proteins is regulated by numerous other interactions such as those with the tumor suppressor proteins p21 and $p16^{INK4a}$. (See color plate p. xxix).

to generate two daughter cells, mitosis can begin. The onset of mitosis is driven by the cdc2/cyclinB complex. Major phosphorylation targets for this complex include histone H1 and lamin proteins. Modification of these proteins, together with numerous other modifications, drive chromosome condensation and disruption of the nuclear membrane. RNA synthesis is also switched off. Chromosome condensation begins late in G2; at this stage of the cell cycle, simple staining of DNA within the nuclei reveals the wormlike appearance of the condensing chromatin. The condensation process involves an axial retraction of the chromatin loops and

condensation along the chromosome axis. Chromosome condensation occurs during prophase and continues for about 30 min; as this process nears completion the nuclear membrane breaks down, marking the transition from interphase to mitosis. By late prophase, the chromosomes are fully condensed and are interacting with the mitotic microtubules. Microtubules function to adjust the position of the chromosomes until they are all aligned, at metaphase, across the center of the cell.

Chromosome condensation generates an elongated structure, which ensures that the kinetochores located within the centromeres of each chromosome are accessible to the microtubules that will eventually drag each daughter chromatid to the site where the two daughter nuclei will form. As the efficacy of this process is critical to the maintenance of genomic integrity, a mitotic checkpoint ensures that the chromatids are only able to separate once all kinetochores are attached via microtubules to the spindle. Prior to this point, a G2 checkpoint verifies that the replication process is complete. Once the kinetochores are fully engaged, the cohesin proteins that hold the sister chromatids together are released and the chromatids move from the metaphase plate toward the spindle poles; prior to this, topoisomerase II serves to resolve interlocked DNA loops that arise during replication. Microtubule and kinetochore-associated motor proteins, together with the dynamic properties of the microtubules themselves, provide the motive forces that reposition the daughter chromatids within the cell. The mitotic phase that involves migration of sister chromatids toward the spindle poles is called *anaphase* and this is followed by *telophase*, where chromosome decondensation begins. At this stage, the mitotic cell cycle machinery is destroyed and the new daughter cells form by cytokinesis.

2.3.2 Meiosis

In almost all cells, proliferation involves the duplication of a diploid chromosome set to form a tetraploid set, which is then separated, at mitosis, to form daughter cells that are once more diploid. Meiosis is a related process, which occurs in germ cells and serves to reduce the level of ploidy, so that on completing meiosis cells contain a single, haploid, chromosome set. Meiosis is a complex process, which not only reduces cell ploidy but also allows DNA exchange between homologous chromosomes to stimulate genetic diversity. Perhaps the easiest way to understand meiosis is to imagine the process as two consecutive mitotic events without an intervening S-phase. The cells that undertake meiosis are tetraploid. The first series of events – called *meiotic division one* – follow the same principle as mitosis but include a much longer prophase during which chromosome pairing and recombination take place; in human testes the duration of meiosis is roughly 25 days. The first meiotic prophase – during which time the nucleus remains intact – is traditionally divided into five sequential stages: leptotene, zygotene, pachytene, diplotene, and diakinesis. Homologous chromosome pairs synapse during zygotene and during pachytene, the resulting synaptonemal complex is thought to allow homologous recombination of the paired chromosomes to occur. The molecular details of this recombination process are poorly understood. The recombination complexes dissolve during diplotene to leave homologs that are attached at a small number of chiasma. The first meiotic division then proceeds through metaphase and anaphase and daughter nuclei are formed.

After a brief interphase, the second meiotic division follows according to the pattern described for a normal mitotic division. The overall process generates cells with a haploid DNA content and a small number of novel chromosomes that have arisen through recombination of chromosome homologues that were initially derived from the organisms' male and female predecessors.

3 Nuclear Structure

As the nuclear boundary might be assumed to define the nucleus itself a detailed understanding of the structure and function of the nuclear periphery is required to understand many features that define how nuclei work. In doing this, it is helpful to consider the nuclear periphery as being composed of three components:

1. The nuclear envelope is a double-membrane structure, with an outer membrane that is contiguous with the rough endoplasmic reticulum. Given this continuity, it is not surprising that areas of the outer nuclear membrane can be seen to be studded with ribosomes. The outer and inner lipid bilayers of the double envelope have common components but contain specific proteins that might be nuclear or cytoplasmic in location.
2. Over most of the nuclear surface, the two membranes of the nuclear envelope are separated by a fluid filled space that separates the membranes by about 50 nm. Over a small proportion of the nuclear surface, the inner and outer membranes are fused together. This fusion is seen to occur at structures called *nuclear pores*. These pores are the major routes of transit of any materials that must pass between the nucleus and the cytoplasm.
3. The nuclear face of the inner nuclear envelope is associated with a peripheral nuclear lamina. The lamina is composed of intermediate filaments of the lamin A/C and B proteins and is believed to play a role in the regulation of nuclear envelope structure and nuclear shape. Proteins that bind to the nuclear lamina also interact with chromatin so that through these adaptor proteins the lamina can also play a role in chromatin organization and the regulation of chromatin function.

3.1 Nuclear Pore Structure

A proliferating human cell has a few thousand pores and the number doubles in concert with doubling of chromatin during S-phase. Pores are disrupted during mitosis and the pore proteins are dispersed into about 12 pore subassemblies. Pores are rebuilt from these pore protein complexes during assembly of the new nuclear membrane at the end of mitosis.

In all eukaryotes, nuclear pores provide the gateway between the nucleus and cytoplasm. In mammalian cells, the pores are a megacomplex of about 150 MDa composed of roughly 50 different proteins (Table 3); the yeast pore has some 30 recognized pore proteins and it may be that some of the vertebrate proteins that have been classified as pore components are in fact transport factors. The general structure of a nuclear pore is shown in Fig. 3. The basic features of pore structure are:

1. The major structure of the core complexes is cylindrical with external

Tab. 3 Vertebrate nucleoporins.

Nucleoporin	*Likely location*	*FG repeats?*
vNup153	Nuclear basket ring	FG
vNup50	Nuclear basket	FG
vNup98	Nuclear basket	GLFG
vGle2 (Rae1)		
vNup160	Nuclear side	
vNup133	Nuclear side	
vNup107	Nuclear side	
vNup96	Nuclear side	
sec13		
vNup155	Both sides	
vNup188	Spokes/nuclear	
vNup205	Spokes/nuclear	
vNup93	Spokes/nuclear	
vNup62	Central transporter	FG
vNup58	Central transporter	FG
vNup54	Central transporter	FG
vNup45	Central transporter	FG
vPOM121	Integral membrane	FG
vgp210	Integral membrane	
vNup88 (84)	Cytoplasmic filaments	
vNup214 (CAN)	Cytoplasmic filaments	FG
vGle1	Cytoplasmic filaments	
vNup358 (RanBP2)	Cytoplasmic filaments	FG
RanBP1	Cytoplasmic filaments	
hCG1	Cytoplasmic filaments	

Note: See Vasu, S.K., Forbes, D.J. (2001) Nuclear pores and nuclear assembly, *Curr. Opin. Cell Biol.* **13**, 363–375 for further details.

dimensions of 120 × 80 nm. This complex has a striking eightfold symmetry as well as providing the structure of the channel between the nucleus and cytoplasm, which serves to fuse the inner and outer nuclear membranes. The eight spokes of the channel are held together to form a major scaffold of the pore.

2. These spokes surround a central transporter complex that is ~60-nm high and 40 nm in diameter.
3. The cytoplasmic face of the pore complex has 8 cytoplasmic filaments that extend about 50 nm into the cytoplasm.
4. The nucleoplasmic face of the pore has 8 filaments of 100 to 150 nm that extend into the nucleus. In some situations, these appear to associate at their distal end to generate a basket-like structure at the nuclear face of the pore complex.

3.1.1 Pore Function

Nuclear pores are the gateways between the nuclear and cytoplasmic compartments. Small molecules are able to pass through the pores by simple diffusion. Studies on diffusion kinetics suggest that a cylindrical diffusion channel of ~9 nm exists between the pore spokes and the central transporter complex. Proteins or RNA–protein complexes that are ~50 kDa and smaller are able to pass through this diffusion channel, but larger complexes are not and must rely on active transport systems.

The components of the vertebrate pore complex and their probable location in the complex are shown in Table 3. During active transport, the critical steps involve the initial interaction of the nuclear transport complex – that is, the cargo – with the pore filaments that extend into either the nucleus or cytoplasm. The pore proteins are involved in active transport as described below.

3.2 The Nuclear Lamina

During the 1960s, it was recognized that nuclei from mammalian cells had a protein network between the nuclear envelope and peripheral chromatin. This network or lamina is composed of a branched network of polymerized lamin proteins.

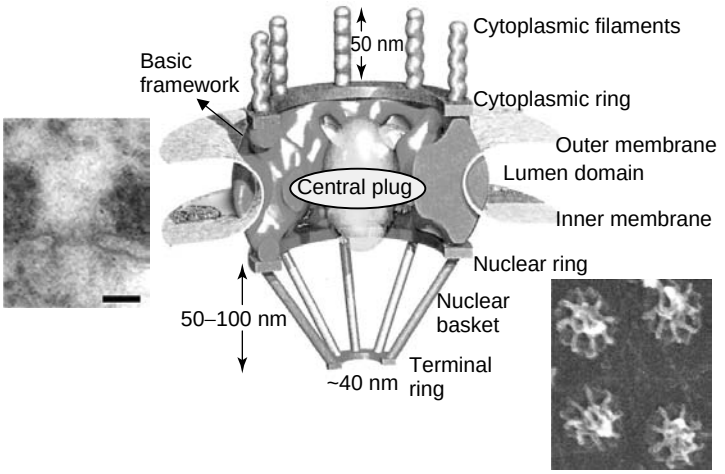

Fig. 3 The nuclear pore. The structure of the nuclear pore complex, shown here in cartoon form, was assimilated using a variety of electron microscopy (em) techniques. Details of the structure and the location of the constituent components are given in the text and in Table 3. The left inset (courtesy Francisco Iborra) shows a classical em thin section prepared in an epoxy resin, the bar is 100 nm. The nucleus occupies the upper half of the image. Note the densely stained heterochromatin lying against the nuclear membrane, the structure of the double membrane and the density of material at the pore. The right inset shows the inner nuclear face of an isolated nuclear membrane, visualized using a high-voltage scanning electron microscope. This shows the structure of the nuclear baskets and emphasizes the eightfold symmetry of the pore complex. The central image is reproduced from Izaurralde, E., Kann, M., Pante, N., Sodeik, B., Hohn, T. (1999) Viruses, microorganisms and scientists meet the nuclear pore, *EMBO J.* **18**, 289–296. With permission of Oxford University Press and the right inset is from Goldberg, M.W., Allen, T.D. (1992) High-resolution electron microscopy of the nuclear envelope – demonstration of a new, regular fibrous lattice attached to the baskets of the nucleoplasmic face of nuclear pores, *J. Cell Biol.* **119**, 1429–1440. With permission of Rockefeller University Press.

The lamins are classified as type V intermediate filaments (IFs), and like all intermediate filament proteins have a specific domain with an α-helical coiled-coil region flanked by nonhelical domains. The central coiled-coil domain is essential for assembly of the filaments. This process occurs initially through dimers and then through higher multimers, such that a mature IF of 10 nm in cross-section has ~32 monomers in its cross-section. During the assembly process, specific alignment of the IF protein monomers gives a repeat structure with ~23 nm periodicity. Major IF proteins in the cytoplasm of mammalian cells, such as vimentin, form very regular 10 nm (diameter) filaments that rarely branch. The nuclear lamins, in contrast, are highly branched and branches of smaller cross-section (i.e. suggesting filaments with fewer monomers in cross-section) are often seen. In addition to the peripheral nuclear lamina, nuclear lamin proteins

are also found within the nuclear interior. *In vivo* studies support the view that these, like the peripheral proteins, are assembled into structures that have rather slow rates of exchange. However, it is still unclear if this represents a lamin filament network that pervades the nucleus or local aggregates of lamin proteins that might serve a function that does not demand a contiguous "nucleoskeleton."

3.2.1 The Lamin Genes

The lamin proteins represent a small protein family that has increased in complexity during metazoan evolution. Simple eukaryotes, such as the yeast *S. cerevisiae*, do not have lamin proteins; perhaps the lamina evolved during the transition from a closed mitosis to open mitosis. The simple multicellular organism *Caenorhabditis elegans* has a single lamin gene, which is expressed in all somatic cells. Mammals have three lamin genes (*LMNA, LMNB1*, and *LMNB2*) that are processed to yield 7 isoforms. RNA processing – alternative splicing – of the *LMNA* primary transcript generates four A-type lamins – lamins A, AΔ10 (a splice variant without exon 10), C, and C2. Three B-type lamins, called *lamins B1–B3*, are encoded by the *LMNB1/2* genes. All vertebrate cells express at least one variant of the B-type lamin proteins. The A-type lamins, in contrast, are expressed primarily in differentiated cells so that patterns of lamins A, AΔ10 and C expression are developmentally regulated. Lamins C2 and B3 are found only in germ cells.

3.2.2 The Lamin Proteins and Filament Assembly

Like other members of the intermediate filament class of proteins, lamins have a central rod domain that consists mainly of heptad repeats. The coiled-coil structure that forms from these repeats drives the interaction of two monomers to form a coiled-coil dimer. This form is the basic building block for lamin filaments. Higher-order filaments form as a consequence of lateral interaction between the rod domains of the dimers. The rod domains are flanked by non-α-helical amino-terminal and carboxy-terminal domains. The latter contains a nuclear localization signal. Lamin proteins are also subjected to posttranslational modifications such as phosphorylation and most importantly are modified by isoprenylation, a modification that targets the lamin filaments to the nuclear envelope. This modification takes place on a special CaaX motif (not found in lamin C). Following cleavage after the cysteine residue the new terminal amino acid is modified by isoprenylation and methyl esterification.

3.2.3 Lamin Function

Various studies identify the structural features as a major role for the lamin proteins. It is notable, for example, that nuclei maintain their shape even after most of the nuclear contents have been removed. This and many other experiments suggest a role for the lamins in maintaining the mechanical stability of nuclei. The importance of maintaining a particular nuclear shape is not clear. Perhaps more important functionally is the possibility that the nuclear lamina can potentially define nuclear sites that have particular functional roles. Many proteins are known that act as adaptors that link the lamina to the nuclear envelope and components within chromatin to the lamina. Heterochromatin – which is transcriptionally inert – is commonly associated with the nuclear periphery and this

location might have important implications for the regulation of gene expression. A-type lamins bind the retinoblastoma protein and other transcriptional repressors and may play important roles in regulating gene expression through spatial organization.

Other experiments support the fascinating possibility that nuclear lamins might be involved in the processes of DNA and RNA synthesis. The experimental evidence involves the introduction into cells – by microinjection – of bacterially expressed lamin proteins that have amino-terminal deletions. The microinjected proteins are engineered so that they interact with the endogenous proteins to disrupt stable filament formation. It has been reported that when the lamin filaments are disrupted in this way, DNA synthesis and RNA polymerase II-dependent RNA synthesis are also disrupted. One explanation for this is that lamin filaments or perhaps smaller local aggregates of the lamin proteins play some structural role within the active centers of DNA and RNA synthesis.

3.2.4 Lamin-associated Proteins

Many proteins are known to interact with the lamins to influence lamin function. These might be classified into two broad groups. Integral proteins of the inner nuclear membrane are the first group and include various isoforms of lamin-associated protein 2 (LAP2), LAP1, lamin B receptor (LBR), emerin, otefin, MAN1, Nurim, nesprin, RING finger-binding protein, A-kinase anchoring protein 149, and p19 (an isoquinoline-binding protein that is also found in the endoplasmic reticulum). The second group contains proteins that associate with the lamins but are not membrane components. This group includes germ cell-less (GCL), young arrest (YA), PP1 phosphatase and the transcription factor Oct1.

The nuclear lamina is known to play a role in chromatin organization and the lamins can be shown to bind DNA elements called *nuclear matrix* and scaffold attachment regions (S/MARs) *in vitro*. The LAP2 isoforms LAP2α and LAP2β bind to chromatin. Chromatin binding involves two motifs on the LAP proteins, a chromatin binding domain and a LEM domain. The LEM domain is a region of about 40 amino acids that is also found in emerin, otefin and MAN1. The LEM domain interacts with other chromatin-associated proteins such as BAF – and might then influence chromatin architecture and function.

3.2.5 Lamin Mutants

Knockout mice for the lamin A/C gene has been developed to study lamin function. Null mice appear normal at birth, but soon after develop severe growth defects. At about one month after birth the null mice display reduced mobility and a stiff walking posture. No mice survive beyond two months. From histological examination, it is clear that the affected mice have skeletal and cardiac muscle wasting and reduced white fat. Fibroblasts from the lamin-A knockout mice show abnormal nuclear blebs, which appear to result from a dramatically reduced amount of lamin proteins. The lamin-associated proteins LAP2 and NUP153 (a nuclear pore complex protein) are also significantly reduced. An abnormal distribution of emerin is also seen, together with some uncharacteristic features of heterochromatin organization. These features of knockout mice are reminiscent of symptoms seen in human patients with Emery Dreifus muscular dystrophy (EDMD). Indeed, mutations in *LMNA* have recently

been show to be causative for the autosomal dominant form of this disease. About 50 mutations within *LMNA* have been described that lead to disease states such as EDMD and closely related disorders. Interestingly, lamin mutations seem not to affect all cell types, and while knockout mice die soon after birth, it is notable that individuals with lamin mutations live into adulthood. This may not be surprising as natural selection allows the accumulation of mutations that are not lethal. It is interesting to speculate that mutant lamins weaken the structure of the lamina to yield fragile nuclei. These might be especially susceptible to damage in cells such as muscle and tendon that are put under significant stress during muscle contraction.

3.3
Biogenesis of the Nuclear Envelope During Cell Proliferation

The process of cell division inevitably demands that a series of profound structural changes will be required to generate two genetically identical daughter cells during mitosis. Prior to mitosis, chromosomes are dispersed throughout the nuclear interior. Before the new daughter cells can form, the chromosomes must condense so one of each of the daughter chromatids (i.e. the two equivalent products of DNA replication for each chromosome) can pass to the daughter cells. The cell cycle proteins, and critically at this part of the cell cycle cdc2 and cyclinB, predominantly through changes in phosphorylation, drive changes in chromatin structure that lead to chromosome condensation. Other changes lead to a concomitant dissolution of the nuclear periphery. The critical events needed to achieve this involve changes in the structure of the nuclear lamina. Changes in phosphorylation of the lamin proteins cause disruption of the lamin filament network. Loss of the chromatin stabilizing properties of the lamin-chromatin interactions allows an increase in chromatin mobility so that chromosome condensation is not impeded. Moreover, as the lamina disrupts, stabilizing contacts between the lamina and nuclear envelope are lost. The new dynamic capabilities of the nuclear envelope then allow the membrane to merge with the envelope networks of the endoplasmic reticulum. During this process the nuclear pores are disrupted into a small number of discrete pore subassemblies. This dissolution of the nuclear membrane and the concomitant condensation of chromosomes marks the early stages of mitosis. While this is by far the most common course of events during a eukaryotic mitosis it should be noted that some simple model eukaryotes – such as the yeast *S. cerevisiae* – have a closed mitosis where the nuclear membrane remains ostensibly intact during chromosome separation.

During mitosis, once the chromosomes have migrated to the poles of the cell, a central constriction in the cell membrane leads to the formation of the two new daughter cells. The cell cycle machinery now drives nuclear reconstruction. In essence, the new nuclei form through a program that is the reverse of the events that preceded the onset of mitosis. As the chromosomes begin to expand, patches on new nuclear lamina are thought to become associated with the chromatin and these then engage the endoplasmic reticulum, which is extruded to surround the chromosomes, eventually to form the new nuclear envelopes. Lamin B appears to associate with chromatin early during the assembly process while lamin A seems to remain in the cytoplasm until nuclear membrane assembly is complete – when

it is imported through the nuclear pores. Assembly and maturation of the lamina continues for an hour or two into the G1 phase. The fluid properties of the nuclear membrane eventually lead to formation of a complete double envelope. While this process is ongoing, nuclear pores are assembled into the envelope. Soon after cell division, a fully operative nuclear envelope is formed.

3.4 Nuclear Transport and Karyopherins

The fact that eukaryotes have evolved to confine the genetic material within a specialized organelle – the nucleus – demands that mechanisms must exist to regulate the passage of information between the nucleus and cytoplasm. This is self-evident as the products of gene transcription must pass to the cytoplasm for translation and all the proteins that control and direct gene function must pass in the opposite direction, moving into the nucleus from the cytoplasm where they are synthesized.

This nuclear-cytoplasmic traffic takes place through the nuclear pores. The detailed structure of these pores has been elucidated and, as we saw earlier, provides a structural framework that can be used to elaborate the mechanisms of nuclear pore-mediated transport. The basic principles of transport are well known, though some details remain to be confirmed. The first point to emphasize is that nuclear pores do not provide an absolute barrier between the nucleus and cytoplasm. In fact, small molecules and proteins that have a size below ~50 kDa are able to diffuse through the pores. Any proteins that are larger than this, and protein nucleic acid complexes undergo either nuclear import or export using specific transport mechanisms. This active process involves nuclear localization and export signals that are recognized by the import/export machinery. The nuclear localization sequence of the SV40 virus large T antigen was the first to be analyzed in detail. In this protein, the amino acid sequence: Pro-Pro-Lys-Lys-Lys-Arg-Lys-Val directs nuclear accumulation of the protein. In other proteins, it is not uncommon to find bipartite localization signals where two short lysine and arginine-rich regions are separated by ~10 intervening amino acids.

The molecular details of the protein transport process are by far the best understood. The key components in this process are the karyopherin proteins that operate as shuttling transport receptors. The simple eukaryotes *S. cerevisiae* is known to have 14 karyopherin proteins and at least 20 have been described in human cells. Karyopherins can be further subdivided into importins and exportins. As the names imply, these are largely dedicated to either the nuclear import or export pathways; though examples are known that can transport different cargoes across the nuclear pores in both directions.

3.4.1 The RanGTP/GDP Cycle

Movement of a molecule across the nuclear pore will generally take place against a concentration gradient and involves ATP hydrolysis. The basis of this process relies on the way in which the karyopherin–cargo complexes interact with the pore proteins and subsequently pass through the pore to be released on the other side. The behavior of importins and exportins during this process is regulated by the small GTPase Ran. More specifically, substrate binding and release by the karyopherins is regulated by the asymmetrical distribution of the GTP- and GDP-bound states of Ran.

These states establish a RanGTP gradient across the nuclear pore, in which Ran inside the nucleus is predominantly GTP-bound, while the one outside is mostly in the GDP-bound state. This gradient is manipulated by a variety of accessory factors that together regulate transport. The most important feature of the RanGTP/GDP cycle is that GTP loading occurs in the nucleus while hydrolysis occurs in the cytoplasm. The Ran guanine nucleotide exchange factor (RanGEF) called RCC1 catalyzes GTP loading, while RanGTP hydrolysis is performed by the RanGTPase activating protein RanGAP together with accessory proteins RanBP1 and RanBP2. RanGAP and RanBP1/2 are restricted to the cytoplasm. This asymmetrical distribution of the Ran regulators establishes a RanGTP gradient such that each molecule of Ran that enters the nucleus will be rapidly converted to the GTP-bound state, while RanGTP moving out of the nucleus will be converted to RanGDP on entering the cytoplasm. The magnitude of this gradient can be estimated using Ran biosensors that use fluorescence resonance energy transfer (FRET) to probe structural changes during transport. This approach suggests that the concentration of free RanGTP in the nucleus is at least 200-fold greater than that in the cytoplasm.

The RanGTP gradient provides the position information that drives nuclear transport. During import, the importins will bind to their substrates in the absence of Ran and after translocation release the substrate on binding RanGTP in the nucleus. Here, the RanGTP operates as a kind of exchange factor and the importin–RanGTP complex that forms is recycled back to the cytoplasm. During export, the exportins only bind to their cargoes in the presence of RanGTP. The RanGTP gradient ensures that this occurs almost exclusively in the nucleus. On the cytoplasmic side of the pore complex, the exportin/RanGTP/cargo complex decays following RanGTP hydrolysis that is driven by RanGAP and RanBP1/2. RanGDP returns to the nucleus in a complex with a Ran importer protein called NTF2. Once inside the nucleus, RanGDP is once again converted to the GTP-bound form, to release NTF2 so that a new cycle of transport can begin.

3.4.2 RNA Export Pathways

In contrast to the well-characterized mechanisms used to regulate protein transport through nuclear pores, many details of the mechanisms for RNA transport remain sketchy. The importance of this should not be underestimated as it clearly represents a critical point of regulation in the control of gene expression. In addition, it is well known that many independent pathways regulate the transport of different classes of RNA–protein complexes – this could be potentially very complex allowing the differential regulation of specific classes of RNA transcript in response to a variety of developmental and growth regulatory cues.

Basic details of the primary mechanisms that control mRNP export have been elucidated. The molecular interactions that underlie this process are dependent on events that take place at the site of mRNA synthesis. The critical components are small proteins that are part of a much larger complex that is deposited on the mRNA during splicing. The complex called the – 20/24 complex because of its location at about 20 nucleotides upstream of the exon–intron junction contains a number of proteins, but includes the protein Aly. Aly interacts specifically with TAP/mex67, which serves as a major export complex for mRNAs. It is important to recognize that

mRNA export is coupled to many of the proceeding steps of gene expression. In broad terms, the pathway that delivers mature mRNAs to the cytoplasm is influenced by promoter strength, the RNA polymerase complex that performs synthesis, transcriptional elongation, as well as down stream processing events such as RNA splicing and polyadenylation. A significant aspect of this pathway reflects the fact that mRNAs are retained at the transcription site until the required processing events have been performed. This is likely to be part of the "quality control" process, which ensures that only authentic mature mRNAs are exported to the cytoplasm.

Apart from the general export factors, many proteins have been described that interact with different classes of RNA to regulate RNA export. Certain factors regulate the export of different classes of mRNA and others are essential for the export of ribosomes and RNA–protein complexes that contain RNAs transcribed by RNA pol III.

3.4.3 Nuclear Transport – Facilitated Diffusion

We have looked in some detail at the molecules involved in nuclear transport without concern for the mechanisms that drive the transport process itself. It was noted above that the nuclear pore is essentially "open" to small molecules. Transport that is mediated by karyopherins is perceived as being a form of directed diffusion – or facilitated translocation. The critical question to address here is how do pores exert a high degree of selectivity on the nuclear transport receptors while excluding proteins that do not contain the appropriate receptor signals. A critical first step requires an association – binding – of the transport receptor to target molecules within the pore complex. These interactions have been characterized in detail. What is not clear, however, is why this primary interaction would inevitably lead to the cargo being directed through the pore. Models of facilitated translocation suppose that selectivity – gating – is achieved through interactions between the transport complex and particular protein motifs in the pore proteins – called *FG-repeats*. The FG-repeats are proposed to provide a sort of hydrophobic plug that severely reduces the efficiency with which molecules that are not destined for transport are able to pass through nuclear pores. The receptor/cargo complex, in contrast, by virtue of its structure and interaction with other pore components is able to diffuse through this region of the pore. Though the details by which selectivity is achieved remain to be refined, changes in pore permeability following treatment with organic solvents are consistent with a role for a hydrophobic meshwork (provided by the FG-repeats) inside the pore channel.

3.5 Nuclear Architecture and Nuclear Compartments

3.5.1 Nuclear Compartments

In mammalian cells, the nuclear membrane defines two major cell compartments – the nucleus and cytoplasm – that have quite distinct structural characteristics (Fig. 4). The cytoplasm gives an impression of being highly structured with many discrete organelles such as mitochondria and lysosomes and has other regions that contain dense clusters of ribosomes (each ribosome is 25 nm in diameter) performing protein synthesis. The nucleus is also structured, but the basic principles that define nuclear structure

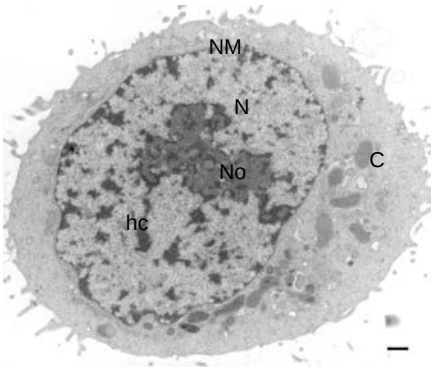

Fig. 4 The mammalian cell. Proliferating HeLa cells (this is a transformed human cell line) were fixed, stained, and processed using standard electron microscopy techniques. Thin sections were then prepared and inspected. This typical cell section shows a central nucleus (N) – typically of about 10 μm diameter in an equatorial section – which is separated from the cytoplasm (C) by a double nuclear membrane (NM). The nuclear membrane is studded with a few thousand nuclear pores that allow molecules to pass between the two major cell compartments. Note that the cytoplasm contains a large number of membrane-bound organelles – the most conspicuous of these, the mitochondria, are dark staining and typically about 1 μm in diameter in this example. The nucleus does not contain any membrane-bound subcompartments. Even so it is highly organized. In this type of image, the most obvious substructure is the nucleolus (No). In this example, a single large nucleolus occupies the center of the nucleus. The nucleolus is densely staining and has three distinct subcompartments: the most densely staining regions are called the *dense fibrillar component*; these structures are generally crescent shaped and a few hundred nm in size. The dense fibrillar components represent the nascent transcripts that surround the much paler fibrillar centers. The processed ribosomal RNAs are assembled into ribosomal subunits in the granular component, which comprises the bulk of the nucleolus. Dark staining clumps of material that lie against the surface of the nucleolus, the nuclear face of the membrane and are scattered throughout the center of the nucleus are dense heterochromatin (hc). The remainder of the nucleus has an amorphous appearance and little obvious structure. This nucleoplasmic region contains the transcribed genes in euchromatin and is composed of a mixture of this open chromatin and the interchromatin material, which is predominantly RNA and protein. The bar is 1 μm.

are quite different. The nucleus does not contain subnuclear organelles – by definition, organelles are enclosed by a membrane. Instead, the nucleus has numerous discrete zones or compartments where particular functions are performed. The nucleolus provides an excellent example of this organizational theme. Indeed, it is not unusual for the nucleoli to be the most conspicuous interphase structure in a mammalian cell.

3.5.2 The Nucleolus

Nucleoli are sites of ribosomal RNA synthesis and processing and also centers of biogenesis for the ribosomal subunits. Clearly defined subcompartments within nucleoli reflect the centers where these different activities are performed. Mammalian cells generally contain 1 to 5 nucleoli; the number will usually reflect the proliferative status of a cell and might be used by histologists as an indicator

of proliferation in cancer diagnosis. Nucleoli range in size from 0.5 to 5.0 μm and are assembled around the active ribosomal RNA genes. Each diploid human cell has about 300 copies of the ribosomal RNA locus. Each locus contains a single transcription unit of 13.3 kb, which is processed posttranscriptionally to generate 3 RNA molecules – the 28S, 18S and 5.8S (in vertebrates) rRNAs – which, together with 5S RNA are the structural RNAs that dictate ribosome assembly and function. The genes are located on 5 human chromosomes at the subtelomeric regions of c13, 14, 15, 21, and 22, where they are found in short tandem arrays – typically as clusters of 3 to 5 rDNA repeats. In most proliferating human cells, only about one-third of the total cell complement of ribosomal genes is active at any time. In proliferating cells, nucleoli form as ribosomal RNA synthesis is reestablished, soon after cell division is complete. Ribosomal RNA genes retain associated transcription factors throughout mitosis and these provide efficient targets for the assembly of transcription complexes at the onset of interphase. As transcription recovers, the components involved are believed to interact to form the characteristic nucleolar morphology. Each active center is defined by a fibrillar center (FC) that has a distinct fibrogranular appearance and is rich in the machinery that is required to generate rRNA. The surface of each FC is coated with a much denser dense fibrillar component (DFC) that contains the nascent transcripts and their associated proteins. The transcribed rRNA genes lie along the interface of the FC and DFC and a single FC will typically have 3 to 5 associated active genes.

Ribosomal RNA genes are the most active transcription units in mammalian cells. It takes about 3 min to complete synthesis of the primary transcript, and polymerases are engaged onto the promoter with such efficiency that each active gene has 100 to 120 engaged RNA polymerase I complexes. In order to achieve this level of synthesis, new initiation events must occur at roughly 1.5-s intervals; the residence time of the major transcription factors in the FC is ~5 s and the residence time for RNA polymerase I on the promoter is roughly 1 s. The primary transcripts are processed to generate mature ribosomal RNAs. This is an extremely complex process that involves a number of specific cleavage events and RNA modifications, typically methylation. These posttranscriptional events are directed by at least 100 special small nucleolar RNA–protein complexes (snoRNPs). The mature rRNAs then associate with the requisite ribosomal proteins and begin their passage from the synthetic center to the cytoplasm. For the early stages – about 1 h – of this migration they remain nuclear and occupy the nucleolar granular component (GC), a morphologically punctate or granular zone that surrounds the active rRNA transcription centers. Eventually, the mature 40 and 60S ribosomal particles pass from the nucleolus and through the nucleoplasm to the nuclear periphery from where they exit the nucleus through nuclear pores.

3.5.3 Nuclear Bodies and Domains

Other nuclear structures have been described as nuclear bodies and domains. Many cell types have a small number (typically 1–5) of coiled or Cajal bodies, which are generally spherical, measure 0.3 to 1 μm across and in cross-section, appear as a network of coiled fibers. Coiled bodies are found in both animals and plants and contain many different proteins, including p80 coilin (the diagnostic

marker), various transcription factors, as well as small ribonucleoproteins. These bodies are not transcriptionally active, and are thought to play a role in the biogenesis of small nuclear RNA–protein complexes (snRNPs) and in the nuclear trafficking of snRNPs and snoRNPs. RNP complexes that are essential for mRNA splicing (U1, U2, U4/U6, and U5 snRNPs), rRNA processing (U3 and U8 snoRNPs) and histone mRNA end formation (U7 snRNPs), all localize to the Cajal bodies. Possible roles for these structures are suggested by the fact that they are often seen to associate with specific internuclear sites, such as the histone loci and U1/U2/U3 gene cluster.

PML bodies are defined by their content of promyelocytic leukemia (PML) protein. They are similar in size to coiled bodies but are usually more abundant – mammalian cells commonly have 10 to 20 PML bodies. PML bodies have also been referred to as the ND10 compartment, PODs (PML oncogenic domains) and Kr bodies. In addition to the PML protein, these structures also contain proteins such as: Sp100; SUMO1; HAUSP; and CBP. Though the precise role of PML bodies is unclear, it has been suggested that they will play some role in transcriptional regulation. Interestingly, these domains are common targets of viral genomes during the early infectious stages in mammalian cells.

Pc-G bodies that contain polycomb group proteins such as RING1, BMI1, and hPc2 have been observed to associate with pericentromeric heterchromatin. Gems (gemini of Cajal bodies) generally lie adjacent to the Cajal bodies and are characterized by the presence of the survival of motor neurons gene product (SMN) and an associated factor Gemin2. Other sites have been described that might be classified as transcription superfactories. The OPT (Oct1/PTF/transcription) domain provides a good example. Immunofluorescence shows that in HeLa cells, the transcription factors Oct1 and PTF are found in 1 to 3 large domains (diameter ~1.5 μm) that appear during G1 phase and disappear in S-phase. Each OPT domain typically contains 2 to 3 transcription sites where BrUTP is incorporated into nascent transcripts, as well as RNA polymerase II, TBP, and Sp1. Some chromosomes, notably chromosome 6, are found to be closely associated with the OPT domains. But while nuclear bodies have raised much interest, the extent to which they define essential functional compartments remains a matter of debate. The Cajal bodies exemplify this, as these structures can be disrupted in living cells or transgenic animals without obvious detriment to the hosts. Hence, the bodies, as structures, might not be absolutely crucial to the functions that are performed by components that are normally found within them.

3.5.4 Nuclear Speckles

Nuclear compartments such as nuclear bodies are not membrane-bound. Instead, they appear to persist because of the "self-assembly" characteristics of their components and architectural constraints imposed by nuclear and chromosome structure (see below). Nuclear speckles/interchromatin granule clusters (IGCs) provide an excellent example of this architectural theme. (Speckles were defined using light microscopy and immunofluorescence and were later found to be equivalent to the IGC that had been visualized using electron microscopy.) This compartment is especially rich in proteins required for RNA splicing – such

as the SR protein, family, which is defined by a specific serine and arginine rich motif – and occupies the interchromatin spaces where the proteins may accumulate to form large, clustered aggregates. Once formed, these aggregates – the speckles or IGCs – are stable in that they occupy a particular area of nuclear space over many hours. This spatial stability is seen even though the components within each site are in a continual state of flux. However, the IGCs are also plastic and can change dramatically under conditions in which RNA synthesis is inhibited and the need for splicing is lost; if cells are treated with the transcriptional inhibitor α-amanitin, the speckles lose their complex surface texture and collapse locally to form more dense spherical aggregates. This and the fact that highly active genes are commonly found at the periphery of speckles imply some functional link. Interestingly, cells that overexpress the serine-arginine (SR) protein kinase cdc2-like kinase (Clk)/STY exhibit disrupted speckles. This kinase hyper-phosphorylates a major IGC component, the nuclear SR proteins, which, as a result move out of the IGCs to diffuse throughout the nucleus. Cells without IGCs show no obvious defects in short-term RNA synthesis. However, under the same conditions RNA FISH to a specific transcript shows that splicing, assessed using a probe to an exon–exon junction, is severely impaired. This implies that splicing factors originating from a diffuse nucleoplasmic pool are not competent to perform splicing in vivo and that the organization of the IGC compartment is critical to this process.

3.5.5 Active Sites – Synthetic Factories

Transcription factories It is generally accepted that mammalian nuclei are highly compartmentalized and that the structure of different nuclear domains is implicit to their function. As we have seen, nucleoli provide the outstanding example of a nuclear compartment that serves a specific function. Nucleoli are dedicated to the efficient synthesis of a single primary RNA transcript and the products generated from this transcript by RNA processing are fundamental to ribosome assembly and function. Within nucleoli, morphologically distinct compartments define the active centers, where, as many as 500 engaged polymerase complexes are found within a typical FC/DFC complex. In terms of structure, the nucleolus represents a nuclear compartment that is dedicated to the synthesis and processing of a specific product.

This might lead us to ask if the same principle applies to transcripts that are synthesized and processed throughout the nucleus. Given the relative sizes and complexity of the corresponding sites, it is not surprising that the architecture of the nucleoplasmic transcription centers is less clearly defined. Classical electron microscopy techniques have shown that gene expression generally occurs at the borders of condensed chromatin in association with perichromatin fibrils. The ability to label the nascent RNA (Fig. 5) using BrUTP *in vitro* or BrU *in vivo* has shown that a proliferative mammalian cell has some 500 to 1000 active sites of RNA polymerase II activity for each haploid chromosome set. As cells have roughly 10 times the number of active RNA polymerase II holoenzyme complexes, each active center must represent a nuclear compartment in which the transcripts from groups of genes are generated and processed together. This spatial coordination of the different steps required to produce mature mRNAs at specific nuclear sites forms the basis of the concept of transcription "factories" where transcripts

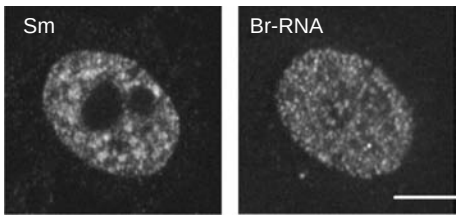

Fig. 5 Nuclear compartments and transcription. HeLa cells growing on glass were permeabilized in an isotonic buffer and transcription performed for 15 min in the presence of BrUTP. Samples were fixed and immuno-labeled for Br-RNA containing sites (right) using a red fluorescence dye and sites containing the auto-immune antigen Sm (left) using a green fluorescent dye. The two colors were recorded in a single 700 nm section of the nucleus using a laser scanning confocal microscope-gray-scale images of the two channels are shown. Note the classical "speckled" distribution of the Sm protein and the complexity of the transcription sites; the two areas that do not stain with Sm are nucleoli. The bar is 5 μm. See Pombo, A., Cook, P.R. (1996) The localization of sites containing nascent RNA and splicing factors, *Exp. Cell Res.* **229**, 201–203 for details. Published with permission of Academic Press.

are polymerized, processed and assembled into the required mRNA–protein complex before being released for export to the cytoplasm. For a typical transcript, events occurring at the transcription site take roughly 15 min to complete.

A likely corollary of this arrangement is that the transcription centers form a functional nuclear compartment that is spatially structured by an association with a component of the interchromatin compartments – perhaps a nucleoskeleton. Hence, during transcription the chromatin loops must be locally dynamic as RNA synthesis forces the gene to progressively associate (i.e. from promoter to 3′ end) with the synthetic center. The idea that genes might be associated with the active center during synthesis and displaced from it thereafter defines a type of spatially determined "transcription cycle" within which the interplay between chromatin dynamics and promoter bound factors will dictate levels of gene expression. The dynamic behavior of transcription factors will also be central to this control. Interestingly, some classes of transcription factor can be shown to associate with chromatin only transiently, whereas others are more stable, and might even remain bound to chromatin throughout mitosis. Factors such as TBP (TATA binding protein), which fall in the second category, are ideal candidates to mark active genes and regulate the association of these genes with transcription sites.

Replication factories DNA replication provides an interesting insight into how nuclear structure and function can be linked. We will see that DNA foci provide the best candidates for any structural units of chromosome structure (below). However, it is also important to recognize that the organization of these structures might be central to both the replication process and how the replication program proceeds. DNA foci contain small groups of replicons, which operate as spatially stable replicon clusters and provide targets for the assembly of replication factories. Like transcription factories, these arise through DNA–protein and protein–protein interactions that generate nuclear sites in which DNA synthesis can be performed with

optimal efficiency. These active centers are also spatially structured through their association with the nucleoskeleton. The modular structure of replicon clusters also provides a mechanism for replication to spread in a predetermined way from one bank of replicons to the next until the S-phase is complete.

Aspects of crosstalk between nuclear structure and function are of particular interest when considering how the S-phase program might be established. The importance of nuclear space in setting up the S-phase program is most evident in *S. cerevisiae*, in which the subnuclear position of replication origins in early G1 phase of the cell cycle determines whether origins are used either early or late in S-phase. In mammalian cells, the early replicons are selected during the first one-third of G1 phase of the cell cycle, prior to the commitment to engage the cell cycle. Two critical events appear to define the replication program and these are believed to depend on the complex interplay of chromosome structure, chromatin organization, and perhaps even global nuclear architecture. Replicons that are activated at the onset of the S-phase and most probably the downstream replication program are established during the first 2 h of G1 phase of the cell cycle. This "temporal decision point" coincides with the period when chromosomal territories reestablish their interphase structure and transcriptional potential. The process of defining an S-phase program is distinct from the process that establishes active origins of DNA replication. This second process also occurs in early G1 phase in mammalian cells and is thought to involve the binding of the origin recognition complex (ORC) to particular chromosomal sites.

In human cells, the assembly of replication factories begins late in the G1 phase of the cell cycle and appears to take about 15 min. The cell cycle machinery is responsible for the cues that initiate both the assembly process and initiation of DNA synthesis. The replication machinery is recruited to appropriate DNA foci – in mammalian cells only about 10% of possible origins are activated at the onset of the S-phase – through an association with ORCs on the DNA. All of the components needed for synthesis are recruited to the factories as a consequence of protein–protein interaction. For many of the replication components, factory-targeting sequences have been defined. As many replicons are replicated in each factory (early S-phase DNA foci contain ~5 replicons and later ones many more) the structure of the complex dictates that chromatin must move into the active center, which itself is static. This view can be confirmed using GFP-PCNA to analyze the spatial dynamics of replication sites in living cells. This demonstrates that the foci, once assembled, tend to be spatially stable, with the PCNA (Fig. 1) molecules that are assembled into the synthetic complex during initiation remaining at the active site until synthesis is complete. Once replication at a site is complete, the active center will disassemble. Often a new active center is then seen to assemble *de novo*, at an adjacent site. This principle underlies a proposed domino effect for a replication program, in which local changes in chromatin architecture dictate how replication proceeds until the S-phase is complete. In mammalian cells, S-phase lasts for about 10 h, while DNA synthesis from a typical origin will continue for about 1 h. As S-phase proceeds, the genome is reproduced in a predictable order so that defined chromosomal regions are duplicated at particular times during early, mid- or late S-phase (Fig. 6).

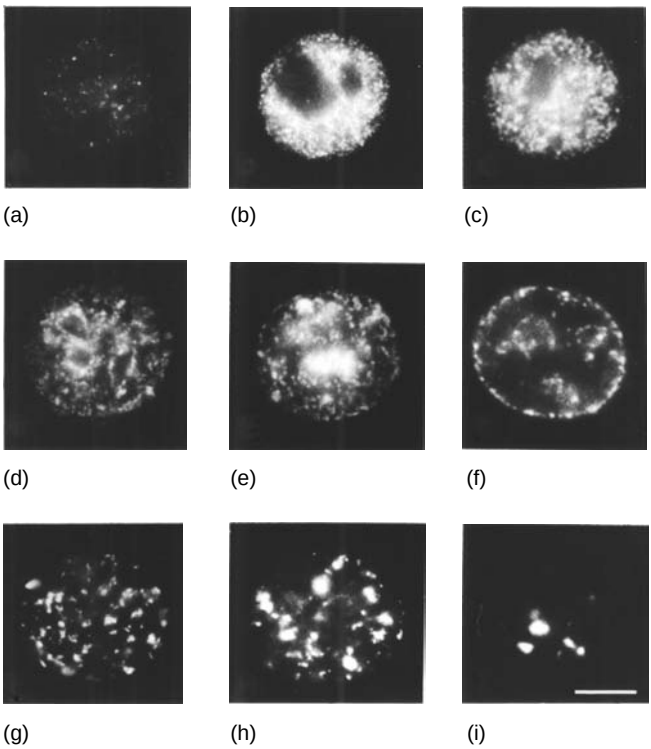

Fig. 6 The S-phase program. During S-phase, different classes of DNA elements are replicated at specific times. Chromatin with the majority of transcribed genes is replicated over the first ~4 h of the S-phase. During this period active sites of DNA synthesis are in discrete foci dispersed throughout the nuclear interior (a–c). These are designated early or type 1 replication patterns. During mid S-phase (d–f), replication switches to peripheral inactive chromatin. These mid S-phase patterns are also referred to as type 2 (d–e) and type 3 (f) patterns. During late S-phase (g–i) the replication of extended blocks of constitutive heterochromatin occurs throughout the nuclear interior. These late S-phase patterns are also referred to as type 4 (g–h) and type 5 (i) patterns. Encapsulated HeLa cells were synchronized at the onset of S-phase and replication sites labeled at 1 h intervals for 10 h. Images shown are replication sites labeled in permeabilized cells – using biotin-dUTP. (Reproduced from Hozak, P., Jackson, D.A., Cook, P.R. (1994) Replication factories and nuclear bodies: the ultrastructural characterization of replication sites during the cell cycle, *J. Cell Sci.* **107**, 2191–2202. With permission of the Company of Biologists. The bar is 5 μm.)

3.6 Chromosome Territories

It has been known for many years that chromosomes occupy discrete regions within interphase nuclei. Over the past 20 years, the idea that interphase chromosomes occupy discrete nuclear "territories" has been confirmed using a number of techniques. A major advance in the analysis in nuclear architecture came from the development of fluorescence in situ hybridization (FISH) techniques for "chromosome painting." This process uses the unique sequences from purified chromosomes as probes to identify the location of particular chromosomes in fixed nuclei. The probe DNA is tagged with a modified DNA precursor analog – such as biotin-dUTP – prior to hybridizing to the fixed cell (with denatured target DNA) and is subsequently visualized using antibodies to biotin and secondary antibodies coupled to a fluorochrome that can be visualized by light microscopy. This technique confirmed that chromosome territories from mammalian cells appear to have distinct borders with little mixing of chromatin from adjacent territories. The territories themselves appear to be subcompartmentalized so that the space occupied by individual chromosome arms and even R and G bands within these arms might be defined within individual territories. The general principle of chromosome territory organization can be confirmed using a variety of DNA labeling techniques, using precursors that can be visualized in both fixed and living cells (Fig. 7).

The existence of chromosome territories begs the question: do interphase chromosomes occupy specific nuclear locations that could influence chromatin function? The answer to this appears to be partly yes and partly no. Territories do not generally occupy specific positions. However, gene activity does appear to influence chromosome location, such that chromosomes with a high density of active genes tend to be located toward the nuclear center, whereas those with a low density tend to lie closer to the nuclear periphery. The smaller chromosomes also show a tendency to occupy the nuclear center; note that for those chromosomes with ribosomal RNA gene clusters, location is determined by their association with nucleoli, which generally tend to be situated centrally in mammalian nuclei. Studies on cell populations show that individual chromosomes do not assume precise locations. However, when an individual mitotic event is followed, the arrangement of chromosomes in the two daughter cells is very similar to that of the mother nucleus immediately prior to mitosis. This implies that the mechanics of mitosis incorporate a mechanism for maintaining chromosome position.

So interphase chromosomes occupy discrete territories with preferred locations that can be maintained during cell division. However, this should not be taken to confirm that chromosomes are static structures. In fact, this is far from the truth; locally, the chromatin within each chromosome territory can be very dynamic. Moreover, while the vast majority of DNA from an individual territory remains within the defined zone, there are instances where chromatin escapes from its territory to form loops that might spread 1 to 2 μm into the surrounding nucleoplasm. These loops will generally contain a Mbp or more of DNA and commonly have an extruded or elongated appearance suggesting that they might occupy the interchromatin space that lies between adjacent territories. Chromatin loops of this sort are

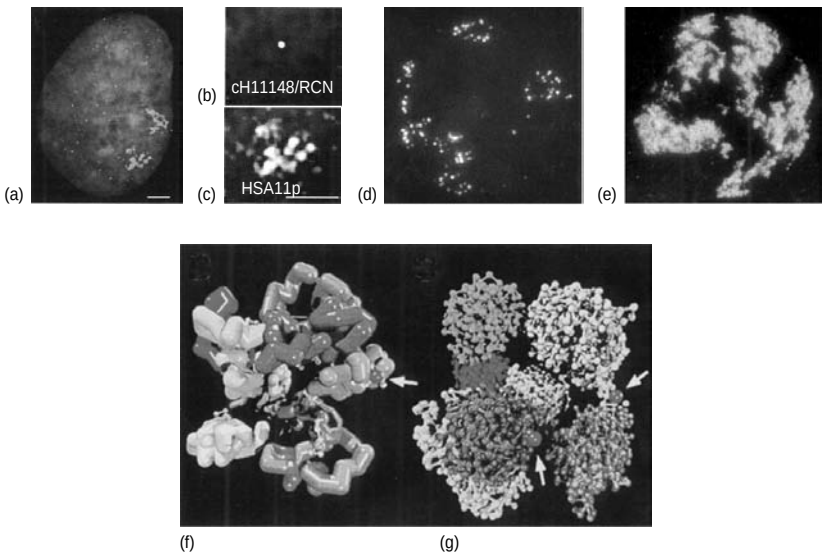

Fig. 7 The structure of chromosome territories. (a–c) show a high-resolution FISH analysis to demonstrate the organization of chromosome territories. In this example (a), the chromosome painting probe was generated from the short arm only of chromosome 11 (HSA11p) and indirectly labeled using a green fluorescent dye. Hybridization was also performed using a much shorter cosmid probe, to the ubiquitously expressed *RCN* gene from the WAGR locus of human chromosome 11, labeled with a red fluorescent dye. The relative positions of the red and green fluorescence are shown in the grayscale images in (b) and (c). Note that the expressed region is located close to the center of the chromosome territory and is separate from the dense chromatin regions, in this primary human fibroblast. The bars are 2.5 µm. From Mahy, N.L., Perry, P.E., Gilchrist, S., Baldock, R.A., Bickmore, W.A. (2002) Spatial organization of active and inactive genes and noncoding DNA within chromosome territories, *J. Cell Biol.* **157**, 579–589, with permission of Rockefeller University Press. (d–e) show the DNA foci that were labeled in HeLa cells at the onset of S-phase for 20 min or 10 h using Bromo-deoxyuridine. The samples were prepared 3 days later, when only 25 to 50% of the chromosome territories within nuclei are labeled. (The loss of labeling is determined by random segregation of the labeled and unlabeled chromatids.) Note that the DNA foci (d) are seen to cluster into discrete nuclear zones – these are chromosome territories – and that this zonal organization is much more distinct when the chromosome are labeled throughout S-phase (e). Though these samples are fixed prior to immunostaining the same observations can be made on living cells after introducing a fluorescent DNA precursor such as FITC-dUTP. (Reproduced from Jackson, D.A., Pombo, A. (1998) Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S-phase in human cells, *J. Cell Biol.* 140, 1285–1295 with permission of Rockefeller University Press. (f–g) show models for the proposed long and short-range architecture of chromosome territories. See Cremer, T., Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells, *Nat. Rev. Genet.* **2**, 292–301 for details. Image courtesy of Gregar Kreth and published with permission of Palgrave-Macmillan Publishing.) (See color plate p. xxx).

believed to reflect the transcriptional status of highly expressed gene-rich chromosomal regions. However, it remains to be established if the chromatin movement involved in the dynamic behavior of this type of chromatin loop reflects any functional association of specific genes with remote active centers – that is, specialized nuclear sites – which lie outside the territorial boundary.

The molecular interactions that define chromosome shape are not yet clear. Many lines of evidence suggest that an appropriate model to describe chromosome structure can be built around the random spatial distribution of DNA foci that contain roughly 1 Mb DNA (Fig. 7). Interestingly, structures of this size correlate with the replication foci, which, in early S-phase at least, correspond with local replicon clusters that are labeled together at specific times of the replication program. Even at the basic structural level this implies that structure and function must be related. How might this type of arrangement impact on chromatin function? In mammalian cells, it appears that the bulk of chromatin is only dynamic over short distances – perhaps ~0.5 μm or less. Beyond this, chromatin dynamics appear to be constrained by chromatin structure and numerous interactions that serve to tether chromatin to a variety of nuclear sites. Interestingly, in yeast, chromatin is so dynamic locally that a locus might move by as much as 0.5 μm during a period of only 10 s. Remarkably, this allows a single locus to pass through at least half of the nuclear volume in only 10 min. Such movements are known to be energy dependent, implying that they are probably related to nuclear activities such as replication or transcription.

3.6.1 Chromosome Structure

Chromosome structure must impose some form of order within chromosome territories and this is likely to influence function. It is generally assumed that the familiar rodlike appearance of metaphase chromosomes relies on the properties of a proteinaceous axis or core. It has been proposed that proteins within such a core might bind DNA to generate a chromosome with arrays of DNA loops of roughly 100 kbp. Simple staining techniques, using silver-based histological procedure, provide the most compelling evidence for an axial core structure, yet remarkably the proteins that might constitute this structure *in vivo* are not known. Isolated metaphase chromosome scaffolds have two predominant proteins – ScI and ScII – which for many years were the best candidates for the core. These are now known to be DNA topoisomerase IIα and SMC (structural maintenance of chromosomes) proteins. These proteins influence DNA topology and mitotic chromosome condensation and are essential for mitosis. However, studies on the behavior of these proteins in living cells do not support the idea that either topoisomerase or SMC proteins form a scaffold that organizes chromosome structure throughout the cell cycle. The likely explanation for this is that the observed structure reflects the complex biophysical properties of the nuclear components that remain associated with DNA throughout mitosis.

3.6.2 Functional Implications of Chromosome Structure and Location

Gene expression from ectopic (i.e. unnatural) chromosomal sites is extremely unpredictable. Genes introduced into inert chromosomal sites are generally inactive, though expression can be modulated by

the availability of heterochromatin proteins. Ectopic genes in active chromatin also behave unpredictably. For example, it is possible to analyze gene expression from artificial gene constructs introduced in predetermined target sites by homologous recombination. However, even when the chromosomal environment is controlled in this way, some genes display unpredictable patterns of ectopic expression, confirming that gene location at least has the potential to influence expression.

The expressional status of a gene can also be shown to influence its nuclear location. Studies on the lymphoid lineage of mammalian cells have shown that the inactivation of gene expression correlates with relocation of a gene to heterochromatic nuclear sites. During this process, the sequence-specific transcription factor Ikaros becomes associated both with the silenced gene and with local centromeric heterochromatin. Ikaros is able to bind both a target promoter and sites within the α-satellite repeats, so providing a means of driving appropriate genes into inactive heterochromatic sites. During B-lymphocyte development, the immunoglobulin H (IgH) and IgK loci are located at the nuclear periphery in hematopoietic progenitors and pro-T cells and in the nuclear interior in pro-B nuclei. The inactive loci are associated with the nuclear lamina and must move to active sites within the nuclear interior before recombination and transcription of the IgH and IgK loci can occur.

3.7 The Principles of Global Nuclear Structure

Thinking about the organization of major nuclear compartments such as nucleoli, the small nuclear bodies, sites of dedicated nuclear function – within factories – and chromosome territories leads us to assess the general principles that underlie nuclear architecture.

3.7.1 Organizing the Cytoplasm

A worthwhile way of approaching this issue is to first appreciate how major cellular activities are organized in the cytoplasm. As we have already seen, mammalian cells have a cytoplasmic compartment that is highly structured, with many membrane-bound organelles. But it is vital to recognize that this apparent structure belies the extremely dynamic nature of the cytoplasmic compartment. It is well known that membranes flow so that lipid bilayers can readily fuse one with the other. The use of GFP (green fluorescent protein), as a tool to study the dynamic behavior of various proteins of interest (each fused to GFP) in living cells, emphasizes this point. The dynamic properties of mitochondria provide an outstanding example: while an electron micrograph of a fixed mammalian cell might suggest that the cytoplasm contains numerous distinct mitochondria, live-cell studies show that these structures continually fuse with and bud from much larger mitochondrial networks. The endoplasmic reticulum, membranes of the Golgi apparatus, and membrane vesicles that lie downstream of the Golgi on the vesicle export pathways, each perform functions that require their components to be in a continual state of flux. However, while the observed intracellular activities might appear to be both dynamic and chaotic, they are in fact highly structured. Membrane-bound vesicles are too large to move inside the cytoplasm by diffusion, and move instead by energy-dependent processes that are directed by cytoplasmic motor proteins. The basic components needed to achieve this are the motor protein complexes themselves and a series of networks

upon which the motor proteins can move. In eukaryotes, the cytoskeleton provides the network that coordinates the movement and location of cytoplasmic vesicles and organelles.

Stated briefly for the sake of clarity, the cytoskeleton of higher eukaryotes is composed of three filamentous networks:

Microtubules (MTs) – are long hollow tubes that measure about 25 nm in diameter. They are composed of multimeric structures that are assembled from heterodimers of α- and β-tubulin. The tubules form with polarity, polymerizing and deploymerizing more rapidly at their + ends. The ends of MTs are usually embedded in a microtubule-organizing center. Typically, in higher cells, this will be the centrosome. MTs are dynamic, with a polymerization rate of $\sim$1 μm min^{-1}. However, MTs are stabilized through interactions with a large number of other proteins – the microtubule associate proteins (MAPs). During mitosis of the cell cycle, phosphorylation of MAPs by the cell cycle kinases has a dramatic impact on MT behavior; particularly with respect to their dynamic properties, which increase 5 to 10 fold. Under these circumstances, the MTs play a vital role in the process of chromosome segregation.

The MTs act as tracks for cytoplasmic motor proteins that fall into two broad classes. Generally speaking, kinesins are plus-end directed MT motor proteins and dyneins are minus-end directed MT motor proteins. Small families of each of these motors are known and these serve specific purposes and ensure that defined cargoes are able to move to appropriate sites throughout the cell.

Microfilaments (MFs) – are much narrower, typically about 8 nm across, helical polymers of a single protein, actin. Like the MTs a complex filamentous network of actin fibers pervades the cytoplasm of most eukaryotic cells. Like MTs, MFs are dynamic and serve as tracks upon which myosin motors can move organelles and vesicles around the cell. Actin filaments play particularly important roles in cell motility and cell architecture.

Intermediate filaments (IFs) – are a much more complex family of proteins that play particular roles in tissue architecture. As this role implies, the IF networks are much more important in multicellular eukaryotes than in their primitive ancestors; simple eukaryotes do not have intermediate filament proteins. IFs are $\sim$10 nm across and intermediate in size between MFs and MTs, hence the name. Members of the protein family have a distinct elongated central domain that interacts as a coiled-coil structure to generate first dimers and then complex multimers that form the filaments. The family in mammals has about 50 members, though different members are expressed in a cell lineage–dependent manner. Keratins, for example, are expressed in the epidermis of the skin and play a vital role in maintaining the mechanical properties of this organ.

Cytoskeleton – is the structure formed from these three networks. The networks are closely associated and cross-linked, through the behavior of associated proteins so that the function of each is dependent on the structure of the others. The cytoskeleton should be considered as an integrated network with cooperative functions.

Even a superficial appraisal of cytoplasmic architecture begs the question: do the essential functions performed by the cytoskeleton have counterparts in the nucleus and if so do any analogous structures exist? This question has been a focus of intense debate over the past 25 years. We have already discussed the function and properties of the nuclear lamina. This network of type V intermediate filament

proteins is by far the best-characterized structural nuclear network. Roles played at the nuclear periphery are particularly clear. What remains more a matter of speculation is whether a diffuse lamin network spreads throughout the nucleoplasm, and if this could then support the nuclear functions that many investigators believe take place at a nuclear matrix.

3.7.2 The Nuclear Matrix – Fact or Fiction

This leads us to assess what the nuclear matrix might be and analyze the functions it might perform. We have seen that nuclei can be plastic, so that they appear to have a complex structure of nuclear compartments, which generally represent centers of nuclear function. These centers have components that are inherently dynamic and in general are not stable entities – this is self-evident given that most are disrupted during mitosis. However, while these compartments might be plastic, it is also interesting to note that the spatial organization of sites such as transcription and replication factories and nuclear speckles is unaffected if nuclei are depleted of chromatin (Fig. 8). This suggests that the compartments are spatially constrained and raises the possibility that a "solid phase" exists within nuclei to provide a platform upon which the active sites assemble. A "nucleoskeleton" has been proposed to fulfil his role.

The nature of the nucleoskeleton and related nuclear matrix remains a matter of debate. Core filaments that could provide a form of continuity throughout the nuclei have been described. These filaments have characteristic morphological features of the intermediate filaments seen throughout the cytoplasm. The best-characterized intermediate filaments within nuclei are undoubtedly the nuclear lamins. However, in addition to roles played by the nuclear lamins at the nuclear periphery, recent studies using GFP-lamin proteins and classical high-resolution immunostaining have suggested that a veil of lamin filaments spreads throughout the nucleus. These observations support the idea that a lamin-based network might pervade the nucleoplasm to provide a framework to coordinate the organization of various nuclear compartments. This idea, however, remains controversial and it is still a matter of debate as to whether the lamins might form a contiguous internal nucleoskeleton. As an alternative, the lamins could feasibly play a structural role in the short-range organization of nuclear compartments such as transcription centers.

3.7.3 Proteins of the Nuclear Matrix

The nuclear matrix is classically described as an amorphous fibrogranular structure that can be isolated from the nuclei by hypertonic treatment following nuclease digestion. By this definition the matrix is perceived as being a product of the nucleoskeleton/nuclear lamina networks and associated proteins. The matrix typically contains many hundreds of different proteins, many of which have been studied in detail. Two of the best characterized – SAF-A (scaffold attachment factor A) and ARBP (attachment region binding protein) – were discovered because of their association with the matrix attached DNA elements. SAF-A turns out to be a major RNA binding protein (hnRNP-U) and ARBP is a protein that binds methylated DNA (MeCP2). SAF-A has been shown to bind p300, a major transcriptional coactivator, and so recruit active genes to the nuclear matrix. ARBP, in contrast, interacts with methylated DNA in MAR elements and through an intermediary that contains Sin3A protein recruits histone deacetylases to generate

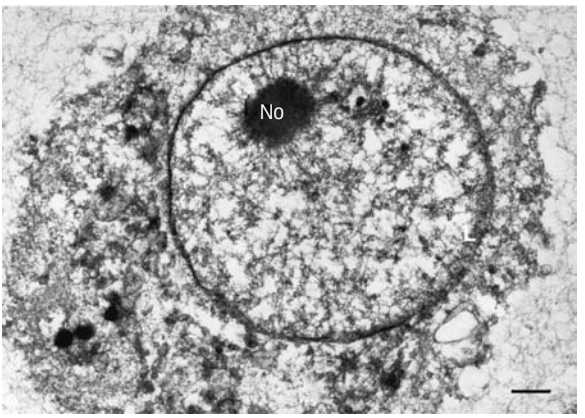

Fig. 8 Nuclear architecture and the nucleoskeleton. To reveal the nucleoskeleton, HeLa cells encapsulated in agarose microbeads were permeabilized and chromatin cut with nucleases so that ~90% DNA could be removed before preparing a resinless electron micrograph. In this ~250-nm thick section, note that a diffuse agarose mesh surrounds the densely stained cell. The spherical, central nucleus is demarcated by a dense nuclear lamina (L). Critically, in this example, nuclear architecture is preserved even though most chromatin – half of the nuclear mass – is removed. The residual nucleoskeleton can be visualized as a diffuse network of coated filaments that pervade the nuclear volume and provide a "solid phase" upon which nuclear compartments such as nucleoli (No), replication and transcription factories and interchromatin granule clusters are assembled. Note that the distribution of transcription and replication centers is preserved even though almost all chromatin has been removed. The bar is 2 μm. (Reproduced from Hozak, P., Jackson, D.A., Cook, P.R. (1994) Replication factories and nuclear bodies: the ultrastructural characterization of replication sites during the cell cycle, *J. Cell Sci.* **107**, 2191–2202. With permission of the Company of Biologists.)

a silenced chromatin state. Another protein, SAF-B, (scaffold attachment factor B) specifically binds to S/MAR regions, interacts with RNA polymerase II (RNA pol II) and a subset of serine-/arginine-rich RNA processing factors (SR proteins). It was proposed that these interactions allow SAF-B to provide a surface for the assembly of the transcription apparatus. As with the vast majority of characterized matrix proteins, these are involved in different aspects of chromatin function. Extending this idea implies that the nucleoskeleton and nuclear matrix (that forms during salt extraction) are an expression of different processes that are performed within the interchromatin space.

Recent studies on the nuclear matrix protein SATB1 (special AT-rich sequence binding 1) emphasize this view. SATB1 is a nuclear matrix component that is found predominantly in thymocytes. In these cells, a nuclear network of SATB1 binds chromatin to form chromatin domains

that are believed to play important roles in gene regulation. SATB1 in association with MARs acts as a transcriptional repressor to regulate gene expression during T-cell development. This protein has the classical properties of a nuclear matrix component and deletion of SATB1 results in defects in the temporal and spatial regulation of many lymphocyte genes, leading to an arrest of T-cell development. SATB1 is known to target chromatin remodeling factors to specific chromatin domains and so provide a mechanism for the long-range regulation of gene expression. SATB1 recruits histone deacetylase complexes containing the NURD chromatin remodeling complex to SATB1-bound sites in the interleukin-2 receptor gene to elicit specific deacetylation of histones within the locus. In addition, SATB1 was shown to target CHRAC and ACF chromatin remodeling complexes to regulate nucleosome positioning over several kbp DNA. SATB1 might also play a direct role in recruiting RNA polymerase II to specific nuclear sites.

3.7.4 Functions at the Matrix

Numerous experiments support the existence of a nuclear matrix that plays central roles in nuclear function. An excellent example is provided by the studies of the behavior of the runt-related transcription factors (RUNX/CBFA/AML), which play essential roles in cellular differentiation and fetal development. A domain within the RUNX transcriptional activators targets the protein to discrete subnuclear foci. Removal of the targeting domain results in lethal hematopoietic and skeletal phenotypes, and implies that for this class of protein the correct nuclear location is critical for the function. The behavior of GFP-RUNX fusion proteins in living cells confirms that RUNX proteins occupy discrete nuclear foci, the majority of which are coincident with transcription sites and associated with a low-salt nuclear matrix. RUNX proteins also interact with Smads, a family of signaling proteins that regulate various developmental and biological processes in response to growth factors. Interestingly, the RUNX proteins are required to target Smads to nuclear sites where transcription is performed. This implies that gene expression involves the *in situ* integration of critical signals through the assembly of regulatory complexes at transcriptionally active subnuclear sites. This work and other studies suggest that activating factors engage appropriate nuclear sites that are competent to perform gene expression and that the process of transcription can only occur once a gene has been recruited to the active center.

4 Integrating Nuclear Structure and Function

Having addressed different aspects of nuclear structure and function we should now explore how these might be integrated and assess if this process influences our perceptions of nuclear organization. The basic principles of chromosome territory organization were outlined earlier. However, it is important to remember that these territories are not homogeneous structures. It is possible, for example, to use three-dimensional FISH techniques to show that condensed chromatin as well as the active chromatin compartment (using specific antibodies for modified histones found in transcribed chromatin) occupy spatially distinct compartments from the transcription factor TFIIH, RNA polymerase II, and hnRNP-U. Interestingly, these proteins are present throughout the

interchromatin space, inside each territory and largely excluded from the domains of chromatin. This approach and other high-resolution studies support the view that transcription centers are restricted to the RNA-rich interchromatin compartment. This interchromatin compartment supports all the major chromatin functions – and it is within this compartment that any nuclear matrix must form.

4.1 Chromatin Structure and Function

While appreciating that the way nuclei are organized might have a profound impact on nuclear function, it is also important to understand how DNA is organized in order to provide an amenable template that supports the major functions. In this respect, it is necessary to understand how DNA is packaged as chromatin and how the properties of this DNA–protein complex impact on function.

To emphasize this point, it is well known that different genes are expressed to quite different extents and that many are expressed in specific cells and at precise times during development. This begs the question: how are different levels of gene expression maintained? For a specific gene, chromatin status and the availability of activating transcription factors will combine to establish an engaged transcription complex that drives RNA synthesis. The role of chromatin structure is clearly implicit in the success of this process. It has been know for many years that active and inactive genes have quite distinct chromatin states and that inert chromatin is able to spread – and so downregulate the previously active genes. Two general features are fundamental in protecting the functional status of the active genes. The first concerns the maintenance of a fluid and active chromatin state through histone modification. The second requires that genes are organized into chromatin domains that behave as the functional units of gene expression. The structure of these domains, at least within the confines of their natural chromosomal locus, might ensure that genes are expressed at the desired levels in appropriate tissues and at the required times during development.

4.1.1 Chromatin Function

Work that determined the genetic code that defines humanity represents a landmark of scientific endeavor. But with this treasure comes the realization that it will not always be possible to decode information held within the genome in a way that might allow us to predict patterns and levels of gene expression in different cell types and at different stages of their development. Some aspects of gene expression control are genetic in origin. Most notably, we have seen how gene expression is activated when transcription factors interact with their cognate recognition motifs in gene promoters and other activating sequences – such as enhancers – to form a complex that recruits the transcription machinery to a gene. This process of sequence-dependent association of specific proteins with defined sequence elements within DNA can be studied in molecular detail, and is conceptually simple to understand. In higher eukaryotes, in particular, these genetic controls are supplemented by a variety of "epigenetic" factors that modulate chromatin structure and both chromosome and nuclear architecture.

4.1.2 DNA Packaging and Chromatin

Somatic mammalian cells are usually diploid; they contain two sets of

chromosomes one inherited from each of the two parents. Humans have 46 chromosomes with about 6×10^9 bp DNA in total. The structure of DNA, with its acid sugar-phosphate backbone, demands that some mechanism must have evolved to neutralize the negatively charged phosphate groups in order to fold DNA so that it might fit inside a mammalian nucleus. Note that this amount of DNA would cover about 2 m if extended, yet must fold to occupy a nucleus that is only ~10 μm in diameter. A family of small (about 10 kDa) arginine and lysine-rich proteins called *histones* are responsible for DNA packaging. The fundamental structural unit of DNA packaging is called the *nucleosome*. The core structure of a nucleosome (Fig. 9) contains ~146 bp DNA wrapped as 1.75 turns over the surface of a histone octamer. The protein core structure contains two sets of the four core histones – called *histones* H2A, H2B, H3 and H4 – arranged with mirror symmetry around the dyad axis of the nucleosome. The histone genes themselves are extremely highly conserved throughout eukaryotes as their proteins must interact in a very precise way to form the nucleosome core. The core structure is in turn defined by the architecture of the individual histone proteins (Fig. 9) while

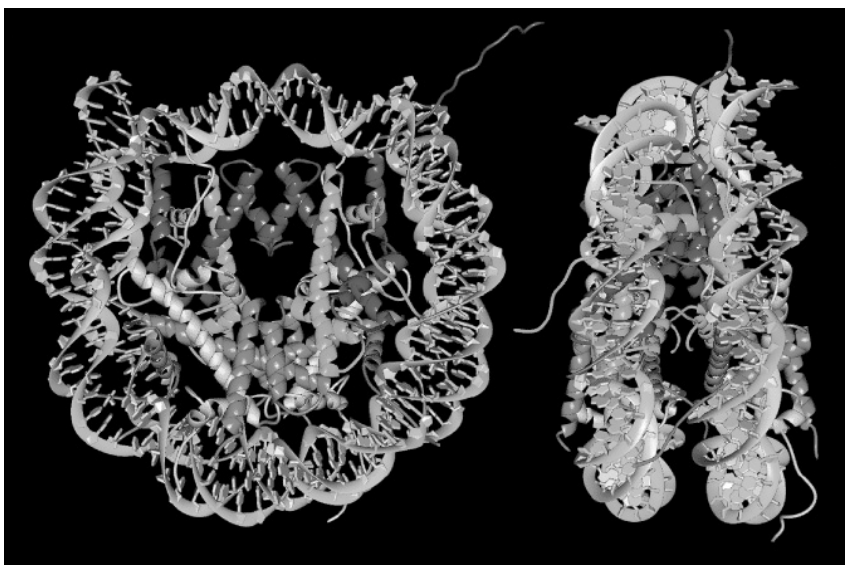

Fig. 9 The nucleosome. High-resolution X-ray crystal analysis can be used to demonstrate the structure of a nucleosome. In this nucleosome core particle, 146 bp of DNA (turquoise and brown) wraps ~1.75 times around the histone octamer. The octamer of each core contains two molecules each of histones H2A (yellow), H2B (red), H3 (blue) and H4 (green). Each of the core histones has a central domain with three α-helical regions separated by short coils. Interactions between these domains within the core determine the core structure. The N-terminal domains are unstructured and can extend to the outside surface of the nucleosome. The dimensions of this core structure are 10×6 nm. (Reproduced from Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 angstrom resolution, *Nature* **389**, 251–260. With permission of Palgrave-Macmillan Publishing.) (See color plate p. xxxi).

the N-terminal domains of the histones associate with DNA and allow interaction between adjacent nucleosomes.

4.1.3 Euchromatin and Heterochromatin

As the chromatin fiber is the template for functions such as DNA replication and RNA transcription, the importance of this DNA–protein complex cannot be overstated. In particular, it is crucial to recognize how modification of chromatin structure impacts on the behavior of chromatin. Broadly speaking, chromatin within a mammalian nucleus can be characterized as euchromatin or heterochromatin. The former is ostensibly composed of chromosomal regions that have a high density of transcriptionally active genes. Euchromatin corresponds to chromosomal regions that can be classified as R-bands using cytological criteria. These are slightly GC-rich (because of GC-islands associated with housekeeping genes), have a high density of Alu-repeats and are duplicated in the early part of S-phase. Heterochromatin, in contrast, has many fewer transcribed genes, occupies chromosomal G- and C-bands and is replicated in the second half of S-phase. Euchromatin and heterochromatin have distinct features that correlate with these characteristics. Perhaps the most diagnostic among these is the ease with which the different chromatin populations can be cut with nucleases. Hence, euchromatin is readily digested with enzymes such as DNase – it is classified as being DNase sensitive – whereas heterochromatin is relatively insensitive. This ease of digestion reflects the accessibility of DNA to DNAse – euchromatin has a loose or open chromatin structure, whereas heterochromatin is more compact. These basic chromatin states correlate with a functional status, which reflects how transcribed genes are packed into chromatin that is modified to allow transcription to occur.

4.1.4 Histone Modifications

As chromatin modification is central to the control of gene expression, it is important to understand at least the basic principles of this process. It has been known for many years that the histones are subject to a wide range of modifications. These include acetylation, methylation, phosphorylation, ADP-ribosylation, and ubiquination. These modifications are used to generate and stabilize the different classes of chromatin, modulate chromatin structure throughout the cell cycle, and control histone turnover. In terms of gene expression, acetylation of the N-terminal domains of histones – particularly H3 and H4 – is of the greatest importance. These lysine-rich domains each have numerous lysine residues that are targets for acetylation. Acetylation serves to reduce the stability of the nucleosome complex so that the DNA is more readily accessible to the transcription machinery.

The mechanisms by which histone acetylation are controlled are extremely complex. Levels of acetylation at any locus are dictated by the combined activities of histone acetyl transferases (HATs) and histone deacetylase complexes (HDACs). The activity of these large protein complexes is determined, in turn, by mechanisms that control their recruitment to different nuclear sites. The protein p300/CBP is a global transcriptional regulator that binds to enhancers within many gene loci and contains a HAT activity that is capable of acetylating specific sites in all the core histones and other transcription factors to stimulate transcription. In addition, complexes such as PCAF interact with p300/CBP, SCR1/ACTR and

other DNA-binding activators to stimulate transcription by histone acetylation, predominantly in histone H3. Components of the multi-subunit general transcription factors TFIID and TFIIIC also have HAT activities that modify histones H3 and H4 to stimulate transcription and other HAT activities are associated with the elongating polymerase holoenzyme complexes. The patterns of histone acetylation that might be developed in response to these different activities can be extremely complex; different HATs have numerous potential acetylation targets and different preferences for the various sites. Major modifications that correlate with an active chromatin status include acetylation of histone H3 at lysines 9 and 14 and H4 at lysine 5. Specificity for particular sites in chromatin is a product of the mechanisms of association of each HAT with chromatin. The bromodomain of these proteins is thought to play a role in this process.

Histone acetylation in euchromatin correlates with gene activity. In contrast, inactive, heterochromatin has lower levels of acetylation and much higher levels of histone methylation and phosphorylation. The human suppressor of variegation protein SUV39H1 encodes a histone methyltransferase (HMT) that selectively methylates histone H3 at lysine 9. This activity is dependent on a SET domain within the protein. This particular histone modification induces high-affinity binding of the heterochromatin protein HP1, through chromodomains. The other major group of heterochromatin proteins, the polycomb group (Pc-G) proteins, are known to recruit protein complexes with histone deacetylase activity. Pc-G proteins and antagonizing proteins of the trithorax group (trx-G) together play a fundamental role in modulating the dynamic transition between inactive and active chromatin states.

4.1.5 ATP-dependent Chromatin Remodeling Machines

Chromatin status and the activity of complexes that catalyze the modification of the histones are also influenced by the activity of a variety of protein complexes that perform ATP-dependent chromatin remodeling. These complexes were first described in studies to understand the control of mating type switching (SWI) and sucrose fermentation (SNF for sucrose nonfermenting) in yeast. Chromatin remodeling was recognized as a major factor in these two processes and has since been shown to be a fundamental regulator of gene expression. Examples of the remodeling machines in human cells include the human SWI/SNF complex, NURD (nucleosome remodeling histone deacetylase complex) and RSF (remodeling and spacing factor). These multiprotein complexes operate through different chromatin binding domains. hSWI/SNF has a bromodomain, NURD chromodomains and RSF a SANT domain.

These chromatin remodeling machines serve to increase the local dynamic properties of chromatin. DNA and histones in chromatin make so many contacts that the nucleosomes they form are inherently stable structures. Nucleosomes can form on almost all stretches of DNA of sufficient length, though the need to fold the DNA duplex over the nucleosome surface does impose constraints on the way chromatin forms. In particular, the center of the dyad axis in a nucleosome has a region of DNA that is distorted or kinked in order to make the necessary contacts with the histones of the nucleosome core. AT bases in DNA are preferred at this location. Other mechanisms exist to position the nucleosome in a specific way – the binding of a factor with DNA prior to establishing local chromatin structure would be an

obvious mechanism. The ATP-dependent chromatin remodeling activities serve to enhance the fluid properties of nucleosomes; in essence, they allow nucleosomes to slide on DNA. The mechanism of this process is not known in detail, but is assumed to reduce the activation energy needed to reposition a nucleosome core.

Hence, the chromatin remodeling machines provide a means by which nucleosomes can be repositioned to allow access to previous inaccessible sites in DNA. This activity will function cooperatively with the histone modification systems described above to modulate and stabilize different chromatin states. The combination of these activities adds a huge complexity to the epigenetic control of chromatin function. Indeed, present estimates suggest that something in the region of 50 chromatin-modifying complexes will collaborate to ensure that chromatin is an extremely complex and structurally dynamic substrate.

4.2 Higher-order Chromatin Folding

The following section touches briefly on chromatin architecture and in particular, on differences in chromatin structure that correlate with chromatin function, notably RNA transcription. With this in mind, it is worth emphasizing that only a very small fraction – about 1% – of the DNA within a mammalian cell is represented by messenger RNA (mRNA) that is decoded to generate proteins in the cytoplasm. About 10 to 20% of the DNA is in the euchromatin that is transcribed to generate the primary transcripts – called *heterogeneous RNA (hnRNA)* – that are subsequently processed to form mature mRNAs. Individual cell types have distinct expression profiles that are determined by their developmental history (differentiation program). About one-third of the roughly 35 000 genes of human cells are housekeeping genes that are expressed in all cell types. The others are expressed in specific, specialized cell types and generally form facultative heterochromatin in cells in which they are not expressed and euchromatin in cells in which they are expressed.

It is self-evident that DNA must be highly folded so as to accommodate 2 m of DNA in a nucleus measuring only 10 μm in diameter (Fig. 10). The wrapping of DNA around the nucleosome core begins the condensation process that is continued by a hierarchy of higher-order DNA folding. Nucleosomal chromatin can be visualized by electron microscopy, as a chromatin fiber of beads on a string; in which each bead or nucleosome is separated from the next by a short stretch of intervening linker DNA; in mammalian cells these linkers are typically 50 to 100 bp in length. As nucleosomes are 10 nm in diameter, this is referred to as the 10-nm chromatin fiber. Inside the cell, euchromatin is loosely folded as a locally chaotic 10-nm chromatin fiber. Interactions between different regions of the fiber might generate local variations in chromatin density – though generally the chromatin will appear diffuse. Heterochromatin, in contrast, is much more highly compacted.

When chromatin is visualized in thin sections of mammalian cells a number of folded forms can be identified. Aggregates of the 10-nm fiber form a fiber with an average diameter of ~30 nm. Historically, a chromatin fiber of this size has been referred to as the chromatin solenoid – in which six nucleosomes are held in one turn of the solenoid by histone H1. Histone H1 is known as a linker histone; it is much richer in heterochromatin, with about 1 copy per nucleosome core compared to

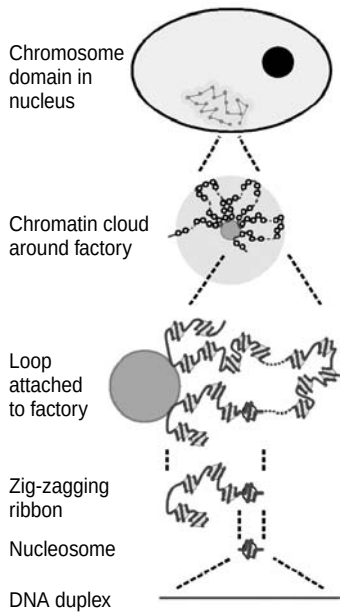

Fig. 10 Chromatin folding. The DNA within a mammalian nucleus is highly folded. The required folding is achieved first by wrapping DNA around the surface of histone complexes to form nucleosomes. Long stretches of nucleosomes then fold to form the higher-order chromatin structures. In the model shown here, DNA foci – see Figure 7 – are depicted as the critical architectural units of higher-order packaging. It is not certain how these structures are formed, but one possibility is that DNA loops are arrayed as stable entities in a way that reflects their functional status at sites of RNA and DNA synthesis. In this example, a cluster of chromatin loops is arrayed around a transcription factory. Within this type of higher-order structure, chromatin might be either open, as euchromatin or more condensed, as heterochromatin. These chromatin states correlate with various histone modifications that either allow chromatin to be transcribed or stimulate local aggregation of the chromatin clouds. (Reproduced from Cook, P.R. (2001) *The Principles of Nuclear Structure and Function*, John Wiley and Sons, New York with permission of John Wiley and Sons.)

about 0.5 copies per nucleosome core in DNase sensitive euchromatin. The solenoid represents the second level of chromatin folding. This structure is not thought to fold according to specific design parameters, however, and is more probably best imagined as a short-range chromatin aggregate that is stabilized by association with chromatin proteins such as histone H1. Further condensation of the chromatin leads to the formation of condensed fibers with diameters in the range of 50 to 80 nm. These are typical of the structures seen in heterochromatin during the interphase and are stabilized by association with heterochromatin proteins such as HP1 and Pc-G proteins. The most highly condensed chromatin fibers are found in mitotic chromosomes where histone modifications that result from the activity of the mitotic cdc/cyclin complexes yield chromatin fibers of 150 to 200 nm. These are folded around the axis of the chromatids, which have an overall diameter of about 600 nm.

4.2.1 Chromatin Domains and Loops

Interactions that allow the stabilization and modulation of chromatin status will clearly have a profound impact on chromatin organization within the nuclei of higher eukaryotes. However, these do not seem to be the only aspects of

chromatin organization that dictate how chromatin is packaged *in vivo*. An important consideration here is the way genes might be organized into chromatin domains – these are usually conceptualized as DNA loops – that might be regarded as the structural units of gene expression.

Evidence for the existence of discrete chromatin domains comes from a number of sources. Nucleases sensitivity is the classical indicator of chromatin structures that correlates with gene expression; general nuclease sensitivity of active loci will often spread throughout domains that cover many kbp up- and downstream of an expressed gene. Chromatin immunoprecipitation experiments – using antibodies to histone variants found in active chromatin – confirm this. Moreover, clear transitions from active to inactive chromatin suggest that chromatin domains are demarcated by elements that determine the boundaries of functional genetic units. As an example, consider a high-resolution analysis of acetylation across three genes in the vicinity of the chicken β-globin locus. This chromosomal region contains a folate receptor gene, a 16-kbp condensed chromatin region, the β-globin gene domain and an adjacent olfactory receptor gene. The condensed chromatin maintains very low levels of histone acetylation at all developmental stages, with similar levels maintained in inactive genes. Much higher levels of acetylation are seen throughout the transcribed gene domains, while chromatin in the vicinity of upstream regulatory sites maintain the highest levels of acetylation. Most significantly, a very strong constitutive focus of hyperacetylation corresponds with an insulator element that appears to define the globin and adjacent folate receptor domains. These observations show how epigenetic modifications influence the chromatin domain structure and also imply that different classes of histone acetyl transferase with different chromatin targets are involved in the control of gene expression.

Various genetic elements have been described that together define the structure of chromatin domains and regulate gene expression. These include: promoter elements, enhancers, LCRs, S/MARs and insulators. Enhancers and LCRs generally contain arrays of transcription factor binding sites that bind appropriate factors to augment levels of gene expression, by directly upregulating rates of transcription. These components contribute to the maintenance of a gene's expressional status and are responsible for establishing a productive spatial configuration in chromatin and perhaps for targeting chromatin to specific nuclear sites. In some cases, LCRs are required to establish natural levels of gene expression from ectopic (unnatural) chromosomal sites.

The nuclear matrix and related scaffold are believed to play important roles in different aspects of chromatin function. S/MARs are DNA elements that clearly augment transcription and may achieve this using AT-rich sequences to modulate superhelical stress that arises during transcription. The importance of these elements in gene regulation is confirmed by the fact that a classical MAR element is known to play a critical role in orchestrating the temporal and spatial expression of many genes during T-cell development. In addition to this, efficient MAR elements have been shown to support long-term gene expression from ectopic chromosomal sites, suggesting that these elements might play a role in targeting gene domains to specific nuclear sites, such as matrix-associated transcription centers, prior to gene expression. These elements are probably distinct from

insulators, which are believed to function as chromatin domain boundaries that generate self-contained units of gene expression.

The combined action of promoter, enhancer, S/MAR and insulator elements will together dictate levels of gene expression. In addition, through their interaction with proteins involved in gene expression, these elements will play an important part in determining how DNA is organized and packaged within the nuclear space.

5 Overview

This chapter sets out to give the reader an overview of our present understanding of the nuclear structure and function. The majority of examples have been drawn from studies on mammalian cells. This was a deliberate choice, as even though many critical advances have been made using simple model systems, the added complexity of mammalian genomes makes them a much more challenging object of study. As the area covered herein represents an extensive field of research, it is inevitable that the treatment of many has been superficial. This should not be seen to reflect my own view of the relative importance of different topics. Instead, I have attempted to deliver a coherent appraisal of this cell organelle with the intention of stimulating interest in the reader so that he or she will feel encouraged to look deeper into this fascinating subject.

In discussing the nucleus, the main focus of attention must be the structure and function of DNA. Little attention is paid to the DNA itself, though the way DNA is packaged in nuclei undoubtedly has a profound influence on chromatin function. The structure of chromatin and more importantly the cellular mechanisms that modulate chromatin fluidity and hence control access of DNA to RNA and DNA polymerases are now well known. Indeed, histone modifications represent a major source of epigenetic information that complements the genetic information defined by DNA sequence. Chromosome structure and global nuclear architecture provide additional sources of epigenetic information, which at least have the potential to influence chromatin function. However, while many specific examples have been reported to show how chromosome and/or nuclear location influence chromatin function, the relative importance of this type of spatial epigenetics remains largely unexplored. Whatever the hierarchies of control turn out to be, there is no denying that the structure of mammalian nuclei and the functions that they perform are very sophisticated indeed.

See also Chromosome Organization within the Nucleus; Developmental Cell Biology; Electron Microscopy in Cell Biology.

Bibliography

General Reference and Books

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2002) *Molecular Biology of the Cell*, 4th edition, Garland Publishing, New York.

Cook, P.R. (2001) *The Principles of Nuclear Structure and Function*, John Wiley and Sons, New York.

Endow, S.A., Glover, D. (1998) *Dynamics of Cell Division*, Oxford University Press, Oxford.

King, R.J.B. (2000) *Cancer Biology*, 2nd edition, Prentice Hall, Harlow.

Lewin, B. (2003) *Genes VII*, Oxford University Press, Oxford.

Lodish, H., Berk, A., Matsudaira, P., Kaiser, C.A., Krieger, M., Scott, M.P., Zipursky, L., Darnell, J. (2003) *Molecular Cell Biology*, 5th edition, WH Freeman and Co, New York.

Tsonis, P.A. (2003) *Anatomy or Gene Regulation*, Cambridge University Press, Cambridge.

Turner, B.M. (2001) *Chromatin and Gene Regulation: Molecular Mechanisms in Epigenetics*, Blackwell Science, Oxford.

White, R.J. (2000) *Gene transcription: Mechanisms and Control*, Blackwell Science Ltd, Oxford.

Wolffe, A. (1998) *Chromatin Structure and Function*, 3rd edition, Academic Press, San Diego.

Research Articles are Reviews

Alvarez, J.D., Yasui, D.H., Niida, H., Joh, T., Loh, D.Y., Kohwi-Shigematsu, T. (2000) The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development, *Genes Dev.* **14**, 521–535.

Barboro, P., D'Arrigo, C., Diaspro, A., Mormino, M., Alberti, I., Parodi, S., Patrone, E., Balbi, C. (2002) Unraveling the organization of the internal nuclear matrix: RNA-dependent anchoring of NuMA to a lamin scaffold, *Exp. Cell Res.* **279**, 202–218.

Baxter, J., Merkenschlager, M., Fisher, A.G. (2002) Nuclear organisation and gene expression, *Curr. Opin. Cell Biol.* **14**, 372–376.

Beaudouin, J., Gerlich, D., Daigle, N., Eils, R., Ellenberg, J. (2002) Nuclear envelope breakdown proceeds by microtubule-induced tearing of the lamina, *Cell* **108**, 83–96.

Bebenek, K., Kunkel, T.A. (2002) Family growth: the eukaryotic DNA polymerase revolution, *Cell Mol. Life Sci.* **59**, 54–57.

Becker, P.B. (2002) Nucleosome sliding: facts and fiction, *EMBO J.* **21**, 4749–4753.

Bell, A.C., West, A.G., Felsenfeld, G. (2001) Gene regulation – insulators and boundaries: versatile regulatory elements in the eukaryotic genome, *Science* **291**, 447–450.

Berezney, R., Mortillaro, M.J., Ma, H., Wei, X., Samarabandu, J. (1996) The nuclear matrix: a structural milieu for genomic function, *Int. Rev. Cytol.* **162A**, 1–65.

Blackwood, E.M., Kadonaga, J.T. (1998) Going the distance: a current view of enhancer action, *Science* **281**, 60–63.

Bode, J., Benham, C., Knopp, A., Mielke, C. (2000) Transcriptional augmentation: modulation of gene expression by scaffold/matrix attached regions (S/MAR elements), *Crit. Rev. Eukaryotic Gene Expr.* **10**, 73–90.

Borden, K.L.B. (2002) Pondering the promyelocytic leukemia protein (PML) puzzle: possible functions for PML nuclear bodies, *Mol. Cell Biol.* **22**, 5259–5269.

Bridger, J.M., Boyle, S., Kill, I.R., Bickmore, W.A. (2000) Re-modelling of nuclear architecture in quiescent and senescent human fibroblasts, *Curr. Biol.* **10**, 149–152.

Bulger, M., Sawado, T., Schubeler, D., Groudine, M. (2002) ChIPs of the beta-globin locus: unraveling gene regulation within an active domain, *Curr. Opin. Genet. Dev.* **12**, 170–177.

Butler, J.E.F., Kadonaga, J.T. (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression, *Genes Dev.* **16**, 2583–2592.

Carter, D., Chakalova, L., Osborne, C.S., Dai, Y.F., Fraser, P. (2002) Long-range chromatin regulatory interactions in vivo, *Nat. Genet.* **32**, 623–626.

Chen, D.Y., Hinkley, C.S., Henry, R.W., Huang, S. (2002) TBP dynamics in living human cells: constitutive association of TBP with mitotic chromosomes, *Mol. Biol. Cell* **13**, 276–284.

Choi, J.Y., Pratap, J., Javed, A., Zaidi, S.K., Xing, L.P., Balint, E., Dalamangas, S., Boyce, B., van Wijnen, A.J., Lian, J.B., Stein, J.L., Jones, S.N., Stein, G.S. (2001) Subnuclear targeting of Runx/Cbfa/AML factors is essential for tissue-specific differentiation during embryonic development, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8650–8655.

Chubb, J.R., Boyle, S., Perry, P., Bickmore, W.A. (2002) Chromatin motion is constrained by association with nuclear compartments in human cells, *Curr. Biol.* **12**, 439–445.

Cleaver, J.E., Crowley, E. (2002) UV damage, DNA repair and skin carcinogenesis, *Front. Biosci.* **7**, D1024–D1043.

Cook, P.R. (1999) The organization of replication and transcription, *Science* **284**, 1790–1795.

Cremer, T., Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells, *Nat. Rev. Genet.* **2**, 292–301.

De, D.N. (2002) Protein constitution of the chromosome axis, *Chromosoma* **111**, 69–79.

Dillon, N., Festenstein, R. (2002) Unravelling heterochromatin: competition between positive and negative factors regulates accessibility, *Trends Genet.* **18**, 252–258.

Dundr, M., Hoffmann-Rohrer, U., Hu, Q.Y., Grummt, I., Rothblum, L.I., Phair, R.D., Misteli, T. (2002) A kinetic framework for a mammalian RNA polymerase in vivo, *Science* **298**, 1623–1626.

Emerson, B.M. (2002) Specificity of gene regulation, *Cell* **109**, 267–270.

Farrell, C.M., West, A.G., Felsenfeld, G. (2002) Conserved CTCF insulator elements flank the mouse and human beta-globin loci, *Mol. Cell Biol.* **22**, 3820–3831.

Femino, A.M., Fay, F.S., Fogarty, K., Singer, R.H. (1998) Visualization of single RNA transcripts in situ, *Science* **280**, 585–590.

Francastel, C., Magis, W., Groudine, M. (2001) Nuclear relocation of a transactivator subunit precedes target gene activation, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12120–12125.

Gasser, S.M. (2002) Visualizing chromatin dynamics in interphase nuclei, *Science* **296**, 1412–1416.

Gerlich, D., Beaudouin, J., Kalbfuss, B., Daigle, N., Eils, R., Ellenberg, J. (2003) Global chromosome positions are transmitted through mitosis in mammalian cells, *Cell* **112**, 751–764.

Gilbert, D.M. (2002) Replication timing and transcriptional control: beyond cause and effect, *Curr. Opin. Cell Biol.* **14**, 377–383.

Goldberg, M.W., Allen, T.D. (1992) High-resolution electron microscopy of the nuclear envelope – demonstration of a new, regular fibrous lattice attached to the baskets of the nucleoplasmic face of nuclear pores, *J. Cell Biol.* **119**, 1429–1440.

Goldman, R.D., Gruenbaum, Y., Moir, R.D., Shumaker, D.K., Spann, T.P. (2002) Nuclear lamins: building blocks of nuclear architecture, *Genes Dev.* **16**, 533–547.

Grande, M.A., van der Kraan, I., de Jong, L., van Driel, R. (1997) Nuclear distribution of transcription factors in relation to sites of transcription and RNA polymerase II, *J. Cell Sci.* **110**, 1781–1791.

Harrington, K.S., Javed, A., Drissi, H., McNeil, S., Lian, J.B., Stein, J.L., van Wijnen, A.J., Wang, Y.-L., Stein, G.S. (2002) Transcription factors RUNX1/AML1 and RUNX2/Cbfa1 dynamically associate with stationary subnuclear domains, *J. Cell Sci.* **115**, 4167–4176.

Hendzel, M.J., Kruhlak, M.J., MacLean, N.A.B., Boisvert, F.M., Lever, M.A., Bazett-Jones, D.P. (2001) Compartmentalization of regulatory proteins in the cell nucleus, *J. Steroid Biochem.* **76**, 9–21.

Heun, P., Laroche, T., Raghuraman, M.K., Gasser, S.M. (2002) The positioning and dynamics of origins of replication in the budding yeast nucleus, *J. Cell Biol.* **152**, 385–400.

Hozak, P., Jackson, D.A., Cook, P.R. (1994) Replication factories and nuclear bodies: the ultrastructural characterization of replication sites during the cell cycle, *J. Cell Sci.* **107**, 2191–2202.

Iborra, F.J., Cook, P.R. (2002) The interdependence of nuclear structure and function, *Curr. Opin. Cell Biol.* **14**, 780–785.

Izaurralde, E., Kann, M., Pante, N., Sodeik, B., Hohn, T. (1999) Viruses, microorganisms and scientists meet the nuclear pore, *EMBO J.* **18**, 289–296.

Jackson, D.A., Pombo, A. (1998) Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S-phase in human cells, *J. Cell Biol.* **140**, 1285–1295.

Jackson, D.A., Iborra, F.J., Manders, E.M., Cook, P.R. (1998) Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei, *Mol. Biol. Cell* **9**, 1523–1536.

Jenuwein, T., Allis, C.D. (2001) Translating the histone code, *Science* **293**, 1074–1080.

Kimura, H., Sugaya, K., Cook, P.R. (2002) The transcription cycle of RNA polymerase II in living cells, *J. Cell Biol.* **159**, 777–782.

Kingston, R.E., Narlikar, G.J. (1999) ATP-dependent remodeling and acetylation as regulators of chromatin fluidity, *Genes Dev.* **13**, 2339–2352.

Koberna, K., Malinsky, J., Pliss, A., Masata, M., Vecerova, J., Fialova, M., Bednar, J., Raska, I. (2002) Ribosomal genes in focus: new transcripts label the dense fibrillar components and form clusters indicative of "Christmas trees" in situ, *J. Cell Biol.* **157**, 743–748.

Kosak, S.T., Skok, J.A., Medina, K.L., Riblet, R., Le Beau, M.M., Fisher, A.G., Singh, H.

(2002) Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development, *Science* **296**, 158–162.

Kumaran, R.I., Muralikrishna, B., Parnaik, V.K. (2002) Lamin A/C speckles mediate spatial organization of splicing factor compartments and RNA polymerase II transcription, *J. Cell Biol.* **159**, 783–793.

Lei, E.P., Silver, P.A. (2002) Protein and RNA export from the nucleus, *Dev. Cell* **2**, 261–272.

Lei, E.P., Silver, P.A. (2002) Intron status and 3′-end formation control cotranscriptional export of mRNA, *Genes Dev.* **16**, 2761–2766.

Levine, A.J. (1997) p53, the cellular gatekeeper for growth and division, *Cell* **88**, 323–331.

Litt, M.D., Simpson, M., Recillas-Targa, F., Prioleau, M.N., Felsenfeld, G. (2001) Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci, *EMBO J.* **20**, 2224–2235.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 angstrom resolution, *Nature* **389**, 251–260.

Mahy, N.L., Perry, P.E., Gilchrist, S., Baldock, R.A., Bickmore, W.A. (2002) Spatial organization of active and inactive genes and noncoding DNA within chromosome territories, *J. Cell Biol.* **157**, 579–589.

Maniatis, T., Reed, R. (2002) An extensive network of coupling among gene expression machines, *Nature* **416**, 499–506.

Martens, J.H.A., Verlaan, M., Kalkhoven, E., Dorsman, J.C., Zantema, A. (2002) Scaffold/Matrix attachment region elements interact with a p300-scaffold attachment factor A complex and are bound by acetylated nucleosomes, *Mol. Cell Biol.* **22**, 2598–2606.

McNally, J.G., Muller, W.G., Walker, D., Wolford, R., Hager, G.L. (2000) The glucocorticoid receptor: rapid exchange with regulatory sites in living cells, *Science* **287**, 1262–1265.

Misteli, T. (2001) The concept of self-organization in cellular architecture, *J. Cell Biol.* **155**, 181–185.

Moir, R.D., Yoon, M., Khuon, S., Goldman, R.D. (2000) Nuclear lamins A and B1: different pathways of assembly during nuclear envelope formation in living cells, *J. Cell Biol.* **151**, 1155–1168.

Nayler, O., Stratling, W., Bourquin, J.P., Stagljar, I., Lindemann, L., Jasper, H., Hartmann, A.M., Fackelmayer, F.O., Ullrich, A., Stamm, S. (1998) SAF-B protein couples transcription and pre-mRNA splicing to SAR/MAR elements, *Nucl. Acids Res.* **26**, 3542–3549.

Nickerson, J.A. (2001) Experimental observations of a nuclear matrix, *J. Cell Sci.* **114**, 463–474.

Pederson, T. (1998) Thinking about a nuclear matrix, *J. Mol. Biol.* **227**, 147–159.

Pederson, T. (1998) The plurifunctional nucleolus, *Nucl. Acids Res.* **26**, 3871–3876.

Pombo, A., Cook, P.R. (1996) The localization of sites containing nascent RNA and splicing factors, *Exp. Cell Res.* **229**, 201–203.

Pombo, A., Jackson, D.A., Hollinshead, M., Wang, Z., Roeder, R.G., Cook, P.R. (1999) Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III, *EMBO J.* **18**, 2241–2253.

Proudfoot, N.J., Furger, A., Dye, M.J. (2002) Integrating rnRNA processing with transcription, *Cell* **108**, 501–512.

Sacco-Bubulya, P., Spector, D.L. (2002) Disassembly of interchromatin granule clusters alters the coordination of transcription and pre-mRNA splicing, *J. Cell Biol.* **156**, 425–436.

Schaufele, F., Enwright, J.F., Wang, X., Teoh, C., Srihari, R., Erickson, R., MacDougald, O.A., Day, R.N. (2002) CCAAT/enhancer binding protein alpha assembles essential cooperating factors in common subnuclear domains, *Mol. Endocrinol.* **15**, 1665–1676.

Schmiesing, J.A., Ball, A.R., Gregson, H.C., Alderton, J.M., Zhou, S.L., Yokomori, K. (1998) Identification of two distinct human SMC protein complexes involved in mitotic chromosome dynamics, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12906–12911.

Spann, T.E., Goldman, A.E., Wang, C., Huang, S., Goldman, R.D. (2002) Alteration of nuclear lamin organization inhibits RNA polymerase II-dependent transcription, *J. Cell Biol.* **156**, 603–608.

Spector, D.L. (2001) Nuclear domains, *J. Cell Sci.* **114**, 2891–2893.

Sporbert, A., Gahl, A., Ankerhold, R., Leonhardt, H., Cardoso, M.C. (2002) DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters, *Mol. Cell* **10**, 1355–1365.

Stratling, W.H., Yu, F. (1999) Origin and roles of nuclear matrix proteins. Specific functions of

the MAR-binding protein MeCP2/ARBP, *Crit. Rev. Eukaryotic Gene Expr.* **9**, 31–318.

Tavormina, P.A., Come, M.G., Hudson, J.R., Mo, Y.Y., Beck, W.T., Gorbsky, G.J. (2002) Rapid exchange of mammalian topoisomerase II alpha at kinetochores and chromosome arms in mitosis, *J. Cell Biol.* **158**, 23–29.

Taylor, W.R., Stark, G.R. (2001) Regulation of the G2/M transition by p53, *Oncogene* **20**, 1803–1815.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., de Laat, W. (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus, *Mol. Cell* **10**, 1453–1465.

van Driel, R., Wansink, D.G., Van Steensel, B., Grande, M.A., Schul, W., de Jong, L. (1996) Nuclear domains and the nuclear matrix, *Int. Rev. Cytol.* **162A**, 151–189.

Vasu, S.K., Forbes, D.J. (2001) Nuclear pores and nuclear assembly, *Curr. Opin. Cell Biol.* **13**, 363–375.

von Hippel, P.H., Jing, D.H. (2000) Structural biology – Bit players in the trombone orchestra, *Science* **287**, 2435–2436.

Verschure, P.J., van der Kraan, I., Enserink, J.M., Mone, M.J., Manders, E.MM., van Driel, R. (2002) Large-scale chromatin organization and the localization of proteins involved in gene expression in human cells, *J. Histochem. Cytochem.* **50**, 1303–1312.

Wallace, H., Ansell, R., Clark, J., McWhir, J. (2000). Pre-selection of integration sites imparts repeatable transgene expression, *Nucleic Acids Res.* **28**, 1455–1464.

Weis, K. (2003) Regulating access to the genome: Nucleocytoplasmic transport throughout the cell cycle, *Cell* **112**, 441–451.

Welcsh, P.L., Owens, K.N., King, M.C. (200) Insights into the functions of BRCA1 and BRCA2, *Trends Genet.* **16**, 69–74.

Wilson, C., Bellen, H.J., Gehring, W.J. (1990) Position effects on eukaryotic gene expression, *Annu. Rev. Cell Biol.* **6**, 679–714.

Wilson, K.L., Zastrow, M.S., Lee, K.K. (2001) Lamins and disease: insights into nuclear infrastructure, *Cell* **104**, 647–650.

Yasui, D., Miyano, M., Varga-Weisz, P., Kohwi-Shigematsu, T. (2002) SATB1 regulates gene expression over long distances via chromatin remodelling, *Nature* **419**, 641–645.

Cell Signaling During Primitive Hematopoiesis

Devorah C. Goldman and Jan L. Christian
Oregon Health and Science University, Portland, OR, USA

1 Primitive Hematopoiesis 430

2 Specification, Expansion and Differentiation of the Primitive Erythroid Lineage 435
2.1 The Signal to become Ventral 436
2.2 Specifying Primitive Hematopoietic Cell Fate 436
2.3 Regulating the Production, Expansion and Survival of Primitive Erythroid Progenitors 437

3 Role of Cell–Cell Interactions in Primitive Hematopoiesis 442
3.1 Extraembryonic Endoderm 442
3.2 Ectoderm 443
3.3 Endothelial Cells 443
3.4 Role of the Microenvironment in Regulating Primitive and Definitive Potential 444

4 Summary and Perspectives 444

Bibliography 445
Books and Reviews 445
Primary Literature 445

Keywords

Hematopoietic Stem Cell (HSC)
A pluripotent cell that gives rise to all of the blood lineages and has the ability to undergo self-renewal.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Hematopoietic Progenitor
Unipotent or multipotent cells that give rise to committed blood precursors that in turn develop into fully differentiated blood cells.

Primitive Hematopoiesis
The first wave of hematopoiesis that occurs during development that generates large numbers of primitive erythrocytes and some primitive leukocytes.

Definitive Hematopoiesis
The hematopoiesis that permanently replaces primitive hematopoiesis during fetal stages to supply all fetal and adult blood lineages.

Primitive hematopoiesis is the first wave of blood development that occurs in development, and it produces the first red and white blood cells of the embryo. Although primitive hematopoiesis produces a largely transient population of primitive blood cells, some blood progenitors that function later in development to produce definitive blood cells also develop in the primitive hematopoietic site. Studies of mouse, chick, zebrafish, and *Xenopus* embryos have revealed that many of the molecular events that regulate primitive hematopoiesis are conserved among vertebrates. In addition, roles for nonhematopoietic tissues in regulating primitive hematopoiesis have been discovered. In this article, we review what is known about vertebrate primitive hematopoiesis at the tissue, cellular, and molecular levels.

1 Primitive Hematopoiesis

Hematopoiesis is the generation of blood cells (erythrocytes and leukocytes) from a common precursor, the hematopoietic stem cell (HSC). During development, two types of hematopoiesis, primitive and definitive, supply the embryo with blood cells (Fig. 1). The first wave of hematopoiesis is termed primitive because it produces blood cells that are primitive in character in that they do not fully resemble their adult counterparts either functionally or molecularly. Primitive blood cells exist transiently, primarily serving the embryo's early requirement for oxygen, until the second and final wave of hematopoiesis, definitive hematopoiesis, occurs. Definitive hematopoiesis continuously produces all the fetal and adult blood lineages, including the myeloid (red cells, macrophages neutrophils, mast cells, platelets) and lymphoid (B-cells and T-cells) lineages throughout the life of the animal. Although there are anatomical differences in where both waves of hematopoiesis occur among vertebrate species, there is a striking conservation of the genes required for these processes.

The most numerous cells generated during primitive hematopoiesis are primitive erythrocytes, large nucleated cells that express embryonic globins and accumulate

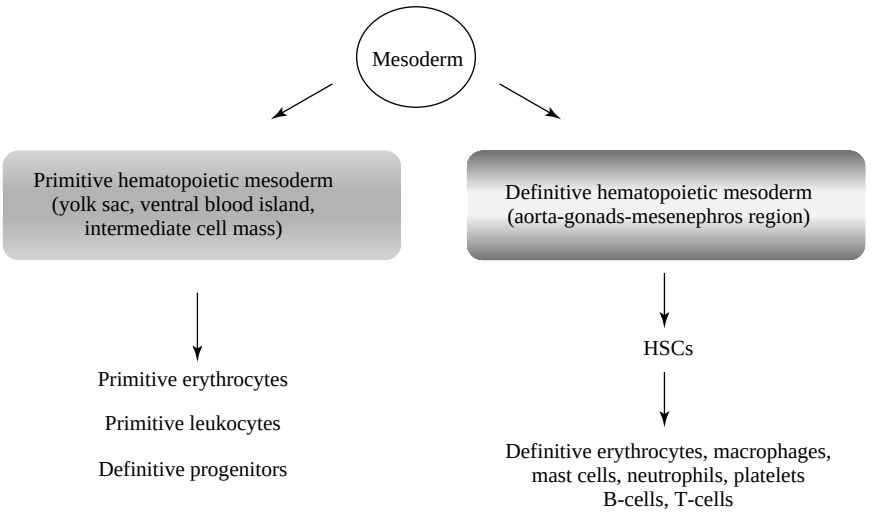

Fig. 1 Embryonic hematopoiesis. This simplified illustration depicts the differentiated cell types that are produced during the first waves of primitive and definitive hematopoiesis, respectively. There are many steps and intermediate cell types in these pathways that are discussed within the text.

hemoglobin during maturation. In contrast, definitive erythrocytes are smaller, contain fetal or adult globins and are enucleated (e.g. in mammals) or have different nuclear morphology (e.g. in *Xenopus*). In all vertebrates, primitive hematopoiesis is initiated during gastrulation in the embryonic or extraembryonic mesoderm, depending upon the species. In mammals and birds, primitive hematopoiesis occurs extraembryonically within the yolk sac mesoderm, whereas in *Xenopus* (African clawed frog) and zebrafish, primitive hematopoiesis occurs embryonically, in a subset of ventrolateral mesoderm known as the *ventral blood islands* and *intermediate cell mass*, respectively. In the mouse, nonhemoglobinized primitive erythroblasts (immature primitive erythrocytes) are first morphologically evident during late gastrulation stages, around embryonic day (E) 7.5 in the yolk sac mesoderm, and they enter the embryo via the circulatory system at E8.25. After E9, primitive erythroid cells are no longer produced by the yolk sac. Within the embryonic blood vessels, primitive erythroid progenitors complete their differentiation in circulation through E13, undergoing mitosis, nuclear condensation, accruement of hemoglobin, and a decrease in cell size. Prior to E12 in the mouse, the only red cells in circulation are derived from the yolk sac. At E12, some liver-derived definitive erythroblasts (that are smaller in size and express adult globins) enter circulation, and within a few days, the primitive erythrocytes are no longer detectable. In the chick, yolk sac mesoderm differentiates into primitive erythroblasts by 1.5 days postfertilization and the erythroblasts enter circulation by day 2, and subsequently mature. In *Xenopus*, ventral blood island-derived primitive erythrocytes enter circulation at stage 35/36, just a few hours after the heart starts beating, and they are the predominant circulating cells until they are replaced by definitive erythrocytes

after metamorphosis. Zebrafish primitive erythroblasts are produced in a synchronous wave in the lateral mesoderm and are morphologically detectable within the intermediate cell mass at 20 hours postfertilization (hpf). Subsequently, they enter embryonic circulation around 24 hpf, where they complete maturation around 5 days postfertilization (dpf). Until 4 dpf, primitive erythrocytes are the only red cells in the blood stream; definitive cells begin to circulate at 5 dpf.

Although primitive hematopoiesis is often defined as the production of primitive erythrocytes, primitive leukocytes are also generated in some species. The exact *in vivo* function of primitive leukocytes has not been well characterized, however. In the mouse, megakaryocyte progenitors are detectable in cultured E7.5 yolk sac cells and primitive megakaryocytes are evident in the yolk sac of intact embryos at E8.5. Primitive megakaryocytes differ significantly from their definitive counterparts by their smaller size, rapid differentiation kinetics, chromosome number, and response to cytokines *in vitro*. It has been suggested that primitive megakaryocytes play a role in maintaining early blood vessel integrity through production of platelets, but this possibility remains to be tested. It is not known whether other vertebrates also produce primitive megakaryocytes. Primitive macrophages are found in all vertebrates and differ significantly from definitive macrophages in many ways, including their phagocytic activity. Electron microscopic studies of the mouse embryo demonstrated the presence of macrophage cells in the yolk sac at E7. Macrophage progenitors are detected in early mouse embryos at the same time and place that primitive erythroid progenitors are first detected, suggesting that the yolk sac is also a source of macrophage progenitors. In *Xenopus*, the ventral blood island gives rise to circulating macrophages but a larger, nonventral blood island–derived macrophage population originates from a region anterior to the heart and migrates interstitially throughout the mesenchyme of the larval body. Recently, it was demonstrated that in zebrafish, a similar population of migrating macrophages develop from a site anterior to the primitive erythroid site. Presumably, macrophages that develop during primitive hematopoiesis function to remove dead cells from the embryo and are involved in tissue remodeling. In summary, although primitive hematopoiesis mostly contributes primitive erythrocytes, it also yields a number of non-erythroid myeloid cells. Further characterization of these primitive myeloid populations will reveal their functions in the early embryo.

Although multiple primitive blood lineages are detected in some vertebrate species, there is no experimental evidence demonstrating that primitive blood cells are derived from HSCs. Markers that uniquely identify HSCs are currently not known; therefore, HSCs are experimentally defined as self-renewing cells that can provide long-term hematopoietic reconstitution of an adult host. Specifically, embryonic HSCs are detected by dissociating embryos or embryo fragments into single cells and transplanting these cells into adults that typically lack their own functional hematopoietic system because they have been irradiated or they carry genetic mutations that impair normal hematopoiesis. If the transplanted embryonic cells contain HSCs, and these HSCs can home to and engraft sites that support hematopoiesis, blood lineages are reconstituted and long-term survival is achieved because the embryo-derived HSCs can self-renew and continuously supply blood

lineage progenitors. In species in which such assays are possible, HSCs are not detected prior to the emergence of primitive progenitors. It is presently unclear, however, whether the inability to detect primitive HSCs is because of experimental limitations or because they simply do not exist.

In addition to primitive erythroid and primitive myeloid progenitors, primitive hematopoietic tissue has also been demonstrated to give rise to definitive progenitors and possibly some definitive HSCs in a species-specific manner. The relationship between primitive hematopoiesis and definitive HSCs in the developing vertebrate embryo, particularly in the mouse, is somewhat contentious. Of some controversy is whether HSCs with definitive potential produced in primitive hematopoietic sites contribute to definitive hematopoiesis *in vivo* or if HSCs with definitive potential are produced *de novo* in the embryo, at sites where primitive hematopoiesis does not occur. Historically, researchers believed that the yolk sac gives rise to HSCs that subsequently seed the embryonic sites of definitive hematopoiesis, the aorta-gonads-mesonephros region and liver, and thus give rise to all definitive lineages. This belief was supported by the observation that hematopoietic cells emerge first in the yolk sac and then later in the liver during embryogenesis. Moreover, it was shown that cultured embryos lacking yolk sacs fail to carry out definitive hematopoiesis. Two subsequent experimental approaches seriously challenged this conclusion. First, it was shown that transplanting quail embryos onto stage-matched precirculatory yolk sacs (Hamburger-Hamilton [HH] stage 9) from chick embryos and allowing them to develop *in vivo* resulted in quail fetuses in which all definitive hematopoietic cells were derived from the quail embryo. The results from this experiment incontrovertibly ruled out any yolk sac contribution to definitive hematopoiesis after HH stage 9 in the chick. Strikingly, in the mouse, HSCs are detected after E10 (~three days after primitive erythroid cells are first detected), first in the aorta-gonads-mesonephros region and subsequently in the yolk sac. Furthermore, fragments of mouse embryo including aorta-gonads-mesonephros mesodermal precursors, isolated from precirculatory embryos, can fully reconstitute a genetically deficient adult host if the precursor-containing fragment is maintained in culture prior to transplantation. Together, these data suggest that the aorta-gonads-mesonephros region is the source of definitive HSCs, and not the yolk sac. Moreover, it has been suggested that some of these aorta-gonads-mesonephros-derived HSCs enter the yolk sac via the circulation, thus explaining the presence of HSCs with definitive potential found later in the yolk sacs of chick and mouse around the time that primitive hematopoiesis diminishes.

Although these and other experiments have led many to conclude that the yolk sac is not a source of definitive HSCs, these data do not completely rule out that there is a primitive contribution to definitive HSCs in the fetus. The failure to detect yolk sac HSCs in mouse embryos before E11 might be explained if HSCs are present at this time, but are not yet competent to respond or home to the adult hematopoietic microenvironment when they are transplanted into adult hosts. For example, definitive HSCs can be detected in E9 – 10 yolk sacs when the yolk sac cells are transplanted into chemically myeloablated livers of newborn hosts. Grafted yolk sac cells are capable of supporting hematopoiesis in the recipients for

nearly a year after transplant. Detection of HSCs from E8 precirculatory murine yolk sacs after a period of coculture with a stromal cell line has also been observed, lending further support to the idea that definitive HSCs are produced in the same location and during the same period of embryogenesis as are primitive progenitors. Interestingly, mouse embryonic stem cells (pluripotent embryo-derived stem cells that are believed to faithfully recapitulate primitive and definitive hematopoiesis *in vitro*) cultured under conditions that promote hematopoietic specification and differentiation, give rise to clonal colonies (i.e. colonies derived from a single cell) that contain both primitive erythrocytes and definitive hematopoietic cells. These data lend some credence to the idea that primitive and definitive hematopoietic cells share a common origin. These data also point to the need for better methods to detect HSCs in embryos. In particular, identifying groups of genes that are expressed and function specifically in HSCs would tremendously aid in HSC detection.

A second possible explanation for the failure to detect yolk sac-derived HSCs is that assays that have been used to assess the origins of HSCs, such as culturing fragments of precirculatory embryos, do not address the possibility that HSCs may be generated in the early yolk sacs of chicks and mice and then migrate interstitially into the embryo prior to the onset of circulation. Interestingly, at the same time that morphological and molecular evidence of hematopoietic development is observed in mouse embryos (E7–7.5), primordial germ cells are also formed in the yolk sac. These primordial germ cells migrate into the embryo before circulation begins, reaching the gonads (in the aorta-gonads-mesonephros region) later in development. Perhaps an early population of HSCs migrates into the embryos with the primordial germ cells. To adequately address the issue of whether the murine yolk sac contributes to functional, definitive HSCs, a reliable method for performing stage-specific *in vivo* lineage analysis of mouse yolk sac derivatives is required. Unfortunately, this will not be a trivial task given the difficulty of *in utero* transplantation, the inability to culture whole mouse embryos past E10.5 and the current lack of molecular tools to direct expression of heritable lineage markers in putative yolk sac HSCs. Until suitable technologies that allow for directly testing the fate of yolk sac hematopoietic cells are created, this issue will likely remain unresolved.

Whereas the presence of definitive HSCs during primitive hematopoiesis is still debated, it is well accepted that some definitive progenitors are generated in the primitive hematopoietic site, also in a species-specific manner. For example, isolated yolk sac cells from mouse embryos can give rise to definitive erythrocytes and other definitive hematopoietic cell types under appropriate culture conditions. The kinetics with which these committed hematopoietic progenitors emerge in the mouse yolk sac, and subsequently in the aorta-gonads-mesonephros and liver suggest that progenitors generated in the yolk sac enter circulation, seed the fetal definitive hematopoietic sites and contribute differentiated definitive blood cells during fetal development. Similarly in *Xenopus*, detailed fate mapping of the ventral blood islands revealed that it contributes not only to primitive erythrocytes, but also to a small but significant number of thymocytes (T-cells) and B-cells in premetamorphic animals. A small contribution of ventral blood island cells to definitive erythrocytes and

thymocytes is also detected postmetamorphosis, when the bulk of definitive hematopoietic cells are derived from the dorsolateral plate. Thus, cells derived from the site of primitive hematopoiesis appear to make at least a temporary contribution of definitive progenitors to the developing embryo.

In summary, primitive hematopoiesis occurs in a unique site and produces the first circulating red cells and leukocytes of the developing embryo. Although differentiated primitive blood cells play only a transient role in development, numerous lines of evidence point to a lasting contribution of definitive progenitors and, potentially, some definitive HSCs that are generated in the primitive hematopoietic site. In the following sections, we will review what is known about the molecular regulation of primitive hematopoiesis including the specification and maturation of primitive hematopoietic cells as well as tissue interactions that influence these events.

2 Specification, Expansion and Differentiation of the Primitive Erythroid Lineage

As discussed above, one of the main functions of primitive hematopoiesis is the generation of large numbers of primitive erythrocytes. In this section, we will briefly review what is known about the molecular regulation of primitive erythropoiesis, including the signals that regulate ventral patterning of the mesoderm, from which primitive blood is derived, the specification of hematopoietic fate within this mesoderm, and the subsequent production, expansion, survival, and differentiation of progenitors into primitive erythrocytes (summarized in Fig. 2). We have elected to focus on primitive erythrocyte differentiation because relatively

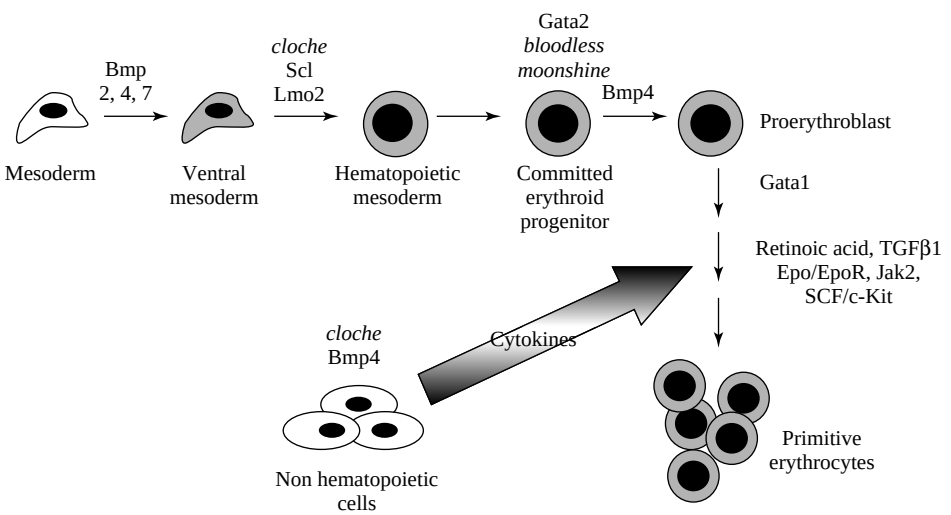

Fig. 2 Primitive erythropoiesis. Genes required for different steps of the erythroid development pathway are shown in plain text, and zebrafish mutants that block erythroid development at various steps are shown in italicized text. For details, see text in Sects. 2 and 3.

little is currently known about primitive leukocyte development.

2.1 The Signal to become Ventral

Primitive and definitive hematopoietic cells are derived from mesoderm, the germ layer of the embryo that gives rise to notochord, kidney, muscle, bone, blood, blood vessels, and heart. Early in development, mesoderm is specified and is subsequently patterned along the dorsal/ventral (D/V; back/belly) axis. The dorsalmost mesoderm is the notochord, whereas the ventralmost mesoderm is the primitive blood. D/V patterning of mesoderm is largely regulated by bone morphogenetic proteins (BMPs), secreted signaling molecules of the transforming growth factor-β (TGF-β) superfamily. Numerous BMPs have been identified in vertebrates; however, existing evidence suggests that *bmp2*, *4*, and *7* are most likely to participate in D/V patterning of nascent mesoderm.

BMPs are proposed to act as morphogens in patterning the mesoderm. A morphogen is a substance that is present in a gradient and elicits a range of specific developmental outcomes (mesoderm patterning in this case) in a concentration-dependent manner. Convincing evidence in support of this theory comes from studies of *Xenopus* and zebrafish embryos in which BMP signaling is incrementally increased or decreased. Increasing the dose of BMP4 in developing embryos (by injecting increasing amounts of *bmp4* mRNA into embryos) leads to formation of mesoderm with increasing ventral character at the expense of dorsal mesoderm. Conversely, decreasing the amount of BMP4 signaling (by injecting increasing amounts of mRNA encoding a dominant-negative BMP receptor) leads to the production of increasingly dorsal mesoderm at the expense of ventral mesoderm. Thus, high levels of BMP signaling specify the most ventral fate (primitive blood), lower levels of BMP signaling specify an intermediate cell fate (pronephros, the functional equivalent of kidney, and definitive blood), and low or absent levels of BMP signaling specify the most dorsal cell fate (notochord, somites) within the mesoderm. Similarly, in zebrafish, a series of BMP signaling mutants has been generated from large-scale mutagenesis screens. Those embryos with the most severe losses of BMP signaling display the most dorsalized mesodermal phenotypes whereas those with the most active BMP signaling display the most ventralized phenotypes. In mouse, *bmp4*, but not *bmp2* or *bmp7*, appears to play a role in the specification of ventral mesoderm. In the absence of *bmp4*, ventral derivatives, including primitive blood and endothelium, fail to form in the yolk sac. Many questions remain as to how BMPs regulate mesodermal pattern. For instance, the direct transcriptional targets of the BMP signaling pathway have not been well characterized. In addition, it is not yet clear how quantitative changes in BMP signaling differentially regulate expression of these target genes. Furthermore, the timing of BMP action on D/V patterning of each mesodermal type during early development and whether BMP acts on each mesodermal type directly or indirectly (for instance, through signaling in surrounding tissues) is not well characterized.

2.2 Specifying Primitive Hematopoietic Cell Fate

After the nascent mesoderm has been patterned along the D/V axis, a subset of the ventral mesoderm is specified as

hematopoietic. Gene targeting in mouse (i.e. the creation of a mouse in which a gene is "knocked-out" or mutated) has identified genes that are required for this process. One such gene is the basic helix–loop–helix transcription factor *scl/tal1*, which is expressed in the nascent ventral mesoderm, prior to overt differentiation of hematopoietic cells in all vertebrates. Mice that are homozygous for a nonfunctional allele of *scl* fail to generate any primitive erythrocytes in otherwise normally patterned embryos. Analysis of gene expression in these mutant embryos, as well as hematopoietic culture assays using mutant tissues and cell lines, demonstrated that *scl* is required for both primitive and definitive hematopoiesis. Consistent with these data, overexpression of *scl* in zebrafish embryos leads to excessive commitment of mesoderm to a primitive hematopoietic fate. Similarly, overexpression of *scl* in dorsal mesoderm of *Xenopus* embryos leads to ectopic production of primitive erythroid cells without otherwise altering the dorsal fate of this tissue. Collectively, these data are consistent with the hypothesis that *scl* functions after the induction and ventral patterning of the mesoderm to specify primitive hematopoietic fate.

The transcription factor *lmo2*, which contains LIM protein interaction domains but no DNA binding domain, is expressed in the prospective hematopoietic mesoderm in a pattern similar to *scl*. Targeting of *lmo2* in mouse confirms a role for *lmo2* in specification of primitive and definitive hematopoietic fate. Overexpression of *lmo2* in *Xenopus* is not sufficient to induce ectopic primitive hematopoiesis, whereas, coexpression of high levels of *lmo2* and *scl* can induce ectopic primitive erythrocytes, suggesting both factors must be present for a cell to adopt a hematopoietic fate. Physical interactions between *lmo2* and *scl* have been detected in mammalian erythroid cell lines, suggesting that *scl* and *lmo2* may form transcriptional complexes that regulate genes required for primitive hematopoiesis.

The regulatory steps that occur between ventralizing the mesoderm and expression of *scl* and *lmo2* are extremely important, but little is known about them. In zebrafish, a spontaneous mutant, *cloche*, has been identified that may reveal much about hematopoietic specification during primitive hematopoiesis. In *cloche* mutants, expression of *scl* and *lmo2* is absent or severely reduced and the intermediate cell mass contains no morphological evidence of the presence of erythroid progenitors. The gene encoded by *cloche* appears to act upstream of *scl* since forced expression of *scl* in *cloche* mutants leads to partial rescue of primitive erythrocytes. Thus, identifying the gene that encodes *cloche* will be a critical step towards gaining a complete understanding of the early events of hematopoietic development, including the regulation of genes required for hematopoietic specification.

2.3 Regulating the Production, Expansion and Survival of Primitive Erythroid Progenitors

Hematopoietic mesoderm gives rise to multipotent and unipotent progenitor cells that must make decisions regarding whether they should proliferate, undergo apoptosis, or differentiate. Each of these decisions is tightly regulated by a combination of extrinsic and intrinsic signals. In the mouse, the cytokine *vegf* and its tyrosine kinase receptor *vegfr2/flk1* have been shown to be required at an early stage for the initial production of primitive erythroid progenitors. Mice mutant

for *vegfr2* lack all blood and blood vessels, possibly because of failure to recruit blood and endothelial progenitors to the correct site where they can receive signals necessary to commit to an erythroid fate. VEGF may also be required for progenitor expansion since cultures of *vegfr2*$^{-/-}$ ES cells generate reduced total numbers of hematopoietic progenitors relative to wild type cells. In zebrafish, *vegfr2* mutants produce normal numbers of primitive erythrocytes, suggesting that either VEGF signaling is not required for hematopoiesis, or that another *vegfr2* locus exists. The latter possibility is feasible as many genes are duplicated in the zebrafish genome. In the chick, VEGFR2 is expressed in prospective hematopoietic mesoderm, and it has been demonstrated that VEGFR2 signaling is required for hematopoietic colony formation in cultures of prospective yolk sac mesoderm. However, addition of VEGF to the culture media does not promote hematopoietic colony formation, suggesting that a different member of the VEGF family or a related cytokine may bind to VEGFR2 and promote hematopoiesis.

A second gene product that is required for primitive erythropoiesis is *gata2*, a member of the GATA family of zinc finger transcription factors. *gata2* is expressed in all vertebrate embryos in a pattern consistent with a role in early hematopoiesis and *gata2*$^{-/-}$ mouse embryos contain two- to sevenfold fewer circulating primitive erythrocytes. Culture of mouse *gata2*$^{-/-}$ embryonic stem cells (ES cells) and yolk sac cells in conditions that favor the expansion of hematopoietic progenitors (both primitive and definitive), revealed that lack of *gata2* caused a decrease in the total number of all progenitors, including erythroid progenitors. *gata2* expression and/or activity may be regulated by cytokine stimulation, as numerous cytokines can induce the expansion of HSCs and progenitors *in vitro* and *in vivo*. Although *gata2* is required for progenitor expansion, a subsequent reduction in *gata2* expression in committed erythroid progenitor cells is required for their differentiation.

Another member of the GATA family, *gata1*, is expressed at relatively early stages of hematopoietic development in all vertebrates, but its function is only required after erythroid progenitor cell commitment as evidenced by differentiation defects in animals that lack functional *gata1*. Numerous genes expressed in developing and mature erythrocytes such as *scl*, and embryonic *globin* genes contain GATA sites in their promoters/enhancers. Mice homozygous for null alleles of *gata1* seemingly lack all circulating primitive erythrocytes. Closer inspection of peripheral blood in these mutants revealed that nonhemoglobinized proerythroblasts are present and that they arrest in development and undergo apoptosis. Similarly, in zebrafish, a point mutation in the *gata1* gene that results in deletion of 79 C-terminal amino acids has been identified as the molecular basis of the *vlad tepes* bloodless phenotype. In these mutants, proerythroblasts do not enter circulation and hematopoietic cells in the intermediate cell mass (IM) do not express terminal erythroid genes.

In the mouse, homozygous mutation of the friend of GATA gene (*fog1*), a zinc finger transcription factor that physically interacts with *gata1*, arrests primitive erythroblast differentiation at the same stage of differentiation as loss of *gata1*. Fog1 and Gata1 protein interaction has been shown to be critical for definitive erythroid cell line differentiation and this interaction requires the presence of the N-terminal zinc finger of Gata1. Consistent with these observations, mice in which the *gata1* gene

has been replaced with a version of *gata1* that contains a single amino acid change that prevents binding to *fog1*, arrest at the same stage of development as *gata1*$^{-/-}$ embryos, suggesting that interactions between Gata1 and Fog1 are a requisite for normal primitive erythroid development. Numerous erythroid cell line studies have also shown that *scl* regulates definitive erythropoiesis through dynamic association with other transcription factors, including the formation of DNA binding complexes containing Lmo2 and Gata1. The recent creation of an adult hematopoietic-specific knockout of *scl in vivo* demonstrates an absolute requirement for *scl* in adult definitive erythroid lineage development. Persistent *scl* expression in erythroid progenitors and circulating primitive erythroblasts suggests that *scl* plays a similar role in primitive erythrocyte differentiation.

A number of zebrafish mutants that lack, or have a severely reduced number of circulating primitive erythroid cells have been identified. Among these mutants, *bloodless* and *cloche* appear to affect progenitor survival or expansion. The *bloodless* mutation acts as a noncell autonomously since wild type cells transplanted into *bloodless* mutant embryos cannot undergo primitive hematopoiesis. This raises the possibility that *bloodless* regulates or encodes a cytokine. In *bloodless* mutants, the onset of *scl* expression is detected at the appropriate time and place in the hematopoietic mesoderm; however, only a few cells continue to express *scl* in the intermediate cell mass and apoptosis is detected at later stages. These data suggest that although hematopoietic cells are specified, they cannot survive. The *cloche* mutation may also regulate progenitor expansion. Transplanted *cloche* mutant cells in wild type embryo hosts can occasionally contribute to the primitive erythroid lineage, but far less efficiently than similarly transplanted wild type cells, suggestive of a defect in erythroid progenitor proliferation. Identifying the genes encoded by *bloodless* and *cloche* will greatly enhance our understanding of how progenitors are maintained.

A second zebrafish mutant, *moonshine*, fails to undergo erythroid progenitor commitment. In *moonshine* mutants, *gata2* expression appears normal in hematopoietic mesoderm, indicating that uncommitted hematopoietic progenitors are present. However, *gata1* expression is absent suggesting that committed erythroid progenitors do not develop in the absence of the *moonshine* gene. The gene encoded by *moonshine* has not yet been reported, and its characterization will undoubtedly provide insights into how hematopoietic gene expression is regulated. Other approaches to identifying genes that are critical in primitive erythroid development have been fruitful. In the mouse, a critical role for BMP4 in erythroid progenitor commitment has been revealed through studies of primitive erythroid differentiation in cultured ES cells. Specifically, hematopoiesis can be blocked at the uncommitted progenitor stage when cells are cultured in the absence of serum. The addition of BMP4 to culture media promotes the expression of *gata1* and primitive erythrocyte differentiation, whereas blocking BMP signaling in erythroid cells by expressing a dominant-negative BMP receptor with an erythroid-specific promoter prevents *gata1* expression and erythroid differentiation. Thus, both *moonshine* and BMP signaling are required for *gata1* expression and for primitive erythrocyte maturation.

Reductions in circulating primitive erythrocytes have been observed in mouse embryos that are defective in some cytokine signaling pathways.

Historically, cytokine function in primitive hematopoiesis has largely been ignored or dismissed, most likely because only a subset of differentiating primitive erythrocytes fail to survive in known cytokine signaling mutants. Moreover, in many cases, the embryos can survive to stages where more drastic effects on definitive hematopoiesis are observed. The erythropoietin receptor (EpoR), a member of the cytokine receptor superfamily, is expressed in the early mouse yolk sac hematopoietic mesoderm, before erythroblasts are detectable, and is also expressed in developing primitive erythroid cells at later stages of development. In yolk sac cultures, addition of EPO to the culture media increases the number of primitive erythrocytes. Conversely, introduction of antisense oligonulceotides that block *epoR* expression reduce the number of differentiated primitive erythrocytes. Mice that are homozygous null for the *epoR* gene apparently initiate hematopoietic development normally, as normal numbers of erythroblasts are present in the yolk sac at E9.5 (at this stage about half of the erythroblasts have entered circulation). However, by E10.5, circulating primitive erythroid cells are reduced by ~80%. In the remaining circulating erythroid cells, differentiation appears to occur normally, suggesting that Epo signaling is specifically required for erythroblast proliferation and/or survival. A two- to threefold reduction in circulating erythrocytes has also been reported in mouse E12 embryos that do not express *c-kit*, a tyrosine kinase receptor that binds stem cell factor/kit ligand. Analysis prior to E12 has not been reported, but a decrease in circulating blood erythrocytes at this stage may indicate a deficit in primitive erythroid survival or proliferation. The Janus Kinase (Jak) family of cytoplasmic tyrosine kinases mediates signaling through cytokine superfamily receptors, including EpoR as well as tyrosine kinase receptors, including c-Kit. A significant reduction in the number of circulating primitive erythrocytes in $jak2^{-/-}$ mice has been reported. Finally, TGF-β1 appears to also play a role in the proliferation of primitive erythroid progenitors. In some $tgf\text{-}\beta 1^{-/-}$ embryos, a large reduction in circulating erythroid cells is observed after E9.5. Together, these data not only highlight the importance of cytokines in primitive erythroid cells but also suggest that some profound differences in cell signaling exist between primitive and definitive hematopoiesis. A number of zebrafish mutants that similarly arrest primitive erythroid development after erythroblasts have entered the bloodstream have been identified and characterized. In these mutants, normal numbers of erythroblasts enter circulation, but they subsequently fail to proliferate, and in some cases, they also fail to maintain hemoglobin expression and do not fully differentiate.

Retinoic acid may also play an important, concentration-dependent role in primitive erythroid cell differentiation. Although numerous retinoic acid receptors have been knocked-out in the mouse without any major effects on primitive or definitive hematopoiesis, increases or decreases in retinoic acid do affect primitive erythropoiesis in *Xenopus* and quail embryos, respectively. For example, incubating *Xenopus* embryos in retinoic acid leads to defects in primitive erythrocyte differentiation and survival without an apparent effect on hematopoietic specification. Vitamin A-deficient quail embryos maintain expression of *gata1* in yolk sac-derived erythroblasts, but form fewer primitive

erythrocytes than wild type embryos and these erythrocytes fail to hemoglobinize. At earlier stages, an increase in apoptosis is detected in hematopoietic cells. These defects can also be mimicked in whole embryo culture by Noggin (a BMP2, −4 antagonist) and primitive erythroid development can be rescued in vitamin A-deficient embryo explants by the addition of BMP4, suggesting that BMP4 is critical for erythroblast survival and differentiation.

Additional evidence for the role of BMPs in regulating primitive erythrocyte survival and differentiation has recently been revealed by altering BMP signaling in *Xenopus* embryos. When BMP signaling is blocked specifically in the ectoderm of developing embryos, by overexpression of RNA encoding a dominant-negative BMP receptor, mesodermally derived erythroid progenitors undergo programmed cell death and very few mature primitive erythrocytes are detected in otherwise normal embryos. Thus, BMP4 signaling is required in ectodermal cells in order to generate an essential survival factor for the erythroid lineage. Upregulation of BMP signaling in the ectoderm by microinjection of RNA encoding a constitutively active BMP receptor also reduces the number of circulating primitive erythrocytes; however, apoptosis does not appear to be responsible for this decrease. Remarkably, a concomitant increase in circulating macrophages is also observed. As previously discussed, the *Xenopus* ventral blood island does make a significant contribution to circulating macrophages in embryos, suggestive of the presence of either multilineage myelopoietic progenitors or macrophage progenitors in the ventral blood island. These results raise the intriguing possibilities that BMP signaling regulates lineage commitment (i.e. the choice to become erythroid or nonerythroid myeloid cells) or specifically regulates the development of the myeloid lineage. Unfortunately, cell culture assays that are routinely used to assay the potential of mammalian and avian hematopoietic progenitors *in vitro* are not technically feasible with amphibian tissues. As discussed in greater detail in the next section, BMPs appear to act indirectly on the ventral blood island, most likely via the production of secondary signals in the ectoderm.

To summarize, numerous transcription factors, cytokines, and signaling molecules have been identified that have essential functions in the development of primitive erythroid cells (Fig. 2). First, mesoderm is ventralized through the actions of BMPs. Next, mesoderm is specified to adopt a primitive hematopoietic fate by processes that are not currently understood, but may be better elucidated once the genetic basis of the *cloche* mutant phenotype is revealed. The production of erythroid progenitors requires the transcription factors *scl* and *lmo2*, whereas maintenance of progenitors requires the transcription factor *gata2*. Signaling molecules that regulate their expression are presumed to exist, but have not been identified. The decision to commit to an erythroid fate is regulated together by *gata1* and *fog1*, and later roles for these and other transcription factors in erythroid differentiation are likely. Expression of *gata1* is regulated at the transcriptional level by cytokines such as BMP4. Cytokines such as BMP4 and EPO have also been shown to play an essential role in proliferation and survival. Numerous zebrafish mutants that block erythroid progenitor commitment and that also fail to maintain normal numbers of circulating primitive erythrocytes have been isolated. Determining which genes

are mutated in these lines will yield insights into how erythroid commitment and survival is achieved, as well as how erythroid-specific gene expression is regulated.

3 Role of Cell–Cell Interactions in Primitive Hematopoiesis

Cell–cell interactions are responsible for regulating many embryonic developmental events. Secreted signaling molecules produced by both hematopoietic and nonhematopoietic cells play a critical role in primitive hematopoiesis through effects on specification, migration, survival, differentiation, and regulating primitive and definitive potential. In the embryo, morphogenetic movements result in hematopoietic cells contacting different cell types, and thus, many microenvironments that potentially influence their development. In this section, we will review tissue–cell interactions that are known to regulate primitive hematopoiesis.

3.1 Extraembryonic Endoderm

During gastrulation in the mouse and the chick, prospective hematopoietic mesoderm migrates into the extraembryonic region, where it comes into direct contact with the extraembryonic endoderm (known in the mouse as *visceral endoderm*, or VE). Studies of cultured mouse ES cells and cultured early-mid gastrula stage mouse embryos have revealed a role for the VE in normal hematopoiesis. For example, hematopoiesis is impaired in embryoid bodies (ES cells that differentiate into structures that resemble mouse gastrulae) that are rendered VE-deficient by a genetic mutation. Likewise, removal of the VE from mid-gastrula stage mouse embryos results in their subsequent failure to specify hematopoietic precursors. Indian hedgehog (Ihh), a member of the Hh family of signaling molecules, is expressed in the VE and can substitute for VE function in hematopoietic development. Although Ihh is sufficient to substitute for the VE-derived signal, it is not required since *Ihh* mutant embryos do not display defects in blood development, possibly owing to genetic redundancy. The effect of Ihh on cultured VE-deficient embryos depends on BMP4 signaling, raising the possibility that VE-derived Ihh functions to maintain early BMP signaling, thereby allowing appropriate hematopoietic specification to occur. A similar requirement for a diffusible extraembryonic endoderm signal in chick primitive erythropoiesis has also been demonstrated, yet the identity of this signal is not known. The VE may also play an additional role in directing the migration of hematopoietic precursors since these cells are the source of *vegf* mRNA. As described earlier, VEGF and related ligands are required for the migration of prospective hematopoietic cells to a site within the yolk sac that is permissive for hematopoietic development. These findings demonstrate that VE-derived signals are essential for the initiation of primitive hematopoiesis, yet their role appears to be short-lived. At the time that morphologically distinct hematopoietic cells can be detected in the mouse yolk sac, isolated yolk sac hematopoietic mesoderm (which includes hematopoietic cells as well as endothelial cells and smooth muscle cell precursors) can differentiate autonomously in culture without the need for any exogenous factors.

3.2 Ectoderm

Numerous lines of evidence point to an essential role of ectodermally derived signals in *Xenopus* primitive erythroid development. Similar to the mouse and chick, gastrulation movements in *Xenopus* embryos bring the prospective hematopoietic tissue into close contact with a different germ layer, the ectoderm. Isolated and cultured prospective ventral mesoderm (posterior ventral blood island precursors) from early gastrula stage embryos develop "blood-like cells" that are not hemoglobinized and do not express globin protein. However, coculture of the ventral mesoderm fragments with fragments of prospective ectoderm from stage-matched embryos leads to robust differentiation of primitive erythrocytes. Disruption of BMP signaling in the prospective dorsal or ventral ectoderm significantly reduces the number of circulating primitive erythrocytes in animals without affecting D/V patterning. Importantly, in embryos in which BMP signaling is disrupted in the prospective ectoderm, specification of hematopoietic cells, as evidenced by *scl* expression, appears to occur normally. Moreover, normal production of erythroid progenitors occurs, as evidenced by wild type levels of *gata1* at early stages of ventral blood island development. However, at subsequent developmental stages, *gata1* expression is not maintained, *globin* expression is absent or significantly reduced, and the number of circulating primitive erythrocytes is severely decreased. In other words, blocking the ability of ectodermal cells to respond to BMPs leads to defects in primitive erythroid differentiation, whereas the specification of primitive erythroid progenitors appears to occur normally. These results strongly suggest that BMP signaling in the ectoderm induces the production of a secondary signal that acts on the prospective hematopoietic tissue to support the survival of primitive erythrocytes. It has been proposed that this secondary signal is a BMP because blocking production of biologically active BMP2, 4, and 7 in the prospective ventral ectoderm also inhibits primitive erythrocyte development, and BMPs positively regulate their own expression. Thus, in theory, inhibiting BMP signaling in the prospective ectoderm could abrogate expression of endogenous BMPs that signal to the hematopoietic mesoderm. Because BMP expression is normally extinguished in the prospective dorsal ectoderm during gastrulation, and inhibiting BMP signaling in the prospective dorsal ectoderm is sufficient to block primitive erythrocyte differentiation, the secondary signal generated by the ectoderm that signals to the hematopoietic progenitors is most likely to include a cytokine or cytokines other than or in addition to BMPs.

3.3 Endothelial Cells

Endothelial cells differentiate in close physical association with primitive hematopoietic cells in all vertebrate species and some evidence suggests that endothelial cells support the differentiation of the primitive erythroid lineage. In the zebrafish, for example, transplantation of wild type cells into cloche mutants showed that in the mutant environment, wild type cells do not maintain *gata1* expression. Given that the cloche mutation also leads to the absence of endothelial cells, it has been suggested that the noncell autonomous effects on *gata1* maintenance

in erythroid progenitors is a result of loss of endothelial cells, which may normally produce factors that support *gata1* expression. Interestingly, coculture of yolk sac hematopoietic cells with an endothelial cell line significantly enhances hematopoietic cell proliferation. Thus, endothelial cells that are in close contact with developing primitive hematopoietic cells are excellent candidates for supplying critical factors that support primitive hematopoiesis.

3.4 Role of the Microenvironment in Regulating Primitive and Definitive Potential

Numerous lines of evidence suggest that the microenvironment determines the primitive or definitive potential of HSCs and progenitors. For instance, HSCs from an adult mouse can be "reprogrammed" to produce primitive cells by aggregating them to mouse blastocysts that are then allowed to develop normally *in utero*. In the resulting chimeric embryos, the HSCs (or cells derived from them) colonize the yolk sac and produce erythrocytes that express embryonic *globin*. In *Xenopus*, heterotopic transplantation of the dorsolateral plate (which contains definitive hematopoietic precursors) to the ventral blood island region results in dorsolateral plate-derived circulating primitive erythrocytes. Conversely, transplanting the ventral blood island to the dorsolateral plate region suppresses production of primitive erythrocytes by the ventral blood island. What regionally determines primitive versus definitive fate? Nearby tissues undoubtedly play a role through their production of extracellular signaling molecules. By using organisms such as *Xenopus* and chick that are experimentally amenable to tissue transplantation studies, it should be possible to identify the tissues that influence primitive and definitive commitment. What regulates primitive versus definitive fate intracellularly? It has recently been shown that forced expression of *hoxb4*, a homeodomain transcription factor, can reprogram primitive progenitors to produce definitive hematopoietic cells. Moreover, numerous *Hox* genes are expressed in definitive but not primitive HSCs and progenitors. Perhaps screens for secreted factors that either suppress or enhance *hoxb4* expression in hematopoietic cells may identify factors that influence primitive and definitive commitment, respectively. At the very least, determining what genes are regulated by *hoxb4* should provide some insight into how definitive potential is achieved at the expense of primitive potential.

4 Summary and Perspectives

Although primitive hematopoiesis is transient, it plays an important role in development by providing the first erythrocytes and leukocytes of the embryo. It remains to be determined if the primitive hematopoietic site contributes long-lasting HSCs with definitive potential; however, it is clear that at least in some species, some definitive hematopoietic progenitors originate there. Numerous genes, including transcription factors, secreted signaling molecules and signal transduction machinery have been shown to play a crucial role in the development of the primitive lineage through loss of function analysis in mouse, gain and loss of function studies in *Xenopus* and chick, and mutagenesis screens in zebrafish. Undoubtedly, additional genes that regulate many aspects of blood development will

be identified in zebrafish blood development mutants, now that the zebrafish genome is nearly completely sequenced. Embryological "cut and paste" experiments as well as perturbing signaling pathways in nonhematopoietic cells by numerous means have further revealed a complex and dynamic role for cell–cell signaling in regulating hematopoietic cell specification, potential (definitive or primitive), cell survival, and differentiation. In spite of a tremendous amount of recent progress in understanding how primitive hematopoiesis is regulated, there is still a long way to go. First, the molecular profile of primitive progenitors, and precursors at discrete stages of the differentiation process must be characterized. With the availability of techniques such as gene expression profiling, it should be possible to identify unique markers for each cell type that will allow for identifying and studying them *in situ*. Moreover, it will help identify genes that determine each cell type's characteristics. Second, a better characterization of cell–cell interactions will help identify factors that are required for various aspects of primitive hematopoiesis. Ultimately, the goal is to understand all of the signaling pathways that regulate primitive hematopoiesis. Once genes that regulate primitive hematopoiesis are identified, determining how they function within regulatory pathways will be an essential step in gaining a complete understanding of hematopoiesis.

See also Biological Regulation by Protein Phosphorylation; Cellular Interactions; Developmental Cell Biology.

Bibliography

Books and Reviews

Amatruda, J.F., Zon, L.I. (1999) Dissecting hematopoiesis and disease using the zebrafish, *Dev. Biol.* **216**, 1–15.

Baron, M. (2001) Induction of embryonic hematopoietic and endothelial stem/progenitor cells by hedgehog-mediated signals, *Differentiation* **68**, 175–185.

Evans, T. (1997) Developmental biology of hematopoiesis, *Hematol. Oncol. Clin. North Am.* **11**, 1115–1147.

Galloway, J.L., Zon, L.I. (2003) Ontogeny of hematopoiesis: examining the emergence of hematopoietic cells in the vertebrate embryo, *Curr. Top. Dev. Biol.* **53**, 139–158.

Orkin, S.H. (1996) Development of the hematopoietic system, *Curr. Opin. Genet. Dev.* **6**, 597–602.

Orkin, S.H., Zon, L.I. (2002) Hematopoiesis and stem cells: plasticity versus developmental heterogeneity, *Nat. Immunol.* **3**, 323–328.

Palis, J., Segel, G.B. (1998) Developmental biology of erythropoiesis, *Blood Rev.* **12**, 106–114.

Palis, J., Yoder, M.C. (2001) Yolk-sac hematopoiesis: the first blood cells of mouse and man, *Exp. Hematol.* **29**, 927–936.

Shepard, J.L., Zon, L.I. (2000) Developmental derivation of embryonic and adult macrophages, *Curr. Opin. Hematol.* **7**, 3–8.

Yoder, M.C. (2002) Embryonic hematopoiesis in mice and humans, *Acta Paediatr. Suppl.* **91**, 5–8.

Primary Literature

Baron, M. (2001) Induction of embryonic hematopoietic and endothelial stem/progenitor cells by hedgehog-mediated signals, *Differentiation* **68**, 175–185.

Belaoussoff, M., Farrington, S.M., Baron, M.H. (1998) Hematopoietic induction and respecification of A-P identity by visceral endoderm signaling in the mouse embryo, *Development* **125**, 5009–5018.

Bertwistle, D., Walmsley, M.E., Read, E.M., Pizzey, J.A., Patient, R.K. (1996) GATA factors and the origins of adult and embryonic blood in Xenopus: responses to retinoic acid, *Mech. Dev.* **57**, 199–214.

Byrd, N., Becker, S., Maye, P., Narasimhaiah, R., St-Jacques, B., Zhang, X., McMahon, J., McMahon, A., Grabel, L. (2002) Hedgehog is required for murine yolk sac angiogenesis, *Development* **129**, 361–372.

Chen, X.D., Turpen, J.B. (1995) Intraembryonic origin of hepatic hematopoiesis in Xenopus laevis, *J. Immunol.* **154**, 2557–2567.

Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J.C., Keller, G. (1998) A common precursor for hematopoietic and endothelial cells, *Development* **125**, 725–732.

Detrich, H.W. III, Kieran, M.W., Chan, F.Y., Barone, L.M., Yee, K., Rundstadler, J.A., Pratt, S., Ransom, D., Zon, L.I. (1995) Intraembryonic hematopoietic cell migration during vertebrate development, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10713–10717.

Dieterlen-Lievre, F. (1975) On the origin of hematopoieitic stem cells in the avian embryo: an experimental approach, *J. Embryol. Exp. Morphol.* **33**, 607–619.

Dosch, R., Gawantka, V., Delius, H., Blumenstock, C., Niehrs, C. (1997) Bmp-4 acts as a morphogen in dorsoventral mesoderm patterning in Xenopus, *Development* **124**, 2325–2334.

Dyer, M.A., Farrington, S.M., Mohn, D., Munday, J.R., Baron, M.H. (2001) Indian hedgehog activates hematopoiesis and vasculogenesis and can respecify prospective neurectodermal cell fate in the mouse embryo, *Development* **128**, 1717–1730.

Eichmann, A., Corbel, C., Nataf, V., Vaigot, P., Breant, C., Le Douarin, N.M. (1997) Ligand-dependent development of the endothelial and hemopoietic lineages from embryonic mesodermal cells expressing vascular endothelial growth factor receptor 2, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5141–5146.

Fujiwara, Y., Browne, C.P., Cunniff, K., Goff, S.C., Orkin, S.H. (1996) Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12355–12358.

Ghatpande, S., Ghatpande, A., Sher, J., Zile, M.H., Evans, T. (2002) Retinoid signaling regulates primitive (yolk sac) hematopoiesis, *Blood* **99**, 2379–2386.

Habeck, H., Odenthal, J., Walderich, B., Maischein, H., Schulte-Merker, S. (2002) Analysis of a zebrafish VEGF receptor mutant reveals specific disruption of angiogenesis, *Curr. Biol.* **12**, 1405–1412.

Hammerschmidt, M., Serbedzija, G.N., McMahon, A.P. (1996) Genetic analysis of dorsoventral pattern formation in the zebrafish: requirement of a BMP-like ventralizing activity and its dorsal repressor, *Genes Dev.* **10**, 2452–2461.

Kabrun, N., Buhring, H.J., Choi, K., Ullrich, A., Risau, W., Keller, G. (1997) Flk-1 expression defines a population of early embryonic hematopoietic precursors, *Development* **124**, 2039–2048.

Kau, C.L., Turpen, J.B. (1983) Dual contribution of embryonic ventral blood island and dorsal lateral plate mesoderm during ontogeny of hemopoietic cells in Xenopus laevis, *J. Immunol.* **131**, 2262–2266.

Kieran, M.W., Perkins, A.C., Orkin, S.H., Zon, L.I. (1996) Thrombopoietin rescues in vitro erythroid colony formation from mouse embryos lacking the erythropoietin receptor, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9126–9131.

Kishimoto, Y., Lee, K.H., Zon, L., Hammerschmidt, M., Schulte-Merker, S. (1997) The molecular nature of zebrafish swirl: BMP2 function is essential during early dorsoventral patterning, *Development* **124**, 4457–4466.

Kumano, G., Belluzzi, L., Smith, W.C. (1999) Spatial and temporal properties of ventral blood island induction in Xenopus laevis, *Development* **126**, 5327–5337.

Liang, D., Xu, X., Chin, A.J., Balasubramaniyan, N.V., Teo, M.A., Lam, T.J., Weinberg, E.S., Ge, R. (1998) Cloning and characterization of vascular endothelial growth factor (VEGF) from zebrafish, Danio rerio, *Biochim. Biophys. Acta* **1397**, 14–20.

Liang, D., Chang, J.R., Chin, A.J., Smith, A., Kelly, C., Weinberg, E.S., Ge, R. (2001) The role of vascular endothelial growth factor (VEGF) in vasculogenesis, angiogenesis, and hematopoiesis in zebrafish development, *Mech. Dev.* **108**, 29–43.

Liao, W., Bisgrove, B.W., Sawyer, H., Hug, B., Bell, B., Peters, K., Grunwald, D.J., Stainier, D.Y. (1997) The zebrafish gene cloche acts upstream of a flk-1 homologue to regulate endothelial cell differentiation, *Development* **124**, 381–389.

Liao, W., Ho, C.Y., Yan, Y.L., Postlethwait, J., Stainier, D.Y. (2000) Hhex and scl function in parallel to regulate early endothelial and

blood differentiation in zebrafish, *Development* **127**, 4303–4313.

Liao, E.C., Trede, N.S., Ransom, D., Zapata, A., Kieran, M., Zon, L.I. (2002) Non-cell autonomous requirement for the bloodless gene in primitive hematopoiesis of zebrafish, *Development* **129**, 649–659.

Lieschke, G.J., Oates, A.C., Paw, B.H., Thompson, M.A., Hall, N.E., Ward, A.C., Ho, R.K., Zon, L.I., Layton, J.E. (2002) Zebrafish SPI-1 (PU.1) marks a site of myeloid development independent of primitive erythropoiesis: implications for axial patterning, *Dev. Biol.* **246**, 274–295.

Lyons, S.E., Lawson, N.D., Lei, L., Bennett, P.E., Weinstein, B.M., Liu, P.P. (2002) A nonsense mutation in zebrafish gata1 causes the bloodless phenotype in vlad tepes, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5454–5459.

Maeno, M., Tochinai, S., Katagiri, C. (1985) Differential participation of ventral and dorsolateral mesoderms in the hemopoiesis of Xenopus, as revealed in diploid-triploid or interspecific chimeras, *Dev. Biol.* **110**, 503–508.

Maeno, M., Ong, R.C., Xue, Y., Nishimatsu, S., Ueno, N., Kung, H.F. (1994) Regulation of primary erythropoiesis in the ventral mesoderm of Xenopus gastrula embryo: evidence for the expression of a stimulatory factor(s) in animal pole tissue, *Dev. Biol.* **161**, 522–529.

Maeno, M., Mead, P.E., Kelley, C., Xu, R.H., Kung, H.F., Suzuki, A., Ueno, N., Zon, L.I. (1996) The role of BMP-4 and GATA-2 in the induction and differentiation of hematopoietic mesoderm in Xenopus laevis, *Blood* **88**, 1965–1972.

Martin, J.S., Dickson, M.C., Cousins, F.M., Kulkarni, A.B., Karlsson, S., Akhurst, R.J. (1995) Analysis of homozygous TGF beta 1 null mouse embryos demonstrates defects in yolk sac vasculogenesis and hematopoiesis, *Ann. N. Y. Acad. Sci.* **752**, 300–308.

Matsuoka, S., Tsuji, K., Hisakawa, H., Xu, M., Ebihara, Y., Ishii, T., Sugiyama, D., Manabe, A., Tanaka, R., Ikeda, Y., Asano, S., Nakahata, T. (2001) Generation of definitive hematopoietic stem cells from murine early yolk sac and paraaortic splanchnopleures by aorta-gonad-mesonephros region-derived stromal cells, *Blood* **98**, 6–12.

McGann, J.K., Silver, L., Liesveld, J., Palis, J. (1997) Erythropoietin-receptor expression and function during the initiation of murine yolk sac erythropoiesis, *Exp. Hematol.* **25**, 1149–1157.

Mead, P.E., Deconinck, A.E., Huber, T.L., Orkin, S.H., Zon, L.I. (2001) Primitive erythropoiesis in the Xenopus embryo: the synergistic role of LMO-2, SCL and GATA-binding proteins, *Development* **128**, 2301–2308.

Mead, P.E., Kelley, C.M., Piedad, P.S., Hahn, O., Zon, L.I. (1998) SCL specifies hematopoietic mesoderm in Xenopus embryos, *Development* **125**, 2611–2620.

Mikkola, H.K., Klintman, J., Yang, H., Hock, H., Schlaeger, T.M., Fujiwara, Y., Orkin, S.H. (2003) Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene, *Nature* **421**, 547–551.

Neave, B., Holder, N., Patient, R. (1997) A graded response to BMP-4 spatially coordinates patterning of the mesoderm and ectoderm in the zebrafish, *Mech. Dev.* **62**, 183–195.

Nikaido, M., Tada, M., Saji, T., Ueno, N. (1997) Conservation of BMP signaling in zebrafish mesoderm patterning, *Mech. Dev.* **61**, 75–88.

Oates, A.C., Brownlie, A., Pratt, S.J., Irvine, D.V., Liao, E.C., Paw, B.H., Dorian, K.J., Johnson, S.L., Postlethwait, J.H., Zon, L.I., Wilks, A.F. (1999) Gene duplication of zebrafish JAK2 homologs is accompanied by divergent embryonic expression patterns: only jak2a is expressed during erythropoiesis, *Blood* **94**, 2622–2636.

Ohinata, H., Tochinai, S., Katagiri, C. (1990) Occurrence of nonlymphoid leukocytes that are not derived from blood islands in Xenopus laevis larvae, *Dev. Biol.* **141**, 123–129.

Palis, J., McGrath, K.E., Kingsley, P.D. (1995) Initiation of hematopoiesis and vasculogenesis in murine yolk sac explants, *Blood* **86**, 156–163.

Palis, J., Robertson, S., Kennedy, M., Wall, C., Keller, G. (1999) Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse, *Development* **126**, 5073–5084.

Parker, L., Stainier, D.Y. (1999) Cell-autonomous and non-autonomous requirements for the zebrafish gene cloche in hematopoiesis, *Development* **126**, 2643–2651.

Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H., Costantini, F. (1991) Erythroid differentiation

in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1, *Nature* **349**, 257–260.

Pevny, L., Lin, C.S., D'Agati, V., Simon, M.C., Orkin, S.H., Costantini, F. (1995) Development of hematopoietic cells lacking transcription factor GATA-1, *Development* **121**, 163–172.

Porcher, C., Liao, E.C., Fujiwara, Y., Zon, L.I., Orkin, S.H. (1999) Specification of hematopoietic and vascular development by the bHLH transcription factor SCL without direct DNA binding, *Development* **126**, 4603–4015.

Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F.W., Orkin, S.H. (1996) The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages, *Cell* **86**, 47–57.

Ransom, D.G., Haffter, P., Odenthal, J., Brownlie, A., Vogelsang, E., Kelsh, R.N., Brand, M., van Eeden, F.J., Furutani-Seiki, M., Granato, M., Hammerschmidt, M., Heisenberg, C.P., Jiang, Y.J., Kane, D.A., Mullins, M.C., Nusslein-Volhard, C. (1996) Characterization of zebrafish mutants with defects in embryonic hematopoiesis, *Development* **123**, 311–319.

Shivdasani, R.A., Mayer, E.L., Orkin, S.H. (1995) Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL, *Nature* **373**, 432–434.

Silver, L., Palis, J. (1997) Initiation of murine embryonic erythropoiesis: a spatial analysis, *Blood* **89**, 1154–1164.

Smith, P.B., Flajnik, M.F., Turpen, J.B. (1989) Experimental analysis of ventral blood island hematopoiesis in Xenopus embryonic chimeras, *Dev. Biol.* **131**, 302–312.

Tsai, F.Y., Keller, G., Kuo, F.C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F.W., Orkin, S.H. (1994) An early haematopoietic defect in mice lacking the transcription factor GATA-2, *Nature* **371**, 221–226.

Tsai, F.Y., Orkin, S.H. (1997) Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation, *Blood* **89**, 3636–3643.

Tsang, A.P., Visvader, J.E., Fujiwara, C.A. Turner, Y., Yu, C., Weiss, M.J., Crossley, M., Orkin, S.H. (1997) FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation, *Cell* **90**, 109–119.

Tsang, A.P., Fujiwara, Y., Hom, D.B., Orkin, S.H. (1998) Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG, *Genes Dev.* **12**, 1176–1188.

Turpen, J.B. (1998) Induction and early development of the hematopoietic and immune systems in Xenopus, *Dev. Comp. Immunol.* **22**, 265–278.

Turpen, J.B., Kelley, C.M., Mead, P.E., Zon, L.I. (1997) Bipotential primitive-definitive hematopoietic progenitors in the vertebrate embryo, *Immunity* **7**, 325–334.

Visvader, J.E., Mao, X., Fujiwara, Y., Hahm, K., Orkin, S.H. (1997) The LIM-domain binding protein Ldb1 and its partner LMO2 act as negative regulators of erythroid differentiation, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13707–13712.

Walters, M.J., Wayman, G.A., Notis, J.C., Goodman, R.H., Soderling, T.R., Christian, J.L. (2002) Calmodulin-dependent protein kinase IV mediated antagonism of BMP signaling regulates lineage and survival of hematopoietic progenitors, *Development* **129**, 1455–1466.

Weinstein, B.M., Schier, A.F., Abdelilah, S., Malicki, J., Solnica-Krezel, L., Stemple, D.L., Stainier, D.Y., Zwartkruis, F., Driever, W., Fishman, M.C. (1996) Hematopoietic mutations in the zebrafish, *Development* **123**, 303–309.

Weiss, M.J., Yu, C., Orkin, S.H. (1997) Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line, *Mol. Cell Biol.* **17**, 1642–1651.

Yoder, M.C., Hiatt, K. (1997) Engraftment of embryonic hematopoietic cells in conditioned newborn recipients, *Blood* **89**, 2176–2183.

Yoder, M.C., Hiatt, K., Dutt, P., Mukherjee, P., Bodine, D.M., Orlic, D. (1997) Characterization of definitive lymphohematopoietic stem cells in the day 9 murine yolk sac, *Immunity* **7**, 335–344.

Yoder, M.C., Hiatt, K., Mukherjee, P. (1997) In vivo repopulating hematopoietic stem cells are present in the murine yolk sac at day 9.0 postcoitus, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6776–6780.

Zhou, Y., Lim, K.C., Onodera, K., Takahashi, S., Ohta, J., Minegishi, N., Tsai, F.Y., Orkin, S.H., Yamamoto, M., Engel, J.D. (1998) Rescue of the embryonic lethal hematopoietic defect reveals a critical role for GATA-2 in urogenital development, *EMBO J.* **17**, 6689–6700.

Cell, Computer Modeling of the: *see* E-Cell: Computer Simulation of the Cell

Cell-free Translation Systems

Takuya Ueda, Akio Inoue and Yoshihiro Shimizu
The University of Tokyo, Kashiwa, Japan

1 History 452
1.1 Beginning 452
1.2 Determination of the Genetic Code 453
1.3 Elucidation of Factors Involved in the Translation Process 453

2 Applications 454
2.1 Protein Production 454
2.2 Synthesis of Artificial Protein 456
2.3 Linkage between Genotype and Phenotype 458

3 Perspectives 460

Bibliography 460
Books and Reviews 460
Primary Literature 460

Keywords

Genetic Code: Relationship between Codons and Amino Acids
CFCF (continuous-flow cell-free) translation system: cell-free translation system developed for protein production by supplementing low molecular weight substrates to cell-free translation.
CECF (continuous-exchange cell-free translation): improved cell-free translation system by the dialysis method.

Suppressor tRNA
tRNA decoding termination codons.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Unnatural Amino Acid
Amino acids that are not used by living organisms.

In vitro Virus
The selection method by RNA molecules covalently bound to the protein product.

Ribosome Display
The selection method for mRNA by the function of the polypeptide displayed on ribosome stalling during the translation process.

A disrupted cell is capable of synthesizing protein from amino acids depending upon the template RNA. This discovery, half a century ago, set the stage for biochemical studies on the gene expression process occurring in cells. The cell-free translation system successfully revealed the function of a number of key molecules participating in the translation system, such as ribosome, mRNA, tRNA, and protein factors and, moreover, elucidated the very complicated mechanism of protein synthesis. Although still frequently utilized in these experimental studies, cell-free translation is beginning to make a mark as an attractive tool for protein production as a potential alternative to an *in vivo* expression system. Moreover, the cell-free translation system is prospective as a method for the synthesis of protein with unnatural amino acids and for the selection of genotypes from peptide libraries.

1 History

1.1 Beginning

Proteins are translated on ribosomes from mRNAs transcribed from DNA. The outline of how translation proceeds had been elucidated by both *in vitro* and *in vivo* approaches for two decades after the discovery of the double-helix structure of DNA. In the 1950s, most studies focusing on the gene expression process were carried out on the basis of a genetic approach using bacteria and phages. The hypothesis of the central dogma, DNA makes the RNA that makes protein, had been evidenced by various observations obtained from *in vivo* experiments. However, the precise mechanism of translation, such as deciphering information on mRNA, had to await a series of verifications by the *in vitro* translation system using disrupted cells. The first demonstration of protein synthesis using cell-extract was described by Zamecnik's group in 1952. The finding that rat liver extract was capable of incorporating amino acids into polypeptides allowed us to elucidate the involvement of individual molecular players in the gene expression process. Thereafter, it was shown in 1955 that the ribonucleoprotein complex, which was designated as ribosome, catalyzes the peptide bond formation. Up to 1960, the construction of bacterial cell-free translation systems had been achieved by several

groups. The addition of viral genomic RNA resulted in a large increase of amino acid incorporation into polypeptide, concluding that RNA serves as a template for protein.

1.2 Determination of the Genetic Code

One of the most remarkable contributions of the cell-free translation system in molecular biology was the determination of the genetic code in the 1960s. Matthaei and Nierenberg demonstrated that the addition of exogenous artificial RNA to the cell-free translation system enhanced the incorporation of particular amino acids into polypeptides. The first demonstration was that, in the presence of the polyurdylic acid (poly U), in *Escherichia coli*, the cell-free translation system was capable of synthesizing polymers of only phenylalanine at higher magnesium concentrations than physiological conditions. This finding in 1960 lead to the determination of the correspondence of the triplet codon to amino acid over the next five years. For instance, it was immediately shown that poly C was able to stimulate proline incorporation into polypeptide, indicating that the codon CCC corresponded to proline. By the various synthesized RNAs consisting of randomized two or three kinds of nucleotides, the stimulation of the incorporation of amino acid into the polypeptide was examined, and the relationship between codons and amino acids was determined using these polynucleotides. However, this approach came to a standstill because statistical speculation resulted in ambiguity.

This problem was solved by the elegant experiment called the *triplet-binding experiment*, in which aminoacyl tRNA charged with cognate amino acids binds synthesized trinucleotide on ribosome and determined the correspondence of amino acids to triplet codons. Simultaneously, the chemically synthesized oligonucleotide with a defined sequence was employed for examination of the incorporation of amino acid into polypeptides. Together with these results, the genetic code table had been established by 1966.

Another contribution of the cell-free translation system was the direct verification that tRNA is an adaptor molecule linking amino acids to codons on mRNA. The existence of an adaptor molecule had been proposed by Crick, which had not been evidenced with any experimental results. In 1957, Hoagland and Ogata independently found that the small soluble RNA molecule, which would later be named tRNA, was charged with amino acids. Nierenberg's triplet-binding experiment leads us to the conclusion that this tRNA molecule plays the role of the adaptor molecule that deciphers comma-less triplet codons on mRNA. The determination of the primary structure of tRNA by Holley demonstrated that tRNA folded in cloverleaf structure fulfills the decoding function based upon the interaction between the anticodon and codon.

1.3 Elucidation of Factors Involved in the Translation Process

To evaluate the role of protein factors involved in the translation process, the cell-free translation system has appeared to be the most useful system. The translation process proceeds in three steps: initiation, elongation, and termination, in which specific protein factors govern a respective refined process. These factors were first discovered to enhance protein synthesis using cell-free translation, and the detailed behavior was elucidated by a well-defined system capable of verifying the elementary

steps. In the case of the elongation process, Elongation Factor (EF)-Tu was first found to be a factor that stimulates poly U-dependent polyphenylalanine synthesis and thereafter it appeared to convey aminoacylated tRNA to the ribosome as a GTP-binding form. EF-Ts were shown to catalyze the exchange of GTP to GDP on EF-Tu. EF-G appeared to catalyze the hydrolysis of the triphosphate of GTP depending on the ribosome and appeared to be involved in the translocation of tRNA from the A-site to the P-site.

In the case of the initiation process, initiation factor-2 (IF-2) was shown to carry initiator tRNA attached by formymethionine onto a 30S subunit of ribosome. Initiation factor-3 (IF-3) was indicated to dissociate 70S ribosome resulting from the termination reaction. Initiation factor-1 (IF-1) was found to enhance the disassembling activity of IF3 by entering an inter-subunit space. Regarding the termination process, release factor-1(RF-1) and release factor-2 (RF-2) enter the A-site of the ribosome depending on the termination codons to catalyze the hydrolysis of peptidyl-tRNA on the P-site of ribosome. RF-1 recognizes UAG and UAA codons, while RF-2 is responsible for UGA and UAA. While the function of these two RFs was verified at an early stage of the study on translation, RF-3 and the ribosome recycling factor (RRF) had remained less understood. Very recently RF-3 was shown to be involved in the release of RF-1 and RF-2 from the ribosome. RRF was first discovered as a factor enhancing the activity of the cell-free translation system and shown to promote the dissociation of mRNA and tRNA from the ribosome.

Similarly, eukaryotic translational factors have been studied by the cell-free translation system. Although most essential factors coincide with bacterial factors, finding a number of additional proteins enhancing the cell-free translation system, especially in the initiation process, suggests that the eukaryotic system appeared to be much more complicated than the bacterial system.

2 Applications

2.1 Protein Production

Although having long served as a useful system for the elucidation of the translation mechanism, the cell-free translation system had been paid only slight attention as a tool for the preparation of protein. Most protein productions at present are based upon the expression system of cloned genes in living cells, in which we sometimes undergo failure, particularly in cases of toxic or unstable proteins. Therefore, it was desired that an *in vitro* gene expression system using the cell-free translation system would compensate for the *in vivo* expression system. However, the low yield of protein product had hindered the utilization of the cell-free translation system. The conventional cell-free translation system using crude extracts, such as *E. coli* S30, rabbit reticulocyte lysate and wheat germ extracts, produces translation reactions for no longer than 1 h in the batch system and synthesized protein products to such a small degree as to be detected only by labeling with radioactive amino acids. To improve productivity, transcription/translation, coupled with using phage RNA polymerases such as T7 or SP6 RNA polymerases, and energy regenerating systems for ATP recycling such as creatine phosphate/creatine kinase or phosphoenolpyruvate/pyurvate kinase

systems, had been developed, though both showed limited results.

However, the short lifetime of the cell-free translation system was overcome by the development of the continuous-flow cell-free (CFCF) system by A Spirin's group in 1988. In CFCF, prolongation of the reaction was achieved by the continuous feeding of amino acids and energy components, and continuous removal of reaction products. The reaction was carried out in a chamber connected to a reservoir to supply the substrates, ATP, GTP, and amino acids, by pumping, and the reactor was separated at the outlet by a membrane facilitating permeation of the synthesized protein and a low molecular weight reaction product (Fig. 1a). By this system, it appeared that the reaction using wheat germ extract proceeds for up to 100 h and was able to synthesize 120 to 240 μg of protein in a 1-mL reaction chamber. The most plausible explanation for this long sustainability is that the shortage of an energy source such as ATP and GTP during the reaction is compensated by the supplement of a feeding solution, considering that in the batch system using cell extracts, the ATPase inevitably dwelling in the extract rapidly consumes ATP. The simpler system without the aid of pumping was developed on the basis of the dialysis method of low molecular weight compounds and designated as the continuous-exchange cell-free (CECF) system (Fig. 1b).

The improvement of productivity and reproducibility of cell-free translation has also been achieved by examining the reaction condition and preparation procedures of crude cell-extract. Preparation of crude extract from polished wheat germ ensures constant production of proteins and prolongation of reaction in the CFCF system. By washing the embryo, endogenous inhibitors of translation, including ribosome inactivating proteins (RIP) such as ricin toxin, are successfully eliminated from extracts, which may result in such drastic

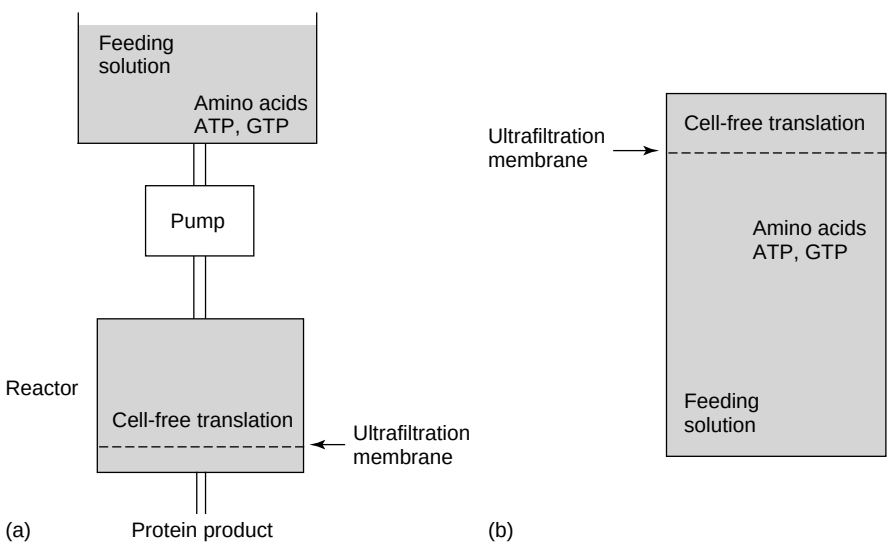

Fig. 1 Schematic illustrations of (a) CFCF and (b) CECF translation systems. A feeding solution containing ATP, GTP, and amino acids is supplied to the reaction chamber in which cell-free translation proceeds, by pumping (CFCF) or diffusion (CECF).

extensions of reaction and high yields of proteins. More recently, a reconstruction of the cell-free translation system using purified components was successfully achieved. Three initiation factors, three elongation factors, three release factors, RRFs, 20 individual aminoacyl tRNA synthetases and methionyl-tRNA transformylase were cloned from the *E. coli* genome for efficient expression in *E. coli* cells and purified. This reconstituted system, designated as the Protein synthesis using recombinant elements (PURE) system, is capable of synthesizing approximately 0.1 mg of proteins in 1-mL volume per hour without any supplemental system such as CFCF and CECF.

2.2 Synthesis of Artificial Protein

One application of cell-free translation is the incorporation of unnatural amino acids into polypeptides. The amino acid–comprising proteins are canonically confined to 20 kinds, barring exceptions such as selenocystein, and therefore *in vivo* systems are unable to synthesize proteins with other amino acids. In 1989, Shultz's group developed a method to incorporate unnatural amino acids at a desired position in polypeptides using the cell-free translation system. In this system, as illustrated in Fig. 2, synthesized suppressor tRNA are first attached by a chemical or enzymatic method to unnatural amino acids and subjected to cell-free translation for DNA templates possessing amber codons at the desired position. The attachment of unnatural amino acids is mostly carried out as follows. An unnatural amino acid is first attached to 3′ OH of adenosine of pCpA dinucleotide by a chemical method. Later, the resultant aminoacylated pCpA is ligated to an amber suppressor tRNA lacking a 3′-terminal CA sequence using

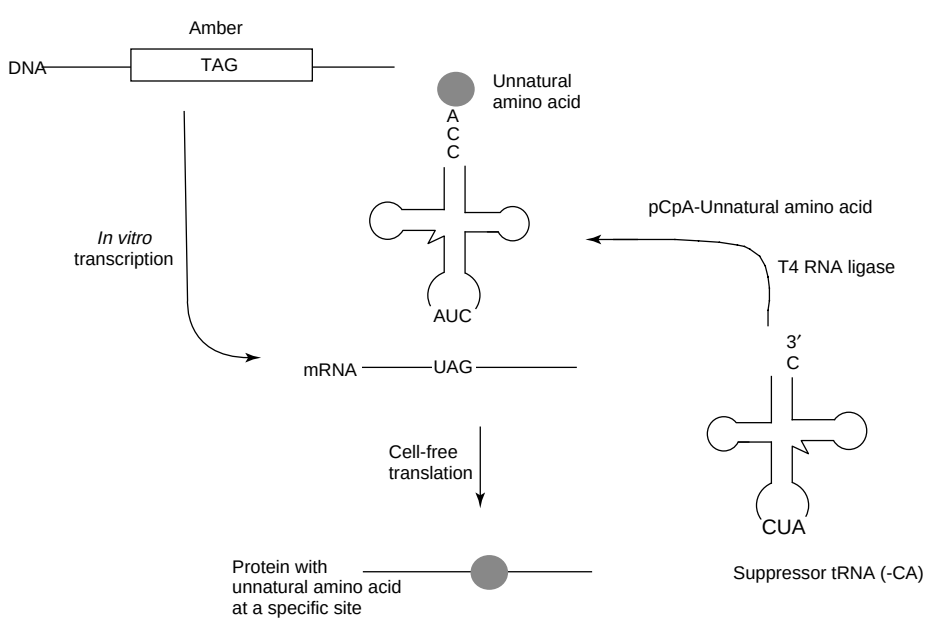

Fig. 2 Strategy for incorporation of unnatural amino acid into polypeptide using the cell-free translation system.

an RNA ligase. If unnatural amino acids are good substrates for aminoacyl tRNA synthetases, they can be attached to tRNA by an appropriate aminoacyl tRNA synthetase. The unnatural amino acid ligated to a suppressor tRNA is further subjected to the cell-free translation system and the codon-forming base pairing with the anticodon of suppressor tRNA is translated into a proteinogenic amino acid.

Although the incorporation of an unnatural amino acid into a polypeptide using cell-free translation is highly advantageous compared to an *in vivo* system, there are several problems to be solved. One problem is the aminoacylation of the suppressor tRNA after one reaction cycle of translation. The suppressor tRNA recharged by an aminoacyl tRNA synthetase during cell-free translation inserts a canonical amino acid at the amber codon, which hinders the efficient incorporation of unnatural amino acid. To solve this problem, tRNAs that are not recognized by an endogenous aminoacyl tRNA synthetase has been developed. For instance, taking advantage of the fact that yeast $tRNA^{Tyr}$ and $tRNA^{Gln}$ do not serve as a substrate for the *E. coli* cognate aminoacyl tRNA synthetase, the suppressor tRNA having the framework of yeast tRNAs is utilized as an adaptor tRNA for unnatural amino acids in an *E. coli* cell-free translation system to prevent recycling of tRNA. *E. coli* $tRNA^{Tyr}$, which is not recognized by the eukaryotic tyrosyl-tRNA synthetase, can be utilized in wheat germ extract and *vice versa*. Another problem is that the presence of RF-1 corresponding to the amber codon in the extract caused the termination to produce an incomplete peptide, reducing the percentage of peptide containing unnatural amino acid. Using the cell-free translation system in which all the components are in a purified state, this competition is circumvented by eliminating RF-1 from the system, achieving almost full suppression efficiency.

Another problem in the amber suppressor strategy is that only one codon in the open reading frame is available for the introduction of unnatural amino acids. To allow the incorporation of several varieties of unnatural amino acids into a single polypeptide, the four-base codon method was developed. In this approach, tRNA having a four-base anticodon is synthesized by transcription reaction using RNA polymerase, and chemically aminoacylated, followed by cell-free translation. By the four-base codon–anticodon method, one is enabled to incorporate multiple unnatural amino acids at plural codons. Obviously, due to competition with endogenous tRNAs, incorporation of an unnatural amino acid into the polypeptide never reaches full yield, but by employing the four-base codons competing with minor three-base codons, the efficiency of incorporation is easily improved, depending on the amount of endogenous tRNA. This artificial genetic code can be expanded by exploiting the five-base anticodon tRNA.

The expansion of the genetic code in cell-free translation was challenged by the utilization of unnatural nucleobases in codon and anticodon correspondence. The utilization of one additional artificial base pairing besides A-U and G-C pairing in the translation process is able to, in principle, extend the canonical 64 codons to 216 ($6 \times 6 \times 6$) codons in the genetic code, which gives rise to a vast possibility for the incorporation of various unnatural amino acids into polypeptides. First, isoC and isoG (Fig. 3a) were introduced in the codon and anticodon, respectively, and the resultant artificial codon corresponded to the unnatural amino acid. Recently,

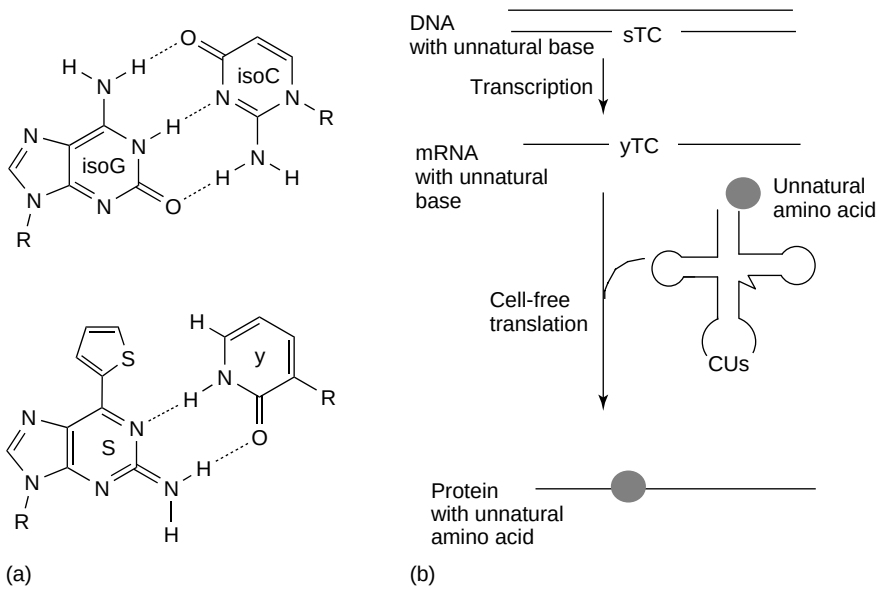

Fig. 3 Structures of unnatural bases utilized in (a) cell-free translation for expansion of genetic code and (b) design of transcription/translation coupled system using unnatural nucleotides for incorporation of unnatural amino acids.

the incorporation of 3-chlorotyrocine was successfully inserted into the desired position of protein utilizing base pairing between 2-amino-6-(2-thienyl) purine (*s*) and pyridin-2-1 (*y*) (Fig. 3b) as follows. First, a DNA fragment in which the *s*-nucleotide was located at the first letter of the desired codon was synthesized for the synthesis of mRNA by transcription reaction in the presence of the *y*-nucleotide. Simultaneously, yeast $tRNA^{Tyr}$ with the *s*-nucleotide at the third position of the anticodon was prepared by ligation of RNA fragments and was aminoacylated with 3-chlorotyrosine using tyrosyl-tRNA synthetase. Using the *E. coli* cell-free translation system, 3-chlorotyrosine was successfully incorporated at the desired position of a target protein. Although this approach looks very expansible, technical difficulties in site-specific introduction of appropriate artificial nucleobases into mRNA and tRNA remain to be solved for the development of a concise and widely distributed system.

2.3 Linkage between Genotype and Phenotype

Cell-free translation provides the potential methodology for linkage between genotype and phenotype. *In vitro* screening of DNA corresponding to RNAs and proteins with certain functions is of significant importance in the field of molecular biology. In the case of functional RNA, such as ribozymes, it is easy to convert to DNA using the reverse transcriptase after the selection procedure designed for the desired function. In contrast, we have no direct system to convert from polypeptide to mRNA. However, the complex formation of nucleic acid and its protein product enables one to perform pseudoreverse translation.

In order to link polypeptide and mRNA, two approaches using cell-free translation are now available: (1) the "*in vitro* virus" method and (2) the ribosome (or polysome)-display method.

In the "*in vitro virus*" method, creation of a covalent linkage of bonding between the mRNA and the protein is executed by cell-free translation depending on mRNA tagged with puromycin, an antibiotic with a structure resembling the 3′ end of aminoacyl tRNA (Fig. 4a). Emerging nascent polypeptide at the P-site is transferred to mRNA-linked puromycin at the A-site, giving rise to a covalent-bonded molecule of the protein product and responsible mRNA. Consequently, through the selection toward a polypeptide portion, mRNA is concurrently isolated, followed by an RT-PCR reaction of the mRNA fragment.

Meanwhile, the ribosome (or polysome)-display method is based upon a non-covalent complex formation of mRNA, ribosome, and polypeptide (Fig. 4b). The mRNA libraries are translated in the cell-free translation system and the reaction is hampered by antibiotics, for instance, chloramphenicol for the system derived from *E. coli*, or simply by cooling on ice. The stalling ribosomes display nascent polypeptides emerging from translating mRNA, and the selection is directed toward arresting the mRNA-ribosome-polypeptide complex. The isolated complex is dissociated with EDTA by splitting the ribosomal subunit, and the eluted mRNA is subjected to the RT-PCR reaction.

To ensure an efficient ribosome display, the stability of the mRNA–ribosome–polypeptide complex should be sustained

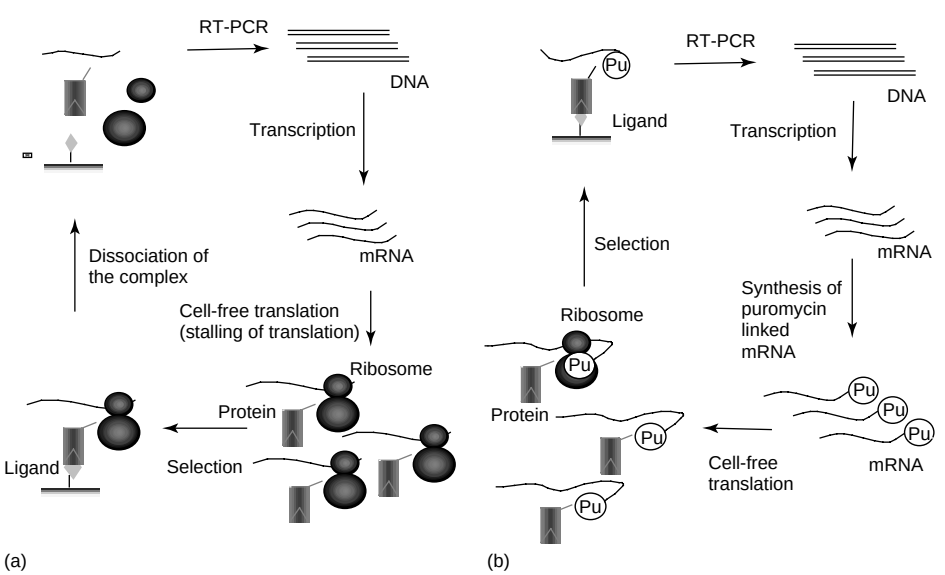

Fig. 4 Schematic drawing of two-selection method of mRNA by protein function. Ribosome display (a) is based on the selection toward attached polypeptide on stalling ribosome. The RNA-ribosome-polypeptide complex is screened by immobilized ligand and RNA derived from a selected complex was amplified by RT-PCR. *In vitro* virus (b) is synthesized puromicin-linked mRNA, which attacks nascent polypeptides attached to tRNA on the ribosome. The covalent complex of mRNA and polypeptide is selected by the ligand and subjected to RT-PCR.

against the presence of RFs in the cell-extract of the cell-free translation system. RFs in cell-free extracts hydrolyze polypeptidyl-tRNA on the P-site, responding to the occupation of the termination codon at the A-site, causing a problematical release of the complex. To prevent this instability, mRNA lacking the termination codon should be designed. The degradation of mRNA during cell-free translation sometimes reduces recovery of the objective mRNA. Thus, nuclease-free cell-free translation systems are eagerly anticipated.

Despite these problems, these two methods allow the mRNA (genotype) to be fished out depending upon the protein function (phenotype) *in vitro*, relatively more conveniently and expeditiously than an *in vivo* system. Moreover, *in vitro* systems are capable of dealing with a huge genetic library comprising 10^{13} molecules, which is an advantage over *in vivo* systems. Particularly, RNA fragments coding for polypeptides with binding properties toward certain ligands, such as antibodies, are enriched in a cloning-independent way by affinity selection from a variety of genetic pools. The multiple cycles of *in vitro* selection and amplification of functional proteins and genes, starting from a randomized pool, can be designated as an *in vitro* evolution system.

3 Perspectives

At present, most protein production processes are carried out utilizing an expression system in living cells. However, we are now able to amplify DNA molecules in a test tube by PCR technology using a DNA polymerase and to synthesize RNA by *in vitro* transcription reaction using an RNA polymerase. Considering that these cell-free systems definitely accelerated an advance in molecular biology, the development of the cell-free translation system will also greatly contribute to studies on the structures and/or functions of proteins. Because of simplicity compared to an *in vivo* expression system, the cell-free translation system is well suited for high-throughput protein production in a postgenome era. Moreover, modification of the translation process is so easily achieved that we will be able to develop an artificial system in addition to synthesis of protein with unnatural amino acids, the *in vitro* virus method and ribosome-display method.

See also DNA Replication and Transcription; Ribosome Structure and Function, High Resolution.

Bibliography

Books and Reviews

Jermutus, L., Ryabova, L.A., Pluckthun, A. (1998) Recent advances in producing and selecting functional proteins by using cell-free translation, *Curr. Opin. Biotechnol.* **9**, 534–548.

Primary Literature

Hirao, I., Ohtsuki, T., Fujiwara, T., Mitsui, T., Yokogawa, T., Okuni, T., Nakayama, H., Takio, K., Yabuki, T., Kigawa, T., Kodama, K., Nishikawa, K., Yokoyama, S. (2002) An unnatural base pair for incorporating amino acid analogs into proteins, *Nat. Biotechnol.* **20**, 177–182.

Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C., Schultz, P.G. (1989) A general method for site-specific incorporation of unnatural amino acids into proteins, *Science* **244**, 182–188.

Piccirilli, J.A., Krauch, T., Moroney, S.E., Benner, S.A. (1990) Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet, *Nature* **343**, 33–37.

Shimizu, Y., Inoue, A., Ueda, T. (2001) Cell-free translation reconstituted with purified components, *Nat. Biotechnol.* **19**, 751–755.

Sisido, M., Hohsaka, T. (2001) Introduction of specialty functions by the position-specific incorporation of nonnatural amino acids into proteins through four-base codon/anticodon pairs, *Appl. Microbiol. Biotechnol.* **57**, 274–281.

Spirin, A.S., Baranov, V.I., Ryabova, L.A., Ovodov, S.Y., Alakhov, Y.B. (1988) A continuous cell-free translation system capable of producing polypeptides in high yield, *Science* **242**, 1162–1164.

Cellular Interactions

Brian A. Koeneman and David G. Capco
School of Life Sciences Cellular and Molecular Biosciences Faculty,
Arizona State University Tempe, AZ, USA

1 Overview of Cytoplasmic Signaling 465
1.1 General 465
1.2 Signals Regulated by Kinases 467

2 Coordination of Complex Signaling Pathways 468
2.1 Molecular Scaffolds 468
2.2 Cell Architecture 470
2.2.1 Action of Signaling Agents on the Cytoskeleton 470

3 Formation of the Fertilization-competent Egg 471
3.1 CaM KII 471
3.2 PKC in the Mammalian Egg 471
3.3 MAPK and MPF 471

4 Activation of the Egg 473
4.1 Remodeling through Changes in Activity of Calcium/Calmodulin-dependent Protein Kinase II 473
4.2 Remodeling through Changes in Activity of Protein Kinase C 476

4.3 Remodeling through Changes in Activity of MPF and MAP Kinase Activity in Activated Eggs 479
5 Concluding Remarks 480

Bibliography 481
Books and Reviews 481
Primary Literature 481

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Keywords

$[Ca^{2+}]_i$ (Concentration of Intracellular-free Calcium)
When the egg is fertilized, there is a large rise in $[Ca^{2+}]_i$ that propagates across the entire cell, and in mammalian eggs this is followed by oscillations in the level of intracellular-free calcium.

CaM KII (Calcium/Calmodulin-dependent Protein Kinase II)
In the egg, it is involved with the meiotic metaphase II to anaphase II/telophase II transition.

Cytoskeleton
A complex network of protein filaments within a cell that continuously reorganizes to adopt different shapes and functions.

Egg Activation
A process that converts an egg into a zygote. It is naturally triggered by fertilization but can be experimentally induced by calcium ionophore.

Signal Transduction
Relaying of an extracellular signal into a cell then amplifying the signal quickly within the cell to create a cellular response.

MAP Kinase (Mitogen-activated Protein Kinase)
Mediates the events triggered by extracellular signals in the egg. Prohibits pronuclear formation and appears to be responsible for the changes in the organizational state of the microtubules (interphase to metaphase).

Molecular Scaffold
Cellular structures that localize elements of assembly pathways and provide a means for different kinase pathways to interact in discrete regions of the cell.

PKC (Protein Kinase C)
General cytoplasmic signal inducing agent in cells. In the egg, it supports the interphase configuration of DNA and is involved with the second polar body formation and cortical granule exocytosis.

The developmental transition of fertilization is a major organizational and functional change, which initiates a rapid series of changes that restructures the egg into the zygote. A series of cytoplasmic signal transduction events mediate these changes after the sperm penetrates the egg, resulting in a rise in $[Ca^{2+}]_i$. The egg undergoes rapid structural changes that result in the extensive remodeling of this specialized

cell. Two pivotal signaling agents that mediate several of these rapid modifications in cell structure are protein kinase C (PKC) and calcium/calmodulin-dependent protein kinase II (CaM KII). Studies indicate that elements from several of the key signaling pathways, including PKC and CaM KII, colocalize on molecular scaffolds (i.e. meiotic spindle) in the egg and may provide a means for these pathways to interact.

1 Overview of Cytoplasmic Signaling

1.1 General

All cells, regardless of whether they are part of tissue in a multicellular organism or an individual cell functioning as an organism, are faced with the challenge of responding to the external environment. The ability to make an appropriate response to external stimuli can be critical to the functioning of a cell and, certainly, in the case of a single-cell organism, the ability of a cell to reproduce. Consequently, over evolutionary time, complex networks of signal systems have coevolved with cells that provide mechanisms, and probably backup mechanisms that permit short term, relatively rapid changes in cells in response to external stimuli. Short-term cytoplasmic signals are typically mediated through kinases, phosphatases, other enzymes, and calcium to mediate relatively rapid changes in cells. This is in contrast to long-term signaling events, which are mediated by changes in gene transcription and can require multiple days to weeks to result in a change in cell function.

There is a diverse array of short-term cytoplasmic signals, but they typically have several features in common. They have the ability to amplify and propagate the signal rapidly from the site of origin, they can be modified rapidly (e.g. phosphorylated/dephosphorylated or bound to calmodulin/calcium), and the signal(s) can be terminated rapidly. Such signals can propagate globally across the entire cell, as is the case with the calcium signals when the egg is fertilized, or they can act in discrete regions in the cell perhaps by binding to molecular scaffolds. Intracellular signals can be mediated through a variety of mechanisms such as enzyme-linked receptors, G-proteins, gap junctions, phospholipases, hormones, or calcium in a network of pathways that can be highly cross-linked. The level of free calcium in cells often has a role in mediating some portion of a signaling pathway. To keep the level of intracellular-free calcium ($[Ca^{2+}]_i$) low, cells have evolved a variety of mechanisms including enzymes and other cell proteins that are differentially regulated by the level of $[Ca^{2+}]_i$. These proteins have coevolved with mechanisms that promote low levels of intracellular-free calcium, so that an elevation of free calcium will activate or inhibit an array of calcium-dependent proteins to provide a system that can affect numerous cellular functions.

The egg, and specifically the mammalian egg, is a useful system for investigating cytoplasmic signaling events, as this unique cell must rapidly undergo a highly ordered set of changes in structure and function as a result of fusion with the sperm. If

these structural and functional changes do not occur correctly, then the organism fails to propagate its genes. Additional features that make mammalian eggs a good system for investigating signaling events include the following: (1) Inherent in the production of (as well as fertilization of) eggs are natural cell cycle arrest points. The natural cell cycle arrest points provide for synchronous release from the cell cycle arrest. This has permitted the analysis of many of the signaling mechanisms involved in the cell cycle. When eggs are employed, these pathways involve the meiotic cell cycle; only after fertilization does the first mitotic cell cycle initiate. Investigations have revealed features unique to the meiotic cell cycle, such as the function of c-mos in the cell cycle arrest. These investigations have also identified many cell cycle signaling events common to both meiosis and mitosis. (2) In mammalian eggs, the DNA is condensed in the form of chromosomes and cannot participate in transcription. Consequently, at this time in development, the regulation of changes in cell function (i.e. from egg to zygote) is under the control of cytoplasmic signaling events and occurs largely at the protein level.

The illustration (Fig. 1) indicates the pathways that have been reported to be engaged as a result of mammalian egg fertilization and cell cycle resumption. Calcium acts at the top of a hierarchy of an extensive signaling cascade and triggers more than one signaling pathway. Furthermore, in mammalian eggs, the initial calcium rise is followed by oscillations in the level of free calcium above the basal level for several hours. A recent study by Ducibella and coworkers show that calcium oscillations drive the progression of several signaling events that convert the egg to a zygote. Once the fertilization-induced rise in intracellular-free calcium occurs, other calcium-dependent pathways are activated as indicated in Fig. 1.

Calcium/calmodulin-dependent protein kinase II (CaM KII) is activated in mammalian eggs, and the activation of CaM

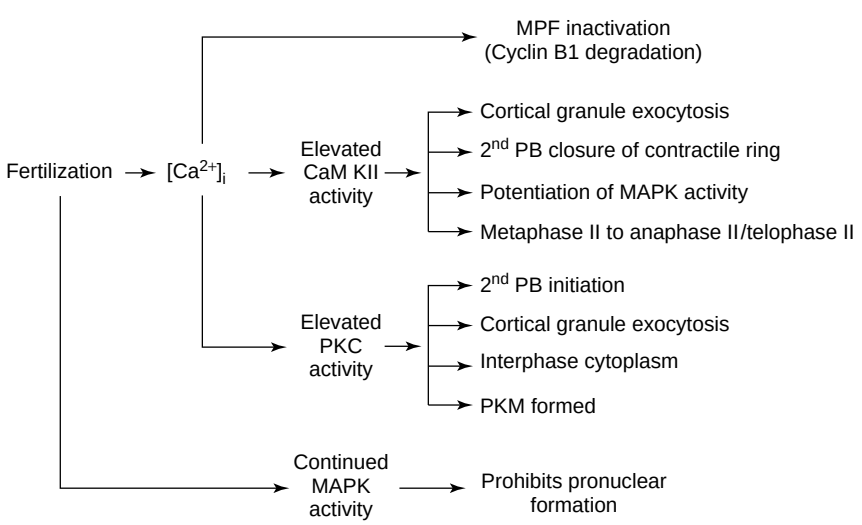

Fig. 1 The illustration indicates pathways that are involved as a result of the rise in intracellular-free calcium that accompanies mammalian fertilization. 2nd PB = second polar body.

KII is dependent on the presence of both calmodulin and calcium. While functions of CaM KII have been demonstrated in eggs from other classes of organisms, there are not many reports concerning CaM KII activation in mammalian eggs. Several roles for CaM KII are summarized in Fig. 1. There is general agreement from many sources that the rise in intracellular-free calcium inactivates maturation promotion factor (MPF) by targeting cyclin B1 for degradation through an ubiquitin-dependent pathway. In addition, Protein kinase C (PKC) acts downstream of the calcium signal and there have been several reports demonstrating this activation at a biochemical or immunocytochemical level in mouse and rat eggs, although the specific isotypes of PKC that act appear to differ depending on the study and species. Once activated, PKC appears to have several roles as summarized in Fig. 1 including the release of PKC's catalytic domain (i.e. PKM). Until recently MAP kinase (mitogen-activated protein kinase) was thought to function independent of the other three pathways in Fig. 1.

1.2 Signals Regulated by Kinases

In the mammalian egg, there are three well-studied kinases that are regulated by calcium. Protein kinase C and calcium/calmodulin-dependent protein kinase II (CaM KII) are activated by calcium, while MPF is inactivated by calcium.

There are at least 12 different isotypes of PKC that range in molecular weight from about 67 to 76 kDa and all phosphorylate serine and threonine residues in protein. The isotypes are organized into three groups referred to as conventional, atypical, and novel by the cofactors required to activate the different isotypes. The conventional PKC isotypes, α, βI, βII, and γ, are activated by diacylglycerol (DAG) and negatively charged phospholipids in a calcium-dependent manner. The atypical PKC isotypes ζ and λ require negatively charged phospholipids but do not require DAG or calcium. The novel PKC isotypes δ, ε, θ, η, and μ do not require calcium but require DAG and phospholipids as cofactors. Some investigators regard PKC μ isotype as a different category of kinase because it has a unique structure. Conventional isotypes are the most likely to act proximally downstream of the calcium signal induced by fertilization of mammalian eggs because they require calcium as a cofactor. It has been shown that cells can contain more than one isotype of PKC and thus, depending on the cofactors generated, members of this kinase family have the potential to influence a variety of different pathways in a single cell as well as have different effects in different cells. The different isotypes can have different substrate specificity allowing one PKC isotype to have different effects on the cell rather than activation of another isotype in the same cell. Work with somatic cells has suggested that at least one of these isotypes, PKC λ, may serve to decode both calcium oscillations and DAG signals. These features are most likely important to mammalian eggs because not only has PKC λ been identified in mouse eggs but at least seven different isotypes of PKC have also been recently identified at the protein level in both mouse eggs and early embryos.

There are several isotypes of calcium/calmodulin-dependent protein kinase II α, β, γ, δ and, the functional enzyme is a heteromultimer composed of several catalytic subunits, but for the kinase to be active not all subunits are required. CaM KII is a large kinase with a molecular weight of 50 kDa and requires an association with calcium

bound to one or more of the calmodulin binding sites. This association activates the kinase and permits phosphorylation on serine/threonine at consensus phosphorylation sites. Once CaM KII is activated it can undergo autophosphorylation on T286/287. Consequently, it can maintain its state of activation after the calmodulin dissociates from CaM KII. CaM KII is thought to be involved with the transition from metaphase to anaphase in both somatic cells and in eggs. There is general agreement from many sources that the rise in intracellular-free calcium in amphibian eggs activates CaM KII, which then activates the ubiquitin-dependent cyclin degradation pathway and also acts on c-mos degradation. At egg activation, an increase in activity of CaM KII can be measured in mammalian eggs. By using antibodies that bind to the β- and δ-isotypes of CaM KII, it has been revealed that specific molecular scaffolds in the egg and early mouse embryo are enriched with CaM KII, as will be discussed in subsequent sections.

One well-studied calcium-dependent event involves inactivation of MPF. MPF is composed of cyclin B1 and $p34^{cdc2}$ and has been indicated as one of the kinases that drive the cell into metaphase. The rise in intracellular-free calcium that accompanies fertilization inactivates MPF by targeting cyclin B1 for degradation through an ubiquitin-dependent proteolytic pathway. MPF activity decreases in the absence of cyclin B1.

2 Coordination of Complex Signaling Pathways

There are numerous signaling pathways involved in the conversion of the egg to the zygote as outlined in an earlier section. These pathways act in a particular temporal sequence and some of the pathways appear to act on the same structure at different times progressively changing the cellular components. This is the case for the formation of the second polar body, which will be described in more detail in a subsequent section. The complexity of the changes that occur is not indicative of a system in which enzymes and specifically kinases are randomly diffused about the cells, but rather of a situation in which kinases and other components have restricted domains within the cells. Two features of the mouse egg are likely to permit the existence of these restricted domains, that is, (1) the meiotic spindle appears to be serving as a molecular scaffold that anchors elements of different signaling pathways; (2) the architecture of the egg itself is highly polarized and may restrict the distribution of signaling components including kinases.

2.1 Molecular Scaffolds

The potential for the meiotic spindle to serve as a molecular scaffold provides the opportunity to study the regulation of, and cross talk among, distinct cytoplasmic signaling pathways in mammalian eggs. MAP kinase pathways and their association with molecular scaffolds have received much study. Such scaffolds, originally identified in yeast as the Ste5 protein, have been shown to bring together many of the upstream regulatory elements in the MAP kinase pathway. Proteins with functional similarities to Ste5 have been identified in higher eukaryotes. Of particular relevance to this review, microtubules have been shown to serve as a molecular scaffold for a variety of signaling pathways including elements of MAP kinase pathways.

For example, the MAP kinases ERK1/2 have been shown to associate with microtubules. In another MAP kinase pathway, the JNK pathway, elements of that pathway (i.e. MLK2, the MAP kinase kinase kinase) associate with the microtubules through kinesin-like molecular motors.

Molecular scaffolds have been investigated in a wide variety of systems and signaling pathways. Results from these studies have led to the identification of the following putative roles for molecular scaffolds: (1) increase the efficiency of cellular reactions by optimally positioning kinases as well as their substrates and products on the scaffold; (2) decrease the response time to a stimulus; (3) enhance some forms of cross talk among pathways while reducing other forms of cross talk (by juxtaposing elements of distinct pathways on the same scaffold); (4) increase substrate specificity and; (5) localize signaling pathways to specific parts of a cell. All of these putative roles would serve useful functions in mammalian eggs since eggs must undergo a highly orchestrated regime of changes rapidly after fertilization and in the absence of new gene transcription.

There are a number of potential advantages in employing microtubules as scaffolds in mammalian eggs. First, in the mammalian egg, microtubules provide a relatively large structure that can serve as a large surface area for the binding of various proteins and components of signaling pathways. Second, the microtubules in M-phase cells exhibit a distinct polarity with their disassembly end embedded in the centrosome and the assembly end closer to the cell equator (for kinetochore and polar microtubules). Should elements of a signaling pathway be associated with a molecular motor protein or activated by something carried on a molecular motor protein, they could be positioned precisely. Third, there are three forms of spindle microtubules, namely, kinetochore, polar, and astral (in mouse eggs astral microtubules are not prominent) that may serve distinct functions for signaling pathways (for example, in the formation of midzone microtubules or in cytokinesis). Fourth, the spindle progresses through distinct, time-dependent changes after fertilization, which can serve to temporarily reposition the scaffold and associated signaling components to distinct locations in the cell at distinct times over the first 1 to 1.5 h postfertilization. Fifth, after fertilization, the meiotic spindle and its derivatives occupy a polarized location in the egg and forming zygote. All of these potential functions of the spindle microtubules as a molecular scaffold regulating cytoplasm signaling pathways could serve to orchestrate the hierarchical regime of structural and functional changes the egg undergoes in a precise spatial and temporal order as it becomes the zygote.

Studies in other systems provide evidence to demonstrate interactions between CaM KII and MAP kinase. For example, CaM KII has been shown to regulate the activity of MAP kinase in smooth muscle tissue although the mechanism of action was not determined. One study examining vascular smooth muscle cells shows that CaM KII acts upstream of ERK 1/2 through a tyrosine kinase. In neurons, CaM KII and MAP kinase colocalize on a scaffold (i.e. the postsynaptic density) and interact. The work on smooth muscle and neurons, together with our own preliminary studies, is very suggestive of an interaction between CaM KII and MAP kinase, or one of the upstream regulators, on the meiotic spindle (i.e. scaffold) in mouse eggs.

2.2 Cell Architecture

2.2.1 Action of Signaling Agents on the Cytoskeleton

The fertilization-competent egg has a variety of molecular scaffolds at distinct locations within the cell. Cortical granules that exocytose after fertilization, create a block to polyspermy and are enriched along the cell periphery of the spherical egg cell. Actin filaments also are enriched at the cell periphery, while unique arrays of intermediate filaments in the form of "cytoskeletal sheets" are in the cell interior and excluded from the cell periphery. The only microtubules in the cell at this time in the cell cycle are the spindle microtubules with their associated DNA in the form of chromosomes. This entire unit (i.e. the spindle and the chromosomes) at meiotic metaphase II is polarized immediately subjacent to the plasma membrane. The egg surface is covered with microvilli except for the region immediately above the meiotic spindle. Even the mitochondria have a polarized distribution with more mitochondria in the hemisphere containing the meiotic spindle.

In mammalian eggs, the microtubule network is acted on by MAP kinase to convert the interphase array of microtubules into the M-phase array of microtubules. The M-phase array of microtubules is the meiotic spindle and this appears to be serving as a molecular scaffold that is involved in the regulation of the complex signaling pathways in the egg.

The actin filament network that is greatly enriched at the cell periphery is a location where many structural changes occur. Correlative data suggest that signaling agents may act on the actin network, such as the localization of active PKC in the cell periphery, shortly after egg activation. However, there is a paucity of direct evidence for this interaction in the mammalian egg.

The intermediate filament network in mammalian eggs is composed of a highly cross-linked network of filaments in the oocyte, fertilization-competent egg, and early embryo. The intermediate filaments are composed of cytokeratins 5, 6, 16, and Z, which are produced by the maternal genome. These cytoskeletal elements referred to as "sheets" serve as a maternal form of intermediate filaments that are involved partly with cell adhesion events during embryonic compaction and extensively with cell adhesion as the blastocyst implants into the uterine wall. This specialized network of intermediate filaments has been identified in the fertilization-competent eggs of a variety of mammals including mouse, rat, hamster, humans, cows, and pigs. PKC acts on these "sheets" as part of the mechanism of downregulating the kinase. As described in more detail in an earlier section, the fertilization-induced rise in calcium activates PKC causing it to translocate to the plasma membrane. At the cell periphery, the PKC signal is downregulated by cleaving the catalytic subunit of PKC from its calcium and membrane binding domain. The catalytic subunit is then free to diffuse into the cell interior (in a form referred to as PKM) where it can phosphorylate substrates in both a membrane- and calcium-independent fashion and then PKM becomes enriched on cytoskeletal sheets. Cytokeratin 16, becomes phosphorylated, along with other components in the sheets, at the same time that this intermediate filament network undergoes extensive remodeling. A recent investigation by Coonrod and coworkers has demonstrated that a peptidylarginine

deiminase-like protein is associated with the cytoskeletal sheets and it becomes active shortly after egg activation where it may serve to aid in the reorganization of the cytoskeletal filaments in the sheets.

3 Formation of the Fertilization-competent Egg

3.1 CaM KII

Several reports have determined that CaM KII is present in the fertilization-competent mouse egg and it also is enriched on the meiotic spindle. A cofactor required for activation of CaM KII, calmodulin, is also present in the egg, but in contrast, very little of it is associated with CaM KII or the meiotic spindle in the metaphase II egg. Calmodulin, in association with CaM KII activity, which results from the basal level of $[Ca^{2+}]_i$ in the egg, has been demonstrated in fertilization-competent eggs by biochemical assays. As will be discussed later, the level of CaM KII activity greatly increases after fertilization.

3.2 PKC in the Mammalian Egg

PKC is another kinase that functions to convert the mammalian oocyte into the fertilization-competent egg. Activation of PKC has been reported to inhibit the disassembly of the germinal vesicle. Thus, one might expect that the activity of PKC would either remain constant, or be reduced, at the time of germinal vesicle disassembly. Several different PKC isotypes exist in fertilization-competent eggs of mice, although there are differences between those represented at the protein level compared to those represented at the RNA level. PKCλ and PKCμ are detected at the RNA level, while at the protein level PKCα, δ, λ, μ, and ξ are detected in the metaphase II egg. The protein level may have more isotypes because the proteins may be of maternal origin. Gangeswaran and Jones also examined mouse PKC isotypes both during the resumption of meiosis to produce the fertilization-competent egg and in the egg itself. However, they did not detect PKC isotypes α, β, γ, ε, η, θ, and ζ but detected PKC isotypes δ and λ in germinal vesicle stage oocytes at both the protein and mRNA level in metaphase II eggs. Raz et al. detected protein and mRNA for PKCα, β, γ, ε, μ, λ, and ζ in oocytes and metaphase II eggs from rat. Only conventional isotypes of PKC (i.e. PKCα, β, and γ) would be candidates for signaling agents that acted proximate to the fertilization-induced $[Ca^{2+}]_i$ since only the conventional isotypes of PKC require calcium as a cofactor for activation. Other PKC isotypes may be active at the time of fertilization, but the function of these isotypes may not be dependent on the generation of a calcium signal.

3.3 MAPK and MPF

Two important kinases that function to convert the oocyte into the fertilization-competent egg are MPF and MAP kinase. At about the time that the oocyte nucleus (i.e. the germinal vesicle) begins to disassemble, the activity of both kinases appears. There is evidence to suggest that in mammalian development, MAP kinase activity appears first, followed by MPF activity, or there is simultaneous activation of MAP kinase and MPF. Data from other investigators have indicated a reverse sequence. As Cyclin B1 is synthesized and

associated with $p34^{cdc2}$ to make active MPF, oocytes arrested in prophase I of meiosis are recruited to resume meiosis. Cyclin degradation is initiated as the oocyte transits between meiosis I and meiosis II. New synthesis of cyclin replenishes the diminished concentration of cyclin and maintains the activity of MPF. This balance of continued degradation and new synthesis of cyclin accompanies the formation of the fertilization-competent egg and explains why, when protein synthesis inhibitors are applied to eggs, the eggs subsequently activate. This is because cyclin continues to be degraded as normal but the inhibitor blocks any new cyclin synthesis. Subsequently, the arrest at meiotic metaphase II cannot be maintained because of the level of MPF. The integrity of a molecular scaffold of the cell, the meiotic spindle, plays an important role in the degradation of MPF within the fertilization-competent egg at meiotic metaphase II. Cyclin degradation is inhibited and the egg remains arrested in meiotic metaphase II if the meiotic spindle is disrupted with a microtubule-disassembling agent. In contrast, the degradation of MPF is much slower between meiosis I and meiosis II and it does not appear to be linked to the integrity of the spindle microtubules. This suggests that in a metaphase II egg, the presence of a molecular scaffold (i.e. the meiotic spindle) is associated with components that provide a mechanism to increase the efficiency of a variety of reactions including cyclin degradation. The colocalization of components on molecular scaffolds in the cell will be discussed below.

Several different MAP kinase pathways have been identified in various cell types functioning through different MAP kinases, for example, ERK 1/2, JNK/SAPK, p38, and ERK 5. The ERK 1/2 pathway has been identified in oocytes and eggs. In this pathway, the upstream activator of MAP kinase is MEK 1/2 (i.e. MAP kinase kinase). Further upstream in this kinase cascade in eggs is c-mos (MAP kinase kinase kinase), whereas in other cell types it is typically one of the Raf isotypes. When oocytes are triggered to resume meiosis from the arrest at prophase I, MAP kinase is activated. MAP kinase is dependent on protein synthesis to be initially activated because if puromycin, an inhibitor of protein synthesis, is applied MAP kinase does not become active in mouse or rat oocytes. However, once MAP kinase is activated, protein synthesis is no longer necessary to maintain MAP kinase activation. C-mos is present in mammalian eggs and is considered a part of cytostatic factor (a factor in the cytoplasm that maintains the arrest at meiotic metaphase II). It is important to note that c-mos is thought to be an upstream activator of MAP kinase and in the absence of c-mos activity, MAP kinase does not become active. Evidence of this activation pathway also comes from studies where the c-mos gene has been disrupted. In such mice, MAP kinase does not become active. Raf, which is a known upstream activator of MAP kinase, is also present in these eggs, but Raf activity was not detected when MAP kinase became active. The activation of MAP kinase requires phosphorylation on both tyrosine and threonine residues to become an active serine/threonine kinase and the application of phosphatase inhibitors speeds MAP kinase activation.

The major changes in organizational state of the microtubules (i.e. the switch from the interphase to the M-phase configuration of microtubules) as well as the disassembly of the germinal vesicle that is required for the formation of the metaphase II egg appear to be due to MAP kinase activation in the oocyte.

Inoue et al. accelerated germinal vesicle disassembly by microinjection of active MAP kinase into the germinal vesicle, which suggests that MAP kinase acts directly on components in the nucleus to mediate nuclear disassembly.

The active form of MAP kinase is enriched on the entire meiotic spindle in the fertilization-competent egg arrested at meiotic metaphase II. To demonstrate the presence of the active form of the kinase, an immunocytochemical assay was performed using antibodies that bind sites on MAP kinase that are phosphorylated only when the kinase is active. By using an antibody that recognized both forms of the kinase (i.e. total MAP kinase and active MAP kinase), the distribution of both the active and total forms of the kinase was compared. Other investigators have also reported association of total MAP kinase with the mammalian meiotic spindle; however, only the poles of the meiotic spindle were reported by the investigators to bind MAP kinase. This may have been observed because the spindles were physically isolated from the metaphase II egg before CaM KII was measured by immunolocalization, and this may have resulted in restriction of MAP kinase to the poles.

4 Activation of the Egg

4.1 Remodeling through Changes in Activity of Calcium/Calmodulin-dependent Protein Kinase II

As previously mentioned, CaM KII's activity increases after fertilization in mouse eggs; however, until recently little was known concerning the function of this activity in mammalian eggs. Johnson et al. demonstrated that CaM KII is enriched throughout the meiotic spindle in the fertilization-competent mouse egg. When either the egg or activated egg is detergent extracted to remove soluble and freely diffusing components from the cell, CaM KII remains behind and appears to be tightly associated with the meiotic spindle microtubules. In contrast, a concentration of calmodulin cannot be detected in association with the spindle in the metaphase II egg after detergent extraction, except for a window of time that is approximately 5 min after egg activation. After activation of an egg with calcium ionophore, calmodulin transiently associates with the detergent-resistant spindle 5 min after the ionophore treatment, but is absent by 15 min posttreatment. Calmodulin would bind to the increase in $[Ca^{2+}]_i$ in the egg caused by a calcium ionophore treatment. Subsequently, one would anticipate that this calcium/calmodulin complex would activate CaM KII by associating with the kinase. This reaction could be representative of the 5 min after ionophore treatment when the transient calmodulin is binding to CaM KII, which is attached to the detergent-resistant spindle. This proposal is supported by Johnson et al., who demonstrated that the level of CaM KII activity almost doubles within 5 min after egg activation, the same window of time that calmodulin is tightly associated with the meiotic spindle. In support of this finding, it was demonstrated by a recent investigation that CaM KII was active during this time interval and that a large portion of the CaM KII that is activated is enriched on the meiotic spindle.

The results described above suggest that activation of CaM KII is specifically involved with some function of the meiotic spindle. The transition of chromosomes

out of metaphase II arrest into anaphase II and subsequently into telophase II is promoted by CaM KII. This was demonstrated in two ways: (1) Living eggs were activated with calcium ionophore after the application of membrane-permeant CaM KII inhibitors, which blocked the transition from metaphase II into anaphase II; (2) When permeabilized eggs were equilibrated in a medium whose free calcium concentration was adjusted to 1 μM (using a calcium/BAPTA buffer), calmodulin was present at a concentration sufficient to establish a 4 : 1 ratio of calcium : calmodulin, and an ATP regenerating system was present; the permeabilized eggs transited through anaphase into telophase with the same timing that would occur in a fertilized living egg. As the ratio of calcium to calmodulin was sequentially dropped to 0 : 1 in the permeabilized egg, the transit to telophase was progressively reduced to zero. When inhibitors to CaM KII were applied, the transit to telophase was blocked, but other kinase inhibitors, including inhibitors to PKC, could not prevent the transit through anaphase into telophase.

The above results strongly indicate that the movement of chromosomes from meiotic metaphase II arrest into anaphase II and subsequently telophase II is regulated by activation of CaM KII. These results also explain investigations discussed earlier, which reported that when PKC is activated in mammalian eggs in the absence of an elevation in the level of intracellular-free calcium, an abnormal transit out of metaphase II occurs. CaM KII is not activated in the absence of a rise in the level of $[Ca^{2+}]_i$ and the chromosomes do not transit into anaphase, but other features of egg activation occur.

It is important to remember that Johnson et al. reported that CaM KII remains on elements of the mammalian meiotic spindle as the cell transits into anaphase and at subsequent stages. Interpolar microtubules form an overlapping assembly in the midzone region during the latter part of M-phase and the microtubules become bundled. Such midzone microtubules contain many spindle-related proteins. It has been proposed that these proteins have a role in the formation of the contractile ring during cleavage. Johnson et al. reported, in accordance with this proposal, that CaM KII appears to transit along the midzone microtubules to the contractile ring of the second polar body and localizes on the contractile ring suggesting it may play an active role in the cytokinetic process. In addition, a recent study by Hatch and Capco demonstrated that the active form of CaM KII associated with the cytokinetic ring, a structure that forms perpendicular to the midzone microtubules. This study confirms and extends on work by earlier investigators who blocked CaM KII activity with drugs. An inhibitor to calmodulin function was employed by Xu et al. and showed that this delayed the formation of the second polar body. Since calmodulin is required to activate CaM KII, the authors speculated that CaM KII activity was needed for the formation of the second polar body. This idea was further supported by two separate studies, which reported that second polar body formation was delayed after inhibitors to CaM KII were employed.

The above observations also address another paradox that was reported when PKC-induced egg activation was investigated. Gallicano et al. reported that after treatment with PKC agonists, the second polar body initially appeared, but then would frequently be absorbed into the egg. When PKC agonists are applied, there is no increase in the level of $[Ca^{2+}]_i$ and

the chromosomes do not transit normally from metaphase into anaphase. Midzone microtubules will not form since the chromosomes do not transit properly into anaphase, which could lead to an incorrect positioning of CaM KII. CaM KII would not be activated even if the kinase did translocate correctly to the contractile ring of the forming second polar body because the $[Ca^{2+}]_i$ levels would not have been elevated. Contractile ring closure may be inhibited in the absence of CaM KII activation. Without a complete separation from the egg, the second polar body may instead become absorbed into the activated egg after a period of time.

As noted in a previous section, CaM KII is tightly associated with the meiotic spindle. However, other cytoplasmic signaling agents including cyclin, c-mos, and MAP kinase have been reported to be associated with at least parts of the spindle in a variety of different systems. The presence of, or localization to, a particular site within a cell does not indicate that a kinase is in its active form. In fact, location of the greatest concentration of a kinase may differ from the site where the active kinase is positioned, as was shown in the study by Hatch and Capco that examined the distribution of CaM KII and MAP kinase in mouse eggs. This study makes use of antibodies that only recognize the active form of the kinase and other antibodies that recognize both the active and inactive forms of the kinase (i.e. the total kinase). The spatial distribution of total CaM KII and total MAP kinase was compared and indicated that the kinases colocalized on the meiotic spindle in the metaphase II egg and on the midzone microtubules after egg activation. However, very little active MAP kinase was present on the midzone microtubules and was only localized on the spindle for 25 min after egg activation. At later postactivation stages, total MAP kinase was still enriched at the midzone microtubules, but at that same location active MAP kinase was not present. In comparison, CaM KII was constantly associated with the spindle microtubules and later the midzone microtubules, whereas active CaM KII was enriched on the spindle only at 5 and 25 min after egg activation, and then enriched on the contractile ring of the second polar body 55 min after egg activation.

Colocalization of CaM KII and MAP kinase on the meiotic spindle provides the opportunity for interaction between the two kinases. This may be particularly important for CaM KII since it becomes active on the spindle 5 min after egg activation at the same time that MAP kinase also is active and associated with the meiotic spindle. An investigation demonstrated that inhibition of CaM KII activity caused a decrease in both the amount of total MAP kinase and active MAP kinase on the spindle. This study suggested that activation of CaM KII (within 5 min of egg activation) serves to potentiate MAP kinase in the active state. There are two forms of MAP kinase in mouse eggs ERK1 and ERK2. Both forms contain consensus phosphorylation sites for CaM KII. These are not the same sites that activate MAP kinase, but phosphorylation by CaM KII on these sites may prevent degradation of MAP kinases. As will be discussed in a later section, it has been proposed that the activity of MAP kinase inhibits pronuclei from forming precociously in the zygote, and after MAP kinase activity drops, pronuclei immediately form. Thus, as the means of regulating the timing of the formation of pronuclei, egg activation induces the activation of CaM KII. This CaM KII activity serves to boost MAP kinase activity

preventing pronuclei from forming too early in the zygote.

CaM KII could have two potential roles in regulating the conversion of the egg into the zygote based on the above studies: (1) to mediate the chromosomal transition from metaphase II arrest into anaphase II and subsequently through telophase II in the fertilization-competent egg and (2) to interact with MAP kinase to potentiate its activity and prevent formation of pronuclei. CaM KII is localized at an optimal site to be involved with both the above-stated roles because it is colocalized with MAP kinase. It subsequently interacts with, and it is localized on, the meiotic spindle, which is a mechanical scaffold in eggs that drives chromosomal movement.

4.2 Remodeling through Changes in Activity of Protein Kinase C

When eggs are treated with PKC agonists, fertilization-competent hamster eggs release from arrest at meiotic metaphase II, initiate second polar body formation, and several hours later pronuclei as well as the Golgi apparatus form (all components that are absent in M-phase cells). A number of the events associated with egg activation are initiated after PKC is activated in eggs using a phorbol ester (i.e. a pharmacologic agent known to activate PKC) or a diacylglycerol, a natural activator of PKC in cells. Even after inducing activation of PKC with agonists while the level of $[Ca^{2+}]_i$ was clamped low with a calcium chelator, these activation events still occurred. However, in this case the meiotic spindle did not undergo a normal metaphase to anaphase transition when the level of $[Ca^{2+}]_i$ was clamped low. Instead, the entire meiotic spindle attempted to translocate into the forming second polar body. When the reverse situation was tested (the level of $[Ca^{2+}]_i$ was elevated by experimental manipulation, and the activity of PKC was clamped low using antagonists) the eggs did not change and remained arrested at meiotic metaphase II. This indicates that PKC acts downstream of the calcium signal, although the specific isotypes of PKC that act appear to differ depending on the study and species. Even though PKC induced a number of changes that are associated with egg activation, some other calcium-dependent feature was needed to cause the normal metaphase to anaphase transition. To confirm that the fertilization-induced event involved PKC at an immunocytochemical level, the investigators monitored the spatial position of several isotypes of PKC by using a PKC reporter dye. In this study, the investigators demonstrated in hamster metaphase II eggs that as a result of the rise in intracellular-free calcium, PKCs translocated from a relatively uniform distribution in the cytoplasm to the cell periphery near the plasma membrane of the egg. Translocation of the reporter dye to the membrane indicated that the kinase was activated. In this study, a DAG, a natural simulator of PKC activity produced *in vivo*, was employed. A DAG was used because the cell itself makes DAGs and contains enzymes that normally metabolize DAG into other compounds. Thus, any DAG when transiently applied is more likely to mimic the normal activation of PKC. Phorbol esters also activate PKC but, in contrast, they are foreign to cells and could likely not be metabolized because they are not normally present in cells. Use of phorbol esters as PKC agonists could lead to abnormal results due to an extended activation of PKC as well as other

perturbations even if the phorbol ester was applied for a brief period.

Mouse eggs also have been used to demonstrate the involvement of PKC in activation. In the study by Gallicano et al., the role of PKC appeared to be identical to that described above for hamster eggs but in addition to application of DAG, sperm inseminated eggs were utilized. It was shown that sperm have the same effect as other PKC agonists, that is, after sperm penetration PKC in the egg translocates to the cell periphery. In the Gallicano et al. study, the investigators used another method, a biochemical assay, to confirm that PKC was translocating to the cell periphery at the same time that the activity of the kinase increased. In addition, at the site of second polar body initiation, PKC was shown to be enriched. Raz et al. used rat eggs to demonstrate that specific PKC isotypes translocate to the cell periphery in mammalian eggs. Upon activation of the kinase in the rat egg, PKC isotypes α, β, and γ translocate to the plasma membrane. These results fit well with the model that proposes that PKC acts downstream of the fertilization-induced elevation in $[Ca^{2+}]_i$ since, noted previously PKC α, β, and γ are calcium-dependent isotypes. This observation supports the earlier work of Gallicano et al. who, using a PKC reporter dye that detects several PKC isotypes, showed, as a result of egg activation and fertilization that PKC became active and translocated to the plasma membrane of mammalian eggs. In addition, a more recent study also supports this model. Luria et al. have examined PKC α, $\beta 1$, and $\beta 2$ in mouse eggs and report that PKCα translocates to the plasma membrane as a result of egg activation and fertilization. Two other laboratories did not detect either PKC$\beta 1$ or PKC$\beta 2$ isotypes at the protein or RNA levels in mouse eggs.

Activated PKC drives the egg into an interphase state of the cell cycle and has been shown to be involved with cortical granule exocytosis in fertilization-competent rat eggs. In the study by Raz et al., PKC was activated by both phorbol esters and DAGs and showed that activation of PKC induced both of these events. Cortical granule exocytosis in mouse eggs has also been reported after PKC is activated. However, since PKC inhibitors cannot block the fertilization-induced exocytosis of cortical granules, this indicates that the PKC pathway and other calcium-dependent signaling pathways also induce cortical granule exocytosis (Ducibella and LeFevre, 1997; Raz et al., 1998c). There is a report that inhibition of CaM KII blocks cortical granule exocytosis, but another study claims that CaM KII has no effect on cortical granule exocytosis. In mammals, the cortical granule exocytosis pathway becomes linked to calcium as the oocyte proceeds through meiosis to become the fertilization-competent egg. As mentioned above, PKC is known to translocate to the cell periphery near the plasma membrane and since the cortical granules are located in this region, PKC may serve as one of the regulators of cortical granule exocytosis. Gallicano et al. have proposed that PKC regulates the remodeling of the egg into the zygote in a very precise temporal and spatial fashion. In that model, active PKC translocates to the plasma membrane, which is induced by the fertilization-induced rise in $[Ca^{2+}]_i$. Once PKC becomes active, it is tethered to the cytoplasmic face of the plasma membrane and can phosphorylate substrates in that vicinity. It may serve to trigger a number of events including cortical granule exocytosis. One of the mechanisms

for active PKC to be downregulated is to cleave its membrane and calcium binding domains from the catalytic subunit. Once cleaved, the catalytic subunit can diffuse away from the plasma membrane while the catalytic subunit remains as an active kinase that is no longer calcium- or membrane-dependent. In this state, it has historically been referred to as PKM. Now that PKM is not membrane tethered, it can phosphorylate elements in the interior of the mammalian egg. In support of this model, Gallicano et al. have demonstrated that after activation of PKC, PKM is generated, and that it phosphorylates components of the cytoskeletal sheets and other substrates in the interior of the egg. PKC has both temporal and spatial precision because PKM can only form after PKC is activated. Thus, as a result of the fertilization-induced elevation in the level of $[Ca^{2+}]_i$, PKC initially modifies the cell periphery and then components in the cell interior are later modified by PKM. Such a model fits well with the restructuring of an egg into a zygote because, by modifying the cell periphery (i.e. cortical granule exocytosis) immediately, polyspermy is prevented and then a more gradual modification of the interior can follow.

By using pharmacologic agents that activate PKC, Moses and Kline have investigated roles for PKC in mouse eggs and have found that when such agents are applied, metaphase II eggs are driven into interphase, a characteristic of egg activation. However, transition into interphase had a different mechanism than an egg was artificially activated by calcium ionophore or inseminated by sperm. Moses and Kline report that the PKC agonist, PMA, caused abnormal and greatly slowed metaphase to anaphase transition. In part, this parallels Gallicano and coworker's results that show that eggs exhibit an abnormal transit to anaphase when PKC is activated by agonists and the level of calcium is clamped low. In agreement with this result, Moses and Kline also show that application of PMA to eggs at meiotic metaphase II does not result in an elevation in the level of $[Ca^{2+}]_i$. These results suggest that the metaphase to anaphase transition requires a component of the cell other than PKC that requires an elevation of the level of $[Ca^{2+}]_i$.

Whether activation of PKC with agonists induces second polar body formation was also addressed by Moses and Kline. The investigators conclude that even though the cell is ultimately driven into interphase, second polar body formation is not induced by activation of PKC. However, two important considerations concerning polar body formation were not taken into account. First, unlike cytokinesis, polar body formation is a two-step process in mammals, with an initial extrusion of cytoplasm (bounded by the plasma membrane) followed by formation of a contractile ring at the base of the extruded cytoplasm. Second, when PMA is applied to mouse eggs, the second polar body initially appears (initial extrusion stage), but frequently, because the contractile ring does not close, the polar body is absorbed into the egg. By the time the cell reaches telophase, the extruded second polar body appears. The cytoplasmic extrusion of the initiated second polar body can be clearly seen in figure 2b of the 1995 publication by Moses and Kline, which shows a telophase stage mouse egg. This inability of the egg to completely close the contractile ring of the second polar body suggests that some other pathway not activated by PMA is responsible for contractile ring closure. As was discussed in the previous section,

it has been suggested that CaM KII is needed for closure of the contractile ring of the second polar body.

By activating PKC with diacylglycerol, Colonna et al. conducted a detailed study to determine the role of PKC on egg activation. Their results indicate that MPF was downregulated by PKC. Colonna et al. demonstrated that the MPF level (monitored as histone H1 kinase activity) drops within 20 min of PKC activation, the chromosomes have a normal transit from metaphase to anaphase in a high percentage of eggs, pronuclei form several hours later, and the second polar body forms. These investigators used a concentration of DAG that did not increase the level of $[Ca^{2+}]_i$, but they did show that the level of $[Ca^{2+}]_i$ increases to about 0.5 μM at higher concentrations of DAG (i.e. 150 μM). This transient increase in the level of $[Ca^{2+}]_i$ may have been caused by the higher concentration of DAG because, as the DAG at this concentration intercalated into the plasma membrane of the egg, it could have affected the lipid packing in the membrane allowing calcium to leak in from the extracellular medium. The results from Moore et al., which differ from Colonna et al., did not detect either a formation of second polar bodies or a decrease in MPF activity (monitored as histone H1 kinase activity). However, to assess the effects on MPF activity, Colonna et al. employed DAGs, the natural activator of PKC in cells, and Moore et al. used phorbol esters to activate eggs. As noted previously in this section, phorbol esters, which are foreign to cells, are not as likely to be properly metabolized and regulated by cells compared to the natural activator of PKC (i.e. diacylglycerols). The unusual results may have been induced by the phorbol esters used by Moore et al.

4.3 Remodeling through Changes in Activity of MPF and MAP Kinase Activity in Activated Eggs

As discussed in an earlier section, cyclin B1 is targeted for degradation through the ubiquitin-dependent pathway by the rise of $[Ca^{2+}]_i$ in mammalian eggs, which results in the rapid inactivation (i.e. within 30 min to 1 h) of MPF, while MAP kinase activity remains elevated for a few hours and then decreases. In eggs pretreated with drugs that stabilize MPF (nocodazole and cyclohexamide), MAP kinase activity still decreased after several hours, which suggests that MPF degradation is not a requirement for the eventual degradation of MAP kinase. Just before formation of two pronuclei – one containing the male chromatin and one containing the female chromatin – MAP kinase is inactivated. The activation of MAP kinase and disassembly of the oocyte nucleus parallels the relationship between inactivation of MAP kinase followed by assembly of the nuclear envelope, that is, the nuclear envelope disassembles when MAP kinase is active. There are three lines of evidence that suggest that MAP kinase activity is incompatible with the existence of a nucleus. (1) Formation of pronuclei is inhibited when a constitutively active form of the upstream activator of MAP kinase (i.e. MAP kinase kinase referred to as MEK) were microinjected into activated eggs. (2) Formation of pronuclei is inhibited when a phosphatase inhibitor (i.e. okadaic acid) is employed to prevent dephosphorylation of MAP kinase, a treatment that maintains MAP kinase activity. (3) The oocyte nucleus disassembles when MAP kinase becomes active in oocytes, and disassembly of the nucleus occurs rapidly when activated

MAP kinase is injected into the oocyte nucleus. Thus, MAP kinase may serve to regulate both the organization of microtubules and the state of the nucleus; when microtubules are in an M-phase configuration and nuclei are absent, MAP kinase is active, whereas nuclei form and an interphase array of microtubules assemble when MAP kinase activity is reduced or absent.

One of the hallmarks of egg activation and functions to prevent polyspermy is cortical granule exocytosis, which is not under the control of MAP kinase or MPF. This was demonstrated by inhibiting protein synthesis in eggs arrested at meiotic metaphase II. It has been shown that a decrease in MPF activity is caused by a decrease in cyclin B1 concentration, which can be decreased by inhibiting protein synthesis that can also slowly decrease MAP kinase activity. Consequently, the egg will be released from arrest at meiotic metaphase II after inhibition of protein synthesis. However, cortical granule exocytosis did not occur after a decrease in MFP and MAP kinase activity, suggesting that the regulation of cortical granule exocytosis is not under the control of MFP or MAP kinase. Since a rise in $[Ca^{2+}]_i$ that accompanies fertilization or egg activation triggers cortical granule exocytosis, this suggests that a calcium-dependent component of the cell other than MPF regulates this event. In support of this assertion, the activation of another kinase, PKC (whose conventional isotypes are calcium dependent) induces cortical granule exocytosis as discussed in Sect. 4.2.

Although c-mos is reported to be an important component of cytostatic factor, an activity present in metaphase II eggs thought to arrest the egg cell cycle at meiotic metaphase II, the activity of c-mos remains elevated well after the time that MPF is degraded and the cell has proceeded into telophase. However, when using double-stranded RNA interference (RNAi) to destroy c-mos in eggs, the eggs activate indicating c-mos clearly has a role in egg activation.

5 Concluding Remarks

Signaling pathways trigger specific changes in the conversion of the egg into the zygote. This review highlighted several signaling pathways that have the potential to interact on molecular scaffolds within the egg. The egg may use these scaffolds to localize components of the various signal pathways and to allow cross talk between different pathways. Scaffolds also provide a means to promote rapid changes in the functioning and modulation of the signaling pathways. In the egg, the meiotic spindle was described as a major molecular scaffold for localizing elements of the signaling pathways to coordinate early developmental events downstream of the calcium signal.

The signaling agents that trigger many of the early events after fertilization have been identified, although clearly more work is needed to define other pathways. Much less is known of the signaling components that trigger events occurring several hours after fertilization and affecting such processes as pronuclear transcription or embryonic genome activation. These will likely be the focus of many research efforts in the future. Investigating the putative interactions between the signaling pathways can lead to improvements in cloning technology applied to agricultural animals. By studying these signaling factors, the success rates of nuclear transfer can be improved, which would make this biotechnology approach more robust.

See also Calcium Biochemistry; Developmental Cell Biology.

Bibliography

Books and Reviews

Burack, W.R., Shaw, A.S. (2000) Signal transducation hanging on a scaffold, *Cur. Opin. Cell Biol.* **12**, 211–216.

Capco, D.G. (1995). Cytoskeletal Mechanisms During Animal Development, in: Capco, D.G. (Ed.) *Current Topics in Developmental Biology*, Vol. 31, Academic Press, San Diego.

Chen, Z., Gibson, T.B., Robinson, F., Silvestro, L., Pearson, G., Xu, B., Wright, A., Vanderbilt, C., Cobb, M.H. (2001b) MAP kinases, *Chem. Rev.* **101**, 2449–2476.

Denouel-Galy, A., Douville, E.M., Warne, P.H., Papin, C., Laugier, D., Calothy, G., Downward, J., Eychene, A. (1997) Murine Ksr interacts with MEK and inhibits Ras-induced transformation, *Curr. Biol.* **8**, 46–55.

Gallicano, G.I., Capco, D.G. (1995) Remodeling of the specialized intermediate filament network in mammalian eggs and embryos during development: regulation by protein kinase C and protein kinase M, *Curr. Top. Dev. Biol.* **31**, 277–320.

Gallicano, G.I., Yousef, M.C., Capco, D.G. (1997a) PKC – a pivotal regulator of early development, *BioEssays* **19**, 29–36.

Kennedy, M.B. (1995). CaM-Dependent PK II (vertebrates), in: Hardie, G., Hanks, S. (Eds.) *The Protein Kinase Facts Book: Protein-Serine Kinases*, Academic Press, San Diego, pp. 131–134.

Margolis, R.L., Andreassen, P.R. (1993) The telophase disc: its possible role in mammalian cell cleavage, *BioEssays* **15**, 201–207.

McGaughey, R.W., Racowsky, C., Rider, V., Baldwin, K., DeMarais, A.A., Webster, S.D. (1990) Ultrastructural correlates of meiotic maturation in mammalian oocytes, *J. Electron Microsc. Tech.* **16**, 257–280.

Murray, A.W., Fournier, A., Hardy, S.J. (1987) Proteolytic activation of protein kinase C: a physiological reaction, *Trends Biosci.* **12**, 55–56.

Sagata, N. (1997) What does mos do in oocytes and somatic cells? *BioEssays* **19**, 13–31.

Soderling, T.R. (1995) Calcium-dependent protein kinases in learning and memory, *Adv. Sec. Mess. Phos. Res* **30**, 175–189.

Toker, A. (1998) Signaling through protein kinase C, *Front. Biosci.* **3**, D1134–D1147.

Whitmarsh, A.J., Davis, R.J. (1998) Structural organization of MAP-kinase signaling modules by scaffold proteins in yeast and mammals, *TIBS* **23**, 481–485.

Yeaman, C., Grindstaff, K.K., Hansen, M.D.H., Nelson, W.J. (1999) Cell polarity: versatile scaffolds keep things in place, *Curr. Biol.* **9**, R515–R517.

Primary Literature

Abbott, A.L., Fissore, R.A., Ducibella, T. (1999) Incompetence of preovulatory mouse oocytes to undergo cortical granule exocytosis following induced calcium oscillations, *Dev. Biol.* **207**, 38–48.

Abraham, S.T., Benscoter, H.A., Schworer, C.M., Singer, H.A. (1997) A role for Ca^{2+}/calmodulin-dependent protein kinase II in the mitogen-activated protein kinase signaling cascade of cultured rat aortic vascular smooth muscle cells, *Circ. Res.* **81**, 575–584.

Araki, K., Naito, K., haraguchi, S., Suzuki, R., Yokoyma, M., Inoue, M., Aizawa, S., Toyoda, Y., Sato, E. (1996) Meiotic abnormalities of c-mos knockout mouse oocytes: activation after first meiosis or entrance in third meiotic metaphase, *Biol. Reprod.* **55**, 11315–11324.

Bailly, E., Pines, J., Hunter, T., Bornens, M. (1992) Cytoplasmic accumulation of cyclin B1 in human cells: association with a detergent-resistant compartment with the centrosome, *J. Cell Sci.* **101**, 529–545.

Brondello, J.M., Pouyssegur, J., McKenzie, F.R. (1999) Reduced MAP kinase phosphatase-1 degradation after p42/p44MAPK-dependent phosphorylation, *Science* **286**, 2514–2517.

Calarco, P.G. (1995) Polarization of mitochondria in the unfertilized mouse oocyte, *Dev. Genet.* **16**, 36–43.

Cao, L.-G., Wang, Y.-L. (1996) Signals from the spindle midzone are required for the stimulation of cytokinesis in cultured epithelial cells, *Mol. Biol. Cell* **7**, 225–232.

Capco, D.G., McGaughey, R.W. (1986) Cytoskeletal reorganization during early mammalian development: analysis using embedment-free sections, *Dev. Biol.* **115**, 446–450.

Capco, D.G., Gallicano, G.I., McGaughey, R.W., Downing, K.H., Larabell, C.A. (1993) Cytoskeletal sheets of mammalian eggs and embryos: a lattice-like network of intermediate filaments, *Cell. Motil. Cytoskeleton* **24**, 85–99.

Capco, D.G. (1996) Specialization in cytoskeletal function during early development, *The Cytoskeleton* **2**, 59–112.

Capco, D.G. (2001). Molecular and biochemical regulation of early mammalian development, *Int. Rev. Cytol.* **207**, 195–235.

Chen, C.-S., Poenie, M. (1993) New fluorescent probes for protein kinase C, *J. Biol. Chem.* **268**, 15812–15822.

Chen, H.-X., Otmakhov, N., Strack, S., Colbran, R.J., Lisman, J.E. (2001) Is persistent activity of calcium/calmodulin-dependent kinase required for maintenance of LTP? *J. Neurophysiol.* **85**, 1368–1376.

Chesnel, F., Eppig, J.J. (1995a) Induction of precocious germinal vesicle breakdown (GVB) by GVB-incompetent mouse oocytes: possible role of mitogen-activated protein kinases rather than $p34^{cdc2}$ kinase, *Biol. Reprod.* **52**, 895–902.

Chesnel, F., Eppig, J.J. (1995b) Synthesis and accumulation of p34cdc2 and cyclin B in mouse oocytes during acquisition of competence to resume meiosis, *Mol. Reprod. Dev.* **40**, 503–508.

Choi, K., Satterberg, Y., Lyons, D.M., Elion, E.A. (1994) Ste5 tethers multiple protein kinases in the MAP kianse cascade required for mating in S. cerevisiae, *Cell* **78**, 499–512.

Cobb, M.H. (1999) MAP kinase pathways, *Prog. Biophys. Mol. Biol.* **71**, 479–500.

Collas, P., Sullivan, E.J., Barnes, F.L. (1993) Histone H1kinase activity in bovine oocytes following calcium stimulation, *Mol. Reprod. Dev.* **34**, 224–231.

Colonna, Tatone, R.C., Francione, A., Rosati, F., Callaini, G., Corda, D., DiFrancesco, L. (1997) Protein kinase C is required for the disappearance of MPF upon artificial activation in mouse eggs, *Mol. Reprod. Dev.* **48**, 292–299.

Dedieu, T., Gall, L., Crozet, N., Sevellec, C., Ruffini, S. (1996) Mitogen-activated protein kinase activity during goat oocyte maturation and the acquisition of meiotic competence, *Mol. Reprod. Dev.* **45**, 351–358.

Ducibella, T., Ukena, T., Karnovsky, M., Anderson, E. (1977) Changes in cell surface and cortical cytoplasmic organization during early embryogenesis in the preimplantation mouse embryo, *J. Cell Biol.* **74**, 153–167.

Ducibella, T., Kurasawa, S., Kopf, G.S., Schultz, R.M. (1993) Regulation of polyspermy block in the mouse egg: maturation dependent differences in cortical granule exocytosis and zona pellucida modifications induced by inositol 1,4,5-triphosphate and an activator of protein kinase C, *Biol. Reprod.* **48**, 1251–1257.

Ducibella, T., LeFevre, L. (1997) Study of protein kinase C antagonists on cortical granule exocytosis and cell-cycle resumption in fertilized mouse eggs, *Mol. Reprod. Dev.* **46**, 216–226.

Ducibella, T., Huneau, D., Angelichio, E., Xu, Z., Schultz, R.M., Kopf, G.S., Fissore, R., Madoux, S., Ozil, J.-P. (2002) Egg-to-embryo transition is driven by differential responses to Ca^{2+} oscillation number, *Dev. Biol.* **250**, 280–291.

Fishkind, D.J., Silverman, J.D., Wang, Y.-L. (1996) Function of spindle microtubules in directing cortical movement and actin filament organization in dividing cultured cells, *J. Cell Sci.* **109**, 2041–2051.

Gavin, A.C., Cavador, J.C., Shorderet-Slatkin, S. (1994) Histone H1 kinase activity, germinal vesicle breakdown, and M phase entry in mouse oocytes, *J. Cell Sci.* **107**, 275–583.

Gallicano, G.I., McGaughey, R.W., Capco, D.G. (1992) Cytoskeletal sheets appear as universal components of mammalian eggs, *J. Exp. Zool.* **263**, 194–203.

Gallicano, G.I., Schwarz, S.M., McGaughey, R.W., Capco, D.G. (1993) Protein kinase C, a pivotal regulator of hamster egg activation, functions after elevation of intracellular free calcium, *Dev. Biol.* **156**, 94–106.

Gallicano, G.I., Larabell, C.A., McGaughey, R.W., Capco, D.G. (1994) Novel cytoskeletal elements in mammalian eggs are composed of a unique arrangement of intermediate filaments, *Mech. Dev.* **45**, 211–226.

Gallicano, G.I., McGaughey, R.W., Capco, D.G. (1995) Protein kinase M, the cytosolic counterpart of protein kinase C remodels the internal cytoskeleton of the mammalian egg during activation, *Dev. Biol.* **167**, 482–501.

Gallicano, G.I., McGaughey, R.W., Capco, D.G. (1997b) Activation of protein kinase C after fertilization is required for remodeling the mouse egg into the zygote, *Mol. Reprod. Dev.* **46**, 587–601.

Gangeswaran, R., Jones, K. (1997) Unique protein kinase C profile in mouse oocytes: lack of calcium-dependent conventional isoforms suggested by rtPCR and Western blotting, *FEBS Lett.* **412**, 309–312.

Ginnan, R., Singer, H.A. (2002) CaM kinase II-dependent activation of tyrosine kinases and ERK1/2 in vascular smooth muscle, *Am. J. Physiol. Cell Physiol.* **282**, C754–C761.

Giovannini, M.G., Blitzer, R.D., Wong, T., Asoma, K., Tsokas, P., Morrison, J.H., Iyengar, R., Landau, E.M. (2001) Mitogen-activated protein kinase regulates early phosphorylation and delayed expression of Ca^{2+}/calmodulin-dependent protein kinase II in long-term potentiation, *J. Neurosci.* **21**, 7053–7062.

Goren, S., Piontkewitz, Y., Dekel, N. (1994) Meiotic arrest in incompetent rat oocytes is not regulated by cAMP, *Dev. Biol.* **166**, 11–17.

Gundersen, G.G., Cook, T.A. (1999) Microtubules and signal transduction, *Curr. Opin. Cell Biol.* **11**, 81–94.

Hanson, P.I., Schulman, H. (1992) Inhibitory autophosphorylation of multifunctional Ca^{2+}/calmodulin-dependent protein kinase analyzed by site-directed mutagenesis, *J. Biol. Chem.* **267**, 17216–17224.

Hashimoto, E., Yamamura, H. (1991) Protease-activated protein kinase C in rat liver, *Int. J. Biochem.* **23**, 507–512.

Hatch, K.R., Capco, D.G. (2001) Co-localization of CaM KII and MAP Kinase on architectural elements of the mouse egg: potentiation of MAP Kinase activity by CaM KII, *Mol. Reprod. Dev.* **58**, 69–77.

Hofmann, J. (1997) The potential for isoenzyme-selective modulation of protein kinase C, *FASEB J.* **11**, 649–669.

Huang, K.-P., Huang, F.L. (1986) Immunocytochemical characterization of rat brain protein kinase C, *J. Biol. Chem.* **261**, 14781–14787.

Inagaki, N., Suzuki, S., Kitai, H., Nakatogawa, N., Kuij, N., Iwahashi, K., Yoshimura, Y. (1997) Effect of KN-62, a selective inhibitor of calmodulin-dependent kinase II, on mouse oocyte activation, *J. Assist. Reprod. Gen.* **14**, 609–616.

Inoue, M., Kishimoto, M.A., Takai, Y., Nishiauka, Y. (1977) Studies on a cyclic nucleotide-independent protein kinase and its proenzyme in mammalian tissues, *J. Biol. Chem.* **252**, 7610–7616.

Inoue, M., Naito, K., Nakayama, T., Sato, E. (1998) Mitogen-activated protein kinase translocates into the germinal vesicle and induces germinal vesicle breakdown in porcine oocytes, *Biol. Reprod.* **58**, 130–136.

Jaken, S. (1990) Protein kinase C and tumor promoters, *Curr. Opin. Cell Biol.* **2**, 192–197.

Johannes, F.J., Prestle, J., Eis, S., Oberhagemann, P., Pfizenmaier, K. (1994) PKC: is a novel, atypical member of the protein kinase C family, *J. Biol. Chem.* **269**, 6140–6148.

Johnson, J., Bierle, B.M., Gallicano, G.I., Capco, D.G. (1998) Calcium/calmodulin-dependent protein kinase II and calmodulin: regulators of the meiotic spindle in mouse eggs, *Dev. Biol.* **204**, 464–477.

Kanaseki, T., Ikeuchi, Y., Sugiura, H., Yamauchi, T. (1991) Structural features of Ca^{2+}/calmodulin-dependent protein kinase II revealed by electron microscopy, *J. Cell Biol.* **115**, 1049–1060.

Katoch, S.S., Su, X.L., Moreland, R.S. (1999) Ca^{2+} and protein kinase C-dependent stimulation of mitogen activated protein kinase in detergent -skinned vascular smooth muscle, *J. Cell. Physiol.* **179**, 208–217.

Kennedy, M.B. (1998) Signal transduction molecules a the glutamatergic postsynaptic membrane, *Brain Res. Rev.* **26**, 243–257.

Kim, I., Je, H.-D., Gallant, C., Zhan, Q., Riper, D.V., Badwey, J.A., Singer, H.A., Morgan, K.G. (2000) Ca^{2+}-calmodulin-dependent protein kinase II-dependent activation of contractility in ferret aorta, *J. Physiol.* **626**, 367–374.

King, R.W., Deshaies, R.J., Peters, J.-M., Kirschner, M.W. (1996) How proteolysis drives the cell cycle, *Science* **274**, 1652–1658.

Kishimoto, A., Nishiyama, K., Nakanishi, H., Uratsuji, Y., Nomura, H., Takeyama, Y., Nishizuka, Y. (1985) Studies on the phosphorylation of myelin basic protein by protein kinase C and adenosine 3′, 5′-monophosphate-dependent protien kinase C, *J. Biol. Chem.* **260**, 12492–12499.

Kubiak, J.Z., Weber, M., de Pennart, H., Winston, N.J., Maro, B. (1993) The metaphase II arrest in mouse oocytes is controlled through microtubule-dependent destruction of cyclin B in the presence of CSF, *EMBO J.* **12**, 3773–3778.

LeFevre, B., Pesty, A., Koziak, K., Testart, J. (1992) Protein kinase C modulators influence meiosis kinetics by not fertilizability of mouse oocytes, *J. Exp. Zool.* **264**, 206–313.

Lehtonen, E., Badley, R.A. (1980) Localization of cytoskeletal proteins in preimplantation mouse embryos, *J. Embryol. Exp. Morph.* **55**, 211–225.

Lindsay, H.D., Whitaker, M.J., Ford, C.C. (1995) Calcium requirements during mitotic cdc2 kinase activation and cyclin degradation in *Xenopus* egg extracts, *J. Cell Sci.* **108**, 3557–3568.

Liu, L., Yang, X. (1999) Interplay of maturation-promoting factor and mitogen-activated protein kinase inactivation during metaphase-to-interphase transition of activated bovine oocytes, *Biol. Reprod.* **61**, 1–7.

Longo, F.H., Chen, D.Y. (1985) Development of cortical polarity in mouse eggs: involvement of the meiotic apparatus, *Dev. Biol.* **107**, 382–394.

Lorca, T., Galas, S., Fesquet, D., Devault, A., Cavadore, J.-C., Dorèe, M. (1991) Degradation of the proto-oncogene product $p39^{mos}$ is not necessary for cyclin proteolysis and exit from meiotic metaphase: requirement for a Ca^{2+}-calmodulin dependent event, *EMBO J.* **8**, 2087–2093.

Lorca, T., Cruzalegul, F.H., Fesquet, D., Cavadore, J.-C., Mèry, J., Means, A., Dorèe, M. (1993) Calmodulin-dependent protein kinase II mediates inactivation of MPF and CSF upon fertilization of *Xenopus* eggs, *Nature* **366**, 270–273.

Lorca, T., Abrieu, A., Means, A., Doree, M. (1994) Ca^{2+} is involved through type II calmodulin-dependent protein kinase in cyclin degradation and exit from metaphase, *Biochim. Biophys. Acta* **29**, 325–332.

Luria, A., Tennenbaum, T., Sun, Q.Y., Rubinstein, S., Breitbart, H. (2000) Differential localization of conventional protein kinase C isotype during mouse oocyte development, *Biol. Reprod.* **62**, 1564–1570.

Marcus, S., Polverino, A., Barr, M., Wigler, M. (1994) Complexes between STE5 and components of the pheromone-responsive mitogen activated protein kinase module, *Proc. Natl. Acad. Sci* **91**, 7762–7766.

Martineau, S.N., Andreassen, P.R., Margolis, R.L. (1995) Delay of HeLa cell cleavage into interphase using dihydrocytochalasin B: retention of a postmitotic spindle and telophase disc correlates with synchronous cleavage recovery, *J. Cell Biol.* **131**, 191–205.

Maro, B., Johnson, M.H., Pickering, S.J. (1984) Changes in actin distribution during fertilization of the mouse egg, *J. Embryol. Exp. Morph.* **81**, 211–237.

Moore, G.D., Kopf, G.S., Schultz, R.M. (1995) Differential effect of activators of protein kinase C on cytoskeletal changes in mouse and hamster eggs, *Dev. Biol.* **170**, 519–530.

Moos, J., Visconti, P.E., Moore, G.D., Schultz, R.M., Kopf, G.S. (1995) Potential role of mitogen-activated protein kinase in pronuclear envelope assembly an disassembly following fertilization of mouse eggs, *Biol. Reprod.* **53**, 692–699.

Moos, J., Xu, Z., Schultz, R.M., Kopf, G.S. (1996a) Regulation of nuclear envelope assembly/disassembly by MAP kinase, *Dev. Biol.* **175**, 358–361.

Moos, J., Kopf, G.S., Schultz, R.M. (1996b) Cycloheximide-induced activation of mouse eggs: effects on cdc2/cyclinB and MAP kinase activities, *J. Cell Sci.* **109**, 739–748.

Morin, N., Abrieu, A., Lorca, T., Martin, F., Doree, M. (1994) The proteolysis-dependent metaphase to anaphase transition: calcium/calmodulin-dependent protein kinase II mediates onset of anaphase in extracts prepared from unfertilized *Xenopus* eggs, *EMBO J.* **15**, 4343–4352.

Morrison, D.K. (2001) KSR: A MAPK scaffold of the Ras pathway? *J. Cell Sci.* **114**, 1609–1612.

Moses, R.M., Kline, D. (1995) Calcium-independent, meiotic spindle-dependent metaphase-to-interphase transition in phorbol ester-treated mouse eggs, *Dev. Biol.* **171**, 111–122.

Moses, R.M., Kline, D., Masui, Y. (1995) Maintenance of metaphase in colcemid-treated mouse eggs by distinct calcium and 6-dimethylaminopurine (6-DMAP)-sensitive mechanisms, *Dev. Biol.* **167**, 329–337.

Muthalif, M.M., Benter, I.F., Uddin, M.R., Malik, K.U. (1996) Calcium/calmodulin-dependent protein kinase II alpha mediates activation of mitogen-activated protien kinase and cytosolic phospholipase A2 in norepinephrine-induced arachidonic acid release in rabbit aortic smooth muscle cells, *J. Biol. Chem.* **271**, 30149–30157.

Nagata, K.I., Puls, A., Futter, C., Aspenstrom, P., Schaefer, E., Nakata, T., Hirokawa, N., Hall, A. (1998) The MAP kinase kinase kinase MLK2 co-localizes with activated JNK along microtubules and associates with kinesin superfamily motor KIF3, *EMBO J.* **17**, 149–158.

Nishizuka, Y. (1988) The molecular heterogeneity of protein kinase C and its implications for cellular regulation, *Nature* **334**, 661–665.

Nislow, C., Lombillo, V.A., Kuriyama, R., McIntosh, J.R. (1992) A plus-end-directed motor enzyme that moves antiparallel microtubules *in vitro* localizes to the interzone of mitotic spindles, *Nature* **359**, 543–547.

Oancea, E., Meyer, T. (1998) Protein kinase C as a molecular machine for decoding calcium and diacylglycerol signals, *Cell* **95**, 307–318.

Oehrlein, S.A., Maelicke, A., Herget, T. (1998) Expression of protein kinase C gene family members is temporally and spatially regulated during neural development in vitro, *Eur. J. Cell Biol.* **77**, 323–337.

Ohta, Y., Ohba, T., Miyamoto, E. (1990) Ca^{2+}/calmodulin-dependent protein kinase II: localization in the interphase nucleus and the mitotic apparatus of mammalian cells, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5341–5345.

Ookata, K., Hisanaga, S., Okano, T., Tachibana, K., Kishimoto, T. (1992) Relocation and distinct subcellular localization of $p34^{cdc2}$ – cyclin B complex at meiosis reinitiation in starfish oocytes, *EMBO J.* **11**, 1763–1772.

Pauken, C.M., Capco, D.G. (2000) The expression and stage-specific localization of protein kinase C isotypes during mouse preimplantation development, *Dev. Biol.* **223**, 411–421.

Posada, J., Cooper, J.A. (1992) Requirements for phosphorylation of MAP kinase during meiosis in Xenopus oocytes, *Science* **255**, 212–215.

Rattner, J.B. (1992) Mapping the mammalian intercellular bridge, *Cell. Motil. Cytoskeleton* **23**, 231–235.

Raz, T., Eliyahu, E., Yesodi, V., Shalgi, R. (1998a) Profile of protein kinase C isozymes and their possible role in mammalian egg activation, *FEBS Lett.* **431**, 415–418.

Raz, T., Ben-Yosef, D., Shalgi, R. (1998b) Segregation of pathways leading to cortical and cell cycle activation in the rat egg, *Biol. Reprod.* **58**, 94–102.

Raz, T., Skutelsky, E., Amihal, D., Hammel, I., Shalgi, R. (1998c) Mechanisms leading to cortical reaction in mammalian eggs, *Mol. Reprod. Dev.* **51**, 295–303.

Reszka, A.A., Seger, D., Diltz, C.D., Krebs, E.G., Fischer, E.H. (1995) Association of mitogen-activated protein kinase with the microtubule cytoskeleton, *Proc. Natl. Acad. Sci.* **92**, 8881–8885.

Reszka, A.A., Bulinski, J.C., Krebs, E.G., Fischer, E.H. (1997) Mitogen-activated protein kinase/extracellular signal-regulated kinase 2 regulates cytoskeletal organization and chemotaxis via catalytic and microtubule specific interactions, *Mol. Biol. Cell.* **8**, 1219–1232.

Schaffer, H.J., Weber, M.I. (1999) Mitogen-activated protein kinases. Specific messages from ubiquitous messengers, *Mol. Cell. Biol.* **19**, 2435–2444.

Schwarz, S.M., Gallicano, G.I., McGaughey, R.W., Capco, D.G. (1995) A role for intermediate filaments in the establishment of the primitive epithelia during mammalian embryogenesis, *Mech. Dev.* **53**, 305–321.

Sobajima, T., Aoki, f., Kohmoto, K. (1993) Activation of miotgen-activated protein kinase during meiotic maturation in mouse oocytes, *J. Reprod. Fertil.* **97**, 389–394.

Sun, Q.Y., Rubinstein, S., Breitbart, H. (1999) MAP Kinase activity is downregulated by phorbol ester during mouse oocyte maturation and egg activation in vivo, *Mol. Reprod. Dev.* **52**, 310–318.

Tachibana, K., Tanaka, D., Isobe, T., Kishimoto, T. (2000) c-Mos forces the mitotic cell cycle to undergo meiosis II to produce haploid gametes, *Proc. Natl. Acad. Sci.* **97**, 14301–14306.

Takai, Y., Kishimoto, A., Inoue, M., Nishizuka, Y. (1997) Studies on a cyclin nucleotide-independent protein kinase and its proenzyme in mammalian tissues, *J. Biol. Chem.* **252**, 7603–7609.

Tapley, P.M., Murray, A.W. (1985) Evidence that treatment of platelets with phorbol esters causes proteolytic activation Ca^{2+} activated, phospholipid-dependent protein kinase, *J. Biochem.* **151**, 419–423.

Tatone, C., Iorio, R., Francione, A., Gioia, I., Colonna, R. (1999) Biochemical and biological effects of KN-93, an inhibitor of calmodulin-dependent protein kinase II, on the initial events of mouse egg activation induced by ethanol, *J. Reprod. Fertil.* **115**, 151–157.

Tombes, R.M., Peppers, L.S. (1995) Sea urchin fertilization stimulates CaM kinase II (multifunctional [type II] Ca^{2+}/CaM kinase). Activity and association with $p34^{cdc2}$, *Dev. Growth Differ.* **37**, 589–596.

Verlhac, M.H., dePennart, H., Maro, B., Cobb, M.H., Clarke, H.J. (1993) MAP Kinase

become stably activated at metaphase and is associated with microtubule-organizing centers during meiotic maturation of mouse oocytes, *Dev. Biol.* **158**, 330–440, 369–374.

Verlhac, M.-H., Kubiak, J.Z., Clarke, H.J., Maro, B. (1994) Microtubule and chromatin behavior follow MAP kinase activity but not MPF activity during meiosis in mouse oocytes, *Development* **120**, 1017–1025.

Verlhac, M.H., Kubiak, J.Z., Weber, M., Geraud, G., Colledge, W.H., Evans, M.J., Maro, B. (1996) Mos is required for MAP Kinase activation and is involved in microtubule organization during meiotic maturation in the mouse, *Development* **122**, 815–822.

Walker, S.D., Murray, N.R., Burns, D.J., Fields, A.P. (1995) Protein kinase C chimeras: catalytic domains of ∀ and $\exists_{II}$ protein kinase C contain determinants for isotype-specific function, *Proc. Natl. Acad. Sci.* **92**, 9156–9160.

Weber, M., Kubiak, J.Z., Arlinghaus, R.B., Pines, J., Maro, B. (1991) c-Mos proto-oncogene product is partly degraded after release from meiotic arrest and persists during interphase in mouse zygotes, *Dev. Biol.* **148**, 393–397.

Webster, S.D., McGaughey, R.W. (1990) The cortical cytoskeleton and its role in sperm penetration of the mammalian egg, *Dev. Biol.* **142**, 61–74.

Wheatley, S.P., Wang, Y.-L. (1996) Midzone microtubules are continuously required for cytokinesis in cultured epithelial cells, *J. Cell Biol.* **134**, 981–989.

Wianny, F., Zernicka-Goetz, M. (2000) Specific interference with gene function by double-stranded RNA in early mouse development, *Nat. Cell Biol.* **2**, 70–75.

Winston, N.J., Maro, B. (1995) Calmodulin-dependent protein kinase II is activated transiently in ethanol-stimulated mouse oocytes, *Dev. Biol.* **170**, 350–352.

Winston, N.J. (1997) Stability of cyclin B protein during meiotic maturation and the first mitotic cell division in mouse oocytes, *Biol. Cell.* **89**, 211–219.

Wright, P.W., Bolling, L.C., Calvert, M.E., Sarmento, O.F., Berkeley, E.V., Shea, M.C., Hao, Z., Jayes, F.C., Bush, L.A., Shetty, J., Shore, A.N., Reddi, P.P., Tung, K.S., Samy, E., Allietta, M.M., Sherman, N.E., Herr, J.C., Coonrod, S.A. (2003) ePAD, an oocyte and early embryo-abundant peptidylarginine deiminase-like protein that localizes to egg cytoplasmic sheets, *Dev. Biol.* **256**, 73–88.

Xu, Z., Abbott, A., Kopf, G.S., Schultz, R.M., Ducibella, T. (1997) Spontaneous activation of ovulated mouse eggs: time-dependent effects on M-phase exit, cortical granule exocytosis, maternal messenger ribonucleic acid recruitment and inositol 1,4,5-triphosphate sensitivity, *Biol. Reprod.* **57**, 743–750.

Zernicka-Goetz, M., Verlhac, M.-H., geraud, G., Kubiak, J.Z. (1997) Protein phosphatases control MAP kinase activation and microtubule organization during rate oocyte maturation, *Eur. J. Cell Biol.* **72**, 30–38.

Zhou, R., Oskarsson, M., Paules, R.S., Schulz, N., Cleveland, D., Woude, G.F.V. (1991) Ability of the c-mos product to associate with and phosphorylate tubulin, *Science* **251**, 671–675.

Chaperones, Molecular

Peter Lund
School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK

1 The Definition of a Molecular Chaperone 489

2 The Need for Molecular Chaperones: Protein Folding *In Vitro* and *In Vivo* 490

3 Classification of Molecular Chaperones 493

4 Mechanisms and Roles of the Major Molecular Chaperones 494
4.1 The Hsp60 Family 494
4.1.1 Introduction to the Hsp60 Family 494
4.1.2 Cellular Roles of the Hsp60 Family 494
4.1.3 Mechanisms of Action of the Hsp60 Family 497
4.1.4 Structure and Function of the Hsp60 Family 499
4.2 The Hsp70 Family 501
4.2.1 Introduction to the Hsp70 Family 501
4.2.2 Cellular Roles of the Hsp70 Family 501
4.2.3 Mechanisms of Action of the Hsp70 Family 502
4.2.4 Structure and Function of the Hsp70/Hsp40 Family 504
4.3 The Hsp90 Family 504
4.3.1 Introduction to the Hsp90 Family 504
4.3.2 Cellular Roles of the Hsp90 Family 504
4.3.3 Mechanisms of Action of the Hsp90 Family 505
4.3.4 Structure and Function of the Hsp90 Family 506
4.4 The Small Heat Shock Proteins 507
4.4.1 Introduction to the Small Heat Shock Proteins 507
4.4.2 Cellular Roles of sHSPs 507
4.4.3 Mechanism of Action of sHSPs 507
4.4.4 Structure and Function of sHSPs 508

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

5 Chaperones and Cellular Processes 508
5.1 Chaperones and Redox Potential 508
5.1.1 Disulfide Bond Formation and Isomerization in Prokaryotes and Eukaryotes 508
5.1.2 Hsp33: A Chaperone with a Redox Switch 510
5.2 Chaperone Networks and Protein Quality Control in the Cell 511
5.2.1 Chaperone Networks and Protein Quality Control in the *E. coli* Cytosol 511
5.2.2 Protein Quality Control in the Endoplasmic Reticulum 513
5.3 Examples of Substrate-specific Chaperones 514
5.3.1 Hsp47 and Collagen Assembly 514
5.3.2 PapD and the Assembly of Bacterial Pili 515

6 Conclusion 516

Acknowledgments 517

Bibliography 517
Books and Reviews 517
Primary Literature 517

Keywords

Chaperonin
One of a member of the family of molecular chaperones that are homologous to the *E. coli* GroEL chaperone, and the mitochondrial and chloroplast Hsp60 chaperones.

Cochaperone
A protein that, while not possessing intrinsic chaperone properties itself, is required for a particular chaperone to function.

Heat Shock
The exposure of cells or organisms to temperatures a few degrees above their normal growth temperature, widely used as an experimental model for inducing cellular stress.

HSP
Abbreviation for ''heat shock protein.'' HSPs are proteins that are induced *de novo* or are upregulated as a result of heat shock.

Molecular Chaperone
Proteins that assist other proteins to reach their final active conformation, but that are not themselves part of this final conformation.

Unfolded Protein
Any protein that is not in its fully folded state. This term encompasses a large range of possible states, from a protein that is nearly but not completely native to one that is a random coil.

Some proteins in the cell are needed to help other proteins reach their final active conformations: these are referred to as molecular chaperones. Molecular chaperones can be grouped into families by sequence similarity, and they act through a variety of mechanisms. They are found in all cell types and in most locations within the cell. Many are induced by stresses such as heat shock. This chapter will review the reasons why molecular chaperones are required in the cell, and then discuss each of the major families of molecular chaperones in turn, in terms of structure, mechanism of action, and cellular role. Most molecular chaperones act in networks with other chaperones, and some of these will be reviewed. Many molecular chaperones interact with a broad range of different substrates, but some are specific to particular proteins, and examples of these will be discussed.

1
The Definition of a Molecular Chaperone

The number and range of proteins now classified as molecular chaperones are very large, and a definition that encompasses all of them is likely to be so vague as to be uninformative. Molecular chaperones must, however, have two properties in order to be defined as such, which are as follows:

1. They must assist other molecules to reach their final active state, by blocking or reversing unfavorable events that would otherwise prevent this state from being reached.
2. They must not be part of the final active structure whose existence they have helped to create.

The name *molecular chaperone* was originally coined to refer to the protein nucleoplasmin, which is required for the correct assembly of nucleosomes, and the choice of the term "chaperone" is based on the conventional use of the word, to refer to a person who prevents inappropriate interactions from occurring that could hinder an otherwise desirable outcome. (The etymology of the word "chaperone" is of interest: it is derived from the French verb "chaperonner," which means "to protect." This, in turn, was originally derived from the Old French noun "chaperon," which meant "a protective hood"). The justification for this choice of name is that for biological molecules such as proteins, getting from the state where they are first synthesized to the state where they are finally active (folded, and in the correct cellular compartment) is a complex process during which many unfavorable interactions can take place. These, if they were allowed to occur, would drastically lower the efficiency with

which active biological material could be formed. Molecular chaperones prevent or reverse such interactions, which explains why they are so important for normal cellular function. Although the original chaperone (nucleoplasmin) is the key to chromatin assembly, the vast majority of chaperones are involved in protein folding and assembly, and it is only these that will be considered in this chapter.

In order for a given protein to be classified as a molecular chaperone, the chaperone activity of the protein should be experimentally demonstrable. Ideally, this demonstration would be made both *in vitro* and *in vivo*, through a combination of biochemical and genetic methods, but in practice this is not always feasible. The kinds of experiments that can demonstrate molecular chaperone activity are discussed below, in the sections on individual classes of molecular chaperone.

2 The Need for Molecular Chaperones: Protein Folding *In Vitro* and *In Vivo*

All life depends upon the ability of proteins to fold into specific conformations. An understanding of the process by which proteins succeed in folding from a linear chain of amino-acyl residues to a functioning tertiary or quaternary structure is thus central to an appreciation of all aspects of cell and organismal biology. The fact that this process occasionally goes wrong, and in doing so can lead to many pathological states, makes the need for such an understanding even more important. Most of the initial studies in this field were done using highly purified proteins, and these gave us many important insights into the essential features of protein folding. Studies of this nature are still continuing. More recently, however, it has been realized that protein folding inside the cell is a more complex process than that studied in the test tube, and many key discoveries have been made by looking at cellular protein folding in more detail. These include the discovery of molecular chaperones, which act in many ways to help proteins overcome some of the intrinsic problems caused by having to fold up in the complex internal environment of the cell. This initial section will thus briefly review our current understanding of protein folding both *in vitro* and *in vivo*, and will discuss in general terms some of the areas where molecular chaperones have been shown to be important or essential.

The earliest work in the area of protein folding was done by Anfinsen, who demonstrated that proteins will spontaneously adopt their native, active conformation in aqueous solution, thus proving that folded proteins are more thermodynamically stable than unfolded proteins. This conclusion was reached by taking purified proteins, denaturing them in suitable solvents that caused them to adopt an unfolded conformation, and then removing the denaturants and observing recovery of activity of the protein in question. No energy source is required in these types of experiments for proteins to refold, thus proving that the folded and active protein must be more stable than the unfolded form.

Although this was an elegant demonstration of the thermodynamically spontaneous nature of protein folding, it was subsequently pointed out by Levinthal that there was an apparent kinetic problem in protein folding. Because the number of possible conformations that can be adopted by a protein is very high indeed, and it takes a finite time for proteins to sample these different conformations,

it would in theory take, on an average, billions of years of searching through all the different conformational options for a given protein molecule to find the most stable conformation. It is known, however, that proteins can fold from the denatured state over a timescale of microseconds to seconds. To resolve this apparent paradox, Levinthal proposed that folding pathways exist, with a limited number of possible conformations existing along the pathways.

This view is still thought to be essentially correct, but the view of a folding pathway has been refined to that of an energy landscape. According to this view, the folding of a protein is analogous to the trajectory taken by a ball that is rolling around on a rubber sheet, which has been deformed by a heavy weight. The ball will eventually finish up at the bottom of the dip caused by the heavy weight; this is equivalent to the protein in its lowest energy conformation. The precise route that the ball takes to get to the bottom of the well depends on a number of factors, such as its initial starting point and the shape of the distorted sheet. Similarly, within a given population of proteins, some molecules may fold very rapidly and others more slowly. The fact that they all finish up in the same final conformation does not imply that they all have to follow precisely the same pathway to get there.

This view of protein folding can be further expanded to consider what events might lead to some proteins failing to reach a fully folded and active conformation. At least two such events can be envisaged, leading respectively to protein misfolding and to protein aggregation. In the case of protein misfolding, the polypeptide chain may reach a metastable state where it is indeed folded but has not reached its correct and active conformation, and cannot proceed to it on a biologically relevant timescale. In the case of protein aggregation, the protein may have formed interactions with other partially folded protein chains such that all these proteins are now held together but are inactive because they have not reached their fully folded states. Interestingly, aggregation appears to be specific in that, when it occurs in mixtures of proteins, similar or identical proteins will tend to aggregate with each other. Aggregation is particularly likely to be a problem where protein concentrations are high, since the likelihood of two proteins interacting with each other before they fold into the native state is greater. Aggregation is potentially a particularly serious problem for proteins, because the forces that drive protein folding (hydrophobic forces) are the same as the forces that lead to protein aggregation. It is also a problem inside the cell, where protein concentrations may be extremely high.

Aggregation and misfolding are not the only problems that may be faced by proteins as they fold within the cell (Fig. 1). Another difficulty for proteins is that many targeting events require proteins to cross membranes into different cellular compartments, and yet it is hard to envisage a more effective barrier to a folded protein than a lipid bilayer. In order to cross membranes, proteins generally need to be prevented from folding but protected from other folding proteins with which they might interact both before and after they have crossed the membrane. Proteins also often need to form interactions with a large array of cofactors and effectors, and may require the presence of these in order to adopt their active conformations. In this event, proteins would need to be synthesized but kept in a receptive state until the cofactor becomes available. Proteins on the secretory pathway in

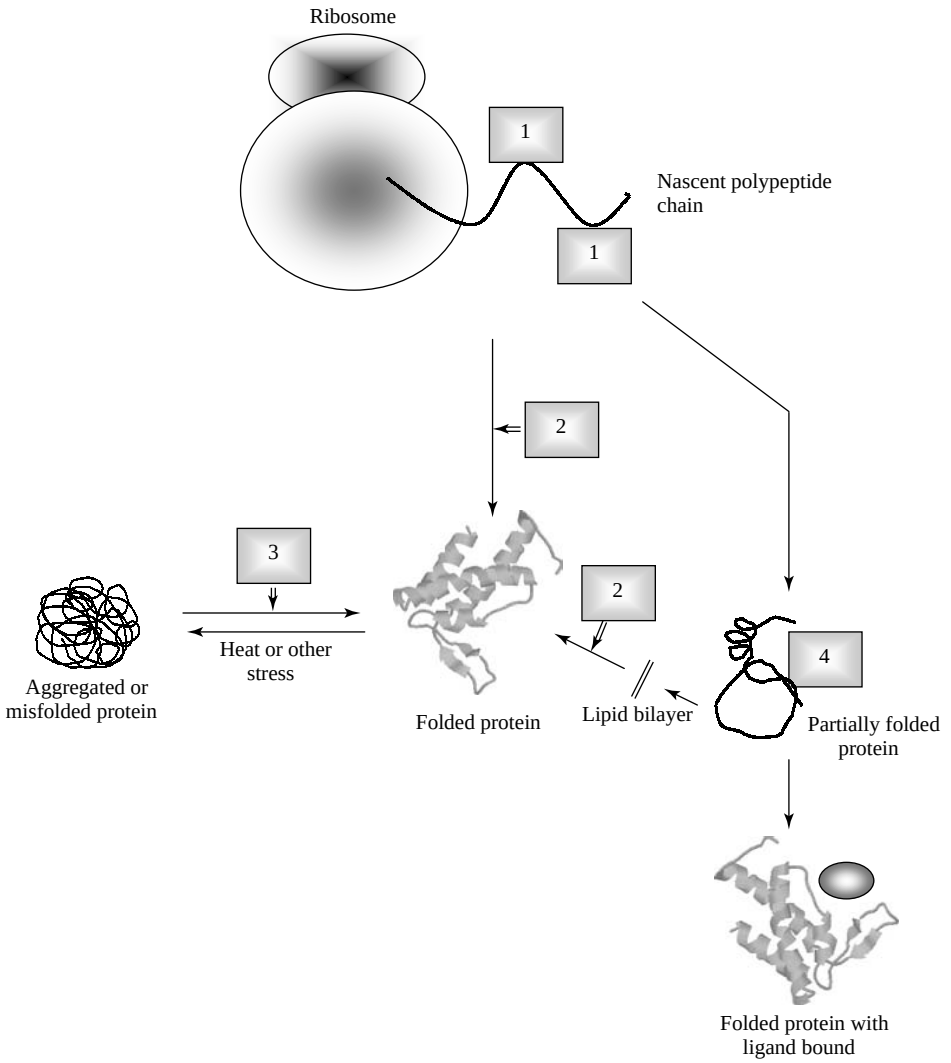

Fig. 1 The general roles of molecular chaperones. Molecular chaperones are involved in many different cellular processes, including protection of nascent proteins on the ribosome (1); folding of newly synthesized or newly membrane-translocated proteins (2); protection and refolding of misfolded, partially unfolded, or aggregated proteins (such as after heat shock (3); and maintenance of proteins in a partially unfolded state (4) for binding effectors or crossing membranes. Other roles are discussed further in the text.

eukaryotic cells have to be modified before leaving the endoplasmic reticulum (ER), and molecular chaperones play an important role here in quality control, ensuring that only those proteins with the correct modifications proceed further down the pathway. Finally, proteins are not particularly stable entities, and hence have a propensity to unfold under certain nonoptimal conditions such as a small

increase in temperature. Unfolding of proteins reintroduces all the potential problems of misfolding and aggregation that exist when the proteins are first synthesized.

Molecular chaperones are proteins that have evolved to deal with all these inherent barriers to the cell's need to have the right proteins in the right place in their active conformations. As later sections will describe, they have been shown to participate in all the examples given above, and in their absence many of these essential cellular functions cannot take place at all. Some chaperones appear to act predominantly by binding to and preventing the aggregation of proteins that have become unfolded; others are able to actively promote folding of proteins that otherwise cannot fold; and yet others can act on aggregated proteins to unfold them in order to allow them another opportunity to refold.

3 Classification of Molecular Chaperones

Molecular chaperones can be described in terms of their cellular locations, their mode of expression, their mode of action, and the proteins or other molecules upon which they act, but they are best classified in terms of sequence similarity. Chaperones that belong to a particular group defined thus turn out, unsurprisingly, to have similar structures and act on similar substrates through related mechanisms. The mechanisms of action and structures of the members of different groups are often quite distinct. The degree of sequence similarity within a group varies between different groups: some molecular chaperones are among the most highly conserved of all proteins, whereas others show very high heterogeneity between members of the same group. It must be remembered that sequence similarity alone is only an indicator of potential molecular chaperone activity, not proof of it, and that direct experimental evidence should always be sought to confirm whether or not a given protein is indeed acting as a chaperone.

The classification of molecular chaperones is complicated by a number of factors, some of which are as follows:

- First, certain proteins have been discovered that do not themselves act directly on protein substrates, but are required for the full chaperone reaction to take place. These are generally referred to as "cochaperones," and as they are central to the chaperoning process they are usually discussed in the same context as molecular chaperones.
- Second, the fact that often proteins from the same chaperone family have been discovered and studied in different organisms or several different cellular locations at roughly the same time has led to a proliferation of names for proteins within a given family, which can be confusing.
- Third, many chaperones are heat-inducible, and as they were often first discovered by virtue of this property, they are referred to as heat shock proteins (HSPs). Unfortunately, this means many chaperone families are labeled as HSPs, even though not all the proteins within that family may be heat shock–inducible. An alternative approach is to use the names of the best studied member to label the group, but this is equally unsatisfactory as the name given to this protein may refer to

the phenotype through which the protein was first discovered, and which may be quite specific to a particular organism.
- Fourth, the name "molecular chaperone" is used in a very catholic sense to refer to any protein with the properties listed in Sect. 1 above, but these may range from extremely broad-spectrum chaperones, which act on a large range of substrates to those that have evolved to act with only one specific substrate.
- Fifth, many proteins are unfortunately described as molecular chaperones with only minimal supporting evidence from *in vitro* and *in vivo* studies.

The Table 1 is not intended to be comprehensive, but illustrates the major known families of molecular chaperones, together with the names of some of their better-studied members, their cellular locations, and a very brief description of their properties. The individual molecular chaperone families shown in the Table are discussed more fully in the following sections.

4 Mechanisms and Roles of the Major Molecular Chaperones

4.1 The Hsp60 Family

4.1.1 Introduction to the Hsp60 Family

The Hsp60 family is one of the best characterized of the molecular chaperone families. It is a large family consisting of highly similar proteins, all with a subunit molecular mass of around 60 kDa. Sequence comparisons have enabled the division of the Hsp60 family into two groups: Group I and Group II. The Group I proteins are found in nearly all bacteria, and also in mitochondria and chloroplasts. The Group II proteins are found in archaea, and in the eukaryotic cytosol. All these proteins have a remarkable multimeric structure: Group I proteins assemble into a "double doughnut" structure with seven subunits in each ring, and the Group II proteins assemble into a similar structure but with eight or nine subunits in each ring, depending on the particular protein and organism. Each ring encloses a cavity or cage. Proteins in the Hsp60 family are often referred to generically as "chaperonins."

4.1.2 Cellular Roles of the Hsp60 Family

The Group I family of the Hsp60 proteins has been the most intensively studied, particularly the protein from *Escherichia coli*, which is referred to as GroEL. Genetic evidence shows that these proteins are essential, both in prokaryotes and eukaryotes, and they are found in all organisms so far studied, with the exception of some mycoplasmas. They are present at fairly high levels in cells (*E. coli* is estimated to contain around 2000 GroEL complexes per cell under normal growth conditions) and are further induced by heat shock. All of them appear to require a cochaperone, called Hsp10, which is a smaller protein of typically around 10 kDa, which itself assembles into a ring with sevenfold symmetry. In *E. coli*, this protein is called GroES, and in *E. coli* and most other bacteria it is encoded in the same operon as the GroEL protein and is also essential.

If the levels of GroEL become too low to sustain cellular growth, or if a temperature-sensitive mutant is shifted to the nonpermissive temperature, a large subset of cellular proteins fails to fold

Tab. 1 Selected examples of the major chaperone families in the cell.

Name of family	*Prokaryotic examples*	*Eukaryotic examples*	*Cellular location*	*Known cochaperones*	*Cellular functions*
Hsp60	GroEL		Cytosol	GroES	Folding of a subset of newly synthesized proteins, folding of a subset of proteins damaged by heat shock and other stresses
	CCT (archaea)		Cytosol	GimC	Unknown; induced by heat shock
		Hsp60	Mitochondria	Hsp10	Import of proteins synthesized in cytosol, folding of proteins, protection against heat shock
		RuBP (rubisco subunit binding protein)	Chloroplasts	Cpn21	Import and folding of proteins synthesized in cytosol, and folding of proteins synthesized within chloroplasts, particularly the large subunit of rubisco
		CCT	Cytosol	GimC	Folding of actin and tubulin, folding of other cytosolic proteins
Hsp70 with Hsp40	DnaK/DnaJ		Cytosol	GrpE	Binding to nascent polypeptides, protection of protein from aggregation at high temperatures, dissolution of aggregates (with ClpB), regulation of heat shock response in *E. coli*
		Hsp70/Hsp40	Cytosol	Fes1, Bag1	Multiple proteins and multiple roles including binding of nascent proteins, protein import into organelles, uncoating of clathrin cages
		BiP/Sec63	ER	SLS1, BAP	Binding of incompletely folded secretory proteins, facilitating retrotranslocation to cytosol for proteasome degradation; note: several different proteins with J-domains interact with BiP
		mtHsp70/Mdj1	Mitochondria	Mt-GrpE	Import and folding of proteins, heat shock protection

(*continued overleaf*)

Tab. 1 (*continued*)

Name of family	*Prokaryotic examples*	*Eukaryotic examples*	*Cellular location*	*Known cochaperones*	*Cellular functions*
Hsp90	HtpG		Cytosol		Unknown, heat shock induced, but dispensable
		Hsp90	Cytosol	Hop, FKBP51, p23	Binding to proteins that require activation with ligand such as steroid hormone, protection against heat shock? Also buffers against genetic variation.
		Grp94	ER		Unknown, probably a role in quality control of protein transport
		TRAP1	Mitochondria		Unknown
Hsp100	ClpB		Cytosol		Works with DnaK to dissipate protein aggregates
		Hsp104	Cytosol		Works with Hsp70 to dissipate protein aggregates
sHSPs	IbpA, IbpB		Cytosol		Probably act to hold denatured proteins for other chaperones to refold
		Hsp25	Cytosol		Probably act to hold denatured proteins for other chaperones to refold
		α-B-crystallin	Eye lens		Keep eye lens proteins in solution
Redox chaperones	DsbA/DsbC		Periplasm		Respectively introduce and isomerize disulfide bonds; require DsbB and DsbD respectively to regenerate active forms
	Hsp33		Cytosol		Activated by oxidation, may protect proteins damaged by oxidation in cytosol
		Protein disulfide isomerase	ER		Introduce and isomerize disulfide bonds; requires Ero1 to regenerate active form
ER chaperones		Calnexin, calreticulin	ER		Act to prevent incorrectly glycosylated proteins from entering the secretory pathway

correctly. Coimmunoprecipitation experiments with antibodies against GroEL have also shown that a large number of different proteins (about 10–15% of all proteins in *E. coli*) bind to GroEL shortly after their synthesis is complete. It is thus thought that GroEL, together with GroES, acts to promote the correct folding of a subset of the proteins present in the bacterial cytosol that would otherwise not fold correctly. After heat shock, an increased association of proteins with GroEL is seen, and this implies that GroEL also acts to help refold proteins, which become wholly or partly denatured by an increase in temperature. The Hsp60 proteins present in mitochondria and chloroplasts are also involved in promoting the folding of some of the proteins that have been newly imported into the organelle after synthesis in the cytosol as well as some of those that are synthesized within the organelle. It was the recognition (in the late 1980s) of the unexpectedly high sequence homology between the chloroplast Hsp60 protein (the major function of which is the folding of the large subunit of rubisco, encoded by the chloroplast genome) and the *E. coli* GroEL protein (identified originally as a protein required for the folding of a component of bacteriophage lambda), which helped to galvanize research into the role and mechanism of action of these proteins.

The role of the Group II proteins is less clear. In archaea, the proteins are heat shock inducible, but not in eukaryotes. Archaea encode one, two, or three proteins in the Group II family, and these assemble into large complexes of two rings with eight- or ninefold symmetry. Eukaryotes encode eight proteins in the Group II family, which assemble into a double-ring complex with eight-fold symmetry. All eight genes are essential in yeast. The protein complex has been referred to variously as CCT, TRiC, and TCP-1. The principal role of this complex in eukaryotes appears to be the folding of the cytoskeletal proteins actin and tubulin, and genetic evidence shows that when the function of the CCT complex is compromised with conditional mutants, cells show disorganized actin, defects in morphogenesis and cell division, and acute sensitivity to microtubule-depolymerizing compounds. Although many other proteins are also substrates for the complex, the complete range of cellular substrates has yet to be defined. Moreover, actin and tubulin are not found in archaea, and the substrates for the archaeal Hsp60 proteins are completely unknown. The Group II Hsp60 proteins do not require a cofactor homologous to the Hsp10 protein used by the Group I proteins, but substrates are delivered to them by a totally unrelated chaperone called GimC or prefoldin, which again is found in both archaea and eukaryotes but not in bacteria.

4.1.3 Mechanisms of Action of the Hsp60 Family

The mechanism of action of the Group I proteins, as typified by the *E. coli* GroEL protein, has been very intensively studied, and yet there remain considerable uncertainties about exactly how the protein works. *In vitro* experiments show that if certain proteins are denatured (by chemical or heat treatment) and are then allowed to refold, the yield of refolded protein is very low, but can be increased to nearly 100% by incubation with GroEL, GroES, and ATP (Fig. 2). This process requires a complex reaction cycle, which is discussed more fully in the next section. During the reaction cycle, the unfolded protein is transiently bound in the cavity formed on one ring by the capping of the ring by the GroES protein. There are

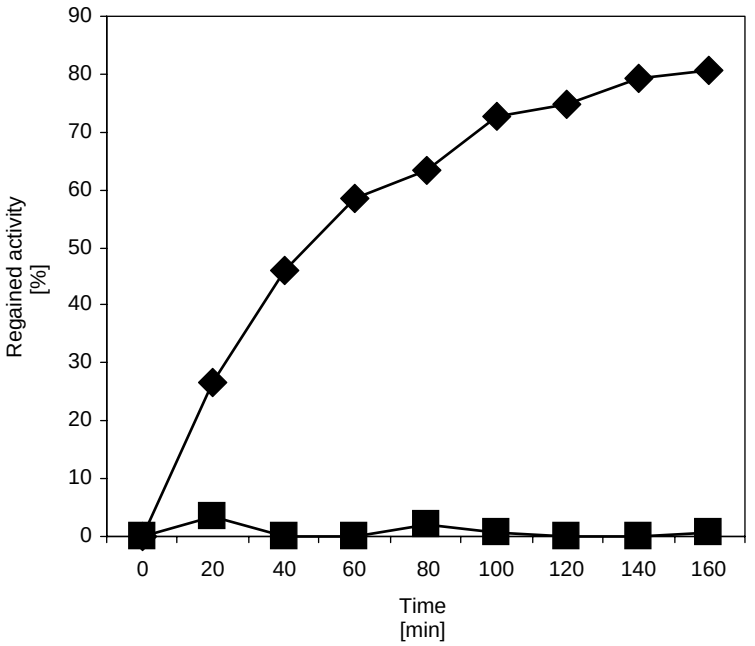

Fig. 2 Refolding of a denatured protein *in vitro* by GroEL, GroES, and ATP. In the example shown here, a substrate protein has been denatured in buffer using heat, and then returned to the temperature at which it is normally active. At suitable time points, the protein is assayed for regain of enzyme activity. The square symbols show the effect in buffer alone, and demonstrate that this particular substrate cannot effectively renature under the conditions shown. The diamond symbols show the effect of having GroEL, GroES, and ATP present in the buffer, with the GroEL and GroES at the same molar concentration as the substrate protein. As can be seen, this addition leads to restoration of roughly 80% of the initial activity of the protein.

at least three different models to explain how this process helps the protein to fold, and these are not mutually exclusive. The first (referred to sometimes as the "Anfinsen cage" model) proposes that the enclosed cavity simply provides a passive and protected environment where the protein can fold without interacting with other folding proteins with which it might otherwise aggregate. A refinement of this model (the "assisted folding" model) suggests that the environment of the cage is important, as the interior surface of the cage is rich in hydrophilic residues that tend to favor the burial of hydrophobic residues in the encapsulated substrate protein, thus accelerating folding. A third model (the "iterative annealing" model) suggests a very different role for GroEL. In this model, the principal role of GroEL is to *unfold* proteins that have become trapped in nonproductive misfolded states, and to redeliver them to the start of the protein folding pathway. It is quite possible that elements of all three models are correct, and that the extent to which any one applies varies with different substrate proteins. Further refinements to these

models may be required following recent demonstrations that GroEL can help fold some proteins that are too large to fit into the cavity of a single ring.

The mechanism of action of the Group II proteins is much less well understood. Actin and tubulin have both been shown to interact with specific subunits of the CCT complex (which are always present in the same order in the complex), and there is evidence that the substrates remain bound throughout the folding cycle, unlike the case in GroEL in which they are discharged into the cavity. It may be that the large domain movements of the complex that take place during the folding cycle act to force the bound unfolded protein into a more compact folded form.

4.1.4 Structure and Function of the Hsp60 Family

It is the remarkable structure of the Hsp60 proteins that has been in part responsible for the degree of interest that they have raised. Members of both the Group I and Group II families have been studied by different structural techniques, notably X-ray crystallography and cryoelectron microscopy, and more recently in the case of GroEL by NMR. Hsp60 proteins assemble into double-ring structures with several subunits in each ring; in the case of CCT from eukaryotes, the relative position of each of these subunits is unique, so that each always has the same neighbors. The individual subunits of proteins from both families consist of three distinct domains: an equatorial domain, which is responsible for the contacts between the two rings and for ATP-binding; a flexible intermediate domain; and an apical domain, which contains the region in which protein binding initially occurs. Moreover, both structures undergo large conformational changes as they pass through the different stages of the protein folding cycle. Again, it is the GroEL protein that is best understood.

The sequence of events that is thought to take place during the folding cycle of the GroEL protein is depicted in Fig. 3. Partially folded protein binds to one ring via a series of hydrophobic contacts with residues at the top of the apical domain. GroES binds to some of the same residues, and when it binds (which requires the presence of bound nucleotide) it displaces the bound peptide into the center of the cavity and simultaneously causes a very large conformational change in the ring to which it binds, the effect of which is to nearly double the size of the cavity. This structure slowly turns over the ATP, and until this reaction is complete, no protein or nucleotide can bind to the opposite ring. Once the ATP has been turned over to ADP, unfolded protein can bind to the opposite ring, followed by ATP and GroES, the net effect of which is to displace the bound substrate, GroES, and ADP, from the opposite ring to which binding first occurred. Thus, both rings are active in binding unfolded proteins, but do not do so at the same time as each other.

The models for action of GroEL discussed above can now be seen in the light of this mechanistic cycle. The Anfinsen cage and assisted folding models predict that the key event in the cycle is the encapsidation of the unfolded protein, which may lead to either the passive or active involvement of the cavity in promoting protein folding. The iterated annealing model sees the main event as being the large conformational change that takes place in the apical domains of GroEL as they move to bind GroES. It is proposed that these act to pull apart misfolded protein, before releasing it again at the start of the protein folding pathway.

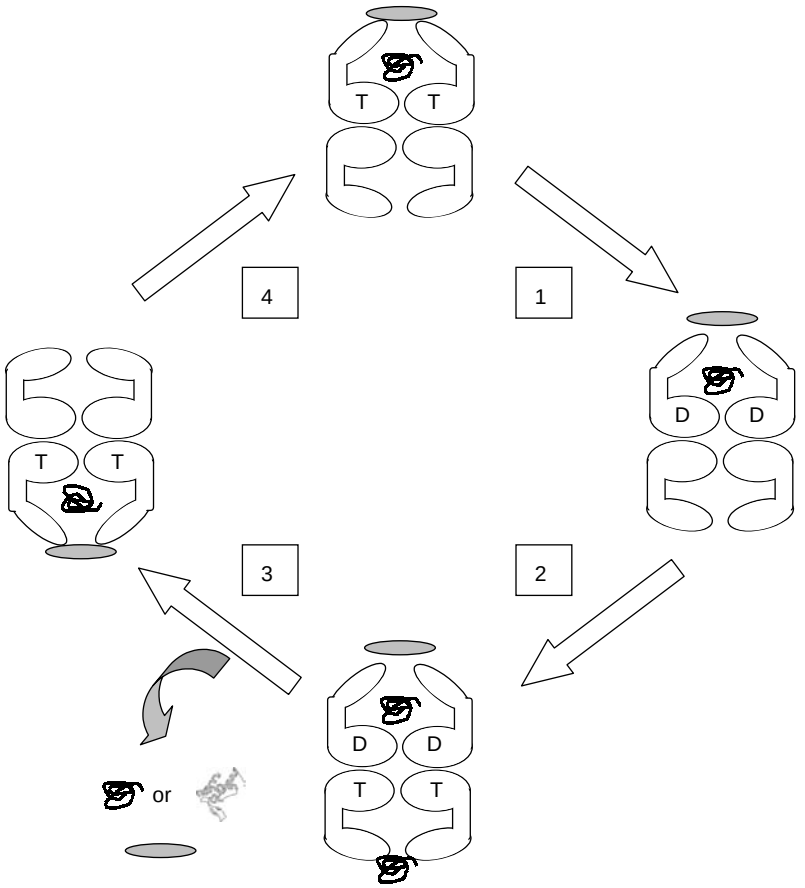

Fig. 3 The proposed reaction cycle of GroEL. GroEL is shown in cross-section, and GroES is shown as a filled ellipse. The complex at the apex of the cycle consists of substrate protein encapsidated in the top (cis) ring of the GroEL complex, which has the ATP (shown as T) bound, and is capped by GroES. The first reaction is the slow hydrolysis of ATP to ADP (reaction 1), which has the effect of weakening the GroES-GroEL interaction. During the time that this reaction occurs, folding of the protein may take place within the cavity. In reaction 2, ATP and another molecule of substrate protein bind to the bottom (trans) ring. This binding results in the release of GroES and ADP from the cis ring (reaction 3), and also the release of the bound substrate, which may now be either folded or still in a partially folded state. The new complex is now essentially the same as the starting one, as can be seen by flipping the complex (step 4) except that the former cis ring is the new trans ring, and vice versa.

The Group II proteins lack a functional GroES cochaperone, but in its place they possess a large helical protrusion at the top of the apical domain that forms a lid to the complex in a way similar to that formed by GroES. It is clear that large domain movements occur in the ATP-binding and hydrolysis cycle with the Group II proteins, but their relationship to the ability of this protein complex to promote the folding of proteins is not yet understood.

4.2 The Hsp70 Family

4.2.1 Introduction to the Hsp70 Family

The Hsp70 proteins comprise a large family, often with many representatives in a given organism (the yeast *Saccharomyces cerevisiae* has 14 Hsp70 homologs). They are found in all eubacteria, all eukaryotes, and most archaea. The Hsp70 proteins are all approximately 70 kDa in size, and show a high degree of sequence similarity. They act in concert with another family of chaperones, the Hsp40 proteins. These are more heterogeneous in size, and include both soluble and integral membrane proteins. All of the Hsp40 family members possess a particular domain (referred to as a J domain, named after DnaJ which is one of the *E. coli* homologs of this protein), which is key to their activity. In eukaryotic cells, both proteins are found in many different cellular environments, including the cytosol, nucleus, mitochondria, chloroplasts, and the lumen of the endoplasmic reticulum. They have a multiplicity of roles, all of which are related to their ability to bind to short stretches of hydrophobic residues in unfolded proteins. As with the Hsp60 proteins, their reaction cycle is driven by the binding and hydrolysis of ATP, but their mode of action appears to be very different. They utilize a wide variety of cochaperones and other factors, and are integrated into a network of interactions with other chaperones in the cell, as described below. Although originally described as heat shock proteins, some members of the family are synthesized constitutively and show no heat shock induction.

4.2.2 Cellular Roles of the Hsp70 Family

The Hsp70 proteins, together with the Hsp40 proteins, fulfill many functions inside the cell, as expected from the large number of homologs that can exist for these proteins in even a single organism. It is clear that not all of these functions have yet been defined. Some of the cytosolic members of the family are involved in binding nascent polypeptides as they emerge from the ribosome, probably thereby preventing them from folding prematurely before the whole chain has been synthesized. Other roles in the cytosol include the uncoating of clathrin from clathrin-coated vesicles, and the delivery of proteins to the mitochondrial and chloroplast import complexes. In the endoplasmic reticulum, the Hsp70 homolog Bip has a key role in ensuring that only correctly folded proteins are delivered to the secretory pathway, and that incorrectly folded proteins are returned to the cytosol for destruction by the proteasome. In the mitochondria and chloroplasts, the organellar homologs of the Hsp70 and Hsp40 proteins have a key role in the uptake of proteins that are synthesized in the cytosol and targeted to organelles, as well as in assisting the folding of proteins encoded by the organellar genomes.

That these roles vary in their importance can be seen from genetic studies. The main Hsp70 homolog in *E. coli*, DnaK, can be deleted, and although the resultant cells show reduced viability and no ability to grow at high or low temperatures, they are capable of growth at normal temperatures. The loss of viability is in part due to the fact that DnaK has a role in regulation of the heat shock response, and if the same deletion is made in *Bacillus subtilis*, an organism where DnaK does not have this role, the effect on viability is reduced. Two interpretations of these data are possible: one is that DnaK fulfills roles of relatively minor importance; the

other is that the roles of DnaK are so important that other proteins have evolved to back them up should DnaK become nonfunctional. At least one of the roles for DnaK, namely, that of binding proteins as they emerge from the ribosome, has been shown to fall into the latter category. Another chaperone, called trigger factor, also has this role, and although loss of either trigger factor or DnaK alone is not lethal, a double mutant is not viable under normal growth conditions (this is referred to as *synthetic lethality*). In yeast, some of the Hsp70 homologs are essential and others are not, and again synthetic lethality is seen for some combinations of mutants in different *hsp70* genes, proving that some of the cellular functions carried out by Hsp70 are essential. Trigger factor does not appear to have a homolog in eukaryotic cells, but another protein called NAC (nascent-polypeptide association complex) is thought to fulfill a similar role.

Another important role that has been demonstrated in *E. coli* and *S. cerevisiae* is not only the prevention of protein aggregation at heat shock temperatures but also the breaking down of any aggregates that may occur. For this activity, another chaperone is required: this is the protein referred to as ClpB in *E. coli* and Hsp104 in *S. cerevisiae*. The structure and precise mechanism of action of this chaperone is not known, but it clearly interacts specifically with Hsp70 in mediating protein disaggregation and subsequent refolding. This interaction is specific: yeast Hsp104 cannot substitute for *E. coli* ClpB. The role of ClpB will be considered further below (Sect. 5.2.1).

Hsp40 proteins share many properties with Hsp70 proteins. In all cases in which an Hsp70 protein is known to act, an Hsp40 partner protein is also involved.

4.2.3 Mechanisms of Action of the Hsp70 Family

Both Hsp70 and Hsp40 act by binding to short, exposed regions of predominantly hydrophobic stretches in unfolded proteins. Proteins bound to Hsp40 or Hsp70 do not appear to undergo any further folding, unlike the situation with the Hsp60 proteins, where the protein released from the complex may be in a very different conformation to the one that was first bound by it. Rather, it appears that the transient binding of these regions is enough to reduce the risk of their aggregation to other hydrophobic patches on proteins in the neighborhood. The importance of this is particularly obvious in those cases in which proteins have to be prevented from folding until the entire protein sequence is available to fold, namely, proteins being synthesized on ribosomes, and proteins being transported across membranes. Similarly, in the case of protein folding in the ER, proteins that still have regions exposed and are hence not ready for transport to the Golgi will bind to the Hsp70 homolog BiP. This protein has an ER retention signal at its C-terminus, and consequently is constantly cycled back to the ER from the early steps of the secretory pathway, together with any unfolded proteins that are bound to it.

Although the nature and role of binding of unfolded protein to the Hsp70 and Hsp40 proteins are different from those pertaining to the Hsp60 proteins, the reaction cycle has some similarities. In particular, as with the Hsp60 proteins, the reaction cycle relies upon the chaperones existing in two states, one with a high affinity for unfolded protein and the other with a low affinity, with the binding of nucleotide mediating the transition between those states. The reaction cycle, shown in Fig. 4 for the case of the *E.*

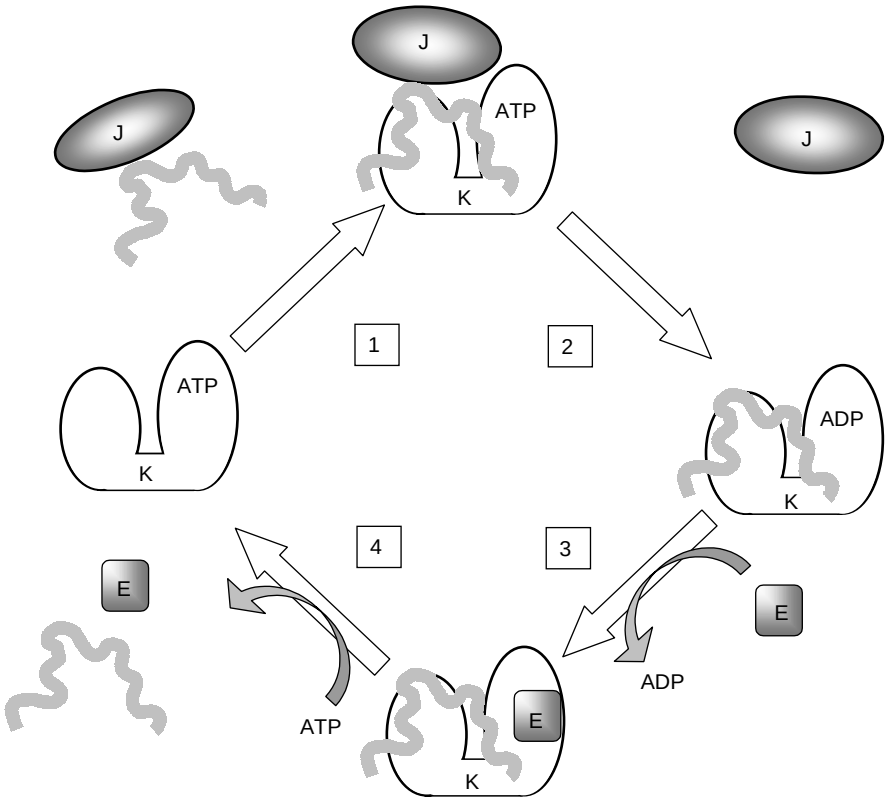

Fig. 4 The proposed DnaK/DnaJ/GrpE reaction cycle. Substrate protein is initially bound to DnaJ (shown as J), which delivers it to DnaK-ATP (step 1). The substrate is transferred to DnaK, and the presence of DnaJ stimulates the ATPase of DnaK to form a stable DnaK-ADP-substrate complex; DnaJ is probably lost at this point (step 2). This complex is stable until the cochaperone GrpE catalyzes the exchange of ADP for ATP, shown here as a two-step reaction where GrpE initially displaces bound ADP (step 3) and is in turn displaced by ATP (step 4). This causes DnaK to release the bound peptide, and the cycle can begin again. It is important to note that several molecules of DnaJ or DnaK may bind one protein. The reaction cycle is best characterized for the *E. coli* proteins shown; in eukaryotes, many different cofactors are involved with cycles involving Hsp70 homologs.

coli proteins DnaK and DnaJ, is thought to begin with the binding of unfolded protein to the Hsp40 partner, from where it is transferred to Hsp70, which is initially in a low-affinity binding state (with ATP bound). The formation of the Hsp70-Hsp40-unfolded protein complex stimulates the Hsp70 ATPase, and the ATP is hydrolyzed to ADP, which in turn has the effect of increasing the affinity of Hsp70 for the unfolded protein, but weakening it for Hsp40, which then dissociates. The resultant Hsp70-ADP-unfolded protein complex is now quite stable. In order for the unfolded protein to be discharged from its bound state, the ADP must be exchanged for ATP to lower the Hsp70 affinity for the unfolded protein. This exchange is a slow reaction but it is speeded up considerably by a cochaperone,

termed GrpE in *E. coli*, which catalyzes the exchange reaction and allows the cycle to be completed with the release of bound protein and the regeneration of the Hsp70-ATP form. GrpE homologs are found in archaea, mitochondria, and chloroplasts, while other proteins that catalyze the same kind of nucleotide exchange, such as Bag-1, are found in eukaryotes.

4.2.4 Structure and Function of the Hsp70/Hsp40 Family

Hsp70 is a two-domain protein, with one domain containing the ATP binding site and the other domain containing the peptide binding site. To date, no structure for the whole protein is available, but structures of the individual domains are known. The peptide binding domain has been crystallized with a bound peptide in place, and this structure shows that the peptide binds to a hydrophobic groove in the chaperone. The bound peptide is held in place by an α-helix, the position of which is thought to change depending on the nucleotide-binding state of the other domain, thus explaining how Hsp70 can change between a low-affinity and a high-affinity state depending on the presence of ATP. Hsp40 proteins can also bind peptides with high affinity, and structural studies on the conserved J domain shared by all these proteins show that again this binding is to a hydrophobic groove. The features that determine affinity of peptides for the grooves on these two proteins are essentially the same – consistent with the model that Hsp40 binds peptides (or, in the cell, stretches of unfolded protein with the appropriate amino acids present) first and then transfers them to Hsp70. The complete structural details of the complex formed by these two proteins are not known, however, so the fine details of this mechanism are not yet understood.

4.3 The Hsp90 Family

4.3.1 Introduction to the Hsp90 Family

The two chaperone families we have looked at so far have benefited from extensive study using simple prokaryotic systems, and many important features of the chaperones are conserved in both prokaryotic and eukaryotic organisms. The situation is somewhat different with the Hsp90 family, a family of well-conserved proteins of approximately 90 kDa, in that studies on prokaryotes show that loss of the *hsp90* gene does not in general produce a strong phenotype, whereas in eukaryotes the gene is essential. In eukaryotes, there are several genes encoding members of this family: the minimum number in complex organisms seems to be two very closely related forms in the cytosol, one in mitochondria, and one in the ER, although in *S. cerevisiae* the ER homolog is not present. Hsp90 proteins are very abundant, and are further induced by heat shock, and they have a rather broad range of cellular functions. Many of these are related to the ability of Hsp90 to hold proteins in a state that is not completely folded or active but which is competent to become so on the reception of a suitable signal.

4.3.2 Cellular Roles of the Hsp90 Family

Many substrates for eukaryotic Hsp90 proteins have been identified, and of these a large proportion are involved in some form of signaling. The best characterized of these are the steroid hormone receptors (SHRs), which are proteins that, when bound to their substrate (steroid hormones), enter the nucleus and stimulate the transcription of specific genes. Hsp90 holds these proteins in an inactive form until the hormone is available, whereupon

it binds to the SHR, which then dissociates from Hsp90, dimerizes, and enters the nucleus. Other proteins, such as the eIF-2α kinase, are bound by Hsp90 and released to become fully folded only after phosphorylation. The fact that Hsp90 is also a heat shock protein suggests that it may also have an important role in protecting cells, and it has been shown *in vitro* that it can bind to and prevent the aggregation of proteins under thermal stress.

The high levels and essential nature of cytosolic Hsp90 point to some other specialized function, and a clue as to what this might be has come from recent remarkable experiments that have shown that Hsp90 may be able to buffer potential phenotypic variation that is present, but not expressed, in all cells. When levels of Hsp90 activity are intentionally lowered, either genetically or by various external treatments, it is found in both plant and animal models that the degree of phenotypic variability seen is increased. Different alterations in phenotype are seen even between organisms that are genetically identical. Thus, it appears that Hsp90 may act to reduce "noise" that is present when organisms develop due to chance fluctuations in environmental conditions or small genetic differences, and this noise can be revealed when Hsp90 activity is compromised. It has even been speculated that this noise suppression may be a mechanism that has evolved to increase the amount of phenotypic variation present and expressed in a population at times of stress, where rapid evolution may be important for survival.

4.3.3 Mechanisms of Action of the Hsp90 Family

The way in which Hsp90 proteins bind to their substrates, and subsequently release them under the control of a suitable signal, is not well understood. What is clear is that the reaction mechanism in the cell is very distinct from those considered above, in that many more cellular cochaperones are involved. The folding cycle for SHRs has been particularly closely studied, and appears to begin with the partially folded SHR binding to Hsp70. Hsp70 can itself associate with a number of different cofactors *in vivo*, and one of these, referred to as Hop or Sti1 depending on the organism, mediates the binding of Hsp70 to Hsp90. The SHR is then transferred to Hsp90 and the Hsp70 and Hop components leave the complex. The SHR is at this stage still not competent to bind steroid hormone. Before it can become so, at least two other components associate with the Hsp90: a protein called p23, and one of a number of proteins that are peptidyl–prolyl cis–trans isomerases (PPIases). Both of these proteins have chaperone activities of their own, and the combined effect of all these proteins coming together is to alter the conformation of the bound SHR such that it can now bind steroid hormone. These reactions are shown in outline in Fig. 5.

A good deal of attention has focused on these latter stages in the binding cycle, as there was for some time a controversy about the role of ATP in the action of Hsp90. Hsp90 tested *in vitro* does not appear to require either binding or hydrolysis of ATP in order to act as a chaperone, but nevertheless contains an ATP binding site, deletion of which is lethal *in vivo*. It now appears that Hsp90 cannot bind to the PPIases or p23 until ATP is bound. As Hsp90 has a potential role in the cell cycle, promoting the folding of some of the kinases needed for cell cycle control, its inhibition in rapidly proliferating cells has been seen as a potential anticancer strategy. Some drugs

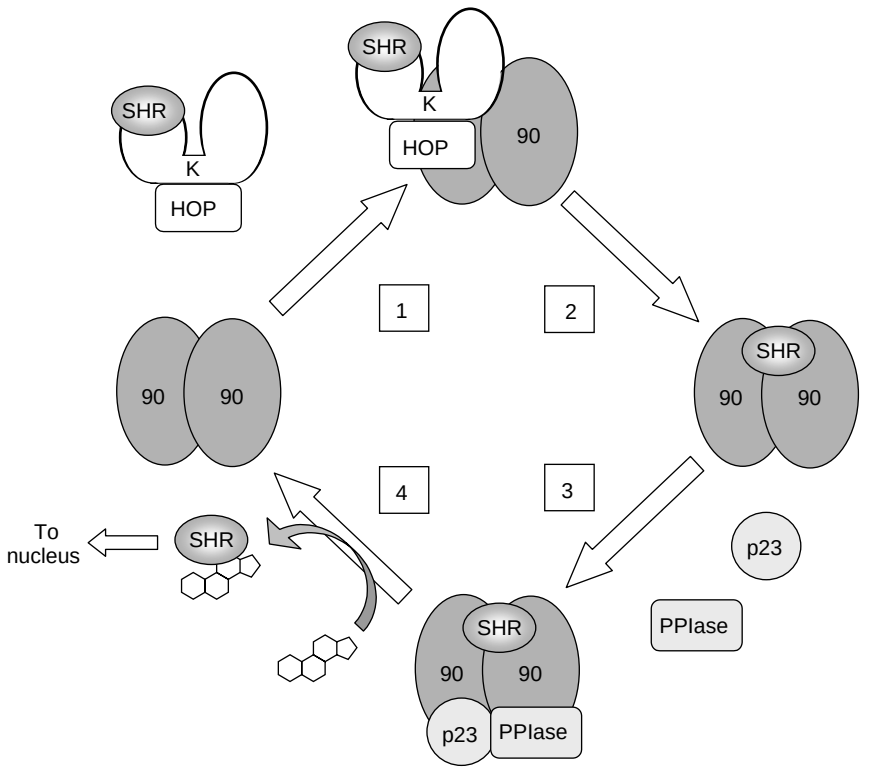

Fig. 5 Outline of the reaction cycle for Hsp90 binding to steroid hormone receptor (SHR). The SHR is delivered to the Hsp90 dimer by Hsp70 in a complex with Hop protein (step 1). SHR is transferred to Hsp90 (step 2), and p23 and PPIase associate (step 3). The precise order of reactions and the stoichiometry of the different proteins are uncertain. SHR can now bind to steroid hormone, which causes it to dissociate from Hsp90 and enter the nucleus to activate gene transcription (step 4).

that have been shown to inhibit Hsp90, in particular, the compound geldanamycin, appear to act as competitive inhibitors of ATP-binding, blocking the access of ATP and hence blocking the later stages in the Hsp90 reaction cycle.

4.3.4 Structure and Function of the Hsp90 Family

The structure of the complete Hsp90 protein is not known, but the structure of the N-terminal domain, which is the main site for binding of ATP, has been solved. As with other chaperones there is evidence of large conformational changes occurring on ATP-binding, but the structural details of these are unknown. Hsp90 is active as a dimer, with the dimerization domain in the C-terminal part of the protein, while the N-terminal regions may have a clamp activity, perhaps with the binding of a nucleotide causing the opening and closing of the clamp. Remarkably, it appears that there are likely to be peptide binding sites and ATP binding sites in both the N- and the C-terminal regions of the protein, but the functional significance of this is not understood.

4.4 The Small Heat Shock Proteins

4.4.1 Introduction to the Small Heat Shock Proteins

The small heat shock proteins (sHSPs) are a much more diverse group of chaperones than those considered above, and are correspondingly less characterized. They range in size from 15 kDa to greater than 40 kDa, and the degree of sequence similarity between them is lower than that for other chaperone families. All possess a conserved domain of approximately 90 amino-acyl residues near the C-terminal end, and all assemble into large oligomers, although again the sizes of these can vary considerably, even for a single sHSP. Organisms often contain numerous genes for sHSPs (plants are particularly abundant in them) and in eukaryotes they have been found in the cytosol, the ER lumen, and in mitochondria and chloroplasts.

4.4.2 Cellular Roles of sHSPs

The precise cellular role of the sHSPs has proven hard to pin down. Strains of *E. coli* or *S. cerevisiae* with all their sHSP genes deleted show very little significant phenotype unless the deletions are combined with other mutations, although in some organisms there is a reduction in thermotolerance when sHSPs are deleted. This lack of phenotype is a surprise, given that the sHSPs are often among the most strongly induced genes by heat shock. Overexpression of sHSPs has been shown to enhance thermotolerance in some cases, which does support the hypothesis that sHSPs have a role to play in survival of temperature stress. They are often found associated with protein aggregates, which have arisen either from the overexpression of proteins that failed to fold correctly, through heat stress or through disease states. There is also evidence that they may play a role in aspects of cytoskeletal and intermediate filament assembly.

A particularly interesting example of the sHSPs is the α-crystallin group of proteins, which are very abundant in the eye lens. Given that the eye lens is effectively a concentrated protein solution, which has to be kept clear for the lifetime of the organism (as cells laid down in fetal development remain in the center of the lens), the avoidance of protein aggregation is of particular importance. Loss of acuity of vision with age, and cataract formation, in particular, are associated with a breakdown in the ability of the α-crystallins to act effectively as chaperones. But defects in α-crystallin function have other consequences as well, showing that its importance is not limited to the eye lens. For example, the inherited disease desmin-related myopathy, which leads to muscle weakness, has been found in some cases to be associated with mutations in one of the α-crystallins, and probably results from defects in intermediate filament assembly.

4.4.3 Mechanism of Action of sHSPs

Despite the lack of a clear *in vivo* role, sHSPs have been demonstrated to have chaperone activity *in vitro*, in that they are able to bind and stabilize unfolded proteins and prevent their aggregation. The range of proteins that sHSPs can bind is very large, with no apparent limits on the size of the protein. There is no clear evidence that sHSPs are able to assist in folding *per se*, and the current view is that they act predominantly as reservoirs of unfolded protein, holding it in a nonaggregated state until conditions more favorable for folding are restored. Refolding of the bound proteins may occur either after the proteins are released into free solution, or

transferred to another chaperone system, such as the Hsp70/Hsp40 system. This is discussed further below, in Sect. 5.2.1. ATP is not required for any of this to occur, further supporting the idea that sHSPs are "holders" rather than "folders." The way in which protein is discharged is not clear, but it may be related to changes in oligomerization and hydrophobicity that occur at different temperatures.

4.4.4 Structure and Function of sHSPs

The sHSPs form oligomers of variable size with fewer than 10 to more than 50 subunits. Moreover, when complexed with bound substrate proteins, their structure and degree of oligomerization may be significantly different from when they have no substrate bound. Several structures have been obtained of free sHSP, and all show oligomers formed roughly into a hollow sphere. The binding sites for unfolded proteins on the sHSPs have not been determined, and there is no detailed structural information for an sHSP-substrate complex.

5 Chaperones and Cellular Processes

A complete understanding of the role of chaperones in the cell is still some way off, but recent work has improved our understanding not only of the individual families of chaperones but also of how they interact with one another and in what way their activities are required for the survival and growth of healthy cells, both under normal and under stressed conditions. The following sections will examine some examples of chaperones and chaperone networks, which are employed for various different processes in the cell. This is not by any means a comprehensive list, but should give at least a flavor of the ways in which chaperones are involved in many different aspects of cell growth.

5.1 Chaperones and Redox Potential

Chaperones have a key role in enabling the cell to maintain active proteins under a variety of redox states, in at least two ways. First, they are directly involved in ensuring that proteins that require disulfide bonds for their activity form those bonds quickly and efficiently. Proteins that catalyze the reactions that make or break disulfide bonds in proteins are very important in maximizing the yield of active folded protein in cells, and many also can act as chaperones on substrates that lack cysteine residues. Second, chaperones exist that are activated by anomalous redox states to help chaperone the folding of those proteins that may themselves have become nonfunctional due to exposure to these states.

5.1.1 Disulfide Bond Formation and Isomerization in Prokaryotes and Eukaryotes

The formation of disulfide bonds in prokaryotes takes place in normal circumstances in the periplasm, the space between the inner and outer membranes in Gram-negative bacteria. (Gram-positive bacteria that have only a single membrane produce very few disulfide-bonded proteins, although interestingly they do express proteins capable of catalyzing disulfide bond formation). There is a significant difference between the redox states of the cytosol (which is highly reducing) and the periplasm (which is oxidizing), and it was originally thought that this difference was sufficient to allow disulfide bond formation to occur at a reasonable rate.

However, proteins have been identified in the periplasm of *E. coli* that catalyze both the formation and removal of disulfide bonds and thus speed up the rate at which proteins that contain such bonds are folded.

The two major proteins in *E. coli* involved in this process are both periplasmic proteins and are called DsbA and DsbC (Fig. 6). DsbA is a powerful oxidant, catalyzing the formation of disulfide bonds between cysteine residues in proteins that have been newly secreted to the periplasm, and itself becoming reduced as a consequence. Although bacteria lacking DsbA are viable, they show enhanced turnover of periplasmic proteins (owing to the presence of proteases in the periplasm that degrade unfolded proteins), lack of motility (owing to the failure of flagella to assemble), and high sensitivity to reducing agents. DsbA apparently binds proteins in a hydrophobic groove, rather analogous to the mechanism seen in some of the

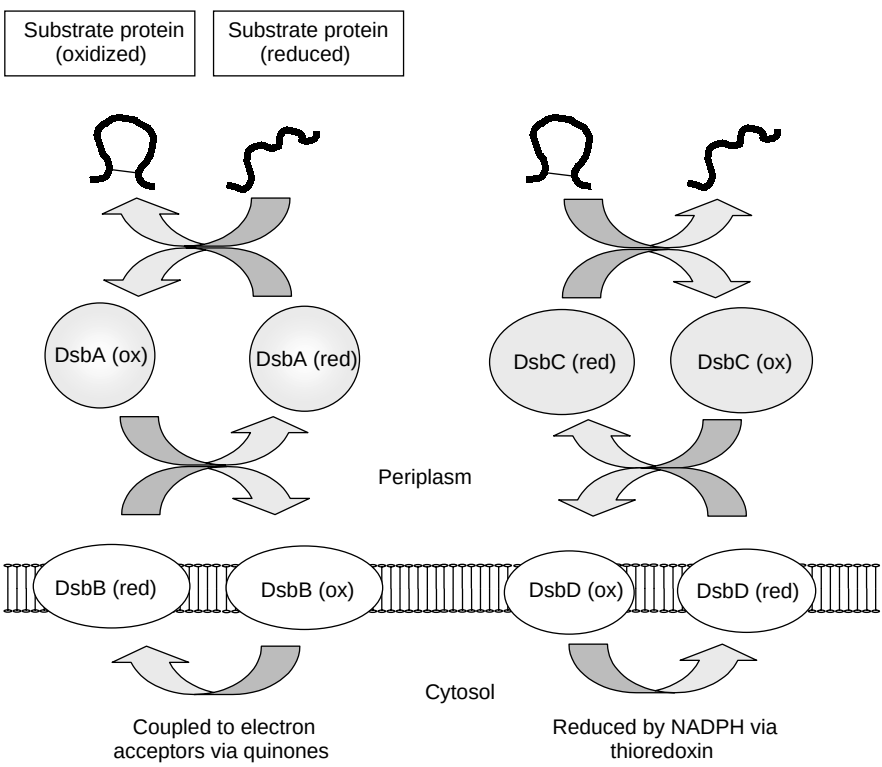

Fig. 6 Mode of action of some of the periplasmic Dsb proteins in *E. coli*. DsbA acts as an oxidase to introduce disulfide bonds into proteins with free thiols. In doing this, it is itself reduced, and must be reoxidized by oxidized DsbB, which in turn becomes reduced. DsbB becomes reoxidized by donating electrons to quinines, which flow to oxygen or other electron acceptors. DsbC in its reduced form acts either as a reductase (the net effect of which is to oxidize DsbC) or an isomerase (where no net redox reaction occurs). In the former capacity, it must be rereduced before regaining its activity; this is done by DsbD, which in turn derives its reducing power ultimately from cytosolic NADPH. Other Dsb proteins also exist in the periplasm of *E. coli* but have been omitted for clarity.

chaperones discussed above, but it is not significantly specific about which cysteines it oxidizes, so both correct and incorrect isomers of the same protein can be formed. The reduced DsbA must be reoxidized for further activity, a process that is mediated by a membrane protein called DsbB, which itself interacts directly with intermediates of the oxidative electron transport pathway to transfer the electrons liberated in disulfide bond formation to quinones and ultimately to either oxygen or, under anaerobic conditions, to other electron acceptors.

The formation of incorrect disulfide-bonded isomers by DsbA points to the need for an isomerase protein that can catalyze the breakage and reformation of disulfide bonds in proteins until the correct isomeric (and active) form is reached. In the *E. coli* periplasm, this role is performed by the protein DsbC. Like protein disulfide isomerase (PDI) (see below), this protein has an innate chaperone activity, in that it can recognize and bind unfolded proteins even when such proteins lack any cysteine residues. It seems likely that DsbC acts to bind incorrectly folded proteins, and then breaks the disulfides if they are present. Disulfides may then be reformed either by subsequent action of DsbA, or by spontaneous reoxidation of the reduced cysteines. In contrast to DsbA, therefore, DsbC must be maintained in a reduced state, and this is mediated by another membrane protein called DsbD. DsbD is in turn maintained in a reduced state by interacting with the cytosolic protein thioredoxin, which itself is kept reduced by metabolically generated NADPH.

The site for disulfide bond formation and isomerization in eukaryotes is the lumen of the endoplasmic reticulum. As is the case with bacteria, disulfide bond formation in newly imported proteins is an enzyme-catalyzed process, the enzyme in this case being protein disulfide isomerase. PDI is by far the most abundant chaperone in the ER. As with bacteria, a membrane protein (Ero1p) exists to restore the active (oxidized) state of PDI after the formation of disulfides has taken place. Reduced Ero1p in turn is reoxidized directly by oxygen, though its activity is also modulated by FAD. In addition to acting as an oxidase, PDI can shuffle the disulfide bonds in a protein after they have formed (hence its name), so it effectively combines the roles of DsbA and DsbC. As with DsbC, it is capable of showing chaperone activity, and so presumably can recognize proteins in which the disulfide bonds are not in the correct position for complete folding. The catalytic site and chaperone site are not the same and indeed are on different domains of the protein. This activity is more important in eukaryotic cells than in prokaryotic ones, as eukaryotic proteins have much larger numbers of disulfide bonds on average, and hence the possibility of forming incorrect disulfide bonds is correspondingly higher.

5.1.2 Hsp33: A Chaperone with a Redox Switch

An intriguing chaperone protein is found in *E. coli*, with homologs in many other species, which is heat shock induced but active only when oxidized. This protein, Hsp33 or HslO, binds very tightly to zinc ions when reduced, coordinating the zinc with four cysteine residues. Under relatively mild oxidizing conditions, the zinc is released and the cysteines form two disulfide bonds. This release causes a substantial change in the conformation of the protein, which significantly increases the amount of exposed hydrophobic area of the protein. In this state, Hsp33 acts as a "holder" rather like the sHSPs, and has

been shown to be capable of protecting a number of substrates from aggregation, whether induced by heat shock or by chemical denaturation. However, it has not yet been possible to demonstrate refolding of the bound protein substrates *in vitro*, implying that *in vivo* Hsp33 may interact either with one of the known chaperone proteins, or with a cofactor which is yet to be identified. The likely function of Hsp33 is protection of proteins, which may be particularly susceptible to oxidative damage *in vivo*, and the rapid activation of the chaperone activity of Hsp33 that takes place when the protein is oxidized, makes it an excellent first line of defense against oxidative stress, which can act until the cell has had time to induce a range of other defenses that will tend to act by restoring the normal redox potential of the cytosol.

5.2 Chaperone Networks and Protein Quality Control in the Cell

As research into molecular chaperones has progressed, it has moved to embrace considerations not only of how the members of individual chaperone families function, but also how they interact with other chaperone families within the cell. Much emphasis is now placed on understanding these interactions or networks. One such network has already been discussed briefly above: the interaction of Hsp70 and Hsp90 proteins in the activation of steroid hormone receptors. Chaperone networks have evolved to enable cells to maximize the efficiency with which they produce folded and active proteins, a process that faces two particular hurdles: the inefficiency implicit in the folding process itself (i.e. the extent to which proteins may finish up becoming misfolded or aggregated), and the fact that folding can be reversed by stresses such as heat shock. These networks also have to be integrated with the cell's mechanisms for disposing of proteins that are not functional. In the next two sections, we will look at two of the best understood of these: the system of protein quality control in the *E. coli* cytosol, and the system in the ER of eukaryotic cells. Neither of these, it must be emphasized, is yet completely characterized.

5.2.1 Chaperone Networks and Protein Quality Control in the *E. coli* Cytosol

It has already been pointed out in the text above that DnaK (one of the Hsp70 protein homologs in *E. coli*) and the unrelated protein trigger factor cooperate in chaperoning proteins that are still in the process of being synthesized on the ribosome. This is an essential process, and the trigger factor and DnaK must significantly overlap in the roles that they play in it, as demonstrated by the fact that loss of both genes together is lethal to the cell, whereas loss of either gene individually is not. It is likely that in normal cells, they act sequentially, with the trigger factor binding to proteins as they first emerge from the ribosome, and DnaK interacting with a subset of these later during synthesis.

Other interactions may exist between the DnaK and the GroEL chaperone systems. *In vitro* studies using purified substrate proteins have demonstrated that proteins can be transferred between these two systems effectively in either direction. Detailed *in vivo* evidence for a network between these two proteins is lacking, however, except inasmuch as a few proteins have been identified by coimmunoprecipitation experiments as being substrates of both systems.

More direct evidence exists to link the sHSPs of *E. coli* with the Hsp70 and Hsp60

major chaperone families. *In vitro* experiments show that unfolded proteins can be held in a nonaggregated but nonactive state bound to the sHSPs, and subsequently transferred to the DnaK and GroEL families for folding. *In vivo* experiments also support the existence of such interactions, although these interactions cannot be essential for viability as even under fairly extreme conditions, loss of the sHSPs from *E. coli* does not cause a severe phenotype. It may be that, as was seen in the case of trigger factor and DnaK, the ability for the cell to ''hold'' transiently unfolded proteins in a refoldable form is so important that more than one protein can act as the ''holder,'' and to date the other proteins that can do this have not been identified.

A third important interaction that exists between chaperones in *E. coli* is that between the Hsp70 protein DnaK and the Hsp100 protein, ClpB. Here, the evidence for an important interaction is much more compelling. It is clear both from *in vitro* and *in vivo* experiments that proteins aggregated by a stress such as heat shock can be disaggregated by the combined actions of DnaK and ClpB, and that the two proteins have very different roles in the process as they cannot substitute for each other. It is likely that ClpB acts upon protein aggregates first, altering their structure to a point where DnaK is able to promote their refolding. ClpB has thus been classified by some workers as an ''unfolder,'' and it is not able to assist the refolding of proteins on its own. Its mode of action is not yet understood, but it is a member of a large family of proteins called the AAA proteins, many of which have a major role in protein quality control. This leads us into the complex area of the overlap between protein refolding and protein degradation.

E. coli (and all other organisms) contains a large number of different proteases whose job it is to remove inactive or damaged proteins from the cell. This process becomes particularly important at high temperatures. A large number of these proteases are, along with ClpB, members of the AAA family of proteins, and some have features reminiscent of molecular chaperones. For example, several of them have an ATP-dependent ability to unfold proteins, and exist as ring-shaped oligomers (usually hexamers). For most of these, unfolding occurs as an essential step prior to degradation of the unfolded protein, either by another site on the same polypeptide that mediated the unfolding, or by another protein that acts in concert with the ''unfoldase.'' The unfoldase activity can, in some cases, give rise to a chaperone-like activity in *in vitro* experiments, although only in the case of ClpB does it appear to have a true role in helping proteins to refold *in vivo*. Degradation by these proteases is reduced by mutations in the DnaK and GroEL chaperone families, implying an interaction, indirect or direct, between the chaperone proteins and the proteases.

Thus, when proteins are unfolded in the *E. coli* cell, they may be captured by ''holders'' for subsequent refolding, or form aggregates that can be disaggregated by ClpB before refolding by DnaK (Fig. 7). In addition, they may be targeted to cellular proteases. Both these processes will help ensure that only active proteins are present in the cell, but the way the balance between them is achieved (i.e. what determines whether a given protein is degraded or refolded) is not known. Clearly, refolding is a preferential option to degradation if the cell is not to lose the investment in ATP that it has already made in producing the protein; equally clearly, refolding is not

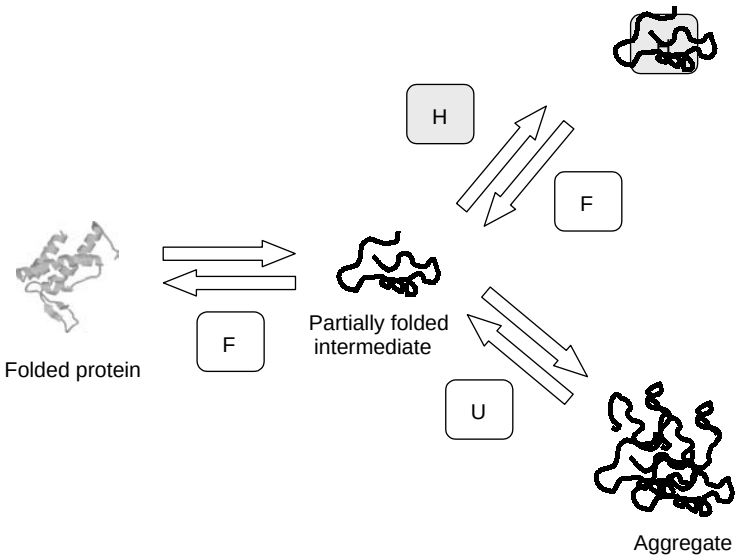

Fig. 7 Holders, folders, and unfolders in a chaperone network. A folded protein may become partially unfolded (for example, due to heat shock). It may refold spontaneously from this state, or may require the assistance of a "folder" chaperone (F), such as the GroEL/ES or DnaK/J systems. The partially unfolded state may interact further with other proteins in the same state to form an aggregate. This can be reversed by the action of an "unfolder" (such as ClpB). Alternatively, a "holder" chaperone (such as an sHSP) may hold the intermediate in a nonactive but nonaggregated state until a folder chaperone can refold it.

always an option, otherwise there would be no need for the cellular proteases to exist.

5.2.2 Protein Quality Control in the Endoplasmic Reticulum

The ER is of particular importance in the quality control of protein folding. It is the site for several important covalent modifications that are key to protein structure, including disulfide bond formation and glycosylation, and proteins pass through the ER en route to the plasma membrane and beyond, where they will be inaccessible to any further modification by chaperones. Thus, a substantial machinery is found within the ER to ensure that proteins are correctly folded and modified, and that leads to the degradation of those that are not. The components of this machinery include proteins that form and isomerize disulfide bonds, a key Hsp70 homolog, several proteins with DnaJ domains, and two proteins called calnexin and calreticulin. Other proteins with chaperone activity also exist in the ER (for example, an Hsp90 homolog called Grp94), although genetic evidence suggests they play a relatively minor role.

The Hsp70 protein homolog (also referred to as BiP, Grp78, and Kar2p) can be shown to associate with incompletely folded proteins using the same rules for binding as other members of the Hsp70 family. BiP has a C-terminal amino acid signal that retains it in the ER by cycles of retrieval from the Golgi body, and so any

protein that has bound to it is also kept in the ER until the sites that BiP recognizes are buried in the folded protein (or until the protein is retrotranslocated back to the cytosol for degradation, a process in which BiP also has a major role). Coimmunoprecipitation and cross-linking experiments have shown that BiP may bind early in the folding pathway, as do other Hsp70 homologs such as DnaK, but other studies support a role later during folding. Complexes recovered by such experiments often include other ER resident chaperones as well as BiP, supporting the model of a network of chaperone proteins acting in concert to promote correct folding. BiP can be acted upon by several different membrane proteins with DnaJ domains (all of which face the lumenal side of the membrane) and these may all be involved in BiP's various functions. As with classic Hsp40 proteins, they can act to stimulate the ATPase activity of their Hsp70 partner (BiP) and thus improve the efficiency of the cycling on and off the complex of partially folded protein. BiP is not the only chaperone in the ER that can interact with partially folded proteins: the PDI shares this property, and thus may be involved in retention of incorrectly folded proteins as well as in the formation and isomerization of disulfide bonds.

A more complex system exists to ensure that glycosylated proteins are in their correct conformation before further transport through the secretory pathway. Glycan chains are added at selected asparagine residues to proteins as they pass into the ER through the membrane. These chains have the formula Glc_3-Man_9-$(GlcNAc)_2$, where Glc is glucose, Man is mannose, and GlcNAc is *N*-acetyl-glucosamine. Shortly after these chains have been added, two of the terminal glucoses are clipped off by the enzymes glucosidase I and II, leaving a single glucose. The presence of this glucose makes the protein bind to one of the two chaperones calnexin and calreticulin, which are respectively transmembrane and lumenal proteins, thus preventing their further transport from the ER. In this state, they can also be acted upon by another PDI called ERp57, which itself binds to both calnexin and calreticulin. The final glucose residue on the bound substrate is removed by glucosidase II, which enables the protein's release from calnexin or calreticulin. If it is correctly folded, it will now exit the ER, but if it is incorrectly folded, a terminal glucose is added back to the protein by another enzyme (UDP-glucose:glycoprotein glucosyl transferase), which only acts on proteins that contain areas that are not folded, and the protein will be rebound by calnexin and calreticulin, thus reinitiating the cycle of protein folding on these chaperones. If after several cycles of this reaction the protein has still failed to fold correctly, it can be targeted for retrotranslocation and degradation in the cytosol.

5.3 Examples of Substrate-specific Chaperones

The chaperones that have been considered above are all examples of what are often referred to as "broad-spectrum" chaperones: that is, they have a range of different substrate proteins with which they interact. Not all chaperones are broad range, however; some interact with only one or a few substrates. There are many examples of these in the literature, two of which are considered here.

5.3.1 Hsp47 and Collagen Assembly

In addition to the various chaperones present in the ER and described in the

previous section, another chaperone protein called Hsp47 is also present. This protein is required for the chaperoning of one class of proteins only: the collagens. Collagens are the most abundant proteins in mammals, and the importance of the action of Hsp47 is shown by the fact that mice with both copies of the *hsp47* gene disrupted do not survive to term and show many abnormalities in collagen processing, which has numerous lethal effects on tissue integrity. How does Hsp47 act? To answer this question, we need to know something about collagen assembly.

Although there are large numbers of different types of collagen, they share certain common features. Collagen is synthesized as procollagen, which contains domains at both the N- and C-termini, which are removed in the formation of the mature collagen molecule. These N- and C-termini form globular domains while the rest of the procollagen molecule forms alpha-helices, three of which coil around each other to form the characteristic collagen triple helix. The alpha-helices consist of repeats of the motif G-X-Y, where Y is often proline (or hydroxyproline, formed in the ER from proline). This molecule is then transported to the Golgi body and ultimately to the cell surface, where the N- and C-terminal domains are removed and collagen triple helices become associated via unknown mechanisms into long fibrils. Hsp47 has been shown to bind to the triple-helical procollagen molecule, and this binding requires the presence of at least one arginine residue in the triple helix. The precise role of the Hsp47 binding is not yet clear, but one promising hypothesis is that it prevents the premature aggregation of procollagen molecules into larger fibrils, that would not be exportable from the ER. Intriguingly, Hsp47 has been shown to be capable of forming trimers, and one speculation is that it may form rings around the triple-helical procollagen molecules that could serve to prevent lateral aggregation with other procollagens.

5.3.2 PapD and the Assembly of Bacterial Pili

The interactions of chaperones and substrates generally lack the degree of specificity that is commonly associated with protein–protein interactions, and it is this lack of specificity that contributes to the broad range of substrates on which chaperones can act. However, the substrate-specific chaperones, such as Hsp47 above, must be able to recognize and act on particular substrates. In one group of proteins, the so-called PapD chaperones, the molecular basis for this recognition is understood, and this gives us useful insights into how such chaperones can act.

PapD chaperones are members of a superfamily of proteins, found in the periplasm of different bacteria, which are involved in the assembly of pili and other surface structures. This gives them considerable applied interest because pili and fimbrae play a major role in the attachment of bacteria to surfaces and in many cases this makes them important virulence factors in disease. More than 30 members of this superfamily are known. PapD is one of the best studied of these and was distinguished as the first chaperone (albeit one whose action is highly specific) for which a crystal structure was obtained.

Certain strains of *E. coli* produce so-called P pili, which are complex structures consisting of an adhesion protein (PapG) at the tip of the pilus, joined via an adaptor protein (PapF) to a string of subunits (of the PapE protein) in an open helical conformation, which in turn are linked by another adaptor protein (PapK) to another

string of subunits (PapA) in a helical cylinder. All these proteins are produced in the cytosol of *E. coli* and exported to the periplasm, and their subsequent assembly into a complete pilus requires the action of the PapD chaperone. In the absence of PapD, the various subunits are degraded by periplasmic proteases. PapD is not, however, part of the assembled pilus, and so it admirably fulfills the definition of a molecular chaperone.

Solving the structure of PapD, both alone and in complexes with proteins (PapK and PapE) whose folding it assists, has demonstrated how this chaperone works. The pilus subunits possess an Ig-fold, which normally consists of seven antiparallel beta sheets. In the pilus subunits, one of these helices is missing, which exposes a deep cleft in the protein. When the protein is in a complex with PapD, an N-terminal beta sheet from PapD fills the space that would normally be occupied by this missing region, thus stabilizing the protein by hiding the hydrophobic residues that would otherwise be exposed. This is referred to as "donor strand complementation," since the completion of the stable Ig-fold is made possible by the "donation" of a β-strand from the chaperone. Indeed, it is likely to be the case that the folding of the individual pilus subunits only takes place in the presence of PapD. Given that this complex is now stabilized by the pilus-chaperone interaction, how does subsequent assembly of the pilus subunit into the growing pilus take place? This assembly happens by a process referred to as "donor strand exchange," where the place occupied by the donated strand of the PapD chaperone is taken by a donated beta strand from the adjacent pilus subunit. This process takes place at an outer membrane pore made up of the protein PapC (sometimes referred to as an "usher"). PapD, complexed to the various different pilus subunits, arrives at the PapC, and donor strand exchange takes place in order to add each new pilus subunit to those already there, which will thus be progressively extruded from the outer membrane of the cell. The order in which the subunits are found is presumably determined by the relative affinities of the different pilus subunits for each other, although this has not been directly proven. Recent results show that the role of PapD in this entire process is far from passive. PapD primes pilus subunits for assembly by holding the groove in an open conformation, which is closed during donor strand exchange, effectively locking the N-terminal extension of each adjacent pilus subunit in place on its neighbor. The interaction between PapD and its substrates is highly specific and, moreover, PapD can be said to be providing steric information required for the correct folding of these substrates.

6 Conclusion

This review has focused on the roles of molecular chaperones in the cell and described some instances of the way in which these roles can be understood in terms of the structures and reaction mechanisms of individual molecular chaperones. It must be emphasized that this is an area of research that has expanded very rapidly over the past decade and a half, and many aspects of molecular chaperone biology have had to be neglected in such a brief discussion. More attention is now being paid to the networks in which molecular chaperones operate and how these overlap with other key areas of the cell's activity,

particularly in the whole area of ''quality control'' – the optimization of protein folding activity under both normal and stressed conditions, and the rapid disposal of proteins which fail to meet the cell's requirements. Attention is also shifting to examining the roles of different chaperones in disease, most notably the protein-folding diseases such as the prion diseases, Alzheimer's, Parkinson's, and others. The possibility that a more complete understanding of the ways in which chaperones are involved in such diseases may lead to new therapies is an exciting one and a powerful incentive to further research in this field.

Acknowledgments

I am grateful to Professor R John Ellis, FRS, for his many useful comments on an earlier draft of this article.

See also Circular Dichroism in Protein Analysis.

Bibliography

Books and Reviews

Buchner, J. (1999) Hsp90 and Co. – a holding for folding, *Trends Biochem. Sci.* **24**, 136–141.

Bukau, B. (1999) *Molecular Chaperones and Folding Catalysts: Regulation, Cellular Functions and Mechanisms*, Taylor & Francis, London.

Graf, P.C.F., Jakob, U. (2002) Redox-regulated molecular chaperones, *Cell. Mol. Life Sci.* **59**, 1624–1631.

Hartl, F.U., Hayer-Hartl, M. (2002) Protein folding – molecular chaperones in the cytosol: from nascent chain to folded protein, *Science* **295**, 1852–1858.

Haslbeck, M. (2002) sHSPs and their role in the chaperone network, *Cell. Mol. Life Sci.* **59**, 1649–1657.

Kadokura, H., Katzen, F., Beckwith, J. (2003) Protein disulfide bond formation in prokaryotes, *Annu. Rev. Biochem.* **72**, 111–135.

Lund, P.A. (2001) *Molecular Chaperones in the Cell*, Oxford University Press, London.

Saibil, H. (2000) Molecular chaperones: containers and surfaces for folding, stabilising or unfolding proteins, *Curr. Opin. Struct. Biol.* **10**, 251–258.

Schrag, J.D., Procopio, D.O., Cygler, M., Thomas, D.Y., Bergeron, J.J.M. (2003) Lectin control of protein folding and sorting in the secretory pathway, *Trends Biochem. Sci.* **28**, 49–57.

Walter, S., Buchner, J. (2002) Molecular chaperones – cellular machines for protein folding, *Angew. Chem., Int. Ed.* **41**, 1098–1113.

Primary Literature

Anfinsen, C.B. (1973) Principles that govern the folding of protein chains, *Science* **181**, 223–230.

Bardwell, J.C.A., McGovern, K., Beckwith, J. (1991) Identification of a protein required for disulfide bond formation *in vivo*, *Cell* **67**, 581–589.

Beckmann, R.P., Mizzen, L.A., Welch, W.J. (1990) Interaction of hsp70 with newly synthesized proteins – implications for protein folding and assembly, *Science* **248**, 850–854.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D.C., Joachimiak, A., Horwich, A.L., Sigler, P.B. (1994) The crystal-structure of the bacterial chaperonin GroEL at 2.8- angstrom, *Nature* **371**, 578–586.

Brinker, A., Pfeifer, G., Kerner, M.J., Naylor, D.J., Hartl, F.U., Hayer-Hartl, M. (2001) Dual function of protein confinement in chaperonin-assisted protein folding, *Cell* **107**, 223–233.

Buchberger, A., Schroder, H., Hesterkamp, T., Schonfeld, H.J., Bukau, B. (1996) Substrate shuttling between the dnaK and GroEL systems indicates a chaperone network promoting protein folding, *J. Mol. Biol.* **261**, 328–333.

Buchner, J., Schmidt, M., Fuchs, M., Jaenicke, R., Rudolph, R., Schmid, F.X., Kiefhaber, T. (1991) GroE facilitates refolding of citrate synthase by suppressing aggregation, *Biochemistry* **30**, 1586–1591.

Burston, S.G., Ranson, N.A., Clarke, A.R. (1995) The origins and consequences of asymmetry

in the chaperonin reaction cycle, *J. Mol. Biol.* **249**, 138–152.

Chappell, T.G., Welch, W.J., Schlossman, D.M., Palter, K.B., Schlesinger, M.J., Rothman, J.E. (1986) Uncoating ATPase is a member of the 70 kilodalton family of stress proteins, *Cell* **45**, 3–13.

Chen, S., Roseman, A.M., Hunter, A.S., Wood, S.P., Burston, S.G., Ranson, N.A., Clarke, A.R., Saibil, H.R. (1994) Location of a folding protein and shape changes in GroEL-GroES complexes imaged by cryoelectron microscopy, *Nature* **371**, 261–264.

Deshaies, R.J., Koch, B.D., Wernerwashburne, M., Craig, E.A., Schekman, R. (1988) A subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides, *Nature* **332**, 800–805.

Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., Bukau, B. (1999) Trigger factor and DnaK cooperate in folding of newly synthesized proteins, *Nature* **400**, 693–696.

Ehrnsperger, M., Graber, S., Gaestel, M., Buchner, J. (1997) Binding of non-native protein to hsp25 during heat shock creates a reservoir of folding intermediates for reactivation, *EMBO J.* **16**, 221–229.

Ellis, R.J. (1994) Molecular chaperones – opening and closing the Anfinsen cage, *Curr. Biol.* **4**, 633–635.

Ellis, R.J. (1999) Chaperonins, *Curr. Biol.* **9**, R352–R352.

Ellis, R.J. (2001) Macromolecular crowding: obvious but underappreciated, *Trends Biochem. Sci.* **26**, 597–604.

Ellis, R.J. (2001) Molecular chaperones: inside and outside the Anfinsen cage, *Curr. Biol.* **11**, R1038–R1040.

Ellis, R.J., Hartl, F.U. (1996) Protein folding in the cell: competing models of chaperonin function, *FASEB J.* **10**, 20–26.

Ewalt, K.L., Hendrick, J.P., Houry, W.A., Hartl, F.U. (1997) In vivo observation of polypeptide flux through the bacterial chaperonin system, *Cell* **90**, 491–500.

Farr, G.W., Scharl, E.C., Schumacher, R.J., Sondek, S., Horwich, A.L. (1997) Chaperonin-mediated folding in the eukaryotic cytosol proceeds through rounds of release of native and nonnative forms, *Cell* **89**, 927–937.

Flynn, G.C., Chappell, T.G., Rothman, J.E. (1989) Peptide binding and release by proteins implicated as catalysts of protein assembly, *Science* **245**, 385–390.

Frydman, J., Nimmesgern, E., Erdjumentbromage, H., Wall, J.S., Tempst, P., Hartl, F.U. (1992) Function in protein folding of TRiC, a cytosolic ring complex containing TCP-1 and structurally related subunits, *EMBO J.* **11**, 4767–4778.

Goloubinoff, P., Christeller, J.T., Gatenby, A.A., Lorimer, G.H. (1989) Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on 2 chaperonin proteins and Mg-ATP, *Nature* **342**, 884–889.

Harrison, C.J., Hayer-Hartl, M., DiLiberto, M., Hartl, F.U., Kuriyan, J. (1997) Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK, *Science* **276**, 431–435.

Hebert, D.N., Foellmer, B., Helenius, A. (1995) Glucose trimming and reglucosylation determine glycoprotein association with calnexin in the endoplasmic-reticulum, *Cell* **81**, 425–433.

Hemmingsen, S.M., Woolford, C., Van der vies, S.M., Tilly, K., Dennis, D.T., Georgopoulos, C.P., Hendrix, R.W., Ellis, R.J. (1988) Homologous plant and bacterial proteins chaperone oligomeric protein assembly, *Nature* **333**, 330–334.

Houry, W.A., Frishman, D., Eckerskorn, C., Lottspeich, F., Hartl, F.U. (1999) Identification of *in vivo* substrates of the chaperonin Groel, *Nature* **402**, 147–154.

Jakob, U., Gaestel, M., Engel, K., Buchner, J. (1993) Small heat-shock proteins are molecular chaperones, *J. Biol. Chem.* **268**, 1517–1520.

Jakob, U., Muse, W., Eser, M., Bardwell, J.C.A. (1999) Chaperone activity with a redox switch, *Cell* **96**, 341–352.

Kozutsumi, Y., Segal, M., Normington, K., Gething, M.J., Sambrook, J. (1988) The presence of malfolded proteins in the endoplasmic-reticulum signals the induction of glucose-regulated proteins, *Nature* **332**, 462–464.

Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M.K., Hartl, F.U. (1992) Successive action of Dnak, Dnaj and Groel along the pathway of chaperone-mediated protein folding, *Nature* **356**, 683–689.

Lee, G.J., Roseman, A.M., Saibil, H.R., Vierling, E. (1997) A small heat shock protein stably binds heat-denatured model substrates

and can maintain a substrate in a folding-competent state, *EMBO J.* **16**, 659–671.

Liou, A.K.F., Willison, K.R. (1997) Elucidation of the subunit orientation in CCT (chaperonin containing TCP1) from the subunit composition of CCT micro-complexes, *EMBO J.* **16**, 4311–4316.

Llorca, O., McCormack, E.A., Hynes, G., Grantham, J., Cordell, J., Carrascosa, J.L., Willison, K.R., Fernandez, J.J., Valpuesta, J.M. (1999) Eukaryotic type II chaperonin CCT interacts with actin through specific subunits, *Nature* **402**, 693–696.

Llorca, O., Smyth, M.G., Carrascosa, J.L., Willison, K.R., Radermacher, M., Martin, J., Horwich, A.L., Hartl, F.U. (1992) Prevention of protein denaturation under heat-stress by the chaperonin Hsp60, *Science* **258**, 995–998.

Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A.L., Hartl, F.U. (1991) Chaperonin-mediated protein folding at the surface of GroEL through a molten globule-like intermediate, *Nature* **352**, 36–42.

Martin, J., Mayhew, M., Langer, T., Hartl, F.U. (1993) The reaction cycle of GroEL and GroES in chaperonin-assisted protein-folding, *Nature* **366**, 228–233.

Mogk, A., Tomoyasu, T., Goloubinoff, P., Rudiger, S., Roder, D., Langen, H., Bukau, B. (1999) Identification of thermolabile *Escherichia coli* proteins: Prevention and reversion of aggregation by DnaK and ClpB, *EMBO J.* **18**, 6934–6949.

Munro, S., Pelham, H.R.B. (1986) An Hsp70-like protein in the ER – identity with the 78 kd glucose-regulated protein and immunoglobulin heavy-chain binding-protein, *Cell* **46**, 291–300.

Prodromou, C., Panaretou, B., Chohan, S., Siligardi, G., O'Brien, R., Ladbury, J.E., Roe, S.M., Piper, P.W., Pearl, L.H. (2000) The ATPase cycle of HSP90 drives a molecular 'clamp' via transient dimerization of the N-terminal domains, *EMBO J.* **19**, 4383–4392.

Prodromou, C., Roe, S.M., O'Brien, R., Ladbury, J.E., Piper, P.W., Pearl, L.H. (1997) Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone, *Cell* **90**, 65–75.

Queitsch, C., Sangster, T.A., Lindquist, S. (2002) Hsp90 as a capacitor of phenotypic variation, *Nature* **417**, 618–624.

Rietsch, A., Belin, D., Martin, N., Beckwith, J. (1996) An *in vivo* pathway for disulfide bond isomerization in *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13048–13053.

Rudiger, S., Germeroth, L., SchneiderMergener, J., Bukau, B. (1997) Substrate specificity of the Dnak chaperone determined by screening cellulose-bound peptide libraries, *EMBO J.* **16**, 1501–1507.

Rye, H.S., Burston, S.G., Fenton, W.A., Beechem, J.M., Xu, Z.H., Sigler, P.B., Horwich, A.L. (1997) Distinct actions of *cis* and *trans* ATP within the double ring of the chaperonin GroEL, *Nature* **388**, 792–798.

Rye, H.S., Roseman, A.M., Chen, S.X., Furtak, K., Fenton, W.A., Saibil, H.R., Horwich, A.L. (1999) GroEL-groES cycling: ATP and non-native polypeptide direct alternation of folding-active rings, *Cell* **97**, 325–338.

Schroder, H., Langer, T., Hartl, F.U., Bukau, B. (1993) DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage, *EMBO J.* **12**, 4137–4144.

Shtilerman, M., Lorimer, G.H., Englander, S.W. (1999) Chaperonin function: folding by forced unfolding, *Science* **284**, 822–825.

Sternlicht, H., Farr, G.W., Sternlicht, M.L., Driscoll, J.K., Willison, K., Yaffe, M.B. (1993) The T-complex polypeptide-1 complex is a chaperonin for tubulin and actin in vivo, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9422–9426.

Tasab, M., Batten, M.R., Bulleid, N.J. (2000) Hsp47: a molecular chaperone that interacts with and stabilizes correctly-folded procollagen, *EMBO J.* **19**, 2204–2211.

Tasab, M., Jenkinson, L., Bulleid, N.J. (2002) Sequence-specific recognition of collagen triple helices by the collagen-specific molecular chaperone Hsp47, *J. Biol. Chem.* **277**, 35007–35012.

Teter, S.A., Houry, W.A., Ang, D., Tradler, T., Rockabrand, D., Fischer, G., Blum, P., Georgopoulos, C., Hartl, F.U. (1999) Polypeptide flux through bacterial Hsp70: Dnak cooperates with trigger factor in chaperoning nascent chains, *Cell* **97**, 755–765.

Trent, J.D., Nimmesgern, E., Wall, J.S., Hartl, F.U., Horwich, A.L. (1991) A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein t-complex polypeptide-1, *Nature* **354**, 490–493.

Tu, B.P., Ho-Schleyer, S.C., Travers, K.J., Weissman, J.S. (2000) Biochemical basis of oxidative protein folding in the endoplasmic reticulum, *Science* **290**, 1571–1574.

Weissman, J.S., Hohl, C.M., Kovalenko, O., Kashi, Y., Chen, S.X., Braig, K., Saibil, H.R., Fenton, W.A., Horwich, A.L. (1995) Mechanism of GroEL action – productive release of polypeptide from a sequestered position under GroES, *Cell* **83**, 577–587.

Weissman, J.S., Kashi, Y., Fenton, W.A., Horwich, A.L. (1994) GroEL-mediated protein-folding proceeds by multiple rounds of binding and release of non-native forms, *Cell* **78**, 693–702.

Wiech, H., Buchner, J., Zimmermann, R., Jakob, U. (1992) Hsp90 chaperones protein folding *in vitro*, *Nature* **358**, 169–170.

Xu, Z.H., Horwich, A.L., Sigler, P.B. (1997) The crystal structure of the asymmetric GroEL-GroES-(ADP)(7) chaperonin complex, *Nature* **388**, 741–750.

Yaffe, M.B., Farr, G.W., Miklos, D., Horwich, A.L., Sternlicht, M.L., Sternlicht, H. (1992) TCP-1 complex is a molecular chaperone in tubulin biogenesis, *Nature* **358**, 245–248.

Zhu, X.T., Zhao, X., Burkholder, W.F., Gragerov, A., Ogata, C.M., Gottesman, M.E., Hendrickson, W.A. (1996) Structural analysis of substrate binding by the molecular chaperone DnaK, *Science* **272**, 1606–1614.

Chemiluminescence and Bioluminescence, Analysis by

Larry J. Kricka
University of Pennsylvania, Philadelphia, PA, USA

1 Chemiluminescence 523

2 Bioluminescence 523

3 Electrochemiluminescence 524

4 Applications 525
4.1 Measurement of Light Emission 525
4.2 Immunoassay, Blotting, and Nucleic Acid Hybridization Assays 527
4.2.1 Labels 528
4.2.2 Detection Reactions 528
4.3 Enzymes, Substrates, and Cofactors 530
4.4 Reporter Genes 531
4.5 Rapid Microbiology 532
4.6 Cellular Luminescence 533
4.7 Toxicity Testing 533

Bibliography 533
Books and Reviews 533
Primary Literature 533

Keywords

Bioluminescence
Light emission from a living organism.

Chemiluminescence
Light emission that occurs in chemical reactions as a result of the decay of chemiexcited species.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Electrochemiluminescence
Light emission that occurs in electrochemical reactions.

Homogeneous Assay
An assay (e.g. immunoassay or nucleic acid hybridization assay) that does not require a separation step.

Immunoassay
An assay that uses specific antibodies or labeled antibodies to detect an antigen (the analyte).

Luciferase
The generic name for an enzyme that catalyzes a bioluminescent reaction.

Luciferin
The generic name for the substrate of a luciferase.

Luminometer
An instrument for detecting and measuring light emission.

Nucleic Acid Hybridization Assay
An assay that utilizes the specific recognition of a nucleic acid probe or labeled nucleic acid probe for a complementary sequence on the target nucleic acid.

Photoprotein
A tightly bound protein–luciferin complex that can be activated to produce light.

Bioluminescent (BL) and chemiluminescent (CL) reactions are used as analytical detection tools in immunoassay, protein and nucleic acid blotting, nucleic acid sequencing, and hybridization assays, and in reporter gene studies. In a BL or CL assay, the intensity or the total light emission is measured and related to the concentration of the unknown analyte. Light is measured quantitatively using a luminometer (photomultiplier tube or charge-coupled device) as a detector or qualitatively by means of photographic or X-ray film. The main advantages of these reactions are simplicity and an analytical sensitivity capable of obtaining results from minute amounts of analyte [attomole–zeptomole (10^{-18}–10^{-21} mol) range]. In immunoassay and nucleic acid hybridization tests, these reactions or their components are used in two ways: as labels and as detection systems for labels of other types (e.g. enzymes). Acridinium esters, acridinium sulfonyl carboxamides, isoluminol derivatives, the photoprotein aequorin, and luciferases from the firefly, marine bacteria, *Vargula* (formerly *Cypridina*), and *Renilla* are the principal luminescent labels. CL and BL detection schemes have been developed for assaying alkaline phosphatase, glucose oxidase, glucose 6-phosphate dehydrogenase, horseradish peroxidase, and xanthine oxidase labels. In addition, simple homogeneous assays are possible with CL acridinium ester and isoluminol labels, as well as with glucose 6-phosphate dehydrogenase labels. CL and BL reactions

can be adapted for the analysis of enzymes, substrates, cofactors, inhibitors, and metal ions, for the study of cellular processes (e.g. phagocytosis), and for toxicity testing. Luciferase genes (*luc, lux*) are now used as reporter genes, and the gene expression is quantitated by measuring the expressed luciferase in a BL assay.

1 Chemiluminescence

Chemiluminescence (CL) is the light emission produced in certain chemical reactions. Excited singlet-state molecules are formed in the reaction via the decomposition of high-energy intermediates such as 1,2-dioxetanes, dioxetanones, peroxides, or endoperoxides, which decay to the ground state, and some of the energy is released as light. The majority of CL reactions are oxidation reactions because these reactions can provide sufficiently energetic intermediates for light emission (e.g. 63.5 kcal/mol for blue light at 450 nm). One of the best-known CL reactions is the oxidation of luminol (5-amino-2,3-dihydro-1,4-phthalazinedione): the decay of singlet excited-state 3-aminophthalate molecules produces a blue CL at 425 nm. Generally, CL reactions are inefficient in aqueous media (CL quantum yield for luminol 0.01; i.e. 1% efficiency) but are improved in aprotic solvents (e.g. dimethyl sulfoxide). Figure 1 illustrates the diversity of compounds that undergo CL reactions (see http://www.shsu.edu/~chm_tgc/chemilumdir/chemiluminescence2.html for movies illustrating chemiluminescent reactions).

2 Bioluminescence

Bioluminescence (BL) is the light emitted from living organisms ("living light"). Many thousands of BL species have been identified (666 genera from 13 phylla), and the firefly (*Photinus pyralis*) is the best-known example (Table 1). Light emission (562 nm) is produced in an ATP-dependent oxidation of firefly luciferin, catalyzed by firefly luciferase (EC 1.13.12.7) (Fig. 2). This highly efficient reaction has a quantum yield approaching unity (100% efficiency). However, the majority of BL reactions are considerably less efficient than the firefly reaction. The color of bioluminescence emissions ranges from violet (395 nm) in the sea coral Thourella, through blue (470 nm) in the jellyfish Periphylla, green (550 nm) in the glow worm Lampyris, to red (705 nm) in the deep-sea fish Malacosteus. Most BL reactions involve an enzyme (a luciferase) and a substrate (a luciferin). Luciferases and luciferins from different BL organisms have different structures and compositions: for example, marine bacterial luciferase is a heterodimer (40 and 37 kDa), and the luciferin is a long-chain aldehyde, whereas Renilla luciferase is a 35 kDa monomer, and the luciferin is an imidopyridazine derivative. Some BL organisms utilize a precharged photoprotein instead of a luciferase; for example, in the jellyfish Aequorea, BL is due to the reaction of calcium ions with aequorin, a stable complex of apoaequorin and coelenterazine (a luciferin) that has been precharged with an oxidant (molecular oxygen). The BL emission is due to an energy transfer to green fluorescent protein (GFP) from excited-state coelenterate oxyluciferin.

3
Electrochemiluminescence

Chemical reactions at the surface of an electrode can also produce excited-state species that decay to the ground state and emit light (electrochemiluminescence). Ruthenium(II) tris(bipyridyl) undergoes an electrochemical reaction with tripropylamine to produce a strongly reducing tripropylamine radical and a strongly oxidizing ruthenium(III) species at the electrode surface (Fig. 3). These react to regenerate ruthenium tris(bipyridyl) in an electronically excited state that decays to the ground state and emits light at 620 nm. Suitably derivatized ruthenium(II) tris(bipyridyl) molecules are effective labels (detection limit 200 fmol/L) in nucleic acid assays of different types.

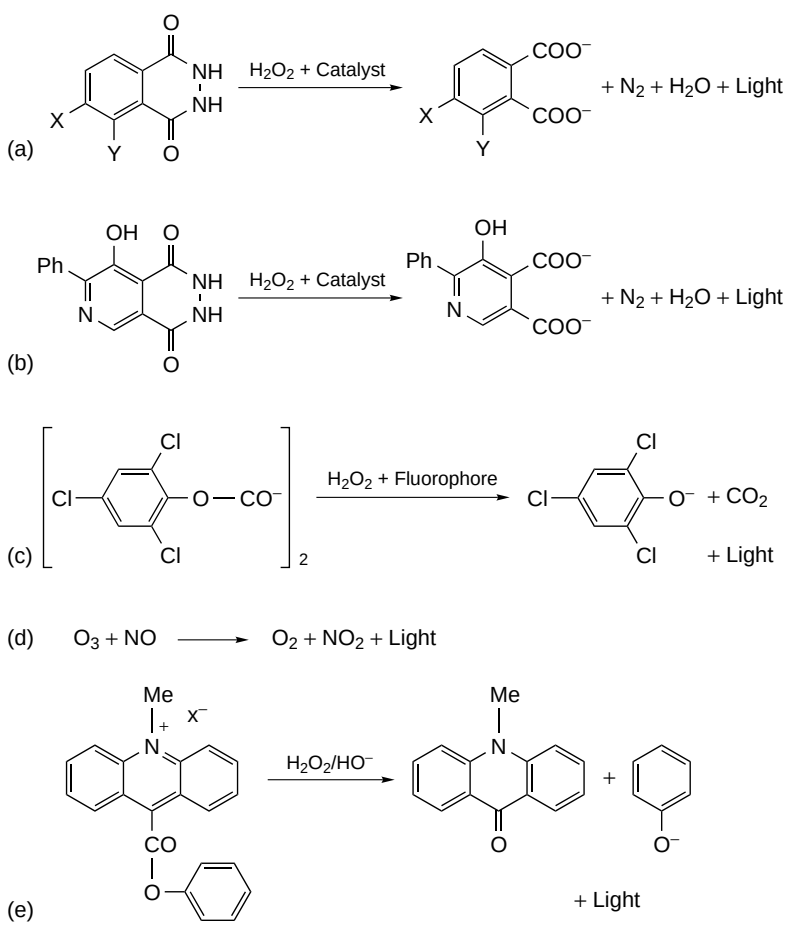

Fig. 1 Examples of chemiluminescent reactions. (a) Luminol X = H, Y = NH_2; isoluminol, X = NH_2, Y = H, (b) 8-hydroxy-7-phenyl-pyrido[3,4-d]pyridazine-1,4(2H, 3H) dione, (c) bis(2,4,6-trichlorophenyl) oxalate, (d) ozone–nitric oxide, (e) acridinium aryl ester, (f) CSPD, (g) siloxene, (h) lophine, and (i) lucigenin.

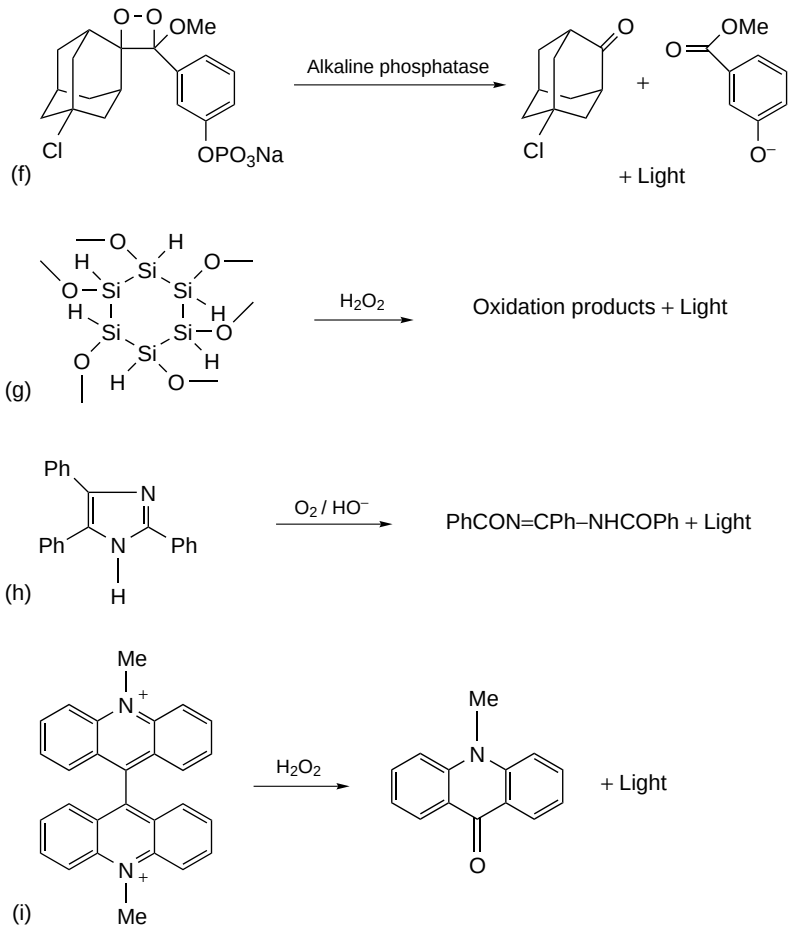

Fig. 1 (*Continued*).

4 Applications

Most assays simply involve mixing the sample with the appropriate CL or BL reagents and measuring or detecting the light emission. CL and BL reactions are sensitive and simple to perform; they utilize nonhazardous reagents, have a wide dynamic range (several orders of magnitude), and are usable in diverse applications (assays for proteins, DNA, RNA, enzymes, substrates, inhibitors, metal ions, and biomass) and in different assay formats (e.g. flow injection, liquid chromatography, extra-laboratory tests) (Table 2).

4.1 Measurement of Light Emission

Light can be measured quantitatively using a luminometer or qualitatively by means of photographic or X-ray film. Quantitative measurements are usually performed using a photomultiplier tube as the detector. This device can be operated in

Tab. 1 Examples of bioluminescent organisms.

Common name	*Phyllum*	*Systematic name*
Moon-night mushroom	Eumycota	*Pleurotus japonicus*
Foxfire		*Armillaria mellea*
Jack-o'-lantern		*Clitocybe illudens*
Sea fir	Cnidaria	*Obelia geniculata*
Sea pansy		*Renilla reniformis*
Sea comb	Ctenophora	*Mnemiopsis leidyi*
Firefly squid	Mollusca	*Watasenia scintillans*
Fireworm	Annelida	*Odontosyllis enopla*
Firefly	Arthropoda	*Photinus pyralis*
Glow worm		*Lampyris noctiluca*
New Zealand glow worm		*Arachnocampa luminosa*
Railroad worm		*Phrixothrix tiemann*
Sea firefly		*Vargula hilgendorfi*
Brittle star	Echinodermata	*Ophiopsila californica*
Angler fish	Chordata	*Melanocetus pelagicus*
Flashlight fish		*Photoblepharon palpebratus*
Hatchet fish		*Argyropelecus spp.*
Lantern fish		*Myctophum spinosum*
Midshipman fish		*Porichthys notatus*

(See http://lifesci.ucsb.edu/~biolum/organism/photo.html for pictures of some of the BL organisms listed in this table).

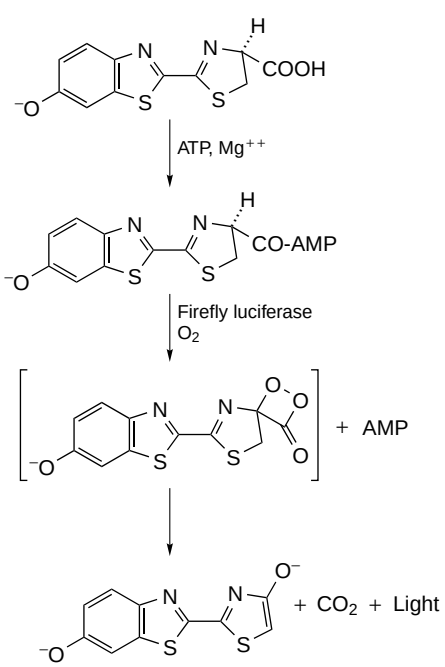

Fig. 2 Bioluminescent firefly luciferase reaction.

the photon-counting mode or the less-sensitive photon-current mode. An alternative is to use a charge-coupled device (CCD) camera to image the light emission from two-dimensional sources. Reaction vessels of different geometries (microtiter plates, membranes) can be imaged and the images stored and manipulated to improve sensitivity via background subtraction techniques. Typically, a CCD detector divides the image into a 512 × 512 array of elements (pixels), but improved resolution can be achieved using sensors that divide the image into a 1024 × 1024 pixel array. The kinetics of the CL or BL light emission is a major consideration in light measurement. Rapid light emission (a flash of light) requires that

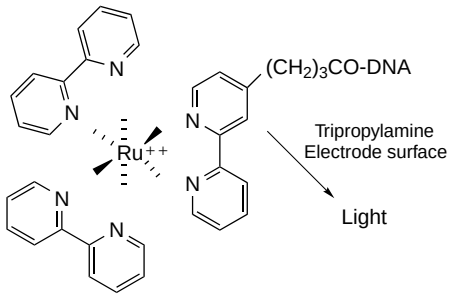

Fig. 3
Electrochemiluminescent ruthenium(II) tris(bipyridyl)–labeled DNA.

Tab. 2 Illustrative examples of applications of chemiluminescent and bioluminescent reactions.

Type of application or test	*Examples*
Antibodies	IgA, IgG, IgE, IgM
Antioxidants	Vitamin E
Biomass	Bacteriuria testing, enumeration of bacteria in milk, soil biomass, shelf-life dating, sperm viability
Chemosensitivity	Biocide assessment, tumor chemosensitivity
Drugs	Digoxin, methotrexate
Environmental	Phenolic herbicides, hygiene monitoring
Enzymes	Adenylate kinase, β-1,4-galactosyltransferase, creatine kinase, galactose 1-phosphate, *N*-acetyl-β-d-glucosaminidase, proteases
Gases	Nitric oxide, nitrogen dioxide, ozone, sulfur dioxide
Hormones	Follitropin, lutropin, thyrotropin, thyroxine, parathyrin
Lighting	Emergency and warning light, aerosols, and smoke
Metals	Ag, Bi, Ca, Ce, Co, Cr, Fe, Hg, K, Li, Mg, Mn, Ni, Os, Pb, Tb, Th, Tl, V, Zr
Oxidants	Hypobromite, hypochlorite, iodine, superoxide
Steroids	Estradiol, estriol, progesterone, testosterone
Substrates and cofactors	ADP, ATP, bile acids, ethanol, galactose, glucose, glucose 6-phosphate, glycerol, lactate, NADH, NADPH, oxaloacetate, phosphocreatine, pyruvate, sorbitol

the reaction be initiated in front of the photodetector using an injector. The wavelength of light emission is a further factor in selecting a detector. Photomultiplier tubes have a wavelength-dependent quantum efficiency for light detection and usually are most sensitive in the blue region (350–400 nm) and least sensitive in the red region (600 nm) of the spectrum.

4.2 Immunoassay, Blotting, and Nucleic Acid Hybridization Assays

The extreme sensitivity of CL and BL assays (detection limits in the attomole–zeptomole range: 10^{-18}–10^{-21} mol) has made them ideal for immunoassay and other applications (Table 3). CL and BL substances are

Tab. 3 Detection limits for labels used in immunoassay and nucleic acid hybridization assays

	Detection limits (zeptomoles: 10−21 mol)[a]	
Label	CL	BL
Acridinium ester	500	
Alkaline phosphatase	1	10
β-D-galactosidase	30	0.2
Glucose oxidase	6 200	
Glucose 6-phosphate dehydrogenase	3	1 000
Horseradish peroxidase	1	400
Isoluminol	50 000	
Xanthine oxidase	3 000	

CL, chemiluminescent assay; BL, bioluminescent assay.

effective alternatives to radioactive labels (^{32}P, ^{125}I), and generally, CL and BL reactions are superior to colorimetric and fluorescent reactions used to assay conventional enzyme labels.

4.2.1 Labels

Luminol, isoluminol, acridinium ester, acridinium sulfonyl carboxamide, firefly luciferase, marine bacterial luciferase, Renilla luciferase, Vargula luciferase, aequorin, and ruthenium(II) tris(bipyridyl) have all been tested as labels. Acridinium esters are one of the principal types of CL label in current use in immunoassay and hybridization assays. A rapid flash of light is produced by reacting the label with a mixture of H_2O_2 and NaOH. Nonseparation (homogeneous) nucleic acid assays can be formatted with an acridinium ester label, as in the "hybridization protection assay." In this three-step assay, first the sample and an acridinium ester–labeled probe are incubated together. Next, the label on an unhybridized probe is hydrolyzed to a nonchemiluminescent product (probe bound to target nucleic acid is protected from hydrolysis), and finally a chemiluminescent signal is obtained from the acridinium ester label on the bound probe by treatment with oxidant in basic conditions (Fig. 4). The recombinant photoprotein aequorin is the most promising BL label. Light emission (rapid flash) is triggered by reacting the aequorin with calcium ions.

4.2.2 Detection Reactions

The two most popular enzyme labels, alkaline phosphatase and horseradish peroxidase, can be quantitated by a range of CL and BL reactions. These enzymes are used in several ways: as direct labels for nucleic acid, indirectly as avidin or streptavidin conjugates with biotin-labeled nucleic acids, or via antigen–antibody coupling, such as anti-digoxigenin conjugates with digoxigenin-labeled nucleic acids.

For alkaline phosphatase, the CL assay using adamantyl 1,2-dioxetane aryl phosphate substrates (e.g. AMPPD®, CSPD®) (Fig. 5) is convenient and very sensitive (detection limit, 1 zmol). In membrane-based assays (e.g. Western

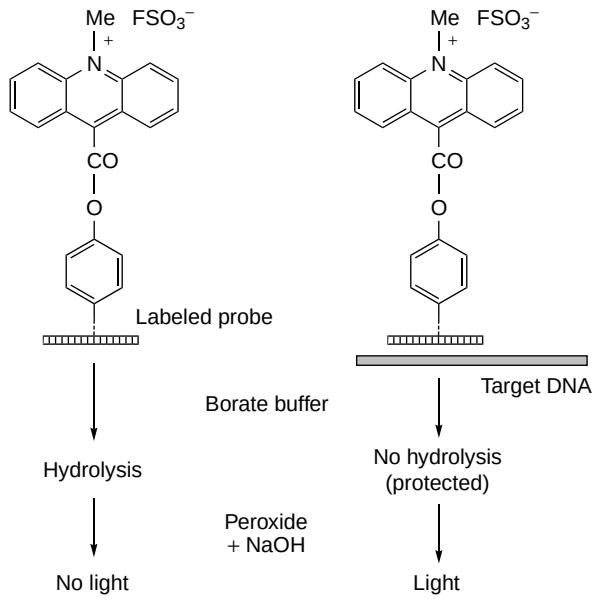

Fig. 4 Hybridization protection assay for detecting target nucleic acid (DNA) using a chemiluminescent acridinium ester label.

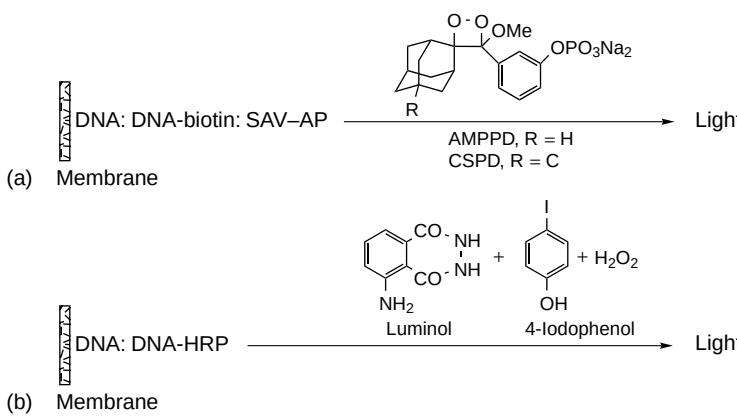

Fig. 5 Chemiluminescent detection of an alkaline phosphatase label using an adamantyl 1,2-dioxetane aryl phosphate (AMPPD®, CSPD®), and detection of a peroxidase label using the enhanced chemiluminescent reaction. AP, alkaline phosphatase; HRP, horseradish peroxidase; SAV, streptavidin.

blot, Southern blot), the nature of the membrane influences the CL signal from bound alkaline phosphatase. The surface of a nylon membrane acts as an enhancer of the CL signal because a dephosphorylated intermediate is sequestered in hydrophobic domains on the nylon. Similar improvements

in reaction kinetics, signal intensity, and signal-to-background ratio can be achieved both in solution and on membranes using certain water-soluble macromolecules [e.g. polyvinylbenzyl (benzyldimethylammonium) chloride]. Several highly sensitive sustained signal CL assays for horseradish peroxidase are available. These are based on (a) an acridan substrate that is converted by peroxidase into an unstable acridinium ester intermediate or (b) an enhanced chemiluminescent assay that uses luminol in conjunction with an enhancer molecule (e.g. 4-iodophenol, 4-hydroxycinnamic acid) (Fig. 5). In addition, sensitive CL or BL detection reactions of varying complexity are available for glucose oxidase, glucose 6-phosphate dehydrogenase, xanthine oxidase, and β-galactosidase labels (Table 4).

CL is also useful in DNA sequencing. For example, it is used to detect biotinylated DNA fragments transferred onto a membrane from a polyacrylamide sequencing gel. The biotin-labeled DNA is reacted with a streptavidin–alkaline phosphatase conjugate, and then the bound enzyme is detected using a 1,2-dioxetane substrate (e.g. CSPD®). Light emission from the membrane surface is imaged onto film or using a CCD camera as shown in Fig. 6.

4.3 Enzymes, Substrates, and Cofactors

The luminol, firefly luciferase, and marine bacterial luciferase reactions are sensitive indicator reactions for the production or consumption of peroxide, ATP, and NAD(P)H respectively. They can be coupled to other reactions involving oxidases, kinases, and dehydrogenases, and can be used to measure any component of the coupled reaction (enzyme, substrate, cofactor). Examples include the assays for alcohol and glucose, shown in Eqs. (1) and (2). Coimmobilization of a luciferase with other enzyme reagents improves the efficiency of the reaction because intermediates are effectively channeled to successive enzymes in the coupled enzyme system.

Tab. 4 Chemiluminescent and bioluminescent detection reactions for enzyme labels.

$$\text{Glucose} + O_2 \xrightarrow{\text{glucose oxidase label}} \text{gluconolactone} + H_2O_2$$

$$H_2O_2 + \text{isoluminol} \xrightarrow{\text{microperoxidase}} \text{light}$$

$$\text{Xanthine} + O_2 \xrightarrow{\text{xanthine oxidase label}} \text{hypoxanthine} + H_2O_2$$

$$H_2O_2 + \text{luminol} + \text{Fe-EDTA} \longrightarrow \text{light}$$

$$o\text{-nitrophenylgalactoside} \xrightarrow{\beta\text{-galactosidase label}} \text{galactose}$$

$$\text{galactose} + NAD^+ + \text{galactose dehydrogenase} \longrightarrow \text{galactonate} + \text{NADH}$$

$$\text{NADH} \xrightarrow{\text{NADH : FMN oxidoreductase} + \text{bacterial luciferase} + \text{FMN} + \text{decanal}} \text{light}$$

$$\text{glucose 6-phosphate} + NADP^+ \xrightarrow{\text{glucose 6-phosphate dehydrogenase label}} \text{glucose} + \text{NADPH}$$

$$\text{NADPH} \xrightarrow{\text{NADPH : FMN oxidoreductase} + \text{bacterial luciferase} + \text{FMN} + \text{decanal}} \text{light}$$

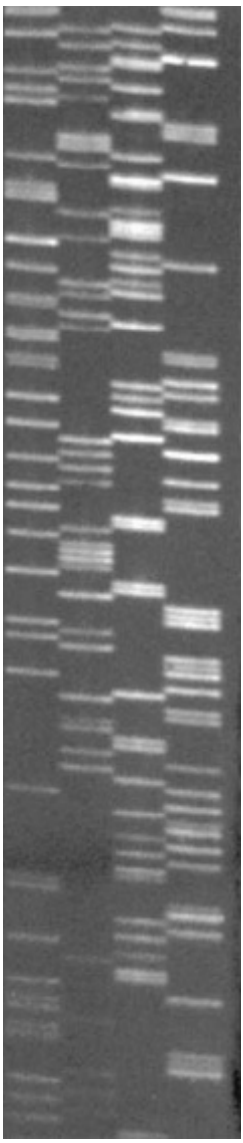

Fig. 6 Charge-coupled device camera image of a chemiluminescent DNA sequencing blot (M13mp18 single-stranded DNA). The image was obtained using a Photometrics Star I CCD camera. Biotinylated DNA sequencing reactions were loaded onto the gel in the following order (left to right): A, C, G, T. CL detection of bound streptavidin–alkaline phosphatase was performed using CSPD®. (Courtesy of Chris Martin, Tropix Inc., Bedford, MA.)

4.4 Reporter Genes

The chloramphenicol acetyltransferase gene (CAT) is used extensively as a reporter gene, but assay of the gene product is slow and requires the use of tritiated reagents.

$$\text{alcohol} + \text{NAD}^+ \xrightarrow{\text{alcohol dehydrogenase}} \text{acetaldehyde} + \text{NADH}$$

$$\text{NADH} + \text{FMN} \xrightarrow[\text{oxidoreductase}]{\text{NAD(P)H : FMN}} \text{NAD}^+ + \text{FMNH}_2$$

$$\text{FMNH}_2 + \text{decanal} + \text{O}_2 \xrightarrow[\text{luciferase}]{\text{marine bacterial}} \text{FMN} + \text{H}_2\text{O} + \text{decanoic acid} + \text{light} \quad (1)$$

$$\text{glucose} + \text{O}_2 \xrightarrow{\text{glucose oxidase}} \text{gluconolactone} + \text{H}_2\text{O}_2$$

$$\text{H}_2\text{O}_2 + \text{luminol} \xrightarrow{\text{peroxidase}} \text{3-aminophthalic acid} + \text{N}_2 + 2\text{H}_2\text{O} + \text{light} \quad (2)$$

The gene for firefly luciferase (*luc*) has been cloned and is an effective alternative to CAT as a reporter gene for studying transcriptional activity of cloned genomic sequences. Luciferase activity in transfected cells is measured by lysing the cells with a detergent (e.g. Triton X-100), adding ATP + Mg^{2+}, injecting firefly luciferin, and measuring the flash of light (detection

limit ca. 1×10^5 molecules). More recently, the luciferase detection reaction has been improved so that a constant light emission is obtained. This is achieved by adding to the assay mixture substances (e.g. coenzyme A, inorganic pyrophosphate, oxidized cytidine triphosphate) that minimize inhibition of the luciferase by oxyluciferin product molecules. The gene for Vargula luciferase, apoaequorin, and the fused marine bacterial luciferase *luxA* and *luxB* genes are also useful as reporter genes. Other alternatives include the CL assay (adamantyl 1,2-dioxetane aryl galactoside substrate, Fig. 7) of β-galactosidase expressed by the *lacZ* gene and the BL assay (firefly luciferin-*O*-phosphate substrate, Fig. 7) of the enzyme expressed by the alkaline phosphatase gene. BL genes are also effective for *in vivo* non-invasive imaging of whole organisms (e.g. plants, animals).

Membrane-permeable luciferin derivatives such as ethyl, hydroxyethyl, and hydroxypropyl esters (Fig. 7) permit non-destructive measurement of firefly luciferase reporter gene expression. These luciferin esters cross the cell membrane and react with intracellular esterases to produce luciferase. This enzyme then reacts with firefly luciferase to give light; the unreacted ester is not a substrate for the luciferase and does not contribute to the signal. Although not a BL reporter gene, the increasingly popular green fluorescent protein (GFP) is derived from a BL organism, the jellyfish *Aequorea victoria*. The advantage of this reporter is that the highly fluorescent expressed GFP can be detected directly using a suitable excitation light (ca. 470 nm) without the need to add reagent or disrupt the cells.

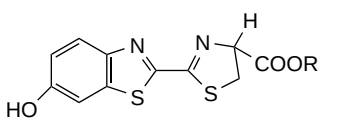

1, R = CH_2—CH_2OH
2, R = CH_2—CH_2—CH_2—OH
3, R = CH_2—CH_3

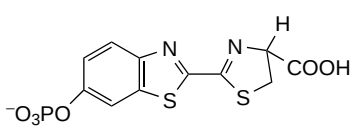

4

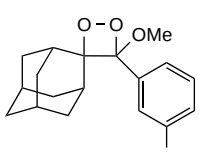

β-D-galactopyranoside

5

Fig. 7 Structure of three membrane-permeable firefly luciferin esters (1, hydroxyethyl; 2, hydroxypropyl; 3, ethyl), (4) bioluminescent alkaline phosphatase substrate firefly luciferin-*O*-phosphate, and (5) the chemiluminescent galactosidase substrate.

4.5 Rapid Microbiology

ATP (adenosine 5-triphosphate) is present in all living cells – a bacterium contains on average 1 fg of ATP – and measurement of ATP using the firefly luciferase reaction is a convenient means of enumerating cells, with a sensitivity of 10^5 colony-forming units per milliliter. The procedure entails first destruction of somatic ATP using the enzyme apyrase (EC 3.6.1.5), followed by lysis of intact cells with a detergent to release ATP, and finally BL assay of the released ATP using a mixture containing firefly luciferase, luciferin, and magnesium ions in the appropriate buffer.

4.6 Cellular Luminescence

Cellular CL is widely used for the study of phagocytosis. Activation of polymorphonuclear neutrophils by phagocytosable particles leads to the production of reactive oxygen species and weak chemiluminescence. Measurement of the weak CL is usually accomplished after enhancement by adding luminol or lucigenin.

4.7 Toxicity Testing

Drug toxicity and tumor chemosensitivity can be readily assessed via the bioluminescent ATP assay using the firefly luciferase reaction. Cell death or growth inhibition in the appropriate cell population that is due to the test drug is determined from the BL signal, which is proportional to the number of cells. Toxic substances can also be detected directly using luminescent bacteria. Contact with toxin inhibits light emission. When dark mutants of the BL marine bacteria *Photobacterium fischeri* are produced by mutagenesis with methyl methane sulfonate, for example, they react with mutagens, and the light-emitting properties of the bacterium are restored. Thus, a versatile and sensitive test for compounds with mutagenic properties is available.

See also Fluorescence Correlation Spectroscopy (FCS).

Bibliography

Books and Reviews

Bronstein, I., Edwards, B., Voyta, J.C. (1989) 1,2-Dioxetanes: novel chemiluminescent enzyme substrates. Applications in immunoassays, *J. Biolumin. Chemilumin.* **4**, 99–111.

Bronstein, I.B., Fortin, J., Stanley, P.E., Stewart, G.S.A.B., Kricka, L.J. (1994) Chemiluminescent and bioluminescent reporter gene assays, *Anal. Biochem.* **219**, 169–181.

Campbell, A.K. (1988) *Chemiluminescence*, Ellis Horwood, Chichester.

DeLuca, M., McElroy, W.D. (1986) *Bioluminescence and Chemiluminescence*, Methods in Enzymology, Vol. 133, Part B, Academic Press, San Diego, CA.

Fletcher, P., Andrew, K.N., Calokerinos, A.C., Forbes, S., Worsfold, P.J. (2001) Analytical applications of flow injection with chemiluminescence detection–a review, *Luminescence* **16**, 1–23.

Kricka, L.J. (1995) *Nonisotopic Probing, Blotting and Sequencing*, 2nd edition, Academic Press, San Diego, CA.

Kricka, L.J. (1991) Chemiluminescent and bioluminescent techniques, *Clin. Chem.* **37**, 1472–1481.

Pringle, M.J. (1993) Analytical applications of chemiluminescence, *Adv. Clin. Chem.* **30**, 89–183.

Stanley, P.E., Kricka, L.J. (Eds.) (2002) *Bioluminescence and Chemiluminescence: Progress and Current Applications*, World Scientific, Singapore.

Ziegler, M.M., Baldwin, T.O. (2000) *Bioluminescence and Chemiluminescence*, Methods in Enzymology, Vol. 305, Part C, Academic Press, San Diego, CA.

Primary Literature

Akhavan-Tafti, H., Arghavani, Z., DeSilva, R., Schoenfelner, B.A, Schaap, A.P. (1994) Applications of Lumigen PS-1 chemiluminescent substrate to immunoassays, in: Campbell, A.K., Kricka, L.J.P. Stanley, P.E. (Eds.) *Bioluminescence and Chemiluminescence Fundamentals and Applied Aspects*, John Wiley & Sons, Chichester, pp. 309–312.

Allen, R.C., Loose, L.D. (1976) Phagocytic activation of a luminol-dependent chemiluminescence in rabbit alveolar and peritoneal macrophages, *Biochem. Biophys. Res. Commun.* **69**, 245–252.

Baldwin, T.O., Berends, T., Bunch, T.A., Holzman, T.F., Rausch, S.K., Shamansky, L., Treat, M.L., Ziegler, M.M. (1984) Cloning

of the luciferase structural genes from *Vibrio harveyi* and expression of bioluminescence in Escherichia coli, *Biochemistry* **23**, 3663–3667.

Blasco, R., Murphy, M.J., Sanders, M.F., Squirrell, D.J. (1998) Specific assays for bacteria using phage mediated release of adenylate kinase, *J. Appl. Microbiol.* **84**, 661–666.

Bronstein, I., Voyta, J.C., Thorpe, G.H., Kricka, L.J., Armstrong, G. (1989) Chemiluminescent assay of alkaline phosphatase applied in an ultrasensitive enzyme immunoassay of thyrotropin, *Clin. Chem.* **35**, 1441–1446.

Chapple, I.L., Mason, G.I., Garner, I., Matthews, J.B., Thorpe, G.H., Maxwell, S.R., Whitehead, T.P. (1997) Enhanced chemiluminescent assay for measuring the total antioxidant capacity of serum, saliva and crevicular fluid, *Ann. Clin. Biochem.* **34**, 412–421.

DiLella, A.G., Hope, D.A., Chen, H., Trumbauer, M., Schwartz, R.J., Smith, R.G. (1988) Utility of firefly luciferase as a reporter gene for promoter activity in transgenic mice, *Nucleic Acids Res.* **16**, 4159.

Kain, S.R., Adams, M., Kondepudi, A., Yang, T.T., Ward, W.W., Kitts, P. (1995) Green fluorescent protein as a reporter of gene expression and protein localization, *Biotechniques* **19**, 650–655.

Kirchner, G., Roberts, J.L., Gustafson, G.D., Ingolia, T.D. (1989) Active bacterial luciferase from a fused gene: expression of a *Vibrio harveyi luxAB* translational fusion in bacteria, yeast and plant cells, *Gene* **81**, 349–354.

Kohen, F., Pazzagli, M., Kim, J.B., Lindner, H.R. (1980) An immunoassay for plasma cortisol based on chemiluminescence, *Steroids* **36**, 421–437.

Kricka, L.J., Wienhausen, G.K., Hinkley, J.E., De Luca, M. (1983) Automated bioluminescent assays for NADH, glucose 6-phosphate, primary bile acids, and ATP, *Anal. Biochem.* **129**, 392–397.

Marshall, J., Molloy, R., Moss, G.W., Howe, J.R., Hughes, T.E. (1995) The jellyfish green fluorescent protein: a new tool for studying ion channel expression and function, *Neuron* **14**, 211–215.

Njus, D., Baldwin, T.O., Hastings, J.W. (1974) A sensitive assay for proteolytic enzymes using bacterial luciferase as a substrate, *Anal. Biochem.* **61**, 280–287.

Simpson, J.S., Campbell, A.K., Ryall, M.E., Woodhead, J.S. (1979) A stable chemiluminescent-labelled antibody for immunological assays, *Nature* **279**, 646, 647.

Thore, A., Ansehn, S., Lundin, A., Bergman, S. (1975) Detection of bacteriuria by luciferase assay of adenosine triphosphate, *J. Clin. Microbiol.* **1**, 1–8.

Thorpe, G.H.G., Kricka, L.J., Moseley, S.B., Whitehead, T.P. (1985) Phenols as enhancers of the chemiluminescent horseradish peroxidase-luminol-hydrogen peroxide reaction: Application in luminescence monitored enzyme immunoassays, *Clin. Chem.* **31**, 1335–1341.

Weeks, I., Beheshti, I., McCapra, F., Campbell, A.K., Woodhead, J.S. (1983) Acridinium esters as high-specific-activity labels in immunoassay, *Clin. Chem.* **29**, 1474–1479.

Wood, K.V., de Wet, J.R., Dewji, N., DeLuca, M. (1984) Synthesis of active firefly luciferase by in vitro translation of RNA obtained from adult lanterns, *Biochem. Biophys. Res. Commun.* **124**, 592–596.

Chicken Genome

Cheryll Tickle[1] *and David W. Burt*[2]
[1] *University of Dundee, Dundee, UK*
[2] *Roslin Institute (Edinburgh), Midlothian, UK*

1 Importance of Chicken Genomics 537

2 Chicken Karyotype 538

3 Maps of the Chicken Genome 538

4 Genetic Linkage Maps 539

5 Physical Maps 540
5.1 Cytogenetic Maps 540
5.2 Radiation Hybrid Maps 542
5.3 Contig Maps 542

6 Genome Sequencing 544
6.1 Sequence Assembly of the Chicken Genome 544
6.2 Comparative Sequencing and Identification of Conserved Regions 545

7 ESTs 546

8 Perspectives 547

Bibliography 547
Books and Reviews 547
Primary Literature 548

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Keywords

BAC (Bacterial Artificial Chromosomes)
A cloning vector for genomic DNA fragments up to 200 kb in length.

Contig Map
A map of the genome based on overlapping genomic clones, usually cloned in either YAC or BAC clones.

EST (Expressed sequence tag)
Expressed sequence tag, a single pass short read of complementary DNA (cDNA) that is generated from a transcribed region of the genome.

FISH
Fluorescence *in situ* hybridization.

Genetic Map
A map of the genome based on the linkage of genetic markers.

Marker
A gene or an anonymous segment of DNA that exists in multiple forms or alleles that can be used as a point of reference when mapping the inheritance of a specific chromosomal segment or a genetic phenotype.

Physical Map
A map of the genome determined directly by various methods including cytogenetics, radiation hybrid analysis and sequencing.

Reference Mapping Population
A pedigreed family used to construct a genetic linkage map, usually within a consortium of laboratories using the same DNA samples and information database.

RH Mapping Panel
A set of radiation hybrid clones used to construct a radiation hybrid (RH) physical map of the genome, usually within a consortium of laboratories using the same DNA samples and information database.

YAC (Yeast Artificial Chromosomes)
A cloning vector for genomic DNA fragments up to 1 Mb in length.

Elucidation of the chicken genome has wide-ranging implications for agriculture and biology. Genetic mapping in chickens has a long history. The recent use of DNA markers has allowed the construction of a consensus genetic map containing 51 linkage groups. Chickens have 39 pairs of chromosomes including microchromosomes. Thirty-two of the linkage groups have now been assigned to specific chromosomes using fluorescence *in situ* hybridization to reveal the location of specific genes on the cytogenetic map. A detailed BAC map of the entire chicken genome is being assembled and will be merged with a whole genome shotgun sequence. The comprehensive collection of chicken expressed sequence tag ESTs will be indispensable for annotating the physical map of the genome. The chicken genome is particularly well placed in the evolutionary tree to provide new insights into vertebrate gene and genome evolution during the past 300 million years.

1 Importance of Chicken Genomics

Chicken genomics have wide-ranging applications because the chicken is important both from agricultural and biological viewpoints. Chickens are arguably the most important source of animal protein for the human population worldwide. The chicken is also the most widely studied avian species and as such provides a warm-blooded vertebrate model. Their relative cheapness and high reproductive rate make them relatively accessible for genetic approaches. Furthermore, because there are both similarities and differences between chickens and humans, genetic comparisons should be particularly illuminating.

From the agricultural viewpoint, the main issues have been to do with features such as growth, meat quality, and egg production, and breeding programs have been selected for these different traits leading to broiler (meat-type) and layer strains. More recently, other traits such as the genetic basis of disease susceptibility and those associated with behavior are coming to the fore. Chickens provide excellent models for investigating quantitative trait loci for these and other parameters that may also be relevant to mammals and even humans.

A surprising number of fundamental biological discoveries have been made in chickens in areas as far apart as virology, immunology, and developmental biology. It was in chickens, for example, that it was first discovered that viruses could cause tumors and that B cells are a distinct class of lymphocytes. The immune system of chickens differs from that of mammals in that antibody diversification is based on the high frequency of gene conversion that occurs at immunoglobulin genes rather than on combinatorial joining, and this provides a novel perspective.

Because of the high frequency of gene conversion at chicken immunoglobulin genes, a chicken cell line, DT40, derived from a chicken bursal lymphoma has proved to be a powerful tool for assessing the function of genes that act cell-autonomously. Homologous recombination between host genomic and transfected DNA in DT40 cells is as efficient as in budding yeast. These cells have been used to probe gene function in several cell biological processes, not only in antibody diversification but also, for example,

in DNA recombination and repair and in signal transduction.

Chickens are also one of the main vertebrate models studied by developmental biologists. A particular advantage is that experimental manipulations can be carried out while the embryo is still in the egg. Classical manipulations consisted of ablating and/or transplanting cells and tissues within chick embryos to study cell–cell interactions and to assay cell determination. A wide range of manipulations is now possible, including the construction of chick/quail chimeras to study cell fate, the grafting of microcarrier beads releasing defined chemicals, such as growth factors and inhibitors, and genetic manipulations. DNA can be readily introduced into chick embryos using retroviral methods and/or electroporation, and these techniques have been used widely to overexpress genes at particular times and locations during embryonic development. A recent report suggests that RNAi should also be applicable to chick embryos opening up the possibility of using chick embryos for high throughput screening of vertebrate gene function.

2
Chicken Karyotype

The chicken genome has a haploid content of 1.2×10^9 bp of DNA and this is divided among 39 chromosomes including not only macrochromosomes, which are cytologically distinct, but also microchromosomes, which are found in all birds and in some reptiles. Chromosomes 1 to 8 are usually classified as macrochromosomes. The sex chromosomes are Z and W; in birds, unlike in mammals, it is the males that are homogametic (ZZ), while the females are heterogametic (ZW). A recent study suggests that the long-standing view that dosage compensation does not occur in birds is incorrect and that the majority of Z-linked genes are dosage compensated. The 30 chicken microchromosomes contain about one-third of the genomic DNA but originally it was not known whether this was genetically inert. A number of studies have now shown that the microchromosomes are in fact gene rich, with recent estimates suggesting that microchromosomes are twice as gene rich as macrochromosomes.

3
Maps of the Chicken Genome

Maps are used in everyday life to find objects (houses, streets, towns, etc.) in different countries around the world. In the same way, gene maps for a particular species tell us the chromosome location of a specific gene. The resolution of any map can vary from simple views of whole counties, highlighting major towns and cities, to high-resolution maps showing individual houses and streets. The same is true of gene maps, from which we may simply know the approximate physical location on a chromosome, as observed in a microscope (a cytogenetic map), or we may actually have the nucleotide sequence of the gene.

There are essentially two types of gene mapping methods: genetic (or genetic linkage) and physical mapping. In genetic mapping, the order and position of genes along specific chromosomes is deduced from the cosegregation of genetic markers in genetic pedigrees. Genes close together on the same chromosome tend to stay together and those further apart may be separated by genetic recombination. The result is an abstract map of each chromosome based on the frequency of

recombination. The genetic map has a direct colinear relationship with the physical map, where genes are mapped directly by a number of methods (FISH (fluorescence *in situ* hybridization), RH (radiation hybrid) mapping, contig building, and DNA sequencing) as shown in Fig. 1. The chicken has a long genome mapping history, starting almost 100 years ago with the pioneering work on Mendelian inheritance and sex determination. The first genetic linkage map for any domesticated animal was also in the chicken.

4 Genetic Linkage Maps

There are two requirements to construct a genetic linkage map: (1) pedigrees in which the cosegregation of genes can be followed and (2) polymorphic genetic markers. Markers must be polymorphic, that is, alternative forms must exist so that their inheritance can be followed among the different members of the pedigree. In the past, genetic markers were often based on differences in morphology (e.g. feather color) or blood groups, but today most are based on DNA-sequence variants. A sequence change may alter the function or activity of a gene; for example, a sequence difference may produce an altered protein product or may alter the expression of the gene. In many cases, such changes may also result in a change in phenotype (e.g. defects in growth hormone receptor gene result in sex-linked dwarfism). However, the most frequent changes used to create gene maps are in noncoding regions such

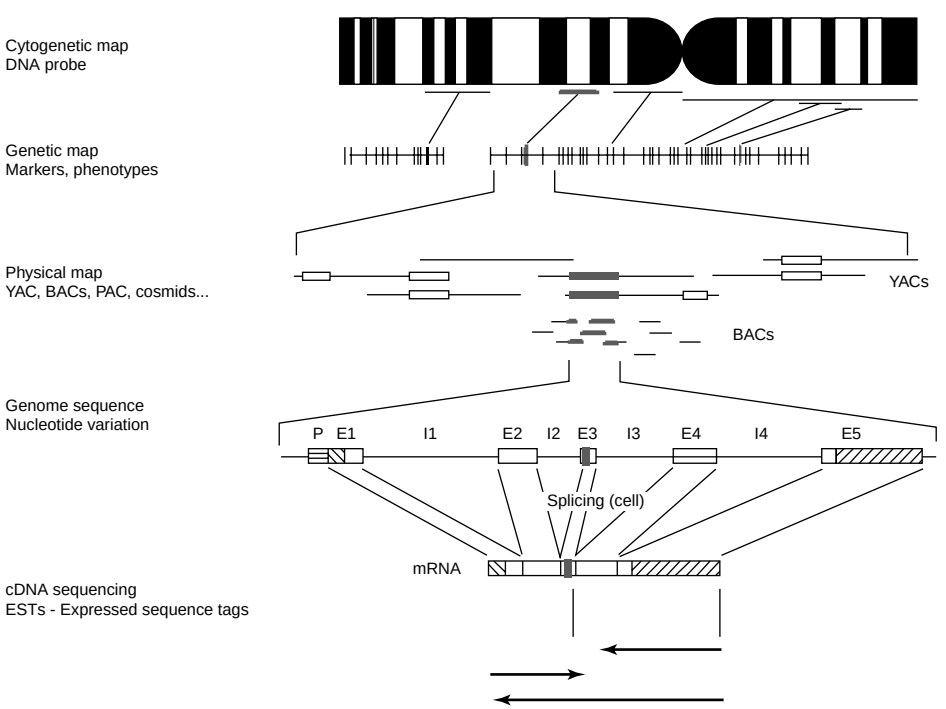

Fig. 1 Diagram showing relationship between genetic and physical maps and ESTs. Different methods of physical mapping are also illustrated.

as introns and may not have any effect on gene function. For example, a frequent genetic marker is based on variation in so-called *microsatellite sequences*. These are common and are based on simple sequence repeats (e.g. ACACACA...) that vary in length from individual to individual and may or may not be associated with a gene; they are therefore called *anonymous genetic markers*.

The first genetic linkage map for chickens based on DNA markers was published by Bumstead and Palyga in 1992 and was based on a cross between two partially inbred White Leghorn lines. Progress in creating a chicken genetic map (a consensus genetic map) accelerated when laboratories decided to pool efforts and concentrate on a few pedigrees (reference mapping families). The most fruitful pedigrees were the East Lansing and Wageningen populations. The East Lansing population consisted of only 52 backcross animals derived from a partially inbred jungle fowl and a highly inbred White Leghorn line. The resolution of this cross was low, between 5 and 7 cM. The Wageningen population consisted of 456 F_2 intercross animals derived from a cross between two broiler dam lines with origins from the White Plymouth Rock breed. With more progeny, the resolution of this cross was higher, about 1 cM. This is equivalent to about 300 kb or 5 to 10 genes. As an example, Fig. 2 shows the genetic map of chromosome 5 derived from the East Lansing and Wageningen reference mapping populations.

Early maps were based on various genetic markers (RFLP (restriction fragment length polymorphism), RAPD (randomly amplified polymorphic DNA), CR1-repeats and classical markers). Since then, most markers have been based on microsatellite sequences and AFLPs (amplified length fragment polymorphisms). Recently, more markers based on SNPs (single-nucleotide polymorphisms) have been used. Since many markers were in common between these various maps, it was possible to calculate a consensus map comprising 2012 loci and spanning 4000 cM. The consensus map still consists of 51 linkage groups, many probably representing the same microchromosome. So far, 32 of the linkage groups have been assigned to a specific chromosome by Griffin and all chromosomes now have a landmark probe.

5 Physical Maps

The resolution of physical mapping varies widely depending on the technique used, from 10 Mb in FISH mapping, 0.1 to 1 Mb in RH mapping, 0.1 to 1 Mb for BAC (bacterial artificial chromosome) and YAC (yeast artificial chromosome) physical maps, and to individual nucleotides in genome sequencing.

5.1 Cytogenetic Maps

The most reliable technique for *in situ* hybridization to reveal physical location on a chromosome has been FISH, where the DNA probe (e.g. a BAC clone) is labeled with a fluorescent dye. The location of the labeled probe is detected under the microscope after it has hybridized to its complementary DNA strand on a specific chromosome. Earlier experiments using radioactive *in situ* hybridization tended to be incorrect due to the technical challenges of the method. When combined with the availability of large genomic clones, FISH

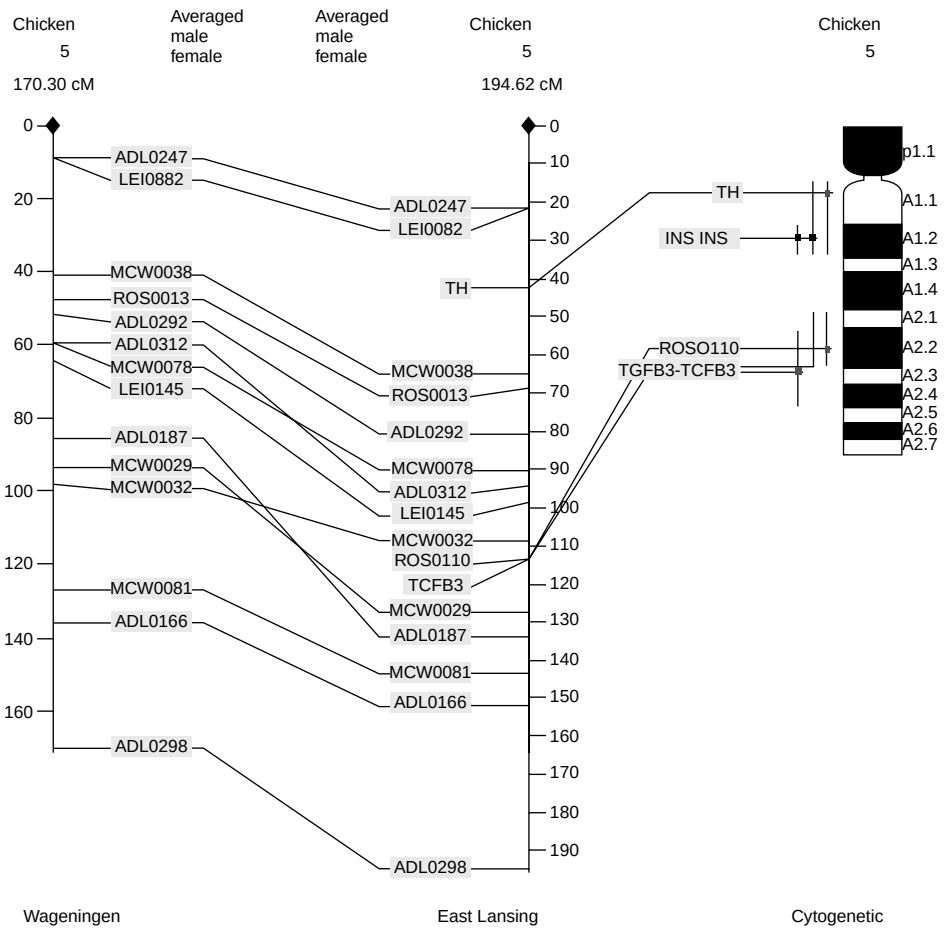

Fig. 2 Integration of genetic and physical maps of chromosome 5 (Arkdb-CHICK, June 2003).

has been a fast and simple method of mapping genes in recent years. The definition of a standard karyotype for the chicken by Ladjali-Mohammedi et al. has helped to exchange and compare data between different laboratories, in particular, for the larger chromosomes 1 to 8 and the sex chromosomes Z and W. Recent developments in the isolation of BAC clones and the isolation of individual microchromosomes have increased the possibility of a universal set of DNA or "Landmark probes" specific to each chromosome, including all 30 microchromosomes. In Fig. 3, chromosome paints derived from isolated chicken chromosomes have been used to identify homologous chromosomes in the turkey genome. To date, 218 genes have been assigned by FISH to a specific chromosome and 37 genes have been mapped to nonspecific microchromosomes. Through these experiments, genetic linkage maps have been assigned to specific chromosomes (e.g. Fig. 2 shows an integrated view of chromosome 5).

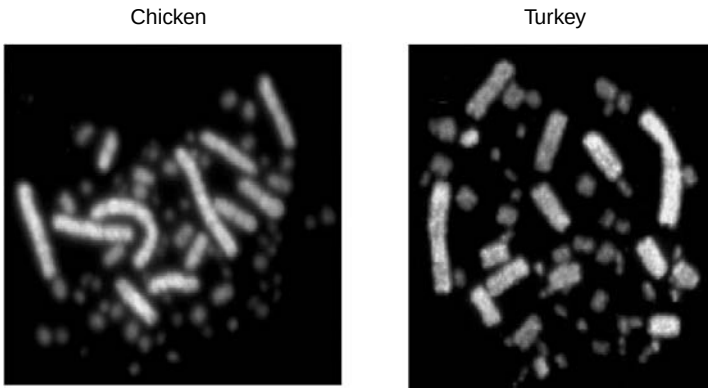

Fig. 3 Chromosome painting of turkey chromosomes with chromosome paints derived from chicken chromosomes 1 to 5, and the Z chromosome (kindly provided by Professor Johannes Wienberg). (See color plate p. xxi).

5.2 Radiation Hybrid Maps

A major limitation of genetic markers is the need to be polymorphic so that we can track their inheritance. About 10 years ago, the use of radiation hybrid mapping panels in humans and mice increased the rate of gene mapping significantly. In this method, the genome of the species to be mapped is fragmented by ionizing radiation. These DNA fragments are rescued by fusion with a host cell line to create panels of cells, each retaining a specific set of donor chromosomal fragments. Much like genetic mapping, genes close together on a specific chromosome will tend to be found in the same RH cell line. The frequency of coretention can be used to establish the order and distance between genes. In this method, since only the presence or absence of gene marker is required, there is no need to identify polymorphic markers; only a simple PCR test is required. In contrast to the success with other species, the construction of an RH panel in the chicken had met with limited success until recently. Early results for chromosome 7 mapped 77 loci, which are colinear with the corresponding genetic map. Currently, the chicken consortium are mapping over a thousand PCR-based anonymous and gene markers on this panel, which should prove to be an important tool to complement the BAC-based physical mapping.

5.3 Contig Maps

A contig map is constructed using a method that starts by fragmenting and cloning the genome in large segments of 100 to 1000 kb. Then various methods are used to identify overlapping genomic clones, and the data are finally assembled into a set of overlapping clones or contigs. Ideally, the process would end with a single contig for each chromosome, but usually the process defines hundreds of smaller contigs. The final number depends on the number of genomic clones analyzed (usually 100 000 clones) and the success rate for cloning difficult regions such as centromeres and telomeres.

Two types of large insert genomic libraries have been constructed in the chicken. The first was a YAC library that provided eightfold coverage of the chicken

genome. Such clones are difficult to maintain and analyze, so most contig-building projects nowadays use the simpler and more stable BAC system. In preparation for sequencing the chicken genome, many laboratories have constructed a number of BAC libraries with various enzyme combinations. The library by Crooijmans et al. was constructed with the restriction enzyme *Hind*III and contains 50 000 clones or fivefold coverage of the genome. Recently, three BAC libraries were constructed (*Hind*III, *Eco*RI, and *Bam*HI), each consisting of about 40 000 clones.

Methods used to create contigs vary and depend on using a simple and rapid method to identify a shared characteristic, such as the presence of a PCR product (STS or sequence tagged site), or more simply, an overlapping DNA fingerprint. These fingerprints can be generated in sophisticated ways (e.g. using AFLPs) or more simply, by restriction digests of the BAC clones (*Hind*III, *Bam*HI, *Eco*RI, etc.). Because the genomic inserts in these clones are large, many DNA fragments are observed for each clone after restriction digestion. The images are usually captured using imaging software and the results analyzed semiautomatically.

Currently, partial BAC contigs have been built for chicken chromosomes 10, 13, and 24, and when integrated with other data (sequence sampling data, genetic maps, RH maps, etc.), reveal gene maps of high resolution. Zhang et al.

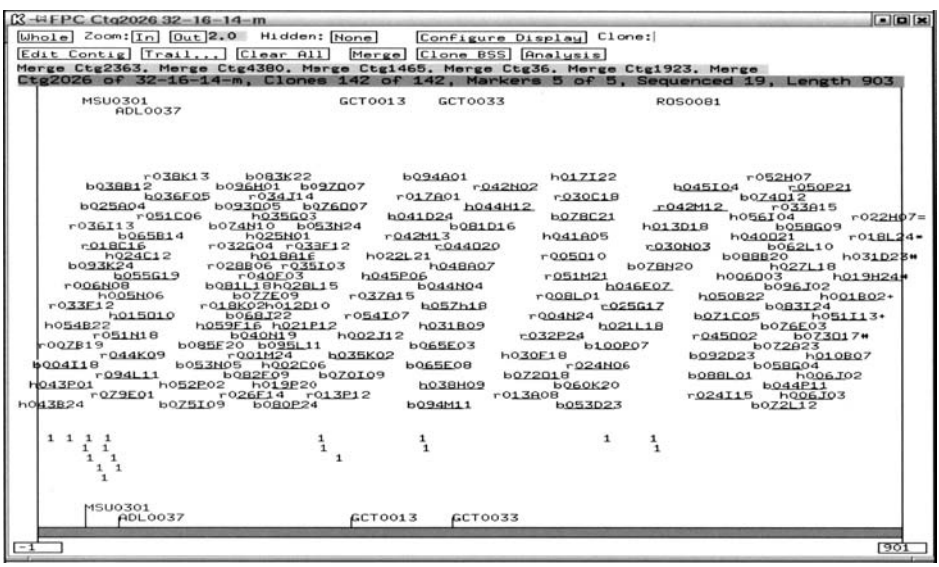

Fig. 4 Example of BAC contigs (ctg87, 1829 kb) for the chicken physical map. Example of chicken BAC contig anchored to the chromosome 1 genetic map of the chicken genome. This contig consists of 142 clones from three source BAC libraries, contains 903 unique fingerprint bands, and is estimated to span 4.01 Mb. The clones prefixed with h were from the HindIII BAC library, with b from the BamHI BAC library and with r from the EcoRI BAC library. The contig was anchored to the region around 361 cM of the chromosome 1 genetic map using five DNA markers – MSU0301, ADL0037, GCT0013, GCT0033, and ROS0081. The highlighted clones indicate the positive clones identified by DNA marker hybridization. (Reproduced by kind permission of Chengwei Ren, Mi-Kyung Lee, Bo Yan, Jerry B. Dodgon, and Hong-Bin Zhang).

(http://hbz.tamu.edu/) are fingerprinting over 100 000 chicken BAC clones (Fig. 4), which, when combined with the Wageningen BAC data, will result in a detailed BAC map of the entire chicken genome. This map will serve to order the whole genome shotgun (WGS) of the chicken genome.

6 Genome Sequencing

6.1 Sequence Assembly of the Chicken Genome

Owing to its interest as an important anchor species for evolutionary studies, there has been a rapid increase in genome sequencing of selected regions of the chicken genome. Such efforts will end with a nearly complete genome sequence in March 2004. Currently, there are 1811 entries in EMBL (June 2003) with information on gene structure in the chicken. In addition, there are a number of sequence sampling (BAC and cosmid clones) and BAC-end sequencing efforts. Of more interest are sequenced regions larger than 20 kb in length and since June 2003, the number in the nucleotide database is 133, of which 90% represent complete BAC clones. In general, these pilot genome-sequencing projects reveal that chicken genes share the same exon/intron structure as their mammalian orthologs. However, the genes in the chicken are more compact (2.5-fold), often with smaller introns. Also, the repeat content of the chicken is far less (10–15%) compared to mammals (50%), making WGS a practical option.

Extensive sequencing of large segments in the chicken has been limited: T-cell receptor beta-chain constant region (AF110982), T-cell receptor alpha chain constant region (U83833), beta-globin gene cluster (L17432), *MHC* complex (AL954802, AL023516), chemokine gene clusters, *SCL* locus, class II cytokine receptor gene cluster (AF082667), and to specific regions in the chicken orthologous to human chromosomes 7, 11, and 14. The general conclusions have been that the chicken regions tend to be two to three times smaller and that gene order is conserved over regions of at least 1 cM or 300 kb. The largest sequence is for a region orthologous to human chromosome 7q31.2. This sequence is 414 kb and covers 5 overlapping BAC clones. The order of the genes is identical (*TES, CAV1, MET, CAPZA2,* and *ST7*) to that found in the human genome, which covers a 1-Mb region. Another example is the 305-kb IGF2 region on chicken chromosome 5, also identical in gene order to its human counterpart (*IGF2, INS, TH, ASCL2,* and *CD81*).

The chicken genome-sequencing project will benefit from the experience gained from the human and mouse genome projects, and will include a mixture of targeted and whole genome sequencing. Currently, a physical map based on a 20-fold genome coverage of overlapping BAC clones is being assembled between laboratories at Washington University, Texas A&M, Michigan State University, and Wageningen University. This map will be integrated with other chicken mapping resources, including FISH mapped clones, genetic maps, and RH panels. Low sequence coverage of selected BAC clones, and BAC and Fosmid end sequences will be used to merge the BAC map with an eightfold whole genome shotgun. Preliminary assemblies based on 2.5 Mb of finished sequence have shown that this assembly approach should be very efficient

for the chicken. The high repeat content (~50%) in mammalian genomes has been a major problem with the WGS strategy; however, in the chicken, the repeat content is only 10 to 15%. This was shown dramatically in the repeat plots of finished genome sequences of human, mouse, and chicken. Also, more than 75% of WGS sequences from chicken BAC clones were contained in a single contig compared to only 25% in the mouse. The sequence will be based on a female (the heterogametic sex in chicken) inbred Jungle Fowl used to create the East Lansing genetic linkage map.

6.2 Comparative Sequencing and Identification of Conserved Regions

The chicken has often been used as an outgroup species in many evolutionary studies, in particular, in the study of the evolution of genes and genomes. Recently, there has been a rise in comparative genome analysis between long sequences in mammals (mostly human and mouse) and the chicken as a tool to identify conserved regions with the potential to function as regulatory sequences. This type of analysis began on a smaller scale in earlier gene sequencing studies in which conserved sequence motifs were found in both mammals and chicken. For example, it was identified that *NFκB, PEA1, PEA3*, and *C/EBP* transcription factor binding sites are associated with chicken inducible nitric-oxide synthase. It was shown that transcription binding sites in the 5′-flanking region of the *TGFB3* gene are conserved in chickens and humans. It was shown that the lens-preferred alphaA-crystallin gene contains a conserved stretch in the 5′-flanking region of the chicken and mouse genes. The 5′-half of this region has consensus binding sites for *AP-1* and other transcription factors. Furthermore, the mouse and chicken alphaA-crystallin genes are expressed with lens specificity using a similar assortment of transcription factors but with a different physical arrangement of their respective cis-elements within the promoter region.

A detailed study of the *SCL* locus compared 36 kb around the chicken locus with orthologous regions in human and mouse. In all the three species, the *SIL* and *MAP17* genes flanked the *SCL* gene. A detailed analysis also confirmed a number of previously characterized enhancer sequences and predicted a new enhancer 23-kb downstream of the *SCL* promoter start site. Another example is the comparison of the *alpha-globin* gene in a number of species (human, mouse, chicken, and Fugu) spanning 400 million years of evolution. In this study, several conserved motifs were identified in all the four species that suggested a common conserved function. A recent study on *SOX2* gene expression in the chick identified five functional-enhancer sequences in a 50-kb region flanking the chicken gene, after electroporation of embryos with various constructs. The exciting finding was that these functionally identified *SOX2* enhancers corresponded exactly to the extragenic sequence blocks conspicuously conserved between chicken and mammals, but which were not detected by sequence comparison among mammals. These large-scale studies require new bioinformatic tools such as *PipMaker*. An example of the type of analysis is shown in Fig. 5 in which a region on human chromosome 19p13.1 (contig NT_011295) is compared with orthologous regions on chicken and mouse chromosomes. The exciting thing about this is that the chick sequence is so well conserved

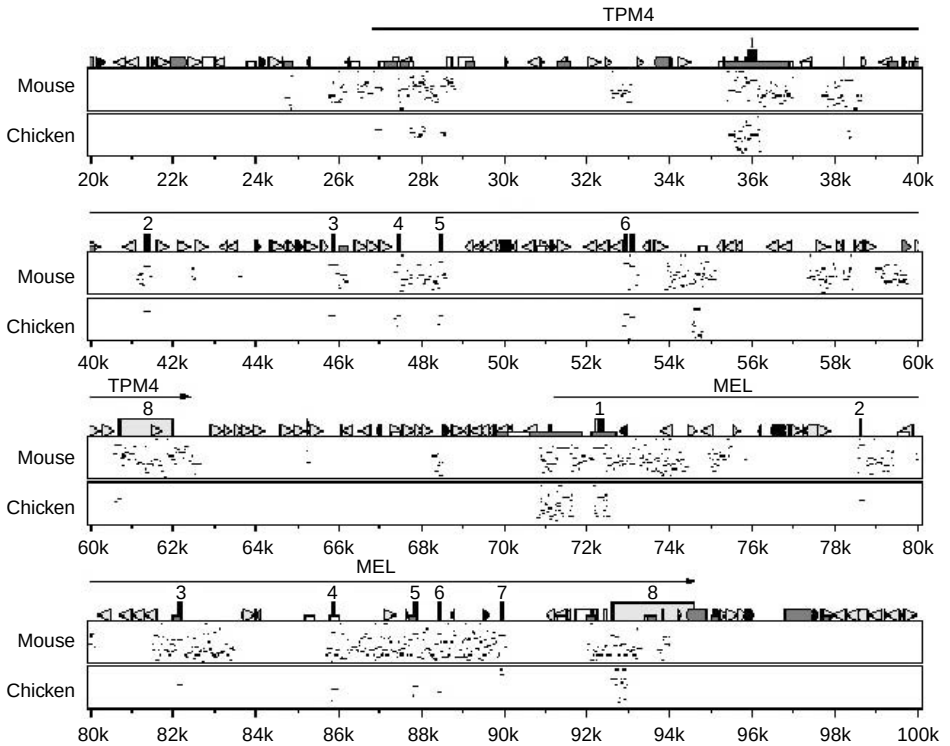

Fig. 5 Comparison of a segment of human chromosome 19p13.1 (contig NT_011295), with orthologous regions in the chicken and mouse genomes using the PipMaker program.

that most exons are visible as conserved sequences even at the nucleotide level.

7 ESTs

There is a large collection of chicken expressed sequence tags (ESTs) available (450 795 dbEST release 081503), which is an indispensable tool for annotating the physical map of the genome. EST programs include those carried out at the University of Delaware and the GSF (see www.chicken-genome.org for a complete summary). A large EST resource was funded in the United Kingdom by the BBSRC. This recent BBSRC project collected 340 000 ESTs generated from libraries taken from 21 different tissues, both adult and embryonic. Analysis of the data from this project gives an estimate of 35 000 chicken genes. Although the chicken genome is only about 40% of the size of the human genome, it is currently estimated to contain about the same number of genes. Furthermore, the microchromosomes (a third of the chicken genome) contain twice as many genes as the macrochromosomes. Thus, the chicken genome is compact.

Annotation of the EST data from the BBSRC project revealed that about 40% of clustered sequences have orthologs in known sequence databases. Furthermore, the chicken EST resource showed strong

matches to 90% of the human disease genes in the "morbid map" – a subset of human genes linked to known hereditary diseases. The EST resource is likely to contain additional chicken genes with, as yet, no known orthologs, or perhaps even genes that are avian specific. Ten thousand cDNAs from EST collections are currently being sequenced. These EST resources provide exciting opportunities for gene expression studies and for devising new tests for gene function using chick embryos.

8 Perspectives

During the past decade, the progress in chicken genomics has benefited from the sharing of data. ChickNET (http://dev.chicken-genome.org/) is one organization that tries to encourage this ethic. Using common resources, such as genetic mapping populations, genetic markers, physical probes, BAC libraries, cDNA and EST resources, RH mapping panels, microarrays, and the genome sequence, it will be possible to obtain different views of the chicken genome. A wealth of genome data for the chicken has already been produced using these resources and tools, and this will continue in the future. A number of genome mapping databases have been produced to store and manage such data. These databases include structural information, such as genetic maps, cytogenetic maps and banded karyotypes, RH maps, BAC and other clone-based maps, and the gene models based on EST sequences and other predictive tools. In addition, there will be the need to integrate more functional information, such as expression data (ESTs, *in situ* hybridization, etc.), mapping of genetic mutations, complex traits (QTL, Quantitative Trait Loci mapping) and phenotype using other tools (e.g. *RNAi*). Finally, such a genome view of the chicken needs to be compared and integrated with the views on other genomes, such as human, fish, and so on.

See also Genetics, Molecular Basis of; Genomic Sequencing (Core Article).

Bibliography

Books and Reviews

Bitgood, J.J., Somes, R.G. Jr. (1993) Gene Map of the Chicken (*Gallus gallus* or *G. domesticus*), in: O'Brien, S.J. (Ed.) *Genetic Maps*, 6th edition, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, pp. 4332–4342.

Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R., Wilson, S.A., Hubbard, S.J. (2002) A comprehensive collection of chicken cDNAs, *Curr. Biol.* **12**, 1965–1969.

Burt, D.W. (2003) Comparative Genomics, in: Muir, W.M., Aggrey, S.E. (Eds.) *Poultry Breeding and Biotechnology*, CABI International, Wallingford, Oxon, UK, pp. 563–606.

Burt, D.W., Pourquié, O. (2003) Chicken genome – science nuggets to come soon, *Science* **300**, 1669.

Groenen, M.A.M., Crooijmans, R.P.M.A. (2003) Structural Genomics: Integrating Linkage, Physical and Sequence Maps, in: Muir, W.M., Aggrey, S.E. (Eds.) *Poultry Breeding and Biotechnology*, CABI International, Wallingford, Oxon, UK, pp. 497–536.

Schmid, M., Nanda, I., Guttenbach, M., Steinlein, C., Hoehn, H., Schartl, M., Haaf, T., Weigend, S., Fries, R., Buerstedde, J.M., Wimmers, K., Burt, D.W., Smith, J., A'Hara, S., Law, A., Griffin, D.K., Bumstead, N., Kaufman, J., Thomson, P.A., Burke, T.A., Groenen, M.A.M., Crooijmans, R.P.M.A., Vignal, A., Fillon, V., Morisson, M., Pitel, F., Tixier-Boichard, M., Ladjali-Mohammedi, K.,

Hillel, J., Mäki-Tanila, A., Cheng, H.H., Delany, M.E., Burnside, J., Mizuno, S. (2000) First report on chicken genes and chromosomes 2000, *Cytogenet. Cell Genet.* **90**, 169–218.

Walter, M.A., Goodfellow, P.N. (1995) Irradiation and fusion gene transfer, *Mol. Biotechnol.* **3**, 117–128.

Primary Literature

Abdrakhmanov, I., Lodygin, D., Geroth, P., Arakawa, H., Law, A., Plachy, J., Korn, B., Buerstedde, J.M. (2000) A large database of chicken bursal ESTs as a resource for the analysis of vertebrate gene function, *Genome Res.* **10**, 2062–2069.

Bateson, W. (1902) Experimental studies with poultry, *Poult. Rep. Evol. Commun. R. Soc.* **1**, 87–124.

Bateson, W., Punnett, R.C. (1906) Experimental studies in the physiology of heredity, *Poult. Rep. Evol. Commun. R. Soc.* **3**, 11–30.

Bateson, W., Punnett, R.C. (1908) Experimental studies in the physiology of heredity, *Poult. Rep. Evol. Commun. R. Soc.* **4**, 18–35.

Brown, W., Hubbard, S.J., Tickle, C.A., Wilson, S.W. (2003) The chicken as a model for large-scale analysis of vertebrate gene function, *Nat. Rev. Genet.* **4**, 87–98.

Buerstedde, J.M., Takeda, S. (1991) Increased ratio of targeted to random integration after transfection of chicken B cell lines, *Cell* **67**, 179–188.

Buitenhuis, A.J., Crooijmans, R.P.M.A., Bruijnesteijn van Coppenraet, A., Veenedaal, M.A.M., Groenen, M.A.M., van der Poel, J.J. (2002) Improvement of the comparative map of chicken chromosome 13, *Anim. Genet.* **33**, 249–254.

Bumstead, N., Palyga, J. (1992) A preliminary linkage map of the chicken genome, *Genomics* **13**, 690–697.

Burt, D.W., Dey, B.R., Paton, I.R., Morrice, D.R., Law, A.S. (1995) The chicken TGF-beta3 gene: genomic structure, promoter analysis and chromosomal location, *DNA Cell Biol.* **14**, 111–123.

Burt, D.W., Bruley, C.K., Dunn, I., Jones, C.T., Ramage, A., Law, A.S., Morrice, D.R., Paton, I.R., Smith, J., Windsor, D., Sazanov, A., Fries, R., Waddington, D. (1999) Dynamics of chromosome evolution, clues from comparative gene mapping in birds and mammals, *Nature* **402**, 411–413.

Cheng, H.H., Levin, I., Vallejo, R.L., Khatib, H., Dodgson, J.B., Crittenden, L.B., Hillel, J. (1995) Development of a genetic map of the chicken with markers of high utility, *Poult. Sci.* **74**, 1855–1874.

Clarke, M.S., Edwards, Y.J.K., McQueen, H.A., Meek, S.E., Smith, S., Umrania, Y., Warner, S., Williams, G., Elgar, G. (1999) Sequence sampling chicken cosmids: a methodology for genome screening, *Gene* **227**, 223–230.

Crittenden, L.B., Provencher, L., Santangelo, L., Levin, I., Abplanalp, H., Briles, R.W., Briles, W.E., Dodgson, J.B. (1993) Characterisation of a Red Jungle Fowl by White Leghorn backcross reference population for molecular mapping of the chicken genome, *Poult. Sci.* **72**, 334–348.

Crooijmans, R.P.M.A., Dijkhof, R.J.M., Veenendaal, T., van der Poel, J.J., Nicholls, R.D., Groenen, M.A.M. (2001) The gene orders on human chromosome 15 and chicken chromosome 10 reveal multiple inter- and intrachromosomal rearrangements, *Mol. Biol. Evol.* **18**, 2102–2109.

Crooijmans, R.P.M.A., Van Oers, P.A.M., Strijk, J.A., Van Der Poel, J.J., Groenen, M.A.M. (1996) Preliminary linkage map of the chicken (*Gallus domesticus*) genome based on microsatellite markers: 77 new markers mapped, *Poult. Sci.* **75**, 746–754.

Crooijmans, R.P.M.A., Vrebalov, J., Dijkhof, R.J.M., van der Poel, J.J., Groenen, M.A.M. (2000) Two-dimensional screening of the Wageningen chicken BAC library, *Mamm. Genome* **11**, 360–363.

Elnitski, L., Riemer, C., Petrykowska, H., Florea, L., Schwartz, S., Miller, W., Hardison, R. (2002) PipTools: a computational toolkit to annotate and analyze pairwise comparisons of genomic sequences, *Genomics* **80**, 681–690.

Fillon, V., Morrison, M., Zoorob, R., Auffray, C., Douaire, M., Gellin, J., Vignal, A. (1998) Identification of 16 microchromosomes by molecular markers using two-colour fluorescent *in situ* hybridization (FISH), *Chromosome Res.* **6**, 307–314.

Flint, J., Tufarelli, C., Peden, J., Clark, K., Daniels, R.J., Hardison, R., Miller, W., Philipsen, S., Tan-Un, K.C., McMorrow, T., Frampton, J., Alter, B.P., Frischauf, A.M., Higgs, D.R. (2001) Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster, *Hum. Mol. Genet.* **10**, 371–382.

Gibbs, M., Dawson, D.A., McCamley, C., Wardle, A.F., Armour, J.A., Burke, T. (1997) Chicken microsatellite markers isolated from libraries enriched for simple tandem repeats, *Anim. Genet.* **28**, 401–417.

Göttgens, B., Barton, L.M., Gilbert, J.G., Bench, A.J., Sanchez, M.J., Bahn, S., Mistry, S., Grafham, D., McMurray, A., Vaudin, M., Amaya, E., Bentley, D.R., Green, A.R., Sinclair, A.M. (2000) Analysis of vertebrate SCL loci identifies conserved enhancers, *Nat. Biotechnol.* **18**, 181–186.

Griffin, D.K., Haberman, F., Masabanda, J., O'Brien, P., Bagga, M., Sazanov, A., Smith, J., Burt, D.W., Ferguson-Smith, M., Wienberg, J. (1999) Micro- and macro-chromosome paints generated by flow-cytometry and microdissection: tools for mapping the chicken genome, *Cytogenet. Cell Genet.* **87**, 278–281.

Groenen, M.A.M., Cheng, H.H., Bumstead, N., Benkel, B., Briles, E., Burt, D.W., Burke, T., Dodgson, J., Hillel, J., Lamont, S., Ponce de Leon, F.A., Smith, G., Soller, M., Takahashi, H., Vignal, A. (2000) A consensus linkage map of the chicken genome, *Genome Res.* **10**, 137–147.

Herbergs, J., Siwek, M., Crooijmans, R.P.M.A., Van der Poel, J.J., Groenen, M.A.M. (1999) Multicolour fluorescent detection and mapping of AFLP markers in chicken (*Gallus domesticus*), *Anim. Genet.* **30**, 274–285.

Hu, J., Mungall, C., Law, A.S., Papworth, R., Nelson, J.P., Brown, A., Simpson, I., Leckie, S., Burt, D.W., Hillyard, A.L., Archibald, A. (2001) The ARKdb, genome databases for farmed and other animals, *Nucleic Acids Res.* **29**, 106–110.

Hu, W., Myers, C., Kitzer, J., Plaff, S., Bushman, F. (2002) Inhibition of retroviral pathogenesis by RNA interference, *Curr. Biol.* **12**, 1301–1311.

Hughes, A.L., Hughes, M.K. (1995) Small genomes for better flyers, *Nature* **377**, 391.

Hutt, F.B. (1936) Genetics of the fowl. VI. A tentative chromosome map, *Neue Forschung für Tierzucht und Abstammung. (Duerst Festschrift)*, pp. 105–112.

Ilagan, J.G., Cvekl, A., Kantorow, M., Piatigorsky, J., Sax, C.M. (1999) Regulation of alphaA-crystallin gene expression. Lens specificity achieved through the differential placement of similar transcriptional control elements in mouse and chicken, *J. Biol. Chem.* **274**, 19973–19978.

Jennen, D.G.J., Crooijmans, R.P.M.A., Kamps, B., Acar, R., Veenendaal, A., Van der Poel, J.J., Groenen, M.A.M. (2002) A comparative map of chicken chromosome 24 and human chromosome 11, *Anim. Genet.* **33**, 205–210.

Kaufman, J., Milne, S., Gobel, T.W., Walker, B.A., Jacob, J.P., Auffray, C., Zoorob, R., Beck, S. (1999) The chicken B locus is a minimal essential major histocompatibility complex, *Nature* **401**, 923–925.

Knorr, C., Cheng, H.H., Dodgson, J.B. (1999) Application of AFLP markers to genome mapping in poultry, *Anim. Genet.* **30**, 28–35.

Kumar, S., Hedges, B.A. (1998) Molecular timescale for vertebrate evolution, *Nature* **392**, 917–920.

Kwok, C., Korn, R.M., Davis, M.E., Burt, D.W., Critcher, R., McCarthy, L., Paw, B.H., Zon, L.I., Goodfellow, P.N., Schmitt, K. (1998) Characterization of whole genome radiation hybrid mapping resources for non-mammalian vertebrates, *Nucleic Acids Res.* **26**, 3562–3566.

Ladjali-Mohammedi, K., Bitgood, J.J., Tixier-Boichard, M., Ponce de Leon, F.A. (1999) International system for standardized avian karyotypes (ISSAK): standardized banded karyotypes of the domestic fowl (*Gallus domesticus*), *Cytogenet. Cell Genet.* **86**, 271–276.

Law, A.S., Archibald, A.L. (2000) Farm animal genome databases, *Briefings Bioinf.* **1**, 151–160.

Lee, M.K., Ren, C.W., Yan, B., Cox, B., Zhang, H.B., Romanov, M.N., Sizemore, F.G., Suchyta, S.P., Peters, E., Dodgson, J.B. (2003) Construction and characterization of three BAC libraries for analysis of the chicken genome, *Anim. Genet.* **34**, 151–152.

Levin, I., Crittenden, L.B., Dodgson, J.B. (1993) Genetic map of the chicken Z chromosome using random amplified polymorphic DNA (RAPD) markers, *Genomics* **16**, 224–230.

Levin, I., Santangelo, L., Cheng, H.H., Crittenden, L.B., Dodgson, J.B. (1994) An autosomal linkage map of the chicken, *J. Hered.* **85**, 79–85.

Lin, A.W., Chang, C.C., McCormick, C.C. (1996) Molecular cloning and expression of an avian macrophage nitric-oxide synthase cDNA and the analysis of the genomic 5′-flanking region, *J. Biol. Chem.* **271**, 11911–11919.

McQueen, H.A., McBride, D., Miele, G., Bird, A.P., Clinton, M. (2001) Dosage compensation in birds, *Curr. Biol.* **11**, 253–257.

McQueen, H.A., Fantes, J., Cross, S.H., Clark, V.H., Archibald, A.L. (1996) CpG islands of chicken are concentrated on microchromosomes, *Nat. Genet.* **12**, 321–324.

Morisson, M., Lemiere, A., Bosc, S., Galan, M., Plisson-Petit, F., Pinton, P., Delcros, C., Feve, K., Pitel, F., Fillon, V., Yerle, M., and Vignal, A. (2002) ChickRH6: a chicken whole-genome radiation hybrid panel, *Genet. Sel. Evol.* **34**, 521–533.

Reboul, J., Gardiner, K., Monneron, D., Uze, G., Lutfalla, G. (1999) Comparative genomic analysis of the interferon/interleukin-10 receptor gene cluster, *Genome Res.* **9**, 242–250.

Smith, J., Bruley, C.K., Paton, I.R., Dunn, I., Jones, C.T., Windsor, D., Morrice, D.R., Law, A.S., Masabanda, J., Sazanov, A., Waddington, D., Fries, R., Burt, D.W. (2000) Differences in gene density on the chicken macrochromosomes and microchromosomes: a tool for gene discovery in vertebrate genomes, *Anim. Genet.* **31**, 96–103.

Tirunagaru, V.G., Sofer, L., Cui, J., Burnside, J. (2000) An expressed sequence tag database of T-cell-enriched activated chicken splenocytes: sequence analysis of 5251 clones, *Genomics* **66**, 144–151.

Toye, A.A., Schalkwyk, L., Lehrach, H., Bumstead, N. (1997) A yeast artificial chromosome (YAC) library containing 10 haploid chicken genome equivalents, *Mamm. Genome* **8**, 274–276.

Uchikawa, M., Ishida, Y., Takemoto, T., Kamachi, Y., Kondoh, H. (2003) Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals, *Dev. Cell.* **4**, 509–519.

Wang, K., Gan, L., Kuo, C.L., Hood, L. (1997) A highly conserved apoptotic suppressor gene is located near the chicken T-cell receptor alpha chain constant region, *Immunogenetics* **46**, 376–382.

Chimpanzee Genome

Ingo Ebersberger
Institute for Bioinformatics, Heinrich-Heine-University Düsseldorf, Germany

1 Human and Great Ape Phylogeny 553
1.1 Phylogenetic Analyses Based on Protein Comparisons 555
1.1.1 Immunological Comparison of Proteins 555
1.1.2 Electrophoretic Comparison of Proteins 555
1.1.3 Amino Acid Sequence Comparison 556
1.2 Phylogenetic Analyses Based on DNA Sequence Comparisons 556
1.2.1 Comparative DNA–DNA Hybridization 557
1.2.2 Mitochondrial DNA 558
1.2.3 Nuclear DNA 558
1.3 Genetic Diversity 559

2 Comparison of the Chimpanzee and Human Genomes 561
2.1 Cytogenetic Differences 561
2.2 Subchromosomal Rearrangements and Repeat Content 562
2.3 Substitutional DNA Sequence Differences 563
2.4 Insertions and Deletions 564

3 Chimpanzees in Evolutionary Research 565
3.1 DNA Sequence Evolution 565
3.2 From Genotype to Phenotype 567
3.3 Comparative Analysis of Gene Expression 568

4 Chimpanzees in Biomedical Research 569
4.1 Are Chimpanzees Good Model Organisms? 569
4.2 Biomedical Differences Between Humans and Chimpanzees 570

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

4.3 Chimpanzees and Infectious Diseases 571
4.4 Ethical Considerations 572

5 The Chimpanzee Genome Project 573
5.1 The Quest for Differences 573
5.2 Sequencing Strategy and Requirements 573
5.3 Human, Chimpanzee, and... Future Perspectives 575

Bibliography 576
Books and Reviews 576
Primary Literature 576

Keywords

Genetic Distance
An estimate of the amount of nucleotide substitutions between species. Genetic distance serves as a measure of the time elapsed since two species last shared a common ancestor.

Great Apes
Collective name for the orangutans (*Pongo pygmaeus*) from Borneo and Sumatra and for the African apes: chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), and gorillas (*Gorilla gorilla*).

Lineage Sorting
The distribution of alleles to the daughter populations upon a split of the ancestral population, for example, during speciation.

Monophyly
A group of species in a phylogenetic tree is monophyletic if all included species are descendants from a common ancestor to which no other species outside the group can be traced. Monophyletic groups are also referred to as clades or branches of a phylogenetic tree.

Phenotype
The complete set of organizational properties of an individual. It resembles the expressed information encoded in the genotype.

Phylogeny
A hypothesis about the evolutionary relationships of species expressed in a tree. All parts of a phylogenetic tree are inferred except the tips, which represent extant species or fossil remains.

On the basis of comparative molecular studies, chimpanzees are the closest living relatives of humans. On the level of their genomes, this is reflected in a nearly identical chromosomal organization and an average DNA sequence difference only about 10 times higher than that between any two humans. This renders the comparison of chimpanzees and humans, and the resulting catalog of their genetic differences, suitable to gain relevant insights into human evolution and human disease. The collection of genetic differences between both species as a whole comprises an ideal data set to analyze how DNA sequences change over time on a high-resolution scale. Of further interest are those differences that are located in the functional regions of the genome. They form the genetic basis of the distinct biological properties of humans and chimpanzees and thus might bear an answer to the question "What makes us human?" More practically, these differences provide access to the molecular factors that account for the different disease spectra in humans and chimpanzees, and might, therefore, open up new therapeutic approaches to fight human disease.

The relevance of chimpanzees as a subject in evolutionary and biomedical research is obvious. However, their close relationship to humans demands ethical guidelines that clearly distinguish chimpanzees from common genetic model organisms in laboratory research.

1 Human and Great Ape Phylogeny

For more than a century, scientists have been busy reconstructing the phylogenetic relationship of humans and the great apes. On the basis of the presence of shared morphological or molecular characters, individual species can be grouped to clades of an evolutionary tree. At the end of the nineteenth century, morphological evidence led to the generally accepted view that the great apes together with gibbons (*Hylobates* sp. and *Symphalangus* sp.) and humans share a common ancestor from which no other living organism descended. In phylogenetic terms, they represent the monophyletic group of hominoids within the primates (Fig. 1).

While this view has remained unchallenged since then, the relative order in which the lineages to the individual hominoid species emerged – often referred to as the "branching order" – provoked intense debate. From the shared consensus that the gibbon lineage diverged first, several competing hypotheses were proposed concerning the branching order of the remaining species. According to one favored scenario, the African apes (chimpanzees, bonobos, and gorillas) branched off second, allying humans with orangutans (Fig. 2a). An alternative view already taken by Darwin and Huxley positions the African apes closer to humans than the orangutans. Obviously, this hypothesis requires a further decision whether the closest relatives of chimpanzees and bonobos are the gorillas – as most morphologists believed – (Fig. 2b) or whether either of the species is closer to humans (Fig. 2c, d). The limited number of phylogenetic informative morphological characters as well as their sometimes

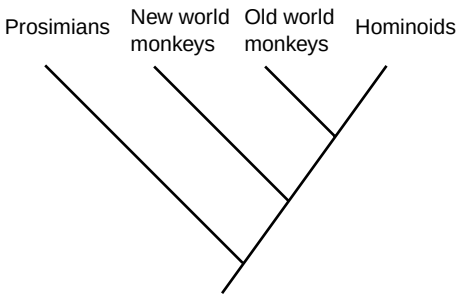

Fig. 1 A phylogenetic tree for the anthropoid primates. The prosimians combine the Lemur species from Madagascar, the nocturnal Lorises from South and Southeast Asia and from Africa, and the Tarsiers from certain islands of Southeast Asia. Generally, prosimians seem to be more ancestral in their morphology compared to the other three groups. The New World monkeys are confined to South and Central America and are, with one exception, the owl monkey, active during the day. They differ from the third group, the Old World monkeys typically in having a flat nose and separate nostrils. Old World monkeys such as rhesus macaques and baboons are found in Africa as well as in certain parts of Asia. The hominoids combine the gibbons (Asia), the orangutans (Asia), the African apes (gorillas, chimpanzees, bonobos), and humans.

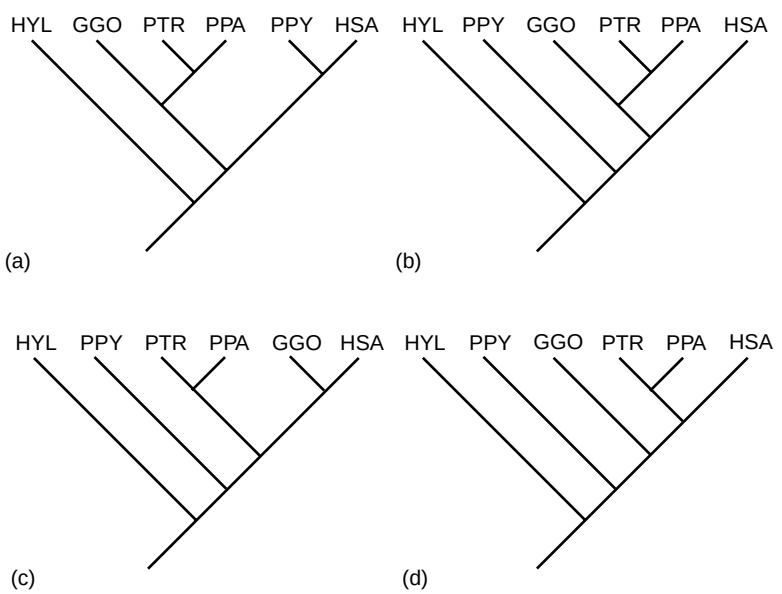

Fig. 2 Alternative phylogenies of the hominoids. HYL: gibbons; PPY: orangutans; GGO: gorillas; PTR: chimpanzees; PPA: bonobos; HSA: humans.

problematic interpretation rendered morphologists unable to conclusively solve the question of hominoid phylogeny.

In the second half of the twentieth century, the expansion of molecular biology provided a new and presumably more powerful approach to phylogenetic studies. The genetic distance between species as reflected in the amount of structural differences between their proteins and nucleic acids was introduced as a more objective measure of their evolutionary relationship. Under the assumption that the accumulation of such differences behaves like a clock

that is synchronized in all species under consideration, those species with a smaller genetic distance are more closely related.

1.1 Phylogenetic Analyses Based on Protein Comparisons

1.1.1 Immunological Comparison of Proteins

The first attempt to resolve primate phylogeny using molecular evidence was based on immunological comparisons of their serum proteins. As the underlying principle, an antiserum against a protein from species X is obtained by immunizing, for example, rabbits or chickens. The degree of similarity between the protein of species X and the corresponding protein in a second species Y is then determined by the extent of cross-reactivity of anti-X-serum and Y-protein. A quantitative measure of similarity is achieved since the antiserum comprises a mixture of antibodies, each recognizing a distinct surface structure (epitope) of the protein used for immunization. The more closely two species are related, the more similar are their proteins and thus, the more epitopes are shared among them. As a consequence, the extent of cross-reactivity increases.

Initial immunological comparisons of primate proteins used antisera against whole blood serum, a rather poorly defined protein mixture. Their results supported the rough classification of anthropoid primates into the major groups already suggested by morphologists (Fig. 1). However, no further resolution was possible since the antigenic properties of hominoid blood sera are too similar to each other to be differentiated by these means.

In later years, increasing understanding of the molecular processes of immune response in combination with the ability to generate antisera against individual proteins facilitated the development of more sensitive immunological assays. On the basis of such improved methods, the African apes and not orangutans were assigned as the closest relatives of humans. But subsequent determination of the branching order of the individual African ape species and humans failed for mainly the following reason. For proteins that differ only slightly, the correlation between the observed immunological difference and the genetic distance is not absolute. As an example, protein A from species X differs from the corresponding protein in species Z in three positions. By chance, none of these differences affect the antigenic properties of the respective proteins; therefore, no difference will be detected in an immunological assay with anti-A_X-serum. Protein A from a third species, Y, however, differs only in a single position from A_X which, again by chance, affects an epitope recognized by the anti-A_X-serum. The resulting lower cross-reactivity with anti-A_X-serum erroneously suggests a higher genetic distance between species Y and X than between species Z and X and thus, a grouping of X and Z to the exclusion of Y. This fuzziness in combination with a substantial experimental effort for the individual assays renders immunological comparisons not suitable to infer the evolutionary relationships of closely related species.

1.1.2 Electrophoretic Comparison of Proteins

Proteins contain amino acids with basic or acidic side chains and thus are charged molecules under most pH-conditions. This circumstance is exploited to separate proteins on a matrix according to their size and their net charge by application of

an electric field. Additional resolution can be achieved by a preceding separation of the proteins according to their isoelectric point – the pH where the net charge is 0 (2D electrophoresis). Variants of a particular protein that differ in their net charge because of sequence differences involving charged amino acids can be identified by their different mobility on the matrix. The application of protein electrophoresis in phylogenetic analyses is based on the fact that the degree of difference in the mobility pattern between related proteins approximately reflects their extent of amino acid sequence difference. Thus, similar to immunological comparisons, protein electrophoresis provides an estimate of the genetic distance between species that forms the basis of phylogenetic conclusions.

The most extensive study to reconstruct the phylogeny of hominoids by means of protein electrophoresis was based on a 2D electrophoresis of 383 fibroblast proteins. Concordant to conclusions from immunological distances of the hominoids, it supports an early separation of the gibbons followed by the lineage leading to orangutans. Notably, the study suggested a smaller genetic distance between humans and chimpanzees than between either species and gorillas. This lends support to the model that humans are the closest relatives to chimpanzees and bonobos (Fig. 2d).

However, similar to the immunological comparison of proteins, electrophoretic analyses can detect only a certain subset of amino acid differences between proteins. Only about 30% of all amino acid substitutions alter the charge and, therefore, the mobility pattern of a protein during electrophoresis. For closely related species, this poses the same resolution problem as already discussed for the immunological comparison of proteins.

1.1.3 Amino Acid Sequence Comparison

The direct comparison of their amino acid sequences represents the most accurate way to assess the degree of difference between two proteins. Even though experimentally elaborative, it provides the opportunity to also trace those changes in the protein sequence that neither affect the immunological properties nor the charge of a protein.

Despite this better resolution, the phylogenetic insights from comparative sequence analyses of hominoid proteins are limited. A data set of 1271 amino acid positions from 9 different polypeptides revealed an average similarity between human and chimpanzee proteins of 99.6%. A similar analysis with a smaller data set comparing humans, chimpanzees, and gorillas determined the average amino acid sequence similarity between humans and gorillas and between chimpanzees and gorillas to be slightly lower (99.3%). Even though this implies a closer relationship between humans and chimpanzees to the exclusion of gorillas, the number of informative positions was far too low to allow decisive conclusions.

Hindered by the disproportionate experimental effort, this data set has never been substantially enlarged. Rather, the focus of evolutionary biologists shifted from the comparison of proteins to the comparison of DNA sequences between species.

1.2 Phylogenetic Analyses Based on DNA Sequence Comparisons

At least two reasons suggest that comparative analyses of DNA sequences are more suitable for reconstructing the phylogeny

of closely related species than the comparison of proteins. First, the molecular clock of DNA sequences ticks faster than that of protein sequences and, therefore, provides a better resolution in time. This is because in coding sequences of genes, three types of positions can be distinguished: nondegenerated sites where any of the three possible nucleotide substitutions is nonsynonymous, that is, it alters the encoded amino acid. Twofold-degenerated sites where only two of the three possible nucleotide substitutions are nonsynonymous. And finally, fourfold-degenerated sites where no nucleotide substitution changes the encoded amino acid. Thus, only a subset of DNA sequence changes translates into changes of the respective protein. Furthermore, a change of the protein sequence generally interferes with protein function and is therefore commonly removed by purifying selection. Therefore, the accumulation of nonsynonymous DNA sequence differences and the consequent accumulation of amino acid sequence differences in proteins is considerably slow. In contrast, nucleotide sequence changes that leave the encoded protein unaltered are less likely subject to purifying selection and thus, accumulate differences at a substantially higher rate. Second, comparative molecular analyses can be extended to those regions in the genome that are not protein-coding, and in the optimal case, not functional at all. The latter sequences should be most informative for phylogenetic analyses of closely related species since they are likely to evolve free of selective constraint.

1.2.1 Comparative DNA–DNA Hybridization

DNA occurs in nature usually as a double helix where the two complementary strands are connected by hydrogen bonds formed between guanine and cytosine, and adenine and thymine respectively. Heating DNA duplexes breaks the hydrogen bonds and the two strands are separated – the DNA is "denatured" or "melted". The progress of this process can be quantified as it gradually increases the UV absorption at 260 nm of the DNA solution. Upon cooling, complementary DNA strands will rehybridize to duplex DNA. Performing the rehybridization in the presence of DNA from a second species, heteroduplexes can be formed where the two complementary strands stem from different species.

The temperature required to separate 50% of the DNA duplexes, also called the *melting temperature* (T_m), increases, depending on a number of factors: (1) The guanine and cytosine content (GC content) of the DNA: guanine and cytosine pair with three hydrogen bonds whereas adenine and thymine pair only with two hydrogen bonds; (2) the length of the analyzed fragments; (3) the degree of sequence similarity between the two strands forming the duplex. Accordingly, controlling for base composition and duplex length, the difference between the melting temperature of DNA heteroduplexes and that of the corresponding homoduplexes – both strands stem from the same species – (ΔT_m) serves as a measure for the genetic difference between species. For closely related species, the correlation between ΔT_m and the number of mismatches between the complementary strands of a DNA duplex is approximately linear.

An application of this method to the problem of hominoid phylogeny confirmed the conclusions from protein comparisons that gibbons, followed by orangutans, are the most diverged species. Furthermore, additional support was given to the view that humans and chimpanzees

shared a common ancestor after the separation of the gorilla lineage.

However, conclusions based on hybridization experiments have been treated with caution since ΔT_m is likely to be partially determined by factors of uncertain identity. The unclear influence of experimental conditions and of genomic organization such as the presence of duplications, which facilitates hybridization of nonhomologous DNA sequences, were among the main points of critique.

1.2.2 Mitochondrial DNA

In the pre-PCR era, the use of nucleic acids for phylogenetic analyses was substantially complicated by the need to reproducibly isolate the region of interest – usually a fragment of a few hundred base pairs (bp) – from a genome several million to few billion base pairs in size. To minimize this problem, initial focus was laid on the genome of mitochondria, which has a size of only approximately 16 000 bp in humans, and is generally present in more than 1000 copies per cell.

Isolated mitochondrial genomes from the hominoid species have been compared by two different methods. Restriction enzymes – enzymes that cut DNA at specific recognition sites – can be used to test for the presence and relative position of their respective cleavage site in a DNA sequence (the phylogenetic informative character in this approach). On the basis of restriction maps of their mitochondrial genomes, chimpanzees were placed – in contrast to results from other molecular comparisons – closer to gorillas than to humans. However, little statistical support was provided for that conclusion and alternative phylogenies were almost as likely.

Similar to the situation with comparative analyses of proteins, the most accurate way to determine the exact degree of difference between nucleic acids is to directly compare their sequences. Parts, and later, the sequences of the entire mitochondrial genome of hominoids were determined and compared. In these studies, chimpanzees appear again to be closer related to humans than to gorillas.

1.2.3 Nuclear DNA

Meanwhile, modern techniques of molecular biology provide a rapid and reproducible access to the DNA sequence of almost any region in a genome. As a consequence, a variety of data sets have been collected from the nuclear genome that address the question of hominoid phylogeny with particular regard to the relationship of humans, chimpanzees, and gorillas (Table 1). For practical reasons, early studies compared DNA sequences of coding regions from genes. Later, increasing focus was laid on the comparison of noncoding and potentially nonfunctional DNA sequences such as introns, pseudogenes, or intergenic regions. These DNA sequences appear to be the best possible data source for phylogenetic studies on closely related species. However, as can be clearly seen in Table 1, the supported phylogenetic relationship of humans, chimpanzees, and gorillas varies according to the analyzed genomic region.

Apparently, two factors have prevented the hominoid phylogeny from being determined by means of molecular comparisons: first, the insufficient accuracy and resolution of most comparative methods, and second, ambiguous information content in the molecular data itself. In order to determine which of the alternative phylogenies represents the true order in which the lineages to the individual hominoid species emerged (species tree), it is necessary to understand why phylogenetic trees estimated from DNA

Tab. 1 Comparative DNA sequence data sets to resolve the phylogeny of humans, chimpanzees, and gorillas.

Number of genomic loci	*Total compared length [bp]*	*Numbers of loci supporting phylogeny*[a]			
		(H,C)G	*(H,G)C*	*(C,G)H*	*(H,C,G)*
45	46 855	23	8	8	6
53	24 237	31	10	12	0
51	62 530	26	10	12	3

Notes: H: human; C: chimpanzee; G: gorilla.
[a]The grouped species shared a common ancestor after the separation of the third species. (H,C,G) represents those loci that could not resolve the phylogeny (trifurcation).

sequences (gene trees) can differ from the species tree. Mainly two problems need to be considered: first, sampling only a finite number of nucleotides from a locus produces a stochastic sampling error and thus, bears a probability to infer an incorrect gene tree; second, the model where the time point at which corresponding DNA sequences from different species last shared a common ancestor is identical to the time point of speciation is too simplistic. Figure 3 shows an example where the presence of sequence variants of the analyzed locus in the ancestral population in combination with lineage sorting can cause gene tree–species tree incongruities. Especially for species where the amount of genetic polymorphisms in their ancestral population was high and the time between subsequent speciation events was short – both seem to apply to the common ancestor of chimpanzees, humans, and gorillas – lineage sorting renders phylogenetic conclusions based on only a single locus unreliable. Therefore, a joint analysis of multiple independent DNA sequence data sets was required to eventually end the debate concerning the phylogeny of hominoids with sufficient statistical support. Humans are the closest relatives of chimpanzees and based on molecular evidence, both species separated approximately 4.6 to 6.2 million years ago. The likewise estimated date for the speciation of gorillas ranges from 6.2 to 8.4 million years before present.

1.3 Genetic Diversity

DNA sequence evolution is a continual process in time. Therefore, DNA sequence

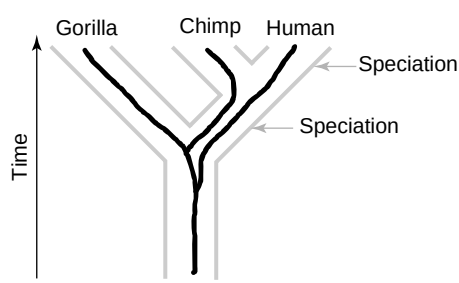

Fig. 3 Gene tree–species tree incongruities due to ancestral polymorphisms and lineage sorting. The gray and black lines represent the species tree (true phylogeny) and the gene tree respectively.

differences exist not only as divergence between species but also as diversity within species. In analogy to the use of divergence to reconstruct a phylogenetic tree that helps infer the evolutionary relationships among species, diversity data can be used to model a within-species tree representing the relationships of sequence variants at a given locus in a population. Such a genealogical tree can help clarify the demographic history of an individual species. Following a genealogical tree backwards in time, the genetic lineages represented by the individual alleles will gradually join until only a single lineage remains. This lineage from which all extant sequence variants are derived is termed the *most recent common ancestor* (MRCA). Making an assumption about the mutation rate and the long-term population size of the species – diversity increases with an increase of either parameter – the estimated time point of the MRCA can be inferred and reflects the amount of diversity present in the extant population.

For chimpanzees and humans, the date of their MRCAs as assessed from single nucleotide polymorphism data is 1 900 000 years and 500 000 years before present respectively. Accordingly, contemporary chimpanzees are two to fourfold more diverse compared to contemporary humans. What can explain the differing amounts of genetic diversity between both species? Since there is clear indication that the mutation rate in chimpanzees and humans is approximately the same, a difference in their population history is proposed as the underlying cause. Contemporary humans are, presumably, descendants of a small African population that subsequently expanded and spread over the entire world. Owing to the small size of the founding population, the majority of DNA sequence variation present in the ancestral human population has been lost, and thus, the genetic diversity in present day humans is low. Chimpanzees, in contrast, are believed to have not undergone such a substantial population size decrease. Therefore, a larger proportion of old genetic lineages could contribute to their present day diversity.

From the analysis of genetic diversity in chimpanzees, a second interesting aspect emerges. According to the location of their habitats in Africa, three groups of chimpanzees are distinguished: western, central, and eastern chimpanzees. From the perspective of DNA sequence variation of the mitochondrial genome and at a Y chromosomal locus, each of the three groups appear to be genetically unique. No shared DNA sequence variants for these loci have been observed, which indicate an independent evolution of the three chimpanzee populations. As a consequence, they were given the status of a subspecies: western chimpanzee (*Pan troglodytes verus*), central chimpanzee (*P. t. troglodytes*) and eastern chimpanzee (*P. t. schweinfurthii*). However, no such genetic uniqueness is seen at autosomal and X chromosomal loci. This was initially taken as an indication that the respective chimpanzee populations do not evolve independently and raised doubts on the biological significance of their classification into different subspecies. Recently, an alternative explanation has been proposed that would be consistent with an independent evolution of the chimpanzee populations. Subsequent to the reproductive isolation of populations, shared ancestral polymorphisms will gradually become extinct by genetic drift, that is, the stochastic loss of alleles caused by the fact that some individuals in a generation

have no offspring whereas others have several. Eventually, no ancestral alleles will be retained and the respective populations are genetically unique. It is now crucial for the case of the chimpanzee populations that the time needed to complete this process depends on the effective population size and thus varies among the analyzed loci. Mitochondrial and Y chromosomal loci will have removed ancestral sequence variants first, since their effective population size is the smallest – only one sequence variant is passed on to the next generation. X chromosomal loci will need on an average triple the time since they have a threefold higher effective population size – two copies in females, and one copy in males. And the time for autosomal loci to become genetically unique is about four times that of mitochondrial and Y chromosomal loci since on an average four copies – two copies each in males and females – are passed on to the next generation. As a consequence, the presence of shared sequence variants at biparentally inherited loci in the three chimpanzee populations in contrast to their genetic uniqueness at mitochondrial and Y chromosomal loci could indicate that the subspecies started to evolve independently – in evolutionary scales – only recently. Therefore, the time was not sufficient to establish genetic uniqueness at all loci.

2 Comparison of the Chimpanzee and Human Genomes

Early comparative studies of their proteins and DNA sequences have already provided evidence that chimpanzees and humans are genetically highly similar. However, differences exist that are informative in two ways. They help understand how genomes diverge subsequent to reproductive isolation of two populations. And they help identify those changes that form the genetic basis of the phenotypic properties differentiating chimpanzees from humans.

Four categories of differences between the genomes of chimpanzees and humans will be discussed in the following text.

2.1 Cytogenetic Differences

The microscopic comparison of chimpanzee and human chromosomes revealed a small number of cytogenetic differences, that is, differences in either chromosome number or differences in presence, order, and position of large chromosomal segments. Their detection was greatly simplified by chromosome banding techniques as well as fluorescence *in situ* hybridization (FISH), methods that allow a specific and reproducible labeling of individual chromosomal regions. The most apparent difference between the genomes of chimpanzees and humans is the number of chromosome pairs per cell. While chimpanzees – as well as gorillas and orangutans – have 24 pairs, humans have only 23. A comparison of the chromosome banding patterns between the species revealed that human chromosome 2 is present as two separate chromosomes (12 and 13) in chimpanzees (Fig. 4a). A subsequent molecular characterization revealed that a fusion at the ends (telomeres) of two ancestral chromosomes gave rise to human chromosome 2.

Seven additional chromosomes, human chromosomes 4, 5, 9, 12, 15, 16, and 17 differ between humans and chimpanzees by an inverted arrangement of chromosomal segments. In all cases, the inversion breakpoints are located on either side

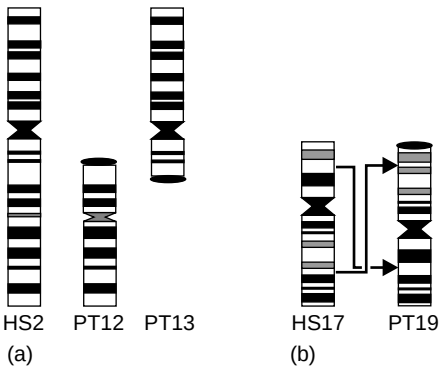

Fig. 4 Examples of cytogenetic differences between humans and chimpanzees. Chromosomes and chromosome bands are not drawn to scale. (a) Human chromosome 2 (HS2) is present as two separate chromosomes in chimpanzees (PT12, PT13). (b) Human chromosome 17 (HS17) and chimpanzee chromosome 19 (PT19) differ by a pericentric inversion. Arrows mark the inversion breakpoints.

of the centromere (pericentric inversion) (Fig. 4(b)). To date, the inversion breakpoints of a single human chromosome, human chromosome 17, have been analyzed on the DNA sequence level, and the breakpoints in three additional chromosomes, human chromosome 4, 9, and 12, have been mapped to intervals of 1 to 2 Mb. So far, no gene was found to be structurally modified by these inversions.

Finally, a number of chromosomes differ between chimpanzees and humans in the amount of constitutive heterochromatin, chromosomal regions that remain condensed throughout the cell cycle and generally do not contain genes.

Cytogenetic differences are discussed to comprise evolutionary important changes that contributed to the initial separation of chimpanzees and humans. Somewhat in contrast to this view, currently no evidence exists that the function of a gene has been affected by any of the observed chromosomal rearrangements. However, pericentric inversions can establish reproductive isolation between populations without the necessity of an altered gene function. In a heterozygous state, an individual carries one normal and one inverted copy of the chromosome, pericentric inversions can substantially interfere with the fertility of the individual. Crossing-over during meiosis in the inverted region generates recombination products that are generally incompatible with life. Thus, it can be speculated that the pericentric inversions differentiating the chromosomes of contemporary chimpanzees and humans comprised early reproductive barriers that facilitated the separation of the two species.

2.2 Subchromosomal Rearrangements and Repeat Content

In contrast to the well-defined extent of cytogenetic differences between humans and chimpanzees, the amount of genomic differences whose size is below the resolution of cytogenetic methods is still unclear. Subtelomeric regions, those genomic segments that are located directly adjacent to the chromosome ends, appear to be dynamic structures that undergo rapid changes over short evolutionary distances. For example, a subtelomeric region formed by a 32-bp repeat unit present at many chromosome ends of chimpanzees and almost all chromosome ends of gorillas is not found in humans. Similarly, a genomic locus that exists only as a single copy in the subtelomeric region of chimpanzee chromosome 17 is found in

eight additional chromosomes in humans. The reason for these rapid changes in the subtelomeric region of chromosomes is still unclear.

From the analysis of the human genome draft sequence, it emerged that about 5% of the DNA sequence – preferentially in the regions around the centromeres – represents evolutionary recent duplications of chromosomal segments with a DNA sequence identity of >90%. Interestingly, initial calculations reveal that more than 7000 exons are transcribed from these duplicated segments. Thus, gene duplications due to segmental duplications provide good candidates for species-specific evolution. Comparing the duplication pattern for human chromosome 22 between humans and chimpanzees, differences are seen, which imply that loss and gain of duplicated elements is an ongoing process in human and chimpanzee evolution. However, comparative analyses focusing on segmental duplications have just started. Therefore, the exact amount of differentially duplicated segments in the chimpanzee and human genomes as well as the number and identity of possibly involved genes remains to be determined.

Around 45% of the human genome consists of repetitive DNA sequences due to the expansion of various groups of mobile elements in the genome. As an example, a particular class of repeats termed *Alu-elements* are present in more than one million copies in the human genome comprising ~10% of the entire DNA sequence. Even though the overall mobility of transposable elements has markedly decreased over the past 35 to 50 million years, transposition is still ongoing in the human and, presumably, also in the chimpanzee genome.

Repetitive elements can propagate genetic differences between species by mainly two ways. First, the insertion of individual transposable elements into regulatory or coding regions of genes can alter expression and function of the affected genes. As an example, humans differ from chimpanzees – and from all other primates – by an inactivation of a gene *CMAH* whose product is essential for the generation of a certain glycoprotein. This inactivation is due to a human-specific insertion of an Alu-element that replaced a 92-bp long exon and, in addition, causes a frameshift in the subsequent coding sequence. The gene modified thus is expressed into a truncated and nonfunctional protein and is responsible for the only proven biochemical difference between chimpanzee and human cells to date. Second, intrachromosomal crossing-over mediated by direct or inverted repetitive elements can cause deletions and inversions, respectively. For example, exon 35 of the human tropoelastin gene has been deleted presumably by a recombination event between two flanking Alu-elements.

In summary, several factors can account for genetic differences between humans and chimpanzees on a subchromosomal scale. However, a determination of the extent to which they each contribute to the total amount of differences between both genomes must await more comprehensive studies.

2.3 Substitutional DNA Sequence Differences

The most frequently detected difference between chimpanzee and human DNA sequences is caused by substitutions of individual nucleotides. Initial studies suggested that per 200 compared nucleotide positions, only about 3 differ between humans and chimpanzees, equivalent to a mean sequence difference of around

1.5%. Many of these estimates were based on the comparison of short stretches of DNA from only one or few regions of the genome. However, different regions in the human genome differ in aspects such as gene content, transcription frequency, base composition, and recombination frequency all of which are suspected to influence the evolutionary rate of DNA sequences. Thus, sampling of DNA sequences from a multitude of loci distributed throughout the genome is required to obtain a relevant estimate of the average amount of substitutional DNA sequence differences between two species. On the basis of genome-wide comparisons of DNA sequences between humans and chimpanzees, their average amount of DNA sequence difference was recently determined to be 1.2%. The results furthermore reveal that indeed genomic regions vary substantially in their divergence among species (Fig. 5).

In absolute numbers, a mean DNA sequence difference of 1.2% corresponds to a total of about 40 million substitutional DNA sequence differences between the human and chimpanzee genomes. However, only a tiny proportion of the human genome – about 3% – is believed to be relevant for gene function. This suggests that less than (only) 1.2 million substitutional differences account for the distinct phenotypes of humans and chimpanzees.

2.4 Insertions and Deletions

Processes that are capable of removing nucleotides from DNA or adding nucleotides to DNA (deletions and insertions respectively) comprise a further source of genetic difference between species. Insertions and deletions mediated by transposable elements or by duplications of chromosomal segments have already been

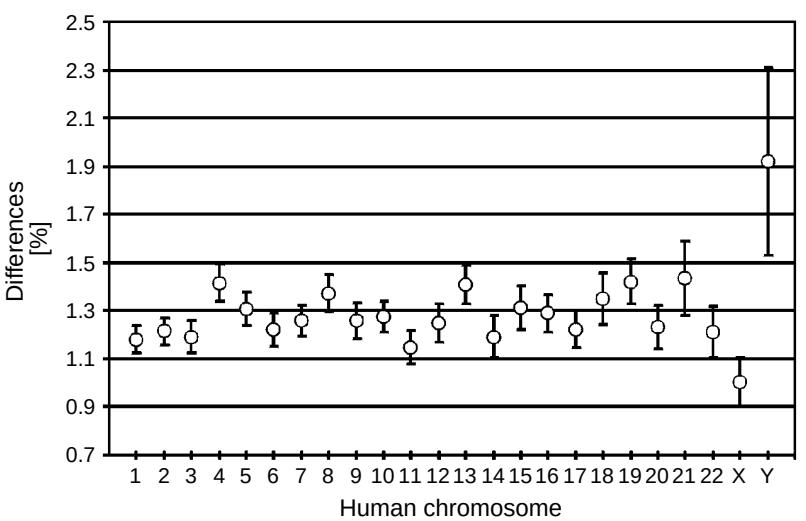

Fig. 5 Mean DNA sequence differences between humans and chimpanzees by human chromosome. Bars indicate 95% confidence intervals. [Reprinted with permission from Ebersberger, I., Metzler, D., Schwarz, C., Paabo, S. (2002) Genomewide comparison of DNA sequences between humans and chimpanzees, *Am. J. Hum. Genet.* **70**, 1490–1497.]

discussed above (Sect. 2.2). Each of these events generally affects a large number of DNA sequence positions; however, they presumably comprise considerably rare occurrences during human and chimpanzee evolution. Insertions and deletions extending only over a few nucleotides apparently occur far more frequently in DNA sequence evolution. A comparison of around only 700 kbp of genomic DNA sequence between humans and chimpanzees has detected already over 1000 individual insertions and deletions. About half of these extend over only 1 to 2 nucleotides and 95% do not exceed 20 bp in length.

Currently, it is estimated that a total of about 4% of DNA sequence in the chimpanzee and human genomes has no counterpart in the respective other species. This indicates that insertions and deletions contribute substantially both in the number of events and in the number of affected positions to the DNA sequence differences between species. Since insertions and deletions also appear to be commonly located in functional regions of the genome, they need to be considered as likely causes for phenotypic differences between humans and chimpanzees.

3 Chimpanzees in Evolutionary Research

Prior to describing the relevance of chimpanzees in evolutionary analyses – I will only focus on biological evolution and omit the complex topic of the evolution of culture – the term *evolution* itself shall be briefly discussed. In the broadest sense, evolution can be regarded as mere change of distinct characters over time. In the more specific view of biologists, evolution represents heritable changes in a population that spread over many generations. The only changes that can be stably inherited comprise modifications in the DNA sequence. Thus, in a strict sense, biological evolution is, ultimately, synonymous to DNA sequence evolution and evolutionary research aims at identifying the factors that determine emergence and accumulation of DNA sequence changes. In reality, however, it is far more practicable to study biological evolution at two levels: molecular evolution comprised of the change of DNA sequences over time, and the evolution of phenotypes leading to the differentiation of individual species. This distinction appears to be indispensable for the following reasons. The genetic basis of the vast majority of observed phenotypic differences between two species is generally, and in the better case, not well understood. Therefore, it is not feasible to study their evolution on the DNA sequence level. Furthermore, the majority of DNA sequence changes, at least in the genomes of mammals, has no effect on the phenotype since approximately 97% of the genome has no obvious function. As a consequence, the evolution of DNA sequences is in major parts independent of the evolution of phenotypes.

However, in order to arrive at a comprehensive understanding of biological evolution, the interdependence of DNA sequence evolution and phenotypic evolution needs to be determined. Here, the analysis of chimpanzees in comparison to humans is likely to provide relevant new insights.

3.1 DNA Sequence Evolution

The advantage of DNA sequence comparisons between chimpanzees and humans to shed light on the factors that determine DNA sequence evolution is twofold.

First, corresponding DNA sequences in the two species differ only in about 1.2% of the positions. This reduces the risk that individual positions have changed more than once in one species, or parallel in both species since the sequences last shared a common ancestor. Such multiple or parallel substitutions blur the correlation between the true and observed number of changes and comprise a substantial problem in the evolutionary analysis of DNA sequences. Second, the close relationship between chimpanzees and humans justifies the hope that factors determining rate and mode of DNA sequence change do not differ substantially between the two species. The approach to analyze DNA sequence evolution by a DNA sequence comparison between chimpanzees and humans is straightforward. Different regions in the human genome vary in divergence to the chimpanzee. Any genomic feature that is found to covary with the divergence can serve as a candidate to influence DNA sequence evolution. However, in order to correctly interpret the data, one pitfall needs to be regarded that has been already mentioned in Sect. 1.3. Genomic regions can differ in their divergence between humans and chimpanzees even though they evolve at the same rate. This problem arises since the time point at which the human and chimpanzee DNA sequences last shared a common ancestor is not the same throughout the genome. It is required to take this into account in order to identify regions that truly differ in their evolutionary rate.

When DNA sequences from different regions of the chimpanzee genome are compared to the corresponding sequences in humans, it is seen that different chromosomes vary in the amount of DNA sequence differences between humans and chimpanzees (Fig. 5). Interestingly, the extent of divergence for the individual chromosomes is correlated with the evolutionary time they spend in the male germ line. The X chromosome, which spends the least time in the male germ line, displays also the least divergence (1.0%), the Y chromosome, which is confined to the male germ line, displays the highest divergence (1.9%), and the divergence of the autosomes, which spend an intermediate time in the male germ line ranges between that of the sex chromosomes (1.2%). Currently, it is assumed that a mutation rate about threefold higher in the male germ line compared to the female germ line is responsible for this observation. However, it is noteworthy that the evolutionary rates also differ among autosomes that spend the same amount of time in the male and female germ line respectively. This indicates the existence of further factors that influence the accumulation of DNA sequence differences in the human genome. Recently, a second genomic feature has been found to covary with DNA sequence divergence between humans and chimpanzees. Regions with a high recombination rate also display a high divergence between the two species, while regions with a lower recombination rate have diverged less. This can be taken as an indication that the process of recombination is mutagenic and thus, drives DNA sequence evolution.

The amount of comparative DNA sequence data between chimpanzees is still limited. However, more extensive studies are likely to give further insights into how different factors interact to shape the rate and pattern of DNA sequence evolution in the genomes of humans and chimpanzees.

3.2
From Genotype to Phenotype

Maybe the most widely noticed field in evolutionary research is concerned with the emergence of new phenotypes that contribute to the uniqueness of individual species. Mainly three aspects are in the focus of interest: (1) When did the new phenotype arise? (2) Was it chance or selective advantage that caused its fixation in the species? (3) What are the genetic changes that underly the particular phenotype? While the answers to the first two questions are mainly of academic interest, the correlation between genotype and phenotype is – as will be discussed in Sect. 3.3 – of immediate practical use.

When comparing humans and chimpanzees, the impression is that both species are strikingly alike, yet show a certain set of well-defined differences. Meanwhile, an extensive catalog of phenotypic differences between humans and chimpanzees has been collected. Among the most well-known examples are bipedal walking, enlarged brain size, and spoken language, which distinguish humans from chimpanzees and all other primates. This set of phenotypic differences is opposed by a considerably low amount of genetic differences (see Sect. 2), which provides an excellent starting point for the correlation of genotype and phenotype. Two approaches can be chosen for such a venture. In a bottom-to-top strategy, the genomes of humans and chimpanzees are screened for differences that affect gene function. This can be either changes in the coding region of genes resulting in a structural change of the gene product, or in regions that control the expression pattern of a gene. While the identification of differences potentially affecting gene function is considerably simple, the subsequent determination of the likely effect on the phenotype is experimentally extremely demanding. In the first step, the function of the affected gene must be determined, and in the next step, the influence of the genetic change on the respective function of the gene product must be determined. This latter step will be especially problematic for such changes that affect the expression pattern of genes rather than the structure of the gene product. Eventually, inferences of the likely phenotypic consequence of the genetic change can be made. The reverse approach, a top-to-bottom strategy, starts from a recognized phenotypic difference between humans and chimpanzees. Certain genes that are known or suspected to contribute to the phenotype of interest are assigned as candidates whose DNA sequence can be screened for differences between humans and chimpanzees. Once species-specific differences are found for a candidate gene, the subsequent procedure resembles that of the bottom-to-top strategy with the subtle but essential difference that a particular phenotype has been already correlated with the gene.

To date, the study of candidate genes has been successfully applied at least in two studies, one of which originated from the observation of a structural difference between the cell surfaces of humans and great apes. It could be shown that this is due to an inactivation of an enzyme on the human lineage caused by a deletion in the corresponding gene (see Sect. 2.4). Even though this comprises a marked biochemical difference between human and chimpanzee cells with possible implications for the intercellular cross talk, the relevance of this difference for human evolution is still unclear. A second study focused on a gene *FOXP2*, which when mutated, causes a distinct type of speech impairment in humans.

A comparison of this gene between humans, chimpanzees, and other primates revealed that it carries two human-specific nucleotide substitutions, each of which translates into an amino acid substitution in the encoded protein. This becomes especially interesting in the light that this protein is evolutionarily highly conserved. Between humans and mice, only three amino acid changes are observed; two of which occur on the human lineage. Furthermore, there is indication that this gene has been the target of selection during recent human evolution. Both observations in combination can be judged as an indication that the respective gene could have a certain relevance for the evolution of spoken language in humans.

These examples demonstrate that, indeed, the comparison of humans and chimpanzees is capable of detecting genetic changes that can be correlated with phenotypic differences between both changes. This justifies the hope that in a step-by-step process, even the genetic frameworks of traits as complex as (spoken) language can be identified.

3.3 Comparative Analysis of Gene Expression

The phenotypic appearance of a species is not only dependent on the functional integrity of its genes but also on their correct activation with respect to time, place, and extent. Genetic changes that alter this construction plan of a species provide a powerful tool to achieve substantial phenotypic changes over a very short time period. In view of the low genetic difference between humans and chimpanzees, it was considered in early comparative studies that changes in a few "master genes" – genes that control gene expression at certain key points of development – account for the phenotypic differences between both species.

To date, such master changes, if they exist, remain to be identified. However, as a first step, evidence was provided that the pattern of gene expression indeed varies in a species-specific way. When the expression pattern of nearly 18 000 genes is analyzed for several human and chimpanzee individuals, it is found to be more similar among individuals of the same species than between species (Fig. 6). Starting from this observation, follow-up studies have to determine whether the evolution of gene expression is mainly due to regulatory changes in individual genes or whether a cluster of functionally related genes changes their expression in a concerted manner due to a switch in a shared master gene. Furthermore, those individual genes that have substantially changed

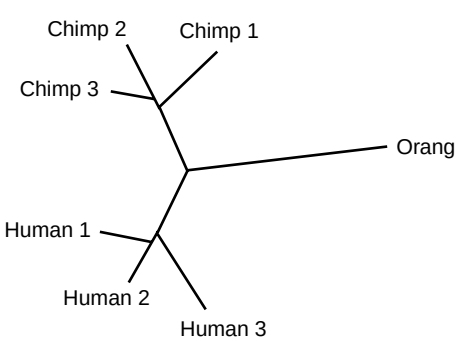

Fig. 6 Relative extent of the differences in the expression pattern within and between species. The expression pattern jointly determined from nearly 18 000 genes is found to be more similar within humans and chimpanzees respectively, than between both species.

their expression pattern in humans or chimpanzees comprise candidates for phenotypic differences between both species and thus, should be the subject of comparative studies as described in the previous section (Sect. 3.2).

4 Chimpanzees in Biomedical Research

The cure or prevention of diseases is one of the major interests in human research. In the study of disease-causing processes on the cellular and molecular level, nonhuman primates in general, and chimpanzees in particular, can be regarded as important model organisms. Their similarity to humans renders them susceptible to a number of pathogens correlated with human diseases. Thus, the infection of primates under controlled laboratory conditions allows the study of the onset and progress of the respective disease with a resolution far beyond what can be achieved by observatory studies in humans. Furthermore, as a result of the growing understanding of disease-causing processes, new therapeutical strategies and medicines will be developed that specifically target individual pathways in the human metabolism. Therefore, the traditional model organisms that are rather distantly related to humans may no longer be suitable to test the efficacy and safety of these medicines. Rather, model organisms such as the chimpanzee are required that resemble the human metabolism as closely as possible. Eventually, certain human diseases and pathogens appear to affect chimpanzees either to a lesser extent than humans, or not at all. The identification of the underlying molecular basis of this reduced susceptibility will provide starting points for new therapeutical approaches to cure the respective disease.

4.1 Are Chimpanzees Good Model Organisms?

The arguments listed above comprise a clear vote for a wide use of chimpanzees in biomedical research. Accordingly, numerous research facilities founded chimpanzee colonies, and in the 1980s, breeding programs were started to insure that enough chimpanzees were available for research needs. However, it appears that the similarity between chimpanzees and a (hypothetical) good model organism for biomedical research is very limited. Such an organism would be rather mouse-sized than chimp-sized. Not only is that desirable due to the limitation of laboratory space and budget for food, it also greatly facilitates handling of the animals during the experiments. Who ever has seen somebody pulling chimps by their tail out of the cage? And this is not because their tails are pretty much nonexisting! Furthermore, a good model organism would have a generation time and a life span substantially shorter than that of chimpanzees. The benefits would be twofold. First, it increases the value of the model organism for research on transgenic animals. The specific inactivation of genes in an organism (knockout) as well as their overexpression or exchange with functional homologs from other species (knockin) gains increasing importance in the study of function and evolution of genes and gene products. A minimal time span between the genetic modification of the individual – this generally happens at a very early stage in development – and the latest possible observation of the corresponding phenotypic effect is desirable for both kinds of studies. The latter time point is obviously

determined by the life span of the model organism. Furthermore, once an individual carrying the desired genetic modification has been generated – which is still a costly, time-consuming, and experimentally elaborative venture – its maintenance as a laboratory strain is generally desirable. This, however, is only feasible when the generation time of the respective species is in a reasonable range of at most a year. A slowly developing, late reproducing, and long living species such as chimpanzees is thus, not particularly well suited for standard research on transgenic animals. Second, an early (natural) death of the research animal elegantly solves the problem of what to do with the animals once the experiments are over. Chimpanzees, however, have an average life span of around 40 to 50 years, which is well beyond the duration of most experimental studies. The relevance of this problem becomes immediately obvious when one imagines a group of chimpanzees infected with a human pathogen that need to be taken care of once they are no longer useful for experimental studies.

Clearly, strong arguments exist to use chimpanzees as experimental animals in biomedical research. However, the practical disadvantages are apparent. Together with a number of ethical concerns that I will discuss below (Sect. 4.4), the use of chimpanzees as a standard experimental model organism is clearly voted against. Meanwhile, this has been widely recognized and the number of chimpanzees held as laboratory animals has substantially decreased over the past years. In the case of Europe, invasive experimental studies on chimpanzees have entirely ended.

4.2 Biomedical Differences Between Humans and Chimpanzees

Practical experience has shown that chimpanzees are not suitable as experimental model organisms for studying human diseases, or for testing the efficacy and safety of new pharmaceuticals. Nevertheless, a certain relevance remains for chimpanzees in biomedical research. From long-term observations in primate centers and zoos, differences in susceptibility and severity of certain human diseases between humans and chimpanzees have emerged. A few examples are listed in Table 2. Although the evidence is in parts fragmentary and vague – in certain cases it

Tab. 2 Differences in the disease spectra between humans and chimpanzees.

Medical condition	*Humans*	*Chimpanzees*	*Evidence*
HIV progression to AIDS	Common	Very rare	Definite
Late complication in hepatitis B/C	Frequent	Uncommon	Definite
Malaria[a]	Susceptible	Resistant	Definite
Menopause	Universal	Rare	Definite
Influenza A symptomatology	Moderate to severe	Mild	Likely
Alzheimer's disease pathology	Common	Rare	Likely
Myocardial infarction	Common	Uncommon	Likely
Epithelian cancers	Common	Rare	Likely

[a]Caused by *Plasmodium falciparum* infection.

has mere anecdotal character – the medical relevance of involved diseases such as AIDS, hepatitis, or certain types of cancer is sufficient to draw attention to this issue. Differences in the disease spectrum between two species can be attributed to two different causes: differences in environmental influences such as diet and differing genetic predispositions. In recent years, the environmental conditions for captive chimpanzees have become increasingly similar to those of humans. This, however, had no substantial effect on the differences of their respective disease spectra. Therefore, the existence of crucial genetic differences between the two species must be assumed whose identification comprises the main relevance of biomedical studies on chimpanzees. Conveniently, the identification of candidates for these genetic differences can be achieved by a comparison of the genome sequences of chimpanzees and humans and does not require experimental studies on chimpanzees.

As an outlook, the medical consequences of biomedical studies on chimpanzees are likely to be revolutionary. Once the genetic basis is understood as to why this species is less susceptible to certain human diseases, cure or even prevention of these diseases can be imagined by mimicking the chimpanzee phenotype in well-defined disease-relevant parts of human biology.

4.3 Chimpanzees and Infectious Diseases

Maybe the most widely recognized biomedical difference between chimpanzees and humans is their different susceptibility to AIDS upon infection with HIV-1 (Human Immunodeficiency Virus 1), the most common cause of AIDS in humans. While chimpanzees rarely display AIDS-like syndromes, the progression to AIDS in humans is common. This difference is generally explained by the evolutionary history of HIV-1. It is suggested that HIV-1 originated from the chimpanzee virus SIV (Simian Immunodeficiency Virus) that crossed the species border only about 50 years ago. Thus, the low susceptibility to AIDS in chimpanzees could be due to a long coevolution of virus and host in a way that contemporary chimpanzees get infected by the virus but due to an efficient control of the virus load do not develop the disease. Humans, in contrast, who have become exposed to the virus only recently, lack the immunological ability for virus control and thus, suffer from AIDS. Obviously, the missing development of the disease renders chimpanzees inadequate as a model to study the progress of AIDS. Other primate species, such as the rhesus macaque appear to be more suitable since AIDS-like symptoms have been observed in these species subsequent to infection with SIV. However, the genetic reasons for the adapted immune response that cause the relative resistance against AIDS in chimpanzees are of considerable interest. They might point toward novel therapeutic approaches to prevent the progression to AIDS in HIV-infected humans. Initial insights suggest that differences in the allele combination in the major histocompatibility complex – a collection of genes correlated with immune response – between humans and chimpanzees is responsible for their varying susceptibility to AIDS.

Similar to the case of infection with HIV-1 and progression to AIDS, chimpanzees reduce the risk of long-term complications of hepatitis infections by an effective control of the virus load in the organism. Thus, there appears to be a general capability in

chimpanzees to deal more effectively with virus infections than humans. It will be the main task of future studies to determine the exact elements of the virus–host relationship in chimpanzees that account for this better virus control.

In conclusion, although chimpanzees are not suited as experimental model organisms, there appears to be sufficient reason for the comparison of humans and chimpanzees to become a major field in biomedical research. However, one should be aware that the insights from past biomedical studies in chimpanzees have not met the high initial expectations. The risk remains that the biomedical relevance of genetic differences between humans and chimpanzees is too difficult to interpret just by observatory studies. As a consequence, a substantially increased experimental effort might be required whose realization ultimately depends on the establishment of new experimental systems that replace chimpanzees as a model organism. Such systems should combine a maximum of the advantages provided by chimpanzees as a model organism with a minimum of its practical and ethical concerns.

4.4 Ethical Considerations

Chimpanzees, like all other great apes, are an endangered species that, due to human interaction, face the extinction of their wild populations. From this perspective, it seems somewhat odd to discuss their use for experimental research beneficial primarily to humans. However, chimpanzees breed well in captivity, which insures that enough animals are available for research without affecting the wild populations. In addition, it can be argued that every study, even those with fatal outcome for the participating individual, eventually benefits chimpanzees as a species. For example, vaccines and drugs to cure human diseases developed on the basis of experimental research on chimpanzees will be effective also in chimpanzees. A scenario can be imagined where these pharmaceuticals help save a wild chimpanzee population whose survival is endangered by such a disease. In combination, both arguments would, presumably, be sufficient for a well-controlled use of almost any other species – even one that faces its extinction – as an experimental animal. However, the situation is more complex for chimpanzees. An increasing number of comparative studies reveal that there is no clear-cut qualitative difference distinguishing humans from chimpanzees in a number of areas that have initially been regarded as "specific to humans." Many points of similarities exist in cognition and psychology between chimpanzees and humans. For example, it was repeatedly suggested that chimpanzees can understand and use rudimentary forms of language – even though they cannot speak due to anatomical limitations – and there is certainly tool usage in chimpanzees. In addition, only recently, it was appreciated that chimpanzees have cultural traditions similar to humans. Given these insights and given the high genetic similarity between humans and chimpanzees, it might be worthwhile to reconsider the clear-cut distinction in "human being" and "animal". The resulting implications for the use of chimpanzees in research are obvious and well recognized. In brief, only such studies should be done with chimpanzees that would be done with humans who are considered not to be able to provide informed consent. As a legal consequence, the government of New Zealand has extended three basic human rights

to chimpanzees: (1) the right not to be deprived of life; (2) the right not to be subjected to torture and cruel treatment; and (3) the right not to be subjected to medical or scientific research that is not in the best interest of this individual.

Whether these ventures and agreements are sufficient to protect chimpanzees from abuse in human research will be seen in the future. As soon as comparative studies of humans and chimpanzees lead to fundamental questions that can be solved neither by DNA sequence comparisons nor by transgenic experiments in model systems, research proposals will arise that include the generation of transgenic – that will be ultimately "humanized" – chimpanzees. Therefore, it is an ethical imperative to accompany an extensive comparative study on the chimpanzee genome with a program that studies ethical, legal, and social implications of such research.

5 The Chimpanzee Genome Project

Even under careful consideration of all practical and ethical limitations that are correlated with comparative studies between chimpanzees and humans, the insights based on the availability of a chimpanzee genome sequence are likely to be substantial. As a consequence, the US National Human Genome Research Institute (NHGRI) has recently assigned chimpanzees the top priority for whole-genome sequencing, a decision that might be a pioneer for future research on nonhuman primates. The choice of sequencing the genome of chimpanzees rather than that of the rhesus macaque – the most widely used primate in biomedical research – seems to put more weight on the evolutionary aspect of a human–nonhuman primate comparison than on its immediate value for experimental biomedical studies.

5.1 The Quest for Differences

With respect to the list of eukaryotes whose genome has been entirely sequenced, the choice of determining the genome sequence of chimpanzees comprises the first step into the next phase of comparative genomics. Comparing the genome sequences from yeast, worm, fly, mouse, and human led to the identification of shared, and thus, evolutionarily conserved elements. Follow-up studies mainly aimed at determining the function of these conserved elements and generating a rough view on the genetic framework underlying organismal organization. However, adding chimpanzees to that list shifts the focus ultimately from conserved genomic elements toward the differences between two genomes. As a consequence, the initial outcome of a genome comparison between humans and chimpanzees will be a comprehensive map of their DNA sequence differences. As a second step, the functional consequences of these differences need to be analyzed and, eventually, a fine-scale view on the functional organization of the human genome will emerge.

5.2 Sequencing Strategy and Requirements

Sequencing an entire genome with more than three billion bases is still an enterprise that requires careful planning and strategical considerations. Recent sequencing efforts, including the genomes of *Drosophila*, human, and mouse clearly demonstrated that the whole-genome shotgun sequencing strategy is the fastest and

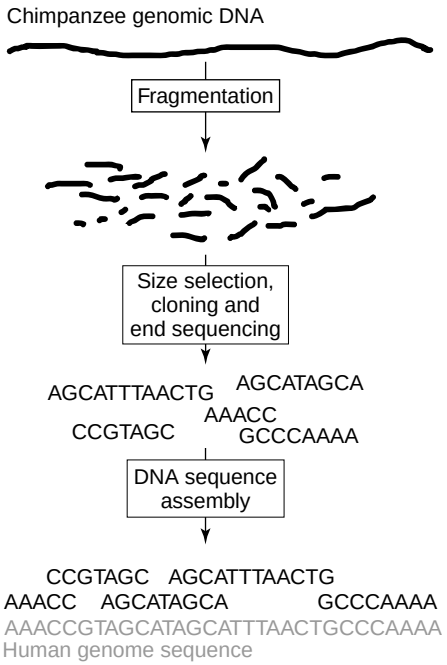

Fig. 7 Whole-genome shotgun strategy to sequence the chimpanzee genome. The entire genome of an individual is fragmented at random positions and genomic libraries with a defined insert size are generated. Subsequently, random clones are picked from the genomic libraries and the DNA sequences of their insert ends are determined. In the assembly step, the individual DNA sequences are aligned to reconstruct a contiguous consensus sequence that represents the original genomic sequence of the chimpanzee. This step is greatly facilitated for the chimpanzee genome since the human genome sequence can serve as a guideline for the assembly.

most efficient way to determine an initial draft sequence of a genome. The general principle of this strategy is outlined in Fig. 7. While the generation of the shotgun sequence reads is considerably straightforward, their assembly to a comprehensive consensus sequence comprises the crucial and most problematic step. This step, however, will be greatly facilitated for the chimpanzee genome. Given the high similarity between both genomes, it will be in most cases straightforward to align the chimpanzee DNA sequence fragments to the human genome sequence. However, special care must be given to those genomic regions that differ between the two species. Human regions not present in chimpanzees will cause gaps in the alignment, while chimpanzee DNA sequences not present in humans will fail to align against the human genome. In addition, recent duplications as well as inversions of genomic regions comprise further obstacles on the way toward a comprehensive draft version of the chimpanzee genome.

Following the shotgun phase of a genome project, remaining gaps and uncertainties in the draft sequence must be removed in the so-called finishing phase. This sequence-finishing is tedious, time-consuming and thus, expensive. However, it is assumed that the majority of comparative studies between chimpanzees and humans will not substantially benefit from a finished, rather than a draft version of the chimpanzee genome. Thus, it is currently considered to initially leave the chimpanzee genome sequence in a draft status. It is estimated that such a draft sequence will cover more than 90% of the chimpanzee genome.

In conclusion, a first draft of the chimpanzee genome sequence can be generated with considerably low effort and is likely to be available by the end of 2003.

However, there is one important issue when working with this draft sequence that played only a very minor role in previous comparative studies of genome sequences. Errors in the draft sequence of the chimpanzee genome, but also in the human genome sequence due to errors in the assembly process or due to sequencing errors suddenly gain relevance. An attempt to identify conserved regions between distantly related species is conservative with respect to the quality of the draft sequence. Regions that have been found conserved between species using early versions of the draft sequence have a high probability of remaining conserved after sequencing errors in the draft sequence have been corrected. It is self-evident that the situation is certainly more difficult when the focus lays on DNA sequence differences between genomes as in the case of the human–chimpanzee comparison. To minimize this problem, every DNA sequence position in the draft version of the chimpanzee genome should get assigned a confidence value to allow a quick assessment of its reliability.

5.3 Human, Chimpanzee, and. . . Future Perspectives

With the availability of the chimpanzee genome draft sequence, the first step into the next stage of comparative genomics will be completed. A comprehensive catalog of genetic differences between humans and chimpanzees will emerge that ultimately reveals those changes responsible for the distinct phenotypic traits of both species. The identification of those presumably few functional changes and the dating of their presumed occurrence will be the main task of future studies. To accomplish this goal, it will be necessary to exactly determine the extent and pattern of genetic variation both within humans and within chimpanzees. This not only helps in identifying those genuine genetic differences between both species that cannot be attributed to polymorphisms in either species but also helps pinpoint those regions in the genomes of chimpanzees and humans for which fewer than average sequence variants exist. Such a reduction of diversity can comprise signatures of selection during recent human and chimpanzee evolution and provide hints for the likely importance of changes that are close by in functional regions of the genome.

Eventually, the detection of genuine genetic differences between humans and chimpanzees can only then be interpreted in a meaningful way when the lineage on which the corresponding change occurred has been determined. Therefore, the genome of a third, more distantly related species is required as an outgroup that helps infer whether an observed DNA sequence difference between humans and chimpanzees is due to a change on the human or the chimpanzee lineage. Presumably, a second anthropoid primate would be optimal as such an outgroup. In order to make full use of the genomic comparison of humans and chimpanzees, the genome sequence of the outgroup species should be determined immediately following the completion of the chimpanzee genome sequence. Likely candidate species are two Old World monkeys regularly used in biomedical research: the rhesus macaque (*Macaca mulatta*), and the baboon (*Papio hamadryas*), or the orangutan (*P. pygmaeus*). Which of these species to choose depends largely on future plans to use primates as experimental animals.

See also Genetics, Molecular Basis of; Genomic Sequencing (Core Article).

Bibliography

Books and Reviews

Bontrop, R.E. (2001) Non-human primates: essential partners in biomedical research, *Immunol. Rev.* **183**, 5–9.

Gagneux, P., Varki, A. (2001) Genetic differences between humans and great apes, *Mol. Phylogenet. Evol.* **18**, 2–13.

Goodman, M. (1999) The genomic record of Humankind's evolutionary roots, *Am. J. Hum. Genet.* **64**, 31–39.

Hacia, J.G. (2001) Genome of the apes, *Trends Genet.* **17**, 637–645.

Kaessmann, H., Paabo, S. (2002) The genetical history of humans and the great apes, *J. Intern. Med.* **251**, 1–18.

King, M.C., Wilson, A.C. (1975) Evolution at two levels in humans and chimpanzees, *Science* **188**, 107–116.

Olson, M.V., Varki, A. (2003) Sequencing the chimpanzee genome: insights into human evolution and disease, *Nat. Rev. Genet.* **4**, 20–28.

Primary Literature

Arnason, U., Gullberg, A., Janke, A. (1998) Molecular timing of primate divergences as estimated by two nonprimate calibration points, *J. Mol. Evol.* **47**, 718–727.

Arnason, U., Gullberg, A., Janke, A., Xu, X. (1996a) Pattern and timing of evolutionary divergences among hominoids based on analyses of complete mtDNAs, *J. Mol. Evol.* **43**, 650–661.

Arnason, U., Xu, X., Gullberg, A. (1996b) Comparison between the complete mitochondrial DNA sequences of Homo and the common chimpanzee based on nonchimeric sequences, *J. Mol. Evol.* **42**, 145–152.

Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., Eichler, E.E. (2002a) Recent segmental duplications in the human genome, *Science* **297**, 1003–1007.

Bailey, J.A., Yavor, A.M., Viggiano, L., Misceo, D., Horvath, J.E., Archidiacono, N., Schwartz, S., Rocchi, M., Eichler, E.E. (2002b) Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22, *Am. J. Hum. Genet.* **70**, 83–100.

Britten, R.J. (2002) Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13633–13635.

Caccone, A., Powell, J.R. (1989) DNA divergence among hominoids, *Evolution* **43**, 925–942.

Casane, D., Boissinot, S., Chang, B.H., Shimmin, L.C., Li, W. (1997) Mutation pattern variation among regions of the primate genome, *J. Mol. Evol.* **45**, 216–226.

Chen, F.C., Li, W.H. (2001) Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees, *Am. J. Hum. Genet.* **68**, 444–456.

Chou, H.H., Hayakawa, T., Diaz, S., Krings, M., Indriati, E., Leakey, M., Paabo, S., Satta, Y., Takahata, N., Varki, A. (2002) Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11736–11741.

Chou, H.H., Takematsu, H., Diaz, S., Iber, J., Nickerson, E., Wright, K.L., Muchmore, E.A., Nelson, D.L., Warren, S.T., Varki, A. (1998) A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11751–11756.

deGroot, N.G., Otting, N., Doxiadis, G.G., Balla-Jhagjhoorsingh, S.S., Heeney, J.L., van Rood, J.J., Gagneux, P., Bontrop, R.E. (2002) Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11748–11753.

Ebersberger, I., Metzler, D., Schwarz, C., Paabo, S. (2002) Genomewide comparison of DNA sequences between humans and chimpanzees, *Am. J. Hum. Genet.* **70**, 1490–1497.

Enard, W., Khaitovich, P., Klose, J., Zollner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore, E., Varki, A., Ravid, R., Doxiadis, G.M., Bontrop, R.E., Paabo, S. (2002a)

Intra- and interspecific variation in primate gene expression patterns, *Science* **296**, 340–343.

Enard, W., Przeworski, M., Fisher, S., Lai, C., Wiebe, V., Kitano, T., Monaco, A., Paabo, S. (2002b) Molecular evolution of FOXP2, a gene involved in speech and language, *Nature* **418**, 869–872.

Felsenstein, J. (1987) Estimation of hominoid phylogeny from a DNA hybridization data set, *J. Mol. Evol.* **26**, 123–131.

Ferris, S.D., Wilson, A.C., Brown, W.M. (1981) Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2432–2436.

Gagneux, P., Wills, C., Gerloff, U., Tautz, D., Morin, P.A., Boesch, C., Fruth, B., Hohmann, G., Ryder, O.A., Woodruff, D.S. (1999) Mitochondrial sequences show diverse evolutionary histories of African hominoids, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5077–5082.

Gibbons, A. (1998) Which of our genes make us human?, *Science* **281**, 1432–1434.

Goldman, D., Giri, P.R., O'Brien, S.J. (1987) A molecular phylogeny of the hominoid primates as indicated by two-dimensional protein electrophoresis, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3307–3311.

Goodman, M. (1962) Evolution of the immunologic species specificity of human serum proteins, *Hum. Biol.* **34**, 104–150.

Goodman, M., Koop, B.F., Czelusniak, J., Fitch, D.H., Tagle, D.A., Slightom, J.L. (1989) Molecular phylogeny of the family of apes and humans, *Genome* **31**, 316–335.

Goodman, S. (2001) Europe brings experiments on chimpanzees to an end, *Nature* **411**, 123.

Hasegawa, M., Kishino, H., Yano, T. (1987) Man's place in Hominoidea as inferred from molecular clocks of DNA, *J. Mol. Evol.* **26**, 132–147.

Hayakawa, T., Satta, Y., Gagneux, P., Varki, A., Takahata, N. (2001) Alu-mediated inactivation of the human CMP-N-acetylneuraminic acid hydroxylase gene, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11399–11404.

Horai, S., Satta, Y., Hayasaka, K., Kondo, R., Inoue, T., Ishida, T., Hayashi, S., Takahata, N. (1992) Man's place in Hominoidea revealed by mitochondrial DNA genealogy, *J. Mol. Evol.* **35**, 32–43.

Ijdo, J.W., Baldini, A., Ward, D.C., Reeders, S.T., Wells, R.A. (1991) Origin of human chromosome 2: an ancestral telomere-telomere fusion, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9051–9055.

Johnson, M.E., Viggiano, L., Bailey, J.A., Abdul-Rauf, M., Goodwin, G., Rocchi, M., Eichler, E.E. (2001) Positive selection of a gene family during the emergence of humans and African apes, *Nature* **413**, 514–519.

Kaessmann, H., Wiebe, V., Paabo, S. (1999) Extensive nuclear DNA sequence diversity among chimpanzees, *Science* **286**, 1159–1162.

Kehrer-Sawatzki, H., Schreiner, B., Tanzer, S., Platzer, M., Muller, S., Hameister, H. (2002) Molecular characterization of the pericentric inversion that causes differences between chimpanzee chromosome 19 and human chromosome 17, *Am. J. Hum. Genet.* **71**, 2.

McConkey, E.H., Fouts, R., Goodman, M., Nelson, D., Penny, D., Ruvolo, M., Sikela, J., Stewart, C.B., Varki, A., Wise, S. (2000) Proposal for a human genome evolution project, *Mol. Phylogenet. Evol.* **15**, 1–4.

McConkey, E.H., Goodman, M. (1997) A human genome evolution project is needed, *Trends Genet.* **13**, 350–351.

McConkey, E.H., Varki, A. (2000) A primate genome project deserves high priority, *Science* **289**, 1295–1296.

Monfouilloux, S., Avet-Loiseau, H., Amarger, V., Balazs, I., Pourcel, C., Vergnaud, G. (1998) Recent human-specific spreading of a subtelomeric domain, *Genomics* **51**, 165–176.

Muchmore, E.A. (2001) Chimpanzee models for human disease and immunobiology, *Immunol. Rev.* **183**, 86–93.

Muchmore, E.A., Diaz, S., Varki, A. (1998) A structural difference between the cell surfaces of humans and the great apes, *Am. J. Phys. Anthropol.* **107**, 187–198.

Muller, S., Wienberg, J. (2001) "Bar-coding" primate chromosomes: molecular cytogenetic screening for the ancestral hominoid karyotype, *Hum. Genet.* **109**, 85–94.

Nickerson, E., Nelson, D.L. (1998) Molecular definition of pericentric inversion breakpoints occurring during the evolution of humans and chimpanzees, *Genomics* **50**, 368–372.

O'hUigin, C., Satta, Y., Takahata, N., Klein, J. (2002) Contribution of homoplasy and of ancestral polymorphism to the evolution of genes in anthropoid primates, *Mol. Biol. Evol.* **19**, 1501–1513.

Royle, N.J., Baird, D.M., Jeffreys, A.J. (1994) A subterminal satellite located adjacent to

telomeres in chimpanzees is absent from the human genome, *Nat. Genet.* **6**, 52–56.

Ruvolo, M. (1997) Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets, *Mol. Biol. Evol.* **14**, 248–265.

Ruvolo, M., Pan, D., Zehr, S., Goldberg, T., Disotell, T.R., Dornum, M.V. (1994) Gene trees and hominoid phylogeny, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8900–8904.

Satta, Y., Klein, J., Takahata, N. (2000) DNA archives and our nearest relative: the trichotomy problem revisited, *Mol. Phylogenet. Evol.* **14**, 259–275.

Satta, Y. (2001) Comparison of DNA and protein polymorphisms between humans and chimpanzees, *Genes Genet. Syst.* **76**, 159–168.

Schmid, C.W., Marks, J. (1990) DNA hybridization as a guide to phylogeny: chemical and physical limits, *J. Mol. Evol.* **30**, 237–246.

Schwartz, J.H. (1984) The evolutionary relationships of man and orangutans, *Nature* **308**, 501–505.

Sibley, C.G., Ahlquist, J.E. (1984) The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization, *J. Mol. Evol.* **20**, 2–15.

Sibley, C.G., Comstock, J.A., Ahlquist, J.E. (1990) DNA hybridization evidence of hominoid phylogeny: a reanalysis of the data, *J. Mol. Evol.* **30**, 202–236.

Stone, A.C., Griffiths, R.C., Zegura, S.L., Hammer, M.F. (2002) High levels of Y-chromosome nucleotide diversity in the genus Pan, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 43–48.

Thomas, J.W., Touchman, J.W. (2002) Vertebrate genome sequencing: building a backbone for comparative genomics, *Trends Genet.* **18**, 104–108.

Toder, R., Grutzner, F., Haaf, T., Bausch, E. (2001) Species-specific evolution of repeated DNA sequences in great apes, *Chromosome Res.* **9**, 431–435.

Varki, A. (2000) A chimpanzee genome project is a biomedical imperative, *Genome Res.* **10**, 1065–1070.

Vigilant, L., Paabo, S. (1999) A chimpanzee millennium, *Biol. Chem.* **380**, 1353–1354.

Whiten, A., Goodall, J., McGrew, W.C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C.E., Wrangham, R.W., Boesch, C. (1999) Cultures in chimpanzees, *Nature* **399**, 682–685.

Winter, H., Langbein, L., Krawczak, M., Cooper, D.N., Jave-Suarez, L.F., Rogers, M.A., Praetzel, S., Heidt, P.J., Schweizer, J. (2001) Human type I hair keratin pseudogene phihHaA has functional orthologs in the chimpanzee and gorilla: evidence for recent inactivation of the human gene after the Pan-Homo divergence, *Hum. Genet.* **108**, 37–42.

Wise, C.A., Sraml, M., Rubinsztein, D.C., Easteal, S. (1997) Comparative nuclear and mitochondrial genome diversity in humans and chimpanzees, *Mol. Biol. Evol.* **14**, 707–716.

Yunis, J.J., Prakash, O. (1982) The origin of man: a chromosomal pictorial legacy, *Science* **215**, 1525–1530.

Chirality in Biology

Ronald Bentley
University of Pittsburgh, Pittsburgh, PA, USA

1 Chirality – An Introduction 582

2 Molecular Chirality in Biology 582
2.1 Chirality Descriptors: the *R*/*S* Notation 582
2.2 Chirality Descriptors: the D/L Notation 584
2.3 Axial Chirality 585
2.4 Amino Acids 587
2.5 Carbohydrates 590
2.6 Other Natural Products 592

3 Physiological Responses to Enantiomers 594
3.1 Taste 594
3.2 Odor 595
3.3 Insect Pheromones 599
3.4 Plant Growth Regulators 600
3.5 Chiral Drugs 600

4 Chiral Recognition 604

5 Enzyme Specificity with Chiral Substrates 606

6 Prostereoisomerism 607

7 Chirality in Biological Structures and Behavior 611
7.1 Structures 611
7.2 Behavioral Asymmetries 613

8 Origin of Chirality on Earth 614

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Bibliography 615
Books and Reviews 615
Primary Literature 615

Keywords

Chiral
Describing an object or a molecule whose image in a plane mirror cannot be superposed on the original.

Chiral Drug
A pharmaceutical agent that exists in two enantiomeric forms. In many cases, the two forms will have different physiological activity.

Chiral Methyl
A methyl group with the three hydrogen isotopes, ^{1}H, ^{2}H, ^{3}H. When combined with another group, R, a chiral molecule is obtained: $R\text{-}C^{1}H^{2}H^{3}H$.

Chiral Phosphorous Atom
A phosphorous atom linked to the three oxygen isotopes, ^{16}O, ^{17}O, ^{18}O, and to another group, usually OR: hence, $RO-P-^{16}O^{17}O^{18}O$.

Chiral Recognition
Discrimination between enantiomers achieved by enzymes or biological receptor molecules and by some chemical catalysts and reagents.

Configuration
Three-dimensional arrangement of atoms or groups of atoms at an element of chirality; arrangements resulting from rotation about single bonds are excluded.

Diastereoisomer
Stereoisomers not in an object-/mirror-image relationship.

D/L Notation
An arbitrary convention to denote configuration about a chiral element in α-amino acids, some α-hydroxy acids, and carbohydrates.

Enantiomer
Stereoisomers that are in an object-/mirror-image relationship.

Homochiral
A molecule or assembly of molecules in which the components have like chirality sense. Normal protein amino acids have L-configuration; for naturally occurring carbohydrates, the D-configuration predominates. The antonym is heterochiral.

Isomer
Compounds having the same molecular formula but differing in the nature or sequence of bonding of their atoms, or in the spatial arrangement of their atoms.

Optical Activity
The property, possessed by some substances (or solutions of some substances) of rotating the plane of plane-polarized light. For a chiral substance, one enantiomer will produce dextrorotation, (+), and the other, levorotation (−).

Prochiral (Prostereoisomerism)
Describing a molecule that yields a chiral structure by replacement of one of two chemically like groups by a different group. Thus prochiral Caabc yields chiral Cabcd by replacement of one "a" by "d".

Racemic
A mixture of equal amounts of two enantiomers of a compound; this mixture is without optical activity. Hence, racemization, the conversion of an enantiomer to a racemic mixture, is characterized by the time-dependent loss of optical activity.

R/S Notation
An unambiguous method to specify configuration of chiral compounds on the basis of the ordering of atoms or groups of atoms (by a sequence rule) when the molecular model is viewed from a defined direction.

Stereoisomer
Isomers differing only in the spatial arrangement of their atoms.

An object, structure, or molecule for which the image in a plane mirror cannot be superposed on the original is chiral; for example, the two human hands. Chiral structures or growth forms are common in nature and in manufactured objects. A typical chiral molecule contains a carbon atom linked to four different groups, Cabcd. Such molecules exist in two forms of opposite handedness that are termed *enantiomers*. Most metabolites are chiral, and enzymes usually exhibit a high degree of specificity toward enantiomers. A molecule containing two chemically like "a" groups, Caabc, is described as prochiral; replacement of one "a" group by "d" leads to a chiral molecule, Cabcd. Enzymes usually discriminate between the two chemically like groups in a prochiral molecule. Enantiomers can have very different physiological actions, for example, taste or odor. With chiral pharmaceutical agents, the enantiomers may also have different physiological actions. Physiological differences arise from the phenomenon of chiral recognition – the ability of biological receptors, enzymes, and some catalysts and reagents to discriminate between the enantiomers of a given compound. Although many hypotheses account for the homochirality observed on planet Earth, the original process or processes leading to this situation remain unclear.

1
Chirality – An Introduction

The earliest people must have recognized that one human hand, although very similar, cannot be superposed on the other hand and that the reflected image (e.g. on a water surface or in a mirror) of one hand was superposable on the other. In modern terminology, the two human hands are in a mirror-image relationship. The "handedness" of biological structures, molecules, objects, and so on, generates much interest. For understanding, it is necessary to define right and left. Although children are taught to distinguish hands, feet, eyes, ears, and so on, as right or left, a precise definition of right or left in terms of direction is difficult if not impossible. The normal definition is a statistical one – the right hand is the stronger, or that which is used preferentially by the majority. Most of the world's population favors the right hand, and populations with the majority preferentially using the left hand are not known.

A comprehensive terminology describing handedness is based on the words chiral (adjective) and chirality (noun) introduced by Lord Kelvin in 1894 as follows: "I call any geometric figure, or group of points *chiral* and say that it has *chirality*, if its image in a plane mirror, ideally realized, cannot be brought to coincide with itself." Beginning in about 1958, they began to be very widely used. Although based on the Greek word "$\chi\varepsilon\iota\rho$" for hand, the definition of chirality requires only the observation of a mirror image without requiring knowledge of right or left. The two human hands are chiral regardless of their names.

While chirality usually involves three-dimensional structures, it is possible in two dimensions. Thus, the mirror image of the letter F cannot be superposed on the original by translation or rotation in the plane; F is chiral in two dimensions. The letter E is achiral; mirror image and original are superposable.

Chirality, a very pervasive phenomenon in biology, operates at two levels. The first is that of easily observed structures such as the hands. The second concerns atoms and molecules. Since the macrobiological structures depend on enzyme reactions for their formation, the chirality of metabolites, both large and small, determines the chirality of the larger structures. Hence, a brief review of general stereochemical principles will be given before discussing the biological implications. In this article, all of the common amino acids will in some cases be described by the conventional three-letter abbreviations.

2
Molecular Chirality in Biology

2.1
Chirality Descriptors: the *R*/*S* Notation

Isomers are compounds having the same molecular formula but differing in the nature or sequence of bonding of their atoms, or in the spatial arrangement of their atoms. Those isomers differing in their spatial atomic arrangements are called *stereoisomers*. Stereoisomers showing a mirror-image relationship are termed *enantiomers*, whereas stereoisomers not in a mirror-image relationship are *diastereoisomers*. Enantiomers show the same general chemical properties but differ in their interaction with polarized light. One enantiomer will produce a right-handed dextro (+) rotation, the other a left-handed levo (−) rotation. Enantiomers frequently exhibit different properties in biological systems.

Diastereoisomers differ in their general chemical properties.

The major biological concern is with the chirality of carbon compounds. The simplest organic chiral compound is formed by a carbon atom linked to four different noncarbon atoms as in CHBrClF. The spatial arrangements in such a compound are termed *configurations,* and if the precise arrangement is known, it is described as an absolute configuration (arrangements resulting from rotation about single bonds are excluded). Absolute configurations are specified unambiguously by the Cahn–Ingold–Prelog system. The four different atoms or groups of atoms attached to the carbon atom are first arranged in a carefully described priority sequence. The "sequence rules" are based on parameters such as atomic number and atomic mass. To be unambiguous, they are complex but are readily available in texts of organic chemistry or stereochemistry. For many groups of biological interest, the following sequence will be adequate; a > b indicates that a has a higher priority than b:

$$\mathrm{Br} > \mathrm{Cl} > \mathrm{SH} > \mathrm{F} > \mathrm{OCH_3} > \mathrm{OH}$$
$$> \mathrm{NHCOCH_3} > \mathrm{NH_2} > \mathrm{COOR} > \mathrm{COOH}$$
$$> \mathrm{CHO} > \mathrm{CH_2OH} > \mathrm{C_6H_5} > \mathrm{C_2H_5}$$
$$> \mathrm{CH_3} > {}^3\mathrm{H} > {}^2\mathrm{H} > {}^1\mathrm{H}$$

A model of the molecule is then viewed with the atom or group of lowest priority pointing away from the observer. If the remaining atoms or groups of atoms, when examined in order of decreasing priority, show a clockwise or right-handed arrangement, the absolute configuration is denoted as *R* (rectus). A counterclockwise, left-handed path yields the *S* (sinister) absolute configuration. For CHBrClF, the priority is Br > Cl > F > H, and the two possible enantiomeric models are shown in Fig. 1(a, b). The (*R*) enantiomer (Fig. 1a)

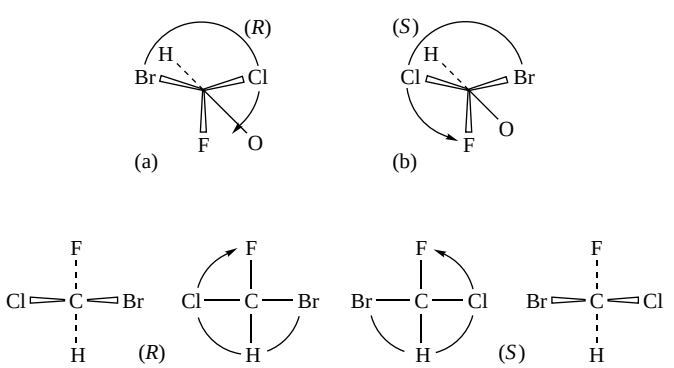

Fig. 1 Enantiomers of CHBrClF. In this and subsequent figures a "dashed" line indicates a bond below the paper plane and a "wedge" shape indicates a bond above the paper plane. Structures (a) and (b) show the molecular models viewed to determine (*R*) or (*S*) configuration; O = eye of observer. Note that the sequence, Br → Cl → F, describes a right-handed path in (a) and a left-handed path in (b). The bottom drawings determine configuration using projection drawings with the group of lowest priority, H, at the bottom. For each enantiomer, there is a three-dimensional representation and a conventional projection formula.

is levorotatory and, conversely, the (*S*) enantiomer (Fig. 1b) is dextrorotatory.

Three-dimensional molecular arrangements are often represented as two-dimensional "projection" drawings by carefully following conventions established by Emil Fischer. In a projection formula, a vertical line indicates a bond pointing below the plane of the paper, while a horizontal line indicates a bond pointing above the plane of the paper; see CHBrClF enantiomers (Fig. 1). If the group of lowest priority is at the bottom of a projection formula, the configuration can be determined from the handedness of the remaining three groups (Fig. 1).

For chirality specification, any projection formula can be transformed so that the group of lowest priority is at the bottom; hence the handedness can be ascertained as just described. This transformation must involve two interchanges of any two groups; such a process retains the original configuration (a single interchange of any two groups inverts the original configuration). For example, (+)-glyceraldehyde is conventionally written with the most oxidized group located at the top (Fig. 2a). The priority sequence is OH > CHO > CH_2OH > H. If the groups H and CH_2OH are interchanged so that the lowest priority H is at the bottom, a structure with inverted configuration is obtained (Fig. 2b). A second interchange of CHO and OH restores the original configuration (Fig. 2c). "Reading" the handedness of the sequence, OH > CHO > CH_2OH, gives the configuration of (+)-glyceraldehyde as (*R*).

2.2 Chirality Descriptors: the D/L Notation

Certain groups of metabolites having structures with identical handedness – α-amino acids, α-hydroxy acids, and carbohydrates – are often described by the D/L notation. A projection formula is drawn with the most oxidized group (usually CHO or COOH) placed at the top. The next carbon atom in the vertical chain has invariant groups, H and NH_2 for amino acids, and H and OH for hydroxy acids. Two configurational "standards", glyceraldehyde and serine, are used (Fig. 3). If in such formulae, the NH_2 or OH group is to the right, the configuration is assigned as D; if to the left, the assignment is L. If necessary, the modifications D_g and D_s are used to emphasize a relationship with D-glyceraldehyde or D-serine. D/L descriptors refer only to configurations and not to optical rotation. The classical 20 protein amino acids have the same L-configuration

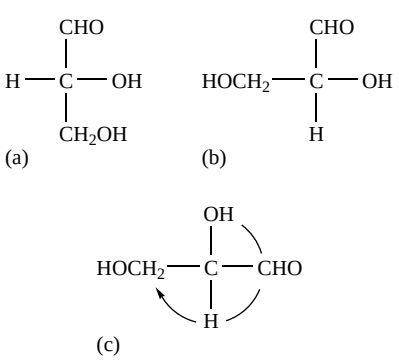

Fig. 2 Transformation of projection formulae. The structure drawn is that of (+)-glyceraldehyde. One interchange of two groups, (a) → (b), inverts the configuration, and a second, (b) → (c), restores the original configuration. As shown in (c), (+)-glyceraldehyde has the (*R*) configuration.

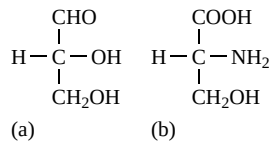

Fig. 3 Configurational standards.
(a) = D-glyceraldehyde [also (*R*), see Fig. 2] and (b) = D-serine.

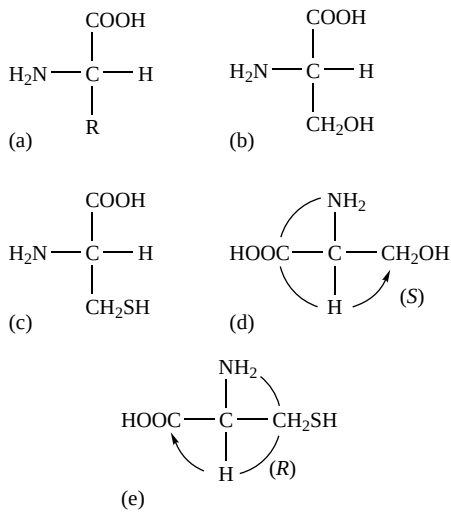

Fig. 4 Amino acid configurations.
(a) = L-configuration of the 19 chiral, protein amino acids: R indicates the side chains; (b) = L-serine; (c) = L-cysteine; (d) = L-serine; also (*S*)-serine; (e) = L-cysteine, also (*R*)-cysteine.

(Fig. 4a). As indicated, L-serine also has the (*S*) configuration (Fig. 4b, d). Because of the vagaries of the sequence rule, not all L-amino acids are (*S*); for instance, L-cysteine has (*R*) configuration (Fig. 4c, e).

Carbohydrates have multiple carbon atoms carrying H and OH groups. D/L assignments are based on the arrangement in a projection formula of the H and OH groups at the highest numbered chiral carbon; the numbering starts with the most oxidized group, CHO, C-1 (Fig. 5). Projection formulae for carbohydrates are not informative since the linear carbon chain would normally have a zigzag arrangement; they were devised many decades before computer drawing programs for chemical structures were available. More realistic representations use either vertical or horizontal zigzag carbon chains (Fig. 5d, e). Since carbohydrates usually adopt cyclic structures in aqueous solutions, alternate structures are necessary (e.g. Haworth or Mills structures).

2.3 Axial Chirality

The chirality type so far described is termed *central chirality*. Another type, axial chirality, occurs in planar structures and in compounds of the allene type, abC=C=Cab. Some natural products are allenes; one example is mycomycin: $HC{\equiv}C{-}C{\equiv}C{-}CH{=}C{=}CH{-}CH{=}CH{-}CH{=}CH{-}CH_2{-}COOH$. This compound is a toxic antibiotic. Moreover, a sex attractant for male Dried Bean beetles is $CH_3{-}(CH_2)_6{-}CH_2{-}CH{=}C{=}CH{-}CH{=}CH{-}COOCH_3$. Grasshopper ketone,

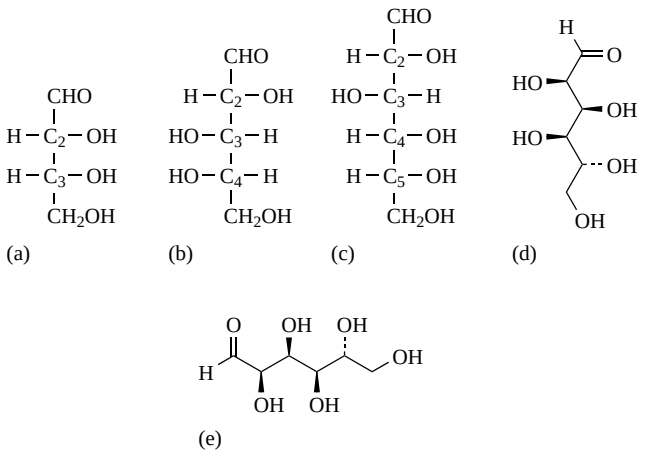

Fig. 5 Carbohydrate configurations. (a) = D-erythrose; (b) = L-arabinose; (c) = D-glucose. The D/L assignments are based on the configurations at C-3 for (a), C-4 for (b), and C-5 for (c); these are the chiral carbon atoms with the highest number. For D-glucose, zigzag configurations are shown as (d) (vertical) and (e) (horizontal).

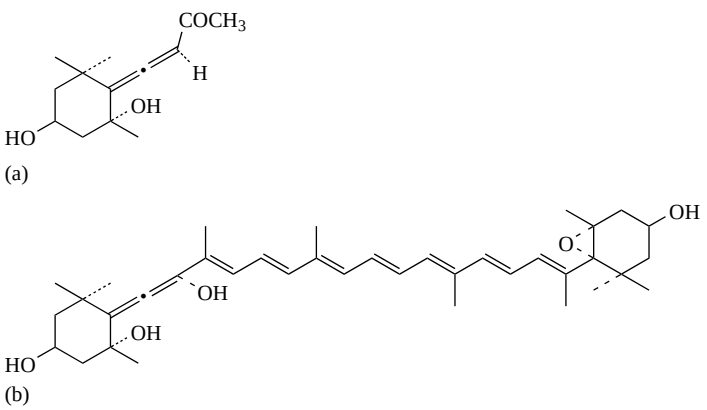

Fig. 6 Naturally occurring allenes. (a) = "grasshopper ketone"; (b) = neoxanthin (from *Lycopersicon* roots).

(Fig. 6a), is used by flightless grasshoppers to repel predatory ants.

Certain carotenoids such as neoxanthin (Fig. 6b) are allenes. Some eicosanoids form allene oxides from fatty acid hydroperoxides by the action of allene oxide synthase (EC 4.2.1.92, hydroperoxide dehydratase). In turn, these allene oxides can undergo cyclization to cyclopentenone derivatives by allene oxide cyclase (EC 5.3.99.6). A specific example of an allene oxide pathway is the biosynthesis of jasmonic acid from α-linolenic acid (Fig. 7). Certain tertiary α-allenic amines inhibit mitochondrial monoamine oxidase type B and the allene group chirality has

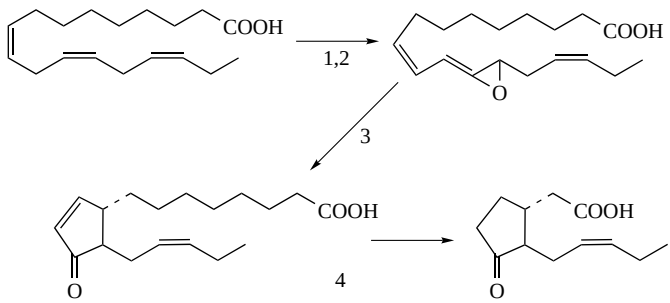

Fig. 7 Biosynthesis of jasmonic acid from α-linolenic acid via an allene oxide. Enzymes are (1) lipoxygenase; (2) allene oxide synthase; (3) allene oxide cyclase; (4) reduction and β-oxidation.

a profound action on the inactivation rate; (*R*)-allenes are up to 200-fold more potent than the (*S*)-allenes.

2.4 Amino Acids

Our biological world is often described as homochiral – thus, the protein amino acids belong to the L configurational series. If only weight is considered, this is certainly true since the tonnage of L-amino acids found in protein is overwhelming; about 18% of the wet weight of a mammal is protein. Of the 20 "classical" amino acids involved in ribosomal protein synthesis, glycine is achiral, 17 contain a single chiral center at the α-position, and threonine and isoleucine contain a second chiral center. The necessary aminoacyl-tRNA ligase enzymes are generally very specific for L-amino acids during protein synthesis. When examined *in vitro*, some reactivity with D-amino acids has been observed; for example, the tyr-tRNA ligase from *Escherichia coli* can utilize D-tyrosine.

The so-called 21st ribosomally incorporated amino acid is selenocysteine, $HSe-CH_2-CH(NH_2)-COOH$, found in bacteria, eukaryotes, Archae, and animals. Its formation is complex involving a unique material, $tRNA^{Sec}$. This is ligated with L-serine and the L-seryl-$tRNA^{Sec}$ is converted to selenocysteinyl-$tRNA^{Sec}$ with the aid of selenophosphate. Chirality at the α-position of L-serine is apparently lost in this pyridoxal phosphate-dependent process, but it is usually assumed that the selenocysteine in protein has the L-configuration. A 22nd ribosomally incorporated amino acid, pyrrolysine, present in certain Archae and eubacteria, is $HOOC-CH(NH_2)-(CH_2)_4-NH-CO-R$ (R is apparently (4*R*,5*R*)-4-substituted-pyrroline-5-carboxylate). The R group is attached via the carboxyl group at the $\varepsilon-NH_2$ of lysine. The substituent at position 4 could be methyl, ammonium, or hydroxyl. This amino acid is clearly of the L-configuration, derived from L-lysine. The pyrroline ring structure may be related to D-proline.

Some posttranslational racemization of L- to D-amino acids is possible in proteins. In long-lived proteins, such as lens crystallins, tooth dentine and enamel, and brain myelin, aspartate and serine racemize at a rate of about 0.15% per annum (for aspartate). This process can be used to determine the age of fossil proteins and of long-lived animals if care is taken to

prevent or correct for racemization during isolation.

Posttranslational modification also occurs in the synthesis of the heptapeptide, dermorphin, obtained from skin secretions of certain frogs. An all-L peptide is formed initially by the ribosomal mechanism but an L-alanine residue is converted to D-alanine, probably by dehydrogenation and rehydrogenation, to yield L-Tyr-D-Ala-L-Phe-Gly-L-Tyr-L-Pro-L-Ser (NH_2). The opioid activity of this peptide is greater than that of morphine. Similarly, ribosomally produced lantibiotics (from gram-positive bacteria) contain D-alanine and D-aminobutyric acid units. The material nisin, approved for use as a food preservative, contains 10 of 34 amino acid residues as one lanthionine unit and four β-methyllanthionine units. These units have the following structures; if R = H, the material is lanthionine and if R = CH_3, it is β-methyllanthione. The starred chiral center (*) has the D-configuration: HOOC−C*H(NH_2)−CH(R)−S−CH_2−CH(NH_2)−COOH. These D-residues may derive from L-serine (lanthionine) and L-threonine (β-methyl-lanthione) by dehydrogenation to dehydroalanine and dehydrobutyrine residues respectively, followed by addition of L-cysteine. The L-configuration is lost in the dehydrogenation and L-cysteine addition results in the formation of a D-configuration.

Serine has unusual roles. Several membrane proteins are "glutamate receptors" and mediate the effects of this amino acid in neurotransmission. One group termed *N*-methyl-D-aspartate (NMDA – see later) receptors plays a role in memory acquisition, learning, and neurological disorders. Both glycine and D-serine bind to the NMDA receptor. In brain, D-serine is formed by the action of a serine racemase, a pyridoxal phosphate-dependent protein. This brain racemase has been cloned from astrocytes and it is now clear that D-serine has an important role as a neurotransmitter. The incubation of D-serine-containing neurons with D-amino acid oxidase to destroy D-serine greatly reduces the activation of the NMDA receptor. The racemase activity requires divalent cations (e.g. Ca^{2+}, Mg^{2+}), and Ca^{2+} may regulate serine racemase activity. In mutant mice lacking D-amino acid oxidase activity, high levels of D-serine were observed in the forebrain (100–400 nmol g wet tissue^{-1}) with only low levels in the pituitary and pineal glands. The D-serine levels in the cerebellum and the medulla oblongata were 10-fold greater in mutants than in controls. For D-alanine, the levels in mutant mice were higher than those in controls for all brain regions.

Important roles are becoming apparent for D-aspartate. It had been used as a probe for high-affinity L-glutamate/L-aspartate uptake sites and in studies of L-glutamate transport. Its *N*-methyl derivative, *N*-methyl-D-aspartate (NMDA), was known to activate the NMDA receptors. These phenomena were treated as "unphysiological" since in mammalian metabolism, D-amino acids were regarded as "unnatural". However, free D-aspartate has now been found in the tissues of many invertebrates and vertebrates. It occurs in nervous tissues (chicken, rat, human), in embryo and adult brain and cerebrospinal fluid (humans), and in the pituitary gland and gonads (many mammals). One of the highest observed levels of D-aspartate, 114 ± 18 nmol g wet tissue^{-1}, was in rat adenohypophysis gland; high levels are also present in embryonic tissue.

NMDA itself has now been identified in neuroendocrine tissues. Previously, it was known to be present in the bivalve, *Scapharca broughtonii*. The NMDA levels in

rat nervous tissues and endocrine glands increase following administration of D-aspartate. Experiments with homogenates have shown an enzymatic methyl group transfer from *S*-adenosylmethionine to D-aspartate yielding NMDA. The highest level of transferase activity was in the hypothalamus with significant levels in rat hippocampus, adenohypophysis, brain, and liver.

D-aspartate and NMDA may influence hormone release in various cells. In the case of prolactin (PRL) and the prolactin releasing factor (PRF), NMDA, synthesized in both brain and hypothalamus, stimulates the release of PRF in the hypothalamus. In turn, PRL is then released in the hypophysis. D-aspartate also acts directly on the adenohypophysis, further reinforcing PRL release. D-aspartate occurs at the level of 120 nmol mL^{-1} in testicular venous blood plasma with lower levels in other components; the values are all higher than those for peripheral blood plasma (6 nmol mL^{-1}). The distribution of D-aspartate in rat testes is quite different from that of testosterone. The role of D-aspartate in the male reproductive tract remains unclear, although it may be involved in steroid synthesis in rat testis.

D-Enantiomers of several amino acids occur in various species of hyperthermophilic *Archaea* (e.g. species of *Desulfurococcus*, *Pyrococcus*, and *Thermococcus*). Remarkably, in the case of aspartate, almost 50% of the total content was the D-enantiomer (for the ratio, D-aspartate/total D- + L-aspartate, values ranged from 43.0 to 49.1%). Significant levels of racemization were also observed for alanine, leucine, lysine, and phenylalanine. Aspartate racemase activity was detected in crude extracts from various strains and an aspartate racemase gene was cloned and sequenced from *Thermococcus sp.* strain KS-8. Since peptidoglycan – a usual source of bacterial D-alanine (see below) – is not present in the *Archaea*, the function of D-amino acid enantiomers in these organisms is unclear. Several D-amino acids occur in insects. D-Alanine is present in hemolymph of milkweed bugs (*Oncopeltus fasciatus*), and free and combined forms of D-serine occur in silkworms and earthworms.

D-Amino acids function in the complex structures of some bacterial cell walls. These peptidoglycans (mureins) are polysaccharides with alternating units of *N*-acetyl-D-glucosamine and *N*-acetylmuramic acid. The latter units are linked to a tetrapeptide containing at least one D-amino acid. In *Staphylococcus aureus*, the tetrapeptide is L-ala-D-isoglu-L-lys-D-ala (in isoglutamate, the peptide bond is formed with the γ-carboxyl group). In *E. coli*, *meso*-diaminopimelate replaces L-lysine, and depending on the specific organism, other variations are possible. The capsular substance of anthrax bacillus (*Bacillus anthracis*), *Bacillus subtilis*, and other bacteria consists entirely of poly-D-glutamate.

D-Amino acids also occur in relatively small peptides and related structures produced by microorganisms, for example, penicillin, in which the penicillamine unit ($\beta\beta$-dimethylcysteine) has the D-configuration. Penicillin has three chiral centers; two of them derive from a tripeptide precursor, δ(L-α-aminoadipoyl)-L-cys-D-valine. The gramicidins are antibiotic bacterial peptides synthesized by a multienzyme complex rather than by the ribosomal mechanism. In gramicidin A, a channel-forming ionophore, the peptide chain, contains 15 amino acid residues; with the exception of an achiral glycine unit, units with D- and L-configurations

alternate:

L-Val-Gly-L-Ala-D-Leu-L-Ala-D-Val-L-Val-

D-Val-(L-Trp-D-Leu)$_3$-L-Trp

The N-terminal L-valine is modified by a formyl group and the C-terminal L-tryptophan by ethanolamine. In the antibiotic, actinomycin D (*Streptomyces* strains) there are two D-valine components and the unusual amino acids, sarcosine, and *N*-methyl-valine. Bacitracin A, produced by *B. subtilis* and *B. licheniformis* and present in some nonprescription antibiotic creams, is a dodecapeptide with 4 D-amino acids (aspartate, glutamate, ornithine, phenylalanine). A very simple antibiotic with a single chiral center is D-cycloserine (D-4-amino-3-isoxazolidinone), $C_3H_6N_2O_2$. It is produced by a *Streptomycete sp.*, ultimately from L-serine by way of a ureido derivative:

L-Ser →→ *O*-ureido-L-ser →

O-ureido-D-ser → D-cycloserine.

2.5 Carbohydrates

Carbohydrates are also cited as examples of homochirality. Again, the sheer weight of D-glucose, D-ribose, and D-deoxyribose on planet Earth is enormous; however, there are substantial amounts of some L-carbohydrates, for example, L-arabinose. There have been several chemical or biochemical syntheses of "unnatural" carbohydrates, for example, L-glucose, L-galactose, and L-ribose. A biochemical method involves the use of D-galactose oxidase, EC 1.1.3.9. The type reaction is the oxidation of D-galactose at C-6, thus producing D-*galacto*-hexodialdose and H_2O_2, but some other alcohols are oxidized. To prepare L-galactose, the *meso* compound, galactitol (Fig. 8b, $R^1 = OH$, $R^2 = H$) is derived by reduction of D-galactose (Fig. 8a, $R^1 = OH$, $R^2 = H$). Treatment of galactitol with galactose oxidase yields L-galactose (Fig. 8c, $R^1 = OH$, $R^2 = H$). This provides a two-step racemization, D-galactose → L-galactose. Oxidation of L(+)-glucitol (Fig. 8b, $R^1 = H$, $R^2 = OH$) yields L-glucose (Fig. 8c, $R^1 = H$, $R^2 = OH$). L(+)-glucitol is prepared by reduction of the readily available D-gulonolactone (Fig. 8a, $R^1 = H$, $R^2 = OH$). The oxidation efficiencies were increased by addition of catalytic amounts of ferricyanide.

L-Glucose has, apparently, not been isolated as a natural product; however, some microorganisms can oxidize it. Very few physiological actions have been attributed to it or its derivatives. Thus, penta-*O*-acetyl-β-L-glucopyranose caused a biphasic increase in plasma insulin concentration

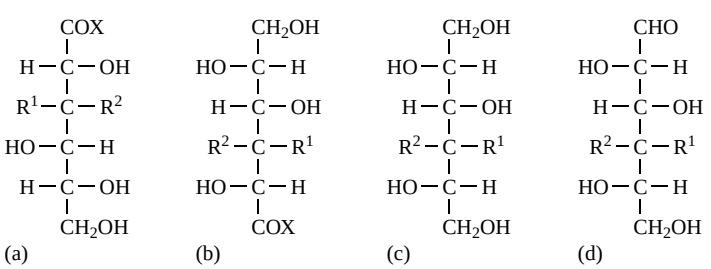

Fig. 8 Synthesis of L-galactose and L-glucose. In the second structure from the left, (a) has been rotated through 180°. See text for details.

when injected into anesthetized rats. In mice, L-glucose increased production of advanced glycation end products and there was accelerated aging. Mice injected with L-glucose also showed an improved memory effect reversed by peripherally acting cholinergic drugs. The L-glucose effect may involve facilitation of acetylcholine synthesis and/or its release. Commercially available L-glucose is used as a nonmetabolizable analog for D-glucose in transport and other metabolic studies. The derivative, *N*-methyl-L-glucosamine, is a component of streptomycin and is biosynthesized from D-glucose.

D-Galactose predominates with L-galactose derivatives in some polysaccharides. The gelling material, agar-agar, contains about 70% of agarose. In this heteropolymer, units with D- and L-configurations alternate; the two monomers are β-D-galactopyranose and 3,6-anhydro-α-L-galactopyranose. Liganded agarose derivatives are used in affinity chromatography. L-Galactose has a role in ascorbic acid biosynthesis in plants (see later). Both enantiomers of 6-deoxygalactose, trivial name fucose, occur widely – D-fucose in plant glycosides and L-fucose in seaweed polysaccharides and in the cell wall matrix of higher plants. L-Fucosyl residues are also present in animal glycoproteins and glycolipids including human erythrocyte surface antigens.

Among other hexoses and their derivatives, residues of L-iduronic acid are present in chondroitin sulfate B (dermatan sulfate), a glycosaminoglycan present in ground substances, connective tissue, and so on; a second component is *N*-acetyl-D-galactosamine. Alginic acid – a hydrophilic polysaccharide found in brown seaweeds, especially the California giant kelp (*Macrocystis pyrifera*) and horsetail kelp (*Laminaria digitata*) – has a complex interrupted structure of stretches of α-L-gulopyranosyluronic acid and β-D-mannopyranosyluronic acid residues. The sodium salt is used extensively as an emulsifier and thickener.

The rather rare 6-deoxytalose, 6-deoxy-3-*O*-methylglucose, and fucosamine occur as both enantiomeric forms. In two cases, enantiomers have different trivial names; the D- and L-forms of 3,6-dideoxygalactose are respectively abequose (*Salmonella* lipopolysaccharides) and colitose (*E. coli* lipopolysaccharides). The corresponding forms of 3,6-dideoxymannose are tyvelose (lipopolysaccharides of some gram-negative bacteria) and ascarylose (a glycolipid component in *Ascaris* eggs and *Yersinia pseudotuberculosis*).

The pentoses, D-ribose and 2-deoxy-D-ribose, occur in large amounts in nucleic acid structures. Although generally assumed not to occur in nature, isolation of L-ribose from water-soluble polysaccharides of *Adonis vernalis* (pheasant's eye, ox's eye) has been reported. Mutants of *E. coli* and some other bacteria metabolize L-ribose by an NADPH-linked reductase to ribitol; subsequently, ribitol pathway enzymes convert ribitol to D-ribulose (D-*erythro*-pentulose). An L-ribose isomerase, converting L-ribose to L-ribulose (L-*erythro*-pentulose), has been isolated from *Acinetobacter sp*. The gene for this enzyme was cloned into *E. coli* and sequenced; the recombinant enzyme was purified to homogeneity. Further metabolism presumably involved the pentose phosphate phosphoketolase pathway: L-ribulose $\rightarrow$ L-ribulose 5-phosphate $\rightarrow$ D-xylulose 5-phosphate $\rightarrow$ acetyl phosphate + glyceraldehyde 3-phosphate.

The next most abundant pentose is D-xylose, present in the xylans; xylan is the

next most abundant carbohydrate after cellulose. The rare L-xylose occurs in some antibiotics and lipopolysaccharides (e.g. of *Xanthomonas campestris*). With arabinose, the L-enantiomer is common in many plant products (e.g. gums, hemicelluloses, and glycosides) and in bacterial polysaccharides. The also-rare D-enantiomer occurs in some plant glycosides and bacterial polysaccharides. D-Lyxose is very rare but is present in alkali-labile glycolipids of *Mycobacterium smegmatis*. Naturally occurring derivatives of L-lyxose are 2-*O*-methyl-lyxose, a component of antibiotics such as everninomycin B and D, and 3-*O*-methyl-lyxose, present in lipopolysaccharides of *Pseudomonas maltophilia*. The streptomycin component, L-streptose, is 5-deoxy-3-formyl-L-lyxose, and a material in *Coxiella burnettii* lipopolysaccharides is 3-*C*-hydroxymethyl-L-lyxose (dihydrodihydroxystreptose). Some strains of *E. coli* utilize the lyxose enantiomers. One microorganism, *Aerobacter aerogenes* strain PRL-R3, utilizes all eight aldopentoses (i.e. four each of the D- and L-configurations) and the four pentitols (D- and L-arabinitol, ribitol, xylitol) as well.

Vitamin C (L-ascorbic acid, 2-keto-L-gulono-γ-lactone) occurs only as the L-enantiomer. It is biosynthesized in plants by way of L-galactose as follows: D-glucose 6-phosphate → GDP-D-mannose → GDP-L-galactose → L-galactose 1-phosphate → L-galactose → L-galactono-1,4-lactone → L-ascorbic acid. *Saccharomyces cerevisiae* synthesizes L-ascorbic acid when incubated with L-galactose, L-galactono-1,4-lactone, and L-gulono-1,4-lactone. However, incubation with D-glucose, D-galactose, D-mannose, or D-arabinose gave D-erythroascorbic acid, a five-carbon analog of L-ascorbic acid.

Animals use a different pathway to make vitamin C and in some, the capacity is lacking (e.g. primates and guinea pigs). In the rat, conversion of D-glucose to L-ascorbic acid requires "inversion" of the carbon chain sequence so that C_1 to C_6 of D-glucose appear as C_6 through C_1 in ascorbate.

2.6 Other Natural Products

Many natural products occur both as enantiomers and, sometimes, as racemates or *meso* structures; only a few examples can be given. The case of tartaric acid is typical. In the nineteenth century, a material named paratartaric acid was sometimes formed in wine production, and Pasteur separated crystals of sodium ammonium paratartrate into two enantiomorphic forms by "crystal picking." One of the free tartaric acids obtained from the salts was dextrorotatory, the other levorotatory. In other words, paratartaric acid was a racemic DL mixture. The L(+) (or 2*R*,3*R*) enantiomer is found in some fungi and bacteria and is particularly abundant in grape juice. The D(−) (or 2*S*,S3) enantiomer and the *meso* form have limited distribution.

Lactic acid formed in muscle glycolysis is the L(+) enantiomer, (*S*)-(+)-2-hydroxypropanoic acid. D(−)-lactic acid is present in bacteria as the peptidoglycan component, *N*-acetylmuramic acid, muramic acid being the 3-*O*-D-lactyl ether of D-glucosamine. Racemic DL acid is found in sour milk, molasses, and some fruit juices. Bacterial lactate fermentations produce either the D- or L-enantiomer or the racemate depending on the organisms used.

Alkaloids, terpenes, and many other secondary metabolites frequently occur in both enantiomeric forms or as racemates. Examples of alkaloids that occur as (+), (−), and racemic forms are the tobacco

alkaloid, nornicotine, and the lupinane alkaloids, lupanine and sparteine. The important alkaloid, reticuline, is involved in a number of biosynthetic processes. (*S*)-(+)-Reticuline occurs in *Anona reticulata* and other plants and is an intermediate for biosynthesis of, among others, berberine, chelidonine, laudanosine, and narcotine. *Papaver somniferum* contains both enantiomers depending on the age of the plants, and (*R*)-(−)-reticuline is the major precursor in morphine alkaloid biosynthesis.

Among terpenes, it is almost the rule rather than the exception for materials to occur in enantiomeric and/or racemic forms. Common materials such as borneol, camphor, carvone, citronellol, fenchone, limonene, menthol, pinene, and terpineol occur as both enantiomers; racemates are known for borneol, camphor, and limonene. Some biosynthetic information is available. Purified enzymes from sage (*Salvia officinalis*) and tansy (*Tanacetum vulgare*) convert achiral geranyl bisphosphate to a linalyl bisphosphate, LPP, probably as an enzyme-bound intermediate. The sage system forms (3*R*)-(−)-LPP and tansy, (3*S*)-(+)-LPP. Subsequently, (3*R*)-(−)-LPP is converted to (1*R*,4*R*)-(+)-bornyl bisphosphate and the (3*S*)-(+)-LPP to the (1*S*,4*S*) enantiomer. Hydrolysis of the bisphosphates yields the two enantiomeric borneols. Two further enzymes from sage form enantiomeric pinenes. The (+)-pinene cyclase converts achiral geranyl bisphosphate to (+)-α-pinene and (+)-camphene, and the (−)-pinene cyclase forms mainly (−)-α-pinene, (−)-β-pinene and (−)-camphene.

Since racemization may occur during isolation of chiral natural products, precautions are necessary. Hyoscyamine readily undergoes racemization in the tropic acid component and partial racemization occurs during isolation. For pharmaceutical use, racemization is completed by alkali treatment yielding the widely used atropine (racemic hyoscyamine containing (±)-tropic acid). The mydriatic action resides only in the (−)-tropic acid component.

A few natural products are "quasi-racemates," a mixture of equimolar amounts of two substances with slightly different structures and of opposite chirality. An early example was a benzoquinone from *Dalbergia nigra* (Fig. 9a, b). Narcisamine, from King Alfred daffodils, was a quasi-racemate of (−)-demethylgalantha-

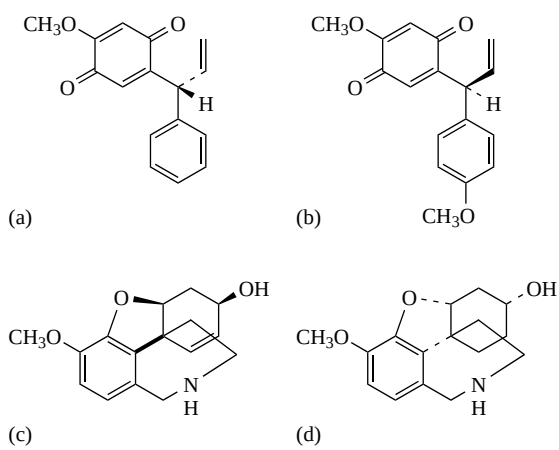

Fig. 9 Naturally occurring quasi-racemates. (a), (b) = benzoquinone from *Dalbergia nigra*; (c), (d) = narcissamine from King Alfred daffodils.

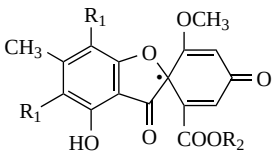

Fig. 10 Spiro polyketide metabolites. The chiral, spiro carbon is indicated by the symbol ..

mine and (+)-demethyldihydrogalanthamine (Fig. 9c, d).

In some polyketide metabolites of fungi, the chiral center is a spiro carbon atom (Fig. 10). When $R_1 = Cl$, $R_2 = H$, the compound is named *erdin* and occurs as a racemate in *Aspergillus terreus*; this organism also produces geodin, $R_1 = Cl$, $R_2 = CH_3$, as the (+) enantiomer. Bisdechlorogeodin, $R_1 = H$, $R_2 = CH_3$, occurs as the (+) enantiomer in *Penicillium frequentans* and as the (−) enantiomer in *Oospora sulfurea-ochracea*.

3 Physiological Responses to Enantiomers

Of the human senses, probably only hearing lacks a connection with chirality. Enantiomeric differences are common for taste and smell, and sight and touch can detect mirror-image relationships.

3.1 Taste

Classically, taste qualities in mammals included sweet, bitter, salty, and sour. The quality "umami" has now been added, describing the meatlike, savory taste of L-glutamate. Enantiomeric taste differences for amino acids were first noted in 1886 when D-asparagine was found to be sweet, while L-asparagine was tasteless. For the following, the D-enantiomers are sweet, the L-forms being bitter or flat: histidine leucine, phenylalanine, tryptophan, and tyrosine. Alanine is an exception since the L-enantiomer is sweet. Arginine, aspartate, isoleucine, lysine, proline, serine, and threonine are essentially tasteless, for both enantiomers. Achiral glycine is sweet as the name implies.

The remarkable umami taste is responsible for much of the flavor of soy sauce and monosodium L-glutamate, MSG, which is widely used in the food industries. In amounts greater than about 3.0 g, L-glutamate may show other physiological effects since some individuals experience the "Chinese restaurant syndrome" – a general feeling of tightness, facial pressure, burning sensations, and headache. In contrast, D-glutamate with at most a flat taste is relatively inert physiologically.

Another remarkable taste quality is that of the methyl ester of *N*-L-α-aspartyl-L-phenylalanine (trade name, Aspartame), a material some 160 times as sweet as sucrose. Interestingly, rodents do not react to Aspartame. It is widely used but should be avoided by individuals with phenylketonuria. The DD enantiomer, as well as the DL and LD diastereoisomers have bitter tastes.

Specific taste receptors have been identified in humans and animals. Receptors designated as TRB (taste receptor, family B) are G-protein-coupled receptors (GPCR); human receptors were identified by searching for GPCR-encoding genes in human genome regions implicated in taste perception.

A receptor in mammals, designated T1R1 +3, is a heteromer of the taste-specific T1R1 and T1R3 GPCRs. This receptor responds to the classical 19 chiral amino acids as L- but not D-forms, and the effect is potentiated by IMP; it is probably a constituent of the umami response. Sequence differences in T1R receptors between human and mouse influence the selectivity and specificity of the taste response. A different receptor, T1R2 +3 is activated by those D-amino acids that taste sweet and attract mice; L-amino acids did not cause this activation.

A complex system occurs in a preparation from catfish (*Ictalurus punctatus*) taste epithelium where L-alanine is a potent taste stimulus; D-alanine has a lesser effect. For a portion of the observed response, there may be a common receptor/transduction process, but in addition there appear to be independent processes for each enantiomer. The Caribbean spiny lobster (*Panulirus argus*) contains populations of chemosensory receptors that are differentially sensitive to D- and L-alanine. Of 77 neurons tested, 44% and 34% were classified as L-alanine and D-alanine-sensitive respectively.

Taste differences are generally rather subtle, especially for carbohydrates. While sucrose might be regarded as the epitome of sweet substances, it has been reported that "at the threshold level, sucrose yields a tactual sensation, which may be called bitter, medicinal, sour, and so on, but as the concentration is increased, the taste becomes bitter, then bitter-sweet, and finally, purely sweet." Apparently, "enantio-sucrose" (containing L-glucose and L-fructose) has never been prepared or tasted. Carbohydrate solutions are generally complex equilibrium mixtures of anomeric pyranose/furanose and straight-chain forms. Some taste differences for anomers and equilibrium solutions are known. There are also temperature and concentration effects (e.g. with sucrose) and considerable variation from one individual to another.

There is no strong evidence for general enantiomeric differences among carbohydrates. Thus, L-mannose was rated less sweet than D-mannose by five individuals, but three others said the reverse and still others found D-mannose to be bitter. Identical "sweetness scores" have been claimed for enantiomers of glucose, mannose, galactose, fructose, and xylose. In fact, there have been proposals to use L-hexoses as nonnutritive sweetners. While syntheses for L-glucose are available (see earlier) they are presumably too expensive to compete with materials such as aspartame. Penta-*O*-acetyl derivatives of various monosaccharides are not sweet; the following were actually rated as bitter: penta-*O*-acetyl-D-glucose (both anomers), penta-*O*-acetyl-α-D-mannopyranose, and penta-*O*-acetyl-β-L-glucopyranose.

"Pseudo"-carbohydrates, in which $-CH_2-$ replaces $-O-$ in the ring structure, also have sweet tastes, but again, there are ambiguities. Five people found the enantiomers of pseudo-β-fructopyranose to be as sweet as ordinary D-fructose; three out of five judged the D-enantiomer to be somewhat sweeter than the L. The thiopyranoid analogs of α-D-glucopyranose and β-D-fructopyranose are significantly sweeter than the usual forms.

3.2 Odor

The subject of enantiomeric odor differences has been controversial. Unfortunately, there is considerable variation in human olfactory sensitivity, and the nose is generally very sensitive to small

levels of impurities in materials under investigation. Testing differences also arise from variations in individual physiology, past experience with a given odor, and fatigue among members of a testing panel. Nevertheless, careful experiments have established a spearmint odor for (*R*)-carvone and a caraway odor for (*S*)-carvone. These materials are the major components of respectively oil of peppermint and oil of caraway. Important evidence was obtained by ingenious chemical inversions of one enantiomer into the other, the inversions resulting in very specific changes of odor; one sequence was (*R*)-(−)-carvone (spearmint) → (*S*)-(+)-carvone (caraway) → (*R*)-(−)-carvone (spearmint). The inversion was repeated similarly, starting with (*S*)-(+)-carvone.

Recent, careful testing has confirmed that the carvone enantiomers are distinguishable by odor, and this was also true for enantiomers of α-pinene and limonene. Most members of this panel (20 subjects), however, could not distinguish enantiomers of 2-butanol, camphor, β-citronellol, fenchone, menthol, rose oxide, and α-terpineol. The discriminatory ability between enantiomers is apparently substance-specific and is not a general phenomenon. There are probably specific enantioselective odor receptors for some but not all volatile enantiomers. Squirrel monkeys showed discriminatory abilities similar to those of humans; unlike humans, they could distinguish the fenchone enantiomers.

Rats readily discriminated the carvone enantiomers and, unlike humans, also distinguished enantiomers of fenchone, 2-butanol, and 2-octanol. In a study of the rat olfactory bulb, each enantiomer produced a widespread and distinct pattern of active glomeruli. Each member of an enantiomeric pair preferentially activated one or more glomeruli. In mole rats (*Spalax ehrenbergi*), some differential responses were observed to enantiomers of carvone, citronellol, and fenchone. The animals responded to one enantiomer but were indifferent to or did not smell the other [(*S*)-(+)-carvone, (−)-citronellol, (+)-fenchone]. (*R*)-(−)-Carvone was attractive to both sexes and females were attracted to (−)-fenchone. Both sexes were repelled by (+)-citronellol. Free-flying honeybees discriminated carvone, β-citronellol, limonene, menthol, and α-pinene enantiomers but not those of 2-butanol, camphor, fenchone, rose oxide, and α-terpineol.

For the *cis*-methyl jasmonates, two diastereoisomeric pairs have been tested (Fig. 11). Only the (3*R*,7*S*) and (3*R*,7*R*) diastereoisomers have an odor described as "typically floral, jasminic, slightly fruity," and the odor of natural "methyl jasmonate" is due largely to the (3*R*,7*S*)-(+) material (Fig. 11a). The (3*S*,7*R*) and (3*S*,7*S*) materials were odorless.

The "musk" odor is important in perfumery, and several synthetic odorants are used commercially. For the naturally occurring (3*R*)-(−)-muscone (Fig. 11b), the musk note was "very nice, rich, and powerful" but for the enantiomer, (3*S*)-(+)-muscone, the note was "poor and less strong." The compound shown in Fig. 11(d) "has an animalic tonality with camphoraceous aspects," while the enantiomer (Fig. 11e) had a musk note "with a sandalwood tonality." Both are found in *Angelica* root oil in a ratio of 29 : 72 for structures (d) and (e) (Fig. 11). In another case, (12*R*,9Z)-12-methyl-14-tetradec-9-enolide (trade name, Nirvanolide, Fig. 11c) possessed an "intense musky, fruity, powdery odor with lactonic nuances." Its enantiomer was odorless. The commercially important polycyclic

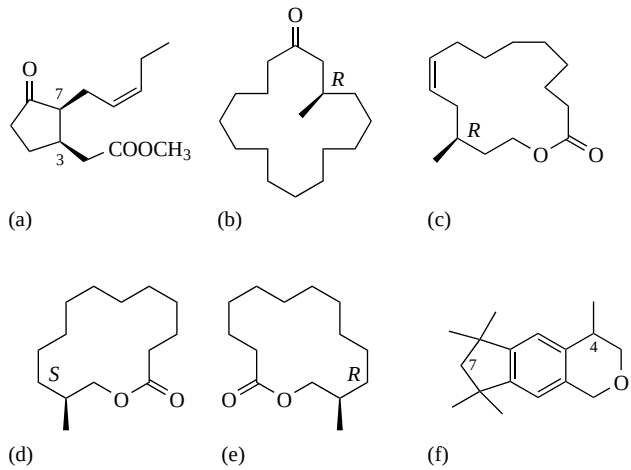

Fig. 11 Enantiomeric fragrances. (a) = (3*R*,7*S*) methyl jasmonate; (b) = (3*R*)-muscone; (c) = (12*R*)-Nirvanolide; (d), (e) = *Angelica* root oil compounds; (f) = Galaxolide.

musk odorant with the trade name Galaxolide (Fig. 11f) has four diastereoisomers rated as follows: (4*S*,7*R*)-(−) – very pleasant musk character; (4*R*,7*S*)-(+) – weak, uncharacteristic; (4*S*,7*S*)-(−) – musky, but dry aspects; and (4*R*,7*R*)-(+) – very weak, mainly fruity.

A new flavor material, 3-mercapto-2-methylpentanol, has an odor described as brothlike, leeklike, and sweaty – quite a combination! It has been isolated from volatiles of raw onions and in some processed foods. With two chiral centers, there are four stereoisomers; while all apparently have the same general odor, there are remarkable differences in their odor thresholds in water and air. The two *anti* enantiomers (2*R*,3*S*) and (2*S*,3*R*) (Fig. 12a) had thresholds in water of 0.03 to 0.04 μg L^{-1}. In air, the (2*R*,3*S*) material has an extremely low threshold, 0.00007 to 0.0002 ng L^{-1} and is "among the most potent flavor compounds." Its less-potent enantiomer has air threshold of 0.003 to 0.007 ng L^{-1}. The air threshold values for the *syn* enantiomers (2*R*,3*R*) and (2*S*,3*S*) (Fig. 12b) could not be determined since they were contaminated with 0.1 to 0.3% of the very potent *anti* compounds.

Another odorant with two chiral centers is "whiskey lactone," 2-methyl-4-octalone, extracted from oak woods during the

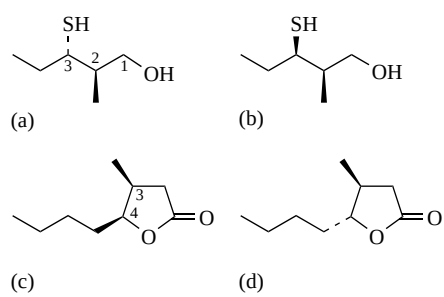

Fig. 12 Enantiomers of 3-mercapto-2-methylpentanol and "whiskey lactone." (a) = *anti* (2*R*,3*S*)-3-mercapto-2-methylpentanol; (b) = *syn* (2*R*,3*R*) diastereoisomer; (c) = (3*S*,4*S*)-2-methyl-4-octalone; (d) = (3*S*,4*R*) diastereoisomer.

aging process for wines and spirits. The major natural odor material is the (3*S*,4*S*)-*cis*-compound (Fig. 12c) with a coconut, lactonelike odor. Some flowery, lactonelike (3*S*,4*R*)-*trans*-component is also present (Fig. 12d). Substitution of one or both oxygen atoms with sulfur gave significant odor changes.

While many enantiomeric odor discriminations involve pleasant, floral, or fruity odors, nature also produces many foul odors. For instance the microbial product, geosmin, *trans*-1,10-dimethyl-*trans*-decalol, is known for its unpleasant moldy, earthy odor, with the (−) enantiomer being more odoriferous than the (+) form. The odor thresholds in water are also very different: (−), 9.5 ± 1.3 ng L^{-1}, and (+), 78 ± 12 ng L^{-1}. Enantio-MGDC analysis (see later) of wines with an earthy odor revealed significant levels of (−)-geosmin. This enantiomer was also found in some grape juices, probably from the growth of microorganisms.

Gas chromatography has been a valuable tool in studying the volatile components of the widely used essential oils. Column packings able to separate enantiomers (''chiral chromatography'') have greatly expanded the scope and value of this technique, as have ''hyphenated'' techniques, using mass spectrometry; for example, enantioselective multidimensional gas-chromatography mass spectrometry (enantio-MDGC-MS), and gas-chromatography pyrolysis-isotope mass spectrometry. One group, using various gas chromatographic techniques, has published a series of papers titled ''Stereoisomeric Flavor Components'' that reached number LXXXII in 1999.

These techniques yield important basic scientific information and have practical applications. By comparing an authentic natural oil sample with other samples, the presence of unnatural, often chemically synthetic, components can be detected. The important lavender oil has several odorants, including linalool and linalyl acetate; (*R*)-(−)-linalool has the characteristic lavender odor, (*S*)-(+)-linalool, has a petitgrain-like odor (i.e. an odor of bitter or sour orange). For these materials, genuine lavender oil must contain at least 85% enantiomeric excess of (*R*)-(−)-linalool and at least 95% enantiomeric excess of (*R*)-(−)-linalyl acetate. In a recent study, five samples of ''lavender oil'' were found to be adulterated with racemic, synthetic materials (see Table 1).

Tab. 1 Enantiospecific analysis of samples of oil of lavender.

Sample	*Linalyl acetate*		*Linalool*	
	(R)	*(S)*	*(R)*	*(S)*
1	52.9[a]	47.1	70.7	29.3
2	55.7	44.3	55.5	44.6
3	51.8	48.2	62.0	38.0
4	52.0	48.0	69.7	30.3
5	53.3	46.7	60.8	39.2
Authentic	95.0 (minimum)		85.0 (minimum)	

[a]The values given are percentage compositions.

Essential oils have been used in folk medicine, and it is a common belief that odors can influence mood, alertness, and sexual arousal (witness the appeal of aromatherapy). Some work concerns the effects of chiral fragrances on the human autonomic nervous system. In one study, inhalation of (+)-limonene (odor is orange) gave increased systolic blood pressure, subjective alertness, and restlessness; inhalation of (−)-limonene (odor is turpentine) increased systolic blood pressure without influencing psychological parameters. Similar responses were obtained with carvone enantiomers. The effects were complex and both pharmacological and psychological mechanisms were probably involved.

In studies of linalool enantiomers, differences in perception and responses elicited by a sensory test and forehead surface potential wave measurements were measured. (*R*)-(−)-linalool, administered after hearing environmental sounds, gave a favorable impression, and there was a tendency for the beta wave to decrease. With (*S*)-(+)-linalool, there was an unfavorable impression and the beta wave increased. Different and complex results were obtained following mental work. These preliminary results showed that the responses to the two enantiomers depended, in addition, on tasks assigned to the subjects.

Enantio-MDGC-MS has been used in some human metabolic studies. Methylcitric acid, MCA, occurs in body fluids of individuals with propionic acidemia, methylmalonic aciduria, or holocarboxylase deficiency. MCA has two chiral centers; of the four possible stereoisomers, only the (2*S*,3*R*) and (2*S*,3*S*) diastereoisomers were detected in patients and controls, in proportions varying from about 30 to 70%. The differences may be due to physiological effects on the citrate synthase enzyme forming MCA from oxaloacetate and propionyl-CoA.

3.3 Insect Pheromones

For chiral insect pheromones, the nature of the chirality generally has a profound effect on the observed physiological response. Thus, with the female spider sex pheromone, 3-hydroxybutyric acid, only the (3*R*) structure is functional. All of the following situations have been observed:

1. As just noted, activity is present in only one enantiomer, the other being indifferent (e.g. (+)-exobrevicomin, pine sawfly pheromone) or inhibitory (e.g. (+)-disparlure, (*S*)-*cis*-verbenol).
2. All stereoisomers are active (e.g. grandisol, seudenol).
3. Different species may differentiate enantiomers, for example, (*S*)-(+)-ipsdienol used by *Ips paraconfusus* and the *R* enantiomer by *I. calligraphus*. A more complex situation occurs with scarab beetles. For *Anomala osakana*, the female produces the male attractant, (*S*,*Z*)-5-(1-decenyl)oxacyclopentan-2-one. However, in *Popillia japonica* the (*R*) enantiomer is attractive to males. These activities are inhibited by the respective enantiomers. In field tests with *A. osakana*, catches of males in traps baited with the (*S*) enantiomer were completely inhibited by as little as 5% of the (*R*) form. The reciprocal behavioral agonist–antagonist situation provides a species-specific chemical signal with the use of a single chiral structure. Interestingly, these two beetles share a common habitat in Japan.

4. Activity may require both enantiomers. For honeybee queen mandibular pheromone (a mixture of compounds), both enantiomers of 9-hydroxy-2-decenoic acid must be present for full activity. Another example is sulcatol.

3.4 Plant Growth Regulators

With chiral plant growth regulators, one enantiomer of a chiral structure is usually the most potent. Thus, for α-phenoxypropionic acid, the (*R*) enantiomer has the most powerful herbicidal activity. The naturally occurring form of abscisic acid is (*S*), and the (*R*) form is significantly less potent in inhibiting root growth of germinating barley; there is little difference between these two enantiomers in other physiological effects. The growth regulator, paclobutrazol (β-[(4-chlorophenyl)-methyl]-α-(1,1-dimethylethyl)-1H-1,2,4-triazole-1-ethanol), also has antifungal activity. Of the four stereoisomers, the most effective in growth regulation is the (2*S*,3*S*)-(−) form (Fig. 13a), whereas the enantiomer (2*R*,3*R*)-(+) is most effective as a fungicide.

Jasmonic acid and its conjugates are plant signaling substances synthesized in response to external stimuli (e.g. wounding). The (3*R*,7*R*)-(−) enantiomer is particularly common (Fig. 13b). In barley leaves, treatment with jasmonic acid yields upregulation of specific genes and downregulation of housekeeping genes. The (−) enantiomers of the methyl ester and amino acid conjugates are more active than the (+) forms; conjugates with an L-amino acid also tend to be more active. The responses of the jasmonate-responsive genes to the various chiral structures are complex.

Physiological effects can vary with different strains of the same plant. For the synthetic auxin indole-3-succinic acid, a 10^{-8} M concentration of the (*R*) enantiomer is most potent as a root growth promoter for the SF (swingtime) strain of *Fuschia hybrida*, with the opposite enantiomer being more potent for the VF (variegated) strain at all concentrations.

3.5 Chiral Drugs

Reactions to pharmaceutical agents may be influenced significantly if chiral elements are present. Early in the twentieth century, enantiomers and racemates of hyoscyamine and epinephrine were distinguished by quantitative differences in their physiological actions. However, there was for many years little concern about the possible impact of chirality on widely used drugs. After World War II, attention was focused on the role of chirality in pharmacology and medicinal chemistry, and there is now an enormous literature including conferences and symposia on the topic of "chiral drugs." For instance, a query to PubMed on this topic

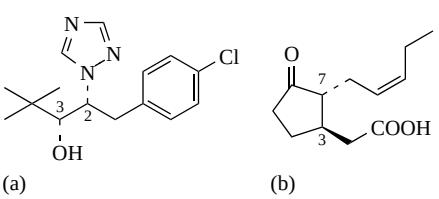

Fig. 13 Chiral plant growth regulators. (a) = (2*S*,3*S*) stereoisomer of paclobutrazol; (b) = (3*R*,7*R*)-jasmonic acid.

in October 2002 elicited 703 "hits." The topic is extremely complex and only a brief account is possible here. A complication is that pharmacologists consider two main areas – pharmacodynamics and pharmacokinetics. The first concerns actual physiological effects caused by a drug and the nature of the binding to receptor sites as well as correlation of chemical structure with physiological activity. Pharmacokinetics concerns differences in the rate of transport of the drug to the receptor sites, in clearance rates, and the possibility of transformation to other products that themselves might or might not have a physiological action.

Medicinal drugs are of three general types: some such as aspirin are achiral and of no concern here; chiral drugs obtained by chemical synthesis are usually racemic mixtures (e.g. fluoxetine, trade name Prozac); chiral drugs obtained from natural sources such as plants or from fermentations are usually single enantiomers (e.g. morphine, penicillin). There has been considerable interest among pharmaceutical companies in developing single-enantiomer forms or in switching from a racemic mixture already in use to a single enantiomer – the so-called *racemic switch* (a racemic switch may also provide a new drug when patents are expiring). The stakes are high. In 2001, worldwide sales of formulated pharmaceutical products totaled $410 billion; single-enantiomer drugs accounted for sales of $147 billion (36%).

In 1992, the US Food and Drug Administration, FDA, issued a policy statement for the development of stereoisomeric drugs that was mainly concerned with chiral drugs and the need for appropriate testing of enantiomers. However, under certain circumstances, the development and use of racemic mixtures was to be allowed. This FDA statement recognized three pharmacodynamic cases:

Case 1. Both drug enantiomers have similar, desirable pharmacological effects. This is the case with ibuprofen, a nonsteroidal antiinflammatory agent, sold as a racemate over the counter under various brand names (e.g. Advil, Motrin, Nuprin). It contains a single chiral center. Both enantiomers are antiinflammatory, but (*S*)-(+)-ibuprofen (Fig. 14a) is the more active and lacks some side effects. A racemic switch has led to the availability of (*S*)-ibuprofen in Austria and Switzerland, but this form was not approved for the United States. Ibuprofen also shows an interesting pharmacokinetic property. During metabolism, the (*R*)-(−) enantiomer undergoes configurational inversion forming the (*S*)-(+)-ibuprofen. Thus, administration of racemic ibuprofen to humans leads to excreted material in which the enantiomer composition is 70% (*S*) and 30% (*R*).

Another Case 1 situation is warfarin. When used as an anticoagulant in humans, the enantiomers are equipotent. However, warfarin is also used as a selective rodenticide, and in this application (*S*)-(−)-warfarin (Fig. 14b) is about 6 to 7 times more potent than the (*R*) form. There may be a pharmacokinetic effect – the (*S*) form has an approximately twofold longer half-life (15.4 ± 2.8 h) than the (*R*) enantiomer.

For the serotonin (5-hydroxytryptamine) uptake inhibitor fluoxetine (trade name, Prozac), both enantiomers also have an almost equipotent blocking action. In this case, there was an economic reason for pursuing a racemic switch – the patent on the racemate began to expire in 2001. With the single-enantiomer (*R*) form (Fig. 14c), there was a small but significant

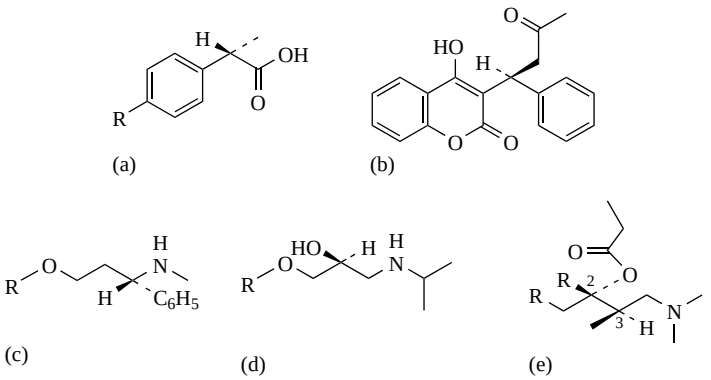

Fig. 14 Chiral drugs. (a) = (*S*)-ibuprofen, R = $CH_3CH(CH_3)CH_2$–; (b) = (*S*)-warfarin; (c) = (*R*)-fluoxetine (trade name, Prozac), R = CF_3-C_6H_4–; (d) = (*S*)-propranolol, R = $C_{10}H_7$–; (e) = (2*R*,3*S*)-propoxyphene (as hydrochloride, trade name Darvon), R = C_6H_5–.

cardiac side effect at high dosage, and development was discontinued.

Case 2. One enantiomer has the desired physiological activity, the other does not. For propranolol, a β-adrenergic blocking agent, the (*S*)-(−) compound (Fig. 14d) is much more effective than the enantiomer; the activity ratio, S : R, is 100 : 1. However, like many drugs, propranolol has other activities. As a local anesthetic, as a regulator of thyroxine metabolism, in membrane stabilization, and for psychiatric use, the enantiomers are equipotent. Propranolol shows complex, enantioselective pharmacokinetic behavior. The (*S*) enantiomer accumulates selectively in the heart with the (*R*) enantiomer remaining largely in blood and being stereoselectively metabolized.

Case 3. The two drug enantiomers have different pharmacological activities one of which may be either desirable or not. Both enantiomers of propoxyphene have useful, different activities. The (2*S*,3*R*)-(+) enantiomer is an analgesic (trade name of hydrochloride, Darvon, Fig. 14e), while the (2*R*,3*S*)-(−) enantiomer is marketed as a naphthalene sulfonate (trade name, Novrad) for use as an antitussive. Darvon read in reverse is Novrad, but the two names are not exact mirror images.

Thalidomide (trade name, Contergan) is often considered as a Case 3 compound but the actual situation is complex. Use of the racemic mixture as a sedative and antinausea agent by pregnant women (although not in the United States) beginning in about 1956 was responsible for the birth of the so-called *thalidomide babies* – worldwide about 10 000 cases of phocomelia – before it was withdrawn from the world market. Early animal work suggested that the teratogenic action was associated with (*S*)-(−) thalidomide (Fig. 15a), which underwent metabolism successively to *N*-phthaloylglutamine (Fig. 15b, R = NH_2) and *N*-phthaloylglutamic acid (Fig. 15b, R = OH). The latter compounds with L(*S*) configurations were teratogenic and the D(*R*) metabolites were harmless. However, this work has apparently never been confirmed independently, and there is a difficulty. Thalidomide undergoes

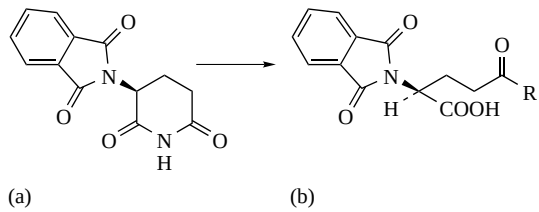

Fig. 15 Thalidomide metabolism. (a) = (*S*)-(−)-thalidomide; (b) = *N*-phthaloyl-L-glutamine, R = NH_2; (b) = *N*-phthaloyl-L-glutamic acid, R = OH.

rather rapid racemization under physiological conditions. In rabbits, for example, thalidomide isolated from plasma 2 h after intravenous injection was completely racemized. It has actually been stated, "that it is practically impossible to demonstrate stereoselectivity in any *in vivo* biological effect of thalidomide" (but see below). With dynamic capillary electrokinetic chromatography, the enantiomerization barrier, $\Delta G^{\#}$, was found to be 102 ± 1 kJ mol^{-1} at pH 8.

Physiological effects of intravenous injection (rather than *per os*) of thalidomide enantiomers into male volunteers have been investigated. Although the enantiomers have low solubility (ca. 0.3 mg mL^{-1}) in 5% glucose (pH 4–5) solutions, preparations were obtained for injections containing 0.2 mg mL^{-1}. [The racemate has lower solubility – 0.07 mg mL^{-1} – and is a likely racemic compound (each crystal contains equal numbers of each enantiomer) rather than a racemic mixture (any crystal contains only one enantiomeric form)]. The solutions formed less than 1% of the opposite enantiomer on 9 days of storage. The results confirmed earlier observations that sedation and sleep were related to blood concentrations of (*R*)- but not (*S*)-thalidomide. Immunological effects (see below) have been attributed to the (*S*) enantiomer. These observations appear to confirm the earlier work assigning teratogenic activity to the (*S*)-(−) enantiomer.

To a surprising extent, racemic thalidomide (only available on prescription) has made a comeback because of immunomodulatory and antiangiogenic properties. It has been used to treat severe recurrent aphthous stomatitis in AIDS patients and to treat systemic manifestations of erythrema nodosum leprosum in patients receiving recombinant interferon-γ treatment for lepromatous leprosy. Moreover, racemic thalidomide inhibits replication of human immunodeficiency virus Type 1 and has a modest ability to inhibit tumor necrosis factor-α, an immune system modulator. Some structural analogs of thalidomide are more effective than thalidomide itself.

Several anesthetics are chiral; this is the case for the inhalation anesthetics halothane, enflurane, and isoflurane, all used as racemates. Some work on isoflurane indicated enantiomer effects on pond snail nerve channels, where the (*S*) enantiomer (Fig. 16a) was twice as active as the (*R*) form in triggering an anesthetic-activated potassium current. In mice and rats, inhalation experiments with isoflurane enantiomers indicated a greater effect of the (*S*) form, but there were also contradictory results. It is likely that any enantiomer-selective effects for volatile anesthetics will be relatively weak.

More definitive results have been observed with intravenous anesthetics. One material, etomidate, is actually used as the single (*R*)-(+) enantiomer (Fig. 16b).

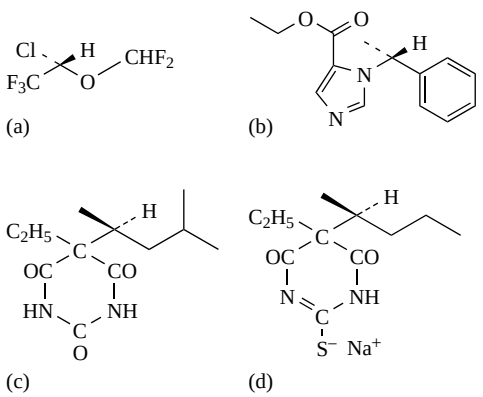

Fig. 16 Chiral anesthetic agents. (a) = (*S*)-isoflurane; (b) = (*R*)-etomidate; (c) = (*S*)-5-ethyl-5(1,3-dimethylbutyryl)-barbital; (d) = (*S*)-thiopental.

It is about five times as active as the (*S*)-(−) form. For the barbiturate compounds used as anesthetics and hypnotics, many agents show Case 3 pharmacodynamic action; a depressant activity is observed in the (*S*) forms, while an excitatory activity is associated with the (*R*) enantiomer. Thus, for 5-ethyl-5(1,3-dimethylbutyryl)-barbital, the (*S*) form (Fig. 16c) is a depressant and the (*R*) form is a convulsant; the racemate is also a convulsant and is not suitable for clinical use. In the clinically useful thiopental sodium, studies in mice and human volunteers indicate greater potency for the (*S*) form (Fig. 16d). However, in animals it yields a lower therapeutic ratio. It has a shorter terminal half-life being cleared and metabolizes more rapidly.

Drug enantiomers show many pharmacokinetic differences such as rates of adsorption and/or transport to receptor sites and in rates of metabolism to various end products. For more details, the specialist literature should be consulted. Further complications such as differences in animals versus humans and differences depending on sex and genetic makeup can arise.

One consequence of the interest in chiral drugs has been a lot of work of a chemical nature – for instance, the June 10, 2002 issue of *Chemical and Engineering News* features "Chiral Chemistry" as a cover story. Syntheses of chiral materials have been devised using asymmetric chemical catalysis, or enzymes, or microbial systems. Moreover, many chromatographic techniques have been developed for chiral separations, both analytically and for preparative purposes. This work is beyond the scope of this review.

4 Chiral Recognition

Enantiospecific physiological effects imply that enzymes or drug receptors achieve "chiral recognition". Much work has focused on the Easson–Stedman three-point attachment model, proposed initially to account for drug enantioselectivity. Given three specific binding sites in a plane, only one enantiomer can bind to the surface (Fig. 17). However, it is now clear that in the absence of specific constraints, the minimal requirement for chiral recognition is interaction between eight atomic centers, involving four different nonplanar centers on each interacting molecule. The interactions may

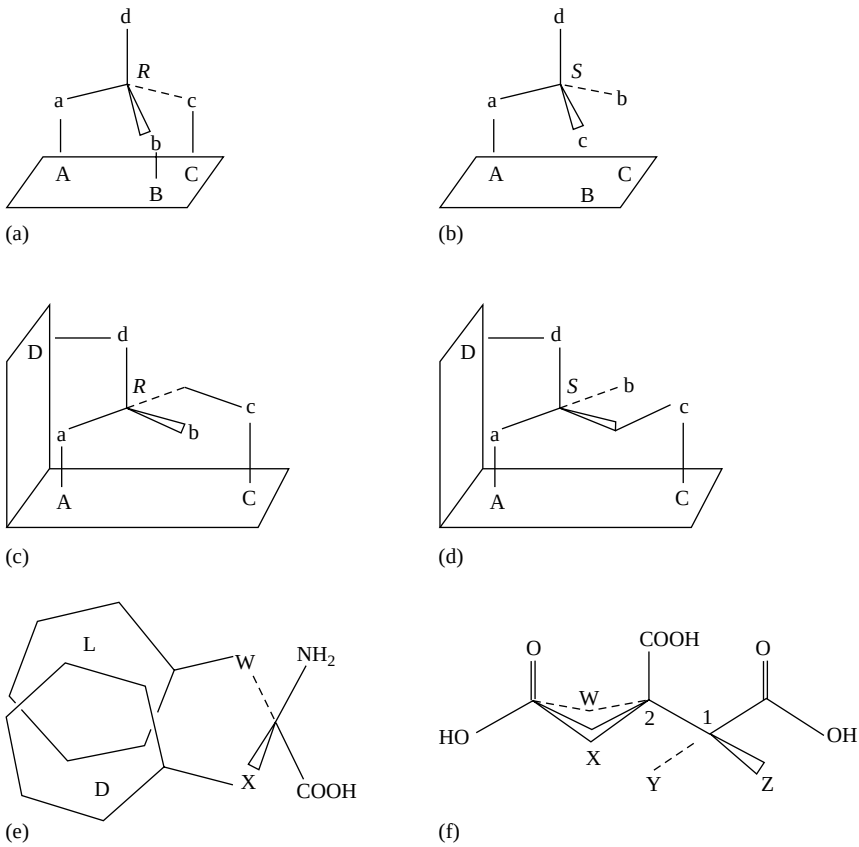

Fig. 17 Models for chiral recognition. (a), (b) show the conventional Easson–Stedman model for Cabcd (sequence assumed to be a > b > c > d). Three binding sites, A, B, and C, are in a plane and it is assumed that the substrate (drug) cannot approach from below (interior). Only the (*R*) enantiomer binds at all three sites, (a). (c), (d) diagram a modified model with three binding sites, A, C, and D, in three dimensions. The substrate (drug) is Cab(-CH_2-c)d. Under these circumstances, both (*R*) and (*S*) enantiomers can bind at three sites, (c), (d). A representation of mirror-image packing for phenylalanine enantiomers is drawn at (e). For D-phenylalanine, W = H, X = CH_2; for L-phenylalanine, W = CH_2, X = H. The aromatic rings (double bonds omitted) are drawn in the plane of the paper with NH_2 above and COOH below this plane. (f), Representation of mirror-image packing for isocitrate enantiomers. For (1*R*,2*S*)-isocitrate, W = CH_2, X = H, Y = OH, and Z = H. For the (1*S*,2*R*) enantiomer, W = H, X = CH_2, Y = H, and Z = OH. For both phenylalanine and isocitrate, the CH_2 group provides a flexibility that makes the mirror-image packing possible.

be either attractive or repulsive. In other words, the minimal requirement is for a four contact-point (four-point attachment) model. In the Easson–Stedman model, it was implicitly assumed that the drug or enzyme substrate could approach the protein surface only from above and not from below, that is, from the "interior" of the protein. Hence, this was a special case of the four-point attachment model. There

is a specific constraint (fourth condition) involved in the interaction.

In enzyme–substrate or drug–receptor interaction, there must be a diastereoisomeric relationship at some point in the pathway or in an intermediate compound/complex. If a chiral enzyme is denoted as (−E) and two substrate enantiomers as (+S) and (−S), the complexes (−E)(+S) and (−E)(−S) will be diastereoisomeric. Since diastereoisomers (unlike enantiomers) have different chemical properties, differentiation between (+S) and (−S) will, in principle, be possible. The occurrence of diastereoisomeric geometries is a necessary but not sufficient condition for chiral recognition.

In the reaction of phenylalanine ammonia lyase, both phenylalanine and substrate analog enantiomers bound to the enzyme; it was proposed that they occupied the same site with "mirror-image packing." This possibility has been confirmed for isocitrate enantiomers in reaction with isocitrate dehydrogenase. These appear to be special situations deriving from the specific structures of the enzyme substrates (see Fig. 17). If three contact points are involved in mirror-image packing, clearly a fourth condition or site must be involved for chiral recognition as noted previously.

5
Enzyme Specificity with Chiral Substrates

Enzymes are polymers of many chiral amino acids. Thus, the relatively small enzyme, ribonuclease ($M_r = 13\,700$), has 109 amino acid residues with a single chiral center, 12 with 2 chiral centers and 3 residues of achiral glycine. Put in simple terms, this and other enzymes are highly chiral catalysts. Many texts contain statements such as the following actual example: "It has long been known that natural enzymes act only on one chiral form (enantiomer) of a chiral substrate." While in general enzymes are often highly specific for one enantiomer, there are also many documented cases where specificity is not observed or is not absolute. An extensive listing of various situations and reaction types was given previously and only a few examples will be noted here.

One highly specific case is D-amino acid oxidase (EC 1.4.3.3). Although L-amino acids are not substrates, several inhibit the oxidation of D-alanine (but not L-alanine). Presumably, these inhibitors bind at the normal active site. The amino acyl-tRNA ligases (synthetases, EC 6.1.1.X) apparently show absolute specificity for L-amino acids during ribosomal protein synthesis. These enzymes are of interest since they ligate two chiral substrates. When examined *in vitro*, there is some relaxation of the specificity as noted earlier for tyrosine-$tRNA^{tyr}$ ligase (EC 6.1.1.1). For arginine and proline ligases (EC 6.1.1.19 and EC 6.1.1.15 respectively), the D-enantiomers are inhibitory.

Examples of enzymes with lack of enantiospecificity are alcohol dehydrogenase (EC 1.1.1.1), where the liver enzyme reacts with both enantiomeric forms of butan-2-ol and octan-2-ol (although there are rate differences), and with nicotine dehydrogenase (EC 1.5.99.4), converting both nicotine enantiomers to the (*R*)- and (*S*)-6-hydroxynicotines. Further metabolism of the 6-hydroxynicotine enantiomers requires separate (*R*)- and (*S*)-6-hydroxynicotine oxidases (EC 1.1.3.6 and EC 1.5.3.5 respectively).

A remarkable achievement was the 1992 chemical synthesis of both enantiomeric forms of the dimeric enzyme HIV-1 protease; aminobutyrate was an isosteric replacement for cysteine at monomer

positions 67 and 95. The all-D and all-L proteins were in every respect mirror images, apart from enzymatic activity. They were enantiospecific in cleavage of a hexapeptide analog of a normal cleavage site. The all-D analog 2-aminobenzoyl-D-thr-D-ile-D-nle-D-*p*-nitrophe-D-gln-D-arg($CONH_2$) was cleaved only by D-enzyme; conversely, the L-enzyme cleaved only the all-L hexapeptide (nle = norleucine).

In several cases, there are separate enzymes for enantiomeric substrates; examples include D- and L-amino acid oxidases, D- and L-lactate dehydrogenases, D- and L-arabinose 1-dehydrogenases, and the two 6-hydroxynicotine oxidases previously noted. Such enzymes acting in concert would provide a racemase possibility, for example, for lactate dehydrogenases:

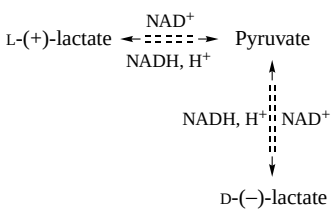

In other cases, a single enzyme functions as a racemase and is, therefore, an example of an enzyme utilizing both enantiomers. Racemase enzymes are especially well known for amino acids (e.g. serine racemase noted earlier) and hydroxyacids (e.g. lactate, mandelate). An important enzyme involved in propionyl-CoA metabolism interconverts the (*R*) and (*S*) forms of methylmalonyl-CoA (Fig. 18). Although often termed *methylmalonyl-CoA racemase*, this enzyme is technically not a racemase since the five chiral centers of the CoA component are (presumably) unchanged. The more appropriate name is methylmalonyl-CoA epimerase (EC 5.1.99.1).

6 Prostereoisomerism

Many enzyme substrates have the general achiral structure, Caabc; there are two chemically like atoms or groups of atoms, a, and two different atoms or groups of atoms, b and c. Although these structures are achiral, the two a groups are in different stereochemical environments. For convenience, the a groups will be distinguished as a′ and a″. When a′ is observed as if to determine chirality by the Cahn–Ingold–Prelog method, the sequence a″ → b → c has a right-handed sense (Fig. 19a). Similarly, with a″ in the remote position, the sequence a′ → b → c has a left-handed sense (Fig. 19b). If each a group is imagined to be replaced by a further group, d, the two enantiomers of a chiral compound, Cabcd, are obtained (see below).

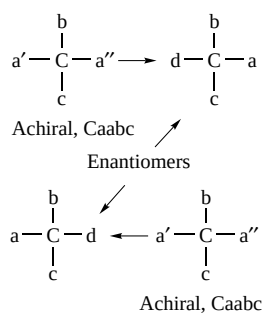

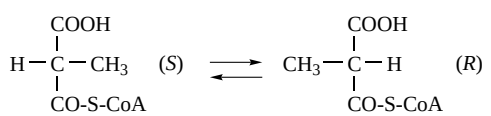

Fig. 18 Methylmalonyl-CoA epimerase. The chiral centers in the CoA unit do not undergo inversion during the epimerization.

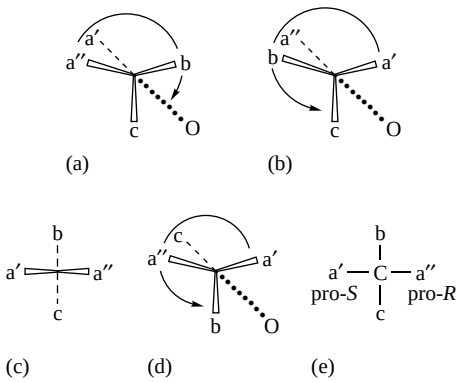

Fig. 19 Prochirality in Ca′a″bc. The different stereoisomeric environments for a′ and a″ are shown as (a) and (b). (c) = conventional three-dimensional representation of cCa′a″bc. (d) = prochirality determination if the sequence is a′ > a″ > b > c; O = eye of observer. The sequence, a′ → a″ → b, is left-handed, hence the promoted group, a′, is pro-*S*. (e) = final arrangement.

For that reason, Caabc is often described as a prochiral structure; it is preferable, however, to refer to the situation as prostereoisomerism.

A chiral reagent, usually an enzyme, is able to differentiate between a′ and a″ in Ca′a″bc. This possibility was rationalized by Ogston in terms of a three-point attachment model exactly as for chiral recognition (at the time of his suggestion, Ogston was unaware of the Easson–Stedman model). It is now apparent that the three-point attachment model is a special case and that in the absence of specific constraints (e.g. substrate cannot approach from the interior of the protein), eight center interactions are required – a four-location model.

Stereochemical descriptors for prostereoisomeric ligands derive from an extension of the Cahn–Ingold–Prelog system. One a group, say a′, is arbitrarily assigned a higher priority in the sequence rule than the other, a″. If the "normal" sequence for Caabc is a > b > c, the "modified" sequence becomes a′ > a″ > b > c. The model is then viewed in the usual way from the side remote from the lowest priority group, c. If the sequence, a′ → a″ → b is right-handed, the "promoted" group a′ is assigned the pro-*R* descriptor and a″ is assigned the pro-*S* descriptor (Fig. 19c–e).

The descriptors can also be obtained from projection models, as described previously for chiral centers. If the two a groups are hydrogen atoms, the descriptors are written as H_R and H_S. A specific case concerns the two $CH_2{-}COOH$ groups of citric acid; the modified sequence is OH > COOH > $C'H_2{-}C'OOH$ > $C''H_2{-}C''OOH$ and the descriptors are derived as shown in Fig. 20. Moreover, in citric acid, each CH_2 group contains H_R and H_S possibilities. In the important aconitase reaction, it is H_R of the pro-*R* methylene group that is removed along with OH in the formation of *cis*-aconitate (Fig. 20).

Replacement of H_R and H_S, for example, in $CH_3{-}CH_2{-}OH$, by ^2H leads to the two enantiomers of [1-^{2}H]-ethanol; optical activity has been demonstrated for

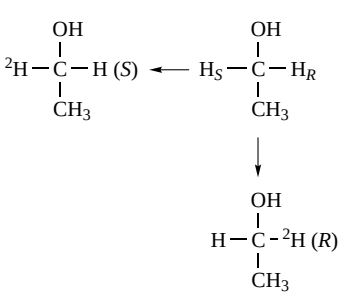

the enantiomers; (*S*) is (−) and (*R*) is (+).

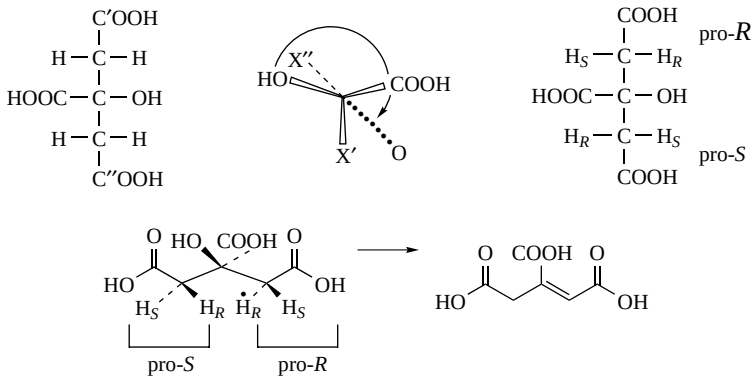

Fig. 20 Prochirality considerations for citric acid. The three structures at the top show the assignments for the CH_2COOH groups and the hydrogens of the CH_2 groups; X′ = $CH_2C'OOH$, X″ = $CH_2C''OOH$; O = eye of observer. The "promoted" priority sequence is HO > COOH > $CH_2C'OOH$ > CH_2 $C''OOH$. Derivations for H_R and H_S are not shown. The bottom line shows the action of aconitase (citrate → H_2O + *cis*-aconitate) The H_R proton (identified as •) is removed from the pro-*R* methylene group.

Fig. 21 Alcohol dehydrogenase stereochemistry. (a) = the reducing cofactor, NADPH; R^1 = rest of structure, R^2 = $CONH_2$. The H_A (=H_R) hydrogen is transferred to form (*S*)-[2-^{2}H]-ethanol.

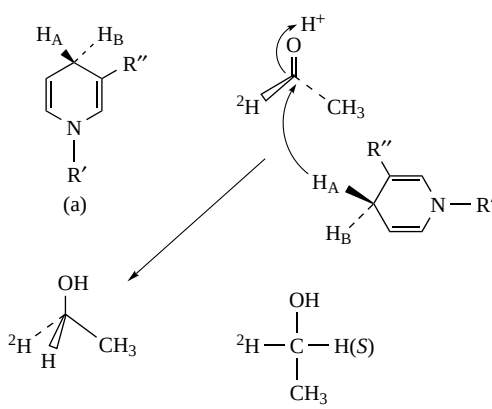

The "faces" or "sides" of trigonal carbon atoms can also be distinguished. For instance, the enzymatic reduction of [1-^{2}H]-acetaldehyde is stereospecific with respect to the face of the carbonyl group (Fig. 21). If the enzyme used is alcohol dehydrogenase, there is a further factor – the two hydrogens at C-4 of the reducing coenzyme, NADH, are prostereoisomeric. They are often described as H_A and H_B, corresponding respectively to H_R and H_S. For citrate synthase, catalyzing the formation of citrate from oxaloacetate and acetyl-CoA, the "incoming" CH_2COOH unit, derived from acetyl-CoA, occupies the pro-*S* position in citrate – the carbonyl of oxaloacetate was attacked from only one side (Fig. 22). Stereochemical descriptors to define the sides and faces of trigonal carbon atoms have been devised; the specialist literature should be consulted.

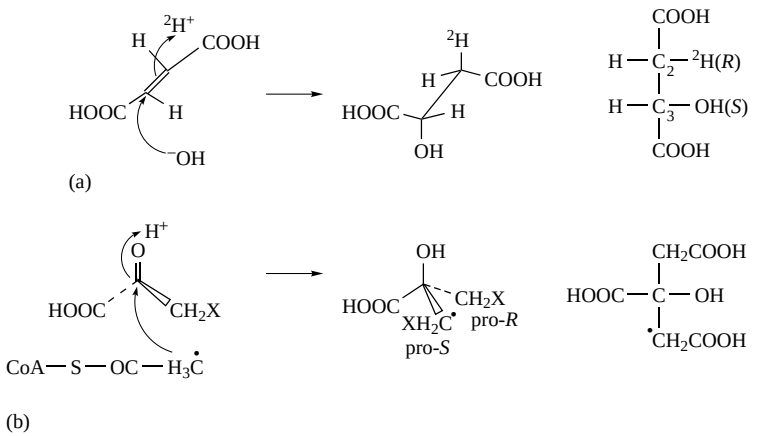

Fig. 22 The fumarase and citrate synthase reactions. (a) The top line shows fumarase addition of ^2H_2O to fumarate leading to production of (2*R*,3*S*)-[2-^{2}H]malate. (b) The lower line shows the addition of acetyl-CoA to oxaloacetate; X = COOH. Use of ^{13}C-labeled acetyl-CoA (^{13}C identified as •) yields ^{13}C in the pro-*S* position of citrate.

Similar considerations apply to addition reactions at double bonds, C=C. Thus, in the action of fumarase on fumarate in ^2H_2O, the OH group attacks from below the double-bond plane and H^+ enters from above the plane, thus leading to formation of (2*S*,3*R*)-[3-^{2}H]-malic acid (Fig. 22).

Since ^{2}H is readily available at 100% concentration, $^2H-CH_2-COOH$ can be prepared. In this compound, the two ^{1}H atoms are in a prostereoisomeric relationship. If a further substitution of one ^{1}H by 100% ^{3}H were practicable, a chiral compound would be formed, [^{3}H,^{2}H,^{1}H]-acetate, and theoretically optical activity would be observable. Because of the very high level of radioactivity associated with 100% ^{3}H, it is unlikely that this statement will be put to the proof. However, normal tracer levels of ^{3}H can be used to explore the chirality of methyl groups containing all three of the hydrogen isotopes. The two enantiomeric acetates (Fig. 23a, b) can be synthesized either chemically or by combined enzymatic and chemical methods; their chirality is demonstrated as follows. The acetate samples are converted to the CoA derivatives and the latter are used as substrates in the malate synthase reaction (EC 4.1.3.2; glyoxylate + acetyl-CoA → L-(*S*)-malate). In this reaction, a marked isotope effect dictates removal of ^{1}H in preference to either ^{2}H or ^{3}H. In turn, the malate samples are treated with fumarase (EC 4.2.1.2, fumarate hydratase, L-(*S*)-malate → H_2O + fumarate) and the fumarate is examined for ^{3}H radioactivity (Fig. 23).

The overall result is that from (*R*)-[^{1}H,^{2}H,^{3}H]-) acetate, ^{3}H is predominantly retained in the isolated fumarate (ca. 80% retention) and from the (*S*) enantiomer, ^{3}H is predominantly lost (ca. 20% retention). While loss/retention of ^{3}H is not 100%, such completeness is not expected; the actual value depends on the nature of the isotope effect in the malate synthase reaction. This malate synthase/fumarase system demonstrates the chirality of the methyl group and provides a convenient assay method for the chirality of acetate

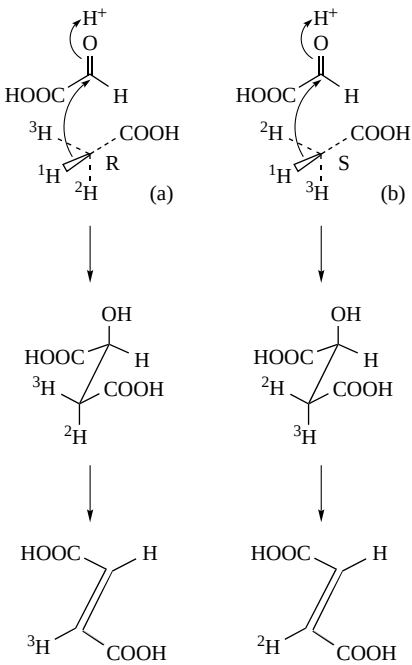

Fig. 23 Analysis of chiral acetates. The first reaction is that of malate synthase, glyoxylate + acetyl-CoA → malate (for simplicity, the samples are shown as acetate rather than as acetyl-CoA). An isotope effect dictates removal of ^{1}H as shown. The second reaction is that catalyzed by fumarase – removal of the elements of water with defined *trans* stereochemistry. The possible fumarate products are shown on the last line.

samples. There have been many investigations of chiral methyl groups in biochemical processes.

In a similar way, chiral phosphates, containing ^{16}O, ^{17}O, and ^{18}O, and chiral phosphorothioates, containing S, ^{16}O, and ^{18}O, have been utilized in several enzymatic reactions. Thus, in the action of adenylate kinase, a phosphorothioate analog of ATP was used to show that the reaction proceeds with overall inversion of configuration (Fig. 24). Similarly, with the use of all three of the oxygen isotopes, glucose phosphorylation by hexokinase was also shown to proceed with inversion of configuration.

7 Chirality in Biological Structures and Behavior

7.1 Structures

Chirality occurs in many manufactured objects (e.g. screws, propellers, shoes) and is very common in the world of nature. When viewed externally, the bodies of humans and most animals show an overall bilateral symmetry. A mirror plane can be imagined at right angles to the body, thus bisecting the nose and legs. Usually, this symmetry is imperfect with slight differences say between the left and the right hands (the right hands of right-handed people are usually larger than the left hands). Irregularities are often seen in faces. Faces artificially constructed from two left or two right sides are markedly

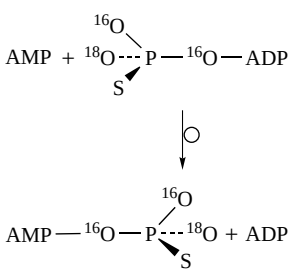

Fig. 24 The mechanism of action of adenylate kinase, EC 2.7.4.3. The catalyzed reaction is ATP + AMP = ADP + ADP. ATP is shown as a phosphorothioate analog containing ^{16}O, ^{18}O, and S (double bonds are omitted). The "looped arrow" indicates that the reaction involves inversion of configuration at the phosphorous atom.

different from the original. Faces on some statues are actually mirror symmetric and have been described as having an "eerie beauty."

The internal arrangement of body organs, however, lacks symmetry. As the embryo develops, the heart, spleen, and stomach normally assume the left side of the chest cavity, while the major lobe of the liver and the gall bladder locate on the right side of the abdomen. The left lung usually has two lobes, the right, three. In the condition of *situs inversus*, there is a transposition of the viscera, the liver being on the left, the heart on the right. This condition occurs with a frequency of 1 per 10 000 births. It seems surprising that individuals with *situs inversus* generally are right-handed. Of 160 cases of *situs inversus* in which handedness was known, only 11 were left-handed.

A rarer situation is that of individuals who are much more symmetrical than normal. In "right isomerism," both lungs have three lobes, the heart has two right atrial appendages, the liver is central, and the spleen is absent. For a male patient, the testicles are of the same length (the right testicle is usually longer). In "left isomerism," the lungs are two-lobed, the heart has two left atrial appendages, the liver is central, and the individual possesses multiple spleens. Individuals with either isomerism have serious problems with heart function.

Mutant mice showing *situs inversus* are known. The *iv* mutation (*inversus viscerum*) gives about 50% of cases of *situs inversus*. The different *inv* mutation (abbreviation of "inversion of embryonic turning") yields 100% inversion; these mice die early in life.

Of the primary axes of the embryo (anterioposterior, dorsoventral, and left/right asymmetry), much attention has been focused on the left/right asymmetry axis in recent years. Answers are being sought, primarily at the genetic level, to questions about what determines the asymmetry and how does it develop. Several genes or gene clusters have been identified, but there still is relatively little information about the gene products. The subject is too complex for discussion here.

There are many exceptions to overall bilateral symmetry; only a few examples can be given. Lobsters have one small limb for clutching and a large one for tearing. In 50% of the population, the small limb is on the left and in the other 50% it is on the right. Similarly, in male fiddler crabs the large claw or pincer is to the left in 50% of the population and to the right in the other 50%. In male narwhals, a large ivory tusk (up to 10 feet long) pierces the upper lip on the left side of the body and shows a striated spiral pattern from right to left. In birds, the New Zealand wry-billed plover has a beak that is always bent to the right-hand side of the bird. Similarly, a breed of European finches, termed *crossbills*, has an upper mandible pointing to the bird's right and the lower to the left. This crossed arrangement may facilitate removal of a seed from a pinecone directly with the mouth. North American crossbills have the opposite asymmetry. Clearly, the arrangement provides some advantage, whatever direction it takes.

Chirality is well illustrated in mollusk shells. In most cases, the helical direction is right-handed, but some whelks (e.g. the Florida lightning whelk) are left-handed. Rare examples of "sinistral," left-handed shells, for normally right-handed types, are prized by collectors. In India, the left-handed form of the chank shell (*Turbinella pyrum*) is considered sacred by Hindus. Very ancient statues portray

the god, Vishnu, as holding a left-handed chank shell

In tobacco plants, leaves alternate along the stem in either a left-handed or a right-handed spiral. In large populations, the left : right ratio is 1 : 1. Vining plants coil upward in a left- or right-handed spiral characteristic of the species; honeysuckle is right-handed and bindweed is left-handed.

Even microorganisms can show chiral growth patterns. In *B. subtilis*, one structural possibility consists of filaments of many individual cells with a double-stranded helical arrangement. The helix chirality and degree of twist are influenced by growth conditions. There is a range from "maximum tight left-handedness" to "maximum tight right-handedness." Some mutants have only one helical chiral sense. In a truly bizarre case, one fungal species of the genus *Laboulbenia* parasitizes beetles. It occurs only on a single joint of the left hind leg; no other part, including the right leg, is involved.

7.2 Behavioral Asymmetries

That the two parts of the brain have different functions became apparent from studies of brain-damaged individuals. Early work was carried out in France by Marc Dax, with more extensive studies by Pierre Paul Broca (1824–1880), a surgeon. Broca determined that language facility and action are usually associated with the left hemisphere of the brain – in his words (1865), "On parle avec l'hémisphère gauche." In the condition of Broca's aphasia (or motor aphasia), the power of expression by writing, speaking, or signs is lost. In 1995, another form of mental performance, perfect pitch, was attributed to the left brain. Still more recently, left-hemisphere cerebral specialization was reported for the babbling of babies, either those acquiring English or those acquiring French. The right hemisphere is involved with processes of perception. Hence, right-brain damage involves not only sight but also other senses such as touch, sound, and taste. Much has been written on the left-brain/right-brain situation and specialized texts should be consulted for further details. The situation is complex and there are exceptions to the generalizations just noted.

Perhaps even more complex is the subject of handedness; again, much has been, and continues to be, written on this subject. In the Western nations, a little over 10% of the population is left-handed (defined as those using the left hand for at least half of their tasks). Left-handedness is more common among women than men, and individuals can be inconsistent in their hand usage. Similar preferences exist for the use of feet, eyes, and ears. There has been much speculation concerning the origin of handedness. A genetic component is probably involved and left-handedness in humans (but not in animals) runs in families. Studies of twins have posed some problems for genetic hypotheses. For instance, with identical twins, about one in five pairs shows nonidentical handedness, one twin being right-handed, the other, left-handed.

There are two genetic models, one proposed by Annett, the other by McManus. In brief, Annett postulates a gene, *RS+*, that displaces the chance distribution to the right (the right-shift model), whereas McManus postulates two alleles, *D (Dextral)* and *C (Chance)*, at a single autosomal locus. The *DD* genotype produces exclusively right-handed individuals, while the

CC genotype results in "fluctuating asymmetry," that is, a 1 : 1 distribution. The *DC* genotype yields 25% left-handedness. The original hypotheses have been modified and subject to criticism and the topic cannot be summarized simply. It is noteworthy that in a 1991 Symposium, "Biological Asymmetry and Handedness," a paper by McManus (The Inheritance of Left-handedness) was extensively discussed by the participants. The recorded comments comprise about 7000 words. The genetics of left/right asymmetry in bodily organs are under study, but apparently none of the identified genes have been implicated in behavior.

8
Origin of Chirality on Earth

Many attempts to explain the emergence of life on planet Earth and the establishment of homochirality for the major macromolecules, nucleic acids, proteins, and carbohydrates have led to little agreement. One difficulty is that experimental chemical approaches to prebiotic conditions produce racemic compounds if a chiral center is involved. Two general possibilities are (1) that the first significant molecules for evolution of life were already formed with one chiral sense in place or (2) that initially produced racemic mixtures were subjected to chiral enrichment by a selection process.

Actual mechanisms under consideration are legion-interactions with polarized β-radiation from nuclear decay, interactions with circularly polarized light, selective adsorption, or polymerization on crystals or clays, stochastic sorting of conglomerate crystals, enantioselective magnetochiral photochemistry, stirring, and chiroselective amplification. A recent development is the suggestion that life may have arisen in high-temperature environments with prebiotic synthesis in an environment of iron sulfide and hot magmatic exhalations.

Another general suggestion is that extraterrestrial organic compounds may have been transported to Earth and used for evolutionary processes. Certain amino acids are definitely present in meteorites, most famously, the Murchison. There has been some evidence concerning possible enantiomeric excess in such amino acids, particularly those that do not occur in proteins. There are, however, conflicting reports, and the possible contamination of meteorite material by terrestrial material is real. To overcome such difficulties, instrumental approaches to detecting chirality in, for example, amino acids, are being prepared for space missions to Mars and comets.

Another major area of discussion relates to the possibility that earthly homochirality reflects the fundamental structure of matter itself, in particular, the nonconservation of parity with the weak force as carried by W^+ and Z^0 particles. The weak force is the only one that is chiral. The universe has an intrinsic "left-handedness" – it is made of matter and not antimatter, and the parity-violating differences between enantiomers might determine biomolecular chirality. The handedness of particles implies that L and D molecules are actually diastereoisomers; the true enantiomer of, say, L-alanine would be D-alanine made from antimatter. Some authors believe that the possible selectivity in such processes is too small and yet others view the initial choice as a random event.

For life as we know it, all D-proteins and nucleic acids with L-ribose and L-deoxyribose would (presumably) have been equally acceptable – this point of view is bolstered by the work previously described with all-D HIV protease. Perhaps

somewhere in the Universe there is, indeed, a ''looking-glass'' world in which Alice would find that looking-glass milk containing lactose, with L-glucose and L-galactose components, was good to drink.

See also Bioorganic Chemistry; Circular Dichroism in Protein Analysis.

Bibliography

Books and Reviews

Aboul-Enein, H.Y., Wainer, I.W. (Eds.) (1997) *The Impact of Stereochemistry on Drug Development and Use*, Wiley, New York.

Annett, M. (2002) *Handedness and Brain Asymmetry. The Right Shift Theory*, Psychology Press, Hove, UK.

Bentley, R. (1969, 1970) *Molecular Asymmetry in Biology*, Vols. I and II, Academic Press, New York.

Brown, C. (Ed.) (1991) *Chirality in Drug Design and Synthesis*, Academic Press, San Diego.

Buckingham, J., Hill, R.A. (1987) *Atlas of Stereochemistry*, Vols. I and II and Supplement, 2nd edition, Chapman & Hall, London.

Casy, A.F., Dewar, G.H. (1993) *The Steric Factor in Medicinal Chemistry. Dissymmetric Probes of Pharmacological Receptors*, Plenum Press, New York.

Close, F. (2000) *Lucifer's Legacy: The Meaning of Asymmetry*, Oxford University Press, Oxford.

Eliel, E.L., Wilen, S.H., Doyle, M.P. (2001) *Basic Organic Stereochemistry*, Wiley-Interscience, New York.

Lough, W.J., Wainer, I.W. (Eds.) (2002) *Chirality in Natural and Applied Science*, Blackwell, Oxford.

McManus, C. (2002) *Right Hand, Left Hand. The Origins of Asymmetry in Brains, Bodies, Atoms, and Cultures*, Weidenfeld & Nicolson, London.

Primary Literature

Armstrong, D.W., Liu, Y.-S., He, L., Ekborg-Ott, K.F., Barnes, C.L., Hammer, C.F. (2002) Potent enantioselective auxin: indole-3-succinic acid, *J. Agric. Food Chem.* **50**, 473–476.

Avetisov, V., Goldanskii, V. (1996) Mirror symmetry breaking at the molecular level, *Proc. Natl. Acad. Sci. USA* **93**, 11435–11442.

Bailey, J., Chrysostomou, A., Hough, J.H., Gledhill, T.M., McCall, A., Clark, S., Ménard, F., Tamura, M. (1998) Circular polarization in star-formation regions: implications for biomolecular homochirality, *Science* **281**, 672–674.

Bentley, R. (1995) From optical activity in quartz to chiral drugs: molecular handedness in biology and medicine, *Perspect. Biol. Med.* **38**, 188–229.

Bentley, R. (1997) Are Fischer projection formulas really necessary? *Biochem. Educ.* **25**, 216–220.

Bentley, R. (1998) The transformation of linear to cyclic monosaccharide structures, *Biochem. Educ.* **26**, 205–210.

Bentley, R. (2003) Diastereoisomerism, contact points, and chiral selectivity: a four-site saga, *Arch. Biochem. Biophys.* **414**, 1–12.

Bilke, S., Mosandi, A. (2002) Authenticity assessment of lavender oils using GC-P-IRMS: $^{2}H/^{1}H$ isotope ratios of linalool and linalyl acetate, *Eur. Food Res. Technol.* **214**, 532–535.

Bock, G.R., Marsh, J. (Eds.) (1991) *Biological Asymmetry and Handedness*, Ciba Foundation Symposium 162, Wiley, Chichester.

Burke, D., Henderson, D.J. (2002) Chirality: a blueprint for the future, *Br. J. Anaesth.* **88**, 563–576.

Capdevila, J., Vogan, K.J., Tabin, C.J., Belmont, J.C.I. (2000) Mechanisms of left-right determination in vertebrates, *Cell* **101**, 9–21.

Cody, G.D., Boctor, N.Z., Filley, T.R., Hazen, R.M., Scott, J.H., Sharma, A., Yoder, H.S. (2000) Primordial carbonylated iron-sulfur compounds and the synthesis of pyruvate, *Science* **289**, 1337–1340.

D'Aniello, G., Tolino, A., D'Aniello, A., Errico, F., Fisher, G.H., Di Fiore, M.M. (2000) The role of D-aspartic acid and *N*-methyl-D-aspartic acid in the regulation of prolactin release, *Endocrinology* **141**, 3862–3870.

Eriksson, T., Björkman, S., Roth, B., Höglund, P. (2000) Intravenous formulations of the enantiomers of thalidomide: pharmacokinetics and initial pharmacodynamic characterization in man, *J. Pharm. Pharmacol.* **52**, 807–817.

Floss, H.G., Tsai, M.-D. (1979) Chiral methyl groups, *Adv. Enzymol.* **50**, 243–302.

Frey, P.A. (1989) Chiral phosphorothioates: stereochemical analysis of enzymatic substitution at phosphorous, *Adv. Enzymol.* **62**, 119–201.

Garcia-Castro, M.I., Vielmetter, E., Bronner-Fraser, M. (2000) *N*-Cadherin, a cell adhesion molecule involved in establishment of embryonic left-right asymmetry, *Science* **288**, 1047–1051.

Hancock, R.D., Galpin, J.R., Viola, A. (2000) Biosynthesis of L-ascorbic acid (vitamin C) by *Saccharomyces cerevisiae*, *FEMS Microbiol. Lett.* **186**, 245–250.

Hanein, D., Geiger, B., Addadi, L. (1994) Differential adhesion of cells to enantiomorphous crystal faces, *Science* **263**, 1413–1416.

Hanson, K.R. (1981) Phenylalanine ammonia-lyase: mirror-image packing of D- and L-phenylalanine and D- and L-transition state analogs into the active site, *Arch. Biochem. Biophys.* **211**, 575–588.

Hazen, R.M., Filley, T.R., Goodfriend, G.A. (2001) Selective adsorption of L- and D-amino acids on calcite: implications for biochemical homochirality, *Proc. Natl. Acad. Sci. USA* **98**, 5487–5490.

Heuberger, E., Hongratanaworakit, T., Böhm, C., Weber, R., Buchbauer, G. (2001) Effects of chiral fragrances on human autonomic nervous system parameters and self-evaluation, *Chem. Senses* **26**, 281–292.

Holowka, S., Pettitto, L.A. (2002) Left hemisphere cerebral specialization for babies while babbling, *Science* **297**, 1515.

Hough, J.H., Bailey, J.A., Chrysostomou, A., Gledhill, T.M., Lucas, P.W., Tamura, M., Clark, S., Yates, J., Ménard, F. (2001) Circular polarization in star-forming regions: possible implications for homochirality, *Adv. Space Res.* **27**, 313–322.

Imai, K., Fukushima, T., Santa, T., Homma, H., Huang, Y., Sakai, K., Kato, M. (1997) Distribution of free D-amino acids in tissues and body fluids of vertebrates, *Enantiomer* **2**, 143–145.

Kraft, P., Fráter, G. (2001) Enantioselectivity of the musk odor sensation, *Chirality* **13**, 388–394.

Lahav, N., Nir, S., Elitzur, A.C. (2001) The emergence of life on earth, *Biophys. Mol. Biol.* **75**, 75–120.

Laska, M., Galizia, C.G. (2001) Enantioselectivity of odor perception in honeybees (*Apis mellifera carnica*), *Behav. Neurosci.* **115**, 632–639.

Laska, M., Liessen, A., Teubner, P. (1999) Enantioselectivity of odor perception in squirrel monkeys and humans, *Am. J. Physiol.* **277**, (*Regulatory Integrative Comp. Physiol.*) R1098–R1103.

Laska, M., Teubner, P. (1999) Olfactory discrimination ability of human subjects for ten pairs of enantiomers, *Chem. Senses* **24**, 161–170.

Leal, W.S. (1996) Chemical communication in scarab beetles: reciprocal behavioral agonist-antagonist activities of chiral pheromones, *Proc. Natl. Acad. Sci. USA* **93**, 12112–12115.

Lee, S.B., Mitchell, D.T., Trofin, L., Nevanen, T.K., Söderlund, H., Martin, C.R. (2002) Antibody-based bio-nanotube membranes for enantiomeric drug separations, *Science* **296**, 2198–2200.

Levin, M., Mercola, M. (1998) The compulsion of chirality: toward an understanding of left-right asymmetry, *Genes Dev.* **12**, 763–769.

MacDermott, A. (1995) The weak force and the origin of life and self- organization, in: Chela-Flores, J., Chadha, M., Negrón-Mendoza, A., Oshima, T. (Eds.) *Chemical Evolution: Self-Organization of the Molecules of Life*, A. Deepak Publishing, Hampton, VA, pp. 237–249.

Makonkawkeyoon, S., Limson-Pobre, R.N.R., Moreira, A.L., Schauf, V., Kaplan, G. (1993) Thalidomide inhibits the replication of human immunodeficiency virus type 1, *Proc. Natl. Acad. Sci. USA* **90**, 5974–5978.

McCarthy, A., Brown, N.A. (1998) Specification of left-right asymmetry in mammals: embryo culture studies of stage of determination and relationships with morphogenesis and growth, *Reprod. Toxicol.* **12**, 177–184.

Mercola, M., Levin, M. (2001) Left-right asymmetry determination in vertebrates, *Annu. Rev. Cell Dev. Biol.* **17**, 779–805.

Mesecar, A.D., Koshland, D.E. (2000) Sites of binding and orientation in a four-location model for protein stereospecificity, *IUBMB Life* **49**, 457–466.

Mizanur, R.M., Takata, G., Izumori, K. (2001) Cloning and characterization of a novel gene encoding L-ribose isomerase from *Acinetobacter* sp. strain DL-288 in *Escherichia coli*, *Biochim. Biophys. Acta* **1521**, 141–145.

Morgan, D., Goodship, J., Essner, J.J., Vogan, K.J., Turnpenny, L., Yost, H.J., Tabin, C.J., Strachan, T. (2002) The left-right determinant inversin has highly conserved ankyrin

repeat and IQ domains and interacts with calmodulin, *Hum. Genet.* **110**, 377–384.

Moss, G.P. (1996) Basic terminology of stereochemistry (IUPAC Recommendations 1996), *Pure Appl. Chem.* **68**, 2193–2222.

Nelson, G., Chandrashekar, J., Hoon, M.A., Zhao, G., Ryba, N.J.P., Zuker, C.S. (2002) An amino-acid taste receptor, *Nature* **416**, 199–202.

Nicolau, D.P., West, T.E. (1990) Thalidomide treatment of severe recurrent aphthous stomatitis in patients with AIDS, *DICP Ann. Pharmacother.* **24**, 1054–1056.

Nishino, H., Kosaka, A., Hembury, G.A., Shitomi, H., Onuki, H., Inoue, Y. (2001) Mechanism of pH-dependent photolysis of aliphatic amino acids and enantiomeric enrichment of racemic leucine by circularly polarized light, *Org. Lett.* **3**, 921–924.

Oh, S.P., Li, E. (2002) Gene-dosage-sensitive genetic interaction between inversus viscerum (*iv*), *nodal*, and activin type IIB receptor (*ActRIIB*) genes in asymmetrical patterning of the visceral organs along the left-right axis, *Dev. Dyn.* **224**, 279–290.

Opitz, J.M., Utkus, A. (2001) Comments on biological asymmetry, *Am. J. Med. Genet.* **101**, 359–369.

Pizzarello, S., Cronin, J.R. (2000) Non-racemic amino acids in the Murray and Murchison meteorites, *Geochim. Cosmochim. Acta* **64**, 329–338.

Ribó, J.M., Crusats, J., Sagués, F., Claret, J., Rubires, R. (2001) Chiral sign induction by vortices during the formation of mesophases in stirred solutions, *Science* **292**, 2063–2066.

Rikken, G.L.J.A., Raupach, E. (2000) Enantioselective magnetochiral photochemistry, *Nature* **405**, 932–935.

Rodier, C., Laurent, C., Szopa C., Sternberg, R., Raulin, F. (2002) Chirality and the origin of life: in situ enantiomeric separation for future space missions, *Chirality* **14**, 527–532.

Rubin, B.D., Katz, L.C. (2001) Spatial coding of enantiomers in the rat olfactory bulb, *Nat. Neurosci.* **4**, 355–356.

Saghatelian, A., Yokobayashi, Y., Soltani, K., Ghadri, M.R. (2001) A chiroselective peptide replicator, *Nature* **409**, 797–801.

Sampaio, E.P., Moreira, A.L., Sarno, E.N., Malta, A.M., Kaplan, G. (1992) Prolonged treatment with recombinant interferon γ induces erythrema nodosum leprosum in lepromatous leprosy patients, *J. Exp. Med.* **175**, 1729–1733.

Schmarr, H.-G., Eisenrich, W., Engel, K.-H. (2001) Synthesis and analysis of thio-, thiono-, and dithio-derivatives of whiskey lactone, *J. Agric. Food Chem.* **49**, 5923–5928.

Schon, P., Tsuchiya, K., Lenoir, D., Mochizuki, T., Guichard, C., Takai, S., Maiti, A.K., Nihei, H., Weil, J., Yokayama, T., Bouvagnet, P. (2002) Identification, genomic organization, chromosomal mapping and mutation analysis of the human INV gene implicated in left-right axis development and biliary atresia, *Hum. Genet.* **110**, 157–165.

Seife, C. (2002) Shadowy 'Weak Force' steps into the light, *Science* **297**, 184–187.

Sheldon, R.A. (1993) *Chirotechnology: Industrial Synthesis of Optically Active Compounds*, Marcel Dekker, New York.

Strand, A., Kvernberg, K., Karlsen, A.M., Liaaen-Jensen, S. (2000) Geometrical *E*/*Z* isomers of (6*R*)- and (6*S*)-neoxanthin and biological implications, *Biochem. Syst. Ecol.* **28**, 443–455.

Sugawara, Y., Hara, C., Aoki, T., Sugimoto, N., Masujima, T. (2000) Odor distinctiveness between enantiomers of linalool; difference in perception and responses elicited by sensory test and forehead surface potential wave measurement, *Chem. Senses* **25**, 77–84.

Thiemann, W.H., Rosenbauer, H., Meierhenrich, U.J. (2001) Conception of the "chirality-experiment" on ESA's mission ROSETTA to comet P46/Wirtanen, *Adv. Space Res.* **27**, 323–328.

Topiol, S., Sabio, M. (1989) Interactions between eight centers are required for chiral recognition, *J. Am. Chem. Soc.* **111**, 4109–4110.

Trapp, O., Schoetz, G., Schurig, V. (2002) Stereointegrity of thalidomide: gas-chromatographic determination of the enantiomerization barrier, *J. Pharm. Biomed. Anal.* **27**, 497–505.

Vandenabeele-Trambouze, O., Geffard, M., Bodet, D., Despois, M., Dobrijevic, M., Loustalot, M.-F.G., Commeyras, A. (2002) Antibodies directed against L and D valine using a chemical derivatizing reagent for the measurement of their enantiomeric ratio in extraterrestrial samples: first- step production and characterization of antibodies, *Chirality*, Published on-line, May 21, 2002, DOI 10.1002/chir.10116

Wang, H., Wolosker, H., Morris, J.F., Pevsner, J. Snyder S.H., Selkoe, D.J. (2002) Naturally occurring free D-aspartate is a nuclear component of cells in the mammalian hypothalamo-neurohypophyseal system, *Neuroscience* **109**, 1–4.

Welch, C.J. (2001) Formation of highly enriched microenvironments by stochastic sorting of conglomerate crystals: a plausible mechanism for generation of enantioenrichment on the prebiotic earth, *Chirality* **13**, 425–427.

Wheeler, G.L., Jones, M.A., Smirnoff, N. (1998) The biosynthetic pathway of vitamin C in higher plants, *Nature* **393**, 365–369.

Widder, S., Lüntzel, C.S., Dittner, T., Pickenhagen, W. (2000) 3-Mercapto-2-methylpentan-1-ol, a new powerful aroma compound, *J. Agric. Food Chem.* **48**, 418–423.

Wolosker, H., Blackshaw, S., Snyder, S.H. (1999) Serine racemase: a glial enzyme synthesizing D-serine to regulate glutamate-*N*-methyl-D-aspartate neurotransmission, *Proc. Natl. Acad. Sci. USA* **96**, 13409–13414.

Wright, C.V. (2001) Mechanisms of left-right asymmetry: what's right and what's left ? *Dev. Cell* **1**, 179–186.

Yost, H.J. (2001) Establishment of left-right asymmetry, *Int. Rev. Cytol.* **203**, 357–381.

Chlamydomonas

Jean-David Rochaix
University of Geneva, Geneva, Switzerland

1 The Organism 621

2 Three Autonomous Genetic Systems 623
2.1 Nuclear Genome 624
2.2 Chloroplast Genetic System 624
2.3 Mitochondrial Genome 628

3 Recent Technical Advances 629
3.1 Nuclear Transformation 629
3.2 Chloroplast Transformation 631

4 Chlamydomonas as a Model System 631
4.1 Function and Assembly of the Photosynthetic Apparatus 631
4.1.1 Mutations Affecting Structural Photosynthetic Genes 632
4.1.2 Nuclear Mutations Affecting Posttranscriptional Steps in the Expression of Chloroplast Genes 634
4.1.3 Metal Ion Control of Photosynthetic Gene Expression 636
4.1.4 Adaptation of the Photosynthetic Apparatus to Changes in Light Conditions 637
4.1.5 Heteroplasmicity of the Chloroplast Genome 637
4.2 Function and Assembly of the Flagellar Apparatus 638

5 Perspectives 641

Bibliography 642
Books and Reviews 642
Primary Literature 642

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30544-0

Keywords

BAC
Bacterial artificial chromosome, large cloned segment of chromosomal DNA that can be propagated in bacteria.

Cosmid
A plasmid that contains phage lambda cos sites and therefore can be packaged *in vitro* into phage coats.

DNA Complexity
The total length of different DNA sequences contained within a given DNA preparation; the DNA complexity is usually given in kilobase pairs (kb; 1 kb = 1000 base pairs).

EST
Expressed sequence tag, sequence of gene fragment obtained by high-throughput sequencing.

Intron
A segment of DNA within a gene that is transcribed but removed from the RNA transcript by splicing together the flanking coding sequences (exons).

Linkage Group
All loci that are linked together by genetic criteria, as an entity, and correspond to a chromosome.

Open Reading Frame
A DNA sequence that consists of a series of triplets (codons) coding for amino acids and potentially encodes a protein.

Splicing
Process of removing introns from a precursor RNA and joining of flanking exons.

The highly diversified, polyphyletic genus Chlamydomonas comprises more than 450 species of unicellular flagellated photosynthetic algae. These organisms are especially attractive for several areas of research in cell and molecular biology because of their small size, fast growth rate, and short sexual cycle. However, only a few species have been actively used for molecular genetic studies. They include the heterothallic species *C. reinhardtii* and *C. smithii*, which are in the line of descent leading to the multicellular species *Volvox*; the sibling species pair *C. eugametos* and *C. moewusii*, both very distinct from *C. reinhardtii*; and the homothallic species *C. monoica*. This article focuses on *C. reinhardtii* because this alga has emerged as the organism of choice for most investigations. *Chlamydomonas eugametos* and *C. moewusii* have also

been used for studies on mating and chloroplast inheritance. However, since in contrast to *C. reinhardtii* both these species are obligate photoautotrophs, their use is more limited. *Chlamydomonas monoica* is particularly suited for the analysis of the genetic control of sexuality, and several mutants defective in mating-type or zygote formation have been isolated and characterized.

Major technical advances have been achieved recently with *C. reinhardtii*. They include the development of reliable transformation methods for the nuclear, chloroplast, and mitochondrial compartments. Gene tagging and nuclear gene rescue of mutants with genomic libraries are also feasible. *Chlamydomonas* genomics projects have created large sets of ESTs that have greatly facilitated the cloning of nuclear genes of known function, and the *Chlamydomonas* nuclear genome sequence has been determined recently.

With these new tools, *C. reinhardtii* has become a powerful model system in which it is possible to perform an extensive molecular genetic dissection of fundamental biological processes, in particular photosynthesis, organellar biogenesis, and flagellar function and assembly.

1 The Organism

The characteristic features of *C. reinhardtii* are its unique cup-shaped chloroplast, which occupies nearly half of the cell volume, and its two flagella (Fig. 1). In this organism, photosynthetic function is dispensable, provided a reduced carbon source such as acetate is included in the growth medium. The alga can, therefore, be grown under three different regimes: phototrophic growth with CO_2 assimilated through photosynthesis as the unique carbon source, heterotrophic growth in the dark with acetate, and mixotrophic growth in the light with acetate. These growth properties have been used extensively to isolate and maintain numerous mutants deficient in photosynthetic activity. The cell division cycle of this alga can also be synchronized by subjecting cells to alternate light and dark cycles.

Haploid vegetative cells of *C. reinhardtii* exist as mating-type$^+$ or mating-type$^-$, determined by alternative alleles of a nuclear gene, and can be propagated indefinitely through mitotic divisions (Fig. 2). Gametogenesis is usually induced when vegetative cells are starved for nitrogen. Gametes formed from synchronously grown *C. reinhardtii* display several characteristic features such as loss of ribosomes, alterations of chloroplast morphology, starch accumulation, and reduced photosynthetic activity. The mixing of gametes of opposite mating type is followed by a rapid adhesion of their flagella. Since an individual gamete can adhere to more than one cell of opposite mating type, clumps of gametes are formed. This response appears to be mediated through gamete-specific glycoproteins called *agglutinins*, which are associated with the flagellar membrane and can be released as vesicles from the flagellar tips. This agglutination triggers flagellar-tip activation, loss of cell walls, and activation of mating structures, in which specialized regions differentiated at the cell anterior

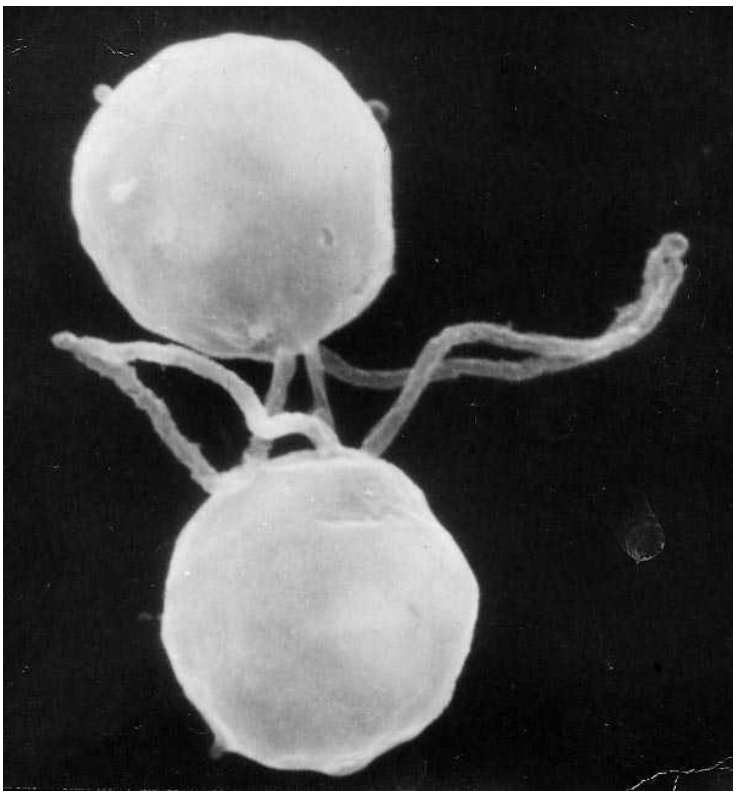

Fig. 1 Scanning electron micrograph of two *C. reinhardtii* gametes of opposite mating-type during the initial stage of mating (Courtesy of U. Goodenough).

give rise to the fertilization tube of the mating-type$^+$ cell. This series of complex reactions ultimately leads to the fusion of gametes and, subsequently, of their nuclei and chloroplasts. The resulting zygote matures into a thick-walled zygospore, which can be induced to undergo meiosis and produce a tetrad consisting of four haploid spores. Vegetative diploids can also be selected from matings or poly(ethylene glycol)-mediated cell fusions. They divide mitotically and allow one to determine whether a mutation is dominant or recessive. Vegetative diploids have also proven to be useful for isolating chloroplast and mitochondrial mutations affecting photosynthesis and respiration respectively. The most frequent mutations of this type are nuclear and recessive and, therefore, are not expressed in diploid cells, whereas chloroplast and mitochondrial mutations will segregate to yield mutant isolates.

Recently established methods of cytoduction allow the transfer of chloroplast or mitochondrial markers into a defined nuclear genetic background in *C. reinhardtii*. In this process, fusion of cytoplasms of the two parents occurs in the absence of nuclear fusion. This leads to the appearance of cytoductants containing the nuclear genome of one parent and cytoplasmic components of both parents. During further divisions, segregation occurs and the cytoplasmic marker of one

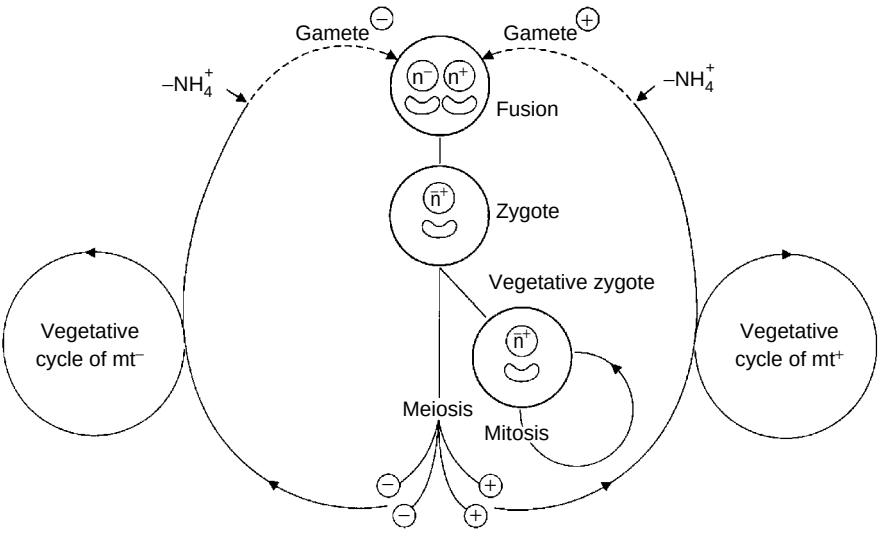

Fig. 2 Life cycle of *C. reinhardtii*: mt, mating-type; n, nucleus; the chloroplast is shown below the nucleus. See text for explanation. [Reproduced with permission from J.-D. Rochaix, FEMS *Microbiol. Rev.* **46**: 13–34 (1987).]

parent becomes associated with the nucleus from the other.

2 Three Autonomous Genetic Systems

Like higher plants, *Chlamydomonas* contains three genetic systems, located in the nucleocytosol, the chloroplast, and the mitochondria. In contrast to nuclear genes, which are transmitted to the offspring in a Mendelian fashion, chloroplast and mitochondrial DNAs are usually transmitted uniparentally from the mating-type$^+$ and mating-type$^-$ parent respectively. In *C. eugametos* and *C. moewusii*, both organellar genomes are transmitted uniparentally from the mating-type$^+$ parent (Table 1).

Tab. 1 Three genetic systems in *Chlamydomonas*.

System	*Species*	*Complexity of genome [kb]*	*Number of linkage groups*	*Copy number*	*Inheritance*
Nucleus	*reinhardtii*	10^5	17		Me
Chloroplast	*reinhardtii*	204	1	80	up$^+$
	eugametos	343	1		up$^+$
	moewusii	292	1		up$^+$
Mitochondria	*reinhardtii*	15.8	1	50	up$^-$
	eugametos	20.2	1		up$^+$
	moewusii	18.5	1		up$^+$

Notes: Me: Mendelian inheritance; up$^+$, up$^-$: uniparentally inherited from the mating-type$^+$ and mating-type$^-$ parent respectively.

2.1 Nuclear Genome

The complexity of the nuclear genome of *C. reinhardtii* has been estimated at 1×10^5 kb. At least 17 nuclear linkage groups have been identified with numerous markers (http://www.biology.duke.edu/chlamy_genome/nuclear_maps.html). These include auxotrophic mutations, drug-resistant markers, mutations affecting photosynthesis directly or indirectly, flagellar function, and mating. The use of restriction fragment length polymorphisms (RFLPs) to follow the segregation of specific DNA fragments in the genome has greatly accelerated genetic mapping. Efforts to establish RFLP maps for several of the nuclear linkage groups of *C. reinhardtii* are under way. A large collection of BAC clones covering the nuclear genome has been established and was used to sequence the entire genome. Telomere sequences from *C. reinhardtii* have been isolated and characterized. They are found at the chromosome ends with an average length of 300 to 350 bp, and they consist of the simple invariant repeat TTTTAGGG, which is very similar to its counterpart in *Arabidopsis thaliana*. Telomere-associated sequences, defined as the sequences close to the telomere repeats, appear to be useful markers for establishing RFLP maps.

Many nuclear genes of *C. reinhardtii* have been characterized. Genes that are highly expressed (e.g. those encoding proteins of the photosynthetic and flagellar apparatus or the heat shock proteins) display a strong codon bias. In general, cytidine (C) or guanine (G) residues at the third position of the codons are most common and, if possible, also in the first position, reflecting the high GC content (63%) of the nuclear DNA. However, genes that are less expressed, such as cytochrome c_6 and arylsulfatase, show a considerably more balanced codon usage. An A-rich sequence resembling the TATA box can usually be recognized in the upstream region of the nuclear genes of *C. reinhardtii* with an adjacent GC-rich stretch. In all the cases examined, a putative polyadenylation recognition motif TGTAA is found at the 3′ end, 10 to 15 nucleotides upstream of the polyadenylation site. Another distinctive feature of the nuclear genes of *C. reinhardtii* is the presence of multiple introns. As an example, the gene of one heat shock protein, hsp70, contains six introns, whereas its counterpart in plants contains only one intron, or none at all.

2.2 Chloroplast Genetic System

The chloroplast genome of *C. reinhardtii* has recently been sequenced and found to consist of 204 kb circular molecules. It is, therefore, larger than the plastid genomes of land plants, which range between 120 and 160 kb. Although the informational content of this DNA is low, about 0.2% of the cell DNA complexity, it constitutes 10 to 15% of the cellular DNA mass because it is present in approximately 80 copies per cell, which are arranged into 8 to 10 nucleoids within each chloroplast. The chloroplast DNA of *C. reinhardtii* is AT-rich and, therefore, its density differs considerably from that of nuclear DNA. This property can be used for purifying chloroplast DNA from total cell DNA by density gradient centrifugation. The striking differences in codon usage between nuclear genes and chloroplast genes suggest that these genomes are distinct in origin, in agreement with the endosymbiotic theory, according to which chloroplasts have evolved from photosynthetic prokaryotes that took up

an internal symbiotic relationship with the ancestor of eukaryotic plants and green algae. Other similarities between chloroplasts and photosynthetic bacteria supporting this theory are discussed in the paragraphs that follow.

Sequencing of the chloroplast genome of *C. reinhardtii* has revealed that it contains 99 genes. They include genes involved in photosynthesis, genes of the components of the chloroplast gene expression system, genes involved in posttranslational steps, and other genes of unknown function (Fig. 3, Table 2). Many of these genes are also found in the plastid genomes of higher and lower plants. Although the chloroplast genome of *C. reinhardtii* is larger than those of land plants, it has lost numerous coding regions that are still present in other chloroplast genomes. The large size of the chloroplast genome of *C. reinhardtii* is due to the presence of many

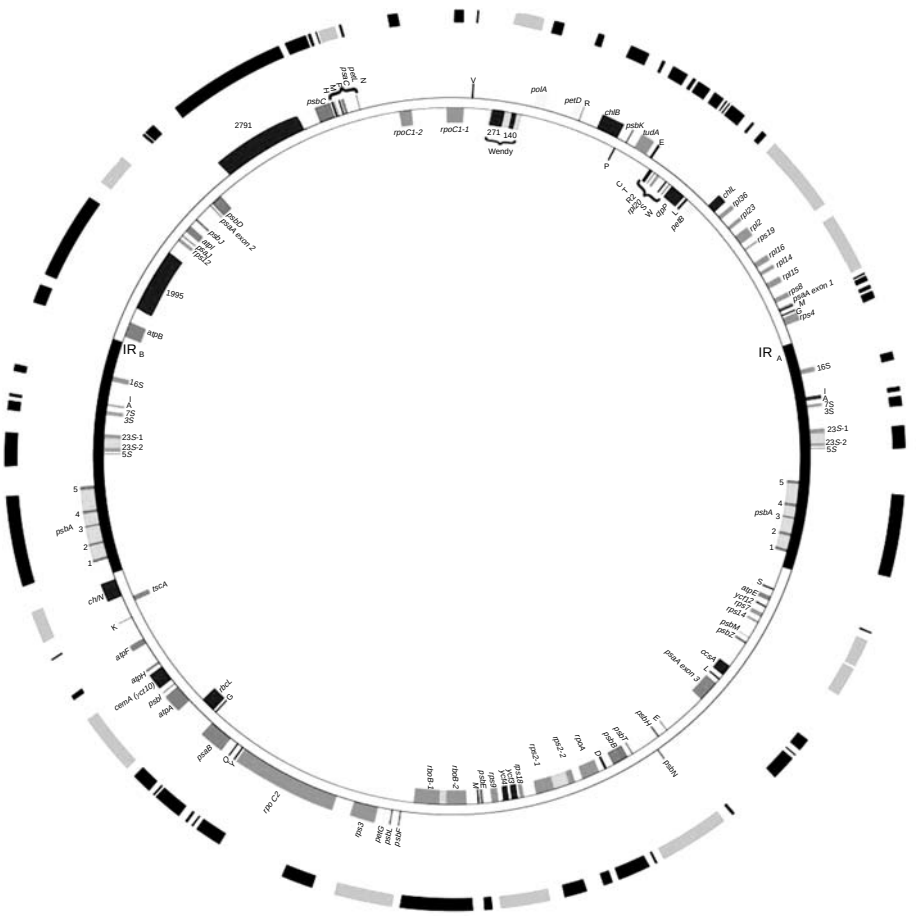

Fig. 3 Map of the chloroplast genome of *C. reinhardtii* with transcribed regions. See Table 2 for gene nomenclature. The outer circle indicates the transcribed regions: monocistronic RNAs (black), polycistronic RNAs (gray). [Reproduced with permission from Lilly et al. *Plant Cell* **14**, 2681–2708 (2002).]

Tab. 2 Chloroplast genes of *C. reinhardtii*.

Genes involved in chloroplast gene expression		
16S	16S rDNA	16S ribosomal RNA
23S, 7S, 3S	23S[a], 7S, 3S rDNA	23S, 7S, 3S ribosomal RNA
5S	5S rDNA	5S ribosomal RNA
Trn	A, C, D, E, E1, E2, F, G(2), H, I, K, L(2), M(3), N, P, Q, R, R2, S(2), T, V, W, Y	tRNAs
Rpl	2, 5, 14, 16, 20, 23, 36	Proteins of large ribosomal subunit
Rps	2, 3, 4, 7, 8, 9, 12, 14, 18, 19	Proteins of small ribosomal subunit
Rpo	A, B, C1a, C1b, C2	Subunits of RNA polymerase
tufA		Elongation factor EF-Tu
Genes involved in photosynthesis		
psa	A[a], B, C, J Yc3, Yf4	Subunits of photosystem I Assembly factors of photosystem I
Psb	A[a], B, C, D, E, F, H, J, K, L, M, N, T, Z	Subunits of photosystem II
Pet	A, B, D, G, L	Subunits of cytochrome b_6f complex
Atp	A, B, E, F, H, I	Subunits of ATP synthase
rbcL		Large subunit of ribulose 1,5-bisphosphate carboxylase
Chl	B, L, N	Subunits of light-independent protochlorophyllide oxidoreductase
Genes involved in other functions		
clpP		Subunit of ATP-dependent protease
ccsA		Involved in heme lyase
cemA		Protein required for facilitated CO_2/HCO_3 transport
Genes of unknown function		
ycf12	Conserved in other plastid genomes	Function unknown
ORFs	27 ORFs specific to *C. reinhardtii*	Function unknown

[a]Genes containing introns.
The arrangement of these genes on the chloroplast genome of *C. reinhardtii* is indicated in Fig. 3.

short dispersed repeats that account for more than 20% of its sequence. Moreover, the chloroplast genes of *C. reinhardtii* are usually larger than their land plant counterparts. While the organization of these genes on the chloroplast genomes of plants has been conserved to a remarkable extent, this conservation does not extend to Chlamydomonas. Even among *C. reinhardtii*, *C. eugametos*, and *C. moewusii*, the arrangement of chloroplast genes is rather different. The rearrangements of the chloroplast genome appear to be correlated with, and probably mediated through, the short dispersed repeats.

Besides the numerous genes that are found in plants and Chlamydomonas, several chloroplast genes have been identified in algae that are not present in the chloroplast genomes from land plants.

The *tuf*A gene encoding EF-Tu is located in the chloroplast genome in *C. reinhardtii* and in the nuclear genome in all higher and lower plants examined. Three chloroplast genes encoding subunits of the enzyme involved in light-independent protochlorophyllide reduction are present in *C. reinhardtii* but not in higher plants. Interestingly, these genes are also present in the chloroplasts of lower plants and pines that, together with *C. reinhardtii*, are capable of synthesizing chlorophyll in the absence of light. Finally, a few open reading frames (ORFs) of unknown function and unique to *C. reinhardtii* have been identified. Chloroplast genes that are absent from *C. reinhardtii*, but present in land plants include those encoding several ribosomal proteins and subunits of the NADH dehydrogenase. This enzyme is thought to mediate electron transfer from reduced stromal components to the plastoquinone pool of the photosynthetic electron transfer chain.

The chloroplast DNA is associated with small basic proteins and appears to be under torsional stress. Both relaxing and supercoiling topoisomerases have been detected in the chloroplast, and inhibition of the latter with novobiocin reduces torsional stress and alters chloroplast transcription *in vivo*. As in several prokaryotic organisms, the chloroplast DNA–dependent RNA polymerase of *C. reinhardtii* is sensitive to rifampicin. Some chloroplast promoters resemble typical prokaryotic promoters with characteristic −10 and −35 boxes. However, several chloroplast promoters lack the −35 box and can be as short as 22 bp. Transcription rates of chloroplast genes vary almost 100-fold. RNA stability can also vary considerably from one transcript to the other and, for all transcripts tested, the stability is significantly higher in dark- than in light-grown cells. Hence, changes in the abundance of chloroplast transcripts are regulated in a gene-specific manner and depend on both environmental and endogenous factors.

Splicing is an important posttranscriptional step in chloroplast gene expression. Most chloroplast introns fall into two major classes, group I and group II, first identified in yeast mitochondrial genes. Members of each class possess a characteristic secondary structure. Whereas introns are present in many chloroplast genes from plants, only three chloroplast genes from *C. reinhardtii* contain introns. Group I introns occur in the 23S ribosomal RNA and *psb*A genes of *C. reinhardtii*. Most of these highly structured introns are able to self-splice *in vitro*. Another unusual property of some of these group I introns is their ability to move at the DNA level, provided a suitable integration site is available. This process, called *intron homing*, was first discovered in yeast mitochondria. In crosses between intron-containing and intronless parents, all the progeny inherit the intron. Intron homing has also been found to occur in bacteriophage T4, in the slime mold *Physarum*, and in the chloroplast and mitochondria of *Chlamydomonas*. All the mobile group I introns contain an ORF encoding a protein with double-stranded DNA endonuclease activity that cleaves the intronless allele at the site at which the intron is inserted. Cleavage triggers transposition of an intron copy to this target site and leads to efficient spreading of the intron to all target sites available. Besides cis-splicing there is trans-splicing in the chloroplast of *C. reinhardtii* during the maturation of the *psa*A message (cf. Sect. 4.1.1).

The chloroplast translational apparatuses of *Chlamydomonas* and plants share

several features with those of prokaryotic organisms: chloroplast ribosomes sediment at 70S and are sensitive to the same spectrum of antibiotics as prokaryotic ribosomes. There is considerable sequence identity between the chloroplast RNAs and bacterial ribosomal RNAs, and some chloroplast translation initiation factors can substitute to some extent for their *E. coli* counterparts *in vitro*. There are, however, some differences. Although several chloroplast genes of *C. reinhardtii* have sequences resembling the Shine–Dalgarno sequences preceding the initiation codon, this is not a general rule. In contrast to *E. coli*, the spacing between the putative Shine–Dalgarno sequence and the initiation codon is not highly conserved. In further contrast to *E. coli*, in which most of the translated messages are polycistronic, chloroplast genes are usually transcribed and translated as monocistronic mRNAs in *C. reinhardtii*.

An unusual feature of *C. reinhardtii* is that expression of the chloroplast *clpP* gene appears to involve protein splicing or posttranslational polypeptide ligation. The *clpP* gene encodes one subunit of the chloroplast ATP-dependent protease. The other subunit is encoded by the nuclear genome. Together, the 5′ and 3′ parts of the *clpP* gene encode a polypeptide that is homologous and colinear with the *clpP* proteins of other organisms. However, the 858 bp middle part of the *clpP* gene does not contribute to the 22 kDa *clpP* subunit. This extra sequence is not excised at the RNA level because it does not display structural features typical of chloroplast introns and because a single transcript accumulates that includes both the *clpP* coding sequence and the extra sequence. Therefore, this region is thought to be removed by protein splicing.

2.3 Mitochondrial Genome

The mitochondrial genome of *C. reinhardtii* consists of 15.8-kb linear molecules and is considerably smaller than the mitochondrial genomes of plants, which range between 200 and 2400 bp. Sequencing of the entire mitochondrial DNA of *C. reinhardtii* has revealed that it encodes eight protein genes, three tRNA genes, and two ribosomal RNA genes (Fig. 4). The proteins encoded by the mitochondrial genome of *C. reinhardtii* are cytochrome *b*, subunit 1 of cytochrome oxidase, five subunits of NADH dehydrogenase, and a reverse transcriptase–like protein whose function is not yet known. The ends of the mitochondrial DNA form an inverted repeat. The most unusual feature of this DNA is that the two mitochondrial ribosomal RNA genes are split into several smaller coding modules, which are scattered over nearly half the mitochondrial genome and are interspersed with protein and tRNA genes. Expression of these multiple gene fragments leads to the accumulation of low- molecular weight subribosomal RNAs that can base-pair with each other to form the characteristic secondary structure of the large and small ribosomal RNAs. The presence of only three mitochondrial tRNA genes strongly suggests that the remaining mitochondrial tRNAs are nucleus-encoded and imported into the mitochondria, as has been demonstrated for some plant mitochondrial tRNAs. Another surprising feature is that, in marked contrast to other organisms, the genes for subunits II and III of cytochrome oxidase and for any of the subunits of ATP synthetase are not encoded by the mitochondrial genome of *C. reinhardtii*.

None of the mitochondrial genes of *C. reinhardtii* contain introns. However, the

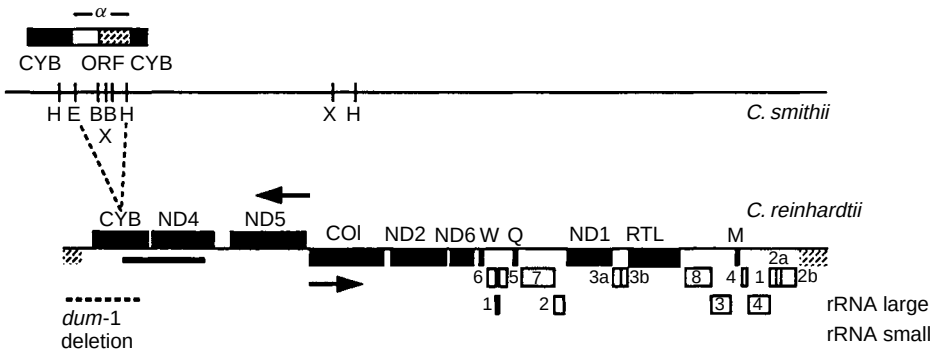

Fig. 4 Physical map of the mitochondrial genomes of *C. reinhardtii* and *C. smithii*. Genes for mitochondrial proteins: CYB, cytochrome *b*; COI, subunit I of cytochrome oxidase; RTL, reverse transcriptase–like protein; ND 1–6, subunits of NADH dehydrogenase. The dispersed arrangement of the ribosomal RNA gene modules is shown; W, Q, M are genes of tRNA Trp, Gln, Met respectively. Restriction site polymorphisms are indicated for *Bam*HI (B), *Eco*RI (E), *Hpa*I (H), and *Xba*I (X). α refers to the intron within the cytochrome *b* gene in *C. smithii* that contains an open reading frame (ORF). Arrows indicate direction of transcription. (Reproduced with permission from E.H. Harris, *Genetic Maps*, Vol. 6, Cold Spring Harbor Press, Plainview, NY, 1993, pp. 2157–3169.)

mitochondrial cytochrome *b* gene of *C. smithii*, which is interfertile with *C. reinhardtii*, contains a mobile group I intron that has properties similar to those of the chloroplast ribosomal intron described in Sect. 2.2. The mitochondrial genomes of *C. eugametos* and *C. moewusii* consist respectively of 24 and 22 kb circular DNA molecules that are colinear with one another. Although these two mitochondrial genomes appear to contain the same set of protein-coding genes as the mitochondrial DNA of *C. reinhardtii*, the genes are strikingly different in their organization.

Mutants with deletions in the apocytochrome *b* gene have been isolated and characterized. These mutants require light for growth and are unable to grow on acetate medium in the dark, indicating that cytochrome *b* is dispensable in the light. Mitochondrial mutants resistant to the respiratory inhibitors myxothiazol and mucidin have been shown to carry point mutations in the cytochrome *b* gene.

3 Recent Technical Advances

A major breakthrough in the area of molecular studies on *Chlamydomonas* was the establishment of reliable methods for transformation of the nuclear, chloroplast, and mitochondrial compartments, and the establishment of a large EST collection.

3.1 Nuclear Transformation

Nuclear transformation can be easily performed with cell wall–deficient *C. reinhardtii* mutants by vortexing the cells with glass beads and DNA or by electroporation. Transformation of the walled strains can also be achieved with the same techniques after removing the wall with autolysin, a glycoprotein that is secreted when gametes of both mating types are mixed together. The most widely used host strains are *arg7*, deficient in argininosuccinate 1yase, and *nit-1*, deficient in nitrate reductase.

Selection is performed on medium lacking arginine or ammonium, respectively, by transformation with the corresponding wild-type genes. It is also possible to transform wild-type strains using markers that confer resistance to several drugs such as emetine, phleomycin, paromomycin, and spectinomycin. The latter three are of bacterial origin and need to be driven by Chlamydomonas promoters for expression in the algal cells. Although several foreign drug-resistant markers have been used successfully in Chlamydomonas cells, efficient expression of foreign genes remains a difficult task. It is also possible to introduce additional genes into the nuclear genome by cotransformation. In most cases, transformation occurs through nonhomologous recombination, as the transforming DNA appears to integrate at random sites of the nuclear genome. This property has been used successfully for tagging genes. In this approach, new mutations are induced through the integration of the transforming DNA into nuclear genes. The bacterial vector sequences in the transforming DNA can then be used as a probe for isolating the mutated gene. Two mobile elements of *C. reinhardtii* – Gulliver, which resembles classical transposable elements, and *TOC1*, which is related to retrotransposons – have been characterized and can also be used for nuclear gene tagging. The high nuclear transformation yield has made it possible to perform gene-rescue experiments by complementing nuclear mutations with genomic cosmid libraries of *C. reinhardtii*.

While nuclear transformation is efficient when cloned *C. reinhardtii* nuclear genes are used as the transforming DNA, it has been difficult to express foreign genes efficiently in this organism even when they are fused to *C. reinhardtii* promoters and 3′ untranslated regions of abundant mRNAs. It is not yet clear whether this is due to the biased codon usage of nuclear genes of *C. reinhardtii*, to methylation of the foreign DNA sequences, or to other factors. Codon bias is clearly important as shown by the successful expression of GFP (green fluorescent protein) when its gene is reconstructed with the *C. reinhardtii* codon usage. Moreover, it is possible to express chimeric genes in the nuclear compartment, provided all gene sequences originate from *C. reinhardtii*. Transcriptional fusions between the *cabII-1* and the *nit-1* genes, coding for a light-harvesting chlorophyll a/b binding protein and the nitrate reductase respectively have been introduced into the nuclear genome and expressed. These chimeric constructs consist of the *cabII-1* promoter region fused at its transcriptional initiation site to the *nit-1* gene. In wild-type cells, expression of the *cabII* gene is strongly increased in the light, whereas the *nit-1* gene is repressed in the presence of ammonium and is expressed in media containing nitrate. Expression of the *nit-1* gene requires a positive-acting regulator encoded by the *nit-2* gene. In the transformants, expression of the chimeric *cabII-1/nit-1* gene was also stimulated by light. However, the *nit-1* mRNA accumulated to a much lower level than the *cabII-1* mRNA. Expression could be slightly stimulated by adding further upstream sequences of the *cabII-1* promoter to the chimeric construct. Expression of the chimeric gene occurred in the presence of ammonium and no longer depended on the *Nit-2* product. This is a useful property since many mutations in *C. reinhardtii* have been isolated in a *nit-1 nit-2* genetic background. Similar chimeric genes have been produced with the tubulin β_2 promoter and the gene encoding periplasmic arylsulfatase. As observed for the tubulin

gene, transcription of the chimeric gene was strongly increased after cells had been deflagellated, although the level of mRNA from the chimeric genes was only 5 to 10% of that from the endogenous β_2 tubulin gene.

3.2 Chloroplast Transformation

Chloroplast transformation of *C. reinhardtii* can be achieved when cells are bombarded with DNA-coated tungsten particles from a particle gun. Chloroplast mutants carrying a defective photosynthetic gene are usually transformed with the corresponding wild-type gene, which integrates into the chloroplast genome by homologous recombination. Alternatively, nonphotosynthetic markers such as ribosomal RNA genes with mutations, conferring resistance to streptomycin and spectinomycin, can be used for selection. Resistance to these antibiotics can also be obtained with the bacterial *aadA* (aminoglycoside adenyl transferase) gene fused to a chloroplast promoter and 5′ leader region. These tools have opened the door for genetic engineering of the chloroplast genome. It is now possible to perform chloroplast gene disruptions and site-directed mutagenesis, and to insert and express foreign genes (e.g. chimeric GUS constructs) at specific sites in the chloroplast genome.

4 Chlamydomonas as a Model System

4.1 Function and Assembly of the Photosynthetic Apparatus

The primary reactions of photosynthesis occur at the thylakoid membranes, in which light energy is collected and converted into chemical energy through charge separations across the membrane and a series of complex oxidoreduction reactions. Ultimately, the process leads to the formation of an electrochemical gradient across this membrane and the production of ATP and NADPH, both of which are required to drive the Calvin cycle, which results in CO_2 fixation and the synthesis of carbohydrates. Four multimolecular complexes are involved in these primary reactions: photosystem II and photosystem I and their associated light-harvesting systems; the cytochrome b_6f complex; and ATP synthetase (Fig. 5). Table 2 lists some of the major subunits of photosystems II and I and their genes.

As in higher plants, the biosynthesis of the photosynthetic apparatus of *C. reinhardtii* occurs through the concerted action of two genetic systems located in the nucleus and chloroplast respectively. Several subunits of the photosynthetic complexes are encoded by the chloroplast genome and translated on 70S chloroplast ribosomes. The remaining subunits are encoded by nuclear genes and translated on cytosolic 80S ribosomes as precursors with a transit peptide at the N-terminal end, which targets the protein to the chloroplast compartment. Upon import of the protein into the chloroplast, the transit peptide is cleaved by a stromal peptidase. Proteins targeted to the thylakoid lumen have a bipartite transit sequence: the N-terminal part acts as a chloroplast transit sequence and is cleaved in the stroma; the C-terminal part always contains a hydrophobic region that is required for translocation across the thylakoid membrane, and this region is cleaved by a second protease, which is tightly associated with the thylakoids. In the final steps, chloroplast- and nuclear-encoded subunits are assembled either in

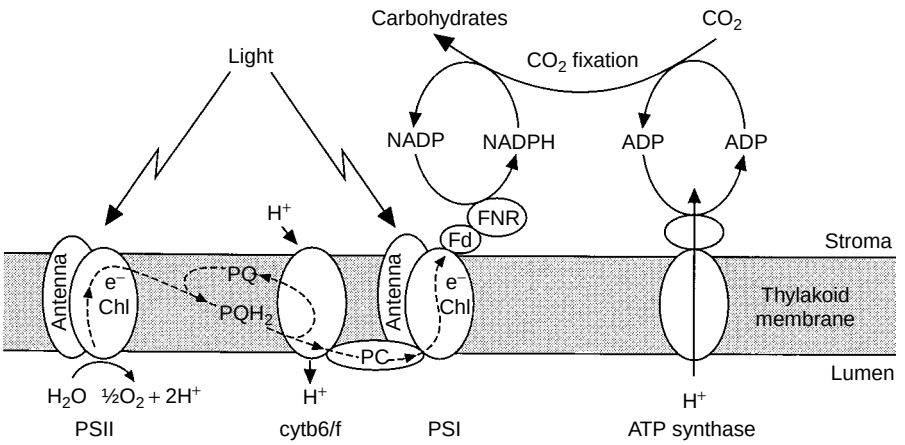

Fig. 5 Scheme of photosynthetic electron transport chain showing the four complexes – photosystem II (PSII) and photosystem I (PSI) with their associated chlorophyll antennae, the cytochrome *b6/f* complex (cytb6/f), and the ATP synthetase. Fd, ferredoxin; FNR, ferredoxin-NADP reductase; PQ, plastoquinone; PQH_2, plastoquinol; chl, chlorophyll.

the stroma or on the thylakoid membranes to form functional protein complexes.

Because photosynthetic function is dispensable when *C. reinhardtii* cells are grown on a medium containing acetate, it has been possible to isolate numerous nuclear and chloroplast mutants deficient in photosynthetic activity. An important feature of the thylakoid membrane is that the physical–chemical properties of its numerous pigments depend critically on the functional state of photosynthetic complexes. Fluorescence has, therefore, been used as a powerful, noninvasive method for screening photosynthetic mutants. Mutations affecting photosynthesis fall into two major classes. The first includes mutations within genes encoding components of the photosynthetic system. The second class includes mostly nuclear mutations that act indirectly on photosynthesis. Many of these mutations affect genes whose products are required for the proper expression of chloroplast genes. These factors appear to act at several posttranscriptional levels (e.g. chloroplast RNA stability, RNA processing and splicing, translation) and, most probably, at the level of assembly of the photosynthetic complexes. Surprisingly, the number of the nuclear loci involved in chloroplast gene expression is quite large and most of their products appear to act in a gene-specific manner (Fig. 6).

4.1.1 Mutations Affecting Structural Photosynthetic Genes

Analysis of photosynthetic mutants has revealed that several are affected in the structural genes of the photosynthetic complexes. Characterized chloroplast mutations comprise deletions, sequence duplications, and single–base pair changes. Among the latter, herbicide-resistant mutants have been especially useful for probing the structure–function relationship of the D1 subunit of photosystem II. Earlier studies had revealed that the absence of any of the subunits from the core complex of photosystem II or I usually leads to a drastic destabilization of the other subunits of the complex. It

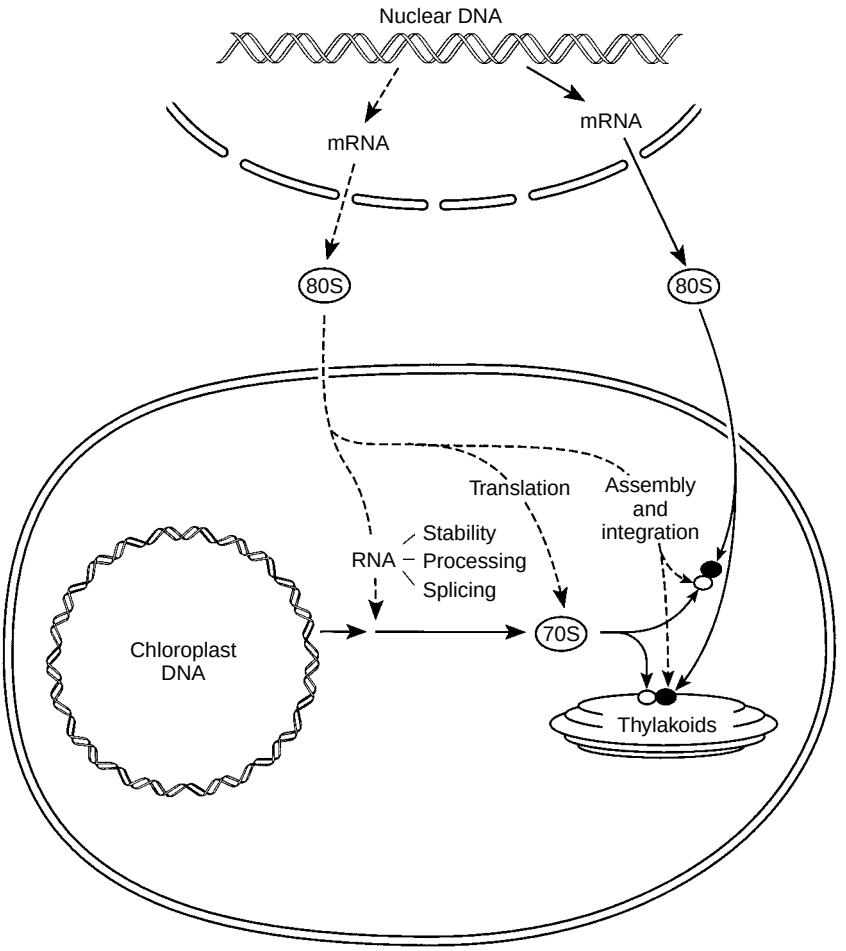

Fig. 6 Biosynthesis of the photosynthetic apparatus. The nucleus and chloroplast are shown in the upper and lower parts, respectively; 70S and 80S represent the chloroplast and cytosolic ribosomes. Synthesis of structural photosynthetic proteins is indicated by continuous lines. The roles of nuclear-encoded factors in posttranscriptional steps of chloroplast gene expression are shown by dashed lines.

has been possible to perform targeted disruptions of several of the smaller photosystem I and II genes in the chloroplast genome of *C. reinhardtii* using the *aad*A expression cassette. It is noteworthy that most of these gene disruptions have a more drastic effect on the stability of the complexes than that of similar disruptions in cyanobacteria. For example, disruption of either the *psbO* or the *psbK* gene in cyanobacteria still allows for photoautotrophic growth in cyanobacteria, but a similar disruption results in the destabilization of photosystem II in *C. reinhardtii*. Similar differences in the stability of the photosystem I complex between *C. reinhardtii* and cyanobacteria have been observed, with mutants carrying directed, specific amino acid substitutions in the PsaC protein, the apoprotein of

the terminal electron acceptors of photosystem I. One possibility is that the clearing system, which recognizes misfolded protein complexes and degrades them subsequently, works more efficiently in *C. reinhardtii* than in cyanobacteria. Site-directed mutagenesis of chloroplast genes encoding subunits of photosystem II, photosystem I, and the cytochrome b_6f complex through chloroplast transformation has been performed extensively and has provided new and important insights into the structure–function relationship of these complexes.

While chloroplast transformation occurs in all cases examined through homologous recombination and allows one to perform gene disruptions or site-directed mutagenesis in the chloroplast genome, nuclear transformation appears to lead, in most cases, to random insertions within the nuclear genome. Attempts to disrupt nuclear genes through homologous recombination have not yet succeeded. However, it has been possible to use transforming DNA as a random mutagen and, thus, to inactivate specific nuclear genes by using suitable screens. This was first achieved with the nuclear *atp*C gene encoding the *CF1* γ subunit of ATP synthetase. Wild-type cells were transformed with a nonfunctional homologue of the *atp*C gene in the presence of carrier DNA. After enrichment for photosynthetic mutants and an immunoscreen, a mutant was isolated that lacked the γ-subunit because its *atp*C gene had been rearranged. Similarly, cells from the *arg*7 mutant were cotransformed with the argininosuccinate lyase gene and the *psa*F gene that had been mutated by three small internal deletions. Transformants were selected for arginine prototrophy and screened for fluorescence patterns typical of photosystem I mutants, and by DNA analysis. One transformant that was obtained lacked *Psa*F and had suffered a large rearrangement at the *Psa*F locus. It appears, therefore, that transformation is a valuable tool for disrupting specific nuclear genes of *C. reinhardtii*, provided suitable screens for the mutant phenotype are available.

4.1.2 Nuclear Mutations Affecting Posttranscriptional Steps in the Expression of Chloroplast Genes

Posttranscriptional steps are important in the control of chloroplast gene expression. *C. reinhardtii* offers unique advantages for genetic analysis of posttranscriptional events. Several nuclear mutations destabilize specifically defined chloroplast mRNAs. Mutations of this type have been reported for *psbB*, *psbD*, *atpA*, *atpB*, and *rbcL* mRNA. In these mutants, all other chloroplast transcripts accumulate normally. It has been possible to identify the target region for *psbD* and *psbB* RNA degradation in nuclear mutants that do not accumulate these mRNAs. For this purpose, either the 5′ or 3′ untranslated regions of these mRNAs were fused to the reporter *aadA*, which confers spectinomycin resistance. These chimeric genes were introduced into the chloroplast genome of a wild-type strain and the transformants of mating-type$^+$ were crossed to either of the original mutants of mating-type$^-$. Because of the uniparental inheritance of the chloroplast genome, all the offspring from these crosses inherit the chloroplast chimeric gene, whereas the nuclear mutations segregate 2 : 2. If the 5′ untranslated region of *psbD* or *psbB* is sufficient to confer instability to the chimeric RNA, there will be cosegregation between spectinomycin sensitivity and photosynthetic deficiency. This cosegregation has indeed been shown to occur among the progeny, and the loss of chimeric RNA in

the mutant nuclear background has been confirmed. Recently, the genes that are deficient in these two mutants have been cloned and characterized. In both cases, they encode proteins consisting mainly of 9–10 TPR (tetratricopeptide) repeats that are present in a wide range of proteins with distinct biological function and that appear to play an important role in protein–protein interactions.

A molecular genetic analysis is also feasible at the level of RNA processing or splicing. The *psaA* gene, encoding one of the reaction center subunits of photosystem I, consists of three exons that are widely scattered on the chloroplast genome of *C. reinhardtii*. Each exon is flanked by sequences characteristic of group II introns, and maturation of the *psaA* message depends on two trans-splicing reactions. Surprisingly, intron 1 itself has a tripartite structure: the 5′ part is fused to exon 1; the middle part is encoded by a locus, *tscA*, that is distant from both exons 1 and 2; and the 3′ part is fused to exon 2 (Fig. 7). The second *psaA* intron has a bipartite

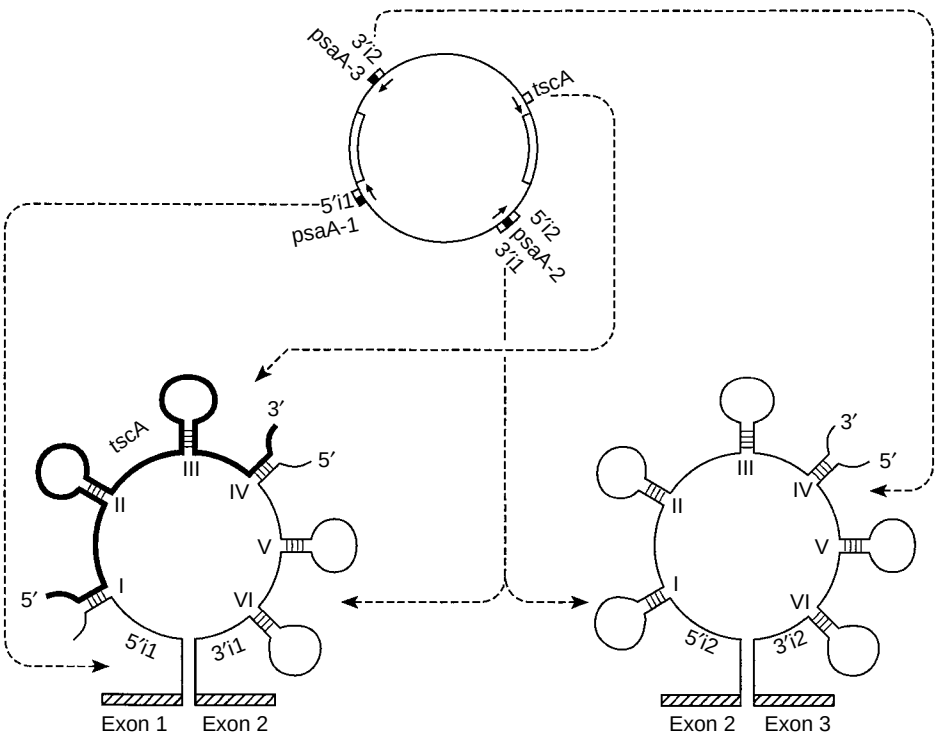

Fig. 7 Maturation of the chloroplast *psaA* mRNA of *C. reinhardtii*. Upper part: the three exons *psaA-1*, *-2*, *-3* and their flanking group II intron sequences (5′i1, 3′i1, 5′i2, 3′i2) and *tscA* (not drawn to scale) are shown on the circular chloroplast genome. Arrows indicate the direction of transcription, and the two open boxes represent the two segments of the chloroplast inverted repeat. Lower part: secondary structure of the split-group II intron of *psaA* showing the six characteristic domains. Intron 1 (left) has a tripartite structure: the *tscA* RNA (heavy line) base-pairs with the 5′ region of intron 1 within domain I, completes the catalytic core by forming domains II and III, and pairs with the 3′ part of intron 1 within domain IV. Intron 2 (right) has a bipartite structure; pairing occurs within domain IV.

structure. Numerous mutants affected in *psaA* trans-splicing have been isolated and can be grouped into three classes. In the first are mutants unable to splice exons 1 and 2, the second comprises mutants unable to splice exons 2 and 3, and the third consists of mutants unable to perform either trans-splicing reaction. At least 14 nuclear loci are involved in this complex maturation pathway. Whether the unusual split structure of the *psaA* gene reflects an ancient gene structure or whether it was created through division of an ancestral *psaA* gene by DNA rearrangements remains an open question. The nuclear genes of several of the factors involved in trans-splicing have been cloned and characterized. Some of them resemble known proteins involved in RNA metabolism such a pseudouridine synthases, although the enzymatic activity is not required for trans-splicing. Whether these proteins have a dual role in chloroplast RNA maturation is not yet clear. Some of these factors are part of high molecular weight RNA protein complexes that are reminiscent of snRNPs in eukaryotic cells, although no significant sequence identity is apparent between the subunits of these complexes.

Mutants affected in chloroplast translation have also been examined. These mutants were first identified on the basis of pulse labeling of cells in the presence of an inhibitor of cytoplasmic protein synthesis. Under these conditions, only polypeptides translated in the chloroplast compartment are labeled. From the labeling patterns, it is, however, difficult to distinguish mutants truly affected in translation from those having increased protein turnover. By using a strategy resembling the one based on the use of chimeric genes outlined earlier for the nuclear mutant deficient in *psbD* RNA accumulation, it has recently been shown that some of the mutants are indeed affected at the level of initiation of translation. Several of the trans-acting factors have been recently identified (Table 2).

4.1.3 Metal Ion Control of Photosynthetic Gene Expression

The copper-containing protein plastocyanin catalyzes electron transfer between the cytochrome b_6f complex and photosystem I (Fig. 5). Growth of terrestrial plants in copper-deficient habitats leads to symptoms of copper deficiency. However, *Chlamydomonas*, like other aquatic algae and cyanobacteria, remains photosynthetically competent under copper-deficient growth conditions. This is because, in the absence of copper, the alga expresses the heme-containing cytochrome c_6, an alternate electron carrier to plastocyanin. Although plastocyanin is still expressed under conditions of copper deficiency, the apoprotein is highly unstable and is rapidly degraded. Expression of cytochrome c_6 proceeds only in the absence of copper and is regulated at the transcriptional level. The cupric ion–responsive promoter of the cytochrome c_6 gene displays unique metal specificity and high sensitivity; in addition, it governs reciprocal control of synthesis of two proteins, cytochrome c_6 and plastocyanin, which are distinct although functionally equivalent. On the basis of the observation that the amount of copper ions needed to repress cytochrome c_6 transcription matches the amount of plastocyanin in *Chlamydomonas* cells, it has been proposed that repression of transcription requires a copper-binding factor that is titrated by copper only after plastocyanin has accumulated to a level needed for photosynthesis. Therefore, the control of cytochrome c_6 expression results mainly from a direct cellular sensing of available copper rather than indirectly as

a consequence of loss of plastocyanin or decreased photosynthetic activity in copper-deficient cells.

4.1.4 Adaptation of the Photosynthetic Apparatus to Changes in Light Conditions

Like other algae and land plants, *C. reinhardtii* has the remarkable ability to adapt its photosynthetic machinery to changes in light quantity and quality. Photosynthetic organisms can dissipate the light excitation energy by photochemistry, by fluorescence, or by nonphotochemical quenching. The latter occurs under excessive illumination, is triggered by the increased proton gradient across the thylakoid membrane, and involves a reversible structural modification of the carotenoids through the xanthophyll cycle. This leads to an increased heat dissipation of the excitation energy and lowers the fluorescence emission and the photooxidative damage within the photosynthetic reaction centers. A genetic approach has identified several factors involved in nonphotochemical quenching and has provided direct evidence for the involvement of the xanthophylls in this process.

Photosystem II and photosystem I act in series in the photosynthetic electron transport chain, and they are connected to two distinct antennae systems with different light absorption properties. Upon a change in the spectral quality of the exciting light, a reorganization of the antennae occurs, insuring a balanced excitation of the two photosystems and hence an optimal photosynthetic quantum yield. This process is called *state transition* and involves the displacement of the antenna of photosystem II to photosystem I under conditions in which photosystem II is overstimulated relative to photosystem I. A key step of this mechanism is the activation of a kinase that specifically phosphorylates the N-terminal end of the LHCII proteins. The activation is triggered through a signal transduction chain that involves the redox state of the plastoquinone pool and a functional cytochrome b_6f complex. Attempts to isolate the kinase by biochemical means have failed. However, a genetic approach has identified several mutants that are deficient in state transition and blocked in the phosphorylation of LHCII. Recently, the gene deficient in one of these mutants has been isolated and characterized and has been found to encode a thylakoid-associated kinase. Mutants of this sort offer promising possibilities for identifying the different factors involved in state transition.

4.1.5 Heteroplasmicity of the Chloroplast Genome

It is generally assumed that chloroplast genomes consist of identical copies of single DNA molecules. The predominantly uniparental inheritance of chloroplast genomes could easily maintain such homoplasmicity. Although most heteroplasmic markers segregate rapidly to form homoplasmic cells, cases of stable heteroplasmicity can occur in the chloroplast of *C. reinhardtii*.

Mutants deficient in photosynthetic activity have been shown to arise from nonsense mutations within the chloroplast *rbcL* gene encoding the large subunit of ribulose 1,5-bisphosphorate carboxylase/oxygenase. In these mutants, the holoenzyme is undetectable because of rapid degradation of the unassembled and truncated subunits. Photosynthetically competent suppressors of these nonsense mutants have been found to be heteroplasmic, giving rise to both mutant and suppressor cells during divisions under

nonselective growth conditions. No homoplasmic suppressor segregants could be obtained even after repeated cloning or crosses under phototrophic growth conditions. In contrast, photosynthetic-deficient segregants quickly became homoplasmic under heterotrophic growth conditions. The molecular basis of the heteroplasmicity has been determined for a suppressor of an amber (UAG) mutation. In this suppressor, the original mutation is still present in all chloroplast DNA copies. However, 70% of the copies of tRNA *trp* have the tryptophan CCA anticodon changed to the amber-specific CUA. Under phototrophic growth conditions, therefore, stable heteroplasmicity can arise as a balanced polymorphism of suppressor and wild-type alleles of a tRNA gene within the chloroplast genome. In this case, the suppressor allele restores ribulose 1,5-bisphosphate carboxylase, whereas the wild-type allele is required for normal protein synthesis. When photosynthesis is not required, the suppressor tRNA allele is lost because of random segregation, giving rise to homoplasmic photosynthetic-deficient segregants.

As mentioned earlier, it is possible to disrupt chloroplast genes encoding components of the photosynthetic apparatus using transformation and the *aadA* expression cassette that confers spectinomycin resistance in the chloroplast. The transformants that are obtained usually become homoplasmic after a few cloning steps. However, disruptions of chloroplast genes that are essential under all growth conditions never give rise to homoplasmic transformants. In this case, stable heteroplasmicity is maintained as long as spectinomycin is present in the growth medium. Stable heteroplasmicity can, therefore, be used for identifying chloroplast genes encoding essential functions.

4.2 Function and Assembly of the Flagellar Apparatus

Chlamydomonas reinhardtii possesses two flagella, located at the anterior end of the cell, that are assembled on basal bodies. During cell division, basal bodies migrate to the interior of the cell and function as centrioles by organizing the spindle apparatus. The flagellar system of *Chlamydomonas* has proven to be particularly well suited for studying microtubule assembly and function, and motility. This is because flagellar biosynthesis can be readily synchronized, and numerous mutants affected in the function and assembly of the flagellar apparatus have been isolated. These mutants can be separated into two major classes: those with abnormal or no motility, usually called *paralyzed mutants* (with the acronym pf), and those defective in flagellar assembly (fla).

Extensive ultrastructural and biochemical studies have revealed that the flagellae consist of a set of nine outer doublets, each consisting of two microtubules A and B, and a central pair of microtubules (Fig. 8). Outer and inner arms arise from A microtubules of the outer doublets. The outer arms, which comprise dyneins and large multisubunit ATPases, extend toward the B tubules, where they act to generate interdoublet sliding. Outer-arm dyneins consist of α-, β-, and γ-heavy chains and of intermediate and light chains. Partial cDNA sequences of the γ-heavy chain and of an unidentified heavy chain of *C. reinhardtii* have revealed that these proteins have at least two conserved domains corresponding to the ATP hydrolytic site and to a region related to the microtubule binding domain of the kinesin superfamily. A mutant lacking outer arms could be rescued by transformation with a genomic clone

Fig. 8 Cross section of the *C. reinhardtii* flagellum showing the nine outer doublets with inner and outer dynein arms and the radial spokes extending toward the central pair of tubules; bar indicates 0.05 m. (Reproduced with permission from Witman, in *Ciliary and Flagellar Membranes*, R.A. Bloodgood, Ed., Plenum Press, New York, 1990, pp. 1–30.)

encoding a 69 kDa intermediate chain, thus showing that a defect in this chain can interfere with outer-arm assembly. The α-heavy chain does not appear to be required for stable outer-arm assembly because a mutant missing the α-chain and its associated 16 kDa light chain is still able to form a partially functional truncated outer arm. Two morphologically different types of inner arm have been identified. However, biochemical and genetic studies indicate the existence of at least five different types of inner arms, consisting of homodimers and heterodimers of six inner-arm heavy chains, and of smaller polypeptides. Mutants lacking the outer arms or some of the inner arms are still mobile, showing that none of these arms is essential for interdoublet sliding. Since the former are affected in flagellar beat frequency and the latter in the waveform, however, it is clear that functional differences exist between the arms.

The most abundant flagellar proteins are the α- and β-tubulins, each of which is encoded by two genes. Both β-genes encode identical proteins, while the products of the two α-genes differ by two amino acids. Besides the tubulins, as many as 250 to 300 distinct flagellar polypeptides can be resolved. Analysis of many paralyzed mutants has revealed that some are deficient in specific axonemal components. The defects observed in these mutants have been found to cosegregate with the mutant phenotype. Mutants lacking the central pair of

tubules have rigid flagellae, while mutants lacking functional radial spokes are usually paralyzed. Radial spokes participate in the regulation of dynein arm activity and, thus, in flagellar bending. The radial spoke consists of a thin stalk attached to the A tubule of the outer doublet and of a head that projects close to the central pair of microtubules. A set of 17 axonemal polypeptides appears to be associated with radial spokes because radial spokeless mutants lack all these polypeptides. Several of the corresponding genes have been isolated and characterized.

A common feature observed in many paralyzed mutants is the loss of a set of polypeptides corresponding to an entire complex, presumably because loss of one subunit leads to increased turnover of the other subunits or because the missing subunit is important for anchoring the complex. In some cases, it has been possible to identify the specific polypeptide that is affected by the mutation by performing a dikaryon rescue experiment. During wild-type crosses, biflagellate gametes of plus and minus mating-type produce quadriflagellate temporary dikaryons. It was originally shown that flagellar motility is restored in pairwise crosses between paralyzed mutants and the wild-type, suggesting that the polypeptides contributed by both parents can be assembled into functional flagellar complexes after mating. This approach has been extended by labeling the proteins of the mutant and by mating the labeled mutant cells with unlabeled wild-type cells. Restoration of flagellar function is allowed to occur in the presence of a protein synthesis inhibitor to block new protein synthesis. Under these conditions, the polypeptide that is deficient in the mutant appears only in the unlabeled form, since it is derived from the wild-type parent, whereas the other normal proteins from the mutant are labeled.

Dikaryon rescue has also been used to study the polarity of assembly of the flagellar microtubules. In these studies, transformants that expressed an epitope-tagged α-tubulin were mated to recipient cells that had half-length flagellae. Upon fusion, the shorter pair of flagellae grew to full length using a common pool of precursors that included epitope-tagged tubulin. Immunodetection methods revealed that the flagellar microtubules elongate by addition of new subunits at their distal end. Similar results have been obtained with the assembly of radial spoke proteins onto radial spokeless full-length flagella, suggesting that both tubulin and radial spoke proteins are transported to the distal end of the flagellum prior to their assembly into flagellar structures.

Because flagellar structure has been conserved throughout evolution, results obtained with *Chlamydomonas* are relevant for understanding human diseases. These include primary ciliary dyskinesis that affects the motility of cilia, polycystic kidney disease that involves, in some cases, a defect in the assembly of the primary cilia, and retinitis pigmentosa that causes retinal degeneration through a defect in transport of proteins through the connecting cilium of the photoreceptor cells and thereby leads to blindness. Several of the *Chlamydomonas* flagellar proteins are remarkably similar to the human proteins associated with some of these diseases.

Other mutations affecting the size of the flagella have also been examined. Some of these mutants have short flagella but otherwise normal axonemal structures. Others, called *stumpy mutants*, have very short, nonfunctional flagella. Mutants lacking flagella, called *bald mutants*, are deficient in

basal bodies. Mutants with a variable number of flagella and basal bodies have also been isolated. One of them, *vfl2*, has a single amino acid change in centrin, a 20 kDa calcium binding protein involved in the formation of fibers that show calcium-dependent contraction. The *vfl2* mutant has structural defects in the fibrous structures in which centrin is present: the nucleobasal body connectors, the distal striated fibers that link the two basal bodies together, and the flagellar transition region between the basal bodies and their flagella. This mutant is also defective in basal body localization and/or segregation.

While most of the nuclear mutations affecting flagellar function and assembly map on various chromosomes, several are linked together on a linkage group called the *uni* linkage group. The name is derived from the *uni* mutation, which affects the assembly of one basal body and, consequently, leads to the appearance of uniflagellated cells. The high degree of clustering of functionally related genes on the *uni* linkage group is unusual for eukaryotic chromosomes.

Fragments of the *uni* linkage group have been cloned by taking advantage of the restriction fragment length polymorphisms that exist between *C. reinhardtii* and *C. smithii*. Two genetically marked strains of these two species were crossed, and the progeny containing the *uni* linkage markers from the *C. smithii* parent and the other nuclear markers from the *C. reinhardtii* parent were selected and backcrossed with the *C. reinhardtii* parent. By repeating these backcrosses several times, it has been possible to construct a strain with a *uni* linkage group derived from *C. smithii* in a nuclear genetic background that derives mostly from *C. reinhardtii*. A repetitive element present in both parental strains was then used to detect RFLPs linked to the *uni* linkage group and to clone DNA fragments of the *uni* linkage group. These fragments were used as probes to determine that the *uni* chromosome is a linear DNA molecule consisting of 6 to 9 megabase pairs.

5 Perspectives

Chlamydomonas has emerged as an attractive model system for studies of the molecular and cellular biology of eukaryotic photosynthetic cells. This alga can be manipulated with relative ease at the biochemical, molecular, and genetic levels. It is, at present, the only organism in which nuclear, chloroplast, and mitochondrial transformations are feasible. The ability to introduce genetically modified genes into *C. reinhardtii* mutant strains in which the endogenous gene is defective can provide important insights into the structure–function relations of the corresponding proteins. In this respect, this alga will remain valuable for studies on flagellar function and assembly, especially for understanding the complex functional interrelationships between the various flagellar subcomplexes *in vivo*. An important point is that the use of *C. reinhardtii* can also be extended to the understanding of human diseases associated with flagellar or ciliary dysfunction.

Because the photosynthetic apparatus of *C. reinhardtii* is very similar to its homologue in higher plants, this alga is also an excellent system for studying the biogenesis and function of the photosynthetic complexes, and for investigating more generally the genetic interactions between the nuclear, chloroplast, and mitochondrial compartments. Other areas of research for which *Chlamydomonas* is

uniquely suited are phototaxis, cell wall synthesis, mating reactions and gametogenesis, and the metabolism of carbon, nitrogen, and sulfur. It is anticipated that the forthcoming sequencing of the nuclear genome of *C. reinhardtii* will further enhance the power of this unicellular alga for biological studies.

See also Genetics, Molecular Basis of; Genomic Sequencing (Core Article).

Bibliography

Books and Reviews

Boynton, J.E., Gillham, N.W., Newman, S.M., Harris, E.H. (1992) Organelle Genetics and Transformation of *Chlamydomonas*, in: Herrmann, R.G. (Ed.) *Plant Gene Research, Cell Organelles*, Springer-Verlag, Berlin, pp. 3–64.

Curry, A.M., Rosenbaum, J.L. (1993) Flagellar radial spoke: a model molecular genetic system for studying organelle assembly, *Cell Motil. Cytoskeleton* **24**, 224–232.

Dutcher, S.K. (1989) Linkage group XIX in *Chlamydomonas reinhardtii* (Chlorophyceae): Genetic Analysis of Basal Body Function and Assembly, in: Stein Taylor, J.R., Coleman, A.W., Goff, L.G. (Eds.) *Algae as Experimental Systems*, Liss Inc., New York, pp. 39–53.

Harris, E.H. (1989) *The Chlamydomonas Sourcebook*, Academic Press, San Diego, CA.

Johnson, K.A., Rosenbaum, J.L. (1991) Basal bodies and DNA trends, *Cell Biol.* **1**, 145–149.

Luck, D.J. (1984) Genetic and biochemical dissection of the eucaryotic flagellum, *J. Cell Biol.* **98**, 789–794.

Mayfield, S.P. (1990) Chloroplast gene regulation: interaction of the nuclear and chloroplast genomes in the expression of photosynthetic proteins, *Curr. Opin. Cell Biol.* **2**, 509–513.

Rochaix, J.-D. (1995) *Chlamydomonas reinhardtii* as the photosynthetic yeast, *Annu. Rev. Genet.* **29**, 209–230.

Rochaix, J.-D., Goldschmidt-Clermont, M., Merchant, S. (Eds.) (1998) The Molecular Biology of Chloroplasts and Mitochondria, Advances in Photosynthesis, Vol. 7, Kluwer academic Publisher, Dordrecht, Boston, London.

Shrager, J., Hauser, C., Chang, C.W., Harris, E.H., Davies, J., McDermott, J., Tamse, R., Zhang, Z., Grossman, A.R. (2003) *Chlamydomonas reinhardtii* genome project. A guide to the generation and use of the cDNA information, *Plant Physiol.* **131**, 401–408.

Primary Literature

Bastia, D., Chiang, K.S., Swift, H., Siersma, P. (1971) Heterogeneity, complexity, and repetition of the chloroplast DNA of Chlamydomonas reinhardtii, *Proc. Natl. Acad. Sci. USA* **68**, 1157–1161.

Bennoun, P. (1982) Evidence for a respiratory chain in the chloroplast, *Proc. Natl. Acad. Sci. USA* **79**, 4352–4356.

Bennoun, P., Levine, R.P. (1967) Detecting mutants that have impaired photosynthesis by their increased level of fluorescence, *Plant Physiol.* **42**, 1284–1287.

Boer, P.H., Gray, M.W. (1988) Scrambled ribosomal RNA gene pieces in Chlamydomonas reinhardtii mitochondrial DNA, *Cell* **55**, 399–411.

Boudreau, E., Takahashi, Y., Lemieux, C., Turmel, M., Rochaix, J.-D. (1997) The chloroplast ycf3 and ycf4 open reading frames of Chlamydomonas reinhardtii are required for the accumulation of the photosystem I complex, *EMBO J.* **16**, 6095–6104.

Boynton, J.E., Gillham, N.W., Harris, E.H., Hosler, J.P., Johnson, A.M., Jones, A.R., Randolph-Anderson, B.L., Robertson, D., Klein, T.M., Shark, K.B., Sanford, J. (1988) Chloroplast transformation in Chlamydomonas with high velocity microprojectiles, *Science* **240**, 1534–1538.

Cerutti, H., Johnson, A.M., Gillham, N.W., Boynton, J.E. (1997) Epigenetic silencing of a foreign gene in nuclear transformants of Chlamydomonas, *Plant Cell* **9**, 925–945.

Choquet, Y., Stern, D.B., Wostrikoff, K., Kuras, R., Girard-Bascou, J., Wollman, F.A. (1998) Translation of cytochrome f is autoregulated through the 5′ untranslated region of petA mRNA in Chlamydomonas

chloroplasts, *Proc. Natl. Acad. Sci. USA* **95**, 4380–4385.

Choquet, Y., Wostrikoff, K., Rimbault, B., Zito, F., Girard-Bascou, J., Drapier, D., Wollman, F.A. (2001) Assembly-controlled regulation of chloroplast gene translation, *Biochem. Soc. Trans.* **29**, 421–426.

Chua, N.H., Gillham, N.W. (1977) The sites of synthesis of the principal thylakoid membrane polypeptides in Chlamydomonas reinhardtii, *J. Cell. Biol.* **74**, 441–452.

Danon, A., Mayfield, S.P. (1991) Light regulated translational activators: identification of chloroplast gene specific mRNA binding proteins, *EMBO J.* **10**, 3993–4001.

Davies, J.P., Yildiz, F., Grossman, A.R. (1994) Mutants of Chlamydomonas with aberrant responses to sulfur deprivation, *Plant Cell* **6**, 53–63.

Day, A., Schirmer-Rahire, M., Kuchka, M.R., Mayfield, S.P., Rochaix, J.-D. (1988) A transposon with an unusual arrangement of long terminal repeats in the green alga Chlamydomonas reinhardtii, *EMBO J.* **7**, 1917–1927.

Debuchy, R., Purton, S., Rochaix, J.-D. (1989) The argininosuccinate lyase gene of Chlamydomonas reinhardtii: an important tool for nuclear transformation and for correlating the genetic and molecular maps of the ARG7 locus, *EMBO J.* **8**, 2803–2809.

Delosme, R., Olive, J., Wollman, F.A. (1996) Changes in light energy distribution upon state transitions: an *in vivo* photoacoustic study of the wild type and photosynthesis mutants from Chlamydomonas reinhardtii, *Biochim. Biophys. Acta* **1273**, 150–158.

Depege, N., Bellafiore, S., Rochaix, J.-D. (2003) Role of chloroplast protein kinase Stt7 in LHCII phosphorylation and state transition in Chlamydomonas, *Science* **299**, 1572–1575.

Durrenberger, F., Rochaix, J.-D. (1991) Chloroplast ribosomal intron of Chlamydomonas reinhardtii: *in vitro* self-splicing, DNA endonuclease activity and *in vivo* mobility, *EMBO J.* **10**, 3495–3501.

Ferris, P.J. (1989) Characterization of a Chlamydomonas transposon, Gulliver, resembling those in higher plants, *Genetics* **122**, 363–377.

Ferris, P.J., Goodenough, U.W. (1994) The mating-type locus of Chlamydomonas reinhardtii contains highly rearranged DNA sequences, *Cell* **76**, 1135–1145.

Finazzi, G., Rappaport, F., Furia, A., Fleischmann, M., Rochaix, J.-D., Zito, F., Forti, G. (2002) Involvement of state transitions in the switch between linear and cyclic electron flow in Chlamydomonas reinhardtii, *EMBO Rep.* **3**, 280–285.

Fuhrmann, M., Oertel, W., Hegemann, P. (1999) A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in Chlamydomonas reinhardtii, *Plant J.* **19**, 353–361.

Goldschmidt-Clermont, M. (1991) Transgenic expression of aminoglycoside adenine transferase in the chloroplast: a selectable marker of site-directed transformation of chlamydomonas, *Nucleic Acids Res.* **19**, 4083–4089.

Goldschmidt-Clermont, M., Choquet, Y., Girard-Bascou, J., Michel, F., Schirmer-Rahire, M., Rochaix, J.-D. (1991) A small chloroplast RNA may be required for trans-splicing in Chlamydomonas reinhardtii, *Cell* **65**, 135–143.

Goldschmidt-Clermont, M., Girard-Bascou, J., Choquet, Y., Rochaix, J.-D. (1990) Trans-splicing mutants of Chlamydomonas reinhardtii, *Mol. Gen. Genet.* **223**, 417–425.

Guergova-Kuras, M., Boudreaux, B., Joliot, A., Joliot, P., Redding, K. (2001) Evidence for two active branches for electron transfer in photosystem I, *Proc. Natl. Acad. Sci. USA* **98**, 4437–4442.

Herrin, D.L., Bao, Y., Thompson, A.J., Chen, Y.F. (1991) Self-splicing of the Chlamydomonas chloroplast psbA introns, *Plant Cell* **3**, 1095–1107.

Hill, K.L., Merchant, S. (1995) Coordinate expression of coproporphyrinogen oxidase and cytochrome c_6 in the green alga Chlamydomonas reinhardtii in response to changes in copper availability, *EMBO J.* **14**, 857–865.

Huang, C., Wang, S., Chen, L., Lemieux, C., Otis, C., Turmel, M., Liu, X.Q. (1994) The Chlamydomonas chloroplast clpP gene contains translated large insertion sequences and is essential for growth, *Mol. Gen. Genet.* **244**, 151–159.

Kathir, P., LaVoie, M., Brazelton, W.J., Haas, N.A., Lefebvre, P.A., Silflow, C.D. (2003) Molecular map of the Chlamydomonas reinhardtii nuclear genome, *Eukaryot. Cell* **2**, 362–379.

Kindle, K.L. (1990) High-frequency nuclear transformation of Chlamydomonas reinhardtii, *Proc. Natl. Acad. Sci. USA* **87**, 1228–1232.

Kropat, J., Oster, U., Rudiger, W., Beck, C.F. (1997) Chlorophyll precursors are signals of chloroplast origin involved in light induction of nuclear heat-shock genes, *Proc. Natl. Acad. Sci. USA* **94**, 14168–14172.

Kruse, O., Nixon, P.J., Schmid, G.H., Mullineaux, C.W. (1999) Isolation of state transition mutants of Chlamydomonas reinhardtii by fluorescence video imaging, *Photosynth. Res.* **61**, 43–51.

Kuroiwa, T., Kawano, S., Nishibayashi, S., Sato, C. (1982) Epifluorescent microscopic evidence for maternal inheritance of chloroplast DNA, *Nature* **298**, 481–483.

Levine, R.P. (1960) Genetic control of photosynthesis in Chlamydomonas reinhardtii, *Proc. Natl. Acad. Sci. USA* **46**, 972–977.

Luck, D., Piperno, G., Ramanis, Z., Huang, B. (1977) Flagellar mutants of Chlamydomonas: studies of radial spoke-defective strains by dikaryon and revertant analysis, *Proc. Natl. Acad. Sci. USA* **74**, 3456–3460.

Marshall, W.F., Rosenbaum, J.L. (2001) Intraflagellar transport balances continuous turnover of outer doublet microtubules: implications for flagellar length control, *J. Cell. Biol.* **155**, 405–414.

Matagne, R.F., Remacle, C., Dinant, M. (1991) Cytoduction in Chlamydomonas reinhardtii, *Proc. Natl. Acad. Sci. USA* **88**, 7447–7450.

Maul, J.E., Lilly, J.W., Cui, L., dePamphilis, C.W., Miller, W., Harris, E.H., Stern, D.B. (2002) The Chlamydomonas reinhardtii plastid chromosome: islands of genes in a sea of repeats, *Plant Cell* **14**, 2659–2679.

Merchant, S., Bogorad, L. (1986) Regulation by copper of the expression of plastocyanin and cytochrome c552 in Chlamydomonas reinhardtii, *Mol. Cell. Biol.* **6**, 462–469.

Moseley, J., Quinn, J., Eriksson, M., Merchant, S. (2000) The Crd1 gene encodes a putative di-iron enzyme required for photosystem I accumulation in copper deficiency and hypoxia in Chlamydomonas reinhardtii, *EMBO J.* **19**, 2139–2151.

Nishimura, Y., Misumi, O., Kato, K., Inada, N., Higashiyama, T., Momoyama, Y., Kuroiwa, T. (2002) An mt(+) gamete-specific nuclease that targets mt(−) chloroplasts during sexual reproduction in C. reinhardtii, *Genes Dev.* **16**, 1116–1128.

Niyogi, K.K., Bjorkman, O.B., Grossman, A.R. (1997) Chlamydomonas Xanthophyll cycle mutants identified by video imaging of chlorophyll fluorescence quenching, *Plant Cell* **9**, 1369–1380.

Pazour, G.J., Baker, S.A., Deane, J.A., Cole, D.G., Dickert, B.L., Rosenbaum, J.L., Witman, G.B., Besharse, J.C. (2002) The intraflagellar transport protein, IFT88, is essential for vertebrate photoreceptor assembly and maintenance, *J. Cell. Biol.* **157**, 103–113.

Perron, K., Goldschmidt-Clermont, M., Rochaix, J.-D. (1999) A factor related to pseudouridine synthases is required for chloroplast group II intron trans-splicing in Chlamydomonas reinhardtii, *EMBO J.* **18**, 6481–6490.

Petracek, M.E., Lefebvre, P.A., Silflow, C.D., Berman, J. (1990) Chlamydomonas telomere sequences are A+T-rich but contain three consecutive G-C base pairs, *Proc. Natl. Acad. Sci. USA* **87**, 8222–8226.

Preble, A.M., Giddings Jr., T.H., Dutcher, S.K. (2001) Extragenic bypass suppressors of mutations in the essential gene BLD2 promote assembly of basal bodies with abnormal microtubules in Chlamydomonas reinhardtii, *Genetics* **157**, 163–181.

Purton, S., Rochaix, J.-D. (1994) Complementation of a Chlamydomonas reinhardtii mutant using a genomic cosmid library, *Plant Mol. Biol.* **24**, 533–537.

Randolph-Anderson, B.L., Boynton, J.E., Gillham, N.W., Harris, E.H., Johnson, A.M., Dorthu, M.P., Matagne, R.F. (1993) Further characterization of the respiratory deficient dum-1 mutation of Chlamydomonas reinhardtii and its use as a recipient for mitochondrial transformation, *Mol. Gen. Genet.* **236**, 235–244.

Rivier, C., Goldschmidt-Clermont, M., Rochaix, J.-D. (2001) Identification of an RNA-protein complex involved in chloroplast group II intron trans-splicing in Chlamydomonas reinhardtii, *EMBO J.* **20**, 1765–1773.

Sager, R. (1954) Mendelian and non-Mendelian inheritance of streptomycin resistance in Chlamydomonas reinhardtii, *Proc. Natl. Acad. Sci. USA* **40**, 356–363.

Salvador, M.L., Klein, U., Bogorad, L. (1993) 5′ sequences are important positive and negative determinants of the longevity of

Chlamydomonas chloroplast gene transcripts, *Proc. Natl. Acad. Sci. USA* **90**, 1556–1560.

Takahashi, Y., Goldschmidt-Clermont, M., Soen, S.Y., Franzen, L.G., Rochaix, J.-D. (1991) Directed chloroplast transformation in Chlamydomonas reinhardtii: insertional inactivation of the psaC gene encoding the iron sulfur protein destabilizes photosystem I, *Embo J* **10**, 2033–2040.

Tam, L.-W., Lefebvre, P.A. (1993) Cloning of flagellar genes in Chlamydomonas reinhardtii by DNA insertional mutagenesis, *Genetics* **135**, 375–384.

Thompson, R.J., Mosig, G. (1987) Stimulation of a Chlamydomonas chloroplast promoter by novobiocin in situ and in E. coli implies regulation by torsional stress in the chloroplast DNA, *Cell* **48**, 281–287.

Umen, J.G., Goodenough, U.W. (2001) Chloroplast DNA methylation and inheritance in Chlamydomonas, *Genes Dev.* **15**, 2585–2597.

Vallon, O., Bulte, L., Dainese, P., Olive, J., Bassi, R., Wollman, F.A. (1991) Lateral redistribution of cytochrome b6/f complexes along thylakoid membranes upon state transitions, *Proc. Natl. Acad. Sci. USA* **88**, 8262–8266.

Yu, W., Spreitzer, R.J. (1992) Chloroplast heteroplasmicity is stabilized by an amber-suppressor tryptophan tRNA(CUA), *Proc. Natl. Acad. Sci. USA* **89**, 3904–3907.

Zerges, W., Girard-Bascou, J., Rochaix, J.-D. (1997) Translation of the chloroplast psbC mRNA is controlled by interactions between its 5′ leader and the nuclear loci TBC1 and TBC3 in Chlamydomonas reinhardtii, *Mol. Cell. Biol.* **17**, 3440–3448.

Zhang, H., Herman, P.L., Weeks, D.P. (1994) Gene isolation through genomic complementation using an indexed library of Chlamydomonas reinhardtii DNA, *Plant Mol. Biol.* **24**, 663–672.

Zhao, H., Lu, M., Singh, R., Snell, W.J. (2001) Ectopic expression of a Chlamydomonas mt+− specific homeodomain protein in mt- gametes initiates zygote development without gamete fusion. *Genes Dev.* **15**, 2767–2777.

Zito, F., Finazzi, G., Delosme, R., Nitschke, W., Picot, D., Wollman, F.A. (1999) The Qo site of cytochrome b_6f complexes controls the activation of the LHCII kinase, *EMBO J.* **18**, 2961–2969.

Chromosome Organization within the Nucleus

Wallace F. Marshall
Dept. of Molecular, Cellular, and Developmental Biology Yale University, New Haven, CT 06520

1 Introduction 3

2 Chromosome Positioning 3
2.1 Analysis of Polytene Chromosomes 3
2.2 Fluorescence *In Situ* Hybridization (FISH) 4
2.3 Interactions with the Nuclear Envelope 4
2.4 Retention of Mitotic Chromosome Geometry 5

3 Molecular Interactions That Specify Nuclear Architecture 6
3.1 Nuclear Lamins 6
3.2 Nuclear Pores 6
3.3 Nuclear Matrix 6

4 Interphase Chromatin Movement 7
4.1 Evidence for Chromatin Immobilization 7
4.2 Evidence for Chromatin Diffusion 7
4.3 Evidence for Active Chromatin Movement 8

5 Regulation of Gene Expression by Nuclear Architecture 8
5.1 Role of Nuclear Envelope Interactions in Gene Silencing 8
5.2 Gene Gating 8
5.3 Nuclear Compartmentalization and Gene Expression 9

6 Role of Nuclear Architecture in Chromosome Interactions 9
6.1 Regulation of Interactions by Constrained Diffusion 9
6.2 Robertsonian Translocations 10
6.3 Chromosome Rearrangements in Cancer 10

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

7 **Conclusions** 11

Acknowledgments 11

Bibliography 11
Books and Reviews 11
Primary Literature 11

Keywords

Constrained Diffusion
Type of Brownian motion in which an object undergoes random, thermally driven movements, just as in free diffusion, except that the object is prevented from diffusing outside a limited region of space. Chromatin motion during interphase appears to be dominated by this type of motion.

FISH (Fluorescence In Situ Hybridization)
A method of visualizing the position of individual chromosome loci during interphase by hybridization with an exogenous DNA probe that is either itself fluorescently labeled, or that contains an epitope that can be detected with fluorescent probes.

Heterochromatin
Regions of chromatin containing mostly simple-sequence repeated DNA that is highly compacted during interphase. Most heterochromatin is transcriptionally inactive.

Nuclear Envelope
The double bilayer that surrounds the nucleus. The ER derived from endoplasmic reticulum, is organized by a protein-based network containing nuclear lamins, and is penetrated by pores called *nuclear pore complexes*, which allow import and export of proteins and RNA molecules.

Nuclear Matrix
A putative internal protein-based network that criss-crosses the nuclear lumen and may help organize chromatin.

Rabl Configuration
The polarized organization of chromosomes in the nucleus, with centromeres clustered at one end of the nucleus and telomeres at the other. The Rabl configuration is a remnant of the anaphase chromosome configuration.

The interphase nucleus provides the structural context for chromosome biology, including gene expression and recombination. The organization of chromosomes within the nucleus is not random. Instead, chromosomes are organized by specific constraints, including interactions of specific loci with the nuclear envelope. This

interaction may involve both nuclear lamins and nuclear pore complexes. Specific nuclear envelope attachments, together with a persisting remnant of the anaphase chromosome configuration, leads to chromosomes being organized into nonoverlapping territories with specific orientations. Superimposed on this organization is a high degree of chromosome mobility driven by diffusion. Because chromosomes are constrained by nuclear envelope attachments, their diffusion is constrained, with each locus able to explore only a limited subregion of the nucleus. This constrained mobility, together with the nonrandom positional organization of chromosomes, predicts a high degree of nonrandomness in the pattern of interchromosomal interactions.

1 Introduction

The genome is packaged within the nucleus on many levels. Not only is the DNA wound around nucleosomes and compacted into chromatin, the chromosomes themselves are arranged within the nucleus in a defined pattern. This nuclear organization is set up and maintained by interactions between chromosomes and the nuclear envelope (NE) and possibly an internal nuclear matrix. However, chromosomes are not absolutely fixed, but are able to undergo a certain degree of constrained motion, which allows interactions between chromosomes to take place. This combination of specific positioning and constrained movement leads to a dynamic view of nuclear organization in which large-scale chromosome organization can play a decisive role in regulating gene expression and chromosome interactions.

2 Chromosome Positioning

2.1 Analysis of Polytene Chromosomes

Determining the arrangement of interphase chromosomes is a challenging problem, because during interphase the chromosomes decondense and thus become difficult to visualize. Early dramatic evidence for nonrandom chromosome arrangement came from analysis of *Drosophila* polytene chromosomes in three dimensions. Polytene chromosomes are large and compact and easy to image inside living cells. Moreover, the well-known banding pattern of polytene chromosomes allowed the position of each locus, defined cytologically by banding pattern, to be determined in three-dimensional space, thus projecting the one-dimensional genetic map onto a three-dimensional map of the nucleus. These studies were among the first studies ever to use three-dimensional fluorescence microscopy. On the basis of this type of analysis, it was found that the arrangement of chromosomes was different in different nuclei, indicating that the position of a locus was not rigidly predetermined. But although the arrangements were variable, they were not entirely random either. For one thing, the chromosomes followed a polarized orientation in the nucleus, with centromeres at one end of the nucleus, and telomeres at the other. Both the polarized arrangement of chromosomes and their separation into distinct domains probably reflects a persistence of

their arrangement in mitosis, and suggests that the chromatin did not undergo extensive rearrangement within the nucleus following decondensation in telophase. In addition, the chromosomes occupied separate domains and were not intertwined. Finally, chromosomes contacted the NE at particular loci. These studies in polytene chromosomes showed chromosomes were organized according to two fundamental organizational principles: specific contacts with the NE and persistence of the mitotic chromosome arrangement.

2.2 Fluorescence *In Situ* Hybridization (FISH)

Following the initial work in polytene chromosomes, great interest was created in testing whether the same conclusions would hold in ordinary nonpolytene chromosomes during interphase. Measuring interphase chromosome arrangements in normal cells required FISH in which DNA probes specific for a given locus are hybridized onto intact nuclei and detected using fluorescent probes. One potential danger in using FISH is that the harsh conditions needed for hybridization can often disrupt the organization of the nucleus. This is particularly true of methods that involve air-drying or acid treatments during fixation, and it is always worthwhile, in assessing the value of a given FISH analysis, to examine the protocol in detail. It is likely that some of the published studies that have failed to observe reproducible nuclear organization failed because of overly harsh sample preparation that destroys the three-dimensional structure. Certainly, any procedure in which nuclei are flattened or spread will, by definition, destroy the three-dimensional structure. Fortunately, the study of nuclear organization by FISH has been greatly facilitated by the development of fixation and hybridization procedures that preserve the three-dimensional structure of the nucleus. Many such studies, summarized below, have shown that the two main results of the polytene studies, namely, NE interactions and persistence of mitotic arrangement, also hold true in other cell types.

2.3 Interactions with the Nuclear Envelope

Many studies have used FISH to show specific chromosome regions are localized to the NE, which probably accounts for the earlier observations using electron microscopy that revealed multiple sites of chromatin–NE contact in the nucleus. In most cases, we do not yet know what DNA sequences determine this interaction, although several studies have now suggested that silenced chromatin may confer NE attachment (see below).

A priori, if a given locus is found near the NE, this could either reflect an interaction of the chromatin with the NE, or alternatively it might simply reflect a nonrandom radial positioning within the nucleus due to either the intrinsic folding of the chromosomes or perhaps due to exclusion from the nuclear interior of loci that cannot interact with some internal structure. However, all available evidence suggests that the localization of chromatin near the NE involves an actual physical interaction between chromatin and some component of the NE. Polytene chromosomes have been clearly shown to be physically stuck onto the NE at discrete sites. This conclusion is also supported by studies of interphase chromatin movement *in vivo* in which loci near the NE tend to move significantly less than more internal loci.

It is likely that loci targeted to the NE will find themselves in a highly specific subenvironment within the nucleus, which might have properties different from those in the nuclear interior. Certainly, such loci would have the opportunity to interact with many proteins of the NE that would be effectively invisible to loci that were found in the nuclear interior. Direct interactions with NE proteins are indeed thought to play a role in gene silencing, as discussed below. However, in addition to a direct effect on the NE-targeted loci, the attachment of chromosomes to the NE forces the rest of the chromosome into a nonrandom radial positioning, in which even loci not actually attached to the NE will tend to occupy specific radial distances relative to the surface. This nonrandom radial positioning has been systematically explored in *Drosophila*, but has also been seen in mammalian cells. This nonrandom radial position is likely to have significant effects on chromosome interactions (see Sect. 6 below).

2.4 Retention of Mitotic Chromosome Geometry

During anaphase, chromosomes become aligned such that all centromeres cluster together as they move to the pole, while the telomeres trail behind. When the chromatin decondenses in telophase, this arrangement should result in a polarized nucleus in which telomeres are at one side of the nucleus and centromeres at another. This is generally referred to as the *Rabl Configuration* after its discoverer, Carl Rabl. This Rabl configuration has been clearly demonstrated using FISH in a wide range of cell types. The Rabl orientation affects not just the position of centromeres and telomeres, but of all the chromatin in between. This has been most clearly shown in *Drosophila* embryos, in which loci spanning a chromosome arm were shown to occupy a precise position along the nuclear axis, in an order that follows precisely their linear order along the chromosome. The Rabl arrangement is most dramatic in cells that are dividing actively, and tends to become less obvious in cells that have arrested in the cell cycle. This is likely to reflect the slow diffusive motion of chromatin during interphase, such that the longer a cell remains arrested following division, the more the mitotic chromosome arrangement becomes distorted.

In addition to aligning during mitosis, chromosomes also become condensed into spatially separated objects. When the chromatin decondenses in telophase, this spatial separation tends to force the chromosomes into nonoverlapping regions within the nucleus. Indeed, a variation on FISH called chromosome painting, in which entire chromosomes are visualized using a pool of DNA probes, has shown that interphase chromosomes generally remain separated into nonoverlapping "territories."

The combined effect of specific NE interactions and the retention of mitotic arrangement is to set up a highly nonrandom arrangement of chromosomes within the nucleus, such that different loci tend to occupy specific subregions of the nucleus. Because this positioning is a by-product of large-scale chromosome arrangements, there tends to be a lot of variability in the position of individual loci, such that the nonrandomness of nuclear architecture is best viewed as a statistical trend, rather than as an absolute predetermined arrangement.

3
Molecular Interactions That Specify Nuclear Architecture

Nuclear architecture is maintained by interactions with the NE and possibly an internal nuclear matrix. Many studies over the past few decades have revealed a number of candidate molecules that might account for these interactions. Such studies will really be the key to understanding the function of nuclear architecture, since in principle, the identification of the molecules involved should allow the architecture to be perturbed in a reproducible way and the effects of this perturbation tested.

3.1
Nuclear Lamins

What part of the NE is responsible for binding chromatin? The nuclear lamins, a family of intermediate filament-like proteins that form a network on the inner surface of the NE, are able to bind specific DNA sequences *in vitro* as well as histones. Several lamin-associated proteins can bind DNA or chromatin. It remains unclear, however, whether these interactions really occur in living cells. Moreover, since lamins are thought to be involved in reassembling the nucleus around chromosomes during telophase, it is important to distinguish between chromatin–lamin interactions that are involved solely in nuclear reassembly. A FISH analysis of nuclear organization during nuclear reassembly in *Drosophila* embryos has shown that the loci that interact with the NE during interphase are not the same as the loci that interact with the NE during nuclear reassembly.

3.2
Nuclear Pores

Nuclear pore complexes (NPCs) also appear capable of interacting with chromatin. For example, one NPC protein binds DNA *in vitro*. Moreover, the telomere protein yKu70 can bind to the NPC associated protein Mlp2, which is attached to the NPC via an interaction with the nucleoporin Nup145. All three of these proteins are needed for telomeres to interact with the NE. In electron microscopy studies of the NE, the nuclear pores are seen to extend long processes into the nuclear interior, and these may be likely candidates for the specific structures that interact with chromatin. Because both the nuclear lamins and the NPCs can associate with chromatin, it is critical to distinguish which set of molecules is required for a given chromatin–NE interaction.

3.3
Nuclear Matrix

In addition to interactions with the NE, chromatin might also interact with an internal structure within the nucleus called the *nuclear matrix*. The nuclear matrix was first defined as the insoluble residue that remains after isolated nuclei are heavily extracted, and so its relevance to actual structures *in vivo* has never been very clear. A network of fibers has been seen in nuclei that might correspond to the matrix, but only when using unusual resinless embedding procedures for electron microscopy, which raise concerns about potential artifacts. Two proteins in *Drosophila*, CP60 and CP190, appear to form a network within nuclei of living *Drosophila* embryos, and this network persists in the shape of the nucleus even after the NE has broken down and the chromosomes have congressed to

the metaphase plate. CP60 and CP190 are components of the nuclear matrix defined biochemically, and thus may represent a true matrix that exists as a coherent structure in living cells. This possibility deserves more detailed investigation.

4 Interphase Chromatin Movement

The only way for chromosomes to be held in a defined arrangement is for their mobility to be constricted. Otherwise, one would expect the chromatin to diffuse throughout the nucleus, like polymers in a polymer melt, and rapidly randomize any initially nonrandom configuration.

4.1 Evidence for Chromatin Immobilization

Early cytological evidence suggested that interphase chromosomes were fairly immobile as judged by the similarity of chromosome arrangements in successive mitoses by the separation of parental genomes in the first few divisions after fertilization. Modern biophysical methods, including photobleaching and *in vivo* chromosome visualization, have confirmed the early impression that chromosome movement is constricted during interphase.

4.2 Evidence for Chromatin Diffusion

However, more detailed measurement of interphase chromatin movement in living cells, based on an application of submicron single-particle tracking techniques to individual loci tracked using a green fluorescent protein (GFP) tagging system, revealed that interphase chromatin does in fact undergo substantial and rapid diffusion, but this diffusion is constrained such that a given locus can only diffuse within a small subregion of the nucleus. In yeast and *Drosophila*, chromatin moved with a diffusion constant of approximately 10^{-11} cm^2 s^{-1} because of thermally driven Brownian motion. But this diffusion was confined to a subregion with a radius of 0.3 μm for yeast and 0.9 μm for *Drosophila*, roughly equal to 1to 5% of the nuclear volume. Following this first analysis of chromatin diffusion in yeast and *Drosophila*, subsequent analyses have confirmed the result that interphase chromatin can diffuse within a confined region in a wide range of cell types. More detailed analysis has shown that the details of chromatin diffusion depend on the timescale, and that at very short timescales chromatin movement is highly constrained, whereas at longer timescales, chromatin is freer to diffuse over larger distances. This suggests that chromatin diffusion at short timescales is constrained by interaction with a structure (e.g. the nuclear lamina) that itself undergoes slow random movements. Nuclear pores have been shown to turn over at a slow but detectable rate such that if chromatin were constrained by interactions with NPCs, then the slow rearrangement of the nuclear pores could account for the relaxation of chromatin constraint on long timescales. The idea that chromatin diffusion is constrained via interactions with the NE is supported by two recent studies that found that diffusion of chromatin loci associated with the NE is significantly more constrained than other loci. In the most dramatic example, Chubb and coworkers showed that NE-associated chromatin has exactly the same diffusion constant as other chromatin, but is confined to a much smaller subregion of the nucleus.

4.3
Evidence for Active Chromatin Movement

In addition to this constrained diffusion, it also appears that loci can undergo active directed movements, at least in some cases. When centromeres in animal cells were tracked *in vivo*, it was found that the centromeres occasionally experienced persistent directed movements in the interphase nucleus. Another study showed large directed shifts in gene position during the mid-blastula transition in *Drosophila* embryos, possibly driven by the force of heterochromatin condensation.

5
Regulation of Gene Expression by Nuclear Architecture

5.1
Role of Nuclear Envelope Interactions in Gene Silencing

Heterochromatin, the highly condensed transcriptionally silenced regions of the genome, is generally seen near the NE in electron microscopy studies. Indeed, insertion of a block of heterochromatin into a euchromatic position is sufficient to locate the whole region onto the NE. More specific FISH studies of individual chromosomes have found that gene-poor chromosomes with reduced overall transcription tend to be NE associated. These results raise the possibility that the NE might play some role in establishment of silenced heterochromatin. Indeed, targeting a reporter gene to the NE by means of an NE-interacting DNA-binding fusion protein results in silencing of the gene, and induced transcriptional activation of a reporter locus caused the locus to dissociate from the NE and move to the nuclear interior. Mutations in the telomere-binding protein dimer yKu70/yKu80 and the NPC associated proteins Mlp1/Mlp2 cause telomeres to dissociate from the NE and result in loss of telomeric silencing, again suggesting that NE interactions play some role in gene silencing. The SIR3 and SIR4 silencing proteins in yeast are associated with the NE, which suggests a simple model that genes associated with the NE experience a higher local concentration of silencing proteins.

Another potential example of the NE playing a role in gene repression comes from studies of the lamin-associated protein 2 beta (LAP2β), an integral membrane protein component of the nuclear lamina. Ectopic expression of LAP2β and its binding partner GCL (germ cell-less) in cells that do not normally express GCL, reduces transcriptional induction by the E2F-DP complex, which binds GCL. This raises the possibility that Lap2β and GCL could recruit E2F-regulated genes to the NE where they would become silenced.

5.2
Gene Gating

Although most studies suggest that the NE exerts a negative effect on gene expression, a possible positive correlation between gene expression and NE localization was suggested by the increased nuclease sensitivity of NE-associated chromatin. Since increased nuclease accessibility usually correlates with gene expression, these studies suggested that NE-associated chromatin might be enriched for active genes. Since mRNA must be exported through NPCs, it was proposed that perhaps active genes were targeted to the NE in order to facilitate transcript export, a hypothesis known as *gene gating*. Gene gating is particularly appealing in the case of

genes whose transcripts are nonrandomly localized relative to the nucleus: if the corresponding gene was nonrandomly localized to the NE near the site of future message accumulation, then locally increased export through the nearby nuclear pores could explain the ultimate asymmetric transcript localization. The fact that some transcript localization is perturbed in a nuclear lamin mutant might tend to support this model.

Although these studies raised the possibility that active genes might be targeted to the NE, in most cases it has been found that active genes are, in fact, nonrandomly localized to the nuclear interior and not to the NE. As for a role of the NE in transcript localization, a careful study of many pair-rule genes in *Drosophila*, whose transcripts are specifically localized relative to the nucleus, revealed no relation between the position of the gene in the nucleus and the position of transcript accumulation. These studies suggest that gene gating probably does not play any significant role in transcript targeting or export.

5.3 Nuclear Compartmentalization and Gene Expression

Heterochromatin tends to be nonuniformly distributed in cells. In most cell types, the heterochromatin aggregates into a small number of large domains, and when euchromatic genes become silenced by heterochromatic insertions (a phenomenon called *position effect variegation* in *Drosophila*), the genes become localized to these large heterochromatin domains. This relocalization probably plays a causal role: when a euchromatic reporter gene was artificially targeted to the heterochromatin domain in *Drosophila* by flanking chromosome pairing sites, the gene became silenced. Moreover, chromosome rearrangements that impede localization of a gene to the heterochromatin domain have the effect of reducing silencing. Thus, a probable model is that the heterochromatin domain creates a "bad neighborhood" for transcription of euchromatic genes, and that when genes are brought into this neighborhood they tend to become silenced.

This model may also explain the ability of certain sequences (known as *insulators*) to protect nearby euchromatic genes from silencing by heterochromatin. The gypsy insulator protects genes from silencing by nearby heterochromatin, and its function requires the Su(hw) and mod(mdg4) proteins. These two proteins localize to the NE, as does the gypsy insulator itself. Su(hw) mutants that prevent insulator function cause the insulator sequence to dissociate from the NE. These results imply that Su(hw) protein tethers insulator sequences to the NE, and more importantly, that possibly this tethering could be critical in insulator function. One way this would work would be to lock genes in place near the NE, preventing them from relocating from their initial position into the silenced heterochromatin compartment.

6 Role of Nuclear Architecture in Chromosome Interactions

6.1 Regulation of Interactions by Constrained Diffusion

Many aspects of DNA metabolism involve interactions between widely separated loci.

For example, recombination frequently involves an interaction between sites on two entirely different chromosomes. Another example is transvection, in which a promoter on one chromosome can drive transcription of a coding region on another chromosome. These interactions do not occur at a distance, but instead require the interacting sites to move into close physical proximity. If the organization of genes in the nucleus is nonrandom, then we can predict that the ability of genes to interact with each other is likewise nonrandom. It is verified experimentally that interactions between different loci during recombination and DNA break repair is highly nonrandom. As an obvious example, corresponding loci on homologous chromosomes will be expected to occupy similar regions of the nucleus. Similarly, nucleolar loci are more likely to interact because they are nonrandomly close together from the outset.

The strong influence that nuclear architecture can have on chromosome interactions can be best seen during the establishment of somatic homolog pairing, which is best known in *Drosophila* but also occurs sporadically in humans and yeast. An analysis of initiation of somatic pairing in *Drosophila* showed that the loci that paired first tended to start out close to each other within the nucleus, while loci that started out far apart tended to pair later. In fact, the pairing could be modeled precisely assuming that chromatin motion was driven by constrained diffusion and taking into account the known nuclear architecture in the *Drosophila* embryo. Thus, as suggested by the studies of Robertsonian translocations, it appears that position within the nucleus plays an important determinative role in interactions between chromosomes.

We can therefore propose a general model: because chromatin diffusion is constrained to a small subregion of the nucleus, only loci whose regions of confinement overlap will be able to interact. Loci that start out far apart will be prevented from interacting, since their regions of confinement do not overlap, so that physical contact is impossible. Thus, the relative positions between loci, as dictated by the nuclear architecture, can play a decisive role in determining which interactions can occur.

6.2 Robertsonian Translocations

Robertsonian translocations, spontaneous reciprocal translocations of whole chromosome arms, provide an illustrative example of the influence that nuclear organization might have on chromosome interactions. The distribution of breakpoints in such Robertsonian translocations is highly nonrandom, such that translocations in which both breakpoints occur on acrocentric chromosomes containing nucleolar organizing regions (NORs) are highly favored. Since cells generally have just a single nucleolus, chromosomes with an NOR will tend to be nonrandomly close in the nucleus, thus biasing the distribution of recombination breakpoints.

6.3 Chromosome Rearrangements in Cancer

Perhaps the most important instance of nuclear architecture influencing chromosome rearrangements occurs in chromosome translocations that lead to leukemia. A translocation between the BCR and ABL loci on chromosomes 22 and 9 gives rise

to a translocation called the Philadelphia chromosome, and the breakpoint encodes a novel fusion protein called BCR-ABL. The majority of patients with chronic myelogenous leukemia (CML) contain this rearrangement in their genomes. Why is this rearrangement so common? FISH measurements show that BCR and ABL loci are in fact nonrandomly close together in the nucleus, which may account for the high frequency of recombination between these two loci. Another rearrangement important for cancer involves the RET and H4 loci. An inversion with breakpoints in RET and H4 is found in many cases of radiation-induced thyroid cancer. As with BCR and ABL, RET and H4 are nonrandomly close together in normal cells, again suggesting that it is the spatial proximity of these loci that allows them to recombine so frequently.

7 Conclusions

Clearly, a more detailed understanding of nuclear architecture is critical for the progress of molecular medicine. Understanding how the specific positioning of genes occurs, and what influence this positioning has on gene expression and recombination, will lead to a clearer understanding of chromosome behavior in interphase. The key is to uncover the molecular players that establish and maintain nuclear architecture.

Acknowledgments

Wallace F. Marshall was funded by a Leukemia and Lymphoma Society Special Fellowship.

See also Cell Nucleus Biogenesis, Structure and Function; Electron Microscopy in Cell Biology.

Bibliography

Books and Reviews

Comings, D.E. (1980) Arrangement of chromatin in the nucleus, *Hum. Genet.* **53**, 131–143.

Cremer, T., Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells, *Nat. Rev. Genet.* **2**, 292–301.

Fung, J.C., Sedat, J.W., Marshall, W.F. (2000) Homologous DNA Interactions: Organization of the Interphase Nucleus, *The Encyclopedia of Life Sciences*, Nature Publishing Group, London, pp. 171–175

Marshall, W.F. (2002) Order and disorder in the nucleus, *Curr. Biol.* **12**, R185–R192.

Primary Literature

Abney, J.R., Cutler, B., Fillbach, M.L., Axelrod, D., Scalettar, B.A. Chromatin dynamics in interphase nuclei and its implications for nuclear structure, *J. Cell Biol.* **137**, 1459–1468.

Andrulis, E.D., Neiman, A.M., Zappulla, D.C., Sternglanz, R. (1998) Perinuclear localization of chromatin facilitates transcriptional silencing, *Nature* **394**, 592–595.

Baricheva, E.A., Berrios, M., Bogachev, S.S., Borisevich, I.V., Lapik, E.R., Sharakhov, I.V., Stuurman, N., Fisher, P.A. (1996) DNA from Drosophila melanogaster β-heterochromatin binds specifically to nuclear lamins in vitro and the nuclear envelope in situ, *Gene* **171**, 171–176.

Berezney, R., Coffey, D.S. (1974) Identification of a nuclear protein matrix, *Biochem. Biophys. Res. Commun.* **60**, 1410–1417.

Blobel, G. (1985) Gene gating: a hypothesis, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 8527–8529.

Boveri, T. (1909) Die blastomerenkerne von Ascaris megalocephala und die Theorie der Chromosomenindividualität, *Arch. Zellforschung* **3**, 181–268.

Boyle, S., Gilchrist, S., Bridger, J.M., Mahy, N.L., Ellis, J.A., Bickmore, W.A. (2001) The spatial

organization of human chromosomes within the nuclei of normal and emerin-mutant cells, *Hum. Mol. Genet.* **10**, 211–219.

Brown, K.E., Guest, S.S., Smale, S.T., Hahm, K., Merkenschlager, M., Fisher, A.G. (1997) Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin, *Cell* **91**, 845–854.

Burgess, S.M., Kleckner, N., Weiner, B.M. (1999) Somatic pairing of homologs in budding yeast: existence and modulation, *Gene Dev.* **13**, 1627–1641.

Capco, D.G., Wan, K.M., Penman, S. (1982) The nuclear matrix: three-dimensional architecture and protein composition, *Cell* **29**, 847–858.

Carter, K.C., Bowman, D., Carrington, W., Fogarty, K., McNeil, J.A., Fay, F.S., Lawrence, J.B. (1993) A three-dimensional view of precursor messenger RNA metabolism within the mammalian nucleus, *Science* **259**, 1330–1335.

Carvalho, C., Pereira, H.M., Ferreira, J., Pina, C., Mendonca, D., Rosa, A.C., Carmo-Fonseca, M. (2001) Chromosomal G-dark bands determine the spatial organization of centromeric heterochromatin in the nucleus, *Mol. Biol. Cell* **12**, 3563–3572.

Chubb, J.R., Boyle, S., Perry, P., Bickmore, W.A. (2002) Chromatin motion is constrained by association with nuclear compartments in human cells, *Curr. Biol.* **12**, 439–445.

Chung, H.M., Shea, C., Fields, S., Taub, R.N., Van de Ploeg, L.H., Tse, D.B. (1990) Architectural organization in the interphase nucleus of the protozoan Typanosoma brucei: location of telomeres and mini-chromosomes, *EMBO J.* **9**, 2611–2619.

Conklin, E.G. (1901) The individuality of the germ nuclei during the cleavage of the egg of crepidula, *Biol. Bull.* **2**, 257–265.

Cremer, T., Cremer, C., Baumann, H., Lüdtke, E.K., Sperling, K., Teuber, V., Zorn, C. (1982) Rabl's model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments, *Hum. Genet.* **60**, 46–56.

Cremer, M., von Hase, J., Volm, T., Brero, A., Kreth, G., Walter, J., Fischer, C., Solovei, I., Cremer, C., Cremer, T. (2001) Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells, *Chromosome Res.* **9**, 541–567.

Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P., Bickmore, W.A. (1999) Differences in the localization and morphology of chromosomes in the human nucleus, *J. Cell Biol.* **145**, 1119–1131.

Csink, A.K., Henikoff, S. (1996) Genetic modification of heterochromatic association and nuclear organization in Drosophila, *Nature* **381**, 529–531.

Daigle, N., Beaudouin, J., Hartnell, L., Imreh, G., Hallberg, E., Lippincott-Schwartz, J., Ellenberg, J. (2001) Nuclear pore complexes form immobile networks and have a very low turnover in live mammalian cells, *J. Cell Biol.* **154**, 71–84.

Dernburg, A.F., Broman, K.W., Fung, J.C., Marshall, W.F., Philips, J., Agard, D.A., Sedat, J.W. (1996) Perturbation of nuclear architecture by long-distance chromosome interactions, *Cell* **85**, 745–759.

Duband-Goulet, I., Courvalin, J.C. (2000) Inner nuclear membrane protein LBR preferentially interacts with DNA secondary structures and nucleosomal linker, *Biochemistry* **39**, 6483–6488.

Edelmann, P., Bornfleth, H., Zink, D., Cremer, T., Cremer, C. (2001) Morphology and dynamics of chromosome territories in living cells, *Biochim. Biophys. Acta* **1551**, M29–M40.

Francastel, C., Walters, M.C., Groudine, M., Martin, D.I. (1999) A functional enhancer suppresses silencing of a transgene and prevents its localization close to centromeric heterochromatin, *Cell* **99**, 259–269.

Funabiki, H., Hagan, I., Uzawa, S., Yanagida, M. (1993) Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast, *J. Cell Biol.* **121**, 961–976.

Fung, J.C., Marshall, W.F., Dernburg, A., Agard, D.A., Sedat, J.W. (1998) Homologous chromosome pairing in Drosophila melanogaster proceeds through multiple independent initiations, *J. Cell Biol.* **141**, 5–20.

Furukawa, K. (1999) LAP2 binding protein 1 (L2BP1/BAF) is a candidate mediator of LAP2-chromatin interaction, *J. Cell Sci.* **112**, 2485–2492.

Galy, V., Olivo-Marin, J.C., Scherthan, H., Dove, V., Rascalou, N., Nehrbass, U. (2000) Nuclear pore complex in the organization of silent telomeric chromatin, *Nature* **403**, 108–112.

Gerasimova, T.I., Byrd, K., Corces, V.G. (2000) A chromatin insulator determines the nuclear localization of DNA, *Mol. Cell* **6**, 1025–1035.

Gindullis, F., Peffer, N.J., Meier, I. (1999) MAF1, a novel plant protein interacting with matrix attachment region binding protein MFP1, is located at the nuclear envelope, *Plant Cell* **11**, 1755–1768.

Goldberg, M., Harel, A., Brandeis, M., Rechsteiner, T., Richmond, T.J., Weiss, A.M., Gruenbaum, Y. (1999) The tail domain of lamin Dm0 binds histones H2A and H2B, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2852–2857.

Gotta, M., Laroche, T., Formenton, A., Maillet, L., Scherthan, H., Gasser, S.M. (1996) The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type Saccharomyces cerevisiae, *J. Cell Biol.* **134**, 1349–1363.

Grogan, J.L., Mohrs, M., Harmon, B., Lacy, D.A., Sedat, J.W., Locksley, R.M. (2001) Early transcription and silencing of cytokine genes underlie polarization of T helper cell subsets, *Immunity* **14**, 205–215.

Guillemin, K., Williams, T., Krasnow, M.A. (2001) A nuclear lamin is required for cytoplasmic organization and egg polarity in Drosophila, *Nat. Cell Biol.* **3**, 848–851.

Gunewardena, S., Rykowski, M.C. (2000) Direct evidence for interphase chromosome movement during the mid-blastula transition in Drosophila, *Curr. Biol.* **10**, 285–288.

Haaf, T., Ward, D.C. (1995) Rabl orientation of CENP-B box sequences in Tupaia belangeri fibroblasts, *Cytogenet. Cell Genet.* **70**, 258–262.

Henikoff, S., Jackson, J.M., Talbert, P.B. (1995) Distance and pairing effects on the brown dominant heterochromatic element in Drosophila, *Genetics* **140**, 1007–1017.

Heun, P., Laroche, T., Shimada, K., Furrer, P., Gasser, S.M. (2001) Chromosome dynamics in the yeast interphase nucleus, *Science* **294**, 2181–2186.

Hill, R.J., Whytock, S. (1993) Cytological structure of the native polytene salivary-gland nucleus of Drosophila melanogaster – a microsurgical analysis, *Chromosoma* **102**, 446–456.

Hilliker, A.J. (1985) Assaying chromosome arrangement in embryonic interphase nuclei of Drosophila melanogaster by radiation induced interchanges, *Genet. Res.* **47**, 13–18.

Hochstrasser, M., Mathog, D., Gruenbaum, Y., Saumweber, H., Sedat, J.W. (1986) Spatial organization of chromosomes in the salivary gland nuclei of Drosophila melanogaster, *J. Cell Biol.* **102**, 112–123.

Hoefers, C., Baumann, P., Hummer, G., Jovin, T.M., Arndt-Jovin, D.J. (1993) The localization of chromosome domains in human interphase nuclei. Three-dimensional distance determinations of fluorescence *in situ* hybridization signals from confocal laser scanning microscopy, *Bioimaging* **1**, 96–106.

Hutchison, N., Weintraub, H. (1985) Localization of DNAase I-sensitive sequences to specific regions of interphase nuclei, *Cell* **43**, 471–482.

Jin, Q.W., Fuchs, J., Loidl, J. (2000) Centromere clustering is a major determinant of yeast interphase nuclear organization, *J. Cell Sci.* **113**, 1903–1912.

Kozubek, S., Lukásová, E., Marecková, A., Skalníková, M., Kozubek, M., Bártová, E., Kroha, V., Krahulcová, E., Slotová, J. (1999) The topological organization of chromosomes 9 and 22 in cell nuclei has a determinative role in the induction of t(9,22) translocations and in the pathogenesis of t(9,22) leukemias, *Chromosoma* **108**, 426–435.

Laroche, T., Martin, S.G., Gotta, M., Gorham, H.C., Pryde, F.E., Louis, E.J., Gasser, S.M. (1998) Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres, *Curr. Biol.* **8**, 653–656.

LaSalle, J.M., Lalande, M. (1996) Homologous association of oppositely imprinted chromosomal domains, *Science* **272**, 725–728.

Lopez, J.M., Wolfner, M.F. (1997) The developmentally regulated Drosophila embryonic nuclear lamina protein 'Young Arrest' (fs(1)Ya) is capable of associating with chromatin, *J. Cell Sci.* **110**, 643–651.

Luderus, M.E., den Blaauwen, J.L., de Smit, O.J., Compton, D.A., van Driel, R. (1994) Binding of matrix attachment regions to lamin polymers involves single-stranded regions and the minor groove, *Mol. Cell Biol.* **14**, 6297–6305.

Manuelidis, L., Borden, J. (1988) Reproducible compartmentalization of individual chromosome domains in human CNS cells revealed by in situ hybridization and three-dimensional reconstruction, *Chromosoma* **96**, 397–410.

Marshall, W.F., Dernburg, A.F., Harmon, B., Agard, D.A., Sedat, J.W. (1996) Specific interactions of chromatin with the nuclear envelope: positional determination within the nucleus in Drosophila melanogaster, *Mol. Biol. Cell* **7**, 825–842.

Marshall, W.F., Straight, A., Marko, J.F., Swedlow, J., Dernburg, A., Belmont, A., Murray, A.W., Agard, D.A., Sedat, J.W. (1997) Interphase chromosomes undergo constrained diffusional motion in living cells, *Curr. Biol.* **7**, 930–939.

Masuzawa, N., Urata, Y., Yagi, K., Ashihara, T. (2000) Constrained, random, and independent motion of texas-red labeled chromatin in living interphase PtK2 cells, *Acta Histochem. Cytochem.* **33**, 419–427.

Moenkhaus, W.J. (1904) The development of the hybrids between Fundulus heteroclitus and Medidia notata with especial reference to the behavior of the maternal and paternal chromatin, *Am. J. Anat.* **3**, 29–65.

Murray, A.B., Davies, H.G. (1979) Three-dimensional reconstruction of the chromatin bodies in the nuclei of mature erythrocytes from the newt Triturus cristatus: the number of nuclear envelope-attachment sites, *J. Cell Sci.* **35**, 59–66.

Neves, H., Ramos, C., da Silva, M.G., Parreira, A., Parreira, L. (1999) The nuclear topography of ABL, BCR, PML, and RAR alpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation, *Blood* **93**, 1197–1207.

Nikiforova, M.N., Stringer, J.R., Blough, R., Medvedovic, M., Fagin, J.A., Nikiforov, Y.E. (2000) Proximity of chromosomal loci that participate in radiation-induced rearrangements in human cells, *Science* **290**, 138–141.

Nili, E., Cojocaru, G.S., Kalma, Y., Ginsberg, D., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Berger, R., Shaklai, S., Amariglio, N., Brok-Simoni, F., Simon, A.J., Rechavi, G. (2001) Nuclear membrane protein LAP2beta mediates transcriptional repression alone and together with its binding partner GCL (germ-cell-less), *J. Cell Sci.* **114**, 3297–3307.

Oegema, K., Marshall, W.F., Sedat, J.W., Alberts, B.M. (1997) Two proteins that cycle asynchronously between centrosomes and nuclear structures: Drosophila CP60 and CP190, *J. Cell Sci.* **110**, 1573–1583.

Palladino, F., Laroche, T., Gilson, E., Axelrod, A., Pillus, L., Gasser, S.M. (1993) SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres, *Cell* **75**, 543–555.

Quick, P. (1980) Junctions of polytene chromosomes and the inner nuclear membrane, *Experientia* **36**, 456–457.

Robinett, C.C., Straight, A.F., Li, G., Wilhelm, C., Sudlow, G., Murray, A., Belmont, A.S. (1996) *In vivo* localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition, *J. Cell Biol.* **135**, 1685–1700.

Sachs, R.K., Hlatky, L.R., Trask, B.J. (2000) Radiation-produced chromosome aberrations: colourful clues, *Trends Genet.* **16**, 143–146.

Sass, G.L., Henikoff, S. (1999) Pairing-dependent mislocalization of a Drosophila brown gene reporter to a heterochromatic environment, *Genetics* **152**, 595–604.

Selvin, P.R., Scalettar, B.A., Langmore, J.P., Axelrod, D., Klein, M.P., Hearst, J.E. (1990) A polarized photobleaching study of chromatin reorientation in intact nuclei, *J. Mol. Biol.* **214**, 911–922.

Shelby, R.D., Hahn, K.M., Sullivan, K.F. (1996) Dynamic elastic behavior of alpha-satellite DNA domains visualized in situ in living human cells, *J. Cell Biol.* **135**, 545–557.

Skaer, R.J., Whytock, S., Emmines, J.P. (1976) Intranuclear electrophoresis of the chromatin of living cells, *J. Cell Sci.* **21**, 479–496.

Sukegawa, J., Blobel, G. (1993) A nuclear pore complex protein that contains zinc finger motifs, binds DNA, and faces the nucleoplasm, *Cell* **72**, 29–38.

Sun, H.B., Yokota, H. (1999) Correlated positioning of homologous chromosomes in daughter fibroblast cells, *Chromosome Res.* **7**, 603–610.

Talbert, P.B., LeCiel, C.D., Henikoff, S. (1994) Modification of the Drosophila heterochromatic mutation brown dominant by linkage alterations, *Genetics* **136**, 559–571.

Taniura, H., Glass, C., Gerace, L. (1995) A chromatin binding site in the tail domain of nuclear lamins that interacts with core histones, *J. Cell Biol.* **131**, 33–44.

Therman, E., Susman, B., Denniston, C. (1989) The nonrandom participation of human acrocentric chromosomes in Robertsonian translocations, *Ann. Hum. Genet.* **53**, 49–65.

Tumbar, T., Belmont, A.S. (2001) Interphase movements of a DNA chromosome region modulated by VP16 transcriptional activator, *Nat. Cell Biol.* **3**, 134–139.

Vazquez, J., Belmont, A.S., Sedat, J.W. (2001) Multiple regimes of constrained chromosome motion are regulated in the interphase

Drosophila nucleus, *Curr. Biol.* **11**, 1227–1239.

Vourc'h, C., Taruscio, D., Boyle, A.L., Ward, D.C. (1993) Cell cycle-dependent distribution of telomeres, centromeres, and chromosome-specific subsatellite domains in the interphase nucleus of mouse lymphocytes, *Exp. Cell Res.* **205**, 142–151.

Wilkie, G.S., Shermoen, A.W., O'Farrell, P.H., Davis, I. (1999) Transcribed genes are localized according to chromosomal position within polarized Drosophila embryonic nuclei, *Curr. Biol.* **9**, 1263–1266.

Wu, C.T., Morris, J.R. (1999) Transvection and other homology effects, *Curr. Opin. Genet. Dev.* **9**, 237–246.

Ye, Q., Worman, H.J. (1994) Primary structure analysis and lamin B and DNA binding of human LBR, an integral protein of the nuclear envelope inner membrane, *J. Biol. Chem.* **269**, 11306–11311.

Zink, D., Cremer, T., Saffrich, R., Fischer, R., Trendelenburg, M.G., Ansorge, W., Stelzer, E.H. (1998) Structure and dynamics of human interphase chromosome territories *in vivo*, *Hum. Genet.* **102**, 241–251.

Chromosome, Microdissection and Microcloning

Nabil G. Hagag
State University of New York, Stony Brook, NY, USA

1 Chromosome Structure and Organization 20

2 Preparation of Chromosomes for Microdissection 22
2.1 Aspects of Preparing Mammalian Chromosomes 22
2.1.1 Sources of DNA 22
2.1.2 Hypotonic Solution 23
2.1.3 Chromosome Fixation 23
2.1.4 Aging 24
2.1.5 GTG Banding for Chromosome Microdissection 24
2.2 Preparation of Dipteran Chromosomes 24

3 Methods of Chromosome Microdissection 25
3.1 Equipment 25
3.2 Methods 26
3.2.1 Oil Chamber Method 26
3.2.2 Videomicroscope Method 27
3.2.3 Laser Microbeam Method 29
3.3 Dissection and Collection of Chromosome Fragments by the Videomicroscopy Method 30

4 Methods of Cloning and Analysis of Microdissected DNA 31
4.1 Cloning Methods 31
4.1.1 Microcloning 33
4.1.2 Linker Adapter with PCR Amplification 34
4.1.3 PCR Amplification of Chromosomal DNA 35
4.2 Analysis of Cloned DNA 38

5 Laser-capture Microdissection (LCM) 40

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

6 Applications of Chromosome and Tissue Microdissection 43
6.1 Direct Analysis of the PCR Product of Microdissected Chromosome Fragments 43
6.2 Recombinant DNA Libraries Generated from Microdissected Chromosome Fragments 44
6.3 Isolation of Transcribed Sequences Using Microdissected DNA as a Probe 44
6.4 Study of Chromosome Abnormalities in Cancer Cells 45
6.5 Investigation of Other Specialized Chromosomal Structures 46
6.6 Gene Transfer Using Chromosome Fragments 47
6.7 Coupling Microdissection with Microarrays and Proteomics Technologies (DNA/Protein Microchips) 48
6.8 Incorporation of Microdissection Techniques in Modern Pathology 50

7 Summary 52

Bibliography 52
Books and Reviews 52
Primary Literature 52

Keywords

BAC
Bacterial artificial chromosome based cloning vector capable of accommodating large fragments of genomic DNA.

Chromosomal Band
Variation in the staining of DNA due to differential binding of dyes along the length of the chromosome.

Chromosome
A self-replicating genetic structure that contains the cellular DNA representing the linear array of genes.

Chromosome Fragment
A small portion of chromosomal DNA obtained by means of physical excision using microdissection.

Gene Mapping
Determining the relative position of one or more genes on chromosomes.

Karyotype
A photomicrograph of individual chromosomes arranged in a standard format showing the number, size, and shape of each chromosome.

Megabase
One million base pairs of nucleotides.

LCM
Laser-capture microdissection. A devise that delivers laser pulses to dissect tissues or chromosomes. Pulses can be tuned in variable micron diameter.

Metaphase Chromosomes
Condensed chromosomes at metaphase during the cell cycle.

Microarray
A technology that assesses gene expression by arraying PCR-amplified cDNA clones or genes at high density on derivatized glass microscope slides (known as a DNA chip). Relative expression levels are simultaneously determined by specific probe hybridization.

Microdissection
A method to physically cut and collect chromosomal fragments by means of very fine needles and micromanipulators or by a laser microbeam.

Polymerase Chain Reaction(PCR)
A technique by which target DNA sequences may be replicated (amplified) selectively *in vitro*.

Proteomics
The global analysis of expressed proteins in a cell or tissue phenotype in attempt to establish the relationship between the genome sequence, the expressed proteins, and the protein–protein interactions.

YAC
A yeast artificial chromosome based cloning vector capable of accommodating larger pieces of genomic DNA (average insert >100 kb).

This article reviews the pertinent aspects of chromosome structure, microdissection techniques, and methods of microcloning. Recent applications of this technology are discussed. Microdissection is a specialized aspect of cell microsurgery by which one can remove a chromosome fragment from a cell in metaphase. The DNA material thus obtained can be used for molecular cloning experiments. Coupling this technology with *in vitro* DNA amplification using polymerase chain reaction (PCR) and gene-transfer techniques have made possible the performance of biochemical and biological experiments using small amounts of DNA obtained from chromosome fragments. The emerging technology of laser-capture microdissection (LCM) and its potential applications in global gene expression and profiling (proteomics) is also introduced.

1
Chromosome Structure and Organization

DNA is coiled in eukaryotic chromosomes in a hierarchical fashion with several levels, to achieve the highest degree of condensation. Chromosome organization is described in the following orders of DNA folding (from lowest to highest): naked double-stranded DNA, DNA coiled around core nucleosomes, the 30 nm chromatin fiber, the 250 nm chromatin fiber containing DNA loops, helical coiling of the 250 nm chromatin fiber, chromosome bands, and finally chromosome regions. At the first level of DNA folding, linear double-stranded DNA is coiled around a histone core octamer (two copies each of H2A, H2B, H3, and H4). The DNA undergoes two left-handed superhelical turns around the core histone octamer to form a *nucleosome*. Adjacent nucleosomes are connected by "linker" DNA, which may be 8 to 114 bp long in different eukaryotic species. Nucleosomes are folded in the form of a simple solenoid structure that generates a 30 nm *chromatin fiber*. There are six nucleosomes per turn (approximately 1.2 kb DNA). In both the interphase and metaphase nucleus, the 30 nm fiber is folded in a loop structure. The chromatin loops contain between 5 and 100 kb of DNA. One model predicts radial arrays of ten 30 kb loops yielding an array of about 300 kb – that is, approximately 250 nm in width. The chromatin loops are anchored to a nonhistone protein structure that is referred to as the *nuclear matrix* or the *chromosome scaffold*. The metaphase chromosome scaffold is positioned along the central axis of the chromatid. *Topoisomerase II,* a major component of the scaffold structure, is localized along the length of the chromatid. Topoisomerase II appears to behave as a "loop fastener" in addition to its catalytic function during replication and transcription. DNA segments that interact with the scaffold (scaffold-associated regions: SAR) or with the nuclear matrix (matrix-associated regions: MAR) are 0.6 to 1.0 kb in length and are presumed to form the base of the loop structure. The metaphase scaffol can be visualized microscopically by using treatments that partially deplete chromosomes of histones. Under these conditions, the scaffold has been observed to undergo a helical coiling resulting in a further ninefold compaction of the chromosome in metaphase. This structural feature accounts for the zigzag appearance of chromatids and probably represents partially uncoiled structures.

During interphase, the extended, uncoiled chromosome scaffold structure yields chromatids approximately 250 nm in width, whereas the condensed, coiled scaffold in metaphase yields a width of 700 nm. After metaphase, there is selective decondensation (uncoiling) of euchromatic chromosome regions. Specific mammalian chromosomes can be identified by the pattern of transverse bands produced by a fluorescent dye or by Giemsa stain. Chromosome bands have become topographical landmarks used to map genes, inherited traits, and chromosome structural abnormalities, and as reference points for large-scale genomic sequencing projects. Chromosome bands represent not only structural but also functional compartmentalization of the genome.

The number of visible bands (e.g. Giemsa dark bands or "dark G-bands") depends on the exact stage in the cell cycle at which chromosomes are examined. Chromosomes in midprophase will yield approximately 2000 bands in a human karyotype. Later, in early metaphase, when chromosomes are condensed, only 450 to 800 G-bands will be visible. An

average dark G-band in prophase contains approximately 1.5 Mb of DNA, whereas an average-sized light G-band in metaphase contains 10 to 30 Mb. Therefore, condensed metaphase bands are more heterogeneous (i.e. composed of multiple band types) than those found in midprophase. Each metaphase G-band contains about 15 to 50 fg of DNA. Chromosome bands can be visualized using other fluorochromes. Quinacrine mustard and its derivatives have specificity for chromatin regions that are rich in deoxyadenine (dA) and deoxythymidine (dT). Quinacrine staining produces a banding pattern (Q-bands) that is different from that due to fluorochromes with affinity for deoxycytidine (dC) and deoxyguanidine (dG) residues (e.g. chromomycin A3 and mitramycin). Treatment of fixed chromosomes with trypsin followed by staining with Giemsa (GTG banding) yields dark (G) and light (reverse or R) bands. R-bands usually correlate with bright bands detected with quinacrine (Q-bands). Giemsa staining of human chromosomes after heat treatment results in preferential staining of centromeric and paracentromeric regions, as well as regions of chromosomes 1, 9, and 16 and the telomere of the Y chromosome; these patterns are referred to as *C-bands*. C-bands are regions of constitutive heterochromatin and remain condensed throughout the cell cycle, except for a short period of time during DNA replication.

The structural compartmentalization of chromosomes into transverse bands obtained with Giemsa and fluorescent dyes can also be seen by treating chromosome spreads with the restriction endonuclease *Hae*III, as well as with *in situ* hybridization using repeat-sequence DNA probes. Long interspersed element (LINE) repeats are found predominantly in G-bands, whereas short interspersed element (SINE) repeats or *Alu* sequences are found in R-bands. The differing base composition of these two repeat DNA groups accounts for the higher GC content of R-bands in contrast to higher AT content of G-bands. C bands do not contain SINEs or LINEs but are made of highly repetitive sequences called *satellite DNA sequences*.

The structural organization of chromosome bands has profound implications for the regulation of gene expression. Most housekeeping genes and the majority of tissue-specific genes are found in R-bands. In this regard, clustering of nonmethylated CpG dinucleotides (HTF islands), which is usually found 5′ to "housekeeping" genes, appears to be restricted to R-bands. The staining techniques likely to generate bands for microdissection are summarized in a number of student cytogenetic manuals. Other chromosomal structural regions such as the centromeres and telomeres are also very amenable to microdissection. The primary constriction in mammalian chromosomes is called the *centromere*, and its location is useful in the preliminary grouping of human chromosomes. The centromere is composed of constitutively condensed chromatin, α-satellite DNA, composed of multiple copies of highly diverging basic repeat units of approximately 171 bp. Tandem repeats of α-satellite DNA are further organized into micro repeat units ranging in length from 0.5 to 10 Mb. Human chromosomes can be distinguished from one another by their divergent α-satellite sequences.

The chromosome ends, or *telomeres*, are specialized structures that are essential for the maintenance of chromosome integrity. Telomeres are likely to be important in the spatial positioning of the chromosomes in the nucleus as well. The primary structure of telomeres is a conserved G/C-rich repeat

unit of the general formula D (T/A_{1-4} dG_{1-8}). The terminal repeat number varies considerably in different species, from a few hundred base pairs in ciliates to 10 kb in humans, and up to 150 kb in mice. This review of chromosome structure was confined to essential background, and more detailed discussion of this subject can be found elsewhere.

2 Preparation of Chromosomes for Microdissection

Methods of preparing chromosomes and chromosome bands have been empirically developed by cytogeneticists over the last four decades. These methods, however, must be modified when chromosomes are to undergo microdissection. Conventional banding and staining techniques result in degradation of chromosomal DNA and hence render the DNA unsuitable for biochemical and biological studies. In fact, it is evident that improvement in GTG banding is associated with increased DNA degradation. A number of factors contribute to DNA breakdown during the preparation of chromosome spreads. DNA appears to be more unstable (i.e. susceptible to denaturation and DNAse digestion) in condensed metaphase chromatin than it is in interphase chromatin. Furthermore, at each step in the preparation of metaphase chromosomes – that is, hypotonic swelling, fixing, and "aging" – DNA degradation is facilitated.

Chromosome fixation and aging contribute most significantly to chromosomal DNA degradation. Therefore, it is essential in microdissection experiments to consider chromosomes as units of linear DNA rather than cytogenetic structures.

Several critical aspects of chromosome preparation for microdissection experiments are worth mentioning. These include the enrichment of preparations for metaphase spreads, hypotonic treatment, chromosome fixation, aging, staining, and storage of chromosomal DNA. When planning for microdissection, it is ideal to have as many metaphase spreads as possible on one coverslip. Only a fraction of the spreads are amenable to dissection because of the poor extension of chromosomes in some spreads, overlapping chromosomes, or disadvantageous orientation of the chromosome of interest. The number of metaphase spreads may be increased by synchronization in the cell cycle to capture the maximum number of cells in mitosis. In most instances, adequate numbers of mitotic cells can be obtained from short- or long-term monolayer cultures without drug synchronization by plating cells sparsely and harvesting mitotic cells when cells are in a log phase of growth. If necessary, the thymidine "block" and "release" method may be used. The use of synchronizing agents such as antimetabolites (e.g. methotrexate), cytotoxic drugs (e.g. actinomycin D), or other chemicals capable of inducing DNA damage is not advised.

2.1 Aspects of Preparing Mammalian Chromosomes

2.1.1 Sources of DNA

Several sources can be used to prepare chromosomal DNA for microdissection. If whole-blood microculture is used, metaphase spreads are prepared from peripheral blood T lymphocytes as follows. Venous blood is drawn into a heparinized tube or into a syringe containing 0.1 mL of heparin stock solution (5000 IU mL^{-1}).

A few drops of heparinized blood are mixed with 5 mL of prewarmed chromosome 4A medium or growth medium containing 0.2 mL of phytohemagglutinin solution in sterile polypropylene tubes. Cultures are set up in duplicate and placed in an incubator at 37 °C for 72 h. A stock solution of colcemid (diacetylmethylcolchicine) is added to each tube to a final concentration of 0.05 μg mL^{-1} and incubated for an additional 30 min. The cell pellet is prepared by centrifugation at 800 *g* for 10 min and resuspended in warm (37 °C) 5 mL hypotonic solution (0.075 M KCl) and incubated at 37 °C for 10 min. The cell pellet is prepared again by centrifugation at 800 *g* and is resuspended in 5 mL of cold 3 : 1 methanol–acetic acid fixative on ice. Cells are dropped onto large, (35 mm × 50 mm) cold, wet coverslips using a Pasteur pipette. Coverslips are examined first on a phase microscope to determine whether the spreads are at the correct density. The coverslips are kept at −70 °C in ethanol for a few hours or until use. Immediately before use, coverslips are immersed in ice-cold distilled water. If monolayer tissue culture cells are used to prepare chromosomes, the cells are seeded at low density onto plastic petri dishes for 24 h; then the medium is removed and replaced with fresh medium. Colcemid solution is added 48 to 72 h later and cultures are incubated for 10 min to 1 h at 37 °C in a 5% CO_2 incubator. Cells are removed from the culture plate or flask and processed as described for lymphocytes.

Among the numerous effective mitotic spindle inhibitors, colcemid is particularly useful, as it arrests cells at the M/G2 interphase. Since chromosomes become more condensed and shorter at the end of metaphase, short incubation times with colcemid (10–60 min) tend to yield more extended chromosomes. This is preferable, particularly when one is dissecting small chromosomes or attempting to dissect minor bands. The use of agents that promote chromosome extension and enhance DNA nicking (e.g. ethidium bromide) should be avoided.

2.1.2 Hypotonic Solution

The hypotonic solution swells the nucleus and cytoplasm, breaks intrachromosomal connections, and allows a better separation of chromosomes when the cells are smashed on a coverslip. Hypotonic treatment of cells results in degradation of DNA from chromosome spreads. Therefore, the shortest length of exposure to hypotonic solution is preferable. Lymphocyte cultures usually can be treated for 10 min at 37 °C to produce adequate spreads, whereas monolayer cell lines are usually treated for 30 min at 37 °C. Prolonged hypotonic treatment results in cell lysis and dispersion of chromosomes over a large area. Less than optimum hypotonic treatment results in tightly packed metaphase spreads that may not be suitable for microdissection.

2.1.3 Chromosome Fixation

Chromosome fixation serves to arrest chromosomes at a specific stage in the cell cycle with little distortion of morphological detail. Optimum fixation also rids the spreads of cytoplasmic debris. The most commonly used fixative is a mixture of methanol and acetic acid, at a ratio of 3 : 1 or higher. Acetic acid, which has excellent penetrability, will precipitate nucleoprotein, extract considerable amounts of chromosomal proteins, and remove cytoplasmic constituents from metaphase spreads. Mixtures of acetic acid and methanol produce minimal swelling or condensation of chromosomes. Acetic

acid fixation for chromosome microdissection studies has one major adverse effect: it causes depurination and nicking of DNA. Depurinated DNA is more susceptible to enzymatic digestion as well as to the DNA degradation associated with chromosome aging. Therefore, a number of precautions and modifications of standard fixation protocols should be observed. Since acid depurination is a hydration reaction, acetic acid and methanol should be free of water. Reagents should be aliquoted into small bottles and kept well sealed. All steps in chromosome fixation and spreading should be done at 0 to 4 °C. Care should be taken to avoid prolonged exposure of chromosomes to fixative. Depending on the cell type, cells may be kept in fixative for a few seconds to 10 min before spreads are dropped onto coverslips.

2.1.4 Aging

Once metaphase spreads have been fixed and dropped onto coverslips, the chromosomes must be aged to obtain optimum GTG banding. Aging is accomplished by keeping chromosome spreads at room temperature for 3 to 5 days or at 56 °C overnight. The precise physicochemical basis for the aging of chromosomes is unknown, but it is evident that chromosomal DNA that has become depurinated by acid fixation grows increasingly single stranded and is degraded during the aging process. Therefore, when possible, it is preferable to microdissect fresh, uniformly stained (solid staining) or unstained chromosomes. This procedure is certainly applicable to the microdissection of telomeric or centromeric regions of chromosomes, which are readily recognizable by size and morphology. Similarly, the microdissection of marker chromosomes and double minutes can be performed with unstained material.

2.1.5 GTG Banding for Chromosome Microdissection

GTG banding of metaphase chromosomes is the most commonly used technique when chromosomes must be stained for identification of a specific chromosome or chromosome region. Coverslips are removed from ethanol and are placed in a sterile Coplin jar containing pH 6.8 buffer at room temperature. The coverslips are immediately removed from the buffer and are placed in sterile 0.5% trypsin/EDTA solution (in Hanks buffered salt solution), and incubated at room temperature for 1 to 3 min. If solid staining is desired, the trypsin/EDTA step is omitted. The coverslips are washed for 1 s in pH 6.8 buffer and then placed directly into Giemsa stain. Staining may take 15 to 90 s. The coverslips are washed twice in pH 6.8 buffer and air-dried. The optimum time for trypsin treatment and staining should be determined experimentally. After a coverslip has been air-dried, DNA will begin aging. Therefore, to obtain high-quality DNA, it is recommended that microdissection be performed immediately after staining. Also, to avoid contaminating slides with foreign DNA, the staining procedure should be performed in a biological containment hood. Sumner has reviewed chromosome banding.

2.2 Preparation of Dipteran Chromosomes

Drosophila chromosomes may be prepared from dipteran salivary gland for microdissection. Salivary glands are dissected in insect Ringer's solution and are spread on a siliconized coverslip in a drop of 45% acetic acid. The tissue is squashed between two siliconized coverslips. The coverslips are then held in place with forceps and quick-frozen in liquid nitrogen.

They are separated by flipping them apart with a razor blade. The coverslip containing the squash preparation is placed in three successive ethanol washes (70%, 95%, absolute) for 1 min and either kept in ethanol or air-dried; it may then be kept in a moisture-free, airtight box until use. Conditions of aging, fixing, and storage described for mammalian chromosomes should be used when preparing *Drosophila* chromosomes.

3 Methods of Chromosome Microdissection

The methods used in microdissection and isolation of chromosome fragments are an extension of the methods and instrumentation originally used for cell microsurgery. Many of the cutting needles and nanoliter volume pipettes were described nearly a half-century ago by De Fonbrune. De Fonbrune also described the tools for fashioning micropipettes, such as micromanipulators and the microforge. Edström and his colleagues later reported the first definitive chromosome microdissection experiment using chromosome spreads in an oil chamber. The oil chamber was placed on an upright microscope and microdissections and DNA extractions were performed using curved micropipette tips and nanoliter volumes. Subsequently, microcloning experiments were done in nanoliter aqueous drops under oil. While this was an effective method, particularly in skilled hands, Hagag and Viola and others found the use of an inverted microscope and straight needles to be technically easier and more generally applicable. Hagag and Viola described and improved the method by utilizing an inverted microscope equipped with a high-magnification system. Other groups have described microscopes equipped with a laser microbeam for chromosome microdissection. The laser microbeam has evolved into a new technology known as *laser-capture microdissection* (LCM). LCM allows the collection of a minute amount of DNA and RNA from a pure cell population in a heterocomplex tissue. A brief account of each of these methods is described below.

3.1 Equipment

The equipment for chromosome microdissection includes a microscope, a pair of micromanipulators, a micropipette puller, a microinjector, micropipette holders, and glass tubing to make microdissection needles and nanoliter transfer pipettes. A microforge and a micropipette grinder are optional for fashioning specialized micropipettes. An ideal microscope (upright or inverted) should have the following features: first, a high-intensity light source with a short and efficient light path; second, a high numerical aperture phase and fluorescence optics. The condenser should permit a sufficient working distance (10–20 mm) and should allow easy interchange between the various optics during manipulations. Third, several specimen holders that permit easy access to chromosomes on the slide are required. These holders should be easily attachable to the stage and capable of rotating freely to align chromosomes with microdissection pipettes. Other accessories (camera ports, video ports, etc.) must not interfere with accessibility to the micromanipulators and micropipettes. At least one micromanipulator is required for microdissection of chromosomes. A second manipulator may be handy for the simultaneous manipulation of the chromosomal fragments after

dissection (e.g. fragment collection, transfer to aqueous drops). Three-dimensional mechanical or hydraulic micromanipulators are best suited for this application. A hydraulic drive system with three-axis motion and single-direction angular movement controls ensures smooth, precise, vibration-free movement during operation. Most micromanipulators are equipped with a fine drive joystick to reduce hand movement during dissection, separation, and collection of fragments.

Microdissection needles or micropipettes are produced from borosilicate capillary tubes (1 mm o.d., 0.65 mm i.d.) using a two-stage micropipette puller (e.g. Narshige PP83). A micropipette puller helps prepare needles consistently with a specifically programed tip diameter. Dissecting needles and volumetric micropipettes are drawn to a long taper with a tip diameter of approximately 0.5 μm. The pipette ends must be flexible enough to permit the tip of the needle to accept pressure without breaking. This balance between needle flexibility and sharpness of the tip must be determined empirically by the investigator. The diameter of the micropipette can be estimated using a microscope and a microgrid (e.g. Graticules Ltd., England). The dissecting needle is fitted into a microinjector (e.g. Eppendorf) holder or a microsyringe. The micropipettes are connected with Tygon or Teflon tubing to the injector. The micropipette, tubing, and syringe may be filled with silicone fluid (which is inert for most biochemical procedures), to provide a closed injection–suction system. Chromosome fragments are cut and pooled into an aqueous drop of 1 to 2 mL. The collection droplet is placed under a small volume of silicone oil on a depression slide or a moist chamber next to the dissection slide.

3.2 Methods

3.2.1 Oil Chamber Method

Microdissection using the oil chamber method was first described for dissecting and microcloning of polytene chromosomes and has been applied since to the dissection of mouse and human metaphase chromosomes. The oil chamber consists of a one-piece glass slide that measures $70 \times 35 \times 6\ mm^3$ with a rectangular groove 25 mm wide and 3 mm deep. Chromosome spreads are prepared on $12 \times 30 \times 0.17\ mm^3$ coverslips. Glass chambers and coverslips are washed in hydrochloric acid and distilled water before use. The coverglass is placed on the groove (chromosome spreads side down). All microdissection and microcloning operations are performed on the lower surface of the coverslips using various dissection needles and volumetric pipettes. The space between the coverslips and the bottom of the chamber is filled with liquid paraffin. The liquid paraffin used in the oil chamber also contains phenol–chloroform droplets and buffers to be used for extraction, solubilization, and restriction enzyme digestions and subsequent microcloning. In this method, investigators used an upright phase contrast microscope equipped with a long working distance condenser to view through the thick oil chamber. Focusing was done by moving the stage. Microdissection was performed under a high-power dry objective (e.g. 40×). All other manipulations, such as micropipetting, transfer of DNA fragments to aqueous droplets, and DNA extractions, were done with low-power objectives (10 and 20×). Figure 1 represents the oil chamber apparatus and microcloning steps.

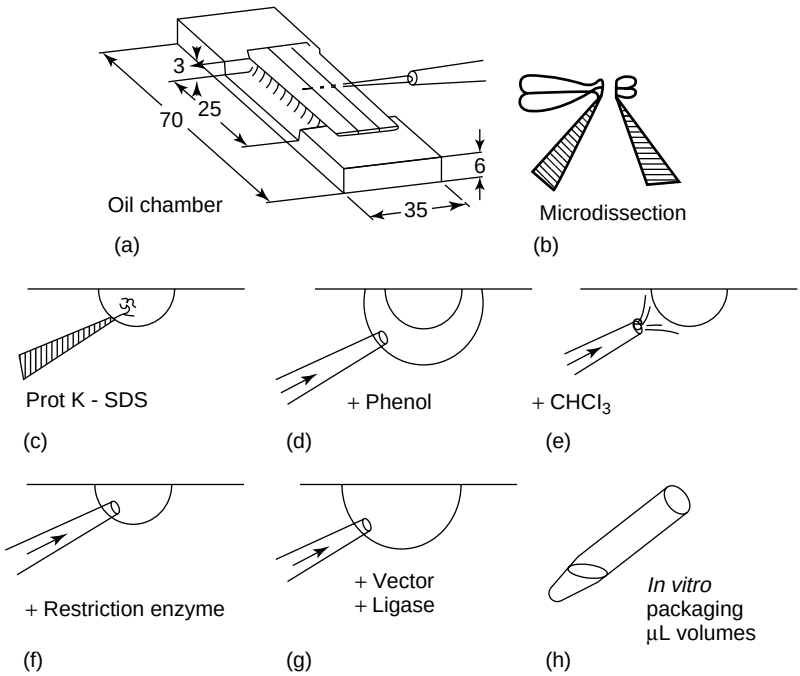

Fig. 1 Oil chamber microdissection apparatus: (a) dimensions of the oil chamber (mm) and (b)–(h) steps of microcloning operations.

3.2.2 Videomicroscope Method

We have described a chromosome workstation that can be used for microdissection and *in situ* hybridization experiments (Fig. 2a). The station features an inverted microscope equipped with a fixed stage, a low-level light video camera, and a video monitor. A 12 V, 100 W halogen lamp and a high-focal length phase-contrast condenser were added to enhance illumination and obtain a better resolution of chromosomal GTG bands. In addition, a new microscope stage was designed, featuring a mechanical stage with *XY* coordinates graduation and a circular rotating stage in the center (Fig. 2b). The circular stage accommodates the holders for the dissection and collection slides. For easy relocation of a specific chromosomal spread, the center stage is also graduated 360°. The circular stage offers the advantage of rapidly positioning target chromosomes perpendicular to the dissecting needle tip regardless of their orientation in the spread on the slide. Also, the stage accepts large coverslips (35 mm × 50 mm), thus maximizing the number of spreads on a slide as well as the number of chromosome fragments collected per session.

Video cameras have an inherent magnification factor that may range from 10× to 200×. For example, the video camera series QX-104 (VSP Inc.) and the charge-coupled device (CCD) video camera CCD-72 (Dage Corp.) have an intrinsic magnification factor of 40×. When coupled with a microscope using a 40 or 63× objective, we were able to achieve magnification of metaphase chromosomes on the

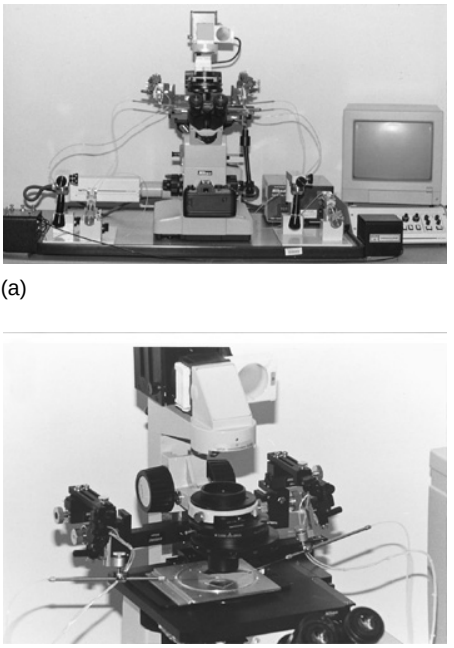

(a)

(b)

Fig. 2 (a) Videomicroscope microdissection apparatus. This system is composed of the Nikon Diaphot microscope, CCD 72 video camera, high-resolution display monitor, two micromanipulators, two microsyringes, and the rotating microdissection/microinjection stage. The entire system is placed on a tabletop vibration-free system. (b) Close-up showing the rotating stage and the micromanipulators mounted on the microscope stage.

video monitor by a factor of roughly 1600 to 2500×. Moreover, with the available high-resolution TV/video camera microscope couplers (e.g. 2× HR200 series coupler), an additional 0.5 to 2× magnification factor can be added. When a CCD camera with a 2× coupler was used, the magnification was further raised to about 3200 to 5000× using 40 and 63× dry objective lenses respectively. The CCD cameras are better for this application because they are ultrasensitive to low light and their parafocality and focusing can be adjusted electronically. The resulting image is displayed on a VGA monitor or an equivalent color monitor. The image on the monitor can be further improved by the appropriate choice of microscope TV/video coupler, interference contrast filter, lens, and resolution of the display monitor. The videomicroscope is placed on a tabletop isolation platform (e.g. Vibraplan, Kinetics Systems, Inc.). These platforms are passive air-mount templates that float freely when inflated to the load-supporting pressure, thus minimizing vibrations transmitted to the microscope and accessories. Microdissection is performed while observing the chromosome on the video monitor. In this manner, it is possible to visualize secondary and minor bands that are below microscope limits of resolution for a light microscope. Moreover, the system's real-time design enables one to observe on the monitor the various micromanipulations, which may serve as a useful teaching tool as well. Using the videomicroscope, it is possible to cut chromosome fragments less than 0.5 μm long (i.e. a single band of a human chromosome). The precision of the dissection of a specific chromosomal band is substantially improved by the two- to threefold enhanced magnification. For example, a human chromosome 1 is

about 7 μm long and measures roughly 2.8 mm long on an objective lens of a light microscope at 40× magnification. With the videomicroscope, it would measure 2.24 cm on the display monitor. Thus, an average GTG band of 0.4 μm will measure about 2 mm using a phase or light microscope and it is relatively difficult to resolve bands corresponding to 10 to 30 Mb of chromosomal DNA. By using the videomicroscope and a fine cutting needle, fragments containing as little as 10 Mb (about 0.4 μm) can be readily resolved.

The videomicroscope microdissection method offers a number of other advantages. The use of an inverted microscope permits convenient working distance for performing the micromanipulations. The system overcomes light- and contrast-limited situations by means of the picture enhancement provided by the video camera and monitor.

Furthermore, microdissection in air makes it easier to see the needle when dissecting chromosomes.

3.2.3 Laser Microbeam Method

The laser microbeam microdissection of Monajembashi et al. is based on the principle that at high photon densities, light can liquefy, evaporate, or break down optically active biological material. Even chemical bonds are cleaved when a biological material is heated locally to a few thousand degrees for nanoseconds to microseconds. The laser microbeam apparatus consists of an excimer laser as a primary source of laser light (e.g. Lambda Physik EMG 103 MSC). An additional dye laser (Lambda Physik FL 2002) is utilized to improve the beam quality and select a specific wavelength. A circular disk system of concentric rings is used to deliver high-pulse energies at a repetition rate of about 10 Hz. Energy densities of more than 10^{14} W/cm^2 can be achieved by selecting a ring with a radial distance that corresponds to the wavelength in use. The pulses are directed into an inverted microscope via the fluorescence illumination path and are focused through the objective into the object plane. The pulses of the laser, 20 ns in length, are directed into a microscope using optics similar to that of a fluorescence microscope (e.g. Ultrafluar 100, which has a numerical aperture of 0.85 by Zeiss).

The microbeam laser system provides pulse energies between 1 and 10 mJ at a wide range of wavelengths (320–800 nm) that is optimum for several dyes and is particularly suitable for working with ultraviolet light. A UV wavelength is appropriate for microdissection of chromosomal DNA because the damaging effect of the laser is limited to an area nanometers away from the cutting region. Therefore, the secondary damage to DNA is negligible. Proponents of this method indicate that DNA situated in chromosome slices dissected by this procedure using UV laser above 300 nm will retain its biological integrity and is highly suitable for subsequent microcloning procedures. Damage of DNA by stray light is expected only for wavelengths well below 300 nm. The principle of this technique has recently been applied to the microdissection and cloning of both human chromosome fragments and *Drosophila* polytene chromosome bands and several other species.

Chromosomes of human lymphocytes were prepared by standard procedures on cover glass, treated by the ring system with a single laser pulse. Each slice was, on average, 0.5 μm thick. The slices were taken up by a microdrop and used as is for cloning experiments. A schematic of the laser microbeam microscope is shown in Fig. 3.

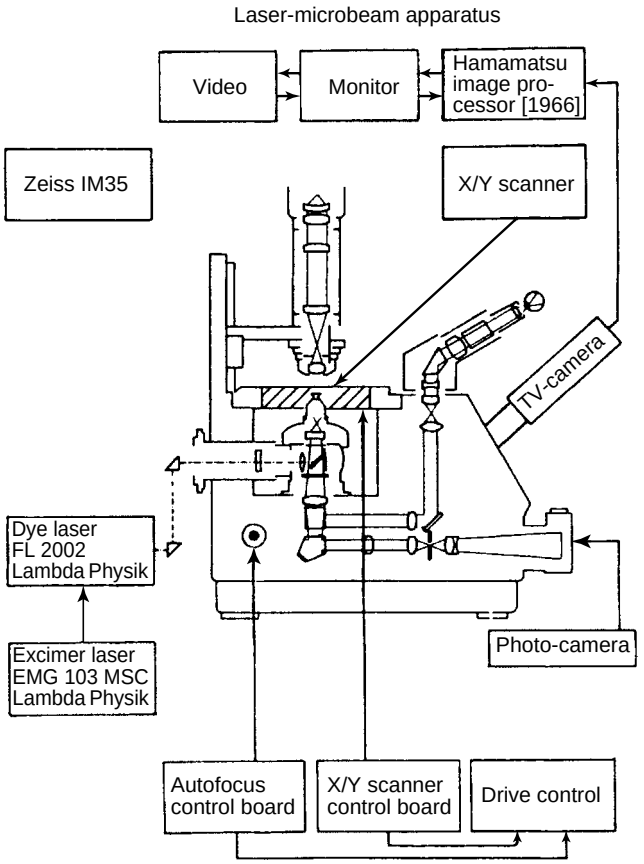

Fig. 3 Diagram of the laser-microbeam microdissection apparatus. The microinstruments necessary for fragment collection and manipulation (not depicted) are essentially the same as described earlier.

3.3 Dissection and Collection of Chromosome Fragments by the Videomicroscopy Method

Approximately two days before microdissection, cultures are initiated to prepare high-mitotic index cells according to standard procedures. Metaphase spreads are prepared a few hours before microdissection and stored in ethanol. Chromosomes are GTG band–stained immediately before the scheduled microdissection. Volumetric micropipettes (i.d. ≃1–2 μm) and straight microdissection needles (o.d. = 0.2–0.5 μm) are prepared at precalibrated settings on a micropipette puller, filled with paraffin oil, and mounted on the appropriate holders. For chromosomal DNA collection, 200 μL of filter-sterilized paraffin or silicone oil is placed on a depression slide or moist chamber near the dissection slide. Similarly, nanoliter drops of proteinase K/SDS and Tris-HCl solutions are placed under oil using volumetric pipettes. Also, several 1 to 2 mL drops of aqueous Tris-HCl/EDTA buffer are placed adjacent

to the proteinase K drop using another volumetric pipette. The aqueous droplets will be used later for washing and solubilizing extracted DNA. For easy manipulation, all these steps should be performed at low magnification (10×). Coverslips containing the metaphase spreads are secured into the customized slide holder, spreads side up, and mounted onto the rotating stage. Usable spread(s) are identified and marked at 10× magnification for easy orientation of the chromosome of interest. The dissecting needle is placed within a few micrometers of the chromosome of interest in the metaphase spread and at 10× magnification. The system is switched to high magnification (40–60×) and video display (3200–5000×). After the desired chromosome has been identified, the GTG band to be dissected is approached at an angle of about 30 to 45° and the dissection needle is brought up into the field of the spread. Dissections are preferably performed in cells with well-spread chromosomes and when the chromosome of interest is located in the periphery of the metaphase spread. This precaution minimizes cross-contamination with other chromosomal material.

Microdissection is performed with the aid of the micromanipulator by moving the micropipette tip back and forth across the chromosome band of interest. Under these conditions, the chromosomes are readily cut across, and the dissected region is detached from the slide. Each dissection leaves a clear zone in the surrounding cytoplasm, thus allowing easy and reliable verification of the dissected chromosomes. Fragments that are not identifiable are discarded. The selected band is scraped off the slide, and the fragment is pushed to the side with a micropipette. A drop of oil is expelled from the micropipette to cover the fragment and provide a closed system during suction and collection of the fragment. The fragment is secured to the tip of the micropipette by applying suction and a holding pressure that is maintained during fragment transfer. The microscope is then switched to low magnification and the depression slide (containing DNA collection and extraction drops) is brought into the field of view. The micropipette is gently introduced into the oil drop containing the aqueous droplets, while the holding pressure is maintained. The fragment is expelled into the droplet containing proteinase K/SDS solution by reversing the micrometer into the injection mode. The fragment should be visible in the drop at this point. This process is repeated until a sufficient number of fragments have been collected. A schematic representation of the sequence of these steps appears in Fig. 4. An example of microdissection of human chromosome 3 at bands 3p13, 3p14, 3p21, and 3q21 using the videomicroscopy method is illustrated in Fig. 5. The time required to microdissect a specific chromosome region depends on the quality and number of metaphase spreads on the coverslip. With some experience, 10 to 15 fragments can be dissected per hour. Approximately 30 to 100 fragments are sufficient for cloning experiments.

4 Methods of Cloning and Analysis of Microdissected DNA

4.1 Cloning Methods

The first experiments to successfully generate recombinant genomic DNA clones directly from microdissected eukaryotic chromosome fragments were performed in the early 1980s by Jan-Erik Edström and

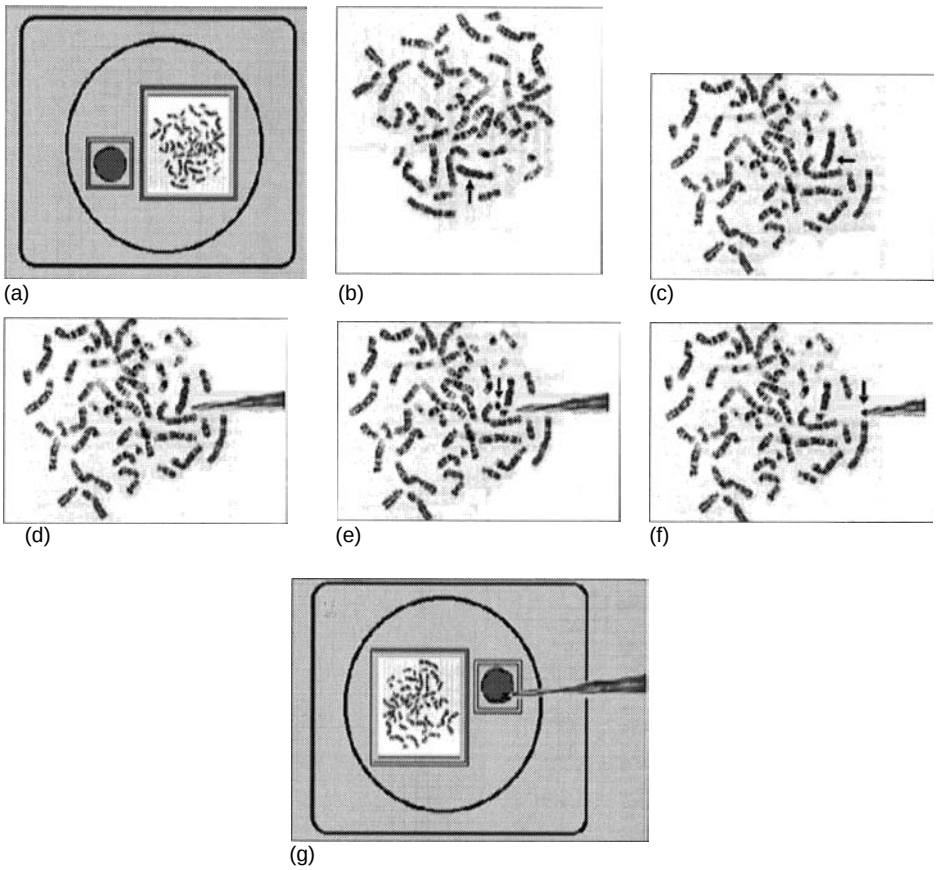

Fig. 4 Sequential steps in chromosome microdissection. (a) Locate a chromosome spread on the dissection slide at 10× magnification, (b) switch to higher magnification to identify the target chromosome for microdissection, (c) rotate the slide to orient the target chromosome perpendicular to the dissection micropipette, (d) lower the dissection micropipette and position the tip proximal to the target chromosomal band (e) cut the chromosomal fragment by scraping across the band in one forward motion, (f) apply suction to the pipette to collect and secure the fragment to the tip of the micropipette, (g) lift the micropipette slowly; rotate the stage to bring the depression slide into the field of view. Lower the pipette, while holding suction, into the aqueous drop under the oil. Eject the fragment into the drop by switching to injection mode.

his colleagues at the European Molecular Biology Laboratory in Heidelberg. The methodology was an outgrowth of the microchemistry work that Edström had developed in the 1960s. During the next twenty years, new technologies emerged and facilitated both the performance of precise chromosome microdissection and the efficiency of cloning chromosomal DNA. A major contribution to these advances was the introduction of PCR technology. PCR permits the amplification of minute quantities of template DNA and generates a product in quantities sufficient for genetic and molecular analysis. Procedures for preparing

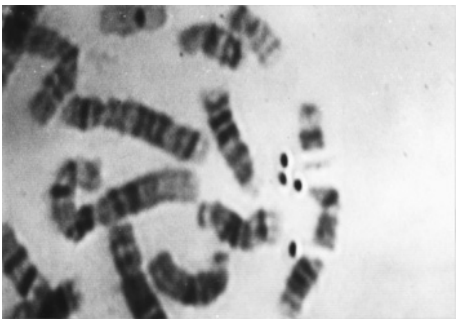

Fig. 5 Example of fragments microdissected from human chromosomes 3 using videomicroscopy. Magnification is 3200× on the display monitor. The fragments cut represent from top to bottom bands 3p21, 3p14, 3p13, and 3q21. Each fragment measures approximately 0.5 μm and contains 10–15 Mbp.

metaphase chromosomes were modified to minimize DNA degradation as reviewed earlier. When care is taken to avoid prolonged acid fixation, microcloning yields recombinant clones with relatively larger DNA inserts. Finally, methods such as videomicroscopy were incorporated to provide higher magnification and resolution of minor chromosome bands.

To date, three distinct approaches can be described for generating recombinant DNA from microdissected chromosomal fragments. In the first and earlier method, DNA was extracted from chromosome fragments in a nanoliter aqueous drop and digested with *Eco*RI restriction enzyme, whereupon the digestion products were directly cloned into a λ-phage vector (this method is often called *microcloning*). The second method is an extension of the first and utilizes PCR. Microdissected DNA is first digested with a restriction enzyme; then, digested DNA is enzymatically ligated either to a linearized plasmid vector or to oligonucleotide linker primers. The cloned inserts are then amplified by PCR using vector-specific primers or the linker primers. This method generates a large number of recombinant clones (usually ≥20 000). DNA from a number of human chromosome bands of considerable clinical interest has been cloned in this way, as discussed in Sect. 4.2. The third method involves the direct PCR amplification of chromosomal DNA using carefully designed oligonucleotide primer sequences that are often termed *universal* or random primers, and more recently using degenerate oligonucleotide primers (DOP-PCR). The universal primers method is based on *Alu* repeat sequences that are naturally present in the genome. The prerequisite of using universal or repeat-sequence primers is that they occur frequently in the genome being studied. The PCR product is then used to probe a more complete recombinant library (cDNA or genomic). The library chosen for screening should contain clones with large inserts [e.g. cosmid, bacterial/yeast artificial chromosome (BAC/YAC) clones]. Hence if a PCR reaction is symmetrical and unbiased, a more complete chromosomal region–specific library can be generated by using the PCR product as a probe. Since our focus is microdissection and microcloning, a more detailed account of these methods is given in Sect. 4.2, with emphasis on the role of PCR and fluorescent *in situ* hybridization (FISH).

4.1.1 Microcloning

To initiate the direct cloning of microdissected chromosomal DNA, the DNA was extracted, first with phenol–chloroform, then with proteinase K, and then digested

with restriction enzyme, *Eco*RI. The smaller fragments generated (estimated to be 3–4 kb) were ligated to a λ-phage vector molecule, a reaction that is not dependent on large quantities of insert DNA and thus could be applied to microcloning. The biochemical steps necessary for cloning (phenol–chloroform extraction, restriction enzyme digestion, ligation to cleaved vector) were performed in aqueous drops under oil. The small quantity of recombinant DNA was then added to the packaging mixture and the recombinant phages were replicated in *E. coli*. Using this technique (microdissection and microcloning), genomic clones were obtained from *Drosophila melanogaster* polytene chromosomes, *D. hydei* lampbrush DNA loops, and mouse and human metaphase chromosomes. In these studies, the generation of multiple clones from specific genetic regions has expedited the identification and isolation of a number of important genes (e.g. the t complex in mice and the Kruppel gene in *Drosophila*). The microcloning method requires a large number of fragments (≥100) to provide sufficient starting material for cloning. However, the library of recombinant clones generated with this method usually contains DNA inserts (≤1 kb) that represent only a small fraction (<5%) of the total DNA sequences in the microdissected region.

In addition to low cloning efficiency, the microcloning method has a number of other problems. The use of a phage DNA vector often results in self-religation of phage DNA and the presence of spontaneous phage mutants (false positives). This problem can be minimized by increasing the ratio of insert to cloning vector and by using highly purified phage vectors. Reducing the initial ligation reaction volume to a few nanoliters would increase the cloning efficiency.

To increase the number of recombinant DNA clones, it is necessary to monitor carefully conditions that promote DNA degradation (e.g. excessive acid fixation of chromosomes during preparation of spreads, chromosome aging on slides before staining). Chromosomes should be prepared immediately before microdissection. All slides containing chromosome spreads, handling pipettes, and solutions must be thoroughly cleaned and sterilized. Slides should be examined for the presence of contaminating microorganisms before use.

4.1.2 Linker Adapter with PCR Amplification

In the second method that involves ligation of microdissected chromosomal DNA with a plasmid vector or a linker adapter and PCR amplification, the chromosomal DNA is first extracted with phenol–chloroform and digested with proteinase K in nanoliter drops. Genomic DNA is digested with an enzyme such as *Rsa*1 (or *Hpa*II) that generates fragments of average length of an optimal size for PCR amplification (1000–2000 bp). Another advantage of using *Rsa*1 enzyme is that its recognition site occurs infrequently in DNA repeat sequences, thus minimizing the preferential cloning of repeat sequences. Next, DNA is ligated to a plasmid vector (e.g. pUC 18) and then enzymatically amplified, using in the vector universal primer binding sites that flank the insert DNA cloning site. These primers, the M13 forward and reverse primers, are as follows:

M13 forward sequencing primer:

5′-ACTGGCCGTCGTTTTAC-3′

M13 reverse sequencing primer:

3′-GTCCTTTGTCGATACTG-5′

The cloned DNA is amplified using PCR with M13 primers. Therefore, only one pair of oligonucleotide primers is used for the simultaneous amplification of many different DNA inserts. PCR amplification is performed according to the following thermal profile: 25 one-minute cycles of denaturation at 94 °C (first cycle, 5 min), 1 min of annealing at 55 °C, and 2 min of primer extension at 72 °C. Following the initial amplification, the DNA inserts are cleaved with a second enzyme (*Eco*RI) that flanks the *Rsa*1 site. The inserts are subsequently subcloned into a second plasmid vector and propagated in the appropriate bacterial host to generate the band-specific recombinant minilibrary. This procedure was first employed to clone specific genes from human chromosome 8 from Langer–Giedion syndrome patients. This cloning method offers several advantages. Cloning in plasmid vectors is easier than cloning in phages and can be more rapidly characterized. Also, the smaller size of the vector renders DNA inserts in plasmid vectors more amenable to enzymatic amplification and direct sequencing than they are in phage vectors. The introduction of PCR allows the amplification of a population of clones that would otherwise go undetected. Finally, the use of PCR significantly reduces the number of chromosomal fragments initially required for cloning.

A similar method in which the microdissected DNA was ligated to a linker adapter molecule instead of a plasmid vector followed by PCR amplification was described. The linker adapter consisted of a 24-nucleotide molecule with a terminus complementary to the *Mbo*1 enzyme recognition sequence. The *Mbo*1 linker adapters have the following sequence:

5′-GATCTGTACTGCACCAGCAAATCC-3′
3′-ACATGACGTGGTCGTTTAGG-5′

The linker adapter is formed by annealing a mixture containing 100 μg mL^{-1} of each oligonucleotide at 58 °C for 1 h. Prior to the annealing step, the 24-mer oligonucleotide component carrying the 5′ *Mbo*1 complementary overhanging end is phosphorylated. This step is necessary to prevent the formation of tandem arrays of linkers during ligation, which may interfere with the amplification of chromosomal DNA. Following chromosomal DNA extraction and digestion as described earlier, the DNA is ligated to the *Mbo*1 linker adapter at 19 °C for 16 h. PCR amplification is carried out as follows: 35 one-minute cycles of denaturation at 94 °C, primer annealing (2 min, 52 °C), and primer extension (3 min, 73 °C). The amplified DNA is subsequently cloned into *Bam*HI linearized Bluescript plasmids after extraction and gel purification. The library is then propagated in competent bacteria.

4.1.3 PCR Amplification of Chromosomal DNA

The third method of microcloning employs PCR amplification of microdissected chromosomal DNA fragments without prior manipulation. The PCR reaction product is examined for self-completeness, is radioisotopically labeled, and is used to probe a more complete recombinant library. In this direct method, DNA sequences from a specific chromosomal region are isolated by amplification of the microdissected DNA with PCR, using a set of universal primers containing a series of degenerate base sequences. The amplification product would represent a heterogeneous population of DNA sequences. Direct PCR amplification of microdissected DNA has been recently employed to isolate genes from band 2F of

the *Drosophila* polytene X chromosome using degenerate "universal" primers. More recently, libraries of specific human chromosome bands were constructed using universal primers and PCR. The use of *Alu*-specific primers to screen for human genes in complex DNA mixtures has been described. Probes generated from microdissected chromosomal DNA fragments with *Alu*-PCR have been used to localize specific chromosomal sequences in YAC recombinant clones. They were also used to generate sequence-tagged sites (STSs) from discrete chromosomal regions. Direct PCR amplification of microdissected DNA has a few advantages over the methods that employ cloning first. Direct amplification of picogram amounts of DNA obviates the time-consuming steps that involve restriction enzyme digestion, ligation, and subcloning. This approach minimizes the loss of minute starting quantities of DNA during the various manipulations.

Most important, the clones constituting the final library contain chromosome-specific DNA inserts of considerably larger size, depending primarily on the recombinant library that was screened. This method does not involve cloning; hence the size of the library depends entirely on the efficiency and completeness of PCR amplification.

The completeness of the PCR product is dependent on the sequence of primers used and the PCR reaction conditions (e.g. annealing temperatures during the early cycles, primer extension times), concentration of deoxynucleoside triphosphates, the quality of DNA polymerase, and finally the number of cycles (usually 60–90) required to amplify sufficient amounts of DNA inserts containing unique sequences.

The term *universal primers* refers to a mixture of oligonucleotide sequences that lack absolute complementarity to the target template sequences. They contain multiple degenerate bases. When one is using these primers in PCR amplification and selecting lower PCR annealing temperatures (46–48 °C), complementary sequences that contain significant homologies can be amplified. For example, the following primer was recently used to generate a probe library from *Drosophila* polytene chromosome fragments:

5′-TTGCGGCCGCATTNNNNTTC-3′

With complete degeneracy at positions 4 to 7 from the 3′ end, authentic DNA sequences from the *Drosophila* genome were isolated. A similar primer has been described in PCR amplification of DNA from human chromosomes. This 22-mer, however, contains complete degeneracy at 6 nucleotides (positions 11–16), with the following sequence:

5′-CCGACTCGAGNNNNNNATGTGG-3′

In general, a nonspecific primer should have the following structure from 5′ to 3′ ends: two or more protective (cap) nucleotides (e.g. TT and CCGA in the primers just described), four to six nucleotides with complete degeneracy, and finally, a terminal 3′ sequence that occurs at high frequency in the genome of the species in which DNA is amplified, such as TTC and ATGTGG in the *Drosophila* and human genomes respectively. Furthermore, a restriction endonuclease site (e.g. *Not*I and *Xho*I, in the just-described primers respectively) may be added for easy recognition of the PCR product and subsequent cloning in an appropriate vector. The *Alu* repeats (SINEs) and L1 repeats (LINEs) are short interspersed repeated DNA elements distributed throughout the genomes of primates. *Alu* elements are dimeric in

structure and are composed of two tandomly arranged halves. *Alu* repeats are approximately 300 bp long and are present at a copy number in excess of 10^6 per haploid human genome equivalent. The L1 repeats, however, are present in 10^4 to 10^5 copies per genome. *Alu* and L1 repeats occur preferentially in light Giemsa bands and dark Giemsa bands respectively.

Oligonucleotide primers based on human *Alu* sequences and L1 sequences were designed and used in PCR amplification of human DNA sequences from hamster–human hybrid cells containing a single human chromosome. The following primers were utilized:

Alu-559: 5′-AAGTCGCGGCCGCTTGCA-GTGAGCCGAGAT-3′

Alu-517: 5′-CGACCTCGAGATCT(C/T)-(G/A)GCTCACTGCAA-3′

L1Hs: 5′-CATGGCACATGTATACATA-TGTAAC(T/A)AACC-3′

A drawback of direct amplification using degenerate primers, however, is that minute quantities of any contaminant DNA will also be amplified along with the chromosomal DNA. Therefore, certain precautions are necessary. Chromosome spreads should be dropped on clean, sterile coverslips. All solutions and even the oil used in the microdissection system should be filter-sterilized and placed under UV light for several hours or overnight before use.

Also, glass cutting needles should be acid/alcohol washed and sterilized under UV light. Pipettes should be designated for use in chromosome fragment amplification only. Occasionally, the *Thermus aquaticus* DNA polymerase used in PCR reactions becomes contaminated with foreign DNA (bacterial or yeast) that served as the PCR template. Therefore, the lots of enzyme should be tested before use in microdissection experiments. No PCR product should be detectable after three consecutive rounds of amplification in the absence of template DNA. Finally, it is customary to run a "water control" (reaction without DNA template) in tandem with each PCR amplification to ensure that the test reaction DNA product was generated from chromosomal fragments.

Direct amplification of microdissected chromosomal DNA also has unique problems. Often, one or more amplified products cannot be visualized on agarose gel after one round of PCR cycle. This situation can be rectified, however, by increasing the number of chromosome fragments in the PCR reaction and increasing the primer concentration ($\geq$1 μg). More important, decreasing the annealing temperature of the initial thermal cycles of PCR (27–30 °C) was found to enhance the amplification product. Running an additional round or two of PCR amplification is recommended. This is accomplished by removing an aliquot (1/10th) from the first round and repeating the PCR. Chromosomal DNA that is precut with restriction enzyme (e.g. *Hpa*II) would generate a homogeneous population of fragments, amenable to uniform PCR amplification. A more troublesome problem – the isolation of large number of clones containing repeat sequences – is the result of the combination of an abundance of these sequences throughout the human genome and the low stringency of PCR annealing temperatures. The problem can be minimized (but not avoided) by fine-tuning the annealing temperature in the initial PCR cycles to a point that biases the PCR toward amplification of unique sequences. Alternatively, following the generation of PCR product it may be useful to perform

suppression hybridization of the product to *Cot*1 or genomic DNA to remove repeat sequences before the product is used in the screening of recombinant libraries.

4.2 Analysis of Cloned DNA

Once a chromosomal region–specific recombinant DNA library has been constructed, it should be tested for its completeness and complexity. One should also verify whether the cloned DNA originated from the microdissected fragment of interest. First, DNA insert size and range must be determined. Second, the percentage of clones containing repeat and unique sequences must be calculated. Third, the percentage of the total microdissected DNA fragments that has been cloned is determined. Fourth, cloned sequences are localized to the respective chromosomal regions by *in situ* hybridization or Southern hybridization to region-specific DNA.

For determination of the average DNA insert size and range, a number of clones (30–50) are picked at random, and the insert DNA is released from the cloning vector by digestion with the appropriate restriction enzyme and is fractionated on an agarose or polyacrylamide gel electrophoresis apparatus using appropriate DNA markers. The restriction enzyme digestion of each of the randomly picked clones provides a pattern that is indicative of the number of unique clones in the library. Estimation of the average size and range of cloned inserts as well as determination of the number of independent clones is used to calculate the completeness of the library.

Next, the random clones can be used to identify the portion of recombinant clones that contain repeat sequences. DNA inserts are transferred to a nitrocellulose filter and probed with total genomic human DNA that is radioisotopically labeled by nick translation. If the library were generated in a completely unbiased manner, a large number (~80%) of these clones would contain repeat sequences and would result in positive hybridization. Unique sequences are determined by digesting total genomic DNA with several restriction enzymes (e.g. *Bam*HI, *Eco*RI, *Pst*I) and performing a Southern hybridization. Recombinant DNA inserts from individual clones are radioisotopically labeled and used to hybridize to restriction polymorphic genomic fragments under conditions that favor unique sequences (65–68 °C).

An ideal library generated from a chromosome fragment would consist of a series of recombinant DNA clones containing overlapping sequences that span the entire chromosomal region. The size of a library of completely random fragments of genomic DNA needed to ensure representation of a particular sequence of interest is dictated by the size and number of the cloned inserts and the size of the chromosomal fragment originally dissected. The following formulas are used to calculate the percentage of DNA that would be cloned from a chromosome fragment. Theoretically, the number of clones that is required to achieve a library of a 99% probability of containing a particular sequence is

$$N = \frac{\ln(1 - P)}{\ln(1 - I/G)}$$

where P is the desired probability, I is the insert size, G is the genome size of the chromosome fragment, and N is the necessary number of clones. For example, to generate a recombinant library of average insert size 2.4 kb (a typical insert size following *Eco*RI digestion), the number of clones necessary to ensure

a 99% probability of cloning the entire 10 Mb fragment is determined as follows:

$$N = \frac{\ln(1 - 0.99)}{\ln[1 - (2.4 - 10^3/10^7)]} = 19\,000$$

The actual number of clones obtained would be determined by factors such as the efficiency of the cloning method, the cloning capacity of the vector, the relative number and size of clonable and nonclonable DNA sequences in the target DNA, and the degree of DNA degradation during the process of chromosome preparation and library construction. Therefore, the percentage of the total DNA that has been cloned would be determined as follows:

$$\%\ \text{fragment DNA cloned} = \frac{\text{experimental number of clones obtained} - \text{mean average insert size}}{\text{fragment DNA content}} \times 100$$

Evolutionarily conserved DNA sequences are indicative of potential structural gene sequences. To determine the *potential structural genes* in the library, the gene(s) of interest can be hybridized to genomic DNA from a number of species using relatively stringent hybridization conditions. The relatedness of a sequence among various species is readily identified by its hybridization to genomic DNA from these species. Genomic DNA from a number of species is now commercially available on biomembrane blots (e.g. Biomap Evo blots, BIOS), sometimes also referred to as *Zooblots*. These blots contain genomic DNA from human, cat, dog, sheep, pig, cow, rabbit, rat, and mouse. Potential structural genes are identified as those that hybridize to DNA sequences from at least three species. Once a structural gene has been recognized, its sequence can be used to isolate the full-length expressed gene from a cDNA library of the appropriate tissue. Alternatively, structural genes may be identified by directly hybridizing cloned genomic DNA to cDNA libraries under conditions that suppress hybridization of repeat sequences. Following the detailed analysis of the recombinant clones, it is also necessary to perform *in situ* hybridization to confirm whether the clones map to the microdissected chromosomal region of interest. In addition to identifying the chromosome on which a gene lies, fluorescence *in situ* hybridization (FISH) helps position the gene of interest at a resolution of about 1 to 2×10^7 bp (0.5%) of the genome. Genomic clones as well cDNA clones can be used in FISH. Unique sequences contained in large-insert DNA probes, such as cosmid and YAC clones, can be used directly in FISH following the suppression hybridization of repetitive sequences first with unlabeled genomic DNA or repeat sequences (e.g. *Cot*1 DNA).

The fluorescent labeling system is based on the modification of the DNA probe with biotin. The probe DNA is labeled using nick translation. Biotin is detected by its binding to avidin or streptavidin coupled to enzymes such as horseradish peroxidase or to fluorochromes. Fluorescent *in situ* hybridization of metaphase chromosome spreads involves two steps:

1. Denaturing the two strands of the chromosomal DNA contained in a metaphase spread on a coverslip, followed by hybridization to a denatured fluorescent-labeled oligonucleotide probe.
2. Localization of the hybridization signal to a specific chromosomal band region.

Significant amplification of the signal can be achieved using a second layer of a biotinylated antiavidin antibody and a third

layer of fluorochrome-coupled avidin. The DNA to be used for *in situ* work should be of the highest quality. The probe should be devoid of repetitive sequences. Any repeat sequences that are present must be removed by subcloning into a suitable vector or by using suppression hybridization in the detection reaction. All other components for *in situ* hybridization are readily available in kits from biochemical companies specializing in molecular biology products.

There are at least three methods for the identification of specific chromosomes and chromosomal bands. GTG banding followed by FISH and propidium iodide fluorescence staining is commonly used. In this case, FISH is performed according to standard procedures, whereupon photomicrographs from FISH and GTG banding are compared to identify the chromosome(s) harboring the gene of interest. A second method involves FISH hybridization followed by R-banding using Hoechst dye. Chromosomes are stained with Hoechst 33258 dye for 15 min after FISH, UV-irradiated for 20 min at 365 nm, and then counterstained with propidium iodide. FISH signals and R-banding using propidium iodide or DAPI can be visualized simultaneously. The third method takes advantage of the newly described chromosome painting probes. A target chromosome can be readily identified using a specific painting probe with a different fluorochrome that issues from the gene to be mapped. Mapping of a new gene to the chromosomal region from which it was dissected is a crucial step before any further molecular analysis is done. The success of an *in situ* experiment depends on several critical factors. The factors that contribute significantly to the fluorescent signal include the specific activity of the probe used, its length, and the effective concentration of probe per slide. Furthermore, conditions for chromosome preparation, banding, and staining must be optimized accordingly. Often, chromosomes do not band well enough to identify individual chromosomes. Overstaining with the counterstaining dye may quench or diminish the hybridization signal. Distortion of chromosomal bands following the FISH procedure can be minimized by performing GTG banding first. Finally, the problem of background noise can be considerably alleviated by selecting narrow excitation band optical filters.

5
Laser-capture Microdissection (LCM)

Molecular examination of pathologically altered cells and tissues at the DNA, RNA, and protein level has revolutionized research and diagnostics in pathology. However, the inherent heterogeneity of primary tissues with an admixture of various reactive cell populations can affect the outcome and interpretation of molecular studies. Microdissection of tissue sections and cytological preparations is being increasingly used for the isolation of homogeneous, morphologically identified cell populations, thus overcoming the obstacle of tissue complexity.

In conjunction with sensitive analytical techniques, such as the polymerase chain reaction, microdissection has aided in the precise *in vivo* examination of cell populations, such as carcinoma *in situ* or the malignant cells of Hodgkin's disease, which are otherwise inaccessible for conventional molecular studies. To further implement this concept, the National Cancer Institute developed laser-capture microdissection. LCM is a technique that permits the isolation of morphologically

defined cell populations from specific microscopic regions in complex primary tissues using tunable laser beams. Once captured, the DNA, RNA or protein can be extracted from the isolated cells for molecular analyses. The minute samples may be analyzed by conventional PCR, reverse-transcription (RT)-PCR or polyacrylamide gel electrophoresis for specific macromolecular changes.

LCM is based on the selective adherence of visually targeted cells and tissue fragments to a thermoplastic polymer membrane coating [ethylene vinyl acetate (EVA)]. The membrane is attached to a rigid support and is placed in contact with the dehydrated tissue section. The system consists of an inverted microscope, a solid-state near-infrared laser diode, a laser control unit, and a joystick-controlled microscope stage with a vacuum chuck for slide immobilization, a CCD camera, and a color monitor. The LCM microscope may be connected to a personal computer for additional laser control and image acquisition. The thermoplastic membrane used for the transfer of selected cells has a diameter of approximately 6 mm and is mounted on an optically clear cap, which fits on a standard 0.5 mL microfuge tube for subsequent tissue processing. The cap is suspended on a mechanical transport arm and placed on the desired area of the dehydrated tissue section under standard pressure. After visual selection of the desired cells guided by the positioning beam, a selected cell cluster is precisely activated by a near-infrared laser pulse that results in focal melting of the EVA membrane. The absorption maximum of the membrane is near the wavelength of the laser. The melted polymer expands into the section and fills the hollow spaces present in the tissue. The polymer resolidifies within milliseconds and forms a composite with the tissue. The adherence of the tissue to the activated membrane exceeds the adhesion to the glass slides and allows selective removal of the desired cells. Laser impulses between 0.5 and 5 ms in duration can be repeated multiple times across the whole cap surface, which allows the rapid isolation of large numbers of cells. The selected tissue fragments are harvested by simple lifting of the membrane, which is then transferred to a microfuge tube containing the buffer solutions required for the isolation of the molecules of interest.

The minimum diameter of the laser beam of an LCM microscope is currently 7.5 μm, the maximum diameter 30 μm. Because most of the energy is absorbed by the membrane, the maximum temperatures reached by the tissue upon laser activation are in the range of 90 °C for several milliseconds, thus leaving biological macromolecules of interest intact.

After the first description of the initial NIH LCM, it was soon followed by the development of a commercially available LCM microscope. Currently, the second generation of this instrument (PixCell II; Arcturus Engineering, Mountain View, California, USA) illustrated in Fig. 6 is on the market. The technique has proven to be an effective routine microdissection technique for subsequent macromolecular analysis in many laboratories around the world. LCM has been applied to a wide range of cell and tissue preparations including paraffin wax–embedded material and fresh and frozen biopsies.

A prerequisite for this technique is the availability of histological dyes that do not interfere with downstream analysis of the sampled genetic material. Ehrig and colleagues have examined the effect of four histological nuclear dyes (methyl green, hematoxylin, toluidine blue O, azure B)

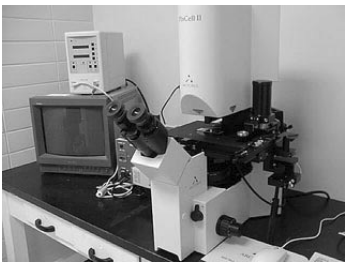

(a)

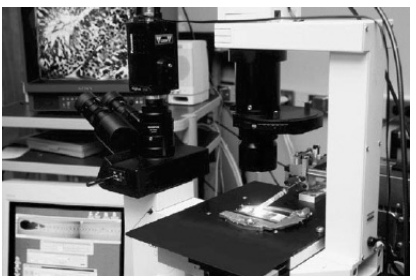

(b)

Fig. 6 (a) Arturus PixCell II Laser-capture Microdissection System that consists of an Olympus inverted microscope with 4×, 10×, 20×, and 40× objectives, near-infrared diode laser, user-selectable 7.5 mm, 15 mm, and 30 mm spot sizes, CCD color camera and monitor, image-archiving workstation. (b) Close-up at the laser beam to capture a single cell.

on TaqMan polymerase chain reaction amplification of genomic DNA prepared from fixed and frozen tissue for laser-capture microdissection. They concluded that DNA recovery from LCM-sampled tissue is independent of the histological stain chosen to highlight nuclear detail. The use of immunohistochemical stains allows the selection of cells according to phenotypic and functional characteristics. Depending on the starting material, DNA, good quality mRNA, and proteins can be extracted successfully from captured tissue fragments.

More recently, LCM has been refined to capture single cells in a specimen. The capture of individual cells with a new NIH LCM microscope epi-irradiates the EVA polymer overlying individual cells with 1-ms laser pulses focused to 6 µm. A computer-controlled arm precisely positions a 40-µm-wide strip of a cylindrical EVA surface onto a sample with a light contact force (ca. 0.1 g). The small contact force and contact area on the film on the sample diminishes nonspecific transfer to negligible levels.

With the advent of LCM, cDNA libraries are being developed from pure cells obtained directly from stained tissue, and microhybridization arrays of thousands of genes can now be used to examine gene expression in microdissected human tissue biopsies of various specific pathological lesions, especially malignant neoplasms. Similarly, the use of LCM is sufficient to isolate nanogram quantities of high-quality RNA. Together with the use of several internal standards and microcapillary electrophoresis of input RNA, two rounds of linear molecular amplification have been used to generate sufficient quantities of labeled target for hybridization to high-density oligonucleotide expression arrays. Results demonstrate that the technique is reproducible and generates only modest biasing of the original transcript population.

For further details regarding applications of LCM, the reader is encouraged to refer to the review article by Fend and Raffeld. It is an excellent source for determining the pros and cons of LCM versus conventional microdissection methods, and the limitation of technology. The review also discusses the current applications with references to DNA analysis, gene expression analysis, global

expression profiling, and single-cell applications. With regard to the technology itself, protocols, sample preparation, and signing up for training, visiting the following NIH website is recommended. (http://dir.nichd.nih.gov/lcm/LCM.HTM)

6 Applications of Chromosome and Tissue Microdissection

A recent three-year period literature review of the potential uses of DNA (RNA and protein) obtained by microdissection resulted in more than 600 reports on the application of chromosome and tissue microdissection to generate recombinant DNA clones from various human, mouse, rat, plant, and a number of other species. With the inclusion of LCM, there is a plethora of articles on the use of RNAs and proteins derived predominantly from human tissue archives and biopsies to delineate several diseases. The applications of microdissection can be grouped into eight categories.

1. Direct analysis of microdissected chromosomal DNA following PCR amplification.
2. Generation of recombinant DNA libraries from specific chromosomal regions.
3. Isolation of expressed genes localized to specific regions by probing cDNA libraries with amplified microdissected DNA from that region.
4. Study of chromosome abnormalities in cancer cells
5. Investigation of other specialized chromosomal structures
6. Gene transfer using chromosome fragments
7. Coupling microdissection with microarrays and proteomics technologies (DNA/protein microchips)
8. Incorporation of microdissection techniques in modern pathology

6.1 Direct Analysis of the PCR Product of Microdissected Chromosome Fragments

If the sequence of the gene of interest is known, gene-specific oligodeoxynucleotide primers can be synthesized and used in a PCR diagnostic for that gene. DNA from a microdissected fragment serves as a PCR template using gene-specific primers to localize (or exclude) a gene from that chromosome region. Also, if the spatial relationship of multiple genes in a specific chromosome region is known, target sequences can be amplified by PCR to detect structural alterations (deletions or rearrangements). In these experiments, a limited number of chromosome fragments from one region are dissected and target sequences are amplified using site-specific oligonucleotide primers. The inability to generate a PCR product using DNA from the affected region (in the presence of appropriate positive controls) is indicative of the absence or rearrangement of the gene in that region. Using this method, it was possible to map the breakpoints on chromosome 11 that are present in a number of translocations associated with human leukemia. When a translocation is identified by karyotype analysis but the chromosome of origin of the derivative chromosome is ambiguous, analysis of the microdissected translocation site may be helpful. One strategy involves the microdissection of the region of translocation, followed by *in vitro* amplification (PCR) using nonspecific primers. The PCR product is fluorescently labeled and hybridized to a normal karyotype to identify the chromosome that participated in the translocation. This method (called

micro-FISH) is rapid and may help identify complex karyotypes.

A new multicolor-banding technique has been developed that allows the differentiation of chromosome region–specific areas at the band level. The technique is based on the use of differently labeled overlapping microdissection libraries. The changing fluorescence intensity ratios along the chromosomes are used to assign different pseudocolors to specific chromosome regions. Weimer et al. have used that advantage and have further refined micro-FISH by combining the techniques of fluorescence *in situ* hybridization (FISH) and chromosomal microdissection in one experiment. Rearranged chromosomes were first identified by multicolor-FISH; the fluorophore-labeled derivative chromosomes were directly isolated by microdissection and reverse painted to identify the breakpoints. Thus, identification of the composition, origin, and breakpoints of rearranged chromosomes was rapidly accomplished.

6.2 Recombinant DNA Libraries Generated from Microdissected Chromosome Fragments

Recombinant DNA libraries can be generated from microdissected chromosomal fragments either by direct microcloning or by PCR-based methods as described earlier. Although methods in current use yield libraries that represent only a small fraction of unique sequences within the dissected region, these recombinant clones can be valuable as entry sites into poorly defined chromosomal regions. Also, they may be used as probes to identify and localize clones from genomic libraries that contain large DNA inserts. Chromosome microdissection has been used increasingly in genomic projects to "fill in" specific chromosome regions that are poorly mapped. Microdissection of a poorly defined chromosomal region, followed by cloning and sequencing unique sequence clones, could generate a pool of potential landmarks or tracts of DNA sequences 200 to 250 bp in length. Such entities, called *sequence-tagged sites* (STSs), are potentially useful in establishing a physical map (at the base sequence level) of human chromosomes. Chromosome microdissection may be useful in constructing "contig maps" (a series of clones representing overlapping regions of the genome). Microdissected DNA is cloned and amplified by PCR and used as a probe to screen genomic libraries that contain larger DNA inserts (e.g. cosmid, BAC or YAC libraries) under conditions that suppress repeat-sequence DNA.

6.3 Isolation of Transcribed Sequences Using Microdissected DNA as a Probe

Mapping expressed genes to specific chromosome regions can also be accomplished using radioisotopically labeled microdissected DNA as probes in hybridization experiments with appropriate cDNA libraries. This calls for sequential screening of cDNA libraries using plaque hybridization methods. More recently, methods such as the capture approach utilizing biotin-labeled probes bound to streptavidin with magnetic beads are useful. A large number of inheritable diseases and traits have been mapped to chromosome regions using genetic linkage methods. The isolation of causative genes can be rapidly accomplished by using chromosome microdissection to generate a library restricted to DNA sequences from the region associated with the disease. Once

clones that are within a million base pairs of the disease locus have been isolated, other molecular biological techniques may be used to isolate the gene of interest. As an example, chromosome microdissection was used to isolate a novel oncogene (eIF-5A2 initiation factor) from human ovarian cancer. In this disease amplification of 3q25–q26 constitutes one of the most frequent chromosomal alterations. By hybrid selection using 3q26 band-specific DNA, transcribed sequences (among which eIF-5A2) were isolated from a primary ovarian cancer containing high-copy number 3q26.

Table 1 lists a number of disease-related genes that were isolated by using microdissection and cloning methods.

Similar chromosome band–specific expression fragments were obtained from probe pools generated by microdissection of murine, bovine, swine, equine, and rice chromosomes. The pool DNA was used as probes to screen cDNA libraries of the respective species.

6.4 Study of Chromosome Abnormalities in Cancer Cells

Many nonrandom chromosome abnormalities are associated with specific cancer cell types. Microdissection may help advance our knowledge about the structure of such aberrant chromosomes and their behavior during mitosis, as well as delineate cytogenetic aberrations that cannot be identified by conventional chromosome analysis. Translocation sites may be microdissected directly and cloned to expedite the isolation of the causative chimeric (fusion) gene. Also, regions corresponding to a deletion site may be dissected from the normal chromosome. Clones may be used as probes in *in situ* hybridization experiments to delineate whether the

Tab. 1 Recombinant DNA libraries from microdissected chromosome regions associated with inheritable disorders in human.

Disease	*Chromosome region*
Adenomatous polyposis coli	5q21
Beckwith–Wiedmann syndrome	11p15.5–pter
Cystic fibrosis	7q
Down's syndrome	21q
Fragile X syndrome	Xq27–Xq28
Langer–Giedion syndrome	8q24.1
Meningioma	22q12–q13
Neurofibromatosis type II	22
Prader–Willi syndrome	15q11.1–12
Trichorhinophalangeal syndrome type I	8q23–qter
Wilms' tumor/anoridia	11p13
Waardenburg syndrome type I	2q33–qter
Multiple endocrine neoplasia type 2A	10p11.2–q21.1
Small-cell lung cancer and renal cell carcinoma	3p14–p21
Angelman syndrome	16p11.2
Ovarian cancer	3q25–q26
Acute nonlymphocytic leukemia	4p
Germ cell Tumors	12p

isolated sequences fall within the region encompassed by the deletion. For example, using a chromosome 9p–specific microdissection probe showed chromosome 9p anomaly resulting from two translocations [t(9; 18)(p12; p11.2) and t(1; 9)(p12; p12)] in polycythemia vera patients. Similarly, chromosome microdissection and reverse painting of aberrant chromosome 2 revealed a 2q33–q35 deletion in lung hypoplasia.

6.5 Investigation of Other Specialized Chromosomal Structures

Recombinant libraries from microdissected chromosomes can be very useful in genetic analysis of specialized chromosome structures, in genomic sequencing projects, and for the characterization of disease-related genetic loci. Specialized chromosome regions are particularly amenable to microdissection and cloning. These regions include common structures such centromeres and telomeres and the regions adjacent to these structures, short arms of acrocentric chromosomes, translocations, and fragile sites. Also, the structural basis of light and dark Giemsa banding as well as of the banding patterns observed with newer staining methods can be investigated.

In addition to chromosome band–specific libraries (probes), microdissection in combination with reverse-painting fluorescence *in situ* hybridization (FISH) has proven to be a very effective method to identify the origin of less common but important structures such as homogeneous staining regions (HSR), marker chromosomes, ring chromosomes, small accessory chromosomes (SACs), and directed isolation and mapping of microsatellite sequences.

Regions of gene amplification (homogeneous staining regions) are relatively large structures that can be microdissected easily for cloning and molecular analysis. For example, the chromosomal origin of HSR in neuroblastoma was identified as coamplification of chromosome bands 2p13 and 2p24 by chromosome microdissection and FISH. A microdissection-specific probe isolated from HSR in cells of transitional-cell carcinoma of the bladder revealed a new region (1p31–p32) that is amplified in primary tumors in addition to the previously known regions involving amplifications in chromosomes 6p22, 7p11, 9p23–pter.

In myelodysplastic syndromes (MDS), karyotypic aberrations identify subgroups of patients with distinct clinical–morphological features and can be relevant in risk assessment of developing leukemia. Often, conventional cytogenetic analysis is not sufficiently informative due to the presence of numerous unrecognizable markers. Microdissection of these markers followed by reverse-paint normal karyotypes may resolve questions regarding their origin in cases of MDS in patients previously treated for cancer. Ring chromosomes are another example of a rare constitutional abnormality with inconsistent phenotypic features. The ring chromosomes were characterized using microdissection in combination with degenerate nucleotide-primed polymerase chain reaction (PCR) and reverse painting (micro-FISH). This method made it possible to determine the chromosomal origin of the ring chromosomes in detail in several diseases and thus to compare the phenotypes of similar cases. Reverse painting of microdissected chromosome 6 probe onto normal metaphase spreads revealed a small deletion of the terminal region of 6q26qter in cases of acute

myelogenous leukemia. Microdissection of a ring chromosome in another AML-M5 case disclosed a (8 : 15) translocation with overrepresentation of 8q22. The origin of a ring chromosome in a myxoid malignant fibrous histiocytoma (MFH) was also determined by microdissection and FISH analyses. Hybridization of the microdissected labeled DNA to normal metaphase cells revealed that the signal localized only to 20q. Three signals were seen in the tumor cells using either the microdissected 20q probe or the chromosome 20 centromeric probe, indicating the involvement of both the long arm and the centromere in the ring chromosome. Small accessory chromosomes (SACs) are often associated with the centromeres of other chromosomes. The occurrence of more than two SACs in a single individual is extremely rare. Recently, a case of multiple SACs was reported in two different tissues of a moderately mentally retarded male. Microdissection combined with regular FISH demonstrated that SACs were derived from the centromeres of different chromosomes.

6.6 Gene Transfer Using Chromosome Fragments

Gene transfer into tissue culture cells or embryos has become an exceptionally useful method for the functional analysis of the regulatory and coding regions of genes. These methods are also used to induce phenotypic changes in cells as well as in whole organisms. The preferred vehicle for gene transfer and expression is a cloned gene (usually a cDNA with the appropriate flanking regulatory regions, promoters, and enhancers). Alternatively, whole genomic DNA has been transferred, usually with dominant selective markers. Gene-transfer experiments using microdissected fragments from a specific chromosome region would have several potential advantages. Transfer of large contiguous DNA stretches from microdissected chromosome fragments allows the gene to be presented "in context" – that is, with its own promoters, enhancers, control regions, and introns. These DNA sequences may be critical to the appropriate cell-type-specific expression of the transgene. The gene or locus of interest does not have to be identified, isolated, and cloned prior to gene transfer. This ability would be particularly advantageous in cases of a disease-related locus that was mapped genetically to a chromosome band by linkage analysis but was not further characterized. The transfer of genes into embryos through the mediation of chromosome fragments may be a useful tool in generating animal disease models and in confirming genetic linkage studies that map disease-causing genes to specific regions. To be transferred to an animal, however, the disease must behave as a positive dominant-acting genetic trait; it may not be associated with a structural deletion. An obvious disadvantage of chromosome fragment–mediated gene transfer is that only single-copy gene transfer is technically feasible. There are several reasons for this limitation. One would be the size and amount of DNA accepted by the recipient cell and the labor-intensiveness of the methodology. Two, physical manipulation of large stretches of microdissected DNA and microinjection would undoubtedly result in shearing of DNA into smaller fragments. Although this method will result in position-independent integration, the transferred DNA will be subject to DNA methylation, controlled by genotype-specific modifiers, as occurs with other transgenes. Richa and Lo reported the first

gene-transfer experiment in 1989. Human centromeric DNA was successfully transferred to fertilized mouse ova and was detected in blastocysts, 15-day embryos, and a mouse was derived from an injected ovum. Microinjection of insoluble or partially solubilized chromatin particles is likely to be an inefficient vehicle for gene transfer. A microscope stage such as that designed by Hagag and Viola would overcome this problem. The circular rotating stage accommodates a depression slide to collect chromosome fragments and a culture dish to contain fertilized ova (Fig. 2b). The stage permits the collection of a large number of fragments. The DNA is solubilized using trypsin or proteinase K in a small aqueous drop that is placed under oil. Soluble DNA is then used in microinjection of ova during the same setting.

6.7 Coupling Microdissection with Microarrays and Proteomics Technologies (DNA/Protein Microchips)

The microarray expression analysis system has become one of the most widely used functional genomic tools. The system has features that make it widely used for profiling mRNA expression. Microarrays are constructed by mechanically arraying PCR-amplified cDNA clones or genes representing the collection of genes to be assayed at a high density on derivatized glass microscope slides. A relatively simple *x-y-z* stage robotic system is used to spot an entire set of genes from a microbial genome or tens of thousands of eukaryotic cDNA clones. The microarrays are queried in cohybridization assays using two or more fluorescently labeled probes prepared from the cellular phenotype of interest.

Hybridization is assayed using a confocal laser scanner to measure fluorescence intensities, which allows simultaneous determination of the relative expression levels of all the genes represented in the array. Expressed sequence tags (EST) are single-pass, partial sequences of cDNA clones used for gene discovery, identification, and mapping in humans and other organisms. Generally, cDNA clones are selected to represent as many unique transcripts as possible. The two most widely used ESTs are UniGene at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/UniGene) and the Institute for Genomic Research (TIGR) Human Gene Index (http://www.tigr.org/tbd/hgi/hgi.html). Readers new to this technology may consult the excellent concise guide by Hegde et al. The article offers a discussion of protocols, instrumentation, and data analysis. Several microarray systems are commercially available in the market. The ScanArray 3000 (GSI Lumonics, Watertown, MA, USA) and Biorad VersArray (www.bio-rad.com/versarray) are examples. The affymetrix (Santa Clara, California) DNA chip is widely used.

A recent study has put to use the combination of microarrays and microdissection to elucidate the distinct molecular changes that occur at each progressive stage of prostate cancer. The cDNA generated from prostate cancer (tumor) and benign tissue (reference) samples were labeled with distinguishable fluorescent dyes and hybridized to a cDNA microarray consisting of known sequence motifs, each representing a gene of interest. The cDNA microarray was analyzed using a scanner, and fluorescence ratios were determined for each gene. Ratios were then imported into a database and analyzed using statistical algorithms. The data surveying method

assigns color intensity to ratios of gene expression: for example, shades of red represent genes that are upregulated in prostate cancer (1–4 fold increase); shades of green are downregulated (1–4 fold decrease); black, genes that are unchanged between tumor and benign tissue; gray, missing elements (not present). Selected genes were then validated at the transcript level using Northern blot or RT-PCR or at the protein level with immunoblots.

Similarly, the feasibility of combining LCM and cDNA microarray hybridization has been demonstrated in several reports showing reproducible differences in gene expression between normal and pathologic tissue samples. Furthermore, a combination of LCM, cDNA arrays, and real-time quantitative PCR was used to show altered gene expression patterns at various stages of breast cancer progression. Several expression libraries from purified cell populations derived from microdissected normal, preneoplastic, and neoplastic breast, prostate, ovary, lung, and liver tissues with large numbers of partially sequenced clones are already available at the CGAP website. (http://www.ncbi.nlm.nih.gov/ncicgap)

Knowledge of the gene sequence or the quantity of gene expression using DNA chips is not sufficient to predict the biological nature and function of a protein. The identification and quantitation of expressed proteins is more challenging. Proteomics is the global analysis of expressed proteins (including posttranslational modifications) and seeks to establish the relationship between genome sequence, expressed proteins, protein–protein interactions, and cell and tissue phenotype. The benefit of global protein analysis is immense because identification of distinctive protein signatures associated with cell function may provide novel therapeutic targets, molecular markers of disease, and increased understanding of determinants of cell phenotype.

Proteomics-based profiling uniquely allows delineation of global changes in protein expression patterns resulting from transcriptional and posttranscriptional control, posttranslational modifications, and shifts in proteins between different cellular compartments. Some of the current technologies for proteome profiling and the application of proteomics to the analysis of tumor tissues are being used concurrently with microdissection. Given that comprehensive expression profiles obtained using genomics and proteomics are highly complementary, a combined approach to profiling may well uncover expression patterns that could not be predicted using a single approach. Hanash used LCM and high-throughput microarrays/proteomics to study in parallel with Alzheimer's disease and showed that consistent with cDNA microarray evidence, the expression of proteins involved in synaptic activities was also altered in the brains of early Alzheimer's disease cases. The study demonstrated that protein chip arrays can be used to examine sequential changes in distinctive gene expression patterns in the brain as a function of the progression of Alzheimer's disease.

Currently, the most widely used proteomic tool for analysis of disease is two-dimensional protein gel electrophoresis (2-DE), which can display protein expression patterns to a high degree of resolution. 2D-PAGE is being complemented by new, exciting approaches to multiparametric protein characterization. As an alternative, an equally powerful technology is being employed to generate protein expression patterns from whole tissue extracts. The technology uses surface-enhanced

laser desorption/ionization (SELDI) that allows the retention of proteins on a solid-phase chromatographic surface (ProteinChip Array) with direct detection of retained proteins by time of flight-mass spectrometry (TOF-MS). The use of rapid, high-throughput mass spectrometric–based fingerprints of peptides and proteins may prove valuable for new molecular classification of human tumors and disease stages. Coupled with laser-capture microdissection, high-density protein arrays, antibody arrays, and small molecular arrays could have a substantial impact on proteomic profiling of human malignancies.

Differentially expressed proteins in normal and neoplastic tissues were recently analyzed by von Eggeling et al. by comparing the protein expression patterns generated using SELDI-based TOF-MS in cases of renal cell carcinoma, peripheral and central tumor tissue, as well as four microdissected cases of cervical intraepithelial neoplasia and three microdissected cases of cervix uteri carcinoma. By applying ProteinChip array technology, it was possible to investigate simultaneous and complex changes associated with tumor development and progression at the protein level in several cancers. In their landmark article, Petricoin et al. declared that the ultimate goal of proteomics is to translate the bench-side promise into bedside reality. It is to characterize the information flow through protein networks. This information can be a cause, or a consequence, of disease processes. Using cancer as a model disease, clinical proteomics applications are being developed to detect cancer earlier, to discover the next generation of targets and imaging biomarkers, and finally to tailor the therapy to the patient.

6.8 Incorporation of Microdissection Techniques in Modern Pathology

The last decade has witnessed a "molecular" revolution in pathology. Demonstrating that transcription of specific single genes or small gene sets and their protein products by *in situ* hybridization and immunocytochemistry has become routine in diagnostic and experimental pathology laboratories. A perhaps greater revolution is imminent with the application of more recently established and emergent technologies. These include new approaches to polymerase chain reaction (PCR); simultaneous studies of multiple genes and their expression using oligonucleotide and cDNA arrays; serial analysis of gene expression (SAGE); expressed sequence tag (EST) sequencing, subtractive cloning, and differential display; high-throughput sequencing; comparative genomic hybridization, multiplex fluorescence *in situ* hybridization (FISH) (spectral karyotyping); reverse chromosome painting; and knockout and transgenic organisms. Laser microdissection and micromachining, and new methods in bioinformatics, "data mining," and data visualization are added to the list. The use of any combination of these molecular methods will profoundly change diagnosis, prognosis, and treatment targeting oncology and will elucidate fundamental mechanisms of neoplastic transformation. Individual susceptibility to specific diseases will become assessable and screening will be refined.

Microdissection of chromosomal translocations, breakpoints, HSR, marker and ring chromosomes, micronuclei, and reverse-painting normal metaphase chromosomes have shed light on the origin of these aberrations. Table 2 presents a

Tab. 2 Disease-specific chromosomal regions mapped by microdissection methods.

Disease	*Chromosomal region*
Adenocarcinoma of the lung	9q, 16p
Acute myelogenous leukemia (ETO gene)	8q22
Autism/mental retardation	4q21
Basal cell carcinoma	9q22.3
Cervical carcinoma	6p21.3
Colorectal carcinoma	5q, 18q
Congenital heart defect	6p22.1–6pter
Ductal carcinoma breast	11q22–q23
Esophageal adenocarcinoma	4q32–33
Glioma	10q23.3
Melanoma	10q24–25
Ovarian cancer primary	3q25–q26
Ovarian cancer metastatic	3p25–p26
Pancreatic carcinoid	8p12–22
Prostate carcinoma	16q
Psychomotor/growth retardation	13q22–q32, 5q31.3q33.3
Small-cell lung carcinoma	3q13.2
Urinary bladder carcinoma (EGFR gene)	7p11–p12

sample of the numerous lesions/diseases for which chromosomal aberrations have been mapped by various microdissection techniques. In some instances, a specific gene has been identified as being potentially involved in the etiology of the disease. In other instances inactivation or absence of tumor suppressor genes are being sought as the cause of the disease.

How is functional genomics likely to impact on the pathology laboratory? Important aspects of functional genomics in the post genome era include the LCM, DNA and cDNA microarrays, proteomic methods, collaborative human tissue banking, tissue microarrays, and pathobioinformatics. The role of mass spectroscopy in the analysis of RNA, DNA, and protein has expanded. The recent combination of LCM and high-throughput cDNA microarrays provides a unique opportunity to generate gene expression profiles of cells from various stages of tumor progression as it occurs in the actual neoplastic tissue milieu. This combination was put to use to monitor *in vivo* gene expression levels in purified normal, invasive, and metastatic breast cell populations from a single patient. Expression profiles were verified by real-time quantitative PCR and immunohistochemistry and were proven to be generally applicable to the study of progression of malignancy. Rubin has used LCM, cDNA microarrays, and tissue microarrays to validate the presence of differentially expressed genes in various stages of prostate cancer. Pasinetti et al. conducted a large-scale gene expression profiling study using LCM and a number of tumors and their normal counterparts (e.g. tumor and normal oral epithelium). About 600 genes were found to be associated with oral cancer. These included oncogenes, tumor suppressor genes, transcription factors,

xenobiotic enzymes, metastatic proteins, differentiation markers, and genes that have not been implicated in oral cancer. A conglomerate of such genes may create a database that provides a verifiable global profile of gene expression during carcinogenesis.

7 Summary

Chromosome and tissue microdissection and microcloning technologies have proven to be powerful tools that made possible the isolation of new genes from chromosomal regions in several human genetic disorders. Three different techniques to perform chromosome microdissection and a number of alternatives to isolate and characterize microdissected DNA are presented. Also, the more recent technology of laser-capture microdissection is discussed. LCM has added to our ability to isolate DNA, RNA, and protein from a minute number of homogeneous cell populations. Advances in PCR and FISH methods have substantiated the use of microdissection in many applications. Indeed, the combination of these technologies has facilitated the isolation of several novel genes from infinitesimal amounts of DNA or tissue, a task that would not be possible using conventional methods. More recently, microdissection is being coupled with microspot technologies (DNA and protein microchips). These combinations will undoubtedly expedite the identification of disease-specific genes and markers via global expression profiling.

See also Gene Distribution in the Human Genome; Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Ausubel, F.M., Brent, R., Kingston, R.K., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987) *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Interscience, New York.

Celis, J.E., Graessmann, A., Loyter, A. (1986) *Microinjection and Organelle Transplantation Techniques: Methods and Applications*, Academic Press, London.

Hagag, N., Viola, M. (1993) *Microdissection and Cloning: A Practical Guide*, Academic Press, New York.

Hegde, P., Lee, N., Quackenbush, J. (2000) A concise guide to cDNA microarray analysis, *Biotechniques* **29**, 548–562.

Kornberg, R., Lorch, Y. (1992) Chromatin structure and transcription, *Annu. Rev. Cell Biol.* **8**, 563–587.

Nelson, W., Pienta, K.J., Barrock, E.R., Coffey, D.S. (1986) The role of the nuclear matrix in the organization and function of DNA, *Annu. Rev. Biophys. Chem.* **15**, 457–475.

Polak, J.M., McGee, J.O. (1990) *In Situ Hybridization: Principles and Practice*, Oxford University Press, New York.

Rooney, D.E., Czepulkowski, B.H. (1986) *Human Cytogenetics, A Practical Approach*, IRL Press, Oxford, Washington, DC.

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Vols. 1–3, Cold Spring Harbor Laboratory Press, Plainview, New York.

Sumner, A.T. (1990) *Chromosome Banding*, Unwin & Hyman, London.

Primary Literature

DeFonbrune, P. (1949) *Technique de Micromanipulation, Monographes de l'Institut Pasteur*, Masson, Paris.

Ehrig, T., Abdulkadir, S.A., Dintzis, S.M., Milbrandt, J., Watson, M.A. (2001) Quantitative amplification of genomic DNA from histological tissue sections after staining with nuclear dyes and laser capture microdissection, *J. Mol. Diagn.* **3**(1): 22–25.

Fend, F., Raffeld, M. (2000) Laser capture microdissection in pathology, *J. Clin. Pathol.* **53**(9), 666–672.

Hanash, S.M. (2001) Global profiling of gene expression in cancer using genomics and proteomics, *Curr. Opin. Mol. Ther.* **3**(6), 538–545.

Hedge, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J.E., Snesrud, E., Manualides, L., Chen, T.L. (1990) A unified model of eukaryotic chromosomes, *Cytometry* **11**, 8–25.

Monajembashi, S., Cremer, C., Cremer, T., Wolfrum, J., Greulich, K.O. (1986) Microdissection of human chromosomes by a laser microbeam, *Exp. Cell Res.* **167**, 262–265.

Petricoin, E.F., Zoon, K.C., Kohn, E.C., Barret, J.C., Liotta, L.A. (2002) Clinical proteomics: translating benchside promise into bedside reality, *Nat. Rev. Drug Discovery* **1**(9), 683–695.

Richa, J., Lo, C. (1989) Introduction of human DNA into mouse eggs by injection of dissected chromosome fragments, *Science* **245**, 175–177.

Rubin, M.A. (2001) Use of laser capture microdissection, cDNA microarrays, and tissue microarrays in advancing our understanding of prostate cancer, *J. Pathol.* **195**(1), 80–86.

Pasinetti, G.M., Ho, L. (2001) From cDNA microarrays to highthroughput proteomics, Restorative Neurol. *Neurosci.* **18**(2–3), 137–142.

Von Eggeling, F., Junker, K., Fiedel, W., Wollscheid, V., Durst, M., Calussen, U., Ernst, G. (2001) Mass spectrometry meets chip technology: a new proteomic tool in cancer research? *Electrophoresis* **22**(14), 2898–2902.

Weimer, J., Kiechle, M., Arnold, N. (2000) FISH- microdissection (FISH-MD) analysis of complex chromosome rearrangements, Cytogenet. *Cell Cytogenet.*, **88**(1–2), 114–118.

Circular Dichroism in Protein Analysis

Zhijing Dang and Jonathan D. Hirst
School of Chemistry, University of Nottingham, University Park, Nottingham, UK

1 Chirality and Polarized Light 57

2 Electronic Circular Dichroism 58
2.1 Far Ultraviolet 59
2.2 Near Ultraviolet 61

3 Vibrational Circular Dichroism 61

4 Estimation of Secondary Structure Content 62

5 Experimental Aspects 63
5.1 Instrumentation 63
5.2 Measurements 64
5.3 Solvents and Temperature Effects 65
5.4 Stopped-flow 65
5.5 Synchrotron Radiation 66
5.6 Fast Time-resolved Techniques 66

6 Spectral Characteristics of Elements of Protein Structures 67
6.1 α-helix 67
6.2 3_{10}-helix 68
6.3 β-sheet 68
6.4 β-turn 69
6.5 Random Coil 69
6.6 Polyproline Helices 70
6.7 Extrinsic Chromophores 70
6.8 Membrane Proteins 70

7 Conclusion and Outlook 71

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

Bibliography 71
Books and Reviews 71
Primary Literature 72

Keywords

Chiral
Possessing a nonsuperposable mirror image (like left and right hands).

Chromophore
Chemical moiety that absorbs light through excitation of one of its electrons.

Circular Dichroism
The differential absorption of left- and right-circularly polarized light.

Circularly Polarized Light
The resultant from combining two beams of linearly polarized light (*vide infra*) of equal intensity, but orthogonal polarizations, with a phase difference of a quarter of a wavelength.

Linearly Polarized Light
A beam of light in which the electric fields of all the photons oscillate in the same plane (also called *plane-polarized light*).

Optical Rotation
The rotation of plane-polarized light by a chiral compound (equivalent to circular birefringence).

Proteins are chiral objects. Elements of local, regular structure, called *secondary structure*, within proteins have a distinct handedness. For example, α-helices are mostly right-handed. Proteins absorb ultraviolet light, exciting electrons into high-energy states. Plane-polarized light (such as that created by Polaroid sunglasses) is light that propagates in only a single plane. It is possible to generate circularly polarized light. Left- and right-handed circularly polarized light may be considered as two components of plane-polarized light. When plane-polarized light passes through a chiral medium, its plane of polarization is rotated. If the light excites an electronic transition, a chiral molecule will absorb the left- and right-handed circularly polarized components differently. This differential absorption is termed *circular dichroism*. Circular dichroism spectroscopy may be used to estimate quantitatively the secondary structure content of proteins, giving the fraction of residues in α-helices, β-sheets, turns, and in coil conformations. It is widely used to characterize proteins under equilibrium conditions and to measure the kinetics of protein folding and unfolding.

1
Chirality and Polarized Light

At a molecular level, much of matter is asymmetrical, like our left and right hands, which are mirror images. In the world of organic chemistry, many compounds have chiral centers, typically carbon atoms bonded to four different atoms or groups. Chiral compounds have two forms (left- and right-handed), which cannot be superimposed on each other by rotation, although they have the same chemical composition. The asymmetric character of chiral compounds induces different interactions with electromagnetic radiation, which gives rise to optical activity.

A beam of natural light is a set of electromagnetic waves propagating in a single direction, but with all possible orientations perpendicular to the direction of propagation. When it passes through a linear polarizer, only a single orientation is selected and we get linearly polarized light. It may be divided into two chiral components of equal intensity, left- and right-circularly polarized light (Fig. 1). The electric vector of each circularly polarized component rotates along the direction of propagation. The overall effect of these two vectors is a planar sinusoidal wave.

Circular dichroism (CD) spectroscopy is the differential absorption of left- and right-circularly polarized light by a chiral medium. When linearly polarized light passes through achiral molecules, the absorption of left- and right-circularly polarized light is equal and no CD spectrum results. However, when the medium comprises chiral molecules, such as peptides, proteins, or DNA, there is differential absorption. The resultant of these two components is an ellipse with a specific optical rotational angle (Fig. 1). CD spectroscopy is based on the observation of optical rotation or circular birefringence, which can reflect detailed information about the three-dimensional conformation of biomolecules. The differential absorption arises because the left- and right-handed circularly polarized components experience different refractive indices on passing through chiral molecules.

The absorbance of light by a sample is given by the Beer–Lambert law:

$$A = \log\left(\frac{I_0}{I}\right) = \varepsilon c l \qquad (1)$$

where A is the absorption, c is the concentration in mol dm^{-3}, I_0 is the intensity of the incident light, l is the pathlength in cm, and ε is the molar absorption coefficient in units of dm^3 mol^{-1} cm^{-1}. The differential absorption is then

$$\Delta A = A_{\mathrm{L}} - A_{\mathrm{R}} = (\varepsilon_{\mathrm{L}} - \varepsilon_{\mathrm{R}})cl = \Delta\varepsilon cl \qquad (2)$$

where $\Delta\varepsilon$ is the difference in the molar absorption coefficients for left- and right-circularly polarized components. This may also be described, removing the linear dependence on pathlength and solute concentration, by

$$\Delta\varepsilon = \varepsilon_{\mathrm{L}} - \varepsilon_{\mathrm{R}} = \frac{\theta}{cl} \qquad (3)$$

where θ is the ellipticity in degrees. The ellipticity is equal to the ratio of the minor radius to the major radius of the now elliptically polarized light, that is A/B, as shown in Fig. 1, and it is related to the differential absorption by

$$\theta = 32.98\ \Delta A \qquad (4)$$

For a peptide or protein, the mean residue ellipticity in units of degree cm^2 dmol^{-1} is defined as

$$[\theta] = 100\,\frac{\theta}{cl} = 3298\,\Delta\varepsilon = \theta \times 0.1 \times \frac{MRW}{cl} \qquad (5)$$

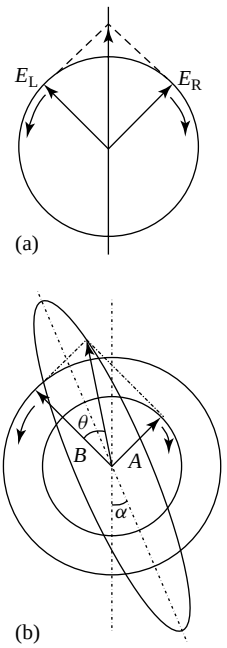

Fig. 1 (a) shows linearly polarized light with equally intense left- and right-polarized components (E_L and E_R). (b) Shows circularly polarized light with unequal left- and right-circularly polarized components. θ is the ellipticity and is the ratio of the minor axis to the major axis of the ellipse (A/B); α is called the *optical rotation*.

where c is the concentration in mg ml^{-1}; MRW is the mean residue weight, which is equal to molecular mass/$(N-1)$; N is the number of amino acid units and is typically one more than the number of peptide bonds, giving rise to the factor of $N-1$. The mean residue ellipticity is independent of the pathlength and concentration.

Although optical rotation is a dispersive measurement requiring the measurement over the complete spectrum, experimentally the measurements are confined to a relatively narrow wavelength range. In principle, CD can be measured with any frequency of electromagnetic radiation. However, in practice, most CD spectra are recorded in the ultraviolet–visible (UV–vis) region arising from electronic transitions. Since the early 1980s, new developments in optical and electronic technology have enabled optical activity measurements to be extended into the vibrational spectrum using both infrared and Raman techniques. We will focus on electronic CD, which is more commonly used, and will give only a brief description of vibrational optical activity.

2 Electronic Circular Dichroism

CD spectra arise from transitions that are electrically and magnetically allowed. When an electron of a molecule is excited from the ground state to a higher energy state, there is a change in the electric field of the radiation or the magnetic field or both. The change of the electric field causes a linear rearrangement of the electrons. The net linear displacement of charge during the transition is called the *electric transition dipole moment*, μ. The magnetic field change induces a circular rearrangement of electron density. The net circulation of charge is termed the *magnetic transition dipole moment*, m. In electronic absorption spectroscopy, bands arise from transitions that are either electrically allowed and magnetically forbidden or magnetically allowed and electrically forbidden. The oscillator strength, f, is proportional to the square of the electric transition dipole moment, μ^2, and corresponds to the area under a band in an electronic absorption spectrum. The rotational strength of a band in a CD spectrum is analogous

to the oscillator strength in electronic absorption spectroscopy. It is related to the product of the electric and magnetic transition moments and corresponds to the area under the band in the CD spectrum.

Electronic CD spectra may be divided into the far UV region (180–250 nm), the near UV region (250–300 nm), and the UV–vis region (300–700 nm). Absorption in the far UV region is dominated by the excitation of electrons on amide chromophores. It furnishes information on secondary structure (Fig. 2). In the near UV region, the CD arises from side-chain chromophores, such as aromatic groups and disulfide bonds. In this region, CD spectra can provide information on protein tertiary structure. The UV–vis region reflects the conformation and environment of the protein. Transitions in this region arise from chromophoric prosthetic groups, such as the heme group.

2.1 Far Ultraviolet

Peptide bonds are the key chromophores in proteins. To appreciate the origin of the CD of proteins, one needs to understand the electronic structure of the peptide bond. The valence shell of the peptide bond comprises four π electrons in three π orbitals (π_+, π_0, and the unoccupied π^*) and two electron pairs in two lone-pair orbitals (n and n′). The nπ^* and $\pi_0\pi^*$ electronic transitions of the peptide chromophore occur in the far UV region. The lowest energy transition is the nπ^* transition, which is polarized along the carbonyl bond. It is an excitation of an electron from a lone pair on the oxygen atom to an antibonding π orbital and occurs at 210 to 230 nm. It is magnetically allowed and electrically forbidden (Fig. 3). The transition energy is sensitive to the extent of hydrogen-bonding to the oxygen atom. In the lower symmetry of

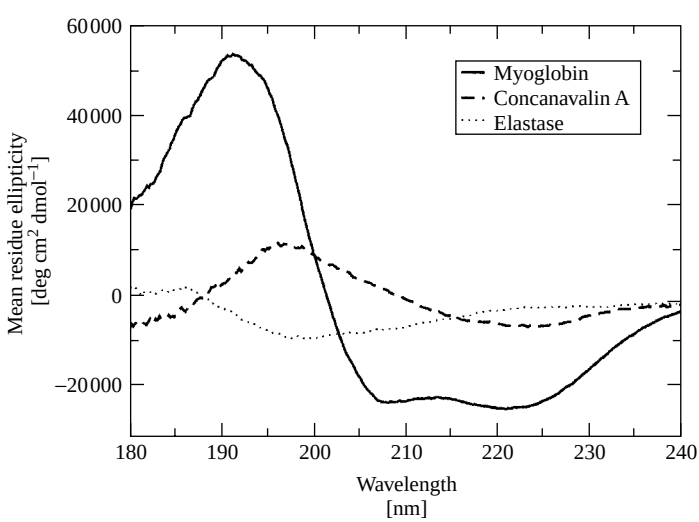

Fig. 2 The standard CD spectra for α-helix, β-sheet, and random coil, as typified by the proteins myoglobin, concanavalin A, and elastase respectively; the spectra are plotted from the freely available experimental data.

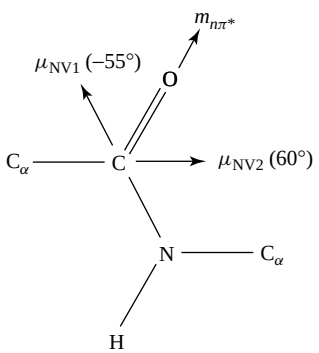

Fig. 3 The nπ^* and $\pi\pi^*$ transitions of the peptide chromophore in the far UV region. μ is the electrical transition dipole moment. *m* is the magnetic transition dipole moment.

a protein, the nπ^* transition is not strictly forbidden, but it is very weak. In solvents that lack hydrogen-bond donors, the nπ^* transition appears near 230 nm, while in aqueous solution, it is around 210 nm. The n$'\pi^*$ transition occurs at ~140 nm and is polarized perpendicular to the carbonyl bond.

The $\pi\pi^*$ transitions at 190 and 140 nm are electrically allowed. The first $\pi\pi^*$ transition, the NV_1 band, involves excitation of an electron from the nonbonding π orbital, π_0, to the π^* orbital. In secondary amides, the NV_1 band is observed around 185 to 190 nm, whereas in tertiary amides, it is closer to 200 nm. The second $\pi\pi^*$ transition, the NV_2 band, at 140 nm involves electronic excitation from the bonding π orbital, π_+, to the π^* orbital. The transition dipole moments of the two $\pi\pi^*$ transitions in secondary amides are approximately parallel (NV_1) and perpendicular (NV_2) to the C–N bond. The electric transition dipole moment of the $\pi_0\pi^*$ transition (NV_1) is about 3.1 D and that of the $\pi_+\pi^*$ transition (NV_2) is about 1.8 D. In an α-helix, the electric dipole coupling of the $\pi_0\pi^*$ (NV_1) transitions on neighboring residues results in two bands in the CD spectrum, one at ~190 nm and the other, a long-wavelength component, at ~208 nm. Interactions among the amide nπ^* and $\pi\pi^*$ transitions in a peptide are relatively delocalized and are significantly affected by environmental or local perturbations.

The arrangement of peptide bonds in a protein largely determines its CD spectrum in the far UV. The main-chain dihedral angles define the relative orientation of the backbone chromophores. If the dihedral angles for each amino acid residue were the same, the protein would form a perfectly regular helix (a β-strand may be viewed as a very narrow helix). Alternatively, an arbitrary set of (sterically allowed) dihedral angles would generate a random coil conformation. In reality, most secondary structures lie between these two extremes. The origins of the electronic CD of proteins can be understood by dividing the biopolymer into subunits, which do not have significant electronic exchange among them. This basic concept underlies theoretical calculations of protein CD spectra. Protein CD spectra arise from three types of interactions between $\pi\pi^*$ and nπ^* transitions of amides. Exciton interactions occur between degenerate or nearly degenerate $\pi\pi^*$ transitions on different peptide groups. An nπ^* transition and a $\pi\pi^*$ transition on different peptide groups may mix; this is also referred to as *μ–m coupling*. Finally, the nπ^* and $\pi\pi^*$

transitions on a single peptide group may mix – an interaction known as the *one-electron effect*. In an infinite helix, exciton interactions of $\pi\pi^*$ transitions on different peptide groups lead to three allowed transitions, one polarized parallel to the helix axis and two degenerate (equal energy) bands polarized perpendicular to the helix axis.

2.2 Near Ultraviolet

The near UV region, 250 to 300 nm, is also called the *aromatic region*. Because the peptide absorption in this area is at least an order of magnitude weaker than in the far UV region, a small number of aromatic chromophores have a distinct advantage. Bands in this region are due to $\pi\pi^*$ transitions from the aromatic side chains of histidine, phenylalanine, tyrosine, and tryptophan and $n\sigma^*$ transitions from disulfide groups. The indole of tryptophan has two or more transitions in the range of 240 to 290 nm with a combined maximum extinction coefficient $\varepsilon_{max}(279\ \text{nm}) \sim 5600\ \text{dm}^3\ \text{cm}^{-1}\ \text{mol}^{-1}$; tyrosine has one transition with $\varepsilon_{max}(274\ \text{nm}) \sim 1400\ \text{dm}^3\ \text{cm}^{-1}\ \text{mol}^{-1}$ and phenylalanine has one transition with $\varepsilon_{max}(258\ \text{nm}) \sim 190\ \text{dm}^3\ \text{cm}^{-1}\ \text{mol}^{-1}$. The cystine disulfide bond absorbs from 250 to 270 nm with $\varepsilon_{max} \sim 300\ \text{dm}^3\ \text{cm}^{-1}\ \text{mol}^{-1}$. The near UV CD also reflects the environment of aromatic groups. For example, the most red-shifted tyrosine feature should be due to tyrosine(s) in a highly polarizable environment, while the most blue-shifted are due to exposure to solvent. Although tryptophan has the most intense transition in this region, usually proteins have few tryptophans compared to the other aromatic groups, so the region tends not to be dominated by tryptophan transitions.

The disulfide group of cystine in many extracellular proteins has two well-characterized $n\sigma^*$ electronic transitions in the near UV region. The wavelengths depend on the dihedral angle of the disulfide band. This dihedral angle is normally approximately 90° in proteins, and the two $n\sigma^*$ transitions are degenerate, giving a single broad absorption band near 260 nm. As the dihedral angle deviates from 90°, the degeneracy is broken and one transition shifts to higher energy, with the other moving to lower energy. Disulfide contributions to a near UV CD spectrum are distinguishable from those of aromatic side chains, because the former are much broader.

3 Vibrational Circular Dichroism

Vibrational optical activity consists of two complementary spectroscopic areas, vibrational circular dichroism, and vibrational Raman optical activity. Both spectra show the differential response to left- versus right-circularly polarized radiation due to a vibrational transition in a chiral molecule. Vibrational circular dichroism (VCD) arises from simultaneous changes in the electric and magnetic dipole moments of the molecule due to the nuclear motion. The vibrational transition is between the g_0 and g_1 vibrational sublevels of the ground electronic state. The transition dipole coupling is highly dependent on peptide secondary structure. Because the interpretation of VCD depends on band shape, it is less susceptible to error from frequency shifts arising from nonstereochemical sources.

The amide I band (amide I′ in D_2O), due to primary C–O stretch, is the most

characteristic and the easiest to identify with VCD. Its frequency is relatively free of overlapping transitions from other parts of the molecule. An α-helix has a major positive peak at $\sim$1640 cm^{-1}. An antiparallel β-sheet has a split amide I′ band, with a major, sharp positive peak at $\sim$1615 to 1620 cm^{-1} and a weaker positive one at $\sim$1690 cm^{-1}. Random coil conformations give a broad positive peak at $\sim$1650 cm^{-1} and a positive trough at $\sim$1620 cm^{-1}, which is quite universal in band shape and intensity in most oligo- and polypeptides.

The amide II band is a mixture of N−H deformation and C−N stretch. The differences among α-helix, β-sheet, and random coil are less clear than those with the amide I band. The distinct feature of the amide II VCD for the α-helix is a dominant negative band; for β-sheet one observes a negative couplet and for random coil there is a weaker intensity positive band. The amide III band of an α-helix is mostly positive in the region of $\sim$1350 to 1250 cm^{-1}, and for β-sheet it is mostly negative in the same region. The amide III band is more sequence dependent, which affects its ultimate utility for peptide backbone conformational analyses.

VCD provides complementary data to electronic CD. The spectra are richer in information, but the instrumentation is not as developed or as widely used as that for electronic CD. The rotational strengths of vibrational transitions are much less intense than those of electronic transitions, which makes it a challenge to achieve the necessary signal/noise ratio. The technique has clear promise, but for the remainder of our discussion, we shall focus on electronic CD.

4 Estimation of Secondary Structure Content

The three-dimensional conformation of a protein can be specified by the dihedral angles of the backbone. Hydrogen-bonding patterns and particular values of dihedral angles lead to regular, local elements of structure, known as *secondary structure*. Secondary structures can be grouped into α-helices, β-sheet, β-turn, and random coil. As most peptides and proteins are mixtures of these four basic secondary structures, the overall mean residue ellipticity at a specific wavelength, $[\theta(\lambda)]$, is the sum of contributions from each secondary structure. It can be calculated by the equation:

$$[\theta(\lambda)] = f_\alpha[\theta(\lambda)]_\alpha + f_\beta[\theta(\lambda)]_\beta + f_\gamma[\theta(\lambda)]_\gamma + f_R[\theta(\lambda)]_R + \text{noise} \quad (6)$$

which is based on the fraction of each secondary structure type: α-helix (f_α), β-sheet (f_β), β-turn (f_γ), and random coil (f_R). As a first approximation, the CD spectrum of a protein is the sum of the appropriate percentages of each component spectrum, but the relative arrangements of structural units and motifs may also contribute to the observed CD spectrum.

There are many computer programs that can analyze CD spectra to estimate protein secondary structure content. The major methods are either algorithm-based or rely on databases. The former includes constrained least squares, normalized least squares, singular value deconvolution, parameterized fit and neural networks. Frequently used software packages are SELCON3, CONTIN, and Varslc. Database approaches are available in SELCON3, CONTIN, and SELCON. Most of these facilities can be found on the web site: http://www.cryst.bbk.ac.uk/cdweb.

The variable selection-self-consistent program SELCON provides analysis based on a database of 17 proteins. In SELCON, the proteins in the database are arranged in an order of increasing root-mean-square difference from the spectrum; the spectra least like the spectrum of interest are deleted to increase the speed. Then the program utilizes the observation that prediction improves when the protein is included in the basis set. The initial guess of the structure of the protein is made, and the secondary structure of the protein is then determined. The solution replaces the initial guess and the process is repeated until self-consistency is attained. This program gives a fairly accurate estimation of α-helix, β-sheet, and β-turn of globular proteins based on CD spectra between 240 and 200 nm. However, it overestimates α-helix and underestimates β-sheet when polypeptides have high β-sheet content.

The neural network program, K2D, uses a learnt mapping between spectra and secondary structure content to analyze new experimental spectra. It was trained on input data, which were CD spectra represented as 41 inputs corresponding to the intensities at the wavelengths between 240 and 200 nm. The output units formed a self-organized map, which could be analyzed to give fractional weights of protein secondary structures. In the training phase, the weighting of each wavelength is altered, until the error between the calculated and actual secondary structure content for the training data is minimized. For a new spectrum, the data are fed through the network using the adjusted weights to give an estimate of the secondary structure content. The K2D program predicts the secondary structure content fairly well, but does not give a matched spectrum. A different type of neural network was used, in which the output units gave directly the fractional secondary structure content: α-helical, parallel and antiparallel β-sheet, β-turn, and others. The network was trained on 13 spectra, each represented as 83 inputs corresponding to the intensities from 178 to 260 nm.

Each of these methods and programs has its advantages and disadvantages. Even when a method provides a good match between the calculated spectrum and the experimental data, it may not necessarily provide the best estimation of protein secondary structure content. Among these methods, SELCON may perform better than the others. Most of these programs usually estimate α-helical content well, but underestimate β-sheet and β-turn contents.

5 Experimental Aspects

5.1 Instrumentation

Nowadays, measuring CD is routine work in many laboratories. There are four manufacturers of CD instruments. Aviv and Associates (Lakewood, New Jersey) has Cary 61 and 62 CD instruments with a photoelastic modulator, modernized electronics, and a computer control system. JASCO (Easton, Maryland and Tokyo) makes J710 and J720 spectrometers. Jobin-Yvon (Longjumeau, France) manufactures the JY Mark VI and OLIS (On-Line Instrument Systems, Bogart, Georgia) offers the Cary 61 and 62 CD instruments. CD accessories for stopped-flow instruments are available from Applied Photophysics Ltd. (Leatherhead, UK) and Biologic (Grenoble, France).

Most instruments use a single beam. In Eq. (2), the intensity of the incident

light, I_0, does not appear for calculation of absorbance, so there is no need for a reference beam in measurements. However, a new double-beam instrument made by OLIS using sample and reference beams permits direct calculation of the CD signal. This technique reduces drift over time and factors such as lamp fluctuations are negated, since they affect both beams equally.

The light source is usually a 150 or 450 W xenon arc, which provides a continuous spectrum from the edge of the vacuum UV into the infrared region. Linearly polarized light from a monochromator passes through a modulating device, a photoelastic modulator, which converts the plane-polarized light to circularly polarized light, alternating between left- and right-circularly polarized light at the modulation frequency. In the detection end of the spectrometer, the light intensity is also modulated at the same frequency as the incident light.

A different modulation method, digital subtraction, is used by OLIS. The unique aspect of this method is that both of the two diverging orthogonally polarized beams are rotated by 180°. Then, the digitized signals from the two photomultiplier tubes are combined in such a way as to yield the CD signal directly and to cancel the contributions of noise. So, in this system, the absolute CD is measured and no calibration is required. However, one shortcoming is that the two beams pass through different parts of the sample and the cuvette; the significance of this is still uncertain.

The cooling system of the instruments is usually water or a narrow bore central heating system pump with car radiator cooling fluid. If it is possible, an air-cooled lamp may give better performance. Nitrogen flow is needed to avoid the creation of ozone, which reacts with surfaces of the mirrors. Nitrogen may also keep oxygen from entering the sample compartment, as it absorbs the incident light.

High quality quartz cuvettes, which transmit the full wavelength range of UV–vis light, are required for CD experiments. Cuvettes have rectangular and cylindrical shapes. A cylindrical cuvette has lower birefringence than a rectangular shaped one. However, rectangular cells are more often used especially in titration experiments. Different pathlength cuvettes are used for measurements at different wavelengths. Ten and 1 mm pathlengths cannot be used for wavelengths under 200 and 190 nm respectively. For shorter wavelengths, and for more strongly absorbing solvents at longer wavelengths, cells of 0.1 or 0.05 mm are necessary. For these cells, as the total volume is small, a higher concentration is demanded to keep absorbance of more than 1.0.

5.2 Measurements

For measurements in the far UV, solvent absorbance is not negligible, even with aqueous solutions. It is necessary to check the CD signal as a function of concentration to have a maximum absorbance of less than 1.5. If a linear relationship of absorbance as a function of concentration is not there, then there are solute–solute interactions in the sample. A baseline spectrum should be collected of solvent and buffer under the same conditions as the sample spectrum.

The spectrometer scans from longer to shorter wavelengths. Initially, the examining wavelength should be set 20 nm longer than that demanded by the experiment and it should be established that in this region

the absorbance is zero. Otherwise, it may be necessary to clean the cuvette and test again. The signal-to-noise ratio depends on many factors: it increases as $n^{1/2}$ (where n is the number of measurements), $I^{1/2}$ (where I is the intensity of light beam), and $\tau^{1/2}$ (where τ is the measuring time for each point). The CD instrument must be calibrated regularly using a standard sample. The most commonly used standard for visible and UV measurements is (+)-10-camphorsulfonic acid (CSA).

5.3 Solvents and Temperature Effects

The amount of sample required for CD experiments is relatively small compared to other absorption spectroscopic techniques. Water is the most frequently used solvent for CD measurements because of its excellent UV transparency. Fluorinated alcohols, such as 2,2,2-trifluoroethanol (TFE) and 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP), are good solvents as well. However, secondary structure, especially of peptides, can be influenced by solvents, for example, trifluoroethanol promotes α-helical conformations and nonpolar solvents, such as lipids and destabilize α-helices. MeCN is a good organic solvent. Phosphate buffers are also useful due to their low absorbance. However, some common salts and buffer components do significantly interfere with the far UV spectra and are best avoided.

Temperature may have a significant effect on CD spectra. When the free energy, difference between two conformations of a molecule, ΔG, is not too much higher than RT in the range of accessible temperatures (where R is the universal gas constant), a mixture of both conformers can be expected in solution. In this case, modifications of the experimental conditions can lead to a change in the measured rotational strength and can be calculated by

$$R_{\text{obs}} = \frac{R_1 + R_2 \exp(-\Delta G/RT)}{1 + \exp(-\Delta G/RT)} \qquad (7)$$

where R_1 and R_2 are the rotational strengths of these two conformers. Even when there is no a temperature-sensitive conformational change, there are temperature effects on CD spectra. For short helices, the temperature coefficient is small, but it increases significantly with helical length.

5.4 Stopped-flow

Stopped-flow CD is used for kinetic measurements of conformational transitions in proteins and in protein folding studies. It has been applied in the near UV and the far UV. The principle of stopped-flow is straightforward. The apparatus uses a drive motor to mix together rapidly two solutions, contained in separate drive syringes, into a mixing cell. The solutions then flow into the observation cell displacing the previous contents with freshly mixed reactants. A stop syringe is used to limit the volume of solution expended with each experiment and abruptly stops the flow. The flow of solution into the stop syringe causes the plunger to move back and trigger data collection. Then CD spectra may be measured as a function of time.

The primary objective of the stopped-flow experiments is to investigate transient changes in chiroptical properties associated with chemical, biochemical, and biophysical reactions, such as structural changes undergone during protein folding on a millisecond (ms) time scale and refolding. In many cases, secondary

structure is formed rapidly within 2 to 5 ms, and tertiary structure is formed slowly. The two-state folding of a small protein may occur within 3 ms. Stopped-flow measurements require relatively few combined shots to provide folding kinetics data. It can be made over a wide wavelength range, allowing the collection of time-resolved CD spectra. These can be analyzed to provide time-courses of the development of individual secondary structure types. However, the limitation of stopped-flow is the "dead-time" of the mixing process, which is typically more than 2 ms. In order to study faster folding kinetics, temperature-jump methods can be employed to shorten the observing time to 10 ~ 20 ns with laser-induced rapid heating (discussed in Sect. 5.6). In general, stopped-flow only provides average information of the secondary structure of protein; it cannot specify precisely where the secondary structure is formed.

5.5 Synchrotron Radiation

The X-ray source for measuring an X-ray absorption spectrum is synchrotron radiation. It covers all the wavelengths of the electromagnetic spectrum with an intensity over 100 times higher than the conventional X-ray tubes. When charged particles (electrons or positrons) from a linear accelerator with a speed close to that of light are injected into a storage ring under high vacuum, the magnetic fields from the strong bending magnets around the ring force the accelerated electrons to follow the ring in the curved sections. The high-energy particles hit the curved parts and lose part of their energy as synchrotron radiation, which is emitted tangential to these curved sections.

Synchrotron radiation CD has been developed for more than 20 years, but it has been used only recently in protein secondary and tertiary structure analysis. Synchrotron radiation CD provides higher photon fluxes of linearly polarized light particularly at short wavelengths than conventional sources. There is no need of a polarizer and the measuring limit extends to the cut-off of the photoelastic modulators (usually 120–130 nm) with much higher signal-to-noise ratio. So, CD spectra can be collected to a lower wavelength limit, allowing more accurate secondary structure content on folding motifs and a high-intensity light source may work well in the presence of salts, buffers, and detergents as well. Collection times for synchrotron radiation CD spectra can be considerably shorter than with conventional instruments.

5.6 Fast Time-resolved Techniques

In order to study protein folding on short time scales, fast time-resolved CD has been developed. It is now possible to resolve motions of chromophores well into the subpicosecond time scale. Time-resolved spectroscopy determines the spectral properties of a material at various instants of time after a stimulus has been applied. There are two main approaches: signal gating by sampling the output of a fast-responding photodetector and probing with a short-duration pulse of light. The limitations of the former are that few detectors operate at long wavelengths and the signal-to-noise ratio would be unacceptable. The pump-probe technique overcomes these limitations. It provides high temporal resolution of repeatable, photostimulated events.

Time-resolved CD can provide information on structural rearrangement and folding for both smaller chiral species and large macromolecules in different environments. Conventional time-resolved CD spectroscopy is limited by the time scale of the polarization modulator employed. Now, with shorter-pulse He–Ne lasers and a photoelastic modulator to alternate the pump laser polarizations without limiting the time response of the measurement, it is possible to detect the inherently weak difference between two large absorption signals, which are ellipsometric and are subject to significant optical artifacts. The limitation is the pulse length of the grating forming (pump) laser. Measurements can be made on microsecond, nanosecond, and picoseconds timescales. Although picosecond resolution is achievable, there are some technical problems, including intensity instability, polarization scrambling, and tunability limitations.

6 Spectral Characteristics of Elements of Protein Structures

CD is commonly used to probe the secondary conformations of proteins to study protein folding and to investigate interactions of proteins with small molecules such as achiral molecules whose induced CD is due to their interaction with the protein. Changes in CD can be used to provide evidence for conformational changes and to determine equilibrium constants. A few examples of conformational transitions are the α-helix to β-sheet conformational switch in the prion protein as the concentration of sodium dodecyl sulfate is altered; the unfolding of the coiled-coil structure of fibrinogen binding protein as temperature increases; the increasing helical content of troponin-C on binding Ca^{2+} and the binding of molybdate by the protein ModE, which gives large changes in the Trp signal at 292 nm. Far UV and near UV CD can be used to monitor protein unfolding upon the alteration of pH or the addition of a denaturant, such as urea or guanidinium chloride (GdmCl). CD can also be used to examine the interactions between multiple domains of proteins, for example, the flavin domain and the heme domain of flavocytochrome P450-BM3 give characteristic changes in the environment of the cofactors when they are linked in the intact enzyme.

6.1 α-helix

The right-handed α-helix (Fig. 4) has main-chain hydrogen bonds between amino acid residues and positions i and $i+4$ along the backbone chain. There is a 0.15-nm translation and a 100° rotation between two consecutive peptide units. Its helical pitch (the number of residues multiplied by the distance between α-carbon atoms on neighboring residues) is 0.54 nm. The α-helix has three distinctive bands in the far

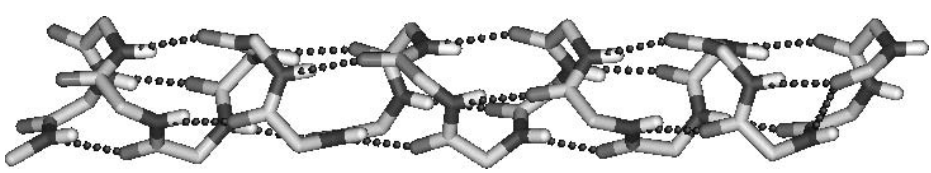

Fig. 4 The main chain of a polypeptide in an α-helical conformation, with hydrogen bonds denoted by dashed lines.

UV CD spectrum: two negative bands with separate minima of similar magnitude at 222 nm (the $n\pi^*$ transition) and 208 nm (the parallel $\pi\pi^*$ exciton band) and a positive band, the NV_1 band, at 190 nm due to the antiparallel $\pi\pi^*$ exciton band. The ratio $[\theta]_{208}/[\theta]_{222}$ is sensitive to the backbone dihedral angles.

The average fractional helicity, f_H, of an N-residue peptide may be related to the observed mean residue molar ellipticity at 222 nm by

$$f_H = \frac{[\theta]_{222}}{[\theta_{H\infty}]_{222}(1-k/N)} \qquad (8)$$

where $[\theta_{H\infty}]_{222}$ is the ellipticity of an infinite, completely helical peptide, k is an end-effect correction, which is ~ 3, and N is the number of residues. Estimated values for $[\theta]_{222}$ of an infinite, completely helical peptide range from $-37\,000$ to $-44\,000$ deg cm^2 dmol^{-1}. However, anomalously intense bands have been observed in some peptide models under certain experimental conditions.

6.2 3_{10}-helix

The 3_{10}-helix has $(i, i+3)$ main-chain hydrogen bonds. There are three representative dihedral angles (φ, ψ) for 3_{10}-helices, type I $(-60°, -30°)$; type II $(-54°, -28°)$, and type III $(-44°, -33°)$. The three types of 3_{10}-helices are anticipated to have different CD spectra based on first-principles calculations. The CD spectrum for a type I 3_{10}-helix is similar to that for an α-helix, but a little weaker band at 222 nm relative to the 208 nm band, and the $[\theta]_{222}/[\theta]_{208}$ ratio ranges from about 0.3 to 0.85; this ratio is greater than 1 for an α-helix. Type II has the strongest experimental foundation; type I and III have weaker and stronger $n\pi^*$ transitions than type II, respectively. It may be possible to distinguish reliably between 3_{10}-helix and α-helix by CD, but it is still a matter of debate.

6.3 β-sheet

β-sheets are composed of β-strands, which can be thought of as helices with two residues per turn. Typical backbone dihedral angles are $\phi = -120°$ and $\psi = +120°$ producing a translation of 3.2 to 3.4 Å per residue for residues in antiparallel and parallel strands respectively. An antiparallel arrangement is shown in Fig. 5. The CD spectra for β-sheet proteins are less intense and have fewer features than those of α-helical proteins, with a negative band near 217 nm, a positive band near 195 nm, and another negative band near 180 nm. However, the CD spectra for β-sheet proteins have much greater variability than those for α-helices: for example, the intensities

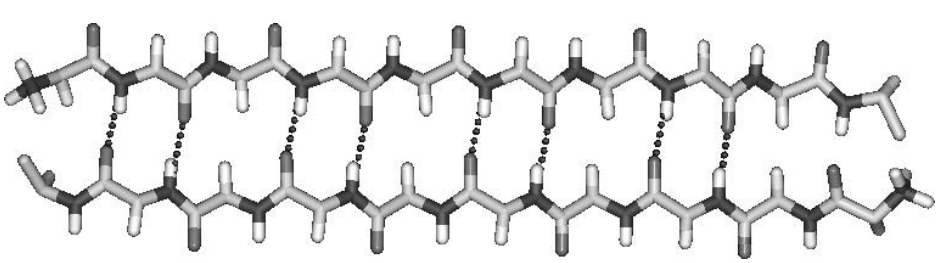

Fig. 5 The main chain of a polypeptide in an antiparallel two-stranded β-sheet, with hydrogen bonds depicted by dashed lines.

of the two long-wavelength bands, their ratio, and the wavelength of the positive band. These can be affected by side chains, solvents, and other environmental factors. Structural factors may also be important; β-sheets can be antiparallel, parallel, or mixed, intra- or intermolecular, and are also twisted to varying extents. Theoretical studies suggest that the extent of twisting of β-sheets is probably more important than the distinction between antiparallel and parallel sheets.

The wavelength at which the intensity of the CD spectrum changes sign is the most convenient criterion for differentiating parallel and antiparallel β-sheets. Crossover at ~178 nm indicates antiparallel β-sheet, at ~192 nm, parallel β-sheet, and at ~187 nm, a mixture of parallel and antiparallel β-sheets. The effects of twisting on β-sheet CD are predicted to include large differences in the magnitudes of the two long-wavelength bands, and a red-shift of both the positive maximum and the short-wavelength crossover from positive to negative CD.

Unlike α-helix content, it is not possible to estimate β-sheet content in proteins reliably from CD spectra because of the twisted and planar β-sheet showing quantitative differences between parallel and antiparallel β-sheet. One method is to calculate the difference between the CD at 217 nm and 195 nm. For 100% β-sheet, it is $50-55 \times 10^3$ deg cm^2 dmol^{-1}. This type of estimate may be reasonable, as long as the β-sheet is not strongly twisted.

6.4 *β*-turn

There are eight types of β-turn, three of which are common in proteins. The CD spectra of β-turn are rather varied. The type II β-turn has a strong negative band between 180 and 190 nm, a strong positive band between 200 and 205 nm, and a weak, red-shifted band at 225 nm due to the $n\pi^*$ transition. Generally, type II CD spectra are similar to β-sheet CD, except that the maximum is red-shifted by 5 to 10 nm. The CD spectra for type I and III turns are similar to that of the α-helix, with a negative $n\pi^*$ band and a negative $\pi\pi^*$ couplet.

6.5 Random Coil

CD spectroscopy is sensitive to precise protein conformation, but most random coils are very flexible. So, there are less common features in their CD spectra, except a strong negative band near 200 nm and a weak band at longer wavelengths, but the latter band may be either positive or negative. Poly(Glu) and poly(Lys) are frequently used models of unordered polypeptides. Their neutral aqueous solution spectra show two characteristic features: a strong negative band near 200 nm ($[\theta]_{max} \sim -40\,000$ deg cm^2 dmol^{-1}) and a significant band shift between 200 nm and 191 nm. When the pH is set to neutralize the side chains, or the ionic strength is increased by added salt, or the temperature is increased, the very weak negative band in the 235- to 240-nm region is observable. Urea and guanidinium chloride (GuCl) can have a complicated effect on CD spectra. At low concentrations, GuCl causes a decrease in the long-wavelength CD bands of the charged polypeptides by shielding the charges on the side chains and decreasing the electrostatic stabilization of the helical peptide.

There are two different explanations for unordered polypeptides CD spectra. In the conventional view, random coil polypeptides exist as an ensemble of an enormous

number of different conformers, with the relative geometry of nearest neighbors determined statistically. Another opinion is that fully ionized poly(Glu) and poly(Lys) at low temperature and low ionic strengths are in an extended helix conformation, which is similar to the left-handed 3_1 helix of poly(Pro)II helix. Theoretical calculations suggest that the parallel-polarized $\pi\pi^*$ transition occurs as a positive peak at 209 nm and that the perpendicular-polarized component is at 201 nm with a negative rotational strength. In general, random coil structures with different sequences and under different conditions can give a variety of CD curves. However, it seems likely that unordered polypeptides that have a positive band near 218 nm may have a substantial amount of poly(Pro)II helical conformation.

6.6 Polyproline Helices

The homopolymer poly(Pro) can adopt two different conformations. Poly(Pro) I is a right-handed helix with *cis*-peptide bonds throughout and it is stable only in relatively nonpolar solvents such as *n*-propanol. Poly(Pro)II is a left-handed 3_1-fold helix with trans residues. These two types of polyproline helices have different CD spectra. Poly(Pro)I has a strong positive band at ~215 nm and a slightly weaker negative at ~195 nm. Poly(Pro)II has a weak positive band at ~230 nm and a strong negative band at ~205 nm of ellipticity $\sim -60\,000$ deg cm^2 dmol^{-1}.

6.7 Extrinsic Chromophores

The CD spectra above 300 nm can be used to study transitions in chromophoric prosthetic groups, metal ions, inhibitors, or substrate analogues. Proteins lacking prosthetic groups do not exhibit absorption or CD bands above 300 nm, except possibly for the tail of a disulfide transition or a tryptophan band just above 300 nm. The CD spectra of disulfides have a characteristic tail stretching beyond 300 nm. There are three mechanisms contributing to the CD of chromophores bound to proteins. An extrinsic chromophore may be inherently chiral. There may be coupling between electronic transitions on the extrinsic chromophore and chromophores in the protein. It is also possible that transitions of differing symmetry on the extrinsic chromophore may mix because of the electrostatic field of the protein. The last two mechanisms depend on both the extrinsic chromophore geometry and also the structure of the protein.

6.8 Membrane Proteins

There are two kinds of membrane proteins: extrinsic (or peripheral) and intrinsic (or integral). Extrinsic membrane proteins may be removed from the membrane, or solubilized, by mild treatment, such as shaking with a dilute salt solution. Intrinsic membrane proteins cannot be removed from the membrane without treatment that destroys the membrane structure, such as dissolving it with detergent. Membrane proteins are difficult to study in NMR or X-ray crystallography, since the use of detergents may affect the conformation. So, CD spectroscopy is a useful method for investigating the structure of membrane proteins.

Membrane proteins tolerate mildly disruptive detergents without loss of activity. To analyze the CD spectra of membrane proteins requires the known structures

of membrane proteins (determined by X-ray diffraction), of which there are relatively few. Nevertheless, it is possible to estimate secondary structures content from CD spectra of membrane proteins using approaches such as convex constraint analysis. Five component spectra have been identified. They are two different types of α-helices (the α-helix in the soluble domain and the α_T-helix for transmembrane α-helix), β-sheet, β-turn, and unordered conformation. The characteristics of CD spectra of α-helix, β-sheet, β-turn, and random coil have been discussed above. The α_T-helix has a positive band in the range of 195 to 200 nm. The intensity of the 208-nm band is slightly more negative than that of the 222-nm band. α_T-helix has a larger rotational strength than an α-helix in aqueous conditions. Membrane proteins are immersed in a much lower dielectric medium than soluble proteins, which may explain some differences. Another important factor may be that the average chain length of an α_T-helix is 25 residues, which is about twice that of an α-helix in a soluble domain.

CD spectra can give much information about membrane proteins, such as secondary structure, fold motifs, conformational changes, environmental effects, folding, and insertion into membranes. However, there are potential artifacts of differential scattering, absorption flattening, and wavelength shifts, which may affect the CD spectra. CD spectra of membrane proteins often exhibit various degrees of distortions in shapes, intensities, and/or positions of the CD bands, and shifts in crossover points. There have been several experimental and theoretical approaches for differentiating the CD bands from the artifacts of the membrane protein CD spectra.

7 Conclusion and Outlook

CD spectroscopy is a useful technique for detecting protein secondary structure quickly and quantitatively. It is also useful in following protein folding/unfolding processes with stopped-flow or faster time-resolved techniques. The development of transient VCD may present a significant advance in quantifying the timescales of motions of biological systems. Currently, there is ongoing research to improve the accuracy of estimates of protein secondary structure content. Making such estimates more quantitative remains a challenge, as there are many factors that may affect the intensity and wavelength of CD spectra. Temperature and solvents are important external factors; dihedral angles and hydrogen-bonding patterns are key internal factors. All of these can change the relative orientation of $n\pi^*$ and $\pi\pi^*$ transitions, which in turn can lead to different CD spectra. In general, CD spectroscopy is a rapid technique, which is sensitive to protein secondary structure. More reliable analysis of CD spectra may emerge from ongoing experimental and theoretical studies.

See also Chirality in Biology.

Bibliography

Books and Reviews

Berova, N., Nakanishi, K., Woody, R.W. (1994) *Circular Dichroism: Principles and Applications*, VCH Publishers Inc., New York.

Fasman, G.D. (1996) *Circular Dichroism and the Conformational Analysis of Biomolecules*, Plenum Press, New York.

Greenfield, N.J. (1996) Methods to estimate the conformation of proteins and polypeptides

from circular dichroism data, *Anal. Biochem.* **235**, 1–10.

Rodger, A., Norden, B. (1997) *Circular Dichroism and Linear Dichroism*, Oxford University Press, Oxford.

Woody, R.W. (1995) Circular dichroism, *Methods Enzymol.* **246**, 34–71.

Primary Literature

Andrade, M.A., Chacon, P., Merelo, J.J., Moran, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network, *Protein Eng.* **6**, 383–390.

Besley, N.A., Hirst, J.D. (1999) Theoretical studies toward quantitative protein circular dichroism calculations, *J. Am. Chem. Soc.* **121**, 9636–9644.

Böhm, G., Muhr, R., Jaenicke, R. (1992) Quantitative analysis of protein for UV circular dichroism spectra by neural networks, *Protein Eng.* **5**, 191–195.

Brahms, S., Brahms, J. (1980) Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism, *J. Mol. Biol.* **138**, 149–178.

Chang, C.T., Wu, C.S.C., Yang, J.T. (1978) Circular dichroic analysis of protein conformation: inclusion of the beta-turns, *Anal. Biochem.* **91**, 13–31.

Chin, D.H., Woody, R.W., Rohl, C.A., Baldwin, R.L. (2002) Circular dichroism spectra of short, fixed-nucleus alanine helices, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15416–15421.

Compton, L.A., Johnson, W.C. Jr. (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication, *Anal. Biochem.* **155**, 155–167.

Dang, Z., Hirst, J.D. (2001) Short hydrogen bonds, circular dichroism, and over-estimates of peptide helicity, *Angew. Chem., Int. Ed.* **40**, 3619–3621.

Fasman, G.D. (1995) The measurement of transmembrane helices by the deconvolution of CD spectra of membrane proteins: a review, *Biopolymers* **37**, 339–362.

Fitts, D.D., Kirkwood, J.G. (1957) The optical rotatory dispersion of the α-helix, *Proc. Natl. Acad. Sci. U.S.A.* **43**, 1046–1052.

Holzwarth, G., Doty, P. (1965) The ultraviolet circular dichroism of polypeptides, *J. Am. Chem. Soc.* **87**, 218–228.

Johnson, W.C. Jr. (1990) Protein secondary structure and circular dichroism: a practical guide, *Proteins: Struct., Funct., Genet.* **7**, 205–214.

Lewis, J.W., Goldbeck, R.A., Kliger, D.S., Xie, X., Dunn, R.C., Simon, J.D. (1992) Time-resolved circular dichroism spectroscopy: experiment, theory, and applications to biological systems, *J. Phys. Chem.* **96**, 5243–5254.

Lobley, A., Whitmore, L., Wallace, B.A. (2002) Dichroweb: an interactive website for the analysis of protein secondary structure from circular dichroism spectra, *Bioinformatics* **18**, 211–212.

Luchins, J., Beychok, S. (1978) Far-ultraviolet stopped-flow circular dichroism, *Science* **199**, 425–426.

Luo, P., Baldwin, R.L. (1997) Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water, *Biochemistry* **36**, 8413–8421.

Manning, M.C., Illangasekare, M., Woody, R.W. (1988) Circular dichroism studies of distorted α-helices, twisted β-sheets, and β-turns, *Biophys. Chem.* **31**, 77–86.

Manning, M.C., Woody, R.W. (1991) Theoretical CD studies of polypeptide helices: examination of important electronic and geometric factors, *Biopolymers* **31**, 569–586.

Moffitt, W. (1956) Optical rotatory dispersion of helical polymers, *J. Chem. Phys.* **25**, 467–478.

Nitta, K., Segawa, T., Kuwajima, K., Sugai, S. (1977) Application of stopped-flow circular dichroism to the study of the unfolding of proteins, *Biopolymers* **16**, 703–706.

Pancoska, P., Bitto, E., Janota, V., Urbanova, M., Gupta, V.P., Keiderling, T.A. (1995) Comparison of and limits of accuracy for statistical analyses of vibrational and electronic circular dichroism spectra in terms of correlations to and predictions of protein secondary structure, *Protein Sci.* **4**, 1384–1401.

Pancoska, P., Keiderling, T.A. (1991) Systematic comparison of statistical analyses of electronic and vibrational circular dichroism for secondary structure prediction of selected proteins, *Biochemistry* **30**, 6885–6895.

Provencher, S.W., Glockner, J. (1981) Estimation of protein secondary structure from circular dichroism, *Biochemistry* **20**, 33–37.

Schellman, J.A. (1968) Symmetry rules for optical rotation, *Acc. Chem. Res.* **1**, 144–151.

Sreerema, N., Venyaminov, S.Y., Woody, R.W. (1999) Estimation of the number of helical and strand segments in proteins using CD spectroscopy, *Protein Sci.* **8**, 370–380.

Sreerema, N., Woody, R.W. (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism, *Anal. Biochem.* **209**, 32–44.

Sreerama, N., Woody, R.W. (2000) Estimation of protein secondary structure from CD spectra: comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set, *Anal. Biochem.* **282**, 252–260.

Toniolo, C., Polese, A., Formaggio, F., Crisma, M., Kamphuis, J. (1996) Circular dichroism spectrum of a peptide 3_{10}-helix, *J. Am. Chem. Soc.* **118**, 2744–2745.

Wallace, B.A., Janes, R.W. (2001) Synchrotron radiation circular dichroism spectroscopy of proteins: secondary structure, fold recognition and structural genomics, *Curr. Opin. Chem. Biol.* **5**, 567–571.

Wallimann, P., Kennedy, R.J., Kemp, D.S. (1999) Large circular dichroism ellipticities for N-templated helical polypeptides are inconsistent with currently accepted helicity algorithms, *Angew. Chem., Int. Ed.* **38**, 1290–1292.

Woody, R.W. (1992) Circular dichroism and conformation of unordered polypeptides, *Adv. Biophys. Chem.* **2**, 37–79.

Yang, J.T., Wu, C.S., Martinez, H.M. (1986) Calculation of protein conformation from circular dichroism, *Methods Enzymol.* **130**, 208–269.

Cloning of DNA Markers: *see* Genomic DNA Libraries, Construction and Applications

Combinatorial Phage Antibody Libraries

Michael B. Zwick, R. Anthony Williamson, and Dennis R. Burton
The Scripps Research Institute, La Jolla, CA, USA

1 Overview 76

2 Combinatorial Antibody Libraries 77
2.1 The Early Libraries: The Lambda Phage System 77
2.2 Constructing Antibody Libraries on the Surface of a Phage 79

3 Selection from Antibody Phage Libraries 84

4 Applications of Combinatorial Antibody Libraries 85
4.1 Antibodies from Immune Donors 85
4.1.1 Antibodies against Viruses and Prions 88
4.1.2 Antibodies against Tumor-associated Antigens 92
4.1.3 Intracellular Antibodies or "Intrabodies" 93
4.2 Antibodies in the Absence of Immunization 93
4.2.1 Naive Libraries 93
4.2.2 Semisynthetic Libraries 94

5 Summary 95

Bibliography 96
Books and Reviews 96
Primary Literature 96

Keywords

Antibody Repertoire
The array of different antibody molecules that may be synthesized by the immune system.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

Combinatorial Antibody Library
A large ensemble of antibody molecules expressed in a given vector, generally derived by the independent cloning of single heavy- and light-chain genes in the vector.

Panning
An affinity-based selection procedure for the isolation of an antibody phage with specificity for a desired antigen or epitope.

Phage Display
The expression of proteins or peptides on the surface of a filamentous bacteriophage.

Evolutionary pressure has driven the need for immune systems of considerable complexity and finesse. Indeed, to function efficiently, host immunity must be equal to the diversity and ingenuity of organisms that threaten the well-being of the animal. Antibody-mediated immune defense plays a pivotal role in this process. This article discusses the development and application of combinatorial antibody libraries, an approach to the generation of human monoclonal antibodies that began in the early 1990s and has generated increasing interest, both academic and commercial, ever since.

1 Overview

Antibody molecules are characterized by their ability to engage in high-affinity interactions of narrowly defined specificity. These properties make antibodies appealing as potential therapeutic agents. There are in fact several examples in which serum antibody preparations, derived from individuals with high serum titers against particular antigens or pathogens, have proven useful in the clinic. For example, the incidence of hemolytic disease of the newborn has decreased dramatically in countries where anti-rhesus D prophylaxis is practiced. There are, however, a number of problems inherent to hyperimmune serum that prevent its wider application. First, it is practically available only in limited supply. Second, because it is a blood product, hyperimmune serum cannot be guaranteed completely safe for the recipient. Outbreaks of hepatitis in Ireland and Eastern Europe, traced back to contaminated batches of anti-rhesus D hyperimmune sera, have underlined this concern. Most importantly, however, serum preparations are an extremely inefficient way of delivering useful antibodies. Even in the most favorable circumstances, only a small proportion of the antibody in such sera will possess the desired specificity.

The arrival of hybridoma technology presented the opportunity to produce large-scale quantities of rodent monoclonal antibodies of a desired specificity, but these antibodies elicit a strong immune response in humans and have proven to be of limited use for therapy. Moreover, the generation of human monoclonal

antibodies using hybridoma technology and Epstein–Barr virus transformation of human B-cells has met with only limited success, particularly when one considers the size of the available antibody repertoire in man. Consequently, much effort has been directed toward "humanizing" existing mouse monoclonal antibodies through grafting murine antigen-binding sequences (complementarity-determining regions) into a human antibody framework. This methodology is useful insofar as the immunogenicity of the final construct is considerably lower than that of the original murine molecule. Significantly though, extra manipulation of framework residues is often required to attain the affinity of the original mouse antibody, and the process is labor-intensive and time consuming.

Nevertheless, humanized monoclonal antibodies are being increasingly used in the clinic including in the treatment of breast cancer, lymphomas, Crohn's disease, rheumatoid arthritis, and graft rejection and in the prevention of respiratory syncytial virus (RSV) infection in premature infants. Combinatorial antibody libraries provide a rich source of diverse and fully human antibodies without the need for manipulation and so are expected to be used increasingly for generating antibodies for the clinic. Indeed, the first human antibody developed from phage libraries has recently been approved for human use.

2 Combinatorial Antibody Libraries

2.1 The Early Libraries: The Lambda Phage System

In the search for an alternative approach for generating human monoclonal antibodies, two developments were critical. The first was the demonstration that the antibody antigen-binding fragments Fab and Fv could be cloned and expressed in bacteria. To ensure correct folding and assembly of the antibody fragments (essential for retention of the binding characteristics of the parent molecule), the translated proteins were shepherded to the bacterial periplasm under the guidance of appropriate bacterial leader sequences. Second, and equally important to the emergence of combinatorial antibody libraries, was the use of the polymerase chain reaction (PCR) and a family of oligonucleotide primers to amplify antibody genes from a mixed population of antibody-producing cells. Taken together, these techniques meant that one could, by using an appropriate vector, express in bacteria the antigen-binding domains of an array of antibodies of any class and from any animal for which one had the antibody sequence for the construction of appropriate PCR primers: in effect, antibody repertoire cloning.

The first combinatorial antibody libraries were prepared from hyperimmunized mice and were constructed in the lytic bacteriophage lambda. Total RNA taken from mouse spleen was reverse-transcribed, and antibody heavy chains (Fd of IgG1, V_H and C_{H1} domains) and light chains (V_L and C_L domains) were PCR-amplified and independently cloned into modified lambda phage vectors. These heavy and light chain libraries were combined, following restriction digestion and religation, to create a single vector capable of expressing Fab fragments derived from the randomly scrambled antibody light and heavy chain genes. To identify Fab fragments specific for the immunizing antigen, the combinatorial library was screened using a filter lift plaque assay. Here, soluble Fabs, released during plaque

formation, were transferred to nitrocellulose filters. Subsequently, the filters were probed with radiolabeled antigen, and plaques formed by a phage encoding antigen-binding Fab were identified by autoradiography.

The successes of the mouse system were followed up by studies seeking to apply the methodology for cloning human antibodies (Fig. 1). The first human combinatorial Fabs were derived from a library prepared using peripheral blood lymphocytes (PBLs) of an individual recently boosted with tetanus toxoid. No specific Fabs, however, were obtained from a similar but nonboosted tissue donor, despite the high anti-tetanus toxoid serum titer in this individual. This failure underscores an important point in constructing human antibody libraries: in the absence of recent boosting, PBLs are a poor source of RNA from specific antibody-producing

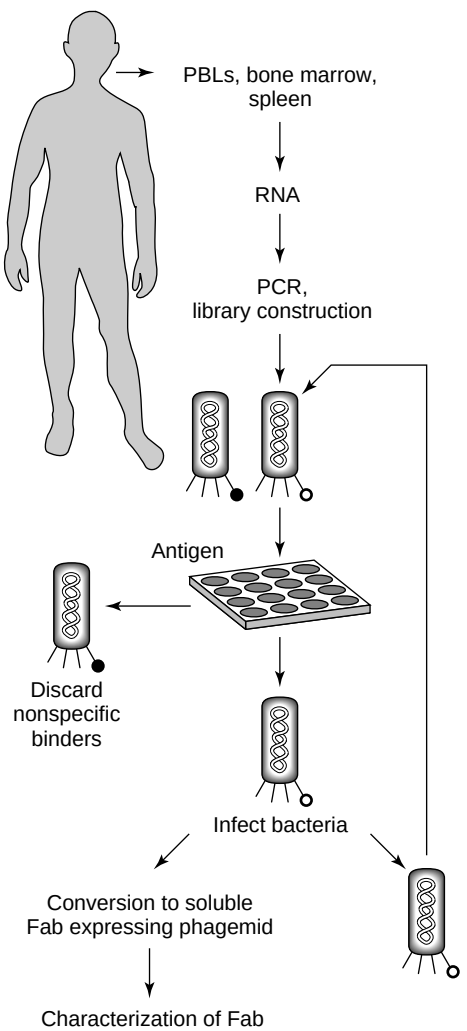

Fig. 1 Strategy for cloning human monoclonal Fab fragments from phage display combinatorial libraries. RNA prepared from antibody-producing tissue sources (e.g. bone marrow and peripheral blood) is reverse-transcribed, and the light chain and Fd portion of the heavy chain are amplified using PCR. The amplified genes are then cloned sequentially into a phagemid vector and "rescued" to a phage display library in which each phage expresses Fab on its surface. The library is then "panned" over an immobilized antigen, and specific phage Fabs are converted to a soluble Fab-expressing system for further characterization. [Adapted from Burton, D.R. and Barbas, C.F. (1993) *Chem. Immunol.* **56**, 112–126.]

plasma cells. In the resting state, plasma cells will be found in higher numbers in the bone marrow, and this is often a preferable starting point for library construction.

These early experiments illustrated the validity of the combinatorial approach. Despite the disruption of the original *in vivo* heavy–light chain combinations, the frequency of productive Fabs was higher than what may have been envisioned *a priori*. This result can be explained by chain promiscuity – the ability of a particular light or heavy chain to accept a number of different heavy or light chains respectively, while retaining antigen binding.

The lambda phage system, however, was hampered by its inefficient screening procedure. This method is laborious and restrictive of the number of clones that may be examined, which is a limiting factor when working with libraries prepared from sources other than a hyperimmunized one. Furthermore, screening is confined to antigens that are available in large quantities and in a purified form. What was needed was a library system in which positive clones could be affinity-selected by minute quantities of antigen. To address these problems, several groups developed systems to display antibody fragments on the surface of the Ff filamentous phage. We will concentrate here on the pComb3 phagemid vector system. Other systems are described in detail elsewhere.

2.2 Constructing Antibody Libraries on the Surface of a Phage

In vivo, antibody immunity is driven largely by selection. Particular B-cells may be triggered to replicate and differentiate on the basis of recognition of a given antigen by the antibody displayed on their surface. B-cells displaying an antibody that has no specificity for the same antigen will remain passive. In this manner, the immune system is able to home in on an antigenic determinant. Pioneering studies suggested a means by which selection from combinatorial libraries might approximate to native immunity. The work demonstrated ligand-driven selection, or biopanning, from peptide libraries displayed on the surface of a nonlytic filamentous phage. It was reasoned that monoclonal antibodies presented in a similar fashion might also be identified from within a library of many antibody phages by interaction with immobilized antigen. The general scheme for cloning monoclonal antibodies from combinatorial libraries on the surface of a phage is outlined in Fig. 1.

Many different fragments of antibodies have been displayed on phages including Fabs, single-chain Fv (scFv) fragments, and single-domain antibody fragments (including human and camelid antibodies). The absence of antibody constant regions in Fv fragments means that these constructs will not efficiently pair to form an antibody-binding pocket. Therefore, the variable domains must be joined by means of linker sequences using PCR primers. Single-chain antibodies have, however, been shown to reproduce the binding properties of the corresponding whole antibody with varying degrees of efficiency. Note that there are two different types of scFvs: "light chain first" and "heavy chain first," and the order can subtly affect the binding properties and stability of the scFv. Another feature of the scFv is that different linker lengths can affect the interaction between the heavy- and light-chain fragments. For example, very short linkers can restrict the heavy-chain fragment such that it will prefer to associate with the light chain of another scFv more readily than with its own light chain, thus producing a

dimer called a *diabody* or a trimer called a *triabody*.

Both Fab and scFv formats are commonly used, and there can be both advantages and disadvantages of an scFv form. The use of splicing overlap extension PCR in scFv library construction reduces the number of restriction enzymes and cloning steps, which may lead to better library diversity. In addition, scFvs are typically produced in much higher yields in *Escherichia coli* than their Fab counterparts. ScFvs may be advantageous when a smaller molecule is desirable, such as for probing narrow cavities. Finally, scFvs, with their tendency to oligomerize, may be exploited for multivalent interactions, although this feature is often a disadvantage when oligomerization is unwanted. Hence, Fabs may be preferable to scFvs if stability is an issue, or if the intention is eventually to generate a whole antibody molecule.

Structurally, the Ff (includes f1, fd, and M13) filamentous phage particle is composed of a closed circular single-stranded DNA genome encapsulated by a long thin protein coat. The majority of the phage particle is composed of approximately 2700 copies of major coat protein 8 (g8p or pVIII), which are distributed along the length of the virion. About five copies of a minor coat protein 3 (g3p or pIII) are arranged at the tail of the phage. The pIII has two "N-terminal domains" (N1 and N2) and one C-terminal domain (CT) that are separated by glycine-rich linkers. The N2 domain binds to F′ pili, facilitating infection of male *E. coli*, and N1N2 additionally serves to confer immunity from superinfection by other Ff phages. The membrane-bound CT domain serves to cap the protein to the trailing end of the phage filament. Of the other minor coat proteins, pVI is also present in about five copies with pIII at the tail of the phage, with about five copies each of pVII and pIX deployed at the head.

All the coat proteins on the filamentous phage have been used as fusions for displaying polypeptides on the phage coat, and, to date, all but pVI have been used to display antibody fragments (pVI displays polypeptides via its C-terminus and has been used primarily for cDNA libraries). Although antibody display by pVIII is possible, it has not gained widespread use (see below). An Fab library has been described in which the heavy and light chains are fused to pVII and pIX respectively, and an scFv library based on pIX has been successfully screened to yield novel antibodies. Display by pIX may have the advantage that because the antibody is displayed at the opposite end of the phage from pIII, infectivity is less likely to be affected. Nevertheless, the most common choice by far for antibody display is pIII, and pIII display has proven to be extremely robust.

Filamentous phage are nonlytic and, unlike their lytic counterparts, new virus particles are not formed in the cytoplasm of the host cell for release by lysis. Instead, virions are extruded through the outer membrane, leaving the bacterial cell intact. The extrusion process begins as the filamentous phage particle assembles on the inner bacterial membrane. Then, a pool of pIII molecules is secreted into the inner membrane for incorporation into an emerging phage particle. These pIII molecules are anchored in the membrane via their CT domains, with the remainder of the sequence extending into the periplasmic space. Of the coat proteins, only pIII and pVIII have leader sequences at their N-termini, which are cleaved off in the periplasm during phage assembly.

Drawing from the studies of antibody expression in bacteria described earlier, it seemed reasonable that an antibody fused to the N-terminus of pIII should be able to fold correctly in the environment of the periplasm. The tethered antibody would then be assimilated into the extruding virion and displayed on the surface.

In the earliest phage display experiments, peptides and proteins were fused to pIII by cloning directly the corresponding genes into the phage genome. Consequently, every copy of pIII on the phage surface carried the fusion. In the context of antibody display, the result is multivalent binding and a significant avidity effect, which dramatically hampers the efficiency of selection between antibody clones of different affinities. Thus, monovalent display is preferable for selecting clones with the highest affinity. However, multivalent display of antibody has proved useful in selecting antibodies against a number of targets including those on the cell surface wherein the phage becomes internalized much more efficiently.

To regulate the number of fusion molecules displayed per phage particle, phagemid systems have been developed. Phagemids are simply plasmids containing filamentous phage regulatory elements (typically f1 is used). This sequence contains a filamentous phage DNA origin of replication and a DNA packaging signal. Upon coinfection of a cell with an M13 helper phage, the regulatory element of the phagemid sequesters the replicative machinery provided by the helper phage to package single-stranded phagemid DNA into an infectious phage particle. By introducing into the phagemid a segment of DNA encoding the pIII fusion, subsequent superinfection with the helper phage results in the production of native pIII encoded by the helper phage and the pIII fusion encoded by the phagemid. Tagging a bacterial leader sequence to the front of the pIII fusion serves to direct the translated protein to the bacterial inner membrane. A competition is therefore established between these two forms of the coat protein for incorporation into newly formed phage particles. The resultant phages generally carry one copy of the fusion.

Phagemid vectors have also been constructed to generate antibody fusions with pVIII. This coat protein is present in greater abundance than pIII and covers the entire length of the phage particle; display of up to 24 copies of Fab fused to pVIII has been described. However, there are significant limitations in the size of the protein that can be displayed by pVIII, which are presumably imposed by steric constraints during the phage assembly process. Thus, certain antibodies may not get displayed by pVIII, or the number of copies of antibody per phage may vary considerably leading to poor discrimination of antibodies of high and low affinity. The limits of pVIII display and the associated lack of control over antibody valence during selection have made pVIII display a less favored choice for antibody library construction.

The first vector system developed to display Fab fragments on the surface of filamentous phages via pIII fusion was pComb3. Similar but improved versions of this vector, pComb3H and pComb3X, have since been described, and their use in constructing combinatorial antibody libraries will be examined here by way of example. During the construction of an antibody library, the light and heavy (Fd) chains are cloned sequentially into the vector, effectively randomizing the light and Fd pairings of the recombinant Fabs. In pComb3H, both antibody chains are under the polycistronic control of a single

*lac*Z promoter. The heavy chain possesses a *pel*B leader sequence, whereas the light chain has an *omp*A leader sequence to direct secretion into the periplasmic space in *E. coli*. The Fd chain is fused with a sequence encoding the CT domain of pIII. A five–amino acid tether (GGGGS) is positioned between the Fd and the pIII to minimize interaction. The Fd fragment is thus cotranslated with the pIII CT domain and, following translocation to the periplasm, the fusion protein is anchored to the inner membrane via the pIII domain, with the antibody projecting into the periplasm (Fig. 2). The light chain is then able, given the redox potential of the periplasmic environment, to assemble on the heavy chain to yield a functional Fab fragment. The inclusion of an f1 intergenic region in the vector permits superinfection with an M13 helper phage to be followed by the packaging of single-stranded, phagemid-carrying genes encoding heavy- and light-chain sequences (Fig. 3). Competition between native pIII, comprising the N-terminal domain essential for virion infectivity, and the Fab–pIII fusion yields a phage particle carrying native pIII and monomerically displaying Fab at the head of the filamentous virion. The pComb3X vector is nearly identical to pComb3H except that a hexahistidine and a hemagglutinin (HA) tag are fused to the heavy chain, followed by an amber stop codon and the *gene3* sequence, such that soluble Fab or scFv can be produced readily in nonsuppressor strains of *E. coli* without the need for excising

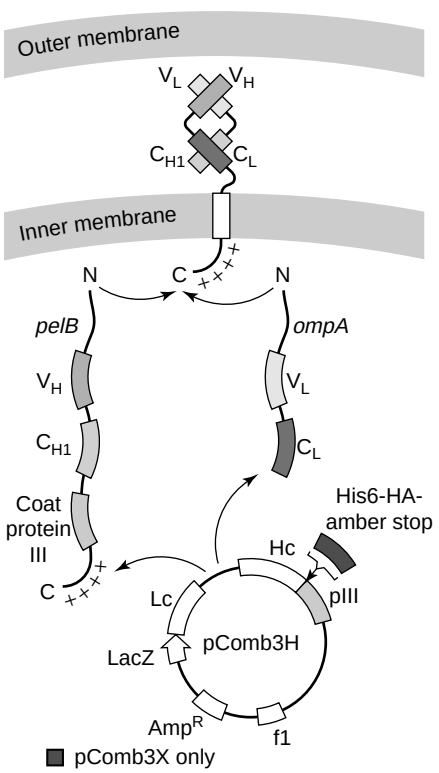

Fig. 2 The proposed pathway of Fab assembly in the bacterial periplasm. Polycistronic expression of cloned light chain (Lc) and Fd (Hc, heavy chain)/pIII fusion is controlled by a single *lac* promoter/operator sequence in pComb3H and pComb3X. Both vectors are improved versions of the original pComb3 vector and have enhanced stability. The pComb3X vector is nearly identical to pComb3H except that a hexahistidine and HA tag followed by an amber stop codon have been inserted in-frame between Fd and *gene3*, allowing for facile detection of the cloned single chain Fvs. The translated antibody is directed to the periplasmic space by *omp*A (Lc) and *pel*B (Hc) leader sequences. Here, the light chain assembles onto the heavy-chain template, which is anchored to the inner membrane by the pIII fusion. (From Barbas et al. (1991) *Methods: A Companion to Methods in Enzymology*, Vol. 2, Academic Press, Orlando, FL, pp. 119–124.)

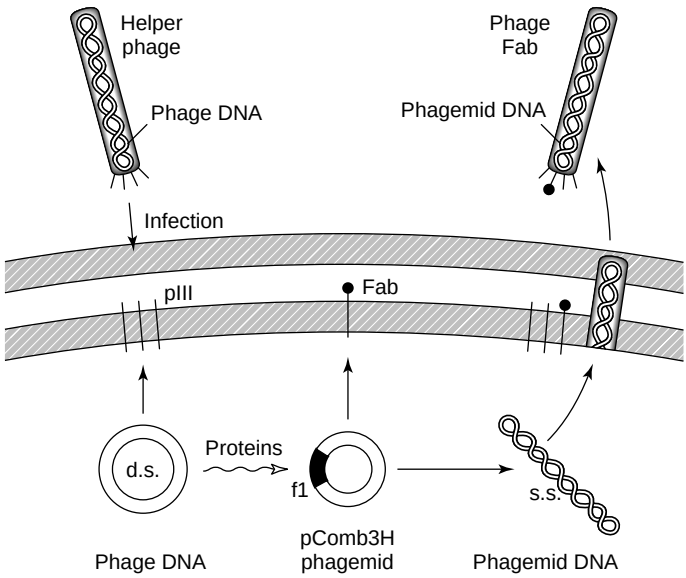

Fig. 3 Helper phage rescue of pComb3H phagemid yielding a phage display library. Incorporation of the f1 origin of replication into the pComb3H phagemid means that upon coinfection with an M13 helper phage, single-stranded pComb3H DNA will be packaged and extruded from the bacterial cell as an assembled phage. As the assembling phage is extruded from the cell, it is "capped" by both the native pIII (provided by the helper phage) and the pIII–Fab complex (derived from pComb3H). The resulting phage has Fab displayed primarily monovalently on its surface and contains the corresponding antibody genes inside. [Adapted from Burton, D.R. and Barbas, C.F. (1993) *Immunomethods*, **3**, 155–163.]

gene3. We note that in some phage display systems, the entire *gene3* sequence is incorporated rather than just the CT domain, in the phagemid fusion. Since one function of the amino-terminal half of pIII is to prevent superinfection of *E. coli* by other filamentous phages, in this approach the expression of the phagemid must be halted for helper phage infection and induced for phage display, thus complicating the selection procedure.

Some antibody fragments are poorly displayed on phages. To increase the display of such antibodies on the surface of the phage, some creative strategies have been devised. For example, overexpressing certain chaperones has facilitated increased display and stability of poorly folding antibody fragments. In another example, a random-mutagenesis-selection approach resulted in increased display of antibodies and other polypeptides, the success of which could either be traced to beneficial mutations in the antibody itself or to the leader peptide sequence. Finally, a general strategy for increasing antibody display on pIII has been simply to use a specially designed helper phage that has a defective *gene3*, such that oligovalent display of phagemid-encoded antibody is enforced.

Some alternatives to phage display have been developed in recent years, and two of them deserve particular mention. Yeast

has been used successfully to display scFvs on their surface. Potential advantages of yeast display are that flow cytometry can be used to sort the positive and negative binding clones, and equilibrium binding experiments have shown a good correlation between binding affinity and selection allowing fine discrimination between clones of similar affinity. Ribosome display has also been used to select antibodies in a totally cell-free system. The potential advantages of ribosome display are the large and diverse repertoires that can be achieved (up to $\sim 10^{14}$) and flexibility in selection conditions. The principles of selection for virtually all combinatorial antibody libraries are similar, so because phage display is the most common mode of display, we will next consider the selection from antibody phage libraries.

3 Selection from Antibody Phage Libraries

The strength of phage display lies in the ability to select positive clones occurring in a library at frequencies of one in a million or less. This selection takes place over consecutive rounds of panning in which positive phages are progressively enriched against a background of unwanted clones. In effect, panning is performed by affinity purification of an antibody phage against an antigen immobilized on a solid surface such as a column matrix or a well in an enzyme-linked immunosorbent assay (ELISA) setup. Proteins on the surface of whole cells may also be panned in this manner in tissue-culture plates or using flow cytometry. Phages may also be bound to biotinylated antigen in solution and then captured on streptavidin-coated magnetic beads. In each case, nonantigen-bound phages are removed by thorough washing and bound phages are recovered by elution from the antigen, generally with acid. Eluted phages are infected into *E. coli*, amplified, and reselected in further rounds of panning (see Fig. 4). When monovalent antibody display is used, typically three to five rounds of panning will select a population of clones with the highest affinity for antigen.

The panning process is unlike the antibody response to antigen *in vivo* in which mutation and recombination occur simultaneously with selection. Thus, an antibody library can be screened by alternating antigens in successive rounds to select for cross-reactive antibodies without worrying about "eliciting" *de novo* antibodies to each new antigen. Moreover, it is possible to pan with mutant proteins, unfolded proteins and peptide fragments that are either unsuitable for immunization or are simply not immunogenic. Since panning is not limited by a requirement for purified antigen, phage libraries may be selected against an enormous range of molecules. Panning experiments performed using lysates of virus-infected cells and intact cells indicate that the technique is sensitive to antigen present in small amounts. Different panning strategies are detailed in Section 4.

Antibody clones isolated at the phage level from the final round of panning need to be examined as soluble Fab to further investigate their binding properties. In the pComb3H vector, the transition from phage display to soluble expression is made after removal of the g3 sequence by restriction digestion and religation of the vector. Soluble Fab, which accumulates in the periplasmic space, is released following lysis of the bacteria. If necessary, recombinant Fab may be purified to

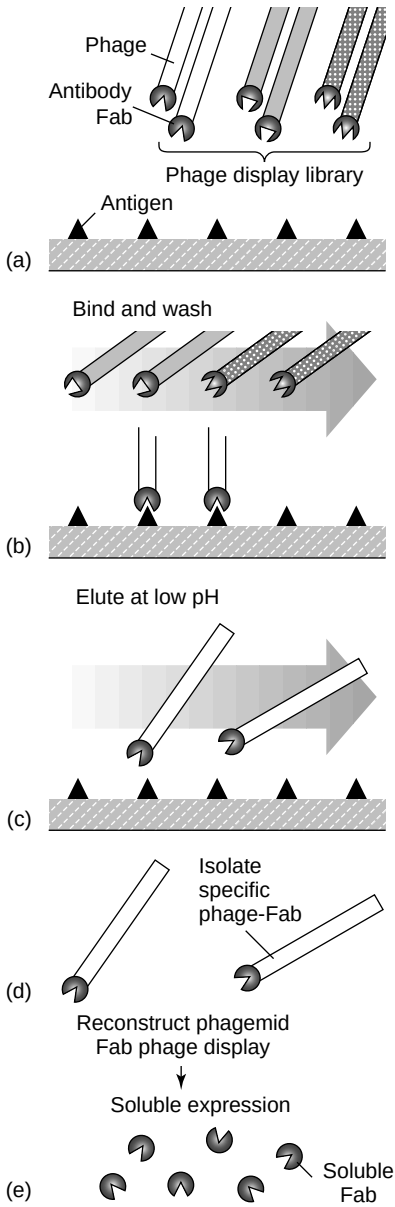

Fig. 4 Panning of the combinatorial phage display library. Phages are incubated with immobilized antigen (a), and unbound phages are removed by washing (b). The remaining phages are eluted in a low pH solution or in the presence of excess soluble antigen (c, d). Eluted phages are then reinfected into bacteria, reamplified, and reapplied to the antigen. Three or four repetitions of steps (a) to (d) result in a progressive enrichment for antigen-specific phage-Fab. Phages from (d) are converted to the phagemid form of the vector, the DNA prepared, and the *gene3* sequence excised. Religation then yields a reconstructed phagemid, which may be used to transform bacteria for the production of soluble Fab (e). [Adapted from Burton, D.R. (1992) *Hosp. Prac.* **27**, 67–74.]

homogeneity from bacterial supernatants in a single step by passage over an anti-Fab affinity column.

For many applications of antibodies, whole immunoglobin molecules are preferred. To this end, a number of vector systems have been assembled, which facilitate rapid subcloning from phage display vectors to eukaryotic expression systems. This presents the opportunity for the investigator to create whole antibodies of any isotype from a phage-selected Fab by joining variable region sequences to the appropriate constant regions contained within the eukaryotic vector.

4 Applications of Combinatorial Antibody Libraries

4.1 Antibodies from Immune Donors

Recombinant antibodies isolated from pComb3-based combinatorial libraries were first prepared from antigenically boosted mice and humans. Specific Fabs have also been generated from humans who have not been actively boosted but who have at some time had contact with

the antigen (e.g. an infectious agent). Preparing combinatorial antibody libraries from immune individuals is advantageous in that the antibody gene repertoires of these donors have been biased in favor of antibodies binding to particular antigens in the course of a natural immune response. In addition, if the subject has been challenged with the antigen on multiple occasions, the available antibodies will probably be of high affinity following rounds of affinity maturation. Combinatorial antibody libraries prepared from a single immune donor can provide a glimpse or a rough "snapshot" of an individual's antibody repertoire. This allows a unique opportunity to examine the human humoral immune response to any number of pathogens at the monoclonal level. However, one large caveat should be borne in mind. Library construction shuffles heavy and light chains so that in theory no knowledge of heavy–light chain pairing *in vivo* is available. In practice, many heavy chains may be combined with light chains related to those to which it was paired *in vivo,* but this is extremely difficult to prove for any given case.

One application has been to obtain monoclonal antibodies with antiviral activity, in particular, those antibodies that are best able to neutralize viruses by binding to their surface and diminishing or eliminating their ability to infect host cells. (There are additional and very important effector functions of antibodies, such as their ability to recruit complement and effector cells, but this topic is outside the scope of this review.) In the next section, we will focus on antibody-mediated neutralization, with the acknowledgement that effector functions of antibodies have been found to play a critical role in fighting infectious diseases. Indeed, all neutralizing antibodies (of the appropriate isotype) have effector function capability.

Another potential application of antibody immune libraries is to attempt to understand what deficiencies might exist in the antibody response against certain pathogens, or the nature of immune pathologies in certain tissues. For example, if mostly nonneutralizing antibodies are elicited against a virus, to which epitopes do they bind? What isotypes of antibodies are elicited, how long are the third complementarity-determining regions in the heavy chains (CDR H3s), and what is the degree of somatic mutation? In another example, an antibody library may be constructed from peripheral blood in a tumor or at a site of autoimmune inflammation, so that antibodies selected against these tissues might identify potentially therapeutic or pathologic specificities.

There is no doubt that there are many human antibody specificities of great diagnostic and therapeutic potential that may be cloned from immune donors. When making antibody libraries from immune sources, however, the donor and tissue source should be carefully chosen. The primary prognostic indicator is demonstrable serum antibody reactivity with the antigen to be studied. Serum antibody levels are thought to be predictive of antibody secretion, which in turn reflects the level of specific RNA in plasma cells, and it is this factor that will largely dictate the composition of the library. Although specific antibodies have been isolated from libraries constructed from donors with comparatively low serum titers, libraries prepared from donors with higher titers tend in general to yield a more diverse panel of Fabs. To clone human antibody genes from reverse-transcribed cDNA, human primer sets have been described, which allow the creation of very

diverse libraries covering all functional germ line V genes.

Also critical to library preparation is the choice of tissue source from which specific mRNA can be accessed. In principle, peripheral blood, bone marrow, spleen, tonsil, or lymph node can be used as a starting point for library construction. In humans, bone marrow is known to be a major reservoir of plasma cells that serve to maintain circulating antibody titers and has proven to be an excellent starting point for library construction. In contrast to bone marrow, the number of antibody-producing cells in the peripheral circulation is normally very low. However, if a potential donor is exposed to secondary contact with the antigen, the level of specific plasma cells increases dramatically within around three days postexposure, returning to a resting level a few days later. It has been demonstrated that libraries constructed from peripheral blood lymphocytes obtained during this time frame can yield specific antibodies. Recombinant Fabs have also been isolated from libraries prepared from spleen and tonsil, although these and lymph tissues are not so readily obtained.

To date, the major application of combinatorial libraries derived from immune donors has been in the area of infectious diseases, particularly viruses, although human antibodies have also been selected against bacteria, protozoans, and self and tumor antigens. With respect to viruses, current clinical management of viral infections continues to be inadequate, and at present there are only a handful of antiviral drugs in clinical use in a restricted number of disease states. A significant number of studies, however, suggest that human monoclonal antibodies may prove to be a viable alternative. First, it is well known that specific antibodies are capable of extremely efficient neutralization of a broad spectrum of viruses *in vitro* through a variety of mechanisms. In addition, numerous studies with human serum antibody preparations or rodent monoclonal antibodies have clearly shown the utility of antibodies in preventing an array of viral diseases, if these agents are administered before or shortly after exposure to the pathogen. In a smaller number of cases, neutralizing antibodies have been found to resolve established viral infections, an activity once considered to be mediated exclusively by cytotoxic T-lymphocytes. In humans, the lives of patients suffering from Argentine hemorrhagic fever, induced by Junin virus, can be saved by administration of pooled human sera containing a high titer of Junin virus neutralizing antibodies within eight days of the onset of symptoms. In an animal model of RSV, infection-neutralizing antibodies, administered intravenously or via aerosol at the height of infection, efficiently cleared the virus from the lungs of cotton rats. Antibodies have also been shown to clear viral infections from immunocompromised mice, including persistent virus infections of neurons, which are not recognized by CD8+ cytotoxic T-lymphocytes because the cells are deficient in molecules of the class I major histocompatibility complex. The potential of antibodies has also been demonstrated in passive transfer studies in animal models in which sterilizing immunity against viral challenge has been achieved.

Nevertheless, most commonly, an antibody does not provide sterilizing immunity against natural viral infection. Rather, the prevailing view has been that an antibody acts by reducing the viral load to allow the innate and the adaptive cellular arm of immunity to clear the remaining infection. Indeed, neutralizing antibodies are essential to the success of the vaccines against

major viral infections. In summary, there is a significant body of evidence underlining the proven and potential role of antibodies in prevention and therapy of viral infection. Sections 4.1.1 and 4.1.2 describe the application of phage antibody libraries in the generation of recombinant Fabs against clinically important pathogens and tumor-associated antigens.

4.1.1 Antibodies against Viruses and Prions

We have extensive experience in recovering and characterizing antibodies against viruses (particularly HIV-1) and prion proteins and will therefore present some of these antibodies as examples in this section. One of the first human phage display antibody libraries taken from an immune source was prepared using bone marrow from an individual who had been infected for several years with HIV-1 but was asymptomatic. As would be expected, the patient showed strong serological reactivity against the HIV-1 surface glycoprotein gp120. By panning the library phage against recombinant gp120 immobilized on ELISA wells, it was possible to isolate a large number of recombinant Fabs reacting with the panning antigen, but not with wells coated with bovine serum albumin.

DNA sequences from the variable domains of these 33 positive clones revealed that 10 clones possessed unique heavy chains and 20 clones possessed unique light chains. The sequencing data revealed a phenomenon termed *chain promiscuity*; that is, heavy chains that were identical, or very closely related, were seen pairing with light chains of very a different sequence. This observation made it clear that the heavy- and light-chain pairings found in library-generated Fabs are not necessarily those encountered *in vivo*. All the specific Fabs were subsequently shown to compete with soluble CD4 for binding to gp120. This accords with the finding that more than 50% of the gp120 reactivity present in the donor serum may be inhibited by CD4. Moreover, a cocktail of 3 Fabs was able to inhibit more than 50% of gp120 serum reactivity from a selection of seropositive donors. These observations indicate that despite chain promiscuity, antibodies generated from combinatorial libraries are relevant to and reflective of the *in vivo* response.

Ultimately, the more interesting Fab antibodies are those that are not only able to neutralize viruses *in vitro* but protect against infection *in vivo*. To determine their ability to inhibit HIV-1 infection, Fabs were assessed in p24 ELISA and syncytia assays. Most Fab clones showed little or no neutralizing activity, but one group of closely sequence-related Fabs was able to neutralize 50% of the virus in both assays at a concentration of approximately 1 $\mu g\ mL^{-1}$. One of these Fab clones (termed *b*12) has been linked to Fc to generate whole IgG1 (see Section 3). In this form, it has been shown to be the most effective anti-HIV-1 antibody produced to date, combining potent neutralizing activity with broad strain cross-reactivity against a diverse panel of virus primary isolates. Significantly, passive administration of IgG1 b12 to macaques has been shown to completely protect against vaginal challenge of the simian-HIV hybrid ($SHIV_{162P4}$). In this experiment, the serum titer of IgG1 b12 required for sterilizing immunity was roughly 100 times greater than the 90% neutralization titer of the challenge virus. The X-ray crystal structure of whole IgG1 b12 has now been determined and is the first structure to be determined of an entire human IgG1 molecule (Fig. 5). The long CDR H3 loop is predicted to be important for its anti-HIV-1

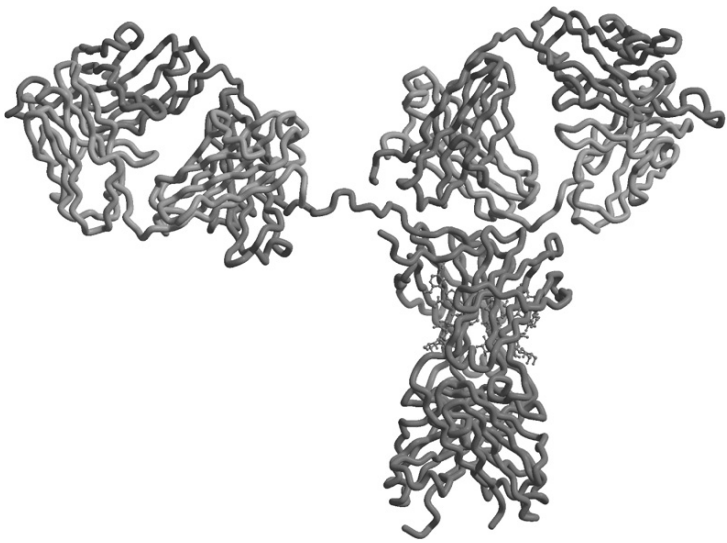

Fig. 5 Tube representation of the structure of the whole IgG1 b12, a broadly neutralizing human antibody against HIV-1. b12 was isolated from an Fab phage display library prepared from the bone marrow RNA of an asymptomatic HIV-1-seropositive individual. The structure was the first to be determined of an intact whole human IgG molecule. The light chains are shown in pink and the heavy chains in blue. Carbohydrate residues in the Fc portion of the molecule are also shown (see color plate p. xxii).

activity, and alanine-scanning mutagenesis of the CDR H3 loop of b12 supports this prediction.

Several other novel antibodies against HIV-1 have been selected by the use of phage display of which we will mention three. One antibody, Fab Z13, was selected using a synthetic peptide corresponding to the membrane-proximal region on the ectodomain of the transmembrane protein, gp41. This antibody and two others, 2F5 and 4E10, all bind to a similar region on gp41, and all of them are capable of cross-neutralizing primary isolates of HIV-1. Another antibody, Fab X5, was selected against a ternary complex of CCR5 (HIV-1 coreceptor), CD4, and gp120. Fab X5 has been shown to neutralize a wide range of primary isolates of HIV-1. Interestingly, the whole IgG X5 is actually poorer at neutralizing many viral isolates than its Fab counterpart, a property that has been attributed to steric constraints in accessing its epitope. A third antibody, scFv 4KG5, has been selected, which binds to a novel epitope on gp120 comprising the V1, V2, and V3 variable loops. The binding of 4KG5 to gp120 is inhibited by all anti-CD4 receptor binding site (CD4BS) antibodies tested ($n = 14$), and yet, its binding is moderately enhanced in the presence of the broadly neutralizing antibody b12 (described above). Thus, 4KG5 has helped distinguish, at the molecular level, b12 from other nonneutralizing anti-CD4BS antibodies.

Following the generation of the first recombinant Fabs against HIV-1, two more libraries were constructed from the bone marrow of additional HIV-1-positive asymptomatic donors. Analysis

of donor serum taken concomitantly with bone marrow samples indicated that these patients possessed measurable antibody titers against a plethora of different viruses. The libraries were then independently panned against antigens from the various pathogens to enrich for antigen-binding clones. Recombinant Fabs against HIV-1, RSV, herpes simplex virus (HSV) 1 and 2, varicella zoster virus (VZV), cytomegalovirus (CMV), and rubella were isolated from a single library, attesting to the antibody diversity captured during library construction. Essentially, existing antibody responses demonstrable at the serological level could be cloned and characterized at the molecular level.

Some of these Fab fragments were obtained by panning against lysates of virus-infected cells, rather than with recombinant or purified viral antigens. As a consequence, these experiments often generated antibodies against abundant, highly immunogenic antigens that do not elicit neutralizing antibodies in a natural immune response. To increase the probability of selecting neutralizing antibodies when purified antigen preparations are not available, a simple antigen-capture technique was developed. In this procedure, antigens known to contain important neutralizing epitopes (NE) are captured (typically from a lysate of virus-infected cells) onto a solid surface by an antibody specific for a nonneutralizing epitope (NNE) on the same antigen. Unbound material is then washed away, and panning is performed against the remaining captured antigen. In this way, only Fabs recognizing epitopes of the captured antigen (other than that occupied by the capturing antibody) are selected. Figure 6 gives an example taken from a capture-panning of the HSV glycoprotein D.

Two recombinant antibodies with exceptional neutralizing properties have been isolated against RSV and HSV. The RSV-specific antibody was derived from a library prepared from a HIV-1-positive donor with high serum antibody titer to recombinant RSV surface glycoprotein FG. Panning the library against FG yielded the antibody RSV 19, which was shown to potently neutralize a diverse collection of field isolates. RSV 19 was later evaluated as an Fab for therapeutic efficacy in a murine model of RSV infection, and when administered

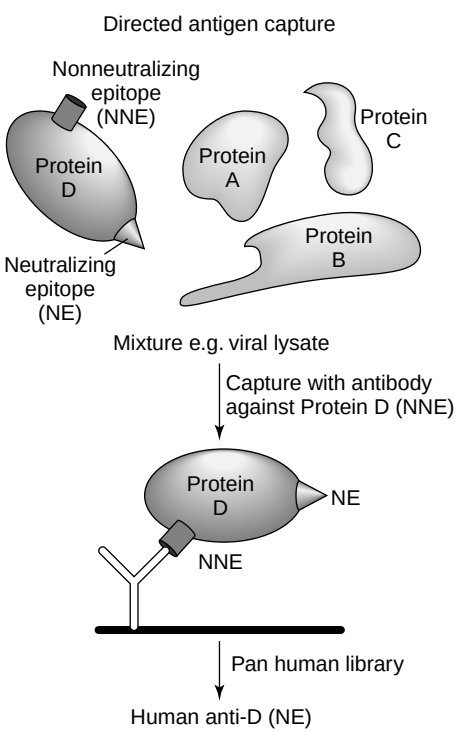

Fig. 6 Antigen capture prior to library panning. Library panning may be focused on a neutralizing epitope of a given antigen by first capturing the antigen, using a mouse antibody or a recombinant Fab that recognizes a nonneutralizing epitope. The method is advantageous when recombinant or purified forms of an antigen are not available.

in the form of an aerosol directly into the lungs of infected mice at the time of peak viral replication, it was able to reduce the viral titer by a factor of 5000. These findings suggest that recombinant human monoclonal Fabs may prove useful in the therapy of RSV disease and other respiratory conditions in which virus replication occurs largely in the luminal lining of the respiratory tract.

The HSV-specific antibody was selected from a library constructed from a different HIV-1-positive donor whose serum had demonstrable antibody titers to HSV-1 and -2. Even though only one heavy-chain sequence was recovered by panning against both viruses, all the antibodies studied reacted well with HSV-1 and -2. Only one clone, Fab 8, which recognizes the HSV envelope glycoprotein D, was able to consistently neutralize the virus in a plaque-reduction assay. HSV Fab 8 has also been shown to both completely prevent plaque development in cell monolayers 72 h post infection, establishing that the Fab can neutralize after attachment of the virus to the host cell, and to protect mice against vaginal challenge with the virus.

The filoviruses, which cause hemorrhagic fever, are characterized by their extreme lethality (up to ~80% of those infected die) and have gained much media attention in recent years. Human Fab phage display libraries were constructed from the bone marrow from two donors who recovered from infection with Ebola virus during the 1995 outbreak in Kikwit, Democratic Republic of the Congo. Initially, only weakly binding Fabs could be selected by radiation-inactivated infected cell lysates. By contrast, soluble Ebola virus glycoprotein (SGP) and irradiated whole virions were used to select Fabs with high affinity to both SGP and nucleoprotein (NP). Of these antibodies, Fab KZ52, which binds to GP, had the greatest reactivity against live Ebola virus–infected Vero cells as shown by immunofluorescence. Fab KZ52 was engineered as a whole IgG1 molecule and, significantly, neutralized Ebola virus at 0.4 $\mu g\ mL^{-1}$ in a plaque reduction assay. IgG1 KZ52 was also shown to protect guinea pigs from lethal Ebola Zaire virus challenge. KZ52 protected at 25 $mg\ kg^{-1}$ was administered 1 h prior to (preexposure) or 1 h following (postexposure) virus challenge. Although sterilizing protection was not observed, plasma viremia was reduced by approximately three orders of magnitude. Whether human recombinant monoclonal antibodies against the filoviruses will be useful as prophylactic agents remains unknown.

Prion diseases such as scrapie and bovine spongiform encephalopathy (''mad cow disease'') are fatal disorders of protein conformation, which are characterized by progressive degeneration of the central nervous system. Mad cow disease has had enormous negative impact on the beef and cattle industry in the United Kingdom, and worldwide restrictions have been placed on the import of beef from the United Kingdom. The insidious transmissibility of prion diseases and the lack of a reliable and sensitive diagnostic test for prions has created a demand for highly sensitive agents to accurately diagnose prion protein in animals, animal products, and in humans. The molecular basis of prion propagation is generally believed to be that the pathogenic or ''scrapie'' form of the prion protein (PrP^{Sc}) acts as a template to convert the normal cellular prion protein (PrP^{C}) to the PrP^{Sc} form. Although the primary amino acid sequences of PrP^{C} and PrP^{Sc} are identical, numerous structural studies suggest that the conversion of PrP^{C} to PrP^{Sc} involves an extensive conformational change with substantial acquisition

of β-sheet content. Elicitation of antibodies against prion protein is difficult due to host tolerance of endogenous PrP^C. This problem was overcome by immunizing mice in which the *PrP* gene was ablated ($Prnp^{0/0}$), and antibodies were retrieved from phage display libraries prepared from the bone marrow and spleens of these mice. Fabs specific for both discontinuous and linear epitopes covering the majority of the murine prion molecule were isolated, most of which were cross-reactive with human PrP. Surprisingly, although the mice were immunized with infectious PrP fragment preparations, Fabs against only the cellular form of PrP were isolated (Fig. 7). The antibodies generated have given insight into the conformational changes associated with PrP. For example, several of the antibodies were used to show that PrP that is sparsely immobilized on a biosensor chip surface can undergo spontaneous rearrangement to a conformation similar to PrP^{Sc}. More importantly, one of the Fabs, D18, has been shown to bind tightly to cell-surface PrP^C and inhibit PrP^{Sc} formation in a dose-dependent manner. These results may eventually lead to the prevention or treatment of prion diseases by anti-PrP antibodies.

4.1.2 Antibodies against Tumor-associated Antigens

Combinatorial antibody libraries are being increasingly used in cancer research in which an increasing number of tumor-associated antigens are being identified as potential targets for disease intervention. The successes in treatment of certain cancers with monoclonal antibodies such as in the treatment of relapsed HER2 breast malignancy by Herceptin (Genentech), and B-cell follicular lymphoma by Rituximab (IDEC/Genentech), have set a precedence that could be followed by library-derived antibodies in the near future. There are many examples of phage display–selected antibodies against tumor-associated antigens, but we will mention only a few due to limitation of space. Of particular note are studies in which phage display–selected antibodies

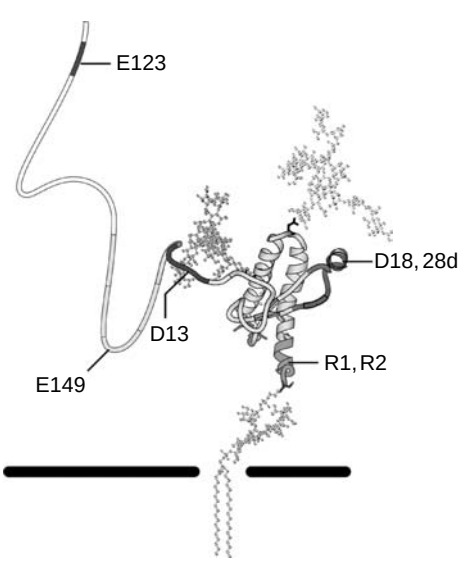

Fig. 7 Regions of prion protein (PrP) bound by individual antibodies selected from phage display libraries. Each antibody epitope has been assigned with a rectangle of different color as indicated on the figure. Beginning at the N-terminus, E123: purple; E149: grey; D13: blue; 3F4: yellow; 13A5: black line; D18 and 28d: red; R1 and R2: green. The cell membrane is shown as a black line at the bottom of the figure. The glycosylated side chains attached to Asn 181 and Asn197 are shown in orange and yellow respectively. The glycosyl phosphatidyl inositol protein (GPI) group anchoring PrP to the cell surface is shown in light blue. The antibody D18 has been shown to inhibit the formation of the infectious PrP^{Sc} form of the protein in a dose-dependent manner. [Adapted from Leclerc et al. (2003) *J. Mol. Biol.* **326**, 475–83.] (See color plate p. xxv.)

have given insight into the nature of the antibody response within tumors. For example, phage display libraries have been prepared from tumor-infiltrating B-cells revealing an oligoclonal antibody response against the intracellular autoantigen actin that becomes translocated to the cell surface upon apoptosis. Another group, using a similar approach, provides evidence for a bias toward intracellular antigens in the local antitumor immune response in a colorectal patient. Antibodies have also been selected against a number of gene products frequently found to be important in oncogenesis. An antibody library prepared from a colorectal cancer patient was used to isolate antibodies against the tumor suppressor p53. In another study, an antibody was selected against a stage-specific melanoma-associated antigen, and the antibody was used to identify a novel target, a 23 kDa glycoprotein, p23.

4.1.3 Intracellular Antibodies or "Intrabodies"

A relatively recent application of antibodies has been their use within the cell to modulate cellular function. Certain antibodies are able to bind to their target in the relatively oxidizing environment within the cytoplasm, and they can either act by inactivating a deleterious biochemical pathway or by initiating or enhancing a desirable one. In some cases, existing specificities can be engineered as scFvs specifically for intracellular expression, for example, by removing Cys residues and then selecting highly stable and functional variants using phage display. A number of experiments have demonstrated the ability of intrabodies to combat HIV-1 *in vitro*. Potential therapeutic applications of intrabodies hinge upon the potential of gene therapy, an area in which there has been a growing foundation of basic science and progress.

4.2 Antibodies in the Absence of Immunization

Isolating antibodies from an immune source has the advantage that the molecules will already have been selected by the host. However, there are many molecules of importance (e.g. self or highly conserved antigens) against which no high-affinity antibodies will be produced in the normal human host. Libraries prepared from "nonimmunized" or unimmunized sources provide the possibility of isolating these specificities. Moreover, given that libraries can be made approximating to the size of an animal's entire antibody repertoire, antibodies to a vast number of different antigens could become accessible from a single source.

4.2.1 Naive Libraries

Two main types of nonimmunized libraries have been described. In the first, "naive" antibody repertoires have been prepared using V genes from the IgD and IgM antibody classes. These antibody gene pools are less biased by the immune history of the host than those encoding IgG. Antibodies against a range of different molecules, including haptens and a number of self antigens, have been selected from such libraries. The antibodies have been shown to be highly specific, and, in some cases, of high (nM) affinity. More generally, however, they are of much lower affinity (1×10^5 to 1×10^6 M^{-1}) than those typically obtained from IgG libraries. In addition, from libraries of about 10^8 members, often only a small number of different clones of the desired specificity are selected. One of the most difficult

problems associated with this type of nonimmunized library is ensuring that the gene pool from which the library is made is truly naive. In fact, most V_H and V_L genes taken from IgM and light chain mRNA have already been biased in the host, precluding the selection of antibody against many antigens.

Nevertheless, some highly specific antibodies against diverse targets have been selected from nonimmune libraries. In one example, a nonimmunized library was screened against a diverse panel of antigens, and specific antibodies (scFv) were isolated in every case, including antibodies against botulinum neurotoxin, *Chlamydia trachomatis* elementary bodies, and the tumor marker Erb-B2; antibodies to the latter of which had dissociation constants below 1 nM. Another nonimmune library of $>10^{10}$ independent clones has been described from which has been isolated both cross-reactive and highly specific Fabs against the highly related glycoprotein hormones, human chorionic gonadotropin, human luteinizing hormone, and human follicle-stimulating hormone. Yet another nonimmune library was engineered using p9 display, and it performed comparably to many p3-displayed libraries.

4.2.2 Semisynthetic Libraries

Another source for nonimmunized libraries consists of antibodies that are encoded strictly by germ line genes. Such semisynthetic libraries are constructed from naturally occurring antibody genes in which some or all of the CDRs are composed of synthetic gene segments, and typically the CDR3 region of the heavy chain is composed of a randomized oligonucleotide sequence. A synthetic library has been described in which 49 V_H germ line genes were used as a starting point for library construction. These germ line heavy chains were given randomized CDR3 regions of five or eight residues and were combined with a single germ line lambda light chain gene. The library size of 2×10^7 clones represented only a fraction of the potential diversity in this system. However, this limitation must be considered in conjunction with the knowledge that immune systems themselves can represent only a fraction of the potential diversity at any one time, and that even germ line antibody genes are themselves biased by evolutionary pressures. To date, many synthetic antibody phage libraries have been described, some with considerable diversity (e.g. "Griffiths library," $\sim 6.5 \times 10^{10}$).

The importance of antigen recognition of CDR sequences, particularly the CDR3 of the heavy chain, makes them an ideal target for introducing segments of synthetic sequence to generate very large numbers of different specificities. Since naturally occurring heavy-chain CDR3 segments vary in length from 2 or 3 to 26 residues, the diversity that may be generated by randomization of this region alone is virtually limitless. In this sense, semisynthetic libraries may be considered to be more truly diverse and unbiased than those created from a naive source. However, although this aspect of library technology offers many possibilities, here too there are limits to diversity. For example, the complete randomization of 16 amino acids in a single heavy-chain CDR will alone require a library of more than 10^{20} clones to fully represent it. This figure is much greater than the transformation efficiency of *E. coli* (10^8–10^{10} clones per microgram of vector DNA). The probability of identifying desired specificities may, however, be increased by custom-designing semisynthetic libraries using

targeted mutagenesis and codon optimization. In addition, to mimic the affinity maturation seen in a secondary response *in vivo*, antibodies selected from both nonimmune and semisynthetic repertoires can be genetically manipulated *in vitro* to improve their affinity and/or modify their specificity. For example, it is possible to increase the affinity of these clones through the random mutagenesis of selected CDRs of the heavy chain followed by reselection against antigen. Additionally or alternatively, one can take a particular heavy or light chain and "cross" or reclone it into a light- or heavy-chain library respectively in the hope that it will pair with a more suitable partner. These strategies can often improve binding constants by more than two orders of magnitude, thus enabling high-affinity antibodies to be generated from naive sources.

Several companies are now in existence, which engineer and screen naive combinatorial antibody libraries that have been optimized to be very large and diverse. Fully human monoclonal antibodies have been selected against therapeutically relevant targets. In fact, Knoll Pharmaceuticals and the Cambridge Antibody Technologies Group have developed the first human antibody using phage libraries, which has been approved for human use and is marketed by Abbott Laboratories under the name Humira™ (http://www.fda.gov/ohrms/dockets/ac/03/briefing/3930B1_02_A-Abbott-Humira.pdf).

5 Summary

In the short time since their inception, combinatorial phage antibody libraries have been rapidly established as one of the most efficient routes to the generation of human monoclonal antibodies. Libraries prepared from immunized sources allow the selection of antibodies generated during a natural immune response. Moreover, the antibodies generated will normally be of high affinity following rounds of selection *in vitro*. Despite the scrambling of light and heavy-chain pairs inherent to library construction, antibodies selected from immune libraries have repeatedly been shown to be broadly reflective of the donor's serological reactivity. In addition, efficient selection procedures permit the isolation of antibodies occurring at very low frequencies in the library. Hence, library technology grants the investigator previously unavailable access to the human antibody response. Significantly, the problem of eliciting human antirodent responses, an enormous limitation to early therapeutic applications for rodent monoclonal antibodies, has been overcome as fully human antibody libraries are now readily available.

Libraries constructed from immune donors have already yielded a number of antibodies highly efficient in virus neutralization. Detailed information on the antibody response to infectious pathogens obtained through antibody libraries may be applied in the design and assessment of candidate vaccines as well as in elucidating the mechanisms by which viruses are neutralized by antibodies.

Unimmunized libraries, whether derived from naive repertoires or from synthetic antibody sequences, have almost unlimited potential to generate antibody specificities against any antigen. In some cases, these libraries have generated antibodies with affinities similar to those found in a secondary immune response *in vivo*. The resulting antibodies – having already been cloned at the

library-construction stage – are readily refined by mutagenesis. In combination with emerging array technologies, combinatorial antibody libraries are being used to identify binders for a multitude of novel targets and can be tested in high-throughput screens. These massive screening efforts are likely to yield new and improved diagnostics and therapeutics to further expand the already impressive role of antibodies in the clinic.

See also Antibody Molecules, Genetic Engineering of; Bioorganic Chemistry; Design and Application of Synthetic Peptides; Medicinal Chemistry.

Bibliography

Books and Reviews

Barbas, C.F. III, Burton, D.R., Scott, J.K., Silverman, G.J. (2001) *Phage Display: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Burton, D.R. (2002) Antibodies, viruses and vaccines, *Nat. Rev. Immunol.* **2**, 706–713.

Drewe, E., Powell, R.J. (2002) Clinically useful monoclonal antibodies in treatment, *J. Clin. Pathol.* **55**, 81–85.

Hemming, V.G. (2001) Use of intravenous immunoglobulins for prophylaxis or treatment of infectious diseases, *Clin. Diagn. Lab. Immunol.* **8**, 859–863.

Holt, L.J., Enever, C., de Wildt, R.M., Tomlinson, I.M. (2000) The use of recombinant antibodies in proteomics, *Curr. Opin. Biotechnol.* **11**, 445–449.

Marvin, D.A. (1998) Filamentous phage structure, infection and assembly, *Curr. Opin. Struct. Biol.* **8**, 150–158.

Rondon, I.J., Marasco, W.A. (1997) Intracellular antibodies (intrabodies) for gene therapy of infectious diseases, *Annu. Rev. Microbiol.* **51**, 257–283.

Primary Literature

Barbas, C.F. III, Crowe J.E. Jr., Cababa, D., Jones, T.M., Zebedee, S.L., Murphy, B.R., Chanock, R.M., Burton, D.R. (1992) Human monoclonal Fab fragments derived from a combinatorial library bind to respiratory syncytial virus F glycoprotein and neutralize infectivity, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10164–10168.

Better, M., Chang, C.P., Robinson, R.R., Horwitz, A.H. (1988) Escherichia coli secretion of an active chimeric antibody fragment, *Science* **240**, 1041–1043.

Bothmann, H., Plückthun, A. (1998) Selection for a periplasmic factor improving phage display and functional periplasmic expression, *Nat. Biotechnol.* **16**, 376–380.

Broders, O., Breitling, F., Dubel, S. (2003) Hyperphage. Improving antibody presentation in phage display, *Methods Mol. Biol.* **205**, 295–302.

Burioni, R., Williamson, R.A., Sanna, P.P., Bloom, F.E., Burton, D.R. (1994) Recombinant human Fab to glycoprotein D neutralizes infectivity and prevents cell-to-cell transmission of herpes simplex viruses 1 and 2 in vitro, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 355–359.

Burton, D.R., Barbas, C.F. III, Persson, M.A.A., Koenig, S., Chanock, R.M., Lerner, R.A. (1991) A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10134–10137.

Burton, D.R., Pyati, J., Koduri, R., Sharp, S.J., Thornton, G.B., Parren, P.W.H.I., Sawyer, L.S.W., Hendry, R.M., Dunlop, N., Nara, P.L., Lamacchia, M., Garratty, E., Stiehm, E.R., Bryson, Y.J., Cao, Y., Moore, J.P., Ho, D.D., Barbas, C.F. (1994) Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody, *Science* **266**, 1024–1027.

Coomber, D.W., Ward, R.L. (2001) Isolation of human antibodies against the central DNA binding domain of p53 from an individual with colorectal cancer using antibody phage display, *Clin. Cancer Res.* **7**, 2802–2808.

Crowe, J.E. Jr., Murphy, B.R., Chanock, R.M., Williamson, R.A., Barbas, C.F. III, Burton, D.R. (1994) Recombinant human respiratory syncytial virus (RSV) monoclonal antibody Fab

is effective therapeutically when introduced directly into the lungs of RSV-infected mice, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1386–1390.

D'Souza, M.P., Livnat, D., Bradac, J.A., Bridges, S., The AIDS Clinical Trials Group Antibody Selection Working Group, Collaborating Investigators, (1997) Evaluation of monoclonal antibodies to HIV-1 primary isolates by neutralization assays: performance criteria for selecting candidate antibodies for clinical trials, *J. Infect. Dis.* **175**, 1056–1062.

de Haard, H.J., van Neer, N., Reurs, A., Hufton, S.E., Roovers, R.C., Henderikx, P., de Bruine, A.P., Arends, J.W., Hoogenboom, H.R. (1999) A large non-immunized human Fab fragment phage library that permits rapid isolation and kinetic analysis of high affinity antibodies, *J. Biol. Chem.* **274**, 18218–18230.

De Kruif, J., Terstappen, L., Boel, E., Logtenberg, T. (1995) Rapid selection of cell subpopulation-specific human monoclonal antibodies from a synthetic phage antibody library, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3938–3942.

de Wildt, R.M., Tomlinson, I.M., Ong, J.L., Holliger, P. (2002) Isolation of receptor-ligand pairs by capture of long-lived multivalent interaction complexes, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8530–8535.

Gao, C., Mao, S., Kaufmann, G., Wirsching, P., Lerner, R.A., Janda, K.D. (2002) A method for the generation of combinatorial antibody libraries using pIX phage display, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12612–12616.

Gram, H., Marconi, L.A., Barbas, C.F. III, Collet, T.A., Lerner, R.A., Kang, A.S. (1992) In vitro selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 3576–3580.

Griffiths, A.D., Williams, S.C., Hartley, O., Tomlinson, I.M., Waterhouse, P., Crosby, W.L., Kontermann, R.E., Jones, P.T., Low, N.M., Allison, T.J., Prospero, T.D., Hoogenboom, H.R., Nissim, A., Cox, J.P.L., Harrison, J.L., Zaccolo, M., Gherardi, E., Winter, G. (1994) Isolation of high affinity human antibodies directly from large synthetic repertoires, *EMBO J.* **13**, 3245–3260.

Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H.R., Plückthun, A. (1998) Ribosome display efficiently selects and evolves high-affinity antibodies in vitro from immune libraries, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14130–14135.

Hansen, M.H., Nielsen, H.V., Ditzel, H.J. (2002) Translocation of an intracellular antigen to the surface of medullary breast cancer cells early in apoptosis allows for an antigen-driven antibody response elicited by tumor-infiltrating B cells, *J. Immunol.* **169**, 2701–2711.

Hoogenboom, H.R., Winter, G. (1992) By-passing immunisation. Human antibodies from synthetic repertoires of germline V_H gene segments rearranged in vitro, *J. Mol. Biol.* **227**, 381–388.

Horton, R.M., Hunt, H.D., Ho, S.N.P.J.K., Pease, L.R. (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension, *Gene* **77**, 61–68.

Huse, W.D., Sastry, L., Iverson, S.A., Kang, A.S., Alting-Mees, M., Burton, D.R., Benkovic, S.J., Lerner, R.A. (1989) Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda, *Science* **246**, 1275–1281.

Jestin, J.L., Volioti, G., Winter, G. (2001) Improving the display of proteins on filamentous phage, *Res. Microbiol.* **152**, 187–191.

Kieke, M.C., Cho, B.K., Boder, E.T., Kranz, D.M., Wittrup, K.D. (1997) Isolation of anti-T cell receptor scFv mutants by yeast surface display, *Protein Eng.* **10**, 1303–1310.

Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wolle, J., Plückthun, A., Virnekas, B. (2000) Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides, *J. Mol. Biol.* **296**, 57–86.

Kohler, G., Milstein, C. (1975) Continuous cultures of fused cells secreting antibody of predefined specificity, *Nature* **256**, 495–497.

Larrick, J.W., Danielsson, L., Brenner, C.A., Abrahamson, M., Fry, K.E., Borrebaeck, C.A.K. (1989) Rapid cloning of rearranged immunoglobulin genes from human hybridoma cells using mixed primers and the polymerase chain reaction, *Biochem. Biophys. Res. Commun.* **160**, 1250–1256.

Leclerc, E., Peretz, D., Ball, H., Sakurai, H., Legname, G., Serban, A., Prusiner, S.B., Burton, D.R., Williamson, R.A. (2001) Immobilized prion protein undergoes spontaneous

rearrangement to a conformation having features in common with the infectious form, *EMBO J.* **20**, 1554–1574.

Li, J., Pereira, S., Van Belle, P., Tsui, P., Elder, D., Speicher, D., Deen, K., Linnenbach, A., Somasundaram, R., Swoboda, R., Herlyn, D. (2001) Isolation of the melanoma-associated antigen p23 using antibody phage display, *J. Immunol.* **166**, 432–438.

Liu, B., Marks, J.D. (2000) Applying phage antibodies to proteomics: selecting single chain Fv antibodies to antigens blotted on nitrocellulose, *Anal. Biochem.* **286**, 119–128.

Marks, J.D., Hoogenboom, H.R., Bonnert, T.P., McCafferty, J., Griffiths, D., Winter, G. (1991) By-passing immunization. Human antibodies from V-gene libraries displayed on phage, *J. Mol. Biol.* **222**, 581–597.

Maruyama, T., Rodriguez, L.L., Jahrling, P.B., Sanchez, A., Khan, A.S., Nichol, S.T., Peters, C.J., Parren, P.W.H.I., Burton, D.R. (1999) Ebola virus can be effectively neutralized by antibody produced in natural human infection, *J. Virol.* **73**, 6024–6030.

Moulard, M., Phogat, S.K., Shu, Y., Labrijn, A.F., Xiao, X., Binley, J.M., Zhang, M.Y., Sidorov, I.A., Broder, C.C., Robinson, J., Parren, P.W. H.I., Burton, D.R., Dimitrov, D.S. (2002) Broadly cross-reactive HIV-1-neutralizing human monoclonal Fab selected for binding to gp120-CD4-CCR5 complexes, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6913–6918.

Mullinax, R.L., Gross, E.A., Amberg, J.F., Hay, B.N., Hogrefe, H.H., Kubitz, M.M., Greener, A., Alting-Mees, M., Ardourel, D., Short, J.M., Sorge, J.A., Shopes, B. (1990) Identification of human antibody fragment clones specific for tetanus toxoid in a bacteriophage lambda immunoexpression library, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8095–8099.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction, *Cold. Spring. Harb. Symp. Quant. Biol.* **51**, 263–273.

O'Connell, D., Becerril, B., Roy-Burman, A., Daws, M., Marks, J.D. (2002) Phage versus phagemid libraries for generation of human monoclonal antibodies, *J. Mol. Biol.* **321**, 49–56.

Orlandi, R., Gussow, D.H., Jones, P.T., Winter, G. (1989) Cloning immunoglobulin variable domains for expression by the polymerase chain reaction, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 3833–3837.

Parren, P.W.H.I., Marx, P.A., Hessell, A.J., Luckay, A., Harouse, J., Cheng-Mayer, C., Moore, J.P., Burton, D.R. (2001) Antibody protects macaques against vaginal challenge with a pathogenic R5 simian/human immunodeficiency virus at serum levels giving complete neutralization in vitro, *J. Virol.* **75**, 8340–8347.

Peretz, D., Williamson, R.A., Kaneko, K., Vergara, J., Leclerc, E., Schmitt-Ulms, G., Mehlhorn, I.R., Legname, G., Wormald, M.R., Rudd, P.M., Dwek, R.A., Burton, D.R., Prusiner, S.B. (2001) Antibodies inhibit prion propagation and clear cell cultures of prion infectivity, *Nature* **412**, 739–743.

Persson, M.A., Caothien, R.H., Burton, D.R. (1991) Generation of diverse high-affinity human monoclonal antibodies by repertoire cloning, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2432–2436.

Plotkin, S.A. (2001) Immunologic correlates of protection induced by vaccination, *Pediatr. Infect. Dis. J.* **20**, 63–75.

Roovers, R.C., van der Linden, E., Zijlema, H., de Bruine, A., Arends, J.W., Hoogenboom, H.R. (2001) Evidence for a bias toward intracellular antigens in the local humoral anti-tumor immune response of a colorectal cancer patient revealed by phage display, *Int. J. Cancer* **93**, 832–840.

Saphire, E.O., Parren, P.W.H.I., Pantophlet, R., Zwick, M.B., Morris, G.M., Rudd, P.M., Dwek, R.A., Stanfield, R.L., Burton, D.R., Wilson, I.A. (2001) Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design, *Science* **293**, 1155–1159.

Sastry, L., Alting-Mees, M., Huse, W.D., Short, J.M., Sorge, J.A., Hay, B.N., Janda, K.D., Benkovic, S.J., Lerner, R.A. (1989) Cloning of the immunological repertoire in *Escherichia coli* for generation of monoclonal catalytic antibodies: construction of a heavy chain variable region-specific cDNA library, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5728–5732.

Sblattero, D., Bradbury, A. (2000) Exploiting recombination in single bacteria to make large phage antibody libraries, *Nat. Biotechnol.* **18**, 75–80.

Sheets, M.D., Amersdorfer, P., Finnern, R., Sargent, P., Lindqvist, E., Schier, R., Hemingsen, G., Wong, C., Gerhart, J.C., Marks, J.D.

(1998) Efficient construction of a large nonimmune phage antibody library: the production of high-affinity human single-chain antibodies to protein antigens, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6157–6162.

Skerra, A., Plückthun, A. (1988) Assembly of a functional immunoglobulin F_v fragment in *Escherichia coli*, *Science* **240**, 1038–1041.

Smith, G.P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface, *Science* **228**, 1315–1317.

Vaughan, T.J., Williams, A.J., Pritchard, K., Osbourn, J.K., Pope, A.R., Earnshaw, J.C., McCafferty, J., Hodits, R.A., Wilton, J., Johnson, K.S. (1996) Human antibodies with subnanomolar affinities isolated from a large non-immunized phage display library, *Nat. Biotechnol.* **14**, 309–314.

Yang, W.P., Green, K., Pinz-Sweeney, S., Briones, A.T., Burton, D.R., Barbas, C.F. III (1995) CDR walking mutagenesis for the affinity maturation of a potent human anti-HIV-1 antibody into the picomolar range, *J. Mol. Biol.* **254**, 392–403.

Zwick, M.B., Labrijn, A.F., Wang, M., Spenlehauer, C., Ollmann Saphire, E., Binley, J.M., Moore, J.P., Stiegler, G., Katinger, H., Burton, D.R., Parren, P.W.H.I. (2001) Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41, *J. Virol.* **75**, 10892–10905.

Cytochrome P450

Michael R. Waterman[1] *and Irina A. Pikuleva*[2]
[1] *Vanderbilt University School of Medicine, Nashville, Tennessee, USA*
[2] *University of Texas Medical Branch, Galveston, Texas, USA*

1 Characteristics of P450s 102

2 P450 Functions 104

3 How Many P450s Are There? 108

4 P450 Structure 109

5 Regulation of P450 Activities 111

6 Future Directions 112

Bibliography 114
Books and Reviews 114
Primary Literature 114

Keywords

Endogenous Substrates
Substrates for enzymatic (P450) reactions that are natural compounds synthesized within the organism harboring the enzyme.

Exogenous Substrates
Substrates for enzymatic (P450) reactions that are derived from the environment surrounding the organism harboring the enzyme.

Hemoprotein
A protein containing the prosthetic group, heme.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

Monooxygenase
Enzyme that reduces O_2: one oxygen atom is used in substrate modification (often hydroxylation), while the other is used to form water.

Superfamily
A group of proteins that are evolutionarily related as determined by their primary amino acid sequence.

In the original issues of this Encyclopedia in 1996, there was the article (Volume 1) on cytochrome P450 monooxygenases. This new article follows the same format as that used previously. However, a great deal of interesting information on P450s has appeared in the last six to seven years. Accordingly, it is the goal of the authors to present the basic principles of the P450 system as outlined in the previous article, while at the same time taking advantage of new discoveries to provide more depth in understanding these fascinating enzymes.

Cytochrome P450 is the generic name applied to a large superfamily of hemoprotein, mixed-function oxidases (monooxygenases) that metabolize a structurally diverse group of exogenous and endogenous organic substrates. The name is derived from the prominent absorption band observed at 450 nm following reduction of the heme iron and its coordination with carbon monoxide. These enzymes are widely distributed among all biological kingdoms and they catalyze a reaction of the general nature

$$O_2 + \underset{\text{Biological reducing equivalents}}{NAD(P)H} + \underset{\text{Organic substrate}}{AH} \xrightarrow{P450} \underset{\text{Hydroxylated organic product}}{AOH} + H_2O + NAD(P)^+$$

In the case of exogenous substrates, this reaction is most frequently involved in an organism's effort to detoxify foreign compounds (xenobiotics) from the environment. In the case of endogenous substrates, the reaction is generally involved in the production of biologically more active compounds (viz. steroid hormones) from less active compounds (viz. cholesterol).

1 Characteristics of P450s

In 1962, Ryo Sato and Tsuneo Omura characterized an unusual pigment in a subcellular fraction of rabbit liver, the endoplasmic reticulum, which showed an intense absorbance band at 450 nm upon reduction and coordination with carbon monoxide. They tentatively named the material "pigment 450 nm" or "cytochrome P450." The latter name

stuck even though cytochromes P450, being enzymes, do not fit the true definition of cytochromes. Subsequently, the pigment in question was shown to contain protoheme, the same prosthetic group found associated with many other hemoproteins, most notably hemoglobin and myoglobin. Hemoglobin and myoglobin also show an intense absorbance upon reduction and binding to CO, but with these proteins the maximum is at 420 nm. Iron has six coordination positions: four coordination positions in protoheme are occupied by the pyrrole nitrogens; in hemoglobin and myoglobin, the fifth coordination position is provided by the imidazole group of the proximal histidine; and the sixth position is available to bind gaseous ligands such as O_2 or CO. It is now apparent that in P450s, the fifth coordination position is provided by a thiolate group from the amino acid cysteine. Although the Nomenclature Committee of the International Union of Biochemistry suggests the name "heme-thiolate protein" rather than "cytochrome P450," common usage in the field is simply P450, dropping the "cytochrome."

Shortly after Sato and Omura tagged this hemoprotein as cytochrome P450, Ronald Estabrook, Otto Rosenthal, and David Cooper at the University of Pennsylvania assigned a function to a cytochrome P450 in the endoplasmic reticulum of the adrenal cortex by demonstrating with CO inhibition and photochemical action spectroscopy its ability to hydroxylate progesterone at C-21. We now know that there are many different P450s and that they catalyze the general reaction type shown above.

Returning to our analogy with hemoglobin, we recognize that a fundamental difference between P450 and hemoglobin is that hemoglobin binds O_2 reversibly and transports it unchanged, while P450 reduces O_2, one atom of oxygen being inserted into the substrate and the other used for the formation of water. The redox potentials of these two hemoproteins are quite different, that of P450 being -300 mV or lower, while that of hemoglobin is $+110$ mV. Presumably, the electron-donating capacity of the thiolate ligand can facilitate the reduction of O_2, making P450s potent monooxygenases. Not all heme-thiolate proteins are P450s, but we will find that P450s consist of a very large number of proteins (over 2,400 identified to date) found in plants, bacteria, birds, fishes, insects, mammals, and probably all other forms of life.

For the heme iron to reduce oxygen, cytochrome P450 must receive electrons from reduced pyridine nucleotides (either NADH or NADPH) via an electron transport system. In bacteria, P450s are soluble proteins; in plants, insects, and animals, they reside as integral members of subcellular membranes. Virtually all soluble bacterial P450s, as well as those forms localized in mitochondrial membranes of eukaryotes, receive electrons via the following pathway:

$$\text{NADPH(NADH)} \longrightarrow \underset{\text{(FAD – containing)}}{\text{ferredoxin reductase}}$$

$$\longrightarrow \underset{\text{(often 2Fe-2S)}}{\text{ferredoxin}} \longrightarrow \text{P450}$$

Electrons are transferred to the P450 one at a time from the ferredoxin. In mitochondria, P450s are localized to the inner mitochondrial membrane, with ferredoxin reductase and ferredoxin residing as soluble proteins in the matrix. In eukaryotes, the majority of P450s are localized in the endoplasmic reticulum as integral membrane proteins. Reducing equivalents are transferred to these P450s via a microsomal flavoprotein quite distinct from

ferredoxin reductase:

$$\text{NADPH} \longrightarrow \text{NADPH P450 reductase (FAD + FMN containing)} \longrightarrow \text{P450}$$

where FAD is flavin adenine dinucleotide and FMN is flavin mononucleotide. A small number of variations in the reductase systems do exist amongst the P450 superfamily. In at least one case, a soluble bacterial P450 utilizes the microsomal type of reductase system that is attached to the C-terminus of the P450 domain to form a fusion protein. Another type of P450 fusion protein having a ferredoxin domain attached to the P450 domain is also found in bacteria. In another example found in fungi, reduced pyridine nucleotide interacts directly with the P450.

In summary then, P450s are protoheme-containing, mixed-function oxidases in which a cysteine thiolate ligand coordinates the polypeptide chain to the heme iron. In bacteria, these proteins are soluble, while in higher life-forms they are integral membrane proteins. Virtually all P450s require additional proteins to transfer electrons from reduced pyridine nucleotides to their heme iron.

2 P450 Functions

P450s metabolize a wide range of both endogenous and exogenous (foreign from the environment) compounds. These compounds generally are lipophilic, and it is predicted that the P450 superfamily can metabolize more than 10,000 known compounds. Endogenous substrates include steroids, fatty acids, prostaglandins, leukotrienes, bile acids, and biogenic amines, while exogenous substrates include environmental pollutants and pharmaceuticals. P450s catalyze the oxidative, peroxidative, and reductive metabolism of such compounds. In the broadest sense then, we can group P450s into two classes: those that metabolize endogenous compounds and those that metabolize exogenous substrates (Table 1). The general catalytic mechanism for oxygenation by P450s of both types of substrate can be represented by the cycle shown in Fig. 1. Binding of substrate in the P450 active site has been shown in some cases to raise the redox potential of the heme iron to about −180 mV, making it more susceptible to reduction by NADPH.

In animals, the forms involved in endogenous substrate metabolism tend to be localized in specific functional sites such as steroidogenic tissues (adrenal, gonads, placenta) or a specific organ (kidney, liver) in which they participate in key endogenous pathways such as steroid hormone biosynthesis (mineralocorticoids, glucocorticoids, sex hormones), bile acid biosynthesis, vitamin D_3 activation and inactivation, and synthesis of key metabolites from arachidonic acid. These reactions lead to the production of regulators of many biological processes such as reproduction and vascular activity, to name only two. Other potential roles for P450s include biosynthesis of neurosteroids and activation or inactivation of important compounds for growth and development including derivatives of vitamin A. Certain P450-dependent endogenous functions such as cholesterol biosynthesis are widely distributed. Cholesterol biosynthesis is important in most, if not all, animal cells for membrane biogenesis and is dependent on lanosterol 14α-demethylase cytochrome P450. A key developmental role for P450-dependent activities is

Tab. 1 Classification of P450s according to substrate metabolism.

Endogenous substrate metabolism
Examples
Steroid hormone biosynthesis
Vitamin D_3 metabolism
Cholesterol biosynthesis
Bile acid biosynthesis
Reactions
Hydroxylations
C—C bond lysis
Exogenous substrate metabolism
Examples
Environmental pollutant metabolism
Drug metabolism
Insect metabolism of plant toxins
Reactions
Hydroxylations
N-Dealkylations
O-Dealkylations
Alcohol oxidation
Alkane oxygenations

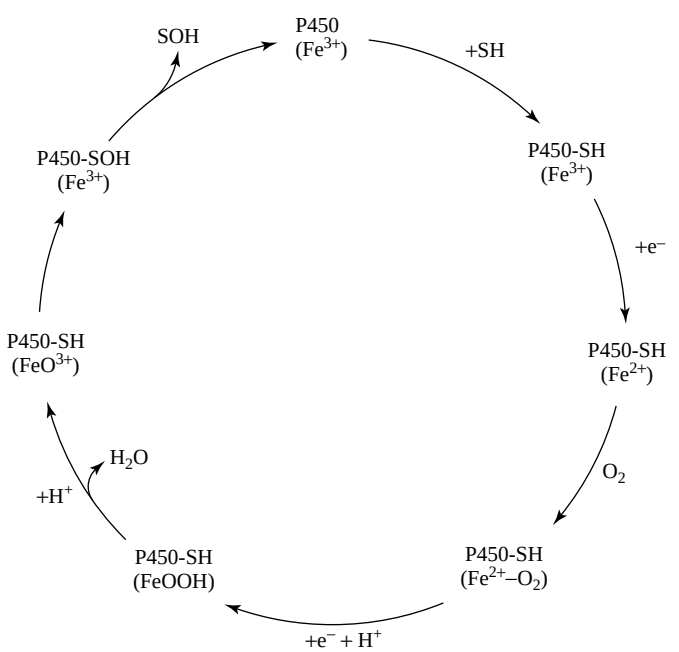

Fig. 1 Generalized catalytic cycle for cytochromes P450.

the requirement for testosterone at the correct time during fetal life for the appearance of the male phenotype. Without timely testosterone production, the 46,XY male karyotype displays female sex characteristics, and two different P450s (one mitochondrial and one microsomal) are required for testosterone biosynthesis from cholesterol. In the male songbird, testosterone produced in the testis crosses the blood–brain barrier, where it is converted to estrogen by the aromatase cytochrome P450 in the brain, which activates the formation of the song center in the brain.

Inherited genetic diseases are known to be associated with mutations in the P450-encoding genes involved in endogenous substrate metabolism. For example, the group of diseases known as congenital adrenal hyperplasia results from defects in steroid hormone biosynthesis in the gonads and adrenal. Depending on the specific *P450* gene affected, alterations in these pathways lead to changes in mineralocorticoid, glucocorticoids, or sex hormone levels. Another P450-dependent pathway subject to genetic defects is bile acid biosynthesis. Recently, a mutation in a fatty acid hydroxylase P450 expressed in tubules in the kidneys has been found to be associated with human hypertension. Aromatase P450 is required for estrogen biosynthesis and, in addition to expression in the ovaries, it is found in adipose (fat) tissue, in which it may be a key locus in breast cancer. Thus, members of the class of P450s involved in endogenous substrate metabolism play important roles in development and homeostasis and are the loci for a number of disease states in humans.

Forms of P450 involved in exogenous substrate metabolism are generally more widely distributed and have broader substrate specificities than forms involved in endogenous substrate metabolism. Exogenous substrate metabolism is found in every organ including the skin, while the highest concentration and largest number of exogenous substrate–metabolizing P450s is in the liver. Some forms in this class are widely distributed, such as that which metabolizes polycyclic aromatic hydrocarbons (P4501A1 – this nomenclature is defined in the next section), which can be found in almost every cell type. Others are present in only a limited number of cells such as the olfactory epithelium.

In bacteria, P450-dependent xenobiotic (foreign compound) metabolism can involve compounds that serve as important carbon sources for these organisms. A particular strain of *Pseudomonas putida* grows on camphor following its metabolism via several enzymatic steps, including one catalyzed by the soluble cytochrome P450cam. In insects, P450-dependent xenobiotic metabolism can provide protection against plant toxins produced to ward off pests. The larvae of the black swallowtail butterfly are able to feed on carrot and parsley plants because they can rapidly induce a particular P450 that metabolizes the xanthotoxin produced by the plants. Many other insects cannot generate such a defense in the warfare between plants and insects and thus cannot feed on naturally protected plants. Taxol, an important anticancer agent that disrupts microtubule formation, is produced by the Pacific yew tree via a P450-dependent pathway as a defense against pests. Animals including humans contain a broad spectrum of different P450s capable of biotransforming pollutants and drugs. These transformations (generalized in Table 1) involve countless drugs, chemical carcinogens, and environmental contaminants. From an evolutionary point of view, it is interesting to consider how these forms

arose. Our modern chemical environment is less than 200 years old. Nevertheless, forms such as P4501A1 are found in fishes and all mammals. This P450 metabolizes polycyclic aromatic hydrocarbons, which can arise from burning organic matter. Perhaps such forms were first produced during volcanic eras and have persisted to present times.

Much of the P450-dependent xenobiotic metabolism serves the purpose of detoxification. Hydroxylated products are more soluble than the original substrates and also serve as good substrates for phase II detoxification enzymes, which conjugate these products for easier elimination. Obviously, this ability is beneficial to humans, just as detoxifying xanthotoxin is beneficial to the black swallowtail butterfly larvae. In certain cases, however, products produced by P450 metabolism can be more toxic than the substrate. For example, benzo(a)pyrine, a polycyclic aromatic hydrocarbon in tobacco smoke that also arises from charbroiling meat, is converted by P450-dependent biotransformation into a product that is carcinogenic and may play a key role in lung cancer. Thus, in certain instances, products of P450 biotransformation can be mutagenic or carcinogenic.

The pharmaceutical industry is particularly interested in the roles P450s play in drug metabolism. This metabolism obviously contributes to the efficacy of pharmaceuticals. In drug metabolism studies using laboratory animals, inbred strains are generally used in which all individuals contain fairly constant levels of different forms of drug-metabolizing P450s. Humans are outbred and show considerable individual variation in drug metabolism. Each individual can be considered to have a unique drug-metabolizing profile in part because of the person's individual P450 pattern, which is dependent on genetic (sex, age) and environmental (nutrition, exposure) factors. The human genome has revealed genomic evidence for 57 P450s to date, a mixture of forms metabolizing endogenous and exogenous substrates. Of this latter group, six found in human liver are primary P450s used to study P450 metabolism of potential pharmaceuticals. In some cases, different forms of P450 metabolize the same chemical substrate by different reactions, leading to varied patterns of products from the same drug. Thus, individual variation in the levels of different P450s can significantly affect drug metabolism, particularly in instances of combined drug therapy. Genetic contributions to individual variations (drug metabolism polymorphisms) exist within the human population as a result of variations in the levels of specific forms of P450 or mutations within specific enzymes. Thus, certain individuals, classed as poor metabolizers for specific drugs, do not clear these drugs as efficiently as the general population and experience side effects reminiscent of drug overdose. A goal of research in drug metabolism is to develop noninvasive screening procedures that will permit estimation of individual P450 profiles. Thus, individual treatment regimens could be prescribed providing both health and economic benefits to patients.

While the classification of P450s into two groups as used here is helpful, ancestral P450s probably have existed for more than 3.5 billion years. This longevity suggests that the drug-metabolizing forms of P450 could play important roles in regulating key endogenous compounds used in growth, development, and homeostasis. In most cases, no biological connection between members of the exogenous substrate–metabolizing class and specific endogenous substrates (e.g. retinoids) has

yet been made. However, because drug-metabolizing P450s have a broad substrate specificity, it is clear that, *in vitro*, these P450s can metabolize various endogenous compounds. Thus, it will not be surprising to see the distinction between these two classes of P450 become less sharp in the future.

3 How Many P450s Are There?

The P450 superfamily is very large, having members in most if not all life-forms. In 1996, only 400 CYPs were known and now the P450 superfamily contains more than 2400 enzymes. The majority of these have been identified in genomic sequences and for the vast majority, the function is unknown. Classification of members of the superfamily based on function has not been possible, originally because enzymes from the same superfamily in different species have different activities and now because we have far too many genes with unknown activities. Rather, a systematic classification for P450s has been developed on the basis of amino acid sequence (http://drnelson.utmem.edu/cytochrome.P450.html). Before development of this classification by Dan Nebert and David Nelson, the field was in disarray because individual investigators selected names for the different P450s, and in cases in which discovery occurred in two laboratories at about the same time, there were two names for the same CYP. The nomenclature involves a number that identifies the gene family, a letter that identifies the subfamily, and another number that identifies the member of the subfamily. CYP2C23 indicates gene family 2, subfamily C of gene family 2, and member 23 of subfamily C of gene family 2. The subfamily member designation is sequential based on order of discovery. CYP2C23 and CYP2C24 are rat P450s, while CYP2C25 is from the Syrian hamster. Gene family members have about 40% amino acid sequence identity or greater, while subfamilies have about 55% amino acid identity or greater. Certain gene families that have very restricted enzymatic activity and are found to have the same function in different organisms do not have subfamilies. CYP51 (sterol 14α-demethylase) is an example because it is involved in sterol biosynthesis in many organisms in which it removes the 14α-methyl group from the cyclized sterol ring. Another such example is CYP17 (17α-hydroxylase/17,20-lyase) found in the gonads of animals, birds, and fish serving an identical role in androgen biosynthesis. However, the P450s in the CYP2C subfamily generally have different substrate specificities from those in the CYP2B subfamily. It would not even be surprising to learn that CYP2C23 and CYP2C24 have different substrate specificities. Sometimes, a single amino acid change can profoundly change the specificity of a P450. In the mouse, the Cyp2A subfamily (Phe209 $\rightarrow$ Leu) changes the specificity of mouse coumarin hydroxylase (Cyp2A5) to steroid 15α-hydroxylase (Cyp2a4). For mouse P450s Cyp is used, while for P450s in all other species the designation CYP is used.

CYP51 is expressed in all kingdoms of life, and the approximately 50 known CYP51 sequences contain 40 conserved amino acids. Obviously, sterol 14α-demethylation is a highly conserved enzymatic activity. Another highly conserved enzymatic activity is ω-hydroxylation of fatty acids. However, in this case different gene families are involved in different phyla. There is not enough amino acid conservation between different

fatty acid hydroxylases to catalog them together in the same gene family. Perhaps this difference between sterol 14α-demethylation and fatty acid ω-hydroxylation is because three successive monooxygenation steps are required for 14α-demethylation and only one for fatty acid hydroxylation. Only three amino acid residues are conserved in all P450s, the cysteine that provides the thiolate ligand for the heme iron and the EXXR sequence that forms a charge pair in the K-helix of P450s. The number of P450s are very different from organism to organism and Table 2 provides selected examples that emphasize this point.

As indicated earlier, knowing the sequence of a P450 does not necessarily provide information on its function. A variety of different heterologous expression systems including bacteria, yeast, insect cells, and animal cells have been applied to P450 research. Thus, as new P450s are identified by cloning, they can be overexpressed and their enzymatic properties examined with a variety of potential substrates. Even following such characterization, however, it is not possible to know specific P450 function with certainty in the absence of biological data, including tissue localization and physiological regulation of activity. Directed approaches to identify rare P450s are difficult. Universal probes do not exist for P450s, either nucleotide or antibodies. Use of subfamily-specific PCR primers does permit identification of new P450s within a limited subset. P450s contain approximately 500 amino acids, of which fewer than 10 are highly conserved, several of these being located near the heme-binding cysteine. In studies aimed at cloning P450 genes responsible for flower color, degenerate PCR primers complementary to sequences in the heme binding region were very effective in identifying a number of different P450 forms.

4 P450 Structure

In the 1996 article, it was reported that four soluble bacterial P450 structures were known. This number has grown to at least nine. In addition, we now know the structure of a soluble fungal P450, and most

Tab. 2 Expression of P450s in different organisms.

Organism	*Number of CYPs*
Human	57
Mouse	60
Rice	458
Arabidopsis	273
Caenorhabditis elegans	81
Drosophila melagaster	84
Streptomyces cerevisae	3
Escherichia coli	0
Mycobacterium tuberculosis	20
Mycobacterium emegmatus	44
Mycobacterium leprae	1
Streptomyces coelicolor	18
Streptomyces avermitilus	33

exciting, the structures of mammalian, membrane-bound CYPs are becoming available. The first is a rabbit microsomal steroid hydroxylase, CYP2C5, and a commercial firm has reported structure determination of two human hepatic CYPs (2C9 and 3A4). Because eukaryotic CYPs are integral membrane proteins and are consequently more hydrophobic, they are more difficult to crystallize. However, that is true of proteins in general where more than 2,000 structures of soluble proteins are available, yet only about 50 membrane protein structures have been solved. The recent success in structure determination of eukaryotic P450s has required modification by mutagenesis to generate less hydrophobic molecules. We can expect additional structures of human P450s to appear in the near future.

Generally speaking, there is little sequence conservation between *CYP* gene families. However, the CYP structures all look the same from a distance, assuming what is now called the *P450 fold*. Closer analysis reveals, however, subtle but important differences. More than 10 years ago, Gotoh described in the CYP2 family six "substrate recognition sequences" (SRS). These regions are predicted to interact with the substrate and are hypervariable between gene families but are more conserved within P450s that carry out similar or the same reactions. Having considerable structural information on P450s, we now think of these regions in terms of secondary structural elements. SRS1 is generally found in the region of B′- (or B)helix and BC loop, SRS2 in the C-terminal region of the F-helix, SRS3 in the N-terminal region of G-helix, SRS4 in the I-helix, SRS5 in β1–4, and SRS6 in β4–2. The lengths of these secondary structural elements vary between P450 families to accommodate substrates of different shape, size, and polarity. From structural studies, we know that the active site in P450s is sequestered inside the protein molecule, and it is not clear at present how the substrate gains access to the enzyme's active site. Only in two cases has the structure of the same CYP plus and minus substrate been determined – the bacterial enzymes P450cam and P450BMP. In the case of P450BMP (CYP102A1), there is a channel (substrate access channel) that connects the surface of the molecule and the buried active site, which is open in the substrate-free form of the enzyme. Binding of the substrate causes significant conformational changes and results in a closure of the substrate access channel. This is in contrast to P450cam (CYP101), in which there is no obvious route to the active site and substrate binding is accompanied by only minor conformational changes.

P450s in the endoplasmic reticulum have a very hydrophobic amino-terminal region, which serves as a signal anchor sequence. All such P450s are inserted co-translationally into the endoplasmic reticulum membrane by the signal recognition particle pathway, and this amino-terminal anchor participates both in association of the P450 with the membrane and in directing the proper folding pathway for the protein. However, removal of this hydrophobic tail from the primary sequence of the P450 does not generally lead to synthesis of a soluble P450. Rather, such truncated P450s are still found to be associated with the membrane, indicating that other hydrophobic stretches in the P450 molecule participate in membrane association. Even P450s in the mitochondrion do not have discernible hydrophobic signal anchors. They, however, have amino-terminal precursor extensions that are removed proteolytically as the unfolded P450 is taken up

and inserted into the mitochondrial membrane before final folding occurs. There is no sequence homology of this extension between different mitochondrial P450s and its length varies significantly from one protein to another. Like other precursor sequences, however, mitochondrial P450 precursor segments contain charged amino acids. All membrane-bound P450s are proposed to have a similar mode of association with the membrane via a large hydrophilic domain anchored to the lipid bilayer through several noncontiguous portions of the polypeptide chain, with some of the membrane-interacting areas being common to both mitochondrial and microsomal P450s. It has also been suggested that in membrane-bound P450s the substrate may enter directly from the lipid bilayer, thus providing a rationale as to how hydrophobic substrates can avoid traveling through the aqueous cytosol.

Other than addition of protoheme, posttranslational modification is not a frequent event in P450s. There have been reports that phosphorylation is important in modifying the function of specific forms of P450, but this has not been firmly established. Glycosylation of aromatase cytochrome P450 is known to occur, although this modification apparently has no effect on the activity of the enzyme. Glycosylation of other P450s is not common. Heme binding remains the one posttranslational event occurring in all P450s. How it occurs, however, is not at all understood. Since the last step of heme biosynthesis (iron insertion into protoporphyrin IX) occurs in the mitochondrion, mitochondrial P450s might find heme quite readily in the matrix. How heme gets to the endoplasmic reticulum for insertion into P450s and other heme proteins is a mystery.

5
Regulation of P450 Activities

While certain P450-dependent activities, such as the lanosterol 14α-demethylase step in cholesterol biosynthesis and that of 27-hydroxylase in bile acid biosynthesis, have important constitutive regulatory components, other P450 activities are regulated by signal transduction pathways. Much of this regulation is at the transcriptional level. For example, peptide hormones from the anterior pituitary regulate transcription of P450s involved in steroid hormone biosynthesis. Adrenocorticotropin (ACTH), via cAMP-dependent pathways, controls the transcription of genes in the adrenal cortex, encoding both mitochondrial and microsomal P450s in the steroid hormone biosynthetic pathway, as well as related proteins such as ferredoxin. ACTH also controls the availability of substrate (cholesterol) in the inner mitochondrial membrane, thereby having the dual effect of regulating enzyme levels and substrate availability. As another example, parathyroid hormone regulates P450-dependent vitamin D_3 metabolism via cAMP, presumably at the transcriptional level.

When considering regulation of P450 activities, there are four different regulatory components to be evaluated: development, tissue-specific, constitutive, and inducible. Not all P450s are inducible, but the other three factors of this multifactorial regulatory scheme probably apply to all eukaryotic P450s. We do not yet have a good view of the multifactorial regulation of most P450 enzymes, the steroid hydroxylases being the best understood. In addition to peptide hormone–dependent regulation through cAMP, developmental/tissue-specific regulation is dependent on an orphan nuclear

receptor transcription factor known as SF-1, which also plays a key role in organogenesis of the adrenal and gonads. Constitutive regulation of steroid hydroxylase levels is mediated by growth factors and cytokines.

Many forms of P450 are induced by substrates. Camphor, the substrate for P450cam, regulates the level of this P450 and other enzymes involved in camphor metabolism. It has been noted that xanthotoxin produced by carrot and parsley plants induces P450 levels in certain butterfly larvae, permitting them to metabolize this toxin and therefore feed on the plants. Drugs and pollutants that serve as substrates for many P450s can function to induce the levels of the specific forms that detoxify them. In a well-studied example, polycyclic aromatic hydrocarbons, some of which are substrates, induce the level of P4501A1 through a specific receptor that binds a unique DNA sequence associated with CYP1A1 when the ligand (polycyclic aromatic hydrocarbon) is bound. The same regulatory process for P4501A1 exists in fish and all animals, raising the possibility that there is an endogenous ligand for this receptor. Polyacrylic aromatic hydrocarbons derived from burning organic matter must have been in the environment for a very long time. Modern polycyclics such as dioxin are potent inducers of this P450 by this receptor-mediated mechanism, which probably has existed for millions of years. Therefore, the question remains as to whether P4501A1 might be regulated by a naturally occurring compound that might also be a substrate for this P450. Barbiturates such as phenobarbital enhance levels of certain forms of drug-metabolizing P450s, while peroxisome proliferators such as clofibrate induce levels of yet other P450s. Pharmaceutical companies not only need to know what forms of P450 metabolize potential new drugs and what the resultant metabolites are but they also need to be aware of the forms of P450 that are induced by such compounds, because alteration in the P450 profile by induction can alter the pattern of drug metabolism. Overall, the levels of P450s in all organisms are regulated by environmental conditions. In higher organisms, developmental, tissue-specific, and endogenous systems also contribute to the existing levels of P450s. Since there is little post-translational modification of P450s that influences enzymatic activity, levels of the monooxygenases themselves must be key in controlling activity. Transcription is a very important player in this aspect of P450 function, although translation that has not been studied as extensively as transcription may also be important.

6 Future Directions

This section, in 1996, began with three questions that were considered to be particularly important from the standpoint of the pharmaceutical industry.

- How many human forms of P450 are there?
  From the human genome we now know that there appear to be 57 human CYPs.
- What are the overlapping substrate specificities of the major P450 forms to a variety of drugs?
  Of course, this is a question that will forever be an ongoing one because of the development of new drugs. However, we now know the number of CYPs and, in most cases, they can be expressed in heterologous systems providing sufficient enzyme for analysis of substrate specificity. Thus, the methods are clearly

available to test the ability of different human CYPs to metabolize a common substrate. Most often, this analysis will not include all 57 human CYPs but rather the 6 major drug-metabolizing forms in human liver (3A4, 2C9, 2C19, 2E1, 2A1, 2D6).

- How are the major human forms of P450 regulated?
  This has been a rapidly growing area of research over the past seven years. Particularly, discovery of transcription factors such as PXR, CAR, LXR, along with further understanding of the mechanism by which AhR functions to induce CYPs through polycyclic aromatic hydrocarbons, have begun to bring rationale answers to this question.

The other two questions in 1996 are

- What genetic and environmental factors control the individual levels of these enzymes, and can noninvasive screening tests be developed to determine an individual's P450 profile?
- Can designer drugs be developed for targeting specific P450 active sites?

We do not have clear answers to these at this time. However, development of microarray technologies and methods for solving the structures of human CYPs should assure some answers to these questions by the time the next volume is written. Many of the same questions are also being asked by the agricultural industry in its search for new herbicides and pesticides.

The role of the P450 superfamily in growth and development will be one of the major issues addressed in academic laboratories during the next decade, along with the elucidation of the three-dimensional structure of eukaryotic P450s and the biochemistry of the regulation of *P450* gene expression. In addition, questions that remain to be answered focus on substrate accessibility and binding, particularly to P450s of the endoplasmic reticulum, and association of these P450s with limiting amounts of P450 reductase in the membrane.

One of the major thrusts during this decade will be in application of the unique chemistry catalyzed by P450s for commercial purposes. The recent success in producing cortisol by insertion of a mammalian P450–dependent metabolic pathway in yeast is very exciting and opens the door to other future efforts. At the same time, the solution of the *Streptomyces coelicolor* genome demonstrating the presence of 18 *CYP* genes opens the way to modification of P450s leading to novel secondary metabolites. The opportunity will soon exist to make novel and more potent antibiotics by modifying specific *CYP* gene(s) in particular streptomycetes. The use of microorganisms for expression of specific forms of P450 may also prove important in pollutant control. Organisms engineered to metabolize compounds such as polycyclic aromatic hydrocarbons may be very important in protecting the environment. In the context of fine chemical or pharmaceutical synthesis and in pollution control, the ability to change the catalytic activity of a specific P450 could be particularly useful. The realization of designer P450s engineered to catalyze specific activities may not be far in the future, since solution of the three-dimensional structure of eukaryotic P450s would seem to be a lesser obstacle than before.

Finally, transgenic plants expressing P450s that provide protection against insects and herbicides are presently being tested. Also, transgenic plants are being developed to modify flower color and for the production of commercially important

oils. Efforts to produce blue roses by modifying the P450 profile are under way. The flavor spearmint is produced by a plant much less susceptible to disease than that producing peppermint. Knocking out one *P450* gene and inserting another should in principle generate peppermint flavor in the spearmint plant, thereby lowering the cost of peppermint flavorings.

Many fundamental details concerning P450 systems at the biochemical and cell biology levels remain to be elucidated, and we can anticipate continued vigorous activity in these areas. The discovery of new forms of P450, particularly in plants and insects, will proceed at a rapid pace. Finally, applications of P450-catalyzed chemistry to commercial projects, including synthesis of chemicals, pollution control, and agriculture, are becoming important directions of this area of investigation.

See also Bioorganic Chemistry; Bioinorganic Chemistry.

Bibliography

Books and Reviews

Johnson, E.F., Waterman, M.R. (Eds.) (1996) *Methods in Enzymology* Vol. 272. *Cytochrome P450* (Part B). Academic Press, Orlando, FL.

Johnson, E.F., Waterman, M.R. (Eds.) (2002) *Methods in Enzymology*, Vol. 357. *Cytochrome P450* (Part C). Academic Press, Orlando, FL.

Nelson, D.R., Kamataki, T., Waxman, D.J., Guengerich, F.P., Estabrook, R.W., Feyereisen, R., Gonzalez, F.J., Coon, M.J., Gunsalius, I.C., Gotoh, O., Okuda, K., Neberg, D.W. (1993) The P450 superfamily. Update on new sequences, gene mapping, accession number, early trivial names of enzymes, and nomenclature, *DNA Cell. Biol.* **1**, 1–51.

Ortiz de Montellano, P. (1995) *Cytochrome P450: Structure, mechanism and biochemistry* (2nd Edition). Plenum Press, New York.

Ortiz de Montellano, P. (2003) *Cytochrome P450: Structure, mechanism and biochemistry* (3rd Edition). Plenum Press, New York (in preparation).

Primary Literature

Guengerich, F.P., et al. (2002) Cytochrome P450 enzymes in the generation of commercial products, *Nat. Rev. Drug. Discov.* **1**, 359–366.

Hasemann, C.A., Kurumbail, R.G., Bodupalli, S.S., Peterson, J.A., Deisenhofer, J. (1995) Structure and function of cytochrome P450: Comparative analysis of three crystal structures, *Structure* **3**, 41–62.

Williams, P.A., Cosme, J., Sridhar, V., Johnson, E.F., McRee, D.E. (2000) Mammalian microsomal cytochrome P450 monooxygenase: Structural adaptations for membrane binding and functional diversity, *Molecular Cell* **5**, 121–131.

Cytokines: Interleukins

Anthony Meager
Division of Immunobiology, National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Herts, UK

1 Background and Historical Perspective 118

2 Interleukin Proteins and Genes 121
2.1 Interleukin Proteins 121
2.2 Interleukin Genes 123

3 Interleukin Receptors 127
3.1 Receptor Structure 127
3.2 Intracellular Signaling Pathways 131

4 Biological Activities of Interleukins 133
4.1 Interleukin-1 134
4.2 Interleukin-2 135
4.3 Interleukin-3 135
4.4 Interleukin-4 137
4.5 Interleukin-5 138
4.6 Interleukin-6 139
4.7 Interleukin-7 139
4.8 Interleukin-8 140
4.9 Interleukin-9 140
4.10 Interleukin-10 140
4.11 Interleukin-11 141
4.12 Interleukin-12 141
4.13 Interleukin-13 142
4.14 Interleukin-14 142
4.15 Interleukin-15 142
4.16 Interleukin-16 142
4.17 Interleukin-17 142

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

4.18 Interleukin-18 143
4.19 Interleukin-19 143
4.20 Interleukin-20 143
4.21 Interleukin-21 143
4.22 Interleukin-22 143
4.23 Interleukin-23 144
4.24 Interleukin-24 144
4.25 Interleukin-25 144
4.26 Interleukin-26 144
4.27 Interleukin-27 144
4.28 Interleukin-28 144
4.29 Interleukin-29 145

5 Interleukin Physiology 145

6 Pathophysiology and Disease Correlates 147

7 Clinical Uses of Interleukins 148

8 Concluding Remarks 150

Acknowledgments 150

Bibliography 151
Books and Reviews 151

Keywords

Cytokine
One of a class of inducible biologically active proteins that exercise specific, receptor-mediated effects in target cells or in the cytokine-producing cells themselves.

Interleukin
One of several different cytokines acting between leukocytes and other cell types, which has a variety of stimulatory activities that regulate immune, inflammatory, and hematopoietic responses.

Hematopoietin
One of a subgroup of interleukins having an α-helical bundle structure.

Hematopoietin Receptor
One of a class of structurally related cell surface receptors for the hematopoietin subgroup of interleukins.

Hematopoiesis
The process of populating and replacing circulating erythrocytes and leukocytes from stem cells contained within the bone marrow.

Macrophage
Phagocytic leukocyte found in various tissues, which is important in nonspecific cellular immunity and antigen presentation, and which is a major producer of and responder to interleukins.

T lymphocyte (T cell)
A type of white blood cell capable of responding to foreign antigens and thus of mediating cellular immunity, which it does by secreting a variety of interleukins and other cytokines to activate leukocytes and other cells.

Homolog
A structurally related interleukin molecule whose activity is mediated by a receptor that is common to all homologs within an interleukin family.

Paralog
A structurally related interleukin molecule whose activity is mediated by a receptor distinct from those of other paralogs within an interleukin family.

The harmonious regulation of vital physiological processes, for example, replenishment of mature blood cells from bone marrow stem cells, termed *hematopoiesis*, and the activation of defense mechanisms against pathological microbes and injury has been shown to depend on the production and action of a variety of secreted biologically active proteins, collectively known as cytokines. Central among cytokines is a class of mediators, largely involved in the regulation of immune, inflammatory, and hematopoietic functions, designated interleukins primarily on the basis that interleukins are produced mainly by leukocytes and act locally on other leukocytes in surrounding tissues. Each interleukin, of which there are now 29 designated ones, exercises a spectrum of biological activities via specific cell surface receptors. Overlaps of biological activities have been found to be common among different interleukins and to result in many cases from the sharing of receptor components. On binding their cognate interleukins, receptors activate intracellular signaling pathways leading to the transcription of nuclear genes and expression of proteins necessary to commit the cell to a number of contingent events and responses according to the particular interleukin bound and cell type. Interleukins, however, probably rarely act alone and *in vivo* form complex interactive networks both among themselves and with other cytokines. Such complicated intercellular communications systems have made it difficult to precisely define the biological roles of interleukins in health and disease. In certain diseases, for example, cancer, some exogenously administered interleukins induce beneficial responses, but it has

rarely been possible to dissociate their desirable pharmacological activities from their undesirable pharmacological activities, which often give rise to severe side effects. In fact, many disease symptoms, for example, fever, hypotensive shock, have been demonstrated to be strongly associated with the presence of endogenously produced interleukins, and this is leading to clinical evaluation of several interleukin antagonists in both acute and chronic diseases.

1 Background and Historical Perspective

From the beginning of the twentieth century, there has been a growing realization that many of the biological processes in multicellular organisms are regulated by extracellular factors. Together such factors constitute an elaborate, interactive communication system that governs the cellular and physiological responses at several levels. In particular, the immune and neuroendocrine systems of higher animals, such as mammals, are subject to regulation by secreted factors that can act both locally (paracrine action) and at a distance (endocrine action) on cells bearing cognate receptors. The word "cytokine" was coined in 1974 by Dr Cohen to describe any soluble factor produced by both lymphoid and nonlymphoid cells that exercise specific effects in its target cells. Since then, the definition of cytokine has become restricted to nonendocrine, biologically active, protein mediators involved in (1) cell proliferation and thus tissue development and repair and (2) cellular function, which is required for the maintenance of homeostatic and defense mechanisms. Cytokines may be perceived to be analogous to polypeptide hormones except, in contrast to hormones, cytokines mainly act locally rather than at a distance. However, in many respects the distinction between cytokines and hormones is blurred.

In today's classification, interferons, interleukins, colony stimulating factors, and polypeptide growth factors are considered to belong to the cytokine superfamily. The word "interleukin" was coined in 1979 at the 2nd International Lymphokine Workshop in Ermatingen, Switzerland (quite possibly interleukin was born from a corruption of Interlaken, another town in Switzerland!), to apply to soluble mediators produced by activated T lymphocytes that acted in a paracrine fashion on "responder" lymphocytes. However, the definition of interleukin was quickly widened to include soluble mediators produced by a wide variety of cell types that acted primarily on lymphocytes and other cell types within the immune system. Regrettably, prior designation of some of the earlier characterized cytokines, for example, interferon gamma (IFNγ), has effectively excluded them from the interleukin family to which they truly belong, and the inclusion/exclusion of newly discovered cytokines to the interleukin family has often been fairly arbitrary.

Historically, the biological activities that are now attributed to interleukins were first uncovered as early as the 1940s. Then, for example, a fever-inducing substance isolated from "neutrophils" was called *granulocyte pyrogen* (GP); it had similar properties to endogenous pyrogen (EP), a substance isolated from the blood of rabbits made febrile by the injection of

bacteria. Since then and particularly in the late 1960s and early 1970s, many other biological activities have been defined, being originally ascribed to poorly characterized, soluble factors. Thus, lymphocyte activating factor (LAF) was the name given by immunologists to a macrophage-derived substance that enhanced mitogen-driven lymphocyte proliferation. A substance secreted by such mitogen (lectin)-driven lymphocytes was shown to specifically support the proliferation of T lymphocytes and was therefore named *T-cell growth factor* (*TCGF*). Subsequently, a number of other growth and differentiation factors acting on specific cell lineages were described.

By the late 1970s, it was beginning to be realized that some of these soluble factors, which had been separately designated, were in fact either similar or identical substances. For example, the production of EP and LAF could be shown to be stimulated by the same agents and their molecular weights (mw), isoelectric points (pI), and other properties were broadly similar. Their purification initially showed that they contained similar active, but heterogenous, proteins with pIs ranging between 5 and 8. Following the coining of "interleukin" in 1979, LAF was redesignated as interleukin-1. TCGF became interleukin-2. This naturally started the quest for "discovering" more interleukins and during the 1980s, with the major advance in rDNA technology, many "new" interleukins were cloned. Some of these were clearly identical to specific cell growth factors, which had been partially characterized earlier; others turned out to be novel biologically active proteins. The new and "old" names for interleukins are summarized in Table 1. New interleukins continue to be identified right up to the present day.

Once interleukins were cloned and made available in large quantities via production in recombinant bacteria, their biochemistry and biology could be and has been thoroughly investigated. A common characteristic that has emerged from such studies is that most, if not all, interleukins have more than one biological activity and there is frequently an overlap between the spectrum of activities of one particular interleukin and another interleukin or a member of the cytokine superfamily, such as an interferon or a colony stimulating factor. This indicates a significant redundancy in the biological functions of interleukins and of cytokines in general. Such redundancy suggests that interleukins trigger convergent signaling pathways in cells, and the more recent cloning of interleukin receptors has shed some light on why this should be so. For instance, it is now known that some interleukins share nonspecific receptor components besides those components that are necessary for specific interleukin binding. Furthermore, while interleukin receptors do not contain cytoplasmic kinase domains themselves, it appears increasingly likely that they are associated with a limited number of nonreceptor kinases, which mediate the phosphorylation of nuclear transcription factors. The latter are activated by this process to bind to interleukin-responsive elements of interleukin-inducible genes and activate their transcription. The subsequent expression of interleukin-inducible gene mRNAs and protein synthesis leads to cellular responses. If two (or more) interleukins trigger the induction of a common set of genes, then the observed cellular response, for example, mitogenesis, will probably be similar. However, cells have regulatory mechanisms that control the expression of interleukin receptors and thus their responsiveness to particular interleukins. Some interleukin receptors are restricted to specific cell types or lineages,

Tab. 1 Interleukins and former nomenclature.

Interleukin	*Former names*
Interleukin-1α (IL-1α)	Endogenous pyrogen (EP); lymphocyte activating factor (LAF); hemopoietin-1 (HP-1)
Interleukin-1ß (IL-1ß)	As for IL-1α; osteoclast-activating factor (OAF)
Interleukin-1 receptor antagonist (IL-1ra)	–
Interleukin-2 (IL-2)	T-cell growth factor-1 (TCGF-1); killer helper factor (KHF)
Interleukin-3 (IL-3)	Multiple colony stimulating factor (multi-CSF); hemopoietic cell growth factor (HCGF)
Interleukin-4 (IL-4)	B-cell stimulating factor-I (BSF-I); T-cell growth factor-II (TCGF-II); macrophage activation factor (MAF)
Interleukin-5 (IL-5)	T-cell replacing factor-I (TRF-I); eosinophil differentiation factor (EDF); IgA-enhancing factor
Interleukin-6 (IL-6)	B-cell stimulating factor-2 (BSF-2); hybridoma-plastocytoma growth factor (HPGF); hepatocyte stimulating factor (HSF)
Interleukin-7 (IL-7)	Lymphopoietin-1 (LP-1); thymocyte growth factor (THGF); pre-B-cell growth factor (PBGF)
Interleukin-8 (IL-8)[a]	Macrophage-derived neutrophil chemotactic factor (MDNCF); neutrophil activating factor-1 (NAF-1)
Interleukin-9 (IL-9)	T-cell growth factor-III (TCGF-III); p40; mast-cell enhancing activity
Interleukin-10 (IL-10)	Cytokine synthesis inhibitory factor (CSIF)
Interleukin-11 (IL-11)	Adipogenesis inhibitory factor (AGIF)
Interleukin-12 (IL-12)	Natural killer stimulatory factor (NKSF); cytotoxic lymphocyte maturation factor (CLMF)
Interleukin-13 (IL-13)	P-600
Interleukin-14 (IL-14)	High molecular weight B-cell growth factor (HMW-BCGF)
Interleukin-15 (IL-15)	–
Interleukin-16 (IL-16)	Lymphocyte chemoattractant factor (LCF)
Interleukin-18 (IL-18)	Interferon gamma inducing factor
Interleukin-22 (IL-22)	IL-10-related T-cell-derived inducible factor (IL-TIF)
Interleukin-24 (IL-24)	Melanoma differentiation antigen-7 (MDA-7)
Interleukin-26 (IL-26)	AK155

[a]IL-8 is a member of the small cytokine or chemokine superfamily. There are two main subgroups distinguished on the basis of the positions of cysteine residues. IL-8 is a member of the CXC Ligand (CXCL) subgroup, which also includes platelet factor 4 (PF4), ß-thromboglobulin (ß-TG), gro-α, gro-ß, IP-10, and so on. The CC Ligand (CCL) subgroup includes monocyte chemoattractant protein-1 (MCP-1), RANTES, macrophage inflammatory proteins (MIP-1α MIP-1ß), I-309, and so on.

and their expression may also change according to the stage of cell development or differentiation.

Overall, the regulation of cellular responses by interleukins is complex, and probably even more so where they act as mixtures of interleukins or with other cytokines or other noncytokine mediators. The complex interaction of interleukins and other mediators, which is envisaged to occur *in vivo*, has often obscured the physiological roles of interleukins. However, do they have any major physiological roles or do they have one or more accessory or subordinate roles? The fact that most interleukins are inducibly, rather

than constitutively, produced and mainly affect the working of the immune system, suggests they have been evolved primarily for triggering host defense mechanisms against infectious microorganisms. They have been clearly shown to regulate both cell-mediated and humoral immunity. As such, they could also be important in antitumor mechanisms and in chronic degenerative diseases.

The role(s) of many interleukins in preventing or combating infectious or invasive diseases has been supported by numerous studies carried out in experimental animal model systems. The antitumor effect of interleukin-2 (IL-2) has looked, for example, to be the most efficacious in causing tumor regression in allogenic and xenogenic tumors in mice. These studies have stimulated clinical interest in interleukins, and in recent years many clinical trials to evaluate interleukins as anticancer agents have been carried out. IL-2 has been used extensively, but as with IFNα, its usefulness in treating cancer has appeared limited. It also causes severe side effects. Less is known generally about what clinical use other interleukins might offer; however, some such as IL-1 look to be too toxic. It is thus not at all certain that the therapeutic application of individual interleukins, which now number up to interleukin-29 (IL-29), will significantly affect the outcome or the management of clinical diseases.

2 Interleukin Proteins and Genes

2.1 Interleukin Proteins

There are now 29 distinct biologically active mediators that have been classified as interleukins (Table 1). However, the actual number of molecularly distinct proteins is larger than 29 because some interleukins are comprised of more than one molecular species, for example, IL-1, IL-8, IL-17. A comparable situation exists for the type I interferon (IFN) family, where for human IFNα in particular, there are 12 related molecular species known as subtypes. In the case of IL-1, initially three structurally related proteins, two of which are biologically active, that is IL-1α and IL-1ß, and the third, which is an inhibitor of IL-1α/ß actions, known as IL-1 receptor antagonist (IL-1ra), were identified. The latter is the only known interleukin without agonistic activity and which behaves as a competitive inhibitor of an interleukin. More recently, seven more structural homologs of IL-1 have been described, including one that is more commonly known as IL-18. Other examples of structurally related interleukins that nevertheless are designated as differently numbered interleukins, for example, IL-10, IL-19, IL-20, IL-22, IL-24, and IL-26, (IL-28, IL-29), are now known. In one or two cases, the interleukin name has proved less appropriate, for example, IL-8, which is now clearly just one of a large family of chemokines.

The interleukins exhibit a wide variety of primary structures, sizes, and post-translational modifications. At the tertiary level, however, the interleukins generally fall into but a few categories. Excluding the IL-1 and IL-8 families, the majority of interleukins are α-helical proteins, which although unrelated in amino acid sequences, fold up as 4-6-α-helix bundles. This α-helical bundle structure is common to IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-9, IL-10, IL-11, IL-12, IL-13, IL-15, and most of the newly described interleukins (Fig. 1). The structures of the

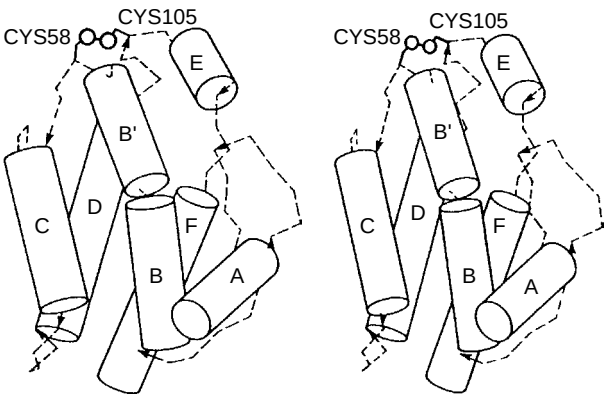

Fig. 1 Schematic stereo drawing of IL-2; helices are represented as cylinders and are lettered sequentially from the N-terminus. (Reprinted with permission from Brandhuber et al. (1987) *Science* **238**, 1707. Copyright 1987 by American Association for the Advancement of Science.)

receptors for these interleukins also have common features (discussed in Sect. 3.1). These interleukins have sometimes been referred to as *hematopoietins* and their receptors as hematopoietin receptors (HR). Interestingly, the hematopoietin family extends to other noninterleukin mediators such as erythropoietin (EPO) and growth hormone (GH) as well as a number of cytokines, including granulocyte colony stimulating factor (G-CSF), granulocyte-macrophage colony stimulating factor (GM-CSF), leukemia inhibitory factor (LIF), oncostatin M (OSM), and ciliary neurotrophic factor (CNTF). In contrast, IL-1ß and other IL-1 family members are nonhelical proteins with about 60% of residues in ß-strands, in a compact configuration known as a ß-trefoil (Fig. 2). IL-1α and ß and IL-18 are also unlike the other interleukins in having long N-terminal leader sequences (prodomains) rather than the recognizable signal polypeptides of secreted interleukins. Their processing and the exit of mature IL-1α, IL-1ß, or IL-18 from cells is therefore quite different from the other interleukins. A specific enzyme, a cysteine protease originally called IL-1ß converting enzyme but now designated caspase 1, which cleaves pro-IL-1ß to release the mature (active) IL-1ß protein, has been characterized. The IL-8 protein has a structure that incorporates both α-helical and ß-strand domains. The three-dimensional structures of the latest additions to the interleukin family are not known in detail, but are predicted to be mainly α-helical.

Another variable characteristic of interleukins is whether they are active in the monomeric or dimeric configuration. Most are active as monomers, but IL-5, IL-8, IL-10, (and IL-19, IL-20, IL-22, IL-24, IL-26) and IL-12 (and IL-23, IL-27) are dimers. Glycosylation is also a variable feature of interleukins; IL-1α and ß are not glycosylated, whereas most of the other interleukins have either O-linked and/or N-linked oligosaccharide side chains. The molecular characteristics of interleukins are summarized in Table 2.

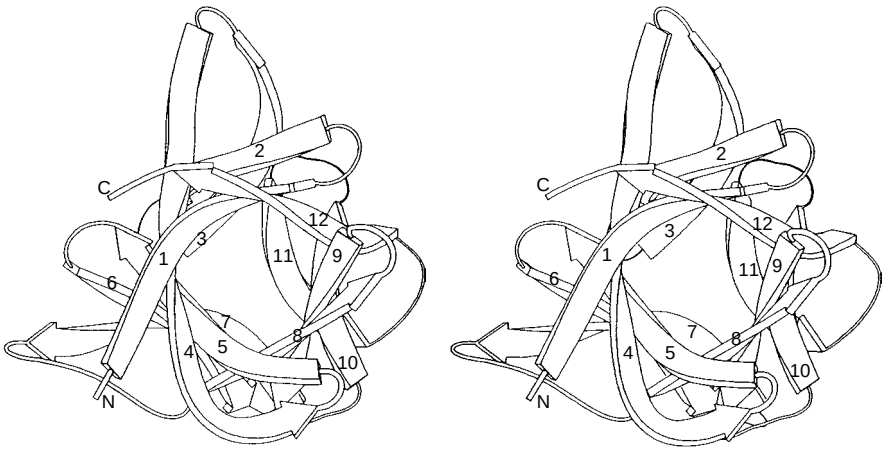

Fig. 2 Stereo cartoon of IL-1ß. The twisted arrows represent ß-strands and they are numbered sequentially from the N-terminal. The view is down the axis of the barrel formed by six of the ß-strands. (Reprinted with permission from Priestle et al. (1988) *EMBO J*, **7**, 339. Copyright EMBO/IRL Press Ltd, UK.)

2.2 Interleukin Genes

The fact that interleukin-like molecules can be found in invertebrates indicates that interleukin genes probably arose very long ago (>500 million years). The precise ancestry of interleukin genes is not known, but it is clear that they have been conserved throughout evolution from an apparent ancient origin. The structures of interleukin genes are complex, being composed of several exons and introns (cf the intronless type I IFN genes). For example, the human IL-1ß gene, which is located on the long arm of chromosome 2, contains 7 exons and 6 introns. However, many of the other interleukin genes share a common 4-exon-3-intron structure, for example, IL-2, IL-4, IL-5, which is also found in the GM-CSF and IFNγ genes. The gene for IL-6 has 5 exons and 4 introns, a feature it shares with the G-CSF gene. The interleukin genes are widely distributed among human chromosomes (Table 2), but there are several locations where interleukin families and their receptors are clustered, for example, the genes for IL-3, IL-4, IL-5, IL-9, IL-12p40, IL-13, GM-CSF, monocyte-colony stimulating factor (M-CSF) and its receptor are linked on the long arm (q) of chromosome 5. The equivalent site in the mouse genome is located on chromosome 11.

The expression of interleukin genes is dependent upon the activation of cellular transcription factors and the binding of these to the response elements located in the 5′ flanking regions of the coding DNA. The promoters and enhancers of interleukin gene transcription have been rather less studied than, for example, those of the IFNß gene. For instance, while IL-1α, IL-1ß, and IL-1ra genes are activated by common inducers such as bacterial lipopolysaccharide (LPS) and phorbol esters, there are cellular mechanisms that give rise to differential expression of their respective mRNAs. These mechanisms are poorly understood, but are probably cell type specific. Inducers of IL-1 genes,

Tab. 2 Molecular characteristics of human interleukins and their receptors.

Interleukin	*Precursor, no. of AA*	*Mature protein, no. of AA.*	*Glycosylation*	*Disulfide bonds*	*Monomer (M) or dimer (D)*	*Chromosome assignment of gene and organization (exon/intron)*	*Receptors*
IL-1α	271	159	–	0	M	2 (7/6)	80 kDa type I receptor
IL-1ß	269	153	–	1	M	2 (7/6)	68 kDa type II receptor
IL-1ra	177	152	N-linked		M	2	
IL-2	153	133	O-linked	1	M	4 (4/3)	55 kDa α-chain 75 kDa ß-chain 65 kDa γ-chain
IL-3	152	133	N-linked	1	M	5 (5/4)	80 kDa α-chain 120 kDa ß-chain
IL-4	154	129	N-linked	3	M	5 (4/3)	130 kDa binding protein; other component?
IL-5	134	114	N-linked	1	D	5 (4/3)	80 kDa α-chain 120 kDa ß-chain
IL-6	212	184	N-linked	2	M	7 (5/4)	80 kDa α-chain 130 kDa ß-chain - (gp130)
IL-7	177	152	N-linked	3	M	8	68 kDa binding protein
IL-8	99	72	–	2	D	4 (c-x-c) 17 (c-c) (4/3)	58 kDa binding proteins; two types at least
IL-9	144	126	N-linked	5	M	5 (5/4)	64 kDa binding protein

IL-10	178	160	Not glycosylated but contains one site of N-linked glycosylation	2 (5-cysteines)	D	1 (5/4)	110 kDa binding protein
IL-11	199	178	–	0 (2-cysteines)	M	19	150 kDa α-chain 130 kDa ß-chain - (gp130)
IL-12							Two similar chains
p35	253	197	N-linked	3 (7-cysteines)	D (Heterodimer)	3 5	related to gp130 named IL-12Rβ1 & IL-12Rβ2
p40	326	306	N-linked	5			
IL-13	132	112	N-linked	2	M	5	Two IL-13Rα chains (HR family) in combination with components shared with IL-4
IL-14	498	483	N-linked	?	M		
IL-15	162	114	N-linked	2	M	4	Specific IL-15Rα chain, with IL-2Rβ & IL-2Rγ shared with IL-2

A growing number of additional interleukins have now been characterized; however, by and large, these belong to the previously characterized interleukin families. For instance, several IL-1 homologs have been identified, the genes of which are all on chromosome 2 close to IL-1α, IL-1β, and IL-1ra genes; in contrast, the gene for IL-18, an IL-1 paralog, is located on chromosome 11. Genes for IL-10 paralogs IL-19, IL-20, and IL-24 are located at chromosome 1q32 along with the IL-10 gene, while those for IL-22 and IL-26 are located on chromosome 12q15, close to the IFNγ gene. IL-10 and related paralogs are all V-shaped homodimers with topology resembling that of IFNγ. Receptors for IL-10 family members are constituted by specific pairs of class II cytokine receptor chains. IL-28 and IL-29 are related to both IL-10 and type I IFN families; however, their genes resemble the intron–exon structure of IL-10-type genes rather than the intronless type I IFN genes and are located on chromosome 19q13, away from either IL-10- and paralog genes, located as above, or type I IFN genes on chromosome 9. The receptor for IL-28/29 is a distinct class II cytokine receptor pair composed of IL-10Rβ and CRF2-12 and different from type I IFN receptors. IL-21 is a 4-helix cytokine most closely to IL-15; IL-23 and IL-27 are heterodimers structurally related to IL-12, IL-23 sharing with IL-12 its p40 subunit and thus one of its receptors, IL-12Rβ1.

for example, LPS, phorbol myristate acetate (PMA), increase protein kinase C (PKC) activity, which in turn activates the nuclear transcription factors, NFκB and AP-1 (composed of the protooncogene products c-jun and c-fos). Binding motifs for NFκB and AP-1 are present in the promoter regions of IL-1 genes. It is probable that transcription of IL-1 genes is regulated by activated NFκB and AP-1, although presently it is not understood how their transcription is differentially regulated. Other factors such as intracellular levels of cyclic AMP (cAMP) could be important in this respect. Increased concentrations of cAMP have been found to enhance IL-1ß expression, both at the transcriptional and posttranscriptional level. Interestingly, it is clear that IL-1α and ß can themselves induce IL-1 gene expression, probably utilizing the same PKC signaling pathway and NFκB/AP-1 transcription factors as other IL-1 inducers. However, in macrophages, LPS is by far the strongest inducer of IL-1 gene expression.

In contrast, IL-2 gene expression is dependent on mitogenic or antigenic stimulation of T lymphocytes. Gene activation requires at least three transcription factors, AP-1, and two others designated NFAT-1 and NFIL-2A, which themselves are subject to regulatory mechanisms. The signaling pathway via PKC appears to be involved in IL-2 gene induction. IL-1, which also utilizes the PKC pathway and is known to enhance IL-2 synthesis, probably acts to augment/amplify the cooperation of nuclear transcription factors at the level of the IL-2 gene. The promoter region of the IL-6 gene contains the NFκB and AP-1 binding motifs, similar to those found in IL-1 genes, plus a cAMP-responsive element (cRE).

It is known that other interleukin genes are responsive to the same or similar inducers that lead to IL-1 gene activation, for example, IL-6, IL-8, while others are responsive to the same inducers that lead to IL-2 gene activation, for example, IL-3, IL-4, IL-5, IL-7, IL-9, IL-10, and IL-13. Synthesis of the latter interleukins is almost entirely restricted to T lymphocytes, but regulation at the transcriptional level can determine which interleukins are produced. Mature T-helper (h) $CD4^+$ lymphocytes, for example, can be separated into two or more functional subsets. The Th1 subset, which is responsible for initiating responses against intracellular pathogens and delayed-type hypersensitivity, produces IL-2 and IFNγ, whereas the Th2 subset, which preferentially induces antibody-mediated responses, produces IL-4, IL-5, IL-6, IL-10, and IL-13. Other interleukins, for example, IL-3, and cytokines, for example, GM-CSF, tumor necrosis factor alpha (TNFα) may be produced by either subset (Table 3). More immature Th lymphocytes, designated Th0, appear to be able to produce an unrestricted range of interleukins and cytokines. The mechanisms by which interleukin genes are switched on and off in Th lymphocytes are not fully understood, but are partly dependent on differentiation and maturation signals mediated by other interleukins and cytokines, for example, IL-12, IFNγ, and pathogenic triggers. For example, allergens and certain parasite antigens stimulate the development of Th2 lymphocytes.

The production of interleukins is also likely to be subject to posttranscriptional and posttranslational controls. As mentioned earlier, the processing of IL-1α and ß is complex, involving specific enzyme cleavage of IL-1 31 kDa precursors and transport of the mature, active 17.5 kDa IL-1α and ß proteins through the cytoplasm to cell membrane (neither IL-1 precursors

Tab. 3 Interleukins and cytokines secreted by T-helper CD4$^+$ cell subsets and cytotoxic CD8$^+$ T lymphocytes.

Interleukin/cytokine	*Th1*	*Th2*	*Th0*	*Cytotoxic T lymphocytes (CTL)*
Interferon-γ (IFNγ)	++	−	++	++
IL-2	++	−	++	±
IL-3	++	++	++	+
IL-4	−	++	++	−
IL-5	−	++	++	−
IL-6	−	++	+	−
IL-9	?	?	+	?
IL-10	−	++	++	?
IL-13	−	++	?	?
GM-CSF	++	+	+	++
Tumor necrosis factor-α (TNFα)	++	+	?	+
Tumor necrosis factor-ß (TNFß)	++	−	?	+

nor mature forms appear to enter the endoplasmic reticulum as would a "normal" secreted protein). Some IL-1α appears to remain "anchored" in the cell membrane or the cell associated for several hours before being released, whereas IL-1ß is quickly "secreted". The processing of IL-1α and ß is unusual and for all of the other interleukins, including IL-1ra, processing, posttranslational modification, for example, glycosylation, and secretion occurs through the regular route for a protein containing an N-terminal signal sequence, that is, via the endoplasmic reticulum and the Golgi apparatus.

The expression of all the interleukin genes is tightly regulated. Interleukin mRNAs are usually made for only short periods of time following induction. The interleukin genes then become hyporesponsive to further stimulation, probably due to increases in the level of repressor complexes. Synthesis of interleukins is also of a relatively short duration, but can depend on the stability of the mRNA. In addition, the level of synthesis will be affected (modulated) by the presence of other active regulatory agents [(for example, immunosuppressive hormones (prostaglandins, glucocorticoids) and cytokines (transforming growth factor beta TGF-β)].

The major cell sources and inducers of particular interleukins are detailed in Table 4.

3 Interleukin Receptors

3.1 Receptor Structure

As discussed in Sect. 2.1, many interleukins have similar α-helical bundle structures and therefore it is not surprising that their cell surface receptors also share common structural features, particularly in their N-terminal extracellular binding domains. The receptors for IL-3, IL-4, IL-5, IL-6, IL-7, IL-9, IL-12, IL-13, IL-15, and IL-21 have been classified under the generic name of hematopoietin receptor (HR). Alternatively, they are known as type I cytokine receptors. Two

Tab. 4 Major cell sources and inducers of interleukins.

Interleukin	*Cell source*	*Inducers*
IL-1α/ß and IL-1ra	Monocytes/macrophages; endothelial cells; fibroblasts; most cell types	Bacterial lipopolysaccharide (LPS); phorbol esters; calcium ionophore; muramyldipeptide; IL-1α/ß; TNFα
IL-2	T cells	Mitogens, e.g. plant lectins and bacterial enterotoxins; antigens
IL-3	T cells; mast cells; thymic epithelial cells; keratinocytes	Mitogens and antigens (T cells) antibody cross-linking of FcR for IgE (mast cells)
IL-4	T cells; mast cells; basophils; B cells; bone marrow stromal cells	Mitogens and antigens; antibody cross-linking of FcR for IgE
IL-5	T cells; mast cells	Mitogens and antigens; antibody cross-linking of FcR for IgE
IL-6	T cells; macrophages; fibroblasts, hepatocytes; endothelial cells	LPS; phorbol esters; IL-1α/ß; TNFα/ß; calcium ionophore
IL-7	Bone marrow stromal cells; fetal liver cells	Cytokines; TGFß, PDGF and IL-1α/ß
IL-8	Wide variety of cell types, including macrophages; T cells; endothelial cells	LPS; phorbol esters; IL-1α/ß; TNFα/ß; calcium ionophore
IL-9	T cells	Mitogens and antigens
IL-10	T cells; macrophages; B cells; keratinocytes	Mitogens and antigens; LPS; EBV (transformed B cells)
IL-11	Stromal fibroblasts; trophoblasts	Cytokines; TGFß, PDGF and IL-1α/ß
IL-12	Macrophages; dendritic cells	LPS; mitogens; phorbol esters
IL-13	T cells	Mitogens and antigens
IL-14	T cells	Mitogens and antigens
IL-15	Macrophages;epithelial cells	Type I interferons
IL-16	T cells; monocytes; eosinophils	?
IL-17	T cells	Mitogens and antigens
IL-18	Macrophages	LPS
IL-19	PBMC; T- and B-cell lines	LPS; GM-CSF
IL-20	Keratinocytes; skin cells	?
IL-21	T cells	Mitogens and antigens
IL-22	T cells	LPS; IL-9
IL-23	Macrophages; dendritic cells	LPS
IL-24	Th2 cells; melanoma cell lines	IFNβ and phorbol esters
IL-25	Th2 cells; mast cells	Mitogens; antigens; IgE
IL-26	T cells	Herpesvirus transformation
IL-27	Macrophages; dendritic cells	LPS
IL-28	Dendritic cells	Virus infection; poly (I).(C)
IL-29	Dendritic cells	Virus infection; poly (I).(C)

components of the high-affinity IL-2 receptor also belong to the HR family as do the receptors for G-CSF, GM-CSF, EPO, GH, LIF, CNTF, and OSM. In addition, the two nonspecific receptor chains that are required for IL-3, IL-5, and GM-CSF receptor function and for IL-6, IL-11, LIF, CNTF, and OSM receptor function respectively, are HR family members.

The basic structural characteristics of HR are two extracellular domains of approximately 100 amino acids, the one closest to the cell membrane containing a WSXWS motif and the outer N-terminal domain containing a number of conserved cysteine residues (Fig. 3). As new members of the HR family were identified, it became clear that either (1) these two domains could be duplicated or (2) unrelated domains that were immunoglobulin-like or fibronectin-like could be added to extend the extracellular portion of the receptor. In the case of the GH receptor, which has the simplest two domain structure (Fig. 3), X-crystallography has recently shown that each of the 100 amino acid stretches consists of seven ß-strands that are folded to form a sandwich of two antiparallel sheets. However, only the N-terminal domain makes contact with the GH molecule. In fact, the GH molecule is able to bind to two GH receptors, thus effecting their dimerization, a process that is likely to be required for signal transduction via the C-terminal intracellular domains.

The G-CSF receptor probably functions in a similar manner. However, many of the interleukin receptors are more complex in that they are composed of more than one component (subunit). The IL-2 receptor, for example, has three subunits, the first (IL-2Rα) being a glycoprotein unrelated to the HR family and the second (IL-2Rß) and third (IL-2Rγ) being HR family members. All three subunits can bind IL-2 with low affinity, the complexes of α- and ß- and ß- and γ-subunits bind IL-2 with intermediate affinity, but only the complex of α-, ß-, and γ-subunits can form a high-affinity receptor for IL-2, leading to signal transduction and internalization of the ligand–receptor complex (IL-2Rγ has been shown to be a subunit of IL-4, IL-7, IL-13, IL-15, and IL-21 receptors).

In the case of IL-3, IL-5, and GM-CSF, a different receptor system has evolved whereby ligand binding to a ligand-specific receptor subunit (α-chain) induces the association of this complex with a nonspecific receptor subunit (ß-chain) that enables signal transduction to take place. Both α- and ß- chains are members of the HR family (Fig. 3). The α-chains have only small cytoplasmic domains, insufficient for signaling, and therefore the presence of the ß-chain, which does have a large cytoplasmic domain, is required for signal transduction. This receptor system explains why IL-3, IL-5, and GM-CSF share many biological activities since where cells express all three ligand-specific α-chains, signaling can only take place by interaction with the single common, nonspecific, ß-chain yielding the same cellular response.

A similar receptor system exists for IL-6 and IL-11 and the cytokines LIF, CNTF, and OSM. Here the nonspecific ß-chain has been characterized as a 130-kDa transmembrane glycoprotein (gp130) that resembles the G-CSF receptor in overall structure (Fig. 3). Only in the presence of gp130, which itself cannot bind IL-6, can a high-affinity receptor complex be formed by interaction of the IL-6 binding α-chain and signal transduction ensues. gp130 is also the ß-chain for IL-11, LIF, CNTF and OSM (gp130 can weakly bind OSM). The IL-6Rα chain may be cleaved from the cell membrane to form a soluble receptor for IL-6 (sIL-6Rα), and this can also interact with IL-6 and gp130 to trigger signal transduction. Interestingly, IL-12 has two components, a p40 subunit that is homologous to sIL-6Rα and a p35 subunit that resembles a typical helical cytokine. In effect, it is an interleukin that carries its own specific α-chain around

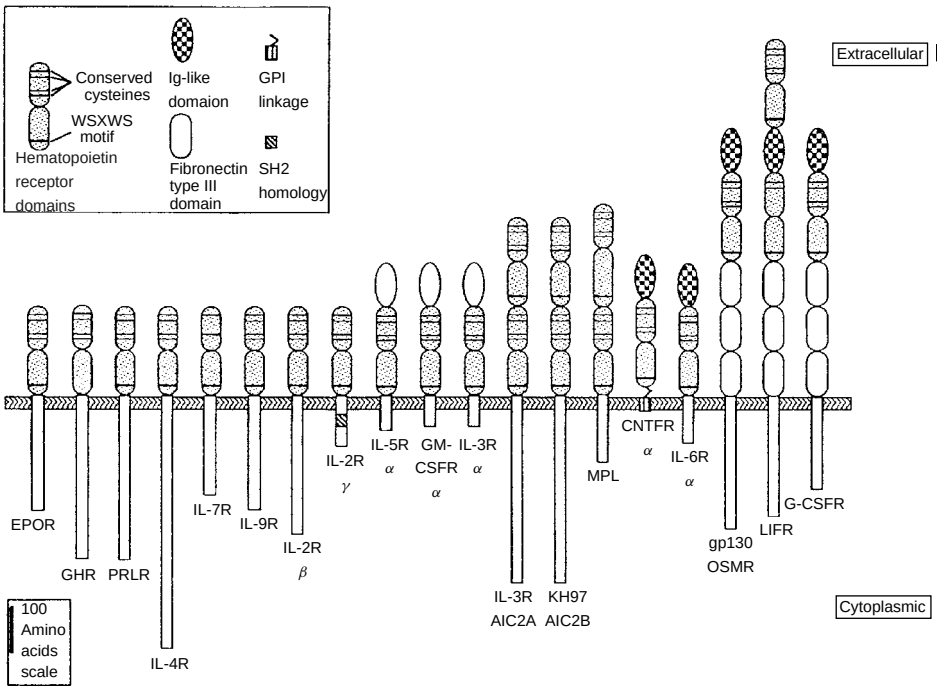

Fig. 3 Schematic receptor structures of the hematopoietin superfamily. Modular elements are as shown in the box. (Reprinted with permission of Cosman (1993) *Cytokine* **5**, 95 Copyright Academic Press Ltd.)

with it. The high-affinity receptor for IL-12 is composed of two related ß-type receptor chains, IL-12Rß1 and IL-12Rß2, each independently having low affinity for IL-12. IL-12 p40 interacts primarily with IL-12Rß1, while p35 interacts with IL-12Rß2. IL-23, which contains the p40 subunit in common with IL-12, is able to bind only to IL-12Rß1, and probably a second, as-yet unidentified, receptor subunit is required to form the high-affinity IL-23 receptor. The receptors for IL-10 and related interleukins (IL-19, IL-20, IL-22, IL-24, and IL-26) also have two subunits, a long chain with a large intracellular domain and a smaller accessory chain, which have been classified on a structural basis as type II cytokine receptors. The latter, of which 12 members (cytokine receptor family 2, CRF2-1 to CRF2-12) are now characterized, include receptors for interferons.

Unlike the two or three component receptors of most interleukins, there are two distinct single-chain receptors for IL-1 molecules. The first of these to be characterized, now known as the type I IL-1 receptor (IL-1RI), has widespread tissue distribution and is found on T cells, fibroblasts, keratinocytes, endothelial cells, synovial lining cells, chondrocytes, and hepatocytes. The second receptor, the type II IL-1 receptor (IL-1RII), is found on B cells, neutrophils, and bone marrow cells. The two receptors are about 40% related in their N-terminal extracellular domains, contain three immunoglobulin (Ig) loop structures and are therefore

included in the Ig superfamily rather than the hematopoietin superfamily. They have more recently been classified with Toll-like receptors, a family of 10 proteins involved in regulating innate immunity. The larger 80-kDa IL-1RI has an intracellular domain (217 amino acids) of sufficient size to effect signal transduction, whereas the smaller 68-kDa IL-1RII has only a short intracellular C-terminal tail and is probably not able to act as a signal transducer on its own. There is a large body of evidence demonstrating that IL-1RI is functional and is required to mediate IL-1 activities, but the biological significance of the "incomplete" IL-1RII is poorly understood. Recent evidence suggests IL-1RII acts as an inactive "decoy" receptor to bind excess IL-1ß. Expression of IL-1RI is inducible or highly regulated and this probably to some extent controls the responsiveness to IL-1α and ß, although it is known that low (5%) receptor occupancy is sufficient to trigger intracellular events. Both IL-1α and ß bind with similar affinity to IL-1RI and their binding is competitively inhibited by IL-1ra. Shortly after binding, the receptor-ligand complex is internalized and may be translocated to the nucleus.

The receptors for IL-8 and related molecules belong to a separate receptor family, that of the so-called G-protein-coupled receptor superfamily. These receptors contain seven membrane-spanning helices that couple to guanine nucleotide binding proteins (G-proteins). For IL-8, there are two distinct receptors, type I and type II (or type A and B), which are about 74% related in amino acids. It is probable that IL-8 and other members of the c-x-c subgroup of chemokines utilize these two receptors, whereas the c-c subgroup (includes MCP-1, RANTES, MIP-1α/ß) have recently been shown to bind to another G protein–coupled receptor, which has approximately 33% homology to the IL-8 receptors.

3.2 Intracellular Signaling Pathways

None of the interleukin receptors has an integral protein kinase. However, it is now known that they are associated through their intracellular domains, either directly or indirectly via adaptor proteins, with "nonreceptor" protein kinases to enable signal transmission to occur. For example, the intracellular domain of IL-1RI interacts with an adaptor molecule known as MyD88, which in turn couples to an "interleukin-1 receptor associated kinase" (IRAK). This serine kinase phoshorylates a signal transducer belonging to the tumor necrosis factor receptor associated factor (TRAF) family and this leads to the activation of nuclear transcription factors, such as AP-1 (jun/fos) and NFκB; these in turn activate IL-1 responsive genes. It is known that IL-1 can potentially induce the expression of a whole catalog of genes ranging from those of other interleukins and cytokines, for example, IL-2 to IL-8, GM-CSF, TNFα and its own family genes, that is, IL-1α/ß and IL-1ra to a wide variety of enzymes (for example, manganous superoxide dismutase, cyclooxygenase, tissue plasminogen activator, collagenase), oncogenes (*c-fos, c-jun, c-myc*), cell adhesion molecules (intercellular adhesion molecule-1 (ICAM-1), vascular cell adhesion molecule-1 (VCAM-1)) and the 25-kDa IL-2 receptor chain, IL-2Rα (tac antigen). Many of these proteins are only specifically induced in certain cell types. However, mechanisms underlying such differential expression of IL-1 inducible genes remain not well understood.

It appears probable that individual interleukin receptors can be coupled to one

or more intracellular signaling pathways, usually depending on the cell type. For example, IL-1α/ß through binding to IL-1R1 can evoke PKC activity in T cells and mouse NIH 3T3 fibroblasts, whereas in human foreskin fibroblasts, the cAMP-dependent protein kinase (PKA) is activated. In Th2 cells, both PKA and PKC are evoked by IL-1. It has been reported that IL-1 triggers the formation of several second messengers, including arachidonic acid metabolites via phospholipase A_2, prostaglandins via the cyclooxygenase pathway, and ceramide via sphingomyelin. Ceramide may be responsible for activating a serine kinase, distinct from PKC, which is involved in phosphorylating the epidermal growth factor (EGF) receptor and causing a lowering of the affinity of this receptor for EGF.

As mentioned above, one of the proteins induced by IL-1 is the IL-2Rα and this, when expressed at the cell surface in combination with the two other IL-2 receptor chains, IL-2Rß and IL-2Rγ, forms the high-affinity receptor for IL-2. In contrast to IL-1RI, high-affinity IL-2R appears not to connect to signaling pathways involving phosphatidyl inositol hydrolysis, Ca^{2+} mobilization, PKC or PKA. However, IL-2 stimulation is known to result in tyrosine phosphorylation of several cytoplasmic proteins. It is now known that the intracellular domains of IL-2Rß and IL-2Rγ respectively, bind via specific tyrosine residues, the Janus tyrosine kinases JAK1 and JAK3. The JAKs constitute a family of receptor-activated kinases, which phosphorylate the members of a distinct class of transcription factors known as *signal transducers and activators of transcription* (*STATs*). Once activated, STATs translocate to the nucleus to activate transcription of IL-2 responsive genes. Increased expression of several oncogenes, including *c-myc, c-myb, c-jun, c-fos,* and src-related protein tyrosine kinase (PTK), which are probably involved in cell proliferation, are found in IL-2 activated cells. Notably, other interleukin and cytokine genes are also activated, but their expression depends on the phenotype of the lymphoid cell. For example, the Th-lymphocyte subsets Th1 and Th2 each express a defined spectrum of interleukins and cytokines (Table 3).

The high-affinity receptors for IL-4, IL-7, IL-9, IL-15, and IL-21 all share the IL-2Rγ chain, and therefore bind JAK3 and function in a similar way to the IL-2R. For example, high-affinity IL-15R also requires IL-2Rß and thus intracellular signaling pathways are likely to be identical to those triggered by IL-2. However, IL-15Rα is different from IL-2Rα and more widely expressed, suggesting a broader activity profile for IL-15. Most other high-affinity interleukin receptors are composed of two, rather than three chains – one α- and one ß-chain. In the case of IL-3 and IL-5 receptors, which share a common ß-chain, the latter is associated with JAK2, which activates STAT5. Similarly, for IL-6 and IL-11 receptors, the ß-chain associates primarily with JAK1, although other JAK family kinases might also be involved in signal transduction. In contrast, for IL-10 receptors, and those for the related interleukins IL-19, IL-20, IL-22, IL-24, IL-26, IL-28, and IL-29, both α- and ß-chains are associated with JAKs, in particular with TYK2 and JAK1 and 2; a similar association of JAK2 and TYK2 with the two receptor chains of IL-12R and IL-23R is also found. It is evident therefore that due to the sharing of JAKS and STATs, the intracellular pathways from the majority of interleukin receptors will have much in common, and that as

a consequence activities can be shared among several interleukins. However, since interleukins are in the main locally produced intercellular mediators, such activities are dependent on the cells present in the vicinity of their release expressing appropriate receptors, having functional intracellular signaling pathways and having transcriptionally activatable "interleukin-responsive" genes.

For example, IL-6 is an inducer of acute-phase proteins in hepatocytes and this has proved to be a useful system for studying the regulation of transcription of acute-phase proteins. A nuclear transcription factor, NF-IL-6, has been identified in IL-6 stimulated hepatocytes that binds to the promoter regions of acute-phase protein genes. The NF-IL-6 transcription factor is highly homologous to a liver-specific transcription factor C/EBP, a member of the so-called basic leucine-zipper family. These leucine-zipper proteins are known to bind to DNA as dimers through a leucine-zipper structure that is required for dimerization and that an adjacent basic region makes direct contact with DNA. Serine phosphorylation of NF-IL-6 is required to activate it to bind to the IL-6-response elements in acute-phase protein genes. Most of these genes share a common consensus sequence (CTGGGAA(T)), which also appears to be important in IL-6-dependent acute-phase protein gene activation. The acute-phase proteins induced by IL-6 include C-reactive protein (CRP), serum amyloid A, haptoglobulin, α1-antichymotrypsin, fibrinogen. These are largely restricted to hepatocytes, but as IL-6 is active in many different cell types, it is expected that IL-6 stimulation will lead to the enhanced transcription of a number of other genes, for example, *c-jun, c-fos*.

Members of IL-8 family bind to G-protein-coupled receptors. Binding of IL-8 to these receptors present on neutrophils triggers a large rise in intracellular Ca^{2+} and activation of PKC, suggesting that these changes are involved in signal transduction leading subsequently to neutrophil activation and the chemoattractant response. G-proteins, or heterotrimeric GTP-binding regulatory proteins, have been clearly implicated in signal transduction; pertussis toxin (*Bordetella pertussis* islet-activating protein), which blocks the activation of certain G-proteins, inhibits the IL-8 stimulated chemoattractant response. The activated G-proteins have been found to activate phosphatidylinositol phospholipase C and Ca^{2+} mobilization, subsequent PKC activation, secretion of granular enzymes, and activation of respiratory burst, and the generation of superoxide anion.

4 Biological Activities of Interleukins

There are now 29 recognized interleukins, three colony stimulating factors (G-CSF, GM-CSF, and M-CSF), and many more cytokines, for example, TNFα, LIF, OSM, CNTF, transforming growth factor ß (TGF-β), stem cell or steel factor and interferons (IFN), to cite a noninclusive list, that can affect the proliferation, differentiation, and function of cells. With the number of these biologically active mediators now in the hundreds, the number of possible combinations is astronomic. *In vivo*, it is probable that interleukins and other cytokines form the basis of a complex interactive communication network with the overall cellular responses being dependent on integrated assimilation of multiple signals. *In vitro*, it has been rarely possible to

mimic the *in vivo* situation and experimentally a highly reductionist approach using isolated molecules, isolated cells, and cell lines has formed the basis for studying the biological activities of individual interleukins. Such work has provided an important base, but is unsatisfactory in the long run as the true roles of interleukins cannot be accurately predicted. Only studies in whole animals and intact tissues will provide answers on the physiological importance of interleukins.

This section will first concisely review the biological activities of interleukins, starting with IL-1, that have been defined from *in vitro* experimentation. However, the vast amount of literature information precludes a totally comprehensive survey. Secondly, some examples of interleukin interactions will be briefly described and thirdly, the *in vivo* biological roles of interleukins, where information is available, will be outlined.

4.1 Interleukin-1

An unusually large number of biological activities have been attributed to IL-1. Broadly speaking, the activities of IL-1α and IL-1ß are qualitatively the same, and both IL-1α and ß are antagonized by IL-1ra. Currently, the activities of other members of the IL-1 superfamily, excluding IL-18, are not known. In general, IL-1α and ß behave as the primary activators of cells, readying them for secondary stimuli. As such, they have been included in a category of mediators called *competence factors*. These are considered to be elements of cell activation pathways, which commit cells to a series of stimulatable, contingent events. In this context, IL-1 has been shown to augment antigen activation of T lymphocytes and to potentiate the proliferation of hematopoietic progenitors. IL-1 has also been demonstrated to induce the synthesis of secondary acting mediators, the so-called *progression factors* that stimulate cells to undergo further proliferative or differentiating events. For example, IL-1 induces platelet-derived growth factor (PDGF), a potent mitogenic factor, in fibroblasts, GM-CSF in endothelial cells, IL-2 in T lymphocytes and IL-1, IL-6, IL-8, and TNFα in monocytes/macrophages. In addition, IL-1 can induce or regulate the expression of receptors and other cell surface molecules, for example, IL-2Rα in T lymphocytes, ICAM-1, and VCAM-1 in endothelial cells. IL-1, at least *in vitro*, can stimulate acute-phase protein synthesis in hepatocytes, although it is not as potent as IL-6 (NB possibly the effects of IL-1 on hepatocytes occur as the result of intermediate production of IL-6). At relatively high doses, IL-1 induces bone and cartilage resorption *in vitro* cell systems. In contrast, at low doses, IL-1 may promote osteoblast proliferation and transiently stimulate collagen synthesis. IL-1 induces prostaglandin (PGE_2) synthesis by synovial cells.

In hematopoiesis, that is, the generation of mature blood cells from bone marrow stem cells, IL-1 can act as activator of early progenitors, and in combination with other interleukins, for example, IL-3, IL-6, it probably stimulates proliferation and differentiation of the various cell lineages. It may synergize with IL-6 for IL-2 synthesis by activated T lymphocytes, with IL-4 for B-cell activation and Ig isotype regulation, and with IL-2 or IFN for augmenting natural killer (NK) cell activity.

In vivo, IL-1 is known to induce fever (IL-1 was originally called *endogenous pyrogen* (*EP*)), sleepiness, and anorexia. IL-1 also has been shown to affect the neuroendocrine system, principally by

increasing pituitary adrenocorticotropic hormone (ACTH) and endorphin levels as well as glucocorticoids. Administration of IL-1 causes neutrophilia, probably due to IL-1-mediated induction of GM-CSF from endothelial cells. The latter are activated and sticky for lymphocytes and monocytes due to increased ICAM-1 expression. There are effects on liver function leading to increased amino acid turnover and hyperlipidemia. By itself, but particularly in combination with TNFα and/or bacterial LPS, high levels of IL-1 can lead to profound hypotension, myocardial suppression, shock, and death.

The biological activities of IL-1 and other interleukins are summarized in Table 5.

4.2 Interleukin-2

In contrast to IL-1, IL-2 is much more limited in its biological activities, due largely to the restricted expression of high-affinity IL-2 receptors to a relatively few cell types. Mitogen- or antigen-activated mature T lymphocytes express high-affinity IL-2R and subsequent interaction with IL-2 leads to clonal proliferation. Many T-cell lines and clones, specific for particular antigens, remain wholly or partially dependent on the presence of exogenous IL-2 for proliferation. In addition, antigen-independent murine cytotoxic T-lymphocyte line (CTLL) requires IL-2 for continuous growth. NK cells, contained within the large granular lymphocyte (LGL) population, express IL-2Rß and not only proliferate in response to IL-2 but also exhibit enhanced cytolytic activity. In the presence of high IL-2 concentrations, the so-called lymphokine-activated killer (LAK) cells emerge from resting populations of lymphoid cells. LAK cells, which are cytolytic for some tumor cells, can be induced *in vitro* and *in vivo*. IL-2 can also act on activated B lymphocytes to stimulate both their proliferation and the induction of Ig synthesis.

Besides its growth-promoting activity, IL-2 stimulates T cells to secrete a range of other interleukins and cytokines (Table 3). While the principal role of these interleukins and cytokines is to act as "helper factors" for the growth and differentiation of leucocytes other than T lymphocytes, they could also be involved in the differentiation of the latter. Thus, IL-2 may indirectly regulate T-lymphocyte differentiation, for example, in the maturation of CTL or LAK to express cytolytic activity.

In vivo, IL-2 induces lymphoid hyperplasia, for example, increases in mature T cells, neutrophilia, and eosinophilia; the latter are probably indirectly caused by IL-2 induced interleukin/cytokine synthesis by activated T cells. LAK cells appear rapidly following the infusion of IL-2. However, high doses of IL-2 induce undesirable side effects such as hypotension, oliguria, fluid retention, progressive dyspnoea (difficulty in breathing), atrial arrhythmias, thrombocytopenia, and vascular leak syndrome.

4.3 Interleukin-3

IL-3 has been called the pan-specific interleukin because of its highly pleiotropic activities. Like IL-2, IL-3 acts mainly as a specific growth factor, principally affecting the proliferation and differentiation of erythroid and myeloid lineages. It has been shown to be particularly effective in stimulating the proliferation of early erythroid/myeloid progenitors, for example, in the formation of granulocyte-erythroid-macrophage-megakaryocyte colony-forming units

Tab. 5 The biological activities of interleukins.

Interleukin	*Principal activities*
IL-1α/ß	– Activates mature T cells to produce IL-2 and express IL-2R. – Costimulant of proliferation and activation of B cells and hematopoietic progenitor cells. – Induces proinflammatory cytokines, e.g. TNFα, IL-6, and other mediators depending upon cell type. – Increases expression of endothelial cell adhesion molecules. – Induces neutrophil accumulation *in vivo*. – Induces fever and hypotension.
IL-1ra	– Inhibits the activities of IL-1α/ß by competing for IL-1R.
IL-2	– Stimulates T-cell proliferation and differentiation. – Activates NK and LAK cells. – Promotes proliferation of and Ig secretion in activated B cells. – Causes vascular leak syndrome *in vivo* at high doses.
IL-3	– Acts in combination with lineage-restricted cytokines to stimulate proliferation and differentiation of hematopoietic progenitors of macrophages, neutrophils, basophils, eosinophils, mast cells, megakaryocytes, and erythrocytes. – Supports the proliferation of early multipotential bone marrow stem cells.
IL-4	– B-cell growth and differentiating factor, induces IgE synthesis. – Generates Th2 cell subset from naive Th0 cell population. – Antagonist of IFNγ.
IL-5	– Stimulates generation of eosinophils from hematopoietic precursors. – Possible involvement in B-cell differentiation and Ig class switching.
IL-6	– Growth factor for many transformed and tumor cells. – Costimulant of proliferation in thymocytes and IL-3-dependent hematopoietic progenitors, and of IL-2 production in mature T cells. – Terminal B-cell differentiation factor. – Hepatocyte stimulating factor; inducer of acute-phase protein synthesis.
IL-7	– Supports the proliferation of immature B- and T cells. – Possible involvement in proliferation and activation of mature T cells.
IL-8	– Chemoattractant for neutrophils and T cells. – Stimulates neutrophil activation and degranulation.
IL-9	– Augments mast-cell proliferation response to IL-2. – Sustains antigen-independent growth of certain Th cells.
IL-10	– Inhibitor of cytokine/interleukin synthesis by Th1 cells. – Suppressant of macrophage functions, including down regulation of inflammatory cytokine production. – Enhances B-cell proliferation and Ig synthesis. – Growth factor for B-cell tumors.
IL-11	– Synergistic factor for IL-3-dependent proliferation of bone marrow progenitors (similar to IL-6 in this respect), particularly megakaryocyte colonies. – Can act on hepatocytes to induce acute-phase protein synthesis.
IL-12	– Stimulates the differentiation of naive Th0 cells into the Th1 cell subset and thereby is an initiator of cell-mediated immunity. – Induces IFNγ production with IL-18 as costimulant. – Stimulates the proliferation and activity of NK cells and mature T cells.
IL-13	– Induces B-cell proliferation and differentiation (similar to IL-4). – Inhibits inflammatory cytokine synthesis by monocytes/macrophages.

Tab. 5 (*continued*)

Interleukin	*Principal activities*
IL-14	– Inducer of proliferation of activated B cells (but not resting B cells). – Inhibits Ig secretion by mitogen-stimulated B cells.
IL-15	– Inducer of proliferation of T cells and T-cell lines. – Activates NK and LAK cells.
IL-16	– Lymphocyte chemoattractant factor; stimulates migratory response in $CD4^+$ lymphocytes, monocytes and eosinophils.
IL-17	– Induces IL-6 and IL-8 production in fibroblasts; neutrophil recruitment. – Enhances cell surface antigen and adhesion molecule expression.
IL-18	– Induces T-cell IFNγ production in combination with IL-12. – Enhances T-cell cytotoxicity and inflammatory responses.
IL-19	– Similar to IL-10, but poorly defined as yet.
IL-20	– Involved in skin differentiation and keratin expression.
IL-21	– Regulation of NK cell activation and differentiation.
IL-22	– Activates synthesis of acute-phase proteins in liver cells.
IL-23	– Similar to IL-12, but poorly defined as yet.
IL-24	– Induction of apoptosis and inhibition of proliferation in tumor cells.
IL-25	– Induces expression of Th2 type immunosuppressive interleukins.
IL-26	– Not known.
IL-27	– Similar to IL-12, but poorly defined as yet.
IL-28/29	– Induce antiviral activity.

(GEMM-CFU) and in the stimulation of erythroid burst-forming units (BFU-E) and megakaryocyte progenitors. There is some evidence that IL-3 also acts on very early multipotential progenitor cells. IL-3 may be considered to act as a progression factor following cell activation with competence factors, such as IL-1. IL-3 may also stimulate the proliferation/differentiation of more mature cells of the myeloid lineage, for example, by inducing macrophage precursors to express cytokine receptors, such as M-CSF-R. There is experimental data showing that IL-3 has a growth-supporting activity for murine mast cells, but it is not clear that this is so for human mast cells.

In vivo, release of IL-3 by injected WEHI-3B tumor cells results in the stimulation of all of the various types of hematopoietic cells predicted from the *in vitro* investigations, that is, increases in numbers of myeloid cells, erythroid cells, mast cells, and megakaryocytes. Administration of IL-3 to mice, primates, and humans yields broadly similar effects; the progenitors of mast cells, neutrophils, and macrophages are increased.

4.4 Interleukin-4

IL-4 was originally discovered by its action on B lymphocytes and was the first B-cell stimulating factor (BSF-1, Table 1) to be characterized. B lymphocytes, which express surface Ig, were found only to proliferate in response to anti-Ig if IL-4 (BSF-1) was present. It would appear that IL-4 is a B-lymphocyte competence factor, perhaps the counterpart of IL-1 for T lymphocytes. As such, IL-4 is involved in B-lymphocyte activation, rather than proliferation, resulting in, for example, the increased expression of

cell surface molecules, such as major histocompatibility complex (MHC) class II antigens, low-affinity IgE receptor (CD23) and surface IgM. IL-4 also regulates which Ig isotype is synthesized and secreted by mature B cells; it appears to be directly responsible for switching on IgG1 and IgE synthesis *in vitro*. This action is opposed by IFNγ, which preferentially induces the synthesis of IgG2a.

Additionally, IL-4 can act on other cell lineages. In particular, it drives the differentiation of naive T lymphocytes into the Th2-lymphocyte subset, possibly by blocking the synthesis of its own natural antagonist, IFNγ. In some antigen-independent T-cell lines, IL-4 can act as a proliferation signal, similar to IL-2. In contrast, IL-4 has been reported to antagonize the stimulatory actions of IL-2, for example, in LAK cells. IL-4 has also been shown to affect hematopoiesis, especially the development of mast cells and eosinophils from progenitor cells. It has a variety of actions in monocytes/macrophages, for example, upregulation of low-affinity IgE receptor (CD23), enhancement of MHC class II antigens, inhibition of spontaneous and induced IL-1, IL-6, IL-8, and TNFα synthesis, but stimulation of G-CSF and M-CSF, and in granulocytes, fibroblasts, epithelial, and endothelial cells. IL-4 has many activities in common with IL-13 with which it shares receptor components (see Sect. 4.13).

In vivo, IL-4 release results in the generation of large numbers of eosinophils (through induction of IL-5 production), mast cells, and macrophages. In transgenic mice, expressing IL-4 in the thymus and T cells, thymic hypoplasia, eosinophilia, and eye inflammation (possibly due to activated macrophages) have been found.

4.5 Interleukin-5

It is known that the control of eosinophil numbers is T cell dependent. In steady state hematopoiesis, relatively few eosinophils are found in the circulation. In the presence of IL-3 and GM-CSF, only small numbers of eosinophil colonies develop from GEMM-CFU progenitor cells. However, infection by parasites, such as helminths, dramatically increases eosinophil numbers and the factor responsible for eosinophil proliferation and differentiation has been identified as T-cell-derived IL-5. The main action of IL-5 is to stimulate the proliferation and maturation of eosinophil precursors (Eo-CFU); IL-5 appears to be lineage specific and does not induce the proliferation of other myeloid lineage precursors. This could be linked to the restricted IL-5Rα chain expression to eosinophils. IL-5 activity in eosinophils is probably enhanced by the actions of other interleukins, such as IL-3 and IL-4.

In addition to its growth-promoting and differentiating effects in immature eosinophils, IL-5 also acts on mature eosinophils, for example, increasing phagocytosis, killing of antibody-coated tumor cells, stimulation of superoxide (O_2^-) production. IL-5 has been demonstrated to induce the differentiation of murine- but not human-activated B lymphocytes. This activity is only apparent if the B lymphocytes have been activated by a priming stimulant such as LPS, anti-immunoglobulin, or specific antigen. IL-5 alone preferentially enhances IgA production, while specific combinations of IL-5, IL-2, and IL-4 appear to regulate the amount of IgG1 isotype synthesis.

In vivo, IL-5 causes eosinophilia, an expected result from its *in vitro* activities. In

transgenic mice that constitutively express IL-5 in their T cells, massive eosinophilia with eosinophil infiltration of tissues was observed, although in all other respects such mice remained normal and healthy.

4.6 Interleukin-6

The biological activities that are associated with IL-6 are very wide ranging – it has been shown to be very pleiotropic, indicating that IL-6R are expressed on many cell types. Originally, it was described as a B-cell stimulating factor (BSF-2, Table 1) and it is clear that IL-6 is a growth and differentiation factor for a variety of B-lymphoid cells. It is a growth factor for many murine antibody-secreting hybridomas, for some human myelomas, and for low-density cultures of human B cells transformed with Epstein–Barr virus (EBV). IL-6 is also a late-acting differentiating factor in mature B lymphocytes; their terminal differentiation into Ig-secreting plasma cells appears to be upregulated by IL-6. For T cells, IL-6 probably acts as a costimulant (secondary signal) for proliferation of thymocytes and IL-2 production by mature T cells. In hematopoiesis, IL-6 acts as a costimulant for IL-3-dependent proliferation of multipotential cells and induces maturation of megakaryocytes leading to increased platelet number. It may also stimulate the proliferation of myeloid leukemic blast cells and keratinocytes.

A major activity of IL-6 is that of a "hepatocyte stimulating factor", principally as an inducer of acute-phase proteins. IL-6 induces a variety of acute-phase proteins, including fibrinogen, α-1-antichymotrypsin, α1 acid glycoprotein and haptoglobulin, from the hepatoma cell line, Hep2G, and primary hepatocytes (which also produce C-reactive protein (CRP), serum amyloid A and α-1-antitrypsin)). *In vivo*, levels of IL-6 are positively correlated with levels of CRP. Serum levels of IL-6 also correlate well with fever. In transgenic mice that constitutively express IL-6 in their B cells, IgG1 plasmacytosis is developed with the infiltration of plasma cells into lung, spleen, and kidney. In contrast, infection of bone marrow cells with an IL-6 expressing recombinant retrovirus resulted in a fatal myeloproliferative disease with massive neutrophil infiltration of lungs, liver, and lymph nodes, but no plasmacytosis.

4.7 Interleukin-7

IL-7, which is produced by bone marrow stromal cells, appears to be principally involved in stimulating the growth and differentiation of cells of the lymphoid lineages. In particular, it has been shown to induce the proliferation of precursors of B lymphocytes, the so-called pro-B-cells, *in vitro*. In addition, IL-7 can stimulate the proliferation of thymocytes; it has been demonstrated to act synergistically with IL-1 in this respect. Probably, IL-7 occupies a pivotal role in T-cell development, but it also may be involved in differentiation of mature T lymphocytes. Here IL-7 may act rather like IL-6 as a costimulant of IL-2 production by T lymphocytes. Possibly IL-7 can also substitute IL-2 as a mitogen in the generation of CTL and LAK cells.

In vivo, it is likely that IL-7 activities mirror those found *in vitro*. For instance, in transgenic mice expressing IL-7 in lymphoid organs, increased numbers of both B and T cells were observed. In some transgenic mice, the continuous secretion of IL-7 has led to lymphoproliferative disorders, including B and T lymphomas.

4.8 Interleukin-8

IL-8 is the name given to just one of a number of related chemokines that are induced by inflammatory stimuli, for example, LPS, IL-1, and that act as chemoattractants for particular cell types, neutrophils, in the case of IL-8. Thus, release of IL-8 from infiltrating or tissue macrophages specifically recruits neutrophils to sites of injury or infection. IL-8 also acts on neutrophils to induce respiratory burst responses and degranulation, resulting in the superoxide ion (O_2^-) production and the release of lysosomal enzymes. In addition, IL-8 has been demonstrated to be a chemotactic and activating agent for about 10% of T cells and some basophils *in vitro*. However, it is not chemotactic for blood monocytes, the role for this being assigned to another member of the IL-8 superfamily, namely, the monocyte chemoattractant protein-1 (MCP-1). (NB IL-8 is a member of the c-x-c intercrine α-subgroup, while MCP-1 is a member of the c-c intercrine ß-subgroup).

IL-8 has been recovered from inflammatory sites such as psoriatic lesions, the synovial fluid of rheumatoid arthritis patients, and the bronchial lavage fluid taken from patients with acute respiratory distress syndrome. Injection of IL-8 leads to local neutrophil infiltration in both rabbits and humans. If repeatedly injected, IL-8 appears to cause neutrophil accumulations in joints and lungs, and may be associated with damage to cartilage and pulmonary inflammation.

4.9 Interleukin-9

IL-9 is produced by Th1 cells and has been shown to sustain the antigen-independent growth of certain Th-cell clones and lines *in vitro*. Subsequently, mouse IL-9 has also been shown to enhance the growth of bone marrow–derived mast-cell lines in response to IL-3, and that of fetal thymocytes in response to IL-2. In addition, human and mouse IL-9 act to enhance erythroid burst-forming (BFU-E) activity and human IL-9 acts to stimulate proliferation of the human megakaryoblastic leukemic cell line, MO7E. Human IL-9 can synergize with stem cell factor (SCF) and/or EPO in the stimulation of growth of this megakaryoblastic cell line and megakaryocytic progenitors.

An IL-9-dependent cell line that was transfected with the IL-9 gene to give deregulated IL-9 expression became highly tumorigenic in mice. This suggests that uncontrolled expression of IL-9 can support T-cell proliferation *in vivo*, and may be a transforming event involved in the development of certain T-cell lymphomas.

4.10 Interleukin-10

IL-10 is produced by activated Th2 lymphocytes, macrophages, and B lymphocytes and was originally characterized as a cytokine synthesis inhibitory factor (CSIF) because it inhibited IFN-γ synthesis in T cells. It also appears that IL-10 actively suppresses the development of the Th1 subset of Th lymphocytes. This probably occurs as a result of the immunosuppressive activity of IL-10 on macrophages leading to the inhibition of IL-12 production. IL-12, produced by activated macrophages and dendritic cells, stimulates Th1 lymphocyte proliferation. IL-10, in possible synergy with IL-4 and transforming growth factor ß, also blocks the production of IL-1 and TNFα by macrophages and the development of macrophage cytotoxicity and thus

can be considered as an anti-inflammatory agent. In contrast, IL-10 enhances B-cell proliferation and Ig synthesis.

Th1 cells are generally host-protective for several infectious diseases. Therefore, the suppression of Th1 cell activities will favor the growth of the infecting agent. In this context, it is interesting that EBV and several other large DNA viruses encode (viral) IL-10 molecules with limited homology to IL-10, but which act through IL-10R and thus have the same activities as IL-10. These viral IL-10 homologs appear to induce immunosuppression, thus providing an immune-evasion mechanism for such viruses. In contrast, the paralogous IL-10-related proteins, IL-19, -20, -22, -24 -26, -28, and -29 with similar limited homology to IL-10, act through distinct heterodimeric receptors composed of receptor chains belonging to the type II cytokine receptor family, which also includes receptors for interferons.

4.11 Interleukin-11

IL-11, produced by bone marrow–derived stromal cells, appears to be intimately involved in hematopoiesis and, in particular, the generation of megakaryocytes. *In vitro* studies suggest that IL-11 is capable of directly supporting the proliferation of committed murine myeloid progenitors and, like IL-6 and G-CSF, acts synergistically with IL-3 to shorten the Go phase of the cell cycle in early progenitors. Although IL-11 has no inherent megakaryocytic colony stimulating activity, it can synergize with IL-3 in stimulating human and murine megakaryocyte colony formation.

In many respects, the activities of IL-11 are very similar to those of IL-6. For example, like IL-6, IL-11 has been found to promote an increase in the number of Ig-secreting B cells. It has also been shown to stimulate the proliferation of an IL-6-dependent murine plasmacytoma cell line, and to act as an autocrine growth factor for human megakaryoblastic cell lines. Lastly, IL-11 appears to stimulate the synthesis of hepatic acute-phase proteins, but is less effective than IL-6.

4.12 Interleukin-12

The actions of IL-12 are primarily on T cells and NK cells. It appears to be a necessary factor in the generation of Th1 lymphocytes from naive or uncommitted T cells. In addition, there are results that indicate that IL-12 stimulates IFNγ production by T cells and NK cells by cooperation with either IL-1, or IL-18, or TNFα. The presence of IFNγ may also favor the development of the Th1 subset by inhibiting the production of IL-10 by macrophages. IL-12 appears to be an important factor for the differentiation and maturation of dendritic cells.

The emerging experimental data for IL-12 therefore strongly suggest that it plays an important part in the initiation of immune responses by providing a link between natural resistance mediated by phagocytic cells, for example, macrophages, and NK cells and adaptive immunity mediated by Th cells, CTL, and B cells. In support of this proposed role, recombinant IL-12 has recently been shown to cure mice infected with the parasite, Leishmania major. A neutralizing antibody against IFNγ abrogated the curative effect of IL-12 indicating that IL-12 was acting by stimulating T-cell and/or NK-cell IFNγ-production *in vivo*, and the development of a protective Th1-cell immune response. Conversely, generation of Th2 cells and associated immune responses was inhibited.

4.13 Interleukin-13

IL-13 bears approximately 30% homology with IL-4 and appears to act very similarly to IL-4. In particular, IL-13, like IL-4, induces IgE and IgG4 synthesis by human B cells. In addition, IL-13 inhibits inflammatory cytokine production by macrophages and monocytes, another activity characteristic of IL-4. Recent studies indicate that IL-13 and IL-4 share receptor components, that is, the IL-4Rα and IL-13Rα1 chains, which together form high-affinity receptors for either IL-4 or -13, thus offering an explanation of their similar activities. Emerging experimental evidence suggests a more complex picture with receptors for IL-13 being constituted from among four components, including IL-13Rα1, IL-13Rα2, IL-4Rα, and IL-2Rγ.

4.14 Interleukin-14

IL-14 is a high molecular weight B-cell growth factor. However, it appears to act as a costimulant, that is, it induces the proliferation of activated B cells, but not resting B cells. It has been speculated that human IL-14 fulfills a similar role to murine IL-5 for B cells. IL-14 has also been demonstrated to inhibit Ig synthesis by mitogen-activated B cells.

4.15 Interleukin-15

IL-15 is a 14 to 15-kDa glycoprotein with shared bioactivities, but no sequence homology, with IL-2. Its biological activities are remarkably like those of IL-2 and include the induction of proliferation of the established T-cell line CTLL.2 and mitogen-stimulated peripheral blood mononuclear cells (PBMC) as well as the generation of cytolytic NK and LAK cells. The overlap of IL-15 activities with those of IL-2 is probably largely accounted for by their sharing both the IL-2Rß and IL-2Rγ chains as components of their respective high-affinity receptors.

4.16 Interleukin-16

IL-16 (or lymphocyte chemoattractant factor) is an atypical interleukin, possibly produced as a degradation product of a larger protein of 42 kDa present in lymphocytes, which probably utilizes CD4 receptors. Nevertheless, IL-16 has been shown to stimulate a migratory response in $CD4^+$ lymphocytes, monocytes, and neutrophils, upregulate IL-2 receptors and class II MHC antigens in resting T lymphocytes and exhibit some anti-HIV activity.

4.17 Interleukin-17

IL-17 is the prototypic member of a family of structurally related IL-17 proteins (6 human members A–E) produced by T lymphocytes. IL-17 appears to act like IL-1 by inducing IL-6, IL-8, and G-CSF synthesis and enhancing expression of the cellular adhesion molecule ICAM-1 in human fibroblasts. Thus, IL-17 is representative of those cytokines involved in initiating inflammatory responses and innate host defense mechanisms, for example, the pulmonary recruitment of neutrophils and microbial host defense in response to the lung pathogen *Klebsiella pneumoniae*. It may also promote the development of hematopoietic precursors.

4.18 Interleukin-18

IL-18, formerly known as *interferon gamma inducing factor* (*IGIF*) and now known to be structurally related to IL-1, appears to be mainly involved in the augmentation of Th1 cellular immune responses to invading pathogens. Its capacity to induce IFNγ is reliant on the costimulation of T lymphocytes by IL-12, whose synthesis is induced by bacterial LPS in monocytes/macrophages. It has also been shown to enhance Fas ligand-mediated cytotoxicity of cloned murine Th1 cells, but not Th0 or Th2 cells. Its activities are mediated by IL-1 receptor-related protein (IL-1Rrp), an Ig superfamily member related to IL-1RI, and can be inhibited by a naturally occurring, specific IL-18 binding protein.

4.19 Interleukin-19

IL-19, one of several IL-10 paralogs, is produced mainly by LPS-stimulated monocytes and macrophages and in common with IL-10 is expected to exert immunoregulatory functions that influence the activities of many of the cell types in the immune system. It appears probable that IL-19 will, like IL-10, act as a feedback inhibitor of proinflammatory cytokine, for example, IL-1, TNFα synthesis and thus initiate immunosuppressive effects.

4.20 Interleukin-20

IL-20, one of several IL-10 paralogs, is in contrast to IL-10 and IL-19 produced mainly by skin cells, and is thus expected to target distinctive activities on typical skin cells such as keratinocytes, which express its specific heterodimeric receptor, IL-20R1/IL-20R2. Overexpression of IL-20 has been shown to result in severe skin abnormalities, including hyperkeratosis, hyperproliferation, and aberrant epidermal differentiation.

4.21 Interleukin-21

IL-21, which is most closely related to IL-15, is one of the several structurally related cytokines, including IL-2, which is secreted by activated $CD4^+$ T cells as part of the normal response to foreign antigens. Its main role appears to be as a differentiation factor for NK cells. It has been shown to promote IFNγ synthesis, which acts to enhance macrophage killing of microbial pathogens, and to prevent apoptosis of precursor NK cells, allowing them to become terminally differentiated and fully activated.

4.22 Interleukin-22

IL-22, one of several IL-10 paralogs, formerly known as IL-10-related T-cell-derived inducible factor (IL-TIF) is primarily produced by activated $CD4^+$ T cells, but unlike immunoregulatory IL-10, its target cell types include mesangial, neuronal, and liver cells (hepatocyctes). Its action on hepatocytes (and hepatoma cell lines) appears similar to that of IL-6 in that it results in the upregulation of a number of acute-phase proteins. It should therefore probably be included with other proinflammatory cytokines as a prime mediator of the acute-phase response.

4.23
Interleukin-23

IL-23, produced mainly by macrophages and dendritic cells, is a novel heterodimeric cytokine constituted by the p40 subunit of IL-12 and a p19 subunit that is most closely related to the p35 subunit of IL-12. It is evident that IL-23, besides sharing one structural subunit with IL-12, also shares IL-12Rβ1 of the high-affinity IL-12R as part of IL-23R. However, a second specific IL-23R component is indicated for the recognition of p19 suggesting IL-23 has distinct cell targets from IL-12. For instance, although IL-23 in common with IL-12 stimulates IFNγ production, IL-23 is a strong inducer of memory T-cell proliferation, while IL-12 is not. In addition, IL-23, in contrast to IL-1, for example, appears to act more broadly as an end-stage effector cytokine via direct actions on macrophages.

4.24
Interleukin-24

IL-24, one of several IL-10 paralogs, formerly known as melanoma differentiation-associated antigen 7 (MDA7), was originally found to be expressed in differentiated melanoma cells, but is now also known to be expressed by a variety of cell types, including T lymphoblasts. IL-24 appears to act through IL-20R and thus is expected to have similar cell targets to IL-20. It has been demonstrated to have antiproliferative activity against cultured tumor cell lines, including those derived from melanomas.

4.25
Interleukin-25

IL-25, also known as SF20, is a novel cytokine belonging to the IL-17 family, which is mainly produced by Th2 cells. It has been shown to induce expression of the immunosuppressive interleukins, IL-4, IL-5 and IL-13, required for Th2-like responses and may thus be an important mediator of allergic disease.

4.26
Interleukin-26

IL-26, one of several IL-10 paralogs, also previously known as AK155, is expressed at high levels in activated T cells, monocytes, and various T-derived leukemic cell lines. However, little is yet known about its cell targets and biological activities.

4.27
Interleukin-27

IL-27 is a newly identified "IL-12-like" heterodimeric cytokine, also similar to IL-23. It is constituted by Epstein–Barr virus-induced gene 3 (EBI3) protein, a protein that is widely expressed and related to IL-12 p40 subunit and a further homolog of IL-12 p35 known as p28. IL-27 is the product of activated antigen-presenting cells (macrophages and dendritic cells) and stimulates the rapid clonal expansion of naïve but not memory $CD4^+$ T cells. It also strongly synergizes with IL-12 to trigger IFNγ production by naïve $CD4^+$ T cells.

4.28
Interleukin-28

IL-28 consists of two near identical proteins, IL-28A and B, which are distantly related to both the IL-10 and IFNα families, which are induced by viral infection. They interact with a heterodimeric class II cytokine receptor that consists of IL-10Rβ chain and an orphan class II receptor

chain designated IL-28Rα. IL-28 induces interferon-stimulated response elements (ISRE), but not via IFN receptors, and mediates moderate antiviral activity in response to virus infection. This has led to IL-28A and B being also named as IFNs, IFN-lambda2, and -lambda3, respectively. In common with IFNs, IL-28 upregulates several known IFN-responsive genes, including MxA, 2-5A synthetase, and class I MHC antigen.

4.29 Interleukin-29

IL-29 shows 81% sequence homology to IL-28 and acts in a very similar way. It also has been shown to induce IFN-responsive genes and exert antiviral activity, and thus is alternatively designated IFN-lambda1.

5 Interleukin Physiology

Interleukins are mainly, but not exclusively, inducible mediators of nonspecific host resistance and adaptive immunity, that is, they are stimulators of host-protective mechanisms against infectious and invasive diseases. The biological effects of most interleukins are readily observable *in vivo*. However, the administration of high doses of a single interleukin by abnormal routes may also create imbalances that perturb homeostatic mechanisms resulting in aberrant biological and physiological responses. Therefore, studies in which deficiencies of interleukin production or action can be investigated can provide perhaps a better guide to the physiological role(s) of individual interleukins. Such studies can either be done by blocking interleukin action, for example, with anti-interleukin antibodies, antireceptor antibodies, or soluble receptors, or by examining the effects of genetic deficiencies in interleukin protein or receptor. The former are probably the less satisfactory of the two approaches since interleukin production needs to be stimulated before an effect of an inhibitor of interleukin action can be demonstrated. Generally speaking, it is not possible to induce the production of a single interleukin; several interleukins and cytokines are normally produced in response to a stimulant. In addition, most antibodies used will be foreign (xenogenic) to the responding animal, and this may result in problems of an immunological nature.

There have been very few natural genetic deficiencies in interleukin genes and interleukin-receptor genes identified so far. However, it is now possible to artificially create knockout mutations in mice by using the strategy of gene targeting in embryonic stem (ES) cells. Increasing numbers of different knockout mice are being created, and it is perhaps surprising that some of them apparently develop normally, are born alive, and can grow and survive for weeks, if not months. Possibly, compensatory molecules or mechanisms are induced in these knockout mice. In the IL-2 knockout or null mouse, for example, development is normal with no effects being observed on the early development of thymocytes and inactivation of T lymphocytes within the thymus. However, a lack of response of mature T cells to polyclonal activators together with a lack of T cell–mediated "help", that is, interleukin production, for B-cell growth and differentiation were manifest in these mice. There were also excessively high levels of IgG1 and IgE found. About half of these mice became severely immunocompromised and died between four and nine weeks from

birth. The remainder developed an inflammatory bowel disease, probably resulting from an abnormal immune response to a normal antigenic stimulus. In man, a child with a defect in IL-2 mRNA production (but not due to a defect in the IL-2 gene) was found to be more susceptible to infections. In addition, a few patients suffering from a particular form of severe combined immunodeficiency disease (SCID), which has been linked to the X chromosome (thus XSCID), are characterized by a lack of expression of IL-2Rγ. XSCID patients are therefore unable to form high-affinity IL-2 (or IL-4, -7, -9, -15, -21) receptors and their immunocytes cannot respond to IL-2 or the other interleukins that require IL-2Rγ. In total, these studies provide strong evidence for IL-2 playing a key role in the regulation of immune and inflammatory responses. In contrast, either IL-15 or IL-15Rα knockout mice are lymphopenic with selective loss particularly of NK and CD8$^+$ T-cell subsets, rendering these mice susceptible to vaccinia virus infection. Thus, this phenotype is distinctive from that of IL-2 knockouts and indicates a role different to IL-2 for IL-15, despite the latter sharing two of the high-affinity IL-2R components (see Sect. 4.15).

Other knockout mice that are interesting in this respect are those of the IL-4 and IL-10 null phenotypes. In IL-4 knockout mice, it has been found that they generally are unable to mount a Th2 cell type of immune response. For instance, IgE was undetectable in these mice. In addition, when these mice were parasitized with the nematode *Nippostrongylus brasiliensis*, only a weak eosinophilia was observed. Since eosinophil proliferation is dependent on IL-5 (see Sect. 4.5), this finding indicates that IL-4 is required for inducing Th2 cell interleukin production, that is, IL-4, IL-5, IL-10, and so on. In the case of IL-10 knockout mice, it might be expected that these animals are also less able to mount a Th2 cell immune response, since loss of IL-10, which actively suppresses macrophage "help" (IL-12) for Th1 cell generation, would relieve the inhibition on Th1 cell development. However, the main observable effect in IL-10 knockout mice is chronic bowel inflammation where mice are not kept under specific pathogen-free conditions. This finding indicates that IL-10 acts as a regulator of immune responses stimulated by enteric antigens in the intestine. In the absence of IL-10, such immune responses are uncontrolled leading to bowel inflammation, probably due to a loss of IL-10-mediated suppression of macrophage cytotoxicity and release of inflammatory cytokines, for example, IL-1, TNFα. In support of this hypothesis, it has been shown that normal mice become more sensitive to LPS-induced endotoxic shock by treatment with anti-IL-10 antibodies, while lethal endotoxemia and elevated serum TNFα levels were suppressed by injected IL-10.

Another interesting cohort of interleukin knockouts that is currently emerging is that with knockouts of IL-12 or IL-23. Knockouts of IL-12 p40 (which lack both IL-12 and IL-23) are more immunocompromised than knockouts of IL-12 p35 (which lack only IL-12). However, knockouts of IL-23 p19 (which lack only IL-23) manifest a phenotype that can be distinguished from the p35 knockout. The p19 null mice can still generate Th1 cells and IFNγ, but are susceptible to experimental autoimmune encephalitis (EAE), whereas p35 null mice cannot generate Th1 cells, but are highly susceptible to EAE. Such studies indicate the importance of the interleukin recognition components of their cognate receptors in determining their functional roles.

An alternative method for determining *in vivo* roles of interleukins involves neutralizing their biological activity with specific antibodies, soluble receptors, and other antagonists, that is, essentially preventing the interaction of the interleukin with its cognate receptors, also generates interesting results. To date, however, the abrogation of interleukin activities has mostly been carried out under conditions in which pathological conditions have been induced, for example, microbial challenge, tissue injury, graft-versus-host disease, or autoimmune disease. From these studies, for example, for IL-1, it has been shown that neutralizing antibodies suppress cell-mediated immunity and increase susceptibility to pathogens such as *Listeria monocytogenes*. In this case, the IL-1ra and soluble IL-1R can also produce similar effects. Such work indicates a protective role for IL-1. Interestingly, supportive evidence for this has been revealed by a series of recent findings that members of certain virus families, particularly the poxvirus family, encode soluble homologs of interleukins or their receptors, suggesting that some viruses have evolved means of countering host-defense mechanisms that would otherwise inhibit their replication. For example, vaccinia and cowpox viruses encode a soluble form of IL-1RII that can compete for IL-1 binding to cell surface receptors and thus reduce IL-1 actions. Besides this, the IL-1RII homolog, cowpox virus has also been found to encode for an inhibitor, a serpin, of the IL-1ß converting enzyme, and this too would be expected to inhibit an IL-1ß-driven immune response from being mounted. The example of the Epstein–Barr virus encoded homolog of IL-10 has been previously discussed under Sect. 4.10.

6 Pathophysiology and Disease Correlates

There are a large number of examples of interleukins having pathophysiological roles or disease correlates, regrettably far too many to be covered here, which may help understand the complex and versatile nature of interleukins in health and disease. Starting with IL-1, it is clear that when produced at low levels it has a protective role, but when produced at high levels, which spill into the circulation, it produces a predominance of proinflammatory effects that are correlated with disease states such as hypotensive shock and sepsis. High levels of IL-1 can induce bone and cartilage resorption and degradation and IL-1ß has been found in the synovial fluid and serum of rheumatoid arthritis (RA) patients. Other inflammatory cytokines such as TNFα can also be found in RA synovial fluids. Serum or plasma IL-6 levels can be greatly increased by bacterial infections and result in acute-phase responses. IL-8 levels are raised in psoriasis scales. IL-10 levels can be raised in malignant B-cell lymphomas in AIDS patients.

Studies in transgenic mice where expression of interleukin genes is under the control of constitutive promoters have also contributed evidence of pathophysiological effects. For example, IL-2 expression in many body organs of transgenic mice leads to baldness and interstitial pneumonia due to an inflammatory infiltrate. In contrast, IL-2 expression in the pancreas leads to a lethal pancreatitis, but not diabetes. Constitutive expression of IL-5, IL-6, and IL-7 can result in the expected hyperplasias, that is, eosinophilia (IL-5), plasmacytosis (IL-6), lymphocytosis (IL-7). However, the site and control of IL-6 expression, for example, is important in determining the

outcome. When IL-6 is expressed under the control of the human Ig heavy chain promoter in B cells, the result is IgG1 plasmacytosis and infiltration of plasma cells into lung, spleen, and kidney. If, however, IL-6 is expressed under the control of the human keratin (K14) promoter in skin and tongue, there is growth retardation, poor hair growth, and epidermal scaliness (tail, paws), but no changes typical of psoriasis. By contrast, if bone marrow cells are infected with an IL-6-expressing recombinant retrovirus, a fatal myeloproliferative disease with massive neutrophil infiltration of lungs, liver, and lymph nodes develops within four weeks of engraftment, but no plasmacytosis occurs. As a further example of the diverse outcomes of constitutive interleukin expression, it has been found that IL-4 expression in the thymus and T cells results in thymic hypoplasia, inflammatory ocular lesions (blepharitis), and raised IgE and IgG1 levels, whereas IL-4 expression in B cells results in raised IgE and IgG1, but no thymic hypoplasia or blepharitis. Such differences indicate an importance of where or in which cells an interleukin is expressed, together with the quantitative and temporal elements, on the outcome.

7
Clinical Uses of Interleukins

The biological activities of interleukins have suggested that some may be of clinical use as therapeutic agents in human diseases as shown in Table 6. For example, the stimulating activity of IL-2 on the cytotoxic function of NK and LAK cells against tumor cells *in vitro* suggested that IL-2 had potential as an anticancer agent. Thus, once adequate amounts of recombinant IL-2 (rIL-2) became available, clinical trials to evaluate rIL-2 in different human malignancies were carried out. It was quickly established that high-dose IL-2 therapy caused severe side effects, for example, hypotension, oliguria, fluid retention, breathing difficulties, heart problems, with the dose-limiting side effect being the so-called *vascular leak syndrome*. In the latter, fluid extravasation and subsequent edema (swelling) takes place in the pleural and peritoneal cavities. The underlying cause appears to be the adherence of IL-2-activated lymphocytes to vascular endothelial sites, that is, the linings of blood vessels, resulting in holes in the endothelial cell layer being produced through which fluids leak. Despite these severe side effects, some tumor responses were found in a limited number of malignancies, including renal cell carcinoma, melanoma, and non-Hodgkins lymphoma. However, overall response rates with IL-2 as a single agent have been disappointingly low. New strategies involving combinations of IL-2 with other cytokines (e.g. IFNα, TNFα), antitumor monoclonal antibodies, cytotoxic drugs, or LAK cells removed by leukopheresis, activated by IL-2 and grown *ex vivo* before reinfusing back into the patient, have or are being tested in order to improve efficacy. The most "successful" approach has been that of adoptive cellular therapy where activated LAK cells are combined with high doses of IL-2. However, this is a highly aggressive antitumor therapy, which has serious complications for patients, with major life-threatening side effects. Despite initial tumor regressions following this therapy in "hard-to-treat" cancers such as colorectal cancer and malignant melanoma, in most instances remissions were not durable. A more sophisticated approach in which tumor-infiltrating lymphocytes (TIL) are obtained, IL-2

Tab. 6 Disease correlates and possible scope for therapeutic/clinical intervention using interleukin inhibitors.

Interleukin	*Disease correlates*	*Possibilities for therapeutic/ clinical intervention*
IL-1	– Gram-negative septacemia,	IL-1ra, soluble IL-1R (other anti-endotoxic shock, hypotension inflammatory agents, e.g. anti-TNFα)
	– Rheumatoid arthritis	As above
	– Multiple sclerosis	,,
	– Kawasaki syndrome (inflammation of veins)	,,
	– Myeloid leukemia	,,
	– Insulin-dependent diabetes mellitus (IDDM)	,,
IL-2	– Lymphoid leukemia	Soluble IL-2R (also anti-TNFα antibody)
	– Systemic lupus erythematosus (SLE)	,,
	– Graft-versus-host disease	,,
IL-3	– Cerebral malaria	Anti-IL-3, anti-GM-CSF, anti-IFNγ
IL-4	– Allergy/asthma	Anti-IL-4, soluble IL-4R
	– IDDM	,,
	– Blepharitis (inflammatory eye lesion)	,,
IL-5	– Allergy/asthma	Anti-IL-5
	– Eosinophilia	,,
IL-6	– Gram-negative septicemia	IL-1ra, soluble IL-1R, anti-TNFα
	– Multiple myeloma	Anti-IL-6, soluble IL-6R
IL-7	– Lymphoid leukemia	Anti-IL-7, soluble IL2R
IL-8	– Psoriasis	Signal transduction inhibitors
	– Erythroderma	,,
IL-10	– Burkitt's lymphoma	Anti-IL-10
	– Malignant B-cell lymphomas in AIDS patients	,,
IL-11	– Megakaryocytic leukemia	Anti-IL-11
IL-12	– Multiple sclerosis	Anti-IL-12
	– Autoimmune diseases	,,
IL-20	– Psoriasis	Anti-IL-20

activated and expanded *ex vivo*, and then reinfused back into the patient with the expectation that the TIL would home back into tumors has shown some promise, but again is complicated by side effects and less-than-durable responses. Novel gene therapy strategies involving the transduction of patients' lymphocytes or fibroblasts with the IL-2 gene leading to constitutive IL-2 production are now being tested.

So far, other interleukins despite their proven immunoregulatory activities *in vitro*, have not yet found much favor among clinicians for the treatment of human diseases. IL-1, for example, has been shown to enhance cellular proliferation and immune responses, but given in high-dose schedules it is probably too profoundly toxic for safe use. Similarly, IL-4 can act as a stimulator of lymphocytes, but its augmentation of IgE secretion and thus the potential for inducing allergic responses may disqualify it from being used clinically.

IL-12 has been proposed as an antitumor agent, but has proved to be quite toxic when administered systemically to cancer patients.

In patients with chronic disease, the presence of certain interleukins and their potential involvement in inflammatory and degradative processes suggests that countering their activities could produce clinical benefits. For example, IL-1 is frequently found in the synovial fluids of RA patients and is quite possibly associated with inflammatory symptoms. Thus, intervention using anti-IL-1 antibodies, IL-1ra, or soluble IL-1R is suggested as a therapeutic approach for RA. For autoimmune diseases such as type 1 diabetes and multiple sclerosis, a similar strategy employing IL-1 antagonists is also suggested. These antagonists could additionally be potentially useful in the treatment of acute illness, for example, bacterial sepsis and shock, as well as perhaps in other conditions such as asthma and acute respiratory distress syndrome (ARDS). So far, clinical phase II trials of IL-1ra in sepsis and RA patients have led to encouraging results and, although IL-1ra was administered in high doses, there were few, if any, side effects.

In several types of hematological malignancies, tumor cells secrete interleukins or other cytokines that may act as autocrine growth factors. For example, IL-2 and IL-7 may be produced by lymphoid leukemia cells, IL-6 by multiple myeloma cells and IL-10 by B-lymphoma (e.g. Burkitt's lymphoma) cells. Such interleukin production therefore suggests a case for investigating whether antagonists of these interleukins would inhibit tumor cell proliferation. There is as yet little clinical work done in the cancer area with interleukin antagonists, and it will be for future clinical trials to determine if they will be successful.

8 Concluding Remarks

Interleukin biology is highly complex due to biochemical redundancy, pleiotropic activities, and the numerous interactions among interleukins themselves and with other biological effector molecules. While some biological activities of individual interleukins shown *in vitro* can be demonstrated *in vivo*, there remain questions about types of expressing and responding cells, sites, levels and duration of expression, developmental stages, and genetic background. Advances in molecular genetics, for example, knockout mice, have provided means for identifying physiological roles for some interleukins, but in most instances it is still not certain that such roles can be compensated for by other interleukins or cytokines, or how such roles are integrated within the host intercellular communication network as a whole. The present fragmentary evidence for defined physiological roles for the majority of interleukins remains an obstacle for translating the activities of these powerful biological molecules into clinically useful treatments of human diseases.

Acknowledgments

6I am indebted to Miss Deborah Kirk for the excellent typing of this chapter.

See also Bioorganic Chemistry; Medicinal Chemistry.

Bibliography

Books and Reviews

Baxter, A., Ross, R. (Eds.) (1991) *Cytokine Interactions and Their Control*, John Wiley & Sons, Chichester, UK.

Callard, R.E., Gearing, A.J.H. (Eds.) (2001) *The Cytokine Factsbook*, 2nd edition, Academic Press, London, UK.

Cosman, D. (1993) The hematopoietin receptor superfamily, *Cytokine* **5**, 95.

Fickenscher, H., et al. (2002) The interleukin-10 family of cytokines, *Trends Immunol.* **23**, 89.

Horst Ibelgaufts Cytokines Online Pathfinder Encyclopaedia. A COPE free servive on the World Wide Web. (2003).

Ibelgaufts, H. (1995) *Dictionary of Cytokines*, VCH, Weinheim (Germany) and New York (USA).

Meager, A. (1998) *The Molecular Biology of Cytokines*, John Wiley & Sons, Chichester, UK.

Mire-Sluis, A., Thorpe, R. (Eds.) (1998) *Cytokines*, Academic Press, San Diego (USA) and London (UK).

Stahl, N., Yancopoulos, G.D. (1993) The alphas, betas, and kinases of cytokine receptor complexes, *Cell* **74**, 587.

Thomson, A.W. (Ed.) (1992) *The Molecular Biology of Immunosuppression*, John Wiley & Sons, Chichester, UK.

Thompson, A.W. (Ed.) (1998) *The Cytokine Handbook*, 3rd edition, Academic Press, London, San Diego.

Cytomegalovirus and Varicella–zoster Virus Vaccines

Lawrence A. Hunt
University of Louisville School of Medicine, Louisville, KY, USA

1 Introduction to Human Cytomegalovirus and Varicella-zoster Virus 155
1.1 Human Diseases from Primary and Recurrent Infection 155
1.2 Naturally Acquired Protective Immunity 157
1.3 Efficacy of Live-attenuated Varicella Vaccine in Preventing Chickenpox 157
1.4 Vaccine Strategies to Prevent Congenital CMV and Disease in Immunocompromised Adults 159
1.5 Vaccine Strategies to Prevent Zoster in the Elderly and Immunocompromised Adults 159

2 Virion Structure and Composition 160
2.1 Viral DNA Genomes and Specific Gene Sequences for Recombinant Vaccines 160
2.2 Immunogenic Proteins for Eliciting Neutralizing Antibody 161
2.3 Immunogenic Proteins for Stimulating Cytotoxic T cells 164

3 Protein Subunit Vaccines 165
3.1 Natural and Recombinant CMV gB Protein 165
3.2 Other Recombinant and Chimeric CMV Proteins 168
3.3 Recombinant VZV gE, gB, Tegument, and Chimeric Proteins 168

4 Live-recombinant Virus Vaccines 169
4.1 Recombinant Poxviruses Expressing CMV or VZV Genes 169
4.2 Recombinant Adenoviruses Expressing CMV Genes 172
4.3 Recombinant VZV Expressing Heterologous Viral Genes 173

5 Plasmid DNA Vaccines for VZV and CMV 173

6 Clinical Testing of New Vaccines 175

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

6.1 Single or Combined Vaccine Strategies in Human Clinical Trials 175
6.2 Animal Models Including Animal Homologues of Human CMV and VZV 176
6.3 Quantitation of Immune Responses in Immunized Animals and Humans 177

7 Concluding Remarks 178

Bibliography 179
Books and Reviews 179
Primary Literature 179

Keywords

Adjuvant
Natural or chemical substance that enhances antigen-stimulated antibody production by B cells and/or activation of T cells.

Antigen
Foreign substance that can stimulate a specific immune response, usually specific antibody; specific determinants on the antigen that are bound by the antibody are known as epitopes.

Live Vaccine
Replication-competent viral or bacterial vaccine; either an attenuated form of a natural infectious agent or a recombinant agent expressing a foreign gene.

Neutralizing Antibody
Antibody to the surface protein of virus particles that renders the virion noninfectious.

Seropositive versus Seronegative
Presence of antibody to a specific pathogen in the serum due to past infection or immunization (seropositive) versus lack of antibody to a specific pathogen in the serum because of no past infection or immunization.

Structural versus Nonstructural Protein
Viral protein species present in the virion (structural) versus viral protein species synthesized in a virus-infected cell but not incorporated into the virion, such as viral DNA polymerase of herpesviruses.

Subunit Vaccine
Purified component of an infectious agent that is incapable of replication in the host but can stimulate a protective immune response; for viruses, natural or recombinant protein.

Virion
Complete infectious virus particle.

Classical live-attenuated vaccines were developed several decades ago for two clinically important human herpesviruses, varicella-zoster virus and human cytomegalovirus. The varicella vaccine has been licensed for immunization of normal children in Japan, Korea, and the United States for a number of years. In contrast, live-attenuated cytomegalovirus has only been tested in a limited number of studies of organ transplant recipients and healthy adult volunteers, and has not yet been considered for licensure. A number of experimental cytomegalovirus vaccines have been developed using molecular biology and recombinant DNA technology, because of theoretical concerns about the long-term safety and effectiveness of the currently available live cytomegalovirus vaccines. Some of the same techniques are also being applied to varicella-zoster virus to develop improved vaccines that could supplement or replace the current varicella vaccine or provide protection against viral reactivation and zoster. The major emphasis of high-technology vaccine development has been on the viral glycoproteins because they are the targets of neutralizing antibody, and includes both recombinant protein subunit vaccines and live-recombinant virus vaccines. Other virion and nonstructural viral proteins are also being considered as immunogens for cytotoxic T-cell responses.

1 Introduction to Human Cytomegalovirus and Varicella-zoster Virus

1.1 Human Diseases from Primary and Recurrent Infection

Varicella-zoster virus (VZV) and human cytomegalovirus (CMV) are both human members of the family *Herpesviridae*, and the vast majority of humans have been infected with both CMV and VZV by the time they reach adulthood. Along with the six other human members, such as herpes simplex virus (HSV) types 1 and 2 and Epstein–Barr virus, these viruses have a common clinical feature of viral latency and reactivation following primary virus infection. Medical complications of reactivation are most severe in immunodeficient or immunosuppressed individuals.

VZV is the cause of the common childhood illness known as chickenpox or varicella, and this disease is characterized by a distinct vesicular rash on the trunk, head, and limbs. VZV is a highly contagious virus, easily spread by respiratory secretions of infected individuals several days before the appearance of skin lesions. Complications of varicella include pneumonia, secondary bacterial infections, and inflammation of the brain, known as encephalitis. These complications may occur in persons with normal immune function and may be life-threatening, but are even more common and serious in immunocompromised persons. Prior to introduction of the varicella vaccine in the

United States, there were approximately 4 million cases of varicella each year in children, 11 000 hospitalizations, and 100 deaths each year, with half of the deaths in children and half in adults infected for the first time with VZV.

Passive immunization with high-titer human antibody against VZV, known as varicella-zoster immune globulin (VZIG), and antiviral therapy with acyclovir, an antiviral drug that inhibits the viral DNA polymerase, can prevent or significantly reduce disease if administered within a few days after exposure to VZV.

The same virus also causes the disease known as shingles or herpes zoster, which is most common in the elderly. This disease results from reactivation of the virus that has been latent in sensory ganglia for decades, with the skin lesions and accompanying pain usually localized to an area of skin innervated by a specific sensory ganglion. The major complication of herpes zoster is the ''postherpetic neuralgia'', which may persist for months in elderly patients. Specific antiviral chemotherapy or passive immunization with VZIG is much less effective in treating reactivated, secondary VZV infection compared to primary infection and varicella.

In contrast to chickenpox/varicella in nonvaccinated children, most primary CMV infections of children or young adults are mild or inapparent, with a mononucleosis-like syndrome being one of the more common symptomatic outcomes in healthy young adults. Past infection with CMV, whether asymptomatic or symptomatic, is indicated by the individual's seropositivity in a laboratory assay for anti-CMV antibody. The site of viral latency following primary CMV infection is not as well characterized as for VZV, but presumably involves white blood cells known as monocytes and polymorphonuclear leukocytes. Serious infections are limited to two major groups: infants infected during pregnancy when virus from the infected mother crosses the placental and infection or reactivation in immunocompromised persons such as immunosuppressed transplant recipients or persons with AIDS. An estimated 1% of all newborns have congenital CMV, but less than 1/10th of these CMV-infected newborns have serious, symptomatic infection at birth. Up to 15% of the congenital infections that are asymptomatic at birth may have later neurological sequelae, especially hearing defects. Intrauterine infection of the fetus during the primary infection of the mother, as compared to reinfection or reactivation in the CMV-seropositive mother, carries the highest risk of serious disease and involves multiple organ systems. The virus is commonly spread among children in preschool and day-care centers because the virus is shed in saliva and urine of asymptomatic children, and it can also spread to CMV-seronegative females of childbearing age who work in these centers or have a child in preschool or day care.

Similarly, the risk of serious and life-threatening CMV disease in immunosuppressed recipients of kidney or liver transplants is lower for seropositive persons undergoing CMV reactivation or reinfection versus seronegative persons acquiring primary CMV infection from the transplanted organ. For CMV-seronegative patients who will be receiving organ transplants from CMV-seropositive donors, initiation of antiviral therapy prior to the transplant can reduce the severity of CMV infection transmitted from the donor tissue. Reactivation of CMV in seropositive patients undergoing bone marrow transplantation or in patients with human immunodeficiency virus and AIDS is also a significant and

sometimes life-threatening disease involving the lungs, liver, or eyes and requires antiviral therapy. The drug of choice is ganciclovir, a compound that inhibits CMV replication by specifically inhibiting the viral DNA polymerase, but can often display significant toxicity.

1.2 Naturally Acquired Protective Immunity

The acquired immunity to CMV following primary infection is typically sufficient to prevent disease in immunocompetent individuals. However, asymptomatic reinfection with different antigenic strains of CMV or reactivation of latent CMV can occur in healthy, CMV-seropositive adults. The antibody and cytotoxic T-cell responses induced by prior CMV infection of females appear insufficient to prevent virus reactivation during pregnancy and congenital infection, but the resulting congenital infections are almost always asymptomatic. Transplacental antibody can also protect newborns and infants from infection or disease after birth until this maternal antibody disappears from circulation.

Immunocompetent children and adults acquire life-long immunity to varicella/chickenpox following this initial infection with VZV, even though the viral genetic information persists in a latent form in nerve tissue known as sensory ganglia. Specific antibody capable of neutralizing virions can be highly protective and can prevent initial infection with VZV, as demonstrated by the protective effects of maternally acquired antibody in preventing varicella/chickenpox in most infants less than 12 months of age. This antibody is transported across the placenta from mother to fetus late in pregnancy and is a natural form of passive immunization against a large number of infectious organisms. Specific cytotoxic T-cell responses are necessary in addition to neutralizing antibody in controlling infection and limiting disease once the virus infection has occurred. These acquired cytotoxic T-cell responses to VZV proteins are essential in preventing or delaying reactivation of latent virus, consistent with the highest incidence of zoster/shingles in the elderly with a natural decrease in the activity of the immune system.

1.3 Efficacy of Live-attenuated Varicella Vaccine in Preventing Chickenpox

The most commonly used varicella vaccine is a live-attenuated VZV strain that was developed in the 1970s from a human clinical isolate of virus from a patient with typical varicella (Table 1). Long-term propagation and multiple passaging of this Oka strain in human diploid fibroblast cell culture led to virus that replicated better in cell culture but was attenuated for replication and pathogenicity in humans, thereby causing infection without disease in most healthy persons when administered by intramuscular or subcutaneous infection. The licensed Oka vaccine strain of VZV can be distinguished from corresponding clinical isolates of VZV on the basis of restriction endonuclease analysis of the viral DNA genome. In addition, determination of the complete DNA sequences of the Oka vaccine virus and its parental virus has recently been completed to allow detailed comparisons and identification of specific gene alterations responsible for the *in vivo* attenuation of the vaccine strain.

Extensive studies of the varicella vaccine in large numbers of healthy children have demonstrated a high degree of safety and

Tab. 1 Important events in the history of varicella vaccine.

Year	*Event*
1974	Development of live-attenuated Oka-strain vaccine in Japan
1979	Initiation of large-scale human trials of Oka varicella vaccine in United States
1995	Licensing of Oka varicella vaccine in United States for healthy children aged 12 months and older and in VZV-seronegative adults
1995	First use of Oka varicella vaccine as live-recombinant vaccine (R-Oka) for protection against other human viruses in animal clinical trials
1996	First human clinical trial of heat-inactivated Oka vaccine in United States for VZV-seropositive adults aged 55 years and older, to decrease incidence and severity of zoster
1998	Initiation of "Shingles Prevention Study," a large-scale clinical trial in United States of live-attenuated Oka vaccine for VZV-seropositive adults aged 55 years and older to decrease the incidence and severity of zoster
2002	Completion of the DNA genome sequences of the Oka vaccine and its parental virus, for mapping of potential attenuation mutations

efficacy in inducing neutralizing antibody to VZV and protection from varicella disease and complications. Approximately 90% of VZV-seronegative healthy children and adults who received a single dose of Oka varicella vaccine in prelicensure clinical trials were protected from illness when naturally exposed to VZV, and essentially all of those vaccinated individuals who did have "breakthrough" varicella after natural exposure to VZV were protected from severe disease. Several studies of varicella outbreaks in preschool and day-care centers since the Varivax (Merck & Co.) vaccine was licensed in the spring of 1995 suggest that the level of protection against varicella in some populations may be much lower than 90%, especially if the children were immunized at 12 months rather than at 15 to18 months of age, but the disease in vaccinated children was milder than in unvaccinated children in the same outbreak. Detailed clinical studies indicate that there is an inverse relationship between the titers of anti-VZV antibody at six weeks postvaccination and the long-term incidence of "breakthrough" varicella, so that specific immunity induced by the vaccine does correlate with protection. The Oka varicella vaccine has not been licensed for immunocompromised children who are most at risk for dangerous disease, although these children could benefit most from the varicella vaccine. The live-attenuated varicella vaccine is somewhat less effective in this population and may actually cause a mild form of varicella.

A second generation of the Oka/Merck varicella vaccine, known as Varivax II, has the same safety and efficacy characteristics of Varivax, but has the advantage of being refrigerator-stable. A similar vaccine has been developed by SmithKline Beecham Biologicals and is known as VARILRIX. There are incentives to develop safer alternatives to these current live-attenuated vaccines because of the ability of the Oka vaccine virus to establish latent infection and subsequently reactivate, and because of the unknown duration of immunity from a single childhood immunization compared to infection

with "wild-type" virus. Two other theoretical concerns with live-attenuated human viruses are potential reversion to virulence during replication in humans and potential oncogenicity in humans. Although the former phenomenon clearly occurred in another human viral vaccine, the live-oral Sabin polio vaccine, there has been no demonstration of similar reversion with the Oka varicella vaccine. Even though several other human herpesviruses are associated with specific human cancers, such as B-cell lymphomas and nasopharyngeal carcinoma with Epstein–Barr virus and Kaposi's sarcoma with human herpesvirus-8, there is no evidence for oncogenicity associated with either natural infection by wild-type VZV or immunization with the live-attenuated varicella virus.

1.4 Vaccine Strategies to Prevent Congenital CMV and Disease in Immunocompromised Adults

The primary target for a safe and effective CMV vaccine is CMV-seronegative females of childbearing age, to prevent symptomatic congenital infection. Another target is CMV-seronegative children or adults who are immunocompromised and at greater risk of CMV disease. Attenuated and laboratory-adapted strains of CMV were developed in the 1970s by the same procedures as those used for the Oka varicella vaccine. One of these, the Towne strain, has been clinically tested in humans, but has not been shown to be effective in preventing primary infection with VZV. The Towne strain may be too highly attenuated, as suggested by the fact that experimental infection does not result in latent infection with potential reactivation.

A number of alternate vaccine strategies are being developed for CMV and also for improved and safer VZV vaccines, on the basis of biochemical and immunological knowledge of the important viral protein components of these viruses and the viral genes encoding these potential vaccine components. One major and current approach for CMV vaccines is the production of recombinant protein subunit vaccines in recombinant bacteria, insect cells, or mammalian cell culture. A second major approach is the production of recombinant viruses that express the CMV gene and can be used as live-recombinant vaccines, including recombinant vaccinia and canarypox viruses and adenoviruses. A third approach is the creation of less-attenuated and possibly more-immunogenic live-attenuated CMV vaccines by constructing genetic chimeras between the Towne strain and a virulent strain. Finally, the newer technology of plasmid DNA vaccines is being applied to CMV and other human herpesviruses.

1.5 Vaccine Strategies to Prevent Zoster in the Elderly and Immunocompromised Adults

Because the time interval between varicella and zoster/shingles can be six decades or longer and correlates with age-induced decreases in immune function, it may be worthwhile to stimulate and enhance cellular immunity against VZV in seropositive adults. This immunization of individuals who have already been infected and still have the virus in the latent form is classified as "therapeutic" immunization, in contrast to the typical "prophylatic" immunization of uninfected individuals with licensed human viral vaccines. One strategy is to immunize seropositive adults with the same dose as children or higher doses of the live Oka varicella vaccine, and this is currently being tested in a large-scale,

long-term clinical trial. In seropositive but immunocompromised individuals, such as candidates for transplantation and immunosuppression, inactivated varicella virus or recombinant protein subunit vaccines may be safer alternatives even though they are generally less effective in stimulating cytotoxic T cells.

2 Virion Structure and Composition

All of the members of the human herpesvirus family have a common structure for the virus particle, or virion, as shown in Fig. 1. From the inside to the outside, there are four structures: an inner core containing the viral genetic information (DNA genome) associated with viral protein, a protein shell with icosahedral symmetry known as an icosahedral capsid, an amorphous layer containing viral proteins known as the tegument or matrix, and a surface envelope composed of a cell-derived lipid bilayer and multiple species of viral glycoprotein. The single-letter alphabetical nomenclature for individual envelope glycoproteins is based upon homology to envelope glycoproteins of herpes simplex virus-1 (HSV-1). The envelope is the most fragile component of the virion, and removal or destruction of the envelope with detergents or organic solvents such as ether or chloroform renders the virion noninfectious. The corresponding capsid, tegument/matrix, and envelope glycoproteins of CMV and VZV serve similar functions in virus attachment to cells, intracellular replication, and virus assembly and egress. However, the corresponding CMV and VZV proteins have distinct amino acid sequences and can be distinguished by reactivity with VZV-specific versus CMV-specific antibody.

2.1 Viral DNA Genomes and Specific Gene Sequences for Recombinant Vaccines

The viral genome for all herpesviruses is a linear molecule of double-stranded DNA. This genome encodes all these virion-associated structural proteins as well as additional nonstructural viral proteins that may function as intracellular enzymes and regulatory proteins. The 230 kb DNA genome of CMV is among

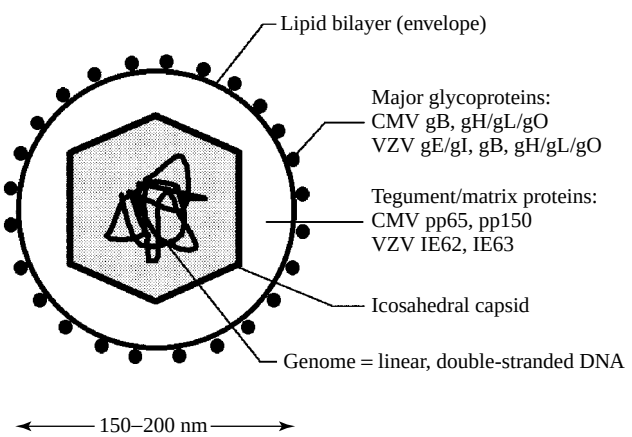

Fig. 1 Structure and composition of CMV and VZV virions.

the largest of all human viruses. The complete nucleotide sequence has been determined for both the AD-169 and Towne laboratory-adapted strains, and this sequence contains approximately 200 distinct open reading frames (ORFs) capable of coding for approximately 200 distinct polypeptides. In contrast, the VZV DNA genome is 125 kb in length, and the complete nucleotide sequence determined for several strains predicts 71 ORFs for potential viral polypeptides. As shown in Fig. 2, herpesviral genomes have the unusual characteristic of repeated DNA sequences at one or both ends and at an internal site (TR and IR), which results in recombinational inversion of unique long (U_L) and/or unique short (U_S) sequences during intracellular DNA replication. Each of the four isomeric forms of the linear DNA genome for CMV and the two isomers for VZV can be packaged with equivalent efficiency into new virions and exhibit similar infectivity. Determination of the location and nucleotide sequences of individual genes in the large DNA genomes of VZV and CMV (Fig. 2) has permitted detailed structural analysis of the deduced amino acid sequences for immunologically important viral proteins. This has greatly facilitated expression of these genes into recombinant protein or inclusion into live-recombinant vaccines.

2.2 Immunogenic Proteins for Eliciting Neutralizing Antibody

A general principal of viral immunology is that surface proteins of the virion are the target antigens for protective antibody responses. In the case of lipid-enveloped

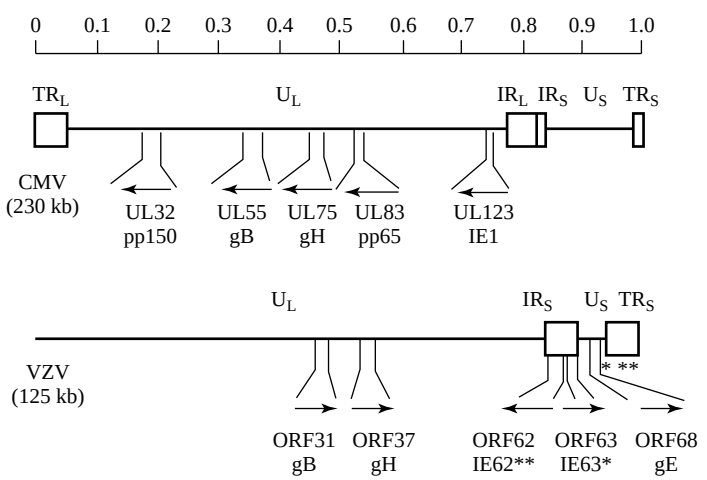

Fig. 2 Structure of the DNA genome for CMV and VZV and location of genes for important viral antigens. Only one of four possible isomers for CMV and one of the two possible isomers for VZV are displayed. The top scale corresponds to genetic map units (0.0–1.0). Open boxes represent inverted repeat sequences at the ends of the unique long (U_L) and unique short (U_S) genome segments, and arrows indicate the orientation of coding sequences, which are identified by gene number and gene product. The single and double asterisks indicate the positions of inverted duplication of *IE62* and *IE63* genes at the right end of the VZV genome.

viruses such as VZV and CMV, these neutralizing antibodies are elicited and bound by the envelope glycoproteins, whereas antibody produced to internal virion proteins such as tegument and capsid proteins is not protective and is only useful for confirming VZV or CMV infection. Helper T-cell responses are necessary for production of antibody against protein antigens, so that T-cell responses to both surface and internal virion proteins can have an indirect role in antibody production.

The major glycoprotein species for CMV are gB and the gH/gL/gO heterotrimer, and the major species for VZV are the gE/gI heterodimer, gB, and the gH/gL/gO heterotrimer (Fig. 1). CMV and VZV gB glycoproteins may also exist in oligomeric forms in their native conformation on the surface of the virion, either as a dimer or a tetramer. The characteristics of these viral gene products are summarized in Table 2. The original electrophoretic analyses of purified CMV and VZV detected a larger number of electrophoretically distinct envelope glycoproteins, but subsequent immunological analyses with monoclonal antibodies demonstrated that these multiple species actually

Tab. 2 Characteristics of major immunogenic proteins of CMV and VZV.

Protein (gene)	*Viral function*	*Immunogenicity*
CMV gB (UL55)	Most abundant glycoprotein in envelope; virion fusion and penetration into cell	Principal target of neutralizing antibody, and target for helper and cytotoxic T cells
CMV gH (UL75)	Second most abundant glycoprotein in envelope, oligomer with CMV gL and gO; fusion	Target of neutralizing antibody and helper/cytotoxic T cells
CMV pp65 (UL83)	Major component of tegument; phosphoprotein with protein kinase activity and role in intracellular regulation	Principal target for helper and cytotoxic T cells
CMV pp150 (UL32)	Second most predominant component of tegument; phosphoprotein	Significant target of helper and cytotoxic T cells
CMV IE1 (UL123)	Nonstructural protein; immediate-early protein important in regulation	Significant target for cytotoxic T cells
VZV gE (ORF68)	Most abundant glycoprotein in envelope; oligomer with gI	Major target of neutralizing antibody
VZV gB (ORF31)	Second most abundant envelope glycoprotein; virion entry and fusion	Target of neutralizing antibody
VZV gH (ORF37)	Glycoprotein component of envelope; oligomer with VZV gL and gO	Target of neutralizing antibody
VZV IE62 (ORF62)	Major component of tegument; immediate-early regulatory protein with role in replication; major altered gene product in Oka vaccine vs parental virus strain	Major target of cytotoxic T cells
VZV IE63 (ORF63)	Component of tegument; immediate-early protein with regulatory role in intracellular replication	Major target of cytotoxic T cells

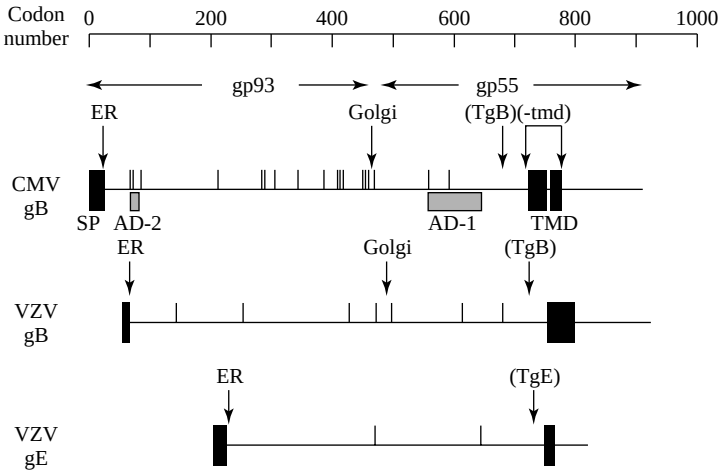

Fig. 3 Linear structure of major CMV and VZV envelope glycoproteins. These diagrams are based upon the derived amino acid sequences from the nucleic acid sequences of the corresponding genes indicated in Fig. 2. The solid black boxes represent hydrophobic domains: amino-terminal signal peptide (SP) and membrane-anchor/transmembrane domain (TMD). Short vertical lines indicate the positions of potential asparagine-linked oligosaccharides. Vertical arrows indicate the positions of cotranslational cleavage by the cellular signal peptidase (ER), posttranslational cleavage of gB into mature subunits (gp93 and gp55 for CMV) by a cellular endoprotease (Golgi), and removal of the membrane-anchor/TMD by gene truncation after codon 680 of CMV gB (TgB) or codon 511 of VZV gpE (TgpE). Another recombinant and genetically modified form of gB lacks the transmembrane-anchor domain (-tmd) but maintains the carboxy-terminus of full-length glycoprotein. Crosshatched rectangles below the CMV gB polypeptide indicate the positions of an amino-terminal linear neutralization epitope in the gp93 subunit (antigenic domain-2 = AD-2) and an immunodominant linear neutralization domain in the gp55 subunit (antigenic domain-1 = AD-1).

represent a more limited set of gene products with heterogeneity due to oligosaccharide modifications and posttranslational proteolytic cleavage. For example, the CMV *gB* gene encodes a precursor glycoprotein of approximately 130 to 160 kDa and processed subunits of 93 to 116 kDa and 55 kDa (gp93 and gp55 proteins in Fig. 3). These glycoproteins contain multiple asparagine-linked oligosaccharides that are added and processed by host cell machinery, and some of the proteins may also contain some serine- or threonine-linked oligosaccharides. Both viruses contain additional minor glycoprotein species, but their contribution to the induction of neutralizing antibody in natural CMV and VZV infections is less well characterized.

The amino acid sequences deduced from the DNA gene sequences (Fig. 2) for the predominant CMV and VZV glycoproteins are indicative of type I integral membrane proteins, with an amino-terminal hydrophobic signal peptide (SP) necessary for transport of the protein into the

lumen of the endoplasmic reticulum (ER) and another hydrophobic, transmembrane domain (TMD) closer to the carboxy terminus, which functions as a membrane anchor (Fig. 3). More detailed analysis of these amino acid sequences for CMV gB suggests that the first of two hydrophobic peptides is actually a fusion domain on the noncytoplasmic side of the membrane and the second is the actual transmembrane anchor. The removal of the SP is usually a cotranslational event or posttranslational event in the ER, whereas the cleavage of CMV and VZV gB into mature subunits occurs late in processing after transport from the ER to the Golgi membranes. The positions of potential sites for cotranslational addition of asparagine-linked oligosaccharides in the ectodomain between the SP and TMD correspond to Asn-X-Ser/Thr, where X is any amino acid except proline. CMV gB, especially the gp93 subunit, is among the most highly glycosylated of all viral glycoproteins, with approximately 15 asparagine-linked oligosaccharides in the gp93 subunit compared to only 4 potential oligosaccharides on the corresponding subunit of the VZV gB glycoprotein and 2 potential oligosaccharides in the entire ectodomain of VZV gE. These multiple oligosaccharides may help the virus evade host immune responses by preventing access of neutralizing antibodies to antigenic determinants on the viral polypeptide. Multiple cysteine residues in the ectodomain of these major VZV and CMV glycoproteins form intramolecular disulfide bonds, which are presumed to be very important in the overall structure and expression of antigenic determinants in the native molecules.

In contrast to typical cellular genes for membrane glycoproteins, there are no intervening sequences (introns) within the coding sequences for the CMV and VZV glycoproteins, so that DNA fragments can be directly isolated by restriction endonuclease digestion from viral genome DNA, rather than having to isolate viral mRNA and then produce complementary DNA (cDNA) prior to recombinant expression. The important linear and continuous neutralizing epitopes of the CMV gB protein have been specifically mapped using synthetic or recombinant peptides on the basis of the deduced amino acid sequence: the immunodominant domain, known as *antigenic domain 1* (AD-1), is within the gp55 subunit and an additional epitope is located near the amino terminus of the gp93 subunit, known as *antigenic domain-2* (AD-2) (Fig. 3). Discontinuous and conformational epitopes, unlike continuous and linear epitopes, cannot be duplicated by synthetic or recombinant peptides, and are usually lost after protein denaturation or protein misfolding. Neutralization epitopes for the other CMV and VZV envelope glycoproteins are less well characterized.

2.3 Immunogenic Proteins for Stimulating Cytotoxic T cells

Virus-specific cytotoxic T cells are important in recognizing and destroying virus-infected host cells, so that these cellular factories for production of new virus are shut down. Cytotoxic T cells have specific T-cell receptors on their surface that recognize peptide fragments of foreign proteins that were synthesized in the target cells, proteolytically processed, and transported to the surface in a complex with class I major histocompatibility antigens. These viral proteins can be either structural proteins that get assembled into virions (Fig. 1) or nonstructural viral proteins present only in infected cells.

Although the gB and gH envelope glycoproteins of CMV can be targets of cytotoxic T cells, the major target for cytotoxic T cells in infected humans is the pp65 tegument protein. Other important proteins for CMV-specific cytotoxic T-cell responses include the pp150 tegument protein and the IE1 immediate-early and nonstructural protein. IE1 is the first viral protein synthesized after reactivation of CMV from latency. The IE1 protein is encoded by one of the few CMV genes with intron sequences and spliced mRNA, with the second and third exons encoding short peptides also present at the amino terminus of several other IE proteins and the fourth exon encoding the unique immunodominant peptide sequences of IE1.

The relative importance of glycoproteins for neutralizing antibody and tegument proteins for cytotoxic T cells is also characteristic of VZV: two immediate-early proteins that are synthesized early in infection and present in the tegument layer of virions are the major targets for VZV-specific cytotoxic T-cell responses. One of these, IE62, has a number of amino acid mutations in the attenuated Oka varicella vaccine strain compared to the parental virus strain, although differences in IE62-specific cytotoxic T-cell responses in humans with natural VZV infection versus immunization with the Oka varicella vaccine have not been described.

3 Protein Subunit Vaccines

Because of potential concerns with the long-term safety and effectiveness of live-attenuated herpesvirus vaccines, there are major efforts under way in developing and testing protein subunit vaccines for CMV, and more limited studies are also under way for VZV. Nonreplicating and inactivated viral vaccines have the major advantage of safety compared to live-attenuated viral vaccines, but the disadvantage of requiring multiple vaccine doses to obtain helper T-cell responses and levels of antibody production by B cells that are comparable to those obtained with a single inoculation of a live-attenuated vaccine. In addition, these nonreplicating protein-based vaccines require a chemical or biological adjuvant. Physical or chemical inactivation of whole virus particles has been used in the production of a number of licensed human vaccines, including those for rabies, polio, and hepatitis A viruses. However, protein antigens purified from virions can be just as effective in stimulating the production of neutralizing antibody, while possessing fewer side effects or biosafety risk from residual infectious virions or viral genomes. With modern recombinant DNA technology, immunogenic viral proteins can be produced in mass quantities at lower expense than from virions or virus-infected mammalian cell cultures.

3.1 Natural and Recombinant CMV gB Protein

Nonrecombinant gB protein can be isolated and purified from the envelope of virions, cell membranes of CMV-infected human diploid fibroblast cells, or cell-associated subviral particles known as *dense bodies*. These dense bodies, so named because of their electron-dense appearance in electron microscopy of thin sections of infected cells, contain the viral envelope and tegument proteins, but lack the capsid and viral DNA genome. In addition, a number of recombinant techniques are

Tab. 3 CMV gB antigen and subunit vaccines.

Source of Protein	***Characteristics of Protein (Fig. 3)***
Virions purified from CMV-infected cells	Full-length glycoprotein, cleaved by cell protease into gp93/gp55 subunits
Envelope glycoprotein from CMV-infected cells	Full-length glycoprotein, mixture of uncleaved and gp93/gp55 subunits
Dense bodies purified from CMV-infected cells	full-length glycoprotein, mostly uncleaved
Recombinant *E. coli* bacteria	Nonglycosylated polypeptides containing either AD-1 or AD-2 sequences
Recombinant adenovirus-infected human epithelial cells	Full-length glycoprotein, mostly uncleaved
Recombinant baculovirus-infected insect cells	Full-length or truncated, soluble glycoprotein [TgB], mostly uncleaved
Recombinant vaccinia virus–infected cells	Full-length glycoprotein, mostly uncleaved
Recombinant, genetically modified Chinese hamster ovary (CHO) cells	Transmembrane anchor-deleted [gB(-tmd)], soluble and secreted glycoprotein

being used to produce full-length and genetically deleted or truncated forms of gB from UL55 gene sequences (Table 3).

A number of plasmid vectors are available for high-level, inducible expression of eukaryotic gene sequences in recombinant *E. coli* bacteria as fusion proteins or nonchimeric polypeptides representing potential antigenic domains of viral proteins. There are several potential problems in expressing viral glycoprotein genes in immunogenic protein with the same neutralizing epitopes as the natural protein, including lack of addition of asparagine-linked oligosaccharides in bacteria, and potential toxicity of the hydrophobic amino-terminal SP of eukaryotic proteins to bacteria. A cell lysate of recombinant *E. coli* containing a nonglycosylated, SP-deleted form of CMV gB was able to induce low levels of neutralizing antibody in mice but lacked some of the important antigenic determinants of the native CMV gB. A recombinant *E. coli*-derived polypeptide representing AD-1 of CMV gB has the correct antigenic structure if a key intramolecular disulfide bond is present, even though the polypeptide lacks the two oligosaccharide chains of the native protein (Fig. 3).

Recombinant baculovirus-infected insect cells have become a popular and economical system for producing large quantities of glycosylated protein antigens. The major strategy in the construction of recombinant baculoviruses is the replacement of coding sequences of the polyhedrin (PH) protein with coding sequences of the desired foreign protein, because the polyhedrin gene is expressed in huge quantities late in viral infection but is unnecessary for virus replication in cell culture. The initial step in the construction of recombinant baculoviruses is to clone the viral coding sequences into an *E. coli* bacterial plasmid that contains a multiple restriction site for cloning downstream of the polyhedrin promoter and is flanked by baculovirus sequences that allow reciprocal recombination with the genome. The *E. coli*

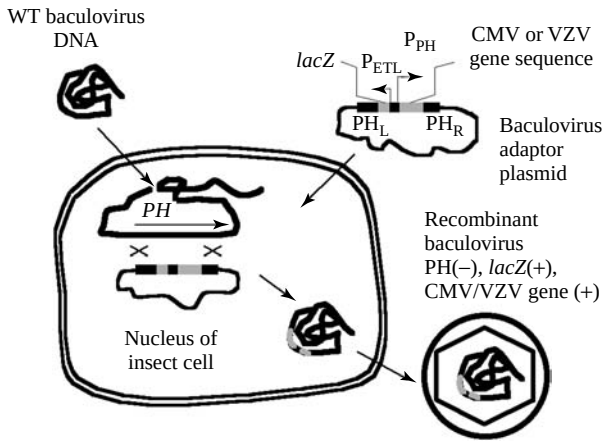

Fig. 4 Production of recombinant baculoviruses to express CMV and VZV glycoprotein genes. This diagram displays the rare reciprocal recombination event that occurs in cells cotransfected with the full-length, wild-type DNA genome and adaptor plasmid DNA containing the foreign herpesviral gene. *PH* refers to the polyhedrin gene, PH_L and PH_R to left and right baculovirus recombination sequences in the plasmid, and P_{ETL} and P_{PH} refer to specific baculovirus promoters for gene transcription. *lac*Z is the coding sequence for the bacterial β-galactosidase protein used in the positive selection of recombinant virus.

β-galactosidase gene (*lac*Z) under control of another baculovirus promoter is also included for positive selection of recombinant virus (Fig. 4). Recombinant baculoviruses are produced by cotransfection of full-length viral genome DNA and the adaptor plasmid containing the herpesviral gene sequences, followed by selection of blue-colored plaques containing rare recombinant viruses that result from intracellular reciprocal recombination.

The major difference in posttranslational processing in insect cells versus mammalian cells is in the final stages of asparaginyl-oligosaccharide processing in Golgi vesicles: high-mannose precursor oligosaccharide structures ($Man_{5-9}GlcNAc_2$-Asn) can be converted into complex, acidic structures [(NeuNAc $\pm$ Gal-GlcNAc)$_{2-4}$Man$_3$GlcNAc$_2$($\pm$Fuc)-Asn] in mammalian cells, but are converted into more simple mature structures (Man$_3$GlcNAc$_2$-Asn) in insect cells. This oligosaccharide difference does not appear to adversely affect the structure and conformation of the polypeptide and may actually make some of the linear or conformational epitopes even more accessible. In contrast, asparagine-linked oligosaccharides in yeast can be processed to large oligomannose structures known as mannan, and this modification can lead to insolubility and loss of antigenic determinants for recombinant glycoproteins.

The glycosylated, full-length gB protein of CMV has been purified from recombinant baculovirus-infected insect cells and demonstrated to induce neutralizing antibodies in mice. Cleavage into gp93

and truncated gp55 subunits is inefficient in the baculovirus-infected insect cells, even when gB proteins are efficiently transported through Golgi membranes. Truncated versions of gB lacking the membrane-anchor and cytoplasmic domain (TgB in Fig. 3) or with more extensive internal deletions in heavily glycosylated domains have also been produced in recombinant baculovirus-infected insect cells, and still maintain the antigenic determinants of both the immunodominant gp55 domain and the amino-terminal gp93 linear neutralization epitope (AD-1 and AD-2 in Fig. 3). The removal of the transmembrane anchor also allows the protein to be secreted from cells in a soluble form that is more easily isolated and purified.

Recombinant viral glycoproteins expressed in mammalian cells are usually similar, if not identical, to the corresponding natural protein from CMV-infected cells with respect to co- and posttranslational modifications and antigenic structure, but this must be balanced with the higher cost of mammalian cell culture compared to the baculovirus-insect cell system. Chinese hamster ovary (CHO) cell lines are commonly used in genetic studies because of the ease of manipulation and growth properties in selective media. Genetically modified CHO cells with recombinant CMV UL55 sequences were used to produce a membrane-anchor-deleted CMV gB subunit vaccine known as *cytomegalovirus glycoprotein vaccine* (Chiron). This gB(-tmd) glycoprotein (Fig. 3) is secreted from the CHO cells as a mixture of cleaved subunits and uncleaved glycoprotein, and this secretion still occurs when the gp93/gp55 endoproteolytic cleavage site is eliminated by site-specific mutagenesis.

3.2 Other Recombinant and Chimeric CMV Proteins

The gH envelope glycoprotein is also a candidate for protein subunit vaccines in combination with gB, rather than being used in place of gB. However, coexpression of the *gL* gene product may be necessary for efficient processing and intracellular transport of gH in recombinant baculovirus-infected insect cells or genetically modified CHO mammalian cells, because gL serves as a protein chaperon in transport from the ER to the Golgi and plasma membranes.

Although recombinant proteins are usually not good immunogens for stimulating specific cytotoxic T cells, a chimeric polypeptide containing an immunodominant peptide from CMV pp65 linked to a nonviral peptide and a chimeric CMV IE1-pp65 protein may both be immunogenic in stimulating both helper T-cell and cytotoxic T-cell responses.

3.3 Recombinant VZV gE, gB, Tegument, and Chimeric Proteins

Truncated and soluble VZV gE and gB glycoproteins have also been expressed as immunogenic glycoproteins in the recombinant CHO mammalian cell or baculovirus-infected insect cell systems (Table 4). Although *gE* (ORF68) and *gI* (ORF67) genes can be coexpressed to produce heterodimers equivalent to the virion glycoprotein, expression of *gI* has not been necessary for efficient production and secretion of soluble and immunogenic gE in CHO cell cultures. A genetically modified CHO cell line was also used to produce a chimeric gE-IE63 protein that combines the neutralizing antibody-inducing properties of gE with the dominant T-cell-inducing properties of the immediate-early

Tab. 4 VZV protein antigens and subunit vaccines.

Source of protein	*Characteristics of protein (Fig. 3 for gE or gB)*
Virions purified from VZV-infected cells	Full-length gE or gB; IE62 or IE63 tegument proteins
Recombinant baculovirus-infected insect cells	Complex of truncated, soluble gE (TgE) and gI
Recombinant, genetically modified CHO cells	Truncated, soluble gE or gB secreted from cells (TgE or TgB)
Recombinant, genetically modified CHO cells	Chimeric, soluble gE-IE63 protein secreted from cells
Recombinant vaccinia virus–infected cells	Full-length gE, gB, or gH/gL oligomer
Recombinant yeast Ty virus–like particles	Chimeric proteins with fragments of gE

and tegument protein when used to vaccinate guinea pigs.

Another relatively new vaccine technology has been recently applied to immunogenic proteins of VZV: the production of recombinant, hybrid Ty virus–like particles in yeast. These particles are devoid of nucleic acid and contain only the yeast transposon protein modified to contain polypeptide fragments of VZV gE or other viral proteins, so that they are safer than live-recombinant or live-attenuated vaccines but have the potential for stimulating both neutralizing antibody and cytotoxic T-cell responses. Immunization studies in mice and guinea pigs have demonstrated the immunogenicity of these recombinant gE-Ty particles.

4 Live-recombinant Virus Vaccines

Live-recombinant vaccines have the advantage of inducing both neutralizing antibody plus cytotoxic T-cell responses, and single inoculations are usually as effective as multiple doses of a nonreplicating subunit vaccine. The immunogenicity of membrane-anchored versus truncated and soluble forms of viral envelope glycoproteins may vary depending upon the amount of membrane-anchored protein that reaches the cell surface or soluble form that is secreted from cells. Thus, both forms may need to be analyzed for each individual viral glycoprotein species to determine the most immunogenic live-recombinant vaccine. The live-recombinant viral vaccines that have been developed to express CMV or VZV genes are summarized in Table 5. Cell cultures infected with these live-recombinant viruses are important sources of recombinant protein for use in subunit vaccines or immunological assays (Tables 3 and 4).

4.1 Recombinant Poxviruses Expressing CMV or VZV Genes

Vaccinia virus, a member of the poxvirus family, has been extensively developed as a live-recombinant vector over the last two decades because this naturally attenuated virus had been used as a smallpox vaccine in humans for over a century, culminating in the global eradication of smallpox in 1977. Immunization with vaccinia virus had a much higher level of complications

Tab. 5 Experimental live-recombinant vaccines for CMV and VZV.

Gene product	*Recombinant vector*	*Immune response*
CMV gB	Adenovirus	Neutralizing antibody in guinea pigs; gB-specific cytotoxic T-cell response in mice
CMV IE1 (exon 4)	Adenovirus	IE1-specific cytotoxic T-cell response in mice
CMV gB	Canarypox virus	Antibody and cytotoxic T-cell response in guinea pigs; priming and enhancement of antibody response to subsequent CMV (Towne) vaccine in humans
CMV pp65	Canarypox virus	Antibody and cytotoxic T-cell response in humans
CMV gB	Vaccinia virus	Neutralizing antibody in mice
VZV gE or IE62	Herpes simplex virus-1	Antibody response in mice
VZV gE and/or gB	Vaccinia virus	Neutralizing antibody and T-cell response to gE or gB in mice, guinea pigs, or rabbits
VZV gH and gL	Vaccinia virus	Neutralizing antibody to gH in mice and guinea pigs
VZV IE62	Vaccinia virus	T-cell response and protection of guinea pigs against challenge with guinea pig–adapted VZV

than other licensed human vaccines, even in persons with normal immune systems. In the event that smallpox immunization is renewed on a large scale in response to potential use of smallpox in bioterrorism, immunocompromised persons such as transplant recipients and persons infected with human immunodeficiency virus are at especially high risk of serious and life-threatening complications if vaccinated or accidentally exposed by contact with skin lesions of an immunized individual. The Western Reserve (WR) strain has been commonly used in construction of live-recombinant vaccinia viruses and was derived from a human vaccine strain by neuroadaptation in mice. Several more-highly attenuated strains are available that would be safer for production of recombinant vaccinia-derived vaccines for human use. A live-recombinant vaccinia virus expressing the rabies virus glycoprotein gene has been licensed for special use as a rabies vaccine against terrestrial wildlife, including raccoons in the eastern United States and foxes in western Europe.

The construction and isolation of recombinant vaccinia viruses for expression of human viral genes is similar in some respects to the process for recombinant baculoviruses: the foreign gene sequences are cloned downstream of a vaccinia promoter in an adaptor plasmid, and recombinant virus is produced by rare reciprocal recombination and replacement of the vaccinia thymidine kinase (*TK*) gene with the foreign gene plus the *lac*Z indicator gene (Fig. 5). Nonrecombinant, TK-positive virus is negatively selected in the presence of a toxic nucleoside analogue, and recombinant, *lacZ*-positive virus is positively selected by the appearance of

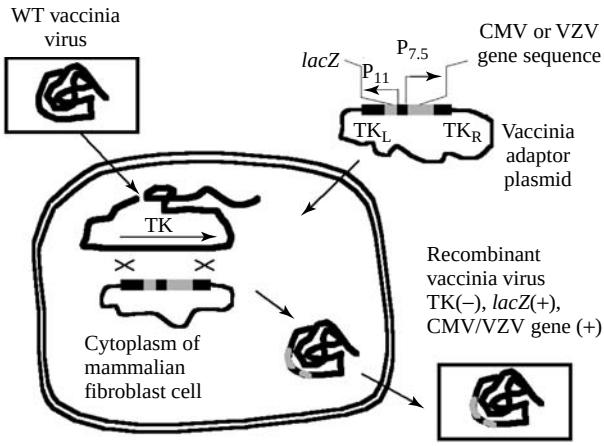

Fig. 5 Production of recombinant vaccinia viruses to express CMV and VZV genes. This diagram displays the rare reciprocal recombination event that occurs in cells infected with wild-type vaccinia virus and cotransfected with adaptor plasmid DNA containing the foreign herpesviral gene. Abbreviations and symbols are the same as those described in the caption to Fig. 4, except that TK refers to the vaccinia thymidine kinase DNA sequences and enzymatic activity, and P_{11} and $P_{7.5}$ refer to vaccinia virus-specific transcriptional promoters.

blue plaques. The vaccinia *TK* gene is unnecessary for virus replication in cell culture, and inactivation of the *TK* gene by this negative selection process further attenuates the vaccinia virus, so that complications in humans with TK-negative virus may be less than the complications in those with wild-type vaccinia virus.

The large vaccinia DNA genome (185 kb) can accommodate at least 25 kb of additional foreign DNA sequences without significant deleterious effects on virus replication, so that multiple foreign virus sequences can potentially be expressed in the same virus, such as both the *CMV gB* (UL55) gene for induction of neutralizing antibodies and the *CMV pp65* (UL83) gene for induction of cytotoxic T cells. Several different recombinant vaccinia viruses have been constructed to express CMV or VZV glycoprotein and tegument genes, and they have been tested in experimental animals (Table 5). These recombinant viruses are also useful for production of recombinant protein in cell culture and for production of specific target cells necessary for assaying virus-specific cytotoxic T cells in cell culture.

An avian poxvirus, canarypox virus (ALVAC), has also been genetically engineered to express human herpesvirus proteins, because the inability to replicate in mammalian cells makes the avian viruses much safer than recombinant vaccinia viruses for use in humans. Pasteur Merieux Connaught has manufactured ALVAC-CMV gB and -CMV pp65 viruses and clinically tested them in CMV-seronegative volunteers (Table 5). Because of the extreme attenuation of canarypox virus in human cells, ALVAC-CMV gB is relatively nonimmunogenic when used alone, and may be best when used in combination with other

vaccines in "prime-boost" immunization strategy.

4.2 Recombinant Adenoviruses Expressing CMV Genes

The idea of using live-recombinant adenoviruses as vaccines is based upon the extensive experience of using enteric-coated human adenovirus types 4 and 7 to immunize military recruits for prevention of acute respiratory disease. The enteric coating prevents the virulent adenovirus from replicating in the respiratory tract and causing disease, but allows the virus to replicate in the gastrointestinal tract after uncoating in the stomach, where it induces good systemic and mucosal immunity without causing disease. These vaccines have the additional advantage of inducing mucosal IgA antibody, which can block the virus soon after entrance into the body, rather than just serum antibody, which blocks infection at a later generalized stage.

In contrast to the large capacity of recombinant vaccinia virus for additional foreign DNA, the 36 kb DNA genome of human adenovirus type 5, the type most commonly used in construction of recombinant adenoviruses, can only accommodate approximately 2 kb additional DNA without genetic instability and inefficient genome packaging into the adenoviral icosahedral capsid. Because approximately 2 kb of nonessential *E3* gene sequences can be deleted, foreign gene inserts of 4 kb can be accommodated during the construction of replication-competent recombinant adenoviruses. One method for construction of these recombinant adenoviruses utilizes an adaptor plasmid containing the final 40% of the viral genome, with the foreign coding sequences inserted downstream of the viral E3 promoter (Fig. 6). These recombinant viral sequences are cotransfected or electroporated into human epithelial cells in culture along with an adenoviral DNA fragment containing the first 76% of the genome, and infectious adenovirus is isolated after several days. These E3-deleted recombinant virions are replication-competent, in contrast to the E1-deleted replication-defective recombinant adenovirus vectors used in experimental gene therapy. The essential *E1* gene is located at the left end of the

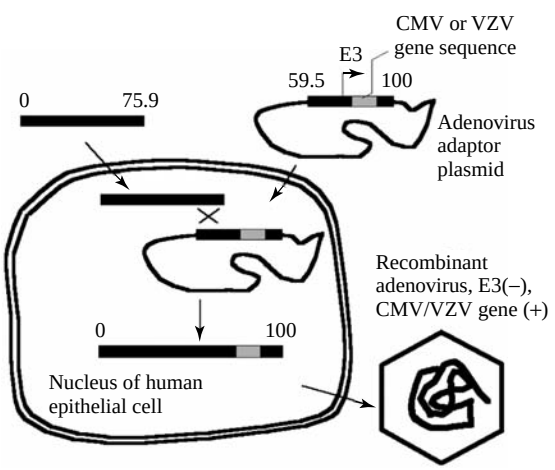

Fig. 6 Production of recombinant, replication-competent adenoviruses to express *CMV* and *VZV* genes. This diagram displays the rare reciprocal recombination event that occurs in cells cotransfected with overlapping fragments of the complete adenovirus DNA genome (0–100 map units), with one plasmid-associated fragment (59.6–100 map units) containing herpesviral gene sequences in place of adenoviral E3 coding sequences downstream of the E3 promoter. The infectious icosahedral virion of the recombinant adenovirus is displayed at the bottom right.

adenovirus genome (map position 1–11 in Fig. 6), and recombinant viruses with this sequence deleted can only replicate in special cell lines that constitutively produce the viral *E1* gene products.

Replication-competent adenoviruses that expressed the full-length *CMV gB* (UL55) gene or part of the *IE1* (UL123) nonstructural gene were capable of eliciting neutralizing antibody and/or cytotoxic T-cell responses following intraperitoneal or intranasal inoculation into experimental animals (Table 5). These immune responses occurred despite the fact that adenovirus infection is incomplete in these nonhuman hosts, presumably because the adenoviral E3 promoter is expressed at an early time after infection even when intracellular replication is incomplete. In related studies with another human herpesvirus, a recombinant adenovirus expressing the *gB* gene of HSV-1 was able to protect mice against a lethal challenge with mouse-adapted HSV-1. Human trials with enteric-coated live-recombinant adenovirus type 5 will be necessary to demonstrate the efficacy of this vaccine strategy for humans.

4.3 Recombinant VZV Expressing Heterologous Viral Genes

To take advantage of the extensive clinical history and relative safety of the Oka varicella vaccine in humans, this live-attenuated virus strain has been modified by recombinant DNA technology to express genes for other protective and immunogenic viral proteins. These include the hepatitis B virus surface antigen, the gD glycoprotein of HSV-2, and the human immunodeficiency virus envelope glycoprotein. These live-recombinant varicella vaccines, known as R-Oka vaccines, have not yet been tested in humans but have been shown to be immunogenic in guinea pigs, an experimental animal that is susceptible to infection by wild-type and attenuated VZV. In the case of the R-Oka-HSV-2 gB vaccine, immunized guinea pigs were protected against a challenge infection by HSV-2. Live-recombinant virus vectors have also been developed for HSV-1, and recombinant HSV-1 expressing *VZV gE* or *IE62* genes has been shown to be immunogenic in mice. One major disadvantage of using live-recombinant vaccines derived from attenuated VZV or HSV is that the immune response to the foreign gene product would be limited and minimal in immunized individuals who had prior immunity to the herpesvirus vector from either natural infection or vaccination.

5 Plasmid DNA Vaccines for VZV and CMV

The most revolutionary development in vaccines against human virus infection and disease utilizes direct inoculation with plasmid DNA containing foreign gene sequences under the control of eukaryotic transcriptional signals, first reported in animal immunization studies with influenza viral genes in 1993. This technology has since been applied to a large number of other viral genes and genes of other infectious agents, and would be most applicable to viral genes that have been shown to induce protective antibody or cytotoxic T-cell responses when expressed by live-recombinant viral vaccines.

The basic requirements for the plasmid DNA vaccines are the same as those for plasmids used for transient gene expression in transfected mammalian cells in culture: upstream promoter and enhancer sequences recognized

Tab. 6 Experimental plasmid DNA vaccines for CMV and VZV.

Gene product	*Immune responses*
CMV gB, TgB, and gB(-tmd) (Fig. 3)	Neutralizing antibody in mice, with higher titers with gB(-tmd) or TgB vs full-length gB; priming for booster immunization with purified gB protein subunit vaccine
CMV pp65	Cytotoxic T-cell response and nonneutralizing antibody
CMV pp65 with kinase-deficient mutation	Cytotoxic T-cell response similar to that with wild-type gene
VZV TgB and TgE (Fig. 3)	Strong neutralizing antibody against gB in mice, but only weak antibody response to gE
VZV gE or TgE	Antibody responses to both in mice, but different antibody classes indicative of Th1 (gE) vs Th2 (TgE) helper T-cell responses
VZV gE and IE62	Antibody and T-cell proliferative responses to both gene products in mice

by host cell transcriptional machinery, coding sequences with in-frame ribosomal initiation and termination codons, and downstream polyadenylation and RNA termination sequences. Routes of inoculation of plasmid DNA in animal studies have included epidermal, intramuscular, mucosal, and intravenous.

These DNA vaccines are safer than live-recombinant vaccines because they do not replicate once they have entered the cell nucleus. Also, the same plasmid DNA vector can be used multiple times with either the same or different foreign gene sequences, without the immunized host developing immunity to the vector. The production and purification of bacterial plasmid DNA is also much less expensive than the purification of protein for subunit vaccines or the production and purification of live-recombinant virus vaccines.

The DNA vaccines designed to prevent primary human CMV infection have employed the coding sequences for the major targets of neutralizing antibody and cytotoxic T cells in humans (Table 6). Plasmid DNA encoding a truncated gB glycoprotein (equivalent to TgB in Fig. 3) was able to induce significant levels of gB-specific neutralizing antibody in a majority of mice after three intramuscular doses, and half of the mice inoculated with a pp65-expressing plasmid developed pp65-specific cytotoxic T-cell responses. In related studies with another human herpesvirus, mice inoculated with a plasmid expressing the HSV-1 gB protein developed high levels of gB-specific antibody and were protected from a lethal challenge with HSV-1.

Current studies with DNA vaccines for VZV (Table 6) have utilized genes encoding full-length or truncated forms of both gE and gB envelope glycoprotein in order to induce neutralizing antibody in mice. Also, immunization with plasmid DNA–containing sequences for IE62 has induced T-cell responses in mice. Similar studies will need to be conducted in guinea pigs because these animals are susceptible to infection with human VZV and they provide an experimental model for directly analyzing whether vaccine-induced immune responses are protective.

Optimization of immune responses to DNA vaccines for VZV or CMV may require a multidose combination of plasmid DNA immunization along with live-attenuated virus or protein subunit vaccines, with initial testing in animals prior to phase I human clinical trials.

6 Clinical Testing of New Vaccines

6.1 Single or Combined Vaccine Strategies in Human Clinical Trials

The live-attenuated CMV strain (Towne), known as *cytomegalovirus vaccine* live and currently produced by Pasteur Merieux Connaught, has undergone a number of human clinical trials over the last two decades (Table 7). Although immunogenic and safe, this vaccine has not been effective in preventing primary infection with CMV except in experimental studies using low-dose challenge with the virulent Toledo strain of CMV, unlike the efficacy of the live-attenuated varicella vaccine in preventing natural infection with VZV. The most disappointing results involved the immunization of CMV-seronegative females of childbearing age, who were not protected from naturally acquiring CMV infection from their own child. The most positive results have been in reducing the

Tab. 7 Human clinical trials of CMV and VZV[a] vaccines.

Year of report	*Vaccine*	*Experimental vaccine population*
1982	Live-attenuated CMV (Towne)	CMV-seronegative pediatric nurses of childbearing age
1984, 1988, 1991, 1994	Live-attenuated CMV (Towne)	CMV-seronegative and seropositive renal transplant candidates
1984, 1996	Live-attenuated CMV (Towne)	Healthy CMV-seronegative adults
1989	Live-attenuated CMV (Towne)	Healthy CMV-seronegative adults, subsequently challenged with low-passage isolate of CMV
1990	Nonrecombinant CMV gB protein from virus-infected cells	CMV-seronegative and seropositive adults
1998	Live-attenuated CMV (Towne)	CMV-seronegative women of childbearing age
1999	Recombinant CMV gB(-tmd) protein from CHO cells	CMV-seronegative healthy adults aged 18 to 50 years
1999, 2002	Recombinant canarypox-CMV gB virus	CMV-seronegative adults, subjected to booster immunization with CMV (Towne) or recombinant gB(-tmd) protein
2001	Recombinant canarypox-CMV pp65 virus	CMV-seronegative adults
2002	Recombinant CMV gB(-tmd) protein from CHO cells	CMV-seronegative children aged 12 to 35 months
1996,1998, 2000, 2001	Live-attenuated Oka varicella vaccine	Healthy VZV-seropositive adults aged 55 years and older
1996, 2000	Heat-inactivated Oka varicella vaccine	Healthy VZV-seropositive adults aged 55 years and older
1997, 2002	Heat-inactivated Oka varicella vaccine	VZV-seropositive bone marrow transplant candidates or recipients

[a] Does not include prelicensure trials of Oka varicella vaccine.

severity of CMV disease in seronegative recipients of organ transplants from CMV-seropositive donors.

One effort to increase the immunogenicity of the Towne vaccine has involved combining a priming immunization with replication-defective canarypox virus expressing CMV gB and subsequently boosting with the Towne strain. The MedImmune biotechnology company has been replacing parts of the Towne strain DNA genome with sequences from the low-passage Toledo strain in order to create a less-attenuated and potentially more-immunogenic vaccine. Initial human studies of these chimeric Towne–Toledo strains of CMV are under way.

A more recently developed experimental strategy for immunization against CMV in humans is the gB subunit vaccine produced in genetically modified CHO cells by the Chiron Corporation. This recombinant gB protein [gB(-tmd) in Fig. 3] is administered by intramuscular injection along with an oil–water emulsion of squalene as adjuvant, and has proved to be both safe and immunogenic after two or three doses in a number of small-scale trials in seronegative adults and children (Table 7). A human clinical trial using prior or simultaneous immunization with the live-recombinant canarypox-CMV gB virus did not improve the immune response to the Chiron gB subunit vaccine. It has not yet been demonstrated whether the induced levels of gB-specific neutralizing antibody will be protective against primary infection with CMV. The limited effectiveness of similar recombinant glycoprotein subunit vaccines in preventing sexual transmission and infection by HSV-2 glycoproteins has tempered the hope and excitement for this vaccine strategy against CMV and other human herpesviruses.

Recent human clinical trials of varicella vaccines have focused upon the use of the live-attenuated Oka varicella vaccine and heat-inactivated Oka vaccine to prevent or reduce the severity of herpes zoster/shingles in elderly seropositive adults who are at risk of VZV reactivation after decades of latency (Tables 1 and 7). These vaccines appear to be effective in boosting VZV-specific immunity without significant risk in healthy adults aged 55 years and older. Ongoing large-scale human trials should indicate whether this boost in immunity translates into long-term protection against zoster and postherpetic neuralgia.

The inactivated varicella vaccine, which carries no risk in immunodeficient recipients, has also been tested in recipients of bone marrow transplants, because these individuals are at especially high risk of serious VZV reactivation and disease. Immunization prior to the transplant and in the three months following the transplant has been shown to dramatically decrease the incidence of zoster, so that use of inactivated varicella vaccine may soon be a standard procedure in this special patient population.

6.2 Animal Models Including Animal Homologues of Human CMV and VZV

There is no nonhuman animal host for human CMV replication and disease; so animals that are experimentally immunized with CMV protein subunit or live-recombinant vaccines cannot be directly challenged with human CMV to determine whether antibody and/or cytotoxic T-cell responses are protective. There are homologous cytomegaloviruses in mice (MCMV) and guinea pigs (GPCMV), and infections of the corresponding hosts

with these two animal herpesviruses result in diseases with some similarity to human CMV disease. The GPCMV-guinea pig system is probably a better biological and clinical model system than the MCMV-mouse system because transplacental transmission and congenital disease has been demonstrated for GPCMV.

Experimental GPCMV vaccine strategies have included subunit vaccines composed of purified viral envelope glycoproteins, live-vaccines using tissue-culture passaged and attenuated GPCMV, and DNA vaccines encoding GPCMV gB or pp65. The purified glycoprotein and live-attenuated vaccines provided significant immunity to female guinea pigs immunized prior to pregnancy, and reduced the incidence and severity of congenital GPCMV in pregnant animals challenged with virulent GPCMV. Several vaccine strategies with MCMV vaccines have also been successful in protecting mice from challenge with lethal or sublethal doses of virulent MCMV. These vaccines include live-attenuated MCMV, DNA vaccines encoding the MCMV homologues of pp65 and IE1, and recombinant vaccinia virus encoding a chimeric protein with a key IE1 peptide recognized by cytotoxic T cells.

A model system with guinea pigs has been developed with a guinea pig–adapted strain of human VZV and can therefore be used for VZV-challenge following immunization with protein subunit, live-recombinant, or plasmid DNA vaccines. For example, a live-recombinant vaccinia virus expressing the gene for the IE62 tegument protein of VZV or a recombinant baculovirus-derived gE protein was able to protect guinea pigs against challenges with the guinea pig–adapted strain of VZV. There is a human VZV-like virus in monkeys, simian varicella virus, but this model system has not been extensively developed or utilized in varicella vaccine studies.

6.3 Quantitation of Immune Responses in Immunized Animals and Humans

The ideal assay for quantitation of protective immune responses following immunization with human CMV or VZV vaccines is the challenge of immunized animals or humans with a defined amount of infectious natural virus, with determination of whether infection and disease is reduced in vaccinated individuals. Experimental challenges of humans with CMV or VZV are usually not allowed for ethical reasons, so that experimental challenge with virulent human virus following immunization is possible only with the guinea pig model with human VZV. Thus, most of the quantitation of immune responses to protein subunit, live-recombinant, and DNA vaccines for CMV and VZV has utilized indirect assays for protective immune responses: quantitation of neutralizing antibody and quantitation of helper/cytotoxic T-cell responses.

Neutralizing antibody titers are more reliable than titers of antigen binding from simple enzyme-linked immunosorbent assays (ELISAs). Neutralizing antibody assays are performed by mixing serial dilutions of serum from immunized animals and control, unimmunized animals with a defined amount of infectious CMV or VZV and then inoculating human fibroblast cells in culture and quantitating a reduction in viral plaques, viral antigen production, or viral cytopathic effects. Neutralizing antibody from human clinical trials with CMV or VZV vaccines can be quantitated in the same cell culture assays as those used for antibody from experimental animals. These virus neutralization

assays are usually performed in the presence and absence of complement because many antibodies to CMV and VZV glycoproteins require complement for neutralization activity. This complement-dependent neutralization is characteristic of some herpesviruses but uncharacteristic of neutralizing antibodies for most other lipid-enveloped human viruses, such as measles, influenza, and rabies viruses.

Specific helper T-cell responses are usually determined by removing peripheral blood mononuclear cells from immunized humans or spleen cells from immunized animals, and quantitating cell proliferation following incubation in the presence or absence of specific viral antigen, typically recombinant viral protein. Quantitation of cytotoxic T-cell responses to CMV or VZV antigens in the immunized animals or humans requires antigen-presenting cells from the same host or genetically identical host that are synthesizing CMV or VZV proteins. Recombinant vaccinia viruses expressing specific *CMV* or *VZV* genes are useful for creating these target cells in cell culture because of the wide host range of vaccinia virus for different species and tissue types.

Protective immune responses in humans following immunization with new vaccines for CMV or VZV can also be quantitated in short-term studies of select high-risk populations, such as seronegative candidates or recipients of organ transplants or bone marrow transplants, or in long-term studies of seronegative but immunocompetent individuals, by following the natural incidence of virus infection and disease in the immunized population versus a corresponding nonimmunized population. Conclusive evidence of long-term immunological protection and safety is necessary before a human CMV vaccine can be licensed for general use or a new human VZV vaccine is licensed to replace or supplement the current live-attenuated varicella vaccine.

7 Concluding Remarks

As use of the licensed Oka varicella vaccine in normal childhood immunization increases to high levels, the number of cases of varicella in children and susceptible adults should decrease significantly and almost all cases of serious complications of varicella can be eliminated. The incidence and severity of zoster due to reactivation of the Oka strain of VZV in future decades may be significantly lower than that with wild-type VZV, especially if a booster immunization dose is recommended in the future. For the hundreds of millions of VZV-seropostive adults who had varicella as children prior to the availability of this vaccine, therapeutic immunization with live or inactivated Oka varicella vaccines or other vaccines that stimulate cytotoxic T-cell responses to VZV has the potential for reducing or eliminating zoster and the accompanying postherpetic neuralgia.

The future benefits of CMV vaccines are less certain in regard to congenital CMV and CMV-related disease in immunocompromised adults. Induction of protective immune responses in CMV-seronegative humans, especially in females prior to reaching the age of childbearing, may require a combination of several modern vaccine strategies unless a less-attenuated and more-immunogenic live-attenuated CMV strain is successful in clinical trials. Current vaccine strategies are beneficial in preventing severe CMV disease in the limited population of CMV-seronegative candidates for transplantation even if immunization does not prevent primary infection.

See also Medicinal Chemistry; RNA Virus Genome Packaging.

Bibliography

Books and Reviews

Gershon, A.A. (2001) Live-attenuated varicella vaccine, *Infect. Dis. Clin. North Am.* **15**, 65–81.

Levin, M.J. (2001) Use of varicella vaccines to prevent herpes zoster in older individuals, *Arch. Virol. Suppl.* **17**, 151–160.

National Institute of Allergy and Infectious Diseases (2002) *The Jordan Report 20th Anniversary: Accelerated Development of Vaccines* 2002, National Institutes of Health, U.S. Department of Health and Human Services [www.niaid.nih.gov/dmid/vaccines/jordan20/].

Pass, R.F., Burke, R.L. (2002) Development of cytomegalovirus vaccines: prospects for prevention of congenital CMV infection, *Semin. Pediatr. Infect. Dis.* **13**, 196–204.

Plotkin, S.A. (2001) Vaccination against cytomegalovirus, *Arch. Virol. Suppl.* **17**, 121–134.

Plotkin, S.A., Orenstein, W.A., Zorab, R. (Eds.) (1999) *Vaccines*, 3rd edition, W. B. Saunders, Philadelphia.

Richman, D.D., Whiteley, R.J., Hayden, F.G. (Eds.) (2002) *Clinical Virology*, 2nd edition, ASM Press, Washington, DC.

Takahashi, M. (2001) 25 years' experience with the Biken Oka strain varicella vaccine: a clinical overview, *Paediatr. Drugs* **3**, 285–292.

Temperton, N.J. (2002) DNA vaccines against cytomegalovirus: current progress, *Int. J. Antimicrob. Agents* **19**, 169–172.

World Health Organization (2001) *The WHO Position Paper on Varicella Vaccines*, WHO [www.who.int/vaccines-diseases/diseases/PP_Varicella.shtml].

Primary Literature

Abendroth, A., Slobedman, B., Springer, M.L., Blau, H.M., Arvin, A.M. (1999) Analysis of immune responses to varicella zoster viral proteins induced by DNA vaccination, *Antiviral Res.* **44**, 179–192.

Adler, S.P., Hempfling, S.H., Starr, S.E., Plotkin, S.A., Riddell, S. (1998) Safety and immunogenicity of the Towne strain of cytomegalovirus vaccine, *Pediatr. Infect. Dis.* **17**, 200–206.

Adler, S.P., Plotkin, S.A., Gonczol, E., Cadoz, M., Meric, C., Wang, J.B., Dellamonica, P., Best, A.M., Zahradnik, J., Pincus, S., Berencsi, K., Cox, W.I., Gyulai, Z. (1999) A canarypox vector expressing cytomegalovirus (CMV) glycoprotein B primes for antibody responses to a live attenuated CMV vaccine (Towne), *J. Infect. Dis.* **180**, 843–846.

Ampofo, K., Saiman, L., LaRussa, P., Steinberg, S., Annunziato, P., Gershon, A. (2002) Persistence of immunity to live attenuated varicella vaccine in healthy adults, *J. Clin. Infect. Dis.* **34**, 774–779.

Baghian, A., Chouljenko, V.N., D'Auvergne, O., Newman, M.J., Baghian, S., Kousoulas, K.G. (2002) Protective immunity against lethal HSV-1 challenge in mice by nucleic acid-based immunization with herpes simplex virus type-1 genes specifying glycoproteins gB and gD, *J. Med. Microbiol.* **51**, 350–357.

Berencsi, K., Gonczol, E., Endresz, V., Kough, J., Takeda, S., Gyulay, Z., Plotkin, S.A., Rando, R.F. (1996) The N-terminal 303 amino acids of the human cytomegalovirus envelope glycoprotein B (UL55) and the exon 4 region of the major immediate early protein 1 (UL123) induce a cytotoxic T-cell response, *Vaccine* **14**, 369–374.

Berencsi, K., Gyulay, Z., Gonczol, E., Pincus, S., Cox, W.I., Michelson, S., Kari, L., Meric, C., Cadoz, M., Zahradnik, J., Starr, S., Plotkin, S. (2001) A canarypox vector-expressing cytomegalovirus (CMV) phosphoprotein 65 induces long-lasting cytotoxic T cell responses in human CMV-seronegative subjects, *J. Infect. Dis.* **183**, 1171–1179.

Bernstein, D.I., Schleiss, M.R., Berencsi, K., Gonczol, E., Dickey, M., Khoury, P., Cadoz, M., Meric, C., Zahradnik, J., Duliege, A.M., Plotkin, S. (2002) Effect of previous or simultaneous immunization with canarypox expressing cytomegalovirus (CMV) glycoprotein B (gB) on response to subunit gB vaccine plus MF59 in healthy CMV-seronegative adults, *J. Infect. Dis.* **185**, 686–690.

Bia, F.J., Miller, S.A., Lucia, H.L., Griffith, B.P., Tarsioi, M., Hsiung, G.D. (1984) Vaccination

against transplacental cytomegalovirus transmission: vaccine reactivation and efficacy in guinea pigs, *J. Infect. Dis.* **149**, 355–362.

Bold, S., Ohlin, M., Garten, W., Radsak, K. (1996) Structural domains involved in cytomegalovirus glycoprotein B-mediated cell fusion, *J. Gen. Virol.* **77**, 2297–2302.

Boppana, S.B., Rivera, L.B., Fowler, K.B., Mach, M., Britt, W.J. (2001) Intrauterine transmission of cytomegalovirus to infants of women with preconceptional immunity, *N. Engl. J. Med.* **344**, 1366–1371.

Bourne, N., Schleiss, M.R., Bravo, F.J., Bernstein, D.I. (2001) Preconception immunization with a cytomegalovirus (CMV) glycoprotein vaccine improves pregnancy outcome in a guinea pig model of congenital CMV infection, *J. Infect. Dis.* **183**, 59–64.

Britt, W., Fay, J., Seals, J., Kensil, C. (1995) Formulation of an immunogenic human cytomegalovirus vaccine: responses in mice, *J. Infect. Dis.* **171**, 18–25.

Britt, W.J., Vugler, L., Butfiloski, E.J., Stephens, E.B. (1990) Cell surface expression of human cytomegalovirus (HCMV) gp55-116 (gB): use of HCMV-recombinant vaccinia virus-infected cells in analysis of the human neutralizing antibody response, *J. Virol.* **64**, 1079–1085.

Cadoz, M., Strady, A., Meignier, B., Taylor, J., Tartaglia, J., Paoletti, E., Plotkin, S. (1992) Immunisation with canarypox virus expressing rabies glycoprotein, *Lancet* **339**, 1429–1432.

Corey, L., Langenberg, A.G., Ashley, R., Sekulovich, R.E., Izu, A.E., Douglas, J.M., Handsfield, H.H., Warren, T., Marr, L., Tyring, S., DiCarlo, R., Adimora, A.A., Leone, P., Dekker, C.L., Burke, R.L., Leong, W.P., Straus, S.E., Chiron HSV Vaccine Study Group (1999) Recombinant glycoprotein vaccine for the prevention of genital HSV-2 infection: two randomized controlled trials, *JAMA* **282**, 331–340.

De Val, M., Schlicht, H.J., Volkmer, H., Messerle, M., Reddehase, M.J., Koszinowski, U.H. (1991) Protection against lethal cytomegalovirus infection by a recombinant vaccine containing a single nonameric T-cell epitope, *J. Virol.* **65**, 3641–3646.

Drulak, M.W., Malinoski, F.J., Fuller, S.A., Stewart, S.S., Hoskin, S., Duliege, A.M., Sekulovich, R., Burke, R.L., Winton, S. (2000) Vaccination of seropositive subjects with CHIRON CMV gB subunit vaccine combined with MF59 adjuvant for production of CMV immune globulin, *Viral Immunol.* **13**, 49–56.

Emery, V.C., Hassan-Walker, A.F., Burroughs, A.K., Griffiths, P.D. (2002) Human cytomegalovirus (HCMV) replication dynamics in HCMV-naive and -experienced immunocompromised hosts, *J. Infect. Dis.* **185**, 1723–1728.

Endresz, V., Burian, K., Berencsi, K., Gyulay, Z., Kari, L., Horton, H., Virok, D., Meric, C., Plotkin, S.A., Gonczol, E. (2001) Optimization of DNA immunization against human cytomegalovirus, *Vaccine* **19**, 3972–3980.

Endresz, V., Kari, L., Berencsi, K., Kari, C., Byulai, Z., Jeney, C., Pincus, S., Rodeck, U., Meric, C., Plotkin, S.A., Gonczol, E. (1999) Induction of human cytomegalovirus (HCMV)-glycoprotein B (gB)-specific neutralizing antibody and phosphoprotein 65 (pp65)-specific cytotoxic T lymphocyte responses by naked DNA immunization, *Vaccine* **17**, 50–58.

Frankenberg, N., Pepperl-Klindworth, S., Meyer, R.G., Plachter, B. (2002) Identification of a conserved HLA-A2-restricted decapeptide from the IE1 protein (pUL123) of human cytomegalovirus, *Virology* **295**, 208–216.

Frey, S.E., Harrison, C., Pass, R.F., Yang, E., Boken, D., Sekulovich, R.E., Percell, S., Izu, A.E., Hirabayashi, S., Burke, R.L., Duliege, A.M. (1999) Effects of antigen dose and immunization regimens on antibody responses to a cytomegalovirus glycoprotein B subunit vaccine, *J. Infect. Dis.* **180**, 1700–1703.

Galil, K., Brown, C., Lin, F., Seward, J. (2002) Hospitalizations for varicella in the United States, 1988 to 1999, *Pediatr. Infect. Dis. J.* **21**, 931–935.

Galil, K., Lee, B., Strine, T., Carraher, C., Baughman, A.L., Eaton, M., Montero, J., Seward, J. (2002) Outbreak of varicella at a day-care center despite vaccination, *N. Engl. J. Med.* **347**, 1909–1915.

Gallez-Hawkins, G., Lomeli, N.A., Li, X., Yao, Z.Q., La Rosa, C., Diamond, D.J., Zaia, J.A. (2002) Kinase-deficient CMV pp65 triggers a CMV pp56 specific T-cell immune response in HLA-A*0201. Kb transgenic mice after DNA immunization, *Scand. J. Immunol.* **55**, 592–598.

Garcia-Valcarcel, M., Fowler, W.J., Harper, D.R., Jeffries, D.J., Layton, G.T. (1997) Induction of neutralizing antibody and T-cell responses to varicella-zoster virus (VZV) using Ty-virus-like

particles carrying fragments of glycoprotein E (gE), *Vaccine* **15**, 709–719.

Gerna, G., Percivalle, E., Baldanti, F., Revello, M.G. (2002) Lack of transmission to polymorphonuclear leukocytes and human umbilical vein endothelial cells as a marker of attenuation of human cytomegalovirus, *J. Med. Virol.* **66**, 335–339.

Gomi, Y., Sunamachi, H., Mori, Y., Nagaike, K., Takahashi, M., Yamanishi, K. (2002) Comparison of the complete DNA sequence of the Oka varicella vaccine and its parental virus, *J. Virol.* **76**, 11447–11459.

Gonczol, E., Berencsi, K., Pincus, S., Endresz, V. Meric, C., Paoletti, E., Plotkin, S.A. (1995) Preclinical evaluation of an ALVAC (canarypox)-human cytomegalovirus glycoprotein B vaccine candidate, *Vaccine* **13**, 1080–1085.

Gonczol, E., deTaisne, C., Hirka, G., Berencsi, K., Lin, W.C., Paoletti, E., Plotkin, S.A. (1991) High expression of human cytomegalovirus (HCMV)-gB protein in cells infected with a vaccinia-gB recombinant: the importance of the gB protein in HCMV immunity, *Vaccine* **9**, 631–637.

Gonczol, E., Ianacone, J., Ho, W.Z., Starr, S., Meignier, B., Plotkin, S. (1990) Isolated gA/gB glycoprotein complex of human cytomegalovirus envelope induces humoral and cellular immune-responses in human volunteers, *Vaccine* **8**, 130–136.

Gyulai, Z., Endresz, V., Burian, K., Pincus, S., Toldy, J., Cox, W.I., Meric, C., Plotkin, S., Gonczol, E., Berencsi, K. (2000) Cytotoxic T lymphocyte (CTL) responses to human cytomegalovirus pp65, IE1-exon4, gB, pp150, and pp28 in healthy individuals: reevaluation of prevalence of IE1-specific CTLs, *J. Infect. Dis.* **181**, 1537–1546.

Harrison, C.J., Britt, W.J., Chapman, N.M., Mullican, J., Tracy, S. (1995) Reduced congential cytomegalovirus (CMV) infection after maternal immunization with a guinea pig CMV glycoprotein before gestational primary CMV infection in the guinea pig model, *J. Infect. Dis.* **172**, 1212–1220.

Hasan, U.A., Harper, D.R., Wren, B.W., Morrow, W.J. (2002) Immunization with a DNA vaccine expressing a truncated form of varicella zoster virus glycoprotein E, *Vaccine* **20**, 1308–1315.

Hata, A., Asanuma, H., Rinki, M., Sharp, M., Wong, R.M., Blume, K., Arvin, A.M. (2002) Use of an inactivated varicella vaccine in recipients of hematopoietic-cell transplants, *N. Engl. J. Med.* **347**, 26–34.

Haumont, M., Jurdan, M., Kangro, H., Jacquet, A., Massaer, M., Deleersnyder, V., Garcia, L., Bosseloir, A., Bruck, C., Bollen, A., Jacobs, P. (1997) Neutralizing antibody responses induced by varicella-zoster gE and gB glycoproteins following infection, reactivation or immunization, *J. Med. Virol.* **53**, 63–68.

He, H., Rinaldo, C.R., Morel, P.A. (1995) T cell proliferative responses to five human cytomegalovirus proteins in healthy seropositive individuals: implications for vaccine development, *J. Gen. Virol.* **76**, 1603–1610.

Heineman, T.C., Connelly, B.L., Bourne, N., Stanberry, L.R., Cohen, J. (1995) Immunization with recombinant varicella-zoster virus expressing herpes simplex virus type 2 glycoprotein D reduces the severity of genital herpes in guinea pigs, *J. Virol.* **69**, 8109–8113.

Huber, M.T., Compton, T. (1999) Intracellular formation and processing of the heterotrimeric gH-gL-gO (gCIII) glycoprotein envelope complex of human cytomegalovirus, *J. Virol.* **73**, 3886–3892.

Hwang, E.S., Kwon, K.B., Park, J.W., Kim, D.J., Park, C.G., Cha, C.Y. (1999) Induction of neutralizing antibody against human cytomegalovirus (HCMV) with DNA-mediated immunization of HCMV glycoprotein B in mice, *Microbiol. Immunol.* **43**, 307–310.

Jacquet, A., Haumont, M., Massaer, M., Garcia, L., Mazzu, P., Daminet, V., Gregoire, D., Jacobs, P., Bollen, A. (2002) Immunogenicity of a recombinant varicella-zoster gE-IE63 fusion protein, a putative vaccine candidate against primary infection and zoster reactivation, *Vaccine* **20**, 1593–1602.

Kamiyama, T., Sato, H., Takahara, T., Kageyama, S., Shiraki, K. (2000) Novel immunogenicity of Oka varicella vaccine vector expressing hepatitis B surface antigen, *J. Infect. Dis.* **181**, 1158–1161.

Kimura, H., Wang, Y., Pesnicak, L., Cohen, J.I., Hooks, J.J., Straus, S.E., Williams, R.K. (1998) Recombinant varicella-zoster virus glycoproteins E and I: immunologic responses and clearance of virus in a guinea pig model of chronic uveitis, *J. Infect. Dis.* **178**, 310–317.

Klein, M., Schoppel, K., Amvrossiadis, N., Mach, M. (1999) Strain-specific neutralization of human cytomegalovirus isolates by human sera, *J. Virol.* **73**, 878–886.

Krause, P.R., Klinmann, D.M. (2000) Varicella vaccination: evidence for frequent reactivation of the vaccine in healthy children, *Nat. Med.* **6**, 451–454.

Kuhn, E.M., Stolte, N., Matz-Rensing, K., Mach, M., Stahl-Henning, C., Hunsmann, G., Kaup, F.J. (1999) Immunohistochemical studies of productive rhesus cytomegalovirus infection in rhesus monkeys (Macaca mulatta) infected with simian immunodeficiency virus, *Vet. Pathol.* **36**, 51–56.

Kutinova, L., Hainz, P., Ludvikova, V., Maresova, L., Nemeckova, S. (2001) Immune response to vaccinia virus recombinants expressing glycoproteins gE, gB, gH, and gL of varicella-zoster virus, *Virology* **280**, 211–220.

La Rosa, C., Wang, Z., Brewer, J.C., Lacey, S.F., Villacres, M.C., Sharan, R., Krishman, R., Crooks, M., Markel, S., Maas, R., Diamon, D.J. (2002) Preclinical development of an adjuvant-free peptide vaccine with activity against CMV pp65 in HLA transgenic mice, *Blood* **100**, 3681–3689.

Lantto, J., Lindroth, Y., Ohlin, M. (2002) Non-germ line encoded residues are critical for effective antibody recognition of a poorly immunogenic neutralization epitope on glycoprotein B of human cytomegalovirus, *Eur. J. Immunol.* **32**, 1659–1669.

Lau, Y.L., Vessey, S.J., Chan, I.S., Lee, T.L., Huang, L.M., Lee, C.Y., Lin, T.Y., Lee, B.W., Kwan, K., Kasim, S.M., Chan, C.Y., Kaplan, K.M., Distefano, D.J., Harmon, A.L., Golie, A., Hartzel, J., Xu, J., Li, S., Matthews, H., Sadoff, J.C., Shaw, A. (2002) A comparison of safety, tolerability and immunogenicity of Oka/Merck varicella vaccine and VARILRIX in healthy children, *Vaccine* **20**, 2942–2949.

Levin, M.J., Ellison, M.C., Zerbe, G.O., Barber, D., Chan, C., Stinson, D., Jones, M., Hayward, A.R. (2000) Comparison of a live attenuated and an inactivated varicella vaccine to boost the varicella-specific immune response in seropositive people 55 years of age and older, *Vaccine* **18**, 2915–2920.

Li, S., Chan, I.S., Matthews, H., Heyse, J.F., Chan, C.Y., Kuter, B.J., Kaplan, K.M., Vessey, S.J., Sadoff, J.C. (2002) Inverse relationship between six week postvaccination varicella antibody response to vaccine and likelihood of long term breakthrough infection, *Pediatr. Infect. Dis. J.* **21**, 337–342.

Liu, Y.N., Curtsinger, J., Donahue, P.R., Klaus, A., Optiz, G., Cooper, J., Karr, R.W., Bach, F.H., Gehrz, R.C. (1993) Molecular analysis of the immune response to human cytomegalovirus glycoprotein B. I. Mapping of HLA-restricted helper T cell epitopes on gp93, *J. Gen. Virol.* **74**, 2207–2214.

Lowe, R.S., Keller, R.M., Keech, B.J., Davison, A.J., Whang, Y., Morgan, A.J., Kieff, E., Ellis, R.W. (1987) Varicella-zoster virus as a live vector for the expression of foreign genes, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3896–3900.

Lowry, P.W., Koropchak, C.M., Choi, C.Y., Mocarski, E.S., Kern, E.R., Kinchington, P.R., Arvin, A.M. (1997) The synthesis and immunogenicity of varicella-zoster virus glycoprotein gE and immediate-early protein IE62 (IE62) expressed in recombinant herpes simplex virus-1, *Antiviral Res.* **33**, 187–200.

MacDonald, M.R., Li, X.Y., Stenberg, R.M., Campbell, A.E., Virgin, H.W. (1998) Mucosal and parenteral vaccination against acute and latent murine cytomegalovirus (MCMV) infection using an attenuated MCMV vaccine, *J. Virol.* **72**, 442–451.

Marshall, G.S., Fenger, D.P., Stout, G.S., Knights, M.A., Hunt, L.A. (1996) Processing of human cytomegalovirus glycoprotein B in recombinant adenovirus-infected cells, *J. Gen. Virol.* **77**, 1549–1557.

Marshall, G.S., Li, M., Stout, G.G., Louthan, M.V., Duliege, A.M., Burke, R.L., Hunt, L.A. (2000) Antibodies to the major linear neutralizing domains of cytomegalovirus glycoprotein B among natural seropositives and CMV subunit vaccine recipients, *Viral Immunol.* **13**, 329–341.

Marshall, G.S., Rabalais, G.P., Stout, G.G., Waldeyer, S.L. (1992) Antibodies to recombinant-derived glycoprotein B after natural human cytomegalovirus infection correlate with neutralizing activity, *J. Infect. Dis.* **165**, 381–384.

Marshall, G.S., Ricciardi, R.P., Rando, R.F., Puck, J., Ge, R.W., Plotkin, S.A., Gonczol, E. (1990) An adenovirus recombinant that expresses the human cytomegalovirus major envelope glycoprotein and induces neutralizing antibodies, *J. Infect. Dis.* **162**, 1177–1181.

Massaer, M., Haumont, M., Garcia, L., Mazzu, L., Bollen, A., Jacobs, P., Jacquet, A. (1999) Different neutralizing antibody responses to varicella-zoster virus glycoproteins

B and E following naked DNA immunization, *Viral Immunol.* **12**, 227–236.

Masuoka, M., Yoshimuta, T., Hamada, M., Okamoto, M., Fuminmori, T., Honda, J., Oizumi, K., Itoh, K. (2001) Identification of the HLA-A24 peptide epitope within cytomegalovirus protein pp65 recognized by CMV-specific cytotoxic T lymphocytes, *Viral Immunol.* **14**, 369–377.

Mitchell, D.K., Holmes, S.J., Burke, R.L., Duliege, A.M., Adler, S.P. (2002) Immunogenicity of a recombinant human cytomegalovirus gB vaccine in seronegative toddlers, *Pediatr. Infect. Dis. J.* **21**, 133–138.

Morello, C.S., Ye, M., Spector, D.H. (2002) Development of a vaccine against murine cytomegalovirus (MCMV), consisting of plasmid DNA and formalin-inactivated MCMV, that provides long-term, complete protection against viral replication, *J. Virol.* **76**, 4822–4835.

Panicali, D., Davis, S.W., Weinberg, R.L., Paoletti, E. (1983) Construction of live vaccines by using genetically engineered poxviruses: biological activity of recombinant vaccinia virus expressing hemagglutinin, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5364–5368.

Pass, R.F., Duliege, A.M., Boppana, S., Sekulovich, R., Percell, S., Britt, W., Burke, R.L. (1999) A subunit cytomegalovirus vaccine based on recombinant envelope glycoprotein B and a new adjuvant, *J. Infect. Dis.* **180**, 970–975.

Pepperl, S., Munster, J., Mach, M., Harris, J.R., Plachter, B. (2000) Dense bodies of human cytomegalovirus induce both humoral and cellular immune responses in the absence of viral gene expression, *J. Virol.* **74**, 6132–6146.

Pepperl-Kindworth, S., Frankenberg, N., Plachter, B. (2002) Development of novel vaccine strategies against human cytomegalovirus infection based on subviral particles, *J. Clin. Virol.* **25**, S75–S85.

Plotkin, S.A., Higgins, R., Kurtz, J.B., Morris, P.J., Campbell, D.A., Shope, T.C., Spector, S.A., Dankner, W.M. (1994) Multicenter trial of Towne strain attenuated virus vaccine in seronegative renal transplant recipients, *Transplantation* **58**, 1176–1178.

Plotkin, S.A., Huang, E.S. (1985) Cytomegalovirus vaccine virus (Towne strain) does not induce latency, *J. Infect. Dis.* **152**, 395–397.

Plotkin, S.A., Smiley, M.L., Friedman, H.M., Starr, S.E., Fleisher, G.R., Wlodaver, C., Dafoe, D.C., Friedman, A.D., Grossman, R.A., Barker, C.F. (1984) Towne-vaccine-induced prevention of cytomegalovirus disease from renal transplantation, *Lancet* **1**, 528–530.

Plotkin, S.A., Starr, S.E., Friedman, H.M., Gonczol, E., Weibel, R.E. (1989) Protective effects of Towne cytomegalovirus vaccine against low-passage cytomegalovirus administered as a challenge, *J. Infect. Dis.* **159**, 860–865.

Robinson, H.L., Hunt, L.A., Webster, R.G. (1993) Protection against a lethal influenza virus challenge by immunization with a haemagglutinin-expressing plasmid DNA, *Vaccine* **11**, 957–960.

Roy, D.M., Grundy, J.E., Emery, V.C. (1993) Sequence variation within neutralizing epitopes of the envelope glycoprotein B of human cytomegalovirus: comparison of isolates from renal transplant recipients and AIDS patients, *J. Gen. Virol.* **74**, 2499–2505.

Sachs, G.W., Simmons, R.L., Balfour, H.H. (1984) Cytomegalovirus vaccine: persistence of humoral immunity following immunization of renal transplant candidates, *Vaccine* **2**, 215–218.

Saiman, L., LaRussa, P.J., Steinberg, S.P., Zhou, J., Baron, K., Whittier, S., Della-Latta, P., Gershon, A.A. (2001) Persistence of immunity to varicella-zoster virus after vaccination of healthcare workers, *Infect. Control Hosp. Epidemiol.* **22**, 279–283.

Schleiss, M.R., Bourne, N., Jensen, N.J., Bravo, F., Bernstein, D.I. (2000) Immuogenicity evaluation of DNA vaccines that target guinea pig cytomegalovirus proteins glycoprotein B and UL83, *Viral Immunol.* **13**, 155–167.

Schoppel, K., Schmidt, C., Einsele, H., Hebart, H., Mach, M. (1999) Kinetics of the antibody response against human cytomegalovirus-specific proteins in allogeneic bone marrow transplant recipients, *J. Infect. Dis.* **178**, 1233–1243.

Seward, J.F., Watson, B.M., Peterson, C.L., Mascola, L., Pelosi, J.W., Zhang, J.X., Maupin, T.J., Goldman, G.S., Tabony, L.J., Brodovicz, K.G., Jumaan, A.O., Wharton, M. (2002) Varicella disease after introduction of varicella vaccine in the United States, 1995–2000, *JAMA* **287**, 606–611.

Sharrar, R.G., LaRussa, P., Galea, S.A., Steinberg, S.P., Sweet, A.R., Keatley, R.M., Wells, M.E., Stephenson, W.P., Gershon, A.A. (2000) The postmarketing safety profile of varicella vaccine, *Vaccine* **19**, 916–923.

Smith, G.L., Mackett, M., Moss, B. (1983) Infectious vaccinia virus recombinants that express hepatitis B virus surface antigen, *Nature* **302**, 490–495.

Spaete, R.R., Thayer, R.M., Probert, W.S., Masiarz, F.R., Chamberlain, S.H., Rasmussen, L., Merigan, T.C., Pachl, C. (1988) Human cytomegalovirus strain Towne glycoprotein B is processed by proteolytic cleavage, *Virology* **167**, 207–225.

Speckner, A., Kropff, B., Knor, S., Mach, M. (2000) The antigenic domain 1 of human cytomegalovirus glycoprotein B contains an intramolecular disulphide bond, *J. Gen. Virol.* **81**, 2659–2663.

Stanberry, L.R., Spruance, S.L., Cunningham, A.L., Bernstein, D.I., Mindel, A., Sacks, S., Tyring, S., Aoki, F.Y., Slaoui, M., Denis, M., Vandepapeliere, P., Dubin, G., GlaxoSmithKline Herpes Vaccine Efficacy Study Group. (2002) Glycoprotein-D-adjuvant vaccine to prevent genital herpes, *N. Engl. J. Med.* **347**, 1652–1661.

Tackaberry, E.S., Dudani, A.K., Prior, F., Tocchi, M., Sardana, R., Altosaar, I., Ganz, P.R. (1999) Development of biopharmaceuticals in plant expression systems: cloning, expression and immunological reactivity of human cytomegalovirus glycoprotein B (UL55) in seeds of transgenic tobacco, *Vaccine* **17**, 3020–3029.

Takahashi, M., Otsuka, T., Okuno, Y., Asano, Y., Yazaki, T. (1974) Live vaccine used to prevent the spread of varicella in children in hospital, *Lancet* **2**, 1288–1290.

Taylor, J., Trimarchi, C., Weinberg, R., Languet, B., Guillemin, F., Desmettre, P., Paoletti, E. (1991) Efficacy studies on a canarypox-rabies recombinant virus, *Vaccine* **9**, 190–193.

Terada, L., Niizuma, T., Ogita, S., Kataoka, N. (2002) Responses of varicella zoster virus (VZV)-specific immunity in seropositive adults after inhalation of inactivated or live attenuated varicella vaccine, *Vaccine* **20**, 3638–3643.

Trannoy, E., Berger, R., Hollander, G., Bailleux, F., Heimendinger, P., Vuillier, D., Creusvaux, H. (2000) Vaccination of immunocompetent elderly subjects with a live attenuated Oka strain of varicella zoster: a randomized, controlled, dose-dependent trial, *Vaccine* **18**, 1700–1706.

Ulmer, J.B., Donnelly, J.J., Parker, S.E., Rhodes, G.H., Felgner, P.L., Dwarki, V.J., Gromkowski, S.H., Deck, R.R., DeWitt, C.M., Friedman, A., Hawe, L.A., Leander, K.R., Martinez, D., Perry, H.C., Shiver, J.W., Montgomery, D.L., Liu, M.A. (1993) Heterologous protection against influenza by injection of DNA encoding a viral protein, *Science* **259**, 1745–1749.

Urban, M., Klein, M., Britt, W.J., Hassfurther, E., Mach, M. (1996) Glycoprotein H of human cytomegalovirus is a major antigen for the neutralizing humoral immune response, *J. Gen. Virol.* **77**, 1537–1547.

Vazquez, M., LaRussa, P.S., Gershon, A.A., Steinberg, S.P., Freudigman, K., Shapiro, E.D. (2001) The effectiveness of the varicella vaccine in clinical practice, *N. Engl. J. Med.* **344**, 955–960.

Vaz-Santiago, J., Lule, J., Rohrlich, P., Jacquier, C., Gilbert, N., Le Roy, E., Betbeder, D., Davignon, J.L., Davrinche, C. (2001) Ex vivo stimulation and expansion of both CD4(+) and CD8(+) T cells from peripheral blood mononuclear cells of human cytomegalovirus-seropositive donors by using a soluble recombinant chimeric protein, IE1-pp65, *J. Virol.* **75**, 7840–7847.

Wang, J.B., Adler, S.P., Hempfling, S., Burke, R.L., Duliege, A.M., Starr, S.E., Plotkin, S.A. (1996) Mucosal antibodies to human cytomegalovirus B occur following both natural infection and immunization with human cytomegalovirus vaccines, *J. Infect. Dis.* **174**, 387–392.

Welsh, M.D., Harper, D.R., Garcia-Valcarcel, M., Fowler, W.J., Aitken, C., Jeffries, D.J., Layton, G.T. (1999) Ability of yeast Ty-VLPs (virus-like particles) containing varicella-zoster virus (VZV) gE and assembly protein fragments to induce in vitro proliferation of human lymphocytes from VZV immune patients, *J. Med. Virol.* **59**, 78–83.

Wise, R.P., Salive, M.E., Braun, M.M., Mootrey, G.T., Seward, J.F., Rider, L.G., Krause, P.R. (2000) Postlicensure safety surveillance for varicella vaccine, *JAMA* **284**, 1271–1279.

Yao, Z.Q., Gallez-Hawkins, G., Lomeli, N.A., Li, X., Molinder, K.M., Diamond, D.J., Zaia, J.A. (2001) Site-directed mutation in a conserved kinase domain of human cytomegalovirus-pp65 with preservation of cytotoxic T lymphocyte targeting, *Vaccine* **19**, 1628–1635.

Ye, M., Morello, C.S., Spector, D.H. (2002) Strong CD8 T-cell responses following

coimmunization with plasmids expressing the dominant pp89 and subdominant M84 antigens of murine cytomegalovirus correlate with long-term protection against subsequent viral challenge, *J. Virol.* **76**, 2100–2112.

Cytoplasmic Signaling: *see* Cellular Interactions

Denaturation of DNA

R. D. Blake
University of Maine, Orono, USA

1 Denaturation 189
1.1 Denaturation of DNA 189

2 Analytical Methods for Following the Denaturation Process 190
2.1 Changes in Absorbance with Temperature 190
2.2 Other Methods 192

3 Factors Affecting the Stability of DNA 193
3.1 Intrinsic Factors 194
3.1.1 (G + C) Content 194
3.1.2 Distribution of Nearest Neighbors 196
3.1.3 Sequence Complexity 197
3.1.4 Conformational States of Neighboring Domains 197
3.1.5 DNA Length and Concentration 197
3.2 Extrinsic Factors 198
3.2.1 The Dependence of Melting Temperature on Counterion Type and Concentration 198
3.2.2 pH 200
3.3 Denaturing Solvents 201

4 Utility of Denaturation Studies 203
4.1 Thermodynamic Analyses 203
4.1.1 Simulation of Melting 203
4.1.2 Denaturation Maps 205
4.1.3 Conditions for Equilibrium Denaturation 205
4.2 Hybridization 205
4.2.1 Stringency 206
4.2.2 The Potential for Degradation 206
4.3 PCR 206

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

4.4 Sequence Variances 206
4.4.1 Identification of Point Mutations and Mispair Thermodynamics 206
4.5 Profiles of Genomic Base Distributions 207

5 Denaturation of RNA 207

Bibliography 211
Books and Reviews 211
Primary Literature 211

Keywords

Degradation
Breakdown of covalent bonds supporting the polynucleotide chains, caused by shearing action, low pHs, and high temperatures. The process is generally irreversible.

Stacking Forces
The forces between stacked base pairs in DNA and bases in RNA.

Melting Curve
The variation with temperature of any property, for example, absorption at 260 nm, sensitive to the native structure of DNA and RNA.

T_M
The melting temperature of DNA or RNA, determined from the midpoint of the change in the dependent variable of the melting curve, for example, absorption at 260 nm, with temperature.

Helix–Coil Transition
The point of conversion of a double-stranded helical DNA to a disordered single-stranded DNA, sometimes described (inappropriately) as random coils.

Denatured DNA and RNA are disordered nonnative states brought about by any of the many means of disturbing noncovalent bonds supporting their native conformations: heat, nonaqueous denaturing solvents, low or high pHs, low ionic strengths. The denaturation process is generally reversible.

1
Denaturation

Denaturation studies and, in particular, melting curves have many uses. They have been important for establishing conditions for the formation of DNA *heteroduplexes* and defining levels of *stringency*, denaturation mapping by electron microscopy, amplification of DNA by the *polymerase chain reaction* (PCR), determining DNA sequence variances of homologous segments and tandem and interspersed repetitive elements, profiling genomic DNA populations and estimating the relative amounts and sequence complexities of satellite repetitive elements, and determination of population and evolutionary relationships. They have been particularly important for confirming aspects of structure and sources of stability as well as for determining thermodynamic quantities associated with those sources, where temperature is an important variable.

The familiar dictum for the relationship of structure and function is especially appropriate to DNA. The strands of DNA form double helical threads of inconceivable great lengths, well suited to their role as repositories of enormous amounts of genetic information in the lineal sequences of nucleotide residues. The structures of ribosomal, transfer, and messenger RNAs, on the other hand, are strikingly different from DNA, consisting of single strands that fold into a variety of compact structural elements with shapes that enable them to serve as agents of recognition by proteins and antibiotics, gene expression, gene transport, transcription attenuation, translation regulatory agents, templates, and catalysts.

Protein and enzyme-induced alterations of the conformations of DNAs and RNAs are natural processes that, however, are not taken into account in definitions of native and denatured states. It is the practice to define the "native" conformations of these molecules as those existing in isolation and under physiological conditions. These molecules can be *denatured* by shifting the order–disorder equilibrium away from that favoring the native conformation by increasing the temperature or, less commonly, by the addition of denaturing solvents. Denaturation is usually a reversible process, so that if physiological conditions are restored, the native structures and biological function are restored. However, if covalent bonds are broken, the molecule is *degraded*. Unlike denaturation, degradation is usually irreversible.

1.1
Denaturation of DNA

Base–base hydrogen bonds maintain the alignment of the chains, but stacking forces appear to be the major source of support. Both are sensitive to conditions and, in particular, to temperature. Stacking forces are responsible for local variations in stability, but hydrogen bonds between bases are responsible for the high level of *cooperativity* within the molecule, resulting in an abrupt approach to the equilibrium boundary between native and denatured states, the helix–coil boundary. The high level of cooperativity reflects an unusually strong dependence of the structural state of each pair on the structural state of its neighbors, an effect that sometimes involves many hundreds of base pairs. In addition to stacking and hydrogen bonds, DNA may derive some small support from solvation effects and the occasional water bridge. On the debit side, entropic effects from restraints placed on rotation about torsional bonds and ionic forces counteract the stabilizing forces

of stacking and hydrogen bonds. Every phosphodiester group (·····O-P[O_2^-]-O·····) that links each nucleotide residue has a negative charge that creates strong repulsive forces that weaken the helix. Besides high temperatures, solvents that lessen the forces of support or promote the destabilizing forces act to shift the helix–coil equilibrium.

With exception for temperature or salt extreme organisms, conditions in cells are within the normal range of stability of DNA. The molecule undoubtedly experiences small conformational palpitations from normal mode thermal motions, but remains native because of a favorable free energy and the *cooperative* nature of the secondary interactions. Thermal motions rarely force covalently joined atoms very far from their equilibrium positions; however, occasionally they are sufficiently turbulent to cause some transient dissociation of noncovalent bonds and base pairs, resulting in a local "breathing" action by the two strands. Under physiological conditions, the probability that a base pair will breathe at any instant in time is only 10^{-5} to 10^{-6}, or one transient opening every 200 000 base pairs. Breathing is a random process, perhaps occurring with slightly greater probability at regions of lower stability, possibly making these regions more accessible to mutagens and therefore as "hot spots" of mutation.

The amount of breathing action increases with increasing temperature, gradually weakening the helix. Nevertheless, cooperative support from neighboring residues acts to restore the helix, so the structure remains essentially native almost up to the melting temperature. If the system is returned to low temperatures, the helix is seen to be unaltered. If, however, the temperature is increased further, thermal motions become more turbulent, lifetimes of pairs decrease, and eventually large numbers of pairs begin breathing in concert. At the melting temperature, T_M, (Fig. 1) where the free energy approaches zero and the equilibrium constant for the helix–coil transition approaches unity, the polynucleotide chains begin to separate, leading to nonhelical single-stranded structures. The abrupt approach to the transition caused by the cooperativity of interactions is seen in the unusual sharpness of thermally induced denaturation or "melting" curves. If all noncovalent bonds are broken, the chains assume the conformation of *random coils*. However, denaturation is rarely, if ever, that complete; rather, the single-stranded DNA assumes a complex disordered state with many regions of (1) single-strand stacking; (2) double-stranded hairpins; and (3) a few segments in which the structure may be random and dynamic.

2 Analytical Methods for Following the Denaturation Process

2.1 Changes in Absorbance with Temperature

Perhaps the most popular method for monitoring the denaturation process is by *electronic absorption spectroscopy*. Intense absorption bands with large extinction coefficients in the ultraviolet region are due to allowed *electronic transition moments* for the bases near 260 nm. The molar absorption coefficient for the helix, $e_{260\ nm}(P)$, is $\sim 7(\pm 0.2) \times 10^3$ L (mol × cm)$^{-1}$, more than 40% *hypochromic* to the constituent nucleotides. During denaturation, hydrogen bonds joining base pairs are broken allowing the bases to unstack, leading to an increase in the

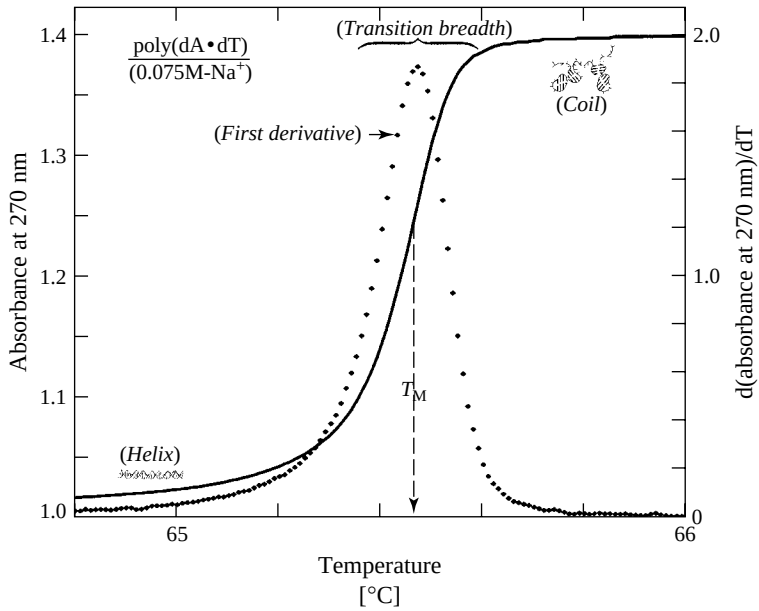

Fig. 1 Thermal denaturation (melting curve) obtained by the variation in absorption at 270 nm with increasing temperature of poly(dA·dT), a synthetic double helical DNA in a buffer consisting of 0.075 M Na^+. The bell-shaped curve of points represents the first derivative, dA_λ/dT. The T_M is usually defined as the temperature corresponding to half the hyperchromic effect. Alternately, it is the peak temperature of the first derivative curve. The free energy approaches zero at the melting temperature, T_M, where the equilibrium constant for the helix–coil transition approaches unity.

molar absorption coefficient to $\sim 10^4$ L $(\text{mol} \times \text{cm})^{-1}$. This value is close to but still less than the value for the constituent nucleotides, since some small amount of residual stacking of the unpaired bases persists beyond the denaturation temperature of the helix. The percent hypochromicity, h (%) at 260 nm, is sometimes given by

$$h_\lambda(\%) = [1 - (\varepsilon_{\text{native},\lambda}/\varepsilon_{\text{denatured},\lambda})] \times 100 \quad (1)$$

where $\varepsilon_{\text{denatured},\lambda}$ is the molar absorption above T_M. As noted by the left-hand scale of Fig. 1, the observed increase in hyperchromicity for the melting of poly(dA·dT), a synthetic double helical DNA, is ~40%. With such large changes, denaturation can be monitored with considerable precision on nucleotide residue concentrations in the micromolar range and, if the precision of data is adequate, denaturation can be represented in the first derivative, dA_λ/dT, which is better for resolving subtransitions.

DNAs of minimal sequence complexity such as poly(dA·dT) (Fig. 1) exhibit the sharpest melting curves. The melting temperature, T_M, determined from the midpoint of the transition or from the peak of the first derivative curve, can be taken as a measure of the thermal stability of the DNA. The transition breadth is an indication of the sequence complexity and

variation in the fractional (G + C) content, denoted by $F_{(G+C)}$, since the properties of (G·C) are identical to those of (C·G) and $F_{(G+C)} = 1 - F_{(A+T)}$.

A further advantage of absorption spectroscopy is that the characteristics of the base composition of natural DNAs of quasi-random sequence can be determined at different temperatures by spectral decomposition. The dissociation of (A·T) base pairs contributes more than 4x more to the absorption change at 260 nm than (G·C) pairs, while almost the reverse is true at 282 nm. Both pairs contribute equally at 270 nm; therefore, this is the preferred wavelength for expressing the ordinate of melting curves as the fraction of base pairs that have denatured.

2.2 Other Methods

Methods for monitoring the denaturation of DNA can be classified as (1) short-range physical; (2) long-range physical; (3) chemical; and (4) biological methods. A partial list of the physical methods that have been used is given in Table 1. Chemical methods include those that probe differences in the reactivity of the bases in the helical and nonhelical states, while biological methods include techniques

Tab. 1 Selected physical methods for using or following denaturation.

Methods that depend on short-range physical properties:

1. *Absorption spectroscopy*.
   Thomas, R. (1954) *Biochim. Biophys. Acta* **14**, 231; Doty, P. and Rice, S.A. (1955) *Biochim. Biophys. Acta* **16**, 446–448; (Some of the first DNA denaturation curves were carried out by taking advantage of the large hyperchromic effect associated with the destacking of base pairs).
2. *Circular dichroism*.
   Gray, D.M., Ratliff, R.L., & Vaughan, M.R. (1992) *Meth. Enzymol.* **211a**, 389. (CD measures the difference in molar extinction of the left and right circularly polarized components of monochromatic light in the region of 200 to 290 nm. As such, it is used to detect the spatial asymmetry and helical handedness as the chromophoric base pairs twist one way or the other around the major axis).
3. *Infrared spectroscopy*.
   Taillanier, E. & Liquier, J. (1992) *Meth. Enzymol.* **211a**, 307.
4. *Raman spectroscopy*.
   Peticolas, W.L. & Evertsz, E. (1992) *Meth. Enzymol.* **211a**, 335. (In the 500 to 1800 cm^{-1} region, IR and Raman reflect changes in *vibrational* and *rotational* modes affected by changes in hydrogen bonding and stacking interactions. Changes involving polar groups give the most intense absorptions and are useful for following the dissociation of hydrogen bonding. This method suffers from spectral interference by the solvent water).
5. *Nuclear magnetic resonance*.
   Feigon, J., Sklenár, V., Wang, E., Gilbert, D.E., Macaya, R.F. & Schultze, P. (1992) *Meth. Enzymol.* **211a**, 235. [One-dimensional NMR spectra (in the radio frequency range $\sim 10^{10}$ Hz) from 9.5 to 0.5 ppm are used mainly for studies of the exchangeable imino and amino resonances occurring during denaturation].
6. *Differential scanning calorimetry*.
   Breslauer, K.J., Freire, E. & Straume, M. (1992) *Meth. Enzymol.* **211a**, 533.

(*continued overleaf*)

Tab. 1 (*continued*)

Methods that depend on long-range physical properties:

1. *Viscometry.*
   Zamenhof, S., Alexander, H.E. & Leidy, G. (1953) *J. Exp. Med.* **98**, 373; Doty, P. & Rice, S.A. (1955) *Biochim. Biophys. Acta* **16**, 446; Rice, S.A. & Doty, P. (1957) *J. Amer. Chem. Soc.* **79**, 3937. (The first thermal denaturation profiles of DNA and bacterial "transforming activity" were obtained from changes in viscosity).
2. *Sedimentation velocity.*
   (Changes in these hydrodynamic properties reflect the reduction in frictional coefficient during denaturation when DNA changes from a rigid rod into flexible chains. These methods are generally more difficult and cumbersome than spectral methods, requiring greater amounts of monodispersive material).
3. *Dynamic light scattering.*
   Soda, K. & Wada, A. (1984) *Biophys. Chem.* **20**, 185; Delrow, J.J., Heath, P.J., Fujimoto, B.S. & Schurr, J.M. (1998) *Biopolymers* **45**, 503; Protozanova, E. & Macgregor, R.B. Jr. (2000) *Biophys. Chem.* **84**, 137.
4. *Electron microscopy.*
   Inman, R.B. (1967) *J. Mol. Biol.* **28**, 103; Borovik, A.S., Kalambet, Y.A., Lyubchenko, Y.L., Shitov, V.T. & Golovanov, E.I. (1980) *Nucl. Acids Research* **8**, 4165. (Partially denatured regions are fixed at intermediate temperatures within the melting region by reaction of the imino and amino groups with the bifunctional reagent glyoxal, thereby preventing reformation of the helix).
5. *Gradient gel electrophoresis.*
   Fischer, S.G. & Lerman, L.S. (1979) *Meth. Enzymol.* **68**, 183; Abrams, E.S. & Stanton, V.P. Jr. (1992) *Meth. Enzymol.* **212b**, 71; Cariello, N.F. & Skopek, T.R. (1993) *Mutat. Res.* **288**, 103; Fodde, R. & Losekoot, M. (1994) *Hum. Mutat.* **3**, 83; Wartell, R.M., Hosseini, S.H., Powell, S. & Zhu, J. (1998) *J. Chromatogr.* **806**, 169.
6. *Denaturing HPLC.*
   Xiao, W. and Oefner, P.J. (2001) *Hum. Mutat.* **17**, 439; Huber, C.G., Premstaller, A., Xiao, W., Oberacher, H., Bonn, G.K. & Oefner, P.J. (2001) *J. Biochem. Biophys. Meth.* **47**, 5.

such as S1 nuclease sensitivity, reaction with antibodies, and transformation methods.

3 Factors Affecting the Stability of DNA

A number of factors from within and without the DNA affect its sensitivity to denaturation: (1) (G + C) content; (2) distribution of nearest neighbors in the DNA; (3) sequence complexity; (4) local sequence and conformational states of neighboring domains; (5) presence of wobble or other mispairs; (6) DNA length; (7) supercoiling; (8) pH; (9) temperature; (10) counterion concentration and type; and (11) various salts and nonaqueous solvents. The use of temperature for denaturing DNA is particularly convenient, figuring prominently in most studies of denaturation. Melting temperatures depend on the *intrinsic* characteristics of base composition ($F_{(G+C)}$), sequence, and chain length of the DNA specimen. Studies of the dependence of T_M on various *extrinsic* factors have been useful in the evaluation of energetics. Assuming that the fraction, f, of base pairs in the ordered (helix, h) and disordered (coil, c) states are equal at T_M, midway through the transition, $(f^c/f^h) = (f^c/[1 - f^c]) = 1$, where the free energy is given by

$$\Delta G_M = -RT\ln(f^c/f^h) = 0$$
$$= \Delta H_M - T_M \Delta S_M \qquad (2)$$

Since $T_M = \Delta H_M/\Delta S_M$, the absolute magnitudes of enthalpic and entropic contributions to stability cannot be determined from T_M alone.

3.1 Intrinsic Factors

3.1.1 (G + C) Content

T_M exhibits a linear dependence on the fraction of (G·C) base pairs, $F_{(G+C)}$. At physiological monovalent

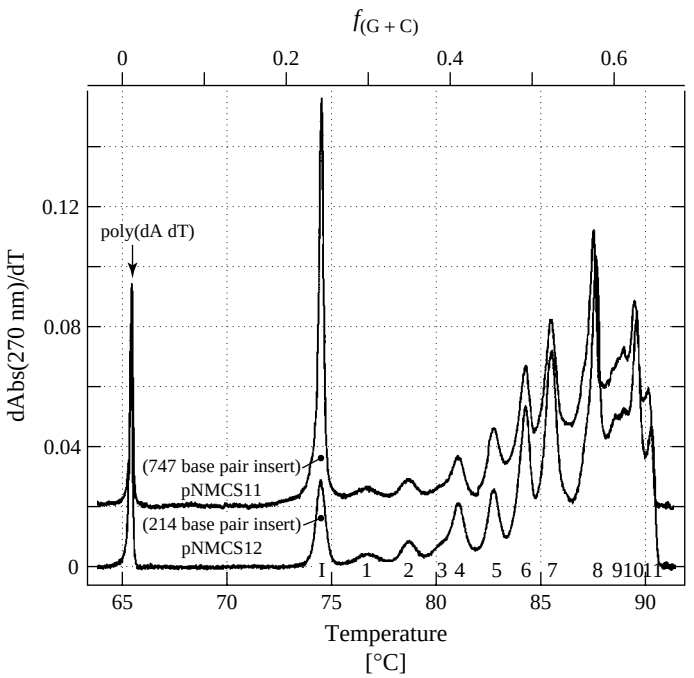

Fig. 2 The normalized first derivative of the change in absorption at 270 nm with temperature, $dA_{270\ nm}/dT$, of linearized plasmid DNAs pNMCS11 (5166 bp) and pNMCS12 (4633 bp), in which a small amount of poly(dA·dT) was added as marker. At this wavelength, the change in absorption for dissociation of (A·T) base pairs is equal to the change for (G·C) pairs, so that denaturation curves reflect changes in numbers of base pairs with temperature. With exception for the lengths of the inserts that produce the prominent subtransitions at 74.46 °C, pNMCS11 and pNMCS12 are circular plasmids with identical sequences, grown on *Escherichia coli*, isolated and linearized by enzymatic bilateral scission at the same site immediately adjacent to the insert. The insert consists of the repetitive sequence, $[AAGTTGAACAAAT]_N AAGTTGAA$, [23% (G + C)], where $N = 57$ (747 bp) in pNMCS11 and 16 (214 bp) in pNMCS12. The insert divides two very (G + C)-rich regions that serve as strong helical boundaries for the insert. Since the cut was made next to the insert, it melts according to the case I mechanism in Table 2, and the t_m of the subtransitions for the insert in the two plasmids are identical. The $[Na^+] = 0.075$ M, and the rate of heating was 6.00 °C h^{-1}.

cation concentration (0.15 M-M^+), the dependence, derived from (Eq. 12) below, is given by,

$$T_M = 40.0 \cdot F_{(G+C)} + 70.04\,^\circ C \qquad (3)$$

This expression is applicable to individual DNAs or to segments of a single DNA.

As shown in Fig. 2, melting of natural DNAs occurs with denaturation of whole stretches or *domains* of base pairs. The defining characteristic of a domain is that it dissociates all at once, implying the existence of thermodynamic boundaries. In contrast with the sharp melting of DNAs of low sequence complexity, natural DNAs of quasi-random sequence melt over a range of 15 to 18 °C and exhibit the intermediate *subtransitions* summarized in Table 2. The melting curves in Fig. 2, where a small amount of poly(dA·dT) was added as marker, are for two plasmid DNAs of identical sequence over most of their 4.6 kbp length. As can be seen by the number of partially resolved subtransitions, melting takes place in intermediate steps between 74 to 91 °C, corresponding to the denaturation of sequence domains of 5 to 500 base pairs of 23 to 70% (G + C), approximated by the upper horizontal scale (Eq. 12). Domain boundaries are associated with changes in the stability and fractional (G + C) compositions since (G·C) pairs are ~20% stronger than (A·T) pairs. Sizes of domains are as small as a few pairs and sometimes larger than 500, depending on the base composition and level of *sequence complexity* of the domain and its immediate neighborhood. Complete denaturation involves the dissociation of many domains in an independent multistate fashion. Each plasmid DNA used in Fig. 2 was constructed with an insert of the same (A + T)-rich oligomeric repeat of 13 base pairs, $[AAGTTGAACAAAT]_N AAGTTG]$,

Tab. 2 Intermediate Subtransitions of DNA Helix–Coil Transitions.

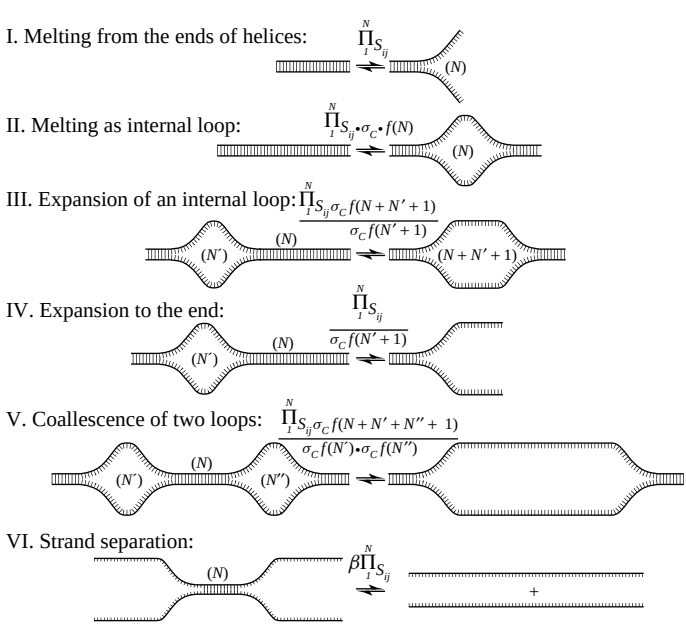

but different numbers of repeats, that is, different values of N. Although the inserts in these two plasmids differ by more than 500 base pairs in length, the domain *subtransition* melting temperature t_m are identical, as is their transition breadths. There is reasonably good agreement between the *subtransition* fractional (G + C) contents $f_{(G+C)}$ of all domains determined by spectral decomposition, sequence analysis, and expression (3). The agreement is not quantitative, however, due to nearest–base pair–neighbor and neighboring-conformational effects (Table 2).

3.1.2 Distribution of Nearest Neighbors

Expression (12) is applicable to the melting of long, natural DNAs with sequences that can be expected to exhibit a random distribution of the frequencies of all 10 unique nearest-neighbor base pairs. There are $4^2 = 16$ nearest-stacked neighbors, but 10 are unique, since complementarity ensures that

$$[A-A] = [T-T], [A-G] = [C-T],$$

$$[A-C] = [G-T], [C-A] = [T-G],$$

$$[G-A] = [T-C], [G-G] = [C-C].$$

DNAs of biased frequencies of neighbor pairs exhibit corresponding biases in T_M. As summarized in Table 3, each neighbor pair has a different stacking energy.

Neighbor-pair biases can be readily detected in the T_M of synthetic DNAs of biased neighbor frequencies. Although both have the same $F_{(G+C)}$, the T_M for alternating co-poly[d(A – T)·d(A – T)] is almost 6° lower than for the homo-poly(dA·dT) duplex. Also, poly[d(G – A)·d(T – C)] and poly[d(G – C)·d(G – C)] melt lower by several degrees and poly[d(G – T)·d(A – C)] melt higher than predicted by the empirical Eq. (12). In instances in which the neighbor frequencies of a DNA are biased, perhaps because the specimen is short, it should be recognized that T_M depend not as directly on $F_{(G+C)}$ than they do on the mole fraction of the 10 nearest neighbors, $f_{ij}/\Sigma f_{i,j}$, where ij denotes the stack of pair i on its neighbor j, and i, j denotes the numbers of the 10 neighbor pairs in the sequence.

Tab. 3 Nearest stacked and paired neighbor energies.

	Neighbor	$dT_{ij}/d\log[Na^+]$ (°C)	(1.0 M Na^+) K	ΔH_{ij} kcal(mol-bp)$^{-1}$	ΔS_{ij} cal(° mol-bp)$^{-1}$
1	A·T/T·A	21.00	355.01	−8.00	−24.15
2	T·A/A·T	20.11	359.88	−8.31	−24.64
3	A·T/A·T	19.78	362.24	−8.45	−24.86
4	G·C/A·T	17.76	372.65	−9.13	−25.89
5	C·G/A·T	17.10	376.34	−9.36	−26.21
6	A·T/G·C	16.87	377.59	−9.44	−26.33
7	A·T/C·G	16.21	381.12	−9.67	−26.65
8	G·C/G·C	14.18	391.65	−10.34	−27.52
9	C·G/G·C	13.20	397.70	−10.72	−28.00
10	G·C/C·G	13.20	397.77	−10.72	−27.99

Values of $T_{ij}(Na^+)$ at different conditions of $[Na^+]$ are determined from $T_{ij}(Na^+) = (dT_{ij}/d\log_{10}[Na^+]) \cdot \log_{10}[Na^+] + T_{ij}(1.0M - Na^+)$, K.

3.1.3 Sequence Complexity

Equivalent but different operational definitions of sequence complexity include the total length of the nonrepetitive sequence, or the number of ways in which oligonucleotide sequence elements can be arranged, or the information content of the sequence. Sequences of lowest complexity, such as poly(dA·dT), denature over the narrowest range of temperatures (<0.1 °C), as can be seen in Figs. 1 and 2. The complexity of the repetitive insert in the plasmid of Fig. 2 is slightly greater and accordingly denatures over a slightly greater temperature range (<0.8 °C), while the plasmid DNA melts over 15 °C.

3.1.4 Conformational States of Neighboring Domains

Melting temperatures of domain subtransitions, t_m, depend on the conformational states of neighboring domains. Despite a difference of more than 500 base pairs in size, the t_m of the (A + T)-rich repetitive insert domains of the two plasmids in Fig. 2 are identical (74.5 °C) because they have the same repeat sequence and occur at the ends of the helix (case I in Table 2) where they denature independently of length and of the penultimate (G + C)-rich domain. Both *circular* plasmids were cut and linearized at a site adjacent to the inserts, positioning the insert domain at the end of the helix. However, if the plasmids are cut and linearized at a remote site that positions the insert domains internally, they melt as loops with a higher t_m (case II in Table 2). The t_m of domains that melt as internal loops depend on the size of the loop and are greater than the t_m for the same domains when melting from the end of the helix. As an internal domain, the larger insert of the plasmid pNMCS11 melts at 75.0 °C, while the smaller insert melts at 76.3 °C. A higher t_m and sharper melting for loops is due to their greater probability of renaturing and their lower conformational entropy, since the two ends of the domain are constrained by the two helical domains that border it. Since larger loops have a lower probability of renaturing than smaller loops, the Δt_m has a reciprocal dependence on loop size: smaller for large loops than for small loops, all other energetic factors being equal.

Figure 3 shows matching helix-end (case I) and helix-internal (case II) domain subtransitions for a different 200 bp (A + T)-rich repetitive domain that was installed in the same location of the same plasmid as that of Fig. 2, but now designated pNMCS22. This insert denatures at 73.08 °C when the plasmid is linearized adjacent to the domain but at 75.31 °C, more than 2 °C higher and more sharply when linearized at a remote site, forcing the domain to denature as an internal loop.

3.1.5 DNA Length and Concentration

The T_M of DNAs greater than ~3000 bp vary independently of length or concentration since the strands stay together until the last, most stable domain is denatured, allowing the two strands to diffuse away from each other, leading to a nonequilibrium situation. (This is the reason, incidentally, why the melting curve of poly(dA·dT) in Fig. 1 is slightly asymmetric, and why the final subtransitions of the melting of the plasmids in Fig. 2 are so abrupt). Short DNAs show a dependence on length and concentration because of their greater translational entropy. The effects of length on helix stability for a series of oligomers of identical neighbor-pair frequency and concentration follow a relationship of the form

$$1/T_M(N) = A + B/(N-1) \qquad (4)$$

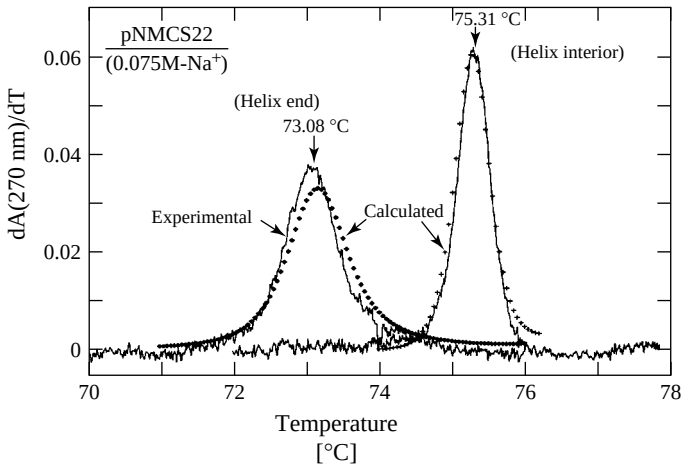

Fig. 3 Matching subtransitions for the insert sequence $[AAGTTGAACAAAAAAT]_{12}AAGTTGAA$ in pNMCS22, when melting according to the case I mechanism from helix ends ($t_m = 73.08\,^{\circ}C$) and case II mechanism from the helix interior ($t_m = 75.31\,^{\circ}C$). Except for the insert, the sequence of pNMCS22 is identical to those of pNMCS11 and pNMCS12 in Fig. 2. The continuous solid line represents experimental curves of $dA_{270\ nm}/dT$, while the curves of symbols represent the statistical–mechanical curves, $d\theta_{bp}/dT$, calculated with thermodynamic values in Table 3 and in the text. The pNMCS22 plasmid was linearized adjacent to the insert sequence for both experimental and calculated curves exhibiting the case I mechanism, while for the case II mechanism, the plasmid was linearized 800 bp away, leaving the insert in the interior of the helix with (G + C)-rich boundaries.

for oligonucleotides of length N, where A and B are constants that depend on the thermodynamic properties of the infinite polymer and on conditions. The denominator $(N-1)$ corresponds to the number of stacking interactions in an oligomer duplex of length N.

The stability of short helices also depends on oligonucleotide strand concentration, c, according to the relationship

$$1/T_M(c) = a - b \ln c \qquad (5)$$

where a and b are constants. Integration between concentrations $c_{N,1}$ and $c_{N,2}$ for oligomers of the same length N leads to

$$\frac{1}{T_{m,N}(c_{N,1})} = \frac{1}{T_{m,N}(c_{N,2})} + \frac{R}{(N-1)\cdot \Delta H_m} \cdot \ln\left[\frac{c_{N,1}}{c_{N,2}}\right] \qquad (6)$$

thus avoiding any direct consideration of the properties of the infinite polymer, whose stability is both length and concentration independent.

3.2 Extrinsic Factors

3.2.1 The Dependence of Melting Temperature on Counterion Type and Concentration

T_M exhibits a linear dependence on the logarithm of the cation concentration. Owing to the high density of negative charges, cations are attracted to DNA in large

numbers where they screen and neutralize the negative charges on phosphate oxygens O1P and O2P, reducing the effective charge on the phosphates from $-1e^-$ to about $-0.24e^-$. If the cation is a monovalent alkali metal ion M^+, where M^+ is Li^+, Na^+, $K^+, \ldots$, theory indicates that the local cation concentration close to the DNA exceeds 1 M, regardless of the bulk concentration.

The difference between the fraction of counterions condensed to native and denatured DNA, $0.76 - 0.39 = 0.37$, together with the difference in Debye–Hückel screening brought about by these counterions leads to the release of large numbers of condensed cations from the helix as DNA is converted from the high–charge density helical state to the much lower–charge density coil state, representing the physicochemical basis for the linear dependence of melting temperatures on $\log[M^+]$ as well as the strong power dependence on salt concentration of second-order *renaturation* rates, k_2 [L (mol·s)$^{-1}$]:

$$(\partial \log k_2 / \partial \log[Na^+]) = 3.6 \qquad (7)$$

The loss of M^+ from the helix is represented by an apparent equilibrium constant, K_M^{app}, for the reaction scheme

$$\mathrm{DNA^h} - r^{\mathrm{h}}_{\mathrm{M^+}} \cdot \mathrm{M^+} \rightleftharpoons \mathrm{DNA^c} - r^{\mathrm{c}}_{\mathrm{M^+}} \cdot \mathrm{M^+} + \Delta r_{\mathrm{M^=}} \cdot \mathrm{M^+} \qquad (8)$$

where $\Delta r_{M^+} = (r^{h}_{M^+} - r^{c}_{M^+})$ is the number of monovalent counterions released from association with the average-size cooperatively melting domain. An increase of M^+ therefore favors the helical state while a decrease shifts the equilibrium away from the *helical* state and toward the denatured *coil* state. The amount released is relatively constant over the range $0.001 < [M^+] < 0.3$ M/L.

A limiting law has been constructed by Manning in the formalism of the *counterion condensation theory* that serves the same purpose for dilute DNA solutions as does the Debye–Hückel theory for simple electrolyte solutions. This theory, successful in explaining many of the physical properties of polyelectrolytes, is an approximation of the classical Poisson–Boltzmann theory traditionally used in numerical computations of ion–ion distribution functions and thermodynamic properties of electrolyte solutions. In the counterion condensation theory, DNA is modeled as a rigid rod with a uniform and continuous line charge, which leads to condensation by large numbers of counterions. Counterions also reduce electrostatic interactions between the fixed charges through Debye–Hückel screening, further reducing the range that would be predicted from Coloumb's law.

The key parameter leading to the spontaneous neutralization of unit charges, e, resident on the polyelectrolyte is the *linear charge density*, ξ,

$$\xi = \left(\frac{e^2}{4\pi\varepsilon_o\varepsilon_r k_B T}\right) \cdot \frac{1}{b} \qquad (9)$$

where ε_o is the permittivity constant of free space (8.8542×10^{-12} C^2 Nm^{-2}), ε_r is the dielectric constant of the aqueous solution (78.3 at 25 °C), k_B is the Boltzmann's constant (1.380662×10^{23} J K^{-1}), and b is the length of the charge separation along the mean contour axis of the polyelectrolyte. The quantity $e^2/(4\pi\varepsilon_o\varepsilon_r k_B T)$ in this expression has dimensions of length (m), the so-called *Bjerrum characteristic length*, with a value of 7.13 Å for aqueous solutions at 25 °C. Since ξ represents the ratio of the characteristic length and b, the charge density parameter is dimensionless. For B-form DNA, the mean spacing

of charges, b^{h}, is 3.34 Å/2 = 1.7 Å. Consequently, the value of ξ^{h} for native DNA is 7.14 Å/1.7 Å = 4.2 and ~1.7 for denatured DNA.

The magnitude of the charge density parameter ξ determines the extent of counterion condensation. The condensed phase of associated counterions is the consequence of thermodynamic instability when ξ exceeds $|e|^{-1}$, which is resolved by local association of counterions. If $\xi < 1$, no condensation occurs. In the case of DNA, $\xi^{h} = 4.2$, therefore counterion condensation occurs, which effectively neutralizes the fixed charges from a charge density of ξ to $|e|^{-1}$. $(\xi^{h})^{-1}$ represents the *saturated charge fraction*, which for the helix is 1/4.2 = 0.24 (0.12 for divalent counterions). The fraction of counterions that are condensed per charge is therefore given by

$$\theta^{h}_{M^+} = [1 - (\xi^{h})^{-1}] = 0.76 \qquad (10)$$

According to theory, a fraction, 0.76, of the phosphate groups of B-DNA are neutralized by a dense cloud of univalent counterions, corresponding to a local concentration of ~1.2 M-M^+. The condensed counterion phase fades rapidly with increasing radial distance from the helix. The corresponding fraction for the denatured single-strand coil state is 0.39, so the immediate concentration of counterions is only ~0.21 M surrounding denatured DNA.

Predictions of the counterion condensation theory have been confirmed in virtually every respect by experiment, including the prediction of a semilogarithmic dependence of stability on $[M^+]$. The variation of the apparent equilibrium constant, K_M^{app}, with melting temperature for the reaction scheme of (8) is given by the Manning-Record equation, in which electrostatic factors affecting denaturation are expressed explicitly

$$\frac{dT_M}{d\ln a^+} \approx \frac{d(T_M^{-1})}{d\ln a^+} = \frac{2RT_M^2}{\Delta H_M^o}(\Delta\theta'_{M^+}) \qquad (11)$$

where ΔH_M^o is the mean enthalpy, a^+ the *activity* (or molarity if concentrations are low) of monovalent cation, and $\Delta\theta'_{M^+}$ the fraction of condensed counterion *per mole of nucleotide residue*. Evaluation of $\Delta\theta'_{M^+}$ or ΔH_M^o is therefore possible from slopes taken from the dependence of T_M on $\ln[M^+]$. The observed salt effect, $dT_M/d\ln[M^+]$, is almost entirely entropic in origin as helices jettison large numbers of condensed counterions upon melting.

The dependence of T_M on $[Na^+]$ has been combined with the dependence of T_M on $F_{(G+C)}$ (3) into a single empirical relationship

$$T_M = 193.08 - (3.09 - F_{(G+C)}) \times (34.47 - 2.83\ln[Na^+])\ ^{\circ}C. \qquad (12)$$

which, for the same reasons explained above for expression (3), is applicable only to large DNAs of quasi-random distributions of neighbor-pair frequencies.

The dependence of T_M on monovalent counterion type follows the order $Rb^+ \approx Cs^+ > Li^+ \approx K^+ > Na^+$.

3.2.2 pH

pK_a of the four bases indicate that the range of DNA stability is limited to pHs between ~4 and 9.4. On the basis of their pK_a, in Table 4 titration curves for guanine, cytosine, and thymine frame the pH region between 4 and 9.4 as the region most hospitable to the native structure. There are practical reasons for avoiding pHs below ~5. Protonation leads to denaturation and precipitation, while still lower pHs greatly increase the rates of depurination and chain

Tab. 4 pK_a of nucleosides and nucleotides.

Nucleoside	*Site*	pK_a	pK_a *of the nucleotide 3′-phosphate*	pK_a *of the nucleotide 5′-phosphate*
Adenosine	N1	3.52	3.70	3.88
Cytidine	N3	4.17	4.43	4.54
Guanosine	N1	9.42	9.84	10.00
Uridine	N3	9.38	9.96	10.06
2′-Deoxythymidine	N3	9.93	–	10.47

cleavage. Denaturation at high pH is less problematic and represents a good method for separating the two strands from one another. Provided the strands differ in purine/pyrimidine ratio from unity, they can be separated by CsCl buoyant density gradient centrifugation at pH $\sim$ 11, where thymine and guanine residues are titrated.

3.3 Denaturing Solvents

A large number of different salts and miscible organic solvents have been investigated for their effects on the denaturation of DNA. The results are summarized in Table 5. These agents can affect the stability of DNA in two ways, by destabilizing the helix or by interacting favorably with the coil. Evidence indicates that many organic solvents interact with the bases, favoring the denatured coil state in the *helix–coil* equilibrium. Denaturing effectiveness is related to the size of the hydrophobic alkyl substituents on alcohols (*n*-butyl > *n*-propyl > ethyl > methyl), carbonates, amides, and ureas, while the addition of hydrophilic groups reduces the denaturing effectiveness of a solvent (ethyl alcohol > ethylene glycol; *n*-propyl alcohol > glycerol; cyclohexyl alcohol > inositol).

Formamide, a popular and well characterized denaturant, lowers melting temperatures by 2.4 to 2.9 °C per mole of formamide (C_F) depending on the $F_{(G+C)}$,

$$dT_M/dC_F = 0.453F_{(G+C)} - 2.88 \quad (13)$$

and the state of hydration. The inherent cooperativity of melting is unaffected by the denaturant (Fig. 4). Although dT_M/dC_F fits to a linear dependence on (G + C) content, the denaturing effect is consistent with a (G + C)-independent alteration in the apparent equilibrium constant for the *helix–coil* transition. Formamide appears to have a destabilizing effect on the helical state, where sequence-dependent variations in local hydration patterns lead to small variations in sensitivity to the denaturant. The effect is attributable to the displacement by formamide of weakly bound water in the vicinity of the helix, while the phenomenological effects of displacement are equivalent to lowering the bulk counterion concentration.

The integrity of the double helix is also sensitive to the nature of the anion of certain concentrated salts, and to a lesser extent, the cation of the salt. Salts that are most effective in shifting the *helix–coil* equilibrium are also most effective in solubilizing the bases. The following selected anions are listed in order of decreasing effectiveness in destabilizing the double

Tab. 5 Organic reagents and their concentrations (*M*) that give a 50%-level of denaturation of T_4 DNA ($F_{(G+C)} = 0.34$) at 73 °C and an ionic strength of 0.043 M.

Aliphatic alcohols:							
Methyl	3.5	*n*-propyl	0.54	*tert*-butyl	0.60	*tert*-amyl	0.39
Ethyl	1.2	Allyl	0.50	Isobutyl	0.45	Ethylene glycol	2.2
Isopropyl	0.90	*sec*-butyl	0.62	*n*-butyl	0.33	Glycerol	1.8
Thio alcohols:							
Dithioglycol	2.2						
Cyclic alcohols:							
Cyclohexyl	0.22	Inositol	1.5	*p*-methoxyphenol	0.09		
Benzyl	0.09			Phenol	0.08		
Other cyclic compounds:							
Aniline	0.08	Purine	0.13	γ-butyrolactone	0.55		
Pyridine	0.09	1,4-dioxane	0.64	3-aminotriazole	0.42		
Amides:							
Formamide	1.9	Acetamide	1.1	Propionamide	0.62	Glycolamide	1.1
N-ethylformamide	1.0	*N*-ethylacetamide	0.88	Butyramide	0.46	Thioacetamide	0.32
N,*N*-dimethylformamide	0.60	*N*,*N*-dimethylacetamide	0.60	Hexanamide	0.17	δ-valerolactam	0.34
Ureas:							
Urea	1.0	Ethyleneurea	0.53	Thiourea	0.41	Carbohydrazide	1.0
Ethylurea	0.60	1,3-dimethylurea	1.0	Allylthiourea	0.28		
tert-butylurea	0.22			Ethylenethiourea	0.32		
Carbamates:							
Urethane	0.50	*N*-methylurethane	0.38	*N*-propylurethane	0.24		
Other compounds:							
Cyanoguanidine	0.21	Glycine	2.2	Sulfamide	1.1	Acetonitrile	1.2
Tween 40	>20%	Triton X-100	>10%				

Source: Levine et al. (1963) *Biochemistry*, **2**, 168.

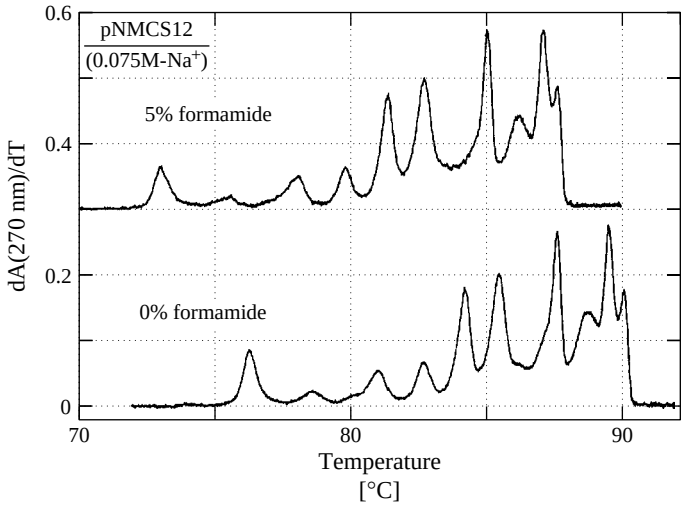

Fig. 4 Melting curves of pNMCS12 DNA (4,633 bp), in 0 and 5% of the denaturant formamide.

helix (or increasing effectiveness in supporting the helix): $Cl_3CCOO^- > SCN^- > I^- > CH_3COO^- > Br^- > Cl^-$, while the order for cations is $Li^+ > Na^+ > K^+ \gg Mg^{2+}$.

4 Utility of Denaturation Studies

4.1 Thermodynamic Analyses

Knowledge of the forces that govern the conformation and stability of DNA is needed to understand and model the behavior of this molecule. The equilibrium concentrations of individual base pairs in all possible conformational substates must be known in order to develop plausible models for the mechanics and dynamic behavior of DNA during replication, repair, and transcription. Studies of the thermal denaturation of DNA are also important adjuncts to structural studies, providing the means for identifying various molecular sources of stability in the helix, and for quantitating the thermodynamic characteristics of those sources.

4.1.1 Simulation of Melting

Denaturation profiles of specific sequences can be simulated through the application of statistical–mechanical methods. Given suitable thermodynamic values for the various parameters, it is possible to simulate melting and predict the macroscopic properties of DNA from its sequence alone by averaging over all possible solutions for the mechanical behavior of the molecule at the molecular level. A considerable amount of both qualitative and quantitative agreement with experiment has been achieved with calculated melting curves.

During melting, six intermediate nonhelical states can be identified, represented by cases I to VI in Table 2, and distinguishable by the physical states of neighboring

domains. The simplest case is represented by melting from the ends of the helix (case I). s_{ij} is assigned as the stability constant for stacking pair i onto j at the end of a helical domain, and, in the absence of any extraordinary long-range energetic factors, by s_{ij}^N for a domain of N pairs. s_{ij} is related to temperature by

$$s_{ij} = \exp[-\Delta G_{ij}/RT] = \exp[-(\Delta H_{ij} - T \cdot \Delta S_{ij}/RT)] \quad (14)$$

where values for ΔH_{ij}^* and ΔS_{ij}^* are listed in Table 3.

In long DNAs, denaturation is primarily internal with formation of closed loops, represented by case II in Table 2. At least three factors affect the thermodynamics and temperatures at which internal domains dissociate: (1) the previously mentioned costs of unstacking each base pair in the domain; (2) the energetic costs of interrupting the helix; and (3) the energetic level of the loop:

$$K_{eq} = s_{ij}^N \sigma_c f(N) \quad (15)$$

The cooperativity parameter, σ_c, represents the weighting factor for helix interruption:

$$\sigma_c = \exp(-\Delta G_c/RT_M) \quad (16)$$

with a value of $\sim 4 \times 10^{-6}$, so that $\Delta G_c \approx 7$ kcal mol^{-1} interruptions. The statistical weight for imposing rotational and translational constraints on N denatured residues of a closed loop is a significant item in the equilibrium constant:

$$f(N) = (N+1)^{-1.7\pm0.2} \quad (17)$$

With these values for the various parameters, it is possible to quantitatively account for three of the six intermediate equilibria in Table 2. Cases IV to VI frequently show indications of nonequilibrium behavior owing to the abrupt increase in translational freedom, similar to that from strand dissociation. The melting behavior of domains isolated between two denatured loop regions is sometimes in nonequilibrium and is abrupt. The parameter governing strand association is given by

$$\beta = K \cdot N_\alpha \quad (18)$$

where N_α is the length of the terminal domain with α translational degrees of freedom, and K includes all factors independent of N_α, and

$$\beta = \theta_{\text{int}} - 3 \quad (19)$$

where θ_{int} represents the average fraction of intact base pairs for DNAs with at least one intact pair. β has a value of $\sim$0.0015 L mol^{-1}.

A popular statistical thermodynamical model for denaturation is the canonical one-dimensional lattice with loop entropy, in which linked nucleotide residues exist in either the paired or unpaired state for each configuration of states. The free energy for the kth configuration is given by

$$\Delta G_k = -RT \ln Z_k \quad (20)$$

while the probability that the chain will assume the kth configuration is Z_k/Z. Z is the partition function over all configurations. The fraction of residues in the paired state is proportional to the statistical weights,

$$\theta_{\text{bp}} = \Sigma(N_k Z_k/Z)/N_{\text{len}} \quad (21)$$

where N_k is the number of paired residues in the kth configuration. Z_k is the product of the weighting factor in (17).

A considerable effort has been made in the evaluation of the numerous parameters in the statistical–mechanical model, and

quantitative agreement with experiment can be achieved for the melting of most domains. Calculated melting curves can be obtained with the program MELTSIM that reads in the DNA sequence and, using Poland's recursion formulae, calculates conditional probabilities at each temperature that the $(m-1)$th residue is in the paired state, given that the mth residue is paired. Following this, the unconditional probabilities that the mth residue is in the paired state are calculated. The unconditional probabilities yield $\theta_{bp}(T)$, and, thereby, $d\theta_{bp}(T)/dT$, which is proportional to normalized values of experimental $dA_{270\ nm}/dT$. A comparison of the experimental (line) and calculated (points) subtransitions for the 200 bp repetitive insert in pN/MCS22, $[AAGTTGAACAAAAAAT]_{12}AAGTTGAA$, is illustrated in Fig. 3.

4.1.2 Denaturation Maps

Denaturation maps correlate sequence position with conformational state, helix, or coil, at all temperatures, and are produced at the same time that melting curves are calculated. If the maps are accurate, they can be useful for determining the denaturation of particular base sequences and identifying specific functional regions from the sequence database. Thus, coding and noncoding sequences are readily identifiable from denaturation maps. Coding regions exhibit what has been described as a *homostabilizing propensity* with unusually large melting domains. Noncoding regions exhibit greater variability of base composition and smaller domains. Denaturation maps have also been used to demonstrate that melting temperatures and mean (G + C) base compositions of coding regions are significantly higher than those of flanking, intron, and intergenic regions, making possible the delineation of coding regions in uncharacterized DNA sequences.

4.1.3 Conditions for Equilibrium Denaturation

If quantitative significance is required of the T_M or transition breadth, experimental melting curves must be reproducible and represent true helix–coil equilibria at all temperatures. Usually, this means that curves are carried out slowly, with suitable delays at each temperature, to ensure that the property being monitored such as absorption at 260 nm is unchanging. Nonequilibrium changes are more prone to occur with decreasing ionic strengths, because the reformation of the helix during the helix–coil transition has a strong power dependence on cation concentration. As a rule of thumb, when absorbance values are monitored in continuous fashion as temperature is steadily increased, the heating rate should not exceed 10 °C per hour when the counterion concentration is in the neighborhood of 0.075 M Na^+. Rates can be faster at higher $[Na^+]$, but must be slower at lower $[Na^+]$. Below 0.02 M Na^+, the rate must be so slow that it becomes difficult to obtain equilibrium curves.

4.2 Hybridization

Denaturation is necessary for producing hybrids of complementary DNA strands from different source material, forming heteroduplexes for Southern blots or for the detection of evolutionary or mutational differences, and for joining complementary single-stranded ends for cloning DNA fragments.

4.2.1 Stringency

Studies of denaturation are required to define the level of stringency for a particular hybridization scheme. Stringency is a qualitative attribute that attempts to define the optimal conditions for hybridization between DNA strands from different sources. It refers to a set of conditions that favors the kinetic discrimination between two or more competing DNA strands during hybridization. As the temperature or solvent conditions for renaturation (hybridization) approach those for denaturation, hybrids of less perfect alignment are disfavored because of their lower stability. It is usually assumed that stringency increases as the incubation temperature T approaches the melting temperature T_M, $(T_M - T) \rightarrow 0$, so that chains of lesser sequence similarity are prevented from kinetically competing with more perfect regions of sequence and slowing up the hybridization process. As has been shown, overall, T_M are not simple measures of the stabilities of specific DNA sequences, particularly if the DNA is short. Melting curves such as those in Fig. 2 consist of innumerable intermediate states where a single T_M has no obvious significance for the hybridization of a specific domain sequence. Discrimination between two or more competing strands during renaturation is probably better achieved at room temperature at high ionic strength, which favors ready strand displacement of strands of lower similarity and stability, while at the same time avoids the hazards of thermal degradation.

4.2.2 The Potential for Degradation

Depurination of denatured strands is of significant concern when DNA is denatured by temperature. Denaturation should be conducted as rapidly as physically possible, preferably by microwave heating, in order to minimize the exposure of the single strands to the degradative conditions of high temperatures because the rate of depurination increases substantially above 50 °C and is often accompanied by spontaneous strand cleavage. Many procedures for denaturing DNA call for heating in boiling water for 2 to >10 minutes; however, the incubation time need only be long enough to ensure that thermal equilibrium has been achieved since denaturation at the high temperatures of ~100 °C is almost instantaneous.

4.3 PCR

Repeating cycles of denaturation and synthesis with the thermally stable *Taq* polymerase from *Thermus aquaticus* is the basis for the amplification of minute quantities of DNA by the PCR, a simple alternative or complement to cloning.

4.4 Sequence Variances

Measurements of levels of sequence variation within satellite elements and multicopy genes are determined from δT_M for the difference in thermal stability of homo- and heteroduplexes. Mispairs in heteroduplexes contribute to a lower overall stability, where the magnitude of δT_M is proportional to the fraction of *mispairs*. The current semiquantitative relationship finds that a 1.7% divergence in sequence yields a δT_M of 1.0 °C, however, curves of short heteroduplexes may exhibit a different dependence of δT_M for specific differences in mispaired neighbors.

4.4.1 Identification of Point Mutations and Mispair Thermodynamics

Denaturing gradient gel electrophoresis (DGGE), temperature gradient gel

electrophoresis (TGGE), and denaturing high-performance liquid chromatography (DHPLC) are sensitive methods for detecting small variations in local stability of the duplex caused by mispairs. These techniques have the capacity to detect differences in methylation of C residues in otherwise identical DNA specimens. The occurrence of point mutations implies formation of mispairs as intermediates during replication and repair of genomic DNAs. In some cases, the structures of these non-Watson–Crick mispairs require little distortion of the duplex, so their occurrence may go undetected, leading to spontaneous substitutions. TGGE has been used effectively to measure the effect of local DNA structure on mispair stacking. Recent studies have focused on the stability and structure of non-Watson–Crick mismatches, demonstrating, for example, that mispairs with 5′(C·G) and 3′(A·T) neighbors occur in the order $G{\cdot}T > G{\cdot}A > G{\cdot}G > A{\cdot}G \geq T{\cdot}G > A{\cdot}A = T{\cdot}T > A{\cdot}C > C{\cdot}A, T{\cdot}C, C{\cdot}T > C{\cdot}C$. The order for other pairs and mispairs is given in Table 6.

4.5 Profiles of Genomic Base Distributions

Determinations of base distributions in total nuclear DNA are useful for what they reveal about the overall content and pattern of different sequence elements in the genome as well as what they reveal of quantitative differences between species. Melting curves of genomic DNAs yield detailed profiles of base distributions, since they reflect the mean (G + C) contents of segments only 350 ± 150 bp in length. Bacterial DNAs are ~1000 times longer than the plasmid DNAs that produced the complex melting curves in Fig. 2, and therefore they exhibit a virtual continuum of subtransitions and distributions of domain (G + C) contents. Observed melting curves of bacterial DNAs are almost Gaussian in shape with variances of only 8% (G + C), reflecting the highly streamlined genomes expected of rapidly dividing single-celled organisms.

Sequence complexities of warm-blooded vertebrates are much larger and far more cluttered and less streamlined than those of bacteria, and, accordingly, exhibit broader bell-shaped denaturation curves with sharp bands. These bands represent different amounts and types of low-complexity satellite sequence arrays that vary in (G + C) contents among even closely related species. The curves of three *artiodactyla* (class *Mammalia*), two cervids, and a bovid in Fig. 5 are typical. The profiles, adjusted in integrated areas to cellular DNA contents, are superimposable at low temperatures and fractional base contents below 40% (G + C), but exhibit a number of distinctive species-specific sharp bands at higher temperatures and (G + C) contents. Each contains at least three major satellites that have been designated I, II, and III. The 8.85 pg of nuclear DNA per diploid cell of *Alces alces* is more than 25% higher than that in other cervids due to the presence of more than three million copies of a 1-kbp satellite I, seen as a prominent band of 50% (G + C). Sequence variations are only 1 to 2% owing to a mechanism of concerted evolution involving nonreciprocal recombination.

5 Denaturation of RNA

Although the *covalent* structures of the polydeoxyribonucleate chains of DNA and polyribonucleate chains of RNA are almost

Tab. 6 Relative stabilities of Watson–Crick pairs and mispairs.

Nearest neighbor pairs		***Relative stabilities of these intervening pairs and mispairs***
5′	***3′***	
(G·C)	(T·A)	C·G > G·C > T·A > A·T > G·G > A·G > G·A > G·T > T·G > A·A > C·A > T·T ≥ A·C > T·C ≥ C·T > C·C
(G·C)	(G·C)	C·G > G·C > T·A > A·T > G·A > G·G > T·G > A·G > G·T > A·A = T·T > C·A > C·T > A·C > T·C > C·C
(C·G)	(A·T)	G·C > C·G > A·T > T·A > G·T > G·A > G·G > A·G ≥ T·G > A·A = T·T > A·C > C·A, T·C, C·T > C·C
(T·A)	(T·A)	G·C > C·G > T·A > A·T > G·T > G·G, A·G > G·A, T·G > A·A > T·T > A·C > T·C > C·C, C·A > C·T

Source: Ke, S-H & Wartell, R.M. (1993) *Nucl. Acids Res.* **21**, 5137–5143.

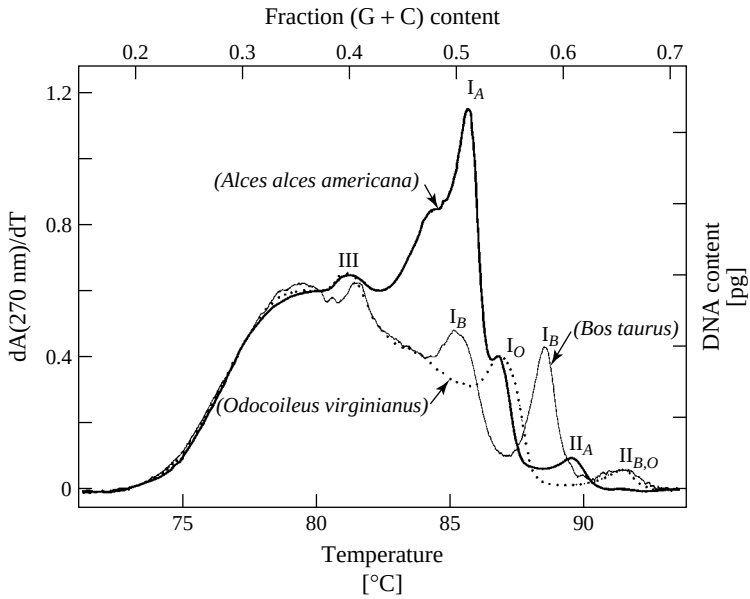

Fig. 5 Denaturation profiles of total nuclear DNA from three *artiodactyla* (class *Mammalia*), two cervids *Odocoileus virginianus* (North American white-tailed deer) and *A. alces americana* (North American moose), and a bovid, *Bos taurus* (domestic ox). The integrated area under each profile has been normalized to cellular DNA contents (pg). All three profiles exhibit bands arising from the presence of at least three major satellites that have been designated I, II, and III. The 8.85 pg of nuclear DNA per diploid cell of *A. alces* is more than 28% higher than a close cervid *O. virginianus* (6.9 pg per cell) and 18% higher than *B. taurus* (7.5 pg per cell), due to the extraordinary presence of more than three million copies of a 1-kbp satellite I, seen as a prominent band of 50% (G + C).

the same, the way these chains are synthesized, their lengths, and the way *secondary* and *tertiary* noncovalent forces act to support *native* structures are very different. In DNA, the two strands are aligned and held together over their entire lengths by Watson–Crick *hydrogen bonds* between complementary bases projecting laterally from each chain and by vertical *stacking forces* between base pairs. The shorter single-stranded chains of RNAs fold back on themselves through secondary and tertiary interactions to form *bulges, hairpin stem loops, pseudoknots, coaxial duplexes,* and *base triples*. Not surprisingly, the denaturation of these two informational biopolymers are quite different. When structures of natural RNAs of quasi-random sequence are heated, stacking and hydrogen-bonding interactions dissociate over a range of more than 70 °C, a five-fold greater range than the melting of duplex DNAs. Synthetic single-stranded homopolymeric RNAs melt over 100 °C (Fig. 6), a greater range than natural RNAs because they are completely devoid of the short hairpin structural elements responsible for cooperative behavior.

Single-strand regions of RNAs, regions in which base residues are partially stacked but unpaired, are flexible, permitting the formation of large

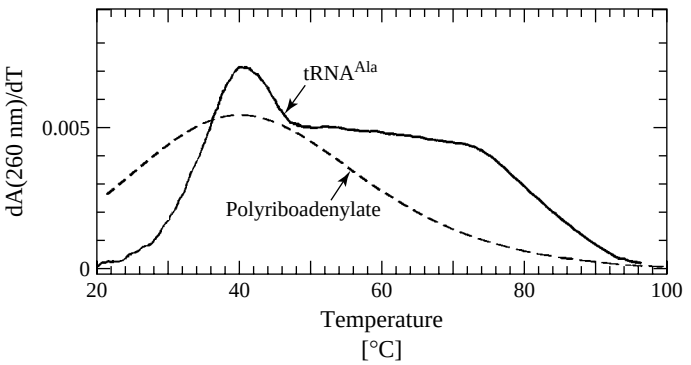

Fig. 6 Derivative melting curves of high- and low-complexity RNAs, obtained at 260 nm in 0.075 M Na^+. The solid line is the melting curve of alanine tRNA, and the broken line is the melting curve of the synthetic homopolymeric RNA polyadenylate. While melting curves of natural RNAs are normally insensitive to monovalent cation concentrations, they exhibit increasing cooperativity with increasing concentrations of some divalent cations.

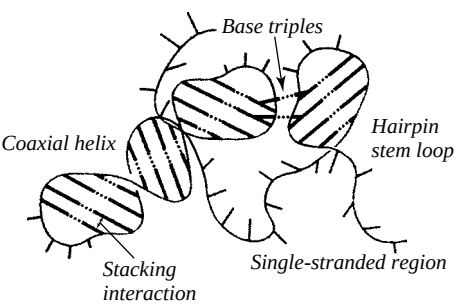

numbers of stem loops, bulges, base triples, and so on, that fold into complex shapes through interactions at the tertiary level.

The makeup of RNAs from small structural elements is apparent from the way they melt, which takes place at lower temperatures and over much wider ranges than DNA. The example of the melting of $tRNA^{Ala}$ is shown in Fig. 6. The range of melting increases with decreasing sequence complexity of the RNA as opportunities decrease for formation of hairpin stem loops. This dependence is just opposite the sharp melting of low-complexity DNA duplexes (Fig. 1).

Sharpness of melting is governed by two factors: (1) the change in enthalpy for the transition and (2) the magnitude of the equilibrium constant for creating an interruption in a linear array of stacked bases (16). The stacking enthalpy is only marginally lesser for RNAs than for DNAs; however, the interruption constant, a measure of the degree of cooperativity, is ~ 0.5, much larger than that for DNAs, $\sim 10^{-6}$. The interruption constant corresponds to the reduction in entropy for restricting rotation about *N*-glycosidic bonds linking a base to a sugar. In single-stranded RNAs, the conformational states of stacked bases are essentially independent of the states of their neighbors, so that any detectable cooperativity during melting must reflect the denaturation of individual hairpin stem loops (Fig. 6). Thus, while the denaturation of single-stranded RNA chains is characterized by strong stacking interactions, the average length of stacked segments, given

by $1 + \sigma^{-1/2}$, is only 2 to 3 at physiological temperatures.

See also Bioorganic Chemistry.

Bibliography

Books and Reviews

Breslauer, K.J., Freire, E., Straume, M. (1992) Calorimetry: a tool for DNA and ligand–DNA studies, *Methods Enzymol.* **211**, 533–567.

Cantor, C.R., Schimmel, P.R. (1980) *Biophysical Chemistry. Part II. Techniques for the Study of Biological Structure and Function*, Freeman, San Francisco, pp. 1139–1165.

Gotoh, O. (1983) Prediction of melting profiles and local helix stability for sequenced DNA, *Adv. Biophys.* **16**, 1–52.

Palecek, E. (1991) Local supercoil-stabilized DNA structures, *Crit. Rev. Biochem. Mol. Biol.* **26**, 151–226.

Piskur, J., Rupprecht, A. (1995) Aggregated DNA in ethanol solution, *FEBS Lett.* **375**, 174–178.

Plum, G.E., Breslauer, K.J., Roberts, R.W. (1999) Thermodynamics and Kinetics of Nucleic acid Association/Dissociation and Folding Processes, Chapter 2, in: Barton, D., Nakenishi, K. (Eds.) Comprehensive Natural Products Chemistry, Vol. 7, Pergamon Press, New York, pp. 15–33.

Poland, D., Scheraga, H.A. (1970) *Theory of Helix–Coil Transitions in Biopolymers*, Academic Press, New York.

Riesner, D., Romer, R. (1973) Thermodynamics and Kinetics of Conformational Transitions in Oligonucleotides and tRNA, in: Duchesne, J. (Ed.) *Physico–Chemical Properties of Nucleic Acids*, Academic Press, New York, pp. 237–318.

Slatko, B.E. (1994) Thermal cycle dideoxy DNA sequencing, *Methods Mol. Biol.* **31**, 35–45.

Turner, D.H. (2000) Conformational Changes, Chapter 8, in: Bloomfield, V.A., Crothers, D.M., Tinoco, I. Jr. (Eds.) *Nucleic Acids*, University Science Books, Sausalito, CA, pp. 259–334.

Wada, A., Suyama, A. (1986) Local stability of DNA and RNA secondary structure and its relation to biological functions, *Prog. Biophys. Mol. Biol.* **47**, 113–157.

Wada, A., Yubuki, S., Husimi, Y. (1980) Fine structure in the thermal denaturation of DNA: high temperature resolution spectrophotometric studies, *Crit. Rev. Biochem. Mol. Biol.* **9**, 87–144.

Wartell, R.M., Benight, A.S. (1985) Thermal denaturation of DNA molecules: a comparison of theory with experiment, *Phys. Rep.* **126**, 67–107.

Primary Literature

Abrams, E.S., Stanton, V.P. Jr. (1992) Use of denaturing gradient gel electrophoresis to study conformational transitions in nucleic acids, *Methods Enzymol.* **212**, 71–104.

Applequist, J. (1966) True phase transitions in macromolecules of the DNA type, *J. Chem. Phys.* **45**, 3459.

Applequist, J., Damle, V. (1963) Theory of the effects of concentration and chain length on helix–coil equilibria in two–stranded nucleic acids, *J. Chem. Phys.* **39**, 2719–2721.

Azbel, M.Y. (1980) DNA sequencing and helix–coil transition. II. Loop entropy and DNA melting, *Biopolymers* **19**, 81–93.

Blake, R.D., Bizzaro, J.W., Blake, J.D., Day, G.R., Delcourt, S.G., Knowles, J., Marx, K.A., SantaLucia, J. Jr. (1999) Statistical mechanical simulation of polymeric DNA melting with MELTSIM, *Bioinformatics* **15**, 370–375.

Blake, R.D., Delcourt, S.G. (1998) Thermal stability of DNA, *Nucleic Acids Res.* **26**, 3323–3332.

Borovik, A.S., Kalambet, Y. A., Lyubchenko, Y. L., Shitov, V.T., Golovanov, E. I. (1980) Equilibrium melting of plasmid ColE1 DNA: electron microscopic visualization, *Nucleic Acids Res.* **8**, 4165–4184.

Crothers, D.M. (1968) Calculation of melting curves for DNA, *Biopolymers* **6**, 1391–1404.

Crothers, D.M., Kallenbach, N.R., Zimm, B.H. (1965) The melting transition of low–molecular–weight DNA: theory and experiment, *J. Mol. Biol.* **11**, 802–820.

Doty, P., Rice, S.A. (1955) Denaturation of deoxypentose nucleic acid, *Biochim. Biophys. Acta* **16**, 446–448.

Daniel, R.M., Cowan, D.A. (2000) Biomolecular stability and life at high temperatures, *Cell Mol. Life Sci.* **57**, 250–264.

Felitsky, D.J., Record, M.T. Jr. (2003) Thermal and urea-induced unfolding of the marginally

stable lac repressor DNA-binding domain: a model system for analysis of solute effects on protein processes, *Biochemistry* **42**, 2202–2217.

Felsenfeld, G., Sandeen, G. (1962) The dispersion of the hyperchromic effect in thermally induced transitions of nucleic acids, *J. Mol. Biol.* **5**, 587–610.

Fixman, M., Freire, J.J. (1977) Theory of DNA melting curves, *Biopolymers* **16**, 2693–2704.

Gray, D.M., Ratliff, R.L., Vaughan, M.R. (1992) Circular dichroism spectroscopy of DNA, *Methods Enzymol.* **211**, 389–406.

Herskovits, T.T., Singer, S.J., Geiduschek, E.P. (1961) Nonaqueous solutions of DNA: denaturation in methanol and ethanol, *Arch. Biochem. Biophys.* **94**, 99–114.

Inman, R.B. (1973) Denaturation mapping of DNA, *Methods Enzymol.* **29**, 451–458.

Kajiyama, T., Miyahara, Y., Kricka, L.J., Wilding, P., Graves, D.J., Surrey, S., Fortina, P. (2003) Genotyping on a thermal gradient DNA chip, *Genome Res.* **13**, 467–475.

Ke, S.-H., Wartell, R.M. (1993) Influence of nearest neighbor sequence on the stability of base pair mismatches in long DNA: determination by temperature–gradient gel electrophoresis, *Nucleic Acids Res.* **21**, 5137–5143.

Lyubchenko, Y.L., Vologodskii, A.V., Frank–Kamenetskii, M.D. (1978) Direct comparison of theoretical and experimental melting curves for RF11 ϕX174 DNA, *Nature* **271**, 28–31.

Mandel, M., Marmur, J. (1968) Use of ultraviolet absorbance–temperature profiles for determining guanine plus cytosine contents of DNA, *Methods Enzymol.* **12**, 195–206.

Manning, G.S. (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides, *Q. Rev. Biophys.* **11**, 179–246.

Marmur, J., Doty, P. (1962) Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature, *J. Mol. Biol.* **5**, 109–118.

Porschke, D. (1974) A direct measurement of the unzippering rate of a nucleic acid double helix, *Biophys. Chem.* **2**, 97–101.

Porschke, D., Eigen, M. (1970) Co-operative non-enzymatic base recognition. I. Thermodynamics of the helix-coil transition of oligoriboadenylic acids at acidic pH, *J. Mol. Biol.* **53**, 123–141.

Record, M.T. Jr., Anderson, C.F., Lohman, T. (1978) Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity, *Q. Rev. Biophys* **11**, 103–178.

SantaLucia, J. Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest–neighbor thermodynamics, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1460–1465.

Springer, M.S., Davidson, E.H., Britten, R.J. (1992) Calculation of sequence divergence from the thermal stability of DNA heteroduplexes, *J. Mol. Evol.* **34**, 379–382.

Suyama, A., Wada, A. (1983) Correlation between thermal stability maps and genetic maps of double–stranded DNAs, *J. Theor. Biol.* **105**, 133–145.

Wilson, W.D., Tanious, F.A., Fernandez-Saiz, M., Rigl, C.T. (1997) Evaluation of drug-nucleic acid interactions by thermal melting curves, *Methods Mol. Biol.* **90**, 219–240.

Siino, J.S., Yau, P.M., Imai, B.S., Gatewood, J.M., Bradbury, E.M. (2003) Effect of DNA length and H4 acetylation on the thermal stability of reconstituted nucleosome particles, *Biochemistry* **42**, 3481–3491.

Thomas, R. (1954) Denaturation of deoxyribonucleic acid, *Biochim. Biophys. Acta* **14**, 231–240.

Wartell, R.M., Hosseini, S.H., Moran, C.P. (1990) Detecting base pair substitutions in DNA fragments by temperature–gradient gel electrophoresis, *Nucleic Acids Res.* **18**, 2699–2705.

Wells, R.D., Larson, J.R., Grant, R.C., Shortle, B.R., Cantor, C.R. (1970) Physicochemical studies on polydeoxyribonucleotides containing defined repeating nucleotide sequences, *J. Mol. Biol.* **54**, 465–497.

Yen, W.-S., Blake, R.D. (1980) Analysis of high–resolution melting (thermal dispersion) of DNA. Methods, *Biopolymers* **19**, 681–700.

Dendritic Cells

Manfred B. Lutz, Thomas Berger, Erwin Schultz, and Alexander Steinkasserer
Department of Dermatology, University Hospital of Erlangen, Hartmannstr, 14, D-91052 Erlangen, Germany

1 Introductory Remarks 215
1.1 Professional Antigen-presenting Cells 215

2 DC Subsets 216
2.1 Primary DC *In Vivo* 217
2.2 *In Vitro* Generation of DC from Hematopoietic Precursor Cells 217

3 Antigen Recognition and Uptake 218

4 Antigen Processing and Presentation 219
4.1 Endosomal Antigen Processing for Presentation on MHC II Molecules 219
4.2 Antigen Processing and Presentation of Cytoplasmic Antigens on MHC I Molecules 219
4.3 Cross-presentation 220
4.4 Glycolipid Antigens and Presentation on CD1 Molecules 220

5 Antigen Delivery Systems 220
5.1 Peptide and Protein Loading 220
5.2 Apoptotic and Necrotic Cells 221
5.3 Antibody Mediated Uptake 222
5.4 Exosomes 222
5.5 Genetic Manipulation of DCs 222
5.5.1 DNA 222
5.5.2 RNA 222
5.5.3 Viral Vectors 223

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

6 DC as Immunostimulators 224
6.1 DC Maturation 224
6.2 DC Migration 224
6.3 Costimulation and Cytokines 224
6.4 DC Interaction with B Cells and NK Cells 225

7 DC and Tolerance Induction 225
7.1 DC Inducing T-cell Anergy 225
7.2 DC Inducing Regulatory T Cells 226
7.3 Deletion 227

8 DC and Virus Interactions 227
8.1 HIV-1 and SIV 227
8.2 Measles Virus 227
8.3 Lymphocytic Choriomeningitis Virus 228
8.4 Human Papillomaviruses 228
8.5 Herpes Simplex Virus Type 1 228
8.6 Cytomegalovirus 228
8.7 Dengue Virus 229
8.8 Ebola Virus 229
8.9 Epstein–Barr Virus 229
8.10 Influenza Virus 229

9 DC as Tumor Vaccines 229
9.1 Clinical Studies in Cancer Patients 229
9.2 Melanoma 230
9.3 Solid Tumors 230
9.4 Virus-associated Malignancies 231
9.5 Other Malignancies 231
9.6 Hematological Malignancies 231
9.7 DC Vaccination in Infectious Diseases 232

Bibliography 232
Books and Reviews 232
Primary Literature 233

Keywords

Clinical Studies
Methods of testing new therapeutic strategies in patients.

Dendritic Cells
Professional antigen-presenting cells derived from bome marrow and present mainly in lymphoid tissue that are specialized for the uptake of particulate material by phagocytosis.

Immunity
The process of the induction of immune responses.

Tolerance
The process of the inhibition and/or down modulation of immune responses.

Tumor Vaccine
A vaccination agent against a special type of cancer.

Virus
An infectious agent consisting of nucleic acid enclosed in a protein coat and capable of replicating within the host cell and spreading from cell to cell.

Dendritic cells (DC) are the most potent antigen-presenting cells known today. They are the only antigen-presenting cells able to induce naïve T helper and cytotoxic T cells and are thus also known as *natures adjuvant*. In this review article, we will cover several aspects regarding the biological function of DC including antigen processing and presentation as well as antigen delivery systems including viral and non-viral systems. Specific chapters will focus on the function of DC as immunostimulators and as inductors of tolerance. A specific section will describe the interaction between DC and viruses including viral escape strategies. The final chapter will focus on the use of modified DC vaccines in the fight against cancer and virus-associated malignancies.

1 Introductory Remarks

1.1 Professional Antigen-presenting Cells

Dendritic cells (DCs), together with B cells and macrophages, belong to the family of professional antigen-presenting cells. These cells have the capacity to take up extracellular material not only for nutritional purposes but also to activate cells of the adaptive immune system such as T cells. They are able to enzymatically digest proteins into small peptides that are then presented on MHC molecules for antigen presentation to T cells. B cells process and present antigen in secondary lymphoid organs to stimulate antigen-specific T helper cells, which in turn provide feedback signals by surface ligands and cytokines. Both macrophages and DCs reside in almost all peripheral tissues as sentinel cells of the immune system to respond to an inflammatory environment or invading microorganisms. Upon immunological challenge, the functional differences between macrophages and DCs becomes apparent. Both cell types take up microbial organisms but for different purposes. Macrophages continuously ingest pathogens to inactivate them and continue their task within the

tissue until the infection is cleared. In contrast, DCs take up pathogens only once to transport them away from the site of infection to the draining lymph nodes and present them to T cells (Fig. 1).

2 DC Subsets

With respect to their hematopoietic development, several subsets of DCs have been proposed. DCs could be generated from many cell-type-restricted precursors, as shown for natural killer (NK), T, and B cells, neutrophils, monocytes, and plasmacytoid cells, but not for mast cells. Moreover, DC precursors can develop along the myeloid or lymphoid lineage and then still convert to the other respective lineage type of DCs in mice. Within secondary lymphoid organs of mice $CD4^-$ $CD8\alpha^-$, $CD4^+$ $CD8\alpha^-$, and $CD4^-$ $CD8\alpha^+$, subsets have been described. The $CD4^+$ $CD8\alpha^-$ DC are considered as "myeloid" DC, and the $CD4^-$ $CD8\alpha^+$ DC subset as "lymphoid" DC. However, typical lymphoid markers can be found on "myeloid" DC as well. In contrast, no markers typical for myeloid cells are found on lymphoid" DC. Nevertheless, there are some functional differences between all of these subsets. Some of the differences might also be a consequence of the organ in which they reside at their immature stage, rather than a result of lineage-restricted commitments.

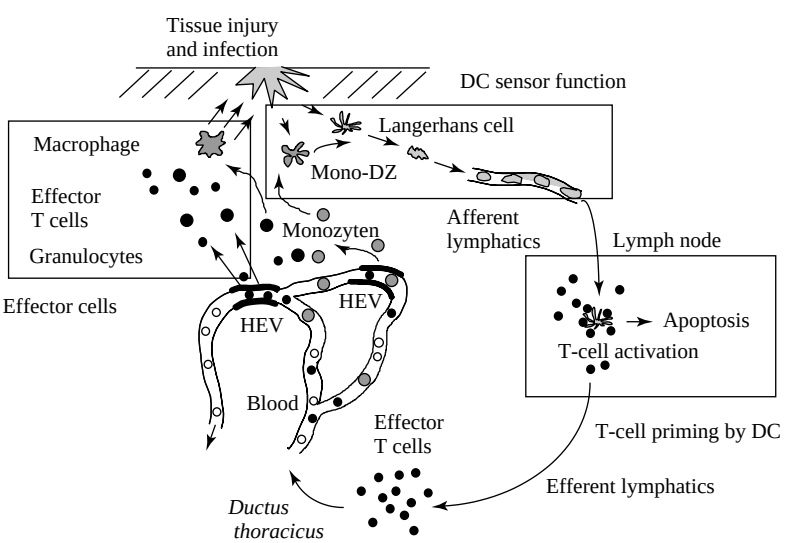

Fig. 1 Dendritic cell functions during an infection. When tissue injury allows the invasion of microbes, dendritic cells are in first line to take up the antigens, transport them into the draining lymph nodes, and present them to the T cells. Activated T cells emigrate from the lymph node and travel through the blood until they reach areas of inflamed endothelium. There the activated T cells transmigrate together with monocytes and granulocytes toward the infection. Monocytes then decide to develop into macrophages, which help to clear the infection, or into dendritic cells, which migrate to the lymph node for T-cell stimulation. This loop can be continued until the resolution of the infection.

2.1 Primary DC *In Vivo*

DCs are present in almost all organs, despite immune-privileged sites such as the central nervous system, testis, ovaries, and others, and are enriched in surface organs such as the skin and mucosa. DCs can be isolated not only from solid nonlymphoid and lymphoid organs but also from the blood or lymph.

In the skin, two DC populations can be distinguished: Langerhans cells (LCs) of the epidermis and dermal DCs. LCs are probably the best investigated primary DC type. As all hematopoietic cell types, LCs have their origins in the bone marrow where precursor cells are formed before they are released into the blood and home in onto the skin. There, the precursors remain within the dermis, divide, and subsequently migrate to the epidermis to continuously maintain the mesh of the residing LCs. LCs seem to reflect a special subset of DCs as their formation depends on the presence of transforming growth factor β-1 (TGF-β) and the transcription factor Id2, as the respective gene-deficient mice lack epidermal langerhan cells (ELC) while other DC populations are present.

In human blood, at least two major subsets of DC can be isolated: myeloid $CD11c^{+}$ $CD123^{dim}$ DC and plasmacytoid $CD11c^{-}$ $CD123^{+}$ DC. The myeloid DC subset is highly similar to DC that can be generated *in vitro* from blood monocytes (Sect. 2.2). The plasmacytoid DCs are a rather young population that was initially designated DC2 as it was found that after T-cell stimulation, predominantly T helper type 2 (Th2) responses resulted. More recent data showed, however, that they can also induce Th1 responses depending on a viral maturation stimulus. The major effector function of plasmacytoid DC may be to enhance antiviral immune responses by producing large amounts of interferon-α.

In human tonsils, a secondary lymphoid organ, myeloid and plasmacytoid DC can also be detected, most of which are in an immature activation state, though a few are in a mature activation state. The distribution in the mouse is very similar except that another subpopulation expressing CD8α is also present. This marker is absent in all DC populations analyzed in humans. Murine $CD8\alpha^{+}$ "lymphoid" DCs are specifically potent in cross-presenting antigens or apoptotic cells.

2.2 *In Vitro* Generation of DC from Hematopoietic Precursor Cells

The study of DCs is heavily influenced by the accessibility of tissues in the respective species studied. While all *in vivo* populations of DCs are rare, the generation of large numbers from hematopoietic precursor cells represent the most common tools that are used to study DC biology. The most accessible way to investigate the human myeloid DC is their generation from $CD14^{+}$ peripheral blood monocytes with granulocyte-macrophage colony-stimulating factor (GM-CSF) plus IL-4. Also, $CD34^{+}$ hematopoietic stem cells can be used for DC generation, but because of their lower frequency in the blood, their expansion is required before DC generation can start. Human plasmacytoid DCs were only developed from immediate blood precursors with IL-3 but do not expand. However, IL-3 is also a growth factor for myeloid DC in mouse and man. In the mouse, flt-3L seems to be the major growth factor to expand not only the plasmacytoid DC but also the myeloid DC. For the generation of murine myeloid DC, the most

effective way is to culture bone marrow precursors with GM-CSF. LCs can also be generated *in vitro* in the presence of TGF-β1.

3
Antigen Recognition and Uptake

Tissue resident DCs are in a so-called *immature differentiation state* and can be viewed as immunological sensors with high susceptibility to microbial products and the capacity to phagocytose them. At the endocytic immature state, DCs are poor presenters of antigens. Microbial challenge initiates their maturation and migration into the draining lymph node where they now have downregulated the endocytosis and upregulated the presentation of antigens transported from the peripheral organ (Sect. 4).

To differentially recognize dangerous or infectious nonself material and to ignore harmless self-antigens, DCs express specialized surface receptors. The recognition of infectious products leads to DC maturation. A wide variety of such receptors is expressed on immature DCs. Some of them initiate phagocytic mechanisms, others stimulate DC maturation and migration. However, receptor-independent mechanisms for antigen uptake have also been described.

Macropinocytosis is a receptor-independent mechanism to engulf liquid vesicles for soluble antigen capture, which is permanently active and is predominantly used by DCs and not by other professional antigen-presenting cells. Uptake of apoptotic material also seems to occur continuously in immature DCs and is mediated by a special set of receptors including $\alpha_V\beta_5$ integrin and CD36. However, both macropinocytosis and uptake of apoptotic material do not induce DC maturation, although such antigens are presented (see below).

Receptor-dependent recognition can occur *indirectly* through complement receptors (CR) binding to microbes opsonized with soluble complement factors from the serum or Fc-receptors (FcR) binding to serum antibodies.

Direct binding of antigens is mediated through so-called *antigen pattern recognition receptors* (PRRs) that recognize evolutionarily conserved pathogen-associated molecular patterns (PAMPs). One major family of such receptors has been described as the toll-like receptors (TLR). So far, 10 different TLRs have been described. They recognize different products (cell wall, DNA) of diverse types of microbes (bacteria, viruses, and parasites). The most prominent family member is TLR4, which has been identified as the endotoxin/lipopolysaccharide (LPS) receptor. TLR engagement stimulates DC maturation, which includes upregulation of MHC class I and II products, costimulatory molecules, and cytokine secretion.

For the triggering of endocytosis by DCs, other receptors are responsible. As already mentioned, the indirect mechanisms of antigen recognition by FcR and CR efficiently mediate the ingestion of the antibody- or complement-bound material. For direct recognition of antigens, some C-type lectin family members, such as the macrophage mannose receptor recognizing glycosylated structures of microbes, or the scavenger receptor family, are involved in endocytosis. When compared with fluid-phase uptake by macropinocytosis, antigens are directed to intracellular compartments that are 100 to 10 000 times more efficient when receptors are employed for

antigen uptake. Such efficient mechanisms for antigen sensing and uptake by immature DCs are a prerequisite for the excellent antigen presentation by mature DCs.

4 Antigen Processing and Presentation

4.1 Endosomal Antigen Processing for Presentation on MHC II Molecules

DCs recognize and phagocytose antigens but do not immediately process them to provide antigen presentation as observed for B cells. The antigens seem to be stored in intracellular compartments where enzymatic digestion is prevented by regulation of the pH. Proteolytic enzymes (cathepsins) are active at lower pH (<5.5) but the vesicles remain at a pH of 6.5 to 7. Storage of antigens might be useful, as the microbial products need to be transported by the DCs into the draining lymph node. This DC migration depends on the distance from the site of infection and the localization of the lymph node but takes around 8 to 24 h in mice.

Endocytosed material is located within the cytoplasm in early and late endosomes. These compartments then become enriched with cathepsins and MHC II molecules derived from the trans-Golgi-network. At this stage, MHC II molecules are protected from random peptide loading by binding to the invariant chain (Ii). For cleavage of proteins into peptides that can fit into the antigen-binding groove of MHC II molecules, acidification by proton pumps is required. This acidification of endosomal compartments might be delayed during the migration of DCs from peripheral organs to the T-cell areas of the lymph node and then released by T-cell contact. Then, sequential and highly coordinated processes of peptide editing by the cathepsins F, L, and S occur, and the replacement of MHC II-bound Ii is controlled by MHC II-like chaperone molecules HLA-DM and HLA-DO (human) or H-2DM and H-2DO (mouse). Finally, the peptide-loaded MHC II molecules are transported along microtubular structures to the cell surface, where they can now present the formerly endocytosed microbial products to $CD4^+$ T cells.

4.2 Antigen Processing and Presentation of Cytoplasmic Antigens on MHC I Molecules

Not all infectious agents can be sensed, endocytosed, and processed for loading of MHC II molecules. Viruses escape endocytosis by their fusion with the cell membrane to release their content into the cytoplasm, and intracellular immune survey takes place. Intracellular proteins can be ubiquitinylated and then enzymatically cleaved within proteasomes, which represent protease complexes. This can occur with self-proteins, tumor-derived proteins, or viral proteins. DC maturation also initiates the formation of immunoproteasomes, which employ additional combinations of proteases. The peptides released are then transported into the endoplasmic reticulum by transporters associated with antigen processing (TAP). There, the freshly synthesized MHC I molecules are still associated with other molecules that protect the MHC I groove from uncontrolled peptide loading. After the regulated opening of the groove by removal of calreticulin, tapasin associates with the MHC I/peptide complex and is transported to

the cell surface for presentation to $CD8^+$ T cells.

4.3 Cross-presentation

The term *cross-presentation* describes the phenomenon of antigen presentation across the typical rules of MHC restriction. This means that exogenous antigens that are classically endocytosed and then presented by MHC II molecules can also appear on MHC I molecules or vice versa. It has been shown that MHC I–restricted peptides within the endoplasmic reticulum can enter autophagosomes, which fuse with endosomes or lysosomes and there meet MHC II molecules for peptide loading. By far, more data are available for the reverse phenomenon, but the exact mechanism by which endosomal antigens enter the cytoplasm is still unclear.

Cross-presentation of phagocytosed apoptotic or virus-infected cells and also of endocytosed protein antigen on MHC I molecules can be observed by all subtypes of DCs *in vitro*. In mice, only the $CD8\alpha^+$ subset seems to cross-present all kinds of antigen. It is unclear whether there are functional differences between *in vitro* generated and *in vivo* generated $CD8\alpha^-$ myeloid DC, or if this is due to the antigen transport pathways and architecture of secondary lymphoid organs.

4.4 Glycolipid Antigens and Presentation on CD1 Molecules

DCs can not only present peptide fragments of proteins to $CD4^+$ and $CD8^+$ T cells but also possess a third way of stimulating other immune cells by antigen presentation on CD1 molecules. CD1 molecules show a high structural similarity to MHC I molecules but do not form hypervariable regions. They are subdivided into two groups. Group 1 CD1 genes (*CD1a*, *b*, *c*) are found in humans but not in mouse, and their expression is restricted to DCs. Only one group-2 CD1 gene (*CD1d*) is found, which is now also present in mice. *CD1d* molecules have a more broad distribution on several cell types. The antigens presented by CD1 molecules are lipid or glycolipid fragments and the cell type able to recognize these complexes are natural killer T (NKT) cells which express T-cell receptors (TCR) like T cells but also markers of NK cells. NKT cells have been described to play tolerogenic or immunogenic roles within the immune system. However, their precise tasks are unclear to date as, so far, only mycobacterial antigens have been described as natural ligands that can bind to CD1 molecules.

5 Antigen Delivery Systems

5.1 Peptide and Protein Loading

The identification of tumor-associated antigens was followed by the determination of immunodominant peptides, which are recognized on tumor cells by T cells. A broad array of tumor-specific peptides presented by different HLA class I molecules and recognized by $CD8^+$ CTLs has been identified, and, recently, several $CD4^+$ helper T-cell epitopes have been added. These defined tumor peptides can be readily synthesized in GMP quality and used to load onto *ex vivo* generated DCs. The sequence of the naturally occurring peptides can be modified by substitution of single

amino acids to improve their binding to a given HLA molecule or the affinity of the HLA-peptide complex for the T-cell receptor. Vaccination with peptide-pulsed DCs has been shown to induce both peptide-specific $CD8^+$ and $CD4^+$ T cells in healthy volunteers and even in advanced cancer patients. Although straightforward and technically easy, the peptide-based approach has some major drawbacks. First, the choice of peptides is restricted to the HLA-typing of the patient, at least for HLA class I peptides, which are less promiscuous binders than HLA class II peptides. Second, vaccination with peptide-pulsed DCs should only induce a T-cell response directed against a limited number of tumor antigens, which may not be sufficient to effectively combat the tumor. In this scenario, the tumor might escape the immune response directed against a small array of peptides and the emergence of antigen-loss tumor cell variants may occur. Third, repetitive vaccinations using relatively high peptide concentrations may favor the induction of low-affinity T cells, which are unable to recognize the tumor cells.

In addition to peptides, one could also use recombinant proteins in order to deliver defined tumor-associated antigens into DCs. This approach, however, requires good expression and purification systems and since these recombinant proteins have to be generated under GMP-conditions, they are most likely very expensive.

5.2 Apoptotic and Necrotic Cells

To optimize the antitumor effects of DC-based immunotherapy, it is tempting to allow the DCs to present the whole antigenic spectrum of a given tumor. This strategy should lead to the induction of an antitumor T-cell response directed against a broad array of tumor antigens including antigens derived from tumor-specific mutations potentially relevant for oncogenesis. Thus, the probability of tumor escape via loss of antigen should be reduced. One intriguing way is to let the DCs phagocytose whole tumor cells or their fragments, resulting in presentation and cross-presentation of tumor antigens on both MHC class I and MHC class II molecules. This would allow the simultaneous induction of tumor-specific CTLs and $CD4^+$ helper T cells, which play a critical role in antitumor immunity.

Several concerns have been raised regarding this approach. First, it is often difficult to obtain sufficient quantities of autologous tumor material from patients. The use of allogeneic tumor cell lines may present an alternative to overcome this problem and even amplify the immune response by activation of alloreactive T cells. Second, immunizing with DCs loaded with whole tumor cell preparations bears the potential risk of inducing autoimmunity, as the DCs will not only present tumor-specific antigens but also numerous self-peptides. However, it has been suggested that immature DCs induce tolerance to self-antigens derived from apoptotic cells. Thus, most patients should be tolerant to many self-peptides that DCs can present to T cells after having phagocytosed apoptotic cells. Third, a reliable monitoring of the immune response may be difficult in the latter case when compared to the use of defined tumor antigens. Fourth, T cells may be generated recognizing peptides only processed by the immunoproteasome expressed by mature DCs and not by the constitutive proteasome expressed by tumor cells. Therefore,

the tumor cells would not be recognized by these T cells unless expression of the immunoproteasome by tumor cells is induced by IFN-γ production, for example, by $CD4^+$ helper T cells at the tumor site. Finally, the optimal preparation of antigen, for example, tumor cell lysates, necrotic or apoptotic material, still has to be defined.

5.3 Antibody Mediated Uptake

Another approach yet to be looked at in clinical trials is the delivery of antigens as immune complexes to FcR-bearing DCs, which results in the formation of both MHC class II and I peptide complexes and *in vitro* can induce cytotoxic as well as helper T cells. It has been demonstrated in the mouse that immune complex delivery to DCs is particularly potent if inhibitory CD32b FcR are circumvented. Clearly, this antigen delivery method is promising as it has the potential to generate multiple epitopes for both MHC class I and II epitopes, thus lifting the restriction to HLA molecules and defined HLA-binding peptides.

5.4 Exosomes

Exosomes may be an attractive means to load DCs with antigen. These are small membrane vesicles resulting from fusion of the plasma membrane with multivesicular endosomes. They contain adhesion and costimulatory molecules, MHC products, and heat-shock proteins and are secreted by various cell types including dendritic and tumor cells. Exosomes derived from peptide-pulsed DCs induce antitumor immune responses in mice. Tumor-specific CTLs can be activated using DCs loaded with exosomes derived from tumor cells.

5.5 Genetic Manipulation of DCs

5.5.1 DNA

An alternative approach is the genetic modification of DCs. Here, one could directly transfect DCs with plasmid DNA coding for full-length tumor antigens, which then present the relevant antigens to human T cells. Plasmids can be readily constructed to not only encode a tumor antigen but also other sequences that lead to better antigen processing and T-cell stimulation. One major drawback with the transfection of DCs with plasmid-derived DNA is the fact that the transduction efficiency is relatively low. Thus, alternative strategies have to be looked into.

5.5.2 RNA

The transfection of RNA coding for defined antigens is an attractive alternative, particularly due to the development of effective electroporation protocols as a substitute for the variable transfection with naked RNA. RNA transfection also does not require costly production of GMP quality proteins and antibodies, and appears advantageous over naked DNA transfection. RNA transfection leads only to transient antigen expression that is, however, sufficient for antigen processing. RNA, in contrast to DNA, does not integrate into the genome, and therefore is much easier to work with from a regulatory point of view. RNA transfection provides an opportunity to test many defined tumor antigens (including universal ones such as survivin and telomerase). In addition, apart from defined RNA, one could also use total tumor RNA, which has the advantage, just like the approach with apoptotic and or necrotic cells, that many not-yet defined tumor antigens will be expressed and one has no restrictions regarding the patients

HLA-haplotype. However, also in this case, questions regarding autoimmunity have to be addressed.

5.5.3 Viral Vectors

The possibility of manipulating viral genomes by biotechnological techniques, together with the recent cloning of tumor-associated antigens has sparked an interest in using genetically modified viruses to express tumor-associated antigens. An increasing number of modified viral vectors have been designed for gene therapy purposes. Transduction of DCs is currently also considered a promising approach for cellular immunotherapy. This optimism has been hampered somewhat as the use of various viruses as vectors has led to the discovery of viral effects disturbing the function of DCs as antigen-presenting cells. Several virus strains that are highly interesting for the *ex vivo* transduction of human monocyte–derived DCs, that is, vaccinia virus (VV), herpes simplex virus (HSV), adenovirus (AV), influenza virus (IV), and lentivirus (LV) have been described. In addition, retroviral vector systems mediating transduction of human $CD34^+$ cell-derived DCs have also been reported.

Recently, several viral vectors have been introduced successfully in the *ex vivo* transduction of DCs, showing a very high transduction efficiency of over 90%. Since antigen concentration seems to be a crucial parameter for the induction of potent immune responses, high expression levels in DCs may represent a clear advantage.

Most of the vectors used have been adopted from the field of gene therapy, where several vector types have been designed in order to achieve the long-term expression of modified genes in nonproliferating cells. However, not all viral vectors are suitable for the *ex vivo* transduction of DCs. One of the main obstacles is the induction of cellular immune responses against viral antigens and the transgene itself, which leads to the destruction of genetically modified target cells, and the preexisting or generated neutralizing antibodies that prevent the initial or booster immunization. The aim of DC-based immunotherapy is not the long-term expression of genes in DCs but rather the induction of strong cellular and humoral immunity against an inserted tumor-associated antigens. To reach this goal, the strict avoidance of any perturbation of DC function appears critical. Furthermore, in gene-replacement therapy, one has to deal with the problem of specific cell targeting in order to induce the expression of the desired gene in a selected tissue or cell. This, however, does not represent a problem for the *ex vivo* transduction of DCs in culture.

However, when viral vectors are used to transduce DCs, the viral mechanisms of immune evasion must be taken into account. Viral escape strategies can either be designed to limit the effect of the preexisting immune response or to hamper the generation of antiviral immunity. Viruses employ various strategies to alter immune responses, including numerous mechanisms influencing the complement system and CTL and NK cell activity.

Considering the dominant role of DCs in the induction of antiviral immune responses, it is not surprising that viruses target DCs in order to block immune responses. In fact, general immunosuppressive phenomena observed, for example, during measles virus (MV) infection can be attributed to the effective interference with DC functions. Consideration of these interactions plays a role in the future development of viral vectors that potently

transduce DCs with no or only minimal inhibitory properties.

6 DC as Immunostimulators

6.1 DC Maturation

A plethora of stimuli have already been described to induce DC maturation. Most can be attributed to microbial stimuli (e.g. bacteria, viruses, parasites, or their products), proinflammatory cytokines (e.g. IL-1, IL-6, TNF-α), T-cell surface molecules (e.g. CD154), or tissue damage (cell necrosis, activated platelets, and extracellular matrix components). The diversity of signals implies that there is well-tuned coordination and combination of these signals for instructing a DC about the precise nature of the infection and the adequate immunological response required. Some of the signals are "weaker," others are "stronger" when applied to cultured DC. Gene-chip microarray techniques have already shown that the proinflammatory cytokine TNF-α and the bacterial cell wall component LPS differentially induce a large number of genes. Inflammatory stimuli such as TNF-α have been proven weaker stimuli than microbial products. Also, exclusive CD40 signaling by the CD40-ligand on T cells does not lead to the induction of IL-12 production *in vivo*. This might indicate that an inflammation or T-cell contact to a DC must not necessarily lead to an immune response in the absence of a pathogen. Moreover, peripheral presentation of self-antigens should actively induce tolerance as not all autoreactive T cells are deleted in the thymus. Transgenic expression of TNF-α in the pancreas did cause insulitis but not autoimmunogenic diabetes.

6.2 DC Migration

In vivo, activation or maturation of the DC occurs mainly in peripheral tissues where microbial infections enter the organism. As the surface organs' skin and mucosa are highly challenged by microbes, the density of DCs in these organs is higher than in others. Upon infection or inflammation, DCs transport and process the captured antigens and transport them via the afferent lymphatics into the T-cell areas of the draining lymph nodes. DC are guided along their way by interacting with extracellular matrix components and directed by the recognition of the chemokines ELC and SLC by the CCR7 chemokine receptor.

6.3 Costimulation and Cytokines

During the process of DC maturation, the number of MHC I and II molecules on the cell surface are upregulated to improve antigen presentation. At the same time, DCs also upregulate the costimulatory molecules CD80 (B7-1) and CD86 (B7-2). Thereby, T cells receive signal 1 through the TCR and signal 2 through CD28, which is enough to trigger their proliferation and IL-2 production. At the same time, the T cells also upregulate activation markers such as CD154 (CD40L), which then ligate the CD40 molecule on the DC as a feedback signal. This can prolong the survival of the DC and improve their cytokine production. For the complete differentiation of $CD4^+$ T cells into T helper cells, additional instruction beyond the signals 1 and 2 are necessary. Polarization into Th1 cells can be provided by IL-12p70 production by DC and is triggered through TLR

ligation on DC. The signals instructing T cells to polarize toward Th2 are less clear. The Th2 driving cytokine IL-4 is not typically produced by the DC, and therefore the absence of IL-12 might be sufficient to enable a default polarization toward Th2. Similar to $CD4^+$ T cells, $CD8^+$ T cells can be primed by DCs. This feature is important for antiviral immunity and antitumor immune responses and is discussed below.

6.4 DC Interaction with B Cells and NK Cells

DC are not exclusively prime naive T-cell responses *in vivo*, but interact with other cell types of the adaptive and innate immune system. The DC has been proposed to transport intact protein antigens into the lymph nodes for B-cell recognition, and, at the same time, to present processed antigen to T cells and thereby form a triple-cell aggregate. Whether this is the rule or the exception needs further proof. Recently, it has been demonstrated that DC can directly activate marginal zone B cells in the spleen.

An interesting observation is also the signal exchange of DC with NK cells. Both cell types seem to regulate each other as NK cells can kill DCs after bone marrow transplantation and also synergize their activities by NK cell-derived IL-4 for DC and DC-mediated augmentation of NK cell killer activity against tumor cells.

7 DC and Tolerance Induction

While DC are widely accepted as stimulators of immune responses, it becomes increasingly evident that they are also major players in maintaining immune homeostasis and preventing autoimmune reactions. First of all, DCs play a decisive role in the selection of T cells in the thymus as they are more effective in presenting self-antigens than thymic epithelial cells. Thymic selection is, however, incomplete as autoimmune-reactive T-cell clones can be isolated from the circulation of healthy individuals. Therefore, additional peripheral tolerance mechanisms must exist and five have been described so far: ignorance, immune deviation, anergy, regulatory T cells, and deletion. The latter three will be mentioned here.

7.1 DC Inducing T-cell Anergy

If no DC stimulation occurs by inflammation or microbial contact, DCs remain immature, express only low amounts of MHC II, and no costimulatory molecules. Such immature DCs can also be found in secondary lymphoid tissues where naive T cells permanently pass by during their normal turnover. Antigen presentation in the absence of costimulation can lead to clonal T-cell anergy. As all isolation procedures of DC will stimulate them so that they rapidly upregulate MHC II and costimulators, active inhibition of maturation is needed to study the antigen presentation capabilities of immature DCs *in vitro*. Many treatments or protocols have been described to block maturation, and indeed, the induction of T-cell anergy or hyporesponsiveness could be shown in many cases.

Immature DCs residing in peripheral tissues would fulfil the prerequisites for T-cell anergy induction, but, under steady state conditions, there are few T cells detectable in the peripheral organs such as the skin. This situation changes when an antigenic peptide is

injected intravenously into mice; antigen-specific T cells spread over many peripheral organs. Such T cells might be anergized when arriving at a noninflammatory peripheral site upon recognizing tissue-derived self-antigen presented by immature DC.

Recently, the classical concept of T-cell anergy has been challenged because of the fact that CTLA4-deficient mice were incapable of generating T-cell anergy. Thus, the induction of T-cell anergy seems to require a negative costimulatory signal through CTLA4. This led to a new model of T-cell anergy where a low density of B7 molecules on the surface of an antigen-presenting cell leads to preferential engagement of CTLA4 and not CD28 due to its 20 to 50 times higher affinity to the B7 molecules. In this respect, it is a prerequisite that immature DC express at least some B7-1 or B7-2 molecules for the induction of T-cell anergy.

7.2 DC Inducing Regulatory T Cells

Regulatory T cells have been defined by their capacity to actively suppress functional activation of immunoreactive T cells. They are constitutively present in the circulation of healthy individuals and mice at about 3 to 10% of all T cells. This $CD4^+$ $CD25^+$ subset seems to derive directly from the thymus and regulates mainly by cell contact or a TGF-β-dependent mechanism. Another subset of regulatory T cells seems to be inducible from the naive T-cell pool. When naive T cells were repetitively restimulated in the presence of IL-10, such T cells started producing IL-10 by themselves and are therefore usually termed *$CD4^+$ IL-10^+ regulatory T cells*. For the induction of the regulatory properties of both subsets, antigen-specific activation through their TCR is required, but the effector mechanisms may partially act antigen-independent. This might especially apply for the IL-10 produced by the $CD4^+IL{-}10^+$ subset.

Regulatory T cells expand slowly after activation and this has been used to term them anergic. Anergy, however, means functional inactivation after a single suboptimal stimulation of T cells, while the induction of regulatory T cells needs repetitive stimulation in the presence of costimulation and results in functionally active T cells, although with immunosuppressive properties. Some recent data indicate that anergic T cells can be developed into regulatory T cells after repetitive stimulation.

Which DC type is able to induce regulatory T cells? For human DCs, it has been reported that immature monocyte-derived DC (Mo-DC) are able to generate regulatory T cells *in vitro*. Such Mo-DC are, however, more mature with respect to their expression of costimulatory molecules than tissue resident DCs, such as ELCs. *In vivo* experiments in mice indicate that the type of DCs inducing $CD4^+IL{-}10^+$ regulatory T cells is rather mature. The induction of $CD4^+IL{-}10^+$ cells *in vivo* might occur through steady state migratory DC, which permanently drain from peripheral tissues into the lymph nodes. These DC transport self-antigens for tolerogenic presentation. As migration, antigen processing, and presentation of tolerogenic DCs is overlapping with immunogenic DC maturation, we term such DCs *semi-mature*. The difference between tolerogenic and immunogenic mature DCs might partially result from their capacity to secrete proinflammatory cytokines, which can only be observed after microbial challenge and results in

"fully mature" DCs that prime T-cell responses.

7.3 Deletion

The complete elimination of autoreactive T cells is the decisive mechanism of T-cell selection in the thymus, but deletion of T cells can also be observed in the periphery. During the resolution of an immune response, an expanded T-cell clone with a certain antigen specificity declines in numbers by showing signs of apoptosis. When too many T cells have to compete for DC presenting the relevant antigen, the T cells might not be restimulated appropriately and might die because of neglect. This scenario also occurs when the antigen disappears after elimination of infectious microbes. In contrast, prolonged exposure to fully mature DC can lead to "activation-induced cell death" (AICD) of T cells.

The initial *in vivo* studies showing deletion of $CD8^+$ T cells as a consequence of tolerogenic cross-presentation found that bone marrow–derived antigen-presenting cells were responsible for this effect. It is now clear that $CD8\alpha^+$ DC were the responsible cell type in this experimental setting. The authors assumed that an immature stage of DC might be responsible for the cross-tolerization. Only recently, *in vitro* studies suggest that partial maturation with TNF-α and PGE_2 is required for cross-tolerance against apoptotic material with deletion of $CD8^+$ T cells as the consequence. This finding is in line with the concept that for "conventional" processing of exogenous antigen to $CD4^+$ cells, partial maturation of DC is required in the same way as for the cross-processing and cross-presenting machinery.

8 DC and Virus Interactions

8.1 HIV-1 and SIV

DC are the best antigen-presenting cells known to date and thus DCs play a crucial role in the induction of antiviral immune responses. On the other hand, many viruses are capable of developing a vast armamentarium in order to interfere with DC biology, thereby blocking antiviral immune responses. Thus, DCs often play a dual role during viral infections, influencing both pathogenesis as well as protective mechanisms.

This has also been shown in the case of HIV-1 and SIV, where HIV-1 infects selected DC populations. In addition, HIV-1 uses DCs as a vector to reach the lymphoid tissues where it infects T cells. The cell surface molecule DC-SIGN has been identified as a DC-specific HIV-1 receptor, which plays a key-role in the dissemination of HIV-1 by DCs. On the other hand, DCs are efficient antigen-presenting cells for HIV-1 and SIV antigens to both $CD4^+$ and $CD8^+$ T cells. In addition, DCs can cross-present nonreplicating viral antigens on class I MHC molecules, thereby allowing the stimulation of $CD8^+$ T cells after the uptake of antibody-coated HIV-1 and dying infected T cells.

8.2 Measles Virus

Measles virus (MV) infection is a major cause of childhood mortality in developing countries and is accompanied by a severe immunosuppression of the infected host. However, virus-specific immunity is efficiently induced, leading to viral clearance

and long-term induction of protective immunity. On the one hand, DCs play a crucial role in the induction of antiviral immune responses, while on the other hand, MV infects DCs, thereby interfering with DC biology. Inhibition of stimulated IL-12 production from MV-infected DCs might affect T-cell responses, favoring Th2 and suppressing Th1 responses. In addition, viral protein expressed on the cell surface of infected DCs most likely also inhibits T-cell responses.

8.3 Lymphocytic Choriomeningitis Virus

DCs also express the receptor (α-dystroglycan) for lymphocytic choriomeningitis virus (LCMV), Lassa fever virus (LFV), and several other arenaviruses. It has been shown that LCMV strains that bind to the receptor with high affinity replicate in the majority of DCs, causing a generalized immunosuppression, and establishing a persistent infection. This immune suppression is caused by the loss of MHC class II molecules and costimulatory molecules on the cell surface of the DCs. Furthermore, the migratory capacity of DCs is hampered by LCMV infections. It is noted, however, that LCMV strains that bind with low affinity rarely replicate in DCs and generate a potent anti-LCMV response, which is able to clear the infection.

8.4 Human Papillomaviruses

DCs also play a crucial role in the case of human papillomavirus (HPV) infection. Here, DCs (including LCs) are pivotal for the induction of T-cell dependent immunity. This, however, depends exclusively on cross-presentation of viral antigens by DCs. This exogenous pathway of MHC class I–restricted antigen presentation plays an important role in the generation of antiviral immunity against several viruses.

8.5 Herpes Simplex Virus Type 1

Herpes simplex virus type 1(HSV-1) have developed several immune escape strategies, and it is clear that many are directed against DCs. HSV-1 infected DC are defective in upregulating costimulatory molecules, do not produce cytokines, and do not acquire responsiveness to chemokines, which are required for migration to secondary lymphoid organs. Furthermore, HSV-1 downmodulates DC functions by interfering with the expression of DC-specific molecules such as CD83. In fact, CD83 is completely degraded in HSV-1 infected DCs, which coincides with a clearly reduced T-cell stimulation, representing yet another new escape strategy.

8.6 Cytomegalovirus

Cytomegalovirus (CMV) infection is characterized by host immunosuppression and multiorgan involvement. CMV-infected DCs display reduced immune stimulatory capacity.

Thus, in CMV-infected DCs, MHC molecules are downregulated, leading to a reduced antigen-presenting capacity. In addition, the apoptosis-inducing ligands CD95L (FasL) and tumor necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL) are upregulated, thereby enabling CMV-infected DCs to delete activated T lymphocytes. Interestingly, as already observed for HIV-1, DC-SIGN

serves as a receptor for CMV, thereby transmitting the virus to permissive cells. Blocking DC-SIGN by specific antibodies inhibited DC infection by primary CMV isolates, and expression of DC-SIGN or its homolog DC-SIGNR rendered susceptible cells permissive to CMV infection.

8.7 Dengue Virus

It has been demonstrated that dengue virus infects and replicates in immature human myeloid DCs. Exposure to live dengue virus led to maturation and activation of both the infected and surrounding, uninfected DCs and stimulated production of TNF-a and IFN-a. Activation of the dengue virus-infected DCs was blunted compared to the surrounding, uninfected DCs, and dengue virus infection induced low-level release of interleukin-12 p70 (IL-12 p70), a key cytokine in the development of cell-mediated immunity. It seems that DCs are the early, primary target of dengue virus in natural infection and that this infection regulates the antiviral immune response.

8.8 Ebola Virus

Ebola virus is a highly lethal pathogen responsible for several outbreaks of hemorrhagic fever. It has been shown that the primate lentiviral binding C-type lectins DC-SIGN and L-SIGN act as cofactors for cellular entry by the Ebola virus. Furthermore, DC-SIGN on the surface of DCs is able to function as a trans-receptor, binding Ebola virus-pseudotyped lentiviral particles and transmitting infection to susceptible cells. Our data underscore a role for DC-SIGN and L-SIGN in the infective process and pathogenicity of Ebola virus infection.

8.9 Epstein–Barr Virus

DCs also play a crucial role in Epstein–Barr virus (EBV) infection. There is growing evidence that DCs, and not EBV-transformed B cells, are responsible for protective antiviral immunity and that EBV nuclear antigen 1 (EBNA1) is the crucial antigen regarding resistance against all types of EBV-associated malignancies.

8.10 Influenza Virus

The exposure of human DCs to IV leads to an efficient infection *in vitro* as shown by the expression of the viral antigens HA and NS1. However, this infection does not lead to rapid cell death and only small amounts of infectious virus particles are produced. Interestingly, infected DC, but not macrophages, can induce CTL responses from purified blood $CD8^+$ cells in the absence of exogenous cytokines.

9 DC as Tumor Vaccines

9.1 Clinical Studies in Cancer Patients

The initial clinical DC-vaccination studies were performed primarily as phase I/II trials to demonstrate feasibility, immunogenicity, and lack of toxicity. Some studies revealed T-cell responses to the applied antigens and also clinical improvement even in advanced cancer patients. The heterogeneity of treated patients, vaccination schemes, DC sources, and generation

protocols as well as the lack of standardized methods to determine a successful induction of immunity makes it difficult to draw firm conclusions about the efficacy of DC-based immunotherapy. Nevertheless, many of these studies provided valuable information and a brief summary of some examples is given below.

9.2 Melanoma

The first human tumor-associated antigen was identified in melanoma and currently this is the best-studied malignancy in the context of DC-therapy. The pioneering work with peptide-loaded, mo-DCs has proven the concept of DC-vaccination to elicit tumor-specific CTL responses even in advanced cancer patients. Although these patients were heavily pretreated, clinical responses could be observed in some patients. In another clinical study, melanoma patients were vaccinated with immature, peptide- or tumor-lysate-loaded DCs. Several clinical responses were observed, but a detailed immunological readout was lacking. CD34-derived DCs have also been used in melanoma vaccination trials. Whereas in one study only moderate immune and minor clinical responses were achieved, strong immune responses as well as tumor regressions have been observed in a recent study. Recently, using mature, peptide-pulsed mo-DCs, a potent induction of tumor-specific $CD4^+$ T-cell responses could be detected in melanoma patients accompanied with a stabilization of disease in 8 out of the 16 patients evaluated. Taken together, DC-based immunotherapy in melanoma patients is particularly promising, as melanoma is an immunogeneic tumor where a multitude of tumor antigens have been described. Apart from minor side effects (fever, swelling of lymph-nodes, cutaneous reactions at the vaccination site) or the development of vitiligo in a few patients, no major toxicity was reported. Further efforts are clearly required in order to enhance the immune response, for example, by simultaneous activation of different immune cells (T cells, NK, and NKT cells).

9.3 Solid Tumors

In contrast to melanoma, many solid tumors are considered less immunogeneic and only few TAAs have been identified for these tumors. Clinical studies have been performed using DCs loaded with peptides as well as with crude tumor extracts or tumor-derived RNA.

CEA is a self-antigen overexpressed in colorectal, lung, and breast cancer. In a phase I study, on vaccinating patients with advanced CEA-expressing malignancies with peptide-pulsed cells or DCs, minor responses and stabilization of disease were observed. Other overexpressed self-antigens are Her2/neu, which is expressed in up to 30% of human breast and ovarian carcinoma, and MUC1, which is found in an aberrant form in these tumors. Patients with advanced breast or ovarian cancer were vaccinated with DCs pulsed with Her2/neu and MUC1-derived peptides after a high-dose chemotherapy. T-cell responses could be detected in 50% of patients.

Renal cancer is relatively resistant to chemotherapy; however, spontaneous remissions have been observed as well as responses to unspecific immunotherapy with IL-2. For specific immunotherapy, DCs have been pulsed with tumor lysate.

Prostate cancer is another promising candidate for DC-based immunotherapy,

since serum markers to monitor the tumor burden and defined antigens such as prostate acid phosphatase, prostate-specific antigen, and prostate-specific membrane antigen are known. Several HLA- A*0201-restricted peptides are available and have been used in a number of clinical trials with mo-DCs. Others have pulsed blood-derived DCs with a GM-CSF/PAP fusion protein and reported a significant decrease in serum PSA-levels, reflecting the reduction of tumor burden in some patients. Recently, 21 patients with metastatic prostate cancer were vaccinated with DCs pulsed with the xenoantigen mouse PAP. All patients developed T-cell responses to mouse PAP and 11 also developed responses to the human homolog. Six out of 21 patients showed a stabilization of disease. Vaccination with xenoantigens may break tolerance to self-antigens and increase the clinical efficacy of DC-based immunizations.

9.4 Virus-associated Malignancies

Targets for immune-intervention strategies in these malignancies are different viral antigens. Since they are foreign to the immune system, the challenge is rather to circumvent viral immune escape mechanisms than to induce autoimmunity. Examples for virus-associated malignancies include hepatocellular carcinoma (Hepatitis B or C virus), Burkitt's lymphoma and nasopharyngeal carcinoma (Epstein–Barr virus), squamous cell cancer of the skin, and cervical carcinoma (human papilloma virus). Studies to vaccinate against HPV-associated cervical cancer have been performed with non-DC vaccines. Some immunological effects have been reported, which, however, were only detectable after intensive *in vitro* restimulation of T cells.

9.5 Other Malignancies

Owing to the immune-privileged status of the central nervous system (CNS), it is questionable as to whether DC-based immunotherapy can be applied to patients with these tumors or not. In fact, in a recent study, patients suffering from malignant glioblastoma or astroglioma were vaccinated with immature mo-DCs pulsed with peptides eluted from autologeous tumor cells. Two patients showed $CD8^+$ infiltrates within the tumor, but no tumor regressions occurred. There is also evidence that DCs pulsed with crude tumor lysate can induce immunity and clinical responses. A study in pediatric cancer patients showed a substantial tumor regression in a child with fibrosarcoma after DC-vaccination. Tumor-specific $CD8^+$ lymphocytes and some clinical effects were also demonstrated in a recent study using tumor lysate-pulsed DCs in gynecological malignancies. However, it is difficult to monitor the successful antigen-loading and T-cell response using lysates from tumors where defined antigens have not yet been described.

9.6 Hematological Malignancies

Myeloma and B-cell lymphoma respond initially to standard chemotherapy, but are finally incurable. For both tumors, idiotypic proteins have been used as specific antigens. B-cell lymphoma was the first human malignancy to be targeted by a DC-based vaccination. Isolated blood DCs were pulsed with KLH and the patient-specific idiotype-protein (Id), which is expressed on the surface or secreted by the tumor cells. This early DC-trial with four patients showed that administration

of DCs was safe and immunogeneic with priming to the neoantigen KLH. In a subsequent study, 35 patients with B-cell lymphoma were treated with Id-pulsed DCs. T-cell and humoral anti-Id responses as well as durable tumor regressions or stabilization were reported for the majority of patients. In myeloma, idiotype- and Id-KLH-pulsed DCs were used to vaccinate patients after high-dose chemotherapy and stem-cell transplantation. Although the majority of these patients developed potent KLH responses, only minor immune responses against the tumor antigens or clinical remissions could be observed.

9.7 DC Vaccination in Infectious Diseases

The prevention of infectious diseases through vaccination represents one of medicine's greatest triumphs. However, many infectious agents such as HIV, dengue virus, *Mycobacterium tuberculosis*, *Plasmodium malariae*, and numerous others escape the immune system and standard vaccination approaches fail. The underlying immune evasion mechanisms may be attributed to a great extent to interference of the microbial pathogen with DCs and their function, for example, inhibition of DC maturation. Augmentation of the immune responses by *ex vivo* generated DCs is therefore an attractive consideration. However, DC-vaccination trials have not been performed so far. The feasibility of this approach is supported by *in vitro* studies with human cells as well as *in vivo* models in the mouse system. Induction of antiparasitic CTLs has been demonstrated *in vitro* with DCs loaded with helminthic proteins and protective immunity against bacteria has been achieved by DC-vaccination in mice. Protection against pulmonary infection with *Pseudomonas aeruginosa* following immunization with *P. aeruginosa*-pulsed DCs has also been demonstrated. One of the best-studied animal models in this regard is leishmaniasis. Lessons learned from this model may have important implications for future DC-based vaccination strategies in infectious diseases.

See also Antigen Presenting Cells (APCs).

Bibliography

Books and Reviews

Alcami, A., Koszinowski, U.H. (2000) Viral mechanisms of immune evasion, *Immunol. Today* **21**, 447–455.

Amigorena, S., Bonnerot, C. (1999) Fc receptor signaling and trafficking: a connection for antigen processing, *Immunol. Rev.* **172**, 279–284.

Austyn, J.M. (1998) Dendritic cells, *Curr. Opin. Hematol.* **5**, 13–15.

Burdin, N., Kronenberg, M. (1999) CD1-mediated immune responses to glycolipids, *Curr. Opin. Immunol.* **11**, 326–331.

Du, X., Thompson, P., Chan, E.K., Ledesma, J., Roe, B., Clifton, S., Vogel, S.N., Beutler, B. (1998) Genetic and physical mapping of the Lps locus: identification of the toll-4 receptor as a candidate gene in the critical region, *Blood Cells, Mol., Dis.* **24**, 340–355.

Figdor, C.G., van Kooyk, Y., Adema, G.J. (2002) C-type lectin receptors on dendritic cells and Langerhans cells, *Nat. Rev. Immunol.* **2**, 77–84.

Heath, W.R., Carbone, F.R. (2001) Cross-presentation in viral immunity and self-tolerance, *Nat. Rev. Immunol.* **1**, 126–134.

Jenne, L., Schuler, G., Steinkasserer, A. (2001) Viral vectors for dendritic cell-based immunotherapy, *Trends Immunol.* **22**, 102–107.

Lutz, M.B., Schuler, G. (2002) Immature, semi-mature, and fully mature dendritic cells: Which signals induce tolerance or immunity? *Trends Immunol.* **23**, 445–449.

Medzhitov, R. (2001) Toll-like receptors and innate immunity, *Nat. Rev. Immunol.* **1**, 135–145.

Poltorak, A., Smirnova, I., He, X., Liu, M.Y., Van Huffel, C., Birdwell, D., Alejos, E., Silva, M., Stahl, P.D., Ezekowitz, R.A. (1998) The mannose receptor is a pattern recognition receptor involved in host defense, *Curr. Opin. Immunol.* **10**, 50–55.

Reid, S.D., Penna, G., Adorini, L. (2000) The control of T cell responses by dendritic cell subsets, *Curr. Opin. Immunol.* **12**, 114–121.

Roncarolo, M.G., Bacchetta, R., Bordignon, C., Narula, S., Levings, M.K. (2001) Type 1 T regulatory cells, *Immunol. Rev.* **182**, 68–79.

Schuler, G., Lutz, M.B., Bender, A., Thurner, B., Röder, C., Young, J., Romani, N. (1999) A guide to the isolation and propagation of dendritic cells, in: Lotze, M.T., Thomson, A.W. (Eds.) *Dendritic Cells: Biology and Clinical Applications*, Academic Press, San Diego, pp. 515–533.

Shevach, E.M., McHugh, R.S., Piccirillo, C.A., Thornton, A.M. (2001) Control of T-cell activation by CD4+ CD25+ suppressor T cells, *Immunol. Rev.* **182**, 58–67.

van Parijs, L., Perez, V.L., Abbas, A.K. (1998) Mechanisms of peripheral T cell tolerance, *Novartis Found. Symp.* **215**, 14–15; discussion 14–20, 33–40.

Walker, L.S., Abbas, A.K. (2002) The enemy within: keeping self-reactive T cells at bay in the periphery, *Nat. Rev. Immunol.* **2**, 11–19.

Primary Literature

Abbas, A.K. (1996) Die and let live: eliminating dangerous lymphocytes, *Cell* **84**, 655–657.

Albert, M.L., Pearce, S.F., Francisco, L.M., Sauter, B., Roy, P., Silverstein, R.L., Bhardwaj, N. (1998) Immature dendritic cells phagocytose apoptotic cells via alphavbeta5 and CD36, cross-present antigens to cytotoxic T lymphocytes, *J. Exp. Med.* **188**, 1359–1368.

Alvarez, C.P., Lasala, F., Carrillo, J., Muniz, O., Corbi, A.L., Delgado, R. (2002) C-type lectins DC-SIGN and L-SIGN mediate cellular entry by Ebola virus in cis and in trans, *J. Virol.* **76**, 6841–6844.

Arrighi, J.F., Hauser, C., Chapuis, B., Zubler, R.H., Kindler, V. (1999) Long-term culture of human CD34(+) progenitors with FLT3-ligand, thrombopoietin, stem cell factor induces extensive amplification of a CD34(−)CD14(−) and a CD34(−)CD14(+) dendritic cell precursor, *Blood* **93**, 2244–2252.

Bender, A., Bui, L.K., Feldman, M.A., Larsson, M., Bhardwaj, N. (1995) Inactivated influenza virus, when presented on dendritic cells, elicits human CD8+ cytolytic T cell responses, *J. Exp. Med.* **182**, 1663–1671.

Balazs, M., Martin, F., Zhou, T., Kearney, J. (2002) Blood dendritic cells interact with splenic marginal zone B cells to initiate T-independent immune responses, *Immunity* **17**, 341–352.

Banchereau, J., Palucka, A.K., Dhodapkar, M., Burkeholder, S., Taquet, N., Rolland, A., Taquet, S., Coquery, S., Wittkowski, K.M., Bhardwaj, N., Pineiro, L., Steinman, R., Fay, J. (2001) Immune and clinical responses in patients with metastatic melanoma to CD34(+) progenitor-derived dendritic cell vaccine, *Cancer Res.* **61**, 6451–6458.

Banchereau, J., Steinman, R.M. (1998) Dendritic cells and the control of immunity. *Nature* **392**, 245–252.

Baumeister, T., Rossner, S., Pech, G., de Bruijn, M.F., Leenen, P.J., Schuler, G., Lutz, M.B. (2003) Interleukin-3Ralpha+ myeloid dendritic cells and mast cells develop simultaneously from different bone marrow precursors in cultures with interleukin-3, *J. Invest. Dermatol.* **121**, 280–288.

Berlyn, K.A., Schultes, B., Leveugle, B., Noujaim, A.A., Alexander, R.B., Mann, D.L. (2001) Generation of CD4(+) and CD8(+) T lymphocyte responses by dendritic cells armed with PSA/anti-PSA (antigen/antibody) complexes, *Clin. Immunol.* **101**, 276–283.

Bickham, K., Munz, C., Tsang, M.L., Larsson, M., Fonteneau, J.F., Bhardwaj, N., Steinman, R. (2001) EBNA1-specific CD4+ T cells in healthy carriers of Epstein-Barr virus are primarily Th1 in function, *J. Clin. Invest.* **107**, 121–130.

Borkowski, T.A., Letterio, J.J., Farr, A.G., Udey, M.C. (1996) A role for endogenous transforming growth factor beta 1 in Langerhans cell biology: the skin of transforming growth factor beta 1 null mice is devoid of epidermal Langerhans cells, *J. Exp. Med.* **184**, 2417–2422.

Brossart, P., Wirths, S., Stuhler, G., Reichardt, V.L., Kanz, L., Brugger, W. (2000) Induction of cytotoxic T-lymphocyte responses

in vivo after vaccinations with peptide-pulsed dendritic cells, *Blood* **96**, 3102–3108.

Brasel, K., De Smedt, T., Smith, J.L., Maliszewski, C.R. (2000) Generation of murine dendritic cells from flt3-ligand-supplemented bone marrow cultures, *Blood* **96**, 3029–3039.

Caux, C., Vanbervliet, B., Massacrier, C., Durand, I., Banchereau, J. (1996) Interleukin-3 cooperates with tumor necrosis factor alpha for the development of human dendritic/Langerhans cells from cord blood CD34+ hematopoietic progenitor cells, *Blood* **87**, 2376–2385.

Cella, M., Facchetti, F., Lanzavecchia, A., Colonna, M. (2000) Plasmacytoid dendritc cells activated by influenza virus and CD40L drive a potent Th1 polarization, *Nat. Immunol.* **1**, 305–310.

Cella, M., Jarrossay, D., Facchetti, F., Alebardi, O., Nakajima, H., Lanzavecchia, A., Colonna, M. (1999) Plasmacytoid monocytes migrate to inflamed lymph nodes and produce large amounts of type I interferon, *Nat. Med.* **5**, 919–923.

Chow, A., Toomre, D., Garrett, W., Mellman, I. (2002) Dendritic cell maturation triggers retrograde MHC class II transport from lysosomes to the plasma membrane, *Nature* **418**, 988–994.

De Bruijn, M.L., Schuurhuis, D.H., Vierboom, M.P., Vermeulen, H., de Cock, K.A., Ooms, M.E., Ressing, M.E., Toebes, M., Franken, K.L., Drijfhout, J.W., Ottenhoff, T.H., Offringa, R., Melief, C.J. (1998) Immunization with human papillomavirus type 16 (HPV16) oncoprotein-loaded dendritic cells as well as protein in adjuvant induces MHC class I-restricted protection to HPV16-induced tumor cells, *Cancer Res.* **58**, 724–731.

Dhodapkar, M.V., Krasovsky, J., Steinman, R.M., Bhardwaj, N. (2000) Mature dendritic cells boost functionally superior CD8(+) T-cell in humans without foreign helper epitopes, *J. Clin. Invest.* **105**, R9–R14.

Dhodapkar, M.V., Steinman, R.M., Sapp, M., Desai, H., Fossella, C., Krasovsky, J., Donahoe, S.M., Dunbar, P.R., Cerundolo, V., Nixon, D.F., Bhardwaj, N. (1999) Rapid generation of broad T-cell immunity in humans after a single injection of mature dendritic cells, *J. Clin. Invest.* **104**, 173–180.

Dubey, P., Hendrickson, R.C., Meredith, S.C., Siegel, C.T., Shabanowitz, J., Skipper, J.C., Engelhard, V.H., Hunt, D.F., Schreiber, H. (1997) The immunodominant antigen of an ultraviolet-induced regressor tumor is generated by a somatic point mutation in the DEAD box helicase p68, *J. Exp. Med.* **185**, 695–705.

Ebner, S., Hofer, S., Nguyen, V.A., Furhapter, C., Herold, M., Fritsch, P., Heufler, C., Romani, N. (2002) A novel role for IL-3: human monocytes cultured in the presence of IL-3 and IL-4 differentiate into dendritic cells that produce less IL-12 and shift Th cell responses toward a Th2 cytokine pattern, *J. Immunol.* **168**, 6199–6207.

Fernandez, N.C., Flament, C., Crepineau, F., Angevin, E., Vivier, E., Zitvogel, L. (2002) Dendritic cells (DC) promote natural killer (NK) cell functions: dynamics of the human DC/NK cell cross talk, *Eur. Cytokine Netw.* **13**, 17–27.

Fong, L., Brockstedt, D., Benike, C., Breen, J.K., Strang, G., Ruegg, C.L., Engleman, E.G. (2001) Dendritic cell-based xenoantigen vaccination for prostate cancer immunotherapy, *J. Immunol.* **167**, 7150–7156.

Geijtenbeek, T.B., Kwon, D.S., Torensma, R., van Vliet, S.J., van Duijnhoven, G.C., Middel, J., Cornelissen, I.L., Nottet, H.S., KewalRamani, V.N., Littman, D.R., Figdor, C.G., van Kooyk, Y. (2000) DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells, *Cell* **100**, 587–597.

Gough, P.J., Gordon, S. (2000) The role of scavenger receptors in the innate immune system, *Microbes Infect.* **2**, 305–311.

Grouard, G., Rissoan, M.C., Filgueira, L., Durand, I., Banchereau, J., Liu, Y.J. (1997) The enigmatic plasmacytoid T cells develop into dendritic cells with interleukin (IL)-3 and CD40-ligand, *J. Exp. Med.* **185**, 1101–1111.

Greenwald, R.J., Boussiotis, V.A., Lorsbach, R.B., Abbas, A.K., Sharpe, A.H. (2001) CTLA-4 regulates induction of anergy in vivo, *Immunity* **14**, 145–155.

Guery, J.C., Sette, A., Appella, E., Adorini, L. (1995) Constitutive presentation of dominant epitopes from endogenous naturally processed self-beta 2-microglobulin to class II-restricted T cells leads to self-tolerance, *J. Immunol.* **154**, 545–554.

Hacker, C., Kirsch, R.D., Ju, X.S., Hieronymus, T., Gust, T.C., Kuhl, C., Jorgas, T., Kurz, S.M., Rose-John, S., Yokota, Y., Zenke, M. (2003) Transcriptional profiling

identifies Id2 function in dendritic cell development, *Nat. Immunol.* **4**, 380–386.

Halary, F., Amara, A., Lortat-Jacob, H., Messerle, M., Delaunay, T., Houles, C., Fieschi, F., Arenzana-Seisdedos, F., Moreau, J.F., Dechanet-Merville, J. (2002) Human cytomegalovirus binding to DC-SIGN is required for dendritic cell infection and target cell trans-infection, *Immunity* **17**, 653–664.

Henri, S., Vremec, D., Kamath, A., Waithman, J., Williams, S., Benoist, C., Burnham, K., Saeland, S., Handman, E., Shortman, K. (2001) The dendritic cell populations of mouse lymph nodes, *J. Immunol.* **167**, 741–748.

Hernando, J.J., Park, T.W., Kubler, K., Offergeld, R., Schlebusch, H., Bauknecht, T. (2002) Vaccination with autologous tumour antigen-pulsed dendritic cells in advanced gynaecological malignancies: clinical and immunological evaluation of a phase I trial, *Cancer Immunol. Immunother.* **51**, 45–52.

Hochrein, H., O'Keeffe, M., Wagner, H. (2002) Human and mouse plasmacytoid dendritic cells, *Hum. Immunol.* **63**, 1103–1110.

Hsu, F.J., Benike, C., Fagnoni, F., Liles, T.M., Czerwinski, D., Taidi, B., Engleman, E.G., Levy, R. (1996) Vaccination of patients with B-cell lymphoma using autologous antigen-pulsed dendritic cells, *Nat. Med.* **2**, 52–58.

Inaba, K., Inaba, M., Romani, N., Aya, H., Deguchi, M., Ikehara, S., Muramatsu, S., Steinman, R.M. (1992) Generation of large numbers of dendritic cells from mouse bone marrow cultures supplemented with granulocyte/macrophage colony-stimulating factor, *J. Exp. Med.* **176**, 1693–1702.

Inaba, K., Turley, S., Yamaide, F., Iyoda, T., Mahnke, K., Inaba, M., Pack, M., Subklewe, M., Sauter, B., Sheff, D., Albert, M., Bhardwaj, N., Mellman, I., Steinman, R.M. (1998) Efficient presentation of phagocytosed cellular fragments on the major histocompatibility complex class II products of dendritic cells, *J. Exp. Med.* **188**, 2163–2173.

Iyoda, T., Shimoyama, S., Liu, K., Omatsu, Y., Akiyama, Y., Maeda, Y., Takahara, K., Steinman, R.M., Inaba, K. (2002) The CD8+ dendritic cell subset selectively endocytoses dying cells in culture and in vivo, *J. Exp. Med.* **195**, 1289–1302.

Kalergis, A.M., Ravetch, J.V. (2002) Inducing tumor immunity through the selective engagement of activating Fc gamma receptors on dendritic cells, *J. Exp. Med.* **195**, 1653–1659.

Kamath, A.T., Pooley, J., O'Keeffe, M.A., Vremec, D., Zhan, Y., Lew, A.M., D'Amico, A., Wu, L., Tough, D.F., Shortman, K. (2000) The development, maturation, turnover rate of mouse spleen dendritic cell populations, *J. Immunol.* **165**, 6762–6770.

Kruse, M., Rosorius, O., Kratzer, F., Stelz, G., Kuhnt, C., Schuler, G., Hauber, J., Steinkasserer, A. (2000) Mature dendritic cells infected with herpes simplex virus type 1 exhibit inhibited T-cell stimulatory capacity, *J. Virol.* **74**, 7127–7136.

Lutz, M.B., Kukutsch, N., Ogilvie, A.L., Rößner, S., Koch, F., Romani, N., Schuler, G. (1999) An advanced culture method for generating large quantities of highly pure dendritic cells from mouse bone marrow, *J. Immunol. Methods* **223**, 77–92.

Morel, S., Levy, F., Burlet-Schiltz, O., Brasseur, F., Probst-Kepper, M., Peitrequin, A.L., Monsarrat, B., Van Velthoven, R., Cerottini, J.C., Boon, T., Gairin, J.E., Van den Eynde, B.J. (2000) Processing of some antigens by the standard proteasome but not by the immunoproteasome results in poor presentation by dendritic cells, *Immunity* **12**, 107–117.

Murphy, G., Tjoa, B., Ragde, H., Kenny, G., Boynton, A. (1996) Phase I clinical trial: T-cell therapy for prostate cancer using autologous dendritic cells pulsed with HLA-A0201-specific peptides from prostate-specific membrane antigen, *Prostate* **29**, 371–380.

Nagata, Y., Ono, S., Matsuo, M., Gnjatic, S., Valmori, D., Ritter, G., Garrett, W., Old, L.J., Mellman, I. (2002) Differential presentation of a soluble exogenous tumor antigen, NY-ESO-1, by distinct human dendritic cell populations, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10629–10634.

Nair, S.K., Boczkowski, D., Morse, M., Cumming, R.I., Lyerly, H.K., Gilboa, E. (1998) Induction of primary carcinoembryonic antigen (CEA)-specific cytotoxic T lymphocytes in vitro using human dendritic cells transfected with RNA, *Nat. Biotechnol.* **16**, 364–369.

Nair, S.K., Hull, S., Coleman, D., Gilboa, E., Lyerly, H.K., Morse, M.A. (1999) Induction of carcinoembryonic antigen (CEA)-specific cytotoxic T-lymphocyte responses in vitro using autologous dendritic cells loaded with CEA peptide or CEA RNA in patients with metastatic malignancies expressing CEA, *Int. J. Cancer* **82**, 121–124.

Nestle, F.O., Alijagic, S., Gilliet, M., Sun, Y., Grabbe, S., Dummer, R., Burg, G., Schadendorf, D. (1998) Vaccination of melanoma patients with peptide- or tumor lysate-pulsed dendritic cells [see comments], *Nat. Med.* **4**, 328–332.

Palucka, A.K. (2000) Dengue virus and dendritic cells, *Nat. Med.* **6**, 748–749.

Pooley, J.L., Heath, W.R., Shortman, K. (2001) Cutting edge: intravenous soluble antigen is presented to cd4 t cells by cd8(−) dendritic cells, but cross-presented to cd8 t cells by cd8(+) dendritic cells, *J. Immunol.* **166**, 5327–5330.

Randolph, G.J. (2001) Dendritic cell migration to lymph nodes: cytokines, chemokines, and lipid mediators, *Semin. Immunol.* **13**, 267–274.

Reis e Sousa, C., Stahl, P.D., Austyn, J.M. (1993) Phagocytosis of antigens by Langerhans cells in vitro, *J. Exp. Med.* **178**, 509–519.

Rissoan, M.-C., Soumelis, V., Kadowaki, N., Grouard, G., Briere, F., de Waal Malefyt, R., Liu, Y.-J. (1999) Reciprocal control of T helper cell and dendritic cell differentiation, *Science* **283**, 1183–1186.

Salio, M., Cella, M., Suter, M., Lanzavecchia, A. (1999) Inhibition of dendritic cell maturation by herpes simplex virus, *Eur. J. Immunol.* **29**, 3245–3253.

Sevilla, N., Kunz, S., Holz, A., Lewicki, H., Homann, D., Yamada, H., Campbell, K.P., de La Torre, J.C., Oldstone, M.B. (2000) Immunosuppression and resultant viral persistence by specific viral targeting of dendritic cells, *J. Exp. Med.* **192**, 1249–1260.

Sallusto, F., Cella, M., Danieli, C., Lanzavecchia, A. (1995) Dendritic cells use macropinocytosis and the mannose receptor to concentrate macromolecules in the major histocompatibility complex class II compartment: downregulation by cytokines and bacterial products, *J. Exp. Med.* **182**, 389–400.

Salomon, B., Cohen, J.L., Masurier, C., Klatzmann, D. (1998) Three populations of mouse lymph node dendritic cells with different origins and dynamics, *J. Immunol.* **160**, 708–717.

Schuler-Thurner, B., Schultz, E.S., Berger, T.G., Weinlich, G., Ebner, S., Woerl, P., Bender, A., Feuerstein, B., Fritsch, P.O., Romani, N., Schuler, G. (2002) Rapid induction of tumor-specific type 1 T helper cells in metastatic melanoma patients by vaccination with mature, cryopreserved, peptide-loaded monocyte-derived dendritic cells, *J. Exp. Med.* **195**, 1279–1288.

Schneider-Schaulies, S., ter Meulen, V. (2002) Triggering of and interference with immune activation: interactions of measles virus with monocytes and dendritic cells, *Viral Immunol.* **15**, 417–428.

Schultz, E.S., Chapiro, J., Lurquin, C., Claverol, S., O. Burlet-Schiltz, Warnier, G., Russo, V., Morel, S., Levy, F., Boon, T., Van den Eynde, B.J., van der Bruggen, P. (2002) The production of a new MAGE-3 peptide presented to cytolytic T lymphocytes by HLA-B40 requires the immunoproteasome, *J. Exp. Med.* **195**, 391–399.

Shortman, K., Liu, Y.J. (2002) Mouse and human dendritic cell subtypes, *Nat. Rev. Immunol.* **2**, 151–161.

Steinman, R.M., Turley, S., Mellman, I., Inaba, K. (2000) The induction of tolerance by dendritic cells that have captured apoptotic cells, *J. Exp. Med.* **191**, 411–416.

Strobel, I., Berchtold, S., Gotze, A., Schulze, U., Schuler, G., Steinkasserer, A. (2000) Human dendritic cells transfected with either RNA or DNA encoding influenza matrix protein M1 differ in their ability to stimulate cytotoxic T lymphocytes, *Gene Ther.* **7**, 2028–2035.

Strobl, H., Riedl, E., Scheinecker, C., Bello, F.C., Pickl, W.F., Rappersberger, K., Majdic, O., Knapp, W. (1996) TGF-beta 1 promotes in vitro development of dendritic cells from CD34+ hemopoietic progenitors, *J. Immunol.* **157**, 1499–1507.

Summers, K.L., Hock, B.D., McKenzie, J.L., Hart, D.N. (2001) Phenotypic characterization of five dendritic cell subsets in human tonsils, *Am. J. Pathol.* **159**, 285–295.

Thurner, B., Haendle, I., Roeder, C., Dieckmann, D., Keikavoussi, P., Jonuleit, H., Bender, A., Maczek, C., Schreiner, D., von Den Driesch, P., Broecker, E.B., Steinman, R.M., Enk, A., K.E., Schuler, G. (1999) Vaccination with mage-3A1 peptide-pulsed mature, monocyte-derived dendritic cells expands specific cytotoxic T cells and induces regression of some metastases in advanced stage IV melanoma, *J. Exp. Med.* **190**, 1669–1678.

Thurnher, M., Klocker, H., Papesh, C., Ramoner, R., Radmayr, C., Hobisch, A., Gastl, G., Romani, N., Ebner, S., Böck, G., Bartsch, G. (1997) Dendritic cells for the

immunotherapy of renal cell carcinoma, *Urol. Int.* **59**, 1–5.

Timmerman, J.M., Czerwinski, D.K., Davis, T.A., Hsu, F.J., Benike, C., Hao, Z.M., Taidi, B., Rajapaksa, R., Caspar, C.B., Okada, C.Y., van Beckhoven, A., Liles, T.M., Engleman, E.G., Levy, R. (2002) Idiotype-pulsed dendritic cell vaccination for B-cell lymphoma: clinical and immune responses in 35 patients, *Blood* **99**, 1517–1526.

Toes, R.E., Ossendorp, F., Offringa, R., Melief, C.J. (1999) CD4 T cells and their role in antitumor immune responses, *J. Exp. Med.* **189**, 753–756.

Wolfers, J., Lozier, A., Raposo, G., Regnault, A., Thery, C., Masurier, C., Flament, C., Pouzieux, S., Faure, F., Tursz, T., Angevin, E., Amigorena, S., Zitvogel, L. (2001) Tumor-derived exosomes are a source of shared tumor rejection antigens for CTL cross-priming, *Nat. Med.* **7**, 297–303.

van der Bruggen, P., Traversari, C., Chomez, P., Lurquin, C., De Plaen, E., Van den Eynde, B., Knuth, A., Boon, T. (1991) A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma, *Science* **254**, 1643–1647.

Van Tendeloo, V.F., Ponsaerts, P., Lardon, F., Nijs, G., Lenjou, M., Van Broeckhoven, C. Van Bockstaele, D.R., Berneman, Z.N. (2001) Highly efficient gene delivery by mRNA electroporation in human hematopoietic cells: superiority to lipofection and passive pulsing of mRNA and to electroporation of plasmid cDNA for tumor antigen loading of dendritic cells, *Blood* **98**, 49–56.

Villadangos, J.A. (2001) Presentation of antigens by MHC class II molecules: getting the most out of them, *Mol. Immunol.* **38**, 329–346.

Yu, J.S., Wheeler, C.J., Zeltzer, P.M., Ying, H., Finger, D.N., Lee, P.K., Yong, W.H., Incardona, F., Thompson, R.C., Riedinger, M.S., Zhang, W., Prins, R.M., Black, K.L. (2001) Vaccination of malignant glioma patients with peptide-pulsed dendritic cells elicits systemic cytotoxicity and intracranial T-cell infiltration, *Cancer Res.* **61**, 842–847.

Zhong, L., Granelli-Piperno, A., Choi, Y., Steinman, R.M. (1999) Recombinant adenovirus is an efficient and non-perturbing genetic vector for human dendritic cells, *Eur. J. Immunol.* **29**, 964–972.

Design and Application of Synthetic Peptides

Gregory A. Grant
Washington University School of Medicine, St. Louis, MO, USA

1 Overview 240

2 Design and Structure 242

3 Nomenclature 244

4 Synthesis 246
4.1 The Solid-phase Method 247
4.2 Chemistry 247

5 Evaluation 250

6 Applications and Perspectives 253
6.1 Antigenic and Immunogenic Uses and Vaccines 253
6.2 Antisense Peptides 256
6.3 Peptide Libraries 257
6.4 Other Applications 257

Bibliography 258
Books and Reviews 258
Primary Literature 259

Keywords

Peptide Bond
The amide bond formed between the α-carboxyl group and the α-amino group of two adjacent amino acid residues in a peptide.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

Peptide
The common usage for a shortened form of polypeptide. Peptide refers to any organic compound composed at least partially of amino acids linked by peptide bonds. The distinction between peptides and proteins is arbitrary with the borderline often placed at 50 to 100 amino acids.

Residue
An amino acid residue is an amino acid in a peptide linkage. Its structure is the same as that of the free amino acid except that it has lost a molecule of water consisting of a hydrogen from the α-amino group and a hydroxyl from the α-carboxyl group.

Epitope
The term *epitope* is used to operationally define that portion of a protein that interacts specifically with some component of a biological system. With respect to the immune system, an epitope was initially defined as that portion of a protein that is recognized by an antibody. However, the usage has now been expanded to include portions of proteins that interact in other ways, such as epitopes that interact with T-cell receptors to induce a heightened antibody response toward another epitope.

A peptide consists of two or more amino acids linked together by amide bonds. Synthetic peptides are peptides that have been produced with organic, synthetic techniques to either mimic naturally occurring peptides or segments of naturally occurring proteins, or to interact in a specific manner with biological systems. Synthetic peptides are important molecules for biomedical research and medicinal chemistry that are useful for understanding biological systems and that have important applications in the development of vaccines and therapeutic agents and the process of *de novo* drug discovery.

1 Overview

Synthetic peptides are organic compounds, generated by synthetic chemical techniques, which are composed of two or more amino acids linked together by a peptide bond. They are often intended to resemble or duplicate naturally occurring peptides or segments of naturally occurring proteins. Synthetic peptides can be linear (**I**), cyclic (**VII**), or branched (**VI**) (bold roman numerals refer to the structures in Fig. 1) and are usually composed of, but not limited to, the 20 naturally occurring L-amino acids or their D-amino acid counterparts (**IV**). Synthetic peptides may also contain non–amino acid constituents (**III**) as well as chemical modifications of the natural side chains (**VI**, **IX**) or the polypeptide backbone structure (**V**). Synthetic peptides are important as tools for research in the biological and biomedical sciences, and a significant number

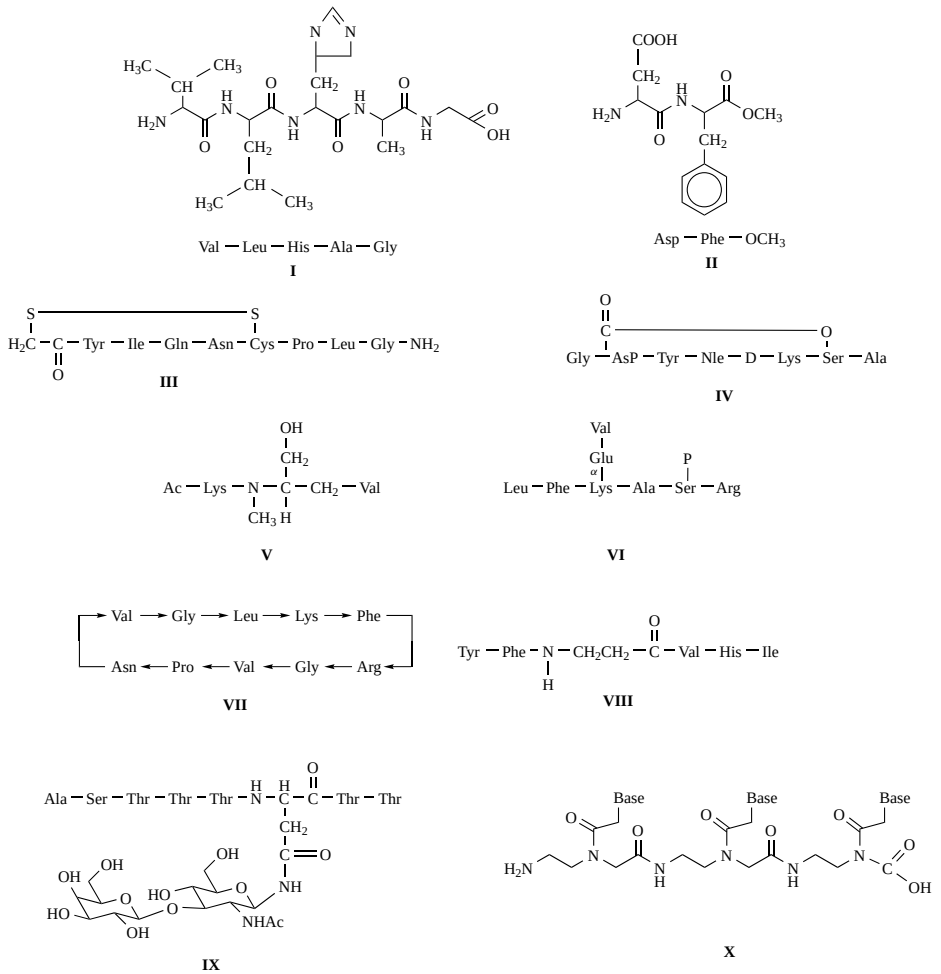

Fig. 1 Examples of synthetic peptides that demonstrate many of the features of chemistry and its nomenclature described in the text. The structures are referred to in the text by their roman numerals: **I**, a typical linear pentapeptide composed of α-peptide bonds and its three letter symbol designation; **II**, the commercial sweetener, aspartame; **III**, [β-mercaptopropionic acid[1]] oxytocin containing a terminal non–amino acid constituent and illustrating the disulfide linkage for clarity; **IV**, a heptapeptide containing a nontypical amino acid (Nle, norleucine) a D-amino acid and an intramolecular lactone bond; **V**, a tetrapeptide containing an N-terminal acetyl group, an alkylated α-amino group, and a fully reduced α-carboxyl or alkane constituent as part of the backbone structure; **VI**, a hexapeptide containing a phosphorylated serine residue and a branched chain connected by a peptide bond between the ε-amino group of lysine and the α-carboxyl group of glutamic acid; **VII**, circular depiction of a cyclized decapeptide; **VIII**, a hexapeptide containing a β-peptide bond; **IX**, an octapeptide with a glycosyl group as a sidechain modification; **X**, an example of a peptide nucleic acid (note the unusual nature of the backbone).

of synthetic peptides have pharmaceutical and commercial uses. For instance, synthetic polypeptide hormones have useful clinical applications, and the commercial sweetener aspartame, a dipeptide (N-L-α-aspartyl-L-phenylalanine-1-methyl ester, **II**), is the basis for an entire industry and is perhaps the best-known synthetic peptide.

Naturally occurring peptides have played crucial roles in life processes since the very beginnings of living organisms. Large peptides, generally referred to as proteins, serve in catalytic, structural, recognitive, and signal transduction processes in both plants and animals. Small peptides have equally important functions as hormones, growth factors, neurotransmitters, and cytokines.

The science of producing and using synthetic peptides is generally recognized as having its beginnings in the late 1800s and early 1900s and is associated mainly with the work of Theodore Curtius and Emil Fischer. Fischer, who is generally regarded as the father of peptide chemistry, was the first to introduce the concepts of peptides and polypeptides and present protocols for their synthesis. Over the years, many investigators have contributed to the advancement of the production and use of synthetic peptides. Perhaps the greatest of these was the introduction of the solid-phase method of synthesis by Bruce Merrifield in 1963. This development, along with improvements in chemistry, instrumentation, and analytical techniques by a large number of other investigators, has made it possible to obtain useful amounts of synthetic peptides on a relatively routine basis.

The genesis of a synthetic peptide can be viewed as being composed of several discrete steps ultimately culminating in its use or application. These steps consist of the design of the peptide to serve a specific purpose, its chemical synthesis, and the evaluation of the integrity of the product, which includes its purification from the initial crude product of the synthesis.

2
Design and Structure

The design of a particular synthetic peptide is often based on its intended use and on considerations of the synthetic chemistry. With modern synthetic techniques it is possible to produce peptides containing in excess of 100 amino acids, although those shorter than 50 are much more common and easier to synthesize. Short peptides tend not to exhibit preferred or stable solution conformations but rather possess a large degree of flexibility. As peptides increase in length, they have a greater tendency to adopt elements of secondary structure, such as helices and β-sheet structures connected by discrete turns, which impart an overall decrease in flexibility. These are structures generally found in naturally occurring proteins and large peptides that are responsible for their specific biological interaction. From a knowledge of the features that contribute to the formation of these structures, peptides can be specifically designed to contain them.

The physical, chemical, and physiological properties of synthetic peptides are determined by the nature of their constituent amino acids. In nature, the range of structure of these amino acids is vast. In addition to the 20 common ribosomally incorporated "protein" amino acids, there are enzymatically formed, posttranslational modifications to these amino acids found in proteins as well as hundreds of naturally occurring, enzymatically synthesized amino acids not found in proteins.

Add to this the imagination of a synthetic chemist, and the possibilities are almost limitless.

In its most basic sense, a synthetic peptide consists of at least two amino acids joined by a peptide bond, which is an amide bond formed between the amino group of one amino acid and the carboxyl group of an adjacent amino acid (Fig. 2). Peptides composed of more than two amino acids possess a number of these amide bonds in series known as the *peptide backbone*. This backbone structure is an important contributor to the overall peptide structure. The resonance properties of the carbon–nitrogen peptide bond give it a substantial double bond character. As a result the amide bond is flat, with the carbonyl carbon, oxygen, nitrogen, and amide hydrogen all lying in the same plane. Thus, no free rotation occurs around the carbon–nitrogen bond. The torsional angle of that bond, called ω, is defined by the atoms $C\alpha-C(O)-N-C\alpha$, and because of the double bond character and the large energy barrier to rotation (25 kcal mol^{-1}), there are two rotational isomers: trans ($\omega = 180°$) and cis ($\omega = 0°$). The trans-isomer possesses the lower energy and is generally found for all peptide bonds not involving proline. The energy of the trans X-Pro bond is somewhat elevated and the barrier to rotation is lowered. Thus, proline-containing peptides will often exhibit cis–trans isomerization.

In addition to ω, the peptide backbone can be completely described with two additional angles, ϕ and ψ. ϕ is the angle described by $C(O)-N-C\alpha-C(O)$ and ψ is described by $N-C\alpha-C(O)-N$. By convention, the angle ϕ is 180° when the two carbonyl carbons are trans to each other and the angle ψ is 180° when the two amide nitrogens are trans to each other.

Peptide and protein structure is organized in several levels. The order of amino acids, or amino acid sequence, is referred to as the primary structure. The secondary structure is the folding up of the primary structure into discrete elements such as helices, sheets, and turns. The tertiary structure describes how the elements of secondary structure found in a single peptide chain interact with each other to produce a final three-dimensional arrangement. The design of synthetic peptides is basically concerned with primary and secondary structure; the concern with secondary structure becoming more important as the size of the peptide increases. With the largest synthetic polypeptides,

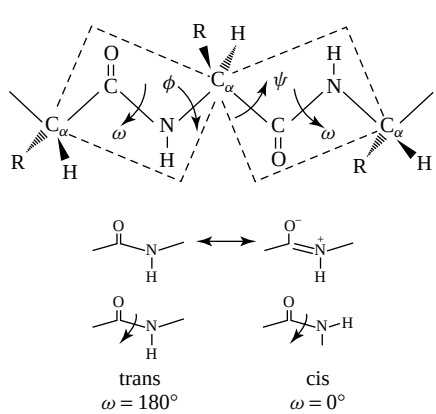

Fig. 2 A stretch of polypeptide demonstrating the planar nature of the peptide bond, the angles that describe the conformation of the polypeptide backbone, and the chirality of L-amino acids. The resonance structures of the peptide bond and cis- and trans-peptide bonds are also shown.

elements of tertiary structure may also be important.

Secondary and tertiary structures are determined by, and can be influenced by a variety of interactions. Basically, secondary and tertiary structures develop from noncovalent interactions such as hydrogen bonds, salt bridges, and hydrophobic interactions resulting from main chain and side chain atoms interacting with each other and with bulk solvent.

The synthetic approach allows the introduction of novel or unusual components such as D-amino acids (**IV**), other unnatural amino acids (**IV**, **X**), ester or alkyl backbone bonds in place of the normal amide bond (**V**), N- or C-alkyl substituents (**V**), amino acids in which the amino group is not attached to the α-carbon (**VIII**, **X**), side chain modifications (**VI**, **IX**), and constraints such as disulfide bridges (**III**) and side chain amide or ester linkages (**IV**). The results of such changes may be higher biological potency, greater stability resulting in longer biological lifetime, or the ability to specifically interact with or covalently label a biological macromolecule or receptor for localization or structure–function relationship studies.

Peptide solubility is another potentially important concern in the design of a synthetic peptide. If the intention is to work with a peptide in aqueous media, it is important that the peptide has sufficient polar character for solubility. This is a point that might especially be overlooked when a peptide is made based on a sequence found internally in a protein. The solubility properties of a particular peptide may be vastly different from the protein when it is removed from the context of the surrounding protein structure. The peptides, composed completely of uncharged amino acids, tend to be less soluble in aqueous environments and their insolubility tends to increase as the size of the peptide increases. Therefore, it is advisable to incorporate charged amino acids into a synthetic peptide when possible. Histidine, lysine, and arginine promote solubility at pHs where their side chains tend to be positively charged and aspartic acid and glutamic acid promote solubility at pHs where their side chains tend to be negatively charged. The contribution of the charge properties of the α-amino and α-carboxyl groups should also be taken into account. In borderline cases, uncharged, polar groups can also promote solubility.

3 Nomenclature

A proper use of nomenclature for describing peptides is important. The International Union of Pure and Applied Chemists (IUPAC), in conjunction with the International Union of Biochemists (IUB), have formulated a set of guidelines for peptide nomenclature (*J. Biol. Chem.* (1985) **260**, 14–42). These guidelines are too extensive to be described here in their entirety, but some of the more common conventions are presented.

The sequence of amino acids within a peptide is usually designated with either the one-letter or three-letter amino acid symbol (Table 1). The amino acid symbols denote the L-configuration unless designated with a D or DL preceding the symbol (**IV**). In the case of linear polypeptides, the convention is to express the peptide left to right from the amino-terminus to the carboxy-terminus (**I**). Thus, a pentapeptide might be represented either as V-L-H-A-G or as Val-Leu-His-Ala-Gly. The α-amino group is assumed to be at

Tab. 1 Common symbols for amino acids in expressing peptide structure.

Residue	*Three-letter symbol*	*One-letter symbol*[a]
Alanine	Ala	A
Asparagine[b]	Asn	N
Aspartic acid[b]	Asp	D
Arginine	Arg	R
Cysteine	Cys	C
Glutamic acid[b]	Glu	E
Glutamine[b]	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V
Hydroxylysine	Hyl	
4-Hydroxyproline	Hyp	
α-Aminobutyric acid	Abu	
Norvaline	Nva	
Norleucine	Nle	
Homoserine	Hse	
Homocysteine	Hcy	
Ornithine	Orn	
Citrulline	Cit	
α-Aminoisobutyric acid	Aib	
γ-Carboxyglutamic acid	Gla	

[a]Symbols for less common amino acids should be defined in the individual publications in which they appear.
[b]Asx or B and Glx or Z are used when the distinction between aspartic acid/asparagine and glutamic acid/glutamine is uncertain. This convention is most commonly applied in sequence analysis.

the left-hand side of the amino acid symbol when using hyphens and at the point of the arrow when using arrows for special cases such as cyclic peptides (**VII**, and see below). Normally, when the termini are free amino or carboxyl groups, no additional designation of this is necessary. However, in order to avoid confusion or ambiguity, the state of the termini is often indicated (**III**). Free termini are designated with a H– at the amino end and an –OH at the carboxyl end so that the above peptide would be written as H-Val-Leu-His-Ala-Gly-OH. Similarly, a C-terminal amide or aldehyde group would be designated by $-NH_2$ or –H respectively.

Occasionally, the names of small peptides will be written out. In this case, the names of all amino acids contributing carboxyl groups to peptide bonds end in -yl, and the amino acid with the terminal carboxyl group retains its original name. The peptapeptide shown above and in **I** would thus be expressed as valylleucylhistidylalanylglycine.

Cyclic peptides are usually expressed on a single line of text, similar to linear peptides, with a line or "extended bond" drawn to connect the joined amino acids that lead to the cyclic nature of the peptide. If the amino acids are joined between side chains, such as in disulfide bonds or side chain lactams, the line starts and ends in the middle of the amino acid name or abbreviation, such as

Gly-Val-Glu-Ala-Leu-Ile-Lys-Ala-Thr.

If the structure drawn in this manner is ambiguous, additional atoms can be illustrated, as shown in Fig. 1 (**III**, **IV**), to clarify the meaning.

If the connection is between the α-amino and α-carboxyl groups of the termini, the line is drawn from the sides of the names, such as

Gly-Val-Glu-Ala-Leu-Ile-Lys-Ala-Thr.

In this case, the N→C direction is read left to right. If the peptide is written out in a circular fashion, as if it was written along the circumference of a circle or a loop, then the N→C direction must be designated by arrows (**VII**). Finally, cyclic peptides are sometimes designated as such with the prefix *cyclo*, such as cyclo(Gly-Val-Glu-Ala-Leu-Ile-Lys-Ala-Thr), to indicate the same structure as immediately above.

Synthetic analogs of naturally occurring peptides that have specific names associated with them represent a special case. These include substances such as insulin, oxytocin, enkephalin, vasopressin, and bradykinin to name a few. For example, the polypeptide hormone, oxytocin, has the structure

Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly-NH2,

where the cysteine residues form a disulfide bridge. If one or more amino acids are replaced in the synthesis of an oxytocin analog, the new amino acids and their positions are placed in brackets before the name. Thus, [Phe^2, Asn^4]-oxytocin refers to an oxytocin analog in which the tyrosine at position 2 and the glutamine at position 4 have been replaced with phenylalanine and asparagine respectively. The prefix *endo* is used to designate the insertion of an extra amino acid into the chain. Thus, *endo*-Ala^{7a}-oxytocin denotes an alanine insertion between Pro^7 and Leu^8. Similarly, deletion of an amino acid is designated by the prefix *des*, for example, *des*-Ile^3-oxytocin denotes that Ile at position 3 has been deleted. Peptide sequences that form fragments of a longer sequence that has a trivial name are designated by the trivial name followed by numbers giving the positions of the first and last amino acids and then a designation describing the number of amino acids in the fragment. For example, the peptide Tyr-Ile-Gln is a segment of oxytocin (see above) and can be designated as oxytocin-(2–4)-tripeptide.

4 Synthesis

Peptides can be synthesized either in solution, which is the classical approach, or by solid-phase methods. With the exception of specialized applications in which solution phase chemistry is used

because it is easier, less expensive, or the necessary chemistry is not amenable to the solid-phase method, most peptides made for research purposes are synthesized with the solid-phase method that was introduced by Merrifield in 1963. This technique, for which he was awarded the Nobel Prize in Chemistry in 1984, was primarily responsible for opening the way to the widespread use of synthetic peptides in chemical and biomedical investigations.

Solid-phase methods are readily automated and most solid-phase syntheses performed today are done with the aid of machines. A variety of peptide synthesizers are commercially available for batchwise and continuous flow operations as well as for the synthesis of multiple peptides within the same run.

4.1 The Solid-phase Method

The solid-phase approach to peptide synthesis is presented in Fig. 3. Basically, it consists of anchoring the growing peptide to an insoluble support or resin. This is usually accomplished through the use of a chemical handle that links the support to the first amino acid at the carboxyl terminus of the peptide. Subsequent amino acids are then added, one at a time, in a stepwise manner until the peptide is fully constructed. The solid-phase method allows excess reagents and soluble reaction by-products to be easily removed by filtration and washing, thus doing away with the need to isolate or crystallize the product after each step.

Synthesis of the growing polypeptide chain is accomplished by formation of an amide bond between the free α-amino group of the receiving amino acid (anchored to the support) and the activated α-carboxyl group of the incoming amino acid (in solution). Note that synthesis proceeds from the carboxy end to the amino end. To prevent unwanted side reactions with reactive amino acid side chains and self-coupling of the α-carboxyl group of one incoming amino acid to the α-amino group of another incoming amino acid, these groups must be protected at the appropriate times during synthesis. As depicted in Fig. 3, the group protecting the α-amino group of the amino acid being coupled must ultimately be removed, after coupling is complete, in order to accept the next amino acid during the next cycle. Chemical groups that allow this are referred to as "temporary protecting groups". The reactive amino acid side chains, on the other hand, must remain permanently protected throughout the synthesis. Thus, chemical groups that accomplish this are referred to as "permanent protecting groups". At the end of the synthesis, the completed peptide is cleaved from the resin support and the permanent protecting groups are removed, usually in a single step process, to yield the desired peptide product.

4.2 Chemistry

A number of different chemical approaches for the synthesis of peptides on solid phase supports have been investigated and are in use today. Generally, two levels of protecting group stability are required. The permanent protecting groups must be such that they are stable during the repetitive conditions of removing the temporary protecting groups from the growing peptide chain. However, the permanent protecting groups, as well as the attachment handle holding the peptide to the support, must ultimately be removed under conditions

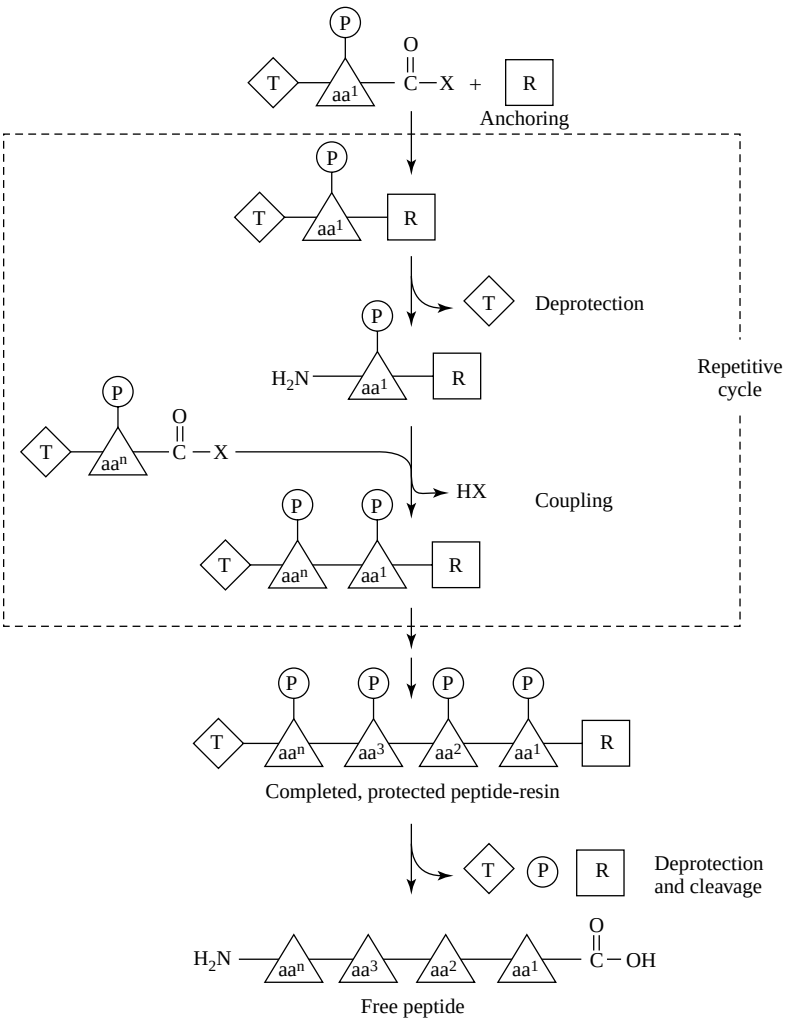

Fig. 3 Stepwise solid-phase synthesis of linear peptides. R, resin or insoluable polymeric support; T, temporary protecting group; P, permanent protecting group; X, denotes activated carboxyl group; $aa^1 \ldots aa^n$, amino acid residues numbered from the C-terminus.

Fig. 4 Common protection schemes for solid-phase peptide synthesis (a) Protection scheme for solid-phase synthesis based on the Boc group utilizing graded acidolysis. The Boc group is removed at each step by trifluoracetic acid (TFA). The permanent side chain protecting groups and the PAM linkage are cleaved simultaneously by hydrofluoric acid (HF) or other strong acid. (b) Mild orthogonal protection scheme for solid-phase synthesis based on the Fmoc group. The Fmoc group is removed at each step by the base catalyzed elimination mechanism, usually using piperidine, as shown. Permanent side chain–protecting groups and the HMP/PAB linkage are cleaved by treatment with TFA. (Figures are reprinted with permission from Freeman, W. H. and Fields, G. B., et al. in Grant, G. A. (Ed.) (2002) *Synthetic Peptides: A User's Guide*, 2nd edition, Oxford University Press, Oxford, New York).

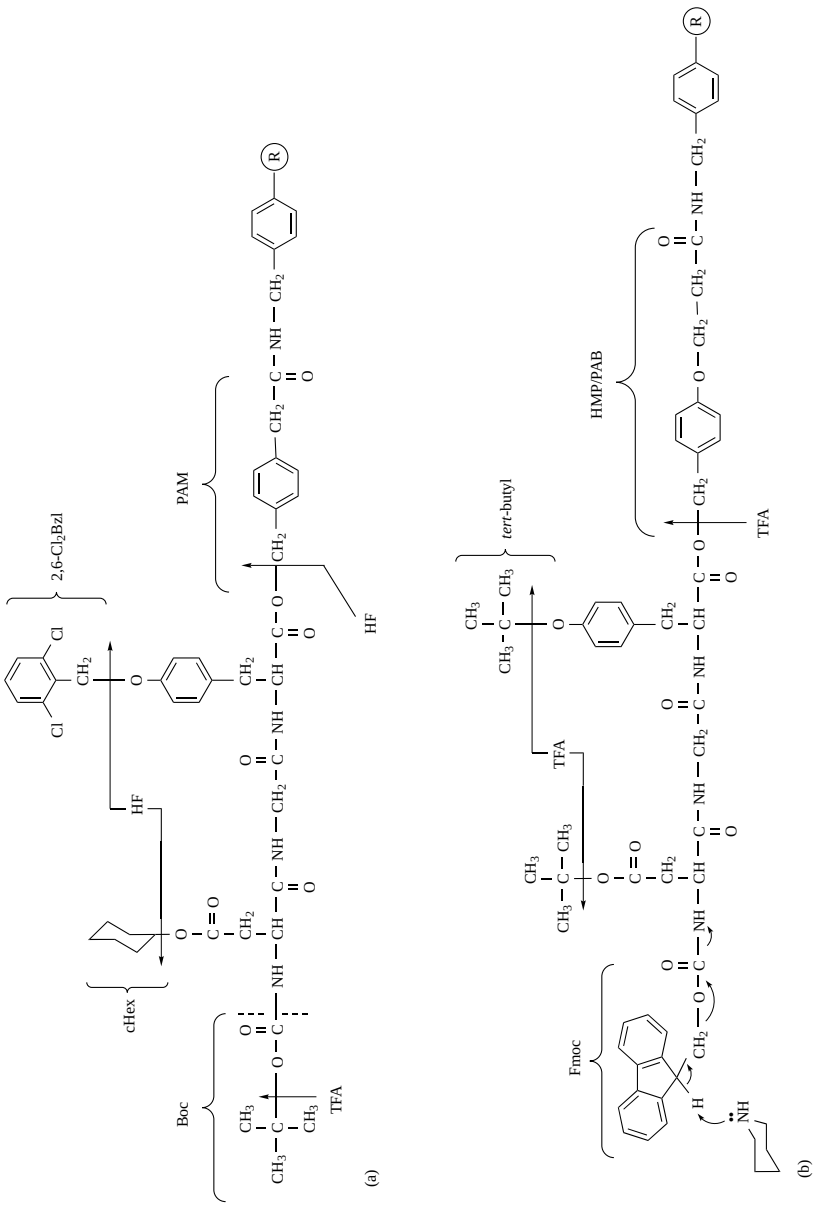
2,6-Cl2Bzl
Cl
Cl
CH2
O
CH2
HF
cHex
O
C=O
CH2
Boc
CH3
CH3
CH3
O
C
O
C
O
NH
CH
C
O
NH
CH2
C
O
NH
CH
C
O
O
TFA
PAM
CH2
CH2
C
O
NH
CH2
R
HF
(a)
Fmoc
CH2
O
C
O
NH
CH
C
O
NH
CH2
C
O
NH
CH
C
O
O
CH3
CH3
C
CH3
O
CH2
TFA
CH3
CH3
C
CH3
O
C=O
CH2
TFA
tert-butyl
CH3
CH3
C
CH3
O
CH2
HMP/PAB
O
CH2
CH2
C
O
NH
CH2
R
TFA
H
NH
(b)

that do not harm the underlying peptide structure. Two general methods have emerged as the most common and are depicted in Fig. 4. They are referred to as either "Boc", which uses the tert-butyloxycarbonyl group, or "Fmoc", which uses the 9-fluorenylmethyloxycarbonyl group for temporary α-amino protection. Today, the Fmoc chemistry is probably used more often than the Boc chemistry largely because the permanent protecting groups and the resin handle can be removed under milder conditions with reagents that do not present as much of a potential hazard as those used for Boc chemistry. However, the Boc chemistry remains a vital and useful chemistry that is still used by many peptide chemists.

A wide variety of permanent protecting groups are available for these two general approaches and a few of them are depicted in Fig. 4 and Table 2. The Boc strategy relies on graded acidolysis to selectively differentiate between temporary and permanent protection. Temporary protecting groups are removed by treatment with trifluoroacetic acid while cleavage from the resin and removal of permanent protecting groups is usually accomplished with anhydrous hydrofluoric acid or other strong acid. The Fmoc strategy is a milder orthogonal approach in that the two classes of protecting groups are removed by different chemical mechanisms such that either class can be removed while preserving the other class. The temporary Fmoc group is removed by a base-catalyzed β-elimination mechanism, usually employing piperidine. Removal of permanent protecting groups as well as cleavage from the support is accomplished by acidolysis, usually with trifluoroacetic acid.

Formation of the peptide bond between the carboxyl group of the incoming amino acid and the amino group on the peptide-resin requires that the carboxyl group be activated chemically. There are currently four major coupling strategies that utilize either active esters, preformed symmetrical anhydrides, acid halides, or generation of the activated acid *in situ* with a variety of different reagents.

Peptides modified at the α-amino group can be easily produced by treatment with an acylating agent, such as acetic anhydride, prior to removal of the permanent protecting groups. Resins are also available for both Boc and Fmoc strategies that will directly yield peptides with either a free acid or an amide function at the carboxy-terminus.

In some instances, larger peptides have been formed from smaller ones by the use of segment condensation and chemoselective ligation techniques. In either case, two or more smaller peptides are synthesized in such a way that they can subsequently be linked together to form a large peptide. However, these procedures are not common and are still relatively difficult to perform.

5 Evaluation

Although the synthesis of any given peptide is often considered routine and can be performed with standard protocols on automated instruments, the chemistry used to produce synthetic peptides is complex and can be influenced by a number of factors that cannot always be foreseen or controlled. As a result, one must never assume that the final product is correct until it is proven to be so.

Homogeneity of the product and correct covalent structure are the two main aims in the production of synthetic peptides. The initial product obtained from the

Tab. 2 Some commonly used side chain protecting groups for solid-phase peptide synthesis.

Amino acid	*Compatible with boc protection*	*Compatible with fmoc protection*
Asp/Glu	Benzyl (OBzl) 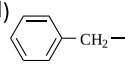	*tert*-Butyl (*Ot*-Bu)
Ser/Thr	Benzyl (Bzl)	*tert*-Butyl (*t*-Bu) 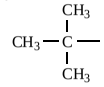
Tyr	2-Bromobenzyloxycarbonyl (2-BrZ) 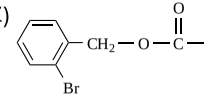	*tert*-Butyl (*t*-Bu)
Lys	2-Chlorobenzyloxycarbonyl (2-ClZ) 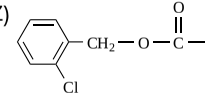	*tert*-Butyloxycarbonyl (Boc) 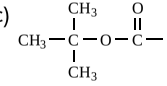
His	Benzyloxymethyl (Bom) 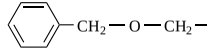	Triphenylmethyl (Trt)
Cys	4-Methylbenzyl (Meb) 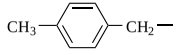	Triphenylmethyl (Trt) 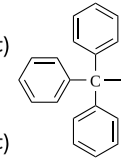
Asn/Gln		Triphenylmethyl (Trt)
Arg	4-Toluenesulfonyl (Tos)	2,2,5,7,8-Pentamethyl chroman-6-sulfonyl (Pmc) 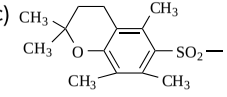
Trp	Formyl (CHO) (on N^{in})	

stepwise solid-phase peptide synthesis will contain a variety of by-products in addition to the desired peptide. These products usually consist of deletion sequences generated during synthesis, and chemically modified peptides generated either during synthesis or during the cleavage and deprotection step. A successful synthesis is one in which these products are minimal and the intended peptide is present in sufficient yield to allow its purification from the mixture. Today, virtually all peptide purifications can be accomplished by high-pressure liquid chromatography (HPLC) using either reverse-phase or ion-exchange columns.

A common problem encountered in peptide synthesis is the production of deletion sequences. These come about as a result of either incomplete coupling or incomplete deprotection at any given cycle. Incomplete coupling of the incoming amino acid leaves a free α-amino group on the peptide chain that will persist into the next cycle and be able to couple with the next incoming residue. Incomplete deprotection results in the retention of a protected α-amino group that can subsequently be deprotected in the next cycle and thus be able to receive the incoming amino acid at that cycle. Both cases result in a finished peptide that is missing one amino acid residue. Deletion sequences due to incomplete coupling are probably the most common and the easiest to manage. The elongation of deletion peptides due to incomplete coupling can be halted by the use of a capping step. Capping simply refers to the reaction of the free α-amino group remaining at the end of a coupling step with a reagent such as acetic anhydride that derivatizes the α-amino group in such a way that it can no longer be elongated during subsequent cycles. Although capped peptides represent contaminants that will have to be removed from the final product, they are generally easier to separate from the full-length product than are deletion peptides, which may differ from the full-length product by only one amino acid.

Unwanted chemically modified peptides result either from incomplete deprotection of the side chains or from random adducts formed at some point during synthesis or cleavage and deprotection. As long as they do not represent an overwhelming proportion of the total product, they can usually be separated from the desired product without the need to identify their structure. When unwanted side-products show up as major species, either the chemical protocol or instrument operation needs to be checked and modified. In this regard, it is important to note that the isolation of a single major product from a synthesis does not guarantee that it is the intended product. Facile modifications are not uncommon and there are many instances in which the wrong peptide was isolated and used in experimentation because its structure was not verified.

There are numerous analytical techniques available for the evaluation of synthetic peptides. The techniques that have been commonly used include analytical high performance liquid chromatography (HPLC), capillary zone electrophoresis (CZE), amino acid compositional analysis, UV spectroscopy, sequence analysis, and mass spectrometry.

Today, mass spectrometry is most often used as the primary means of characterizing synthetic peptides. Efficient characterization of synthetic peptides is best obtained by a combination of analytical reverse-phase HPLC and mass spectrometry. Analytical reverse-phase HPLC provides evidence for the homogeneity of

the peptide product by presenting a picture of the number of peptide species present based on their chromatographic elution characteristics. Multiple peaks indicate heterogeneity and the area under the peaks is an indication of the relative amounts in each peak. Note, however, that a single main peak suggests relative homogeneity but does not exclude the possibility of comigrating species. Both electrospray ionization mass spectrometry (ESMS) and matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) can be used for the accurate determination of peptide mass. The position of modifications and deletions, if present, can be identified by sequencing with either chemical methods (Edman chemistry) or tandem mass spectrometry. In this respect, sequencing is most easily accomplished by collision-induced decomposition on ion-trap or tandem instruments such as Quadrupole-Time of Flight (Q-TOF) instruments. Amino acid analysis is the best technique for the quantitation of the product. It specifically measures the amount of amino acids present and is accurate to within ±5 to 10%. Gravimetric techniques will also measure the amount of salt and adsorbed water present in the peptide preparation. This can be substantial, up to 50% of the total, thus causing as much as a twofold error.

6 Applications and Perspectives

Peptides have become an important class of molecules in biomedical research and medicinal chemistry. Already a number of synthetic hormones or their analogs have found use as therapeutic agents. Synthetic peptides show great potential in being effective in regulating such physiological processes as blood pressure, neurotransmission, reproduction, and endocrine functions, to name a few. Synthetic peptide substrates for enzymes such as proteases, kinases, and phosphatases have been used successfully to study enzyme kinetics and mechanism and develop effective inhibitors. Synthetic peptides are seeing great success as antigens for the preparation of polyclonal and monoclonal antisera and offer the potential for more effective vaccines. The process of *de novo* drug discovery is enhanced through the development of peptide libraries. Epitope mapping of proteins with synthetic peptides has identified regions important for biological activity. Synthetic peptides have been instrumental in receptor localization, characterization, and isolation. Finally, even whole enzymes have been produced by synthetic methods. A comprehensive discussion of the many and diverse uses for synthetic polypeptides is obviously beyond the scope of this article. Therefore, the following sections attempt to briefly present some of the more recent and interesting applications in the field.

6.1 Antigenic and Immunogenic Uses and Vaccines

The ability of synthetic peptides to immunologically mimic regions of whole proteins has resulted in their extensive use as antigens and probes of protein epitopes. Antibodies that are raised against a native protein bind to relatively small areas on the surface of that protein and can often recognize and bind to the same antigenic sequence in the absence of the rest of the protein. Thus, highly antigenic sequences or epitopes can be mapped in

proteins by synthesizing overlapping peptides, spanning the full sequence of the protein and testing them for binding of the antibody.

Many gene products have been identified and isolated by prediction of antigenic regions in the sequence of a protein that was initially deduced from a gene sequence. Synthetic peptides corresponding to these predicted antigenic regions are used to produce antibodies that in turn are used to specifically interact with the intact protein product for identification and isolation. Although this approach has proven very successful, the main problem in its implementation is the prediction of the protein's antigenic sequences. Although there is no guarantee of success, certain general guidelines for choosing surface epitopes have been described. First, use of a hydropathy profile locates areas that are more hydrophilic. This is based on the assumption that the surface of a protein is exposed to solvent and will thus be hydrophilic in nature. If you only have a short section to work with, choose areas that are charged or polar. In general, the immediate amino and carboxyl ends of a protein may tend to be more solvent accessible than the interior sequence. These regions may be better candidates if you have no other criteria. Secondly, since short stretches of sequence in proteins do not have exposed termini that can become charged, synthesize the peptide with α-amino and α-carboxyl blocking groups such as N-terminal acetyl groups and C-terminal amide groups. Keep in mind that the peptide must be sufficiently polar to be soluble in aqueous solution. Peptides of 10 to 15 residues are usually adequate for the production of antibodies, and including a cysteine residue at either end can be helpful for cross-linking to carrier proteins via heterobifunctional cross-linking reagents (see Fig. 5). The use of a carrier protein is usually necessary because small peptides, in general, are not sufficiently antigenic to produce an immune response on their own. This is because they generally lack T-cell epitopes that can be provided by the carrier. Antibodies are produced by B-cells that recognize or respond to certain epitopes. T-cells also recognize epitopes, but these epitopes are different from those recognized by B-cells. In the presence of T-cell epitopes or antigens, T-cells stimulate B-cells to boost their antibody production. Thus, to produce a good antisera, both B- and T-cell epitopes should be present.

In addition to linking polypeptide antigens to carrier proteins, two purely synthetic approaches have also been shown to be successful in producing antisera (Fig. 5). One approach, using the concept of a template-assembled synthetic protein (TASP) involves synthesizing both B- and T-cell peptide epitopes as branches on a β-sheet peptide backbone template. Another approach, known as the multiple antigen peptide system (MAPS), uses a peptide core of radially branching lysine residues covalently connected to a conventional solid support for peptide synthesis. Peptide antigen sequences are built on the α- and ε-amino groups of the terminal multiple lysine residues. The resultant structure consists of multiple, tightly packed peptide antigen sequences on a single structure. Although no apparent T-cell epitopes are present, MAPS antigens have been shown to produce a strong immunogenic response that results in high titer antisera for the peptide antigen. This is thought to be due to the local high concentration of peptide presented.

One very exciting area that is based on the use of synthetic peptides as antigens is in the development of vaccines.

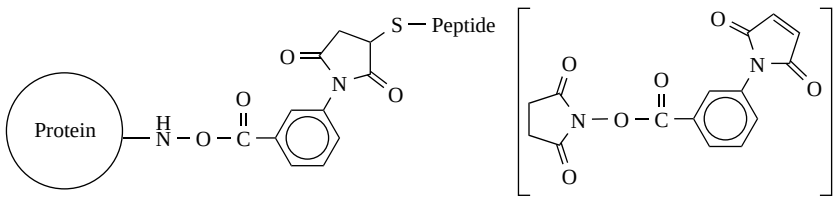

Carrier protein linked via heterobifunctional reagent

(a)

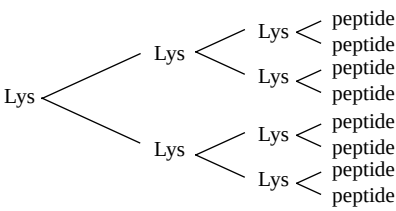

Multiple antigenic peptide system (MAPS)

(b)

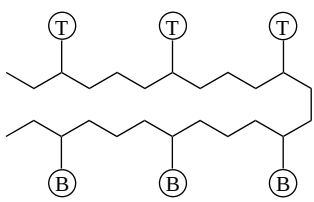

Template assembled synthetic protein (TASP)

(c)

Fig. 5 Methods for the use of synthetic peptides as antigens to produce high titer antisera. (a) a synthetic peptide/carrier protein conjugate using the heterobifunctional cross-linking reagent *m*-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS, in brackets). (b) multiple antigenic-peptide system with peptide antigen linked to a polylysine core. (c) incorporation of both T- and B-cell peptide epitopes on a β-sheet peptide template.

Synthetic peptide-based vaccines are actively being pursued as potential antiviral, antibacterial, and antiparasitic agents for such diseases as foot-and-mouth disease, poliovirus, hepatitis B, influenza, HIV, feline leukemia virus, malaria, diphtheria, and cholera to name a few. Prevention of some carcinomas, such as melanoma, is also being explored by this technique on the basis of segments of melanoma cell specific proteins that are presented antigenically to the host's immune system. These vaccines are based on the observations that viruses and microorganisms present major, protein-derived antigenic determinants on their surfaces, that many bacterial diseases stem from protein or peptide toxins secreted by the bacteria, and that cancer cells express antigens that are recognized as foreign to the host. If effective antisera can be raised against these regions, they have the potential to

neutralize the virus or toxic agent. Major impediments to developing effective peptide-based vaccines are the inability to definitively recognize the proper determinants to be used as a basis for the synthetic peptide antigen and the problem of serological diversity. With viruses, for example, selecting peptides that will be effective as vaccines is more difficult than just developing antibodies that will react with viruses. Virus particles possess a large number of different epitopes, but not all will produce antibodies that will neutralize virus infectivity. Even when a major neutralizing epitope is identified, diversity within that epitope can be a major problem. For instance, foot-and-mouth disease virus has as many as seven different serotypes throughout the world, and vaccination with one does not necessarily protect against the others. This is exacerbated by the fact that variation can exist within each serotype; many viruses display a much larger number of serotypes, and some display the ability to mutate fairly rapidly. However, if invariant regions can be found among the various subtypes, the number of synthetic peptide-based vaccines needed to protect against the agent can be reduced and vaccine development becomes feasible. The prospect of developing vaccines from synthetic peptides has the distinct advantages that the antigen is chemically well defined, relatively stable, and is free from possible biological contamination that can be a major problem with vaccines from biological sources.

6.2 Antisense Peptides

The area of antisense peptides is very new and still the subject of controversy that stems mainly from the fact that the experimental observations have not yet been adequately explained theoretically.

An antisense peptide is a peptide whose sequence is deduced from the nucleotide sequence that is complementary to the nucleotide sequence which codes for a naturally occurring peptide. In general, although it has not been shown to be the case in all instances, antisense peptides have shown selectivity in binding to their sense counterparts and have displayed unusual or surprising biological activity.

The basis for the unusual interaction between sense and antisense peptides appears to stem from the genetic code. In general, the codons for hydrophobic or hydrophilic amino acids on the coding strand of DNA are complemented by codons for hydrophilic or hydrophobic amino acids, respectively, on the opposite strand. Furthermore, reading of the complementary DNA strand in either direction displays the same pattern of hydrophobicity in the resultant amino acid sequence since the middle base of the codon seems to specify the hydropathic nature of the amino acid.

One of the most stunning successes in the use of antisense peptides is the demonstration that antibodies raised against an antisense peptide of adrenocorticotropin hormone (ACTH) were successful in recognizing and purifying the ACTH receptor. In other words, the antibodies, acting like an image of an image, interacted with the receptor in a manner similar to ACTH. Although many other similar successes have been reported, it is important to note that many failures have also been seen. Thus, the exact nature of the interaction is not well understood but the area of antisense peptides continues to be an exciting and growing field.

6.3 Peptide Libraries

The process of developing new synthetic peptides that are useful for diagnostic or therapeutic purposes usually involves the costly and time-consuming process of synthesizing hundreds of peptide analogs based on the structure of an original active peptide. In spite of recent advances that have improved this process, the task of screening all possible combinations of even short peptides has been impractical. For instance, there are 64 000 000 possible hexapeptides (20^6) for all combinations of the 20 naturally occurring protein amino acids.

Recently, the development of synthetic peptide libraries has greatly aided this task. Basically, the libraries are generated by synthesis of a mixture of a large number of peptides of defined length in which each position, or a defined number of positions, contains a random mixture of amino acids. These heterogeneous mixtures are then screened with some appropriate assay system, and the reactive sequences are isolated and determined. Several different methods for the generation and screening of synthetic peptide libraries have been described.

6.4 Other Applications

One approach to identifying, locating, and isolating receptors to polypeptide ligands such as peptide hormones and opioids is to establish chemical cross-links between the peptide ligand and the receptor subsequent to their binding. An ever-increasing number of heterobifunctional cross-linking reagents, many with photoactivatable reactive groups for *in situ* activation, are being developed for this purpose which can be attached to the biologically active peptides at a variety of locations. Once attached to the peptide, these reagents are capable of producing covalent cross-links between the peptide and the receptor. If the synthetic peptide is labeled, the label can be followed in isolating the receptor and identifying the area of contact in the peptide – the receptor complex.

Peptide nucleic acids are compounds that are synthesized with a peptide-like backbone with DNA- and RNA-derived bases as side chains (**X**). These polyamide oligomers appear to be capable of invading the duplex strand of DNA at specific locations determined by the identity of the bases, causing displacement of one strand, and disrupting transcription elongation of that stretch of DNA. As such, they display potential as antisense and antigene agents that may be superior to some phosphodiester backbone-based antisense agents with regard to stability and bioefficacy.

The use of peptidomimetics is a relatively new but growing area related to the field of synthetic polypeptides. Although the mimetic structure is often not strictly a peptide because it may completely lack peptide bonds, the structure is based on that of a peptide. Since the three-dimensional shape of a peptide determines how it will interact with its biological target, producing specific stable conformations in synthetic peptides is one of the most challenging areas in the design of biologically active peptides. In general, small peptides tend not to exhibit stable or even preferred conformations in solution. The general idea with peptidomimetics is to synthesize a compound that mimics some structural or conformational property that a peptide might assume biologically or that mimics some stable, relatively rigid

structure found in proteins, such as β- or γ-turns. The approach in synthesizing peptidomimetics is usually to replace part or all of the normal peptide structure with nonpeptidelike chemical structure so that the resulting conformation is stabilized and mimics the original peptide in space. The final mimic often has very little chemical resemblance to the peptide, but displays stable three-dimensional elements of conformation which is similar to that assumed by the peptide. Some approaches that have been taken in this regard have been to replace peptide bonds with rigid structures that maintain bond distance and side chain orientation or introduction of intramolecular cross-links to stabilize turns or side chain positions.

The area of peptidomimetics has been very actively pursued in rational drug design and pharmaceuticals. Although success has been limited owing to the difficulty in accurately mimicking the exact active biological conformation, some advances have been made and will no doubt continue since peptidomimetics have many potential applications. One successful example is the design of a tricyclic, conformationally constrained lactam that is among the most potent angiotensin converting enzyme (ACE) inhibitors known.

Finally, whole, biologically active proteins have been produced by purely synthetic methods. One striking example of this is the complete chemical synthesis of the human immunodeficiency virus-1 (HIV-1) protease that possesses full biological activity. Even more striking is the demonstration that the HIV-1 protease synthesized from all D-amino acids displays specificity for substrates with opposite chirality to the natural substrate.

The above sections only touch on some of the diverse and potentially very useful applications of synthetic peptides. From what has been accomplished with synthetic polypeptides already, it seems clear that they will continue to play an increasingly larger role in biomedical research and in the treatment of human, animal, and plant disease.

See also Bioorganic Chemistry; Combinatorial Phage Antibody Libraries; Medicinal Chemistry.

Bibliography

Books and Reviews

Atherton, E., Sheppard, R.C. (1989) *Solid Phase Peptide Synthesis: A Practical Approach*, IRL Press, Oxford.

Bodanszky, M. (1988) *Peptide Chemistry, A Practical Textbook*, Springer-Verlag, Berlin.

Bodanszky, M., Bodanszky, A. (1984) *The Practice of Peptide Synthesis*, Springer-Verlag, Berlin.

Grant, G.A. (Ed.) (2002) *Synthetic Peptides: A User's Guide*, 2nd edition, Oxford University Press, Oxford, New York.

Jones, J. (1991) *The Chemical Synthesis of Peptides*, Clarendon Press, Oxford.

Kates, S.A., Albericio, F. (Eds.) (2000) *Solid-Phase Peptide Synthesis: A Practical Guide*, Marcel Dekker, New York.

Sewald, S., Jakubke, H.-D. (2002) *Peptides: Chemistry and Biology*, Wiley-VCH, Germany.

Smith, B.J. (Ed.) (1997) *Protein Sequencing Protocols*, Humana Press, Totowa, New Jersey, USA.

Stewart, J.M., Young, J.D. (1984) *Solid Phase Peptide Synthesis*, Pierce Chemical Company, Rockford, IL.

Van Regenmortel, M.H.V., Briand, J.P., Muller, S., Plaue, S. (1988) in: Burdon, R.H., Kippenberg, P.H. (Eds.) *Synthetic Polypeptides as Antigens*, In the Laboratory Techniques in Biochemistry and Molecular Biology Series, Elsevier, Amsterdam.

Wieland, T., Bodanszky, M. (1991) *The World of Peptides: A Brief History of Peptide Chemistry*, Springer-Verlag, Berlin.

Primary Literature

Addona, T., Clauser, K. (2002) De novo peptide sequencing via manual interpretation of MS/MS spectra, *Current Protocols in Protein Science*, John Wiley & Sons, New York, pp. **16**.11.1–**16**.11.19.

Angeletti, R.H., Bonewald, L.F., Fields, G.B. (1997) Six-year study of peptide synthesis, *Methods Enzymol.* **289**, 697–717.

Bergman, M., Zervas, L. (1932) Ueber ein allgemeines Verfahren der Peptid-Synthese, *Ber. Dtsch. Chem. Ges.* **65**, 1192–1201.

Berk, S.C., Rohrer, S.P., Degrado, S.J., Birzin, E.T., Mosley, R.T., Hutchins, S.M., Pasternak, A., Schaeffer, J.M., Underwood, D.J., Chapman, K.T. (1999) A combinatorial approach toward the discovery of non-peptide, subtype-selective somatostatin receptor ligands, *J. Combinatorial Chem.* **1**, 388–391.

Blalock, J.E., Bost, K.L. (1986) Binding of peptides that are specified by complementary RNAs, *Biochem. J.* **234**, 679–683.

Blalock, J.E., Smith, E.M. (1984) Hydropathic anti-complementarity of amino acids based on the genetic code, *Biochem. Biophys. Res. Commun.* **121**, 203–207.

Bodanszky, M., Ondetti, M.A., Levine, S.D., Williams, N.J. (1967) Synthesis of secretin II. The stepwise approach, *J. Am. Chem. Soc.* **89**, 6753–6757.

Bodanszky, M., Williams, N.J. (1967) Synthesis of secretin I. The protected tetradecapeptide corresponding to sequence 14–27, *J. Am. Chem. Soc.* **89**, 685–689.

Bost, K.L., Blalock, J.E. (1986) Molecular characterization of a corticotropin (ACTH) receptor, *Mol. Cell. Endocrinol.* **44**, 1–9.

Carr, S.A., Annan, R.S. (1999) Overview of peptide and protein analysis by mass spectrometry, *Current Protocols in Protein Science*, John Wiley & Sons, New York, pp. **16**.1.1–**16**.1.27.

Crabb, J.W., West, K.A., Dodson, W. Scott, Hulmes, J.D. (1997) Amino acid analysis, *Current Protocols in Protein Science*, John Wiley & Sons, pp. **11**.9.1–**11**.9.42.

Edman, P., Begg, G. (1967) A protein sequenator, *Eur. J. Biochem.* **1**, 80–91.

Fischer, E. (1902) Ueber einige Derivate des Glykocolls Alanins und Leucins, *Ber. Dtsch. Chem. Ges.* **35**, 1095–1106.

Fischer, E. (1903) Synthese von Derivaten der Polypeptide, *Ber. Dtsch. Chem. Ges.* **36**, 2094–2106.

Fischer, E. (1906) Untersuchungen über Aminosäuren, Polypeptide, und Proteine, *Ber. Dtsch. Chem. Ges.* **39**, 530–610.

Fischer, E. (1907) Synthese von Polypeptiden XVII, *Ber. Dtsch. Chem. Ges.* **40**, 1754–1767.

Grant, G.A. (1999) Synthetic peptides for the production of antibodies that recognize intact proteins, *Current Protocols in Protein Science*, John Wiley & Sons, pp. **18**.3.1–**18**.3.14.

Harington, C.R., Mead, T.H. (1935) Synthesis of glutathione, *Biochem. J.* **29**, 1602–1611.

Henzel, W.J., Stults, J.T. (1999) Matrix-assisted laser desorption/ionization time-of-flight mass analysis of peptides. *Current Protocols in Protein Science*, John Wiley & Sons, pp. **16**.2.1–**16**.2.11.

Hirschmann, R., Nutt, R.F., Veber, D.F., Vitali, R.A., Varga, S.L., Jacob, T.A., Holly, F.W., Denkewalter, R.G. (1969) Studies on the total synthesis of an enzyme V. The preparation of enzymatically active material, *J. Am. Chem. Soc.* **91**, 507–508.

Hruby, V.J., Al-Obeidi, F., Kazmierski, W. (1990) Emerging approaches in the molecular design of receptor selective peptide ligands: conformational, topographical and dynamic considerations, *Biochem. J.* **268**, 249–262.

Hruby, V.J., Matsunaga, T.O. (2002) Applications of synthetic peptides, in: Grant, G.A. (Ed.) *Synthetic Peptides: A User's Guide*, Vol. 5, Oxford University Press, Oxford, New York, pp. 292–376.

Milton, R.C., Milton, S.C., Kent, S.B. (1992) Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show reciprocal chiral substrate specificity, *Science* **256**, 1445–1448.

Moore, R.E., Young, M.K., Lee, T.D. (2000) Protein identification using a quadrupole ion trap mass spectrometer and SEQUEST database matching, *Current Protocols in Protein Science*, John Wiley & Sons, pp. **16**.10.1–**16**.10.9.

Nair, S.A., Kim, M.H., Warren, S.D., Choi, S., Songyang, Z., Cantley, L.C., Hangauer, D.G. (1995) Identification of efficient pentapeptide substrates for the tyrosine kinase pp60c-src, *J. Med. Chem.* **38**, 4276–4283.

Nielsen, P.E. (2001) Peptide nucleic acid: a versatile tool in genetic diagnostics and

molecular biology, *Curr. Opin. Biotechnol.* **12**, 16–20.

Nutt, R.F., Brady, S.F., Darke, P.L., Ciccarone, T.M., Colton, C.D., Nutt, E.M., Rodkey, J.A., Bennett, C.D., Waxman, L.H., Sigal, I.S., Anderson, P.S., Veber, D.F. (1988) Chemical synthesis and enzymatic activity of a 99 residue peptide with a sequence proposed for the human immunodeficiency virus protease, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7129–7133.

Pinilla, C., Appel, J.R., Blanc, P., Houghten, R.A. (1992) Rapid identification of high affinity peptide ligands using positional scanning synthetic peptide combinatorial libraries, *BioTechniques* **13**, 901–905.

Pinilla, C., Appel, J.R., Blondelle, S.E., Dooley, C.T., Eichler, J., Ostresh, J.M., Houghten, R.A. (1994) Versatility of positional scanning synthetic combinatorial libraries for the identification of individual compounds, *Drug Dev. Res.* **33**, 133–145.

Reim, D.F., Speicher, D.W. (1997) N-terminal sequence analysis of proteins and peptides, *Current Protocols in Protein Science*, John Wiley & Sons, pp. **11**.10.1–**11**.10.38.

Rodda, S.J., Geysen, H.M., Manson, I.J., Schaafs, P.G. (1986) The antibody response to myoglobin-I. Systematic synthesis of myoglobin peptides reveals location and substructure of species dependent continuous antigenic determinants, *Mol. Immunol.* **23**, 603–610.

Schneider, J., Kent, S.B.H. (1988) Enzymatic activity of a synthetic 99 residue protein corresponding to the putative HIV-1 protease, *Cell* **54**, 363–368.

Schwyzer, R., Sieber, P. (1963) Total synthesis of adrenocorticotrophic hormone, *Nature* **199**, 172–174.

Sifferd, R.H., du Vigneaud, V. (1935) A new synthesis of carnosine, with some observations on the splitting of the benzyl group from carbobenzoxy derivatives and from benzylthio ethers, *J. Biol. Chem.* **108**, 753–761.

Tam, J.P., Spetzler, J.C. (1997) Multiple antigen peptide system, *Methods Enzymol.* **289**, 612–637.

Tuchscherer, G., Servis, C. Corradin, G., Blum, U. Rivier, J., Mutter, M. (1992) Total chemical synthesis, characterization, and immunological properties of an MHC class I model using the TASP concept for protein de novo design, *Protein Sci.* **1**, 1377–1386.

du Vigneaud, V., Ressler, C., Swan, J.M., Roberts, C.W., Katsoyannis, P.G., Gordon, S. (1953) The synthesis of an octapeptide amide with the hormonal activity of oxytocin, *J. Am. Chem. Soc.* **75**, 4879–4880.

Developmental Cell Biology

Frank John Dye
Western Connecticut State University, Danbury, CT, USA

1 Historical Perspective 264
1.1 Origins of Cell Biology 264
1.2 Origins of Developmental Biology 264
1.3 Relationship between Cell and Developmental Biology 266

2 Cell Activities Underlying Development 268
2.1 Intracellular Signal Transduction 268
2.1.1 Receptors 268
2.1.2 Intracellular Signal Transduction 270
2.1.3 Transcription Factors 271
2.2 Cell Signaling 274
2.2.1 Juxtacrine Signaling 274
2.2.2 Autocrine Signaling 274
2.2.3 Paracrine Signaling 275
2.2.4 Endocrine Signaling 276
2.2.5 Growth Factors 276
2.3 Cell–Cell Interaction 277
2.3.1 Cell Junctions 277
2.3.2 Cell Signals 278
2.4 Cell–Matrix Interaction 278
2.4.1 Integrin Signaling 279
2.4.2 Cell Movement 279
2.4.3 Tubulogenesis 284

3 Cell Differentiation 287

4 The Cell Cycle and Development 288

5 Organogenesis 290

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

5.1 Chalones 292

6 Stem Cells 293

7 Cloning 297

Bibliography 299
Books and Reviews 299
Primary Literature 300

Keywords

Cell Cycle
A series of phases, designated G1, S, G2, and M, that a cell sequentially passes through between the time it is formed by cell division and the time it itself divides.

Chalones
Proposed soluble factors produced locally by an organ as it develops that act to limit further growth of that organ once it reaches its appropriate size.

Clone
Generally, a population of cells derived from a single cell by mitosis.

Embryonic Induction
A developmental phenomenon wherein one part of the embryo determines the fate of another part of the embryo.

Embryonic Stem (ES) Cells
Stem cells derived from the embryo.

Growth Factors
Extracellular polypeptide signal molecules that can stimulate a cell to grow or proliferate.

Intracellular Signal Transduction
Cascades of chemical reactions used by cells to convey information from the cell surface to the interior of the cell (e.g. the nucleus) where the information is acted upon; the cell makes use of intracellular signal transduction pathways.

Organogenesis
Organ formation.

Pluripotent Cell
A cell that is capable of giving rise to several of the cell types of an organism.

Receptors
Proteins that bind other molecules or ions, and, as a result of the binding, influence cellular activity.

Reproductive Cloning
Cloning for reproductive purposes; that is, producing new organisms that are genetically identical to the donor of the nuclear material.

Stem Cell
A cell that undergoes mitotic cell division to give rise to the same type of cell; at some point, stem cells leave the pool of mitotically dividing cells to begin a process of cell differentiation.

Therapeutic Cloning
Transferring nuclear material from a donor somatic cell into an enucleated egg cell and harvesting embryonic stem cells (ESCs) from the resulting blastocyst.

Transcription Factors
Proteins that attach to DNA at a gene regulatory site and by so doing influence the rate of transcription of a specific gene.

Developmental cell biology is that part of biology concerned with the consequences of what cells do in the context of developing organisms. Developmental biologists are concerned with events before fertilization, for example, gametogenesis, and after birth (or its equivalent in nonmammalian species), for example, metamorphosis and regeneration. The range of species in the purview of the developmental biologist covers, essentially, the entire living world – animals, plants, protists, fungi, and even prokaryotes. Using the human example, cell proliferation is required to create the trillions of cells of which we are composed; cell differentiation provides the more than 200 different kinds of cells identified so far in the human body, and morphogenesis shapes the body form characteristic of the human species. Our understanding of the molecular and cellular mechanisms that underlie these phenomena has been progressing at an ever-increasing rate, a testimonial to the army of cell and developmental biologists at work across the planet. Understanding the regulation of the cell cycle, apoptosis, cell differentiation, cell signaling, cell signal transduction, and cell movement, to mention a few such phenomena; understanding the integration of their underlying molecular and cellular mechanisms; and understanding the developmental consequences of their operation, makes the early twenty-first century an exciting time to be a developmental cell biologist. It is also a hopeful time for those people suffering the ravages of dreadful diseases, such as cancer, and for those people awaiting the promise of regenerative medicine.

1 Historical Perspective

Recently, we have been reminded that cell biology involves investigation of the molecular basis of how cells work, whereas developmental biology is often concerned with understanding how processes are integrated to make the organism. Furthermore, even though the boundaries between cell and developmental biology are once again beginning to blur, differences still exist between cell and developmental biology and we have not returned to a situation in which cell and developmental biology completely overlap.

1.1 Origins of Cell Biology

The credit for the Cell Theory, of the early nineteenth century (1838/1839), is given to Schwann and Schleiden. According to Churchill, Matthias Schwann's cell theory was, above all else, an embryological theory of the fine structure of the body; normal differentiation of tissues and organs followed a normal *Entwicklungsgeschichte* (developmental history) of the constituent cells; abnormal growths and neoplasias followed an abnormal cellular *Entwicklunsgeschichte*. It was no accident that both Schwann and Müller (Johannes Müller, German physiologist and comparative anatomist) quickly saw the implications of the cell theory for human pathology. During the middle of the same century, Rudolph Virchow laid the foundation for cellular pathology, replacing humoral pathology, which for centuries was purported to explain disease states. During the last quarter of the nineteenth century, the fundamental details of mitosis and meiosis at the light microscope level were accumulated.

1.2 Origins of Developmental Biology

The nineteenth century also saw the laying of the foundations of descriptive and comparative embryology, and the end of the same century provided the beginnings of experimental embryology.

Karl Ernst von Baer (1792–1876), an Estonian naturalist and pioneer embryologist, is considered by some to be the father of modern embryology. He first described the mammalian egg (1827), and, with his microscope, he could see that the mammalian egg was an undifferentiated mass containing no preformed creature. He believed that the mature egg, or the early developing embryo, possesses certain definite areas having different qualities, each of which contributes to the formation of a particular structure or several structures (1828, 1837). Von Baer became the first to trace the egg to the embryo.

In 1828, he published the *Epistle*, in which he originally set forth his discovery of the ovum and *Uber Entwickelungsgeschichte der Tiere* (Developmental History of Animals), which became a standard text of embryology. Four major advances in embryology were included in this book: (1) the discovery of the mammalian ovum, (2) the germ-layer theory, (3) the law of corresponding stages in the development of the embryo, and (4) the discovery of the notochord.

Also regarded as the father of comparative embryology, he was the first to make embryology truly comparative. Together with Pander (Christian Pander, Russian zoologist, regarded as a founder of the science of embryology), he formulated the germ layer concept as a structural fact for vertebrate embryology. Karl Ernst von Baer's *Development of Animals* (*Entwickelungsgeschichte der*

Tiere – Beobachtung und Reflexion) added the final stroke against the preformation doctrine; he showed differentiation to be a progressive process. Von Baer enriched embryology in three directions: (1) he set a higher standard for all work in embryology and lifted the science to a higher level, (2) he established the germ layer theory, and (3) he made embryology comparative. For a long time after von Baer, the aim of embryologists was to trace the history of the germ layers.

According to Hickman, the very tiny ova of mammals escaped de Graaf's (Regnier de Graaf, seventeenth-century Dutch embryologist, after whom the Graafian follicle is named) eyes, but von Baer brought mammalian reproduction in line with that of other animals by detecting the ova and their true relationship to the follicles. One of the first achievements of the greatest embryologist of all time was this important discovery.

Again, according to Hickman, von Baer developed the biogenetic law, which states that the embryos of higher and lower forms tend to resemble each other more closely the further one goes back in their development. In comparing corresponding stages in the development of different kinds of embryos, von Baer advanced four propositions, which together became incorporated in the biogenetic law: (1) in development, general characters appear before special characters, (2) from the more general characters are developed the less general and finally the special, (3) in the course of development, an animal of one species diverges continuously from that of another, and (4) a higher animal during development passes through stages that resemble stages *in the development of* lower animals. This is a conservative and sounder conception of the biogenetic law than that of Ernst Haeckel (German embryologist), because von Baer did *not* maintain that the embryos of higher forms resemble the *adults* of lower organisms. Von Baer did not relate these observations to evolution; Haeckel (1834–1919) made the connection and popularized the phrase "ontogeny recapitulates phylogeny."

Wilhelm Roux (1850–1924), a German embryologist and a student of Haeckel, is considered to be a pioneer in (and by some, the founder of) experimental embryology.

Roux visualized a complex machine already functional in the fertilized egg, which suggested preformation. To be transmitted from cell to cell, Roux reasoned, the machine must in some way break down into its component parts during the process of cell division so that the parts can be transmitted to another cell, and reassemble there in some way and again become functional.

On the basis of his experimental evidence, of rotating eggs in their gravitational field and finding that the eggs develop normally, Roux concluded that eggs are self-differentiating rather than being driven by external conditions. By 1885, Roux was generating a general theory about the causes of embryonic development based on his idea of self-differentiation. He suggested that the nucleus holds all of the qualities for individual formation offering a theory of qualitative cell division, that is, each division actually separates off differential nuclear materials into different daughter cells; he said, the process is like producing a mosaic.

In the early 1890s, E. B. Wilson worked with the German embryologist Hans Driesch at the Zoological Station in Naples. Wilson, Driesch, and others did not gravitate toward Roux's emphasis on nuclear division to explain development; rather, they continued to look at the internal structure and the patterns of cell division

for clues to the causes of embryonic differentiation. Wilson (1904) carried out isolation experiments with *Patella coerulea*, a mollusc, and found that the blastomeres possess within themselves all the factors that determine (1) the form and rhythm of cleavage and (2) the differentiation that they undergo, wholly independent of their relationship to the remainder of the embryo. He also carried out experiments (1904) where he removed polar lobes from *Dentalium*, a mollusc, and determined that its mosaic development was based on the segregation of specific morphogenetic determinants into specific blastomeres; for example, Wilson concluded that the polar lobe cytoplasm contained the mesodermal determinants. August Weismann came to conclusions similar to those of Roux, but went on to offer a larger theory of development. Weismann's theory was based on the assumption that physical hereditary units exist within the nucleus and he postulated a mechanism for the separation of those units.

Roux's famous experiments, the so-called half-embryo experiments (1888), which seemed to demonstrate preformation, were carried out by killing, with a hot needle, one of the two blastomeres of a frog embryo; only one half of an embryo developed and Roux concluded that determination had already occurred in the two-cell stage. He recognized that his view was the leading alternative to Eduard Pflüger's (German physiologist 1829–1910) hypothesis that external conditions cause differentiation.

The goal of Hans Driesch (1867–1941) was to reduce embryology to the laws of physics and mathematics. He carried out *isolation* experiments (1892) with sea urchin embryos; he isolated blastomeres from the 2-cell, 4-cell, and 8-cell stages and found that individual isolated blastomeres developed into complete larvae; this result was not expected from the early experimental work of Roux.

Roux inaugurated (1894) the physiological approach to embryology (*Entwicklungsmechanik*, developmental mechanics). He insisted that embryology would no longer be the servant of evolutionary studies, but, rather, embryology would assume its role as an independent experimental science. He believed that many of the conclusions drawn from the investigation of ontogeny (development) would throw light on the phylogenetic (evolutionary) process. Roux implied that he considered experimentation the only legitimate method for biological science. He founded his own new journal, *Wilhelm Roux's Archiv für Entwicklungsmechanik der Organismen* (*Roux's Archives of Developmental Biology*), in which he was noted for exercising a heavy editorial hand. Roux himself, in 1881 had, in *Der Kampf der Theile*, expressed the idea of mutual interaction between embryonic parts.

Hans Spemann (1869–1941), a German zoologist was the 1935 recipient of a Nobel Prize in Physiology or Medicine for his discovery of the organizer effect in embryonic development. He wrote the classic book, *Embryonic Development and Induction* (1938). According to Saha, although Spemann was not the first scientist to introduce the concept of induction into embryological thought, he, in the course of his lens induction studies, popularized and refined the concept, rendering it as one that might serve as a unifying principle for vertebrate development.

1.3 Relationship between Cell and Developmental Biology

Albrecht von Koelliker (1817–1905) traced the origin of tissues from the segmenting

ovum through the developing embryo. He encompassed the whole field of embryology and published, in 1861 and again in 1876, a general treatise on vertebrate embryology of high merit. He is said to have done more to establish the cell theory than anyone else; he was one of the first to recognize the ovum as a cell (1840, doctoral thesis) and the first to point out that spermatozoa were cells. He earned his doctorate with work that traced the development of spermatozoa from stem cells to their release from the testes (1841). His *Entwicklungsgeschichte* was the first book on comparative embryology; it was he who first applied the cell theory to embryology.

The German biologist, August Weismann (1834–1914), one of the great synthesizers of biology, in 1887, recognized that the doubling of the chromosome number that occurs in fertilization would lead to unmanageable numbers of chromosomes unless a means was at hand to counteract this buildup (meiosis, of course, is that process). He proposed the first testable alternative to Darwin's theory of pangenesis. In 1883, Weismann proposed the germplasm theory, which maintained that different nuclear determinants entered different cells, and only in the nuclei of germ cells were all types of determinants retained. His hypothesis proposed continuity of the germplasm, diversity of the somatoplasm, and that the germ line was totally independent of the somatic cells. E. B. Wilson claimed that his classic textbook grew out of Weismann's hypothesis.

Thus, by the dawn of the twentieth century, although not generally referred to as such at the time, cell biology and developmental biology were burgeoning disciplines of biology.

Early in the twentieth century, Edmund Beecher Wilson (1856–1939), an American zoologist, specializing in cytology and embryology, wrote *The Cell in Development and Heredity*, 3rd edition (1925). This book is correctly considered to be a classic in, at least, the fields of cytology (cell biology) and embryology (developmental biology). Testimonial to the classic nature of Wilson's book is the observation that it is still referenced early in the twenty-first century.

According to Gilbert, the experiments of Spemann and his students framed the questions that experimental embryologists asked for most of the twentieth century, and, further, the discoveries associated with these inductive processes have provided some of the most exciting moments in contemporary science.

In the editorial of the article, Where Cell Biology and Developmental Biology Meet, in the inaugural issue of the journal, *Developmental Cell*, **1** (1): 1, Siegel and Sweet propose that it really does seem as though we are beginning to understand how cells function, both individually and as part of multicellular organisms, and that there is growing interaction and overlap between cell biology and developmental biology.

The commentary by Dawes–Hoang and Wieschaus in the same issue of this journal (2001) suggests that the development of tissue culture techniques, initiated by Ross Harrison in 1910, enabled biologists to study cells *in vitro*, and that this further separated cytology from embryology and has remained an important difference between them. They further propose that, with a shift in interest from the structural composition of cells to the processes taking place within them and the molecular function of the components involved, cytology grew into the

field of cell biology; and with the success of genetic and molecular approaches coupled with a focus on the concepts of induction, patterning, and gene expression, embryology expanded into the intellectual discipline of developmental biology. They further propose that in areas where cell and developmental biology are successfully interfacing (e.g. the growth cone and signal transduction pathways), developmental biology is becoming increasingly more defined and molecular and cell biology has expanded to incorporate broader aspects of function for the proteins of interest.

Edward B. Lewis (USA) and Christiane Nüsslein-Volhard (Germany), and Eric Wieschaus (USA) were corecipients of the 1995 Nobel Prize for Physiology or Medicine for the discovery of genes that control the early stages of the body's development.

During the second half of the twentieth century, the merging of cell and developmental biology was reflected in the titles of books and journals (e.g. *Annual Review of Cell Biology* became *Annual Review of Cell and Developmental Biology, In Vitro* became *In Vitro Cell and Developmental Biology*), and, in 2001, Cell Press established a new journal, *Developmental Cell*.

Just as one endeavored to explain cellular phenomena at the molecular level, during the twentieth century it became increasingly clear that development needed to be explained at the cellular level, including not only intracellular signal transduction but also cell–cell communication and cell–extracellular matrix (ECM) interaction. Attempts to explain the mechanisms and control of cell proliferation, cytodifferentiation, embryo patterning, and morphogenesis necessitated molecular understanding.

2 Cell Activities Underlying Development

2.1 Intracellular Signal Transduction

Developmentally important cellular activities include cell division and differential gene expression; in a multicellular organism, these cellular activities, including signals from other cells, are under the control of the cell's environment. Although some of these signals, for example steroid hormones, are nonpolar and are able to pass through the plasma membrane of the cell, other signals are polar, for example, the hormone epinephrine, and attach to receptors in the plasma membrane. In order to convey information carried by the signal from the cell surface to the interior of the cell (e.g. the nucleus) where the information is acted upon, the cell makes use of cascades of chemical reactions that make up intracellular signal transduction pathways.

2.1.1 Receptors

Receptor proteins are proteins that bind other molecules or ions and as a result of the binding, influence cellular activity, for example, insulin receptors in the cell surface and progesterone receptors in the cytoplasm.

Androgen insensitivity syndrome is the result of a mutation in which XY individuals (chromosomally male) do not make a functional testosterone *receptor*. Testicular feminization is a result of a sex-linked mutation; tissues that normally respond to androgens (male hormones, primarily testosterone) do not and, as a consequence, affected individuals, although carrying the XY sex chromosome makeup, are phenotypically females.

EGF or epidermal growth factor receptors are tyrosine kinases and activate the MAP kinase pathway (an intracellular signal transduction pathway important in the control of cell growth and proliferation).

Enzyme-linked receptors are cell surface receptors (e.g. tyrosine kinases, serine/threonine kinases) found in plasma membranes, which upon binding their ligand (e.g. a paracrine factor) activate enzyme activity. These receptors are important in developmentally significant intracellular signal transduction pathways.

Ephrin proteins are the main proteins involved in *restriction* of neural crest–cell migration. Eph receptors are cell-surface receptors used by neural crest cells to recognize the ephrin proteins; binding of ephrin proteins activates the tyrosine kinase domains of the Eph receptors in neural crest cells; these kinases probably phosphorylate proteins that *interfere* with the actin cytoskeleton which is critical for cell migration.

The, now archaic, Fertilizin Theory of Fertilization held that the binding of fertilizin (from jelly coats of sea urchins) with antifertilizin (fertilizin receptors on the heads of sea urchin sperm) was the basis of species specificity in sea urchin fertilization.

Toll is a maternal gene in *Drosophila* that encodes a receptor protein, which is only activated in the future ventral region of the embryo, by a ligand, called the *spätzle fragment*, which is only produced in the ventral perivitelline space. Activation of the toll receptor generates a signal that causes a maternal gene product in the adjoining cytoplasm, the dorsal protein encoded by the *dorsal* gene, to enter nearby nuclei; this dorsal protein is a transcription factor with a vital role in organizing the dorsoventral axis of the embryo.

Torso is a maternal gene in *Drosophila* that plays a key role in specifying the structures at the extreme ends of the anteroposterior axis of the embryo (e.g. the acron at the anterior end and the telson at the posterior end). *Torso* encodes a receptor protein, which is uniformly distributed throughout the egg plasma membrane, but is only activated at the ends of the fertilized egg because the protein ligand for the receptor is present there only. The ligand for the torso protein, which distinguishes the termini, is synthesized and secreted by both posterior and anterior follicle cells, but not by other follicle cells.

Osteoprotegerin or osteoclastogenesis inhibitory factor is a member of the TNF receptor superfamily; together with its ligand, it regulates conversion of a macrophage stem cell into an osteoclast.

Retinoic acid receptors (RARs) is a group of transcription factors, active only when they have bound retinoic acid; the RARs bind to specific enhancer elements in the DNA called *retinoic acid response* elements. Some of the *Hox* genes have retinoic acid response elements in their promoters. Retinoic acid, a derivative of vitamin A, is a small-molecule-inducing factor, which can enter cells freely by diffusion; its receptor is itself a transcription factor, which upon binding retinoic acid, translocates to the nucleus to activate its target genes. Retinoic acid is a morphogen-controlling pattern formation in chick and amphibian limbs; it is also likely to play a role in posteriorizing the neural tube; and it has been shown to activate the expression of more posterior *Hox* genes. Homeotic changes are seen when mouse embryos are exposed to teratogenic doses of retinoic acid. Human infant anomalies caused by inadvertent exposure to retinoic acid include absent or defective ears, absent or small jaws, cleft palate, aortic

arch abnormalities, thymic deficiencies, and abnormalities of the central nervous system. Inside the developing embryo, vitamin A and 13-*cis*-retinoic acid become isomerized to the developmentally active forms of retinoic acid, *all-trans*-retinoic acid, and 9-*cis*-retinoic acid.

The nuclear receptor superfamily consists of transcription factors that are also intracelluar receptors stimulated by lipophilic ligands including steroids, thyroid hormones, and retinoic acid; they have a hormone-binding domain, a DNA-binding domain, and a transcription activation domain. The receptor is normally complexed with a heat shock protein called hsp90 and thereby retained in the cell cytoplasm; on binding the hormone, the hsp90 is displaced, and the factors can form active dimers that move to the nucleus and bind to target genes. Examples include the retinoic acid receptors and the ecdysone receptors.

RU-486 (mifepristone) is a synthetic steroid that blocks the progesterone receptor from binding progesterone, which stops the uterine wall from thickening and prevents the implantation of a blastocyst; used as an alternative to surgical abortion.

Downregulation is a reduction in responsiveness of a cell to a stimulus following the first exposure: this is often produced by a reduction in the number of available receptors expressed on the surface, which can result from internalization of the ligand–receptor complex or from decreased expression of the receptor, for example, the cell-surface receptors for several growth factors are internalized following the growth factor binding and are eventually degraded in lysosomes. The effect of this process is to remove the receptor–ligand complexes from the plasma membrane terminating the response of the cell to growth factor stimulation. Classically, the concept referred to hormone receptors but contemporary usage includes other cell-surface receptors.

Upregulation is a process that increases ligand–receptor interactions due to an increase in the number of available receptors; for example, EGF-induced upregulation of cyclin D1 protein in NIH 3T3(M17) fibroblasts.

A morphogen receptor gradient is a gradient of those molecules that recognize the morphogen; the expression of the *Brachyury* and *goosecoid* genes has been correlated with the number of activin receptors on each cell that binds activin.

2.1.2 Intracellular Signal Transduction

Signal transduction is converting a signal from one form to another, for example, converting a cAMP signal to a kinase signal as an intracellular signal transduction pathway progresses into the interior of a cell.

The JAK/STAT signal transduction pathway is an example of an intracellular signal transduction pathway that is important in development. This pathway is extremely important in the differentiation of blood cells and in the activation of the casein gene during milk production. The STAT pathway is very important in the regulation of human fetal bone growth.

The phosphoinositol pathway is a common intracellular signal transduction pathway. During sea urchin fertilization, when bindin (a ligand on the surface of the spermatozoon) binds to the bindin receptor on the surface of the egg, a G-protein is activated. The G-protein, in turn, activates the enzyme phospholipase C, which hydrolyzes its substrate, phosphatidylinositol bisphosphate (PIP_2) into two products, diacylglycerol (DAG) and inositol triphosphate (IP_3). The DAG activates protein kinase C, which in turn, activates a H^+/Na^+

exchanger, which by simultaneously transporting H^+ out of the egg and Na^+ into the egg, causes alkalinization of the egg's cytoplasm. Meanwhile, the IP_3 binds to a receptor on the membrane of the endoplasmic reticulum (ER), causing the ER to release Ca^{++} into the cytosol of the egg. Together, alkalinization of the cytoplasm and elevated Ca^{++} concentration in the egg cytoplasm cause activation of the egg.

The RTK (receptor tyrosine kinase) signal transduction pathway is an intracellular signal transduction pathway important in development and was one of the first pathways to unite various areas of developmental biology. In migrating neural crest cells of humans and mice, the RTK pathway is important in activating the microphthalmia transcription factor (Mitf) to produce pigment cells.

Another intracellular signal transduction pathway important in development is the Smad signal transduction pathway. In vertebrates, this pathway appears to be activated by the TGF-ß superfamily ligand nodal in those cells responsible for the formation of the mesoderm and for specifying the left–right axis in vertebrates.

An intracellular signal transduction pathway used in different cells in different ways is the Wnt-signal transduction pathway; additionally, its components can have more than one function in the cell; it is involved in segment polarity in *Drosophila*, axis formation in *Caenorhabditis elegans*, and in axis formation in *Xenopus*.

The hedgehog signal transduction pathway is an important pathway in limb and neural differentiation in vertebrates.

Phytochromes are plant pigments that transduce photoperiodic signals from the external environment.

Cross-talk interactions occur between seemingly independent intracellular signal transduction pathways, for example, cooperation between two pathways resulting in the formation of the AP-1 transcription factor, which turns on interleukin 2 transcription during lymphocyte differentiation.

Homologous signal transduction pathways are composed of homologous proteins arranged in a homologous manner. Homologous pathways form the basic infrastructure of development; however, the targets of these pathways may differ among organisms, for example, the dorsal-cactus pathway used in *Drosophila* for specifying dorsal–ventral polarity is also used by the mammalian immune system to activate inflammatory proteins. When homologous developmental pathways are used for the same function in both protostomes and deuterostomes, they are said to have deep homology.

2.1.3 Transcription Factors

A transcription factor is a molecule that attaches to DNA at a gene regulatory site and by so doing influences the rate of transcription of a specific gene; it is a protein that regulates the transcription of genes, often, but not exclusively, by binding to cis-regulatory elements (promoters and enhancers). Transcription factors come in two general types: basal transcription factors (e.g. TFIID, TFIIA, etc.), which attach to gene promoters, and cell-specific transcription factors (e.g. Pax6, Mitf, etc.), which bind to gene enhancers. Trans-regulatory factors are usually transcription factors. There are four major families of transcription factors: homeodomain, basic helix–loop–helix (bHLH), basic leucine zipper (bZip), and zinc finger.

The trans-activating domain of a transcription factor is that domain concerned with activating or suppressing the gene's transcription; activation domains

of transcription factors often contain many acidic amino acids forming an "acid blob," which accelerates the formation of the general transcription complex.

A trans-regulatory factor is usually a soluble protein, produced at a gene distant from the chromosome region that it affects; they are usually transcription factors that bind to cis-regulatory elements (promoters and enhancers).

Transcriptional regulation is regulation of transcription of RNA from DNA, under the control of RNA polymerase, transcription factors, promoters, and enhancers. A transcription initiation complex is, collectively, the RNA polymerase and the associated basal transcription factors that attach to a gene promoter.

A battery is a group of genes regulated by the same transcription factor.

Zygotic genes are those genes expressed by the embryo itself, that is, the embryo's genes as opposed to maternal genes expressed in the embryo. In early *Drosophila* development, most of the zygotic genes, first activated along the anteroposterior and dorsoventral axes set up by maternal genes, encode transcription factors, which, then activate more zygotic genes.

Combinatorial regulation is the control of gene transcription by two or more transcription factors; the spatial patterns of gene expression are often delimited by the combined action of transcription factors.

Competence is the ability to respond to an inductive signal or the reactive state permitting development in response to a stimulus, as part of an embryo in response to an inductive signal, or the responsiveness to the signals that turn on the relevant combination of transcription factors.

Specification is the early part of commitment to cell fate, where the fate of a cell depends on environmental cues, but is flexible and can still be altered in response to environmental signals. Determination is the second part of commitment to cell fate; the fate of a cell has become cell-intrinsic and is irreversible; the cell's commitment to eventually differentiate into a specific cell type and no other (transdifferentiation or metaplasia is the transformation of one form of *adult* tissue to another, as seen for example in Wolffian regeneration). Blastomeres are determined by cytoplasmic segregation or by interactions among blastomeres. In a molecular sense, determination means that the cells have lost their responsiveness, or competence, to the signals that originally turned on the relevant combination of transcription factors.

The basic helix–loop–helix (bHLH) family is one of the four major families of transcription factors; transcription factors in this family function in muscle and nerve specification, for example, MyoD.

The homeodomain family is one of the four major families of transcription factors, including the Hox subfamily, which play an important role in axis formation, for example, Hoxa-1, Hoxb-2, and so on. The bicoid protein is a member of the homeodomain family of transcriptional activators and activates the *hunchback* gene by binding to regulatory sites within the promoter region.

The hunchback protein acts, in the *Drosophila* embryo, as a morphogen for the next stage of patterning. To establish a clear anteroposterior gradient of zygotic hunchback protein, nanoprotein inhibits translation of posterior maternal *hunchback* mRNA. The hunchback protein, itself a transcription factor, acts as a morphogen to which other gap genes respond.

The basic leucine zipper (bZip) transcription factors with a structure similar to that of bHLH proteins are dimers, each

of whose subunits contains a basic DNA-binding domain at the carboxyl end, followed closely by an alpha-helix containing several leucine residues. The two subunits of the protein interact through leucine-containing repeat sequences. CCAAT enhancer binding protein (C/EBP) is a bZip protein that can bind to the DNA sequence CCAAT; it plays a role in adipogenesis similar to that of bHLH proteins in myogenesis. The bZip family is one of four major families of transcription factors; transcription factors in this family function in liver differentiation and fat cell specification, for example, C/EBP, AP1, and so on.

The zinc finger family is one of the four major families of transcription factors and includes the nuclear hormone receptor subfamily; transcription factors in this family function in secondary sex determination, for example, glucocorticoid receptor, estrogen receptor, testosterone receptor, and so on. Zinc finger proteins constitute a distinct class of transcription factors in which a DNA-binding polypeptide loop or "finger" forms through a coordination complex of zinc with cysteine and histidine residues at the base of the loop. These transcription factors make up a large and diverse group of proteins, for example, Krüppel in the early *Drosophila* embryo, WT-1 in the kidney, and Krox20 in the rhombomeres of the hindbrain.

The *Pax* genes are a subfamily of homeobox genes in vertebrates, which contain a homeobox and another conserved domain known as paired; these genes encode transcription factors with various functions in development. *Pax6* is a transcription factor belonging to the homeodomain family, necessary for mammalian eye, nervous system, and pancreas development.

During the development of bilaterally symmetrical organisms, three axes arise: (1) the anteroposterior axis: in *Drosophila*, this axis is laid down in the egg, (2) the dorsoventral axis: in *Drosophila*, this axis is laid down in the egg, and (3) the proximodistal axis. In *Drosophila*, the anteroposterior axis becomes divided into a number of unique regions on the basis of the overlapping and graded distributions of different transcription factors; the patterning of the dorsoventral axis involves two gradients, one of the dorsal protein and one of the decapentaplegic protein, with high points at the opposite ends of the dorsoventral axis. Patterning along the dorsoventral and anteroposterior axes divides the embryo into a number of discrete regions, each characterized by a unique pattern of zygotic gene activity.

The anterior visceral endoderm is one of two signaling centers in the mammalian embryo; it works together with the node (Hensen's node, "the organizer") to form the forebrain; it expresses several genes necessary for head formation, including genes for transcription factors, *Hesx-1*, *Lim-1*, and *Otx-2*, and for the paracrine factor Cerberus. These two signaling centers in the mammalian embryo are responsible for creation of the whole body, for example, the node produces Chordin and Noggin and works together with the anterior visceral endoderm to form the forebrain.

Tbx4 is a transcription factor that appears to be critical in instructing chick limbs to become hindlimbs and Tbx5 is a transcription factor that appears to be critical in instructing chick limbs to become forelimbs.

Historically, the bicoid protein gradient provided the first reliable evidence for the existence of the morphogen gradients that had been postulated to control pattern formation. The bicoid protein is a transcription factor that acts as a morphogen;

it switches on certain zygotic genes at different threshold concentrations, thus initiating a new pattern of gene expression along the anteroposterior axis. The bicoid protein is a member of the homeodomain family of transcriptional activators and activates the *hunchback* gene by binding to regulatory sites within the promoter region. *Bicoid* mRNA is localized in the anterior end of the unfertilized egg of *Drosophila*; after fertilization, it is translated (bicoid protein is absent from the unfertilized egg) and the bicoid protein diffuses from the anterior end, forming a concentration gradient along the anteroposterior axis.

Developmentally important transcription factors have DNA binding sites by which they bind to specific sites on DNA.

ß-catenin, the major candidate for the factor that forms the Nieuwkoop center in the dorsal-most vegetal cells of the amphibian blastula, is a transcription factor that can associate with other transcription factors to give them new properties; the β-catenin/Tcf3 (a ubiquitous transcription factor) complex appears to bind to the promoters of several genes whose activity is critical for axis formation.

CBFA1 is a transcription factor that appears to be able to transform mesenchyme cells into osteoblasts; the *Cbfa1* gene is activated in mesenchymal cells by bone morphogenetic proteins.

2.2 Cell Signaling

Signaling between and among cells is necessary for the order, that is the organism, to emerge from the potential chaos, which is multicellularity. Signaling between cells of the anterior pituitary gland and the gonads orchestrates development of gametes, and signaling between cells may result in cell death known as *apoptosis*.

2.2.1 Juxtacrine Signaling

Juxtacrine signaling is a mode of cell–cell communication in which signaling molecules are retained on the surface of the signaling cell, and they interact with receptor proteins on adjacent cell surfaces, for example, interaction between bride of sevenless protein and its receptor sevenless.

Ephrins interact directly with the 14 known Eph receptors, but there is some specificity: ephrin-A ligands preferentially bind to EphA receptors and ephrin-B ligands to EphB receptors. The five ephrin-A ligands are anchored to the cell surface through a glycosyl phosphatidylinositol (GPI) moiety, whereas the three ephrin-B ligands are anchored through a transmembrane segment. Expression patterns of Eph receptors and ephrin ligands determine which ligands interact with which receptors in particular tissues. This in turn dictates the precise spatial arrangement of signal activation, given that receptor- and ligand-bearing cells must be next to each other for the ephrin/Eph receptor association to take place (juxtacrine signaling). The ephrin/Eph receptor interaction is unusual in other respects. Signals are propagated not only downstream of the Eph receptor but also downstream of the ephrin ligand. Through this bidirectional mode of signaling, two neighboring cells with distinct identities can reciprocally influence each other. Thus, for example, ephrin- and Eph receptor-bearing cells move away from each other to occupy distinct spatial domains and are prevented from intermingling, as seen in Fig. 1.

2.2.2 Autocrine Signaling

An autocrine growth factor is a growth factor that is made by the same cell (target

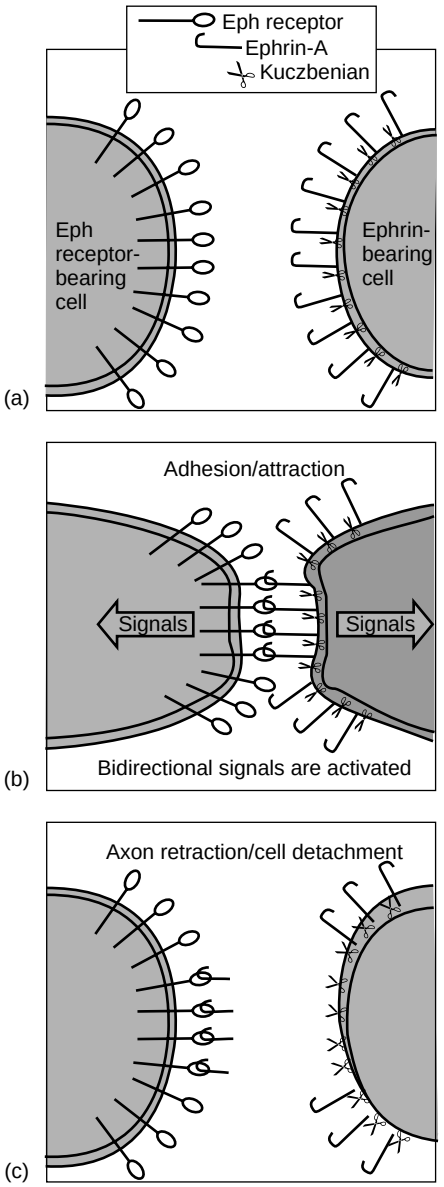

Fig. 1 Repulsive activities. (a) A cell bearing the Eph receptor comes into close proximity with an ephrin-bearing cell. (b) The receptor and ligand form multivalent adhesive complexes that transmit signals in both cells. (c) Some of these signals produce changes in the cytoskeleton that enable the two cells to pull apart after activation of the metalloprotease Kuzbanian and cleavage of ephrin. Pasquale (2000).

cell) that responds to it, for example, platelet-derived growth factor (PDGF). Transforming growth factor alpha (TGF-alpha) is an autocrine growth factor that stimulates development of the epidermis made by the basal cells and stimulates their own division.

It has been suggested that the HGF/SF (hepatocyte growth factor/scatter factor) could play a role in muscle development in an autocrine way. As a rule, HGF/SF is produced by mesenchymal cells, while its receptor, the tyrosine kinase encoded by the *met* proto-oncogene, is expressed by the neighboring epithelial cells in a paracrine fashion. They showed that both *HGF/SF* and *met* are coexpressed by undifferentiated C2 mouse myoblasts.

In growing cells, the autocrine loop is active as the receptor exhibits a constitutive phosphorylation on tyrosine that can be abrogated by exogenously added anti-HGF/SF neutralizing antibodies. The transcription of *HGF/SF* and *met* genes is downregulated when myoblasts stop proliferating and differentiate. The coexpression of *HGF/SF* and *met* genes is not exclusive to C2 cells since it has been assessed also in other myogenic cell lines and in mouse primary satellite cells, suggesting that HGF/SF could play a role in muscle development in an autocrine way. The autocrine loop needs to be downregulated during differentiation, since the constitutive activation of met kinase is incompatible with myogenesis.

The explosive proliferation of placental cytotrophoblast cells in response to PDGF, which these cells themselves produce, is an example of autocrine signaling.

2.2.3 Paracrine Signaling

Paracrine signaling is a mode of cell–cell communication in which signaling molecules (paracrine factors) act as local

mediators and only affect cells in the immediate environment of the signaling cell. These factors are the inducing factors of the classical experimental embryologists. Most paracrine factors fall into one of the four major families: Hedgehog, Wnt, TGF-ß, fibroblast growth factor (FGF). Keratinocyte growth factor (KGF), also known as FGF 7, is a growth factor needed for epidermal production, a paracrine factor produced by fibroblasts of the underlying dermis.

Stem cell factor (SCF) is a paracrine protein that promotes cell division in numerous stem cell populations; mice lacking SCF or its receptor (c-Kit protein) are sterile (no germ cells), white (no pigment cells), anemic (no red blood cells), and immunodeficient (no lymphocytes).

According to Pescovitz et al., spermatogenesis is a complex process that is both remarkable and enigmatic. While circulating hormones clearly play an important role in initiating and regulating the process, the Sertoli cell barrier prevents most substances from entering the seminiferous tubule compartment and directly influencing germ cell development. Therefore, the testis cannot rely solely upon the delivery of circulating hormones, nutrients, and growth factors, but must independently produce its own regulatory substances. A rapidly increasing number of testicular factors that appear to modulate spermatogenesis in a paracrine fashion are now being identified. These discoveries are beginning to contribute to our understanding of the intricate network of testicular cell–cell interactions that control male reproduction.

2.2.4 Endocrine Signaling

Endocrine signaling is a mode of cell–cell communication in which signaling molecules (endocrine factors/hormones) are released into the circulatory system and may affect cells (target cells) that are some distance from the signaling cell, for example, the effect of anterior pituitary gland hormones on the gonads. Endocrine factors are molecules (hormones) that function in endocrine signaling, for example, estrogens, testosterone, and progesterone. Cells that produce hormones are called endocrine cells and are often found in the endocrine glands, for example, pituitary glands, thyroid glands, and adrenal glands. Endocrine disrupters are exogenous chemicals that interfere with the normal function of hormones, for example, diethylstilbesterol.

2.2.5 Growth Factors

Growth factors are extracellular polypeptide signal molecules that can stimulate a cell to grow or proliferate. In the presence of growth factors, retinoblastoma protein (Rb), one factor maintaining the G_0 state of the cell cycle, becomes phosphorylated and hence deactivated. In the absence of Rb, a transcription factor called E2F becomes active and initiates a cascade of gene activation culminating in the resynthesis of cyclins, Cdks, and other components necessary to initiate the S-phase of the cell cycle.

Transforming growth factor ß (TGF-ß) is a family of protein growth factors, including the activins, bone morphogenetic proteins (BMPs), decapentaplegic protein (in *Drosophila*), and Vg1 protein (in *Xenopus*); their actions include potentiating or inhibiting responses of most cells to other growth factors, regulating the differentiation of some cell types, and acting as inductive signals in embryonic development. TGF-ß was originally discovered as a mitogen secreted by "transformed" (cancer-like) cells.

Activin is a member of the TGF-ß family of growth factors; it plays a key role in the induction of dorsal mesoderm in *Xenopus* embryos and is secreted by vegetal cells. Cytokines are growth factors that regulate blood cells and lymphocytes.

The cell-surface receptors for several growth factors are internalized following growth factor binding and are eventually degraded in lysosomes; the effect of this process is to remove the receptor–ligand complexes from the plasma membrane terminating the response of the cell to growth factor stimulation.

Apoptosis may be initiated by a withdrawal of growth factors from the cell or by an active response to a signal.

Fibroblast growth factor is a family of protein growth factors, first identified as mitogens for fibroblasts in tissue culture. FGF stimulates proliferation of many cell types, inhibits differentiation of various types of stem cells, and acts as inductive signals in embryonic development; for example, it plays a key role in the induction of ventral mesoderm in *Xenopus* embryos; it is secreted by vegetal cells.

The G_1 checkpoint is the checkpoint normally controlled by growth factors. In the absence of growth factors, the cell will leave the cycle and enter a state called G_0 in which Cdks and cyclins are absent.

The developmental pathway taken by the descendants of a pluripotential stem cell depends on which growth factors it is exposed to, which in turn is determined by the stromal cells to which it is exposed. Stem cells in the spleen become predominantly erythroid while those in the bone marrow become predominantly granulocytic. Stromal cells of the bone marrow create hematopoietic inductive environments (HIMs) by their ability to bind hematopoietic growth factors; without these growth factors, the stem cells die.

Insulin-like growth factor I (IGF-I) is a protein growth factor that promotes cell survival, stimulates cell metabolism, and collaborates with other growth factors to stimulate cell proliferation.

Platelet-derived growth factor is a protein growth factor that stimulates proliferation of connective tissue cells and some neuroglial cells; in angiogenesis, it is necessary for the recruitment of pericyte cells (mesenchymal cells found around a capillary). The explosive proliferation of placental cytotrophoblast cells in response to PDGF, which these cells themselves produce, is an example of autocrine signaling.

2.3 Cell–Cell Interaction

Cells may interact with each other in a variety of ways. In addition to cells interacting through cell signaling of the juxtacrine, autocrine, paracrine, and endocrine varieties, cells may make physical contact with each other through a variety of cell junctions, that is, tight junctions, gap junctions, and desmosomes.

2.3.1 Cell Junctions

Tight junctions are a type of adhering junctions typically found near the apical surfaces of epithelial cells. In addition to compartmentalizing parts of the multicellular animal body, tight junctions also create the necessity of transcytosis (movement of substances through, rather than around, cells) and by doing so, enable cells to regulate the passage of substances from one compartment into another.

Gap junctions are a type of communicating junction arising from organization of connexin proteins into conduits through which the cytoplasmic matrices

of neighboring cells establish continuity. Ions and small molecules may readily pass through these proteinaceous conduits, and provide the opportunity for metabolic coupling of the participating cells.

Gap junctions are collections of transmembrane channels that directly link the cytosols of two adjoining cells. They are present in almost all cell types of multicellular tissues in organisms ranging from mesozoa to humans, making them arguably the most ubiquitous and evolutionarily ancient form of intercellular communication in the animal kingdom. Hertzberg and coauthors (all leaders in the field) review recent progress in the identification and molecular characterization of the constituents of gap junctions, the determination of how gap junction channels are assembled and gated, and the elucidation of the myriad roles gap junctions play *in vivo*.

Desmosomes are like rivets holding adjacent epithelial cells together and, by doing so, provide the epithelium with structural integrity.

In a review article, Knust and Bossinger pointed out that although in recent years we have enormously increased our knowledge of the composition and formation of intercellular junctions in the epithelia of various organisms, we are still far from fully understanding the mechanisms by which the complex architecture of epithelial cells, or even the relatively simpler polarized organization of nonepithelial cells, is established and preserved. However, genetic and reverse genetic approaches in model organisms and extensive cell biological studies will be important in dissecting the sequence of events that promote the polarization of a cell. Complemented by data obtained from targeted gene inactivation experiments in the mouse, we will expand our knowledge of epithelial development and function. Additional information obtained from the analysis of other polarized cell types will further our knowledge of the process. Neurons, which transmit signals in one direction, migrating cells, and egg cells or embryonic blastomeres, which undergo asymmetric divisions, all rely on a polarized phenotype to fulfill their specialized tasks. The basic mechanisms used to establish and maintain polarization may be conserved and the same protein complexes may be used repeatedly to build different sorts of polarized cells.

2.3.2 Cell Signals

A pathway in the context of cell–cell signaling consists of the components required for sending, receiving, and transduction of a signal, including one or more ligands, membrane-associated receptors, intracellular signal transducers, and transcription factors.

2.4 Cell–Matrix Interaction

Growing, quiescent, and dying cells may all be found within the same microenvironment during morphogenesis. As it turns out, pattern formation is driven by differentials in cell growth and viability. The results of Chen et al. suggest that living cells can filter the same set of chemical inputs (activation of integrin and growth factor receptor signaling) to produce different functional outputs (growth vs apoptosis) as a result of local mechanical deformation of the cell or nucleus. By sensing their degree of extension or compression, cells may be able to monitor local changes in cell crowding or the extracellular matrix (ECM); (e.g. due to enhanced ECM remodeling or local application of cell tension) and thereby

couple changes in ECM extension to expansion of cell mass within the local tissue microenvironment. On the other hand, involution of tissue may be promoted in other microenvironments by inducing rapid breakdown of ECM and associated cell retraction. During malignant transformation, progressive loss of shape-dependent regulation may also lead to cell survival in the absence of ECM extension, unrestricted mass expansion, and hence neoplastic disorganization of tissue architecture.

2.4.1 Integrin Signaling

Cells reside in a protein network, the ECM, which they secrete and mold into the intercellular space. The ECM exerts profound control over cells. The effects of the matrix are primarily mediated by integrins, a family of cell surface receptors that attach cells to the matrix and mediate mechanical and chemical signals from it. As Giancotti and Ruoslahti point out, these signals regulate the activities of cytoplasmic kinases, growth factor receptors, and ion channels, and control the organization of the intracellular actin cytoskeleton. Furthermore, many integrin signals converge on cell cycle regulation, directing cells to live or die, to proliferate, or to exit the cell cycle and differentiate, as seen in Fig. 2.

Most integrins recognize several ECM proteins, and individual ECM proteins (e.g. fibronectin, laminins, collagens, and vitronectin) bind to several integrins. Integrins can signal through the cell membrane in either direction – extracellular binding activity of integrins is regulated from the inside of the cell, while the binding of the ECM elicits signals that are transmitted into the cell.

Adherent cells must be anchored to an appropriate ECM to survive; depending partly on signals from the matrix, they either proliferate or exit the cell cycle and differentiate. This anchorage requirement is lost in neoplastic cells. See Fig. 3.

The cytoplasmic domains of integrins always lack enzymatic activity, and thus, they transduce signals by associating with adaptor proteins that connect the integrin to the cytoskeleton, cytoplasmic kinases, and transmembrane growth factor receptors. Integrin signaling and assembly of the cytoskeleton are intimately linked. As integrins bind to ECM, they become clustered in the plane of the cell membrane and associate with a cytoskeletal and signaling complex that promotes the assembly of actin filaments. The reorganization of actin filaments into larger stress fibers, in turn, causes more integrin clustering, thus enhancing the matrix binding and organization by integrins in a positive feedback system (Fig. 2). As a result, ECM proteins, integrins, and cytoskeletal proteins assemble into aggregates on each side of the membrane. Well-developed aggregates detectable by immunofluorescence microscopy are known as *focal adhesions and ECM contacts*. Thus, integrins serve as integrators of the ECM and the cytoskeleton, as seen in Fig. 4, the property for which integrins are named.

2.4.2 Cell Movement

Cell shape changes and cell movements underlie the morphogenetic movements that create the form of the developing organism. Examples of the burgeoning literature on this topic are given in the published work of Keller, Adler, and Hall.

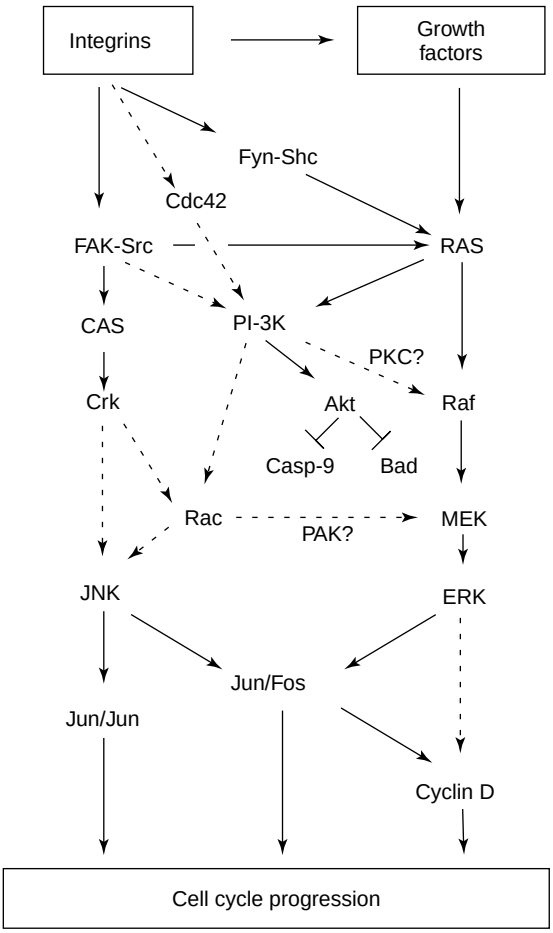

Fig. 2 Major signaling pathways that are known (solid arrows) or presumed (dashed arrows) to be coordinately regulated by integrins and growth factor receptors. These pathways control immediate-early gene expression, the cell cycle machinery, and cell survival. Giancotti and Ruoslahti (1999).

Polarized cell movements shape the major features of the vertebrate body plan during development. The head-to-tail body axis of vertebrates is elongated in embryonic stages by "convergent extension" tissue movements. During these movements, cells intercalate between one another transverse to the elongating body axis to form a narrower, longer array. Recent discoveries show that these polarized cell movements are controlled by homologs of genes that control the polarity of epithelial cells in the developing wing and eye of *Drosophila*.

A major breakthrough is the recent discovery that homologs of genes controlling the polarity of hairs on the epidermal cells of *Drosophila* wings also control the polarized cell motility underlying the morphogenic movements that shape the vertebrate body plan. These movements, known as *convergence and extension* or *convergent extension*, narrow (converge) the mediolateral aspect and elongate (extend) the anterior–posterior aspect of the vertebrate embryo, as seen in Fig. 5, and thereby establish its morphological and functional polarity, with

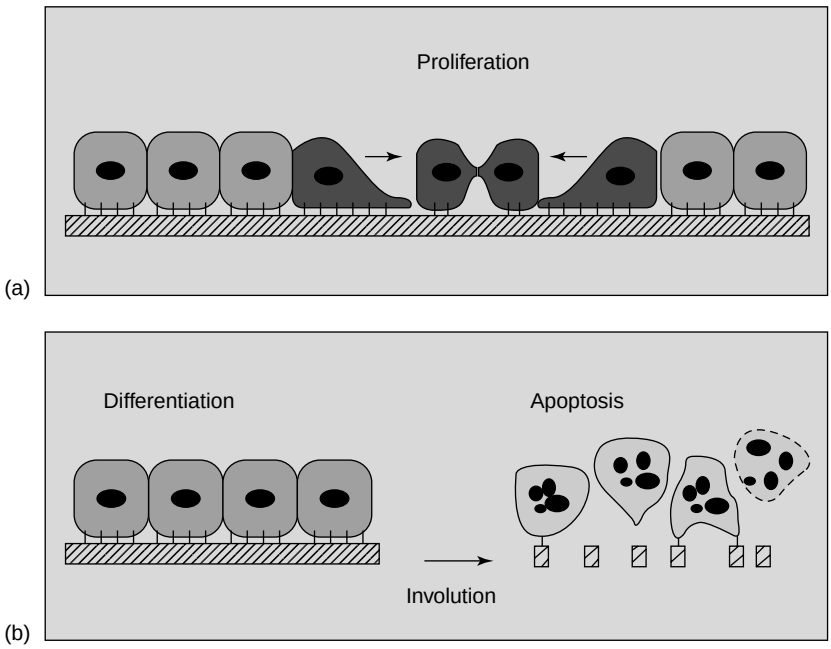

Fig. 3 Cell survival and cell proliferation require interaction with the extracellular matrix. (a) Epithelial cells in some tissues, such as skin and gut, are continuously renewed from stem cells that rest on a basement membrane. Neighboring cells migrate into the space left empty by cells that have moved away to differentiate. (b) Certain epithelia, such as those of the mammary gland and prostate, are not continuously renewed. In this case, interaction with the matrix appears to promote differentiation. During involution, the basement membrane is dissolved by proteolysis, and the cells undergo apoptosis. Giancotti and Ruoslahti (1999).

a head on one end and a tail on the other.

The polarity of cells within the plane of the epithelium of *Drosophila* is regulated by the planar cell polarity (PCP) pathway. This pathway controls the polarity of the single hair on each epidermal cell of the fly wing and the polarity of ommatidia in the compound eye.

Elements of the PCP pathway controlling epithelial cell polarity in *Drosophila* also regulate the polarized cell motility that shapes the body plan in vertebrates. This finding provides a genetic and molecular basis for a polarized cell behavior underlying the development of polarity at the level of the organism and thereby advances our understanding of the genetic encoding of the three-dimensional form. The cell biological mechanisms of how elements of the PCP pathway interact to build a form-generating machine are yet to be resolved.

The actin cytoskeleton mediates a variety of essential biological functions in all eukaryotic cells. In addition to providing a structural framework around which cell shape and polarity are defined, its dynamic properties provide the driving force for cells to move and to divide. Understanding the biochemical mechanisms

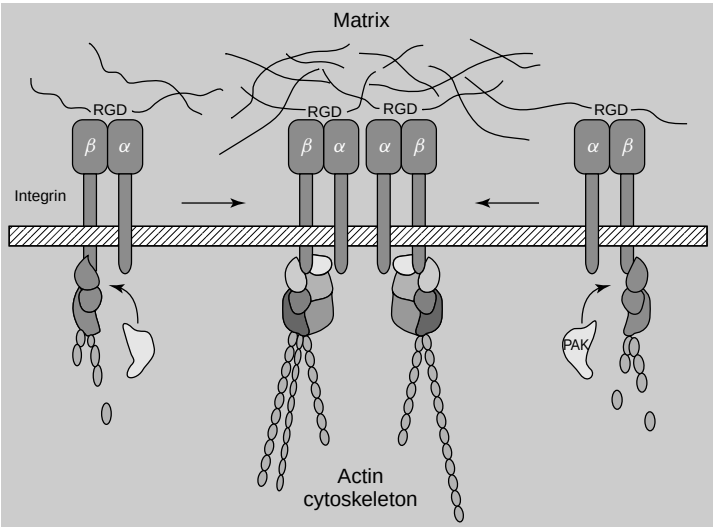

Fig. 4 Matrix binding promotes integrin clustering and association with the cytoskeleton. This in turn promotes further integrin clustering and matrix organization in a positive feedback system. RGD, Arg-Gly-Asp integrin-binding motif; Tal, talin; Pax, paxillin; Vin, vinculin; CAS, p130CAS. Giancotti and Ruoslahti (1999).

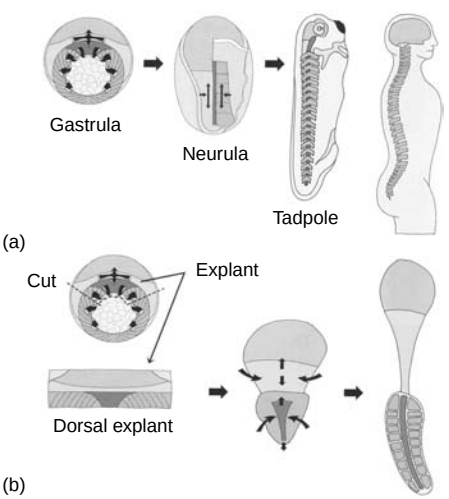

Fig. 5 Convergent extension movements elongate the anterior–posterior axis of the vertebrate body plan. (a) The notochordal (red) and somitic (pink) tissues turn inside, and converge (narrow) and extend (lengthens) in the gastrula and neurula stages of the frog embryo. The overlying presumptive hindbrain and spinal cord (blue) tissues converge and extend coordinately but on the outside of the embryo. These movements push the head away from the tail and elongate the body axis of the tadpole. Similar movements elongate the body axis of mammals. (b) Cultured explants of the same tissues also converge and extend, showing that convergent extension movements are driven by internal forces. Keller (2002). (See color plate p. xxiii.)

that control the organization of actin is thus a major goal of contemporary cell biology, with implications for health and disease. Members of the Rho family of small guanosine triphosphatases have emerged as key regulators of the actin cytoskeleton, and furthermore, through their interaction with multiple target proteins, they ensure coordinated control of other cellular activities such as gene

transcription and adhesion. Observations suggest that members of the Rho GTPase family are key regulatory molecules that link surface receptors to the organization of the actin cytoskeleton. See Fig. 6.

The JAK/STAT pathway is required in an unusually broad set of developmental decisions, including cell proliferation, cell fate determination, cell migration, planar polarity, convergent extension, and immunity. There is increasing evidence that the versatility of this pathway relies on its cooperation with other signal transduction pathways.

In a review, the components of the JAK/STAT pathway in model organisms and what is known about its requirement in cellular and developmental processes are discussed. In particular, they emphasize recent insights into the role that this pathway plays in the control of cell movement. In the past few years, studies from model organisms have established that the

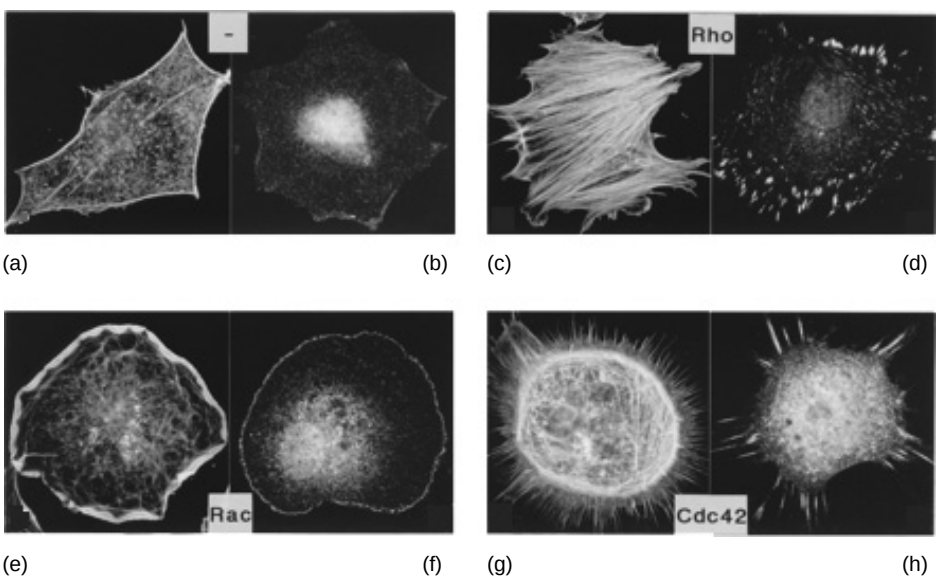

Fig. 6 Rho, Rac, and Cdc42 control the assembly and organization of the actin cytoskeleton. (a) Quiescent, serum-starved Swiss 3T3 fibroblasts (-) contain very few organized actin filaments or (b) vinculin-containing integrin adhesion complexes. The effects of Rho, Rac, or Cdc42 activation in these cells can be observed in several different ways such as with the addition of extracellular growth factors, microinjection of activated GTPases, or microinjection of guanosine diphosphate (GDP)–guanosine triphosphate (GTP) exchange factors. (c) Addition of the growth factor lysophosphatidic acid activates Rho, which leads to stress fiber and (d) focal adhesion formation. (e) Microinjection of constitutively active Rac induces lamellipodia and (f) associated adhesion complexes. (g) Microinjection of FGD1, an exchange factor for Cdc42, leads to formation of filopodia and (h) the associated adhesion complexes. Cdc42 activates Rac; hence, filopodia are intimately associated with lamellipodia, as shown in (g). In (a), (c), (e), and (g), actin filaments were visualized with rhodamine phalloidin; in (b), (d), (f), and (h), the adhesion complexes were visualized with an antibody to vinculin. Scale: 1 cm = 25 μm. (Figure courtesy, Kate Nobes).

JAK/STAT pathway is a central component of the signal transduction network that controls cell proliferation, fate, and movement. The outcome of JAK/STAT signaling in development depends on its collaboration with other signals and/or pathways. For example, during cell migration, the JAK/STAT signal may connect to the Rac/cytoskeleton pathway, and in the establishment of planar polarity, the JAK/STAT pathway cooperates with other pathways such as Wnt and N to produce a secondary signal.

Hou et al. ask the question, how can the same JAK/STAT pathway be used to produce completely different developmental outcomes in different tissue contexts? They then respond that the answer to this question will require an understanding of the array of transcription factors and molecules expressed in specific cells (history of the cell) as well as a description of the multiple signals received by a single cell (combinatorial signaling). Characterization of the genes that are regulated by JAK/STAT as well as the identification of signaling pathways that cooperate with JAK/STAT signaling are promising avenues toward understanding how this pathway can be involved in such a wide array of functions. See Fig. 7.

2.4.3 Tubulogenesis

Many organs of the body, for example, kidneys and testes, have a tubular nature. It is not surprising, therefore, that a burgeoning area of developmental cell biology would be the origin of tubulogenesis. The following information is taken from a recent review of tubulogenesis, in which specific references may be found.

Regarding the cellular events in tube formation, initial investigations of the cell biology of tubulogenesis suggested that the cellular mechanisms of tube formation might be as diverse as the gross differences in morphogenesis implied.

One of the earliest and most intensively studied systems, primary neural tube formation in chicks, pointed to the importance of cell shape changes. The first morphological event is the elongation of cells in the apicobasal dimension, with subsequent apical narrowing and basal expansion to create wedge-shaped cells that promote epithelial bending and wrapping. This and other investigations of epithelial morphogenesis led many to focus on cytoskeletal and cell shape changes in tubulogenesis. However, studies of other systems suggested a role for cell death in creating a luminal cavity, and early studies of thyroid follicle cells and blood vessel endothelial cells noted a possible role for cytoplasmic vesicles in lumen formation.

Tubulogenesis processes appeared to have little in common mechanistically. This view began to change over the past few years as modern cell biological tools were applied to cell culture systems in which the cellular events in tube formation could be carefully analyzed. Results from a number of systems, like the MDCK and endothelial cell culture systems converged on a critical role for apical membrane biogenesis and vesicle fusion in creating a lumen. During this same period, studies of tube growth and maturation in several invertebrate and vertebrate organs, including respiratory and renal organs, also directed attention to morphological and molecular events at the apical surface.

The concept of a common pathway of tubulogenesis began to emerge. Despite the wide variety of tubulogenesis mechanisms, emerging descriptions of the cellular and molecular events of tubulogenesis in a number of diverse systems have converged on apical membrane

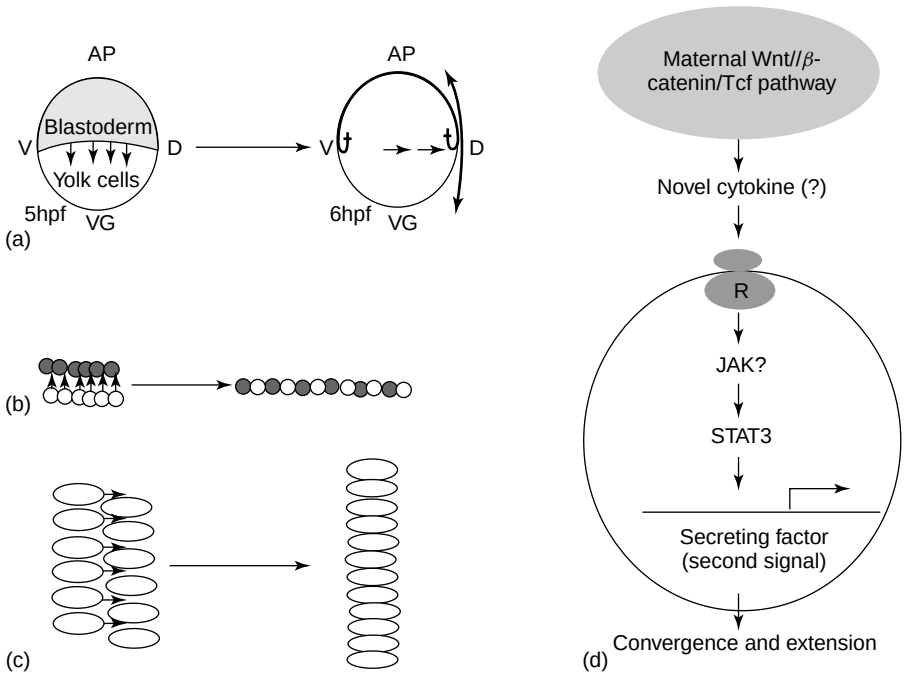

Fig. 7 Gastrulation movements in zebrafish. (a) During the early zebrafish development, four cell movements occur. First, blastomeres move down the surface of the centrally located yolk cell from the animal pole toward the vegetal pole (green arrows), a process called *epiboly*. Shortly after 6 h of development, 3 additional cell movements begin. Internalization (blue arrows) results in the formation of a localized thickening of cells known as the *germ ring*. Convergence of cells (red arrows) toward the future dorsal side of the embryo creates the embryonic shield, and extension (orange arrows) causes an elongation of the embryonic shield. (b) Epiboly is a radial intercalation process; deeper cells move up and insert themselves between more superficial neighbors, which leads to the spreading and thinning of the tissue. (c) Cells intercalate along the D/V axis during the dorsal convergence of lateral mesendodermal (LME) cells. (d) An activation of STAT3 on the dorsal side is regulated by the maternal Wnt/-catenin pathway and requires a Tcf transcription factor. An unidentified upstream molecule, possibly a cytokine or growth factor capable of activating STAT3 in early embryogenesis, is proposed to be one of the targets of the maternal Wnt/-catenin pathway. No JAK has yet been shown to be involved in STAT3 activation. The JAK/STAT signal may regulate the expression of a second signal molecule in prechordal mesendodermal (PCME) cells. The LME cells possibly sense the local gradient of the second signal and converge dorsally. AP, animal pole; VG, vegetal pole; V, ventral; D, dorsal; hpf, hr postfertilization. Hou et al. (2002). (See color plate p. xxiv.)

biogenesis, secretion, and other events at the apical surface as critical steps in tube formation and growth. Although the full process is not understood for any system, and the genes and molecules involved are just beginning to be characterized, the available data allow us to piece together a general mechanistic framework for tubulogenesis (see Fig. 8).

One of the surprising findings of early descriptions of tubular organ development is that, like chick neural tube formation, the tubes comprising some organs are not all created in the same way. The

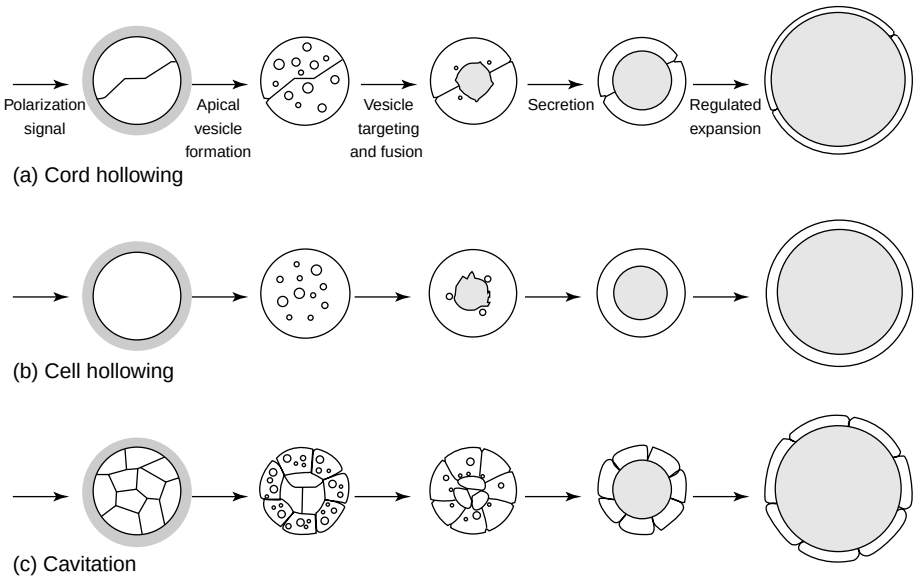

Fig. 8 A common pathway of tubulogenesis. Cells about to undergo (a) cord hollowing, (b) cell hollowing, or (c) cavitation follow a similar series of cellular events to create tubes. After receiving a polarization signal (green shading) that sets apicobasal polarity, the cells establish a basal surface (green) and generate vesicles carrying apical membrane antigens (red circles). The vesicles are targeted to the prospective apical region, where they fuse with the existing membrane or each other to form a lumen. Continued vesicle fusion and apical secretion expands the lumen. Expansion is a regulated process that continues until it is inhibited by a tube-size sensor set to trigger at a specific value for each tube. The sensor feeds into apical membrane biogenesis and secretory controls, as well as mechanisms for building apical structural support. During cavitation, interior cells fail to receive the polarizing signal and are left within the developing lumen and subsequently eliminated by cell death. Lubarsky and Krasnow (2003). (See color plate p. xxi.)

best-known example is the blood vessel development, where some vessels arise by the cord hollowing mechanism described above and others form by budding from the walls of existing vessels or by extensive remodeling of existing vessels. Similarly, in the mammalian kidney, the collecting ducts and distal tubules arise by sprouting from existing tubes or sacs, whereas proximal tubules arise from condensations of mesenchymal cells that assemble into tubes *de novo* and connect to the existing network. In the *Drosophila* tracheal system too, different mechanisms are employed at different stages of branching to generate the tubular network. If each of these mechanisms of tubulogenesis is not fundamentally different but represents a distinct variation on the general tubulogenesis process, then it is easier to envision how such tubular networks evolved: by the emergence of new variants of the tubulogenesis pathway that create tubes of specific sizes or shapes or different functional characteristics within a network. And, if it were easy to generate such variants during evolution, this would help explain how differences arose among related species in the tubulogenesis processes that form the neural tube and other organs such as the heart.

Although we are beginning to pinpoint some of the critical cellular and molecular events of tubulogenesis and can formulate a speculative general model for the process, we expect rapid progress in this area from a large number of systems using a variety of experimental approaches. A field of tube morphogenesis is emerging that not only seeks to answer the fundamental questions of how tubes are made and shaped but also how these processes go awry in human diseases associated with tube structural defects. Further investigations of the model systems and exploration of human diseases will continue to provide insights into the molecules and mechanisms of tube morphogenesis. These may in turn suggest ways to repair tubular defects or engineer better tubes.

3 Cell Differentiation

As the field of cell differentiation has generated vast literature, a model of cell differentiation, the role of pRb (retinoblastoma protein) in differentiation will be considered here.

Cell differentiation is characterized by permanent withdrawal from the cell cycle. Evidence for a role of pRb in cell differentiation came from Rb1 knockout mice ($Rb^{-/-}$), which die *in utero*. The correlation between death of $Rb^{-/-}$ mice and tissue-specific peak levels of pRb in normal mice suggested an essential role of pRb in development.

The expression levels of pRb protein in early mouse development are highest in brain, blood-forming cells of the liver, megakaryocytes, and skeletal muscle. Loss of pRb interferes with differentiation and can reactivate the cell cycle even in terminally differentiated cells.

The active role of pRb in differentiation has been studied in a number of cell lines inducible for differentiation. In each instance, an early event in differentiation was the dephosphorylation of pRb, which correlated with cell cycle arrest in G_0/G_1, a prerequisite to enter the differentiation pathway.

During differentiation of muscle cells, pRb accumulates in the nucleus and forms complexes with muscle-forming transcription factors such as MyoD and myogenin, thus preventing pRb rephosphorylation and locking the cell in the differentiated state. However, inactivation of pRb reverses the differentiated phenotype and allows cells to reenter the cell cycle.

The role of the E2F transcription factors is important in exiting the cell cycle and entering the differentiation pathway; these transcription factors are targets of pRb during growth suppression. When HL60 cells are induced to differentiate the ratio of pRb/p130 to the total E2F pool, shifted to a surplus of pRb/p130, resulting in complete sequestration of all E2Fs. In addition to E2F-1 to -3, E2F-4 was shown to bind pRb under these circumstances suggesting that pRb levels might be critical for the decision to enter S-phase or undergo differentiation. Overexpression of E2F-1 blocked terminal differentiation of megakaryocytes resulting in cell accumulation and apoptosis.

The role of p107 in differentiation seems to be more controversial than that of pRb. During myocyte differentiation, the expression levels usually go down, but in $pRb^{-/-}$ myoblasts, p107 is able to bind MyoD and compensate for myogenic differentiation. However, p107 is unable to hold the differentiated phenotype, since upon serum stimulation, it is downregulated and the cells reenter the cell cycle, an effect not observed in $pRb^{+/+}$ cells.

4
The Cell Cycle and Development

The relationship between the cell cycle and development is a topic that is of great interest to developmental cell biologists. Since the classic cell cycle, G1, S, G2, M was worked out in the 1950s it has been intensively studied by cell biologists. Also, it has long been known that the cell cycle of embryos undergoing cleavage is not classical, but is an abbreviated cycle consisting of S alternating with M. This embryonic (before gastrulation) cell cycle, using maternal gene products and not providing time (G1) for cell growth, results in cleavage of the zygote into an ever-increasing number of blastomeres of progressively decreasing size, that is, multicellularity is achieved at the expense of cell size. The midblastula transition of early *Xenopus* development has been used to study the conversion of S/M to G1/S/G2/S.

In a recent review of a book (Bock et al. (2001)) on the cell cycle and development, much of the current knowledge in developmental control of the cell cycle on many levels and in a variety of organisms is summarized, and is the prime source of the following information. The symposium on which the book is based is concerned with exploring the interplay between the cell cycle and the developmental programs employed by various organisms.

How and why does the "standard" cell cycle change in developing organisms, and what are the mechanisms that bring about these changes? How and why is the progression through the major cell cycle phases affected? What influences a cell's decision to stop proliferating and to differentiate at a certain point in the developmental program, and how is differentiation achieved?

Cell cycle alteration and regulation during development is a fundamental topic in considering the relationship between the cell cycle and development.

Xenopus oocytes and early embryos are excellent systems for studying the relationship between the cell cycle and development. Oocyte maturation is a hormonally triggered developmental switch that involves precise controls on M-phase entry and exit. *Xenopus* embryos bracketing the midblastula transition provide a developing system for the study of the sequential establishment of cell cycle checkpoints and the connection of signaling pathways to regulation of cell cycle progression(Fig. 9).

Further work in the *Xenopus* system should continue to provide fundamental insights into the interface between cell cycle control and developmental decisions.

Studying mechanisms that drive cell cycle transitions in developing *Xenopus* have included paying special attention to *Xenopus* oocyte maturation, as well as to the development of checkpoint control of the cell cycle. Studying the effect of DNA damage in the early embryo reveals that DNA damage has no effect on early *Xenopus* embryos, but when damaged embryos reach the onset of gastrulation, rapid and synchronous apoptosis occurs, implying that the DNA damage checkpoint is activated at this stage. Studies of early mouse embryos have focused on cell cycle transitions undergone during the first mitotic cleavages.

The degree to which the cell cycle can be altered to serve the changing needs of the developing organism is provided by studying *Drosophila* development. During *Drosophila* embryonic development, cell cycle progression is mediated exclusively by cell-type-specific transcriptional enhancers. Such control makes sense during this developmental phase, since

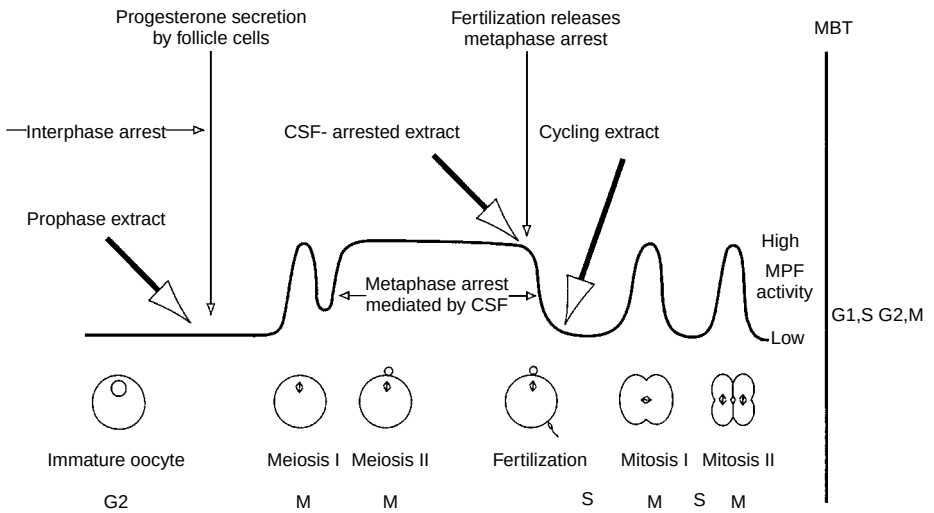

Fig. 9 Schematic diagram of early *Xenopus* development. The figure shows relative levels of MPF (cyclinB/Cdc2) activity from the beginning of oocyte maturation until after fertilization. Maller et al. (2001).

Drosophila embryonic cells undergo minimal growth between divisions. There is no need for cells to pause for growth following mitosis; the key event during this developmental phase is patterning, and this is quickly and effectively accomplished by allowing exclusive control of cell proliferation by pattern-inducing transcriptional enhancers. As development progresses to the larval phase, the developmental objective changes, and so too does the nature of the cell cycle. The major objective of *Drosophila* larval tissues is cell growth rather than cell proliferation; larval cells undergo repeated rounds of endoreplication, an effect that may be designed to increase the total gene dosage in a given cell in an effort to support drastically increased cell size. In such cells, the cell cycle is controlled by nutrition; adequate nutrition triggers further rounds of DNA synthesis. In the undifferentiated larval imaginal cells, a combination of the two regulatory schemes mentioned above is observed. Imaginal disk cells in the developing larva must undergo both growth and patterning, and exhibit a "classical" G1/S/G2/M cell cycle, with a growth-sensitive cell cycle checkpoint at the G1/S boundary and a patterning checkpoint at the G2/M boundary. The developmental program thereby neatly adjusts the cell cycle to promote patterning without growth, growth without patterning, or growth and patterning simultaneously, as each is required for proper development.

Closely related to cell cycle alteration during development is the topic of differentiation and cessation of proliferation. Eventually, most developing cells stop proliferating and specialize. What is the mechanism by which the transition from proliferation to specialization is achieved and timed? The asymmetrical nature of cell divisions has been considered, in the context of *Drosophila* neural progenitor cells, with focus on the protein interactions involved as well as the mechanisms of asymmetric cell division. Timing of differentiation has also been considered in

Drosophila with special attention paid to the global versus cell-intrinsic nature of the timing mechanisms, and in *C. elegans*, focusing on the tight control of postembryonic cell division by temporal regulation genes of the heterochronic gene pathway.

The timing of cell cycle exit and proliferation in mammalian oligodendrocyte development is a process notable for the autonomous control displayed by individual cells *in vitro*. Sister oligodendrocyte precursor cells plated in individual microwells differentiate at approximately the same time. Moreover, the differentiation mechanism does not count cell divisions as illustrated by the observation that at lower temperature conditions, cells divide more slowly, but differentiate after fewer divisions. It therefore appears that oligodendrocyte differentiation is controlled by a cell-intrinsic timing mechanism.

Regarding the molecular underpinnings of the cellular mechanisms involved, asymmetric spindle positioning in the developing *C. elegans* embryo has been studied in the context of the forces exerted by the cell to induce such asymmetry.

5 Organogenesis

A recent paper (Arnone (2002)) considers the current status of our knowledge about the regulation of organogenesis. Differentiated groups of cells make up tissues and organs. Each organ has a characteristic structure and function, which emerges during the development of the embryo, and thus needs to be studied on an individual basis. Many tissue-specific factors that regulate transcription and induce differentiation have been identified, but how they fit into the general phenomenon of organogenesis remains unclear. It is also likely that additional important factors have not yet been identified. Future research into the genesis of various tissues and organs will be designed to identify these as-yet-unknown factors and to integrate these and the currently known factors into a general understanding of organogenesis. Practical applications of this knowledge will include repairing and/or replacement of patients' compromised organs, probably beginning with the patient's own stem cells.

It has been proposed that the relative affinities of organ-identity transcription factors for different target sites control the timing of expression of downstream genes during development. How are organs, typically composed of many different cell types spatially and functionally organized into a unit, initiated and formed during embryonic development? The processes of morphogenetic pattern formation that underlie the construction of organs always have the same initial event – specification of the part of the embryo where organogenesis will begin. The process of organogenesis has a prerequisite step: the specification of spatially defined regulatory domains that promote the differentiation programs. This commonplace mechanism enables the recognition and definition of regulatory fields as discrete territories of specific gene activities.

A model for the regulation of organogenesis has been proposed, as shown in Fig. 10.

Selector genes control the formation and identity of the various fields. Field-specific genes are a special class of selector genes that have the unique property of directing the formation of complex, specialized structures such as organs; they are usually necessary and often sufficient to establish organ identity. Examples of such organ-identity genes include

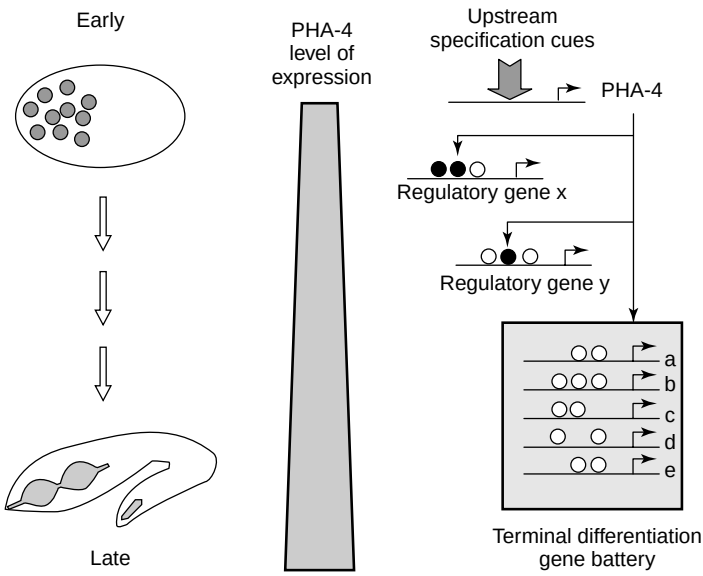

Fig. 10 The product of the organ-identity gene *pha-4* controls pharynx formation in the *C. elegans* embryo. PHA-4 protein (shown in orange or red on the left side of the figure) is present in the nuclei of essentially all pharyngeal cells throughout development. PHA-4 expression levels (shaded wedge at the center) increase as development progresses. PHA-4 directly regulates most pharyngeal genes (red arrows). Both early-acting regulatory genes (*x* and *y*) and late-acting structural genes involved in terminal differentiation (a to e in gray box) are shown on the right side of the figure, bearing one or more PHA-4-binding sites in their cis-regulatory elements (horizontal black lines). In the model proposed, the relative affinities of PHA-4 protein for different targets (open and closed circles for low- and high-affinity sites respectively) modulate the timing of pharyngeal gene expression throughout the genetic regulatory network. Arnone (2002). (See color plate p. xxvi.)

Pax6/eyeless which is required for eye formation in *Drosophila* imaginal discs and *Pit1*, which, together with *Gata2*, controls pituitary differentiation. How do organ-identity genes execute their function?

A frequently observed developmental feature offers a clue: organ-identity genes often act both early and late in the process of organogenesis. The same transcriptional regulators that are necessary for the initial specification of a certain organ are also required for the transcription of structural genes involved in terminal differentiation. What mechanisms allow a differentiation gene battery to be activated only at the end of the process of specification when the genes that regulate them are expressed and active earlier?

In experimental terms, one must find both early and late targets of these genes and study their cis-regulatory systems. Such an approach to study pharynx formation in the nematode *C. elegans* was taken. The precursors of the pharynx are generated by two distinct genetic pathways that eventually converge on the *pha-4* locus. *Pha-4* loss-of-function mutants fail to produce pharynx cells. Conversely,

ectopic *pha-4* expression produces excess pharyngeal cells. *Pha-4* encodes a *FoxA* homolog that is expressed in all pharyngeal precursor cells at the time they are restricted to a pharyngeal fate. In sum, *pha-4* specifies pharyngeal organ identity in *C. elegans*. What are the mechanisms that act downstream of this gene to mediate pharyngeal organogenesis?

By incubating labeled complementary DNA derived from *C. elegans* embryos with microarrays comprised of probes representing 11,917 *C. elegans* genes, and characterizing the resulting candidate pharyngeal genes, it was found that 76 genes are selectively expressed in the pharynx. The cis-regulatory regions of eight of these genes were characterized by reporter-construct analysis using green fluorescent protein. Pharynx-specific expression of all the eight reporter constructs was found to be dependent on the integrity of putative PHA-4 binding sites present in varying copy number within their cis-regulatory elements. Mutations affecting one or more of these binding sites eliminated, dramatically reduced, or delayed the pharynx-specific reporter gene expression.

Expression of these pharynx-specific genes is probably the result of *pha-4*-specific activity. The use of a *pha-4* temperature-sensitive mutant demonstrated that *pha-4* activity was also required for pharynx-specific expression of reporter constructs. Shifting from permissive to restrictive temperature (or *vice versa*) resulted in reduction (or reactivation) of pharynx-specific gene expression. PHA-4 was able to bind *in vitro* to the putative PHA-4 binding elements found in these pharynx-specific promoters.

It was observed that PHA-4 binds to the various target sites with different affinities, suggesting a possible regulatory mechanism for controlling the temporal onset of expression of the genes bearing these target sites. Their results offer a very simple model for regulation of pharynx formation – the product of the organ-identity gene *pha-4* directly activates most of the genes downstream of the process of initial specification, and the relative affinities of PHA-4 for different target sites modulate the timing of gene expression throughout the regulatory network. Consistent with this hypothesis, PHA-4 is expressed at low levels in the early stages of *C. elegans* development, and its expression progressively increases thereafter.

5.1 Chalones

The topic of chalones was recently (Gamer et al. (2003)) put into a historical and contemporary context. This topic relates to the phenomenon of organogenesis. How does an organ ascertain that it has reached the appropriate size during development? It was proposed that the existence of soluble factors produced locally by an organ as it develops act to limit further growth of that organ once it reaches its appropriate size. These soluble factors were named chalones; however, interest in them faded away when these factors could not be purified.

Recently, the discovery of a new TGF-like molecule, Gdf8 (also called *myostatin*), that is expressed in all skeletal muscles, was reported; it was found that, after birth, mice developed from *Gdf8* null embryos exhibited a dramatic and widespread increase in skeletal muscle mass with twice the amount of muscle. It has been shown that GDF8, secreted by differentiating myoblasts, inhibits the proliferation of neighboring muscle precursor cells, arresting them in the G1 phase of the cell cycle. The ability of GDF8 to keep muscle size in check revived the idea of chalones.

Wu et al. demonstrated that during embryonic neurogenesis, GDF11, a TGF-like molecule that is almost identical to GDF8, may act as a chalone in the nervous system. Unlike *Gdf8, Gdf11* expression is not specific for a single tissue. They reasoned that GDF11 could be the inhibitory signal they had been searching for that controlled the number of neurons that differentiate from neural progenitors during embryogenesis. Using mouse olfactory epithelium (OE) as a model, it was shown that Gdf11 and its receptors are expressed by neurons and neuronal progenitors in the OE and that exogenous GDF11 inhibits OE neurogenesis *in vitro* by inducing *p27Kip1* expression and reversible cell cycle arrest in neuronal progenitors. They also demonstrated that mice lacking a functional *Gdf11* gene have increased numbers of neuronal progenitors and neurons in the OE, while mice lacking follistatin, a GDF11 antagonist, show the reverse effect, a dramatic decrease in OE neurogenesis. They conclude that the negative autoregulatory action of GDF11 is strikingly similar to the action of GDF8 in skeletal muscle, demonstrating that similar strategies are used to establish and maintain proper cell number in nervous tissue and muscle, as seen in Fig. 11.

Experiments involving GDF8 and GDF11 demonstrate that molecules fitting some of the requirements of Bullough's chalones are produced and act locally. How this local activity relates to the regulation of organ size still remains a mystery. It seems that chalones are back with a vengeance.

6 Stem Cells

A stem cell is a cell that undergoes mitotic cell division to give rise to the same type

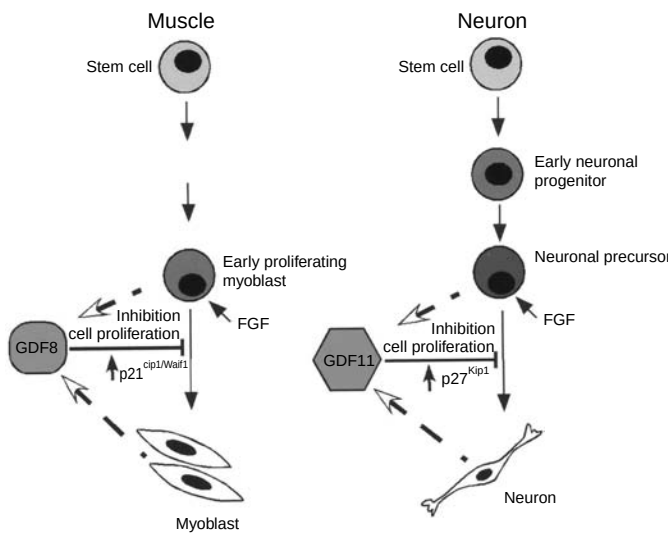

Fig. 11 Regulation of tissue size by GDFs. During neurogenesis and myogenesis, the negative autoregulatory action of GDFs on cell proliferation sets a limit on the number of mature neurons and muscle cells, maintaining control of organ size. Wu et al. (2003). (See color plate p. xxii.)

of cell. At some point, stem cells leave the pool of mitotically dividing cells to begin a process of cell differentiation. Stem cells are, in effect, an embryonic population of cells, continually producing cells that can undergo further development within an adult organism. The path of differentiation that a stem cell descendant enters depends on the molecular milieu in which it resides, for example, erythrocytes, granulocytes, neutrophils, platelets, and lymphocytes share a common precursor cell – the pluripotential hematopoietic stem cell.

The AGM region consists of the aorta, gonads, and mesonephros region. It is the domain, near the aorta, where, in fishes, mammals, and frogs, the definitive hematopoietic cells are formed. These mesodermal regions of the day-11 mouse embryo are where pluripotential hematopoietic stem cells (HSC) and CFU-s (colony-forming unit of the spleen) cells can be found. These blood cell precursors will colonize the liver and constitute both the fetal and adult circulatory system. Around the time of birth, stem cells from the liver populate the bone marrow, which then becomes the major site of blood formation throughout adult life.

The term pluripotent describes a cell that is capable of giving rise to several of the cell types of an organism, for example, the pluripotent stem cells of the bone marrow give rise to all classes of blood cells and neural crest cells can differentiate into different cell types, such as cholinergic neurons and adrenergic neurons, depending on their location. The pluripotent epidermal stem cell is a stem cell whose offsprings become epidermis, sebaceous gland, or hair shaft; the pluripotential hematopoietic stem cell is the common precursor cell shared by erythrocytes, basophils, eosinophils, mast cells, monocytes, tissue macrophages, osteoclasts, neutrophils, platelets, and T- and B-lymphocytes; and the pluripotent neural stem cell (NSC) is a stem cell whose offspring become neural and glial cells. The stratum germinativum, the innermost layer of the epidermis, contains mitotic stem cells required for the continual replenishment of the epidermis.

Hemangioblasts are the common precursor cells shared by blood vessels and blood cells; true blood cell stem cells of the chick embryo are derived from the splanchnopleure and formed within nodes of mesoderm that line the mesentery and the major blood vessels. Hematopoietic stem cells are those cells that generate the different types of blood cells. Hematopoietic inductive environments (HIMs) are regions where blood-forming cells undergo determination; the developmental pathway taken by the descendants of a pluripotential stem cell depends on which growth factors it is exposed to, which, in turn, is determined by the stromal cells to which it is exposed. Stem cells in the spleen become predominantly erythroid while those in the bone marrow become predominantly granulocytic. Stromal cells of the bone marrow create HIMs by their ability to bind hematopoietic growth factors; without these growth factors, the stem cells die.

Lineage-restricted stem cells are cells that can produce only one type of cell in addition to renewing itself, for example, the BFU-E (burst-forming unit, erythroid) is a lineage-restricted stem cell; it can form only one cell type, CFU-E (colony-forming unit, erythroid), in addition to itself. CFC-Meg is a lineage-restricted stem cell, derived from the myeloid precursor cell; it gives rise to megakaryocytes, which, in turn, give rise to platelets. The CFU-GM lineage-restricted stem cell, derived

from a CFU-M, L, gives rise to the granulocyte precursor cell and the monocyte. Osteoclasts, cells that have the ability to break down already-formed bone, play an important role in bone remodeling. Osteoclasts come from the CFU-GM, the same stem cell as macrophages and granulocytes; the growth factor interleukin 6 (IL-6) stimulates the production of osteoclasts; the production of IL-6 is inhibited by estrogen and testosterone. Osteoprotegerin, osteoclastogenesis inhibitory factor, is a member of the TNF receptor superfamily; together with its ligand, it regulates conversion of a macrophage stem cell into an osteoclast.

CFU-M, L is the colony-forming unit of the myeloid and lymphoid cells, a pluripotential hematopoietic stem cell, from which both the CFU-S and lymphocytes are derived. This cell gives rise to the CFU-S (blood cells) and the CFU-L (lymphocytes). CFU-S, a pluripotential hematopoietic stem cell, is capable of forming erythrocytes, granulocytes, and platelets. However, the immediate progeny of the CFU-S are lineage-restricted stem cells, for example, the BFU-E (burst-forming unit, erythroid) can form only one cell type in addition to itself.

The myeloid precursor cell is the pluripotent stem cell, which gives rise to three kinds of lineage-restricted stem cells: (1) CFU-GM, (2) CFC-Meg, and (3) BFU-E. The lymphoid precursor cell is the pluripotent stem cell, which gives rise to pre-T and pre-B cells; these cells, in turn, give rise to T-cells and B-cells, respectively; the terminally differentiated cells of this lineage are activated T-cells and plasma cells, respectively.

Embryonic germ (EG) cells are primordial germ cell (PGC)-derived cells; these pluripotent embryonic stem cells (ESCs) may be produced in culture by treating PGCs with stem cell factor, leukemia inhibition factor, and basic FGF2. Stem cell factor (SCF) is a paracrine protein that promotes cell division in numerous stem cell populations; mice lacking SCF or its receptor (c-Kit protein) are sterile (no germ cells), white (no pigment cells), anemic (no red blood cells), and immunodeficient (no lymphocytes). Fibroblast growth factor, a family of protein growth factors, first identified as mitogens for fibroblasts in tissue culture, stimulates proliferation of many cell types, inhibits differentiation of various types of stem cells, and acts as inductive signals in embryonic development. EG cells have the potential to differentiate into all the cell types of the body; they are often considered to be embryonic stem cells, ignoring the distinctiveness of their origin.

Embryonic stem cells are stem cells derived from the embryo, for example, derived from normal mouse inner mass cells cultured *in vitro*. Use of these cells to make chimeras allows investigators to introduce an engineered gene with a known mutation into ES cells and then to breed mice that have the same mutation in all their cells.

Cord blood is blood obtained from the umbilical cord at the time of delivery, a rich source of stem cells.

The oogonium (plural, oogonia) is a type of female germ cell that functions as a stem cell until it begins to undergo cell differentiation and becomes an oocyte. In humans, all oogonia have differentiated into primary oocytes before the birth of a female; however, in frogs oogonia persist throughout reproductive life as a stem cell population, multiply and produce eggs for the next breeding system. Spermatogonia are male germ cells that act as stem cells and may undergo meiosis and cell differentiation to become spermatozoa (gametes). Spermatogonia

are spermatogenic (sperm producing) cells. Various types of spermatogonia are recognized in humans, that is, A_1, A_2, A_3, A_4, intermediate, and B. It is thought that each of the A types are stem cells, capable of renewal; intermediate spermatogonia are committed to becoming spermatozoa and divide to form type B spermatogonia; type B spermatogonia are the precursors of spermatocytes. In 1840, Albrecht Koelliker realized that sperm from the male and ova from the female were cells; he earned his doctorate with work that traced the development of spermatozoa from stem cells to their release from the testes.

Nurse cells of the *Drosophila* ovary are produced by stem cells undergoing four mitotic divisions (one of the resulting 16 cells, per stem cell, becomes the oocyte; the others become nurse cells). The nurse cells produce large quantities of proteins and RNAs that are exported into the egg through cytoplasmic bridges. Bicoid mRNA is made by nurse cells located next to the anterior end of the developing oocyte and is transferred from them into the egg.

A meristem is a part of a plant characterized by dividing cells; these regions of plants contain stem cell populations that produce cells, some of which go on to differentiate into plant tissues and some of which constantly renew the stem cell population. Generally, two varieties are found: (1) apical meristems, found at the tips of plant parts, for example, at the growing ends of roots and shoots and (2) lateral meristems, found on the sides of plants, for example, vascular cambium and cork cambium are lateral meristems found under the bark of tree trunks. The primary root of a flowering plant has three meristematic regions: the procambium, the ground meristem, and the protoderm, which give rise to primary tissues.

Primary tissues are plant tissues produced directly by the activity of apical meristems; the procambium gives rise to the primary xylem, vascular cambium (a lateral meristem), and primary phloem; the ground meristem gives rise to the pericycle, endodermis, and cortex; and the protoderm gives rise to the epidermis.

In the procedure known as therapeutic cloning, nuclear material from a donor somatic cell is transferred into an enucleated egg cell and ESCs are harvested from the resulting blastocyst; if the ESCs can be coaxed to differentiate into functioning tissues, the new cells may be introduced into the donor to regenerate diseased tissue and even organs. An objective of therapeutic cloning is to derive immunologically compatible embryonic stem cells.

Recently, citing new data, that is, the first genome-wide transcript analysis of ESCs, HSCs, and NSCs produced by transcriptional profiling of purified stem cell populations it was pointed out that the new data sets begin to establish the nature of "stemness." Burns and Zon (2002) suggest an alternative to the Till and McCulloch stochastic model of 1961, the multilineage priming model of May and Ever, a more modern model for stem cell fate that suggests that stem cells express a low level of lineage affiliated genes, which are amplified or repressed upon differentiation, that is, it may be the specific combination of genes rather than individual genes that endow the unique properties of stem cells, priming them to accept specific environmental cues.

It was noted that, despite these recent advances, the intrinsic program induced by environmental cues has not been molecularly described and that further analysis is clearly required to ascertain the relevance of the identified core stem cell genes, suggesting that the functionally

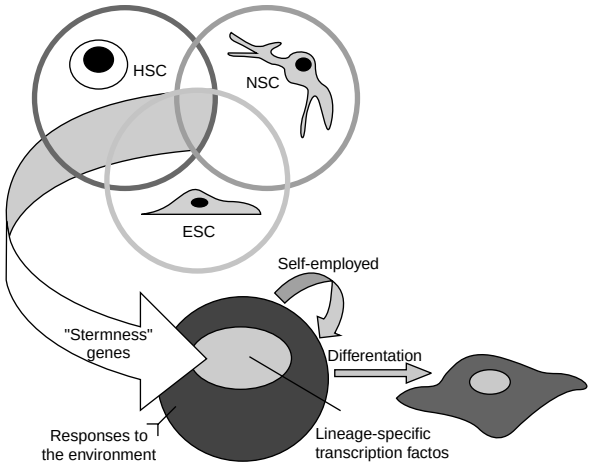

Fig. 12 Theoretically, all stem cells (ESC, HSC, and NSC) may share a core set of ''stemness'' genes that enable both self renewal and differentiation. These internal programs may be regulated by environmental cues. Each individual stem cell also expresses a set of individual lineage-specific genes. The specific combination of common and lineage-specific transcripts and the biased distribution of internal factors may lead to stem cell asymmetric division. Burns and Zon (2002).

important genes will likely be uncovered by comparing stem cells in different lineages and species. See Fig. 12.

7 Cloning

A clone may be defined as (1) a population of cells derived from a single cell by mitosis or (2) a group of organisms produced by asexual reproduction, for example, vegetative propagation in plants, parthenogenesis in animals, nuclear transfer techniques involving the transfer of somatic cell nuclei of identical genetic constitution into enucleated eggs. These definitions notwithstanding, cloning may be applied at the molecular and nuclear levels, as well as at the cellular and organismal levels.

In molecular cloning (gene cloning), there is amplification (by the polymerase chain reaction) of single genes to a chemically useful quantity. The polymerase chain reaction (PCR) is a highly efficient, *in vitro*, enzymatic method of cloning (making many identical copies of) specific pieces of DNA.

Positional gene cloning is a method used to find genes that are involved in normal human development and whose mutations cause congenital developmental malformations. Pedigree analysis highlights a region of the genome where a particular mutant gene is thought to reside; by making a DNA map of that area, one may find a region of DNA that differs between people who have the mutation and people who lack it; then by sequencing that region of the genome, the gene can be located, for example, the discovery of the human *Aniridia* gene. Such candidate gene mapping was used to study Waardenburg sydrome type 2, that is, pedigree

analysis was used to determine that Waardenburg sydrome type 2 is caused by a mutation in a gene on the small arm of chromosome 3, between bands 12.3 and 14.4; using a probe made from the mouse microphthalmia gene, it was found that a human homolog (*MITF*) of the mouse microphthalmia gene mapped to exactly the same region as Waardenburg syndrome type 2.

In the procedure known as knockout or targeted gene knockout, the targeted disruption of function in a cloned gene, the gene is usually introduced back into the organism in such a way as to replace a normal gene copy, by homologous recombination, in order to analyze the effect of the mutation. Interestingly, in many cases knocking out a normal gene has had no effect at all, even when the gene was thought to be essential. Two genes exhibit genetic redundancy when the absence of the function of one gene causes little or no effect because a second gene can operate in its place.

In cellular cloning, the founder cell gives rise to a particular group of cells, for example, a cell clone consists of all surviving descendants of a single cell, which is called the founder cell of the clone.

In therapeutic (or research) cloning, nuclear material from a donor somatic cell is transferred into an enucleated egg cell and ESCs are harvested from the resulting blastocyst. If the ESCs can be coaxed to differentiate into functioning tissues, the new cells might be introduced into the donor to regenerate diseased tissue and even organs. An objective of therapeutic cloning is to derive immunologically compatible embryonic stem cells. Reproductive cloning is cloning for reproductive purposes, that is, producing new organisms that are genetically identical to the donor of the nuclear material. Cloning, as applied to organisms, is the process (natural or artificial) of making several, genetically identical, independent organisms from a single initial organism.

In 1936, a "somewhat fantastical" experiment was suggested to transfer a nucleus from a differentiated cell into an enucleated egg. In 1952, Briggs and King developed the techniques for the frog (*Rana*); they transferred blastula stage cell nuclei into enucleated eggs and were able to obtain complete tadpoles. It was already shown that blastula cell nuclei were pluripotent, but what about nuclei from more advanced stages? In 1956, Briggs and King used somatic cell nuclei of tailbud-stage tadpoles (vs germ cell nuclei of tailbud-stage tadpoles); they demonstrated (1) an apparent decrease in nuclear potency and (2) apparently the decrease in nuclear potency was stable and tissue specific, as demonstrated by nuclear cloning, for example, endoderm nuclei appear to be good at making endoderm, but are restricted in making ectoderm and mesoderm. Progressive restriction of nuclear potency during development appeared to be the general rule. However, another possible explanation for limited potency would be chromosomal abnormalities. Gurdon, using the frog *Xenopus*, used intestinal endoderm of feeding tadpoles and seemed to demonstrate that these nuclei were totipotent. However, King criticized Gurdon's work; subsequently, Gurdon cultured epithelial cells from adult frog foot webbing and by serial transplantation was able to obtain tadpoles, but these died before feeding. These amphibian-cloning experiments suggested (1) that a general restriction of potency occurred with progressive development and (2) that the differentiated cell genome was remarkably potent. Debate persisted about the totipotency of

such nuclei, but there was little doubt that they were extremely pluripotent.

Dolly, the sheep, was the first mammal cloned from a cell of an adult animal. She was a result of an experiment to test the suitability of different sources of cells for nuclear transfer. She was derived from cells that had been taken from the udder of a six-year old Finn Dorset ewe and then cultured in the laboratory. Individual cells were then fused with unfertilized eggs from which the genetic material had been removed and 29 of these "reconstructed eggs," each now with a nucleus from the adult animal, were then implanted in surrogate Blackface ewes. One gave rise to a live lamb, Dolly, some 148 days later. Other cloned lambs were derived in the same way from cells taken from embryonic and fetal tissue. Dolly was euthanized in February 2003 due to respiratory disease.

Somatic cell nuclear transfer (SCNT) is the technique whereby nuclear material from a donor somatic cell is transferred into an enucleated egg cell, generally, as part of a cloning technique.

Clonal analysis is a technique used in developmental biology as a form of fate mapping, in which a single cell is labeled and the position and cell types of its progeny identified at a later stage. The *Minute* technique uses cells of *Drosophila* carrying a mutation in the *Minute* gene, which, therefore, grow more slowly than those in the wild type. By using flies heterozygous for the *Minute* mutation, clones can be made in which a mitotic recombination event has generated a marked cell that is normal because it has lost the *Minute* mutation, and so is wild type. This normal cell proliferates faster than the slower growing background *Minute* heterozygous cells, and thus large clones of the marked cells are produced.

Clonal restriction is found in insect compartments defined by clonal restriction lines, in which lines are not transgressed by cell clones generated after a certain stage of development.

Cystocytes are the cells of a clone of cells derived from the oogonium by meroistic oogenesis in insects; ring canals are connections between such clones.

Clonal selection theory is the concept that each animal first randomly generates a vast diversity of lymphocytes and then those cells that react against the foreign antigens that the animal actually encounters are specifically selected for action (form a clone).

See also Brain Development; Cellular Interactions; Electron Microscopy in Cell Biology; Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Bock, G., Cardew, G., Goode, J.A. (Eds.) (2001) *The Cell Cycle and Development*, Novartis Foundation Symposium 237, John Wiley & Sons, West Sussex, UK.

Bullough, W.S. (1965) Mitotic and functional homeostasis: a speculative review, *Cancer Res.* **25**, 1683–1727.

Churchill, F.B. (1991) The Rise of Classical Descriptive Embryology, in: Gilbert, S.F. (Ed.) *A Conceptual History of Modern Embryology*, The Johns Hopkins University Press, Baltimore, 1–23.

Dye, F.J. (2002) *Dictionary of Developmental Biology and Embryology*, Wiley-Liss, New York.

Gamer, L.W., Nove, J., Rosen, V. (2003) Return of the chalones, *Dev. Cell* **4**, 143–144.

Gilbert, S.F. (2000) *Developmental Biology*, 6th edition, Sinauer, Sunderland, MA.

May, G., Enver, T. (2001) Lineage Commitment and Self-Renewal of Blood Stem Cells, in:

Zon, L.I. (Ed.) *Hematopoiesis: A Developmental Approach*, Oxford University Press, New York, 61–81.

Oberst, A. (2002) The cell cycle and development, *Nature* **9**, 227–228.

Saha, M. (1991) Spemann Seen Through a Lens, in: Gilbert, S.F. (Ed.) *A Conceptual History of Modern Embryology*, The Johns Hopkins University Press, Baltimore, 91–107.

Wilson, E.B. (1925) *The Cell in Development and Heredity*, 3rd edition, Macmillan, New York.

Primary Literature

Adler, P.N. (2002) Planar signaling and morphogenesis in Drosophila, *Dev. Cell* **2**, 525–535.

Arnone, M.I. (2002) Bringing order to organogenesis. *Nat. Genet.* **30**, 348–350.

Burns, C.E., Zon, L.I. (2002) Portrait of a stem cell, *Dev. Cell* **3**, 612–613.

Campbell, K.H.S., McWhir, J., Ritchie, W.A., Wilmut, I. (1996) Sheep cloned by nuclear transfer from a cultured cell line, *Nature* **380**, 64–66.

Chen, C.S., Mrksich, M., Huang, S., Whitesides, G.M., Ingber D.E. (1997) Geometric control of cell life and death, *Science* **30**, 1425–1428.

Gaudet, J., Mango, S.E. (2002) Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4, *Science* **295**, 821–825.

Giancotti, F.G., Ruoslahti E. (1999) Integrin signaling, *Science* **285**, 1028–1033.

Hall A. (1998) Rho GTPases and the actin cytoskeleton, *Science* **279**, 509–514.

Hertzberg, E.L. (2000) Gap Junctions in the New Millennium: More Than Pretty Pictures, *Gap Junctions 30: Advances in Molecular and Cellular Biology*, JAI Press, Stamford, CT, ix–xii.

Herwig, S., Strauss, M. The Retinoblastoma Protein – A Master Regulator of Cell Cycle, Differentiation and Apoptosis, http://hucellbiol.mdc-berlin.de/review0_text.ht 1997.

Hou, S.X., Zheng, Z., Chen, X., Perrimon, N. (2002) The JAK/STAT pathway in model organisms: emerging roles in cell movement, *Dev. Cell* **3**, 765–778.

Ivanova, N., Dimos, J.T., Schaniel, C., Hackney, J.A., Moore, K.A., Lemischka, I.R. (2002) A stem cell molecular signature, *Science* **298**, 601–604.

Keller, R. (2002) Shaping the vertebrate body plan by polarized embryonic cell movements, *Science* **298**, 1950–1954.

Knust, E., Bossinger, O. (2002) Composition and formation of intercellular junctions in epithelial cells, *Science* **298**, 1955–1959.

Lubarsky, B., Krasnow, M.A. (2003) Tube morphogenesis: making and shaping biological tubes, *Cell* **112**, 19–28.

Maller, J.L., Gross, S.D., Schwab, M.S., Finkielstein, C.V., Taieb, F.E., Qian, Y.-W. (2001) Cell Cycle Transitions in Early Xenopus Development, in: Bock, G., Cardew, G., Goode, J.A. (Eds.) *The Cell Cycle and Development*, John Wiley & Sons, Chichester, UK, 58–78.

Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R.C., Melton, D.A. (2002) Stemness transcriptional profiling of embryonic and adult stem cells, *Science* **298**, 597–600.

Till, J., McCulloch, E. (1961) A direct measurement of the radiation sensitivity of normal mouse bone marrow cells, *Radiat. Res.* **14**, 213–222.

Wu, H.H., Ivkovic S., Murray R.C., Jaramillo S., Lyons K.M., Johnson J.E., Calof, A.L. (2003) Autoregulation of neurogenesis by GDF11. *Neuron* **37**, 197–207.

Diabetes Insipidus, Molecular Biology of

Alexander Oksche[1,2], Peter Pohl[2], Gerd Krause[2], and Walter Rosenthal[1,2]
[1] *Institut für Pharmakologie, Charité – Universitätsmedizin Berlin, Campus Benjamin Franklin, Thielallee 67–73, Berlin, Germany*
[2] *Forschungsinstitut für Molekulare Pharmakologie, Robert-Roessle-Str. 10, Berlin, Germany*

1 Vasopressin 304
1.1 Structure, Synthesis, and Release of Vasopressin 304

2 Vasopressin Receptors 306
2.1 Structure and Expression of Vasopressin Receptors 306
2.2 Signal Transduction of G protein-coupled Vasopressin Receptors 307

3 Aquaporins 309
3.1 Requirements for Water Channel Proteins 309
3.2 Structural Features Responsible for Water Selectivity 309
3.3 The Main Aquaporins in Kidney 309
3.4 The AQP2 Shuttle 311

4 Diabetes Insipidus 312
4.1 Central Diabetes Insipidus 312
4.1.1 Autosomal Dominant Central Diabetes Insipidus 313
4.1.2 Autosomal Recessive Central Diabetes Insipidus 314
4.2 Nephrogenic Diabetes Insipidus 314
4.2.1 X-linked Nephrogenic Diabetes Insipidus 315
4.2.2 Autosomal Recessive Nephrogenic Diabetes Insipidus 315
4.2.3 Autosomal Dominant Nephrogenic Diabetes Insipidus 317
4.3 Treatment of Diabetes Insipidus 317
4.3.1 Treatment of Central Diabetes Insipidus 317
4.3.2 Treatment of Nephrogenic Diabetes Insipidus 318

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

5 Animal Models 319
5.1 Animal Models of Central Diabetes Insipidus 319
5.2 Animal Models of Nephrogenic Diabetes Insipidus 319

6 Outlook 319

Bibliography 320
Books and Reviews 320
Primary Literature 321

Keywords

Diabetes Insipidus
Inherited or acquired disease characterized by polyuria and polydipsia.

Central Diabetes Insipidus
Diabetes insipidus caused by insufficient release of vasopressin from the neurohypophysis.

Nephrogenic Diabetes Insipidus
Diabetes insipidus caused by a resistance of the kidney toward the action of vasopressin.

Vasopressin
Nonapeptide synthesized in hypothalamic nuclei and released from the posterior pituitary. The main biological activity is conservation of water (via vasopressin V_2 receptors); therefore it is also termed *antidiuretic hormone* (ADH). Vasopressin also stimulates the contraction of vascular smooth muscle cells, glycogenolysis in liver cells (via vasopressin V_{1a} receptors), release of ACTH from the anterior pituitary (via vasopressin V_{1b} receptors), and secretion of the von Willebrand factor (via vasopressin V_2 receptors). In addition, vasopressin acts as a neurotransmitter in the central nervous system (via V_{1a} and V_{1b} receptors), regulating the adaptive and social behavior.

Oxytocin
Nonapeptide homologous to vasopressin, which is also produced in hypothalamic nuclei and released from the posterior pituitary. In the pregnant uterus, it potentiates the strength and frequency of smooth muscle contractions. After birth, stimulation of the nipple results in the release of oxytocin, which in turn causes contraction of myoepithelial cells of the mammary ducts and the ejection of milk.

G protein-coupled Receptors
Superfamily of membrane receptors, characterized by a common seven transmembrane-spanning configuration. Activation of the receptors by extracellular ligands causes stimulation of heterotrimeric G proteins.

Heterotrimeric G-proteins
Guanine nucleotide regulatory protein complexes composed of α and $\beta\gamma$ subunits. Upon activation by G protein-coupled receptors, they influence the activity of a large variety of effector molecules such as adenylyl cyclase, phospholipases, and ion channels.

Aquaporins
Integral membrane proteins with six transmembrane-spanning domains, allowing water to cross the plasma membrane rapidly. In man, 10 different aquaporins are known, of which seven are expressed in the renal tubular epithelium and the renal collecting duct. Aquaporins are also found in the epithelium and endothelium of brain, lung, and intestine.

The neurohypophyseal hormone vasopressin acts via three different G protein-coupled receptors. The main physiological response elicited by vasopressin is antidiuresis, which is mediated by vasopressin V_2 receptors in the kidney. Further responses, such as vasoconstriction and the release of the adrenocorticotropic hormone (ACTH), are mediated by vasopressin V_{1a} and vasopressin V_{1b} receptors respectively.

Diabetes insipidus is a hereditary or acquired disease characterized by polyuria and polydipsia. It is either caused by an insufficient release of vasopressin from the neurohypophysis (central diabetes insipidus) or by a resistance of the kidney toward the action of vasopressin (nephrogenic diabetes insipidus).

In the case of central diabetes insipidus, autosomal dominant and autosomal recessive modes of inheritance are known. In both traits, mutations in the vasopressin-neurophysin II gene (encodes the vasopressin precursor) are responsible for the disease. For congenital nephrogenic diabetes insipidus, three modes of inheritance are known: X-chromosomal recessive, autosomal recessive, and autosomal dominant. While the X-chromosomal trait is caused by mutations in the vasopressin V_2 receptor gene, the autosomal forms are due to mutations in the aquaporin-2 gene (encodes the vasopressin-regulated water channel).

Analysis of mutant vasopressin-neurophysin II, vasopressin V_2 receptor, and aquaporin-2 proteins revealed impaired protein transport as the most common cause of the disease. The transport defect is either caused by (1) retention of misfolded proteins in the endoplasmic reticulum or (2) misrouting of correctly folded proteins.

1
Vasopressin

1.1
Structure, Synthesis, and Release of Vasopressin

Vasopressin is a cyclic nonapeptide (Fig. 1a). Its amino acid sequence is highly conserved among mammals. Only swine and marsupials have a lysine residue at position 8 (Lys-vasopressin, LVP), whereas all other mammals display an arginine residue (Arg-vasopressin, AVP). AVP is highly homologous to oxytocin (Fig. 1b). The genes encoding AVP and oxytocin evolved from an anchestral gene by gene duplication. Both genes share a common gene structure and are only separated by 11 000 and 8 000 bps of noncoding DNA in rat and man respectively.

AVP and oxytocin are encoded as precursor proteins and are synthezised in magnocellular neurons of the supraoptic and paraventricular nuclei of the hypothalamus. The AVP precursor consists of a signal peptide, the hormone moiety (AVP), a carrier protein (neurophysin II), and a copeptin (Figs. 2 and 3). The oxytocin precursor comprises a signal peptide, the hormone moiety (oxytocin) and a carrier protein (neurophysin I). The neurophysins (neurophysin I and II) function not only as carrier proteins to protect the hormones from proteolytic degradation within the secretory vesicles but are also crucially involved in correct folding and processing of the precursor proteins. The function of the copeptin encoded by the *AVP-NPII* gene remains elusive.

In man, Arg-vasopressin (AVP) serum levels normally range from 2.5 to 5 pg mL^{-1}. The secretion of AVP is under the control of osmoreceptors in the hypothalamus and of baroreceptors located in the heart and the carotid artery. The osmoreceptors in the hypothalamus are activated by minute increases in the plasma osmolality ($\geq$1%), which in turn stimulate the release of small quantities of AVP (changes of serum AVP are less than or equal to 1 pg mL^{-1}). Greater changes in plasma osmolality cause larger increases in plasma AVP of up to 10 pg mL^{-1}.

Relatively large changes in blood volume are necessary to stimulate AVP release

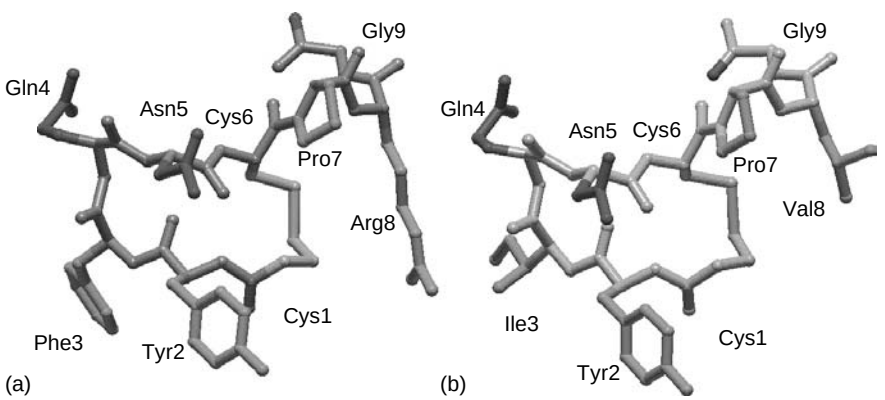

Fig. 1 Structure of cyclic nonapeptides vasopressin and oxytocin. Homology model of vasopressin (a) based on the crystal structure of oxytocin (b; Brookhaven 3D PDB: 1XY1). Amino acids are depicted in the three letter code. The numbers are the positions of amino acids within the molecules.

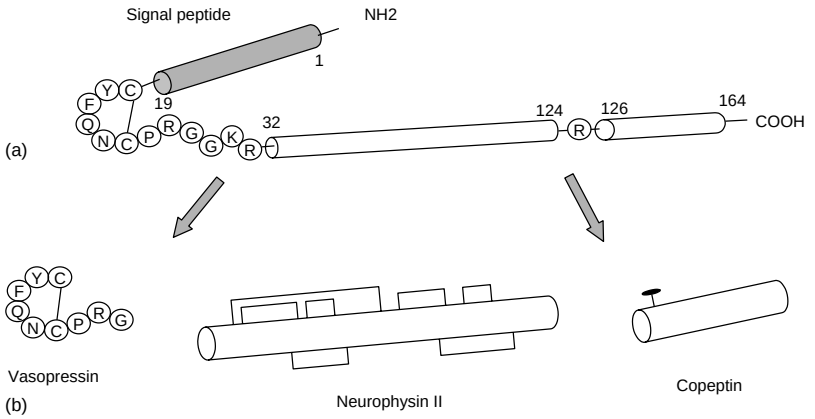

Fig. 2 Generation of vasopressin from the vasopressin-neurophysin II precursor protein. (a) Structure of the AVP-NPII precursor protein; (b) mature vasopressin, neurophysin II, and copeptin. Posttranslational modifications shown are intramolecular disulfide bonds in the vasopressin and neurophysin II moiety, glycosylation of the copeptin.

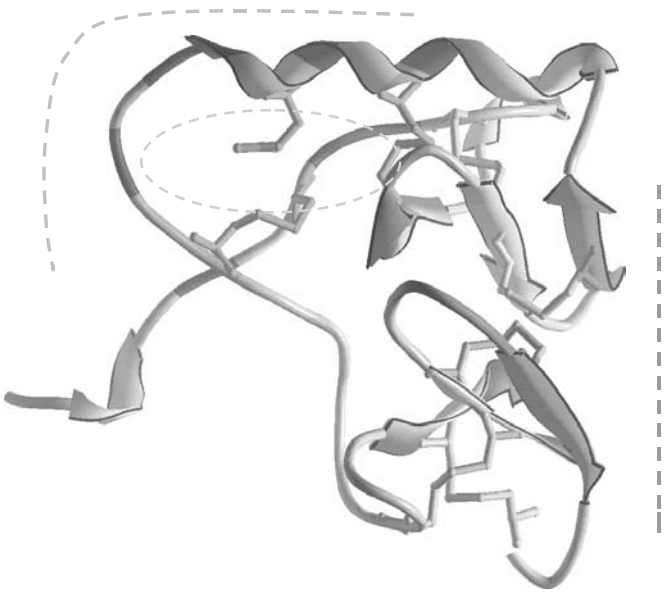

Fig. 3 Crystal structure of bovine neurophysin II (Brookhaven 3D PDB: 2BN2). *Blue* – amino acids interacting with vasopressin, *blue circle* – docking site of vasopressin, *yellow* – disulfide bonds, *green* – amino acids, involved in the dimerization of NPII, *green dotted line* – contact site of NP II dimers, *red* – amino acids involved in the tetramerization of NPII, *red dotted line* – contact site of the tetramer. Dimerization of two NPII molecules has been experimentally verified. Tetramerization was found in the crystal structure, but its physiological significance has not yet been determined. (See color plate p. xxvii.)

via baroreceptors. While a decrease in the blood volume of less than 10% is insufficient to stimulate AVP secretion, the acute loss of more than 10% dramatically increases AVP secretion: Plasma levels of AVP can reach 100 pg mL^{-1} if the blood loss exceeds 20 to 30%. While antidiuresis is maximally stimulated at about 10 pg mL^{-1} AVP, vascular effects such as vasoconstriction are observed only at higher AVP levels (10–100 pg mL^{-1}).

2 Vasopressin Receptors

2.1 Structure and Expression of Vasopressin Receptors

AVP exerts its action via four distinct receptors belonging to two different types of membrane proteins. Three receptors, namely vasopressin V_{1a}, V_{1b}, and V_2 receptors (V_{1a}, V_{1b}, and V_2 receptors), belong to the superfamily of G protein-coupled receptors. These proteins share a common topology with an extracellular N terminus, an intracellular C terminus and seven transmembrane domains (Fig. 4). The G protein-coupled vasopressin receptors are encoded by three different genes and show a distinct pattern of tissue expression (Table 1). The binding site for AVP is formed by the transmembrane domains and the first and second extracellular loops. The docking of AVP elicits a conformational change, which in turn leads to the activation of heterotrimeric G proteins. The activated G proteins in turn can modulate the function of effector molecules, such as adenylyl cyclase, phospholipase Cβ, or diverse ion channels. The receptor-stimulated physiological response depends on the cellular environment and the nature of the activated G proteins (see Sect. 2.2).

The fourth vasopressin receptor, the vasopressin-activated calcium-mobilizing receptor (VCAM-1), is a single transmembrane receptor comprising about 780 amino acids. It has been identified in man, rat, and rabbit. VCAM-1 does not possess any homology to the G protein-coupled vasopressin receptors. It belongs to a new

Tab. 1 Synoposis of the chromosomal localization, tissue expression, and activated signal cascades of vasopressin receptors.

Type of receptor	*Chromosomal localization*	*Exon number*	*Signaling*	*Site of expression*
V_{1a}	12q14-q15	2	G_q, G_i, $G_{12/13}$	Vessels (smooth muscle cells) Liver (hepatocytes) Kidney (interstitial cells)
V_{1b}	1q32	2	G_q, ($G_{12/13}$)	Adenohypophysis (basophilic cells) Adrenal gland (chromaffine cells) Pancreas (β cells)
V_2	Xq28	3	G_s	Kidney (principal cells) Vessels (endothelium)
VACM-1	11q22-q23	20	Ca^{2+} ↑, G-protein-independent	Renal collecting duct Endothelial cells

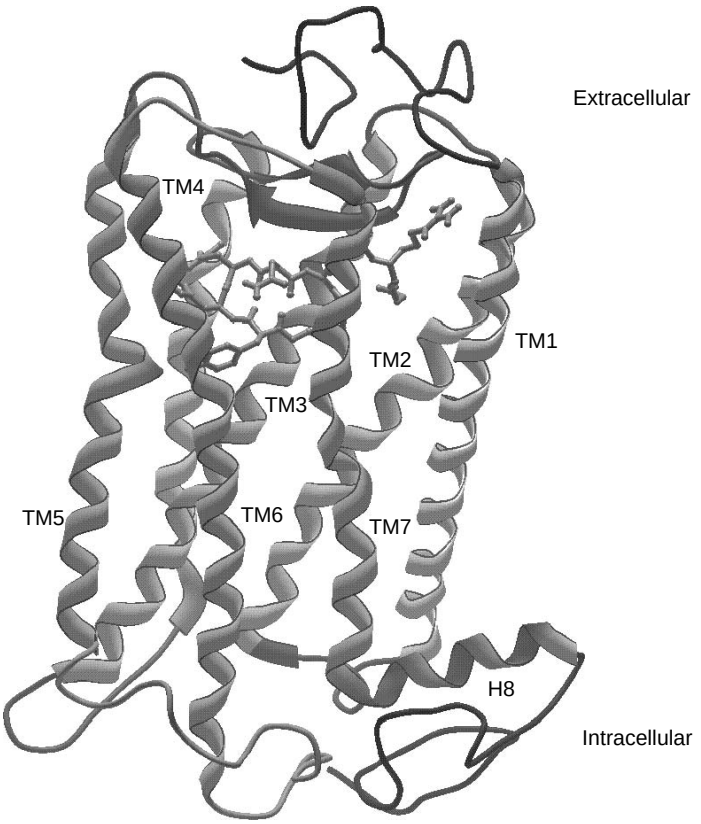

Fig. 4 Homology model of the ligand-bound vasopressin V_2 receptor based on the 3D model of bovine rhodopsin (Brookhaven 3D PDB: 1L9H). Seven transmembrane helices are labeled TM1 to TM7. For clarity, the transmembrane helices in the front are in cyan and the transmembrane helices at the back are gray. *Dark blue lines* – N and C tail, *red lines* – 1st extracellular/intracellular loop, *magenta lines* – 2nd extracellular/intracellular loop, *green lines* – 3rd extracellular/intracellular loop. *Yellow* – synthetic agonist DDAVP. A short helix (H8) is found between TM7 and the C tail. (See color plate p. xxviii.)

family of peptide receptors, designated as cullins. VACM-1 shows homology with cullin-5 of *Caenorhabditis elegans*, a protein involved in the control of the cell cycle, and when mutated, contributes to tumor progression. Binding of AVP leads to an increase in calcium by a yet unknown mechanism.

In the rat, a further AVP-sensitive receptor, the dual angiotensin II/AVP (AngII/AVP) receptor, has been described. Up to now, the existence of this receptor in any other species has not been reported.

2.2 Signal Transduction of G protein-coupled Vasopressin Receptors

The V_{1a} receptor mainly activates G proteins of the $G_{q/11}$ and the $G_{12/13}$

family. Depending on the cell cycle, V_{1a} receptors also have the ability to activate G proteins of the G_i family. The V_{1b} receptor activates G proteins of the $G_{q/11}$ family and the V_2 receptor stimulates the G protein G_s.

Activation of G proteins of the G_q family leads to a stimulation of the phospholipase Cβ, which catalyzes the formation of inositoltriphosphate (IP_3) and diacylglycerol (DAG). IP_3 and DAG function as second messengers. DAG activates multiple isoforms of protein kinase C, which in turn phosphorylate a great number of diverse proteins at serine and threonine residues, thereby regulating the functional activity of the proteins. IP_3 mediates the release of calcium from the smooth endoplasmic reticulum via the IP_3 receptor. The increase of intracellular calcium stimulates the release of ACTH in basophilic cells (via the V_{1b} receptor) of the anterior pituitary and mediates the constriction of vascular smooth muscle cells (via the V_{1a} receptor, Fig. 5). In addition, the stimulation of $G_{12/13}$ has a synergistical action on smooth muscle cell contraction, resulting in a prolonged vascular smooth muscle contraction.

In the case of the V_2 receptor, the activation of G_s leads to the stimulation of the adenylyl cyclase, resulting in the generation of the second messenger cyclic adenosine monophosphate (cAMP). High levels of cAMP stimulate the cAMP-dependent protein kinase (PKA), which in turn phosphorylates several proteins, among

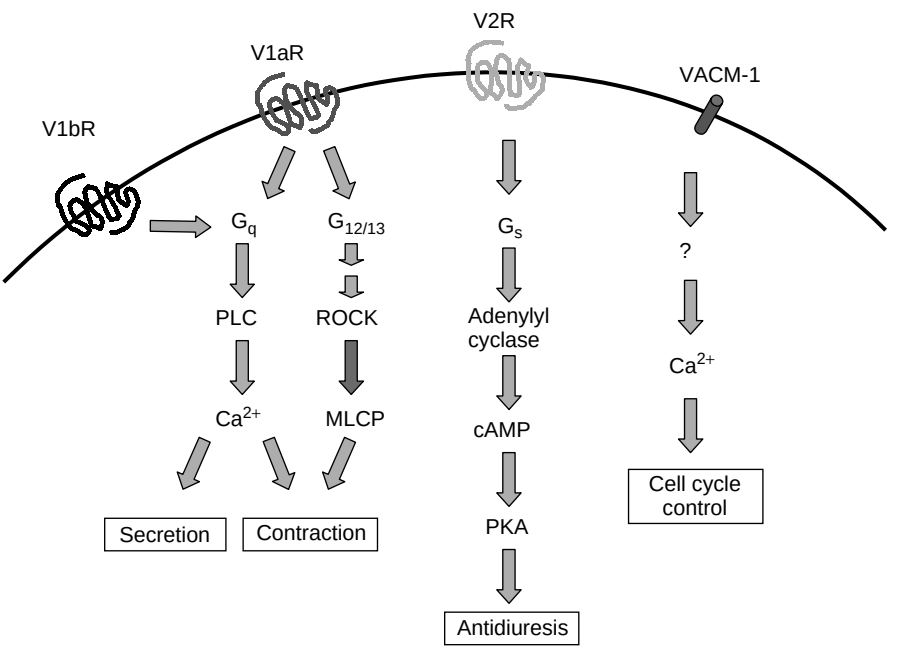

Fig. 5 Synopsis of signaling cascades activated by vasopressin receptors. Shown here are the three heptahelical G protein-coupled vasopressin receptors ($V_{1a}R$, $V_{1b}R$, and V_2R) and the single transmembrane spanning vasopressin-sensitive calcium-mobilizing receptor (VACM-1). *Green arrows* – activation, *red arrow* – inhibition. (See color plate p. xxix.)

others the AVP-sensitive water channel aquaporin-2 (AQP2) and the small G protein Rho A. The cAMP/PKA-mediated signaling events finally result in the translocation of submembranously stored AQP2-harboring vesicles into the apical plasma membrane (Fig. 5; see Sect. 3.3 below).

3 Aquaporins

3.1 Requirements for Water Channel Proteins

In the collecting duct of the kidney, the apical membrane must maintain large chemical and osmotic gradients and therefore be effectively impermeable to ions and most other small molecules. This is achieved by high concentrations of glycosphingolipids and sphingomyelin in the outer leaflet of the epithelial lipid bilayer. Expression and regulated insertion of specific water-permeable transmembrane channels, that is, aquaporins (AQP), represent an elegant way for the organism to regulate the water permeability of these critical barrier interfaces. In response to AVP, the abundance of aquaporins in the apical membrane of principal cells increases, and the epithelia can become about 50-fold more water-permeable. AQPs have little importance in slow water transport, which occurs across plasma membranes for housekeeping functions such as volume regulation. Rather, they are important for rapid vectorial water transport in response to osmotic gradients created by continuous flow or active salt pumping. Consequently, water-transporting proteins are expressed not only in renal collecting duct but also in other fluid-transporting epithelia and endothelia.

3.2 Structural Features Responsible for Water Selectivity

The members of the water channel protein family are homotetramers containing four independent water channels. Aquaporins are characterized by two tandem repeats consisting each of three transmembrane helices and a short loop (Fig. 6). The latter bears the signature motif Asn-Pro-Ala (NPA). The atomic structure of various AQP family members suggests that the NPA motif is crucial for water selectivity. Meeting with their positive ends in the middle of the membrane, the dipoles of the two short loops force the water molecules passing in single file to reorient (Fig. 7). The reorientation disrupts the hydrogen bonds between the two central water molecules. This mechanism is assumed to inhibit Grotthuss proton diffusion involving the chemical exchange of hydrogen nuclei between hydrogen-bonded water molecules in the channel. As revealed by molecular dynamic simulation, another highly conserved residue, a positively charged amino acid Arg187 (Arg-195 in AQP1), may be even more important for the exclusion of protons and ions from the water pathway. Pointing into a channel that is so narrow that water molecules cannot pass each other, its positive charge electrostatically prevents any cation (including H_3O^+) from entering the channel (Fig. 7).

3.3 The Main Aquaporins in Kidney

The archetypal member of the aquaporin family, AQP1, is highly abundant in the proximal tubule, the thin descending limb of Henle and the descending vasa recta. It is exclusively expressed in segments of the nephron that are constitutively and highly

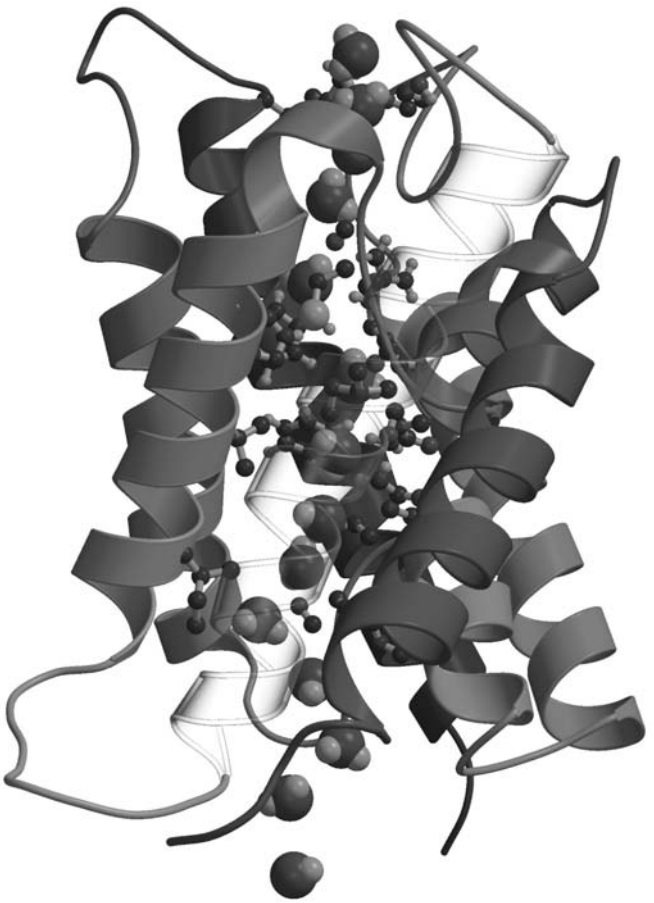

Fig. 6 The structure of human AQP1. The water channel protein is responsible for the efficient water transport in kidney proximal tubule. The pore is formed by six transmembrane helices. Two of the connecting loops fold back into the protein forming two short helices (*blue*). Water molecules (*red* – oxygen, *white* – hydrogen atoms) cross the channel in a single-file configuration. The pore is so narrow that water molecules cannot pass each other. The sequence-related water-transporting protein in the kidney collecting duct, AQP2, is expected to have a similar atomic structure of the water pathway. The figure was kindly provided by B. de Groot (modified from de Groot, B.L., Grubmuller, H. (2001) Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF, *Science* **294**, 2353–2357). (See color plate p. xxxi.)

water permeable. The protein facilitates water movement across both the apical and basolateral membranes. AQP3 and AQP4, in contrast, are found only in the basolateral but not in the apical membrane of collecting duct principal cells. Using a sodium concentration gradient as a driving force, water enters these cells via AQP2

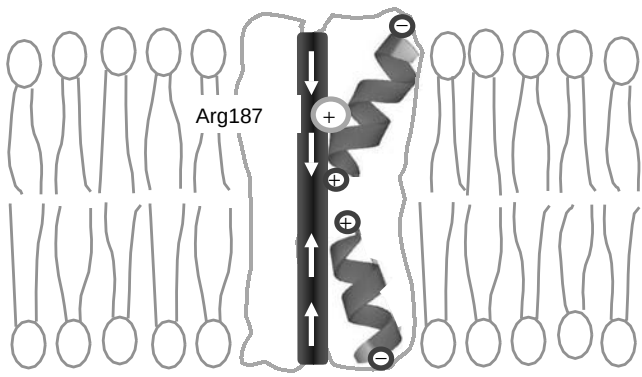

Fig. 7 The selectivity filter of AQP2. The water pathway is shown in blue. Water molecules cross the channel in an oriented fashion. In the middle of the membrane, the water dipoles (*white arrows*) undergo a rotation. It is governed by partial charges from the dipoles of the short helices. Thus, the structure of the central pore region is designed to disrupt the hydrogen bonds between neighboring water molecules. Thereby, protons are prevented from crossing the channel by the Grotthuss mechanism (by structural diffusion along the water wire). The positively charged Arg187 is a barrier for cation movement (including H_3O^+). (See color plate p. xxiii.)

and crosses to the interstitial fluid through AQP3 and AQP4. AQP2 abundance in the apical plasma membrane is hormonally regulated, thus determining the water permeability of the collecting duct. Long- and short-term regulation mechanisms can be distinguished. Long-term AVP treatment increases the expression level of AQP2 and AQP3 in parallel. Short-term regulation is restricted to AQP2. It is accomplished by the AQP2 shuttle (see Sect. 3.4). AQP4 appears not be regulated or at least not as markedly as AQP2 and AQP3.

3.4 The AQP2 Shuttle

Acute regulation of body water balance requires mechanisms that are much faster than the augmentation of protein expression levels. The rapid increase in collecting duct water permeability, that is, the fast increase of AQP2 abundance in the plasma membrane, results from exocytosis of AQP2-containing intracellular vesicles (Fig. 8). In the first step, AVP binds to the V_2 receptor expressed on the basolateral membrane of renal principal cells. Stimulation of the V_2 receptor leads to a rise of cAMP and cytoplasmic Ca^{2+}. The subsequent phosphorylation of AQP2 initiates a long-range AQP2 redistribution from vesicles distributed throughout the cell to the apical plasma membrane. The prominent terminal web of actin filaments in the apical part of principal cells appear to be involved in the AQP2 translocation, since disruption of microfilaments with cytochalasins stimulates the plasma membrane insertion. The final step is the exocytotic insertion of AQP2 into the plasma membrane resulting in an increase in the osmotic water permeability (Fig. 8). Trafficking to the plasma membrane is impaired by most mutations in AQP2, even though some of the mutants will act as functional water channels (see Sect. 4.2).

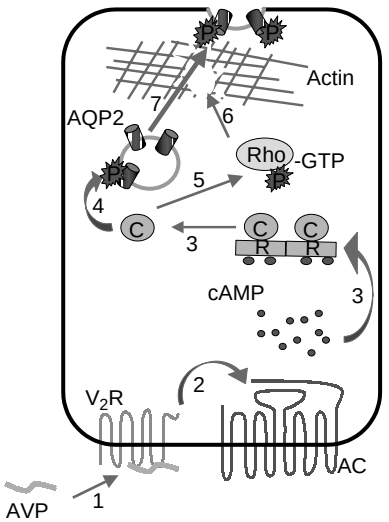

Fig. 8 The AQP2 shuttle. In a first step (1), the antidiuretic hormone AVP binds to the V_2 receptor (V_2R) expressed in principal cells, which are the main epithelial cell type of the medullary collecting duct. Stimulation of the V_2 receptor on the basolateral surface of principal cells activates the adenylyl cyclase (2). The increased cyclic AMP levels, in turn, activate PKA (3). Substrates of PKA are AQP2 (4) and Rho (5), which are phosphorylated. The phosphorylation of Rho favors the depolymerization of the F-actin cytoskeleton (6) and allows translocation of AQP2-bearing vesicles to the cell membrane (7). Subsequent fusion of the vesicles with the plasma membrane increases the membrane abundance of AQP2, thereby facilitating osmotic water transport.

4 Diabetes Insipidus

Diabetes insipidus is an acquired or inherited disorder of water homeostasis, characterized by polyuria and excessive thirst. It is either caused by an insufficient release of AVP from the posterior pituitary (central diabetes insipidus) or by a resistance of the kidney to the action of AVP (nephrogenic diabetes insipidus). The urinary volume exceeds 50 mL kg^{-1} bodyweight in 24 h. The specific gravity and the osmolality of the urine are less than 1010 g L^{-1} and 300 mosmol L^{-1} respectively. In severe cases, the volume exceeds 200 mL kg^{-1} bodyweight in 24 h, and the urinary osmolality is less than 100 mosmol L^{-1}. In these cases, the patients life is ruled by frequent drinking and by micturition every 30 to 60 min both day and night. The large amounts of urine may cause a dilation of the urinary tract resulting in hydronephrosis and renal insufficiency. Infants are under particular threat, since insufficient fluid intake results in severe dehydration with plasma osmolality and plasma sodium exceeding 320 mosmol L^{-1} (>10% increase in osmolality) and 160 mmol L^{-1} respectively. Recurrent episodes of dehydration result in a failure to thrive and may cause mental retardation. Both sequelae can be completely prevented by sufficient fluid intake.

4.1 Central Diabetes Insipidus

Central diabetes insipidus is characterized by an inadequate or no release of AVP from the posterior pituitary. Frequent causes of central diabetes inspidus are brain tumors, pituitary surgery, and severe head injuries, which disrupt the axonal projections of the AVP-producing magnocellular neurons. Depending on the neuronal lesion, diabetes is transient (regeneration of the magnocellular neurons) or persistent (degeneration of neurons). In idiopathic central diabetes insipidus, which typically manifests in childhood and early adult life, the cause is unknown.

Congenital forms of central diabetes insipidus are rare. A comprehensive list of known mutations in the *AVP-NPII* gene is presented on the web (http://www.medicine.mcgill.ca/nephros/avp_npii.html). In most of the cases, an autosomal dominant mode of inheritance is found. So far, only three cases of autosomal recessive central diabetes insipidus have been reported.

4.1.1 Autosomal Dominant Central Diabetes Insipidus

The congenital forms of central diabetes insipidus show an autosomal dominant mode of inheritance. The penetrance is almost 100%. While newborns and infants rarely present with symptoms, polyuria and polydipsia typically become apparent in children older than one year (the onset can vary between 1 and 30 years of age).

Genetic analysis of the autosomal dominant central diabetes insipidus have revealed about 40 different mutations within the *AVP-NPII* gene. The majority of mutations (33 of 40) are found in the region encoding the neurophysin II moiety. Five mutations lead to an altered amino acid sequence of the signal peptide, and one mutation causes an amino acid substitution within the AVP moiety (Fig. 9). The great majority of AVP-NPII mutants is characterized by amino acid substitutions in the neurophysin II moiety. This may be explained by the fact that neurophysin II is not only a carrier protein for AVP but also of essential importance for the correct folding and processing of AVP in the precursor.

The molecular mechanisms underlying the dominant mode of inheritance are still not fully understood. The delayed onset of the disease argues against a dominant negative effect of the mutant AVP-NPII precursor. It is more likely that secondary events elicited by mutant AVP-NPII proteins contribute to the pathomechanisms of the disease.

Several lines of evidence suggest that the mutant AVP-NPII precursor proteins have neuropathic or neurotoxic effects on AVP-producing magnocellular neurons. This is supported by neurohistological studies, in which selective

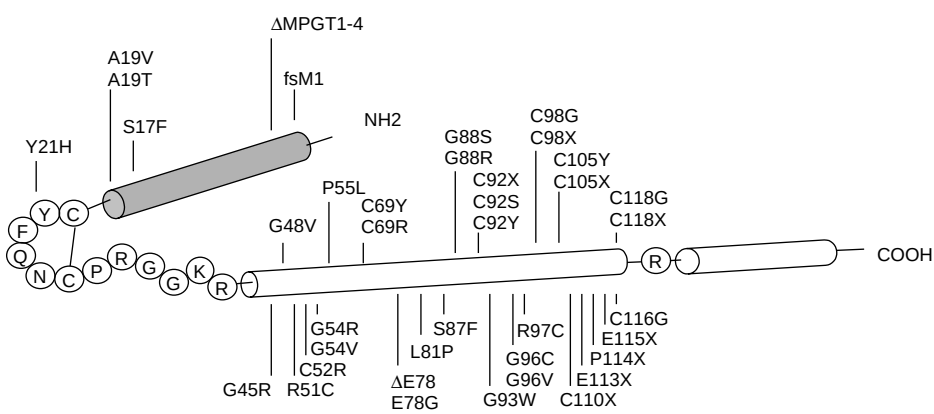

Fig. 9 Synopsis of AVP-NPII mutants responsible for autosomal dominant central diabetes insipidus. The mutants are depicted by the native amino acid (single letter code), the position in the AVP-NPII precursor protein, and the replacing amino acid. Δ – deletion of amino acids, fs – indicates nucleotide deletions causing a shift in the reading frame.

degeneration of AVP-producing but not of oxytocin-producing magnocellular neurons was found. *In vitro* analysis of differentiated neuronal cell lines that stably express mutant AVP-NPII precursor proteins demonstrated strong accumulation of AVP-NPII mutants within the ER and progressive cell death. The mechanisms involved in this process have not been explored in detail. However, recent studies on different secretory and membrane proteins suggest that ER accumulation causes cellular stress (unfolded protein response), which results in a number of effects including an increased synthesis of chaperones. The chaperones help to prevent the aggregation of misfolded proteins. If the native state of proteins is not achieved in subsequent folding cycles, degradation via the ubiquitin-proteasome system is initiated. However, this degradation can be impaired by several factors. For example, some mutant proteins have a high probability of forming aggregates spontaneously, as found for certain mutants of serine protease inhibitors. The formation of such aggregates prevents degradation of the misfolded proteins and is cytotoxic. Similarily, the excessive synthesis of a misfolded protein (like mutant rhodopsin in retinitis pigmentosa) also affects the cellular degradation machinery by unknown mechanisms.

4.1.2 Autosomal Recessive Central Diabetes Insipidus

So far, only three cases of congenital diabetes insipidus are known that show an autosomal recessive mode of inheritance. In all three cases, the patients presented with symptoms shortly after birth. In one case, a homozygous splice site mutation was identified. In a second case, a compound heterozygous mutation (a splice site and a nonsense mutation) in the *AVP-NPII* gene was found. In a third case, a homozygous *AVP-NPII* gene mutation encodes a mutant AVP, in which proline at residue 7 is replaced by leucine. This AVP mutant has a 30-fold lower affinity to the V_2 receptor than the native AVP. The parents, both heterozygous for the mutation, were reported to be healthy.

4.2 Nephrogenic Diabetes Insipidus

Nephrogenic diabetes insipidus is characterized by the resistance of the kidney to the action of AVP. Patients show polyuria and excessive thirst despite the presence of normal or elevated amounts of plasma AVP. Acquired forms of nephrogenic diabetes insipidus are much more frequent than inherited forms. The most common cause of nephrogenic diabetes insipidus is lithium, which is used for the treatment of bipolar manic-depressive illness. Polyuria can be found in up to 20% of patients receiving lithium. In most cases, polyuria is reversible after cessation of the therapy. However, in some patients, renal insufficiency can develop despite the cessation of lithium therapy. Patients presenting with impaired renal function as a result of long-term lithium therapy (creatinine serum levels >2.5 mg dL^{-1}) are under high risk of developing endstage renal failure.

Hypokalemia and hypercalcemia are metabolic disorders associated with less severe forms of nephrogenic diabetes insipidus, which are fully reversible when serum potassium or calcium values are normalized. In some instances, the antifungal amphotericin B or the microtubule inhibiting agents vincristine and colchicine can cause severe, but reversible nephrogenic diabetes insipidus.

Congenital diabetes insipidus is a rare disease with an estimated incidence of 8.8 per 1 000 000 males. Two genes have

been identified to be responsible for the disease: the genes encoding the G protein-coupled vasopressin V_2 receptor and the AVP-sensitive water channel AQP2. Inactivating mutations of the V_2 receptor show an X-linked mode of inheritance. Mutations of the AQP2 show an autosomal recessive or autosomal dominant trait. The onset and the symptoms are indistinguishable in X-linked and autosomal nephrogenic diabetes insipidus. Thus, an interruption of the signal cascade at the receptor or the effector level has identical consequences.

4.2.1 X-linked Nephrogenic Diabetes Insipidus

To date, more than 170 different mutations within the V_2 receptor gene that cause X-linked nephrogenic diabetes insipidus have been identified (refer to http://www.medicine.mcgill.ca/nephros/avpr2.html for a comprehensive list of reported mutations). The mutations are scattered throughout the coding region without evidence for a hypermutable region. About 50% of the mutations are missense mutations, leading to the exchange of a single amino acid. Roughly 10% are nonsense mutations resulting in a premature stop of the polypeptide synthesis. The remaining 40% of the mutations are deletions and insertions. Strikingly, most of the single amino acid substitutions (Fig. 10, green circles) are localized within the transmembrane regions. Further clusters of amino acid exchanges are found in the first extracellular loop and in the N- and C-terminal parts of the second extracellular loop.

In recent yeras, a large number of V_2 receptor mutants have been characterized for their functional defects. In general, three main defects have been identified: (1) a failure to reach the cell surface, (2) a decrease or loss of AVP-binding, and/or (3) a decrease or loss in G protein-coupling activity. The most common defect found in about two-thirds of the investigated mutants is an impaired transport to the cell surface, most likely mediated by the quality control system of the ER. The quality control system consists of a variety of different chaperones, which recognize and bind misfolded proteins. Thereby, the chaperons prevent the exit of misfolded proteins from the ER. If the native conformation cannot be formed, the misfolded proteins are finally degraded via the ubiquitin-proteasome system.

Amino acid substitutions leading to a defect in the binding of AVP are predominantly found in the first and second extracellular loop or at the junction of the transmembrane region and the extracellular space. This observation highlights the importance of the first and second extracellular loop of the V_2 receptor for the binding of AVP. In photoaffinity labeling experiments, threonine 103 and arginine 106 of the bovine V_2 receptor have been identified to be of crucial importance in ligand binding (Fig. 10, asterisks).

So far, only few V_2 receptor mutants have been identified that show a reduced ability or are unable to activate the G protein. In one mutant (in which arginine 137 at the junction of the second transmembrane domain, and the second intracellular loop is replaced by histidine; R137H), the loss of G protein-coupling could be assigned to a constitutive binding of the mutant receptor to the adapter protein β-arrestin.

4.2.2 Autosomal Recessive Nephrogenic Diabetes Insipidus

Genetic defects in the *AQP2* gene account for about 10% of congenital nephrogenic diabetes insipidus. The human *AQP2* gene

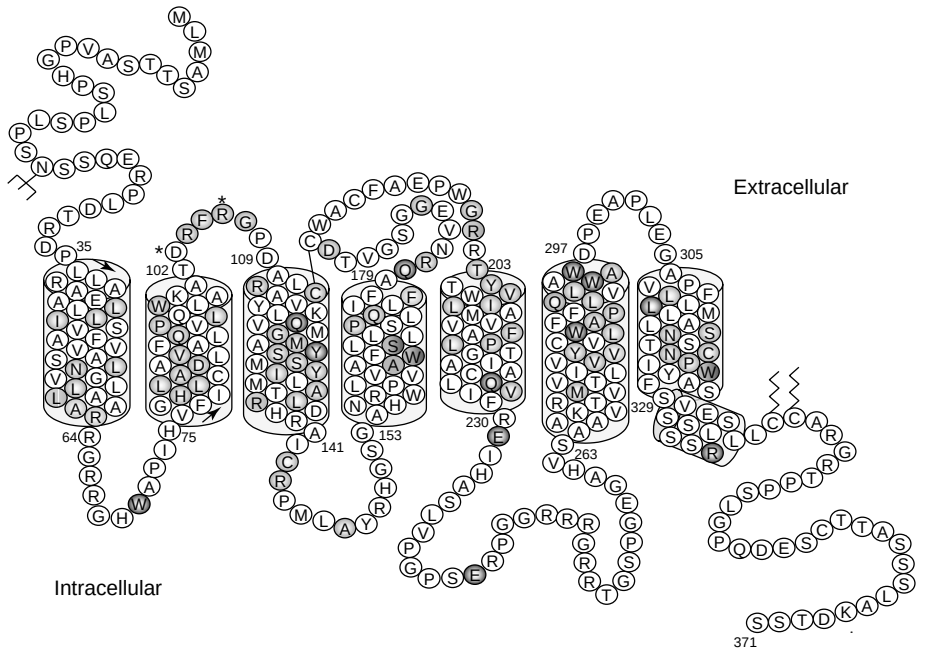

Fig. 10 Synopsis of vasopressin V_2 receptor mutants responsible for X-linked nephrogenic diabetes insipidus. Diagram of the human vasopressin V_2 receptor. Amino acids are shown in the single-letter code. The arrows in the first and second transmembrane domain show the succession of amino acids. *Stars* – indicate amino acids involved in ligand binding as demonstrated by photoaffinity labeling, *green circles* – missense mutations, *red circles* – nonsense mutations, *yellow circles* – in-frame deletions, *blue circle* – in-frame insertion, *yellow/green circles* – missense mutations and in-frame deletions found at this position, *red/green circles* – missense and nonsense mutations found at this position. Posttranslational modifications shown are glycosylation at asparagine 22, disulfide bonds between cysteine 112 and 192, and palmitoylation at cysteines 341 and 342. (See color plate p. xxx).

is located in the chromosome region 12q13 and has four exons and three introns. The majority of mutations show an autosomal recessive mode of inheritance, although autosomal dominant traits (see Sect. 4.2.3) are also possible. Most of the mutations are missense mutations, followed by frameshift mutations and seldomly occurring nonsense and splice site mutations. The mutations are located throughout the gene, affecting all transmembrane helices and connecting loops (Fig. 11). A comprehensive list of known *AQP-2* gene mutations is presented on the web (http://www.medicine.mcgill.ca/nephros/aqp2.html). Mutations found in the two loops bearing the signature motif NPA are likely to destroy the water-transporting part of the protein. For example, substitution of Arg187 for Cys disrupts the normal folding of the narrow constriction in the aqueous pore (Fig. 7). All mutant AQP2 proteins are retained in the ER (Fig. 12), although some of the mutants appear to be able to function as water channels. Thus, misrouting of mutant AQP2 proteins is the major cause of autosomal recessive nephrogenic diabetes insipidus.

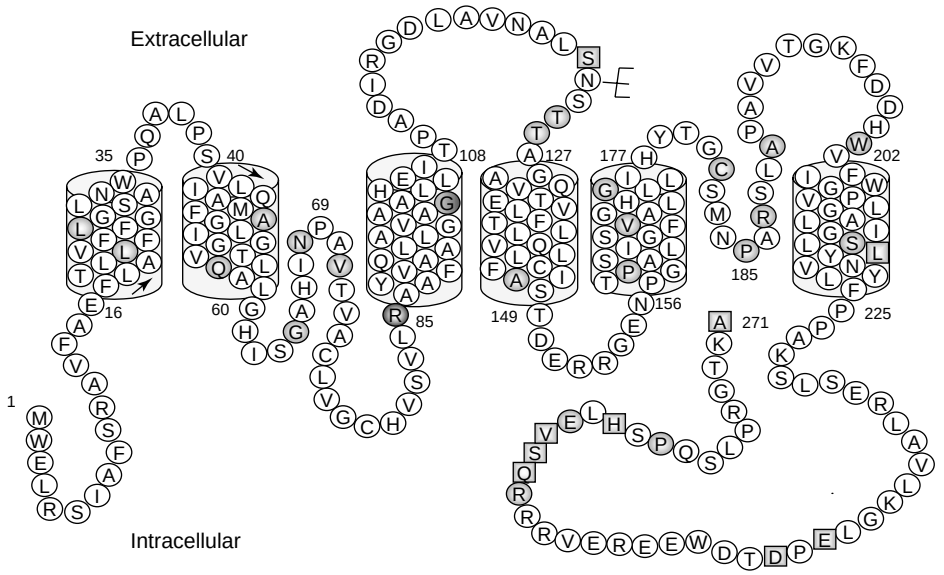

Fig. 11 Synopsis of mutant AQP2 water channel responsible for autosomal nephrogenic diabetes insipidus. Diagram of the human AQP2. Amino acids are shown in the single-letter code. The arrows in the first and second transmembrane domain show the succession of amino acids. Posttranslational modifications shown are glycosylation at asparagine 123. Mutations responsible for autosomal recessive diabetes insipidus: *Green circles* – missense mutations, *green squares* – nucleotide deletions causing a shift in the reading frame, *red circles* – nonsense mutations, *red/green circle* – missense and nonsense mutations found at this position. Mutations responsible for autosomal dominant nephrogenic diabetes insipidus: *blue circles* – missense mutations, *blue squares* – nucleotide deletions/insertions causing a shift in the reading frame. (See color plate p. xxix.)

4.2.3 Autosomal Dominant Nephrogenic Diabetes Insipidus

In some instances, nephrogenic diabetes insipidus shows autosomal dominant inheritance. In these cases, mutant AQP2 proteins inhibit the normal function of the coexpressed wild-type AQP2. Here, heterooligomerization between the mutant AQP2 and wild-type AQP2 occurs, which is not found for "recessive" AQP2 mutants. However, the heterotetramers do not show AVP-regulated translocation to the apical plasma membrane (dominant negative effect of mutant AQP2). The AQP2 mutant in dominant nephrogenic diabetes insipidus, E258K-AQP2, for example, is a functional water channel but it is retained in the Golgi complex. Misrouting to late endosomes/lysosomes or to the basolateral membrane may also occur (Fig. 12). All dominant forms of nephrogenic diabetes insipidus are due to mutations in the intracellular C terminus near the site that regulates exocytosis.

4.3 Treatment of Diabetes Insipidus

4.3.1 Treatment of Central Diabetes Insipidus

Central diabetes insipidus can be efficiently treated by the substitution of AVP. Since AVP also has vasopressor activity, the selective V_2

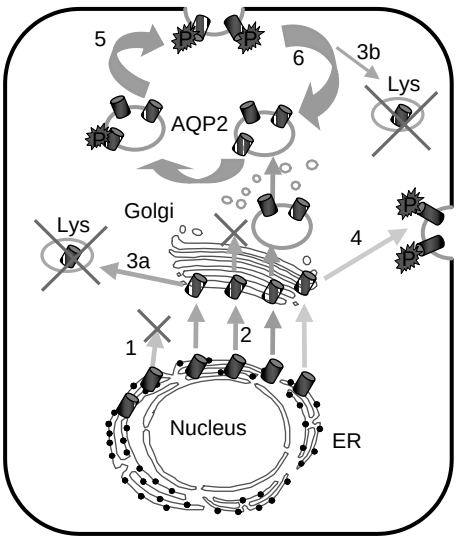

Fig. 12 Synopsis of routing defects of mutant AQP2 water channels. AQP2 mutants associated with autosomal recessive nephrogenic diabetes insipidus are retained within the endoplasmic reticulum (1, *ochre arrow*). Since the "recessive" AQP2 mutants exist as monomers, they do not affect the sorting of the wild-type AQP2. In contrast, AQP2 mutants associated with autosomal dominant nephrogenic diabetes insipidus affect sorting of the wild-type AQP2, since dominant AQP2 mutants and wild-type AQP2 form heterotetramers. So far, three different sorting defects of "dominant-negative" AQP2 mutants have been characterized: the mutant E258K-AQP2 is retained in the Golgi apparatus (2, *blue arrows*), the mutant ΔG727-AQP2 is constitutively transported to lysosomes (3a and 3b, *orange arrows*), and the mutant insA779-AQP2 undergoes PKA-dependent translocation to the basolateral plasma membrane (4, *cyan arrows*). Normal trafficking of AQP2 (*green arrows*) is not finished with its exocytic insertion (5) but also includes its endocytic retrieval (6). (See color plate p. xxvii.)

receptor agonist 1-deamino-8-D-arginine vasopressin (DDAVP, Minirin®) is preferred for hormonal replacement therapy. No serious adverse reactions of this treatment have been reported so far. Doses range from 5 to 60 μg each day.

4.3.2 Treatment of Nephrogenic Diabetes Insipidus

The most important treatment for nephrogenic diabetes insipidus is sufficient water intake adjusted to the increased urinary output. When recurrent episodes of hypernatremia are efficiently prevented, affected children develop normally. In the long run, control of the urinary output is of major importance. The large urinary volumes may lead to the dilation of the urinary system resulting in hydronephrosis and renal insufficiency. Currently, urinary volumes can be reduced by thiazide diuretics and/or inhibitors of the prostaglandin synthesis (inhibitors of cyclooxygenases – COXs). Thiazide diuretics cause a depletion of body sodium and a reduction in the glomerular filtration rate. As a consequence, the efficiency of thiazides in reducing urinary output is impaired by increased dietary sodium intake. Therefore, patients with congenital nephrogenic diabetes insipidus should receive a low sodium diet. Nutrition guidelines and patient-oriented information on nephrogenic diabetes insipidus are available on the web (http://www.ndif.org). Nonselective COX inhibitors reduce the glomerular filtration rate, thereby leading to a reduced urinary output. However, in patients with congestive heart failure, COX inhibitors impair renal function. In addition, nonselective COX inhibitors, like indomethacin, are associated with severe adverse drug reactions, such as gastrointestinal bleeding and headache. Such adverse reactions may not be seen with selective inhibitors of COX-2. Recent studies provided evidence that selective COX-2 inhibitors reduce urinary output in another congenital polyuric disease (antenatal Bartter/hyperprostaglandin E

syndrome) without evidence for serious adverse reactions. At present, it is not clear whether selective COX-2 inhibitors will be safe and efficient drugs for the treatment of nephrogenic diabetes insipidus.

5 Animal Models

5.1 Animal Models of Central Diabetes Insipidus

The Brattleboro rat shows central diabetes insipidus due to a spontaneous mutation in the *AVP-NPII* gene. Only rats homozygous for this mutation have diabetes insipidus, whereas rats heterozygous for the mutation are normal. Thus, Brattleboro rats show an autosomal recessive mode of inheritance. This model has been of greatest importance for the understanding of AVP-mediated antidiuresis and still is of value for the analysis of AVP-induced gene expression. However, the model does not contribute to the understanding of the cellular pathomechanisms underlying the autosomal dominant form of central diabetes insipidus.

5.2 Animal Models of Nephrogenic Diabetes Insipidus

To establish a model of X-linked nephrogenic diabetes insipidus, a premature stop mutation was introduced into codon 242 of the V_2 receptor gene (*E242X*). This mutation was originally identified in patients suffering from X-linked nephrogenic diabetes insipidus. The mutated gene encodes a V_2 receptor with only five instead of seven transmembrane domains. Male mice, hemizygous for the E242X mutant V_2 receptor are polyuric, but die within three weeks after birth. So far, treatment of these mice with thiazides and/or COX inhibitors did not improve survival. To establish a model of autosomal recessive nephrogenic diabetes insipidus, a missense mutation was introduced into codon 126 of the murine *AQP2* gene, which leads to the replacement of threonine 126 by methionine (T126M). This T126M mutant was identified in patients suffering from autosomal recessive nephrogenic diabetes insipidus. Mice homozygous for AQP2-T126M die within two weeks after birth.

6 Outlook

While central diabetes insipidus can be adequately treated by the administration of AVP or DDAVP, treatment of nephrogenic diabetes insipidus still poses a serious problem. Although urinary volume can be reduced by about one-third using thiazide diuretics or COX inhibitors, patients still pass large amounts of dilute urine. Frequent micturition, even throughout the night, affects the quality of life. The discovery of the genetic defects underlying nephrogenic diabetes insipidus allowed a firm diagnosis of the patients. However, this knowledge did not result in an improvement in the therapy.

An obstacle in improving the therapy of diabetes insipidus is a high degree of diversity in the underlying genotypes. There is no predominant mutation, as found, for example, in cystic fibrosis: Here, more than 60% of all patients encode the ΔF508 CFTR mutant, so that

research can focus on this particular mutation. Despite these obstacles, a new therapeutic concept has been presented recently, which could be useful in patients encoding misfolded V_2 receptor mutants. Using the selective V_2 receptor antagonist SR 121463, a functional rescue of several different misfolded V_2 receptor mutants has been demonstrated. In line with this biochemical analysis, a clinical trial with patients suffering from congenital nephrogenic diabetes insipidus demonstrated that the V_1 receptor antagonist SR49059 reduced urinary output by 50% when compared to untreated controls.

It is assumed that the pharmacological agents support proper receptor folding, thus allowing the exit from the ER. Therefore, drugs with an ability to promote receptor folding are also termed *pharmacological chaperones* or *pharmacochaperones*. Whether these pharmacochaperones have the potential to be used broadly in nephrogenic diabetes insipidus remains to be seen. Another potential option for treatment, applicable to all patients with X-linked nephrogenic diabetes insipidus, is the stimulation of the cAMP/PKA-signaling cascade downstream of the V_2 receptor. PDEs, which degrade intracellular cAMP, represent one possible drug target. However, earlier clinical trials with rolipram, a selective inhibitor of the type 4 PDE, did not result in the reduction of the urinary output. Finally, gene therapy might be considered as a further treatment option. However, from the current point of view, gene therapy, in general, lacks safety.

See also Genetics, Molecular Basis of; Medicinal Chemistry.

Bibliography

Books and Reviews

Bankir, L. (2001) Antidiuretic action of vasopressin: quantitative aspects and interaction between V1a and V2 receptor-mediated effects, *Cardiovasc. Res.* **51**, 372–390.

Barak, L.S., Oakley, R.H., Laporte, S.A., Caron, M.G. (2001) Constitutive arrestin-mediated desensitization of a human vasopressin receptor mutant associated with nephrogenic diabetes insipidus, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 93–98.

Breyer, M.D., Ando, Y. (1994) Hormonal signaling and regulation of salt and water transport in the collecting duct, *Annu. Rev. Physiol.* **56**, 711–739.

Carrell, R.W., Lomas, D.A. (2002) α1-antitrypsin deficiency – a model for conformational diseases, *N. Engl. J. Med.* **346**, 45–53.

Chevet, E., Cameron, P.H., Pelletier, M.F., Thomas, D.Y., Bergeron, J.J. (2001) The endoplasmic reticulum: integration of protein folding, quality control, signaling and degradation, *Curr. Opin. Struct. Biol.* **11**, 120–124.

Jackson, E.K. (2001) Vasopressin and Other Agents Affecting the Renal Conservation of Water, in: Hardman, J.G., Limbird, L.E. (Eds.) *Goodman & Gilman's – The Pharmacological Basis of Therapeutics*, 10th edition, McGraw-Hill, New York, pp. 789–808.

Klussman, E., Maric, K., Rosenthal, W. (2000) The mechanisms of aquaporin control in the renal collecting duct, *Rev. Physiol. Biochem. Pharmacol.* **141**, 33–95.

Nielsen, S., Frokiaer, J., Marples, D., Kwon, T.H., Agre, P., Knepper, M.A. (2002) Aquapo rins in the kidney: from molecules to medicine, *Physiol. Rev.* **82**, 205–244.

Oksche, A., Rosenthal, W. (1998) The molecular basis of nephrogenic diabetes insipidus, *J. Mol. Med.* **76**, 326–337.

Robertson, G.L. (1992) Vasopressin, in: Seldin, D.W., Giebisch, G. (Eds.) *The Kidney*, Lippincott Williams & Wilkens, Philadelphia, pp. 1595–1613.

Scharrer, E., Scharrer, B. (1954) Neurosekretion, in: Hrsg.: Bargmann, W. *Handbuch der mikroskopischen Anatomie des Menschen*, Bd VI/5, Springer-Verlag, Berlin, Göttingen, Heidelberg, pp. 953–1066.

Primary Literature

Agre, P., Preston, G.M., Smith, B.L., Jung, J.S., Raina, S., Moon, C., Guggino, W.B., Nielsen, S. (1993) Aquaporin CHIP: the archetypal molecular water channel, *Am. J. Physiol.* **265**, F463–F476.

Arthus, M.F., Lonergan, M., Crumley, M.J., Naumova, A.K., Morin, D., De Marco, L.A., Kaplan, B.S., Robertson, G.L., Sasaki, S., Morgan, K., Bichet, D.G., Fujiwara, T.M. (2000) Report of 33 novel AVPR2 mutations and analysis of 117 families with X-linked nephrogenic diabetes insipidus, *J. Am. Soc. Nephrol.* **11**, 1044–1054.

Bebok, Z., Mazzochi, C., King, S.A., Hong, J.S., Sorscher, E.J. (1998) The mechanism underlying cystic fibrosis transmembrane conductance regulator transport from the endoplasmic reticulum to the proteasome includes Sec61beta and a cytosolic, deglycosylated intermediary, *J. Biol. Chem.* **273**, 29873–29878.

Bence, N.F., Sampat, R.M., Kopito, R.R. (2001) Impairment of the ubiquitin-proteasome system by protein aggregation, *Science* **292**, 1552–1555.

Bergeron, C., Kovacs, K., Ezrin, C., Mizzen, C. (1991) Hereditary diabetes insipidus: an immunohistochemical study of the hypothalamus and pituitary gland, *Acta. Neuropathol. Berl.* **81**, 345–348.

Beuret, N., Rutishauser, J., Bider, M.D., Spiess, M. (1999) Mechanism of endoplasmic reticulum retention of mutant vasopressin precursor caused by a signal peptide truncation associated with diabetes insipidus, *J. Biol. Chem.* **274**, 18965–18972.

Bichet, D.G., Arthus, M.F., Lonergan, M., Hendy, G.N., Paradis, A.J., Fujiwara, T.M., Morgan, K., Gregory, M.C., Rosenthal, W., Didwania, A., Antaramian, A., Birnbaumer, M. (1993) X-linked nephrogenic diabetes insipidus mutations in North America and the Hopewell hypothesis, *J. Clin. Invest.* **92**, 1262–1268.

Bichet, D.G., Ruel, N., Arthus, M.F., Lonergan, M., Hendy, G.N., Bichet, D.G. (1990) Rolipram, a phosphodiesterase inhibitor, in the treatment of two male patients with congenital nephrogenic diabetes insipidus, *Nephron* **56**, 449–450.

Birnbaumer, M., Seibold, A., Gilbert, S., Ishido, M., Barberis, C., Antaramian, A., Brabet, P., Rosenthal, W. (1992) Molecular cloning of the receptor for human antidiuretic hormone, *Nature* **357**, 333–335.

Boton, R., Gaviria, M., Batlle, D.C. (1987) Prevalence, pathogenesis, and treatment of renal dysfunction associated with chronic lithium therapy, *Am. J. Kidney Dis.* **10**, 329–345.

Brakch, N., Rholam, M., Boussetta, H., Cohen, P. (1993) Role of β-turn in proteolytic processing of peptide hormone precursors at dibasic sites, *Biochemistry* **32**, 4925–4930.

Braverman, L.E., Mancini, J.P., McGoldrick, D.M. (1965) Hereditary idiopathic diabetes insipidus: a case report with autopsy findings, *Ann. Intern. Med.* **63**, 503–508.

Burnatowska-Hledin, M., Lazdins, I.B., Listenberger, L., Zhao, P., Sharangpani A., Folta, V., Card, B. (1999) VACM-1 receptor is specifically expressed in rabbit vascular endothelium and renal collecting tubule, *Am. J. Physiol.* **276**, F199–F209.

Byrd, P.J., Stankovic, T., McConville, C.M., Smith, A.D., Cooper, P.R., Taylor, A.M. (1997) Identification and analysis of expression of human VACM-1, a cullin gene family member located on chromosome 11q22-23, *Genome Res.* **7**, 71–75.

Charmandari, E., Brook, C.G.D. (1999) 20 years of experience in idiopathic central diabetes insipidus, *Lancet* **353**, 2212–2213.

Chen, L.Q., Rose, J.P., Breslow, E., Yang, D., Chang, W.R., Furey, W.F. Jr., Sax, M., Wang, B.C. (1991) Crystal structure of a bovine neurophysin II dipeptide complex at 2.8 A determined from the single-wavelength anomalous scattering signal of an incorporated iodine atom, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4240–4244.

Cheng, A., van Hoek, A.N., Yeager, M., Verkman, A.S., Mitra, A.K. (1997) Three-dimensional organization of a human water channel, *Nature* **387**, 627–630.

Christensen, S., Kusano, E., Yusufi, A.N., Murayama, N., Dousa, T.P. (1985) Pathogenesis of nephrogenic diabetes insipidus due to chronic administration of lithium in rats, *J. Clin. Invest.* **75**, 1869–1879.

Deen, P.M., Verdijk, M.A., Knoers, N.V., Wieringa, B., Monnens, L.A., van Os, C.H., van Oost, B.A. (1994) Requirement of human renal water channel aquaporin-2 for vasopressin-dependent concentration of urine, *Science* **264**, 92–95.

de Groot, B.L., Grubmuller, H. (2001) Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF, *Science* **294**, 2353–2357.

Fassina, G., Chaiken, I.M. (1988) Structural requirements of peptide hormone binding for peptide-potentiated self-association of bovine neurophysin II, *J. Biol. Chem.* **263**, 13539–13543.

Fushimi, K., Sasaki, S., Marumo, F. (1997) Phosphorylation of serine 256 is required for cAMP-dependent regulatory exocytosis of the aquaporin-2 water channel, *J. Biol. Chem.* **272**, 14800–14804.

Fushimi, K., Sasaki, S., Yamamoto, T., Hayashi, M., Furukawa, T., Uchida, S., Kuwahara, M., Ishibashi, K., Kawasaki, M., Kihara, I., Marumo, F. (1994) Functional characterization and cell immunolocalization of AQP-CD water channel in kidney collecting duct, *Am. J. Physiol.* **267**, F573–F582.

Gelman, M.S., Kannegaard, E.S., Kopito, R.R. (2002) A principal role for the proteasome in ER-associated degradation of misfolded intracellular CFTR, *J. Biol. Chem.* **277**, 11709–11714.

Guillon, G., Butlen, D., Cantau, B., Barth, T., Jard, S. 1982 Kinetic and pharmacological characterization of vasopressin membrane receptors from human kidney medulla: relation to adenylate cyclase activation, *Eur. J. Pharmacol.* **85**, 291–304.

Ito, M., Jameson, J.L., Ito, M. (1997) Molecular basis of autosomal dominant neurohypophyseal diabetes insipidus. Cellular toxicity caused by the accumulation of mutant vasopressin precursors within the endoplasmic reticulum, *J. Clin. Invest.* **99**, 1897–1905.

Johnston, J.A., Ward, C.L., Kopito, R.R. (1998) Aggresomes: a cellular response to misfolded proteins, *J. Cell Biol.* **143**, 1883–1898.

Kamsteeg, E.J., Wormhoudt, T.A.M., Rijss, J.P.L., van Os, C.H., Deen, P.M.T. (1999) An impaired routing of wild-type aquaporin-2 after tetramerization with an aquaporin-2 mutant explains dominant nephrogenic diabetes insipidus, *EMBO J.* **18**, 2394–2400.

Kaufmann, J.E., Oksche, A., Wollheim, C.B., Günther, G., Rosenthal, W., Vischer, U.M. (2000) Vasopressin-induced von Willebrand factor secretion from endothelial cells involves V2 receptors and cAMP, *J. Clin. Invest.* **106**, 107–116.

Klussmann, E., Maric, K., Wiesner, B., Beyermann, M., Rosenthal, W. (1999) Protein kinase A anchoring proteins are required for vasopressin-mediated translocation of aquaporin-2 into cell membranes of renal principal cells, *J. Biol. Chem.* **274**, 4934–4938.

Kojro, E., Eich, P., Gimpl, G., Fahrenholz, F. (1993) Direct identification of an extracellular agonist binding site in the renal V2 vasopressin receptor, *Biochemistry* **32**, 13537–13544.

Land, H., Schutz, G., Schmale, H., Richter, D. (1982) Nucleotide sequence of cloned cDNA encoding bovine arginine vasopressin-neurophysin II precursor, *Nature* **295**, 299–303.

Liard, J.F. (1989) Peripheral vasodilatation induced by a vasopressin analogue with selective V2-agonism in dogs, *Am. J. Physiol.* **256**, H1621–H1626.

Liebenhoff, U., Rosenthal, W. (1995) Identification of Rab3-, Rab5a- and synaptobrevin II-like proteins in a preparation of rat kidney vesicles containing the vasopressin-regulated water channel, *FEBS Lett.* **365**, 209–213.

Lorenz, D., Krylov, A., Hahm, D., Hagen, V., Rosenthal, W., Pohl, P., Maric, K. (2003) Cyclic AMP is sufficient for triggering the exocytic recruitment of aquaporin-2 in renal epithelial cells, *EMBO Rep.* **4**, 88–93.

Klussmann, E., Tamma, G., Lorenz, D., Wiesner, B., Maric, K., Hofmann, F., Aktories, K., Valenti, G., Rosenthal, W. (2001) An inhibitory role of Rho in the vasopressin-mediated translocation of aquaporin-2 into cell membranes of renal principal cells, *J. Biol. Chem.* **276**, 20451–20457.

Krylov, A.V., Pohl, P., Zeidel, M.L., Hill, W.G. (2001) Water permeability of asymmetric planar lipid bilayers: leaflets of different composition offer independent and additive resistances to permeation, *J. Gen. Physiol.* **118**, 333–340.

Ma, T., Yang, B., Gillespie, A., Carlson, E.J., Epstein, C.J., Verkman, A.S. (1998) Severely impaired urinary concentrating ability in transgenic mice lacking aquaporin-1 water channels, *J. Biol. Chem.* **273**, 4296–4299.

Mannucci, P.M., Aberg, M., Nilsson, I.M., Robertson, B. (1975) Mechanism of plasminogen activator and factor VIII increase after vasoactive drugs. *Br. J. Haematol.* **30**, 81–93.

Marr, N., Bichet, D.G., Lonergan, M., Arthus, M.F., Jeck, N., Seyberth, H.J., Rosenthal, W., Oksche, A., van Oos, C.H., Deen, P.M.T. (2002) Misrouting of wild-type aquaporin-2 to late endosomes/lysosomes after heterotertramerization with an aquaporin-2 mutant explains dominant nephrogenic diabetes insipidus, *Hum. Mol. Genet.* **11**, 779–789.

Markowitz, G.S., Radhakrishnan, J., Kambham, N., Valeri, A.M., Hines, W.H., D'Agati, V.D. (2000) Lithium nephrotoxicity: a progressive combined glomerular and tubulointerstitial nephropathy, *J. Am. Soc. Nephrol.* **11**, 1439–1448.

Marples, D., Christensen, S., Christensen, E.I., Ottosen, P.D., Nielsen, S. (1995) Lithium-induced downregulation of aquaporin-2 water channel expression in rat kidney medulla, *J. Clin. Invest.* **95**, 1838–1845.

Marples, D., Frokiaer, J., Dorup, J., Knepper, M.A., Nielsen, S. (1996) Hypokalemia-induced downregulation of aquaporin-2 water channel expression in rat kidney medulla and cortex, *J. Clin. Invest.* **97**, 1960–1968.

Matsumura, Y., Uchida, S., Rai, T., Sasaki, S., Marumo, F. (1997) Transcriptional regulation of aquaporin-2 water channel gene by cAMP, *J. Am. Soc. Nephrol.* **8**, 861–867.

Morello, J.P., Salahpour, A., Laperriere, A., Bernier, V., Arthus, M.F., Lonergan, M., Petaja-Repo, U., Angers, S., Morin, D., Bichet, D.G., Bouvier, M. (2000) Pharmacological chaperones rescue cell-surface expression and function of misfolded V2 vasopressin receptor mutants, *J. Clin. Invest.* **105**, 887–895.

Mulders, S.M., Bichet, D.G., Rijss, J.P., Kamsteeg, E.J., Arthus, M.F., Lonergan, M., Fujiwara, M., Morgan, K., Leijendekker, R., van der Sluijs, P., van Os, C.H., Deen, P.M.T. (1998) An aquaporin-2 water channel mutant which causes autosomal dominant nephrogenic diabetes insipidus is retained in the Golgi complex, *J. Clin. Invest.* **102**, 57–66.

Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J.B., Engel, A., Fujiyoshi, Y. (2000) Structural determinants of water permeation through aquaporin-1, *Nature* **407**, 599–605.

Nielsen, S., Terris, J., Smith, C.P., Hediger, M.A., Ecelbarger, C.A., Knepper, M.A. (1996) Cellular and subcellular localization of the vasopressin-regulated urea transporter in rat kidney, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5495–5500.

Oksche, A., Schülein, R., Rutz, C., Liebenhoff, U., Dickson, J., Müller, H., Birnbaumer, M., Rosenthal, W. (1996) Vasopressin V2 receptor mutants that cause X-linked nephrogenic diabetes insipidus: analysis of expression, processing, and function, *Mol. Pharmacol.* **50**, 820–828.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M., Miyano, M. (2000) Crystal structure of rhodopsin: A G protein-coupled receptor, *Science* **289**, 739–745.

Rascher, W., Rosendahl, W., Henrichs, I.A., Maier, R., Seyberth, H.W. (1987) Congenital nephrogenic diabetes insipidus-vasopressin and prostaglandins in response to treatment with hydrochlorothiazide and indomethacin, *Pediatr. Nephrol.* **1**, 485–490.

Rittig, S., Robertson, G.L., Siggaard, C., Kovacs, L., Gregersen, N., Nyborg, J., Pedersen, E.B. (1996) Identification of 13 new mutations in the vasopressin-neurophysin II gene in 17 kindreds with familial autosomal dominant neurohypophyseal diabetes insipidus, *Am. J. Hum. Genet.* **58**, 107–117.

Rosenthal, W., Antaramian, A., Gilbert, S., Birnbaumer, M. (1993) Nephrogenic diabetes insipidus. A V2 vasopressin receptor unable to stimulate adenylyl cyclase, *J. Biol. Chem.* **268**, 13030–13033.

Rosenthal, W., Seibold, A., Antaramian, A., Lonergan, M., Arthus, M.F., Hendy, G.N., Birnbaumer, M., Bichet, D.G. (1992) Molecular identification of the gene responsible for congenital nephrogenic diabetes insipidus, *Nature* **359**, 233–235.

Saparov, S.M., Kozono, D., Rothe, U., Agre, P., Pohl, P. (2001) Water and ion permeation of aquaporin-1 in planar lipid bilayers: major differences in structural determinants and stoichiometry, *J. Biol. Chem.* **276**, 31515–31520.

Schmale, H., Bahnsen, U., Richter, D. (1993) Structure and expression of the vasopressin precursor gene in central diabetes insipidus, *Ann. N. Y. Acad. Sci.* **689**, 74–82.

Schülein, R., Zühlke, K., Krause, G., Rosenthal, W. (2001) Functional rescue of the nephrogenic diabetes insipidus-causing vasopressin V2 receptor mutants G185C and

R202C by a second site suppressor mutation, *J. Biol. Chem.* **276**, 8384–8392.
Tamarappoo, B.K., Verkman, A.S. (1998) Defective aquaporin-2 trafficking in nephrogenic diabetes insipidus and correction by chemical chaperones, *J. Clin. Invest.* **101**, 2257–2267.
Valtin, H., Sawyer, W.H., Sokol, H.W. (1965) Neurohypophysial principles in rats homozygous and heterozygous for hypothalamic diabetes insipidus (Brattleboro strain), *Endocrinology* **77**, 701–706.
Wade, J.B., McCusker, C., Coleman, R.A. (1986) Evaluation of granule exocytosis in toad urinary bladder, *Am. J. Physiol.* **251**, C380–C386.
Willcutts, M.D., Felner, E., White, P.C. (1999) Autosomal recessive familial neurohypophyseal diabetes insipidus with continued secretion of mutant weakly active vasopressin, *Hum. Mol. Genet.* **8**, 1303–1307.
Wood, S.P., Tickle, I.J., Treharne, A.M., Pitts, J.E., Mascarenhas, Y., Li, J.Y., Husain, J., Cooper, S., Blundell, T.L., Hruby, V.J., Buku, A., Fishman, A.J., Wyssbrod, H.R. (1986) Crystal structure analysis of deamino-oxytocin: conformational flexibility and receptor binding, *Science* **232**, 633–636.
Yang, B., Gillespie, A., Carlson, E.J., Epstein, C.J., Verkman, A.S. (2001) Neonatal mortality in an aquaporin-2 knock-in mouse model of recessive nephrogenic diabetes insipidus, *J. Biol. Chem.* **276**, 2775–2779.
Yun, J., Schöneberg, T., Liu, J., Schulz, A., Ecelbarger, C.A., Promeneur, D., Nielsen, S., Sheng, H., Grinberg, A., Deng, C., Wess, J. (2000) Generation and phenotype of mice harboring a nonsense mutation in the V2 vasopressin receptor gene, *J. Clin. Invest.* **106**, 1361–1371.

Diabetes Treatment: *see* Adipocytes

Dictyostelium Discoideum: Cellular Slime Mold

Annette Müller-Taubenberger[1] *and Markus Maniak*[2]
[1] *Max-Planck-Institute of Biochemistry, Martinsried, Germany*
[2] *University of Kassel, Kassel, Germany*

1 *Dictyostelium* as a Model Organism 327
1.1 *Dictyostelium* – a Social Amoeba 327
1.2 Phylogeny of *Dictyostelium* 329
1.3 Current Focus of *Dictyostelium* Research 330

2 The *Dictyostelium discoideum* Genome 330
2.1 General Organization 330
2.2 The Mitochondrial Genome 331
2.3 The Extrachromosomal rDNA Palindrome 332
2.4 Complex Repetitive Elements 332
2.5 Nuclear Plasmids 333
2.6 Small Noncoding RNAs 333
2.7 Sequencing Projects 333
2.7.1 The *Dictyostelium* Genome Project 333
2.7.2 The *Dictyostelium* cDNA Project 335
2.8 Transcriptome and Proteome Approaches in *Dictyostelium* 336

3 Tools for Molecular Manipulation 336
3.1 Transformation and Selection 337
3.2 Transformation Vectors 337
3.3 Pathways of Molecular Manipulation 338
3.3.1 Gain-of-Function Approaches 339
3.3.2 Loss-of-Function Mutations 340
3.3.3 Random Mutagenesis 341
3.4 Complex Mutants 342

4 Outlook 343

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

Acknowledgments 343

Bibliography 343
Books and Reviews 343
Primary Literature 344

Keywords

Amoebozoa
Eukaryotes can be assigned to one of eight major groups: excavates, discicristates, heterokonts, alveolates, cercozoa, plants, amoebozoa, or opisthokonts. Amoebozoa comprise lobose amoebas, dictyostelid slime molds, plasmodial slime molds, protostelid slime molds, and pelobionts.

Chemotaxis
Oriented movement toward a source of chemoattractant.

Development
Sequential changes in the gene expression program leading to cells with different properties.

Model Organism
Several organisms whose genomes are sequenced have been chosen by the National Institutes of Health (NIH) as model organisms for biomedical research to study the molecular mechanisms of cell motility, signal transduction, and cell-type differentiation and developmental processes.

Transformation
Introduction of foreign DNA into a cell in order to produce a protein or to interfere with gene expression.

Dictyostelium discoideum is a tractable system for the study of basic processes in cell and developmental biology. These processes include signal transduction, chemotaxis, motility, cytokinesis, phagocytosis, and aspects of development such as cell sorting, pattern formation, and cell-type differentiation. *Dictyostelium* has unique advantages for studying fundamental biological questions, since it is particularly amenable to both genetic and biochemical analyses. *Dictyostelium* has been chosen, along with several other organisms with known genomes, as a model system to analyze the functions of the identified genes. Biological questions in *Dictyostelium* have mainly

addressed the function of individual gene products, but more systematic approaches have now been launched on the basis of the knowledge derived from genome-, proteome-, and cDNA-projects. Recent studies on the genetics of *Dictyostelium* have opened new directions in biomedical research. This review summarizes the current knowledge on the *Dictyostelium* genome, updates the state of the *Dictyostelium* genome project, and concentrates on the molecular methods currently used to explore the functions of individual gene products.

1 *Dictyostelium* as a Model Organism

1.1 *Dictyostelium* – a Social Amoeba

The first reported species of the so-called cellular slime molds, *Dictyostelium mucorides*, was described by the mycologist Oskar Brefeld in 1869. Kenneth Raper subsequently introduced *D. discoideum* into the laboratory in 1935, and it has since become the most intensively studied species of the social amoebae. Other species of slime molds, such as *Polysphondylium pallidum*, have been poorly characterized. Cellular slime molds live as unicellular amoebae in decaying leaves of cool forest soil, where they feed on bacteria and other microorganisms by phagocytosis (Fig. 1a, b and Fig. 2d). *Dictyostelium discoideum* has been isolated in North America (Virginia, North Carolina, Massachusetts), in Europe, and in Japan.

As long as food is provided, *Dictyostelium* cells divide by mitotic cleavages (Fig. 2b, c). Starvation initiates a developmental program, which is accompanied by dramatic changes in gene expression. Starving cells stream toward a common center by chemotaxis, responding to and emitting cAMP signals (Fig. 1c and 2e). When about 100 000 amoebae have aggregated, they organize into a slug-shaped entity that can migrate toward light and can sense heat gradients (Fig. 1d). The slug then ceases migration and lifts off the substratum in a process called *culmination*. During this process, the cells differentiate into spore and stalk cells and form a fruiting body of 1 to 2 mm in height (Fig. 1e). Under appropriate nutritive conditions, new amoebae hatch from the spores (Fig. 2a) and the life cycle is started again. Alternative possibilities of resisting unfavorable environmental conditions are either the production of microcysts, a process about which very little is known, or macrocyst formation. During the latter process, cells of two mating types fuse under defined conditions, which include low ionic strength and starvation, and eventually a giant cell is formed. All three developmental options – formation of fruiting bodies, microcysts, and macrocysts involve the synthesis of cellulose walls that protect the cells against their environment.

The aspects of *Dictyostelium* development make it particularly amenable for research purposes. The developmental pathway is easily induced and reasonably fast, as it is completed within 24 h. Spores provide a convenient stage for long-term storage of cells, but the developmental cycle is not obligatory, since vegetative

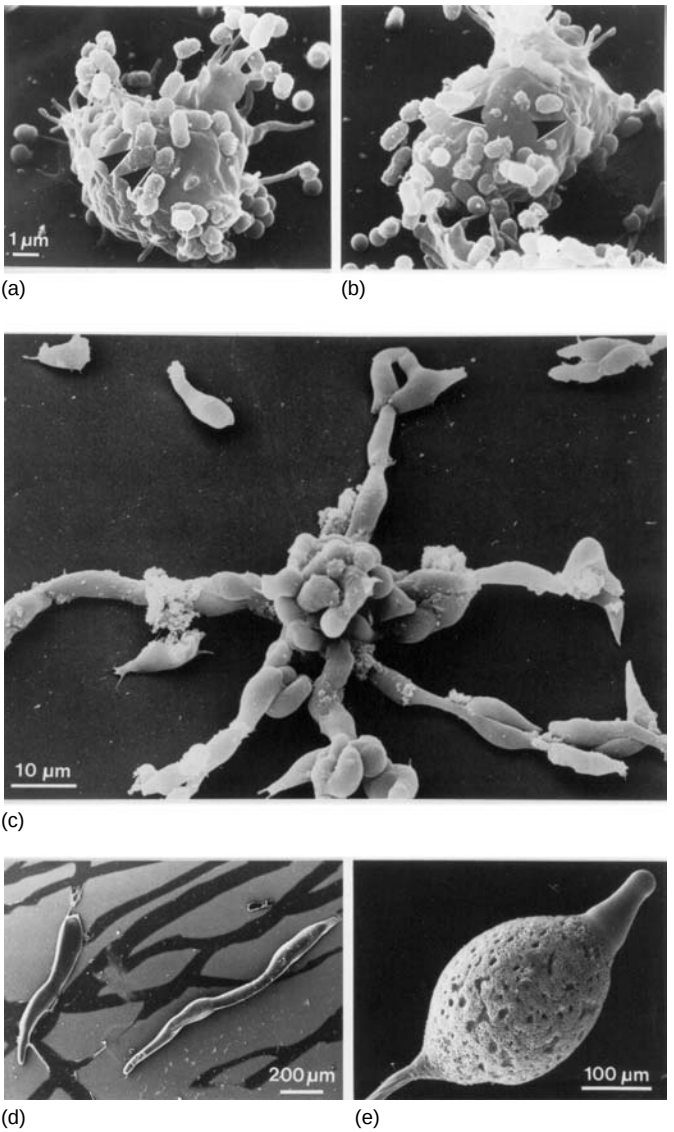

Fig. 1 Scanning electron micrographs illustrating different stages of the *Dictyostelium* life cycle: (a) and (b) amoebae feeding on bacteria by phagocytosis; (c) aggregating cells; (d) slugs moving on a solid substratum; (e) spore head of an almost completely differentiated fruiting body (Figure courtesy: M. Claviez).

cells can be frozen and maintained indefinitely in the laboratory. One of the great advantages of *Dictyostelium* is that it can be grown in the laboratory, not only on bacterial lawns but also on defined or semidefined liquid media, which enables mass production of genetically identical cells.

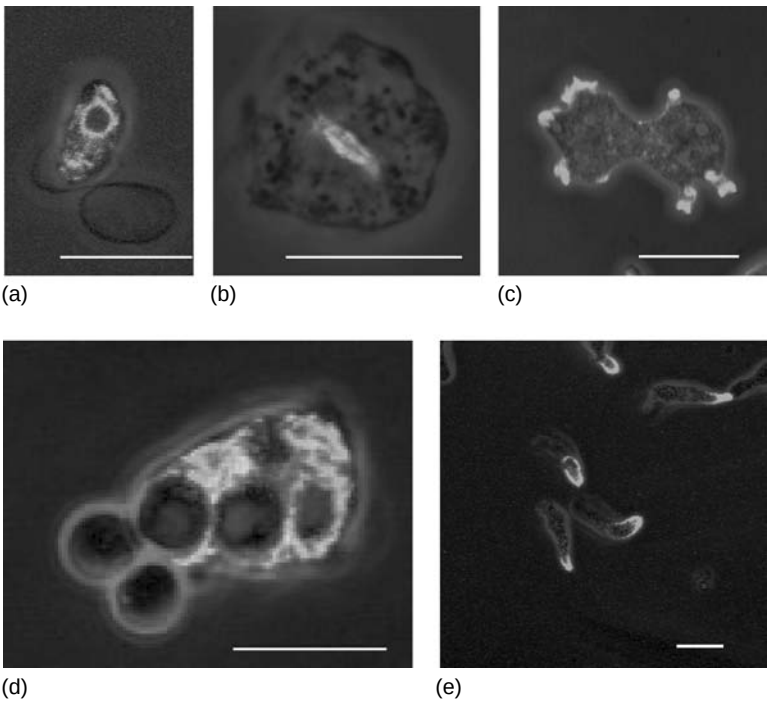

Fig. 2 Confocal micrographs of *Dictyostelium* cells illustrating (a) spore germination; (b) mitosis; (c) cytokinesis; (d) phagocytosis; (e) chemotaxis. GFP-tagged proteins were expressed in *Dictyostelium* to localize them within the cells during specific processes. (a) Calnexin-GFP was used as a specific marker for the endoplasmic reticulum (ER) and was imaged during germination of spores. The lower spore did not express the fusion protein. (b) Metaphase stage of mitosis; chromosomes were visualized by GFP-histone (green) and the spindle apparatus by antibodies directed against α-tubulin (red). (c) GFP-coronin, an actin-binding protein in the cell cortex, localizes to the distal cell poles during cytokinesis. (d) A *Dictyostelium* cell during the uptake of bacteria; GFP-calreticulin, a lumenal protein within the ER, was expressed to study the involvement of the ER in phagocytic cup formation. (e) *Dictyostelium* cells expressing a GFP-tagged kinase (DPAKa), which localizes to the rear end of the cells. Cells were moving in a gradient of cAMP toward the upper left corner. With the exception of (b) all images were taken using live cells. Bars, 10 μm (see color plate p. xxxiv).

1.2 Phylogeny of *Dictyostelium*

It is important to note that *Dictyostelium* amoebae can be considered as a prototype of a eukaryotic – a not yet specialized cell. Although historically called a *slime mold, Dictyostelium* bears little in common with other microorganisms like yeast or filamentous fungi, which have lost some or many proteins and pathways that are present in higher eukaryotes and also in *Dictyostelium*.

The phylogenetic position of *Dictyostelium* was, until recently, regarded as quite controversial. According to phylogenetic analyses based on combined protein data sets, *Dictyostelium* species are

now classified as Amoebozoa, positioned within the tree of life next to the Opisthokonts (Animalia and Fungi) and more distant from plants that diverged earlier. Accordingly, *Dictyostelium* genes show higher similarities to those of vertebrates than to those of plants or fungi.

1.3 Current Focus of *Dictyostelium* Research

The brief summary of the *Dictyostelium* life cycle already points to some main topics of research with this organism:

- chemotaxis
- signal transduction
- cytoskeletal research
- mitosis and cytokinesis
- phagocytosis
- mechanism of uptake of pathogenic bacteria like *Legionella*
- cell adhesion
- development, for example, cell differentiation and proportioning
- natural chimerism for investigation of issues relating to coexistence and conflict between cells
- homologs of human disease genes
- pharmacological studies

As a result of the continued work in these areas, *Dictyostelium* has been chosen by the National Institutes of Health (NIH, Bethesda) as a model organism for the study of the molecular mechanisms of signal transduction, cell motility, developmental processes and cell differentiation (http://www.nih.gov/science/models/).

DictyBase (http://www.dictybase.org/) is a web server that provides a comprehensive online resource for information on *Dictyostelium* research and is recommended for those who would like to get the most recent information on *Dictyostelium*. The web pages contain, for instance, genomics information, video images as well as protocols and methods. Furthermore, it provides a link to the most complete *Dictyostelium*-specific literature database available and a link to the currently built-up *Dictyostelium* stock center, which stores and distributes *Dictyostelium* strains and mutants as well as plasmids.

2 The *Dictyostelium discoideum* Genome

2.1 General Organization

The *Dictyostelium* genome is about 34 Mb in size and contains 6 chromosomes ranging from 4 to 7 Mb. Extrachromosomal elements comprise the mitochondrial DNA and the ribosomal DNA (rDNA) palindrome, which has been mistaken sometimes as a seventh chromosome. The chromosomes are believed to be acrocentric, with the centromere embedded in a cluster of long terminal repeat (LTR) retrotransposons (DIRS-1) composed of more than 40 elements.

The base composition of *Dictyostelium* is extremely biased toward a high adenine/thymine (A + T) content. On an average, the A + T content is 78%; in intergenic regions it may be as high as 98%. In coding regions, A and T residues are highly preferred in the third, or wobble position of the triplet. Coding sequences rather frequently contain repeats of the triplet AAC that encode depending on the reading frame polyasparagine, polyglutamine or polythreonine stretches. It is unclear if there is any functional role of these stretches in the respective proteins.

Dictyostelium genes are often simple in structure. Introns are infrequent and

their average length is about 130 bases, with larger introns of 1000 to 2000 bases occurring very rarely. Exon/intron junctions have the consensus splice sites AG/**GT**AAGT and AT**AG**/, with the bases printed in bold being absolutely conserved. The 3′ untranslated regions are on average not longer than approximately 150 bases. They contain a signal for 3′ processing and polyadenylation that corresponds to the mammalian AATAAA in most cases. Promoters are 1 to 2 kb in length.

The gene density with about 1 gene per 2.6 kb is quite high and similar to that of *Schizosaccharomyces pombe* (1 gene per 2.5 kb). For comparison, *Saccharomyces cerevisiae* has a gene density of 1 gene per 2.0 kb, whereas for human genes, the density is much lower and varies substantially across the genome between 6 and 23 per Mb.

Throughout most of its life cycle, *Dictyostelium* is a haploid organism, and therefore mutants can be easily generated (see Sect. 3). Nonetheless, one should keep in mind that a *Dictyostelium* cell is normally in the G2 phase of the cell cycle in which two sister chromatids already exist. Since the S-phase follows rather rapidly after mitosis, the G1 phase is extremely short.

2.2 The Mitochondrial Genome

Mitochondria of *Dictyostelium* carry a circular DNA molecule (mtDNA). The mtDNA contig is 55 564 bp in length and has been assembled from sequences contaminating the random shotgun reads of genomic DNA (http://dictygenome.bcm.tmc.edu) and by a primer walking strategy. It is significantly larger than mammalian mitochondrial genomes and hence more typical for protists (http://www.ncbi.nlm.nih.gov/genomes/static/euk_o.html). The A + T content of mtDNA, at 73%, is slightly lower than the one of genomic DNA. During vegetative growth, mitochondrial DNA can comprise up to 40% of the total cellular DNA.

The mitochondrial genome is transcribed into eight major, polycistronic RNA precursors encoding polypeptides, ribosomal RNAs (rRNAs), and interspersed transfer RNAs (tRNAs), which are processed into mono-, di-, or tri-cistronic RNAs. The maturation of polycistronic transcripts involves cleavage, using tRNAs as excision signals. The processing mechanisms are therefore similar to those reported for mammalian and fungal mitochondria and only rarely observed in higher plants. The complete sequence of the mtDNA revealed genes encoding 2 rRNAs, 18 tRNAs, 10 subunits of the NADPH dehydrogenase complex, apocytochrome b, 3 subunits of the cytochrome oxidase (COX), 4 subunits of the ATP synthase complex, 15 ribosomal proteins, and 5 other ORFs. In addition, a mitochondrial small (ms) RNA was identified in *Dictyostelium*. The msRNA is 129 nucleotides in length and is encoded by a mitochondrial gene. Although the predicted structure resembles that of 5S rRNA, the sequence shows only little homology with the mitochondrial 5S rRNA and the RNA is apparently not associated with ribosomes.

Sequence, gene content, and genome organization of the mitochondrial DNA of *Dictyostelium* displays similarities with the mitochondrial DNA of *Acanthamoeba castellanii*, based on sequence composition, gene content, and organization. In terms of phylogeny, it is important to note that in both organisms, the genes encoding COX1/2 are fused and the gene order within the ribosomal protein cluster is similar.

2.3 The Extrachromosomal rDNA Palindrome

In *Dictyostelium*, the genes that encode the RNA components of the ribosomes and the rRNAs are encoded by highly amplified multicopy families of identical genes that are carried on extrachromosomal palindromic elements that comprise up to 20% of the nuclear DNA. The sequence of the entire 88-kb rDNA element, which is present in about 100 copies per haploid genome, was determined recently. A chromosomal copy of the rDNA on chromosome 4 could provide the master copy for the stable transmission of the extrachromosomal elements. In nuclei of nocodazole-treated *Dictyostelium* cells that were arrested in mitosis, the extrachromosomal elements were found to form compact clusters. These clusters could ensure the efficient segregation of the rRNA to daughter cells during mitosis. They resembled true chromosomes in size and may thus explain previous cytological observations of a seventh chromosome in *Dictyostelium*.

2.4 Complex Repetitive Elements

Repetitive sequence elements are found in most genomes. They can be simple or complex, and if they are mobile, they are called *transposable elements*. Transposable elements carry direct or inverted repeat sequences at their ends and replicate within the genome. Transposable elements have no obvious function for the host cells, but probably contribute to genomic variability due to their transposition frequency within the genome.

The *Dictyostelium* genome hosts a series of complex repetitive elements that can be grouped into several classes. Together, these elements account for approximately 10% of the whole genome. Most complex repeats in *Dictyostelium* fit into commonly known classes of transposable elements. Some of them have been renamed recently and a summary on complex repeats in *Dictyostelium* has been published recently. The repeats are grouped into LTR retrotransposons, non-LTR retrotransposons, and DNA transposons.

LTR retrotransposons comprise DIRS-1 and *gypsy* class elements. DIRS-1 (also named Tdd1), had already been detected in the early days of exploration of the *Dictyostelium* genome. DIRS-1 is the most frequently found retrotransposon and makes up more than 3% of the genome. It encodes a retrovirus-like reverse transcriptase, and is flanked by inverted repeats. DIRS repeats are clustered in the vicinity of centromeres. DGLT-A and skipper are *gypsy* class transposable elements carrying direct repeats at their ends.

Non-LTR transposons are classified into TRE3 and TRE5 elements, so named as they insert in a position- and orientation-specific manner into the genome and integrate preferentially about 50 bp upstream (TRE5) or about 100 bp downstream (TRE3) of tRNA genes. On chromosome 2, for instance, TREs have spread and are found adjacent to about 75% of the tRNA loci present. TRE5 can be also used for gene-targeted retrotransposition.

DNA transposons are elements that are excised and inserted into new genomic locations without amplification. This new class of elements, including Tdd-4, Tdd-5, and DDT transposons, has only recently been identified in *Dictyostelium*. Thug elements have no coding capacity and carry terminal inverted repeats.

2.5 Nuclear Plasmids

Endogenous plasmids have been isolated from many wild-type strains of *D. discoideum*, for instance, NC4 or Ohio, or from related *Dictyostelium* species like *D. purpureum*. These plasmids range in size from about 1.3 to 30 kb and have copy numbers of 50 to 300 molecules per cell. The *Dictyostelium* plasmids are located within the nucleus and have a nucleosome structure that is typical for histone-associated DNA. The number of copies of each plasmid is dependent upon the individual plasmid, and remains stable over time. This copy number is independent of the genetic background of the strain and is constant relative to the degree of ploidy. Unlike the 2 μm circle of *Saccharomyces*, the different strains of *Dictyostelium* harbor different, nonhomologous plasmids.

In *D. discoideum*, two major plasmid families are distinguished – Ddp1 and Ddp2, and newly found plasmids can be grouped into either class. Previous studies have shown that Ddp1 carries an origin of replication within a 543-bp region, and its plasmid-encoded gene products are essential for long-term maintenance of the plasmid but not for replication. Ddp2 carries two elements that are essential for its extrachromosomal replication. It encodes the so-called Rep protein, a transactivator that interacts with a cis-acting DNA sequence, which is located in a 592-bp region of an inverted repeat. This element ensures extrachromosomal replication when the transacting factor is expressed. Studies showed that the Rep protein is not essential for autonomous replication, but it is likely to be involved in transcriptional regulation, control of DNA replication, and amplification of plasmids at low copy number.

2.6 Small Noncoding RNAs

Recently, several small noncoding RNAs (ncRNAs) were identified in *Dictyostelium* by utilizing an unbiased shotgun approach in combination with bioinformatics. Small RNAs of about 50 to 500 nucleotides length were isolated, followed by construction of cDNA libraries representing full-length small RNAs. Different classes of RNAs were identified, for example, putative small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNA), and a signal recognition particle RNA (SRP-RNA). In addition, an abundant novel class of RNAs (~50 to 65 nucleotides) was identified, which had previously not been identified in any other organism. These RNAs have conserved 5′ and 3′ ends that are complementary to each other and are predicted to base-pair and to form a stem-structure. Such structures are common motifs within different classes of functional ncRNAs, suggesting a biological role for these novel RNAs. Interestingly, these RNAs seem to be downregulated during development, suggesting a biological role for these novel RNAs.

2.7 Sequencing Projects

2.7.1 The *Dictyostelium* Genome Project

The sequencing and analysis of the genome of *Dictyostelium* (strain AX4) is an international collaboration between research groups in Germany, the United Kingdom, and the United States of America (see Table 1) (see also http://www.dictybase.org/). The project was started in 1998, and currently all

Tab. 1 Dictyostelium chromosomes and status of the genome project.

Chromosome	*Size [Mb]*	*Number of genes*	*Status*	*Contributing groups*
1	5.5	1784	Finished	1, 2
2	6.9	2799	Published	1, 2
3	5.3	2000	Shotgun complete/finishing	1, 2, 3
4	6.2	2300	Shotgun complete/finishing	3, 4
5	5.3	2000	Shotgun complete/finishing	3, 4, 5, 6
6	4.1	1500	Finishing	1–7

Notes: Sizes of the chromosomes are based on YAC mapping. The number of protein-coding genes per individual chromosome was either predicted from the completed sequence (chromosomes 1 and 2), or estimated on the basis of chromosomal sizes and gene density determined for chromosomes 1 and 2. Groups contributing to the *Dictyostelium* sequence consortium: (1) University of Cologne, Cologne, Germany (http://www.uni-koeln.de/dictyostelium/); (2) Genome Sequencing Center, Jena, Germany (http://genome.imb-jena.de/dictyostelium/); (3) Baylor College of Medicine, Houston, United States of America (http://dictygenome.bcm.tmc.edu/); (4) Sanger Institute, Hinxton, United Kingdom (http://www.sanger.ac.uk/Projects/D_discoideum/); (5) University of Dundee, Dundee, United Kingdom; (6) Medical Research Council, Cambridge, United Kingdom (http://www.mrc-lmb.cam.ac.uk:80/happy-homepage.html); (7) Pasteur Institute, Paris, France.

contributing groups have now pooled their data sets for a whole genome assembly, suggesting that the complete genome will be available in 2004. Goals of the genome project are to establish a comprehensive gene catalog and to provide functional analysis of these genes using, for instance, microarrays.

To sequence the *Dictyostelium* genome, a whole chromosome shotgun was employed following the separation of the chromosomes using pulse-field electrophoresis. DNA segments were subcloned either into yeast artificial chromosomes (YACs), or the DNA was sheared and cloned into pUC vectors. Shotgun sequences were assembled by a seed- and BLAST-based cyclic assembly strategy. To order the contigs along the chromosomes, HAPPY map, an *in vitro* PCR-based method that examines the cosegregation of sequence-tagged site markers in a panel of randomly broken genomic DNA, was employed. The program Phusion enabled computer-assisted readpair-based assembly. Previously mapped genes and a circular YAC library were also used to validate the assembly. The total coverage of the genome is now about 10-fold. The high A + T content, the presence of low-complexity intergenic regions, and the large number of complex repetitive elements complicated the sequencing and assembly of the *Dictyostelium* genome significantly.

Gene prediction and annotation of chromosomes 1 and 2 was initially performed using GlimmerM, a partially trained program, and continued with GeneID, trained with 160 previously known *Dictyostelium* genes. The latter version was fine-tuned for *Dictyostelium* genes taking into account features common to promoters, splice signals, and polyA signals. Gene ontology system (http://www.geneontology.org/) was used for a computer-assisted classification of genes homologous to genes of other organisms. About 70% of the genes match

known sequences and can be grouped into categories describing cellular processes and/or molecular functions.

For the *Dictyostelium* genome, a total number of about 12 000 genes is expected (see Table 1). This is approximately twice as much as in yeast, and close to the number of genes in *Drosophila* (13 600). The estimated gene number in *Dictyostelium* also provides a good estimate for the number of encoded proteins (the proteome), as only a few genes are known that use differential promoters, alternative splicing, or regulated 3′ processing to make alternative gene products.

An overview on the *Dictyostelium* chromosomes and the status of the genome project is given in Table 1. Chromosome 1 is 5.5 Mb in size encompassing 1784 protein-coding genes. Its sequence analysis is completed. Chromosome 2 is the largest of the *Dictyostelium* chromosomes comprising 6.9 Mb and representing about 20% of the genome. It was the first *Dictyostelium* chromosome for which the complete sequence was reported. Analysis revealed that it codes for 2799 predicted protein-coding genes and 73 transfer RNA genes. For chromosome 3, shotgun sequencing is finished and the assembly has started. Chromosomes 4 and 5 comigrate in the pulse-field electrophoresis and were therefore processed together. Chromosome 6 is the smallest one with 4.1 Mb and approximately 1500 genes are expected. In an analysis of the structural features of this chromosome, DIRS repeat clusters of about 50 to 180 kb were detected at each chromosome's end, supporting the assumption that *Dictyostelium* chromosomes are telocentric.

2.7.2 The *Dictyostelium* cDNA Project

Several Japanese universities are involved in a collaborative project to identify and analyze expressed genes in *Dictyostelium* (http://www.csm.biol.tsukuba.ac.jp/cDNAproject.html). The goal of the project is to identify and characterize genes of *Dictyostelium* in distinct stages of growth and development to provide essential information regarding the structure, function, timing, and composition of expressed genes.

Initially, the project started with EST (expressed sequence tag) libraries constructed from size-fractionated growth- and slug-stage specific cDNAs. Sequences were analyzed by BLAST and PHRAP programs to obtain a set of nonredundant ESTs. The EST libraries were of great value for discovering genes and for analyzing their expression profile. To complement the gene identification by large-scale EST analysis, full-length cDNA libraries were constructed by an oligo-capping method starting with RNA from vegetative, aggregation, migration, and culmination stages. About 90% of the cDNA clones obtained in each set were full length. The sequences of more than 70 000 cDNA clones were obtained and analyzed. From the data sets, 8361 nonredundant sequences and 6718 assigned genes were determined (contig version 5.0). Partial cDNA clones from vegetative (VS-clones), developmental (SL- and SS-clones), and sexual (FC-, FCL-, and FC-IC-clones) stages as well as full-length clones are available and can be requested for conducting further work.

Furthermore, the cDNA project is directly linked to ATLAS, a spatial expression database containing results from *in situ* hybridization studies. Clones of the cDNA project have been used in microarray analyses in order to characterize both developmentally regulated and cell-type specific genes. In addition, whole-mount *in situ* hybridization was employed to study changes in the expression pattern of prestalk cells

based on the abundance of established cell-type specific genes from microarray results.

2.8 Transcriptome and Proteome Approaches in *Dictyostelium*

The developmental cell cycle of *Dictyostelium* is accompanied by a series of coordinated morphological and physiological changes. There is a general interest as to whether these transitions can be correlated with changes in gene expression and also protein levels. One way to examine if coregulated genes share functional characteristics, or if the expression pattern reflects the cell's physiology, is to apply transcriptional profiling using microarrays.

This approach was recently applied to *Dictyostelium*, using a collection of 7385 cDNA clones that represent about two-thirds of the expressed protein encoding genes. This analysis revealed that major developmental events coincide with prominent transcriptional changes. A significant change in the expression of more than 25% (2000 to 3000) of the genes takes place during the aggregation stage. This compares with changes in the preaggregation period and the finger stage of morphogenesis in which only little changes were detected. The transcriptional profile itself seems to be robust, as it is independent of the previous nutritional state and the strain investigated. Microarrays have also been used successfully to identify new cell-type specific genes preferentially expressed either in prestalk or prespore cells that can serve as useful markers for the characterization of cell types, or to investigate gene regulation during early development.

Proteomics involves the systematic separation, identification, and characterization of proteins present in a cell, tissue, or other biological sample. The collection of proteins may be correlated with a certain physiological state or might be helpful to determine diseases at a molecular level. In studies of single-cell organisms like *Dictyostelium*, however, proteomic approaches can also be used to analyze the proteins associated with entire organelles like phagosomes, or prespore secretory vesicles to enhance the understanding of the endocytic pathway or the mechanism of regulated secretion. SWISS-2DPAGE (http://www.expasy.org/ch2d/publi/dicty.html) offers a published 2D-PAGE map of *Dictyostelium*.

3 Tools for Molecular Manipulation

As a model system, *Dictyostelium* is used to investigate the functions that underlie biological processes. In the past, this had been done by mutagenizing cells with radiation or chemicals like nitrosoguanidine, and selecting mutants of interest for further investigation. Some of these mutants were subjected to additional rounds of mutations to build up elaborate regulatory networks, but details of the molecules underlying these processes often remained unknown. Many research groups have tried to complement this type of mutant with expression libraries, but this approach has frequently failed, although there is one report on a recent success. At present, most molecular studies investigate the function of individual proteins, but more systematic approaches have been launched on the basis of knowledge derived from genome-, proteome- and cDNA-projects.

3.1 Transformation and Selection

The career of *Dictyostelium* as a molecular model system started almost 20 years ago, when Nellen and coworkers developed transformation vectors and devised a protocol for transformation and selection of axenic cells. In principle, this method is similar to a technique used for transformation of mammalian cells, in which cells adhering to a culture dish are overlaid with a calcium precipitate containing many copies of the transforming DNA. Cells are then osmotically stressed for a few minutes by the addition of glycerol, which results in the internalization of the DNA, and its integration into the nuclear genome. This method has been continuously improved and is still used in many labs around the world. Subsequently, two alternative methods have been established. Electroporation allows the introduction of single or a few vector copies into the cell. This technique is important if vectors containing libraries are to be transformed. More recently, a "gene gun" has been applied for transient transfection experiments. Here, metal particles are coated with DNA and "shot" into the cells. In principle, this method bears the potential to target the mitochondrial genome as well as the nuclear DNA.

Transformation efficiencies of these methods are, however, low, giving in the order of 10^{-5} cells that carry the desired foreign DNA per experiment. These rare transformants can be enriched by metabolic selection for mutants in the biosynthetic pathway of nucleotides. Alternatively, transformants are selected for resistance to antibiotics like hygromycin, bleomycin, geneticin (G418), and blasticidin, of which the last two are most commonly used. Different antibiotics require different amounts of enzyme for their inactivation. Therefore, the transforming DNA is found to be amplified to varying degrees in the resistant cells.

3.2 Transformation Vectors

We now need to consider the different characteristics of the vectors used for transformation. If the plasmid bears a replication origin specific for *Dictyostelium* and encodes accessory factors (see Sect. 2.5), it remains extrachromosomal and is propagated autonomously. The majority of circular plasmids, however, lack these features. Accordingly, they are integrated into the genome and segregate with the chromosomes. A detailed analysis by Barth et al. has revealed that when a circular plasmid enters the nucleus, it becomes replicated by a rolling circle mechanism and the multimeric product is integrated into a single site into the genome. If more than one plasmid enters a cell, which is a known and sometimes desired consequence of calcium precipitation, they may end up in different chromosomal locations. The final copy number obtained after amplification varies between 1 and 1000, but the selection regimes using different antibiotics result in the survival of low copy numbers ($\ll 10$) for blasticidin and high copy numbers (50–200) for geneticin. Integration is also inevitable if linear DNA fragments are used for transformation instead of circular plasmids.

There are two applications for the molecular manipulation of cells: one is to obtain cells with additional features (gain-of-function mutant). The other reason is to construct strains lacking a certain cellular property (loss-of-function mutant). There are many ways to achieve these goals,

and most of them require plasmids that allow a certain degree of control over the expression levels in time and space. Some plasmids use naturally regulated promoters from the *discoidin* I *gamma* or *actin* 6 genes, which are much more active in axenically growing cells than in cells feeding on bacteria, but a promoter active in bacterially growing cells is also available. Promoters that provide easier control over their expression level are, for example, repressible by the natural chemoattractant folate or inducible by the bacterial antibiotic tetracycline. For mRNA expression, specifically in the aggregation stage of the *Dictyostelium* life cycle, a vector is available that uses the promoter of the cell adhesion protein contact sites A. This promoter can be superinduced by extracellular pulses of the chemoattractant cAMP. In later phases of development, cell-type specific promoters can be used like those expressed in prespore cells and different types of prestalk cells. However, for many applications, a nearly constitutive promoter like the one from the *actin* 15 gene may suffice.

3.3 Pathways of Molecular Manipulation

The possibilities of genetic manipulation that have been exploited in *Dictyostelium* cells to date are summarized in Fig. 3.

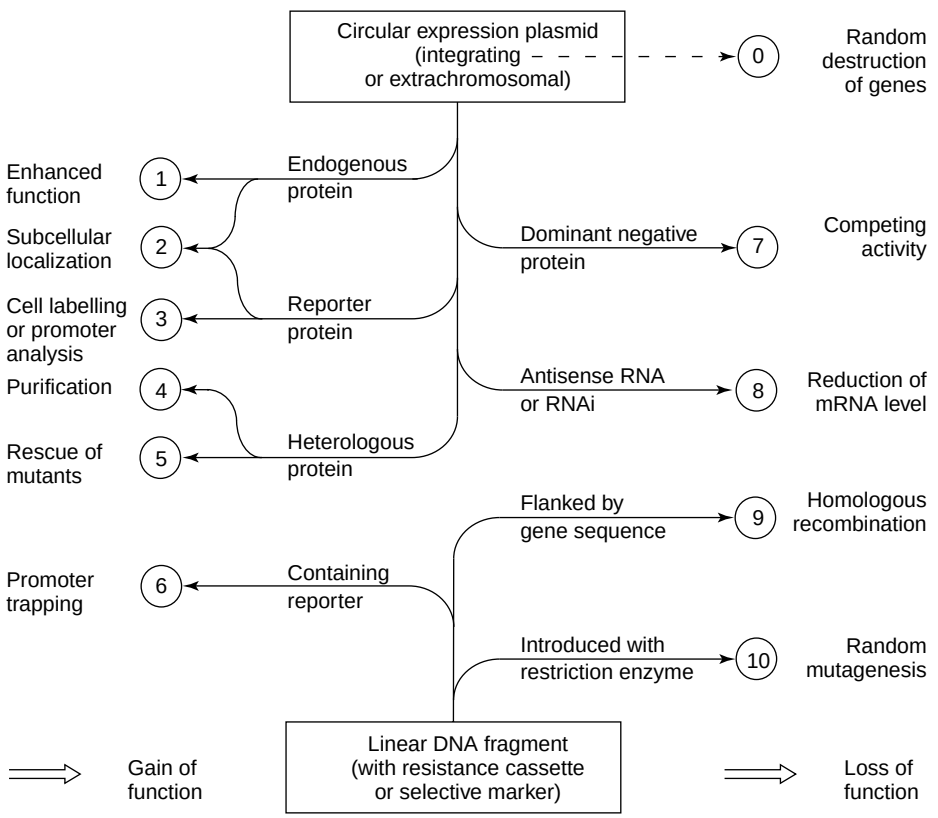

Fig. 3 Pathways of molecular manipulation. See text for detailed description.

Below, we discuss the individual options step-by-step, using applications illustrated with one or a few citations rather than providing a complete listing of the available literature.

3.3.1 Gain-of-Function Approaches

The aim of transforming *Dictyostelium* cells with expression vectors is to express either homologous native or mutated genes, often containing an introduced marker protein. *Dictyostelium* can also be used to express genes derived solely from heterologous organisms. Because many transformation vectors integrate randomly into the genome and may accidentally destroy genes (Fig. 3, path 0), it is advisable to analyze a number of clones obtained from independent transformations for the absence of growth defects and developmental abnormalities.

Overexpression of endogenous proteins In many cases, endogenous proteins were overexpressed just to increase the efficiency of their purification. Looking at the phenotypes of overexpressing cells, one is also tempted to interpret observations as possible clues for protein function (Fig. 3, path 1). This may indeed be feasible in an ideal situation in which the phenotype of overexpressing cells is the reverse of cells lacking the protein. Unfortunately, this correlation only holds true if the effect that is studied is directly caused by the protein under investigation. In contrast, there are also reports revealing that overexpression of a protein and its loss produce similar phenotypes. Results of this sort suggest that for many biological processes there is an ideal concentration of the components involved and tipping the balance to either side perturbs the molecular pathway.

Expression of reporter proteins Many different reporter proteins have been expressed in *Dictyostelium* (Fig. 3, path 3). These experiments served to provide quantitative measures for promoter activity and were used to dissect functional promoter elements. Here, the firefly luciferase, bacterial beta-galactosidase, and chloramphenicol acetyltransferase were most frequently used for quantification. On the other hand, reporter proteins were used to label entire cells for studies on their behavior in the multicellular stage of development. Beta-galactosidase was the protein of choice for a long time, because it can be easily detected and its lifetime can be manipulated. More recently, the green fluorescent protein (GFP) and its variants, which can be visualized without fixing and staining cells, have taken over, thus allowing the fates of living cells to be tracked. The GFP reporter protein has proven to be especially useful for producing chimeras with endogenous proteins to study their subcellular localization (Fig. 2 and Fig. 3, path 2). Using this approach, virtually all known compartments and organelles have been tagged to allow their fluorescence-mediated detection in living cells.

Expression of (other) heterologous proteins The relative ease of expressing genes from other organisms conferring resistance to antibiotics or acting as reporters has stimulated interest to use *Dictyostelium* cells as a eukaryotic expression system (Fig. 3, path 4). It was found, however, that it was necessary to change the first few codons of heterologous genes into *Dictyostelium* codon usage, especially if G or C residues are frequent in the wobble position. Then, antigens from parasites or viruses, or proteins of medical importance can be successfully expressed. The

Dictyostelium system may also become important to express functional membrane proteins and glycoproteins. The requirements for glycosylation have been recently established and various glycosylation mutants are available. Most recently, one important prerequisite for large-scale production has been achieved when Flaschel and coworkers succeeded in growing cells in a conventional bioreactor to high densities in an improved medium. In addition to these more technological aspects, expression of heterologous genes has been used to rescue *Dictyostelium* cytoskeletal mutants with proteins from other organisms like plants or animals (Fig. 3, path 5).

3.3.2 Loss-of-Function Mutations

In order to study the role of a protein *in vivo*, it is often illuminating to investigate cells that lack its function. This goal can be achieved using one of three targeted approaches.

Dominant negative effects In rare cases, there is detailed information available on the protein's crystal structure, or previous work in other organisms has identified functionally important amino acids. Then one can reconstruct single point mutations creating dominant negative mutants (Fig. 3, path 7), which are generally used for the functional analysis of small GTPases involved in regulating the cytoskeleton, vesicle trafficking steps, or other events in signal transduction. Similarly, one can guess whether a domain of the protein like the myosin tail will interfere with the endogenous protein forming inactive complexes or a myosin power-relay mutant will block the actin target by high-affinity interaction.

Antisense RNA and RNA interference The expression of antisense RNA has been used to interfere with gene expression posttranscriptionally, thus creating a type of gene knockout strain. In this method, a fragment of the gene of interest is inserted in opposite orientation behind a promoter of an expression plasmid so that it is transcribed into the noncoding strand. It was assumed that this RNA forms a base-paired hybrid molecule that becomes degraded within the cell, since neither endogenous mRNA nor antisense RNA could be detected in successful gene ablation attempts (Fig. 3, path 8). More recently, the novel approach of RNA interference (RNAi) was established for *Dictyostelium* (Fig. 3, path 8). Here, a self-complementary, hairpin-structured RNA molecule is transcribed, which acts both as a signal and cofactor in the pathway of enzyme-mediated destruction of the corresponding endogenous mRNA. Antisense RNA and RNAi techniques have two major advantages. First, they are suitable for the elimination of all of the mRNAs and protein products of a gene family, and secondly they provide the chance to isolate hypomorphous mutants in essential genes, which express just the minimum amount of the gene product required for cell survival.

Gene disruption by homologous recombination Undisputedly, the most precise tool to study a protein's function *in vivo* is the production of a null mutant. This has become feasible when De Lozanne and Spudich attempted to overexpress a truncated myosin molecule and the resulting phenotype indicated a total loss of function. Closer analysis of the mutant revealed that the vector had integrated by homologous recombination and disrupted the myosin gene. Since then, many genes have

been ablated using resistance cassettes or metabolically selectable markers that replace an essential part of the gene (Fig. 3, path 9). Technically, the assembly of gene replacement constructs has become easier by two novel techniques. The first one is based on a PCR approach that works without conventional cloning steps such as ligation and transformation into *Escherichia coli* hosts. The second method for constructing replacement fragments uses the reaction of Tn5 *in vitro* transposition into a genomic fragment carried in a plasmid, allowing the subsequent isolation of disruption vectors after transformation of *E. coli*.

Because gene replacement requires two independent crossover events within the two flanking homologous sequences, it is a relatively inefficient event. To enhance the success rate, a linear DNA fragment rather than a circular plasmid is used. To prevent a possible circularization *in vivo*, the ends of the transforming DNA can be modified by dephosphorylation or the addition of dideoxynucleotides. In addition, a negative selection regime can be used to reduce random insertion: a gene encoding a tRNA molecule suppressing the most frequent stop codon in *Dictyostelium* (UAA) is appended outside the homologous regions of the gene. After transformation and integration, only those cells that have lost this lethal factor by a crossover event will survive.

Gene replacement strategies do not inevitably lead to all-or-none results. A couple of approaches allow for the construction of conditional mutants or cells producing residual protein activity, which is especially useful for the investigation of essential gene products. Parts of the *NSF* gene, which plays a role in the essential process of membrane fusion, have been randomly mutated, and the endogenous copy of the gene has been replaced with this library to select a temperature-sensitive mutation. Using an even more ambitious approach, the chromosomal copy of the DNA-binding factor CMBF has been exchanged with a mutated version containing an in-frame UAG stop codon within the coding region. This UAG codon is rarely used by native *Dictyostelium* genes, therefore, the cells apparently tolerate the expression of a cognate suppressor tRNA; but because suppression is inefficient, the cells contained strongly reduced amounts of CMBF, so that its function could be finally investigated.

3.3.3 Random Mutagenesis

Because random approaches produce large numbers of mutants, it is especially important to develop efficient selection or screening regimes. The easiest way is, of course, to select for mutants that survive a drug treatment or are able to develop in the presence of a drug. Elegant strategies have been used to enrich mutants impaired in phagocytosis or cell adhesion. If changes in the expression of a single gene are the output of a mutagenesis experiment, colony blot techniques can be used to detect the expression of a protein by antibodies or by hybridization to endogenous mRNA. If none of these approaches can be used, mere visual inspection of clones growing on bacterial lawns may identify mutants that are defective in the completion of development.

Promoter traps While the molecular approaches discussed so far all rely on some knowledge of the gene product to be manipulated, there are some methods that are suited to target even unknown genes in a genome-wide manner. One of these methods is the random knockin strategy of a promoterless reporter gene to identify stage-specific and developmentally

regulated transcription units by virtue of the gain in reporter activity (Fig. 3, path 6). Of course, by applying the appropriate selection regimes, this method can also be used for the isolation of loss-of-function mutations.

Restriction enzyme-mediated integration For functional analysis of a biological process, it is essential to saturate the genome with mutations in addition to tagging genes for subsequent identification. In this respect, the method of restriction enzyme-mediated integration (REMI) has proven to be especially useful. For REMI, a linearized vector is transformed into the cells, together with a restriction enzyme producing compatible ends. The enzyme cleaves chromosomal DNA at random positions and the linearized vector integrates into the restriction site during the repair process. This method was initially applied to isolate developmental mutants. Subsequently, the protocol was optimized through integration of vectors bearing the blasticidin resistance cassette into the more frequent DpnII sites of the *Dictyostelium* genome. To recover the mutated gene, a preparation of genomic DNA is digested with a restriction enzyme that cuts outside the vector and is then ligated and transformed into *E. coli* cells. Thus, sequences flanking the insertion site are recovered and the resulting plasmid can be immediately used to reconstruct the mutation in a wild-type background to confirm the original phenotype.

Shotgun antisense A conceptually related approach to producing random mutations is the transformation of cells with an expression library of *Dictyostelium* cDNAs inserted in reverse orientation with respect to the promoter, so that antisense RNA molecules are produced (see also Fig. 3, path 8). The antisense strategy is associated with a couple of problems, because it favors the isolation of genes that are overrepresented in the antisense library. However, this shotgun approach has been used to successfully identify genes involved in the completion of the developmental cycle of *Dictyostelium*.

3.4 Complex Mutants

Basically, the strategies discussed above lead to cells with a single genetic alteration. When a mutant strain constructed by homologous recombination is complemented with a plasmid that restores expression of the missing protein, for example, from a different promoter, a second level of complexity is reached. However, to obtain such cells, two different selectable markers have already been employed. Therefore, the construction of even more complex mutants is a challenging task. In some cases, knowledge about the chromosomal organization of the genes in question may help. It was found that all four members of the GP138 gene family were arranged in tandem, so a total tetra knockout could be created in one go. In most cases, however, the genes to be analyzed are spread throughout the genome. If they turn out to reside on different chromosomes, one can use parasexual genetics to produce crosses between mutants as used previously for the combination of chemically induced mutations. More recently Insall and King have devised protocols for producing diploids in axenic cultures and subsequent haploidization. They showed for the first time the successful combination of two gene disruption mutants obtained by molecular methods.

In order to construct multiple knockouts using the same resistance mechanism for each single step, the cre/lox system is currently exploited in a couple of laboratories. In this approach, the resistance cassette is directly flanked by two bacteriophage-derived *loxP* sequences that are recognized by the cre recombinase. After successful gene disruption, the resistance cassette can be excised by the cre enzyme to provide a strain that is again sensitive toward the selective antibiotic, which can be used in another round of targeted homologous recombination. Finally, one should not forget that when the goal is not disrupting multiple genes but rather expressing two different proteins within a single cell, cotransformation may be the method of choice. In this approach, two different plasmids, which can even bear the same selectable marker, are transformed simultaneously. Irrespective of whether the precipitation and osmotic shock method or electroporation is used, many cells will receive both plasmids that integrate independently into different chromosomal locations. They may accordingly be amplified to different degrees. In the end, one can choose among a number of clones and identify the ones that produce the desired expression levels of both proteins.

4 Outlook

For roughly a century, *Dictyostelium* has served as an excellent model system for biological research. During this time, the fundamental goals of research using this organism have changed. Initial studies in areas of developmental biology such as the determination of cell fate or the maintenance of body proportion has become more focused on the underlying topics such as cell adhesion processes and signal transduction. The last decade has seen the opening up of both the entire genome and the regulation of genes within the genome. Valuable information gained over this period is now being employed for functional studies of this simple eukaryote that is equally amenable to both genetic manipulation and biochemical analysis. It will contribute significantly to basic research in areas such as biomedicine. Results obtained with *Dictyostelium* will continue to influence the research in other fields and vice versa. This will ensure the continued use of *Dictyostelium* to solve the questions posed by modern biology.

Acknowledgments

We would like to thank Fredrik Söderbom (Swedish University of Agricultural Sciences, Uppsala, Sweden) for communicating results on small RNAs in *Dictyostelium*, Ludwig Eichinger (University of Cologne, Germany) for providing information on the genome project prior to publication, and Wolfgang Nellen (University of Kassel, Germany) for comments on the manuscript.

See also Genetics, Molecular Basis of; Genomic Sequencing (Core Article).

Bibliography

Books and Reviews

Eichinger, L. (2003) Revamp a model – status and prospects of the *Dictyostelium* genome project, *Curr. Genet.* **44**, 59–72.

Kessin, R.H. (2001) *Dictyostelium*, Cambridge University Press, Cambridge.

Loomis, W.F. (Ed.) (1982) *The Development of Dictyostelium*, Academic Press, New York.

Maeda, Y., Inouye, K., Takeuchi, I. (Eds.) (1997) *Dictyostelium – A Model System for Cell and Developmental Biology*, Universal Academy Press, Tokyo.

Primary Literature

Abe, T., Langenick, J., Williams, J. (2003) Rapid generation of gene disruption constructs by in vitro transposition and identification of a *Dictyostelium* protein kinase that regulates its rate of growth and development, *Nucl. Acids Res.* **31**, e 107.

Adachi, H., Hasebe, T., Yoshinaga, K., Ohta, T., Sutoh, K. (1994) Isolation of *Dictyostelium discoideum* cytokinesis mutants by restriction enzyme-mediated integration of the blasticidin-S resistance marker, *Biochem. Biophys. Res. Commun.* **205**, 1808–1814.

Adessi, C., Chapel, A., Vincon, M., Rabilloud, T., Klein, G., Satre, M., Garin, J. (1995) Identification of major proteins associated with *Dictyostelium discoideum* endocytic vesicles, *J. Cell Sci.* **108**, 3331–3337.

Ashktorab, H., Welker, D.L. (1988) Establishment of the nuclear location in *Dictyostelium discoideum* plasmids, *Gene* **65**, 41–49.

Aubry, L., Maeda, M., Insall, R., Devreotes, P.N., Firtel, R.A. (1997) The *Dictyostelium* mitogen-activated protein-kinase ERK2 is regulated by ras and cAMP-dependent protein-kinase (PKA) and mediates PKA function, *J. Biol. Chem.* **272**, 3883–3886.

Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., Doolittle, W.F. (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data, *Science* **290**, 972–977.

Barth, C., Fraser, D.J., Fisher, P.R. (1998) Co-insertional replication is responsible for tandem multimer formation during plasmid integration into the *Dictyostelium* genome, *Plasmid* **39**, 141–153.

Barth, C., Greferath, U., Kotsifas, M., Tanaka, Y., Alexander, S., Alexander, H., Fisher, P.R. (2001) Transcript mapping and processing of mitochondrial RNA in *Dictyostelium discoideum*, *Curr. Genet.* **39**, 355–364.

Blaauw, M., Linskens, M.H., van Haastert, P.J. (2000) Efficient control of gene expression by a tetracycline-dependent transactivator in single *Dictyostelium discoideum* cells, *Gene* **252**, 71–82.

Blusch, J., Morandini, P., Nellen, W. (1992) Transcriptional regulation by folate-inducible gene expression in *Dictyostelium* transformants during growth and early development, *Nucleic Acids Res.* **20**, 6235–6238.

Buczynski, G., Bush, J., Zhang, L.Y., Rodriguez-Paris, J., Cardelli, J. (1997) Evidence for a recycling role for Rab7 in regulating a late step in endocytosis and in retention of lysosomal enzymes in *Dictyostelium discoideum*, *Mol. Biol. Cell* **8**, 1343–1360.

Burns, C.G., Reedy, M., Heuser, J., De Lozanne, A. (1995) Expression of light-meromyosin in *Dictyostelium* blocks normal myosin-II function, *J. Cell Biol.* **130**, 605–612.

Bush, J., Temesvari, L., Rodriguez-Paris, J., Buczynski, G., Cardelli, J. (1996) A role for a Rab4-like GTPase in endocytosis and in regulation of contractile vacuole structure and function in *Dictyostelium discoideum*, *Mol. Biol. Cell* **7**, 1623–1638.

Cappello, J., Handelsman, K., Lodish, H.F. (1985) Sequence of *Dictyostelium* DIRS-1: An apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence, *Cell* **43**, 105–115.

Chang, W.T., Gross, J.D., Newell, P.C. (1995) Trapping developmental promoters in *Dictyostelium*, *Plasmid* **34**, 175–183.

Chang, A.C.M., Hall, R.M., Williams, K.L. (1991) Bleomycin resistance as a selectable marker for transformation of the eukaryote, *Dictyostelium discoideum*, *Gene* **107**, 165–170.

Chen, P., Chaudoir, B.M., Trybus, K.M., Chisholm, R.L. (1999) Expression of chicken gizzard RLC complements the cytokinesis and developmental defects of *Dictyostelium* RLC null cells, *J. Muscle Res. Cell Motil.* **20**, 177–186.

Clarke, M., Kayman, S.C., Riley, K. (1987) Density-dependent induction of discoidin-I synthesis in exponentially growing cells of *Dictyostelium discoideum*, *Differentiation* **34**, 79–87.

Cornillon, S., Pech, E., Benghezal, M., Ravanel, K., Gaynor, E., Letourneur, F., Brückert, F., Cosson, P. (2000) Phg1p: A nine-transmembrane protein superfamily member involved in *Dictyostelium* adhesion and phagocytosis, *J. Biol. Chem.* **275**, 34287–34292.

Cox, E.C., Vocke, C.D., Walter, S., Gregg, K.Y., Bain, E.S. (1990) Electrophoretic karyotype for

Dictyostelium discoideum, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8247–8251.

De Lozanne, A., Spudich, J.A. (1987) Disruption of the *Dictyostelium* myosin heavy chain gene by homologous recombination, *Science* **236**, 1086–1091.

Detterbeck, S., Morandini, P., Wetterauer, B., Bachmair, A., Fischer, K., MacWilliams, H.K. (1994) The prespore-like cells of *Dictyostelium* have ceased to express a prespore gene-analysis using short-lived beta-galactosidases as reporters, *Development* **120**, 2847–2855.

Dingermann, T., Reindl, N., Werner, H., Hildebrandt, M., Nellen, W., Harwood, A., Williams, J., Nerke, K. (1989) Optimization and in situ detection of *Escherichia coli* ß-galactosidase gene expression in *Dictyostelium discoideum*, *Gene* **85**, 353–362.

Dingermann, T., Troidl, E.M., Broker, M., Nerke, K. (1991) Expression of human antithrombin-III in the cellular slime mould *Dictyostelium discoideum*, *Appl. Microbiol. Biotechnol.* **35**, 496–503.

Dittrich, W., Williams, K.L., Slade, M.B. (1994) Production and secretion of recombinant proteins in *Dictyostelium discoideum*, *Biotechnology* **12**, 614–618.

Dynes, J.L., Firtel, R.A. (1989) Molecular complementation of a genetic marker in *Dictyostelium* using a genomic DNA library, *Proc. Natl. Acad. Sci U.S.A.* **86**, 7966–7970.

Early, A.E., Williams, J.G. (1987) Two vectors which facilitate gene manipulation and a simplified transformation procedure for *Dictyostelium discoideum*, *Gene* **59**, 99–106.

Egelhoff, T.T., Brown, S.S., Manstein, D.J., Spudich, J.A. (1989) Hygromycin resistance as a selectable marker in *Dictyostelium discoideum*, *Mol. Cell. Biol.* **9**, 1965–1968.

Eichinger, L., Noegel, A.A. (2003) Crawling into a new era-the *Dictyostelium* genome project, *EMBO J.* **22**, 1941–1946.

Emslie, K.R., Miller, J.M., Slade, M.B., Dormitzer, P.R., Greenberg, H.B., Williams, K.L. (1995) Expression of the rotavirus sa11 protein vp7 in the simple eukaryote *Dictyostelium discoideum*, *J. Virol.* **69**, 1747–1754.

Faix, J., Dittrich, W., Prassler, J., Westphal, M., Gerisch, G. (1995) pDcsa vectors for strictly regulated protein-synthesis during early development of *Dictyostelium discoideum*, *Plasmid* **34**, 148–151.

Fasel, N., Begdadirais, C., Bernard, M., Bron, C., Corradin, G., Reymond, C.D. (1992) *Dictyostelium discoideum* as an expression host for the circumsporozoite protein of *Plasmodium falciparum*, *Gene* **111**, 157–163.

Fey, P., Cox, E.C. (1997) Gene trapping with GFP – the isolation of developmental mutants in the slime-mold *Polysphondylium*, *Curr. Biol.* **7**, 909–912.

Fey, P., Stephens, S., Titus, M.A., Chisholm, R.L. (2002) SadA, a novel adhesion receptor in *Dictyostelium*, *J. Cell Biol.* **159**, 1109–1119.

Freeze, H.H., Yeh, R., Miller, A.L., Kornfeld, S. (1983) The mod A mutant of *Dictyostelium discoideum* is missing the α1,3-glucosidase involved in asparagine-linked oligosaccharide processing, *J. Biol. Chem.* **258**, 14880–14884.

Gerisch, G., Hagmann, J., Hirth, P., Rossier, C., Weinhart, U., Westphal, M. (1985) Early *Dictyostelium* development: control mechanisms bypassed by sequential mutagenesis, *C.S.H. Symp.* **50**, 813–822.

Gerisch, G., Müller-Taubenberger, A. (2003) GFP-fusion proteins as fluorescent reporters to study organelle and cytoskeleton dynamics in chemotaxis and phagocytosis, *Methods Enzymol.* **361**, 320–337.

Glöckner, G., Eichinger, L., Szafranski, K., Pachebat, J.A., Bankier, A.T., Dear, P.H., Lehmann, R., Baumgart, C., Parra, G., Abril, J.F., Guigo, R., Kumpf, K., Tunggal, B., Cox, E., Quail, M.A., Platzer, M., Rosenthal, A., Noegel, A.A. (2002) Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*, *Nature* **418**, 79–85.

Glöckner, G., Szafranski, K., Winckler, T., Dingermann, T., Quail, M.A., Cox, E., Eichinger, L., Noegel, A.A., Rosenthal, A. (2001) The complex repeats of *Dictyostelium discoideum*, *Genome Res.* **11**, 585–594.

Griffiths, K.R., Champion, A.C., Vesey, G., Williams, K.I. (1996) Isolation of glycosylation mutants in *Dictyostelium discoideum* using flow-cytometry, *Cytometry* **25**, 133–143.

Haberstroh, L., Firtel, R.A. (1990) A spatial gradient of expression of a cAMP-regulated prespore cell type specific gene in *Dictyostelium*, *Genes Dev.* **4**, 596–612.

Han, S.I., Friehs, K., Flaschel, E. (2003) Cultivation of *Dictyostelium discoideum* on an improved synthetic medium in a conventional bioreactor, *Process Biochem.* in press.

Hata, T., Takahashi, M., Tanaka, Y., Urushihara, H. (2001) Total tetra knockout of GP138

multigene family implicated in cell interactions in *Dictyostelium discoideum*, *Gene* **271**, 33–42.

Heikoop, J.C., Grootenhuis, P.D.J., Blaauw, M., Veldema, J.S., Van Haastert, P.J.M., Linskens, M.H.K. (1998) Expression of a bioactive, single-chain choriogonadotropin in *Dictyostelium discoideum*, *Eur. J. Biochem.* **256**, 359–363.

Hellstern, S., Dammann, H., Husain, Q., Mutzel, R. (1997) Overexpression, purification and characterization of *Dictyostelium* calcineurin-A, *Res. Microbiol.* **148**, 335–343.

Hildebrandt, M. (1997) Antisense Application in *Dictyostelium*: A Lower Eukaryotic Model System, in: Lichtenstein, C., Nellen, W. (Eds.) *Antisense Technology – A Practical Approach, IRL Press, Oxford, UK*.

Hodgkinson, S. (1995) GFP in *Dictyostelium*, *Trends Genet.* **11**, 327–328.

Howard, P.K., Ahern, K.G., Firtel, R.A. (1988) Establishment of a transient expression system for *Dictyostelium discoideum*, *Nucleic Acids Res.* **16**, 2613–2623.

Hughes, J.E., Ashktorab, H., Welker, D.L. (1988) Nuclear plasmids in the *Dictyostelium* slime molds, *Dev. Genet.* **9**, 495–504.

Hughes, J.E., Podgorski, G.J., Welker, D.L. (1992) Selection of *Dictyostelium discoideum* transformants and analysis of vector maintenance using live bacteria resistant to G418, *Plasmid* **28**, 46–60.

Hughes, J.E., Welker, D.L. (1989) Copy number control and compatibility of nuclear plasmids in *Dictyostelium discoideum*, *Plasmid* **22**, 215–223.

Insall, R.H., King, J. (2003) Parasexual genetics of *Dictyostelium* gene disruptions: identification of a ras pathway using diploids, *BMC Genet.* **4**, 12.

Iranfar, N., Fuller, D., Loomis, W.F. (2003) Genome-wide expression analyses of gene regulation during early development of *Dictyostelium discoideum*, *Eucaryot. Cell* **2**, 664–670.

Iranfar, N., Fuller, D., Sasik, R., Hwa, T., Laub, M., Loomis, W.F. (2001) Expression patterns of cell-type-specific genes in *Dictyostelium*, *Mol. Biol. Cell* **12**, 2590–2600.

Jermyn, K.A., Duffy, K.T., Williams, J.G. (1989) A new anatomy of the prestalk zone in *Dictyostelium*, *Nature* **340**, 144–146.

Jung, E., Gooley, A.A., Packer, N.H., Karuso, P., Williams, K.L. (1998) Rules for the addition of O-linked N-acetylglucosamine to secreted proteins in *Dictyostelium discoideum – in-vivo* studies on glycosylation of mucin muc1 and muc2 repeats, *Eur. J. Biochem.* **253**, 517–524.

Kalpaxis, D., Zündorf, I., Werner, H., Reindl, N., Boy-Marcotte, E., Jacquet, M., Dingermann, T. (1991) Positive selection for *Dictyostelium discoideum* mutants lacking UMP synthase activity based on resistance to 5-fluoroorotic acid, *Mol. Gen. Genet.* **225**, 492–500.

Karakesisoglou, I., Schleicher, M., Gibbon, B.C., Staiger, C.J. (1996) Plant profilins rescue the aberrant phenotype of profilin-deficient *Dictyostelium* cells, *Cell Motil. Cytoskel.* **34**, 36–47.

Kay, R.R., Williams, J.G. (1999) The *Dictyostelium* genome project: an invitation to species hopping, *Trends Genet.* **15**, 294–297.

Kimmel, A.R., Firtel, R.A. (1985) Sequence organization and developmental expression of an interspersed, repetitive element and associated single-copy DNA sequences in *Dictyostelium discoideum*, *Mol. Cell. Biol.* **5**, 2123–2130.

Knecht, D.A., Loomis, W.F. (1987) Antisense RNA inactivation of myosin heavy chain gene expression in *Dictyostelium discoideum*, *Science* **236**, 1081–1085.

Konfortov, B.A., Cohen, H.M., Bankier, A.T., Dear, P.H. (2000) A high-resolution HAPPY map of *Dictyostelium discoideum* chromosome 6, *Genome Res.* **10**, 1737–1742.

Kuspa, A., Loomis, W.F. (1992) Tagging developmental genes in *Dictyostelium* by restriction enzyme-mediated integration of plasmid DNA, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8803–8807.

Kuspa, A., Loomis, W.F. (1994) Transformation of *Dictyostelium* – gene disruptions, insertional mutagenesis, and promoter traps, *Methods Mol. Genet.* **3**, 3–21.

Kuspa, A., Loomis, W.F. (1996) Ordered yeast artificial chromosome clones representing the *Dictyostelium discoideum* genome, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5562–5566.

Kuspa, A., Sucgang, R., Shaulsky, G. (2001) The promise of a protist: the *Dictyostelium* genome project, *Funct. Integr. Genomics* **1**, 279–293.

Kuwayama, H., Obara, S., Morio, T., Katoh, M., Urushihara, H., Tanaka, Y. (2002) PCR-mediated generation of a gene disruption construct without the use of DNA ligase and plasmid vectors, *Nucleic Acids Res.* **30**, E2.

Leiting, B., Lindner, I.J., Noegel, A.A. (1990) The extrachromosomal replication of *Dictyostelium*

plasmid Ddp2 requires a cis-acting element and a plasmid-encoded trans-acting factor, *Mol. Cell. Biol.* **10**, 3727–3736.

Leiting, B., Noegel, A.A. (1991) The ble Gene of *Streptoalloteichus hindustanus* as a new selectable marker for *Dictyostelium discoideum* confers resistance to phleomycin, *Biochem. Biophys. Res. Comm.* **180**, 1403–1407.

Li, G., Alexander, H., Schneider, N., Alexander, S. (2000) Molecular basis for resistance to the anticancer drug cisplatin in *Dictyostelium*, *Microbiology* **146**, 2219–2227.

Lim, R.W.L., Fechheimer, M. (1997) Overexpression, purification, and characterization of recombinant *Dictyostelium discoideum* calcium-regulated 34,000-dalton F-actin bundling protein from *Escherichia coli*, *Protein Expr. Purif.* **9**, 182–190.

Liu, T.Y., Williams, J.G., Clarke, M. (1992) Inducible expression of calmodulin antisense RNA in *Dictyostelium* cells inhibits the completion of cytokinesis, *Mol. Biol. Cell* **3**, 1403–1413.

Maeda, M., Sakamoto, H., Iranfar, N., Fuller, D., Maruo, T., Ogihara, S., Morio, T., Urushihara, H., Tanaka, Y., Loomis, W.F. (2003) Changing patterns of gene expression in *Dictyostelium* prestalk cell subtypes recognized by in situ hybridization with genes from microarray analyses, *Eukaryotic Cell* **2**, 627–637.

Maniak, M., Saur, U., Nellen, W. (1989) A colony-blot technique for the detection of specific transcripts in eukaryotes, *Anal. Biochem.* **176**, 78–81.

Manstein, D.J., Schuster, H.P., Morandini, P., Hunt, D.M. (1995) Cloning vectors for the production of proteins in *Dictyostelium discoideum*, *Gene* **162**, 129–134.

Manstein, D.J., Titus, M.A., De Lozanne, A., Spudich, J.A. (1989) Gene replacement in *Dictyostelium*: generation of myosin null mutants, *EMBO J.* **8**, 923–932.

Martens, H., Novotny, J., Oberstrass, J., Steck, T.L., Postlethwait, P., Nellen, W. (2002) RNAi in *Dictyostelium*: the role of RNA-directed RNA polymerases and double-stranded RNase, *Mol. Biol. Cell* **13**, 445–453.

May, T., Kern, H., Müller-Taubenberger, A., Nellen, W. (1989) Identification of a cis-acting element controlling induction of early gene expression in *Dictyostelium discoideum*, *Mol. Cell Biol.* **9**, 4653–4659.

Metz, B.A., Ward, T.E., Welker, D.L., Williams, K.L. (1983) Identification of an endogenous plasmid in *Dictyostelium discoideum*, *EMBO J.* **2**, 515–519.

Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M., Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H., Williams, J., Maeda, M., Takeuchi, I., Ochiai, H., Tanaka, Y. (1998) The *Dictyostelium* developmental cDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development, *DNA Res.* **5**, 335–340.

Morrison, A., Marschalek, R., Dingermann, T., Harwood, A.J. (1997) A novel, negative selectable marker for gene disruption in *Dictyostelium*, *Gene* **202**, 171–176.

Nellen, W., Firtel, R.A. (1985) High-copy-number transformants and co-transformation in *Dictyostelium*, *Gene* **39**, 155–163.

Nellen, W., Hildebrandt, M., Mahal, B., Möhrle, A., Kröger, P., Maniak, M., Oberhauser, R., Sadiq, M. (1992) Mechanisms of gene regulation by endogenous and artificially introduced antisense RNA, *Biochem. Soc. Trans.* **20**, 750–754.

Nellen, W., Saur, U. (1988) Cell-cycle dependent transformation competence in *Dictyostelium discoideum*, *Biochem. Biophys. Res. Commun.* **154**, 54–59.

Nellen, W., Silan, C., Firtel, R.A. (1984) DNA-mediated transformation in *Dictyostelium discoideum*: regulated expression of an actin gene fusion, *Mol. Cell Biol.* **4**, 2890–2898.

Noegel, A., Metz, B.A., Williams, K.L. (1985) Developmentally regulated transcription of *Dictyostelium discoideum* plasmid Ddp1, *EMBO J.* **4**, 3797–3803.

Ogawa, S., Yoshino, R., Angata, K., Iwamoto, M., Pi, M., Kuroe, K., Matsuo, K., Morio, T., Urushihara, H., Yanagisawa, K., Tanaka, Y. (2000) The mitochondrial DNA of *Dictyostelium discoideum*: complete sequence, gene content and genome organization, *Mol. Gen. Genet.* **263**, 514–519.

Palmieri, S.J., Nebl, T., Pope, R.K., Seastone, D.J., Lee, E., Hinchcliffe, E.H., Sluder, G., Knecht, D., Cardelli, J., Luna, E.J. (2000) Mutant Rac1B expression in *Dictyostelium*: effects on morphology, growth, endocytosis, development, and the actin cytoskeleton, *Cell Motil. Cytoskel.* **46**, 285–304.

Pi, M., Morio, T., Urushihara, H., Tanaka, Y. (1998) Characterization of a novel small RNA encoded by *Dictyostelium discoideum*

mitochondrial DNA, *Mol. Gen. Genet.* **257**, 124–131.

Primpke, G., Iassonidou, V., Nellen, W., Wetterauer, B. (2000) Role of cAMP-dependent protein kinase during growth and early development of *Dictyostelium discoideum*, *Dev. Biol.* **221**, 101–111.

Rivero, F. (2002) mRNA processing in *Dictyostelium*: sequence requirements for termination and splicing, *Protist* **153**, 169–176.

Robinson, D.N., Spudich, J.A. (2000) Dynacortin, a genetic link between equatorial contractility and global shape control discovered by library complementation of a *Dictyostelium discoideum* cytokinesis mutant, *J. Cell Biol.* **150**, 823–838.

Rosen, E., Sivertsen, A., Firtel, R.A. (1983) An unusual transposon encoding heat shock inducible and developmentally regulated transcripts in *Dictyostelium*, *Cell* **35**, 243–251.

Rupper, A., Grove, B., Cardelli, J. (2001) Rab7 regulates phagosome maturation in *Dictyostelium*, *J. Cell Sci.* **114**, 2449–2460.

Sauer, B. (1998) Inducable gene targeting in mice using the Cre/lox system, *Methods* **14**, 381–392.

Schindl, M., Wallraff, E., Deubzer, B., Witke, W., Gerisch, G., Sackmann, E. (1995) Cell-substrate interactions and locomotion of *Dictyostelium* wild-type and mutants defective in three cytoskeletal proteins: a study using quantitative reflection interference contrast microscopy, *Biophys. J.* **68**, 1177–1190.

Schneider, N., Weber, I., Faix, J., Prassler, J., Müller-Taubenberger, A., Köhler, J., Burghardt, E., Gerisch, G., Marriott, G. (2003) A Lim protein is involved in the progression of cytokinesis and the regulation of the mitotic spindle, *Cell Motil. Cytoskel.* **56**, 130–139.

Shah-Mahoney, N., Hampton, T., Vidaver, R., Ratner, D. (1997) Blocking the ends of transforming DNA enhances gene targeting in *Dictyostelium*, *Gene* **203**, 33–41.

Shammat, I.M., Gonzales, C.M., Welker, D.L. (1998) *Dictyostelium discoideum* nuclear plasmid Ddp6 is a new member of the Ddp2 plasmid family, *Curr. Genet.* **33**, 77–82.

Shammat, I.M., Welker, D.L. (1999) Mechanism of action of the Rep protein from the *Dictyostelium* Ddp2 plasmid family, *Plasmid* **41**, 248–259.

Shaw, D.R., Richter, H., Giorda, R., Ohmachi, T., Ennis, H.L. (1989) Nucleotide sequences of *Dictyostelium discoideum* developmentally regulated cDNAs rich in (AAC) imply proteins that contain clusters of asparagine, glutamine, or threonine, *Mol. Gen. Genet.* **218**, 453–459.

Slade, M.B., Emslie, K.R., Williams, K.L. (1997) Expression of recombinant glycoproteins in the simple eukaryote *Dictyostelium discoideum*, *Biotech. Gen. Eng. Rev.* **14**, 1–35.

Spann, T.P., Brock, D.A., Lindsey, D.F., Wood, S.A., Gomer, R.H. (1996) Mutagenesis and gene identification in *Dictyostelium* by shotgun antisense, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5003–5007.

Srinivasan, S., Traini, M., Herbert, B., Sexton, D., Harry, J., Alexander, H., Williams, K.L., Alexander, S. (2001) Proteomic analysis of a developmentally regulated secretory vesicle, *Proteomics* **1**, 1119–1127.

Sucgang, R., Chen, G., Liu, W., Lindsay, R., Lu, J., Muzny, D., Shaulsky, G., Loomis, W., Gibbs, R., Kuspa, A. (2003) Sequence and structure of the extrachromosomal palindrome encoding the ribosomal RNA genes in *Dictyostelium*, *Nucleic Acids Res.* **31**, 2361–2368.

Sutoh, K. (1993) A transformation vector for *Dictyostelium discoideum* with a new selectable marker bsr, *Plasmid* **30**, 150–154.

Szafranski, K., Glockner, G., Dingermann, T., Dannat, K., Noegel, A.A., Eichinger, L., Rosenthal, A., Winckler, T. (1999) Non-LTR retrotransposons with unique integration preferences downstream of *Dictyostelium discoideum* tRNA genes, *Mol. Gen. Genet.* **262**, 772–780.

Thompson, C.R., Bretscher, M.S. (2002) Cell polarity and locomotion, as well as endocytosis, depend on NSF, *Development* **129**, 4185–4192.

Tsiavaliaris, G., Fujita-Becker, S., Batra, R., Levitsky, D.I., Kull, F.J., Geeves, M.A., Manstein, D.J. (2002) Mutations in the relay loop region result in dominant-negative inhibition of myosin II function in *Dictyostelium*, *EMBO Rep.* **3**, 1099–1105.

Urushihara, H. (2002) Functional genomics of the social amoebae, *Dictyostelium discoideum*, *Mol. Cells* **13**, 1–4.

Van Driessche, N., Shaw, C., Katoh, M., Morio, T., Sucgang, R., Ibarra, M., Kuwayama, H., Saito, T., Urushihara, H., Maeda, M., Takeuchi, I., Ochiai, H., Eaton, W., Tollett, J., Halter, J., Kuspa, A., Tanaka, Y., Shaulsky, G. (2002) A transcriptional profile of multicellular development in *Dictyostelium discoideum*, *Development* **129**, 1543–1552.

Venter, J.C., and the Human Genome Consortium (276 co-authors) (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Voith, G., Kramm, H., Zündorf, I., Winkler, T., Dingermann, T. (1998) Expression of the rat muscarinic receptor gene m3 in *Dictyostelium discoideum*, *Pharmazie* **53**, 707–710.

Wallraff, E., Gerisch, G. (1991) Screening for *Dictyostelium* mutants defective in cytoskeletal proteins by colony immunoblotting, *Methods Enzymol.* **196**, 334–348.

Welker, D.L., Williams, K.L. (1982) A genetic map of *Dictyostelium discoideum* based on mitotic recombination, *Genetics* **102**, 691–710.

West, C.M., Loomis, W.F. (1985) Absence of a carbohydrate modification does not affect the level or subcellular localization of three membrane glycoproteins in modB mutants of *Dictyostelium discoideum*, *J. Biol. Chem.* **260**, 13803–13809.

Wetterauer, B., Morandini, P., Hribar, I., Murgia-Morandini, I., Hamker, U., Singleton, C., MacWilliams, H.K. (1996) Wild-type strains of *Dictyostelium discoideum* can be transformed using a novel selection cassette driven by the promoter of the ribosomal v18 gene, *Plasmid* **36**, 169–181.

Wetterauer, B., Salger, K., Demel, P., Koop, H. (2000) Efficient transformation of *Dictyostelium discoideum* with a particle inflow gun, *Biochim. Biophys. Acta* **1499**, 139–143.

Williams, R.S., Eames, M., Ryves, W.J., Viggars, J., Harwood, A.J. (1999) Loss of a prolyl oligopeptidase confers resistance to lithium by elevation of inositol (1,4,5) trisphosphate, *EMBO J.* **18**, 2734–2745.

Winckler, T. (1998) Retrotransposable elements in the *Dictyostelium discoideum* genome, *Cell. Mol. Life Sci.* **54**, 383–393.

Winckler, T., Dingermann, T., Glöckner, G. (2002) *Dictyostelium* mobile elements: strategies to amplify in a compact genome, *Cell. Mol. Life Sci.* **59**, 2097–2111.

Winckler, T., Trautwein, C., Tschepke, C., Neuhauser, C., Zündorf, I., Beck, P., Vogel, G., Dingermann, T. (2001) Gene function analysis by amber stop codon suppression: CMBF is a nuclear protein that supports growth and development of *Dictyostelium* amoebae, *J. Mol. Biol.* **305**, 703–714.

Witke, W., Nellen, W., Noegel, A. (1987) Homologous recombination in the *Dictyostelium* α-actinin gene leads to an altered mRNA and lack of the protein, *EMBO J.* **6**, 4143–4148.

Yan, J.X., Tonella, L., Sanchez, J.C., Wilkins, M.R., Packer, N.H., Gooley, A.A., Hochstrasser, D.F., Williams, K.L. (1997) The *Dictyostelium discoideum* proteome – the SWISS-2DPAGE database of the multicellular aggregate (slug), *Electrophoresis* **18**, 491–497.

Zeng, C., Anjard, C., Riemann, K., Konzok, A., Nellen, W. (2000) gdt1, a new signal transduction component for negative regulation of the growth-differentiation transition in *Dictyostelium discoideum*, *Mol. Biol. Cell* **11**, 1631–1643.

Zhang, N., Long, Y., Devreotes, P.N. (2002) Ege A, a novel C2 domain containing protein, is essential for GPCR- mediated gene expression in *Dictyostelium*, *Dev. Biol.* **248**, 1–12.

DNA Damage Induced by Ultraviolet Radiation

David L. Mitchell
The University of Texas M.D. Anderson Cancer Center
Department of Carcinogenesis Smithville, Texas, USA

1 Effects of UVR on Living Systems 353

2 The Electromagnetic Spectrum 354

3 DNA Photoproducts 355
3.1 Pathways for Dissipating Absorbed Energy 355
3.2 Types of DNA Photoproducts 355
3.3 DNA Photoproduct Distribution 358

4 Analytical Techniques 361
4.1 Detecting and Quantifying DNA Photoproducts 361
4.2 Quantifying DNA Repair 363

5 DNA Damage Tolerance Mechanisms 364
5.1 Photoprotection 365
5.2 Photoenzymatic Repair 366
5.3 Excision Repair 368
5.3.1 Excision Repair Pathways 368
5.3.2 Temporal and Spatial Distribution of Excision Repair 371
5.3.3 Developmental and Genetic Regulation of Excision Repair 372

6 Biological Consequences of DNA Repair 376
6.1 Photoaging 377
6.2 Immunosuppression 377
6.3 Photocarcinogenesis 378
6.4 Effects on the Environment 379

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

Bibliography 380
Books and Reviews 380
Primary Literature 380

Keywords

Dimerization
Covalent bonding of adjacent purine or pyrimidine bases resulting from the resolution of the unstable electronic configuration created by absorption of a photon within the UVR range of light.

Endonuclease
An enzyme, often associated with the excision repair process that cleaves a phosphodiester bond to either the 3′ or 5′ side of a base or abasic site in DNA.

Nucleotide Excision Repair
Replacement of damaged DNA with nascent, corrected DNA involving recognition, unwinding, incision, excision, resynthesis, and ligation.

Photohydrate
A type of monobasic photoproduct to which a hydroxyl (–OH) group has been added to the 5 or 6 carbon of a pyrimidine base.

Photoenzymatic Repair (PER)
A light-dependent process involving the enzyme-catalyzed reversal of CPDs.

Photoproduct
A stable change in DNA structure involving a single base or adjacent bases resulting from the dissipation of absorbed UVR.

Reactive Oxygen Species (ROS)
Highly reactive, short-lived chemical species containing oxygen formed by the interaction of ionizing or nonionizing radiation with water.

Ultraviolet Radiation (UVR)
Wavelengths of the electromagnetic spectrum ranging from 190 nm to 400 nm; UVR is subdivided into vacuum UV (190–240 nm), UVC (240–290 nm), UVB (290–320 nm), and UVA (320–400 nm).

Xeroderma Pigmentosum
A rare autosomal recessive disease, displaying sun sensitivity resulting in progressive degenerative changes in sun-exposed portions of the skin and eyes, often leading to neoplasia.

It has been known for more than a hundred years that ultraviolet radiation (UVR) can have adverse effects on the growth and development of living organisms. Degradation of stratospheric ozone and the resultant increase in UVB radiation at the earth's surface has focused our attention on the public health and environmental effects of solar DNA damage. With the advent and advancement of molecular biology over the past three decades, our understanding of the biochemical foundation of the cytotoxic, mutagenic, and carcinogenic effects of UVR has greatly increased. Armed with an arsenal of powerful techniques, today's researchers are able to probe the intricate structure of UVR damage induction and repair in DNA. Indeed, recent advances in deciphering DNA repair in bacteria, yeast, and humans have greatly increased our understanding of the molecular mechanisms underlying this complex and essential process. In concert with public awareness and political action, the adverse effects of global stratospheric deozonation can be mitigated by the development and application of basic photobiological research.

1 Effects of UVR on Living Systems

UVR damage in cellular DNA can be resolved in several ways (Fig. 1), depending on the type of lesion produced, the genomic location of the damage, the type of cell affected, and the developmental state of the cell. It is plausible that some lesions are not perceived as different from normal DNA by the cell. However, it is more probable that, unless the damage is repaired, it will disrupt the normal operation of vital cellular processes such as DNA replication or transcription. In this event, cell proliferation will cease or, if the lesion is situated in a gene required for an essential metabolic function, the cell will die.

Some types of photodamage are more effective than others at blocking the progression of DNA or RNA polymerases. Types of DNA damage that facilitate lesion bypass by the replication machinery may allow misincorporation of an incorrect complementary base, thus producing a mutation. A mutation may have several outcomes. First, it may be benign, neither altering the genetic code nor affecting normal metabolism. Second, the mutation may create a truncated or partial RNA transcript and a dysfunctional protein. If the protein is essential, then the mutation is lethal. Finally, it may result in activation of an oncogene or inactivation of a tumor suppressor gene, resulting in the initiation of cell transformation and carcinogenesis.

Although a direct role of DNA damage in replication and transcription is well documented, other pathways in which DNA damage can affect cell rate and metabolism have been described and operate through various signal transduction pathways. For instance, DNA damage in the skin resulting from extended exposure to solar UVR is implicated in immunosuppression, erythema (sunburn), and photoaging. On a more global scale, increased DNA damage in the environment resulting from decreases in stratospheric ozone has been linked to reduced productivity in marine ecosystems, effects on food crops, and livestock and has been implicated in the global amphibian decline.

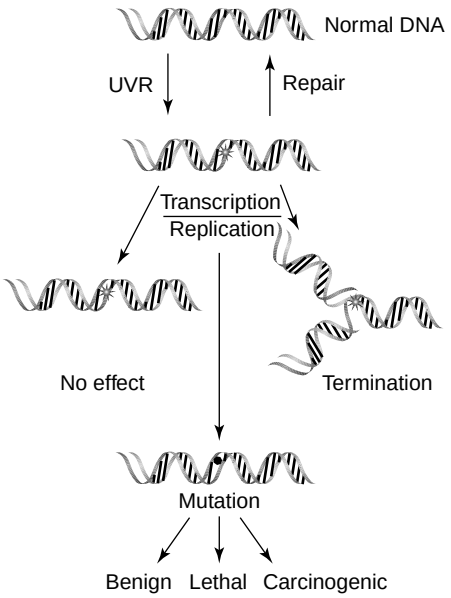

Fig. 1 Biological consequences of UVR. Molecular events that lead to cell killing, mutation induction, and carcinogenesis are shown.

2 The Electromagnetic Spectrum

The sun emits a broad array of energies at wavelengths ranging through 11 orders of magnitude from 1 nm to 100 m (Fig. 2). Much of this energy is biologically irrelevant; long wavelength, low energy emissions such as far infrared and microwaves are highly unlikely to impart sufficient energy to influence biochemical reactions. Likewise, visible and near infrared wavelengths, although very important for vision and photosynthesis, are not considered hazardous to living organisms. Finally, ionizing radiations such as high-energy particles, X-rays, and gamma rays are expended by atomic collisions in the upper atmosphere. Hence, UVR emerges as the most biologically relevant component of the solar spectrum with respect to DNA damage.

The UV spectrum is divided into three regions: UVC (240–290 nm), UVB (290–320 nm), and UVA (320–400 nm). Because of their high energies and availability, germicidal (UVC) lamps have historically been used for photochemical and photobiological studies. However, since the earth's atmosphere absorbs UVR wavelengths below 300 nm, other components of the solar spectrum, such as UVB and UVA, are of more concern. Although the UVA and UVB constitute a negligible portion of the sun's energy, they are primarily responsible for the pathological effects of solar radiation. For this reason, recent research has been directed more toward characterizing DNA damages produced by these biologically relevant solar wavelengths.

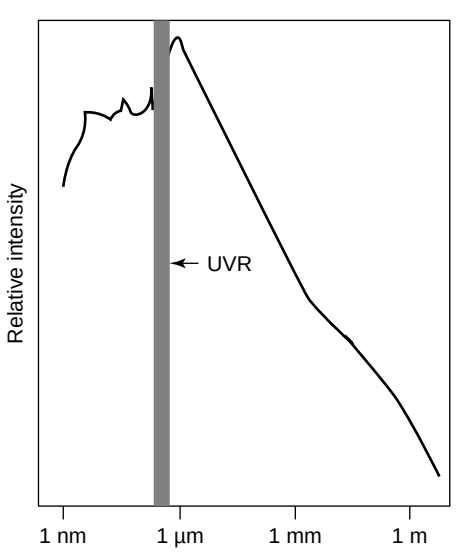

Fig. 2 The electromagnetic spectrum. Energy emitted by the sun is shown for wavelengths between 1 Å and 100 m. The ultraviolet spectrum (190–400 nm) is indicated by the shaded line.

3
DNA Photoproducts

3.1
Pathways for Dissipating Absorbed Energy

Photon absorption rapidly converts a pyrimidine base to an excited state (Fig. 3). This event promotes an electron in a filled bonding π-orbital in the singlet ground state into a higher energy, empty π^* antibonding orbital, thus initiating photoproduct formation. Formation of an excited base occurs within 10^{-12} s after photon absorption. Various pathways are then available for resolution of this unstable electronic configuration. The major pathway involves rapid dissipation of the energy of the excited singlet base to the ground state (10^{-9} s) by nonradiative transition or by fluorescence, yielding heat or light in the process. Secondly, the excited base can react with other molecules to form unstable intermediates (i.e. free radicals) or stable photoproducts. Finally, there is a low probability that intersystem crossing, a nonradiative pathway, can transfer a base from the excited singlet state to the excited triplet state. The lifetime of the triplet state is several orders of magnitude longer than the excited singlet state (10^{-3} s), increasing the chance of photoproduct formation. Formation of the triplet state can be greatly enhanced in the presence of sensitizers such as acetone or acetophenone. The cyclobutane pyrimidine dimer (I in Fig. 4) forms through the excited triplet state, whereas, other photoproducts form by other mechanisms.

3.2
Types of DNA Photoproducts

The most abundant photoproduct induced by UVR is the cyclobutane pyrimidine

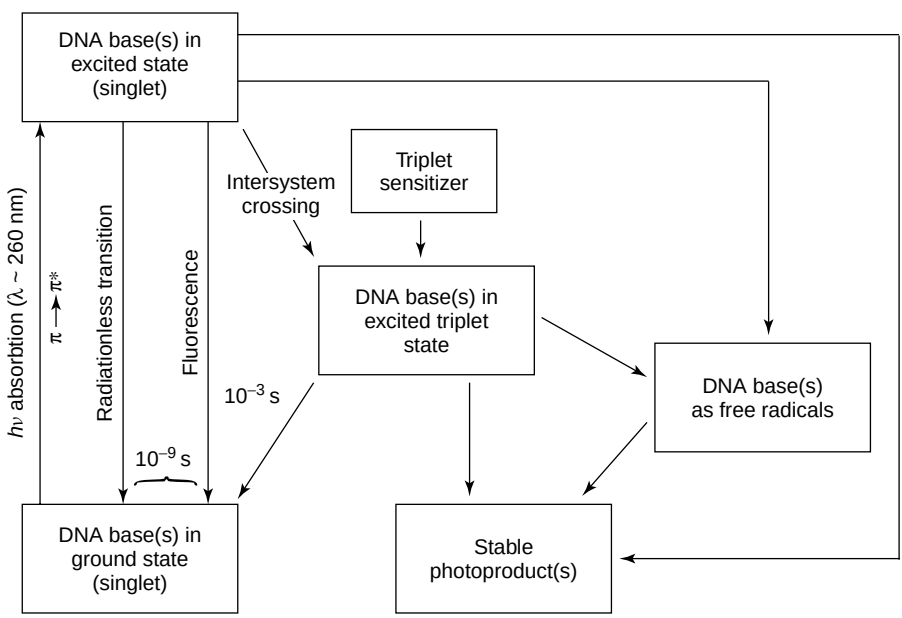

Fig. 3 Pathways of dissipation of energy absorbed by UVR. Schematic of the routes of UVR absorption and resolution leading to photoproduct formation.

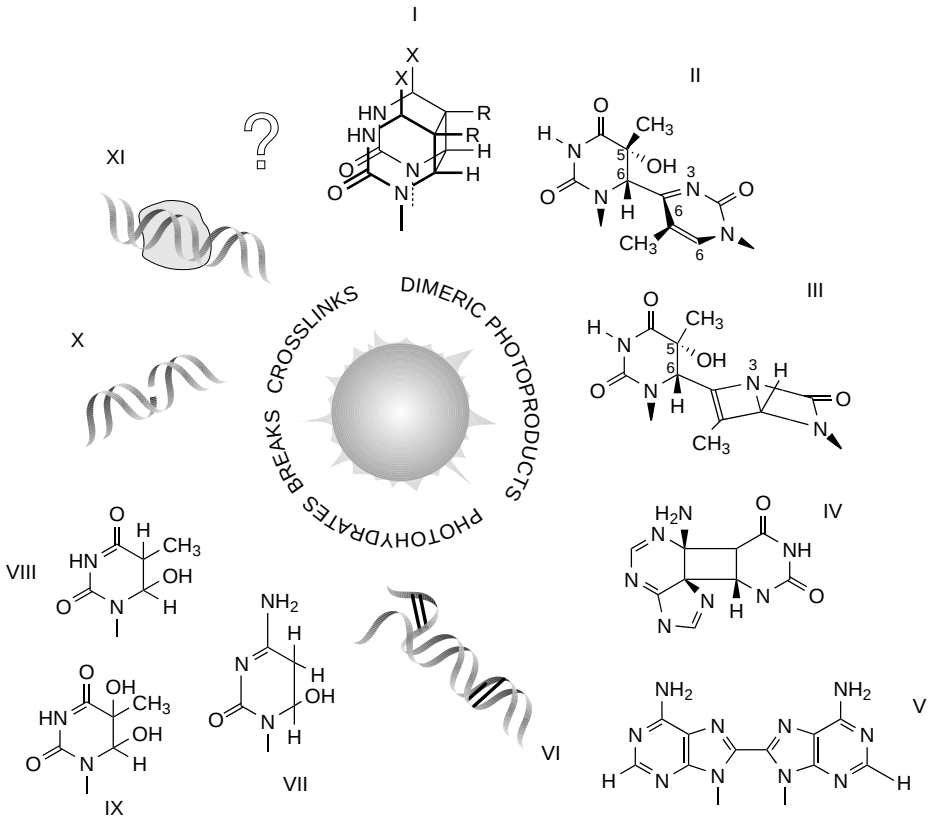

Fig. 4 DNA photoproducts. Types and structures of known photoproducts induced in DNA by solar UV radiation. I = CPD; II = (6-4)PD; III = Dewar pyrimidinone; IV = adenine–thymine dimer; V = 8,8 adenine dehydrodimer; VI = inter- and intrastrand DNA cross-links; VII = 6-hydroxy-5,6-dihydrocytosine; VIII = 6-hydroxy-5,6-dihydrothymine; IX = 5,6-dihydroxy-5,6-dihydrothymine (thymine glycol); X = single-strand break; XI = DNA-protein cross-link.

dimer or CPD (I) linked by a cyclobutyl ring between the 5 and 6 carbons of adjacent thymine (T) or cytosine (C) bases (Fig. 5). Of the four possible CPD combinations (T<>T, C<>T, T<>C, and C<>C), the T<>T dimer occurs with greatest frequency at 40 to 60% of the sum of cytosine-containing dimers. However, at wavelengths above 300 nm, the photochemical alterations of cytosine, rather than thymine, may account for most biological effects. The primary reason for this is that at wavelengths below 300 nm, saturation of CPD-containing cytosine occurs at a lower fluence and at a lower level than that of thymine homodimers (T<>T). Hence, whereas T<>T dimers form at a 1.5-fold *greater* frequency than dimers containing cytosine after UVC irradiation, they occur with one-third *less* frequency after UVB irradiation.

The pyrimidine(6-4)pyrimidone dimer or (6-4)PD (II) is the second most prevalent type of UV damage, occurring at 5 to

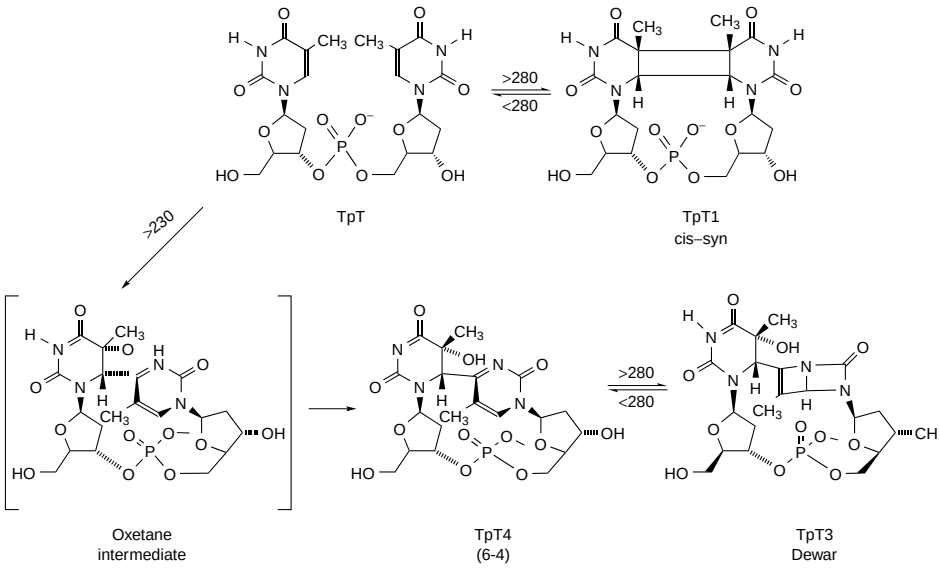

Fig. 5 Reaction scheme for the formation of the major UV photoproducts photochemical reactions in a dipyrimidine DNA sequence leading to the formation of the CPD (TpT1), (6-4)PD (TpT4), and Dewar pyrimidinone (TpT3). (Redrawn from Taylor and Cohrs, *J. Am. Chem. Soc.* **109**: 2834–2835 (1987), with permission).

50% of the frequency of the CPD. In this photoproduct, the 6 carbon of a 3′ pyrimidine (either cytosine or thymine) is linked to the 4 carbon of a 5′ pyrimidine (primarily cytosine) by way of an unstable oxetane or azetidine intermediate. Unlike CPDs, the formation of (6-4)PDs does not proceed via an excited triplet state (Fig. 3), hence is less dependent on the photochemical behavior of thymine bases in DNA. Infrared spectroscopic and nuclear magnetic resonance analyses of the (6-4)PD indicate that the plane of the 3′ cytosine base is shifted 90° relative to the 5′ thymine.

As shown in the reaction scheme for dimer formation (Fig. 5), absorption of UVB light by the (6-4)PD results in the formation of a structurally related photoisomer, the Dewar pyrimidinone (III in Fig. 4). Rare dimers containing purine, such as the thymine–adenine dimer (IV) and the 8,8-adenine dehydrodimer (V) have also been isolated from heavily irradiated DNA. Dimerizations between nonadjacent pyrimidines either on the same strand or on opposite strands can produce intra- or interstrand cross-links (VI). It should be kept in mind that the relationship between photoproduct frequency and its biological effect depends on the intrinsic capability of that lesion to kill cells or to induce mutations. Hence, even though a photoproduct may occur at a very low frequency, its structure and location may elicit a potent biological effect.

Compared with photoproducts involving two pyrimidines, those involving a single base occur in DNA at very low frequencies (≈1–2% of dimers) after exposure to UVR. Because of its instability, the photobiology of the cytosine photohydrate (VII) has been

difficult to study using the conventional analytical procedures for quantifying DNA damage. It should be kept in mind, however, that the half-life of the cytosine photohydrate is not trivial (i.e. 24 h at room temperature) and this lesion may indeed pose a viable biological threat to a proliferating cell system. Recent data using sequencing analysis indicate that saturated thymine photoproducts (VIII and IX), such as the thymine glycol, do not occur in UVC- or UVB-irradiated DNA at a significant rate, hence, they do not appear to contribute significantly to the biological effects of UVR.

In contrast with the induction of DNA damage resulting from the direct absorption of UVC light, UVA and UVB may produce damage indirectly through highly reactive chemical intermediates. Similar to ionizing radiation, these wavelengths generate reactive oxygen species (ROS) that can form, for example, singlet oxygen, hydroxyl and peroxyl radicals, and superoxide anions which in turn can react with DNA to form single base damage, such as 8-oxodeoxyguanosine, as well as strand breaks and DNA-protein cross-links (X and XI in Fig. 4). Although the number of cytosine photohydrates compared to CPDs and (6–4)PDs remains constant throughout the UVR, the relative proportion of single-strand breaks and protein DNA cross-links increases significantly in DNA irradiated with UVB and UVA light.

3.3 DNA Photoproduct Distribution

A major pathway of cell killing by UVR results from DNA damage that inhibits DNA synthesis or transcription of genes essential to cell growth and metabolism. Recent advances in biotechnology have greatly increased our understanding of the diverse nature of DNA damage in mammalian systems, particularly regarding the spatial and temporal distribution of UV photoproducts in mammalian chromatin. Photochemical and biochemical data indicate that the distribution of UVR damage is nonrandom, determined by DNA sequence and DNA–protein interactions.

Using biochemical and immunological techniques, (6-4)PDs were quantified at approximately 23% of the level of CPDs in purified DNA irradiated *in vitro*. However, the relative occurrence of (6-4)PDs and CPDs varies widely, depending on base content of DNA, state of methylation, and chromatin structure. For example, the ratio of CPDs and (6-4)PDs correlates with the AT : GC ratio with CPDs occurring in *Clostridium perfringens* DNA (30% AT : GC) > *Escherichia coli* DNA (50% AT : GC) > *Micrococcus luteus* DNA (70% AT : GC). Using DNA sequencing techniques, the incidence of (6-4)PDs and CPDs in fragments of the *lacI* and *supF* genes is highly variable and sequence-dependent. Specifically, CPDs occur predominantly at T–T, less frequently at T–C and C–T, and least often at C–C sites. In contrast, (6-4)PDs can occur with 100-fold greater frequency than CPDs at certain dipyrimidine sites. Specifically, (6-4)PDs form predominantly at T–C, less frequently at T–T and C–C and rarely at C–T sites. Cytosine methylation significantly increased the yield of (6-4)PDs after exposure to UVC and of CPDs and (6-4)PDs after exposure to UVB. Hence, both types of dimers are preferentially induced at 5-methylcytosine bases in DNA of cells exposed to solar UVR.

The distribution of UV photoproducts is influenced by secondary structure mediated through DNA–protein interactions; namely, chromatin structure modulates sites of CPD induction in core nucleosome

DNA. Using the 3′–5′ exonuclease activity of T4 DNA polymerase to map UV photoproducts, it was shown that they do not occur randomly along the DNA helix, but are situated with a periodicity of 10.3 bases. Sites of maximum photoproduct induction in core DNA correspond to sites in which the phosphate backbone is farthest from the core histone surface. The limited flexibility of the DNA helix resulting from histone–DNA interactions is thought to inhibit photoproduct formation.

The relative induction of (6-4)PDs and CPDs has been measured in nucleosome (core) and internucleosome (linker) DNA in UV-irradiated human chromatin. CPDs form in equal amounts per nucleotide in core and linker DNA, whereas (6-4)PDs occur with sixfold greater frequency per nucleotide in linker DNA. These data indicate that (6-4)PD formation is strongly influenced by DNA–protein interactions, occurring more readily in DNA that is less tightly bound to nucleosomal proteins. These studies suggest that (6-4)PDs may be more readily induced in actively transcribing genes compared to quiescent genes or in the overall genome in human fibroblasts. Preferential induction of (6-4)PDs in functionally active regions of chromatin would serve to increase the lethal potential of this type of photodamage. Indeed, repair of these lesions (see below) may be facilitated by their predominance in nonnucleosomal (metabolically active) regions of the genome; DNA damage in more open chromatin conformations (CC) may be more accessible to excision repair enzymes compared to damage induced in more compact regions.

Mutagenicity of UV photoproducts and low fidelity DNA polymerases The progression of replicative class B polymerases, including Pol A, D, and E are blocked by most types of photoproducts. Depending on their structures, many photoproducts are, however, bypassed during DNA replication to different degrees by damage-specific class-Y DNA polymerases including Pol H, I, and K that have low fidelities due to expanded active sites that facilitate read-through of noninformative sequence information (e.g. abasic sites). Pol H has the greatest capacity to replicate a large variety of DNA lesions and preferentially inserts adenine in the nascent strand opposite the lesion (called the *A rule*). Pol I preferentially inserts guanine and is capable of replicating C-containing photoproducts. Pol H or I, therefore, can insert bases opposite dipyrimidine photoproducts, but the 3′ complementary base will be mismatched either by an erroneous insertion or the distortion caused by the photoproduct. The absence of Pol H results in increased mutagenesis and has the most recognizable pathological consequence in the human disease Xeroderma pigmentosum variant complementation group.

This mechanism has two important implications regarding the mutagenicity of different photoproducts. First, mutations will most often occur where cytosine is a component of the photoproduct since insertion of adenine opposite thymine is a correct and nonmutagenic event. Hence, most CPDs, because they form between two thymine bases, are not mutagenic. Second, the more distortive a lesion is, the more likely it will block DNA synthesis and will result in a lethal rather than a mutagenic event. Since the (6-4)PD is considerably more distortive than the CPD (i.e. it causes a 47° as opposed to a 7° helical bend) it is more likely to be lethal rather than mutagenic. Because damage bypass and base insertion depend on a

variety of conditions, both CPDs and (6-4)PDs contribute to tumorigenesis in a complex manner.

The role of specific photoproducts in UV mutagenesis in human cells has been investigated with the use of shuttle vectors. In these systems, UV-irradiated simian virus SV40-based plasmids are transfected into mammalian host cells (e.g. mouse and human) where they are replicated. The plasmids are subsequently recovered, amplified in bacteria, and analyzed for mutations by DNA sequencing. Sites of mutations can then be compared with sites of photoproduct induction in the target sequence. Results of these studies indicate that sites of transition mutations correlate with sites of increased (6-4)PD induction (Table 1). In the *supF* gene inserted into the mouse L-cell chromosome and in the endogenous *APRT* gene of Chinese hamster ovary (CHO) cells, most of the mutations consisted of C→T transitions occurring at T–C and C–C sequences. Owing to the strand-specificity of repair (see below), a bias was observed between mutations in the coding and the noncoding strands of expressed genes that differed according to the repair capacity of the cells.

Because CPDs and (6-4)PDs can both form at sequences shown to be mutation hotspots in shuttle vectors, the identity of the mutagenic lesion was also tested by photoenzymatic repair (PER) (see below) of the *supF* sequence in plasmids prior to transfection. Specific photoreversal of CPDs reduced the mutation frequency in normal cells by 75% and in XP group A cells by 90%. These results are not consistent with the model developed in *E. coli* and suggest that (6-4)PDs may be less mutagenic in human cells. A comparison of photoproduct yields, rates of repair, and mutations in the *PGKI* and p53 genes,

Tab. 1 UVC-induced mutations observed in shuttle vector pZ189 replicated in XP or normal human cells.

Mutations	***Number of plasmids with base changes***	
	XP	***Normal***
Independent plasmids sequenced	61 (100%)	89 (100%)
Point mutations		
Single base substitution	47 (77%)	48 (53%)
Tandem base substitutions	12 (20%)	16 (18%)
Multiple base substitutions	1 (2%)	24 (28%)
Base insertions and deletions		
Single base insertion	0	2
Single or tandem base deletions	1	3
Types of single or tandem base substitutions and number of changes		
Transitions	67 (94%)	61 (75%)
GC to AT	66 (93%)	59 (73%)
AT to GC	1 (1%)	2 (2%)
Transversions	4[d] (6%)	20 (25%)
GC to TA	0[d]	8 (10%)
GC to CG	1 (1%)	5 (6%)
AT to TA	3 (4%)	6 (8%)
AT to CG	0	1 (1%)

however, has shown that regions of high UV-induced mutation can be caused by either or both high photoproduct yield and low repair. A combination of initial yields and rates of repair that leave a high net-persistent load of photoproducts in a particular site appear to be directly related to the mutational yield. Using ligation-mediated polymerase chain reaction (LMPCR) that allows precise location of damaged bases, the photoproduct distribution in exons 1 and 2 of three *ras* proto-oncogenes was mapped and no correlation between photoproduct frequency and mutation induction in codon 12 of H-*ras* and K-*ras* was found. DNA repair at individual nucleotides in the p53 tumor suppressor gene was highly variable and sequence-dependent with slow repair observed at seven of the eight positions associated with mutations. UV-induced mutations in the p53 gene are a probable step in the formation of squamous cell carcinoma (SCC) and may arise at DNA repair "coldspots" rather than photoproduct "hotspots."

4 Analytical Techniques

4.1 Detecting and Quantifying DNA Photoproducts

Methods for quantifying photoproducts in DNA rely on (1) separation of modified bases by chromatography, (2) enzymatic or biochemical incision of DNA at sites of photoproducts, or (3) antibody binding to structural damage in DNA (Table 2). Damage caused by UVR was first detected in DNA nearly 40 years ago using paper chromatography. Chromatographic measurements require that DNA photoproducts either be labeled during cell proliferation by incorporation of a radioactive pyrimidine precursor, such as [^{3}H]-thymidine, or postlabeled using [$\delta-^{32}$P] ATP after enzymatic hydrolysis. After UV-irradiation, the DNA is extracted, purified, and reduced to individual bases by acid or enzymatic hydrolysis. Photoproducts are then separated from unmodified bases by two-dimensional paper, thin-layer, or high-performance liquid chromatography (HPLC). A standard procedure for quantifying 8-oxodeoxyguanosine residues in DNA after hydrolysis is HPLC combined with electrochemical detection of the lesion (HPLC-EC).

Enzymatic or chemical incision of DNA at sites of damage and subsequent quantification of the resultant single-strand breaks by various means can also be used to measure UV photoproducts. Sites of CPDs can be cleaved with UV endonuclease V from T4 phage, an enzyme that specifically recognizes this type of damage. UvrABC exinuclease, a partial excision repair complex purified from *E. coli*, and an ultraviolet damage endonuclease (UVDE) purified from yeast, cleave DNA at or a few bases to either side of CPDs or (6-4)PDs in DNA. Both broad-spectrum enzymes have been used to measure overall levels of DNA damage in UV-irradiated DNA and nondimer damage remaining after the removal of CPDs or (6-4)PDs by enzymatic photoreactivation. Other enzymes involved in base excision repair (BER) have been used to cleave DNA at sites of nondimer photodamage, including endonuclease III (nth protein) and formamidopyrimidine glycosylases (fpg protein) purified from *E. coli* as well as a human endonuclease that specifically cleaves 8-oxodeoxyguanosine residues in DNA (hOGG). Such enzymes have been used to cleave DNA at cytosine and thymine photohydrates, fragmented

Tab. 2 Analytical techniques for quantifying UVR DNA damage and repair.

Chromatography
Paper/thin-layer 2-D chromatography
High-performance liquid chromatography (HPLC)
Gas chromatography/mass spectrometry (GC/MS)
Strand break/Molecular weight determinations
Supercoil relaxation assays (*in vitro*)
Alkaline gradient sedimentation
Denaturing gel electrophoresis
Microcapillary electrophoresis
Gene specific (Southern blot analysis)
Maxam-Gilbert sequencing
PCR-based assays:
Quantitative PCR
Ligation-mediated PCR
Immunoassays:
Immunoprecipitation
Radioimmunoassay
Enzyme-linked immunosorbent assay
Immunoslot blot/immunodot blot
Immunoelectronmicroscopy
Single-cell assays
Comet assay
Immunohistochemistry/immunocytochemistry
Excision repair assays:
Incision assays
DNA repair synthesis
Excision kinetics (endonuclease/antibody)
Transient expression assays

bases and sugars, photo-oxidation products, as well as unidentified photoproducts occurring at sites of (6-4)PDs and modified purines.

Several analytical techniques have been developed around the ability to specifically break DNA at sites of photoproducts. A single-strand nick at the site of a lesion in a supercoiled molecule converts that molecule from its compact supercoiled form to a relaxed form that can be easily resolved and quantified using gel electrophoresis. Other techniques required separation of the DNA strands and molecular weight determinations of the resultant single-strand fragments. Alkaline sucrose gradient sedimentation, denaturing gel electrophoresis, and microcapillary electrophoresis are approaches used to quantify endonuclease digestion at sites of UV damage. As with the chromatographic techniques, analyses of photodamage as strand breaks require radioactive or fluorimetric labeling to visualize and quantify DNA. A variation on this approach has combined alkaline gel electrophoresis with Southern blotting to detect and quantify strand-break frequencies in specific DNA sequences (genes) and has led to the discovery and description of transcription-coupled repair (TCR) (see below). Gene-specific repair

is also possible using quantitative PCR in which the ability of DNA photoproducts or strand breaks produced at sites of DNA photoproducts to block the progression of Taq polymerase is exploited. A further refinement of this technique combined with PCR technology has led to the development of ligation-mediated PCR (LMPCR) and the analysis of DNA damage at base resolution in cellular DNA. Previously, the sequence-specificity of photoproduct induction was limited to *in vitro* analyses using Maxam–Gilbert sequencing. Although all of these techniques are more sensitive than chromatography, their sensitivity depends on the maintenance of high-molecular weight DNAs throughout the purification and analysis.

Over the past 20 years, antisera raised against UV-irradiated DNA have proven to be powerful and versatile reagents for the study of various photochemical and photobiological phenomena. Polyclonal and monoclonal antibodies have been produced, which bind a variety of photodamage, including the CPD, (6-4)PD, Dewar pyrimidinone, and thymine glycol. A variety of immunological assays have been developed and have provided unique insights into photochemical and photobiological problems. Prominent within the extensive arsenal of immunological approaches adapted to the analysis of DNA damage and repair is immunoprecipitation of radiolabeled DNA fragments, enzyme-linked immunosorbent assays (ELISA), radioimmunoassays (RIA), immunoslot blots, and immunodot blots. The first and last of these techniques measure direct binding of antibody to damaged DNA; ELISA and RIA are competitive binding assays with considerably greater sensitivity and flexibility. A technique has even been developed in which CPDs were visualized in DNA using electron microscopy. Each technique has its own unique attributes and applications. Unlike chromatographic techniques, immunological assays do not require chemical or enzymatic degradation of DNA prior to analysis; unlike endonucleolytic assays, their sensitivities do not depend on the molecular weight (or purity) of sample DNA.

Many of the approaches discussed above can be adapted to the analysis of UV photoproducts in single cells, thereby greatly increasing the sensitivity of the assay. Strand breaks associated with DNA damage have been visualized in single cells using the "Comet assay" in which cells are embedded in agarose and the DNA extruded from the nuclei in an electric field. The resultant image, visualized with fluorescent dyes, resembles a comet with the head of the comet representing the nucleus and tail, the extruded DNA. The length and quantity of DNA in the tail is used to quantify the number of strand breaks and, hence, the amount of DNA damage. Antibodies to UV photoproducts have also been applied to CPD and (6-4)PD analyses in individual mammalian and bacterial cells. Secondary antibodies tagged with a fluorescent dye (e.g. fluorescein) bind to the primary damage-specific antibodies and allow visualization and quantification of the signal using fluorescence microscopy and image analysis software.

4.2 Quantifying DNA Repair

Removal of individual photoproducts can be assayed directly using the techniques described above. Techniques have also been developed to monitor various stages of the excision repair process. Transient incisions associated with the initial step of excision repair (Figs. 7 and 8) can be measured using alkaline elution. Alkaline

lysis of cells on filters results in denaturation of the DNA, which then elutes at a rate dependent on its molecular weight. Smaller DNA fragments containing more incisions elute faster than larger fragments; by measuring the molecular weight of the eluting fragments, the rate of repair incision can be estimated. Incorporation of base analogs (e.g. 1-ß-D-arabinofuranosylcytosine), which result in premature termination of repair synthesis, cause strand break accumulation. To discriminate between strand breaks associated with DNA repair and DNA replication, semiconservative DNA synthesis can be inhibited by specific inhibitors (e.g. hydroxyurea) or nonreplicating senescent (G_0) or terminally differentiated nonproliferating cell populations can be used. Studies with normally UV-resistant mammalian cells show a rapid accumulation of strand breaks within the first few (0–15) minutes after UV-irradiation followed by gradual restoration to the original molecular weight within an hour. The Comet assay has also been used to estimate the extent of incision associated with DNA repair.

As with incision assays, protocols for measuring the resynthesis stage of excision repair must utilize cells in which normal DNA replication has been inhibited. Unscheduled DNA synthesis (UDS) provided the first direct evidence for the DNA excision repair process in mammalian cells. Cells were grown for a short period (pulsed) in [^{3}H]-thymidine and then examined autoradiographically (as silver grains on a photographic plate). Cells undergoing DNA synthesis displayed intense labeling of their nuclei, whereas nonreplicating cells undergoing repair (unscheduled) synthesis showed much reduced labeling. Repair synthesis can also be determined using buoyant density centrifugation of DNA. In this technique, the base analog 5-bromodeoxyuridine (BUdR) is incorporated into DNA in place of thymidine. Regions containing this base have a higher density than the surrounding DNA. Hence, DNA that has replicated semiconservatively (i.e. produced one entire new strand) migrates as a hybrid density band on cesium chloride gradients. Repair synthesis occurring in very small patches ($\approx$30 bases) migrates to a position of normal density DNA. The amount of radioactivity [(^{3}H)-labeled BUdR] incorporated into unreplicated DNA is a measure of the total amount of repair synthesis.

The DNA repair capacity of a cell can also be determined using transient expression assays. Several of these assays have been developed, each varying only in the "reporter" gene utilized. In this assay, a plasmid containing a gene that can be expressed in a specific host cell is UV-irradiated and transfected into a host cell using one of a variety of transfection protocols. The ability of this gene to recover expression of a certain product (e.g. chloramphenicol acetyl transferase, β-galactosidase, luciferase) reflects the ability of the cell to repair the DNA damage induced in that sequence. Transient expression assays thus measure gene-specific repair.

5 DNA Damage Tolerance Mechanisms

The amount of photodamage present in cellular DNA at any given time depends not only on the extent of UV light absorbed by the DNA but also on the ability of the organism to repair the damage. UV damage tolerance mechanisms vary and may include combinations of photoprotection and DNA repair. The DNA repair strategy

utilized depends on the type of photoproduct encountered and the organism at risk. DNA repair is composed, in whole or part, of at least two well-studied mechanisms, photoenzymatic repair and nucleotide or base excision repair (Figs. 6–8).

5.1 Photoprotection

Nearly all organisms have behaviors or natural features that lower exposure of DNA to solar UVR and reduce the amount of photodamage. Human behavior includes wearing clothes, hats, sunglasses, and sunscreens, and general occupational and recreational life-styles. Many plants and animals have similar strategies. Biological components such as bark, cuticles, skin, fur, feathers, scales, and shells can strongly attenuate or eliminate the absorption of UVR by internal areas of cells or organisms. Habitat selection is also a significant factor. Shielding by terrain, foliage, water, or another organism can reduce UVR exposure. Actively seeking shade to avoid direct sunlight is also a common biological response.

Morphology may also affect the amount of radiation absorbed by cellular DNA. Small cells and organisms are more susceptible to the physical damage caused by ionizing radiation. Recent studies on marine plankton indicate that UVR damage may likewise be mediated by morphological factors. The amount of DNA damage induced in phytoplankton and bacterioplankton by UVB light correlates with the size of the individual species. In smaller cells, the distance between the cell surface and nuclear DNA is less than in larger cells. The shortened light path reduces refraction and absorption by cytoplasmic components and increases the amount of UVR reaching the nucleus. Similar principles can be applied to multicellular organisms. In humans, for example, the ratio of exposed surface area to internal tissues and organs is reduced when compared to smaller animals.

Several mechanisms for biochemical protection from UVR occur in nature. Pigmentation in the outer layers of plants and animals modulates the intensity and wavelength distribution of light reaching internal components. In some cases, light exposure induces pigment production. For example, the UVR absorbing compounds melanin and anthocyanin are produced in human skin and plants respectively, after exposure to solar UV. Melanin has multiple functions that can be either harmful or beneficial to the organism. On the one hand, it can be photoprotective, acting as a natural sunscreen, although not a very good one, and as a scavenger of active chemical species produced by solar UVR, although again not a very effective one. On the other hand, melanin can itself produce active radical species that can damage DNA and has the capacity to bind to drugs in ways that can either benefit or harm the cell. The beneficial role of melanin in humans is best illustrated by the observed reduction in skin cancers such as basal and squamous cell carcinomas and protection against sunburn in dark-skinned peoples. In black skin, melanosomes, the cellular compartments in which melanin is sequestered, are oval, single, and densely packed with melanin and clustered around the nucleus. UVR is blocked from the nucleus and free radicals are absorbed. In contrast, in white skin, melanosomes are round, aggregated, and lightly melanotic. UVR can bypass the clusters to enter the nucleus.

In addition to pigments, colorless UV-absorbing compounds have also been identified as possible UV-protective chemicals.

Concentrations of flavenoids in terrestrial plants, mycosporine amino acids (MAAs) in fungi, and mycosporine-like compounds in marine organisms all correlate with UV exposure history. Over 25 different mycosporine-like compounds have been identified in marine algae and invertebrates. These molecules display different UVA and UVB absorption profiles, with several compounds present in a single species. Combinations of these ''natural sunscreens'' provide broadband UV coverage and optimize protection from solar UVR. Because of these attributes, these compounds are being investigated for human applications.

5.2 Photoenzymatic Repair

Photoenzymatic repair is a light-dependent process involving the enzyme-catalyzed reversal of CPDs (Fig. 6). This was the first DNA repair mechanism recognized, discovered simultaneously by Albert Kelner and Renato Dulbecco in 1949. PER involves binding of a low molecular weight protein to a CPD followed by absorption of photons within the UVA/visible range of light, reversal of the damage, and, finally, release of the enzyme. PER appears to be a ubiquitous repair mechanism widespread throughout the plant and animal kingdoms (Table 3).

Although it is a relatively simple enzymatic process, PER displays much diversity in various biological systems. There are numerous organisms in which it has not been found, including several bacteria, such as *Haemophilus influenzae*, *Diplococcus pneumoniae*, *Bacillus subtilis*, and *Micrococcus radiodurans*, some algae and higher plants, such as *Chaetoceros convolutus*, *Thalassiosira australis*, *Phaseolus aureus* and *Haplopappus gracilis*, the nematode, *Caenorhabditis elegans*, and in placental, but not marsupial, mammals. Differences in action spectra, constitutive levels of photolyase, and the presence of required cofactors may account for the diverse PER efficiencies among organisms. The wavelength dependence for maximum PER efficiency is ≈380 nm in *E. coli* and *Euglena*, ≈400 nm in *Neurospora*, and ≈440 nm in *S. griseus*. The lack of PER in many organisms and its evolutionary conservation in systems never exposed to sunlight (e.g. soil or enteric bacteria, Blind cave fish)

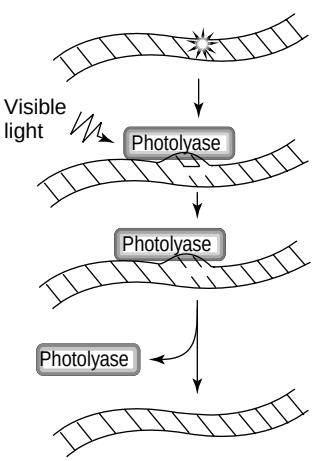

Fig. 6 Enzymatic photoreactivation. Generalized enzymatic pathway for the repair of UVR damage in DNA by photolyase. Required cofactors are not shown.

Tab. 3 Plants and animals capable of removing cyclobutane dimers by enzymatic photoreactivation (PER).

Group	***Species***	***Common name***
Cyanobacteria	*Plectonema boreanum*	Blue-green alga
	Anacystis nidulans	Blue-green alga
Schizomycophyta	*Escherichia coli*	Colon bacteria
	Streptomyces griseus	Soil actinomycete
Eumycophyta	*Saccharomyces cerevisiae*	Baker's yeast
	Neurospora crassa	Bread mold
Euglenophyta	*Euglena gracilis*	[a]
Angiospermae	*Phaseolus vulgaris*	Pinto bean
	Phaseolus lunatus	Lima bean
	Arabidopsis thaliana	[a]
	Zea mays	Maize
	Nicotiana tabacum	Tobacco
Gymnospermae	*Gingko biloba*	Gingko tree
Protozoa	*Paramecium Aurelia*	Paramecium
	Tetrahymena pyriformis	[a]
Mollusca	*Physa sp.*	Pond snail
Echinodermata	*Arabacia punctulata*	Sea urchin
	Echinarachnius parma	Sand dollar
Arthropoda	*Angasta Kiihniella*	Flower moth
	Gecarcinus lateralis	Land crab
	Artemia salina	Brine shrimp
	Homarus americanus	Lobster
Insecta	*Trichoplusia ni*	Cabbage looper
	Drosophila melanogaster	Fruit fly
Teleosts	*Haemulon sciuris*	Blue-striped grunt
	Pimephales promelas	Fathead minnow
	Anoptichthys jordani	Blind cave fish
	Tautoga onitis	Tautog
	Tautogolabrus adsperus	Cunner
	Xiphopherus variatus	Variable platyfish
	Carassius auratus	Goldfish
Amphibia	*Bufo marinus*	Cowflop toad
	Xenopus laevis	African Clawed toad
	Rana pipiens	Leopard frog
Reptilia	*Terrapene Carolina*	Box turtle
	Iguana iguana	Iguana lizard
	Gekko gekko	Gecko lizard
Aves	*Gallus gallus*	Domestic chicken
Marsupialia	*Didelphis marsupialis*	American opossum
	Caluromys derbianus	Woolly opossum
	Potorous tridactylis	Rat kangaroo
Mammalia	(*Homo sapiens?*)	Human

[a] No common name.

suggests photolyase may function in other molecular processes, such as excision repair. In support of this idea, it has been shown that *E. coli* photolyase stimulates excision repair *in vitro* and a DNA-binding protein associated with excision repair in human cells (XPE protein) has been identified with properties similar to a yeast photolyase. It has been shown that PER is inducible in fish cells, both in culture and *in vivo* (in caudal fins). Preillumination with visible light 8 h prior to damage induction significantly predisposes cells to CPD removal by PER. These studies provided the first clear demonstration of an inducible repair system in eukaryotes

Photoenzymatic repair of (6-4)PDs has also been observed in many organisms including insects (*Bombyx mori* and *D. melanogaster*), reptiles (*Crotalus atrox*), amphibia (*X. laevis*), and fish (*Xiphophorus* spp.). The (6-4)PD but not its Dewar isomer is the substrate for these photolyases and the enzyme has an average action spectrum peak at 400 nm. The efficiency of repair per incident photon is very low compared with CPD photolyases, and it appears that the (6-4)PD photolyase converts the photoproduct to unmodified bases, probably through an oxetane intermediate.

5.3 Excision Repair

DNA excision repair is a genetically complex and phenotypically diverse system; its competency is dependent on the species and developmental state of the individual as well as on the type of photoproduct encountered and its genomic distribution. The excision repair pathway is thought to proceed by the following fundamental steps: (1) recognition of the lesion as a structural distortion in the DNA helix; (2) unwinding or other activity to disassociate the lesion from chromatin proteins and provide accessibility to repair enzymes; (3) incision of the DNA backbone at or near the site of the lesion; excision and resynthesis of the DNA around the damaged site; and (4) ligation of the single-strand nick remaining after disengagement of the DNA polymerase complex. This process is very versatile, correcting various classes of chemical damage in addition to those induced by UVR.

5.3.1 Excision Repair Pathways

Excision repair may proceed by one of two routes (Figs. 7 and 8). Base excision repair (BER) is initiated by enzymatic recognition of the lesion and scission of the bond between the damaged base and its associated deoxyribose sugar, a process called aglycosylation. BER-initiating enzymes fall into two distinct categories: the hydrolytic dRPases and AP lyases. Examples of AP lyases include (1) endonuclease III which repairs several types of base damage including pyrimidine photohydrates, (2) pyrimidine dimer glycosylases from T4 phage-infected *E. coli* and *M. luteus*, which have the unique capability to recognize the 5′ pyrimidine of a CPD as a modified base, and (3) FaPy-DNA glycosylase from *E. coli*, which recognizes and cleaves various types of radiation damage including 8-oxodeoxyguanosine. After removal of the base damage, a 3′ AP lyase or 5′ AP endonuclease cleaves the phosphodiester bond adjacent to the abasic site. Enzymes that release 5′-deoxyribose-5-phosphate (dRP) residues from preincised AP-DNA have been collectively termed DNA-deoxyribophosphodiesterases (dRPases). To prepare the strand for resynthesis, the 3′ diesterase activity associated with AP lyase or the 3′–5′ exonuclease activity associated with dRPase and pol β digests

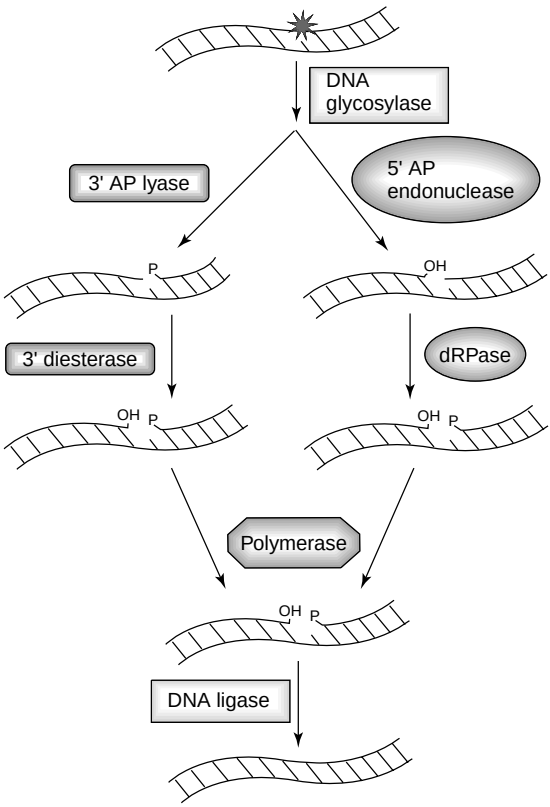

Fig. 7 Base excision repair. Pathway for generalized BER is shown including DNA damage cleavage by glycosylases, damage removal by endonuclease digestion, and gap-filling by DNA polymerase and DNA ligase.

the remaining abasic site. The damaged strand is removed and resynthesized in *E. coli* by pol I and in mammalian cells by pol β and the XRCC 1 gene product. The remaining nick is repaired by DNA ligase returning the DNA duplex to its original state. In addition to the short-patch repair described, BER may also proceed by a long-patch repair mechanism in which 2–13 new bases are inserted. The primary difference between short- and long-patch repair is the utilization of different polymerases and flap endonuclease (Fen1) for excision and resynthesis.

The nucleotide excision repair (NER) pathway recognizes a broad spectrum of UVR-induced photoproducts as well as other "bulky adducts" that may be induced by genotoxic agents. This process has been well studied in *E. coli* and is thought to be the primary excision repair pathway in eukaryotes as well. As with BER, enzymes that recognize and bind the helical distortion created at the damaged site initiate the NER process. In prokaryotes, a fairly simple multiprotein complex consisting of the Uvr A, B, C, and D scans the DNA for helical distortions, assembles at the lesion site, cleaves the DNA on both sides of the lesion leaving a gap, and is displaced to repeat the process. A DNA polymerase fills the gap and the strand is ligated to restore the DNA duplex to its original integrity.

In eukaryotes, NER is the most important repair process concerned with UVR

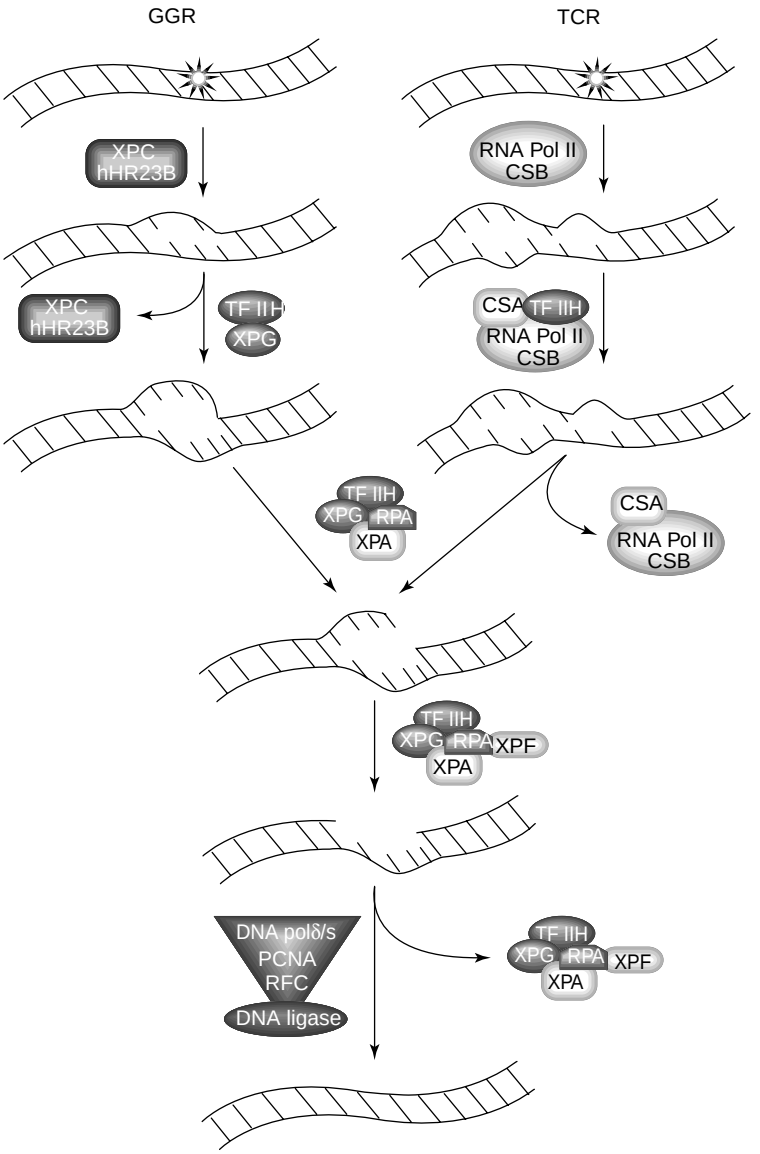

Fig. 8 Nucleotide excision repair. Generalized pathways for global genome (GGR) and transcription-coupled repair (TCR) in eukaryotes. Enzymes, DNA damage–binding proteins, and complexes are shown for DNA-damage recognition, stabilization, incision, excision, and resynthesis.

damage. Two major pathways of NER are known: transcription-coupled repair (TCR) and global genome repair (GGR) (Fig. 8). These NER pathways remove CPDs and (6-4)PDs in DNA and replace the damaged site with a newly synthesized polynucleotide patch approximately 29 bases in length. TCR removes damage more rapidly

from the transcribed strands of transcriptionally active genes, whereas GGR acts more slowly on nontranscribed regions and is regulated by p53 through control of XPE p48 expression. An increased excision rate in active genes (TCR) may also occur for (6-4)PDs, but this is less easily resolved against the greater overall rate of excision of these photoproducts in the genome as a whole. The differences in excision from active versus inactive genes occurs because a basal transcription factor, TFIIH, regulates basal transcription by RNA polymerase II and plays a major role in repair (Fig. 8). A detailed study of the promoter and first exons of the *PGKI* gene has indicated that excision is slow in regions of promoter binding but increases immediately after the transcription start site.

Many of the genes that regulate TCR are associated with the human disorders Xeroderma pigmentosum (XP), Cockayne syndrome (CS), trichothiodystrophy (TTD), and other disorders (see Sect. 5.3.3). Two of the helicases in TFIIH correspond to the *XPB* and *XPD* genes, and others are known to play a role from their analogs in the yeast transcription factor B. TCR also involves the CS genes *CSA* & *CSB, XPG*, the mismatch repair gene *MLH2*, and the breast cancer susceptibility gene *BRCA1*.

In Fig. 9, excision repair of damaged sites in mammalian chromatin is shown. The first step involves mechanisms that recognize damage and dissociate nucleoproteins to make the DNA accessible to repair enzymes. After recognition and dissociation, the excision repair pathway proceeds as in Fig. 8, followed by reassembly and repackaging of the chromatin complex. The actual number of damaged sites involved in excision repair at any one time is very small, comprising only about 1% of the lesions produced in DNA irradiated with a low UV fluence. Excision is, therefore, rate-limited by the enzymes involved in the early stages of DNA repair which dissociate the DNA from the nucleosome scaffolding and provide accessibility to excision enzymes.

5.3.2 Temporal and Spatial Distribution of Excision Repair

The efficiency of the excision repair process is dependent on the organism, the type of damage encountered, and where it occurs in the genome. In Fig. 10, repair curves for CPD and (6–4)PD excision are shown for several organisms representing the plant and animal kingdoms. The initial rates and extent of CPD repair (Fig. 10(a)) show great diversity, ranging from very rapid removal in primitive eukaryotes, such as bread mold and yeast, to much slower repair in vertebrates such as fish, frogs, and mammals. With few exceptions, the rate of (6-4)PD excision is much faster than the CPD (Fig. 10(b)). Notable exceptions include the common bread mold (*N. crassa*), the nematode *C. elegans*, and the epidermis of the marsupial *Monodelphis domestica*. Rapid repair of the (6-4)PD may be a consequence of its structure and its nonrandom distribution in chromatin. The 90° orientation of the 3′ base of the (6-4)PD (Fig. 4) produces a segment of DNA with considerable distortion in the phosphodiester backbone, local denaturation, and disruption of normal base pairing. This structure is bound with 10-fold greater affinity by the UvrABC exinuclease of *E. coli* and blocks the progression of DNA polymerase about 20-fold more efficiently than the CPD.

Rapid repair of the (6-4)PD may also belie the structural and functional role of chromatin in DNA repair (Fig. 9). Insight into the relationship between DNA repair and chromatin structure has come from studies using a yeast minichromosome

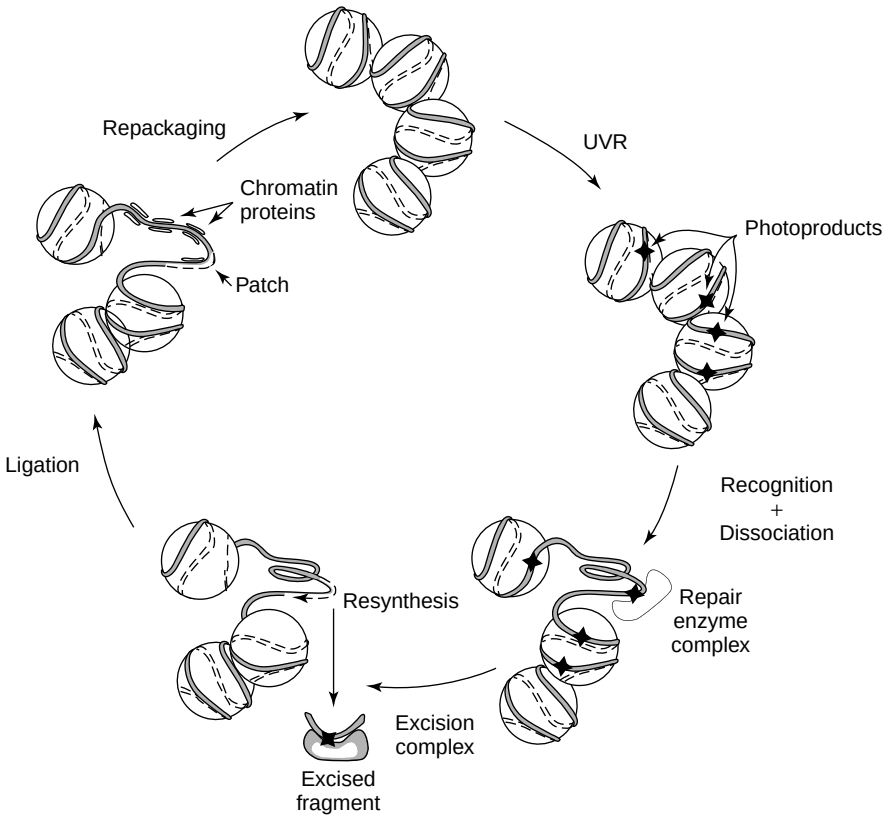

Fig. 9 Excision repair in mammalian chromatin. Repair is initiated by photoproduct recognition and dissociation of damaged DNA from nucleoproteins to make the DNA more accessible to repair enzymes. This is followed by the sequential incision by an endonuclease, excision, and resynthesis by a DNA polymerase, and sealing of the patch by a polynucleotide ligase. The repaired segment is finally reassembled and repackaged in the original nucleosomal structure. (Redrawn from Cleaver and Mitchell, 1993).

as a model system. Regions of this chromosome containing unstable nucleosomes are more readily repaired than regions in which the nucleoprotein structure is more constant. These data imply that "open" chromatin configurations are more accessible and more repairable than compact, heterochromatic regions of the genome. As mentioned, (6-4)PDs are induced with sixfold greater frequency in linkers compared to nucleosomal core DNA. The nonrandom induction of the (6-4)PDs in more accessible regions of chromatin may, therefore, facilitate its rapid removal by the excision repair complex.

5.3.3 Developmental and Genetic Regulation of Excision Repair

The efficiency of DNA repair depends on the developmental state of a cell, tissue or organism. Both developmental and genetic attributes have made the nematode *C. elegans* a popular organism for probing many biological problems, including DNA repair. A striking phenotype of certain

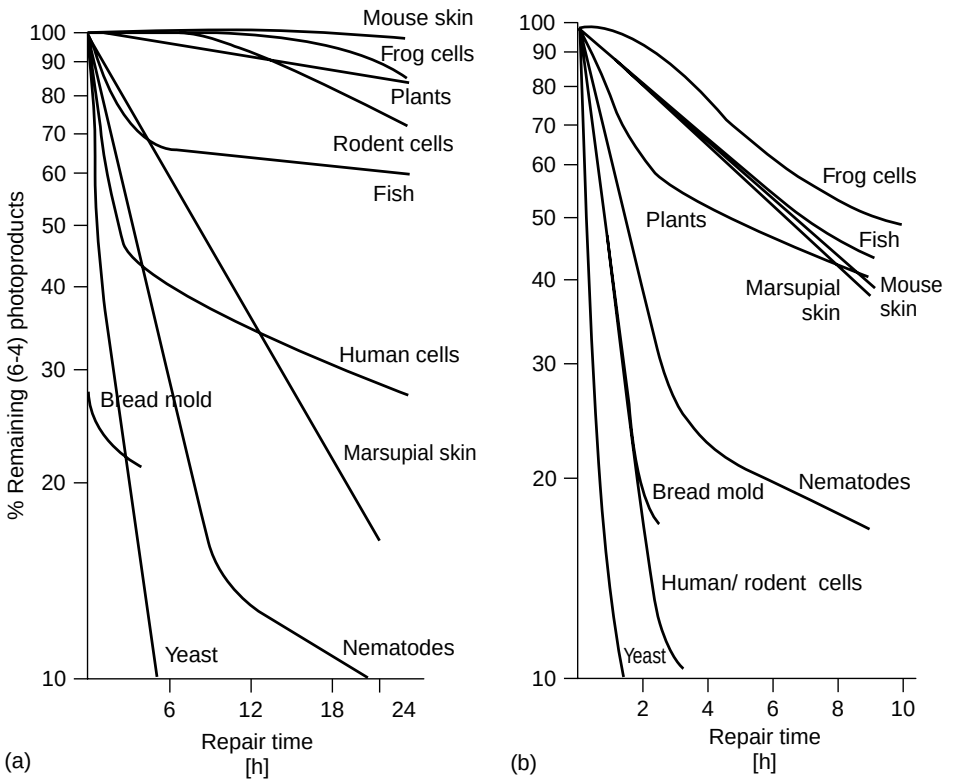

Fig. 10 Kinetics of CPD and (6-4)PD repair, radioimmunoassay of DNA repair of CPDs in various organisms. Organisms include *N. crassa* (bread mold), *S. cerevisiae* (yeast), *C. elegans* (nematode), *M. domestica* (marsupial), *Xiphophorus variatus* (fish), *A. thaliana* (plant), *X. laevis* (frog).

C. elegans mutants is the manifestation of stage-specific variation in UV hypersensitivity. For example, *rad-3* mutant nematodes are moderately hypersensitive when irradiated as embryos, but larvae are highly hypersensitive. Unlike wild-type nematodes, whose repair capacity remains relatively constant, the excision repair capacity of *rad-3* nematodes varies substantially throughout development. *Rad-3* embryos display a reduced capacity for excision repair, yet larvae assayed 24 h later are almost totally deficient in excision repair. The correlation between UV hypersensitivity and reduced excision repair capacity suggests that DNA repair in the nematode is developmentally regulated.

Evidence also suggests that DNA repair is developmentally regulated in higher eukaryotes. Excision repair of UV damage in cultured fish cells is very low, although significant excision repair is observed in primary embryonic cells. Similar regulation is seen in mice; exposure to photoreactivating light has no effect on CPD removal in adult mouse epidermal DNA or mouse fibroblasts in culture, yet significantly reduces the number of these lesions in the epidermis of neonatal mice; and excision (dark) repair of CPDs is much

more efficient in early embryos compared to adult mice.

DNA damage resulting from UVR is more efficiently repaired in proliferating basal cells in rodent and human epidermis than in terminally differentiated keratinocytes. A compelling generality one can derive from the developmental pattern of DNA repair in organisms and tissues is that more rapidly proliferating systems (i.e. stem cells, embryos, fetuses, larvae) have increased capacities for self-correction. Hence, those organisms most vulnerable to the lethal and mutagenic effects of genetic change are best able to tolerate it. DNA repair exhibits different developmental profiles in different organisms; thus, both ontogeny and phylogeny of DNA repair can vary substantially.

Selection of UV-sensitive rodent cells and the study of human sunlight-sensitive disorders have identified a large array of genetic loci that control the response of mammalian skin to UVR damage (Table 4). These loci are all characterized by significant increases in sensitivity to UVC or UVB radiation and include the excision repair cross-complementing (ERCC) series in rodent cells, XP, CS, TTD, and basal cell nevus syndrome (BCNS).

With the exception of BCNS, all of these disorders represent increased sensitivity to UVB and UVC wavelengths due to recessive mutations. BCNS exhibits a unique sensitivity to UVB radiation and is a dominant disorder. The recessive disorders can be considered as subsets within a large family of genes that regulate human cell DNA repair. These subsets are not mutually exclusive; CS complementation group C is identical to XP complementation group B; TTD overlaps with group D; XP group B corresponds to ERCC3 as well as to CS group B; XP group D corresponds to ERCC2.

XP is a rare autosomal recessive disease that occurs at a frequency of about 1 : 250 000 in the United States. Affected

Tab. 4 Complementation groups in XP and UV-sensitive Chinese hamster ovary (CHO) cells.

Group	*Human chromosome location*	*Central nervous system repair disorders*	*Relative DNA repair [%]*
Xeroderma pigmentosum			
A	9q34.1	Yes	2–5
B	2q21	Yes	3–7
C	3q25	No	5–20
D	19q13.2	Yes	25–50
E	11p11-12	No	50
F	16q13.1	No	18
G	13q32.3	Yes	<2
Variant	6p21	No	100
CHO (*ERCC*)			
1	19q13.2	–	Low
2 (XPD)	19q13.2	–	Intermediate
3 (XPB)	2q21	–	Intermediate
4 (XPF)	16q13.1	–	Low
5 (XPG)	13q 32.3	–	Intermediate

patients (homozygotes) have sun sensitivity resulting in progressive degenerative changes of sun-exposed portions of the skin and eyes, often leading to neoplasia. Some XP patients have, in addition, progressive neurological degeneration. Obligate heterozygotes (parents) are generally asymptomatic. The median age of onset is 1 to 2 years, with skin rapidly taking on the appearance of that seen in individuals with many years of sun exposure. Pigmentation is patchy, and the skin shows atrophy and telangiectasia with development of basal cell carcinomas (BCC) and squamous cell carcinomas (SCC). The frequency of cancers is about two thousand times that seen in the general population less than 20 years of age, with an approximate 30-year reduction in life span.

Cells from patients with XP excise CPDs and (6-4)PDs at reduced rates of 0 to 90% of normal, except for the variant group, which has a near-normal rate. The reductions are similar in all tissues thus far investigated, including skin *in vivo*, peripheral lymphocytes, fibroblasts, liver cell cultures, and tumor cells. There are seven complementation groups among patients who are deficient in excision repair, and an eighth, the XP variant, has a defect in replication of damaged DNA. Compared with normal cells, cells from XP groups A and D are very sensitive to the lethal effects of UV light and are unable to repair the CPD and the (6-4)PD. XP group A cells also have a reduced capacity to repair the Dewar pyrimidinone.

Group C is one of the largest groups and is often referred to as the "classic" form of XP. The patients show only skin disorders, which vary considerably in severity, depending on the climate. Tumors of the tongue have been observed in several patients. Cells have low but heterogeneous levels of excision repair (10–20% of normal) and are less sensitive to killing by UVR and chemical carcinogens than cells in groups A and D. This group is unique among the XP groups in that the repair defect is restricted to GGR with TCR functioning normally (Fig. 8). This raises the dilemma that high rates of cell killing, somatic mutation, and cancer from UV light in XP group C are associated with repair deficiencies in the nontranscribed regions of the genome. This in turn suggests that activating rather than silencing mutations may be important or that mutations may, in fact, arise from unrepaired lesions in the nontranscribed strand of active genes.

Group E is a rare group that exhibits mild symptoms and residual levels of repair that are between 50 and 100% of normal and lack a DNA-binding protein, p48, that is associated with a larger protein, p127. XPE patients carry mutations in the p48 gene. The role of this protein is unclear but it is dependent on p53 and involved in repair of nontranscribed regions of DNA. The absence of a homologous protein in rodent cells has been invoked to explain the reduced capacity of these cells in nontranscribed DNA.

XP variant cells have normal excision repair but lack a DNA low fidelity class-Y DNA polymerase, *hRAD30A* (Pol H) that is required for accurate bypass of CPDs during DNA replication. In its absence, XP variant cells are very susceptible to UV-induced mutagenesis and exhibit essentially the same symptoms as other XP patients. Carcinogenesis from UV damage in XP patients arises therefore from the loss of either NER capacity or Pol H; both lead to an increase in the amount of persistent DNA damage that becomes the substrate for error-generation (mutations, gene rearrangements, deletions etc.).

Cockayne syndrome is an autosomal recessive disease characterized by cachectic dwarfism, retinal abnormalities, microcephaly, deafness, neural defects, and retardation of growth and development after birth. Carcinomas of the skin as a result of hyper-photosensitivity are not seen in patients with CS, setting this disease apart from XP. CSB knockout mice however are sensitive to UVB carcinogenesis. Characteristic cellular changes in CS include a failure of DNA and RNA synthesis to recover to normal levels after UV-irradiation. The excision of DNA photoproducts from total genomic DNA of CS cells is normal, but repair of transcriptionally active genes is reduced. The CS gene products are involved in coupling excision repair to transcription, but their precise function is not yet clear. They may be involved in the ubiquitination and degradation of stalled RNA pol II at damaged sites.

Cockayne syndrome and XP group C therefore make an interesting contrast. CS cells repair only transcriptionally inactive genes, whereas XP group C cells repair only transcriptionally active genes. They show a similar increased sensitivity to cell killing, indicating that all regions of the genome must be repaired for normal survival. But only XP group C shows elevated mutagenesis and carcinogenesis, indicating that defective repair of transcriptionally inactive genes is more important for these endpoints. Mice with defects in XP or CS genes or both show elevated carcinogenesis, which raises questions of the different roles of NER in humans and mouse strains.

Trichothiodystrophy is a rare autosomal recessive disorder characterized by sulfur-deficient brittle hair and ichthyosis. Hair shafts split longitudinally into small fibers, and this brittleness is associated with levels of cysteine/cystine in hair proteins that are 15 to 50% of those in normal individuals. Patients also show physical and mental retardation of varying severity and often have an unusual facial appearance, with protruding ears and a receding chin. Mental abilities range from low normal to severe retardation. Several categories of the disease can be recognized on the basis of UV sensitivity and DNA repair. The most severe cases have repair deficiencies and complementation properties that place them in XP groups B and D. Most cases represent mutations in components of the transcription factor TFIIH and the symptoms of this disease indicate a role for this factor in development and hair growth.

6
Biological Consequences of DNA Repair

Air pollution has resulted in global decreases in stratospheric ozone concentrations and an increase in the amount of harmful solar radiation reaching the earth's surface. The effects of increased UVB radiation on the human population are complex: the obvious and direct consequences include increased skin cancer and accelerated aging; less obvious and more indirect effects include deterioration of world food crops and natural ecosystems. The role of UVR DNA damage as a mutagenic precursor of some types of skin cancer is well documented and such damage is undoubtedly involved in cell death and growth inhibition in environmental systems. Its importance in photoaging and immunosuppression is, however, less well understood, although data suggest that DNA damage may provide an important signaling event in these processes.

6.1 Photoaging

Photoaging of skin results from chronic exposure to sunlight and is distinct from chronological aging. Changes in the skin induced by chronic exposure to solar UVR can occur well before signs of chronic skin aging. Signs of photoaging occur exclusively on sites of the body that are exposed to sunlight such as the neck, face, forearms, and legs (in women) and are manifested as a leathery appearance, increased wrinkle formation, loss of elasticity, increased fragility, impaired wound healing, and the formation of lentigines (sunspots) and diffuse hyperpigmentation. Because of differential attenuation of UVR in the skin (i.e. longer wavelengths penetrate deeper), it is believed that the shorter solar wavelengths (UVB) are responsible for photoaging effects in the epidermis and longer wavelengths (UVA) for dermal effects. Hence, direct damage in DNA, such as CPDs and (6-4)PDs may play a role in epidermal photoaging and ROS generated by UVA may act through various signal transduction pathways that may or may not be mediated by DNA damage. Transcription factors such as AP-1 or NFκB are known to stimulate matrix metalloproteinases (MMPs) that degrade matrix proteins (e.g. collagen) within the dermis and result in loss of tissue integrity. A more likely connection between DNA damage and photoaging may involve mitochondrial DNA (mtDNA). Mitochondria are not able to repair CPDs or (6-4)PDs and recent evidence shows mutation frequencies in mtDNA at significantly greater frequencies than that observed in genomic DNA of sun-exposed skin cells. In addition to a putative role for UVB-induced damage, a central role of UVA-induced ROS has been postulated in generating the most common deletion mutations found in these cells. Such deletions have been associated with loss of mitochondrial function and the induction of the metalloproteinase MMP-1.

6.2 Immunosuppression

The immunosuppressive properties of solar UVR have been well studied and have important implications with regard to the epidemiology and control of infectious diseases as well as the progression of various types of cancer. It has been known for some time that UVR inhibits the immunological rejection of transplanted tumors. Two major areas of UVR-related immunosuppression have been under study, including the localized suppression of contact hypersensitivity, a delayed type hypersensitivity, and systemic immunosuppression, which is induced by high UVB, doses affecting immune responses at distant, nonexposed sites of the body. The molecular mechanisms underlying these phenomena are complex and some considerable attention has been directed toward identifying the targets of UVR immunosuppression and the role of DNA damage. Two chromophores have been identified in the epidermises, which participate in photoimmunosuppression, including urocanic acid and DNA. Urocanic acid accumulates in the stratum corneum and is converted from a trans- into a cis-form upon exposure to UVR. Removal of the stratum corneum has been shown to inhibit immunosuppression. Direct damage caused by UVB is considered to play a major role in UV-mediated immunosuppression. Accelerated removal of CPDs by exogenous application of a liposome cream containing DNA repair enzyme (i.e. T4 endonuclease V) resulted in complete abrogation of the immunosuppressive effects

of UVB. Enhanced repair by this topical treatment also suppressed cytokine release of interleukin-10 (IL-10) and tumor necrosis factor (TNF) by keratinocytes, both of which are implicated in the immune response. Suppression of DNA repair by the cytokine IL-12 may also play a role in immunosuppressed individuals.

6.3 Photocarcinogenesis

Carcinogenesis often appears to proceed by a multistep process, the first being an initiation event with subsequent promotional events that can often occur much later. One view of carcinogenesis correlates initiation with the induction of a small number of somatic mutations in a few critical genes, and promotion with further alterations in gene expression, copy number, and karyotype. Carcinogenesis appears to involve a process that destabilizes the genome with resulting variations in the activity of a large number of genes that change and develop over time. These include genes for cell-cycle regulation, checkpoint and damage response, detoxifying carcinogenic chemicals, DNA repair, and some 50 or more dominantly acting proto-oncogenes activated by mutation, deletion, translocation, or amplification, and tumor suppressor genes whose loss may contribute to the development of cancer. Two molecular mechanisms can be important in the initiation of carcinogenesis: activation of proto-oncogenes and inactivation of tumor suppressor genes and both are vulnerable to the lethal and mutagenic effects of UVR.

Skin cancers include at least three categories of tumor type: squamous cell carcinoma, basal cell carcinoma, and melanoma. These have distinct origins, molecular changes and etiology, and eventual outcomes, and each has characteristic genetic and solar components to their development. SCCs and BCCs appear to originate from different locations in the skin, and melanoma from the pigment cells. The initial damage produced by solar UVB to the skin is eliminated either by repair or by proliferation and exfoliation from the skin's surface. Some cells in the skin also die by apoptosis following exposure ("sunburn cells") resulting in the elimination of damaged cells. Stem cells for the epithelium are thought to reside in the bulge region of the hair follicles, but there are also secondary stem cells at the base of each column of epidermal transit amplifying cells in the epidermal proliferative units. Whether SCCs and BCCs come from different stem cells is currently unknown. A very low frequency of quiescent cells has been observed to retain DNA damage for long periods seemingly without repair or proliferation. The carcinogen-retaining cells may be stem cells, or damaged cells with the potential to become mutants once stimulated to proliferate.

The progression of molecular changes involved in SCC appears to be initiated by inactivating mutations in p53 that result in expanding clones in the sun-exposed areas of the skin that are initially confined within the proliferating units. These clones can be very frequent and can break out of the confines of the columnar structure of the proliferative units after chronic UVB irradiation. Over 90% of SCCs are reported to have p53 mutations that are characteristic of UV exposure, being in dipyrimidine sequences with notable frequencies of C–C to T–T changes. Several investigations have also identified activating mutations in the Ha-*ras* and N-*ras* oncogenes at codon 61, from solar UVR. However, although over 75% of UV-induced mutations are

generally C→T transitions at T–C or C–C CPD sites, the Ha-*ras* and N-*ras* activation occurred in tumors at a T–T site and are transversions not previously identified in model culture systems.

BCCs are exemplified by the hereditary disease, BCNS. The defect in BCNS appears to be in the human homolog of the *Drosophila* gene *PATCHED* (*PTC*), which is involved in embryonic patterning as well as in determining the fate of multiple structures in the developing embryo. Both somatic mutations in sporadic BCCs and single allele mutations in patients with BCNS, have provided strong evidence that this tumor suppressor is important in BCC tumorigenesis. The patched protein (PTC) is a transmembrane receptor, and with the coreceptor membrane protein Smoothened (SMO), regulates signal transduction by the extracellular protein Hedgehog (hH) that binds to PTC. Mutations in *SMO* have been identified in BCCs. Also, more than 50% of BCCs and 90% of SCCs contain mutations in p53 with a specific signature induced by ultraviolet light and many other gene amplifications and deletions have been detected.

Melanoma occurs in a hereditary form among approximately 5 to 12% of all patients. Although this is clinically and histologically indistinguishable from nonfamilial melanoma, there are differences in the age of diagnosis, lesion thickness, and frequency of multiple lesions. There are two known melanoma predisposition genes, CDKN2A and CDK4. CDKN2A functions as a cell-cycle regulatory protein that inhibits the activity of the cyclin D1 cyclin-dependent kinase 4 (CDK4) or 6 (CDK6) complex, which then act as negative regulators of growth by arresting cells at the G1 phase of the cell cycle.

6.4 Effects on the Environment

Global decreases in stratospheric ozone have been observed in recent years as a consequence of air pollution from natural or anthropogenic sources, such as volcanic eruptions and hydrocarbon (fuel) combustion. There is widespread concern about the effects of such atmospheric changes on the biosphere. Reduced ozone concentrations affect the intensity and spectral distribution of UVR at the earth's surface and up to at least 20 m deep in the oceans by increasing the amounts of biologically harmful wavelengths of UVB.

An increase in solar UV, as postulated from models of stratospheric ozone depletion, could significantly alter populations of marine and terrestrial plant and animal communities. Selection by UVR would result in reductions of UV-sensitive species and expansion of more UV-resistant species or clones. Trends of this sort would inevitably produce a more UV-resistant community; however, the consequences of changes in taxonomic structure on other parameters relevant to the ecosystem, such as primary production and trophic dynamics, are problematic. Decreases in total primary production have been documented during Antarctic ozone depletion and are expected at other latitudes.

The annual occurrence of springtime ozone depletion over the Antarctic provides a unique setting for monitoring the effects of increased UVB radiation on natural populations. Studies have been conducted on the induction and repair of UVB damage in various Antarctic bacterioplankton, phytoplankton, larval echinoderm, and fish species. The results indicate that great diversity in the photobiology of individual species exists with photoproduct

induction and repair varying considerably. Indeed, DNA-damage frequencies can vary a hundredfold and the ability of individual species to excise DNA photodamage ranges from negligible to near complete removal within 6 h. Parallel to this diverse molecular response, the ability of these organisms to survive the effects of UVB light is also extremely variable, correlating with DNA repair capacity. It is apparent that, within a cross-section of the Antarctic marine community, relatively sensitive and resistant phytoplankton species coexist.

Marine plankton are an important component of the oceanic food chain and are directly responsible for much of the world's oxygen production. Knowledge of the molecular and biological responses to UVB light at the species level will enable us to predict changes in the biomass and population diversity of marine ecosystems. Using plankton as a model, insight into the ultimate effects of stratospheric deozonation on other marine and terrestrial ecosystems can be extrapolated. Knowledge of the molecular biology of DNA damage and repair in selected plants and animals at different stages of their development will also allow us to identify species at risk and predict the consequences of deozonation on man and the earth's environment.

Bibliography

Books and Reviews

Björn, L.O., Moan, J., Young, A.R. (Eds.) (1993) *Environmental UV Photobiology*. Plenum Press, London.

Cleaver, J.E., Mitchell, D.L. (1993) Ultraviolet radiation carcinogenesis, in: Holland, J.F., Frei, E. III, Bast, R.C., Kufe, D.W., Morton, D.L., Weichselbaum, R.R., (Eds.) *Cancer Medicine*, 3rd edition, Lea and Febiger, Philadelphia, pp. 245–254.

Friedberg, E.C., Hanawalt, P.C. (Eds.) (1981–1988) *DNA Repair: A Laboratory Manual of Research Procedures*, Vols. 1–3. Marcel Dekker, Inc., N.Y.

Friedberg, E.C., Walker, G.C., Seide, W. (1995) *DNA Repair and Mutagenesis*. ASM Press, Washington, D.C.

Henderson, D.S. (Ed.) (1999) *DNA Repair Protocols, Eukaryotic Systems*. Humana Press, Totowa, N.J.

Morrison, H. (Ed.) (1990) *Bioorganic Photochemistry, Vol. 1: Photochemistry and the Nucleic Acids*. John Wiley and Sons, Inc. New York, N.Y.

Pfeifer, G.P. (Ed.) (1996) *Technologies for Detection of DNA Damage and Mutations*, Plenum Press, New York, N.Y.

Wang, S.Y. (Ed.) (1976) *Photochemistry and Photobiology of Nucleic Acids*, Vols. I and II. Academic Press, N.Y.

Primary Literature

Ananthaswamy, H.N., Pierceall, W.E. (1990) Molecular mechanisms of ultraviolet radiation carcinogenesis. *Photochem. Photobiol.* **52**, 1119–1136.

Beehler, B.C., Przybyszewski, J., Box, H.B., Kulesz-Martin, M.F. (1992) Formation of 8-hydroxydeoxyguanosine within DNA of mouse keratinocytes exposed in culture to UVB and H2O2. *Carcinogenesis* **13**, 2003–2007.

Berg, R.J., Rebel, H., van der Horst, G.T., van Kranen, H.J., Mullenders, L.H., van Vloten, W.A., de Gruijl, F.R. (2000) Impact of global genome repair versus transcription-coupled repair on ultraviolet carcinogenesis in hairless mice. *Cancer Res.* **60**, 2858–2863.

Bernstein, E.F., Brown, D.B., Urbach, F., Forbes, D., Del Monaco, M., Wu, M., Katchman, S.D., Uitto, J. (1995) Ultraviolet radiation activates the human elastin promoter in transgenic mice: a novel in vivo and in vitro model of cutaneous photoaging. *J. Invest. Dermatol.* **105**, 269–273.

Beukers, R., Berends, W. (1960) Isolation and identification of the irradiation product of thymine. *Biochim. Biophys. Acta* **41**, 550–551.

Birch-Machin, M.A., Tindall, M., Turner, R., Haldane, F., Rees, J.L. (1998) Mitochondrial

DNA deletions in human skin reflect photo- rather than chronologic aging [see comments]. *J. Invest. Dermatol.* **110**, 149–152.

Bohr, V.A., Smith, C.A., Okumoto, D.S., Hanawalt, P.C. (1985) DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. *Cell* **40**, 359–369.

Boorstein, R.J., Hilbert, T.P., Cadet, J., Cunningham, R.P., Teebor, G.W. (1989) UV-induced pyrimidine hydrates in DNA are repaired by bacterial and mammalian DNA glycosylase activities. *Biochemistry* **28**, 6164–6170.

Brash, D.E., Haseltine, W.A. (1982) UV-induced mutation hotspots occur at DNA damage hotspots. *Nature* **298**, 189–192.

Breter, H.J., Weinblum, D., Zahn, R.K. (1974) A highly sensitive method for the estimation of pyrimidine dimers in DNA by high-pressure liquid cation-exchange chromatography. *Anal. Biochem.* **61**, 362–366.

Brown, D.W., Libertini, L.J., Suquet, C., Small, E.W., Smerdon, M.J. (1993) Unfolding of nucleosome cores dramatically changes the distribution of ultraviolet photoproducts in DNA. *Biochemistry* **32**, 10527–10531.

Cadet, J., Voituriez, L., Hahn, B.S., Wang, S.Y. (1980) Separation of cyclobutyl dimers of thymine and thymidine by high-performance liquid chromatography and thin-layer chromatography. *J. Chromatogr.* **195**, 139–145.

Ceccoli, J., Rosales, N., Tsimis, J., Yarosh, D.B. (1989) Encapsulation of the UV-DNA repair enzyme T4 endonuclease V in liposomes and delivery to human cells. *J. Invest. Dermatol.* **93**, 190–194.

Chan, G.L., Doetsch, P.W., Haseltine, W.A. (1985) Cyclobutane pyrimidine dimers and (6-4) photoproducts block polymerization by DNA polymerase I. *Biochemistry* **24**, 5723–5728.

Chu, G., Chang, E. (1988) Xeroderma pigmentosum group E cells lack a nuclear factor that binds to damaged DNA. *Science* **242**, 564–567.

Cleaver, J.E. (1968) Defective repair replication of DNA in xeroderma pigmentosum. *Nature* **218**, 652–656.

Coohill, T.P., Jacobson, E.D. (1981) Action spectra in mammalian cells exposed to ultraviolet radiation. *Photochem. Photobiol.* **33**, 941–945.

Cook, P.R., Brazell, I.A. (1976) Detection and repair of single-strand breaks in nuclear DNA. *Nature* **263**, 679–682.

Demple, B., Linn, S. (1980) DNA N-glycosylases and UV repair. *Nature* **287**, 203–208.

Drobetsky, E.A., Grosovsky, A.J., Glickman, B.W. (1987) The specificity of UV-induced mutations at an endogenous locus in mammalian cells. *Proc. Natl. Acad. Sci. USA* **84**, 9103–9107.

Dulbecco, R. (1949) Reactivation of ultraviolet inactivated bacteriophage by visible light. *Nature* **613**, 949.

Edenberg, H.J. (1976) Inhibition of DNA replication by ultraviolet light. *Biophys. J.* **16**, 849–860.

Fears, T.R., Scotto, J. (1983) Estimating increases in skin cancer morbidity due to increases in ultraviolet radiation exposure. *Cancer Invest.* **1**, 119–126.

Fitzpatrick, T.B., Sober, A.J. (1985) Sunlight and skin cancer [editorial]. *N. Engl. J. Med.* **313**, 818–820.

Francis, A.A., Lee, W.H., Regan, J.D. (1981) The relationship of DNA excision repair of ultraviolet-induced lesions to the maximum life span of mammals. *Mech. Ageing Dev.* **16**, 181–189.

Franklin, W.A., Doetsch, P.W., Haseltine, W.A. (1985) Structural determination of the ultraviolet light-induced thymine-cytosine pyrimidine-pyrimidone (6-4) photoproduct. *Nucleic Acids Res.* **13**, 5317–5325.

Gale, J.M., Nissen, K.A., Smerdon, M.J. (1987) UV-induced formation of pyrimidine dimers in nucleosome core DNA is strongly modulated with a period of 10.3 bases [published erratum appears in Proc Natl Acad Sci USA 1989 Nov; 86(21):8386]. *Proc. Natl. Acad. Sci. USA* **84**, 6644–6648.

Glazer, P.M., Greggio, N.A., Metherall, J.E., Summers, W.C. (1989) UV-induced DNA-binding proteins in human cells. *Proc. Natl. Acad. Sci. USA* **86**, 1163–1167.

Glazer, P.M., Sarkar, S.N., Summers, W.C. (1986) Detection and analysis of UV-induced mutations in mammalian cell DNA using a lambda phage shuttle vector. *Proc. Natl. Acad. Sci. USA* **83**, 1041–1044.

Glickman, B.W., Schaaper, R.M., Haseltine, W.A., Dunn, R.L., Brash, D.E. (1986) The C-C (6-4) UV photoproduct is mutagenic in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **83**, 6945–6949.

Hanawalt, P.C. (1994) Transcription-coupled repair and human disease. *Science* **266**, 1957–1958.

Hariharan, P.V., Cerutti, P.A. (1977) Formation of products of the 5,6-dihydroxydihydrothymine type by ultraviolet light in HeLa cells. *Biochemistry* **16**, 2791–2795.

Hart, R.W., Setlow, R.B., Woodhead, A.D. (1977) Evidence that pyrimidine dimers in DNA can give rise to tumors. *Proc. Natl. Acad. Sci. USA* **74**, 5574–5578.

Hauswirth, W., Wang, S.Y. (1977) Excited state processes and solution conformation of dipyrimidine adducts. *Photochem. Photobiol.* **25**, 161–166.

Herndl, G.J., Muller-Niklas, G., Frick, J. (1993) Major role of ultraviolet-B in controlling bacterioplankton growth in the surface layer of the ocean. *Nature (London)* **361**, 717–719.

Horsfall, M.J., Lawrence, C.W. (1994) Accuracy of replication past the T-C (6-4) adduct. *J. Mol. Biol.* **235**, 465–471.

Jackson, D.A., Balajee, A.S., Mullenders, L., Cook, P.R. (1994) Sites in human nuclei where DNA damaged by ultraviolet light is repaired: visualization and localization relative to the nucleoskeleton. *J. Cell Sci.* **107**, 1745–1752.

Jagger, J. (1981) Near-UV radiation effects on microorganisms. *Photochem. Photobiol.* **34**, 761–768.

Jiang, N., Taylor, J.S. (1993) In vivo evidence that UV-induced C→T mutations at dipyrimidine sites could result from the replicative bypass of cis-syn cyclobutane dimers or their deamination products. *Biochemistry* **32**, 472–481.

Jonason, A.S., Kunala, S., Price, G.J., Restifo, R.J., Spinelli, H.M., Persing, J.A., Leffell, D.J., Tarone, R.E., Brash, D.E. (1996) Frequent clones of p53-mutated keratinocytes in normal human skin [see comments]. *Proc. Natl. Acad. Sci. USA* **93**, 14025–14029.

Kantor, G.J., Setlow, R.B. (1981) Rate and extent of DNA repair in nondividing human diploid fibroblasts. *Cancer Res.* **41**, 819–825.

Karentz, D., Cleaver, I.E., Mitchell, D.L. (1991) Ultraviolet photobiology of Antarctic Bacillariphyceae (Diatoms): Molecular and biological responses to UV-B radiation. *J. Phycol.* **27**, 326–341.

Kligman, L.H., Akin, F.J., Kligman, A.M. (1985) The contributions of UVA and UVB to connective tissue damage in hairless mice. *J. Invest. Dermatol.* **84**, 272–276.

Kobayashi, N., Muramatsu, T., Yamashina, Y., Shirai, T., Ohnishi, T., Mori, T. (1993) Melanin reduces ultraviolet-induced DNA damage formation and killing rate in cultured human melanoma cells. *J. Invest. Dermatol.* **101**, 685–689.

Kondo, S. (1975) DNA repair and evolutionary considerations. A search for a general principle in nuclear biology with use of radiation as a probe. *Adv. Biophys.* **7**, 91–162.

Kripke, M.L., Cox, P.A., Alas, L.G., Yarosh, D.B. (1992) Pyrimidine dimers in DNA initiate systemic immunosuppression in UV-irradiated mice. *Proc. Natl. Acad. Sci. USA* **89**, 7516–7520.

Lamola, A.A., Yamane, T. (1967) Sensitized photodimerization of thymine in DNA. *Proc. Natl. Acad. Sci. USA* **58**, 443–446.

Lang, H., Luck, G. (1973) Ultraviolet-light-induced conformational changes in DNA. *Photochem. Photobiol.* **17**, 387–393.

Lebkowski, J.S., Clancy, S., Miller, J.H., Calos, M.P. (1985) The lacI shuttle: rapid analysis of the mutagenic specificity of ultraviolet light in human cells. *Proc. Natl. Acad. Sci. USA* **82**, 8606–8610.

Legrand-Poels, S., Bours, V., Piret, B., Pflaum, M., Epe, B., Rentier, B., Piette, J. (1995) Transcription factor NF-kappa B is activated by photosensitization generating oxidative DNA damages. *J. Biol. Chem.* **270**, 6925–6934.

Lehmann, A.R., Arlett, C.F., Broughton, B.C., Harcourt, S.A., Steingrimsdottir, H., Stefanini, M., Malcolm, A., Taylor, R., Natarajan, A.T., Green, S., et al. (1988) Trichothiodystrophy, a human DNA repair disorder with heterogeneity in the cellular response to ultraviolet light. *Cancer Res.* **48**, 6090–6096.

Lehmann, A.R., Oomen, A. (1985) Effect of DNA damage on the expression of the chloramphenicol acetyltransferase gene after transfection into diploid human fibroblasts. *Nucleic Acids Res.* **13**, 2087–2095.

Lesko, S.A., Li, W., Zheng, G., Callahan, D., Kaplan, D.S., Midden, W.R., Strickland, P.T. (1989) Quantitative immunofluorescence assay for cyclobutyldithymidine dimers in individual mammalian cells. *Carcinogenesis* **10**, 641–646.

Levine, L., Seaman, E., Hammerschlag, E., Van Vunakis, H. (1966) Antibodies to photoproducts of deoxyribonucleic acids

irradiated with ultraviolet light. *Science* **153**, 1666–1667.

Lindahl, T. (1982) DNA repair enzymes. *Annu. Rev. Biochem.* **51**, 61–87.

Lubin, D., Jensen, E.H. (1995) Effects of clouds and stratospheric ozone depletion on ultraviolet radiation trends. *Nature* **377**, 710.

Lucas, C.J. (1972) Immunological demonstration of the disappearance of pyrimidine dimers from nuclei of cultured human cells. *Exp. Cell Res.* **74**, 480–486.

Lutgerink, J.T., de Graaf, E., Hoebee, B., Stavenuitez, H.F., Westra, J.G., Kriek, E. (1992) Detection of 8-hydroxyguanine in small amounts of DNA by 32P postlabeling. *Anal. Biochem.* **201**, 127–133.

Lyons, M.M., Aas, P., Pakulski, J.D., van Wassbergen, L., Mitchell, D.L., Miller, R.V., Jeffrey, W.H. (1998) Ultraviolet radiation induced DNA damage in coral reef microbial communities. *Mar. Biol.* **14**, 215–222.

Malloy, K.D., Holman, M.A., Mitchell, D., Detrich, H.W., 3rd (1997) Solar UVB-induced DNA damage and photoenzymatic DNA repair in Antarctic zooplankton. *Proc. Natl. Acad. Sci. USA* **94**, 1258–1263.

Mann, D.B., Springer, D.L., Smerdon, M.J. (1997) DNA damage can alter the stability of nucleosomes: effects are dependent on damage type. *Proc. Natl. Acad. Sci.* **94**, 2215–2220.

Mera, S.L., Lovell, C.R., Jones, R.R., Davies, J.D. (1987) Elastic fibres in normal and sun-damaged skin: an immunohistochemical study. *Br. J. Dermatol.* **117**, 21–27.

Mitchell, D.L., Haipek, C.A., Clarkson, J.M. (1985) (6-4)Photoproducts are removed from the DNA of UV-irradiated mammalian cells more efficiently than cyclobutane pyrimidine dimers. *Mutat. Res.* **143**, 109–112.

Mitchell, D.L., Hartman, P.S. (1990) The regulation of DNA repair during development. *Bioessays* **12**, 74–79.

Mitchell, D.L., Nairn, R.S. (1989) The biology of the (6-4) photoproduct. *Photochem. Photobiol.* **49**, 805–819.

Moore, P.D., Bose, K.K., Rabkin, S.D., Strauss, B.S. (1981) Sites of termination of in vitro DNA synthesis on ultraviolet- and N-acetylaminofluorene-treated phi X174 templates by prokaryotic and eukaryotic DNA polymerases. *Proc. Natl. Acad. Sci. USA* **78**, 110–114.

Mori, T., Matsunaga, T., Hirose, T., Nikaido, O. (1988) Establishment of a monoclonal antibody recognizing ultraviolet light-induced (6-4) photoproducts. *Mutat. Res.* **194**, 263–270.

Nakabeppu, Y., Sekiguchi, M. (1981) Physical association of pyrimidine dimer DNA glycosylase and apurinic/apyrimidinic DNA endonuclease essential for repair of ultraviolet-damaged DNA. *Proc. Natl. Acad. Sci. USA* **78**, 2742–2746.

Painter, R.B. (1977) Rapid test to detect agents that damage human DNA. *Nature* **265**, 650–651.

Patrick, M.H. (1970) Near-U.V. photolysis of pyrimidine heteroadducts in *E. coli* DNA. *Photochem. Photobiol.* **11**, 477–485.

Patterson, M., Chu, G. (1989) Evidence that xeroderma pigmentosum cells from complementation group E are deficient in a homolog of yeast photolyase. *Mol. Cell Biol.* **9**, 5105–5112.

Pedrini, A.M., Ciarrocchi, G. (1983) Inhibition of Micrococcus luteus DNA topoisomerase I by UV photoproducts. *Proc. Natl. Acad. Sci. USA* **80**, 1787–1791.

Peters, J., Jagger, J. (1981) Inducible repair of near-UV radiation lethal damage in *E. coli*. *Nature* **289**, 194–195.

Pfeifer, G.P., Drouin, R., Riggs, A.D., Holmquist, G.P. (1991) In vivo mapping of a DNA adduct at nucleotide resolution: detection of pyrimidine (6-4) pyrimidone photoproducts by ligation-mediated polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **88**, 1374–1378.

Povirk, L.F. (1977) Localization of inhibition of replicon initiation to damaged regions of DNA. *J. Mol. Biol.* **114**, 141–151.

Protic-Sabljic, M., Kraemer, K.H. (1985) One pyrimidine dimer inactivates expression of a transfected gene in xeroderma pigmentosum cells. *Proc. Natl. Acad. Sci. USA* **82**, 6622–6626.

Radany, E.H., Friedberg, E.C. (1980) A pyrimidine dimer-DNA glycosylase activity associated with the v gene product of bacteriophage T4. *Nature* **286**, 182–185.

Radicella, J.P., Dherin, C., Desmaze, C., Fox, M.S., Boiteux, S. (1997) Cloning and characterization of hOGG1, a human homolog of the OGG1 gene of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **94**, 8010–8015.

Regan, J.D., Cook, J.S., Lee, W.H. (1968) Photoreactivation of amphibian cells in culture. *J. Cell Physiol.* **71**, 173–176.

Rhoades, D.F., Wang, S.Y. (1971) Photochemistry of polycytidylic acid, deoxycytidine, and cytidine. *Biochemistry* **10**, 4603–4611.

Ribeiro, E.A., Sutherland, J.C. (1991) Quantitative gel electrophoresis of DNA: resolution of overlapping bands of restriction endonuclease digests. *Anal. Biochem.* **194**, 174–184.

Sancar, A., Smith, F.W., Sancar, G.B. (1984) Purification of *Escherichia coli* DNA photolyase. *J. Biol. Chem.* **259**, 6028–6032.

Setlow, R.B. (1966) Cyclobutane-type pyrimidine dimers in polynucleotides. *Science* **153**, 379–386.

Setlow, R.B. (1978) Repair deficient human disorders and cancer. *Nature* **271**, 713–717.

Setlow, R.B., Regan, J.D., German, J., Carrier, W.L. (1969) Evidence that xeroderma pigmentosum cells do not perform the first step in the repair of ultraviolet damage to their DNA. *Proc. Natl. Acad. Sci. USA* **64**, 1035–1041.

Smith, K.C., O'Leary, M.E. (1967) Photoinduced DNA–protein cross-links and bacterial killing: a correlation at low temperatures. *Science* **155**, 1024–1026.

Smith, P.J., Paterson, M.C. (1981) Abnormal responses to mid-ultraviolet light of cultured fibroblasts from patients with disorders featuring sunlight sensitivity. *Cancer Res.* **41**, 511–518.

Smith, R.C., Prezelin, B.B., Baker, K.S., Bidigare, R.R., Boucher, N.P., Coley, T., Karentz, D., MacIntyre, S., Matlick, H.A., Menzies, D., Ondrusek, M., Wan, Z., Waters, K.J. (1992) Ozone depletion: ultraviolet radiation and phytoplankton biology in Antarctic waters. *Science* **255**, 952–959.

Strauss, B., Searashi, T., Robbins, M. (1966) Repair of DNA studied with a nuclease specific for UV-induced lesions. *Proc. Natl. Acad. Sci. USA* **56**, 932–939.

Sutherland, J.C., Griffin, K.P. (1981) Absorption spectrum of DNA for wavelengths greater than 300 nm. *Radiat. Res.* **86**, 399–409.

Taylor, J.S., Lu, H.F., Kotyk, J.J. (1990) Quantitative conversion of the (6-4) photoproduct of TpdC to its Dewar valence isomer upon exposure to simulated sunlight. *Photochem. Photobiol.* **51**, 161–167.

Todo, T., Takemori, H., Ryo, H., Ihara, M., Matsunaga, T., Nikaido, O., Sato, K., Nomura, T. (1993) A new photoreactivating enzyme that specifically repairs ultraviolet light-induced (6-4)photoproducts. *Nature* **361**, 371–374.

Tommasi, S., Swiderski, P.M., Tu, Y., Kaplan, B.E., Pfeifer, G.P. (1996) Inhibition of transcription factor binding by ultraviolet-induced pyrimidine dimers. *Biochemistry* **35**, 15693–15703.

Tornaletti, S., Pfeifer, G.P. (1994) Slow repair of pyrimidine dimers at p53 mutation hotspots in skin cancer [see comments]. *Science* **263**, 1436–1438.

Udono, M.U., Hori, M., Hirayama, M., Yoshida, H., Suematsu, T., Mori, T., Matsunaga, T., Mizuno, T., Nikaido, O. (1993) Visualization of ultraviolet light-induced thymine dimers in DNA by immunoelectron microscopy. *Photochem. Photobiol.* **57**, 752–754.

Venema, J., Mullenders, L.H., Natarajan, A.T., van Zeeland, A.A., Mayne, L.V. (1990) The genetic defect in Cockayne syndrome is associated with a defect in repair of UV-induced DNA damage in transcriptionally active DNA. *Proc. Natl. Acad. Sci. USA* **87**, 4707–4711.

Witkin, E.M. (1966) Radiation-induced mutations and their repair. *Science* **152**, 1345–1353.

Wood, R.D. (1989) Repair of pyrimidine dimer ultraviolet light photoproducts by human cell extracts. *Biochemistry* **28**, 8287–8292.

Wood, R.D. (1997) Nucleotide excision repair in mammalian cells. *J. Biol. Chem.* **272**, 23465–23468.

Yamane, T., Wyluda, B.J., Shulman, R.G. (1967) Dihydrothymine from UV-irradiated DNA. *Proc. Natl. Acad. Sci. USA* **58**, 439–442.

Zelle, B., Lohman, P.H. (1979) Repair of UV-endonuclease-susceptible sites in the 7 complementation groups of xeroderma pigmentosum A through G. *Mutat. Res.* **62**, 363–368.

Ziegler, A., Leffell, D.J., Kunala, S., Sharma, H.W., Gailani, M., Simon, J.A., Halperin, A.J., Baden, H.P., Shapiro, P.E., Bale, A.E., et al. (1993) Mutation hotspots due to sunlight in the p53 gene of nonmelanoma skin cancers. *Proc. Natl. Acad. Sci. USA* **90**, 4216–4220.

DNA Libraries

Chengcang Wu, Zhanyou Xu, and Hong-Bin Zhang
Department of Soil and Crop Sciences, and Institute for Plant Genomics and Biotechnology, 2474 TAMU, Texas A&M University, College Station, Texas 77843-2474, USA

1 History and Development 388
1.1 Recombinant DNA Technologies 389
1.2 Cloning Vectors of DNA Libraries 390
1.3 Transformation 394
1.4 Host Systems 395

2 Types of DNA Libraries 396
2.1 cDNA Libraries 396
2.2 Genomic DNA Libraries 398

3 Construction of DNA Libraries 402
3.1 cDNA Libraries 403
3.1.1 Preparation of Double Strand cDNA 403
3.1.2 Insertion of cDNA into a Plasmid Vector 403
3.1.3 Transformation of the Recombinant Plasmid into *E. coli* 405
3.1.4 Normalization of the cDNA Library 406
3.2 Small-insert Genomic DNA Libraries 407
3.3 Large-insert Genomic DNA Libraries 407
3.3.1 Preparation of HMW Genomic DNA in LMP Agarose Plugs 408
3.3.2 Preparation of Vectors 413
3.3.3 Partial Digestion of HMW DNA Embedded in LMP Agarose Plugs with a Restriction Enzyme 414
3.3.4 Size Selection of DNA Fragments on a CHEF Gel 415
3.3.5 Ligation of the HMW DNA Fragments into the Cloning Vector 416
3.3.6 Transformation of the HMW Recombinant DNA Molecules into *E. coli* Host Cells 416

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

3.3.7 Characterization and Storage of Bacteria-based Large-insert Genomic DNA Libraries 417

4 Applications of DNA Libraries 417
4.1 Gene Discovery, Expression, and Function Studies 418
4.2 Gene Mapping and Cloning 419
4.3 Genome Physical Mapping 419
4.4 Genome Sequencing 421
4.5 Gene Engineering and Gene Therapy 422

5 Concluding Remarks 423

Bibliography 424
Books and Reviews 424
Primary Literature 424

Keywords

DNA Library
A collection or bank of DNA clones generated from DNA of an organism that is either stored in bulk or in array in microplates. The relationship between the clones can be established by sequencing, by restriction mapping and/or by cross-hybridization.

Complementary DNA Library (cDNA Library)
A collection of clones of cDNA synthesized from RNAs present in a cell or an organism by reverse transcription *in vitro* using an enzyme named *reverse transcriptase*. In general, cDNA libraries are constructed from messenger (m) RNA.

Expression Library
A library of DNA clones that are created in an expression vector in order that the genes present in the library can be directly expressed in their host cells.

Genomic DNA Library
A collection of clones made from a set of DNA fragments representing the entire genome of an organism. The fragments can be generated either by partial digestion of source DNA with one or more restriction enzymes, or by physical shearing followed by fragment-end repair. Small-insert genomic DNA libraries are traditionally referred to as those cloned in the plasmid, phage, cosmid, or fosmid vectors, with an average insert size of up to 45 kbp. However, recent studies have shown that the plasmid, cosmid, and fosmid vectors are also capable of cloning and stably maintaining DNA fragments of greater than 300 kbp. The clones of small-insert genomic DNA libraries are often stored in bulk, and thus the libraries are unordered libraries. Large-insert genomic DNA libraries include those cloned in the bacteriophage P1 (P1),

bacteriophage P1-derived artificial chromosome (PAC), bacterial artificial chromosome (BAC), plant-transformation-competent binary BAC (BIBAC), plasmid-based clone (PBC) and yeast artificial chromosome (YAC) vectors, with an average insert size of 100 kb or greater. The clones of large-insert genomic DNA libraries are often arrayed in 96- or 384-well microplates, and thus the libraries are ordered libraries.

Bacteriophage λ
A phage that infects bacteria such as *Escherichia coli*. It has a complex set of regulatory mechanisms to determine whether it will lysogenize or lyze the host bacteria. The bacteriophage lysogenizes if it quietly inserts its DNA into the bacterial genome to become dormant and to be reproduced whenever the host bacterium reproduces. The bacteriophage lyzes the host cell if it hijacks the bacterium's cellular machinery to reproduce itself in order to infect more bacteria, causing the host bacterium to self-destruct shortly after infection. Bacteriophage λ is widely used as a cloning vector for genomic DNA and cDNA library construction.

Plasmid
A small circular genetic element that is stably inherited independent of the chromosome(s) of the host cell. Plasmids have been found in bacteria and fungi. They are widely used as cloning vectors to introduce foreign DNA into both bacterial and yeast cells and for genomic DNA and cDNA library construction.

Cosmid
Artificially constructed plasmid-based cloning vector containing the cos sequence of phage-λ. Therefore, cosmids can be packaged in the phage particles for infection into *E. coli*, permitting cloning of DNA fragments of up to 45 kbp in bacterial hosts.

Cloning Vector
A DNA molecule originating from a virus, a plasmid, or the cell of a higher organism into which a foreign DNA fragment of appropriate size can be integrated without loss of the vector's capacity for self-replication; vectors are used to introduce foreign DNA into host cells where it can be reproduced in a large quantity. Examples are plasmids, cosmids, BACs, PACs, BIBACs, PBCs, and YACs; vectors are often recombinant molecules containing DNA sequences from several sources. In general, a desirable DNA cloning vector should contain elements essential for DNA replication, multiple cloning sites, and selection for both transformants and recombinants (the clones containing inserts).

DNA libraries have been invaluable tools for molecular biology research. They have facilitated the study of the genes and genomes of complex organisms by providing a method of producing large amounts of easily manipulated DNA for the investigation of any region of interest from the cloned DNA representative of the entire genome. A DNA library is a collection of DNA clones constructed *in vitro* by inserting the fragments of source DNA into a cloning vehicle or vector, which contains all

elements required for the delivery of the foreign DNA fragments and stable propagation as a plasmid, a bacteriophage or an artificial chromosome in the host cells. Depending on their DNA source, DNA libraries can be categorized as cDNA libraries, genomic DNA libraries, and oligo libraries. The genomic DNA libraries could further be classified into plasmid, phage, cosmid, P1, BAC, PAC, BIBAC, PBC, and YAC libraries, according to the cloning vectors used. Genomic DNA libraries could be constructed with insert sizes from several up to a million base pairs (bp), depending on research purposes and cloning technologies. Many strategies and techniques have been developed and used for construction of DNA libraries from plants, animals, insects and microbes. DNA libraries are essential resources and tools for all research areas of molecular cell biology and molecular medicine, from gene discovery, cloning, and characterization to gene therapy.

1 History and Development

Recombinant DNA molecules, cloning vectors, and transformation of recombinant DNA into host cells are essential for construction of DNA libraries. The discovery and isolation of the restriction-modification system and DNA ligase had, for the first time, made it possible to reconstruct DNA into recombinant DNA molecules *in vitro*, and thus to construct DNA libraries. Although the restriction-modification system in *E. coli* was first reported in 1953 and DNA ligase was first isolated in 1967, it was not until the early 1970s that biological scientists started to develop recombinant DNA technologies using plasmid and bacteriophage λ as vectors. DNA fragments of interest were enzymatically digested and ligated into a replication-competent, receptive vector to form recombinant DNA clones. The population of individual recombinants was then reintroduced into appropriate host cells to form a DNA library.

In the past 30 years, the techniques and tools for DNA library construction have been developed and significantly advanced. As a result, the size of the DNA molecules that scientists are able to manipulate has been increased from 100 to 10 000 kbp, and the recombinant DNA molecules between cloning vectors and source DNA fragments for library construction can be constructed from dozens of base pairs to 2000 kbp in insert size. To facilitate construction of DNA libraries for different research purposes, a variety of DNA cloning vectors, including plasmid-based, phage-based and artificial chromosome-based, have been developed. Several methods have been developed to transform the recombinant DNA into host cells, including the chemical ($CaCl_2$) method with heat shock, the biological method via bacteriophage packaging and infection, and the physical method using electroporation. The efficiency of transformation has been increased from 10^5 to 10^{10} colony-formed units (cfu) or plaque-formed units (pfu) per microgram of vector DNA. Many different strains of host cells, including bacteria, yeast, and mammalian cells have been developed and used for DNA library construction.

1.1 Recombinant DNA Technologies

The quality of the DNA to be cloned is crucial for the construction of high-quality DNA libraries. Not only should the source DNA be pure, but the fragments of the DNA should also be large enough in size for library construction. This is especially true for the generation of clonable fragments using restriction enzymes (see below). In general, the sizes of the source DNA fragments should be at least 3- to 4-fold larger than the average insert size of the target DNA library. The widely used conventional DNA isolation methods only allow isolation of genomic DNA with a majority of fragments ranging from 50 to 150 kbp in size. The DNA isolated with these methods is suitable for the construction of DNA libraries with average insert sizes of 45 kbp or smaller; however, it is not suitable for the construction of DNA libraries with larger insert sizes such as bacterial artificial chromosome (BAC) and P1-derived artificial chromosome (PAC) libraries. Therefore, the techniques for preparation of high-molecular-weight (HMW) or megabase-size DNA were developed in the late 1980s. In the conventional DNA isolation methods, whole cells, protoplasts, or nuclei are isolated from tissues and lysed, from which naked DNA are purified. However, to isolate HMW DNA, the cells, protoplasts, or nuclei must be embedded in a low-melting-point (LMP) agarose matrix. Cell lysis, DNA purification, and subsequent manipulation must be carried out in the agarose matrix. Use of the agarose matrix has allowed the isolation of DNA with fragment sizes greater than 10 000 kbp. The isolation of DNA by using the megabase-size DNA isolation methods has enabled the construction of DNA libraries with average insert sizes of up to 1000 kbp. Although whole cells have widely been used to isolate HMW DNA from animals, yeast, and microbes, whole cells and protoplasts are not the optimal source for the isolation of HMW DNA from plants, because the plant metabolic substances in the cytoplasm often interact with the DNA, making the DNA no longer digestible with restriction enzymes. Therefore, the nuclei isolation method has been widely adopted for preparation of HMW DNA from plant species. It is this method that has allowed construction of large-insert DNA libraries from different plant species.

Agarose gels are a widely used matrix for DNA fragment separation according to size. In laboratories, DNA fragments are often fractionated on conventional agarose gels by electrophoresis. However, the conventional agarose gel electrophoresis does not allow separation of DNA fragments greater than 50 kbp. To facilitate manipulation of DNA fragments greater than 50 kbp, therefore, a pulsed-field gel electrophoresis (PFGE) system has been developed to resolve DNA fragments between several kilobase pairs and 10 000 kbp in size.

Generation of clonable fragments of an acceptable size range from the isolated DNA is an essential step for recombinant DNA construction and DNA library generation. DNA fragments for library construction can be generated by either physical shearing, or by partial digestion of DNA with a restriction enzyme (enzymatic method). The physical shearing method has proven useful for small-insert shotgun library construction for genome sequencing, whereas the enzymatic partial digestion method has widely been used for construction of large-insert DNA libraries, such as yeast artificial chromosome (YAC),

BAC, binary BAC (BIBAC), PAC and plasmid-based clone (PBC) libraries. The physical shearing method is independent of the distribution and frequency of particular restriction sites. Therefore, the fragments generated with this method are most randomly distributed in the genome, and the library constructed from such fragments is the most representative of the entire genome. However, the sheared DNA fragments have to be repaired by end-blunting, or modified by adding restriction site-containing synthetic linkers prior to cloning. In contrast, the enzymatic partial digestion method is dependent upon the distribution and frequency of particular restriction sites. Since the distribution of the restriction sites of a particular enzyme is not necessarily uniform throughout a genome, the library constructed from the partial digestion of DNA with a single restriction enzyme usually does not have an optimal representation of the source DNA. To overcome or minimize such problems for research, particularly involving whole-genome physical mapping and genome sequencing, two or more DNA libraries are constructed from the partial digestion of DNA with individual restriction enzymes that are complementary in the base compositions of their restriction sites. A major advantage of the enzymatic method is that the partially restricted fragments can be directly cloned, that is, no further enzymatic modification of the restricted fragment ends is necessary. For library construction, especially large-insert DNA library construction, at least four methods have been used to generate clonable DNA fragments: (1) varying the concentration of the restriction enzyme, (2) varying the time of digestion with the same amount of the restriction enzyme, (3) varying the concentration of a restriction enzyme cofactor (e.g. Mg^{2+}), and (4) varying the ratio of the restriction enzyme to the corresponding methylase.

Clonable DNA fragments generated with any of the currently used methods vary widely in size. Therefore, the clonable DNA fragments generated are often subjected to size selection. So far, at least two types of matrices, agarose gels, and sucrose gradients, have been used for DNA fragment size selection. However, agarose gels are the most popularly used matrix for this purpose. To select for DNA fragments smaller than 40 kbp, the conventional agarose gel electrophoresis is routinely used. Pulsed-field gel electrophoresis is the method of choice to select for DNA fragments larger than 40 kbp. The size-selected DNA fragments are then ligated into a DNA cloning vector for library construction.

1.2 Cloning Vectors of DNA Libraries

Since the advent of recombinant DNA technology in the early 1970s, hundreds of DNA cloning vectors have been constructed. Their versatility is represented not only by their cloning capacity as described above, but also by cloning sites, host range and functions, which appear to be limited only by the ingenuity and imagination of inventors. Whereas insertion of foreign DNA fragments into the first plasmid pSC101 was limited to a single restriction enzyme cleavage site and to *E. coli* as a host, most of the vectors currently in use carry multiple restriction-enzyme cleavage sites (polylinker sites) and have been modified for use with most common laboratory microorganisms. Many vectors now include dual origins of replication that enable them to be shuttled between host as dissimilar as *E. coli*, yeast, and even mammalian cells. Function-specific

vectors are now available that enable investigators to streamline particular research applications, such as to clone a genomic DNA fragment and a structural gene, to identify an open reading frame, to identify and isolate regulatory DNA sequences, for example, promoters, and terminators, to overexpress useful RNAs and proteins, and to determine the nucleotide sequence of genes, DNA segments, and the entire genome of an organism.

The DNA cloning vectors can be categorized into three groups – plasmid-based, phage-based, and artificial chromosome-based vectors. For instance, plasmid, cosmid, fosmid, BAC, PAC, and PBC are all plasmid-based vectors; bacteriophage λ, M13, and P1 are phage-based vectors; and YAC and mammalian artificial chromosome (MAC) are artificial chromosome-based vectors (Table 1).

A vector that is desirable for DNA library construction must have at least one element for DNA replication, a multiple cloning site and markers or mechanisms for selection of transformants and the clones containing recombinant DNA. Vectors that also provide additional functions, such as transformation via *Agrobacterium* in plants, have additional elements, such as T-DNA. Phage-based vectors, bacteriophage λ and P1, for example, use the capacity of packaging recombinant DNA into the phage heads for transformant and recombinant DNA clone selection, because the recombinant DNA molecules that are significantly larger or smaller than the natural genome size of the phage cannot be packaged into the phage heads. Most of the modern plasmid-based vectors use antibiotics for transformant selection and colony color (e.g. *LacZ* gene) or clone growth ability (e.g. *SacB* gene) for recombinant clone selection. Artificial chromosome-based vectors use auxotrophic complementation for transformant selection and colony color for recombinant clone selection. The biological method is usually used to transform the recombinant DNA constructed in the phage-based vectors, whereas the biological, chemical, and physical methods are used to transform the recombinant DNA constructed in the plasmid-based and artificial chromosome-based vectors (see below).

The cloning capacity of the vectors, that is, the insert size of foreign DNA that a vector is able to clone and stably propagate in the host cells, is not yet clear. The cloning capacity of the phage-based vectors is limited due to the limited capacity for packaging recombinant DNA into the phage heads; however, the cloning capacity of plasmid and artificial chromosome-based vectors seem to vary greatly. When the plasmid and cosmid vectors were initially developed, they were considered to be capable of cloning only up to 45 kbp. Recent studies have documented that they are capable of cloning and stably maintaining DNA fragments of over 300 kbp in *E. coli*. The increase in cloning capacity of the plasmids and cosmids was achieved with the recent development of technologies for HMW recombinant DNA construction and transformation. As new technologies for large DNA fragment cloning continue to develop, it can be predicted that the cloning capacity of the plasmid and cosmid vectors will be further increased, because *E. coli*, the widely used hosts of the vectors, is capable of stably maintaining and propagating at least 1500 kbp of foreign DNA.

Plasmids were first reported as molecular cloning vectors in 1973. This marked the beginning of a new era of molecular biology research and laid the groundwork for what was to become known

Tab. 1 Cloning vectors and systems of DNA libraries.

Vector type[a]	*Cloning capacity*[b] *[kbp]*	*Examples*[c]	*Structure*	*Year of appearance*[d]	*Host*	*Growth mode on agar plate*	*Number of copies/cell*
Plasmid-based							
Plasmid	0.05–300	pSLJ1711, pUC	Circular	1977	*E. coli*	Colonies	3–100
Cosmid	0.1–300	pCLD04541	Circular	1981	*E. coli*	Colonies	3–5
Fosmid	0.1–300	pFOS1	Circular	1992	*E. coli*	Colonies	1–2
BAC	0.1–300	pBeloBAC11	Circular	1992	*E. coli*	Colonies	1–2
BIBAC	0.1–300	BIBAC2	Circular	1996	*E. coli*	Colonies	1–2
PAC	0.1–300	pCYPAC2	Circular	1994	*E. coli*	Colonies	1–2
PBC	0.1–300	pCLD04541	Circular	1998	*E. coli*	Colonies	1–50
Shuttle plasmid	NA	Yip, Yep, Ycp	Circular	1979	*E. coli* and *S. cerevisiae*	Colonies	1–300
Phage-based							
M13	0.1–3.5	M13mp18	Circular	1981	*E. coli*	Plaques	NA
Lambda	10–20	EMBL, FIX	Linear	1983	*E. coli*	Plaques	NA
P1	50–100	pAd10SacBII	Linear	1990	*E. coli*	Plaques	1–2
Artificial chromosome-based							
YAC	Up to 2000	pYAC4	Linear	1987	*S. cerevisiae*	Colonies	1
MAC	Up to 5000	α7C5	Linear	1998	Mammalian cells	Colonies	1

[a]YAC and MAC are also shuttle vectors.
[b]The cloning capacity of plasmid and artificial chromosome-based vectors varies, depending on the cloning technologies.
[c]Many different vectors have been developed for different research purposes within each category.
[d]The year indicates the year of appearance of the modern form of the vector, but not the first reported use of the vector type. "NA" indicates "Not Available."

as the biotechnology industry. Previously, the plasmid vectors were used only to construct DNA libraries with insert sizes ranging from several hundreds to several thousand base pairs. Although bacteriophage λ can package approximately 50 kbp of DNA into its heads, the phage itself requires at least 30 kbp of essential genes and elements for viability, which limits the cloning capacity of phage λ vector to no more than 20 kbp.

Until the early 1990s, cosmids were the widely used vectors for cloning large-insert DNA fragments in *E. coli*. They allowed cloning of DNA fragments up to 45 kbp. Cosmid vectors had not only the advantage of the high efficiency of cos-mediated bacteriophage λ packaging for transformation but also had the advantage of small plasmid vectors. To facilitate reproduction of the recombinant DNA in the host cells, most, if not all of the cosmids previously developed are maintained as multiple copies per cell. However, while the DNA libraries cloned in the cosmid vectors were extensively used in genome research, the rearrangements of the cloned DNA were observed in some host *E. coli* strains and attributed to the multiple copies of the cosmids within a host cell. Therefore, fosmid vectors were designed from the single-copy F plasmid of *E. coli*. As with cosmids, fosmids also contain cos sites for cos-mediated phage λ packaging; therefore, they have the same cloning capacity as cosmids. Even though fosmids appeared to be more stable than the cosmid clones, instability of fosmid clones was also observed in some *E. coli* host strains. Therefore, further studies were needed to determine the relationships of clone stability with the copy number of vectors and host strains.

Cosmid cloning vectors had played an important role in large-insert DNA library construction in the early stages of genome research. However, the insert sizes (up to 45 kbp) of the cosmid DNA libraries were insufficient for genome research of species with large genomes. Therefore, the artificial chromosome-based YAC cloning system was developed in 1987. Using yeast as the host, YAC vectors permitted the cloning of fragments greater than 1000 kb, which represented a revolutionary advance in cloning capacity. However, several problems, such as high level of chimeric clones, insert rearrangement and difficulty in purifying cloned insert DNA, were observed in this system. These problems limited the utility of YAC libraries for genome research.

The development of alternative systems that would be capable of cloning and maintaining large DNA fragments in bacteria, therefore, became imperative. In the early 1990s, the P1, BAC, and PAC systems were developed. The P1 vectors were derived from the bacteriophage P1 with a genome size of about 100 kbp. Since the P1 recombinant DNA is transformed into the host *E. coli* cells by the same procedure as in bacteriophage λ, namely, packaging and infection, the P1 vectors have a cloning capacity of up to 100 kbp. The BAC vectors were based on the single-copy F factor and derived from the fosmid vectors. Because BAC constructs are transformed into the host *E. coli* cells by electroporation, instead of $CaCl_2$/heat shock, BACs having insert sizes of over 300 kbp could be obtained. The modification of the P1 vector combined with the BAC electroporation technology led to the development of the PAC system. Therefore, the PAC vectors have a cloning capacity similar to the BAC vectors. In comparison with the YAC system, despite their smaller insert sizes, BACs and PACs are much more stable in the host cells, much lower in chimerism,

and more easily purified from the host chromosome DNA, and therefore, have been rapidly adopted for genome research.

Unlike small-insert clones, the entire insert of the large-insert BACs and PACs are difficult to subclone. Therefore, it is desirable to develop a cloning system that is not only capable of cloning and stably propagating foreign DNA fragments as large as BACs and PACs, but can also directly be transformed into plants. Therefore, an *Agrobacterium*-mediated, plant-transformation-competent BIBAC system was developed in 1996. In addition to the features of the BAC vectors, the BIBAC vector has a T-DNA cassette to facilitate BAC insert transfer into plants via *Agrobacterium*. As expected, the BIBAC vector is capable of stably cloning and propagating DNA fragments as large as BACs and PACs, and also capable of transferring the BAC inserts into plants. This has significantly streamlined the use of the large-insert clones for positional cloning, functional analysis of genomic sequences, and genetic engineering.

BAC, PAC, and BIBAC are all based on a single-copy "artificial chromosome." Nevertheless, the conventional plasmids or cosmids that were previously designed for different research purposes were later found to have the same capacity as the BAC, BIBAC, and PAC vectors for large fragment cloning and stable maintenance in *E. coli* strain DH10B, the host strain of almost all BAC and PAC libraries constructed to date. The breakthrough in the cloning capacity of the conventional plasmid and cosmid vectors clearly resulted from the use of the HMW recombinant DNA technique and the electroporation technology for transformation as in BACs and PACs. Since the plasmids and cosmids are not single-copy as the BAC and PAC vectors, the high stability of the large-insert PBCs in the host cells indicated that the stability of bacteria-based large-insert clones is not contingent on the unique copy number of the clones in a host cell. These findings have not only established concepts of large DNA fragment cloning in bacteria but have also indicated that many of the plasmid-based vectors previously developed for different biological research purposes could be used as vectors for large-insert DNA library construction in *E. coli*.

1.3 Transformation

DNA library construction also involves techniques of transformation by which the recombinant DNA constructed *in vitro* is delivered into the host cells where it is propagated. As recombinant DNA technologies and DNA cloning vectors evolved, techniques for several transformation methods have been developed and used for DNA library construction. There are three transformation methods – chemical ($CaCl_2$/heat shock), biological (*in vitro* packaging and infection), and physical (electroporation). As described above, although the biological method is routinely used to transform the phage-based vectors such as bacteriophages λ and P1, all three methods are used to transform the plasmid- and artificial chromosome-based vectors such as plasmids, cosmids, fosmids, BACs, PACs, PBCs, YACs, and MACs. The chemical and biological transformation methods were developed when the recombinant DNA research was initiated in the 1970s. At present, they are still widely used for DNA cloning and library construction. During the past 30 years, many procedures have been developed, and the transformation efficiency has been

increased several-fold. Now it is routine to obtain a transformation efficiency of 10^9 cfu μg^{-1} plasmid vector DNA using these transformation methods. The electroporation transformation method was developed a decade ago. It was this method that made it possible to construct large-insert DNA libraries in bacteria. The electroporation transformation method is much simpler than the previously developed chemical and biological methods and has a much higher transformation efficiency, greater than 10^{10} cfu μg^{-1} plasmid vector DNA. Therefore, it has been rapidly adopted for construction of different DNA libraries, including cDNA libraries, large-insert BAC, PAC, BIBAC, and PBC libraries and small-insert plasmid libraries.

1.4 Host Systems

Host cells that are competent for DNA library construction must be capable of stably maintaining and propagating cloned foreign DNA, and have high transformation efficiencies. As the recombinant DNA technologies developed, several host systems that are suitable for DNA cloning and library construction have been developed. According to the sources of the hosts, they can be categorized into bacterial, *Saccharomyces cerevisiae* (yeast), mammalian, and other organism hosts (Table 2). Bacterial *E. coli* is the most widely used host system for DNA library construction. Wild-type *E. coli* strains are highly proficient in recombination, which causes deletions and rearrangements through intramolecular recombination and intermolecular recombination as well, in the case of multicopy vectors at the high repetitive sequences cloned. Poor clonability also results from incompatibility of the foreign DNA with the natural host restriction and modification systems. To reduce the ability of the bacterial host to modify cloned DNA fragments in such ways, many *E. coli* mutations have been identified or created in

Tab. 2 Examples of the host systems for DNA libraries.

Host	*Strain*	*Genotype*	*Compatible with vector types*
E. coli	DH5α	F^- Φ80d*lacZ*ΔM15 Δ(*lacZY-argF*)U169 *deoR recA1*	Plasmid, phage, cosmid
		endA1 hsdR17(r_K^- m_K^+) *supE44* λ^- *thi-1 gyrA96 relA1*	Fosmid
	DH10B	F^- *mcrA* Δ(*mrr-hsdRMS-mcrBC*)Φ80d*lacZ* ΔM15	Plasmid, phage, cosmid, fosmid,
		Δ*lacX74 deoR recA1 endA1 araD139*	P1, BAC, PAC, BIBAC, PBC
		Δ(*ara leu*) 7697 galUλ^- *rpsL nupG*	
S. cerevisiae	AB1380	MATa ψ + ura3-52 trp1 ade2-1 his5 lys2-1 can1-100	Plasmid, YAC
Mammalian cells	HT1080 (ATCC CCL 121)	NA	MAC

"NA" indicates "Not Available."

host systems available for DNA library construction. For example, the host *E. coli* strains DH5α and DH10B and their derivatives are widely used for the construction of plasmid, bacteriophage and cosmid libraries, and DH10B is routinely used for the construction of BAC, PAC, BIBAC and PBC libraries.

The budding yeast *S. cerevisiae* is a host system that is widely used for DNA cloning and library construction in yeast. Yeast is a single-cell eukaryotic organism and thus has all the basic features and mechanisms of eukaryotes. Not only is yeast as well studied as *E. coli*, and amenable to large-scale isolation techniques, but it is able to carry out some eukaryote-specific processes, such as posttranslational processing of eukaryotic proteins, eukaryotic DNA replication, telomeric function or functional chromosomes. These features and mechanisms have rendered the yeast host system more advantageous than the bacterial host system for functional analysis of eukaryotic genes. A number of different *S. cerevisiae* host strains are available to meet various library applications for studying insertion, deletion, alteration, and expression of genes. For YAC library construction, the primary requirement of a host is its ability to be transformed at reasonably high frequencies with large-insert DNA. AB1380 and a few other strains have been found suitable for this purpose.

In addition, mammalian cells have been used as a host system for MAC library construction. *Bacillus subtilis* has the ability to secrete proteins into growth medium and has been used as a host system for industrial production of recombinant proteins. Baculoviruses that infect insect cells have gained attention as a host system for production of glycosylated protein at very high levels.

2 Types of DNA Libraries

In the past 30 years, many DNA libraries have been developed for different molecular biology research purposes. The libraries can be categorized into cDNA, genomic DNA, and oligo libraries by their source DNA; plasmid, phage, cosmid, fosmid, P1, BAC, PAC, BIBAC, PBC, YAC, and MAC libraries by their cloning vectors; or general, binary, and expression libraries by their functions. Both cDNA and genomic DNA libraries could be general, binary, or expression libraries, whereas cDNA libraries could be plasmid, phage, or cosmid libraries and genomic DNA libraries could be plasmid, phage, cosmid, fosmid, P1, BAC, PAC, BIBAC, PBC, YAC, or MAC libraries.

2.1 cDNA Libraries

cDNA libraries are those DNA libraries constructed from DNA that are reverse-transcripted from and complementary to an organism's RNA pools. cDNA can be prepared from any source RNA, including coding and noncoding RNA, isolated from single-cell organisms, cultured cells and isolated tissues. However, almost all cDNA libraries are generated from total mRNAs in a cell representing all protein-encoding genes that are being expressed in the cell at a particular time, developmental stage, or treatment. The polyadenylated mRNA had previously undergone posttranscriptional editing and removal of any intervening sequences (introns). Thus, the copy of a gene in an mRNA or cDNA usually contains an uninterrupted sequence encoding the gene product.

A high-quality cDNA library should not only contain the sequences of all

genes expressed at the time when the RNA samples are collected but should also be full length and balanced in the copy number of every gene expressed. The probability that a given mRNA is represented in a cDNA library can be calculated by the following formula:

$$p(x) = [1-(1-f)_n]$$

where n is the number of recombinant clones and f is the frequency of a particular mRNA. Since a typical somatic cell contains approximately 0.6 pg of mRNA, approximately 500 000 mRNA molecules are present in a cell ($0.6\ \text{pg}/11 \times 10^{-7}$ pg/1 mRNA molecule) if the average size of an mRNA is 2000 bases (11×10^{-7} pg). Accordingly, the probability of obtaining a rare mRNA (one copy per cell, $f = 1/500\,000$) in a library of 5 000 000 and 1 000 000 recombinant clones, would be 99.9 and 86.5% respectively.

In order to construct a high-quality cDNA library, it is essential to isolate full-length mRNAs encoding full-length protein sequences in the subject cells or tissues. Libraries made from fragmented mRNA would not preserve the full open reading frames of the genes, thus rendering the resultant partial-length cDNA libraries less useful. Message RNA tends to be a transient species by design as cells have developed complex posttranscriptional mechanisms to regulate gene expression through mRNA decay. The mRNA decay problem can be overcome through specialized procedures that seek to inactivate ribonuclease, an enzyme that cleaves RNAs, concomitantly with mRNA extraction from intact tissues.

Another concern of a high-quality cDNA library is the copy number balance of each gene expressed in the library. Such copy number-balanced cDNA libraries are often obtained through a process named normalization of cDNA populations. The mRNAs of a typical somatic cell are distributed into three frequency classes: prevalent, intermediate, and rare (complex). The classes at the two extremes, representing 10% and 40 to 45% of the mRNA species respectively, include the members occurring at vastly different relative frequencies. In general, the most prevalent class consists of approximately 10 mRNA species, each represented by 5000 copies per cell, whereas the class of high complexity comprises 15 000 different species, each represented by only 1 to 15 copies. Therefore, for a variety of research purposes, a normalization procedure is needed to bring the frequency of each clone in a cDNA library within a narrow range.

Plasmids, phages, or cosmids have widely been used as cloning vectors for cDNA library construction. Many procedures have been developed for cDNA library construction, such as directionally cloned cDNA libraries followed by normalization. cDNA libraries have served as, and will continue to be, one of the most important resources for gene mapping, discovery, cloning, expression, and functions (see below). cDNA libraries can be screened either by detection of cDNA inserts using polymerase chain reaction (PCR)- or hybridization-based methods, or by detection of RNA or protein expression products encoded by the cDNA inserts. PCR- or hybridization-based library screening methods are the same in principle for screening cDNA libraries and genomic libraries. For PCR-based library screening, super pools, plate pools, column pools, and row pools must be generated from the library and used as templates of PCR. The PCR primers are designed and synthesized according to the sequences of the existing genes. For the

hybridization-based library screening, the library should be blotted onto nylon membranes or filters. Either synthetic oligos (e.g. overlapping oligos named *Overgos*) designed according to the gene sequences or the partial or entire sequences of the genes could be used as probes. The probes are labeled with one or more radioactive or fluorescent nucleotides and hybridized to the library that is blotted onto the nylon membranes. Colonies or plaques that contain the inserts complementary to the probe can then be detected and isolated. The cDNA expression libraries can be screened on the basis of affinity of the target gene product to antibodies, proteins, or small-molecule ligands. In addition, functional properties can be screened for, such as catalysis or induction of transcription of a reporter gene.

2.2 Genomic DNA Libraries

In contrast to cDNA libraries, genomic DNA libraries are constructed from DNA sequences derived from an organism's chromosome(s) or entire genome. Early genomic DNA libraries were constructed in the plasmid, bacteriophage, and cosmid vectors, and had average insert sizes of up to 45 kbp. These libraries are suited for isolation and characterization of most genes, but not for the genes or gene clusters that span a large physical distance in the species with large genomes. Cloning and characterization of such genes or gene clusters require libraries that accommodate larger inserts. Furthermore, recent genome research and analysis require DNA libraries with even larger inserts. Therefore, YAC libraries that have average insert sizes of over 1000 kbp were developed for several plant and animal species. Nevertheless, the YAC libraries have high levels of chimerism, are unstable in inserts, and difficult to purify cloned insert DNA, all of which have limited their utility in genome research. The advent of bacteria-based large-insert cloning systems, BAC, PAC, BIBAC, and PBC, has shed light on large-insert genomic DNA library construction and its applications in genome research. Although the insert sizes of the BAC, PAC, BIBAC, and PBC libraries do not approach those of YAC libraries, they are stable in the bacterial hosts and have few chimeric clones. Most importantly, high-quality and high-quantity insert DNA of the large-insert bacterial clones can readily be purified by using the simple plasmid-based DNA preparation procedures (Fig. 1).

Three methods, PCR-based DNA amplification, physical shearing, and enzymatic digestion, have been used to prepare clonable DNA fragments for construction of genomic DNA libraries. PCR products are used to construct genomic DNA libraries specific to the sequences containing the primer binding sites, for which specific plasmid-based vectors have been developed. As described above, the physical shearing method generates a series of partially overlapping fragments distributed evenly along a genome. Thus, it has widely been used to produce DNA shotgun libraries for genome sequencing. This method has contributed to the genome sequencing of *Arabidopsis*, human, *Drosophila*, mouse, and rice, and will continue to be one of the most important tools for genome sequencing.

Restriction enzymes, either frequent (4 bp) or rare (6 or 8 bp) cutters, have also been used to produce DNA fragments for construction of genomic DNA libraries of

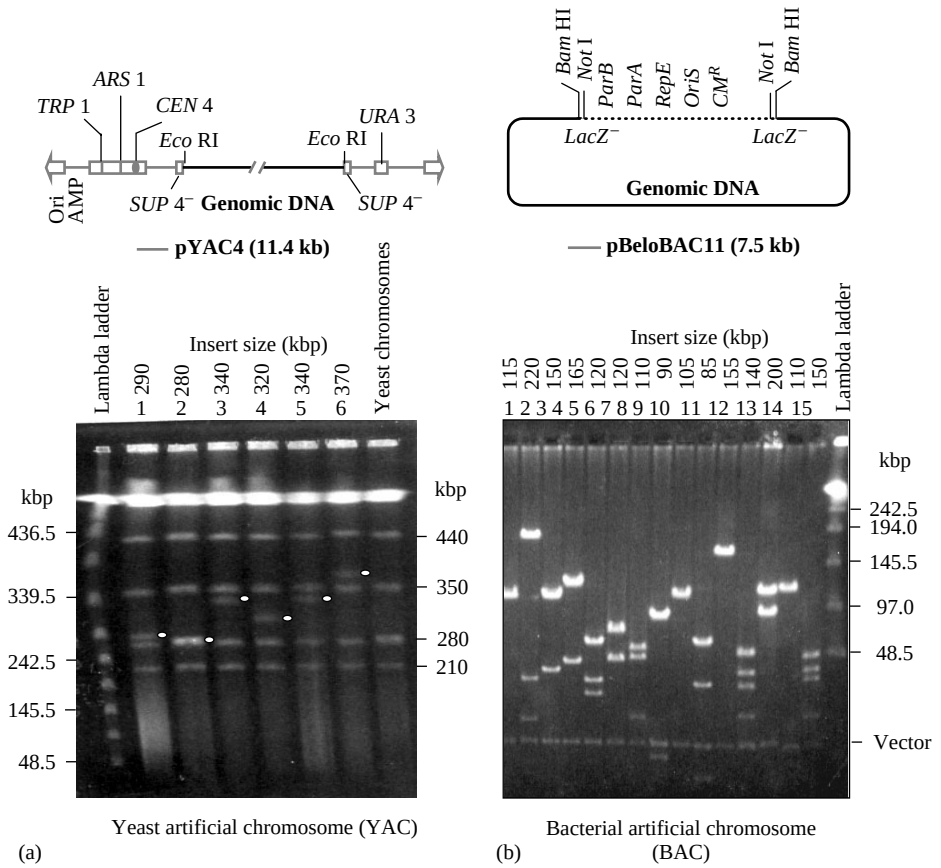

Fig. 1 (a) A yeast artificial chromosome (YAC) constructed in the YAC vector pYAC4 (top) and YACs fractionated on a pulsed-field gel (bottom). In the YAC construct, the terminal arrows indicate the telomeres, *CEN4* is the centromere of the artificial chromosome; and the bold-faced *Eco* RI is the cloning site. For YAC analysis, YAC DNA was isolated, subjected to PFGE, stained, and photographed. The YACs on the gel are indicated by bullets, and the bands appearing in all lanes are the yeast chromosomes. The YAC clones have a size range from 280 to 370 kbp; (b) A bacterial artificial chromosome (BAC) constructed in the BAC vector pBeloBAC11 (top) and BACs fractionated on a pulsed-field gel (bottom). In the BAC construct, the bold-faced *Bam* HI is the cloning site. For BAC analysis, BAC DNA was isolated, digested with *Not* I to release the inserts, subjected to PFGE, stained, and photographed. The band appearing in all lanes was derived from the BAC vector and the remaining bands were from the insert DNA. The BAC clones have an insert size range from 85 to 220 kbp.

different insert sizes. Although the complete digestion of genomic DNA with a restriction enzyme can be used for genomic DNA library construction, the clones of the resulting library do not overlap, which is undesirable for many aspects of molecular biology research. Therefore, most genomic DNA libraries were constructed from the partial digestions of source genomic DNA. The partial digestion method has been widely used in the construction of large-insert genomic DNA libraries. Because

large DNA fragments become dramatically more sensitive to physical shearing as their sizes increase, making it impossible to maintain the integrity of chromosome size in DNA molecules, specialized protocols have been developed that enable the cloning of DNA fragments of several hundred kilobase pairs, such as BAC, PAC, BIBAC, PBC, and YAC libraries.

The development of large-insert DNA libraries has not only facilitated cloning and characterization of genes and gene clusters that span large genomic regions but, importantly, it has also reduced the number of clones needed for a complete genomic DNA library, and thereby allowed each clone to be arrayed in multiple-well (384 or 96) microplates. In contrast, small-insert genomic DNA libraries are often maintained in bulk because of the prohibitively large number of clones needed for a complete library. The ability to array each individual clone of a large-insert DNA library in microplates has significantly facilitated isolation of clone(s) of interest from the library. Moreover, each arrayed clone has a permanent address that can be easily shared among investigators who have access to the library, along with any available information such as genes and nucleotide sequences contained in the clone. As such, data about a specific clone can rapidly accumulate, and the clone becomes more valuable through the contribution of the whole research community. At the present time, arrayed, large-insert genomic DNA libraries in forms of arrayed plates, high-density clone filters and/or DNA pools from a variety of plants, animals, human, and microbes are publicly available at several genomic resource facilities, such as http://hbz.tamu.edu, http://informa.bio.caltech.edu/lib_status.html, http://www.resgen.com/, http://www.chori.org/bacpac, and http://www.genome.clemson.edu/groups/bac/libraries_genomic.html.

Large-insert genomic DNA libraries have been widely used for a variety of molecular biology research studies ranging from gene discovery, cloning, and genome physical mapping to genome sequencing (see below). Studies showed that the representation of a genomic DNA library for the entire genome is of significance for use of the library in molecular biology research. Theoretically, the probability (p) that a given genomic sequence is present in a library can be estimated by a formula as follows:

$$p = 1 - e^{-n}$$

where n is the number of genome equivalents of the library, which equals the product of the number of total clones and the average insert size divided by the genome size. Table 3 shows the number of clones that are needed to have a 99.9% probability of obtaining a particular DNA sequence from a genomic DNA library with an average insert size of 50, 100, 150 or 500 kbp for *Arabidopsis, Drosophila*, rice, human, and wheat respectively. For a genome of a given size, the larger the clone inserts, the fewer the clones that are needed for a genomic DNA library. Large inserts are also significant for the efficiency and success of genome research, such as chromosome walking, and regional or whole-genome physical mapping and sequencing. The larger the inserts the fewer the steps, and the higher the likelihood of successfully walking a genomic region. The probability of success of completing a chromosome walk increases linearly as average insert sizer of the library increases from 50 to 150 kbp; however, there is little increase in probability of chromosome walking success once the average insert size exceeds 150 kbp.

Moreover, studies further showed that the theoretical (calculated) probability of genome coverage of a large-insert genomic DNA library constructed with a single restriction enzyme is about 15% lower than the genome coverage estimated by library screening. This implies that when the library is screened with random DNA markers, only about 85% of the markers are represented by one or more positive clones, whereas the remaining 15% are not represented in the library. The studies also showed that the genome coverage of the library could not be increased significantly by simply increasing the number of clones in the library. This is due to at least two notable reasons. First, some genomic regions may be lethal or poorly clonable, maintained and/or propagated in the host system; to clone these DNA sequences, a different vector and host system is needed. Second, the restriction sites of the enzyme used for the library construction are not uniformly distributed along the genome; consequently, the genomic regions having too many or too few restriction sites of the enzyme are difficult to clone because the small (<50 kbp) and large (>400 kbp) DNA fragments generated by partial digestion are removed during the size selection of large-insert DNA library construction. Therefore, it is recommended that at least two restriction enzymes with different nucleotide contents (AT-rich and GC-rich) in their restriction sites be used for large-insert genomic DNA library construction. This approach would minimize the bias of clone distribution along the genome, and thus achieve libraries of truly high genome coverage, because the large-insert DNA libraries constructed with this approach are shown to be complementary to each other, similar to mechanically sheared shotgun libraries.

In addition, unlike small-insert clones, the entire inserts of large-insert clones are difficult to subclone for functional analysis of the inserts by genetic transformation. This is because it is difficult to find an enzyme whose restriction sites flank the cloning site of the vector, but are absent in the large insert of the clone. Although the restriction enzyme *Not* I, a rare 8-bp cutter, is used to subclone the inserts of large-insert clones into transformation-competent binary vectors, it is present at an average frequency of one restriction site in every 65 536 bp in a genome, which is smaller than the 150-kb average insert size of a large-insert DNA library. Studies showed that the frequency of the *Not* I sites is much higher in the genomes of monocotyledonous plant species than those of dicotyledonous plant species. Therefore, it is desirable to construct large-insert DNA libraries in binary vectors that are capable of stably cloning and maintaining large DNA inserts, and that are also competent for large DNA fragment transformation. If the large-insert binary libraries are used for whole-genome physical mapping and sequencing, the ability of the binary clones to directly transform plants would significantly streamline the functional analysis of genomic sequences by genetic transformation.

Great efforts are needed to construct a large-insert DNA library for a species with a huge genome, for instance, common wheat (16 000 Mbp/1C) (see Table 3). According to the above formula, the higher the genome coverage the higher the probability of obtaining a clone of interest from the library. However, the increase in library genome coverage would increase the cost and effort that go into library production, storage, and manipulation. Therefore, an efficient genome coverage

Tab. 3 The number of clones needed for a 99.9% coverage genomic DNA library with an average insert size of 50, 100, 150, and 500 kbp respectively, for several species.

Species	*Genome size*	*Average insert size [kbp]*			
	[Mb/1C]	*50*	*100*	*150*	*500*
Arabidopsis	145	2.0×10^4	1.0×10^4	6.7×10^3	2.0×10^3
Drosophila	180	2.5×10^4	1.3×10^4	8.3×10^3	2.5×10^3
Rice	430	5.9×10^4	3.0×10^4	2.0×10^4	5.9×10^3
Human	3000	4.1×10^5	2.1×10^5	1.4×10^5	4.1×10^4
Wheat	16 000	2.2×10^6	1.1×10^6	7.4×10^5	2.2×10^5

of a large-insert DNA library would be helpful for use of the library in genome research. Although a DNA library with 4.7 x genome equivalents is equivalent to a 99% library according to the above formula, it is recommended that the total number of clones with an average insert size of at least 150 kb derived from two or more restriction enzymes and equivalent to a maximum of 10 x haploid genomes is sufficient for most genome research purposes, including global genome physical mapping and genome sequencing. This is supported by the whole-genome shotgun sequencing of *Haemophilus influenzae*, *Mycoplasma genitalium*, *Anopheles gambiae*, and the euchromatic genome portion of the *D. melanogaster*, and by whole-genome physical mapping of *Arabidopsis*, rice, and soybean, in which the numbers of clones equivalent to 6 – 10 x haploid genomes were used.

3 Construction of DNA Libraries

Construction of cDNA and genomic DNA libraries are essentially the same in principle; both include preparation of clonable DNA fragments, preparation of vectors, ligation of the clonable DNA fragments into the vectors, transformation of the ligation mixture into host cells, and library storage. The major difference between cDNA and genomic DNA library constructions is the preparation of clonable DNA fragments. For cDNA library construction, the clonable cDNAs are prepared from mRNA, whereas for genomic DNA library construction, the clonable DNA fragments are prepared from genomic DNA. The method that is used to transform the DNA/vector ligation into host cells depends on the vectors used for the library construction. For the construction of plasmid, M13, cosmid, fosmid, BAC, PAC, BIBAC, and PBC libraries, the electroporation method is widely used to transform the ligations into *E. coli* host cells, although the cosmids and fosmids were initially designed to transform into *E. coli* cells through the phage *in vitro* packaging and infection method. For the construction of bacteriophage λ and P1 libraries, the phage *in vitro* packaging and infection method is used to transform the ligation into *E. coli* host cells. For YAC library construction, the ligation is transformed into yeast host cells using the chemical method through yeast spheroplasts. Small-insert DNA libraries are usually maintained in bulk, while large-insert DNA libraries are arrayed in 96- or 384-well microplates. Nevertheless, small-insert DNA libraries

for special research purposes, such as expressed sequence tag (EST) sequencing and genome sequencing, are also arrayed in 96- or 384-well microplates. Outlined below is the construction of cDNA and genomic DNA libraries to demonstrate the principles and methods of DNA library construction.

3.1 cDNA Libraries

cDNA libraries are crucial tools for the study of tissue-specific genes and their expression. A customized cDNA library should contain a stable representation of the mRNA present in a specific tissue at the time of tissue collection. Four steps are often involved in the construction of a typical cDNA library (Fig. 2).

3.1.1 Preparation of Double Strand cDNA

1. *From mRNA to double strand cDNA*: "Reverse transcription" is a process whereby genetic information contained in an mRNA is converted back into a double-stranded DNA form. This DNA is named complementary DNA, or cDNA. The enzyme responsible for this process is an RNA-dependent DNA polymerase called reverse transcriptase. Reverse transcriptases have traditionally been isolated from viruses whose genomes are actually in an RNA form and must be converted to duplex DNA during propagation. These viruses typically carry a functional reverse transcriptase along with their mRNA genetic component when they infect cells. One of the most common commercially available reverse transcriptases is purified from moloney murine leukemia virus (MMLV). MMLV uses mRNA as a template, but requires a primer because it can extend a DNA primer, but cannot synthesize one.
2. One of the useful features of eukaryotic mRNA is the presence of the 3′ poly A tracks. They provide a unique method to synthesize cDNA complementary to the original mRNA strand from a variety of different eukaryotic mRNAs using poly dT as a single primer. When the first strand of cDNA is synthesized, the mRNA/cDNA hybrid duplex is developed (see Fig. 2).
3. Nicks in the RNA half of the hybrid molecule can be introduced via the action of the enzyme RNase H. This enzyme exhibits endonucleolytic cleavage of the RNA moiety of the RNA/DNA hybrids, as well as 5′ → 3′ and 3′ → 5′ exoribonuclease activity. In other words, it will nick the RNA and then proceed to digest it back in both directions.
4. These RNA fragments can now serve as primers for the second strand cDNA synthesis by *E. coli* Polymerase I (Pol I). This enzyme also translates the "nicks" to remove the RNA primers.
5. T4 DNA ligase ligates the nicked DNA and develops the whole second strand cDNA. Therefore, the double strand (ds) cDNA is completed. The ds DNA has either a residual 5′ RNA cap region, or a gap at the 5′ end of the original mRNA strand.

3.1.2 Insertion of cDNA into a Plasmid Vector

In order to complete the construction of a cDNA library, the cDNA needs to be maintained and propagated. This can be accomplished by inserting the cDNA into an appropriate plasmid, cosmid, or phage. Here we use a plasmid vector as an example because plasmid vectors used

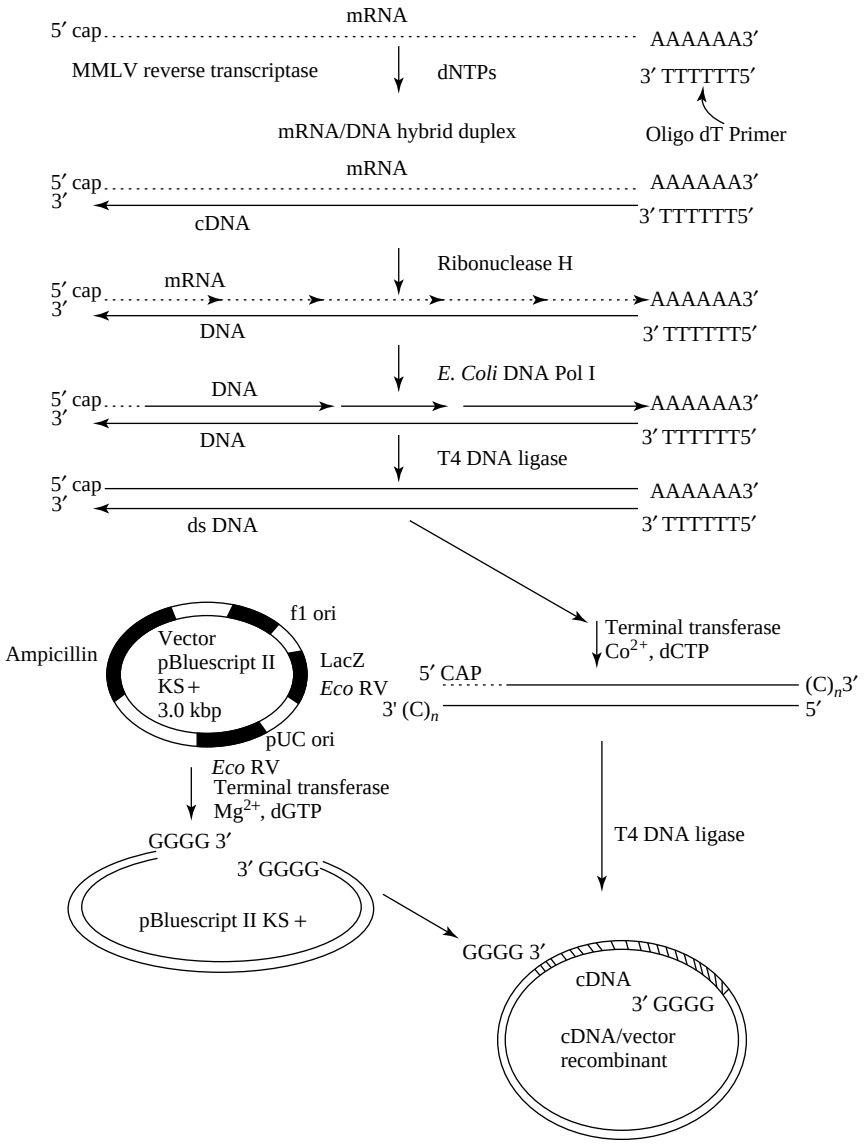

Fig. 2 A general procedure for the construction of cDNA/vector recombinants for cDNA library construction. pBluescript II KS+ is the cloning vector; "DNA Pol I" indicates "DNA Polymerase I;" "dNTPs" indicates "dATP, dCTP, dGTP and dTTP;" and "ds DNA" indicates "double strand DNA."

in combination with the extremely high-efficiency electroporation transformation are becoming the method of choice for cDNA library construction. There are two ways of preparing the synthesized cDNA into clonable cDNA – homopolymeric tailing and linker addition. Homopolymeric tailing is completed by an enzyme called

terminal transferase, which is an unusual DNA polymerase found only in a type of eukaryotic cell called a *prelymphocyte*.

1. In the presence of a divalent cation, terminal transferase catalyzes the addition of dNTPs to the 3′-hydroxyl termini of DNA. In the presence of Mg^{2+}, a purine (G or A) is added, and in the presence of Co^{2+}, a pyrimidine (C or T) is added.
2. Depending on the reaction conditions, anywhere from three to several thousand nucleotides can be added.
3. One nucleotide (e.g. dCTP) is added to the 3′ ends of the ds cDNA using the terminal transferase in the presence of Co^{2+}; the plasmid vector is digested with a restriction enzyme (e.g. *Eco* RV); the complementary nucleotide (e.g. dGTP) is added to the one already added to the cDNA at the 3′ ends of the vector; and it is digested using the terminal transferase in the presence of Mg^{2+}. Now the tailed vector and cDNA are mixed, annealed and ligated into cDNA/vector recombinant DNA (also called ligation mixture). The molar ratio of insert : vector for the ligation is from 8–15 : 1, of which routinely 50 to 95% of the transformants are clones containing recombinant DNA.

Alternatively, ds cDNA can be cloned into a library vector through the addition of "linkers" to the ds cDNA. Linkers are short oligonucleotides (~18 to 24 mers), which are typically palindromic and contain a single or repeated restriction endonuclease recognition sequence (e.g. *Eco* RI site). The palindromic nature allows the linker oligonucleotide to self-hybridize to form a blunt-ended duplex. If the ends of the cDNA are blunt, then the linker can be ligated to both ends to introduce terminal restriction cloning sites. The steps of linker addition are as follows:

1. Treat the cDNA with S1 nuclease to remove possible 5′ cap mRNA fragment remaining in cDNA duplex.
2. Convert potential "ragged" ends to blunt by treatment with Pol I. The Pol I will fill in 5′ overhangs and chew back 3′ overhangs.
3. Methylate cDNA at the potential internal *Eco* RI sites of the cDNA by treatment with *Eco* RI methylase (plus S-adenosyl methionine).
4. Ligate *Eco* RI-containing linkers to the end-blunted, methylated cDNA using T4 DNA ligase.
5. Cut the linkers with *Eco* RI restriction endonuclease and remove the excess linkers from cDNAs by agarose gel electrophoresis.
6. Ligate cDNA to a cloning vector prepared using *Eco* RI restriction endonuclease, followed by a process called dephosphorylation (see below for vector preparation; see above for the ligation molar ratio of insert : vector = 8–15 : 1).

3.1.3 Transformation of the Recombinant Plasmid into *E. coli*

The procedure of transferring the recombinant plasmid, phage, or cosmid into host cells is called transformation. Before transformation, the host cell must be treated to make the cell competent to accept foreign DNA such as plasmid or cosmid. Although the recombinant DNA can be transformed into the host cells through different methods, we use the electroporation method here as an example due to its simplicity and high transformation efficiency ($>10^{10}$ cfu per μg vector DNA).

1. Make agar plates containing 25 g LB (Luria Broth) base, 15 g agar, appropriate antibiotics, 75 μL of 200 mg mL^{-1} IPTG (isopropylthio-β-D-galactoside) and 3 mL of 20 mg mL^{-1} x-gal (5-bromo-4-chloro-3-indoyl-β-D-galactoside) per liter.
2. Transform the ligated DNA into the *E. coli* strain DH5α electrocompetent cells (Life Technologies) by electroporation using the Cell Porator and Voltage Booster System (Life Technologies) at the following settings:

Cell Porator	*Voltage Booster*
Voltage	400 V
Capacitance	330 μF
Impedance	Low ohms
Charge rate	Fast
Resistance	4 kΩ

3. Collect the cells into 1 mL SOC medium and incubate at 37 °C with shaking at 250 rpm for 1 h to allow the cells to recover and fully develop resistance to the antibiotics.
4. Plate the recovered cells onto the agar plates made at the first step. Incubate at 37 °C for 24 h to allow the colony colors to fully develop. The white clones are the recombinant clones. Titer the transformed cells and calculate the number of clones needed for the cDNA library. cDNA libraries are usually stored in bulk as glycerol stocks. For some special research purposes, the cells are plated on the selective agar medium, and individual clones are arrayed into 96- or 384-well microplates containing freezing medium [LB broth plus 1x freezing buffer (36 mM K_2HPO_4, 13.2 mM KH_2PO_4, 1.7 mM citrate, 0.4 mM $MgSO_4$, 6.8 mM $(NH_4)_2SO_4$, 4.4% (v/v) glycerol) and appropriate antibiotics] and maintained in a −80 °C freezer for long-term storage.

3.1.4 Normalization of the cDNA Library

A fundamental problem in cDNA library construction is the occurrence of redundant sequences. This occurs because mRNA titers and their derived cDNA clones vary over several orders of magnitude in a particular tissue. The range of gene expression may be as high as 200 000 mRNA molecules per cell or as low as <1 per cell on average, with perhaps 30% of the genes expressing at levels of less than 10 mRNA molecules per cell at any given time. This implies that several million clones are needed to have a reasonable chance of finding a specific low-abundance transcript. In large-scale cDNA sequencing for these low-abundance classes, strategies are needed to normalize the cDNA populations. Since the goal is to construct a genome-wide cDNA library and generate as many unique ESTs as possible, normalization is required even though it cannot completely remove the members of gene families. There is, as yet, no single optimal strategy available to maximize the relative abundance of different sequences for normalization of a cDNA library. Each of the different strategies has advantages and disadvantages that also depends on the scope and purpose of a project. Currently, three approaches are used in the cDNA or cDNA library normalization.

1. *Screening with abundant sequences:* The screening probes can either be total cDNA derived from mRNA of the same tissue as the library, or the pools of common-sequence clones (10–20 frequent sequence clones).
2. *Reassociation:* The methods of using several variations of reassociation of single-stranded plasmid preparations

and synthesized insert DNAs could remove the redundant copies of the same clones. This procedure has been used in many plant EST programs including the soybean EST project.

3. *Hybridization to total genomic DNA:* Single-stranded biotin-dUTP labeled genomic DNA has been used to bind to streptavidin-coated magnetic beads to select first-strand cDNA according to relative genome representation. Upon elution, conversion to double-stranded cDNA and cloning, libraries enriched for rarer mRNA sequences are generated.

3.2 Small-insert Genomic DNA Libraries

Currently, the cloning efficiency for the systems such as plasmid, bacteriophage, cosmid, and fosmid is extremely high. Only a small amount of DNA is required when constructing a small-insert genomic DNA library using current transformation technology such as electroporation. A small-insert genomic DNA library can be easily and rapidly generated by a single researcher in any molecular laboratory. Small-insert genomic DNA libraries are essential for genome sequencing, fine-scale mapping and genome analysis. These libraries are often screened by directly screening replicas of randomly plated libraries. Physical shearing and enzymatic partial digestion of source DNA are used to generate clonable DNA fragments for library construction. To construct shotgun libraries from genomic DNA by the physical shearing method, genomic or other source DNA is sonicated from 1 to 20 s with a sonicator on ice and end-repaired by the enzyme Mung Bean Nuclease, followed by size-fractionation on conventional agarose gels. The fractions from 0.4 to 1.2 kb and from 2 to 5 kb are independently cloned into plasmid vectors such as Ready-To-Go TM pUC18 Smal1/BAP (Phamarcia Biotech, #27-5266-01). They are referred to as short-fragment and long-fragment shotgun libraries, respectively. The procedure for construction of a small-insert library by partial digestion is similar to that of large-insert library construction, except that the procedure for preparing the DNA fragments is different. The molar ratio of insert : vector for small-insert library ligation is usually 8–15 : 1, while that for large-insert DNA library construction is usually in the range of 1 : 4–10.

3.3 Large-insert Genomic DNA Libraries

Unlike what is required for the construction of cDNA and small-insert genomic DNA libraries, the *in vitro* construction of large-insert genomic DNA libraries (YAC, P1, BAC, PAC, BIBAC, and PBC) requires the manipulation of extremely long DNA fragments. The most difficult aspect of large-insert recombinant construction is the minimization of random shearing of the genomic DNA used for the insert. Electrophoresis analysis of very large fragments of DNA has been made possible by the advent of PFGE systems in which two separate electric fields are generated across the gel field at different angles, alternating with each other for defined periods of time. One type of PFGE system called the contour clamped homogeneous electric field (CHEF) provides a much more uniform electric field, thereby reducing distortion in DNA migration and improving resolution.

Two systems have been developed for the construction of genomic DNA libraries with insert sizes of 100 kbp or larger: yeast- (YAC) and bacteria (BAC, PAC,

BIBAC, and PBC)-based systems. However, the bacteria-based large-insert library construction system has become more prominent in genome research. The details about YAC and YAC library construction are described in Loehrlein and Ronald's chapter on Yeast artificial chromosome (YAC) construction and applications, pp. 291–302, elsewhere in this series. The construction of bacteria-based large-insert genomic DNA libraries is described in the following sections (Fig. 3):

1. Preparation of HMW genomic DNA in LMP agarose plugs
2. Preparation of vectors
3. Partial digestion of HMW DNA embedded in LMP agarose plugs with a restriction enzyme
4. Size selection of DNA fragments on a CHEF gel
5. Ligation of the HMW DNA fragments into the cloning vector
6. Transformation of the HMW recombinant DNA molecules into *E. coli* host cells
7. Characterization and storage of bacteria-based large-insert genomic DNA libraries

3.3.1 Preparation of HMW Genomic DNA in LMP Agarose Plugs

HMW (>600 kbp) DNA is essential for large-insert genomic DNA library construction. Unlike conventional-size (<200 kbp) DNA preparation, HMW DNA must be protected from physical shearing during preparation. One of the popularly used approaches for doing so is to isolate protoplasts (plants and fungi), cells (animals and microbes), or nuclei (plants, animals, and insects), and then embed them in LMP agarose in the form of plugs or microbeads. The cells are lysed and the DNA is purified in the LMP agarose plugs or microbeads. The resulting HMW DNA is carefully manipulated in LMP agarose plugs or microbeads. Since the preparation of HMW DNA from nuclei is straightforward, it results in low contamination with cytoplast DNA, and is economical and suitable for the preparation of HMW DNA from a wide variety of plant and animal species, the nuclei method is widely used for the preparation of HMW DNA in plants and animals. Here we introduce several procedures for preparation of HMW DNA from plants, animals, insects, and fungi.

3.3.1.1 Preparation of HMW nuclear DNA from plants

Plant materials Plant leaves or whole plants of divergent species, including grasses, legumes, vegetables, and trees, can be used as materials for preparation of HMW DNA by this method. The tissues can be either frozen in liquid nitrogen and stored at −80 °C or kept fresh on ice before use.

Reagents

10 x homogenization buffer (HB) stock: 0.1 M Trizma base, 0.8 M KCl, 0.1 M EDTA, 10 mM spermidine, 10 mM spermine, adjust pH from 9.4 to 9.5 with NaOH. The stock is stored at 4 °C.

1 x HB: A suitable amount of sucrose is mixed with a suitable volume of 10 x HB stock. The final concentration of sucrose is 0.5 M and HB stock is 1 x. The resultant 1 x HB is stored at 4 °C. Before use, β-mercaptoethanol is added to 0.15%.

20% Triton X-100 1 x HB: Triton X-100 is mixed with 1 x HB without β-mercaptoethanol to 20%. The solution is stored at 4 °C.

Wash Buffer (1 x HB plus 0.5% Triton X-100): The buffer is prepared by mixing

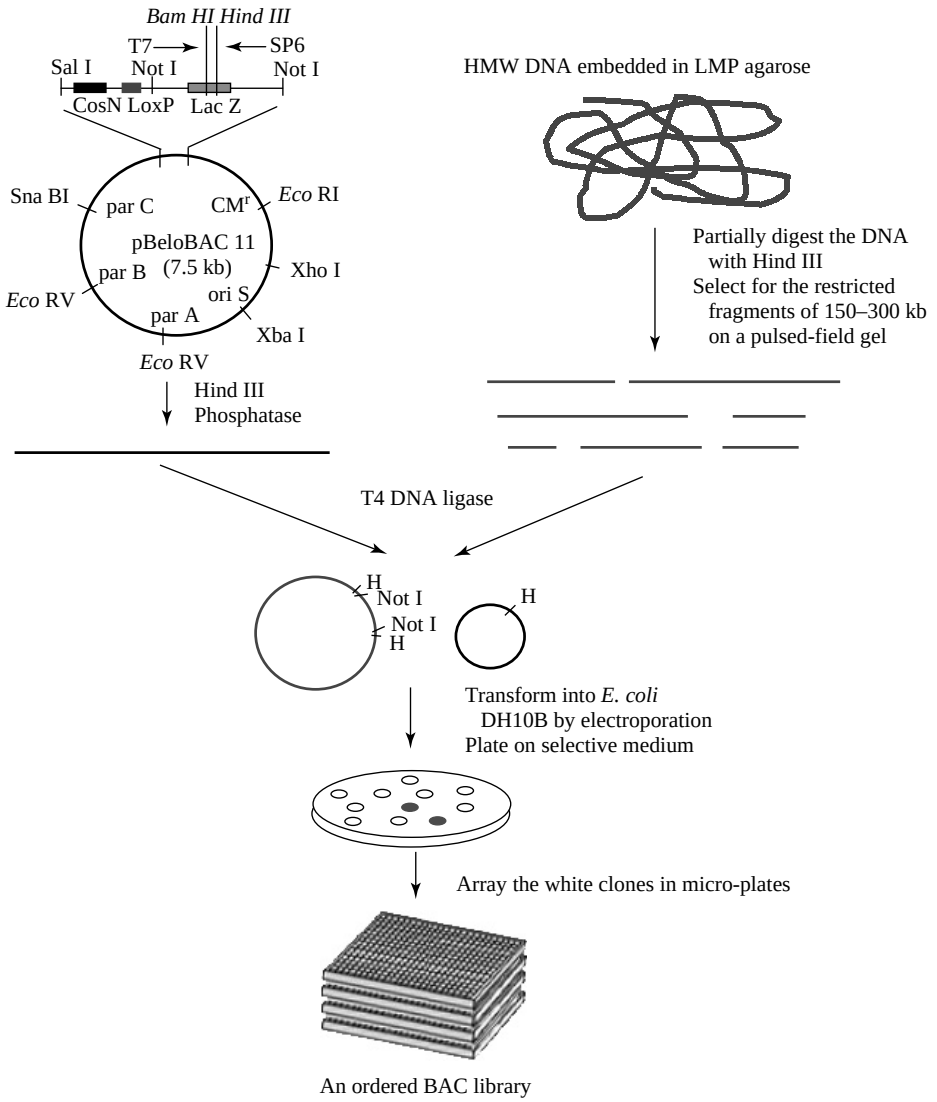

Fig. 3 A general procedure for construction of bacteria-based large-insert, ordered genomic DNA libraries. The bacteria-based large-insert genomic DNA libraries are referred to as BAC, PAC, BIBAC, and PBC libraries. pBeloBAC11 is the cloning vector with two cloning sites, *Bam* HI and *Hind* III (bold). In the procedure, *Hind* III is used for the library construction, and the colorless (white) clones are the clones containing inserts (recombinants).

1 x HB without β-mercaptoethanol with 20% Triton X-l00 in 1 x HB and stored at 4 °C. Before use, β-mercaptoethanol is added to 0.15%.

Lysis buffer: 0.5 M EDTA; pH 9.0 to 9.3, 1% sodium lauryl sarcosine, and 0.1 to 0.5 mg mL^{-1} proteinase K. The proteinase K powder is added just before use.

Procedure

1. Grind about 50 g of the tissue into fine powder in liquid nitrogen with a mortar and pestle, and immediately transfer the powder into an ice-cold 1000-mL beaker containing 800 to 1000 mL ice-cold 1 x HB plus 0.15% β-mercaptoethanol and 0.5% Triton X-100. Gently swirl the contents with a magnetic stir bar for 10 min on ice and filter into six ice-cold 250-mL centrifuge bottles through two layers of cheesecloth and one layer of miracloth by squeezing with gloved hands.
2. Pellet the homogenate by centrifugation with fixed-angle rotor at 1800 g at 4 °C for 20 min, discard the supernatant fluid, and add approximately 1 mL of ice-cold wash buffer to each bottle. Gently resuspend the pellet with the assistance of a small paintbrush soaked in ice-cold wash buffer, combine the resuspended nuclei from all bottles into a 40-mL centrifuge tube and, finally, fill the tube with ice-cold wash buffer. If particulate matter remains in the suspension, filter the resuspended nuclei into the 40-mL centrifuge tube through two layers of miracloth by gravity, centrifuge the contents at 57 g, 4 °C for 2 min to remove intact cells and tissue residues, and transfer the supernatant fluid into a fresh centrifuge tube. Pellet the nuclei by centrifugation at 1800 g, 4 °C for 15 min in a swinging bucket centrifuge.
3. Wash the pellet 1 to 3 additional times by resuspension in wash buffer using a paintbrush followed by centrifugation at 1800 g, 4 °C for 15 min. Note that this step is necessary to minimize the contamination of cytoplast organelles in the nuclei. After the final wash, resuspend the pelleted nuclei in a small amount (about 1 mL) of 1 x HB without β-mercaptoethanol, count the nuclei, if possible, under the contrast phase of a microscope, bring to approximately 5×10^7 nuclei mL^{-1} with addition of the 1 x HB without β-mercaptoethanol, and store on ice. The concentration of nuclei can also be estimated empirically, a concentration of nuclei that is just transparent under light is estimated to be $5–10 \times 10^7$ nuclei mL^{-1}. The concentration of nuclei varies, depending on the genome sizes of different species. In general, 5 to 10 µg DNA of 100-µL plug are suitable for large-insert DNA library construction.
4. Prepare 10 mL of 1% LMP agarose in 1 x HB without β-mercaptoethanol and Triton X-100, cool down to 45 °C and maintain in a 45 °C water bath before use. Prewarm the nuclei to 45 °C in a 45 °C water bath (about 5 min), mix with an equal volume of the prewarmed 1% LMP agarose in 1 x HB without β-mercaptoethanol and Triton X-100 using a cut-off pipette tip. Aliquot the mixture into ice-cold plug molds on ice with the same pipette tip, at 100 µL per plug. When the agarose is completely solidified, transfer the plugs into 5 to 10 volumes of lysis buffer. Incubate the agarose plugs in the lysis buffer for 24 to 48 h at 50 °C with gentle shaking.
5. Wash the plugs once in 0.05 M EDTA, pH 8.0 for 1 h on ice, and store in 0.05 M EDTA, pH 8.0, at 4 °C. The DNA at this step can be stored at 4 °C for 1 year without significant degradation.

3.3.1.2 **Preparation of HMW DNA from animals** Preparation of HMW DNA from animals is much easier and simpler

than the preparation of HMW DNA from plants. Animal cells can be directly embedded into LMP agarose plugs after briefly washing. Although the nuclei of animal cells can be used, the whole cells are frequently used for HMW DNA preparation. Here, the preparation of chicken HMW DNA plugs is used as an example.

Materials Whole chicken blood cells collected in heparinized tubes are used.

Procedure

1. Centrifuge the chicken blood cells at 1000 g, room temperature for 5 min, discard the supernatant and resuspend the cells in the phosphate-buffered saline (PBS) buffer (138 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM NaH_2PO_4, pH 7.4). Repeat this 2 to 3 times and then bring the concentration of cells to $5–10 \times 10^7$ cells mL^{-1}.
2. Prewarm the cell suspensions at 45 °C for 5 min in a water bath, and mix with equal volume of 1% LMP agarose in PBS molten and maintained in a 45 °C water bath. Aliquot into ice-cold 100-μL plug molds with a pipette using a cut tip and keep the molds on ice for 15 min to allow the plugs to completely solidify. Transfer the plugs into the lysis buffer (0.5 M EDTA, pH 9.0, 1% lauryl sarcosine, and 0.5 to 1 mg mL^{-1} proteinase K) at approximately 1 mL lysis buffer per 100-μL plug. Incubate at 50 °C with gentle shaking for 24 to 48 h.
3. Wash the plugs once in 0.05 M EDTA, pH 8.0 for 1 h on ice, and store in 0.05 M EDTA, pH 8.0, at 4 °C. The DNA at this step can be stored at 4 °C for 1 year without significant degradation.

3.3.1.3 **Preparation of HMW DNA from insects** Preparation of HMW DNA from insects is similar to those from plants and animals. Here, the preparation of mosquito and worm HMW DNA are used as examples.

Mosquito

Materials The L1 larvae of mosquito are used as materials. Either fresh or liquid nitrogen-frozen L1 larvae are suitable for preparation of HMW DNA.

Procedure

1. Grind the fresh or frozen L1 larvae into fine powder in liquid nitrogen with a mortar and pestle, and transfer the powder into the suspension solution (100 mM NaCl, 200 mM sucrose and 10 mM EDTA, pH 8.0), or add 10 to 20 mL of the suspension solution into the mortar and incubate at room temperature until the frozen sample and solution just melt. Filter the cell suspensions through 1 to 2 layers of miracloth into a centrifuge tube on ice to remove the large pieces of cell debris and spin at 2000 g, 4 °C for 10 min to collect the cells. Discard the supernatant and resuspend the cells in about 0.5 to 1.0 mL of the suspension solution, depending on the amount of the sample used. From 500 mg fresh weight of mosquito larvae, resuspend the cell pellet in about 1.0 mL of the suspension solution and make about 20 100-μL plugs.
2. Prewarm the cell suspensions at 45 °C for 5 min in a water bath, and mix with an equal volume of molten

1% LMP agarose in the suspension buffer maintained in a 45 °C water bath. Aliquot into ice-cold 100-μL plug molds using a cut pipette tip and keep the molds on ice for 15 min to allow the plugs to completely solidify. Transfer the plugs into the lysis buffer (0.5 M EDTA, pH 9.0, 1% lauryl sarcosine, and 0.5 to 1 mg mL^{-1} proteinase K) at approximately 1 mL lysis buffer per 100-μL plug. Incubate at 50 °C with gentle shaking for 24 to 48 h.

3. Wash the plugs once in 0.05 M EDTA, pH 8.0 for 1 h on ice, and store in 0.05 M EDTA, pH 8.0, at 4 °C. The DNA at this step can be stored at 4 °C for 1 year without significant degradation.

Worms

Materials Depending on species, cells, tissues, or whole body (larvae, pupae, or adults) may be used as materials.

Procedure

1. Grind the fresh or frozen tissues or entire body into fine powder in liquid nitrogen with a mortar and pestle, transfer the powder into ice-cold STE buffer (0.1 M NaCl, 10 mM Tris-HCl, 10 mM EDTA, pH 8.0) with 30 mM β-mercaptoethanol and kept on ice for 10 min with gentle stirring. Filter the cell suspension through 1 to 2 layers of miracloth into a centrifuge tube on ice to remove the large pieces of cell debris and spin at 5000 rpm, 4 °C for 10 min to pellet the cells. Wash the cells in the STE buffer without β-mercaptoethanol once and collect the cells as above. Discard the supernatant and resuspend the cells in 1 to 2 mL STE buffer (without β-mercaptoethanol) per gram of the tissues or whole bodies. This allows the making of 20 to 40 of the 100-μL LMP plugs.
2. Prewarm the cell suspensions at 45 °C for 5 min in a water bath, and mix with an equal volume of molten 1% LMP agarose in the STE buffer maintained in a 45 °C water bath. Aliquot into ice-cold 100-μL plug molds using a cut pipette tip and keep the molds on ice for 15 min to allow the plugs to completely solidify. Transfer the plugs into the lysis buffer (0.5 M EDTA, pH 9.0, 1% lauryl sarcosine, and 0.5 to 1 mg mL^{-1} proteinase K) at approximately 1 mL lysis buffer per 100-μL plug. Incubate at 50 °C with gentle shaking for 24 to 48 h.
3. Wash the plugs once in 0.05 M EDTA, pH 8.0 for 1 h on ice, and store in 0.05 M EDTA, pH 8.0, at 4 °C. The DNA at this step can be stored at 4 °C for 1 year without significant degradation.

3.3.1.4 **Preparation of HMW DNA from fungi** Preparation of HMW DNA from fungi is similar to that from plant protoplasts. Fungal spheroplasts can be generated by treatment of cultured fungal cells with Novozym and then washed briefly.

Materials Cultured cells are used as materials.

Procedure

1. Collect the fungal cell culture by centrifugation at 1000 g, room temperature for 4 min. Wash the cells in 10 mL 1 M sorbitol per gram of fungi, resuspend in 10 mL SCE (1 M sorbitol, 20 mM sodium citrate pH 5.8,

10 mM EDTA) and add Novozym to 10 mg/mg cells. The mixture is incubated with shaking at 100 rpm for 90 min at room temperature. Harvest the spheroplasts by filtering successively through cheesecloth and Nytex nylon mesh that has a 25-μ pore size, followed by pelleting at 2000 g, 4 °C for 10 min. Wash the spheroplasts twice with 1 M sorbitol and then resuspend in 1 M sorbitol at about 3×10^9 cells mL^{-1} to obtain about 6 μg DNA in a 100-μL plug.

2. Prewarm the cell suspensions at 45 °C for 5 min in a water bath, and mix with an equal volume of molten 1% LMP agarose in the SCE buffer maintained in a 45 °C water bath. Aliquot into ice-cold 100-μL plug molds using a cut pipette tip and keep the molds on ice for 15 min to allow the plugs to completely solidify. Transfer the plugs into the lysis buffer (0.5 M EDTA, pH 9.0, 1% lauryl sarcosine, and 0.5–1 mg mL^{-1} proteinase K) at approximately 1 mL lysis buffer per 100-μL plug. Incubate at 50 °C with gentle shaking for 24 to 48 h.
3. Wash the plugs once in 0.05 M EDTA, pH 8.0 for 1 h on ice, and store in 0.05 M EDTA, pH8.0, at 4 °C. The DNA at this step can be stored at 4 °C for 1 year without significant degradation.

3.3.2 Preparation of Vectors

Several plasmid-based vectors, including BAC, PAC, BIBAC, and PBC, have been widely used for bacteria-based large-insert genomic DNA library construction (Table 1). Preparation of the vectors is one of the most critical steps for successful construction of large-insert DNA libraries. The vector DNA to be used for library construction must be very pure, completely digested (but not overdigested) and completely dephosphorylated (>95% recombinant clones in a ligation test). The vector DNA to be used for library construction is isolated and purified from 1 to 2 L overnight culture by the alkaline lysis method, followed by two rounds of cesium chloride/ethidium bromide gradient ultra-centrifugations. Below is a widely used procedure for the preparation of vectors for library construction, including cDNA, small-insert, and large-insert genomic DNA libraries.

1. Completely digest 20 μg of the vector DNA using an appropriate restriction enzyme (here, a 5′-overhang restriction enzyme is used as an example) and check the digestion on a 1% agarose gel to ensure that the digestion is complete. Note that overdigestion may lead to an increase of the percentage of clones with no inserts (insert-empty clones).
2. Extract the digest with an equal volume of saturated phenol/chloroform (1 : 1), and spin at 10 000 rpm, room temperature for 5 min. Precipitate the DNA by adding 1/10 volume of 3 M NaAC, pH 5.2 and 2 volumes of 100% ethanol and incubating at −80 °C for 10 min, followed by centrifugation at 10 000 rpm for 15 min. Discard the supernatant fluid, wash the pellet carefully with 70% ethanol, air-dry, dissolve in 100 μL H_2O, and measure the concentration of the DNA using λDNA of known concentration on a 1% agarose gel.
3. Dephosphorylate the linearized vector with calf intestinal alkaline phosphatase (CIAP) at 37 °C for 30 min, at a ratio of 1.0 unit/pmole (3.89 units of CIAP/10 μg DNA for pBeloBAC11, 1.23 units of CIAP/10 μg DNA for BIBAC2, or 1.0 units of CIAP/10 μg

DNA for pCLD04541). Stop the reaction immediately by adding 4 μL 0.5 M EDTA, pH 8.0, 20 μL 10% SDS and 40 μL of 1 mg mL^{-1} proteinase K in cold TE to the tube, followed by incubation at 56 °C for 30 min.

4. Cool down to room temperature, extract once with an equal volume of saturated phenol and then once with an equal volume of saturated phenol/chloroform/iso-amyl alcohol (25 : 24 : 1). Precipitate the DNA by adding 1/10 volume of 3 M NaAC, pH 7.0 and 2 volumes of 100% ethanol and incubating at −80 °C for 10 min, followed by centrifugation at 10 000 rpm for 15 min.
5. Discard the supernatant fluid, wash the pellet carefully with 70% ethanol, air-dry, dissolve in 200 μL H_2O, and measure the concentration of the DNA using λDNA of known concentration on a 1% agarose gel. Adjust the concentration of the DNA to 10 ng/μL for pBeloBAC11, and to 40 ng μL^{-1} for BIBAC2 or pCLD04541, aliquot the DNA and store in a −20 °C freezer. The vector DNA in the −20 °C freezer is good for cloning for at least three months. To ensure that the vector is properly dephosphorylated, ligate the vector DNA into the λDNA digested with the same restriction enzyme. When a molar rate of 4 vector : 1 digested λDNA is used and the ligation mixture is transformed into *E. coli* DH10 cells by electroporation, the percentage of recombinant (white) clones should be greater than 95%.

3.3.3 Partial Digestion of HMW DNA Embedded in LMP Agarose Plugs with a Restriction Enzyme

Before use, the LMP agarose plugs containing HMW DNA should be washed 3 times for 1 h each in 10 to 20 volumes of ice-cold TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0) plus 0.1 mM phenylmethyl sulfonyl fluoride (PMSF), and 3 times for 1 h each in 10 to 20 volumes of ice-cold TE on ice. As described above, several methods have been used to partially digest the HMW DNA embedded in LMP agarose plugs with a restriction enzyme. Described below is the partial digestion strategy of varying the concentration of the enzyme used at a fixed digestion time. A series of small pilot digestions is commonly carried out to determine the optimal concentration range of the enzyme required to generate DNA fragments desirable for library construction.

1. Slice each LMP agarose plug containing HMW DNA into nine equal-sized pieces per 100-μL plug, and equilibrate the pieces in the restriction-enzyme reaction buffer containing 2 mM spermidine and 1 mM dithiothreitol (DTT) at 3 mL 100-μL^{-1} plug, on ice for 1 h, changing the buffer once. Then, transfer three plug slices each into 1.5-mL microtubes containing 170 μL reaction buffer with 0.5 mg ml^{-1} BSA, 2 mM spermidine, 1 mM DTT, and a concentration series of the enzyme (0–20 units/tube). Incubate on ice for 60 to 100 min, depending upon the compactness of the HMW DNA plugs, to allow the enzyme access to the DNA in the agarose.
2. Transfer the digestion reactions into a 37 °C water bath and incubate for exactly 8 min. Immediately transfer the reactions onto ice and then stop them by adding 1/10 volume of 0.5 M EDTA, pH 8.0. Keep the reactions on ice.
3. Analyze the partial digestions by loading the plug slices on a 1% agarose

gel in 0.5 x TBE and subjecting to the PFGE using the CHEF DRIII under the following conditions: 11 °C (setting for the cooler), 80 (pump setting), 120° angle, 6 V cm^{-1} and 50-s switch time for 18 h. The partial digestion condition that generates restricted fragments in a size range of 100 to 400 kbp is selected for large-scale partial digestion for the construction of DNA libraries with an average insert size of about 150 kbp.

4. Perform large-scale partial digestion (usually 6–10 100-μL plugs) using the optimal digestion conditions determined above.

3.3.4 Size Selection of DNA Fragments on a CHEF Gel

After partial digestion, the DNA fragments must be selected to obtain the fragments that are desirable in size range for library construction and to remove the smaller fragments that can compete more effectively for vector ends. To this end, PFGE has been widely used for size selection; almost all YAC and bacteria-based large-insert DNA libraries constructed to date have used this method for size selection. Furthermore, to maximize the removal of the small DNA fragments, many scientists perform two rounds of size selection by PFGE.

1. *First size selection*: Prepare a 1% agarose CHEF gel in 0.5 x TBE with a sample trough at the top portion of the gel and 2 L of 0.5 x TBE running buffer, and precool the buffer to 11 °C in the electrophoresis chamber. Load the plug slices containing the partially digested DNA into the sample trough, and size markers (lambda concatemers and/or yeast chromosomes) in the wells on both sides of the trough. Seal the sample trough and marker wells with 1% molten agarose gel. Run the CHEF gel using the CHEF DRIII system at 6.0 V cm^{-1}, 11 °C, 120° angle, and 90-s switch time in 0.5 x TBE for 16 h. Cut both the sides of the gel that contain the size markers and the edge of the DNA sample with a clean microscope glass slide coverslip and keep the center of the gel on ice. Stain and photograph the gel pieces containing the samples adjacent to the size markers alongside a scale ruler. Using the photograph as reference, locate the gel zone containing the DNA fragments ranging from 100 to 400 kbp in the unstained portion of the gel, excise it, and cut it into three sections horizontally, 100 to 200 kb, 200 to 300 kb and 300 to 400 kb. Recover the size-selected DNA in each gel section by electroelution in a dialysis tube (1/4 in in diameter and molecular weight exclusion limit = 12 000–14 000 Da) in 0.5 x TBE using the CHEF DRIII at 6 V cm^{-1}, 11 °C, and 35-s switch time for 4 h, followed by reversing the polarity of the current for 60 s. The DNA fragments can be used for ligation and library construction after dialysis in 0.5 x TE (see below), or are carefully transferred from the dialysis tubing to a microfuge tube with a wide-bore pipette tip for the second size selection.
2. *Second size selection*: Prepare a 1% agarose CHEF gel in 0.5 x TBE with a sample trough at the top portion of the gel and 2 L of 0.5 x TBE running buffer, and precool the buffer to 11 °C in the electrophoresis chamber. Load size markers on both sides of the sample trough, seal the marker wells with 1% molten agarose gel, and place in the electrophoresis chamber containing the 0.5 x TBE buffer. Mix

the DNA fragments from the first size selection with the gel-loading dye by inverting the tube gently, and load into the sample trough. Run the CHEF gel at 4 V cm^{-1}, 11 °C and 5-s switch time in 0.5 x TBE for 6 to 8 h. Locate and excise the gel zone containing the compressed DNA fragments, and electroelute the DNA from the gel as above.

3. Dialyze the DNA fragments electroeluted from the gel in the same dialysis tubing against 1 L of 0.5 x TE for 3 h at 4 °C with one change of 0.5 x TE every hour. Carefully collect the DNA from the dialysis tubing to a 1.5-mL microtube with a wide-bore pipette tip for ligation and library construction. The immediate ligation of the DNA into a vector following dialysis often gives a much better result.

3.3.5 Ligation of the HMW DNA Fragments into the Cloning Vector

The concentration of the DNA collected from the second size selection is routinely 1 to 8 ng μL^{-1}. The concentration of the DNA needs to be estimated before preparing the ligation reactions.

1. Estimate the concentration of the insert DNA on a 1% conventional agarose minigel using the known concentration of λDNA as standard.
2. Set up the ligation using the following criteria: Molar ratio of vector : insert = 4–10 : 1, the final concentration of insert DNA in the ligation mixture = 1–2 ng μL^{-1} and T4 DNA ligase = 1 unit per 50 μL ligation. To control the ratio between recombinant and nonrecombinant clones in the transformants, the molar ratio of vector:insert can be adjusted, with a smaller ratio for a higher recombinant/nonrecombinant clone ratio.
3. Incubate the ligation at 16 °C for 6 to 10 h. The ligation can immediately be used for transformation, or aliquoted and stored in a −20 °C, non frost-free freezer. The ligation can be stored for several months without significant problems. Although it was observed that the storage of the ligated DNA might slightly reduce the transformation efficiency, it did not significantly reduce the insert sizes of the clones.

3.3.6 Transformation of the HMW Recombinant DNA Molecules into *E. coli* Host Cells

Although recombinant DNA can be transformed into host cells through several methods, almost all existing bacteria-based large-insert genomic DNA libraries, including BAC, PAC, BIBAC, and PBC, were constructed using the electroporation method due to its extremely high transformation efficiency. Similarly, almost all existing bacteria-based large-insert genomic DNA libraries are hosted in *E. coli* strain DH10B and its derivatives. This is because it has been proven that the large-insert clones can be stably maintained and propagated in these strains; other strains may not be capable of stably maintaining and propagating large-insert genomic DNA clones in their cells.

1. Mix 1 to 2 μL of the ligation mixture with 20 μL of *E. coli* DH10B electroporation-competent cells and transform into the cells using the Cell Porator and Voltage Booster electroporation system (Life Technology) at the following settings:

Cell Porator	*Voltage Booster*
Voltage	350–400 V
Capacitance	330 μF
Impedance	Low ohms

Charge rate	Fast
Resistance	4 kΩ

2. Transfer the electroporated cells in 1 ml of SOC medium and incubate at 37 °C, 250 rpm for 1 h to allow the antibiotics-resistant gene to be fully expressed, and then plate onto LB agar medium containing appropriate antibiotics and/or amendments for transformant and recombinant selection. For the vectors, such as pBeloBAC 11 and pCLD04541, that use *Lac*Z gene for recombinant selection, incubate at 37 °C for 36 h to allow the colony color (blue) to fully develop, whereas for the vectors, such as BIBAC2, that use the *Sac*B gene for recombinant selection, incubate at 37 °C for 20 h. The colorless (white) colonies constructed with the *Lac*Z gene-containing vectors are expected to contain inserts (recombinants), whereas all of the clones of the *Sac*B gene-containing vectors growing on the selective medium are expected to contain inserts.

3.3.7 Characterization and Storage of Bacteria-based Large-insert Genomic DNA Libraries

Analysis of the clones resulting from a particular ligation is essential before large-scale transformation and library assembly are undertaken. The parameters usually used for this purpose include average insert size and distribution, and the percentage of empty clones. In general, a large-insert genomic DNA library such as is desirable for genome research should have a large average insert size (≥150 kbp), a low percentage of insert-empty clones (≤5%) and a reasonable distribution of clones in insert sizes (providing a better genome coverage). Therefore, at least 100 randomly selected clones obtained from each ligation should be analyzed.

1. Inoculate the recombinant clones on the selective medium in 5 mL of LB medium with appropriate antibiotics and incubate at 37 °C with shaking at 250 rpm overnight. Isolate DNA using the alkaline lysis method, digest with *Not* I, a rare cutting enzyme, and run on a 1% pulsed-field gel in 0.5 x TBE at 6 V cm^{-1}, 11 °C, 5-s initial switch time and 15-s final switch time with a linear ramp, for 16 h. Stain the gel with ethidium bromide, photograph and estimate the insert size of each clone by adding the sizes of all DNA fragments derived from the clone insert and the number of clones with no inserts.
2. Select the ligation that gives the largest insert sizes, fewer insert-empty clones, and higher transformation efficiency for large-scale transformation and library assembly. Array the recombinant clones into 384-well microtiter plates containing 50 μL of freezing medium (see above) per well with appropriate antibiotics, grow at 37 °C overnight and store at −80 °C for long-term maintenance. Usually, the library is replicated and stored in separate −80 °C freezers to guard against unforeseen disasters.

4 Applications of DNA Libraries

DNA libraries are indispensable and crucial tools for molecular biology research. They have been used in every aspect of molecular biology. Summarized below are a few examples of their applications.

4.1 Gene Discovery, Expression, and Function Studies

cDNA libraries have been essential resources for gene discovery, expression monitoring, and function analysis. cDNA sequence tags, popularly known as ESTs, can be used not only for the analysis of gene structure, expression, and function but also for the study of genome composition, organization, and structure. Projects for developing ESTs for a large number of genes have emerged in the past few years as one of the approaches to rapid gene discovery. In this approach, a large number of cloned cDNAs are randomly chosen from the cDNA libraries constructed from different tissues or treatments, and sequenced from one or both ends of the clone inserts. The nucleotide sequences and translations of the ESTs are used as tags of the corresponding genes to compare with the sequences of known functions available in the public databases for identification of homologs and to provide useful information for gene identification.

Currently, a few genomes of model plant and animal species, such as *Arabidopsis*, *Drosophila*, rice, mosquito, and human have been sequenced. The availability of these sequences provides us with a detailed view of the gene content and genome organization of the species. Yet, the degree to which the gene content, gene number, and genome organization are conserved among related species remains unresolved. It will require the sequencing of genomes from many other species to answer these questions. However, because there are so many species with relatively larger genome sizes compared with those of model genomes, and also considering the associated high cost of sequencing, it is unlikely that we will have the full genomic sequences for many other species in the near future. Large-scale EST sequencing thus provides a less expensive alternative approach, which can reveal a substantial portion of the expressed genes of a genome at a fraction of the cost of a genomic sequencing project. Therefore, extensive EST projects have been performed or are under way in a wide variety of animal and plant species. With all of these EST and genomic sequencing projects, it is anticipated that once most genes are identified and their clones become available, the process of gene identification and characterization will be accelerated several-fold.

There are tens of thousands of genes and their products (e.g. RNA and proteins) in a given organism that function in a complicated and orchestrated way to create the mystery of life. However, classic procedures in molecular biology generally work in a "one gene one experiment" fashion, which means that the throughput is very limited and the "whole picture" of gene function is difficult to obtain. In the past few years, a new technology, now known as DNA microarrays or DNA chips, has been invented, in which a large number of DNA clones (e.g. cDNA clones or synthesized oligonucleotides) are arrayed by high-speed robotics on glass slides or nylon substrates, and fluorescently labeled target nucleic acid samples are used to hybridize the DNA microarray, allowing high-throughput study of gene expression, gene function, and gene identity. This technique also promises to monitor genes at the whole-genome level even on a single chip so that researchers can have a better picture of the interactions among thousands of genes simultaneously.

4.2 Gene Mapping and Cloning

Depending on the type and amount of information available for target genes, DNA libraries can be used for gene mapping and cloning in different ways. When the information regarding the function of a target gene and its protein product is available, a method called functional cloning could be used. Knowledge of gene function assists investigators to develop strategies to purify the gene product and link it with a phenotype, develop antibodies, derive partial amino acid sequence, and then identify cDNA clones containing the target gene by screening an expression cDNA library using antibodies or degenerate oligonucleotides derived from the protein sequence. When a cDNA clone is identified, isolated, and sequenced for verification that it corresponds to the desired gene, it is then used as a probe to screen a genomic library and thereby to isolate its cognate gene. The cDNA clone could also be used for studying the gene expression in different cell types, tissues, and developmental stages.

The candidate gene method is another approach to isolating a gene of interest from a DNA library based on the existing information about the gene. This method is based on a survey of previously identified genes that seem to perform the same function altered in a mutant. Clone(s) for the candidate gene can be isolated from a DNA library by screening the library using the candidate gene previously identified from other species as a probe. The genetic position of the candidate gene and mutation analysis can be further used to confirm the candidacy of the gene.

On the other hand, there is no available prior knowledge of RNA and protein products for many genes. These genes are known only by their phenotypes. To clone such genes, an approach named *positional cloning* or *map-based cloning* has been developed. By using a genetic cross, a phenotype can be assigned to a map interval by identifying two flanking genetic markers. Once the chromosomal localization of the gene locus has been established with DNA markers, genomic DNA libraries, particularly large-insert DNA libraries, for example, BAC or YAC libraries, are screened with the DNA markers to identify overlapping genomic DNA clones covering the marker locus. The clones are then analyzed, assembled into an overlapping clone contig and mapped against the target gene. If the contig does not cover the locus of the target gene, new probes are developed from the ends of the contig and used to rescreen the library. The new clones are used to extend the contig and are mapped against the target gene. Using this strategy of chromosome walking, the clone(s) containing the target gene can be isolated. Once a clone containing the target gene is identified, it can be transformed into a mutant of the gene. The genetic complementation of the clone to the mutant phenotype will indicate that the clone contains the target gene.

The completion of major sequencing efforts, such as in *Arabidopsis* and human, promises to alter the classical paradigm by providing the identity and syntenic relationships between the genes. These new tools and knowledge will further accelerate the process of novel gene identification.

4.3 Genome Physical Mapping

DNA libraries, especially large-insert, ordered DNA libraries, are essential sources

for genome physical mapping. Clone (P1, BAC, BIBAC, PAC, PBC, and YAC)-based physical maps are the centerpiece of genomics research. They are essential not only for an efficient and thorough approach to large-scale genome sequencing but also for performing efficient map-based cloning and associating candidate genes with biological, medical, or agronomic traits. Several methods have been developed and used to construct genome-wide physical maps from large-insert genomic DNA libraries. These include DNA marker-based chromosome landing, random DNA clone fingerprinting and contig assembly, and BAC end sequencing and fingerprint analysis.

DNA marker-based chromosome landing is a hybridization- and/or PCR-based strategy for genome physical mapping, depending on the types of DNA markers to be used for landing the large-insert clones of the source library to the target linkage map (see below). DNA markers are selected from a molecular genetic linkage map and used as probes or PCR primers to screen the large-insert source library. Overlapping clones are assembled into contigs according to the probes to which they hybridize, or the primer pairs with which they produce PCR products. The contigs are extended and anchored to the genetic linkage map according to the position of the DNA markers on the genetic map. This physical mapping strategy was first used in the development of YAC-based physical maps of the human, mouse, rice and *Arabidopsis thaliana* genomes, and recently, in the development of the BAC-based physical map of the euchromatic portion of the *Drosophila* genome.

If the markers are PCR-based, such as sequence-tagged site (STS) and simple sequence repeats (SSRs), the large-insert clones of the source library are anchored to the genetic map by screening the source library with PCR, in which the library is fabricated into DNA pools, including multiple-plate superpools, single-plate pools, column pools and row pools. Alternatively, the synthetic oligos or PCR products of the STSs and SSRs can also be used as hybridization probes to land large-insert clones to the target linkage map by the hybridization-based strategy. The sequences of the STSs and SSRs are helpful for anchoring sequence assemblies to chromosomes (linkage groups) through the bioinformatics approach. However, if the DNA markers are hybridization-based, such as RFLPs (restriction fragment length polymorphisms) and synthetic Overgos, the large-insert clones of the source library are anchored to the genetic linkage map by screening the source library with hybridization probes. In this method, the library is robotically gridded onto high-density colony filters.

A significant advantage of this physical mapping strategy is the development and integration of the physical map with linkage genetic maps in the same process. The integration of the physical map with a genetic map will associate the large-insert clones of the map with useful traits, which will greatly expedite isolation of the genes by positional cloning. However, it is a prerequisite to have a high-density, evenly distributed DNA marker genetic map and a genomic DNA library of very large inserts in order to develop a whole-genome physical map with this strategy. First of all, although high-density DNA marker maps have been developed for the genomes of human and several model species, they are not available for many of the plant and animal species of economic importance. Second, bacteria-based large-insert clones have much smaller insert sizes than YACs. The smaller insert sizes

of the bacteria-based large-insert clones require highly saturated DNA marker maps for the development of a BAC-based physical map using the DNA marker-based chromosome landing strategy.

Random DNA clone fingerprinting and contig assembly is a restriction fragment fingerprint-based strategy for genome physical mapping. This strategy is based on the hypothesis that the clones derived from the same regions of a genome share similar restriction patterns (a large number of common bands). The restriction pattern of a clone is designated as the "fingerprint" of the clone. DNAs are purified from random clones of the source genomic DNA library and digested with one or more restriction enzymes. The restricted fragments are either directly fractionated on gel matrices, or end-labeled and then fractionated on gel matrices to generate fingerprints for each clone. The fingerprints of the clones are visualized and digitized using computers. Overlapping clones are assembled into contigs according to the similarity of the restriction patterns or fingerprints of the clones. The advent of BACs, PACs, BIBACs, and PBCs has facilitated application of the clone fingerprint analysis strategy in the development of BAC-based physical maps of large complex genomes. So far, whole-genome BAC and/or BIBAC-based physical maps have been developed for *Arabidopsis, Drosophila*, rice, human, and mouse.

BAC end sequencing and fingerprint analysis is a combined BAC end sequencing and restriction fingerprinting strategy for genome physical mapping. This strategy has been used in the sequencing projects of *Arabidopsis, Drosophila*, and human to select clones as substrates for sequencing. In this strategy, genome sequencing and physical mapping are performed simultaneously. Both ends of BACs equivalent to about 15 x haploid genomes are first sequenced and entered into computers as a database. Then, seed BACs are randomly selected from the source BAC library, sequenced, and entered into the BAC end database. The sequences of the seed BACs are analyzed against the BAC end sequence database, and the BACs whose ends share common sequences with the seed BACs are selected. Finally, the BACs that are potentially overlapping the seed BACs are fingerprinted and assembled into contigs. According to the fingerprint contigs and BAC end sequences, the minimally overlapping BACs are selected as substrates for continuous sequencing. In this strategy, although no preexisting sequence-ready, BAC-based physical map is required for genome sequencing, a similar amount of effort is needed to develop the physical map as in the genome sequencing progresses. Nevertheless, the global BAC-based physical map would not be available for other applications in genomics research until the genome is completely sequenced. This would be a great loss, considering the utility of the physical map in genomics research.

4.4 Genome Sequencing

Many strategies have been developed and used in large-scale genome sequencing in the past decade; however, shotgun sequencing has become the most dominant strategy. Two alternative approaches, the whole-genome shotgun and the clone-by-clone shotgun, have been developed and used for sequencing large, complex genomes. The whole-genome shotgun approach often involves the following steps: high-quality genomic DNA is purified and randomly sheared; the sheared fragment

ends are enzymatically repaired, size-selected for size-appropriate fragments (e.g. 0.4–2 kbp and 5–20 kbp), and cloned into plasmid-based vectors (e.g. pUC vectors), respectively; template DNA is purified from randomly selected clones, and sequences are generated from both ends of each clone; finally, sequencing data is edited and assembled into contigs with computers. In this approach, the short insert (0.4–2 kbp) clones are used for draft sequence production and contig assembly. The end sequences of the relatively long insert (5–20 kbp) clones are used to link sequence contigs into long-range sequence "scaffolds" and to span long-range repetitive sequences. The "contigs" referred to here are groups of overlapping sequence reads or assemblies without any interruptions, and the "scaffolds" referred to are collections of two or more sequence contigs linked by the relatively large-insert clones. This approach has been used in the sequencing of the human and rice genomes. In the clone-by-clone shotgun approach, HMW genomic DNA is partially digested, size-selected and constructed into large-insert genomic DNA libraries, such as BAC and PAC libraries. A PAC- or BAC-based integrated genetic and physical map of the genome is constructed by using one or more of the physical mapping strategies described above. Minimally overlapping (also referred to as "tiling path") clones that completely cover the entire genome are selected from the physical map and used as substrates for shotgun sequencing, one clone at a time. This approach has been used in the sequencing of the *Arabidopsis* and human genomes. Although both approaches have been used in different large-scale sequencing projects of large, complex genomes, it is obvious that genomic DNA libraries including large-insert DNA libraries are indispensable for both the methods.

4.5 Gene Engineering and Gene Therapy

One of the many applications of DNA libraries is genetic engineering and gene therapy. Genetic engineering and gene therapy have the same basic premise: introduction of DNA clones containing specific new genetic material into an organism's genome. There are many variations of the technology to achieve this goal, depending upon the organism and situation. The basic strategy is to place the new genetic material (e.g. a gene for a new desired function) into the cell, and use the cell's natural machinery to integrate this material into the genome. There are some successful cases for both animal and plant gene engineering.

Animal eggs or embryonal stem (ES) cells can be directly microinjected with gene expression DNA clones or fragments to create a genetically engineered animal. For example, to produce human β-interferon on a large scale, transgenic sheep or cows are used for producing a large amount of casein protein secreted into milk by the mammary gland. The human β-interferon gene has been attached to casein synthesis and secretion signals. High levels of the protein (e.g. 10 g per liter of milk) can be expressed in some animals.

Transgenic plants can be created by the introduction of genes via infection by *Agrobacterium tumefaciens*, a bacterium that can transfer DNA of its Ti plasmid into plant cells, where it integrates into the host plant genome. To date, the main goals of plant engineering are to induce tolerance to herbicides and resistance to insect pests, and to improve crop quality. For example,

a gene from the *Agrobacterium* codes for phosphinothricin-N-acetyltransferase (*PAT*) that inactivates the herbicide Basta. The *PAT* gene is inserted into the toxin gene of the Ti plasmid, transferred into plants via the *Agrobacterium*, and then integrated into the plant genome. When plants are engineered to contain this gene they will have resistance to the glufosinate ammonium herbicides. This allows selective killing of weeds by herbicides.

Gene therapy also involves the introduction of a new gene or the modification of an existing gene (itself or its activity) in cells. Germ-line cell therapy is not allowed in humans; however, somatic cell therapy has been used to treat cystic fibrosis, severe combined immune deficiency (SCID), tumors and other diseases. This kind of gene therapy can be applied to the whole body (*in vivo* therapy) or to the cells removed from a patient (*ex vivo* therapy). In the latter, the engineered cells are returned to the patient.

Recombinant retroviruses or adenoviruses can be used as DNA vectors for gene therapy. Retroviruses can accept up to 7 kbp of introduced DNA and infect only dividing cells. Because of this, they are used in *ex vivo* therapies. Since they integrate into the host genome, their effect is long lasting, but insertional mutagenesis is a potential problem. Retroviruses are used in the treatment of SCID. In contrast, adenoviruses can accommodate larger genes and do not integrate into the host genome. Their effect is transient. The main problem with adenoviruses is the immune response they elicit (used in the treatment of cystic fibrosis and a 1-antitrypsin deficiency). In another approach, recombinant gene clones are encapsulated in liposomes (lipid micels). This method produces no side effects but is less efficient in transferring DNA to target cells. It does not cause an immune reaction. Liposomes can be used *in vivo* and *ex vivo*, and carry any size of DNA fragments. The first gene therapy for a human disease was successfully achieved for SCID by introducing the missing gene of Adenosine Deaminase (ADA) into the peripheral lymphocytes of a 4-year-old girl and returning the modified lymphocytes to her system (1990). As of January 2000, more than 350 gene therapy protocols have been approved in the United States.

5 Concluding Remarks

Technologies for development of DNA libraries are evolving extremely rapidly. Different types of DNA libraries have been developed with different DNA sources, functions, and cloning vectors. The advent of large-insert DNA libraries such as YAC, BAC, PAC, BIBAC, and PBC has revolutionized chromosome walking, whole-genome physical mapping and sequencing of large, complex genomes. Further dramatic improvements in the technology of DNA library construction are to be expected. The applications and potential uses of various DNA clones containing all kinds of genes have made it clear that DNA libraries continue to be indispensable genetic resources and tools for every aspect of molecular research from molecular biology, molecular genetics and genomics to human gene therapy.

See also Body Expression Map of Human Genome; Genetics, Molecular Basis of; Genomic DNA Libraries, Construction and Applications.

Bibliography

Books and Reviews

Bolivar, F., Backman, K. (1979) Plasmids of *Escherichia coli* as cloning vectors, *Methods Enzymol.* **68**, 245–267.

Hieter, P., Connelly, C., Shero, J., McCormick, M.K., Antonarakis, S., Pavan, W., Reeves, R. (1990) Yeast Artificial Chromosomes: Promises Kept and Pending, in: Davies, K.E., Tilghman, S.M. (Eds.) *Genome Analysis: Genetic and Physical Mapping*, Vol. 1, Cold Spring harbor laboratory Press, Cold Spring Harbor, New York, pp. 83–120.

Hohn, B. (1979) *In vitro* packaging of λ and cosmid DNA, *Methods Enzymol.* **68**, 299–309.

Ioannou, P.A., de Jong, P.J. (1994) Construction of Bacterial Artificial Chromosome Libraries Using the Modified P1 (PAC) System, in: Dracopoli, N.C., Haines, J.L., Korf, B.R., Moir, D.T., Morton, C.C., Seidman, C.E., Seidman, J.G., Smith, D.R. (Eds.) *Current Protocols in Human Genetics*, John Wiley & Sons, New York, pp. 5.15.1–5.15.24.

Kenan, D., Keene, J., Henderson, B. (1999) DNA Libraries, in: Thomas, E.C. (Ed.) *Encyclopedia of Molecular Biology*, John Wiley, New York, pp. 711–712.

Zhang, H.-B. (2000) Construction and Manipulation of Large-Insert Bacterial Clone Libraries – Manual, *Texas A&M University, TX, USA* (http://hbz.tamu.edu).

Zhang, H.-B., Wu, C. (2001) BAC as tool for genome sequencing, *Plant Physiol. Biochem.* **39**, 195–209.

Zhang, H.-B., Woo, S.-S., Wing, R.A. (1996) BAC, YAC and Cosmid Library Construction, in: Foster, G.D., Twell, D. (Eds.) *Plant Gene Isolation: Principles and Practice*, John Wiley & Sons, New York, pp. 75–99.

Primary Literature

Blaese, R.M., Anderson, W.F., Culver, K.W., Rosenberg, S.A. (1990) Treatment of severe combined immune deficiency disease (SCID) due to adenosine deaminase (ADA) deficiency with autologous lymphocytes transduced with human ADA gene, *Protocol #2, Hum. Gene Ther.* **1**, 327–362.

Blattner, F.R., Williams, B.G., Blechl, A.E., et al. (1977) Charon phages: Safer derivatives of bacteriophage λ for DNA cloning, *Science* **196**, 161–169.

Burke, D.T., Carle, G.F., Olson, M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors, *Science* **236**, 806–812.

Cohen, S.N., Chang, A.C.Y., Boyer, H.W., Helling, R.B. (1973) Construction of biologically functional bacterial plasmids *in vitro*, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3240–3244.

Coulson, A., Sulston, J., Brenner, S., Karn, J. (1986) Toward a physical map of the genome of the nematode *C. elegans*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7821–7825.

Grant, S.G.N., Jessee, J., Bloom, F.R., Hanahan, D. (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1213–1217.

Hamilton, C.M. (1997) A binary-BAC system for plant transformation with high-molecular-weight DNA, *Gene* **200**, 107–116.

Hamilton, C., Frarey, A., Lewis, C., Tanksley, S.D. (1996) Stable transfer of intact high molecular weight DNA into plant chromosomes, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9975–9979.

Ikeno, M., Grimes, B., Okazaki, T., et al. (1998) Construction of YAC-based mammalian artificial chromosomes, *Nat. Biotech.* **16**, 431–439.

Jenkins, R.E., Pennington, S.R. (2001) Arrays for protein expression profiling: towards a viable alternative to two-dimensional gel electrophoresis, *Proteomics* **1**, 13–29.

Kim, U.J., Shizuya, H., de Jong, P.J., Birren, B., Simon, M.I. (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector, *Nucleic Acids Res.* **20**, 1083–1085.

Messing, J. (1983) New M13 vectors for cloning, *Methods Enzymol.* **101**, 20–78.

Murray, N.E., Murray, K. (1974) Manipulation of restriction targets in phage λ to form receptor chromosomes for DNA fragments, *Nature* **251**, 476–481.

Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based

vector, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8794–8797.

Soares, M.B., Bonaldo, M.F., Jelenc, P., Su, L., Lawton, L., Efstratiadis, A. (1994) Construction and characterization of a normalized cDNA library, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9228–9232.

Sternberg, N. (1990) Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 103–107.

Tao, Q., Zhang, H.-B. (1998) Cloning and stable maintenance of DNA fragments over 300 kb in *Escherichia coli* with conventional plasmid-based vectors, *Nucleic Acids Res.* **26**, 4901–4909.

Wheals, A.E., Rose, A.H., Harrison, J.S. (Eds.) (1995) The Yeast, *Yeast Genetics*, Vol. 6, 2nd edition, Academic Press, London.

Zhang, H.-B., Zhao, X.-P., Ding, X.-L., Paterson, A.H., Wing, R.A. (1995) Preparation of megabase-size DNA from plant nuclei, *Plant J.* **7**, 175–184.

DNA Repair in Yeast

Errol C. Friedberg and Paula L. Fischhaber
The University of Texas Southwestern Medical Center, Dallas, TX, USA

1 Yeast as a Eukaryotic Model 429

2 DNA Repair by the Reversal of Base Damage 429
2.1 Enzymatic Photoreactivation (EPR) 429
2.1.1 The Utility of DNA Photolyases 431
2.2 Repair of Alkylation Damage by 06-Methylguanine-DNA Methyltransferase 431

3 DNA Repair by the Excision of Base Damage 432
3.1 Base Excision Repair (BER) 432
3.1.1 DNA Glycosylases 433
3.1.2 AP Endonucleases 434
3.1.3 Repair Synthesis, Excision and DNA Ligation 436
3.1.4 DNA Ligation 436
3.2 Nucleotide Excision Repair (NER) 437
3.2.1 NER Mediated by Dual Damage-specific Endonucleases 437
3.2.2 Proteins for NER are Required for Transcription by RNA Polymerase II 438
3.2.3 NER Mediated by a Single Damage-specific Endonuclease Activity 438
3.3 Mismatch Repair 439

4 DNA Damage Tolerance by Translesion DNA Synthesis (TLS) 440

Bibliography 442
Books and Reviews 442
Primary Literature 442

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

Keywords

Enzymatic Photoreactivation
A process in which pyrimidine dimers, produced by ultraviolet (UV) irradiation of DNA, are repaired by their enzyme-catalyzed monomerization.

DNA Photolyase
The enzyme that catalyzes enzymatic photoreactivation of pyrimidine dimers in DNA.

Cyclobutane Pyrimidine Dimers
Covalently joined adjacent pyrimidines in DNA produced by exposure of DNA to ultraviolet radiation.

O6-Methylguanine-DNA Methyltransferase
An enzyme that specifically transfers methyl groups from the methyltransferase O6 position of guanine in DNA to itself.

Base Excision Repair
A DNA excision repair mode, whereby damaged bases are excised as the free bases.

Nucleotide Excision Repair
A DNA excision repair mode whereby damaged bases are excised as nucleotide structures.

Mismatch Repair
A DNA repair mode whereby mispaired bases are excised from DNA.

DNA Damage Tolerance
Any mechanism by which base damage in DNA is effectively bypassed during DNA replication without its removal prior to the bypass.

DNA repair comprises a series of biochemical processes by which alterations of the chemical structure of DNA are eliminated and the DNA is restored to its native state. By strict definition, the term excludes cellular responses by which DNA damage is tolerated but not eliminated from the genome. However, popular usage has incorporated DNA damage tolerance as a mode of DNA repair in a biologically meaningful sense. Hence, for completeness, a particular mode of damage tolerance called *translesion DNA synthesis* (TLS) is included in this discussion.

DNA repair occurs by a variety of distinct biochemical pathways. These differ primarily on the basis of the particular type of damage under consideration. The damage can affect both, the nitrogenous bases and the sugar-phosphate backbone of DNA. The most important form of DNA damage in the latter category is strand breaks. However, this discussion is confined to the repair of damage to the nitrogenous bases in DNA, the informational components of the genetic material. The detailed mechanisms of known DNA repair pathways vary in different organisms. This article considers DNA repair pathways in the yeast, *Saccharomyces cerevisiae*.

1 Yeast as a Eukaryotic Model

Yeasts are eukaryotes, since within each cell the genetic material is contained in a morphologically discrete nucleus. However, yeast cells are not organized into tissues and organs. Thus, they are classified as lower eukaryotes. Nonetheless, the essential parameters of nucleic acid metabolism in yeast cells are very similar to those in higher eukaryotes (including mammals). Thus, yeasts have served as informative models for exploring many aspects of the molecular biology of eukaryotic DNA metabolism.

Yeasts have been shown to be a tractable organism for genetic studies and for recombinant DNA technology. Yeast cells are also easy and inexpensive to grow in the laboratory and have a short life cycle. The latter property provides the further advantage that they multiply rapidly relative to mammalian cells. The two yeasts most widely used in research are the budding yeast *Saccharomyces cerevisiae* and the fission yeast *Schizosaccharomyces pombe*. However, these two yeasts are as evolutionarily distant from one another as each of them is from *Homo sapiens*. This review is confined to a consideration of DNA repair in the yeast *S. cerevisiae*. Thus, the terms yeast and *S. cerevisiae* are used interchangeably here.

2 DNA Repair by the Reversal of Base Damage

Several types of base damage can be repaired by single step biochemical reactions during which specific chemical alterations are directly reversed. The two best-studied examples of this mode of DNA repair in yeast are the enzymatic photoreactivation (EPR) of cyclobutane pyrimidine dimers (CPD) and of [6–4] photoproducts caused by UV radiation, and the repair of O6-methylguanine and O4-methylthymine in DNA caused by exposure to alkylating agents.

2.1 Enzymatic Photoreactivation (EPR)

When living cells are exposed to UV light at wavelengths of ~260 nm (near the absorption maximum for DNA), the absorbed radiation can produce chemical alterations of the bases, resulting in the formation of distinctive photoproducts. The

quantitatively major photoproduct in DNA arises from the covalent joining of adjacent pyrimidines through saturation of the 5,6 double bonds. These altered bases are called *cyclobutane pyrimidine dimers* (CPD) because the covalent linkage between them comprises a cyclobutane ring structure. A second predominant photoproduct is called the *pyrimidine–pyrimidone* (PP) or [6–4] photoproduct ([6–4] PP), in which adjacent pyrimidines are covalently joined through a 6,4 linkage.

Repair enzymes in yeast called *photoreactivating enzymes* or *DNA photolyases* that specifically recognize either CPD or [6–4] PP in DNA catalyze their reversal, generating the native dipyrimidine state (Fig. 1). Both CPD-DNA photolyase and [6–4] PP-DNA photolyase catalyze photochemical reactions that require the presence of light at wavelengths ranging from ~350 to 450 nm. The absorption of light at photoreactivating wavelengths and the specific photochemistry required to monomerize CPD and [6–4] PP is effected by chemical entities called *chromophores* that are noncovalently bound components of all DNA photolyases (Fig. 1). The specific binding of DNA photolyases to substrate sites in DNA does not require the presence of the chromophores or photoreactivating light. Hence, the complete photolyase reaction can be written as follows.

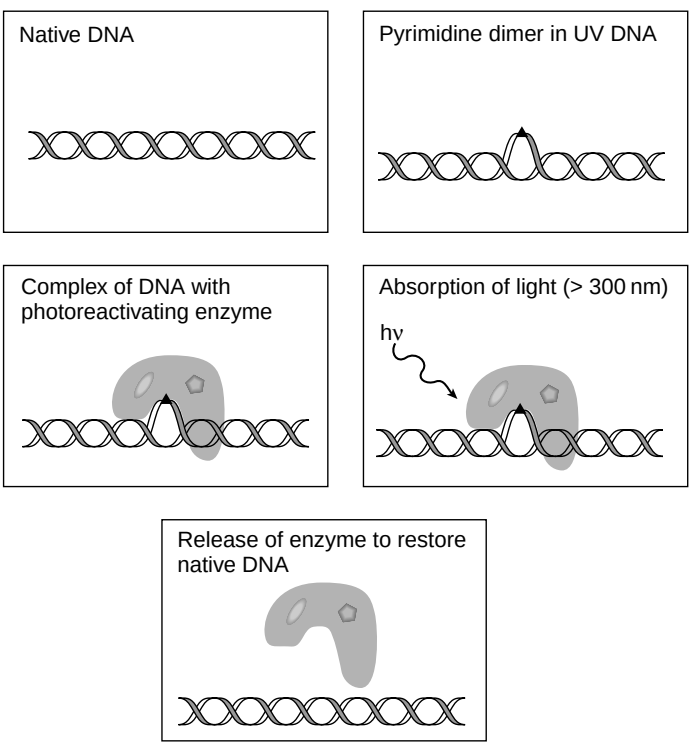

Fig. 1 Schematic representation of the reversal of pyrimidine dimers in DNA by yeast photoreactivating enzyme (pyrimidine dimer-DNA photolyase).

$$E + DNA[Py<>Py] \xrightarrow{\text{Dark reaction}} E - DNA[Py<>Py] \xrightarrow{\text{Light reaction}} E + DNA[PyPy]$$

2.1.1 The Utility of DNA Photolyases

The substrate specificity of DNA photolyases for CPD and [6–4] PP has facilitated the use of this class of DNA repair enzymes as specific probes for the presence of these photoproducts in DNA. Since DNA exposed to UV radiation contains multiple types of photoproducts, it is possible to correlate a particular biological endpoint with the presence of CPD or [6–4] PP, or by showing that the endpoint is mitigated by their removal by enzymatic photoreactivation. This is particularly useful for studies in living cells. The removal of CPD or [6–4] PP from UV-irradiated DNA by incubating DNA in the presence of DNA photolyase *in vitro* also enriches for other types of photoproducts, which may be the primary focus in DNA repair studies. Finally, when used under appropriate conditions, the tight binding affinity of DNA photolyases for CPD and [6–4] PP respectively, in DNA, allows for the detection and quantitation of these lesions.

2.2 Repair of Alkylation Damage by 06-Methylguanine-DNA Methyltransferase

A second ubiquitous example of the repair of specific types of base damage in DNA is that effected by the enzyme O6-methylguanine-DNA methyltransferase. The methylation of oxygen and nitrogen atoms in the nitrogenous bases in DNA is a frequent consequence of the exposure of DNA to reactive alkylating agents, including some that are generated in living cells. Many sites in each of the four bases are sensitive to the covalent addition of small alkyl groups. However, alkylations at O6-G and O4-T are uniquely available for repair by O6-methylguanine-DNA methyltransferase. During the course of this repair reaction, the methyl groups (which can alter the normal base pairing properties of DNA during replication and hence promote mutations) are transferred and covalently attached to a specific cysteine in the enzyme, resulting in the formation of a 5-methylcysteine residue in the polypeptide, thereby inactivating it. Hence, the enzyme is consumed in the reaction in which it participates and acts stoichiometrically rather than catalytically. Such enzymes are called *suicide enzymes*. This novel enzyme also removes ethyl and butyl groups from the O6 position of guanine and the O4 position of thymine in DNA, though less efficiently.

The yeast enzyme is encoded by a gene called *MGT1*. Not unexpectedly, *mgtl* mutants are hypersensitive to the lethal and mutagenic effects of alkylating agents. Such mutants also manifest an increased rate of spontaneous mutagenesis, suggesting that alkylation of O6-G and O4-T occurs naturally in yeast. Natural sources of such spontaneous alkylation damage to DNA have in fact been identified. The amino acid sequence of the yeast enzyme is highly conserved. In all O6-methylguanine-DNA methyltransferase proteins studied, including that from yeast, the same cysteine residue is used as the methyl acceptor species. This cysteine residue is present in the highly conserved penta-amino acid sequence Pro-Cys-His-Arg-Val/Ile. It is not obvious what structural features of alkylated

guanine and thymine provide the exquisite substrate specificity of O6-methylguanine-DNA methyltransferase, although it has been noted that both the O6 position of guanine and the O4 position of thymine occupy the major groove of double-stranded DNA. However, O6-methylguanine-DNA methyltransferase does not catalyze the removal of alkyl groups in other positions on bases, which also project into the major groove. Thus, enzyme specificity is apparently also conferred by oxygen atoms of the nitrogenous bases carrying bound alkyl groups.

3 DNA Repair by the Excision of Base Damage

The fundamental paradigm of all excision repair is the physical removal of the damaged, incorrect (with respect to base pairing), or inappropriate (with respect to normal DNA function) bases, and the restoration of normal sequence information by using the opposite intact DNA strand as a template for DNA synthesis. Hence, all excision repair modes require template-directed repair DNA synthesis, that is, double-stranded DNA.

All excision repair modes share in common a requirement for the enzyme-catalyzed removal of damaged bases. In some cases, the damaged (or inappropriate) bases are excised as free bases. This mode of excision repair is called *base excision repair* (BER) (Fig. 2). In other cases, the damaged bases are removed as part of a larger entity, typically an oligonucleotide fragment, which in yeast is ~30 nucleotides long. This mode of repair is called *nucleotide excision repair* (NER) (Fig. 3). There are indications from studies on the yeast *S. pombe* that NER can occur by several biochemically distinct modes. A form of NER that operates specifically on mispaired bases and which is specifically directed to the strand containing the incorrect base is called *mismatch repair* (*MMR*) or *mismatch correction* (Fig. 4).

3.1 Base Excision Repair (BER)

The substrates amenable to BER are determined by the range of specificity exhibited by repair enzymes called *DNA glycosylases* (see below). In general, BER operates on bases that are chemically altered in ways that result in minimal, if any, distortion of the DNA double

Fig. 2 Schematic model of base excision repair (BER) of DNA. The figure shows the presence of a uracil residue in DNA that arose by the deamination of cytosine. This residue is specifically recognized by uracil-DNA glycosylase, which catalyzes its removal, leaving an apurinic (AP) site in the DNA. This site is attacked by an AP endonuclease activity (such as the yeast APN1, human APE or HAP 1 activities), resulting in hydrolysis of the phosphodiester bond 5′ to the AP site. Two possible schemes are shown for further repair of the DNA. On the left, DNA polymerase β inserts a single correct nucleotide and at the same time a dRpase activity excises the 3′ deoxyribose-phosphate moiety generated by the AP endonuclease. DNA ligase (which in human cells is believed to act in concert with XRCC1 protein) ligates the DNA, thereby completing the DNA repair. The scheme shown on the right suggests that, in some cases, a different DNA polymerase may begin synthesis at the site of the nick created by the AP endonuclease, peeling back a 5′ single-stranded region and eventually generating a flap structure. This flap is specifically recognized by an enzyme called Rad27 protein in yeast and DNase IV, MF1 or FEN-1 in mammalian cells. DNA ligation, then, completes the repair as before. (Reproduced with permission from T. Lindahl.)

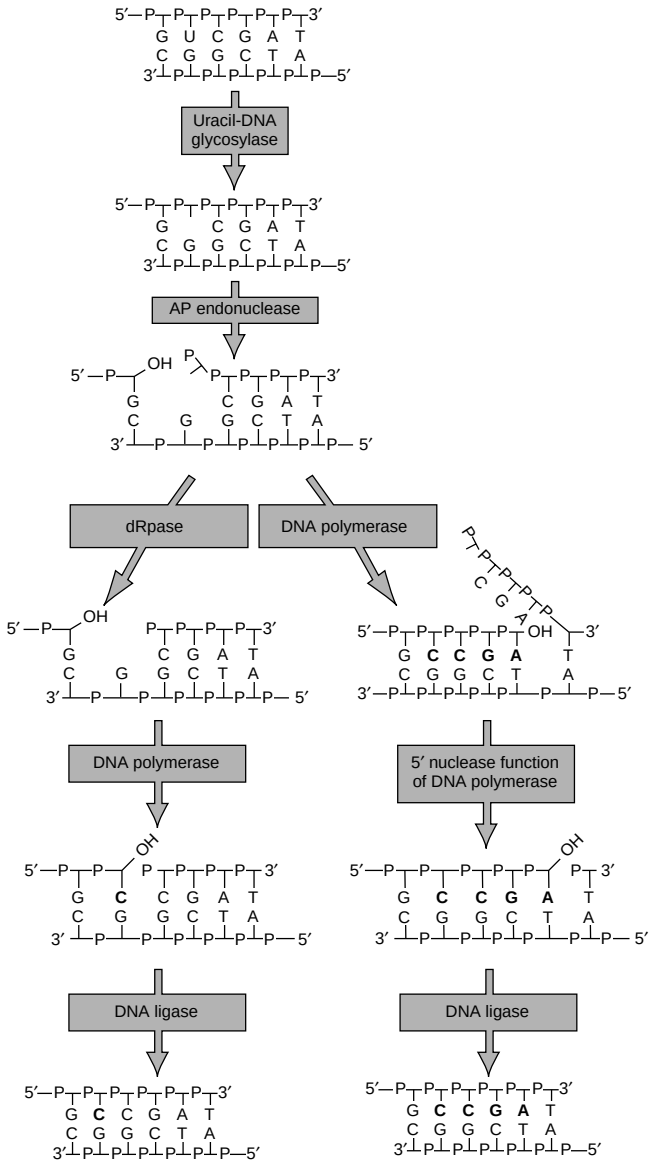

helix. Many such forms of base damage result from the presence of chemically reactive oxygen species (ROS) in cells. Hence, oxidative base damage is heavily dependent on BER for its repair. Such damage is believed to be an important form of spontaneous mutations.

3.1.1 DNA Glycosylases

During BER, the excision of damaged bases is achieved by the enzyme-catalyzed hydrolysis of the *N*-glycosyl bond that links the base to the rest of the DNA molecule (Fig. 2). Enzymes endowed with this catalytic function are called *DNA*

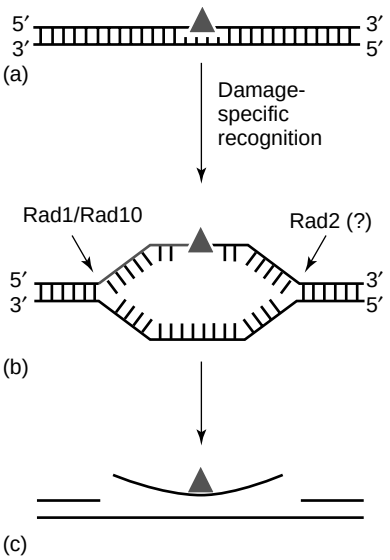

Fig. 3 Cartoon showing the postulated mechanism of bimodal incision of DNA during Global NER in yeast. Following the specific recognition of base damage by the GNER machinery, a "bubble" structure is somehow generated. The Rad1/Rad10 enzyme recognizes the junction between double-stranded and single-stranded DNA, and cleaves the 3′ single-stranded DNA. The Rad2 endonuclease similarly recognizes the junction between double-stranded and single-stranded DNA, and cleaves the 5′ single-stranded DNA. The opposite (undamaged) strand must somehow be protected, since *in vivo* incisions are only made on the damaged strand. (Reproduced with permission from E. C. Friedberg, G. C. Walker and W. Siede.)

glycosylases. Yeast and most other living organisms have multiple DNA glycosylases, each of which is more or less specific for different types of base damage. DNA glycosylases are monomeric proteins with no requirement for divalent cations or other cofactors. The first DNA glycosylase discovered is called *uracil-DNA glycosylase* because it exclusively recognizes uracil residues in DNA.

DNA glycosylases are not only endowed with the ability to recognize particular bases but also to only recognize these in the context of deoxyribose linkages. Thus, uracil-DNA glycosylase does not remove uracil from RNA. Some DNA glycosylases have a more relaxed substrate specificity. For example, a DNA glycosylase called 3-methyladenine-DNA glycosylase II not only recognizes 3-methyladenine residues in DNA but also recognizes 7-methylguanine and 3-methylguanine. A 3-methyladenine-DNA glycosylase activity has been identified in yeast cells. The protein is encoded by a gene called *MAG*, which is upregulated when yeast cells are exposed to alkylating agents. Curiously, as is the case with the yeast *PHR1* gene (see above), induction of *MAG* expression also occurs following DNA damage by agents such as UV radiation, which do not result in the formation of alkyl groups in DNA.

3.1.2 AP Endonucleases

The release of free bases catalyzed by DNA glycosylases results in the generation of sites of base loss (apurinic or apyrimidinic sites) in the genome (Fig. 2). The repair of AP sites and hence the completion of the process of BER requires the participation of several other enzymes. Sites of base loss, generated by the action of DNA glycosylases or by spontaneous hydrolysis of N-glycosyl bonds, are attacked by a specific class of endonucleases called apurinic/apyrimidinic (AP) endonucleases. All known AP endonucleases catalyze the hydrolysis of the phosphodiester bond 5′ to the site of base loss generating nicks with 3′-OH and 5′-Phosphate termini (Fig. 2). *In vitro*, some DNA glycosylases (such as the pyrimidine

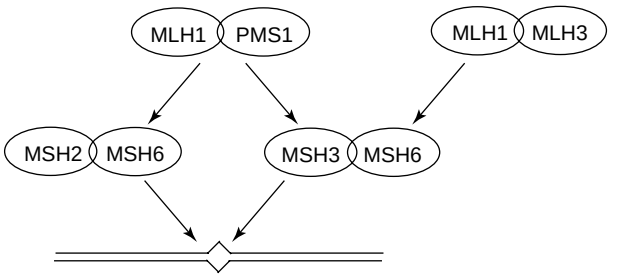

E.coli protein homolog	*S. cerevisiae* proteins	*S. cerevisiae* known activity or probable function
MutS	MSH2–MSH6 (MutSα) MSH3–MSH6 (MutSβ)	Binding to mispaired bases or loops
MutL	MLH1–PMS1 (MutLα) MLH1–MLH3 (MutLβ) MLH1–MLH2 ?	Increased affinity of MutSα or MutSβ to mismatched or looped DNA MLH2 presumed to function analogously to MLH3
	Exo1	DNA nuclease that interacts with MSH2, MSH3, and MLH1. Possible functional equivalent of MutH in *E. coli*

Fig. 4 Top panel: A Schematic of function among the known yeast mismatch repair factors. The MSH2–MSH6 complex or the MSH3–MSH6 complex bind mismatched or looped DNA. The MLH1–PMS1 or MLH1–MLH3 complexes bind MSH2-MSH6 or MSH3-MSH6, enhancing affinity for the mismatch site. Bottom panel: Table describing the known or presumed protein factors involved in yeast mismatch repair.

dimer-DNA glycosylase discussed above) are able to effect cleavage of the 3′-phosphodiester bond at AP sites through a so-called AP lyase activity.

The quantitatively major AP endonuclease activity in yeast is variously designated as yeast AP endonuclease I, yeast DNA diesterase, yeast 3′-phosphoglycoaldehyde

(3′-PGA) diesterase, or yeast AP endonuclease/3′-diesterase. The enzyme has a molecular weight of ~40.5 kDa and is a metalloenzyme. Atomic absorption spectroscopy has revealed the presence of 3.3 Zn^{++} ions per molecule of protein. Nuclease activity is also stimulated by Co^{++} ions. In addition to its AP endonuclease activity, yeast AP endonuclease can excise a variety of 3′ esters in DNA including 3′-phosphoglycoaldehyde, 3′-phosphoryl groups and 3′-α, β unsaturated aldehydes. Like all AP endonucleases, the enzyme hydrolyses DNA 5′ to AP sites, generating 3′-OH and 5′-deoxyribose- phosphate termini.

The structural gene that encodes this AP endonuclease is designated *APN1*. Strains carrying disruptions or deletions of the *APN1* gene are hypersensitive to killing by a variety of oxidative and alkylating agents, which damage DNA. Additionally, such mutants have a spontaneous mutation rate about an order of magnitude greater than that observed in wild-type yeast cells.

3.1.3 Repair Synthesis, Excision and DNA Ligation

The incision of DNA 5′ to sites of base loss provides a normal 3′OH nucleotide terminus for the initiation of repair synthesis of DNA (Fig. 2). The yeast repair synthesis during BER is effected by an enzyme called *DNA polymerase* δ, but a related enzyme, DNA polymerase ε, may also carry out this function. It is not known what specific conditions dictate the preferential utilization of a particular DNA polymerase. The repair synthesis reaction is limited. Only ~2 to 4 nucleotides are inserted. To achieve this, the DNA strand with the 5′-phosphate terminal AP site must be displaced by a limited nick translation. As shown in Fig. 2, displacement of a single strand with a 5′ terminus by repair synthesis would generate a three-stranded flap structure. A yeast enzyme has been identified, which specifically recognizes flap structures with 5′ single-strand tails and removes the single-stranded flap. This protein is encoded by a gene called *RAD27*. The substrate specificity of the Rad27 nuclease has promoted the notion that repair synthesis precedes excision of the AP site during BER as shown in Fig. 2. However, an equally plausible model is that the 5′ terminal AP site is excised first and repair synthesis follows this event.

Figure 2 also shows an alternative repair synthesis/excision reaction for which there is substantial evidence in mammalian cells. In this reaction, a single nucleotide is believed to be inserted by DNA polymerase β. The 5′ terminal AP site is then excised by an enzyme called DNA deoxyribophosphodiesterase (dRpase). Hence, the repair synthesis patch during this putative mode of BER is one-nucleotide long. A dRpase activity has not been identified in yeast. A gene that encodes a polypeptide with the properties of mammalian DNA polymerase β has been cloned from yeast. However, deletion of this gene produces no recognizable phenotype. This does not necessarily mean that the polymerase has no role in DNA repair, since in its absence the alternative longer patch repair synthesis mode may operate exclusively.

3.1.4 DNA Ligation

The final event in all excision repair reactions is the joining of newly inserted nucleotides with 3′-OH termini to the 5′-phosphate terminus of the extant DNA. This joining reaction is catalyzed by an enzyme called *DNA ligase*. A DNA ligase encoded by the yeast *CDC9* gene has been purified and characterized. However, there

are indications that yeast cells contain at least one other DNA ligase activity.

3.2 Nucleotide Excision Repair (NER)

Unlike BER, in which different repair enzymes (DNA glycosylases) recognize different forms of base damage, NER is in general characterized by the recognition of multiple diverse types of base damage by a single repair apparatus. Yeast has been a particularly informative model for deciphering the many genetic and biochemical complexities of NER in eukaryotes. Parallel studies in yeast and mammalian cells have shown that the genes and proteins involved in NER are conserved, and that the detailed biochemistry of this process in these organisms is different from that in bacteria. This structural and functional conservation has allowed for reliable extrapolations from yeast to man, and vice versa, and at the risk of some inaccuracy, we will continue to make such extrapolations in the ensuing discussions. Very recent studies in the yeast *S. pombe* have also provided evidence that NER can occur by more than one biochemical pathway in eukaryotic cells (see later discussion). A fundamental issue yet to be resolved is what determines the broad substrate specificity of the NER apparatus. It appears that two fundamental DNA elements must be present for a DNA lesion to be recognized by NER: the presence of a covalent modification of the DNA, either on a nucleobase or on the phosphodiester backbone, and loss of DNA base pairing. DNA damage recognition is the first step in NER.

3.2.1 NER Mediated by Dual Damage-specific Endonucleases

This excision repair mode is characterized by enzyme-catalyzed incision of the affected DNA strand both 5′ and 3′ to sites of base damage at relatively precise distances from such sites, generating oligonucleotide fragments ~30 nucleotides long, which are subsequently excised (Fig. 3). During NER, the 5′ incisions are located ~22 nucleotides from the sites of damage while the 3′ incisions are located just five nucleotides from these sites (Fig. 3). This precise spacing of nicks is believed to result from the formation of a discrete "repair bubble" ~30 nucleotides long by unwinding of the DNA duplex around the site of base damage (Fig. 3). This repair bubble is generated by the action of two DNA helicases, Rad3, and Rad25 (Ssl2) that unwind DNA in opposite directions. The two DNA helicases are part of a larger protein complex called TFIIH that participates in both NER and RNA polymerase II transcription (see later discussion).

Other proteins specifically recognize and cleave double-strand/single-strand junctions in DNA with polarities that are specified by the directionality of the single-stranded DNA. Hence, the 5′ endonuclease for NER in yeast, consisting of a heterodimer of the products of the *RAD1* and *RAD1O* genes, cleaves the 3′ single-stranded DNA at double-strand/single-strand junctions *in vitro* (Fig. 3). In contrast, the 3′ endonuclease, consisting of the monomeric product of the yeast *RAD2* gene, specifically cleaves the 5′ single-stranded DNA at double-strand/single-strand junctions (Fig. 3). So the substrate specificity for incision of damaged DNA during NER is determined by the presence of duplex-/single-stranded DNA junctions generated by the TFIIH helicases rather than by the damaged bases themselves.

Besides the Rad1, Rad1O, Rad2, Rad3, and Ss12 proteins, the nucleotide excision machinery in yeast includes at

least six other proteins designated Rad4, Radl4, Ssl1, Tfb1, Tfb2, Tfb3, and Tfb4 (Fig. 5). Rad14 protein binds preferentially to damaged DNA *in vitro*. The Rad4 and Rad23 proteins may be involved in the recognition of base damage as well, and the Rad7 and Radl6 genes are believed to be required for the post-incision excision of oligonucleotide fragments. All of the proteins required for NER in yeast are believed to be assembled into a large multiprotein complex called the *nucleotide excision repairosome* (Fig. 5).

3.2.2 Proteins for NER are Required for Transcription by RNA Polymerase II

As mentioned above, Rad3, Ss12, Ssl1, Tfb1, Tfb2, Tfb3, and Tfb4 proteins are components of a protein complex called TFIIH, which in concert with several other transcription factors is required for the initiation of basal transcription by RNA polymerase II. It has been demonstrated that Rad3, Ss12, Ssl1, and Tfbl are indispensable for NER *in vitro* and it is likely that the same is true of the Tfb2 and Tfb3 proteins. Hence, all six of these proteins are probably bifunctional, with obligatory roles in Global NER and in RNA polymerase II transcription.

3.2.3 NER Mediated by a Single Damage-specific Endonuclease Activity

Recent studies in the yeast *S. pombe* have identified a novel endonuclease that recognizes CPD and [6–4] photoproducts, as well as various bulky chemicals in DNA. The enzyme nicks DNA uniquely 5′ immediately adjacent to sites of damage. The events that follow the 5′ incision event are not yet known. It is likely that the physical excision of the base damage (presumably as part of an oligonucleotide fragment) is effected by events associated

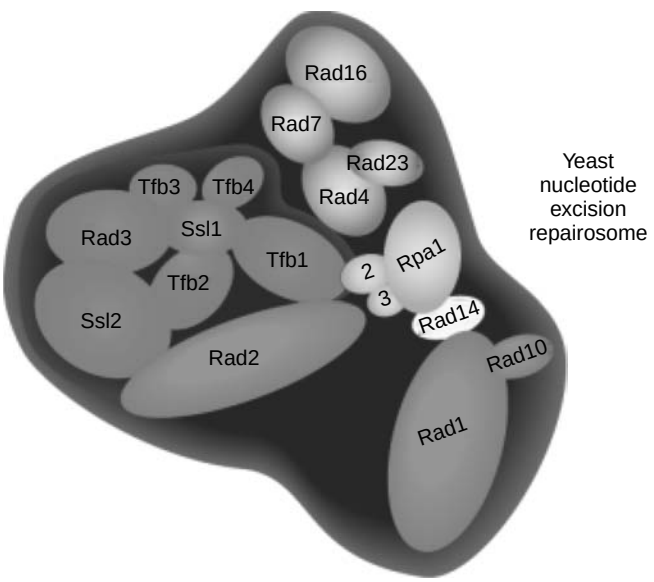

Fig. 5 A cartoon showing all the known protein subunits of the fully assembled yeast nucleotide excision repairosome. The subcomplex comprising Rad3, Ssl2, Tfb1, Tfb2, Tfb3, and Tfb4 proteins is called TFIIH.

with repair synthesis and nick translation, without a requirement for a second endonuclease.

3.3 Mismatch Repair

Mismatch repair refers to biochemical events that result in the excision of mismatched bases from DNA. The repair of mismatches can occur by either a BER mode or a NER mode, that is, the mismatched bases can be excised as free bases or as nucleotides. The BER mode of MMR is limited to certain mismatches and has not been identified in yeast. In contrast, the NER mode of MMR is ubiquitous and appears to be mechanistically conserved from prokaryotes to humans. During semiconservative DNA synthesis in *E. coli*, newly replicated DNA strands are methylated at GATC tetranucleotide sequences. The kinetics of this process results in a period of transient undermethylation of these strands. The MMR system exploits this chemical difference to provide strand specificity for the excision of mispaired bases. In *E. coli*, the presence of both mismatched bases and undermethylated (at GATC sequences) DNA results in the binding of proteins encoded by the *mutS* and *mutL* genes. The formation of a bound MutS/MutL complex activates a latent endonuclease in a protein encoded by the *mutH* gene, resulting in the incision of the newly replicated DNA strand at an unmethylated GATC sequence. This nick can occur either 5′ or 3′ to a mismatch and initiates degradation of the DNA between the nick and the mismatch, resulting in excision of the latter. The excision reaction, which is dependent on the MutS and MutL proteins, a DNA helicase and an exonuclease, results in the excision of a tract of DNA that extends from the site of the mismatch to the site of incision at the nonmethylated GATC sequence.

Homologous systems for MMR have been identified in yeast and in higher eukaryotes. Yeast DNA is not methylated, hence the molecular basis for the strand specificity during MMR in this organism is unknown. Six yeast genes have been shown to be homologous to the *E. coli MutS* gene, designated *MSH1 – MSH6* (for Mut S Homolog) (Fig. 4). The MSH4 and MSH5 proteins are not believed to function in MMR, but rather in meiotic recombination events. MSH1 protein is targeted to mitochondria. The MSH2, MSH3, and MSH6 proteins bind mismatched and looped-out DNA bases as heterodimeric protein complexes. MSH2 and MSH6 form a complex known as MutSα, while MSH3 and MSH6 form MutSβ (Fig. 4). Homologs of *E. coli* MutL protein include MLH1, MLH2, MLH3, and PMS1. Several of these factors also form heterodimeric protein complexes with each other that interact with the MutSα and MutSβ complex *in vitro*. The MLH1-PMS1 complex is known as MutLα, which interacts with both MutSα and MutSβ. MLH1 and MLH3 form the complex MutLβ, which interacts with MutSβ.

MutSα binds most readily to single base mismatches while MutSβ prefers longer looped-out regions of DNA. MutLα increases the affinity of either MutSα or MutSβ for mismatched or looped DNA. Additionally, ATP binding to MutLα induces a conformational change. In *E. coli*, MutL increases the translocation of MutS in traversing from the mismatch site to the nearest site of hemimethylation that establishes strand specificity, the nonmethylated strand being the newly synthesized DNA and hence the strand requiring correction. As already mentioned, the same mechanism cannot hold for yeast and it has

been postulated that a MutLα conformational change may recruit the necessary factors for subsequent events of MMR such as strand discrimination.

The other required factors for MMR have not been determined, but likely include EXO1, which interacts with MSH2, MSH3, and MLH1. EXO1 is a DNA nuclease but bears homology to the XPG/FEN1 family rather than *E. coli* MutH. The EXO1 relationship to MMR has also been suggested genetically through epistasis analysis. Hence, EXO1 may function as the exonuclease required for removal of mismatched DNA or loops. Additionally, Proliferating Cell Nuclear Antigen (PCNA) has been shown to interact both with MutSα and MutSβ through MSH6 and MSH3, respectively. PCNA is also capable of disrupting MutSα-MutLβ-mismatched DNA complexes. Interaction with PCNA suggests that yeast MMR is influenced by the DNA replication machinery, which perhaps affords a basis for strand discrimination.

4 DNA Damage Tolerance by Translesion DNA Synthesis (TLS)

Various types of base damage can impede or completely arrest normal high-fidelity DNA replication, thereby compromising the ability of cells to traverse the cell cycle during cell division. To avoid cell death, cells have evolved multiple mechanisms to bypass the arrested DNA replication machinery (Fig. 6). One of these utilizes several specialized DNA polymerases that have the ability to bypass template lesions that block DNA replication by the process of translesion DNA synthesis. Higher eukaryotes are endowed with at least nine such polymerases; the genome of *S. cerevisiae* reveals open reading frames for just three of these called Rev 1, DNA polymerase ζ (polζ) and DNA polymerase η (polη). Rev1 may not be a true translesion polymerase since its catalytic activity is almost exclusively confined to the incorporation of cytidine into DNA. Hence, Rev1 is also referred to as dCMP transferase. Polη and Polζ are typical translesion polymerases. Like all such enzymes, these two polymerases are endowed with a structure that facilitates their ability to replicate past various types of DNA lesions. This relaxed fidelity results in markedly reduced accuracy for nucleotide incorporation when copying either damaged or undamaged DNA. Hence, these polymerases are highly error prone and can generate mutations. Their potential for mutagenic DNA synthesis presumably requires that they do not normally access the genome when not required for translesion DNA synthesis.

Polη is highly conserved in higher eukaryotes and in human cells has been identified as the product of the *XP V* gene, which when inactivated, results in the disease xeroderma pigmentosum (XP). XP is characterized by extreme sensitivity to sunlight with a markedly increased predisposition to cancer in exposed areas of the skin. Like the human enzyme, yeast polη is able to bypass thymine–thymine dimers and possibly other CPDs. In doing so, it incorporates adenine opposite the dimerized thymines with the same relative accuracy as it incorporates adenine opposite native adjacent thymine pairs. It has been similarly demonstrated to bypass other types of base damage with relative accuracy. Hence, it would appear that polη has evolved to bypass many types of base damage with relative accuracy in *S. cerevisiae*. The biological role of polζ is not known. *In vitro*, this enzyme, like polζ,

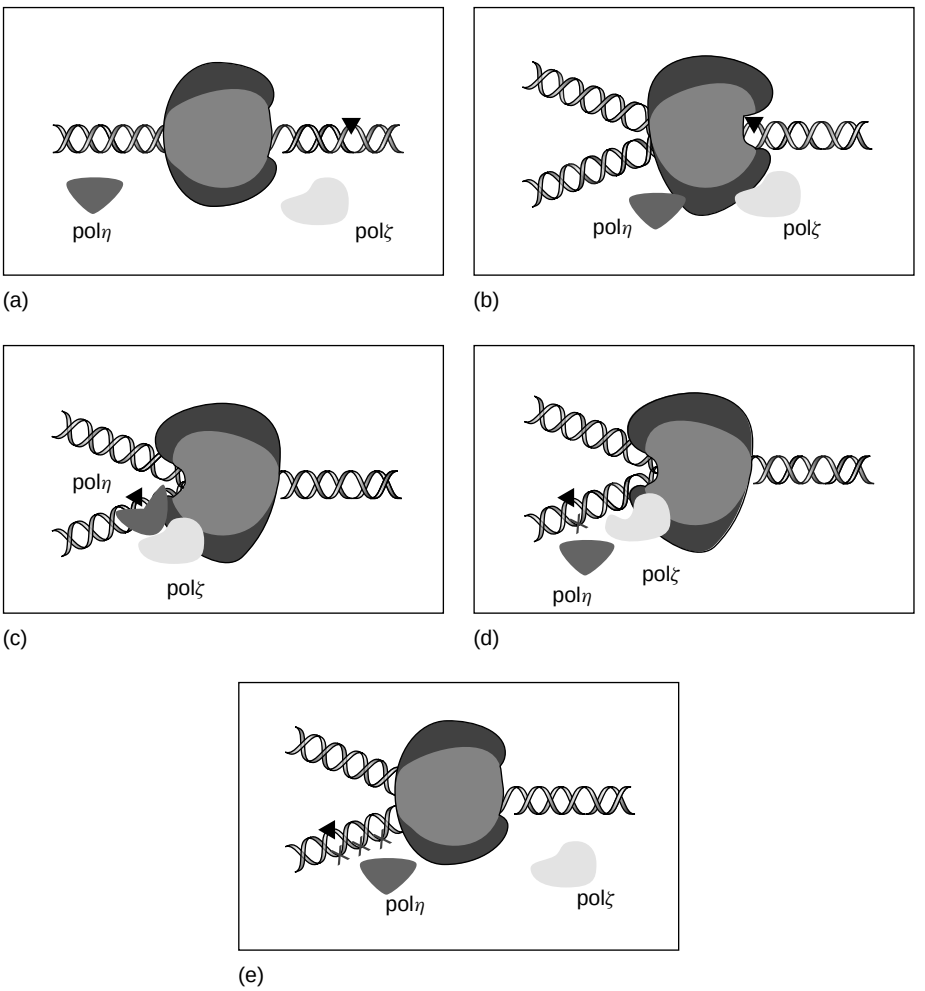

Fig. 6 Diagrammatic representation of TLS by two specialized polymerases in yeast, Polη and Polζ. These polymerases overcome blocks in replication caused by a DNA lesion (a) The DNA replicative machinery stalls immediately behind a site of base damage (triangle). (b) The two specialized polymerases (η and ζ) bind to the arrested replication complex. (c) This interaction promotes a conformational change in the arrested replication machinery placing Polη in direct proximity to the site of base damage, allowing DNA synthesis across the lesion. (d) Polη dissociates and allows Polζ to complete the process of replicative bypass by incorporating several more nucleotides. Once the lesion has been completely bypassed, the polζ dissociates and the replicative machinery assumes DNA replication. Changes to the DNA sequence are incorporated as a result of this process, known as *translesion synthesis*.

can bypass various types of base damage. It remains to be determined whether polζ merely serves as a backup TLS polymerase function in yeast, or has some other function. Studies on TLS polymerases in general suggest that polζ may function primarily, if not exclusively, to extend DNA synthesis beyond sites of TLS by

enzymes like polη. Thus, TLS is considered to be a two-step process involving the incorporation of one or two nucleotides opposite sites of base damage by one TLS enzyme, followed by extension of the primer terminus for a limited number of nucleotides. These two DNA synthesis events may well be coupled.

The precise mechanism whereby the replication machinery switches places with one or more TLS polymerase at the primer terminus of a site of arrested DNA replication is unknown. Nor is it known how a particular TLS enzyme or enzyme pair is selected for the bypass of a particular lesion. The mutations generated by TLS are not necessarily always deleterious. Genomic infidelity is a required backdrop for Darwinian evolution.

See also Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Friedberg, E.C., Walker, G.C., Siede, W. (1995) *DNA Repair and Mutagenesis*, ASM Publications, Washington DC.

Dianov, G., Lindahl, T. (1994) Reconstitution of the Base Excision-Repair Pathway, *Curr. Biol.* **4**, 1069–1076.

Aboussekhra, A., Wood, R.D. (1994) Repair of UV-Damaged DNA by Mammalian Cells and *Saccharomyces cerevisiae*, *Curr. Biol.* **4**, 212–220.

Prakash, S., Sung, P., Prakash, L. (1993) DNA Repair Genes and Proteins of *Saccharomyces cerevisiae*, *Annu. Rev. Genet.* **27**, 33–70.

Friedberg, E.C., Siede, W., Cooper, A.J. (1991) Cellular Responses to DNA Damage in Yeast, In *The Molecular and Cellular Biology of the Yeast Saccharomyces. Genome Dynamics, Protein Synthesis, and Energetics.* Broach, J.R., Pringle, J.R., Jones, E.W. (Eds.) pp. 147–192. Cold Spring Harbor, NY.

Modrich, P. (1991) Mechanisms and Biological Effects of Mismatch Repair. *Ann. Rev. Genet.* **25**, 229–254.

Bebenek, K., Kunkel, T.A. (2002) Family growth: the eukaryotic DNA polymerase revolution, *Cell Mol. Life Sci.* **59**, 54–57.

Cunningham, R.P. (1996) DNA repair: how yeast repairs radical damage, *Curr. Biol.* **6**, 1230–1233.

Friedberg, E.C., Walker, G.C., Siede, W. (1995) *DNA Repair and Mutagenesis*, ASM Press, Washington, DC.

Friedberg, E.C. (1997) *Correcting the Blueprint of Life: An Historical Account of the Discovery of DNA Repair Mechanisms*, Cold Spring Harbor Laboratory Press, Plainview, NY.

Goodman, M.F. (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes, *Annu. Rev. Biochem.* **71**, 17–50.

Gros, L., Saparbaev, M.K., Laval, J. (2002) Enzymology of the repair of free radicals-induced DNA damage, *Oncogene* **21**, 8905–8925.

Modrich, P., Lahue, R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology, *Annu. Rev. Biochem.* **65**, 101–133.

Prakash, S., Prakash, L. (2000) Nucleotide excision repair in yeast, *Mutat. Res.* **451**, 13–24.

Prakash, S., Sung, P., Prakash, L. (1993) DNA repair genes and proteins of Saccharomyces cerevisiae, *Annu. Rev. Genet.* **27**, 33–70.

Sancar, G.B. (2000) Enzymatic photoreactivation: 50 years and counting, *Mutat. Res.* **451**, 25–37.

Primary Literature

Acharya, S., et al. (1996) hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13629–13634.

Bankmann, M., Prakash, L., Prakash, S. (1992) Yeast RAD14 and human xeroderma pigmentosum group A DNA-repair genes encode homologous proteins, *Nature* **355**, 555–558.

Bardwell, A.J., Bardwell, L., Tomkinson, A.E., Friedberg, E.C. (1994) Specific cleavage of model recombination and repair intermediates by the yeast Rad1-Rad10 DNA endonuclease, *Science* **265**, 2082–2085.

Bardwell, L., et al. (1994) Yeast RAD3 protein binds directly to both SSL2 and SSL1 proteins:

implications for the structure and function of transcription/repair factor b, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3926–3930.

Bawa, S., Xiao, W. (1999) Methionine reduces spontaneous and alkylation-induced mutagenesis in Saccharomyces cerevisiae cells deficient in O6-methylguanine-DNA methyltransferase, *Mutat. Res.* **430**, 99–107.

Bebenek, K., et al. (2001) 5′-Deoxyribose phosphate lyase activity of human DNA polymerase iota in vitro, *Science* **291**, 2156–2159.

Bertrand, P., Tishkoff, D.X., Filosi, N., Dasgupta, R., Kolodner, R.D. (1998) Physical interaction between components of DNA mismatch repair and nucleotide excision repair, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14278–14283.

Chang, W.H., Kornberg, R.D. (2000) Electron crystal structure of the transcription factor and DNA repair complex, core TFIIH, *Cell* **102**, 609–613.

Conconi, A., Bespalov, V.A., Smerdon, M.J. (2002) Transcription-coupled repair in RNA polymerase I-transcribed genes of yeast, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 649–654.

Datta, A., Hendrix, M., Lipsitch, M., Jinks-Robertson, S. (1997) Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9757–9762.

Dianov, G., Lindahl, T. (1994) Reconstitution of the DNA base excision-repair pathway, *Curr. Biol.* **4**, 1069–1076.

Drotschmann, K., et al. (1999) Mutator phenotypes of yeast strains heterozygous for mutations in the MSH2 gene, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2970–2975.

Eide, L., et al. (1996) Base excision of oxidative purine and pyrimidine DNA damage in Saccharomyces cerevisiae by a DNA glycosylase with sequence similarity to endonuclease III from Escherichia coli, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10735–10740.

Evans, E., Sugawara, N., Haber, J.E., Alani, E. (2000) The Saccharomyces cerevisiae Msh2 mismatch repair protein localizes to recombination intermediates in vivo, *Mol. Cell* **5**, 789–799.

Flores-Rozas, H., Kolodner, R.D. (1998) The Saccharomyces cerevisiae MLH3 gene functions in MSH3-dependent suppression of frameshift mutations, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12404–12409.

Frit, P., et al. (2002) Transcriptional activators stimulate DNA repair, *Mol. Cell* **10**, 1391–1401.

Gellon, L., Werner, M., Boiteux, S. (2002) Ntg2p, a Saccharomyces cerevisiae DNA N-glycosylase/apurinic or apyrimidinic lyase involved in base excision repair of oxidative DNA damage, interacts with the DNA mismatch repair protein Mlh 1p. Identification of a Mlh 1p binding motif, *J. Biol. Chem.* **277**, 29963–29972.

Glassner, B.J., Rasmussen, L.J., Najarian, M.I., Posnick, L.M., Samson, L.D. (1998) Generation of a strong mutator phenotype in yeast by imbalanced base excision repair, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9997–10002.

Guillet, M., Bolteux, S. (2002) Endogenous DNA abasic sites cause cell death in the absence of Apn1, Apn2 and Rad1/Rad10 in Saccharomyces cerevisiae, *EMBO J.* **21**, 2833–2841.

Guzder, S.N., et al. (1994) DNA repair gene RAD3 of S. cerevisiae is essential for transcription by RNA polymerase II, *Nature* **367**, 91–94.

Guzder, S.N., Habraken, Y., Sung, P., Prakash, L., Prakash, S. (1995) Reconstitution of yeast nucleotide excision repair with purified Rad proteins, replication protein A, and transcription factor TFIIH, *J. Biol. Chem.* **270**, 12973–12976.

Guzder, S.N., Habraken, Y., Sung, P., Prakash, L., Prakash, S. (1996) RAD26, the yeast homology of human Cockayne's syndrome group B gene, encodes a DNA-dependent ATPase, *J. Biol. Chem.* **271**, 18314–18317.

Guzder, S.N., Sung, P., Bailly, V., Prakash, L., Prakash, S. (1994) RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription, *Nature* **369**, 578–581.

Guzder, S.N., Sung, P., Prakash, L., Prakash, S. (1993) Yeast DNA-repair gene RAD 14 encodes a zinc metalloprotein with affinity for ultraviolet-damaged DNA, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5433–5437.

Guzder, S.N., Sung, P., Prakash, L., Prakash, S. (1997) Yeast Rad7-Rad16 complex, specific for the nucleotide excision repair of the nontranscribed DNA strand, is an ATP-dependent DNA damage sensor, *J. Biol. Chem.* **272**, 21665–21668.

Guzder, S.N., Sung, P., Prakash, L., Prakash, S. (1998) Affinity of yeast nucleotide excision repair factor 2, consisting of the Rad4 and

Rad23 proteins, for ultraviolet damaged DNA, *J. Biol. Chem.* **273**, 31541–31546.

Guzder, S.N., Sung, P., Prakash, L., Prakash, S. (1999) Synergistic interaction between yeast nucleotide excision repair factors NEF2 and NEF4 in the binding of ultraviolet-damaged DNA, *J. Biol. Chem.* **274**, 24257–24262.

Haber, J.E., Ray, B.L., Kolb, J.M., White, C.I. (1993) Rapid kinetics of mismatch repair of heteroduplex DNA that is formed during recombination in yeast, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3363–3367.

Habraken, Y., Sung, P., Prakash, L., Prakash, S. (1993) Yeast excision repair gene RAD2 encodes a single-stranded DNA endounclease, *Nature* **366**, 365–368.

Habraken, Y., Sung, P., Prakash, L., Prakash, S. (1994) Holliday junction cleavage by yeast Rad1 protein, *Nature* **371**, 531–534.

Habraken, Y., Sung, P., Prakash, L., Prakash, S. (1995) Structure-specific nuclease activity in yeast nucleotide excision repair protein Rad2, *J. Biol. Chem.* **270**, 30194–30198.

Habraken, Y., Sung, P., Prakash, L., Prakash, S. (1997) Enhancement of MSH2-MSH3-mediated mismatch recognition by the yeast MLH1-PMS1 complex, *Curr. Biol.* **7**, 790–793.

Hamilton, K.K., Kim, P.M., Doetsch, P.W. (1992) A eukaryotic DNA glycosylase/lyase recognizing ultraviolet light-induced pyrimidine dimers, *Nature* **356**, 725–728.

Haracska, L., et al. (2001) Roles of yeast DNA polymerases delta and zeta and of Rev1 in the bypass of abasic sites, *Genes Dev.* **15**, 945–954.

Haracska, L., Prakash, S., Prakash, L. (2002) Yeast Rev1 protein is a G template-specific DNA polymerase, *J. Biol. Chem.* **277**, 15546–15551.

Haracska, L., Prakash, S., Prakash, L. (2003) Yeast DNA polymerase zeta is an efficient extender of primer ends opposite from 7,8-dihydro-8-Oxoguanine and 06-methylguanine, *Mol. Cell. Biol.* **23**, 1453–1459.

Haracska, L., Yu, S.L., Johnson, R.E., Prakash, L., Prakash, S. (2000) Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase eta, *Nat.Genet.* **25**, 458–461.

Heale, S.M., Petes, T.D. (1995) The stabilization of repetitive tracts of DNA by variant repeats requires a functional DNA mismatch repair system, *Cell* **83**, 539–545.

Hosfield, D.J., Guan, Y., Haas, B.J., Cunningham, R.P., Tainer, J.A. (1999) Structure of the DNA repair enzyme endonuclease IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis, *Cell* **98**, 397–408.

Iaccarino, I., et al. (1996) MSH6, a Saccharomyces cerevisiae protein that binds to mismatches as a heterodimer with MSH2, *Curr. Biol.* **6**, 484–486.

Jelinsky, S.A., Estep, P., Church, G.M., Samson, L.D. (2000) Regulatory networks revealed by transcriptional profiling of damaged Saccharomyces cerevisiae cells: Rpn4 links base excision repair with proteasomes, *Mol. Cell. Biol.***20**, 8157–8167.

Johnson, R.E., Haracska, L., Prakash, S., Prakash, L. (2001) Role of DNA polymerase zeta in the bypass of a (6–4) TT photoproduct, *Mol. Cell. Biol.***21**, 3558–3563.

Johnson, R.E., Kovvali, G.K., Prakash, L., Prakash, S. (1995) Requirement of the yeast RTH1 5′ to 3′ exonuclease for the stability of simple repetitive DNA, *Science* **269**, 238–240.

Johnson, R.E., Kovvali, G.K., Prakash, L., Prakash, S. (1996) Requirement of the yeast MSH3 and MSH6 genes for MSH2-dependent genomic stability, *J. Biol. Chem.* **271**, 7285–7288.

Johnson, R.E., Trincao, J., Aggarwal, A.K., Prakash, S., Prakash, L. (2003) Deoxynucleotide triphosphate binding mode conserved in Y family DNA polymerases, *Mol. Cell. Biol.* **23**, 3008–3012.

Johnson, R.E., Washington, M.T., Haracska, L., Prakash, S., Prakash, L. (2000) Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions, *Nature* **406**, 1015–1019.

Johnson, R.E., Washington, M.T., Prakash, S., Prakash, L. (1999) Bridging the gap: a family of novel DNA polymerases that replicate faulty DNA, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12224–12226.

Johnson, R.E., Yu, S.L., Prakash, S., Prakash, L. (2003) Yeast DNA polymerase zeta (zeta) is essential for error-free replication past thymine glycol, *Genes Dev.* **17**, 77–87.

Kirkpatrick, D.T., Petes, T.D. (1997) Repair of DNA loops involves DNA-mismatch and nucleotide-excision repair proteins, *Nature* **387**, 929–931.

Kunz, B.A., et al. (1994) Specificity of the mutator caused by deletion of the yeast structural gene (APN1) for the major apurinic endonuclease, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8165–8169.

Lee, S.K., Yu, S.L., Prakash, L., Prakash, S. (2001) Requirement for yeast RAD26, a homolog of the human CSB gene, in elongation by RNA polymerase II, *Mol. Cell. Biol.* **21**, 8651–8656.

Lee, S.K., Yu, S.L., Prakash, L., Prakash, S. (2002) Requirement of yeast RAD2, a homolog of human XPG gene, for efficient RNA polymerase II transcription. Implications for Cockayne syndrome, *Cell* **109**, 823–834.

Lee, S.K., Yu, S.L., Prakash, L., Prakash, S. (2002) Yeast RAD26, a homolog of the human CSB gene, functions independently of nucleotide excision repair and base excision repair in promoting transcription through damaged bases, *Mol. Cell. Biol.* **22**, 4383–4389.

Ling, H., Boudsocq, F., Woodgate, R., Yang, W. (2001) Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication, *Cell* **107**, 91–102.

Masutani, C., et al. (1999) The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta, *Nature* **399**, 700–704.

Meadows, K.L., Song, B., Doetsch, P.W. (2003) Characterization of AP lyase activities of Saccharomyces cerevisiae Ntg1p and Ntg2p: implications for biological function, *Nucleic Acids Res.* **31**, 5560–5567.

Meier, A., Livingstone-Zatchej, M., Thoma, F. (2002) Repair of active and silenced rDNA in yeast: the contributions of photolyase and transcription-couples nucleotide excision repair, *J. Biol. Chem.* **277**, 11845–11852.

Mol, C.D., lzumi, T., Mitra, S., Tainer, J.A. (2000) DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected], *Nature* **403**, 451–456.

Naegeli, H., Modrich, P., Friedberg, E.C. (1993) The DNA helicase activities of Rad3 protein of Saccharomyces cerevisiae and helicase II of Escherichia coli are differentially inhibited by covalent and noncovalent DNA modifications, *J. Biol. Chem.* **268**, 10386–10392.

Nash, H.M., et al. (1996) Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily, *Curr. Biol.* **6**, 968–980.

Ni, T.T., Marsischky, G.T., Kolodner, R.D. (1999) MSH2 and MSH6 are required for removal of adenine misincorporated opposite 8-oxoguanine in S. cerevisiae, *Mol. Cell* **4**, 439–444.

O'Donovan, A., Davies, A.A., Moggs, J.G., West, S.C., Wood, R.D. (1994). PG endonuclease makes the 3′ incision in human DNA nucleotide excision repair, *Nature* **371**, 432–435.

Ohmori, H., et al. (2001) The Y-family of DNA polymerases, *Mol. Cell* **8**, 7–8.

Pavlov, Y.I., et al. (2002) Correlation of somatic hypermutation specificity and A-T base pair substitution errors by DNA polymerase eta during copying of a mouse immunoglobulin kappa light chain transgene, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9954–9959.

Pavlov, Y.I., Newlon, C.S., Kunkel, T.A. (2002) Yeast origins establish a strand bias for replicational mutagenesis, *Mol. Cell* **10**, 207–213.

Prolla, T.A., Pang, Q., Alani, E., Kolodner, R.D., Liskay, R.M. (1994) MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast, *Science* **265**, 1091–1093.

Russell, S.J., Reed, S.H., Huang, W., Friedberg, E.C., Johnston, S.A. (1999) The 19S regulatory complex of the proteasome functions independently of proteolysis in nucleotide excision repair, *Mol. Cell* **3**, 687–695.

Saparbaev, M., Prakash, L., Prakash, S. (1996) Requirement of mismatch repair genes MSH2 and MSH3 in the RAD1-RAD10 pathway of mitotic recombination in Saccharomyces cerevisiae, *Genetics* **142**, 727–736.

Sassanfar, M., Dosanjh, M.K., Essigmann, J.M., Samson, L. (1991) Relative efficiencies of the bacterial, yeast, and human DNA methyltransferases for the repair of O6-methylguanine and O4-methylthymine. Suggestive evidence for O4-methylthymine repair by eukaryotic methyltransferases, *J. Biol. Chem.* **266**, 2767–2771.

Schauber, C., et al. (1998) Rad23 links DNA repair to the ubiquitin/proteasome pathway, *Nature* **391**, 715–718.

Silvian, L.F., Toth, E.A., Pham, P., Goodman, M.F., Ellenberger, T. (2001) Crystal structure of a DinB family error-prone DNA polymerase from Sulfolobus solfataricus, *Nat. Struct. Biol.* **8**, 984–989.

Singh, K.K., Samson, L. (1995) Replication protein A binds to regulatory elements in yeast DNA repair and DNA metabolism genes, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4907–4911.

Singh, K.K., Sigala, B., Sikder, H.A., Schwimmer, C. (2001) Inactivation of Saccharomyces

cerevisiae OGG1 DNA repair gene leads to an increased frequency of mitochondrial mutants, *Nucleic Acids Res.* **29**, 1381–1388.

Singhal, R.K., Prasad, R., Wilson, S.H. (1995) DNA polymerase beta conducts the gap-filling step in uracil-initiated base excision repair in a bovine testis nuclear extract, *J. Biol. Chem.* **270**, 949–957.

Strand, M., Earley, M.C., Crouse, G.F., Petes, T.D. (1995) Mutations in the MSH3 gene preferentially lead to deletions within tracts of simple repetitive DNA in Saccharomyces cerevisiae, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10418–10421.

Sugawara, N., Paques, F., Colaiacovo, M., Haber, J.E. (1997) Role of Saccharomyces cerevisiae Msh2 and Msh3 repair proteins in double-strand break-induced recombination, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9214–9219.

Svejstrup, J.Q., et al. (1995) Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome, *Cell* **80**, 21–28.

Swanson, R.L., Morey, N.J., Doetsch, P.W., Jinks-Robertson, S. (1999) Overlapping specificities of base excision repair, nucleotide excision repair, recombination, and translesion synthesis pathways for DNA base damage in Saccharomyces cerevisiae, *Mol. Cell. Biol.* **19**, 2929–2935.

Szafraniec, K., Borts, R.H., Korona, R. (2001) Environmental stress and mutational load in diploid strains of the yeast Saccharomyces cerevisiae, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1107–1112.

Tishkoff, D.X., et al. (1997) Identification and characterization of Saccharomyces cerevisiae EXO1, a gene encoding an exonuclease that interacts with MSH2, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7487–7492.

Tran, P.T., Simon, J.A., Liskay, R.M. (2001) Interactions of Exo 1p with components of MutLalpha in Saccharomyces cerevisiae, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9760–9765.

Trincao, J., et al. (2001) Structure of the catalytic core of S. cerevisiae DNA polymerase eta: implications for translesion DNA synthesis, *Mol. Cell* **8**, 417–426.

Vance, J.R., Wilson, T.E. (2002) Yeast Tdp 1 and Rad1-Rad10 function as redundant pathways for repairing Top1 replicative damage, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13669–13674.

van der Kemp, P.A., Thomas, D., Barbey, R., de Oliveira, R., Boiteux, S. (1996) Cloning and expression in Escherichia coli of the OGG1 gene of Saccharomyces cerevisiae, which codes for a DNA glycosylase that excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-N-methylformamidopyrimidine, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5197–5202.

Vande Berg, B.J., Sancar, G.B. (1998) Evidence for dinucleotide flipping by DNA photolyase, *J. Biol. Chem.* **273**, 20276–20284.

Wang, Z., et al. (1994) Transcription factor b (TFIIH) is required during nucleotide-excision repair in yeast, *Nature* **368**, 74–76.

Wang, T.F., Kleckner, N., Hunter, N. (1999) Functional specificity of MutL homologs in yeast: evidence for three Mlh1 based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13914–13919.

Wang, Z., Wu, X., Friedberg, E.C. (1993) Nucleotide-excision repair of DNA in cell-free extracts of the yeast Saccharomyces cerevisiae, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4907–4911.

Washington, M.T., Helquist, S.A., Kool, E.T., Prakash, L., Prakash, S. (2003) Requirement of Watson-Crick hydrogen bonding for DNA synthesis by yeast DNA polymerase eta, *Mol. Cell. Biol.* **23**, 5107–5112.

Washington, M.T., Johnson, R.E., Prakash, S., Prakash, L. (1999) Fidelity and processivity of Saccharomyces cerevisiae DNA polymerase eta, *J. Biol. Chem.* **274**, 36835–36838.

Washington, M.T., Johnson, R.E., Prakash, S., Prakash, L. (2001) Mismatch extension ability of yeast and human DNA polymerase eta, *J. Biol. Chem.* **276**, 2263–2266.

Washington, M.T., Johnson, R.E., Prakash, L., Prakash, S. (2003) The mechanism of nucleotide incorporation by human DNA polymerase eta differs from that of the yeast enzyme, *Mol. Cell. Biol.* **23**, 8316–8322.

Washington, M.T., Prakash, L., Prakash, S. (2001) Yeast DNA polymerase eta utilizes an induced-fit mechanism of nucleotide incorporation, *Cell* **107**, 917–927.

Washington, M.T., Prakash, L., Prakash, S. (2003) Mechanism of nucleotide incorporation opposite a thymine-thymine dimer by yeast DNA polymerase eta, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12093–12098.

Xiao, W., Samson, L. (1992) The Saccharomyces cerevisiae MGT1 DNA repair methyltransferase gene: its promoter and entire

coding sequence, regulation and in vivo biological functions, *Nucleic Acids Res.* **20**, 3599–3606.

Xiao, W., Singh, K.K., Chen, B., Samson, L. (1993) A common element involved in transcriptional regulation of two DNA alkylation repair genes (MAG and MGT1) of Saccharomyces cerevisiae, *Mol. Cell. Biol.* **13**, 7213–7221.

Yang, W. (2003) Pruning DNA. Structure-specific endonucleases (XPF/Rad1/Mus81), *Structure (Camb)* **11**, 365–366.

Yu, S.L., Johnson, R.E., Prakash, S., Prakash, L. (2001) Requirement of DNA polymerase eta for error-free bypass of UV-induced CC and TC photoproducts, *Mol. Cell. Biol.* **21**, 185–188.

DNA Replication and Transcription

Yusaku Nakabeppu[1], *Hisaji Maki*[2], *and Mutsuo Sekiguchi*[3]
[1] *Kyushu University, Fukuoka, Japan*
[2] *Nara Institute of Science and Technology, Nara, Japan*
[3] *Biomolecular Engineering Research Institute, Osaka, Japan*

1 DNA Replication 451
1.1 Structural Aspects of DNA Replication 451
1.1.1 Replicon: Unit of Replication 451
1.1.2 Replication Fork 452
1.2 Biochemistry of DNA Replication 453
1.2.1 DNA Polymerase 454
1.2.2 DNA Helicase and Primase 456
1.2.3 Other Replication Proteins 457
1.3 Regulation of DNA Replication 458
1.3.1 Initiation Mechanisms in Prokaryotes 458
1.3.2 Initiation Mechanisms in Eukaryotes 460

2 Transcription 461
2.1 RNA Polymerase and Transcriptional Apparatus 461
2.1.1 General RNA Polymerase Architecture 461
2.1.2 Transcription Initiation 463
2.1.3 Promoter Escape 465
2.1.4 Transcription Elongation 465
2.2 Transcription Unit 466
2.2.1 Bacterial Transcription Unit 467
2.2.2 Eukaryotic Transcription Unit 467
2.3 Regulation of Transcription 469
2.3.1 Transcriptional Regulation in Bacteria 469
2.3.2 Transcriptional Regulation in Eukaryotes 470

Bibliography 472
Books and Reviews 472
Primary Literature 472

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

Keywords

Origin (ori)
A unique site on the chromosome from which DNA replication starts in either one or both directions.

Replication Fork
A moving front of DNA replication at which a double-stranded DNA helix is separated into two newly replicated helices.

DNA Polymerase
An enzyme that catalyzes the formation of DNA chain by adding deoxyribonucleotides to the 3′-hydroxyl end of a preexisting DNA strand.

Promoter
A region on a DNA molecule in which an RNA polymerase binds and initiates transcription.

Transcription Factor
A class of regulatory protein that binds to a promoter or to a nearby sequence of DNA to facilitate or prevent initiation of transcription.

RNA Polymerase
An enzyme that transcribes an RNA molecule from one strand (template strand) of a DNA molecule.

Genetic information of organisms is kept in DNA in the form of an array of nucleotide sequences, and the cell must precisely replicate its chromosomal DNA. The self-complementary nature of the double-stranded DNA molecule is the basis for accurate replication, and organisms are equipped with DNA polymerase enzyme, which adds deoxyribonucleotides to preexisting DNA chains in the $5' \rightarrow 3'$ direction, along the template strand of DNA. One strand (the leading strand) is synthesized continuously while the other (the lagging strand) is synthesized discontinuously and is then sealed.

To extract information from DNA, a limited region of the DNA, which constitutes the gene or the gene cluster (operon), is transcribed into RNA. This process is catalyzed by RNA polymerase. As is the case with DNA polymerase, this enzyme is composed of multiple subunits and catalyzes the addition of ribonucleotides to the growing RNA chain in the $5' \rightarrow 3'$ direction, along the template DNA. Both the startpoint and the frequency of transcription are determined by the promoter, a specific sequence of DNA to which RNA polymerase and transcription regulatory proteins can bind.

1 DNA Replication

Precise transmission of chromosomal DNA from generation to generation is crucial to cell propagation. This can only be achieved when chromosomal DNA is accurately replicated, providing two copies of the entire genome for faithful distribution into each daughter cell. To this end, the cell has special mechanisms to maintain the high fidelity of DNA replication, to segregate the replicated DNA, and to coordinate DNA replication and cell division tightly within the cell cycle.

Interest and research in DNA replication have been fueled by increasing concerns with mutagenesis, evolution, carcinogenesis, and cancer therapy. Spontaneous mutations, apparently a major force in evolution, arise largely from errors during DNA replication. Cancer cells, having multiple mutations in genes controlling the cell cycle, escape the tight regulation of cell cycle and start their chromosomal DNA replication abnormally. This uncoupling of cell division and DNA synthesis may frequently induce chromosome disorders and additional mutations, thus accelerating the malignancy. Most anticancer treatments involve inhibition of DNA replication in cancer cells.

Many basic concepts and techniques in modern biotechnology are derived from research on DNA replication. The development of cloning vectors has depended on the knowledge of the replication mechanisms of bacterial plasmids and phages. Genetic engineering, DNA sequencing, and the polymerase chain reaction (PCR) utilize replicative enzymes, such as DNA polymerase and DNA ligase.

1.1 Structural Aspects of DNA Replication

The double-stranded nature of DNA provides the basis for its semiconservative replication. The two strands of the parental duplex contain complementary sequence information. DNA synthesis by a polymerase also makes use of this complementarity in generating the daughter duplex. A faithful copy can be made by complementary base pairing between the free nucleotide substrate and the parental DNA strand template. Thus, the daughter duplex consists of one parental strand and one new strand. Complementary base pairing also allows for the correction of replication errors by several repair mechanisms.

1.1.1 Replicon: Unit of Replication

DNA replication does not start at random locations but at particular sites, called the *origins of DNA replication*. A DNA whose replication starts from an origin and proceeds bidirectionally or unidirectionally to terminus site(s) is called a *replicon*, a unit of DNA replication. In bacterial cells, the circular chromosome contains a unique origin and DNA replication proceeds bidirectionally from the origin to the terminus (Fig. 1(a)). Therefore, the whole bacterial genome (4700 kb for *Escherichia coli*) is a single replicon. This genome structure is also common to most bacterial plasmids and some phages, although some of them are replicated unidirectionally.

On the other hand, eukaryotic cells contain multiple replication origins on each chromosome (Fig. 1(b)). In order for the eukaryotic cell to replicate its huge genome (2900 Mb per human haploid genome) in a relatively short period (about 7 h for the S-phase in a cultured animal cell), many replicons function simultaneously. Each replicon is 40 to 100 kb in size

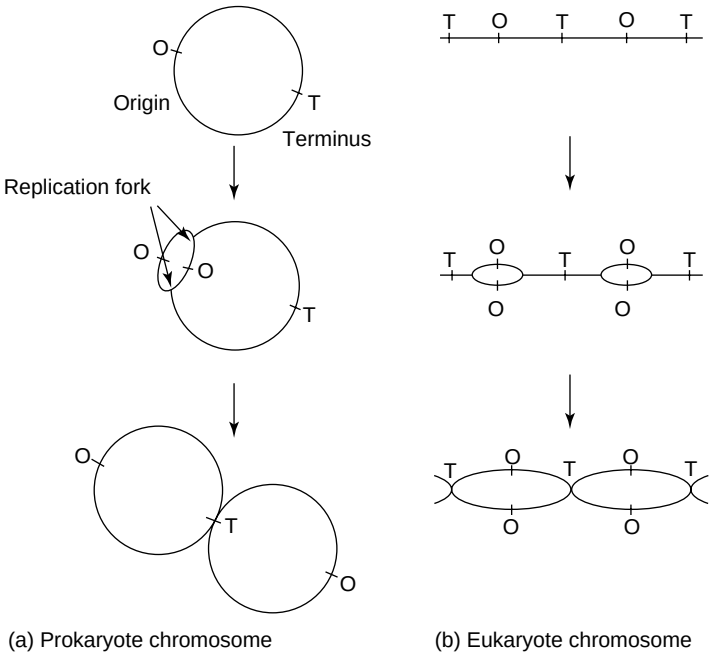
O
Origin
T
Terminus
Replication fork
O
O
T
O
T
O
T

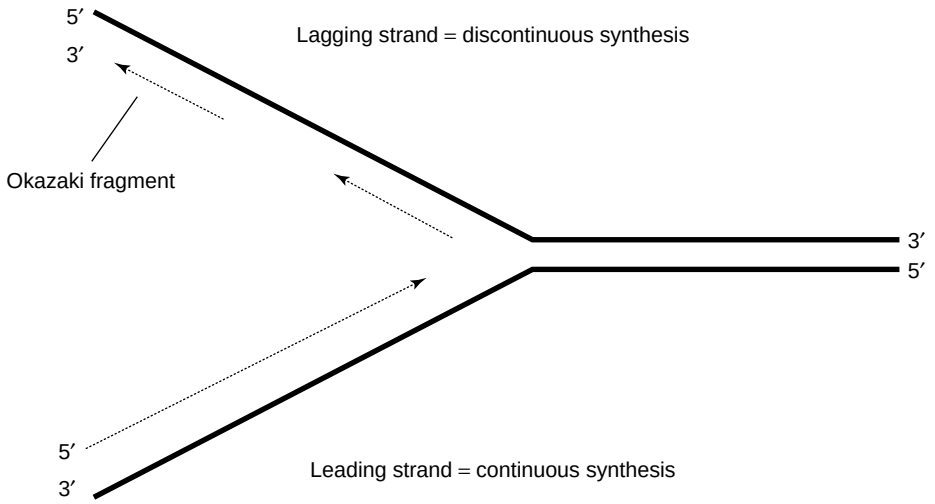

Fig. 2 Semi-discontinuous DNA synthesis at the replication fork.

1.2 Biochemistry of DNA Replication

DNA replication is a complex process involving numerous enzymes at the replication fork. In *E. coli*, more than 20 different proteins participate in DNA replication. These were identified by screening of mutants defective in DNA replication and by purifying enzymes required for *in vitro* DNA synthesis. On the basis of their biochemical roles in different stages of chromosomal DNA replication, these proteins are classified into several categories: (1) proteins involved in the initiation of replication: DnaA, HU, DNA gyrase, single-stranded DNA binding protein (SSB), DnaB, DnaC, RNaseH1, and RNA polymerase; (2) proteins required for chain elongation and connecting Okazaki fragments: DnaG (primase), DNA polymerase III (Pol III) holoenzyme, SSB, DNA polymerase I (Pol I), RNaseH1, and DNA ligase; (3) a swivel to relieve torsional stress in advance of the replication fork: DNA gyrase; (4) proteins for the termination of replication and segregation of daughter molecules: terminus binding protein (Tus), DNA gyrase, and DNA topoisomerase IV; (5) primosome proteins required for reformation and restart of replication fork when it accidentally collapses: PriA, PriB, PriC, DnaT, DnaB, DnaC, and DnaG. Among these proteins, the DnaB helicase, primase (DnaG), and Pol III holoenzyme are the basic components acting at the replication fork, forming a multiprotein complex called a *replisome* (Fig. 3).

Replication of many bacterial plasmids and small phages requires most of the host-encoded replication proteins, although the initiation step of these extra chromosomal replicons differs from that at the host origin. Large phages such as T4 and T7 encode their own replication proteins, whose activities resemble those of the *E. coli* replication proteins. In eukaryotic cells, chain elongation, termination, and segregation appear to be carried out by activities very similar to the *E. coli* replication proteins. Proteins involved in the

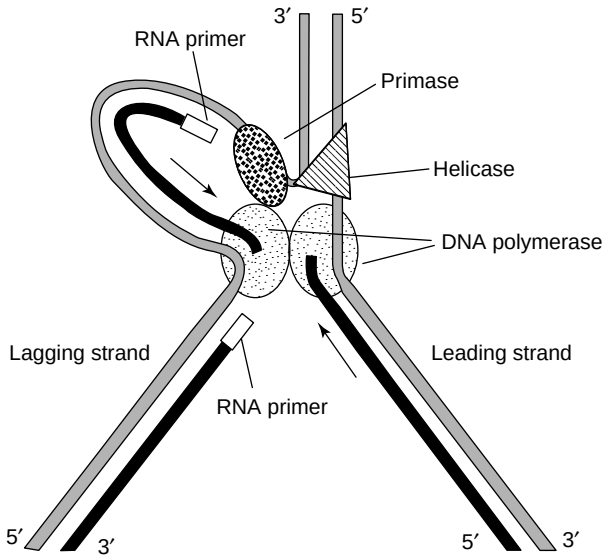

Fig. 3 Hypothetical structure of replication machinery. Primase, DNA helicase, and twin DNA polymerases form a "replisome."

initiation of chromosomal replication in eukaryotic cells have also been identified.

1.2.1 **DNA Polymerase**

DNA polymerase plays a primary role in replication, catalyzing synthesis of a DNA chain complementary to the template DNA strand as follows:

$$(\mathrm{dNMP})_n + \mathrm{dNTP} \longrightarrow$$
$$(\mathrm{dNMP})_{n+1} + \mathrm{PP_i}$$

The reaction absolutely requires three components: (1) deoxynucleosidetriphosphate (dNTP) as a substrate; (2) single-stranded DNA as a template; and (3) a short RNA or DNA as a primer. The primer must be annealed to the template by base pairing and its 3′-terminus must possess a free 3′-OH group. Chain growth is exclusively 5′ → 3′ since the polymerization mechanism is a nucleophilic attack by the 3′-OH group of the primer on the α-phosphate of the incoming dNTP. The products are a new primer, longer by one nucleotide, and an inorganic pyrophosphate. 2′,3′-dideoxy NTP can be incorporated by some DNA polymerases, but blocks chain elongation because it lacks a free 3′-OH group. This chain termination property is utilized in some DNA sequencing methods.

Both prokaryotic and eukaryotic cells possess multiple numbers of distinct DNA polymerases (Table 1), which differ in their structures as well as in their biochemical and biological functions. A 3′ → 5′ exonuclease activity is associated with all replicative DNA polymerases except the eukaryotic Pol α. The fidelity of DNA synthesis is increased by the 3′ → 5′ exonuclease activity, which allows for removal of nucleotides incorrectly inserted by the polymerase.

In *E. coli*, Pol III holoenzyme is the major replicative polymerase for both leading-

Tab. 1 DNA polymerases found in prokaryote and eukaryote.

Organism	*DNA polymerase*	*Processivity*	*Editing exonuclease*	*Function*	*Unique features*
E. coli	Pol I	Low	Yes	Repair	5′ → 3′ exonuclease associated
	Pol II	Medium	Yes	TLS	Structural homology to eukaryotic Pol α
	Pol III	High	Yes	Replication	Asymmetric dimer with twin active site
	Pol IV	Low	No	TLS	Capable of induction of frameshifts
	Pol V	Low	No	TLS	Activated by RecA protein
Mammal	Pol α	Medium	No	Replication	Primase associated
	Pol β	Low	No	Repair	Small single-polypeptide
	Pol γ	High	Yes	Mitochondrial replication	Preference to ribohomopolymer template
	Pol δ	High	Yes	Replication, repair	PCNA-dependent processivity
	Pol ε	High	Yes	Replication, repair	PCNA-independent processivity
	Pol ζ	Low	No	TLS	Mismatch extender polymerase
	Pol η	Low	No	TLS	Error-free bypass DNA synthesis
	Pol ι			TLS	Most error-prone polymerase
	Pol κ	Medium	No	TLS	Related to *E. coli* Pol IV
	Pol λ	Low	No	Repair in meiosis?	Homologous to Pol β
	Pol μ	Low	No	Somatic hyper mutation?	Expressed preferentially in lymphoid

Note: TLS: translesion DNA synthesis.

and lagging-strand synthesis. Pol I participates in lagging-strand synthesis by eliminating primer RNAs and also has a role in repair synthesis. The other three polymerases, Pol II, Pol IV, and Pol V, are involved in translesion DNA synthesis and are induced when the chromosome DNA is damaged. The Pol III holoenzyme is a huge multiprotein complex that consists of 10 distinct polypeptides. This enzyme extends the DNA chain with a high processivity (>500 kb of DNA can be

continuously synthesized without dissociation of polymerase from the template) and a high catalytic efficiency (the velocity of chain elongation is 1000 nucleotides second at 37 °C). The catalytic core, composed of α-, ε-, θ-subunits, contains the polymerase activity and a $3' \rightarrow 5'$ exonuclease for proofreading. τ-subunit connects two of the catalytic core subassemblies. The remaining six auxiliary subunits enhance the processivity of the core by clamping it onto the template. They also promote the repeated association of the polymerase necessary for discontinuous synthesis of the lagging strand. Among the auxiliary subunits, a dimer of β-subunit shows a ring-shaped structure and acts as a sliding clamp. Loading of the β-clamp onto the DNA is facilitated by a five-subunits subassembly called *γ-complex*. Structural analysis of the Pol III holoenzyme and studies on a reconstituted replication fork suggest that the holoenzyme is a functionally asymmetric dimer with twin polymerase active sites: one half of the dimer has high processivity and might be a polymerase for continuous synthesis of the leading strand; the other half has recycling capacity needed for the lagging-strand synthesis. Thus, a single molecule of Pol III holoenzyme acts at the replication fork catalyzing concurrently both leading and lagging-strand synthesis.

In eukaryote cells, nuclear DNA replication requires three polymerases, Pol α, Pol δ, and Pol ε. Pol γ is a sole polymerase participating in mitochondrial DNA replication. Pol β, a small single polypeptide, functions in base excision repair. There are many other polymerases that were recently identified. Some of them are involved in translesion DNA synthesis, but biological roles of the remaining polymerases are unknown. Among the replicative polymerases, Pol α is unique in possessing a primase activity, the only such activity so far identified in eukaryotic cells, suggesting that Pol α may play a role in the priming of DNA synthesis. However, DNA synthesis by Pol α, lacking the $3' \rightarrow 5'$ exonuclease activity, is inaccurate and shows a low processivity. These enzymatic characteristics make Pol α a poor candidate for the major replicative polymerase. On the other hand, Pol δ and Pol ε both possess a $3' \rightarrow 5'$ exonuclease activity and are able to carry out highly processive DNA synthesis with the aid of a sliding clamp (PCNA) and a clamp loader (RF-C). Therefore, these polymerases are better suited for chromosome replication. Yeast mutant strains defective in the $3' \rightarrow 5'$ exonuclease activity of either Pol δ or Pol ε increase spontaneous mutation rate, suggesting that both polymerases are responsible for chain elongation during chromosome replication. However, the division of labor between the polymerases remains ambiguous.

1.2.2 DNA Helicase and Primase

DNA polymerase cannot replicate a duplex DNA without assistance. This enzyme requires single-stranded DNA as a template and cannot synthesize a chain without an RNA or DNA primer annealed to the template. Two other enzymes enable polymerase to work on a duplex DNA. One is a DNA helicase, which opens up the duplex at the replication fork to provide a single-stranded template. The other is a primase, which synthesizes a short RNA to prime DNA chain elongation.

Several DNA helicases have been identified from *E. coli* and its phages, including DnaB, T7 gp4, and T4 gp41. Biochemical characterization of these activities *in vitro* DNA replication systems suggests that the primary replicative helicase binds

to and moves on the lagging-strand template in the 5′ → 3′ direction unwinding the duplex as it goes. This helicase action requires ATP hydrolysis. In addition, genetic studies indicate that these helicases are essential for DNA replication and growth of the cell or its phages, respectively. Another common property of the replicative helicases is an intimate association with a primase. The T7 gp4 has both primase and helicase activity within the same polypeptide. T4 gp41 greatly enhances the primase activity of T4 gp61. A similar functional interaction has been observed between DnaB protein and *E. coli* primase (DnaG protein).

Although several DNA helicase activities have been found in eukaryotic cells, the replicative helicase has not been clearly identified. Large T-antigen encoded by the SV40 virus is required for the replication of SV40 DNA and has a 3′ → 5′ helicase activity. Moreover, human DNA Pol α, which contains a primase subunit, has a specific affinity to the large T-antigen helicase. Most recently, MCM2-7 proteins that form a ring-shaped complex around DNA and that are required for chain elongation during chromosomal replication were suggested to be a replicative helicase.

Primases can start a new chain when copying a duplex DNA, but, like DNA polymerases, require a single-stranded DNA as the template. Primases can utilize dNTPs in place of rNTPs, although they absolutely require an rNTP (ATP in most cases) to initiate the primer. Under physiological concentrations, rNTPs are preferentially used during primer synthesis. In prokaryotes, primer synthesis is initiated at preferred sequences on the template: 3′-GTC for *E. coli* primase (DnaG), 3′-CTG(G/T) for T7 phage primase (gp4), and 3′-TTG for T4 phage primase (gp61). The sequence preference of the initiation of primer synthesis by eukaryotic primases is uncertain. All primases found in various organisms extend the primer chain in the 5′ → 3′ direction to a particular length: 10 to 12 nucleotides by DnaG protein, 4 nucleotides by T7 gp4, 5 nucleotides by T4 gp61, and 8 to 12 nucleotides by eukaryotic primases.

A functional interaction between primase and replicative helicase has been extensively studied using *E. coli* replication proteins. On most templates, DnaG exhibits a very feeble priming activity, which can be greatly enhanced if DnaB first binds that DNA. This stimulation of primase activity is further increased when the DnaB helicase is activated to its processive form at the replication fork. The activation of DnaB protein is a key step in the initiation of DNA replication and is catalyzed in several ways (see Sect.1.3.2).

1.2.3 Other Replication Proteins

In addition to the components of the replisome, several auxiliary proteins are needed for DNA replication. Single-stranded DNA binding protein (SSB) protects the exposed template and stabilizes the structure of the replication fork. *E. coli* mutants defective in SSB are inviable; DNA synthesis in these conditional mutants stops quickly under nonpermissive conditions. *E. coli* contains about 300 molecules of SSB per cell, at a level sufficient to cover about 1400 nucleotides of the lagging-strand template at each replication fork. Eukaryotic counterpart to the *E. coli* SSB has been identified in the *in vitro* replication system. This protein, RF-A, consists of three distinct subunits, binds tightly to ssDNA, and is required for the helicase action of the virus-encoded T-antigen at the SV40 origin.

Enzymes to remove the primer RNA and fill the gap, such as RNaseH1 and

DNA polymerase I of *E. coli*, are essential for the sealing of Okazaki fragments by DNA ligase. Mutants defective in either DNA polymerase I or DNA ligase show a massive accumulation of short Okazaki fragments under restrictive conditions. In eukaryote, a $5' \rightarrow 3'$ exonuclease called *FEN1* removes the primer RNA.

DNA topoisomerases function as swivels, to relieve torsional stress produced ahead of the replication fork. From the analysis of mutants, it appears that DNA gyrase (a type II topoisomerase from *E. coli*) acts as the swivel in prokaryotic cells. In eukaryotes, either type I or type II topoisomerases can provide the swivel action needed. Another important role of topoisomerases in replication is the decatenation of daughter molecules. This function is provided by type II topoisomerases in both prokaryotic and eukaryotic cells and allows the separation of the daughter molecules prior to segregation.

1.3 Regulation of DNA Replication

Initiation of DNA replication is a major determinant of the cell cycle, and thus is tightly regulated in all organisms. Although signals that turn on DNA replication in eukaryotic and prokaryotic cells are different, how these signals are transduced to the initiation machinery is one of the most important issues in cell biology. The mechanisms of initiation have been detailed using systems that initiate and replicate a minichromosome *in vitro*. These various mini-chromosomes contain replication origins from bacterial chromosomes (*E. coli oriC*), bacteriophages (λ *ori*, P1 phage *ori*), and plasmids (ColE1 *ori*, F plasmid *oriS*). In contrast to the prokaryotic origins, the sequences required for an origin of replication vary significantly between different eukaryotic organisms. Chromosomal origins have been isolated from yeast as well as from several eukaryotic viruses. However, attempts to identify such autonomously replicating segments from animal or plant chromosomes have been unsuccessful. Instead, a more direct approach, using electron microscope and two-dimensional gel electrophoresis techniques, has been taken to identify replication origins in the higher eukaryotes. Although it has been difficult to develop *in vitro* systems for eukaryotic DNA replication, genetic studies with yeast mutants that affect initiation processes of chromosomal replication have successfully revealed the initiation mechanisms of eukaryotic DNA replication.

1.3.1 Initiation Mechanisms in Prokaryotes

A DNA segment carrying a replication origin can be isolated as a mini-replicon. DNA fragments can be linked to a selectable marker and introduced into the cell. Those fragments that allow for the maintenance of the marker are likely to contain an origin for autonomous replication of the DNA. Analysis of deletion and base-substitution mutants of the mini-replicon for their replication capacity will define a minimal region of the replication origin and essential motifs in that region. Replication origins that have been examined in this way can be classified into two groups from their structural similarities (Fig. 4). One class, to which most prokaryotic replicons belong, consists of 4 to 7 repeats of an initiator-protein binding site and an AT-rich region. The other group (ColE1 plasmid *ori*, T7 primary *ori*, T4 primary *oriA*) carries a transcriptional promoter instead of the initiator-protein binding sites.

The reconstitution of initiation in the test tube has elucidated several events during this process as follows: (1) recognition of the origin; (2) opening of a particular region in the origin; (3) loading of a replicative helicase onto the single-stranded DNA region, and (4) initial priming for leading- and lagging-strand synthesis. Some bacterial plasmids and phages are replicated unidirectionally. A major determinant for uni- or bidirectionality is the number of replicative helicases loaded at the origin. One helicase can form one replication fork. Thus, bidirectional replication involves two replicative helicases.

The initiation mechanism at the origin carrying the initiator-protein binding site has been depicted from studies on *in vitro* replication of *oriC* plasmid DNA (Fig. 4(a)). At least eight proteins are required for the *in vitro* reconstitution of the *oriC* minichromosome. A key event is the formation of the initial complex in which the initiator protein (DnaA protein) tightly binds to its 9 bp recognition sequence in *oriC*. Following this stage, an AT-rich region is opened in an ATP-dependent manner, and the DnaB helicase and SSB are loaded onto the single-stranded region to form the prepriming complex. Primase and DNA polymerase III holoenzyme then start the DNA synthesis. Several plasmids (F-factor and R-factor) and bacteriophages (λ, P1) in this class of replicon encode their own initiator proteins that specifically recognize and bind the origin sequence. Most of these origin-binding proteins can open the duplex DNA by themselves but fail to load the DnaB helicase. To do this, they contain one or two DnaA recognition sequences in their origins and utilize the DnaA protein as a landmark for entrance and activation of the replicative helicase.

Replicons without an initiator protein use transcription for origin recognition and duplex opening (Fig. 4(b)). Initiation of ColE1 plasmid replication requires RNA synthesis started 555 bp upstream from the transition point between RNA and DNA. The transcript (RNA II) forms an RNA–DNA hybrid around the transition point, is processed by RNaseH1, and serves as the initial primer for the leading strand DNA synthesis. DNA pol I extends the leading-strand DNA by about 400 nucleotides. This initial chain elongation, resulting in duplex opening, does not involve any helicase action but

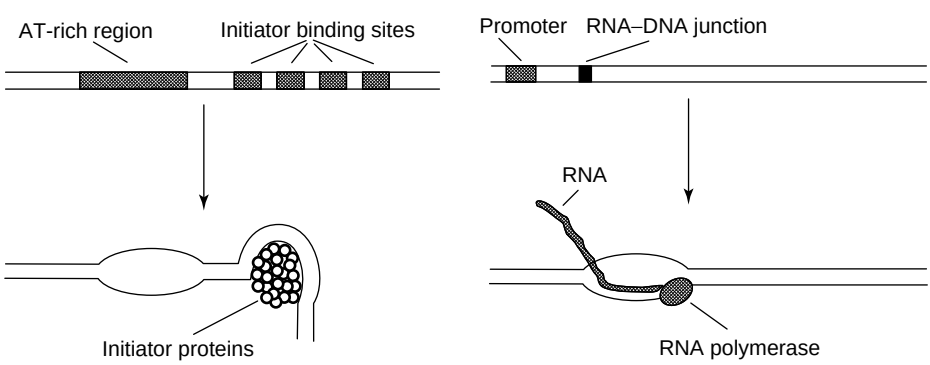

(a) Origin with initiator-protein binding site (b) Origin with transcriptional promoter

Fig. 4 Two types of replication origins and related mechanisms of initiation.

requires that the template be negatively supercoiled. The loading of DnaB helicase is facilitated by primosome proteins, which are assembled into a complex on a specific DNA segment (primosome assembly site) about 150 bp downstream from the origin.

1.3.2 Initiation Mechanisms in Eukaryotes

The yeast *Saccharomyces cerevisiae* is the only eukaryote in which there is a detailed understanding of the sites used for initiation of chromosomal DNA replication. In yeast, these sites are called *ARS*s (autonomously replicating sequences) because they were initially identified by their ability to confer high-frequency transformation and self-replication on plasmids introduced into cells. By these criteria, there are 200 to 400 ARS elements in the yeast genome. ARS elements are relatively small, about 100 to 150 bp, and consist of one or more copies of an essential 11 bp-long AT-rich ARS consensus sequence, as well as several other less conserved elements. In other organisms, the sequences required to direct the initiation of DNA replication are more complex. In the fission yeast, *Schizosaccharomyces pombe*, initiation of replication requires origin sequences spread over at least 800 to 1000 bp. Although these sequences contain several AT-rich sequences of 20 to 50 bp that are important for origin function, they do not exhibit the strong sequence similarity observed for the *S. cerevisiae* ARS elements. Metazoan origins are still less well defined and can extend over thousands of base pairs of DNA. In addition, the sites of initiation are not always tightly linked within these regions.

Eukaryotic origins of replication direct the formation of a number of protein complexes leading to the assembly of two bidirectional DNA replication forks. These events are initiated by the formation of the pre-replicative complex (pre-RC) at the origins of replication during G1 (Fig. 5). Pre-RC formation involves the ordered assembly of a number of replication factors including ORC, Cdc6, Cdt1, and Mcm2–7 proteins. The ARS consensus sequence is the binding site for the multisubunit origin recognition complex (ORC), which binds constitutively throughout the cell cycle and is essential for initiation. Several proteins are recruited to the ARS during the G1 phase of the cell cycle to form a pre-RC, including the multisubunit minichromosome maintenance (MCM) complex, which has ATPase and

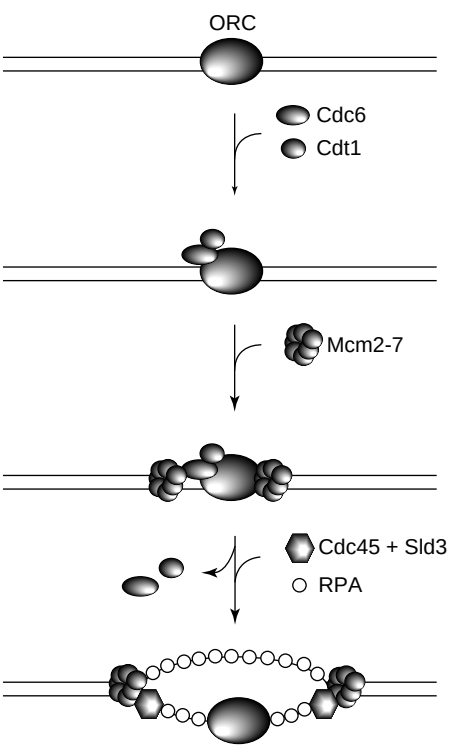

Fig. 5 Initiation mechanisms of eukaryotic DNA replication.

helicase activity and is needed for both initiation and fork progression. The regulation of pre-RC formation is a key element of the mechanisms coordinating DNA replication with the cell cycle. Once formed, this complex awaits activation by at least two kinases that trigger the transition to DNA replication. As with the formation of the pre-RC, the transition to replication involves the ordered assembly of additional replication factors that facilitate unwinding of the DNA at the origin and culminates in the association of the multiple eukaryotic DNA polymerases with the unwound DNA.

2 Transcription

The transfer of information from DNA to protein begins with the synthesis of RNA molecules in a process called *transcription*. In this process of transcription, the genetic information carried in DNA is transferred to several kinds of RNA molecules (mRNA, tRNA, rRNA, and others), so that an mRNA encoding a polypeptide in its turn translates the information in the four-letter language of the nucleic acids to the twenty-letter language of the amino acids with the coordinated actions of tRNA and rRNA as part of the mechanism of protein synthesis.

2.1 RNA Polymerase and Transcriptional Apparatus

DNA-dependent RNA polymerase, which catalyzes the polymerization of RNA from ribonucleotides in a DNA-dependent manner, is the enzyme responsible for the transcription of DNA into RNA. The RNA polymerase binds to a specific DNA sequence, called the *promoter*, and unwinds the duplex DNA for about one turn of the helix to expose a short stretch of single-stranded DNA so that complementary base pairing can be made with the incoming ribonucleotides. The enzyme joins two of the ribonucleoside triphosphate monomers and then moves along the DNA strand, extending the growing RNA chain in the $5' \rightarrow 3'$ direction until it encounters a second special sequence, called the *terminator*, which signals where the RNA synthesis should stop. After the transcription is completed, each RNA chain is released from the DNA template as a free, single-stranded RNA molecule. In addition to RNA polymerase, other protein factors are required for efficient transcription, and they are responsible for determining the transcription efficiency of specific transcription units.

2.1.1 General RNA Polymerase Architecture

RNA polymerases are large multisubunit enzymes, which consist of 5 to 15 subunits and whose sizes are up to 0.6 MDa. In eubacteria and archaebacteria, a single type of RNA polymerase synthesizes all classes of RNA molecules, except for the primer RNA, which is synthesized by primase. The bacterial RNA polymerases consist of 5 subunits (2 α-subunits and one each of β, β', and ω), and a few more transcription factors, such as σ, ρ, and NusA for *E. coli* RNA polymerases, are required to initiate transcription from the proper start site in the promoter of duplexed DNA template (Table 2).

There are three known distinct types of RNA polymerases in eukaryotic cells. RNA polymerase I (pol I) makes ribosomal RNA and pol II mostly makes messenger RNA, while pol III makes small RNA molecules, such as transfer

Tab. 2 Subunit composition of *E. coli* RNA polymerase and transcription factors.

Subunit	*Gene*	*Number of amino acids*	*Molecular weight [Da]*	*Function*
α	*rpoA*	329	36 512	Connecting $\beta\beta'$ subunits
β	*rpoB*	1342	150 618	Catalyzing RNA synthesis substrate nucleotide binding
β'	*rpoC*	1407	155 163	Template binding and association with σ-subunit
ω	*rpoZ*	91	10 230	Regulation of transcription
σ70	*rpoD*	613	70 263	Recognition of general promoters
ρ	*rho*	419	46 974	Termination
NusA	*nusA*	494	54 536	Elongation, termination

Tab. 3 Eukaryotic RNA polymerases[a].

Polymerase	*Location*	*Genes transcribed*	*Polymerase activity in cells [%]*
pol I	Nucleolus	Genes for 28S, 5.8S and 18S ribosomal RNAs (rRNAs) Protein-coding genes	50–70
pol II	Nucleoplasm	Most genes (U1, U2, U4, U5) of small nuclear RNAs (snRNAs) Genes for tRNAs, 5S rRNA	20–40
pol III	Nucleoplasm	Small nucleolar RNAs (snoRNAs) Small cytopasmic RNAs (scRNAs)	10

[a]Mitochondria and chloroplast contain distinctive RNA polymerases.

RNA, 5S ribosomal RNA, and a few other small RNA molecules (Table 3). Among eukaryotic RNA polymerases, yeast *S. cerevisiae* RNA polymerases have so far been the most extensively characterized. Each RNA polymerase molecule consists of two large subunits, 160 ~ 200 kDa and 130 ~ 150 kDa, which share amino acid sequence homology with the β'- and β-subunits of the bacterial enzymes respectively, and a collection of smaller polypeptides. In pol II, Rpb3 and Rpb11 share amino acid sequence homology with the bacterial α-subunit, while Rpb6 shares a homologous sequence with the ω-subunit. Among the three types of eukaryotic RNA polymerases, five subunits (Rpb5, 6, 8, 10, and 12) are shared. pol I and pol III share two additional subunits (AC19 and AC40), both of which share amino acid sequence homology with the bacterial α-subunit.

As shown in Fig. 6, *S. cerevisiae* pol II and *Thermus aquaticus* RNA polymerase share a general architecture, which consists of five core subunits, β', β, α/α, and ω or their counterparts in the pol II (Table 4).

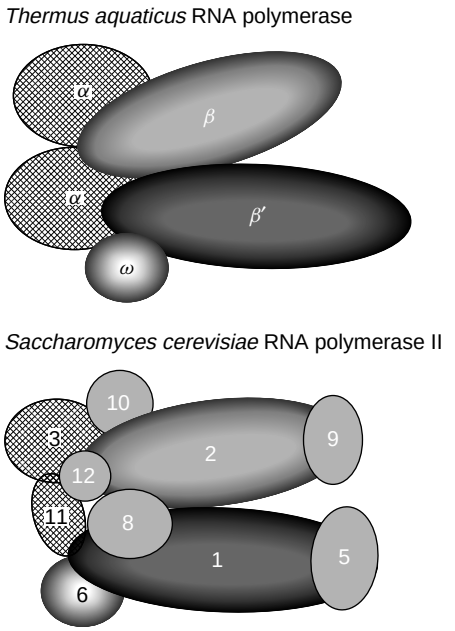

Fig. 6 General architecture of RNA polymerases. Schematic presentation for RNA polymerase subunits in bacterial (α, α, β, β', ω) and eukaryotic enzymes (Rbp1–12).

The bacterial β' and β subunits or the yeast Rpb1 and 2 subunits form the central masses with a deep cleft between them, respectively. α-homodimer in the bacterial enzyme or Rpb3–Rpb11 heterodimer in the yeast pol II anchor the two large subunits, while the fifth subunit ω and Rpb6 support and stabilize either the large subunit β or Rpb1.

2.1.2 Transcription Initiation

None of the eukaryotic RNA polymerases can initiate transcription by themselves; rather, they require a group of general transcription initiation factors for initiation at proper sites. pol II requires several general initiation factors (TFIIA, IIB, IID, IIE, IIF, IIH), and the formation of a functional preinitiation complex on the core promoters, which are defined as the minimal sequences required to recruit the appropriate RNA polymerase and initiate transcription, involves ordered

Tab. 4 Subunit structure of RNA polymerase II.

Subunit	*Amino acid residues (% identity to human)*			
	Human	*S. cerevisiae*	*A. thaliana*	*Notes*
Rpb1	1970	1733(49)	1860(55)	[a]Bacterial β'
Rpb2	1174	1224(56)	1188(65)	[a]Bacterial β
Rpb3	275	318(39)	319(39)	[a]Bacterial α
Rpb4	142	221(23)	138(37)	
Rpb5	210	215(44)	205(54)	[b]RNA pol I, II, III
Rpb6	127	155(47)	144(47)	[a]Bacterial ω
Rpb7	172	171(43)	176(53)	
Rpb8	150	146(35)	146(47)	[b]RNA pol I, II, III
Rpb9	125	122(38)	114(58)	
Rpb10	67	70(69)	71(75)	[b]RNA pol I, II, III
Rpb11	117	120(46)	116(54)	[a]Bacterial α
Rpb12	58	70(33)	51(53)	[b]RNA pol I, II, III

[a]Homologous subunit in bacterial RNA polymerase.
[b]Common subunits for eukaryotic RNA polymerase I, II, III.

interactions of an array of general initiation factors on the core promoters (Fig. 7).

TFIID, a high molecular weight complex comprising TATA-binding protein (TBP) and more than 10 TBP-associated factors ($TAF_{II}s$), binds to the TATA box of the promoter; then TFIIA and B enter to form the DAB complex. The non-phosphorylated form of RNA polymerase II (pol IIa), together with TFIIF, binds to the DAB complex on the promoter, followed by association of TFIIE and TFIIH to form a closed preinitiation complex. The TFIIH XPB DNA helicase, composed

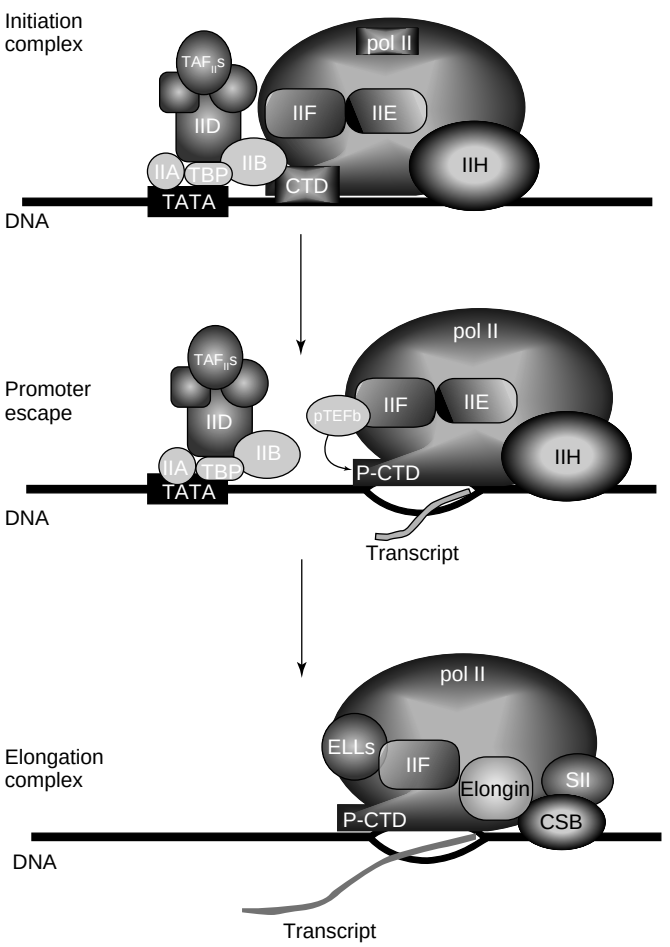

Fig. 7 Formation of a functional initiation complex on a TATA-containing promoter by general initiation factors. After completion of the final initiation complex, pol II initiates transcription and moves away from the promoter and dissociates from the DAB complex after synthesizing 10- to 15-nucleotide-long transcripts (promoter escape). During the promoter escape step, the elongation complex is released from the promoter to start elongation by phosphorylation of the CTD by pTEFb. P-CTD: phosphorylated CTD.

of nine subunits, and two of which, XPB and XPD, are catalytic subunits for the DNA helicase, converts the closed preinitiation complex to both an initiation- and an escape-competent transcriptional intermediate (an open complex) in a single ATP-dependent step, probably by unwinding the DNA downstream from the transcription start site, and the synthesis of the first phoshodiester bond of nascent transcript begins.

RNA polymerase III requires different combinations of transcription initiation factors to form a preinitiation complex on the promoters of 5S RNA, tRNA, and other small RNA genes. Three pol III specific initiation factors, TFIIIA, B, and C, are required for transcription of the 5S RNA genes, TFIIIB and C for the tRNA genes, and TFIIIB and TFIID for the U6 gene. TFIII A and C, as well as TFIID, recognize and bind to specific DNA elements while TFIIIB has by itself no sequence–specific interaction with DNA. RNA polymerase I also requires several transcription factors to form an active initiation complex on the promoter of rDNA.

2.1.3 Promoter Escape

After initiating transcription, pol II moves away from the promoter and dissociates from the DAB complex after synthesizing 10- to 15-nucleotide-long transcripts (promoter escape, Fig. 7).

The largest subunit of pol II (Rpb1) contains an unusual repetitive domain, which is not found in the homologous large subunits of pol I and III or in the bacterial β'-subunit. This carboxyl-terminal domain (CTD), called the *tail* of pol II, consists of multiple heptapeptide repeats of the consensus sequence YSPTSPS. The heptapeptide sequence is found in the largest subunits of most eukaryotic RNA polymerase II. The deletion of most or all of the CTD is lethal for yeast, *Drosophila*, and mouse cells, thereby indicating that the CTD has an essential role in transcription. The CTD of RNA polymerase II in the initiation complex is hypophosphorylated, while pTEFb (positive transcription elongation factor b), a cyclin-dependent CTD kinase complex composed of CDK9 and one of several cyclins such as T1, T2, and K, highly phosphorylates the CTD during the step of promoter escape, and thus promotes a dissociation of the elongation complex of pol II from the DAB complex.

2.1.4 Transcription Elongation

During RNA chain extension, RNA polymerase II forms a stable elongation complex, and in the complex downstream DNA duplex enters the enzyme between a pair of mobile "jaws" formed by Rpb1, Rpb5, and Rpb9 subunits and then extends through the cleft toward the active site for RNA synthesis with Mg^{2+}. The DNA–RNA hybrid extends upward beyond the active site and the template strand is locked near the catalytic site in place by a clamp formed by parts of, Rpb1, Rpb2, and Rpb6. In the active site, the nontemplate strand of DNA is held within the Rpb2 subunit and the RNA transcript; it then extrudes through the complex via a channel formed by parts of both of the two large subunits, and the growing RNA end is located above the pore/secondary channel, from which nucleoside triphosphates (NTPs) may be supplied during RNA synthesis (Fig. 8).

mRNA synthesis is carried out efficiently in eukaryotic cells at rates of 1200 to 1500 nucleotides/min. However, purified mammalian pol II extends the RNA chain *in vitro* at rates of only 300 to 400 nucleotides/min, thus indicating

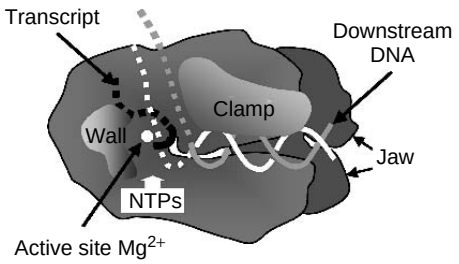

Fig. 8 The RNA polymerase II elongation complex. During transcript (black line) extension, RNA polymerase II forms a stable elongation complex in which a downstream DNA duplex (gray lines) enters the enzyme between a pair of mobile "jaws" and extends through the cleft toward the active site for RNA synthesis. The DNA–RNA hybrid extends upward beyond the active site and the template strand (white line) is locked near the catalytic site. The pore/secondary channel where the nucleoside triphosphates (NTPs) enter is shown.

that the elongation step of RNA synthesis by eukaryotic RNA polymerase II *in vivo* has to be supported by various transcription factors or elongation factors to achieve higher rates of elongation. Furthermore, unusual template structures, damage on the template, or imbalance in the cellular ribonucleotide pool, which cause either a pause or an arrest of transcription, can be overcome by factors that directly or indirectly interact with the elongation complex. Pausing occurs when RNA polymerase II undergoes modest and reversible backsliding on the DNA template, thus resulting in a misalignment of 3′-OH terminus of the nascent transcript with the catalytic site, while irreversible backsliding of the enzyme causes arrest since the misalignment becomes even more severe.

Elongation factors fall into two broad functional classes based on their abilities to either prevent transient pausing by pol II or reactivate arrested pol II. The majority of elongation factors including ELLs, Elongin, and CSB are capable of suppressing transient pausing by pol II. TFIIF, one of the general transcription factors also has an activity to suppress transient pausing. In addition, the phosphorylation of the pol II CTD by pTEFb is required to support the efficient elongation by pol II, as well as for the promoter escape, as described above (Fig. 7).

Among many elongation factors, only the SII family elongation factor has a capacity to reactivate the arrested pol II by triggering the polymerase-associated endoribonuclease that cleaves the nascent transcript upstream of its 3′-OH terminus, thus resulting in the formation of a new 3′-OH terminus. The newly generated 3′-OH, which is correctly positioned to the catalytic site of the enzyme, can thus be reextended.

During transcript elongation, pol II is hypothesized to face a major obstacle-nucleosome in which the DNA is tightly wrapped into the chromatin. It is likely that a class of chromatin elongation factors, such as an elongator, alters the structure of the nucleosome through histone acetylation, chromatin remodeling, and histone dissociation, thereby facilitating the elongation step by pol II.

2.2 Transcription Unit

Transcription takes place in limited regions of genomic DNA, and only one of the two DNA strands is used as a template. The promoter, an oriented DNA sequence along the template DNA, determines the startpoint of the region to be

transcribed and also determines which of the two strands is copied. A transcription unit extends to the terminator, at which the RNA polymerase stops adding nucleotides to the growing RNA chain. The critical feature of the transcription unit is that it constitutes a stretch of DNA expressed via the production of a single RNA molecule. A transcription unit may include only one or several genes. A typical transcription unit of *E. coli* and various transcription units for three eukaryotic RNA polymerases are shown in Fig. 9.

2.2.1 Bacterial Transcription Unit

Bacterial promoters are identified by two short conserved sequences centered at −35 and −10 relative to the startpoint. In *E. coli*, TTGACA and TATAAT are consensus sequences for the −35 and the −10 boxes of the promoters, respectively, which are recognized by the RNA polymerase holoenzyme carrying σ^{70}. Promoters whose −10 and −35 sequences closely approximate the consensus sequences are strong promoters, and the distance separating the two sequences, usually 16 to 18 bp, is also important. Bacterial RNA polymerase terminates transcription at two types of sites; factor-independent sites contain a GC-rich hairpin followed by a run of U residues, while ρ factor–dependent termination requires ρ-specific sequences.

2.2.2 Eukaryotic Transcription Unit

In eukaryotes, the promoters recognized by each type of RNA polymerase are distinct. The promoters for RNA polymerase I are located upstream from the transcription startpoint, and the sole product of transcription by pol I is a large precursor that contains the sequences of the major rRNA (28S, 5.8S, 18S). Termination occurs at a discrete site and apparently requires few factors.

The promoter elements that are necessary and sufficient for the specific initiation by RNA polymerase II with its general factors are referred to as either minimal or core promoter elements, and are located around the transcription startpoint. The most common core elements are the TATA box, which is present about 20 to 30 bases upstream of the transcription start site, the initiator element (Inr) located at the start site, and a downstream promoter element (DPE), which is present about 30 bases downstream of the transcription start site. Specific genes may contain at least one or all of these elements, which may function in conjunction with other cis-elements such as UAS (upstream activating sequence) or enhancers, which positively regulate transcription initiation.

The termination capacity of RNA polymerase II is regulated through the polyadenylation of the transcript. The signal sequence for polyadenylation, AAUAAA, is located 10 to 30 nucleotides upstream of the polyadenylation site, which is often immediately after the dinucleotide 5′-CA-3′ and is followed 10 to 20 nucleotides later by the GU-rich region. When the elongation complex of pol II, to which the cleavage and polyadenylation specificity factor (CPSF) and the cleavage stimulation factor (CstF) bind through the CTD, transcribes across the polyadenylation signal sequence, CPSF binds to the signal sequence and CsTF binds GU-rich sequence on the transcript. Next, the transcript is cleaved to create a new 3′ end, to which the poly (A) tail is added by poly(A) polymerase associated with the bound CPSF and CstF, followed by the termination of transcription.

For RNA polymerase III, there are two types of promoters: the promoters for U6

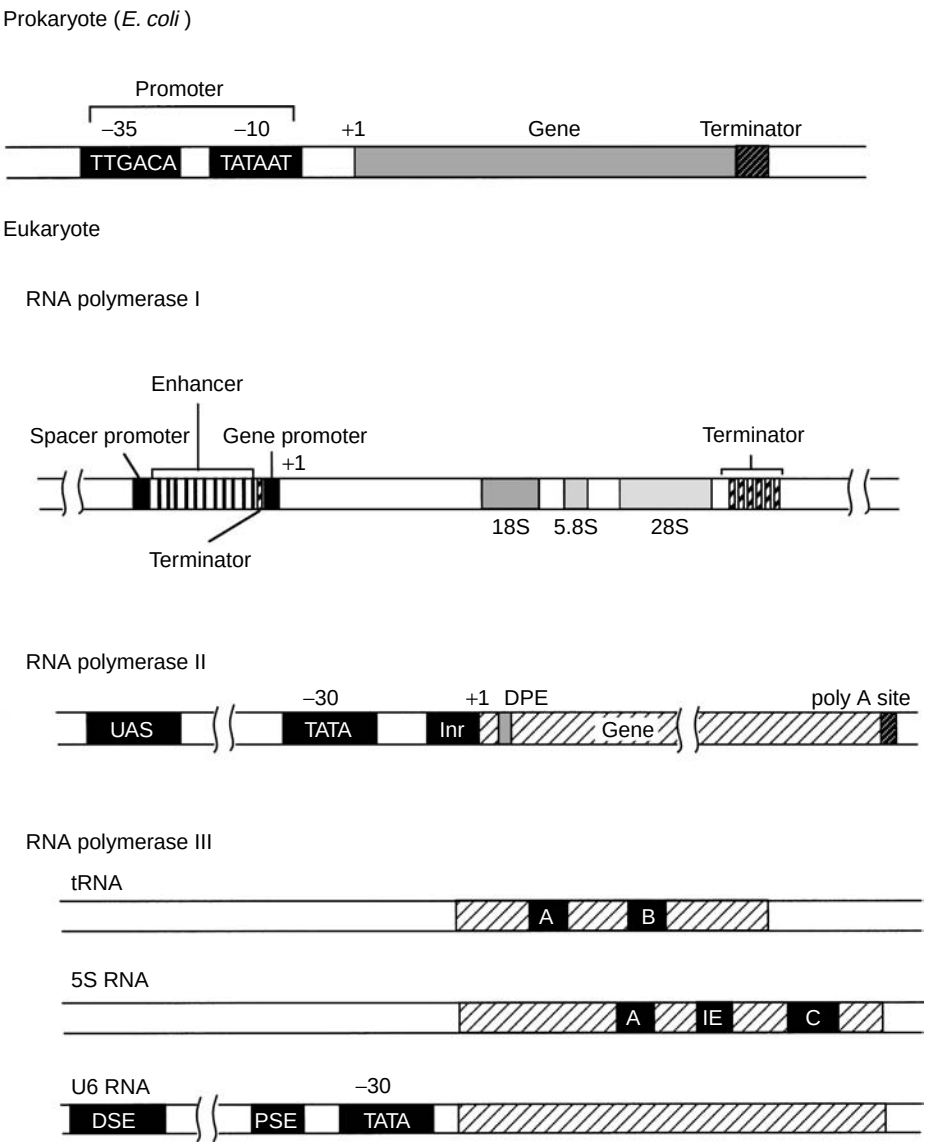

Fig. 9 Various types of transcriptional units in prokaryotic and eukaryotic cells. Inr, the initiator element; UAS, upstream activating sequence; DPE, a downstream promoter element; A, B, IE, C, intragenic cis-elements in genes for tRNA and 5S RNA; DSE, the distal sequence element; PSE, proximal sequence element.

are upstream from the startpoint, while those of the 5S RNA and tRNA genes lie downstream from the startpoint. The external promoter for the U6 gene consists of the TATA box and the distal and proximal sequence elements (DSE and PSE), which enhance promoter activity. Intragenic A and B boxes in the tRNA

gene or the B box in the 5S RNA gene can both be recognized by, and bound by, TFIII C, while the IE and C boxes in 5S RNA are involved in interaction with TFIIIA. Transcription by RNA polymerase III terminates within a run of 4 U residues embedded in a GC-rich region, in a manner similar to that observed in the prokaryotic factor-independent terminators.

2.3 Regulation of Transcription

Transcriptional regulation is the primary means of controlling gene expression, and it is essential for the developmental program of multicellular organisms, as well as for the switching on and off of gene expression in response to intra- or extracellular signals. In any organism, RNA synthesis can be regulated by mainly controlling the efficiency of initiation.

2.3.1 Transcriptional Regulation in Bacteria

In bacterial cells, there are several different mechanisms by which the modulation of transcription initiation is achieved. Various σ-factors modify the binding ability of RNA polymerase to specific promoter sequences, thus enabling it to initiate at a different set of promoters.

In a set of genes whose expression is coinduced in response to intra- or extracellular signals, a common repressor protein binds to specific DNA sequences, called an *operator*, in a group of promoters and represses their transcription by inhibiting RNA polymerase binding to the promoters. After responding to a specific signal, such repressor molecules are inactivated, and the RNA polymerase can thereby initiate transcription from the promoters. For example, the LexA protein, which represses the SOS-regulon, is proteolytically inactivated by an activated form of the RecA protein in response to DNA damage. The modulations of genes (an operon) involved in the biosynthesis of a specific metabolite are mediated by the repressors that directly sense the intracellular status of the product. For example, the LacI repressor protein binds to the operator region of the *lac* operon promoter in the absence of lactose or its metabolite, thereby prohibiting transcription initiation on the promoter, and it dissociates from the promoter after binding to a metabolite of lactose, thus allowing for the transcription of the *lac* operon.

Promoters requiring positive activator proteins for efficient expression are characterized by a poor fit to the consensus sequence in the −35 region. The protein activators bind to the region near the −35 box, and stimulate initiation either by improving the affinity of RNA polymerase for the sequence or by increasing the rate of isomerization of the initiation complex. The complex of cAMP with catabolite activator protein (CAP) enhances the affinity of RNA polymerase to the *lac* promoter about 20-fold. Ada-protein-accepted alkyl groups from DNA damaged by alkylating agents as a repair reaction, in turn, strongly bind to Ada boxes in the promoters of Ada regulon, such as *alkA, aidB* and *ada* itself, and enhance the affinity of RNA polymerase to the promoters, thus establishing an adaptive status to the alkylating agents.

In addition to the regulation of initiation, the level of transcription can be regulated by the premature termination of transcription or attenuation. For example, in most bacterial amino acid biosynthesis operons, an early termination of transcription occurs at the attenuator site that lies from 100 to 200 bases downstream of the start site, when an abundance of the charged

cognate tRNA signals a sufficiency of the amino acid. Attenuation in these operons is regulated in a complex mechanism involving charged cognate tRNA, ribosome, and ρ-protein.

2.3.2 Transcriptional Regulation in Eukaryotes

In multicellular organisms, each cell selects a set of genes from their genome to be expressed at a given time and under specific stimuli, and this selectivity and specificity of gene expression is a fundamental mechanism for controlling the cell fate such as cell proliferation, differentiation, and cell death as well as cellular functions. In eukaryotic cells, transcription by RNA polymerase II, a process for the synthesis of mRNA encoding polypeptides, which are major determinants of diversity of each cell, is regulated more dynamically than transcription by pol I and III that provide rather common elements, such as rRNA and tRNA for cell function.

The genes in the nuclear genome are usually packaged into chromatin consisting of nucleosomes and other chromatin proteins, and their promoters are transcriptionally inactive because of their limitations for access by the transcription apparatus (Fig. 10). An assortment of regulatory elements for RNA polymerase II transcription can be scattered both upstream and downstream of the transcription start site for a gene. Each gene has a particular combination of positive and negative regulatory cis-elements that are uniquely arranged regarding the number, type, and spatial array. Usually, cis-elements are arrayed within several hundred base pairs of the initiation site, but some elements can exert their control over much greater distances (1 to 30 kb). These elements are binding sites for sequence-specific DNA binding proteins, such as AP-1 (JUN/FOS heterodimer), that either directly or indirectly mediate activation or repression of transcription of the gene. The region near the startpoint, in particular the TATA box (if there is one), is responsible for selecting the exact startpoint and it is called the *promoter*, as described above. The elements further upstream determine the efficiency with which the promoter is used and are known as *UAS* (*upstream activating sequence*) or enhancers (Fig. 9).

In order to transcribe a particular gene, the chromatin structure has to be remolded, thus enabling various DNA binding proteins and transcription apparatuses to access a particular promoter (Fig. 10). First, inactive genes are localized to condensed chromatin domains or so-called heterochromatin compartments, in which each promoter is assembled in regularly spaced nucleosomes. A sequence-specific DNA binding protein binds to a particular cis-element, and then recruits a specific chromatin-remodeling complex such as SWI/SNF with other proteins through protein–protein interaction. The chromatin-remodeling complex alters the twist and writhe of nucleosomal DNA thereby affecting its accessibility to DNA binding proteins. Following chromatin remodeling, the prebound protein or another DNA binding protein on the second cis-element recruits another class of chromatin modifiers with histone acetyltransferase (HAT) activity, such as PCAF (p300/CBP-associated factor), CBP (CREB binding protein), and p300. HATs acetylate specific lysine residues in histone amino–termini and weaken the interactions between the histone octamer and DNA, thus further facilitating decondensation of nucleosome structure. The retention of SWI/SNF complex on nucleosomal promoter is significantly stabilized by histone acetylation.

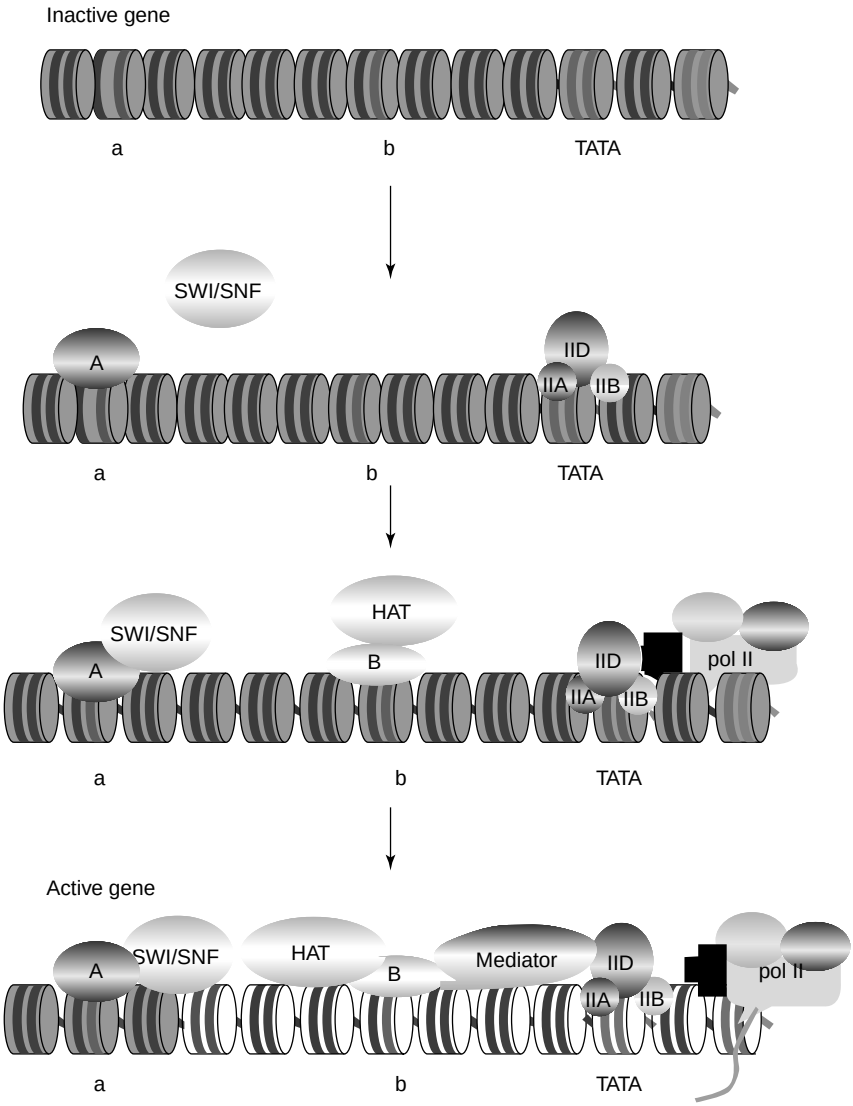

Fig. 10 Transcription activation in eukaryotes. Genes in the nuclear genome are usually packaged into chromatin consisting of nucleosomes and other chromatin proteins. In order to transcribe a particular gene, the chromatin structure has to be remolded, and thus various DNA binding proteins (A and B) can bind cis-elements (a and b) and further promote the formation of the transcription initiation complex on a particular promoter (TATA).

In addition to the acetylation of histones, phosphorylation and methylation of histones or non-histone proteins are likely to play important roles in both chromatin modification and control of gene expression.

Either during or after the remodeling or modification steps of chromatin, a

partial initiation complex (TFIIA/B and TBP) or a complete initiation complex (TFIIA/B, TBP/TFIID, and pol II) with other general factors are recruited. Again, the assembly of the initiation complex is facilitated through histone modification by the promoter-targeted HAT. Finally, a mediator complex such as DRIP, ARC, TRAP, and/or coactivator such as p300/CBP or p160/SRC are recruited to the initiation complex through protein–protein interaction with DNA binding proteins such as CREB or nuclear receptors, which specifically bind to one of the cis-elements on the promoter. The mediator or coactivator is likely to be a bridge between a given DNA binding protein or activator and pol II via direct associations with specific subunits of the complex, thus ensuring the formation of a functional preinitiation complex on the core promoter.

The repression of a certain gene expression is also regulated through the modification of chromatin by histone deacetylases (HDACs), which remove acetyl groups at lysine residues in histones, which are transfered by HATs. The deacetylation of histones causes a particular region of chromatin to be condensed as heterochromatin, which thus results in the repression of gene expression.

See also Alternatively Spliced Genes; Genetics, Molecular Basis of; Genomic Sequencing (Core Article).

Bibliography

Books and Reviews

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2002) *Molecular Biology of the Cell*, 4th edition, Garland Publishing, New York.

Bell, S.P., Dutta, A. (2002) *Annu. Rev. Biochem.* **71**, 333–374.

Brown, T.A. (2002) *Genomes*, 2nd edition, BIOS Scientific Publishers, Oxford.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., Kornberg, R.D. (1999) Architecture of RNA polymerase II and implications for transcription mechanism, *Science* **288**, 640–649.

Cramer, P. (2002) Multisubunit RNA polymerases, *Curr. Opin. Cell Biol.* **12**, 80–97.

Dvir, A., Conaway, J.W., Conaway, R.C. (2001) Mechanism of transcription initiation and promoter escape by RNA polymerase II, *Curr. Opin. Genet. Dev.* **11**, 209–214.

Featherstone, M. (2002) Coactivators in transcription initiation: here are your orders, *Curr. Opin. Genet. Dev.* **12**, 149–155.

Kornberg, A., Baker, T.A. (1992) *DNA Replication*, 2nd edition, Freeman, New York.

Lindahl, T., Sedgwick, B., Sekiguchi, M., Nakabeppu, Y. (1988) Regulation and expression of the adaptive response to alkylating agents, *Annu. Rev. Biochem.* **57**, 133–157.

Marians, K.J. (1992) Prokaryotic DNA replication, *Annu. Rev. Biochem.* **61**, 673–720.

Nakabeppu, Y., Ryder, K., Nathans, D. (1988) DNA binding activities of three murine jun proteins: Stimulation by Fos, *Cell* **55**, 907–915.

Rachez, C., Freedman, L.P. (2001) Mediator complexes and transcription, *Curr. Opin. Cell Biol.* **13**, 274–280.

Waga, S., Stillman, B. (1998) *Annu. Rev. Biochem.* **67**, 721–751.

Zahng, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., Darst, S.A. (1999) Crystal structure of thermus aquaticus core RNA polymearse at 3.3 A resolution, *Cell* **98**, 811–824.

Primary Literature

Andrulis, E.D., Werner, J., Nazarian, A., Erdjument-Bromage, H., Tempst, P., Lis, J.T. (2002) The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila, *Nature* **420**, 837–841.

Arai, K., Arai, N., Shlomai, J., Kornberg, A. (1980) Replication of duplex DNA of phage

ϕX174 reconstituted with purified enzymes, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3322–3326.

Bar-Nahum, G., Nudler, E. (2001) Isolation and characterization of sigma(70)-retaining transcription elongation complexes from Escherichia coli, *Cell* **106**, 443–451.

Bell, S.D., Jackson, S.P. (2001) Mechanism and regulation of transcription in archaea, *Curr. Opin. Microbiol.* **4**, 208–213.

Bell, S.P., Stillman, B. (1992) ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex, *Nature* **357**, 128–134.

Bhaumik, S.R., Green, M.R. (2002) Differential requirement of SAGA components for recruitment of TATA-box-binding protein to promoters in vivo, *Mol. Cell. Biol.* **22**, 7365–7371.

Bjorklund, S., Buzaite, O., Hallberg, M. (2001) The yeast mediator, *Mol. Cells* **11**, 129–136.

Boube, M., Joulia, L., Cribbs, D.L., Bourbon, H.M. (2002) Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man, *Cell* **110**, 143–151.

Buck, S.W., Sandmeier, J.J., Smith, J.S. (2002) RNA polymerase I propagates unidirectional spreading of rDNA silent chromatin, *Cell* **111**, 1003–1014.

Bushnell, D.A., Kornberg, R.D. (2003) Complete, 12-subunit RNA polymerase II at 4.1-A resolution: Implications for the initiation of transcription, *Proc. Natl. Acad. Sci. U.S.A* **100**, 6969–6973.

Butler, J.E., Kadonaga, J.T. (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression, *Genes Dev.* **16**, 2583–2592.

Cairns, J. (1963) The bacterial chromosome and its manner of replication as seen by autoradiography, *J. Mol. Biol.* **6**, 208–213.

Cashel, M., Hsu, L.M., Hernandez, V.J. (2003) Changes in conserved region 3 of Escherichia coli sigma 70 reduce abortive transcription and enhance promoter escape, *J. Biol. Chem.* **278**, 5539–5547.

Conaway, R.C., Brower, C.S., Conaway, J.W. (2002) Emerging roles of ubiquitin in transcription regulation, *Science* **296**, 1254–1258.

Dallmann, H.G., Kim, S., Pritchard, A.E., Marians, K.J., McHenry, C.S. (2000) Characterization of the unique C terminus of the *Escherichia coli* τ DnaX protein. Monomeric C-τ binds α and DnaB and can partially replace τ in reconstituted replication forks, *J. Biol. Chem.* **275**, 15512–15519.

Davis, J.A., Takagi, Y., Kornberg, R.D., Asturias, F.A. (2002) Structure of the yeast RNA polymerase II holoenzyme: mediator conformation and polymerase interaction, *Mol. Cells* **10**, 409–415.

Dvir, A., Conaway, J.W., Conaway, R.C. (2001) Mechanism of transcription initiation and promoter escape by RNA polymerase II, *Curr. Opin. Genet. Dev.* **11**, 209–214.

Eissenberg, J.C., Ma, J., Gerber, M.A., Christensen, A., Kennison, J.A., Shilatifard, A. (2002) dELL is an essential RNA polymerase II elongation factor with a general role in development, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9894–9899.

Ellison, V., Stillman, B. (2001) Opening of the clamp: an intimate view of an ATP-driven biological machine, *Cell* **106**, 655–660.

Falkenberg, M., Gaspari, M., Rantanen, A., Trifunovic, A., Larsson, N.G., Gustafsson, C.M. (2002) Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA, *Nat. Genet.* **31**, 289–294.

Freiman, R.N., Tjian, R. (2003) Regulating the regulators: lysine modifications make their mark, *Cell* **112**, 11–17.

Fuller, R.S., Kaguni, J.M., Kornberg, A. (1981) Enzymatic replication of *Escherichia coli* chromosome, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7370–7374.

Funnell, B.E., Baker, T.A., Kornberg, A. (1987) In vitro assembly of a prepriming complex at the origin of the *E. coli* chromosome, *J. Biol. Chem.* **262**, 10327–10334.

Gill, G. (2001) Regulation of the initiation of eukaryotic transcription, *Essays Biochem.* **37**, 33–43.

Gonzalez, F., Delahodde, A., Kodadek, T., Johnston, S.A. (2002) Recruitment of a 19S proteasome subcomplex to an activated promoter, *Science* **296**, 548–550.

Grummt, I. (1999) Regulation of mammalian ribosomal gene transcription by RNA polymerase I, *Prog. Nucleic Acid Res. Mol. Biol.* **62**, 109–154.

Hartzog, G.A. (2003) Transcription elongation by RNA polymerase II, *Curr. Opin. Genet. Dev.* **13**, 119–126.

Hiasa, H. Marians, K.J. (1999) Initiation of bidirectional replication at the chromosomal origin is directed by the interaction between

helicase and primase, *J. Biol. Chem.* **274**, 27244–27248.

Hilleren, P., Parker, R. (1999) Mechanisms of mRNA surveillance in eukaryotes, *Annu. Rev. Genet.* **33**, 229–260.

Hirota, Y., Ryter, A., Jacob, F. (1968) Thermosensitive mutants of *E. coli* affected in the processes of DNA synthesis and cellular division, *Cold Spring Harbor Symp. Quant. Biol.* **33**, 677–693.

Huberman, J., Riggs, A.D. (1968) On the mechanism of DNA replication in mammalian chromosomes, *J. Mol. Biol.* **32**, 327–341.

Jacob, F., Brenner, S., Cuzin, F. (1963) On the regulation of DNA replication in bacteria, *Cold Spring Harbor Symp. Quant. Biol.* **28**, 329–348.

Katayama, T., Kurokawa, K., Crooke, E., Sekimizu, K. (1998) The initiator function of DnaA protein is negatively regulated by the sliding clamp of the *E. coli* chromosomal replicase, *Cell* **94**, 61–71.

Kim, D.K., Yamaguchi, Y., Wada, T., Handa, H. (2001) The regulation of elongation by eukaryotic RNA polymerase II: a recent view, *Mol. Cells* **11**, 267–274.

Kitagawa, R., Ozaki, T., Moriya, S., Ogawa, T. (1998) Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for *E. coli* DnaA protein, *Genes Dev.* **12**, 3032–3043.

Kornberg, T., Gefter, M.L. (1971) Purification and DNA synthesis in cell-free extracts: properties of DNA polymerase II, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 761–764.

Kugel, J.F., Goodrich, J.A. (2002) Translocation after synthesis of a four-nucleotide RNA commits RNA polymerase II to promoter escape, *Mol. Cell. Biol.* **22**, 762–773.

Lee, J., Chastain, P.D., Kusakabe, T., Griffith, J.D., Richardson, C.C. (1998) Coordinated leading and lagging strand DNA synthesis on a minicircular template, *Mol. Cell* **1**, 1001–1010.

Lee, S.K., Yu, S.L., Prakash, L., Prakash, S. (2002) Requirement of yeast RAD2, a homolog of human XPG gene, for efficient RNA polymerase II transcription. Implications for Cockayne syndrome, *Cell* **109**, 823–834.

Maki, H., Maki, S., Kornberg, A. (1988) DNA Polymerase III holoenzyme of *Escherichia coli*. IV. The holoenzyme is an asymmetric dimer with twin active sites, *J. Biol. Chem.* **263**, 6570–6578.

Marahrens, Y., Stillman, B. (1992) A yeast chromosomal origin of DNA replication defined by multiple functional elements, *Science* **255**, 817–823.

Masternak, K., Peyraud, N., Krawczyk, M., Barras, E., Reith, W. (2003) Chromatin remodeling and extragenic transcription at the MHC class II locus control region, *Nat. Immunol.* **4**, 132–137.

Masukata, H., Tomizawa, J. (1990) A mechanism of formation of a persistent hybrid between elongating RNA and template DNA, *Cell* **62**, 331–338.

McHenry, C., Kornberg, A. (1977) DNA polymerase III holoenzyme of *Escherichia coli*. purification and resolution into subunits, *J. Biol. Chem.* **252**, 6478–6484.

Meselson, M., Stahl, F.W. (1958) The replication of DNA in *E. coli*, *Proc. Natl. Acad. Sci. U.S.A.* **44**, 671–682.

Myers, L.C., Kornberg, R.D. (2000) Mediator of transcriptional regulation, *Annu. Rev. Biochem.* **69**, 729–749.

Okazaki, T., Kurosawa, Y., Ogawa, T., Seki, T., Shinozaki, K., Hirose, S., Fujiyama, A., Kohara, Y., Machida, Y., Tamanoid, F., Hozumi, T. (1979) Structure and metabolism of the RNA primer in the discontinuous replication of prokaryotic DNA, *Cold Spring Harbor Symp. Quant. Biol.* **43**, 203–219.

Pillai, B., Verma, J., Abraham, A., Francis, P., Kumar, Y., Tatu, U., Brahmachari S.K., Sadhale, P.P. (2003) Whole genome expression profiles of yeast RNA polymerase II core subunit, Rpb4, in stress and nonstress conditions, *J. Biol. Chem.* **278**, 3339–3346.

Reeve, J.N. (2003) Archaeal chromatin and transcription, *Mol. Microbiol.* **48**, 587–598.

Salinas, F., Benkovic, S.J. (2000) Characterization of bacteriophage T4-coordinated leading- and lagging-strand synthesis on a minicircle substrate, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7196–7201.

Selleck, W., Howley, R., Fang, Q., Podolny, V., Fried, M.G., Buratowski, S., Tan, S. (2001) A histone fold TAF octamer within the yeast TFIID transcriptional coactivator, *Nat. Struct. Biol.* **8**, 695–700.

Shamoo, Y., Steitz, T.A. (1999) Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex, *Cell* **99**, 155–166.

Shilatifard, A., Conaway, R.C., Conaway, J.W. (2003) The RNA polymerase II elongation complex, *Annu. Rev. Biochem* **72**, 693–715.

Shim, E.Y., Walker, A.K., Shi, Y., Blackwell, T.K. (2002) CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the C. elegans embryo, *Genes Dev.* **16**, 2135–2146.

Stukenberg, P.T., Studwell-Vaughan, P.S., O'Donnell, M. (1991) Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme, *J. Biol. Chem.* **266**, 11328–11334.

Svejstrup, J.Q. (2002) Mechanisms of transcription-coupled DNA repair, *Nat. Rev. Mol. Cell. Biol.* **3**, 21–29.

Tan, Q., Prysak, M.H., Woychik, N.A. (2003) Loss of the Rpb4/Rpb7 Subcomplex in a Mutant Form of the Rpb6 Subunit Shared by RNA Polymerases I, II, and III, *Mol. Cell. Biol.* **23**, 3329–3338.

Taube, R., Lin, X., Irwin, D., Fujinaga, K., Peterlin, B.M. (2002) Interaction between P-TEFb and the C-terminal domain of RNA polymerase II activates transcriptional elongation from sites upstream or downstream of target genes, *Mol. Cell. Biol.* **22**, 321–331.

Tekotte, H., Davis, I. (2002) Intracellular mRNA localization: motors move messages, *Trends Genet.* **18**, 636–642.

Tougu, K., Marians, K.J. (1996) The interaction between helicase and primase sets the replication fork clock, *J. Biol. Chem.* **271**, 21398–21405.

Tsurimoto, T., Melendy, T., Stillman, B. (1990) Sequential initiation of lagging and leading strand synthesis by two different polymerase complexes at the SV40 DNA replication origin, *Nature* **346**, 534–539.

Waga, S., Stillman, B. (1994) Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication *in vitro*, *Nature* **369**, 207–212.

Weinreich, M., Liang, C., Stillman, B. (1999) The Cdc6p nucleotide-binding motif is required for loading mcm proteins onto chromatin, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 441–446.

Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., Bell, S.P., Aparicio, O.M. (2001) Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins, *Science* **294**, 2357–2360.

Xiao, T., Hall, H., Kizer, K.O., Shibata, Y., Hall, M.C., Borchers, C.H., Strahl, B.D. (2003) Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast, *Genes Dev.* **17**, 654–663.

Yamazaki, K., Aso, T., Ohnishi, Y., Ohno, M., Tamura, K., Shuin, T., Kitajima, S., Nakabeppu, Y. (2003) Mammalian elongin A is not essential for cell viability but required for proper cell cycle progression with limited alteration of gene expression, *J. Biol. Chem.* **278**, 13585–13589.

Young, B.A., Anthony, L.C., Gruber, T.M., Arthur, T.M., Heyduk, E., Lu, C.Z., Sharp, M.M., Heyduk, T., Burgess, R.R., Gross, C.A. (2001) A coiled-coil from the RNA polymerase β' subunit allosterically induces selective nontemplate strand binding by σ^{70}, *Cell* **105**, 935–944.

Zyskind, J.W., Smith, D.W. (1980) Nucleotide sequence of the *S. typhimurium* origin of DNA replication, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2460–2464.

DNA Typing Analysis

Chantal J. Frégeau[1], *Rémy A. Aubin*[2], *Bruce Budowle*[3], *and Ron M. Fourney*[1]
[1] *Royal Canadian Mounted Police, National DNA Data Bank, Ottawa, Ontario, Canada*
[2] *Health Canada, Centre for Biologics Research, Ottawa, Ontario, Canada*
[3] *Federal Bureau of Investigation, Washington, DC, USA*

1 Principles 480
1.1 DNA Types and Regions of Variability Relevant to Individual Identification 480
1.2 Specimen Processing 483
1.3 Specimen Analysis 486
1.4 Data Processing 490
1.5 Quality Assurance and Quality Control 491

2 Applications 492
2.1 Criminal and Forensic Investigation 492
2.1.1 Human Identity 492
2.1.2 Animal Identity 495
2.1.3 Plant Identity 496
2.1.4 Fungal Identity 497
2.1.5 Bacterial Identity 497
2.1.6 Viral Identity 497
2.2 Clinical Medicine 497
2.3 Parentage Studies 498
2.3.1 Human Testing 498
2.3.2 Testing of Nonhuman Animal Species 498
2.4 Molecular Evolution 499
2.4.1 Evolution of Human and Other Animal Species 499
2.4.2 Evolution of Plants 499

3 Perspectives 500
3.1 DNA Typing versus Conventional Methods 500
3.2 Future Directions 500

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

Bibliography 501
Books and Reviews 501
Primary Literature 501

Keywords

AmpFLP (Amplified Fragment Length Polymorphism)
A technique based on PCR amplification of specific VNTRs with core repeats ranging between 8 and 70 bp in length, using sequence-specific oligonucleotide primers that flank the tandem repeat regions. This strategy generates discrete profiles of amplification products (amplicons) that correspond to the alleles of the locus. Profiles with one band are denoted as homozygous. Profiles with two bands are scored as heterozygous.

AP-PCR (Arbitrarily Primed Polymerase Chain Reaction)
A variation of the RAPD strategy (see below) wherein discrete amplification patterns are generated using single oligonucleotide primers 10 to 50 nucleotides in length to amplify genomic DNA by PCR. In the first two cycles of amplification, annealing of the primers is carried out under nonstringent conditions to allow the primers to anneal to multiple regions along the DNA template. The method generates multiple band patterns.

DAF (DNA Amplification Fingerprinting)
Another variation of the RAPD technique wherein single oligonucleotide primers of arbitrary sequence as short as five nucleotides in length are used to amplify DNA by PCR. This technique also generates multiple band profiles.

ISSR (Inter Simple Sequence Repeat)
In this technique, nucleotide sequence information from microsatellite regions (simple sequence repeat elements) is used to design oligonucleotide primers that will serve to amplify regions of DNA mapping between microsatellites (SSRs). An anchor made up of a stretch of degenerate nucleotides (between 3 and 10 nucleotides) is also attached either at the 5′ or 3′ end of the primers to increase the specificity of the amplification process. This technique generates complex multiple band patterns typically consisting of more than 200 band signatures.

Microsatellites and Minisatellites
Microsatellites are nucleotide sequence elements composed of tandemly repeated core repetitive sequence blocks of 2 to 7 bp in length distributed throughout the genomes of higher eucaryotes. Minisatellites are similar in nature but are composed of 8 to 70 bp core repeat sequences.

RAPD (Random Amplified Polymorphic DNA)
A procedure used to detect nucleotide sequence polymorphisms in DNA using a single primer of arbitrary nucleotide sequence. PCR generates several discrete amplification products when a single primer (10 to 21 bases in length) is designed to anneal, in forward and reverse orientations, to two different sites on DNA within a distance that allows amplification.

Sequencing
A method used to determine the sequence (linear order) of nucleotide bases along a strand of DNA.

SNP (Single-nucleotide Polymorphism)
Variation between individuals at certain nucleotide base positions within the genome. SNPs may arise as a consequence of single-nucleotide substitution, insertion, or deletion.

STR (Short Tandem Repeat)
Also referred to as *simple sequence repeat* (SSR) or *microsatellite*. STRs are tandem repeat regions 2 to 7 bp in length, scattered throughout genomes. PCR procedures are routinely used to amplify these regions for the generation of distinct allelic profiles.

VNTR (Variable Number of Tandem Repeats)
Regions of the genome that contain a variable number of end-to-end duplications of a series of identical or almost identical nucleotide sequence motifs (typically between 2 and 70 bp each). The number of tandem repetitions translates into length polymorphisms that characterize the allele profile for a given genomic region (locus).

DNA typing analysis consists in generating and interpreting allelic profiles for the purpose of identifying an individual, a representative from a given group or species, or to determine the origin/provenance of biological material such as cells, clinical specimens, and forensic evidence. The process capitalizes on length differences or sequence variations (collectively referred to as *polymorphisms*) associated with discrete regions within the genomes of living organisms. These genetic differences are examined using a variety of DNA profiling strategies that utilize different modes of detection to reveal the polymorphic nature of the target DNA segments. The choice of method adopted is largely dictated by the nature, quantity, and physical state of the sample to be identified. Each of these considerations, in addition to the current state of knowledge regarding the sequence, structure, and organization of an organism's nuclear and organelle genomes, will also determine what type of polymorphisms can be surveyed. Polymorphic allele patterns can be represented as bar code–like profiles, peaks on tracings, photographic bars or dots on X-ray films, or colored spots on solid supports or computer monitors. DNA typing can be applied to any biological material that contains DNA such that current applications reach across every taxonomic kingdom.

DNA typing analysis is particularly powerful because no two individuals or strains, with the exception of genetically identical twins and clones, share the same allele complement for multiple polymorphic loci. In addition, with the advent of DNA typing strategies based on the polymerase chain reaction (PCR), investigators have advanced beyond conventional restriction fragment length polymorphism (RFLP) analysis and can carry out reliable, accurate, precise, and sensitive identifications from only minute quantities of starting biological material. This, coupled with the surprising chemical stability and longevity of DNA, has had a major impact in forensic science for the resolution of violent crimes or the identification of human skeletal remains from the scenes of mass disasters. Today, DNA typing analysis is employed in almost every discipline under the biological sciences. It continues to prove instrumental to studies on human origins and diversity, in clinical medicine to resolve the identity of specimens, ascertain the provenance of transplantable tissues and organs, and expedite the identification of viral and microbial pathogens, and in guiding the manner in which animal, plant, and insect breeding programs are being conducted and followed.

This article describes the general principles underlying DNA typing analysis, summarizes and compares currently available methods for specimen identification, and provides an overview of applications in the field of forensics and the biological sciences.

1 Principles

1.1 DNA Types and Regions of Variability Relevant to Individual Identification

With the notable exception of the RNA viruses, DNA constitutes the molecule of heredity in all living organisms. In eucaryotes, the majority of a cell's genetic material and gene set (genome) resides in the nucleus and is referred to as *nuclear genomic DNA*. Plants and animals also contain organelles – chloroplasts (restricted to plants) and mitochondria – with their own genomes. In procaryotes, the sum total of genomic DNA is not confined to a nuclear compartment but rather is attached to the inner side of the cell membrane as a nucleoid.

DNA typing analysis relies on genetic variations that manifest themselves as tandem arrays of repetitive sequences or single-nucleotide sequence polymorphisms. Tandem repeats of repetitive DNA sequence blocks detected in human nuclear genomic DNA are generally subdivided into minisatellites and microsatellites according to the size of the core repeat unit and to the overall length of the tandem array. Similar subsets of tandem repetitions have been detected and characterized in the nuclear genomic DNA of other animals and plants. Depending on the DNA region (or locus), each repeat element can consist of a minimum of 2 to perhaps 70 bp. The number of tandem repetitions for each allele, on the other hand, can vary from one to several hundred – hence their collective designation as "variable number of tandem repeats

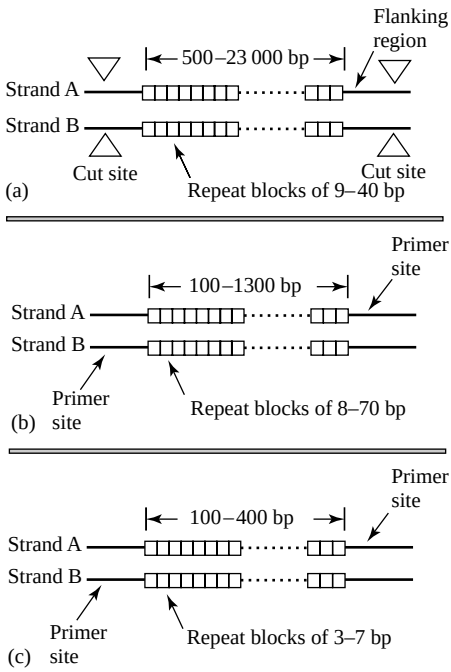

Fig. 1 Three types of VNTR polymorphisms routinely detected in human nuclear genomic DNA: (a) restriction fragment length polymorphism (RFLP), (b) amplified fragment length polymorphism (AmpFLP), and (c) short tandem repeat (STR) polymorphism. For all three types of length polymorphisms, the ability to distinguish one individual from another depends on the variations in the number of tandem repeat elements present in each of the alleles of a particular locus or genomic region. VNTRs detected by the RFLP approach are attributed to repetitive nucleotide sequence blocks ranging between 9 and 40 base pairs (bp) in length that are repeated 60 to 450 times in tandem. DNA segments containing VNTRs are produced after digestion with a site-specific restriction enzyme (i.e. *Hae*III) that cleaves within the regions immediately flanking the VNTR. The procedure generates allelic DNA segments in the range of 500 bp to 23 kb in size. Because of their large size, VNTR-containing DNA fragments cannot be amplified by PCR and must be resolved by conventional agarose gel electrophoresis. By contrast, VNTRs detected by the AmpFLP approach are considerably smaller and consist of repeat blocks typically ranging between 8 and 70 bp and repeated 10 to 40 times, depending on the allele. AmpFLP VNTRs can be generated by PCR amplification with specific VNTR-flanking oligonucleotide primer pairs. This process routinely yields DNA size fragments (i.e. amplicons) of 100 to 1300 bp in size that are easily resolved on polyacrylamide gels. STRs, on the other hand, arise from repeat blocks in the 3- to 7-bp range, which are repeated 6 to 44 times in tandem. Because of their small overall repeat lengths – typically in the range of 100 to 400 bp per allele – STR loci lend themselves readily to PCR amplification either alone or as combinations using several locus-specific primer pair sets in a process known as "multiplex PCR" and to automated high-throughput analysis. STR amplicons are best resolved by polyacrylamide or capillary gel electrophoresis. [Modified from Duncan and Tracey (1997).]

(VNTRs)" (Fig. 1). Investigators rely on these length variations to construct DNA allele profiles. This can be achieved by restriction fragment length (RFLP), amplified fragment length (AmpFLP), or short tandem repeat (STR) polymorphism analysis. The extent of length variation across the allele set for a given locus provides a measure of the usefulness of the VNTR for identification purposes. Thus, the greater the variation encountered within a population of individuals, the greater the "polymorphism information content" and consequent utility of the VNTR locus. For the study of VNTRs, DNA regions containing different numbers of tandem repeats

digested with a restriction endonuclease or amplified by PCR are fractionated by size using electrophoresis through a sieving medium such as agarose (RFLP) or polyacrylamide (AmpFLP and STR). Following electrophoresis and processing for detection, the relative positions of bands corresponding to distinct alleles are determined relative to known DNA size standards. By combining several polymorphic loci in a single analysis, investigators can obtain very high "powers of discrimination" (i.e. a statistical measure of the unique character of the genetic profile generated from the biological sample).

Recent genomic mapping and sequencing initiatives in humans and other organisms have identified single-nucleotide polymorphisms (SNPs) as one of the most common forms of genetic variation among individuals. SNPs account for approximately 85% of all known polymorphisms in the human genome and are more frequent than microsatellite markers. They occur on average every 1000 nucleotide positions as opposed to every 300 to 500 kbp for microsatellites. The large size of the human nuclear genome (6.2×10^9 bp) predicts the existence of millions of SNPs and thousands of microsatellite markers. Individual SNPs, however, are either detected or altogether absent and are consequently less informative than VNTR markers. Nonetheless, this simple "biallelic" variation greatly facilitates high-throughput analysis and detection using DNA microarrays. Recent estimates suggest that a subset of 50 to 100 SNPs could be used to achieve the same power of discrimination as that provided by the 13 STR core loci currently in use by the majority of forensic DNA laboratories in North America and Europe.

SNP typing of organelle genomic DNA like that of the mitochondrion (mtDNA) provides yet another means to help establish the identity of biological specimens of human and other animal species origin. In humans, the exercise exploits nucleotide sequence variations within two hypervariable regions (HV1 and HV2) of the noncoding section (known as the *mitochondrial control region* or *D-loop*) of the mtDNA genome. In other animals, species identification has been achieved using the mitochondrial cytochrome b, 16S, and 12S ribosomal DNA (rDNA) nucleotide sequences. MtDNA analysis constitutes a very sensitive assay because thousands of copies of the comparably small mtDNA circular genome (16 500 bp) are present in each somatic cell as opposed to two copies (one maternal, the other paternal) of nuclear genomic DNA. MtDNA is also reported to be more resistant to degradation by cellular enzymes or enzymes secreted by microorganisms located in the vicinity of the biological specimen. Because mtDNA polymorphisms are detected within much smaller segments of DNA, this technique is more successful on highly degraded specimens (i.e. aged skeletal material, mummified soft tissue and hair shafts, sterilized canned foods, charred fragmented bones). MtDNA is maternally inherited and does not undergo recombination thereby increasing the range of reference material available for the identification of human remains (i.e. distant maternal relatives become potential reference sources for comparison). Although nuclear genomic DNA profiling is highly informative and represents the method of choice for the majority of DNA identification requirements, mtDNA analysis is recognized as a validated, robust, reliable, and well-established DNA profiling system for some of the most challenging and degraded samples such as archeological

and mass disaster specimens, where nuclear DNA profiles are either impossible to obtain or uninformative.

DNA typing in plants using chloroplast genomic DNA is becoming more frequent. Chloroplasts are found in large numbers within each plant cell. Chloroplasts are protected by the plant cell wall and by the plant cell and organelle lipid bilayer membranes such that chloroplast genomic DNA tends to resist breakdown from microorganisms and environmental insult. Nucleotide sequence variations recorded within the transfer RNA gene complex of the chloroplast genome have been successfully applied to the identification of plant species.

1.2 Specimen Processing

In principle, and largely as a result of significant advances in PCR amplification strategies, any living or nonliving biological specimen or material containing minute quantities of relatively intact DNA fragments can be used for DNA typing analysis. At present, severely degraded samples containing DNA, no longer than 60 to 100 bp in length, can be analyzed. Until very recently, common sources of DNA routinely processed for medical and forensic applications included blood, saliva, semen, solid tissues, plant foliage, and seeds. With PCR, however, the list of samples submitted for DNA testing has been expanded to include bone, teeth, hair, epithelial cells from the skin, and even sweat from human or other animals. In addition, forensically relevant specimens such as fingernails, cigarette butts, chewing gum, bite marks on food items, licked envelope flaps and stamps, and handled objects such as telephone receivers, wine glasses, beverage cans, and beer bottles (all of which can be swabbed to collect transferred epithelial cells) have become amenable to analysis. Other samples such as human and animal feces, urine, and sputum as well as canned foods can also constitute rich sources of DNA.

Biological samples destined for forensic DNA analysis are often subjected to a variety of environmental insults. Consequently, these samples create a special challenge and it is often only following collection that conditions can be controlled to preserve the integrity of the specimens. Ensuring the integrity of clinical or forensic samples is best achieved by cold storage [freezer (−20 °C or − 70 °C) or liquid nitrogen] in clean, dry environments. Such conditions reduce contamination by microorganisms and the likelihood of degradation of DNA by cell lysis and DNAase activity. Alternatively, specimens such as bones or tissues can be immersed in preservation solutions at the site of collection and stored at room temperature for prolonged periods of time before being processed for DNA analysis. Biological samples (blood, buccal swabs, or hair) such as those destined for inclusion in convicted offender DNA data banks can be collected by direct deposition on specially treated filter paper (i.e. FTA® paper; Fitzco Inc., Minneapolis, MN, USA) that inhibits the growth and biological properties of blood-borne pathogens (e.g. HIV and hepatitis B and C) and stabilizes DNA for long-term storage7 at ambient temperatures.

Various DNA extraction procedures can be used to obtain essentially pure preparations of high molecular weight DNA in good yields from human or animal tissues. Although specific DNA extraction procedures are required for bones, teeth, and hair, the enzyme proteinase K is generally used to promote cell lysis and digest proteins in a sample in the presence of

the detergent sodium dodecyl sulfate. Digested proteins, lipids, and other debris and contaminants can be removed by organic extraction using a mixture of phenol and chloroform. Traces of residual phenol are removed by extraction with chloroform and isoamyl alcohol. High molecular weight DNA is then recovered in essentially pure form (although RNA may also be present without consequence to the final analysis) by ethanol precipitation in the presence of salt. An alternative method for collecting phenol-extracted DNA consists in selective ultrafiltration through special membranes (Centricon-100 or Microcon-100 ultrafiltration cartridges; Amicon Corp., Beverly, MA) that concentrate DNA while removing impurities by centrifugal dialysis. DNA obtained by the organic extraction method is double stranded and is therefore compatible with both RFLP and PCR-based procedures.

Other methods that bypass organic extraction for the recovery of DNA from cell lysates include (1) selective capture of DNA on magnetic beads followed by elution in a buffered solution and (2) the use of cation-exchange resins such as Chelex 100 to bind and partition positively charged molecules from cell lysates on the resin and allow the negatively charged DNA to remain in solution. Both these methods are only compatible with PCR-based assays because cell lysis is carried out at elevated temperatures (95 °C and 100 °C), which result in the denaturation of DNA to its single-stranded form.

Disruption of plant cells to release DNA poses a special challenge. Plant tissues differ from those of animals in that they are surrounded by rigid lignin-rich cell walls that are very difficult to lyse. Methods optimized to obtain the best yields and the purest DNA preparations possible typically involve further purification steps to remove phenolic substances, polysaccharides, and humic acids that are often present along with DNA and can inhibit downstream restriction enzyme digestion and PCR amplification reactions.

When biological samples are collected and stored on small, designated areas of FTA® paper (microbial, animal and plant cell cultures, blood, hair bulb cells, buccal swabs), deposited cells are lysed on contact with the chemicals present in the paper. The DNA remains trapped between the fibers and is subsequently purified by washing away the chemical components and cellular debris that could inhibit subsequent assays. The washed immobilized DNA can be recovered on circular punches (1.2 mm, 2 mm, or 3.7 mm in diameter) and transferred directly into a microcentrifuge tube or a 96-well assay plate. This simple procedure allows automation of sample preparation for PCR amplification with robotic workstations. In addition, the use of the FTA® paper represents a simple and inexpensive approach to long-term storage or archiving of biological samples at room temperature.

Some DNA extraction methods have been designed to address special circumstances. In forensic investigations, for example, the analyst may be presented with specimens containing mixtures of male and female cells such as in the case of oral, vaginal, and anal swabs obtained from sexual assault victims where sperm and female epithelial cells are often present. To separate the male and female cells from one another, specific DNA extraction methods are used. The standard ''preferential lysis'' method exploits the different resistance properties of sperm and vaginal cells against proteolytic digestion. The female cells are lysed in extraction buffer lacking the reducing agent dithiothreitol (DTT), and the DNA is isolated from the

supernatant. The remaining sperm pellet is then lysed by adding extraction buffer supplemented with DTT and detergents, which release the male DNA into the solution, and allows for subsequent isolation and concentration. Recently, enriched male cell fractions have been obtained using methods such as fluorescence activated cell sorting (FACS), which takes advantage of differences in the cell size, shape, surface antigen properties, and DNA content (haploid for spermatocytes, diploid for vaginal cells). Another approach used to enrich for male cells consists in capturing sperm cells using specific antibodies raised against sperm head or tail antigens, which have been covalently linked to magnetic beads. Following washes to remove cellular debris and contaminating female cells, the male cells can be eluted and processed for DNA typing.

The integrity of a DNA preparation can be estimated by visual inspection with the help of ultraviolet light following agarose gel electrophoresis and staining with intercalating dyes such as ethidium bromide and SYBR® Green I. Intact, high molecular weight DNA migrates as a single large band at the top of a gel, whereas partially degraded DNA produces a long smear composed of large to small fragments (>20 000 bp to a few hundred base pairs for human DNA). DNA molecular weight ladders run alongside the sample lanes are used as reference size markers. For many PCR-based DNA typing assays, an evaluation of DNA degradation is not performed in order to conserve potentially limiting quantities of DNA. Rather, the extent of DNA degradation can be assessed indirectly by examining the outcome of the PCR amplification. Partial profiles as well as failure to produce a profile are often indicative of DNA degradation when the presence of an inhibitor has been discounted.

Quantification of the amount of DNA present in the extract is required to optimize the DNA typing assay, particularly if the latter involves PCR amplification. Knowing the initial quantity of DNA template can reduce undue consumption of biological specimen and reagents and can prevent the generation of potential PCR artifacts under suboptimal conditions. In general, two categories of procedures are available to estimate the quantity of DNA. The first category defines methods that are not specific to the human species. Such methods estimate the total amount of DNA in an extract. For example, the quantity of DNA can be estimated by measuring the absorbance of a DNA solution at a wavelength of 260 nm using a spectrophotometer. Using a quartz cell with a path length of 1 cm, a concentration of 50 μg DNA/mL yields an A_{260} of 1.0. The relative purity of the DNA sample can be assessed by including a measurement for the absorbance at 280 nm. By doing so, one searches for traces of phenol and residual amounts of aromatic amino acids signifying the presence of residual proteins. The investigator then derives an absorbance ratio for A_{260}/A_{280}. A ratio of 1.8 is usually indicative of a clean DNA preparation. Alternatively, yield gels can be used where the sample DNA and a series of DNA standards of known concentrations are compared by agarose gel electrophoresis and ethidium bromide or SYBR® Green I staining. Total DNA quantification can also be performed in the presence of the intercalating fluorescent reagent Picogreen® that binds primarily to double-stranded DNA. In a typical assay, samples of unknown quantities of DNA are compared to a series of serially diluted DNA standards after being mixed for

5 min at room temperature with the dye in 96-well assay plates. Light emission (fluorescence) measurements are then made using a fluorimeter. This method is particularly sensitive and can detect DNA down to 0.05 ng/μL.

For the quantification of human-specific DNA, an aliquot of the DNA sample is denatured (converted into single-stranded molecules) in alkali along with serial dilutions of DNA standards. Each sample is then adsorbed and covalently bound to a charge-modified nylon membrane. The immobilized DNAs are hybridized to a radiolabeled, chemiluminescent or colorimetric, primate-specific α-satellite DNA probe (p17H8) fragment that targets highly repetitive human DNA sequences located at the D17Z1 chromosome locus. This method is capable of detecting subnanogram amounts of human DNA. A recent approach eliminates the need for DNA immobilization on membranes and uses solution hybridization of DNA probes targeting highly repetitive human Alu DNA elements coupled to an enzyme-linked chemiluminescence reaction to estimate the quantity of human DNA in the sample.

Quantification from processed biological samples immobilized on FTA® filter paper is typically not carried out because a punched disc of uniform diameter (1.2 mm, 2 mm, or 3.7 mm) and depth contains a relatively consistent quantity of DNA bound to the matrix and produces equally reproducible amplification profiles for specific DNA VNTR targets.

1.3 Specimen Analysis

For many years, RFLP-based DNA typing was considered the standard method for most identification purposes. This method has since been replaced by PCR-based analytical strategies, particularly for human identification. RFLP analysis is still widely used in plant and animal studies. The RFLP method produces excellent, valid, reliable, and easily interpretable typing results (Fig. 2). Despite these advantages, RFLP analysis has limitations. For instance, each analysis requires a minimum of 10 to 50 ng of high-molecular weight genomic DNA to permit hybridization to each probe used to reveal the banding patterns. The RFLP procedure is also time-consuming and labor intensive, often taking weeks to complete when full analysis requires multiple rounds of hybridizations to several probes to build a discriminating allele profile (i.e. a genotype).

To circumvent these limitations, AmpFLP and STR-based methods were initially developed to complement RFLP typing in humans. Both strategies focus on short VNTR regions (8 to 70 bp for AmpFLPs and 2 to 7 bp for STRs) that are amenable to amplification by PCR. Amplicons produced are typically 150 to 1500 bp and 100 to 500 bp for AmpFLPs and STRs respectively. It should be noted, however, that most loci targeted for PCR analysis are much less polymorphic than the hypervariable regions used for RFLP testing. STR markers, in general, have far fewer repeats than AmpFLP markers and offer less variation at any one locus, but, despite this characteristic, the power of STRs lies in the ability to combine them into a single amplification assay (i.e. a multiplex PCR reaction). Multiplex STR analysis, therefore, represents a major advance in DNA typing by providing enhanced power of discrimination and automation capabilities. In a typical scenario, paired sets of specific oligonucleotide primers flanking short polymorphic VNTR regions

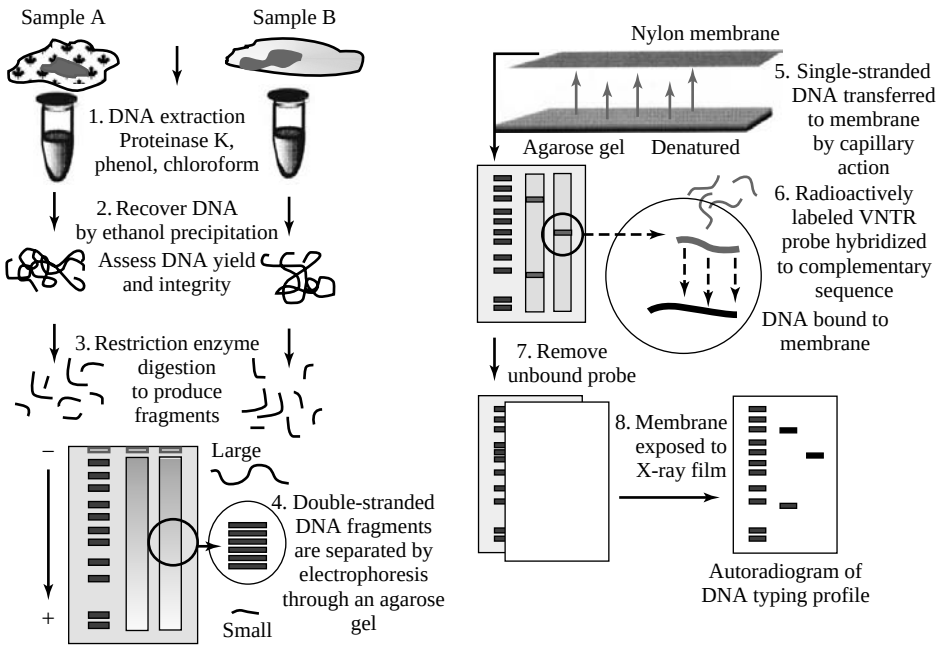

Fig. 2 Restriction fragment length polymorphism–based DNA typing process. Genomic DNA is first released from a biological sample by whole-cell lysis using a proteolytic enzyme (proteinase K) in the presence of a detergent (sodium dodecyl sulfate). Denatured proteins, peptides, and lipids are then removed by one or more organic extractions with a mixture of buffered phenol and chloroform. Genomic DNA is finally recovered by ethanol precipitation and both the yield and integrity of the DNA are assessed. The DNA is then cleaved with a robust site-specific restriction endonuclease (*Hae*III) to generate a series of DNA fragments of different lengths. These resulting fragments are resolved according to their sizes (lengths) by electrophoresis through an agarose gel. The negatively charged DNA fragments migrate through the agarose sieve toward the positive electrode located at the bottom of the electrophoresis unit. Once resolved, the series of DNA fragments is transferred by capillary action from the gel to a nylon membrane by a process known as *Southern blotting*. During this process, an alkaline transfer solution is used to unwind (denature) the DNA strands of each fragment and carry them along with the flow of transfer solution out of the gel and onto a charge-modified nylon membrane on which DNA strands will bind covalently. The membrane is then exposed to a radiolabeled or chemiluminescent complementary DNA probe that hybridizes with the membrane-bound denatured DNA fragments that contain target VNTR nucleotide sequences. Following a series of extensive washing steps to remove unbound probe, the membrane is placed in intimate contact with photographic X-ray film that is sensitive to the emission of β particles (decay products of the radioactive phosphorus atoms) or photons (generated by a subsequent light-emitting chemical reaction to detect hybridized probe). After an appropriate exposure time (minutes, hours, or days), the film is developed to produce a photographic image of the relative positions of the radiolabeled or chemiluminescent probe/target DNA hybrids on the membrane. The probes can be stripped off the membrane without removing the bound DNA, and the process can be repeated using a different probe that binds to a different set of DNA fragments. The composite data set of all DNA patterns generated using different probes is referred to as the *genetic profile* or *genotype* of the particular sample.

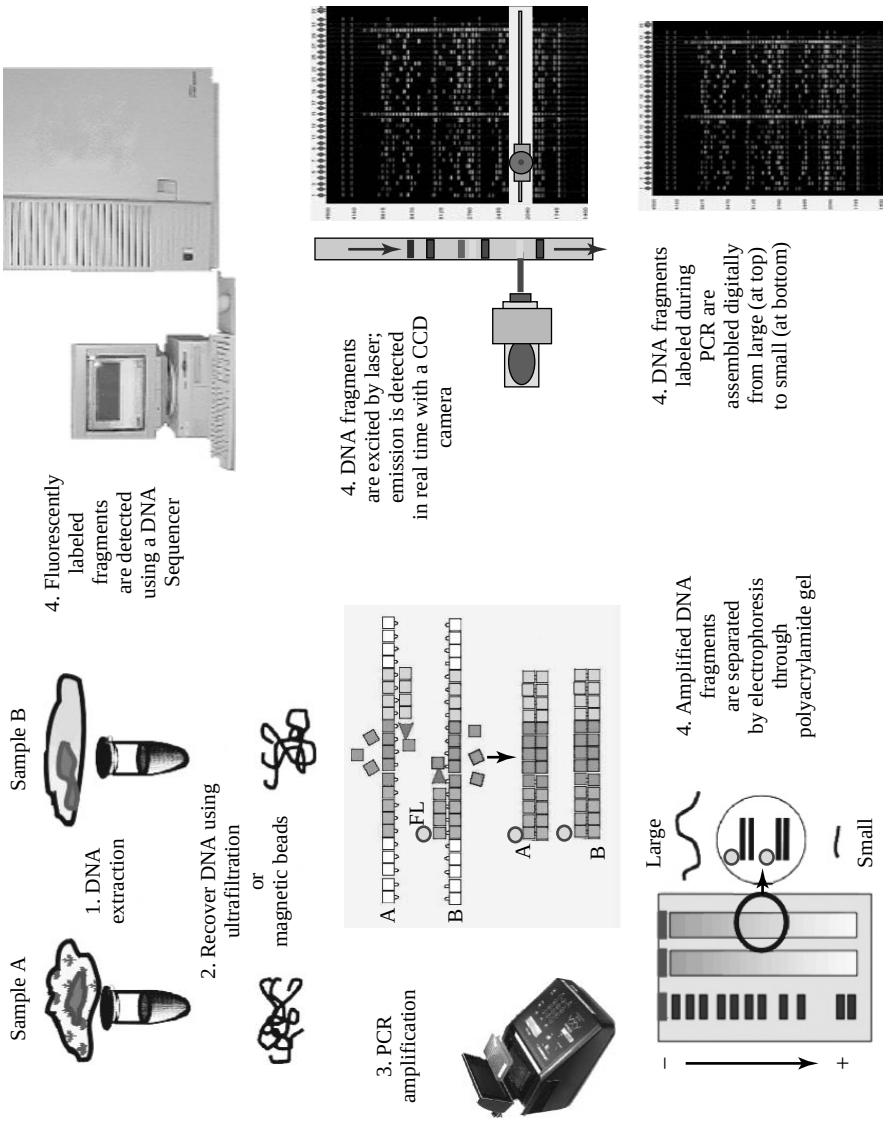
Sample A
Sample B
1. DNA extraction
2. Recover DNA using ultrafiltration or magnetic beads
3. PCR amplification
A
B
FL
A
B
4. Fluorescently labeled fragments are detected using a DNA Sequencer
4. DNA fragments are excited by laser; emission is detected in real time with a CCD camera
4. DNA fragments labeled during PCR are assembled digitally from large (at top) to small (at bottom)
4. Amplified DNA fragments are separated by electrophoresis through polyacrylamide gel
Large
Small
–
+

are used to amplify allelic DNA fragments that are then resolved according to their sizes by electrophoresis through polyacrylamide or other suitable synthetic polymer sieving matrices. The allelic DNA fragments are revealed using silver staining or fluorescence detection (Fig. 3).

A wide variety of biological samples have been successfully analyzed (genotyped) using the STR-based approach. In particular, allele profiles can be obtained from highly degraded DNA (in general about 100 bp in size) by simply redesigning PCR primer pairs to anneal closer to the STR sequence block. For a small percentage of biological samples from which only limiting amounts of highly degraded DNA are available, alternative typing methods, such as mtDNA analysis, have been developed and successfully applied.

Other types of nucleotide sequence polymorphisms can be detected in specific genes or noncoding regions of the genome by exploiting "allele-specific oligonucleotide (ASO) hybridization" probes. When these are used in a "dot-blot" format, the target DNA sequence is amplified by PCR and the amplification products are fixed to membranes. The DNA is then incubated with labeled ASOs, which hybridize only to DNA sequences that contain their exact complement. If hybridization occurs, the dot will be highlighted. If the amplified DNA does not correspond to the proper allele, the spot will remain blank. Alternatively, dot blots can be configured such that it is the

Fig. 3 Short tandem repeat–based DNA typing process. Genomic DNA is first released from a biological sample by whole-cell lysis using a proteolytic enzyme (proteinase K) in the presence of a detergent (sodium dodecyl sulfate). The DNA in solution is then recovered by ultrafiltration through specialized membrane units, or, alternatively, can be captured on magnetic beads to yield high-purity DNA ready for PCR amplification. Amplification is carried out in a thermocycler. During this process, multiple copies of STR-containing DNA regions are generated by a heat-stable DNA polymerase (i.e. *Taq* DNA polymerase) that uses sequence-specific oligonucleotide primer pairs that bind to complementary regions flanking the VNTR (shown in orange). The PCR process consists of three simple steps. In the first, the double-stranded DNA molecule is heat denatured (strands A and B unwind from each other at 95 °C). In the second, primers bind to their complementary sequences as the temperature is lowered (50 °C–65 °C), and in the third, daughter DNA strands are synthesized by the *Taq* DNA polymerase as the enzyme extends the hybrid DNA template–primer complex as the temperature is augmented to an optimum (72 °C). The first amplification cycle generates two copies of the original VNTR region. Repeated cycles of amplification generate millions of copies of the VNTR region according to a 2^n logarithmic rule where n represents the number of amplification cycles. Multiple primer sets labeled with different fluorochromes (denoted as FL in the diagram) can be used simultaneously ("multiplex PCR" approach) to amplify multiple VNTR regions from different chromosomal loci within the genome. The resulting colored fragments are resolved according to their individual sizes by polyacrylamide gel electrophoresis. The size fractionation and detection of the amplified fragments occur simultaneously in an analytical instrument called a *DNA sequencer*. During electrophoresis, the DNA fragments migrate toward the bottom of the gel and reach the detection window where an argon laser excites the fluorochrome attached to one of the 5′ ends of the amplified DNA fragments. The photons emitted are captured in real time by a charged-coupled device (CCD) camera housed in the instrument and calibrated to the emission spectra of the fluorochromes. A gel image of the position of each of the STR-containing DNA fragments is then produced against a reference allelic ladder. The allelic ladder represents a composite migration set of the many STR-containing fragments detected in the general population. The composite array of all STR signatures amplified for a given sample is referred to as the *genetic profile* of that sample. (See color plate p. xxxii.)

ASOs that are fixed to the membrane, and the amplified DNA is allowed to hybridize to the immobilized probe sequences. This "reverse dot-blot" approach is routinely used for HLA-DQα typing, "polymarker" profiling and mtDNA typing.

Nucleotide sequence polymorphisms can also be detected in plants and animals using a scheme for generating "randomly amplified polymorphic DNA" (RAPD) segments. The ensuing methods are known as *arbitrary primed polymerase chain reaction* (AP-PCR) or *DNA amplification fingerprinting* (DAF). AP-PCR and DAF typing systems rely on PCR primers of arbitrary sequence. The latter are allowed to anneal to the flanking regions of target DNA regions. Annealing must occur over reasonable distances in order to facilitate amplification. Each system differs according to (1) the length of the primers used, (2) amplification conditions, and (3) the way in which amplified products are resolved and detected. Amplification will not occur if the template sequence differs significantly from the primer sequence. The main advantage of RAPD methods is that nucleotide sequence information is not required to design primers for a particular genomic DNA. Nucleotide sequence information is paramount, by contrast, for successful design and application of AmpFLP and STR-based typing strategies.

ISSR-PCR typing provides yet another alternative method to derive highly reproducible allele profiles without knowing anything about the nucleotide sequence of an organism's genome. ISSR-PCR is particularly useful in plant and nonhuman animal DNA typing exercises. The assay is somewhat more specific than the RAPD approach since higher annealing temperatures (usually in the vicinity of 50 °C) and longer primers are used. ISSR-generated profiles are usually preferred over those obtained from RFLP and RAPD strategies because the great number of signature amplification products produced per primer (typically more than 200 bands) facilitates identification across many species. ISSR-PCR is not used for human identification because locus-specific assays are available.

The genetic profiles (genotypes) generated by each of the aforementioned DNA typing methods can be visualized and/or scored with the help of colored dyes, fluorochromes or fluorescent dyes, radioisotopes, chemiluminescence detection, or staining with silver-based compounds. Today, the vast majority of DNA typing exercises rely on chemiluminescence and fluorescence as sensitive, less hazardous, and readily automated detection methods.

1.4 Data Processing

Genetic profiles can be represented as (1) colored, fluorescent, or photographic dots; (2) band patterns corresponding to the position of DNA fragments following electrophoretic separation; or (3) direct nucleotide sequence information (e.g. AATCGTACCTGATCC). While such data can be evaluated visually, computer-assisted analyses greatly facilitate allele resolution, detection, pattern interpretation, data storage, and statistical evaluation, particularly when data streams are abundant.

For dot-blot profiles, most comparative interpretations are performed by "eye" against known standards. The data are generally transcribed manually into a computer-managed database. However, since most dots are colored or visible on X-ray film, densitometric scanning systems can be used to automate the

scoring and storage of DNA profiles for subsequent analysis.

VNTR analysis relies on size and position as fundamental allele characteristics. The first generation of data acquisition systems used a simple ruler to determine the relative size of unknown DNA allele fragments by comparing their migrations with those of known DNA size standards on autoradiograms. In the early days of forensic human identification, semiautomated computer-assisted image-capture systems (which act essentially as electronic rulers) quickly replaced this rudimentary method of data analysis. This improvement reduced labor, tedium, and subjectivity and enhanced reproducibility. At present, fluorescence detection systems allow labeled DNA fragments to be detected, analyzed, and downloaded to databases in real time. The inclusion of internal size standards also allows PCR-amplified allelic signatures to be accurately and reliably scored, while sophisticated software packages assign each amplified DNA segment an allele designation number.

Analyses of nucleotide sequence polymorphisms, such as those applied to mtDNA, are typically carried out using commercial fluorescence-based nucleotide sequencing kits. Here again, specialized data collection and analysis software suites are required to expedite the reading and interpretation of DNA sequences.

High-throughput fluorescence analysis of SNP signatures is currently performed on DNA microarrays (a miniaturized high-density format of the dot-blot system). In this assay, which is now fully automated, a series of oligonucleotide DNA probes corresponding to individual SNPs are spotted as a two-dimensional high-density array on the surface of a glass slide. The arrays then serve as baits when hybridized to fluorescently labeled amplified DNA fragments. The presence of an SNP signature is revealed by matching the fluorescence signal to the coordinates of the oligonucleotide probe on the array. Data capture and interpretation is achieved using dedicated scanners and computer software. The principal advantages of automated sample processing and data-acquisition platforms are the reduction of time, labor, and human error in the transcription of data. Data files can be accessed and further studied to determine the statistical significance and linkage properties of one or more SNPs.

1.5 Quality Assurance and Quality Control

Quality control management is paramount to obtaining quality results. It is also a key factor in establishing uniform reliability of data and practices between laboratories. Laboratory quality assurance consists in verifying and documenting that all procedures relevant to DNA typing exercises have been performed by skilled, highly trained personnel. The validity of DNA typing data relies on correct identification of true "nonmatches" as well as true "matches" from prospective DNA typing profiles. The reliability of a test reflects its reproducibility under defined conditions and its ability to transcend different laboratories and practitioners. Quality assurance must encompass all significant aspects of the DNA typing process. These include personnel education and training, documentation of records, data analysis, quality control of reagents and equipment, technical controls, proficiency testing, reporting of results, and auditing of laboratory procedures. The principles and practices of quality assurance and the development of appropriate standards and guidelines for

DNA typing emerged initially from the experience of clinical laboratories. These have now been carefully defined through consensus between forensic laboratories represented in the North American Non-Regulatory Federal Initiative, the Scientific Working Group on DNA Analysis Methods (SWGDAM), the European DNA Profiling Group (EDNAP), and the European Network of Forensic Science Institutes (ENFSI). Recently, some laboratories performing forensic analyses have developed a quality assurance system on the basis of international standards such as the ISO/IEC 17025. It is important to note that quality assurance guidelines in forensic science must retain enough flexibility to accommodate the unique nature and diversity of forensic samples as well as future advances in recombinant DNA technology, molecular biology, genetics, and instrumentation.

Most laboratories engaged in DNA typing analyses seek accreditation. Accreditation represents the formal recognition of the competence and quality of the laboratory and secures international recognition that the quality of the work performed meets exacting standards.

2 Applications

2.1 Criminal and Forensic Investigation

2.1.1 Human Identity

Violent crimes such as sexual assault or murder are often associated with multiple acts of aggression and defense that lead to the transfer of DNA-containing material between assailant and victim. Tissue, blood, or other biological material recovered from the crime scene, therefore, provide a source of forensic evidence that can help determine the circumstances of the event as well as the provenance of the specimens (Fig. 4). Accident or homicide victims, unidentifiable from their physical features, can be identified provided "typeable" DNA can be isolated from the victim's remains and matched with the allele profile generated from DNA extracted from, for example, hair roots or scalp cells collected from his or her hairbrush. If the victim's parents or other blood relatives are available, parentage testing may also be performed (Fig. 5). Genetic relatedness diminishes over generations. Biological parents provide a direct genetic link to the victim, whereas maternal or paternal aunts or grandparents share fewer alleles with the victim. Sisters and brothers of the victim can be either very informative or noninformative depending on which alleles are inherited by each member of the family. Genetic relatedness across multiple alleles can generally provide a combined DNA profile that allows the likelihood of relatedness to be determined. In essence, the more DNA tests performed or the closer the relatedness of potential blood relatives, the higher the probability of determining identity. In cases where nuclear genomic DNA may be lacking or unusable, mtDNA or loci from the nonrecombinogenic section of the male Y chromosome can be exploited to identify and link family members.

Reference samples often collected from victims, their parents, or other close relatives include whole blood and buccal swabs. Archival pathology specimens (such as paraffin-embedded tissue or histological sections mounted on glass slides), despite their small size and degree of processing, can serve as adequate resource material for the identification of individuals.

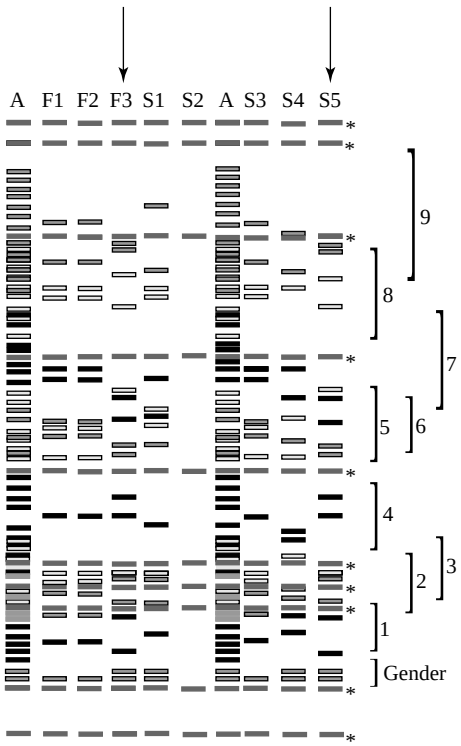

Fig. 4 STR analysis of a sexual assault case. In sexual assault cases, evidence samples such as oral, vaginal, and anal swabs are routinely collected and submitted for DNA typing analysis. Often, swabs present as a mixture of the assailant's sperm and the victim's epithelial cells. Both cell types can be separated from one another using a ''preferential lysis'' procedure that eventually releases crude genomic DNA from the epithelial cells in the first ''wash'' and second fractions (referred to as fractions F1 and F2 respectively), and crude genomic DNA from sperm cells in fraction F3. Each fraction is then subjected to sequential rounds of phenol/chloroform extractions to yield high-purity DNA that can then be used for PCR amplification using nine sets of STR primer pairs in a ''multiplex'' analysis. To facilitate real-time detection and sizing during electrophoresis on the DNA sequencing instrument, each set of allelic DNA fragments is labeled with its own fluorochrome ''tag'' during PCR amplification. Gel images are produced by the instrument and downloaded to a computer workstation where the analyst scrutinizes the data and interprets the allele profiles using specialized software suites. Individual STR allele signatures are scored against an internal DNA standard (marked with asterisks). The latter is a collection of red fluorescently tagged DNA fragments of known sizes that are included with each ''run'' sample on the gel in order to correct for minor variations in electrophoretic migration. The internal standard is also used to generate a calibration curve to estimate the size of each STR fragment. The allelic ladder (denoted as A in the diagram) is a composite of the alleles most frequently encountered for each of the nine STR loci surveyed in individuals sampled at random in the general population. Each band in the ladder has been sequenced so that the exact number of tandem repeats displayed in these fragments is known. The ladder is used to determine the number of tandem repeats present in each of the amplified fragments generated from biological samples. Appropriate positive and negative controls are always incorporated to verify the efficiency and accuracy of the extraction and amplification processes (referred to as S1 and S2 respectively, in the diagram). Because no amplified products are detected in lane S2, one can assert that the components used in the amplification reaction were not contaminated with any external source of DNA. As a rule, in processing a criminal case for DNA analysis, the evidentiary lanes on a gel are separated into ''questioned'' and ''known'' samples. The ''knowns'' usually correspond to DNA from fresh blood samples taken from victims (referred to here as S3) and suspects (referred to here as S4 and S5 for suspects 1 and 2 respectively). These ''knowns'' are compared to the ''questioned'' samples obtained from crime scenes, which, in this case, were vaginal swabs from the victim. Visual inspection and statistical analysis of the electrophoregram reveal matches between the F3 sperm cell fraction and the known sample from suspect 2 (S5) for all nine STR DNA regions examined. This is, therefore, a case of ''inclusion'' for suspect 2 (S5). Suspect 1 (S4), on the other hand, is ''excluded'' from being the perpetrator of this criminal offense. Statistical analyses are conducted to calculate the probability of a random match, that is, to estimate the odds of finding, by random chance alone, an unrelated individual, other than suspect 2, who would also produce matches on all nine DNA regions examined.

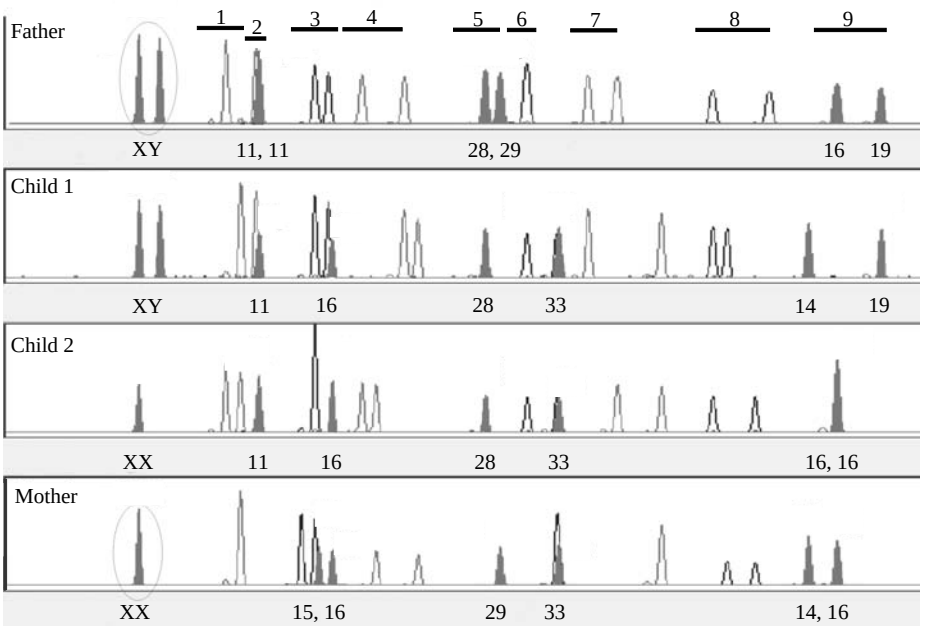

Fig. 5 Paternity analysis using short tandem repeats. Genetic profiles generated from blood samples or buccal swabs are shown for a typical family. Profiles are represented by a series of peaks that correspond to the sets of STR-containing allelic DNA fragments that migrated through a polyacrylamide gel in the DNA sequencing unit. The shorter STR fragments appear on the left side of the panel as they are the first to pass in front of the detection window of the instrument. The larger STR fragments appear on the right side of the tracing because their large size has delayed their migration and appearance in front of the detection window. In this particular example, nine STR loci (denoted as 1–9 on the diagram) have been amplified simultaneously in a "multiplex" PCR reaction. The test also includes an amplification reaction that provides an indication of the sex of the person from whom the sample originated. The presence of two peaks indicates a male signature (X,Y), whereas the presence of only one peak indicates a female (X,X) genotype. Since half of the genome of an individual is contributed from the mother and the other half is provided by the father, each of the nine STR loci examined has two potential signature (allelic) components (two peaks), one inherited from each parent. This provides a total of 18 points of comparison, which makes the test highly discriminating. It is therefore possible to identify which half of the children's DNA was contributed by which parent. In this particular example, child 1 has inherited allele 11 from his father and allele 16 from his mother for the STR region referred to as 2. For STR region 5, child 1 has inherited allele 28 from his father and allele 33 from his mother. For STR region 9, child 1 has inherited allele 14 from his mother and allele 19 from his father. Brothers and sisters (siblings) are likely to share one or more allele signatures as similar alleles may have been inherited from the mother or the father. This is in fact the case for the children of this family who have inherited the same alleles from their parents for STR regions 2 and 5. Conversely, siblings may not share any alleles. This is demonstrated for STR region 9 where child 1 has inherited allele 14 from his mother and allele 19 from his father and child 2 has inherited the other alleles from his mother and father, that is, alleles 16 and 16. Parentage tests such as this one have been an invaluable resource for establishing kinship and identity in the wake of the Swissair and World Trade Center tragedies.

With the evolution of analytical methods, the role of DNA profiling in mass disasters has become increasingly important as demonstrated in the wake of the TWA 800 and Swissair 111 plane crashes and terrorist attacks on the World Trade Center. In each of these cases, allele profiling enabled the identification of human remains where all other methods had failed. Incidents such as these have brought DNA profiling to the forefront of public awareness.

DNA methods have also facilitated the identification of casualties of war and other armed conflicts. Confidence in this approach is such that the military in some countries now request soldiers to provide a biological sample for the derivation of a reference DNA profile, should the need for a match comparison arise in the future.

DNA analysis has provided assistance in some unusual circumstances where forensic evidence was originally thought not to exist. Investigators have found that insects that had fed on blood from a victim could be used to recover the human DNA evidence that eventually led to the resolution of the case. Analysis of hematophagous arthropod (i.e. biting insects) blood meals and excreta may indeed be rare, but such examinations can prove instrumental in identifying the assailant or excluding one or more potential suspects from the scenes of violent crimes (e.g. rape, homicide, child abuse).

The automation of the STR-based DNA typing process for biological samples collected on FTA® paper, paper strips, or cotton swabs has played an important role in allowing the construction of databases containing the DNA profiles of criminals convicted of designated violent crimes, such as sexual assault, homicide, and armed robbery, around the world. The use of simple commercial collection kits allows biological samples to be collected safely and without risk of cross-contamination. In countries where criminal DNA databases have been implemented, DNA profiles from convicted offenders are compared to the DNA profiles generated from biological samples collected from crime scenes. A DNA data bank can link a convicted offender's allele profile to more than one crime scene (which suggests a serial offender) and can link known offenders to crime scenes where the perpetrator was unknown. Data banks such as these have proven extremely useful in assisting law enforcement agencies to build cases against suspected offenders, to focus investigations, to protect the public, and to exonerate those falsely accused. International law enforcement agencies and the justice community in general are, therefore, experiencing first hand the benefits of DNA profiling. The success of a database is directly related to the number of samples registered. When more samples are loaded into the data bank, there is a better chance of generating a match to crime-scene evidence or to offenders who have already been convicted. As new DNA typing technologies are developed, measures will have to be put in place to ensure that existing databases are not rendered obsolete.

2.1.2 Animal Identity

There are numerous examples in the forensic record where cases were resolved by pivotal DNA evidence derived from nonhuman biological samples. Common examples include confirming or excluding an animal as being the cause of a traffic accident associated with high material damage and personal injury. Dog hair left as evidence at the site of a motorcycle accident, for example, can be used to

incriminate the animal and establish the cause of an otherwise unsolvable accident. In addition, saliva from canine bite wounds on a deceased individual can be used to trace the animal responsible for a fatal attack.

Poaching of animals on the endangered species list is a criminal offense. Species identification as well as authentication of various processed biological materials can be achieved using DNA methods. A dried processed skin suspected to be that of a poached tiger was certified as an imitation derived from cattle skin painted to look like tiger skin. A similar approach was used to demonstrate that trace evidence recovered from the remains of an animal found at an illegal poaching site in a national park matched that of the confiscated frozen meat.

The protection of endangered animal species requires inexpensive, accurate, reliable, rapid, and easily adaptable strategies like those offered by PCR-based DNA analysis techniques to verify identity. Protein-based identification techniques become less reliable with heat-treated products, such as canned or smoked fish. Processed seafood products (like caviar and other meat preparations) often contain material from several species even though regulations stipulate that the contents of a container should be unique. Laws also require that products be labeled with official names to discourage fraud. The use of genotyping methods to determine the authenticity of samples can assist international conservation and legal agencies to better manage and preserve commercial species populations while protecting species at risk. PCR-based methods give wholesalers alternatives to crude identification methods and, more importantly, protect consumer groups against paying exorbitant prices for mislabeled food products.

2.1.3 Plant Identity

DNA typing analysis has been useful in the identification of regulated or prohibited plant material, such as Cannabis (marijuana). PCR-based methods are so sensitive that they allow detection of the presence of *Cannabis sativa* DNA on the skin of persons who recently handled both leaf and resinous material.

The traditional method of detecting the fraudulent propagation of ornamental cultivars required much time and effort, principally because it was based on the phenotypic analysis of plants grown to the flowering stage. This laborious approach is rapidly being superseded through implementation of databases of reference cultivar DNA profiles. By way of example, the rose is one of the most economically important landscape or cut-flower ornamental species in the world. In France, efforts are being deployed to derive reference genotypes for the more than 1500 rose varieties maintained in the French reference collection by PCR-based methods. PCR-based assays have also assisted in solving cases of unauthorized sale of spurious chili seeds raised by an individual and marketed under the brand name of an elite variety. In the same vein, standard-quality rice has been sold as the premium-priced basmati rice. Similar typing methods were successfully applied to expose the fraud.

VNTR markers have also proven effective for genotyping individual trees in forest crime prosecutions. Illegal tree harvesting destroys habitat and reduces diversity within parkland and protected ecological sectors. Theft of standing timber continues to represent a growing problem in the western part of Canada where red

cedar (*Thuja plicata*) is particularly targeted by thieves because of its high commercial value. Forensic experts are now using microsatellite markers to match illegally harvested tree stumps to suspected stolen wood. In addition, chloroplast DNA analysis allows discrimination of species from dry wood samples. On a more sombre note, the RAPD technique has been successfully used to link seedpods recovered from the truck of a suspect's car to the paloverde tree under which a murder victim had been buried.

2.1.4 Fungal Identity

Several truffle species are among the most valuable mushrooms on the market because of their aroma and culinary qualities. They represent a valuable source of agricultural income in some areas of southern Europe. DNA profiling allows rapid control of production quality and identification of truffle fruit bodies in fresh or canned preparations. The recent occurrence of Asian truffles morphologically similar to *Tuber melanosporum* on the European market raises two kinds of problems – one is fraud, because Asian truffles are cheaper and of lower quality, the other is ecological in nature, because Asian black truffles may be used (knowingly or not) as inoculum for the production of seedlings destined to be planted in Europe.

2.1.5 Bacterial Identity

Forensic investigations often require that the presence of ethanol in postmortem blood and tissue samples be determined. However, it is well documented that a variety of microbes can produce significant amounts of ethanol under postmortem conditions. DNA analysis allows investigators to identify the microbial species present in tissues, blood, urine, and other body fluid samples from victims in order to help determine if the cause of a fatal motor vehicle or airplane accident was indeed attributed to the consumption of alcoholic beverages. In such cases, sets of PCR primers are usually designed to hybridize to the 16S ribosomal RNA genes in order to help identify microbial species. More recently, VNTR and STR typing were employed to identify the strain of anthrax delivered through mail by terrorists in New York, Washington DC, and Florida.

2.1.6 Viral Identity

When sexual assault results in HIV transmission, a criminal charge of conduct endangering life, in addition to rape, may become an option for the prosecution if the carrier was aware of his infection status. In such a case, direct sequencing of the viral genome is used to demonstrate that the strain infecting the perpetrator matches that of the victim. Other cases of criminal viral transmission that have been resolved by DNA analysis include those of a dentist with AIDS who infected several clients and HIV-infected surgeons who transmitted the virus to their patients.

2.2 Clinical Medicine

The ability to discern whether a specimen originates from a particular source is often critical to diagnosis, to laboratory analysis, and to exercises where the endpoint is quality assurance. For example, because the results of a positive drug test could have significant economic and social consequences, the number of submissions of adulterated or substituted urine specimens has increased. DNA analysis acts as a safeguard by allowing confirmation that the urine that tested positive for the drug of abuse was in fact provided by the donor.

DNA analysis in the clinical laboratory is paramount to ascertain the provenance of tissue biopsy specimens (i.e. normal vs malignant tissue) and blood specimens where either could be mislabeled at the time of collection. Under such circumstances, accurate specimen identification prevents misdiagnosis and unnecessary treatment. VNTR-based methods have recently been used to demonstrate that a malignant melanoma that had suddenly arisen in a kidney transplant recipient had in fact originated from the tissue of the donor.

Other applications in medicine include genetic counseling, tracing the percentage of donor versus recipient cells in bone marrow transplants, determination of possible "contamination" of fetal chorionic villi sample (CVS) tissue with maternal tissue, and the confirmation of twin zygosity.

The application of STR genotype analysis to demonstrate the occurrence of needle sharing provides an excellent tool to evaluate the safety of the needle-exchange program as an adjunct to attempts at reducing the risk of HIV transmission in intravenous drug users. Since common injection practices leave sufficient quantities of trace DNA that can be efficiently amplified by PCR, investigators can easily profile the history of use for each syringe.

DNA profiling offers rapid and accurate identity testing in human and other animal cell lines registered in large repositories. This is particularly important because a significant fraction of research results could be misinterpreted if the cell lines used turn out not to be of the specified origin. Cross-contamination of cell lines continues to plague repositories and constitutes an unacceptably frequent cause of data correction or retraction. DNA methods allow comparison of parental and derived cell lines and ensure their authenticity.

The ability to identify the victims of blood-sucking insects is valuable to medical and forensic entomology. Studying variations in the biting rates and insect preference for different people could have important consequences for the epidemiology of vector-borne diseases. Detailed knowledge about the patterns of the mosquito's feeding behavior are used to develop improved disease control strategies by taking into account patterns of potential vector spread within and between communities and by concentrating intervention strategies on those individuals most at risk of sustaining bites and contracting infectious diseases.

2.3 Parentage Studies

2.3.1 Human Testing

Parentage can be determined for child custody and immigration cases. This is easily established because the bands in the offspring's allele profile must, by definition, have been inherited from the biological parents. Barring the rare instance of mutations, the presence of bands not found in either parental profile is indicative of nonparentage. A dramatic demonstration of a parentage study using STR markers and mtDNA sequencing was the identification of the remains of the lost members of the Romanov family in Russia.

2.3.2 Testing of Nonhuman Animal Species

Paternity testing has become an important tool for breeders of dogs, sheep, horses, and cattle. Biological relationships can be confirmed in artificial insemination and embryo transfer programs for

domesticated animals. DNA profiling is also used in breeding programs for populations of endangered species with small gene pools. Animals exhibiting the greatest number of DNA profile differences are taken as possessing the greatest degree of genetic diversity and are therefore selected for breeding. DNA profiling has also proven effective in identification of stolen or lost animals as well as in the establishment of relevant breeding (trait) markers. Recently, a set of bovine SNPs has been developed and is expected to facilitate the high-throughput genotyping of prize animals. High-resolution genetic markers such as microsatellites can provide detailed information about the mating habits of animal species such as alligators, turtles, chimpanzees, and Gypsy moths. Such analyses can reveal the reproductive dynamics of local populations and can provide much needed insight on population genetics and ecology.

2.4 Molecular Evolution

2.4.1 Evolution of Human and Other Animal Species

DNA typing analysis has revolutionized the study of human evolution. Numerous population studies examining VNTRs and nucleotide sequence polymorphisms have formed the basis of exciting new theories about human origins, diversity, and cultural expansions from the distribution and frequency of genes and genetic markers found around the world today.

Many museums routinely archive animal tissue for molecular, systematic, and taxonomic studies. The vast majority of museum specimens predates the advent of molecular techniques and consists of dried skins, skeletons, and formalin-fixed or ethanol-preserved material. The success of DNA recovery depends largely on the condition of the specimen. Small bones, only a few millimeters in length are a surprisingly excellent source of relatively intact DNA. Such material increases the value of museum specimens to systematic biologists. In fact, one of the first animal DNA analyses performed using PCR consisted in typing a museum skin specimen of the now extinct quagga.

The evolution of populations can be traced, for example, by analyzing DNA from well-preserved (i.e. dried) museum specimens or from insects naturally embedded in amber for millions of years. Base-pair sequences and allele-profile bands can be compared with those of modern relatives. PCR-based markers are being used extensively for reconstructing the phylogenies of various species. Such analyses provide much needed information regarding the fine timescale over which closely related species have diverged and about what sorts of genetic variations are associated with the emergence of species.

2.4.2 Evolution of Plants

Microsatellite markers have proven to be extremely valuable in the analysis of gene pool variations of crops during the process of cultivar development and in the classification of their germplasm. It is extremely important to study the genetic composition of the germplasm of existing modern-day cultivars and to compare the data with those of their ancestors and related species because studies such as these provide insight on speciation and divergence patterns.

Plant improvement, either by natural selection or through the efforts of breeders, has always relied upon creating,

evaluating, and selecting the right combination of alleles. Microsatellite markers are efficiently employed for "tagging" several traits (like yield, disease resistance, stress tolerance, and seed quality) that are beneficial to a breeding program.

3 Perspectives

3.1 DNA Typing versus Conventional Methods

The introduction of genetic methods to the field of human identification has all but replaced long-standing biochemical and serological techniques. The conventional methods of isoenzyme gel electrophoresis and blood antigen typing are admittedly rapid, simple, and definitive for exclusions, but they are much less powerful than DNA typing analysis for determining the probability of inclusion. If a mother and a presumptive father both share blood type O, their offspring cannot express the antigens for blood types A, B, or AB. Thus, the presumptive father is readily excluded as one of the biological parents in this example. However, if the offspring is of blood type O, exclusion of a man who is not the biological father but is also of blood type O (or heterozygous for the O allele) is considerably more difficult if the investigator relies solely on serological methods.

Additional key elements that have contributed to the adoption of DNA as the substrate for identity testing across every taxonomic kingdom include the ready availability of the hereditary material, the highly polymorphic nature of genomes, the continuity of the genetic script within each cell of an organism, the chemical stability of DNA relative to that of enzymes and other proteins, and the technical ability to amplify DNA regions in the laboratory when only a minimal amount of biological material, even partially degraded, is available.

DNA evidence is a matter of probabilities. In the absence of mutation, when two samples do not share similar DNA allele profiles, one can state with confidence that neither could have originated from the same individual. In this regard, the genetic evidence leading to the exclusion of a suspect sample is absolute. If two samples share the same genotype, the question now becomes, "What is the significance of the match?" If a sufficient number of polymorphic DNA markers are analyzed, then the power of discrimination becomes very high and the statistical probability of a random match becomes extremely low.

3.2 Future Directions

DNA typing technologies have consistently evolved over the past fifteen years. The trend is expected to continue and will probably be fueled by the various genome characterization and sequencing initiatives currently under way around the globe. The quest to unravel the secrets that lie within the genetic scripts of living organisms will no doubt lead to the discovery of novel polymorphisms and the development of selected sets of SNPs with exquisite powers of discrimination. Significant efforts are already being deployed toward the development of more efficient, effective platforms and analytical instrument systems to facilitate high-throughput surveys of SNPs and other elements of genetic variation. Prime examples of recent advances leading to the wider acceptance of SNP-based identification strategies include automation and the

design of prototype ''lab-on-a-chip'' microanalytical devices capable of performing cell separations, ultrarapid PCR amplification, liquid stream electrophoresis, resolution of DNA fragments, DNA sequencing, real-time data capture, and downstream processing. Parallel innovations in microarray design and high-density spotting are also anticipated to enhance our ability to expose genetic variation between individuals as a matter of routine practice both in centralized analytical facilities and in the field. With respect to the necessity of being able to work with limiting amounts of biological material, refinement of whole genome amplification strategies remains a priority so that forensic scientists may one day perform DNA analyses on samples containing a single genome equivalent. Equally important for single-cell or microanalysis will be the development of reiterative processes that will allow typing runs to be repeated many times on the same sample in order to sustain accuracy, reproducibility, and statistical significance. Whatever future developments hold in store, collectively, all of them will converge on ensuring that DNA typing processes can be performed at low cost to the end user with greater speed and greater efficiency without compromising accuracy or profile quality.

See also Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Bennett, P. (2000) Microsatellites, *Mol. Pathol.* **53**, 177–183.

Budowle, B., Moretti, T., Smith, J., DiZinno, J. (2000) *DNA Typing Protocols: Molecular Biology and Forensic Analysis, BioTechniques Books BioForensic Sciences Series*, Eaton Publishing, Natick, MA.

Burczak, J.D., Mardis, E. (Eds.) (2000) *Polymorphism Detection and Analysis, BioTechniques Books Update Series*, Eaton Publishing, Natick, MA.

Butler, J. (2001) *Forensic DNA Typing – Biology & Technology Behind STR Markers*, Academic Press, San Diego, CA.

Duncan, G.T., Tracey, M.L. (1997) Serology and DNA Typing, in: Eckert, W.G. (Ed.) *Introduction to Forensic Sciences*, 2nd edition, CRC Press, Boca Raton, FL, pp. 233–293.

Hajeer, A., Worthington, J., John, S. (Eds.) (2000) *SNP and Microsatellite Genotyping – Markers for Genetic Analysis, BioTechniques Books*, Eaton Publishing, Natick, MA.

Holland, M.M., Parsons, T.J. (1999) Mitochondrial DNA sequence analysis Validation and use for forensic casework, *Forensic Sci. Rev.* **11**, 21–50.

Joshi, S.P., Ranjekar, P.K., Gupta, V.S. (1999) Molecular markers in plant genome analysis, *Curr. Sci.* **77**, 230–240.

Moxon, E.R., Wills, C. (1999) DNA microsatellites: agents of evolution? *Scientific Am.* **280**, 94–99.

Popping, B. (2002) The application of biotechnological methods in authenticity testing, *J. Biotechnol.* **98**, 107–112.

Rudin, N., Inman, K. (2002) *An Introduction to Forensic DNA Analysis*, 2nd edition, CRC Press, New York.

Primary Literature

Alford, R.L., Hammond, H.A., Coto, I., Caskey, C.T. (1994) Rapid and efficient resolution of parentage by amplification of short tandem repeats, *Am. J. Hum. Genet.* **55**, 190–195.

Andréasson, H., Asp, A., Alderborn, A., Gyllensten, U., Allen, M. (2002) Mitochondrial sequence analysis for forensic identification using pyrosequencing technology, *BioTechniques* **32**, 124–133.

Andréasson, H., Gyllensten, U., Allen, M. (2002) Real-time DNA quantification of nuclear and mitochondrial DNA in forensic analysis, *BioTechniques* **33**, 402–411.

Arana, A., Soret, B., Lasa, I., Alfonso, L. (2002) Meat traceability using DNA markers: application to the beef industry, *Meat Sci.* **61**, 367–373.

Baker, C.S., Cipriano, F., Palumbi, S.R. (1996) Molecular genetic identification of whale and dolphin products from commercial markets in Korea and Japan, *Mol. Ecol.* **5**, 671–685.

Bartlett, S.E., Davidson, W.S. (1992) FINS (Forensically Informative Nucleotide Sequencing): a procedure for identifying the animal origin of biological specimens, *BioTechniques* **12**, 408–411.

Bataille, M., Crainic, K., Leterreux, M., Durigon, M., de Mazancourt, P. (1999) Multiplex amplification of mitochondrial DNA for human and species identification in forensic evaluation, *Forensic Sci. Int.* **99**, 165–170.

Batzer, M.A., Deininger, P.L. (2002) Alu repeats and human genomic diversity, *Nat. Rev./Genet.* **3**, 370–379.

Beamonte, D., Guerra, A., Ruíz, B., Alemany, J. (1995) Microsatellite DNA polymorphism analysis in a case of an illegal cattle purchase, *J. Forensic Sci.* **40**, 692–694.

Birch, C.J., McCaw, R.F., Bulach, D.M., Revill, P.A., Carter, J.T., Tomnay, J., Hatch, B., Middleton, T.V., Chibo, D., Catton, M.G., Pankhurst, J.L., Breschkin, A.M., Locarnini, S.A., Bowden, D.S. (2000) Molecular analysis of human immunodeficiency virus strains associated with a case of criminal transmission of the virus, *J. Infect. Dis.* **182/3**, 941–944.

Bossier, P. (1999) Authentication of seafood products by DNA patterns, *J. Food Sci.* **64**, 189–193.

Budowle, B., Baechtel, F.S., Comey, C.T., Giusti, A.M., Klevan, L. (1995) Simple protocols for typing forensic biological evidence: chemiluminescent detection for human DNA quantitation and RFLP analyses and manual typing of PCR amplified polymorphisms, *Electrophoresis* **16**, 1559–1567.

Chakraborty, R., Stivers, D.N., Su, B., Zhong, Y., Budowle, B. (1999) The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems, *Electrophoresis* **20**, 1682–1696.

Chow-Shaffer, E., Sina, B., Hawley, W.A., De Benedictis, J., Scott, T.W. (2000) Laboratory and field evaluation of polymerase chain reaction-based forensic DNA profiling for use in identification of human blood meal sources of *Aedes aegypti* (Diptera: Culicidae), *J. Med. Entomol.* **37**, 492–502.

Clayton, T.M., Whitaker, J.P., Maguire, C.N. (1995) Identification of bodies from the scene of a mass disaster using DNA amplification of short tandem repeat (STR) loci, *Forensic Sci. Int.* **76**, 7–15.

Constable, J.L., Ashley, M.V., Goodall, J., Pusey, A.E. (2001) Noninvasive paternity assignment in Gombe chimpanzees, *Mol. Ecol.* **10**, 1279–1300.

Corach, D., Sala, A., Penacino, G., Sotelo, A. (1995) Mass disasters: rapid molecular screening of human remains by means of short tandem repeats typing, *Electrophoresis* **16**, 1617–1623.

Deguilloux, M.-F., Pemonge, M.-H., Petit, R.J. (2002) Novel perspectives in wood certification and forensics: dry wood as a source of DNA, *Proc. R. Soc. London, Ser. B* **269**, 1039–1046.

Del Rio, S.A. (1998) Cost-Effectiveness in Sample Processing Using the FTA™ Treated Stain Card for High Throughput, *Proceedings of the 8th International Symposium on Human Identification*, Scottsdale, AZ, Sept. 17–20 1997, Promega Corporation, Madison, WI, pp. 64–69.

Dequeker, E., Ramsden, S., Grody, W.W., Stenzel, T.T., Barton, D.E. (2001) Quality control in molecular genetic testing, *Nat. Rev. /Genet.* **2**, 717–723.

DeSalle, R., Birstein, V.J. (1996) PCR identification of black caviar, *Nature* **381**, 197–198.

Dirks, W., MacLeod, R.A.F., Jäger, K., Milch, H., Drexler, H.G. (1999) First searchable database for DNA profiles of human cell lines: sequential use of fingerprint techniques for authentication, *Cell. Mol. Biol.* **45**, 841–853.

Edwards, A., Civitello, A., Hammond, H.A., Caskey, C.T. (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats, *Am. J. Hum. Genet.* **49**, 746–756.

Frégeau, C.J., Aubin, R.A., Elliott, J.C., Gill, S.S., Fourney, R.M. (1995) Characterization of human lymphoid cell lines GM9947 and GM9948 as intra- and interlaboratory reference standards for DNA typing, *Genomics* **28**, 184–197.

Frégeau, C.J., Bowen, K.L., Bowen, J.H., Gaudette, B.D., Fourney, R.M. (1998) National Casework and National DNA Database: The Royal Canadian Mounted Police Perspective, in: Olaisen, B., Brinkmann, B., Lincoln, P.J. (Eds.) *Progress in Forensic Genetics 7, Proceedings of the 17th Congress of the International Society for Forensic Haemogenetics, International Congress Series 1167*, Oslo, Norway, Sept. 2–6 1997, Elsevier Science BV, Amsterdam, the Netherlands, pp. 541–543.

Frégeau, C.J., Fourney, R.M. (1993) DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification, *BioTechniques* **15**, 100–119.

Frégeau, C.J., Leclair, B., Bowen, K., Porelle, F., Fourney, R.M. (2002) The National DNA Data Bank of Canada – A Laboratory Bench Retrospective on the First Year of Operation, in Olaisen, B., Brinkmann, B., Lincoln, P.J. (Eds.) Progress in Forensic Genetics 9, *Proceedings of the 19th Congress of the International Society of Forensic Genetics, International Congress Series 1239*, Münster, Germany, Aug. 28–Sept. 1 2001, Elsevier Science BV, Amsterdam, the Netherlands, pp. 621–625.

Frégeau, C.J., Vanstone, H., Borys, S., McLean, D., Maroun, J.A., Birnboim, H.C., Fourney, R.M. (2001) AmpFlSTR® Profiler Plus™ and AmpFlSTR® COfiler™ analysis of tissues stored in GenoFix™, a new tissue preservation solution for mass disaster DNA identification, *J. Forensic Sci.* **46**, 1180–1190.

Gigliano Siniscalco, G., Di Finizio, A. (1999) The *Cannabis sativa L.* fingerprint as a tool in forensic investigations, *Bull. Narc.* **49** and **50** (1 and 2), 129–137.

Gill, P. (2001) An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes, *Int. J. Leg. Med.* **114**, 204–210.

Gill, P., Ivanov, P.L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E., Sullivan, K. (1994) Identification of the remains of the Romanov family by DNA analysis, *Nat. Genet.* **6**, 130–135.

Gill, P., Kimpton, C.P., Urquhart, A., Oldroyd, N., Millican, E.S., Watson, S.K., Downes, T.J. (1995) Automated short tandem repeat (STR) analysis in forensic casework – a strategy for the future, *Electrophoresis* **16**, 1543–1552.

Harbison, S.A., Hamilton, J.F., Walsh, S.J. (2001) The New Zealand DNA databank: its development and significance as a crime solving tool, *Sci. Justice* **41**, 33–37.

Heaton, M.P., Harhay, G.P., Bennett, G.L., Stone, R.T., Grosse, W.M., Casas, E., Keele, J.W., Smith, T.P.L., Chitko-Mckown, C.G., Laegreid, W.W. (2002) Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle, *Mamm. Genome* **13**, 272–281.

Hellmann, A., Rohleder, U., Schmitter, H., Wittig, M. (2001) STR typing of human telogen hairs – a new approach, *Int. J. Leg. Med.* **114**, 269–273.

Holt, C.L., Buoncristiani, M., Wallin, J.M., Nguyen, T., Lazaruk, K.D., Walsh, P.S. (2002) TWGDAM validation of AmpFlSTR™ PCR amplification kits for forensic DNA casework, *J. Forensic Sci.* **47**, 66–96.

Horswell, J., Cordiner, S.J., Maas, E.W., Martin, T.M., Sutherland, B.W., Speir, T.W., Nogales, B., Osborn, A.M. (2002) Forensic comparison of soils by bacterial community DNA profiling, *J. Forensic Sci.* **47**, 350–353.

Ikegaya, H., Iwase, H., Sugimoto, C., Yogo, Y. (2002) JC virus genotyping offers a new means of tracing the origins of unidentified cadavers, *Int. J. Leg. Med.* **116**, 242–245.

Ivanov, P.L., Wadhams, M.J., Roby, R.K., Holland, M.M., Weedn, V.W., Parsons, T.J. (1996) Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II, *Nat. Genet.* **12**, 417–420.

Iudica, C.A., Whitten, W.M., Williams, N.H. (2001) Small bones from dried mammal museum specimens as a reliable source of DNA, *BioTechniques* **30**, 732–736.

Kasai, K., Nakamura, Y., White, R. (1990) Amplification of a variable number of tandem repeats (VNTR) locus (pMCT118) by the polymerase chain reaction (PCR) and its application to forensic science, *J. Forensic Sci.* **35**, 1196–1200.

Kereszturya, L., Rajczya, K., Lászikb, A., Gyódia, É., Pénzes, M., Falus, A., Petrányia, G.G. (2002) Combination of DNA-based and conventional methods to detect human leukocyte antigen polymorphism and its use for paternity testing, *Am. J. Forensic Med. Pathol.* **23**, 57–62.

Kloosterman, A.D. (2001) Credibility of forensic DNA typing is driven by stringent quality standards, *Accredit. Qual. Assur.* **6**, 409–414.

Kocagöz, T., Yilmaz, E., Özkara, S., Kocagöz, S., Hayran, M., Sachedeva, M., Chambers, H.F. (1993) Detection of *Mycobacterium tuberculosis* in sputum samples by polymerase chain reaction using a simplified procedure, *J. Clin. Microbiol.* **31**, 1435–1438.

Krawczak, M. (1999) Informativity assessment for biallelic single nucleotide polymorphisms, *Electrophoresis* **20**, 1676–1681.

Kreike, J., Kampfer, S. (1999) Isolation and characterization of human DNA from mosquitoes (Culicidae), *Int. J. Leg. Med.* **112**, 380–382.

Kumar, L.D., Kathirvel, M., Rao, G.V., Nagaraju, J. (2001) DNA profiling of disputed chilli samples (*Capsicum annum*) using ISSR-PCR and FISSR-PCR marker assays, *Forensic Sci. Int.* **116**, 63–68.

Kupfer, D.M., Chaturvedi, A.K., Canfield, D.V., Roe, B.A. (1999) PCR-based identification of postmortem microbial contaminants – a preliminary study, *J. Forensic Sci.* **44**, 592–596.

Landegren, U., Nilsson, M., Kwok, P.-Y. (1998) Reading bits of genetic information: methods for single-nucleotide polymorphism analysis, *Genome Res.* **8**, 769–776.

Lantz, P.-G., Matsson, M., Wadström, T., Rådström, P. (1997) Removal of PCR inhibitors from human faecal samples through the use of an aqueous two-phase system for sample preparation prior to PCR, *J. Microbiol. Methods* **28**, 159–167.

Lazaruk, K., Wallin, J., Holt, C., Nguyen, T., Walsh, P.S. (2001) Sequence variation in humans and other primates at six short tandem repeat loci used in forensic identity testing, *Forensic Sci. Int.* **119**, 1–10.

Lebo, R.V., Maher, T., Farrer, L., Fenerci, E.Y., Milunsky, J.M. (2001) Highly polymorphic short tandem repeat analyses clarify complex molecular test results, *Diagn. Mol. Pathol.* **10**, 179–189.

Leclair, B., Frégeau, C.J., Aye, M.T., Fourney, R.M. (1995) DNA typing for bone marrow engraftment follow-up after allogeneic transplant: a comparative study of current technologies, *Bone Marrow Transplant.* **16**, 43–55.

Leclair, B., Frégeau, C.J., Bowen, K.L., Borys, S.B., Elliott, J., Fourney, R.M. (2000) Enhanced Kinship Analysis and STR-Based DNA Typing for Human Identification in Mass Disasters, in: Sensabaugh, G.F., Lincoln, P.J., Olaisen, B. (Eds.) *Progress in Forensic Genetics 8, Proceedings of the 18th Congress of the International Society for Forensic Haemogenetics, International Congress Series 1193*, San Francisco, USA, Aug. 17–21 1999, Elsevier Science BV, Amsterdam, the Netherlands, pp. 91–93.

Legrand, B., de Mazancourt, P., Durigon, M., Khalifat, V., Crainic, K. (2002) DNA genotyping of unbuffered formalin fixed paraffin embedded tissues, *Forensic Sci. Int.* **125**, 205–211.

Linch, C.A., Whiting, D.A., Holland, M.M. (2001) Human hair histogenesis for the mitochondrial DNA forensic scientist, *J. Forensic Sci.* **46**, 844–853.

Lord, W.D., DiZinno, J.A., Wilson, M.R., Budowle, B., Taplin, D., Meinking, T.L. (1998) Isolation, amplification, and sequencing of human mitochondrial DNA obtained from human crab louse, *Pthirus pubis* (L.), blood meals, *J. Forensic Sci.* **43**, 1097–1100.

Luckenbach, C., Gärtner, V.H., Seidl, C., Ritter, H. (1996) Tumor Inoculation Between Two Unrelated Human Individuals: STR Analysis of Paraffin-Embedded Tissue Section, in Carracedo, A., Brinkmann, B., Bär, W. (Eds.) *Advances in Forensic Haemogenetics 6, Proceedings of the 16th Congress of the International Society for Forensic Haemogenetics*, Santiago de Compostela, Spain, Sept. 12–16 1995, Springer-Verlag, Berlin, Germany, pp. 192–194.

Machuca, R., Jørgensen, L.B., Theilade, P., Nielsen, C. (2001) Molecular investigation of transmission of human immunodeficiency virus type 1 in a criminal case, *Clin. Diagn. Lab. Immunol.* **8**, 884–890.

MacLeod, R.A.F., Dirks, W.G., Matsuo, Y., Kaufmann, M., Milch, H., Drexler, H.G. (1999) Widespread intraspecies cross-contamination of human tumor cell lines arising at source, *Int. J. Cancer* **83**, 555–563.

Martin, P.D., Schmitter, H., Schneider, P.M. (2001) A brief history of the formation of DNA databases in forensic science within Europe, *Forensic Sci. Int.* **119**, 225–231.

Masters, J.R., Thomson, J.A., Daly-Burns, B., et al. (2001) Short tandem repeat profiling provides an international reference standard for human cell lines, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8012–8017.

Micka, K.A., Sprecher, C.J., Lins, A.M., Comey, C.T., Koons, B.W., Crouse, C., Endean, D., Pirelli, K., Lee, S.B., Duda, N., Ma, M., Schumm, J.W. (1996) Validation of multiplex polymorphic STR amplification sets developed for personal identification applications, *J. Forensic Sci.* **41**, 582–590.

Moor, D., Brodmann, P., Nicholas, G., Eugster, A. (2002) Polymerase chain reaction (PCR) for the detection of king bolete (*Boletus edulis*) and slippery jack (*Suillus luteus*) in food samples, *Eur. Food Res. Technol.* **214**, 340–345.

Moretti, T.R., Baumstark, A.L., Defenbaugh, D.A., Keys, K.M., Smerick, J.B., Budowle, B. (2001) Validation of short tandem repeats

(STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples, *J. Forensic Sci.* **46**, 647–660.

Morgante, M., Hanafey, M., Powell, W. (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes, *Nat. Genet.* **30**, 194–200.

Müller, S., Flekna, G., Müller, M., Brem, G. (1999) Use of canine microsatellite polymorphisms in forensic examinations, *J. Hered.* **90**, 55–56.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., White, R. (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping, *Science* **235**, 1616–1622.

Natarajan, P., Trinh, T., Mertz, L., Goldsborough, M., Fox, D.K. (2000) Paper-based archiving of mammalian and plant samples for RNA analysis, *BioTechniques* **29**, 1328–1333.

O'Brien, S.J. (2001) Cell culture forensics, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7656–7658.

Olaisen, B., Stenersen, M., Mevåg, B. (1997) Identification by DNA analysis of the victims of the August 1996 Spitsbergen civil aircraft disaster, *Nat. Genet.* **15**, 402–405.

Pádár, Z., Egyed, B., Kontadakis, K., Füredi, S., Woller, J., Zöldág, L., Fekete, S. (2002) Canine STR analyses in forensic practice – observation of a possible mutation in a dog hair, *Int. J. Leg. Med.* **116**, 286–288.

Parson, W., Pegoraro, K., Niederstätter, H., Föger, M., Steinlechner, M. (2000) Species identification by means of the cytochrome b gene, *Int. J. Leg. Med.* **114**, 23–28.

Poetsch, M., Seefeldt, S., Maschke, M., Lignitz, E. (2001) Analysis of microsatellite polymorphism in red deer, roe deer, and fallow deer – possible employment in forensic applications, *Forensic Sci. Int.* **116**, 1–8.

Prakash, S., Patole, M.S., Ghumatkar, S.V., Nandode, S.K., Shinde, B.M., Shouche, Y.S. (2000) Mitochondrial 12S rRNA sequence analysis in wildlife forensics, *Curr. Sci.* **78**, 1239–1241.

Radtkey, R., Feng, L., Muralhidar, M., Duhon, M., Canter, D., DiPierro, D., Fallon, S., Tu, E., McElfresh, K., Nerenberg, M., Sosnowski, R. (2000) Rapid, high fidelity analysis of simple sequence repeats on an electronically active DNA microchip, *Nucleic Acids Res.* **28**, e17 i–vi.

Rajora, O.P., Rahman, M.H., Dayanandan, S., Mosseler, A. (2001) Isolation, characterization, inheritance and linkage of microsatellite DNA markers in white spruce (*Picea glauca*) and their usefulness in other spruce species, *Mol. Gen. Genet.* **264**, 871–882.

Reeder, D.J. (1999) Impact of DNA typing on standards and practice in the forensic community, *Arch. Pathol. Lab. Med.* **123**, 1063–1065.

Replogle, J., Lord, W.D., Budowle, B., Meinking, T.L., Taplin, D. (1994) Identification of host DNA by amplified fragment length polymorphism analysis: preliminary analysis of human crab louse (Anoplura: Pediculidae) excreta, *J. Med. Entomol.* **31**, 686–690.

Romualdi, C., Balding, D., Nasidze, I.S., Risch, G., Robichaux, M., Sherry, S.T., Stoneking, M., Batzer, M.A., Barbujani, G. (2002) Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms, *Genome Res.* **12**, 602–612.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, *Science* **239**, 487–491.

Saiki, R.K., Walsh, P.S., Levenson, C.H., Erlich, H.A. (1989) Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6230–6234.

Sajantila, A., Lukka, M. (1994) Proving Paternity Using Formalin-Fixed, Paraffin-Embedded Tissue Samples From a Deceased Father, in: Bär, W., Fiori, A., Rossi, U. (Eds.) *Advances in Forensic Haemogenetics 5, Proceedings of the 15th Congress of the International Society for Forensic Haemogenetics*, Venezia, Italy, Oct. 13–15 1993, Springer-Verlag, Berlin, Germany, pp. 232–234.

Salas, A., Lareu, V., Calafell, F., Bertranpetit, J., Carracedo, Á. (2000) mtDNA hypervariable region II (HVII) sequences in human evolution studies, *Eur. J. Hum. Genet.* **8**, 964–974.

Savolainen, P., Arvestad, L., Lundeberg, J. (2000) A novel method for forensic DNA investigations: repeat-type sequence analysis of tandemly repeated mtDNA in domestic dogs, *J. Forensic Sci.* **45**, 990–999.

Schoell, W.M.J., Klintschar, M., Mirhashemi, R., Pertl, B. (1999) Separation of sperm and vaginal cells with flow cytometry for DNA

typing after sexual assault, *Obstet. Gynecol.* **94**, 623–627.

Séjalon-Delmas, N., Roux, C., Martins, M., Kulifaj, M., Bécard, G., Dargent, R. (2000) Molecular tools for the identification of *Tuber melanosporum* in agroindustry, *J. Agric. Food Chem.* **48**, 2608–2613.

Shrestha, S., Strathdee, S.A., Brahmbhatt, H., Farzadegan, H., Vlahov, D., Smith, M.W. (2000) Short tandem repeat methodology for genotypic identification of single-person versus multi-person use of syringes, *AIDS* **14**, 1507–1513.

Sparkes, R., Kimpton, C., Watson, S., Oldroyd, N., Clayton, T., Barnett, L., Arnold, J., Thompson, C., Hale, R., Chapman, J., Urquhart, A., Gill, P. (1996) The validation of a 7-locus multiplex STR test for use in forensic casework I. Mixtures, ageing, degradation and species studies, *Int. J. Leg. Med.* **109**, 186–194.

Sullivan, K.M., Walton, A., Kimpton, C., Tully, G., Gill, P. (1993) Fluorescence-based DNA segment analysis in forensic science, *Biochem. Soc. Trans.* **21**, 116–120.

Tsongalis, G.J., Wu, A.H.B., Silver, H., Ricci, A. Jr. (1999) Applications of forensic identity testing in the clinical laboratory, *Am. J. Clin. Pathol.* **112**((Suppl. 1)), S93–S103.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting, *Nucleic Acids Res.* **23**, 4407–4414.

Waye, J.S., Presley, L.A., Budowle, B., Shutler, G.G., Fourney, R.M. (1989) A simple and sensitive method for quantifying human genomic DNA in forensic specimen extracts, *BioTechniques* **7**, 852–855.

Weber, J.L., May, P.E. (1989) Abundant class of human polymorphisms which can be typed using the polymerase chain reaction, *Am. J. Hum. Genet.* **44**, 388–396.

Wells, J.D., Sperling, F.A.H. (2001) DNA-based identification of forensically important Chrysomyinae (Diptera: Calliphoridae), *Forensic Sci. Int.* **120**, 110–115.

Whitaker, J.P., Clayton, T.M., Urquhart, A.J., Millican, E.S., Downes, T.J., Kimpton, C.P., Gill, P. (1995) Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples, *BioTechniques* **18**, 670–677.

White, E., Hunter, J., Dubetz, C., Brost, R., Bratton, A., Edes, S., Sahota, R. (2000) Microsatellite markers for individual tree genotyping: application in forest crime prosecutions, *J. Chem. Technol. Biotechnol.* **75**, 923–926.

Wilkinson, M., Linacre, A.M.T. (2000) The detection and persistence of *Cannabis sativa* DNA on skin, *Sci. Justice* **40**, 11–14.

Wilson, M.R., Polanskey, D., Butler, J., DiZinno, J.A., Replogle, J., Budowle, B. (1995) Extraction, PCR amplification, and sequencing of mitochondrial DNA from human hair shafts, *BioTechniques* **18**, 662–669.

Zhang, D., Besse, C., Cao, M.Q., Gandelin, M.H. (2001) Evaluation of AFLPs for Variety Identification in Modern Rose (Rosa hybrida L.), in: Doré, C., Dosba, F., Baril, C. (Eds.) *Acta Hortic 546, Proceedings of the International Symposium on Molecular Markers for Characterizing Genotypes and Identifying Cultivars in Horticulture*, 2000, International Society for Horticultural Science, The Hague, the Netherlands, pp. 351–357.

DNA, Repressor–Operator Recognition

Peter G. Stockley and Simon E.V. Phillips
University of Leeds, Leeds, UK

1 Principles 508
1.1 The Operon Hypothesis 508
1.2 The Structures of Repressor Proteins 511
1.3 Repressor–Operator Complexes 513
1.4 General Principles of Repressor–Operator Interactions 515

2 Techniques 519
2.1 *In Vitro* Techniques for Studying Repressor–Operator Interactions 519
2.2 *In Vivo* Techniques for Studying Repressor–Operator Interactions 523

3 Applications 526

4 Perspectives 526

Bibliography 527
Books and Reviews 527
Primary Literature 527

Keywords

Corepressor (Inducer)
A small coeffector molecule whose presence increases (decreases) the affinity of a repressor for its operator site.

Direct Readout
A situation in which amino acid side chains of the repressor protein make direct contacts with the edges of DNA base pairs to detect their sequence.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

DNA-binding Motif
A protein structural element primarily responsible for making sequence-specific contacts with DNA (e.g. the helix-turn-helix motif or the β-ribbon).

Indirect Readout
A situation in which repressor–operator complex formation involves sequence-dependent distortions to the DNA conformation.

Transcriptional control is the most important mechanism of differential gene expression in the vast majority of organisms. Repressors control transcription by binding to specific sequences (operator sites) in genomic DNA and sequestering those regions from RNA polymerase, hence preventing transcriptional initiation. Repressor–operator binding is controlled by the physiological state of the cell, either directly by changes in the active repressor concentration or indirectly by sensing of the levels of key metabolites. Operator binding is highly specific and is achieved by a large number of distinct molecular interactions. A combination of structural studies and molecular genetics has led to rapid recent progress in understanding this field. Repressor–operator interactions provide simple tools to regulate the expression of recombinant proteins in a variety of organisms such as bacteria and yeast.

1 Principles

1.1 The Operon Hypothesis

In the 1950s, geneticists realized that bacterial cells are able to regulate the expression of genes encoding enzymes controlling particular metabolic pathways depending on their physiological state. For instance, the enzymes enabling *Escherichia coli* to metabolize lactose are not constantly present in the cell, but their synthesis can be induced by the addition of lactose to the medium, the extent of induction being dependent on the levels of glucose, the preferred energy source (Fig. 1). To investigate the mechanism behind such enzyme induction, a series of elegant genetic experiments was performed by a group based in France, led by Jacques Monod, François Jacob, and their colleagues. Their results led to the formulation of the *operon hypothesis*, which has become the basic model for transcriptional control of gene expression. In general, regulated genes (or groups of genes) contain two sequence elements at their 5′ ends: a promoter, which defines the start sequence for the bacterial RNA polymerase, and an operator, which is the target sequence of a repressor protein. Binding of repressor to the operator interferes with either binding of the RNA polymerase, or its initiation of transcription, leading to repression of the enzymes encoded by the downstream genes (Fig. 2). The genes for enzymes involved in related processes, such as lactose utilization, may be clustered and regulated by a single promoter–operator site; that is, they

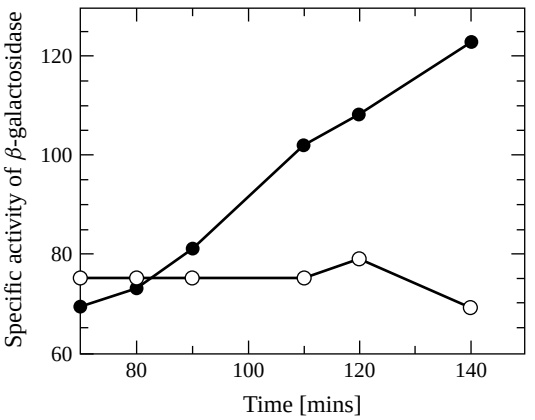

Fig. 1 Graph of the induction of β-galactosidase activity in bacterial cells in the presence (curve with solid points) and absence (open points) of the inducer, isopropylthiogalactoside (IPTG). The levels of induction of the other structural genes in the operon show identical behavior.

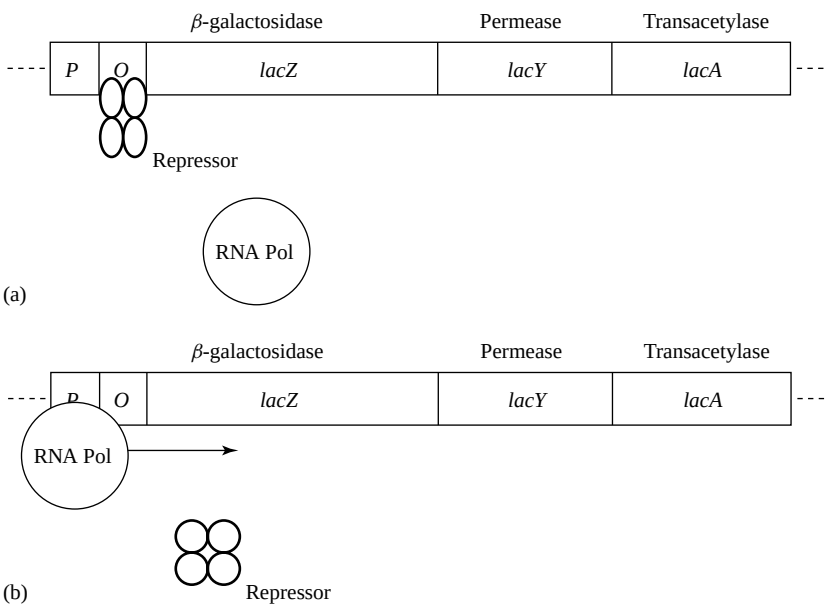

Fig. 2 Simplified diagram of the repression system in the *lac* operon. The DNA of the operon is shown on a block divided into three structural genes (coding for the enzymes required to metabolize lactose), the promoter site *P* where RNA polymerase (RNA Pol) binds, and the operator site *O*. Dotted lines indicate DNA flanking the operon. (a) The uninduced state in the absence of lactose (or allolactose). The repressor binds tightly to the operator, impeding access of RNA polymerase. (b) The induced state, where allolactose is present and binds to the repressor causing a change in its structure. As a result, the repressor can no longer bind to the operator, allowing RNA polymerase access to the promoter.

constitute an operon. It is also possible to have several separate promoter–operator sites regulated by the same repressor, a grouping that has been termed a *regulon*.

Repression often occurs in response to an environmental stimulus, such as the presence of the sugar lactose, by the binding of small coeffector molecules to the repressor proteins. Such coeffectors can increase or decrease the affinity of the repressor for its operator site; that is, they can act as either corepressors or inducers of gene expression. In the case of the *lac* operon, the presence of lactose in the growth medium leads to the production of a breakdown product, allolactose, which binds to the LacR repressor protein. This causes an allosteric conformational change in LacR, preventing the protein from binding to DNA. In the absence of allolactose, the repressor binds tightly to the *lac* operator, preventing transcription initiation by RNA polymerase. In the presence of allolactose, the repressor dissociates from the DNA, allowing transcription to begin, and mRNAs for the *lac* operon structural genes are produced. These encode enzymes for lactose metabolism such as lactose permease, which actively sequesters the available lactose into the cell and β-galactosidase, which breaks down the lactose into its component parts (Fig. 3).

In bacteriophage lambda, repression controls the switch between lytic and lysogenic growth, and this phage has

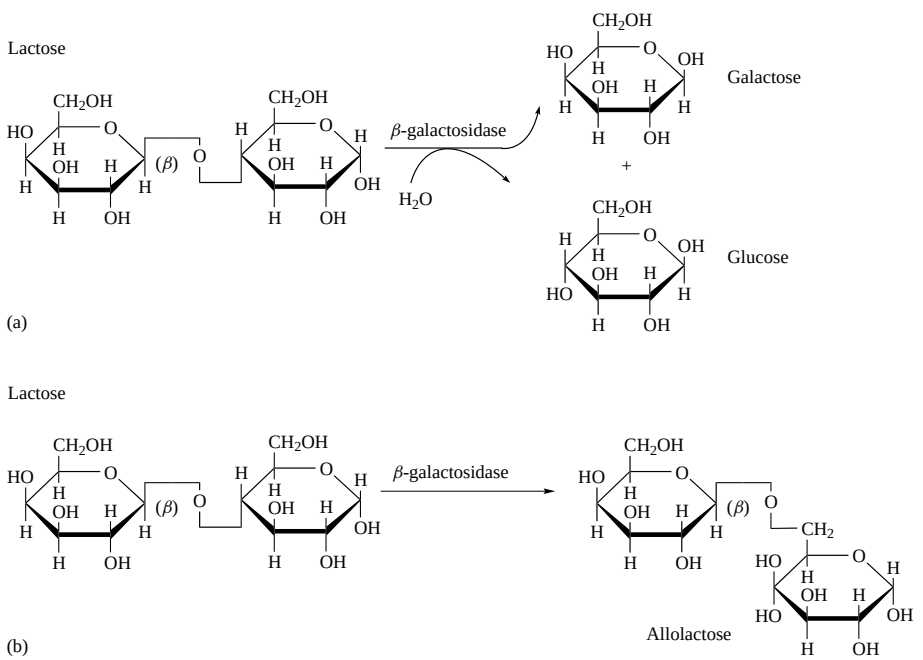

Fig. 3 The two reactions catalyzed by β-galactosidase. (a) Cleavage of the disaccharide lactose yields the hexoses galactose and glucose, which are both readily metabolized to provide energy for the cell. (b) The minor isomerization reaction produces allolactose, which acts as an inducer, binding to LacR and preventing it from binding to its operator site. Eventually, in the absence of lactose, the allolactose concentration falls, restoring the repressed state.

been one of the major model systems for studying the control of gene expression.

1.2 The Structures of Repressor Proteins

In the early 1980s, a number of three-dimensional structures of repressor proteins were determined by X-ray diffraction techniques. The first three proteins for which structures became available were phage lambda Cro, the *E. coli* cyclic AMP receptor protein (CRP), and lambda repressor (cI). Comparison of the polypeptide chain topology in these cases shows that although each protein has a unique overall fold, they all share a particular structural element, consisting of two α-helices linked by a rather unusual β-turn (Fig. 4). This helix-turn-helix (hth) motif is located on the surface of the protein in each case. All three proteins are symmetric dimers, and the two copies of the motif, one per subunit, are positioned such that the distance between them is approximately 34 Å. This is also the separation of the successive turns of the major groove of B-form DNA along one face of the DNA duplex. The observation immediately suggested that the hth was the DNA-binding site, a proposal subsequently confirmed by the so-called helix swap experiment, in which the amino acid residues exposed on the surfaces of the second helices from two phage repressors, 434 (λ) cI and P22 cI were swapped using site-specific mutagenesis. The resulting hybrid proteins acquired the operator binding properties of the newly inserted sequences; that is, the hybrid λcI protein bound to P22 operators and vice versa (Fig. 5). As a result of such experiments, the second helix of the hth motif became known as the recognition helix, implying that

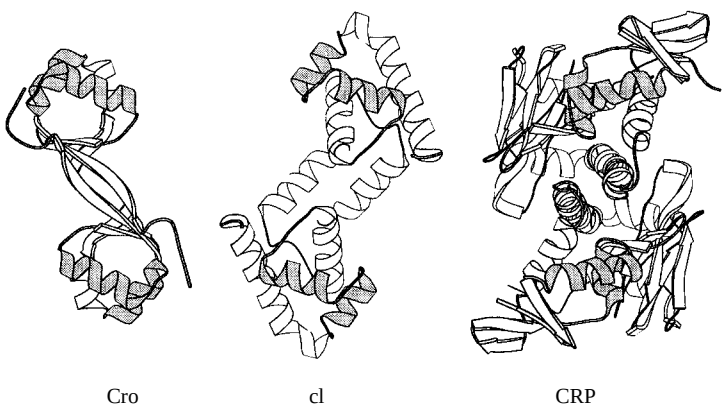

Fig. 4 Structural cartoons of three proteins, the bacteriophage lambda repressors Cro and cI repressors, and *E. coli* CRP, containing the conserved helix-turn-helix motif. (In the case of λcI, only the DNA-binding domain is shown.) Although the three structures are obviously quite different, pairs of hth motifs (shaded) are found in all of them, with a spacing corresponding to that of successive turns of the major groove of B-DNA. In each case, a DNA duplex bound to the protein would lie with its helical axis approximately vertical. In this and subsequent structural figures, coils represent α-helices and ribbons with arrowheads, β-strands. (Drawn with computer program MOLSCRIPT: Kraulis, P.J. (1991) *J. Appl. Crystallogr.* **24**, 946.)

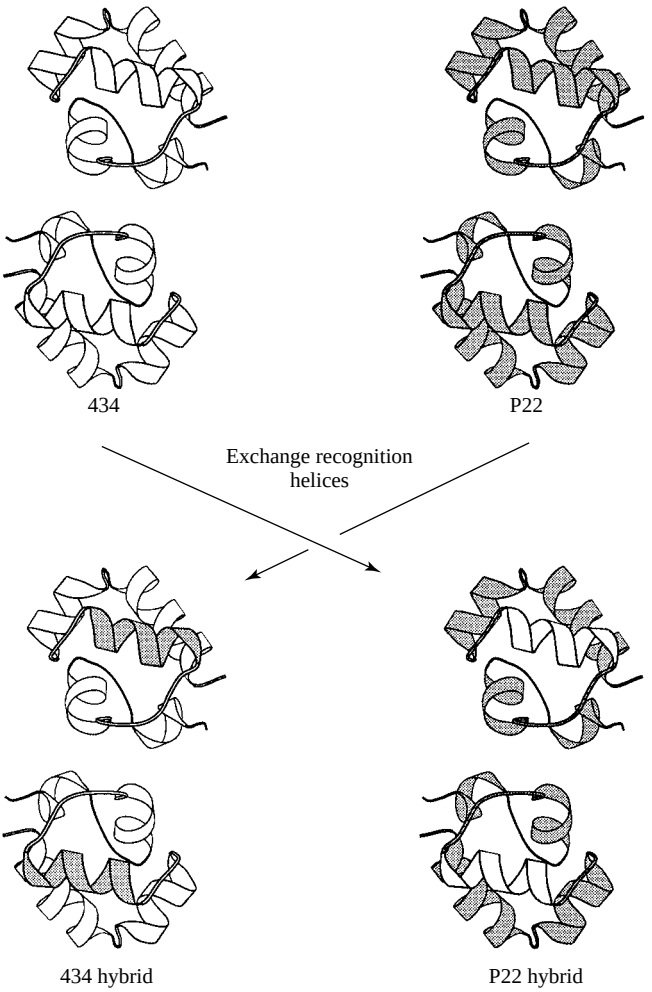

Fig. 5 The helix swap experiment. Recognition helices are exchanged between repressors 434 (open) and P22 (shaded), which have almost identical structures. After the swap, each repressor loses the ability to recognize its own operator, but will bind to that of the other repressor.

all operator recognition resided in it. As the three-dimensional structures of many more repressor–operator complexes have become available, this idea has proved to be an oversimplification (see Sect. 1.4).

The initial observation of a structural feature conserved between three repressor proteins was unexpected because there is very little amino acid sequence similarity between them. However, the structural studies allowed the identification of key residues within the hth motif (e.g. the presence of glycine in the turn). It was then possible to construct an algorithm to search the protein sequence database for potential matches with the hth motif.

The result was an impressive list of several hundred proteins, all believed to bind DNA, having the required sequence features of the hth motif. It is now clear that the hth motif defines a large family of proteins from both prokaryotic and eukaryotic cells sharing a conserved DNA-binding motif.

Although extremely common, the hth is not the only protein structural motif capable of binding to DNA in a sequence-specific fashion. Early model-building studies had suggested that both α-helices and antiparallel two-stranded β-ribbons have the structural features necessary to interact with duplex DNA. Two distinct families of proteins having β-ribbon DNA-binding motifs have now also been characterized. These are based on the *Bacillus stearothermophilus* HU protein, which is a nonsequence-specific DNA-binding protein, and the *E. coli* MetJ protein, the repressor of methionine biosynthesis. A number of other DNA-binding motifs have been identified in proteins regulating transcription in eukaryotic cells. It is likely that other families having novel DNA-binding motifs will be characterized in the near future.

1.3 Repressor–Operator Complexes

Operator sites generally contain palindromic sequences allowing dimeric repressor proteins to align with their twofold axes along the sequence dyad, thus generating two symmetrical sets of protein–DNA interactions. Repressors consisting of higher multimers such as LacR, which is a tetramer, or ArgR, the *E. coli* arginine repressor, which is a hexamer, nevertheless seem to be organized into pairs of subunits and retain the symmetry of the interaction with their operator sites. Figure 6 shows the structures of two repressors, one each from the hth and β-ribbon families. In both cases, the repressor–operator complex is formed by insertion of the DNA-binding motif into the major groove of the DNA duplex. A large number of specific intermolecular contacts are made between the DNA and the repressors, including hydrogen bonds and electrostatic interactions.

Both the TrpR (the *E. coli* tryptophan repressor) and MetJ repressors regulate the flux through biosynthetic pathways and must therefore alter the levels of repression at their various operons in response to the physiological state of the cell. This is achieved by the binding of two molecules of L-tryptophan to TrpR, and of *S*-adenosylmethionine (AdoMet), a derivative of L-methionine, to MetJ. The resultant repressor–corepressor complexes (holorepressors) have much higher affinities for operator DNA than the corresponding free repressors (aporepressors). However, the molecular mechanisms of these increases in affinity are quite distinct. In apoTrpR, the hth DNA-binding motifs are rather flexible and tend to lie too close together to fit easily into the adjoining major grooves of B-form DNA. Binding of L-tryptophan results in a concerted conformational change and general stiffening of the repressor dimer that stabilizes a structure in which the spacing of the hth motifs is more complementary to that of the DNA major groove (Fig. 7). MetJ, on the other hand, does not appear to alter its conformation significantly when it binds its corepressor. Indeed, AdoMet lies in a pocket on the opposite side of the molecule from the one that binds to DNA. In this case, it appears that the increase in affinity upon

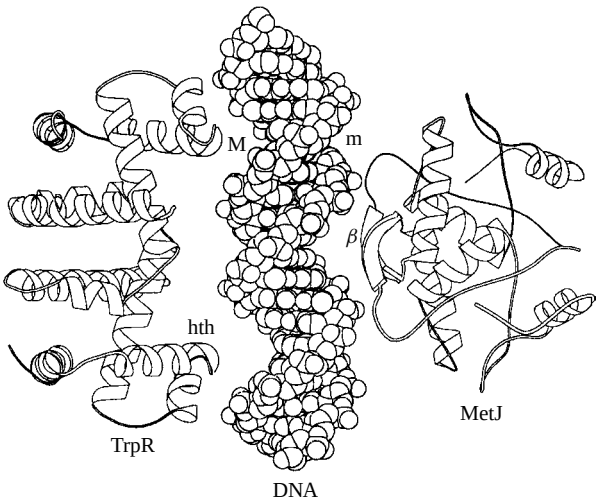

Fig. 6 Two turns of B-DNA flanked by structural representations of the active forms of left, the *E. coli* TrpR (tryptophan); right, MetJ (methionine) repressors. Both repressors are oriented correctly for DNA binding but have been pulled away slightly. As TrpR approaches the DNA, its two hth motifs can dock into adjacent turns of the major groove (M). As MetJ approaches, its β-ribbon (β) can dock into the central portion of the major groove. Corepressor molecules have been omitted for clarity.

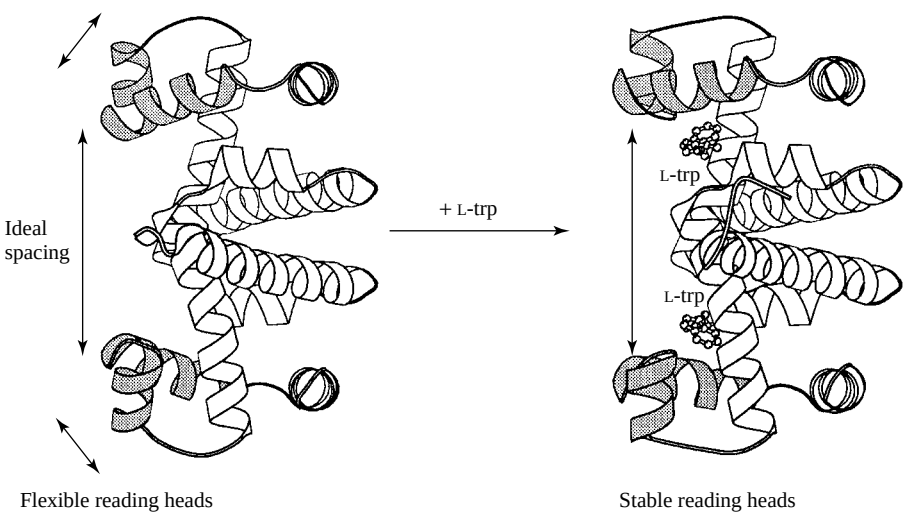

Fig. 7 Activation of TrpR repressor by L-tryptophan binding. In the absence of L-tryptophan, the hth reading heads are flexible and can collapse toward the middle of the repressor. When L-tryptophan is bound, the structure is stabilized, with the reading heads ideally spaced for DNA binding.

corepressor binding is due to a long-range electrostatic effect. Each AdoMet molecule carries a net positive charge, which helps to neutralize the negative charge on the DNA phosphate backbone, thus raising the affinity of the repressor for its operator. The differences in the molecular mechanisms of the corepressor effects in these two cases show that more than one mechanism has evolved to regulate repressor affinity for operator sites, and others may yet be discovered.

1.4 General Principles of Repressor–Operator Interactions

A bacterium such as *E. coli* has a double-stranded DNA genome consisting of more than four million base pairs, which at any one time will be partially replicated. Repressor proteins therefore must locate a few specific operator site(s) among a background of a large number of potential nonspecific sites elsewhere on the DNA. Indeed, since duplex DNA has a periodic structure, for any defined target size, translation along the DNA by just one base pair creates a new potential operator site. Target sites that occur once per genome are hidden among millions of nonspecific sites. To generate a target operator site that is unique in the genome, the repressor must be able to recognize a specific sequence of reasonable length. For instance, there is a 1 : 256 probability that a random 4 bp sequence will match a defined unique sequence by chance (i.e. there will be >10 000 such matches in the *E. coli* genome). However, the chances of a random match fall to roughly one in a million for sequences 10 bp long. Most operator sites characterized to date are between 12 to 20 bp in length and thus likely to be unique within bacterial genomic DNAs.

The sequence specificity of the repressor–operator interaction appears to be achieved in at least two distinct ways. The first relies on the fact that not all the functional groups on the bases are directly involved in Watson–Crick pairing, but are exposed in the major and minor grooves. Modeling studies suggested that individual base pairs could be recognized by formation of a series of hydrogen-bond contacts from amino acid side chains to these functional groups. Comparison of the two kinds of base pairs showed that discrimination would be more effective in the major groove because of the number and diversity of the contacts that could be made (Fig. 8). Such interactions have been observed experimentally in many repressor–operator complexes. This type of interaction has now been termed "direct readout," since the protein makes direct contacts with the edges of operator base pairs (Fig. 9). The modeling studies had suggested that bidentate side chains, such as glutamine and arginine, could make specific contacts to particular base pairs in 1 : 1 interactions, giving a simple protein-DNA recognition code. In the observed structures of repressor–operator complexes, however, there are more complex interactions in which side chains often contact more than one base pair. Thus, there is no simple code relating amino acid to base sequences. Rather, complexes contain a constellation of interacting side chains, main chain atoms, and nucleotide bases with many contacts contributing to the overall specificity.

A second type of interaction, termed "indirect readout," has also been identified: when complex formation involves sequence-dependent conformational distortion of DNA, increased intermolecular

complementarity (or interaction) results. An example of this type of interaction is shown in Fig. 10 for MetJ. The MetJ repressor is a dimer, and two of these bind to the minimum 16 bp operator site, which consists of two 8 bp sequence repeats (dA-GACGTCTAGACGTCT). Each repressor interacts, via direct readout, with the base

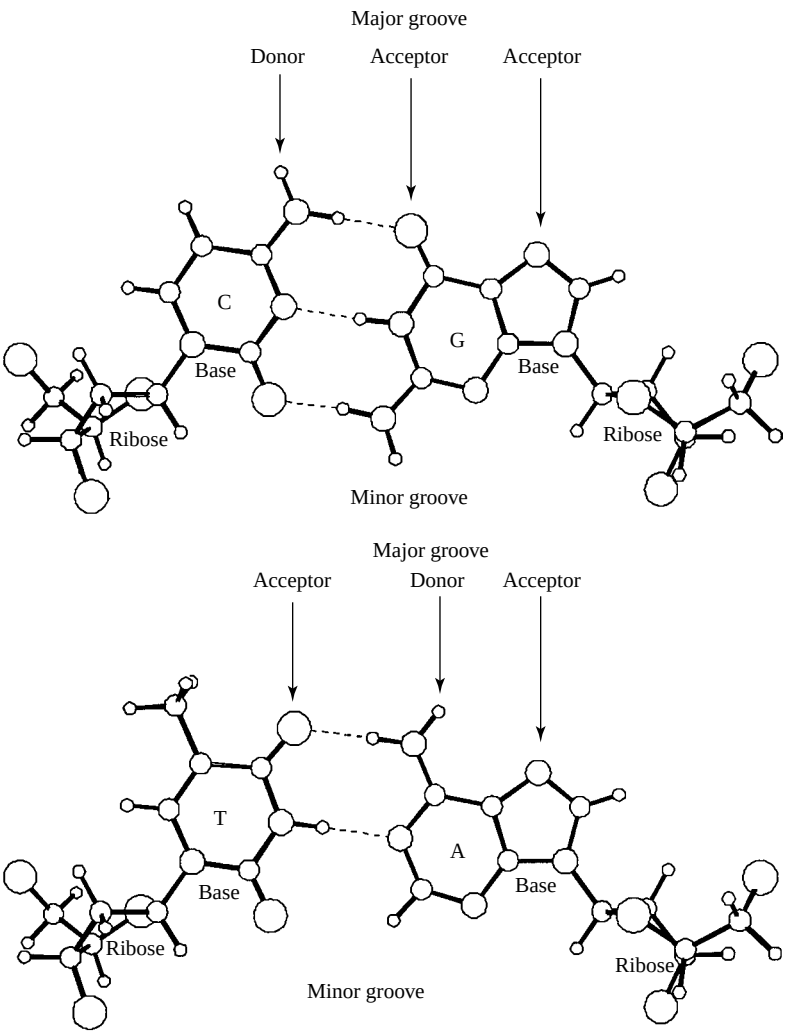

Fig. 8 The two possible kinds of base pair viewed along the DNA helix axis. The smallest circles represent hydrogen; the largest, oxygen; and the second largest, nitrogen atoms. The remaining circles represent carbon atoms. The upper edges of the bases are exposed in the major groove, and each has a unique pattern of hydrogen-bond donors and acceptors. Each base pair can be reversed, of course, leading to two further different patterns when exact positioning is taken into account. The sugar-phosphate backbone links adjacent base pairs through the outer edges of the riboses.

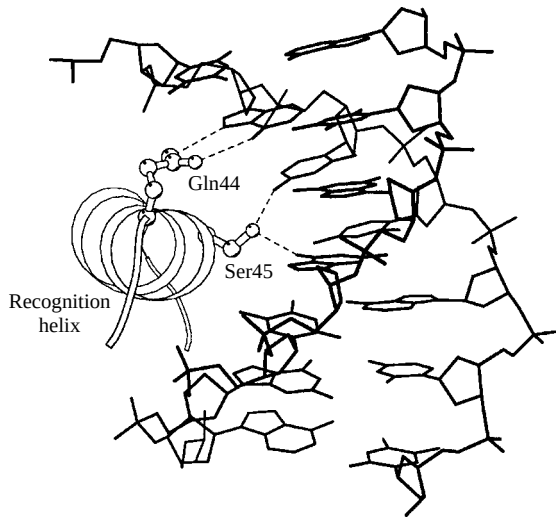

Fig. 9 The recognition helix of λcI repressor lying in the major groove of its operator. Of the 10 amino acid residues in the helix, the side chains are shown as balls and sticks only for the two that make direct hydrogen-bond contacts to the bases (dotted lines). Gln44 has a hydrogen-bond donor and acceptor, allowing it to make two bonds to the acceptor and donor present on the A base (see Fig. 8).

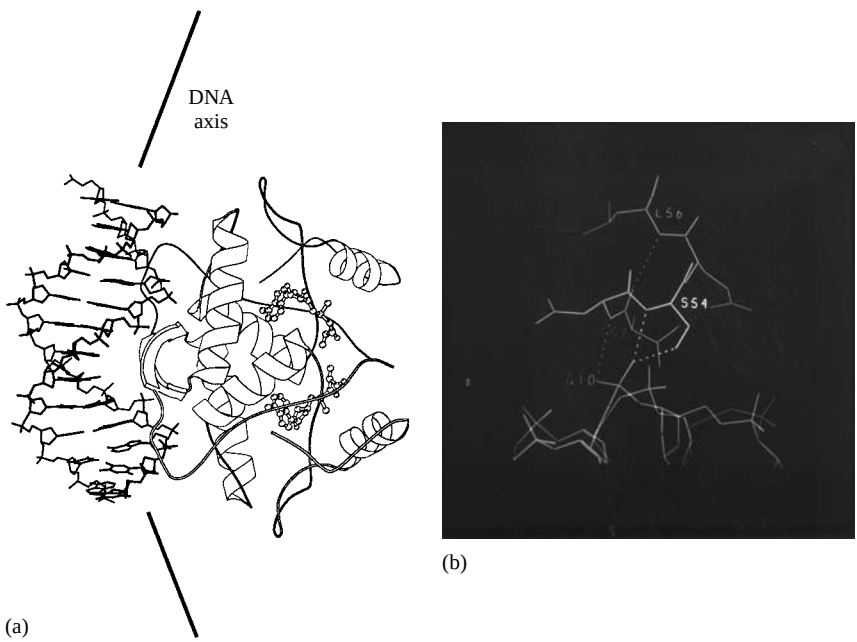

Fig. 10 (a) Bending of DNA when bound to MetJ repressor, allowing it to follow the protein surface more closely. (b) Close-up view of DNA distortion in the MetJ repressor–operator complex. One end of an α-helix in the repressor (blue) makes hydrogen-bond contacts (dotted lines) to a phosphate group in the operator (green). If the DNA were in the undistorted B-conformation (red), the phosphate would be out of range of this interaction. [Redrawn from *Nature*, **359**, 287–393 (1992).] (See color plate p. xxxv.)

pairs within each operator half-site. The half-site is bent toward the protein, wrapping snugly along its surface, thus increasing the structural complementarity of the components of the complex (Fig. 10(a)). However, the overall repressor–operator interaction is also sensitive to the sequence at the junction between the two half-sites, 5′-CT·AG-3′. The T.A base step at the junction is easily distorted away from the B-form conformation because of the relatively poor stacking energy of pyrimidine bases on purines, and especially T on A. In the MetJ–operator complex, this step is overwound, resulting in displacement of the phosphate 5′ to the T.A step from its expected B-DNA position by about 1.5 Å (Fig. 10(b): see color plate p. xxxv). The repressor can then make a number of hydrogen bonds to this phosphate, which it could not do with the phosphate in the standard B-DNA conformation.

The problem remains of how repressor proteins locate their specific operators in the presence of the large background of competing sequences against which they must discriminate. Essentially all sequence-specific DNA-binding proteins will also bind to noncognate DNA sequences, but with much lower affinity. Thus, in the cell, most repressors would be expected to be bound either specifically or nonspecifically to the genomic DNA molecule. The interactions in nonspecific complexes appear to be mostly electrostatic, the protein interacting primarily with the negatively charged phosphodiester backbone. Hydrogen bonds to base pair edges may occasionally also be made, together with contacts owing to DNA conformational distortion. Only at the true operator site, however, will the entire constellation of contacts be "correct." At other sites, in general, the interactions will never be strong enough to prevent the repressor from dissociating. It is notable that many repressors that have been characterized bind to operator sites that are slightly suboptimal in terms of binding affinity (i.e. in the number of intermolecular contacts that can be made). Presumably this is because the cell requires rapid dissociation of the less tightly bound repressor–operator complexes when environmental conditions call for derepression. Clearly, there must be a sliding scale of affinities between very poor (noncognate sites) and high affinity (but not necessarily perfect) operators.

Recognition of the ability of repressors to bind nonspecific DNA sequences and the problem of locating a single unique sequence among millions of competing sites raises the question of the time taken by the repressor to locate its operator. Experimental measurements of the kinetics of repressor–operator complex formation (see Sect. 2.1) show that the process is extremely rapid despite these difficulties. It appears that repressors can use one or both of two approaches to solve the search problem. In each case, the mechanism relies on the repressor maintaining contact with the DNA during the movement from one potential operator site to another. The first is termed "sliding" and involves lateral movement of the repressor along the DNA. At each interval (1 bp) during the slide, the repressor is, therefore, able to check a new sequence for a match to its operator. The second mechanism is termed "intersegment transfer" and involves the transient binding of two different segments of DNA duplex by one repressor molecule. The repressor can then "hop" between different regions of the DNA molecule. In reality, it appears that a combination of these two mechanisms may be used, dramatically increasing the speed of operator site location (Fig. 11).

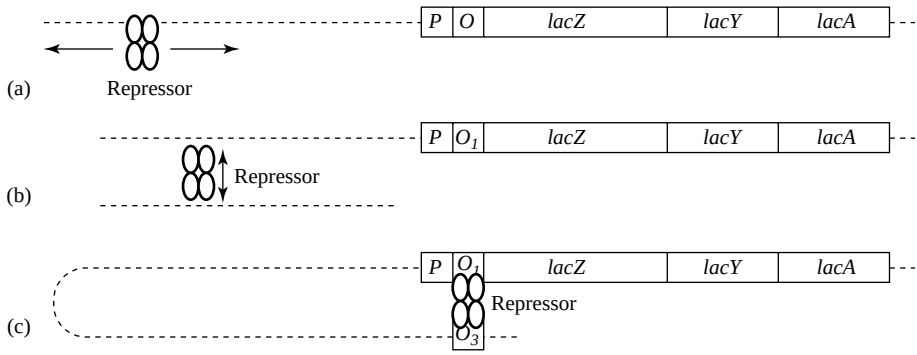

Fig. 11 Cartoon illustrating facilitated diffusion. (a) Lac repressor bound nonsequence-specifically to DNA can search for its operator site by sliding along the DNA in either horizontal direction (arrows). (b) Lac repressor can "jump" between sections of the bacterial chromosome via an intersegment transfer in which transiently two DNA duplexes are bound. The repressor can remain bound to either duplex when the complex dissociates, as indicated by the arrows. (c) The doubly bound complex is functionally important for repression in the *lac* system. Both DNA segments bound are operator sites, and the result is the formation of a loop of DNA between them.

If the two sections of DNA duplex bound in an "intersegment transfer" are part of the same molecule, a loop of DNA is created between the contacted sequences. Oligomeric repressors with more than two subunits appear to use such loops to help achieve their physiological effect. This has been demonstrated most clearly for the tetrameric LacR protein. Analysis of the sequences around the 5′ promoter region for the *lac* operon identified a major (high affinity) LacR binding site (primary operator), which is contacted by two hth motifs from two of the four protein subunits in the repressor. There are also two minor (lower affinity) sites, one further upstream in the nontranscribed sequences and the second downstream from the major operator and actually within the coding region of the first gene of the operon. Either of these secondary sites can also be contacted by the hth motifs of the two remaining LacR subunits forming a loop between the major and minor binding sites. Several possible loops can thus be made, since the repressor can bind only two of the three sites at once. *In vivo*, all the possible loops do form roughly in proportion to the overall affinities of their respective repressor–DNA complexes, and each contributes to the overall level of repression observed.

2 Techniques

The experimental approaches used to study repressor–operator interactions can be distinguished by whether they are *in vitro* or *in vivo* techniques.

2.1 *In Vitro* Techniques for Studying Repressor–Operator Interactions

Repressors have many advantages for *in vitro* studies. For instance, the very high specificity of their interaction with target

operators allows experiments to be carried out with unfractionated protein extracts as well as with purified samples. Normally each bacterial cell would contain at most a few hundred repressor molecules, so studies are most often carried out using recombinant proteins expressed to high levels to facilitate purification. This in turn has led to a dramatic increase in our understanding of the structures and mechanisms of action of these proteins. The techniques of X-ray diffraction and nuclear magnetic resonance (NMR) spectroscopy have been used to determine the three-dimensional structures of repressors and their complexes (see Sect. 1.2). Both techniques require milligram quantities of essentially homogeneous protein or protein–DNA complex. The results make the effort required to produce such samples well worthwhile, allowing a detailed molecular understanding of phenomena such as sequence-specific DNA binding. Although such structural studies are enormously useful in identifying regions of the components that are interacting in the complexes (e.g. the hth motif), they do not provide any information on the interaction energy. It is therefore important to correlate the structural studies with functional experiments aimed at determining the importance of particular interactions to the overall free energy of complex formation. This can be done by varying the nucleotide sequences of test operator fragments and/or by using repressor protein mutants.

The affinities for operator sites are often in the nanomolar range. Thus, experiments require only tiny amounts of protein and usually are performed in conditions of protein excess over operator-containing DNA fragments, which are usually radioactively labeled. One of the commonest techniques used to monitor complex formation is the gel-retardation assay, which separates repressor–operator complexes from unliganded operator DNA fragments by means of electrophoresis in polyacrylamide gels. The repressor–operator complexes are retarded relative to protein-free DNA because of changes in the size/shape/charge of the molecule (Fig. 12). The amount of complex present can then be determined

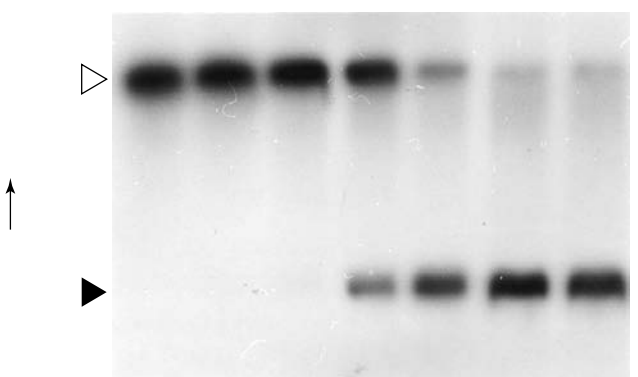

Fig. 12 A photograph of an autoradiograph from a gel retardation assay involving the *met* repressor; arrow indicates direction of electrophoresis. The positions of protein-free operator DNA and the repressor–operator complex are marked by open and solid arrowheads, respectively. Repressor concentrations decrease from left to right.

by autoradiography if radioactively labeled DNA is used. The assay has the advantage of revealing the presence of multiple protein–DNA complexes if they form, but retarded species are observed only if the half-life of the complex is greater than the time required to run the sample into gel (usually ~60 s). Under conditions of protein concentration excess over the DNA target fragment, the concentration of repressor producing 50% retardation of DNA is approximately equal to the equilibrium dissociation constant K'_d allowing quantitative comparisons of repressor–operator affinity to be made.

An alternative method involves fractionation of the repressor–operator complexes by filtration through nitrocellulose, which binds tightly to proteins and their DNA complexes but allows protein-free DNA to pass through. The filtration step is very fast, so problems due to rapid dissociation of the complexes are avoided. It is not possible, however, to determine the relative amounts of multiply-bound DNA fragments, and there are problems associated with the differential retention on the filters of complexes with different stoichiometries.

Although useful for comparative quantitation of repressor–operator interactions, the *in vitro* assays provide no information on the precise sequence of the operator that is being recognized. A series of procedures is available to determine the positions of tight interaction between repressors and the components of the DNA duplex. Some of these techniques can be used in a footprinting mode (i.e. applied to a preexisting complex); others must be used in an interference mode. In the latter, the DNA site is partially modified by a specific reagent before complex formation, and the molecules still capable of being bound by repressor are separated from those in which the modification has interfered with complex formation. In both cases, the experiments are usually performed with DNA radioactively labeled at one end, and the products are analyzed on sequencing gels next to a sequence ladder of the parent fragment for ease of identifying the positions of modification (Fig. 13). Examples of footprinting reagents are DNase 1, which makes single-stranded cuts in the backbone of the target DNA; dimethyl sulfate (DMS), which methylates the N-7 position of guanine in the major groove; hydroxyl radicals, which attack the sugar residues along the duplex; and methidium propyl EDTA, which intercalates between bases before cleaving via a hydroxyl radical mechanism. Binding-interference experiments are often performed with ethyl nitrosourea, which ethylates the DNA phosphate groups, or with hydroxyl radicals or dimethyl sulfate.

The kinetics of operator location can be studied by means of both the nitrocellulose and gel retardation assays. The rate of association or dissociation is monitored, while the size of the target fragment, the salt concentration of the solution, or the overall concentration of DNA is varied. Rates of sliding are affected by the first two variables but will not be affected by DNA concentration. Intersegment transfer, however, is affected by DNA concentration because two duplexes must be bound simultaneously. Care must be taken to characterize the complexes being formed, since quantitative analysis of the binding data is possible only for well-characterized complexes.

Recently a new technique was introduced to measure the kinetics of biomolecular interactions directly, using

the phenomenon of surface plasmon resonance (SPR). The physical principles of this technique are beyond the scope of this article. There are, however, commercial biosensor systems available in which SPR generates a real-time electronic signal proportional to the refractive index of a buffer flowing over the face of a sensor chip to which are immobilized target molecules (e.g. DNA operator fragments) for complex formation. Potential ligands such as repressors are then introduced to the buffer flowing across this surface. When molecules bind to the immobilized target, the resultant refractive index changes give rise to a change in the SPR signal, which can be analyzed to yield a value

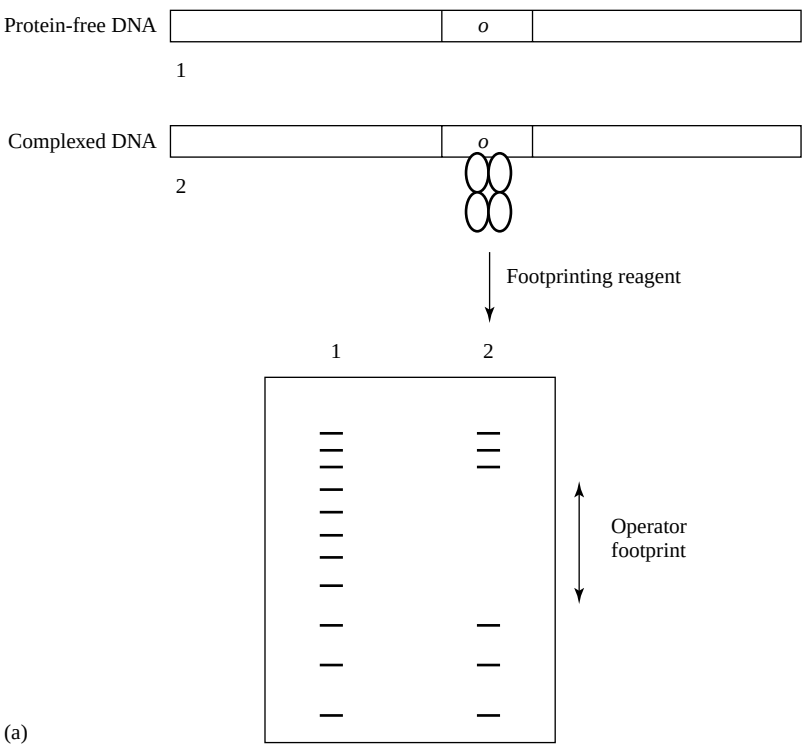

Fig. 13 (a) Footprinting with a reagent that cleaves the DNA at every nucleotide position. The DNA fragment is uniquely radiolabeled at one end, and the resultant DNA fragments are analyzed on a sequencing gel. Formation of the repressor–operator complex protects the bound region from cleavage and results in a blank section on the sequencing gel, which shows the sizes of all the cleaved fragments with respect to the radiolabel at one end.
(b) Modification/binding-interference assay. Radiolabeled, protein-free operator DNAs are modified at base, sugar, or phosphate positions at a level producing approximately one modification per molecule. These DNAs are then used in a gel retardation assay (see Fig. 12) and the bound and free DNAs recovered, cleaved at the sites of modification, and analyzed on a sequencing gel as in (a). In this assay, the blank section of the gel indicates the sites at which modification prevented formation of the repressor–operator complex.

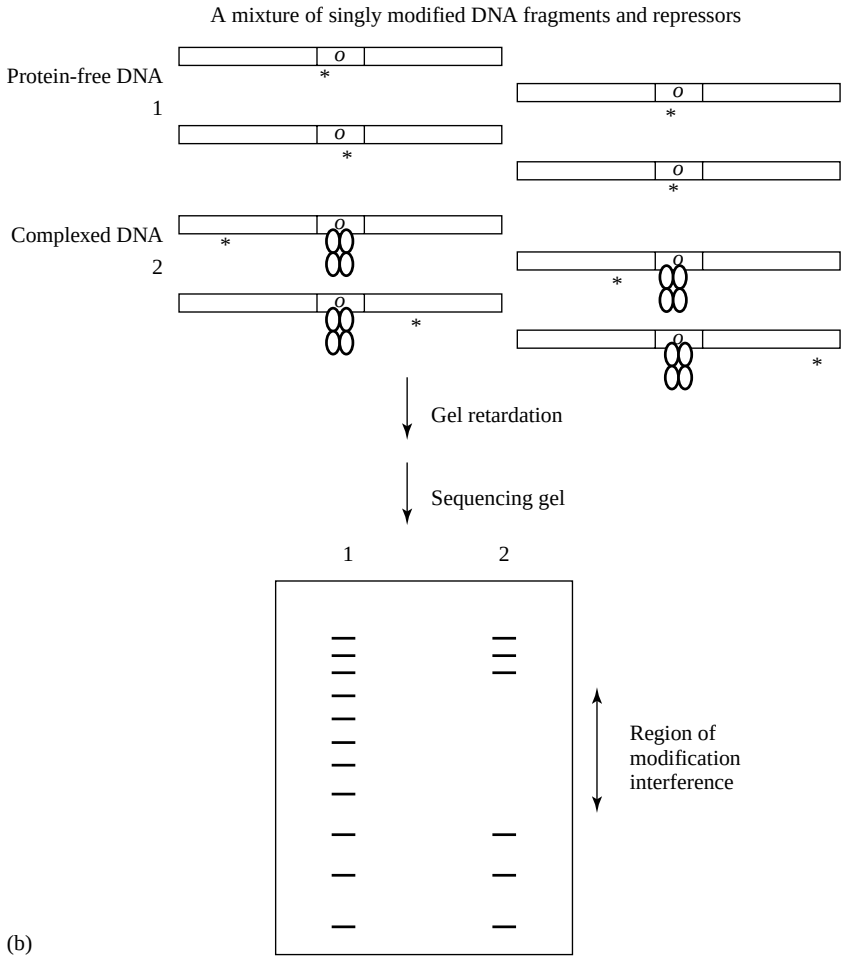

(b)

Fig. 13 (*Continued*).

for the association rate constant. The reverse occurs during removal of the bound ligands, giving information on the dissociation rate constant. This technique has the advantage of not involving separation of equilibrium mixtures, not requiring radioactively labeled molecules, and working well with concentrations of macromolecules in the physiological range. Its applications within the field of repressor–operator recognition should grow rapidly in the next few years.

2.2 *In Vivo* Techniques for Studying Repressor–Operator Interactions

Perhaps the most common technique for monitoring repression *in vivo* involves coupling a promoter–operator site to a reporter gene whose product is easily monitored (e.g. β-galactosidase). The levels of repression at a particular site can then be determined by enzymatic assay after lysis of the cells. Techniques have also been

developed to footprint repressor–operator complexes *in vivo*, using reagents (e.g. DMS) that can penetrate living cells. A major problem with both approaches is controlling the concentrations of repressor and operator sites, since they, or their genes, are usually carried on plasmids of variable copy number.

A dramatically new approach to understanding repressor–operator interactions *in vivo* is the use of arrays of immobilized gene fragments to determine the effects of mutational changes on the levels of mRNA transcripts. The arrays can be of whole genes or short oligonucleotide fragments and they can encompass all known open reading frames in an organism. For model organisms whose genomes have been sequenced, these transcriptome analyses allow us to investigate the interconnections between metabolic pathways. An example of this type of analysis for the MetJ repressor is shown in Fig. 14. Two bacterial strains are used; a wild-type

(a)

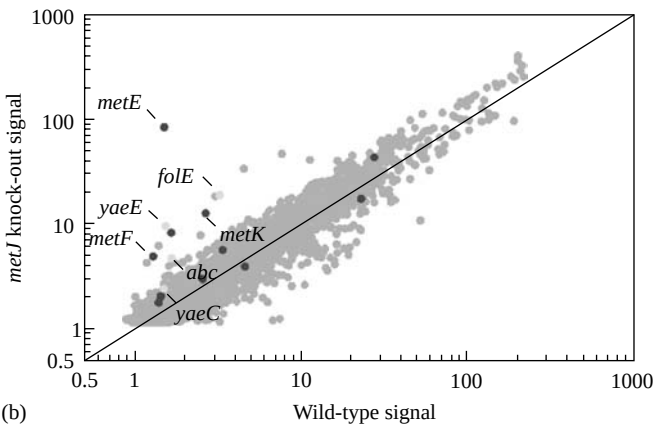

(b)

Fig. 14 (a) An example of a hybridized slide showing multiple fluorescent dyes. Columns of spots represent multiple copies of the same target oligonucleotide. Targets whose mRNAs are expressed at different levels in the mutant of interest appear nonblue. (b) Example of a scatterplot of *metJ* knockout versus wild-type (in the presence of methionine). Blue indicates points corresponding to *met* genes, whereas green corresponds to the *folE* and the methionine transport genes, *abc*, *yaeE*, and *yaeC*. The positions of a selection of individual genes are also shown. (c) A schematic showing the steps in the methionine biosynthetic pathway of *E. coli*. The arrows and their sizes show the genes that are regulated above the background as judged from the oligonucleotide hybridization experiment (see color plate p. xxxvi, xxxvii).

reference strain and one containing a mutation within a gene of interest, in this case *metJ*. Cells are grown to a fixed optical density and then total mRNAs are extracted from lysed cells. RNA from each strain is then converted to DNA via reverse transcription using random DNA primers carrying specific fluorophores. By using different fluorophores for the RNAs from each strain, the relative amounts of every mRNA in the cell can be assessed by hybridizing both sets of labelled DNA simultaneously against the array of genomic fragments (Fig. 14(a)). A fluorescence scan of the immobilized, hybridized array can then be used to determine the relative amounts of each mRNA from control and mutant strains. In practice, the amount of each fluorescent dye bound to each point in the array is plotted along an intensity axis such that mRNAs whose levels are identical in the two strains should lie along a diagonal. Genes that are overexpressed relative to the control show as points above this line, while those that are underexpressed are below it (Fig. 14(b); Table 1).

For the *E. coli* methionine repressor, we would expect that mutation of the *metJ* gene, eliminating the repressor protein from the cell, would cause permanent depression of all the MetJ-regulated genes, many of which had been mapped by classical genetic techniques. The operator sites of these various genes are scattered throughout the *E. coli* chromosome; that is, *met* is an example of a regulon. The sequences of the natural operators also vary significantly from the consensus sequence discussed above. The transcriptome analysis shows that although regulated by the same repressor protein, the genes of the methionine biosynthesis pathway are controlled to very different degrees by this system. Measurement of the repressor affinity for these natural operators using

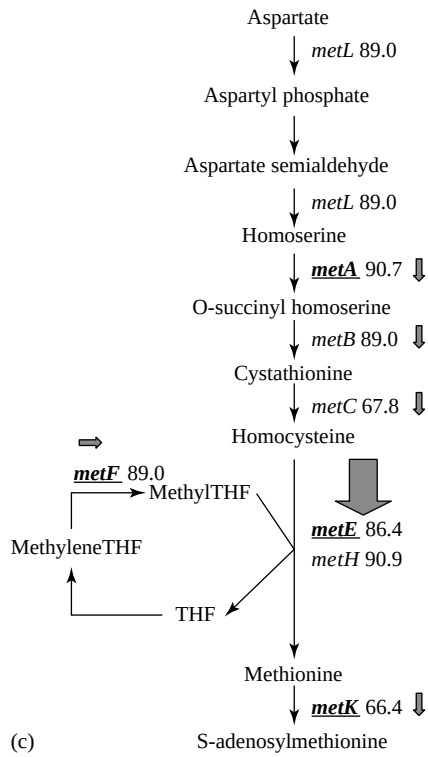

Fig. 14 (*Continued*).

techniques described above suggests that despite their sequence differences *in vitro* they all have roughly similar affinities. These results suggest that it is the rate of initiation by the RNA polymerase that determines the levels of repression observed. It is very noticeable also that the strongest repression levels occur toward the end of the *met* pathway (Fig. 14(c)) where most of the intermediates would be turned into methionine or AdoMet. Reduced repression ratios elsewhere in the pathway might have deleterious effects on the use of earlier intermediates in other biosynthetic pathways that would still need to operate when *met* was being repressed.

As well as the classical *met* genes, these studies identify a series of other genes that are clearly regulated by the repressor. One

Tab. 1 Gene expression ratios in the *metJ* knockout strain versus wild-type as determined using gene arrays. The column on the right provides the averages of three independent experiments.

Gene	*Repression ratios*			
	Exp. 1	*Exp. 2*	*Exp. 3*	*Avg.*
metE	38.4	34.6	40.0	37.7
metF	6.6	8.2	8.4	7.8
metK	4.0	5.0	4.8	4.6
metA	3.9	3.5	4.1	3.8
metB	2.7	2.6	3.3	2.8
metC	2.0	2.8	2.8	2.6
metL	2.2	3.0	2.2	2.5
metR	1.3	1.6	1.3	1.4
metH	1.1	1.5	1.4	1.3
metG	0.7	1.1	1.0	0.9
metJ	0.4	0.6	0.5	0.5
folE	2.9	4.2	6.8	4.6

of these is *folE*, which controls the biosynthesis of the folate cofactor of the enzymes providing the methyl group that is added to homocysteine. Other regulated genes are involved in transporting methionine into the cell, implying that the levels of free methionine are tightly regulated and that only under conditions of relative starvation would the cell begin to transport methionine from the surrounding medium. Such studies will clearly offer unique insights into the roles of repressor–operator interactions in gene regulation in future.

3 Applications

The primary applications of repressor–operator interactions are in the regulated expression of recombinant proteins in bacteria. The strength of the promoter is also an important consideration. A particularly strong promoter that is used for such experiments is *tac*, which is a hybrid between the *trp* and the *lac-uv5* promoters and is regulated by the LacR repressor. Expression of the regulated gene is induced by the addition of nonhydrolyzable analogs of lactose such as isopropylthiogalactoside (IPTG) to the medium. In this system, the expression occurs at a low level even in the uninduced state, since the promoter is very strong, and if the protein product is poisonous to the host cell, the effect can be deleterious. A common alternative system involves the regulation of the lambda promoter–operator site by a temperature-sensitive mutant λcI repressor. Shifting the bacteria to the nonpermissive temperature results in rapid derepression of the regulated gene.

4 Perspectives

The wealth of information that exists about repressor–operator complexes is being used as a design database for the synthesis of novel reagents capable of binding to the DNA sequence-specifically. There may well be applications for these reagents in the therapeutic control of gene expression, for example, in the treatment of cancers or retroviral diseases such as AIDS. Alternatively, chemically modified versions of several of the repressors described here have been used to achieve sequence-dependent destruction of DNAs. The modifications often involve coupling of reagents capable of cleaving the DNA via reactions similar to those used in footprinting experiments (see Sect. 2.1). The underlying principles of DNA-protein recognition, which are emerging from the study of repressor–operator complexes, will also assist in the design of drugs directed at altering the transcription of particular genes.

See also Alternatively Spliced Genes; Bioorganic Chemistry; Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Aggarwaal, A.K., Harrison, S.C. (1990) DNA recognition by proteins with the helix-turn-helix motif, *Annu. Rev. Biochem.* **59**, 933–969.

Lilley, D.M.J. (Ed.) (1995) *DNA-protein: Structural Interactions*, IRL Press at OUP, Oxford.

Müller-Hill, B. (1996) *The lac Operon: A Short History of a Genetic Paradigm*, Walter de Gruyter, Berlin.

Old, I.G., Phillips, S.E.V., Stockley, P.G., Saint Girons, I. (1991) Regulation of methionine biosynthesis in the enterobacteriaceae, *Prog. Biophys. Mol. Biol.* **56**, 145–185.

Phillips, S.E.V. (1994) The β-ribbon DNA recognition motif, *Annu. Rev. Biophys. Biomol. Struct.* **23**, 671–701.

Ptashne, M., Gann, A. (2002) *Genes & Signals*, Cold Spring Harbor Press, Cold Spring Harbor, New York.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

Sauer, R.T., Jordan, S.R., Pabo, C.O. (1990) λ repressor: a model system for understanding protein-DNA interactions and protein stability, *Adv. Protein Chem.* **40**, 1–61.

Primary Literature

Anderson, W.F., Ohlendorf, D.H., Takeda, Y., Matthews, B.W. (1981) Structure of the *cro* repressor from bacteriophage λ and its interaction with DNA, *Nature* **290**, 754–758.

Anderson, W.F., Takeda, Y., Ohlendorf, D.H., Matthews, B.W. (1982) Proposed α-helical super-secondary structure associated with protein-DNA recognition, *J. Mol. Biol.* **159**, 745–751.

Beamer, L.J., Pabo, C.O. (1992) Refined 1.8 Å crystal structure of the lambda repressor-operator complex, *J. Mol. Biol.* **227**, 177–196.

Besse, M., Von Wilcken-Bergmann, B., Müller-Hill, B. (1986) Synthetic lac operator mediates repression through lac repressor when introduced upstream and downstream from lac repressor, *EMBO J.* **5**, 1377–1381.

Brennan, R.G., Roderick, S.L., Takeda, Y., Matthews, B.W. (1990) Protein-DNA conformational changes in the crystal structure of a λ Cro-operator complex, *PNAS (USA)* **87**, 8165–8169.

Carey, J. (1991) Gel retardation, *Methods Enzymol.* **208**, 103–117.

Carey, J., Lewis, D.E.A., Lavoie, T.A., Yang, J. (1991) How does *trp* repressor bind to its operator? *J. Biol. Chem.* **266**, 24509–24513.

Fickert, R., Müller-Hill, B. (1992) How lac repressor finds lac operator in vitro, *J. Mol. Biol.* **226**, 59–68.

He, Y.Y., Garvie, C.W., Elworthy, S., Manfield, I., McNally, T., Lawrenson, I.D., Phillips, S.E.V., Stockley, P.G (2002) Structural and functional studies of an intermediate on the pathway to operator binding by *E. coli*, *J. Mol. Biol.* **320**, 39–53.

He, Y.Y., McNally, T., Manfield, I., Navratil, O., Old, I.G., Phillips, S.E.V., Saint-Girons, I., Stockley, P.G. (1992) Probing *met* repressor-operator recognition in solution, *Nature* **359**, 431–433.

Joachimiak, A., Haran, T.E., Sigler, P.B. (1994) Mutagenesis supports water mediated recognition in the trp repressor-operator system, *EMBO J.* **13**, 367–372.

Jordan, S.R., Pabo, C.O. (1988) Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions, *Science* **242**, 893–899.

Karlsson, R., Roos, H., Fagerstam, L., Persson, B. (1994) Kinetic and Concentration Analysis Using BIA Technology, *Methods; A Companion to Methods in Enzymology*, Vol. 6, Academic Press, pp. 99–110.

Koudelka, G.B., Carlson, P. (1992) DNA twisting and the effects of non-contacted bases on affinity of 434 operator for 434 repressor, *Nature* **355**, 886–889.

Koudelka, G.B., Harbury, P., Harrison, S.C., Ptashne, M. (1988) DNA twisting and the affinity of bacteriophage 434 operator for bacteriophage 434 repressor, *PNAS (USA)* **85**, 4633–4637.

Lawson, C.L., Carey, J. (1993) Tandem binding in crystals of a trp repressor/operator half-site complex, *Nature* **366**, 178–182.

Levandoski, M.M., Tsodikov, O.V., Frank, D.E., Melcher, S.E., Saeker, R.M., Record, M.T. (1996) Co-operative and anti-co-operative

effects in binding of the first and second plasmid Osym operators to a LacI tetramer: evidence for contributions of non-operator DNA binding by wrapping and looping, *J. Mol. Biol.* **260**, 697–717.

Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G., Lu, P. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer, *Science* **271**, 1247–1254.

McKay, D.B., Steitz, T.A. (1981) Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA, *Nature* **290**, 744–749.

Mondragon, A., Harrison, S.C. (1991) The phage 434 Cro/O_R1 complex at 2.5 Å resolution, *J. Mol. Biol.* **219**, 321–334.

Oehler, S., Eismann, E.R., Krämer, H., Müller-Hill, B. (1990) The three operators of the *lac* operon cooperate in repression, *EMBO J.* **9**, 973–979.

Ogata, R.T., Gilbert, W. (1979) DNA-binding site of lac repressor probed by dimethylsulphate methylation of *lac* operon, *J. Mol. Biol.* **132**, 709–728.

O'Shannessy, D.J., Bringham-Burke, M., Soneson, K.K., Hensley, P., Brooks, I. (1993) Determination of rate and equilibrium binding constants for macromolecular interactions using surface plasmon resonance: use of non-linear squares analysis methods, *Anal. Biochem.* **212**, 457–468.

Otwinowski, Z., Schevitz, R.W., Zhang, R.G., Lawson, C.L., Joachimiak, A., Marmorstein, R.Q., Luisi, B.F., Sigler, P.B. (1988) Crystal structure of *trp* repressor/operator complex at atomic resolution, *Nature* **335**, 321–329.

Pabo, C.O., Lewis, M. (1982) The operator-binding domain of λ repressor: structure and DNA recognition, *Nature* **298**, 443–447.

Parsons, I.D., Persson, B., Mekhalfia, A., Blackburn, G.M., Stockley, P.G. (1995) Probing the molecular mechanism of action of co-repressor in the *E.coli* methionine repressor-operator complex using surface plasmon resonance, *Nucleic Acids Res.* **23**, 211–216.

Phillips, S.E.V., Manfield, I., Parsons, I., Davidson, B.E., Rafferty, J.B., Somers, W.S., Margarita, D., Cohen, G.N., Saint-Girons, I., Stockley, P.G. (1989) Co-operative tandem binding of *met* repressor of *Escherichia coli*, *Nature* **341**, 711–715.

Phillips, K., Phillips, S.E.V., (1994) Electrostatic activation of *E.coli* methionine repressor, *Structure* **2**, 309–316.

Phillips, S.E.V., Stockley, P.G. (1994) Similarity of *E.coli met* and *trp* repressors, *Nature* **368**, 106.

Phillips, S.E.V., Stockley, P.G. (1996) Structure and function of *Escherichia coli met* repressor: similarities and contrasts with *trp* repressor, *Philos. Trans. R. Soc. London, Series B* **351**, 527–535.

Rafferty, J.B., Somers, W.S., Saint-Girons, I., Phillips, S.E.V. (1989) Three dimensional crystal structures of the *Escherichia coli* Met repressor with and without co-repressor, *Nature* **341**, 705–710.

Raumann, B.E., Brown, B.M., Sauer, R.T. (1994) Major groove DNA recognition by β-sheets: the ribbon-helix-helix family of gene regulatory proteins, *Curr. Opin. Struct. Biol.* **4**, 36–43.

Rogers, D.W., Harrison, S.C. (1993) The complex between phage 434 repressor DNA-binding domain and operator site OR3: structural differences between consensus and non-consensus half-sites, *Structure* **1**, 227–240.

Saint-Girons, I., Belfaiza, J., Guillou, Y., Perrin, D., Guiso, N., Barzu, O., Cohen, G.N. (1986) Interactions of the *E.coli* methionine repressor with the *metF* operator and with its co-repressor, S-adenosylmethionine, *J. Biol. Chem.* **261**, 10936–10940.

Schevitz, R.W., Otwinowski, Z., Joachimiak, A., Lawson, C.L., Sigler, P.B. (1985) The three-dimensional structure of *trp* repressor, *Nature* **317**, 782–786.

Schultz, S.C., Shields, G.C., Steitz, T.A. (1991) Crystal structure of a CAP-DNA complex: the DNA is bent by 90°, *Science* **253**, 1001–1007.

Seeman, N.C., Rosenberg, J.M., Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins, *PNAS (USA)* **73**, 804–808.

Somers, W.S., Phillips, S.E.V. (1992) Crystal structure of the *met* repressor-operator complex at 2.8 Å resolution: DNA recognition by β-strands, *Nature* **359**, 387–393.

Stockley, P.G., Baron, A.J., Wild, C.M., Parsons, I.D., Miller, C.M., Holtham, C.A.M., Baumberg, S. (1998) Dissecting the molecular details of prokaryotic transcriptional control: the methionine and arginine

repressor proteins, *Biosens. Bioelectron.* **13**, 637–650.

Urpi, L., Tereshko, V., Malinina, L., Huynh-Dinh, T., Subirana, J.A. (1996) Structural comparison between the d(CTAG) sequence in oligonucleotides and trp and met repressor-operator complexes, *Nat. Struct. Biol.* **3**, 325–328.

Van Duyne, G.D., Ghosh, G., Maas, W.K., Sigler, P.B. (1996) Structure of the oligomerization and L-arginine binding domain of the arginine repressor of *Escherichia coli*, *J. Mol. Biol.* **256**, 377–391.

von Hippel, P.H., Berg, O.G. (1989) Facilitated target location in biological systems, *J. Biol. Chem.* **264**, 675–678.

Wharton, R.P., Brown, E.L., Ptashne, M. (1984) Substituting an α-helix switches the sequence specific DNA interactions of a repressor, *Cell* **38**, 361–369.

Wild, C.M., McNally, T., Phillips, S.E.V., Stockley, P.G. (1996) Effects of systematic variation of the minimal *Escherichia coli met* consensus operator site: *In vivo* and *in vitro met* repressor binding, *Mol. Microbiol.* **21**, 1125–1135.

Zhang, R.-G., Joachimiak, A., Lawson, C.L., Schevitz, R.W., Otwinowski, Z., Sigler, P.G. (1987) The crystal structure of *trp* aporepressor at 1.8 Å shows how binding tryptophan enhances DNA affinity, *Nature* **327**, 591–597.

DNA–Protein Interactions

Sylvie Rimsky and Malcolm Buckle
Enzymologie et Cinétique Structurale. LBPA. UMR 8113 du CNRS,
Ecole Normale Supérieure, 61 Avenue du Président Wilson, 94235 Cachan cedex
France

1 Protein–DNA Interactions 533
1.1 Structural Composition of Nucleoprotein Complexes 533
1.1.1 Protein Structural Motifs 533
1.1.2 DNA Structures Recognized by DNA-Binding Proteins 534
1.1.3 The Thermodynamics of the Formation of Nucleoprotein Complexes 535

2 The Role of Nucleoprotein Complexes in Gene Expression 537
2.1 The Core Transcription Apparatus 537
2.1.1 The Initiation Complex 537
2.1.2 The Kinetics of Promoter Recognition and Escape 538
2.2 Bacterial Chromatin Structure 541

3 General Methods for Studying Nucleoprotein Complexes 543
3.1 Chemical and Enzymatic Footprinting 543
3.2 Electromobility Shift Assay 545
3.3 Cross-linking of Nucleoprotein Complexes 547
3.3.1 Formaldehyde Cross-linking 547
3.3.2 Laser UV Cross-linking and Photofootprinting 548
3.4 Surface Plasmon Resonance (SPR) 551
3.5 Microcalorimetry 555
3.6 Fluorescence 556
3.6.1 Fluorescence Intensity Measurements 556
3.6.2 Fluorescence Anisotropy 556
3.6.3 Fluorescence Energy Transfer (FRET) 557
3.7 Electron Microscopy and Atomic Force Microscopy 557

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

Bibliography 559
Books and Reviews 559
Primary Literature 559

Keywords

Activator
A protein that accelerates transcription, generally at the level of promoter recognition, by binding to specific sites that may be at, near, or distant from the promoter region.

Core Transcription Complex
Machinery responsible for initiating the transcription of one strand of DNA from a promoter sequence into an RNA molecule. This complex contains RNA polymerase.

Genome
The information coded in DNA for the expression of all the proteins specific to the phenotype of a given organism.

Promoter
The minimal sequences needed to initiate correct transcription. In prokaryotes, this consists of consensus −35 and −10 regions, spacer regions (18 bp ± 2) between the two, UP (upstream) elements, and upstream activator sequences. In eukaryotes, it often contains TATA box elements and initiation sequences but there is more variation with respect to the nature of the putative TATA region (which may, therefore, be CAAT, GC, or octamer), also these promoters are often assisted by enhancer elements.

Transcription Repressor
A molecule that inhibits transcription generally by sterically interfering with the formation of initiation complexes.

Transcription Factors
Proteins required for recruiting RNA polymerase to the transcription start site to form a preinitiation complex.

Proteins interact with DNA to form nucleoprotein complexes that mediate a host of important cellular processes. Nucleoprotein complexes are at the heart of many processes that include DNA recombination and repair, gene transcription, pathology, viral infection, and in defining DNA accessibility through chromosome structure. The interplay between proteins and DNA is dynamic and must be understood in a time-resolved fashion. Whilst a structural analysis of the individual components of such active complexes is essential, it has to be interpreted within the changing

environment of a cell as it responds to alterations in the environment or as it transits through a preprogrammed cycle of growth, differentiation, and finally but not always, apoptosis. While the number of processes involving nucleoprotein complexes is large, and the mechanisms by which they carry out these functions are many and complex, there is, however, a restricted known set of the structure–function relationships involved. In this chapter, we will attempt to sketch out some of the more common of these, always relating this to possible dynamic changes that occur during the establishment of the nucleoprotein complex or take place during its function. We will pay particular attention to complexes involved in gene regulation at the level of DNA organization and at the level of the control of transcription where the information contained in the genome is decoded to produce the RNA complement that constitutes the transcriptome and hence, the potential proteome of a cell at any given moment.

We have restricted our analysis to nucleoprotein complexes involved in processes in prokaryotic organisms for reasons of expediency, clarity, and because it falls within our own personal range of interest. However, the molecular basis for the formation of specific nucleoprotein complexes is clearly "system" independent and will be found across the whole range of living organisms; to paraphrase Jacques Monod "the same mechanisms present in *E. coli* will be at work in an elephant." It's clearly logistically more favorable to study the former than the latter!

1 Protein–DNA Interactions

1.1 Structural Composition of Nucleoprotein Complexes

Proteins are not mere collections of amino acids strung together in a series of chains or globular conglomerates but are modular entities consisting of a select set of three-dimensional structures arranged so as to provide a specific range of functions. In the same sense, DNA is most emphatically not a rod consisting of four bases in a sugar-phosphate backbone double helix, but rather possesses a well-defined local and global structure capable of extensive structural flexibility and of engaging in a myriad of interactions with proteins. Although the potential for the types of interactions is enormous, over the years recurring patterns of specific arrangements have been observed.

1.1.1 Protein Structural Motifs

Protein recognition of DNA sequences involves a wide range of contacts including nonpolar, direct, and indirect (water mediated) hydrogen bonding with DNA bases and direct and water mediated contacts with the sugar-phosphate backbone of the DNA. These interactions may occur in the minor groove or in the major groove or indeed in both grooves simultaneously. The protein structural motifs engaged in these interactions include helix-turn helix, basic leucine zipper (coiled coil) and zinc finger structures, using α-helices, β-sheets, and loops to effect the recognition. A relatively recent study showed that of the three types of interactions characterizing

complexes involving protein residues and DNA, namely, van der Waals contacts, hydrogen bonds, and water mediated bonds, two thirds fell into the category of van der Waals contacts and the remainder was more or less evenly distributed between the other two modes. This rather surprising observation suggests that the simple definition regarding protein–DNA interactions as being specific or nonspecific needs to be examined in some detail. A step in this direction has been made by taking into account the observation that in general a protein is presented to a nucleic acid so as to promote one or both of two distinct types of interaction:

- indirect readout in which the protein recognizes structural characteristics on the DNA;
- direct readout in which there is generally a direct hydrogen-bond distance interaction between amino acids and a given base on the DNA.

The three types of interactions outlined above, however, are in fact present in both these categories and indeed a given nucleoprotein complex may involve both direct and indirect readout. Thus, the distinction between sequence-specific recognition and structural recognition becomes blurred by the dependence of DNA structure on base sequence. The local structure is going to influence the accessibility of bases to the protein. Van der Waals contacts occur mainly with the phosphate backbone of the DNA; some contacts are made with the sugars and a few directly to bases. Water mediated contacts make up nearly 15% of all contacts observed to date and hydrogen bonds, simple, bidentate, and complex, are found essentially between the phosphate backbone and amino acid residues. A rule of thumb is that while the direct hydrogen bonds reflect specific interactions, water mediated and van der Waals contacts act as stabilizing factors and may be useful in allowing the protein in indirect readout of the DNA. However, to date, there are no data that support a simple general code for protein–DNA recognition.

1.1.2 DNA Structures Recognized by DNA-Binding Proteins

There are many types of DNA structures reflecting the myriad ways in which a rod of DNA can be spatially organized through the disposition of the bases with the sugar-phosphate backbone and the orientation of the major and minor grooves. A typical list of such structures would include bent or curved DNA, four-way cruciform junctions, locally melted "bubbles," forked junctions, and so forth. We are going to concentrate on protein binding to bent or curved DNA sequences to investigate some of the salient features of DNA structure recognition by proteins. The insertion of short runs of adenines (A tracts) phased with the double helix in a DNA molecule generally induces a bend by a mechanism that remains elusive but apparently has to accommodate the more or less universally accepted notion that the center of curvature is toward the minor groove of the A tract and toward the major groove of intervening sequences. In other words, one would expect compression of the minor groove within the A tract and compression of the major groove of adjoining sequences. A key question relates to how this structure lends itself to recognition by certain proteins. Answering this question will require a concerted approach that takes into consideration the thermodynamic, kinetic, and structural aspects of the situation.

1.1.3 The Thermodynamics of the Formation of Nucleoprotein Complexes

Since protein–DNA interactions are competitive and noncovalent, the rates and extents of formation of complexes are affected by macromolecular concentrations and by the concentrations of low molecular mass competitors such as electrolytes. There are essentially five interactions that are important in determining the stability and specificity of nucleoprotein complexes. These are the hydrophobic effect, the polyelectrolyte effect, aromatic ring interactions, ionic interactions, and hydrogen bonding interactions. The hydrophobic effect in which the association of nonpolar surfaces is driven by the release of water molecules, and the polyelectrolyte effects that derive from the reduction of the polyelectrolyte charge density at low salt concentrations essentially help reduce the unfavorable noncovalent interaction between uncomplexed macromolecules and a solvent and hence, effectively promote the formation of macromolecular complexes.

Thermodynamically, the strategies adopted to attain a favorable free energy of binding (ΔG_{bind}) are rather varied. Simply stated, the measure of the change in work energy of a system during a reaction, that is, during the formation of a nucleoprotein complex, involves an enthalpic change (ΔH°) and an entropic change $-T\Delta S^\circ$, given by the Gibbs relationship

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \qquad (1)$$

The formation of a nucleoprotein complex should be accompanied by a decrease in ΔG°, that is, $\Delta G^\circ < 0$. It is thus obvious that an enthalpically driven reaction ($\Delta H^\circ < 0$) compensates for an unfavorable entropic change ($-T\Delta S^\circ > 0$) and vice versa. An interesting observation shows that whilst ΔH° and $-T\Delta S^\circ$ values for a number of nucleoprotein complexes involving specific DNA recognition can vary across a broad range ($\sim$60 kcal mol^{-1}), the resulting ΔG° values lie within a small range. This observation that ΔH° and ΔS° do not vary independently has important repercussions. In Table 1, are collected the various molecular processes that are associated with either ΔH° or ΔS°. Although this table is incomplete and oversimplified, it illustrates the combinations available in the play-off between $-T\Delta S^\circ$ and ΔH°.

Strain is obviously a major factor to be taken into consideration. Inducing a distortion on the DNA is clearly enthalpically unfavorable. To unstack a single base pair costs between 7 and 11 kcal mol^{-1}, in the

Tab. 1 Generalized enthalpic or entropic favorability of the main groups of interactions in nucleoprotein and protein–protein contacts. Strain means any configuration for a given atom, molecule or group that deviates from the minimal potential energy for that entity in the free protein or DNA.

Process	*Effect on ΔH°*	*Effect on $-T\Delta S^\circ$*
Hydrogen-bond contacts	Favorable	–
Ion-pair contacts	Favorable	–
Nonpolar contacts	Favorable	–
Desolvation	Unfavorable	Favorable
Strain	Unfavorable	Favorable

case of open complex formation between RNA polymerase and a promoter sequence involving the unstacking of at least 12 bp this could cost up to 77 kcal mol^{-1}. It is clearly expensive and at first sight unfavorable for proteins to choose a binding process that induces strain in the DNA. An attractive proposition as to why proteins have evolved this mechanism is that the induced distortion in the DNA can then be used for other, related processes. In the case of the formation of open complexes with RNA polymerase, the apparent cost of the distortion may already have been met to some extent by topological distortions induced in the DNA by the prior binding of activators.

It seems, however, that nucleoprotein complexes that distort the DNA benefit from a ΔS° driven change in the face of an unfavorable ΔH°, whereas protein–DNA complex formation that does not appreciably distort the DNA is driven by a favorable ΔH° against an unfavorable ΔS°.

How do induced changes in the protein influence the binding characteristics for a given complex? Notwithstanding the many examples that now exist of three-dimensional structures for nucleoprotein complexes, there remains a dearth of concrete data that provide direct insight into this question. Clearly, there is reduced translational and rotational freedom for both the protein and the DNA in the nucleoprotein complex and this will lead to an unfavorable ΔS°. However, in cases in which little distortion of the DNA occurs, then some induced changes in the protein most probably occur and would thus be associated with an unfavorable entropic change. In other words, there may be associated folding changes in the protein during DNA binding, in order to adapt to a given existing structure on the DNA. However, there are no hard and fast rules for predicting the extent of protein rearrangements during binding, and few precise experiments that provide information on this process.

It is worthwhile at this point to examine the phenomena in thermodynamics known as entropy–enthalpy compensation. The fact is that both ΔH and ΔS are temperature dependent. The variation of enthalpy with temperature at constant pressure is given by

$$C_p = \left(\frac{dH}{dT}\right)_p \tag{2}$$

C_p is the molar specific heat capacity for a single compound and is in fact the amount of heat required to raise the temperature of 1 mol by 1 K. Another way of thinking about C_p is to see it as the capacity of a substance to absorb heat energy without increasing the kinetic energy of the substance and thus without increasing the temperature.

Any change in the heat capacity (ΔC_p) between reactants and products in a reaction therefore can be expressed as

$$\Delta C_p = \frac{d(\Delta H)}{dT} \tag{3}$$

Similarly, entropy changes depend on C_p, at equilibrium, $\delta H = T\delta S$, so

$$C_p = \frac{dH}{dT} = \frac{T\,dS}{dT} \tag{4}$$

$$\frac{\Delta C_p}{T} = \frac{d(\Delta S)}{dT} \tag{5}$$

From Eqs. (3) and (5) we can see that integrating from absolute zero to the temperature of interest we can get

$$H = \int C_p\,dT \text{ and } S = \int \frac{C_p}{T\,dT} \tag{6}$$

Thus, if there is a change in heat capacity during a process then both ΔH

and ΔS will be temperature dependent, since $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$, then these effects will effectively cancel, thus producing what is called entropy–enthalpy compensation.

It is curious that most protein–DNA interactions are associated with a strong negative ΔC_p change. Since as mentioned above, DNA-binding proteins may or may not distort the DNA and/or undergo some induced fit during binding, any distortion will presumably be associated with a ΔC_p effect, the entropy–enthalpy compensation may thus play a crucial role and the real synergy that exists between the effect of ΔH and ΔS will need to take into consideration the relationships described by Eq. (6). However, to date, insufficient data are available in order to establish the ground rules for a complete thermodynamic understanding of the formation of nucleoprotein complexes but as the following sections illustrate, techniques are developing that should furnish this information.

2 The Role of Nucleoprotein Complexes in Gene Expression

Gene expression requires intimate interactions between proteins and nucleic acids, and therefore represents an ideal context in which to discuss many of the concepts involved in the formation of nucleoprotein complexes. We are going to concentrate on two aspects of this process; first the process whereby RNA polymerases involved in prokaryotic transcription recognize specific regions of DNA to form competent transcription complexes, and secondly how the spatial organization of the DNA affects this process in the cell.

2.1 The Core Transcription Apparatus

2.1.1 The Initiation Complex

Transcription is the process whereby one of the two strands of DNA are copied into an RNA molecule. The decision to transcribe a given region of DNA is at the core of the regulation of gene expression and thus is a fundamental biological process. The enzymes responsible for transcription are the RNA polymerases. In prokaryotes, RNA polymerase is a relatively large multi-subunit enzyme consisting of 4 subunits $((\alpha)_2, \beta', \beta, \omega, \sigma)$ in the core enzyme and 5 subunits $(((\alpha)_2), \beta', \beta, \omega, \sigma)$ in the holoenzyme. The core enzyme (378 784 Da) consists of the two β (150 625 Da) and β' (155 145 Da) subunits, the dimeric form of the α (36 507 Da) subunit involved in interactions with activators and specific regions on the DNA and the ω subunit. The holoenzyme (449 044 Da), the form that actively recognizes promoter sequences responsible for housekeeping genes, contains the σ^{70} subunit (70 260 Da). Recent three-dimensional structures for the core enzyme and holoenzyme in complexes with DNA have shed considerable light on the structural orientation of the RNA polymerase machinery. However, since the process is essentially a dynamic one, the mechanism by which a σ^{70} containing RNA polymerase operates in order to locate its DNA counterpart, locally separate the two DNA strands, and precisely position the catalytic center at the start site for transcription, is not understood at the molecular level. Without a σ-subunit, *Escherichia coli* RNA polymerase is unable to recognize a promoter sequence. In agreement with the direct contact model originally proposed, several genetic studies have indicated that σ^{70} interacts in a

bipartite manner via its conserved regions 2.4 and 4.2, with two hexameric consensus sequences (separated by 17 (±1 bp)) located around positions −10 and −35 relative to the transcription start site. More recent biochemical evidence has corroborated the interaction with the −10 region and direct cross-linking data have revealed the presence of intimate contacts made by σ^{70} with the DNA in the −35 region.

2.1.2 The Kinetics of Promoter Recognition and Escape

The kinetics of promoter formation is best characterized in prokaryotic systems and we shall first describe this in some depth. We are using this as an example of the predominantly dynamic aspect of protein–DNA interactions in a very basic and important cellular context. Most processes in the cell are not at equilibrium, and understanding these processes requires an appreciation of the kinetics involved and the evolution of the complexes over time, especially in a cellular context. The basic pathway leading to mRNA production is depicted in the following scheme:

$$\mathrm{R} + \mathrm{P} \underset{k_{-1}}{\overset{k_1}{\longleftrightarrow}} \mathrm{RP_c} \underset{k_{-2}}{\overset{k_2}{\longleftrightarrow}} \mathrm{RP_I} \underset{k_{-3}}{\overset{k_3}{\longleftrightarrow}} \mathrm{RP_O}$$

$$\xrightarrow{\mathrm{NTP}} \mathrm{RP_{init}} \xrightarrow{\mathrm{NTP}} \mathrm{RP_E}$$

In this scheme R, P, and RP represent the RNA polymerase, the promoter containing DNA, and the various RNA polymerase-promoter complexes respectively. RP_C denotes the initial closed complex. RP_I includes all subsequent species in which the DNA in the promoter region remains closed. RP_O signifies complexes that have undergone promoter opening. RP_{init} denotes those species performing the reiterative synthesis of short abortive products. The ensuing escape from the promoter leads to the formation of a stable elongation complex, RP_E. The formation of the initial closed complex (RP_C) involves the specific recognition of conserved promoter sequences. While the −35 hexamer is certainly involved in the initial recognition process, the upstream portion of the −10 element has been implicated as well, since both regions are specifically recognized by σ^{70} as double-stranded DNA. The rate of formation of the RP_C complex may be accelerated by base nonspecific binding mechanisms that allow for two-dimensional sliding along the DNA, "hopping" or intersegment transfer between juxtaposed regions of DNA. However, the estimated association rate constants remain in line with a process of three-dimensional diffusion. Moreover, the equilibrium between free RNA polymerase (R) and the RP_C complex (K_B) is extremely rapid at many promoters, meaning that the subsequent isomerization step, and not initial complex formation, is rate-limiting.

However, many promoters are not saturated with RNA polymerase *in vivo*, and can thus be activated at K_B through increased recruitment of the RNA polymerase to the promoter. This most often involves the establishment of contacts between DNA-bound activating proteins and the RNA polymerase that serve to increase binding affinity, although additional DNA–RNA polymerase contacts may also serve this purpose. Activators involved in this step often bind to the DNA upstream of the −35 region, at sites centered on the same side of the DNA helix (the activator protein, CRP – cyclic AMP receptor protein – for example). Thus, RNA polymerase may form favorable interactions between its α-CTD and the activator, which increase initial promoter binding. Alternatively, the

α-CTD can mediate an increase in recognition via contacts with a UP element. Or, DNA-binding proteins (such as the factor for inversion stimulation, the protein FIS) may serve to induce a DNA bend, which brings upstream sequences closer to the RNA polymerase allowing for additional RNA polymerase-DNA interactions. Intrinsic bends in the upstream DNA region can also serve to facilitate promoter binding, presumably by permitting similar interactions between the RNA polymerase and upstream DNA sequences. Such interactions with the upstream region may be sufficient to cause topological distortion in the DNA, thereby explaining the change in linking number observed during the formation of RP_C. Indeed a change of -0.2 (from -1.6 to -1.4 in linking number, or the number of times that two strands cross over) was measured for this transition, and has been interpreted as resulting from DNA bending or conformational change. The ensuing torsion in the DNA could be exploited by the RNA polymerase later in the transcription cycle. In support of this idea, it has been shown that superhelical density also plays a role in determining K_B at some promoters.

The formation of the RP_I complex is believed to include an induced torque in the DNA and/or bending, which introduces an unwinding twist near the transcription start site. Investigations into the effects of supercoiling on promoter usage have suggested that there is an early topological untwisting of 0.3–0.5 of a helical turn, which takes place during the first isomerization (i.e. before promoter opening). These data have lead to the attractive hypothesis that the formation of RP_I aligns the -35 and -10 hexamers on the same side of the DNA helix. Evidence for DNA bending in this complex is provided by scanning atomic force microscopy, native gel band shift analysis and the DNase I hyperreactivity of bases in the vicinity of -38, -48, and the spacer region.

The transition to an open complex (RP_O) is accompanied by DNA strand separation, which is thought to begin within the -10 region followed by propagation downstream past the transcription start site ($+2$ or $+3$). This melting of between 1 and 1.5 turns of DNA thus renders the template strand single stranded and accessible to the initiating NTP (as well as chemical probes such as dimethyl sulfate (DMS), diethyl pyrocarbonate (DEP) and $KMnO_4$). However, the DNA backbone between -10 and $+20$ remains protected from HO radical attack, and the DNase I footprint is unchanged from that of RP_I, indicating that promoter opening occurs without any additional displacement of the RNA polymerase on the DNA. Also the nontemplate strand remains inaccessible to chemical attack.

No external energy source is necessary for the transition to RP_O, making it likely that the RNA polymerase not only lowers the activation energy for DNA strand separation but forms favorable interactions in the open complex as well. As mentioned previously, it has been postulated that some of the required energy comes from the release of a stressed RNA polymerase-DNA intermediate. This intermediate is presumed to be RP_I. Upon release of this strained species, one would expect conformational changes within the RNA polymerase and/or DNA. In support of this idea, during the RP_I to RP_O transition, there was a change noted in the direction of curvature of the DNA entering and exiting the RNA polymerase and in the linking number (from -1.2 in RP_I to -1.8 in RP_O).

In addition to this role in the facilitation of strand opening, interactions with the RNA polymerase are also believed to stabilize the single-stranded region. Indeed, recent studies have demonstrated that the σ^{70} subunit is competent to recognize specifically single-stranded oligodeoxynucleotides corresponding to the nontemplate strand −10 sequence. The binding affinity measured for this sequence was higher when the region was single stranded than for the analogous double-stranded fragment, thereby implicating a single-strand specificity in this interaction

Once the DNA strands have been locally melted and the stable open complex formed, the template strand becomes accessible to incoming nucleoside triphosphates. The nucleotide at position +1 on the template strand will, by Watson–Crick base pairing, determine the identity of the 5′ end of the transcript. The binding of the initiating NTP to the active site serves to stabilize the RNA polymerase–promoter complex, even in the absence of hydrolysis or bond formation. Transcript elongation requires the catalysis of phosphodiester bond formation between the hydroxyl at the current 3′ end, and the subsequent NTP, which will be determined by base pairing with the following nucleotide.

However, a significant proportion of the initiating complexes do not immediately go on to synthesize productive transcript. Instead, these complexes repetitively form and then release small transcripts (up to 10 nt) without either leaving the promoter region or releasing RNA polymerase-promoter contacts. This reiterative process, termed abortive initiation, results in the cyclic formation of short nonproductive mRNAs. These small transcripts dissociate from the stably bound polymerase, leaving the active site free once again to bind the initiating NTP. At each early nucleotide position, the decision between transcript release and continued elongation relies on a kinetic competition between the two processes. It follows that the duration of RNA polymerase stalling on a promoter is determined by the strength of RNA polymerase-promoter contacts as well as flanking promoter sequence. In this way, a promoter with sequences optimized for stability at preceding steps can be hindered by these same contacts during promoter clearance. Therefore, each functional promoter must represent a delicate equilibrium between stable complex formation and facile escape.

Productive mRNA synthesis is only achieved after the RNA polymerase ceases this process and moves downstream of the promoter region. This promoter clearance is thought to require the release of the specific RNA polymerase contacts with the promoter region, accomplished by the separation of the core enzyme from σ^{70} during the RP_{init} to RP_E transition. In accordance with this idea, the size of the DNase I footprint produced by RP_E is considerably diminished in comparison to initiating complexes (20–30 bp vs 65–95 bp). The observed decrease in promoter protection coincides with promoter clearance, indicating that core-σ dissociation occurs at this step. The transcription bubble is also enlarged during this transition, from ~10 to between 17 and 18 nucleotides. The requirement for additional melting of the DNA in the downstream region may also explain the importance of downstream sequence composition in promoter escape.

The resultant elongating complex is incredibly salt tolerant, resisting up to 350 mM NaCl. Thus, the core polymerase in RP_E binds DNA in a fashion that is both extremely stable to dissociation and yet capable of rapid translocation along

the DNA. There is an obvious parallel between this RNA polymerase activity and that displayed by the DNA polymerases. However, there is no apparent analog for the "sliding clamp" apparatus for the bacterial RNA polymerases. Numerous studies have probed the conformational changes undergone during the transition by the core subunits themselves. As such it has been proposed that "jaw-like" regions of the β' subunit close around the DNA forming close interactions with the downstream sequences.

2.2 Bacterial Chromatin Structure

In simple terms, the decision to express a given gene is, as we have seen above, taken at the level of recognition of promoter sequences by the transcription machinery. This is clearly a complex process involving a number of proteins, some of which are molecular motors, working in a concerted kinetic fashion. However, *in vivo*, this dynamic process takes place against the backdrop of the DNA organization within the cell. Although a lot has been achieved in the study of eukaryotic chromosome structure, our general understanding of the organization and expression of genetic information remains incomplete. The progress of research in bacteria has been hampered in part by a low degree of DNA compaction (albeit around 1000-fold) and resultant instability of the bacterial chromatin. Studies conducted in the paradigm organism *E. coli* revealed that most of the structural proteins of the bacterial nucleoid (Fig. 1) are synthesized in a growth phase-dependent manner.

The relative abundance of several major nucleoid-associated proteins (Fig. 1) in exponential phase is FIS > HU > H-NS > IHF > Lrp, whereas the order of abundance on entry in stationary phase changes to Dps > IHF > HU > H-NS > Lrp > FIS. In organisms like *E. coli* in which HU is composed of two subunits (α and β), the subunit composition also changes as a function of growth phase. This change in the chromosomal protein composition on transition from exponential growth into stationary phase is accompanied by silencing of genome functions. Thus, these proteins not only define the architecture of the bacterial chromosome but also specifically change the pattern of gene expression in a growth phase-dependent manner. This coupling between the modulation of chromosome structure and the pattern of gene expression is further reinforced by the topological fluctuations of DNA during the growth phase and the strong dependence of bacterial promoter activity on DNA topology.

Although the existence of a link between the metabolic reorganization of the cell and the transitions in the architecture of the bacterial chromosome is widely accepted, how the alterations in cellular physiology are coupled with alterations in the topology of the chromosome remains largely unexplored. We will briefly discuss the role of one of the proteins involved in this process – the H-NS protein – in the light of its DNA-binding properties linked to its function as a transcription repressor.

H-NS is an atypical global regulator involved in the expression of genes, which seem to share the common property of being involved in the cells response to changes in the external environment. Modification of the *hns* gene modifies or completely abolishes the expression of around 200 proteins in *E. coli*. H-NS is involved in such diverse functions as the response to osmolarity changes, the use of certain carbon sources, antibiotic resistance, ribosomal RNA regulation,

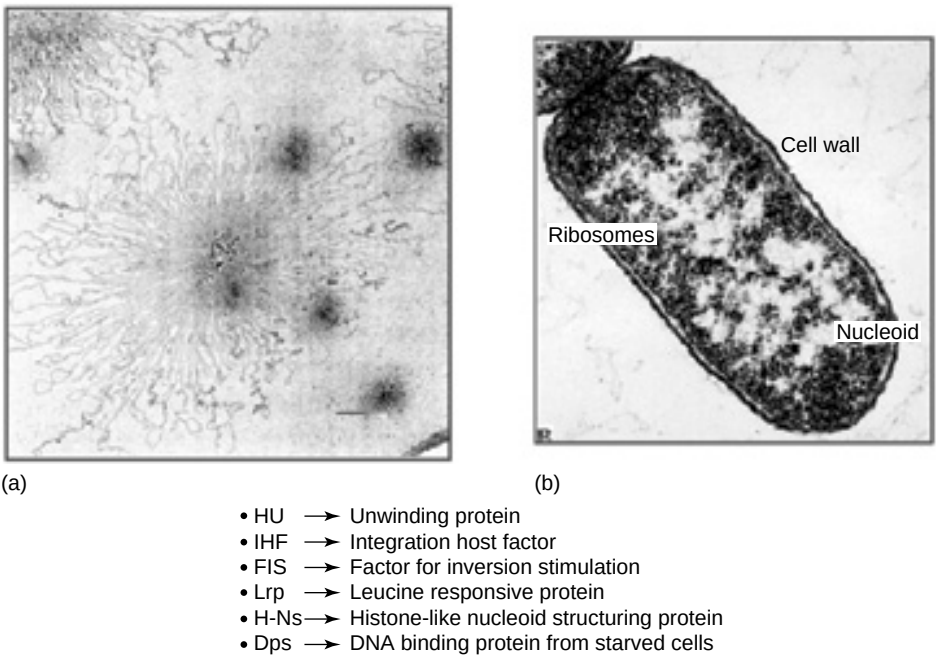

Fig. 1 Structural proteins of the bacterial nucleoid. Electron micrographs of (b) a complete *E. coli* bacterium showing the DNA condensed (the white material) within the cell) and (a) the same cell exploded after salt treatment showing the DNA now completely unwound, note that the horizontal blue bar in (a) shows the length of the original prior to disruption. The organization of the DNA is mostly carried out by the proteins listed (see color plate p. xxxviii).

pili formation, and the activation of virulence genes in pathogenic bacteria. H-NS affects transcription in a differential manner depending on the growth phase, and consequently is strongly involved in the processes of virulence and silencing in bacteria.

Although H-NS does not exhibit a high DNA sequence specificity, H-NS recognizes and preferentially binds to DNA-containing phased A_n tracts that are intrinsically curved. A number of H-NS responsive promoters have indeed been shown to contain regions of intrinsic DNA curvature located either upstream or downstream of the transcription start point. The available data support the idea that the presence of such DNA sequences *in vivo* renders proximal promoters specifically responsive to transcription modulation by H-NS. A crucial question therefore concerns on the nature of the nucleoprotein complexes involving H-NS. A working model for H-NS binding has been proposed in which differential binding to relatively high and low affinity sites occurs. The formation of an H-NS/DNA complex at a region containing high affinity binding sites constitutes a first step of nucleation; this is followed by a buildup of H-NS binding to lower affinity sites on the DNA leading eventually to total occlusion of the DNA by a process of propagation. The high affinity sites clearly require the presence of a singular structural feature and the helical phasing of the recognized

motifs with the periodicity of the DNA. Note, therefore, that specificity of binding is a function of both the intrinsic structure and the deformability of the DNA. An essential feature of this model is that this propagation step is highly dependent on the formation of a "correct" nucleation event. How this binding mode is related to the repression of transcription at proximal or distal promoters remains an object of intensive research; however, a plausible mechanism is that the formation of the nucleation step leads to a facilitated coverage of nonspecific sites on the DNA and that interactions between H-NS bound at "specific" and "nonspecific" sites are responsible for the formation of higher order structures that either physically occlude RNA polymerase from promoter regions (even promoter regions that are not within high affinity H-NS binding-site regions) or provide torsional resistance to the RNA polymerase induced DNA transformations required for the formation of competent transcription complexes. The higher order structure produced by H-NS is most probably not a simple filament that would in general be insufficiently compacted to explain the observed condensation of the DNA, nor would a simple filament explain the phenomenon of silencing in which H-NS is capable of repressing transcription at promoters quite distal to the sites of binding at concentrations incommensurate with the complete coverage of a filamentous DNA. Quite clearly, H-NS is a complex transcriptional regulator exerting its effects through an unusual binding mechanism and forming higher order nucleoprotein complexes. It is to be expected that this mechanism will turn out to be even more intricate when the full nature of interactions between H-NS and other proteins within the bacterial chromatin organization machinery are unraveled.

3 General Methods for Studying Nucleoprotein Complexes

3.1 Chemical and Enzymatic Footprinting

The reactivity of DNA to various chemical and enzymatic reagents has been extensively utilized to probe the local and global structure of DNA in nucleoprotein complexes. The wealth of information that this has yielded has been relatively recently extended by the possibility of applying certain of the techniques in a time-resolved fashion. We will describe the main techniques used in what we call a static (i.e. at equilibrium) and a dynamic (i.e. time resolved) context. We will also try to illustrate attempts to access changes in protein conformation during nucleoprotein complex formation.

Table 2 contains a nonexhaustive list of the reagents that target DNA.

In general, a nucleoprotein complex is treated under specific conditions with the reagent and the products of the reaction examined by denaturing electrophoresis and phosphor imager densitometry. The choice of a specific reagent depends on the nature of the nucleoprotein complex. If knowledge of the position of a protein on a given DNA molecule is required, then DNAse I, exonuclease III, and the more precise hydroxyl radical approach are used. These techniques may also furnish information concerning the quantification of the reaction. The results obtained from attacks by DNAse I are referred to as footprints for the simple reason that the DNA in nucleoprotein complexes will not be accessible to the DNAse I and hence will not be cleaved in those regions of the DNA to which the protein is bound. DNase I, because it cleaves via the minor groove,

Tab. 2 Reagents that target nucleic acids.

Reagent	*Nature*	*Target*
Dnase I	Enzymatic	Phosphodiester bond in sugar-phosphate backbone. Attack via binding in the minor groove. Sequence independent, structure dependent
Exonuclease III	Enzymatic	3′–5′ exonuclease activity, removes mononucleotides. Sequence dependent, structure independent.
Hydroxy radicals	Chemical	Phosphodiester bond in sugar-phosphate backbone. Sequence independent, structure dependent.
Dimethyl sulfate (DMS)	Chemical	N-7 of guanines in major groove. Also cytosines when unpaired in single-stranded DNA. Sequence dependent, structure dependent.
Diethyl pyrocarbonate (DEP)	Chemical	Carboxyethylates N-7 of imidazole ring of purine residues in distorted DNA (deviation from classical B-form). Sequence dependent, structure dependent.
Potassium permanganate ($KmnO_4$)	Chemical	Oxidizes nonsaturated double bonds to vicinal diols, reacts preferentially with thymines. Sequence dependent, structure dependent.
Osmium tetroxide (OsO_4)	Chemical	Unstacked thymines. Sequence dependent, structure dependent.
1,10-Phenanthroline-copper	Chemical	Cleaves phosphodiester backbone through minor groove. Sequence independent, structure dependent.
Singlet oxygen	Chemical	Generated by irradiation of eosin isothiocyanate in the presence of Tris, singlet oxygen diffuses into the DNA and modifies bases. Sequence independent, structure dependent.
UV laser	Photonic	Photochemical rearrangement of bases, leads primarily to pyrimidine dimer formation and other rearrangements generally involving guanine residues. Sequence dependent, structure dependent.

will produce enhanced cleavage at those regions of the DNA possessing a widened minor groove. Thus, by looking at DNAse I cleavage patterns one can deduce, not only that region occupied by the protein but also those regions in which the DNA has been distorted leading to a widened minor groove.

Sometimes it is useful to probe changes in the local conformation of the DNA. In these cases, reagents that react with unusual configurations are used. Potassium permanganate, for example, which oxidizes unsaturated bonds in pyrimidines, reacts almost exclusively with thymines in which unstacking between a thymine and an adjacent base on the same DNA strand has exposed the 5–6 double bond containing the vicinal diols. Oxidation of thymines produces a labile DNA molecule that may be cleaved at the sugar-phosphate bond adjacent to the modified base. Such modifications may then be visualized, as for DNaseI cleavage, by the use of end-labeled DNA and separation on denaturing polyacrylamide sequencing gels. Consequently, where there is a local change in DNA structure (such as local melting of the DNA strands, characteristic of the formation of an initiating transcription complex

with RNA polymerase), the extent of distortion may be determined by the relative reactivity of the thymines to $KMnO_4$. Similarly, the reactivity of cytosines (DMS), adenines (DEP) (Table 2), and thymines (OsO_4) may be examined in different nucleoprotein contexts. 1,10-Phenanthroline-copper is a reagent that intercalates in the minor groove of DNA and can, under specific conditions, generate radicals that then cleave the sugar-phosphate backbone. It, therefore, provides information on the relative geometry of the DNA, that is, the minor groove, at specific points on the DNA. One can generate singlet oxygen molecules, which then migrate into the DNA and modify bases; the modified bases render the DNA suitable for cleavage in a similar fashion to the chemical modification ($KMnO_4$ etc.), and thus the site of modification may be simply ascertained.

The use of a reagent is limited by two considerations, the spatial resolution provided by the probe (DNAse I, for example, needs to bind in the minor groove and also requires a specific binding region on the DNA, which therefore limits the spatial resolution of the ensuing footprint), and the degree of perturbation of the nucleoprotein complex by the probe itself. Eosin isothiocyanate, used to generate singlet oxygen, actually very strongly perturbs many nucleoprotein complexes and thus one has to be sure that the structure that is being hit by the reagent is not in fact an artifact due to rearrangements in the nucleoprotein complex induced by the reagent itself.

The judicious use of these types of reagents provides useful information about the local structure of the DNA in these complexes. Indeed, these reagents have provided enormous insights into many, by now classical structures, ranging from cruciforms, and Holliday junctions to transcription initiation complexes.

In cases in which the formation of a nucleoprotein complex occurs over a timescale in excess of tens or hundreds of milliseconds, quite a few of the techniques listed in Table 2 are sufficiently rapid that they may be used to probe changes taking place during this process. This is an extremely important and powerful approach since many of the processes involving nucleoprotein complexes are practically irreversible and knowledge of the kinetics of these processes is essential for a full understanding of the mechanisms involved. Table 3 illustrates how three of the techniques used to probe DNA may be used in a time-resolved fashion.

3.2 Electromobility Shift Assay

The basis of the electromobility shift or more commonly called band shift assay is that protein–DNA complexes remain intact during migration in gel electrophoresis and thus appear as distinctly more slower moving bands than the free DNA fragment. The assay is simple and quick, and the use of radioactive binding-site DNA makes it highly sensitive. It can be used for both highly sequence-specific and nonspecific proteins, such as histones. Band shift assays can be used quantitatively to estimate dissociation constants for protein–DNA complexes. Additionally, band shift experiments can be used to visualize protein–protein interactions between a DNA-binding protein and other non-DNA-binding proteins. The binding of a second protein to a protein–DNA complex to form a triple complex is visualized by a further retardation of mobility that is called a *super shift*. Super shift experiments can

Tab. 3 Some time-resolved methods to study the kinetics of DNA–protein interactions.

Method	*Target and spatial resolution*	*Time resolution*	*Advantages*	*Disadvantages*
UV laser photofootprinting	van der Waals contacts, T-dimers	~10 ns	ns pulses no need to add any reagents	DNA damage only < 2 Å low quantum yield
DNase I footprinting	Protection and hypersensitive sites	30 ms	Strong signal	Needs Mg^{2+} large footprint
OH X-ray footprinting	Solvent accessibility of bases (amino acids)	3 ms	Single base resolution no need to add any reagents	Sensitive to quenchers such as glycerol and Tris buffer

also be used to assay the binding of a second DNA-binding protein to the DNA or the binding of a second molecule of DNA to the protein.

The band shift assay can be used simply to estimate dissociation constants for nucleoprotein complexes involving just one protein and its binding site. Two methods are generally used; in the first the protein concentration needs to be known accurately, whereas in the second method both the protein and the DNA concentrations are required. In both methods, the amount of free DNA and the amount of DNA bound in the protein–DNA complex have to be estimated either by measuring the intensity of a band in an auto radiograph using techniques such as classical densitometry, or more commonly nowadays, phosphor imaging. An important caveat when using the band shift assay is that association rate constants may be altered by the concentrating influence of the so-called cage effect of the polyacrylamide gel and that dissociation of the complexes may be drastically affected by the change in buffer conditions as the complex enters the gel. The net result may well be important deviations from equilibrium and thus band shift assays are most useful for comparative studies rather than for obtaining absolute equilibrium values.

The first method involves the use of a low DNA concentration (below the dissociation constant) compared to a relatively large range of protein concentration. Since, in a simple situation in which a protein P binds to a DNA-binding site D, the concentration of nucleoprotein complex [PD] is related to the dissociation constant K_D by

$$K_D = \frac{[P][D]}{[PD]} \text{ such that when}$$

$$[D] \leq\leq K_D \text{ then } [P]_{free} \cong [P]_{total}$$

and so $K_D = ([P]_{total} \cdot [D])/[PD]$ in other words, when the concentration of DNA is equal to the concentration of the protein–DNA complex (50% complex formation) then the dissociation constant is given by the total protein concentration.

The second method for determining dissociation constants requires the preparation of a range of different concentrations of a one-to-one molar mix of protein and DNA. The proportion of DNA in the complex at the different concentrations depends upon the ratio between the

total DNA concentration and the dissociation constant. The most convenient way to perform this measurement is to titrate (labeled) DNA at high concentrations (i.e. at least 1000 times the estimated dissociation constant) with increasing amounts of protein. Analysis of a band shift gel allows the determination of the point at which a one-to-one complex is formed. Subsequently, a one-to-one complex, formed under these conditions, is diluted serially and the analysis repeated at different DNA concentrations to estimate the percentage of complex. The concentration of complex is dependent upon the dissociation constant and the concentration of DNA, since in the equilibrium

$$\mathrm{P} + \mathrm{D} \longleftrightarrow \mathrm{PD}$$

$$K_{\mathrm{D}} = \frac{[\mathrm{P}]_{\mathrm{free}} \cdot [\mathrm{D}]_{\mathrm{free}}}{[\mathrm{PD}]} \tag{7}$$

and

$$[\mathrm{D}]_{\mathrm{total}} = [\mathrm{PD}] + [\mathrm{D}]_{\mathrm{free}} \tag{8}$$

since the required condition is that

$$[\mathrm{P}]_{\mathrm{total}} = [\mathrm{D}]_{\mathrm{total}} \tag{9}$$

then

$$[\mathrm{P}]_{\mathrm{free}} = [\mathrm{D}]_{\mathrm{free}} \tag{10}$$

and

$$K_{\mathrm{D}} = \frac{([\mathrm{D}]_{\mathrm{total}} - [\mathrm{PD}])^2}{[\mathrm{PD}]} \tag{11}$$

so

$$[\mathrm{PD}]^2 - (2[\mathrm{D}]_{\mathrm{total}} + K_{\mathrm{D}})[\mathrm{PD}] + [\mathrm{D}]^2_{\mathrm{total}} = 0 \tag{12}$$

then

$$[\mathrm{PD}] = \frac{(2[\mathrm{D}]_{\mathrm{total}} + K_{\mathrm{D}}) - \sqrt{(2[\mathrm{D}]_{\mathrm{total}} + K_{\mathrm{D}})^2 - 4[\mathrm{D}]^2_{\mathrm{total}}}}{2} \tag{13}$$

It must be emphasized that the apparent K_{D} values obtained are primarily of use in a comparative sense rather than to allow a rigorous thermodynamic analysis.

3.3 Cross-linking of Nucleoprotein Complexes

3.3.1 Formaldehyde Cross-linking

Formaldehyde is a highly reactive reagent that produces protein–DNA and protein–protein cross-links between macromolecules in close contact. The distance between the cross-linking groups of protein and DNA should be around 2 Å suggesting an interaction at the range of van der Waals radii. Therefore, the formaldehyde cross-linking procedure allows identification of proteins and protein domains in close proximity to the DNA. Formaldehyde cross-linking had been widely used for studies of protein–DNA interactions in chromatin *in vivo* as described in detail. Although the reaction of formaldehyde with nucleotides and DNA is well characterized, the exact mechanism of the cross-linking reaction between proteins and DNA is not well known. It was shown that formaldehyde reacts with the amino groups of cytosines, guanines, and adenines and the imino groups of thymines and probably guanines. In proteins, potential candidates for cross-linking are lysine, arginine, tryptophan, and histidine residues. The first stage includes the reaction with amino or imino groups and results in the formation of methylol derivatives that react with the adjacent second reacting group. In conditions in which short cross-linking times are used and cross-links occur only in the melted DNA regions, the imino groups of thymine (or guanine) bases are highly probable targets of cross-linking reaction. This assumption is based on two known observations:

- Kinetic studies demonstrated slow formaldehyde reaction rates with amino

groups while reaction rates of imino group were fast.

- Reactions with the amino groups of purines can take place with double-stranded DNA as well as with melted DNA regions, while the reaction with imino groups of thymine occurs only in melted DNA regions.

The product of the first reaction step (methylol derivative) is very unstable while the final cross-linked derivatives are stable at 24 °C. Cross-linked complexes of RNA polymerase subunits can be incubated for several hours at room temperature without significant degradation. At 50 °C, slow degradation of protein–DNA cross-links was detected after 20 min of incubation in Tris-HCl containing buffer ($t_{1/2}$ = 90 min). Incubation for 2 min at 95 °C leads to full disruption of the cross-links.

The identification of proteins or subunits cross-linked under different experimental conditions is crucial. Two methods provide simple and direct ways to characterize these contacts. The first method requires polyclonal antibodies to individual proteins or subunits and is a modification of the method used for fractionation of cross-linked chromatin. The method includes precipitation of the cross-linked complexes with subunit-specific antibodies and purification on protein A-Sepharose affinity chromatography. Analysis of bound and nonbound fractions on SDS-PAGE allows the identification of those subunits involved in any particular cross-linked complex. The second method utilizes genetically engineered subunits containing histidine-tags at the carboxyl end or the amino-terminal end. In this case, the cross-linked complexes must be denatured by urea and subjected to fractionation on Ni^{2+}-NTA agarose affinity columns.

3.3.2 Laser UV Cross-linking and Photofootprinting

The intrinsic photoreactivity of DNA and RNA is a useful parameter to measure local DNA structure and identify regions in contact with proteins in given nucleoprotein complexes. In order to satisfy two main criteria of nonperturbation and selectivity, an optimum source of photons must be rapidly delivered to the sample. This is possible by using a UV laser providing high-energy photons (266 nm) in a brief pulse (5 ns) on a small volume of material (10–20 μL). The minimal experimental set up that is required is shown in Fig. 2.

The laser source is a neodymium laser consisting of Nd^{3+} ions at low concentration in yttrium-aluminum-garnet (YAG = $Y_3Al_5O_{12}$) that produces a continuous monochromatic beam at 1064 nm. In Q-switching mode, in conjunction with nonlinear crystals, this beam is quadrupled in frequency to give a homogenous polarized source of photons at 266 nm (approximately 10^{17} photons per pulse proving a dose of around 20–30 mJ per pulse). Since the beam has a diameter of approximately 6 mm, this gives complete irradiation of a solution of 10 to 20 μL at the bottom of an Eppendorf tube. The optimal beam form is a Gaussian rather than a "doughnut" configuration. The final 266 nm beam is selected using an array of dichroic mirrors and directed onto the sample using a Pellin–Brocca Prism (Fig. 2). To all intents and purposes, only nucleotide bases are excited at 266 nm. At the photon fluxes used in the setup described in Fig. 2, each base in a solution of DNA, for example, will see at least one photon carrying approximately 4 eV. Under the conditions used here, the ensuing photochemistry is monophotonic and the DNA is not ionized. This is generally confirmed by the absence of nicks

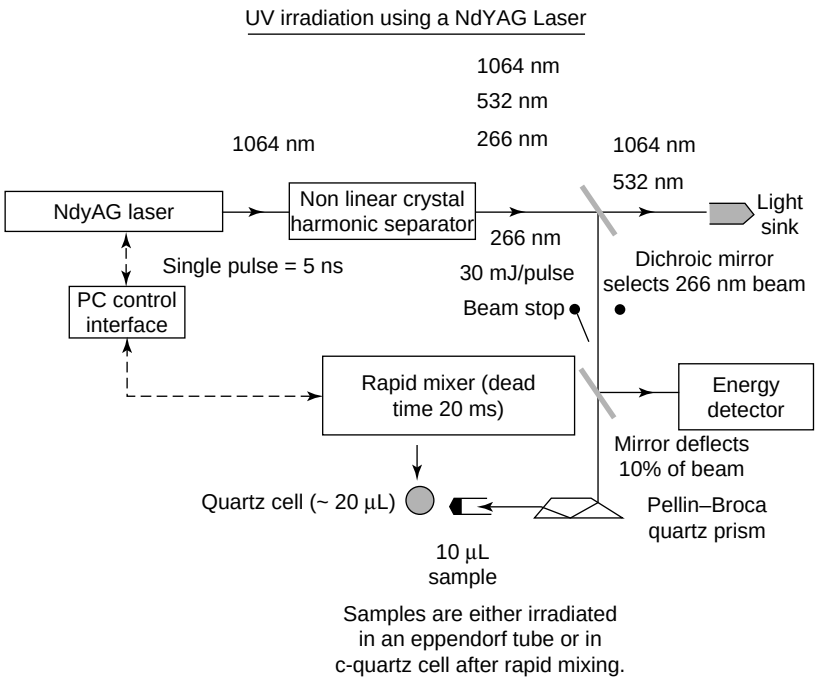

Fig. 2 Typical UV laser set up for the irradiation of nucleoprotein complexes. Complexes may be irradiated in an Eppendorf tube or as they pass through a mixing cell in a stopped flow apparatus.

on DNA strands following irradiation. It should be noted that the ionization energy for thymines is 7.6 eV in water. Furthermore, we can assume that in simple terms, the excited state for the base follows the profile shown in Fig. 3.

Irradiation thus produces excited singlet states of very short (picosecond) duration and a very small proportion of triplet states with longer (μs) lifetimes. In consequence, the photo reactive species are relatively short lived and 98 to 99% are consumed within the 5 ns duration of the laser pulse. What then is the fate of the excited species? There are a number of potential photoreactive products but three in particular are of interest here. The first is the formation of pyrimidine dimers. Adjacent nonsaturated conjugated 5–6 positions in the pyrimidine ring can, following excitation, form covalent cyclic products, notably 4–6 and 5–6 cyclobutanes. The quantum yield for the 4–6 cyclobutanes is around 4%, which essentially means that for a 100 bp DNA, there will be one thymine dimer per molecule following a single pulse at 266 nm. The second event concerns modification of guanine residues. The nature of this reaction is somewhat obscure. The quantum yield may be relatively high and the overall chemistry may require a biphotonic event. The final photochemical event is perhaps the most interesting and involves the formation of covalent links between excited bases and amino acids. The conditions for this cross-linking are that the reactive species are within van der Waals radii distance of each other.

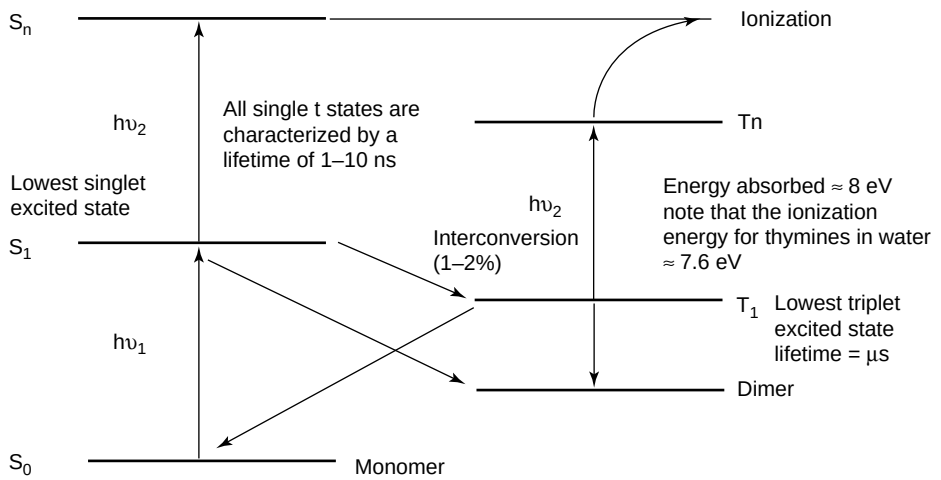

Fig. 3 Electronic states and lifetimes of excited states of atoms in nucleic acids following irradiation. The S1 state is that which is most commonly involved either in cross-linking with amino acids or undergoes internal rearrangement to structures such as pyrimidine dimers, for example.

These limitations arise essentially because of the very short lifetimes of the reactive species (ps). This limitation is of course present for pyrimidine dimer formation and represents one of the major advantages of the technique, namely, that it is extremely rapid, thus reporting configurations that exist before perturbations occur. A simple approach to determine the extent of UV photo-cross-linking to a DNA substrate is to separate the protein–DNA adduct(s) from non–cross-linked DNA by electrophoresis on sodium dodecyl sulfate (SDS)-polyacrylamide gels (SDS-PAGE). Quantification of the separated radio-labeled species is then performed by densitometry, for example, using a PhosphorImager. For most purposes, complexes are formed using 5′-end labeled oligonucleotides (i.e. labeled with (γ-^{32}P)-ATP and polynucleotide kinase). After single-pulse UV irradiation, the samples are separated by SDS-PAGE electrophoresis. The gel matrix should be optimized to allow a clear separation between the protein–DNA adduct and the DNA substrate. The electrophoretic migration of the UV-induced adduct will be dependent on its apparent mass, which, in most cases, equates directly to the sum of the covalently linked components (i.e. the apparent mass of the cross-linked DNA strand plus that of the attached polypeptide chain, calculate approximately 325 Da for each nucleotide of the cross-linked strand). If the DNA fragment is too large to be separated clearly, then an alternative methodology must be considered, for example, by first fragmenting the DNA or by degrading it with a nuclease.

When analyzing cross-linking data, it is important to perform measurements over a range of protein concentrations. This is essential when comparing the relative cross-linking efficiencies of different DNA substrates, since it is the saturation isotherm and not the absolute level of DNA cross-linking that contains the information

on binding affinity. Generally, titrations of fixed concentrations of the oligonucleotide are performed over a gradient of increasing protein concentration. A typical SDS-PAGE gel (visualized by autoradiography) and resulting binding curve analysis is shown in Fig. 4

The progressive fractional saturation of potential DNA-interacting sites is signaled by the increase in adduct formation, measured in terms of the percentage of the total DNA cross-linked (S_{obs}). Maximal cross-linking (S_{max}) is obtained under saturating concentrations of protein in which all the DNA-binding sites are occupied. Adduct formation may thus be represented simply as

$$S_{obs} = \theta . S_{max} = \theta . \lambda_c \qquad (14)$$

where θ ($= S_{obs}/S_{max}$) denotes the fractional saturation of the DNA lattice and λ_c represents the quantum yield of adduct obtained at saturation of the lattice with ligand, a constant factor dictated by the photochemical reactivity of the oligonucleotide sequence occluded and the number and nature of protein–DNA contacts formed. At 266 nm, λ_c is highly dependent on the base composition of the occluded site. Comparison of the data obtained for two independent proteins (gene32 protein and HIV-1 integrase), both of which are capable of binding non-selectively to ssDNA, indicate that the reactivity of each base may be ordered (relative to dT = 1) as T(1) $\gg$ C(0.04/0.08) $>$ A(0.03/0.04) $\gg$ G($\sim$0). Thus, as a general "rule-of-thumb," pyrimidine residues are more reactive than their purine counterparts, with d(T) by far the most efficient at producing protein–DNA cross-links. Of course, in order for an adduct to form, a close contact must exist between an amino acid and the reactive base. While most amino acids appear capable of cross-linking, the photochemistry is poorly understood and the influence of the side chains on the efficiency of the reaction cannot be predicted for a complex nucleoprotein structure. Least-squares analysis of graphical representations of the degree of cross-linking parameter θ as a function of ligand concentration (Fig. 4) then allows the fractional saturation to be derived. This in turn may then be used to determine binding isotherm characteristics from which an apparent equilibrium constant K_{obs} and apparent stoichiometry may be obtained from curves of the type shown in Fig. 4.

A simple technique for establishing those bases having undergone a specific photoreactivity is that of primer extension. Following irradiation, the bases on a given DNA molecule will have been photomodified and the assumption is that for any given DNA molecule, only one photoactive event will have taken place. Figure 4 illustrates the methodology involved. Essentially, a DNA polymerase such as the Klenow fragment is used to carry out primer extension of end-labeled oligomers along an irradiated template. The Klenow fragment terminates extension at position $n - 1$ with respect to a photomodified base. The photomodification may be pyrimidine dimer formation; intrinsic photomodification of a given base (e.g. G) or the presence of a covalent link with a protein. Products may be easily visualized on denaturing sequencing gels and the extent of photomodification determined using phosphor imager densitometry.

3.4 Surface Plasmon Resonance (SPR)

SPR is a technique that paradoxically measures apparent rate constants for the interaction between an immobilized

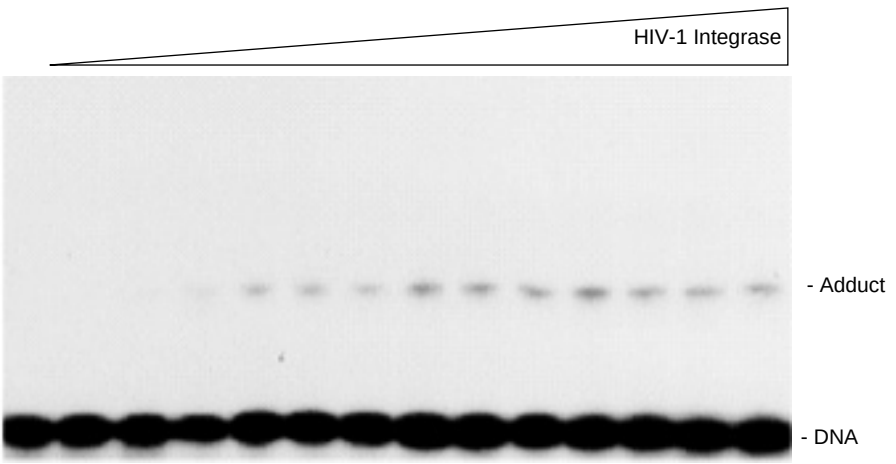

(a)

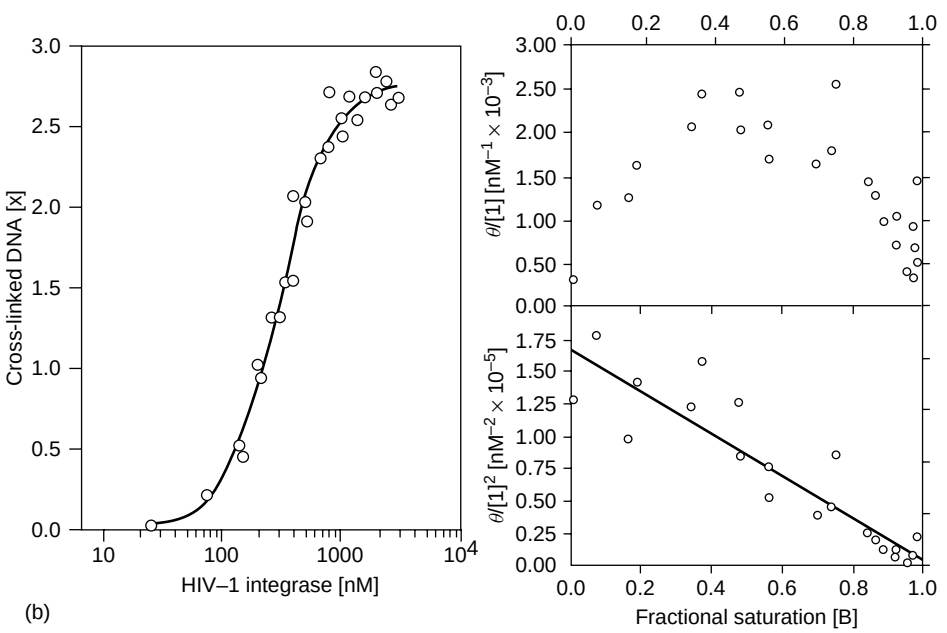

(b)

ligand and an analyte in solution in order to calculate equilibrium binding constants. The general scope and utility of this approach has been covered in a number of reviews. It is, however, worthwhile here to review the basic principles behind the technique in order to understand its application to studying nucleoprotein complexes. The SPR principles used by the main commercial supplier of this technology (Biacore™) are illustrated in Fig. 5.

Fig. 4 Laser-mediated UV cross-linking of the HIV-1 integrase to DNA. The HIV integrase protein was incubated with a 21-bp duplex DNA molecule (25 nM) for 10 min at 21 °C prior to exposure to a 5-ns pulse of 266 nm UV light. (a) Cross-linked adducts were resolved from un–cross-linked DNA by SDS-polyacrylamide gel electrophoresis. The autoradiograph shows part of a typical titration experiment performed by varying the integrase concentration between 0 and 3 mM. (b) Data points represent the quantification (by phosphor image densitometry) of gels such as the one shown in (a). The sigmoidal curve, shown in semilog form (*panel A*), represents the best fit of the data by nonlinear least-squares procedures in terms of the total HIV-1 integrase concentration and corresponds to a Hill coefficient of 1.90. In *panel B*, the data are transformed, in terms of fractional saturation, to Scatchard form for a first-order (*top*) and second-order (*bottom*) reaction. In this case, the data points are corrected for the bound protein assuming a density of 10 monomers per LTR at saturation. The Hill coefficient is 1.94 after transformation.

Light from a laser source arriving through a prism at a gold surface at the angle of total internal reflection (θ) induces a nonpropagative evanescent wave that penetrates into the flow cell opposite the prism. The intensity of the reflected light is continuously monitored. At a given angle (λ) dependent upon the refractive index of the solution in the flow cell, resonance between the evanescent wave and free electrons in the gold layer results in a reduction in the intensity of reflected light. The change in angle of reduced intensity ($\Delta\lambda$) reflects changes in the refractive index (n) of the solution in the flow cell immediately adjacent to the gold layer. A dextran surface coupled to the gold layer allows immobilization of ligands

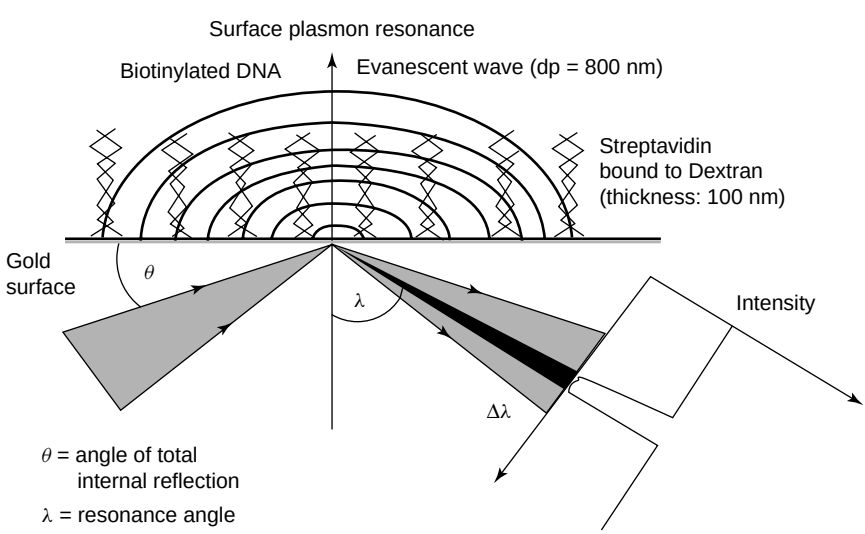

Fig. 5 Basic principles of surface plasmon resonance. In this example, DNA molecules are immobilized through a biotin–streptavidin interaction at the dextran/gold interface. Changes in refractive index as molecules bind to the surface are measured by alterations in the angle λ at which extinction of reflection occurs.

within the evanescent field. In physical terms, the detection system of a typical SPR machine consists of a monochromatic, plane-polarized light source and a photodetector that are connected optically through a glass prism. A thin gold film (50 nm) thick, deposited on one side of the prism, is in contact with the sample solution; this gold film is in turn covered with a long-chain hydroxyalkanethiol, which forms a monolayer (approximately 100-nm thick) at the surface. This layer essentially serves as an attachment point for carboxymethylated dextran chains that create a hydrophilic surface to which ligands can be covalently coupled. Light incident to the back of the metal film is totally internally reflected on to the diode-array detector. A property of this situation is that a nonpropagative evanescent wave penetrates into the solution side of the prism away from the light source. Free electrons in the gold layer enter into resonance with the evanescent wave. In fact, such resonance implies that the amplitude vector characterizing a transversal wave propagating along the gold surface ($\overrightarrow{ksp}$) is equal to the component ($\overrightarrow{kx}$) of the evanescent wave. Since $\varepsilon = n^2$, then if ω is the frequency of the wave and c the speed of light then

$$|\mathbf{kx}| = \frac{\omega}{\mathrm{c}}\sqrt{\frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + \varepsilon_2}} \tag{15}$$

Furthermore, given that for the evanescent wave:

$$|\mathbf{kx}| = \frac{\omega}{\mathrm{c}}\sin\theta\sqrt{\varepsilon_g} \tag{16}$$

When resonance occurs, $|\mathbf{ksp}| = |\mathbf{kx}|$ and the intensity of the reflected light decreases at a sharply defined angle of incidence, the SPR angle, given by the simple expression:

$$\sin^2\theta_0 = \frac{\varepsilon_1\varepsilon_2}{\varepsilon_g(\varepsilon_1 + \varepsilon_2)} \tag{17}$$

Thus θ_0, the SPR angle at which a decrease in the intensity of reflected light occurs, measures the refractive index of the solution in contact with the gold surface and is dependent on several instrumental parameters, for example, the wavelength of the light source and the metal of the film. When these parameters are kept constant, the SPR angle shifts are dependent only on changes in refractive index of a thin layer adjacent to the metal surface. Any increase of material at the surface will cause a successive increase of the SPR angle, which is detected as a shift of the position of the light intensity minimum on the diode array. This change can be monitored over time, thus allowing changes in local concentration to be accurately followed. The SPR angle shifts obtained from different proteins in solution have been correlated to surface concentrations determined from radio-labeling techniques and found to be linear over a wide range of surface concentrations. The instrument output, the resonance signal, is indicated in resonance units (RU); in the Biacore configuration, 1000 RU correspond to a 0.1° shift in the SPR angle and for an average protein this corresponds to a surface concentration change of about 1 ng mm^{-2}. The machine measures refractive index changes (Δn) at or near a surface and relates these to changes in local concentration. This relationship is given by the Clausius–Mossotti form

$$\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{N}{3\varepsilon 0}\left(\frac{\alpha + \mu}{kT}\right) \tag{18}$$

of the Debye equation

$$\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{N.\alpha}{3.\varepsilon 0} \tag{19}$$

where ε is the real part of the dielectric constant or permittivity constant related to the refractive index by $\varepsilon = n^2$, N is the number density given by $N_a\rho/M_a$ (N_a = Avogadro's number, ρ = the density and M_a the molecular mass). It is assumed that $\Delta n/\Delta C$ is a constant.

In general, either the DNA or the protein may be immobilized on the surface and the choice will depend on the nature of the interaction to be studied. It is, however, technically more expedient to immobilize DNA. Whilst a variety of techniques exist for the immobilization of DNA on the dextran surface, the most efficient for the majority of protein/DNA interactions is the use of immobilized streptavidin that can then interact with a suitably end-labeled DNA molecule. While it is often possible to obtain useful binding data pertaining to a specific nucleoprotein complex, it is critical to recall that SPR measurements monitor a steady state rather than equilibrium, that is, a reaction involving two components may form an initial binary complex that then "evolves" through intramolecular interactions into a final complex having different dissociation characteristics than that of the intermediate(s) present during the reaction. In many cases now, SPR has furnished useful data concerning specific nucleoprotein complexes and even in complicated cases, such as promoter recognition by RNA polymerase, SPR can furnish useful binding data that correlates well with other footprinting techniques.

An interesting application of this technique toward a more dynamic approach was carried out by using SPR to follow the polymerization of a reverse transcriptase along an immobilized single-stranded DNA template. The resulting change in SPR signal could be closely correlated with the incremental progression of the reverse transcriptase along the immobilized DNA template as it replicated the corresponding RNA or DNA strand. This approach has also been extended to observing RNA polymerases transcribing an immobilized double-stranded DNA template into an RNA molecule and again could be correlated with changes in the population of polymerases on the surface and the appearance of the RNA molecule.

3.5 Microcalorimetry

Microcalorimetry, like SPR, is a non-invasive technique for determining the magnitude of a specific process, such as the formation of a nucleoprotein complex. Microcalorimetry, however, in contrast to SPR, provides access to real thermodynamic measurements of a system at equilibrium. A system at thermodynamic equilibrium is of course defined as one where $\Delta G = 0$. This state is also defined by the equilibrium constant K, where, for a complex AB formed between A and B, $K = [AB]/[A][B]$ (in which case K is known as the *association* or *affinity constant*). This ratio, reflecting the relative probabilities of finding the molecules in a bound or free state, is related to ΔG° by the simple expression

$$K = e^{\left(\frac{-\Delta G^\circ}{RT}\right)} \tag{20}$$

and since

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ, \tag{21}$$

$$K = e^{\left(\frac{-\Delta H^\circ}{RT}\right)} + e^{\left(\frac{-\Delta S^\circ}{R}\right)} \tag{22}$$

This is a form of the familiar van't Hoff equation that allows calculation of ΔH and ΔS by studying the variation of K as a function of temperature.

Microcalorimetry, however, directly measures heat energy changes during a reaction. There are essentially two approaches in use, differential scanning calorimetry (DSC) and isothermal titration calorimetry (ITC). DSC studies changes that occur as the temperature of a reaction is altered, whereas ITC measures thermal changes during the mixing of reactants. Both approaches provide measurements of ΔC_p and thus by the application of relationships like those given in Eqs. (20), (21), and (22) allow calculation of ΔH and ΔS (and thus ΔG and K).

We mentioned above that nucleoprotein complexes are composed of noncovalent interactions and as such the forces that are involved are therefore relatively weak such that the enthalpy changes involved are of the order of 1–25 kcal mol^{-1}. It is to the credit of calorimetric technology and those who use the approach that they are able to measure the changes in the order of differences of 10^{-6} °C associated with these weak changes.

3.6 Fluorescence

Fluorescence-based studies are becoming increasingly popular especially amongst the more biophysically oriented research groups. The advantages offered by this technique include the possibility of being able to carry out studies in solution in a true equilibrium fashion and in many cases in a time-resolved fashion. The dynamic range of sensitivity extends from milli to pico molar and the increased availability of diverse reagents extends the versatility of the approach. There are essentially three fluorescence-based strategies that are currently in use with respect to nucleoprotein complexes.

3.6.1 Fluorescence Intensity Measurements

Changes in fluorescence intensity originally relied on changes in the intrinsic fluorescence of protein upon binding to nucleic acids. Alternatively, fluorescence modified bases such as 2-aminopurine and ethenedeoxy-adenosine or extrinsically labeled nucleic acids have also been extensively used. In general, this latter approach has been favored because of the relatively high quantum yield (i.e. the number of reactant molecules that react for every photon absorbed) allowing good sensitivity.

3.6.2 Fluorescence Anisotropy

A phenomenon that is in fact the basis for dichroism is that the absorption of light by a chromophore depends upon the relative orientation of the absorbing chromophore with respect to the polarization of the light. When the incident light is polarized, then absorption will clearly be more likely for those molecules with chromophores whose transition dipole moments are parallel to the direction of polarization. By transition dipole moment we mean that, for a molecule to be able to absorb or emit a photon of frequency ν, it must possess, albeit transiently, a dipole oscillating at ν. Indeed, it is a property of this phenomenon that the intensity of the transition associated with either absorption or emission of a photon is proportional to the square of the transition dipole moment. In other words, for a transition dipole moment making an angle θ with respect to the direction of light polarization, the molar absorption coefficient ε_0 (also referred to as the extinction coefficient) is given by

$$\frac{\varepsilon_0}{\varepsilon_1} = \cos^2\theta \qquad (23)$$

By using light pulses of short duration and by measuring the fluorescent emission using polarizers, the intensity of emission parallel (I_a) can be compared with the intensity of the emission perpendicular (I_b) to the incident light.

This is very useful since the emission anisotropy may be obtained simply by using the expression

$$A = \frac{I_a - I_b}{I_a + 2I_b} \tag{24}$$

Anisotropy decays exponentially, that is,

$$A_{(t)} = A_{(0)} e^{t/\rho} \tag{25}$$

where ρ is the rotational relaxation time. It is clear that very short ($\sim 10^{-9}$ s) rotational relaxation times may be measured using this technique. Once one has a relaxation time, then it is very straightforward to calculate the hydrated volume V_h for the chromophore-bearing molecule using the simple expression

$$\rho_{(t)} = \frac{\eta V_h}{kT} \tag{26}$$

where η is the viscosity of the solution (J K^{-1} m^{-1} s^{-1}), k is the Boltzmann constant (1.38066×10^{-23} J K^{-1}), and T is the temperature in degrees Kelvin.

The rotational diffusion properties of a labeled oligonucleotide alter during complex formation with a protein thus providing information not only about the nature of the complex being formed but also providing the means of carrying out time-resolved analysis of the formation of the nucleoprotein complex. This latter is an increasingly powerful technique.

3.6.3 Fluorescence Energy Transfer (FRET)

FRET is the nonradiative exchange of excitation energy between two chromophores. In fact, an energy donor in an excited state transfers energy to an acceptor by means of intermolecular long-range dipole–dipole coupling. The quantum yield of transfer, E, depends on the distance (R) between the dyes and is given by

$$E = \frac{1}{1 + ([R]/[R]_0)^6} \tag{27}$$

In practice, although this depends to an extent on the nature of the fluorophore, the effective distance at which the phenomena may be useful is of the order of 4 to 6 nm. So one can see that this technique can be used in a nucleoprotein complex to measure distances between select chromophores and indeed has been applied in looking at changes in nucleic acid structure during complex formation or at putative contact points between RNA polymerases and promoter sequences in DNA.

3.7 Electron Microscopy and Atomic Force Microscopy

Electron and near field microscopy possess the capacity to visualize and follow the average behavior of a population of molecules observed individually and in the case of nucleoprotein complexes provide information concerning DNA curvature and flexibility intrinsic or induced by a protein. In classical electron microscopy (EM), molecules are adsorbed onto a carbon film, which may lead to some loss of three-dimensional information but generally will conserve the macromolecules conformational properties. EM generates data in two broad fields: imaging and analysis. Global electron irradiation of a specimen produces images on a fluorescent screen, a photographic film, or a TV camera. DNA or protein–DNA complexes may be directly adsorbed onto

carbon films or mica supports. Visualization is generally carried out by the use of contrasting procedures that do not interfere with the interactions under study. Thus, samples may be either stained using an evaporated metal coating (such as uranyl acetate) or by using an annular dark field in which the electron beam is tilted to provide a tilted illumination of the specimen and the image resulting from those electrons scattered at wide angles by the heavy atoms present in the sample. Contrast is obtained by impeding direct electrons and allowing only scattered electrons through the aperture. In all cases, direct images of DNA molecules or nucleoprotein complexes may be obtained (Fig. 6).

A useful parallel technique to EM is cryo-EM, which enables visualization

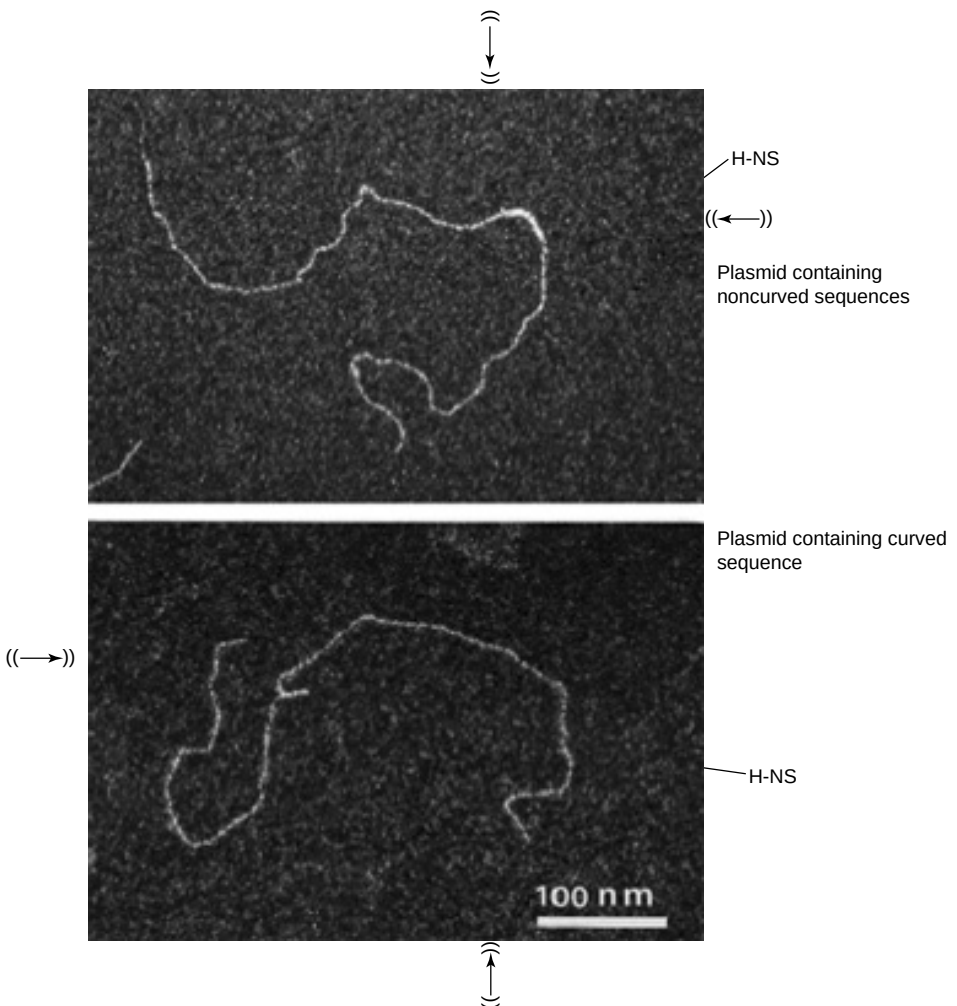

Fig. 6 Electron microscopy of nucleoprotein complexes. H-NS /DNA complexes as visualized by electron microscopy.

of nucleic acids, for example, without staining. A solution of the macromolecules is rapidly frozen to produce a thin film of vitrified water. By using EM, it is possible, in the correct configuration, to visualize molecules floating in the ensuing amorphous solution.

Scanning force microscopy provides a complementary technique to EM possessing high resolution and contrast. In a simple configuration known as contact mode, a flexible cantilever carrying a sharp tip is scanned across a surface. Deflections of the tip as it pans across the sample on the surface may be translated into a topographic image. A second operating mode known as tapping mode exists in which the cantilever oscillates at a high frequency during the scan. The surface topography is then reconstructed from changes in the oscillating amplitude of the cantilever. Nucleoprotein complexes are generally deposited onto a flat mica surface in a suitable buffer, then rinsed and dried under conditions optimized so as not to disturb the conformation of the complexes under study. In practice, this is a major restriction to the technique.

In general, in images formed by 512 × 512 pixels on a 2 μm scan size, each pixel represents a DNA of 3.9 nm or about 11 bp. In the case of nucleoprotein complexes, changes in the measured contour length of DNA molecules due to the presence of bound protein may be interpreted in terms of specific bending or deformation of the DNA.

See also Footprinting Methods to Examine the Structure and Dynamics of Proteins.

Bibliography

Books and Reviews

deHaseth, P.L., Zupancic, M.L., Record, M.T. Jr. (1998) RNA polymerase-promoter interactions: the comings and goings of RNA polymerase, *J. Bacteriol.* **180**(12), 3019–3025.

Hill, J.J., Royer, C.A. (1997) Fluorescence Approaches to Study of Protein-Nucleic Acid Complexation, in: Braud L., Johnson M.L. (Eds.) *Methods in Enzymology, Fluorescence Spectroscopy*, Vol. 278, Academic Press, pp. 278–416.

Ishihama, A. (1999) Modulation of the nucleoid, the transcription apparatus, and the translation machinery in bacteria for stationary phase survival, *Genes Cells* **4**(3), 135–143.

Kneale, G..G. (1994) DNA-Protein Interactions: Principles and Protocols, *Methods in Molecular Biology*, Vol. 30, Humana Press, Totowa, NJ.

Moss, T. (2001) DNA-Protein Interactions: Principles and Protocols, *Methods in Molecular Biology*, 2nd edition, Vol. 148, Humana Press, Totowa, NJ.

Murakami, K.S., Darst, S.A. (2003) Bacterial RNA polymerases: the wholo story, *Curr. Opin. Struct. Biol.* **13**(1), 31–39.

Ptashne, M., Gann, A. (1997) Transcriptional activation by recruitment, *Nature* **386**(6625), 569–577.

Sauer, R.T. (1991) *Protein-DNA Interactions Methods in Enzymology*, Vol. 208, Academic Press, London, UK.

Travers, A. (1996) Transcription: building an initiation machine, *Curr. Biol.* **6**(4), 401–403.

Travers, A. (1997) DNA-protein interactions: IHF–the master bender, *Curr. Biol.* **7**(4), R252–R254.

Travers, A.A., Buckle, M. (2000) *DNA-Protein Interactions: A Practical Approach*, Oxford University Press, New York.

Primary Literature

Adelman, K., Brody, E.N., Buckle, M. (1998) Stimulation of bacteriophage T4 middle transcription by the T4 proteins MotA and AsiA occurs at two distinct steps in the transcription cycle, *Proc. Natl. Acad. Sci. U.S.A.* **95**(26), 15247–15252.

Azam, T.A., Ishihama, A. (1999) Twelve species of these nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity, *J. Biol. Chem.* **274**(46), 33105–33113.

Bloch, V., Yang, Y., Margeat, E., Chavanieu, A., Auge, M.T., Robert, B., Arold, S., Rimsky, S., Kochoyan, M. (2003) The H-NS dimerization domain defines a new fold contributing to DNA recognition, *Nat. Struct. Biol.* **10**(3), 212–218.

Buckle, M., Buc, H. (1989) Fine mapping of DNA single-stranded regions using base-specific chemical probes: study of an open complex formed between RNA polymerase and the lac UV5 promoter, *Biochemistry* **28**(10), 4388–4396.

Buckle, M., Pemberton, I.K., Jacquet, M.A., Buc, H. (1999) The kinetics of sigma subunit directed promoter recognition by E. coli RNA polymerase, *J. Mol. Biol.* **285**(3), 955–964.

Buckle, M., Williams, R.M., Negroni, M., Buc, H. (1996) Real time measurements of elongation by a reverse transcriptase using surface plasmon resonance, *Proc. Natl. Acad. Sci. U.S.A.* **93**(2), 889–894.

Losick, R., Pero, J. (1981) Cascades of sigma factors, *Cell* **25**(3), 582–584.

Luscombe, N.M., Thornton, J.M. (2002) Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity, *J. Mol. Biol.* **320**(5), 991–1009.

Mukhopadhyay, J., Kapanidis, A.N., Mekler, V., Kortkhonjia, E., Ebright, Y.W., Ebright, R.H. (2001) Translocation of sigma (70) with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA, *Cell* **106**(4), 453–463.

Naryshkin, N., Revyakin, A., Kim, Y., Mekler, V., Ebright, R.H. (2000) Structural organization of the RNA polymerase-promoter open complex, *Cell* **101**(6), 601–611.

Pabo, C.O., Nekludova, L. (2000) Geometric analysis and comparison of protein-DNA interfaces: why is there no simple code for recognition? *J. Mol. Biol.* **301**(3), 597–624.

Record, M.T. Jr., Ha, J.H., Fisher, M.A. (1991) Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA, *Methods Enzymol.* **208**, 291–343.

Rimsky, S., Zuber, F., Buckle, M., Buc, H. (2001) A molecular mechanism for the repression of transcription by the H-NS protein, *Mol. Microbiol.* **42**(5), 1311–1323.

Travers, A., Muskhelishvili, G. (1998) DNA microloops and microdomains: a general mechanism for transcription activation by torsional transmission, *J. Mol. Biol.* **279**(5), 1027–1043.

Zinkel, S.S., Crothers, D.M. (1990) Comparative gel electrophoresis measurement of the DNA bend angle induced by the catabolite activator protein, *Biopolymers* **29**(1), 29–38.

Down Syndrome, Molecular Genetics of

Charles J. Epstein
University of California, San Francisco, CA, USA

1 Phenotype of Down Syndrome 562

2 Cytogenetics 563

3 Structure of Chromosome 21 563

4 Phenotypic Mapping 564

5 Pathogenesis 565

6 Animal Models 565
6.1 Trisomic Mice 566
6.2 Transgenic Mice 568
6.3 Gene Expression Analyses 569

Bibliography 569
Books and Reviews 569
Primary Literature 570

Keywords

Acrocentric
A type of chromosome in which the centromere is very close to one end.

Autosomes
All of the chromosomes in the genome except for the sex chromosomes (X and Y).

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

Centromere
The structure within each chromosome at which the fibers required to move the chromosomes during meiosis or cell division (mitosis) attach.

Meiosis
The process within the germ cells during which genetic recombination occurs, and in which the number of chromosomes within the egg or sperm is reduced from the 46 found in somatic cells to 23.

Trisomy
The presence in the genome of three rather than two copies of a specific chromosome.

Down syndrome (DS) is the commonest of the genetically caused forms of mental retardation. It occurs with a frequency of approximately 1 per 800 to 1 per 1000 live births, and its incidence increases with increasing maternal age. DS is caused by the presence of an extra chromosome 21 within the genome, which, in turn, results in a 50% increase in the expression of the genes contained on the chromosome. By mechanisms currently undefined, the increased expression of several genes on human chromosome 21 results in a syndrome characterized by mental and growth retardation, a distinctive set of major and minor congenital malformations, a variety of cellular abnormalities, and, later in life, by the development of Alzheimer disease.

1 Phenotype of Down Syndrome

The most immediately apparent, if not the most serious, manifestations of Down syndrome (DS) are the many minor dysmorphic features that collectively constitute its distinctive physical phenotype. Salient among these are upslanting palpebral fissures, epicanthic folds, flat nasal bridge, brachycephaly, short broad hands, incurved fifth fingers, loose skin of the nape of the neck, open mouth with protruding tongue, and transverse palmar creases. Although any individual with DS will have many of the characteristic features and can be easily recognized as having the disorder, none of these features is present in all persons with DS. DS affects the nervous system in three principal ways: hypotonia, which occurs in virtually all newborns and infants; delayed psychomotor development in infancy and mental retardation throughout life; and neuronal degeneration during the adult years. The latter process, which is pathologically identical with Alzheimer disease (presenile and/or senile dementia), results in significant pathologic changes in the brain and may further compromise the already impaired mental functioning. DS is associated with two types of major congenital malformations. Most frequent (about 40%) is congenital heart disease, usually of the endocardial cushion type or one of its variants. Gastrointestinal tract abnormalities occur in

about 4.5% of individuals, more than half of whom have duodenal stenosis or atresia. Structural abnormalities of the thymus and functional defects in T-cell function leading to an increased susceptibility to infection are present, and there is a 10- to 18-times normal incidence of childhood leukemia with a frequent occurrence of acute megakaryoblastic leukemia.

2 Cytogenetics

Down syndrome is the phenotypic manifestation of trisomy 21. As such, it occurs when a third copy of chromosome 21 is present in the genome, either as a free chromosome or as part of a Robertsonian fusion chromosome (in which the long arms of two acrocentric chromosomes are joined at the centromeres). Except in 2 to 4% of cases in which mosaicism, with two populations of cells, one diploid and one trisomic, exists, all cells of the body are trisomic. Although these cytogenetic abnormalities involve most or all of chromosome 21, there are rare cases involving translocations in which only part of the long arm of the chromosome is triplicated. Depending on the region of the chromosome that is involved, such cases may or may not express the classical DS phenotype.

It has long been recognized that the risk of having a child with DS increases with maternal age and that the distribution of maternal age in the population of women having children is the primary determinant of the overall incidence of DS. A variety of estimates of the incidence of DS in the newborn population have been made, and most recent figures are in the vicinity of 1 per 1000. The figures can be broken down to provide maternal age–specific rates per 1000 live births: 0.6 at 20 years, 1.0 at 30 years, 2.6 at 35 years, 9.1 at 40 years, 24.9 at 44 years, and 41.2 at 46 years. Analyses using DNA markers have shown that maternal nondisjunction, the failure of paired chromosomes to separate properly, accounts for 92% of all cases of trisomy 21, 65% occurring at meiosis I, 23% at meiosis II, and 3% postzygotically (at mitosis). Eight percent of cases are paternal in origin, 3% at meiosis I and 5% at meiosis II. A reduced rate of recombination is associated with meiosis I errors and an increased rate with meiosis II errors.

3 Structure of Chromosome 21

Chromosome 21 is an acrocentric chromosome with a genetic length estimated to be 46 cM. It is the smallest of the human autosomes constituting approximately 1.7% of the length of the haploid genome. In physical terms, chromosome 21 is an acrocentric chromosome with the centromere very close to one end and with a very small short arm. The short arm (21p) terminates in a satellite region that may vary in size. Proximal to the satellite is the stalk (secondary constriction) that, as the nucleolar organizer region (NOR), contains multiple copies of the ribosomal RNA genes (*RNR4*) and stains characteristically with silver. The degree of silver staining appears to be a representation of the molecular activity of the ribosomal RNA genes that the chromosome contains. The region of 21p adjacent to the centromere contains highly repeated DNA sequences that consist of the satellite (including alphoid) and the "724" families of sequences. None of

these gene families is unique to chromosome 21. It is believed that these families of repeated gene sequences may be involved in the juxtaposition or association of the satellite regions (satellite association) of the acrocentric chromosomes during mitotic metaphase and with the formation of the nucleolus during interphase.

The major part of chromosome 21 is the long arm (21q) that has a characteristic banding pattern consisting of 3 or 4 bands at low resolution and as many as 11 dark and light bands resolvable by prometaphase banding. With one possible exception, all genes of known function (other than those for ribosomal RNA) are located on this arm of chromosome 21, and only this arm is essential for normal development and function. The presence of a Robertsonian fusion chromosome in which the short arms of two acrocentric chromosomes (sometimes both chromosomes 21) are deleted does not cause detectable abnormalities if the genome is otherwise balanced.

The sequence of the long arm of chromosome 21 was completed in 2000 with an estimated 99.7% coverage. This arm has a length of 33.7 Mbp and is relatively gene-poor, with less than half the number of identified genes found on the similarly sized chromosome 22. In the original annotation, it was judged to contain approximately 127 known genes, 98 predicted genes (of which 69% have no similarity to known proteins), and 59 pseudogenes. These numbers have been repeatedly revised as annotation of the sequence has proceeded, and the most recent figures are 178 confirmed and 36 predicted genes. There is evidence that not all expressed genes have been identified and that some predicted genes do not exist. Among the known genes are at least 10 kinases, 5 genes in the ubiquitination pathway, 5 cell adhesion molecules, 7 ion channels, 5 members of the interferon receptor family, and several transcription factors. About 22.4% of the chromosome consists of interspersed Alu sequences and LINE1 elements.

4 Phenotypic Mapping

Cases in which only part of chromosome 21 is triplicated have been intensively studied to arrive at a phenotypic map that will permit a correlation between particular phenotypic features of DS and specific regions or loci on the chromosome. The consensus of studies carried out before molecular markers became available is that the full DS phenotype, as manifested by mental retardation, congenital heart disease, characteristic facial appearance, hand anomalies, and dermatoglyphic changes, appears when band 21q22 is duplicated, and of this, subbands 21q22.1 and probably 21q22.2 are required. Molecular analysis has been used to define the extent of the triplication of regions of chromosome 21, and the results have been used to generate phenotypic maps. Although interpretations vary, it seems clear that many of the phenotypic features appear to be associated with imbalance in the region surrounding and distal to *D21S55*. However, there is also evidence for contributions of genes outside the *D21S55* region, especially in the proximal part of 21q, although the region responsible for congenital heart disease seems to be confined to the distal part of 21q. The genes that, when present in extra copies, contribute to impaired cognition appear to be located in several regions of the chromosome.

5
Pathogenesis

The immediate consequence of an aneuploid state is a gene dosage effect for each of the loci present on the unbalanced chromosome or chromosome segment. Such gene dosage effects have been reported for several chromosome 21 loci. With the exception of the amyloid precursor protein (APP) in the brain from fetuses with DS, and possibly a few other chromosome 21–encoded genes, the measured increase in activities or concentrations in trisomic cells is close to the theoretically expected value of 1.5. Taken in the aggregate, these results confirm the existence of quite precise dosage effects in cells aneuploid for chromosome 21. However, changes in the synthesis of proteins coded for by genes elsewhere in the genome have been observed, presumably as the result of genomic dysregulation resulting from the increased expression of chromosome w31 genes. However, this dysregulation is limited and certainly does not affect the vast majority of expressed genes.

On the basis of studies in animal models (see the following section), decreased platelet serotonin uptake and prostaglandin synthesis, and abnormal myoneural junctions, all of which occur in DS, have been ascribed to increased activity of CuZn-superoxide dismutase (SOD1). With these exceptions, it has not been possible to relate any component of the DS phenotype to overexpression of specific loci. However, a great amount of attention has been devoted to the pathogenesis of the cognitive deficits and Alzheimer disease in DS. The effects of increased expression of several of the loci on chromosome 21 have been considered, and it has been possible to construct plausible explanations for how overexpression of each locus might affect the development, function, and integrity of the central nervous system. A recent case in point is the gene for cystathionine synthase, overexpression of which in DS results in depressed levels of plasma homocysteine. It has been suggested that the altered homocysteine metabolism could compromise the folate-dependent resynthesis of methionine and thereby create a functional folate deficiency, the consequences of which on the central nervous system function could be many. However, in considering such possibilities, it is important to keep in mind that whereas imbalance of an individual locus could well have significant effects on the central nervous system, it is unlikely that it will provide a unitary explanation for the entire range of deficits found in DS. Nevertheless, it is possible that the development of Alzheimer disease in DS is related to the overproduction of APP, perhaps superimposed on an intrinsically defective nervous system.

6
Animal Models

The fact that many of the consequences of aneuploidy in humans arise during the period of morphogenesis and affect the development and function of the central nervous system places special stumbling blocks in the way of their investigation. Research on events occurring during gestation, especially early gestation, is both technically impractical and, currently and for the foreseeable future, ethically and legally impossible. Similarly, there are severe limitations to the study of the central nervous system function in living individuals, and experimental genetic alterations are impossible. It is for this reason that

interest has turned to the development of animal models for DS.

Since ultimate concern is with humans and with human disease, we require models that will duplicate the human condition – in developmental and functional terms – as closely as possible. No model can be an exact one, since no other organism duplicates the human with respect to all of his or her biologic and genetic attributes. Nevertheless, models based on other mammals, which share numerous biologic similarities with humans, seem most appropriate. Primates with the genetic and clinical equivalent of trisomy 21 and DS have been observed but do not offer significant advantages over humans for investigational purposes. Therefore, the mouse has been the model animal of choice for several reasons, including ease of manipulation and genetic control and the similarities to humans in the processes of morphogenesis and probably of central nervous system function, in neurobiological if not psychological terms. Furthermore, despite considerable rearrangement of the mammalian genome, sizable chromosomal regions carrying many genes remain intact and structurally similar in both humans and mice.

Models based on the mouse are not without their problems. It may not always be easy or even possible to obtain postnatally viable animals when sizable regions of the genome are unbalanced. Furthermore, even if it were possible, these animals will never be able to duplicate the higher central nervous system functions, such as cognition, which are so vulnerable to the effects of aneuploidy in humans. As difficult as it is to define mental retardation in an aneuploid human, it is even more difficult to specify the proper functional homology in an aneuploid mouse. Nevertheless, elementary forms of learning are common to all animals with an evolved nervous system, and there appears to be a conservation of the relevant biochemical mechanisms. Therefore, the mouse can be considered to be a legitimate model for learning and memory in humans.

6.1 Trisomic Mice

The modeling of human trisomy 21 in the mouse started at the two ends of the possible spectrum of such models. On one end is the trisomic mouse with an extra copy of mouse chromosome 16, the mouse chromosome that most closely resembles human chromosome 21. At the other extreme is the transgenic mouse with an extra copy or copies of a single human chromosome 21 gene. More recently, however, several intermediate possibilities, with duplications of only parts of mouse chromosome 16 or the insertion of mouse or human yeast artificial chromosome (YAC) transgenes, have been developed.

With perhaps one exception, all the loci in human 21q from the centromere to the *MX1* locus in the distal part of band 21q22.3, approximately 130 to 150 in number, map to the distal part of mouse chromosome 16 (see http://www.informatics.jax.org). The remaining loci, not on mouse chromosome 16, have been mapped to mouse chromosomes 10 and 17. These loci are outside the region shown by phenotypic mapping to be responsible for the facial features of DS. In addition to coding regions, there are also a large number of sequences conserved between the human and mouse genomes that appear to have a functional significance but cannot be identified as protein-coding or RNA genes.

Because of the genetic discrepancies between mouse chromosome 16 and human chromosome 21, trisomy for all of mouse chromosome16, while reproducing much of the genetic imbalance associated with human trisomy 21, results in an overall degree of imbalance that is more extensive than in human trisomy and in an overall phenotype that is more severe. For this reason, the results of studies with mouse trisomy 16 (Ts16) must be viewed with concern insofar as modeling DS is concerned. To correct this situation, segmentally trisomic mice (designated Ts65Dn) in which only the human chromosome 21 orthologous region is triplicated were produced. In these animals, approximately 115 of the original number of 225 genes on human chromosome 21q are represented. The trisomic animals do not have the major anomalies present in Ts16. Rather, their salient phenotypic abnormalities are confined to the craniofacial structures and the central nervous system, although they do exhibit male sterility. As in persons with DS, there is a reduction in the size and granule cell density of the cerebellum, a reduced sensitivity to pain, elevation of brain *myo*-inositol, and abnormalities of the dendritic spines. Although motor learning is normal, Ts65Dn mice have major deficits in learning and behavior, and show abnormalities in long-term potentiation (LTP) and depression and in hippocampal cAMP generation. The pattern of behavioral and learning deficits suggests impairment of both hippocampal spatial learning and of prefrontal functions, and the LTP and cAMP abnormalities are also indicative of hippocampal dysfunction. Further evidence for abnormality of the hippocampus of the Ts65Dn mouse is provided by the presence of giant dendritic spines with aberrant connections in hippocampal neurons and by suppression of LTP by inhibitory inputs. The finding in Ts65Dn mice of an age-dependent loss of function and atrophy of cholinergic neurons in the medial septum of the basal forebrain, which correlates with the loss of cognitive function and appears to be related to an abnormality of nerve growth factor transport, is also consistent with hippocampal dysfunction. It is of note, in this regard, that a reduction in Trk-A immunoreactive cholinergic neurons has been observed in the nucleus basalis of persons with DS.

A second segmentally trisomic mouse, Ts1Cje, is functionally trisomic for the region of mouse chromosome 16 distal to *Sod1*. This trisomy involves about 87 chromosome 21–orthologous genes and still includes the region responsible for the major phenotypic abnormalities in DS. Their learning deficits are similar to but somewhat less severe than those of the Ts65Dn animals. Similarly, there is a reduction in the volume of the cerebellum, but granule cell density is much less affected. However, Ts1Cje animals do not appear to exhibit degeneration of the cholinergic neurons, suggesting that the gene(s) responsible for this phenotype are located in the region of difference between the two segmental trisomies, a region containing about 30 chromosome 21–orthologous genes. A third segmentally trisomic mouse, designated Ms1Ts65, carries a triplication of this region of difference. These animals have only minimal learning deficits, but they are still not normal.

Taken together with the comparison of Ts65Dn and Ts1Cje, the results with Ms1Ts65 indicate that the major genes responsible for the behavioral and learning abnormalities of Ts65Dn mice are located distal to *Sod1*. These studies constitute the first step in phenotypic mapping of the segmental-trisomy DS models

by a subtractive approach in which a known phenotype (as in Ts65Dn) is compared with the phenotype obtained after reduction of the region of trisomy (as in Ts1Cje). The same approach has been used to compare the alterations in craniofacial morphogenesis produced by Ts65Dn and Ts1Cje. Both trisomics demonstrate alterations in the formation of the craniofacial skeleton that are similar to what is observed in DS, with the changes in Ts1Cje being largely the same as those in Ts65Dn. This again demonstrates that the major genes responsible for these skeletal abnormalities in the Ts65Dn mice are located distal to *Sod1*. The ability to generate these various segmental trisomics and to compare their phenotypes directly provides a precedent for the systematic creation of smaller and smaller segmental trisomics and the dissection of the trisomic phenotype.

6.2 Transgenic Mice

In contrast with the formation of trisomic animals, an approach that permits the analysis of the effects of increased dosage of individual human chromosome 21 genes is the construction of transgenic mice. Several strains of such transgenic mice carrying *SOD1*, the gene for human CuZnSOD, have been made and shown to have from 1.6- to 6-fold increased activities of CuZnSOD in the brain and other tissues. Several effects of this increased activity of CuZnSOD that resemble changes found in persons with DS have been found. Transgenic mouse platelets have diminished levels of serotonin and decreased rates of serotonin accumulation that are attributed to a diminished pH gradient across the granule membrane. There is a reduction in prostaglandin E_2 and prostaglandin D_2 biosynthesis in transgenic primary fetal cells and in the cerebellum and hippocampus. The neuromuscular junctions of the transgenic tongue muscles are abnormal, with atrophy, degeneration, withdrawal and destruction of terminal axons, development of multiple small terminals, decreased ratio of terminal axon area to postsynaptic membrane, and hyperplastic secondary folds. Transgenic mice also have anatomical abnormalities of the thymus and premature thymic involution. Impairment of LTP has also been reported.

Many of the abnormalities in transgenic tissues just discussed are presumed to be mediated by alterations in the metabolism of oxygen free radicals resulting from the increased CuZnSOD activity (this enzyme mediates the first step in the ultimate conversion of superoxide anions to water).

Limited work has been done with other types of transgenic mice carrying individual human chromosome 21 genes or their murine homologues. Although they do not develop the full pathology of AD, mice expressing the human APP-751 isoform gene do have diffuse and amorphous extracellular deposits of β-amyloid that resemble those of early AD and manifest deficits in spatial learning. Transgenic mice overexpressing the Ets2 transcription factor develop craniofacial and other skeletal anomalies, with brachycephaly and short necks, thought to resemble anomalies found in DS and mouse Ts16. However, similar anomalies are not found in any of the segmental trisomics of chromosome 16, even though *Ets2* is present in three copies. Overexpression of *HMG14* results in abnormalities of the thymus and epithelial cysts, and mice transgenic for the mouse single-minded gene, *mSim2*, have impaired spatial learning.

The transgenes just described involve loci present on mouse chromosome 16 as

well as human chromosome 21. Therefore, the absence of similar findings in trisomic mice involving the same regions is of concern. The following two loci, however, are represented on mouse chromosome 10, and the trisomy 16 models are, therefore, not relevant. Transgenic mice overexpressing the gene for the protein subunit S100β were reported to be hyperactive, to have learning impairments, and to manifest astrocytosis, axonal proliferation, and a significant loss of dendrites in the hippocampus. Mice overexpressing liver-type phosphofructokinase had elevated activity during fetal life but not during adult life.

Another approach to the generation of transgenic models for DS has been the insertion of genomic segments larger than individual genes in the form of yeast artificial chromosomes. A mouse produced in this manner had abnormalities of learning and behavior too, but less severe than that observed with the segmental trisomies, Ts1Cje and Ts65Dn. Since the major and possibly only gene in the integrated segment of human DNA was the human minibrain gene (*DYRK*), it was suggested that imbalance of this locus is responsible for the observed abnormalities. Support for this inference is provided by mice transgenic for the *Dyrk1A* gene. These animals display delayed neuromotor development and impairment in spatial learning with defective reference memory, findings considered indicative of hippocampal and prefrontal cortex dysfunction.

6.3 Gene Expression Analyses

The cloning and sequencing of both the mouse and human genomes has made it possible to undertake *in situ* hybridization studies of the expression of chromosome 21–orthologous genes in diploid mouse embryos, fetuses, and postnatal animals. Broad expression of the human chromosome 21 orthologues is present in early embryos and becomes more restricted as development proceeds. The number of genes expressed and their patterns of expression differ among tissues, but it is possible to identify several clusters of coexpressing genes. By correlating expression patterns with specific features of the DS phenotype, candidate genes have been proposed for congenital heart and intestinal defects.

See also Brain Development; Genetics, Molecular Basis of; Medicinal Chemistry.

Bibliography

Books and Reviews

Antonarakis, S.E., Lyle, R., Chrast, R., Scott, H.S. (2001) Differential gene expression studies to explore the molecular pathophysiology of Down syndrome, *Brain Res. Rev.* **36**, 265–274.

Capone, G.T. (2001) Down syndrome: advances in molecular biology and the neurosciences, *Dev. Behav. Pediatr.* **22**, 40–59.

Dierssen, M., Fillat, C., Crnic, L., Arbonés, M., Flórez, J., Estivill, X. (2001) Murine models for Down syndrome, *Physiol. Behav.* **73**, 859–871.

Epstein, C.J. (1986) *The Consequences of Chromosome Imbalance. Principles, Mechanisms, and Models*, Cambridge University Press, New York.

Epstein, C.J. (Ed.) (1993) *The Phenotypic Mapping of Down Syndrome and Other Aneuploid Conditions*, Wiley-Liss, New York.

Epstein, C.J. (2001) Down Syndrome (trisomy 21), in: Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D. (Eds.) *The Metabolic and Molecular Bases of Inherited Disease*, 8th edition, McGraw-Hill, New York, pp. 1123–1256.

Epstein, C.J., Hassold, T., Lott, I.T., Nadel, L., Patterson, D. (Eds.) (1995) *Etiology and Pathogenesis of Down Syndrome*, Wiley-Liss, New York.

Galdzicki, Z., Siarey, R., Pearce, R., Stoll, J., Rapoport, S.I. (2001) On the cause of mental retardation in Down syndrome: extrapolation from full and segmental trisomy 16 models, *Brain Res. Rev.* **35**, 115–145.

Hassold, T., Hunt, P. (2001) To err (meiotically) is human: the genesis of human aneuploidy, *Nat. Rev. Genet.* **2**, 280–291.

Pueschel, S.M., Rynders, J.E. (Eds.) (1982) *Down Syndrome. Advances in Biomedicine and the Behavioral Sciences*, Ware, Cambridge, MA.

Reeves, R.H., Baxter L.L., Richtsmeier, J.T. (2001) Too much of a good thing: mechanisms of gene action in Down syndrome, *Trends Genet.* **17**, 83–88.

Primary Literature

Altafaj, X., Dierssen, M., Baamonde, C., Marti, E., Visa, J., Guimerà, J., Oset, M., González, J.R., Flórez, J., Fillat, C., Estivill, X. (2001) Neurodevelopmental delay, motor abnormalities and cognitive deficits in transgenic mice overexpressing *Dyrk1A* (*minibrain*), a murine model of Down's syndrome, *Hum. Mol. Genet.* **10**, 1915–1923.

Chrast, R., Scott, H.S., Papasavvas, M.P., Rossier, C., Antonarakis, E.S., Barras, C., Davisson, M.T., Schmidt, C., Estivill, X., Dierssen, M., Pritchard, M., Antonarakis, S.E. (2000) The mouse brain transcriptome by SAGE: differences in gene expression between P30 brains of the partial trisomy 16 mouse model of Down syndrome (Ts65Dn) and normals, *Genome Res.* **10**, 2006–2021.

Cooper, J., Salehi, A., Delcroix, J.-D., Howe, C.L., Belichenko, P.V., Chua-Couzens, J., Kilbridge, J.K., Carlson, E.J., Epstein, C.J., Mobley, W.C. (2001) Failed retrograde transport of NGF in a mouse model of Down's syndrome: reversal of cholinergic neurodegenerative phenotypes following NGF infusion, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10439–10444.

Dermitzakis, E.T., Reymond, A., Lyle, R., Scamuffa, N., Ucla, C., Deutsch, S., Stevenson, B.J., Flegel, V., Bucher, P., Jongeneel, C.V., Antonarakis, S.E. (2002) Human chromosome 21 gene expression atlas in the mouse, *Nature* **420**, 578–582.

Ema, M., Ikegami, S., Hosoya, T., Mimura, J., Ohtani, H., Nakao, K., Inokuchi, K., Katsuki. M., Fujii-Kuriyama, Y. (1999) Mild impairment of learning and memory in mice overexpressing the mSim2 gene located on chromosome 16: an animal model of Down's syndrome, *Hum. Mol. Genet.* **8**, 1409–1415.

Engidawork, E., Lubec, G. (2001) Protein expression in Down syndrome brain, *Amino Acids* **21**, 331–361.

Gitton, Y., Dahmanane, N., Balk, S., Ruiz, I., Altaba, A., Neidhardt, L., Scholtze, M., Herrmann, B.G., Kahlem, P., Benkahla, A., Schrinner, S., Yildirimman, R., Herwig, R., Lehrach, H., Yaspo, M.-L. (2002) A gene expression map of human chromosome 21 orthologues in the mouse, *Nature* **420**, 586–600.

Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.-S., Toyoda, A. et al. (2000) The DNA sequence of human chromosome 21, *Nature* **405**, 311–319.

Holtzman, D.M., Bayney, R.M., Li, Y.W., Khosrovi, H., Berger, C.N., Epstein, C.J., Mobley, W.C. (1992) Dysregulation of gene expression in mouse trisomy 16, an animal model of Down syndrome, *EMBO J.* **11**, 619–627.

Hyde, L.A., Frisone, D.F., Crnic, L.S. (2001) Ts65Dn mice, a model for Down syndrome, have deficits in context discrimination learning suggesting impaired hippocampal function, *Behav. Brain Res.* **118**, 53–60.

Korenberg, J.R., Chen, X., Schipper, R., Sun, Z., Gonsky, R., Gerwehr, S., Carpenter, N., Daumer, C., Dignan, P., Disteche, C., Graha, J.M. Jr., Hudgins, L., McGillivray, B., Miyazaki, K., Ogasawara, N., Park, J.P., Pagon, R., Pueschel, S., Sack, G., Say, B., Schuffenhauer, S., Soukup, S., Yamanaka, T. (1994) Down syndrome phenotypes: the consequences of chromosome imbalance, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4997–5001.

Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P.A., Gingeras, T.R. (2002) Large-scale transcriptional activity in chromosomes 21 and 22, *Science* **296**, 916–919.

Lamb, N.E., Feingold, E., Savage, A., Avramopoulos, D., Freeman, S., Gu, Y., Hallberg, A., Hersey, J., Karadima, G., Pettay, D., Saker, D., Shen, J., Taft, L., Mikkelsen, M., Petersen, M.B., Hassold, T., Sherman, S.L. (1997) Characterization of susceptible chiasma configurations that increase the risk for maternal

nondisjunction of chromosome 21, *Hum. Mol. Genet.* **6**, 1391–1399.

Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome, *Nature* **420**, 520–562.

Mural, R.J., Adams, M.D., Myers, E.W., Smith, H.O., Miklos, G.L.G., Wides, R. et al. (2002) A comparison of whole-genome shotgun-derived mouse chromosome 116 and the human genome, *Science* **296**, 1661–1671.

Pinter, J.D., Eliez, S., Schmitt, J.E., Capone, G.T., Reiss, A.L. (2001) Neuroanatomy of Down's syndrome: a high resolution MRI study, *Am. J. Psychiatry* **158**, 1659–1665.

Pletcher, M.T., Wiltshire, T., Cabin, D.E., Villanueva, M., Reeves, R.H. (2001) Use of comparative physical and sequence mapping to annotate mouse chromosome 16 and human chromosome 21, *Genomics* **74**, 45–54.

Pogribna, M., Melnyk, S., Pogribny, I., Chango, A., Yi, P., James, S.J. (2001) Homocysteine metabolism in children with Down syndrome: in vitro modulation, *Am. J. Hum. Genet.* **69**, 88–95.

Reeves, R.H., Irving, N.G., Moran, T.H., Wohn, A., Kitt, C., Sisodia, S.S., Schmidt, C., Bronson, R.T., Davisson, M.T. (1995) A mouse model for Down syndrome exhibits learning and behaviour deficits, *Nat. Genet.* **11**, 177–184.

Reymond, A., Camargo, A.A., Deutsch, S., Stevenson, B.J., Parmigiani, R.B., Ucla, C., Bettoni, F., Rossier, C., Lyle, R., Guipponi, M., de Souza, S., Iseli, C., Jongeneel, C.V., Bucher, P., Simpson, A.J., Antonarakis, S.E. (2002) Nineteen additional unpredicted transcripts from human chromosome 21, *Genomics* **79**, 824–832.

Reymond, A., Friedli, M., Henrichsen, C.N., Chapot, F., Deutsch, S., Ucla, C., Rossier, C., Lyle, R., Guipponi, M., Antonarakis, S.E. (2001) From PREDs and open reading frames to cDNA isolation: revisiting the human chromosome 21 transcription map, *Genomics* **78**, 46–54.

Reymond, A., Marigo, V., Yaylaoglu, M.B., Leoni, S., Ucla, C., Scamuffa, N., Cacciopoli, C., Dermitzakis, E.T., Lyle, R., Banfi, S., Eichele, G., Antonarakis, S.E., Ballabio, A. (2002) Human chromosome 21 gene expression atlas in the mouse, *Nature* **420**, 582–586.

Richtsmeier, J.T., Zumwalt, A., Carlson, E.J., Epstein, C.J., Reeves, R.H. (2002) Craniofacial anomalies in segmentally trisomic mouse models for Down syndrome, *Am. J. Med. Genet.* **107**, 317–324.

Sago, H., Carlson, E.J., Smith, D.J., Rubin, E.M., Crnic, L.S., Huang, T.-T., Epstein, C.J. (2000) Genetic dissection of the region associated with behavioral abnormalities in mouse models of Down syndrome, *Pediatr. Res.* **48**, 606–613.

Sinet, P.M., Théophile, D., Rahmani, Z., Chettouh, Z., Blouin, J.L., Prieur, M., Noel, B., Delabar, J.M. (1994) Mapping of the Down syndrome phenotype on chromosome 21 at the molecular level, *Biomed. Pharmacother.* **48**, 247–252.

Weitzdoerfer, R., Fountoulakis, M. (2002) Reduction of actin-related protein complex 2/3 in fetal Down syndrome brain, *Biochem. Biophys. Res. Commun.* **293**, 836–841.

Drosophila Genome

Robert D.C. Saunders
Department of Biological Sciences, The Open University, Milton Keynes, UK

1 Introduction 575

2 An Overview of the *Drosophila* Genome 576

3 The Chromosomes of *Drosophila* 576
3.1 Mitotic Chromosomes 576
3.2 Polytene Chromosomes 577
3.3 Chromosome Aberrations 579
3.3.1 Using Polytene Chromosomes to Map Genes 580

4 The *Drosophila* Genome Project 581
4.1 Euchromatin and the Genes 581
4.1.1 Duplicated Genes, Pseudogenes 581
4.2 Transposable Elements 582
4.3 Heterochromatin 584
4.4 Long-range Genome Structure 584

5 Genetic Analysis in *Drosophila* 585
5.1 Mutagenesis 585
5.1.1 Reverse Genetics 586
5.1.2 Classes of Mutant Allele 586
5.1.3 Position Effect Variegation 587
5.1.4 Interfering RNA 587
5.1.5 Targeted Mutagenesis 587
5.2 Genetic Mapping 587
5.2.1 Microsatellite and RFLP Markers 588

6 Prospects 589

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

Bibliography 589
Books and Reviews 589
Primary Literature 589

Keywords

Polytene Chromosome
Highly replicated chromosomes in an interphase-like chromatin conformation, with each chromatid precisely synapsed with its sisters. This results in a banded structure, permitting high-resolution mapping by light microscopy.

Heterochromatin
A term describing those regions of the chromosome that are highly condensed in metaphase chromosomes and undergo late replication in S-phase. Heterochromatin generally corresponds with those regions of the genome containing simple sequence repetitive, or satellite, DNA. There are few genes in heterochromatic regions.

Euchromatin
Euchromatin corresponds to the single copy and middle repetitive DNA of the genome, in which most of the genes are located. Unlike heterochromatin, euchromatin does polytenize.

In situ Hybridization
Cloned sections of DNA can be accurately mapped to polytene chromosomes by *in situ* hybridization, a process in which chemically or radioactively labeled nucleic acid is hybridized to denatured chromosomes on a microscope slide. Signals are detected immunochemically or by application of a radioactive emulsion.

Balanced Lethal Stocks
Recessive lethal stocks are maintained as balanced stocks, in which the population consists of flies heterozygous for the mutant chromosome and for a corresponding balancer chromosome. Balancer chromosomes possess multiple inversions, which suppress recombination, and several genetic markers, including a dominant visible, and are themselves recessive lethal or sterile, so that balancer homozygotes are either inviable or sterile.

Drosophila melanogaster is arguably one of the most important model organisms available to the modern researcher. This chapter seeks to review the structure of the *Drosophila* genome and the main genetic and molecular techniques currently in use. The *Drosophila* genome has recently been sequenced, and this chapter sets the biology of *Drosophila* in the context of recent updates to the sequence annotation.

1 Introduction

Drosophila melanogaster is arguably the genetically best-characterized metazoan organism. Since its adoption for genetic analysis, its many advantages as a laboratory animal have led to its use as a model system for a wide variety of biomedical researches. The *Drosophila* genome is relatively small, approximately 170×10^6 bp, and is distributed between four pairs of chromosomes. The polytene chromosomes of *Drosophila* have provided a physical map for genetic studies since they were mapped in the 1930s, and continue to be a powerful tool in molecular genetic analysis.

This article describes the features of the *Drosophila* genome relevant to genetic and molecular studies in *Drosophila*.

In the early years of this century, *D. melanogaster* (then known as *Drosophila ampelophila*) was adopted by Thomas Hunt Morgan as an experimental organism in the then new science of genetics. *Drosophila* proved to be an ideal organism for genetic research: it has a short life cycle, 10 to 14 days at 25 °C, and very modest requirements in terms of husbandry. In fact, a modern *Drosophila* laboratory may maintain a thousand or more genetically distinct stocks. Many of the basic concepts of genetics were established using the *Drosophila* species, including important advances in population genetics and evolutionary theory, particularly with *Drosophila pseudoobscura* and related species. However, it is *D. melanogaster* that has made the biggest impact on modern molecular genetic research. With the advent of molecular cloning technology, *Drosophila* found itself ideally placed to benefit from integrated genetics and molecular biology.

As a model organism, *Drosophila* has been an essential tool for the elucidation of basic genetic processes, common to many eukaryotes.

1. For many areas of biomedical research: Examples include aging, neurobiology, embryology, pattern formation, signal transduction, and cell cycle regulation. In particular, research in these areas is facilitated by the sophisticated genetic analysis possible with *Drosophila*.
2. For genome mapping and analysis: *D. melanogaster* has a moderately sized genome, 170×10^6 bp, in comparison with those of yeasts, nematode worms, the mouse, humans, and other species in which large-scale molecular mapping is carried out. Table 1 shows the haploid genome sizes of these species. Thus, the *Drosophila* genome is ideal for use as a model for genome mapping in higher eukaryotes, and additionally has features such as polytene chromosomes, which simplify mapping.
3. Finally, many medically and economically important pests and disease vectors are members of the order Diptera. Examples are the anopheline mosquitoes (malaria vectors), aedine mosquitoes (vectors of yellow and dengue fevers), and the tsetse flies (vector of trypanosomiasis, or sleeping

Tab. 1 The haploid genome sizes of some species.

Organism	*Genome size [$\times 10^6$ bp]*
Escherichia coli	4
Saccharomyces cerevisiae	16
Schizosaccharomyces pombe	15
Caenorhabditis elegans	100
Drosophila melanogaster	170
Homo sapiens, Mus musculus	3,000

sickness) to name but a few. Several features of the biology and life cycle of these insects mean they are not good research organisms, from the point of view of laboratory culture and genetics. *Drosophila* is a model system for the molecular genetic analysis of these insects.

In this article, I shall review the genetics and molecular biology of *D. melanogaster*, with respect to its use as a model system.

2 An Overview of the *Drosophila* Genome

The haploid genome of *D. melanogaster* is relatively small, 170×10^6 bp, compared with the typical mammalian genome of approximately 3×10^9 bp. *Drosophila* genomic DNA is unmethylated, in contrast to most other higher eukaryotes. The composition of the *Drosophila* genome in terms of single copy and repetitive DNA fractions is presented in Table 2.

About 21% of the genome comprises highly repetitive simple sequence DNA, known as *satellite DNA*. Satellite DNA is present in large blocks of repeated short (5 to 10 bp) units principally located close to centromeric regions. There are several classes of satellite, distinguishable by the repeat unit sequence, and these classes often have chromosome-specific distribution. One satellite, the 1.688 or 359 bp satellite, is atypical, comprising arrays of repeated 359 bp units confined to the X chromosome. In general, satellite DNA is concentrated in heterochromatin, which will be discussed below.

Tab. 2 The composition of the *Drosophila melanogaster* genome.

Single copy sequences	61%	103.7×10^6
Satellite DNA	21%	35.7×10^6
rDNA, histones, etc.	3%	5.1×10^6
Transposable elements	9%	15.3×10^6
Foldback DNA	6%	10.2×10^6
Total genome size	–	170×10^6

A substantial proportion, 18%, of the genome is moderately repetitive. Some of this repetition is due to the ribosomal RNA and histone gene families, but much is accounted for by mobile DNA, or transposable elements. Transposable elements are found in variable, dispersed locations within the genome. Several transposable elements have proven to be very useful in the experimental manipulation and analysis of *Drosophila*, particularly the *P* element, which is described in detail in Sect. 4.2.

The composition of the *Drosophila* genome has, of course, been largely redefined by the successes of the *Drosophila* genome projects, and in particular that of the Berkeley Drosophila Genome Project (BDGP). This will be discussed in some detail below.

3 The Chromosomes of *Drosophila*

3.1 Mitotic Chromosomes

The mitotic chromosome complement of *D. melanogaster* is illustrated diagrammatically in Fig. 1. There are two metacentric autosomes, chromosomes 2 and 3; one tiny dotlike autosome, chromosome 4; and the heteromorphic sex chromosomes, the acrocentric X chromosome and the submetacentric Y chromosome. Sex determination in *D. melanogaster* is based on the ratio of X chromosomes to autosomes. XX individuals have a ratio of 1, and develop as

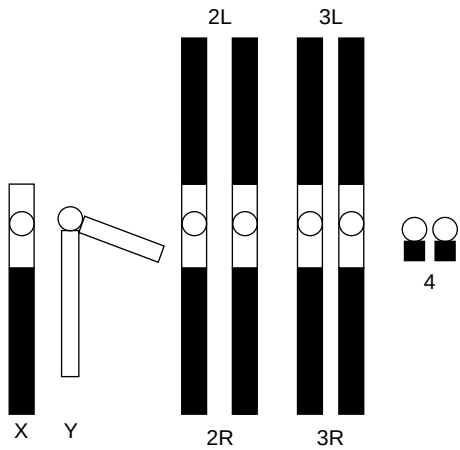

Fig. 1 Diagrammatic representation of the diploid complement of mitotic chromosomes of *D. melanogaster*. The karyotype shown is XY (male). Circles and white boxes represent centromeres and heterochromatin respectively.

females, while XY individuals have a ratio of 0.5, and develop as males. Thus, male *Drosophila* are hemizygous for X-linked loci. XO individuals are male, but sterile, since the Y chromosome carries factors essential for male fertility.

Chromatin can be subdivided on microscopic criteria into heterochromatin and euchromatin, with heterochromatin being characterized by having a highly condensed state and late replication in S-phase. Modern molecular genetic research has shown that, in general, the genes lie in the euchromatin, while the heterochromatin is composed principally of simple sequence repetitive DNA. Indeed, the term heterochromatin is often used as a synonym for satellite DNA. There are exceptions to this rule, including the male fertility factors of the Y chromosome mentioned above, and a few autosomal loci.

3.2 Polytene Chromosomes

In common with many species of Diptera, *Drosophila* possesses polytene chromosomes in several tissues, most notably the larval salivary glands. The salivary glands of third instar larvae are easy to identify and dissect, and their chromosomes are simple to prepare for analysis. Polytene chromosomes are very large, highly replicated chromosomes that are invaluable in genetic research using *Drosophila* (and many other Diptera) as a model.

Polytene chromosomes are found in many species of Diptera including *Drosophila*, where the easiest to work with are found in the salivary glands of the third instar larva. The polytene chromosomes of *Drosophila* represent the euchromatic portion of the genome, in a highly amplified condition, in which each chromatid is replicated approximately 1000-fold, and each copy is precisely aligned, resulting in a highly reproducible transverse banding pattern along the chromosomes (Fig. 2). The pericentromeric heterochromatin, and the heterochromatic Y chromosome remain unpolytenized and are located together with the centromeres in the chromocenter, to which each of the chromosome arms are attached. Thus, the polytene chromosome complement is roughly star shaped, with each major chromosome arm of approximately equal length radiating from the chromocenter.

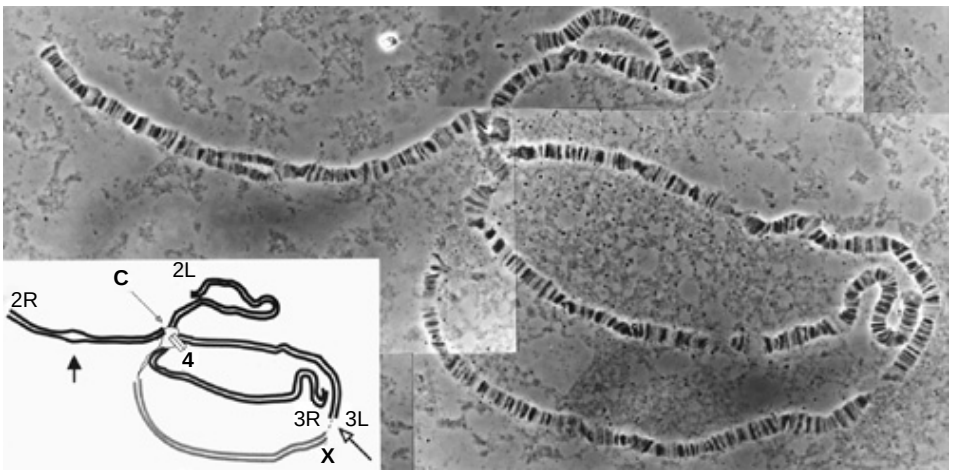

Fig. 2 The polytene chromosomes of *D. melanogaster*. Phase contrast micrograph of a preparation of *D. melanogaster* polytene chromosomes. The specimen has been fixed in 17% lactic acid, 50% acetic acid, and is unstained. Each chromosome arm extends from the chromocenter. Inset: diagrammatic representation of the photograph. Each chromosome arm is labeled. The X chromosome is shaded in light gray, chromosome 2 in dark gray, and chromosome 3 in black, with chromosome 4 in white. **C**: chromocenter, **X**: X chromosome, 2L, 2R: left and right arms of chromosome 2, 3L, 3R: left and right arms of chromosome 3, **4**: chromosome 4. The solid arrow points to a region of asynapsis in arm 2L. The open arrow points to ectopic pairing between the telomeres of 3L and X.

The arm corresponding to chromosome 4 is about 5% of the length of the others. This arrangement is shown in Fig. 2, with a diagrammatic interpretation in the inset.

In the 1930s, the importance of polytene chromosomes for genetic research was realized and highly detailed cytogenetic maps were created on the basis of the banding pattern of the chromosomes. The map devised by Calvin Bridges in the mid-1930s is basically that used to this day, and provides a reference system of map locations understandable by all researchers possessing a copy of the map. Each major arm is divided into 20 divisions numbered 1 to 20 (X chromosome), 21 to 60 (second chromosome), and 61 to 100 (third chromosome). The tiny fourth chromosome is allocated two divisions, 101 and 102. The telomeric divisions are 1, 21, 60, 61, 100, and 102. The numbered divisions are each further broken down into six subdivisions, lettered A to F. This basic map was further refined by Bridges and Bridges to include a number for each band on the map. Thus, each band of the polytene chromosome complement can be unambiguously identified (at least in principle).

Genes identified by mutation can be accurately mapped to the polytene chromosomes using a huge collection of chromosome rearrangements, such as deficiencies, inversions, duplications, transpositions, and translocations. At the simplest level, this involves determining whether a mutant allele is located within a given deficiency (see Sect. 3.3.1).

The resolution afforded by the use of polytene chromosomes is very high;

features of 5 to 100 kb can be recognized on the map by light microscopic analysis, which can rapidly provide positional information at a precision useful by molecular cloning standards. *In situ* hybridization can map cloned segments of DNA to single band resolution, several orders of magnitude better than that easily possible with metaphase chromosomes, and this technique is consequently invaluable for many gene cloning strategies and physical genome mapping projects.

3.3 Chromosome Aberrations

A great many chromosome aberrations, affecting all chromosomes have been recovered and analyzed. The gene and chromosomal aberration data sets of the *Drosophila* database FlyBase contain details of many thousands of chromosomal aberrations. These include the following:

a. *Deficiencies* (Fig. 3). Deficiency is the terminology used in *Drosophila* research for a chromosomal deletion. In a deficiency chromosome, a segment of the chromosome is deleted. In practice, the largest deficiency of practical utility is about one polytene map division. Flies carrying deficiencies larger than this have severely reduced fitness and viability. Furthermore, there are a number of haploinsufficient loci in the genome, which are inviable when hemizygous (e.g. when uncovered by a deficiency). Telomeric deficiencies can be unstable, unless the specialized telomeric sequences have been replaced.

b. *Inversions* (Fig. 3). Inversion chromosomes are those in which a section has been inverted with respect to the wild-type gene order. Inversion and normal chromosomes synapse to form inversion loops, in both meiotic and polytene chromosomes. Chromosomes bearing inversions suppress recombination within the inverted region when heterozygous with wild-type sequence chromosomes. *Balancer chromosomes* are chromosomes carrying several inversions, together with

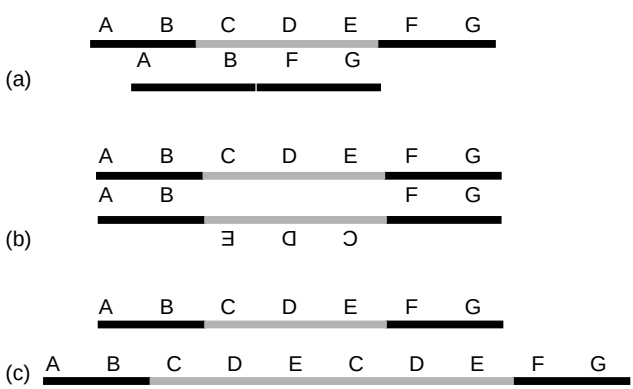

Fig. 3 Chromosome rearrangements. In each panel, a standard order chromosome bearing loci A–G is shown above the rearranged chromosome. The region involved in the rearrangement is shaded. (a) Deficiency. Loci C, D, and E have been deleted. (b) Inversion. Loci C, D, and E have been inverted. (c) Duplication. Loci D, D, and E have been duplicated in tandem.

suitable dominant and recessive markers, and are used to effectively suppress recombination along the whole chromosome. The use of balancer chromosomes is described in Sect. 5.1. Chromosome inversions that contain the centromere are referred to as *pericentric inversions* and those that do not are *paracentric inversions*.

c. *Duplications.* (Fig. 3). These chromosomes carry an additional segment derived from a homologous or other chromosome.
d. *Translocations.* Translocations involve at least two chromosome breaks, and involve the reciprocal interchange of chromosome sections between two or more nonhomologous chromosomes. More than two chromosomes can be rearranged.
e. *Transpositions.* Essentially, a transposition is where a chromosome section is inserted in another chromosome, involving three breaks.
f. *Compound chromosomes.* These chromosomes are derived from the normal chromosome complement by breakage and fusion of chromosome arms. Compound X chromosomes (two X chromosomes sharing a single centromere) affect the pattern of inheritance of normal X and Y chromosomes: in flies of normal karyotype, Y chromosomes are patroclinous (inherited from the father), while in a typical compound X stock, they are matroclinous (inherited from the mother). Compound chromosomes involving the attachment of X and Y chromosomes, whole autosome fusions, and whole autosome arm fusions can be constructed.

The genetic technologies for constructing novel aberrations and compound chromosomes are well established for *D. melanogaster*. Translocations can be combined to create segmental aneuploids, synthetic deficiencies or duplications, when necessitated by the lack of conventional deficiencies for a particular chromosome region. Recombination between inversion chromosomes can also be used to synthesize deficiencies and duplications.

3.3.1 Using Polytene Chromosomes to Map Genes

The principle of deficiency mapping a recessive mutant allele is to test whether a deficiency chromosome "uncovers" the mutation by examining the phenotype of a stock carrying both the deficiency chromosome and the recessive allele-bearing chromosome. If the mutant phenotype is expressed, the locus must lie within the region deleted in the deficiency chromosome (Fig. 4). In addition to simple

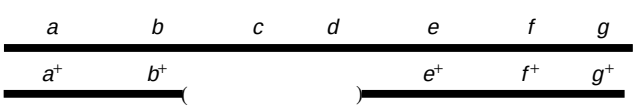

Fig. 4 Deficiency mapping. This figure illustrates the principle of deficiency mapping. A chromosome bearing recessive mutation at several loci (*a* to *g*) is heterozygous with a deficiency chromosome. The region deleted is indicated by parentheses. In this case mutants at loci *c* and *d* will be uncovered by the deficiency, and will display a mutant phenotype. Conversely, loci *a*, *b*, *e*, *f*, and *g* are heterozygous with wild-type alleles on the inversion chromosome and do not display mutant phenotype.

deficiency chromosomes, the *Drosophila* researcher can create "synthetic deficiencies," by combining other, often more complex, rearrangements. If a recessive allele's homozygous phenotype is identical to its hemizygous phenotype, it is likely to be an amorphic (null) allele, while if the hemizygous phenotype is stronger, it is likely to be a hypomorphic allele (see Sect. 5.1.2).

In some cases, a chromosome rearrangement inactivates or otherwise affects the expression of a gene, and can be used to pinpoint that gene's location both cytologically and molecularly.

The polytene chromosome map location of a gene is often a vital prelude to both further genetic analysis and its molecular cloning. This is discussed further in Sect. 3.2.

4 The *Drosophila* Genome Project

The *Drosophila* genome was sequenced by a collaboration of the Berkeley Drosophila Genome Project (BDGP), Celera Genomics, and the European Drosophila Genome Project (EDGP). Since the publication of the original draft sequence in 2000, two further revisions have been published, the most recent at the time of writing being Release 3.0. In Release 3.0, the euchromatin is represented by high quality sequence of 116.8 Mbp, with mere X gaps – in fact chromosome arm 3R is represented by a single contiguous sequence.

Identification of genes in the *Drosophila* genome has been achieved by combined strategy, in which data gleaned by conventional genetic and molecular analyses have been combined with purely computer-based techniques of *in silico* gene prediction. The input of data from conventional sources has been of particular utility in identifying alternatively spliced gene transcripts, and defining the 5′ and 3′ ends of genes.

The total of predicted genes in the *Drosophila* genome is 13 676. This is, perhaps, a rather small number; the nematode worm *C. elegans* appears to have around 19 000 genes, and morphologically at least appears a simpler organism.

4.1 Euchromatin and the Genes

98% of identified genes are located in the euchromatin, with the remainder lying in heterochromatic sequences. The total number of euchromatic genes is 13 379. The average number of exons per gene is 4.6, very similar to the genes of *C. elegans* and the plant *Arabidopsis thaliana*, but considerably fewer than the genes of the human genome (8.9).

Some interesting observations concerning the structure of *Drosophila* genes can be made. Firstly, nesting of genes is common: about 7.5% of genes lie within introns of other genes. Generally, such genes lie in opposite orientation to the gene in which they are located. Similarly, 15% of identified genes overlap another gene. There are 26 cases of genes that are interleaved with another gene. Finally, 48 examples of dicistronic genes – in which two genes with distinct protein products produce a single transcript – have been identified.

4.1.1 Duplicated Genes, Pseudogenes

There are comparatively few examples of duplicated genes and pseudogenes in the *Drosophila* genome. There are a mere 17 examples of pseudogenes identified by the

BDGP, of which 15 are evidently generated by a recombinational mechanism, as they contain introns, and of the remaining examples, one appears to be the result of retrotransposition (it has a poly-A tail and no introns), while the other is so scrambled that its origin cannot be determined.

4.2 Transposable Elements

Transposable elements, or transposons, are segments of DNA that are able to excise from the genome, or replicate, and insert at a second location. These elements fall into the middle repetitive faction of the genome, since they are generally present in copy numbers varying from a few tens to several hundred. *Drosophila* possesses a great many transposable elements (there are 1572 in Release 3 of the *Drosophila* genome sequence), classified into several families on the basis of structure and transposition mechanism. Thus, different strains of *D. melanogaster* have different chromosomal locations for each type of transposable element. The transposition activity of transposable elements can result in a genetic syndrome known as *hybrid dysgenesis*: high mutation rates, high frequency of chromosome rearrangements, sterility, and male recombination (male recombination does not normally occur in *Drosophila*). In general, each type of transposable element displays particular preferred sites of insertion, determined by DNA sequence and other less well-understood factors.

The *P* element is the best-characterized transposable element. Genetically, the activity of this element is seen as hybrid dysgenesis resulting from the cross of *P* element containing (P strain) males with females lacking *P* elements (M strain). Crossing M males with P females does not cause hybrid dysgenesis. Molecular analysis revealed the *P* element to be a 2.9-kb segment of DNA that encodes two proteins: a transposase, and a transposase repressor, depending on alternative splicing of four exons. Transposase is normally only expressed in the germ line. Additionally, the 31-bp terminal repeats of the element and other factors (not encoded by the element) are also required for transposition.

Cloned *P* elements may be injected into *Drosophila* embryos, and will subsequently integrate into the genome. By *in vitro* genetic manipulation, a rich variety of artificial *P* elements have been constructed for use in inserting exogenous DNA into the *Drosophila* germ line. These elements do not themselves encode transposase (so that they will be stable once integrated into the genome) and therefore require a source of transposase, which can be supplied by a defective *P* element encoding transposase, but lacking one or both of the terminal repeats. A more elegant method is to inject embryos of a *Drosophila* strain carrying a stably integrated element expressing transposase. Examples of *P* element derivatives are described below.

1. *Transformation vectors.* One use of germ-line transformation is to provide direct experimental evidence of the identity of a cloned gene. This can be done by determining whether or not a given segment of DNA identified in a cloning project can rescue a mutant phenotype when introduced into the genome by *P* element mediated germ-line transformation. Transformation vectors are defective elements that lack transposase, but possess the terminal repeats. In addition, they have a

selectable marker gene so that genetically transformed flies can be selected or recognized. Typical marker genes include the *Drosophila* gene *white*$^+$, and the bacterial antibiotic resistance gene *neo*^{R}. *white*$^+$ restores wild-type eye color in individuals of *white*$^-$ background, and *neo*^{R} confers resistance to the antibiotic G418, to which *Drosophila* are normally sensitive.

2. *Mutagenesis.* Insertion of a *P* element, or another transposable element often affects that gene's pattern of expression. It is therefore possible to make use of transposon mobilization as a mutagen. Mutants induced in this way present a considerable advantage when the mutated gene is to be cloned. Targeted mutagenesis will be discussed in Sect. 5.1.1.
3. *Enhancer trapping.* *P* elements that contain the *E. coli* *Lac*Z gene lacking a functional promoter sequence have been constructed. Upon insertion, these elements express *Lac*Z only if there is a neighboring gene with an enhancer able to direct expression of the *Lac*Z transgene. *Lac*Z expression is conveniently detected with a histochemical or an immunological assay. Owing in part to the high gene density of the *Drosophila* genome, a large proportion of enhancer trap elements do express *Lac*Z, often in an intricate temporal and spatial pattern that is frequently informative about the nearby gene's identity or function. Enhancer trapping is therefore an efficient means of detecting genes.
4. *The Gal4-UAS system.* This system comprises two distinct element insertions. The first carries the *S. cerevisiae* *Gal4* gene, which encodes a transcription factor that activates transcription from an upstream activating sequence (UAS_{Gal4}). The second element carries the transgene of interest, under the control of the UAS_{Gal4}, and which is therefore expressed only in cells expressing Gal4. The Gal4 expression may be controlled by a specific promoter, or by neighboring genomic enhancers (in which case the Gal4 element functions as an enhancer trap). This bipartite system is extremely flexible, since a large collection of lines expressing Gal4 in distinct spatial and temporal patterns is available. These can be combined with any UAS element carrying a transgene, which may be a reporter gene such as *LacZ*, or a mutant gene predicted to have a dominant phenotype. One exciting possibility is to express toxin genes (such as ricin A chain) in specific cell types, to investigate the developmental consequences of ablation of particular cells.
5. *FLP–FRT.* This is also a two-component system, in which one element carries the gene encoding the *S. cerevisiae* 2 μm plasmid fragment length polymorphism (FLP) recombinase, and the other the target sequences for this enzyme (FRT). The expression of FLP can be controlled in several ways as can the Gal4 described above. Cells in which FLP is expressed will rearrange sequences flanked by the FRT sites. Depending on the orientation of the two FRT sequences, this can be an inversion or an excision, enabling a gene between the sequences to be turned on or off.
6. Elements bearing visible marker genes can be used as dominant markers for the purpose of generating chromosomal deficiencies. For example, a w^- stock (*w* mutants have white eyes) bearing a $P[w^+]$ element inserted at, for example, 96A will have normal reddish-brown eyes. Following X irradiation,

flies with this chromosome region deleted will have lost the $P[w^+]$ element, and will have white eyes. These w^- stocks can be analyzed by polytene cytogenetics to reveal which, if any, carry a chromosomal deletion.

4.3 Heterochromatin

Heterochromatin is characterized by particular staining patterns that reflect differential condensation in comparison to euchromatin. Typical characteristics include late replication and a higher level of condensation than seen for euchromatin. Typically, heterochromatin is located in pericentromeric regions, and near telomeres. Furthermore, the Y chromosome is entirely heterochromatic. Heterochromatin does not polytenize, and is not represented in the polytene chromosome maps; it forms the chromocentre, from which the polytenized chromosome arms radiate. Detailed cytological analyses of mitotic chromosomes have permitted banding maps of the heterochromatin to be constructed on the basis of differential staining patterns; such maps are of low resolution, consisting of 61 defined bands.

Sequencing the DNA corresponding to heterochromatin presents considerable difficulty because of its highly repetitive nature. For this reason, the BDGP's heterochromatic sequence remains draft quality – of the estimated 59 Mb heterochromatin (to which must be added the 41 Mb Y chromosome, which is entirely heterochromatic), an unfinished draft-quality assembly of 20.7 Mb has been produced and distributed over 2597 scaffolds. Nevertheless, several conclusions may be drawn.

Some 297 protein-coding genes have been identified within heterochromatin. This total includes several loci known from genetic analyses to map within heterochromatin. Several genes that do not encode proteins have been identified; these include 5.8S, 2S, 18S, and 28S rRNA genes.

The general organization of heterochromatic DNA is in blocks of approximately 20 kb to 1 Mb consisting of satellite DNA, separated by sections of between 5 and 50 kb of more complex sequences that contain a high density of transposable elements.

4.4 Long-range Genome Structure

That the structure of the *Drosophila* genome involves long-range interactions is in no doubt, as clearly indicated by genetic phenomena, such as transvection, in which regulatory interactions between homologous chromosomes are revealed. One of the outcomes of the *Drosophila* genome sequence was the ability to examine the transcriptional activity of the whole genome with reference to specific treatments or biological processes, by microarray studies. Analyses of several microarray experiments encompassing over 80 experimental conditions has revealed that there appears to be considerable clustering of genes which share expression profiles.

Over 20% of the *Drosophila* genes surveyed appeared to be clustered in groups of 10 to 30 genes with similar expression patterns. This is likely to reflect a regulation at the level of chromatin structure. However, there appeared to be no correlation with the banding pattern of polytene chromosomes, or with

sites of known attachment to the nuclear scaffold.

5 Genetic Analysis in *Drosophila*

Despite the availability of the complete genome sequence, genetic analytical techniques remain very important. Many of the genes identified in the genome have no known mutant alleles. Examination of mutant phenotype, often in combination with mutation of other genes, has the potential to reveal their biological function.

5.1 Mutagenesis

A typical genetic screen for mutations, in this case autosomal recessive lethals, is shown in Fig. 5. A large variety of mutagens are available, which can differ widely in the nature of the lesions produced. For example, ethylmethanesulphonate (EMS), administered in the food, causes predominantly single base pair changes, whereas X rays induce chromosomal deletions and other rearrangements. The use of transposable elements as mutagens is described in Sect. 4.1.

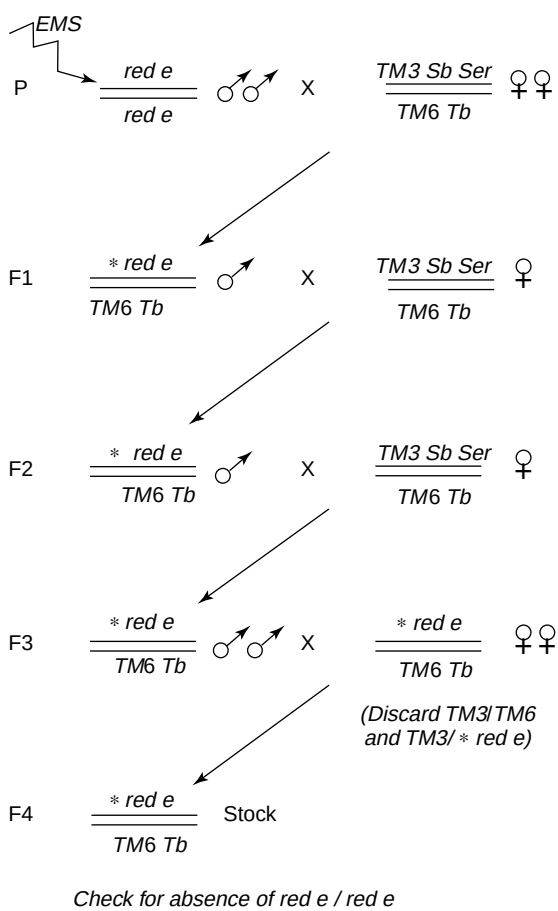

Fig. 5 Mutagenesis scheme. This scheme is designed to recover autosomal recessive lethal mutations. With minor modification, however, other types of phenotype may be screened for. Full descriptions of marker mutations and balancer chromosomes can be found in *The Genome of Drosophila melanogaster*. All females in crosses are virgin. The markers in this figure are: *red*: red malpighian tubules (wild-type tubules are whitish, this recessive mutation also results in brown eyes in the adult); *ebony* (*e*): adult has black body, recessive; *Stubble* (*Sb*): dominant mutation affecting adult bristle morphology; *Serrate* (*Ser*): dominant mutation causing serrated wing margins; *Tubby* (*Tb*): dominant mutation causing tubby larvae and adults, but more easily scored in larvae. The mutagenized chromosome is indicated by an asterisk.

Male flies are treated with mutagen, and mated *en masse* to virgin females heterozygous for two balancer chromosomes. All subsequent matings are single-pair matings, again with virgin females. The target chromosome is marked with recessive markers, to facilitate later genetic manipulations (the mutagenized chromosome is labeled with an asterisk in subsequent crosses in Fig. 5). Note that the F1 and F2 crosses are identical – this is necessary when the mutagen, such as EMS, causes lesions requiring one or more cell divisions to become fixed. The F3-cross yields a balanced stock in which flies homozygous for the mutagenized chromosome (easily recognized because of the recessive markers) can be examined for the phenotype of interest. In the case of recessive lethals, this class would be absent. Balancer chromosomes have multiple inversions, which prevent recombination from occurring, and carry a recessive lethal mutation to render balancer chromosome homozygotes inviable. In general, balancer chromosomes carry a dominant marker, so they can be followed in genetic crosses. Because balancer chromosomes suppress recombination, these features will not be separated by recombination. Some X chromosome balancer chromosomes carry a female sterile mutation rather than a lethal, so that the balancer chromosome will be male viable.

Many classes of mutant phenotype can be discovered, using appropriate genetic systems, from dominant temperature-sensitive lethals, to simple visible mutations, from female sterile mutants to behavioral mutations.

5.1.1 Reverse Genetics

Often, the *Drosophila* researcher is in the position of having cloned a gene, but is without a corresponding mutant. Unlike other model organisms, there is no efficient means of targeted mutagenesis in *Drosophila*. A useful technique in this regard is to use *P* element mutagenesis as described in Sect. 4.2 and detect insertions in the gene of interest using PCR with one gene-specific primer and one *P* element specific primer.

As an alternative to PCR screening *P* element insertions, saturation mutagenesis of a chromosomal region known to include the gene of interest, and spanned by a small chromosome deficiency can be undertaken. Mutants displaying an appropriate phenotype could then be examined molecularly, to investigate whether the gene in question is mutated.

A very powerful technique of targeted mutagenesis has been developed by Golic and colleagues. It is clear that it is of general utility, but it is uncertain at the time of writing whether it will be widely adopted by the *Drosophila* research community.

5.1.2 Classes of Mutant Allele

Several classes of mutant allele may be recovered following mutagenesis. The simplest is the loss-of-function mutation, either complete (null mutation or amorph) or partial (hypomorph). Amorphs and hypomorphs are generally recessive, and can be formally distinguished by comparing their phenotypes in combination with a deficiency (see Sect. 3.3.1). Other allelic states are hypermorphs, in which there is a gain of gene function, and neomorphs, in which a novel function is acquired. Neomorphs are generally dominant alleles. Mutants of any or all classes can be recovered for a given gene, using appropriate mutagenesis protocols.

An additional complication is that an allele's phenotypic effect may be temperature-sensitive. At its extreme, this can be seen as a true conditional

mutant, in which environmental conditions can entirely suppress the mutant phenotype.

5.1.3 Position Effect Variegation

As described earlier, most genes are located within euchromatin. In comparison with euchromatin, heterochromatin has unusual properties, connected with differences in chromatin organization and composition. These properties influence gene expression. When a gene, for example *white*, is relocated by chromosome rearrangement so that it is located close to heterochromatin, its expression is affected by the heterochromation. The heterochromatin appears to "heterochromatinize" its neighboring euchromatin. This occurs to different extents in different clonal cell lineages within the developing organism. In the case of a *white* allele mutated by an inversion bringing it into close proximity of the centric heterochromatin, it may be active in some ommatidia of the compound eye, and inactive in others, resulting in a mottled eye with red and white facets (*white* mutants have white eyes rather than the wild-type reddish color). The *white* allele w^{m4} is an example. There are loci that display the opposite effect; an example is *light*, which maps to the pericentric heterochromatin of chromosome arm 2L: rearrangements that relocate *light* to euchromatic regions result in reduced levels of expression. Mutants that enhance or suppress PEV are known, and characterization of these genes is illuminating with respect to chromatin structure.

5.1.4 Interfering RNA

Perhaps the single most important molecular biological technique to have been developed in recent years is that of RNA interference. *Drosophila* researchers have been able to apply this new technology in several ways – firstly by injecting embryos or transfecting cultured cells with double-stranded RNA corresponding to a gene under investigation, and subsequently examining the phenotypic consequences, and secondly by generating transgenic strains, which encode a transcript suitable for genetic interference.

5.1.5 Targeted Mutagenesis

Rong and Golic published a method for targeted mutagenesis, in which *in vitro* mutated genes are inserted in the *Drosophila* genome by P-element techniques and then substituted for the wild type allele by a recombination based procedure. A detailed description of this procedure is out of the scope of this article, but can be found in the cited references. If widely adopted, this technique has great potential.

5.2 Genetic Mapping

Genetic, or recombination mapping, in *D. melanogaster* has a long history. The theoretical basis of recombination mapping was devised and implemented using *Drosophila* in 1913. The map unit, or centimorgan (cM) corresponds to one percent recombination. Over short map distances, these values are additive, but over longer distances, complications arise from multiple crossovers. Correlating the cytogenetic and recombination map positions of a number of loci has demonstrated that the recombination rate varies along the chromosome: recombination is effectively suppressed close to telomeres and centromeres. Thus, 1 cM represents a larger stretch of DNA near the telomere or centromere than in the middle of a chromosome arm.

Under normal circumstances, there is no recombination in *D. melanogaster* males. The special circumstances in which male recombination is observed include hybrid dysgenesis. Additionally, recombination in the tiny fourth chromosome does not occur, except in triploid females.

Recombination mapping is now generally undertaken as a starting point to cytological mapping, the localization of a gene on the polytene chromosome map, and its subsequent molecular cloning. The correlation between the polytene chromosome map and the recombination map permits the approximate cytogenetic map position to be determined from recombination map data. The recombination map order of loci can be important in cloning projects.

A typical scheme for mapping a recessive lethal mutation is shown in Fig. 6. The first two crosses in this scheme are carried out *en masse*, while subsequent crosses are single-pair matings. Each recombinant stock generated in this scheme can be classified on the basis of which markers are present on the recombinant chromosome, and whether the lethal is present. This permits easy assignation of the lethal to one interval between markers. Further recombination mapping may be undertaken, concentrating on that interval, to yield a more precise map position.

5.2.1 Microsatellite and RFLP Markers

Genetic mapping in *Drosophila* has rarely used microsatellites and restriction fragment length polymorphism (RFLP) markers, due to the availability of many visible markers and chromosome rearrangements, permitting the easy localization of genes to a molecularly

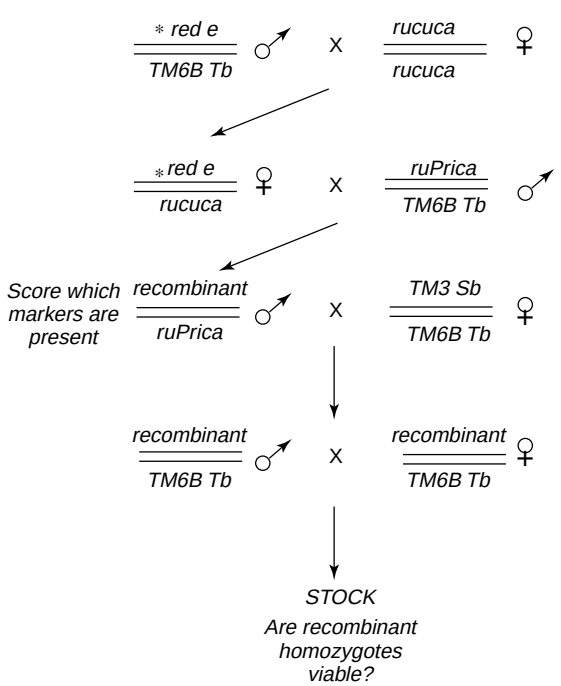

Fig. 6 *Recombination mapping scheme.* Markers are as described in the legend to Fig. 5, and are as follows. The chromosome *rucuca* is multiply marked with the following recessives: *ru* (*roughoid*: eyes are rough in appearance); *h* (*hairy*: extra hairs on wing veins); *th* (*thread*: aristae are threadlike, instead of branched); *st* (*scarlet*: adult eye is bright red); *cu* (*curled*: wings curled upwards); *sr* (*stripe*: stripe on dorsal surface of thorax); *e*; *ca* (*claret*: adult eyes are ruby red). Flies homozygous for both *st* and *ca* have orange eyes. The chromosome *ruPrica* is as *rucuca*, but in addition contains the dominant marker *Prickly* (*Pr*), which causes the adult bristles to be shortened, with twisted tips. As in Fig. 5, all females are virgin.

relevant precision. However, microsatellites and RFLPs are of central importance for recombination mapping in genetically less tractable insects, such as mosquitoes.

6 Prospects

To some extent, *Drosophila* is a late starter with respect to the development of molecular physical maps. This is undoubtedly due to having had an excellent physical map dating from the 1930s, the polytene chromosome map. It is an enduring legacy to the mapping talents of Calvin Bridges that his polytene chromosome map has continued to be relevant to researchers undertaking detailed molecular genetic studies undreamt of by the early geneticists.

See also Genetics, Molecular Basis of; Genomic Sequencing (Core Article)

Bibliography

Books and Reviews

Adams, M.D., Sekelsky, J.J. (2002) From sequence to phenotype: reverse genetics in *Drosophila melanogaster*, *Nat. Rev. Genet.* **3**, 189–198.

Ashburner, M. (1989) *Drosophila, A Laboratory Handbook*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Duffy, J.B. (2002) GAL4 system in Drosophila: a fly geneticist's Swiss army knife, *Genesis* **34**(1–2), 1–15.

Lefevre, G. (1976) A Photographic Representation and Interpretation of the Polytene Chromosomes of *Drosophila Melanogaster* Salivary Glands, in: Ashburner, M., Novitski, E. (Eds.) *The Genetics and Biology of Drosophila*, Vol. 1a, Academic Press, New York, pp. 31–66.

Sorsa, V. (1988) *Chromosome Maps of Drosophila*, CRC Press, Boca Raton, Florida.

St Johnston, D. (2002) The art and design of genetic screens: *Drosophila melanogaster*, *Nat. Rev. Genet.* **3**, 176–188.

The FlyBase Consortium (2003). The FlyBase database of the Drosophila genome projects and community literature, *Nucleic Acids Res.* **31**, 172–175, http://flybase.org/

Primary Literature

Adams, M.D., Celniker, S.E., et al. (2000) The genome sequence of *Drosophila melanogaster*, *Science* **287**, 2185–2195.

Agard, D.A., Sedat, J.W. (1983) Three-dimensional architecture of a polytene nucleus, *Nature* **302**, 676–681.

Ajioka, J.W., Garza, D., Johnson, D., Carulli, J.P., Jones, R.W., Hartl, D.L. (1990) Genome Evolution Analyzed by Cloning Large Fragments of *Drosophila* DNA in Yeast Artificial Chromosomes, *Molecular Evolution*, A. R. Liss.

Ajioka, J.W., Smoller, D.A., Jones, R.W., Carulli, J.P., Vellek, A.E.C., Garza, D., Link, A.J., Duncan, I.W., Hartl, D.L. (1991) *Drosophila* genome project: one-hit coverage in yeast artificial chromosomes, *Chromosoma* **100**, 495–509.

Arbeitman, M.N., et al. (2002) Gene expression during the life cycle of Drosophila melanogaster, *Science* **297**(5590), 2270–2275.

Ashburner, M., et al. (1999) An exploration of the sequence of a 2.9-Mb region of the genome of *Drosophila melanogaster*: The *Adh* region, *Genetics* **153**, 179–219.

Ballinger, D., Benzer, S. (1989) Targeted gene mutations in *Drosophila*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9402–9406.

Benos, P.V., et al. (2001) From first base: the sequence of the tip of the X chromosome of *Drosophila melanogaster*, a comparison of two sequencing strategies, *Genome Res.* **11**, 710–730.

Bentley, A., MacLennan, B., Calvo, J., Dearolf, C.R. (2000) Targeted recovery of mutations in Drosophila, *Genetics* **156**, 1169–1173.

Bergman, C.M., et al. (2002) Assessing the impact of comparative genomic sequence data on the functional annotation of the

Drosophila genome, *Genome Biol.* **3**(12), RESEARCH0086-6.

Bibikova, M., Golic, M., Golic, K.G., Carroll, D. (2002) Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases, *Genetics* **161**(3), 1169–1175.

Bier, E., et al. (1989) Searching for pattern and mutation in the *Drosophila* genome with a *P-laxZ* vector, *Genes Dev.* **3**, 1273–1287.

Bieschke, E.T., Wheeler, J.C., Tower, J. (1998) Doxycycline-induced transgene expression during *Drosophila* development and aging, *Mol. Gen. Genet.* **258**, 571–579.

Boutanaev, A.M., Kalmykova, A.I., Shevelyov, Y.Y., Nurminsky, D.I. (2002) Large clusters of co-expressed genes in the Drosophila genome, *Nature* **420**(6916), 666–669.

Bridges, C.B. (1935) Salivary chromosomes. With a key to the banding of the chromosomes of *Drosophila melanogaster*, *J. Hered.* **26**, 60–64.

Bridges, C.B., Bridges, P.N. (1939) A new map of the second chromosome: a revised map of the right limb of the second chromosome of *Drosophila melanogaster*, *J. Hered.* **30**, 475–476.

Bridges, P.N. (1941) A revision of the salivary gland 3R-chromosome map of *Drosophila melanogaster*, *J. Hered.* **32**, 299–300.

Celniker, S.E., Wheeler, D.A. et al. (2002) Finishing a whole-genome shotgun: release 3 of the Drosophila melanogaster euchromatic genome sequence, *Genome Biol.* **3**(12), RESEARCH0079-9.

Clemens, J.C., Worby, C.A., Simonson-Leff, N., Muda, M., Maehama, T., Hemmings, B.A., Dixon, J.E. (2000) Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6499–6503.

Cooley, L., Kelley, R., Spradling, A.C. (1988) Insertional mutagenesis of the *Drosophila* genome with a single P element, *Science* **239**, 1121–1128.

Danilevskaya, O.N., Petrov, D.A., Pavlova, M.N., Koga, A., Kurenova, E.V., Hartl, D.L. (1992) A repetitive DNA element, associated with telomeric sequences in *Drosophila melanogaster*, contains open reading frames, *Chromosoma* **102**, 32–40.

Deak, P., Omar, M.M., et al. (1997) P element insertion alleles of essential genes on the third chromosome of *Drosophila melanogaster*: Correlation of physical and cytogenetic maps in chromosomal region 86E-87F, *Genetics* **147**, 1697–1722.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., Tuschl, T. (2001) Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate, *EMBO J.* **20**(23), 6877–6888.

Engels, W.R., Preston, C.R. (1984) Formation of chromosome rearrangements by P factors in *Drosophila*, *Genetics* **107**, 657–678.

Garza, D., Ajioka, J.W., Burke, D.T., Hartl, D.L. (1989) Mapping the *Drosophila* genome with yeast artificial chromosomes, *Science* **246**, 641–646.

Garza, D., Ajioka, J.W., Carulli, J.P., Jones, R.W., Johnson, D.H., Hartl, D.L. (1989) Physical mapping of complex genomes, *Nature (Product review)* **340**, 577–578.

Gavrila, L., Ecovoiu, A.A., et al. (2001) Localizing genes in *Drosophila melanogaster* polytene chromosomes by fluorescence *in situ* hybridization, *J. Cell Mol. Med.* **5**(1).

Giordano, E., Rendina, R., Peluso, I., Furia, M. (2002) RNAi triggered by symmetrically transcribed genes in *Drosophila melanogaster*, *Genetics* **160**, 637–648.

Gloor, G.B., Nassif, N.A., Johnson-Schlitz, D.M., Preston, C.R., Engels, W.R. (1991) Targeted gene replacement in Drosophila via P element-induced gap repair, *Science* **253**(5024), 1110–1117.

Gong, W.J., Golic, K.G. (2003) Ends-out, or replacement, gene targeting in Drosophila, *Proc. Natl. Acad. Sci. U.S.A.* **100**(5), 2556–2561.

Hartl, D.L. (1992) Genome Map of *Drosophila Melanogaster* Based on Yeast Artificial Chromosomes, in: Tilghman, Davies (Eds.) *Genome Analysis*, Vol. 4, Cold Spring Harbor Laboratory Press, Cold spring Harbor, pp. 39–69.

Hartl, D.L., Ajioka, J.W., Cai, H., Lohe, A., Lozovskaya, E.R., Smoller, D.A., Duncan, I.A. (1992) Towards a *Drosophila* genome map, *Trends Genet.* **8**(2), 70–75.

Hartl, D.L., Nurminsky, D.I., Jones, R.W., Lozovskaya, E.R. (1994) Genome structure and evolution in *Drosophila*: applications of the framework P1 map, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6824–6829.

Hochstrasser, M., Mathog, D., Gruenbaum, Y., Saumweber, H., Sedat, J.W. (1986) Spatial organization of chromosomes in the salivary gland nuclei of *Drosophila melanogaster*, *J. Cell Biol.* **102**, 112–123.

Hoheisel, J.D., Lennon, G.G., et al. (1991) Use of high coverage reference libraries of *Drosophila melanogaster* for relational data analysis. A step towards mapping and sequencing of the genome, *J. Mol. Biol.* **221**, 903–914.

Hoskins, R.A., et al. (2002) Heterochromatic sequences in a Drosophila whole-genome shotgun assembly, *Genome Biol.* **3**(12), RESEARCH0085-5.

Huang, A.M., Rehn, E.J., Rubin, G.M. (2000) Recovery of DNA Sequences Flanking P-element Insertions: Inverse PCR and Plasmid Rescue, in: Sullivan, W., Ashburner, M., Hawley, R.S. (Eds.) *Drosophila Protocols*, CSHL Press, Cold Spring Harbor, pp. 429–437.

Jack, J.W., Judd, B.H. (1979) Allelic pairing and gene regulation: a model for the *zeste-white* interaction in *Drosophila*, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1368–1372.

Judd, B.H. (1988) Transvection: allelic crosstalk, *Cell* **53**, 841–843.

Kafatos, F.C., Louis, C., Savakis, C., Glover, D.M., Ashburner, M., Link, A.J., Siden-Kiamos, I., Saunders, R.D.C. (1991) Integrated maps of the *Drosophila* genome: progress and prospects, *Trends Genet.* **7**, 155–161.

Kaiser, K., Goodwin, S.F. (1990) ''Site-directed'' transposon mutagenesis of *Drosophila*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1686–1690.

Kaminker, J.S., et al. (2002) The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective, *Genome Biol.* **3**(12), RESEARCH0084-4.

Kennerdell, J.R., Carthew, R.W. (1998) Use of dsRNA-mediated genetic interference to demonstrate that *frizzled* and *frizzled 2* act in the Wingless pathway, *Cell* **95**, 1017–1026.

Kennerdell, J.R., Carthew, R.W. (2000) Heritable gene silencing in *Drosophila* using double-stranded RNA, *Nat. Biotechnol.* **18**, 896–898.

Kornberg, T. (2002) Another arrow in the Drosophila quiver, *Proc. Natl. Acad. Sci. U.S.A.* **99**(15), 9607–9608.

Laird, C.D., McCarthy, B.J. (1969) Molecular characterization of the *Drosophila* genome, *Genetics* **63**, 865–882.

Leach, T.J., Chotkowski, H.L., Wotring, M.G., Dilwith, R.L., Glaser, R.L. (2000) Replication of heterochromatin and structure of polytene chromosomes, *Mol. Cell Biol.* **20**(17), 6308–6316.

Lindsley, D.L., Zimm, G.G. (1992) *The Genome of Drosophila Melanogaster*, Academic Press, San Diego.

Lozovskaya, E.R., Petrov, D.A., Hartl, D.A. (1993) A combined molecular and cytogenetic approach to genome evolution in *Drosophila* using large-fragment DNA cloning, *Chromosoma* **102**, 253–266.

Madueno, E., et al. (1995) A physical map of the X chromosome of *Drosophila melanogaster*: cosmid contigs and sequence tagged sites, *Genetics* **139**, 1631–1647.

Mathog, D., Sedat, J.W. (1989) The three-dimensional organization of polytene nuclei in male *Drosophila melanogaster* with compound XY or ring X chromosomes, *Genetics* **121**, 293–311.

Mathog, D., Hochstrasser, M., Gruenbaum, Y., Saumweber, H., Sedat, J.W. (1984) Characteristic folding pattern of polytene chromosomes in *Drosophila* salivary gland nuclei, *Nature* **308**, 414–421.

Miklos, G.L., Rubin, G.M. (1996) The role of the genome project in determining gene function: insights from model organisms, *Cell* **86**(4), 521–529.

Miklos, G.L.G., Healy, M.J., Pain, P., Howells, A.J., Russell, R.J. (1984) Molecular and genetic studies on the euchromatin-heterochromatin transition region of the X chromosome of *Drosophila melanogaster*, *Chromosoma* **89**, 218–227.

Miklos, G.L.G., Yamamoto, M.-T., Davies, J., Pirrotta, V. (1988) Microcloning reveals a high frequency of repetitive sequences characteristic of chromosome 4 and the β-heterochromatin of *Drosophila melanogaster*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2051–2055.

Misra, S., et al. (2002) Annotation of the Drosophila melanogaster euchromatic genome: a systematic review, *Genome Biol.* **3**(12), RESEARCH0083-RESEARCH0083.

Myers, E.W., et al. (2000) A whole-genome assembly of *Drosophila*, *Science* **287**, 2196–2204.

Ohler, U., Liao, G.C., Niemann, H., Rubin, G.M. (2002) Computational analysis of core promoters in the Drosophila genome, *Genome Biol.* **3**(12), RESEARCH0087-7.

Painter, T.S. (1933) A new method for the study of chromosome rearrangements and the plotting of chromosome maps, *Science* **78**, 585–586.

Painter, T.S. (1934) A new method for the study of chromosome aberrations and the plotting of chromosome maps in *Drosophila melanogaster*, *Genetics* **19**, 175–188.

Peter, A., et al. (2002) Mapping and identification of essential gene functions on the X chromosome of *Drosophila*, *EMBO Rep.* **3**, 34–38.

Potter, C.J., Turenchalk, G.S., Xu, T. (2000) *Drosophila* in cancer research, *Trends Genet.* **16**(1), 33–39.

Roman, G., Endo, K., Zong, L., Davis, R.L. (2001) P[Switch], a system for spatial and temporal control of gene expression in Drosophila melanogaster, *Proc. Natl. Acad. Sci. U.S.A.* **98**(22), 12602–12607.

Rong, Y.S., Golic, K.G. (2000) Gene targeting by homologous recombination in *Drosophila*, *Science* **288**, 2013–2018.

Rong, Y.S., Golic, K.G. (2001) A targeted gene knockout in Drosophila, *Genetics* **157**(3), 1307–1312.

Rong, Y.S., Titen, S.W. et al. (2002) Targeted mutagenesis by homologous recombination in D. melanogaster, *Genes Dev.* **16**(12), 1568–1581.

Rørth, P. (1996) A modular misexpression screen in *Drosophila* detecting tissue specific phenotypes, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12418–12422.

Rørth, P., Szabo, K. et al. (1998) Systematic gain-of-function genetics in *Drosophila*, *Development* **125**, 1049–1057.

Scalenghe, F., Turco, E., Edstrom, J.E., Pirrotta, V., Melli, M. (1981) Microdissection and cloning of DNA from a specific region of *Drosophila melanogaster* polytene chromosomes, *Chromosoma* **82**, 205–216.

Schubeler, D., Scalzo, D., Kooperberg, C., van Steensel, B., Delrow, J., Groudine, M. (2002) Genome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing, *Nat. Genet.* **32**(3), 438–442.

Sharakhov, I.V., Fillippova, M.A., et al. (1993) beta-Heterochromatin in Drosophila – molecular organization and functioning – characterization of the dna sequence from the d-melanogaster proximal beta-heterochromatin associated with the nuclear envelope, *Genetika* **29**(3), 393–402.

Siden-Kiamos, I., et al. (1990) Towards a physical map of the *Drosophila melanogaster* genome: mapping of cosmid clones within defined genomic divisions, *Nucelic Acids Res.* **18**, 6261–6270.

Smoller, D.A., Petrov, D., Hartl, D.L. (1991) Characterization of bacteriophage P1 library containing inserts of *Drosophila* DNA of 75–100 kilobase pairs, *Chromosoma* **100**, 487–494.

Spellman, P.T., Rubin, G.M. (2002) Evidence for large domains of similarly expressed genes in the *Drosophila* genome, *J. Biol.* **1**, 5.

Spierer, A., Spierer, P. (1984) Similar level of polyteny in bands and interbands of *Drosophila* giant chromosomes, *Nature* **307**, 176–178.

Spradling, A.C., Stern, D., Beaton, A., Rhem, E.J., Laverty, T., Mozden, N., Misra, S., Rubin, G.M. (1999) The Berkeley Drosophila genome project gene disruption project: single P-element insertions mutating 25% of vital Drosophila genes, *Genetics* **153**(1), 135–77.

Stapleton, M., et al. (2002) The Drosophila gene collection: identification of putative full-length cDNAs for 70% of D. melanogaster genes, *Genome Res.* **12**(8), 1294–1300.

Stapleton, M., et al. (2002) A Drosophila full-length cDNA resource, *Genome Biol.* **3**(12), RESEARCH0080-0.

Stark, A., Brennecke, J., Russell, R.B., Cohen, S.M. (2003) Identification of *Drosophila* microRNA targets, *Public Library Sci. Biol.* **1**(3), 1–13.

Stebbins, M.J., Yin, J.C.P. (2001) Adaptable doxycycline-regulated gene expression systems for *Drosophila*, *Gene* **270**, 103–111.

Stebbins, M.J., Urlinger, S., Byrne, G., Bello, B., Hillen, W., Yin, J.C.P. (2001) Tetracycline-inducible systems for *Drosophila*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10775–10780.

Toba, G., Ohsako, T., Miyata, N., Ohtsuka, T., Seong, K.-H., Aigaki, T. (1999) The gene search system: a method for efficient detection and rapid molecular identification of genes in *Drosophila melanogaster*, *Genetics* **151**, 725–737.

Tomancak, P., et al. (2002) Systematic determination of patterns of gene expression during Drosophila embryogenesis, *Genome Biol.* **3**(12), RESEARCH0088-8.

Tower, J., Karpen, G.H., Craig, N., Spradling, A.C. (1993) Preferential transposition of Drosophila *P* elements to nearby chromosomal sites, *Genetics* **133**, 347–359.

van Roessel, P., Brand, A.H. (2000) GAL4-mediated ectopic gene expression in *Drosophila*, in: Sulliva, W., Ashburner, M.,

Hawley, R.S. (Eds.) *Drosophila Protocols*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 439–447.

White, K.P., Rifkin, S.A., Hurban, P., Hogness, D.S. (1999) Microarray analysis of *Drosophila* development during metamorphosis, *Science* **286**, 2179–2184.

Yamamoto, M.-T., Mitchelson, A., Tudor, M., O'Hare, K., Davies, J.A., Miklos, G.L.G. (1990) Molecular and cytogenetic analysis of the heterochromatin-euchromatin junction region of the *Drosophila melanogaster* X chromosome using cloned DNA sequences, *Genetics* **125**, 821–832.

Drug Bioavailability, Distribution and Clearance Prediction

William Bains
Choracle Ltd., 37 The Moor, Melbourn, Royston, Herts SG8 6ED, UK

1 Introduction 597

2 Approaches to Modeling 600
2.1 Descriptors 601
2.2 Approaches to Building Models 607
2.2.1 Types of Models: Mechanistic Models 607
2.2.2 Types of Models: Statistical Models 608
2.2.3 Problems with Modeling and Overcoming them 608
2.2.4 Nonlinear and "Artificial Intelligence" Approaches 609
2.3 The Problem of Data 611
2.3.1 Datasets and Statistical Reliability 612
2.4 Conclusions on Modeling 613

3 Mode of Deployment 613

4 Accuracy 614

5 Oral Bioavailability 616
5.1 Simple Rules for OB 617
5.2 Advanced Type V Models for Oral Bioavailability 618
5.3 Models of Components of OB 619
5.3.1 Gut Physiology, Solubility, pK_a, and LogP 619
5.3.2 Solubility 619
5.3.3 Intestinal Absorption 621
5.3.4 Physicochemical Methods 623
5.3.5 CaCo-2 Model 624
5.3.6 Conclusion for Passive OB and IA 624
5.3.7 P-glycoprotein 625
5.3.8 Other Transporters 626

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

5.3.9 Metabolism and Intrinsic Clearance 626
5.3.10 Conclusion of P-gp, Transporters, and CYPs 628

6 Distribution 629
6.1 Volume of Distribution and Protein Binding 629
6.2 Blood–brain Barrier 630
6.3 Conclusion on Distribution 633

7 Integration of Predictive Models 633

8 Overall Conclusion and a Caution 634

Acknowledgments 634
Note 634

Bibliography 635
Books and Reviews 635
Primary Literature 635

Keywords

ADME, oral bioavailability, drug design, blood–brain barrier, P-gp, CYP, virtual biology, in silico, DMPK

Prediction of how drugs are absorbed, metabolized, distributed round the body, and excreted is increasingly important to the design of new medicines. However, it is too complex to model rigorously. Prediction algorithms are primarily based on databases of known examples, which are mined for simple rules, more complex mathematical algorithms, or rules sets. Many modeling techniques have been applied, and many ways of describing a molecule to the computer have been tested. I discuss the application of these methods for predicting whether a drug will be absorbed through the intestines, including whether it will be soluble, transported in or out by cell membrane transporters, and whether it will be metabolized, and for predicting penetration into the brain. Advanced statistical and machine learning techniques can generate good algorithms for approximating these quantities. However, predictions are not yet good enough to integrate into a "virtual human". The main limitation is that the algorithms need more experimental data to encompass the complexity and variability of biological systems, as well as better, more biologically relevant ways of describing the molecules to the computer.

1 Introduction

Drugs must have many properties apart from their pharmacological action and lack of toxicity if they are to be suitable as components of medicines. The ideal medicine should be easy to take, quick to act on the relevant tissue, and be required in a dosing regime that is easy for the patient to accommodate and remember, which usually means a once-a-day tablet. Whether a specific chemical can achieve this depends on whether it can be absorbed through the gut wall, once it has been absorbed, where it goes to, how long it stays there, and how it is broken down or excreted from the body. Together, these properties comprise the DMPK (drug metabolism and pharmacokinetics) or, more broadly, ADME (absorption, distribution, metabolism, and excretion) properties of the drug. Structural features of the molecules that confer suitable ADME properties are said to make a molecule "druglike". This article discusses the extent to which these properties can be predicted by computer software, how this is achieved, and whether such predictions can be improved to the point where they become mainstream practice in pharmaceutical discovery.

In the 1970s and 1980s, around 40% of compounds entering clinical development failed because of inappropriate pharmacokinetic parameters (although this figure is dominated by antiinfectives, whose efficacy is easy to test). This figure is substantially lower today (probably in the order of 5%) because such parameters are tested much earlier in drug discovery. However, drug discovery and development is not getting faster or cheaper. Testing ADME properties during discovery and development identifies failures earlier, but this does not seem to reduce the number of those failures and hence the time needed to replace them with more successful molecules.

The ultimate parameters in which we are interested are the integrated concentration of the drug over time after a given dose (area under curve – AUC), which provides a measure of the time for which the drug will be present in the target tissue at a therapeutically useful concentration but not at a toxic one, and the half-life, which provides an indication of how often dosing must occur to maintain a therapeutically useful concentration. (Usually, AUCs refer to blood concentration, which is taken as a substitute for tissue concentration.) The components of this are generally called ADME.

- The amount of drug taken into the animal (colloquially, "on board"), which for any enteral route of delivery is related to its absorption.
- Its concentration, a function of how it is distributed around the tissue of the body, that is, its distribution around the body.
- How much and how fast it is removed by metabolism.
- How fast and by what route the drug and its metabolites are excreted from the body.

The four basic components of ADME are not themselves irreducible or independent and can be analyzed in terms of less complex processes, as illustrated in Fig. 1; however, as Fig. 1 suggests, there is no simple one-to-one relationship between the less complex processes and the physiological parameters that we want to know. For convenience, the whole-organism properties are usually referred

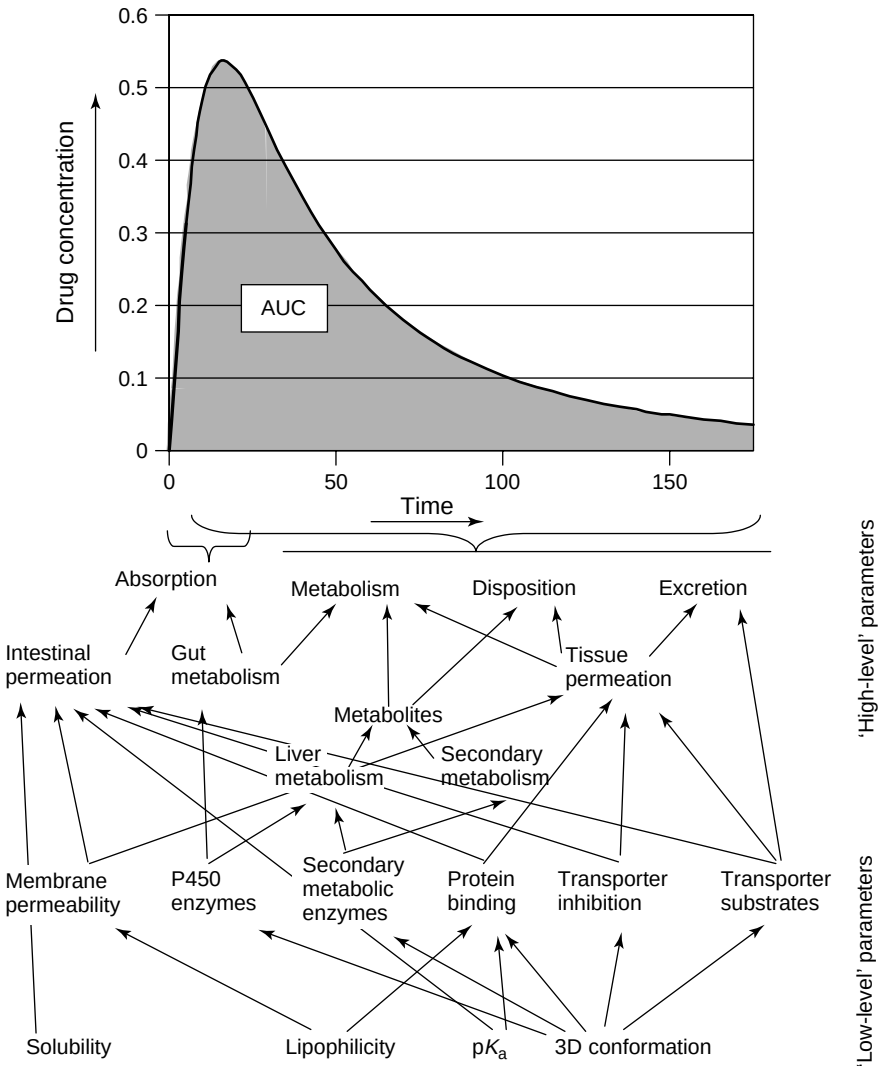

Fig. 1 Cartoon showing the concentration: time curve for a "typical" orally absorbed drug, the area under curve integrated time/concentration measure, and the relationship between some of its component parameters.

to as *high level* and the apparently simpler cellular and molecular properties as *low level*. The general study of how a drug's concentration changes in the body is called *pharmacokinetics*.

There are three approaches to estimating ADME properties.

Extrapolate from animal studies This is the most obvious route, and although it is not practical to continue to study compounds one at a time in animals, dosing several compounds at once into animals (cassette dosing) can extend the capabilities of animal studies substantially.

Results from rodents can be mapped onto results for man through allometric scaling algorithms, on the basis of relative mass, fluid volumes, blood flow, and other parameters.

The problem with this approach is that animal studies rarely reflect the result in man even qualitatively. It is now well known that drug metabolism is qualitatively different in the common laboratory animals and in man, with the enzyme complement differing in presence, level, and specificity: if these differences are completely characterized, then they can be corrected for, and human hepatic clearance can be predicted from scaling models. Higher-level parameters can correlate or can be apparently completely unconnected: for example, there is little correlation between overall oral bioavailability (OB) in man and in rat or dog (two commonly used model animals – see Fig. 2) due to radical differences in gut physiology, and although the fraction absorbed in monkeys and man is similar, it is more difficult from a practical point of view to test new drugs in higher primates than in human volunteers. The only appropriate model for human bioavailability as a whole is man.

Animal experiments are therefore now used to derive lower-level parameters, such as metabolism through specific enzymes, diffusion across the intestinal wall, passive diffusion across the blood–brain barrier, and so on. These are then input into a model of human physiology to provide quantitative human predictions.

Measurement of low-level properties If animal experiments are being used to measure low-level parameters, which can then be integrated into models of human physiology, can the low-level parameters not be measured directly? A recent development has been to measure "lower-level" properties in *in vitro* assays, and then to integrate

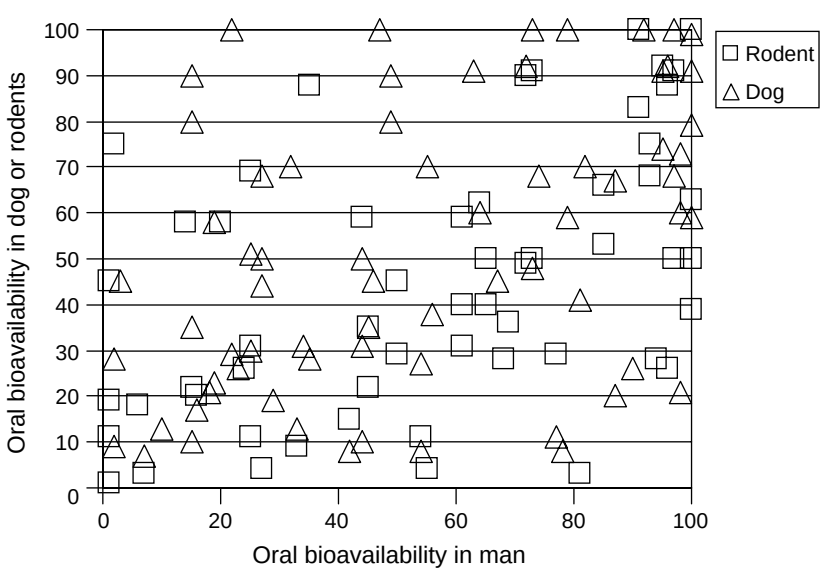

Fig. 2 Relationship between the observed oral bioavailability of individual drugs in man (*x*-axis) and in rat and dog (*y*-axis). Data from Grass and Sinko 2002.

them into an estimate of whole-organism properties. A wide range of surrogate, *in vitro* assays has been developed, many suitable for high-throughput screening (HTS) execution; some are specifically mentioned below. However, none are very satisfactory. Simplistic extrapolation from *in vitro* assays to man is very error prone, and so the lower-level results must be put into a sophisticated mathematical context to give useful predictions in man.

Prediction *in silico* The third approach therefore seeks to predict the ADME properties of molecules by computer (*in silico*) alone, or in combination with the results of HTS ADME assays. Prediction of a molecule's physiological properties before it has even been synthesized is ideal, allowing only molecules with a high chance of clinical success to be made at all.

Substantial advances have been made in determining the structure of the enzymes of drug metabolism, and this aspect of ADME will soon be amenable to "classical" structure-based drug-design techniques. This article therefore focuses on computational approaches to the more complex problems of absorption and distribution and the related issue of penetration of compounds across the blood–brain barrier. There is currently little work using computers to predict excretion *per se*.

2 Approaches to Modeling

The interaction of chemicals with other well-defined molecular targets can be modeled using one of the well-known molecular modeling techniques. This has limited relevance for ADME prediction. Unlike pharmacological properties, ADME properties are the result of interactions of the drug with a very large number of molecules in the body, including macromolecular complexes, the extracellular matrix, as-yet uncharacterized molecular pumps, and highly promiscuous active sites such as those for CYPs and P-gp (discussed below), whose structures are not well characterized. Even if the identity and nature of the "target" proteins that give rise to an ADME endpoint were known, ADME is the result of interactions with many such "targets"; so any results from molecular modeling on single targets must be integrated to give a physiological result. Because the interactions between them are highly nonlinear, any small errors in chemical prediction can cause extremely large errors when input into the physiological model. So a variety of other approaches are used to link low-level properties of molecules with their higher-level physiological effects. These can be broadly classified as statistical or mechanistic models.

Any model that posits a mechanism and then derives a mathematical description of it to predict the outcome is mechanistic. Docking models are a molecular-level example of mechanistic models: they seek to reflect the binding of particular molecular shapes to each other in mathematics. The Michaelis–Menten equation for enzyme kinetics or the computer models for weather prediction are also mechanistic – the equations reflect what happens in the real world. In physiology, these are often called *physiological models*.

Statistical models connect an observed starting position with an observed outcome using equations that can be shown statistically to predict that outcome but without (necessarily) reflecting the "real" mechanism involved. Public opinion polls

are of this sort – they can accurately predict consumer preference for candy bars, holidays, and leaders without attempting to model the human mind. Statistical models are sometimes called *black-box* models: we treat the model (and the process that it is modeling) as a "black-box" into which we put starting conditions and from which we get answers. In mechanistic models, we can look inside the box to see the mechanics.

For ADME modeling, we can apply either type of model to high- or low-level components of the overall physiology. There are five options for combining these approaches:

I A mechanistic model could be built of the whole process on the basis of the properties of the molecules involved.
II A statistical model could be built of subprocesses on the basis of the properties of the molecules involved, and the subprocesses are then linked mechanistically into the whole process.
III Experimentally determined parameters for subprocesses could be linked into a mechanistic model of the whole process.
IV Experimental results for subprocesses could be linked into a statistical model of the whole process.
V A statistical model can be built of the whole process.

These will be referred to as Type-I to Type-V models below (this is an arbitrary classification).

As discussed, Type-I models are impractical. ADME is a complex interaction of many molecules and processes, some of which are unknown, most of which cannot be modeled accurately. So, a truly mechanistic model of how a drug interacts with a whole animal is many years in the future. Most supposed Type-I models are in fact Type-II models.

2.1 Descriptors

To apply any of the strategies discussed in the previous section, we need to describe the molecule to the computer. Molecules as drawn by chemists are convenient for human perception but not for computational analysis, so we have to identify features of the molecule that can be quantified, such as their molecular weight or the number of specific groups. Such numbers are termed *descriptors*.

A critical aspect of any modeling process is defining appropriate descriptors – successful modeling usually results from the correct combination of descriptors and modeling approach, not from either alone. In their search for new descriptors, researchers have examined ever-more complex (and hence computationally "expensive") ways of analyzing a molecule. This enhances knowledge of what descriptors are useful: fast approximations to the best descriptors can then be developed, resulting in a useably fast method.

The most "mechanistic" approach to describing molecular structure is to model the structure on the basis of our understanding of the properties of the atoms that make it up. This approach is broadly called *computational chemistry*, modeling the properties of molecules themselves. Computational methods include molecular mechanics and quantum mechanical methods.

Molecular mechanics treats atoms as points linked by bonds, and optimizes the bond length and point positions to known values of bond length and angle, and Coulombic and van der Waals' interactions

between the points. This requires accurate parameterization for the atoms and bond types involved, derived from X-ray crystal structures and spectroscopic data from known compounds, but this is not limiting for druglike molecules. Molecular mechanics is useful in ADME prediction for optimizing the likely 3D conformation of a molecule for subsequent 3D shape analysis.

Quantum mechanics calculations treat the electrons and nuclei of the molecule as independent entities. Because the Schroedinger equation (the fundamental description of any quantum mechanical system) is effectively unsolvable for molecules of more than two particles, a number of simplifications are used. These include semiempirical methods (which make a number of simplifying assumptions and then correct for the errors introduced with data from experiment), density functional models (which calculate the effects of electrons on each other in terms of a simple model of electrons in a "gas"), and MP2, which uses a combination of simplified ground-state and excited-state wave functions to approximate the complex state in a molecule. Quantum mechanical calculations enable electron densities, dipoles, and energies of the highest-occupied or lowest-unoccupied molecular orbitals (HOMO and LUMO respectively) to be calculated, and some ADME calculations have used these parameters. Commonly used quantum-derived measures also include polarizability, polarity, and dipole moment.

The other approach that uses "rigorous" knowledge is the expert-based approach, where human expertise identified specific structures as having ADME relevance. This is rarely very helpful, except in the general detection of molecules as "druglike" (see below). There are a limited number of widely used basic drug structures, and they have some specific ADME properties that can guide the expert in their evaluation. However, as the thrust of pharmaceutical research is to generate genuinely new molecules, by definition these will not be very similar to old ones.

Many conceptually simple parameters are used in ADME modeling. For, example, molecular weight (MW) is a very simple and surprisingly useful descriptor, probably because it is mechanistically linked to diffusion kinetics and to polarizability. Other simple parameters include the following:

- *LogP*. This is the log of the partition coefficient of the molecule between water and octanol and is a measure of how lipophilic the molecule is. There are dozens of algorithms to estimate LogP. The most commonly used is CLogP, a value calculated by the Daylight software using a method of counting the number of a small number of specific, small fragments in the molecule and adding their observed LogP contributions to a set of correction factors to generate the estimate of the molecule as a whole. CLogP has effectively replaced LogP in the ADME prediction arena as a measure of the lipophilicity of a molecule.

- *LogD*. The distribution coefficient, LogD, is a partition measure corrected for the ionization of the species in water (and hence is pH sensitive).

There are a variety of closely related methods, such as the "lipoaffinity" measure that does not include O or N atoms, that give slightly better performance for predictions of blood–brain barrier (BBB) penetration.

- Other partition coefficients are also used, notably LogPcyc, the ratio of LogP to the partition between cyclohexane and water.

• *PSA*. Polar surface area is a measure of how much of the interactions of the molecule with the solvent are likely to involve polar groups. A molecule has no "surface" *sensu stricto*: the surface concerned is the outer van der Waals radii of the atoms and so is dependent on conformation. Thus, a full computation of PSA should involve dynamic simulation of the molecule for several tens of picoseconds and sampling all the conformers that result. Some workers have found that calculating PSA for a single (unstrained) conformer ("fast PSA") is nearly as accurate and works as well as full dynamic PSA for ADME calculations.

These are descriptors of the whole molecule. Other commonly used descriptors seek to characterize the presence of specific chemical structures in the molecule. Looking for the presence of one chemical group on its own is rarely very useful for ADME prediction, although it is a powerful method for eliminating compounds that are toxic or reactive, and is often used as an initial filter to remove such compounds automatically from large, automatically generated libraries of synthetic candidates (for example, as implemented in Vertex' REOS software – Rapid Elimination Of Swill).

While single fragments or groups are not very informative, collections of them taken together can be more useful. Molecules can be characterized by an array of descriptors relating to the presence of many different fragment types. For ADME prediction, the presence of several types of groups is counted and the counts combined. Counting methods can count any of

- atoms;
- atoms in a chemical context (thus, for example, distinguishing carbon sp^2 or sp^3 hybridized contexts ("single" or "double" bonded), carbonyl, amide, and so on);
- bond types;
- simple groups defined by connectivity (i.e. which atoms are connected to which, without regard for their spatial arrangement, also sometimes called *topological descriptors*), as is done by the software developed by Amedis Pharmaceuticals;
- known chemical groups (phenyl, pyridyl, indole, or more complex groups such as β-lactam and benzodiazepine).

The last two of these categories – counting topological fragments and known chemical groups – are known as key-based approaches. They produce a long string, called a *key* or *fingerprint*, where each bit represents whether a specific fragment is present or absent from a molecule. The fingerprints of different molecules can then be compared very quickly and the presence of specific subfragments determined. Fingerprint-based methods have been found to be generally better as a basis for searching databases of molecules with associated pharmacokinetic data than whole-molecule descriptors (i.e. molecules with similar fingerprints are more likely to have similar pharmacokinetic properties), which has encouraged their use in ADME modeling.

Approaches based on counting fragments and molecular groups were pioneered by CASE and MultiCASE, initially applied to toxicity prediction. The fragments need not be what a chemist would consider a distinct portion of a molecule but can be defined solely by the fact that the atoms in it are connected (see examples in Fig. 3.) Amedis Pharmaceuticals and Topliss et al. are unusual in relying almost exclusively on counting descriptors,

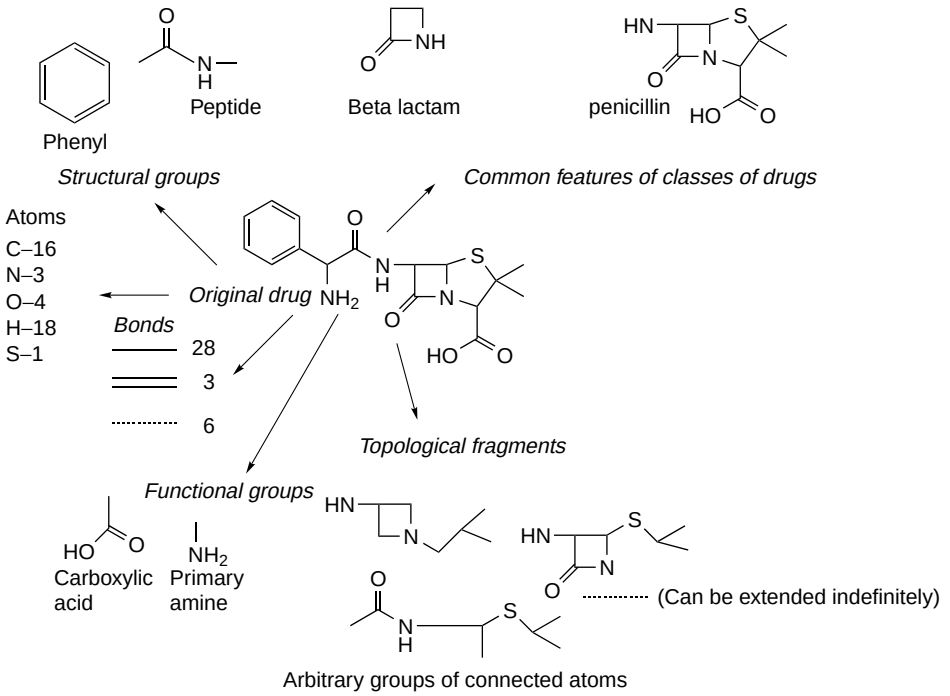

Fig. 3 Atoms, functional groups, fragments. Illustration of the types of way that a single drug structure (center) can be decomposed into (clockwise from top left) recognized chemical fragments, counts of atoms or bond types, groups with specific chemical functionality, topological fragments (groups of atoms that are connected in the original molecule, but which do not confer any special chemical significance), and fragments that are characteristic of drug classes.

together with highly nonlinear modeling methods, for ADME prediction.

If a "fragment" is not intuitively obvious, then how do you define what a fragment is? There are hundreds of ways of conceptually fragmenting a molecule, and even automated methods can give different results depending, for example, on whether phenyl rings are coded as bonded with true aromatic bonds or a Kekule structure. In practice, then, fragmentation is part of a specific modeling process, and the final algorithm must define what the fragment is and how to look for it in a molecule.

Counts can be combined into connectivity indices, which attempt to capture the "style" of the molecule – extended or compact, with or without multiple ring systems, and so on. A simple shape index can count the fraction of atoms connected to others (other than hydrogen) by one, two, or three bonds. Other more complex, information theory–derived measures measure the "randomness" with which the components of the molecules are assembled. All these are very rapid to compute and are based on 2D structure only.

A related and commonly used descriptor is the electrotopological state index computed by Molconn-Z. This combines a topological state index (which is a measure of the connectivity of the molecule)

with terms for electronic interactions of the atoms. Each atom type is assigned a valence electronegativity, and this affects connected atoms to an extent diminished by distance.

CLogP (and many other LogP algorithms) and PSA calculations are fragment-based methods. For CLogP, the algorithm contains a database of fragments and their contributions to LogP. CLogP adds the contribution to LogP from a small number of basic fragments, derived from measurements, and then applies correction factors related to other structural features.

Related to the presence or absence of specific groups is the idea of chemical similarity. This also describes a molecule, but in terms of how similar it is to another molecule. This is usually achieved by a scoring matrix, where each of the identified groups in two molecules is compared, and the match is the number of co-occurring groups divided by the total number of groups. The Tanimoto coefficient is one such similarity measure.

A related idea is that of "chemical space"–plotting molecules according to chemical characteristics in a space in which the dimensions are the descriptors of the molecules. Neighbouring molecules in chemical space are more similar than distant ones. Many statistical methods discussed below seek to partition molecules in chemical space into groups that reflect ADME properties: whether this works depends on what the dimensions of the space are, that is what the descriptors are.

Such simplistic models are not widely used in pharmacological structure–activity relationships (SAR) because (apart from computational chemical indices such as dipole moment) they do not capture the 3D shape of the molecule, which is clearly important for its interaction with a receptor. For ADME properties, 3D SAR has been found to be generally less useful. However, some 3D descriptor sets have been successfully applied. Ekins et al. have done a number of studies using the CoMFA technique implemented in Sybyl. In CoMFA (comparative molecular field analysis), the surface features of a molecule in space are determined by placing it on a grid and then imagining that a test atom was at each point of that grid, and calculating the steric and electrostatic interactions between the test atom and the molecule. The result is a series of numbers that describe key aspects of the molecule's shape and electrostatic characteristics. Other simpler calculations look at aspects of the molecule such as "shadow area" (the maximum area of shadow a model of the molecule would cast on a plane as it rotates) and ellipsoidal parameters (how far it is from spherical).

These are all quite well established methods, and no new "ADME-specific" descriptors have been developed for some years. There are indications that a "biological similarity score" – scoring how similar a molecule is to the aggregate of all small molecules in the body – might be valuable: however, no robust test on this has been performed yet.

Many of these descriptor types are calculated by industrial or commercial programs. A selection of the more specialist programs used in ADME descriptor generation and modeling are listed in Fig. 4. "Standard" descriptors such as PSA, molecular mechanical descriptors, and molecular conformation are calculated by all professional molecular modeling packages.

Program	Functions	Selected reference	Web address
QSAR programs, and programs that generate descriptors			
VolSurf	Converts 3D interaction energy grid maps into a variant of descriptors, specifically targeted at ADME prediction	Cruciani et al. (2000)	http://www.moldiscovery.com
CLogP (Part of the Daylight suite)	Most popular program for calculating LogP	Hansch and Leo (1979)	http://www.daylight.com
Molconn-Z	Calculate connectivity and eletrotopological state indices	Keir and Hall (1999)	http://www.eslc.vabiotech.com/molconn
Absolv	Calculates solubility and OB using "Abraham descriptors"	Abraham and Le (1999)	http://www.sirius-analytical.com
Cerius	Wide-ranging modeling environment, includes a range of SAR modules including quantum mechanics, pharmacophore, fingerprint and GA	-	http://www.accelrys.com/cerius2/
Pharma algorithms	Range of fragment-based modeling and ADME-prediction tools	Zmuidinavicius et al. 2002	http://www.ap-algorithms.com
Sybyl	Chemical database with a variety of SAR methods, including "CoMFA"	Kim et al. 1998	http://www.tripos.com
Programs that perform ADME prediction			
MULTIcase	Fragment-based SAR program, used industrially, primarily to predict toxicity, but also applied to other ADME prediction	Klopman (1984)	http://www.multicase.com
GastroPlus	Physiologically based simulation of gastrointestinal absorption	Agoram et al. (2001)	http://www.simulations-plus.com
MetabolExpert	Expert system-based prediction of metabolism and metabolites	-	http://www.compudrug.hu
Meteor	Rules-based prediction of metabolism and metabolites	Langowski and Long (2002)	http://www.chem.leeds.ac.uk/luk/meteor
iDEA	ADME simulation system, primarily focused on OB prediction, based on large database.	Timony et al. (2000)	http://www.lionbiosciences.com
QikProp	Calculation of several molecular properties including solubility and CaCo-2 permeation	Jorgernssen and Duffy (1999)	http://schrodinger.com

Fig. 4 List of available computer programs mentioned in the text.

2.2 Approaches to Building Models

A mathematical model links a description of the molecule – a list of descriptors – to the endpoint. As discussed, there are two general types of model, mechanistic and statistical. To create either, the form of the equation must be defined, and then the parameters in the equation given specific numerical values. For mechanistic models, the form is predefined and the "only" problem is to parameterize it so that it gives the correct numerical result. For statistical models, the form of the equation must often be discovered as well.

2.2.1 Types of Models: Mechanistic Models

Type-III models (purely mechanistic models based on measured values of low-level parameters) are well known. For this to be deemed truly mechanistic, the model must accurately reflect the mechanism (not just the performance) of the *in vivo* situation.

These can include single-parameter models where one *in vitro* surrogate assay measurement is correlated with an observed *in vivo* effect. An example is the widespread use of the CaCo-2 colon carcinoma cell line as a model for intestinal absorption. More complex Type-III models seek to take low-level pharmacokinetic parameters measured in animals and *in vitro* and predict the human pharmacokinetics from them, or take human data from one dose and extrapolate it to other doses or other patient groups: this is the conventional development of a pharmacokinetic model for a drug. Such models can consist of a set of differential equations that describe the rate of change of the system or a simulation of the system.

The rate of change of compound concentration in these models depends on many things – apart from the starting concentration of compounds, these can include physiological conditions, rates of bulk flow, surface areas and enzyme activities, all of which may themselves change. The set of rate equations can therefore be very complex and not always amenable to formal algebraic solution. It is common therefore to solve them by simulation – taking the starting position of the system, calculating the rate of change on the basis of that starting position for all the components, making those changes for a very small time step (milliseconds in the case of physiology, femtoseconds in the case of molecular modeling), and then recalculating the conditions of the whole system by adding those changes to the starting conditions. Such simulations can, if necessary, include the fact that molecules are discrete things, and cannot be divided, so the rate of flux through a channel can be one molecule or two molecules per time step but not 1.5 molecules. Such modeling can be by Monte Carlo simulation, where the starting conditions are varied at random according to known variation in the inputs to what is being modeled – thus one run might assume the patient was 67 kg in mass, then next that they were 75, the next 82, and so on. A succession of such runs, producing different outputs, will give a measure of the range of predicted results as well as the average.

Clinical trials design software performs a similar function, designing trials that will optimize the information arising from a clinical trial based on prior knowledge of ADME properties. This is amply discussed in standard pharmacology textbooks and will not be mentioned further here.

How mechanistic models work depends on the mechanisms believed to be

involved. (We should note that the mechanism believed to be involved is sometimes not the correct one, but the model can work nevertheless.) By contrast, all statistical models have some common properties and fall into different classes.

2.2.2 Types of Models: Statistical Models

If we do not know what form of equation links the descriptors to the endpoint, then there are a variety of statistical approaches that can provide this for us. The result is a model that is statistically valid but does not claim to reflect the underlying mechanism of the modeled process. The modeling process must select which descriptors to use, and how to link them to provide a final answer. Often, the choice of descriptors is made by the scientist, which reduces the problem (but which assumes that the scientist has made the right choice!)

The relationship between a set of descriptors and a set of endpoints is developed using a number of examples – the training set. These can be inherently involved in the modeling process or applied by the user afterwards, approaches that are called *supervised* or *unsupervised* respectively. Supervised learning techniques, such as neural nets, develop an algorithm on the basis of information about what the "correct" answer is. Unsupervised techniques develop algorithms based on the structure of the descriptors alone.

Such statistical modeling is also called *quantitative structure–property relationship* (*QSPR*) *mapping*, a variant on quantitative structure–activity relationship (QSAR) (activity). The two are exactly the same in that the activity is a subset of all the properties that a molecule has.

QSAR is a well-established set of techniques for pharmacological targets, which identifies key aspects of the drug molecule that must be present to interact with the target protein, and hence by inference the key aspects of the target protein that interact with the drug. For ADME, modeling is not so clearly related to the structure of a protein target. However, the many tools developed for QSAR in drug pharmacology discovery have been used in QSPR discovery in ADME.

2.2.3 Problems with Modeling and Overcoming them

For complex statistical modeling problems, three types of problem are common.

- *The curse of dimensionality*. There are many inputs to a problem (in our case, as many descriptors as we can imagine) and an effectively infinite number of combinations of them. How can the method sort through them for the few that do not only match the training data but can also generalize to new, "unseen" test data? (This is also related to "overfitting".)

- *Multiple local minima*. There are many "OK" solutions to a problem but few very good ones – however, it is usually not possible to adapt many of the OK ones to be a very good one. So how do you avoid the algorithm from being trapped in an "OK" solution, a local minimum in the solution space?

- *Overfitting*. A relationship between any number of descriptors and an endpoint can be derived with enough effort, but usually such a solution will be specific for those examples and will not apply more generally – it will not "generalize". How can overfitting and lack of generalization be traded off against modeling power?

There are many statistical techniques that can be used successfully to overcome these problems and match ADME data to molecular descriptors. This article will not review them all, except to mention that simple

nonlinear mapping methods have been very widely used and are often as effective as very sophisticated techniques. However, six types of more advanced techniques are widely used in the ADME research literature and will be summarized here.

Two well-known linear statistical techniques, partial least squares (PLS) and principal components analysis (PCA), have been widely applied to ADME datasets.

PLS (partial least squares) assumes that the endpoint is a linear function of the descriptors and adjusts constants in that equation to minimize the square of the differences between predicted and actual values. Simple versions of it are commonly implemented in spreadsheet and other statistical programs to provide "lines of best fit".

PCA (principal components analysis) performs a similar task. It plots the distribution of molecules in chemical space and then rotates the axes of that space so that as much as possible of the variability lies in the first dimension, as much as possible of the remaining variability lies in the second dimension, and so on. The new axes are called *principal components* and are linear combinations of the original axes. This then is a method of transforming the original descriptions of the molecule to focus the variability present in the original descriptions of the molecule into a smaller number of descriptions. We then observe whether the variability in the resulting reduced space is related to differences in biological response (i.e. whether there are simple shapes that we can draw in the transformed space to cleanly separate "actives" from "inactives").

2.2.4 Nonlinear and "Artificial Intelligence" Approaches

A limitation of PLS and PCA is that they are linear techniques – they only discover solutions

$$\text{Solution} = \text{A}\cdot V_1 + \text{B}\cdot V_2 + \text{C}\cdot V_3. + \cdots \text{ etc}$$

where V_x are variable parameters and A, B, and C are constants. The techniques seek to find what the useful values for the constants are (if possible, determining that they should best be left as zero, i.e. that variable is not relevant). This is limiting, in that we might expect our endpoint to be dependent on products of parameters, powers of parameters, or other nonlinear functions. Nonlinear techniques could produce solutions of the sort

$$\begin{aligned}\text{Solution} = {} & \text{A}_1\cdot V_1 + \text{A}_2\cdot V_1^2 + \text{A}_3\cdot V_1^3 \\ & + \text{A}_4\cdot V_1^4 + \cdots + \text{B}_1\cdot V_2 + \text{B}_2\cdot V_2^2. \\ & + \cdots. \text{ etc.}\end{aligned}$$

While there are a limited number of possible solutions for a linear function, there is a far wider space of possible nonlinear functions. Linear techniques usually come supplied with reliable and rapid one-step methods for finding optimal solutions – with nonlinear methods, the solution techniques are invariably less reliable and more computationally expensive, so searching them thoroughly is effectively impossible. In addition, nonlinear techniques have an increased risk that any model resulting from a nonlinear framework has "overfitted" to the training data it was presented with. So there is a trade-off in using more complex, nonlinear techniques, balancing the power of the method to discover better solutions against the risk that they find a solution that only

fits the data we have in hand, and will not generalize.

A powerful and increasingly widely used modeling method for nonlinear modeling is support vector machines (SVM). SVMs are related to PCA, but rather than partitioning data in the chemical space ("feature space") directly, they partition it in a feature space that is to an extent nonlinear (although all of its boundary segments are linear) related to the input data space. Using Kernel functions, this can be done quickly without doing computation in that high-dimensionality space, which is claimed to overcome the computational expense of the curse of dimensionality (although not the problem of overfitting). A variety of SVM software is available for free download, which has contributed to its rapid take-up. SVMs have been used for carcinogenicity prediction, as well as QSAR, DNA sequence analysis, and some ADME problems.

Also widely used is a set of techniques called *recursive partitioning*, which can be highly nonlinear. Here, the data set is partitioned using one variable into two sets that differ for the endpoint. Each set is then partitioned again, but the criteria for partitioning can be different for the two sets. Each of the resulting sets is then partitioned again, and so on until fairly uniform sub-subsets are found.

Neural nets are a matrix method modeled on how nerve networks work and are very effective at "learning" complex patterns in nonlinear problems. They have been applied to conventional QSAR problems. They are characterized by their connectivity – how many input nodes, hidden nodes, and output nodes they have. In general, nets with a large number of input or hidden nodes are very hard to "train" in a way that generalizes to data other than the training set, and attempting to train neural nets on very large number of input descriptors (>50) is extremely difficult.

More effective means of training neural networks include prior selection of inputs using genetic algorithms (GAs) (see below) or Monte Carlo/Hastings techniques, which overcome (to some extent) the limitation of training a net using very large descriptor sets.

Genetic algorithms have no such limitation. They "evolve" a solution in the computer. The problem is described to the machine in numbers (as with other techniques), and the computer generates a solution through a process analogous to natural evolution: a number of possible solutions are generated, tested for their effectiveness at predicting known answers, and the best are retained, recombined, and "mutated" to generate new solutions, which are then tested again, and so on. While this process is very simple in concept, there are many parameters that need to be optimized for its application to a specific problem, including the size and structure of the population of problems, the nature and relative frequency of the genetic operators (mutation and/or recombination), how the fitness function operates, and the overall form of the solution. A type of GA called *genetic programing* (GP) overcomes this by evolving the form of the solution as well as its parameters, but at the cost of requiring yet further optimization parameters to be specified. GAs are particularly efficient at exploring enormous "problem spaces", where the number of parameters to the solution is huge and hence the ability of conventional statistical methods to explore them all is very limited. However, their application to life sciences has been curiously limited. They have been applied to molecular conformation analysis, ligand docking, and *de*

novo structure design (where a new structure is "evolved" to fit the criterion that it fits a known pharmacophore or docking site).

GAs have been used in pharmacophore mapping and in "evolving" new compounds to fit a known target site. Their use in ADME prediction is more restricted to a genetic algorithm for generating pharmacophores included in the Cerius2 software, some isolated research "demonstrations", and the application of GPs to ADME prediction by Amedis Pharmaceuticals.

Both ANNs (artificial neural nets) and GAs are often called *black-box* methods, methods that produce a solution, but the form of that solution and the method by which it was produced is opaque to the user. Hybrid forms, where a GA is used to pick the variables used to input into an ANN, are also known.

Complex, nonlinear methods such as GA have a termination problem – how do you know when the solution you have found is good enough? Usually, this is done by reference to a test set; however, this has statistical problems that are discussed below.

2.3 The Problem of Data

With the exception of *ab initio* quantum chemical calculations, all modeling of the properties of potential drug molecules is based on data about the modeled parameter from other drug molecules, and as rigorous *ab initio* quantum calculation of the many-particle body that is a drug molecule is impossible, this means that all modeling is based on experimental data. The importance of good data on which to base models cannot therefore be overemphasized. What that data is depends on the type of model being developed.

If the model is based on fundamental chemical properties (a chemically mechanistic model in the terminology above), then the data are the modeling parameters for the chemical modeling software, and as this is independent of the biological context of the chemical, it is also independent of any requirement for biological data. However, few of the modeling results below are like this: nearly all are developed by reference to actual biological results. Even such highly mechanistic models as the Abraham model of intestinal absorption have to be "parameterized" (i.e. have the constants in the equation adjusted) by reference to actual biological data. So real experimental measures of the parameters that we want to model are critical.

This is a problem because these data are scarce, because they are expensive in time and resources to acquire – which is why we want to model those endpoints instead of measuring them. Existing datasets are usually biased in terms of the chemicals examined and the range of endpoints that are found. For example, very few measurements of OB include drugs with very low OB, but it is precisely the modeling of very low OB compounds (so that they can be avoided) that we are interested in. Data is also always "noisy", in part because the difficulty in accurate measurements of complex ADME endpoints introduces experimental error, in part because the variation in biological systems introduces real variability, and in part because different groups measure complex parameters in different ways, with different results. The result can be enormous ranges in apparently quantitative measurements (see Fig. 5). In addition, compilation and transcription of the data is itself error prone,

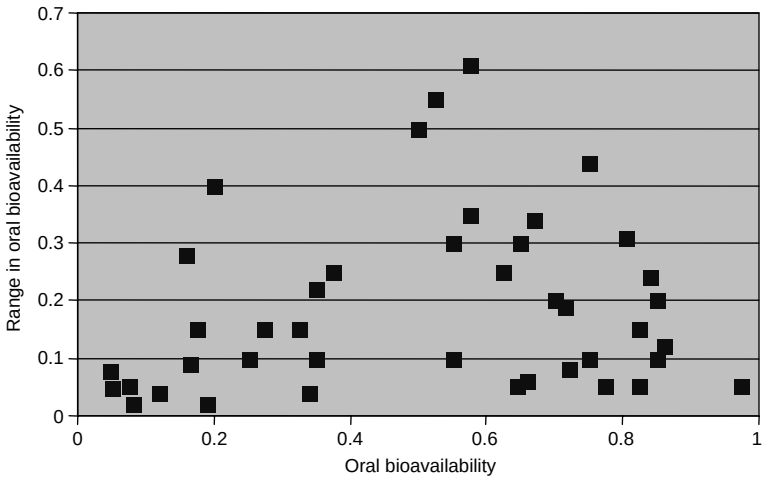

Fig. 5 Plot of the range of reported human oral bioavailability values for 50 drugs for which a range is described. *x*-axis – OB values. *y*-axis – range in OB values. Data from Goodman and Gilman, and Dollery.

so databases can be even noisier than the data from which they are derived. White and Sverlenko found several errors and a number of duplications of data in apparently carefully compiled data sets for blood–brain barrier penetration, which had been widely used in literature studies.

Thus, a major part of any modeling program is actually compiling the necessary data, and as a result, the databases themselves are valuable. Most ADME modeling results use previously compiled data, often published in the literature, as the cost of generating enough data within one research program would be prohibitive. An exception is the metabolism prediction work done by Cyptotex, who run their own screens to generate consistent data for CYP metabolism and CaCo-2 permeation. This will not be discussed further as there is no "magic technique" for getting good quality data by experiment or literature; it is a matter of effort and scientific rigor.

2.3.1 Datasets and Statistical Reliability

All the statistical modeling techniques seek to fit a mathematical model to a set of examples collected, as above. However, we do not actually want an equation that represents the data we already know – we want one that predicts the desired endpoint for a compound we do not know about. The model must be able to generalize to new compounds. As mentioned above, it is easy to use powerful statistical matching methods to fit a complex equation to what amounts to a set of points on a graph; however, there is a point when the equation becomes overfitted – it fits the known data extremely well but does not generalize.

To avoid this, it is essential to split the data into at least two sets – a training set, which is used to optimize the algorithm, and a test set, which is not used as part of the training process but which is kept separate until training is over, when it is used to see whether the resulting algorithm can predict results for molecules that were not used in its construction.

Division into training and test sets can be done by any method, providing that method is not related to the training itself.

Ideally, the training and test sets should be of similar size; however, when data is limited, it is common practice to make the training set larger than the test set. The problem with this is that, if the result is poor, it can be blamed on the fact that the small test set happens to include an unrepresentative collection of molecules; the temptation is then to select a new test set and train again. This is invalid, as the way you select the test set is influenced by the training result, that is, it is not actually independent. A similar invalidity arises when "outliers" are identified from the results of training, excluded, and then training performed again.

An only partially valid approach is the "leave-one-out" approach, where one molecule is left out of the training set, training performed, and the match of the one excluded molecule is determined, then another left out and the process repeated. Overall, there are no molecules that can be said to be independent of the training process, although for each test, the "test set of one" is independent. Such "leave-one-out" statistics are common in ADME prediction as they allow maximal use of the sparse data sets in training.

Evolutionary techniques have an additional problem of when to stop, as mentioned. Termination can be directed by performance against the test set – when the models being generated stop generalizing, you know you have gone too far. However, this means that you are using the test set to control how you do the modeling – it is not truly independent. So for nonlinear, data-driven methods, a third set of data is needed that is not involved in the training process at all, with which to validate that the result produced is general. It is rare for any ADME studies to use this level of rigor, because there is too little data to split it into three data sets.

A related concept is shuffling the observed values between different molecules and seeing if the training algorithm still claims to a good model. If the target values that the method is predicting are actually predicted by the molecular structure, then the shuffled results should give no training. This is used more to validate that the training method is producing meaningful answers than to confirm the generality of one of those answers.

2.4 Conclusions on Modeling

Modeling, therefore, is a balance between the amount of data available, the relevance of the descriptors, and the power of the modeling technology. Enhancing the modeling technology power has risks as well as benefits, so it is always preferable to find more data (if it is available) or use more relevant descriptors (if you can invent them). A strategy for developing ADME models, or indeed any other statistical model, is to start with simple modeling techniques and move to the more complex ones only when the simple ones have proven inadequate.

3 Mode of Deployment

These technologies are applied in one of three broad ways. In all cases, computer-based approaches are used to support decision-making, rather as decision methods on their own.

- *For library selection*. A decision that the medicinal chemist has to make early in the lead development phase of

chemistry is which "lead" to use as the basis for development of focused libraries of compounds for further testing and development of a SAR. Predictive ADME programs can be used against all the possible compounds in a "virtual library" to prioritize libraries for the future medicinal properties of any lead that will emerge from them. This requires only rough categorization of individual compounds to be effective.

- *For synthesis selection*. A subsequent step to this, having chosen a library scaffold, is to select the subset of the possible library that will be synthesized. This requires accurate classification or semiquantitative prediction of endpoints.

- *For compound selection*. When faced with a small number of candidates for preclinical workup, fully quantitative models can be used to prioritize which to take into this expensive and time-consuming stage of development.

The majority of users at the start of the twenty-first century see predictive ADME in the first of these roles – as a coarse filter in early chemical design. This means calculating results for thousands or millions of molecules.

This has important implications for the speed of the model. We must distinguish between generating the model (which can take months or years) and delivering the model. A model delivered into the hands of chemists must be able to help their decision-making process and so must be able to process all the molecules that they want to process overnight or faster (and ideally as fast as they can generate them). The limit on algorithm speed is usually descriptor calculation (as the algorithms themselves simply combine the descriptors in simple equations that can be coded in a few hundred instructions or less). So there is a premium on descriptors that are fast to calculate. Thus, simple counts of chemical groups based on 2D structures are much to be preferred to rigorous quantum chemical calculations or computation of all possible conformers of a molecule.

In all cases, it is helpful if the algorithm being used is interpretable, so that the output of the algorithm is easy to relate to features of the molecule that a chemist can then alter. In this respect, PSA is an excellent parameter – to a good approximation, it is just the sum of the nonbonded "surface" of the N and O atoms and the hydrogens attached to them, so a result that derives from an adverse PSA measure can be changed by "just" altering those aspects of the molecule's chemistry. Few other descriptors are that simple, and when they are, the algorithms that integrate them into an ADME prediction are so nonintuitive that the value of the descriptor transparency is lost.

In practice, a completely opaque black-box algorithm can still be used as an *in silico* experiment, into which the chemist puts new candidate designs and from which he/she obtains estimates of how those designs will perform in reality.

4 Accuracy

Computer predictions need to be accurate, and, equally importantly, we need to know how accurate they are so that we can weight decisions made on the prediction appropriately. In fact, ADME prediction is already as accurate at predicting the human result as *in vitro* surrogate assays (see, for example, Boobis et al., 2002).

However, more is expected of it, as it is being compared not to a chemist's intuition but to actual experimental data, which by definition are 100% accurate even when, as is often the case with animal models of human ADME, they are misleading.

Accuracy is generally described by an R^2 (correlation) measure for quantitative measures, a percentage accuracy for qualitative ones (here designated C_a), or Spearman's rank coefficient, which determines the extent to which the compounds are sorted into the correct order, even if the absolute values for the predicted parameter are not accurate. (Rank ordering is almost as good as absolute accuracy for some applications, where the objective is to find the best N of a set of molecules.) Because many models are data-driven, it is possible to overfit them so that R^2 is very close to 1; however, these models perform very poorly on new molecules that are not part of their training set. A more useful measure is therefore the square correlation with a set of molecules that has not been involved in the training, which is usually designated Q^2. C_a is calculated for an independent test set, not the training set. (We note above that this needs to be an *independent* test set, not a test set that has actually been involved in the training process.)

To compare different algorithms, they must be applied to the same set of molecules, because accuracy also implies accuracy in a specific chemical space, not across all possible chemicals – this is rarely done.

All models divide into local, and general or global models. The former apply only to molecules of a particular chemical type, formally ones that fall into a specific volume of chemical space. Global models claim to predict results for all molecules. In reality, they too are limited, in this context to all druglike molecules: to expect a model meant to predict the properties of carbon-based, druglike molecules to generalize to, say, uranium hexafluoride is unreasonable. Thus, it is important to state the chemical space over which a model is valid. Too often this is not done until the model is found to fail to predict known results for a new type of molecule.

An indication that a supposedly global model is actually a local one is when its performance is substantially worse on a new set of molecules, even if they are apparently broadly the same as the set for which it was developed. This is a common observation: the results for several OB models reported by Bains et al. were over 90% accurate for the molecules used in this study, but this fell to 80% for both their model and the other models used as comparisons in the study when run on a different set of molecules. It is not surprising that no model is truly global – as mentioned above, non-druglike molecules are likely to fall outside the chemical space of a model developed for drugs. However, we should note that even within the chemical space of "druglike" molecules, there are many domains, and most models do not address them all equally well.

A related fact is that no model is perfect. Biology is full of nonlinearities and exceptions. For example, all prediction systems for OB and metabolism predict that cyclosporin and erythromycin will be almost completely orally unavailable and rapidly metabolized if injected (unless these systems encapsulate specific "exception rules" about these drugs). In fact, both are well absorbed and have usefully long serum half-lives. It may be that the increasing sophistication of medicinal chemistry leads to this type of highly functionalized molecule becoming common in

synthetic drug discovery as well as in "natural products", in which case many of the models discussed below will have to be rebuilt. So models will have to either include sufficient understanding of biology to predict when complex molecules such as erythromycin will generate "exceptions" or have explicit rules in them to deal with them *ad hoc*. At the moment, there is insufficient ADME data available to allow either of these solutions to be implemented; exceptional molecules just give the "wrong" answer.

With this armoury of techniques available, what models of ADME properties have been built? I will concentrate on models that predict absorption (oral bioavailability) and distribution. Models of metabolism (especially hepatic metabolism) are in flux, and, with the availability of X-ray structures for the CYP enzymes, will soon be put on a more rigorous, mechanistic footing, although they are mentioned below for their contribution to OB. Excretion is still not modeled to any significant extent: it is measured in animals and then extrapolated to man.

5 Oral Bioavailability

Oral bioavailability is defined by the European Committee for Proprietary Medicinal Products (CPMP) as the extent and the rate at which a substance or its active moiety is delivered from a pharmaceutical (dose) form and becomes available in the general circulation. A major target for modeling is oral bioavailability, the extent to which an orally administered drug ends up in the systemic circulation. All types of models (except Type-I) have been developed to predict oral bioavailability.

The barrier between the contents of the gut and the systemic bloodstream is one of the most critical in the body. Animals are exposed to many potentially dangerous agents on their skin, exposed mucosae, and in their lungs, but only in the gut do they deliberately take in potentially poisonous materials in concentrated form, and then treat them so as to release all their toxic contents.

The barrier is in two parts. The cells of the intestinal epithelium form a barrier through which few molecules can penetrate. Those that do are liable to metabolism in those cells. Molecules passing through the epithelium into the blood are collected into the hepatic portal vein that flows to the liver, which contains a more extensive set of xenobiotic metabolizing enzymes. Only molecules that pass both barriers accumulate in the blood. Overall OB is the combined effect of how rapidly, and in what concentration, a drug is delivered to the intestine, transit time in the intestine, dissolution rates and extents, transfer across the gut wall, metabolism in the gut wall, and metabolism in the liver (also called *first-pass metabolism*).

Oral bioavailability is operationally defined as the ratio of the integrated plasma concentration (AUC) of a drug after an oral dose to that after intravenous injection. However, this can be hard to determine as it requires the fate of all the drug that is tracked – this means detecting unabsorbed drug in urine and faeces, and all the metabolites of the drug in blood, urine, faeces, and tissue. It can also be corrected for solubility effects if the solubility in the conditions of the intestine are known. However, these corrections are complex and time-consuming, so usually the OB is calculated simply from the plasma AUC for the "parent" drug.

5.1 Simple Rules for OB

Many studies have sought to find simple Type-V models that are based on a few rules or molecular motifs to predict oral bioavailability. By far, the most generally used is Lipinski's Rule of Five. Lipinski developed four "rules", which are used as a classificatory model of whether a compound will have "good" OB or not. They were derived essentially by inspection of 2,287 compounds with United States Adopted Name (USAN) or International Non-proprietary Name (INN) designations (as opposed to just chemical names), because such names are usually assigned prior to entry into phase II clinical trials, meaning that these compounds had satisfactorily completed phase I clinical trials, which evaluate safe dosage levels and include assessments of ADME/toxicity properties.

The rules are as follows:

1. ClogP < 5
2. Molecular weight < 500
3. H-bond donors < 5
4. H-bond acceptors < 10

The first two are primarily associated with membrane and aqueous solubility. The second two are independently predictive of OB. If a compound breaks any two (or more) of these rules, it has a low chance of being orally bioavailable.

These have the enormous advantage that they are all easy to compute, and all of them have immediate and obvious interpretations for the medicinal chemist. Subsequent extensions have included a maximum number (usually 5) for the number of rotatable bonds (NRB).

The Rule of Five remains one of the most effective *in silico* predictors of OB. There have been many projects to extend the Rule of Five, either enhancing its performance by including more parameters or by turning what is a rather fuzzy classifier into a quantitative estimate of OB. Generally, these have not been taken up: several extensions and optimizations of it have not proven more useful to the chemist "at the bench", although they are statistically more accurate. These include refinement of the definition of hydrophilicity, and quantitative equations that link LogP and hydrogen bonding to a semiquantitative estimate of OB.

To an extent, this is an effect of organizational inertia. The Rule of Five was the first popularized rule set that was (1) easy to compute, (2) interpretable in chemical terms, and (3) backed by a major pharmaceutical company (Pfizer, Lipinski's employer at the time). The Rule of Five is now widely used as a screening criterion for compounds being considered for chemical research, and in Pfizer, they are automatically checked for every compound designed. This has resulted in a reduction in average molecular mass and an increase in hydrophilicity of compounds entering synthesis, a reversal of a 20-year trend away from "druglike" compounds. It is the aim of many ADME prediction software projects to be as widely implemented and as successful as the Rule of Five.

Exceptions to Rule of Five occur. Some are molecules that mimic biological molecules, and hence are the subject of specific transporter systems, such as the macrolide antibiotics mentioned above. Egan et al. showed that other failures of Rule of Five, which is essentially a linear (additive) rule, could be explained by nonlinear combinations of essentially the same variables. For example, highly

lipophilic compounds are rarely soluble – adding an ionizable center would, under simplistic linear rules, make the compounds less likely to have good OB, but because they enhance solubility, they actually enhance OB. However, the number of examples of each exception is too small at the moment to build systematic models from them in their own right.

The only major extension to the simple rule sets for overall OB calculation is the use of PSA calculations in OB prediction. This is primarily a model for intestinal absorption, and not of overall OB, as discussed below. AstraZeneca have examined the role of the Lipinski descriptors in screening out compounds with poor OB, and have found that PSA and NRB are not relevant when MW and LogP filters are used. While in theory, and for a very diverse set of molecules, considering PSA and NRB enhances the accuracy of these rules, in practice, for the molecules being considered by drug discovery departments, all the information in PSA and NRB that is relevant to OB prediction is captured in the other Lipinski parameters, and specifically, a major contributor to PSA is H-bond number, which is already captured in Rule of Five.

Other group counting methods have had limited success. Counting specific chemical moieties does not lead to good prediction of compounds that are orally available, but can be used to eliminate drugs that clearly cannot be orally available but can be generated by automated chemical design programs. This allows the automatic elimination of groups such as phosphate and quaternary nitrogen from compound libraries being designed for drugs that are to be orally available.

A related, but much fuzzier, rule set is that which determines "drug-likeness". The idea of a "druglike" molecule has been alluded to above, but what exactly is it? A "druglike" molecule is one that a chemist would view as being similar in structure to the structure of successful drugs, and this usually reflects a combination of high intestinal absorption (IA) and low first-pass clearance. There has been less success in defining "druglike" in simple algorithmic terms. In general, drugs are not chemically reactive, do not contain a number of groups that are the subject of rapid metabolism (such as phenols), are rigid structures, and have no more than two or three asymmetric centers (this makes synthesis easier, and has little to do with their biology). So far, "druglike" has tended to be a very subjective measure and one that has not been automated in a consistent way.

5.2 Advanced Type V Models for Oral Bioavailability

Further development beyond qualitative guidelines such as the Rule of Five has required more complex mathematical models, breakdown of the problem into simpler, lower-level problems, or both. The conceptually simplest approach is to take similar datasets to that used by Lipinski and apply more sophisticated modeling tools to it.

This has been performed by several workers with modest success. Yoshida and Topliss used a variant of discriminant analysis to derive a model of overall OB that classifies compounds into one of four classes – <20%, 20 to 49%, 50 to 79%, and >80% oral bioavailability. This uses 18 descriptors, including LogD, $(\mathrm{LogD})^2$, $(\mathrm{LogD}_{\mathrm{pH=6.5}})$ to $(\mathrm{LogD}_{\mathrm{pH=7.4}})$, and occurrence of a wide range of molecular fragments such as lactam and benzodiazepine rings that classify compounds

into different classes, and groups such as primary and secondary alcohol, allylic oxidation sites, and esters that are prone to chemical or enzymatic degradation. The descriptors were selected as potentially relevant before modeling. This is a linear method, but a requirement for $(\text{LogD})^2$ is fulfilled by including $(\text{LogD})^2$ as an explicit descriptor. Bains et al. use GP to select an algorithm that combined topological fragment descriptors into a classification into high–low classes or the Topliss four classes. Performance was broadly similar, with the GP-derived algorithm performing slightly better and both well outperforming Lipinski. However, this is probably at the expense of globality of the models, as they rely on detecting fragments that are present in orally available drugs. This is particularly true of Topliss, who explicitly coded existing drug classes into his descriptor library. By comparison, Bains et al. used an entirely automatic technique to generate a very large molecular fragment library, from which the GP systems selected those fragments most relevant to prediction. The fragments that the GP "selects" are very diverse and may in fact reflect simple molecular parameters, an "improved" version of LogP rather than anything more subtle.

5.3 Models of Components of OB

Modeling OB as a target in its own right has had limited success, and in any case, if such a model predicts that a molecule is poorly absorbed, it is not immediately obvious what to do about it. So more attention has been paid to modeling components of OB, which can then be treated separately or integrated into a Type II model.

Widely modeled components are the physicochemistry and physiology of getting solid drug into solution in the gut, the absorption of dissolved molecule from the gut into the blood, and the clearance of the molecule from the hepatic portal blood supply by the liver before it gets to the general circulation.

5.3.1 Gut Physiology, Solubility, pK_a, and LogP

Before the drug molecule gets to the intestinal epithelium, it must get to the intestine and dissolve (not necessarily in that order). Both aspects can be modeled. Modeling gut transit times and volumes is a mature science with a substantial knowledge base. This is a problem of formulating the mechanics of the gut in mathematical terms, and while complicated, it is conceptually straightforward.

5.3.2 Solubility

Prediction of solubility is a very active area of research, reviewed recently by Livingstone. It is complex, as how much drug dissolves in the gut is a kinetic as well as a thermodynamic measure, and the rate of solution depends on particle size, crystal structure, formulation and so on, and on gut physiology.

Clearly, very low solubility will result in poor OB because the concentration of the drug in the intestine will be low and the majority will be excreted undissolved. Poor solubility (meaning a solubility such that the volume of the intestine can only dissolve a very small fraction of the required dose) can also result in very variable absorption, because the amount of drug that is taken up depends on how fast it is removed from the gut and how fast it dissolves to replace that removed, both of which depend on many variables.

Variable OB is a worse problem for the pharmaceutics scientist than constant, low OB, as it is impossible to define a dose that will accurately deliver a standard concentration to the blood.

As well as implications for absorbability, low solubility has other implications for the drug discoverer. Very low solubility in water will mean that the compound will give false results in all assay types (either because it does not go into solution, and so give uniformly negative results, or because it remains in suspension as light-scattering or fluorescent "haze" and produce false positives). Estimating solubility, ideally before a compound has been made, is of substantial importance.

It might seem perverse to predict solubility, when all that the scientist has to do is add some drug to water and see how much dissolves. However, even this requires the scientist to make the compound: it would be very valuable to be able to predict the solubility before a compound is even made, thus helping the scientist to decide whether it was worthwhile synthesizing it in the first place. This is the role of *in silico* methods.

There is a lot of solubility data available in the public domain, but much of it relates to small, nonpolar molecules, whose solubility properties may not be relevant to the larger drug molecules. So solubility data is limited for modeling.

For simple, uncharged molecules, aqueous solubility can be predicted on the basis of LogP, which measures the ability of the molecule to penetrate the hydrogen bond network of water. Hansch has derived the now "classical" solubility relationship between solubility and LogP.

$$\text{Log(Sol in water)} = 1.339\ \text{logP} + 0.978$$

which has an $R^2 = 0.874$ for small, liquid organic molecules. Where molecules have significant intermolecular attraction (charge, polarity, mass) or significant interaction with water, estimating solubility is less easy. If the melting point is known for a molecule, then this can be used as a measure of intermolecular forces of all sorts and can be included in equations for solubility. This requires experimental measurement, which, although can be done fast and on a few milligrams of material, still requires the material to be synthesized. Ran et al. generated an equation of this sort of druglike molecules with a root mean square error (RMSE) of 0.76 log units using a training set of 413 compounds. Abraham et al. apply an equation based on solution thermodynamics but based on calculated parameters (see below) to calculate solubility effectively. This solution is marketed commercially as Absolv.

Just as LogP (a solution phase parameter) can be calculated by summing the contributions of fragments of the molecule, so Klopman and Zhu have built a model from 1,168 compounds using fragment contributions to solubility, getting an R^2 of 0.95. Huuskonen has used a set of 30 topological descriptors (including 24 atom-type electrotopological state indices, two path 1 connectivity indices, a flexibility index, a number of hydrogen bond acceptors (HBA), an aromaticity indicator, and an aliphatic hydrocarbon indicator). These have been used to develop multilinear regression models on a total of 1297 molecules ($R^2 = 0.89$, $Q^2 = 0.88$) and a neural net model, which gives an impressive $R^2 = 0.94$, $Q^2 = 0.92$ (although the "Q^2" values derived for this work are based on "leave-one-out" measures rather than matching to a genuinely independent test set). MultiCASE, a fragment-based analysis system, has a module for predicting solubility on the basis of the contribution of individual fragments.

Calculations based on more complex descriptors, especially electrotopological descriptors, can also be used in this mode to obtain R^2 greater than 0.9, as can quantum calculations. However, this suffers from being a very complex model, and hence one that is hard to interpret. Other simpler neural nets perform more poorly, for example, Liu and So based on a neural net model and 7 input variables, mostly topological. The model gives $Q^2 = 0.82$.

Several of these algorithms have been implemented in commercial programs that seek to predict the aqueous solubility of organic compounds. Most do not fare well for any but very simple molecules, or molecules in a single chemical series (for example, alcohols). A widely used one in drug discovery is QikProp, which uses an eight-parameter model

$$\begin{aligned}\mathrm{Log}(S) &= 3.886 - 0.0194.SASA \\ &+ 0.514HBA + 0.578HBD + 1.343(NH) \\ &+ 1.224(CNO) - \frac{116.HBA.\sqrt{HBD}}{SASA} \\ &+ 0.182(rota) - 0.00405.WPSA\end{aligned}$$

where $SASA$ is the solvent accessible surface area, HBA and HBD are the number of hydrogen bond acceptors and donors respectively, (NH) and (CNO) are the number of amine and amide groups, *rota* is the number of rotatable bonds, and $WPSA$ is the weakly polar surface area, a subset of PSA. These parameters come up as common contributors to IA and BBB (see below), which suggests that common processes are involved in solution and biological barrier penetration, even for drugs when solution is not limiting.

However, global solubility models are possible. Pharmacia developed a model for intestinal permeability using an internal compound set, obtaining $R^2 = 0.79$ for the local chemical space but only $R^2 = 0.39$ for global performance across all their compound collections: this decline in performance is typical of the models cited above when they are applied outside their training domains. By including a representative sample of all those collections in the modeling, the same group built a PLS-based model on the basis of 24 equilibrium solubility measurements made on 1,107 compounds (a substantial investment in laboratory work), and produced a model with $R^2 = 0.92$, $Q^2 = 0.91$. This is accurate enough to provide useful quantitative predictions.

A related physicochemical measure is $\mathrm{p}K_\mathrm{a}$, which affects a drug's charge state at specific pHs and hence both its aqueous solubility and its LogP (and hence its ability to dissolve in lipid membranes). It is generally true that a fractional ionization of >10% leads to low intestinal absorption. Commercial $\mathrm{p}K_\mathrm{a}$ predictors exist, but usually work only for molecules with simple, monofunctional groups, and cannot cope with the complexities of druglike molecules. It is also not clear whether the ionization state is truly relevant. Some find that computing it does not improve prediction of OB, and so, except for eliminating multiple charges (such as on the bisphosphonates, a notoriously poorly absorbed group of drugs) from consideration at all, most workers do not seek to include explicit calculation of charge state in their OB models.

5.3.3 Intestinal Absorption

Once in solution, the candidate drug has to cross the intestinal epithelium. The intestinal epithelium is a physicochemical barrier (as is the blood–brain barrier discussed below). The physical barrier

is the lipid membrane of the intestinal epithelial cells: molecules that pass easily from water into lipid and back will penetrate this readily. Thus, measures of lipophilicity and hydrophilicity (related to hydrogen-bonding potential) are often central to predictions of IA.

There are actually two mechanisms by which a drug can pass through the intestinal epithelium – transcellular (crossing the lipid membranes of the cells) and paracellular (permeating across the tight junctions between cells). Paracellular transport is slower because the area of pores is much smaller than the area of the cells; however, it has different requirements for lipophilicity and is not affected by P-glycoprotein (see below). The molecular weight cutoff for effective paracellular transport is about 300 Da. The effect is that for most drugs the transcellular route is the only one of importance, and hence the drug's IA is dominated by its lipophilicity and its ability to avoid the P-glycoprotein export pump.

Prediction of IA has been achieved using several approaches with an accuracy comparable to the accuracy of the *in vitro* permeation model based on the CaCo-2 cell line (discussed further below).

Passage across the membrane can be modeled using molecular modeling methods because, although the membrane is a complex entity containing a huge number of atoms, it is chemically relatively simple if proteins are ignored. Novartis has worked on modeling the lipid membrane by simplifying lipids into connected ellipsoidal "atoms" that reflect the segments of the molecule to simplify the problem. Dimeristoyl phosphatidyl choline is modeled as 11 units – choline, phosphate ester link, glycerol, two acid ester groups, and three side-chain groups per fatty acid. The distribution of a drug across this membrane is well modeled, and assuming that entry into the lipid phase is determined by bulk water concentration and exit from it is determined by the concentration in the outermost layer, this can be used to estimate the rate of crossing. This has not yet proven superior to simple numerical modeling approaches, but unlike them it is open to refinement as computational techniques improve.

Until such improvements are developed, simpler associations of molecular properties with IA remain dominant.

A major contributor to IA is the molecule's ability to diffuse across the cell membrane – highly hydrophilic molecules such as sorbitol have effectively zero IA. LogP has value for IA and OB prediction on its own, and the role of LogP in the Rule of Five reflects this.

Other measures of hydrogen-bond potential and lipophilicity have been used. PSA is a general measure of the degree to which a molecule can interact with water, and hence its hydrophilicity, and is strongly negatively correlated with IA, with a sigmoidal relationship that gives an $R^2 = 0.94$ for a small drug group, but does not fare so well for a larger group of molecules, generating a Q^2 of between 0.7 and 0.8 in several studies.

A variety of studies have compared the value of fast PSA with dynamic PSA in predicting IA. Some have found that full 3D conformer PSA is better than 2D approximations; others have found that this is not so. The differences in conclusion are probably related to minor differences in algorithms and in the biological data sets used to test them.

A PCA analysis of MolSurf outputs (a quantum chemistry program) found that one principal component made up from the number of HBD, the polarity,

the strength of hydrogen bond–accepting oxygen atoms, the hexane : water partitioning coefficient, and the number of nitrogen atoms correlated well with IA. Molconn-Z electrotopological state indices combined with ClogP and molecular size also correlate to IA. There are a number of other measures of hydrogen-bonding potential, some of which are mentioned below under blood–brain barrier penetration: in general, though, they do not add significantly to the performance of the descriptors above.

More complex modeling methods have also generally identified the same types of descriptors as important for IA. Andrews et al. and Zmuidinavicius et al. use a fragment-based model and a recursive partitioning approach with good results: the Zmuidinavicius algorithm is available commercially from Pharma Algorithms. Jurs et al. have applied a hybrid GA/ANN method with good results. The GA selected 6 descriptors from 162 to be inputs for the ANN: they were the number of single bonds, normalized projection of molecule ("shadow area"), charge on donatable hydrogen atoms, fractional surface area on HBAs, fractional surface area HBA × their charge, and cube root of gravitational index (a density measure). The compound set included actively transported compounds, which is impressive in that such compounds are usually excluded as lying outside the modeled chemical space, but may skew the data set.

5.3.4 Physicochemical Methods

It has long been held that, in the absence of specific transporters, IA is essentially a passive process by which molecules dissolve in the cell membrane from the gut side and dissolve from it into the epithelial cells. The role of lipophilicity is the basis of this hypothesis: IA is limited by how much drug can dissolve in the cell membrane. However, Abraham has shown that IA is not a partition-dominated process (i.e. not one in which the partition of compound into a bulklike lipid phase is important), but a diffusion-dominated one. The role of LogP is therefore misleading – OB is not about dissolving in a bulk phase fast, but diffusing through a barrier.

If IA is essentially a diffusion process, can we not model it as such on the basis of an understanding of the physics of diffusion? Abraham and coworkers have successfully applied this concept using a general solvation equation

$$\log(SP) = c + rR_2 + S\pi_2^H + a\alpha_2^H + b\beta_2^H + vV_x$$

where R_2 is the excess molar refraction, π_2^H is the polarizability/dipolarity, α_2^H is the overall hydrogen-bonding acidity, β_2^H is the overall hydrogen-bonding basicity, and V_x is the McGowan volume. This describes solution of a molecule in physical terms, without consideration of the biological mechanism. In a detailed piece of work based on data for 169 drugs, out of 241 researched (a very substantial data collection exercise in its own right), Zhao et al. showed that appropriate parameterization predicts IA quantitatively with R^2 of between 0.75 and 0.8, providing that OB is not dose limited for large molecular weight drugs (i.e. if the rate of solution is not dominating). The parameters are all calculable in their own right from molecular structure. The "Abraham descriptors" have been implemented in a program ABSOLV, which is effective in OB prediction providing the input (and comparative data) is carefully QCed.

Physicochemical methods of this sort are complementary to computational modeling of the lipid membrane – both are based on an understanding of the physics and chemistry of the process that is more "fundamental" than the biology. Both, therefore, are susceptible to improvement either by gathering more biological data to parameterize the models better or from improvements in the model of the underlying processes. By contrast, statistical modeling methods, which today are at least as effective as mechanistic ones, are limited by data and cannot be improved beyond their availability.

5.3.5 CaCo-2 Model

Because IA is complex and expensive to measure, surrogate measures have been developed. The most widely used is the CaCo-2 permeation system, which Artursson described as a model for both passive permeation and P-glycoprotein efflux, although the latter is very variable between experiments and subclones of the cell line. The cell line is derived from a colon cancer, and a monolayer of cells *in vitro* develops tight junctions between the cells that mimic the intestinal epithelium. Permeation assays measure transport of compounds across a monolayer of cells, using a nonpermeant molecule (usually mannitol) as a referent. The MDCK cell line has similar properties, and is also used sometimes.

Despite the cell biological support for the validity of CaCo-2 as a model for IA, and its widespread use in pharmaceutical discovery, CaCo-2 is actually a poor predictor of IA and is not objectively better than most computational models. It works well for comparing the IA of a series of similar molecules, but does not give consistent results between chemical series or between laboratories. Some of these defects can be overcome by more complex methods that control for monolayer quality and boundary layer effects, but the assay is harder and more time-consuming to do, thus removing a lot of the purpose of doing a CaCo-2 screen.

Modeling CaCo-2 itself might therefore seem futile and only useful for drug discovery when an actual CaCo-2 assay is an unavoidable bottleneck in a drug discovery program, or when CaCo-2 is being used as an assay for P-glycoprotein. However, it is also likely that modeling CaCo-2 will give an insight into how to model IA. CaCo-2 data is much easier to obtain than IA data; so as a method of modeling the modeling process, this has substantial value.

CaCo-2 permeation shows the same type of sigmoidal relationship with lipophilicity as IA, with the shape of the curve being modulated by the molecular weight. Several PLS-based models have been constructed, with typical results for small numbers of molecules being $R^2 = 0.85$ to 0.9 using hydrogen-bond count, molecular weight, and lipophilicity terms (usually CLogP). PSA correlates less well with CaCo-2 permeation than with OB in man. More specific improvements to include the strength of hydrogen bonds from nitrogen and oxygen give R^2 of up to 0.935 on small-molecule sets, but do not generalize. The same is observed for IA, suggesting that only very general descriptors are useful for either prediction.

5.3.6 Conclusion for Passive OB and IA

There are active metabolic processes involved in IA and OB, which I consider below. However, for many drugs, it appears that OB can be predicted on the basis of relatively few simple parameters describing size, lipophilicity, and hydrogen-bonding potential. This is plausibly related to the

hypothesis that, for many drugs, IA is the limiting step in OB and is predominantly a diffusion-limited process. Use of sophisticated modeling technology and descriptors has not so far yielded substantially better performance than the simple linear free-energy relationships of Abraham et al. This suggests that either the ideal descriptors for OB and IA have not been identified yet or (more likely) the idea of a unified OB model has been taken as far as it can be. Significant advance will require molecules to be categorized according to whether they passively bind to or are actively transported by specific proteins, interact anomalously with the lipid membrane, or have other molecule-specific behaviors. The first two areas of specific interaction that are likely to be solved are P-gp interaction and CYP metabolism.

5.3.7 **P-glycoprotein**

P-glycoprotein (P-gp) is an ATP-dependent transporter, and sequence analysis shows P-gp to be a member of the ATP-binding cassette (ABC) transporter superfamily. A major contributor to a drug's IA (and BBB penetration discussed below) is the extent to which it is a substrate for P-glycoprotein. Thus, compounds such as doxorubicin and taxol, which superficially look like good candidates for oral drugs, are poorly orally absorbed because they are excellent substrates for P-gp and so are actively pumped out of the gut epithelium and back into the gut lumen.

Compounds can be substrates or inhibitors of P-gp, and the latter themselves can be (but are not always) transported. So a compound can be affected by P-gp, but can also affect others. This is a substantial issue for predicting drug–drug interactions, where the biochemistry of one drug affects that of another other than by way of their pharmacology.

P-gp is also expressed significantly in the testis, suggesting that it might have a role in preventing reproductive toxicity of some compounds. Members of the human *P-gp* gene family are also known as *multidrug resistance* (MDR) genes, owing to their role in conferring drug resistance on cancer cells: overexpression of P-gp allows the cells to pump out cytotoxic anticancer drugs before they can have an effect and so render the cells resistant to a wide range of drugs that are substrates for P-gp. A similar mechanism involving other members of the ABC transporter family operates in some forms of bacterial antibiotic resistance; so modeling P-gp is of interest outside the domain of ADME prediction.

The mechanism of action of P-gp is unknown. It is hypothesized that P-gp may actually "hoover up" drug from the membrane, which would explain why measures of membrane permeability (LogP/PSA combinations) are such effective predictors of BBB and OB even when P-gp, an enzyme, plays such an important role. However, it is not only mechanism as very hydrophilic drugs such as colchicine are also substrates.

P-gp has been modeled *in vitro* and *in silico*. CaCo-2 assays are used as models of P-gp-mediated efflux of drugs in IA and (as mentioned below) other tissue barriers. P-gp is expressed in CaCo-2 cells, although to a variable extent between laboratories, and the extent to which a drug's passage through a CaCo-2 monolayer is affected by P-gp can be judged by measuring its rate of diffusion across the cells in the presence and the absence of a known inhibitor of P-gp, such as verapamil. Such P-gp measurements can themselves be used as guides to drug design (on a general rule that being a P-gp substrate will reduce OB and so is undesirable) or be used as components of other models of OB.

Substantial work has gone into defining what the substrate and inhibitor selectivity of P-gp is. The enzyme tends to be active on large molecules and has a "preference" for hydrophobic, amphipathic molecules carrying a positive charge at physiological pH and containing a planar ring system. Apart from these general constraints, conventional SAR have found it to be very unspecific, and even these "rules" are frequently violated. Many P-gp substrates are uncharged at physiological pH (e.g. colchicine), and several uncharged cyclic and linear hydrophobic peptides and ionophores have recently been described as P-gp substrates.

MultiCASE has been used to identify fragments ("biophores") that may be important for P-gp substrate status, but it is considered unlikely that these are exhaustive or general. Seelig finds that the enzyme has a general requirement for molecules that have 2 or 3 electron donor groups or hydrogen-bond acceptor groups separated by 2.5 Å or 4.6 Å, but many exceptions exist. Ekins et al. have used the Catalyst QSAR modeling software to generate several different models of P-gp on the basis of different modeling approaches and training sets. All the models have a number of hydrophobic features and most have an H-bond feature; however, what these features are and how they are arranged varies with the molecules used to train the models. Some work using a fragment-based GP approach has produced models of P-gp specificity, but these again are fairly local to the chemical space on which they are trained.

Overall, therefore, it is not possible to do more that estimate the chance that a molecule will be a substrate or an inhibitor of P-gp, and it is not practical to estimate the extent to which it will be a substrate or inhibitor at a given concentration. The input of P-gp into IA, OB, or BBB models is therefore usually as a measured input parameter. Whether further structural information on P-gp will improve this is not clear.

5.3.8 Other Transporters

Specific transporter proteins are responsible for efficient uptake of hydrophilic biological molecules that would otherwise not be absorbed from food. These can also transport drugs that have some features in common with their natural substrates. There is some very early limited QSAR results predicting compounds that are transported by specific proteins, which has also been used to attempt to design this property into new drugs. Specific transporters that have been examined include monocarboxylic acid transporters (which are also relevant for BBB penetration), glucose transporters, and peptide transporters. There are also specific transporters in the kidneys that influence clearance rates: some ionic drugs are actively transported into urine by transporters that excrete natural compounds, especially uric acid, choline, histamine, and are transported into bile by transporters that excrete charged bile acids and steroids.

However, in general, modeling of these systems is less advanced than modeling P-gp, and it does not play a significant role in predicting ADME properties yet. This is a very active area of research, and the importance of transporters in absorption, distribution, and excretion means that substantial effort and hence, it is hoped, substantial progress will be made in the next few years.

5.3.9 Metabolism and Intrinsic Clearance

Oral bioavailability is the combined effect of how much drug is absorbed from the

intestine (a combination of passive absorption and transporter effects) and how much is subsequently removed both from the intestinal cells and from the hepatic portal blood supply by liver metabolism before it reaches the systemic circulation. Thus, whether a drug is metabolized, and at what rate it is metabolized, is a critical OB-related ADME issue. Thus, for example, grapefruit juice can enhance the OB of some compounds because it contains powerful inhibitors of CYP3A4, a drug-metabolizing enzyme present in gut and liver that would otherwise degrade the absorbed compound. Related to this is predicting what the metabolite of the drug is and hence whether it is likely to be active or toxic.

Metabolic rate is related to "intrinsic clearance" – the inherent rate at which drugs are cleared from the blood by metabolism, mainly in the liver. Intrinsic clearance is usually estimated *in vitro*, using liver tissue slices, cultured primary liver cells, hepatocyte cell lines, or microsomes. The last of these are membrane fractions from liver tissue and are a widely used reagent: they contain a full complement of P450 enzymes from the species in question and can be quality controlled by reference to known metabolites.

None of these methods are ideal: primary tissue is difficult to obtain and store, microsomes have to be prepared from tissue (although they can be stored frozen), primary hepatocytes vary in their level of differentiation, and hepatocyte cell lines do not usually reflect more than a fraction of the metabolic diversity of the liver. However, these tests are much faster and cheaper than animal metabolism experiments, not least because they can be done on milligrams rather than grams of test material, so they have become standard practice.

The principle clearance is by a family of P450 heme-containing monooxygenases termed *CYPs*. The most important is CYP3A4, which metabolizes over half of all drugs to some extent and is also present in the intestinal mucosa.

In principle, as CYPs are a well-characterized family of enzymes with known modes of action, all the technology of conventional drug design should be able to be brought to bear on them. Lack of crystal structures for the human enzymes has hindered this (the structure of human CYP3A4 was solved in 2002, but is still commercially confidential). However, homology modeling with CYPs from other species has lead to a major modeling effort by Segall et al. using quantum chemical calculations to probe the interaction of the heme with the substrate and molecular mechanical and docking approaches to find which substrates will fit the enzyme. This has shed light on the mechanisms of CYP3A4 interactions with its substrates and in some cases can predict where on the substrates the enzyme will attack (and hence what the product will be). However, for many others, details of the steric interactions between the enzyme and substrate foil modeling attempts, so that (for example) prediction of terfenadine metabolites is completely incorrect. Thus, while these approaches are of value, they still have to be used in conjunction with structure-matching and other heuristic approaches.

Until improved structural chemistry information allows accurate computational chemical calculations of interactions of substrates with CYPs, we must fall back on QSAR approaches. Inspection of the substrates of CYP3A4 and P-gp show that there is substantial overlap, and this has lead to speculation that they are structurally related. It is more likely that this similarity

is because they have evolved specificity to dispose of similar types of molecules (i.e. broadly hydrophobic molecules of the size of plant and bacterial secondary metabolites) and not because of any genuine similarities. Incorporating both separately into pharmacokinetic models is more successful than trying to develop a unified model for these two enzymes.

CoMFA QSAR of hepatocyte and microsome metabolism produces an $R^2 = 0.59$ for quantitative prediction of the rate of metabolism, a challenging modeling target. The CoMFA 3D pharmacophore, depending on training set, combines hydrophobic domain(s), a hydrogen bond acceptor, a ring aromatic feature, and a positively ionizable feature. Use of the Cerius2 GA also produces a descriptor-based model, with particular emphasis on the chlorine atoms in molecules as well as on descriptors consistent with those above. These are features of which one or more have to be present for rapid hepatic clearance. CYP inhibition (formally, pK_i) has been predicted with some success with a combined GA/ANN approach and using GP alone with a fragment-based descriptor set (White et al., Amedis Pharmaceuticals, unpublished). Objectively, these models are as good as rapid *in vitro* screens for CYP metabolism or inhibition at predicting human clearance; however, they are not likely to be widely accepted until they are substantially better.

The other approach to CYP prediction is one widely implemented in toxicology but only minimally in ADME. This is the use of databases of prior examples coupled with some suitable similarity score to find the nearest examples to a test compound. These can be a simple database and search engine or a more sophisticated expert system. These do not have any greater success than the mathematical modeling methods mentioned above and are inherently limited by the need to have humans (with their limited bandwidth and inherent biases) select as well as input the data. However, their substantial advantage for the user is that they can provide explanations for their answers, tracking back "hits" to original literature papers: this provides the user with an understanding of the scientific context of the result as well as a number. It is questionable how real the value of this is, as few will read the original papers in depth to uncover any methodological flaws, unseen biases, or other problems that would render the comparison with their present problem irrelevant. However, subjectively, users feel better about using the system and so it is used. This is a lesson other ADME modelers have been slow to take on board. Systems of this sort include MetabolExpert and Meteor.

5.3.10 Conclusion of P-gp, Transporters, and CYPs

In conclusion, while, in principle, predicting how drugs will interact with single protein targets such as CYP and P-gp is more solvable by a mechanistic, computational chemical approach, this is not yet practical, and the data-driven, mathematical modeling approaches used for higher-level ADME properties are still at least as useful. This is partly because transporters are integral membrane proteins, and hence determining their 3D structure is hard. However, equally problematic is the fact that their binding sites are highly nonspecific, being, in effect, hydrophobic "buckets" with a catalytic site in them, and docking molecules into such ill-defined structures is extremely prone to error. Structural information will improve, but the docking problem is likely to remain.

6 Distribution

Once a drug has successfully negotiated the gastrointestinal tract and the liver, it enters the systemic circulation and from there can go to all the organs of the body. Predicting how this occurs is important to predicting the concentration of the drug at any time in its target tissue or receptor. This breaks down broadly into a system-wide measure of where the drug goes (volume of distribution, V_d) and consideration of how it penetrates specific organs or tissues.

6.1 Volume of Distribution and Protein Binding

Formally, V_d is the dose of the drug divided by its concentration in plasma, that is, the apparent volume of plasma in which the drug is dissolved. It represents a combination of the literal volume of the body that the drug dissolves in and the extent to which it binds to body components. Paradoxically, V_d can be larger than the volume of the patient who received the drug: if the drug binds tightly to (for example) bone or fat, these tissues will extract the drug from plasma, leading to an unexpectedly low concentration and hence a high V_d.

A range of approaches to V_d prediction, including advanced modeling techniques such as ANN and GP, have been quite unsuccessful. This is probably because, although V_d is simple to measure, it is the result of an enormously complex phenomenon of traffic of compound into and out of all the compartments of the body, both through diffusion and partition processes and through active transport. Only complex integrative PK models (see below) that explicitly model many compartments and tissues in the body and include measured values for drug transport and partition come close to modeling V_d successfully.

A major component of such models is plasma protein binding (PPB). Pharmacologists wish to determine PPB in its own right, especially for drugs that bind to plasma proteins very tightly and so are essentially not present in plasma as free drug at all. The extent of PPB may also influence intestinal absorption, an effect modeled in CaCo-2 cell systems. So this is a useful parameter to know and hence to model.

Serum albumin (HSA) is the dominant protein by mass in plasma, being present at 0.6 mM concentration, which is usually far higher than the concentration of any drug. Binding is therefore unusual in that it does not saturate, and very low pK_a can result in significant drug binding. Neutral drugs bind to HSA through a predominantly nonspecific hydrophobic interaction, and this is reflected in modeling results: Colmenartejo et al. use a GA to select among a large number of candidate descriptors and find that the predictions ($R^2 = 0.82$, $Q^2 = 0.73$ for predicted affinity of binding to HSA) are dominated by LogP. Binding has been modeled with quantum mechanics–calculated descriptors – H-bond donors, HOMO energy, PSA, dipole moment, ClogP, and a connectivity measure indicating six-membered rings. Models select these parameters in a way that is related to general shape and overall hydrophobicity. Outliers tend to be acidic compounds that bind to different binding sites in HSA.

Basic drugs bind primarily to α 1-acid glycoprotein; there has been little modeling of binding to this protein.

PPB would be expected to influence V_d substantially, as a major "sink" for drug molecules. This obvious relationship can

be used to derive a formula for V_d on the basis of PPB, protein concentrations in plasma, and similar volumes and concentrations for other compartments, such as specific tissues and the fluid between cells. Lave et al., Lombardo et al., and Poulin and Theil discuss general equations of this sort, relating PPB to V_d, and they are modestly successful: Poulin and Theil obtain $R^2 = 0.61$ for all drugs. However, despite the logic of the approach, in man at least V_d and PPB are almost completely uncorrelated ($R^2 = 0.04$ – see Fig. 6). The equations may therefore work only because PPB is a surrogate for lipophilicity, as many models include an explicit LogP term, and LogP is independently useful as an indicator of V_d.

As PPB and V_d are apparently largely unrelated for most drugs (despite theory), some authors have gone on to examine those limited PPB modes that do substantially affect PK. For the pharmacologist, the critical aspect of PPB prediction is not differentiating 25% from 75% (which can be compensated by dosing) but predicting what drugs will be very tight binders (over 90% bound and especially over 99%), which seriously distort the PK. Kratochwil et al. built a model specifically to address this, to distinguish 90% from 99.9% binding: PPB is significantly more correlated with V_d in these high-binding domains. This application illustrates the need to build models to answer development questions rather than purely scientific ones.

6.2 Blood–brain Barrier

Distribution to well-perfused organs is usually rapid, and the concentration is

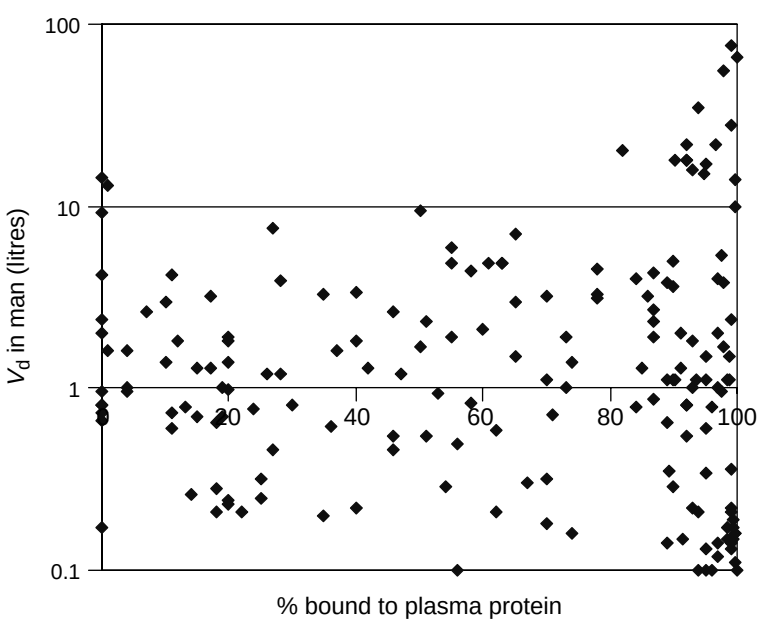

Fig. 6 Plot of fraction of drug bound to plasma protein (*x*-axis) versus volume of distribution (*y*-axis) for 184 compounds for which both parameters are reported in Goodman and Gilman or in Dollery.

simply dependent on blood concentration. However, distribution to some organs and tissues is slower, notably to adipose tissue (fat, where concentrations depend on build up of a reservoir in the fat itself, and hence are strongly modified by lipophilicity) and brain.

Penetration of drugs to the brain is a major development issue. The brain is one of the most biochemically complex organs of the body, but the complement of receptors shares many similarities with receptors outside the central nervous system (CNS). Thus, for example, histamine receptors are almost identical in the brain and the periphery; so the only difference between antihistaminic drugs that sedate (first-generation drugs such as diphenhydramine) and second-generation ones that do not (such as loratidine and terfenadine) is that the former penetrate into the brain and the latter do not. So determining how much of a compound enters the brain is important both for compounds that are meant to have a CNS effect and for those that are not.

Brain penetration is a modeling target because there is a barrier to passive diffusion of compounds out of the blood and into the brain tissue (called the *blood–brain barrier – BBB*), made primarily of the endothelial cells lining the brain capillaries. These are connected by tight junctions and express P-gp that actively pumps lipophilic molecules out of the endothelial cells. Entry into other tissues is not usually the target of *ab initio* modeling in the same way, because they are not protected from the blood by the same type of discrete barrier.

Most drugs enter the brain by passive diffusion – a few are targeted at specific transport mechanisms such as the ferritin transporter, in the expectation of active transport. So properties that enhance diffusion through the biological membrane will enhance BBB penetration (pBBB, almost always expressed as a log scale measure). Consequently, as would be expected, measures of lipophilicity and hydrogen-bonding potential feature substantially in predictions of pBBB. Modeling seeks to model the ratio of a drug to that in whole brain (the usual animal endpoint) or blood to cerebrospinal fluid (CSF), the measurement in man in most cases. Either measurement is practical but slow and expensive. It is also possible to estimate whether a compound enters the brain by examining whether it has any CNS pharmacology, providing that there are receptors in the CNS to which it could bind if it could get there (as with the antihistamine example above).

All attempts to model pBBB are faced with a very limited data set. Partition between blood and brain can only be done in whole animals and requires a substantial number of animals to be sacrificed and analyzed per data point; this is therefore very expensive data. The largest described, compiled by Amedis Pharmaceuticals from literature sources, contains only 167 molecules for which quantitative partition between brain and blood has been measured. Most studies use substantially fewer molecules, and some recycle data sets that have transcription errors or out-of-date and inaccurate values.

In general, pBBB is strongly correlated with linear functions of LogP, LogPcyh (LogP in water/partition coefficient between water and cyclohexane), PSA, and hydrogen-bonding potential. An algorithm based on PSA and molar volume gives $R^2 = 0.71$. LogPcyh is often found to be better than LogPoct when used with other parameters. A good model of pBBB can be derived using LogP, PSA, and H-bond

acceptor count, or other lipoaffinity measures (100 molecules give an $R^2 = 0.73$). Unlike the general finding for OB, several authors have found significantly better correlation of pBBB with a full dynamic PSA calculation than with quick approximations; however, a fast PSA combined with ClogP can provide $R^2 = 0.77$.

PSA alone can be used to predict pBBB: Kelder et al. found that

$$\log(\text{pBBB}) = 1.33 - 0.322\ \text{PSA}.$$

This is unrealistic for all molecules, as it suggests that all molecules with zero PSA have the same pBBB – methane has a pBBB similar to polythene or graphite – but works well within the domain of most druglike molecules, again emphasizing the need to design algorithms to solve practical problems.

Crivori et al. studied the BBB relevance of 72 Volsurf descriptors and found excellent PCA breakdown of compounds into BBB penetrant and partially nonpenetrant compounds without training, which suggests that some of the major chemical features of these molecules (rather than subtle ones) determine their BBB penetration.

With success in physicochemically based models for OB, which assume that the penetration of the gut-epithelial barrier is a diffusion or partition process, the same approach has been used to generate models for pBBB. A model based on the free energy of solvation in water (a calculated thermodynamic measure) achieves an $R^2 = 0.72$. Lombardo et al. used the free energy of solution in water to derive a prediction of pBBB to $R^2 = 0.67$. They considered whether other parameters could improve this and concluded that they could not, which they said was not surprising because the Delta G calculation includes calculation of energy needed to create a cavity in the solvent (reflected in PSA) plus interaction between the molecule and the solvent (reflected in LogP).

On the basis of the understanding that pBBB is primarily a diffusion property, Abraham's group (Platts et al.) derived a version of his general solvation equation to predict pBBB to $R^2 = 0.9$ in a series of very powerful papers. They observe that this model does not predict compounds with carboxylic acid groups well and suggest that these are transported at least in part by the known carboxylic acid transporter.

Two other types of descriptors have been identified as important by other studies. Norinder and Haeberlein have developed the equivalent of the Rule of Five for pBBB: $\text{N} + \text{O} < 5$, $\text{LogP} > (\text{N} + \text{O})$ (N, O = number of nitrogen and oxygen atoms respectively). Applying this gives a C_a of between 0.85 and 0.9. Luco and White et al. (unpublished) have used a range of topological connectivity and atom-type descriptors to build models that had $Q^2 = 0.75$ and $C_a = 0.92$ (for a classification model), respectively. The Luco model again finds the number of nitrogen atoms to be significant, as well as combined measures of hydrogen-bond donor and acceptor counts, differential connectivity, a measure of the information-theoretic complexity of connectivity of the molecule, and the number of different elements in the molecule. The White model is a multidescriptor model built using GP and using dozens of descriptors, but again selected molecular connectivity descriptors and nitrogen atom content as well as PSA as key elements of the model.

Unlike OB, therefore, pBBB modeling results suggest that current approaches and data sets are capable of finding new models to predict brain entry beyond those based on simple diffusion-based or thermodynamic systems and specifically that there is a difference between N

atoms and other polar atoms, and that molecular size, shape, and "floppyness" are important to pBBB prediction.

6.3 Conclusion on Distribution

Unlike OB, it seems likely that there is room for improvement in the descriptor sets and the models used for pBBB prediction. There is also substantial room for improvement in the prediction of other distribution parameters such as V_d, to which advanced modeling methods have been applied only sporadically. Distribution modeling is therefore an area in which substantial advances are likely to be made over the next few years.

7 Integration of Predictive Models

I have stated that true Type-I models of human PK parameters are not feasible, and we have seen that, depending on the endpoint, a variety of Type-V models have been built.

Type-II and Type-III models, in which either modeled or measured subprocesses are integrated mechanistically to provide a model of drug physiology, are widely applied. By contrast, Type-IV models, in which measured inputs are integrated statistically to provide a picture of the whole physiology of the drug, are rare. This is because the top-level interactions and mechanisms of physiology are well understood, and if the parameters of the components of these models can be modeled or measured, there is no need to guess at how macroscopic physiology works.

Integrating low-level measures into high-level systems is well known for OB, where many of the subsystems are well known (even if the specific effects of a particular drug in those systems need to be modeled or measured). For example, Agoram et al. use PLS and ANN models based on fragments and 3D structure–based descriptors to predict MDCK cell permeation, LogP, solubility, and molecular diffusivity. These are then integrated with measured PK parameters to provide the constants in the differential equations for an advanced compartmental absorption and transit (ACAT) model of OB, to determine the rate, extent, and approximate gastrointestinal location of drug liberation (for controlled release), dissolution, passive and carrier-mediated absorption, and saturable metabolism and efflux. The model has 18 GI compartments – stomach, seven small intestine compartments, colon, and nine enterocytes, plus liver, systemic circulation, kidneys, and other compartments, with coupled differential equations for concentration in each compartment. The software operates by numerical integration of the large set of differential equations and provides good results for many drugs (for example, it correctly predicts the dose-dependent bioavailability of propanalol, the effect of grapefruit juice on overall bioavailability of CYP3A4-metabolized drugs, and other "problem" PK outcomes). It is available commercially as GastroPlus.

Leahy et al. at Cyptotex have built integrated models specifically to exploit the data generated by *in vitro* screening assays such as CaCo-2 and the CYP assay. The models include 114 parameters and 14 tissues (lung, muscle, adipose, skin, kidney, brain, gut (9 segments), stomach, spleen, pancreas, liver, heart, kidney, reproductive), with parameters for blood flow, fluid volume, and other physiological

properties gained from the literature. The models can be tailored to different species using the physical parameters of those species and predict half-life and AUC well even for problematical drugs.

Both these examples, and others, require experimental data to tie the models to the real world. Can a purely *in silico* mechanistic model be developed (i.e. a true Type-II model)? At the moment, the answer is definitely "no" for nearly all physiological processes. The nearest approach is in cardiac electrophysiology, where the coupled differential equation models effectively demonstrate the highly nonlinear behavior of the heart to some drugs on the basis of relatively sparse experimental data and substantial literature understanding of the heart. The work of Noble et al. has been developed to the point where it can be used as an adjunct to late-phase clinical trials. Even here, though, models of the underlying low-level processes are not sufficiently accurate to act as inputs to the higher-level model. As an example, HERG IC_{50} can be predicted in a statistical QSAR model to within a log unit (Bains et al.). However, its value needs to be predicted to within a factor of 2 to predict whether HERG blockade will affect the cardiac rhythm. Here, as elsewhere, modeling is limited in part by the technology available but principally by the quality of the data available.

8 Overall Conclusion and a Caution

The statistical modeling tools discussed here have provided some valuable estimators of ADME parameters and will continue to do so. With the possible exception of pBBB, at the moment, all models are almost entirely data-limited. There have been no new "ADME-specific descriptors" since the introduction of the use of PSA and no obvious ones that can be used. Research in this area will therefore work on new theoretically driven approaches (which will come up with completely new ways of describing molecules or biological systems, but will have to wait for the new theories before they can even be conceived), brute-force approaches such as the GP or ANN methods (which can discover new combinations of descriptors on their own, but take enormous amounts of computer power to do so), or better data sets. Ideally, a combination of the second and third options will inform the first, leading to ADME predictions that are sufficiently soundly based in scientific understanding as well as theoretical rigor to be used widely as the basis of decisions.

In the meanwhile, practitioners are reminded to be cautious about the use of the very powerful statistical techniques discussed here. They can find correlations between almost anything. It is essential that such statistical links are tested thoroughly and explored theoretically before any reliance is placed on them.

Acknowledgments

My thanks to Antranig Basman (Amedis) and David Livingstone (ChemQuest) for their comments, and to Robert Scoffin (CambridgeSoft) for many useful discussions and much coding.

Note

Discussion of algorithms, databases, and programs in this article does not imply endorsement of them by the author or VCH.

See also Medicinal Chemistry.

Bibliography

Books and Reviews

Caldwell, J., Gardner, I., Swales, N. (1995) An introduction to drug disposition: the basic principles of absorption, distribution, metabolism, and excretion, *Toxicol. Pathol.* **23**, 102–114.

Clark, D.E., Grootenhuis, P.D.J. (2002) Progress in computational methods for the prediction of ADMET properties, *Current Opin. Drug Discovery Dev.* **5**, 382–390.

Goldstein, G.W., Betz, A.L. (1986) The Blood-brain barrier, *Sci. Am.* **255**, 74–83.

Patrick, G.L. (2001) *An Introduction to Medicinal Chemistry*, 2nd edition, Oxford University Press, New York, Esp, pp. 220–257.

Grass, G.M., Sinko, P.J. (2002) Physiologically based pharmacokinetic simulation modelling, *Adv. Drug Delivery Rev.* **54**, 433–451.

Li, A.P. (2001) Screening for human ADME/Tox drug properties in drug discovery, *Drug Discovery Today* **6**, 357–366.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Delivery Rev.* **23**, 3–25.

Mitchell, T., Showell, G.A. (2001) Design strategies for building druglike chemical libraries, *Curr. Opin. Drug Discovery Dev.* **4**, 314–318.

Noble, D. (2002) Modelling the heart: insights, failures and progress, *BioEssays* **24**, 1155–1163.

Primary Literature

Abraham, M.H. (1993) Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes, *Chem. Soc. Rev.* **22**, 73–83.

Abraham, M.E., Le, J. (1999) The correlation and prediction of the solubility of compounds in water using an amended solvation energy relationship, *J. Pharm. Sci.* **88**, 384–387.

Abraham, M.H., Zhao, Y.H., Le, J., Hersey, A., Luscombe, C.N., Reynolds, D.P., Beck, G., Sherborne, B., Cooper, I. (2002) On the mechanism of human intestinal absorption, *Eur. J. Med. Chem.* **37**, 595–605.

Agoram, B., Woltosz, W.S., Bolger, M.B. (2001) Predicting the impact of physiological and biochemical processes on oral drug bioavailability. *Adv. Drug. Delivery Rev.* **50**, S41–S67.

Ajay, X., Walters, W.P., Murcko, M.A. (1998) Can we learn to distinguish between 'druglike' and 'nondrug-like' molecules? *J. Med. Chem.* **41**, 3314–3324.

Andrews, C.W., Bennett, L., Xu, L.X. (2000) Predicting human oral bioavailability of a compound: development of a novel; quantitative structure-bioavailability relationship, *Pharm. Res.* **17**, 639–644.

Artursson, P., Palm, K., Luthman, K. (1996) Caco-2 monolayers in experimental and theoretical predictions of drug transport, *Adv. Drug Delivery Rev.* **22**, 67–84.

Bains, W., Basman, A., White, C. (2003) HERG binding specificity and binding site structure: evidence from a fragment-based evolutionary computing SAR study, *Prog. Biophys. Mol. Biol.* in press.

Bains, W., Gilbert, R., Sviridenko, L., Gascon, J.-M., Scoffin, R., Birchall, K., Heatherington, L., Harvey, I., Caldwell, J. (2002) Evolutionary computing methods to predict oral bioavailability QSPRs, *Curr. Opin. Drug Discovery Dev.* **5**, 44–51.

Boobis, A., Gundert-Remy, U., Kremers, P., Macheras, P., Pelkonen, O. (2002) In silico prediction of ADME and pharmacokinetics. Report of an expert meeting organised by COST B15, *Eur. J. Pharm. Sci.* **17**, 183–193.

Boyett, M.R., Holden, A.V., Zhang, H. (2001) Engineering virtual cardiac tissue, *Brief Bioinf.* **2**, 233–244.

Caldwell, J. (1981) The current status of attempts to predict species differences in drug metabolism, *Drug Metab. Rev.* **12**, 221–237.

Cashman, J.R. (1996) Drug discovery and drug metabolism, *Drug Discovery Today* **1**, 209–216.

Clark, D.E. (1999) Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena intestinal absorption, *J. Pharm. Sci.* **88**, 807–814.

Colmenarejo, G., Alvarez-Pedraglio, A., Lavandera, J.L. (2001) Cheminformatic models to

predict binding affinities in human serum albumin, *J. Med. Chem.* **44**, 4370–4378.

Crivori, P., Cruciani, G., Carrupt, P.-A., Testa, B. (2000) Predicting blood-brain barrier permeation from three-dimensional molecular structure, *J. Med. Chem.* **43**, 2204–2216.

Cruchiani, G., Pastor, M., Guba, W. (2000) VolSurf: a new tool for pharmacokinetic optimization of lead compounds, *Eur. J. Pharm. Sci.* **11**, S29–S39.

Cruciani, G., Pastor, M., Mannhold, R. (2002) Suitability of molecular descriptors for database mining: a comparative analysis, *J. Med. Chem.* **45**, 2685–2694.

Dunn, W.J., Wold, S. (1990) Pattern Recognition Techniques in Drug Design, in: Hansch, C., Sammes, P.G., Taylor, J.B. (Eds.) *Comprehensive Medicinal Chemistry*, Vol. 4, Pergamon Press, Oxford, pp. 691–714.

Egan, W.J., Merz, K.M., Baldwin, J.J. (2000) Prediction of drug absorption using multivariate statistics, *J. Med. Chem.* **43**, 3867–3877.

Ekins, S., Boulanger, B., Swaan, P.W., Hupcey, M.A.Z. (2002) Towards a new age of virtual ADME/Tox and multidimensional drug discovery, *J. Comput.-Aided Mol. Des.* **16**, 381–401.

Ekins, S., Kim, R.B., Leake, B.F., Dantzig, A.H., Schuetz, E.G., Lan, L.B., Yasuda, K., Shepard, R.L., Winter, M.A., Schuetz, J.D., Wikel, J.H., Wrighton, S.A. (2002) Three-dimensional quantitative structure-activity relationships of inhibitors of P-glycoprotein, *Mol. Pharmacol.* **61**, 964–973.

Ekins, S., Waller, C.L., Swaan, P.W., Cruciani, G., Wrighton, S.A., Wikel, J.H. (2000) Progress in predicting human ADME parameters in silico, *J. Pharmacol. Toxicol. Methods* **44**, 251–272.

Lombardo, F., Blake, J.F., Curatolo, W.J. (1996) Computation of brain–blood partitioning of organic solutes via free energy calculations, *J. Med. Chem.* **39**, 4750–4755.

Grass, G.M. (1997) Simulation models to predict oral drug absorption from in vitro data, *Adv. Drug Delivery Rev.* **23**, 199–219.

Haeberlein, M. (2002) Computational approaches to the prediction of the blood-brain distribution, *Adv. Drug Delivery Rev.* **54**, 291–313 (REF 54291313).

Hansch, C., Leo, A. (1979) *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley-Interscience, New York.

Hansch, C., Quinlan, J.E., Lawrence, G.L. (1968) The linear free energy relationship between partition coefficients and the aqueous solubility of organic liquids, *J. Org. Chem.* **33**, 347–350.

Hardmann, J.G., Limbird, L.E., Molinoff, P.B., Ruddon, R.W., Gilman, A.G. (1996) *Goodman and Gilman's the Pharmacological Basis of Therapeutics*, 9th edition, McGraw-Hill, New York, Esp, pp. 3–28.

Holford, N.H., Kimko, H.C., Monteleone, J.P., Peck, C.C. (2000) Simulation of clinical trials, *Annu. Rev. Pharmacol. Toxicol.* **40**, 209–234.

Huuskonen, J. (2000) Estimation of aqueous solubility for a diverse set of organic compounds based on molecular topology, *J. Chem. Inf. Comput. Sci.* **40**, 773–777.

Huuskonen, J., Salo, M., Taskinen, J. (1997) Neural network modelling for estimation of the aqueous solubility of structurally related drugs, *J. Pharm. Sci.* **86**, 450–454.

Huuskonen, J., Salo, M., Taskinen, J. (1998) Aqueous solubility prediction of drugs based on molecular topology and neural network modelling, *J. Chem. Inf. Comput. Sci.* **38**, 450–456.

Kelder, J., Grootenhuis, P.D.J., Bayada, D.M., De-Ibressine, L.P.C., Ploemen, J.P. (1999) Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs, *Pharm. Res.* **16**, 1514–1519.

Jorgensen, W.L., Duffy, E.M. (2001) Prediction of drug solubility from structure, *Adv. Drug Delivery Rev.* **54**, 355–366.

Kennedy, T. (1997) Managing the drug discovery/development interface, *Drug Discovery Today* **2**, 436–444.

Kier, L.B., Hall, L.H. (1999) *Molecular Structure Description – The Electrotopological State*, Academic Press, San Diego.

Kim, K.H., Greco, G., Novellino, E. (1998) A Critical Review of Recent CoMFA Applications, in: Kubinyi, H., Folkers, G., Martin, Y.C. (Eds.) *3D QSAR in Drug Design*, Vol. 3, Kluwer, UK, pp. 257–278.

King, R.D., Hirst, J.D., Sternberg, M.J.E. (1993) New approaches to QSAR: neural networks and machine learning, *Perspectives Drug Discovery Des.* **1**, 279–290.

Klopman, G. (1984) MultiCASE, a hierarchical computer automated structure evaluation program, *Quant. Struct.-Act Relat.* **11**, 176–184.

Klopman, G., Zhu, H. (2001) Estimation of the aqueous solubility of organic molecules by the group contribution approach, *J. Chem. Inf. Comput. Sci.* **41**, 439–445.

Koza, J.R. (1999) *Genetic Programming III: Darwinian Invention and Problem Solving*, Kaufman, San Francisco.

Kramer, S., Frank, E., Helma, C. (2002) Fragment generation and support vector machines for inducing SARs, SAR QSAR, *Environ. Res.* **13**, 509–523.

Kratochwil, N.A., Huber, W., Muller, F., Kansy, M., Gerber, P.R. (2002) Predicting plasma protein binding of drugs: a new approach, *Biochem. Pharmacol.* **64**, 1355–1374.

Langowski, J.J., Long, A. (2002) Computer systems for the prediction of xenobiotic metabolism, *Adv. Drug Delivery Rev.* **54**, 407–415.

Lave, T., Coassolo, P., Reigner, B. (1999) Prediction of hepatic metabolic clearance based on interspecies metabolic scaling techniques and in vitro – in vivo correlations, *Clin. Pharmacokinet.* **31**, 211–231.

Liu, R., So, S.-S. (2001) Development of quantitative structure-property relationship models for early ADME evaluation in drug discovery. 1) Aqueous solubility, *J. Chem. Inf. Comput. Sci.* **41**, 1633–1639.

Livingstone, D. (2003) Theoretical property predictions, *Curr. Top. in Med. Chem.* **3**, 1171–1192.

Lombardo, F., Obach, R.S., Shalaeva, M.Y., Gao, F. (2002) Prediction of volume of distribution in humans for neutral and basic drugs using physicochemical measurements and plasma protein binding data, *J. Med. Chem.* **45**, 2867–2876.

Luco, L.M. (1999) Prediction of the brain–blood distribution of a large set of drugs from structurally derived descriptors using partial least-squares (PLS) modelling, *J. Chem. Inf. Comput. Sci.* **39**, 396.

Mitchell, M. (1998) *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, MA, USA.

Noble, D. (2002) Modelling the heart – from genes to cells to the whole organ, *Science* **295**, 1678–1682.

Obach, R.S., Baxter, J.G., Liston, T.E., Silber, B.M., Jones, B.C., MacIntyre, F., Rance, D.J., Wastall, P. (1997) The prediction of human pharmacokinetic parameters from preclinical and in vitro metabolism data, *J. Pharmacol. Exp. Ther.* **283**, 46–58.

Oprea, T.I. (2000) Property distribution of drug-related databases, *J. Comput.-Aided Mol. Des.* **14**, 251–264.

Palm, K., Luthman, K., Ungell, A.L., Strandlund, G., Beigi, F., Lundahl, P., Artursson, P. (1998) Evaluation of dynamic polar molecular surface area as predictor of drug absorption: comparison with other computational and experimental predictors, *J. Med. Chem.* **41**, 5382–5392.

Platts, J.A., Abraham, M.H., Zhao, Y.H., Hersey, A., Ijaz, L., Butina, D. (2001) Correlation and prediction of a large blood-brain distribution set – an LFER study, *Eur. J. Med. Chem.* **36**, 719–730.

Podlogar, L., Muegge, I., Brice, L.J. (2001) Computational methods to estimate drug development parameters, *Curr. Opin. Drug Discovery Dev.* **4**, 102–109.

Poulin, P., Theil, F.P. (2002) Prediction of pharmacokinetics prior to in vivo studies. I. Mechanism-based prediction of volume of distribution, *J. Pharm. Sci.* **91**, 129–156.

Ran, Y., Jain, N., Yalkowsky, S.H. (2001) Prediction of aqueous solubility of organic compounds by the general solubility equation (GSA), *J. Chem. Inf. Comput. Sci.* **41**, 1208–1217.

Rooney, K.F., Snoeck, E., Watson, P.H. (2001) Modelling and simulation in clinical drug development, *Drug Discovery Today* **6**, 802–806.

Seelig, A. (1998) A general pattern for substrate recognition by P-glycoprotein, *Eur. J. Biochem.* **197**, 252–261.

Segall, M.D., Payne, M.C., Ellis, S.W., Tucker, G.T., Boyes, R.N. (1997) Ab initio molecular modelling in the study of drug metabolism, *Eur. J. Drug Metab. Pharmacokinet.* **22**, 283–289.

Segall, M.D., Payne, M.C., Ellis, S.W., Tucker, G.T., Boyes, R.N. (1998) An ab initio approach to the understanding of cytochrome P450-ligand interactions, *Xenobiotica* **28**, 15–20.

Stenberg, P., Norinder, U., Luthman, K., Artursson, P. (2001) Experimental and computational screening models for the prediction of intestinal drug absorption, *J. Med. Chem.* **44**, 1927–1937.

Tarbit, M.H., Berman, J. (1998) High-throughput approaches for evaluating absorption, distribution, metabolism, and excretion properties of lead compounds, *Curr. Opin. Chem. Biol.* **2**, 411–416.

Timony, G., Norris, D., Leesman, G., Retajczyk, R., Tran, S., Chen, Y., Lee, Y., Johnson, N., Castelo, J., Lee, K., Christopher, R., Sinko, P., Grass, G. (2000) Development and validation of the physiologically based iDEA predictive model, AAPS 2 Available from: http://www.aapspharmaceutica.com/scientificjournals/pharmsci/am_abstracts/2000/1880.htm.

Vapnik, V. (1995) *The Nature of Statistical Learning Theory*, Springer, Berlin.

Wagener, M., van Geerestein, V. (2000) Potential drugs and non-drugs: prediction and identification of important structural – features, *J. Chem. Inf. Comput. Sci.* **40**, 280–292.

Walters, W.P., Murcko, M.A. (1999) Recognising molecules with druglike properties, *Curr. Opin. Chem. Biol.* **3**, 384–387.

Walters, W.P., Murcko, M.A. (2002) Prediction of 'drug-likeness', *Adv. Drug Delivery Rev.* **54**, 255–271.

Wessel, M.D., Jurs, P.C., Tolan, J.W., Muskal, S.M. (1998) Prediction of human intestinal absorption of drug compounds from molecular Structure, *J. Chem. Inf. Comput. Sci.* **38**, 726–735.

Willett, P. (1995) Genetic algorithms in molecular recognition and design, *Trends Biotechnol.* **13**, 516–521.

Wold, S., Esbensen, K., Geladi, P. (1987) Principal component analysis, *Chem. Intell. Lab. Syst.* **2**, 37–52.

Yoshida, F., Topliss, J.G. (2000) QSAR model for drug human oral bioavailability, *J. Med. Chem.* **43**, 2575–2585.

Zhao, Y.H., Le, J., Abraham, M.H., Hersey, A., Eddershaw, P.J., Luscombe, C.N., Boutine, D., Beck, G., Sherborne, B., Cooper, I., Platts, J.A. (2001) Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure-activity relationship (QSAR) with the Abraham descriptors, *J. Pharmacol. Sci.* **90**, 749–784 (REF 90749784).

Zmuidinavicius, D., Didziapetris, R., Japertas, P., Avdeef, A., Petrauskas, A. (2002) Classification structure-activity relations (C-SAR) in prediction of human intestinal absorption, *J. Pharm. Sci.* **92**, 360–370.

Drugs and Chemicals Biotransformations: *see* Biotransformations of Drugs and Chemicals

E. Coli Genome

Hirotada Mori[1] *and Takashi Horiuchi*[2]
[1] *Faculty of Pharmaceutical Sciences, Kyushu University, Fukuoka, Japan*
[2] *Research and Education Center for Genetic Information, Nara Institute of Science and Technology, Ikoma, Japan*

1 The *E. coli* Genome Sequence 640

2 ORF Predicted from the Complete Genome Sequence of *E. coli* 640

3 Repetitive Sequences, Sites, RNA Genes, etc. 641

4 Post Sequence Genome Project 643
4.1 Transcriptome Analysis 644
4.2 Proteome and Interactome Analysis 644

5 *E. coli* Genome as a Model for Systems Biology 644

6 Useful *E. coli* Websites 645

Bibliography 646
Books and Reviews 646
Primary Literature 646

Keywords

Escherichia coli
A member of γ protepbacteria Enterobacteriaceae normally found in the human gastrointestinal tract among numerous strains. K-12 has been used for most experiments in molecular biology.

Genome
The total complement of genetic material in a cell or individual.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

Functional Genomics
A field of global functional analysis of genome based on complete sequence data.

Since Wollman & Bronfenbrenner introduced this organism in the 1930s, *Escherichia coli* has become one of the most thoroughly studied organisms in biology. The discovery of conjugation by Lederberg and Tatum in the 1940s using sets of defined mutants triggered a revolution in biology through clarification of the concept of "gene." It is a clear testament to the importance of this organism in modern biology that nobody can escape from using *E. coli* to study genes or genomes. For over 50 years, extensive and sustained efforts of many microbiologists, geneticists, biochemists and bioinformatists around the world contributed to the detailed analysis of the functions of many genes and gene products in these bacteria.

1
The *E. coli* Genome Sequence

The *E. coli* genome sequence was determined through independent efforts of both the United States and Japanese groups by using two different strains of *E. coli* – MG1655 and W3110 respectively. These strains diverged from the same ancestral strain about 50 years ago causing slight differences between them, including the large inversion involving the ribosomal RNA genes (see below). The complete genome sequence analysis revealed precise differences between the two strains. Comparison of the genomes of strains MG1655 and W3110 revealed relatively infrequent base substitution mutations, in the order of 10^{-5} per bases in the protein-coding region. On the other hand, dynamic genome rearrangements such as insertions and deletions are not uncommon. In summary, the *E. coli* genome can tolerate large rearrangements, such as insertion, deletion and recombination rather than microscale changes like base substitution. In 2001, pathogenic *E. coli* O157, which is very closely related to *E. coli* K-12, was subjected to genomic sequencing by two groups (in the US and Japan). The results obtained with this pathogenic strain agreed well with the above findings with strain K-12. The *E.coli* O157 acquired its pathogenicity mainly by horizontal gene transfer mediated by a temperate bacteriophage.

2
ORF Predicted from the Complete Genome Sequence of *E. coli*

Extensive studies of a few model organisms including *E. coli* and *Saccharomyces cerevisiae* elucidated the functions of many genes and gene products particularly by using techniques of genetics, biochemistry and physiology, and so on. Complete genome sequence of *E. coli* revealed the presence of about 4400 predicted ORFs. More than 2000 genes have so far been characterized and classified by their function, whereas the rest remain functionally uncharacterized (Fig. 1). Network analysis

Functional category	Number of ORFs	%
Unknown	2001	45.5
Transport/Binding proteins	377	8.6
Energy metabolism	363	8.3
Other categories	300	6.8
Cell envelope	204	4.6
Translation	163	3.7
Central intermediary metabolism	156	3.6
Amino acid biosynthesis	138	3.1
Biosynthesis of cofactors, prosthetic group, carriers	129	2.9
Cellular processes	127	2.9
Regulatory functions	124	2.8
Purines, pyrimidines, nucleosides, and nucleotides	107	2.4
Replication	92	2.1
Fatty acid/Phospholipid metabolism	60	1.4
Transcription	52	1.2

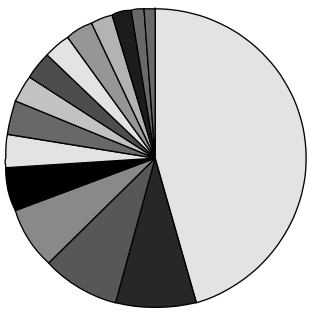

Fig. 1 Functional classification of *E. coli* ORFs.

of mutual functional relationships between genes or gene products has just started.

Accumulation of numerous genomic sequence data accelerated comparative approach, providing the basis for distinction between two ways of genes to be homologous: orthologous genes are homologues between species, and paralogues are those that descend directly from copies of a gene that duplicated within an ancestral genome. Orthologues are therefore good instruments for building phylogenetic trees of organisms, and paralogues are useful for understanding the course of protein evolution. Comparative analysis using the full set of genes in *E. coli* revealed two classes of genes: paralogues and nonparalogues. Sets of paralogous genes could be considered to have descended from ancestral genes by duplication and divergence. The concepts of "protein family" and "module" were discussed by Riley and her colleagues: "module" defined as a long conserved segment, usually larger than 100 amino acid residues, within a protein and an element of sequence that has independent evolutionary history. It may also correspond to a functional unit and contain several domains that are defined as structural units. These functional units are capable of independent existence but in some cases, joined to another module of independent function in a multimodule protein. For example, proteins involved in transport or binding proteins form a large protein family in *E. coli*, and some of them (enzyme, transport protein and transcriptional regulator protein) share modules for binding certain ligands.

3 Repetitive Sequences, Sites, RNA Genes, etc.

Many kinds of repetitive sequences are found within the genome and some of them have important physiological function. Distribution of repeats on the chromosome is not random and seems to be

related to some feature of DNA replication. Repetitive sequences in *E. coli* are encountered in many contexts. Various classes repeats are present in diverse prokaryotes, including *E. coli*. Coding sequences such as ribosomal RNA genes, transfer RNA genes, and insertion sequences are present usually in multiple copies but their copy numbers are relatively low. Other interspersed repetitive DNA sequences are relatively short but abundant and located within intercistronic noncoding sequences. The latter families of repetitive sequences are listed in Table 1.

E. coli has 7 copies of rRNA coding genes (*rrn*) and an additional copy of 5S rRNA coding gene. Other rRNA related DNA sequences named TRIP, showing significant similarity to 5S rRNA, were recently found, which may have important roles related to 5S rRNAs. The rRNA genes are located within half of the chromosome that contains the origin of DNA replication (*oriC*), whereas many of the TRIP sequences are located in the other half of the chromosome. Moreover, *rrn* and some of the related TRIP sequences are distributed symmetrically on the leading strands on both sides of *oriC*. Consistent with this location, transcription of *rrn* operons generally proceeds away from the replication origin (Fig. 2).

Tab. 1 Extragenic highly repetitive sequence families.

Sequence	*Size [bp]*	*Copy number*
BIME	40 ~ 400	~800
Ter	30	63
BoxC	56	36
Iap	29	23
IRU	127	19
RSA	152	6

Repetitive sequences such as rRNA genes provide a driving force for genome rearrangement. A large inversion of the genome between rRNA genes has been documented in *E. coli*, and one such inversion was found in the strain W3110. Although the direction of *rrn* gene transcription is strictly restricted away from the replication origin, the inversion in *E. coli* appears to be stable because the geometric relationship between *rrn* operons and the replication origin are preserved.

Insertion Sequences (IS) represent another family of repetitive elements and cause genetic variation among different strains of *E. coli*: both the abundance and distribution of insertion sequences can vary in different strains. Strain W3110, whose sequence was determined by the Japanese group, contains about 60 ISs

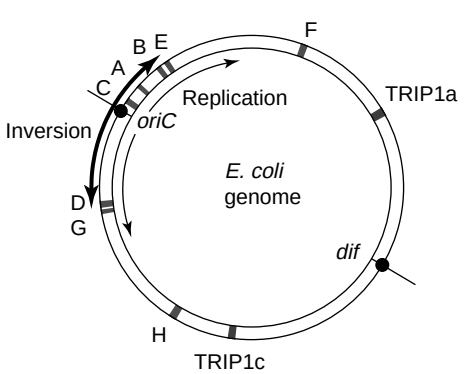

Fig. 2 Ribosomal RNA genes (*rrnA*-H) and related sequences. The arrow represents the inversion loci of W3110 between D (*rrnD*) and E (*rrnE*).

that belong to at least 10 distinct families and at least 10 ISs differ in their location when compared to another strain MG1655, whose sequence was determined by the US group. These results testify to a great variability in both number and family of ISs in closely related strains. Although it is not possible to decide whether these elements were primarily selected on the basis of their beneficial character, or whether they act as genomic parasites, there are a number of examples in which direct selection for certain mutants resulted from transcriptional activation mediated by IS.

Interspersed repetitive sequences represent relatively short (usually less than 500 bp), noncoding, intercistronic and dispersed elements found in bacterial genomes. Six classes of highly repetitive sequences, BIME, IRU, Box C, RSA, iap, and Ter Sequences have been identified (Table 1). Although they are highly repeated, they do not constitute more than 2% of the total *E. coli* DNA and this explains the failure of identification by the classical experimental methods. These sequences were mostly discovered by computer analysis of sequence data.

None of these sequences encode proteins, and are dispersed throughout the chromosome. BIME is the best documented case, originally identified as REP (*R*epetitive *E*xtragenic *P*alindromic), which consists of a 38 b consensus palindromic sequence and can form a stable stem-loop structure with a 5 bp variable loop in the middle. Clusters of REP element with several sequences form a complex mosaic structure called BIME (*B*acterial *I*nterspersed *M*osaic *E*lements). Approximately 800 BIMEs appear to be homogeneously distributed over the entire chromosome. BIMEs are found in extragenic locations at the 3′ ends of operon or between two genes of an operon, but rarely upstream of the first gene of an operon. Specific interactions of DNA gyrase, IHF and DNA polymerase I with BIMEs were reported, although the functional importance is unclear. The BIME-Pol I interaction suggests possible amplification of BIME regions.

4 Post Sequence Genome Project

Upon completion of genome sequencing, efforts were made to construct various resources including a complete set of ORF clones and of deletion mutants. These resources should provide basic tools for systematic functional analysis of the genomes that would eventually develop into systems biology.

Furthermore, genome-wide analysis of transcriptional regulation (transcriptome), protein dynamics (proteome), flow of metabolites (metabolome) have rapidly developed to make the best use of DNA sequence information (''-ome'' is a Greek suffix for ''whole'').

Although *E. coli* is among the most thoroughly studied genetic systems, half of the genes had not been characterized experimentally. One can now identify all of their genetic components, such as protein-coding genes, RNA genes, and regulatory elements. It is, in other words, limited number of genes to be analyzed, although almost half of them remain to be functionally analyzed. Moreover, global analysis of regulatory network of transcription, protein–protein interaction and so on, has just started.

DNA sequence analysis identified about 4400 protein-coding genes in *E. coli*. The power of genetically tractable model organisms is that they can facilitate the analysis of physiological gene function

in vivo. Precise genetic manipulation is particularly important for functional genomics. Genome sequence data permitted to design oligo DNA primers for precise amplification of ORFs and a complete set of histidine tagged ORF clones, fused with and without GFP gene, has been obtained and is available upon request for *E. coli* (reviewed by Mori et al.).

One approach for systematic functional analysis is gene deletion (replacement) by homologous recombination. Targeted gene replacement, once thought to be difficult in *E. coli*, can now be dealt with by using techniques developed by Wanner and colleagues. These systematic approaches to construct resources for functional genomics would rapidly raise *E. coli* to one of the leading organisms in the field of functional genomics.

4.1
Transcriptome Analysis

With the advance of genome research, many novel techniques have been developed including DNA microarray or DNA chip technology that are extremely useful for analysis of global gene expression. Generally, DNA microarray is defined as an orderly arrangement of tens to hundreds of thousands of unique DNA molecules of known sequence usually on a glass slide. Each unique DNA molecule is individually synthesized on a rigid silicon plate (generally called *DNA chip* and developed by Affymetrix Co.) or presynthesized DNA molecules (synthetic oligonucleotides or polymerase chain reaction (PCR) products) are spotted and immobilized on the slide glass. The use of DNA microarray to study *E. coli* gene regulation was first illustrated by Blattner and colleagues, and has rapidly expanded to analyze various aspects of transcriptional regulation.

4.2
Proteome and Interactome Analysis

Technology for global analysis of cellular proteins has been established since 1975 as two-dimensional polyacrylamide gel electrophoresis developed by O'Farrell. And *E. coli* has a long history for protein cataloging in 2D gel. However, new technology called *matrix-assisted laser desorption ionization-mass spectrometry* (MALDI-MS) accelerated protein identification dramatically. Currently, the most widely used method for proteome analysis is 2DE, followed by identification by mass spectrometry in high-throughput. Proteomics has now developed on global scales not only for identification and cataloging of all the proteins expressed in a cell but for global analysis of protein–protein interaction, called interactome, using the yeast two-hybrid system or related methods. As an alternative method for global interaction analysis, His-tagged-pulldown technique widely used as a fusion tag in expression of recombinant proteins has been used to allow affinity purification. A complete set of His-tagged ORF clones has already been established, and the comprehensive protein–protein interaction analysis is now under way. Mass spectrometry is also used for identification of copurified interacting candidate proteins.

5
E. coli Genome as a Model for Systems Biology

Biology itself is now standing at the turning point from the past descriptive biology to the modern quantitative systems biology. Many systematic approaches such as DNA microarray or proteome analysis have been adopted that should help to

understand the cells comprehensively as a system. Among model organisms, *E. coli* is especially important in this field because of enormous accumulation of biological knowledge and experimental methodologies. Scientists can now describe in exquisite detail many of the molecular parts and processes in a cell. Complete genetic information is now available for many organisms in all life kingdoms. Yet, we still do not have a comprehensive computational model of a living cell.

The International *E. coli* Alliance (IECA) was formed to tackle this fundamental biological problem in November 2002. IECA's mission is to consolidate global efforts to understand a living bacterial cell. Scientists around the world are working together to create a complex computer model, integrating all of the dynamic molecular interactions required for the life of a simple, self-replicating cell. An *E. coli* cell model will have immediate practical benefits and will significantly advance the field of computational systems biology. Generation of a computerized *E. coli* cell will add powerful new tools to our existing arsenal of discovery, including virtual experimentation and mathematical simulation. Ultimately, these biological and computational tools could be useful in both drug discovery and in the design of bioenhanced nanomachines. Furthermore, development of a virtual system for experimentation on the *E. coli* cell will be extremely useful for understanding more complex cells and contribute to the development and validation of *in silico* models of human cells. Biology is now changing to become a big science and *E. coli* might become one of the leading organisms in this new biology as was the case in the field of molecular genetics.

6 Useful *E. coli* Websites

As Internet technology advanced, the importance of websites rapidly increased.

Tab. 2 Useful websites of *Escherichia coli*.

http://ecoli.aist-nara.ac.jp	Nara Institute of Science and Technology, Japan
http://www.genome.ad.jp/	Kyoto University, Japan
http://www.shigen.nig.ac.jp/ecoli/pec/index.jsp	The Institute of National Genetics, Japan
http://gib.genes.nig.ac.jp/	The Institute of National Genetics, Japan
http://genome.gen-info.osaka-u.ac.jp/bacteria/o157/	The University of Osaka, Japan
http://www.cifn.unam.mx/Computational_Genomics/regulondb/	Universidad Nacional Autonoma de Mexico, Mexico
http://redpoll.pharmacy.ualberta.ca/CCDB/	University of Alberta, Canada
http://www.uni-giessen.de/~gx1052/ECDC/ecdc.htm	Justus-Liebig-University, Germany
http://genolist.pasteur.fr/Colibri/	Institute Pasteur, France
http://web.bham.ac.uk/bcm4ght6/res.html	The University of Birmingham, UK
http://colibase.bham.ac.uk/	The University of Birmingham, UK
http://biocyc.org/ecocyc/	SRI International, USA
http://www.genome.wisc.edu/	University of Wisconsin – Madison, USA
http://bmb.med.miami.edu/EcoGene/EcoWeb/	University of Miami School of Medicine, USA
http://www.vetsci.psu.edu/ecoli.cfm	Penn State University, USA
http://www.ncbi.nlm.nih.gov/	National Center for Biotechnology Information, USA

The useful websites of *E. coli* are listed in Table 2.

See also Bacterial Pathogenesis, Molecular Basis of; Genetics, Molecular Basis of; Genomic Sequencing (Core Article).

Bibliography

Books and Reviews

Koonin, E.V., Galperin, M.Y. (Eds.) *Sequence – Evolution – Function Computational Approaches in Comparative Genomics*, Kluwer Academic Publishers.

Mahillon, J., Leonard, C., Chandler, M. (1999) *Res. Microbiol.* **150**, 675–687.

Mori, H., Isono, K., Horiuchi, T., Miki, T. (2000) *Res. Microbiol.* **151**, 121–128.

Neidhardt, F.C. (Ed.) *Escherichia coli and Salmonella*, ASM Press.

Tucker, C.L., Gera, J.F., Uetz, P. (2001) *Trends Cell Biol.* **11**, 102–106.

Primary Literature

Allen, T.E., Herrgard, M.J., Liu, M., Qiu, Y., Glasner, J.D., Blattner, F.R., Palsson, B.O. (2003) *J. Bacteriol.* **185**, 6392–6399.

Allen, T.E., Palsson, B.O. (2003) *J. Theor. Biol.* **220**, 1–18.

Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. (1997) *Science* **277**, 1453–1474.

Chen, M., Hofestaedt, R. (2003) *In Silico Biol.* **3**, 0030.

Chiang, D.Y., Brown, P.O., Eisen, M.B. (2001) *Bioinformatics* **17**(Suppl. 1), S49–S55.

Corbin, R.W., Paliy, O., Yang, F., Shabanowitz, J., Platt, M., Lyons, C.E. Jr., Root, K., McAuliffe, J., Jordan, M.I., Kustu, S., et al. (2003) *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9232–9237.

Covert, M.W., Schilling, C.H., Famili, I., Edwards, J.S., Goryanin, II, Selkov, E., Palsson, B.O. (2001) *Trends Biochem. Sci.* **26**, 179–186.

Datsenko, K.A., Wanner, B.L. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6640–6645.

Deane, C.M., Salwinski, L., Xenarios, I., Eisenberg, D. (2002) *Mol. Cell. Proteomics* **1**, 349–356.

Edwards, J.S., Palsson, B.O. (1998) *Biotechnol. Bioeng.* **58**, 162–169.

Edwards, J.S., Palsson, B.O. (2000) *Biotechnol. Prog.* **16**, 927–939.

Edwards, J.S., Palsson, B.O. (2000) *BMC Bioinf.* **1**, 1.

Edwards, J.S., Palsson, B.O. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5528–5533.

Edwards, J.S., Ibarra, R.U., Palsson, B.O. (2001) *Nat. Biotechnol.* **19**, 125–130.

Edwards, J.S., Ramakrishna, R., Palsson, B.O. (2002) *Biotechnol. Bioeng.* **77**, 27–36.

Ehrenberg, M., Elf, J., Aurell, E., Sandberg, R., Tegner, J. (2003) *Genome Res.* **13**, 2377–2380.

Fiehn, O. (2002) *Plant Mol. Biol.* **48**, 155–171.

Figeys, D., Gygi, S.P., Zhang, Y., Watts, J., Gu, M., Aebersold, R. (1998) *Electrophoresis* **19**, 1811–1818.

Galperin, M.Y., Koonin, E.V. (1999) *Genetica* **106**, 159–170.

Gerdes, S.Y., Scholle, M.D., Campbell, J.W., Balazsi, G., Ravasz, E., Daugherty, M.D., Somera, A.L., Kyrpides, N.C., Anderson, I., Gelfand, M.S., et al. (2003) *J. Bacteriol.* **185**, 5673–5684.

Gilson, E., Saurin, W., Perrin, D., Bachellier, S., Hofnung, M. (1991) *Nucleic Acids Res.* **19**, 1375–1383.

Goryanin, I., Hodgman, T.C., Selkov, E. (1999) *Bioinformatics* **15**, 749–758.

Gutierrez-Rios, R.M., Rosenblueth, D.A., Loza, J.A., Huerta, A.M., Glasner, J.D., Blattner, F.R., Collado-Vides, J. (2003) *Genome Res.* **13**, 2435–2443.

Gygi, S.P., Rist, B., Aebersold, R. (2000) *Curr. Opin. Biotechnol.* **11**, 396–401.

Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.G., Ohtsubo, E., Nakayama, K., Murata, T., et al. (2001) *DNA Res.* **8**, 11–22.

Herrgard, M.J., Covert, M.W., Palsson, B.O. (2003) *Genome Res.* **13**, 2423–2434.

Hill, C.W., Harnish, B.W. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7069–7072.

Hutchison, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., Smith, H.O., Venter, J.C. (1999) *Science* **286**, 2165–2169.

Itoh, T., Okayama, T., Hashimoto, H., Takeda, J., Davis, R.W., Mori, H., Gojobori, T. (1999) *FEBS Lett.* **450**, 72–76.
Jiao, Z., Baba, T., Mori, H., Shimizu, K. (2003) *FEMS Microbiol. Lett.* **220**, 295–301.
Kolisnychenko, V., Plunkett, G. III, Herring, C.D., Feher, T., Posfai, J., Blattner, F.R., Posfai, G. (2002) *Genome Res.* **12**, 640–647.
Koonin, E.V., Tatusov, R.L., Rudd, K.E. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11921–11925.
Labedan, B., Riley, M. (1995) *Mol. Biol. Evol.* **12**, 980–987.
Labedan, B., Riley, M. (1995) *J. Bacteriol.* **177**, 1585–1588.
Liang, P., Labedan, B., Riley, M. (2002) *Physiol. Genomics* **9**, 15–26.
Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., Brown, E.L. (1996) *Nat. Biotechnol.* **14**, 1675–1680.
Maharjan, R.P., Ferenci, T. (2003) *Anal. Biochem.* **313**, 145–154.
Masuda, N., Church, G.M. (2003) *Mol. Microbiol.* **48**, 699–712.
Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U. (2002) *Science* **298**, 824–827.
Mrowka, R., Liebermeister, W., Holste, D. (2003) *Nat. Genet.* **33**, 15–16; author reply 16–17.
Natale, D.A., Galperin, M.Y., Tatusov, R.L., Koonin, E.V. (2000) *Genetica* **108**, 9–17.
O'Farrell, P.H. (1975) *J. Biol. Chem.* **250**, 4007–4021.
Oshima, T., Aiba, H., Masuda, Y., Kanaya, S., Sugiura, M., Wanner, B.L., Mori, H., Mizuno, T. (2002) *Mol. Microbiol.* **46**, 281–291.
Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G.D., Maltsev, N. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2896–2901.
Overbeek, R., Larsen, N., Pusch, G.D., D'Souza, M., Selkov, E. Jr., Kyrpides, N., Fonstein, M., Maltsev, N., Selkov, E. (2000) *Nucleic Acids Res.* **28**, 123–125.
Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., et al. (2001) *Nature* **409**, 529–533.
Picataggio, S.K., Templeton, L.J., Smulski, D.R., LaRossa, R.A. (2002) *Methods Enzymol.* **358**, 177–188.
Posfai, G., Kolisnychenko, V., Bereczki, Z., Blattner, F.R. (1999) *Nucleic Acids Res.* **27**, 4409–4415.
Posfai, G., Koob, M.D., Kirkpatrick, H.A., Blattner, F.R. (1997) *J. Bacteriol.* **179**, 4426–4428.
Price, N.D., Reed, J.L., Papin, J.A., Famili, I., Palsson, B.O. (2003) *Biophys. J.* **84**, 794–804.
Ramakrishna, R., Edwards, J.S., McCulloch, A., Palsson, B.O. (2001) *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **280**, R695–R704.
Reed, J.L., Palsson, B.O. (2003) *J. Bacteriol.* **185**, 2692–2699.
Reed, J.L., Vo, T.D., Schilling, C.H., Palsson, B.O. (2003) *Genome Biol.* **4**, R54.
Rhodius, V.A., LaRossa, R.A. (2003) *Curr. Opin. Microbiol.* **6**, 114–119.
Rhodius, V., Van Dyk, T.K., Gross, C., LaRossa, R.A. (2002) *Annu. Rev. Microbiol.* **56**, 599–624.
Richmond, C.S., Glasner, J.D., Mau, R., Jin, H., Blattner, F.R. (1999) *Nucleic Acids Res.* **27**, 3821–3835.
Riley, M., Labedan, B. (1997) *J. Mol. Biol.* **268**, 857–868.
Rudd, K.E. (1998) *Microbiol. Mol. Biol. Rev.* **62**, 985–1019.
Sabatti, C., Rohlin, L., Oh, M.K., Liao, J.C. (2002) *Nucleic Acids Res.* **30**, 2886–2893.
Sabina, J., Dover, N., Templeton, L.J., Smulski, D.R., Soll, D., LaRossa, R.A. (2003) *J. Bacteriol.* **185**, 6158–6170.
Schilling, C.H., Edwards, J.S., Letscher, D., Palsson, B.O. (2000) *Biotechnol. Bioeng.* **71**, 286–306.
Selinger, D.W., Cheung, K.J., Mei, R., Johansson, E.M., Richmond, C.S., Blattner, F.R., Lockhart, D.J., Church, G.M. (2000) *Nat. Biotechnol.* **18**, 1262–1268.
Sharples, G.J., Lloyd, R.G. (1990) *Nucleic Acids Res.* **18**, 6503–6508.
Shen-Orr, S.S., Milo, R., Mangan, S., Alon, U. (2002) *Nat. Genet.* **31**, 64–68.
Soga, T., Ohashi, Y., Ueno, Y., Naraoka, H., Tomita, M., Nishioka, T. (2003) *J. Proteome Res.* **2**, 488–494.
Soga, T., Ueno, Y., Naraoka, H., Matsuda, K., Tomita, M., Nishioka, T. (2002) *Anal. Chem.* **74**, 6224–6229.
Soga, T., Ueno, Y., Naraoka, H., Ohashi, Y., Tomita, M., Nishioka, T. (2002) *Anal. Chem.* **74**, 2233–2239.
Storz, G. (2002) *Science* **296**, 1260–1263.
Tao, H., Bausch, C., Richmond, C., Blattner, F.R., Conway, T. (1999) *J. Bacteriol.* **181**, 6425–6440.
Tatusov, R.L., Galperin, M.Y., Natale, D.A., Koonin, E.V. (2000) *Nucleic Acids Res.* **28**, 33–36.

Tatusov, R.L., Koonin, E.V., Lipman, D.J. (1997) *Science* **278**, 631–637.
Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., Koonin, E.V. (2001) *Nucleic Acids Res.* **29**, 22–28.
Thieffry, D., Salgado, H., Huerta, A.M., Collado-Vides, J. (1998) *Bioinformatics* **14**, 391–400.
Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T.S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C., Hutchison, C.A. III (1999) *Bioinformatics* **15**, 72–84.
Tomita, M. (2001) *Trends Biotechnol.* **19**, 205–210.
Tweeddale, H., Notley-McRobb, L., Ferenci, T. (1998) *J. Bacteriol.* **180**, 5109–5116.
Uchiyama, I. (2003) *Nucleic Acids Res.* **31**, 58–62.
Vanbogelen, R.A. (1996).
Vazquez, A., Flammini, A., Maritan, A., Vespignani, A. (2003) *Nat. Biotechnol.* **21**, 697–700.
Wei, Y., Lee, J.M., Richmond, C., Blattner, F.R., Rafalski, J.A., LaRossa, R.A. (2001) *J. Bacteriol.* **183**, 545–556.
Yang, C., Hua, Q., Baba, T., Mori, H., Shimizu, K. (2003) *Biotechnol. Bioeng.* **84**, 129–144.
Yu, B.J., Sung, B.H., Koob, M.D., Lee, C.H., Lee, J.H., Lee, W.S., Kim, M.S., Kim, S.C. (2002) *Nat. Biotechnol.* **20**, 1018–1023.
Zheng, M., Wang, X., Templeton, L.J., Smulski, D.R., LaRossa, R.A., Storz, G. (2001) *J. Bacteriol.* **183**, 4562–4570.
Zipkas, D., Riley, M. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1354–1358.

E-Cell: Computer Simulation of the Cell

Pawan Dhar[1], *Kouichi Takahashi*[2], *Yoichi Nakayama*[2], *and Masaru Tomita*[2]
[1] *Bioinformatics Institute, Singapore*
[2] *Keio University, Tsuruoka, Japan*

1 Introduction 651

2 Biological Modeling and Simulation Tools 653
2.1 DBSolve 653
2.2 Gepasi 656
2.3 Jarnac 656
2.4 Virtual Cell 656
2.5 A-Cell 656
2.6 BioQUEST 656
2.7 Dynafit 656
2.8 ModelMaker 656
2.9 Metamodel 657
2.10 DMSS 657
2.11 E-Cell 657

3 The E-Cell System 657
3.1 Introduction 657
3.2 Architecture of E-Cell 658
3.2.1 Elements of the Control Panel 659
3.2.2 Elements of the E-Cell Model 659
3.3 Features of E-Cell 2.0 660
3.4 Features of E-Cell 3.0 660
3.5 Advantages of the E-Cell System 660
3.6 Limitations of the E-Cell System 661
3.7 E-Cell with 127 Genes 661
3.8 Applications of the E-Cell System 662
3.9 Simulation of Red Cell Enzyme Deficiencies 663

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30545-9

4 Concluding Remarks 664

Bibliography 665
Books and Reviews 665
Primary Literature 666

Keywords

E-Cell
Short form for Electronic Cell. A tool for modeling and simulating cellular pathways.

Self-supporting Cell
A virtual cell with 127 genes sufficient for survival.

In Silico Modeling
Describes a model building process using computers.

Mycoplasma Genitalium
A microorganism with the smallest number of genes.

Virtual Erythrocyte
An *in silico* model of whole erythrocytic metabolism.

Modeling
A practice of abstracting and interconnecting elements leading to a hypothesis. Models can be qualitative or quantitative or an optimum mix of the two.

Simulation
A process of visualizing a model dynamically.

Cells perform complex tasks with sheer simplicity. They are remarkable, tiny factories where thousands of jobs are performed in parallel and with statistical precision. The grand challenge for the scientific community is to understand the entire structural and functional design of the cell and fine-tune it for scientific applications ranging from basic to pharmaceutical. In a traditional experimental set-up, it is difficult to focus on more than one problem though recent high-throughput techniques have enabled investigation at the ''whole'' organism level. Nevertheless, data coming from these experiments are often noisy and require a large number of replicates to validate a single observation. Furthermore, the statistical treatment of the high-throughput data is also controversial. To overcome the physical and conceptual limitations, we need innovative sets of tools and strategies to address a problem. Recently, a spurt

of articles published in *Nature* (Nov 14, 2002 issue) lends credibility to the idea of using *in silico* approaches for understanding and engineering whole cellular systems. E-Cell is one such tool that has been specifically designed to model and simulate cellular pathways. It has been successfully used to create a self-sustaining cell with 127 genes, that is, just stable enough for survival. In this chapter, some of the basic modeling concepts, their importance, the role of E-Cell, and the future challenges that await the modeling community will be discussed.

1 Introduction

A model is an optimal mix of hypotheses, evidences, and abstractions. Models are the closest replicas of phenomena with sound diagnostic and predictive abilities. They are easily understood, controllable, and analyzable even for large and complex data. The aim of whole-cell modeling is to provide both a conceptual basis and a working methodology for the entire cell by adopting a formal modeling process in the presence/absence of mathematical equations. The purpose of sketching this conceptual scaffold is to identify unknown entities, make predictions, and design experiments targeted at unanswered questions.

A cell by itself is a complete genetic and biochemical reactor holding all the information necessary to sustain life. It offers an ideal middle path between (extreme ends of) atomic interactions and whole organs. By creating a whole-cell model, it is at least theoretically possible to stretch out in either direction. Experimental biology has now reached a stage where future research will heavily depend on the *in silico* contribution. Though static representation of data has traditionally helped develop an overall perspective, it is the dynamic modeling that aids in better understanding of the cellular machinery (Table 1).

Broadly, whole-cell transactions may be classified into enzymatic and nonenzymatic processes. Enzymatic processes represent most of the metabolic events while nonenzymatic processes represent gene expression and regulation, signal transduction, and diffusion. In order to create a complete virtual cell, it is important to have provisions for DNA replication and repair, transcription and its regulation, translation, energy metabolism, metabolism, cell division, signaling pathways, cell membrane dynamics (ion channels, pump, nutrients), and intracellular molecular trafficking with appropriate mathematical representations (Table 2).

Tab. 1 Basics of map construction.

Type of representation	*Description*
Linear chain	Unidirectional flow
Branched chain	Two enzymes participate in one reaction, resulting in different products
Loops	Two branches unite forming an inherent dependency between them
Cycles	Larger loops comprising many intermediates

Tab. 2 Mathematical representation of cellular processes

Process	*Dominant phenomenon*	*Typical computational schemes*
Metabolism	Enzymatic reaction	DAE, S-Systems, FBA
Signal transduction	Molecular binding	DAE, stochastic, diffusion reaction
Gene expression	Molecular binding, polymerization, degradation	OOM, S-Systems, DAE, Boolean, stochastic, Bayesian, rule based
DNA replication	Molecular binding, polymerization	OOM, DAE
Cytoskeletal	Polymerization, depolymerization	DAE, particle dynamics
Cytoplasmic streaming	Streaming	Rheology, finite element method
Membrane transport	Osmotic pressure, membrane potential	DAE, electrophysiology

Notes: DAE: Differential algebraic equations; FBA: Flux balance analysis; OOM: Object-Oriented Modeling.

To accomplish the "big picture," the data should not only be of good quality but should also be treated with intuitive mathematical representations that accurately describe life *in vivo*. It is noteworthy to mention that good data is more of an exception than rule! For modeling metabolic pathways, the data input typically consists of rate constants and concentrations. A metabolic pathway usually consists of a mixture of forward and reverse reactions (uni, bi, ter) of ordered/random types. The inhibitors may be intermediate compounds of the same pathway or external entities. The availability of good data makes the modeling process more or less a straightforward process, but often missing links have to be figured out due to incompleteness of information. Problems in doing so arise mostly due to numerical reasons – stiffness and parameter sensitivity. The main difference between data-to-model and model-to-data approaches is that in the former the starting material is substrate, enzyme, and modifier concentrations (for calculating the velocity curve), while in the latter, kinetic constants and velocity of reaction values are assumed. However, the difference between these two approaches sometimes blurs because in real life situations modeling often involves manual data fitting approaches to match an expected output or hypothesis. The major advantage of carrying out simulations is not only to study the system *per se* but also to extrapolate its behavior, even in the presence of a hypothetical condition, for example, a cell with many essential gene knockouts. Modeling, simulation, and analysis are therefore ideally positioned for integrating experimental knowledge into the *in silico* environment. In addition to demystifying nonintuitive phenomena, simulation lets us test experimentally unfeasible scenarios and can potentially reduce experimental costs. Although wet experiments are indispensable for the advancement of biological knowledge, *in silico* modeling can help shorten knowledge discovery. With the enormous computational power easily available, the challenging part in modeling is more conceptual than physical in nature.

The next section gives an overview of modeling tools, online resources, and databases that the modeling community often makes use of.

2 Biological Modeling and Simulation Tools

A number of promising tools are available for studying gene expression, regulation, and metabolic pathways (Tables 3 and 4). Below is a partial description of a few such tools.

2.1 DBSolve

URL: http://websites.ntl.com/~igor.goryanin/

DBsolve is an integrated development environment for metabolic, enzymatic, and receptor-ligand binding simulation. It is an ordinary differential equation (ODE)

Tab. 3 Cellular databases and pathways.

Name URL	*Features*
1. General Online Maps and Pathways	
IUBMB-Nicholson Minimaps http://www.tcd.ie/Biochemistry/IUBMB-Nicholson/	Comprehensive; describes regulatory and spatial features of substrates and enzymes, available in .gif, .svg and .pdf forms
Boehringer Mannheim Biochemical Pathways http://www.expasy.ch/cgi-bin/search-biochem-index	Comprehensive; covers many organisms, most extensively used by researchers, available in online and paper formats
Kyoto Encyclopedia of Genes & Genomes (KEGG) http://www.genome.ad.jp/kegg/	Huge database on gene sequence, regulatory pathways, metabolism, molecular assemblies, etc.
What Is There (WIT) http://wit.mcs.anl.gov/WIT2/	Covers metabolic pathways of over 25 organisms
Enzyme and Metabolic Pathway (EMP) http://emp.mcs.anl.gov/	Includes metabolic pathways, reaction mechanisms, rate laws, and numeric data from research papers
Biopathways Consortium http://www.biopathways.org/	An open forum for developing technologies and standards for biopathways
EcoCyc http://ecocyc.org/	Houses all the *Escherichia coli* pathways with an aim of creating its functional catalog
PathDB http://www.ncgr.org/pathdb/	Plant metabolic database management system; runs on client server architecture
UM-BBD University of Minnesota Biocatalysis/Biodegradation Database http://umbbd.ahc.umn.edu/	Microbial metabolic pathway database. Also used as a teaching tool
METAVISTA http://www.metabolic-explorer.com	Resource for proteomic profiling, metabolic profiling, and metabolic flux analysis
2. Regulatory Pathways	
KEGG Regulatory Pathways http://www.genome.ad.jp/kegg/regulation.html	An extension of KEGG database

(*continued overleaf*)

Tab. 3 (*continued*)

Name URL	***Features***
BioCarta http://www.biocarta.com/	Interactive web-based resource on gene function and proteomics
Biomolecular Interaction Network Database (BIND) http://www.bind.ca/index.phtml	Describes chemical reactions, conformational changes, and protein and network interactions across various species.
Signal Pathway Database (SPAD) http://www.grt.kyushu-u.ac.jp/spad/	Signal transduction database with emphasis on protein–protein and protein–DNA interactions
Cell Signaling Networks Database (CSNDB) http://geo.nihs.go.jp/csndb/	Contains sequences, structures, functions, and reactions involved in cell signaling
Munich Information Centre for Protein Sequences (MIPS) http://mips.gsf.de/proj/yeast/CYGD/db/index.html	Functional yeast genomic database
GeNet – Gene Networks Database http://www.csa.ru/Inst/gorb_dep/inbios/genet/genet.htm	Developmentally regulated gene networks
EmbryoNet http://www.csa.ru/Inst/gorb_dep/inbios/genet/embryo.htm	Developmentally regulated genetic networks
Genetic Network Maps http://www.csa.ru/Inst/gorb_dep/inbios/genet/access.htm	*Drosophila* embryogenesis network
Wnt Signaling Pathway http://www.stanford.edu/~rnusse/wntwindow.html	*Drosophila* developmentally regulated Wnt signaling pathways
3. Transcription Factors and Expression	
TRANSPATH http://transpath.gbf.de/	Describes pathways involved in regulation of transcription factors
TRANSFAC http://transfac.gbf.de/TRANSFAC/index.html	Transcription factor database
RegulonDB http://www.cifn.unam.mx/Computational_Genomics/regulondb/	Transcription regulation and operon organization database
DBTBS http://elmo.ims.u-tokyo.ac.jp/dbtbs/	*Bacillus subtilis* promoter and transcription factor database
Saccharomyces cerevisiae Promoter Database (SCPD) http://cgsigma.cshl.org/jian/	Database on promoters and mapped regulatory regions of yeast
Axeldb http://www.dkfz-heidelberg.de/abt0135/axeldb.htm	Gene expression database of *Xenopus laevis*
NEXTDB http://nematode.lab.nig.ac.jp/	*Caenorhabditis elegans* expression database
MAGEST http://www.genome.ad.jp/magest/about.html	Ascidian expression database

Tab. 3 (*continued*)

Name URL	*Features*
4. Enzyme Database	
BRENDA http://www.brenda.uni-koeln.de/	The most comprehensive database on biochemical reactions
ExPASy http://www.expasy.ch/	Database on protein sequences and structures
NC-IUBMB http://www.chem.qmw.ac.uk/iubmb/enzyme/	Enzyme nomenclature database
Ligand Chemical Database http://www.genome.ad.jp/dbget-bin/www_bfind?ligand	Database of chemical compounds and reactions in biological systems
NIST Thermodynamic database http://wwwbmcd.nist.gov:8080/enzyme/enzyme.html	Repository on thermodynamics of enzyme catalyzed reactions
PROCAT http://www.biochem.ucl.ac.uk/bsm/PROCAT/PROCAT.html	Specialized for 3-D active site templates of enzymes
5. Scientific Literature Search	
PubMed http://www.ncbi.nlm.nih.gov/entrez/query.fcgi	Huge repository on biomedical literature
Medline http://research.bmn.com/medline/search	The oldest and most comprehensive biomedical literature database
Scirus http://www.scirus.com	Extremely useful metasearch tool developed by Elsevier Science Publishers

Tab. 4 Tools for drawing pathways.

Name of the tool URL	*Target application*
Pathfinder http://bibiserv.techfak.uni-bielefeld.de/pathfinder/	Dynamic visualization of metabolic pathways represented as acyclic graphs
Electric Arc http://home.xnet.com/~selkovjr/ElectricArc/	CAD based, can be used to design abstract graphs to electronic circuits
Biopath http://biopath.fmi.uni-passau.de/index.html	Used for digitizing *Boehringer biochemical pathways*
Pathway Browser http://www-pr.informatik.uni-tuebingen.de/~eiglsper/pathways/	Visualization tool, XML-based, requires Java

based tool that incorporates stoichiometry of chemical reactions too. The main strength of DBsolve is the calculation of steady state, fitting, and optimization options.

2.2 Gepasi

URL: http://www.gepasi.org/

Gepasi simulates the steady state and time-course behavior of reactions over time and space based on stoichiometry and reaction kinetics values. The program is based on ordinary differential equations. It is a very useful tool for conducting metabolic control analysis and linear kinetic stability analysis leading to determination of the steady state of a system.

2.3 Jarnac

URL: http://www.cds.caltech.edu/~hsauro/Jarnac.htm

Jarnac is a cell modeling language for describing metabolic, signal transduction, and gene networks. It is linked to Jdesigner that the user interacts with for modeling a biochemical event.

2.4 Virtual Cell

URL: http://www.nrcam.uchc.edu/

Virtual Cell is a modeling tool that associates biochemical and electrophysiological data with microscopic image data describing subcellular locations. It is based on a strong mathematical foundation, and the results can be analyzed as images. Access to the Virtual Cell modeling software is via the Internet using a Java-based interface.

2.5 A-Cell

URL: http://www.fujixerox.co.jp/crc/cng/A-Cell/

A-Cell is a Windows-based graphical user interface (GUI) for the construction of biochemical reaction models. In addition, it has the capability of importing previously constructed models and combining them with the system.

2.6 BioQUEST

URL: http://omega.cc.umb.edu/~bwhite/ek.html

BioQUEST is a set of building blocks that run on the numerical simulation program "Extend" allowing the user to construct conduct time series analysis of biochemical reactions.

2.7 Dynafit

URL: http://www.biokin.com/

Dynafit is a software for simulating biochemical reactions. It also provides *ViraFit* for analysis of hepatitis C viral data, *BatchKi* (client server tool) for determination of tight-binding enzyme inhibition constants, and *PlateKi* (similar to Batchki) but runs as a stand-alone application.

2.8 ModelMaker

URL: http://www.modelkinetix.com/

ModelMaker allows the modeling of continuous and discontinuous functions and stiff and stochastic systems. It also provides optimization, minimization, Monte Carlo, and sensitivity analysis.

2.9 Metamodel

URL: http://bip.cnrs-mrs.fr/bip10/modeling.htm

MetaModel 3.0 is a DOS-based program for simulating simple biochemical reactions.

2.10 DMSS

URL: http://www.bio.cam.ac.uk/~mw263/ftp/doc/ISMB99.ps

Discrete Metabolic Simulation System (DMSS) does not employ kinetic parameters, stoichiometry matrices, or flux coefficients. Instead, the rate of a reaction is modeled on the basis of competing metabolite concentrations or metabolite affinities to enzymes including metabolite and enzyme concentrations.

2.11 E-Cell

URL: http://www.e-cell.org

E-Cell is a modeling and simulation environment. The basic concepts and applications of the E-Cell system are detailed in sections below.

3 The E-Cell System

3.1 Introduction

The raw material for biological complexity is an immense diversity in components and rules that are employed to create and sustain life. In order to understand the underlying complexity and engineer new systems, it is necessary to create an environment that can translate biology at the level of mathematics. In the early 1990s, the concept of *in silico* biology had just started appearing on the horizon but the scientific community was waiting for proof of the concept. Since none existed at that time, we decided to create one. After many trials and errors, a virtual baby called E-Cell was born.

E-Cell is a short form for Electronic Cell. It is a generic object–oriented environment for modeling and simulating molecular processes of the whole cell in user-definable models, equipped with graphical interfaces that allow observation and interaction. The E-Cell modeling approach links diverse cellular processes such as gene expression, signaling, and metabolism, to form a virtual cell framework. Using E-Cell it is possible to create a model and also translate this model into a simulation environment through mathematical equations. More precisely, however, it is a generic system for constructing object models of the cells that can (optionally) emulate the behavior of numeric equation solvers.

In October 1996, the E-Cell project was started at Keio University, Japan. The first working version of E-Cell was ready within three months and the first virtual cell (*Mycoplasma genitalium*) was developed within a year. On March 2001, the beta version 1.0 of the software was publicly released under open source, and new GUI and peripheral software tools were added. Currently, E-Cell is celebrating its third birthday. Version 1.0 and 3.0 are developed for the Linux environment and version 2.0 for Windows. C++ is used for the components in which runtime efficiency, coding strictness, and flexibility are critical. Python is an ideal language for the other user-side

Tab. 5 E-Cell versions.

E-Cell 1.0 (Linux)	*E-Cell 2.0 (Windows)*	*E-Cell 3.0 (Linux)*
Entire E-Cell application including GUI written in C++	Uses a cocktail of C++, Perl, Cygwin, Java	Core portions (libecs, libemc) and simulation objects (reactors, substances, steppers) written in C++
Peripheral programs (er2eri, ss2er, rd2ch, etc.) written in Perl, Python, and yacc/lex		Most other components, including front end, written in Python

Tab. 6 Public E-Cell mailing lists.

Mailing lists	*Features*	*Address*
E-Cell announce	Very low traffic moderated ML for announcements regarding E-Cell projects	e-cell-announce@e-cell.org
E-Cell users	For free discussions on E-Cell	e-cell-users@e-cell.org http://www.e-cell.org/mailman/listinfo/e-cell-users http://www.e-cell.org/moin/moin.cgi
E-Cell development	For developing E-Cell 3.0	e-cell3-devel@bioinformatics.org

Notes: ML = mailing list.

components, where productivity and readability are demanded (Table 5). E-Cell is an open source project. The entire documentation with source code is available from http://www.bioinformatics.org/E-Cell/. Bioinformatics.org is a nonprofit, academe-based organization committed to opening access to bioinformatics research projects. The publicly available mailing lists have been detailed in Table 6. Since E-Cell is an open source project, the expected third-party contribution includes algorithm modules, GUIs, new language bindings, and mathematical analysis modules. E-Cell (version 3.0) is a highly modularized software that can easily be extended by writing plug-ins.

3.2 Architecture of E-Cell

The description here mainly refers to E-Cell version 1.0. The E-Cell software constructs object models equivalent to a cell system or a part of a cell system employing Structured Substance-Reactor model (SRM). In the SRM, the objects belong to one of the three fundamental object types (Primitives): Substance, Reactor, or System. The Substances represent amounts of a molecular species or other state variables. The Reactors represent cellular phenomena that result in change in the amount or value of the molecular species or the state variables. The Systems are used as

containers for the Primitives representing functional and/or physical compartments. In E-Cell 1.0, a cell model description is composed of two parts: definition of subclasses of the Primitives (mainly of the Reactor) and a rule file. The rule file contains information of (1) a list of the three Primitive objects in the model, (2) relationships among the objects (e.g. stoichiometry of reactions), and (3) parameter values for the objects (e.g. rate constants).

3.2.1 Elements of the Control Panel

The substance window shows the quantity of a selected substance. It also allows the user to alter the quantity at will during the simulation process. The reactor window displays the activity of a selected reaction. The activity of a reaction is defined as the amount of product produced per second in the reaction process. Tracers are windows that plot the concentration of substances with time (Fig. 1).

3.2.2 Elements of the E-Cell Model

In the E-Cell system, substance is a substrate, product, catalyst, or an ion that affects a reaction. Typically, substances include proteins, protein complexes, DNA (genes), RNA, and small molecules. The total number of molecules involved in a reaction is defined as *quantity* while *concentration* describes the amount of substance present in a reaction space (in moles per liter). In the E-Cell, the quantity of a substance equals Avagadro's number times concentration times volume. The E-Cell simulation software uses the number of molecules in a sample to trace a reaction and automatically converts concentration into quantity. The spreadsheet data file needs to be converted to .er text format. It is possible to use macros in the .er file to model complicated systems with ease. In E-Cell 3.0, .er and .eri file formats are no longer used. It uses an XML-based E-Cell model description language. The E-Cell system extracts quantitative information from the rule file, links it up with the equations described in reactors, and plots the velocity curve on tracers.

A special characteristic of E-Cell is the accumulator. The Reserve Accumulator (the default feature) is used when

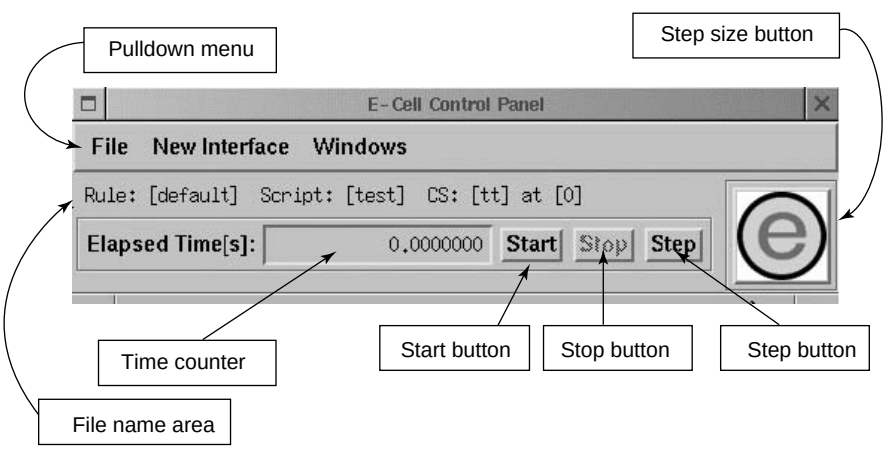

Fig. 1 Elements of the control panel.

decimal fractions are unimportant, for example, when representing an individual cell. However, the Simple Accumulator is used in situations in which the floating-point value is crucial to the interpretation of the results, for example, if the number of molecules is very large or if the simulation represents an "average cell" among a large number of cells. The MonteCarlo Accumulator is used if the simulation requires a high degree of precision in statistical analysis.

3.3 Features of E-Cell 2.0

Recently, Mitsui Knowledge Industry has released the Windows version of E-Cell (ver 2.0). The E-Cell 2.0 is very similar to the E-Cell 1.0 except that the virtual memory function (to show the concentration in the tracer for a long time) is not implemented for now. However, the E-Cell data file contains time, mean value, maximum value, and minimum value as against time and value output of version 1.0.

To run this version, the following supporting software is required:

*m4.exe distributed in Cygwin<http://sources.redhat.com/cygwin/> *Java runtime environment (http://java.sun.com/products/jdk/1.2/), *Perl (http://www.perl.com/pub/a/language/info/software.html), and *C++ (http://www.borland.com/bcppbuilder/freecompiler/)

3.4 Features of E-Cell 3.0

E-Cell 3.0 is currently being developed with an aim of providing the cell simulation community a generic and high performance software environment. It is also Linux-based and has a geometry information interface. It will integrate any sets of different simulation algorithms including the Variable-Process model, differential equation–based, diffusion reaction, and particle dynamics–based approaches. One of the main highlights of the software would be integration of subsystems with different timescales. E-Cell 3.0 allows many components, driven by different simulation algorithms and different timescales, to coexist in the simulation by employing a discrete-event worldview as its fundamental formalism.

The core simulation software of E-Cell 3.0 is a set of extension modules for the Python language interpreter, written in C++/C/Python. It consists of a libecs cell modeling toolkit, an E-Cell microcore (EMC) layer, a Python language binding (PyECS), and other peripheral Python modules. Libecs (code name: Koyurugi) is a generic object-oriented C++ class library for constructing various object-based cell models. One of the special features of Koyurugi is that the constructed cell models themselves work as simulation engines. The E-Cell Micro Core defines interfaces and implementations of the Simulator class, which provides a simple application programming interface (API) of the Koyurugi class library. PyECS is basically a Python binding of the EMC. Libecs, EMC, and a main portion of PyECS are written in C++ language.

3.5 Advantages of the E-Cell System

1. The E-Cell architecture allows users to easily add components to the E-Cell simulation software in order to address individual modeling needs.
2. E-Cell can accommodate many different kinds of simulations rather than following one specific methodology, that is, it can simulate deterministic

or stochastic models, either alone or together. Thus, users are able to model biological systems according to their characteristics, and incorporate diverse methods in the same model. In the first version of E-Cell, this is enabled by Reactors. Reactors are coded in C++, thereby offering users a huge bandwidth for simulating a large variety of reactions. In E-Cell 3.0, we will be able to create new types of Reactors, Substances, and Systems (and Stepper) allowing more flexibility. For example, users will be able to define integration method.

3. E-Cell is custom-made even for people with little/no programming knowledge.
4. E-Cell offers efficient data management through Rule files. This is particularly useful when you have large amounts of data.

3.6 Limitations of the E-Cell System

1. E-Cell 1.0 and 2.0 do not have sophisticated concentration gradient model/simulated 3-D structures and molecular dynamics at the moment. However, this constraint has been overcome in E-Cell 3.0.
2. Ordinary differential equations and algebraic equations can only be calculated explicitly. However, in most cases, users can make adjustments to components to incorporate other calculation methods (such as incorporating a library within the reactor that allows implicit calculation of ODEs).

3.7 E-Cell with 127 Genes

The E-Cell with 127 genes is a hypothetical cell that contains the minimum gene set for survival (Fig. 2). We borrowed the genomic construction from *M. genitalium* to build our first virtual cell to conduct what we call "minimum cellular metabolism." This model takes up glucose from the culture medium using a phosphotransferase system, generates ATP by catabolizing glucose to lactate through glycolysis and

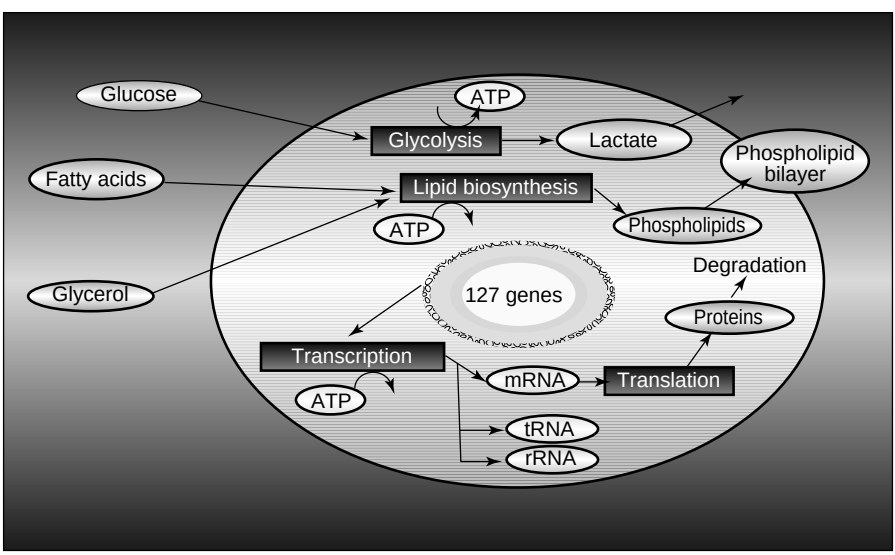

Fig. 2 E-Cell with 127 genes.

fermentation, and exports lactate out of the cell. Enzymes and substrates are spontaneously synthesized and degraded over time to sustain "life". The protein synthesis is implemented by modeling the molecules necessary for transcription and translation, namely, RNA polymerase, ribosomal subunits, rRNAs, tRNAs, and tRNA ligases. The cell also takes up glycerol and fatty acids, and produces phosphatidyl glycerol for membrane structure using a phospholipid biosynthesis pathway. The model cell is "self-supporting", but not capable of proliferating; the cell does not have pathways for DNA replication or the cell cycle. The *Mycoplasma* ammunition we used was formed of genes involved in glycolysis (9), lactate fermentation (1), phospholipid biosynthesis (4), phosphotransferase system (2), glycerol uptake (1), RNA polymerase (6), amino acid metabolism (2), ribosomal L-subunit (30), ribosomal S-subunit (19), rRNA (2), tRNA (20), tRNA ligase (19), initiation factor (4), and elongation factor (1). Overall, we ended up with 98 protein coding genes and 22 RNA coding genes. The rest (7 genes) were imported from other sources. All this was spread out into 495 reaction rules that modeled enzymatic reactions responsible for increasing/decreasing substrate/product quantities, multisubstrate complex formation, transportations of substances, and stochastic processes, for example, transcriptional factors binding factor binding to a specific site of the chromosome.

Mycoplasma genitalium was chosen for constructing a virtual cell with the minimum number of genes for survival, because it has the smallest known genome. Its genomic sequence (580 kb) was determined in 1995. We abstracted *M. genitalium's* gene set to accommodate only those genes that were required for the bare, essential cellular metabolism. At the time of developing the first version of E-Cell, 120 genes from *M. genitalium* were identified and well documented. However, in order to successfully hand-construct a self-sustaining cell, this number just fell short by 7. The shortage was made up by roping in genes from "external sources" – 4 for phospholipid biosynthesis, 1 each for nucleoside phosphate kinase and nucleoside diphosphate kinase, and 1 for tRNA ligase. At that time, the phospholipid biosynthesis and a few other pathways in *M. genitalium* were not well characterized. The information on the kinetic properties of genes and proteins was mostly obtained from the KEGG and BioCyc (previously called EcoCyc) databases.

3.8 Applications of the E-Cell System

The E-Cell with 127 genes predicted for the first time a sudden sharp increase in the ATP level in a glucose-starved cell followed by the equally sharp decrease of ATP. Though the second event was expected, the first one was a big surprise and was later confirmed experimentally. This anomalous situation was explained by the fact that though ATP production was stalled by cutting off glucose levels, it took the cell a little while to consume the intermediates for ATP production, thus, resulting in a sudden increase of ATP. This example demonstrates the potential of *in silico* modeling in generating new information. Furthermore, recent pointers indicate a paradigm shift from experimental biology to *in silico* biology.

Overall, the E-Cell has applications in the following areas:

1. *Metabolic requirements*: The assessment of the metabolic requirements of the

cell is an area E-Cell can successfully address. At present, *M. genitalium* is grown in a complex medium containing several chemically undefined components including fetal bovine serum and also yeast and beef extracts. By combining knowledge of metabolic enzymes present in the cell with information concerning protein transporters of metabolites across the cell membrane, it should be possible, by using the E-Cell model, to evaluate whether a particular defined medium can support growth.

2. *Gene expression*: E-Cell software can be used to decipher gene regulatory networks. We plan to use *M. genitalium* to achieve this objective.

3. *Minimal gene set*: The self-sustaining E-Cell will be further extended to define the minimal set of genes required for a self-replicating cell under a specific set of laboratory conditions.

4. *Clinical applications*: We are currently exploring clinical applications of the E-Cell, for example; diabetes, red blood cell enzyme deficiencies, and so on (see Sect. 3.9).

3.9 Simulation of Red Cell Enzyme Deficiencies

Glucose-6-phosphate dehydrogenase (G6PD) is a key enzyme that produces NADPH in the pentose phosphate pathway (Fig. 3). G6PD converts glucose-6-phosphoric acid into 6-phosphoglucono-1,5-lactone, and generates NADP, which

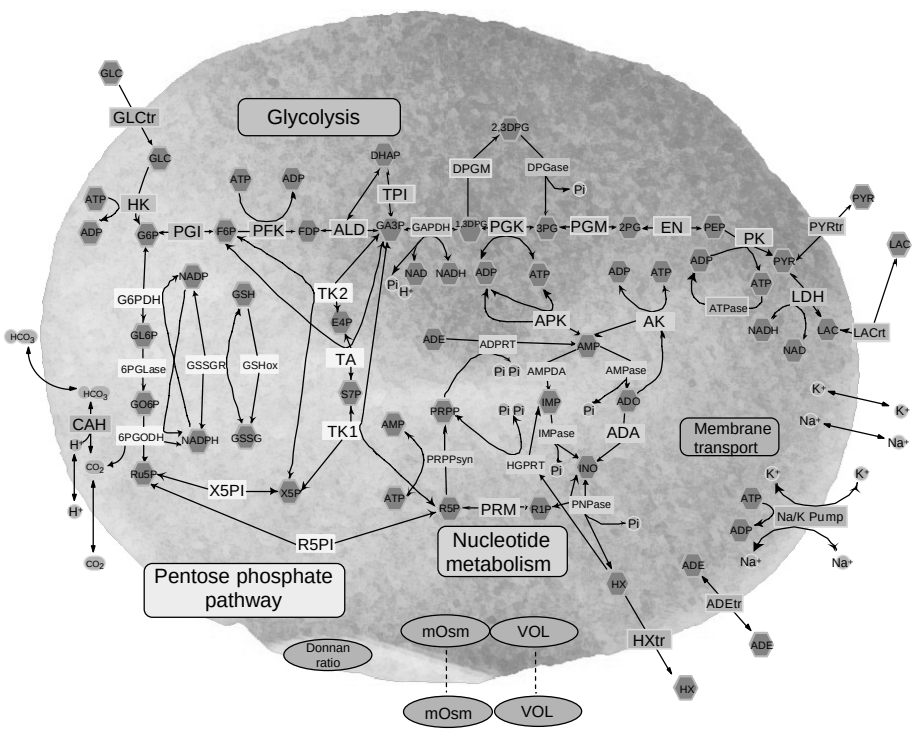

Fig. 3 Whole erythrocyte model.

is then metabolized into ribulose-5-phosphoric acid via 6-phosphogluconic acid generating NADPH in the process. A major function of GSH in the erythrocyte is to eliminate superoxide anions and organic hydroperoxides. Peroxides are eliminated through the action of glutathione peroxidase, yielding oxidized glutathione (GSSG).

We have implemented the G6PD deficiency into the E-Cell model and modified the kinetic parameters in accordance with the biochemical environment of mutant cells from patients reporting this deficiency. The simulation experiments were carried out with steady state concentrations corresponding to those of the normal erythrocyte. Sequential changes in the quantity of NADPH, GSH, and ATP were observed in the simulation experiments. However, the longevity of our computer model estimated by the concentration of ATP turned out to be much shorter than that of the real erythrocyte with G6PD deficiency. This difference was, presumably, due to the lack of pathways producing GSH and the export system of GSSG. After modification, the longevity of the cell and the GSH/GSSG ratio was found to have increased. Our result indicates that these pathways partially compensate for the reduction of GSH and have a role in easing anemia, a condition of G6PD deficiency. This result can be a good explanation for the fact that G6PD deficiency is the most common cause of anemia. If the deficiency with these compensation pathways has no severe disadvantage for survival, it would spread through the population. When the activity of G6PD decreases, the activity of 6-phosphogluconate dehydrogenase increases, compensating for the reduced production of NADPH. However, because either 6-phosphoglucono-1,5-lactone was not supplied or because G6P was deficient, the 6-phosphogluconic acid supply was quickly exhausted and the production of NADPH was stopped. The amount of NADPH started to reduce gradually and was soon exhausted. Subsequently, GSH started to decrease and was soon completely converted into GSSG. Finally, the cells worsened in their metabolic performance and exhausted all the ATP because of the inhibition of rate-determining enzymes (due to low GSH/GSSG). This model of G6PD deficiency correlates well with the clinical situation and can be a test bed for extending the model to other human erythrocyte metabolic disorders (Fig. 4). We are also trying to reconstruct the erythrocyte model with pyruvate kinase (PK) deficiency. On creating an *in silico* model for PK deficiency, the ATP production rate was found to decrease proportionately leading to eventual elimination from the system. Simultaneously, an increase in the concentration of 2-phosphogrycerate, 3-phosphoglycerate, and phosphoenolpyruvate was observed, which is in agreement with the clinical presentation of the PK phenotype.

4 Concluding Remarks

In order to understand the *whole* we must model the *whole*, because a cell is basically systems biology, by default. Given the enormous complexity and data generated by genome, proteome, transcriptome, and metabolome, computer simulations are indispensable for future biological research. Whether or not it is feasible to construct a computer model of a whole living cell is still an open question. Although no attempts were made toward whole-cell modeling until the late

Effect on ATP

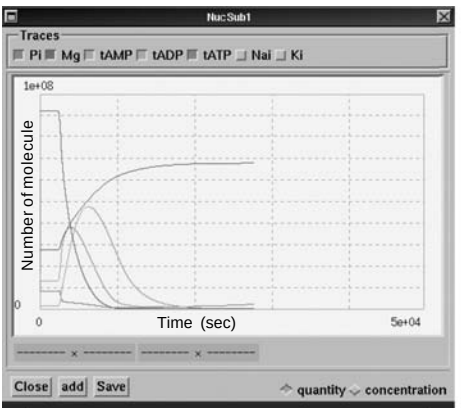

Pi: Inorganic phosphates, Mg: Free magnesium tAMP: total AMP, tADP: total ADP, tATP: total ATP

Effect on metabolic intermediates of glycolysis

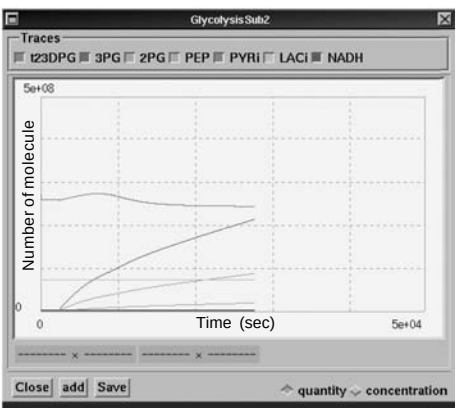

t23DPG: total 2,3-Diphosphogylcerate, 3PG: 3-Phosphoglycerate, 2PG: 2-Phosphoglycerate
PEP: Phosphoenolpyruvate, LACi: Lactate

Fig. 4 Simulation of red cell deficiency.

1990s, the importance of computer simulation of cellular metabolism has been realized since the 1980s. Various cellular transactions like gene expression, cell cycle, signal transduction, and metabolic pathways have been modeled and simulated independently. However, to understand the cross talk among these seemingly ''self-regulating systems'', it is necessary to construct an integrated model of the cell. One of the major problems in constructing large-scale models is a lack of quantitative data. Most of the biological knowledge is qualitative (in the form of pathway maps). The quantitative data available in the published literature is often noisy and not well suited for simulations. Thus, a major challenge is to collect a large amount of highly accurate and, preferably, time series data, construct quantitative models, and train the models with more wet-bench results till simulation matches real life biology. To achieve this objective, we not only need good data but also a novel computational and software engineering approach. This is the biology of the twenty-first century.

See also Animal Biotechnology and Modeling.

Bibliography

Books and Reviews

Hasty, J., McMillen, D., Isaacs, F., Collins, J.J. (2001) Computational studies of gene regulatory networks: *in numero* molecular biology, *Nat. Rev. Genet.* **2**, 268–279.

De Jong, H. (2002) Modeling and simulation of genetic regulatory systems: a literature review, *J. Comp. Biol.* **9**, 67–103.

Hargrove, J.H. (1994) *Kinetic Modeling of Gene Expression*, R.G. Landes Company, Austin, USA.

Tomita, M., Hashimoto, K., Takahashi, K., et al. (2000) The E-cell project: towards integrative simulation of cellular processes, *New Gen. Comput.* **18**, 1–12.

Tomita, M. (2001) Whole cell simulation: a grand challenge of the 21st century, *Trends Biotechnol.* **19**, 205–210.

Ideker, T., Galitski, T., Hood, L. (2001) A new approach to decoding life: systems biology, *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372.

Kitano, H. (2002) Computational systems biology, *Nature* **420**, 206–210.

Hasty, J., McMillen, D., Collins, J.J. (2002) Engineered gene circuits, *Nature* **420**, 224–230.

Primary Literature

Meyers, S., Friendland, P. (1984) Knowledge based simulation of genetic regulation in bacteriophage lambda, *Nucleic Acids Res.* **12**, 1–9.

Koile, K., Overton, G.C. (1989) A qualitative model for gene expression, *Proc. of the 1989 Summer Comp. Simu. Conf.*, pp. 415–421.

Palsson, B.O., Narang, A., Joshi, A. (1989) Computer model of human erythrocyte metabolism, *Prog. Clin. Biol. Res.* **319**, 133–150.

Mendes, P. (1993) GEPASI: a software package for modeling the dynamics, steady states and control of biochemical and other systems, *Comput. Appl. Biosci.* **9**, 563–571.

Sauro, H.M. (1993) SCAMP: a general purpose simulator and metabolic control analysis program, *Comput. Appl. Biosci.* **9**, 563–571.

Novak, B., Tyson, J.J. (1995) Quantitative analysis of a molecular model of mitotic control in fission yeast, *J. Theor. Biol.* **173**, 283–305.

Ni, T.C., Savageau, M.A. (1996) Model assessment and refinement using strategies from biochemical systems theory: application to metabolism in human red blood cells, *J. Theor. Biol.* **179**, 329–368.

McAdams, H.H., Shapiro, L. (1995) Circuit simulation of genetic networks, *Science* **269**, 650–656.

McAdams, H.H., Arkin, A. (1997) Stochastic mechanisms in gene expressions, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 814–819.

Mendes, P. (1997) Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3, *Trends Biochem. Sci.* **22**, 361–363.

Goryanin, I., Hodgman, T.C., Selkov, E. (1999) Mathematical simulation and analysis of cellular metabolism and regulation, *Bioinformatics* **15**, 749–758.

Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C. (1999) E-Cell software environment for whole cell simulation, *Bioinformatics* **15**, 72–74.

Chen, K.C., Csikasz-Nagy, A., Gyorffy, B., et al. (2000) Kinetic analysis of a molecular model of the budding yeast cell cycle, *Mol. Biol. Cell* **11**, 369–391.

Tomita, M. (2001) Whole cell simulation: a grand challenge of the 21st century, *Trends Biotechnol.* **19**, 205–210.

Takahashi, K., Yugi, K., Hashimoto, K., Yamada, Y., Pickett, C.J.F., Tomita, M. (2002) Computational challenges in cell simulation: a software engineering approach, *IEEE Intell. Syst.* **17**, 64–71.

Rao, C.V., Wolf, D.M., Arkin Adam, P. (2002) Control, exploitation and tolerance of intracellular noise, *Nature* **420**, 231–237.

E-Cell: Computer Simulation of the Cell: *see* E-Cell: Computer Simulation of the Cell

Electric and Magnetic Field Reception

Eberhard Neumann
University of Bielefeld, Bielefeld, Germany

1 Concepts 3
1.1 Natural Fields 3
1.2 Definition of Field Reception 4

2 Field and Molecular Motion 4
2.1 Internal and External Fields 4
2.2 Thermal Motion 5
2.3 Freely Mobile Charges 6
2.4 Dipole Orientation and Dipolophoresis 6
2.5 Electric Currents 6

3 Chemical Reaction Equilibria 7
3.1 Dipole Reaction 7
3.2 The Weak-field Approximation 9
3.3 Conformational State Transition 10
3.4 Ligand Binding 10
3.5 Dipole-pairing 10

4 Rate Constants 11
4.1 Dipole Transitions 11
4.2 Ion Equilibrium 12
4.3 Activity Coefficients 13
4.4 Ion Flows 13

5 Amplification Mechanisms 14
5.1 Field Enhancement 14
5.2 Cooperativity and Phase Transition 16

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

5.3 Amplification at High Basis Fields 18
5.4 Comparison of AC/DC Effects 18

Bibliography 20

Keywords

Chemical Electric and Magnetic Field Effects
The consequences of the primary field effects on ions and dipoles, which cause the rate and equilibrium constants of reactive chemical processes and phase transitions to depend on the electric field strength and, to a weaker extent, on the magnetic flux density.

Cooperativity
A mechanism for the simultaneous state transition of a cluster or domain of strongly coupled molecules in a narrow field-strength range.

Electric and Magnetic Field Strength
Measures of the force on electric charges and electric and magnetic dipoles, respectively. The targets are either freely mobile or are fixed ionic and magnetic groups of larger macromolecules.

Interfacial Electric Polarization
Field-induced ion accumulation at interfaces of media with different dielectric constants, leading to ionic charging up of membranes and to large induced transmembrane voltages by small external fields.

Magnetic Induction of Electric Fields
The main effect of time-varying magnetic fields, causing ionic (AC) currents and dipolar displacement currents, both of which interfere with chemical reactivity.

Magnetosomes
Small elongated magnetite (Fe_3O_4, Fe_3S_4) organelles that, in magnetobacteria, serve for geomagnetic field sensing and direction finding and are also found in cells of higher organisms.

Electric field reception (E) and magnetic field reception (B) are defined as the mutual interaction of external fields with the relatively high internal fields of the ionic and the (electric and magnetic, respectively) dipolar components including the cellular magnetosomes involved in sensory processes. The natural E fields are particularly apparent in the electrophysiological function of nerves. Physicochemically, the E and B fields primarily act on charges, changing their positions and motional states. As a consequence, the extent and the rate of chemical reactions, conformational changes, and phase transitions of the field-receiving molecules are dependent on the strength of the E and B fields. Although there are *always* finite chemical shifts in rate and equilibrium constants, weak external fields require amplification processes (e.g. interfacial polarization of larger particles and cooperativity in domain structures).

1 Concepts

In biology, electric and magnetic field reception is commonly associated with sensory processes. The electric fish have developed special electric organs to sense the environment for orientation, navigation, and food search. Electric eels and torpedo fish utilize large electric discharge organs for prey capture. Magnetobacteria contain small magnetite (Fe_3O_4, Fe_3S_4) organelles that serve for direction finding in the earth's magnetic field. Such magnetosomes are also found in higher organisms. It is important to note that biological weak-field reception and behavioral responses are intimately coupled with the nervous system. The receptive and sensing fields of the biosphere are either static (DC) fields or short field pulses of 0.1 to 1 ms duration with low pulse frequencies ($\leq$3 kHz), producing direct currents (DC) or current pulses.

Electrotechnologically, manmade electric and magnetic field devices have become indispensable tools of life. In medicine, in particular, we observe a growing application of the electric and magnetic field apparatus for both diagnostic and therapeutic purposes. On the other hand, it is not yet clear whether the technical fields of power transmission lines and electrical household devices bear potential health risks, even under normal safety conditions. To estimate the possible beneficial or negative effects, it is necessary to understand the physical and chemical interaction mechanisms of electric and magnetic field reception.

Generally, electric and magnetic fields act as forces on charges: electrolyte ions and fixed ionic groups, as well as free electric and magnetic dipolar molecules and fixed dipolar groups of macromolecules, including membrane lipids and cellular magnetosomes. As a consequence of these primary field effects, reactive chemical interactions are dependent on the electric field strength $\mathbf{E}$ (Vm^{-1}) and, to a much weaker extent, on the magnetic field strength $\mathbf{H}$ (Am^{-1}).

1.1 Natural Fields

All life processes have evolved in the natural electric and magnetic force fields of the earth. These geofields are quasi-static or only slowly time-varying. Compared to

possible technical fields, they are classified as weak. The average geoelectric field strength is $|\mathbf{E}_0| \approx 200\ \mathrm{Vm}^{-1}$, the average geomagnetic flux density $|\mathbf{B}_0| = \mu_0\mu|\mathbf{H}_0| \approx 50\ \mu\mathrm{T}$, where $\mu_0 = 1.256 \times 10^{-6}$ Vs/(Am) is the vacuum permeability and the magnetic susceptibility $\chi_{\mathrm{ma}} = \mu - 1$.

Even in its lowest forms, life makes use of electric and magnetic field effects, most obviously, on the microscopic level, in the rapid signal transmission of neurons, involving, in particular, the gated transport of Na^+, K^+, Ca^{2+}, and Cl^--ions.

1.2 Definition of Field Reception

In molecular biology, the term *field reception* appears meaningful only if it comprises more than just the direct interactions of the force fields (and their spatial gradients) with molecules, changing their positional and motional state. An adequate definition of this term must specify at least the next biochemical or biophysical process following the preceding direct force effects. Thus, field reception indispensably includes field effects on the reactive interactions, on the molecular orientations, and on the transport processes of the "field-receiving"ions and electric and magnetic dipoles. Moreover, amplification mechanisms for both small external fields and for locally small chemical shifts are important features of the molecular characterization of field reception.

Briefly, the effects of external E and B fields may be summarized as a cascade of mutual interaction steps:

$$(E, B)_{\mathrm{ext}} \longleftrightarrow (E, B)_{\mathrm{int}} \longleftrightarrow \left| \begin{array}{l} \Delta c,\ \Delta\varphi_{\mathrm{m}} \\ \Delta K/K_0,\ \Delta\beta/\beta_0 \\ \Delta k/k_0 \end{array} \right|$$

The primary step in field reception (and emission) is the mutual interference of the external E and B fields, $(E, B)_{\mathrm{ext}}$, with the local (body and cell) internal fields $(E, B)_{\mathrm{int}}$, provided the external static and low-frequency E fields can penetrate the systems, for instance, via direct (metal) electrode contact. The subsequent steps may occur in parallel (Fig. 1): (1) Mutual changes in electrolyte flows (resulting in local ion concentration changes Δc) and in the internal fields, including the relatively very strong E fields of the cell membranes ($E_{\mathrm{m}} 10^3$–$10^4\ \mathrm{kVm}^{-1}$) corresponding to electric membrane potential differences on the order of $|\Delta\varphi_{\mathrm{m}}| = 0.01$ to 0.2 V and (2) field-induced changes of ligand-binding processes, of conformational transitions, and of cooperative domain phase transitions (e.g. in membranes).

Theoretically, the field-induced changes may be quantified as relative changes ($\Delta K/K_0$, $\Delta\beta/\beta_0$, $\Delta k/k_0$) in the equilibrium constants (K), in the extents of ligand binding and of transitions (β), and in the rate coefficients (k).

2 Field and Molecular Motion

2.1 Internal and External Fields

To understand the effects of external fields on freely mobile ions, on dipolar charge pairs, and on "fixed" ionic or dipolar groups of larger molecules, it is important to realize that these ions and dipoles are themselves the sources of local fields, which, near the charges, have an impressively high field strength (on the order of 3000 kVm^{-1} at body temperature). The local electric fields are largely screened by mobile counterions (ionic atmosphere) or

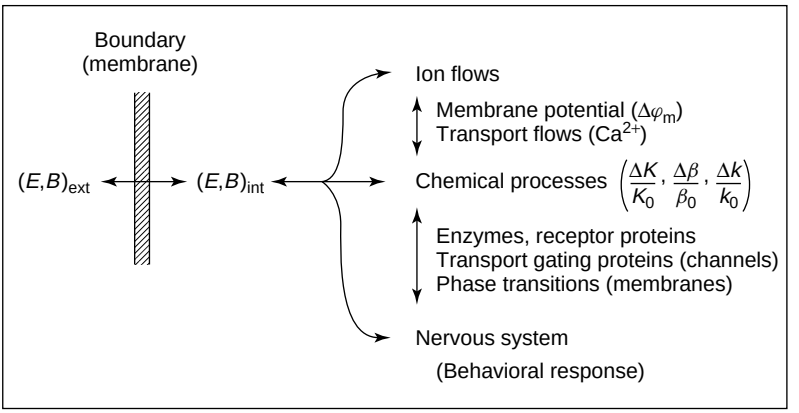

Fig. 1 Interaction scheme for electric (E) and magnetic (B) field reception. Direct current, static, and low-frequency external E fields ($\mathbf{E}_{ext}$) induce noticeable changes in the body's internal fields only if there is direct contact (e.g. via metal electrodes). The surfaces of conducting matter represent no barriers for high-frequency E fields and all magnetic fields. The $(\mathbf{E}, \mathbf{B})_{ext}$-induced changes in $(\mathbf{E}, \mathbf{B})_{int}$ modulate ion flows and chemical processes, in particular, in and on membranes. Weak-field reception and emission by special skin organs always involves the nervous system.

by neutralizing ion pairing. The motion of charges in moving fluids (like the electrolyte ions in the blood) always produces magnetic fields that in turn give rise to electric potential gradients.

External electric and magnetic fields have to compete with the internal biofields. The **E** (DC) fields and the low-frequency (LF) alternating (AC) **E** fields of the air are effectively reduced by conducting material like the skin of the human body. At low frequencies ($\leq$3 kHz), the internally induced electric fields in conductive bodies are therefore extremely small. On the other hand, external magnetic fields are not screened; they permeate bodies. The main effect of time-varying magnetic fields is the induction of time-varying electric fields that cause dissipative ionic (AC) currents as well as electric dipolar displacement (capacitive) currents.

On the level of cells and organelles, the dielectric material of the plasma membrane has condensor properties for the storage of electric energy. External fields cause changes in the ion accumulations at the membrane–solution interfaces. Externally induced interfacial polarizations can lead to enormous amplification of the transmembrane electric field, which in turn can appreciably affect the reactions of membrane proteins and the various transmembrane transport phenomena.

2.2 Thermal Motion

Another important aspect is that the electric and magnetic interaction energy $\Delta_r G_{(F)}$ must compete with the randomizing thermal motion of the species. The thermal noise energy is given by $k_B T$, or on a molar scale, by RT, where k_B is the Boltzmann constant, $R = N_A \cdot k_B$ is the gas constant, N_A the Avogadro constant, and T is the thermodynamic (Kelvin) temperature. In order to be relevant, the electric and magnetic field effects

must emerge from thermal noise and the normal concentration fluctuations.

2.3 Freely Mobile Charges

The **E** and **B** fields are a measure of the respective force line densities. **E** and **B** are defined by the electric and magnetic forces $\mathbf{F}_{el}$ and $\mathbf{F}_{ma}$ exerted on the charge q respectively changing the velocity **v** of free charges and the position of fixed and dipolar charge pairs.

The direction of change is globally given by the charge number z_i (with sign) of the ion or ionic group i such that $q_i = z_i e_o$, where $e_o = 1.6 \times 10^{-19}$ As is the positive elementary charge, and by the drift speed,

$$\mathbf{v}_i = \frac{z_i}{|z_i|} u_i \mathbf{E} \qquad (1)$$

where u_i is the electric mobility. The Coulomb law $\mathbf{F}_{el} = q\mathbf{E}$ and the Lorentz law $\mathbf{F}_{ma} = q(\mathbf{v} \times \mathbf{B})$ yield the total force on q:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \qquad (2)$$

The vector cross product $\mathbf{v} \times \mathbf{B} = |\mathbf{v}||\mathbf{B}| \sin\theta$, where θ is the angle between the vectors **v** and **B**, describes the circular motion of a charge around the **B** lines. Note that $\mathbf{F}_{ma}$ is maximal if **v** and **B** are at right angle and $\mathbf{F}_{ma} = 0$ if **v** and **B** are parallel or if the charge is fixed ($\mathbf{v} = 0$).

2.4 Dipole Orientation and Dipolophoresis

Electric and magnetic dipoles are oriented in homogeneous **E** and **B** fields respectively. The fields exert a torque against the thermal energy $k_B T$. In spatially inhomogeneous fields, dipoles not only rotate into the field direction but also move translationally in the direction of higher or lower field strengths (dipolophoresis, sometimes inappropriately called dielectrophoresis). The total force $\mathbf{F}(r)$ exerted by, say, a radially inhomogeneous field is given by the dot products

$$\mathbf{F}(r) = m_{el} \cdot \frac{\partial \mathbf{E}}{\partial r} + m_{ma} \cdot \frac{\partial \mathbf{B}}{\partial r} \qquad (3)$$

where the partial space derivatives are a measure of the inhomogeneity of the respective fields.

2.5 Electric Currents

The motions of charges and dipoles may be measured as ohmic and capacitive currents. The current density **J** (A/m^2) is related to the field strength by

$$\mathbf{J} = \kappa \cdot \mathbf{E} + \frac{\partial \mathbf{D}}{\partial t} \qquad (4)$$

where κ is the ionic conductivity of the medium and $\mathbf{D} = \varepsilon_0 \varepsilon \mathbf{E}$ is the dielectric displacement vector (covering dipole reorientations), where $\varepsilon_0 = 8.854 \times 10^{-12}$ As/(Vm) is the vacuum permittivity and ε is the (relative) dielectric constant of the homogeneous medium.

In polyelectrolytes like DNA and on polyionic membrane surfaces, an external field can cause displacements of the counterion atmospheres, thereby inducing stationary "ionic flow-dipoles" having appreciably high dipole moments.

The circular-current density vector induced by a homogeneous magnetic field acting on moving charges q_i ($\mathbf{v}_i$) is given by the sum of all ion contributions

$$\mathbf{J} = \kappa \mathbf{E}_{ind} = \kappa \Sigma_i(-\mathbf{v}_i \times \mathbf{B}) \qquad (5)$$

Time-varying magnetic fields induce ion flows by magnetic induction. The voltage (U_{ind}) induced around a closed path of length $\ell = \int d\mathbf{s}$, where $d\mathbf{s}$ is a differential

path element, is given by

$$U_{\text{ind}} = \oint \mathbf{E}\,\mathrm{d}\ell = -\iint \left(\frac{\partial \mathbf{B}}{\partial \mathrm{t}}\right) \mathrm{d}\mathbf{s} \quad (6)$$

where $\partial\mathbf{B}/\partial\mathrm{t}$ is the time derivative of $\mathbf{B}$. Since the dot product $(\partial\mathbf{B}/\partial\mathrm{t}) \cdot \mathrm{d}\mathbf{s} = (\partial|\mathbf{B}|/\partial\mathrm{t})\mathrm{d}|\mathbf{s}|\cos\theta$, U_{ind} is maximal if $\mathbf{B}$ is perpendicular to the path (surface), that is, parallel to $\mathbf{s}$. The magnitude of the $\mathbf{E}(\mathrm{t})$ field of frequency f at a perpendicular distance $\mathbf{r}$ from the center of a path around the $\mathbf{B}$ direction is $|\mathbf{E}_{\text{ind}}| = \pi \mathrm{f}|\mathbf{B}||\mathbf{r}|$ and the current density $|\mathrm{J}| = \kappa\pi \mathrm{f}|\mathrm{B}||\mathrm{r}|$ increases linearly with frequency. Magnetically induced currents are generally about 10^3 times larger than those induced directly by external electric fields $\mathbf{E}_0(f)$.

3 Chemical Reaction Equilibria

In biochemistry, we frequently encounter ligand (L)-induced changes in the conformation of macromolecular binding sites (B). This induced fit is described by the scheme

$$\mathrm{L} + \mathrm{B} \overset{(K_1)}{\rightleftharpoons} \mathrm{LB} \overset{(K_2)}{\rightleftharpoons} \mathrm{LB}' \quad (7)$$

where the conformational transition $\mathrm{B} \rightleftharpoons \mathrm{B}'$ is inherent. The physically more realistic reaction scheme comprises two binding steps and two structural transitions:

$$\begin{array}{ccc} \mathrm{L}+\mathrm{B} & \underset{}{\overset{(K_1)}{\rightleftharpoons}} & \mathrm{LB} \\ (K_4)\ \Big\updownarrow & & \Big\updownarrow\ (K_2) \\ \mathrm{L}+\mathrm{B}' & \underset{(K_3)}{\rightleftharpoons} & \mathrm{LB}' \end{array} \quad (8)$$

The individual equilibrium constants $K_1 = [\mathrm{L}][\mathrm{B}]/[\mathrm{LB}]$, $K_3 = [\mathrm{L}][\mathrm{B}']/[\mathrm{LB}']$ and $K_2 = [\mathrm{LB}']/[\mathrm{LB}]$, $K_4 = [\mathrm{B}']/[\mathrm{B}]$, where the brackets denote molar concentrations, are related by $K_1/K_3 = K_2/K_4$.

The apparent equilibrium constants K, defined as equilibrium concentration ratios, may vary with concentration and with the ionic strength because of the screening effects of ions and other nonreactive interactions. When a reaction is written in the suggestive form $0 = \Sigma_\mathrm{i}\nu_\mathrm{i}\mathrm{J}$, where ν_j is the stoichiometric coefficient of the reaction partner J (positive for a product, negative for an educt), the "really constant" thermodynamic equilibrium constant $K^{\ominus}$ is defined as the equilibrium activity ratio:

$$K^{\ominus} = \Pi a_\mathrm{j}^{\nu_\mathrm{j}} = |K| \cdot Y = \exp\left(\frac{-\Delta_\mathrm{r}\hat{\mathrm{G}}^{\ominus}}{RT}\right), \quad (9)$$

where $\Delta_\mathrm{r}\hat{\mathrm{G}}^{\ominus}$ is the standard value of the transformed Gibbs reaction energy. See Eq. (15).

The thermodynamic activity a_j is a dimensionless quantity expressed as $a_\mathrm{j} = |c_\mathrm{j}|\gamma_\mathrm{j}$, where $|c_\mathrm{j}|$ is the numerical value of the concentration c_j and γ_j is the thermodynamic activity coefficient. Hence, from $K^{\ominus} = |K|Y$, we see that $K = \Pi c_\mathrm{j}^{\nu_\mathrm{j}}$ and $Y = \Pi \gamma_\mathrm{j}^{\nu_\mathrm{j}}$. Note that Eq. (9) is used to relate experimental data via $K^{\ominus}$ to the reaction energetics.

3.1 Dipole Reaction

For the general analysis of field-induced changes, the scheme $0 = \Sigma_\mathrm{j}\nu_\mathrm{j}\mathrm{J}$ is specified in terms of reactive educts $\mathrm{J_r}$ and products $\mathrm{J_p}$ according to

$$\Sigma_\mathrm{r}|\nu_\mathrm{r}|\mathrm{J_r} \underset{k_\mathrm{p}}{\overset{k_\mathrm{r}}{\rightleftharpoons}} \Sigma_\mathrm{p}\nu_\mathrm{p}\mathrm{J_p} \quad (10)$$

where k_r and k_p are the rate coefficients for the forward and backward reaction steps

respectively. The molar reaction dipole moment $\Delta_r M^{\ominus}$ relevant for field effects is defined as the difference

$$\Delta_r M^{\ominus} = M_p - M_r = N_A \langle \Delta_r m \rangle \quad (11)$$

where M_p and M_r are the average molar moments of the products and reactants respectively. The molecular reaction average $\langle \Delta_r m \rangle$ is given by the individual contributions of the reactant and product molecules:

$$\langle \Delta_r m \rangle = \Sigma_r |\nu_r| |\mathbf{m}_r| \langle \cos\theta \rangle - \Sigma_p \nu_p |\mathbf{m}_p| \langle \cos\theta \rangle \quad (12)$$

The brackets $\langle\rangle$ denote orientational averages and θ is the angle between the dipolar axis of a molecule J and the field direction.

Introducing now $F = F_{(E,B)} = |\mathbf{E}|$ or, in the other case, $F = F_{(E,B)} = |\mathbf{B}|$ for the amount of the electric field strength vector $\mathbf{E}$ and that of the magnetic flux density vector $\mathbf{B}$ respectively, the general expression of the (isothermal, isobaric) field-induced change of $K^{\ominus} = \exp[-\Delta_r \hat{G}^{\ominus}/RT]$, see Eq. (9), is given by a van 't Hoff-like expression:

$$\left(\frac{\partial \ln K^{\ominus}}{\partial F}\right)_{p,T} = \frac{\Delta_r M^{\ominus}}{RT} \quad (13)$$

For $F = |\mathbf{E}|$, $\langle \Delta_r m \rangle$ is the electric reaction moment; for $F = |\mathbf{B}|$, $\langle \Delta_r m \rangle$ refers to the molar difference in the magnetic moments. Integration of Eq. (13) yields

$$K^{\ominus} = K_0^{\ominus} \cdot e^X \quad (14)$$

$K_0^{\ominus} = \exp[-\Delta_r G^{\ominus}/RT$ is the K-value at zero field, where $\Delta_r G^{\ominus}$ is the standard Gibbs reaction energy. At $F = 0$, $\Delta_r \hat{G}^{\ominus} = \Delta_r G^{\ominus}$. The "field effect exponent" X refers to the electric or magnetic free energy, respectively:

$$\Delta_r G_{(F)} = -\left(\Delta_r \hat{G}^{\ominus} - \Delta_r \hat{G}^{\ominus}\right) = RT \cdot \ln\left(\frac{K^{\ominus}}{K_0^{\ominus}}\right) \quad (15)$$

and

$$X = \frac{\Delta_r G_{(F)}}{RT} = \frac{\int \Delta_r M^{\ominus}\, dF}{RT} = \ln \frac{K^{\ominus}}{K_0^{\ominus}} \quad (16)$$

The integration boundaries are F and $F = 0$, respectively. If the apparent equilibrium constant K is not independent of the concentrations of the interaction partners or of the ionic strength, extrapolation to zero concentration or zero ionic strength yields the reaction-specific electrothermodynamic quantities $K^{\ominus}$, $K_0^{\ominus}$, and $\Delta_r M^{\ominus}$. Experimentally, it is usually K that is accessible via the fractional extent of a process. Substitution of Eq. (9) into Eq. (13) yields

$$\left(\frac{\partial \ln |K|}{\partial F}\right)_{p,T} = \frac{\Delta_r M^{\ominus}}{RT} - \frac{\partial \ln(Y_E/Y)}{\partial F} \quad (17)$$

The absolute signs refer to cases of K with a concentration dimension. Integration results in

$$K = K_0 \cdot e^X \cdot \frac{Y}{Y_E} \quad (18)$$

where Y is the normal activity factor at $F = 0$ and, purely formally, Y_E is that at F. If Y can be considered independent of F, the approximation of $Y/Y_E = 1$ holds and $K = K_0 \cdot e^X$ is used instead of Eq. (14).

In the case of induced dipole changes caused by the electric displaceability of ionic groups i, in an external field E, for instance, in a gating protein, $\Delta_r M = N_A \cdot \Delta_r \alpha \cdot E_{int}$ may be used. If the internal field E_{int} can be set equal

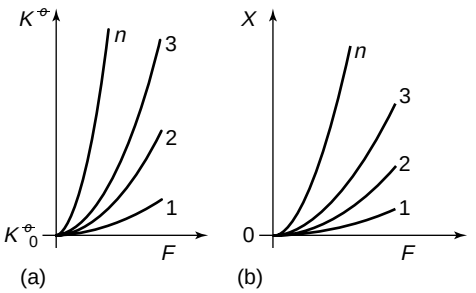

Fig. 2 Field dependences of (a) the equilibrium constant $K^{\ominus} = K_0^{\ominus}e^X$ and (b) the field-effect exponent $X = \int \Delta_r M dF/(RT) \propto F^2$. The steepness $(\partial K/\partial F)_{p,T} = K \cdot \Delta_r M/RT)$ and $(\partial X/\partial F)_{p,T} = \Delta_r M/RT$ both increase with $\Delta_r M$ and with the domain size n. Since the relative change $\Delta\beta/\beta_0$ in the advancement of a chemical process is proportional to $X \propto F^2$, we obtain $\Delta\beta/\beta_0 \propto F^2$. The rate constants k exhibit qualitatively the same field dependences as K; therefore, generally $\Delta k/k_0$ is proportional to F^2.

to E, the molecular reaction polarization $\Delta_r\alpha = \alpha_p - \alpha_r$ reflects an average over the displacements of all groups i. Substitution into Eq. (16) yields (with $F = E$)

$$X = \frac{\int \Delta_r M \, dE}{RT} = \frac{\Delta_r\alpha \cdot E^2}{2k_B T} \qquad (19)$$

In the simplest case, a permanent dipole rotation in a planar membrane is indicated, if $X = \Delta_r M \cdot E/(RT)$ is linearly dependent on E. On the other hand, if X depends on E^2, Eq. (19) is used to determine the average reaction polarizability of an induced dipole mechanism.

3.2 The Weak-field Approximation

When $\Delta_r G_{(F)} \ll RT$, then $X \ll 1$ and the factor e^X in $K = K_0 e^X$ can be expanded into a power series $e^X = 1 + X + \ldots$, where higher order terms are negligibly small. Therefore, for small field effects, Eq. (14) takes the form $K^{\ominus} = K_0^{\ominus}(1 + X)$.

For practical purposes, a relative change $\Delta K/K_0$ is introduced, which readily permits expression of field-induced changes in percent. With $K = K_0 e^X$ as an approximation to $K^{\ominus} = K_0^{\ominus} e^X$, we obtain

$$\frac{\Delta K}{K_0} = \frac{K - K_0}{K_0} = e^X - 1 \qquad (20)$$

The small-field approximation ($X \ll 1$) now reads

$$\frac{\Delta K}{K_0} = X \qquad (21)$$

For small field effects in a random ensemble of permanent or induced dipolar reaction partners, $\langle \Delta_r m \rangle$ is proportional to F. Hence, X is proportional to $F^2/(2RT)$ and thereby $\Delta K/K_0$ is proportional to $F^2/(2RT)$; see Fig. 2. In summary, if $\Delta_r M \cdot F \ll RT$, the field-induced changes in the equilibrium concentrations are proportional to the *square* of the field strength.

The electromagnetodynamic approach so far gives general guidelines for the experimental access to field effects. On the other hand, the analysis clearly shows that it is not so much the magnitude of the individual moments of the reaction partners that matters; rather, the *moment difference* as the reaction moment must be nonzero for finite field effects.

We also see in $K = K_0(1 + X)$ and $X = \int \langle \Delta_r m \rangle dE/k_B T$ that, no matter how small the *difference* $\langle \Delta_r m \rangle$ in the electric or magnetic dipole moments of the

interaction partners is, there is *always a finite change* caused by a field. Of course, the inevitable thermal motion leads to local activity fluctuations $\delta a_j = y_j \delta c_j + c_j \delta y_j$. We must determine the *relevance* of the small, field-induced changes Δc_j and Δy_j in concentration and activity coefficient respectively compared with the random thermal fluctuations $\pm\delta c_j$ and $\pm\delta y_j$.

3.3 Conformational State Transition

In a field, an intramolecular state transition or conformational change of the type.

$$B \rightleftharpoons B', \tag{22}$$

proceeds in the direction of B′, if the dipole moment $m(B') > m(B)$, and hence the following equation holds:

$$\langle \Delta_r m \rangle = \mathbf{m}(B')\langle \cos\theta \rangle - \mathbf{m}(B)\langle \cos\theta \rangle > 0.$$

When the degree of transition is defined as $\beta = [B']/[B_T]$, where $[B_T]$ is the concentration of all binding sites (or all states $[B_T] = [B] + [B']$), we obtain:

$$K = \frac{[B']}{[B]} = \frac{\beta}{1-\beta} = K_0 \cdot e^X \tag{23}$$

$$\beta = \frac{K_0 e^X}{1 + K_0 e^X} \tag{24}$$

The weak-field approximation ($X \ll 1$, $e^X = 1 + X$) yields $\Delta K/K_0 = X$ and

$$\frac{\Delta\beta}{\beta_0} = \frac{\beta - \beta_0}{\beta_0} = X \cdot \frac{1}{1 + K_0} \tag{25}$$

Thus, the relative shifts in both K and in β are linearly dependent on X. If $K_0 \ll 1$, we obtain $\Delta\beta/\beta_0 = X$. Here too, the F-dependence of X yields $\Delta_r M$ or $\Delta_r \alpha$.

3.4 Ligand Binding

For the analysis, the ligand-binding equilibrium is written in the direction of dissociation

$$LB \rightleftharpoons L + B \tag{26}$$

The degree of ligand association is defined by

$$\beta = \frac{[LB]}{[B_T]} = \frac{[L]}{[L] + K} \tag{27}$$

where K is given by the total concentrations $[L_T]$ and $[B_T]$ of ligand and binding site, respectively:

$$K = \frac{[L][B]}{[LB]} = ([L_T] - \beta[B_T]) \cdot \frac{1-\beta}{\beta} \tag{28}$$

Using $(1-a)^{1/2} \approx 1 - \frac{1}{2}a$ if $a < 1$, we obtain

$$\beta \approx \frac{[L_T]}{[L_T] + [B_T] + K} \tag{29}$$

The field-induced relative change $\Delta\beta/\beta_0$ then is

$$\frac{\Delta\beta}{\beta_0} \approx \frac{K_0\left(1 - e^X\right)}{[L_T] + [B_T] + K_0 e^X} \tag{30}$$

The weak-field approximation ($X \ll 1$) yields

$$\frac{\Delta\beta}{\beta_0} \approx -X \cdot \left[1 + \frac{([L_T] + [B_T])}{K_0}\right]^{-1} \tag{31}$$

Here too, the relative change in β is proportional to X. Note that for Eq. (26), we have $\langle \Delta_r m \rangle = (\langle m_L \rangle + \langle m_B \rangle) - \langle m_{LB} \rangle$.

3.5 Dipole-pairing

Frequently, the pair association of dipolar species (B) is found to involve major dipole moment changes $\langle \Delta_r m \rangle$ if the pairing

reduces the dipole moment to zero by antiparallel alignment. Therefore, $\langle \Delta_r m \rangle$ for a dimerization reaction of the type

$$BB \rightleftharpoons 2B \tag{32}$$

is given by $\langle \Delta_r m \rangle = 2\langle mB \rangle - \langle m\ BB \rangle = 2\langle mB \rangle$. For this case, we have

$$\beta = \frac{2[B \cdot B]}{[B_T]} = \frac{[B]}{[B] + K/2} \tag{33}$$

$$K = \frac{[B]^2}{[BB]} = \frac{(1-\beta)^2}{\beta}[B_T] \tag{34}$$

Hence,

$$\beta \approx \frac{[B_T]}{[B]_T + K/2}. \tag{35}$$

The weak-field approximation reads as follows:

$$\frac{\Delta\beta}{\beta_0} \approx -X\left(1 + 2\frac{[B_T]}{K_0}\right)^{-1} \tag{36}$$

describing the decrease in the pair complex with increasing field strength.

4 Rate Constants

When time courses of field-induced processes can be resolved, the time constants or experimental rate coefficients can be analyzed in terms of individual rate constants. Generally, when equilibrium constants are field-dependent, the two rate constants must depend differently on the field strength.

4.1 Dipole Transitions

The equilibrium constant for the general Eq. (10) is given by

$$K = \frac{\Pi_p c_p{}^{\nu_p}}{\Pi_r c_r{}^{|\nu_r|}} = \frac{k_r}{k_p} \quad \text{and} \quad K^{\ominus} = \frac{k_r{}^{\ominus}}{k_p{}^{\ominus}} \tag{37}$$

The reaction moment $\Delta_r m$ may be related to the kinetic activation moments M_a by

$$\Delta_r M^{\ominus} = M_{a,r} - M_{a,p} \tag{38}$$

Similarly, the field effect exponents are connected by

$$X = X_{a,r} - X_{a,p}^{-} \tag{39}$$

Analogous to $K^{\ominus} = K_0{}^{\ominus}\, e^X$, we obtain

$$k^{\ominus} = k_0{}^{\ominus} \cdot e^{X_a} \tag{40}$$

where the field effect factor is defined by

$$x_a = \frac{G_a}{RT} = \frac{\int M_a\, dF}{RT} = \ln\left(\frac{k^{\ominus}}{k_0{}^{\ominus}}\right) \tag{41}$$

The weak-field approximation ($x_a \ll 1$) yields $k^{\ominus} = k_0{}^{\ominus}(1 + x_a)$.

The field-induced relative change of k' is generally given by

$$\frac{\Delta k^{\ominus}}{k_0{}^{\ominus}} = e^{x_a} - 1 \tag{42}$$

Hence, the weak-field approximation reads

$$\frac{\Delta k^{\ominus}}{k_0{}^{\ominus}} = x_a \tag{43}$$

Like the field-induced small shifts in K and β, the relative change in the rate coefficients of dipolar equilibria are linearly dependent on x_a; frequently x_a is proportional to F^2.

If the experimental rate coefficient k is analyzed in terms of the Eyring transition state theory, we have $k = k^{\neq} K^{\neq} \kappa$, where κ is an adjustable parameter and $k^{\neq} = k_B T/h$ is a universal decay constant and $K^{\neq}$ and $\Delta_r M^{\neq}$ are the equilibrium constant and the reaction moment respectively of the assumed rapid preequilibrium. For this case, M_a associated with k is given by $M_a = \Delta_r M^{\neq}$. Since the Eyring approach is not applicable in condensed

matter, any such approach is purely formalistic.

On the other hand, we may equally well assume that the activation moments M_a are determined by the actual dipole moments ($\mathbf{m}$) of the educts and products. Hence,

$$M_{a,p} = N_A \cdot \Sigma_p \nu_p |\mathbf{m}_p| \langle \cos\theta \rangle$$
$$M_{a,r} = N_A \cdot \Sigma_r |\nu_r| |\mathbf{m}_r| \langle \cos\theta \rangle \quad (44)$$

Nonidealities are covered by the activity coefficients. With $K^{\ddagger} = k_r^{\ddagger}/k_p^{\ddagger}$, we formally have

$$k_r = k_r^{\ddagger} \cdot Y_r, \text{ and } k_p = k_p^{\ddagger} \cdot Y_p \quad (45)$$

The field-effect exponent of Eq. (41) is specified analogous to Eq. (18) as

$$x_a = \ln\left(\frac{k^{\ddagger}}{k_0^{\ddagger}}\right) = \ln\left(\frac{k}{k_0}\right) + \ln\left(\frac{Y_E}{Y}\right) \quad (46)$$

where Y refers to the respective activity coefficient term at E and in the absence of E, respectively.

If, by chemical relaxation spectrometry, the rate coefficients can be determined separately, the *activation dipole moments* of the interacting reaction partners can be estimated from the field dependence of the k values.

4.2 Ion Equilibrium

According to Onsager, it is only the dissociation rate constant k_{dis} of the separation of an ion pair that is primarily affected by an electric field $\mathbf{E}$:

$$L^{z_+} \cdot B^{z_-} \underset{k_{ass}}{\overset{k_{dis}}{\rightleftharpoons}} L^{z_+} + B^{z_-} \quad (47)$$

In the linear range, Onsager's theory of very diluted weak electrolytes (z_+/z_-) yields

$$\left(\frac{\partial \ln |k_{dis}^{\Theta}|}{\partial |\mathbf{E}_0|}\right)_{p,T} = \frac{(z_L u_L - z_B u_B)|z_L \cdot z_B| e_0^3}{(u_L + u_B) 8\pi\varepsilon_0 \varepsilon (k_B T)^2} = \gamma \quad (48)$$

where u_L and u_B are the electric mobilities of the separated ionic species; see Eq. (1). For symmetric electrolytes where $z_L = -z_B = |z|$, Eq. (48) is reduced to $\gamma = |z^3| e_0^3/[8\pi\varepsilon_0\varepsilon(k_B T)^2]$.

If the actual dipole moment of the ion pair $L \cdot B$ is not zero (such as the electric dipole moments of the $MgSO_4$ ion pairs) and if the reaction occurs at higher salt concentrations, we use the expressions

$$k_{dis} = k_{dis}^{\ddagger} \cdot Y_d \quad \text{and} \quad k_{ass} = k_{ass}^{\ddagger} Y_a \quad (49)$$

Integration of Eq. (48) yields

$$X_{a,r} = \ln\left(\frac{k_{dis}^{\ddagger}}{k_{d,0}^{\ddagger}}\right) = \gamma \cdot |\mathbf{E}_0| \quad (50)$$

and

$$\ln\left(\frac{k_{dis}}{k_{d,0}}\right) = \gamma |\mathbf{E}_0| - \ln\left(\frac{Y_d^{(E)}}{Y_d}\right) \quad (51)$$

When the dipole moment of the ion pair $L \cdot B$ in the simple reaction $LB = L^+ + B^-$ is dependent on the ionic strength because of screening by ionic "half clouds", we may analyse the k_a data according to, $\ln|k_{ass}| = \ln|k_{a,0}| + X_{a,p} - \ln(Y_a(F)/Y_a)$ with $Y_a(F) = Y_a$ at F. Applying

$$\frac{d \ln |k_{ass}|}{dF} = \frac{M_{a,p}}{RT}, \quad (52)$$

$M_{a,p}$ can be determined. Consistency may be checked using Eq. (38), connecting the field dependencies of K, k_r, and k_p, respectively.

Often, it is useful to define a degree of dissociation by $\alpha = [B]/[B_T]$. If the reaction is symmetric ($[L_T] = [B_T]$), we have:

$$K = \frac{\alpha^2}{1-\alpha}[B_T] \tag{53}$$

The small-field approximation of the relative change in the degree of dissociation is derived analogous to Eq. (31), resulting in:

$$\frac{\Delta\alpha}{\alpha_0} \approx x\left(3 + \frac{K_0}{[B_T]} + \frac{2[B_T]}{K_0}\right)^{-1} \tag{54}$$

Formally, the field dependence of K can be calculated from that of k_{dis} and k_{ass} and formally described as $\Delta K/K_0 = x_{a,r} - x_{a,p}$. See Eqs. (21 and 43).

4.3 Activity Coefficients

Thermodynamic activity coefficients and the theoretical approaches are defined for equilibrium conditions only. The **E** and **B** field effects on electrolytes, however, represent nonequilibrium situations. Under stationary conditions, we may consider the field-induced distortion of screening ionic atmospheres and other local changes in the ion concentration as actual steady state changes αc_i. Purely formally, we introduce the relative change $\Delta\gamma_i/\gamma_i$ in the (theoretically individual) activity coefficient γ_i of the ion (or ionic group) of type i.

The Debye–Hückel approximation for a moderately diluted solution is purely formally given by

$$\ln\gamma_i = \frac{-z_i^2 A' \cdot I_c^{1/2}}{\left(1 + B' I_c^{1/2}\right)} \tag{55}$$

where A' is a (temperature-dependent) constant, B' a parameter containing the encounter distance and

$$I_c = 0.5 \sum_i z_i^2 c_i \tag{56}$$

is the ionic strength (including all ions i).

For elementary reactions between ionic partners like those in Eq. (47), we have in general

$$\ln Y = \frac{2z_+z_-A' \cdot I_c^{1/2}}{\left(1 + B' I_c^{1/2}\right)} \tag{57}$$

The small-field approximation is specified as

$$\ln\left(\frac{Y_E}{Y}\right) = \ln\left(1 + \frac{\Delta Y}{Y}\right) = \frac{\Delta Y}{Y} \tag{58}$$

Because $\ln Y$ is proportional to $I_c^{1/2}$, we obtain for $\Delta I_c = I_c$,

$$\ln\left(\frac{Y_E}{Y}\right) = \left(1 + \frac{\Delta I_c}{I_c}\right)^{1/2} \approx 1 + \frac{\Delta I_c}{2I_c} \tag{59}$$

where $\Delta I_c = \sum_i z_i^2 \Delta c_i/2$ covers the field-induced local changes in the screening ions, which may be measured as a Wien effect (conductivity increase due to F).

Similarly, field effects on rate coefficients may also involve changes in the respective Y coefficients. The concentration changes Δc_i entering into ΔI_c and thus into the relative changes of K, α, and k may also contain contributions from the ionic currents induced by the fields.

4.4 Ion Flows

The electrodiffusive flows of the cellular ions Na^+, K^+, Ca^{2+}, and Cl^- through electrochemically gated channels of membranes are dependent on the local ion concentrations c_i and on induced local field changes.

The flow density vector $\mathbf{i}_i$ (mol/sm^2) of the ion type *i* is given by the Nernst–Planck equation for the flow direction **r**

$$|\mathbf{i}_i| = -D_i \left(\frac{\partial c_i}{\partial r} + \frac{z_i c_i F}{RT} \frac{\partial \varphi}{\partial r} \right) \qquad (60)$$

where D_i is the diffusion coefficient, $F = e_0 N_A$ is the Faraday constant, and $\partial\varphi/\partial r = -E$ is the electric potential gradient at site **r**.

The current density vector is then

$$\mathbf{J} = F \sum_i z_i i_i = \kappa \cdot \mathbf{E} \qquad (61)$$

The relative changes $\Delta J/J_0$ are caused by a change in **E** or in the ion concentrations, or in both. Finally, the change in the transmembrane voltage (or membrane field E_m) may directly change the conformation of a membrane transport–gating protein or of a membrane-associated enzyme.

5 Amplification Mechanisms

Natural electric and magnetic field reception operates at small external field strengths. The molecular reaction moments of the interacting, directly "field-receiving" charges and electric and magnetic dipoles are relatively small ($\Delta_r G_{(F)}/RT \ll 1$). Obviously, there must be amplification mechanisms for both, enhancement of the field (resulting in larger internal fields and ion flows) on the one hand and enhancement of local chemical reactivities on the other hand.

If a chemical amplification cascade enhances a substrate signal from the level $[S_0]$ to the product level $[P_0]$, a relevant external field effect must already change the substrate concentration by a significant amount $\Delta[S]$. The cascade that may contain a network of nonlinear couplings of reaction and diffusion flows will process the relative change $\Delta[S]/[S_0]$ to the output $\Delta[P]/[P_0] = \Delta[S]/[S_0]$. Hence, the relative amplification will not change unless there are significant field effects on the local elements of the cascade itself.

The distribution of electric organs in rows of larger skin areas of electric fish or the strings of magnetosomes in magnetobacteria and other species suggest that geometrical factors are essential in adding up signals from many inputs. The feedback system of many innervated sensor and emitter organs even enables the organisms to differentiate between inanimate ohmic-dissipative matter and capacitively storing biomatter of cells (organelles and lipid/protein membranes). Besides the special structure, form, and shapes of the sensory receptors, there are some general principles for molecular enhancement and amplification processes.

5.1 Field Enhancement

At interfaces of matter with different relative permittivity (dielectric constant ε) such as plasma membranes, applied fields cause ion accumulations, which ionically charge up cells, organelles, and larger macromolecules (like condensers).

If a macromolecule is considered as a spherical microdroplet continuum ε_2 in aqueous solution ($\varepsilon_1 = 80$ at 20 °C), the induced field across the center is given by:

$$\mathbf{E}_{ind} = \frac{3\varepsilon_1}{2\varepsilon_1 + \varepsilon_2} \mathbf{E}_0 \qquad (62)$$

Since usually $\varepsilon_2 < \varepsilon_1$, we obtain $\mathbf{E}_{ind} \leq 1.5\ \mathbf{E}_0$. Note that Eq. (62) is applicable if the conductivity term in the complex dielectric permittivity is negligible.

A much larger field amplification occurs when single cells or organelles

or even clusters of cells and tissue pieces are exposed to external fields ($\mathbf{E}_0$). The ionic interfacial polarization causes induced electric potential differences $\Delta\varphi_{\text{ind}}$ that, in the time range of microseconds to milliseconds, because of the conducting cell interior, occur only across the plasma membranes (see Fig. 3a). For spherical cells, the induced transmembrane potential difference in the direction x of the external field $\mathbf{E}_0$, in the stationary state of ionic interfacial polarization – polarization time $\tau_{\text{pol}} \ll \text{t}$, such that $\Delta\varphi(\text{t}) = \Delta\varphi\left(1 - \text{e}^{-\text{t}/\tau}\right)$ – is given by $\Delta\varphi_{\text{ind}} = -1.5\ \text{f}(\lambda)(\mathbf{E}\cdot\mathbf{a})$. Hence,

$$\Delta\varphi_{\text{ind}}(\theta) = -\frac{3}{2}\,\text{a f}\,(\lambda)|\mathbf{E}_0||\cos\theta| \quad (63)$$

where θ is the angle between the membrane position of the tip of the radius vector $\mathbf{a}$ and the positive x-axis ($\mathbf{E}_0$) through the sphere midpoint (Fig. 3). The conductivity factor $\text{f}(\lambda) \leq 1$ depends on the ratio a/d_{m} and on the conductivities λ_{m}, λ_{in}, and λ_{ex}

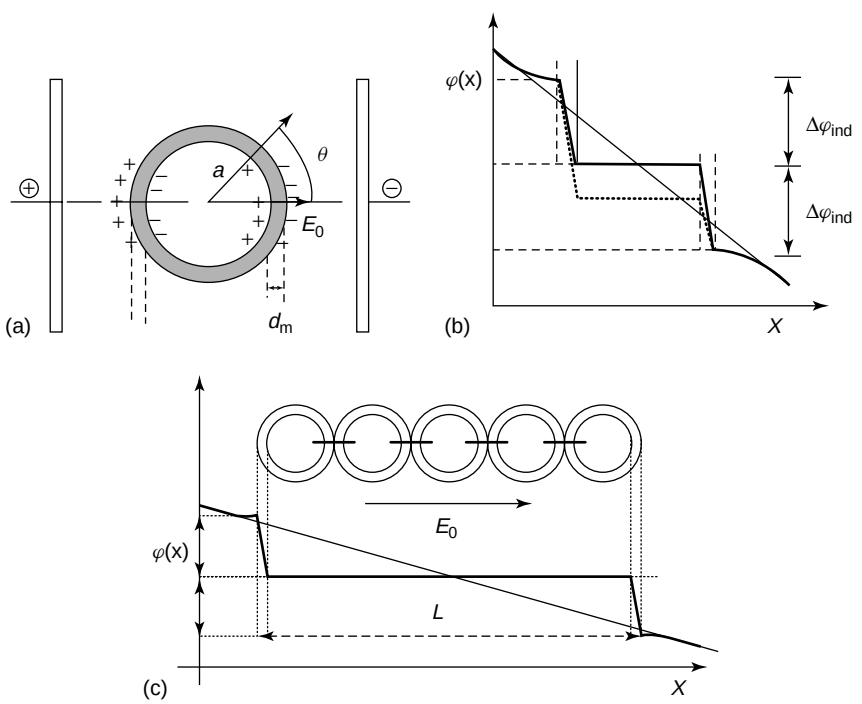

Fig. 3 Field amplification by interfacial polarization through DC and low-frequency **E** fields. (a) Single-cell sphere of the radius vector **a**, and the membrane thickness d_{m}. The angle θ defines the membrane position of the tip of the radius vector **a** relative to the direction of $\mathbf{E}_0$ (x-direction). The dashed fat line refers to $\Delta\varphi_{\text{m}} = \Delta\varphi_{\text{ind}} + \Delta\varphi_{\text{nat}}$ (at $\theta = 180^\circ$, $\theta = 0^\circ$), where in this example, the resting membrane potential $\Delta\varphi_{\text{nat}} < 0$ and also $\Delta\varphi_{\text{ind}} < 0$ (Eq. 64). (b) Electric potential profile $\varphi(\text{x})$ in the direction x of the external field vector $\mathbf{E}_0$ through the pole caps and the midpoint of the sphere. The steep potential drop across the membranes (in the direction of $\mathbf{E}_0$) is given by $\Delta\varphi_{\text{ind}}(\theta) = -1.5\ E_0\text{a}\cdot|\cos\theta|$, if the conductivity of the membrane is very small compared to that of the cell inside and outside and $a/d_{\text{m}} \gg 1$. The field amplification factor at the pole caps is $\text{f}(a) = (E_{\text{ind}}/E_0) = 1.5a/d_{\text{m}}$. (c) Array of cells connected by conducting gap junctions. The field amplification factor is $\text{f}(L) = (E_{\text{ind}}/E_0) = 1.5L/d_{\text{m}}$.

of the membrane, the cell interior, and the outside medium respectively.

If $a/d_m \gg 1$, $\lambda_m \ll \lambda_{ex}$, λ_{in} holds and the $f(\lambda)$ factor is given by $f(\lambda) = 1 - \lambda_m(a/d_m)(\lambda_{ex} + \lambda_{in}/2/(\lambda_{in} \cdot \lambda_{ex})$. At the pole caps, that is, at $\theta = 0°$ (right side) and $\theta = 180°$ (left side in $\mathbf{E}_0$ direction, Fig. 3) hence $\cos\theta = \pm 1$, the induced transmembrane voltage is maximum, and correspondingly

$$|\mathbf{E}_{ind}| = -\frac{\Delta\varphi_{ind}}{d_m} = \frac{3}{2}\frac{a}{d_m}|\mathbf{E}_0| \qquad (64)$$

The amplification factor is defined by

$$f(a) = \frac{|E_{ind}|}{|E_0|} = \frac{3a}{2d} \qquad (65)$$

For a cell radius of $a = 10\,\mu m$ and the membrane thickness $d_m = 5$ nm, $f(a) = 3 \times 10^3$. Hence, interfacial polarization can very appreciably amplify weak external fields; this type of amplification mechanism is the basis for all the electroporation phenomena.

Interfacial polarization not only amplifies external fields. In addition, the induced potential difference is superposed on the natural potential difference of cell membranes, $\Delta\varphi_m \approx -100$ mV, where the electric potential of the outside is arbitrarily taken to be zero. Note that $\Delta\varphi_{nat}$ is independent of θ. The actual transmembrane potential in the direction of $\mathbf{E}_0$ is given by

$$\Delta\varphi = -\left[\frac{3}{2}af(\lambda)|\mathbf{E}_0| + \frac{\Delta\varphi_m}{\cos\theta}\right]|\cos\theta| \qquad (66)$$

Since $E_m(\theta) = -\Delta\varphi_m(\theta)/d_m = E_{ind}(\theta) + E_{nat}$, where $E_{nat} = -\Delta\varphi_{nat}|\cos\theta|/(d_m \cos\theta)$, it is seen that at the left pole cap, $\cos\theta = -1 E_m > E_{ind}$ and at the right pole cap, $(\cos\theta = 1) E_m < E_{ind}$. If $\Delta\varphi_{nat} < 0$, the left cap is hyperpolarized and the right cap is depolarized, and the membrane proteins and transport processes in the two pole caps are differently affected by $\mathbf{E}_0$. Therefore, a longer-lasting DC field may produce appreciably large asymmetric field effects.

At the pole caps of oriented bacteria of length L, the induced transmembrane potential difference is

$$\Delta\varphi_{ind} = -0.5L|\mathbf{E}_0| \qquad (67)$$

and the amplification factor is $f(L) = L/2d_m$. Here too, the field effect is asymmetric at the two pole caps.

Another important example of field amplification by geometrical factors consists of a series of cells connected by gap junctions. Such a string of cells in series resembles a single elongated cell. In an applied external field, the outer pole cap of the head cell and that of the tail cell experience the induced potential drop $\Delta\varphi_{ind} = -0.5L|\mathbf{E}_0|$, where L is the length of the string (Fig. 3b). The amplification factor $f(L)$ increases with L and may become very large. Here too, the field effect is asymmetric with respect to the natural membrane voltage in the head and the tail cell.

5.2 Cooperativity and Phase Transition

In systems with domain structures, the molecules of a domain are cooperatively coupled. A domain changing state or, more specifically, molecular orientations as a whole, is described by

$$B_n \rightleftharpoons B'_n \qquad (68)$$

where n is the number of cooperatively coupled molecules (domain size) and B′ represents the (orientational) state favored in the presence of an applied field because of higher moment.

In the simplest case, the collective transition of a domain is described by the

field-effect exponent

$$X_n = \frac{\Delta_r G_{n(F)}}{RT} = \frac{\int \Delta_r M_n \mathrm{d}F}{RT} \tag{69}$$

where $X_n = nX$ and $\Delta_r M_n = n\Delta_r M$.

For an ensemble of independent domains of different sizes, we may formalize a distribution constant

$$K(\bar{n})^{\ddagger} = K_0^{\ddagger} \cdot \exp(X_n) \tag{70}$$

where $\bar{n}$ is the mean domain size, representing a kind of cooperative amplification factor for the reaction moment $\Delta_r M$.

If $\bar{n} \gg 1$, very steep phase transitions can occur in a very narrow field strength range. Thus, even a weak external field may have a large effect (see Fig. 2).

If a reaction domain with the polarization volume V_n can be specified, the orientational rearrangement within a domain may be subjected to a continuum approach where the polarization P has different values in the two states and B'_n. If $\Delta_r P = P(B'_n) - P(B_n) = N_A \Delta_r p$ is the reaction polarization, straightforward electrodynamics yields

$$\Delta_r M_n = V_n \cdot \Delta_r p \cdot N_A. \tag{71}$$

Specifically, the electric and magnetic polarization differences are given by

$$\Delta P_{el} = \varepsilon_0 \Delta\chi_{el} |\mathbf{E}| = \varepsilon_0 \cdot \Delta\varepsilon |\mathbf{E}|$$
$$\Delta P_{ma} = \mu_0^{-1} \Delta\chi_{ma} |\mathbf{B}| \tag{72}$$

where the $\Delta\chi_{el}$ and $\Delta\chi_{ma}$ are the susceptibility differences, respectively. If the domain refers to the formation of an aqueous pore in a lipid bilayer membrane, $\Delta\chi_{el} = \Delta\varepsilon = \varepsilon_m - \varepsilon_w$ is used (because $\chi_{el} = \varepsilon - 1$), where $\varepsilon_m = \varepsilon(\text{lipid}) \approx 2$ and $\varepsilon_w \approx 80$ are the dielectric constants of the membrane and the pore water respectively.

The field-effect exponents at constant domain volume are derived from Eq. (71):

$$\chi_{el} = \frac{\varepsilon_0 V_n \cdot \Delta\chi_{el}}{2k_B T} g |\mathbf{E}_0|^2 \tag{73}$$

where g is a shape factor [2] and

$$x_{ma} = \frac{V_n \cdot \Delta\chi_{ma}}{2\mu_0 k_B T} |\mathbf{B}|^2 \tag{74}$$

In an ensemble of domains, V_n represents the size average $V_{\bar{n}}$.

More specifically, a domain orientation function ϕ^{OR} may be used, which is different for permanent ($\mathbf{m}_p$) and induced ($\mathbf{m}_{ind}$) dipole mechanisms [5]. For uniaxial permanent dipoles, we have

$$\phi^{OR}_{perm} = 1 - \frac{3}{x_p}\left(\coth x_p - \frac{1}{x_p}\right) \tag{75}$$

where $x = |\mathbf{m}_p| F_{dir}/(k_B T)$. For $F = E$, E_{dir} is the Onsager directing field [3]. If $x_p = 1$, the weak-field approximation yields $\phi^{OR}_p = x_p^2/15$ as the first term of a power series.

The orientation function for induced electric dipoles $\mathbf{m}_{ind} = \alpha \cdot \mathbf{E}_{int}$, where α is the polarizability tensor and $\mathbf{E}_{int}$ is the internal field, is given by

$$\phi^{OR}_{ind} = \frac{3}{4}\left(\frac{e^{x_{ind}}}{x_{ind}^{1/2}\int dy\, e^{y^2}}\right) - \frac{1}{2} \tag{76}$$

where $x_{ind} = \alpha \cdot g_i |\mathbf{E}_0|^2/(2k_B T)$, and $g_i = |\mathbf{E}_{int}|/|\mathbf{E}_0|$ is the shape factor.

If $x_{ind} \ll 1$, the weak-field expression yields

$$\phi^{OR}_{ind} = \frac{2x_{ind}}{15} \propto |\mathbf{E}_0|^2.$$

orientation function are experimentally accessible by electrooptical relaxation techniques.

In the case of domain systems, we have

$$x_p(n) = \frac{n|\mathbf{m}_p|F_{dir}}{k_B T} \tag{77}$$

and

$$x_{ind}(n) = \frac{n \cdot \alpha \cdot g_i |\mathbf{E}_0|^2}{2k_B T}. \tag{78}$$

If $n \gg 1$, the amplification of small molecular field effects may be appreciably large (see Fig. 2).

5.3 Amplification at High Basis Fields

A particular amplification mode resides directly in the field dependence of equilibrium constants K. This enhancement mode appears to be especially important for all membrane processes occurring in the strong natural transmembrane fields.

From Eqs. (13 and 17), it is seen that the steepness of the $K(X)$ relationship, given by

$$\left(\frac{\partial K}{\partial F}\right)_{p,T} = \frac{K(E)\Delta_r M^{\ddagger}}{RT}, \tag{79}$$

depends not only on $\Delta_r M^{\ddagger}$ but also on $K(E)$ itself.

If an external field causes an induced field $|\mathbf{E}_{ind}| = \Delta E_0$ at the level E_{nat} of the resting membrane potential difference ($\Delta\varphi_{nat} = -E_{nat}d_m$), the relative change $\Delta K/K(E_{nat})$ is larger than $\Delta K'/K_0$ at $E_{nat} = 0$ for the same ΔE_0 value (Fig. 4). The inequality $\Delta K/K(E_{nat}) > \Delta K'/K_0$ is due to the field dependence of the reaction moment $\Delta_r M(\propto F)$. Therefore, in terms of the field-effect exponents, we have for the different integration boundaries

$$X(E_{nat} \longrightarrow E_{nat} + \Delta E_0) > X(E_{nat} \longrightarrow E_{nat} - \Delta E_0).$$

If the external force is an oscillatory (AC) field, say, $\mathbf{E}_0(f) = \hat{\mathbf{E}}_0 \sin(2\pi ft)$ where $\hat{\mathbf{E}}_0$ is the peak amplitude, the enforced oscillatory change in the equilibrium constant of a reaction is asymmetric: $\Delta K_+/K(E_{nat}) \neq \Delta K_-/K(E_{nat})$; see Fig. 4. For instance, the induced concentration changes Δc_+ are larger for one AC field direction ($+\Delta E_0$) compared to the Δc_- values of the other direction ($-\Delta E_0$). In a similar way, the change in the forward rate constant is different from that of the reverse rate constant.

5.4 Comparison of AC/DC Effects

In DC fields, freely mobile charges and dipoles are translationally or rotationally net-displaced. In AC fields, however, the motions are oscillatory around a mean position. If the frequency of the AC field is too high, inertia and viscosity may prevent the species from following the variation of the external fields. On the other hand, fields acting on an ensemble of charges and dipoles may enforce and synchronize the individually random motions into a *concerted collective motion* of all the interaction partners. Such an externally enforced cooperation may bias in one direction an otherwise random fluctuation and produce the small net concentration changes involved in weak-field reception.

The interpretation of the experimental data on weak-field effects is still very controversial. The formalism outlined here may help in two ways: (1) in the design of experiments that use both species and sample size as variables, and (2) in the analysis of data in terms of thermodynamic and kinetic parameters consistent with the first principles of chemical reactivity in electric and magnetic fields.

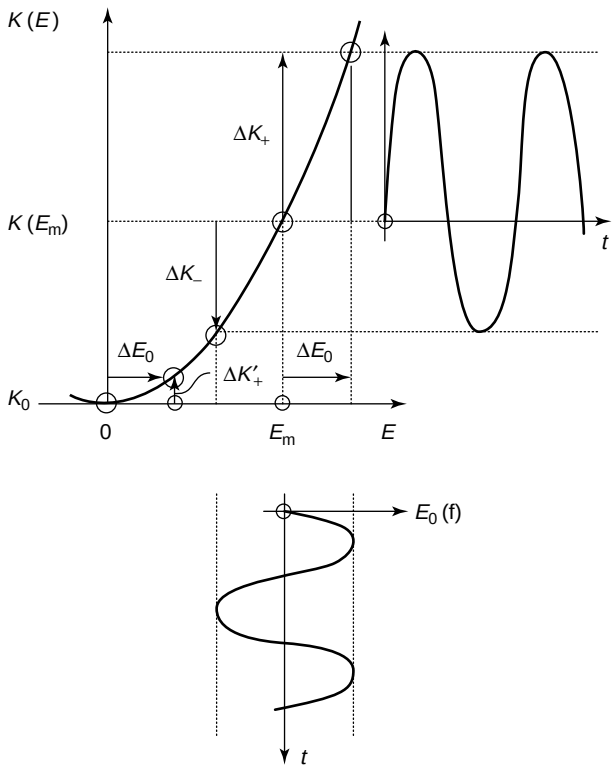

Fig. 4 Amplification at high field and asymmetric field effect for oscillatory external fields (AC). Application of an external field, inducing $E_{\text{ind}} = \Delta|\mathbf{E}_0|$, at $E = 0$, causes only a small change $\Delta K'_+$. If applied at $E = E_{\text{nat}}$ (e.g. at the resting potential $\Delta\varphi_{\text{nat}} \approx 0.1\text{V}$), the same external field causes a much larger change ΔK_+, because $\Delta K/\Delta E = K(E)\Delta_r M/(RT)$ depends on $K(E)$. The asymmetry of the $K(E)$ dependence leads to an asymmetric amplification of K: $\Delta K_+ > \Delta K_-$, if the symmetric external field $\mathbf{E}_{0(\text{f})} = \hat{\mathbf{E}}_0 \cdot \sin \cdot (2\pi \text{f} \cdot \text{t})$ operates at E_{nat}. At $E = 0$, $\Delta K_+ = \Delta K_-$. In a similar way, the rate constant k is asymmetrically dependent on $\mathbf{E}_0(\text{f})$.

There is continued public concern on possible health risks of weak electromagnetic fields, that is, of field intensities that are very small compared to those of the natural geofields.

As indicated here, the problem is not that there are always finite effects that in principle, may either be benefical or bear a risk. The still unresolved question is *whether and how*, if at all, the small and insignificant local changes in the concentrations of the biochemical reaction partners may be amplified such that they emerge out of the natural thermal fluctuations. Despite the various experimental and theoretical activities aiming at knowledge to identify and to understand such mechanisms, no general conclusive and convincing answers can yet be given for the various claims of finite small field-effects.

See also Bacterial Growth and Division.

Bibliography

Böttcher, C.J.F. (1973) *Theory of Electric Polarization*, Vol. 1, Dielectrics in Static Fields, Elsevier, Amsterdam.

Eigen, M., DeMaeyer, L. (1963) Relaxation methods, *Tech. Org. Chem.* **8**(2), 895–1054.

Fredericq, E., Houssier, C. (1973) *Electric Dichroism and Electric Birefringence*, Clarendon Press, Oxford.

Neumann, E. (1986) Chemical electric field effects in biological macromolecules, *Prog. Biophys. Mol. Biol.* **47**, 197–231.

Neumann, E., Sowers, A.E., Jordan, C.A. (Eds.) (1989) *Electroporation and Electrofusion in Cell Biology*, Plenum Press, New York.

Neumann, E. (2000) Digression on chemical electromagnetic field effects in membrane signal transduction–the experimental paradigm of the acetylcholine receptor, *Bioelectrochemistry* **52**, 43–49.

Polk, C., Postow, E. (Eds.) (1986) *Handbook of Biological Effects of Electromagnetic Fields*, CRC Press, Boca Raton, FL.

Electron Microscopy in Cell Biology

Gareth Griffiths
European Molecular Biology Laboratory, Heidelberg, Germany

1 Introduction 24
1.1 Why Do We Need EM in Cell Biology? 24
1.1.1 The Strategy 25

2 Specimen Preparation 27

3 Ambient Temperature EM Methods 27
3.1 Negative Staining of Particles 27
3.2 Positive-negative Staining Approaches 29
3.3 Classical Resin Embedding 31
3.4 Chemical Fixation 32
3.5 Postfixation 33
3.6 Dehydration and Embedding 33
3.7 Methacrylate-based Embedding 34
3.8 Plastic Sectioning 36
3.9 The Tokuyasu Sucrose Embedding – Cryosection Method 37
3.10 On-section Immunogold-labeling Procedure 40
3.10.1 Step 2 – Antibody Labeling 42
3.10.2 Step 3 – Gold 42
3.11 Preembedding Labeling Methods 43
3.12 Immunohistochemical Approaches using Horseradish Peroxidase and Cytochemistry 45
3.12.1 The Use of Microwave Technology for Specimen Preparation 45

4 Cryo-EM Approaches 46
4.1 Vitrification 46
4.2 The Bare-grid Method for Particulate Specimens 46
4.3 Vitrification of Larger Material 47
4.4 The Cellulose Capillary Tube 47

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

4.5 The Fine-needle Biopsy 48
4.6 Hydrated Cryosectioning 49
4.7 Freeze-substitution 49
4.8 Freeze-fracture and Replica Methods 51
4.9 Simple Shadowing Methods 51
4.10 The Kleinschmidt and other EM Methods for Nucleic Acids 52
4.11 EM Autoradiography 53

5 EM-Visualization at Ambient Temperatures 53

6 EM at Cold Temperatures 57
6.1 EM Tomography 58
6.2 Low- and High-resolution SEM 60
6.3 Critical Point Drying 62
6.4 Freeze-drying and Cryo-SEM 64
6.5 Coating Techniques for Cryo-SEM 67
6.6 Scanning Transmission EM 67
6.7 X ray Elemental Microanalysis by EM 67
6.8 Energy-filtering Transmission Electron Microscopy (EFTEM) 70

7 Stereology 70
7.1 Relative Area and Volume 71
7.2 Relative Profile Length and Surface Area 71
7.3 Absolute Volume 72
7.4 Surfaces and Volumes in Practice 72
7.5 Correlative Light and Electron Microscopy 73
7.6 Interpretation in EM 74
7.7 Which Technique for Which Question? 74
7.8 Final Comment 76

Acknowledgments 76

Bibliography 77
Books and Reviews 77
Primary Literature 77

Keywords

Transmission Electron Microscopy (TEM)
Electrons passing through the specimen form the image. Transmission of electrons through the specimen depends on the accelerating voltage, routinely 80 to 120 KV for specimens thinner than 300 nm, but for thicker specimens, or for higher resolution voltages up to 1000 K can be used.

EM grid
A 3-mm diameter, 0.01-mm (10–15 μm) thick disc of metal (usually copper) with regular perforations. The specimen is placed on the grid and those parts of the specimen that lie across the perforations are visualized in the EM.

Scanning Electron Microscopy (SEM)
This approach is used to visualize the specimen surface. An electron probe scans the surface and the secondary electrons are recorded in relation to the moving point of the probe.

Specimen Preparation
The foundation of all EM. These are highly empirical methods that have been developed for all EM approaches, which allow the specimens to be imaged successfully. For TEM, the specimen must either be particulate, generally thinner than 500 nm, or it must be sectioned to that thickness, or thinner, after embedding in plastic, or after cryoimmobilization.

Vitrification
The process whereby liquid water is cooled so rapidly that the water molecules solidify without crystallization to form amorphous or vitreous ice. A vitrified specimen is considered to be cryoimmobilized.

Freeze Substitution
Vitrified specimens are infiltrated at low temperatures with solvents, usually with heavy metal stains such as osmium tetroxide or uranyl acetate. Following this low temperature stabilization, the specimen is infiltrated with resin at low temperature followed by ultraviolet light polymerization below −20 °C, or the specimen temperature is raised to room temperature prior to room temperature resin embedding.

Freeze-Fracture/Replica
A vitrified specimen is planarly split and coated with a thin layer of metal (such as platinum, tungsten, or tantalum). The biological material is digested away and the replica is placed on an EM grid or viewed in the frozen- hydrated state in a SEM. A useful modification, freeze-etching, involves subliming a surface layer of water molecules to reveal details of some internal structures before shadowing. For particles and macromolecular complexes, metal shadowing can be used to form a replica without the freezing step. Glycerol spraying is a related technique where the sample is sprayed into a glycerol suspension and then rotary-shadowed and viewed by TEM.

Immunogold Labeling
Visualization of antigens on sections, or on the surface of isolated particles using (primary) antibodies. These are detected generally using an additional reaction containing gold particles that have bound (secondary) antibodies, or protein-A, that recognize the primary antibodies.

Tomography
The procedure for collecting multiple tilt images of a set of particles, or structures in (mostly) thick sections or isolated particles and collating the information computationally to provide a three-dimensional model of the structure.

Image Analysis
Computing methods used to improve the signal-to-noise ratio of EM images. They are based on averaging the information from two-dimensional data sets and can reveal the three-dimensional structure.

Stereology
A set of sampling tools and geometrical probability–based principles that allow one to estimate relative and absolute three-dimensional structural quantities (such as volume, surface, length and number), and spatial relationships, from sectional images.

Most of the organelles discovered after 1945 were first seen by electron microscopy (EM) in thin sections of resin-embedded cells. Since that time, the technical advances in EM instrumentation and specimen preparation have been enormous, a consequence of physicists, engineers, and biologists working together. These methods now allow cell biologists to visualize unperturbed cell structures over a range of dimensions, from the atomic resolution in the case of single proteins or macromolecular complexes to the range from 2 nm and up in the case of cellular organelles. This level of resolution is far beyond what can be seen by light microscopy. This review will cover the essentials of a range of methods for scientists investigating cell biology using EM. The major emphasis will be on resolution levels poorer than 1 to 3 nm. In all these methods, the most crucial parameter is specimen preparation and, without doubt, the ''gold- standard'' reference technique, which involves vitrification by rapid freezing, in conjunction with cryo-EM.

1 Introduction

1.1 Why Do We Need EM in Cell Biology?

The main goal of this chapter is to convince the reader, especially cell biologists, that electron microscopy (EM) offers them an incredibly powerful set of approaches for analyzing the ultrastructure of cells and their components. Moreover, these methods have been developed to cover a vast scale of detail, ranging from the atomic level to the level of whole organisms. While high-resolution cryo-EM is thriving, the use of lower resolution EM has dropped considerably in recent times in cutting-edge cell biology research. There is no doubt that this downward trend is a direct effect of the availability of increasingly sophisticated state-of-the-art light microscopy (LM) approaches

that can, for example, detect individual fluorescent molecules under ideal conditions. Recently, even the long-considered lower limit to LM resolution (0.2 μm) has been reduced for detecting fluorescent molecules to below 0.1 μm. Such developments, while clearly very important, may even deceive more cell biologists into thinking that LM approaches can now do everything that was earlier considered in the realm of EM.

There is, however, a fundamental flaw to this argument. Even if one can see a single fluorescently labeled molecule in the cell, there is still the fundamental limitation with LM in *relating that signal to the defined structures in the cell*. An EM-thin section can reveal enormous detail of defined structures, such as membranes, filaments, or ribosomes; when this section is labeled with immunogold, the gold particles can be assigned to defined structures with high precision (within 8 to 20 nm). However, at the LM level, the fluorescence signal is mostly seen only as a nondescript blob. Often, one compares the immunolabeled signal from one protein in one color with respect to a compartment reference marker seen in a separate color. However, for such a reference marker to be useful, one needs to have shown first by EM that it is indeed restricted to the structure of interest. Figure 1(a) shows an example of immunofluorescence labeling of the Golgi complex, while Fig. 1(b) reveals how much more information becomes available when the same specimen is visualized at the EM level. Both methods provide important information, but at different, complementary scales.

State-of-the-art LM is nevertheless, quite justifiably at the center stage of cell biology. In contrast to EM analysis, which always presents snap-shots of processes, video analysis by LM can allow one to see dynamic events in real time within cells. One of the "take-home" messages of this review is to point out that for modern cell biology studies one needs *both* LM, to see an overview of the "forest," and EM to see the details of the "trees." I will try to demonstrate that state-of-the-art EM methods should be standard technology in all laboratories involved in molecular cell biology research. It should also be pointed out that the expertise needed to execute these methods at the highest level is rapidly dwindling, as fewer specialist groups are available to keep these methods alive. I hope this chapter will also contribute to an awareness that this situation can, and must, be reversed.

1.1.1 The Strategy

The spectrum of EM techniques of interest to cell biologists is quite broad. For each method, one can easily write a whole book, and many books are indeed available. Several of these methods are technically quite demanding. Given the space limitation and the need to cover almost all techniques, the approach I have taken is to explain the basic concepts behind each approach so that a total beginner could understand how each technique operates and what it is good for. For the most commonly used methods, photographs, and schematic diagrams are provided to show those details that are important for the method to work.

The strategy I have followed is to start with live cells or organisms, or isolated biological material in an aqueous solution and describe the two fundamentally different ways of preparing these specimens for transmission – EM-conventional chemical methods, and the cryo-based approaches. Some of these conventional methods, as well as one partly cryo-approach, can be used for the detection of antigens using

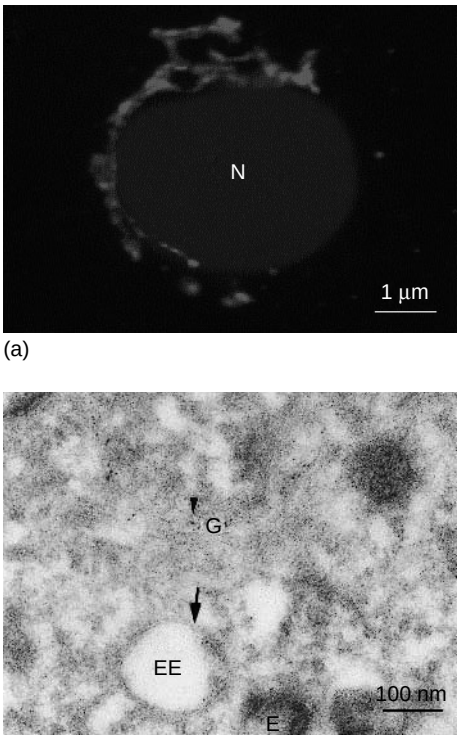

Fig. 1 Comparison of LM and EM. (a) Shows an immunofluorescence micrograph showing labeling for a trans-Golgi protein sialyl transferase. A Hela cell line stably transfected with a construct containing sialyl transferase tagged with a domain from the vesicular stomatitis virus G-protein (provided by Tommy Nilsson) was fixed with 3% formaldehyde followed by a permeabilization with 0.1% Triton X-100. The cell was then labeled with a rabbit antibody against G-protein tag followed by a secondary antibody coupled to rhodamine. The typical ribbon-shaped appearance of the red Golgi complex is evident, next to the nucleus. The latter is revealed using DAPI staining for DNA that was done at the end of immunolabeling reaction. (b) Shows a Tokuyasu cryosection of the same specimen labeled with the same antibody and protein-A gold (arrowhead). Note the significant amount of labeling on the trans-side of the Golgi. There is also sparse label on membranes of some endocytic structures (E). EE – represents the early endosomes, evident by their sparse labeling for 5 nm gold-BSA (arrow) that was internalized for 5 min before fixation. Bars = (a) 10 µm and (b) 100 nm. Micrographs courtesy of Veronika Neubrand and Anja Habermann. (See color plate (p. xxii).)

immunogold labeling; the essence of these methods will be covered.

At the end of these preparation methods, the specimens are placed on an EM grid. The two different ways of imaging these specimens at room temperature or using a cryostage are then described. A recently perfected approach for obtaining three-dimensional models from a series of tilted images by tomography will next be summarized, followed by two sets of methods that visualize (mostly) surface structure, namely freeze-fracture (etching)/replica methods for transmission electron microscopy (TEM) and low- and high-resolution scanning EM. The ability to use scanning transmission EM (STEM) for mass determination and X-ray microanalysis for the detection of some ions will be briefly mentioned. The powerful use of stereology to quantify structural parameters from sections will then be outlined. Following a brief description of methods that can be used to correlate LM with EM data, and a note of caution about interpreting thin section images, I end by offering some guidance on which technique is appropriate for which scientific question.

2 Specimen Preparation

Before a specimen can be visualized by transmission EM, it must fulfil a crucial criterion – it must be *very thin* (less than about 0.4/0.5 μm) for electrons to pass through it. In all transmission-EM approaches, the specimen is mounted on a 3-mm diameter metal (usually copper) grid (0.01 mm in thickness). This supports the specimen between the grid bars and allows electrons to pass through the specimen that lies across the holes in the grid. The grid is usually covered with a thin plastic film (such as formvar or collodion) and preferably coated with a thin layer of evaporated carbon before attaching the specimen. This provides stability by reducing surface charge buildup. The evaporator is a prominent machine in almost all EM laboratories.

In practice, there are two different possibilities for adsorbing the specimen onto the grid. If the specimen is particulate, that is, a discrete entity that is no larger than, at most 0.5 μm in height, it can be directly adsorbed onto the grid. Thin regions of whole cells grown on EM-grids can also be analyzed this way, using a newly developed cryo-EM approach (see Sect. 6.1 on Tomography, below). If, however, the structure is larger or, more commonly, one wants to see structures inside organelles, cells, or tissues, one must embed the specimen, or solidify it by vitrification, and then cut thin (usually 50–300 nm) sections. An alternative approach is to freeze-fracture the specimen and analyze the fractured surface, or after etching, the subsurface topography. Both TEM and scanning electron microscopy (SEM) can be used to visualize these specimens.

The key to the analysis of any cell structure, at any scale, is the ability to "fix" the structure, either by chemical cross-linking, or, more preferably, by arresting it physically in its aqueous environment by freezing. Although chemical fixatives have been used successfully for EM since the early 1940s (see Fig. 17), the introduction of the concept of cooling into the vitrified state has now emerged as the fundamental innovation in EM-specimen preparation. The ability to vitrify a specimen within milliseconds, in conjunction with a physical proof that the specimen is indeed in a vitreous state, now offers the best possible method that provides a fundamental frame of reference for all EM-specimen preparation methods.

Native biological material is extremely fragile and sensitive to electron beam damage. For this reason, a series of methods have been developed to optimize specimen preparation, the *sine qua non* of all EM. While one kind of approach, environmental SEM (ESEM) is starting to provide a means of imaging specimens (such as whole cells) in their aqueous environment, this is still in its infancy. This approach will not be discussed further here.

3 Ambient Temperature EM Methods

3.1 Negative Staining of Particles

The simplest and most rapid approach in EM is "classical" negative-staining (Figs. 2, 4). This approach is useful for proteins, macromolecular complexes, viruses, (small) bacteria, and isolated organelle fractions. Particles are adsorbed onto EM-grids having a suitable surface. This surface is crucial for success and in most cases the grids are covered

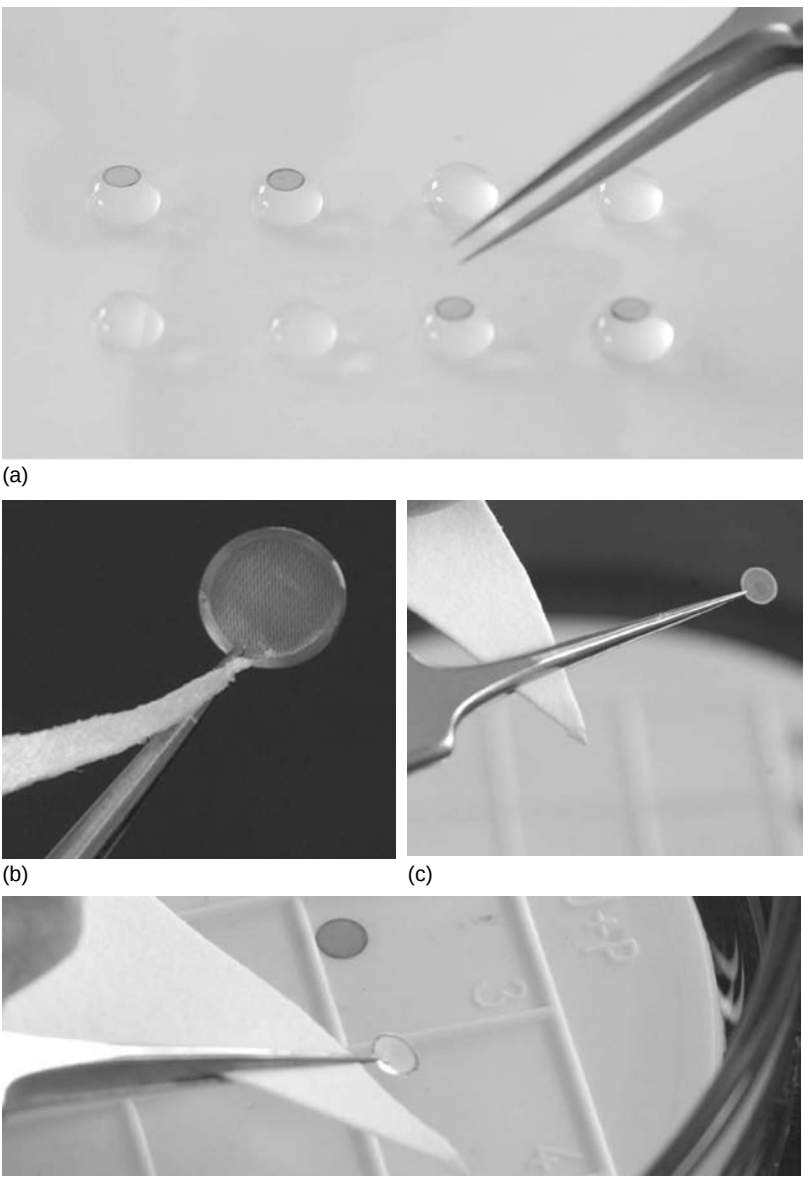

Fig. 2 Negative staining in practice. In (a) the grids are seen floating on solution, they can be moved from one drop to the next using a fine forceps. (b) After the final staining reaction, the grid is lifted from the drop of stain and a filter paper is used to remove the excess stain; (c) and (d) show a simple trick to cleanly remove the dried grid from the forceps. A triangular piece of filter paper piece is inserted in the forceps and is then used to gently push out the grid cleanly onto a suitable support.

with a layer of evaporated carbon on a suitable plastic film (e.g. formvar). It is often an advantage to "glow discharge" grids, a procedure whereby air (in a light vacuum) is ionized by ~500 to 1500 V, thereby making the grid surface hydrophilic (methods are also available to make grids hydrophobic; these are useful for some specimens). Grids are floated on as little as 3 to 5 μL of an aqueous (usually buffered) solution of the specimen for a few seconds to minutes to adsorb the material.

The grid plus absorbed particles are then rinsed rapidly with water and floated on a (1–5%) solution of a heavy metal stain, such as uranyl acetate, sodium phosphotungstate, or ammonium molybdate. If pure water is not desirable, the salts in the buffer (which will react with the heavy metals) can be removed by many rapid changes (1 s each) of the grids on pure stain. After a few seconds to 1 min on the final drop, the grid is removed with forceps and the excess stain is partially removed with filter paper (Fig. 2). A relatively thick layer (~100–300 nm) of stain needs to dry down around the specimen. The staining is referred to as "negative" because the stain fills the spaces around the specimen that are seen outlined in negative contrast. Examples are shown in Fig. 4(a–c). Although technically trivial, there are a number of "tricks," and in the hands of specialists, this approach can provide a relatively high level of resolution/information..

This approach (and the positive staining discussed below) is easily compatible with immunogold labeling before the drying step. One limitation is that only antigens exposed on the surface of particles can be labeled. In some cases, the particles can be treated with reagents that can "open up" the particle. In one such example, the reducing agent dithiothreitol (DTT) was used to expose antigens that are normally buried within vaccinia virus particles (see Fig. 4b).

3.2 Positive-negative Staining Approaches

The adsorption staining procedure developed by Tokuyasu to protect thawed cryosections from drying artifacts is a powerful alternative method that is easier to control and to image than negative staining, especially for beginners. The method is shown in practice in Fig. 3 and schematically in Fig. 8(g). In this approach, the stain adsorbs to some parts of the specimen (positive staining) and can be seen because of the low background contrast. Some negative-staining effects can also be seen, depending on the specimen and the precise conditions used (see the right virus particle in Fig. 4b). This is expected since heavy metal compounds used for negative staining also positively stain biological structures.

This method has several advantages over classical negative staining. (1) The stain adsorbs to selected structures, usually by ionic interactions. Therefore, the differences in intensity of staining in different parts of the structures provide additional information not available by negative stain methods. (2) The fact that the overall concentration of heavy metal stain over the grid is far less than in negative staining means that beam damage, often a serious problem in negative staining, is far easier to deal with. (3) This means that it is also easier to collect large data sets, as for tomography due to the reduction in beam-induced damage. Images such as the one shown in Fig. 4(b) are currently being analyzed by this approach. Membrane organelles collapse far less than with normal negative staining

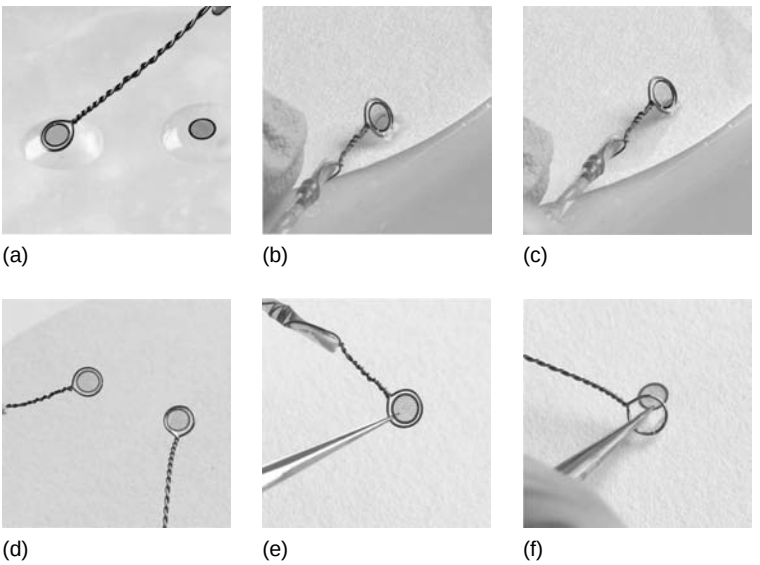

Fig. 3 Methylcellulose embedding in practice shows the basic technique involved in methylcellulose embedding. (a) The grid, having been previously rinsed in distilled water, is placed on a solution of methylcellulose and uranyl acetate on ice. (a–c) The grid is then looped out and the excess stain is removed by touching the loop onto a surface of filter paper (d) The grid is allowed to air-dry. Ideally, the interference colors should be gold, purple, or blue at the end of the technique. (e) and (f) show removal of the grid from the loop.

(4) Finally, by varying the concentration of the stain, one has the possibility in the Tokuyasu approach to vary the physical appearance of the particle; for example, locally high concentration of uranyl acetate over some structures can reveal negatively contrasted thin tubules that are often more difficult to visualize by positive-staining methods. An example is given in Fig. 4(b).

Fig. 4 Examples of negative and positive staining for EM. (a) Shows a classical negative staining using phosphotungstic acid of tobacco mosaic virus (TMV). The typical rod shaped appearance of this virus is seen. Some subunits of the virus are also seen end-on. It is evident that these are spherical in profile. (b) Shows vaccinia virus that has been treated with the reducing agent dithiothreitol (DTT). This treatment is known to break disulfide bonds within the virus. The virus was then embedded using the metal cellulose/uranyl acetate mixture as shown in Fig. 3. The increased stabilization of structures and the partly positive staining by this procedure allows more details of this relatively large virus to be seen. The arrowheads indicate two round tubular projections that have been ejected from the core of the virus by the DTT treatments. We believe these may facilitate virus entry into cells during infection. The particle on the right shows a side-view of another particle. In (c), the same virus preparation can be seen after conventional negative staining using uranyl acetate. Less fine detail is evident after this approach. Nevertheless, the areas where stain enters the particle can be delineated by the heavy metal stain. The arrowheads indicate apparent openings between the virus and the core of the particle. Bars = 100 nm. (a) – Courtesy of Heinz Schwarz; (b) and (c) – from the author.

3.3 Classical Resin Embedding

The majority of thin section analysis worldwide is still carried out using room-temperature specimen preparation techniques in which chemical fixatives, especially glutaraldehyde, are applied. These approaches are structurally acceptable in so far as the structures they describe have been observed by a cryobased approach; this statement is reasonably valid for all known organelles. For new structures, their acceptance should

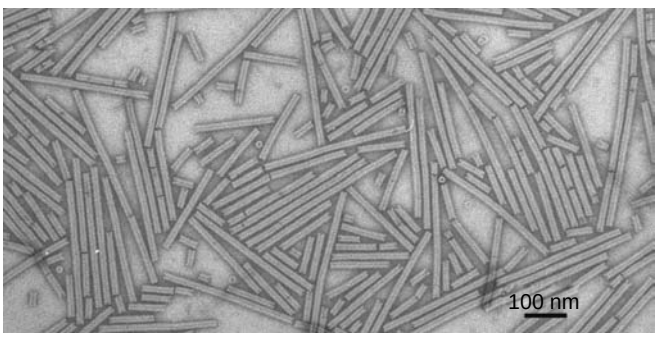

(a)

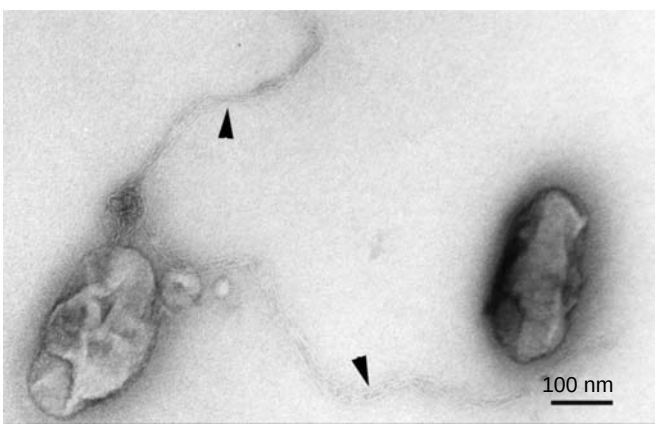

(b)

(c)

be based on confirmation by a cryobased approach, such as freeze-substitution or using hydrated cryosections.

By far the most widely used approach for ultrastructural analysis today, as it has been for the past 50 years, is the epoxy resin method. This is a witch's brew of empirically determined steps for converting fresh tissue, cells, or isolated organelles into a thin section. Although the variety of recipes is greater than in most cookbooks, the general principle is as described below. For analyzing repeatedly the same cells or tissues whose ultrastructure is well known, as in pathology, the conventional plastic-section approach is still the method of choice in most laboratories. Many useful textbooks on EM can be found in many established EM laboratories around the world (usually collecting dust!).

3.4 Chemical Fixation

Glutaraldehyde is the chemical fixative of choice for preserving structure for conventional EM. The introduction of this fixative was, without doubt a revolutionary innovation for EM in cell biology; a number of organelles, such as microtubules, were first seen only after they could be preserved with this reagent. It is a powerful cross-linker of amino groups, especially in proteins and amino-lipids. Although the chemistry involved in this process is very complex, and complicated more by the fact that the cross-linking coincides with the process of cell death, in practice it is very simple to apply the fixative to any "open" cell system; that is, cells that are accessible to a solution without having a barrier, such as a cell wall or a cuticle to block access. If cells are growing in culture, the best results are usually achieved by adding the fixative to the growth medium directly, for the first few minutes, and then switching to buffered fixative. Any barrier to the solutions, such as cell walls, must be dealt with on an individual basis. The best advice is to check how others have prepared the cells of interest. If one wishes to fix a tissue *in situ* in an animal, the best procedure is usually *perfusion*, a nontrivial surgical operation. It is important to respect the old rule of thumb that, for optimal results, the specimen pieces should be less than 1 mm in all dimensions to ensure good penetration of reagents.

When immunolabeling, as opposed to purely structural analyses, is the goal, it is often the case that extensive cross-linking with glutaraldehyde will tend to sterically hinder the access of antibodies to the antigens. A weaker cross-linker, formaldehyde (which, in solution becomes methylene glycol) is therefore used for this purpose, either by itself, or in combinations with (lower concentrations of) glutaraldehyde. For example, in standard protocols for the Tokuyasu method (see below), cells are fixed for 15 to 60 min in 4% formaldehyde and 0.1 to 0.2% glutaraldehyde followed by an overnight (or longer) fixation in 4% formaldehyde in a suitable buffer.

Aldehyde cross-linking generally is a process that releases protons. This lowering of the pH within cells is unlikely to be protected by the buffer since all the routinely used buffers are charged molecules that are thought to enter cells very inefficiently. The cross-linking reactions also consume oxygen.

Although this approach can provide a faithful presentation of ultrastructure, there are many known examples of specimen preparation–induced artifacts. This

usually becomes evident when a less-perturbing cryobased method is used. A classical example is the bacterial "mesosome"; earlier considered a distinct organelle, it is now known to be a fixation artifact, from the use of hydrated cryosections. Another striking example of aldehyde-induced artifacts in the frog retina revealed by high-pressure freezing and freeze-substitution has already been written about.

3.5 Postfixation

For pure structural studies, but generally not for immunolabeling, the aldehyde fixed tissue is "postfixed" in (0.5–1%) osmium tetroxide (OsO_4). One of the very first micrographs using this fixative for EM is shown in Fig. 16. This heavy metal can cross-link lipids and protein and adds general contrast to many organelles, especially membranes (enhancing the trilaminar appearance). It is also clear that artifacts may be induced, the most striking being the proteolysis of structures such as actin filaments. It is generally accepted that, when used at the low temperature for freeze-substitution (see below), the use of OsO_4 is generally free of artifacts and is a useful addition to enhance contrast. As pointed out below, OsO_4 is generally incompatible with the polymerization of many resins used for immunolabeling (such as lowicryl). Concentrations up to 1% in the freeze-substitution schedule may be used without interfering with UV polymerization of these resins. Osmium can also be successfully used with LR white (see below) (see Fig. 7). Many workers add a second postfixation step after the OsO_4 using uranyl acetate, which binds to negatively charged phosphate groups in phospholipids, is generally an excellent stabilizer of lipids. Tissues are often left in aqueous solutions of this stain for, for example, 1 h; an alternative so-called *en bloc* procedure that we often use is to leave specimens overnight in a saturated solution of uranyl acetate in 70% ethanol prior to embedding.

3.6 Dehydration and Embedding

The epoxy resins (e.g. Epon, Spurrs, Araldite) are highly hydrophobic, while many of the methacrylates have a lower hydrophobicity, but are still immiscible with water. For this reason (as in paraffin embedding for histology), the water must be replaced (gradually) with an ascending series of ethanol or acetone (which tends to extract more material than ethanol). At the 100% solvent stage, the resin is gradually mixed with the solvent over a period of a few hours until it can be placed in pure resin. At this stage, the specimen in resin is mounted into an embedding mold (Fig. 5b) or other suitable container (e.g. BEEM capsule) and left overnight at room temperature (at which little polymerization occurs) for complete infiltration, before being allowed to polymerize for about 12 h at 60 to 70 °C. The specimen is now ready to section. The standard epoxy and araldite resins are quite viscous and it is often a problem to penetrate some specimens, such as many plant and insect tissues. In 1969, Arthur Spurr introduced his concoction, known as *Spurr resin* as a low viscosity epoxy resin alternative. This has become the standard resin for many specimens. Alternative low viscosity–embedding media are now also commercially available. For more details there are many classical textbooks available in EM laboratories. For an example of

an image of an epoxy resin section, see Fig. 6.

3.7 Methacrylate-based Embedding

After the introduction of the complete Tokuyasu method in 1978, this approach was the section method of choice for immunolabeling. For the next two to three years, all alternative embedding protocols were far inferior for this purpose. However, in 1980 appeared the first in a series of papers from Eduard Kellenberger's laboratory in Basel, in which an excellent alternative approach became available. After years of painstaking work in collaboration with an industrial partner (Lowi), Carlemalm et al. were successful in introducing the first of the Lowicryl resins, the more hydrophilic K4M, and the more hydrophobic HM20. Its great potential for immunogold labeling was shown already in the first publication. The basic approach is no different to the epoxy protocol outlined above; it is simply a question of which solutions, in which order, and for how long. However, the resins are polymerized by UV light rather than by heat.

Although these resins can be used at room temperature, the goal from the outset was to develop these resins predominantly for use at low temperatures. In the initial publications, an approach known as *progressive lowering of temperature* (PLT) was introduced. Here, starting at the dehydration step, the temperature is gradually lowered to −35 °C. The importance of these resins is that they can be polymerized at this temperature using UV polymerization (although they can also be polymerized by heat). This polymerization usually takes one to three days and often the blocks need additional curing time at room temperature in direct sunlight (without glass).

In 1985, this group introduced new resins (K11M) that could be polymerized down to −60 °C and HM23, which will polymerize even at −80 °C. These became especially useful in combination with freeze-substitution (see below, and Figs. 13, 15b). The important contributions of Bruno Humbel and Martin Müller toward this approach should also be mentioned.

In England, Brian Causton, a chemist, in conjunction with the London Resin (LR) Company, developed an alternative set of resins LR white and LR gold. These resins are now widely used for EM immunolabeling, especially in laboratories that lack

Fig. 5 Plastic sectioning for EM. In (a), a trimmed Epon block is shown in a Leica ultramicrotome being sectioned on a Diatome diamond knife. The 2.5-mm knife-edge is indicated by arrows. A ribbon of sections is seen. This ribbon is only slightly longer in length than the diameter of an EM grid. Normally, one would tease this ribbon away from the knife-edge using a fine eyelash. The grid is then lowered carefully, film side down, onto the ribbon and pressed gently. The grid is ideally rotated on the surface of the water so that the water can run off smoothly from the surface of the grid. The grid is then raised and air dried. An alternative method that avoids section wrinkles is to dip the grid into the water beneath the sections and then carefully raise the grid at about a 45° angle to fish the ribbon. Photograph courtesy of Robert Ranner Leica Microsystems GmbH, Vienna. (b) Shows specimens that have been polymerized in an embedding mold. The specimens are black following osmium treatment (arrows). A piece of paper with specimen details can be conveniently coembedded with the specimen. Courtesy of Anja Habermann and Maj Britt Hansen.

the equipment for cryosectioning. As mentioned, LR white is also compatible with osmium tetroxide treatment and recipes exist for providing beautiful preservation in conjunction with immunogold labeling (Fig. 7).

All the methacrylate-based resins are sectioned in the same way as other plastic-embedded blocks (Fig. 5a). For the more hydrophilic ones, such as Lowicryl K4M, the water level in the trough of the knife boat (Fig. 5a) needs to be lowered, compared to Epon sectioning, to prevent the sections being pulled behind the knife.

All the resins used for EM are potentially toxic, as are the fixatives and appropriate caution should be taken when handling

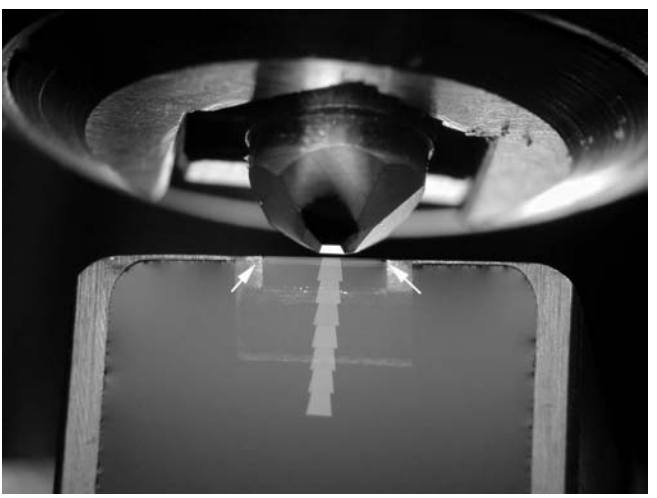

(a)

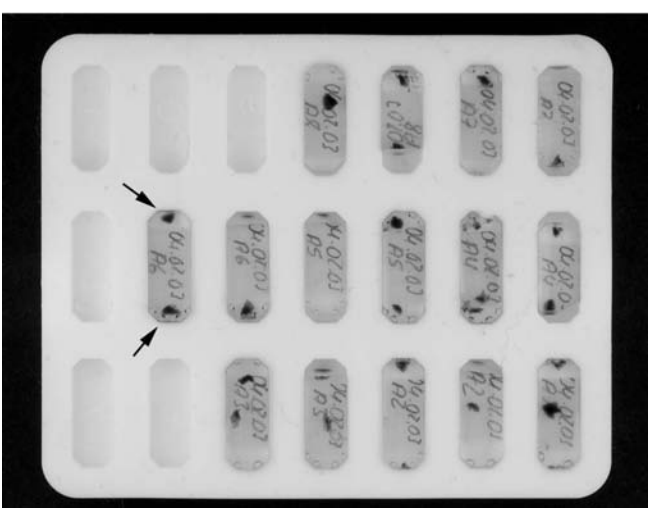

(b)

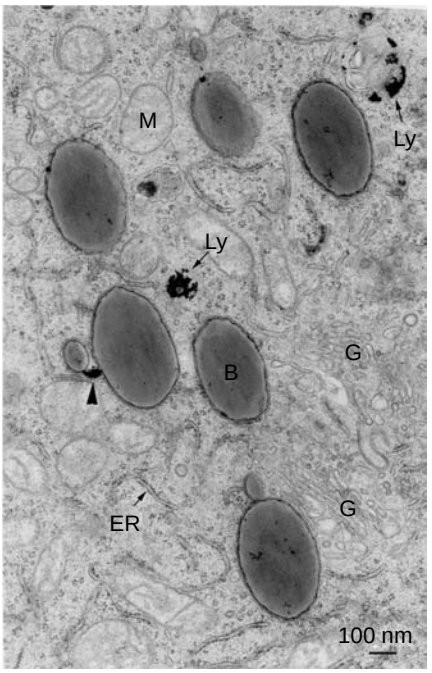

Fig. 6 Example of an epoxy resin section. This shows a section of a J774 macrophage that had internalized BSA-conjugated gold particles for 1 h followed by a chase in medium free of gold for 1 h. Under this condition, the gold is in late endosomes and lysosomes (Ly). Subsequently, 1-μm latex beads (B) were added for 1 h followed by a further 3-h chase. Gold can be seen in late endocytic organelles and in the latex bead phagosomes (arrowhead), after a fusion process. Other organelles, Golgi stacks (G), mitochondria (M), and ER are indicated.

them (eye protection, gloves, hood, etc.). Particular care should be taken when working with the Lowicryl resins that are well known to cause skin problems.

3.8 Plastic Sectioning

The specimen embedded in plastic must next be trimmed to make a small block that can be pyramidal or rectangular at its tip (usually 0.01–0.1 mm, see Fig. 5a). This can be done manually with razor blades but is much more precisely done using commercially available block trimmers that have metal or diamond blades.

The trimmed block is next mounted in an ultramicrotome equipped with a glass knife or, more conveniently using a diamond knife (Fig. 5a). The combination of modern ultramicrotomes and state-of-the-art diamond knives now enables sections as thin as 20 nm to be easily and reproducibly obtained.

As shown in Fig. 5(a), during plastic sectioning, the sections are floated on the surface of (clean) distilled water where they can be manipulated with an eyelash and mounted on a carbon and plastic (e.g. formvar) coated grid. In this simple procedure, the grid surface is brought into contact with the sections, which are adsorbed. The sections are routinely stained by floating the grids on a solution of, first uranyl acetate (which may not be necessary if it was included before embedding), and second with a solution of lead citrate (for a few minutes only) (see Fig. 2a). After rinsing with water and air-drying, the grids are now ready for imaging.

When immunogold labeling is desired, sections of methacrylate-embedded material can be labeled, as described below. In this case, the staining with

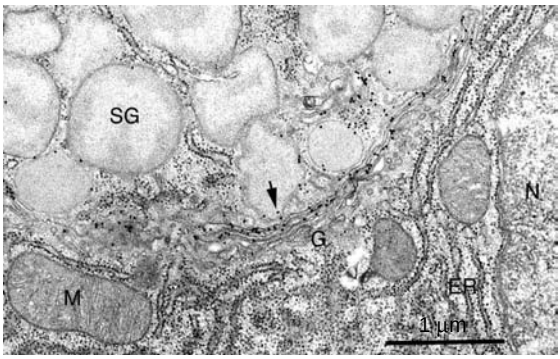

(a)

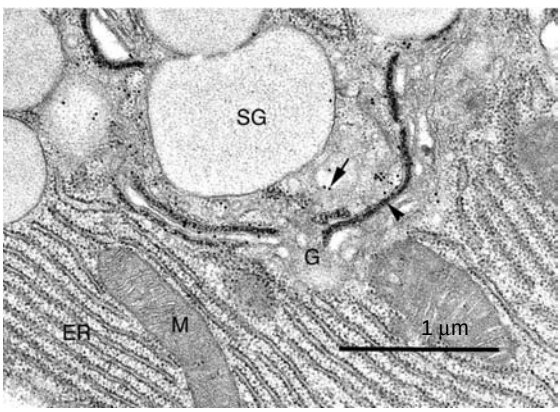

(b)

Fig. 7 Example of embedding in LR White. (a) and (b) show rat parotid acinar cells embedded in LR White. Pieces of parotid tissue from a rat were fixed in a mixture of 0.5% glutaraldehyde and 4% formaldehyde under microwave irradiation, followed by postfixation with 1% osmium tetroxide containing 1.5% potassium ferrocyanide (reduced osmium) under microwave irradiation and continued for 30 min on ice. Tissue blocks were embedded in LR White resin, and ultrathin sections were obtained. The ultrathin section mounted on a nickel grid was pretreated for etching by a saturated aqueous solution of sodium metaperiodate for 30 s, followed by treatment in 1% bovine serum albumin. Immunolabeling was carried out using anti-GF-1. GF-1 is a monoclonal antibody recognizing 105-kDa glycoprotein in the Golgi apparatus of serous exocrine cells. Colloidal gold particles (arrow) can be recognized to distribute on the trans-cisternae of Golgi apparatus (G). The cells in (b) were additionally reacted for the trans-Golgi enzyme thiamine pyrophosphatase (TPPase) subsequent to the primary fixation. This reaction product is seen as an electron-dense precipitate in one trans-Golgi cisterna (arrowhead in (b)). The sections were labeled with an antibody GF-1, revealed by immunogold. The gold particles are seen to be restricted to the trans-side of the Golgi stack, colocalizing with the TPPase reaction product (arrowheads). The ER is indicated, as are the secretion granules (SG), the nucleus (N) and mitochondria (M). Micrographs courtesy of Shohei Yamashina.

heavy metals is carried out *after* the immunolabeling. An advantage of this approach for immunolabeling is that the room-temperature cut sections can be stored dry on grids indefinitely and still be immunolabeled when desired. These sections can also be powerful tools for use at the immunofluorescence level.

3.9 The Tokuyasu Sucrose Embedding – Cryosection Method

An elegant and rapid method for visualizing structure and immunolabeling using cryosections was pioneered by Tokuyasu between the late 1960s and 1978. This is now used extensively worldwide.

In contrast to the hydrated cryosection method (see below), there are two severe restrictions in obtaining cryosections for immunolabeling, in addition to the obvious need for the cryomicrotome. One is that immunolabeling needs to be carried out above 0 °C and the other is that ultrathin cryosections of fresh specimens are destroyed when thawed. For this reason, the specimen must be chemically fixed prior to sectioning, usually with a light, fixation protocol (see above).

So, the role of the sectioning process in this approach is simply to be able to cut thin sections of fixed cells and tissue that can subsequently be immunolabeled while fully hydrated at ambient temperature. While this may be a disadvantage for high-resolution preservation, it also offers a number of practical advantages in comparison to the hydrated cryosectioning method. First, because the material is chemically fixed, the membranes are permeable to high concentrations of sucrose, which is an excellent cryoprotectant. This means that relatively large pieces of specimen (>1 mm^3) can by *vitrified routinely* by simple immersion in liquid nitrogen (which is a relatively poor coolant). Moreover, sucrose (up to 2.3 M) allows the specimen to be more easily and uniformly thin sectioned in a cryoultramicrotome at, for example, −100 to −120 °C. So, although the method involves a vitrification step, it cannot be considered a *bone fide* cryo-EM method. Nevertheless, it is mostly referred to as the cryoimmuno-EM method. The steps in this method are shown schematically in Fig. 8.

The trimmed specimen is cut with a glass knife, or now more commonly the superb new generation of diamond knives designed for cryosectioning. The temperature is usually in the range of −90 to −120 °C. At these temperatures, water obviously freezes and a major difference between cryo and plastic sectioning is that cryosections are cut on a dry knife (i.e. no water trough). The colder the specimen, the harder the block; for cutting sections for LM, the blocks are made softer by raising the temperature from −50 to −80 °C. Many workers also raise the temperature of the block to −80 °C for trimming the block before cutting thin sections at colder temperatures. Plasticity of the block is also an important parameter to vary for optimal sectioning properties, and compounds such as polyvinyl pyrrolidine can be added to the sucrose infusion in order to facilitate the sectioning process. At all temperatures, even the best sections are initially compressed. They are then picked up using a 1- to 2-mm loop of thin wire that contains a drop of 2.3 M sucrose (Fig. 8d). The sucrose drop is rapidly brought into the cryochamber and, upon contacting the sections and brought to room temperature, it decompresses; a spherical cell profile that is significantly compressed after sectioning can now regain its spherical shape. Often, this method leads to overstretching of the sections, causing structural artifacts. For this reason, the use of a mixture of sucrose and methylcellulose was introduced, which could greatly improve structural preservation (see Figs. 1b, 9, 14). Tokuyasu (personal communication) considers the osmotic pressure to be an additional important factor in section pickup. To reduce the impact of high osmotic pressure, the use of a mixture of 1.7 sucrose solution with the detergent Tween 80, at a final concentration of 0.01% to retrieve cryosections embedded in 2 to 2.3 M sucrose was recommended; we have found this approach to be very useful. Section compression can also be minimized by reducing the sectioning angle

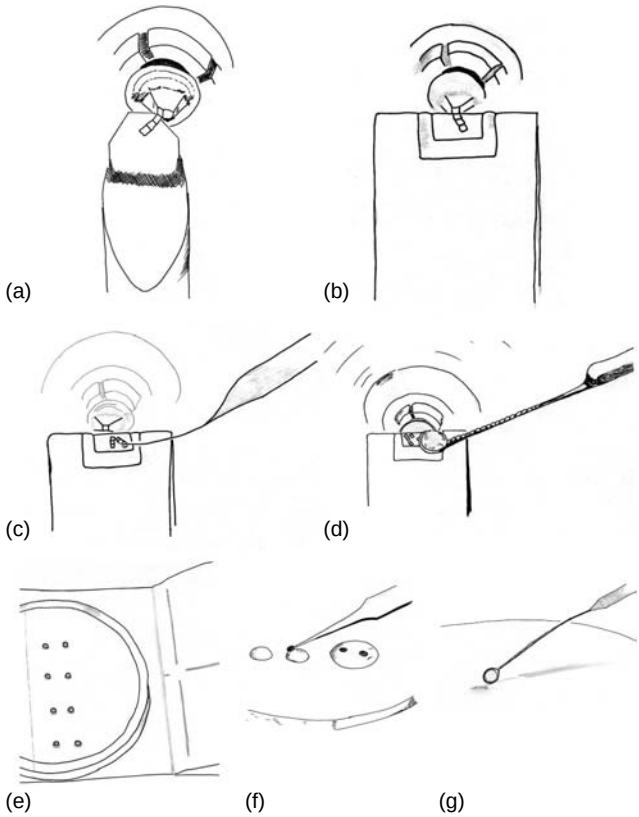

Fig. 8 Schematic illustration of the Tokuyasu method. These drawings, made by Paul Webster, illustrate the basic essential details of the Tokuyasu cryosectioning and labeling method. All steps are carried out within the chamber of the cryoultramicrotome. Prior to step (a), the specimen will have been fixed, infused with sucrose, and placed on a specimen pin. (a) Shows the procedure for trimming the block using a trimming device. Glass knives or special diamond trimmers can also be used for this purpose. In (b), the specimen is sectioned on a glass or, in this case, a diamond knife, and the sections are manipulated using an eyelash probe (c). In (d), the sections are picked up on a loop containing 2.3 M sucrose or a mixture of sucrose and methylcellulose. The loop is then touched onto the surface of a grid and the grid is raised and allowed to be floated (section surface down) on a drop of liquid. (e) Shows the procedure for maintaining a moist chamber around the grids. A petri dish is then layered over the grids, and a piece of moist filter paper is added on one side of the petri dish; (f) and (g) show again the principle of drying the grid after embedding in a methyl cellulose mixture (see also Fig. 3).

of the diamond knife. A new oscillating diamond knife is being developed that already looks highly promising in reducing this problem even further.

Once the sections are thawed on the surface of the drop of sucrose (or the mixture with methylcellulose), the loop is brought to the surface of an EM grid and, upon contact the sections are transferred to the grid surface. From now until the final drying step, the fragile sections must remain hydrated and not be allowed to dry. The universal procedure for immunogold labeling can be applied (see below). It should be noted that in addition to its use for immunolabeling, this method is also an excellent and rapid (1–2 h) method for preserving and visualizing structures in thin sections by EM. It is surprising that this method has not been more widely used, for example, in pathology, as a rapid diagnostic tool.

Either with or without immunolabeling, the final step in the preparation is to dry/embed the sections. Since, as mentioned, the sections are easily destroyed by air-drying, Tokuyasu spent a number of years developing what in the end turned out to be a surprisingly simple approach to protect the sections from collapsing during the drying process. His elegant solution was to float the grids with sections (that were previously floating on pure water) on a mixture of heavy metal stain (usually uranyl acetate) and a polymer (usually methylcellulose, this is more soluble in the cold) for a few minutes. The grids are then looped out using a ~3.5-mm loop, the excess solution is removed by blotting and the grid is air-dried/embedded (Figs. 3, 8). After removing the grids from the loop with a forceps, they can be visualized by EM. This simple procedure revolutionized the use of thawed cryosections for immunolabeling at the EM level.

3.10 On-section Immunogold-labeling Procedure

The localization of specific antigens on isolated particles, or on thin sections can be performed using immunogold-labeling methods. A theoretical drawback of these methods is the need to label the specimen at physiological temperatures, which limits their use for *bone fide* cryobased approaches. Nevertheless, these methods have been very powerful in a

Fig. 9 Example of using the Tokuyasu technique. These images show cryosections of glutaraldehyde-fixed Chinese hamster ovary (CHO) cells showing the distribution of Prion protein using an anti-PrP C monoclonal antibody Fab fragment (R1). (a) Gold labeled PrPC (arrowheads) was found to be highly enriched in the caveolae at the plasma membrane (PM) (not in this picture) and caveolae-containing membrane structures in the trans-Golgi network (TGN) around the pericentriolar region; arrows pointing to centrioles. (b) Endocytosed 5-nm protein-A gold particles (which bind specifically to PrPC; arrowheads) indicate the sites of PrPC molecules that had been taken up via caveolae. These results indicate that typical early endosomes do not intersect significantly with the endocytic pathway of PrPC-containing caveolae. In addition to the typical small tubulo-vesicular structures, many larger multivesicular-bodies (late endosomal/lysosomal profiles, L) were loaded with small gold particles representing endocytosed PrPC, whereas the transferrin receptor (labeled with an antibody and large, 10 nm, gold) localized in nearby and distinct structures. As expected, ER, Golgi complex, mitochondria, and nucleus were not labeled, indicating highly specific labeling in these experiments. G: Golgi complex, n: nucleus. Micrographs courtesy of Peter J. Peters.

wide spectrum of molecular cell biological applications.

The essence of all immunolabeling is the same for all procedures at the LM and EM level. However, the use of sections (for LM or EM) obviates the need for a permeabilization step (see below). Subsequently, a minimum of three incubation steps is required, along with rinsing steps. The first step is to *block nonspecific sites*. All proteins (some more than others) have the capacity to adhere to (most) surfaces and it is generally thought that this is due mostly to ionic and hydrophobic interactions. In practice, one can usually (but not always) block these sites on the sections (and grid) surface by simply floating the grid on a drop of a suitable protein solution for a few minutes. Many groups use different kinds of sera (e.g. goat serum) for this purpose, in which case one must be sure that this does not interfere with the immunolabeling reaction. We prefer to

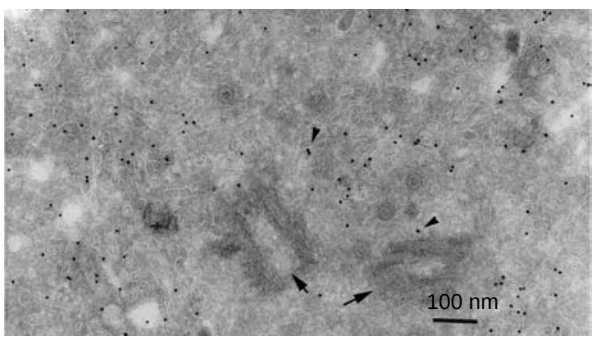

(a)

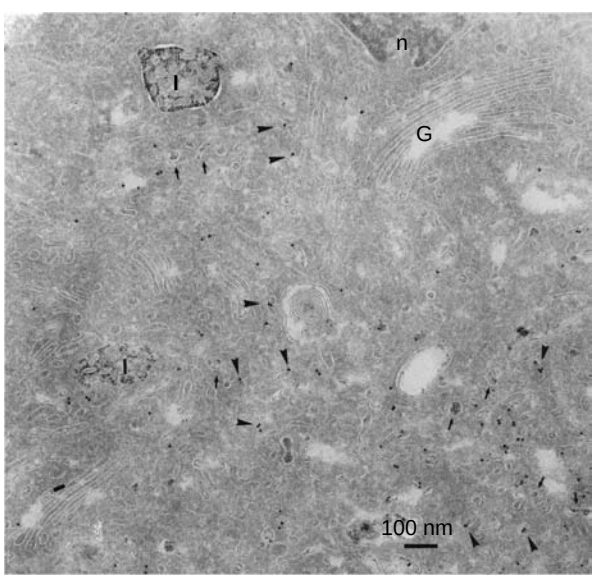

(b)

use a mixture of fish skin gelatin (from an arctic fish, via the Sigma company) and bovine serum albumin. Whatever one uses, it is recommended to dilute the antibodies and gold reagent in this solution for the subsequent labeling steps. In practice, the whole labeling procedure is most conveniently carried out on the surface of a layer of parafilm (see Fig. 2a).

3.10.1 Step 2 – Antibody Labeling

The grid floating on the block solution is transferred to the surface of a 5- to 10-μL drop of the "optimal" dilution of antibody in the blocking solution and left for 30 to 60 min in a humid chamber on the parafilm surface (see Fig. 2a).

The problems of antibody specificity must be seriously considered in evaluating EM (as for LM) immunolabeling. It is important to note that apparent proof of specificity by an immunochemical approach such as Western blotting does *not* necessarily constitute proof of specificity on the surface of a thin section since the conditions facing the antigen are quite different. It is crucial to realize that the complexity of possible interactions on the surface of a section can give rise to subtle and serious artifacts that may fool the observer. It is also important to realize that the numbers of antigens available at the surface of thin sections can be deceptively low.

It is crucial to determine the optimal concentration of antibodies for the best signal-to-noise ratio. Above this level, background labeling can increase significantly. The simplest take-home message to achieve this is to empirically determine the highest concentration of antibody (which, in some cases, may even mean undiluted antibody!) that does not label structures assumed to be free of antigens. The foundation of this assumption is the available cell biological knowledge on the system.

Depending on the gold reagent that will be subsequently used, one may need to use a secondary antibody step. For example, if one uses protein-A gold (which is the preferred reagent in many groups, including ours), and one starts with a mouse antibody (that mostly do not bind protein-A), then an intermediate step of, for example, rabbit anti-mouse is required before applying the protein-A gold. Before and after such a step, a series of rinses on drops of, for example, phosphate-buffered saline (PBS) is required for 10 to 15 min. Alternatively, one can use a secondary antibody conjugated to gold (see below).

3.10.2 Step 3 – Gold

After the rinses following the last antibody step, the grids are floated on the empirically determined concentration of protein-A gold or an IgG gold conjugate. These conjugates can be easily made in the laboratory or, they can be obtained from many commercial suppliers. In the absence of prior information, the optimal concentration of the gold reagent is the highest concentration that gives negligible labeling in the absence of a primary antibody on the sections that have been exposed to blocking solutions.

Colloidal gold particles consist of pure gold, made by reducing gold chloride ($HAuCl{\bullet}2H_2O$). A number of methods for preparing these colloids have been around since the days of Michael Faraday but the method of choice now is that introduced by Slot and Geuze. By a simple procedure involving citrate and tannic acid this method produces uniform, spherical particles of any size between about 3 and 17 nm; up to three different sized particles can be routinely used for triple labeling studies. Many proteins

can be adsorbed fairly stably (but mostly noncovalently onto the surface of gold particles), without loss of activity; the most useful (in our experience) is protein-A. The noncovalent interactions of proteins and other compounds with gold is complex and highly unpredictable; one must be aware of the possibility that any compound bound to gold may elute from the particle surface with increasing storage time. Procedures for single and double labeling have been described. Following the gold incubation, the sections must be rinsed extensively (~15–20 min) and all salts must be washed away with (preferably double or triple) glass-distilled water prior to final contrasting with uranyl and/or lead salts. The basic procedures for plastic and cryosections are described above. It should also be noted that any particulate material that can be adsorbed to grids can be labeled, but on their outer, solvent accessible surfaces. Only if the structures can be "opened up" in some way can one potentially access internally localized antigens.

3.11 Preembedding Labeling Methods

Whereas most EM labeling protocols these days is done using thawed cryosections or methacrylate sections, there is an alternative set of methods referred to as *preembedding labeling* that may be preferred for some applications. The principle here is to start with whole cells or tissue slices (e.g. 20–50 μm thick cryostat or vibratome or tissue/chopper fresh sections) and then treat the material with a solvent or detergent (or even by freeze-thawing) that removes parts of membranes in order to allow antibodies to diffuse into the cell interiors. One simple method that we often use is to put the live cells into distilled water and observe them by light microscopy. When the cells burst (usually after about 10 sfor cultured mammalian cells), fixative is added. In this approach, as in all preembedding methods one accepts structural damage at the outset.

One attraction of the preembedding approach is that at the end of the labeling procedure, the cells are embedded in conventional epoxy resins, that is, technology available in every EM lab for cell biology. Since the ultimate goal is to see the sites of antigens in the context of cell ultrastructure, it is essential to chemically prefix the material before labeling. However, this raises a practical dilemma; if cross-linking is too effective, the antibodies fail to reach many antigens. If, on the other hand, one cross-links too little, cellular ultrastructure is compromised. Since one needs to sequentially introduce antibodies and secondary gold reagents, preembedding labeling is a highly empirical and unpredictable method. Nevertheless, when it works, this approach can provide significant information (see Fig. 10).

The essential steps in a typical preembedding protocol are as follows:

1. The cells/tissue slices are fixed lightly, usually with 2 to 4% formaldehyde (often mixed with a low concentration of glutaraldehyde), usually for less than 30 min.
2. For labeling intracellular antigens, the cells are permeabilized, most often with 0.1 to 0.3% of the detergent Triton X-100 for a few minutes. If the goal is to label antigens on the outer surface of cells this step can be left out. This step, in conjunction with step 1, is difficult to control and is prone to induce artifacts.
3. The permeabilized cells are treated with a protein blocking solution (see labeling of sections above) and then with the

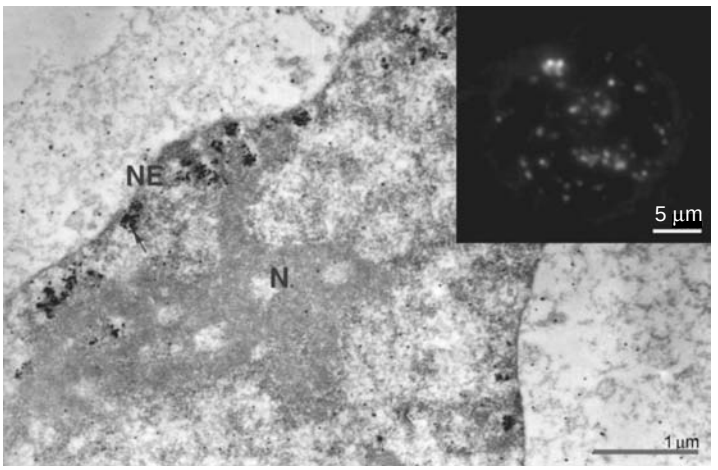

Fig. 10 Preembedding labeling. The replication of chromatin in synchronized HeLa cells has been studied using both light and EM immunodetection. The inset shows individual small fluorescence foci corresponding to DNA replication sites during early S-phase (100 min after release of cells from a double thymidine block of DNA replication). Biotin-16-deoxy-UTP was used as a marker of the newly synthesized DNA. This was delivered into cells by means of hypotonic shift procedure and was visualized by a secondary, fluorescent antibody against biotin. The corresponding EM image is seen in the main figure. For this, the mouse antibiotin was recognized by a 1-nm gold anti-mouse antibody that was subsequently silver-enhanced. The arrow indicates the sites at the periphery of the nucleus (N) where the silver-enhanced gold accumulate, which reflects the DNA replication sites corresponding to the immunofluorescence foci seen in the inset. Micrographs courtesy of Ivan Raska.

optimally determined concentration of antibody (in blocking solution) for periods ranging from 30 min to overnight. The excess unbound antibody must be removed by multiple rinsing steps with buffer (e.g. phosphate-buffered saline, PBS).

4. The cells are incubated with a gold reagent (usually for 30 min to 2 h at room temp) that recognizes the antibody (as for section labeling); also here it is crucial to have the optimally determined concentration of marker. The excess is again rinsed away.
5. The cells can now be fixed with a high concentration of glutaraldehyde (0.5–2% for 30–60 min) in order to protect the structure against the subsequent steps of dehydration and embedding. The remaining procedure is identical to the protocol given above for epoxy resin embedding.

Accessibility of reagents to antigens is often a severe problem in preembedding labeling, with the difficulties increasing as one looks for antigens deeper in the cell, especially within the nucleus. By this approach, a negative result is almost impossible to evaluate; one can never be certain whether or not the reagents had access to the antigens in a defined structure. At the primary antibody level,

antibody fragments such as Fab (which have a significant lower avidity than whole IgG) may improve access. At the gold step, it makes sense to use the smallest sizes of gold. An important innovation here was the commercial availability of 1 to 2 nm gold particles bound with secondary antibodies or protein-A. A number of studies have shown that these penetrate significantly better than 5-nm gold particles. The 1- to 2-nm gold particles are very difficult to see in a typical section. To overcome this problem, metal amplification procedures using silver or gold have long been available and widely used to enhance the size of the gold particles. It should be noted that osmium tetroxide can strip the silver off the gold particles.

At the end of this procedure, thin plastic sections of cells are examined in which gold particles provide an indication of the sites of antigens.

3.12 Immunohistochemical Approaches using Horseradish Peroxidase and Cytochemistry

For preembedding labeling, it is also possible to visualize the bound antibody using a secondary antibody bound to horseradish peroxidase (HRP). This reagent penetrates into cells better than gold particles and has the advantage that the enzymatic properties of HRP can be used to oxidize diaminobenzidine (DAB) into an insoluble polymer that binds Os_5O_4 very well. This cytochemical electron-dense reaction product can be used as an indicator of the site of the antigen. Because of the ability of the reaction product to diffuse, this method is rarely successful for cytoplasmic or nuclear antigens. In the past, it was widely used to localize antigens within the lumen of membrane organelles. However, for a number of reasons, this approach cannot be considered a "state-of-the-art" method for immunolabeling.

HRP has also been widely used as a marker of endocytic organelles, and cells will conveniently take up this compound by endocytosis. More recently, cDNAs encoding for this protein as chimeras with different targeting signals have been used to direct the protein to different biosynthetic compartments, such as the endoplasmic reticulum and the Golgi complex. This approach can be very useful to facilitate structural details of these organelles and is likely to be increasingly important as an additional marker for tomographic studies. It has also been used in an interesting method for visualizing endocytic organelles in whole mount cell preparations.

For structural identification of membrane compartments – an old approach – enzyme cytochemistry is still a useful method, although its use has been declining. It is expected that this approach will be used more often for tomographic studies. An example is provided in Fig. 7.

3.12.1 The Use of Microwave Technology for Specimen Preparation

In 1970, Mayers first introduced the idea of using microwaves to enhance the rate and efficiency of fixative cross-linking. Since that time, this approach has become widely used, not only for the fixation step but also for other steps in specimen preparation. Although the precise effects are still not fully understood, specimen exposure to microwaves results in greatly enhanced rates of penetration of chemicals into tissues as well as chemical reaction rates. In the microwave processor, reactions that take hours to occur by diffusion alone can occur in seconds to a few minutes. It is important to note that

standard kitchen microwave ovens are not recommended, since they are difficult to control and, even with short pulses, tend to overheat the specimen. Specially designed laboratory-grade microwave processors are now available that offer variable, controllable wattage and specimen temperatures. Temperature control is achieved in part by circulating large volumes of water through the microwave chamber to remove excess heat. In our group, following the lead of Paul Webster, we now routinely carry out all steps of epoxy resin embedding using this technology. Whereas the overall process used to take two to three days, it can now be completed in a few hours. The microwave oven is also likely to be an important tool in the future for immunocytochemistry by speeding up the penetration and binding rates of antibodies for immunolabeling. Its role in improving antigen accessibility to antibodies is also being exploited. For an example of tissue sections following microwave-assisted fixation, see Fig. 7. It is likely that this technology will play an increasingly important role in many aspects of EM in the future.

4 Cryo-EM Approaches

4.1 Vitrification

The pioneering ideas of Dubochet on the concept of vitrification was the beginning of a new era in EM. The fact that one requires such a thin specimen for TEM allowed Dubochet and McDowall to vitrify a thin layer of water suspended over the holes on an EM grid. Electron diffraction confirmed that the solidified water was not crystalline but remained amorphous. This breakthrough was the culmination of a whole era of studies that were initiated by Fernandez-Moren in the 1950s.

It is estimated that a freezing rate of 10^5 to 10^6 °C s^{-1} is required to vitrify water at ambient pressures. No method is known that allows cells (or parts of cells) greater than 10 μm in thickness to be vitrified under these conditions. Only at high pressures can this be achieved to a depth of a few hundreds of microns (see below). Owing to the high rates of cooling that are used, it has been estimated that the process of vitrification arrests the *in vivo* state within 100 m s^{-1} of the final perturbation. For this reason, it has evocatively been described as the "solidified *in vivo* state." There are five different cryo-EM methods we shall discuss: (1) the bare-grid method for isolated particles; (2) hydrated cryosections; (3) freeze-substitution; (4) freeze-fracture and (5) cryo-SEM approaches.

4.2 The Bare-grid Method for Particulate Specimens

Viruses, small bacteria, isolated organelles, filaments, macromolecular complexes, and the like can easily be vitrified. The remarkable consequence of the thinness of these specimens is that an almost trivial procedure allows preparations to be routinely vitrified. Following the innovation to use a perforated film grid support, the method involves suspending the specimen on such a grid mounted in a simple guillotine device. For the best results, the atmosphere around the specimen should be kept humid and specialized devices are now available for this purpose. After a brief blotting on one or both surfaces of the specimen, the grid, attached to a forceps, is rapidly shot into a ~2-cm deep chamber of liquid ethane cooled by an

external container of liquid nitrogen. All the subsequent operations such as transporting, storage, mounting in cryoholders and imaging in the electron microscope, are done at liquid nitrogen temperature to avoid devitrification, which occurs above −135 °C. The imaging of cryospecimens requires special care because highly energetic electrons, while passing through a sensitive specimen, can easily destroy the fine structure. The so-called "low dose imaging" has, therefore, been developed (see below). It should be noted that the cryo-EM approach offers the only possibility in EM for visualizing details within the interior of a structure directly. State-of-the-art examples of images obtained by the bare-grid approach are shown in Figs. 11 and 18(a).

This approach has been used for a series of elegant time-resolved analyses, especially by Nigel Unwin and colleagues. For example, the acetylcholine receptor has been visualized in different functional states. The bare-grid approach can also be combined with immunogold labeling and with positive staining using heavy metal salts. This is probably the EM method whose usage is increasing the most in laboratories worldwide.

4.3
Vitrification of Larger Material

The cryopreservation of material larger than 1 μm is a topic that has been extensively investigated. Despite earlier complications, it is now generally accepted that the goal is straightforward; the specimen must be vitrified if it is to provide a faithful reflection of the structure of the *in vivo* state. The hydrated cryosection method (see below) offers the only approach that can allow sections of native material to be evaluated. It is also the only method for unambiguously confirming, via electron diffraction, that the vitreous state has indeed been achieved; although large hexagonal ice crystals can be seen directly without diffraction, the presence of the intermediate cubic ice crystals (up to ~300 nm) cannot. These can also damage specimens.

There are different approaches for vitrifying a layer of biological material thicker than 1 μm; one can consider the more traditional methods such as (1) plunging in ethane or propane; (2) jet freezing; (3) slam-freezing on a cooled silver or copper surface. There is now a consensus that all of these approaches unfortunately, fail in practice to give a layer of vitrification that is more than 5 to 10 μm.

A relatively old method, *high-pressure freezing*, first introduced by Moor, has recently emerged as *the* technique for vitrifying substantial pieces of biological tissue. A reasonable consensus in the field accepts that a layer up to 200 μm (perhaps thicker) can be vitrified by this approach. High-pressure freezers are now available from three commercial companies and the application of this approach is growing slowly but steadily, worldwide in established EM laboratories. It should be emphasized that the technology involved is not trivial and there are many "technical tricks" that can be learned only from the specialists (this is generally true of all EM methods). Examples are shown in Figs. 13–15(b). Below, I would like to point out two very powerful adaptations that, in combination with high-pressure freezing, can be very powerful tools.

4.4
The Cellulose Capillary Tube

Hohenberg has introduced the use of cellulose microcapillary tubes in order

to prepare material for high-pressure freezing (HPF). By capillary action, an aqueous suspension can be drawn into these tubes of 200-μm diameter, which can be easily sealed by force. The cellulose wall allows molecules up to ~10 KDa to diffuse freely through it. Cultured cells can be grown within the tubes on their preferred media prior to freezing (see Fig. 13(a)). For high-pressure freezing, small fragments of these tubes can be conveniently positioned in the specimen holder of the machine. This approach can also be useful with other methods, for example, epoxy resin embedding or the Tokuyasu cryosection method.

4.5 The Fine-needle Biopsy

A second innovation by Hohenberg is potentially of enormous interest to pathologists. A specially designed jet micro-needle system has been developed that can cut

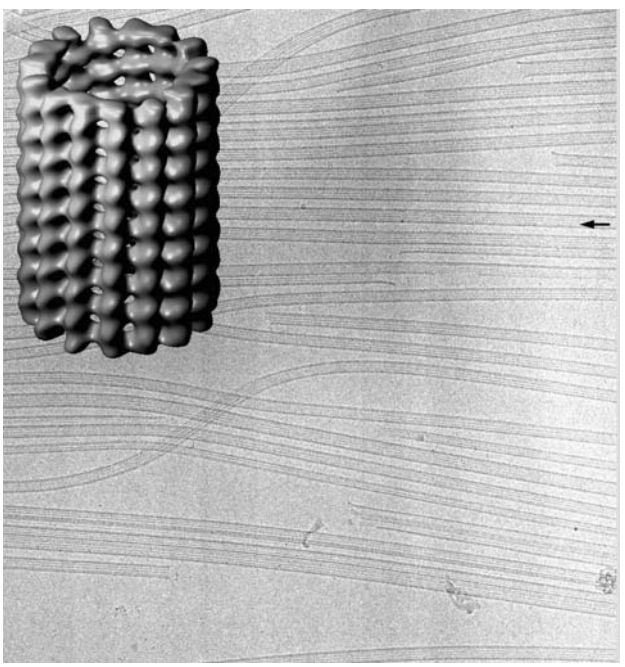

(a)

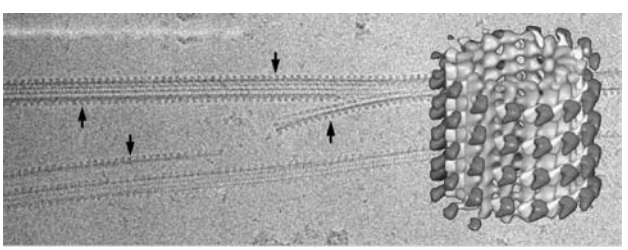

(b) Kinesin decoration

200-μm thick slices of any living tissue. These slices are ideal for high-pressure freezing. The use of this system for taking micro-biopsy specimens from human patients is now a routine (and for the patient apparently painless) procedure by this group and their medical collaborators.

4.6 Hydrated Cryosectioning

Vitrified thick specimens must be sectioned before analysis by TEM. The simplest method in theory, that is by far the most difficult to apply, is the hydrated cryosection method. The theory is indeed simple, one vitrifies a piece of tissue and prepares cryosections (at a temperature below the recrystallization temperature of vitreous water ($<-135\,^{\circ}$C)). These sections can be transferred onto EM-grids kept under liquid nitrogen until they are directly visualized (in the absence of any chemical) at $\sim-160\,^{\circ}$C, or below. In principle, this method allows direct visualization of native structure in the absence of any chemical.

In practice, the technique is quite demanding and is hampered by sectioning problems (cutting artifacts). Foremost among these is the severe compression (routinely 30–50%) that in fact accompanies all sectioning methods (in all other sectioning approaches for EM, the sections have the opportunity to stretch (decompress) on aqueous solutions). It seems, however, that in recent years all the major technical problems have found acceptable solutions and the method is now at the stage where it is ready to fulfil its promises. Moreover, new innovations are on the horizon, such as the vibrating diamond knife already mentioned that can considerably reduce the compression artifact and the use of tomography for providing three-dimensional information from such specimens. For recent examples of this approach, see Fig. 12.

4.7 Freeze-substitution

While the frozen-hydrated cryosection approach is currently something for a specialist, the approach of freeze-substitution offers a much easier-to-apply alternative method for visualizing the (close to) *in*

Fig. 11 Cryo-EM. (a) Cryoelectron micrograph of undecorated microtubules, embedded in vitrous ice. When assembled *in vitro*, microtubules may be formed from different numbers of protofilaments (which can be clearly seen in Fig. 22 after metal shadowing). A computer-assisted 3-D reconstruction in the inset (color) has been carried out on a 15-protofilament microtubule. One of them is marked with an arrow. The number of protofilaments gives rise to different moiree patterns, which can be seen on a careful inspection of the different tubes. For example, 13-protofilament microtubules exhibit very straight lines along the axis, while 15-protofilament microtubules show a characteristic alternating pattern of fuzzy and striated regions along the axis. This pattern allows one to identify different types of microtubules. (b) Cryomicrograph of microtubules decorated with kinesin motor domains; the corresponding model is shown. The motor domains appear as little globular blobs along the microtubule (arrows). These motor domains also have an intrinsic stabilization effect that creates long isolated protofilaments. The inset shows a helical 3-D reconstruction of a microtubule decorated with dimeric ncd domains. Ncd is a kinesin family member that exhibits retrograde directionality, and a unique microtubule-binding pattern, but otherwise shows strong structural similarities to conventional kinesin. It is important to realize that cryo-EM is the only approach that provides information about structural features *within* an intact particle. Images courtesy of Andy Hoenger. (See color plate p. xxiii).

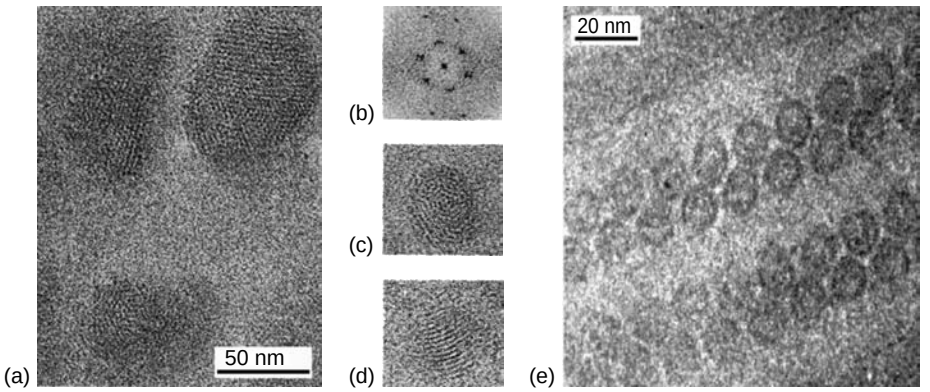

Fig. 12 Hydrated cryosections. In this approach, no chemical treatment whatsoever is used. (a) Shows a hydrated cryosection of stallion sperm chromatin after partial decondensation with 10 mM of the reducing agent DTT. The individual DNA filaments, whose diameter is 2.7 nm, can be clearly seen, as can their organization into a hexagonally arranged liquid crystal; (c) and (d) show different orientations of DNA; a diffractogram is evident in (b) showing the hexagonal spacing of the DNA. (e) Shows a hydrated cryosection through purified nucleosome cores that have self-organized *in vitro* into precisely aligned columns of particles. Both micrographs courtesy of Jacques Dubochet.

vivo structure. Although in this method the specimen is dehydrated and embedded in plastic, a comparison of the data from hydrated cryosections and freeze-substitution has nevertheless convinced most specialists that the latter approach offers many advantages over the former, except at the highest levels of resolution.

Freeze-substitution is an old EM method that went out of favor only to emerge over the past decade as the best routine method for preparing sections of faithfully preserved cellular material, thanks especially to the pioneering work by the groups of Müller and Steinbrecht. It can provide not only relatively high quality structural preservation but also the opportunity to label the sections (in contrast to hydrated cryosections) using immunogold methods.

The fact that *high-pressure freezing*, but no other approach, allows vitrification of up to hundreds of μm of depth in the specimen now makes this approach the method of choice for freeze-substitution.

The principle behind freeze-substitution is to infiltrate the vitrified specimen with solvents at low temperature that, over a period of many hours to days, gradually replace the water with the solvent (often mixed with stabilizing "fixatives," such as OsO_4 (that cross-link lipids as the temperature is raised), or uranyl acetate (that can stabilize the head groups of phospholipids). The approach can also be carried out without any chemical fixatives. The idea here is to start the process of switching the specimen from a water-containing (vitreous) medium to a solvent that is compatible with the plastic embedding media at low temperatures (see below) while the structures of the biomolecules are still rigid. This is routinely done at, or colder than $-80\,^{\circ}C$. It cannot be ruled out that some cubic ice crystals may form under these conditions but even if they do they

are not noticeable at the levels of resolution usually desired.

In freeze-substitution, the tissue is either chemically fixed at low temperature and left to fix as the specimen is slowly warmed up, or left unfixed and embedded in specially developed methacrylate resins (such as Lowicryl HM23) at −70 to −90 °C, using UV polymerization. When labeling of antigens is not required, the specimen is warmed to room temperature, embedded in Epoxy resin, and then sectioned as in the conventional method. Figures 13 and 15 show images made using this approach. An alternative method that is not often used but may be appropriate for some specimens is to combine freeze-drying with freeze-substitution.

4.8 Freeze-fracture and Replica Methods

In the 1960s and early 1970s, freeze-fracture was a widely used approach, especially for the study of membranes. Although now restricted to a few specialist groups in the world, it is still the only method that can allow the investigator to see the insides of membranes. Small pieces of specimen are first vitrified, using a method such as slam-freezing against a cooled copper block or high-pressure freezing. Subsequently, the specimen is fractured, either with a knife, or by physically separating the frozen specimen that is sandwiched between two metal plates. The freshly opened surface is either evaporated with a metal replica directly, to reveal the surface details of the cut specimen, or the specimen is first allowed to warm up briefly to ∼−90 °C in order to etch away a layer of water. In this case, the replica also reveals some subsurface details; this is especially useful for visualizing details of the cytoskeleton, as evident in the pioneering studies of John Heuser, Tom Reese, Nobutaka Hirokawa and others. The underlying tissue is removed from the metal replica by harsh conditions such as immersion in chromic acid and the fine replica is attached to an EM grid and visualized by *ambient temperature* TEM. Thus, although this is a *bona fide* cryobased method, the replica of the frozen surface is finally visualized at room temperature.

A number of methods have been developed for combining freeze-fracture with immunolabeling. An especially interesting approach has already been developed by Kazushi Fujimoto; this method uses the detergent sodium dodecyl sulfate (SDS) to partially digest biological material from the replica while retaining integral membrane proteins whose exposed cytoplasmic domain are accessible for immunolabeling. Figure 16 shows two striking examples using this approach.

4.9 Simple Shadowing Methods

Since the dawn of the EM era, simple metal shadowing methods have proven useful as a substitute for negative staining that show particle surface structure. Any isolated particles, such as those suitable for the bare-grid cryo-EM negative-staining approaches can be usefully examined using this approach. When done well, the appearance of the structure can appear more three-dimensional than with negative staining. In the simplest instances robust structures can be air-dried before evaporating metal in a unidirectional, or rotary fashion. Better results are usually obtained after a more gentle preparation method, such as freeze-drying. Two examples are shown in Fig. 23. Another related method is glycerol spraying whereby the sample in glycerol is sprayed, then rotary shadowed and viewed

by TEM. This approach has been widely used to visualize isolated nucleic acids and attached complexes.

4.10 The Kleinschmidt and other EM Methods for Nucleic Acids

A useful method for visualizing DNA or RNA molecules, and attached complexes was developed by Kleinschmidt in 1959. The molecules are adsorbed to a thin, positively charged protein monolayer spread on a water surface. Cytochrome-C is the most commonly used spreading agent. The film is then picked up on a specimen grid, contrasted with stain (e.g. uranyl acetate) or by rotary shadowing with platinum, and observed in the microscope. This method may not have the resolution

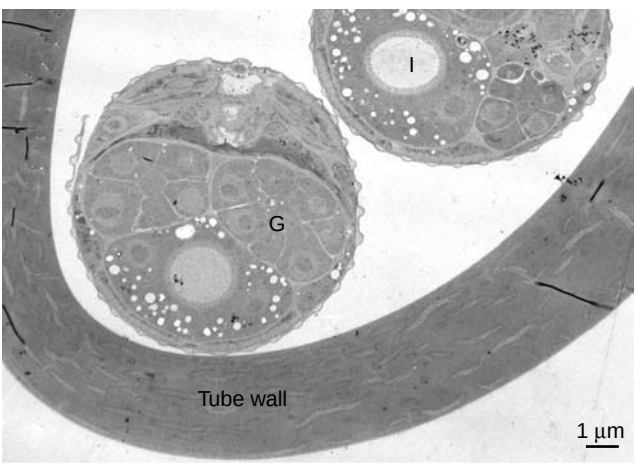

(a)

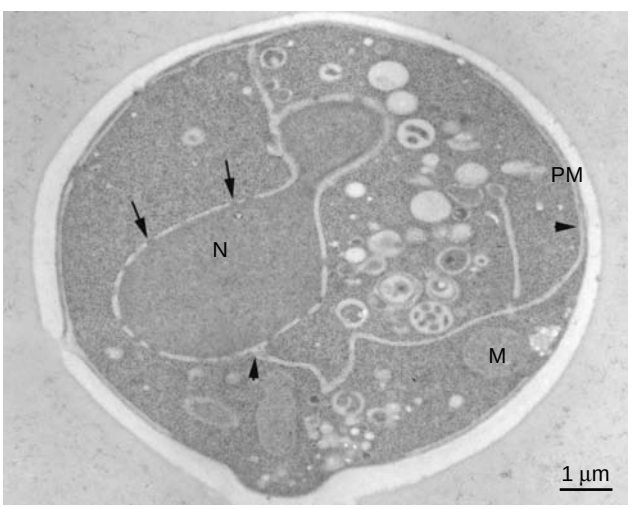

(b)

of the cryo-EM method but is more suitable for estimating parameters such as the length of the fibers.

About 10 years after the introduction of the Kleinschmidt method, the heteroduplex analysis for mapping regions of base sequence homology in nucleic acids was established, particularly in combination with formamide as a denaturating agent. The use of a detergent film (BAC) to replace cytochrome-C later allowed a better resolution of the nucleic acid strands.

4.11 EM Autoradiography

Until the 1970s, the use of EM autoradiography had a central place in cell biological research. In this approach, a radioactive molecule is introduced into the cells/tissues as a "pulse" of signal; the classical example was the use of radioactive amino acids or sugars that are incorporated into newly synthesized proteins or glycoproteins. By removing the radioactive precursor, a period of "chase" allows the tracer to be followed sequentially in the cell. This approach was used to follow the ER to the Golgi pathway for secretory proteins and glycoproteins by the groups of Palade and Leblond in the 1960s. At the end of the experiment, the cells/tissues are chemically fixed and embedded in epoxy resins and sectioned. Subsequently, a thin film of radioactivity- (and light-) sensitive photographic emulsion is layered on the section surface under a red light source in the dark room. The sections must then be incubated for days, weeks, or even months for enough radioactive encounters to interact with the film. At the end of the incubation, the film is developed and visualized under the EM.

Although in principle an excellent method to follow defined molecules, a significant disadvantage is the relatively low resolution, at best 0.5 μm. This, plus the long waiting period, has resulted in its almost total disappearance from the cell biological scene during the past 20 years. Cell biological problems that used to be analyzed by autoradiography are now better resolved by use of specific antibodies; kinetic results can then be achieved by use of specific inhibitors, such as cycloheximide, that can help to synchronize the synthesis of proteins, for example, through the biosynthetic pathway.

5 EM-Visualization at Ambient Temperatures

All EMs are rather sophisticated machines because electrons can only be usefully

Fig. 13 High-pressure freezing and freeze-substitution. Shows freeze-substituted specimens following high-pressure freezing. (a) Shows the nematode *Diploscapter coronata* having been vitrified within a capillary tube (the tube wall is indicated); the freeze-substitution was done using osmium in acetone. I-intestine; G, gonad (b) Shows a section through a high-pressure frozen and freeze-substituted yeast cell (*Saccharomyces cerevisiae*). The specimen was embedded in Lowicryl HM20 resin. Yeast cells are notoriously difficult to preserve by conventional fixation methods due to their robust cell wall. This image shows remarkable preservation. The arrowheads indicate direct continuities between the nuclear envelope and the endoplasmic reticulum. This ER, seen in negative contrast, is continuous with a domain of ER that seems to be very closely attached to the plasma membrane (arrowhead adjacent to the plasma membrane (PM)). The nuclear pores are also evident in negative contrast (arrows). Mitochondria (M) and other organelles are also evident. Both images courtesy of Heinz Schwarz.

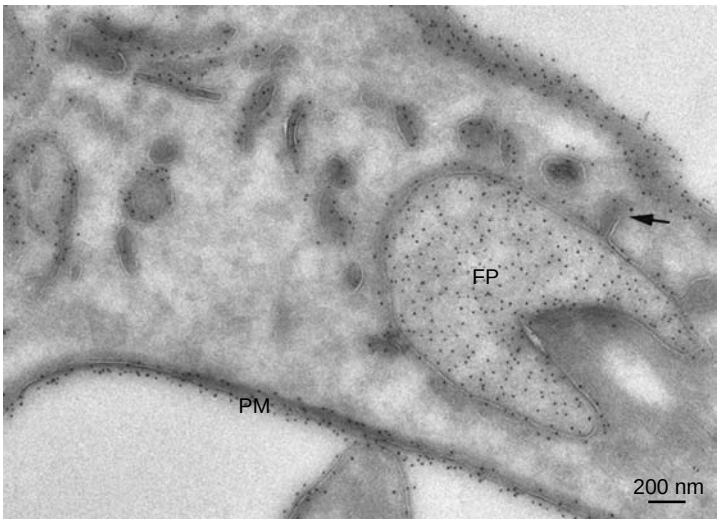

(a)

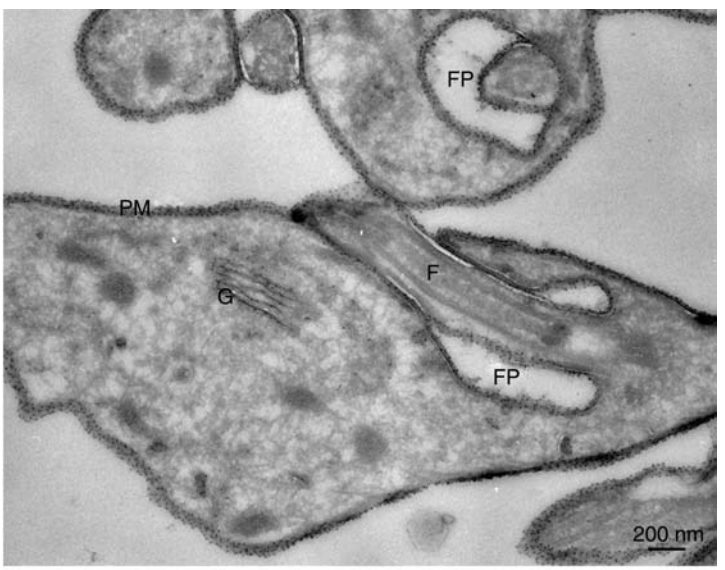

(b)

Fig. 14 Cryo-section immunolabeling of African trypanosomes. These figures show the trypanosome *Trypanasoma brucei*. (a) and (b) show Tokuyasu cryosections of this organism following labeling for the variable surface antigen (VSG). The VSG is seen to be heavily concentrated at the PM, the flagellar pocket (FP, the flagellum is indicated by F), throughout the Golgi complex (G), as well as in ER and endocytic organelles (not indicated). The arrow in (a) shows a labeled endocytic vesicle in the process of budding from the flagellar pocket membrane. Bars–300 nm. Courtesy: see Fig. 15.

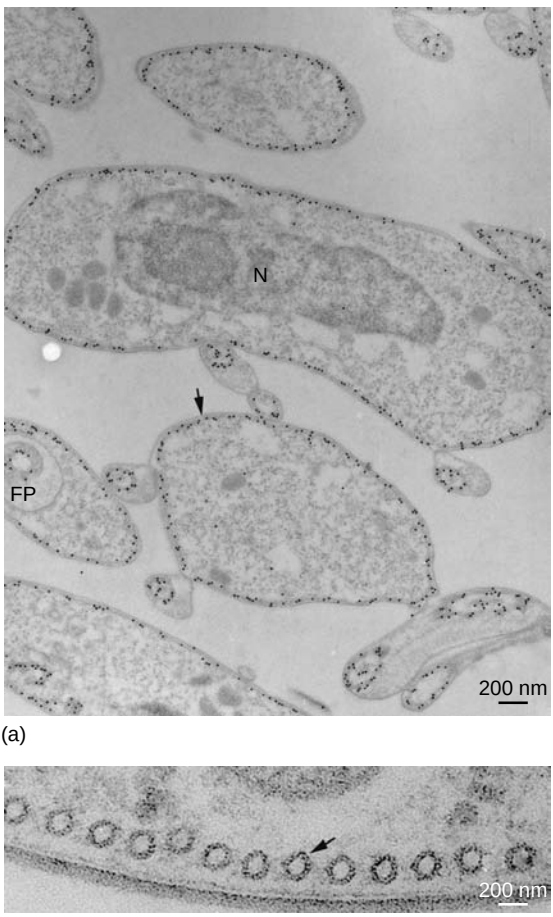

Fig. 15 Plastic-section immunolabeling of African trypanosomes. These figures show the trypanosome *Trypanasoma brucei*. (a) shows a trypanosome that was fixed in 0.05% glutaraldehyde/2% formaldehyde before being embedded in Lowicryl HM 20 after the progressive lowering of temperature method. The sections were labeled with anti-tubulin and donkey anti-mouse gold (12 nm). The array of microtubules beneath the plasma membrane and in the flagellum are strongly labeled (arrow). (b) Shows a trypanosome that was vitrified by high-pressure freezing followed by freeze-substitution in 0.5% osmium tetroxide, 0.5% gallic acid in acetone and room-temperature embedding in Epon. The excellent preservation of the microtubules is evident by the clear delineation of the tubulin protofilaments in cross-section. Micrographs courtesy of Christoph Grünfelder, Peter Overath, and Heinz Schwarz.

transmitted in a rather high vacuum (better than 10^{-3} Pa $\sim 10^{-4}/10^{-5}$ Torr). This high vacuum in the main column also needs to be maintained when a grid is introduced. For this, the grid is first introduced into a low volume prechamber (airlock) that is quickly pumped to vacuum. The grid can then be carefully inserted into the center of the objective (electromagnetic) lens. At the top of the microscope column is an electron source (tungsten filament, lanthanum hexaboride crystal) or, the (most coherent and expensive) field-emission filament. The specimen is observed at acceleration voltages ranging from 80 to 120 kV (standard) to 200 to 400 kV (new ''intermediate-voltage'' microscopes) or up to 1 million volts in the high-voltage EM (of which only a small number are available for biological research worldwide; the others have disappeared due to lack of interest). The recent developments in EM tomography will likely see a reversal of this trend, as methods to investigate thicker sections become more useful (see below).

The emitted electrons are attracted toward the anode leading to a coherent

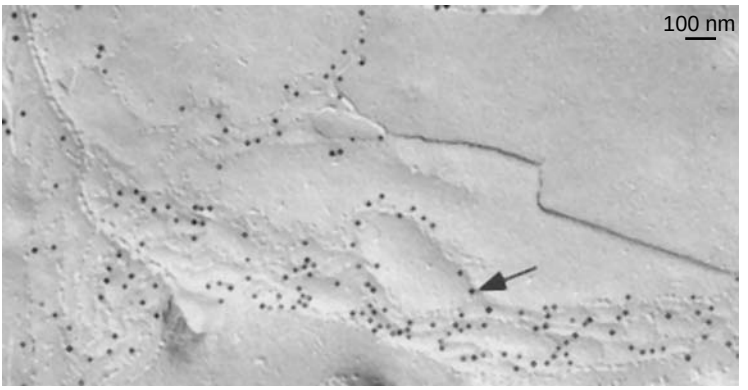

(a)

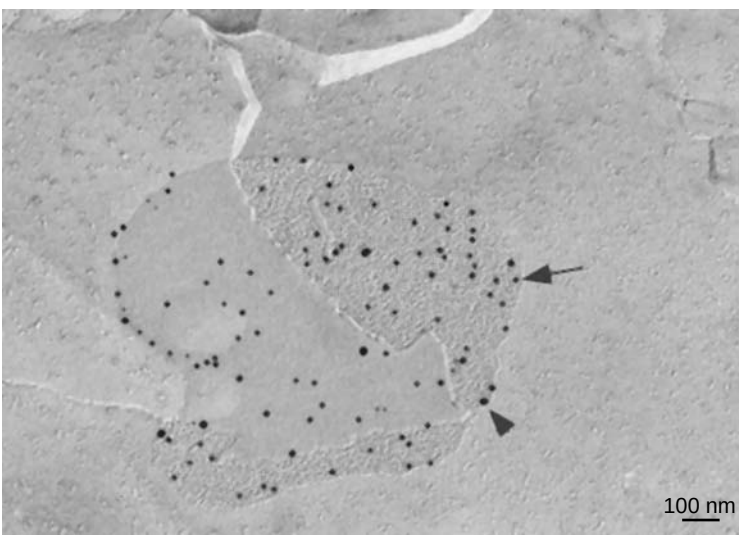

(b)

Fig. 16 Freeze-fracture immunolabeling. (a) Shows chick liver, tight junctions labeled for the tight junction protein occludin while (b) indicates rat liver, gap junctions double-labeled for connexin 32 (large gold particles), and connexin 26 (small particle). The tissue slices were quick-frozen by contact with a copper block cooled with liquid helium. The frozen samples were fractured in a Balzers BAF 400 T freeze-etch unit at −110 °C, replicated by deposition of platinum/carbon (Pt/C) followed by carbon. After thawing and washing with PBS, the pieces of Pt/C replica were transferred to 2.5% SDS containing-buffer for 12 h at room temperature. Subsequently, the replicas were immunogold labeled, rinsed, fixed with glutaraldehyde, rinsed again, and picked up on grids. Micrographs courtesy of Kazushi Fujimoto.

beam of electrons that, with the help of an electromagnetic condenser lens, can then be converged onto the specimen. Electron-dense material (biological material heavy metal stains) scatter the electrons, whereas the electron-lucent parts of the specimen allow the electrons to pass down the column where, after magnification by the objective and projector lenses, they form an image on a phosphorescent screen, film or CCD camera. There are two different types of interactions between electrons and biological material; *elastically scattered electrons* interact with the specimen with a minimum loss of energy and therefore there is no change in the wavelength of the electrons. This type of scattering is due to a simple deviation of the electrons in the atoms in the specimen. In contrast, *inelastically scattered electrons* are those that experience energy loss and change of wavelength during their interactions with the electrons of the encountered atoms. These interactions give rise to the secondary events, such as heating the specimen, mass loss, contrast, X-ray emissions and cathodoluminescence.

When electrons interact with the phosphorescent screen, the green light emitted by fluorescence provides a signal that the human eye sees and the brain needs to interpret. By removing the phosphorescent screen transiently, the electrons can be directed toward photographic emulsion to record the image on film. Cooled CCD cameras are now rapidly replacing film but they are very expensive and still do not have the resolution available on film. It is likely to be only a matter of time before CCD cameras replace film completely.

The inelastically scattered electrons passing through the specimen can damage the specimen, resulting, for example, in *mass loss*, even in plastic sections. However, in practice, beam damage is not a serious problem for plastic or Tokuyasu cryosections, at the usual level of resolution needed. For vitrified, native specimens this is a very serious concern, as we will now discuss.

6
EM at Cold Temperatures

Imaging vitreous specimens, either bulk suspension, or hydrated cryosections, is a lot more difficult than imaging specimens at ambient temperatures. This is really not something that can be taught in a manual. Moreover, one has to deal with quite delicate and expensive equipment. A competent cryo-EM specialist needs to understand a lot about the physics of the processes involved. Native biological sections are extremely sensitive to the electron beam, and under the conditions used for ambient temperature specimens, they visibly boil away within seconds. This necessitates *low-dose imaging*. In the standard approach, the investigator can only surmise that suitable specimens are present at a site selected at low magnification, that is, immediately photographed at higher magnification using the minimum electron dose required for imaging. All focusing and image adjustments are therefore done at sites away from the selected areas. In recent times, the technology available with modern EMs has improved considerably, and is now much more user-friendly.

In most cases, the specimen is unstained and to acquire sufficient contrast to see the details is often a problem. For optimal images, one takes advantage of the increase in phase-contrast that is provided by underfocusing.

Nevertheless, the fact remains that more and more specialists are now routinely recording images of very high information

content with this technology. There is no doubt that at the top level the cryoimaging approach provides the highest resolution available in EM today (see Fig. 11). In suitable specimens, by averaging data from large numbers of micrographs, it can now approach (or reach) the atomic level of resolution (0.3–1 nm) when two-dimensional crystals are available. In many cases where a 1 to 3-nm resolution is obtained, the level of information often suffices to help one to fit higher resolution X-ray crystallographic data of the same structure into a comprehensive model of the structure. In this way, the details of the (high-resolution) X-ray structure are fitted onto its overall shape, which emerges from the (lower resolution) cryo-EM model.

It is important to realize that in order to "solve" a structure to the desired level of resolution, the availability of negatives or online images is merely the beginning of a long, complex, and often tortuous process. In contrast to ambient temperature specimens in which the images are mostly "final," in cryo-EM (as well as in some high-resolution negative stained specimen analysis), there is an increasing number of possibilities to obtain much more relevant structural information by use of computational *image processing* algorithms. The goal of these procedures is to improve signal-to-noise ratio by averaging a large number of structural units. The two main kinds of approach to obtain this goal are the following: (1) *Electron crystallography* (including helical reconstructions) of two-dimensional crystals and (2) *Single particle analysis*, including icosahedral particles, where particles imaged at various orientations are collected, averaged and superimposed into a three-dimensional model.

Examples of cryoimages obtained using sections are shown in Fig. 12 and from whole-mounts of particles in Fig. 11 and 19(a) and of cells in Fig. 19(b).

6.1 EM Tomography

In the past, if one wanted a three-dimensional view of a structure from thin sections it was necessary to use serial sections. The need to cut and maintain in perfect order up to a hundred or more thin sections is technically quite demanding and, until recently it was very difficult to subsequently align images from each section on top of each other in perfect register. In the past few years, a revolutionary alternative approach has started to have a major impact on EM at all levels. From either a single particle or a relatively thick section, *tomography* allows one to obtain a three-dimensional reconstruction of the structure of interest. In practice, a device for tilting the specimen, the goniometer stage, allows the grid to be tilted relative to the beam (under standard conditions + and $-\sim 70^\circ$; new goniometers are now reaching $\pm 80^\circ$). Images of all the projections are collected, usually at intervals of 1 to 2°. By "back-projecting" these images into a three-dimensional volume one can construct the original image in three dimension using now standard computational algorithms. The reader is also referred to a whole volume of the Journal of Structural Biology that was dedicated to the exciting developments in this field (Vol. 138, pp. 1–155, 2002).

There are a number of problems that can limit this approach. First, since there are a large number of images, the effects of beam damage have to be considered. In practice, this is not a serious problem for the "ambient temperature specimens" indicated above: one can take 130 to 140

images, 1° apart. However, for vitrified particles or sections, this problem is still a limitation; in practice one is currently restricted to, at most 100 images before the specimen is irreparably damaged. It should be emphasized that a great deal of effort has been made by a few specialists to make this technology available, with important developments in software, automatic image collection and in specimen preparation.

The second problem is the "missing cone" or wedge due to the practical limitation that one cannot tilt the grid ±90° – there will always be a part of the information that is missed, resulting in loss of resolution. For quasi-regular particles, one can overcome this problem by averaging a number of reconstructions from different viewpoints. For imaging complex structures in sections of cells, it may be possible to combine tomograms from sections that contain the specimen at different orientations.

In addition to being used in high-resolution EM for reconstructing single particles such as ribosomes this approach, in combination with high-pressure freezing and freeze-substitution, is now being applied to analyze the three-dimensional organization of different structures, such as whole yeast cells, the Golgi complex, mitochondria (see Fig. 18). Most of these studies have used relatively thick (−200 −300 nm) plastic sections. It should be noted that for the difficult question of deciding whether or not fine tubules connect certain structures (notably the Golgi complex), this approach can still leave open considerable room for subjectivity. In such cases, the use of heavy metal stains (e.g. the cytochemical reaction product of HRP, or other enzyme reaction products) to fill the lumen of tubules (e.g. by expressing HRP-containing constructs in cells) can greatly aid the analysis.

The ability to visualize *whole cells* directly by a transmission cryo-EM approach can be considered the ultimate goal in electron microscopy. It is now fascinating to consider that, prior to the development of robust methods for preparing thin resin sections of cells, the famous trio of Porter, Claude, and Fullam had made initial attempts in 1945 to visualize cells grown on EM-grids (after osmium tetroxide vapor or formaldehyde primary fixation). Despite the harshness of the preparation (by today's standards), what was observed was striking. In the thinner, peripheral regions of the fibroblast, whose outlines are clearly delineated, one can clearly see mitochondria, filaments and so on (see Fig. 17).

In an important development, the Baumeister group have succeeded in bringing together a whole battery of technical developments to enable the experiment to be redone with state of cryo-EM tomography. In their first publication, they focused on the amoeba, *Dictyostelium,* that in accompanying videos could be seen to move rapidly across the grid just before vitrification. Although one has to accept the (still significant) limitation that only cell projections thinner than ~400 to 500 nm could be clearly analyzed, the details shown are amazing, at a resolution of ~5 nm. It must be realized that this is the first time it has been possible to look directly inside a relatively unperturbed cell at this level of resolution. An example is given in Fig. 19(b).

With more developments, and more widespread use of this approach, (likely aided by higher accelerating voltages and energy filtering) thicker regions of the cell will become more accessible for analysis. The use of hydrated cryosections

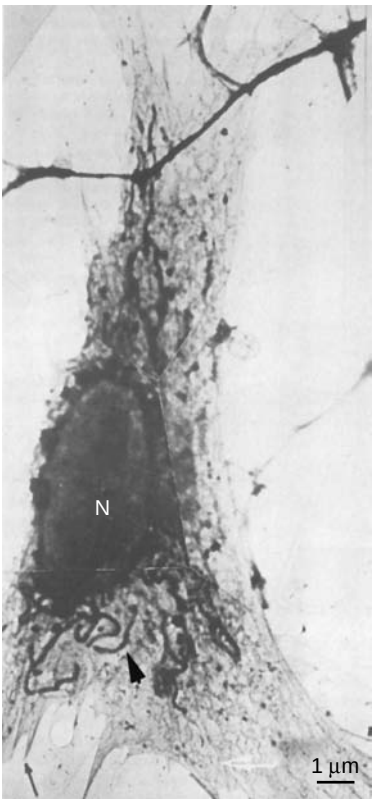

Fig. 17 One of the first micrographs of cells. Shows a famous micrograph taken from the paper by Porter, Claude, and Fullam from the Journal of Experimental Medicine (1945). This micrograph is one of the first ever taken by electron microscopy. The cells were grown on an EM grid, fixed in osmium tetroxide, air-dried, and visualized as a whole mount. Micro spikes (arrow) at the leading edge are evident, as is the nucleus and the worm-shaped mitochondria. A fine reticulum beneath the mitochondria is likely to be the endoplasmic reticulum. This image has already been shown in a wonderful book by Christian de Duve in 1984.

can be expected to complement these efforts. The ability to reconstruct detailed structures (e.g. within the nucleus) will greatly increase the level of information one can obtain from these preparations. Tomography can also greatly simplify the interpretation of the three-dimensional structure from hydrated cryosections.

6.2 Low- and High-resolution SEM

The ability to visualize surface structures by scanning EM is well known, even to many nonscientists, a result of spectacular images of animals, plants, and microorganisms obtained by this approach. Many specimens, such as insects have robust exoskeletons that can withstand beam damage and only a thin layer of metal needs to be evaporated before visualization; for a beautiful example see Fig. 20. This low-resolution SEM approach method is widely used. At a higher level of resolution, parts of the organ of Corti of the ear are seen in Fig. 21.

Over the past 20 years, the use of higher resolution SEM in molecular cell biology has dwindled – only a handful of laboratories currently use this approach. This is indeed surprising, and very disappointing given the proven ability of this approach to see high-resolution details of structures such as nuclear pores, bacteriophages (Fig. 22(a, b)) and actin filaments (Fig. 22c).

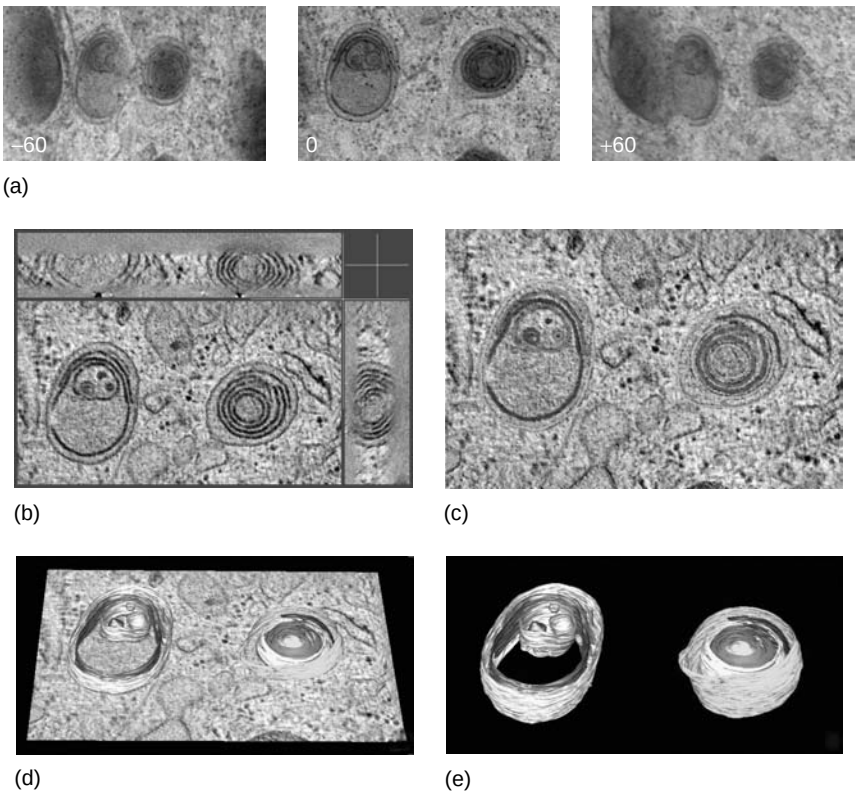

Fig. 18 EM Tomography from plastic thick sections. (a) To obtain a three-dimensional reconstruction of an organelle, a tilted-series of images is recorded. A 250-nm section of multilamellar lysosomes in a mouse dendritic cell is shown after high-pressure freezing, freeze-substitution, and resin embedding. This, relatively thick section was tilted from −60° to 60° along two perpendicular axes to produce a tilt-series of 242 images. This panel shows three of these images, collected at −60, 0, and 60° tilting. The small black dots in the images are 10-nm gold beads on top of the section, which are used as fiducial markers (reference structures) for the three-dimensional reconstruction. (b) After calculating the three-dimensional reconstruction from the tilt-series, the three-dimensional volume can be analyzed by displaying it as thin, so-called tomographic slices. The slices can be made along every possible axis. Three slices through the tomographic volume are shown: top: X-Z, middle: X-Y and right: Y-Z. (c) By manually tracing the membranes of the multilamellar lysosomes in the tomographic volume a three-dimensional model is created. (d) Shows a three-dimensional model top view of one of the tomographic slices, while (e) reveals the three-dimensional model view of the endocytic vesicle. The figure was provided by Jean-Luc Murk and Bruno Humbel. From a study by Murk et al., J. Microsc. in press.

In conjunction with immunogold labeling prior to vitrification, the use of cryo-SEM, especially the remarkable new generation of field-emission (FE) scanning electron microscopes offers the molecular cell biologist a complementary approach to tomography. SEM visualizes the surface of the structure directly in three dimension. In conjunction with the pioneering efforts of a few specialist groups (especially those

of Martin Müller, Paul Walther, Hans Ris and Terry Allen) in developing new approaches for specimen preparation, this method is now capable of extremely high-resolution (1–3 nm) (Fig. 22). A number of publications have shown the great potential of this approach, also for immunogold labeling, especially when gold particles as small as 1 nm are used, that can be visualized in the backscatter mode. In some studies, Fab domains have been used instead of whole IgG molecules to increase the precision of the labeling.

6.3 Critical Point Drying

The most common method to prepare living specimens for SEM has been chemical fixation followed by critical point

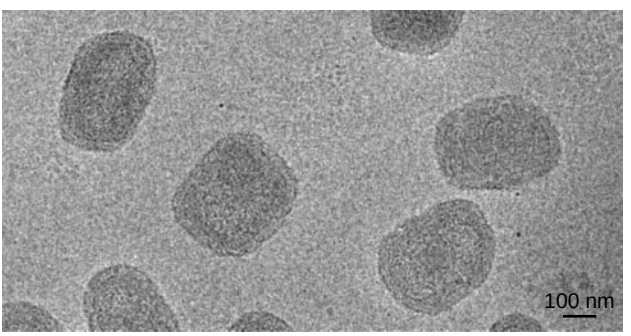

(a)

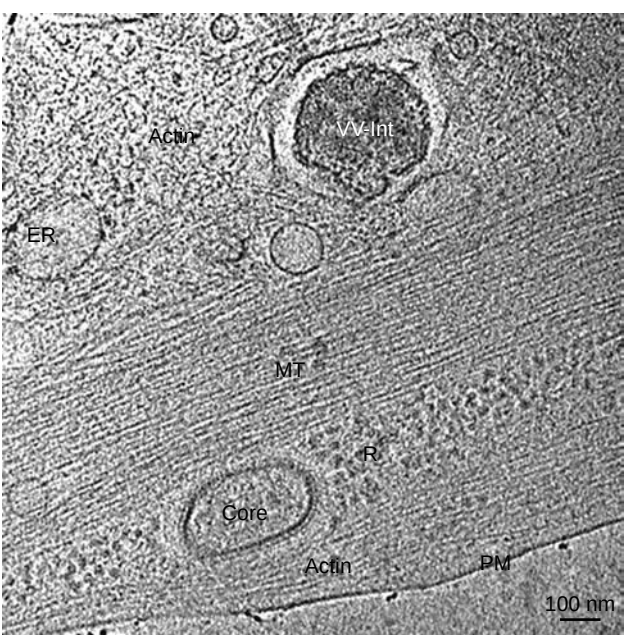

(b)

Fig. 20 Classical SEM. Shows a beautiful image of *Drosophila melanogaster* visualized by conventional SEM. The insect was fixed with glutaraldehyde, rinsed, critically – point-dried and sputter coated using gold-palladium. Fine details of the surface of the insect can be seen in artistic detail. Micrograph courtesy of Jürgen Berger (Max Planck Institute for Developmental Biology, Tuebingen, Germany).

drying (CPD). The basic idea of this approach is to prevent drying artifacts, which occurs when the sample is exposed to the surface tension of the liquid–gaseous phase border. This method has also been extensively used for preparing material for TEM replicas and for epoxy resin embedding. For this, the specimen is chemically fixed and dehydrated in solvents such as ethanol (essentially as for epoxy resin embedding). Later the solvent is replaced by liquid carbon dioxide in the pressure chamber of the critical point drying device. Then, the carbon dioxide is slowly warmed up. This causes the pressure to rise in the closed system. When the temperature and pressure pass above the so-called "*critical point*" of carbon dioxide (31 °C and 74 bar), the pressure is carefully reduced and the sample removed once room pressure is reached. The specimen is thus dried without ever having seen the liquid–gas phase border, which can cause major distortions to a sensitive specimen. Recent applications of the CPD method in cell biology are the preparation of cytoskeleton elements after detergent treatment and TEM replica. As pointed out by Ris, it is very important to remove every trace of water from the specimen before the next stage, which

Fig. 19 Cryo-EM and tomography. (a) Shows cryo-EM of vitrified vaccinia virus using the bare-grid method. This virus is enormously complex as was already evident in Fig 4. Although a tremendous amount of information is present in these particles, at present the detailed three–dimensional organization of the virus remains to be deciphered and the tomographic approach is now in progress. (b) Shows the 2002 equivalent of the experiment already carried out. Hela cells were grown on EM-grids and vaccinia virus was added. The goal of this experiment was to visualize intermediates in the entry of the virus into the cell, a highly complex and not understood process. The grid was vitrified and examined by cryo-EM. This image is a section through a tomogram (not a physical section) showing details that are not approachable by single projections. The upper particle is, we believe, an intermediate (from VV-Int) in the entry of the viral core and is likely to be still outside the cell. In contrast, the viral core, which is transcriptionally competent, having lost its outer membranes is definitely within the cytoplasm. Actin filaments are seen in the lower part of the micrograph, whereas the filaments above the core are microtubules. Ribosomes (R) are also evident, as is the ER. This method is the first approach available for looking inside physically and chemically unperturbed cells by EM at this level of resolution. Bars = 100 nm. Both micrographs courtesy of Marek Cyrklaff; specimen provided by Jacomine Krijnse Locker.

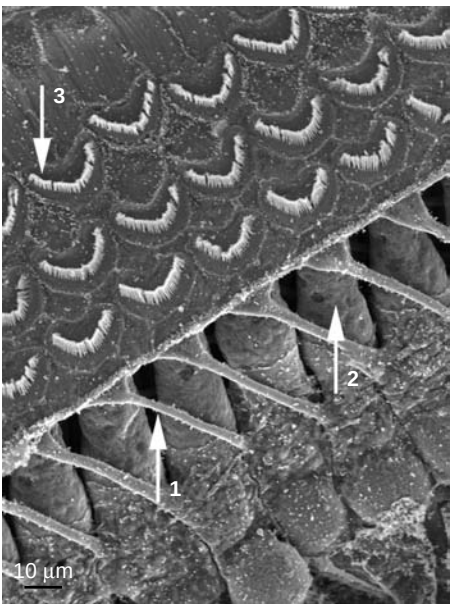

Fig. 21 SEM of the organ of Corti of the mouse ear. These SEM images show a guinea pig cochlea dissected to remove Hensen cells (sensory cells of the ear and the supporting cells). Hensen cells were dissected to give a clear view of the outer hair cells, the sensory cells of the ear (2) and supporting cells. The tops of the outer hair cells have stereocilia attached (3) the chevron-shaped projections) and the main body of the outer hair cell can be seen below. (1) indicates the outer phalangeal cells of Deiter. The specimen was fixed in glutaraldehyde before dissection and then infiltrated with tannic acid and osmium tetroxide. This facilitated examination without metal coating in a field-emission SEM (a Philips XL-30 FESEM operating at 5 kV). This specimen preparation method was a modification of the method of Jongebloed et al. (1999). Micrograph courtesy of Paul Webster.

for SEM involves coating the specimen with a thin layer of metal (see Fig. 23).

6.4 Freeze-drying and Cryo-SEM

The SEM is able to image the physical surface only. In most biological samples, however, the structures of interest are inside cells and tissue. To get access to the structures of interest, the specimen has to be opened by cracking or by sectioning, or the overlaying water must be removed. This is only possible with a solid-state sample. The most straightforward way to solidify a biological sample is by vitrification. In the cryo-SEM, bulk samples can be investigated over a very large range of magnification, from the resolution level of a dissection microscope up to macromolecular resolution. After being immobilized, by cryofixation the samples can be investigated in the cryo-SEM in the fully frozen-hydrated or in the partially freeze-dried ("deep-etched") state. Many artifacts of conventional chemical fixation and dehydration techniques can, thereby, be prevented.

Removal of the vitrified water is possible by sublimation in the vacuum in the microscope or in a preparation chamber. This preparation step is also called "*freeze-etching*" with a partial sublimation of the water, or freeze-drying in the cryo-EM literature. When the water is partially removed, the three-dimensional arrangement of the structures in the cytosol can be investigated. Current methodological research focuses on a better control of the ice-sublimation process.

This method also has a high, and still not fully used potential for the analysis of *in vitro* systems in cell biology. Together with Paul Walther, our group is currently investigating actin-membrane interactions using these cryo-SEM approaches.

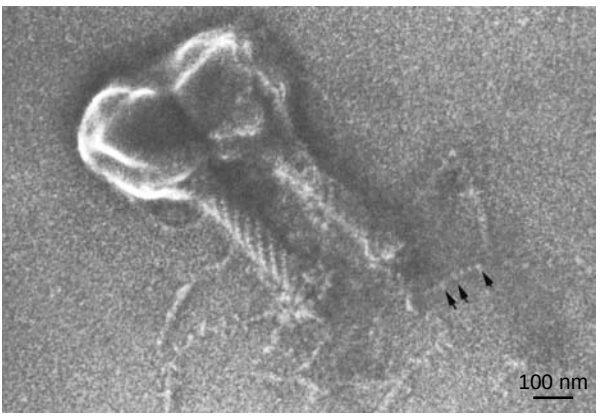

(a)

(b)

(c)

Fig. 22 High-resolution cryoscanning EM. (a) and (b) show the T-even bacteriophage Tull * —46 after freeze-drying, metal coating and visualization by cryofield emission SEM. (a) shows an intact particle, whereas (b) is a particle that became disrupted by spesimen preparation. (a) is additionally labelled with an Fab directed towards the tail protein Y of the phage. Individual Fab molecules can be seen decorating pieces of the isolated tail (small arrows) as well as the helical ribbon of the tail protein. In (b) following disruption, the DNA of the phage is clearly evident (large arrow), as are many other structural details of the inner parts of the virus. (c) shows high-resolution image of an actin filament that had been freeze-dried and coated with tantalum-tungsten. In (a) and (b) by courtesy of Rene Herman, Heinz Schwarz, and Martin Mueller. (c) Courtesy of Paul Walther.

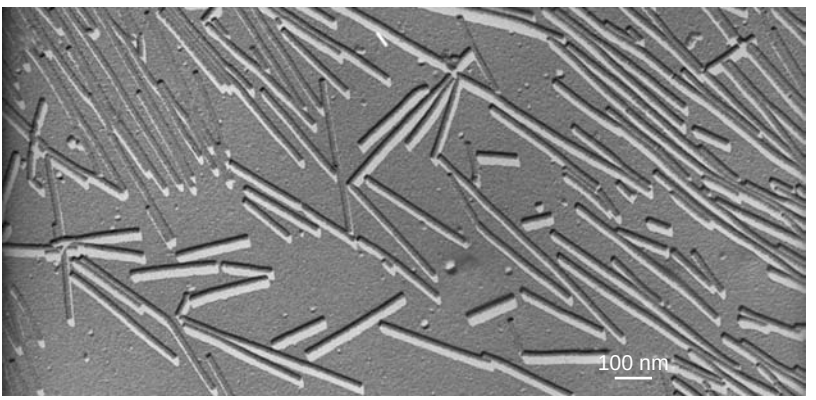

(a)

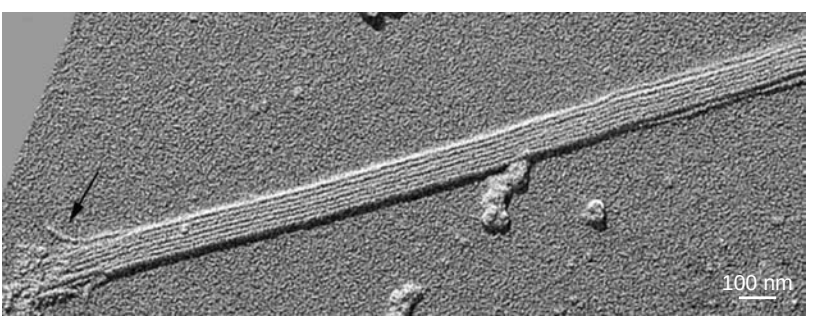

(b)

(c)

6.5 Coating Techniques for Cryo-SEM

Samples for high-resolution SEM are usually coated with a thin metal layer in order to enhance electrical conductivity and to localize the signal on the surface of the sample. Charging of the sample can also be reduced by working at low accelerating voltage of the primary beam (e.g. 2 kV). It was, however, pointed out, that the excitation volume of the secondary electrons is increased when they interact with an insulator. Therefore, it is advantageous to work with coated (shadowed), conductive samples, even when working at low voltages. Relatively thin samples can ideally be supported by a uniform layer of an electrically conductive material, for example, by 1 nm of tungsten. Bulk cryofractured samples, however, are insulators and very beam-sensitive. For these samples, double layer coating has been very useful. The sample is first coated with a contrast-forming layer of heavy metal (platinum or tungsten), both to enhance electrical conductivity and for mechanical stabilization. A carbon layer with a thickness of 5 to 10 nm is then applied on top of the platinum layer. As shown in different studies, this layer drastically reduces the effects of beam damage. For the analysis of these samples, the backscattered electron signal has to be used, which is mainly produced by the metal coat that stays in direct contact with the biological structure of interest.

6.6 Scanning Transmission EM

A specialized variation of transmission EM is scanning transmission electron microscopy (STEM, done mostly in the dark-field mode) in which the specimen on the grid is scanned in a raster pattern with a finely focused electron beam. This offers precise determination of the absolute mass of the specimen irrespective of shape.

6.7 X ray Elemental Microanalysis by EM

X-ray microanalysis of biological specimens can be conducted at two different levels of resolution. Low-resolution microanalysis is typically performed using a scanning electron microscope equipped

Fig. 23 Metal shadowing for EM. (a) Shows the identical TMV particles that were seen by negative staining in Fig. 3. This unpublished image, over 30 years old, was unidirectionally shadowed with platinum. (b) Unidirectional surface shadowing (tantalum/tungsten, elevation angle −45°) of a microtubule stabilized with small amounts of tau-protein. This protein is essentially invisible due to its highly filamentous structure. This leaves a surface pattern typical for microtubules, exhibiting the outer protofilament surface as a continuous axially oriented rim. The arrow shows a protofilament splayed out at the end of the microtubule. (c) Unidirectional surface shadowing of tubulin that coassembled with walls of monomeric kinesin molecules *in vitro*. The presence of these molecules renders the outer surface in a completely different way relative to the undecorated microtubules in (b). Instead of an axial striation, one now observes a perpendicular striation with an axial 8-nm repeat (arrows), which corresponds to one motor domain bound to each α, β-tubulin dimer. This method allows one to clearly distinguish between inner and outer microtubule surface. The inner surface exhibits a 4-nm repeating pattern according to each tubulin monomer. (a) Courtesy of Herman Frank and Heinz Schwarz; (b) and (c) courtesy of Andy Hoenger.

with an X-ray detector. This application is typically used to analyze small particulate inclusions in cells and tissues. The analysis of environmental heavy metal contamination, foreign body ingestion, and occupational exposure are the most typical applications. Examination of cells and tissues in the scanning electron microscope using the backscatter electron imaging mode can also be used to provide low-resolution information, for example, inclusions and particles containing high atomic number atoms can be identified in association with particular cell types.

X-ray microanalysis of biological specimens can also utilize a transmission electron microscope coupled with an X-ray detector, but the demands of specimen preparation and specimen handling are much more challenging. Typically, this high-resolution technique requires the rapid freezing of small, undamaged biological specimens. The biological specimen requirement involves selecting small (less than 1 mm) specimens, which must be solidified using ultra rapid freezing techniques leading to vitrification. Subsequently, the small specimens are mounted in a cryoultramicrotome using either mechanical clamping or a low temperature cryoglue. The ultrathin sections with thicknesses of about 100 to 120 nm are prepared and transferred to formvar-coated double-folding (sandwich) grids in the cryochamber of the cryoultramicrotome. The grid is next transferred to a cryotransfer stage at −170 °C. The cryotransfer stage is inserted into the transmission electron microscope. The transmission electron microscope must be equipped with a liquid nitrogen cooled anticontamination device to prevent the deposition of water vapor onto the specimen from the column of the electron microscope.

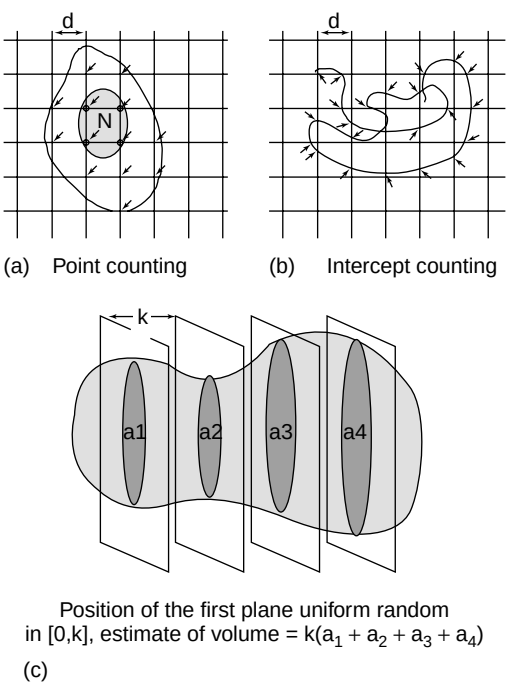

(a) Point counting (b) Intercept counting

Position of the first plane uniform random in [0,k], estimate of volume = $k(a_1 + a_2 + a_3 + a_4)$

(c)

For X-ray microanalysis, the specimen must be warmed to about −100 °C, which allows the water in the specimen to sublime onto the anti contaminator. This freeze-drying step generally requires 45 to 60 min to achieve a specimen that will withstand the high-beam currents necessary for high-resolution X-ray microanalysis. Using this technique allows the analysis of subcellular regions in individual cells such as mitochondria, lysosomes, cytoplasm regions, and nuclear regions.

The critical assumptions in this technique are (1) the freezing process does not translocate diffusible elements from their native location; (2) that the cryosectioning process and freeze-drying of the ultrathin section in the electron microscope also do not translocate diffusible elements from their native sites, and (3) that during analysis under the high electron beam current the diffusible elements also remained in their native location. In practice, these criteria are rarely fulfilled for small, mobile ions of cell biological interest, such as calcium or magnesium. In general, these technical difficulties have limited the applications of this technique in cell biology.

An additional limitation is the lack of sensitivity to small changes in concentrations of sodium, chlorine, potassium, and calcium in the small subregions of cells. Under optimal conditions using freeze-dried thin sections, the limit of detection for X-ray analysis in general is about 100 parts per million. It is very difficult to estimate the water content of the subcellular regions so that the dry weight concentration measurement can be converted to a wet weight concentration that is meaningful to biologists.

Fig. 24 Principles of Stereology. (a) Shows the principle of estimating area using point counts. A transparent lattice grid is randomly positioned over the images of a structure, in this case, a hypothetical profile through a cell, the nucleus N is shown. The point where the edges of the two lines of the lattice meet each other is considered to be a unique point (arrows). One simply counts the number of points over the structure of interest. In this particular example, there are 13 points in total over the cell, four of which follow the nucleus. If this result were seen after averaging over a number of micrographs, this means that the nucleus represents 4/13 = 30.7% of the cell volume. The number of points $P \times d^2$ provides an estimate of the area of each structure on the micrographs. (b) Shows the principle of intercept counting to estimate the boundary length of surface. In this example, a square lattice grid with random orientation and position is placed over a piece of string. The arrows indicate the positions at which the lines of the string transect the lines of the grid: These positions (formally defined by the intersection of the edges of the line and one edge of the string) are referred to as intercepts. The length of the line is then given by $\pi/4 \times I \times d$ where I is the number of intercepts and d is the distance between the lines of the lattice grid. This principle, first introduced by Buffon, is an unbiased estimator of the length of the profile, and can be converted to surface area in three dimensions. In this example, if $d = 1$ cm and there are 21 intercepts, then the length of the string is $\pi(3.142)/4 \times 21(\text{Intercepts}) \times 1\text{ cm} = 16.495$ cm. (c) Shows the principle of the Cavalieri method. The principle here is to section a three-dimensional object with a series of equally spaced sections such that the position of the first section is random within the interval between the sections. In this example, four sections are made, a1–a4, and the volume is then given by the sum of the areas of the object displayed on the sections multiplied by the distance (k) between the sections (section thickness). This method is justifiably considered to be a powerful method for stereology, and has been widely used for many different organs, tissues, cells, and organelles. Diagrams courtesy of John Lucocq.

Theoretically, the water content could be estimated by making measurements before and after freeze-drying; however, fully hydrated specimens in the electron microscope are intolerant of the high-beam currents that are necessary for X-ray microanalysis.

In summary, the most useful methods involving X-ray microanalysis at present are performed in the scanning electron microscope at low resolution for studying occupational exposure, mineralized tissue, and calcification associated with disease processes.

6.8 Energy-filtering Transmission Electron Microscopy (EFTEM)

This approach allows improved imaging of thick sections and frozen-hydrated specimens by removing inelastically scattered electrons. The amount of inelastic scattering increases with specimen thickness (in general quite rapidly) and is dependent on the incident electron energy and on the material the beam is interacting with (in general, the heavier the material, the stronger the interaction). In EFTEM, the electrons are separated according to their energies, making it possible to filter out the blurring effects seen with thick sections, as well as retrieve selected chemical or physical information from the specimen. This additional electron selection results in contrast enhancement for all imaging modes and also provides the possibility of selecting electrons with specific scattering effects for imaging, thereby generating object- or element-specific contrast. For more details as provided by one of the manufacturers of this technology, see (www.leo-em.co.uk/temproducts/principle.htm).

7 Stereology

Although many cell biologists have yet to realize it, a remarkable set of methods has been developed over the past 30 years for quantifying the sizes, contents, and spatial relationships of biological structures. These methods are known as *Stereology*, a rigorous, probability-based discipline for estimating three-dimensional structural quantities (including volume, surface area, length, and number) and spatial relationships from sectional or slice images. The methods are design-based, meaning that they rely on the principle of random sampling (especially systematic random sampling) and offer the twin benefits of no (or minimal) bias combined with high precision. The work needed to implement them is, surprisingly, less than generally expected, so these methods tend also to be highly efficient. The use of stereological methods will be outlined, as will recently developed approaches for quantifying immunogold particles that are used to detect specific antigens.

The *sine qua non* of stereology is the application of geometrical probes for estimating two-dimensional and three-dimensional quantities. Crucial for this goal is unbiased sampling at all levels of specimen preparation and imaging: each specimen (e.g. an organ), and each part of that specimen (e.g. its different tissue compartments), must have an equal chance of being sampled. For some quantities, it is necessary to randomize the orientation of sectioning as well as section position. A number of books and chapters cover these important aspects in great detail. In order to give the flavor of how stereology works, we shall consider one method available for determining the absolute volume of any structure and

two standard methods for determining relative surface area (or length of sectional membrane image) and relative volume (or profile area of sectioned structures). It will be seen that, once the absolute volume of a reference structure (e.g. the nucleus or the cell) is known, it is possible to calculate the absolute total surface and volume of any compartment in the cell using their estimated relative surfaces and volumes. It is equally easy to obtain higher-level quantitative information about cells in tissues, organs, or whole organisms.

7.1 Relative Area and Volume

In order to estimate the areas of the sectional images (profiles) of different structures on micrographs, transparent plastic sheets having a uniform array of "test points" are randomly superimposed over the image (see Fig. 24(a)). The positioning of the overlay with respect to the specimen must be randomized. These days, this can also be conveniently done by computer. Since each point corresponds to a precise area, the total sectional area of any structure is then directly proportional to the average number of points falling over it. For a set of micrographs, the test points are summed and averaged; relative areas can be equated with relative volumes. The relative numbers of test points falling on different structures within the specimen then provide unbiased estimates of the relative profile areas of these structures.

The relative point totals (and relative areas) may be converted into relative volumes (volume densities, volume fractions, or volume proportions) using the principle first described by the geologist, Delesse, in 1848. If the micrographic images themselves represent a set obtained from randomly positioned sites throughout the specimen, the relative number of test points on each structure is not only an estimator of their relative profile areas but also an estimator of their relative volume in the specimen (in fact, the points then sample three-dimensional space). This simple and efficient point-counting approach therefore allows the possibility to characterize a set of cells in terms of their relative contents of nucleus, cytosol, mitochondria, Golgi complex, and any other compartments of interest.

7.2 Relative Profile Length and Surface Area

There is a simple but precise method for estimating the length of a linear trace (line on a sectional image), irrespective of its shape or orientation. On the basis of the principle first observed by Buffon in the nineteenth century, a system of parallel test lines is randomly laid over the profiles of interest. The lines must be randomized in terms of both their position and orientation. This principle can be simply illustrated with a piece of string, which is crumpled (so as to randomize its orientation) and then laid flat on a system of test lines (Fig. 24(b)). The number of times the string is intersected by the test lines (separated by distance, d) is counted and provides a simple and unbiased estimation of the length of string. If the lines on the overlay run in both "horizontal" and "vertical" directions, the length of the string is estimated by the formula $\pi/4 \times I \times d$, where I is the number of intersections and d is the distance between the lines of the grid (Fig. 24(b)). It is easy to visualize the linear traces of membranes on TEM thin sections as the equivalents of pieces of string.

By an analogous three-dimensional approach based on randomly oriented encounters between section planes and cell or organelle membranes, the relative number of intersections between test lines and different membrane compartments allows one to estimate their relative surface areas, that is, surface of membrane per volume of reference space. This simple and efficient intersection counting approach therefore allows the possibility to characterize cells in terms of the surface areas of their different membranes.

7.3 Absolute Volume

Sometimes it is not possible to estimate volume by, say, Archimedes principle of liquid displacement or by weighing, and it is necessary to resort to sectioning. An elegant method for estimating the volume of any structure by sectioning it was invented by Cavalieri in the seventeenth century. It was somehow forgotten until it was rediscovered by Gundersen and was given more general applicability as a stereological principle. The essence of the method is simple. First, the structure is sectioned completely into a uniform random set of slices (Fig. 24(c)). One can analyze each section but, as already mentioned, a great strength of stereology is the concept of sampling. One systematically selects a subset of (e.g. 1 in every 5) sections. On each sampled section, the area is estimated by, say, the test point–counting procedure. The combined area of all the sampled sections is then multiplied by the mean distance between selected sections to obtain the Cavalieri estimate of volume. When proper sampling procedures are followed, this is a remarkably efficient and precise estimator of volume, at scales ranging from a lysosome to a human brain. It can produce volume estimates with a precision of 5%, or less, from just 5 to 6 sections chosen in a systematic random fashion.

7.4 Surfaces and Volumes in Practice

Using the above principles for estimating relative surfaces and volumes, it is easy to appreciate how one could estimate the absolute surface or volume of any cell structure that is larger than the average section thickness (for technical reasons smaller structures are problematical by these methods). For this, one needs to know one quantity in absolute terms. For most cells, biological questions that refer to quantity would pertain to the average volume of the cell (or its nucleus). The *Cavalieri principle* can be used to estimate cell or nuclear volume by either LM or EM. A rather simple method is to stain cell nuclei with a DNA dye, such as Hoechst dye or Dapi. By confocal microscopy, one can optimally section through a set of nuclei and, knowing the distance between sections (given by differences in the focusing plane) and the areas of sections (as above), one can estimate the average nuclear volume by the Cavalieri principle. It is also possible to use 0.5 to 1 μm (semi-thin) section by EM (most nuclei in cells are 10 μm or less in diameter so a relatively small number of sections suffice). At 0.3 μm (or less), the procedure can be done by EM.

Knowing the absolute volume of the nucleus, one would next take random sections at the EM level. For volume estimation, only the position of the sections relative to the specimen needs to be randomized but, for surface estimation, both the position and orientation of the sections must be randomized. At a fairly

low magnification, the overlay of test points can be used to tell us the relative volume of the nucleus to the whole cell. For example, if on average 40 points hit the nuclear profiles and 160 points hit the cytoplasm, the nucleus occupies 40/200 or 20% of the cell volume. If the nuclear volume is known (from above), the absolute volume of the cell is five times that of the nucleus.

The volume of any subcellular compartment of interest can now be estimated by multiplying its volume ratio in the cell (V_V) by total cell volume. Similarly, any surface of interest can be estimated by relating the number of intersections hitting the (membrane) profile of the compartment to the number of points that fall on the entire cell. This gives us the ratio of surface of that membrane to the volume of the cell (S_V). Since $S = S_V \times$ cell volume, this gives an indirect but precise estimate of the absolute surface of the compartment.

There are also ingenious methods (again from Gundersen) for estimating particle *numerical density* (number in a reference volume) or total particle number. Numerical density is estimated by the dissector method, which relies on analysis using two consecutive sections and an unbiased counting frame. If the reference volume is known, an indirect estimate of number is obtained by multiplying this reference volume by particle numerical density. However, number can sometimes be estimated directly using the *fractionator* method. These two approaches have revolutionized the ability of investigators to provide unbiased estimates of the numbers of cells in different tissues, most prominently in various parts of the brain.

Using the test point and intersection counting principles described above, simple methods are available for quantifying immunogold labeling. A major advantage of the most recent developments is the ability to compare observed distributions of gold particles with predicted distributions obtained by counting test points (organelle compartments) or test-line intersections (membrane compartments). The idea here is to use the uniform distribution of the points and intercepts as indicators of randomness. Their predicted distributions correspond to the patterns expected for random labeling. Comparing observed and predicted distributions allows one to identify compartments that are preferentially labeled and to obtain estimates of their labeling densities and relative labeling indices.

A number of stereology courses are offered regularly in Europe and in the United States and the interested reader is well advised to attend such a course. Stereology is not only a set of methods; it offers a completely rigorous way of thinking about good study design in cell and tissue biology. For quantifying structural parameters or for immunolabeling, the use of stereology should be standard. What is most surprising (given the relative lack of its application for cell biology) is that the stereological approach can be so rapidly and efficiently applied.

7.5 Correlative Light and Electron Microscopy

It is not widely appreciated that sections cut for immunolabeling at the EM level can also be very useful tools for labeling at the LM level. This is true for plastic sections such as Lowicryl and for Tokuyasu-thawed cryosections. Since the labeling is restricted to one thin layer on the section surface, this approach can provide better resolution than confocal microscopy. Moreover, multiple fluorescence labeling with different antibodies on

different successive serial sections is possible; since each section (or small number of consecutive sections) is labeled independently there is no restriction on the combinations of antibodies used, for example, three different mouse antibodies can be used, each identified using a different, distinguishable fluorochrome. DNA stains such as Dapi are also very useful since they can be used to find the thin sections by light microscopy. Since the sections are thin, a series of sections covers the same structures and the images from the different antibodies can usually be combined to give pseudo-multiple labeling.

Recently, an elegant method has been developed that allows visualization of GFP-labeled structures in a living cell and then allows to identify the same cell (using an HRP-based approach) at the EM level.

7.6 Interpretation in EM

Seeing and understanding something in three dimension is infinitely easier than trying to interpret two-dimensional (section) images. Thin sections of cells are highly deceptive because they may seduce the investigator into thinking that the inside of the cell is open to a three-dimensional interpretation. In principle, it is, but the fact remains that it takes many years of looking down the microscope before one can be considered a professional. The loss of the third dimension is a more serious problem than is often imagined.

An example was provided of a well known and highly accomplished electron microscopist, who misinterpreted the early thin section images of the ER-cisternae as a system of parallel fibers. This example is useful because a large number of (even experienced) cell biologists automatically think they see a thin "tube" (in 3D) when they visualize profiles of cisternae (especially when the profiles are relatively small).

Another significant challenge facing the beginner in interpreting thin sections is how to average information over many cell profiles, no two of which can be considered really identical! Biases are rampant in the field as one or a few "representative" micrographs are selected for publication. This is especially a problem in presenting immunolabeling images in the absence of quantitation.

7.7 Which Technique for Which Question?

Faced with all of these different possibilities for preparing specimens for EM, the beginner may have difficulty deciding where to start. The first advice is to contact a specialist laboratory. Much will then depend on the scale, whether the starting specimen is isolated vesicles, cells, tissues, or an organism, whose size can scale from a virus or bacterium to a whole organism, such as an insect. Starting at the highest scale, the surface of an organism less than about 1 mm can be visualized directly by SEM. Above this size, the specimen must be physically cut at the first preparation step. An organism as large as 1 mm can be seen at low resolution by conventional ambient temperature SEMs. For higher resolution, the specimen must be available as sets of "particles," whole cells, or sections of cells or tissues. In this case, a cryobased specimen preparation is mostly obligatory.

For TEM, particles smaller than 0.3 to 0.5 μm in height can be directly viewed by cryo-EM or by on-grid negative (positive) staining. The latter should always be

the first approach for any such specimens since it can be done in 5 min with simple technology. Freeze-fracture and shadowing methods can also be applied for isolated particles, especially to see the inner details of membranes in a clearer fashion rather than in the much more limited cross-section profile evident in thin sections. However, the majority of EM studies involve the use of such sections in order to see the detailed ultrastructure within cells. The key question is which approach do I use for (1) structural questions, and (2) immunolabeling questions in the context of ultrastructure.

For purely structural questions for best results, the combination of high-pressure freezing and freeze-substitution is recommended. The hydrated cryosection method may be preferred for some questions when one has access to this technology. If one accepts a potentially lower level of specimen quality one can select one of the two sectioning approaches, plastic embedding or Tokuyasu cryosectioning. Which approach is selected is largely a question of taste; at the top level, the results are often very similar, although the Tokuyasu method has the advantage over conventional embedding in that it avoids solvent dehydration.

For immunolabeling the first question is preembedding versus on-section labeling. Preembedding using immunogold is highly empirical and time-consuming, compared to section labeling. However, when it works, one has the advantage that one has the labeled specimen already embedded in (usually) epoxy resin, which can be cut with standard ultramicrotomes that are widely available. When the labeling is not successful, there is a significantly larger number of variables to adjust than with on-section labeling.

If the antigen is *within* the lumen of a membrane compartment, such as the ER, it makes little sense to try conventional immunogold preembedding labeling because of the difficulties of the antibodies obtaining access to these sites, unless one decides to try to completely open these structures by some means. HRP-based methods can also be used, but one has to accept the fact that this approach is not quantitative and can even be quantitatively misleading.

For the majority of localization questions, on-section labeling is the method of choice. In this approach, each part of every structure has, in theory, an equal choice of being expressed to the antibody on the section surface. In practice, however, the situation is more complex. An important question is whether one should use a plastic-section approach (Lowicryl, LR White etc.) or the Tokuyasu cryosection method. The latter method is faster and recommended for antigens that are less likely to be extracted or artificially mislocalized during the section pickup stage; such artifacts are especially a danger with small, soluble molecules and lipids. It is generally an excellent method for membrane proteins and cytoskeletal proteins; since there is no embedding medium *per se* there is generally more access of antibodies to antigens in the section than is the case for plastic sections, where access is strictly surface-limited. While plastic sections tend to label less than cryosections, they have the advantage that the polymerized resin better ''fixes'' the antigen against redistribution.

In conjunction with freeze-substitution, the low temperature Lowicryl resins offer the ultimate level of specimen preparation that is compatible with immunolabeling.

For small diffusible molecules and lipids, this is absolutely the method of choice. When all methods fail, even with antibodies that give a good signal by light microscopy, the two most likely reasons are, first, lack of accessibility of the antibody to the antigen, even on the surface of the section, and second, the concentration of available antigen on the section surface is simply below detection. A sensible approach is to use similar sections for immunofluorescence analysis that are used for EM; this approach is likely to be more sensitive.

7.8 Final Comment

Given the breadth of techniques available, it is somewhat surprising that the use of more of these EM methods is not sufficiently widespread. Rather than being standard methods in every cell biology institute worldwide, the number of specialist groups has been declining (except for the bare-grid cryo-EM method, which is increasing in use). This problem, and the politics and sociology behind it have recently been discussed in detail. The best possibility to learn these techniques is to visit specialist groups. In addition, practical courses are given in many countries, often sponsored by national EM societies. We teach an annual EMBO-sponsored course in Europe every year. For more details see the websites by Paul Webster and Herb Hagler, which give an excellent overview of many of the techniques mentioned in this chapter. See: (http://www.hei.org/research/scientists/websterbio.htm).and (http://pathcuric1.swmed.edu/Research/haglerEMLab.htm>http://pathcuric1.swmed.edu/Research/haglerEMLab.htm)

Acknowledgments

The need to cover such a broad area of EM obliged me to seek the help of many specialists for this chapter. I would first like to thank my colleagues from the annual EMBO-EM course for their support: Heinz Schwarz, Paul Webster, Herb Hagler, John Lucocq, Marek Cyrklaff, Terry Mayhew, Ivan Raska, Peter Peters, Paul Walther and Kiyoteru Tokuyasu. A number of these friends provided material for figures, acknowledged in the figure legends. Collectively, they also helped me enormously in the writing of the text. I would also like to acknowledge the help of Jacques Dubochet, Bill Earnshaw, Helmut Gnägi, Kenny Goldie, Andy Hoenger, Bruno Humbel, Eduard Kellenberger, Kevin Leonard and Werner Villiger for their input, and, in the case of Bruno Humbel, Jacques Dubochet, Andy Hoenger and Bruno Humbel also for providing beautiful micrographs. Additional excellent images were kindly provided by Kazushi Fujimoto, Shohei Yamashima, Hermann Frank, Christoph Grünfelder, Jean-Luc Murk, Peter Overath, Veronika Neubrand, Anja Habermann and by Robert Ranner of Leica Microsystems, Vienna. Many thanks go also to Herb Hagler, Paul Webster, Heinz Schwarz, Anja Haberman, and Mark Kuehnel for their great support in the preparation of the figures. A special thank to Heinz Schwarz, *to whom this chapter is dedicated*, for giving me the pleasure of seeing his amazing collection of micrographs, the work of over 30 years, that cover almost all the techniques outlined above. Finally, a sad note – In December 2003 Kazushi Fujimoto, a greet pioneer of freeze-fracture immunolabeling passed away at the age of 49. I dedicated this also to his memory.

See also Biogenesis, Structure and Function of Lysosomes; Chromosome Organization within the Nucleus; Electron Microscopy of Biomolecules.

Bibliography

Books and Reviews

DeDuve, C. (1984) *A guided tour of the living cell*, Freeman, New York.

Griffiths, G. (1993) *Fine Structure Immunocytochemistry*, Springer-Verlag, New York.

Hayat, M.A. (1989a) *Principles and Techniques of Electron Microscopy; Biological Applications*, 3rd edition, Macmillan Press, Basingstoke.

Hayat, M.A. (Ed.) (1989b) Colloidal Gold, *Principles, Methods and Applications*, Vols. 1–3, Academic Press. San Diego, New York.

Humbel, B., Müller, M. (1986) *Freeze substitution and low temperature embedding*, in: Müller, M., Becker, R.P., Boyde, A., Wolosewick, J.J. (Eds.) The Science of Biological Specimen Preparation, Scanning Microscope International, AMF O'Hare (Chicago), 175–183.

Roos, N., Morgan, A.J. (1990) Cryopreparation of Thin Biological Specimens for Electron Microscopy: Methods and Applications, *Microscopy Handbooks 21*, Royal Microscopical Society, Oxford University Press, Oxford.

Special Issue on Electron Tomography (2002), *J. Struct. Biol.* **138**(1/2).

Primary Literature

Acetarin, J.D., Carlemalm, E., Villiger, W. (1986) Developments of new lowicryl resins or embedding biological specimens at even lower temperatures, *J. Microsc.* **143**, 81–88.

Adrian, M., Dubochet, J., Lepault, J., McDowall, A.W. (1984) Cryo-electron microscopy of viruses, *Nature* **308**, 32–36.

Adrian, M., Dubochet, J., Fuller, S., Harris, R. (1998) Cryo-negative staining, *Micron* **29**, 145–160.

Allen, T.D., Bennion, G.R., Rutherford, S.A., Reipert, S., Ramalho, A., Kiseleva, E., Goldberg, M.W. (1996) Accessing nuclear structure for field emission, in lens, scanning electron microscopy (FEISEM), *Scanning Microsc. Suppl.* **10**, 149–163.

Allen, T.D., Bennion, G.R., Rutherford, S.A., Reipert, S., Ramalho, A., Kiseleva, E., Goldberg, M.W. (1997) Macromolecular substructure in nuclear pore complexes by in-lens field-emission scanning electron microscopy, *Scanning* **19**, 403–410.

Anderson, T.F. (1951) Techniques for the preservation of three dimensional structure in preparing specimens for the electron microscope, *Trans. N. Y. Acad. Sci. Ser. II* **13**, 130–134.

Baker, T.S., Olson, N.H., Fuller, S.D. (1999) Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs, *Microbiol. Mol. Biol. Rev.* **63**, 862–922.

Baumeister, W. (2002) Electron tomography: towards visualizing the molecular organization of the cytoplasm, *Curr. Opin. Struct. Biol.* **12**, 679–684.

Baumeister, W., Grimm, R., Walz, J. (1999) Electron tomography of molecules and cells, *Trends Cell Biol.* **9**, 81–85.

Bella, J., Rossmann, M.G. (1999) Review: rhinoviruses and their ICAM receptors, *J. Struct. Biol.* **128**, 69–74.

Bellare, J.R., Davis, H.T., Scriven, L.E., Talmon, Y. (1988) Controlled environment vitrification system: an improved sample preparation technique, *J. Electron Microsc. Tech.* **10**, 87–111.

Benedetti, E.L., Dunia, I., Recouvreur, M., Kumar, N.M., Bloemendal, H. (2000) Structural organization of gap junctions as revealed by freeze-fracture and SDS fracture-labeling, *Eur. J. Cell Biol.* **79**, 575–582.

Beuron, F., Hoenger, A. (2001) Structural analsis of the microtubule-kinesin complex by cryo-electron microscopy, *Methods Mol. Biol.* **164**, 235–254.

Bremer, A., Haener, M., Aebi, U. (1998) Negative Staining, in: Celis, J.E. (Ed.) *Cell Biology: A Laboratory Handbook*, Vol. 1, Academic Press, San Diego, pp. 292–299.

Carlemalm, E., Garavito, R.M., Villiger, W. (1980) Advances in Low Temperature Embedding for Electron Microscopy, in: Brederoo, P., Priester, W. (Eds) *Proceedings of the VIIth European Congress on Electron Microscopy*,

Vol. 2, The Hague, The Netherlands, pp. 656–657.

Carlemalm, E., Garavito, R.M., Villiger, W. (1982) Resin development for electron microscopy and an analysis of embedding at low temperature, *J. Microsc.* **126**, 123–143.

Carlemalm, E., Villiger, W., Hobot, J.A., Acetarin, J.D., Kellenberger, E. (1985) Low temperature embedding with Lowicryl resins: two new formulations and some applications, *J. Microsc.* **140**, 55–63.

Causton, B.E. (1984) The Choice of Resins for Immunocytochemistry, in: Polak, J.M., Carndell, I.M. (Eds.) *Immunolabeling for Electron Microscopy*, Elsevier, Amsterdam, pp. 29–37.

Chiu, W., Rixon, F.J. (2002) High resolution structural studies of complex icosahedral viruses: a brief overview, *Virus Res.* **82**, 9–17.

Clegg, J.S. (1982) Alternative Views on the Role of Water in Cell Function, in: Franks, F., Mathias, S. (Eds.) *Biophysics of Water*, Wiley, Chichester, West Sussex, UK, pp. 365–383.

Cooke, C.A., Bernat, R.L., Earnshaw, W.C. (1990) CENP-B: a major human centromere protein located beneath the kinetochore, *J. Cell Biol.* **110**, 1475–1488.

Cooke, C.A., Schaar, B., Yen, T.J., Earnshaw, W.C. (1997) Localization of CENP-E in the fibrous corona and outer plate of mammalian kinetochores from prometaphase through anaphase, *Chromosoma* **106**, 446–455.

Craig, S., Staehelin, L.A. (1988) High pressure freezing of intact plant tissues. Evaluation and characterization of novel features of the endoplasmic reticulum and associated membrane systems, *Eur. J. Cell Biol.* **46**, 81–93.

Cyrklaff, M., Adrian, M., Dubochet, J. (1990) Evaporation during preparation of unsupported thin vitrified aqueous layers for cryo-electron microscopy, *J. Electron Microsc. Tech.* **16**, 351–355.

Dahl, R., Staehelin, L.A. (1989) High-pressure freezing for the preservation of biological structure: theory and practice, *Electron Microsc. Tech.* **13**, 165–174.

Davis, R.W., Davidson, N. (1968) Electron-microscopic visualization of deletion mutations, *Proc. Natl. Acad. Sci. U.S.A.* **60**, 243–250.

Davis, R.W., Simon, M., Davidson, N. (1972) Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids, *Methods Enzymol.* **XXI**, 413–428.

Dubochet, J., Adrian, M., Chang, J.J., Homo, J.C., Lepault, J., McDowall, A., Schutz, P. (1988) Cryo-electron microscopy of vitrified specimens, *Q. Rev. Biophys.* **21**, 129–228.

Dubochet, J., McDowall, A.W. (1981) Vitrification of pure water for electron microscopy, *J. Microsc.* **124**, 3–4.

Dubochet, J., McDowall, A.W., Menge, B., Schmidt, E.N., Lickfeld, K.G. (1983) Electron microscopy of frozen-hydrated bacteria, *J. Bacteriol.* **155**, 381–390.

Dubochet, J., Sartori Blanc, N. (2001) The cell in absence of aggregation artifacts, *Micron* **32**, 91–99.

Dyba, M., Hell, S.W. (1988) Focal spots of size lambda/23 open up far-field fluorescence microscopy at 33 nm axial resolution, *Phys. Rev. Lett.* **88**, 163–901.

Earnshaw, W.C., Cooke, C.A. (1991) Analysis of the distribution of the INCENPs throughout mitosis reveals the existence of a pathway of structural changes in the chromosomes during metaphase and early events in cleavage furrow formation, *J. Cell Sci.* **98**, 443–461.

Echlin, P. (1992) *Low-Temperature Microscopy and Analysis*, Plenum Press, New York, London.

Edelmann, L. (2002) Freeze-dried and resin-embedded biological material is well suited for ultrastructure research, *J. Microsc.* **207**, 5–26.

Elder, H.Y., Wilson, S.M, Johnson, A.D. (1998) Preparative Methods for Quantitative X-ray Microanalysis of Intracellular Elements in Ultrathin Sections for Transmission Electron Microscopy: The Dried, Resin-Embedded Route, in: Celis, J.E. (Ed.), *Cell Biology: A Laboratory Handbook*, Vol. 1, Academic Press, San Diego, pp. 340–346.

Elias, H. (1972) Identification of structure by the common sense approach, *J. Microsc.* **95**, 59–68.

Escaig, J. (1982) New instruments which facilitate rapid freezing at 83 K and 6 K, *J. Microsc.* **126**, 221–229.

Felsmann, M., Stierhof, Y.-D., Humbel, B.M., Otten, M.T. (1992) Detection of ultra-small gold labels with a FEG TEM/STEM, *Beitr. Elektronenmikroskop. Direktabb. Oberfl.* **25**, 15–18.

Fernández-Moran, H. (1959) Cryofixation and supplementary low temperature preparation

techniques applied to the study of tissue ultrastructure, *J. Appl. Phys.* **30**, 2038.

Fernández-Moran, H. (1960) Low temperature preparation techniques for electron microscopy of biological specimens based on rapid freezing with liquid helium II, *Ann. N. Y. Acad. Sci.* **85**, 689–713.

Frank, J. (2001) Ribosomal dynamics explored by cryo-electron microscopy, *Methods* **25**, 309–315.

Frank, J. (2002) Single-particle imaging of macromolecules by cryo-electron microscopy, *Annu. Rev. Biophys. Biomol. Struct.* **31**, 303–319.

Frank, J. (2003) Electron microscopy of functional ribosome complexes, *Biopolymers* **68**, 223–233.

Frank, J., Wagenknecht, T., McEwen, B.F., Marko, M., Hsieh, C.-E., Mannella, C.A. (2002) Three-dimensional imaging of biological complexity, *J. Struct. Biol.* **138**, 85–91.

Frey, T.G., Renken, C.W., Perkins, G.A. (2002) Insight into mitochondrial structure and function from electron tomography, *Biochim. Biophys. Acta* **1555**, 196–203.

Fujimoto, K. (1995) Freeze-fracture replica electron microscopy combined with SDS digestion for cytochemical labeling of integral membrane proteins. Application to the immunogold labeling of intercellular junctional complexes, *J. Cell Sci.* **108**, 3443–3449.

Fujimoto, K. (1997) SDS-digested freeze-fracture replica labeling electron microscopy to study the two-dimensional distribution of integral membrane proteins and phospholipids in biomembranes: practical procedure, interpretation and application, *Histochem. Cell Biol.* **107**, 87–96.

Fujimoto, T., Fujimoto, K. (1997) Metal sandwich method to quick freeze monolayer cultured cells for freeze-fracture, *J. Histochem. Cytochem.* **45**, 595–598.

Fujimoto, K., Noda, T., Fujimoto, T. (1997) A simple and reliable quick-freezing/freeze-fracturing procedure, *Histochem. Cell Biol.* **107**, 81–84.

Fujimoto, K., Pinto da Silva, P. (1988) Macromolecular dynamics of the cell surface during the formation of coated pits is revealed by fracture-flip, *J. Cell Sci.* **91**, 161–173.

Fujimoto, K., Pinto da Silva, P. (1989) Surface views of nuclear pores in isolated rat liver nuclei as revealed by fracture-flip/Triton-X, *Eur. J. Cell Biol.* **50**, 390–397.

Gaietta, G., Deerinck, T.J., Adams, S.R., Bouwer, J., Tour, O., Laird, D.W., Sosinsky, G.E., Tsien, R.Y., and Ellisman, M.H. (2002) Multicolor and electron microscopic imaging of connexin trafficking, *Science* **296**, 503–507.

Geuze, H.J. (2000) A future for electron microscopy in cell biology, *Trends Cell Biol.* **9**, 92–93.

Giberson, R.T., Demaree, R.S. Jr. (1999) Microwave Processing Techniques for Electron Microscopy: A Four Hour Protocol, in: Hajibagheri, N. (Ed.) *Methods in Molecular Biology*, Vol. 117: *Electron Microscopy Methods and Protocols*, Humana Press, Totowa, NJ, USA, pp. 145–158.

Glauert, A.M., Glauert, A.M. (1975) (Eds.) *Practical Methods in Electron Microscopy*, 5 Vols. North Holland, Amsterdam.

Green, J., Griffiths, G., Louvard, D., Quinn, P., Warren, G. (1981) Passage of viral protein through the Golgi complex, *J. Mol. Biol.* **152**, 663–698.

Griffiths, G. (1993) *Fine Structure Immunocytochemistry*, Springer-Verlag, New York.

Griffiths, G.G. (2001) Bringing electron microscopy back into focus for cell biology, *Trends Cell Biol.* **11**, 153–154.

Griffiths, G., Back, R., Marsh, M. (1989) A quantitative analysis of the endocytic pathway in baby hamster kidney cells, *J. Cell Biol.* **109**, 2703–2720.

Griffiths, G., Lucocq, J.M., Mayhew, T.M. (2001) Electron microscopy applications for quantitative cellular microbiology, *Cell Microbiol.* **3**, 659–668.

Griffiths, G., McDowall, A., Back, R., Dubochet, J. (1984) On the preparation of cryosections for immunocytochemistry, *J. Ultrastruct. Res.* **1**, 65–78.

Griffiths, G., Simons, K., Warren, G., Tokuyasu, K.T. (1983) Immunoelectron microscopy using thin, frozen sections: application to studies of the intracellular transport of Semliki Forest virus spike glycoproteins, *Methods Enzymol.* **96**, 466–485.

Griffiths, G, Wepf, R, Wendt, T., Krijnse Locker, J., Cyrklaff, M., Roos, N. (2001) Structure and assembly of vaccinia virus: isolated particle analysis, *J. Virol.* **75**, 11034–11055.

Grünfelder, C.G., Engstler, M., Weise, F., Schwarz, H., Stierhof, Y.D., Boshart M.,

Overath, P. (2002) Accumulation of a GPI-anchored protein at the cell surface requires sorting at multiple intracellular levels, *Traffic* **3**, 547–559.

Gundersen, H.J., Bagger, P., Bendtsen, T.F., Evans, S.M. Korbo, L., Marcussen, N., Moller, A., Nielsen, K., Nyengaard, J.R., Pakkenberg, B., Sorensen, F.B., Vesterby, A., West, M.J. (1988) The new stereological tools: disector, fractionator, nucleator and point sampled intercepts and their use in pathological research and diagnosis, *APMIS* **96**(10), 857–881.

Gundersen, H.J., Bendtsen, T.E., Korbo, L., Marcussen, N., Moller, A., Nielsen, K., Nyengaard, J.R., Pakkenberg, B., Sorensen, F.B., Vesterby, A., West, M.J. (1988) Some new, simple and efficient stereological methods and their use in pathological research and diagnosis, *APMIS* **96**, 379–394.

Gundersen, H.J. (1986) Stereology of arbitrary particles. A review of unbiased number and size estimators and the presentation of some new ones, in memory of William R. Thompson, *J. Microsc.* **143**, 3–45.

Gundersen, J.H., Jensen, E.B. (1987) The efficiency of systematic sampling in stereology and its prediction, *J. Microsc.* **147**, 229–263.

Gundersen, H.J., Jensen, E.B., Kieu, K., Nielsen, J. (1999) The efficiency of systematic sampling in stereology – reconsidered, *J. Microsc.* **193**, 199–211.

Haener, M., Bremer, A., Aebi, U. (1998) Glycerol Spraying/Low Angle Rotary Metal Shadowing, in: Celis, J.E. (Ed.) *Cell Biology: A Laboratory Handbook*, Vol. 1, Academic Press, San Diego, pp. 292–299.

Hagler, H.K., Buja, L.M. (1983) Electron Probe Microanalysis for the Study of Cation Distribution and Localization in Cardiac Muscle, *Methods in Studying Cardiac Membranes*, Vol. II, CRC Press, Boca Raton, pp. 205–215.

Hagler,. K., Buja, L.M. (1990) Subcellular Calcium Shifts in Ischemia and Reperfusion, in: Piper, H. (Ed.), *Pathophysiology of Severe Ischemic Myocardial Injury*, Kluwer Academic Publishers, Dordrecht/Boston/London, pp. 283–296.

Hainfeld, J.F., Furuya, F.R. (1992) A 1.4 nm cluster covalently attached to antibodies improves immunolabeling, *J. Histochem. Cytochem.* **40**, 177–184.

Hannah, M.J., Weiss, U., Huttner, W.B. (1988) Differential extraction of proteins from paraformaldehyde-fixed cells: lessons from synaptophysin and other membrane proteins, *Methods: A Companion Methods Enzymol.* **16**, 170–181.

Harris, J.R., Horne, R.W. (1994) Negative staining: a brief assessment of current technical benefits, limitations and future possibilities, *Micron* **25**, 5–13.

Hasegawa, H., Mizuhira, V., Notoya, M. (2000) Microwave-stimulated fixation and histochemical application to biological specimens, *Acta Histochem. Cytochem.* **33**, 319–340.

Hayat, M.A. (Ed.) (1974–1977) *Electron Microscopy of Enzymes*, Vols. 1–5, Van Nostrand Reinhold, New York.

Hayat, M.A. (1981) *Fixation for Electron Microscopy*, Academic Press, San Diego, London.

Hayat, M.A. (1989a) *Principles and Techniques of Electron Microscopy; Biological Applications*, 3rd edition, Macmillan Press, Basingstoke.

Hayat, M.A. (Eds.) (1989b) Colloidal Gold, *Principles, Methods and Applications*, Vols. 1–3, Academic Press, San Diego, New York.

Henderson, R. (1995) The potential and limitations of neutrons, electrons and X-ray for atomic resolution microscopy of unstained biological molecules, *Q. Rev. Biophys.* **28**, 171–193.

Hermann, R., Müller, M. (1991) High resolution biological scanning electron microscopy: a comparative study of low temperature metal coating techniques, *J. Electron Microsc. Tech.* **18**, 440–449.

Hermann, R., Müller, M. (1992) Towards high resolution SEM of biological objects, *Arch. Histol. Cytol.* **55**, 17–25.

Hermann, R., Pawley, J.B., Nagatani, T., Müller, M. (1988) Double-axis rotary shadowing for high-resolution scanning electron microscopy, *Scanning Microsc.* **2**, 1215–1230.

Hermann, R., Schwarz, H., Müller, M. (1991) High precision immuno-scanning electron microscopy using Fab fragments couples to ultra-small colloidal gold, *J. Struct. Biol.* **107**, 38–47.

Hermann, R., Walther, P., Müller, M. (1996) Immunolabeling in scanning electron microscopy, *Histochem. Cell Biol.* **106**, 31–39.

Heuser, J. (1981) Preparing biological samples for stereomicroscopy by the quick-freeze, deep-etch, rotary-replication technique, *Methods Cell Biol.* **22**, 97–122.

Heuser, J.E. (2000) How to convert a traditional electron microscopy laboratory to digital imaging: follow the "middle road", *Traffic* **1**, 614–621.

Heuser, J.E., Reese, T.S., Dennis, M.J., Jan, Y., Jan, L., Evans, L. (1979) Synaptic vesicle exocytosis captured by quick freezing and correlated with quantal transmitter release, *J. Cell Biol.* **81**, 275–300.

Hippe-Sanwald, S. (1993) Impact of freeze substitution on biological electron microscopy, *Microsc. Res. Tech.* **24**, 400–422.

Hoenger, A., Doerhoefer, M., Woehlke, G., Tittmann, P., Gross, H., Song, Y.H., Mandelkow, E. (2000) Surface topography of microtubule walls decorated with monomeric and dimeric kinesin constructs, *Biol. Chem.* **381**, 1001–1011.

Hohenberg, H. (1989) Replica Preparation Techniques in Immunogold Cytochemistry, in: Verkleij, A.J., Leunissen, J.L.M. (Eds.) *Immunogold Labeling in Cell Biology*, CRC Press, Boca Raton, pp. 155–157.

Hohenberg, H., Kirschning, E., Rutter, G. (1987) Cryo- and micro-techniques: Future tools for electron microscopy, *J. Comp.-Assist. Microsc.* **9**, 69–72.

Hohenberg, H., Mannweiler, K., Müller, M. (1994) High-pressure freezing of cell suspensions in cellulose capillary tubes, *J. Microsc.* **175**, 34–43.

Hohenberg, H., Tobler, M., Müller, M. (1996) High-pressure freezing of tissue obtained by fine-needle biopsy, *J. Microsc.* **183**, 133–139.

Howard, C.V., Reed, M.G. (1998) Unbiased Stereology, *Three-Dimensional Measurement in Microscopy*, Bios Scientific, Oxford.

Hsieh, C.E., Marko, M., Frank, J., Mannella, C.A. (2002) Electron tomographic analysis of frozen-hydrated tissue sections, *J. Struct. Biol.* **138**, 63–73.

Humbel, B.M., De Jong, M.D.M., Müller, W.H., Verkleij, A.J. (1998) Pre-embedding immunolabeling for electron microscopy: An evaluation of permeabilization methods and markers, *Microsc. Res. Tech.* **42**, 43–58.

Humbel, B.M., Konomi, M., Takagi, T., Kamasawa, N., Ishijima, S.A., Osumi, M. (2001) *In situ* localization of β-glucans in the cell wall of Schizosaccharomyces pombe, *Yeast* **18**, 433–444.

Humbel, B., Marti, T., Müller, M. (1983) Improved structural preservation by combining freeze substitution and low temperature embedding, *Beitr. Elektronenmikroskop. Direktabb. Oberfl.* **16**, 585–594.

Humbel, B.M., Müller, M. (1986) Freeze Substitution and Low Temperature Embedding, in: Muller, M., Becker, R.P., Boyde, A., Wolosewick, J.J. (Eds.) *The Science of Biological Specimen Preparation 1985*, SEM, AMF O' Hare, IL, pp. 175–183.

Humbel, B.M., Schwarz, H. (1989) Freeze-Substitution for Immunochemistry, in: Verkleij, A.J., Leunissen, J.L.M. (Eds.) *Immuno-Gold Labeling in Cell Biology*, CRC Press, Boca Raton, FL, pp. 115–134.

Humbel, B.M., Sibon, O.C., Stierhof, Y.D., Schwarz, H. (1995) Ultra-small gold particles and silver enhancement as a detection system in immunolabeling and in situ hybridization experiments, *J. Histochem. Cytochem.* **43**, 735–737.

Hunziker, E.B., Herrmann, W., Schenk, R.K., Mueller, M., Moor, H. (1984) Cartilage ultrastructure after high pressure freezing, freeze substitution, and low temperature embedding. I. Chondrocyte ultrastructure – implications for the theories of mineralization and vascular invasion, *J. Cell Biol.* **98**, 267–276.

Jongebloed, W.L., Stokroos, I., Van der Want, J.J.L., Kalicharan, D. (1999) Non-conductive fixation techniques or redundancy of conductive coating, low kV FE-SEM operation and combined SEM/TEM of biological tissues, *J. Microsc.* **193**, 158–170.

Kaneko, Y., Walther, P. (1995) Comparison of ultrastructure of germinating pea leaves prepared by high-pressure freezing/freeze substitution and conventional chemical fixation, *J. Electron Microsc.* **44**, 104–109.

Kellenberger, E. (1987) The Response of Biological Macromolecules and Supramolecular Structures to the Physics of Specimen Cryopreparation, in: Steinbrecht, R.A., Zierold, K. (Eds.) *Cryotechniques in Biological Electron Microscopy*, Springer, Berlin, pp. 35–66.

Kellenberger, E. (1991) The potential of cryofixation and freeze substitution: observations and theoretical consideration, *J. Microsc.* **161**, 183–203.

Kellenberger, E., Durrenberger, M., Villiger, W., Carlemalm, E., Wurtz, M. (1987) The

efficiency of immunolabel on Lowicryl sections compared to theoretical predictions, *J. Histochem. Cytochem.* **35**, 959–969.

Klar, T.A., Jakobs, S., Dyba, M., Egner, A., Hell, S.W. (1997) Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8206–82010.

Kleinschmidt, A.K. (1968) Monolayer techniques in electron microscopy of nucleic acid molecules, *Methods Enzymol.* **XII**, 361–377.

Knoll, G., Verkleij, A.J., Plattner, H. (1987) Cryofixation of Dynamic Processes in Cells and Organelles, in: Steinbrecht, R.A., Zierold, K. (Eds.) *Cryotechniques in Biological Electron Microscopy*, Springer-Verlag, Berlin, pp. 258–271.

Koberna, K., Stanek, D., Malinsky, J., Eltsoy, M., Pliss, A., Ctrnacta, V., Cermanova, S., Raska, I. (1999) Nuclear organization studied with the help of a hypotonic shift: its use permits hydrophilic molecules to enter into living cells, *Chromosoma* **108**, 325–335.

Kok, L.P, Boon, M.E. (1992) Microwave Cookbook for Microscopists, *Art and Science of Visualization*, 3rd edition, Coulomb Press, Leyden.

Koster, A.J., Grimm, R., Typke, D., Hegerl, R., Stoschek, A., Walz, J., Baumeister, W. (1997) Perspectives of molecular and cellular electron tomography, *J. Struct. Biol.* **120**, 276–308.

Krijnse Locker, J., Griffiths, G. (1999) An unconventional role for cytoplasmic disulfide bonds in vaccinia virus proteins, *J. Cell. Biol.* **144**, 267–279.

Ladinsky, M.S., Mastronarde, D.N., McIntosh, J.R., Howell, K.E., Staehelin, L.A. (1999) Golgi structure in three dimensions: functional insights from the normal rat kidney cell, *J. Cell Biol.* **22**, 1135–1149.

Ladinsky, M.S., Wu, C.C., McIntosh, S., Mcintosh, J.R., Howell, K.E. (2002) Structure of the golgi and distribution of reporter molecules at 20 degrees C reveals the complexity of the exit compartments, *Mol. Biol. Cell.* **18**, 2810–2825.

Leforestier, A., Dubochet, J., Livolant, F. (2001) Bilayers of nucleosome core particles, *Biophys. J.* **81**, 2414–2421.

Liou, W., Geuze, H.J., Slot, J.W. (1996) Improving structural integrity of cryosections for immunogold labelling, *Histochem. Cell Biol.* **106**, 41–58.

Login, G.R., Dvorak, A.M. (1996) *A Practical Guide for Microscopists*, Beth Islael Hospital Boston, Boston, MA, USA.

Lucocq, J. (1994) Quantitation of gold labelling and antigens in immunolabelled ultrathin sections, *J. Anat.* **184**, 1–13.

Maupin-Szaimer, P., Pollard, T.D. (1978) Actin filament distruction by osmium tetroxide, *J. Cell. Biol.* **77**, 837–847.

Mayers, C.P. (1970) Histological fixation by microwave heating, *J. Clin. Pathol.* **23**, 273–275.

Mayhew, T.M. (1991) The new stereological methods for interpreting functional morphology from slices of cells and organs, *Exp. Physiol.* **76**, 639–665.

Mayhew, T.M., Griffiths, G., Habermann, A., Lucocq, J.M., Emre, N., Webster, P. (2003) A simpler way of comparing the labelling densities of cellular compartments illustrated using data from VPARP and LAMP-1 immunogold labelling experiments, *Histochem. Cell Biol.* **119**, 333–341.

Mayhew, T.M., Gundersen. H.J. (1996) If you assume, you can make an ass out of u and me: a decade of the disector for stereological counting of particles in 3D space, *J. Anat.* **188**, 1–15.

Mayhew, T.M., Lucocq, J.M., Griffiths, G. (2002) Relative labelling index: a novel stereological approach to test for non-random immunogold labelling of organelles and membranes on transmission electron microscopy thin sections, *J. Microsc.* **205**, 153–164.

McDonald, K. (1999) High-pressure freezing for preservation of high resolution fine structure and antigenicity for immunolabeling, *Methods Mol. Biol.* **117**, 77–97.

McDonald, K., Morphew, M.K. (1993) Improved preservation of ultrastructure in difficult-to-fix organisms by high pressure freezing and freeze substitution: 1 Drosophila melanogaster and Strongylocentrotus purpuratus embryos, *Micros. Res. Tech.* **24**, 465–473.

McDowall, A.W., Chang, J.J., Freeman, R., Lepault, J., Walter, C.A., Dubochet, J. (1983) Electron microscopy of frozen hydrated sections of vitreous ice and vitrified biological samples, *J. Microsc.* **131**, 1–9.

McDowall, A.W., Hofmann, W., Lepault, J., Adrian, M., Dubochet, J. (1984) Cryo-electron microscopy of vitrified insect flight muscle, *J. Mol. Biol.* **178**, 105–111.

McIntosh, J.R. (2001) Electron microscopy of cells: a new beginning for a new century, *J. Cell Biol.* **153**, 25–32.

Medalia, O., Weber, I., Frangakis, A.S., Nicastro, D., Gerisch, G., Baumeister, W. (2002) Macromolecular architechture in eukaryotic cells visualized by cryoelectron tomography, *Science* **8**, 1209–1213.

Melan, M.A., Sluder, G. (1992) Redistribution and differential extraction of soluble proteins in permeabilized cultured cells, *J. Cell Sci.* **101**, 731–743.

Meyer, H.W., Richter, W. (2001) Freeze-fracture studied on lipids and membranes, *Micron* **32**, 615–644.

Michel, H., Hillmann, T., Müller, M. (1991) Cryosectioning of plant material frozen at high pressure, *J. Microsc.* **166**, 3–18.

Mizuhira, V., Hasegawa, H., Notoya, M. (1993) Microwave Fixation for Histochemistry and Cytochemistry and Immunocytochemistry in Biomedicine, in: Ogawa, K., Barka, T. (Eds.) *Electron Microscopic Cytochemistry and Immunocytochemistry in Biomedicine*, CRC Press, FL, USA, pp. 17–27.

Moffat, K., Henderson, R. (1995) Freeze trapping of reaction intermediates, *Curr. Opin. Struct. Biol.* **5**, 656–663.

Moor, H. (1964) Die Gefrier-Fixation lebender Zellen und ihre Anwendung in der Elektronen-mikroskopie, *Z. Zellforsch.* **62**, 546–548.

Moor, H. (1987) Theory and Practice of High Pressure Freezing, in: Steinbrecht, R.A., Zierold, K. (Eds.) *Cryotechniques in Biological Electron Microscopy*, Springer, Berlin, pp. 175–191.

Moor, H., Riehle, U. (1968) Snap Freezing Under High Pressure: A New Fixation Technique for Freeze Etching, in: Steve Boccarielli, D. (Ed.) *Electron Microscopy 1968*, Vol. 2, *Proc 4th Eur. Conf. Electron Microscopy*, Rome, pp. 33–34.

Müller, S.A., Engel, A. (2001) Structure and mass analysis by scanning transmission electron microscopy, *Micron* **32**, 21–31.

Müller, M., Meister, N., Moor, H. (1980) Freezing in a propane jet and its application in freeze fracturing, *Mikroskopie* **36**, 129–140.

Müller, H., Hermann, R. (1990) Towards High Resolution SEM of Biological Objects, in: Peachy, L.D., Williams, D.B. (Eds.) *Proceedings of the XIIth International Congress of Electron Microscopy*, Vol. 3, San Francisco Press, San Francisco, pp. 4–5.

Müller, M., Moor, H. (1984) Cryofixation of Thick Specimans by High-Pressure Freezing, in: Revel, J.-P., Barnard, T., Haggis, G.H. (Eds.) *The Science of Biological Specimen Preparation*, SEM, AMF O'Hare, IL, pp. 131–138.

Müller, M., Walther, P., Hermann, R., Schwab, P. (1989) SEM Immunocytochemistry With Small (5 to 15 nm) Colloidal Gold Markers, in: Verkleij, A.J., Leunissen, J.L.M. (Eds.) *Immuno-Gold Labeling in Cell Biology*, CRC Press, Boca Raton, pp. 199–216.

Nagatani, T., Saito, S., Sato, M., Yamada, M. (1987) Development of an ultrahigh resolution scanning electron micrscope by means of a field emission source and in-lens system, *Scanning Microsc.* **1**, 901–909.

Nicholson, W.V., Glaeser, R.M. (2001) Review: automatic particle detection in electron microscopy, *J. Struct. Biol.* **133**, 90–101.

Nicolas, M.T., Bassot, J.M., Nicolas, G. (1989) Immunogold labeling of luciferase in the luminous bacterium Vibrio harveyi after fast-freeze fixation and different freeze-substitution and embedding procedures, *J. Histochem. Cytochem.* **37**, 663–674.

Otten, M.T., Stenzel, D.J., Cousens, D.R., Humbel, B.M., Leunissen, J.L.M., Stierhof, Y.-D., Busing, W.M. (1992) High-angle annular dark-field STEM imaging of immunogold labels, *Scanning* **14**, 282–289.

O'Toole, E.T., Winey, M., McIntosh, J.R., Mastronarde, D.N. (2002) Electron tomography of yeast cells, *Methods Enzymol.* **351**, 81–95.

Pawley, J.B., Erlandsen, S.L. (1989) The case for low voltage high resolution scanning electron microscopy of biological samples, *Scanning Microsc.* **3**, 163–178.

Pakkenberg, B., Gundersen, H.J. (1988) Total number of neurons and glial cells in human brain nuclei estimated by the disector and the fractionator, *J. Microsc.* 1–20.

Peters, K.R. (1986) Rationale for the application of thin, continuous metal films in high magnification electron microscopy, *J. Microsc.* **142**, 25–34.

Peters, P.J. (1999) Cryo-Immunogold Electron Microscopy, in: Bonifacino, J.S., Dasso, M., Harford, J.B., Lippincott-Schwartz, J., Yamada, K.M. (Eds.) *Current Protocols in Cell Biology*, John Wiley & Sons, New York, pp. 4.7.1–4.7.12.

Pinto da Silva, P. (1989) Visual Thinking of Biological Membranes: From Freeze-Etching to Label-Fracture, in: Verkleij, A.J., Leunissen, J.L.M (Eds.) *Immuno-Gold Labeling in Cell Biology*, CRC Press, Boca Raton, pp. 179–197.

Plattner, H., Bachman, L. (1982) Cryofixation: a tool in biological ultrastructural research, *Int. Rev. Cytol.* **79**, 237–304.

Polishchuk, R.S., Polishchuck, E.V., Marra, P., Alberti, S., Buccione, R., Luini, A., Mironov, A.A. (2000) *J. Cell Biol.* **148**, 45–58.

Porter, K.R., Claude, A., Fullam, E. (1945) A study of tissue culture cells by electron microscopy: methods and preliminary observations, *J. Exp. Med.* **81**, 233–241.

Posthuma, G., Slot, J.W., Geuze, H.J. (1984) Immunocytochemical assays of amylase and chymotrypsinogen in rat pancreas secretory granules. Efficacy of using immunogold-labeled ultrathin cryosections to estimate relative protein concentrations, *J. Histochem. Cytochem.* **32**, 1028–1034.

Prack, E.R. (1993) An introduction to process visualization capabilities and consideration in the environmental scanning electron microscope (ESEM), *Microsc. Res. Tech.* **25**, 487–492.

Quintana, C. (1994) Cryofixation, cryosubstitution, cryoembedding for ultrastructural, immunocytochemical and microanalytical studies, *Micron* **25**, 63–99.

Raposo, G., Kleijmeer, M.J., Posthuma, G., Slot, J.W., Geuze, H.J. (1997) Immunogold Labeling of Ultrathin Cryosections: Application in Immunology, *Handbook of Experimental Immunology*, 5th edition, Vol. 4, Blackwell, Oxford, pp. 1–11.

Rash, J.E., Yasumura, T., Dudek, F.E. (1998) Ultrastructure, histological distribution, and freeze-fracture immunocytochemistry of gap junctions in rat brain and spinal cord, *Cell Biol. Int.* **22**, 731–749.

Ribrioux, S., Kleymann, G., Haase, W., Heitmann, K., Ostermeier, C., Michel, H. (1996) Use of nanogold- and fluorescent-labeled antibody Fv, *J. Histochem. Cytochem.* **44**, 207–213.

Richter, K., Gnägi, H., Dubochet, J. (1991) A model for cryosectioning based on the morphology of vitrified ultrathin sections, *J. Microsc.* **163**, 19–28.

Ris, H. (1985) The cytoplasmic filament system in critical point-dried whole mounts and plastic-embedded sections, *J. Cell Biol.* **100**, 1474–1487.

Ris, H. (1997) High-resolution field-emission scanning electron microscopy of nuclear pore complex, *Scanning* **19**, 368–375.

Ris, H., Malecki, M. (1993) High-resolution field emission scanning electron microscope imaging of internal cell structures after epon extraction from sections: A New Approach to Correlative Ultrastructural and Immunocytochemical Studies, *J. Struct. Biol.* **111**, 148–157.

Roberts, N., Puddlephat, M.J., McNulty, V. (2000) The benefit of stereology for quantitiative radiology, *Br. J. Radiol.* **783**, 679–697.

Robards, A.W., Sleytr, U.B. (1985) Low Temperature Methods in Biological Electron Microscopy, in: Glauert, A.M. (Ed.) *Practical Methods in Electron Microscopy*, Vol. 10, Elsevier, Amsterdam.

Roos, N., Cyrklaff, M., Cudmore, S., Basco, R., Krijnse-Locker, J., Griffiths, G. (1996) A novel immunogold cryoelectron microscopic approach to investigate the structure of the intracellular and extracellular forms of vaccinia virus, *EMBO J.* **15**, 2343–2355.

Roos, N., Morgan, A.J. (1990) Cryopreparation of Thin Biological Specimens for Electron Microscopy: Methods and Applications, *Microscopy Handbooks 21*, Royal Microscopical Society, Oxford University Press, Oxford.

Roth, J., Bendayan, M., Carlemalm, E., Villiger, W., Garavito, R.M. (1981) The enhancement of structural preservation and imunocytochemical staining in low temperature embedded pancreatic tissue, *J. Histochem. Cytochem.* **29**, 663–671.

Sabatini, D.D., Bensch, K., Barnett, R.J. (1963) Aldehyde fixation for morphological and enzyme histochemical studies with the electron microscope, *J. Cell. Biol.* **17**, 19–58.

Sartori, N., Richter, K., Dubochet, J. (1993). Vitrification depth can be increased more than 10 fold by high pressure freezing, *J. Microsc.* **172**, 55–61.

Sartori Blanc, N., Senn, A., Leforestier, A., Livolant, F., Dubochet, J. (2001) DNA in human and stallion spermatozoa forms local hexagonal packing with twist and many defects, *J. Struct. Biol.* **134**, 76–81.

Schwarz, H. (1994) Immunolabelling of Ultrathin Resin Sections for Fluorescence and Electron Microscopy, in: Jouffrey, B.,

Colliex, C. (Eds.) *Electron Microscopy, 1994 ICEM 13-Paris*, Vol. 3, Les Editions de Physique, Les Ulis, France, pp. 255–256.

Schwarz, H. (1998) Correlative Immunolabelling of Ultrathin Cryosections for Light and Electron Microscopy, in: Benavides, H.A.C., Yacaman, M.J. (Eds.) *1998 ICEM 14-Cancun*, Vol. 4, Institute of Physics Publ., Bristol, Philadelphia, pp. 865–866.

Schwarz, H., Hohenberg, H., Humbel, B.M. (1993) Immuno-Gold Electron Microscopy, in: Hyatt, A.D., Eaton, B.T. (Eds.) *Virus Diagnosis and Research*, CRC Press, Boca Raton, Ann Arbor, London, Tokyo, pp. 350–370.

Schwarz, H., Humbel, B.M. (1989) Influence of fixatives and embedding media on immunolabelling of freeze-substituted cells, *Scanning Microsc. Suppl.* **3**, 57–63.

Sitte, H., Edelmann, L., Neumann, K. (1987) Cryofixation Without Pretreatment at Ambient Pressure, in: Steinbrecht, R.A., Zierold, K. (Eds.) *Cryotechniques in Biological Electron Microscopy*, Springer, Berlin, pp. 87–113.

Sjostrand, F.S. (1967) *Electron Microscopy of Cells and Tissues*, Vol. 1, Instrumentation and Techniques, Academic Press, New York, London.

Slot, W., Geuze, H.J., (1985) A new method of preparing gold probes for multiple-labeling cytochemistry, *Eur. J. Cell Biol.* **38**, 87–93.

Slot, J.W., Posthuma, G., Chang, L.Y., Crapo, J.D., Geuze, H.J. (1989a) Quantitative aspects of immunogold labeling in embedded and in non-embedded sections, *Am. J. Anat.* **185**, 271–281.

Slot, J.W., Posthuma, G., Chang, L.Y., Crapo, J.D., Geuze, H.J. (1989b) Quantitative Assessment of Immuno-Gold Labeling in Cryosections, in: Verkleij, A.J., Leunissen, J.L.M. (Eds.) *Immuno-Gold Labeling in Cell Biology*, CRC Press, Boca Raton, FL, pp. 135–157.

Somlyo, A.P., Himpens, B. (1989) Cell calcium and its regulation in smooth muscle, *FASEB J.* **3**, 2266–2276.

Somlyo, A.P., Schuman, H. (1982) Electron probe and electron energy loss analysis in biology, *Ultramicroscopy* **8**, 219–233.

Spehner, D., Drillien, R., Proamer, F., Hanau, D., Edelmann, L. (2002) Embedding in Spurr's resin is a good choice for immunolabelling after freeze drying as shown with chemically unfexed dendritic cells, *J. Microsc.* **207**, 1–4.

Spurr, A.R. (1969) A low-viscosity epoxy resin embedding media for electron microscopy, *J. Ultrastruct. Res.* **26**, 31–43.

Staehelin, L.A., Giddings, T.H. Jr., Kiss, J.Z., Sack, F.D. (1990) Macromolecular differentiation of Golgi stacks in root tips of Arabidopsis and Nicotiana seedlings as visualized in high pressure frozen and freeze-substituted samples, *Protoplasma* **157**, 75–91.

Steinbrecht, R.A., Zierold, K. Cryotechniques, *Biological Electron Microscopy*, Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo.

Sterio, D.C. (1984) The unbiased estimation of number and sizes of arbitrary particles using the disector, *J. Microsc.* **134**, 127–136.

Steven, A.C., Aebi, U. (2003) The next ice age: cryo-electron tomography of intact cells, *Trends Cell. Biol.* **13**, 107–110.

Stierhof, Y.-D., Humbel, B.M., Schwarz, H. (1991) Suitability of different silver enhancement methods applied to 1 nm colloidal gold particles: an immunoelectron microscopic study, *J. Electron Microsc. Tech.* **17**, 336–43.

Stierhof, Y.-D., Schwarz, H. (1989) Labeling properties of sucrose-infiltrated cryosections, *Scanning Microsc. Suppl.* **3**, 35–46.

Stierhof, Y.-D., Humbel, B.M., Hermann, R., Otten, M.T., Schwarz, H. (1992) Direct visualization and silver enhancement of ultra-small antibody-bound gold particles on immunolabeled ultrathin resin sections, *Scanning Microsc.* **6**, 1009–1022.

Stinchcombe, J.C., Nomoto, H., Cutler, D.F., Hopkins, C.R. (1995) Anterograde and retrograde traffic between the rough endoplasmic reticulum and the Golgi complex, *J. Cell Biol.* **131**, 1387–401.

Stoorvogel, W. (1998) Analysis of the endocytic system by using horseradish peroxidase, *Trends Cell Biol.* **8**, 503–505.

Stowell, M.H., Miyazawa, A., Unwin, N. (1998) Macromolecular structure determination by electron microscopy: new advances and recent results, *Curr. Opin. Struct. Biol.* **8**, 595–600.

Studer, D., Gnaegi, H. (2000) Minimal compression of ultrathin sections with use of an oscillating diamond knife, *J. Microsc.* **197**, 94–100.

Studer, D., Michel, M., Müller, M. (1989) High pressure freezing comes of age, *Scanning Microsc. Suppl.* **3**, 253–269.

Subramaniam, S., Henderson, R. (2000) Crystallographic analysis of protein conformational changes in the bacteriorhodopsin photocycle, *Biochim. Biophys. Acta* **1460**, 157–165.

Svitkina, T.M., Borisy, G.G. (1998) Correlative light and electron microscopy of the cytoskeleton of cultured cells, *Methods Enzymol.* **298**, 570–592.

Svitkina, T.M., Bulanova, E.A., Chaga, O.Y., Vignjevic, D.M., Kojima, A. Vasiliev, J.M., Borisy, G.G. (2003) Mechanism of filopodia initiation by reorganization of a dendritic network, *J. Cell Biol.* **160**, 409–421.

Szczesny, P.J., Walther, P., Müller, M. (1996) Light damage in rod outer segments: the effects of fixation on ultrastructural alterations, *Curr. Eye Res.* **15**, 807–814.

Takizawa, T., Robinson, J.M. (2000) Freeze-fracture cytochemistry: a new fracture-labeling method for topological analysis of biomembrane molecules, *Histol. Histopathol.* **15**, 515–522.

Tamaki, H., Yamashina, S. (1994) Improved method for post-embedding cytochemistry using reduced osmium and L.R. white resin, *J. Histochem. Cytochem.* **43**, 965–966.

Tiedemann, J., Hohenberg, H., Kollmann, R. (1998) High-pressure freezing of plant cells cultured in cellulose microcapillaries, *J. Microsc.* **189**, 163–171.

Tokuyasu, K.T. (1973) A technique for ultracryotomy of cell suspensions and tissues, *J. Cell Biol.* **57**, 551–565.

Tokuyasu, K.T. (1978) A study of positive staining of ultrathin frozen sections, *J. Ultrastruct. Res.* **63**, 287–307.

Tokuyasu, K.T. (1980) Immunochemistry on ultrathin frozen sections, *Histochem. J.* **12**, 381–403.

Tokuyasu, K.T. (1986) Application of cryoultramicrotomy to immunocytochemistry, *J. Microsc.* **143**, 139–149.

Tokuyasu, K.T. (1989) Use of poly(vinylpyrrolidone) and poly(vinyl alcohol) for cryoultramicrotomy, *Histochem. J.* **21**, 163–171.

Torrisi, M.R., Mancini, P. (1996) Freeze-fracture immunogold labeling, *Histochem. Cell Biol.* **106**, 19–30.

Trachtenberg, S. (1998) A fast-freezing device with a retractable environmental chamber, suitable for kinetic cryo-electron microscopy studies, *J. Struct. Biol.* **123**, 45–55.

Unwin, N. (1998) The nicotinic acetylcholine receptor of the torpedo electric ray, *J. Struct. Biol.* **121**, 181–190.

Unwin, N. (2002) Structure of the acetylcholine-gated channel, *Novartis Found. Symp.* **245**, 5–15.

Van Harreveld, A., Crowell, J. (1964) Electron microscopy after rapid freezing on a metal surface and substitution fixation, *Anat. Rec.* **149**, 381–385.

Van Heel, M., Gowen, B., Matadeen, R., Orlova, E.V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M., Patwardham, A. (2000) Single-particle electron cryomicroscopy: towards atomic resolution, *Rev. Biophys.* **33**, 307–369.

Villiger, W. (1993) Low Temperature Embedding Using the Lowicryl Resins, in: Robarts, A.W., Wilson, A.J. (Eds.) *Procedures in Electron Microscopy*, Module 16 : 7, John Wiley & Sons, Chichester, West Sussex, UK, pp. 7.1–7.24.

Vollenweider, H.J., Sogo, J.M., Koller, T. (1975) A routine method for protein-free spreading of double- and single-stranded nucleic acid molecules, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 83–87.

Walther, P. (2003) The potential of high resolution Cryo-SEM in life science, *Hitachi Instrument News*, **40**, 3–8.

Walther, P., Chen, Y., Malecki, M., Steffen Zoran, S.L., Schatten, G.P., Pawley, J.B. (1993) Scanning electron microscopy of high-pressure-frozen sea urchin embryos, *Scanning Microsc.* **7**, 1283–1293.

Walther, P., Müller, M. (1997) Double-layer coating for field-emission cryo-scanning electron eicroscopy – present state and applications, *Scanning* **19**, 343–348.

Walther, P., Müller, M. (1999) Biological ultrastructure as revealed by high resolution cryo-SEM of blockfaces after cryo-sectioning, *J. Microsc.* **196**, 279–287

Walther, P., Wehrli, E., Hermann, R., Müller, M. (1995) Double layer coating for high resolution low temperature SEM, *J. Microsc.* **179**, 229–237.

Watt, I.M. (1997) *The Principles and Practice of Electron Microscopy*, 2nd edition, Cambridge University Press, Cambridge.

Webster, P. (1998) The Production of Cryosections Through Fixed and Cryoprotected Biological Material and Their Use in Immunocytochemistry, in: Hajibagheri, N. (Ed.) *Methods*

in Molecular Medicine Biology, Humana Press, Totowa, NJ, pp. 1–18.

Weibel, E.R. (1980) Stereological Methods, *Practical Methods for Biological Morphology*, Vol. 1, Academic Press, New York.

Wendt, T.G., Volkmann, N., Skiniotis, G., Goldie, K.N., Muller, J., Mandelkow, E., Hoenger, A. (2002) Microscopic evidence for a minus-end-directed power stroke in the kinesin motor ncd, *EMBO J*. **21**, 5969–5978.

Wepf, R., Aebi, U., Bremer, A., Haider, M., Tittmann, P., Zach, J., Gross, H. (1994) High Resolution SEM of Biological Macromolecular Complexes, in: Bailey, G.W., Garratt-Reed, A.J. (Eds.) *Proceedings of the 52nd Annual Meeting of the Microscopy Society of America*, San Francisco Press, San Francisco, pp. 1026–1027.

Westmoreland, B.C., Szybalski, W., Ris, H. (1969) Mapping of deletions and substitutions in heteroduplex DNA molecules of bacteriophage lambda by electron microscopy, *Science* **163**, 1343–1348.

Ziegler, A. (2002) X-ray microprobe analysis of epithelial calcium transport, *Cell Calcium* **31**, 307–21.

Zierold, K. (1991) Cryofixation methods for ion localization in cells by electron probe microanalysis: a review, *J. Microsc.* **161**, 357–366.

Electron Microscopy of Biomolecules

Claus-Thomas Bock[1], *Susanne Franz*[2], *Hanswalter Zentgraf*[2], *and John Sommerville*[3]
[1] *Department of Molecular Pathology, Institute of Pathology, University of Tübingen, Germany*
[2] *German Cancer Research Center, Applied Tumor Virology, Electronmicroscopy, Heidelberg, Germany*
[3] *School of Biology, University of St. Andrews, St. Andrews, United Kingdom*

1 Principles 91

2 Techniques 92
2.1 General 92
2.2 Nucleic Acids 92
2.3 Chromatin 95
2.4 Proteins 99
2.5 Macromolecular Assemblies 103

3 Applications 105
3.1 Use of Antibodies 105
3.2 Detection of Radiolabeled Molecules 107
3.3 Electron Spectroscopic Imaging (ESI) 109
3.4 Image Reconstruction 110

Bibliography 110
Books and Reviews 110
Primary Literature 110
DNA Spreading 110
Long DNA Molecules 110
DNA–DNA Heteroduplex 110
DNA–RNA Heteroduplex 111
Small Chromatin Units 111
Large Chromatin Units (transcriptionally inactive chromatin) 111
Large Chromatin Units (transcriptionally active chromatin) 112

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Negative Staining 112
Immunolabelling of Spread Preparations 112
Immunoelectron microscopy (cell biology and viruses) 113
Image Reconstruction 113
Electron Spectroscopic Imaging (ESI) 114

Keywords

Cryoelectron Microscopy
Imaging of unfixed, unstained biomolecules in a hydrated state after rapid freezing to low temperature (−160 °C).

Heavy-metal Shadowing
Evaporation of a film of heavy metal (e.g. platinum–palladium or tungsten–tantalum) onto a dehydrated preparation. The deposit of metal particles around the biomolecules improves contrast and gives a shadowed, three-dimensional appearance.

High-resolution Autoradiography (EM ARG)
Detection of radiolabeled molecules by coating a stained or shadowed preparation with a film of photographic emulsion. Radioactive emissions hit silver halide crystals in the emulsion, which can be developed into silver grains.

Immunoelectron Microscopy
The application of antibodies to map specific sites on biomolecules. The antibody molecules may be visible after shadowing or negative staining, although detection can be improved by binding of a secondary antibody linked to an electron-dense tag (e.g. ferritin or colloidal gold particles).

Miller Spread
Deposition of dispersed chromatin onto a carbon-coated grid by centrifugation through a denser phase containing sucrose and formaldehyde.

Negative Staining
Instead of staining the biomolecules themselves (positive staining), a solution of heavy-metal salt (e.g. uranyl acetate or sodium phosphotungstic acid) is deposited in the hydrated spaces around and within the molecules.

Replica Casting
Adsorption of biomolecules to a mica surface followed by heavy-metal shadowing and coating of the preparation with a film of carbon. The carbon–metal replica (minus the biomolecules) is then removed from the mica and is mounted on an electron microscope grid for viewing.

Support Film
A thin film of plastic or carbon that is attached to an EM grid and provides a substrate onto which biomolecules can be adsorbed.

Electron microscopy (EM) is a method appropriate for examining details of the sizes and shapes of biological macromolecules and is particularly useful in studying isolated or reconstituted macromolecular assemblies. Small amounts of material (often <1 μg) can be used and prepared in a state suitable for viewing in as short a time as several minutes. The basic procedure involves adsorption of the biomolecules onto a support film, followed by staining or shadowing of the preparation and viewing of the dehydrated molecules *in vacuo* in an electron beam. A resolution of 1 to 2 nm is routine with conventional microscopes. The basic procedure can be adapted to give information about sites of specific epitopes, location of newly synthesized components, internal structures, and atomic composition.

Structures most suitable for EM analysis are nucleoprotein complexes, including ribosomes, spliceosomes, nucleosomes, and virus particles. DNA and RNA molecules are also suitable, and their lengths can be directly related to the number of base pairs or nucleotide residues determined biochemically. In general, it is difficult to visualize small proteins (<50 000 Da), but good detail can be obtained if they are isolated as multimeric complexes or if they can be induced to form filaments or crystalline arrays.

A wide range of applications are available using EM techniques, including virus identification, mapping of hybridized regions in heteroduplexes, detailing of macromolecular interaction, and analysis of the organization of molecular components in replication, transcription, splicing, and translation complexes.

1 Principles

The transmission electron microscope (TEM) consists of a metal column from which air is evacuated and through which a linear beam of electrons is accelerated and focused by electromagnetic lenses. The biomolecules are adsorbed onto a support film, stained and dehydrated (or frozen in an aqueous film on the grid), and introduced into the electron beam through an air lock. Whereas some of the electrons collide with atoms in the specimen, lose energy, and are scattered, the remaining electrons pass through the preparation and are focused to form an image on a phosphorescent screen (for direct viewing) or on a photographic plate (for later examination). Under ideal conditions, a resolution of 0.1 to 0.2 nm can be obtained; however, limitations are imposed by the naturally low masses of atoms (primarily hydrogen, carbon, and oxygen) contained in biomolecules, by distortions and artifacts created during sample preparation, and by radiation damage. Techniques are designed with the following objectives:

1. To minimize distortion by immobilization of the molecules, in their native state, onto an appropriate support.
2. To stabilize molecular complexes by suitable chemical fixation.
3. To improve contrast by staining the preparation with heavy-metal salt or by shadowing with heavy metal.

Irrespective of the method adopted, data derived from the electron microscopic examination of biomolecules should be

entirely consistent with the known biochemical and biophysical properties of the particles. Whenever possible, apparent sizes should be checked by independent measurement of sedimentation rate, electrophoretic mobility, or gel filtration elution. Also, features revealed by EM of isolated molecules should be compared with observations made on them *in situ*, by EM of cell or tissue sections.

2 Techniques

2.1 General

Similar principles of sample preparation (Fig. 1) apply for nucleic acids, proteins, and nucleoprotein complexes, since most of these are less than 20 nm thick and can be adsorbed directly from solution onto the support matrix. The concentration of molecules must be high enough to permit several examples to be viewed together in one field. The efficiency of uptake onto the support is not always predictable, and a range of initial concentrations should be tried. It is important that the biomolecules be held in a solution known to maintain the proper structural features directly prior to applying to the support.

The grid consists of a fine meshwork, usually of copper, which must be coated with a thin support film of plastic (e.g. collodion, parlodion) or carbon. Carbon films are preferable because they are thin (down to 2 nm) and contribute little to the image. However, they are frequently hydrophobic and should be subjected to ionizing gases (by "glow discharging") or should be treated chemically (e.g. with Alcian blue 8GX) to render them hydrophilic before use.

2.2 Nucleic Acids

Nucleic acids are used at a concentration of 1 to 5 μg/μL, and as little as 0.1 μg is required; detailed procedures for working with these molecules may be found in the literature. The molecules can be native DNA or RNA, cloned DNA, single-stranded or double-stranded forms, partially denatured duplexes, or denatured and hybridized structures (Figs. 2–8). A

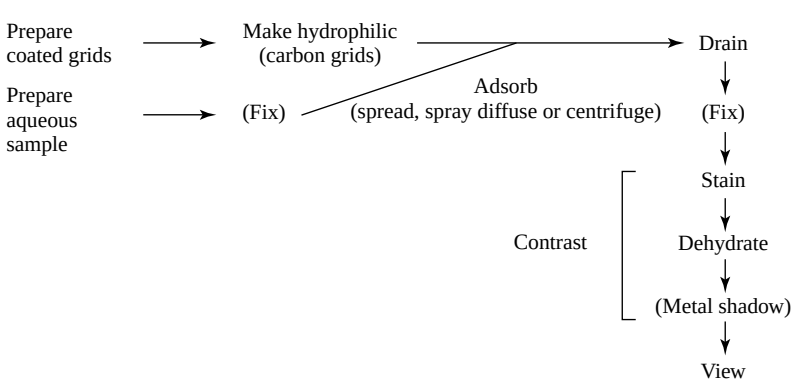

Fig. 1 Flow diagram of the basic procedure. Steps in parentheses are not always used.

key step in their preparation for EM is the spreading out of the molecules so that they can be adsorbed in an extended, nonaggregated form onto the support film. The most successful method of spreading is that devised originally by Kleinschmidt and Zahn (Table 1). The negatively charged nucleic acid is mixed with a basic protein (usually cytochrome c) in a solution appropriate for maintaining the required conformation. This spreading mix, or hyperphase, is spread, via a glass ramp, onto the surface of a second solution, the hypophase, and a molecular monolayer of nucleic acid and protein is formed at the liquid–air interface. Alternatively, the molecular monolayer can be formed by diffusion to the surface of the spreading mix contained in a small vessel or even in a droplet sitting on a hydrophobic surface (Table 2). The nucleic acid–protein complex is adsorbed through brief contact with the support film, stained with ethanolic uranyl acetate, rinsed in ethanol, and air-dried. Protein-free spreading is applied in special circumstances, for instance, to visualize very large DNA molecules that have been stabilized by polyamines (Table 3, Fig. 5). To improve contrast, spread preparations are rotary-shadowing, usually with platinum–palladium, at an angle of 6 to 9° from the plane of the metal vapor. It should be noted that although metal shadowing greatly exaggerates the thickness of the molecules, double-stranded and single-stranded regions can still be differentiated (Figs. 6 and 7).

Cytochrome c should not be used in spreading solutions when experimentally bound proteins are being studied

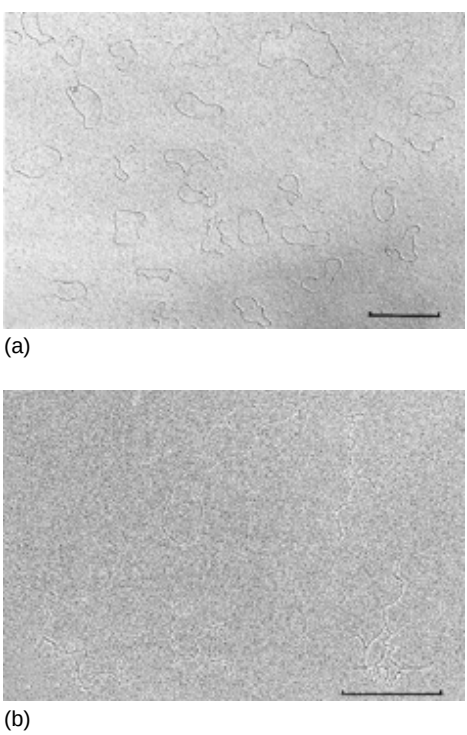

Fig. 2 (a) Simian virus 40 (SV40) DNA spread in the presence of cytochrome c down a glass slide and across a water hypophase. (b) *Xenopus laevis* ribosomal DNA–derived plasmid DNA spread by the cytochrome c microdiffusion droplet technique. The DNA–protein films were picked up on parlodion-coated (a) and carbon-coated (b) copper grids. The use of carbon-coated grids yields a significantly higher proportion of supercoiled molecules and results in a smoother background than is the case for parlodion-coated grids. However, the attachment of molecules to the grid is significantly higher with parlodion. For contrast enhancement, the specimens were rotary-shadowed with platinum/palladium (80 : 20) at an angle of 8°. Note that an additional, very brief, unidirectional shadowing in (a) leads to further contrast enhancement. The bars represent 1 μm.

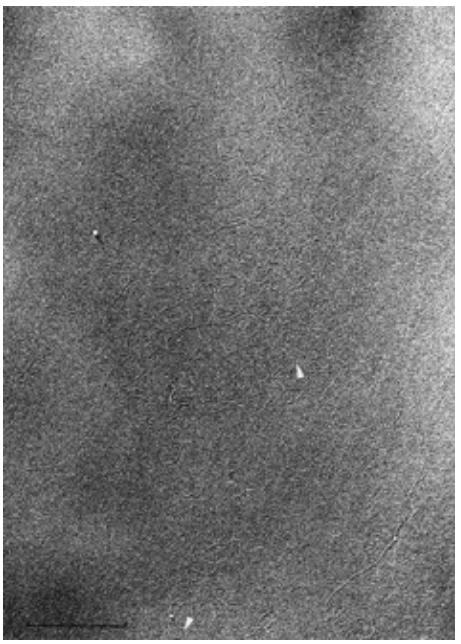

Fig. 3 Herpes simplex virus (HSV) DNA as revealed by the cytochrome c microdiffusion droplet technique, picked up on a carbon-coated grid and shadowed as described in Fig. 2a. Because shearing and stretching forces are minimal when the droplet technique is used, it is possible to spread this extremely large viral genome (150–160 kbp) to its full length. Bacteriophage PM2 circular DNA serves as size markers. The ends of the HSV genome are denoted by arrowheads. Bar represents 1 µm.

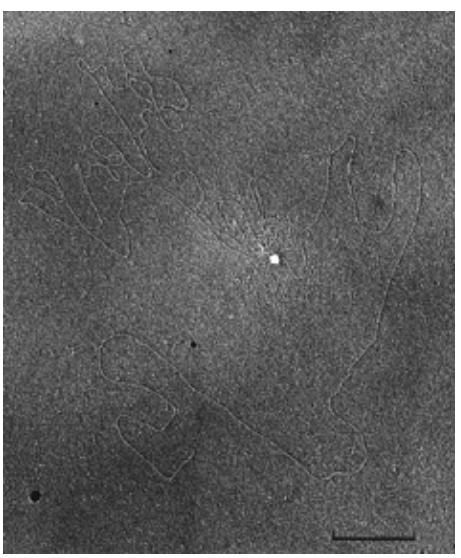

Fig. 4 The "unraveling" of the HSV genome out of the viral core as shown by cytochrome c spreading. Complete intact virus particles (see Fig. 17d) were briefly lysed in water and immediately spread onto a water hypophase. The white dotlike structure anchoring the spread DNA represents the unraveled part of the genome, surrounded by fuzzy material, most probably remnants of the viral core. Bar represents 1 µm.

because it obscures fine details of nucleic acid structure. Thus, to permit detection of proteins bound to specific regions of nucleic acid molecules, other protein-free spreading methods have been devised (Fig. 8e). These methods employ low–molecular weight inorganic substances (e.g. benzyldimethylalkylammonium chloride or ethidium bromide) in place of cytochrome c. In working with

Tab. 1 Sequence of steps for spreading nucleic acids.

1. Mix nucleic acid,[a] cytochrome c, and buffer.[b]
2. Pour hypophase solution[c] into a spreading trough.
3. Insert a glass slide in the trough with one end resting on the rim.
4. Apply spreading solution on the glass slide.
5. Wait until solution has run down the ramp and spread out.
6. Touch the film side of a grid[d] to the surface of the hypophase.
7. Stain the grid.[e]
8. Dehydrate the specimen.
9. Rotary shadow for contrast enhancement.[f]

[a]This can be dsDNA (Figs. 2a and 4), ssDNA (Fig. 6a), RNA (Fig. 6c), or heteroduplex molecules (Fig. 7).
[b]Spreading of ssDNA, RNA, or heteroduplex molecules needs denaturing agents, such as formamide and/or urea (Figs. 6 and 7).
[c]Usually, water or ammonium acetate is used; for highly denaturing conditions, formamide can be included.
[d]Normally, carbon- or parlodion-coated grids are used (Fig. 2).
[e]Ethanolic uranyl acetate is widely used.
[f]Usually, platinum/palladium (80 : 20) is used for rotary shadowing at an angle of 6 to 9°; specimens can also be viewed without rotary shadowing (Fig. 8c).
Note: These preparation steps are also used with slight modifications for protein-free spreading procedures (e.g. in the presence of protamines, Fig. 5, or benzyldimethylalkylammonium chloride, Fig. 8e).

Tab. 2 Sequence of steps for spreading nucleic acids: microdiffusion droplet technique.

1. Mix nucleic acid, cytochrome c, and buffer.[a]
2. Put a small drop of the solution on a hydrophobic surface.
3. Leave for 10 to 30 min for diffusion.
4. Pick up nucleic acid–protein film by allowing the grid[b] to touch the surface of the drop.
5. Stain, dehydrate, and shadow.

[a]The procedure yields best results with double-stranded nucleic acids (Figs. 2b, 3, and 8a–d). For spreading single-stranded nucleic acids, denaturing agents, such as formamide and/or urea, have to be added to the solution.
[b]Carbon- or parlodion-coated grids may be used (Fig. 2).

protein molecules bound to nucleic acids, care must be taken to stabilize the complexes before spreading by cross-linking the molecules, in solution, with glutaraldehyde (Fig. 8) or formaldehyde (for chromatin, see Section 2.3). Good detail of the association can be obtained by making a carbon–metal replica of the molecules bound to the surface of freshly cleaved mica. The use of platinum–carbon or tungsten for shadowing reveals finer detail.

2.3 Chromatin

Chromatin spreading techniques are mostly adapted from the procedure devised by Miller and coworkers (Table 4). Large chromatin units are obtained by

Tab. 3 Sequence of steps for protein-free spreading of long DNA molecules.[a]

1. Dilute DNA (0.25–1.6 Mbp) to 4 to 10 ng/µL in microinjection buffer.[b]
2. Apply a 10 µL sample[c] to a freshly glow-discharged carbon-coated grid.
3. Allow DNA to adsorb for 1 min and remove excess liquid by touching the drop with the edge of a filter paper.
4. Immediately add a 20 µL drop of 2% (w/v) aqueous uranyl acetate solution and leave to stain for 20 s.
5. Remove staining solution by touching with the edge of the filter paper and air-dry grid.
6. Rotary shadow[d] and view.

[a]Appropriate for chromosomal lengths of yeast DNA and artificial chromosomes to check molecular integrity prior to microinjection (see Fig. 6).
[b]10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA (pH 8), 100 mM NaCl, 30 mM spermine, 70 mM spermidine.
[c]Transfer of DNA can be achieved without shearing by using a plastic pipette tip with the end cut off.
[d]For instance, platinum : palladium (80 : 20) at an angle of 5°.

Tab. 4 Sequence of steps for spreading large chromatin units: Miller spreading.

1. Incubate cells, isolated nuclei, chromosomes, or chromatin under low ionic strength conditions for chromatin dispersal.
2. Keep on ice for 15 to 30 min.
3. Glow-discharge carbon-coated grids.
4. Fill the well of the centrifugation chamber with a sucrose solution containing formaldehyde.
5. Insert the carbon-coated grid, which is now hydrophilic.
6. Layer a small volume of chromatin solution on top of the sucrose cushion.
7. Centrifuge the chamber (speed and time depend on the equipment used).
8. Remove grid and dry in Photoflo detergent.
9. Stain with phosphotungstic acid and dehydrate.
10. Rotary shadow.

Note: For details see Figs. 9, 10, and 11.

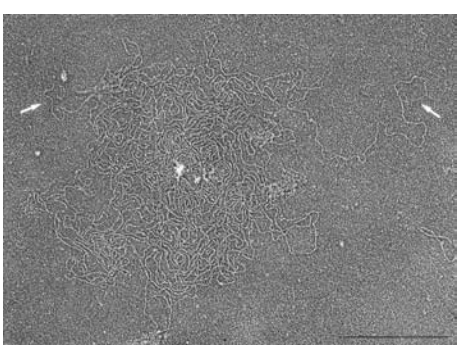

Fig. 5 Full-length chromosomal DNA molecules of approximately 1 Mb isolated by pulse-field gel electrophoresis from *Saccharomyces cerevisiae* strain AB1380 and spread from a buffer containing polyamines and sodium chloride (see Table 3). The long DNA fibres may be stabilized by a coating of polyamines in the presence of sodium chloride and appear to gather into compact bundles. The two ends of a single molecule are indicated by arrows. Bar represents 0.5 µm.

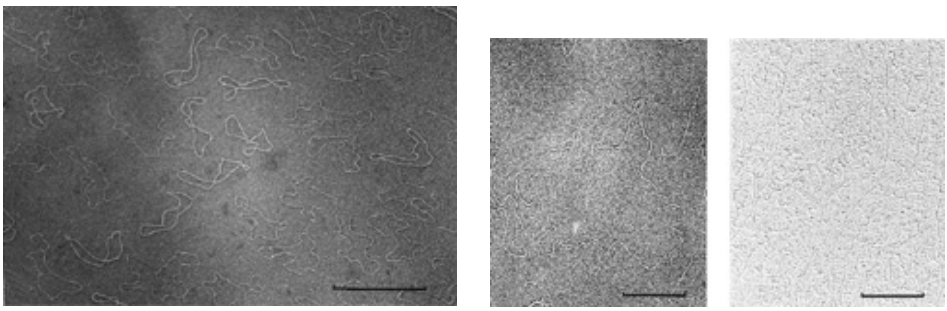

(a) (b) (c)

Fig. 6 (a) Double-stranded and single-stranded DNA of bacteriophage fd spread in 40% formamide, 1 mM EDTA, 10 mM Tris-HCl (pH 8.5) onto a water hypophase. The cytochrome c–DNA film was picked up on a parlodion-coated grid and was shadowed as described in Fig. 2b. The double-stranded DNA molecules are wider with smoother contours than the single-stranded molecules. (b) Replicative intermediate of plasmid pA2Y1, derived from a parvovirus, AAV (adeno-associated virus), spread by the microdiffusion technique in the presence of 10% formamide (parlodion-coated grids, rotary shadowing). The replication eye is denoted by an arrowhead. For contrast enhancement, (a) and (b) are printed in reverse contrast. (c) Tobacco mosaic virus (TMV) RNA was diluted in a spreading solution made from 4 M urea dissolved in "pure" formamide, 1 mM EDTA, and 10 mM Tris-HCl, (pH 8.5), heated to 80 °C (5 min), and spread on a 60 °C water hypophase, picked up on parlodion-coated grids, and shadowed. The bars represent 1 µm in (a) and 0.5 µm in (b) and (c).

Tab. 5 Sequence of steps for spreading small chromatin units.

Part A
1. Dilute chromatin for dispersal under conditions of low ionic strength.
2. Place a drop of chromatin solution on a hydrophobic surface.
3. Place a glow-discharged carbon grid on top of the drop.
4. After 15 min, remove the grid and place it on top of a drop of double-distilled water.
5. After 10 min, transfer it to a second drop of water.
6. After 10 min, remove the grid; negative-stain and air-dry the preparation.
7. Rotary shadow for contrast enhancement (see Figs. 12a, b, and d)
Part B
1. Hold a glow-discharged, carbon-coated grid with forceps.
2. Apply a drop of chromatin solution.
3. Allow the chromatin to adsorb onto the film for 1 to 2 min.
4. Stain and air-dry (alternatively shadow) (see Fig. 12c).

lysing cells or nuclei in a low salt concentration, high pH buffer (often referred to as "pH 9 water"). The chromatin is left to disperse and is then centrifuged at low speed, through a denser solution containing fixative (formaldehyde), onto the surface of a carbon-coated grid. The grid is dipped into a solution of wetting agent (Kodak Photoflo 200) and air-dried. The preparations can be positively stained (with ethanolic phosphotungstic acid), negatively stained (with aqueous uranyl acetate), or metal-shadowed. Chromatin spread from avian erythrocytes

shows the typical nucleosomal configuration of dispersed chromatin (Fig. 9). The features of actively transcribing chromatin can be seen in nucleolar (pre-rRNA) units (Fig. 10). In contrast, nonnucleolar transcription units normally contain more widely spaced nascent transcripts (Fig. 11b).

For small chromatin units, such as chromatin fragments and viral chromatin (minichromosomes), high-speed centrifugation is required to deposit the chromatin onto the coated grid. Alternatively, small chromatin units can be adsorbed directly by floating the grid, carbon film down, on the surface of a droplet containing the sample (Table 5). The nucleosomal configuration of simian virus 40 chromatin (the SV40 minichromosome) is shown in Figs. 12a to c. Supranucleosomal

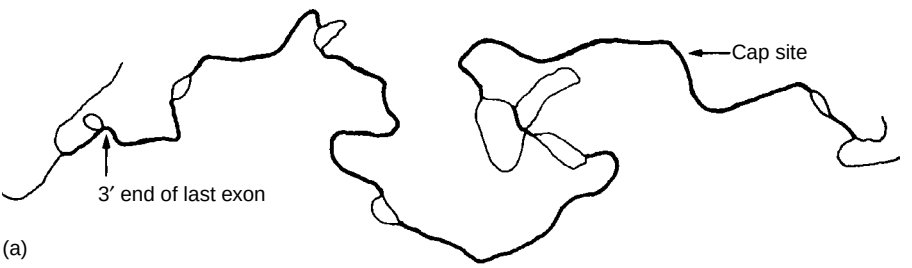

(a)

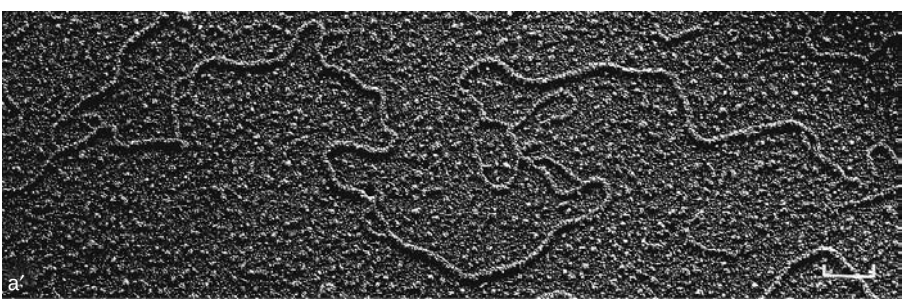

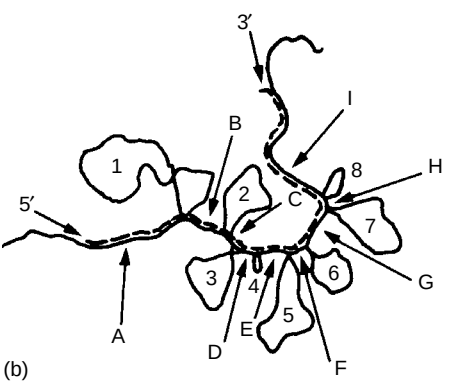

(b)

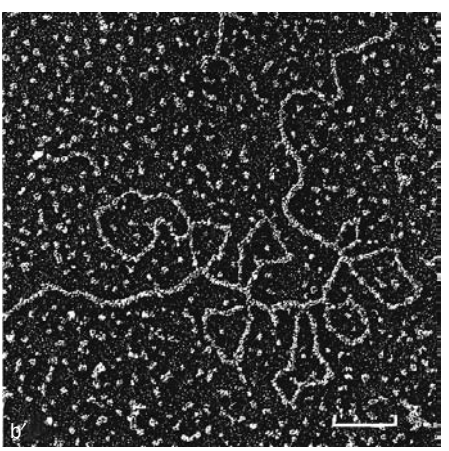

particles – for instance, those obtained by brief nuclease digestion of avian erythrocyte chromatin (Fig. 12d) – are also prepared by direct adsorption.

2.4 Proteins

The most commonly used method for contrast enhancement of proteins is negative staining (Table 6). This method is extremely rapid and involves drying of an aqueous solution of heavy-metal salt (e.g. uranyl acetate) with the protein molecules on the coated grid. The metal salt occupies hydrated regions in and around the biomolecules and, upon drying, forms a dense cast. Since the biomolecules themselves are not stained, but the background is stained, a negative image is created. This technique has been used extensively for examining protein filaments. Negatively stained intermediate filaments, in this example, glial fibrillary acidic protein (GFAP), are shown in Fig. 13a. Details of the arrangement of individual collagen molecules within the collagen fibril can be obtained from preparations like that seen in Fig. 14.

An alternative to the adsorption of proteins from droplets is to spray them, in aerosols, onto the surface of carbon-coated grids. The spraying is most easily achieved

Tab. 6 Sequence of steps for negative staining.

1. Hold a coated grid[a] with forceps.
2. Apply a drop containing biological material to the grid.
3. After 10 to 60 s, drain off excess liquid.
4. Apply a drop of negative stain[b] for 10 to 60 s.[c]
5. Remove liquid and air-dry the preparation.

[a]Carbon, parlodion, or sandwich coats (made from Formvar and carbon) may be used. Glow discharge increases the amount of biological material adsorbed onto the film; plastic films, however, are destroyed by glow discharge.
[b]Uranyl acetate and phosphotungstic acid are widely used for negative staining.
[c]Alternatively, the grid can be transferred through several drops of stain.
Note: For details on the various procedures, see Figs. 13–18.

Fig. 7 (a) A DNA : DNA heteroduplex formed between cloned mouse tyrosine aminotransferase (TAT) and rat TAT. Heteroduplexes were formed after denaturation in 0.1 M NaOH, 20 mM EDTA at 20 °C for 10 min and renaturation in the presence of 50% formamide, 20 mM Tris-HCl (pH 7.2) at 20 °C for 90 min. Heteroduplexes were then prepared from a spreading solution containing 4 M urea dissolved in "pure" formamide, 1 mM EDTA, 10 mM Tris-HCl, (pH 8.5) on water hypophase, followed by rotary shadowing. (a′) The interpretive drawing of (a): the positions of the cap and of the 3′ end of the last exon are indicated. (b) A heteroduplex formed between poly(A+) RNA from bovine muscle epidermis and a genomic clone that contains the gene coding for epidermal bovine keratin (KBla). Heteroduplexes were formed in 70% formamide, 0.3 M NaCl, 1 mM EDTA, 20 mM Tris-HCl (pH 8.0). The sample was kept for 17 h at 62 °C, followed by incubation at 64 °C for 2 to 4 h. The hybrids were prepared for electron microscopy as described in (a). (b′) The interpretative drawing of (b): DNA is represented by a continuous line, RNA by an interrupted line; the 5′ end of mRNA is identified by a change in molecular diameter and the 3′ end by a projecting [i.e. nonhybridized, poly(A) tail]. Exons are denoted by capital letters, introns by arabic numerals. For contrast enhancement, (a) and (b) have been printed as negatives. Bars represent 100 nm.

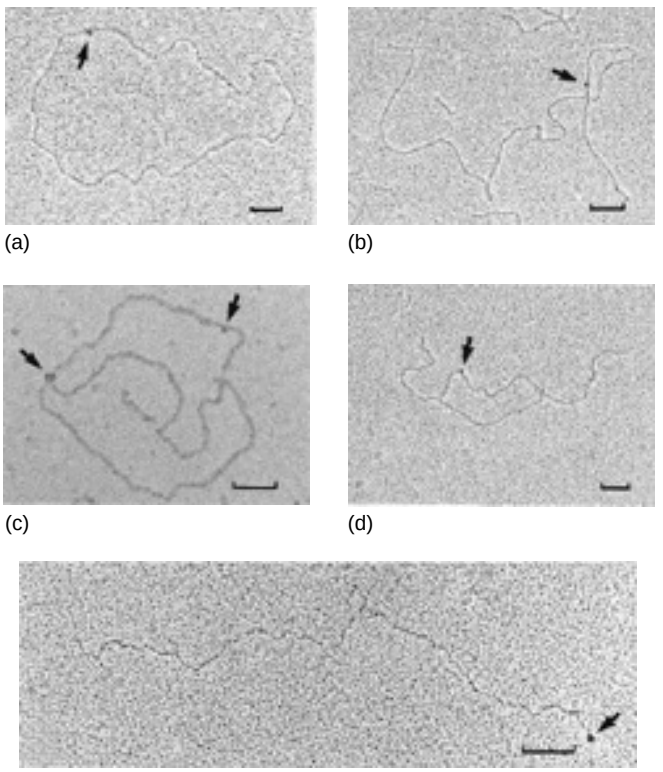

Fig. 8 (a, b) Binding of monoclonal Z-DNA antibodies to DNA from bovine papillomavirus (BPV-1). Antibodies against left-handed Z-DNA were incubated and cross-linked with 0.1% glutaraldehyde (2 h) to the pML2d-BPV-1 plasmid DNA (a). To determine the position of the binding site, the DNA was digested with the single-cutting restriction enzyme Xba I (b). Arrows indicate the position of Z-DNA antibodies on the plasmids. (c, d) Localization of large-tumor antigen (T antigen) on replicative intermediates of SV40 DNA. The molecules were immunostained with either ferritin-labeled protein A (c) or ferritin-labeled goat antimouse antibodies (d) after cross-linking with 0.1% glutaraldehyde (20 min, 37 °C). To localize T antigen on the replicated section, the complex was cleaved with the single-cutting enzyme Bgl I. Arrows indicate the ferritin-labeled T antigen on DNA. These complexes (a–d) were prepared for electron microscopy according to the cytochrome c droplet diffusion technique. The specimens were rotary-shadowed, except the one shown in (c), which is only stained by ethanolic uranyl acetate. (e) DNA-relaxing enzyme molecules linked to single-stranded SV40 DNA (arrow). After enzyme reaction, the complexes were spread from a solution containing 0.25 M NaOH, indicating that the protein dot is stable in alkali and is therefore linked via a covalent bond to the end of the single-stranded DNA. Enzyme–DNA complexes were prepared with a protein-free spreading procedure using benzyldimethylalkylammonium chloride (BAC) and were shadowed. Bars represent 100 nm.

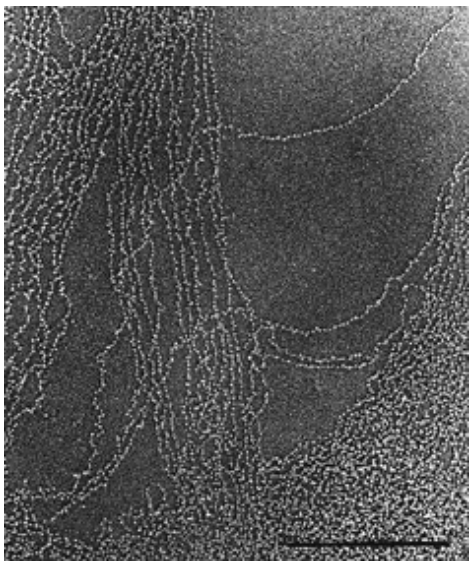

Fig. 9 Appearance of the nucleosomal chromatin configuration of chicken erythrocytes lysed and swollen in 0.5 mM sodium borate buffer (pH 8.8) for 5 min, spread by the Miller technique, positively stained, and metal-shadowed. The dispersed chromatin shows the characteristic nucleosomal "beads-on-a-string" organization. Bar represents 1 μm.

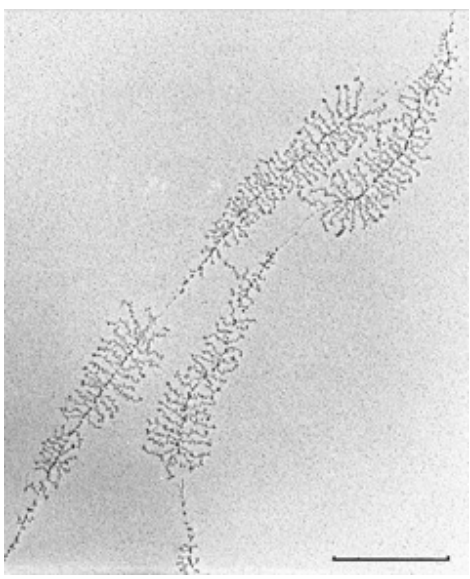

Fig. 10 Organization of nucleolar chromatin from oocytes of the newt *Pleurodeles waltlii*, after spreading using the Miller technique, positive staining, and rotary shadowing. The transcribed chromatin segments are densely packed with RNA polymerases and nascent ribonucleoprotein (RNP) fibrils. The tandemly arranged active pre-rRNA genes are interspersed by the spacer regions. The axis of the spacer is relatively thin and does not show nucleosome-sized particles. Bar represents 1 μm.

by touching the tip of a glass capillary tube, containing both sample and stain, into a stream of nitrogen. Elongate, or rod-shaped, proteins can be mixed with glycerol (and no stain) and sprayed onto the surface of a piece of freshly cleaved mica. The deposit is then shadowed finely and carbon-coated to produce a replica cast of the protein molecules. (Details are provided in the literature.) Unidirectionally

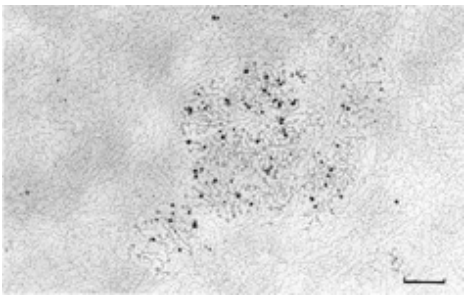

(a)

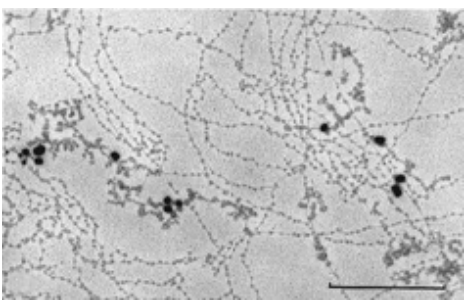

(b)

Fig. 11 High-resolution autoradiography of chromatin spread from mouse P815 cells after labeling for 5 min with [^{3}H]uridine. (a) Radioactivity (recorded as dense silver grains) associated with densely transcribing regions showing the gradient-like features of pre-rRNA transcription units (cf. Fig. 10). The background consists of many strands of nucleosomal chromatin. (b) Radioactivity associated with individual RNP fibrils characteristic of those found in nonnucleolar transcription units. In sparsely transcribing units, nucleosomes can be seen on the chromatin strand between the nascent transcripts. Both preparations were stained with phosphotungstic acid, rotary-shadowed with platinum, and exposed with Ilford L4 emulsion. Bars represent 0.5 µm. (Courtesy of S. Fakan, with permission from Springer-Verlag.)

Fig. 12 Small chromatin units such as SV40 chromatin complexes (a–c) or supranucleosomal particles from chicken erythrocyte chromatin (d) are prepared by direct adsorption to freshly glow-discharged, carbon-coated grids. The chromatin is allowed to adsorb onto the film for 1 to 2 min, stained with 2% uranyl acetate (c), and rotary-shadowed (a, b, d). Under spreading conditions involving buffers of low ionic strength (2 mM EDTA, 1 mM Tris-HCl, pH 8.4), the SV40 nucleoprotein complexes (''minichromosomes'') show the typical ''beads-on-a-string'' morphology of the circular nucleosomal chain (a, b). The foreshortening of DNA in the SV40 ''minichromosomes'' due to nucleosomal organization is demonstrated by comparison with the viral DNA, which is included in the preparation (b). This packaging of the SV40 genome, at first-order level into nucleosomes, results in a foreshortening ratio of about 5.5 : 1. (d) At the second level of packaging, chromatin is organized into supranucleosomal structures. Fractions of supranucleosomal granular subunits, obtained by brief digestion of chromatin with micrococcal nuclease at physiological salt concentration, can be spread after fixation for 15 min at 4 °C with glutaraldehyde (final concentration 0.2%). (e) Nonnucleosomal forms of chromatin beside nucleosomal chains are observed in Miller spread preparations of African green monkey kidney cells (RC37) after infection with herpes simplex virus (HSV). The ''bubblelike'' configurations, thickly coated with a single-strand DNA-binding protein, alternate with unbranched thin intercepts (staining and shadowing as in Figs. 9 and 10). Bars represent 500 nm in (a), (d), and (e), 200 nm in (b), and 50 nm in (c).

shadowed molecules, prepared by the spray/replica technique and representing early stages in the assembly of GFAP filaments, are shown in Fig. 13b.

2.5 Macromolecular Assemblies

Multimeric enzyme complexes, ribosomes, viruses, and other particles can be treated in ways similar to those described for protein filaments or for small chromatin units. Particles in the size range of 10 to 200 nm diameter are excellent targets for negative staining (Figs. 15–18), which often reveals considerable detail of surface structure (Fig. 17). Nevertheless, the size of the particles of dried stain (or of the metal particles after shadowing) limits the resolution of structural

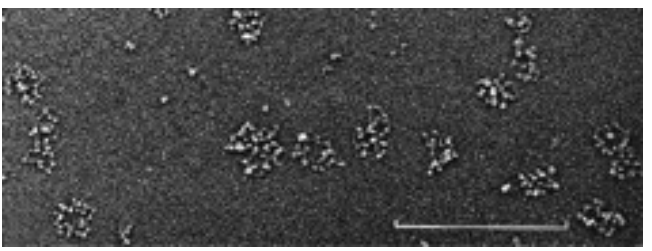

(a)

(b)

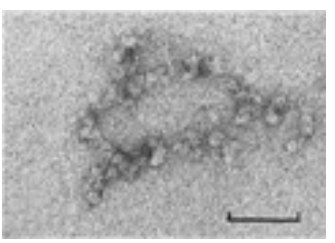

(c)

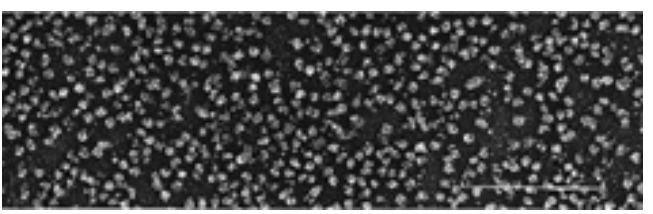

(d)

(e)

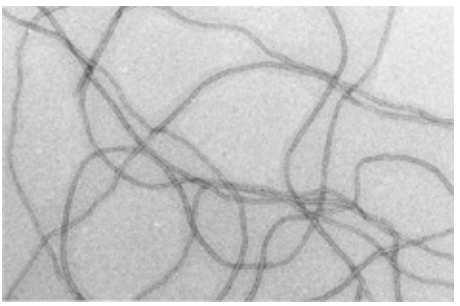

(a)

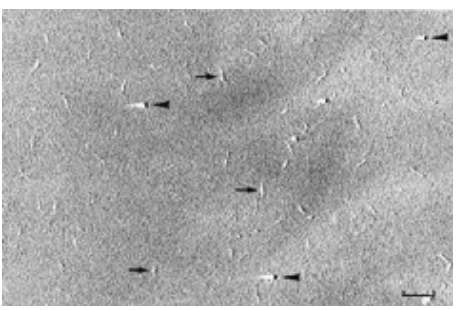

(b)

Fig. 13 Intermediate filaments as visualized by negative staining and partly assembled monomers as visualized by metal shadowing. (a) GFAP intermediate-sized filaments spread onto carbon film and negatively stained with 1% w/v uranyl acetate. GFAP was prepared from spinal cord and assembled *in vitro* by dialysis against the following buffer: 10 mM Tris-HCl (pH 7.0), 1 mM $MgCl_2$, 50 mM NaCl, and 25 mM 2-mercaptoethanol. Note that the intermediate filaments are 10 nm in width and are long with smooth edges. (b) The assembly of intermediate filaments can be arrested at various points along the assembly pathway by choosing the appropriate buffer conditions. In this preparation, GFAP molecules containing chains of four proteins were formed in 10 mM Tris-HCl (pH 8.5). A sample of this was made 50% v/v in glycerol, sprayed onto freshly cleaved mica, and unidirectionally shadowed at 10° with platinum. The replica was then floated on water and picked up onto grids. The molecules of GFAP appear as short rodlets, between 45 and 65 nm long and 2 to 3 nm wide (arrows). Latex beads demonstrate the direction of shadow and locate the position of the droplets on the replica (arrowheads). Both micrographs are shown at the same magnification; bar represents 100 nm. (Courtesy of R. Quinlan and A. M. Hutcheson.)

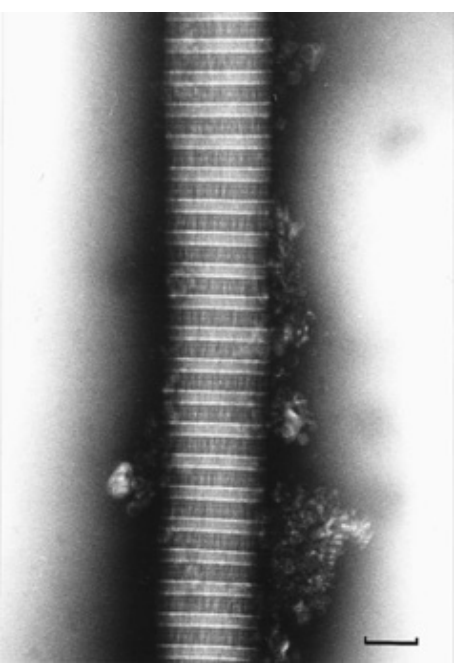

detail. A superior technique cryoelectron microscopy involves the rapid freezing of a thin (100 nm) aqueous film containing the specimen particles. Below −143 °C, water can be held in a vitrified state, which resembles the liquid state and avoids formation of ice crystals. In this condition, macromolecules and macromolecular assemblies can be viewed free of artifacts from fixation, dehydration, and staining. Under-focus phase contrast is used to produce a high-resolution image.

Fig. 14 A collagen fibril, negatively stained with 4% phosphotungstic acid, on a glow-discharged, carbon-coated grid. The staggered arrangement of the collagen molecules in the fibril results in the striated appearance after negative staining. The bars represent 100 nm.

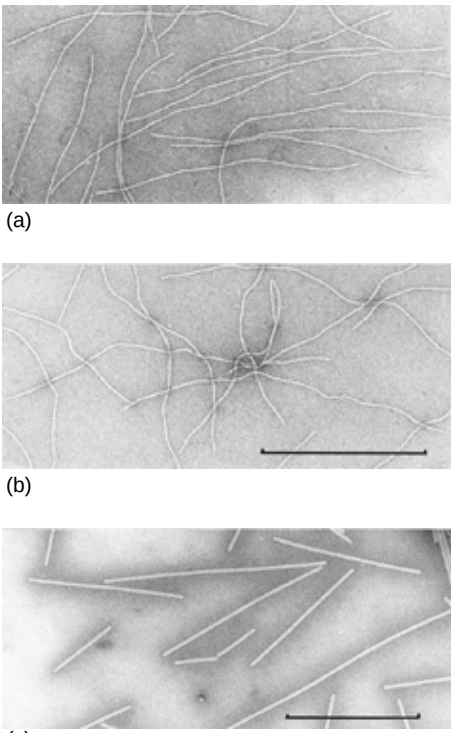

Fig. 15 Examples of filamentous bacteriophages and rodlike viruses. The bacteriophage fd was negatively stained with either 4% phosphotungstic acid (a) or 2% uranyl acetate (b). The tobacco mosaic viruses in (c) are stained with 4% phosphotungstic acid. All preparations were made on carbon-coated, glow-discharged grids. The bars represent 0.5 µm.

3 Applications

3.1 Use of Antibodies

Antibody molecules (IgG and IgM) are large enough to be resolved in metal-shadowed or negatively stained preparations. Monoclonal antibodies can be used to map sites on nucleic acid or protein molecules and to identify individual components within macromolecular assemblies. Antibodies can also be used to build up denser structures around small

Tab. 7 Sequence of steps for immunogold labeling: identification of individual components in macromolecular assemblies.

1. Hold a coated grid[a] with forceps.
2. Apply a drop containing biological material to the grid.
3. After 1 min, wash the grid with 20 drops of double-distilled water from a Pasteur pipette and drain off excess liquid.[b]
4. Place the grid (film side down) on a drop[c] of the primary antibody diluted in phosphate-buffered saline (PBS), 1% bovine serum albumin (BSA).[d]
5. After 30 min, remove the grid and place it on top of a drop of PBS, 1% BSA.
6. After 5 min, transfer the grid to a fresh drop of PBS, 1% BSA; repeat this washing step a third time.
7. After 5 min, transfer the grid to a drop of the gold-labeled secondary antibody[e] diluted in PBS, 1% BSA.
8. After 30 min, place the grid on top of a drop of PBS, 1% BSA and continue as described in step 6.
9. Wash the grid as described in step 3.
10. Stain the specimen with 2% aqueous uranyl acetate for 1 min and air-dry.

[a]The procedure yields best results with glow-discharged carbon-coated grids.
[b]Never drain the grids completely.
[c]Droplets are placed on a piece of Parafilm.
[d]PBS, 1% BSA; filtered through a 0.22 µm Millipore filter.
[e]Alternatively, a protein A–coated colloidal gold probe may be used.

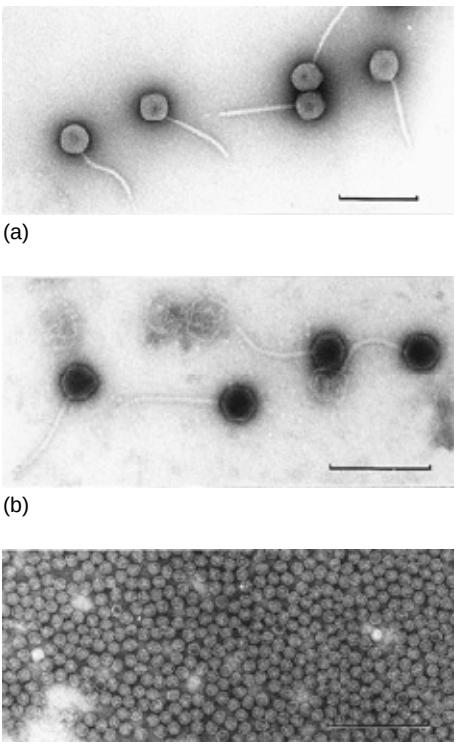

Fig. 16 Examples of tailed (a, b) and globular (c) bacteriophages. Bacteriophage λ (a) and the globular bacteriophage fr (c) were stained with 4% uranyl acetate. The bacteriophage T5 (b) was stained with 2% uranyl acetate. Note the empty phage heads in (b). All preparations were made on freshly glow-discharged, carbon-coated grids. The bars represent 200 nm.

proteins, which are themselves difficult to resolve. Secondary antibodies tagged with electron-dense markers such as ferritin (Figs. 8c and d) or colloidal gold particles are routinely used to improve detection (Table 7). For instance, gold-conjugated secondary antibodies can be used to detect a range of epitopes in recombinant viral coat particles (Fig. 18). A quite different application is the use of anti-biotin to detect biotinylated nucleic acid probes that have been hybridized to complementary sequences in chromatin preparations or *in situ* in isolated chromosomes. In the example shown in Fig. 19, two different antibody-binding sites are enhanced for EM detection by addition of a common secondary antibody tagged with colloidal gold particles. In this way, specific genes can be detected within complex chromatin masses.

Tab. 8 Sequence of steps for high-resolution autoradiography of biomolecules.

1. Radioactively label[a] living cells or subcellular fractions *in vitro*.
2. Isolate molecules or molecular complexes and prepare by direct adsorption on the grid or by spreading techniques.
3. Stain,[b] dehydrate, and rotary shadow.
4. Evaporate a thin layer of carbon (4–8 nm) on the specimen.[c]
5. Coat the specimen with a layer of photographic emulsion.
6. Follow the gold latensification–Elon–ascorbic acid (GEA) procedure.[d]
7. Transfer to Elon–ascorbic acid developer.
8. Transfer into the fixing bath.
9. Wash and air-dry.

[a]Generally, radioactive tritium is used.
[b]Negative or positive staining; shadowing with platinum/palladium.
[c]The fine coat prevents chemical interactions between specimen and emulsion.
[d]The GEA development procedure increases the sensitivity and gives rise to silver grains of rather small size.

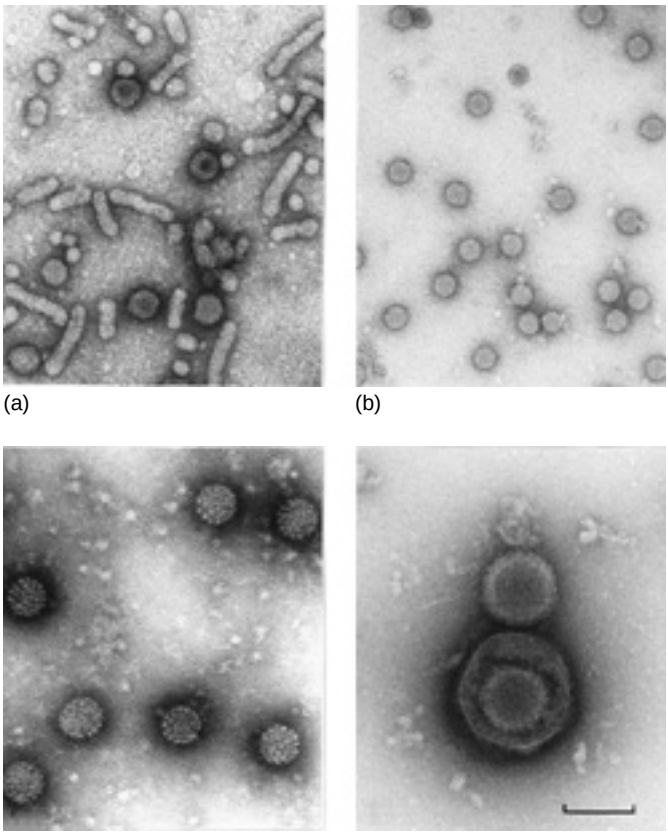

Fig. 17 Details of virus structure revealed in negatively stained preparations on carbon-coated, glow-discharged grids. (a) Hepatitis B viruses (HBV) were stained with 2% uranyl acetate. (b) Polioviruses and (c) bovine papilloma viruses (BPV) were stained with 4% phosphotungstic acid. (d) Herpes simplex virus (HSV) was stained with 2% uranyl formate. HBV was obtained from a patient's serum and reveals, in addition to the mature "Dane" particles, globular and filamentous forms of the viral S protein (a). Polioviruses show a fine-textured surface (b), while BPVs reveal a moruloid structure, which is due to the capsomere architecture (c). This organization is also visible in the HSV capsid, even when surrounded by the envelope (d). The micrographs are magnified to the same scale to give an impression of the size difference between the virus families. Bar represents 100 nm.

3.2 Detection of Radiolabeled Molecules

Although high-resolution autoradiography (EM ARG) is applied mostly to the detection of newly synthesized (radiolabeled) components *in situ* at the cellular level, application is possible with spread molecules and molecular complexes. However, EM ARG is limited to the use

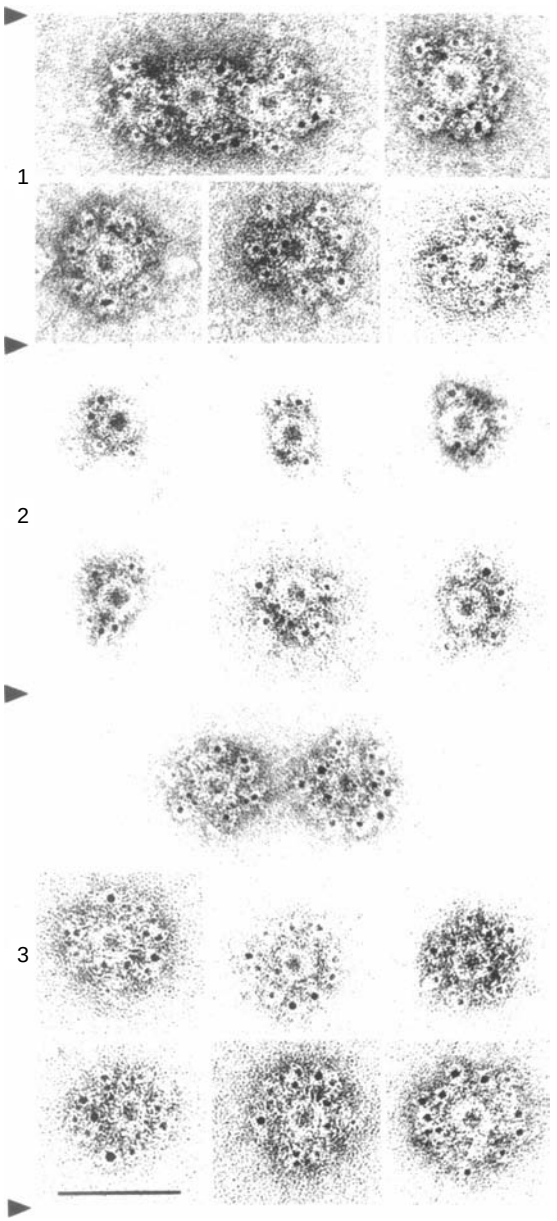

Fig. 18 Detection of different epitopes in recombinant viral core particles. The vector cassette pB–CHG–HBc (C, cytomegalovirus (CMV) promoter; H, histidine tag; G, green fluorescent protein; HBc, core protein of hepatitis B virus) was transfected into HeLa cells. The recombinant core particles were isolated by sucrose gradient centrifugation. Epitopes in the fusion protein were detected by immunoelectron microscopy after adsorption of the particles onto carbon-coated grids. All incubation and washing steps were carried out on the grids. The antibodies used were (1) rabbit anti-HBc/goat antirabbit −5 nm gold; (2) mouse anti-histidine/goat antimouse −5 nm gold; and (3) rabbit anti-GFP/goat antirabbit −5 nm gold. After immunoreactions, the grids were negatively stained with uranyl acetate. Bar represents 100 nm.

of radioisotopes emitting soft β-particles, usually ^{3}H. Molecules that have been radiolabeled either *in vivo* or *in vitro* are prepared for EM by the method most appropriate for that type of sample (Table 8). After staining, the preparation is covered, first with a thin protective coat of carbon and then with a layer of photographic emulsion. Ideally, the emulsion should consist of a homogeneous monolayer of

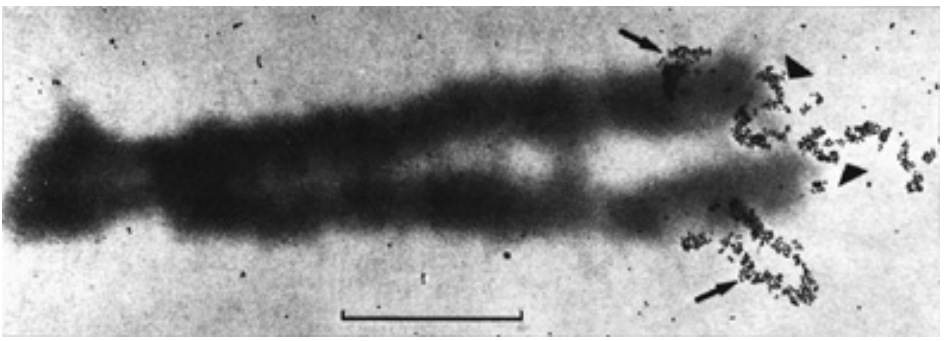

Fig. 19 Whole-chromosome mount from *Xenopus* culture cells after *in situ* hybridization with a mixture of biotinylated DNA encoding tRNA and oocyte-specific 5S RNA. The chromosomes are partially unfolded, and the sites of hybridization are detected using anti-biotin and secondary antibodies tagged with colloidal gold. The gold particles are seen to decorate DNA loops containing tRNA genes (arrows) and 5S RNA genes (arrowheads). The bar represents 2 µm. (Courtesy of S. Narayanswami and B. A. Hamkalo, with permission from Oxford University Press.)

the silver halide crystals. After being kept in the dark, to allow a sufficient amount of radioactive decay, the emulsion is developed, leaving silver grains above the sites of radiolabeling in the molecules. EM ARG is subject to two severe limitations:

1. The accuracy with which the developed silver grain can be located to the source of radiation in the specimen molecule. This depends on the thickness of the preparation and the diameter of the silver halide crystal in the emulsion and results in scattering of grains with a half distance of about 100 nm.
2. The efficiency in detecting radioactive disintegrations. Even molecules labeled to high specific activities require exposure times of generally more than a few weeks.

In spite of these limitations, EM ARG has been used successfully in studying sites of replication in isolated DNA molecules and the location of active transcription complexes in spread chromatin (Fig. 11).

3.3 Electron Spectroscopic Imaging (ESI)

By adapting the electron microscope to include an imaging electron energy filter, positional information can be recovered from electrons that have lost a specific and characteristic amount of energy in colliding with atoms in the specimen molecules. This approach has been used successfully in the location of phosphorus atoms, particularly those in the phosphodiester bonds of nucleic acids. The resulting electron spectroscopic imaging (ESI) has been used to delineate the path of the sugar–phosphate backbone in nucleosomes and the configuration of SRP–RNA (7SL RNA) within the signal recognition (ribonucleoprotein) particle. Along with related techniques in element mapping within biomolecules and their complexes, ESI is becoming increasingly sophisticated and promises many diverse applications.

3.4 Image Reconstruction

Individual molecules, under optimum conditions of spreading and staining, give weak and ill-defined images. To improve on the amount of structural detail, information from many molecules can be combined to smooth out random variations between images. To do this, it is necessary to use molecules (e.g. protein filaments), which contain regular, repeating arrays of subunits. Alternatively, the biomolecules can be induced to form crystalline arrays of regular, tightly packed, oriented units. Electron micrographs of arrays of these types can then be used for image processing to produce enhanced structural detail.

See also Electron Microscopy in Cell Biology.

Bibliography

Books and Reviews

Bazett-Jones, D.P. (1992) Electron spectroscopic imaging of chromatin and other nucleoprotein complexes, *Electron Microsc. Rev.* **5**, 37–58.

Dubochet, J., Adrian, M., Chang, J., Homo, J.-C., Lepault, J., McDowall, A.W., Schultz, P. (1988) Cryoelectron microscopy of vitrified specimens, *Q. Rev. Biophys.* **21**, 129–228.

Harris, J.R. (Ed.) (1991) *Electron Microscopy in Biology: A Practical Approach*, Oxford University Press, London, New York.

McLean, D.M., Wong, K.K. (1984) *Same-Day Diagnosis of Human Virus Infections*, CRC Press, Boca Raton.

Ottensmeyer, F.P. (1986) Elemental mapping by energy filtration: advantages, limitations and compromises, *Ann. N. Y. Acad. Sci.* **483**, 339–353.

Skepper, J.N. (2000) Immunocytochemical strategies for electron microscopy: choice or compromise, *J. Microsc.* **199**, 1–36.

Sommerville, J. (1981) Immunolocalization and Structural Organization of Nascent RNP, in: Busch, H. (Ed.) *The Cell Nucleus*, Vol. 8, Academic Press, New York, pp. 1–57.

Sommerville, J., Scheer, U. (Eds.) (1987) *Electron Microscopy in Molecular Biology: A Practical Approach*, Oxford University Press, London, New York.

van Tuinen, E.J. (1996) Immunoelectron microscopy, *Methods Mol. Biol.* **53**, 407–422.

Primary Literature

DNA Spreading

Bock, C.-T., Zentgraf, H. (1993) Detection of minimal amounts of DNA by electron microscopy using simplified spreading procedures, *Chromosoma* **102**, 249–252.

Egel-Mitani, M., Egel, R. (1972) A rapid visualization of Kleinschmidt-type DNA preparations by phosphotungstic acid, *Z. Naturforsch.* **27**, 480–481.

Kleinschmidt, A.K. (1968) Monolayer techniques in electron microscopy of nucleic acid molecules, *Methods Enzymol.* **12**, 361–377.

Lang, D., Mitani, M. (1970) Simplified quantitative electron microscopy of biopolymers, *Biopolymers* **9**, 373–379.

Spiess, E., Lurz, R. (1988) Electron microscopic analysis of nucleic acids and nucleic acid-protein complexes, *Methods Microbiol.* **20**, 293–321.

Vollenweider, H.H., Sogo, J.M., Koller, T. (1975) A routine method for protein-free spreading of double- and single-stranded nuclei acid molecules, *Proc. Natl. Acad. Sci. USA* **72**, 83–87.

Long DNA Molecules

Montoliu, L., Bock, C.-T., Schütz, G., Zentgraf, H. (1995) Visualization of large DNA molecules by electron microscopy with polyamines: application to the analysis of yeast endogenous and artificial chromosomes, *J. Mol. Biol.* **246**, 486–492.

DNA–DNA Heteroduplex

Davis, R.W., Davidson, N. (1968) Electron microscopic visualization of deletion mutations, *Proc. Natl. Acad. Sci. USA* **60**, 243–250.

Davis, R.W., Simon, M., Davidson, N. (1971) Electron microscope heteroduplex methods for mapping regions of base sequence

homology in nucleic acids, *Methods Enzymol.* **21**, 413–428.

Scherer, G., Bock, C.T., Zentgraf, H. (1990) Heteroduplex between an EMBL3 genomic clone and a gt11 cDNA clone provides internal size standards, *Nucleic Acids Res.* **18**, 4944.

Walfield, A.M., Storb, U., Selsing, E., Zentgraf, H. (1980) Comparison of different rearranged immunoglobulin kappa genes of a myeloma by electron microscopy and restriction mapping of cloned DNA's: implications for allelic exclusion, *Nucleic Acids Res.* **8**, 4689–4707.

Westmoreland, B.C., Szybalski, W., Ris, H. (1969) Mapping of deletions and substitutions in heteroduplex DNA molecules of bacteriophage lambda by electron microscopy, *Science* **163**, 1343–1351.

DNA–RNA Heteroduplex

Chow, L.T., Roberts, J.M., Lewis, J.B., Broker, T.R. (1977) A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA: DNA hybrids, *Cell* **11**, 819–825.

Hsu, M.-T., Ford, J. (1977) A novel sequence arrangement of SV40 late RNA, *Cold Spring Harbor Symp. Quant. Biol.* **42**, 571–576.

May, E., Maizel, J.V., Salzman, N. (1977) Mapping of transcription sites of simian virus 40-specific late 16S and 19S mRNA by electron microscopy, *Proc. Natl. Acad. Sci. USA* **71**, 496–500.

Müller, G., Scherer, G., Zentgraf, H., Ruppert, S., Herrmann, B., Lehrach, H., Schütz, G. (1985) Isolation, characterization and chromosomal mapping of the mouse tyrosin aminotransferase gene, *J. Mol. Biol.* **184**, 367–373.

Schmid, W., Scherer, G., Danesch, U., Zentgraf, H., Matthias, P., Strange, C.M., Röwekamp, W., Schütz, G. (1982) Isolation and characterization of the rat tryptophan oxygenase gene, *EMBO J.* **10**, 1287–1293.

Thomas, M., White, R.L., Davies, R.W. (1976) Hybridization of RNA to double-stranded DNA: formation of R-loops, *Proc. Natl. Acad. Sci. USA* **73**, 2294–2301.

White, R.L., Hogness, D.S. (1977) R-loop mapping of the 18S and 28S sequences in the long and short repeating units of *Drosophila melanogaster* DNA, *Cell* **10**, 177–185.

Small Chromatin Units

Bellard, M., Oudet, P., Germond, J.E., Chambon, P. (1976) Subunit structure of simian virus 40 minichromosomes, *Eur. J. Biochem.* **70**, 543–549.

Bock, C.T., Schwinn, S., Locarnini, S., Fyfe, J., Manns, M.P., Trautwein., C., Zentgraf, H. (2001) Structural organization of the hepatitis B virus minichromosome, *J. Mol. Biol.* **307**, 183–196.

Cremisi, C., Pignatti, P.F., Croissant, O., Yaniv, M. (1976) Chromatin-like structures in polyoma virus and simian virus 40 lytic cycle, *J. Virol.* **17**, 204–211.

Griffith, J.D. (1975) Chromatin structure: deduced from a minichromosome, *Science* **187**, 1202–1203.

Keller, W., Müller, U., Eicken, I., Wendel, I., Zentgraf, H. (1977) Biochemical and ultrastructural analysis of SV40 chromatin, *Cold Spring Harbor Symp. Quant. Biol.* **42**, 227–244.

Müller, U., Zentgraf, H., Eicken, I., Keller, W. (1978) Higher-order structure of simian virus 40 chromatin, *Science* **201**, 406–415.

Scheer, U., Zentgraf, H. (1978) Nucleosomal and supranucleosomal organization of transcriptional inactive rDNA circles in *Dytiscus* oocytes, *Chromosoma* **69**, 243–254.

Strätling, W.H., Müller, U., Zentgraf, H. (1978). The higher-order repeat-structure of chromatin is built up of globular particles containing eight nucleosomes. *Exp. Cell Res.* **117**, 301–311.

Zentgraf, H., Trendelenburg, M.F., Spring, H., Scheer, U., Franke, W.W., Müller, U., Drury, K.C., Runger, D. (1979) Mitochondrial DNA arranged into chromatin-like structures after injection into amphibian oocyte nuclei, *Exp. Cell Res.* **122**, 363–375.

Large Chromatin Units (transcriptionally inactive chromatin)

Finch, J.T., Klug, A. (1976) Solenoidal model for superstructure in chromatin, *Proc. Natl. Acad. Sci. USA* **73**, 1897–1900.

Müller, U., Schröder, C.H., Zentgraf, H., Franke, W.W. (1980) Coexistence of nucleosomal and various non-nucleosomal chromatin configurations in cells infected with herpes simplex virus, *Eur. J. Cell Biol.* **23**, 197–203.

Olins, A.L., Olins, D.E. (1974) Spheroid chromatin units (ν bodies), *Science* **183**, 330–332.

Oudet, P., Gross-Bellard, M., Chambon, P. (1975) Electron microscopic and biochemical evidence that chromatin structure is a repeating unit, *Cell* **4**, 281–300.

Woodcock, C.L.F. (1973) Ultrastructure of inactive chromatin, *J. Cell Biol.* **59**, 368a.

Zentgraf, H., Müller, U., Franke, W.W. (1980) Supranucleosomal organization of sea urchin sperm chromatin in regularly arranged 40–50 mm large granular subunits, *Eur. J. Cell Biol.* **20**, 254–264.

Zentgraf, H., Müller, U., Franke, W.W. (1980) Reversible in vitro packing of nucleosomal filaments into globular supranucleosomal units in chromatin of whole chick erythrocyte nuclei, *Eur. J. Cell Biol.* **23**, 171–188.

Large Chromatin Units (transcriptionally active chromatin)

Foe, V.E., Wilkinson, L.E., Laird, C.D. (1976) Comparative organization of active transcription units in *Oncopeltus fasciatus*, *Cell* **9**, 131–146.

Franke, W.W., Scheer, U., Trendelenburg, M., Spring, H., Zentgraf, H. (1976) Absence of nucleosomes in transcriptionally active chromatin, *Cytobiologie* **13**, 401–434.

Lamb, M.M., Daneholt, B. (1979) Characterization of active transcription units in Balbiani rings of *Chironomus tentans*, *Cell* **17**, 835–848.

McKnight, S.L., Sullivan, N.L., Miller, O.L. (1976) Visualization of the silk fibroin transcription unit and nascent silk fibroin molecules on polyribosomes of *Bombyx mori*, *Prog. Nucleic Acid Res. Mol. Biol.* **19**, 313–318.

Miller, O.L., Bakken, A.H. (1972) Morphological studies of transcription, *Acta endocrinol. Suppl.* **168**, 155–177.

Miller, O.L., Beatty, B.R. (1969) Visualization of nucleolar genes, *Science* **164**, 955–957.

Puvion-Dutilleul, F., Bernadac, A., Puvion, E., Bernard, W. (1977) Visualization of nulear transcriptional complexes in rat liver cells, *J. Ultrastr. Res.* **58**, 107–117.

Samuel, C., Mackie, J., Sommerville, J. (1981) Macronuclear chromatin organization in *Paramecium primaurelia*, *Chromosoma* **83**, 481–492.

Scheer, U. (1978) Changes of nucleosome frequency in nucleolar and non-nucleolar chromatin as a function of transcription: an electron microscopic study, *Cell* **13**, 535–549.

Scheer, U., Franke, W.W., Trendelenburg, M.F., Spring, H. (1976) Classification of loops of lampbrush chromosomes according to the arrangement of transcriptional *complexes*, *J. Cell Sci.* **22**, 503–520.

Scheer, U., Trendelenburg, M.F., Franke, W.W. (1976) Regulation of transcription of genes of ribosomal RNA during amphibian oogenesis, *J. Cell Biol.* **69**, 465–489.

Scheer, U., Zentgraf, H., Sauer, H.W. (1981) Different chromatin structures in *Physarum polycephalum*: a special form of transcriptionally active chromatin devoid of nucleosomal particles, *Chromosoma* **84**, 279–290.

Trendelenburg, M.F., Spring, H., Scheer, U., Franke, W.W. (1974) Morphology of nucleolar cistrons in a plant cell, *Acetabularia mediterrane*, *Proc. Natl. Acad. Sci. USA* **71**, 3626–3630.

Negative Staining

Almeida, J.D., Howatson, A.F. (1963) A negative staining method for cell associated virus, *J. Cell Biol.* **16**, 616–623.

Baumeister, W. (1978) Biological horizons in molecular microscopy, *Cytobiologie* **17**, 246–297.

Brenner, S., Horne, R.W. (1959) A negative staining method for high resolution electron microscopy of viruses, *Biochim. Biophys. Acta* **34**, 103–110.

Harris, R., Horne, R.W. (1991) Negative Staining, in: Harris, R. (Ed.) *Electron Microscopy in Biology: A Practical Approach*, Oxford University Press, London, New York, pp. 203–228.

Horne, R.W., Wildy, P. (1979) An historical account of the development and applications of the negative staining technique to the electron microscopy of viruses, *J. Microsc.* **117**, 103–122.

Smith, K.O. (1967) Identification of Viruses by Electron Microscopy, in: Busch, H. (Ed.) *Methods in Cancer Research*, Academic Press, New York, pp. 545–568.

Spiess, E., Zimmermann, H.-P., Lünsdorf, H. (1987) Negative Staining of Protein Molecules and Filaments, in: Sommerville, J., Scheer, U. (Eds.) *Electron Microscopy in Molecular Biology: A Practical Approach*, Oxford University Press, London, New York, pp. 147–166.

Immunolabelling of Spread Preparations

Bock, C.-T., Schwinn, S., Zentgraf, H. (1995) Diheteroduplex formation using gold

labelled single-stranded PCR fragments and its application in electron microscopy, *Chromosoma* **103**, 653–657.

Broker, T.R., Angerer, L.M., Yen, P.H., Hershey, N.D., Davidson, N. (1978) Electron visualization of tRNA genes with ferritin-avidin: biotin labels, *Nucleic Acids Res.* **5**, 363–383.

Delius, H., van Heerikhuizen, H., Clarke, J., Koller, B. (1985) Separation of complementary strands of plasmid DNA using the biotin-avidin system and its application to heteroduplex formation and RNA/DNA hybridizations in electron microscopy, *Nucleic Acids Res.* **13**, 5457–5469.

Rösl, F., Waldeck, W., Zentgraf, H., Sauer, G. (1986) Properties of intracellular bovine papillomavirus chromatin, *J. Virol.* **58**, 500–507.

Stahl, H., Dröge, P., Zentgraf, H., Knippers, R. (1985) A large-tumor-antigen-specific monoclonal antibody inhibits DNA replication of simian virus 40 minichromosomes in an *in vitro* elongation system, *J. Virol.* **54**, 473–482.

Immunoelectron microscopy (cell biology and viruses)

An, K., Paulsen, A.Q., Tilley, M.B., Consigli, R.A. (2000) Use of electron microscopic and immunogold labeling techniques to determine polyomavirus recombinant VP1 capsid-like particles entry into mouse 3T6 cell nucleus, *J. Virol. Methods* **90**, 91–97.

Arlucea, J., Andrade, R., Alonso, R., Arechaga, J. (1998) The nuclear basket of the nuclear pore complex is part of a higher-order filamentous network that is related to chromatin, *J. Struct. Biol.* **124**, 51–58.

Chang, P., Giddings, T.H. Jr., Winey, M., Stearns, T. (2003) Epsilon-tubulin is required for centriole duplication and microtubule organization, *Nat. Cell Biol.* **5**, 71–76.

Kirkham, M., Muller-Reichert, T., Oegema, K., Grill, S., Hyman, A.A. (2003) SAS-4 is a C. elegans centriolar protein that controls centrosome size, *Cell* **112**, 575–587.

Nermut, M.V., Zhang, W.H., Francis, G., Ciampor, F., Morikawa, Y., Jones, I.M. (2003) Time course of Gag protein assembly in HIV-1-infected cells: a study by immunoelectron microscopy, *Virology* **305**, 219–227.

Portoles, M., Faura, M., Renau-Piqueras, J., Iborra, F.J., Saez, R., Guerri, C., Serratosa, J., Rius, E., Bachs, O. (1994) Nuclear calmodulin/62 kDa calmodulin-binding protein complexes in interphasic and mitotic cells, *J. Cell Sci.* **107**, 3601–3614.

Ueki, S., Citovsky, V. (2002) The systemic movement of a tobamovirus is inhibited by a cadmium-ion-induced glycine-rich protein, *Nat. Cell Biol.* **4**, 478–486.

Urban, S., Schwarz, C., Marx, U.C., Zentgraf, H., Schaller, H., Multhaup, G. (2000) Receptor recognition by a hepatitis B virus reveals a novel mode of high affinity virus-receptor interaction, *EMBO J.* **19**, 1217–1227.

Weiland, E., Bolz, S., Weiland, F., Herbst, W., Raamsman, M.J., Rottier, P.J., De Vries, A.A. (2000) Monoclonal antibodies directed against conserved epitopes on the nucleocapsid protein and the major envelope glycoprotein of equine arteritis virus, *J. Clin. Microbiol.* **38**, 2065–2075.

Image Reconstruction

Andel, F. III, Ladurner, A.G., Inouye, C., Tjian, R., Nogales, E. (1999) Three-dimensional structure of the human TFIID-IIA-IIB complex, *Science* **286**, 2153–2156.

Bottcher, B., Wynne, S.A., Crowther, R.A. (1997) Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy, *Nature* **386**, 88–91.

Mayo, K., Huseby, D., McDermott, J., Arvidson, B., Finlay, L., Barklis, E. (2003) Retrovirus capsid protein assembly arrangements, *J. Mol. Biol.* **325**, 225–237.

Nandhagopal, N., Simpson, A.A., Gurnon, J.R., Yan, X., Baker, T.S., Graves, M.V., Van Etten, J.L., Rossmann, M.G. (2002) The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus, *Proc. Natl. Acad. Sci. USA* **99**, 14758–14763.

Smith, C.L., Horowitz-Scherer, R., Flanagan, J.F., Woodcock, C.L., Peterson, C.L. (2003) Structural analysis of the yeast SWI/SNF chromatin remodeling complex, *Nat. Struct. Biol.* **10**, 141–145.

Sperling, R., Koster, A.J., Melamed-Bessudo, C., Rubinstein, A., Angenitzki, M., Berkovitch-Yellin, Z., Sperling, J. (1997) Three-dimensional image reconstruction of large nuclear RNP (lnRNP) particles by automated electron tomography, *J. Mol. Biol.* **267**, 570–583.

Trus, B.L., Roden, R.B., Greenstone, H.L., Vrhel, M., Schiller, J.T., Booy, F.P. (1997) Novel structural features of bovine

papillomavirus capsid revealed by a three-dimensional reconstruction to 9 A resolution, *Nat. Struct. Biol.* **4**, 413–420.

Zhou, Z.H., He, J., Jakana, J., Tatman, J.D., Rixon, F.J., Chiu, W. (1995) Assembly of VP26 in herpes simplex virus-1 inferred from structures of wild-type and recombinant capsids, *Nat. Struct. Biol.* **2**, 1026–1030.

Electron Spectroscopic Imaging (ESI)

Bazett-Jones, D.P., Brown, M.L. (1989) Electron microscopy reveals that transcription factor TFIIIA bends 5S DNA, *Mol. Cell Biol.* **9**, 336–341.

Bazett-Jones, D.P., Kimura, K., Hirano, T. (2002) Efficient supercoiling of DNA by a single condensin complex as revealed by electron spectroscopic imaging, *Mol. Cell* **9**, 1183–1190.

Boisvert, F.M., Hendzel, M.J., Bazett-Jones, D.P. (2000) Promyelocytic leukemia (PML) nuclear bodies are protein structures that do not accumulate RNA, *J. Cell Biol.* **148**, 283–292.

Czarnota, G.J., Ottensmeyer, F.P. (1996) Structural states of the nucleosome, *J. Biol. Chem.* **271**, 3677–3683.

Grohovaz, F., Bossi, M., Pezzati, R., Meldolesi, J., Tarelli, F.T. (1996) High resolution ultrastructural mapping of total calcium: electron spectroscopic imaging/electron energy loss spectroscopy analysis of a physically/chemically processed nerve-muscle preparation, *Proc. Natl. Acad. Sci. USA* **93**, 4799–4803.

Hendzel, M.J., Boisvert, F., Bazett-Jones, D.P. (1999) Direct visualization of a protein nuclear architecture, *Mol. Biol. Cell* **10**, 2051–2062.

Locklear, L. Jr., Ridsdale, J.A., Bazett-Jones, D.P., Davie, J.R. (1990) Ultrastructure of transcriptionally competent chromatin, *Nucleic Acids Res.* **18**, 7015–7024.

Malecki, M., Hsu, A., Truong, L., Sanchez, S. (2002) Molecular immunolabeling with recombinant single-chain variable fragment (scFv) antibodies designed with metal-binding domains, *Proc. Natl. Acad. Sci. USA* **99**, 213–218.

Electron Spin Resonance of Biomolecules

Betty J. Gaffney
Department of Biological Sciences, BIO Unit 1, Tallahassee, FL, USA

1 Basic ESR Concepts and Spectra 117

2 ESR and the Roles of Naturally Occurring Paramagnetic Biomolecules 119
2.1 Free-radical Enzyme Intermediates 119
2.2 Paramagnetic Metal Ions in Biology 120
2.3 Protein Complexes with Nitric Oxide (NO) 122
2.4 Factors Directing Electron Flow in Energetic Membranes 123
2.4.1 ESR Studies of Photosynthesis 123
2.4.2 ESR Studies of Respiratory Enzymes 124

3 ESR Probes of Biological Structure and Function 124
3.1 Spin Labeling 124
3.1.1 Dynamics of Proteins, DNA, and Membranes 124
3.1.2 Site-directed Spin Labeling (SDSL) 125
3.2 Spin Trapping 127
3.3 ESR Imaging 128
3.3.1 ESR Imaging of Oxygen in Tissues 128
3.3.2 ESR Imaging of NO 129
3.4 Long-range Distance (6–60 Å) Measurements by ESR 129

Bibliography 129
Books and Reviews 129
Primary Literature 130

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Keywords

cwESR
ESR signal detected with radiation turned on continually during recording.

EPR, PMR, EMR
Synonyms, in practice, for ESR: electron paramagnetic resonance, paramagnetic resonance, and electron magnetic resonance.

EPRI
In vivo imaging based on the ESR signal of a paramagnetic molecule.

ESR
Electron Spin Resonance. Spectroscopy involving unpaired electrons in a magnetic field.

ENDOR
Electron-Nuclear Double Resonance spectroscopy. Spectroscopy in which NMR signals are observed as changes in intensity of ESR signals.

ESEEM
Electron Spin Echo Envelope Modulation spectroscopy. A pulsed ESR technique in which nuclear moments modulate time-dependent electron relaxation.

g-factor
Proportionality constant between ESR frequency applied and the magnetic field where ESR absorption occurs.

Hyperfine Splitting
An ESR signal of an unpaired electron is split into multiple signals, one for each dipolar interaction with a unique nuclear spin.

Isotropic or Anisotropic
Uniform or asymmetric, respectively, with regard to coordinates in space.

Pulsed ESR
ESR signal detected as a function of time after a pulse of energy.

Spin Label
Paramagnetic molecule used to provide information about a nonparamagnetic biomolecule with which it interacts; the same term is used in nuclear magnetic resonance imaging for manipulations of nuclei in a portion of the sample to amplify the NMR signal from localized regions.

Spin Trap
A means of capturing information from a radical that has a short lifetime or other properties rendering it unsuitable for ESR so that, when trapped, useful ESR signals result.

Electron spin resonance (ESR) refers to spectroscopy of unpaired electrons, and is sometimes called electron paramagnetic resonance (EPR) or electron magnetic resonance (EMR). The theoretical bases of ESR spectroscopy are similar to those of nuclear magnetic resonance (NMR), except that an electron spin, rather than a nuclear spin, is the focus. Unpaired electrons in biological systems are in much lower abundance than nuclei, so ESR is a technique that focuses on local sites while NMR is more global.

Two electrons are paired, with antiparallel spins, in a single chemical bond. The ESR requirement of an unpaired electron spin is met in biology when (1) a chemical bond is broken homolytically, as in formation of a free-radical enzyme intermediate; (2) there are unfilled valence orbitals, as there are in oxygen, nitric oxide, or many metal ions; and (3) one-electron oxidation or reduction of a nonparamagnetic biomolecule has occurred. Biological subjects for ESR include free-radical enzyme intermediates, metal ions, nitric oxide and some of its complexes, and redox-active cofactors such as quinones and flavins. The range of applications of ESR spectroscopy is not limited to natural sources of unpaired electrons. An ESR probe technique, site-directed spin labeling (SDSL), is widely applied to examine dynamics and folding of biomolecules. Other paramagnetic probes can be used to image oxygen or nitric oxide in biological tissues *in vitro* and *in vivo*.

1 Basic ESR Concepts and Spectra

The term "spin" in the designations of various forms of magnetic resonance spectroscopy refers to a property of electrons (or nuclei). It is the interaction of the spin with magnetic field that leads to separation of energy levels between which spectroscopic transitions occur. Usually, a magnetic field is required for ESR, as well as a source of energy to effect transitions, although a few spin transitions can be detected in the absence of an applied magnetic field.

ESR signals can be detected during continuous application of microwave energy of narrow bandwidth (continuous wave ESR, cwESR) or after a short, broader-band microwave pulse (pulsed ESR). In the former case, the signal is detected by scanning the external magnetic field through the resonance condition; in the latter, the field is fixed for each pulse, but, because of the breadth of many ESR signals, data may be collected at multiple steps in magnetic field. To filter noise, a cwESR signal is recorded with an additional magnetic field modulation and phase-sensitive detection, resulting in an output that is the derivative of the absorption. Absorption and derivative ESR signals are illustrated in Figs. 1(a, b).

Equation (1) gives the resonance condition for a free electron, where ν is the microwave frequency; g_e is Zeeman splitting constant for a free electron; β_e, the Bohr magneton, is proportional to the charge to mass ratio of the electron; and B_0 is the external magnetic field in which the sample resides.

$$\text{Energy} = h\nu = g_e \beta_e B_0 \qquad (1)$$

Internal magnetic fields in molecules alter the simple spectrum shown in Fig. 1(b)

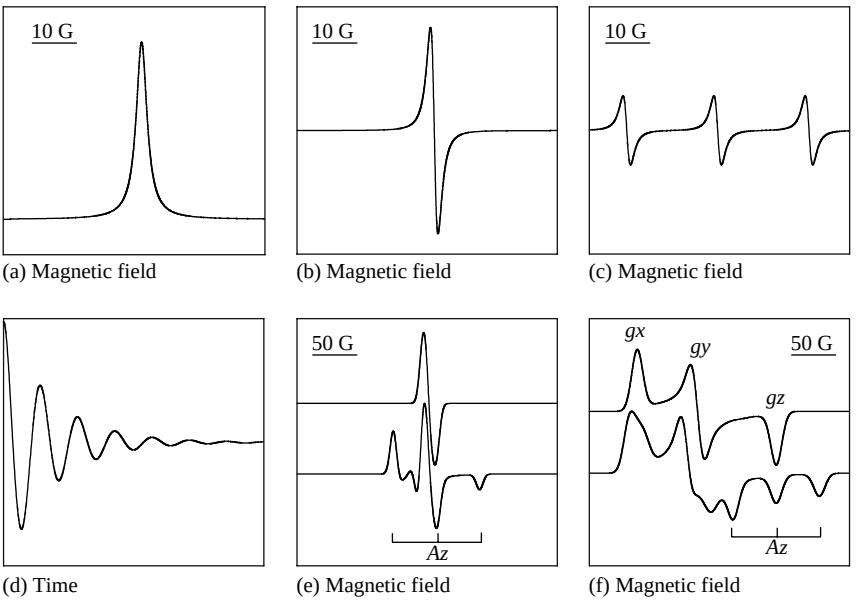

Fig. 1 Basic ESR spectra concepts are illustrated. (a) An isolated, unpaired electron has an absorption ESR signal of Lorentzian shape; (b) in cwESR, the signal shown in (a) is recorded as the derivative of the absorption; (c) when a nuclear spin interacts with the unpaired electron, the signal is split into multiple lines, one for each nuclear spin state (^{14}N spin states are -1, 0, and $+1$); (d) the time-dependent response of an unpaired electron to a pulse is an oscillating decay; (e) and (f) are spectra with parameters typical of spin labels recorded at 9.4 GHz (e) and 94 GHz (f). In (e) and (f), the upper spectra are calculated with only *g*-factor anisotropy, while the calculations of the lower spectra include anisotropic terms for both *g*-factor and hyperfine splitting. The largest hyperfine splitting, *Az*, is shown.

and thus provide the features that allow information about structure and dynamics to be deduced. The local magnetic fields add terms to Eq. (1), and it is convenient to refer to the proportionality between the frequency applied and the magnetic field at which a characteristic feature is seen in the ESR spectrum in terms of an "effective *g*-factor," g', as given in Eq. (2).

$$h\nu = (g_e \beta_e B_0 + \text{other terms}) = g' \beta_e B_0 \tag{2}$$

The "other terms" include, but are not limited to, those of electron spin dipolar interactions with spins of nuclei or other unpaired electrons, spin exchange due to overlap of the orbitals of two electron spins, and terms arising from strong interactions of multiple electrons residing on the same atom. Different terms dominate the ESR spectrum in varied applications. Some terms are described by expressions including magnetic field and other terms are independent of magnetic field. ESR spectra obtained at several different frequencies, ν, help resolve contributions of the different terms.

Dipolar interactions of magnetic nuclei (for instance, hydrogen or nitrogen) with an unpaired electron render ESR spectra sensitive to molecular structure. An unpaired electron residing in a p- or π-orbital on ^{14}N, such as occurs in a spin label or in nitric oxide (NO), has dipolar interactions

of slightly different energies with each of the nitrogen nuclear spin states (^{14}N nuclear spins −1, 0, and +1). As a result, the simple unpaired electron ESR signal becomes a spectrum with three lines, each of one-third the intensity (Fig. 1(c)) of the single line (Fig. 1(b)). The separation of the lines is termed the *hyperfine splitting* of the electron spin by the nuclear spins. Hyperfine splitting by ^{1}H or ^{15}N yields ESR spectra divided into two lines.

The response of spins to a short pulse of energy is recorded over time in pulsed ESR. The time-dependent response may be converted to the cwESR spectrum by Fourier transformation. A Fourier transform of the absorption signal (Fig. 1(a)), which would be detected in the pulsed experiment, is shown in Fig. 1(d). The exponential decay is related to the line width and the oscillations to the offset of the absorption peak from zero magnetic field. ESR and NMR are connected in the variations of pulsed ESR that are termed electron spin echo envelope modulation (ESEEM) and electron-nuclear double resonance (ENDOR), in which electron transitions are modulated by nearby nuclei. After Fourier transformation of the time response, the characteristic frequencies of nuclear spins near the electron spin are displayed in these experiments. Determining the nature, geometric arrangement, and distance of nuclei from the unpaired electron are applications of these pulsed ESR techniques.

Figures 1(a, b) are based on a fictitious "free electron." In a molecule, an unpaired electron responds to an anisotropic (asymmetric) environment, resulting in energies for the magnetic interactions that depend on the orientation of each molecule in the magnetic field. The *g*-factor becomes a matrix (**g**) relating the vectors of magnetic field, B_0, and electron spin, S. The expression of the energy for the hyperfine interaction is $S\,\mathbf{A}\,I$, where electron and nuclear spin vectors are S and I, respectively, and **A** is the interaction matrix. The diagonal elements of the **g**- and **A**-matrices are gx, gy, and gz and Ax, Ay, and Az, respectively. As a result of these interactions, if a frozen solution is studied, the spectrum will be the sum of spectra from molecules at all orientations, and will be broader than the free electron signal. Figures 1(e, f) illustrate the type of spectra that result from a frozen solution of nitroxide molecules. Figure 1(e) is calculated for a frequency of applied radiation (9.41 GHz) 10 times lower than that of Fig. 1(f) (94.1 GHz). The upper spectra show only the contribution from the anisotropic *g*-factor. The features corresponding to gx, gy, and gz can hardly be seen in Fig. 1(e) (upper curve), but they are separated as indicated in Fig. 1(f) (upper curve). This is because the *g*-factor energy term depends on magnetic field. In contrast, the hyperfine term is constant. The lower curves indicate the Az component of the hyperfine splitting; it is centered on gz. Ax and Ay are smaller and are not indicated on the figures. However, it can be seen in Fig. 1(f) (lower curve) that the gx and gy regions are broadened by the smaller Ax and Ay hyperfine splittings.

2 ESR and the Roles of Naturally Occurring Paramagnetic Biomolecules

2.1 Free-radical Enzyme Intermediates

Cofactors and amino acid side chains are known to form radical, or paramagnetic intermediates in enzymes. Ribonucleotide reductases use a pathway of electron transfers to carry out reduction of

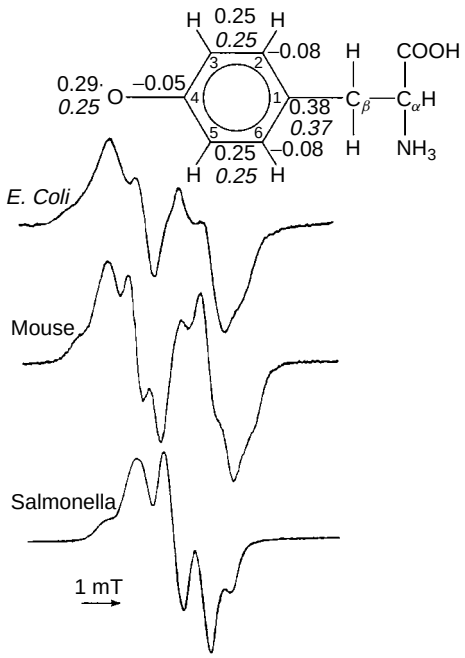

Fig. 2 ESR spectra of tyrosine free radicals of ribonucleotide reductase R2 subunits from different organisms. The ESR spectra at ~9 GHz were recorded at temperatures from 20 to 30 K. The numbers given on the structure are the spin density distributions for the *Escherichia coli* (Roman letters) and *Salmonella typhimurium* (italics) enzymes. (Reprinted with permission from Gräslund, A. and Sahlin, M. (1996) *Ann. Rev. Biophys. Biomol. Struct.* **27**, 259–286.)

ribonucleotides to deoxyribonucleotides. In the enzymes from different sources, the presence of free-radical intermediates on side chains of the amino acids tyrosine, tryptophan, cysteine, and even glycine has been demonstrated by ESR. ESR studies of tyrosine radicals, in this system and also in electron-transfer membranes, show that the unpaired electron density at the various carbon atoms of the tyrosine aromatic ring is fine-tuned by hydrogen bonds to the phenolic oxygen and by the protein environment. ESR spectra also provide information about side-chain torsion angles of the tyrosines giving rise to radicals. Figure 2 gives examples of EMR spectra of tyrosine radicals of ribonucleotide reductases from several sources.

In other proteins, tyrosine radicals participate in the mechanisms of cytochrome oxidase (mitochondrial respiratory complex IV), cyclooxygenase and other heme enzymes. In other proteins, glycine radicals have functional roles in lysine 2,3-aminomutase, pyruvate formate-lyase, and benzylsuccinate synthase. Enzyme-bound substrate and cofactor radical intermediates have also been characterized in enzymes dependent on S-adenosyl methionine, adenosylcobalamin, and arachidonic acid. It is often useful to use isotopic labeling with nuclei having magnetic moments (^{2}H, ^{13}C, ^{15}N, ^{17}O) in the effort to identify the radical site in an enzyme intermediate.

2.2 Paramagnetic Metal Ions in Biology

All metal ions containing unpaired electrons (paramagnetic ions) are, in principle, subjects for study by ESR, but the experimental conditions are quite varied. The paramagnetic ions in proteins, commonly studied by ESR, include manganese^{2+}, copper^{2+}, iron^{3+}, occasionally iron^{2+},

nickel^{3+}, vanadyl ($V = O^{2+}$), and cobalt^{2+}. The conditions for the spectroscopy differ drastically depending on which metal is the subject of study. Manganese ions are sometimes detected with high sensitivity in solutions at room temperature. However, detection of ^{17}O-threonine side-chain ligands to manganese in p21 *ras* required both low temperature and high-frequency ESR to give optimum results. Copper sites also can be detected at room temperature, but sensitivity is improved by conducting the ESR experiments at liquid nitrogen temperature or lower. Studies of iron are almost always done at temperatures near that of liquid helium. The two primary reasons that low temperature is used in metal ion ESR are that relaxation times are too fast for room temperature studies and the signal intensity increases inversely with temperature. The pulsed ESR methods ENDOR and ESEEM are chosen to detect nuclei such as nitrogen or hydrogen bound to, or near, metal ion sites.

Quantitative evaluation of the number of unpaired spins in different sites within a complex biochemical electron-transfer system can be made by ESR and associated studies. For experiments of this type, both calculations of the theoretical spectra and a multifrequency approach in the experiments are needed for interpretation. Recent advances in high-frequency ESR provide improved resolution in samples with multiple ESR-detectable metal sites. For example, Fig. 3 shows a high-frequency ESR spectrum of a frozen solution of copper ion in the site normally occupied by iron in lactoferrin.

The copper spectrum, spread over the field region from ~2.9 to 3.3 T, is well separated from the six-line signal of a manganese impurity near 3.4 T. At lower ESR frequencies, the manganese signal is superimposed on the central portion of the copper signal. Also notable in the high-frequency ESR spectrum shown (Fig. 3) is that the portion (at 2.9–2.95 T) of the spectrum arising from molecules with a unique axis aligned close to the magnetic field direction is well separated from the portion (at 3.2–3.3 T) arising from molecules in which this axis is perpendicular to the field. Additionally, the

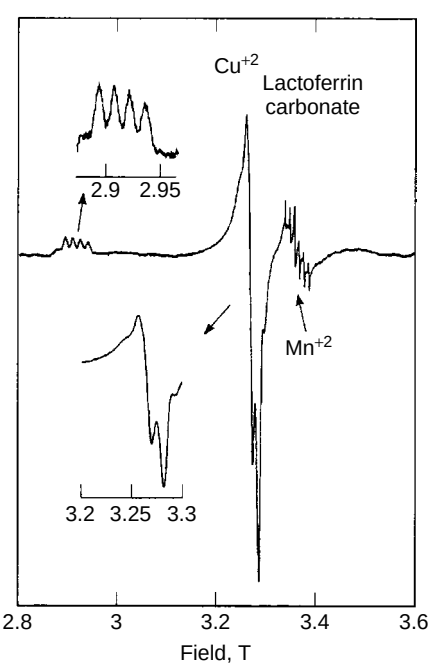

Fig. 3 High-frequency ESR spectrum of dicupric lactoferrin. The spectrum was recorded at 40 K and 94.1 GHz. The two regions from which significant information can be obtained are amplified in insets on the left. The feature on the right is the ESR signal from an impurity of manganese ion in the sample. (Reprinted with permission from Gaffney, B.J., Maguire, B.C., Weber, R.T., Maresch, G.G. (1999) Disorder at metal sites in proteins: a high frequency EMR study, *Appl. Magn. Res.* **16**, 207–222.)

splitting of these regions into four features arises from interaction of the unpaired electron with the four nuclear spin states of the copper nucleus. For copper proteins, very low-frequency EPR is also useful in revealing how many nitrogen ligands to the metal there are. This approach, as well as pulsed ESR, helped elucidate the nature of multiple copper binding sites in the prion protein (PrP), for example.

2.3 Protein Complexes with Nitric Oxide (NO)

Nitric oxide (NO) is a paramagnetic gas and molecular adducts of NO in solution give ESR signals that are employed to determine the pathway of nitric oxide transfers in proteins or the location of NO in tissues by *in vivo* ESR imaging. The characteristic ESR signal of some NO complexes is three lines, broader but otherwise similar to the spectrum shown in Fig. 1(c). When the ^{15}N isotope (nuclear spin states $\pm 1/2$) is incorporated in NO, the ESR spectrum simplifies to two lines. In an NO complex with iron, copper or other metals in proteins, broader signals with no resolved hyperfine splitting may or may not be observed. Three-line hyperfine splitting is seen in NO complexes with 5-coordinate ferrous heme iron, with the NO spin traps: ferrous ion chelated with dithiocarbamate derivatives, and with diverse iron-sulfur-dinitrosyl species *in vivo*.

NO is often used biochemically as an oxygen mimic, but it binds more tightly to metal centers than O_2 does and may alter other ligand geometries. When a single NO combines with ferrous iron, an ESR-detectable complex is formed. The spin 1/2 of NO combines with the even spin (spin = 0 or 2) of ferrous iron to give half-integer spin for the complex (spin = 1/2 or 3/2). Generally, half-integer spins yield well-resolved ESR spectra, whereas spin = 0 is not paramagnetic, and spin = 2 usually does not give resolved signals.

The strength of the Fe–NO interaction with iron in hemoglobin is demonstrated under conditions in which hemoglobin binds NO preferentially at the α-subunits; here, NO acts transaxially to cleave the proximal iron-His bond. ESR spectra of the resulting five-coordinate α-subunit heme have three resolved peaks from ^{14}N-hyperfine interactions. A similar spectrum of the five-coordinate NO complex with heme in myoglobin is shown in Fig. 4.

In contrast, when NO is in a six-coordinate complex with ferrous heme, a broad ESR signal, without hyperfine, is observed. NO activation at heme in guanyl cyclases also involves breaking an Fe-His bond, and the NO-Fe complex is ESR detectable. Interaction of NO or nitrite with oxyhemoglobin, or with methemoglobin, leads to a third species, S-nitrosyl hemoglobin. NO also reacts with

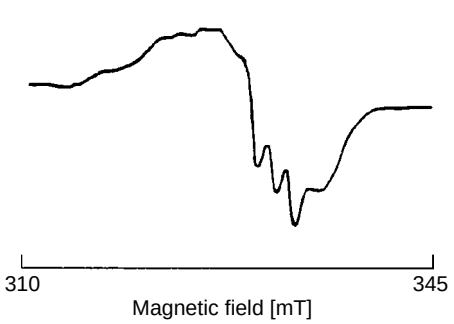

Fig. 4 The ESR spectrum of deoxymyoglobin to which nitric oxide (NO) gas was added. The spectrum was recorded at 77 K and a frequency of 9.1 GHz. (Reprinted with permission from Singel, D. J., Lancaster, J. R. Jr., (1996) *Methods in Nitric Oxide Research*, Feelisch, M., Stamler, J. S. (Eds.) John Wiley & Sons Ltd Ch. 23, pp. 341–356.)

the iron in iron-sulfur proteins. In the case of aconitase, one iron of the 4Fe-4S is lost upon reaction with NO and ESR signals of a protein-bound iron-$(NO)_2$ complex can be seen. This complex is likely an intermediate in the conversion to 3Fe-4S apoaconitase.

2.4 Factors Directing Electron Flow in Energetic Membranes

Mitochondrial and photosynthetic membranes are rich in species giving ESR signals during electron transfer. The full range of recent chemical bonding theory and ESR instrumentation is brought to bear in determining the electron-transfer pathways through the multiple intermediates. ESR studies provide electronic structures of intermediates in electron-transfer pathways, including electron spin density around the ring of an aromatic radical and the overlap of electron orbitals between two paramagnetic intermediates. This information is essential for determining how the surrounding protein matrix influences the path that electrons follow from one paramagnetic intermediate to another. Key electron-transfer species that are observed by ESR methods are summarized in Table 1. The list is not comprehensive. ESR examines the species listed either individually, or in pairs.

2.4.1 ESR Studies of Photosynthesis

The components of photosynthetic electron-transfer pathways, chlorophylls, quinones, and metal ions, are selected for ESR studies by preparations that include optical excitation, selective depletion of some components, and molecular biology. Because electron transfer can be primed by optical excitation in the preparations, it is possible to trap very short-lived intermediates by optical excitation of low-temperature glasses. High-frequency ESR is applied in some cases to improve resolution of ESR spectral components, much as higher frequency NMR gives better resolution of different nuclei (see also Sect. 2.2). In bacterial reaction centers, electrons flow from an excited bacteriochlorophyll dimer to bacteriopheophytin, to quinone A (Q_A), and to quinone B (Q_B). The electronic interactions between these sites are influenced by a $^{2+}$Fe ion on the pathway.

Tab. 1 Some Components of Energetic Membranes Detected by ESR.

Component	*Abbreviation*	*Membrane*
Donor cation (a chlorophyll)	$BChl_2$, $P_{700}^{+\bullet}$ chlorophyll b, $P_{680}^{+\bullet}$	RC, PS I, PS II
Reduced quinones	$Q_A^{-\bullet}$, $Q_B^{-\bullet}$, $Q^{-\bullet}$	RC, PS I, PS II
Iron-sulfur clusters	$[2Fe2S]^{-1}$, $[4Fe4S]^{-1}$, or $[4Fe4S]^{+3}$	PS I, Fd, Cplx III
Manganese clusters	Mn(III) + 3 Mn(IV)	PS II-S_2
Tyr radical	$Tyr^{+\bullet}$	PS II
Ferric heme	heme-Fe^{3+}	Cplx IV
Copper	Cu^{2+}	Cplx IV

Notes: RC: reaction center of purple bacteria; PS I: photosystem I of plants, cyanobacteria, green sulfur bacteria, and algae; PS II: photosystem II of plants, cyanobacteria, and algae; Fd: ferrodoxin; Cplx: one of the electron-transfer complexes of mitochondria; PS II-S_2: the S_2 intermediate of PS II photocycle.

Distances between electron sites on the intermediates, and the term describing electron exchange interactions between them, have now been tabulated for the bacterial system. Good agreement between these ESR measurements and electron-transfer theory is observed.

In plant photosynthetic membranes, ESR is used to examine the manner in which electrons flow ultimately into the photosynthetic oxygen-evolving complex that contains varied numbers of unpaired electrons associated with a cluster of four manganese atoms. A tyrosine radical abstracts an electron and a proton from the manganese cluster and water in steps, leading to oxygen evolution. Determining how the protein environment influences the electronic structure of the tyrosine radical and its coupling to the manganese cluster are subjects of study by the pulsed ESR techniques ESEEM and ENDOR.

2.4.2 ESR Studies of Respiratory Enzymes

In the synthesis of ATP, the mitochondrial respiratory chain employs many of the same cofactors as photosynthetic membranes, although the electron/proton transfers are not photochemically initiated. Instead, reductants or oxidants are used to isolate states with unpaired electrons. ESR contributes to understanding the electronic structure of the oxidized or reduced cofactors. In one example, ESR studies reveal an asymmetric hydrogen-bonding environment for a ubisemiquinone ($Q^{-\bullet}$) bound to bacterial quinol oxidase (QOX), *bo*$_3$. QOX is structurally and functionally related to mammalian cytochrome oxidases. In these experiments, quinones, ^{13}C-labeled individually at the carbonyl carbons, were substituted for the natural ubiquinone in QOX by detergent extraction and reconstitution. Control ESR experiments (^{14}N-ESEEM) demonstrated that the reconstituted, labeled quinones had the same hydrogen-bonding environment as native quinones. The line widths of the high-frequency ESR spectra from ^{13}C-labeled quinones differed, depending on whether the label was located on carbonyl-1 or -4. This shows that the unpaired electron spin distribution is asymmetric in the ubiquinone radical when the radical is bound to its cognate protein. The spin distribution is, in turn, related to an asymmetric hydrogen-bonding environment for the protein-bound $Q^{-\bullet}$.

3 ESR Probes of Biological Structure and Function

3.1 Spin Labeling

Nature does not always provide a free-radical ESR subject at a site of interest. Spin labels are stable, paramagnetic molecules that can be easily tailor-made by organic synthesis to provide selective structural and dynamical information via ESR spectroscopy. Most spin labels are nitroxide molecules. An advantage of spin label ESR studies is that they can be carried out at physiologically relevant temperatures and they are therefore sensitive to molecular motion. Nitroxide labeled nucleic acids, lipids, and enzyme cofactors are among the variations of spin label analogs that have been made. The largest area of applications to proteins is a molecular biology–based approach called site-directed spin labeling (SDSL) for studying the dynamics, folding, and structure of proteins.

3.1.1 Dynamics of Proteins, DNA, and Membranes

The effects of motion of a spin labeled macromolecule can be observed indirectly,

as averaging of features in the ESR spectrum or directly, by time domain ESR experiments. Effects of motion on a spin label ESR spectrum arise because the unpaired electron in the nitroxide molecule is in an asymmetric environment that renders the magnetic interactions sensitive to the orientation of the spin label in the field. When the spin label is rigidly attached to a larger molecule, ESR spectra reflect the motion of the macromolecule. The frequency of the particular ESR spectrometer used and the details of the detection scheme determine the timescale of motions that can be detected using spin label ESR. Motions occurring on timescales from milliseconds to fractions of a nanosecond have been measured. Applications in the area of dynamics include resolving spectra from distinct motional states of muscle, detecting time-dependent structural changes in bacteriorhodopsin, studies of the persistence length for DNA bending and measurements of anisotropic motion of lipids in membranes.

Figure 5 illustrates the sensitivity of ESR to spin label motion in an isotropic paraffin liquid and in a lipid membrane. The nitroxide molecules pictured have shapes that are roughly a sphere, a rectangular prism, and a long cylinder. The three peaks of equal intensity in ESR spectra of the spherical molecule, TEMPO (2,2,6,6-tetramethyl-4-ketopiperidine-1-oxygl), rotating in paraffin are characteristic of isotropic motion. The other spin labels move in paraffin with only slightly different rates about x, y, and z molecular axes (slight broadening of the third peak results). In contrast, motion of the larger spin labels is quite restricted in a lipid bilayer, as indicated by the broader ESR spectra. ESR of the fatty acid spin label reveals that the molecule rotates about the bilayer normal and has some internal flexibility. The lower spectrum from TEMPO in membranes was obtained from a more dilute lipid preparation than the upper one. The additional peak results from some TEMPO in water.

3.1.2 Site-directed Spin Labeling (SDSL)

In SDSL, a reactive nitroxide (e.g. a methanethiosulfonate) reacts with a cysteine side chain that has been placed at one or more selected sites in a protein by site-directed mutagenesis. Figure 6 shows the reaction scheme most often used in the SDSL approach. Secondary structure determination is one application of SDSL, and the approach involves stepwise substitution of nitroxides on a protein structural element. The surface exposure of the SDSL sites is determined through magnetic effects of oxygen or paramagnetic ions on ESR spectra. Interresidue distances between two spin label sites can also be determined as illustrated in Fig. 7 for phage T4 lysozyme. The overall structure of the protein shown in (a) indicates the close proximity of the two spin label side chains at residues 3 and 71. The inset (b) shows ESR spectra of folded, di-spin labeled phage lysozyme (bottom) and the same protein unfolded in urea (top). Spin labels separated by 10 to 25 Å give broadened spectra from which the interresidue distance can be determined. Once a protein derivative has been prepared with two interacting spin labels, spectral changes may be employed to determine refolding kinetics as illustrated in Fig. 7(c). Interresidue distances in membrane proteins, such as rhodopsin, are estimated by the complementary techniques of cysteine-scanning mutagenesis and paired site-directed spin labeling.

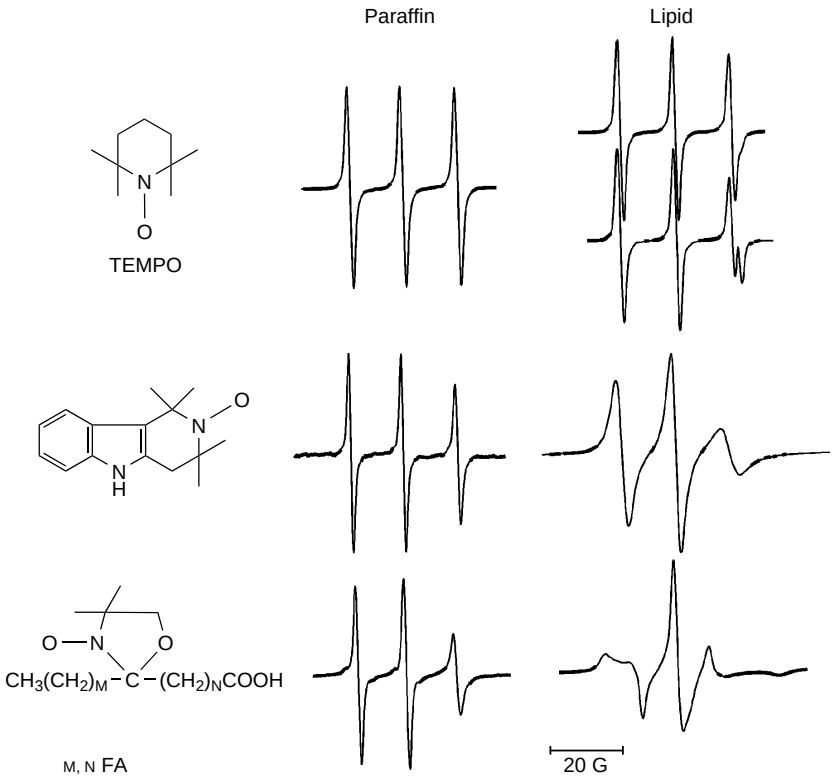

Fig. 5 The effects of restricted motion on the ESR spectra of molecules of different shapes and sizes. The spectra in the left column were obtained using a liquid paraffin solvent (isotropic) and those in the right column were with lipid bilayers (anisotropic fluid), both at 37 °C. Two concentrations of lipid were used for the upper right pair of spectra, a higher concentration in the top and a lower one below. (Redrawn with permission from Gaffney, B.J., Chen, S.C. (1977) *Methods in Membrane Biology*, Korn, E.D. (Ed.) Plenum Press, New York, Vol. 7, pp. 291–353.)

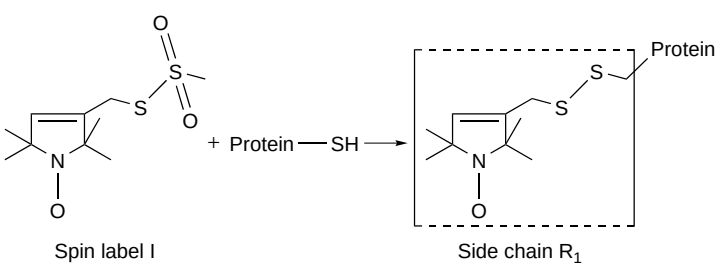

Fig. 6 Introduction of a site-directed spin label. A methanethiosulfonate spin label reacts with the natural, or genetically engineered thiol side chain of a protein to give a site-directed, spin-labeled protein. (Redrawn with permission from Hubbell, W.L., Mchaourab, H.S., Altenbach, C., Lietzow, M.A. (1996) Watching proteins move using site-directed spin labeling, *Structure* **4**, 779–783, Fig. 1.)

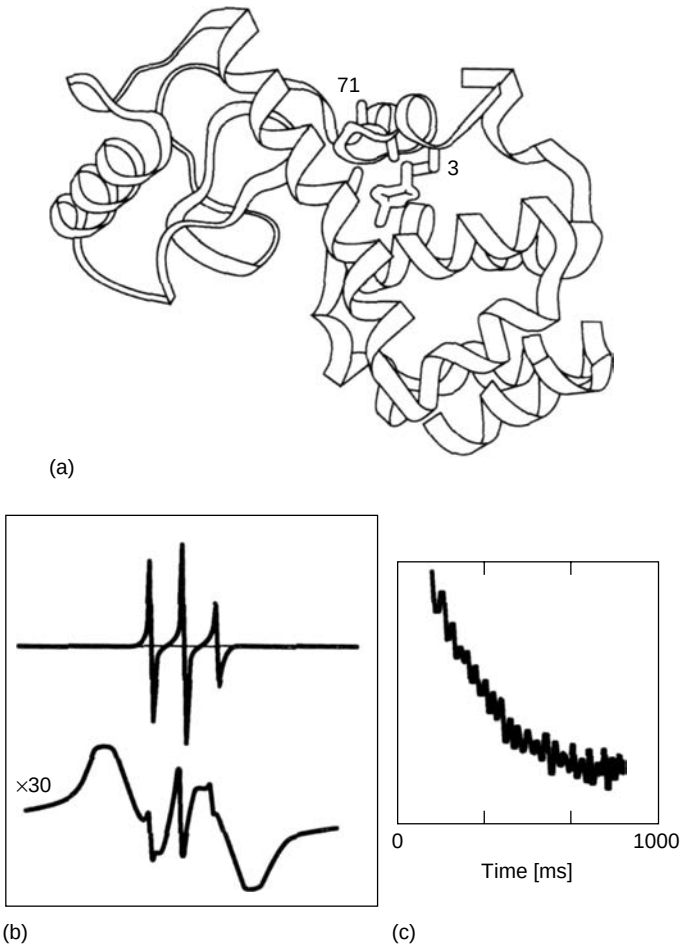

Fig. 7 Spin labeling of phage T4 lysozyme. The structure of a mutant T4 lysozyme in which two free cysteine residues were introduced (I3C and V71C) after reaction with spin labels is shown after reaction with spin labels, as in the scheme of Fig. 6. The proximity of the spin labels leads to the broad ESR spectrum shown in red in the inset (b, lower). The ESR spectrum of the same sample unfolded in urea is sharp and shows little interaction between spin labels (b, upper). The time-resolved rate of refolding can be followed by ESR (c). (Reproduced with permission from Hubbell, W.L., Mchaourab, H.S., Altenbach, C., Lietzow, M.A. (1996) Watching proteins move using site-directed spin labeling, *Structure* **4**, 779–783, Fig. 5.)

3.2 Spin Trapping

Detection of free radicals that occur transiently in cells places perhaps the most challenging demand on ESR, because the radicals have short lifetimes and are of low abundance. Free radicals can be detected indirectly by trapping them to give a longer-lived, secondary radical

in the technique called spin trapping. This technique usually employs nitroxide synthetic precursors to trap unstable free radicals such as xenobiotic radicals, superoxide, and hydroxyl radical. When there is one hydrogen on the carbon next to the nitrogen of a spin trap, the hyperfine splittings in ESR signals of the trapped radical are sensitive to whether an oxygen- or carbon-centered radical was trapped. The scheme below gives the chemistry of a spin trap nitrone reacting with trichloromethyl radical, derived from carbon tetrachloride. This approach has been used to trap similar radicals from halothane *in vivo*.

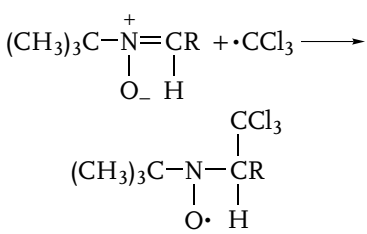

Spin traps can exhibit therapeutic effects *in vivo*, for instance, by inhibition of lipoprotein oxidation. Immunological methods are also used to detect spin-trapped sites on biomolecules.

3.3 ESR Imaging

Relaxation of spins is a dominant theme in ESR because the lifetimes of spins excited to upper energy levels are typically of the order of microseconds. A window on biological rates from 10^9 to about 100 s^{-1} results from the effects of magnetic relaxation on ESR spectra. Collision frequencies of molecules in solution, at biological concentrations, are in the range suitable for ESR relaxation responses. Relaxation of a spin probe depends on the frequency of collisions with paramagnetic oxygen. The simplest manifestation of relaxation is a change in line width of the ESR signal. This interaction is exploited to create images of oxygen levels in tissues. Spin traps selective for NO provide another basis for ESR imaging.

3.3.1 ESR Imaging of Oxygen in Tissues

The ESR spectrum of dissolved oxygen is not suitable for sensitive detection, but the ESR line width of a spin probe does respond to physiological concentrations of oxygen. EPR imaging, a technique under development, relies on spin probe relaxation to provide information about spatial distribution of oxygen in tissues, for instance, in tumors. In a sense, ESR imaging of oxygen concentration in tissues is a spin trap experiment in which the magnetism of oxygen is "trapped" through its magnetic, but not chemical, interactions with a stable spin probe. The degree of broadening of the ESR signal is transformed into an image.

A direct comparison of images of the same mouse tumor obtained by ESR imaging (EPRI) and by MRI has been made. In the experimental design, the leg bearing the tumor is centered in an ESR resonator or an MRI coil for the two imaging modalities. Paramagnetic reagents are applied intravenously in the ESR experiment, at a level similar to that used for MRI contrast agents. Both stable trityl (triphenylmethyl) radicals and nitroxides are used in EPRI, but the extremely narrow signal from trityl probes provides spatial resolution of about 1 mm in a tumor of $7 \times 7 \times 10$ mm. An ESR spectrum of the reagent in each pixel is obtained (spectral-spatial imaging). Separate calibration experiments demonstrate linear response of the ESR line width to oxygen concentration in the range 0 to 75 torr $\pm \sim 3$ torr. ESR and MRI obtain

comparable images of regions of the spatial distribution of oxygen, but the ESR experiment has the advantage that the oxygen concentrations are measured quantitatively. ESR imaging is performed at low frequency because the higher frequencies used in most ESR spectroscopy cannot penetrate far into wet tissues.

3.3.2 ESR Imaging of NO

A focus of ESR imaging has been to examine formation of NO from nitrite in ischemic rat hearts. With increasing durations of ischemia, the myocardial pH drops and this facilitates reduction of nitrite to NO. For example, both *ex vivo* and *in vivo* experiments are conducted in hearts loaded with ~1 mM nitrite and the NO spin trap, ferrous (N-methyl-D-glucamine dithiocarbamate)$_2$. By comparing pH measurements from ^{13}P-NMR with the quantity of trapped NO in the heart, significant NO formation from nitrite is demonstrated. Control experiments with inhibitors of nitric oxide synthase (NOS) and cytochrome oxidase augment these experiments.

3.4 Long-range Distance (6–60 Å) Measurements by ESR

The influence by one spin on the relaxation of another can also extend over a distance. In practice, this makes it possible to measure, by ESR, the distance between two paramagnetic centers, separated by up to 60 Å, in a biomolecule.

The advantages of using ESR methods to measure selected nonbonding distances in biomolecules are that the measurements are usually free of paramagnetic sites other than those involved in the measurement, and the sample can be opaque or semisolid. Recently, methods for ESR distance measurement have been tested and refined. The choice of a method depends on the relative relaxation rates of the two paramagnets between which a distance is sought. Generally, organic radicals such as spin labels and tyrosine radicals have spins with slow relaxation rates, whereas metal ion spins relax faster. Thus, specific methods are chosen for cases in which a slow relaxer is paired with either a second slow, or second fast relaxing paramagnetic species. The range 6 to 60 Å is a conservative estimate of distances that can be determined by ESR studies of two interacting paramagnets.

See also Bioorganic Chemistry; Free Radicals in Biochemistry and Medicine.

Bibliography

Books and Reviews

Berliner, L.J., Eaton, G.R., Eaton, S.S. (Eds.) (2001) Distance measurements in biological systems by EPR, *Biological Magnetic Resonance*, Vol. 19, Plenum Press, New York, and earlier volumes in this series.

Cammack, R. (Ed.) (1985) The applications of electron spin/paramagnetic resonance in biochemistry; a colloquium in honor of Dr. Helmut Beinert, *Biochem. Soc. Trans.* **13**, 541–634.

Eaton, G., Eaton, S., Salikhov, K. (1997) *Foundations of Modern EPR*, World Scientific, Singapore.

Ehrenberg, A., Malmström, B.G., Vänngård, T. (1967) *Magnetic Resonance in Biological Systems*, Werner-Gren Center International Symposium Series, Vol. 9, Pergamon Press, Oxford.

Feher, G. (2002) My road to biophysics: picking flowers on the way to photosynthesis, *Annu. Rev. Biophys. Biomol. Struct.* **31**, 1–44.

Frey, P.A. (2001) Radical mechanisms of enzymatic catalysis, *Annu. Rev. Biochem.* **70**, 121–148.

Hubbell, W.L., Altenbach, C., Hubbell, C.M., Khorana, H.G. (2003) Rhodopsin structure, dynamics, and activation: a perspective from crystallography, site-directed spin labeling, sulfhydryl reactivity, and disulfide cross-linking, *Adv. Prot. Chem.* **63**, 243–290.

McConnell, H.M., Gaffney-McFarland, B. (1970) Physics and chemistry and spin labels, *Q. Rev. Biophys.* **3**, 91–136.

Pilbrow, J.R. (1990) *Transition Ion Electron Paramagnetic Resonance*, Clarendon Press, Oxford.

Rosen, G.M., Britigan, B.E., Halpern, H.J., Pou, S. (1999) *Free Radicals: Biology and Detection by Spin Trapping*, Oxford University Press, Oxford.

Schultz, B.E., Chan, S.I. (2001) Structures and proton-pumping strategies of mitochondrial respiratory enzymes, *Annu. Rev. Biophys. Biomol. Struct.* **30**, 23–65.

Schweiger, A., Jeschke, G. (2001) *Principles of Pulse Electron Paramagnetic Resonance*, Oxford University Press, Oxford.

Weil, J.A., Bolton, J.R., Wertz, J.E. (1994) *Electron Paramagnetic Resonance, Elementary Theory and Practical Applications*, Wiley-Interscience, New York.

Primary Literature

Al'tshuler, S., Zavoisky, Y.K., Kozyrev, B.M. (1944) New method to study paramagnetic absorption, *Zhurn. Eksperiment. Teoret. Fiziki* **14**, 407–409.

Bar, G., Bennati, M., Nguyen, H.-H.T., Ge, J., Stubbe, J., Griffin, R.G. (2001) High-frequency (140-GHz) time domain EPR and ENDOR spectroscopy: the tyrosyl radical-diiron cofactor in ribonucleotide reductase from yeast, *J. Am. Chem. Soc.* **123**, 3569–3576.

Bellew, B.F., Halkides, C.J., Gerfen, G.G., Griffin, R.G., Singel, D.J. (1996) High frequency (139.5 GHz) electron paramagnetic resonance characterization of Mn(II)-H2^{17}O interactions in GDP and GTP forms of p21 *ras*, *Biochemistry* **35**, 12186–12193.

Burns, C.S., Aronoff-Spencer, E., Legname, G., Prusiner, S.B., Antholine, W.E., Gerfen, G.J., Peisach, J., Millhauser, G.L. (2003) Copper coordination in the full-length, recombinant prion protein, *Biochemistry* **42**, 6794–6803.

Calvo, R., Abresch, E.C., Bittl, R., Feher, G., Hofbauer, W., Isaacson, R.A., Lubitz, W., Okamura, M.Y., Paddock, M.L. (2000) EPR study of the molecular and electronic structure of the semiquinone biradical $Q_A^{-\bullet}$ $Q_B^{-\bullet}$ in photosynthetic reaction centers from *Rhodobacter sphaeroides*, *J. Am. Chem. Soc.* **122**, 7327–7341.

Calvo, R., Isaacson, R.A., Abresch, E.C., Okamura, M.Y., Feher, G. (2002) Spin-lattice relaxation of coupled metal-radical spin-dimers in proteins: application to Fe^{2+}-cofactor ($Q_A^{-\bullet}$, $Q_B^{-\bullet}$, $\phi^{-\bullet}$) dimers in reaction centers from photosynthetic bacteria, *Biophys. J.* **83**, 2440–2456.

Canaan, S., Nielsen, R., Ghomashchi, F., Robinson, B.H., Gelb, M.H. (2002) Unusual mode of binding of human Group IIA secreted phospholipase A2 to anionic interfaces as studied by continuous wave and time domain electron paramagnetic resonance spectroscopy, *J. Biol. Chem.* **277**, 30984–30990.

Carmieli, R., Manikandan, P., Kalb, A.J., Goldfarb, D. (2001) Proton positions in the Mn^{2+} binding site of concanavalin A as determined by single-crystal high-field ENDOR spectroscopy, *J. Am. Chem. Soc.* **123**, 8378–8386.

Cooper, C.E. (Ed.) (1999) Collected papers on biochemistry of nitric oxide, *Biochim. Biophys. Acta Bioenergetics* **1411**, 215–488.

Davydov, R., Makris, T.M., Kofman, V., Werst, D.E., Sligar, S.G., Hoffman, B.M. (2001) Hydroxylation of camphor by reduced oxycytochrome P450cam: mechanistic implications of EPR and ENDOR studies of catalytic intermediates in native and mutant enzymes, *J. Am. Chem. Soc* **123**, 1403–1415.

Debus, R.J., Campbell, K.A., Gregor, W., Li, Z.-L., Burnap, R.L., Britt, R.D. (2001) Does histidine 332 of the D1 polypeptide ligate the manganese cluster in photosystem II? An electron spin echo envelope modulation study, *Biochemistry* **40**, 3690–3699.

Devaux, P., McConnell, H.M. (1972) Lateral diffusion in spin-labeled phosphatidylcholine multilayers, *J. Am. Chem. Soc.* **94**, 4475–4481.

Eaton, S.S., Eaton, G.R. (2002) *Electron Paramagnetic Resonance Techniques for Measuring Distances in Proteins in Structures and Mechanisms: from Ashes to Enzymes*, ACS Symposium Series 827, American Chemical Society, Washington, DC, pp. 321–339.

Edwards, T.E., Okonogi, T.M., Robinson, B.H., Sigurdsson, S.T. (2001) Site-specific incorporation of nitroxide spin-labels into internal sites of the TAR RNA; structure-dependent dynamics of RNA by EPR spectroscopy, *J. Am. Chem. Soc.* **123**, 1527–1528.

Elas, M., Williams, B.B., Parasca, A., Mailer, C., Pelizzari, C.A., Lewis, M.A., River, J.N., Karczmar, G.S., Barth, E.D., Halpern, H.J. (2003) Quantitative tumor oxymetric images from 4D electron paramagnetic resonance imaging (EPRI): methodology and comparison with blood oxygen level-dependent (BOLD) MRI, *Magn. Reson. Med.* **49**, 682–691.

Fu, Z., Aronoff-Spencer, E., Backer, J.M., Gerfen, G.J. (2003) The structure of the inter-SH2 domain of class IA phosphoinositide 3-kinase determined by site-directed spin labeling EPR and homology modeling, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3275–3280.

Gaffney, B.J., Maguire, B.C., Weber, R.T., Maresch, G.G. (1999) Disorder at metal sites in proteins: a high frequency EMR study, *Appl. Magn. Res.* **16**, 207–222.

Gaffney, B.J., Marsh, D. (1998) High-frequency, spin label EPR of non-axial ordering and motion in cholesterol-containing membranes, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12940–12943.

Ge, M., Cohen, J.S., Brown, H.A., Freed, J.H. (2001) ARF6 binding to PIP2-containing vesicles creates defects in the bilayer structure: ESR studies on model membranes (2001), *Biophys. J.* **81**, 994–1105.

Gow, A.J., Benjamin, P., Luchsinger, B.P., John, R., Pawloski, J.R., David, J., Singel, D.J., Jonathan, S., Stamler, J.S. (1999) The oxyhemoglobin reaction of nitric oxide, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9027–9032.

Griffin, M., Muys, A., Noble, C., Wang, D., Eldershaw, C., Gates, K.E., Burrage, K., Hanson, G.R. (1999) XSophe, a computer simulation software suite for the analysis of electron paramagnetic resonance spectra, *Mol. Phys. Rep.* **26**, 60–84.

Griffith, O.H., McConnell, H.M. (1966) A nitroxide-maleimide spin label, *Proc. Natl. Acad. Sci. U.S.A.* **55**, 8–11.

Grimaldi, S., Ostermann, T., Weiden, N., Mogi, T., Miyoshi, H., Ludwig, B., Michel, H., Prisner, T.F., MacMillan, F. (2003) Asymmetric binding of the high-affinity QH-ubisemiquinone in quinol oxidase (bo_3) from *Escherichia coli* studied by multifrequency electron paramagnetic resonance spectroscopy, *Biochemistry* **42**, 5632–5639.

Halpern, H.J., Yu, C., Barth, E., Peric, M., Rosen, G.M. (1995) In situ detection, by spin trapping, of hydroxyl radical markers produced from ionizing radiation in the tumor of a living mouse, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 796–800.

Hille, R., Olson, J.S., Palmer, G. (1979) Spectral transitions of nitrosyl hemes during ligand binding to hemoglobin, *J. Biol. Chem.* **254**, 12110–12120.

Hoffman, B.M. (2003) Electron-nuclear double resonance spectroscopy (And Electron Spin-Echo Envelope Modulation Spectroscopy) in bioinorganic chemistry, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3575–3578.

Högbom, M., Galander, M., Andersson, M., Kolberg, M., Hofbauer, W., Günter Lassmann, G., Nordlund, P., Lendzian, F. (2003) Displacement of the tyrosyl radical cofactor in ribonucleotide reductase obtained by single-crystal high-field EPR and 1.4-Å X-ray data, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3209–3214.

Hoogstraten, C.G., Grant, C.V., Horton, T.E., DeRose, V.J., Britt, R.D. (2002) Structural analysis of metal ion ligation to nucleotides and nucleic acids using pulsed EPR spectroscopy, *J. Am. Chem. Soc.* **124**, 834–842.

Hubbell, W.L., McConnell, H.M. (1968) Spin-label studies of the membranes of nerve and muscle, *Proc. Natl. Acad. Sci. U.S.A.* **61**, 12–16.

Hubbell, W.L., Mchaourab, H.S., Altenbach, C., Lietzow, M.A. (1996) Watching proteins move using site-directed spin labeling, *Structure* **4**, 779–783.

Hustedt, E.J., Beth, A.H. (2001) The sensitivity of saturation transfer electron paramagnetic resonance spectra to restricted amplitude uniaxial rotational diffusion, *Biophys. J.* **81**, 3156–3165.

Kennedy, M.C., Antholine, W.E., Beinert, H. (1997) An EPR investigation of the products of the reaction of cytosolic and mitochondrial aconitases with nitric oxide, *J. Biol. Chem.* **272**, 20340–20347.

Klett, R., Töring, J.T., Plato, M., Möbius, K., Bönigk, B., Lubits, W. (1993) Determination of the g-tensor of the primary donor cation radical in single crystals of *Rhodobacter sphaeroides* R-26 reaction centers by 3-mm high field EPR, *J. Phys. Chem.* **97**, 2015–2020.

Knecht, K.T., DeGray, J.A., Mason, R.P. (1992) Free radical metabolism of halothane *in*

vivo: radical adducts detected in bile, *Mol. Pharmacol.* **21**, 943–949.

Kornberg, R.D., McConnell, H.M. (1971) Inside-outside transitions of phospholipids in vesicle membranes, *Biochemistry* **10**, 1111–1120.

LaConte, L.E.W., Voelz, V., Nelson, W., Enz, M., Thomas, D.D. (2002) Molecular dynamics simulation of site-directed spin labeling: experimental validation in muscle fibers, *Biophys. J.* **83**, 1854–1866.

Lendzian, F., Huber, M., Isaacson, R.A., Endeward, B., Plato, M. (1993) The electronic structure of the primary electron donor cation radical in *Rhodobacter sphaeroides* R-26: ENDOR and TRIPLE resonance studies on single crystals of reaction centers, *Biochim. Biophys. Acta* **1183**, 139–160.

McConnell, H.M. (1956) Effect of anisotropic hyperfine interactions on paramagnetic relaxation in liquids, *J. Chem. Phys.* **25**, 709–711.

McConnell, H.M., Chestnut, D.B. (1958) Theory of isotropic hyperfine interactions in π-electron radicals, *J. Chem. Phys.* **28**, 107–117.

Mustafi, D., Makinen, M.W. (1995) Structure, conformation, and probable mechanism of hydrolysis of a spin-labeled penicillin revealed by electron nuclear double resonance spectroscopy, *J. Am. Chem. Soc.* **117**, 6739–6746.

Norris, J.R., Uphaus, R.A., Crespi, H.L., Katz, J.J. (1971) Electron spin resonance of chlorophyll and the origin of signal I in photosynthesis, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 625–628.

Persson, M., Harbridge, J.R., Hammarström, P., Mitri, R., Mårtensson, L.-G., Carlsson, U., Eaton, G.R., Eaton, S.S. (2001) Comparison of EPR methods to determine distances between spin labels on human carbonic anhydrase II, *Biophys. J.* **80**, 2886–2897.

Qin, P.Z., Hideg, K., Feigon, J., Hubbell, J.L. (2003) Monitoring RNA base structure and dynamics using site-directed spin labeling, *Biochemistry* **42**, 6772–6783.

Ramirez, D.C., Chen, Y.-R., Mason, R.P. (2003) Immunochemical detection of hemoglobin-derived radicals formed by reaction with hydrogen peroxide: involvement of a protein-tyrosyl radical, *Free Radical Biol. Med.* **34**, 830–839.

Reddy, T.J., Iwama, T., Halpern, H.J., Rawal, V.H. (2002) General synthesis of persistent trityl radicals for EPR imaging of biological systems, *J. Org. Chem.* **67**, 4635–4639.

Sefton, B.M., Gaffney, B.J. (1974) Effect of the viral proteins on fluidity of the membrane lipids in sindbis virus, *J. Mol. Biol.* **90**, 343–358.

Steinhoff, H.J., Mollaaghababa, R., Altenbach, C., Hideg, K., Krebs, M., Khorana, H.G., Hubbell, W.L. (1994) Time-resolved detection of structural changes during the photocycle of spin-labeled bacteriorhodopsin, *Science* **266**, 105–107.

Subczynski, W.K., Hyde, J.S. (1981) The diffusion-concentration product of oxygen in lipid bilayers using the spin-label T_1 method, *Biochim. Biophys. Acta* **643**, 283–291.

Subczynski, W.K., Kusumi, A. (2003) Dynamics of raft molecules in the cell and artificial membranes: approaches by pulse EPR spin labeling and single molecule optical microscopy, *Biochim. Biophys. Acta Biomembr.* **1610**, 231–243.

Swamy, M.J., Ramakrishnan, M., Angerstein, B., Marsh, D. (2000) Spin-label electron spin resonance studies on the mode of anchoring and vertical location of the N-acyl chain in N-acylphosphatidylethanolamines, *Biochemistry* **39**, 12476–12484.

Thomas, C.E., Ohlweiler, D.F., Kalyanaraman, B. (1994) Multiple mechanisms for inhibition of low density lipoprotein oxidation by novel cyclic nitrone spin traps, *J. Biol. Chem.* **269**, 28055–28061.

Tommos, C., Tang, X.-S., Warncke, K., Hoganson, C.W., Styring, S., McCracken, J., Diner, B., Babcock, G.T. (1995) Spin-density distribution, conformation, and hydrogen bonding of the redox-active tyrosine Yz in photosystem II from multiple electron magnetic-resonance spectroscopies: implications for photosynthetic oxygen evolution, *J. Am. Chem. Soc.* **117**, 10325–10335.

Török, M., Milton, S., Kayed, R., Wu, P., McIntire, T., Glabe, C.G., Langen, R. (2002) Structural and dynamic features of Alzheimer's A peptide in amyloid fibrils studied by site-directed spin labeling, *J. Biol. Chem.* **277**, 40810–40815.

Trofanchuk, O., Stein, M., Geßner, Ch., Lendzian, F., Higuchi, Y., Lubitz, W., Miyazaki, F. (2000) Single crystal EPR studies of the oxidized active site of [NiFe] hydrogenase from *Desulfovibrio vulgaris*, *J. Biol. Inorg. Chem.* **5**, 36–44.

Willingham, G.L., Gaffney, B.J. (1983) Reactions of spin label cross-linking reagents with

red blood cell proteins, *Biochemistry* **22**, 892–898.

Wu, F., Katsir, L.J., Seavy, M., Gaffney, B.J. (2003) Role of radical formation at Y193 in the allene oxide synthase domain of a lipoxygenase-AOS fusion protein from coral, *Biochemistry* **42**, 6871–6880.

Yonetani, T., Tsuneshige, A., Zhou, Y., Chen, X. (1998) Electron paramagnetic resonance and oxygen binding studies of -nitrosyl hemoglobin. A novel oxygen carrier having NO-assisted allosteric functions, *J. Biol. Chem.* **273**, 20323–20333.

Zweier, J.L., Samouilov, A., Kuppusamy, K. (1999) Non-enzymatic nitric oxide synthesis in biological systems, *Biochim. Biophys. Acta* **1411**, 250–262.

Zweier, J.L., Wang, P., Periannan Kuppusamy, P. (1995) Direct measurement of nitric oxide generation in the ischemic heart using electron paramagnetic resonance spectroscopy, *J. Biol. Chem.* **270**, 304–307.

Electroporation and Electrofusion

DonaldC. Chang
Hong Kong University of Science and Technology, Kowloon, Hong Kong

1 Principles 136
1.1 Electric Field–induced Membrane Breakdown 136
1.2 Uptake of Exogenous Molecules during Electroporation 141
1.3 Electric Field–induced Cell Fusion 142

2 Techniques 144
2.1 Waveforms Used for Electroporation and Electrofusion 144
2.2 Basic Procedures of Electroporation and Electrofusion 146
2.3 *In situ* Electroporation of Attached Cultured Cells 147
2.4 Factors Affecting the Efficiency of Electroporation and Electrofusion 148

3 Applications 150
3.1 Applications of Electroporation in Gene Transfer 150
3.2 Electroinjection of Other Molecules 151
3.3 Study of the Mechanism of Membrane and Cell Fusion 152
3.4 Production of Hybridomas and Antibodies 152

4 Perspectives 153
4.1 Future Applications of Electroporation in Molecular Biology 153
4.2 Electroporation as an Important Tool for Studying Cell Signaling 153
4.3 Use of Electroporation for Gene Therapy 154

Bibliography 155
Books and Reviews 155
Primary Literature 155

Keywords

Electropermeabilization
Using a pulsed electric field to temporally permeabilize the cell membrane.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Electroporation
Same as electropermeabilization.

Electrofusion
Using a pulsed electric field to fuse two adjacent cells.

Electroporation is a phenomenon in which the cell membrane is temporarily permeabilized by exposure to an intense electric field. This phenomenon can be used to introduce a variety of exogenous molecules, particularly DNA, into living cells. Electrofusion is a related phenomenon by which neighboring cells can be induced to fuse by applying a pulse of electric field. Both electroporation and electrofusion are related to the electrical breakdown of the cell membrane. Electroporation is now a principal method of gene transfer, for both prokaryotic and eukaryotic cells. Electrofusion is found to be the most efficient method of cell fusion and has important uses in agriculture and in the production of hybridomas.

1 Principles

1.1 Electric Field–induced Membrane Breakdown

The phenomenon of electroporation has been known for some time; studies can be traced back to the early sixties. The application of electroporation in molecular biology, however, is relatively recent. The first report of using electroporation for gene transfer appeared only in 1982.

Electroporation is the result of membrane breakdown induced by an applied electric field. Suppose a spherical cell with diameter r is exposed to an external electric field for a time period t. The cell membrane will develop an induced membrane potential V_m according to the following equation:

$$V_m = 1.5 Er \cos\theta \{1 - \exp(-t/\tau\tau_c)\} \quad (1)$$

where E is the field strength, θ is the angle between the field and the normal of the membrane, and τ_c is the relaxation time of the cell. For a typical biological cell, τ_c is approximately equal to

$$\tau_c = rC_m(\rho_i + 0.5\rho_o) \quad (2)$$

where C_m is the unit capacitance of the cell membrane and ρ_i and ρ_o are the resistivity of the cytoplasm and the external medium respectively. For small cells, such as red blood cells, τ_c is on the order of 1 μs. Then, when the external electric field is applied for a sufficiently long time period, say 20 μs, the induced membrane potential will reach the steady value

$$V_m = 1.5 Er \cos\theta \quad (3)$$

From this equation, it is apparent that the induced membrane potential is not uniform across the entire cell membrane. The induced membrane potential is highest at

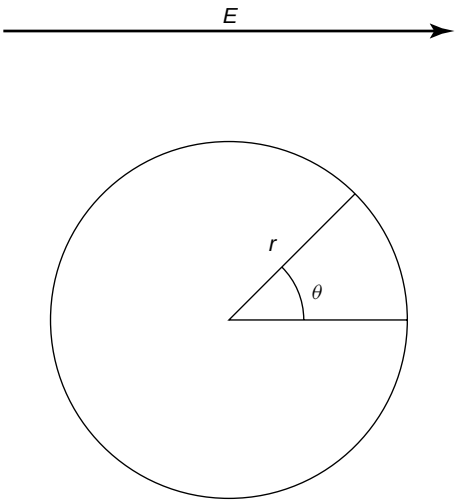

Fig. 1 Schematic diagram showing a spherical cell exposed to an external electric field, *E*.

places where the membrane faces the poles of the electric field (Fig. 1).

The above analysis holds true so long as the cell membrane remains intact and its electrical properties do not change. When the external field strength is so high that the induced membrane potential exceeds a threshold potential, V_{th}, the bilayer membrane will experience an electrical breakdown; the cell membrane can no longer maintain its structural integrity and becomes permeable. This phenomenon is called *electroporation* or *electropermeabilization*. For most cells, the threshold potential is typically on the order of 0.5 to 1 V, which is about 10 times larger than the normal resting potential of an animal cell. The occurrence of electrical breakdown is related to the fact that the cell membrane is a very thin structure. The insulating property of the cell membrane depends on its lipid bilayer, which is only 5 nm thick. When the potential across the cell membrane is increased to 1 V, the electric field within the bilayer reaches 2×10^6 V cm^{-1}. The membrane cannot withstand an electric field of such magnitude and it undergoes a dielectric breakdown.

When the cell membrane is permeabilized by an applied field, molecules that normally could not penetrate the membrane can now pass through easily. These molecules include ions, metabolites, carbohydrates, proteins, nucleic acids, and others. Some of the molecules that enter the cell during the electropermeabilized state could be very large. It has been reported that DNA molecules of sizes up to 150 kbp can be taken up by electropermeabilized cells. The uptake of these very large molecules, however, may involve some complex processes. As for small molecules, such as small carbohydrates or proteins of low molecular weight, it is evident that they can enter the cell by simple diffusion through the permeabilized cell membrane.

If the external electric field is not excessively high, the electropermeabilized cell can recover by resealing its membrane after the external field is removed. Using a model membrane, such as a lipid bilayer, the processes of electroporation

and recovery can be studied by monitoring the changes in the electrical properties of the membrane in response to the application of the external field. For the lipid bilayer, the electroporation process involves three distinct stages:

1. *Nonpermeabilized state*. This is the state of the membrane before the external field is applied. The conductance of the membrane is extremely low.
2. *Permeabilized state*. When the external field is applied, the induced membrane potential (V_m) builds up exponentially. When V_m approaches the threshold potential (V_{th}), the conductance of the membrane increases drastically, indicating that the membrane has become permeabilized to certain ions.
3. *Recovery state*. If the external field is applied only for a brief period and the field strength is not too high, the bilayer will recover by gradually resealing itself after the external electric field is turned off. In this case, the bilayer membrane is said to experience a *reversible breakdown*. The recovery process can be monitored by observing an exponential decrease in the membrane conductance following the termination of the applied field. This recovery process appears to be very rapid; the time constant is generally on the order of submilliseconds. If the external field is applied for too long or if the field strength is too high, the bilayer will not be able to recover and it will rupture even after the applied field is removed. In this situation, the membrane is said to undergo a process called *irreversible breakdown*.

In electroporation of real cell membranes, the situation is more complicated. There is evidence suggesting that the electroporation process may involve several substages. Furthermore, the membrane may undergo secondary structural changes following the removal of the external field. Thus, when an electric field is applied to electroporate cells, the cell membrane may undergo at least four different stages of change:

1. *Nonpermeabilized state*. Before the external field is turned on, the cell membrane is in its normal intact state.
2. *Permeabilized state due to primary structural changes*. After the external field is applied, the induced membrane potential quickly reaches the threshold potential. As a result, the membrane experiences an electrical breakdown and undergoes a structural change to become permeabilized. This electric field–mediated membrane breakdown is commonly referred to as *primary structural changes*.
3. *Permeabilized state due to secondary structural changes*. Once the cell membrane becomes permeabilized, molecules can pass through the membrane to enter or leave the cell. Such flow of material can induce further structural changes in the membrane, making it even more permeable to other molecules. Furthermore, breakdown of the membrane structure in a local region may affect the stability of the membrane structures in neighboring regions and thus could further enlarge the membrane pores. These secondary structural changes can continue even after the removal of the external electric field.
4. *Recovery state*. After the external field has been turned off and the forces causing the pores to expand have subsided, the membrane will repair itself and start to reseal. The mechanism of membrane healing is not yet well understood. However, it is well known that a cell

can repair its own membrane following puncture or scratch. This recovery process takes considerably longer than in the case of a lipid bilayer model membrane. Moreover, the resealing process itself is believed to involve several substages. The larger membrane pores may reseal first, with a time constant on the order of seconds. They, however, do not reseal completely but reduce into residual pores that may linger for considerably longer times. Depending on the cell type, the resealing process may take several minutes to half an hour at room temperature. At lower temperatures, such as at 4 °C, it will take even longer.

The structure and properties of the electric field–induced pores have been investigated actively in the last two decades. The results are still far from complete. There have been different suggestions about the structures of these membrane pores. Some thought that these pores might appear in the form of cracks in the membrane. A few others proposed that they are craterlike round pores. There are also conflicting estimates of the size of the pores. One study concluded that, on the basis of examination of transport properties of the electropermeabilized cell membrane to carbohydrate molecules of different sizes, the membrane pores are very small, with diameters on the order of 1 nm. Another group examined the size of labeled molecules that can be loaded into red cell ghosts and concluded that the size of the membrane pores could reach 10 nm. However, in view of the fact that large macromolecules such as DNA can pass through the cell membrane during the electropermeabilized state, it is suspected that the electric field–induced pores may be significantly larger. We recently used a rapid-freezing electron microscopy (EM) technique to examine the dynamic changes in membrane

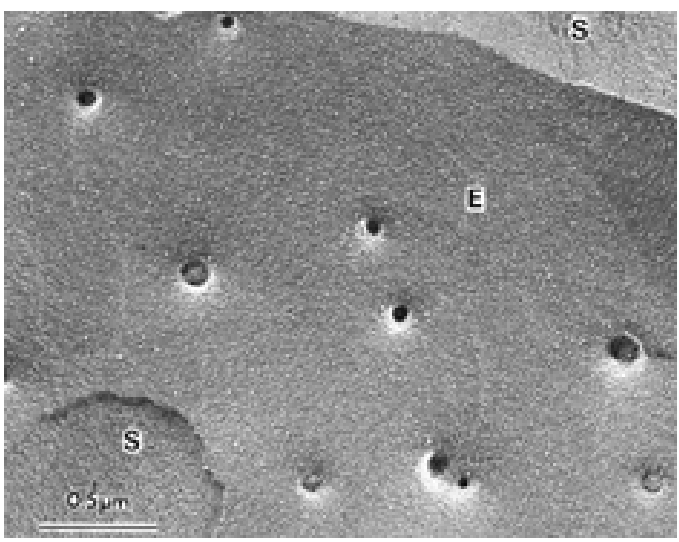

Fig. 2 External membrane face (labeled "E") of an electro permeabilized human red blood cell frozen at 40 ms after the application of an electrical pulse. The frozen extracellular medium is labeled "S."

structure during the electropermeabilized state (Fig. 2). Large volcano-shaped porelike structures with diameters up to 100 nm were observed in the freeze-fractured membrane. Thus, the size of "electropores" could be much larger than that indicated earlier.

When we examined the membrane structure of cells frozen at different times after application of the electric pulse, we found that the membrane pores induced by the electric field underwent significant changes with time. The structure of the red cell membrane frozen at $t = 1$ ms appeared to be smooth and without any unusual structural changes, just like the control membrane. (Here "t" is the time delay between the electric pulse and the time when the specimen was frozen). These observations suggest that, if electropores were formed in the first millisecond following application of the electric field, the diameter of these early pores must be very small (less than the resolution of the freeze-fracture EM, about 2 nm). The earliest time at which we could detect any significant structural changes in the cell membrane was at $t = 2.5$ ms, where deep porelike membrane openings (20–40 nm in diameter) were found in a few E-face membranes of the electropermeabilized red cells. At a later time ($t = 20$ ms), circular membrane openings became more abundant, and the size of the openings appeared to increase. At $t = 40$ ms, porelike membrane openings could be observed in almost all cell membranes examined, and their diameters had expanded to the range of 20 to 120 nm (Fig. 2). At $t = 0.2$ to 1.7 s, circular membrane openings similar to those observed at $t = 40$ ms could be found in most red cells. At $t = 5$ s, the deep porelike structures had almost disappeared and were replaced by numerous pitlike indentations in the membrane, which may be remnants of previous membrane evaginations. These pits might still have an opening in their centers, but such an opening must be very small. Since it was not possible to resolve the shape of such openings with freeze-fracture EM, their diameter had to be less than 3 nm.

On the basis of the time-dependent study of membrane structure using rapid-freezing EM, we believed that the membrane pore induced by the electric field was a dynamic structure whose shape and diameter underwent rapid change with time. The dynamic change in electropores at various stages is conceptually summarized in Table 1.

Thus, the different estimates of pore size reported in the literature may be

Tab. 1 Dynamic change in membrane pores in red blood cells during electroporation.

Time after pulsing	*State of the cell membrane*	*Estimated pore size*
<0 s	Intact membrane	<0.2 nm
0.001–0.01 s	Pore creation due to electrical breakdown	1–10 nm
0.01–1 s	Pore expansion due to material flow	10–100 nm
1–10 s	Membrane starts to reseal; pore shrinking	1–10 nm
0.1–30 min	Membrane continues to reseal; existence of long-lasting residual pores	1 nm
>30 min	Membrane fully recovered	<0.2 nm

partially due to the differences in measurement techniques and partially related to the dynamic nature of the pores, which can appear differently at different stages. The large porelike structures observed in the rapid-freezing EM study may represent transient membrane pores through which DNA molecules can enter the cell. The small pores found in the transport study of carbohydrates, on the other hand, may represent the long-lasting residue pores at the recovery stage. Since the measurements of carbohydrate transport require minutes, most of the large membrane pores would have undergone partial resealing during the measurement period.

1.2 Uptake of Exogenous Molecules during Electroporation

Molecules with a variety of shapes and molecular weights can be taken up by cells during the electropermeabilized state. These molecules include DNA, RNA, proteins, drugs, metabolites, ions, molecular probes, and so on. Depending on the size and charge of the transported molecules, the process of molecular uptake can be different. For small molecules, they could enter cells by simple diffusion during the electropermeabilized state. In studies using either lipid vesicles or red cell ghosts, it has been demonstrated that loading of small molecules, including ions, carbohydrates of low molecular weight, or dyes, during electroporation can be explained simply on the basis of a diffusion model. The uptake of larger molecules such as DNA is more complicated; the process is believed to involve a number of sequential steps. At this point, the exact mechanism of DNA uptake during electroporation is not yet clearly understood, but there is evidence suggesting that the following processes are involved:

1. Before application of the external electric field, a certain fraction of DNA molecules may have already been bound to the cell surface. Such binding could be a required intermediate step for DNA to enter the cell when the cell membrane is electropermeabilized.
2. Application of an external field creates pores on the cell membrane. Some of these pores may have a transient opening that is large enough to allow DNA molecules to pass through.
3. Through an electrophoretic process, the negatively charged DNA molecules may be driven toward these pores, and they enter the cell by the applied electric field.
4. Alternatively, the entire DNA molecule may not actually enter the cell during the electropermeabilized state. Instead, the applied electric field may disrupt the membrane structure to create a phospholipid/DNA complex that would then stimulate the internalization of DNA by endocytosis.

The processes discussed above have been examined in a number of studies and are supported by some of the experimental results. For example, in some systems, the concentration of magnesium ions was found to have a significant effect on gene transfer efficiency. Such a finding may be explained on the basis of DNA binding to the cell surface. Several laboratories, including our own, found that, in order to have an effective gene transfer, DNA must be first added to the cell mixture before applying the electric field. This finding suggests that either the large transient pores have a very short life or DNA must be driven into the cell by the applied field. Using cultured cells grown on a filter

membrane, it was possible to examine the effect of the polarity of the electric field on gene uptake. We found that when DNA was applied on the cathode side, the gene transfection efficiency was much higher. Recently, it was reported that in a single-cell visualization study, fluorescent-labeled plasmids were observed to enter the electropermeabilized cell on the side facing the cathode. This finding suggests that DNA entry may involve an electrophoretic process. At present, however, the existing evidence cannot exclude the involvement of other processes, such as endocytosis.

1.3 Electric Field–induced Cell Fusion

One can apply a pulsed electric field not only to electropermeabilize cells but also to induce cells to fuse. When two neighboring cells are placed very close to each other, applying a pulsed electric field similar to that used in the electroporation process will induce these cells to fuse. This process is called *electrofusion*. The development of electrofusion has a slightly shorter history in comparison with electroporation. Induction of fusion of plant cells by electrical stimulation was first reported in 1979. Electrofusion of several other cell types was reported in the following years. Because of its many important applications, including cell hybridization, animal cloning, and generation of antibody-producing hybridomas, electrofusion quickly became a popular subject of study in the early eighties.

In order to induce cells to fuse, the cells must first be brought in contact with each other. There are a number of methods to achieve this.

1. *By mechanical means*. One can use a specially designed container so that cells are forced to aggregate in close proximity. Or, attached cells can be cultured side by side.
2. *By chemical means*. Cells can be coated with certain chemicals that bring the cells together. For example, one cell type can be coated with biotin while the other with avidin. Or, one cell type may express an antibody against a surface protein of its fusion partner. The immunoreaction would bring the proper pair of cells together.
3. *By dielectrophoresis*. This is probably the most elegant way to bring cells together for electrofusion. When a suspended cell is exposed to an external electric field, it develops an induced dipole moment in parallel to the field. Thus, under a weak electric field, suspended cells would tend to attract each other through the induced dipole–dipole interaction. They often form a long chain of cells (commonly referred to as a *pearl chain*). This process is called *dielectrophoresis* (Fig. 3). The external field used to produce dielectrophoresis is not limited only to a DC field but it can also be an oscillating field. For technical reasons, the dielectrophoresis field used in electrofusion is a continuous oscillating field with a frequency on the order of megahertz; the field strength is usually quite low, in the range of 10 to 20 V cm^{-1}. Such a low field is sufficient to bring about dielectrophoresis while not causing electropermeabilization.

It has been demonstrated in studies using red blood cells that, in electrofusion, two kinds of fusion can occur: "membrane fusion" and "lumen fusion." In membrane fusion, only the membranes of the neighboring cells fuse, while the cytoplasm of the cells remain separate. In lumen fusion, the fusing cells truly fuse their cytoplasm to form a syncytium.

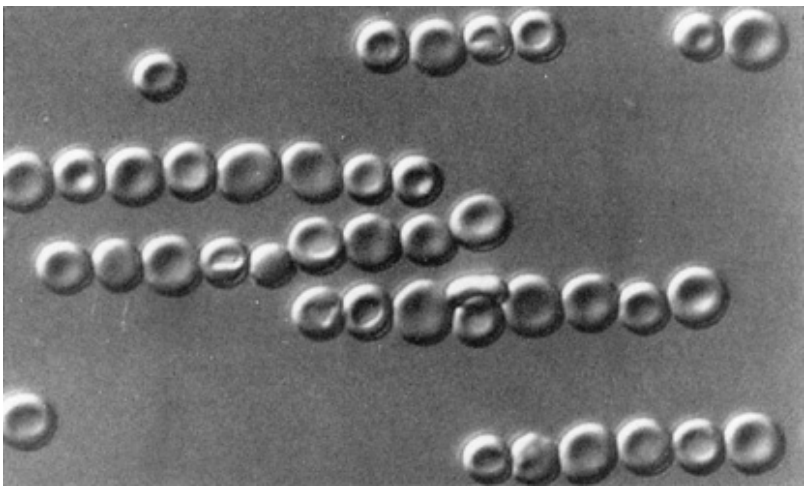

Fig. 3 Formation of pearl chains of human red blood cells, induced by applying a low-intensity AC field (200 V cm^{-1}, 60 kHz). (Reproduced from Chang, D.C., et al. ''Electroporation and Electrofusion Using a Pulsed Radio-Frequency Electric Field''. *In* Chang, D.C., Chassy, B.M., Saunders, J.A. and Sowers, A.E. (Eds.): *Guide to Electroporation and Electrofusion*. Academic Press, Inc. San Diego, CA (1992), p. 319.)

Membrane fusion can be observed using a hydrophobic dye such as DiI, which is selectively distributed in the membrane. If one mixes cells labeled with DiI with unlabeled cells, the dye will be seen to transfer from a labeled cell to an unlabeled cell when the membranes of the two cells fuse. Lumen fusion can be monitored by preloading one of the fusing partners with a fluorescent dye, such as FITC-conjugated dextran. It was discovered that, even after the cell membranes fuse, the lumens of the two cells may or may not fuse. In cell types other than red blood cells, most of the fusion products of electrofusion are of the lumen fusion type.

The time required to fuse two cells using electrofusion varies with the cell type. The electric field is usually applied for a very brief period, about a millisecond. In a few minutes after the application of the electrical pulse, fusion between neighboring cells can be seen. In red blood cells, the fusion process may be completed in about five minutes. For attached mammalian cultured cells, however, the fusion process may take many hours.

The mechanism of electrofusion is not fully understood at present. It is generally believed that the phenomenon of electrofusion is closely related to that of electroporation. In order to induce cells to fuse, the membrane structure of the two neighboring cells must be disrupted so that they can merge together. This electric field–induced disruption of the membrane structure could be very similar to that of the electropermeabilization process. In the literature, several models have been proposed to explain how the electric field may bring about cell fusion.

1. *Joining of electropores in the apposing membranes.* In the early model of electrofusion, it was suggested that when two cells are in close contact during the electroporation state, electropores on the apposing cell membranes may line

up with each other. During the membrane resealing process, the pores may merge to become a continuous intercellular channel; the apposing membranes now fuse at the rim of the merged pore. Such a membrane structure is unstable due to its sharp curvature. As a result, the merged pore expands quickly. This expansion will gradually lead to the fusion of the two cells.

2. *Electric field–induced compaction.* When cells are lined up to form a pearl chain, application of an intense electrical pulse would produce two major effects on the cell membranes. Besides breaking down the membranes, the applied electric field also induces a strong dipolar interaction between the adjacent cells. Such an interaction generates a strong attractive force between the apposing cells and forces the adjacent membranes to come together to make physical contact, which eventually leads to fusion.
3. *Weakening of the hydration force.* In this model, it is argued that there exists a strong repulsive force when two cells are brought into close proximity. This force is originated from the ordering of several water layers at the cell surface. When cells are brought into contact, their respective water layers repel each other. This repulsive force, called *hydration force*, is the major force that prevents spontaneous cell fusion. When an intense electric field is applied, it will disrupt part of the membrane structure and will partially destroy the orderly organization of the cell membrane. As a result, the ordering of the surrounding hydration layer will also be destroyed. The strength of the hydration force is greatly reduced, and the adjacent membranes of the two cells can now come into contact and subsequently fuse.

In the early studies of electrofusion, it was thought that, in order to fuse cells, cells must be first brought into contact before application of the intense electric field. Later experiments demonstrated that application of an intense electric field can induce fusogenic properties even in isolated cells. When cells are first exposed to an intense electric field and then brought together in the absence of the field, some of them can still fuse with each other. Such an electric field–induced fusogenic property was cited as supporting evidence for model #3 discussed above.

2 Techniques

2.1 Waveforms Used for Electroporation and Electrofusion

The external field used in electroporation or electrofusion is always applied in the form of a pulse. Such a pulse can have different shapes, or waveforms, which include (1) rectangular pulse; (2) exponential decay pulse; and (3) oscillating pulse (Fig. 4). In the early stage of development of electroporation and electrofusion, most of the studies used only rectangular pulses. The pulse width, or τ, was usually very short. The major requirement for the pulse width is that it must be longer than the relaxation time of the cell, which is typically on the order of microseconds. Another requirement is that the field must be applied long enough to allow the membrane to undergo an electrical breakdown. There is also a limitation on how long that τ is. In a study using EM to examine structural changes in electropermeabilized cells, it

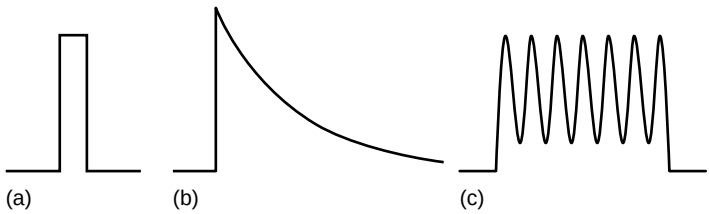

Fig. 4 Major types of waveforms used in electroporation and electrofusion: (a) rectangular pulse, (b) CD pulse generated by discharging a capacitor, and (c) DC-shifted RF pulse.

was shown that, when the pulse width was shorter than 20 μs, cells were more or less intact over a very wide range of applied field strength. If τ is much longer than 20 μs, however, the same field strength caused excessive cell damage. Thus, in most of the early works, the pulse width is usually very short, and the intensity of the applied electric field is relatively high, typically several kilovolts per centimeter for mammalian cells. To electroporate smaller cells, such as bacteria, the required field strength may be as high as 20 kV cm^{-1}.

Recently, the exponential decay pulse has become a more popular waveform for use in electroporation or electrofusion. This pulse is usually generated by discharging a large capacitor that is precharged to a high voltage, and it is thus frequently referred to as a *capacitor discharge (CD) pulse*. The advantage of using the CD pulse is that it requires only simple equipment to generate such a pulse. The disadvantage is that it is difficult to deliver multiple pulses because it takes time to repeat the process of discharging and recharging the capacitor.

Besides being simple in equipment requirements, the CD pulse actually has another important advantage over the rectangular pulse. In 1987, a team of scientists from Stanford University conducted a comparative study in which mammalian cells were transfected with reporter genes using both short CD pulses with high field strength and long CD pulses of low field strength. It was found that the long CD pulse of low field strength gives a much higher transfection efficiency. Results of this work suggest that CD pulses with a large pulse width may be more efficient than narrow rectangular pulses in transfecting mammalian cells. Since then, the CD pulse has become a more popular choice for electroporation use.

More recently, a new type of waveform that utilizes an oscillating field has been developed. It is called a *DC-shifted radio-frequency (RF) pulse*. This RF pulse is found to have several advantages. First, this waveform is more effective in electroporating (or electrofusing) cells of heterogeneous size. As indicated in Eq. (3), the induced membrane potential is proportional to the radius of the cell. Thus, with a given applied electric field, the induced membrane potential will vary depending on the cell size. This can lead to a situation in which large cells may have too high a V_m and suffer excessive damage, while small cells may have too low a V_m such that their membranes do not breakdown. This problem can be solved using the RF pulses. Under an oscillating field, V_m is given by the following equation:

$$V_m = \frac{3Er\cos\theta}{2[1+(\omega\tau_c)^2]^{1/2}} \qquad (4)$$

where ω is the angular frequency of the applied electric field; τ_m and other symbols have been previously defined. At a reasonably high frequency, V_m is almost inversely proportional to τ_c, which is in turn roughly proportional to the cell radius, r. Therefore, under an oscillating field, V_m is not very sensitive to r. As a result, both large and small cells will have a similar V_m under the same field and will both be efficiently electroporated without excessive cell damage.

Secondly, the RF pulse is more effective in inducing membrane breakdown. The oscillating field not only can cause a compression in the cell membrane, as the DC field can, but it can also drive the charged membrane lipids and proteins to undergo an oscillating motion. This motion in turn will create structural fatigue in the membrane and make it more susceptible to electrical breakdown.

Starting from 1990, we have demonstrated in a series of experiments in mammalian cells (including CV-1 cells, COS-M6 cells, and human red blood cells) that the RF waveform was able to provide a significantly higher efficiency in electroporation and electrofusion than the rectangular pulse or the CD pulse. Recently, another study was conducted to compare the RF electroporation with the conventional electroporation method. It was reported that the RF pulses were far more efficient in transfecting the *GFP* gene into a human embryonic kidney cell line.

2.2 Basic Procedures of Electroporation and Electrofusion

The equipment required for electroporating cells is relatively simple, mainly consisting of an electrical pulse generator and a cell chamber (Fig. 5). The major function of the pulse generator is to provide a high-voltage pulse with a large current output. This is necessary because the cell sample may have a very low resistance. The sample chamber not only contains the cell sample but also provides a pair of electrodes for applying the electric field. The electrodes are usually two

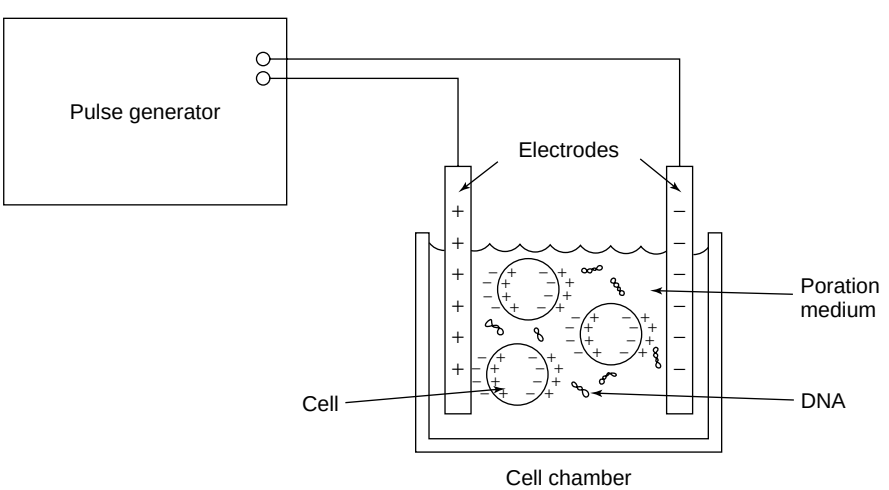

Fig. 5 A schematic diagram showing the basic device required for electroporation. Here, exogenous genes are introduced into cells by applying a pulsed electric field.

parallel metal plates made of platinum, stainless steel, or aluminum, with a typical distance of 0.1 to 0.4 cm.

The procedure of electroporation is also very simple. Suppose one wants to use electroporation to introduce certain recombinant DNA into a cultured cell line. The basic procedures are as follows:

1. When the cells are growing in the mid-log phase, harvest them using the standard cell-detach treatment.
2. Wash the cells at least once with a poration medium (PM) and resuspend them in PM. The composition of PM varies greatly between studies; it is basically an isotonic buffer with the majority of salt replaced by nonelectrolytes, such as sucrose or mannitol. DNA can be added to the cell sample at this time.
3. Load the cell sample into the chamber.
4. Apply the electrical pulse (or pulses) to the sample.
5. Transfer the cells from the sample chamber into the culture plate (or flask) and culture them in their normal conditions.
6. After the cells have been cultured for a sufficient length of time, they can be assayed for expression of the transfected gene or can be put under a selection process.

The procedures of electrofusion are similar to those of electroporation, except that cells must be placed in contact with their fusion partners before applying the intense electric field. Depending on the objective of the fusion, cells can be brought together using various methods, including mechanical, chemical, or electrical means. (See Section 1.3). The buffer used in electrofusion [called *fusion medium* (*FM*)] is also similar to the PM used in electroporation, except that FM must contain very little salt. This is particularly important when electrophoresis is used to bring cells together. The conductivity of the FM must be kept low to avoid generating excessive current in the cell sample.

2.3 *In situ* Electroporation of Attached Cultured Cells

In biomedical research, there is often a need to introduce exogenous molecules other than DNA (such as drugs, enzymes, antibodies, inhibitors, etc.) into living cells. The common method to achieve this aim is microinjection using a glass pipette. This method, however, is technically rather complicated. We have developed a unique method called *in situ electroporation*. Unlike the conventional electroporation method, which works mainly on suspended cells, this method allows us to introduce exogenous molecules into cultured cells in their attached state. This method does not require special sample preparation; cells can be grown on glass coverslips, just like those used for conventional microinjection. A special chamber was made in which an array of parallel Pt wires (with a constant gap) can be laid on top of the coverslip, so that a pulsed electric field can be passed across the cultured cells. (See Fig. 6). Using fluorescent-labeled molecular probes and reporter genes, it was demonstrated that this electroinjection method was highly efficient for a variety of cultured cells, including COS-M6, CV-1, 3T6, and C3H cells. This method is particularly useful for loading Ca^{2+} indicators (e.g. Calcium Green conjugated with dextran) for physiological studies. A fluorescent dye, such a rhodamine-dextran, can be coinjected as a marker. Recently, we have used this *in situ* electroporation method to inject cell-impermeable

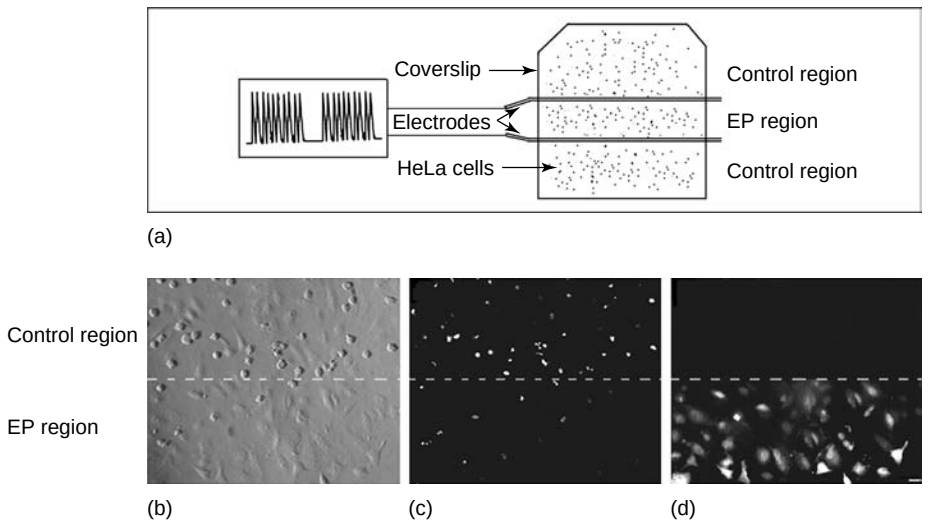

Fig. 6 (a) Schematic diagram showing the basic apparatus of *in situ* electroporation. Only cells between the two electrodes were electroporated and loaded with BAPTA. (b) DIC (differential interference contrast) image of cells at 5 h after UV treatment. Only cells undergoing apoptosis would round up. The dotted line shows the boundary between the EP and control regions. (c) Fluorescence image of nuclei (stained with Hoechst 33342) in the same field. Only apoptotic cells contained condensed chromatin, which appeared as bright spots under this magnification. (d) Fluorescence image of cells in the same field observed with a rhodamine filter, showing that only cells in the EP region were loaded with rhodamine-dextran and BAPTA. Bar, 20 μm.

calcium signal blockers, including BAPTA 1,2-bis-(*o*-aminophenoxy)-ethane-*N*,*N*,*N*′,*N*′-tetra-acetic acid and heparin, into attached HeLa cells for studying the effects of calcium signaling in UV-induced apoptosis. Other groups have also successfully used a different design of *in situ* electroporation to study cell signaling.

This *in situ* electroporation technique has several advantages. First, the experimental procedure was simple. There was no need to trypsinize the cultured cells and wait for the cells to reattach. Second, cell viability was improved owing to the abolishment of the trypsinizing procedure. Third, the gene transfer efficiency was found to be higher using the *in situ* electroporation method in comparison with the conventional electroporation method.

2.4 Factors Affecting the Efficiency of Electroporation and Electrofusion

There are many factors that can affect the efficiency of electroporation. The major ones include the following:

1. *Field strength*. This is probably the most important factor that can affect the efficiency of electroporation and cell viability. In order to electroporate cells, the induced membrane potential (V_m) must be large enough to approach the threshold potential, V_{th}. As indicated in Eq. (3), V_m is directly proportional to the applied field. If the applied electric field is too low, V_m will be much smaller than the threshold potential, and electrical breakdown of the cell

membrane will not occur. Under this situation, cells will not be properly electroporated or will fail to fuse. On the other hand, if the applied field strength is too high, V_m will be much greater than the threshold potential and will lead to irreversible membrane breakdown. The membrane will fail to reseal and the cells will no longer be viable. Thus, selecting a proper field strength is a critical step in both electroporation and electrofusion experiments.

As can be seen in Eq. (3), to generate a fixed V_m, the field strength required is inversely proportional to the radius of the cell. Thus, the field strength used to porate small cells, such as the *Escherichia coli* bacteria, is very high, typically on the order of 10 kV cm^{-1}. For larger animal cells, such as mammalian cultured cells, the applied field strength is typically about 1 kV cm^{-1}.

2. *Pulse width*. The second most important parameter is the pulse width, τ. When τ is too short, there will not be enough time for the membrane to develop an electrical breakdown. Too long a τ, however, will cause excessive damage to the membrane and the cell will not survive. When the rectangular pulse is used, the typical pulse width is less than 1/10th of a millisecond. For CD pulse or RF pulse, the pulse width may be significantly longer, on the order of milliseconds.

   The parameters of field strength and pulse width are found to complement each other. For a given cell type, the optimal value of E will shift upward when the pulse width is reduced.
3. *Number of pulses*. For some cells, a single pulse is sufficient for effective induction of electroporation or electrofusion. In most cell types, however, application of multiple pulses can greatly improve the efficiency of gene transfer.
4. *Growth condition of the cells*. To achieve high efficiency of gene transfer, cells must be harvested during the active growth phase prior to electroporation. If the cultured cells are grown for too long, the transfection efficiency will usually be low. It is recommended to replate the cells one or two days prior to electroporation.
5. *Other factors*. There are other relevant and influential factors, including temperature and the presence of divalent ions in the cell culture medium. For example, the concentration of Mg^{2+} may be critical to the efficiency of electroporation in some cell types. In some studies, it is recommended to add carrier DNA to enhance the efficiency of gene uptake during the electroporation process.

Most of the factors affecting the efficiency of electroporation can also affect the efficiency of electrofusion. There are also other factors that are uniquely important to electrofusion. For example, the alignment of cells can greatly influence fusion efficiency. To fuse cells, one must first bring the cells together before applying the high-intensity pulse. As discussed in Section 1.3, there are different methods of bringing cells together. Among them, the most elegant one is to use a low-amplitude oscillating field, which can induce cells to form a pearl chain by the process of dielectrophoresis. The optimal frequency and amplitude of the applied dielectrophoretic field for pearl chain formation depends on the diameter and the surface-charge properties of each individual cell. Thus, in order to maximize the efficiency of electrofusion, one needs to optimize the pearl chain formation process.

It has been suggested in some studies that certain chemical treatments of the cell surface can considerably increase electrofusion efficiency. For example, it was reported that pretreating red blood cells with pronase can greatly enhance the efficiency of electrofusion. Such an enhancement may be the result of removing some of the glycoproteins at the cell surface and thus allowing the neighboring cells to be closer to each other.

Another factor that can affect fusion efficiency is the osmolarity of the fusion medium. The cell membrane can be expanded using a partial osmotic shock to increase cell volume. Recently, this technique has been applied in the generation of human hybridomas. In this work, cells were treated with a hypoosmotic medium before the intense electric field was applied. It was reported that, with this treatment, the efficiency of electrofusion was greatly enhanced.

In our experience, the waveform of the electric pulse can also significantly affect fusion efficiency. We found that, by using a train of RF pulses, the electrofusion efficiency can be greatly increased as compared to protocols using rectangular pulses.

3 Applications

3.1 Applications of Electroporation in Gene Transfer

The most important application of electroporation is to introduce recombinant DNA into biological cells. Electroporation has been used for introducing genes into many different cell types, both eukaryotic and prokaryotic, including animal cells, plant cells, unicellular organisms such as algae or yeast, parasites, and bacteria. In fact, electroporation is now the method of choice for introducing genes into *E. coli*; the resulting transfection efficiency can be as high as 10^9 colonies per μg of DNA, which is almost thousandfold higher than that normally obtained by conventional methods.

Electroporation can be used for both transient and stable gene transfection in mammalian cells. In 1982, it was first demonstrated by T.K. Wong and E. Neumann that, using electrical pulses, a plasmid DNA containing the simplex virus thymidine kinase (*TK*) gene was successfully introduced into mouse L cells deficient of the *TK* gene. Stable transformants were obtained in this experiment. Later, electroporation was applied to transfect myeloma cells and a neuronal cell line. Within a few years, electroporation was used to introduce exogenous DNA into a variety of cells, including lymphocytes, neuronal cells, fibroblasts, endocrine cells, primary animal cultures, hepatoma cells, hematopoietic stem cells, and so on.

Today, electroporation is widely used in many studies of molecular biology for introducing genes into all sorts of cell types and for a wide variety of purposes. One common application of electroporation-mediated gene transfer is the study of gene regulation. Usually, a reporter gene is inserted into a plasmid vector with specific transcription promoter and enhancer sequences. By comparing the efficiency of gene expression under various conditions, such as the availability of certain transcriptional factors, one can gain knowledge about how the gene of interest is regulated. Alternatively, one may mutate the upstream sequences of a certain gene, fuse it with the reporter gene, and then introduce it into cells by electroporation.

By comparing the resulting level of gene expression under various mutation conditions, one can identify the major promoter or enhancer sequences.

In addition, functional genes can be introduced into mammalian cells by stable transfection using electroporation. This can be done by coinserting a drug-resistance gene with the gene of interest into the cell genome and then subjecting the cell to a selection media containing the specific drug (such as G418). Using this method, a large number of different types of DNA have been transfected into a variety of mammalian cells. For example, the complete human immunoglobulin κ (*Ig*κ) gene was stably transfected into mouse L cells and mouse pre-B-cells by electroporation. Northern blot analysis of RNA indicated that the expression of the *Ig*κ gene is highly tissue-specific.

Not only can electroporation be used for introducing genes into cells to produce stably transfected cell lines but also a similar method can be used to introduce genes directly into eggs or stem cells for the generation of transgenic animals. For example, it has been reported that transgenic zebrafish can be produced using electroporation. Electroporation has also been used to introduce genes into plant cells. For example, using electroporation, genes have been successfully introduced into protoplasts of a number of plants, including carrot, maize, and tobacco leaf mesophyll cells. In addition, this technique can be applied to make transgenic plants. It has been reported that electroporation can be used for the generation of transgenic cereal plants such as rice and sorghum.

In comparison with other methods of gene transfer, such as those using calcium phosphate or retroviruses, electroporation has several distinctive advantages. First, it is a simple method. The procedure takes only minutes, and there is no need for prior treatment of the DNA to be inserted. Secondly, electroporation is efficient. In comparison with the chemical methods using calcium phosphate or DEAE dextran, electroporation usually offers much higher transfection efficiency. Thirdly, since electroporation is a physical method, it is widely applicable to many cell types. Unlike the chemical method or the virus vector, electroporation is less dependent on the surface properties of the cell membrane, and hence, electroporation can be used to introduce genes into a very wide spectrum of different cell types. Finally, electroporation is a relatively "clean" method because it involves little chemical or biological hazard compared with the chemical and retroviral methods.

3.2 Electroinjection of Other Molecules

Besides genes, electroporation can be applied to introduce many other types of molecules into living cells. In fact, the early usage of electroporation was to introduce ions, such as calcium, and small molecules, such as ethyleneglycol-bis-(β-aminoethyl)-*N*,*N*,*N*′,*N*′-tetra-acetic acid (EGTA) and ATP, into cells. Lately, electroporation has also been used to introduce fluorescent-labeled dextran and molecular probes such as Calcium Green-1. Electroporation can also be used to introduce large molecules such as antibodies into cells. For example, using electroporation, monoclonal antibodies have been introduced into HeLa cells and murine lymphoma cells. Furthermore, electroporation has been used for the introduction of a number of active biological compounds, including restriction enzymes, Epstein–Barr virus episomal replicons, substrates of enzymes,

and drugs that block certain signal transduction pathways.

One particular interesting example of using the electroporation technique for studying cell signaling is our recent investigation of the involvement of Ca^{2+} signals in UV-induced programmed cell death. In the last few years, several groups had reported that application of a membrane-permeant form of a Ca^{2+} chelator (BAPTA/AM) failed to prevent cells from entering apoptosis. These findings raised a serious question on whether the Ca^{2+} signal is truly involved in regulating the progression of programmed cell death. To resolve this question, we examined the differential effects of three different Ca^{2+} signal blockers (membrane-permeant BAPTA/AM, membrane-impermeant BAPTA, and heparin) on UV-induced apoptosis in HeLa cells. We found that although BAPTA/AM could not inhibit cell death, the membrane-impermeant form of BAPTA, loaded into the cytosol by electroporation, clearly protected cells from entering apoptosis (Fig. 6). Furthermore, when we injected heparin to block Ca^{2+} release from the endoplasmic reticulum (ER) to the cytosol, apoptosis was greatly suppressed. These findings strongly suggest that elevation of cytosolic Ca^{2+} is part of the signal that drives the progression of apoptosis. The negative result of BAPTA/AM is probably due to its dual effect on subcellular Ca^{2+} distribution; besides suppressing the Ca^{2+} elevation in cytosol, BAPTA/AM can also enter into the ER to reduce the free Ca^{2+} level there. The depletion of Ca^{2+} in ER is believed to stimulate apoptosis and thus would counterbalance the protective effect of BAPTA/AM in suppressing the cytosolic Ca^{2+} elevation.

3.3 Study of the Mechanism of Membrane and Cell Fusion

Electrofusion has been used for studying the mechanism of membrane fusion. Most of these works were done using either a model membrane such as lipid vesicles or cells with simplified membrane structure, such as red blood cells or red cell ghosts. More recently, electrofusion has been used for the study of cytoplasmic fusion. Cytoplasmic fusion is more difficult to study than membrane fusion because the process involved takes a much longer time. The fusing cells must be maintained in their normal physiological state for a long period for observation. Such a requirement excludes some of the common fusogens, like polyethylene glycol (PEG) or a virus, that may have toxic effects. Furthermore, in order to study the dynamic changes in cytoplasmic structures, cell fusion must be induced by a treatment that requires very little time, and the fusion process must be relatively synchronized.

These requirements can be satisfied by using the electrofusion technique. The fusion event can be induced by applying a short pulse of electric field; the fusion efficiency is very high, about 70%. The fusion of all cells exposed to the same electrical pulse would occur almost in synchrony. Because of these advantages, the electrofusion method has been used to study the reorganization of various cytoplasmic components, including cytoskeletons and organelles, during the fusion of mammalian cells.

3.4 Production of Hybridomas and Antibodies

An important application of electrofusion is to make hybridomas for the production

of monoclonal antibodies. For example, mouse myeloma cells can be fused with mouse lymphocytes and then selected for clones producing a specific antibody. Traditional methods using a chemical fusogen, such as PEG, are not as efficient as electrofusion. Several comparative studies have shown that electric field–induced cell fusion can be 1000-fold more efficient than fusion events initiated by chemical means. Lately, there has been a strong interest in making human hybridomas for producing human antibodies, which would obviously be much more useful than animal antibodies for medical purposes. The production of human hybridomas is particularly difficult because, with the conventional fusion methods, the fusion between antibody-producing cells and human plasma cells usually has very low yield. Recently, by using the electrofusion technique, several groups were able to develop new protocols to produce human hybridomas successfully.

4 Perspectives

4.1 Future Applications of Electroporation in Molecular Biology

One of the most important applications of electroporation as a gene transfer technique in molecular biology is gene targeting. At this time, it is difficult to use other methods of gene transfer to conduct homologous DNA recombination. For example, in the retroviral method, the inserted gene is flanked by long terminal repeat (LTR) sequences, which have a tendency to insert the recombinant gene into random sites on the genome. This will defeat the purpose of homologous recombination. Chemical methods, such as those using calcium phosphate, have the disadvantage of inserting multiple copies of the recombinant gene into the genome; the number of copies is uncontrollable. These problems can be avoided by using electroporation as the gene transfer method. With this method, one can control the copy number of recombinant DNA to be inserted into the genome. Also, by linking with proper flanking sequences, the gene introduced by electroporation can be directed to integrate into the proper site of the genome. Hence, electroporation has become a very important tool for gene targeting work.

Another advantage of the electroporation method is that it is less cell-type dependent. Both the chemical methods and the retrovirus method depend on the surface properties of the target cells, and thus, certain cell types are resistant to transfection mediated by chemicals or viruses. It happens that some of the cell types that are highly important for gene therapy, such as lymphocytes and stem cells, are generally resistant to the common retrovirus vectors. Since electroporation is a physical method, it is not particularly sensitive to the composition of the membrane proteins, and, thus, is less cell-type dependent. So, in principle, once the electroporation method is properly optimized, it can be applied for introducing genes into most target cells.

4.2 Electroporation as an Important Tool for Studying Cell Signaling

At present, electroporation has been mainly used in molecular biology. With time, this technique will have the potential to become an important tool in the study of cell biology. Electroporation can be used as

a massive microinjection technique. Currently, microinjection is done by means of a glass pipette, which is tedious, requiring expensive equipment and extensive training. It is also difficult to inject a large number of cells at one time. With electroporation, one can inject thousands or even millions of cells, within a few minutes. Thus, electroporation can become a very efficient injection method.

One of the major uses of electroporation is to introduce molecular probes, such as ion indicators (e.g. for Ca^{2+} or H^{+}) or cell-labeling dyes. At present, most of the calcium indicators, such as fura-2 or fluo-3, are introduced into the cell either by direct injection or by modifying the indicators into triacetoxymethyl ester (AM) forms that can permeate to the cell membrane. However, by the latter methods, the amount of molecules introduced is difficult to control. Certain portions of the AM-form indicator could also be taken up by the intracellular organelles, thus making it more difficult to interpret the experimental results. These problems can be avoided by using electroporation to introduce non-AM-form indicators. Electroporation has been demonstrated to be efficient in injecting several different molecular probes into a wide variety of cells, including animal, plant, and unicellular cells. The amount of molecules injected usually can be controlled by varying the electroporation conditions.

In the future, the most potentially useful application of electroporation is probably the introduction of various substances into cells to study the processes of signal transduction. For example, one can introduce kinases or phosphatases, their substrates, activators, or inhibitors into cells to study a hypothetical signal transduction pathway. It is well known that many of the cell functions are regulated by the phosphorylation and dephosphorylation of certain proteins. The signal pathway usually involves multiple components and a cascade of kinase/phosphatase activities. It is very difficult to study such a pathway. With the electroporation technique, one can introduce reagents into the living cell to specifically inhibit or activate a selected step of the signaling process and then observe the resultant cellular responses. In this way, one can examine the functional role of a specific signaling molecule and can eventually gain a better overall understanding of the entire signal transduction pathway under study.

Similarly, one can use electroporation to introduce many other factors that are important for biological regulations, such as transcription factors, gene products of oncogene or tumor suppressor genes, small G proteins, intracellular receptors of hormones or other signaling molecules, and so on. The ability to introduce these factors into a large number of living cells will open up new opportunities for examining the molecular mechanism of many important cellular functions.

4.3 Use of Electroporation for Gene Therapy

The electroporation method can be applied in many areas of biomedical science. One important use of this method is for gene therapy. Since the target cells in this case are normally obtained from the patient, the quantity of cells available for gene insertion is usually limited. Also, no side effects can be tolerated in such a transfer procedure. At present, most gene therapy studies use virus vectors to insert the recombinant gene. This method creates a high risk for patients because the virus could be reactivated

under certain circumstances (such as infecting by a wild-type virus of similar strain). Being a physical method, the electroporation technique does not have such a risk. By improving the efficiency of gene transfer, electroporation can become the method of choice for gene therapy. It is particularly useful in the explant approach. For example, electroporation can be used to introduce therapeutic genes into bone marrow cells isolated from a patient. After the introduced genes are expressed successfully, cells can be retransplanted into the same patient.

In recent years, a new use of electroporation for medical purpose has been developed; it is the *in vivo* application of electrical pulses for the delivery of drugs or DNA into cells or tissues. *In vivo* electropermeabilization can be achieved by using either invasive electrodes (e.g. needles) or surface electrodes (metal plates). Such electrodes could be used to permeabilize cells in the skin or the skeletal muscle. One important application of this *in vivo* electroporation method is for electrochemotherapy, in which a cell nonpermeant antitumor drug (such as bleomycin) can be introduced directly into the patient's cells in a local region. In the future, besides chemotherapy drugs, one can also use this method to introduce active oligonucleotides directed to specific genes of the patient. A second important application of *in vivo* electroporation is to introduce therapeutic genes directly into the patient's cells. Using electrical pulses of different waveforms, several groups had reported various degrees of success in transferring plasmids into skin cells, liver tissue, skeletal muscle, and tumors during the last several years. Although there are still many technical problems to be overcome, the *in vivo* electroporation technique has already been recognized as an important future tool for gene therapy.

See also Genetic Engineering of Vaccines; Medicinal Chemistry.

Bibliography

Books and Reviews

Chang, D.C., Chassy, B.M., Saunders, J.A., Sowers, A.E. (1992) *Guide to Electroporation and Electrofusion*, Academic Press, San Diego, CA.

Neumann, E., Sowers, A.E. (1989) *Electroporation and Electrofusion in Cell Biology*, Plenum Publishing, New York, pp. 215–227.

Potter, H. (1988) Electroporation in biology: methods, applications and instrumentation, *Anal. Biochem.* **174**, 361–373.

Sowers, A.E. (1987) *Cell Fusion*, Plenum Publishing, New York, pp. 479–496.

Teissie, J., Eynard, N., Gabriel, B., Rols, M.P. (1999) Electropermeabilization of cell membranes, *Adv. Drug Delivery Rev.* **35**, 3–19.

Weaver, J.C. (1993) Electroporation: a general phenomenon for manipulating cells and tissues, *J. Cell. Biochem.* **51**, 426–435.

Zimmermann, U. (1986) Electrical breakdown, electropermeabilization and electrofusion, *Rev. Physiol. Biochem. Pharmacol.* **105**, 176–256.

Primary Literature

Allegretti, J.P., Panje, W.R. (2001) Electroporation therapy for head and neck cancer including carotid artery involvement, *Laryngoscope* **111**, 52–56.

Baron, S., Poast, J., Rizzo, D., McFarland, E., Kieff, E. (2000) Electroporation of antibodies, DNA, and other macromolecules into cells: a highly efficient method, *J. Immunol. Methods* **242**, 115–126.

Boitano, S., Dirksen, E.R., Sanderson, M.J. (1992) Intercellular propagation of calcium waves mediated by inositol trisphosphate, *Science* **258**, 292–295.

Cemazar, M., Sersa, G., Wilson, J., Tozer, G.M., Hart, S.L., Grosel, A., Dachs, G.U. (2002) Effective gene transfer to solid tumors using

different nonviral gene delivery techniques: electroporation, liposomes, and integrin-targeted vector, *Cancer Gene Ther.* **9**, 399–406.

Chang, D.C. (1989) Cell poration and cell fusion using an oscillating electric field, *Biophys. J.* **56**, 641–652.

Chang, D.C., Gao, P.Q., Maxwell, B.L. (1991) High efficiency gene transfection by electroporation using a radio-frequency electric field, *Biochim. Biophys. Acta* **1092**, 153–160.

Chang, D.C., Reese, T.S. (1990) Changes in membrane structure induced by electroporation as revealed by rapid-freezing electron microscopy, *Biophys. J.* **58**, 1–12.

Chu, G., Hayakawa, H., Berg, P. (1987) Electroporation for the efficient transfection of mammalian cells with DNA, *Nucleic Acids Res.* **15**, 1311–1326.

Dezawa, M., Takano, M., Negishi, H., Mo, X., Oshitari, T., Sawada, H. (2002) Gene transfer into retinal ganglion cells by in vivo electroporation: a new approach, *Micron* **33**, 1–6.

Dimitrov, D.S., Sowers, A.E. (1990) Membrane electroporation – fast molecular exchange by electroosmosis, *Biochim. Biophys. Acta* **1022**, 381–392.

Gabriel, B., Teissie, J. (1999) Time courses of mammalian cell electropermeabilization observed by millisecond imaging of membrane property changes during the pulse, *Biophys. J.* **76**, 2158–2165.

Golzio, M., Teissie, J., Rols, M.P. (2002) Direct visualization at the single-cell level of electrically mediated gene delivery, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1292–1297.

Heller, R., Schultz, J., Lucas, M.L., Jaroszeski, M.J., Heller, L.C., Gilbert, R.A., Moelling, K., Nicolau, C. (2001) Intradermal delivery of interleukin-12 plasmid DNA by in vivo electroporation, *DNA Cell. Biol.* **20**, 21–26.

Hibino, M., Shigemori, M., Itoh, H., Nagayama, K., Kinosita, K. Jr. (1991) Membrane conductance of an electroporated cell analyzed by submicrosecond imaging of transmembrane potential, *Biophys. J.* **59**, 209–220.

Holzapfel, C., Vienken, J., Zimmermann, U. (1982) Rotation of cells in an alternating electric field: theory and experimental proof, *J. Membr. Biol.* **67**, 13–26.

Hu, Q., Liang, W., Bao, J., Ping, Q. (2000) Enhanced transdermal delivery of tetracaine by electroporation, *Int. J. Pharm.* **202**, 121–124.

Kalat, M., Kupcu, Z., Schuller, S., Zalusky, D., Zehetner, M., Paster, W., Schweighoffer, T. (2002) In vivo plasmid electroporation induces tumor antigen-specific CD8+ T-cell responses and delays tumor growth in a syngeneic mouse melanoma model, *Cancer Res.* **62**, 5489–5494.

Kinosita, K. Jr., Tsong, T.Y. (1977) Formation and resealing of pores of controlled sizes in human erythrocyte membrane, *Nature* **268**, 438–441.

Klenchin, V.A., Sukharev, S.I., Serov, S.M., Chernomordik, L.V., Chizmadzhev, Y.A. (1991) Electrically induced DNA uptake by cells is a fast process involving DNA electrophoresis, *Biophys. J.* **60**, 804–811.

Li, S., Benninger, M. (2002) Applications of muscle electroporation gene therapy, *Curr. Gene Ther.* **2**, 101–105.

Li, L.H., Sen, A., Murphy, S.P., Jahreis, G.P., Fuji, H., Hui, S.W. (1999) Apoptosis induced by DNA uptake limits transfection efficiency, *Exp. Cell Res.* **253**, 541–550.

Onuma, E.K., Hui, S.W. (1988) Electric field-directed cell shape changes, displacement, and cytoskeletal reorganization are calcium dependent, *J. Cell. Biol.* **106**, 2067–2075.

Potter, H., Weir, L., Leder, P. (1984) Enhancer-dependent expression of human kappa immunoglobulin genes introduced into mouse pre-B lymphocytes by electroporation, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7161–7165.

Pu, Y., Chang, D.C. (2001) Cytosolic Ca(2+) signal is involved in regulating UV-induced apoptosis in hela cells, *Biochem. Biophys. Res. Commun.* **282**, 84–89.

Raptis, L., Brownell, H.L., Vultur, A.M., Ross, G.M., Tremblay, E., Elliott, B.E. (2000) Specific inhibition of growth factor-stimulated extracellular signal- regulated kinase 1 and 2 activation in intact cells by electroporation of a growth factor receptor-binding protein 2-Src homology 2 binding peptide, *Cell Growth Differ.* **11**, 293–303.

Rols, M.P., Femenia, P., Teissie, J. (1995) Long-lived macropinocytosis takes place in electropermeabilized mammalian cells, *Biochem. Biophys. Res. Commun.* **208**, 26–35.

Rols, M.P., Teissie, J. (1998) Electropermeabilization of mammalian cells to macromolecules: control by pulse duration, *Biophys. J.* **75**, 1415–1423.

Sowers, A.E. (1984) Characterization of electric field-induced fusion in erythrocyte ghost membranes, *J. Cell Biol.* **99**, 1989–1996.

Sowers, A.E. (1986) A long-lived fusogenic state is induced in erythrocyte ghosts by electric pulses, *J. Cell Biol.* **102**, 1358–1362.

Steinhardt, R.A., Bi, G., Alderton, J.M. (1994) Cell membrane resealing by a vesicular mechanism similar to neurotransmitter release, *Science* **263**, 390–393.

Stenger, D.A., Hui, S.W. (1986) Kinetics of ultrastructural changes during electrically induced fusion of human erythrocytes, *J. Membr. Biol.* **93**, 43–53.

Sukharev, S.I., Klenchin, V.A., Serov, S.M., Chernomordik, L.V., Chizmadzhev, Y.A. (1992) Electroporation and electrophoretic DNA transfer into cells. The effect of DNA interaction with electropores, *Biophys. J.* **63**, 1320–1327.

Swezey, R.R., Epel, D. (1988) Enzyme stimulation upon fertilization is revealed in electrically permeabilized sea urchin eggs, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 812–816.

Tsong, T.Y. (1991) Electroporation of cell membranes, *Biophys. J.* **60**, 297–306.

Wolf, H., Rols, M.P., Boldt, E., Neumann, E., Teissie, J. (1994) Control by pulse parameters of electric field-mediated gene transfer in mammalian cells, *Biophys. J.* **66**, 524–531.

Wong, T.K., Neumann, E. (1982) Electric field mediated gene transfer, *Biochem. Biophys. Res. Commun.* **107**, 584–587.

Xie, T.D., Sun, L., Tsong, T.Y. (1990) Study of mechanisms of electric field-induced DNA transfection I. DNA entry by surface binding and diffusion through membrane pores, *Biophys. J.* **58**, 13–19.

Zald, P.B., Cotter, M.A. II, Robertson, E.S. (2000) Improved transfection efficiency of 293 cells by radio frequency electroporation, *Biotechniques* **28**, 418–420.

Zhang, L., Nolan, E., Kreitschitz, S., Rabussay, D.P. (2002) Enhanced delivery of naked DNA to the skin by non-invasive in vivo electroporation, *Biochim. Biophys. Acta* **1572**, 1–9.

Zheng, Q.A., Chang, D.C. (1991) High-efficiency gene transfection by in situ electroporation of cultured cells, *Biochim. Biophys. Acta* **1088**, 104–110.

Zimmermann, U., Vienken, J. (1982) Electric field-induced cell-to-cell fusion, *J. Membr. Biol.* **67**, 165–182.

Embryology: *see* Developmental Cell Biology

Endocrinology, Molecular

Franklyn F. Bolander
University of South Carolina, Columbia, USA

1 Introduction 161

2 Hormones 161

3 Receptors 161
3.1 Nuclear Receptors 161
3.2 Membrane Receptors 162

4 Second Messengers 165
4.1 G-proteins and Cyclic Nucleotides 165
4.2 Calcium, Calmodulin, and Phospholipids 168
4.3 Miscellaneous Mediators 170

5 Biological Effects 170
5.1 Nongenomic 170
5.2 Genomic 171

6 Clinical Applications 174
6.1 Endocrine Deficiencies 174
6.1.1 Hormones 174
6.1.2 Receptors 175
6.1.3 Transduction 175
6.2 Endocrine Excesses 175
6.3 Oncogenesis 177
6.3.1 Oncogenes 177
6.3.2 Tumor Resistance to Therapy 178

7 Perspectives 178

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Bibliography 179
Books and Reviews 179
Primary Literature 179

Keywords

Adaptor
A small molecule consisting only of binding domains (e.g. SH2 domains, which bind phosphotyrosine, and SH3 domains, which bind polyproline helices) and used to associate proteins.

G-protein
A GTP-binding protein with intrinsic GTPase activity; it acts as a molecular switch with a built-in timer; it is "on" when bound to GTP and "off" when bound to GDP.

Hormone
A chemical, nonnutrient, intercellular messenger that is effective at very low concentrations.

Hormone Response Element (HRE)
A DNA sequence recognized by a transcription factor that is primarily regulated by hormones.

Oncogene
A gene that usually encodes a component of the growth factor pathway and that has undergone a mutation resulting in constitutive activity and tumor formation.

Endocrinology is the study of hormones, chemicals that cells use to communicate with each other. Traditionally, endocrinology was studied at the organismal level: the effects of hormones on tissue or animal growth, reproduction, or metabolism. Molecular endocrinology seeks to determine the molecular mechanisms for these gross effects: the actions of hormones, direct or through mediators, on enzymes, transport processes, the cytoskeleton, and transcription factors. Such information is vital to understanding how complex metazoans coordinate cellular functions. In addition, this knowledge can be useful in the evaluation and treatment of various endocrine diseases that have genetic bases. Furthermore, many tumors arise when the genes for various growth-promoting hormones or for components of their signaling pathway undergo mutations that render them constitutively active.

1 Introduction

Multicellular organisms have two major coordinating systems: the nervous and the endocrine systems. Both utilize chemical messengers: in the endocrine system, these molecules are called *hormones* and are usually secreted into the circulatory system for general distribution. In the nervous system, the molecules are called *neurotransmitters* and are released into synapses for more precise effects. The distinction is not always this clear; many of the chemicals used by each system and their mechanisms of action are identical. In addition, some hormones are made and act locally in a very defined area, while some neurotransmitters can leak out into the general circulation. Thus, there is a general tendency to consider all chemical messengers as a single functional group.

2 Hormones

Structurally, hormones are extremely diverse; nearly every organic group is represented. Proteins and peptides are the largest group. In addition, there are hormones that are derivatives of amino acids, sterols, fatty acids, phospholipids, nucleotides, and carbohydrates. There are even several gaseous hormones, such as ethylene and nitric oxide.

However, one property divides all hormones into two major groups and will determine their mechanisms of action: water solubility. Hydrophobic ligands have no problem crossing the plasma membrane; as such, these hormones have a direct mechanism of action. In particular, they migrate to the nucleus where they interact with transcription factors. Since their mechanism is primarily genomic, their effects tend to be delayed and long term; they are often involved in developmental processes and long-term adaptation. Hydrophilic hormones cannot cross the plasma membrane and must interact with binding proteins, the receptors. Since these receptors are integral membrane proteins, they must generate another signal on the cytosolic side; if the hormone is the primary messenger, then this subsequent factor becomes a second messenger. These latter messengers often activate kinases that can have acute effects on metabolism and cell structure or phosphorylate transcription factors for more long-term effects.

3 Receptors

3.1 Nuclear Receptors

As noted in Section 2, the receptors for lipophilic hormones are ligand-regulated transcription factors. These receptors are all homologous: the amino terminus possesses a transcription activation domain (TAD), and the center has a DNA binding region. The carboxy terminus binds the hormone, heat shock proteins (hsps), and immunophilins; it also contains a dimerization domain and a second TAD. The hsps are necessary to maintain the receptors in a conformation required for ligand binding, and the immunophilins may be involved in nuclear transport; once the hormone binds, these proteins dissociate (Fig. 1). Ligand binding also induces receptor dimerization, phosphorylation, nuclear translocation (although some receptors are constitutively located in the nucleus), DNA binding, and transcription activation. After

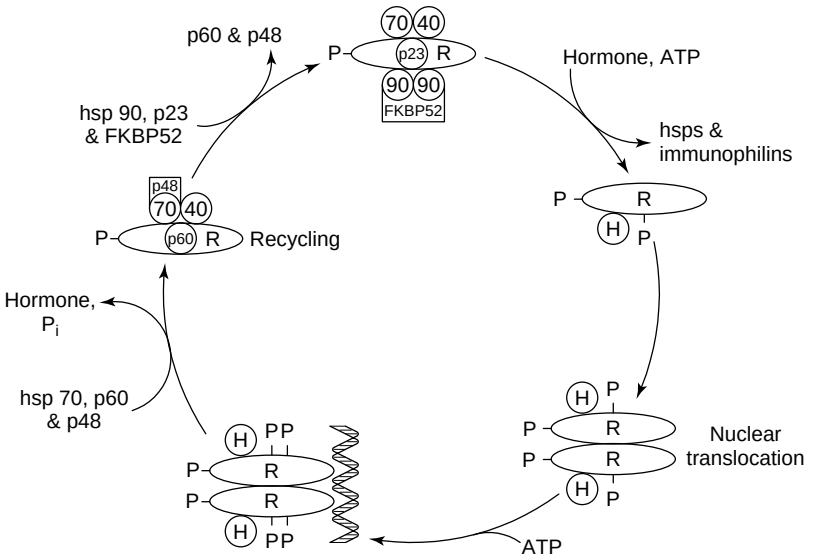

Fig. 1 Activation and recycling of a nuclear receptor: 40, 70, 90, p23, p48, p60, heat shock proteins; FKBP52, an immunophilin; H, hormone; P, phosphorylation site; R, receptor.

the hormone dissociates, the receptor is dephosphorylated and recycled.

The nuclear receptors can be divided into three families on the basis of their structures and the DNA sequences to which they bind. The glucocorticoid family is the most recently evolved group and contains the cortisol, aldosterone, androgen, and progesterone receptors. The members of this family are basically homodimers, require hsp 90, and bind inverted repeats of the hormone response element (HRE) TGTTCT. The thyroid hormone family is the oldest and most diverse group and includes receptors for the thyroid hormone, vitamins A and D, ecdysone, and arachidonic acid. They are most active as heterodimers, do not require hsp 90, and can bind either direct or inverted repeats of TGACC. The estrogen family contains only the estrogen receptor and a few related receptors whose ligands are still unidentified. Its properties lie between the two other groups: it binds the thyroid hormone HRE, but only as inverted repeats; in addition, it forms homodimers and requires hsp 90 like the glucocorticoid family.

3.2 Membrane Receptors

Membrane receptors are considerably more diverse. Four basic superfamilies are recognized: the enzyme-linked receptors, cytokine receptors, G-protein coupled receptors, and ion channel receptors (Fig. 2). The enzyme-linked are the simplest: the basic structure consists of a single protein that traverses the plasmalemma once via an α-helix. The amino-terminal extracellular domain binds the hormone, while the carboxy-terminal cytosolic domain possesses a catalytic site. Ligand binding induces oligomerization and enzyme activation. Three types of enzymatic activity have been identified:

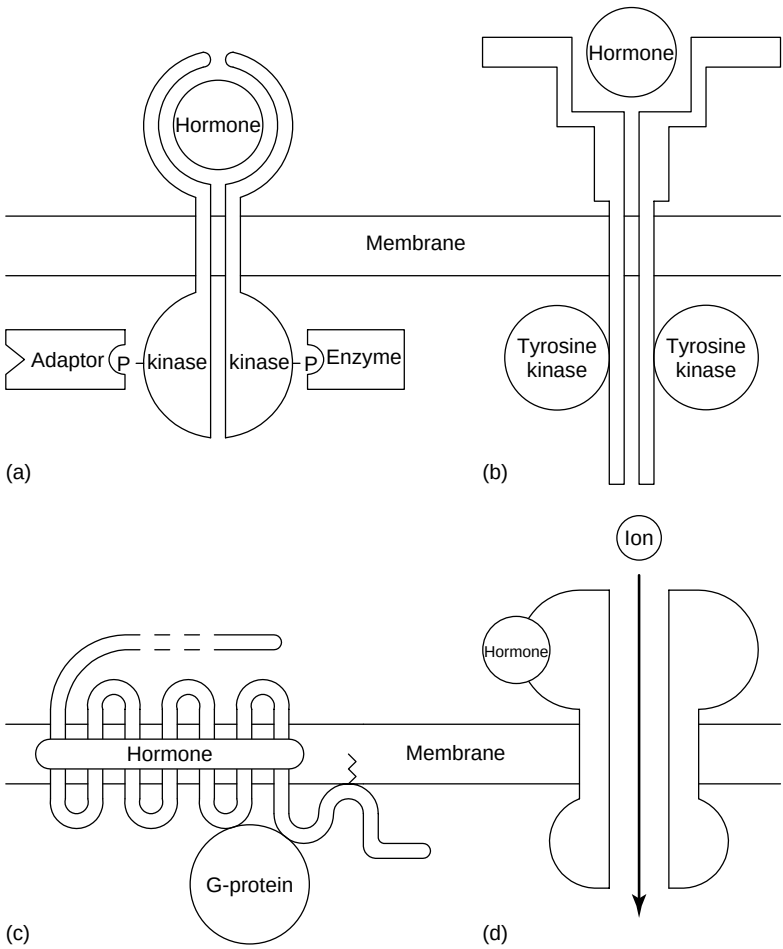

Fig. 2 Schematic representation of several classes of membrane receptors: (a) receptor tyrosine kinase, (b) cytokine receptor, (c) G-protein coupled receptor, and (d) ligand-gated ion channel.

first, the tyrosine kinases are represented by six families (the epidermal growth factor, insulin, platelet-derived growth factor, nerve growth factor, ephrin, and collagen groups); the serine–threonine kinases are found only in the transforming growth factor β family; and the guanylate cyclases generate cyclic GMP (cGMP) in response to atrial natriuretic peptides.

In the case of the receptor tyrosine kinases (RTKs), they serve as the major substrates themselves. Phosphorylated tyrosines are recognized by SH2 domains that are found in many enzymes and adaptors; these latter proteins will bind the autophosphorylated receptors and then mediate many of the biological activities of RTKs (see Section 4).

The cytokine receptors have the same structure as the enzyme-linked binding proteins except that there is no recognizable catalytic site in the cytosolic domain.

The extracellular region is composed of two modified units first identified in fibronectin, a matrix protein. In the class 1 receptors, these units form two seven-stranded β sheets that join at right angles to create a ligand-binding pocket (Fig. 2b). This core may be further embellished by immunoglobulin loops and/or additional fibronectin domains. The class 2 receptors form repeats of five-stranded β sheets that extend over the hormone like fingers.

In the cytosolic juxtamembrane region of the cytokine receptors is a conserved proline-rich region that constitutively binds soluble tyrosine kinases. Like RTKs, these kinases are activated by receptor aggregation induced by ligand binding. As such, the cytokine receptors can be considered RTKs whose catalytic site is located on a separate subunit.

The next superfamily has many names: G-protein coupled receptors (GPCR), seven transmembrane segment receptors, heptahelical receptors, and serpentine receptors. This chapter will use the first term because of its wide usage. The GPCRs are probably the oldest and most diverse of the membrane receptors and mediate both sensory and endocrine transduction. The protein traverses the plasma membrane seven times; several of these transmembrane α helices have conserved prolines that kink the helices, resulting in the formation of a ligand-binding pocket (Fig. 2c). As such, most of the hormones using these receptors are small: for example, the catecholamines, histamines, prostaglandins, and so on. Receptors with larger ligands often have extended amino termini to aid in hormone binding. Several of the intracellular loops bind and activate G-proteins (see Section 4.1), especially the third loop and the juxtamembrane portion of the carboxy terminus. The latter also forms a loop because its midsection is held to the plasma membrane by a palmitic acid that is covalently bound to a conserved cysteine in the carboxy terminus.

The last superfamily includes the ligand-gated ion channels that comprise three families. The cysteine-loop family is a pentamer of homologous subunits; each subunit contributes an α-helix toward forming the wall of the channel. This helix is kinked with the elbow directed toward the center of the channel; furthermore, the bend is populated by hydrophobic amino acids that obstruct the channel in the closed state. Hormone binding causes the helix to rotate; the hydrophobic side chains are moved laterally, and the channel opens. Acetylcholine activates sodium channels; glycine and γ-aminoisobutyric acid, chloride channels; and serotonin, calcium channels.

The voltage-gated channels are the prototype for the P (or H5) family. They are homotetramers; each subunit has six transmembrane helices. The pore is formed by a loop between the last two helices. Two subfamilies from this group are regulated by second messengers: first, the cyclic nucleotide–gated channels are cation pores opened by cAMP or cGMP. The second family gates channels that release internal stores of calcium; they include the inositol 1,4,5-trisphosphate (IP_3) and the cyclic ADP–ribose receptors (see Section 4.2). The glutamate receptors are another subgroup of the P family. They have only three transmembrane helices, and the pore-lining loop is between the first two helices. They are calcium channels.

The structure of the last superfamily, the epithelial sodium channels, is still controversial. They include the P_{2X} purinergic receptor and the FMRF amide-gated channel.

4 Second Messengers

4.1 G-proteins and Cyclic Nucleotides

G-proteins are molecular switches: they are active when GTP is bound to them and inactive when GDP is bound. In addition, they have intrinsic, although weak, GTPase activity; as such, they eventually turn themselves off when they hydrolyze the bound GTP to GDP. In the active state, they can stimulate enzymes, ion channels, and affect the cytoskeleton and vesicular trafficking (Table 1).

For a hormone to "flip" this switch on, it must facilitate the exchange of GDP for GTP. For example, the conversion of Ras.GDP to Ras.GTP is accomplished by a Ras GNP exchange facilitator (RasGEF), which is under hormone regulation (Fig. 3). In fact, there are several different types of RasGEFs, each with its own mechanism of control. In neurons, one RasGEF (called RasGRF1) is directly activated by calcium and a calcium-binding protein, called *calmodulin* (see Section 4.2); on the other hand, phosphorylation by a cAMP-dependent protein kinase will inhibit this GNP exchange facilitator (GEF). Another GEF, CNrasGEF, directly binds cyclic nucleotides and is stimulated by them. A third RasGEF (called mammalian SOS) is activated indirectly: receptor or soluble tyrosine kinases either phosphorylate themselves or some docking protein. An adaptor (Grb2) binds to the phosphotyrosine via its SH2 domain; Grb2 also has an SH3 domain that binds polyproline helices. Such a helix is present in RasGEF so that a complex among a tyrosine phosphorylated protein, Grb2, and RasGEF is formed. This complex activates RasGEF, which then stimulates nucleotide exchange on Ras. Once formed, Ras.GTP binds to Raf, a serine–threonine kinase, and brings it to the plasmalemma where it is stimulated by a still unknown process. Finally, Raf initiates a protein kinase cascade that eventually

Tab. 1 Some G-protein families and their functions.

Group	*Members*	*Function*
Ras superfamily		
Ras	Ras	Mitogenesis
	Ral	Endocytosis, vesicular trafficking
	Rap	Cell adhesion
	Rad, Gem, Kir	Cytoskeletal organization, secretion, Rho kinase inhibition
Rho family	Rho, Rac, Cdc42	Cytoskeletal organization, NADPH oxidase activation
Rab family	Rab	Vesicular trafficking
Ran family	Ran	Chromosomal condensation, nucleocytoplasmic transport, microtubule assembly (spindle)
Arf family	Arf	Endoplasmic reticulum–Golgi vesicle interactions
G-protein trimer	G_s, G_i, G_q, G_{12}, G_{13}	Hormone signal transduction
Elongation factor	EF-Tu	Translation proofreading
Signal recognition particle	SRP54	Transfer of proteins across the endoplasmic reticulum
Dynamin	Dynamin	Assembly of clathrin-coated vesicles vesicular neck constriction

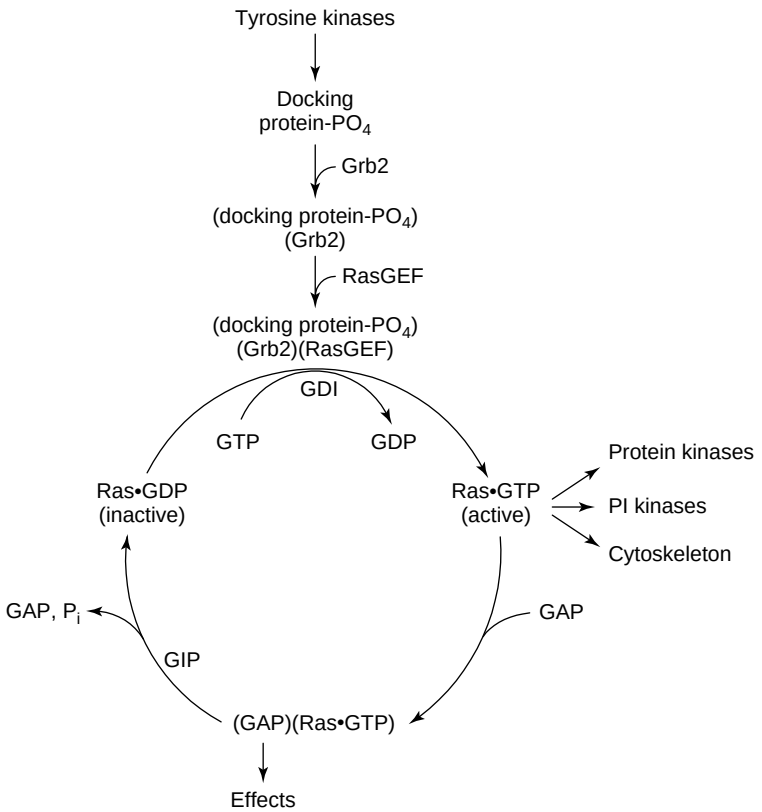

Fig. 3 Activation/inactivation cycle for Ras, a small G-protein: GAP, GTPase activating protein; GDI, GDP dissociation inhibitor; GEF, GNP exchange facilitator; GIP, GTPase-inhibiting protein; Grb2, an adaptor.

activates mitogen-activated protein (MAP) kinase, a critical kinase in transcription regulation (see Section 5.2). Other small G-proteins can directly bind and stimulate protein kinases (Table 2): for example, Rho activation of protein kinase N (PKN) and Rho-associated, coiled-coil forming protein kinase (ROCK); or Rac activation of the p21-activated kinase (Pak).

The GTPase activity of small G-proteins is so weak that they require accessory proteins, the GTPase activating proteins or GAPs. This is another potential site for hormone input: RasGAP can be inhibited by phosphoinositide binding, thereby prolonging the activated state of Ras.GTP. Interestingly, RasGAP, in addition to terminating the Ras signal, may also mediate some of the effects of Ras prior to GTP hydrolysis. Other potential control points include the GDP dissociation inhibitor (GDI), which opposes the GEF, and the GTPase-inhibiting protein (GIP), which opposes the GAP.

The large G-proteins occur as heterotrimers: the α-subunit is the GTPase, while the $\beta\gamma$ dimer is involved with membrane localization and protein association. Although the GTPase activity of α is not as slow as that in the small

Tab. 2 Some multifunctional serine–threonine kinases involved with hormone action.

Kinase	*Abbreviation*	*Structure*	*Activator(s)*
Protein kinase A	PKA	Tetramer	cAMP
Protein kinase B	PKB	Dimer	$PI(3)P_n$
Protein kinase C	PKC	Monomer	Calcium and diacylglycerol
Protein kinase G	PKG	Homodimer	cGMP
Calmodulin-dependent protein kinase II	CaMKII	Dodecamer	Calmodulin
Rho-associated coiled-coil forming protein kinase	ROCK	Monomer	Rho, inhibited by Rad
p21-activated kinase	Pak	Monomer	Rac
Casein kinase 2	CK2	$\alpha\beta$ dimer	Polyamines and phosphorylation
Glycogen synthase kinase 3	GSK3	Monomer	Inhibited by PKB or PKC phosphorylation
Mitogen-activated protein kinase	MAPK	Monomer	Ras and PKC via a kinase cascade

G-proteins, it still has a GAP, called negative regulator of G-protein signaling (RGS). There are four large G-protein families: G_s stimulates adenylate cyclase, G_i inhibits adenylate cyclase, G_q stimulates phospholipase Cβ (PLCβ), and $G_{12/13}$ activates phospholipase ε, protein phosphatase (PP) 5, and a RhoGEF. These G-proteins also directly stimulate several soluble tyrosine kinases and ion channels. The hormone–receptor complex directly accelerates nucleotide exchange, resulting

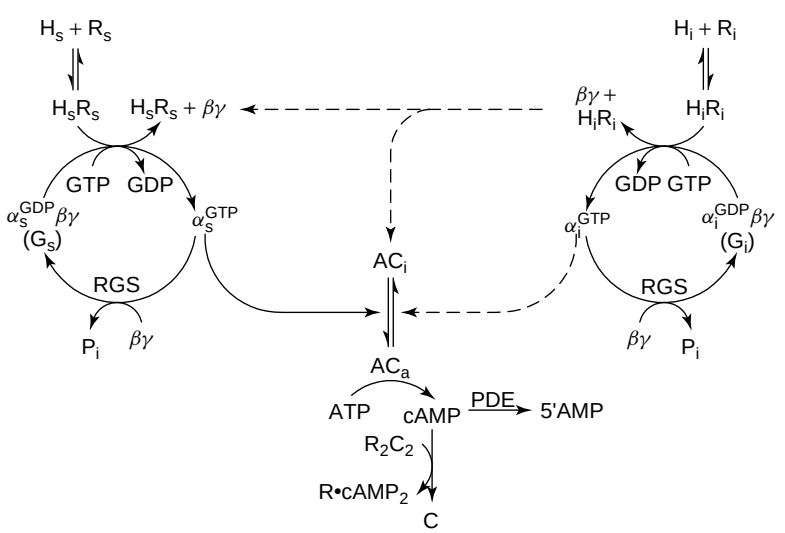

Fig. 4 Regulation of adenylate cyclase by several G-protein trimers. Inhibitory effects are depicted by dashed lines: H_iR_i, adenylate cyclase inhibiting hormone and its receptor; H_sR_s, adenylate cyclase stimulating hormone and its receptor; R_2C_2, PKA tetramer containing two regulatory (R) and two catalytic subunits (C); RGS, negative regulator of G-protein signaling.

in the activation of α_s and the dissociation of $\beta\gamma$ (Fig. 4). As long as GTP remains bound to α_s, adenylate cyclase will be stimulated. However, α_s will eventually hydrolyze GTP and reassociate with $\beta\gamma$. G_i represents a counterregulatory mechanism that operates in a parallel manner except that both α_i and $\beta\gamma$ participate in the inhibition of adenylate cyclase.

The ultimate output to this pathway is cAMP, which can activate several enzymes and ion channels; however, the major effector for cAMP is a serine–threonine kinase (protein kinase A or PKA). This tetramer has two catalytic subunits and two inhibitory subunits (Table 2). cAMP binds to the regulatory subunits and causes them to dissociate, thereby removing the inhibition.

cGMP is another cyclic nucleotide used as a mediator of hormone action; it may be generated by two different pathways. The first pathway, discussed in Section 3.2, involves a hormone that directly binds and activates a membrane-bound guanylate cyclase. cGMP may also be synthesized by a soluble cyclase; this enzyme is activated by nitric oxide whose synthesis is stimulated by hormones that elevate calcium (see Section 4.2). Like cAMP, cGMP stimulates both ion channels and a homologous kinase, protein kinase G or PKG. In this kinase, the regulatory and catalytic subunits are fused into a single protein.

4.2 Calcium, Calmodulin, and Phospholipids

Calcium is an abundant cation in extracellular fluids; in addition, it is concentrated within several cellular organelles, such as mitochondria and elements of the smooth endoplasmic reticulum. However, cytosolic levels are kept extremely low, because many cellular processes are dramatically affected by calcium. As such, hormones can regulate these cellular functions by controlling the cytoplasmic calcium concentrations. The most direct mechanism for elevating calcium would be for hormones to activate ligand-gated calcium channels, such as the glutamate receptor; by merely opening these channels, hormones would cause external calcium to flood the cytoplasm.

However, a major source of hormonally released calcium is internal (Fig. 5, middle). Briefly, hormones stimulate a PLC to hydrolyze a phospholipid (polyphosphoinositide) to diacylglycerol (DG) and the former head group (IP_3). There are several PLC groups distinguished by both their structure and their hormone regulation. PLCγ possesses two SH2 domains through which it binds to autophosphorylated RTKs; this association brings the enzyme into close proximity with its substrate and allows RTKs to phosphorylate and stimulate the PLCγ. On the other hand, PLCβ is allosterically activated by α_q or $\beta\gamma$; PLCδ, by calcium; and PLCε, by Ras, Rap, or G_{12}.

Once released, IP_3 diffuses through the cytosol to its receptor, a calcium channel on the endoplasmic reticulum. IP_3 binding opens the channel and allows the internally stored calcium to enter the cytoplasm. The DG remains in the plasma membrane where it stimulates a calcium-activated, phospholipid-dependent protein kinase, protein kinase C (PKC). Actually, there are three PKC groups that differ in their calcium and DG requirements: the conventional PKCs require both calcium and DG; the novel PKCs require only DG; and the atypical PKCs are activated by alternate second messengers, such as arachidonic acid. DG can also have other effects, including the activation of an acid sphingomyelinase and GEFs for Ras and Rap.

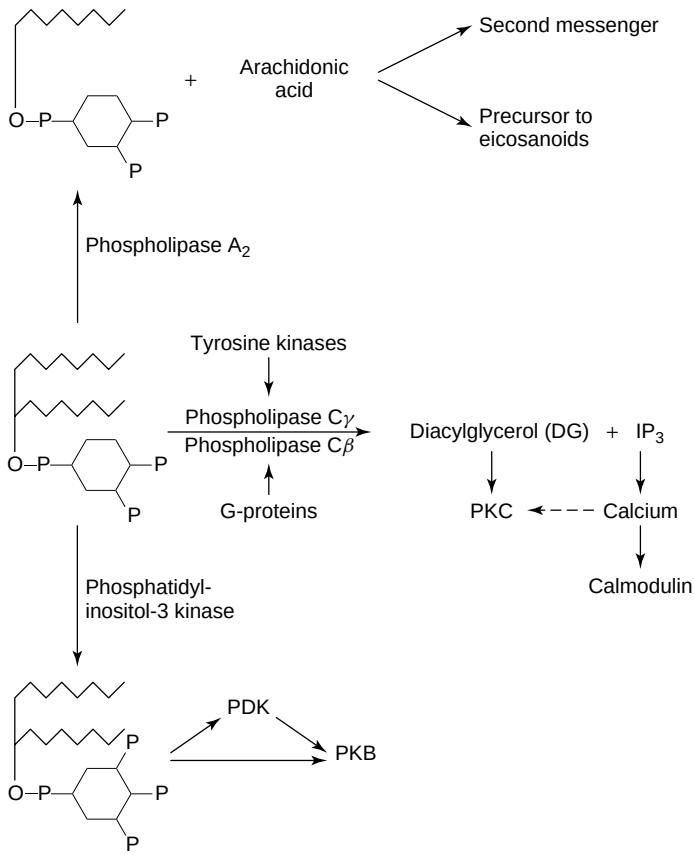

Fig. 5 Brief summary of several phospholipid pathways: IP_3, inositol 1,4,5-trisphosphate; PDK, phosphoinositide-dependent kinase; PKB, protein kinase B; PKC, protein kinase C.

Once elevated, calcium alone can directly affect many enzymes, the cytoskeleton, and other biological processes. However, it frequently acts through a calcium-binding protein, calmodulin (CaM). CaM is a small dumbbell-shaped peptide with the two globular ends separated by an α-helix. Calcium binding to the globular ends allows the groove found in each end to wrap around an α-helix in the target protein. The central α-helix of CaM helps determine binding selectively: it has an acidic and a hydrophobic end and is attracted to target proteins having an exposed α-helix with a basic and a lipophilic end. Such amphipathic helices often inhibit the proteins that possess them, and their masking by CaM prevents this autoinhibition. For example, the CaM-dependent protein kinase II, a general-purpose serine–threonine kinase, has a CaM binding site that blocks both the ATP binding and catalytic sites. In the presence of calcium, CaM binds to this site, prevents it from interfering with ATP and substrate binding, and activates the kinase. There are many other enzymes and channels that operate on the same principle.

Phospholipids may give rise to many other second messengers: for example, almost all growth factors stimulate a phosphatidylinositol-3 kinase, which generates a phospholipid that cannot be hydrolyzed by any known PLC. Rather, the intact molecule is a second messenger that activates several protein kinases, such as the phosphoinositide-dependent kinase (PDK) and protein kinase B (PKB); stimulation of the latter requires both the phospholipid and phosphorylation by PDK (Fig. 5, bottom). PKB is a major antiapoptotic signal.

Phospholipase A_2 can also generate signaling molecules, and it can be stimulated by hormones via MAP kinase phosphorylation and calcium. It liberates the fatty acid in the second position of phosphoglycerides; this is usually arachidonic acid. The free arachidonate may stimulate enzymes or other effectors, or it may be converted to prostaglandins or other eicosanoids (Fig. 5, top). These latter fatty acid derivatives are hormones in their own right: they are secreted, bind to specific membrane receptors, and stimulate various second messenger pathways. However, they usually act locally; such molecules are called *parahormones* and they may adjust a cell's response to other hormones or to local conditions. A lysophospholipid is what remains after the arachidonic acid has been removed, and it is also biologically active.

Finally, another membrane lipid, sphingomyelin, can be the source of several mediators of hormone action. Sphingomyelin is structurally similar to phosphatidylcholine except that it has a serine rather than a glycerol backbone. Acid sphingomyelinase is activated by DG, while the neutral sphingomyelinase is stimulated by an adaptor to the tumor necrosis factor receptor. These enzymes generate ceramide, which can then be converted to sphingosine-1-phosphate by phosphorylation. The former mediates apoptosis, while the latter stimulates proliferation.

4.3 Miscellaneous Mediators

Several other second messenger systems have been postulated, but they have not yet been sufficiently explored to determine their physiological relevance to hormone action. Polyamines are small basic molecules that are required for transcription and translation. They can also activate a protein kinase, casein kinase 2. Hormonal regulation is achieved by the induction of the enzymes for polyamine synthesis; hormones can also stimulate polyamine transport into cells from the surrounding medium.

The hydrogen ion concentration represents another potential mediator of hormone action. All mitogenic stimulation is accompanied by a slight rise in cellular pH. This alkanization has been shown to be necessary, but not sufficient, for subsequent cell division. It is not clear what the output for this signal is; although many cellular reactions are pH sensitive, pH optima are rarely so sharp that a few tenths of a pH unit would make a significant difference in overall activity. Growth factors raise the pH by activating the sodium–hydrogen exchanger 1 (NHE1) by PKC phosphorylation, G-proteins, or calcium–calmodulin binding to an autoinhibitory site.

5 Biological Effects

5.1 Nongenomic

There are two major mechanisms by which hormones can affect cellular processes:

allosterism and phosphorylation. For example, G-proteins can directly bind and alter the activity of both enzymes (e.g. adenylate cyclase, phospholipases, and several kinases) and transporters (e.g. ion and glucose channels and the NHE1). Similarly, cyclic nucleotides can stimulate nucleotide-gated channels and protein kinases. Lipophilic hormones, which have direct access to the cellular interior, have also been shown to bind directly several enzymes, but the physiological relevance of these observations is still controversial.

However, the broadest effects are generally achieved by phosphorylation. Indeed, virtually every known second messenger pathway activates at least one protein kinase and several also affect PPs (Tables 2 and 3); the extent of this modification is controlled by a balance between these two enzyme groups. Phosphorylation is a major regulatory mechanism in metabolism; glycogen metabolism is one of the best-known examples. Hormones that stimulate glycogen breakdown activate PKA, which initiates a protein kinase cascade leading to the phosphorylation and stimulation of glycogen phosphorylase. In addition, PKA and other kinases modify and inhibit glycogen synthase; that is, glycogen breakdown is activated, while its synthesis is blocked. On the other hand, insulin favors glycogen synthesis; it does so by activating the phosphatases that reverse these phosphorylations. Many other metabolic cycles are also regulated by this modification.

In addition, phosphorylation can affect other cellular processes: it can alter the function of various channels and transporters and trigger the breakdown of the cytoskeleton.

5.2 Genomic

Hormones and their mediators can also affect gene expression by the same mechanisms described in the previous section (Table 4). For example, nuclear

Tab. 3 The hormone regulation of several protein phosphatases.

Protein phosphatase (PP)	*Regulation*
Serine–threonine phosphatases	
PP1	Subunit composition affects localization
	Phosphorylation, direct or on inhibitory proteins (± depending upon the kinase)
PP2A	Polyamines, ceramide, phosphorylation, and carboxymethylation (+)
	ATP and several oncogene products can shift PP2A from a serine–threonine to a tyrosine phosphatase
PP2B (calcineurin)	Calcium–calmodulin (+)
	Phosphorylation and several immunosuppressants (−)
PP2C	Unknown
Protein tyrosine phosphatases (PTPs)	
SHP	Allosteric activation when its SH2 domain binds a phosphotyrosine
PTP-1B	Direct phosphorylation by RTKs (+)
PEST-PTP	PKA and PKC phosphorylation (−)

Tab. 4 Some transcription factors mediating hormone action.

Transcription factor	*Abbreviation*	*Primary activator(s)*
cAMP response element binding protein	CREB	PKA phosphorylation
Jun-Fos	AP-1	PKC phosphorylation (indirect)
Myc-Max		CK2 and MAPK phosphorylation
Myocyte enhancer factor 2	MEF2	CaM displaces inhibitor
Nuclear factor of activated T-lymphocytes	NFAT	Dephosphorylation by PP2B, a CaM-regulated phosphatase
Nuclear factor for κ genes of B-lymphocytes	NF-κB	PKC phosphorylation of inhibitor protein (indirect)
Nuclear receptors	GR, ER, TR, etc.	Ligand binding by steroids, retinoids, thyroid hormones, etc.
Serum response factor	SRF	CK2 and MAPK phosphorylation
Signal transducers and activators of transcription	Stat	Direct phosphorylation by receptor or soluble tyrosine kinases
Tubby		Sequestered by polyphosphoinositides at plasma membrane until PLC hydrolysis

receptors are actually transcription factors that are allosterically regulated by their ligands, usually hormones but also second messengers like arachidonic acid. Many other transcription factors are controlled by phosphorylation. The simplest mechanism would be for an RTK or a cytokine receptor with associated soluble tyrosine kinase to phosphorylate directly a transcription factor that would then migrate to the nucleus and activate gene expression. The Stat family is activated by such a mechanism (Fig. 6a).

However, most transcription factors are phosphorylated via second messenger pathways. For example, hormones whose activities are mediated by cAMP first activate G_s; the α subunit then stimulates adenylate cyclase, and the resulting accumulation of cAMP activates PKA. Finally, PKA modifies a cAMP response element binding protein (CREB); in particular, the phosphorylation occurs in a TAD that can then recruit a critical transcription coactivator (Fig. 6b).

Phosphorylation can also occur on accessory proteins: NF-κB is a transcription factor frequently associated with defense genes. It is held in the cytoplasm by an inhibitory subunit, IκB. After being modified by PKC-activated kinase, IκB dissociates, allowing NF-κB to translocate to the nucleus, bind its HRE, and activate transcription (Fig. 6c, right). Phosphorylation is not always stimulatory: NFAT, another transcription factor associated with defense genes, is tonically inhibited by phosphorylation. Activation occurs when hormones elevate calcium concentrations that stimulate PP2B (Table 3), leading to the dephosphorylation of NFAT (Fig. 6c, left).

Many second messengers can directly bind transcription factors (Fig. 6c, left). Calcium binds and inhibits a repressor known as *downstream regulator element antagonist modulator* (*DREAM*). CaM displaces Cain, an inhibitor, from the myocyte enhancer factor 2 (MEF2). Tubby is sequestered at the plasma

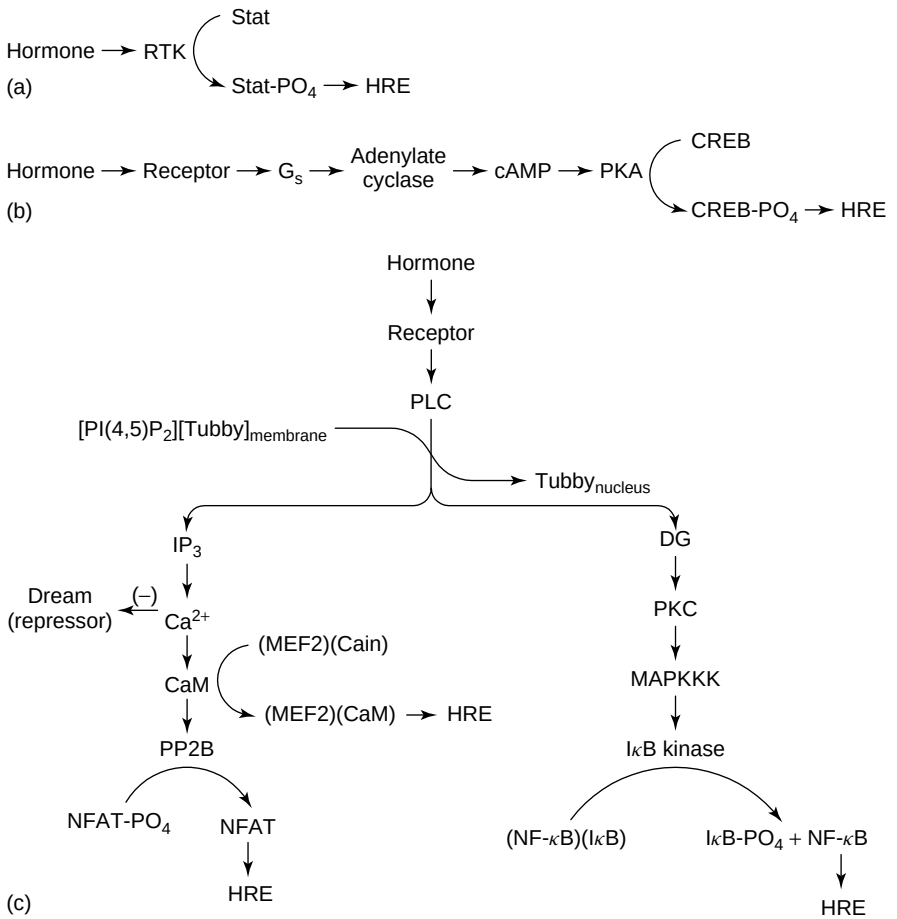

Fig. 6 Activation of several transcription factors by various second messenger pathways: CaM, calmodulin; DG, diacylglycerol; DREAM, downstream regulator element antagonist modulator; HRE, hormone response element; MEF2, myocyte enhancer factor 2; PLC, phospholipase C; RTK, receptor tyrosine kinase. See Tables 2 and 4 for other abbreviations.

membrane by binding to polyphosphoinositide. Hormonal stimulation of PLC hydrolyzes the phospholipid and releases Tubby to migrate to the nucleus and activate transcription.

The reader should be warned that the above discussion is an oversimplification. In fact, many transcription factors may be modified by several different kinases. This cross talk allows for the integration of multiple signals, although it may at times appear confusing. Therefore, this general overview has concentrated only on the dominant regulators of several major pathways.

Finally, hormones can also affect gene expression at posttranscriptional sites. They can induce enzymes that process RNA, binding proteins that stabilize or destabilize mRNA, and activate kinases that can phosphorylate components of the translational machinery.

6 Clinical Applications

Nowhere have the benefits of molecular biology been more obvious than in the diagnosis and management of endocrine diseases. Hormones are so active that their serum concentrations are in the nanomolar range or less, and their receptors rarely exceed more than a few thousand per cell. With such small numbers, conventional protein purification and characterization techniques are inadequate. Most defects in hormones, receptors, and transducers are identified by nucleotide sequencing after amplification by the polymerase chain reaction. Such techniques have been so successful that a review of this size cannot begin to cover all the endocrinopathies whose molecular bases have been determined; however, mention will be made of all the general categories, and a single example of each will be given.

6.1 Endocrine Deficiencies

6.1.1 Hormones

A peptide hormone deficiency can most simply arise from a deletion or mutation in the gene for that hormone. The former occurs with the growth hormone (GH) gene in *isolated GH deficiency, type IA*, and the latter has been reported in the insulin gene in some forms of *diabetes mellitus*. Other mutations may affect the processing of hormones. For example, mutations in the cleavage sites for the signal sequence of parathormone (PTH) or between the different chains of insulin are responsible for *familial isolated hypoparathyroidism* and *hyperproinsulinemia* respectively. Antidiuretic hormone (ADH) concentrates the urine and is synthesized as a polyprotein: the amino terminus forms this hormone, while the carboxy terminus is a transport protein for ADH. Mutations in the transport protein region impair ADH packaging and transport, leading to ADH deficiency (called *familial neurohypophyseal diabetes insipidus*) even though ADH itself is normal. A final processing defect is represented by mutations in the protease responsible for cleaving polyproteins. Adrenocorticotropic hormone (ACTH) is a pituitary hormone that stimulates the adrenal glands to produce steroids; it is synthesized as a polyprotein containing ACTH and several other hormones. The enzyme that cuts out these hormones is defective in *isolated congenital ACTH deficiency*.

Steroids are not coded for by genes; they are synthesized by enzymes. Therefore, steroid deficiencies are actually deficiencies in steroid synthetic enzymes. Defects in virtually every enzyme have been reported and are grouped under the umbrella of *congenital adrenal hyperplasia*. A similar situation exists for the thyroid hormones; defects have been noted both in thyroid peroxidase and its substrate, thyroglobulin, in several forms of *goiter*.

Finally hormones may be deficient because of regulatory defects. Pit-1 is a transcription factor found exclusively in the anterior pituitary gland, and it stimulates transcription of the *GH* gene; mutations in Pit-1 lead to *familial human dwarfism*. Glucokinase represents another regulatory site; this metabolic enzyme phosphorylates glucose as soon as it is transported into the cell. In the insulin-secreting β-cells of the pancreas, it doubles as a glucose sensor; elevated glucose levels stimulate the secretion of insulin, which then activates metabolic pathways that store the glucose. Mutations in glucokinase prevent the β-cell from detecting glucose, and insulin fails to be secreted; this form of

diabetes is called *maturity onset diabetes of the young*.

6.1.2 Receptors

In many endocrinopathies, the hormone is normal in structure and concentration, but the target tissue is unresponsive. The first cellular component that the hormone encounters is its receptor, and many defects in this element have been reported. Mutations for almost every nuclear receptor have been described, but those for the androgen receptor are the most numerous, probably because androgen function is not essential for survival. The default body plan for mammals is female; males are created when the testes produce androgens that masculinize the embryo. Partially active androgen receptors leave this process incomplete, producing a *male pseudohermaphrodite*. Totally inactive mutants result in a phenotypic female; this condition is called *testicular feminization*.

Membrane receptors can also be defective. Mutations in the insulin receptor, an RTK, lead to severe forms of diabetes mellitus; those in the GH receptor, a cytokine receptor, result in *Laron dwarfism*; and defects in the ADH receptor, a GPCR, produce *hereditary nephrogenic diabetes insipidus*. Even mutations in ligand-gated ion channels have been identified: the glycine receptor is a chloride channel whose activation hyperpolarizes, and therefore inhibits, neurons. Defects in this receptor leave stimulatory signals unchecked and are responsible for *hereditary hyperekplexia*, a disease characterized by overstimulation of the muscles and an exaggerated startle response.

6.1.3 Transduction

Examples of defective transduction pathways are much rarer. Many different hormones activate the same mediators; therefore, mutations in these second messenger pathways would have much greater effects than the absence of a single hormone or its receptor. Most reported defects have been in special isoforms restricted to certain tissues or organs so that their effects would be limited. For example, the eye possesses several unique isoenzymes of the phosphoinositide pathway; mutations in any of these components would only result in blindness. Another example is found in the kidneys where ADH stimulates the translocation of a water channel, aquaporin-2, from the cellular interior to the plasmalemma. This allows the kidney tubule to resorb water and concentrate urine. In one form of *nephrogenic diabetes insipidus*, this water channel is mutationally inactivated. ADH and its receptor are normal, but the kidney cannot respond to ADH because the water channel is no longer operational.

One of the few examples of a generalized defect is *pseudohypoparathyroidism, type IA*, where G_s is inactive or absent. Although this defect would affect any hormone utilizing cAMP as a second messenger, the actions of PTH appear to be most severely impacted. PTH maintains serum calcium levels by resorbing this mineral from bone and recovering it from urine. The loss of responsiveness to PTH leads to skeletal abnormalities.

6.2 Endocrine Excesses

Perhaps, more fascinating than loss-of-function mutations are those that actually lead to overactivity. No mutations that increase the intrinsic activity of a hormone have been reported, but mutations affecting hormone processing are known. A mutant preproneuropeptide Y

exhibits enhanced cleavage from its precursor, resulting in higher levels, endothelial proliferation, and an increased risk of atherosclerosis. A mutant fibroblast growth factor 23 is resistant to proteolysis, and its accumulation leads *to autosomal dominant hypophosphatemic rickets*. Altered catabolism also forms the basis for one form of *hyperaldosteronism*. Aldosterone is a steroid that regulates sodium metabolism; excess aldosterone activity causes sodium to be retained and blood pressure to rise. The aldosterone receptor cannot distinguish between aldosterone and cortisol; since cortisol is present in serum at concentrations much higher than aldosterone, cortisol should flood the aldosterone receptor and keep it continuously stimulated. This does not happen because in target tissues, 11β-hydroxysteroid dehydrogenase rapidly metabolizes cortisol before it can reach the receptor. However, a deficiency of this enzyme leads to the overstimulation of the aldosterone receptor by cortisol.

Another form of hyperaldosteronism is caused by a regulatory mutation. The enzymes that synthesize cortisol and aldosterone are very homologous and their genes are contiguous on the chromosome; ACTH stimulates the former via elements in the 5′ end of the gene. Because of the similarity between the genes, they can accidentally pair up during meiosis, resulting in a nonreciprocal crossing-over. In this way, the gene for aldosterone synthase acquires the regulatory region from the other gene and the latter is lost. Because there is no cortisol, ACTH is elevated; but instead of stimulating cortisol synthesis, it stimulates aldosterone synthase. The patient is treated with cortisol, which feeds back to inhibit ACTH secretion; the aldosterone synthase gene, in turn, is shut off. For this reason, the disease is called *glucocorticoid remediable aldosteronism*.

Finally, hormone secretion can be elevated by faulty negative feedback. As noted above, PTH restores serum calcium concentrations by resorbing it from bones and recovering it from urine. As calcium levels rise, PTH secretion shuts off. A mutant calcium sensor blocks this negative feedback and results in the persistent secretion of PTH, which in turn elevates calcium. The heterozygous state, called *familial hypocalciuric hypercalcemia*, manifests only mild symptoms, but the homozygous condition, called *neonatal severe hyperparathyroidism*, produces a more grave disease.

Mutations can also render receptors constitutively active in the absence of their natural ligand. Those in RTKs often produce tumors, since these receptors usually serve growth factors; they are discussed in this context in Section 6.3. An activating mutation in the erythropoietin receptor eliminates the binding site for the phosphatase that normally deactivates it. The hyperactive receptor overstimulates red blood cell production and elevates the hematocrit; this disease is called *primary familial polycythemia*. Several activating mutations have also been reported in GPCRs; they usually occur in either the transmembrane helices or the cytosolic loops. The former are thought to mimic the ligand-occupied state, while the latter improve coupling to G-proteins. Luteinizing hormone (LH) is a pituitary hormone that stimulates the gonads to produce sex steroids. Activating mutations in its receptor can stimulate steroidogenesis prematurely, giving rise to *familial male precocious puberty*. Because women require a second hormone for estrogen production, mutations in the LH receptor alone do not lead to any symptoms.

Finally, activating mutations have been identified in a transducer. G_s inactivates itself when it hydrolyzes GTP to GDP (see Section 4.1); mutations that adversely affect its GTPase activity without altering its ability to interact with its effectors would result in the constitutive activation of G_s. Because G_s is integral to so many pathways, this type of mutation would be lethal if it were genomic; however, somatic mutations would produce a genetic mosaic in which only some cells would possess the defective gene. Individuals with this type of G_s mosaicism have *McCune–Albright syndrome* and have symptoms reflective of overstimulation of all the hormones that utilize cAMP as a second messenger. These hormones include those that stimulate pigment cells (café au lait spots), the gonads (precocious puberty), the thyroid gland (hyperthyroidism), the adrenal glands (adrenal hyperplasia), and bone (dysplasias).

6.3 Oncogenesis

Cancer diagnosis and treatment is another area where molecular endocrinology has made substantial contributions.

6.3.1 Oncogenes

Before defining an oncogene, it would be advantageous to define a protooncogene. A protooncogene is a normal cellular gene that is usually involved in cell growth. If this gene is mutated so that its product is constitutively active, then cell growth may proceed unchecked and tumor formation may occur. Such a modified gene is now called an *oncogene*. Oncogenes may arise spontaneously within the organism's genome or they may be introduced by oncogenic viruses. In the latter case, it is believed that these viruses originally acquired protooncogenes from their hosts and later converted them to oncogenes because they conferred some reproductive advantage on the virus. For example, a rapidly dividing cell may have an abundance of replicative machinery that the virus can then commandeer for itself.

Since most viruses have small genomes, it would be more efficient for the virus to plug into a preexisting proliferative pathway than to control all the processes of cell division itself. As such, most oncogenes are components of the growth factor pathway. Some are GHs, like the platelet-derived growth factor and the fibroblast growth factor. Others are receptors for growth factors; the most abundant groups are the RTKs and the cytokine receptors with their associated tyrosine kinases. Many mutations involve deletions of the extracellular domain and other changes that favor aggregation. Remember, the only function of the ligand in these receptors is to induce oligomerization, which then activates the intrinsic or associated tyrosine kinases. Other mutations may incapacitate autoinhibitory sites within the kinases.

Second messengers are also well represented, especially the small G-proteins such as Ras and various adaptors. In the case of G-proteins, mutations often occur in the GAP binding region; without the ability to interact with GAP, GTP cannot be hydrolyzed and the G-proteins remain permanently active. Finally, several oncogenes are transcription factors. Although most of these factors are constitutively active like the receptors and G-proteins, a few are dominant inhibitors. The active factors are involved with cell proliferation, while the inhibitory ones are involved with either differentiation or defense responses. Because a cell has limited

resources, growth and differentiation often compete with each other; therefore, tumor formation requires both the induction of cell proliferation and the inhibition of differentiation. Examples of factors in the latter group include a mutated thyroid hormone receptor and a mosaic retinoic acid receptor.

6.3.2 Tumor Resistance to Therapy

When hormone target tissues become malignant, they often retain their endocrine sensitivity. Such tumors can be initially treated by hormone therapy: breast cancer may respond to antiestrogens; prostate cancer, to antiandrogens; and leukemias and lymphomas, to glucocorticoids. Hormone therapy is not only effective but also among the most easily tolerated treatments, because of its relative lack of undesirable side effects. Unfortunately, the effectiveness of this therapy is too often short lived as the tumor eventually develops resistance. Quantitative and structural studies of nuclear receptors can help determine which of these tumors will initially respond to hormone therapy and can also explain why many of them later become resistant.

Many resistant tumors have been found to have inactive receptors. The simplest case is seen in leukemias and lymphomas, which undergo apoptosis in response to glucocorticoids. Resistant tumors either lack glucocorticoid receptors or possess defective receptors and, therefore, can no longer respond to these steroids. It is felt that these deletions and mutations arise spontaneously; those that inactivate the receptor offer a selective advantage by allowing the cells that have them to evade the cytotoxic effects of glucocorticoids. Such cells survive and reproduce to create a cancer resistant to these steroids.

On the other hand, breast and prostate cancers are dependent upon sex steroids for survival; as such, they are treated by hormone ablation or antagonists. Therefore, simple deletion or inactivation of the receptor would prove detrimental to the survival of the cancer. Two types of mutations are observed in these cases. First, some mutations render the receptor constitutively active regardless of the presence or absence of hormone agonists or antagonists. Second, mutations in the ligand-binding domain may reduce the binding specificity of the receptor so as to allow other steroids, or even steroid antagonists, to activate the receptor. In either instance, the tumors have either reduced or eliminated hormone dependence.

7 Perspectives

Single cells require careful regulation of metabolism, growth, and reproduction in order to survive in a competitive and dangerous environment. The development of multicellularity introduced the additional problem of intercellular communications to coordinate these processes. Nature's frugal solution was to simply couple external chemical signals to these preexisting intracellular regulators, which then became second messengers. As such, knowledge about molecular endocrinology is knowledge about the very essence of how cells control all their internal functions. In addition, this information provides valuable insights into ourselves via the various disease states that can now be explained by molecular biology. Therefore, molecular endocrinology affords a critical understanding of many basic physiological processes of life.

See also Phospholipids.

Bibliography

Books and Reviews

Bolander, F.F. (1994) *Molecular Endocrinology*, 3rd edition, Academic Press, San Diego, CA.

Conn, P.M., Means, A.R. (Eds.) (2000) *Principles of Molecular Regulation*, Humana Press, Totowa, NJ.

Gomperts, B.D., Tatham, P.E.R., Kramer, I.M. (2002) *Signal Transduction*, Academic Press, San Diego, CA.

Helmeich, E.J.M. (2001) *The Biochemistry of Cell Signalling*, Oxford University Press, Oxford.

Krauss, G. (2001) *Biochemistry of Signal Transduction and Regulation*, 2nd edition, Wiley-VCH, Weinheim.

Primary Literature

Alderton, W.K., Cooper, C.E., Knowles, R.G. (2001) Nitric oxide synthases: structure, function and inhibition, *Biochem. J.* **357**, 593–615.

Anderson, K.E., Jackson, S.P. (2003) Class I phosphoinositide 3-kinases, *Int. J. Biochem. Cell Biol.* **35**, 1028–1033.

Aranda, A., Pascual, A. (2001) Nuclear hormone receptors and gene expression, *Physiol. Rev.* **81**, 1269–1304.

Belville, C., Josso, N., Picard, J.Y. (1999) Persistence of mullerian derivatives in male, *Am. J. Med. Genet.* **89**, 218–223.

Bishop, A.L., Hall, A. (2000) Rho GTPases and their effector proteins, *Biochem. J.* **348**, 241–255.

Bootman, M.D., Lipp, P., Berridge, M.J. (2001) The organisation and functions of local Ca^{2+} signals, *J. Cell Sci.* **114**, 2213–2222.

Bourguet, W., Germain, P., Gronemeyer, H. (2000) Nuclear receptor ligand-binding domains: three-dimensional structures, molecular interactions and pharmacological implications, *Trends Pharmacol. Sci.* **21**, 381–388.

Cantrell, D.A. (2001) Phosphoinositide 3-kinase signalling pathways, *J. Cell Sci.* **114**, 1439–1445.

Carpenter, L.R., Yancopoulos, G.D., Stahl, N. (1999) General mechanisms of cytokine receptor signaling, *Adv. Protein Chem.* **52**, 109–140.

Cohen, L.E., Radovick, S. (2002) Molecular basis of combined pituitary hormone deficiencies, *Endocr. Rev.* **23**, 431–442.

Cohen, P.T.W. (2002) Protein phosphatase 1-targeted in many directions, *J. Cell Sci.* **115**, 241–256.

Crabtree, G.R. (2001) Calcium, calcineurin, and the control of transcription, *J. Biol. Chem.* **276**, 2313–2316.

Delfino, F., Walker, W.H. (1999) Hormonal regulation of the NF-κB signaling pathway, *Mol. Cell. Endocrinol.* **157**, 1–9.

Fukami, K. (2002) Structure, regulation, and function of phospholipase C isozymes, *J. Biochem.* **131**, 293–299.

Gennari, L., Becherini, L., Falchetti, A., Masi, L., Massart, F., Brandi, M.L. (2002) Genetics of osteoporosis: role of steroid hormone receptor gene polymorphisms, *J. Steroid Biochem. Mol. Biol.* **81**, 1–24.

Hamm, H.E. (2001) How activated receptors couple to G proteins, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4819–4821.

Hoeflich, K.P., Ikura, M. (2002) Calmodulin in action: diversity in target recognition and activation mechanisms, *Cell* **108**, 739–742.

Holmberg, C.I., Tran, S.E.F., Eriksson, J.E., Sistonen, L. (2002) Multisite phosphorylation provides sophisticated regulation of transcription factors, *Trends Biochem. Sci.* **27**, 619–627.

Huhtaniemi, I. (2000) Activating and inactivating hormone receptor mutations, *Horm. Res.* **53**(Suppl. 3), 9–16.

Imada, K., Leonard, W.J. (2000) The Jak-STAT pathway, *Mol. Immunol.* **37**, 1–11.

Janssens, V., Goris, J. (2001) Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling, *Biochem. J.* **353**, 417–439.

Klee, C.B., Ren, H., Wang, X. (1998) Regulation of the calmodulin-stimulated protein phosphatase, calcineurin, *J. Biol. Chem.* **273**, 13367–13370.

Lania, A., Mantovani, G., Spada, A. (2001) G protein mutations in endocrine diseases, *Eur. J. Endocrinol.* **145**, 543–559.

Lawlor, M.A., Alessi, D.R. (2001) PKB/Akt: a key mediator of cell proliferation, survival and insulin responses? *J. Cell Sci.* **114**, 2903–2910.

Mayr, B., Montminy, M. (2001) Transcriptional regulation by the phosphorylation-dependent factor CREB, *Nat. Rev. Mol. Cell Biol.* **2**, 599–609.

Moustakas, A., Souchelnytskyi, S., Heldin, C.H. (2001) Smad regulation in TGF-β signal transduction, *J. Cell Sci.* **114**, 4359–4369.

Murakami, M., Kudo, I. (2002) Phospholipase A_2, *J. Biochem.* **131**, 285–292.

New, M.I., Wilson, R.C. (1999) Steroid disorders in children: congenital adrenal hyperplasia and apparent mineralocorticoid excess, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12790–12797.

Ohanian, J., Ohanian, V. (2001) Sphingolipids in mammalian cell signalling, *Cell. Mol. Life Sci.* **58**, 2053–2068.

Ornitz, D.M., Marie, P.J. (2002) FGF signaling pathways in endochondral and intramembranous bone development and human genetic disease, *Genes Dev.* **16**, 1446–1465.

Owen, K., Hattersley, A.T. (2001) Maturity-onset diabetes of the young: from clinical description to molecular genetic characterization, *Best Pract. Res. Clin. Endocrinol. Metab.* **15**, 309–323.

Pierce, K.L., Premont, R.T., Lefkowitz, R.J. (2002) Seven-transmembrane receptors, *Nat. Rev. Mol. Cell Biol.* **3**, 639–650.

Rebollo, A., Martine Castelao A. (1999) Ras proteins: recent advances and new functions, *Blood* **94**, 2971–2980.

Renaud, J.P., Moras, D. (2000) Structural studies on nuclear receptors, *Cell. Mol. Life Sci.* **57**, 1748–1769.

Schenk, P.W., Snaar-Jagalska, B.E. (1999) Signal perception and transduction: the role of protein kinases, *Biochim. Biophys. Acta* **1449**, 1–24.

Schlessinger, J. (2000) Cell signaling by receptor tyrosine kinases, *Cell* **103**, 211–225.

Silberbach, M., Roberts, C.T. (2001) Natriuretic peptide signalling: molecular and cellular pathways to growth regulation, *Cell. Signal.* **13**, 221–231.

Sluyser, M. (1994) Hormone resistance in cancer: the role of abnormal steroid receptors, *Crit. Rev. Oncog.* **5**, 539–554.

Spiegel, A.M. (2000) G protein defects in signal transduction, *Horm. Res.* **53**(Suppl. 3), 17–22.

Strader, C.D., Fong, T.M., Graziano, M.P., Tota, M.R. (1995) The family of G-protein-coupled receptors, *FASEB J.* **9**, 745–754.

Takai, Y., Sasaki, T., Matozaki, T. (2001) Small GTP-binding proteins, *Physiol. Rev.* **81**, 153–208.

Tanoue, T., Nishida, E. (2003) Molecular recognitions in the MAP kinase cascades, *Cell. Signal.* **15**, 455–462.

Tenbaum, S., Baniahmad, A. (1997) Nuclear receptors: structure, function and involvement in disease, *Int. J. Biochem. Cell Biol.* **29**, 1325–1341.

Vajo, Z., Francomano, C.A., Wilkin, D.J. (2000) The molecular and genetic basis of fibroblast growth factor receptor 3 disorders: the achondroplasia family of skeletal dysplasias, Muenke craniosynostosis, and Crouzon syndrome with acanthosis nigricans, *Endocr. Rev.* **21**, 23–39.

Vanderbeld, B., Kelly, G.M. (2000) New thoughts on the role of the $\beta\gamma$ subunit in G protein signal transduction, *Biochem. Cell Biol.* **78**, 537–550.

Weinstein, L.S., Yu, S., Warner, D.R., Liu, J. (2001) Endocrine manifestations of stimulatory G protein α-subunit mutations and the role of genomic imprinting, *Endocr. Rev.* **22**, 675–705.

Whitmarsh, A.J., Davis, R.J. (2000) Regulation of transcription factor function by phosphorylation, *Cell. Mol. Life Sci.* **57**, 1172–1183.

Wilson, R.C., Nimkam, S., New, M.I. (2001) Apparent mineralocorticoid excess. *Trends Endocrinol. Metab.* **12**, 104–111.

Zhu, J., McKeon, F. (2000) Nucleocytoplasmic shuttling and the control of NF-AT signaling, *Cell. Mol. Life Sci.* **57**, 411–420.

Zolnierowicz, S., Bollen, M. (2000) Protein phosphorylation and protein phosphatases, *EMBO J.* **19**, 483–488.

Endocytosis

Hadiya A. Watson[1], *Mark Von Zastrow*[2], *and Beverly Wendland*[1]
[1] *The Johns Hopkins University, Department of Biology, Baltimore, MD, USA*
[2] *University of California San Francisco, Department of Psychiatry, San Francisco, CA, USA*

1 Overview of Endocytosis: Common Themes 184

2 Internalization 186
2.1 Clathrin-mediated Endocytosis 186
2.2 Structure of the Clathrin Coat 188
2.3 Other Protein Components of the Clathrin Coat 188
2.4 Mechanism of Endocytic Cargo Selection 190
2.5 Mechanism of Endocytic Coat Formation 191
2.6 Mechanism of Membrane Curvature and Fission 192

3 Clathrin-independent Endocytosis 193
3.1 Macropinocytosis 193
3.2 Caveolae 194

4 Vesicular Trafficking 194
4.1 Uncoating of CCVs 194
4.2 PM to Cytosol 195
4.3 Cytosol to Early Endosome 195

5 The Early Endosomal System 197
5.1 Sorting Endosomes 197
5.2 Recycling Endosomes 198
5.2.1 Other Proteins Implicated in Recycling 198
5.3 Transport from Early Endosomes 199

6 Endocytosis in Polarized Epithelial Cells 200
6.1 Trafficking Pathways 200

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

6.2 Regulation of Endocytosis in Polarized Epithelia 202
6.3 Transcytosis: the IgA Example 202

7 MVBs and the Late Endosomal System 202
7.1 Late Endosomes 203
7.2 Sorting into MVBs 203

8 Last Stop: the Lysosome 205
8.1 The Lysosome Biogenesis Controversy 206
8.2 Rab7 and Endosome–Lysosome Fusion 207
8.3 Hydrolases, Lamps, and Transporters 207
8.4 Secretory Lysosomes 208

9 Parasites and the Endocytic Pathway 208

10 Exosomes 209

11 Endocytosis and Signaling 210
11.1 Regulated Endocytosis of Signaling Receptors 210
11.2 Functional Consequences of Signaling Receptor Endocytosis 210
11.2.1 Receptor Downregulation 210
11.2.2 Receptor Resensitization 211
11.2.3 Signaling from Endosomes 211

12 Some Diseases of the Endocytic Pathway 211
12.1 Hypercholesterolemia 212
12.2 Defects in Endosomal Acidification and Proximal Tubulopathies 212
12.3 Lysosomal Dysfunction 213
12.4 The Cancer Link 213

13 Conclusion 214

Acknowledgments 215

Bibliography 215
Books and Reviews 215
Primary Literature 216

Keywords

Endocytosis
The process by which cells move extracellular molecules and portions of their own plasma membrane into the cell interior (cytosol). This occurs through formation of a small invagination of the plasma membrane into a bud, which ultimately separates to form an endocytic vesicle in the cytoplasm.

Clathrin
A large protein complex composed of three heavy and three light chains; it provides a coat for a class of endocytic vesicles.

Caveolae
Small flask-shaped invaginations on the plasma membrane that are rich in cholesterol, sphingolipids, and the protein caveolin. These structures are presumed to be relatively static and comprise a major clathrin-independent endocytic pathway.

Adaptor Protein
A class of protein that plays a critical role in connecting the plasma membrane and specific membrane proteins with components of the endocytic machinery and clathrin to ensure efficient endocytosis of particular molecules.

Endosome
An intermediate compartment of the endocytic pathway that is the destination of endocytic vesicles. Sorting events occur at endosomes that dictate the next destination of each protein or molecule that has been internalized.

Multivesicular Body
A late endosome in which sorting events have occurred, which result in the formation of small vesicles within the endosome; these vesicles generally contain membrane proteins that will be degraded upon delivery to the lumen of the lysosome.

Lysosome
An organelle highly enriched in hydrolases that degrade lipids, proteins, carbohydrates, and nucleic acids into their monomer components for reuse in the cell. Lysosomes are frequently the final destination of endocytosed material.

Interaction with the environment is crucial for the survival of cells, through communication with one another via signaling molecules (e.g. hormones) and nutrient acquisition. One major problem a cell must overcome is how to *selectively* carry out exchanges with the external environment and still maintain an internal composition that is chemically distinct. The plasma membrane (PM) makes this selectivity possible. While water and small molecules can cross the lipid bilayer by diffusion or transport through the lipid bilayer, larger macromolecules enter a cell by a different mechanism. Endocytosis is the process by which cells internalize extracellular material by forming intracellular membrane-bound vesicles originating from the PM, and subsequently sort the vesicle contents to the appropriate destination. Internalized material includes membrane proteins, PM lipids, and extracellular fluid, and may also include other cells or apoptotic cell fragments,

parasites, viruses, nutrients, or activated PM signaling complexes. The critical importance of endocytosis to such a wide variety of physiological processes has made it a focus of intensive investigation by scientists with diverse backgrounds and interests. Indeed, studies of the endocytic pathway are among the most highly interdisciplinary and methodologically innovative in all of modern biology, and the field is currently in a state of exciting and rapid advance. In this chapter, we attempt to summarize our current understanding of the endocytic pathway, focusing on recent progress in elucidating fundamental mechanisms of endocytic membrane trafficking and highlighting selected examples of their diverse physiological functions.

1 Overview of Endocytosis: Common Themes

A generic overview of the endocytic pathway is depicted in Fig. 1. Endocytosis begins at the cell surface with the deformation and vesiculation of a region of the PM into which various membrane constituents (cargo) are sequestered.

The newly formed vesicle is then transported to and fuses with a network of organelles called *early endosomes* (EE), where cargo destined for degradation is sorted out from material to be recycled. Recycling of material to the PM can occur from either the sorting endosome or the recycling endosome. Cargo destined for degradation is delivered to late endosomes (LE) (which intersect with the biosynthetic pathway) and then to lysosomes where they are degraded.

The endocytic pathway comprises a complex network of membrane trafficking events consisting of repeated cycles of membrane deformation, coat assembly and disassembly, vesicular transport, and membrane fusion (Fig. 2). Each of these events requires a tightly regulated interplay among numerous protein components. From the recent structural and functional characterization of various proteins involved in endocytosis have emerged several common themes in endocytic trafficking.

One major theme is that many endocytic proteins are *multimodular*; that is, they contain multiple binding sites – many of which are of low affinity – for other proteins or lipids, thus coordinating complex cross-linking events that stabilize a network or web of protein interactions. Among the many binding domains identified, lipid binding modules have received a great deal of recent attention, especially those that bind phosphoinositides localized at the PM, endosomes, lysosome, and Golgi. These proteins include both lipid modifying enzymes and proteins that serve as linkers between membranes and coat constituents.

A second theme in endocytosis is the compartmentalization of endocytic organelles into specific, biochemically distinct domains that underlie the selective sorting of proteins and lipids. Compartmentalization is provided by differential lipid sorting into tubules and large or small vesicles, and by specific localization of members of a group of regulatory molecules known as Rab GTPases, which temporally and spatially coordinate vesicular transport and docking/fusion events throughout the secretory and endocytic pathways. To date, 12 Rab proteins have

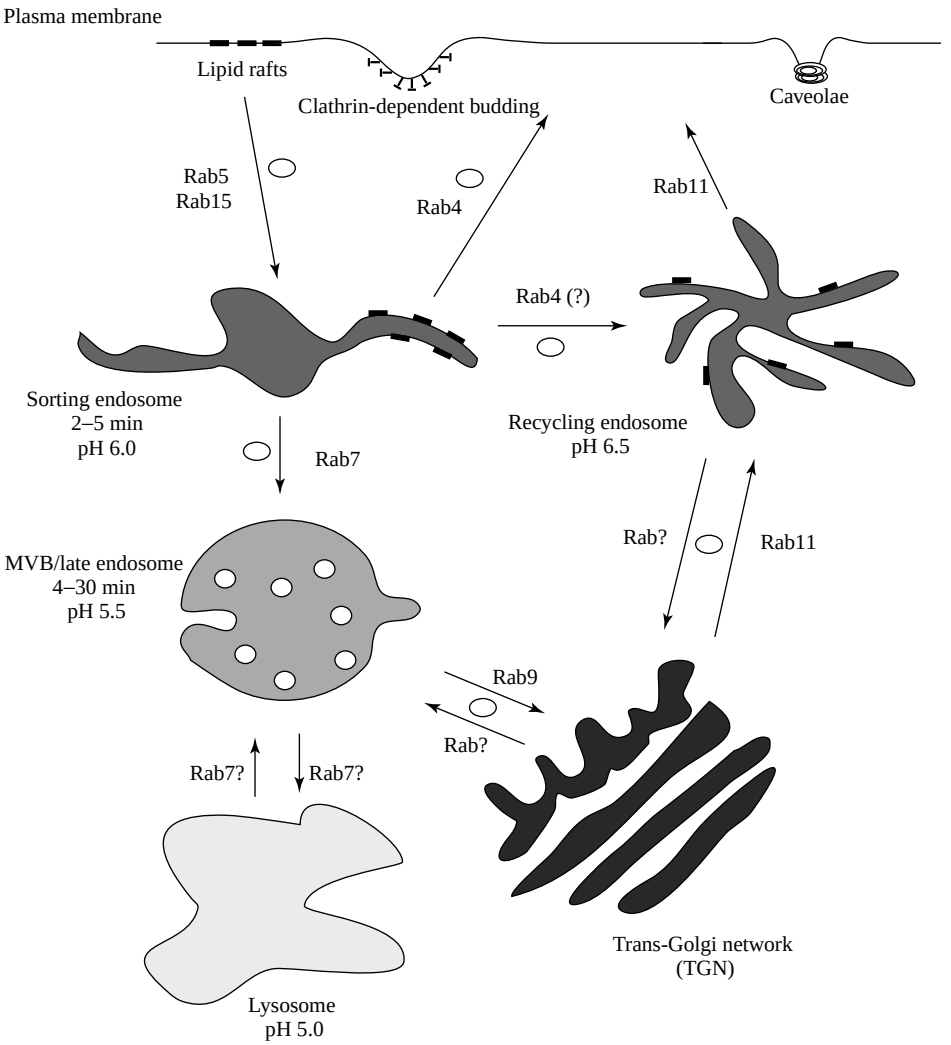

Fig. 1 The endocytic pathway. Endocytosis begins at the PM where membrane invaginations containing cargo give rise to vesicles, which fuse with a network of organelles called *early endosomes*. Here, cargo destined for degradation is sorted out from material to be recycled. Recycling of material to the PM can occur from either the *sorting endosome* or the *recycling endosome*. Cargo destined for degradation is delivered to *late endosomes* (which intersect with the biosynthetic pathway) and then to *lysosomes* where they are degraded.

been localized to the endocytic pathway; some of these are depicted in Fig. 1 (see also Table 1). Rab GTPase activity and membrane localization are tightly regulated by various accessory proteins.

Ubiquitylation modifications have recently been implicated in facilitating various events in the endocytic pathway and thus represent a third theme. At the early steps of endocytosis, monoubiquitylation

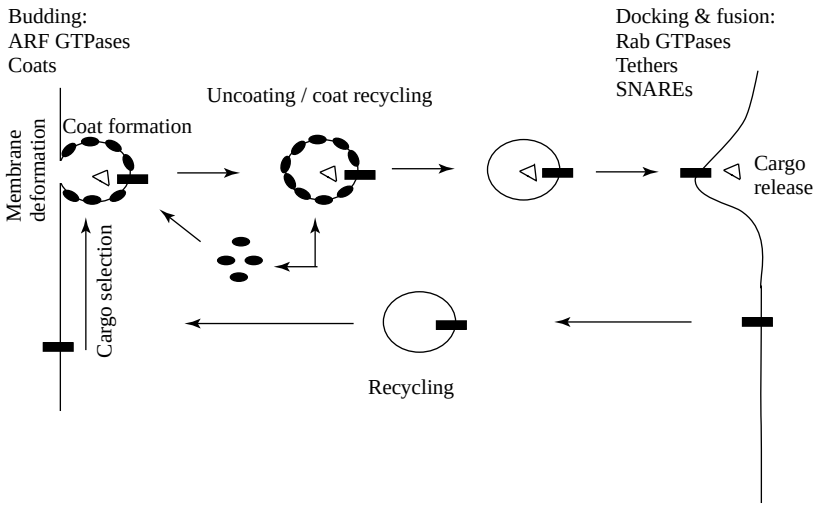

Fig. 2 Membrane trafficking themes in endocytosis. The endocytic pathway comprises a complex network of membrane trafficking events consisting of repeated cycles of cargo selection/membrane deformation, coat assembly and disassembly, vesicular transport, and membrane fusion (based on a figure from Mellman, I., Warren, G. (2000). The road taken: past and future foundations of membrane traffic, *Cell*. **100**, 99–112).

of the cytosolic tail of certain membrane receptors allows them to be recognized by and bind endocytic machinery components. Later in the pathway, ubiquitylation targets certain membrane proteins to the luminal vesicles of LEs. Recent studies have shown that some endocytic machinery components themselves receive covalent ubiquitin modifications, though the function of this remains to be elucidated.

A fourth and final common theme in endocytosis is its relationship to the cytoskeleton. The actin cytoskeleton plays a fundamental role in coordinating the internalization step of endocytosis, as well as in vesicular transport from the PM to EE. The precise role(s) for actin in these events remains to be determined; however, it is clear that a dynamic cycle of polymerization/depolymerization of actin filaments is essential. Microtubules also play important roles, especially at later stages of the endocytic pathway and in polarized epithelial cells, although the microtubule-based motor proteins involved in many of these pathways are not yet known.

This chapter will summarize our current understanding of the cellular and molecular mechanisms of endocytosis, with primary focus on mammalian and yeast systems, and will highlight selected examples illustrating the diverse physiological functions of endocytosis.

2 Internalization

2.1 Clathrin-mediated Endocytosis

The best characterized mechanism of endocytosis is mediated by clathrin-coated pits and vesicles. These structures were first identified by their characteristic

Tab. 1 Proteins involved in endocytosis.

Protein	***Binding properties/function***
I. Cargo recruitment/coat assembly	
Clathrin	Coat protein; driving force for vesicle formation
AP-2	Binds lipids, clathrin, cytosolic tails of certain receptors, accessory factors (e.g. epsin, Eps15); an adaptor protein
Epsin	Binds lipids, clathrin, AP-2, ubiquitin, and Eps15; an accessory protein
Eps15	Binds Epsin, AP-2, ubiquitin, SH3-domain proteins; an accessory protein
AP180/CALM	Binds lipids, clathrin; promotes assembly of clathrin cages
Synaptotagmin	Membrane protein, interacts with SNARE machinery; may serve as "docking site" for clathrin-coat formation
II. Membrane curvature and fission	
Amphiphysin	Binds lipids, clathrin, AP-2, endophilin, dynamin, and synaptojanin; recruits and coassembles with dynamin at bud neck
Endophilin	Binds lipids, synaptojanin, and dynamin; a lysophosphatidic acid acyl transferase that induces membrane curvature
Dynamin	Binds lipids, endophilin, amphiphysin; GTPase that mediates fission of vesicles from the PM
III. Coat disassembly and vesicular transport	
Hsc70	Molecular chaperone; disrupts clathrin–clathrin interactions, disassembles triskelia
Auxilin	DnaJ domain family member; binds clathrin and AP-2; recruits and stimulates Hsc70
Synaptojanin	Binds lipids, amphiphysin, Eps15, and endophilin; a phosphoinositide-5 phosphatase
Dynamin	Interacts with actin cytoskeleton components to mediate vesicular propulsion from the PM
Arp2/3	Stimulates actin nucleation; involved in vesicular movement
IV. Rabs in endosomal sorting	

Protein	***Organelle***	***Function***
Rab4	EE and RE	Recycling from EE and RE
Rab5	PM, vesicles, EE,	PM to EE
Rab7	LE/lysosomes	LE to lysosome (?)
Rab9	LE	LE to TGN
Rab11	RE	RE to PM
Rab15	EE and RE	EE to lysosome; RE to PM
Rab17	ARE	Apical recycling in epithelial cells
Rab25	ARE	Apical recycling in epithelial cells

Note: PM: plasma membrane; EE: early endosome; RE: recycling endosome; LE: late endosome; TGN: trans-Golgi network; ARE: apical recycling endosome.

appearance, observed in transmission electron micrographs of osmium-stained tissue sections, and were subsequently isolated from tissue homogenates using subcellular fractionation techniques. The identification of clathrin heavy chain as the major protein component of the dense membrane coat associated with these endocytic vesicles laid the groundwork for increasingly sophisticated biochemical and structural studies, which continue to this day.

Clathrin-coated vesicles (CCVs) are short-lived intermediates ($t_{1/2} < 2$ min) that are produced in a complex process involving the formation of clathrin-coated invaginations (or coated pits) in the PM, scission of these invaginations from the PM to form endocytic vesicles, and subsequent dissociation of the clathrin-containing coat structure. The "primary" endocytic vesicles thus formed fuse rapidly with other uncoated endocytic vesicles and with the limiting membrane of a heterogeneous population of larger membrane structures, called *early endosomes*, which are located in the peripheral cytoplasm. The coat components liberated into the cytoplasm during the uncoating process are thought to recycle to the PM, where they can participate in subsequent rounds of endocytosis (see Fig. 2).

2.2 Structure of the Clathrin Coat

The structural backbone of the clathrin coat is a "triskelion" composed of three radially arranged spokes composed of clathrin heavy chains (~190 kDa) and three associated light chains (20–25 kDa) (Fig. 3). The aminoterminal domain of clathrin heavy chain forms a globular region at the periphery of each spoke. The carboxyterminal region of each clathrin heavy chain self-assembles to form the central hub of the triskelion structure. One light chain polypeptide binds to the middle portion of each heavy chain. The heavy chains contribute to the major structural integrity of the triskelion. The light chains are thought to serve primarily a regulatory function in the process of clathrin assembly and disassembly. The characteristic hexagonal and pentagonal lattice structure of the clathrin coat, which is elegantly revealed by rapid freeze-etch electron microscopy, is composed of assemblies of interlocking triskelia. Under certain conditions, purified clathrin can form similar latticelike structures spontaneously *in vitro*, suggesting that the clathrin assembly might contribute to the energetics of endocytosis and that the clathrin assembly process must be highly regulated *in vivo*.

2.3 Other Protein Components of the Clathrin Coat

There is a long and growing list of proteins known to associate with the clathrin-coat structure (Table 1). The first such proteins to be identified were the assembly proteins, so-named for their ability to promote the assembly of purified clathrin into lattices *in vitro*, suggesting that they regulate coat formation *in vivo*. Assembly protein complexes termed AP-1 and AP-2 are present in stoichiometric amounts relative to clathrin in purified coated vesicles, but these proteins are segregated in functionally distinct subsets of coated vesicles. AP-2 is associated with endocytic coated vesicles formed at the PM. In contrast, AP-1 is associated with coated vesicles formed at the Golgi apparatus, which mediate transport to the endolysosomal system. AP-1 and AP-2

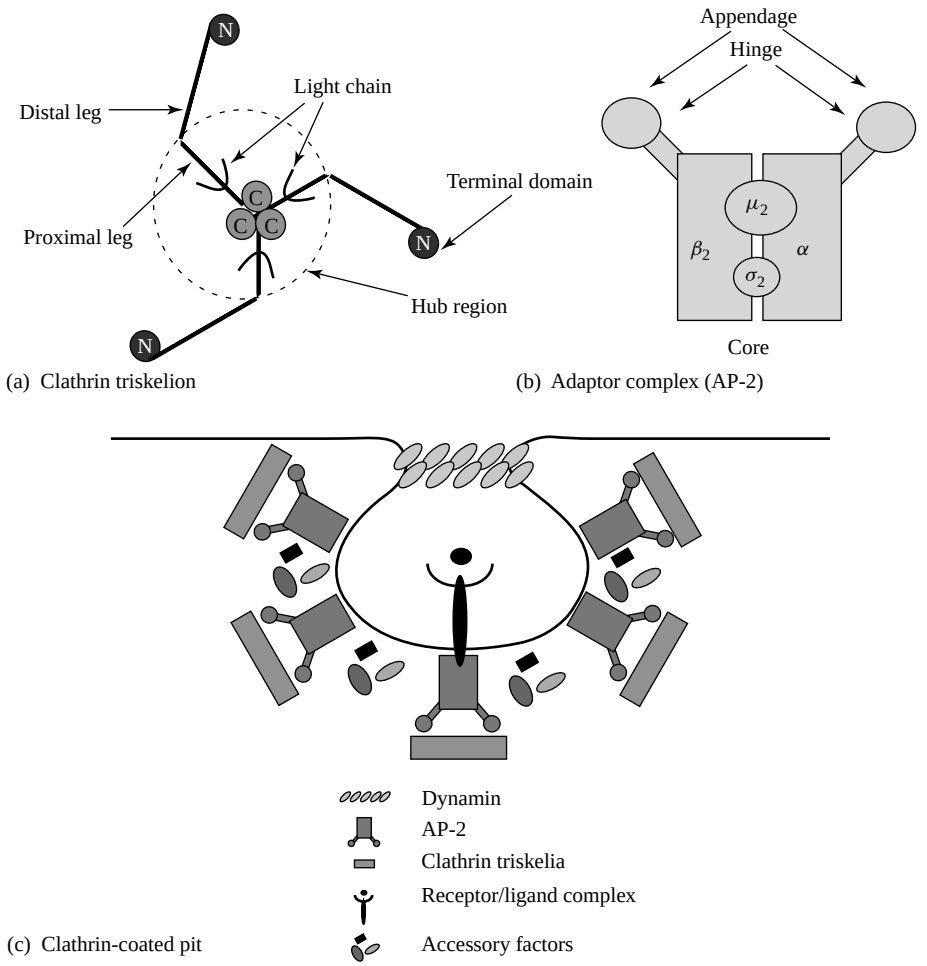

(a) Clathrin triskelion

(b) Adaptor complex (AP-2)

(c) Clathrin-coated pit

Fig. 3 Clathrin-coat components. (a) Structure of a clathrin triskelion. The clathrin triskelion is composed of three radially arranged spokes composed of clathrin heavy chains (~190 kDa) and three associated light chains (20–25 kDa). The aminoterminal domain of clathrin heavy chain forms a globular region at the periphery of each spoke. The carboxyterminal region of each clathrin heavy chain self-assembles to form the central "hub" of the triskelion structure. One light chain polypeptide binds to a central portion of each heavy chain. (b) Structure of AP-2. AP-2 is a heterotetrameric complex consisting of two large subunits (α and β_2; ~100 kDa), one medium subunit (μ_2; ~50 kDa) and one small subunit (σ_2; ~25 kDa). The N-termini of the α and β_2 subunits fold into a compact core structure, and the C-termini form an exposed appendage (or "ear") domain, separated from the core by a flexible hinge region. (c) Simplified diagram of a clathrin-coated pit. Studies indicate that AP-2 is located in a layer between the coated vesicle membrane and the surrounding lattice of assembled clathrin triskelia, suggesting that AP-2 physically links the clathrin lattice to the endocytic membrane. Dynamin localizes to the bud neck to promote fission. Various other accessory factors are depicted.

share a similar heterotetrameric structure, with two large subunits (~100 kDa), one medium subunit (~50 kDa), and one small subunit (~25 kDa).

The subunits of AP-2 are termed α and β_2 (large), μ_2 (medium), and σ_2 (small). Structural and biochemical studies indicate that the aminoterminal portions of the α, β_2, and μ_2 subunits and all of the σ_2 subunits are folded into a compact core (or "brick") structure; the carboxyerminal portion of the α- and β_2-subunits forms an exposed appendage (or ear domain) separated from the core by a flexible hinge region, while the carboxyerminal portion of the μ_2 subunit presents a cargo-binding interface (Fig. 3). The hinge region of the β_2-subunit interacts with clathrin. The core regions are thought to interact with proteins that control the subcellular localization (i.e. PM or Golgi) of APs, and the appendage regions are thought to interact with a number of other proteins that regulate coat formation or function (see Sect. 2.4). Structural studies indicate that AP-2 is located between the coated vesicle membrane and the surrounding lattice of assembled clathrin triskelia, suggesting that AP-2 physically links the clathrin lattice to the endocytic membrane. The AP-2 α-subunit is thought to bind phosphoinositides in the lipid bilayer involved in controlling the dynamics of coated pit assembly at specific sites on the plasma membrane. Recent structural studies suggest that assembled AP-2 contains two phosphoinositide binding sites.

Two other protein complexes, termed AP-3 and AP-4, share sequence homology and a similar heterotetrameric organization with AP-1 and AP-2. However, AP-3 and AP-4 promote clathrin lattice formation weakly or not at all *in vitro*, and whether or not they interact with clathrin coats *in vivo* is controversial. A group of proteins called GGAs (Golgi localized, γ-ear-containing Arf-binding proteins) have sequence homology to classical APs, but exist as monomers and associate with CCVs much less stably than AP-1 or AP-2. A variety of additional proteins associated with varying affinities to clathrin-coated membranes, including dynamin, amphiphysin, synaptojanin, endophilin, intersectin, Eps15, epsin, AP180/CALM, and nonvisual (β-) arrestins. These proteins are proposed to serving a variety of regulatory functions in endocytic coat assembly, membrane dynamics, and cargo selection. In many cases, these proteins associate with several other protein components of the clathrin coat via multiple, low-affinity interactions. A number of coat-associated proteins, in addition to AP-2, also bind to phosphoinositides. The complexity of protein–protein and protein–lipid interactions makes it challenging to define the precise role played by any particular protein interaction and suggests that certain coat-associated proteins may serve multiple functions, and some functions may be performed by multiple proteins. *In vitro* reconstitution experiments and studies in genetically tractable systems like yeast, *Drosophila*, and *Caenorhabitis elegans* will be needed to dissect the physiological role for many of these factors.

2.4 Mechanism of Endocytic Cargo Selection

Clathrin-coated pits exhibit considerable selectivity in their endocytosis of specific membrane cargo. Early insight into the biochemical basis of endocytic cargo selection came from studies of patients suffering from familial hypercholesterolemia (discussed in Sect. 12.1). A severe form

of this disease was associated with a mutation of a tyrosine-containing sequence present in the cytoplasmic tail of the low-density lipoprotein (LDL) particle receptor. Mutation of this tyrosine residue both impaired LDL uptake and changed the distribution of the LDL receptors (LDLRs) from coated pit localization to a uniform distribution in the PM. Similar tyrosine-containing endocytic signals have been identified in a number of other rapidly endocytosing membrane proteins. Additional endocytic signals, such as the dileucine motif, are biochemically distinct from tyrosine-containing sorting signals but also promote the concentration of specific proteins in clathrin-coated membranes. Many endocytic signals, including tyrosine and dileucine-containing sequences, function by binding to the AP-2 assembly protein. Analogous sequences bind to AP-1, and thereby promote packaging of specific membrane cargo into Golgi-derived coated vesicles. For this reason AP-1 and AP-2, named initially for their ability to promote clathrin assembly, are also commonly called *adaptor* proteins, because they link specific cargo to the forming clathrin coat. Phosphorylation of the μ_2-subunit of AP-2 by the adaptor-associated kinase AAK1, enhances binding of tyrosine-based endocytosis signals, indicating that the interaction of adaptor proteins with specific membrane cargo can be regulated.

A number of other clathrin-associated proteins, in addition to AP-2, can serve as endocytic adaptors for recruiting specific membrane cargo into CCVs. For example, nonvisual (β-) arrestins, which bind to G-protein-coupled signaling receptors after ligand-induced activation and phosphorylation, promote endocytosis of receptors via clathrin-coated pits by also binding to clathrin or AP-2 as well as to phosphoinositides. Another clathrin-binding protein, epsin, contains ubiquitin-interacting motifs (UIMs), which may recognize ubiquitylated cargo, as well as Asn-Pro-Pre (NPF) containing sequences that bind to Eps15 homology (EH) domains present in various coat-associated proteins such as Eps15. The currently favored models suggest that epsins function as endocytic adaptors for certain membrane cargo.

2.5 Mechanism of Endocytic Coat Formation

The ability of purified clathrin and AP-2 to self-assemble into extended lattices *in vitro* suggested that clathrin-coat formation *in vivo* is driven by polymerization. However, the initiation of clathrin-coat assembly appears to be highly regulated *in vivo*, as a large fraction of clathrin exists at steady state as "free" triskelions in the cytoplasm, while clathrin coats transiently form and dissociate in specific regions of the PM. The factors controlling the temporal and spatial specificity of clathrin-coat formation are not well understood but are generally believed to involve the initiation of clathrin polymerization at defined "docking sites" in the PM. Synaptotagmin I/II are integral membrane synaptic vesicle proteins that bind to AP-2 via a lipid and Ca^{2+}-binding C2b domain, and are proposed to function as docking sites for clathrin/AP-2 coat formation in neurons to facilitate synaptic vesicle membrane recycling. However, synaptotagmin I/II expression is restricted to neural cell types, and it is not known if a ubiquitously expressed synaptotagmin (many candidates exist) provides an analogous docking site for clathrin-coat initiation in nonneural cells. Another possible type of docking site may involve membrane lipids

such as PtdIns(4,5)P_2, which may be generated in restricted regions of the PM; this phosphoinositide binds a number of coat-associated proteins, including AP-2, dynamin, and epsin.

Although they bind to AP-2, classical endocytic signals present in cargo proteins are not likely to act as docking sites as a primary mechanism for localizing clathrin-coat formation. Overexpressing cargo proteins typically does not affect the density of clathrin-coated pits in the PM; the localization of these integral membrane proteins appears to be determined by the localization of clathrin-coat formation and not vice versa. Nevertheless, membrane cargo that presents tyrosine-containing endocytic signals can promote the association of synaptotagmin with AP-2, suggesting that cargo proteins may contribute to a more complex mechanism regulating coat formation involving multiple protein interactions.

Arf-family GTPases regulate the formation of Golgi-derived (clathrin/AP-1-associated) coated vesicles by a mechanism that may involve Arf-dependent activation of phospholipase D and/or recruitment of GGA proteins, which have been proposed as functioning both in coat formation as well as in cargo recruitment at the Golgi. An analogous GTPase regulating endocytic clathrin/AP-2 coat formation *in vivo* has not been identified, although recent data suggest that Arf6 stimulates the PtdIns(4,5)P2-synthesizing PI5-kinase enzyme that in turn leads to recruitment of AP-2. Endocytic coat formation also appears to be regulated by the phosphorylation state of AP-2. Unassembled AP-2 in the cytosol is phosphorylated on the β_2-subunit. Dephosphorylation of this subunit is thought to occur specifically at the PM and promotes AP-2 assembly with clathrin.

2.6 Mechanism of Membrane Curvature and Fission

Endocytosis via clathrin-coated pits involves the formation of deep membrane invaginations followed by scission of the bilayer to form the nascent clathrin-coated vesicle. These processes of membrane invagination and scission represent distinct steps in a multistep endocytic mechanism. Positive curvature (bulging inward toward the cytoplasm) of invaginated membranes is generally thought to be promoted by self-assembly of the clathrin lattice itself.

A mechanism for generating endocytic membrane curvature was proposed recently on the basis of structural studies of the clathrin coat-associated protein epsin. An aminoterminal portion of epsin was observed to fold into an amphipathic alpha-helix specifically when bound by phosphoinositide. This helical structure inserts into the lipid bilayer and can mediate membrane tubulation *in vitro*. In this case, however, epsin has been suggested to play a role in inducing phosphoinositide-regulated positive curvature to initiate formation of coated pits.

A number of models have been proposed for the mechanism mediating the negative membrane curvature present at the edges of invaginated coated pits, where dynamin, amphiphysin, and endophilin are concentrated. It is suggested that these proteins induce negative curvature by mechanically constricting the membrane in this region. This hypothesis is supported by the ability of these proteins to self-assemble and cause tubulation of liposome membranes *in vitro* and the requirement of endophilin for endocytic membrane invagination *in vivo*. Additionally, purified endophilin possesses

lysophosphatidic acid acyl transferase activity, suggesting that endophilin induces membrane curvature by modifying phospholipids at the site of neck constriction. The extent to which the bilayer-deforming properties of endophilin depend on this enzymatic activity has been debated.

Studies of *shibire* mutant *Drosophila* revealed that they had a mutation in the large GTPase dynamin, which is required for the reuptake of PM from the nerve terminal. Subsequent studies established an essential role of dynamin in endocytic membrane scission in various mammalian cell types. Dynamin self-assembles into rings around the neck of a clathrin-coated pit, suggesting that dynamin may function as a GTP-regulated force-generator to mechanically constrict and perhaps shear the invaginated, coated pit from the PM. Whether dynamin functions as a direct force-generator or instead functions to regulate a distinct mechanochemical effector remains unresolved. It is also not known if a dynamin-related GTPase is involved in endocytosis in yeast, although actin is essential for fission of endocytic membranes in yeast. Various roles have been proposed for actin in clathrin-mediated endocytosis occurring in mammalian cells, but whether or not actin is essential for endocytosis in mammals has remained controversial. A recent study using evanescent field microscopy to image coated pit formation in living mammalian fibroblasts revealed a transient accumulation of polymerized actin around the coated pit just after dynamin recruitment and just prior to vesicle excision from the PM. These data suggest that dynamin coordinates local actin polymerization to provide the mechanical force necessary to propel nascent coated vesicles away from the PM into the cytoplasm.

3 Clathrin-independent Endocytosis

The existence of clathrin-independent endocytic processes was first indicated by morphological observations of endocytic structures not associated with a characteristic electron-dense membrane coat, and has been confirmed using a variety of manipulations that differentially inhibit clathrin-dependent and -independent endocytic mechanisms with varying degrees of specificity. Interestingly, there appears to be regulatory communication between distinct pathways that allow adaptation and upregulation of one pathway when the other is blocked. To further complicate matters, there is evidence of the existence of multiple clathrin-independent endocytic pathways. While none of these mechanisms is understood as well as clathrin-mediated endocytosis, macropinocytosis and endocytosis of caveolae have been characterized in some detail.

3.1 Macropinocytosis

Macropinocytosis is a prominent endocytic pathway in macrophages and can also be induced by growth factors or infectious agents in other cell types. Macropinocytosis is so-named because it is characterized by the uptake of endocytosed cargo in vesicular structures that are typically much larger (up to 5 μm in diameter) than CCVs (<100 nm). Macropinosomes generally form from actin-rich ruffled regions of the PM and, like phagocytosis, macropinocytosis is strongly inhibited by agents such as cytochalasin that disrupt actin polymerization. This is in contrast to the limited or variable effects of such agents on endocytosis involving clathrin-coated pits or caveolae. Macropinocytosis is not inhibited

by overexpression of dominant-negative mutant versions of proteins (dynamin, clathrin, and Eps15) involved in clathrin-mediated endocytosis. The small GTPase Arf6, which modulates actin dynamics at the cell cortex, also appears to regulate the formation of macropinosomes. This activity of Arf6 appears to further distinguish the mechanism of macropinocytosis from clathrin-mediated endocytosis, although Arf6 can also modulate clathrin-mediated endocytosis in some cases.

3.2 Caveolae

Caveolae is a term used to describe small (50–80-nm diameter) invaginations of the PM that were observed initially in electron microscopic studies of endothelial cells and have since been recognized in several other cell types. These structures are enriched in glycosphingolipids and cholesterol and possess a proteinaceous membrane coat that is composed primarily of the cholesterol-binding protein caveolin. Caveolin is an integral membrane protein, unlike clathrin-coat components, which are peripherally associated with the membrane, and is required for the formation and structural integrity of caveolae. While caveolae were initially proposed not to undergo classical endocytosis, but rather to serve as a conduit across closely opposed endothelial membranes, there is increasing evidence that caveolae can indeed undergo dynamin-dependent fission from the PM. However, recent Fluorescence Recovery After Photobleaching (FRAP) analysis using GFP-tagged caveolin indicates that caveolae are highly immobile structures and are thus not likely to play a significant role in constitutive endocytosis. Endocytosis via caveolae appears to be primarily ligand/signal-induced and is a major internalization pathway for certain toxins and viruses. Interestingly, during internalization of the SV40 virus, caveolae may mediate an alternative endocytic pathway leading directly to the ER (endoplasmic reticulum) and bypassing conventional endosomes.

4 Vesicular Trafficking

4.1 Uncoating of CCVs

The events occurring after the initial internalization step have been characterized most thoroughly for endocytosis mediated by CCVs. CCVs must be uncoated to expose proteins important for fusion with EEs, and to allow for recycling of coat components for further rounds of internalization (Fig. 2). Extensive *in vitro* studies have demonstrated that uncoating of CCVs is mediated primarily by two proteins: the molecular chaperone, ATPase Hsc70, and auxilin, a clathrin-binding protein belonging to the Dna J family. Auxilin 1 is neuron specific, while auxilin 2 (aka GAK, or Cyclin G-associated Kinase), a ubiquitously expressed isoform discovered later, contains an AAK1-related protein kinase domain of unknown function. Yeast and *C. elegans* homologs of auxilin have also been identified. In the general mechanism for Hsc70/Auxilin-mediated CCV uncoating, auxilin binds assembled clathrin (and AP-2 in some cases), recruits Hsc70 via its J domain, then stimulates Hsc70's ATPase activity. Hsc70 disrupts clathrin–clathrin interactions to uncoat CCVs, and remains bound to disassembled clathrin, possibly to initiate reassembly, while auxilin is released and recycled. *In vitro* studies

using dominant-negative forms of Hsc70 suggest that it assists in other clathrin-dependent functions including endosomal trafficking. The phosphoinositide-5 phosphatase, synaptojanin 1, also plays a role in CCV uncoating; its enzymatic activity (degrading PtdIns(4,5)P_2 to PtdIns4P) helps disrupt membrane-clathrin coat interactions.

4.2 PM to Cytosol

Clathrin-coated vesicles, after undergoing rapid, chaperone-dependent uncoating in the peripheral cytoplasm, become highly fusogenic. These "nascent" endocytic vesicles fuse rapidly with a pleiomorphic set of endomembranes called *early endosomes* (EEs). EEs are highly dynamic structures in which continuous membrane fusion and fission events can be observed using live-cell microscopy. A large volume of membrane traffic passes through EEs, and cargo molecules are efficiently sorted between divergent membrane pathways after entry into the EEs. Thus, it is believed that the membrane dynamics that occur in the early endocytic pathway are highly specific and regulated.

Shortly after a nascent endocytic vesicle is formed, it moves from the PM to the EEs through a thin region of peripheral cytoplasm that contains a dense network of actin, which may contribute to endocytic vesicle motility in the cortical cytoplasm via actin-based motors, such as myosin I and other myosins, as well as motility driven directly by asymmetric actin polymerization on the endosome membrane. Dynamin may also provide a link between scission of nascent vesicles and the asymmetric stimulation of Arp2/3-dependent actin polymerization for vesicle propulsion. Recent evidence also suggests roles for unconventional myosin VI in endocytosis.

4.3 Cytosol to Early Endosome

Once a nascent endocytic vesicle is in close physical proximity to the EE membrane, a number of mechanisms controlling the specificity and efficiency of the membrane fusion follow. Purified early endocytic membranes fuse efficiently with EEs *in vitro* ("homotypic" fusion), but fusion with later endocytic structures occurs with much lower efficiency ("heterotypic" fusion). A variety of experimental approaches, notably genetic studies of membrane trafficking in yeast and *in vitro* reconstitution of endocytic fusion in extracts prepared from mammalian cells and yeast, have revealed important roles in EE fusion played by tethers, Rab-family GTPases, SNAREs, and phosphoinositides.

N-ethylmaleimide sensitive factor (NSF) (Sec18p), the AAA-family ATPase-mediating disassembly of SNARE complexes is required for membrane transport from early to late endosomes. A number of SNARE proteins are localized selectively to specific populations of endosome membranes, suggesting an important role for SNAREs in controlling the specificity of endocytic membrane fusion. However, to date, the degree to which specific SNARE complex formation determines the selectivity of endocytic membrane fusion *in vivo* has not been determined. Moreover, there is considerable evidence that additional factors contribute to the specificity of endocytic membrane fusion, particularly the Rab family of small GTPases and their associated proteins.

The Rab5 GTPase is selectively localized to EEs and is thought to control the recruitment of a number of other important proteins to the endosome membrane. Mutations that impair GTP hydrolysis on Rab5 (and thus render the protein constitutively active) enhance endocytosis and produce enlarged endosomes. Inactive Rab5 inhibits homotypic fusion of EEs, whereas activated Rab5 promotes this process. Interaction cloning and affinity chromatography methods have identified multiple putative Rab5 effectors. Intriguingly, roles have also been observed for a complex of Rab5-GDI (guanine-nucleotide dissociation inhibitor) in the initial formation of CCVs *in vivo*.

The Rab5 effector, Rabaptin-5, binds to Rab5-GTP in solution and binds to endosomes in a Rab5- and GTP-dependent manner. Rabaptin-5 also binds to the Rab5 GEF Rabex-5, suggesting a synergistic mechanism by which Rab5 activation may be amplified on the endosome membrane. The Rabaptin/Rabex complex can also interact with GGA proteins in a bipartite manner that occludes GGA interaction with clathrin, suggesting that GGAs provide a link between cargo selection in Golgi-derived CCVs and tethering/fusion with endosomes after uncoating. Rabaptin-5 also binds to Rab4, a homologous GTPase that modulates recycling of membrane cargo from EEs to the PM (see Sect. 5); thus, Rabaptin-5 may define functionally distinct endosomal microdomains and coordinate distinct Rab-dependent membrane events involving communication between them. Another cytoplasmic protein that binds preferentially to Rab5-GTP is the early endosome antigen-1 (EEA1), a protein identified as an autoantigen in lupus patients and later shown to regulate endosomal traffic. EEA1 is recruited to the endosome membrane by Rab5-GTP, is essential for efficient fusion of EEs *in vitro* and can promote endosome fusion even in the absence of Rab5 or other cytoplasmic proteins when added to endosomes in high concentration. EEA1 is not detectably associated with nascent endocytic vesicles, yet it is required for efficient fusion of these vesicles with EEs *in vitro*, suggesting a role for EEA1 in determining the directionality of endocytic membrane transport.

Genetic studies of Vps34p in yeast identified a critical role for phosphoinositide 3-kinase in membrane trafficking to the vacuole. Pharmacological inhibitors of type I PI3K activity have profound effects on endocytic trafficking in mammalian cells. EEA1 contains an "FYVE"-type zinc-finger domain (named for homology to Fab1, YOTB, Vac1, and EEA1) that binds with high affinity to 3-phosphoinositides that are enriched in the EE membrane. Efficient recruitment of EEA1 to EEs *in vivo* requires both FYVE domain-mediated binding 3-phosphoinositides and binding of Rab5-GTP via a distinct pair of protein binding domains, leading to synergistic regulation of endosome fusion.

Another binding partner of EEA1 is the endosome-enriched t-SNARE (or Q-SNARE) syntaxin-13. EEA1 can be detected together with syntaxin-13 in a large complex that also contains Rab5 (and several other Rab5-binding proteins). Interestingly, a synthetic peptide that interferes with EEA1/syntaxin-13 binding inhibits EE fusion *in vitro*. These observations suggest that EEA1 may mediate Rab5 and phosphoinositide-dependent membrane fusion by influencing SNARE complex formation on the endosome membrane.

5 The Early Endosomal System

The endosomal system consists of membrane-bound compartments that receive, sort, and release cargo-containing vesicles throughout the endocytic pathway. Early endosomes are the single entry point and first sorting station in the endocytic pathway, and cargo reaches them within 2 to 5 min following internalization. These endosomes are so far known to comprise two subcompartments: sorting endosomes and recycling endosomes.

5.1 Sorting Endosomes

Internalized material is first delivered to the sorting endosome, also referred to as the *tubulovesicular endosome* or the *compartment of uncoupling receptor from ligand* (CURL). This compartment exhibits a high degree of homotypic fusion capacity and is tubulovesicular in morphology, consisting of thin tubules (50–60-nm diameter) that emanate from large vesicles (200–400-nm diameter). The mildly acidic lumen (pH 6.0) of the sorting endosomes promotes the dissociation of many ligands from their recycling receptors. Immediately following dissociation, recycling receptors are sorted away ($t_{1/2} = 2.5$ min) from their ligands and transported through the recycling pathway, while ligands and downregulated receptors are transported through the degradation pathway.

The membranes of sorting endosomes are organized into morphologically and biochemically distinct compartments that are believed to be essential for their sorting functions. For example, recycling receptors are often segregated into tubular regions, while their soluble ligands remain in the lumen of globular/vesicular regions. Regulation of this compartmentalization is thought to depend on the localization of specific molecular machinery. As described above, the Rab family of GTPases comprises a group of regulatory molecules that differentially localize to distinct subdomains of the organelles within the endocytic and exocytic pathways, recruiting specific effector molecules to each membrane domain they inhabit.

Sorting endosomes contain two other Rab GTPases: Rab15 and Rab4. Elferink and colleagues revealed a role for Rab15 in EE function. This GTPase colocalizes with Rab5 on sorting endosomes, and also localizes to recycling endosomes (see Sect. 5.2). Interestingly, the effects on endocytosis caused by overexpression of wild type and mutant forms of Rab15 are in direct contrast to those caused by Rab5 overexpression. Overexpression of constitutively active Rab15 (GTP bound) inhibits internalization of fluid phase marker, while overexpression of constitutively inactive Rab15 mutants (N12I, T22N) increases rates of both internalization and recycling pathways. At sorting endosomes, Rab15 may be an inhibitory GTPase, possibly by negatively regulating Rab5 function. How this regulation occurs has yet to be determined, since interactions between Rab15 and known Rab5 effectors have not yet been identified. Rab4 localizes to the tubular regions of sorting endosomes, which receive cargo within 5 min of internalization. Rab4 mediates "rapid" or direct recycling from the sorting endosome to the PM, which is distinct from the "slower" recycling route taken through the recycling endosomes. Since Rabaptin5 binds both Rab5 and Rab4, it has been suggested as coordinating the directing of cargo between the distinct sorting domains of the sorting endosome.

Differential sorting of lipids between EE subcompartments can occur independently of Rab GTPases. Elegant studies by Maxfield and colleagues showed that different lipid analogs, varying solely in the length and/or saturation of their hydrophobic tails, sorted to distinct endosomal compartments. Specifically, lipid analogs with longer (16-carbon) saturated tails were delivered to LEs, while those with shorter (12-carbon) saturated tails were mainly found in recycling compartments. This type of sorting would provide an energetically economical mechanism for sorting of certain types of molecules; however, as mentioned, protein–protein/protein–lipid interactions most likely provide additional modulation of these events.

5.2 Recycling Endosomes

In addition to membrane lipids, certain membrane receptors are also recycled, which allows for their reuse in further cycles of ligand binding and endocytosis. In fact, while the lysosomal targeting pathway is better characterized, more than 90% of endocytosed material is recycled to the PM. A classic example of a recycling membrane protein is the transferrin receptor (TfR). This receptor binds iron-bound transferrin at the cell surface, and following the release of iron in the mildly acidic sorting endosome, the iron-free TfR-transferrin complex is transported back to the PM. Owing to clustering in clathrin-coated pits, approximately 50% of TfR molecules on the cell surface can be internalized in less than 5 min, after which TfR is recycled.

Two recycling pathways have been identified: one in which the majority of endocytosed transferrin is returned to the PM directly from the sorting endosome through a "rapid" recycling pathway, and another in which the remaining transferrin accumulates in tubulovesicular structures located near the microtubule organizing center, prior to its return to the surface. These pericentriolar structures receive cargo within 2 to 5 min of internalization. They are labeled only with recycling receptors and not with lysosome-targeted receptors (or fluid phase markers) and are slightly more basic than the sorting endosome (pH 6.4–6.5), and were thus distinguished from the sorting endosome as recycling endosomes – also known as the *endocytic recycling compartment* (ERC).

In contrast to Rab4 (described above), Rab11 plays an important role in the "slower" recycling pathway to the PM and colocalizes with Tfn at pericentriolar recycling endosomes. A novel family of Rab11-interacting proteins called *FIPs* was identified recently, which share little homology outside of a conserved Rab11 binding domain, and appear to form mutually exclusive complexes with Rab11 *in vivo*. Unlike typical Rab effector proteins, FIPs do not rely on Rab11 for membrane binding, and it is proposed that these proteins function to recruit Rab11 to specific subdomains of the ERC.

5.2.1 Other Proteins Implicated in Recycling

RME1 A genetic screen in *C. elegans* identified Ce-Rme1 as an EH domain protein required for endocytosis and normal ERC morphology. Chinese hamster ovary (CHO) cells, GFP-tagged Ce-Rme1, and the mouse homolog m-Rme1 (also called *EHD-1*) each localize to the PM and transferrin-positive vesicles, but primarily concentrate at the perinuclear ERC. m-Rme1 with a mutation in the N-terminal nucleotide binding domain (G65R) no

longer localizes to the ERC, and instead is cytoplasmic, whereas a mutation near the EH domain (G429R) not only localizes to, but also alters, the morphology of the ERC. The G429R mutant also retards both transferrin recycling and recycling of the protein TGN38, which travels to the trans-Golgi network (TGN) from the PM through the ERC.

Rcy1 A visual screen of the viable deletion strains of *Saccharomyces cerevisiae* identified the protein Rcy1, which, through a fluorescence assay used to measure real-time recycling in yeast, was demonstrated to be required for recycling of PM lipids and proteins. Interestingly, Rcy1 contains an F-box motif, which is implicated in facilitating an interaction between target proteins and the ubiquitin ligase, SKP1/cullin/F-box (SCF). The precise connection between Rcy1, SCF, and recycling remains to be determined, but it is a compelling prediction given the multiple roles for ubiquitin in regulating endocytosis at many levels.

5.3 Transport from Early Endosomes

Transport of material along the endosomal pathway occurs in two ways. During *vesicular transport*, material is exchanged between relatively stable endosomal organelles in the form of intermediate vesicles. There is evidence to suggest that membrane protein transport from sorting endosomes to recycling endosomes occurs simply by default, following bulk lipid flow. However, numerous studies also provide convincing evidence to support tubulovesicular-mediated transport mechanisms. A recent study identified a sizeable pool of vesicles called *endocytic small vesicles* (ESVs) that derive from the EEs and are involved in recycling. At steady state, these vesicles labeled neither with sorting endosomal markers like Rab5 and EEA1, nor with the recycling endosome marker Rab11, contained Rab4, the v-SNARE cellubrevin, and recycling membrane proteins (e.g. TfR). The presence of Rab4 in ESVs is intriguing, since Sönnichsen et al. found that this GTPase preferentially localizes to tubular regions of the sorting endosome, where recycling material is proposed to segregate. A role for coat proteins in the formation of the ESVs has not been clear, and preliminary data suggested that ESV formation also may not involve Arf GTPase-mediated coat recruitment. ESVs may serve as transport vesicles for recycling material from the sorting endosome to either the recycling endosome or directly to the PM.

In contrast to distinct carriers such as ESVs that traffic from sorting to recycling endosomes, EE to LE traffic occurs via *maturation*, in which lysosomally targeted ligands remain associated with sorting endosomes, which eventually mature into later compartments through acidification and removal of material that is recycled to earlier compartments. *In vitro* studies by Gruenberg and colleagues showed a requirement for the coatomer protein complex (COPI) in the maturation of early to late endosomes via intermediate compartments (or perhaps, distinct transport intermediates) called *endocytic carrier vesicles* (ECVs). COPI, which is recruited by ARF1 GTPase, is better known for its functions in the early secretory pathway. COPI consists of seven proteins (α, β, β', γ, δ, ε, and ζ), and most but not all of these components are required for ECV formation. ECV biogenesis also appears to require endosomal acidification, and COPI/ARF1 association with EEs is pH dependent. Endosomal acidification

is mediated by a vacuolar H^+-ATPase. It is proposed that acidification controls EE functions by stimulating assembly of COPI coats, possibly through a specialized transmembrane pH sensor protein. The physiological consequences of defective endosomal acidification are discussed in Sect 12.

Other studies suggest that EEs may contain two distinct populations of clathrin coats: one involved in budding from tubular structures to form recycling vesicles, and the other, a flat clathrin coat involved in sorting cargo for lysosomal targeting. This finding illustrates the multiple roles a single protein can play in the endocytic pathway and the complexity of cargo sorting and transport mechanisms.

6 Endocytosis in Polarized Epithelial Cells

Polarized epithelial cells have biochemically distinct PM domains at their apical and basolateral surfaces that are separated by tight junctions that form a diffusion barrier between the surfaces. Since approximately 50% of a typical epithelial cell surface is endocytosed per hour, correct polarized sorting in the endocytic pathway is crucial for maintenance of cell polarity. Studies with Madin–Darby canine kidney (MDCK) cells confirmed that epithelial cells are able to internalize material from both apical and basolateral PM domains, which is directed to distinct sets of endosomal organelles. Endosomes found in polarized cells include apical early endosomes (AEE) and basolateral early endosomes (BEE), apical recycling endosomes (ARE), a distinct subapical compartment (SAC) in hepatocytes, and a common endosome (CE).

6.1 Trafficking Pathways

A current model for endocytic trafficking in polarized epithelial cells is depicted in Fig. 4, where specific Rab GTPases known to be associated with each compartment are noted. Basolaterally internalized material, including membrane and fluid, first enters the BEE (pH 5.8). Both recycling (e.g. TfR-Tf) and transcytosing (e.g. pIgR-IgA; see Sect. 6.3) materials are then transported to the CE, while material destined for degradation (e.g. fluid phase markers, LDL, and EGF (epidermal growth factor)) is transported to the LEs. The CE, also known as the *common recycling endosome*, has a tubular morphology and serves as a nexus between apical and basolateral membrane traffic. A role for clathrin and a class of epithelial cell-specific adaptor proteins has been demonstrated for the exit of basolateral recycling material from the CE.

Apically internalized material is first delivered to the AEE. From the AEE, material can travel to the ARE for recycling, to the CE for transcytosis, or to the LE/lysosome for degradation. Rab11 and its effector Rip11 have been implicated in apical recycling. Studies with MDCK cells show that the majority of apically internalized fluid phase markers is either recycled or transcytosed, which may be a reflection of the dependence of apical endocytosis on the GTPase ARF6 and its exchange factor, ARNO (ARF nucleotide binding site opener). In contrast, the majority of basolaterally internalized fluid phase marker accumulates in LEs and lysosomes. These differences in endocytic trafficking pathways among the apical and basolateral surfaces are essential for the maintenance of function and polarity in epithelial cells.

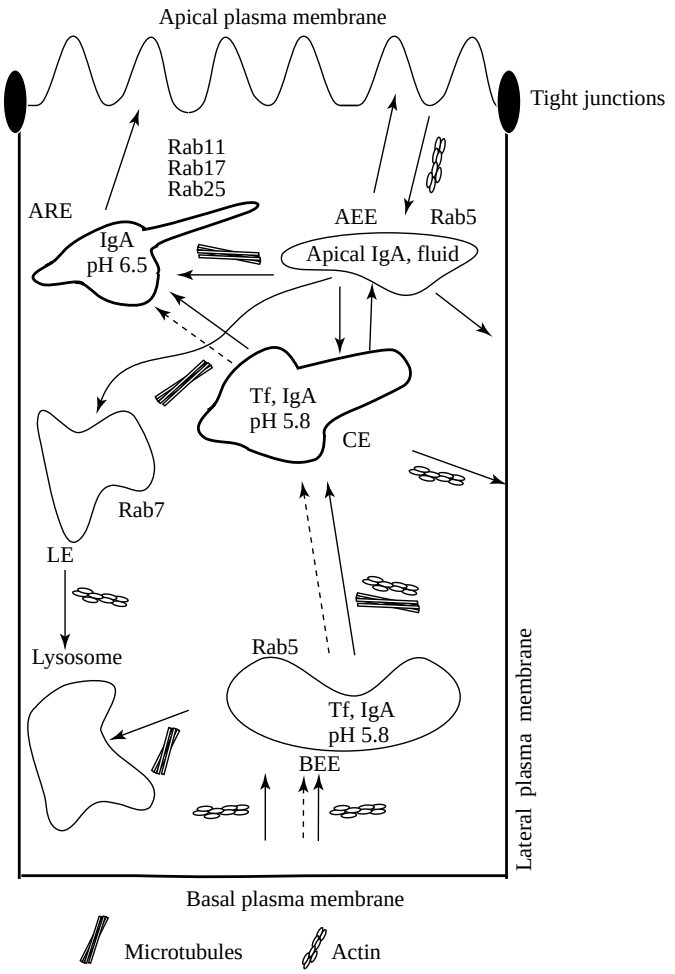

Fig. 4 Endocytic traffic in polarized epithelial cells. Internalization can occur either from the apical or basolateral surface. Internalized material is first delivered to either the AEE or BEE. From the AEE, material can recycle (via the ARE), transcytose (via the CE), or be delivered to the LEs and lysosomes. At the BEE, recycling, transcytosing, and lysosome-directed material are sorted from one another. Recycling from the BEE occurs via the CE. Transcytosing material is delivered from the BEE to the CE, then to the ARE, while lysosome-directed cargo can enter the LE directly from the BEE. Roles for the cytoskeleton (actin or microtubule) are depicted for each step wherever possible, as are resident Rab GTPases.

6.2 Regulation of Endocytosis in Polarized Epithelia

Clathrin-mediated endocytosis (CME) occurs at both apical and basolateral surfaces; however, there are some differences observed between the two subdomains, as already noted above. Studies have also demonstrated differential roles for dynamin in endocytosis at either surface. Two isoforms of dynamin, dyn1 (primarily neuronal) and dyn2 (ubiquitous), differentially localize when expressed in MDCK cells; dyn1 functions at the apical surface, while dyn2 functions at the basolateral surface.

The cytoskeleton plays an important role in endocytic traffic at both membrane domains of polarized epithelial cells. Microtubules, which allow rapid ($\sim$1 μm s^{-1}) movement over long distances, have been implicated in vesicular transport between AEEs and AREs, between AEE/BEEs and LEs, between BEEs and the CE, and during basolateral to apical transcytosis. Actin filaments, which are shorter and support slower movements (0.1 μm s^{-1}), also play an important role in endocytosis. Studies with the actin-perturbing drug cytochalasin D show that apical endocytosis is inhibited, while basolateral endocytosis is unaffected. Interestingly, the actin filament stabilizer, jasplakinolide, stimulated basolateral endocytosis, but did not affect apical endocytosis; this may reflect a difference between the dynamic properties of apical and basolateral actin pools. Differential cytoskeletal dependencies may also underlie differences in internalization rates between the two domains; CME at the apical surface occurs at 20% of the rate (per unit surface area) of that found at the basolateral surface.

6.3 Transcytosis: the IgA Example

The movement of material between the apical and basolateral membrane domains is called *transcytosis*. This process was first shown to transport immunoglobulins (IgG, IgA, and IgM) across epithelial sheets, moving them from the blood into the lumen of the gut to prevent pathogenic invasion of mucosae. Transcytosis is also used to deliver maternal IgG in milk from the gut epithelium to the circulatory system of newborns, and is now known to be a major mechanism by which a variety of macromolecules are delivered to apical surfaces. The transcytotic pathway of the polymeric Ig receptor-dimeric IgA (pIgR-dIgA) complex has been extensively studied and is presented here as an example.

Dimeric IgA (dIgA) is the major immunoglobulin found in mucosal secretions; it is synthesized and secreted by B lymphocytes and then transcytosed across the epithelial cell layer for incorporation into mucus. dIgA first binds the pIgR, located on the basolateral surface (in contact with the blood) of mucosal epithelia. The complex is then endocytosed via clathrin-coated pits, followed by delivery to the BEE. From here, pIgR-dIgA travels to the CE; however, it is presently unclear from this point whether the complex is delivered directly from the CE to the apical surface, or whether it must first enter the ARE. At the apical surface, pIgR is proteolytically cleaved by a thiol-dependent protease, releasing the extracellular portion of the pIgR still bound to dIgA into secretions.

7 MVBs and the Late Endosomal System

For material destined for degradation, processing, or further sorting, the next

step in the endocytic pathway is the late endosomal system, which comprises juxtanuclear, prelysosomal organelles that typically receive cargo within 5 to 30 min following endocytosis. Late endosomes are the major intersection of the biosynthetic and endocytic pathways. As for EEs, LEs are a network of subcompartments consisting of spherical organelles, 0.4 to 0.7 μm in diameter, often containing small luminal vesicles. Though the majority of degradation occurs in lysosomes, evidence suggests that a significant amount of degradation can occur in LEs. LE organelles are variously referred to as endocytic carrier vesicles, late endosomes, multivesicular endosomes and multivesicular bodies (MVBs), and multilamellar bodies. There is still debate about the temporal relationship and defining characteristics of each of these compartments. General characteristics of MVBs/LE are depicted in Fig. 5(a). Like MVBs, LEs contain luminal vesicles; the limiting membrane is relatively depleted of cholesterol and sphingomyelin, while the internal vesicle membranes are highly enriched in lysobisphosphatidic acid (LBPA). Its hydrophobicity and propensity to bend membranes implicates LBPA in the formation of LEs. PI3P is also found on the inner membranes of LEs, mainly in regions poor in LBPA.

7.1 Late Endosomes

In addition to endocytosed cargo, LEs also receive traffic from the TGN consisting of an assortment of newly synthesized lysosomal acid hydrolases and integral membrane glycoproteins. Mannose-6-phosphate Receptors (MPR) sequester mannose-6-phosphate-tagged, soluble lysosomal hydrolases at the TGN, and mediate their sorting into CCVs. GGAs interact with the cytosolic tails of MPRs to coordinate this sorting. It is likely that these TGN-derived vesicles first travel to EEs before entering LEs. Consistent with this, the GGAs were shown recently to interact with a Rababtin-5/Rabex-5 complex on EEs. Thus, GGAs may not only be important for cargo selection but also for proper targeting. In contrast, newly synthesized lysosomal integral membrane proteins (LAMPs, LIMPs, and Igps) may travel directly to LEs through association with AP-3 complex at the TGN.

7.2 Sorting into MVBs

A defining feature of the MVBs is their luminal vesicles. These vesicles form by selective incorporation of a subset of proteins on the limiting membrane of the MVB into small vesicles that invaginate into the lumen, forming luminal vesicles (see Fig. 5(b)). The MVB pathway is required for delivery of both biosynthetic and endocytic membrane proteins to the yeast vacuole. The sorting of specific membrane proteins into luminal vesicles is an important pathway for proteolytic processing and degradation of membrane proteins, but also delivers certain vacuolar hydrolases that become active only after being sorted and processed by this pathway. Genetic studies in yeast have provided significant insight into the mechanisms underlying MVB formation, through the identification of a subset of vacuolar protein sorting (*vps*) mutants that accumulate an exaggerated LE termed the "*class E compartment.*" Recently, members of the class E *vps* family have been implicated in mediating vesicular sorting into the MVB. Vps23, Vps28, and Vps37 form a 350-kDa complex, termed the *endosomal sorting complex*

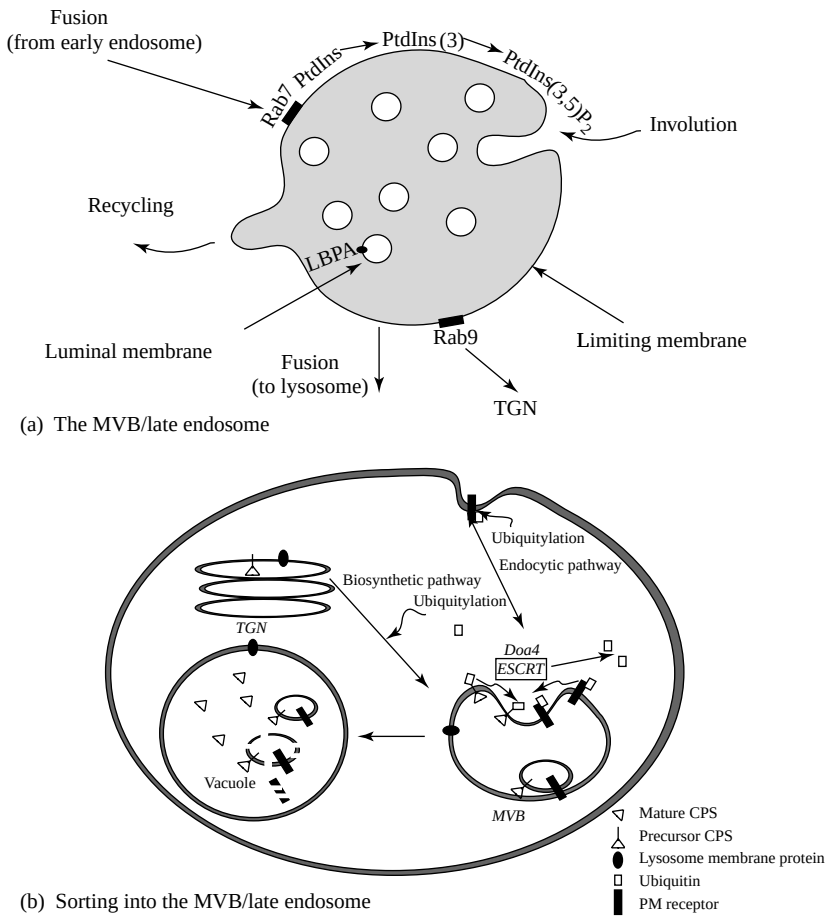

Fig. 5 Structure and sorting of MVB/late endosomes. (a) MVB/late endosomes comprise a network of membrane-bound, spherical structures. These organelles are dynamic and at any point may be involved in fusion with traffic from the early endosome (via Rab7) and the TGN (via Rab9), involution of their limiting membranes to form luminal vesicles, or recycling of limiting membrane components. PI3P localizes to limiting membranes, while LBPA localizes to luminal membranes. (b) Sorting into MVB luminal vesicles is mediated by the ESCRT complexes (I, II, and III). Sorting cargo may include PM receptors or biosynthetic material (e.g. CPS). Ubiquitylation serves as a signal for recognition of cargo by ESCRTs; however, deubiquitylation occurs prior to actual internalization into luminal vesicles.

required for transport (*ESCRT*-I). Vps23 contains a ubiquitin conjugating (UBC)-like domain that is required for its function and is proposed to bind to ubiquitylated proteins. Studies have confirmed that ubiquitylation of cargo is important for its sorting into MVB luminal vesicles in yeast. The mammalian homolog of Vps23 is encoded by the tumor susceptibility gene 101 (*tsg101*), which also contains a UBC-like domain. *tsg101* mutants recycle the epidermal growth factor receptor (EGFR) rather

than targeting it to lysosomes, resulting in excessive growth factor signaling, which may explain why these mutant mice are prone to developing tumors (see Sect. 12.4 below).

Two more yeast ESCRT complexes (ESCRT-II and ESCRT-III) were identified through analysis of physical and genetic interactions among the remaining uncharacterized class E Vps proteins. ESCRT-II is a stable 155-kDa complex formed by Vps22, Vps25, and Vps36, while the ESCRT-III complex forms only upon membrane recruitment and consists of two subcomplexes: Snf7/Vps20 and Vps2/Vps24. In the current model for sorting into and formation of luminal vesicles, Vps27 associates with PI3P and ubiquitylated cargo. This in turn is thought to promote the sequential recruitment and activation of the ESCRT-I and ESCRT-II complexes, which interact with and promote the assembly of a large oligomer of ESCRT-III on the membrane. The activity of the ESCRT complexes together results in concentration of cargo into membrane invaginations on the MVB limiting membrane that buds inwardly to form luminal vesicles. Interestingly, in some cell types, HIV virus recruits ESCRT-I during a late stage in its budding from the cell surface. Consistent with this, each of the Class E *VPS* genes has a homolog in mammalian cells.

A model for sorting biosynthetic and endocytic cargo into the MVB is depicted in Fig. 5(b). Carboxypeptidase S (CPS) is a biosynthetic cargo of the yeast MVB pathway. It is first synthesized as a type II integral membrane protein, termed *precursor CPS* (*pCPS*), and is sorted into MVB luminal vesicles. Following fusion of the limiting membrane of the MVB with the vacuole and delivery of pCPS vesicles into the vacuolar lumen, pCPS is cleaved from its transmembrane anchor to yield a mature, soluble enzyme (mCPS). Ubiquitylation of the cytosolic tail is thought to be the signal responsible for determining that a transmembrane protein will be sorted into an MVB luminal vesicle. As mentioned earlier, ubiquitylation of certain membrane receptors at the cell surface precedes their internalization, and they are believed to remain ubiquitylated until their sorting into MVBs. In yeast, this ubiquitylation is mediated by the ubiquitin ligase, Rsp5 (homolog of mammalian Nedd4). It was recently demonstrated that the ubiquitin ligase, Tul1, ubiquitylates CPS and other biosynthetic cargos for MVB sorting. Interestingly, membrane proteins that contain polar residues in their transmembrane domains are also favored for sorting into MVBs. Prior to their inclusion into MVB luminal vesicles, biosynthetic and endocytic cargo are deubiquitylated by the ubiquitin isopeptidase, Doa4, thereby recycling ubiquitin into the cytoplasm.

8 Last Stop: the Lysosome

Lysosomes – spherical, electron-dense organelles of ~0.5 μm diameter – are the major degradative compartment in the cell, and represent the final destination for endocytosed material. These organelles were discovered about 50 years ago, before the early endosomal compartments of the clathrin-dependent endocytic pathway were recognized. A major distinction that defines the difference between lysosomes and LEs is that lysosomes lack the MPR, which cycles between the TGN and LEs. However, some studies suggest that LEs and lysosomes are in dynamic equilibrium.

8.1 The Lysosome Biogenesis Controversy

The mechanism underlying lysosome formation has been a controversial issue for many cell biologists. Early studies proposed that lysosomes formed from the Golgi; however, it is now believed that these organelles derive from endosomal compartments. The three main hypotheses proposed for lysosome biogenesis are depicted in Fig. 6. The *vesicular transport* model is based on membrane trafficking events used in other steps of the endocytic

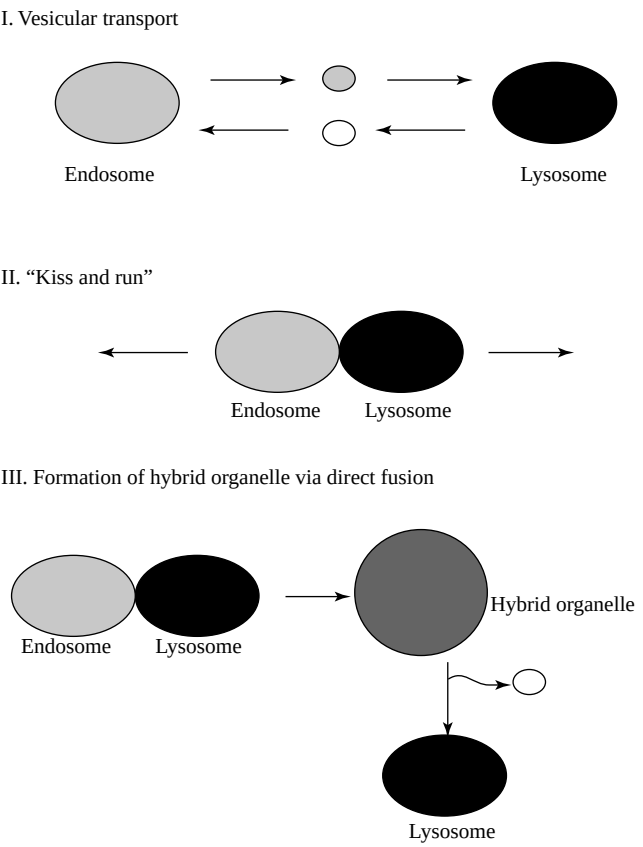

Fig. 6 Models for late endosome–lysosome fusion. Three models exist for late endosome–lysome fusion. (a) Vesicular transport: proposes that late endosomes and lysosomes are stable compartments that receive cargo via fusion with small transport vesicles. There is no experimental evidence to support this model. (b) Kiss and run: transfer of luminal cargo occurs via repeated fusion and fission cycles between LEs and lysosomes. LEs and lysosomes are stable structures, and fusion occurs through transient fusion pores. (c) Direct fusion with formation of a hybrid organelle: fusion between LEs and lysosomes forms a stable hybrid organelle exhibiting intermediate characteristics. There is experimental evidence to support this model.

and secretory pathways and proposes that LEs and lysosomes are stable compartments that receive cargo via fusion with small transport vesicles. While this most likely occurs earlier in the endocytic pathway, there is still no evidence to support this model at later stages.

Another model called *kiss and run* involves repeated fusion and fission cycles between LEs and lysosomes to allow transfer of luminal cargo. This mechanism, proposed by Storrie and Desjardins, was used to explain the differences observed for soluble versus membrane cargo in the rates of delivery between LEs and lysosomes. The third model, *formation of a hybrid organelle via direct fusion*, is best supported by cell-free, *in vitro* fusion assays. When direct fusion was allowed between LEs and lysosomes, a large diameter, membrane-bound organelle was formed with intermediate characteristics between the two compartments and that contained markers for both LE and lysosomes. These studies also provided evidence of LE–lysosome fusion *in vivo* through electron microscopic studies in cultured cells. Also, existence of an LE–lysosome hybrid organelle was observed *in vivo* in normal rat kidney (NRK) cells at steady state, in which approximately 15% of lysosome fusion with LEs formed a hybrid compartment. Thus, LE–lysosome fusion may be the most likely mechanism to occur *in vivo*.

8.2 Rab7 and Endosome–Lysosome Fusion

Work by Bucci and colleagues has demonstrated that the Rab7 GTPase plays a key role in controlling late endocytic traffic. Rab7 localizes to LE–lysosome structures, and overexpression of wild type or GTPase-deficient Rab7 ($Rab7^{Q67L}$) causes aggregation of lysosomes, while expression of dominant-negative $Rab7^{T22N}$ causes their dispersal. A Rab7–GTP binding protein, Rab-interacting lysosomal protein (RILP), which colocalizes with lysosomal markers, was recently identified and is proposed to be an effector protein. Interestingly, in 2001, Jordens et al. showed that RILP expression induces recruitment of dynein–dynactin motor complexes to Rab7-containing LE–lysosome organelles. As other Rab proteins have been shown to interact with motor proteins (e.g. Rab5, Rab6, and Rab27), there may be a general role for Rabs in regulating motor protein recruitment to ensure proper directionality of vesicle transport and fusion.

Formation of LE–lysosome hybrid organelles *in vitro* is ATP, cytosol, and temperature dependent. Cytosol requirements include typical fusion machinery components, including NSF, soluble NSF attachment proteins (SNAP), and a small Rab GTPase. On the basis of EE homotypic fusion studies, a role for SNAREs has also been proposed, though specific SNAREs have yet to be identified. The proteolipid v_0 subunit of the vacuolar H+-ATPase has also been suggested to form a fusion pore necessary for homotypic vacuole fusion. As mentioned earlier, yeast genetics have identifed several proteins involved in vacuolar fusion. Class B and C Vps proteins have been implicated in endosome–vacuole fusion, including t-SNARE machinery (Vam3/Vam7), the Rab GTPase Ypt7, and various Vps proteins that may provide tethering activity.

8.3 Hydrolases, Lamps, and Transporters

There are more than 50 endolysosomal hydrolases including proteases, nucleases, glycosidases, lipases, and phosphatases. Most of these enzymes, like

the yeast vacuolar CPS, are synthesized as precursors that are glycosylated in the ER/Golgi – which may protect them from destruction – followed by removal of a prodomain upon delivery to the lysosome, which generates the mature enzyme. These enzymes are also active only at acidic pHs, and the lysosome lumen is maintained at ~pH 5.0 by vacuolar H^+-ATPase and a redox chain that uses cytoplasmic NADH as an electron donor and oxygen as the final electron acceptor.

LAMPs (lysosomal-associated membrane protein) are the major integral membrane protein constituents of the lysosomal membrane, making up ~50% of the total membrane protein in this organelle. LAMP-1 and LAMP-2 are evolutionarily related type I membrane proteins that have long luminal domains and short cytoplasmic tails which contain lysosome sorting signals. One role of these proteins may be to protect lysosomes from self-hydrolysis and/or prevent escape of hydrolases into the cytoplasm.

Export of proteolysis products, which are recycled for use in anabolic reactions, occurs mainly via transporters. In addition to amino acid transporters, those for carbohydrate monomers, ions, and nucleosides have also been reported. The cysteine transporter is an interesting example. This transporter regulates the flux of cysteine into, rather than out of lysosomes, and its activity may be pH dependent. Cysteine is proposed to serve in regulating the redox potential of lysosomes, as certain enzymes require reducing conditions to be active.

8.4
Secretory Lysosomes

While regulated secretion has long been regarded as unique to specific cell types, the recent finding that conventional lysosomes can fuse with the PM has changed the traditional view of these organelles. Lysosomes that act as Ca^{2+}-dependent secretory granules have been observed in hematopoietic cells such as platelets, neutrophils, macrophages, and cytotoxic T lymphocytes (CTLs). Ca^{2+}-regulated secretion of conventional lysosomes has now been observed in a number of cells, and may serve, in part, to expel undigested lysosomal contents to prevent accumulation of cellular debris.

9
Parasites and the Endocytic Pathway

The endocytic pathway presents several hurdles that parasites must overcome in order to invade and survive successfully in their hosts. First, intracellular parasites must penetrate the PM of the cell, then avoid destruction by the hydrolytic lysosome, and finally, the parasite must employ strategies to evade the immune system. The diversity of intracellular parasites is paralleled by the wide array of solutions to these challenges.

Most commonly, the parasite chooses a professional phagocytic cell as a host, and uses phagocytosis for internalization. For example, the bacterial pathogen *Salmonella* induces robust phagocytosis upon contacting the surface of a macrophage. Other bacteria such as *Listeria* bind to the PM and "zipper up" a membrane compartment around themselves as they dive into the cytoplasm. In contrast, much larger protozoan parasites have resorted to strategies that involve the recruitment of intracellular membranes, ranging from granules, ER, recycling endosomes, or lysosomes, each of which can contribute membranes to the forming parasitiphorous vacuole through fusion with the PM at the point of entry.

Exciting recent evidence suggests that the ER plays a more generally utilized role in phagocytosis, and this finding provides satisfactory explanations for many previously puzzling questions. In this new model for the mechanism of pathogen internalization, the ER provides the bulk of the membrane that envelops the pathogen. Under circumstances in which the PM area is insufficient, the ER is an abundant membrane-rich organelle that can supplement the PM for the purpose of covering the pathogen with a lipid bilayer. This also explains how many pathogens can escape degradation by the endolysosomal pathway; the ER is on the biosynthetic pathway and thus the precursor forms of degradative hydrolases are inactive in this compartment. Finally, an ER-dependent component to phagocytosis potentially explains the source of the ER-resident protein calreticulin at the surface of many cells, and also suggests a mechanism for MHC (major histocompatibility complex) Class I presentation of exogenously derived peptides.

In other cell types such as neutrophils, an ER-independent mode of pathogen uptake via phagocytosis predominates. In this case, some pathogens have devised various clever strategies to avoid being destroyed upon delivery to the hydrolytic lysosome. For instance, some pathogens synthesize factors that prevent the maturation of the endosomal compartment they occupy, others convert their compartment to one of a completely novel identity, and still others escape the membranous compartment entirely and spend their lives swimming naked in the cytosol. Access to nutrients is a concern, and some parasites often introduce transporters or pores into the membrane of the parasitiphorous vacuole to allow for diffusion of small molecules such as sugars, nucleotides, and amino acids from the host cytoplasm.

The protozoan parasite, *Trypanosoma cruzi*, has adopted an alternative strategy that allows it to use nearly any cell type as a host, not just professional phagocytes. Upon interaction with the PM of the host cell, *T. cruzi* recruits lysosomes to the cell periphery and induces a Ca^{2+} signal that triggers local fusion of the lysosomes with the PM (see Sect. 8), followed by entry of *T. cruzi* into the cell. This mechanism of internalization explains how *T. cruzi* internalization can occur in cells treated with the actin-disrupting agent, cytochalasin D, which blocks phagocytosis. Once inside the cell, a brief exposure to the acidic pH within the lumen of the lysosome-like compartment induces *T. cruzi* to proceed through its developmental cycle, culminating in its entry into the cytoplasm where it divides and spreads to other cells.

10 Exosomes

Lysosomes are not the only organelles of the later endocytic pathway that can fuse with the PM; MVBs can also undergo exocytosis. When an MVB containing luminal vesicle fuses with the PM, the "secreted" luminal vesicles are referred to as *exosomes*. While detailed studies of these small membrane vesicles are still in their infancy, exosomes have generated a great deal of excitement in recent years, primarily due to their potential therapeutic roles in stimulating immune responses. Exosomes were first identified in maturing reticulocytes, which are anucleate and devoid of most intracellular organelles. The release of exosomes provides a mechanism for removing extraneous membrane and transmembrane proteins, such as TfR, to form a mature red blood cell.

It is known that many cell types secrete exosomes, but the physiological role is not yet understood in most cases. In contrast, the role for the exosomes secreted by CTLs is better characterized. It appears that CTLs package perforin and granzymes into exosome membranes for secretion into the immunological synapse formed between the CTL and the target cell it will kill. Many questions persist about exosomes, including their *in vivo* function, and the circumstances under which the MVB-containing exosomes fuse with the PM rather than the lysosome. One recent study suggests that as immature dendritic cells become activated, a pathway is stimulated that delivers MHC II-containing organelles (likely exosomes) to the immune synapses.

The greatest degree of excitement surrounding exosomes is derived from their potential use as an antitumorigenic therapy – exosomes purified from B cells or dendritic cells can stimulate CTLs, and clinical trials suggest that they may be an effective adjuvant for antitumor immunizations. On a different note, exosomes have also been proposed to play important roles in developmental events in *Drosophila*, including establishing morphogen gradients and in proteolytic processing and activation of complicated ligand/receptor partners like Delta/Notch, in which each participant is a transmembrane protein.

11 Endocytosis and Signaling

11.1 Regulated Endocytosis of Signaling Receptors

Many signaling receptors are rapidly endocytosed after ligand-induced activation, such as the EGF receptor tyrosine kinase (EGFR) and the beta-2 adrenergic G-protein-coupled receptor (β2AR), which are endocytosed by clathrin-coated pits. There is also evidence for regulated endocytosis of certain signaling receptors via clathrin and/or dynamin-independent mechanisms. Coated pits and vesicles can contain a mixture of signaling receptors, endocytosed in a ligand-regulated manner, and constitutively endocytosed receptors, such as LDL and the TfRs. This led to the proposal that a major mechanism by which endocytosis of signaling receptors is regulated is by ligand-dependent association with constitutively formed clathrin-coated pits. This hypothesis is consistent with many observations and with the proposed biochemical mechanism of β2AR endocytosis, in which nonvisual (β-) arrestins function as adaptor proteins linking ligand-activated and phosphorylated receptors with the clathrin-coat structure. However, there are some data suggesting that signaling receptors endocytose via a subpopulation of clathrin-coated pits, modify the properties of the coated pits in which they are concentrated, or, in some cases, actually promote the formation of new coated pits *de novo*. Therefore, the degree to which signaling receptors function as "inert" endocytic cargo or regulate the endocytic machinery in some way remains a controversial issue.

11.2 Functional Consequences of Signaling Receptor Endocytosis

11.2.1 Receptor Downregulation

One function served by regulated endocytosis of signaling receptors is to rapidly reduce the number of surface receptors accessible to extracellular ligands. In this way, ligand-induced endocytosis is

proposed to serve as a negative feedback mechanism to attenuate (or "downregulate") cellular responsiveness. Downregulation EGFR signaling involves ligand-induced endocytosis of receptors followed by sorting in MVBs and subsequent delivery to lysosomes. Sorting of internalized EGFR in MVBs is mediated by a ubiquitin-dependent mechanism involving mammalian homologs of the yeast class E Vps proteins. Experimental manipulations and naturally occurring mutations that disrupt EGFR sorting by this mechanism result in enhanced growth factor signaling and contribute to tumorigenesis (see Sect. 12.4).

11.2.2 Receptor Resensitization

The functional activity of certain G-protein-coupled receptors, such as the β2AR, can be attenuated rapidly by receptor phosphorylation occurring in the PM, without requiring endocytosis. Ligand-induced endocytosis of the β2AR is proposed to promote the recovery (or "resensitization") of receptor signaling activity, in contrast to the prolonged attenuation of the cell signaling characteristic of downregulation, by delivering phosphorylated receptors to an endosome-associated phosphatase and then to a rapid recycling pathway through which dephosphorylated receptors are returned to the PM in a fully functional state. Nevertheless, ligand-induced endocytosis of many G-protein-coupled receptors (including the β2AR under some conditions) can also lead to receptor trafficking to lysosomes and a more prolonged downregulation of cell signaling. Thus, the degree to which ligand-induced endocytosis via CCVs mediates the functionally opposite processes of resensitization or downregulation depends critically on the sorting of internalized receptors between divergent downstream membrane pathways.

11.2.3 Signaling from Endosomes

In addition to modulating the number or functional activity of signaling receptors in the PM, there is accumulating evidence that certain receptors actually signal from the endosome membrane. Early support for this hypothesis came from biochemical studies suggesting that endocytic vesicles containing internalized receptor tyrosine kinases (such as insulin and EGF receptors) are associated with cytoplasmic proteins that mediate downstream signal transduction from these receptors (such as Grb2 and Src), and more recent studies using fluorescence resonance energy transfer (FRET) imaging indicate that these proteins interact on the endosome membrane. Functional support for this idea comes from studies of neurotrophin signaling via the TrkA receptor tyrosine kinase, which suggest that "signaling endosomes" containing activated receptors and associated signaling molecules carry the receptor-mediated signal for long distances (up to 1 m in certain sensory neurons) from the site of initial receptor activation (in distal axons) to the site of specific effectors (in the cell body) that regulate the expression of genes controlling cell survival.

12 Some Diseases of the Endocytic Pathway

As described earlier, pathogens can exploit the endocytic machinery to gain access into host cells, causing disease. However, a number of genetic diseases result in defects along the endocytic pathway. A few of these diseases are described below.

12.1 Hypercholesterolemia

The majority of the cholesterol required for PM synthesis is acquired through receptor-mediated endocytic uptake. Sixty-five to seventy-five percent of plasma cholesterol is bound to protein in the form of LDL particles for transport through the blood stream, and cells then take up cholesterol via LDLRs. Once in the endosomes, the acidic pH dissociates LDL from its receptor. LDLRs are sorted for recycling, while LDL particles are sorted into the degradation pathway. In lysosomes, cholesteryl esters in the LDL particles are hydrolyzed to free cholesterol, which exits the lysosomes and is available to the cell for new membrane synthesis. Disruption of LDLR-mediated endocytic uptake of cholesterol leads to accumulation of elevated blood cholesterol levels, which can cause formation of atherosclerotic plaques in blood vessels and can lead to strokes and heart attacks. One in 300 to 500 individuals suffers from heterozygous familial hypercholesterolemia (FH), which is associated with mutations in the LDLR gene, as was described earlier. FH patients exhibit abnormally high blood cholesterol levels from birth, throughout adulthood, and often suffer from coronary artery disease and atherosclerosis. A rare autosomal recessive form of hypocholesterolemia (ARH) phenotypically resembles FH, but is not due to mutations in LDLR. Recently, the *ARH* gene was mapped and identified as a putative adaptor protein that contains a phosphotyrosine binding domain (PTB) and interacts directly with the clathrin endocytic machinery. Similar to AP-2, ARH binds to phosphoinositides and clathrin; however, unlike AP-2, ARH can also recognize the LDLR internalization signal, Asn-Pro-X-Tyr (NPXY), through its PTB domain. It is proposed that ARH performs an adaptor function similar to AP-2 for LDLR, and its malfunction disrupts LDL-cholesterol traffic, leading to hypercholesterolemia.

12.2 Defects in Endosomal Acidification and Proximal Tubulopathies

A major function of the proximal tubules of the kidney is filtering hundreds of milligrams of plasma proteins (e.g. albumin and low molecular weight (LMW) proteins), which enter the filtrate via the glomeruli. Since human urine is normally devoid of significant amounts of protein, the reabsorption of proteins by receptor-mediated endocytosis plays an essential role in protein, vitamin, and hormone homeostasis, as well as in the recovery of amino acids. Lignac-de Toni-Debre-Fanconi syndromes (Fanconi syndromes) are disorders that result from the primary dysfunction of the proximal tubule with varying etiologies, but all exhibit defective renal tubular protein reabsorption and tubular proteinuria. These defects are thought to be due primarily to defects in receptor-mediated endocytic pathway in proximal tubules. The apical endocytic apparatus of proximal tubule epithelial cells plays a vital role in the reabsorption and degradation of material from the filtrate. Critical to this endocytic process is endosomal acidification, perturbation of which leads to pathophysiological dysfunction, often resulting in urinary wasting of LMW proteins, electrolytes, and solutes.

The two apical receptors, megalin and cubulin, have been implicated in receptor-mediated endocytosis of ligands such as insulin, HDL, albumin, and vitamin-binding proteins. Indeed, megalin and cubulin knockout mice are defective in tubular

reabsorption, excrete LMW proteins and vitamins in their urine, and are used as animal models for Fanconi syndromes. Proper recycling of megalin and cubulin requires endosomal acidification machinery including the vacuolar H^+-ATPase and CLC-5 chloride ion channels, since dissociation of receptor from ligand in endosomes occurs at low pHs, and is required for receptor recycling. Recently, a connection between endosomal acidification and the Arf6 GTPase/ARNO coat protein factors was identified. Though the precise correlations remain to be determined, Arf6 and ARNO colocalize with vacuolar H^+-ATPase in the apical endosomes of proximal tubules in a pH-dependent manner, a localization that overlaps with that of megalin, Rab5, and Rab11.

12.3 Lysosomal Dysfunction

Numerous diseases occur as a result of lysosomal abnormalities. In fact, mutations causing lysosomal enzyme deficiencies account for approximately 40 different disorders collectively called *Lysosomal Storage Diseases* (*LSDs*). In many cases, lysosomal defects lead to an increase in the size and number of lysosomes, eventually causing malfunction of the affected organ or organs. All LSDs are serious conditions, and most are fatal, resulting in death in infancy or early childhood. Many LSDs are associated with neurodegeneration, and lead to severe clinical manifestations such as mental retardation, progressive cognitive decline, and inappropriate behaviors. Interestingly, these conditions are also phenotypic of common age-related neurological disorders such as Alzheimer's and Parkinson's disease. In fact, it is believed that, since total cellular lysosomal capacity is reduced with brain aging, lysosomal abnormalities may contribute to age-related neuropathogenesis.

12.4 The Cancer Link

The importance of endocytosis in the downregulation of signaling complexes implies, perhaps, obvious connections to cancer. The ErbB family of receptor tyrosine kinases (RTKs) consists of transmembrane receptors including EGFR, Neu/ErbB-2, ErbB-3/HER3, and ErbB-4/HER4. These signaling receptors are involved in epithelial development, renewal, and remodeling, and thus, are major regulators of cell proliferation. c-Cbl is a large, multimodular cytoplasmic protein involved in endocytosis that binds phosphorylated EGFR via an N-terminal tyrosine kinase binding domain. c-Cbl is also an E3 ubiquitin ligase, and ubiquitylates EGFR prior to clathrin-coated pit formation; however, it also colocalizes with EGFR in MVB luminal vesicles. Interestingly, c-Cbl was originally identified as the cellular form (hence the "c") of viral-Cbl (v-Cbl), a transforming protein of the Cas NS-1 retrovirus. v-Cbl contains only the N-terminal region of c-Cbl, which binds to but does not ubiquitylate EGFR; this causes EGFR to be constitutively recycled rather than degraded, and leads to enhanced signaling and oncogenic transformation.

A link to cancer is found in several other endocytic proteins. Chromosomal translocations leading to fusions between three different endocytic proteins and ALL1/HRX-related gene products have been linked to hematopoietic cancers. A fusion between the C-terminal EH domain of AF-1p (the human ortholog of murine Eps15) with the N-terminus of ALL1/HRX

is implicated in two myeloid leukemias. A similar fusion between the N-terminus of ALL1/HRX and the C-terminal domain of SH3p8 (endophilin) occurs in an acute myeloid leukemia. Finally, a fusion containing the full-length CALM protein (ubiquitously expressed AP180 homolog) with its last four amino acids replaced by amino acids 87–1027 of the AF10 gene product (a putative transcription factor) was identified in a lymphoma cell line.

13 Conclusion

The complexity of the endocytic pathway includes requirements for many protein components and numerous membrane compartments, the selection and sorting of cargo that occurs at each step along the pathway, and the intricate regulation of these events that ultimately determines cellular physiology and responses to environmental changes. Many of the lessons learned from molecular cell biological studies of endocytosis also apply to other membrane trafficking pathways, and in fact may foreshadow a greater complexity in these other pathways than was previously appreciated. For example, a similar multiplicity of adaptors and cytosolic factors necessary for the formation of CCVs at the plasma membrane is beginning to be uncovered in studies of CCV formation at the TGN.

In this chapter, we strove to present the current understanding of the molecular mechanisms of endocytosis and to highlight a few of the physiological functions served by these mechanisms; however, there are many important questions that remain controversial and critical breakthroughs that are imminent. Perhaps the most debated area continues to be on the issue of individual compartment identity and whether compartments arise through maturation as opposed to somewhat "static" compartments through which cargo moves via vesicular-mediated transport; recent experimental approaches are yielding fresh insight into this fundamental and long-debated question. It may be informative to consider the endocytic pathway analogously to the biosynthetic pathway: endocytic "hot spots" could correspond to transitional ER exit sites, EEs to the ER–Golgi intermediate compartment, and LEs may mature from one type into another akin to maturation of sequential Golgi cisternae. In this view, the endocytic pathway would utilize both vesicular transport and maturation mechanisms and might rely upon distinct protein coats that direct forward traffic versus retrieval/recycling pathways.

How flow through the endocytic pathway is regulated is another area with many outstanding questions. This will undoubtedly entail studies of protein kinases and phosphatases, lipid modifying enzymes and enzymes that govern protein ubiquitylation, and perhaps other posttranslational modifications not yet anticipated. To understand fully the roles of these regulatory enzymes will require knowledge of how these factors themselves are regulated and what dictates their subcellular localization and substrate specificity.

A final emerging area of great interest for studies of endocytosis is defining the role(s) for the cytoskeleton, particularly the actin cytoskeleton. Many connections between the cytoplasmic endocytic machinery and actin have been uncovered through biochemical and genetic approaches, but the function of the dynamic polymerization/depolymerization cycle in endocytosis remains elusive. Finding the answers will be complicated by the pleiotropic roles

played by actin, and the way in which cells adapt to perturbations of one specific membrane trafficking step by spilling over into alternate pathways.

Our current understanding of endocytosis has been achieved through a wide variety of experimental approaches, and the field of endocytosis continues to be one in which novel and interdisciplinary approaches to cell biological mechanisms are pioneered. Great advances have been made using the tools of structural biology, beginning with seminal studies describing the clathrin lattice, and we anticipate that the pace with which endocytic mechanisms are elucidated at the atomic level will continue to increase. We also anticipate that exciting advances will continue to emerge as live-cell imaging methods facilitating sophisticated four dimensional (simultaneous spatial and temporal) analyses of endocytic proteins in action in the intact cell are developed and applied, along with *in vitro* reconstitution assays that accurately recapitulate individual stages of endocytosis. These approaches, together with increased use of genetic analyses – including newer methods (such as interfering RNA and tissue-specific homologous recombination) that increase the range of organisms in which detailed genetic studies are possible – will eventually lead us to a full and complete appreciation of the inner workings of endocytosis, in particular, and of membrane trafficking processes in general.

Acknowledgments

Hadiya A Watson is supported by a Ford Foundation Dissertation Fellowship. Mark Von Zastrow is supported by grants from the National Institute on Drug Abuse. Beverly Wendland is supported by grants from the Burroughs Wellcome Fund and the National Institutes of Health.

See also Cell Membranes, Transport Across; Phospholipids.

Bibliography

Books and Reviews

Amer, A.O., Swanson, M.S. (2002) A phagosome of one's own: a microbial guide to life in the macrophage, *Curr. Opin. Microbiol.* **5**, 56–61.

Andrews, N.W. (2000) Regulated secretion of conventional lysosomes, *Trends Cell Biol.* **10**, 316–321.

Apodaca, G. (2001) Endocytic traffic in polarized epithelial cells: role of the actin and microtubule cytoskeleton, *Traffic* **2**, 149–159.

Barlowe, C. (2000) Traffic COPs of the early secretory pathway, *Traffic* **5**, 371–377.

Engqvist-Goldstein, A.E., Drubin, D.G. (2003) Actin assembly and endocytosis: from yeast to mammals, *Annu. Rev. Cell Dev. Biol.* **19**, 287–332.

Bahr, B.A., Bendiske, J. (2002) The neuropathogenic contributions of lysosomal dysfunction, *J. Neurochem.* **83**, 481–489.

Boehm, M., Bonifacino, J.S. (2001) Adaptins: the final recount, *Mol. Biol. Cell* **12**, 2907–2920.

Brodsky, F.M., Chen, C.Y., Knuehl, C., Towler, M.C., Wakeham, D.E. (2001) Biological basket weaving: formation and function of clathrin-coated vesicles, *Annu. Rev. Cell Dev. Biol.* **17**, 517–568.

Bryant, N.J., Stevens, T.H. (1998) Vacuole biogenesis in Saccharomyces cerevisiae: protein transport pathways to the yeast vacuole, *Microbiol. Mol. Biol. Rev.* **62**, 230–247.

Brown, M.S., Goldstein, J.L. (1984) How LDL receptors influence cholesterol and atherosclerosis, *Sci. Am.* **251**, 58–66.

Conibear, E. (2002) An ESCRT into the endosome, *Mol. Cell.* **10**, 215–216.

Denzer, K., Kleijmeer, M.J., Heijnen, H.F., Stoorvogel, W., Geuze, H.J. (2000) Exosome: from internal vesicle of the multivesicular body to intercellular signaling device, *J. Cell Sci.* **113**(Pt 19), 3365–3374.

Dupre, S., Volland, C., Haguenauer-Tsapis, R. (2001) Membrane transport: ubiquitylation in endosomal sorting, *Curr. Biol.* **11**, R932–R934.

Gruenberg, J. (2001) The endocytic pathway: a mosaic of domains, *Nat. Rev. Mol. Cell Biol.* **2**, 721–730.

Hurley, J.H., Wendland, B. (2002) Endocytosis: driving membranes around the bend, *Cell* **111**, 143–146.

Kurzchalia, T.V., Parton, R.G. (1999) Membrane microdomains and caveolae, *Curr. Opin. Cell Biol.* **11**, 424–431.

Marsh, M. (Ed.). (2001) *Endocytosis*, Vol. 36, Oxford University Press, Oxford.

Qualmann, B., Kessels, M.M. (2002) Endocytosis and the cytoskeleton, *Int. Rev. Cytol.* **220**, 93–144.

Sorkin, A. (2000) The endocytosis machinery, *J. Cell Sci.* **113**(Pt 24), 4375–4376.

Wendland, B., Emr, S.D., Riezman, H. (1998) Protein traffic in the yeast endocytic and vacuolar protein sorting pathways, *Curr. Opin. Cell Biol.* **10**, 513–522.

Zerial, M., McBride, H. (2001) Rab proteins as membrane organizers, *Nat. Rev. Mol.Cell Biol.* **2**, 107–117.

Primary Literature

Advani, R.J., Bae, H.R., Bock, J.B., Chao, D.S., Doung, Y.C., Prekeris, R., Yoo, J.S., Scheller, R.H. (1998) Seven novel mammalian SNARE proteins localize to distinct membrane compartments, *J. Biol. Chem.* **273**, 10317–10324.

Altschuler, Y., Liu, S., Katz, L., Tang, K., Hardy, S., Brodsky, F., Apodaca, G., Mostov, K. (1999) ADP-ribosylation factor 6 and endocytosis at the apical surface of Madin-Darby canine kidney cells, *J. Cell Biol.* **147**, 7–12.

Anderson, R.G., Goldstein, J.L., Brown, M.S. (1977) A mutation that impairs the ability of lipoprotein receptors to localise in coated pits on the cell surface of human fibroblasts, *Nature* **270**, 695–699.

Aniento, F., Gruenberg, J. (1995) Membrane transport from early to late endosomes, *Cold Spring Harb. Symp. Quant. Biol.* **60**, 205–209.

Aniento, F., Gu, F., Parton, R.G., Gruenberg, J. (1996) An endosomal beta COP is involved in the pH-dependent formation of transport vesicles destined for late endosomes, *J. Cell Biol.* **133**, 29–41.

Apodaca, G., Katz, L.A., Mostov, K.E. (1994) Receptor-mediated transcytosis of IgA in MDCK cells is via apical recycling endosomes, *J. Cell Biol.* **125**, 67–86.

Babst, M., Katzmann, D.J., Estepa-Sabal, E.J., Meerloo, T., Emr, S.D. (2002a) Escrt-III: an endosome-associated heterooligomeric protein complex required for mvb sorting, *Dev. Cell* **3**, 271–282.

Babst, M., Katzmann, D.J., Snyder, W.B., Wendland, B., Emr, S.D. (2002b) Endosome-associated complex: ESCRT-II, recruits transport machinery for protein sorting at the multivesicular body, *Dev. Cell* **3**, 283–289.

Babst, M., Odorizzi, G., Estepa, E.J., Emr, S.D. (2000) Mammalian tumor susceptibility gene 101 (TSG101) and the yeast homologue, Vps23p, both function in late endosomal trafficking, *Traffic* **1**, 248–258.

Beaumelle, B.D., Gibson, A., Hopkins, C.R. (1990) Isolation and preliminary characterization of the major membrane boundaries of the endocytic pathway in lymphocytes, *J. Cell Biol.* **111**, 1811–1823.

Bernard, O.A., Mauchauffe, M., Mecucci, C., Van den Berghe, H., Berger, R. (1994) A novel gene, AF-1p, fused to HRX in t(1;11)(p32;q23), is not related to AF-4, AF-9 nor ENL, *Oncogene* **9**, 1039–1045.

Bishop, N., Horman, A., Woodman, P. (2002) Mammalian class E vps proteins recognize ubiquitin and act in the removal of endosomal protein-ubiquitin conjugates, *J. Cell Biol.* **157**, 91–101.

Boll, W., Ohno, H., Songyang, Z., Rapoport, I., Cantley, L.C., Bonifacino, J.S., Kirchhausen, T. (1996) Sequence requirements for the recognition of tyrosine-based endocytic signals by clathrin AP-2 complexes, *EMBO J.* **15**, 5789–5795.

Bomsel, M., Parton, R., Kuznetsov, S.A., Schroer, T.A., Gruenberg, J. (1990) Microtubule- and motor-dependent fusion in vitro between apical and basolateral endocytic vesicles from MDCK cells, *Cell* **62**, 719–731.

Bomsel, M., Prydz, K., Parton, R.G., Gruenberg, J., Simons, K. (1989) Endocytosis in filter-grown Madin-Darby canine kidney cells, *J. Cell Biol.* **109**, 3243–3258.

Breitfeld, P.P., McKinnon, W.C., Mostov, K.E. (1990) Effect of nocodazole on vesicular traffic to the apical and basolateral surfaces of polarized MDCK cells, *J. Cell Biol.* **111**, 2365–2373.

Bright, N.A., Lindsay, M.R., Stewart, A., Luzio, J.P. (2001) The relationship between lumenal and limiting membranes in swollen late endocytic compartments formed after wortmannin treatment or sucrose accumulation, *Traffic* **2**, 631–642.

Bright, N.A., Reaves, B.J., Mullock, B.M., Luzio, J.P. (1997) Dense core lysosomes can fuse with late endosomes and are re-formed from the resultant hybrid organelles, *J. Cell Sci.* **110**, 2027–2040.

Brown, W.J., DeWald, D.B., Emr, S.D., Plutner, H., Balch, W.E. (1995) Role for phosphatidylinositol 3-kinase in the sorting and transport of newly synthesized lysosomal enzymes in mammalian cells, *J. Cell Biol.* **130**, 781–796.

Bucci, C., Thomsen, P., Nicoziani, P., McCarthy, J., van Deurs, B. (2000) Rab7: a key to lysosome biogenesis, *Mol. Biol. Cell.* **11**, 467–480.

Cantalupo, G., Alifano, P., Roberti, V., Bruni, C.B., Bucci, C. (2001) Rab-interacting lysosomal protein (RILP): the Rab7 effector required for transport to lysosomes, *EMBO J.* **20**, 683–693.

Cao, T.T., Mays, R.W., von Zastrow, M. (1998) Regulated endocytosis of G-protein-coupled receptors by a biochemically and functionally distinct subpopulation of clathrin-coated pits, *J. Biol. Chem.* **273**, 24592–24602.

Christoforidis, S., McBride, H.M., Burgoyne, R.D., Zerial, M. (1999) The Rab5 effector EEA1 is a core component of endosome docking, *Nature* **397**, 621–625.

Clague, M.J., Urbe, S., Aniento, F., Gruenberg, J. (1994) Vacuolar Pase activity is required for endosomal carrier vesicle formation, *J. Biol. Chem.* **269**, 21–24.

Claing, A., Chen, W., Miller, W.E., Vitale, N., Moss, J., Premont, R.T., Lefkowitz, R.J. (2001) beta-Arrestin-mediated ADP-ribosylation factor 6 activation and beta 2- adrenergic receptor endocytosis, *J. Biol. Chem.* **276**, 42509–42513.

Collins, B.M., McCoy, A.J., Kent, H.M., Evans, P.R., Owen, D.J. (2002) Molecular architecture and functional model of the endocytic AP2 complex, *Cell* **109**, 523–535.

Cremona, O., Di Paolo, G., Wenk, M.R., Luthi, A., Kim, W.T., Takei, K., Daniell, L., Nemoto, Y., Shears, S.B., Flavell, R.A., McCormick, D.A., De Camilli, P. (1999) Essential role of phosphoinositide metabolism in synaptic vesicle recycling, *Cell* **99**, 179–188.

Cui, S., Verroust, P.J., Moestrup, S.K., Christensen, E.I. (1996) Megalin/gp330 mediates uptake of albumin in renal proximal tubule, *Am. J. Physiol.* **271**, F900–F907.

D'Souza-Schorey, C., Li, G., Colombo, M.I., Stahl, P.D. (1995) A regulatory role for ARF6 in receptor-mediated endocytosis, *Science* **267**, 1175–1178.

Damke, H., Baba, T., van der Bliek, A.M., Schmid, S.L. (1995) Clathrin-independent pinocytosis is induced in cells overexpressing a temperature-sensitive mutant of dynamin, *J. Cell Biol.* **131**, 69–80.

Dautry-Varsat, A. (2001) Clathrin-Independent Endocytosis, in Marsh, M. (Ed.) *Endocytosis*, Vol. 36, Oxford University Press, Oxford, pp. 26–57.

de Duve, C. (1983) Lysosomes revisited, *Eur. J. Biochem.* **137**, 391–397.

de Melker, A.A., van der Horst, G., Calafat, J., Jansen, H., Borst, J. (2001) c-Cbl ubiquitinates the EGF receptor at the plasma membrane and remains receptor associated throughout the endocytic route, *J. Cell Sci.* **114**, 2167–2178.

de Renzis, S., Sonnichsen, B., Zerial, M. (2002) Divalent Rab effectors regulate the subcompartmental organization and sorting of early endosomes, *Nat. Cell Biol.* **4**, 124–133.

Dell'Angelica, E.C., Mullins, C., Bonifacino, J.S. (1999) AP-4, a novel protein complex related to clathrin adaptors, *J. Biol Chem.* **274**, 7278–7285.

Dell'Angelica, E.C., Puertollano, R., Mullins, C., Aguilar, R.C., Vargas, J.D., Hartnell, L.M., Bonifacino, J.S. (2000) GGAs: a family of ADP ribosylation factor-binding proteins related to adaptors and associated with the Golgi complex, *J. Cell Biol.* **149**, 81–94.

Donaldson, J.G., Radhakrishna, H. (2001) Expression and properties of ADP-ribosylation factor (ARF6) in endocytic pathways, *Methods Enzymol.* **329**, 247–256.

Dunn, K.W., Maxfield, F.R. (1992) Delivery of ligands from sorting endosomes to late endosomes occurs by maturation of sorting endosomes, *J. Cell Biol.* **117**, 301–310.

Dunn, K.W., McGraw, T.E., Maxfield, F.R. (1989) Iterative fractionation of recycling receptors from lysosomally destined ligands in an early sorting endosome, *J. Cell Biol.* **109**, 3303–3314.

Echard, A., Jollivet, F., Martinez, O., Lacapere, J.J., Rousselet, A., Janoueix-Lerosey, I.,

Goud, B. (1998) Interaction of a Golgi-associated kinesin-like protein with Rab6, *Science* **279**, 580–585.

Entchev, E.V., Gonzalez-Gaitan, M.A. (2002) Morphogen gradient formation and vesicular trafficking, *Traffic* **3**, 98–109.

Farsad, K., Ringstad, N., Takei, K., Floyd, S.R., Rose, K., De Camilli, P. (2001) Generation of high curvature membranes mediated by direct endophilin bilayer interactions, *J. Cell Biol.* **155**, 193–200.

Ferguson, S.S., Downey, W.E. III, Colapietro, A.M., Barak, L.S., Menard, L., Caron, M.G. (1996) Role of beta-arrestin in mediating agonist-promoted G protein-coupled receptor internalization, *Science* **271**, 363–366.

Fernandez-Borja, M., Wubbolts, R., Calafat, J., Janssen, H., Divecha, N., Dusseljee, S., Neefjes, J. (1999) Multivesicular body morphogenesis requires phosphatidyl-inositol 3- kinase activity, *Curr. Biol.* **9**, 55–58.

Floyd, S., De Camilli, P. (1998) Endocytosis proteins and cancer: a potential link? *Trends Cell Biol.* **8**, 299–301.

Folsch, H., Ohno, H., Bonifacino, J.S., Mellman, I. (1999) A novel clathrin adaptor complex mediates basolateral targeting in polarized epithelial cells, *Cell* **99**, 189–198.

Ford, M.G., Mills, I.G., Peter, B.J., Vallis, Y., Praefcke, G.J., Evans, P.R., McMahon, H.T. (2002) Curvature of clathrin-coated pits driven by epsin, *Nature* **419**, 361–366.

Fujimoto, L.M., Roth, R., Heuser, J.E., Schmid, S.L. (2000) Actin assembly plays a variable, but not obligatory role in receptor-mediated endocytosis in mammalian cells, *Traffic* **1**, 161–171.

Gagnon, E., Duclos, S., Rondeau, C., Chevet, E., Cameron, P.H., Steele-Mortimer, O., Paiement, J., Bergeron, J.J., Desjardins, M. (2002) Endoplasmic reticulum-mediated phagocytosis is a mechanism of entry into macrophages, *Cell* **110**, 119–131.

Gagnon, A.W., Kallal, L., Benovic, J.L. (1998) Role of clathrin-mediated endocytosis in agonist-induced down-regulation of the beta2-adrenergic receptor, *J. Biol. Chem.* **273**, 6976–6981.

Gaidarov, I., Keen, J.H. (1999) Phosphoinositide-AP-2 interactions required for targeting to plasma membrane clathrin-coated pits, *J. Cell Biol.* **146**, 755–764.

Gaidarov, I., Santini, F., Warren, R.A., Keen, J.H. (1999b) Spatial control of coated-pit dynamics in living cells, *Nat. Cell Biol.* **1**, 1–7.

Galan, J.M., Wiederkehr, A., Seol, J.H., Haguenauer-Tsapis, R., Deshaies, R.J., Riezman, H., Peter, M. (2001) Skp1p and the F-box protein Rcy1p form a non-SCF complex involved in recycling of the SNARE Snc1p in yeast, *Mol. Cell Biol.* **21**, 3105–3117.

Garcia, C.K., Wilund, K., Arca, M., Zuliani, G., Fellin, R., Maioli, M., Calandra, S., Bertolini, S., Cossu, F., Grishin, N., Barnes, R., Cohen, J.C., Hobbs, H.H. (2001) Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein, *Science* **292**, 1394–1398.

Garrus, J.E., von Schwedler, U.K., Pornillos, O.W., Morham, S.G., Zavitz, K.H., Wang, H.E., Wettstein, D.A., Stray, K.M., Cote, M., Rich, R.L., Myszka, D.G., Sundquist, W.I. (2001) Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding, *Cell* **107**, 55–65.

Ghosh, R.N., Maxfield, F.R. (1995) Evidence for nonvectorial, retrograde transferrin trafficking in the early endosomes of HEp2 cells, *J. Cell Biol.* **128**, 549–561.

Goodman, O.J., Krupnick, J.G., Santini, F., Gurevich, V.V., Penn, R.B., Gagnon, A.W., Keen, J.H., Benovic, J.L. (1996) Beta-arrestin acts as a clathrin adaptor in endocytosis of the beta2-adrenergic receptor, *Nature* **383**, 447–450.

Gottlieb, T.A., Ivanov, I.E., Adesnik, M., Sabatini, D.D. (1993) Actin microfilaments play a critical role in endocytosis at the apical but not the basolateral surface of polarized epithelial cells, *J. Cell Biol.* **120**, 695–710.

Grant, B., Zhang, Y., Paupard, M.C., Lin, S.X., Hall, D.H., Hirsh, D. (2001) Evidence that RME-1, a conserved C. elegans EH-domain protein, functions in endocytic recycling, *Nat. Cell Biol.* **3**, 573–579.

Griffiths, G., Hoflack, B., Simons, K., Mellman, I., Kornfeld, S. (1988) The mannose 6-phosphate receptor and the biogenesis of lysosomes, *Cell* **52**, 329–341.

Grimes, M.L., Zhou, J., Beattie, E.C., Yuen, E.C., Hall, D.E., Valletta, J.S., Topp, K.S., LaVail, J.H., Bunnett, N.W., Mobley, W.C. (1996) Endocytosis of activated TrkA: evidence that nerve growth factor induces formation of signaling endosomes, *J. Neurosci.* **16**, 7950–7964.

Gruenberg, J., Maxfield, F.R. (1995) Membrane transport in the endocytic pathway, *Curr. Opin. Cell Biol.* **7**, 552–563.

Gu, F., Gruenberg, J. (2000) ARF1 regulates pH-dependent COP functions in the early endocytic pathway, *J. Biol. Chem.* **275**, 8154–8160.

Hackstadt, T. (2000) Redirection of host vesicle trafficking pathways by intracellular parasites, *Traffic* **1**, 93–99.

Hao, M., Maxfield, F.R. (2000) Characterization of rapid membrane internalization and recycling, *J. Biol. Chem.* **275**, 15279–15286.

Hasson, T. (2003) Myosin VI: two distinct roles in endocytosis, *J. Cell Sci.* **116**, 3453–3461.

Haucke, V., De Camilli, P. (1999) AP-2 recruitment to synaptotagmin stimulated by tyrosine-based endocytic motifs, *Science* **285**, 1268–1271.

Heilker, R., Manning-Krieg, U., Zuber, J.F., Spiess, M. (1996) In vitro binding of clathrin adaptors to sorting signals correlates with endocytosis and basolateral sorting. *EMBO J.* **15**, 2893–2899.

Herskovits, J.S., Burgess, C.C., Obar, R.A., Vallee, R.B. (1993) Effects of mutant rat dynamin on endocytosis, *J. Cell Biol.* **122**, 565–578.

Heuser, J., Kirchhausen, T. (1985) Deep-etch views of clathrin assemblies, *J. Ultrastruct. Res.* **92**, 1–27.

Hicke, L., Riezman, H. (1996) Ubiquitination of a yeast plasma membrane receptor signals its ligand- stimulated endocytosis, *Cell* **84**, 277–287.

Hinshaw, J.E. (2000) Dynamin and its role in membrane fission, *Annu. Rev. Cell Dev. Biol.* **16**, 483–519.

Hinshaw, J.E., Schmid, S.L. (1995) Dynamin self-assembles into rings suggesting a mechanism for coated vesicle budding [see comments], *Nature* **374**, 190–192.

Holstein, S.E., Ungewickell, H., Ungewickell, E. (1996) Mechanism of clathrin basket dissociation: separate functions of protein domains of the DnaJ homologue auxilin, *J. Cell Biol.* **135**, 925–937.

Hopkins, C.R. (1983) Intracellular routing of transferrin and transferrin receptors in epidermoid carcinoma A431 cells, *Cell* **35**, 321–330.

Hunziker, W., Kraehenbuhl, J.P. (1998) Epithelial transcytosis of immunoglobulins, *J. Mamm. Gland Biol. Neoplasia* **3**, 287–302.

Hunziker, W., Male, P., Mellman, I. (1990) Differential microtubule requirements for transcytosis in MDCK cells, *EMBO J.* **9**, 3515–3525.

Ihrke, G., Neufeld, E.B., Meads, T., Shanks, M.R., Cassio, D., Laurent, M., Schroer, T.A., Pagano, R.E., Hubbard, A.L. (1993) WIF-B cells: an in vitro model for studies of hepatocyte polarity, *J. Cell Biol.* **123**, 1761–1775.

Ikonen, E., Parton, R.G. (2000) Caveolins and cellular cholesterol balance, *Traffic* **1**, 212–217.

Jiang, R., Gao, B., Prasad, K., Greene, L.E., Eisenberg, E. (2000) Hsc70 chaperones clathrin and primes it to interact with vesicle membranes, *J. Biol. Chem.* **275**, 8439–8447.

Jin, A.J., Nossal, R. (2000) Rigidity of triskelion arms and clathrin nets, *Biophys. J.* **78**, 1183–1194.

Kaneseki, T., Kadota, K. (1969) A morphological study of the coated vesicle isolated from the nerve endings of the guinea pig brain, with special reference to the mechanism of membrane movements, *J. Cell Biol.* **42**, 202.

Katz, M., Shtiegman, K., Tal-Or, P., Yakir, L., Mosesson, Y., Harari, D., Machluf, Y., Asao, H., Jovin, T., Sugamura, K., Yarden, Y. (2002) Ligand-independent degradation of epidermal growth factor receptor involves receptor ubiquitylation and Hgs, an adaptor whose ubiquitin-interacting motif targets ubiquitylation by Nedd4, *Traffic* **3**, 740–751.

Katzmann, D.J., Babst, M., Emr, S.D. (2001) Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I, *Cell* **106**, 145–155.

Katzmann, D.J., Stefan, C.J., Babst, M., Emr, S.D. (2003) Vps27 recruits ESCRT machinery to endosomes during MVB sorting, *J. Cell Biol.* **162**, 413–423.

Keen, J.H., Willingham, M.C., Pastan, I.H. (1979) Clathrin-coated vesicles: isolation, dissociation and factor-dependent reassociation of clathrin baskets, *Cell* **16**, 303–312.

Kirchhausen, T. (1999) Adaptors for clathrin-mediated traffic, *Annu. Rev. Cell Dev. Biol.* **15**, 705–732.

Kobayashi, T., Gu, F., Gruenberg, J. (1998) Lipids, lipid domains and lipid-protein interactions in endocytic membrane traffic, *Semin. Cell Dev. Biol.* **9**, 517–526.

Koenig, J.H., Ikeda, K. (1989) Disappearance and reformation of synaptic vesicle membrane upon transmitter release observed under

reversible blockage of membrane retrieval, *J. Neurosci.* **9**, 3844–3860.

Krauss, M., Kinuta, M., Wenk, M.R., De Camilli, P., Takei, K., Haucke, V. (2003) ARF6 stimulates clathrin/AP-2 recruitment to synaptic membranes by activating phosphatidylinositol phosphate kinase type I gamma, *J. Cell Biol.* **162**, 113–124.

Lee, E., De Camilli, P. (2002) Dynamin at actin tails, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 161–166.

Lefkowitz, R.J., Pitcher, J., Krueger, K., Daaka, Y. (1998) Mechanisms of beta-adrenergic receptor desensitization and resensitization, *Adv. Pharmacol.* **42**, 416–420.

Leheste, J.R., Rolinski, B., Vorum, H., Hilpert, J., Nykjaer, A., Jacobsen, C., Aucouturier, P., Moskaug, J.O., Otto, A., Christensen, E.I., Willnow, T.E. (1999) Megalin knockout mice as an animal model of low molecular weight proteinuria, *Am. J. Pathol.* **155**, 1361–1370.

Lemmon, S.K. (2001) Clathrin uncoating: Auxilin comes to life, *Curr. Biol.* **11**, R49–R52.

Letourneur, F., Klausner, R.D. (1992) A novel di-leucine motif and a tyrosine-based motif independently mediate lysosomal targeting and endocytosis of CD3 chains, *Cell* **69**, 1143–1157.

Leung, S.M., Ruiz, W.G., Apodaca, G. (2000) Sorting of membrane and fluid at the apical pole of polarized Madin-Darby canine kidney cells, *Mol. Biol. Cell* **11**, 2131–2150.

Lim, S.N., Bonzelius, F., Low, S.H., Wille, H., Weimbs, T., Herman, G.A. (2001) Identification of discrete classes of endosome-derived small vesicles as a major cellular pool for recycling membrane proteins, *Mol. Biol. Cell* **12**, 981–995.

Lin, S.X., Grant, B., Hirsh, D., Maxfield, F.R. (2001) Rme-1 regulates the distribution and function of the endocytic recycling compartment in mammalian cells, *Nat. Cell Biol.* **3**, 567–572.

Lindsay, A.J., McCaffrey, M.W. (2002) Rab11-FIP2 functions in transferrin recycling and associates with endosomal membranes via its COOH-terminal domain, *J. Biol. Chem.* **277**, 27193–27199.

Lippe, R., Horiuchi, H., Runge, A., Zerial, M. (2001a) Expression, purification, characterization of Rab5 effector complex, rabaptin-5/rabex-5, *Methods Enzymol.* **329**, 132–145.

Lippe, R., Miaczynska, M., Rybin, V., Runge, A., Zerial, M. (2001b) Functional synergy between Rab5 effector Rabaptin-5 and exchange factor Rabex-5 when physically associated in a complex, *Mol. Biol. Cell* **12**, 2219–2228.

Liu, P., Rudick, M., Anderson, R.G. (2002) Multiple functions of caveolin-1, *J. Biol. Chem.* **277**, 41295–41298.

Ludwig, T., Griffiths, G., Hoflack, B. (1991) Distribution of newly synthesized lysosomal enzymes in the endocytic pathway of normal rat kidney cells, *J. Cell Biol.* **115**, 1561–1572.

Luzio, J.P., Mullock, B.M., Pryor, P.R., Lindsay, M.R., James, D.E., Piper, R.C. (2001) Relationship between endosomes and lysosomes, *Biochem. Soc. Trans.* **29**, 476–480.

Luzio, J.P., Rous, B.A., Bright, N.A., Pryor, P.R., Mullock, B.M., Piper, R.C. (2000) Lysosome-endosome fusion and lysosome biogenesis, *J. Cell Sci.* **113**, 1515–1524.

Mack, U., Storey, E.L., Powell, L.W., Halliday, J.W. (1985) Characterization of the binding of ferritin to the rat liver ferritin receptor, *Biochim. Biophys. Acta* **843**, 164–170.

Maranda, B., Brown, D., Bourgoin, S., Casanova, J.E., Vinay, P., Ausiello, D.A., Marshansky, V. (2001) Intra-endosomal pH-sensitive recruitment of the Arf-nucleotide exchange factor ARNO and Arf6 from cytoplasm to proximal tubule endosomes, *J. Biol. Chem.* **276**, 18540–18550.

Marshansky, V., Ausiello, D.A., Brown, D. (2002) Physiological importance of endosomal acidification: potential role in proximal tubulopathies, *Curr. Opin. Nephrol. Hypertens.* **11**, 527–537.

Mayor, S., Presley, J.F., Maxfield, F.R. (1993) Sorting of membrane components from endosomes and subsequent recycling to the cell surface occurs by a bulk flow process, *J. Cell Biol.* **121**, 1257–1269.

McBride, H.M., Rybin, V., Murphy, C., Giner, A., Teasdale, R., Zerial, M. (1999) Oligomeric complexes link Rab5 effectors with NSF and drive membrane fusion via interactions between EEA1 and syntaxin 13, *Cell* **98**, 377–386.

McDowall, A., Gruenberg, J., Romisch, K., Griffiths, G. (1989) The structure of organelles of the endocytic pathway in hydrated cryosections of cultured cells, *Eur. J. Cell Biol.* **49**, 281–294.

McLauchlan, H., Newell, J., Morrice, N., Osborne, A., West, M., Smythe, E. (1998) A novel role for Rab5-GDI in ligand sequestration into clathrin-coated pits, *Curr. Biol.* **8**, 34–45.

Mellman, I., Warren, G. (2000) The road taken: past and future foundations of membrane traffic, *Cell* **100**, 99–112.

Meresse, S., Steele-Mortimer, O., Moreno, E., Desjardins, M., Finlay, B., Gorvel, J.P. (1999) Controlling the maturation of pathogen-containing vacuoles: a matter of life and death, *Nat. Cell Biol.* **1**, E183–E188.

Merrifield, C.J., Feldman, M.E., Wan, L., Almers, W. (2002) Imaging actin and dynamin recruitment during invagination of single clathrin-coated pits, *Nat. Cell Biol.* **4**, 691–698.

Merrifield, C.J., Moss, S.E., Ballestrem, C., Imhof, B.A., Giese, G., Wunderlich, I., Almers, W. (1999) Endocytic vesicles move at the tips of actin tails in cultured mast cells, *Nat. Cell Biol.* **1**, 72–74.

Miaczynska, M., Zerial, M. (2002) Mosaic organization of the endocytic pathway, *Exp. Cell Res.* **272**, 8–14.

Mintz, L., Galperin, E., Pasmanik-Chor, M., Tulzinsky, S., Bromberg, Y., Kozak, C.A., Joyner, A., Fein, A., Horowitz, M. (1999) EHD1–an EH-domain-containing protein with a specific expression pattern, *Genomics* **59**, 66–76.

Mishra, S.K., Watkins, S.C., Traub, L.M. (2002) The autosomal recessive hypercholesterolemia (ARH) protein interfaces directly with the clathrin-coat machinery, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16099–16104.

Mostov, K.E., Verges, M., Altschuler, Y. (2000) Membrane traffic in polarized epithelial cells, *Curr. Opin. Cell Biol.* **12**, 483–490.

Mu, F.T., Callaghan, J.M., Steele-Mortimer, O., Stenmark, H., Parton, R.G., Campbell, P.L., McCluskey, J., Yeo, J.P., Tock, E.P., Toh, B.H. (1995) EEA1, an early endosome-associated protein. EEA1 is a conserved alpha- helical peripheral membrane protein flanked by cysteine "fingers" and contains a calmodulin-binding IQ motif, *J. Biol Chem.* **270**, 13503–13511.

Mukherjee, S., Ghosh, R.N., Maxfield, F.R. (1997) Endocytosis, *Physiol. Rev.* **77**, 759–803.

Mukherjee, S., Soe, T.T., Maxfield, F.R. (1999) Endocytic sorting of lipid analogues differing solely in the chemistry of their hydrophobic tails, *J. Cell Biol.* **144**, 1271–1284.

Mullins, C., Bonifacino, J.S. (2001) The molecular machinery for lysosome biogenesis, *BioEssays* **23**, 333–343.

Mullock, B.M., Bright, N.A., Fearon, C.W., Gray, S.R., Luzio, J.P. (1998) Fusion of lysosomes with late endosomes produces a hybrid organelle of intermediate density and is NSF dependent, *J. Cell Biol.* **140**, 591–601.

Mullock, B.M., Perez, J.H., Kuwana, T., Gray, S.R., Luzio, J.P. (1994) Lysosomes can fuse with a late endosomal compartment in a cell-free system from rat liver, *J. Cell Biol.* **126**, 1173–1182.

Newmyer, S.L., Schmid, S.L. (2001) Dominant-interfering Hsc70 mutants disrupt multiple stages of the clathrin-coated vesicle cycle in vivo, *J. Cell Biol.* **152**, 607–620.

Nielsen, E., Severin, F., Backer, J.M., Hyman, A.A., Zerial, M. (1999) Rab5 regulates motility of early endosomes on microtubules, *Nat. Cell Biol.* **1**, 376–382.

Nishi, T., Forgac, M. (2002) The vacuolar (H+)-ATPases – nature's most versatile proton pumps, *Nat. Rev. Mol. Cell Biol.* **3**, 94–103.

Odorizzi, G., Babst, M., Emr, S.D. (1998) Fab1p PtdIns(3)P 5-kinase function essential for protein sorting in the multivesicular body, *Cell* **95**, 847–858.

Odorizzi, G., Pearse, A., Domingo, D., Trowbridge, I.S., Hopkins, C.R. (1996) Apical and basolateral endosomes of MDCK cells are interconnected and contain a polarized sorting mechanism, *J. Cell Biol.* **135**, 139–152.

Oh, P., McIntosh, D.P., Schnitzer, J.E. (1998) Dynamin at the neck of caveolae mediates their budding to form transport vesicles by GTP-driven fission from the plasma membrane of endothelium, *J. Cell Biol.* **141**, 101–114.

Orlando, R.A., Rader, K., Authier, F., Yamazaki, H., Posner, B.I., Bergeron, J.J., Farquhar, M.G. (1998) Megalin is an endocytic receptor for insulin, *J. Am. Soc. Nephrol.* **9**, 1759–1766.

Orth, J.D., Krueger, E.W., Cao, H., McNiven, M.A. (2002) The large GTPase dynamin regulates actin comet formation and movement in living cells, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 167–172.

Pearse, B.M.F. (1975) Coated vesicles from the pig brain: purification and biochemical characterization, *J. Mol. Biol.* **97**, 93.

Pelkmans, L., Helenius, A. (2002) Endocytosis via caveolae, *Traffic* **3**, 311–320.

Pelkmans, L., Kartenbeck, J., Helenius, A. (2001) Caveolar endocytosis of simian virus 40 reveals a new two-step vesicular-transport pathway to the ER, *Nat. Cell Biol.* **3**, 473–483.

Pfeffer, S.R. (1999) Transport-vesicle targeting: tethers before SNAREs, *Nat. Cell Biol.* **1**, E17–E22.

Pillay, C.S., Elliott, E., Dennison, C. (2002) Endolysosomal proteolysis and its regulation, *Biochem. J.* **363**, 417–429.

Pishvaee, B., Costaguta, G., Yeung, B.G., Ryazantsev, S., Greener, T., Greene, L.E., Eisenberg, E., McCaffery, J.M., Payne, G.S. (2000) A yeast DNA J protein required for uncoating of clathrin-coated vesicles in vivo, *Nat. Cell Biol.* **2**, 958–963.

Pisoni, R.L., Thoene, J.G. (1991) The transport systems of mammalian lysosomes, *Biochim. Biophys. Acta* **1071**, 351–373.

Ponka, P., Lok, C.N. (1999) The transferrin receptor: role in health and disease, *Int. J. Biochem. Cell Biol.* **31**, 1111–1137.

Prasad, K., Barouch, W., Greene, L., Eisenberg, E. (1993) A protein cofactor is required for uncoating of clathrin baskets by uncoating ATPase, *J. Biol. Chem.* **268**, 23758–23761.

Raiborg, C., Bache, K.G., Gillooly, D.J., Madshus, I.H., Stang, E., Stenmark, H. (2002) Hrs sorts ubiquitinated proteins into clathrin-coated microdomains of early endosomes, *Nat. Cell Biol.* **4**, 394–398.

Raymond, C.K., Howald-Stevenson, I., Vater, C.A., Stevens, T.H. (1992) Morphological classification of the yeast vacuolar protein sorting mutants: evidence for a prevacuolar compartment in class E vps mutants, *Mol. Biol. Cell* **3**, 1389–1402.

Reggiori, F., Pelham, H.R. (2002) A transmembrane ubiquitin ligase required to sort membrane proteins into multivesicular bodies, *Nat. Cell Biol.* **4**, 117–123.

Ringstad, N., Gad, H., Low, P., Di Paolo, G., Brodin, L., Shupliakov, O., De Camilli, P. (1999) Endophilin/SH3p4 is required for the transition from early to late stages in clathrin-mediated synaptic vesicle endocytosis, *Neuron* **24**, 143–154.

Robinson, L.J., Aniento, F., Gruenberg, J. (1997) NSF is required for transport from early to late endosomes, *J. Cell Sci.* **110**, 2079–2087.

Robinson, M.S. (1987) Coated vesicles and protein sorting, *J. Cell Sci.* **87**, 203–204.

Robinson, M.S., Pearse, B.M. (1986) Immunofluorescent localization of 100K coated vesicle proteins, *J. Cell Biol.* **102**, 48–54.

Roth, T.F., Porter, K.R. (1964) Yolk protein uptake in the oocyte of the mosquito Aedes aegypti, *J. Cell Biol.* **20**, 313.

Rothberg, K.G., Heuser, J.E., Donzell, W.C., Ying, Y.S., Glenney, J.R., Anderson, R.G. (1992) Caveolin, a protein component of caveolae membrane coats, *Cell* **68**, 673–682.

Rubino, M., Miaczynska, M., Lippe, R., Zerial, M. (2000) Selective membrane recruitment of EEA1 suggests a role in directional transport of clathrin-coated vesicles to early endosomes, *J. Biol. Chem.* **275**, 3745–3748.

Santini, F., Gaidarov, I., Keen, J.H. (2002) G protein-coupled receptor/arrestin3 modulation of the endocytic machinery, *J. Cell Biol.* **156**, 665–676.

Santini, F., Keen, J.H. (1996) Endocytosis of activated receptors and clathrin-coated pit formation: deciphering the chicken or egg relationship, *J. Cell Biol.* **132**, 1025–1036.

Schmid, S.L., Fuchs, R., Male, P., Mellman, I. (1988) Two distinct subpopulations of endosomes involved in membrane recycling and transport to lysosomes, *Cell* **52**, 73–83.

Schmid, S.L., Smythe, E. (1991) Stage-specific assays for coated pit formation and coated vesicle budding in vitro, *J. Cell Biol.* **114**, 869–880.

Schu, P.V., Takegawa, K., Fry, M.J., Stack, J.H., Waterfield, M.D., Emr, S.D. (1993) Phosphatidylinositol 3-kinase encoded by yeast VPS34 gene essential for protein sorting, *Science* **260**, 88–91.

Seto, E.S., Bellen, H.J., Lloyd, T.E. (2002) When cell biology meets development: endocytic regulation of signaling pathways, *Genes Dev.* **16**, 1314–1336.

Sever, S., Damke, H., Schmid, S.L. (2000) Garrotes, springs, ratchets, and whips: putting dynamin models to the test, *Traffic* **1**, 385–392.

Sibley, L.D., Andrews, N.W. (2000) Cell invasion by un-palatable parasites, *Traffic* **1**, 100–106.

Simonsen, A., Lippe, R., Christoforidis, S., Gaullier, J.M., Brech, A., Callaghan, J., Toh, B.H., Murphy, C., Zerial, M., Stenmark, H. (1998) EEA1 links PI(3)K function to Rab5 regulation of endosome fusion, *Nature* **394**, 494–498.

Sinai, A.P., Joiner, K.A. (1997) Safe haven: the cell biology of nonfusogenic pathogen vacuoles, *Annu. Rev. Microbiol.* **51**, 415–462.

So, C.W., Caldas, C., Liu, M.M., Chen, S.J., Huang, Q.H., Gu, L.J., Sham, M.H., Wiedemann, L.M., Chan, L.C. (1997) EEN encodes for a member of a new family of proteins containing an Src homology 3 domain and is the

third gene located on chromosome 19p13 that fuses to MLL in human leukemia, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2563–2568.

Sonnichsen, B., De Renzis, S., Nielsen, E., Rietdorf, J., Zerial, M. (2000) Distinct membrane domains on endosomes in the recycling pathway visualized by multicolor imaging of Rab4, Rab5, and Rab11, *J. Cell Biol.* **149**, 901–914.

Stenmark, H., Parton, R.G., Steele-Mortimer, O., Lutcke, A., Gruenberg, J., Zerial, M. (1994) Inhibition of rab5 GTPase activity stimulates membrane fusion in endocytosis, *EMBO J.* **13**, 1287–1296.

Stevens, T.H., Forgac, M. (1997) Structure, function and regulation of the vacuolar (H+)-ATPase, *Annu. Rev. Cell Dev. Biol.* **13**, 779–808.

Strick, D.J., Francescutti, D.M., Zhao, Y., Elferink, L.A. (2002) Mammalian suppressor of Sec4 modulates the inhibitory effect of Rab15 during early endocytosis, *J. Biol. Chem.* **277**, 32722–32729.

Takei, K., Haucke, V., Slepnev, V., Farsad, K., Salazar, M., Chen, H., De Camilli, P. (1998) Generation of coated intermediates of clathrin-mediated endocytosis on protein-free liposomes, *Cell* **94**, 131–141.

Tanaka, K., Matsui, Y. (2001) Functions of unconventional myosins in the yeast Saccharomyces cerevisiae, *Cell Struct. Funct.* **26**, 671–675.

Taunton, J. (2001) Actin filament nucleation by endosomes, lysosomes and secretory vesicles, *Curr. Opin. Cell Biol.* **13**, 85–91.

Tjelle, T.E., Brech, A., Juvet, L.K., Griffiths, G., Berg, T. (1996) Isolation and characterization of early endosomes, late endosomes and terminal lysosomes: their role in protein degradation, *J. Cell Sci.* **109**, 2905–2914.

Tsao, P.I., von Zastrow, M. (2000) Type-specific sorting of G protein-coupled receptors after endocytosis, *J. Biol. Chem.* **275**, 11130–11140.

Ungewickell, E., Branton, D. (1981) Assembly units of clathrin coats, *Nature* **289**, 420–422.

Urbanowski, J.L., Piper, R.C. (2001) Ubiquitin sorts proteins into the intralumenal degradative compartment of the late-endosome/vacuole, *Traffic* **2**, 622–630.

van der Bliek, A.M., Meyerowitz, E.M. (1991) Dynamin-like protein encoded by the Drosophila shibire gene associated with vesicular traffic, *Nature* **351**, 411–414.

van der Sluijs, P., Hull, M., Webster, P., Male, P., Goud, B., Mellman, I. (1992) The small GTP-binding protein rab4 controls an early sorting event on the endocytic pathway, *Cell* **70**, 729–740.

Vigers, G.P., Crowther, R.A., Pearse, B.M. (1986) Location of the 100 kd-50 kd accessory proteins in clathrin coats, *EMBO J.* **5**, 2079–2085.

von Zastrow, M., Kobilka, B.K. (1994) Antagonist -dependent and -independent steps in the mechanism of adrenergic receptor internalization, *J. Biol. Chem.* **269**, 18448–18452.

Wallace, D.M., Lindsay, A.J., Hendrick, A.G., McCaffrey, M.W. (2002) The novel Rab11-FIP/Rip/RCP family of proteins displays extensive homo- and hetero-interacting abilities, *Biochem. Biophys. Res. Commun.* **292**, 909–915.

Wang, G., Deering, C., Macke, M., Shao, J., Burns, R., Blau, D.M., Holmes, K.V., Davidson, B.L., Perlman, S., McCray, P.B., Jr. (2000) Human coronavirus 229E infects polarized airway epithelia from the apical surface, *J. Virol.* **74**, 9234–9239.

Watson, F.L., Heerssen, H.M., Bhattacharyya, A., Klesse, L., Lin, M.Z., Segal, R.A. (2001) Neurotrophins use the Erk5 pathway to mediate a retrograde survival response, *Nat. Neurosci.* **4**, 981–988.

West, M.A., Bright, N.A., Robinson, M.S. (1997) The role of ADP-ribosylation factor and phospholipase D in adaptor recruitment, *J. Cell Biol.* **138**, 1239–1254.

Whiteheart, S.W., Rossnagel, K., Buhrow, S.A., Brunner, M., Jaenicke, R., Rothman, J.E. (1994) N-ethylmaleimide-sensitive fusion protein: a trimeric ATPase whose hydrolysis of ATP is required for membrane fusion, *J. Cell Biol.* **126**, 945–954.

Wiederkehr, A., Avaro, S., Prescianotto-Baschong, C., Haguenauer-Tsapis, R., Riezman, H. (2000) The F-box protein Rcy1p is involved in endocytic membrane traffic and recycling out of an early endosome in Saccharomyces cerevisiae, *J. Cell Biol.* **149**, 397–410.

Wilde, A., Brodsky, F.M. (1996) In vivo phosphorylation of adaptors regulates their interaction with clathrin, *J. Cell Biol.* **135**, 635–645.

Winchester, B., Vellodi, A., Young, E. (2000) The molecular basis of lysosomal storage diseases and their treatment, *Biochem. Soc. Trans.* **28**, 150–154.

Yamashiro, D.J., Tycko, B., Fluss, S.R., Maxfield, F.R. (1984) Segregation of transferrin to a mildly acidic (pH 6.5) para-Golgi compartment in the recycling pathway, *Cell* **37**, 789–800.

Yarden, Y. (2001) The EGFR family and its ligands in human cancer. signalling mechanisms and therapeutic opportunities, *Eur. J. Cancer*. **37**(Suppl. 4), S3–S8.

Zaremba, S., Keen, J.H. (1983) Assembly polypeptides from coated vesicles mediate reassembly of unique clathrin coats, *J. Cell Biol*. **97**, 1339–1347.

Zhang, J.Z., Davletov, B.A., Sudhof, T.C., Anderson, R.G. (1994) Synaptotagmin I is a high affinity receptor for clathrin AP-2: implications for membrane recycling, *Cell* **78**, 751–760.

Zhu, Y., Traub, L.M., Kornfeld, S. (1998) ADP-ribosylation factor 1 transiently activates high-affinity adaptor protein complex AP-1 binding sites on Golgi membranes, *Mol. Biol. Cell* **9**, 1323–1337.

Zuk, P.A., Elferink, L.A. (2000) Rab15 differentially regulates early endocytic trafficking, *J. Biol Chem*. **275**, 26754–26764.

Engineering, Bioprocess: *see* Bioprocess Engineering

Epidemiology, Molecular

Paul A. Schulte[1], *Frederica P. Perera*[2], *and Nathaniel Rothman*[3]
[1] *National Institute for Occupational Safety and Health, Cincinnati, USA*
[2] *Columbia University, NY, USA*
[3] *National Cancer Institute, Rethesda, USA*

1 Principles 226
1.1 Biological Markers of Exposure 227
1.2 Biologic Markers of Effect 228
1.3 Biologic Markers of Susceptibility 228
1.4 Utility of Molecular Epidemiology 229

2 Techniques 229
2.1 Representational Validity of Molecular Biological Markers 230
2.1.1 Validation of the Relationship between Exposure and Dose 230
2.1.2 Validation of the Relationship between Biological Effects and Disease 230
2.1.3 Validation of Markers of Susceptibility 231
2.2 Validation of the Behavior of Molecular Biological Markers 232
2.2.1 Developmental/Characterization Studies 232
2.2.2 Applied Studies 232
2.3 Etiologic Studies 233
2.3.1 Case–control Studies 233
2.3.2 Case–case Studies 233
2.3.3 Prospective Cohort Studies 234
2.4 Molecular Epidemiology and Genomics 234
2.5 Public Health Applications 234

Bibliography 235
Books and Reviews 235
Primary Literature 235

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Keywords

Biological Markers (Biomarkers)
Biochemical, molecular, genetic, immunologic, or physiologic signals of events in biological systems.

Biologic Markers of Effect
A measurable cellular, biochemical, or molecular alteration within an organism that, depending on magnitude, can be recognized as an established or potential health impairment or disease.

Biologic Markers of Exposure
A xenobiotic chemical or its metabolite, or the product of an interaction between a chemical, physical, or biologic agent and some target cell or biomolecule.

Biologic Markers of Susceptibility
An inherited or acquired indicator of the response of an individual or a population to a specific xenobiotic agent.

Molecular epidemiology is the use of molecular biological techniques to identify exposures, effects, or susceptibility factors in studies of human populations. Molecular epidemiology and traditional epidemiology utilize the same paradigm. However, the former presents the opportunity to use the enhanced resolving power of molecular biology in the assessment of exposure–disease relationships.

The resolving power, to elucidate a continuum of events between xenobiotic exposure and disease, can provide stronger approaches to research, prevention, and intervention. One particular aspect of the new resolving power of molecular epidemiology is the application of the products of genomic research to epidemiology to assess the genetic components of disease and the interaction between genetic and environmental factors in disease causation. Conversely, the molecular epidemiologic approach may contribute to genomics research by emphasizing the importance of populations and a population perspective. Genetic biomarkers reflect population dynamics and represent useful tools in uncovering complicated interrelationships between environment, culture, and genetics in human history.

1 Principles

The use of molecular biological techniques in epidemiology provides a potentially powerful tool for medical and public health researchers. These techniques allow for the identification of biological markers (Fig. 1) that can indicate exposure to a xenobiotic agent, reveal a biological effect early in the natural history of disease, or represent unique disease subtypes or susceptibility

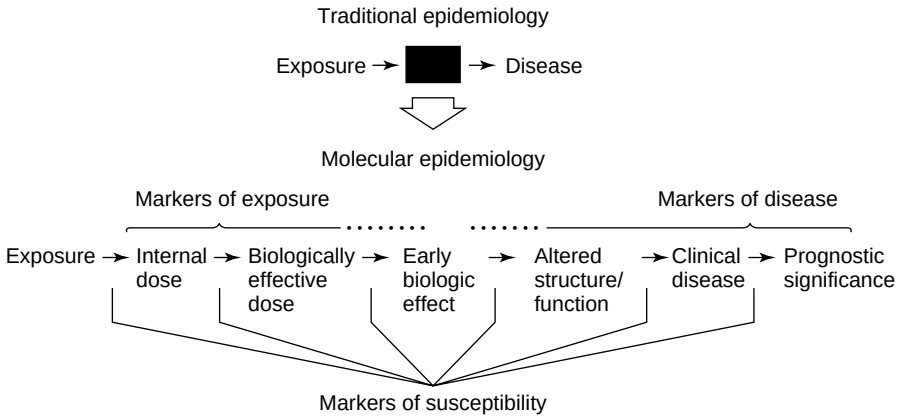

Fig. 1 Enhancement of the traditional epidemiologic paradigm by the use of biological markers resulting in a molecular epidemiologic approach. In traditional epidemiology, the mechanism of action is often a "black box," and associations between an exposure and disease are made by inference. In molecular epidemiology, a continuum between an exposure and a disease is defined, and various markers are identified.

to the development of disease. Although the use of biological markers in epidemiology is not new, the current generation of molecular biological markers enhances past approaches. Epidemiology is an observational science: one makes inferences about disease and health on the basis of comparing groups of people in terms of disease incidence and mortality. Ideally, the groups being compared should be similar in all respects except for the risk factor in question. A benefit of molecular epidemiology is that instead of comparing two groups on the basis of environmental exposure, investigators can compare populations with respect to dose of an environmental agent as measured in critical macromolecules, such as DNA or surrogate proteins for DNA under some conditions, this is presumably a more accurate means of classifying the subjects' true exposure. At the other end of the exposure-disease continuum, instead of using frank disease as an outcome variable in an epidemiologic study, it is possible to use a validated biologic marker of effect.

1.1 Biological Markers of Exposure

The utility of a biological marker of exposure depends, in part, on its half-life, the pattern of the exposure it is measuring (e.g. regular daily exposure vs infrequent episodic exposure), and whether secular trends have occurred in that exposure (e.g. smoking cessation). In addition, the information it provides must be compared with the availability and quality of other sources of data (e.g. questionnaires, environmental monitoring, medical records). Essentially, all exposure measures misclassify some subjects – it is the relative ability of different sources of data to correctly place individuals into exposure categories that must be considered. For example, subjects can generally report average smoking habits and smoking duration in an accurate manner, permitting the calculation of cumulative exposure. Since internal dose markers associated with tobacco smoke have relatively short half-lives, and thus reflect only recent exposure (e.g. cotinine,

a metabolite of nicotine), they have limited utility by themselves to directly assess a smoking–cancer relationship.

In contrast, it is difficult to obtain accurate information about dietary exposure to aflatoxin by using a questionnaire because exposure is sporadic and is present in a spectrum of different food types. In this instance, even a short-term internal dose marker might classify the long-term exposure status of subjects more accurately than questionnaire data. For example, a nested case–control study conducted in Shanghai, China, demonstrated that aflatoxin exposure assessed by measuring several aflatoxin metabolites (and an N_7-guanine aflatoxin adduct) in banked urine samples was associated with an increased risk of hepatocellular carcinoma, while aflatoxin exposure assessed by questionnaire was not associated with elevated risk.

1.2 Biologic Markers of Effect

In terms of disease, techniques are available for identifying biological changes earlier in the continuum between homeostatic response to pathological agents or conditions and development of frank disease. This advance has implications for identifying opportunities for prevention. Biologic markers intermediate between exposure and disease, if validated for disease or risk of disease, can be used to screen people or to allow for early disease detection. These markers can also be used in intervention trials as outcome indicators rather than waiting for subjects to develop disease. There is a popular notion that molecular epidemiology has the potential to contribute to assessment of risk of an individual. However, that point needs clarification. Historically, it has been possible to use an epidemiologic data set, that is, data on a group consisting of sick and healthy people, with and without certain risk factors, to develop risk functions that provide for an estimate of individual risk. This was accomplished in the 1960s using data from the Framingham Longitudinal Study of cardiovascular disease. On the basis of the knowledge of the risk for people with a certain aggregation of characteristics, it was possible to predict the risk for an individual. Molecular epidemiologic approaches provide a means for more confident estimation because more mechanistic information can be utilized. Still, the resultant assessment is only probabilistic determination, not deterministic. While the prediction is that a certain individual is likely to develop disease in the future, there is no guarantee that he or she will develop it. A biomarker may have the utility to screen populations at high risk of disease as part of a primary or secondary prevention effort. However, a substantial amount of information is required before a biomarker can be used for this purpose. In particular, the cumulative probability that an individual will develop disease over a defined period, given a constellation of biologic and nonbiologic risk factors, must be estimated along with a calculation of its uncertainty.

1.3 Biologic Markers of Susceptibility

One of the greatest contributions of molecular epidemiology is the ability to discern the role of host factors, particularly genetic factors, in accounting for variation in response. Why similarly exposed people do not get the same diseases is a target question for molecular epidemiology. In most disease systems, susceptibility markers are being identified and evaluated. These markers can be incorporated

into epidemiologic models as modifiers of the relationship between an exposure and an effect (see Sect. 2.1.3). Traditional epidemiologic approaches can also be enhanced by using molecular genetic techniques to identify host factors that could account for differences in disease risk. Thus, for example, a metabolic polymorphism can be detected from peripheral blood lymphocyte DNA, and groups at potentially greater and lesser risk of disease can be distinguished. The category of markers of susceptibility includes polymorphisms in genes responsible for chemical activation or detoxification, DNA repair, and genomic stability. Susceptibility genes of some types may interact with chemical exposures of very specific types (e.g. cytochrome P450 enzyme subtypes and phase II conjugating enzymes), while others may confer more general susceptibility (e.g. p53 mutations in Li–Fraumeni syndrome). Markers can be measured at the DNA level, if the genetic basis of a polymorphic phenotype has been identified, or at the phenotypic level (e.g. drug probes of hepatic enzyme activity, DNA repair measured in peripheral lymphocytes).

1.4 Utility of Molecular Epidemiology

Molecular epidemiology is a useful temporary blanket term that reflects the reality that increasingly disease, causal exposures, and risk factors are being defined at the molecular level. The utility of this term is that it serves as a signpost for epidemiologists to consider using independent (risk factor) and dependent (outcome) variables that are derived from molecular biological techniques and assays. So, for example, DNA adducts may be used in addition to breathing-zone measurement of a carcinogen, gene mutational patterns may be used as an indicator of disease rather than a nosological death certificate, and genotyping based on the polymerase chain reaction can be used in addition to race and sex to stratify populations for comparisons. In summary, molecular epidemiology has the potential to contribute the following opportunities and capabilities to public health:

- Delineation of a continuum of events between an exposure and a resultant disease
- Identification of exposures to smaller amounts of xenobiotics and enhanced dose reconstruction
- Identification of events earlier in the natural history of clinical diseases and on a smaller scale
- Reduction of misclassification of dependent and independent variables
- Indication of mechanisms by which an exposure and a disease are related
- Better accounting for variability and effect modification
- Enhanced individual and group risk assessments

2 Techniques

If molecular biological markers in epidemiology and public health can be used in the ways described in Sect. 1, they must be demonstrated to be valid in terms of both the assay and the marker. An assay will be valid if it measures what it is expected to measure. A marker will be valid for epidemiologic purposes in two ways. First, it will be valid to the extent that it represents exposure, disease, or susceptibility. Second, it will be valid insofar as the extent of variation in groups with different demographic, behavioral, or medical characteristics is known. A key question

is the prevalence of a particular biomarker (e.g. a mutation) in different ethnic or racial groups, in smokers or drinkers, or in people with various hereditary and acquired diseases. Molecular epidemiologic approaches can be useful in the validation of molecular biological markers in these two ways.

2.1 Representational Validity of Molecular Biological Markers

2.1.1 Validation of the Relationship between Exposure and Dose

Exposure → Internal dose

→ Biologically effective dose

Molecular markers of exposure may be validated by assessing the relation of an exogenous exposure to an internal dose or a biologically effective dose. Critical in validation studies is the need to have an effective exposure assessment. It may be necessary to use a combination of personal and environmental monitoring and questionnaires, record review, and modeling to reconstruct exposure history. The approach also requires an understanding of the pharmacokinetics associated with the particular xenobiotics. Related to this is the need to understand the natural history of the marker and to use it in the validation study. For example, in a study of hydroxyethyl hemoglobin adducts in workers exposed to ethylene oxide, the life span of the erythrocyte, hence the constituent hemoglobin molecule (~4 months), was used as a dosimeter of cumulative exposure. There is also a need to account for factors that might influence the appearance of a molecular biological marker. In the aforementioned studies, when mean values were adjusted for important covariants such as age, cigarette smoking, and education, an exposure–response relationship was found at levels below the permissible exposure level.

2.1.2 Validation of the Relationship between Biological Effects and Disease

Early biologic effect → Altered structure/

function → Clinical disease

Validation information in the biological effects–disease category is limited. The often-repeated question is, "What do the data concerning health and disease mean?" Validation studies of these types are difficult to accomplish because of the temporal factor. Identification of an early effect – that is, an effect in pathogenesis or an effect predictive of disease – generally requires a prospective study design, although cross-sectional clinical studies of diseased and heavily exposed individuals can be used to great advantage. When a prospective design is not used, however, care must be taken to avoid biased associations. This is often difficult; hence prospective studies are the best approach for validation. Prospective studies are expensive and time-consuming, and few are conducted. For example, despite the large number of studies on most of the cytogenetic markers, there is little consensus on their predictive value because most of the studies have been cross-sectional and markers were not linked to disease. Specifically, in epidemiologic terms, predictive value is evaluated in terms of the percentage of those who test positive for a marker and actually develop the disease. Performing the appropriate prospective studies of sister-chromatid exchanges would take a large population and a relatively long time. The best example of such a study is the Nordic prospective study on the relationship between

peripheral lymphocyte chromosome damage and cancer morbidity in occupational groups. Ten laboratories in four Nordic countries participated in a study of a combined cohort of persons (mostly from occupational groups) who had been cytogenetically tested. The cohort is being followed prospectively for cancer morbidity. The cohort comprises 3190 subjects, of whom 1986 subjects (62%) have been scored for chromosome aberrations and 2024 subjects (63%) scored for sister-chromatid exchanges. Preliminary analysis indicates that chromosomal aberrations are associated with cancer. These biologic markers in peripheral lymphocytes represent carcinogenic changes elsewhere in the body. This is the critical criterion of a useful biologic marker. To serve as a valid outcome measure, an intermediate marker must be correlated with disease risk. The criteria for validating intermediate biomarkers have been extensively discussed and include the sensitivity of the marker (i.e. the proportion of subjects who develop cancer and who are positive for the biomarker), the relative risk (a measure of the strength) of the association between the marker and disease, and a judgment about the extent to which the exposure–disease relationship is mediated through a process reflected directly or indirectly by the marker.

2.1.3 Validation of Markers of Susceptibility

Exposure → Susceptibility → Disease

The tools of molecular biology and analytical chemistry have allowed researchers to identify a degree of interindividual variability not previously imagined. Validated biological markers of susceptibility can serve as effect modifiers in epidemiologic studies. Effect modification is a term with statistical and biological aspects. Statistically, the examination of the joint effects of two or more reactors is often discussed in the context of effect modification. It depends on the statistical method (e.g. multiplicative or additive) used to model interaction. From the biological perspective, effect modification contributes to answering the question of why all similarly exposed individuals do not develop a disease. The answer, in part, lies in individual variability in metabolic and detoxification capabilities, their ability to repair genetic damage, or other host factors.

To validate a susceptibility marker, it is important to minimize misclassification, which can occur as a result of laboratory or epidemiologic factors that affect phenotyping or genotyping. Next, it is necessary to demonstrate that the marker increases the risk of disease. The issue of the correlation of acetylation phenotype and bladder cancer from aromatic amines illustrates the concept of susceptibility. Some aromatic amines are detoxified by the enzyme *N*-acetyltransferase, and the slow phenotype of this enzyme has been associated with bladder cancer in exposed individuals. Despite a plethora of studies, the scientific literature is not conclusive on the extent to which being a slow acetylator modifies the risk for bladder cancer in people exposed to aromatic amines. Generally, most studies have been too small, have had weak exposure characterization, and were not designed to allow proper determination of whether exposure or susceptibility was the key factor. An example of how partial validation of a susceptibility marker might occur without using disease as the outcome involved the formation of hemoglobin adducts (which are documented surrogates for DNA adducts

believed to be involved in carcinogenesis) in slow and fast acetylators who had been exposed to 4-aminobiphenyl. Slow acetylators had an average of 1.5-fold greater frequency of adducts than the fast acetylators. Despite these encouraging efforts at validation, few markers of susceptibility have been validated and none are ready for use in population screening.

2.2 Validation of the Behavior of Molecular Biological Markers

Before biomarkers can be used for etiological and prevention research, they need to be validated in populations. This calls for the development of analytical methods for use in large-scale populations. Currently, there is inadequate research support for scaling-up efforts needed for population studies. The validation and scaling efforts discussed here require close collaboration between laboratory scientists and population scientists (clinicians, epidemiologists, industrial hygienists, and exposure assessors). Transitional studies bridge the gap between the development of molecular markers in the laboratory and their application in population-based studies. These studies generally involve the initial evaluation and application of biomarkers in healthy human populations. Their objective is to address issues in sample processing, evaluate assay accuracy and precision, collect information about potential confounders and effect modifiers, and study early biologic effects of selected exposure. Transitional studies can be divided into three broad categories to clarify their distinctive research goals: developmental, characterization, and applied studies. In practice, however, elements of all three types of study are often incorporated into a single field investigation.

2.2.1 Developmental/Characterization Studies

Identification of a promising new molecular biomarker in the laboratory does not mean that the biomarker is ready for use in an epidemiologic study. Other basic issues need to be resolved before its application in human studies can be considered. First, the reliability (i.e. the repeatability of the assay) of a marker must be determined. As long as an assay is reliable, the ordering of subjects by the measure is preserved. Since this is all that is required for studying a marker–disease relationship, reliability, and not accuracy, is of initial importance. Reliability of laboratory assays may be assessed by the analysis of blind replicate samples representative of the range of values likely to be found in human populations. After the assay reliability and accuracy has been determined, it is important to define the optimal conditions for collecting, processing, and storing biological specimens for eventual assay, since, not uncommonly, small variations in these conditions determine the subsequent analyzability of samples. These studies are generally designed to address questions about the presence or the levels of a newly developed marker in the general population. In addition, they serve to identify factors that are confounders or effect modifiers of a marker (e.g. age, gender, medications) that need to be measured and taken into account when applying the marker in subsequent studies.

2.2.2 Applied Studies

Applied studies are investigations performed on subjects with particular patterns of exposure to xenobiotics (e.g.

occupational exposures, smokers) or on patients receiving chemotherapy. In these studies, the biomarker is treated as the outcome variable. At this stage of research, the biomarker has not been shown to predict an increase in risk of disease. The marker, however, can often be used to provide insight into the association between external exposure and biologic processes early in the exposure–disease relationship. Applied studies can help establish the biologic plausibility of associations detected in etiologic studies. Applied studies generally cannot establish a causal relationship between a given exposure, or a given level of exposure, and the risk for developing disease. The results of applied studies using the biomarker as outcome are suggestive only until a marker is shown to predict disease risk, which can be established only by comparing risk of disease in individuals with and without the marker. In these studies, the biomarker may be overly sensitive (i.e. it may respond to low levels of chemical exposures without biological relevance), it may be insensitive, or it may reflect phenomena that are irrelevant to the disease process. Until these relationships have been sorted out, the findings are merely suggestive.

2.3 Etiologic Studies

The major objective of molecular epidemiology is to conduct etiologic and applied research. Etiologic studies [i.e. ecologic, case–control, case–case (also referred to as a case series), prospective cohort, family, twin, and intervention studies] can be distinguished from transitional studies in that the former involve either clinically ill subjects, asymptomatic subjects with early disease, or subjects positive for an intermediate process known to be associated with increased risk of disease (e.g. colon adenomas and risk of colon cancer). Case–control, case–case, and prospective etiologic studies can effectively utilize molecular epidemiologic approaches.

2.3.1 Case–control Studies

A case–control study involves the comparison of cases (people with a particular disease) with controls (people without that disease) for various risk factors. The risk factors could be an exposure, a trait, or a biomarker. Traditionally, case–control studies have involved patients with clinically confirmed disease, identified either through the presence of symptoms or as a result of incidental findings on routine clinical examinations. Increasingly, cases are being defined as asymptomatic subjects whose early preclinical disease has been ascertained by screening (e.g. early breast disease, colon polyps, cervical dysplasia). The case–control study design is used far more frequently than the prospective cohort study design because of its relatively greater efficiency and lower cost. Therefore, maximizing opportunities to creatively integrate biomarkers into case–control studies is important. Because some markers are affected by disease itself, which raises complications of reverse causality, it is important to define which biomarker categories can most effectively be used in this study design.

2.3.2 Case–case Studies

A case–case study involves a series of cases of the same type that are compared on the basis of a particular exposure and a particular biologic characteristic. The accumulation of a spectrum of *ras* oncogene mutations in leukemia cases with benzene exposure compared with those in

leukemia cases without benzene exposure is an example of a case–case study. Such studies have the potential of identifying an exposure-specific effect at the molecular level. Case–case studies, however, cannot be used to estimate the relative risk of disease from a specific exposure. A nondisease control group is required for this purpose. Case–case studies offer great precision in investigating gene–environment interactions when the disease outcome of interest is rare.

2.3.3 Prospective Cohort Studies

Prospective studies involve healthy people, characterized by the presence or the absence of a risk factor, who are followed forward to determine the risk of disease. In prospective studies, the biological samples may be collected from subjects at various times. These samples are either analyzed at the time of collection or banked for later analysis. One way to utilize biologic markers in a prospective study is to follow the groups of subjects forward in time: subjects who develop disease are identified, and premorbid biomarker levels in the group with disease are compared with those without disease. Often, a nested case–control approach is used. Samples from cases and only a sample of the controls (noncases) are analyzed, which considerably reduces the laboratory requirements and costs. Although the prospective study design is by far the most time-consuming and expensive type of observational epidemiologic study, it is the only method available to test the association of biomarkers with disease risk when the markers are transient or may be directly or indirectly affected by disease. Prospective studies may yield banks of biological specimens that are useful for future studies. Large cohort studies initiated in the 1980s and 1990s are, in general, banking most or all fractions of the peripheral blood sample to allow a far wider range of biologic assays to be performed, particularly those that require DNA.

2.4 Molecular Epidemiology and Genomics

Molecular epidemiology and genomics utilize the products of genomic research to identify variation in any gene considered relevant to disease as well as the "genome-wide" genotype for every participant in a study. Additionally, products of genomics research will allow classification of diseases in new ways, possibly leading to new interventions. Evaluation of gene–environment interaction is becoming an increasingly important topic that will be complicated by large numbers of gene–gene and gene–environment interaction candidates. Larger-sized study populations (in thousands) will be required to study more than a few genes and environment factors.

Molecular epidemiology can contribute to addressing the issue of massive amounts of data that result from genomics research. The reduction, summarization, analysis, and interpretation of data from a population perspective are necessary to distinguish homeostatic profiles from pathologic profiles. Molecularly defined subsets may be the bases for new classification or definitions of disease.

2.5 Public Health Applications

Public health practice incorporates the end use of validated biomarkers for risk assessment by government agencies, for population screening (both active and

passive), and for clinical and preventive medical practice. The use of molecular biomarkers in public health practice is still in its infancy, although several very promising markers may soon find their way into the public health arena. The standard principles of biological monitoring and medical screening are applicable to the use of any biomarker, however. These include assay reliability and cost, strength of the association between a marker and disease risk, prevalence of the marker in the population, availability of effective, preventive strategies that can be employed in subjects who are positive for the marker, and a host of ethical considerations, such as informing subjects of test results.

See also Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Gledhill, B.L., Mauro, F. (Eds.) (1990) *New Horizons in Biological Dosimetry*, Wiley-Liss, New York.

Groopman, J.D., Skipper, P.C. (Eds.) (1991) *Molecular Dosimetry and Human Cancer: Analytical, Epidemiological and Social Considerations*, CRC Press, Boca Raton, FL.

Hulka, B.S. (1991) Epidemiological studies using biological markers: issues for epidemiologists, *Cancer Epidemiol. Biomarkers Prev.* **1**, 13–19.

Hulka, B.S., Wilcosky, T.C., Griffith, J.C. (Eds.) (1990) *Biological Markers in Epidemiology*, Oxford University Press, New York.

Millikan, R. (2002) The changing face of epidemiology in the genomics era, *Epidemiology* **13**, 472–480.

Schulte, P.A. (1993) The use of biological markers in occupational health research and practice, *J. Toxicol. Environ. Health* **40**, 359–366.

Primary Literature

Perera, F.P., Weinstein, I.B. (1982) Molecular epidemiology and carcinogen-DNA adduct detection: new approaches to studies of human cancer causation, *Chronic Dis.* **35**, 581–600.

Rothman, N., Stewart, W.F., Schulte, P.A. (1995) Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories, *Cancer Epidemiol. Biomarkers Prev.* **4**, 301–312.

Schatzkin, A., Freedman, L.S., Schiffman, M.H., Dawsey, S.M. (1990) Validation of intermediate endpoints in cancer research, *J. Natl. Cancer Inst.* **82**, 1746–1752.

Schulte, P.A., Perera, F.P. (Eds.) (1993) *Molecular Epidemiology Principles and Practices*, Academic Press, San Diego, CA.

Epigenetic Mechanisms in Tumorigenesis

Maureen K. Bunger and Trevor K. Archer
National Institutes of Health, Research Triangle Park, NC, USA

1 Introduction: The Chromatin–Cancer Connection 239

2 Epigenetic Modifiers 241
2.1 DNA Methylation 241
2.2 Chromatin Architecture and Modifications 245
2.2.1 Histone Acetylation 248
2.2.2 Histone Methylation 251
2.2.3 Histone Phosphorylation 254
2.3 RNA and Posttranscriptional Silencing 255

3 Epigenetics and Tumorigenesis 255
3.1 Methylation-mediated Silencing in Tumorigenesis 255
3.1.1 Methylation of Genes Involved in Cell-cycle Control 256
3.1.2 Methylation of Metastasis-related Genes 260
3.1.3 Methylation of DNA Repair Genes 260
3.1.4 Methylated DNA Binding Proteins 261
3.2 Histone Modifications in Tumorigenesis 261
3.2.1 Histone Methyltransferases 261
3.2.2 Histone Acetyltransferases 263
3.3 Imprinting and Tumorigenesis 264
3.4 Mouse Models 265

4 Clinical Considerations 266
4.1 Detections 266
4.2 Treatments 268
4.2.1 DNA Methyltransferase Inhibitors 268
4.2.2 HDAC Inhibitors 268

5 Conclusions 269

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Bibliography 270
Books and Reviews 270
Primary Literature 270

Keywords

Epigenetics: epi- Gr.
On or above; in addition to; near; or after. The patterns of inheritance that are not produced by changes in DNA sequence or follow the Laws of Mendel. Most actively studied in the context of DNA methylation and histone posttranslational modifications.

Histones
A family of related basic proteins consisting of H1, H2A, H2B, H3, and H4 and several variants that are organized to provide essential scaffolding for DNA. Histones undergo a variety of posttranslational modifications including acetylation, methylation, phosphorylation, and ubiquitination all of which may influence chromatin structure and function.

Nucleosome
Approximately 146 bases of DNA are wrapped 1.8 times around a core octamer that is formed from a tetramer of histones H3 and H4, and two dimers of histones H2A and H2B. The histone H1 is bound to DNA in between each nucleosome and functions to further compact the nucleosomes into higher ordered chromatin structure.

Euchromatin
The gene-rich region of genomes characterized by more open chromatin structure, gene transcription, and poor staining with basic dyes.

Heterochromatin
The gene-poor region of the genome characterized by dense chromatin structure, little gene transcription, and strong staining with basic dyes.

Loss of Heterozygosity
LOH: A genetic event caused by nonhomologous recombination in diploid genomes resulting in three possible outcomes; (1) the structural deletion of a normal allele or loss of the chromosome carrying a normal allele; (2) mitotic recombination between a normal and mutant allele resulting in the formation of a daughter cell homozygous for deleted or inactivated allele; or (3) duplication of chromosome with deleted or inactivated allele accompanied by the loss of the normal chromosome.

There is convincing evidence that tumorigenesis results from mutations in DNA that lead to critical changes in cell metabolism and are inherited as a cell divides itself. However, the process of tumorigenesis can also be strongly influenced by epigenetic factors. Epigenetic inheritance is defined as inheritance that does not follow normal Mendelian patterns and may involve nuclear components in addition to DNA. We now have significant information about the molecular mechanisms that influence epigenetic inheritance. Most of these mechanisms include direct modifications of either DNA itself or the collection of proteins, called *histones*, which intimately associate with DNA. Unlike DNA mutations, epigenetic events are often reversible, making them exciting new targets for therapy. This understanding is leading to the development of specific drugs that target epigenetic events and thus may complement existing cancer treatments.

1 Introduction: The Chromatin–Cancer Connection

The formation of clonally expanding tumors in multicellular organisms is due to complex changes in cellular physiology, which results in failure of cell-growth control. In 1971 it was hypothesized that two molecular events are required to predispose a cell type to become tumorigenic. Since then, our understanding of tumor-related molecular events has expanded tremendously owing to the advent of molecular biology techniques to clone and functionally characterize individual genes. This collective work has broadly defined genes related to tumor formation into two major classes: protooncogenes that stimulate proliferation and inhibit terminal differentiation and tumor suppressor genes that tend to inhibit proliferation and promote terminal differentiation. The classical two-hit hypothesis affecting a tumor suppressor gene is diagrammed in Fig. 1(a). In this scenario, a mutation occurs that inactivates one allele of the tumor suppressor gene leaving only one fully functional allele. If there is a growth advantage from losing the other allele, most often loss of heterozygosity (LOH) occurs leaving only one nonfunctioning allele in the tumor cell.

It is very clear, however, that mutation is but one way to effectively influence gene expression. Aberrant changes in the local chromatin structure surrounding start-sites of expressed genes can also dramatically alter expression levels and patterns. Interestingly, in tumors this so-called "epigenetic" influence on gene expression is also inherited in daughter cells of clonally expanding tumors even when no mutation can be detected. Several mechanisms through which the cell machinery generates epigenetic "marks" that communicate structural changes in the chromatin architecture are now becoming clear. The best characterized is the methylation of DNA at cytosine residues located in the context of CpG dyads found in expression control regions. However, more recent research has focused on histone modifications as a part of a "code" that communicates to chromatin remodeling factors, thereby affecting rates and levels of transcriptional activation. Indeed, emerging evidence suggests that changes in histone modifications may precede or even trigger DNA methylation. While not yet proven to be heritable,

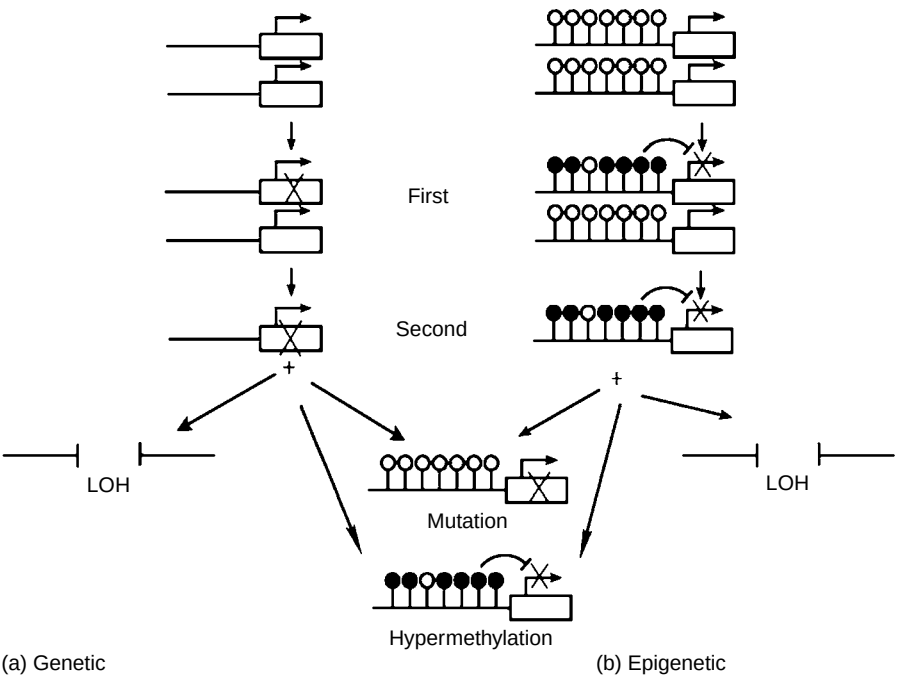

Fig. 1 Genetic and epigenetic two-hit models as the means of silencing a tumor suppressor allele. (a) Classical genetic model of the two-hit hypothesis at a tumor suppressor allele. First hit involves a genetic mutation (marked with an X) of one allele resulting in a single functional allele. Second hit may be the loss of heterozygosity (LOH) or mutation of the second allele resulting in complete loss-of-function protein. (b) Epigenetic model of the two-hit hypothesis at a tumor suppressor allele. The CpG island within the promoter region is unmethylated (○) allowing expression. The first hit results from methylation (●) of CpGs in one allele abrogating its expression leaving only one allele making functional protein. The second hit can be mutation of the unmethylated allele (left), methylation of the second allele (center), or LOH (right).

linkage of histone modifications to DNA methylation renders both to be important in the epigenetic inheritance.

A greater understanding of the importance of epigenetic events in gene expression has allowed a reinterpretation of the two-hit hypothesis to include aberrant epigenetic marks functioning as one or both of the hits. Thus, as diagrammed in Fig. 1(b), a scenario arises in which an epigenetic mark, in this case DNA methylation at a tumor suppressor promoter, silences expression of one allele of a tumor suppressor gene without altering its primary sequence. Again, if there is selective growth advantage by this process, LOH, mutation, or additional hypermethylation occurs at the other allele. Therefore, tumorigenesis can result from any number of combinations of genetic and epigenetic alterations, many of which are discussed below. Because epigenetic marks are often reversible, and most often an epigenetically silenced allele does not also harbor mutations, they are exciting new targets for tumor therapy strategies aimed at reactivating silenced tumor suppressors.

An increasing number of loci that are important in cancer are being determined to be inactivated both by mutation and by silencing through epigenetic means. The focus of this chapter is to outline current knowledge of epigenetic influences on gene expression and the role of these modifications in tumor formation. We will discuss in depth the role of DNA methylation and chromatin modifications in tumor formation and progression; however, we will also briefly discuss another epigenetic modifier of gene expression brought about through RNA-mediated silencing. We first discuss the mechanisms of generating and maintaining epigenetic marks followed by the evidence that epigenetics is intimately linked to tumorigenesis. Finally, we discuss how epigenetic marks can be detected in tumor cells, and current and potential therapies that are directed at cancers influenced by epigenetic events.

2 Epigenetic Modifiers

Epigenetic regulation of gene transcription occurs in at least three layers of the structural hierarchy of the genome. They are on DNA itself, through DNA methylation; at the level of the nucleosome, through histone modifications; and at the level of the chromosome, through chromatin binding structural and remodeling proteins. These "marks" and the proteins responsible for making them can influence gene expression on a global level by compacting or relaxing large segments of the genome or on the local level as coactivators and repressors of inducible gene expression. The formation of these epigenetic marks is enzyme-dependent and is highly regulated. Table 1 notes broadly the types of modifications that occur on DNA, and the individual histone proteins, and the types of cellular enzymes that catalyze the process. Following is a discussion of those processes that play known roles in tumor formation.

2.1 DNA Methylation

The 5′ carbon of cytosines that are found in the context of the CpG or CpNpG nucleotide sequences throughout the genome can be methylated. This produces 5-methylcytosine, often referred to as the *fifth base*. The methylation of DNA has

Tab. 1 Types of epigenetic modifications in chromatin.

Substrate	*Epigenetic marks*	*Residue target*	*Chromatin target*	*Enzyme types*
DNA	Methylation	Cytosines in CpG dyads	DNA	DNA methyltransferase (DNMT)
Histones	Methylation	Lysine and arginine	H3, H4	Histone methyltransferase (HMT)
	Acetylation	Lysine	H3, H4, H2A, H2B	Histone acetyltransferase (HAT)
	Phosphorylation	Serine and threonine	H3, H4, H1	Kinase
	Ubiquitination	Lysine	H2A, H2B	Ubiquitin ligase

been shown to serve at least two major purposes. First, it serves as something akin to a genetic immune system targeting foreign DNAs such as viruses that become incorporated into the genome nonhomologously and silencing their expression. Second, because methylation occurs postreplication, it may serve to mark the parental strand in replication and direct DNA excision repair enzymes to recognize the base in a mismatch that is incorrect. A more recently uncovered role for DNA methylation is a direct influence on both temporal and spatial control of gene expression. It is estimated that 60 to 90% of the CpG dinucleotides in the genome are methylated. The promoters of many highly expressed genes contain CpG islands, where many copies of the CpG dinucleotide are found in a relatively small area near the transcriptional start site. Unlike most genomic CpGs, the CpGs within these islands are mostly nonmethylated.

Normal DNA-methylation patterns are established during embryonic development. DNA in the ovary and sperm is highly methylated. Upon fertilization, the zygotic DNA undergoes rapid demethylation and new methylation patterns are subsequently established near implantation. These patterns continue to be altered throughout differentiation and development. A hallmark of the tumor genome is a decrease in overall DNA methylation. Because DNA methylation is so often associated with repression, this decrease was initially thought to result in aberrant upregulation of viral and somatic protooncogenes or embryonic alleles normally silenced. However, concomitant with this decrease in global methylation is an increase in localized methylation at certain promoter-associated CpG islands. Thus, gene promoters not normally methylated, become methylated and silenced in many tumors. The differentially methylated regions of the genome also occur in coding regions and other intronic regions of genes that can also correlate to expression changes of those genes.

Unlike mutations that are static and irreversible, DNA-methylation patterns are heterogeneous and dynamic throughout tumor formation and growth. Some evidence suggests that methylation spreads to CpG islands as DNA replicates and that the degree of gene silencing is therefore progressive as a tumor ages. The juxtaposition of euchromatic DNA to heterochromatic DNA is maintained by strict boundary elements to prevent spreading of one conformation into the domain of the other. In some cases, specific proteins that bind to these boundary elements to act as barriers are methylation-sensitive. In normal cells, unmethylated CpG islands in euchromatin are often found in the context of less-CpG-dense DNA regions that are highly methylated and compacted. During tumor formation, methylation of CpGs slowly advances into CpG islands from the surrounding compacted chromatin. Recent evidence indicates that specific histone modifications such as acetylation and methylation may differentially mark the regions; however, much remains to be determined regarding the mechanisms that maintain this separation.

Methylation of CpGs occurs by either *de novo* methylation, where unmethylated CpGs become methylated, or by maintenance methylation in newly replicated DNA, where methylation patterns on the parental strand is recognized as a signal to similarly methylate the newly formed daughter strand. Both types of methylation, *de novo* and maintenance, are catalyzed by a group of proteins called *DNA methyltransferases* (DNMTs). The first of

these to be cloned, DNMT1 is more important in maintenance, methylation having preference for hemimethylated DNA as a substrate. Indeed, DNMT1 is recruited to the hemimethylated template within replication forks implicating its role in reading methylation patterns in parent DNA and copying these patterns to the newly formed daughter DNA strand. That DNMT1 is not required for *de novo* methylation came from introducing proviral DNA sequences into cells that lack DNMT1. These sequences were rapidly methylated similar to wild-type cells and was the first proof that additional DNMTs existed. Subsequently, *DNMT2*, *DNMT3a*, and *DNMT3b* genes were cloned by virtue of sequence homology to *DNMT1*. *In vitro* analysis indicated that DNMT3a and 3b proteins preferred unmethylated DNA as a substrate, suggesting these may be critical for the *de novo* DNA-methylation activity seen in *DNMT1*-null cells. Humans having specific mutations in *DNMT3b* have a syndrome known as *immunodeficiency centromeric instability and facial anomalies* (ICF). At the DNA level, this disease is characterized by decreases in methylation at pericentric chromatin. The fourth family member DNMT2, while having significant homology to the other DNMTs has no known role in DNA methylation.

While the exact role of DNMTs in human tumor formation is unclear, several clues can be derived from analysis of animal models generated that lack various DNMTs. There are a total of five reported *DNMT1* mutant alleles in mice generated in different laboratories by gene targeting. Most of these that were completely or nearly completely deficient in DNMT1 activity died in early embryogenesis. However, one of these was only moderately deficient in DNMT1 function and mice were viable as a homozygous adult. This particular animal yielded significant insight into the role of methylation in tumor susceptibility when used in crosses to mice with mutations in alleles that are targets for tumor specific methylation (see Sect. 3).

Animal models generated to lack either DNMT3a or DNMT3b also result in inviable animals, although death occurs at different life stages. *DNMT3a* null animals survive embryogenesis, but are growth-stunted during the postnatal period and most die by about four weeks of age. *DNMT3b* null animals do not survive embryogenesis and die between E9.5 and E16.5 of multiple developmental defects. Embryonic stem (ES) cells derived from these animals lacking either DNMT3a or 3b revealed no significant differences in *de novo* methylation targeted at proviral sequence. However, when both were absent, no methylation of these sequences occurred despite expression of normal levels of DNMT1. Finally, ES cells specifically lacking DNMT3b also are hypomethylated at centromeric repeat sequences. It is important to note that these results do not rule out the role of DNMT1 in de novo methylation and also do not rule out the role of DNMT3s or other as-yet-unknown DNMTs in maintenance methylation.

There are several proposed ways in which DNA methylation can lead to gene repression or silencing (Fig. 2). One occurs when the methylated cytosine blocks recognition of DNA sequence by sequence-dependent transcriptional factors (Fig. 2(b)). For example, it has been shown that a differentially methylated CpG in the breast cancer associated-1 (*BRCA1*) gene promoter can prevent binding of a critical transcription factor necessary for full *BRCA1* gene activation. Alternatively, proteins with methyl-CpG binding domains (MBDs) specifically recognize and

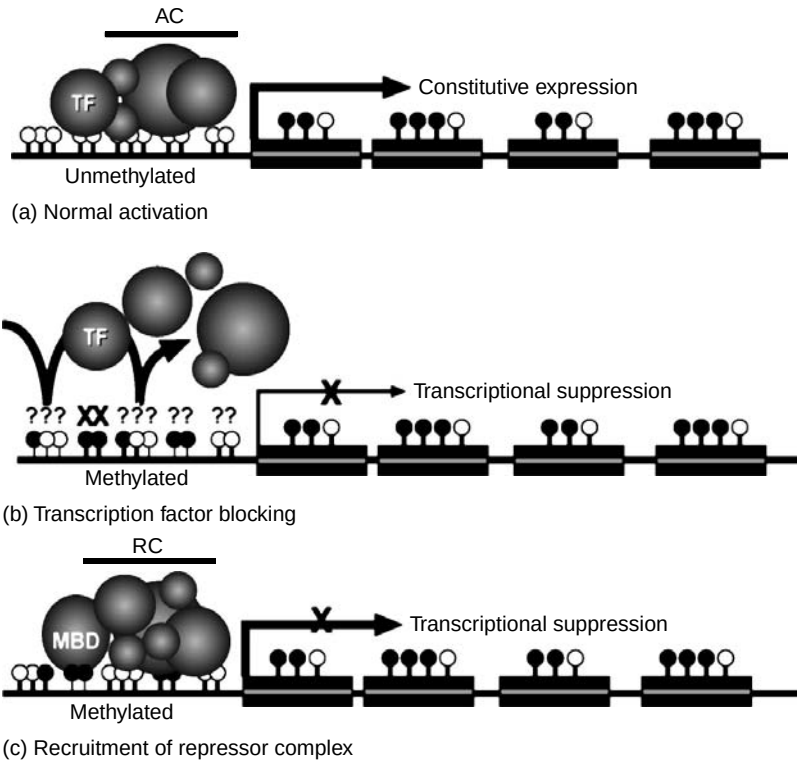

Fig. 2 Potential mechanisms of DNA methylation-induced silencing. (a) Normal transcriptional activation occurs when transcription factors (TF) bind to specific unmethylated (○) sequences in DNA and recruit transcriptional activation complexes (AC) to permit constitutive or inducible gene expression. (b) Methylation causes blocking of TF binding. Aberrant methylation (●) at select CpGs within the same region in tumor subclones (as indicated by XX at the TF binding site) alters the binding of TF and subsequent failure to recruit the AC to the promoter. (c) Aberrant methylation at select CpGs causes recruitment of a methyl binding domain protein (MBD) which then recruits a repression complex (RC) to the promoter inhibiting transcription.

bind to methylated Cs (Fig. 2(c)). Proteins identified that have MBDs include MeCp1, MeCp2, and MBD1, 2, 3, and 4. MeCp1 and 2 also contain transcriptional repression domains that recruit repression complexes (RC) such as Sin3 and Mi-2/NuRD that in turn recruit histone deacetylases (HDACs). These complexes also contain members of the SWI/SNF complexes that may serve to reposition nucleosomes into repressive conformations (see Sect. 2.2). The failure of any one of these complex recruitment pathways could alter the transcriptional context of a significant fraction of the genome.

Methylation of CpGs in certain contexts may lead to mutation in coding regions of genes. The spontaneous deamination of 5-methylcytosine changes the base to a thymine and in the absence of DNA repair causes a heritable sequence change. This accounts for the

significant underrepresentation of CpGs in the context of dinucleotides in mammalian genomes. Occurrence of this change in coding regions of genes can lead to amino acid changes and often truncations. For example, nearly half of all mutations that occur in *p53* in conjunction with tumor formation occur at or are near methylated cytosines within the coding region. Methylations of CpGs within *p53* are the basis for its increased mutation rate when exposed to UV light due to methylation-based changes in the UV absorption spectrum. Furthermore, G to T transversions in *p53* caused by certain carcinogens are more common when Gs are next to a methylated C. Thus, it can be postulated that genes that contain methylated CpGs within their coding regions are potentially more susceptible to mutations caused by environmental factors such as UV light and carcinogens.

Another environmental factor, nutrition, plays a significant role in DNA methylation. Low levels of folate and choline in diet are associated with global DNA hypomethylation and liver cancer in rats. Furthermore, higher folate intake in humans is associated with lower levels of colon cancer and smoking-related lung cancer. One of the potential connections of folate-to-cancer incidence is dependence on folate and folate metabolites to maintain a cellular balance of *S*-adenosyl methionine (SAM), a cofactor for methyltransferases. DNMTs use SAM as a methyl donor group in its enzymatic reaction. Dietary methionine and choline are the building blocks for SAM synthesis and folate functions to prevent accumulation of *S*-adenosyl homocysteine (SAH), the demethylated form of SAM and a potent methyltransferase inhibitor. In the absence of dietary methionine, methionine can be generated from homocysteine using methyl groups derived from either folate or choline. The effect of global changes in the SAM/SAH ratio impacts levels of DNA, RNA, and protein methylation.

2.2 Chromatin Architecture and Modifications

Before discussion on how chromatin is modified to affect gene expression and tumor formation, it is necessary to describe the nature of chromatin itself. In eukaryotes, DNA is compacted into a higher order chromatin structure by association with the family of highly conserved histone proteins. The fundamental unit of chromatin is the nucleosome consisting of ~146 bases of DNA wrapped 1.75 turns around an octomer of histone proteins. The octomer itself consists of two H2A/H2B dimers, and an H3/H4 tetramer that form a tight spool-like structure as determined by NMR and crystallographic studies (Fig. 3(a)). Each nucleosome is linked to the next via a variable length of DNA to which the linker histone H1 is bound (Fig. 3(b)). This histone is thought to be critical for the folding of the chromatin fiber and establishment of the higher order structure. Each of the core histone proteins consists of a globular domain and a tail. The position of the tail domains could not be assigned in nucleosome crystallographic studies, implying that the tail domains do not form an ordered structure. While the minor groove of the double helical DNA faces the nucleosome core, it is thought that these tails may protrude into the major groove of the helical DNA, the surface involved in most protein–DNA interactions. These interactions, influenced by the type and position of modifications of the histone tails, form the basis of the "histone code" hypothesis.

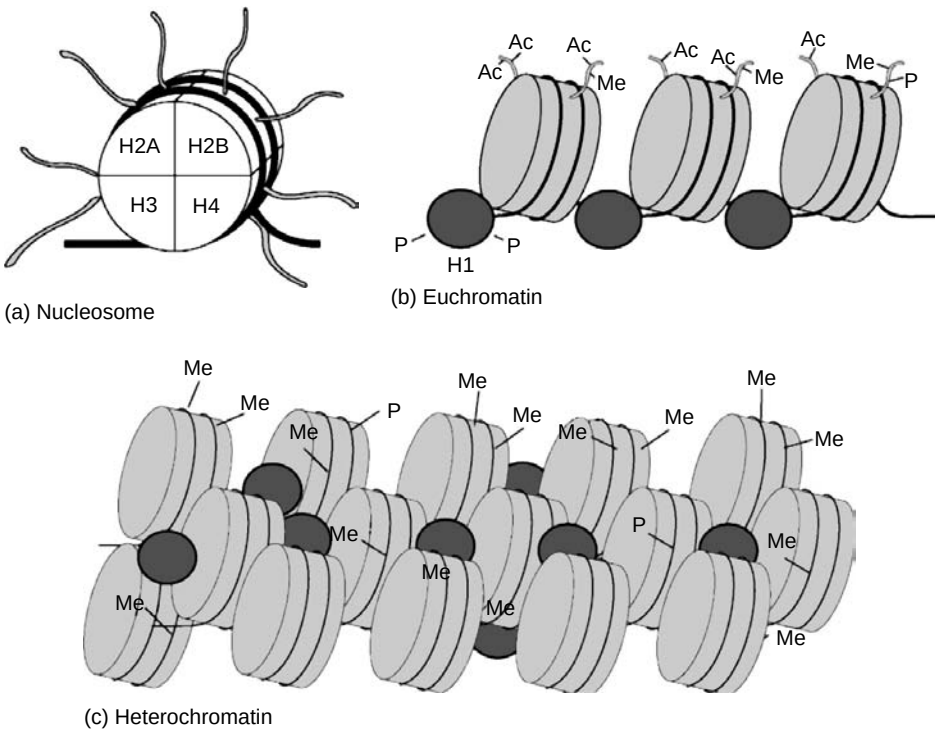

Fig. 3 Chromatin structure and modifications. (a) The nucleosome core particle structure consists of four subunit histones H3, H4, H2A, and H2B. DNA (bold lines) is wrapped around an octomer of these histones approximately twice. Histone tail domains are shown to protrude from the nucleosome and are available for protein–protein interactions with transcriptional and chromatin regulators. (b) Euchromatic DNA forms an open conformation available for binding of transcription factors. Histones in these regions are most often acetylated (Ac), but also methylated (Me) at certain positions. Histone H1, the linker histone, resides in regions of DNA between nucleosomes. H1 makes contact with each nucleosome and may also form protein–protein interactions with transcriptional and chromatin regulators. H1 can be phosphorylated (P) at multiple residues. (c) Heterochromatin forms a closed conformation inhibiting access of transcriptional regulators to DNA. Histones in heterochromatin are most often methylated (Me). DNA methylation is also increased in heterochromatin.

Chromatin structure may be viewed as being organized into two functional modules called *heterochromatin* and *euchromatin*. DNA, modeled into heterochromatin, represents the majority of the genome and consists mainly of repetitive DNA sequences, centromeric, and telomeric chromosome regions, and also includes genes that are differentially expressed depending on any number of other factors. The chromatin structure in heterochromatin is tightly compacted and densely associated with nucleosomes and histone H1 (Fig. 2(c)). The formation of heterochromatin, broadly speaking, prevents expression of unwanted proteins and is heterogeneous from one cell-population to the next. Actively transcribed

or ready-to-be transcribed genes reside in euchromatin. The structure of euchromatin is not compacted, and has fewer nucleosomes that are phased or unevenly spaced to allow a permissive environment for transcription factor binding and transcriptional initiation (Fig. 2(b)). Nucleosomes in euchromatin are more highly acetylated than in heterochromatin. Boundary elements in chromosomes have been found to separate heterochromatin from euchromatin implying that heterochromatin has the potential to spread. Typically, if a normally euchromatic, actively expressed gene is transposed to a region of heterochromatin, the gene will adopt a heterochromatin formation and become silenced.

While heterochromatin is a relatively stable structure, highly methylated and compacted, the structure of euchromatin is constantly being remodeled in response to cellular signaling. Factors that influence this remodeling fall into two classes. The first of these includes a number of large multiprotein chromatin remodeling complexes, the most well characterized being the SWI/SNF complexes; the name derived according to homology with similarly functioning yeast proteins. These complexes regulate accessibility of DNA within nucleosomes to transcription factors by loosening or tightening DNA–histone contacts and shifting histone octomers along DNA. The SWI/SNF complexes can interact with both transcriptional repressors and activators; however, the specificity of such interactions and their role in tumor formation is only just becoming clear. At the core of these complexes is an ATP-dependent DNA helicase that is essential for remodeling activity. In mammals, there are two main remodeling complexes that contain one of two ATP-dependent helicases called Brahma (*Brm*) and Brahma-related-gene-1 (*Brg1*). Both *Brg1* and *Brm* may participate in growth control via interactions with either Rb or BRCA1, as has been demonstrated in tumor cell lines. Moreover, mice deficient for one allele of *Brg1* have increased incidence of leukemia. Furthermore, a component of the SWI/SNF complex called SNF5 is linked to rhabdoid tumors in humans and animal models, a rare but highly aggressive pediatric tumor condition affecting brain, liver, and kidney tissues.

The second class of chromatin modifiers includes proteins that recognize and covalently modify histone tails within nucleosomes. The fact that histones are modified by acetylation, methylation, phosphorylation, and ubiquitination has been known for decades. However, the discovery that histone-modifying enzymes are critical players in the activation of transcription highlighted the dynamic role of histones in the regulation of gene expression. Further insight into these roles of histones in chromatin metabolism came with the advent of immunochemicals that allow visualization of modified histones within the chromatin context. This work has revealed that histones in euchromatin display a distinct range of histone modifications from the ones located in heterochromatin. More important is the yet emerging role of histone modifications in heterochromatin assembly and maintenance of DNA methylation; two events strongly linked to tumorigenesis. The location of most modified residues in the core histones and the linker histone H1 are known (Fig. 4). However, the function of each in modulating gene expression is only beginning to be realized. Interestingly, more and more genes that can modify histones are being linked to tumor formation. The subsections below discuss the role of several key histone

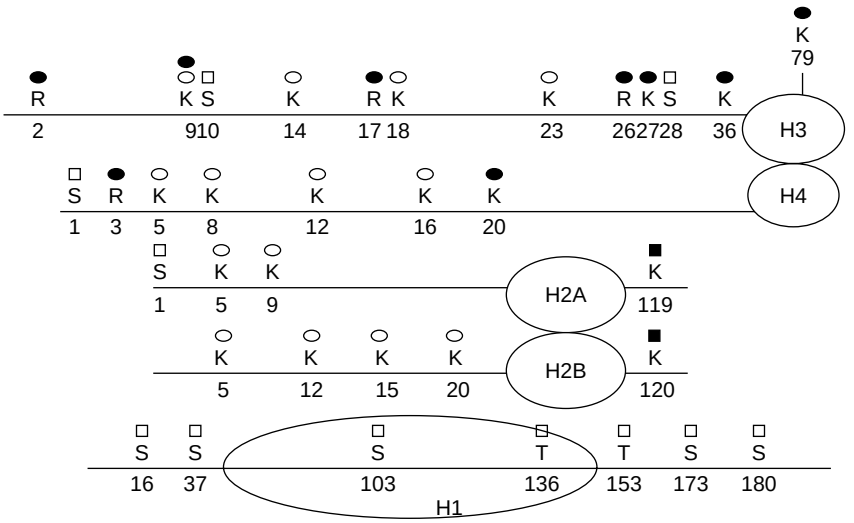

Fig. 4 Positions of modifications in the core histones and linker histone H1. Shown are the amino acids modified in each of the core histone tail regions, N-terminal tails to the left of globular domains and C-terminal tail to the right of globular domains in H2A, H2B, and H1. The types of modifications possible at each site are displayed above each amino acid. Phosphorylation (□), Methylation (●), Acetylation (○), Ubiquitination (■). Individual modifications can influence others in both facilitative and repressive ways.

modifications and the enzymes that catalyze their formation in mediating epigenetic regulation of gene transcription.

2.2.1 Histone Acetylation

Histone acetylation is known to be associated with, although not exclusively, transcriptionally active genes. Two classes of enzymes, histone acetyltransferases (HATs) and histone deacetylases (HDACs) are responsible for modifying the acetylation status of histones. These enzymes are often a part of larger protein complexes associated with transcriptional activation and repression respectively. Acetylation of the N-terminal tail of histones is thought to relax the DNA structure on the histone octamer and allow access to transcriptional machinery. Conversely, deacetylation of histones stabilizes a repressive chromatin structure. Acetylation of lysine residues has been documented in all core histones and at the N-terminus of linker histone, H1. Many of these enzymes are highly conserved in all eukaryotes including yeast, neurospora, and drosophila.

The first functional role of a HAT protein in gene transcription was identified in the yeast homolog of the Gcn5 transcriptional coactivator. Since then, several other proteins identified as transcriptional coactivators, many of which are included in Table 2, also have HAT activity. Currently, it is thought that histone acetyltransferases are recruited to chromatin through interactions with DNA sequence–specific transcription factors. However, acetylation activity can be targeted to specific DNA sequences when HATs also have DNA binding domains, as is the case for activating transcription factor-2 (ATF2). It is important to note that some HATs such

Tab. 2 Acetyltransferase modifiers.

HAT gene	***Histone target***	***Function***	***location***	***Cancer/disease connection***
PCAF	H3, H4	Transcriptional cointegrator	3p22	
CREBBP	All	Transcriptional cointegrator	16p13.3	Leukemia, Rubenstein-Taybi syndrome
P300	All	Transcriptional cointegrator	22q13.2	Leukemia
Tip60	H4, H3, H2A	DNA repair, apoptosis	17p11.2	
MOZ/ZNF220		Transcriptional cointegrator	8p11	Fused to Crebbp in AML
HBO1/MYST2	H3, H4	Binds origin of replication complex	17q21.32	
SRC1/NCOA1	H3, H4	Nuclear receptor coactivator	2p23	
AIB1/NCOA3	H3, H4	Nuclear receptor coactivator	20q12	Multiple tumors, amplified
GRIP1/NCOA2		Nuclear receptor coactivator	8q13.1	
TAFII250	H3, H2A	Transcription, cell-cycle progression	Xq13.1	
MYST1	H3, H4, H2A	Unknown	16p11.1	
ATF2	H4, H2B	Growth factor activated	2q32	
GCN5L2	H3, H2B	Transcriptional cointegrator	17q21	
HAT1	H4, H2A	Newly synthesized H4, H2A	2q31.2–33.1	
MORF	H4, H3	Unknown	10q22.2	Fused to Crebbp in AML

Note: AML: acute myelogenous leukemia.

as CBP, P300, PCAF, and TAFII250 function to modify both histones as well as other members of transcriptional activation complex (Table 2). In the case of CBP and P300, other cellular targets for acetylation such as P53 have been identified and may be equally important in modifying gene expression.

While the acetylation of the core histones is most often studied in the context of transcriptional activation, acetylation plays many roles in chromatin biology. For instance, Lys5 and Lys12 residues of histone H4 are acetylated prior to chromatin assembly and are an important modification for facilitating assembly of newly replicated DNA into chromatin. Once replication completes, this acetylation pattern is erased and new acetylation pattern is established most likely in response to genetic positional information. Histones in transcriptionally active genes are most often acetylated at H3(Lys9) and H3(Lys14). Conversely, acetylation of H4(Lys12) in yeast and flies is found predominantly in heterochromatin or repressed chromatin.

Moreover, as discussed below, acetylation of histones can facilitate and/or repress methylation or phosphorylation of other histone residues and *vice versa*. Among these interactions emerges the critical importance assigned to H3(Lys9), that can be either acetylated or methylated, but not both at the same time, providing what may be a molecular switch point for recruitment of activation versus repression complexes to gene promoters.

Histone acetylation is a dynamic equilibrium resulting from the opposing activities of HATs and histone deacetylases (HDACs). HDACs are defined by genetics into subclasses based on homology to three yeast HDACs, *Rpd3* (Class I: *HDAC1*, 2, 3, and 8), *Hda1* (Class II: *HDAC4*, 5, 6, 7, and 9) and *Sir2* (Class III: *hSIRT1*-7). There are two well-characterized class I HDAC containing complexes, Sin3/HDAC and NuRD/Mi2/NRD that play known roles in transcriptional repression that is dependent on HDAC activity. While the class I HDACs localization in cells is predominantly nuclear, the class II HDACs appear to be shuttled between the nucleus and cytoplasm. Class II enzymes are not found as part of the same complexes as class I enzymes, but do interact with corepressors of certain transcription factors.

How might acetylation status of histones be conveyed to transcriptional machinery? One way could be that acetylation of histones loosens the DNA–histone contact points providing greater accessibility of transcription factors to critical regulatory sequences. Alternatively, regulatory proteins that directly recognize and bind to acetylated lysines may be recruited to acetylated histones. The protein motif that is known to recognize and bind acetylated lysines is called the *bromodomain*. A large family of bromodomain-containing proteins has been identified both functionally and through sequence homology and is included in Table 3. Many HATs themselves contain

Tab. 3 Bromodomain-containing proteins.

Gene	*Description/Function*	*Location*
BAZ1A	Nucleosome remodeling	14q13.2
BAZ1B	Nucleosome remodeling	7q11.23
BS69	Corepressor	10p15.3
CREBBP	Histone acetylation, transcriptional integrator, P53 acetylation	16p13.3
EP300	Histone acetylation, transcriptional integrator, P53 acetylation	22q13.2
GCN5L2	Histone acetylation	17q21.2
PCAF	Histone acetylation	3p24.3
PHIP	Insulin receptor substrate interacting protein	6q14.1
SMAP	Nuclear receptor coactivator	5q31.2
Brg1/SMARCA4	ATP-dependent DNA helicase, nucelosome remodeling	19p13.2
Brm/SMARCA2	ATP-dependent DNA helicase, nucelosome remodeling	9p24.3
SP100	Nuclear body protein	2q37.1
SP110	Localized to the nucleolus	2q36.3
SP140	Nuclear body protein	2q37.1
TAFII250	Transcription complex	Xq13.1
TIF1A	Nuclear receptor coactivator	7q34
TRIM33	Related to TIF1	1p13.2
WDR9	WD-repeat protein	21q22.2

bromodomains, including CREBBP, P300, PCAF, GCN5, and TAFII250. Several bromodomain proteins have unknown functions, but many others are known to play specific roles in chromatin remodeling such as Brg1 and Brm, the catalytic engines of ATP-dependent chromatin remodeling complexes. It is important to note that at least one bromodomain protein, BS69, has known roles in repressor complexes providing further evidence that acetylation may have multiple and diverse functions. One can clearly imagine that through bromodomains, the status of acetylation as a promoter can determine the presence or absence of many key factors in transcriptional control.

2.2.2 Histone Methylation

While methylation of histones was discovered more than three decades ago, its importance in gene regulation and chromatin condensation was not be realized until the late 1990s when researchers began to clone histone methyltransferases functionally from yeast and mammalian cell extracts. Unlike histone acetylation that is subject to the dynamic interplay of HAT and HDAC activity, histone methylation appears relatively static and to date no bona fide histone lysine demethylases have been identified. Methylation can occur on either arginine residues or lysine residues and is catalyzed by several different protein families (Table 4). Many of the lysine directed Histone methyltransferases (HMTs) contain SET domains named for the founding three members in *Drosophila*, Suppressor of variegation effect-39H1 (*Su(Var)3–9*), Enhancer of zeste (*EZH*) and trithorax (*TRX*). These original members of the SET-domain family were cloned by virtue of their influence on *Drosophila Hox* gene regulation (EZH and TRX), or position effect variegation, the concept that the physical nature of surrounding chromatin can dictate whether a gene is expressed or not (Su(Var)3–9). Eventually, the gene products derived from the human homologs of these alleles *Suv39H1*, *EZH2*, and *MLL* were found to possess histone methyltransferase activity. Families of arginine methyltransferases containing SAM binding domains are known to have many cellular targets in addition to histones. The arginine methyltransferase CARM1 was identified by its association with the histone acetyltransferase GRIP1 and preferentially methylates H3(Arg2), H3(Arg17), and H3(Arg36). Finally, the protein DOT1 that does not contain a SET domain or an apparent SAM binding domain methylates H3(Lys79) in the globular domain and may affect telomeric silencing specifically. Because most arginine and lysine methyltransferases utilize SAM as methyl donor, they are possibly subject to the same dietary influences of folate outlined above for DNA methylation (see Sect. 2.1).

Methylation of lysine residues in H3 and H4 tail domains has been associated with either transcriptional activation or repression depending on the residue modified. Methylation of H3(Lys4) is often associated with transcribed genes, whereas methylation of H3(Lys9) is almost invariably associated with heterochromatin. The SET-domain proteins SET7 and MLL both specifically target H3(Lys4) for methylation and this modification is associated with their ability to coactivate transcription. Several lines of evidence point to H3(Lys9) as an epigenetic mark that dramatically alters heterochromatin structure *in vivo*. First, methyltransferases that show specificity for H3(Lys9), such as Suv39H1, are often found in complexes containing HDACs and heterochromatin-associated

Tab. 4 Methyltransferase modifiers.

SET-domain HMTs	***Histone target***	***Function***	***Location***	***Cancer/disease connection***
Suv39H1	H3(Lys9)	Heterochromatin formation	Xp11	
Suv39H2	H3(Lys 9)	Heterochromatin formation	10p12-p14	
EZH2	H3(Lys27)	Transcriptional repressor	7q35	Prostate cancer
MLL1	H3(Lys4)	Transcriptional activator	11q23	Leukemias
ESET	H3(Lys9)	Transcriptional repressor	1q21	
G9A	H3(Lys9, Lys 27)	Transcriptional repressor	6p21.3	
NSD1		Transcriptional coactivator	5q35	Sotos syndrome
NSD2		Silencing and repression	4p16	Multiple myeloma
NSD3		unknown	8p12	Breast tumors
SET8	H4(Lys20)	Cell-cycle regulated	13q11	
SET7	H3(Lys4)	Transcriptional coactivator	4q28	
ASH1	H3(Lys4, 9, 20)	Transcriptional activator	1q21.3	
Non-SET domain HMT's				
CARM1	H3(Arg2, 17, 26)	Transcriptional coactivator	19p13.2	
Dot1	H3(Lys79)	Silencing		
PRMT1	H4(Arg3)	Transcriptional coactivator	19q13.3	

Tab. 5 Chromodomain containing proteins.

Gene	*Description/Function*	*Location*
CBX6	Chromobox protein 6	22q13.1
CBX7	Chromobox protein 7	22q13.1
CDY1	Testis specific chromodomain Y protein 1	q11.2
CDY2	Testis specific chromodomain Y protein 2.	q11.2
CDYL	Chromodomain protein Y-like	6p25.1
CHD1	Helicase, remodeling	5q15
CHD5	Chromodomain helicase DNA binding protein 5	1p36.31
HP1-alpha	Heterochromatin formation, transcriptional repression.	12q13.13
HP1-beta	Heterochromatin	17q21.32
HP1-gamma	Unknown	7p15.2
Mi2-alpha	Deacetylation complex	17p12
Mi2-beta	Deacetylation complex	12p13.31
PC2	Polycomb homolog 2	17q25.3
PC3	Polycomb homolog 3 (CHD8)	17q25.3
Q9BTP1	M-phase phosphoprotein 8	13q12.11
SUV39H1	Suppressor of variegation	Xp11.23
SuV39H2	Suppressor of variegation	10p13
RBP1	Retinoblastoma binding protein	14q22.3

proteins such as HP1. Additionally, cells derived from mice that lack Suv39H1 show decreased heterochromatin assembly and conversely, overexpression of Suv39H1 increases overall heterochromatin in cells. Using Tetrahymena, it has also been shown that mutations in the Suv39H1 homolog or mutation of H3(Lys9) can abolish all histone methylation in this organism.

Similar to acetylation, conveyance of histone methylation marks to the transcriptional machinery may occur through proteins that specifically recognize and bind methylated lysines. The chromodomain (*Chr*omatin *o*rganization *mo*difier) of HP1 protein specifically recognizes and binds to methylated K9 indicating this mark may recruit and/or maintain HP1 in heterochromatin regions. By sequence homology, many other proteins exist that contain chromodomains and it awaits to be seen if the chromodomain is a general motif for binding of methylated lysine residues in histone tails (Table 5).

The dynamic link between histone methylation and DNA methylation is supported by already mentioned correlations of H3(Lys4) and H3(Lys9) with undermethylated and hypermethylated genome regions respectively. Several scenarios arise through which there could be communication between histone modifications and DNA methylation via interactions with the respective enzymes (Fig. 5). First, proteins that recognize methylated cytosines, such as MeCp1, can recruit histone deacetylases to promoter regions (Fig 5(a)). A second possibility is that the methyl binding domains within or in conjunction with histone methyltransferases recognize methylated cytosines and initiate histone methylation (Fig. 5(b)). A third possibility is that chromodomain containing proteins such as HP1 can recognize methylated lysines within nucleosomes and subsequently recruit DNA methyltransferases to CpGs in promoters (Fig. 5(c)). It is important to note

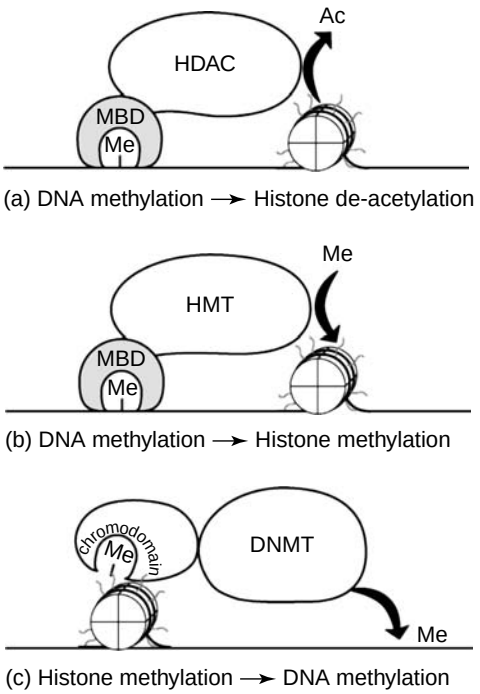

Fig. 5 Relationship between DNA methylation and histone modifications in transcriptional silencing. (a) Proteins with methyl binding domains (MBD; e.g. MeCp2) recognize methylated cytosines in DNA and recruit HDACs that catalyze histone deacetylation. (b) An MBD protein can also function to recruit histone methyltransferases, many of which function to repress transcription through methylation of histones (e.g Suv39H1 and Ezh2 resp). (c) Proteins with chromodomains (e.g. HP-1) recognize methylated lysines in histones and recruit DNAmethyltransferases (DNMT) to chromatin

that these scenarios are not mutually exclusive and that all may exist in a particular cell type or be specific for selected promoters.

2.2.3 Histone Phosphorylation

Phosphorylation of histones H3 and H1 is associated with chromosome condensation during mitosis. However, more recently, the phosphorylation of H3(Ser10) has been associated with activation of immediate-early-response genes such as *c-jun*, *c-fos* and *c-myc*. One H3 directed kinase Rsk-2 is mutated in human Coffin–Lowry syndrome. Phosphorylation of H3 is also catalyzed by MSK1, a kinase that is activated in response to growth factors *in vitro*. As shown in Fig. 4, the linker histone H1 contains up to seven residues that are targets for phosphorylation. The phosphorylation status of H1 undergoes distinct pattern changes throughout the cell cycle that may be regulated in part by the activity of CDK2, a protein kinase often hyperactivated in tumors. Interestingly, recent evidence suggests that phosphorylation status of H1 may influence SWI/SNF remodeling activity and gene activation of certain hormone responsive promoters such as the mouse mammary tumor virus (MMTV) promoter. Many phosphorylation events in histones act cooperatively with acetylation and antagonize methylation. For instance, *in vitro* H3(Ser10) phosphorylation facilitates acetylation of H3(Lys14), but antagonizes methylation of H3(Lys9). This role of H3(Ser10) in gene activation is counterintuitive to the role of phosphorylation in chromatin condensation and it may be that combinatorial changes in other histone modifications may also participate in

this process. Furthermore, sites and levels of phosphorylation in H1 do not appear to be equivalent and the presence of several isoforms of H1 may play specific and contrasting roles in gene activation rather than condensation.

2.3 RNA and Posttranscriptional Silencing

We must also consider another emerging player in epigenetic regulation. Over the past several years, the role of RNA and specifically short RNAs in critical biological processes has become increasingly clear. In plants, *Caenorhabditis elegans*, and neurospora, double-stranded RNAs (dsRNA) sliced into shorter products were shown to target homologous RNA for cleavage and degradation. Researchers immediately recognized the utility of such an endogenous system to allow functional knockouts in these organisms without altering genomic DNA sequence. Called *RNA-interference*, or *RNAi*, this technology can now be performed in mammalian cells with moderate success using short recombinant RNA. The concept of using small interfering RNAs to knockdown gene expression has tremendous translational value as these RNAs may be easier to deliver than typical viral or DNA based gene therapies without as many side effects.

RNA is also an important endogenous modifier of gene expression in mammalian cells. The regulation of two loci, the *H19* and *Xist-Tsix* loci provided the first indication that RNA may play a significant role in epigenetic gene regulation. Both loci code for untranslated RNAs. The *Xist* RNA is required for initiation of X chromosome inactivation and is regulated by balanced expression of *Tsix*. Through gene knockout studies in mice, *H19* RNA is shown to be required for the maintenance of imprinting at several somatic loci. It stands to reason that RNAs such as those generated from *H19* or *Xist* may in fact be structural parts of chromatin. Indeed, in plants, double-stranded RNA functions to mark homologous DNA sequences for methylation and subsequent silencing. Furthermore, in mammalian cells, it is known that the heterochromatin around centromeres is more stable and requires more prolonged treatment with HDAC inhibitors to remove HP1 and relax the chromatin in this region. In a remarkable discovery in 2002, this phenomenon was shown to be dependent on an as-yet-unidentified nuclear RNA since cells that were treated with RNAse rapidly lost HP1 from pericentric chromatin. This result will inevitably lead to a better mechanistic understanding of previous unexplained roles of RNA in epigenetic regulation. This structural role of RNA in chromatin and its impact in tumor development will be a new area of intense research in the years to come.

3 Epigenetics and Tumorigenesis

3.1 Methylation-mediated Silencing in Tumorigenesis

DNA methylation impacts tumor formation by both silencing of methylation-sensitive genes and increasing mutation rates in genes due to spontaneous deamination of methylated residues. For an example of the latter, CpG dyads in P53 are often methylated in normal human tissues and a small number of these CpG dyads account for or are adjacent to the majority of P53-inactivating mutations. While the spontaneous deamination mechanism is

hypothesized to be the most likely cause of CpG mutations, alternative explanations include carcinogen sensitivity and enzyme directed methods of mutating methylated genes (see Sect. 2.1). Alleles that are targets for hypermethylated-mediated silencing in tumors include many genes involved in cell-cycle progression (*P16*^{INK4a}, *Rb*, *P14*^{ARF}, *P73*, *ER*), DNA repair, (*MLH1*, *B4RCA1*), carcinogen metabolism, (*GSTP1*), cell adhesion and metastasis (*CDH1*, *CDH13*) and apoptosis (*DAPK*, *TMS1*). Promoter hypermethylation of some of these tumor suppressor alleles appears to be specific for certain tumor types, as is the case of p73 in lymphoma and BRCA1 in breast cancer. Many targets of methylation associated silencing have clear ties to tumor progression and are included in Table 6.

Many initial studies that mechanistically link methylation to cell growth are the result of analyzing tumor suppressor promoter methylation in tumor cell lines. The evidence that tumor suppressor CpG–island hypermethylation may be an early event in tumor formation is derived from studies on pre-neoplastic and tumor biopsies and resections. The number of alleles associated with methylation-induced silencing continues to grow. For the purpose of this article we have chosen, as archetypes, alleles for which significant progress has been made in determining tumor suppressive mechanism in cells and tumors and which we feel are particularly good examples of the many roles of DNA methylation in tumor formation. Table 6 lists additional known tumor-related genes influenced by methylation and includes information for accessing a database maintained by the National Library of Medicine where additional information for each gene can be found.

3.1.1 Methylation of Genes Involved in Cell-cycle Control

The cyclin dependent kinase inhibitor, CDKN2 or P16^{INK4a}, protein functions by inhibiting cyclin dependent kinases 4 and 6 (Cdk4 and Cdk6) thus inhibiting the phosphorylation of pRB and progression of the cell cycle. The *P16*^{INK4a} gene in tumors is subject to point mutations, deletions, and methylation, and the array of alterations varies by tumor type. The first observation of hypermethylation of the *P16*^{INK4a} CpG island in tumors was made in 1995. Work that followed has identified *P16*^{INK4a} promoter methylation in all tumor types examined. Recent work published in 2003 examined silencing of the *P16*^{INK4a} promoter in a colon tumor cell line and found that DNA methylation is required to maintain silencing of *P16*^{INK4a}. However, *de novo* methylation associated with the silencing of *P16*^{INK4a} promoter was actually preceded by histone methylation. This work provided an interesting new logic behind both the interplay as well as independent actions of these two epigenetic marks.

Familial adenomatous polyposis (FAP) is a familial condition resulting in multiple polyp formation in the lower GI tract and a predisposition of these polyps to develop into carcinoma. The adenomatous polyposis coli (*APC*) gene is located at chromosome 5p21 and mutations can result in the autosomal dominant familial FAP as well as sporadic cancers. However, not in all FAP cases could mutation be detected and in 18% of those that are not mutated, aberrant hypermethylation of the 5′CpG island occurs in association with reduced expression. Loss of APC function results in loss of degradation of the transcriptional activator and oncoprotein β-catenin, leading to increased cell cycle and growth. While the study of

Tab. 6 Expressed genes affected by promoter methylation status.

Gene	*Function*	*Location*	*Tumor types*	*OMIM*[a]
Cell cycle and apoptosis				
TP73	p53 homolog	1p36	Lymphoma, leukemia	601990
P15	Cyclin dependent kinase inhibitor	9p21	Leukemia	600431
P14	MDM2 inhibitor	9p21	Colon, stomach, kidney	600160
P16/INK4a	Cyclin dependent kinase inhibitor	9p21	Multiple	600160
RB	Cell-cycle inhibitor, transcriptional repressor	13q14	Retinoblastoma, multiple	180200
APC	Inhibitor of β-catenin	5q21	Digestive tract	175100
ER	Estrogen receptor	6q25	Breast	133430
DAPK	Kinase, mediator of interferon induced apoptosis	9q34	B-cell lymphoma, lung, colon	600831
STK11/LKB1	Serine/threonine kinase, interacts with p53	19p13	Colon, breast, lung, ovarian	602216
TMS1	Proapoptotic	16p11	Breast	606838
IGF2	Growth factor	11p15	Multiple	147470
TMEFF2/HPP1	Transmembrane protein	2q33	Colon, bladder	605734
DNA Repair				
MLH1	DNA mismatch repair	3p21	Colon, endometrium, stomach	120436
BRCA1	DNA damage repair, transcription, X-inactivation	17q21	Breast, ovary	113705
MGMT	Repairs DNA guanosine-methyl addcuts	10q26	Brain, colon, lung, lymphoma	156569
RASSF1A	RAS effector homolog, DNA repair	3p21	Multiple, lung	605082
Metastasis				
CDH1	E-cadherin, cell–cell adhesion	16q21	Breast, stomach, thyroid, leukemia	192090
TIMP3	Extracellular matrix degradation and cell migration	22q12	Colon, renal, brain	188826
CDH13	H-cadherin, cell adhesion	16q24	Breast, lung	601364
Metabolism				
GSTP1	Conjugation of glutathione, prevention of oxidative damage	11q13	Prostate, breast, kidney	134660

(*continued overleaf*)

Tab. 6 (*continued*)

Gene	*Function*	*Location*	*Tumor types*	*OMIM*[a]
COX2	Cyclooxygenase 2, prostaglandin metabolism	1q25	Colon, stomach	600262
FHIT	Fragile histidine triad, hydrolase	3p14	Clear cell renal, esophogus, stomach, colon	601153
Transcription				
HLTF/SMARCA3	SWI/SNF-related, matrix-associated actin-dependent regulator of chromatin	3q25	Colon	603257
HIC1	Zn-finger transcription factor	17p13	Multiple	603825
RIZ1	Histone/protein methyltransferase	1p36	Breast, liver	601196
Hormone and cytokine responses				
PR	Progesterone receptor	11q22	Breast	607311
AR	Androgen receptor	Xq11	Prostate	313700
SOCS1	Stat inhibitor	16p13	Liver	603597
SYK	Tyrosine kinase involved in immune function	9q22	Breast	600085
Angiogenesis				
VHL	Ubiquitin ligase component, promotes angiogenesis	3p25	Kidney hemangioblastoma, clear cell renal	193300
THBS1	Thrombospondin-1, angiogenesis	15q15	Glioma	188060

Note: RB: retinoblastoma protein; APC: adenomatous polyposis coli; TIMP3: tissue inhibitor of metalloproteases-3; DAPK1: death associated protein kinase; TMS1: target of methylation-induced silencing-1; IGF2: Insulin-like growth factor-2; TMEFF2: transmembrane protein with egf-like and 2 follistatin-like domains; BRCA1: breast cancer-1, early onset; MGMT: *O*-6-methylguanine-DNA methyltransferase; GSTP1: glutathione-*S*-transferase, pi-1; HIC1: hypermethylated in cancer-1; SOCS1: suppressor of cytokine signaling 1; SYK: spleen tyrosine kinase; VHL: Von Hippel–Lindau.

[a]Online Mendelian Inheritance in Man™: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db = OMIM. Website maintained by the National Library of Medicine. Each number is unique identifier of gene-specific information including a review of relevant clinical and basic research and up-to-date links to pertinent primary literature.

APCs involvement in colon polyp and tumor formation led to the identification of methylation as a potential factor, the role for methylation in the progression of polyps to tumors may have been best shown using a mouse model of APC. The APC^{-min} allele in mice is a functional null allele discovered in a mutagenesis screen for colon polyp formation. These animals produce multiple polyps that progress to tumors similar to humans. The *DNMT1* status is a determinant in polyp-formation susceptibility in the APC^{-min} animals. When animals deficient in one copy of *APC* ($APC^{min/+}$) were mated to mice deficient in DNMT1 activity, the resulting animals had reduced intestinal adenomas compared to $APC^{min/+}$ animals with normal DNMT1 activity. This result is concomitant with decreased CpG-island methylation indicating that this decrease in methylation may be the protective mechanism against polyp formation. Therefore, it is possible that treatments to decrease methylation in colon cells may be effective against tumors associated with loss of APC.

The estrogen receptor (ER) is a hormone activated transcription factor that plays a necessary role in cell proliferation of estrogen-responsive tissues such as the uterus and mammary epithelium. In cell lines that have lost ER, readdition of ER protein results in growth inhibition indicating that loss of ER expression in preneoplastic tissues may provide a growth advantage. Estrogen receptor gene has two promoters associated with CpG islands and age-related methylation. Increases in CpG-island methylation at *ER* is linear with age, increasing in human colon tissue from 2 to 20% through the second to the ninth decades of life. Moreover, the degree of methylation is progressive in prostate and colon tumor formation and is consistent with a selective advantage for this modification. Thus, silencing of *ER* is one example of several alleles, also including *MYOD1*, *N33*, and *Pax6*, that undergo both age-related methylation and selection for methylation-induced silencing in tumors. The status of ER in tumors is an important factor in tumor treatment strategies. The loss of ER in breast tumors is often associated with more aggressive cancers and negative clinical outcomes. ER-negative tumors are not responsive to tamoxifen, an anti-estrogen drug found to be a highly effective treatment in ER-positive breast tumors. Furthermore, tumors that escape doxorubicin treatment have often lost ER expression indicating that this particular treatment for ER sensitive tumors may predispose the tumor to become more aggressive. Therefore, the ability to reactivate a silenced ER in tumor cells could be an important advantage in treatment strategy.

The P73 protein is similar in sequence and secondary structure to P53, a known regulator of cell-cycle progression and DNA-damage induced apoptosis. However, the role of P73 may be different and its mechanism of action is not entirely understood. Interestingly, *P73*-promoter hypermethylation is predominant in B-cell and Burkitt's lymphoma; more prominent than mutation or LOH. Among many tumor types, lymphomas and leukemias are the only ones in which *P73* methylation has been detected. It remains to be seen whether methylation of *P73* is causative or correlative in tumor progression in these tissues and why methylation is a more prominent form of P73 inactivation in blood tumors versus solid tumors. One possible reason for this tissue specificity is that normal lymphocytes are imprinted at *P73*, so loss of only the expressed copy

would be necessary to eliminate P73 protein in these cells.

3.1.2 Methylation of Metastasis-related Genes

The E-cadherin protein (CDH1) functions in calcium-dependent cell–cell adhesion and loss of CDH1 in tumors is highly correlated with tumor metastasis, especially in epithelial-derived tumors. While mutations in E-cadherin are found in a few subsets of tumor types, most often, inactivation of E-cadherin occurs through promoter methylation. Methylation frequency is very high in all sporadic tumors examined, including breast, leukemia, and esophageal, at an incidence often greater than 40%. Interestingly, methylation of *CDH1* is inversely correlated with LOH in epithelial tumors and often E-cadherin expression is reactivated in the metastasis indicating that reversal of the *CDH1* silencing is an important part of epithelial tumor progression. Other genes involved in metastasis that are associated with methylation-induced silencing include tissue inhibitor of metalloproteinases (*TIMP3*) and H-cadherin (*CDH13*) (Table 6).

3.1.3 Methylation of DNA Repair Genes

Virtually all cases of hereditary nonpolyposis colon carcinoma (HNPCC) and 13% of sporadic colon carcinomas display genomic microsatellite instability (MSI). The hallmark of MSI are changes in the length of individual single and dinucleotide repeat sequences throughout the genome due to slippage of the replication complex, and failure of the mismatch repair system. When such a repetitive sequence occurs in the coding sequence of a gene, this slippage often results in frameshift mutations. The human homologs of *Escherichia coli mutL* gene, *MLH1* and *MLH2*, are major factors in mismatch repair and hypermethylation-induced silencing of the *MLH1* promoter is found in a majority of colorectal cancers with MSI. Because MLH1 protein is critical in DNA repair, when mutation or silencing occurs, the effect can be genome-wide. Given these associations, it is suspected that a deficiency in methylation activity would be protective against formation of these tumors. When *MLH1* deficient mouse models were made to be deficient in *DNMT1* by mating to a hypomorphic *DNMT*-1 mouse, two opposing results were obtained. First, the incidence of intestinal cancers was decreased consistent with data obtained in human patients interlinking increased DNA methylation concurrent with MSI. However, in these same animals there is a dramatic increase in the incidence of lymphoid tumors. These results indicate that the role of DNA methylation in intestinal cancers is distinct from that involved in lymphoid tumor formation.

BRCA1 is mutated in almost half of familial linked breast cancer cases, although few mutations have been described in sporadic breast cancer. However, recent data indicate that a significant portion of sporadic ovarian and breast tumors are methylated rather than mutated for *BRCA1* and this modification is often associated with LOH. Similar to *P73*, *BRCA1* inactivation by methylation is highly tissue-specific with no methylation reported to be associated with *BRCA1* in colon, leukemia or liver tumors. While the role of BRCA1 protein in either normal cells or progression to tumor cells is poorly understood, BRCA1 protein has been reported to interact with several proteins involved in DNA repair and recombination, implicating a role in mediating this process. Importantly, a recent and

particularly interesting observation indicates a role for BRCA1 in targeting the *Xist* RNA to the X chromosome. In this study, cells that lack BRCA1 were shown to have abnormal X-inactivation and loss of focal H3(Lys9) methylation along the inactivated X chromosome. It will be of interest to determine if BRCA1 plays a more global role in RNA-mediated alterations of chromatin structure such as the recent discovery of RNA-mediated pericentric heterochromatin formation.

BRCA1 represents an example of the impact of promoter methylation on gene activation by interference with specific transcription factor binding (see Fig. 3(b)). In this case, the screening of a series of randomly sampled breast carcinomas revealed the presence of CpG methylation adjacent to the *BRCA1* transcription start site, whereas in control normal tissues these same regions were unmethylated. One such methylated CpG occurs at a CREB (cAMP-responsive element binding) transcription factor binding site in the *BRCA1* promoter. Methylation prevents binding of CREB and the recruitment of transcription factors leading to downregulation of *BRCA1* in a breast tumor cell line. Thus, one possible mechanism of transcriptional inactivation of the *BRCA1* gene within breast tumors may be interference due to methylation of this specific sequence rather than CpG-island hypermethylation and subsequent heterochromatin formation.

3.1.4 Methylated DNA Binding Proteins

The MBD4 protein interacts *in vitro* with MLH and is thought to participate in DNA mismatch repair by removing thymine or uracil from mismatched CpG sites. Mutations in MBD4 are found in colorectal carcinomas, tumors with microsatellite instability, endometrial, and pancreatic tumors with 20 to 25% incidence. A microsatellite "hotspot" in a polyadenine tract in the MBD coding region is often the site where mutations are found. A mouse model deficient in MBD4 had a threefold increase in C to T transitions in CpG dinucleotides throughout the genome. When crossed to APC^{min} allele, MBD4 deficiency increased rate of tumor formation and mutability of *APC* gene. Together these results indicate that MBD4 is required to suppress C to T transitions presumably by recognizing methylated Cs in the parent strand during mismatch repair.

3.2 Histone Modifications in Tumorigenesis

Evidence of the importance of histone-modifying proteins as epigenetic regulators of transcription, chromatin structure, and tumor formation is increasing. Indeed, as mentioned above, silencing of the cell-cycle control protein $P16^{INK4a}$ may be initiated through histone modifications prior to DNA methylation. Here we outline specific examples of known roles that these proteins play in human disease and some of the mouse models generated in attempts to gain more insight into the functions of these proteins. Proteins that modify histones can be inactivated in tumors by point mutation, deletion, translocation, and in some cases methylation. Interestingly, some translocations involving HATs and HMTs result in proteins that gain new functions leading to tumors.

3.2.1 Histone Methyltransferases

By a microarray approach the overexpression of the SET-domain histone methyltransferase EZH2 was recently found to be highly correlated with metastatic prostate tumor tissues as compared with

benign prostate or localized prostate tumors. EZH2 has a SET domain and has recently been shown to have bona fide histone methyltransferase activity targeting H3(Lys27) in nucleosomal DNA. EZH2 associates with HDACs in a complex that also contains the WD-repeat protein EED that is important in maintenance of imprinted X-inactivation. Negative regulation of gene expression in prostate tumor cell lines by EZH2 is dependent on an intact SET domain and HDAC activity. Furthermore, repression of EZH2 in prostate tumor cell lines reduced their cell growth.

Another histone methyltransferase, Suv39H1, is implicated in tumorigenesis by virtue of interaction with Rb in the progression of the cell cycle. The ability of Rb to block E2F transcription of the cyclin E promoter is mediated in part by HP1 and Suv39H1. Suv39H1 exhibits cell-cycle dependent changes in activity, such that it is active in G1, but becomes inactive during G1–S-transition and is reactivated late in S-phase. Although a definitive role of this allele or its homolog *Suv39H2* in human primary tumors has not been demonstrated, results from gene targeting in mice implicates one. Null alleles of *Suv39H1* and *Suv39H2* as well as transgenic animals that overexpress Suv39H1 have been characterized. The results from phenotypic and chromatin analysis of these models is consistent with a role for Suv39H proteins in heterochromatin formation and uncovered a role for Suv39H in chromosomal segregation during mitosis. Individually, *Suv39H1* and *Suv39H2* null mice do not display abnormal phenotypes and are fertile. However, *Suv39H1/H2* double-knockout mice develop late onset B-cell lymphomas that resemble non-Hodgkin's lymphoma in humans. Primary fibroblast cells lacking Suv39H1 and H2 quickly become aneuploid and are prone to aberrant mitosis and chromosomal segregation. Furthermore, mitotic profiles of H3(Lys9) methylation are severely altered or diminished in both fibroblast and bone marrow cells. Overexpression of Suv39H1 in transgenic animals impairs differentiation but, remarkably, can also lead to spontaneous immortalization without crisis in fetal hepatocytes and erythrocytes. The SET domain is also known to be a protein–protein interaction motif and in some cases this may be the primary function of this domain. It remains to be seen if chromatin targets or other cellular targets may be responsible in increasing tumor transformation in these animals.

RIZ1 is a zinc-finger and SET-domain containing tumor suppressor gene identified by virtue of its interaction with Rb. The location of *RIZ1* in the human genome is 1p36, a region often deleted in human cancers. *RIZ1* frame-shift mutations are found in 37% of microsatellite-instable tumors and the *RIZ1* promoter is also a target for hypermethylation-mediated repression. The *RIZ1* locus codes for two protein products, RIZ1 and RIZ2 that are controlled by separate promoters. RIZ2 is smaller than RIZ1 and lacks the SET-domain. Importantly, RIZ1, but not RIZ2 expression is lost in breast cancer, lung cancer, osteosarcoma, hepatoma, neuroblastoma, and colon cancer indicating that the SET-domain portion of RIZ1 is important in its tumor suppressor role. Although RIZ1 contains a SET domain and interacts with several transcription factors including Rb, ER, and GATA3, an HMT activity has not yet been demonstrated. The generation of a null allele of *Riz1* in mice provided perhaps the best evidence of its role as a tumor suppressor. Homozygous *Riz1*-null animals are viable and fertile, but

have increased incidence of multiple tumors especially diffuse B-cell lymphoma, one of the most common non-Hodgkin's lymphomas found in humans. Investigators showed that in cell lines that lack RIZ1, exogenous addition of RIZ1 significantly slowed growth both in cell culture and in the capacity of cells to form tumors when injected subcutaneously into mice. Gene therapy attempts in mice that target MSI+ tumors using RIZ1 expression vectors are encouraging.

The gene for another SET-domain containing protein, *MDS-EVI1* is located in a region of chromosome 3 often involved in translocations in myeloid leukemias. Similar to RIZ1 and RIZ2, the EVI1 and MDS1 protein products are also derived from the same locus under different promoter control. In this case, MDS1-EVI1 contains a SET domain and zinc-finger domain, whereas EVI1 lacks the SET domain. EVI1 acts as an oncogene when found in translocation products acting to increase growth rates in tumor cell lines. However, expression of the *MDS1-EVI1* gene product represses growth. A transgenic mouse model was generated to express the translocation produced *MDS1-EVI1-AML* fusion only in the blood cells. These mice die quickly of acute myelogenous leukemia and can serve as a model system to understand progression and treatment options for this type of leukemia. As with RIZ1, the presence of the SET-domain has not yet been associated with histone modifications.

Many SET-domain proteins are involved in translocation events in tumors. Mixed lineage leukemia (MLL) susceptibility mapped in humans at chromosome 11q23 was subsequently shown to code for the homolog of *Drosophila* TRX protein, a SET-domain containing protein (see Sect. 2.2.2). MLL may present an example of the cross-talk between histone modifications and the SWI/SNF complexes. MLL is a transcriptional effector that targets H3(Lys4) for methylation and has been shown to directly interact with SNF5 and Brg1. The human *MLL* allele is involved in as many as 30 different translocations found in 80% of infant leukemias. These translocations are thought to result in a gain of function for MLL because of the activity of the fusion. Consistent among these translocations is that the new product retains the amino-terminal 1300 amino acids, but loses the plant homeobox domain (PHD), transactivation, and SET domains of MLL. The exact role of many of the fusions in tumorigenesis is yet to be determined, but one recent suggestion noted from experiments using mouse *Mll*-null cells is that the mutant fusions involving *MLL* may cause cells to bypass cellular differentiation–induced downregulation of MLL-targets such as *Hox* genes.

3.2.2 Histone Acetyltransferases

Misdirection of HAT activity occurs as a result of several chromosomal translocations involving HATs. Among the HATs most commonly associated with tumorigenesis is the CREB binding protein (CBP). CBP is a large protein with many functional domains including three putative zinc-finger interaction motifs, an acetyltransferase domain, a plant homeobox domain, and a bromodomain. The locus that codes for CBP, *Crebbp*, is involved in several translocation events. Among these are translocations that create fusions with either *MOZ* (monocytic leukemia zinc-finger protein, t(8;16)(p11;p13)) or *MLL* (t(11;16) (q23;p13.3)) associated with leukemia. Translocations of *Crebbp* involving either *MOZ* or *MLL* result in fusion proteins that gain novel functions that may result in aberrant histone

acetylation. Loss-of-function mutations in *Crebbp* result in the multisymptom Rubenstein–Taybi syndrome associated with an increase in tumor incidence, especially leukemias. Mice that lack CBP also display many of these same symptoms including a predisposition to leukemias. Furthermore, truncations and LOH associated with P300, an acetyltransferase and close homolog of CBP, is implicated in gastric cancers and glioblastomas. It is important to note that both CBP and P300 are known to acetylate many proteins in addition to histones including P53 and other members of the general transcriptional machinery. Thus, the exact role of histone substrates in mediating tumor formation associated with these proteins is unclear.

The "*a*mplified *in* *b*reast protein", AIB1, (ACTR/pCIP/TRAM1/RAC-3/SRC-3) is a member of the P160 subfamily of histone acetyltransferases and was found during a search for genes amplified in breast tumors. AIB1 overexpression is found in more than 60% of breast tumors analyzed and enhances steroid receptor activity by interaction with the steroid receptors themselves. In some cases, amplification of AIB1 is associated with loss of steroid receptor expression or function, which could be a factor in treatment outcomes when using antiestrogens, such as tamoxifen. It is also overexpressed in prostate tumors, ovarian, pancreatic, endometrial cancers, and primary gastric tumors, and is associated with metastases and poorer prognosis. A *Drosophila* homolog of AIB1 was shown to affect cell–cell contact and focal adhesion, but its role in mammalian biology is not completely understood.

The role of HDACs in tumor formation derives from the aberrant targeting of HDACs to promoters as a result of fusions between unrelated transcriptional activators and transcriptional repressors. For example, a transcription factor such as retinoic acid receptor alpha (*RARA*) when fused to promyelocytic leukemia, (*PML*) [t(15;17)(q22;q21)] is no longer hormone responsive, and recruits HDACs to retinoic acid response elements in DNA leading to repression of retinoic acid mediated gene expression and acute promyelocytic leukemia. Conversely, other fusions of AML (Acute myelogenous leukemia) with the transcriptional regulators *TEL1* [t(1;12)(q21;p13)] or *ETO* [t(8;21)(q22;q22)] convert AML from activator to repressor via a gain of HDAC recruitment in the fusion proteins. The role of HDACs in the consequences of these fusions identifies these types of tumor events as good targets for HDAC inhibitor–based treatments.

3.3 Imprinting and Tumorigenesis

Expression of the Wilms' tumor gene, WT, located on human chromosome 11p, is lost in most cases of sporadic and familial Wilms' tumor, a common embryonal tumor that affects the kidney. Within this region of 11p is a cluster of imprinted alleles. In Wilms' tumor-associated LOH on chromosome 11p, there is significant bias for loss of the maternal allele. This implies a growth advantage for the loss of a maternally expressed allele near the *WT* allele. Candidate alleles include the maternally expressed H19 RNA. The paternally expressed insulin-like growth factor 2 (*IGF2*) and *H19* are coordinately regulated in part by differential DNA methylation in a common promoter region. Failure to maintain imprinting at a region of chromosome 11p also leads to biallelic expression of the normally maternally silenced *IGF2*. This effect is tightly linked to the loss of maternally expressed, growth-inhibitory *H19* allele in Wilms' tumor. While early investigations

focused mainly on embryonal and childhood tumors, more recent studies have found that loss of imprinting of IGF2 is associated with most other tumors including hepatocellular, colorectal, brain, and testicular cancers.

3.4 Mouse Models

The ability to manipulate the genome of the common house mouse by gene targeting and transgenesis has propelled the mouse to the forefront in genetic modeling of human disease. Mice are not particularly susceptible to tumor formation and are particularly resistant to tumor metastasis, presumably because of a shorter life span. Therefore, small increases in tumor incidence in mice can be a strong indicator of the importance of a particular process in tumor formation. Conversely, lack of tumor growth can lead to an underestimation of a particular gene's importance. Furthermore, often the spectrum of tumor types resulting from a given mutation in mice is different from those in humans. Therefore, it is critical that results from animals be accepted only in the context of supporting human data from actual tumor biopsies. Nonetheless, the fact that mice get spontaneous tumors at any frequency warrants the development of the mouse as a tool in the war on cancer. Increasing use of methods such as viral oncogene delivery and transgenesis are beginning to "humanize" the mouse in tumor processes. Further development of humanized tumors in mice should have a significant impact on our understanding of tumor progression and treatment research in the future.

In addition to the animal models discussed in the above sections, null alleles have been generated for genes that are affected by methylation such as in *p16*, *p73*, *BRCA1*, *MLH*, *MLL*, *ERα*, and *CDH1*, all revealing their importance in development and tumor suppression. Models have also been generated at alleles for proteins responsible for modification of histones such as *CBP*, *P300*, *GCN5*, *PCAF*, and *G9a* and remodeling proteins such as *Brg1* and *Snf5*. In many mouse models, a heterozygous genotype is not enough to change tumor incidence and often homozygous null genotypes are embryonic lethal. While embryonic lethal phenotypes underscore the importance of a gene in developmental biology, this phenotype prevents the ability of researchers to determine roles of alleles in tumor formation. Many creative ways to circumvent embryonic lethality have been employed including conditional mutagenesis using the Cre-LoxP system of recombination to direct a null genotype to only a single tissue type. In order to determine whether the lack of tumors indicates a protective role, comparisons of phenotypes on tumor-model backgrounds, as done in the case of *Dnmt1* and the crosses to the APC^{min} and *Mlh1* mutant animals, is required (see Sect. 3.1.1 and 3.1.2).

Derivation of homozygous null ES cells and primary embryonic fibroblasts from mice carrying mutations important in establishing and maintaining epigenetic marks can give insight into cell autonomous functions in culture. The use of ES cells in studies of effects of mutations on growth and differentiation has become a welcome trend. ES cells maintain properties of tumor cells such as the ability for continual growth in an undifferentiated state. However, ES cells are stably diploid and lack the plethora of mutations found in most tumor cell lines. Additionally, the capacity of ES to be differentiated into many different cell types will allow definition of a specific role for a mutated gene in

a differentiation-mediated process, not the least important of which includes growth inhibition. For example, the roles of either DNMT1 or the histone methyltransferase G9a in growth only became apparent in differentiated (neuralized) ES, rather than in undifferentiated ES cells indicating a more important role for these proteins in differentiated cells. Furthermore, cell culture models, derived from mutant animals, offer an excellent environment to test the role of epigenetics in foreign DNA recognition and genomic incorporation.

4 Clinical Considerations

4.1 Detections

Methylation of certain promoter regions appears to be an early mark in tumorigenesis. For instance, promoter methylation of $P16^{INK4a}$ is detected in some colorectal adenomas and other precancerous lesions prior to development of the carcinoma. Therefore, it becomes of great importance to further our knowledge of genes, that when methylated can be termed markers of pre-neoplastic tissue, and methods of rapidly determining these changes in small amounts of tissue or body fluids. There are advantages of using methylation rather than classical mutation as a marker for pre-neoplasia. In this regard, mutations occur in many different forms and locations throughout a gene sequence, whereas methylation is specific to CpG dyads and is localized to a relatively small area of a locus. However, hypermethylation is not a ubiquitous characteristic in all tumors, and therefore must be used in conjunction with other neoplastic detection strategies including P53 status. On the other hand, in MSI + tumors where P53 is often normal, methylation status may have the best association with tumor potential.

Many experimental methods have been developed to detect either regional or genome-wide CpG methylation patterns. Excellent and detailed protocols regarding many aspects of methylated DNA were published recently as part of the Methods in Molecular Biology series (see Bibliography). Perhaps the simplest method to detect methylated cytosines in the genome is to take advantage of a reaction between bisulphite and unmethylated cytosines that causes conversion of the cytosine to uracil. Cytosines that are methylated are insensitive to bisulphite. Following treatment of DNA with bisulphite, DNA is amplified by PCR incorporating thymidine in place of the converted uracil. Currently, there are two widely used ways to detect the conversion events. One way that is becoming more popular as reagents become less expensive, is to subclone PCR-products and derive the DNA sequence to look for changes of C's to T's versus a control reaction. This provides a snapshot of all methylated changes within the region amplified and allows quantitation of the number of alleles in a population that are methylated. Another method to detect methylated cytosines is to use specific PCR primers that distinguish between unmethylated cytosines and the bisulphite mediated change. The advantages of this method are cost-effectiveness and rapid results, although the disadvantage is its ability to only detect a single change per reaction. There are additional ways to detect methylated cytosines without bisulphite including the use of methylation-sensitive restriction enzymes directed at specific sites in a promoter followed by Southern blot or ligation-mediated PCR.

In order to gain insight into different methylation patterns of specific tumor types, many researchers use these methods in the candidate gene approach to examine methylation of specific tumor suppressors. Because of the large body of work that had identified tumor suppressors subject to methylation and the sites of these aberrant methylations, this approach has yielded significant insights, for example, into designing new markers for MSI(+) tumors. While the candidate gene approach has yielded many promising leads, it can only render an incomplete picture. The availability of drugs to block promoter methylation, and the availability of cDNA microarrays has allowed more broadly based approaches that provide information not necessarily about promoter methylation, but the downstream effects methylation has on physiology. Such approaches have led to important discoveries in finding additional methylation-sensitive marker alleles, such as those noted above, linking increased EZH2 expression and progression of prostate cancer. Analysis such as these will yield potential PCR-based diagnostics easily adapted to the clinical lab. In addition to gene expression pattern changes, it is also possible to use large-scale methods such as DNA microarrays to detect the presence of methylation at specific promoters. To accomplish this, microarrays of CpG-island sequences derived from the genome database can be hybridized with tumor DNA digested with methylation-sensitive enzymes. The results can detect changes between methylation status in normal and tumor tissues at thousands of CpG islands that may be included in the array.

While the methods outlined above require prior knowledge of sequences of interest, there are also several ways to isolate unknown differentially methylated regions or obtain a global picture of the methylation status in cells. So-called genome-wide approaches include restriction landmark genome scanning in which total genomic DNA is first digested with a methylation-sensitive enzyme, then digested with methylation-insensitive enzymes, and separated in two dimensions. Fragments that differentially migrate can be isolated from the gels and subcloned for identification. A significant advantage of this approach is that it can be adapted to detect LOH in addition to methylation within the same tumors. Restriction landmark genome scanning in combination with computer-aided mapping of the recently completed human genome sequence has been used to successfully identify differentially methylated and deleted genes specific for brain tumors. Moreover, using such genomic approaches, as well as the candidate gene approach, the resulting profile of genes aberrantly methylated in a given tumor may provide a window into the biology of the tumor and allow tailoring of treatment strategies.

Methods for determining histone modification status pose a different set of problems. Possibly the most promising method for detection of global levels of certain histone marks is through the development of antibodies to histone tail domains that contain specific modifications. Antibodies have been developed that specifically recognize methylation, phosphorylation, and acetylation on the individual core histones. The correlation between global levels of these marks and neoplastic status of cells is yet to be determined. Methods that can detect marks on specific loci include chromatin immunoprecipitation where these same antibodies are used to precipitate chromatin followed by gene-specific PCR to determine the quantity of a region that is immunoprecipitated. While they

are promising, the current methods require more material than most standard biopsies provide and are time-consuming. The use of mass spectrometry to detect the phosphorylation status of histone H1 has successfully been used to look at the global dephosphorylation of H1 in cells in response to drug treatments such as dexamethasone. Given the differences in mass of acetyl, methyl, and phospho groups, it is likely that a mass spectrometry method could be adapted to detect the status of these marks on all histones. For histone modifications to be utilized in diagnostics, however, will require that researchers crack the "histone code" and be able find histone modifications that are relatively stable and clonal in tumor populations similar to DNA methylation.

4.2 Treatments

4.2.1 DNA Methyltransferase Inhibitors

The drug 5-Azacytidine (5-Aza) was first discovered to have antimetabolic activities and to be an effective therapeutic agent for acute myelogenous leukemia in the late 1960s and early 1970s. Incorporation of 5-Aza and an analog 5-Aza-2′deoxycytidine (5Aza2d) into DNA in the place of methylated cytosines results in an irreversible binding and inhibition of DNMTs. Unfortunately, this irreversible binding of DNMTs leads to DNA damage. In tumors that have lost the DNA-damage response mechanisms, many more mutations can be clonally introduced into the tumor genome following 5-Aza treatments. Because of the role of P53 protein in initiating cell-death in response to DNA damage, it is not surprising that it was discovered in leukemias, since p53 positive tumors are the most responsive to 5-Aza treatments. Another phenomenon observed following 5-Aza treatment is that cells tend to remethylate DNA over time. Recently, analysis of resilencing of the $P16^{INK4a}$ promoter following demethylation in a colon tumor cell line indicates that histone modifications may initiate this process.

Using DNA microarray technology and differential display, it is known that only a 1–2% change in genome transcription occurs with treatment of tumor cells with 5-Aza indicating that not all hypermethylated alleles are reactivated. However, the modest increases that occur following 5-Aza can be further induced by cotreatment with HDAC inhibitors. While HDAC inhibition alone is not enough to overcome DNA-methylation induced silencing, these results may have greater meaning when evaluating the efficacy of HDAC inhibitors in the future in combination with 5-Aza (see below).

Besides 5-Aza, other agents are being researched for more specific downregulation of DNMTs including targeted antisense RNAs and bended oligonucleotides. These, however, do not appear to be as effective as 5-Aza or 5-Aza2d. Another promising and as-yet-unrealized approach is to target methyl binding proteins. These are often found in larger complexes and may invoke the spreading of methylation from one locus to the next and transmission of methylated status from mother to daughter cells. By designing small-molecule inhibitors to methyl binding proteins, one may be able to prevent recognition and spread of methylated cytosines across replication.

4.2.2 HDAC Inhibitors

The recognition that heterochromatin is profoundly hypoacetylated led to the hypothesis that deacetylase inhibitors may be used to reactivate silenced genes. Historically, HDAC inhibitors were identified

in screens for chemicals that induce differentiation or inhibit cell cycle in tumor cells. These chemicals such as Trichostatin A (TSA) and sodium-*N*-butyrate (butyrate) induce hyperacetylation in cells concomitant with effects on cell-cycle progression. Trichostatin A causes cell-cycle arrest in either G1 or G2 and inhibits differentiation of the neuronal cell line PC12. Furthermore, butyrate also causes cell-cycle inhibition in culture and has been shown to suppress colon tumors in mice. Conversely, HDAC activity can be important in mediating cell-cycle inhibition as in the case of Rb that requires HDACs to inhibit E2F mediation of cell-cycle entry into S-phase.

Inhibiting HDAC activity can relieve repression on certain tumor suppressor promoters in cell culture model systems. Suberoylanilide hydroxyamic acid (SAHA) is in the same class of TSA and inhibits HDAC activity by blocking substrate access to the catalytic domain. Other hydroxiamic acids include *m*-carboxy cinnamic bishydroxamic acid (CBHA) and oxamflatin can be used at submicromolar concentrations and have longer half-lives than SAHA. Trapoxin A, apicidin, and depsipeptide (FR901228) are fungal metabolites that are tetrapeptides and irreversibly bind the catalytic site of HDACs. The cyclic hydroxyamic-acid containing peptide (CHAP) is a synthetic hybrid of tetrapeptides and hydroxyamic acids. Several analogs of CHAP have been tested for activity and have reversible inhibition at nanomolar concentrations. Interestingly, CHAP molecules also have specificity toward HDAC1 and HDAC4 more so than HDAC6. Preclinical studies using mice have proved exciting for butyrates, SAHA, and valproic acid and several HDAC inhibitor drugs such as phenylbutyrate and despeptide have proceeded to clinical trials with encouraging results. Because many DNA methyltransferase complexes also contain HDACs, HDAC inhibitors in combination with methyltransferase inhibitors have been or are being developed as one potential method of preventing the spreading of hypermethylated DNA in tumor cells.

Gene expression analysis following HDAC inhibitor treatments in cell lines indicated less than 10% change in gene expression profile. Although almost all HDAC inhibitors induce p21 and downregulate cyclins A and D and hypophosphorylation of Rb, it may be that HDAC inhibitors work as antitumor agents by their ability to block the cell cycle and induce apoptosis independently of chromatin modification and gene expression. It is possible that there is a critical role for HDACs in other posttranslational modifications of critical proteins in the cell cycle. There are approximately 17 members of the HDAC family with putative HDAC activity; however, only the class 1 HDACs are found in known protein complexes related to transcriptional repression. Because of the broad specificity of HDAC inhibitors in mediating cell-cycle inhibition and apoptosis, it remains to be seen whether individual HDACs are more involved than others. Development of HDAC inhibitors with subtype specificity will allow research to narrow the scope of inhibitor targets important in antitumor effects and from this knowledge produce better and better drugs.

5 Conclusions

In 2003, 50 years after the discovery of the double helix and 2 years after the successful sequencing on the human genome,

it is clear that DNA represents but the initial layer of genetic information. Superimposed on this beautiful structure are the instructions for its packaging into chromatin and specific epigenetic marks on the resultant DNA and histones. The importance of epigenetic changes in cancer are now well accepted on the basis of a broad spectrum of studies utilizing human and animal cell lines, animal models, and human tumors. The application of mechanistic principles to the functions of the RNA and protein factors so far identified represent a significant avenue for research on therapeutic interventions in cancer.

See also Cancer Stem Cells; Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Baylin, S.B. (2002) Mechanisms underlying epigenetically mediated gene silencing in cancer, *Semin. Cancer Biol.* **12**, 331–337.

Johnstone, R.W. (2002) Histone-deacetylase inhibitors: novel drugs for the treatment of cancer, *Nat. Rev. Drug Discovery* **1**, 287–299.

Jones, P.A., Baylin, S.B. (2002) The fundamental role of epigenetic events in cancer, *Nat. Rev. Genet.* **3**, 415–428.

Jones, P.A., Vogt, P.K. (Eds.) (2000) *DNA Methylation and Cancer*, Vol. 249, Springer-Verlag, Berlin Heidelberg, New York.

Knudson, A.G. (2001) Two genetic hits (more or less) to cancer, *Nat. Rev. Cancer* **1**, 157–162.

Mills, K.I., Ramsahoye, B. (Eds.) (2002) *DNA Methylation Protocols*, Vol. 200, Human Press, Totowa, NJ.

Robertson, K.D. (2002) DNA methylation and chromatin – unraveling the tangled web, *Oncogene* **21**, 5361–5379.

Schneider, R., Bannister, A.J., Kouzarides, T. (2002) Unsafe SETs: histone lysine methyltransferases and cancer, *Trends Biochem. Sci.* **27**, 396–402.

Strahl, B.D., Allis, C.D. (2000) The language of covalent histone modifications, *Nature* **403**, 41–45.

van Holde, K.E. (Ed.) (1989) *Chromatin*, Springer-Verlag, New York.

Wolffe, A.P. (1998) *Chromatin: Structure & Function*, Academic Press, San Diego, CA.

Primary Literature

Abbondanza, C., Medici, N., Nigro, V., Rossi, V., Gallo, L., Piluso, G., Belsito, A., Rocisgno, A., Bontempo, P., Puca, A.A., Molinari, A.M., Moncharmont, B., Puca, G.A. (2000) The retinoblastoma-interacting zinc-finger protein RIZ is a downstream effector of estrogen action, *PNAS* **97**, 3130–3135.

Allshire, R. (2002) RNAi and Heterochromatin–a hushed-up affair, *Science* **297**, 1818–1819.

Anzick, S.L., Kononen, J., Walker, R.L., Azorsa, D.O., Tanner, M.M., Guan, X.Y., Sauter, G., Kallioniemi, O.P., Trent, J.M., Meltzer, P.S. (1997) AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer, *Science* **277**, 965–968.

Ayton, P.M., Cleary, M.L. (2001) Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins, *Oncogene* **20**, 5695–5707.

Bachman, K.E., Park, B.H., Rhee, I., Rajagopalan, H., Herman, J.G., Baylin, S.B., Kinzler, K.W., Vogelstein, B. (2003) Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene, *Cancer Cell* **3**, 89–95.

Baylin, S., Bestor, T.H. (2002) Altered methylation patterns in cancer cell genomes: cause or consequence, *Cancer Cell* **1**, 299–305.

Baylin, S.B., Belinsky, S.A., Herman, J.G. (2000) Aberrant methylation of gene promoters in cancer – concepts, misconcepts, and promise, *J. Natl. Cancer Inst.* **92**, 1460–1461.

Baylin, S.B., Herman, J.G. (2000) DNA hypermethylation in tumorigenesis: epigenetics joins genetics, *Trends Genet.* **16**, 168–174.

Baylin, S.B., Herman, J.G. (2001) Promoter hypermethylation–can this change alone ever designate true tumor suppressor gene function, *JNCI Cancer Spectrum* **93**, 664–665.

Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G., Magnuson, T. (2000) A Brg1 null mutation

in the mouse reveals functional differences among mammalian SWI/SNF complexes, *Mol. Cell* **6**(6), 1287–1295.

Burgers, W.A., Fuks, F., Kouzarides, T. (2002) DNA methyltransferases get connected to chromatin, *Trends Genet.* **18**, 275–277.

Cameron, E.E., Bachman, K.E., Myohanen, S., Herman, J.G., Baylin, S.B. (1999) Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer, *Nat. Genet.* **21**, 103–107.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H. Tempst, P., Jones, R.S., Zhang, Y. (2002) Role of histone H3 lysine 27 methylation in polycomb-group silencing, *Science* **298**, 1039–1043.

Chan, M.F., van Amerongen, R., Nijjar, T., Cuppen, E., Jones, P.A., Laird, P.W. (2001) Reduced rates of gene loss, gene silencing, and gene mutation in Dnmt1-deficient embryonic stem cells, *Mol. Cell Biol.* **21**, 7587–7600.

Chen, R.Z., Akbarian, S., Tudor, M., Jaenisch, R. (2001) Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice, *Nat. Genet.* **27**, 327–331.

Christman, J.K. (2002) 5-Azacytidine and 5-aza-2′-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy, *Oncogene* **21**, 5483–5495.

Cuenco, G.M., Nucifora, G., Ren, R. (2000) Human AML1/MDS1/EVI1 fusion protein induces an acute myelogenous leukemia (AML) in mice: a model for human AML, *PNAS* **97**, 1760–1765.

Czvitkovich, S., Sauer, S., Peters, A.H., Deiner, E., Wolf, A., Laible, G., Opravil, S., Beug, H., Jenuwein, T. (2001) Over-expression of the SUV39H1 histone methyltransferase induces altered proliferation and differentiation in transgenic mice, *Mech. Dev.* **107**, 141–153.

DiNardo, D.N., Butcher, D.T., Robinson, D.P., Archer, T.K., Rodenhiser, D.I. (2001) Functional analysis of CpG methylation in the BRCA1 promoter region, *Oncogene* **20**, 5331–5340.

Du, Y., Carling, T., Fang, W., Piao, Z., Sheu, J.-C., Huang, S. (2001) Hypermethylation in human cancers of the RIZ1 tumor suppressor gene, a member of a histone/protein methyltransferase superfamily, *Cancer Res.* **61**, 8094–8099.

Eads, C.A., Danenberg, K.D., Kawakami, K., Saltz, L.B., Danenberg, P.V., Laird, P.W. (1999) CpG-island hypermethylation in human colorectal tumors is not associated with DNA methyltransferase overexpression, *Cancer Res.* **59**, 2302–2306.

Eads, C.A., Nickel, A.E., Laird, P.W. (2002) Complete genetic suppression of polyp formation and reduction of CpG-island hypermethylation in ApcMin/+Dnmt1-hypomorphic mice, *Cancer Res.* **62**, 1296–1299.

Esteller, M. (2002) CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future, *Oncogene* **21**, 5427–5440.

Esteller, M., Corn, P.G., Baylin, S.B., Herman, J.G. (2001a) A gene hypermethylation profile of human cancer, *Cancer Res.* **61**, 3225–3229.

Esteller, M., Fraga, M.F., Guo, M., Garcia-Foncillas, J., Hedenfalk, I., Godwin, A.K., Trojan, J., Vaurs-Barriere, C., Bignon, Y.J., Ramus, S., Benitez, J., Caldes, T., Akiyama, Y., Yuasa, Y., Launonen, V., Canal, M.J., Rodriguez, R., Capella, G., Peinado, M.A., Borg, A., Aaltonen, L.A., Ponder, B.A., Baylin, S.B., Herman, J.G. (2001b) DNA-methylation patterns in hereditary human cancers mimic sporadic tumorigenesis, *Hum. Mol. Genet.* **10**, 3001–3007.

Esteller, M., Risques, R.A., Toyota, M., Capella, G., Moreno, V., Peinado, M.A., Baylin, S.B., Herman, J.G. (2001c) Promoter hypermethylation of the DNA repair gene O(6)-methylguanine-DNA methyltransferase is associated with the presence of G : C to A : T transition mutations in p53 in human colorectal tumorigenesis, *Cancer Res.* **61**, 4689–4692.

Feng, Q., Wang, H., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Struhl, K., Zhang, Y. (2002) Methylation of H3-Lysine 79 is mediated by a new family of HMTases without a SET domain, *Curr. Biol.* **12**, 1052–1058.

Ganesan, S., Silver, D.P., Greenberg, R.A., Avni, D., Drapkin, R., Miron, A., Mok, S.C., Randrianarison, V., Brodie, S., Salstrom, J., Rasmussen, T.P., Klimke, A., Marrese, C., Marahrens, Y., Deng, C.X., Feunteun, J., Livingston, D.M. (2002) BRCA1 supports XIST RNA concentration on the inactive X chromosome, *Cell* **111**, 393–405.

Garcia-Manero, G., Daniel, J., Smith, T.L., Kornblau, S.M., Lee, M.S., Kantarjian, H.M., Issa, J.P. (2002) DNA methylation of multiple

promoter-associated CpG islands in adult acute lymphocytic leukemia, *Clin. Cancer Res.* **8**, 2217–2224.

Goodrich, J., Tweedie, S. (2002) Remembrance of Things Past: chromatin remodeling in plant development, *Annu. Rev. Cell Dev. Biol.* **18**, 707–746.

Hanahan, D., Weinberg, R.A. (2000) The hallmarks of cancer, *Cell* **100**, 57–70.

He, L., Yu, J., Liu, L., Buyse, I., Wang, M., Yang, Q., Nakagawara, A., Brodeur, G., Shi, Y., Huang, S. (1998) RIZ1, but not the alternative RIZ2 product of the same gene, is underexpressed in breast cancer, and forced RIZ1 expression causes G2-M cell-cycle arrest and/or apoptosis, *Cancer Res.* **58**, 4238–4244.

Hendrich, B., Bickmore, W. (2001) Human diseases with underlying defects in chromatin structure and modification, *Hum. Mol. Genet.* **10**, 2233–2242.

Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A., Bird, A. (2001) Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development, *Genes Dev.* **15**, 710–723.

Herman, J.G., Umar, A., Polyak, K., Graff, J.R., Ahuja, N., Issa, J.P., Markowitz, S., Willson, J.K., Hamilton, S.R., Kinzler, K.W., Kane, M.F., Kolodner, R.D., Vogelstein, B., Kunkel, T.A., Baylin, S.B. (1998) Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6870–6875.

Huang, S. (2002) Histone methyltransferases, diet nutrients and tumour suppressors, *Nat. Rev. Cancer* **2**, 469–476.

Issa, J.P., Ottaviano, Y.L., Celano, P., Hamilton, S.R., Davidson, N.E., Baylin, S.B. (1994) Methylation of the oestrogen receptor CpG-island links ageing and neoplasia in human colon, *Nat. Genet.* **7**, 536–540.

Jenuwein, T. (2002) Molecular biology. An RNA-guided pathway for the epigenome, *Science* **297**, 2215–2218.

Jenuwein, T., Allis, C.D. (2001) Translating the histone code, *Science* **293**, 1074–1080.

Jiang, G.L., Huang, S. (2001) Adenovirus expressing RIZ1 in tumor suppressor gene therapy of microsatellite-unstable colorectal cancers, *Cancer Res.* **61**, 1796–1798.

Jones, P.A., Baylin, S.B. (2002) The fundamental role of epigenetic events in cancer, *Nat. Rev. Genet.* **3**, 415–428.

Jones, D.O., Cowell, I.G., Singh, P.B. (2000) Mammalian chromodomain proteins: their role in genome organisation and expression, *BioEssays* **22**, 124–137.

Jones, P.A., Laird, P.W. (1999) Cancer epigenetics comes of age, *Nat. Genet.* **21**, 163–167.

Karpf, A.R., Jones, D.A. (2002) Reactivating the expression of methylation silenced genes in human cancer, *Oncogene* **21**, 5496–5503.

Khochbin, S., Verdel, A., Lemercier, C., Seigneurin-Berny, D. (2001) Functional significance of histone deacetylase diversity, *Curr. Opin. Genet. Dev.* **11**, 162–166.

Kijima, M., Yoshida, M., Sugita, K., Horinouchi, S., Beppu, T. (1993) Trapoxin, an antitumor cyclic tetrapeptide, is an irreversible inhibitor of mammalian histone deacetylase, *J. Biol. Chem.* **268**, 22429–22435.

Klochendler-Yeivin, A., Fiette, L., Barra, J., Muchardt, C., Babinet, C., Yaniv, M. (2000) The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression, *EMBO Rep.* **1**, 500–506.

Klochendler-Yeivin, A., Yaniv, M. (2001) Chromatin modifiers and tumor suppression, *Biochim. Biophys. Acta (BBA) – Rev. Cancer* **1551**, M1–M10.

Knudson, A.G. (2001) Two genetic hits (more or less) to cancer, *Nat. Rev. Cancer* **1**, 157–162.

Knudson, A.G. (2002) Cancer genetics, *Am. J. Med. Genet.* **111**, 96–102.

Kondo, Y., Shen, L., Issa, J.P.J. (2003) Critical role of histone methylation in tumor suppressor gene silencing in colorectal cancer, *Mol. Cell Biol.* **23**, 206–215.

Konishi, N., Nakamura, M., Kishi, M., Nishimine, M., Ishida, E., Shimada, K. (2002) Heterogeneous methylation and deletion patterns of the INK4a/ARF locus within prostate carcinomas, *Am. J. Pathol.* **160**, 1207–1214.

Kornberg, R.D., Lorch, Y. (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome, *Cell* **98**, 285–294.

Kouzarides, T. (1999) Histone acetylases and deacetylases in cell proliferation, *Curr. Opin. Genet. Dev.* **9**, 40–48.

Lachner, M., Jenuwein, T. (2002) The many faces of histone lysine methylation, *Curr. Opin. Cell Biol.* **14**, 286–298.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., Jenuwein, T. (2001) Methylation of histone H3

lysine 9 creates a binding site for HP1 proteins, *Nature* **410**, 116–120.

Ma, H., Baumann, C.T., Li, H., Strahl, B.D., Rice, R., Jelinek, M.A., Aswad, D.W., Allis, C.D., Hager, G.L., Stallcup, M.R. (2001) Hormone-dependent, CARM1-directed, arginine-specific methylation of histone H3 on a steroid-regulated promoter, *Curr. Biol.* **11**, 1981–1985.

Malik, K., Brown, K.W. (2000) Epigenetic gene deregulation in cancer, *Br. J. Cancer* **83**, 1583–1588.

Mancini, D.N., Rodenhiser, D.I., Ainsworth, P.J., O'Malley, F.P., Singh, S.M., Xing, W., Archer, T.K. (1998) CpG methylation within the 5′ regulatory region of the BRCA1 gene is tumor specific and includes a putative CREB binding site, *Oncogene* **16**, 1161–1169.

Marks, P., Rifkind, R.A., Richon, V.M., Breslow, R., Miller, T., Kelly, W.K. (2001) Histone deacetylases and cancer: causes and therapies, *Nat. Rev. Cancer* **1**, 194–202.

Marmorstein, R., Roth, S.Y. (2001) Histone acetyltransferases: function, structure, and catalysis, *Curr. Opin. Genet. Dev.* **11**, 155–161.

Martinez-Delgado, B., Melendez, B., Cuadros, M., Jose Garcia, M., Nomdedeu, J., Rivas, C., Fernandez-Piqueras, J., Benitez, J. (2002) Frequent inactivation of the p73 gene by abnormal methylation or LOH in non-Hodgkin's lymphomas, *Int. J. Cancer* **102**, 15–19.

Matzke, M., Matzke, A.J., Kooter, J.M. (2001a) RNA: guiding gene silencing, *Science* **293**, 1080–1083.

Matzke, M.A., Matzke, A.J.M., Pruss, G.J., Vance, V.B. (2001b) RNA-based silencing strategies in plants, *Curr. Opin. Genet. Dev.* **11**, 221–227.

Merlo, A., Herman, J.G., Mao, L., Lee, D.J., Gabrielson, E., Burger, P.C., Baylin, S.B., Sidransky, D. (1995) 5′ CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers, *Nat. Med.* **1**, 686–692.

Millar, C.B., Guy, J., Sansom, O.J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P.D., Bishop, S.M., Clarke, A.R., Bird, A. (2002) Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice, *Science* **297**, 403–405.

Momparler, R.L., Bovenzi, V. (2000) DNA methylation and cancer, *J. Cell Physiol.* **183**, 145–154.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., Grewal, S.I. (2001) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly, *Science* **292**, 110–113.

Newall, A.E.T., Duthie, S., Formstone, E., Nesterova, T., Alexiou, M., Johnston, C., Caparros, M.-L., Brockdorff, N. (2001) Primary non-random X inactivation associated with disruption of Xist promoter regulation, *Hum. Mol. Genet.* **10**, 581–589.

Nguyen, C.T., Weisenberger, D.J., Velicescu, M., Gonzales, F.A., Lin, J.C., Liang, G., Jones, P.A. (2002) Histone H3-lysine 9 methylation is associated with aberrant gene silencing in cancer cells and is rapidly reversed by 5-aza-2′-deoxycytidine, *Cancer Res.* **62**, 6456–6461.

Nielsen, S.J., Schneider, R., Bauer, U.M., Bannister, A.J., Morrison, A., O'Carroll, D., Firestein, R., Cleary, M., Jenuwein, T., Herrera, R.E., Kouzarides, T. (2001) Rb targets histone H3 methylation and HP1 to promoters, *Nature* **412**, 561–565.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A., Jenuwein, T. (2001) The polycomb-group gene Ezh2 is required for early mouse development, *Mol. Cell Biol.* **21**, 4330–4336.

Okano, M., Bell, D.W., Haber, D.A., Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development, *Cell* **99**, 247–257.

Okano, M., Li, E. (2002) Genetic analyses of DNA methyltransferase genes in mouse model system, *J. Nutr.* **132**, 2462S–2465S.

Paz, M.F., Avila, S., Fraga, M.F., Pollan, M., Capella, G., Peinado, M.A., Sanchez-Cespedes, M., Herman, J.G., Esteller, M. (2002) Germ-line variants in methyl-group metabolism genes and susceptibility to DNA methylation in normal tissues and human primary tumors, *Cancer Res.* **62**, 4519–4524.

Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., Opravil, S., Doyle, M., Sibilia, M., Jenuwein, T. (2001) Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability, *Cell* **107**, 323–337.

Plass, C., Soloway, P.D. (2002) DNA methylation, imprinting and cancer, *Eur. J. Hum. Genet.* **10**, 6–16.

Rice, J.C., Allis, C.D. (2001) Histone methylation versus histone acetylation: new insights into epigenetic regulation, *Curr. Opin. Cell Biol.* **13**, 263–273.

Robert, M.F., Morin, S., Beaulieu, N., Gauthier, F., Chute, I.C., Barsalou, A., MacLeod, A.R. (2003) DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells, *Nat. Genet.* **33**, 61–65.

Roberts, C.W.M., Galusha, S.A., McMenamin, M.E., Fletcher, C.D.M., Orkin, S.H. (2000) Haploinsufficiency of Snf5 (integrase interactor 1) predisposes to malignant rhabdoid tumors in mice, *PNAS* **97**, 13796–13800.

Robertson, K.D. (2001) DNA methylation, methyltransferases, and cancer, *Oncogene* **20**, 3139–3155.

Roth, S.Y., Denu, J.M., Allis, C.D. (2001) Histone acetyltransferases, *Annu. Rev. Biochem.* **70**, 81–120.

Sassone-Corsi, P., Mizzen, C.A., Cheung, P., Crosio, C., Monaco, L., Jacquot, S., Hanauer, A., Allis, C.D. (1999) Requirement of Rsk-2 for epidermal growth factor-activated phosphorylation of histone H3, *Science* **285**, 886–891.

Singh, H., Sekinger, E.A., Gross, D.S. (2000) Chromatin and cancer: causes and consequences, *J. Cell Biochem. Suppl.* **Suppl 35**, 61–68.

Steele-Perkins, G., Fang, W., Yang, X.H., Van Gele, M., Carling, T., Gu, J., Buyse, I.M., Fletcher, J.A., Liu, J., Bronson, R., Chadwick, R.B., de la Chapelle, A., Zhang, X.-k., Speleman, F., Huang, S. (2001) Tumor formation and inactivation of RIZ1, an Rb-binding member of a nuclear protein-methyltransferase superfamily, *Genes Dev.* **15**, 2250–2262.

Strathdee, G. (2002) Epigenetic versus genetic alterations in the inactivation of E-cadherin, *Semin. Cancer Biol.* **12**, 373–379.

Suzuki, H., Gabrielson, E., Chen, W., Anbazhagan, R., van Engeland, M., Weijenberg, M.P., Herman, J.G., Baylin, S.B. (2002) A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer, *Nat. Genet.* **31**, 141–149.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., Shinkai, Y. (2002) G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis, *Genes Dev.* **16**, 1779–1791.

Trinh, B.N., Long, T.I., Nickel, A.E., Shibata, D., Laird, P.W. (2002) DNA methyltransferase deficiency modifies cancer susceptibility in mice lacking DNA mismatch repair, *Mol. Cell Biol.* **22**, 2906–2917.

Turner, B.M. (2002) Cellular memory and the histone code, *Cell* **111**, 285–291.

Usadel, H., Brabender, J., Danenberg, K.D., Jeronimo, C., Harden, S., Engles, J., Danenberg, P.V., Yang, S., Sidransky, D. (2002) Quantitative adenomatous polyposis coli promoter methylation analysis in tumor tissue, serum, and plasma DNA of patients with lung cancer, *Cancer Res.* **62**, 371–375.

Varambally, S., Dhanasekaran, S.M., Zhou, M., Barrette, T.R., Kumar-Sinha, C., Sanda, M.G., Ghosh, D., Pienta, K.J., Sewalt, R.G., Otte, A.P., Rubin, M.A., Chinnaiyan, A.M. (2002) The polycomb group protein EZH2 is involved in progression of prostate cancer, *Nature* **419**, 624–629.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., Martienssen, R.A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi, *Science* **297**, 1833–1837.

Wang, H., Huang, Z.Q., Xia, L., Feng, Q., Erdjument-Bromage, H., Strahl, B.D., Briggs, S.D., Allis, C.D., Wong, J., Tempst, P., Zhang, Y. (2001) Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor, *Science* **293**, 853–857.

West, A.G., Gaszner, M., Felsenfeld, G. (2002) Insulators: many functions, many mechanisms, *Genes Dev.* **16**, 271–288.

Wolffe, A.P. (2001) Chromatin remodeling: why it is important in cancer, *Oncogene* **20**, 2988–2990.

Xu, J., Liao, L., Ning, G., Yoshida-Komiya, H., Deng, C., O'Malley, B.W. (2000) The steroid receptor coactivator SRC-3 (p/CIP/RAC3/AIB1/ACTR/TRAM-1) is required for normal growth, puberty, female reproductive function, and mammary gland development, *PNAS* **97**, 6379–6384.

Yoshida, M., Kijima, M., Akita, M., Beppu, T. (1990) Potent and specific inhibition of mammalian histone deacetylase both in vivo

and in vitro by trichostatin A, *J. Biol. Chem.* **265**, 17174–17179.

Zardo, G., Tiirikainen, M.I., Hong, C., Misra, A., Feuerstein, B.G., Volik, S., Collins, C.C., Lamborn, K.R., Bollen, A., Pinkel, D., Albertson, D.G., Costello, J.F. (2002) Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors, *Nat. Genet.* **32**, 453–458.

Zhang, Y., Reinberg, D. (2001) Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails, *Genes Dev.* **15**, 2343–2360.

Epithelial Morphogenesis

Ronit Wilk[1,2], *Amanda T. Pickup*[1], *and Howard D. Lipshitz*[1,2]
[1] *Program in Developmental Biology, Research Institute, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada*
[2] *Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada*

1 Epithelial Architecture 279
1.1 Cell–cell Junctions 279
1.2 Cell–ECM Contacts 280
1.3 Cell Polarity 282

2 Cellular Basis of Epithelial Morphogenesis 284
2.1 Cell Adhesions and Signal Transduction 284
2.2 The Actin-based Cytoskeleton 284

3 Regulation of Epithelial Morphogenesis: Lessons from *Drosophila* 287
3.1 Epithelial Sheet Movement and Fusion: Dorsal Closure 287
3.2 Branching Morphogenesis and Epithelial Tube Formation: the Tracheal system 290
3.3 Epithelial Patterning and Morphogenesis at Single-cell Resolution: the Retina 292

4 Concluding Remarks 296

Bibliography 297
Books and Reviews 297
Primary Literature 298

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Keywords

Drosophila melanogaster
The vinegar fly (commonly referred to by the misnomer, fruit fly) first systematically studied in 1910 by Morgan and coworkers and arguably still the best-understood multicellular animal from both a genetic and molecular biological perspective.

Epithelium
A layer of polarized cells that are held together by cell-to-cell adhesions and are bounded basally by extracellular matrix.

Extracellular Matrix (ECM)
A network of macromolecules that resides outside of cells and is comprised largely of proteins and polysaccharides that are organized into a meshwork. The ECM interacts with nearby cells via specific transmembrane proteins present in the plasma membrane of these cells.

Metazoa
Animals whose bodies consist of many cells and that pass through embryonic stages of development. Distinct from protozoa, which are unicellular. Sponges, although multicellular, are sometimes classified separately as parazoa, since they contain no mesoderm and have no true organ systems.

Morphogenesis
During development, regeneration, or wound healing, the establishment of form and structure from simple, relatively undifferentiated cells, through cell movement, shape change, migration, differentiation, and programmed cell death.

An epithelium is a layer of polarized cells that are tightly bound together by a variety of junctional complexes and are connected to the extracellular matrix (ECM). Transmembrane proteins link the junctional complexes and the basal ECM to the cytoplasmic signal transduction machinery and the cytoskeleton of the epithelial cells. During development, epithelia are converted from simple two-dimensional sheets into complex, three-dimensional configurations in order to form tissues and organs. These morphogenetic changes are regulated in time and in space and are accompanied by differentiation of the cells within the epithelia into distinct identities. Epithelial morphogenesis is orchestrated by a variety of signaling molecules and transcription factors that coordinate cell shape changes, rearrangement, and migration. Much of the initial understanding of epithelial organization was derived from studies carried out on cultured vertebrate cells or tissues. More recently, genetic analyses in *Drosophila melanogaster* have complemented cell biological approaches and have provided insights into the spatial and temporal regulation of epithelial morphogenesis during development.

1 Epithelial Architecture

Epithelial cells have well-organized connections with neighboring cells and with the ECM. Cell adhesion proteins play a role in the maintenance of cell–cell and cell–ECM contacts as well as in transduction of extracellular signals into and out of epithelial cells. Cell adhesion complexes are made up of three classes of proteins – cell adhesion receptors, ECM proteins, and the cytoplasmic plaque proteins – many of which are conserved in all metazoa. The cell adhesion receptors are usually transmembrane glycoproteins that determine the specificity of cell–cell or cell–ECM interactions by interacting with proteins of the ECM or with other adhesion receptors in nearby cells. Several superfamilies of proteins fall into this category, including cadherins and integrins. ECM proteins are large glycoproteins including collagens, laminins, and fibronectins. These assemble into complex fibrils or other arrays that interact with cell adhesion receptors and thus with the cell surface. The cytoplasmic plaque proteins, such as catenins, are a link between adhesion receptors and the cytoplasmic signal transduction machinery as well as the cytoskeleton. The plaque proteins thus function to transduce signals from the adhesion complexes into the cell.

1.1 Cell–cell Junctions

While many of the molecular players are conserved, the ultrastructure of vertebrate and invertebrate junctional complexes and their location relative to the apico-basal cellular axis differ to some extent. Invertebrate epithelia have two types of junctions – adherens junctions and septate junctions – while three types of vertebrate junctional complexes have been described – tight junctions, desmosomes, and adherens junctions (Fig. 1).

Both vertebrate and invertebrate adherens junctions have cadherins as their major adhesion receptors. In epithelia, the adhesion receptor, E-cadherin, binds β-catenin. In turn, β-catenin binds α-catenin and α-catenin associates with the F-actin cytoskeleton (Fig. 2). In *Drosophila*, the most studied adherens junction is the zonula adherens, a beltlike structure localized at the apico-lateral region of epithelial cells, which contains DE-cadherin, the *Drosophila* homolog of E-cadherin, as well as α- and β-catenin. Adherens junctions are particularly prevalent in epithelia that experience strong contractile forces at their apices, as during invagination of sheets during normal development or wound healing.

Desmosomes are comprised of desmogleins and desmocollins, members of the cadherin superfamily that are linked to the cytoplasm by desmoplakins, plakoglobin (γ-catenin), and other cytoplasmic plaque proteins. Plakoglobin is a close homolog of β-catenin and fulfils some of the same functions. Desmosomes are linked to the intermediate filaments of the cytoskeleton (Fig. 1), creating a network that strengthens the cell. The desmosomal intermediate filament complex is particularly important in tissues that have to absorb high levels of mechanical stress. No intermediate filaments or desmosomes have been found in *Drosophila* or other invertebrates.

Tight or "occluding" junctions are found at the boundary between apical and basolateral cell surfaces (Fig. 1). They function to form a permeability barrier between the apical and basal sides of an epithelium, thus creating separate physiological compartments within a tissue or organ.

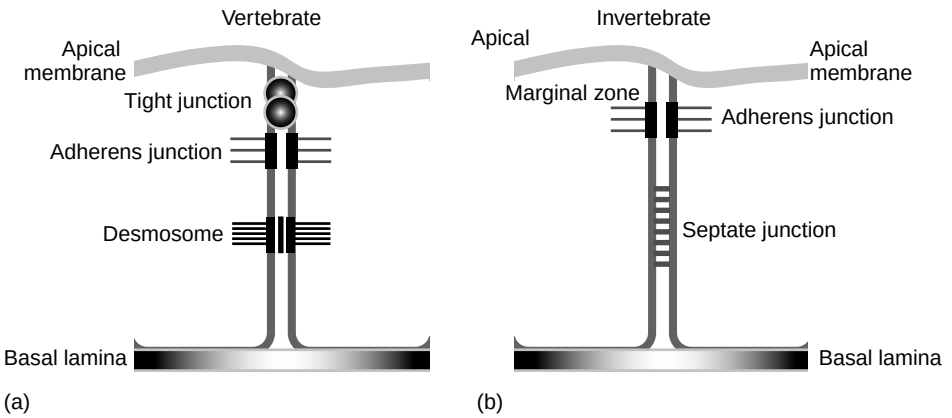

Fig. 1 Comparison of vertebrate and invertebrate cell junctions. (a) Vertebrates have three types of morphologically distinguishable junctions (from apical to basal: tight junctions, adherens junctions, and desmosomes) while *Drosophila* and other invertebrates (b) have two morphologically distinguishable junctions (from apical to basal: adherens junctions and septate junctions). At the molecular level, the adherens junctions of vertebrates and invertebrates are homologous. Homologs of vertebrate tight junction proteins are found in the *Drosophila* marginal zone. In invertebrates, septate junctions fulfil the role of tight junctions in controlling paracellular transport and gating functions. Desmosomes are unique to vertebrates both morphologically and in terms of molecular components. The fibers shown attached to the adherens junctions are F-actin while the fibers attached to the desmosomes are intermediate filaments. Figure based on Muller, H.A. (2000) Genetic control of epithelial cell polarity: lessons from Drosophila, *Dev. Dyn.* **218**, 52–67.

A major component of tight junctions is occludin, a transmembrane protein that contacts cytoplasmic plaque proteins such as zonula occludens protein-1 (ZO-1) and ZO-2, and the ECM. No convincing *Drosophila* occludin homologs have been found. However, *Drosophila* has two genes that encode proteins similar to mammalian claudin – another component of the tight junctions – as well as a homolog of mammalian ZO-1, known as *Polychaetoid*. Furthermore, the homologs of several tight junction proteins (e.g. Polychaetoid, Bazooka) localize to the marginal zone in *Drosophila*, suggesting that the invertebrate marginal zone may share some functions with vertebrate tight junctions. Functionally, but not locationally, tight junctions are similar to *Drosophila* septate junctions in that both control paracellular transport or gating functions.

1.2 Cell–ECM Contacts

Most cell–ECM adhesions share two features: they are mediated by integrins and they interact with the actin cytoskeleton inside the cell. The extracellular ligands that anchor these adhesions include fibronectin, vibronectin, and various collagens. The cells are attached to the collagen-rich interstitial matrix or to the basement membrane, a specialized structure that covers the basal surface of all epithelia. The basement membrane is formed from two layers: the basal lamina and the reticular lamina. The former is closer to the

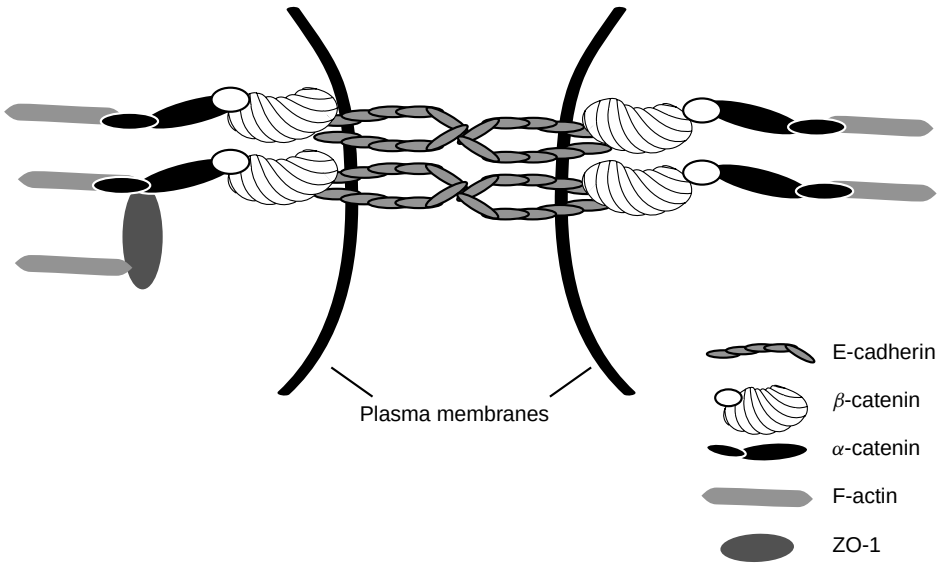

Fig. 2 A simplified adherens junction. The extracellular domain of E-cadherin dimerizes and interacts in a calcium-dependent manner with E-cadherin on neighboring cells. The distalmost segment of the cytoplasmic tail of E-cadherin interacts with β-catenin. α-catenin binds directly to F-actin and to β-catenin, thus linking the transmembrane adhesion receptor to the cytoskeleton. Additionally, α-catenin can bind to proteins such as ZO-1, which are components of tight junctions in vertebrates but are localized to the marginal zone in invertebrates such as *Drosophila*. Figure based on Vasioukhin, V., Fuchs, E. (2001) Actin dynamics and cell–cell adhesion in epithelia, *Curr. Opin. Cell Biol.* **13**, 76–84.

cell and is comprised of glycoproteins such as collagen type IV, laminin, and fibronectin, as well as proteoglycans. The reticular lamina is produced by fibroblasts and contains fibrillar collagens. The molecules that compose the basal lamina are well conserved among metazoa. *Drosophila* has two type IV collagens and four laminins as well as homologs of the glycoprotein nidogen/entactin and the proteoglycan perlecan.

Integrins are heterodimeric transmembrane proteins, composed of α- and β-subunits, whose extracellular domains interact with the ECM while their cytoplasmic tails interact with plaque proteins. The adhesiveness of the extracellular domains of integrin heterodimers is controlled by binding events in the cytoplasmic tails that trigger conformational changes across the plasma membrane (signaling "from the inside to the outside"). Reciprocally, ligand binding to the extracellular domains can trigger an intracellular signal (signaling "from the outside to the inside"). $\alpha\beta$ heterodimeric integrin receptors are found from sponges to mammals. *Drosophila* has two β- and five α-integrin subunits, including αPS1 and αPS2.

Focal adhesions are flat and elongated structures, usually found in the peripheral cytoplasm of cells, which are bound to integrin tails on the one hand and actin filaments on the other (Fig. 3). Focal adhesions comprise a multiprotein plaque that contains vinculin, talin, and paxillin,

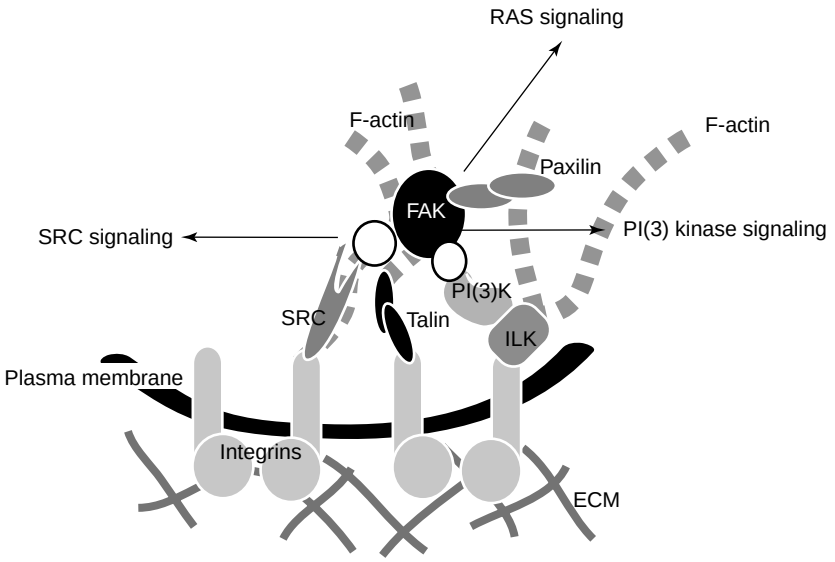

Fig. 3 Signaling and cytoskeletal complexes in focal adhesions. Integrins interact with the extracellular matrix as well as with signaling proteins in the cytoplasm. These include ILK, talin, and SRC – which interact directly with the cytoplasmic tail of integrin – as well as FAK, PI(3) kinase, and paxillin, which interact indirectly. The signaling complexes interact with and regulate the F actin–based cytoskeleton. Figure based on Miranti, C.K., Brugge, J.S. (2002) Sensing the environment: a historical perspective on integrin signal transduction, *Nat. Cell Biol.* **4**, 83–90.

among other proteins. Integrin-associated molecules, such as focal adhesion kinase (FAK), integrin-linked kinase (ILK), or paxillin may bind indirectly to the cytoskeleton through other components of the submembrane plaque, thus regulating the formation or function of focal adhesions. The constituents of focal adhesions are conserved in *Drosophila*, which has genes encoding talin, vinculin, tensin, paxillin, FAK, and ILK.

Fibrillar adhesions are found further from the plasma membrane than focal adhesions. They are elongated dot-like structures that are associated with ECM fibrils, and are typically composed of fibronectin (ECM component), $\alpha_5\beta_1$ integrin (transmembrane receptor), and tensin (cytoplasmic component). In the presence of fibronectin, $\alpha_5\beta_1$ integrin translocates out of the focal adhesion, forming a fibrillar adhesion that promotes fibronectin fibrillogenesis. *Drosophila* does not have fibrillar collagens, any long fibrinogen or fibronectin, and thus lacks fibrillar adhesions.

1.3 Cell Polarity

Epithelial cells have a well-defined apical–basal polarity (Fig. 1). Different domains along the apical–basal axis of an epithelial cell have distinct ECM, membrane, and submembrane cytoskeletal components. Many of these molecules (e.g. cadherins, catenins, and integrins) were first identified in biochemical or

cell biological studies in vertebrates and were subsequently shown to be conserved in metazoa. Along with their role in forming cell–cell and cell–ECM junctions, these proteins are crucial for cell polarity. For example, mutations in *Drosophila* genes such as *armadillo* (β-catenin), *shotgun* (DE-cadherin), and *karst* (β_{Heavy}- or β_{H}-spectrin) lead to defects in the establishment or maintenance of the adherens junctions and of epithelial polarity. Cell–ECM attachments are also important in defining the structural and functional apical–basal polarity of epithelial cells. For example, the E3 and E8 domains of laminin-1 can polarize mammary gland luminal epithelial cells in an α_6 integrin-dependent manner and reconstituted basement membrane can direct cultured mouse pre-Sertoli cells to develop into a polarized epithelium.

Several proteins that play a role in epithelial polarity were first identified on the basis of epithelial defects in *Drosophila* mutants. For example, Crumbs (CRB) is an apically localized transmembrane protein that regulates epithelial polarity in *Drosophila*. Recruitment of Stardust (SDT) and Discs lost (DLT, a PDZ domain-containing protein) – two proteins that are essential for epithelial polarity – to the apical membrane of epithelial cells depends on interactions with the cytoplasmic tail of CRB while, in turn, apical localization of CRB depends on the SDT and DLT proteins.

PDZ domain-containing proteins, including the LAP (Leucine-rich- and PDZ domain-containing) subfamily, are involved in the determination and maintenance of apical–basal epithelial polarity. LAP proteins are thought to form a barrier that keeps apical determinants in the apical domain of epithelial cells. In addition, LAP proteins create and maintain polarity by transporting and anchoring binding partners to particular membrane domains. For example, the human homolog of *Drosophila discs large* (DLG), hDLG, binds kinesin-like motors that interact with the microtubule cytoskeleton to transport and localize binding partners to the membrane. Mutations in the *Caenorhabditis elegans let-413* LAP protein-encoding gene lead to abnormal epithelia with defective adherens junctions, which are mislocalized to more basolateral positions, and a disorganized actin cytoskeleton. Absence of the *Drosophila* LAP proteins Scribble and DLG leads to defects in apical–basal epithelial polarity and defective morphogenesis.

Members of the RHO family of small GTPases (CDC42, RAC1, and RHO), molecular switches that regulate cell signaling pathways and the organization of the actin cytoskeleton, have been implicated in maintaining polarity of epithelial cells. CDC42's role in epithelial apical–basal polarity is in the regulation of polarized transport of secretory vesicles as well as in the stabilization of adherens junctions. RAC1 is involved in the proper formation of tight junctions in epithelial cells. RHO plays an important role in adherens junction formation, intracellular targeting of proteins, and phosphorylation of catenins.

Maintenance of epithelial polarity is crucial for many aspects of epithelial function. For example, transmembrane adhesion molecules and receptors in the basolateral domain facilitate communication with other cells and with the ECM. The apical domain, in contrast, is not in permanent contact with other cells and, typically, faces either the body cavity or the lumen in the case of epithelial tubes. Tight junctions separate the apical and basal domains and are responsible

for lateral cell–cell adhesions as well as for creating a barrier that restricts diffusion of fluids and molecules through the epithelium. Thus, the different sides of an epithelium implement distinct cellular functions while interacting with and maintaining different extracellular environments. If apical–basal polarity fails, cell–cell and cell–ECM connections are affected and epithelial function and/or integrity is therefore compromised.

Interestingly, there is a close link between apical–basal polarity and growth control. Mutations in genes such as *Drosophila scribble, discs large*, and *lethal giant larvae*, which affect epithelial polarity as described above, have also been found to cause uncontrolled cell proliferation. Defects in apical–basal polarity probably cause these cells to lose cell–cell or cell–ECM connections, which in turn allows them to divide.

2 Cellular Basis of Epithelial Morphogenesis

2.1 Cell Adhesions and Signal Transduction

The epithelial shape changes and movements that occur during morphogenesis – as well as the functions of mature, differentiated epithelia – need to be coordinated in time and space. As described in Sect. 1 above, cell–cell and cell–ECM junctions include transmembrane receptors that are linked to the intracellular signaling apparatus. These junctions participate in signal transduction events that regulate cell adhesion, cell motility, and cell shape and thus are crucial for epithelial morphogenesis.

Adherens junctions (described in Sect. 1.1; see Fig. 1 and 2) play a key role in signal transduction within epithelia. The adherens junction component, β-catenin, has roles in cell adhesion, signal transduction, and transcriptional activation. For example, when the WG/WNT signaling pathway is inactive, free β-catenin is phosphorylated by glycogen synthase kinase (GSK) and axin, leading to ubiquitination and degradation of β-catenin. When WG/WNT binds to its receptor, Frizzled (FZ), the Dishevelled (DSH) protein inhibits phosphorylation of β-catenin by GSK. Unphosphorylated β-catenin is stable and accumulates in the nucleus where, together with TCF, it transactivates target genes such as *cyclin D1* and *myc*.

Integrins mediate signal transduction into epithelial cells via focal adhesions, giant membrane–associated protein complexes that contain dozens of stably or transiently associated proteins (see Sect. 1.3 above). Ligand binding, integrin aggregation, and actin polymerization affect the recruitment of different proteins to focal adhesions. These adhesions are linked to webs of intracellular proteins, simultaneously affecting several signal transduction pathways (Fig. 3). For example, the EGFR (epidermal growth factor receptor), PDGFR (platelet-derived GFR), and VEGFR (vascular endothelial GFR) can be activated by focal adhesions, specifically by integrins, thus triggering a RAS-RAF-MAP kinase (MAPK) cytoplasmic signal transduction cascade.

2.2 The Actin-based Cytoskeleton

The assembly and disassembly of filamentous actin (F-actin), a major constituent of the cytoskeleton of all eukaryotic cells, is highly regulated (Fig. 4). For example, proteins function to sequester actin monomers to prevent polymerization

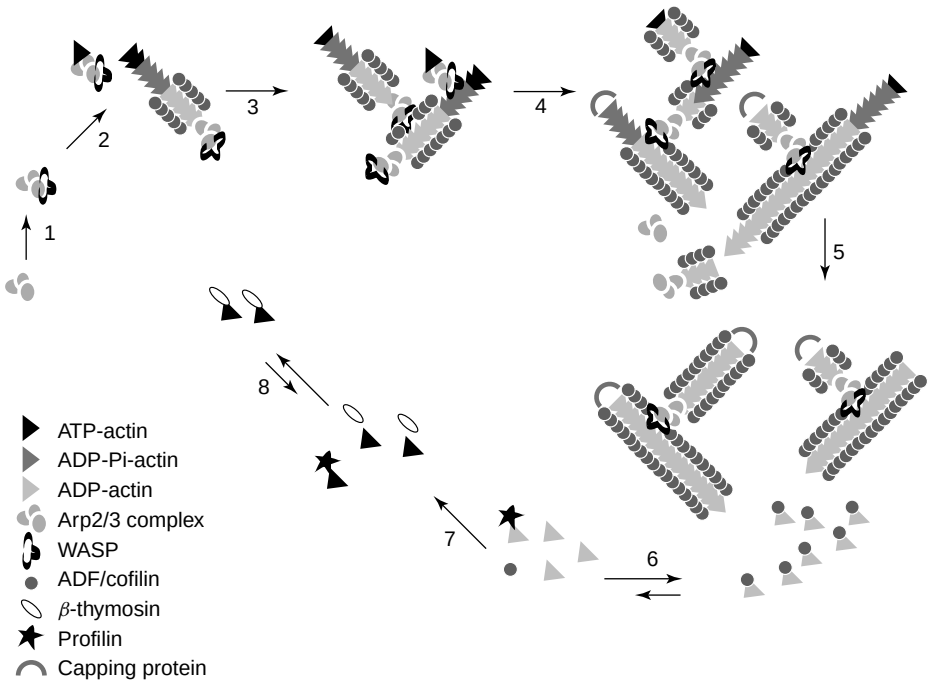

Fig. 4 Actin dynamics. The actin network in the leading edge of cells can promote formation of lamellipodia (shown here) or filopodia via regulated polymerization and depolymerization of F-actin by actin-binding proteins. (1) The Arp2/3 complex is activated upon binding to WASP, which is activated by the small GTPase, CDC42. (2) The active Arp2/3-WASP complex nucleates actin filaments at the free (pointed) end or (3) binds to the side of the filament and nucleates a branch. Reduced dissociation rates of ADP-P_i from actin result in growth of filaments. (4) Capping of the barbed ends of the actin filaments prevents further elongation. (5) ADF/cofilin enhances depolymerization of ADP-actin from the free filament end. (6) The ADF/cofilin/ADP-actin complex that dissociates from the filament is in equilibrium with ADF/cofilin and ADP-actin monomers. (7) Nucleotide exchange on actin monomers is a slow process that is enhanced by profilin and inhibited by ADF/cofilin. (8) ATP-actin monomers are sequestered by β-thymosins to prevent spontaneous nucleation, while simultaneously providing a pool of ATP-actin for filament assembly. Figure redrawn from Chen, H., Bernstein, B.W., Bamburg, J.R. (2000) Regulating actin-filament dynamics *in vivo*, *Trends Biochem. Sci.* **25**, 19–23, with permission from Dr. J. Bamburg.

(β-thymosins), to enhance nucleotide exchange (profilin), to split F-actin to generate more ends (gelsolin, ADF/cofilin also known as the AC family of proteins), or to create a cap at the end of the filament, resulting either in filament growth (Arp2/3 complex) or in enhanced subunit dissociation (gelsolin). Signals that regulate actin dynamics are transmitted through receptor tyrosine kinases, the RHO-family GTPases, and G protein–coupled receptors. These molecules receive and transmit signals that lead the WASP (Wiskott-Aldrich Syndrome protein)-Arp2/3 complex to nucleate F-actin assembly. Both the

molecular components and the regulatory processes of actin dynamics are conserved between *Drosophila* and mammals.

Dynamic changes in F-actin control cell–cell adhesion in epithelia. For example, as described above (Sect. 1.1), in epithelial cells, the adherens junction proteins E-cadherin and β-catenin are linked to the actin cytoskeleton via α-catenin. If the actin cytoskeleton is disrupted or catenin function is lost, intercellular adhesions are not formed or are not maintained. Moreover, the small GTPases, RHO and RAC, function as F-actin organizers and are required for adherens junction formation.

Many morphogenetic processes during development of diverse organisms utilize actin-dependent cytoskeletal extensions to sense the extracellular environment and to transmit signals over long distances. For example, neurons extend filopodia for axon-guidance while, during embryonic dorsal closure, the leading edge cells of the epidermis extend filopodia that contact the extra-embryonic amnioserosa as well as the leading edge cells on the opposite side of the embryo (see Sect. 3.1 below for details).

In cultured epithelial cells, prior to cell–cell contact, actin forms marginal dynamic bundles that undergo rearward flow in a myosin II-dependent manner. Actin polymerization at the free, leading edge of the cell drives lamellipodial protrusions, resulting in the establishment of contacts with neighboring cells. Once the cells contact each other, the rearward actin flow ceases and adherens junction components, such as E-cadherin, accumulate at the site of contact. Actin arcs then form at the edges of the region of cell–cell contact. Subsequently, during the lateral expansion of the initial contact, old marginal actin bundles disassemble and new ones are formed, thus strengthening cell–cell adhesions. If cell–cell contacts are broken *in vitro*, then F-actin is reduced in adherens junctions, and proteins such as α- and β-catenin are lost from the junctions. Thus, establishment and maintenance of epithelial cell–cell contacts in culture requires the coordinated function of actin-filament bundles, myosin II, microtubules, and cell adhesion proteins.

As discussed in Sect. 2.1, proper cytoskeletal organization is necessary for signal transduction into epithelial cells as well as for ECM-to-cell communication. For example, integrin β-subunits regulate the structure of the actin cytoskeleton as well as the activities of the small GTPases, RHO, and RAC. Integrin attachment to the ECM is sufficient for the induction of lamellipodia, filopodia, and focal adhesions through RHO activation. Overexpression of β_3 integrin in hamster ovary cells enhances RHO activity and stress fiber formation (stress fibers are F-actin bundles in which the fibers are loosely cross-linked), whereas overexpression of $\beta 1$ integrin increases RAC and Jun N-terminal kinase (JNK) activity and formation of lamellipodia in which the F-actin fibers form a gel-like network. These results suggest that distinct integrins regulate different processes in epithelial cells by differentially regulating molecules such as RHO and RAC.

Cells also use integrin/actin complexes as mechanosensors to probe the rigidity of the extracellular environment, thus modulating cell–ECM adhesion and triggering signaling events. For example, a local mechanical force induces clustering of integrin molecules, thus leading to the formation of a focal complex at the site of plasma membrane deformation (focal complexes are $\sim$1 μm dotlike adhesions found at the cell periphery,

which are enriched in actin filaments and initiate specific signals). Focal complexes can evolve into stable focal adhesions; both internal and external forces enhance such a conversion. The molecular regulation of cell–ECM contacts and mechanosensation mechanisms are poorly understood. Insights will derive from cell culture on three-dimensional ECM, rather than the more traditional two-dimensional culture methods. *In vivo* studies using high-resolution microscopy in live organisms such as *Drosophila* will also shed light on the molecular regulation of cell–ECM interactions and the transduction of external cues into changes in cytoskeletal structure and function.

Among the cytoskeletal proteins that play a role in epithelial structure and function, spectrin plays a particularly crucial role. The spectrin protein, composed of two α- and two β-subunits, cross-links F-actin. Different spectrin isoforms are found associated with the basolateral versus the apical region of epithelial cells. For example, *Drosophila* has three spectrin subunits, α, β, and β_H. Each of the two β-subunits heterodimerizes with α-spectrin, and the $\alpha\beta$ and $\alpha\beta_H$ heterodimers show nonoverlapping distributions in epithelial cells; $\alpha\beta$-spectrin localizes to the basolateral membrane whereas $\alpha\beta_{\mathrm{H}}$-spectrin is apically enriched in the marginal zone. Mutational analyses have shown that, during *Drosophila* development, $\alpha\beta_{\mathrm{H}}$-spectrin is required for the differentiation of the apical membrane, and that β-spectrin functions independent of α-spectrin in differentiation of the basolateral membrane domain. When β_{H}-spectrin is absent or nonfunctional, in *karst* mutant embryos, the adherens junctions of the ovarian follicle epithelial cells are fragmented, and the cells fail to constrict apically, suggesting a requirement for β_{H}-spectrin in regulation and maintenance of the adherens junctions.

3 Regulation of Epithelial Morphogenesis: Lessons from *Drosophila*

During epithelial morphogenesis, the diverse cellular processes described above have to be coordinated. This is accomplished by exploiting the ability of epithelial cells to respond to extracellular cues and to transduce such signals into changes in cell shape and behavior. Here we consider specific morphogenetic processes in *Drosophila* and describe how genetic and molecular analyses have led to insights into how epithelial morphogenesis is coordinated in the context of a complex, developing organism.

3.1 Epithelial Sheet Movement and Fusion: Dorsal Closure

Following germ-band retraction of the *Drosophila* embryo, a hole is left on the dorsal side of the embryo, where epidermis is absent (Fig. 5). This hole is covered by the amnioserosa, an extraembryonic epithelium. Dorsal closure involves the dorsalward migration of lateral epidermis on both sides of the embryo such that the leading edges of the left and right epidermal epithelia meet and fuse at the dorsal midline, completely covering the amnioserosa and sealing the epidermis around the embryo (Fig. 5). There is little or no cell division or cell rearrangement in the epidermal epithelium during dorsal closure. Almost all of the morphogenetic process is thus accomplished through coordinated cell shape changes within the epithelium; cells lengthen along

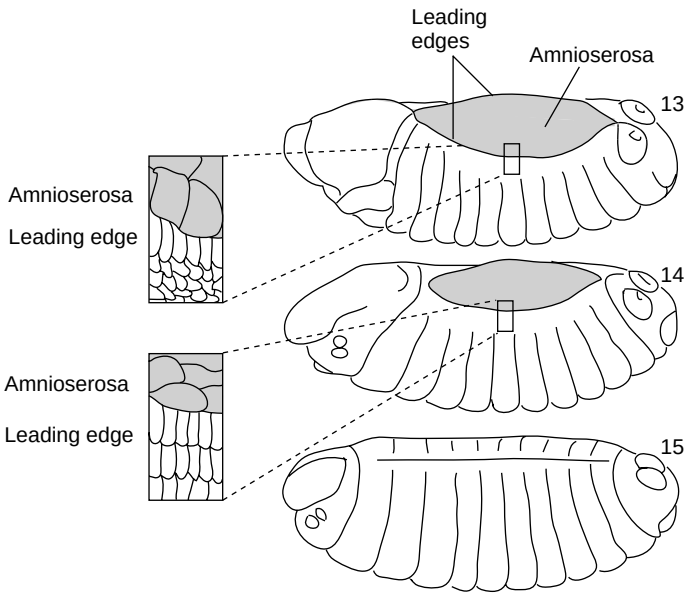

Fig. 5 Dorsal closure in *Drosophila*. Three stages of dorsal closure are shown: (top) initiation, (middle) spreading, and (bottom) suturing. Initiation is first visible in stage 13 embryos (9 to 10 h after fertilization, AF) as the rim of the dorsalmost epidermal cells, the leading edge cells, undergoes dorsoventral elongation. Subsequently, the more lateral epidermal cells also undergo progressive elongation, thus promoting spreading of the lateral epidermal epithelial sheets dorsally. Concomitant with this spreading, the amnioserosa cells (gray) contract across the dorsoventral axis of the embryo. This spreading phase lasts about 2 h, during embryonic stages 13 to 15. When the leading edges of the lateral epidermal sheets reach the dorsal midline, cells have elongated three- to fourfold. During suturing, the sheets meet at the dorsal midline and fuse in register, thus, completely enclosing the embryo within the epidermis. Figure based on Hartenstein, V. (1993) Atlas of *Drosophila* Development, Cold Spring Harbor Laboratory Press, Cold Spring Harbor and Noselli, S. (1998) JNK signaling and morphogenesis in *Drosophila*, *Trends Genet.* **14**, 33–38.

the dorsoventral and shorten along the anteroposterior axis of the embryo. Reciprocally, the amnioserosal epithelial cells shorten along the dorsoventral axis as the epidermal epithelium extends dorsally. Dorsal closure thus exemplifies how coordinated cell shape changes throughout epithelial sheets can drive morphogenesis.

The forces that drive the process of dorsal closure come from both the amnioserosa and the lateral epidermis. This was demonstrated by performing laser ablations of cells in the amnioserosa and/or of cells in the lateral epidermis; neither of these types of ablations inhibits dorsal closure but a combination of both does. A crucial component of the force-generation system is the F-actin/myosin-based "purse string" that is assembled at the apico-dorsal region of the epidermal

leading edge cells adjacent to the amnioserosa. Contraction of this purse string drives dorsalward movement of the epidermal sheet and eventual fusion at the dorsal midline. In myosin II (*zipper*) mutants, dorsal closure fails because the purse string is unable to contract.

Pivotal to the reorganization and contraction of the F actin–based cytoskeleton are RHO, RAC, and CDC42 (discussed in Sects. 1.3 and 2.2 above). Mutants or transgenic flies that express dominant negative versions of these small GTPases show defects in dorsal closure and in F-actin distribution.

Cell shape changes in the leading edge and more lateral epidermal cells are crucial for dorsal closure. These shape changes are coordinated through activity of the BMP/TGFβ-family ligand, DPP, WG/WNT, and JNK signaling. These signals direct the more lateral epidermal cells to change their shape and extend dorsally, in an F actin–dependent process.

Essential for dorsal closure is definition of the boundary between the epidermal leading edge cells and the amnioserosa since the purse string is assembled at this boundary. Recent data suggest that modulation of JNK signaling plays a crucial role in establishing or implementing this boundary. Initially, JNK signaling is present in both the amnioserosa and the epidermal leading edge cells. Prior to dorsal closure, however, JNK signaling is downregulated in the amnioserosa leading to a low/high JNK signaling boundary that coincides with where the purse string assembles. If JNK signaling is not downregulated in the amnioserosa, a high/high JNK signaling situation persists, no purse string is assembled and dorsal closure fails. Reciprocally, if JNK signaling is downregulated in both the amnioserosa and the leading edge cells, then a low/low situation pertains and, again, the purse string is disrupted and dorsal closure fails. Mutant situations in which the high/high situation is converted to a low/high one, lead to rescue of both the purse string and dorsal closure.

The actual zipping up or suturing at the dorsal midline (Fig. 5) is achieved by actin-based mechanisms. During dorsal closure, actin and nonmuscle myosin II accumulate at the leading edge where long and active filopodia form. Leading edge filopodia formation is dependent on the activity of small GTPases and JNK signaling. In order to stop epidermal sheet movement and allow in-register left–right leading edge fusion, filopodia must discriminate between the different segments of the embryo and JNK signaling must be downregulated. During suturing of the epithelial sheets at the dorsal midline, the filopodial contacts lead to the production of precursors of mature adherens junctions.

The low/high JNK signaling boundary may be required to trigger formation of filopodia by the leading edge cells as described above. Reciprocally, it has been proposed that contact of high JNK signaling cells (as when the filopodia from the left and right leading edge cells meet) serves as a "stop" signal that prevents further epithelial sheet movement. In mutants in which JNK signaling is not downregulated in the amnioserosa, a premature stop signal may be received and thus epithelial sheet movement never initiates.

It should be noted that actin-binding molecules such as profilin (see Sect. 2.2 above), encoded by *chickadee*, have been implicated in dorsal closure. Mutant *chickadee* embryos show disrupted actin filopodia similar to those found in embryos lacking JNK activity. This suggests that in the

leading edge cells, the JNK cascade drives the synthesis and/or activation of profilin, which is involved in the production of the filopodia required for dorsal closure.

There are extensive similarities between the epithelial sheet movements that occur during dorsal closure and those that occur during *C. elegans* ventral enclosure and mammalian wound healing. The mechanisms that have been identified as coordinating and directing epithelial sheet movement during dorsal closure are also likely to underlie epithelial sheet movement in these systems. For example, time lapse analysis of live *C. elegans* embryos during ventral enclosure has shown that "leader" cells have active filopodia, that α-catenin accumulates where opposing filopodia make contact, and that this is followed by the formation of mature adherens junctions between the cells that achieve contact. Similar filopodia and F actin–based purse strings have been seen in epithelial cells in culture and during wound healing.

3.2 Branching Morphogenesis and Epithelial Tube Formation: the Tracheal system

During morphogenesis of the *Drosophila* embryonic tracheal system, clusters of epithelial cells invaginate, branch, and migrate in the absence of cell division and cell death. The tracheal system develops from 20 clusters of epidermal cells, 10 each on the left and right surfaces of the embryo, each of which contain ~80 cells (Fig. 6). These clusters invaginate internally and then branch in a stereotypical manner, each giving rise

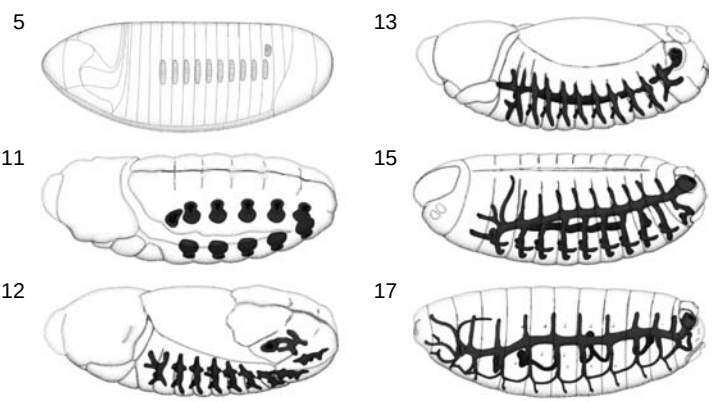

Fig. 6 Tracheal branching morphogenesis in *Drosophila*. The tracheal system develops from 20 placodes, 10 on the surface of each side of the embryo, that are conferred tracheal fate by stage 5 of embryogenesis (2 to 3 h AF). These clusters of cells divide twice to give rise to about 80 cells per placode. Invagination begins at stage 11 (5 to 7 h AF), followed by stereotypic primary branching at stage 12 (7 to 9 h AF). Fusion of specific branches occurs at stages 14 and 15 (10 to 13 h AF), giving a complex tubular system. Secondary and tertiary branching takes place respectively at stages 15 (11 to 13 h AF) and 17 (16 to 20 h AF). Figure modified from Hartenstein, V. (1993) Atlas of *Drosophila* Development, Cold Spring Harbor Laboratory Press, Cold Spring Harbor with permission from the publisher.

to five primary tracheal branches at stage 12 of embryogenesis. Specific tracheal cells elongate and then fuse to create continuous tubes whose lumen allows the passage of oxygen to all the tissues of the organism via secondary and tertiary ("terminal") branches that form a complex tubular network by embryonic stage 17. The apical side of the tracheal epithelium faces the lumen while the basal side is bounded by a collagen IV- and laminin-rich basal lamina. A detailed description of the cellular and molecular mechanisms of tracheal development is available.

Tracheal cell fate is conferred by the basic-HLH transcription factor, Trachealess (TRH), and its binding-partner Tango. The POU-domain protein encoded by *drifter* interacts with TRH to induce tracheal fates. The tracheal clusters are then subpatterned through FGF and DPP signaling; dorsal and lateral branches depend on DPP while ventral branches are determined by FGF.

Primary branching morphogenesis is controlled by one of the two *Drosophila* FGF receptor homologs, breathless (BTL) and its ligand, branchless (BNL). Just before a trachea branches, patches of BNL expression occur in specific clusters of epidermal cells near the tracheal sac. Binding of BNL to BTL induces a RAS-RAF-MAPK signal transduction cascade, resulting in cytoskeletal changes that direct the formation of buds that will, in due course, become branches. Recent work utilizing *in vivo* imaging has shown that DPP and FGF regulate different cellular events in tracheal branching morphogenesis. FGF signaling is required for the formation of filopodia while DPP is important for branch outgrowth. The FGF pathway acts together with cell surface proteins that control cell–cell and cell–ECM interactions during tracheal branching. For example, expression of PS1 integrin (αPS1βPS) specifically in the middle part of the developing tracheal placodes, and PS2 integrin (αPS2βPS) in the visceral mesoderm, is necessary for visceral branch migration. FGF signaling also induces secondary branching genes, including *pointed*, which encodes an ETS domain transcription factor, and *sprouty*, which encodes an inhibitor of branching. Interestingly, the FGF and BMP/TGFβ pathways also function in branching morphogenesis during mammalian lung development.

Specific genes such as *escargot*, which encodes a zinc-finger transcription factor, control fusion of the tracheal tubes by regulating cell shape changes and DE-cadherin production. Before fusion takes place, BTL is activated in a group of "leading" cells at the tip of the branch. The Notch–Delta pathway functions via lateral inhibition to distinguish the "fusion" cells from "nonfusion" cells within the leading cell group. In fusion cells, RHOA and the plakin, Short stop, are required for DE cadherin–dependent apical localization of the cytoskeletal structures that participate in the fusion event.

Terminal branches are long, thin cytoplasmic extensions, that form at the end of embryogenesis and later ramify nonstereotypically in response to low oxygen levels. Cell fate is determined by Notch (N), which is expressed in terminal cells and inhibits other cells from adopting a terminal fate. Terminal branching is guided by levels of oxygen, and by the BNL pathway together with Blistered (also known as Pruned), the serum-response factor homolog in *Drosophila*. Terminal extensions resemble F actin–based filopodia observed in other tissues during morphogenesis and serve to sense the environment as well as to transmit signals to nonneighboring cells.

As epithelia undergo morphogenesis, cell–cell and cell–ECM interactions must be modulated to allow shape changes and rearrangements relative to their neighbors without loss of epithelial integrity. Recent analyses have shown that the Hindsight (HNT) zinc-finger nuclear protein regulates tracheal epithelial integrity during the later stages of branching morphogenesis. In *hindsight* mutants, while primary and secondary branching occurs normally and fusion events initiate, the tracheal epithelium subsequently loses its integrity resulting in epithelial sacs. In *hindsight* mutants, the basal lamina that surrounds the tracheal tubes is abnormal, with an irregular distribution of both collagen IV and laminin, implicating the basal lamina in maintenance of epithelial integrity during branching morphogenesis.

3.3 Epithelial Patterning and Morphogenesis at Single-cell Resolution: the Retina

In *Drosophila*, the adult compound eye comprises an array of approximately 800 unit eyes or ommatidia (Fig. 7). The precise external hexagonal arrangement of these facets reflects a repetitive underlying cellular lattice. Each ommatidium is a stereotypical 19-cell assembly of eight

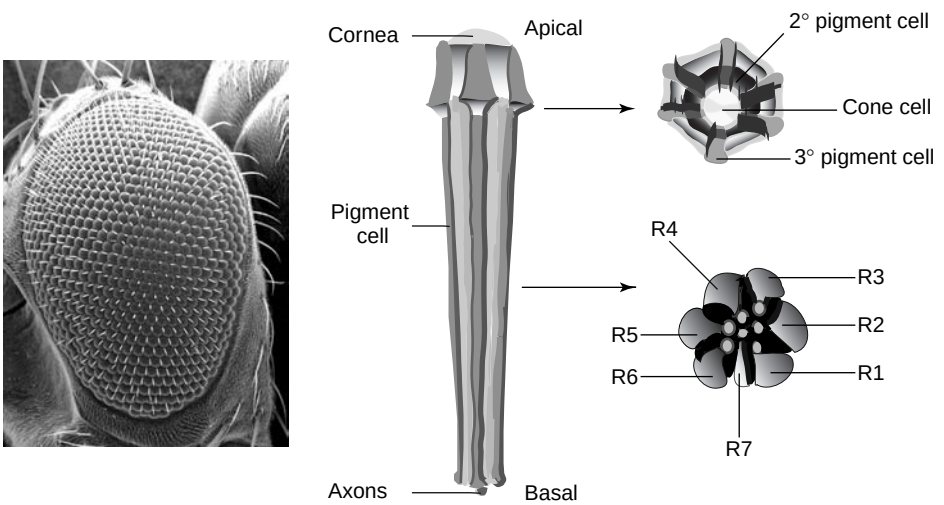

Fig. 7 The adult retina of *Drosophila*. The eye is comprised of about 800 small unit eyes or ommatidia packed in a hexagonal array, with mechanosensory bristles at alternate interommatidial vertices (a scanning electron micrograph is shown on the left). A single ommatidium is shown in longitudinal profile (middle) and cross-sectional views (right). Apically, the cornea is secreted by the cone cells, which extend dorsally and close over the top of the photoreceptor (R) cells during the pupal period. Primary and secondary pigment cells surround the cone and R-cells. At the base of the ommatidium, the cone and pigment cell feet meet to form a cone cell plate and fenestrated basement membrane through which the axons pass. The rhabdomeres of the outer R-cells (R1 to 6) are arranged in a stereotypical trapezoidal pattern around the inner R-cell (R7; the R8 cell is not visible in this plane of section) (lower right). Pigment cells are omitted from the lower cross-section for clarity. Figure based on Wolff, T., Ready, D. F. (1993) Pattern Formation in the *Drosophila* Retina, in: *The Development of Drosophila melanogaster*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 1277–1325.

photoreceptors (R-cells) and 11 accessory cells. R-cells are of three distinct classes: the R1–R6 class that surrounds two central photoreceptors; the distally located central R7 cell and, below it, the more proximally positioned central R8 cell. Each R-cell may be identified by virtue of its photosensitive rhabdomere, which is a dense microvillar extension that extends along the entire length of the ommatidial column. Above the R-cells are four cone cells surrounded by two primary pigment cells. Additional pigment cells form a mesh that encases the core of R-cells, cone cells, and primary pigment cells, and creates a honeycomb-like matrix. Small mechanosensory bristles project from alternate vertices of each ommatidium. At the base of the retina, the cone cells meet to form the cone cell plate; below this plate the feet of some of the pigment cells create the fenestrated membrane, which is a specialized basement membrane. Actin stress fibers anchor the pigment cell feet to specialized rings of ECM material.

The *Drosophila* eye is one of the most intensively studied of all developing organs, in part because its repetitive structure and external location facilitate identification of mutant phenotypes and in part because analyses of mutant versus wild-type tissue in genetic chimeras (''mosaics'') is possible at single-cell resolution. Thus, cell fate specification and morphogenesis within the retinal epithelium are understood in remarkable detail. The description of retinal morphogenesis presented here will, thus, be more detailed than that presented above for dorsal closure and tracheal branching, to emphasize how individual cell types behave.

Cells giving rise to the fly eye are first set aside in the embryo when an ectodermal invagination forms the specialized retinal epithelium known as the eye disc. By the third instar larval stage, the disc has become a two-layered sac composed of columnar epithelium on one side (the disc proper) and a squamous epithelium on the other (the peripodial epithelium). Beginning in the mid-third instar larva, a dorsoventral groove known as the morphogenetic furrow sweeps across the eye disc from the posterior to the anterior side leaving organized clusters of R-precursor cells in its wake. Initiation of the furrow at the posterior margin of the disc is under the concerted control of the secreted morphogens Hedgehog (HH), DPP, and WG/WNT and also requires the EGFR and N-receptor mediated signaling cascades. The developing clusters posterior to the furrow produce HH, which signals to more anterior cells and pushes the furrow across the disc. Morphogenetic events in the furrow are a prelude to subsequent R-cell precursor specification posterior to the furrow. Anterior to the furrow, the unpatterned cells are dividing randomly and they have large apical profiles. Within the furrow, cells stop dividing and synchronize in G_1 of the cell cycle. DPP is expressed in the furrow and it coordinates this cell cycle transition with the movement of the furrow across the disc. Changes in cell shape – a 15-fold reduction in the apical profile and significant cell shortening along the apical–basal axis – are accompanied by the basal migration of the nuclei and cell bodies. This cell shape transformation is caused by the constriction of apical actin cytoskeletal rings and it results in the concentration of apically localized receptors and ligands involved in pattern formation in the furrow. The proteins Act up and Chickadee/profilin regulate the actin polymerization necessary for this cell shape change. Cell shape changes, but not nuclear movement, appear to be coordinated with the timing of photoreceptor

differentiation as *act up* and *chickadee* mutants display premature differentiation.

In the middle of the furrow, cells aggregate to form equidistant clusters of 20 or so cells. These stereotypical ''rosettes'' are the rudiments of future ommatidia and they probably overlap with the so-called intermediate clusters, which express the bHLH transcription factors Atonal (ATO) and Daughterless (DA). The EGFR, the secreted fibrinogen-like Scabrous (SCA) protein, and the receptor/ligand pair N/Delta have all been implicated in a lateral inhibition process whereby rosette cells emit an inhibitory signal, preventing their near-neighbor cells from becoming part of the intermediate group of cells, and an inductive signal to determine the R8 precursor cell. This signal is thought to give rise to the periodic spacing of the rosettes and to whittle the ATO/DA expression down to just one cell, which will later become the R8 cell.

Every two hours, a new row of R precursor–cell clusters emerges from the furrow (Fig. 8). Each rosette transforms into a rearward-pointing curve of 7 to 9 cells – called an arc – and then, typically, seven of these cells zipper shut expelling all excess cells anteriorly. The definitive five-cell precluster containing the R8, R2, R5, R3, and R4 precursor cells forms as the R3 and R4 cells come to contact one another. In a mature larval eye disc, there are about 25 rows that exhibit progressive stages of ommatidial morphogenesis between the furrow and the posterior of the disc. Photoreceptor precursor cell subtype is specified sequentially; first R8 followed by R2/R5, then R3/R4, and then R1/R6. R7 is the final photoreceptor precursor cell to express neural antigens and the last one to differentiate. A series of cell-autonomously acting transcription factors together with secreted growth factors is responsible for establishing the unique identity of each photoreceptor cell.

The array of ommatidia is organized into two mirror-image fields divided by a dorsoventral midline, called the equator. This pattern requires a 90° rotation of the ommatidial clusters behind the furrow in the larval eye disc (Fig. 8). Two vectorial signals are necessary to regulate the direction and extent of rotation as well as a rearrangement of the cells in the early symmetric cluster to create the mature asymmetric cluster. One signal is emitted by cells in the morphogenetic furrow and includes the vesicular transport of SCA to more posterior cells using an actin-mediated mechanism. The other signal involves molecules important for tissue polarity in other epithelia. These include components of the WNT/FZ signaling pathway, namely, the interacting cytoplasmic proteins DSH and Prickle-Spiny Legs; the atypical cadherin, Flamingo/Starry night; the RHOA GTPase, and its kinase dROK, the kinase Misshapen; and a JNK cascade that sets up the distinction between R3 and R4 fates by regulating the level of the Puckered phosphatase and Delta in these cells. Other factors that regulate polarity include the zinc-finger transcription factor, HNT, which may do so by downregulating JNK signaling, and the novel four-pass transmembrane protein, Strabismus/van Gogh. It is not yet clear how the output of these signals controls planar polarity in the eye, but the rotation itself requires the serine/threonine protein kinase Nemo and the uncharacterized Roulette protein. It is likely that the cluster movements and changes in cell shape involve modifications of the cytoskeleton and the ECM, but so far, the only concrete link with these components is a requirement for the cadherins,

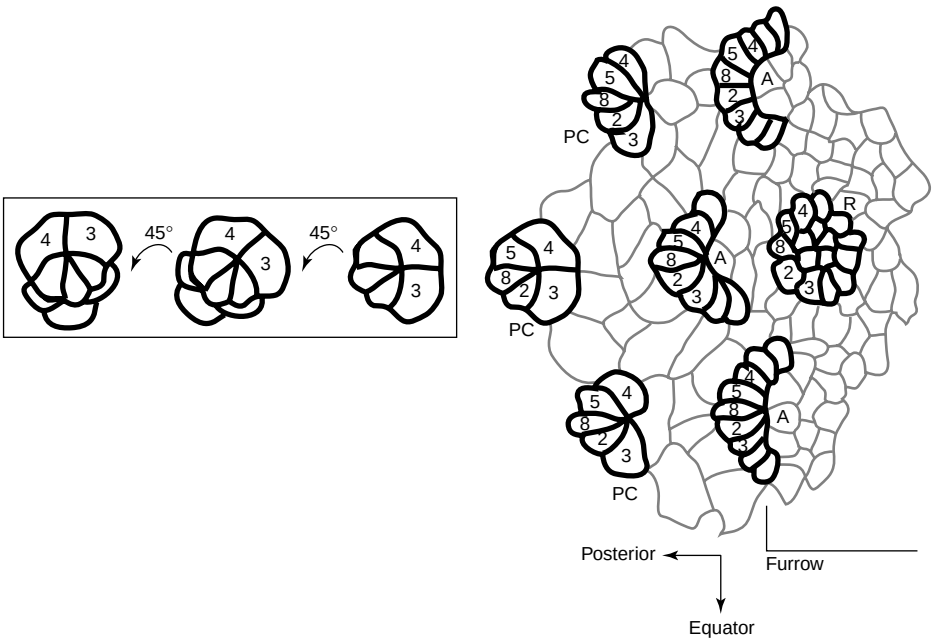

Fig. 8 Ommatidial cluster formation and rotation during *Drosophila* retinal morphogenesis. The diagram shows the apical profiles of the photoreceptor cell precursors (R2, 3, 4, 5, and 8) on the dorsal side of a third instar larval eye disc. Every two hours, another column of clusters is added anteriorly (on the right). In the furrow, clusters of 20 or so cells become "rosettes" (R). These open up anteriorly to form "arcs" (A). The arcs then zipper shut and expel cells on their anterior side to form the "five-cell precluster" (PC) The box on the left is a schematic (compressed over many columns) showing how ommatidial planar polarity is established in the eye disc. Initially, the symmetrical five-cell precluster is arranged so that the R3 precursor is closer to the equator than the R4 precursor. Subsequently, the cluster rotates 90° toward the equator. In its mature form (not shown here), mirror-image symmetry is broken to form a chiral ommatidium as the R4 cell loses contact with the R8 cell and moves to a more equatorial position to give the trapezoidal arrangement of the rhabdomeres seen in the adult ommatidium shown in Fig. 7. Figure based on Wolff, T., Ready, D. F. (1993) Pattern Formation in the *Drosophila* Retina, in: *The Development of Drosophila melanogaster*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 1277–1325.

Dachsous, and Fat, upstream of FZ for correct polarity.

During pupal development, order emerges in the retinal epithelium as specific contacts are made between retinal cells. Most of the remaining cell types, including the bristles and the pigment cells, are recruited during pupal stages in a process involving the N-receptor. Cell fate determination depends on each cell's position and contacts. For example, a cell that contacts the anterior or the posterior cone cell becomes a primary pigment cell, whereas a cell that contacts primaries from two distinct ommatidia becomes a secondary pigment cell. A combination of these kinds of cell–cell interactions together with two waves of programmed cell death transforms a nonpatterned epithelium into a

precise array of functional ommatidial units.

Recent data suggest that photoreceptors develop in a two-step process and that each step requires discrete genetic regulation. In the first step, during larval stages, the cells adopt a particular neuronal fate. During the second step, in pupal stages, these neurons execute their differentiation program and acquire final mature properties including *rhodopsin* gene expression. Once the cone cells have closed above the R-cells, the apical surfaces of the photoreceptors involute into the retinal epithelium to form the adult photosensitive rhabdomeres. During subsequent rhabdomere morphogenesis, the prominent adherens junctions between the R-cells are maintained and drive the longitudinal extension of the apical surfaces down to the retinal floor. Recent work has shown that *Drosophila* CRB is essential for rhabdomere elongation and stalk membrane biogenesis. CRB is a central component of a molecular scaffold (also involving β_H-spectrin, moesin, and FERM, a protein 4.1 superfamily member) that controls adherens junction assembly as described in Sect. 1.3 above. Other molecules associated with adherens junctions have also been implicated in rhabdomere elongation. These include Eyes Closed (EYC), the fly homolog of p47, involved in membrane fusion, and the Canoe protein, which is known to bind F-actin. In *eyc* mutants, β-catenin and E-cadherin fail to concentrate into the adherens junctions during pupal stages and this results in fragmented rhabdomeres with inappropriate contacts in the adult. The Canoe gene product directs straight elongation of the rhabdomere, since mutants possess twisted and fused rhabdomeres.

In general, cell–cell and cell–ECM contacts that are established early in eye development persist into adulthood. Rhabdomeres become aligned via cell–cell and cell–ECM adhesions to the retinal floor at the cone cell plate. It has been shown that the integrins play an important role in the development of the three-dimensional structure of the mature retina. Integrin mutants show several eye phenotypes that originate in the structural failure of the cone cell plate. The integrity of the retinal epithelium is intimately linked to its cellular architecture since many of the components necessary for its development and function – such as the integrins, CRB, HNT, and the photopigment, rhodopsin – are also necessary to prevent photoreceptor degeneration. For this reason, the *Drosophila* eye epithelium provides an excellent model for a genetic dissection of the causes underlying progressive human neurodegenerative diseases.

4 Concluding Remarks

The combination of genetic and molecular analyses in model organisms such as *Drosophila*, with biochemical and cell biological analyses on vertebrate cells in culture, has led to a significant understanding of both the structure of epithelia and the mechanisms by which simple, undifferentiated epithelia are converted into complex, differentiated tissues and organs. Since most of the components of epithelia are evolutionarily conserved, results obtained in one organism or cell type have generalized to others. Three areas of research are likely to predominate and synergize in the future. First, novel methods of cell culture, particularly in three dimensions, will lead to increasingly "*in vivo*–like" models of cell behavior, thus

benefiting cell biological and biochemical strategies. Second, the ability to image epithelial shape changes in living tissues or organisms at very high spatial and temporal resolution, and using specific molecular markers, will lead to increased understanding of dynamic changes during epithelial morphogenesis *in vivo*. Finally, the availability of complete genome sequences, together with mutational and phenotypic analysis of evolutionarily conserved components in invertebrate models (e.g. *Drosophila, C. elegans*) as well as in vertebrates (e.g. zebra fish, mouse), will lead to a comprehensive understanding of the structural and regulatory mechanisms of epithelial morphogenesis during animal development.

See also Developmental Cell Biology.

Bibliography

Books and Reviews

Ashkenas, J., Muschler, J., Bissell, M.J. (1996) The extracellular matrix in epithelial biology: shared molecules and common themes in distant phyla, *Dev. Biol.* **180**, 433–444.

Brown, N.H. (2000) Cell-cell adhesion via the ECM: integrin genetics in fly and worm, *Matrix Biol.* **19**, 191–201.

Burke, R.D. (1999) Invertebrate integrins: structure, function and evolution, *Int. Rev. Cytol.* **191**, 257–284.

Campos-Ortega, J.A., Hartenstein, V. (1985) *The Embryonic Development of Drosophila melanogaster*, Springer-Verlag, Heidelberg.

Chan, H., Bonini, N. (2000) *Drosophila* models of human neurodegenerative disease, *Cell Death Diff.* **7**, 1075–1080.

Chen, H., Bernstein, B.W., Bamburg, J.R. (2000) Regulating actin-filament dynamics *in vivo*, *Trends Biochem. Sci.* **25**, 19–23.

Cohen, S. (1993) Imaginal Disc Development, in: Bate, M., Martinez-Arias, A. (Eds.) *The Development of Drosophila melanogaster*, Vol. 2, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 747–841.

Conacci-Sorrell, M., Zhurinsky, J., Ben-Ze'ev, A. (2002) The cadherin-catenin adhesion system in signaling and cancer, *J. Clin. Invest.* **109**, 987–991.

Cukierman, E., Pankov, R., Yamada, K.M. (2002) Cell interactions with three-dimensional matrices, *Curr. Opin. Cell Biol.* **14**, 633–639.

Garrod, D.R. (1993) Desmosomes and hemidesmosomes, *Curr. Opin. Cell Biol.* **5**, 30–40.

Geiger, B., Bershadsky, A., Pankov, R., Yamada, K.M. (2001) Transmembrane crosstalk between the extracellular matrix-cytoskeleton crosstalk, *Nat. Rev. Mol. Cell Biol.* **2**, 793–805.

Grose, R., Martin, P. (1999) Parallels between wound repair and morphogenesis in the embryo, *Semin. Cell Dev. Biol.* **10**, 395–404.

Gumbiner, B. (1987) Structure, biochemistry, and assembly of epithelial tight junctions, *Am. J. Phys.* **253**, 749–758.

Gumbiner, B.M. (1996) Cell adhesion: the molecular basis of tissue architecture and morphogenesis, *Cell* **84**, 345–357.

Hartenstein, V. (1993) *Atlas of Drosophila Development*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Hynes, R.O. (1992) Integrins: versatility, modulation, and signaling in cell adhesion, *Cell* **69**, 11–25.

Hynes, R.O., Zhao, Q. (2000) The evolution of cell adhesion, *J. Cell Biol.* **150**, 89–95.

Jacinto, A., Martinez-Arias, A., Martin, P. (2001) Mechanism of epithelial fusion and repair, *Nat. Cell Biol.* **3**, E117–E123.

Jamora, C., Fuchs, E. (2002) Intercellular adhesion, signalling and the cytoskeleton, *Nat. Cell Biol.* **4**, E101–E108.

Kiehart, D.P. (1999) Wound healing: The power of the purse string, *Curr. Biol.* **9**, R602–R605.

Machesky, L.M., Insall, R.H. (1999) Signaling to actin dynamics, *J. Cell Biol.* **146**, 267–272.

Manning, G., Krasnow, M.A. (1993) Development of the *Drosophila* Tracheal System, in: Bate, M., Martinez Arias, A. (Eds.) *The Development of Drosophila Melanogaster*, Vol. 1, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 609–685.

Metzger, R.J., Krasnow, M.A. (1999) Genetic control of branching morphogenesis, *Science* **284**, 1635–1639.

Miranti, C.K., Brugge, J.S. (2002) Sensing the environment: a historical perspective on integrin signal transduction, *Nat. Cell Biol.* **4**, 83–90.

Mosher, D.F., Sottile, J., Wu, C., McDonald, J.A. (1992) Assembly of extracellular matrix, *Curr. Opin. Cell Biol.* **4**, 810–818.

Muller, H.A. (2000) Genetic control of epithelial cell polarity: lessons from *Drosophila*, *Dev. Dyn.* **218**, 52–67.

Nelson, W.J., Yeaman, C. (2001) Protein trafficking in the exocytic pathway of polarized epithelial cells, *Trends Cell Biol.* **11**, 483–486.

Noselli, S. (1998) JNK signaling and morphogenesis in *Drosophila*, *Trends Genet.* **14**, 33–38.

Petit, V., Thiery, J.-P. (2000) Focal adhesions: structure and dynamics, *Biol. Cell* **92**, 477–494.

Reifegerste, R., Moses, K. (1999) Genetics of epithelial polarity and pattern in the *Drosophila* retina, *BioEssays* **21**, 275–285.

Simon, M.A. (2000) Receptor tyrosine kinases: specific outcomes from general signals, *Cell* **103**, 13–15.

Tepass, U. (2002) Adherens junctions: new insight into assembly, modulation and function, *BioEssays* **24**, 690–695.

Tepass, U., Tanentzapf, G., Ward, R., Fehon, R. (2001) Epithelial cell polarity and cell junctions in *Drosophila*, *Annu. Rev. Genet.* **35**, 747–784.

Vasioukhin, V., Fuchs, E. (2001) Actin dynamics and cell-cell adhesion in epithelia, *Curr. Opin. Cell Biol.* **13**, 76–84.

Wodarz, A. (2000) Tumor suppressors: linking cell polarity and growth control, *Curr. Biol.* **10**, R624–R626.

Wolff, T., Ready, D.F. (1993) Pattern Formation in the *Drosophila* Retina, *The Development of Drosophila melanogaster*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 1277–1325.

Zahraoui, A., Louvard, D., Galli, T. (2000) Tight junction, a platform for trafficking and signaling protein complexes, *J. Cell Biol.* **151**, F31–F36.

Primary Literature

Adams, C.L., Chen, Y.T., Smith, S.J., Nelson, W.J. (1998) Mechanisms of epithelial cell-cell adhesion and cell compaction revealed by high-resolution tracking of E-cadherin-green fluorescent protein, *J. Cell Biol.* **142**, 1105–1119.

Adams, C.L., Nelson, W.J., Smith, S.J. (1996) Quantitative analysis of cadherin-catenin-actin reorganization during development of cell-cell adhesion, *J. Cell Biol.* **135**, 1899–1911.

Angres, B., Barth, A., Nelson, W.J. (1996) Mechanism for transition from initial to stable cell-cell adhesion: kinetic analysis of E-cadherin-mediated adhesion using a quantitative adhesion assay, *J. Cell Biol.* **134**, 549–557.

Baker, N., Zitron, A.E. (1995) *Drosophila* eye development: Notch and Delta amplify a neurogenic pattern conferred on the morphogenetic furrow by scabrous, *Mech. Dev.* **49**, 173–189.

Baker, N.E., Mlodzik, M., Rubin, G.M. (1990) Spacing differentiation in the developing *Drosophila* eye: a fibrinogen-related lateral inhibitor encoded by *scabrous*, *Science* **250**, 1370–1377.

Baonza, A., Casci, T., Freeman, M. (2001) A primary role for the epidermal growth factor receptor in ommatidial spacing in the *Drosophila* eye, *Curr. Biol.* **11**, 396–404.

Benlali, A., Draskovic, I., Hazelett, D.J., Treisman, J.E. (2000) Act up controls actin polymerization to alter cell shape and restrict hedgehog signaling in the *Drosophila* eye disc, *Cell* **101**, 271–281.

Bershadsky, A., Chausovsky, A., Becker, E., Lyubimova, A., Geiger, B. (1996) Involvement of microtubules in the control of adhesion-dependent signal transduction, *Curr. Biol.* **6**, 1279–1289.

Bilder, D., Li, M., Perrimon, N. (2000) Cooperative regulation of cell polarity and growth by *Drosophila* tumor suppressors, *Science* **289**, 113–116.

Bilder, D., Perrimon, N. (2000) Localization of apical epithelial determinants by the basolateral PDZ protein scribble, *Nature* **403**, 676–680.

Bilder, D., Schober, M., Perrimon, N. (2003) Integrated activity of PDZ protein complexes regulates epithelial polarity, *Nat. Cell Biol.* **5**, 53–58.

Bloor, J.W., Kiehart, D.P. (2002) *Drosophila* RhoA regulates the cytoskeleton and cell-cell adhesion in the developing epidermis, *Development* **129**, 3173–3183.

Borod, E.R., Heberlein, U. (1998) Mutual regulation of decapentaplegic and hedgehog during the initiation of differentiation in the *Drosophila* retina, *Dev. Biol.* **197**, 187–197.

Boube, M., Martin-Bermudo, M.D., Brown, N.H., Casanova, J. (2001) Specific tracheal migration is mediated by complementary expression of cell surface proteins, *Genes Dev.* **15**, 1554–1562.

Boutros, M., Mlodzik, M. (1999) Dishevelled: at the crossroads of divergent intracellular signaling pathways, *Mech. Dev.* **83**, 27–37.

Braga, V.M., Machesky, L.M., Hall, A., Hotchin, N.A. (1997). The small GTPases Rho and Rac are required for the establishment of cadherin-dependent cell-cell contacts. *J. Cell Biol.* **137**, 1421–1431.

Brown, N.L., Paddock, S.W., Sattler, C.A., Cronmiller, C., Thomas, B.J., Carroll, S.B. (1996) Daughterless is required for *Drosophila* photoreceptor cell determination, eye morphogenesis, and cell cycle progression, *Dev. Biol.* **179**, 65–78.

Cadigan, K.M., Jou, A.D., Nusse, R. (2002) Wingless blocks bristle formation and morphogenetic furrow progression in the eye through repression of daughterless, *Development* **129**, 3393–3402.

Cagan, R.L., Ready, D.F. (1989a) The emergence of order in the *Drosophila* pupal retina, *Dev. Biol.* **136**, 346–362.

Cagan, R.L., Ready, D.F. (1989b) Notch is required for successive cell decisions in the developing *Drosophila* retina, *Genes Dev.* **3**, 1099–1112.

Chae, J., Kim, M.J., Goo, J.H., Collier, S., Gubb, D., Charlton, J., Adler, P.N., Park, W.J. (1999) The *Drosophila* tissue polarity gene starry night encodes a member of the protocadherin family, *Development* **126**, 5421–5429.

Chanut, F., Heberlein, U. (1997) Role of decapentaplegic in initiation and progression of the morphogenetic furrow in the developing *Drosophila* retina, *Development* **124**, 559–567.

Choi, K.-W., Benzer, S. (1994) Rotation of photoreceptor clusters in the developing *Drosophila* eye requires the *nemo* gene, *Cell* **78**, 125–136.

Chou, C.F., Chien, C.T. (2002) Scabrous controls ommatidial rotation in the *Drosophila* compound eye, *Dev. Cell* **3**, 839–850.

Clark, E.A., King, W.G., Brugge, J.S., Symons, M., Hynes, R.O. (1998) Integrin-mediated signals regulated by members of the rho family of GTPases, *J. Cell Biol.* **142**, 573–586.

Cukierman, E., Pankov, R., Stevens, D.R., Yamada, K.N. (2001) Taking cell-matrix adhesions to the third dimension, *Science* **294**, 1708–1712.

Curtiss, J., Mlodzik, M. (2000) Morphogenetic furrow initiation and progression during eye development in *Drosophila*: the roles of decapentaplegic, hedgehog and eyes absent, *Development* **127**, 1325–1336.

Das, G., Reynolds-Kenneally, J., Mlodzik, M. (2002) The atypical cadherin flamingo links frizzled and Notch signaling in planar polarity establishment in the *Drosophila* eye, *Dev. Cell* **2**, 655–666.

den Hollander, A.I., ten Brink, J.B., de Kok, Y.J., van Soest, S., van den Born, L.I., van Driel, M.A., van de Pol, D.J., Payne, A.M., Bhattacharya, S.S., Kellner, U., et al. (1999) Mutations in a human homolog of *Drosophila* crumbs cause retinitis pigmentosa (RP12), *Nat. Genet.* **23**, 217–221.

Dominguez, M., Hafen, E. (1997) Hedgehog directly controls initiation and propagation of retinal differentiation in the *Drosophila* eye, *Genes Dev.* **11**, 3254–3264.

Dubreuil, R.R., Wang, P., Dahl, S., Lee, J., Goldstein, L.S. (2000) *Drosophila* beta spectrin functions independently of alpha spectrin to polarize the Na,K ATPase in epithelial cells, *J. Cell Biol.* **149**, 647–656.

Dutta, D., Bloor, J.W., Ruiz-Gomez, M., Vijay Raghavan, K., Kiehart, D.P. (2002) Real-time imaging of morphogenetic movements in *Drosophila* using Gal4-UAS-driven expression of GFP fused to the actin-binding domain of moesin, *Genesis* **34**, 146–151.

Fanto, M., Mlodzik, M. (1999) Asymmetric Notch activation specifies photoreceptors R3 and R4 and planar polarity in the *Drosophila* eye, *Nature* **397**, 523–526.

Fischer-Vize, J.A., Mosley, K.L. (1994) Marbles mutants: uncoupling cell determination and nuclear migration in the developing *Drosophila* eye, *Development* **120**, 2609–2618.

Frankfort, B.J., Mardon, G. (2002) R8 development in the *Drosophila* eye: a paradigm for neural selection and differentiation, *Development* **129**, 1295–1306.

Furuse, M., Itoh, M., Hirase, T., Nagafuchi, A., Yonemura, S., Tsukita, S. (1994) Direct association of occludin with ZO-1 and its possible involvement in the localization of occludin at tight junctions, *J. Cell Biol.* **127**, 1617–1626.

Glise, B., Noselli, S. (1997) Coupling of Jun amino-terminal kinase and decapentaplegic signaling pathways in *Drosophila* morphogenesis, *Genes Dev.* **11**, 1738–1747.

Grawe, F., Wodarz, A., Lee, B., Knust, E., Skaer, H. (1996) The *Drosophila* genes crumbs and stardust are involved in the biogenesis of adherens junctions, *Development* **122**, 951–959.

Gubb, D., Green, C., Huen, D., Coulson, D., Johnson, G., Tree, D., Collier, S., Roote, J. (1999) The balance between isoforms of the prickle LIM domain protein is critical for planar polarity in *Drosophila* imaginal discs, *Genes Dev.* **13**, 2315–2327.

Guillemin, K., Groppe, J., Dücker, K., Triesman, R., Hafen, E., Affolter, M., Krasnow, M. (1996) The pruned gene encodes the *Drosophila* serum response factor and regulates cytoplasmic outgrowth during terminal branching of the tracheal system, *Development* **122**, 1353–1362.

Hacohen, N., Kramer, S., Sutherland, D., Hiromi, Y., Krasnow, M.A. (1998) *Sprouty* encodes a novel antagonist of FGF signaling that patterns apical branching of the *Drosophila* airways, *Cell* **92**, 253–263.

Hanada, T., Lin, L., Tibaldi, E.V., Reinherz, E.L., Chishti, A.H. (2000) GAKIN, a novel kinesin-like protein associates with the human homolog of the *Drosophila* discs large tumor suppressor in T lymphocytes, *J. Biol. Chem.* **275**, 28774–28784.

Harden, N., Lee, J., Loh, H.-Y., Ong, Y.-M., Tan, I., Leung, T., Manser, E., Lim, L. (1996) A *Drosophila* homolog of the Rac- and Cdc42-activated serine/ threonine kinase PAK is a potential focal adhesion and focal complex protein that colocalizes with dynamic actin structures, *Mol. Cell. Biol.* **16**, 1896–1908.

Harden, N., Loh, H.Y., Chia, W., Lim, L. (1995) A dominant inhibitory version of the small GTP-binding protein Rac disrupts cytoskeletal structures and inhibits developmental cell shape changes in *Drosophila*, *Development* **121**, 903–914.

Harden, N., Ricos, M., Ong, Y.M., Chia, W., Lim, L. (1999) Participation of small GTPases in dorsal closure of the *Drosophila* embryo: distinct roles for Rho subfamily proteins in epithelial morphogenesis, *J. Cell Sci.* **112**, 273–284.

Heberlein, U., Singh, C.M., Luk, A.Y., Donohoe, T.J. (1995) Growth and differentiation in the *Drosophila* eye coordinated by hedgehog, *Nature* **373**, 709–711.

Heberlein, U., Wolff, T., Rubin, G.M. (1993) The TGFβ homolog dpp and the segment polarity gene hedgehog are required for propagation of a morphogenetic wave in the *Drosophila* retina, *Cell* **75**, 913–926.

Heiskala, M., Peterson, P.A., Yang, Y. (2001) The roles of claudin superfamily proteins in paracellular transport, *Traffic* **2**, 93–98.

Horsfield, J., Penton, A., Secombe, J., Hoffman, F.M., Richardson, H. (1998) Decapentaplegic is required for arrest in G1 phase during *Drosophila* eye development, *Development* **125**, 5069–5078.

Hou, X.S., Goldstein, E.S., Perrimon, N. (1997) *Drosophila* Jun relays the Jun amino-terminal kinase signal transduction pathway to the Decapentaplegic signal transduction pathway in regulating epithelial cell sheet movement, *Genes Dev.* **11**, 1728–1737.

Issac, D.D., Andrew, D.J. (1996) Tubulogenesis in *Drosophila*: a requirement for the *trachealess* gene product, *Genes Dev.* **10**, 103–117.

Izaddoost, S., Nam, S.C., Bhat, M.A., Bellen, H.J., Choi, K.W. (2002) *Drosophila* crumbs is a positional cue in photoreceptor adherens junctions and rhabdomeres, *Nature* **416**, 178–183.

Jacinto, A., Wood, W., Balayo, T., Turmaine, M., Martinez-Arias, A., Martin, P. (2000) Dynamic actin-based epithelial adhesion and cell matching during *Drosophila* dorsal closure, *Curr. Biol.* **10**, 1420–1426.

Jarman, A.P., Grau, Y., Jan, L.Y., Jan, Y.N. (1993) Atonal is a proneural gene that directs chordotonal organ formation in the *Drosophila* peripheral nervous system, *Cell* **73**, 1307–1321.

Jarman, A.P., Grell, E.H., Ackerman, L., Jan, L.Y., Jan, Y.N. (1994) Atonal is the proneural gene for *Drosophila* photoreceptors, *Nature* **369**, 398–400.

Jarman, A.P., Sun, Y., Jan, L.Y., Jan, Y.N. (1995) Role of the proneural gene, atonal, in formation of *Drosophila* chordotonal organs and photoreceptors, *Development* **121**, 2019–2030.

Jasper, H., Benes, V., Schwager, C., Sauer, S., Clauder-Munster, S., Ansorge, W., Bohmann, D. (2001) The genomic response of the *Drosophila* embryo to JNK signaling, *Dev. Cell* **1**, 579–586.

Johnson, K., Grawe, F., Grzeschik, N., Knust, E. (2002) *Drosophila* crumbs is required to inhibit light-induced photoreceptor degeneration, *Curr. Biol.* **12**, 1675–1680.

Jou, T.S., Schneeberger, E.E., Nelson, W.J. (1998) Structural and functional regulation of tight junctions by RhoA and Rac1 small GTPases, *J. Cell Biol.* **142**, 101–115.

Kaltschmidt, J.A., Lawrence, N., Morel, V., Balayo, T., Fernandez, B.G., Pelissier, A., Jacinto, A., Martinez-Arias, A. (2002) Planar polarity and actin dynamics in the epidermis of *Drosophila*, *Nat. Cell Biol.* **4**, 937–944.

Keely, P.J., Westwick, J.K., Whitehead, I.P., Der, C.J., Parise, L.V. (1997) Cdc42 and Rac1 induce integrin-mediated cell motility and invasiveness through PI(3)K, *Nature* **390**, 632–636.

Kiehart, D.P., Galbraith, C.G., Edwards, K.A., Rickoll, W.L., Montague, R.A. (2000) Multiple forces contribute to cell sheet morphogenesis for dorsal closure in *Drosophila*, *J. Cell Biol.* **149**, 471–490.

Klambt, C. (1993) The *Drosophila* gene *pointed* encodes two ETS-like proteins which are involved in the development of the midline glial cells, *Development* **117**, 163–176.

Klein, G., Langegger, M., Timpl, R., Ekblom, P. (1988) Role of laminin A chain in the development of epithelial cell polarity, *Cell* **55**, 331–341.

Krendel, M.F., Bonder, E.M. (1999) Analysis of actin filament bundle dynamics during contact formation in live epithelial cells, *Cell Motil. Cytoskeleton* **43**, 296–309.

Kroschewski, R., Hall, A., Mellman, I. (1999) Cdc42 controls secretory and endocytic transport to the basolateral plasma membrane of MDCK cells, *Nat. Cell Biol.* **1**, 8–13.

Kumar, J.P., Moses, K. (2001) The EGF receptor and Notch signaling pathways control the initiation of the morphogenetic furrow during *Drosophila* eye development, *Development* **128**, 2689–2697.

Kumar, J.P., Tio, M., Hsiung, F., Akopyan, S., Gabay, L., Seger, R., Shilo, B.-Z., Moses, K. (1998) Dissecting the roles of the *Drosophila* EGF receptor in eye development and MAP kinase activation, *Development* **125**, 3875–3885.

Lee, E.-C., Hu, X., Yu, S.-Y., Baker, N.E. (1996) The *scabrous* gene encodes a secreted glycoprotein dimer and regulates proneural development in *Drosophila* eyes, *Mol. Cell. Biol.* **16**, 1179–1188.

Lee, S., Kolodziej, P.A. (2002) The plakin short stop and the RhoA GTPase are required for E-cadherin-dependent apical surface remodeling during tracheal tube fusion, *Development* **129**, 1509–1520.

Legouis, R., Gansmuller, A., Sookhareea, S., Bosher, J.M., Baillie, D.L., Labouesse, M. (2000) LET-413 is a basolateral protein required for the assembly of adherens junctions in *Caenorhabditis elegans*, *Nat. Cell Biol.* **2**, 415–422.

Leonard, D.S., Bowman, V.D., Ready, D.F., Pak, W.L. (1992) Degeneration of photoreceptors in rhodopsin mutants of *Drosophila*, *J. Neurobiol.* **23**, 605–626.

Liu, S., Calderwood, D.A., Ginsberg, M.H. (2000) Integrin cytoplasmic domain-binding proteins, *J. Cell Sci.* **113**, 3563–3571.

Liu, S., Ginsberg, M.H. (2000) Paxillin binding to a conserved sequence motif in the alpha 4 integrin cytoplasmic domain, *J. Biol. Chem.* **275**, 22736–22742.

Longley, R.L., Jr., Ready, D.F. (1995) Integrins and the development of three-dimensional structure in the *Drosophila* compound eye, *Dev. Biol.* **171**, 415–433.

Ma, C., Zhou, Y., Beachy, P.A., Moses, K. (1993) The segment polarity gene hedgehog is required for progression of the morphogenetic furrow in the developing *Drosophila* eye, *Cell* **75**, 927–938.

Mackay, S., Booth, S.H., MacGowan, A., Smith, R.A. (1999) Ultrastructural studies demonstrate that epithelial polarity is established in cultured mouse pre-Sertoli cells by extracellular matrix components, *J. Electron Microsc. (Tokyo)* **48**, 159–165.

Magie, C.R., Meyer, M.R., Gorsch, M.S., Parkhurst, S.M. (1999) Mutations in the Rho1 small GTPase disrupt morphogenesis and segmentation during early *Drosophila* development, *Development* **126**, 5353–5364.

Martin-Blanco, E., Gampel, A., Ring, J., Virdee, K., Kirov, N., Tolkovsky, A.M., Martinez-Arias, A. (1998) *Puckered* encodes a phosphatase that mediates a feedback loop regulating JNK activity during dorsal closure in *Drosophila*, *Genes Dev.* **12**, 557–570.

Matsuo, T., Takahashi, K., Suzuki, E., Yamamoto, D. (1999) The Canoe protein is necessary in adherens junctions for development

of ommatidial architecture in the *Drosophila* compound eye, *Cell Tissue Res.* **298**, 397–404.

Medina, E., Williams, J., Klipfell, E., Zarnescu, D., Thomas, G., Le Bivic, A. (2002) Crumbs interacts with moesin and beta(Heavy)-spectrin in the apical membrane skeleton of *Drosophila*, *J. Cell Biol.* **158**, 941–951.

Miao, H., Li, S., Hu, Y.-L., Yuan, S., Zhao, Y., Chen, B.P.C., Puzon-McLaughlin, W., Tarui, T., Shyy, J.Y.-J., Takada, Y., Usami, S., Chien, S. (2002) Differential regulation of Rho GTPases by $\beta 1$ and $\beta 3$ integrins: the role of extracellular domain of integrin in intracellular signaling, *J. Cell Sci.* **115**, 2199–2206.

Miller, D.T., Cagan, R.L. (1998) Local induction of patterning and programmed cell death in the developing *Drosophila* retina, *Development* **125**, 2327–2335.

Miyamoto, S., Teramoto, H., Coso, O.A., Gutkind, J.S., Burbelo, P.D., Akiyama, S.K., Yamada, K.M. (1995) Integrin function: molecular hierarchies of cytoskeletal and signaling molecules, *J. Cell Biol.* **131**, 791–805.

Mollereau, B., Dominguez, M., Webel, R., Colley, N.J., Keung, B., de Celis, J.F., Desplan, C. (2001) Two-step process for photoreceptor formation in *Drosophila*, *Nature* **412**, 911–913.

Nobes, C.D., Hall, A. (1999) Rho GTPases control polarity, protrusion, and adhesion during cell movement, *J. Cell Biol.* **144**, 1235–1244.

Noselli, S., Agnes, F. (1999) Roles of the JNK signaling pathway in *Drosophila* morphogenesis, *Curr. Opin. Genet. Dev.* **9**, 466–472.

Oda, H., Tsukita, S. (1999) Dynamic features of adherens junctions during *Drosophila* embryonic epithelial morphogenesis revealed by a D-alpha-catenin-GFP fusion protein, *Dev. Genes Evol.* **209**, 218–225.

Ozawa, M., Ringwald, M., Kemler, R. (1990) Uvomorulin-catenin complex formation is regulated by a specific domain in the cytoplasmic region of the cell adhesion molecule, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4246–4250.

Pankov, R., Cukierman, E., Katz, B.Z., Matsumoto, K., Lin, D.C., Lin, S., Hahn, C., Yamada, K.M., (2000) Integrin dynamics and matrix assembly: tensin-dependent translocation of alpha(5)beta(1) integrins promotes early fibronectin fibrillogenesis, *J. Cell Biol.* **148**, 1075–1090.

Paricio, N., Feiguin, F., Boutros, M., Eaton, S., Mlodzik, M. (1999) The *Drosophila* STE20-like kinase misshapen is required downstream of the frizzled receptor in planar polarity signaling, *EMBO J.* **18**, 4669–4678.

Pellikka, M., Tanentzapf, G., Pinto, M., Smith, C., McGlade, C.J., Ready, D.F., Tepass, U. (2002) Crumbs, the *Drosophila* homolog of human CRB1/RP12, is essential for photoreceptor morphogenesis, *Nature* **416**, 143–149.

Penton, A., Selleck, S.B., Hoffmann, F.M. (1997) Regulation of cell cycle synchronization by decapentaplegic during *Drosophila* eye development, *Science* **275**, 203–6.

Pickup, T.A., Lamka, M.L., Sun, Q., Yip, M.L.R., Lipshitz, H.D. (2002) Control of photoreceptor cell morphology, planar polarity and epithelial integrity during *Drosophila* eye development, *Development* **129**, 2247–2258.

Pignoni, F., Zipursky, S.L. (1997) Induction of *Drosophila* eye development by decapentaplegic, *Development* **124**, 271–278.

Poulson, D.F. (1950). Histogenesis, Organogenesis and Differentiation in the Embryo of *Drosophila* melanogaster (Meigen), in: Demerec, M. (Ed.) *Biology of Drosophila*, Wiley, New York, pp. 168–274.

Price, J.V., Clifford, R.J., Schüpbach, T. (1989) The maternal ventralizing locus torpedo is allelic to faint little ball, an embryonic lethal, and encodes the *Drosophila* EGF receptor homolog, *Cell* **56**, 1085–1092.

Raich, W.B., Agbunag, C., Hrdin, J. (1999) Rapid epithelial-sheet sealing in the *Caenorhabditis elegans* embryo requires cadherin-dependent filopidial priming, *Curr. Biol.* **9**, 1139–1146.

Rawls, A.S., Guinto, J.B., Wolff, T. (2002) The cadherins fat and dachsous regulate dorsal/ventral signaling in the *Drosophila* eye, *Curr. Biol.* **12**, 1021–1026.

Ready, D.F., Hanson, T.E., Benzer, S. (1976) Development of the *Drosophila* retina, a neurocrystalline lattice, *Dev. Biol.* **53**, 217–240.

Reed, B.H., Wilk, R., Lipshitz, H.D. (2001) Downregulation of Jun kinase signaling in the amnioserosa is essential for dorsal closure of the *Drosophila* embryo, *Curr. Biol.* **11**, 1098–1108.

Reichman-Freid, M., Dickson, B., Hafen, E., Shilo, B.Z. (1994) Elucidation of the role of breathless, a *Drosophila* FGF receptor homolog, in tracheal cell migration, *Genes Dev.* **8**, 428–439.

Reifegerste, R., Ma, C., Moses, K. (1997) A polarity field is established early in the development of the *Drosophila* compound eye, *Mech. Dev.* **68**, 69–79.

Ribeiro, C., Ebner, A., Affolter, M. (2002) In vivo imaging reveals different cellular functions for FGF and Dpp signaling in tracheal branching morphogenesis, *Dev. Cell* **2**, 677–683.

Riesgo-Escovar, J.R., Hafen, E. (1997) *Drosophila* Jun kinase regulates expression of decapentaplegic via the ETS-domain protein Aop and the AP-1 transcription factor DJun during dorsal disclosure, *Genes Dev.* **11**, 1717–1727.

Rimm, D.L., Koslov, E.R., Kebriaei, P., Cianci, C.D., Morrow, J.S. (1995) α1(E)-catenin is an actin-binding and -bundling protein mediating the attachment of F-actin to the membrane adhesion complex, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8813–8817.

Riveline, D., Zamir, E., Balaban, N.Q., Schwarz, U.S., Ishizaki, T., Narumiya, S., Kam, Z., Geiger, B., Bershadsky, A.D. (2001) Focal contacts as mechanosensors: externally applied local mechanical force induces growth of focal contacts by an mDia1-dependent and ROCK-independent mechanism, *J. Cell Biol.* **153**, 1175–1186.

Royet, J., Finkelstein, R. (1997) Establishing primordia in the *Drosophila* eye-antennal imaginal disc: the roles of decapentaplegic, wingless and hedgehog, *Development* **124**, 4793–800.

Samakovlis, C., Manning, G., Steneberg, P., Hacohen, N., Cantera, R., Krasnow, M.A. (1996) Genetic control of epithelial tube fusion during *Drosophila* tracheal development, *Development* **122**, 3531–3536.

Sang, T.K., Ready, D.F. (2002) Eyes closed, a *Drosophila* p47 homolog, is essential for photoreceptor morphogenesis, *Development* **129**, 143–154.

Schejter, E.D., Shilo, B.-Z. (1989) The *Drosophila* EGF receptor homolog (DER) gene is allelic to faint little ball, a locus essential for embryonic development, *Cell* **56**, 1093–1104.

Slade, M.J., Coope, R.C., Gomm, J.J., Coombes, R.C. (1999) The human mammary gland basement membrane is integral to the polarity of luminal epithelial cells, *Exp. Cell Res.* **247**, 267–278.

Sonnenfeld, M., Ward, M., Nystrom, G., Mosher, J., Stahl, S., Crews, S. (1997) The *Drosophila tango* gene encodes a bHLH-PAS protein that is orthologous to mammalian Arnt and controls CNS midline and tracheal development, *Development* **124**, 4571–4582.

Spencer, S.A., Powell, P.A., Miller, D.T., Cagan, R.L. (1998) Regulation of EGF receptor signaling establishes pattern across the developing *Drosophila* retina, *Development* **125**, 4777–4790.

Strutt, D.I., Weber, U., Mlodzik, M. (1997) The role of RhoA in tissue polarity and frizzled signalling, *Nature* **387**, 202–295.

Sutherland, D., Samakovlis, C., Krasnow, M.A. (1996) Branchless encodes a *Drosophila* FGF homolog that controls tracheal cell migration and the pattern of branching, *Cell* **87**, 1091–1101.

Sutherland, J.D., Witke, W. (1999) Molecular genetic approaches to understanding the actin cytoskeleton, *Curr. Opin. Cell Biol.* **11**, 142–151.

Tanaka-Matakatsu, M., Uemura, T., Oda, H., Takeichi, M., Hayashi, S. (1996) Cadherin-mediated cell adhesion and cell motility in *Drosophila* trachea regulated by the transcription factor Escargot, *Development* **122**, 3697–3705.

Tanentzapf, G., Tepass, U. (2003) Interactions between the crumbs, lethal giant larvae and bazooka pathways in epithelial polarization, *Nat. Cell Biol.* **5**, 46–52.

Taylor, J., Abramova, N., Charlton, J., Adler, P.N. (1998) Van Gogh: a new *Drosophila* tissue polarity gene, *Genetics* **150**, 199–210.

Tepass, U. (1996) *Crumbs*, a component of the apical membrane, is required for zonula adherens formation in primary epithelia of *Drosophila*, *Dev. Biol.* **177**, 217–225.

Tepass, U., Speicher, S.A., Knust, E. (1990) Crumbs encodes a EGF-like protein expressed on apical membranes of *Drosophila* epithelial-cells and required for organization of epithelia, *Cell* **61**, 787–799.

Thomas, B.J., Gunning, D.A., Cho, J., Zipursky, L. (1994) Cell cycle progression in the developing *Drosophila* eye: roughex encodes a novel protein required for the establishment of G1, *Cell* **77**, 1003–1014.

Tomlinson, A. (1985) The cellular dynamics of pattern formation in the eye of *Drosophila*, *J. Embryol. Exp. Morphol.* **89**, 313–331.

Tomlinson, A., Ready, D.F. (1987) Neuronal differentiation in the *Drosophila* ommatidium, *Dev. Biol.* **120**, 366–376.

Usui, T., Shima, Y., Shimada, Y., Hirano, S., Burgess, R.W., Schwarz, T.L., Takeichi, M., Uemura, T. (1999) Flamingo, a seven-pass transmembrane cadherin, regulates planar cell polarity under the control of frizzled, *Cell* **98**, 585–595.

Van Aelst, L., Symons, M. (2002) Role of Rho family GTPases in epithelial morphogenesis, *Genes Dev.* **16**, 1032–1054.

Vasioukhin, V., Bauer, C., Yin, M., Fuchs, E. (2000) Directed actin polymerization is the driving force for epithelial cell-cell adhesion, *Cell* **100**, 209–219.

Vinson, C.R., Conover, S., Adler, P.N. (1989) A *Drosophila* tissue polarity locus encodes a protein containing seven potential transmembrane domains, *Nature* **338**, 263–264.

Waterman-Storer, C.M., Salmon, W.C., Salmon, E.D. (2000) Feedback interactions between cell-cell adherens junctions and cytoskeletal dynamics in newt lung epithelial cells, *Mol. Biol. Cell* **11**, 2471–2483.

Weber, U., Paricio, N., Mlodzik, M. (2000) Jun mediates frizzled-induced R3/R4 cell fate distinction and planar polarity determination in the *Drosophila* eye, *Development* **127**, 3619–3629.

Wilk, R., Reed, B.H., Tepass, U., Lipshitz, H.D. (2000) The hindsight gene is required for epithelial maintenance and differentiation of the tracheal system in *Drosophila*, *Dev. Biol.* **219**, 183–196.

Wilk, R., Weizman, I., Shilo, B.-Z. (1996) Trachealess encodes a bHLH-PAS protein that is an inducer of tracheal cell fates in *Drosophila*, *Genes Dev.* **10**, 93–102.

Williams-Masson, E.M., Malik, A.N., Hardin, J. (1997) An actin-mediated two-step mechanism is required for ventral enclosure of the *C. elegans* hypodermis, *Development* **124**, 2889–2901.

Winter, C.G., Wang, B., Ballew, A., Royou, A., Karess, R., Axelrod, J.D., Luo, L. (2001) *Drosophila* Rho-associated kinase (Drok) links frizzled-mediated planar cell polarity signaling to the actin cytoskeleton, *Cell* **105**, 81–91.

Wodarz, A., Hinz, U., Engelbert, M., Knust, E. (1995) Expression of crumbs confers apical character on plasma membrane domains of ectodermal epithelia of *drosophila*, *Cell* **82**, 67–76.

Wolff, T., Ready, D.F. (1991) The beginning of pattern formation in the *Drosophila* compound eye: the morphogenetic furrow and the second mitotic wave, *Development* **113**, 841–850.

Wolff, T., Rubin, G.M. (1998) Strabismus, a novel gene that regulates tissue polarity and cell fate decisions in *Drosophila*, *Development* **125**, 1149–1159.

Wood, W., Jacinto, A., Grose, R., Woolner, S., Gale, J., Wilson, C., Martin, P. (2002) Wound healing recapitulates morphogenesis in *Drosophila* embryos, *Nat. Cell Biol.* **4**, 907–912.

Yang, C.H., Axelrod, J.D., Simon, M.A. (2002) Regulation of frizzled by fat-like cadherins during planar polarity signaling in the *Drosophila* compound eye, *Cell* **108**, 675–688.

Yang, L., Baker, N.E. (2001) Role of the EGFR/Ras/Raf pathway in specification of photoreceptor cells in the *Drosophila* retina, *Development* **128**, 1183–1191.

Young, P.E., Richman, A.M., Ketchum, A.S., Kiehart, D.P. (1993) Morphogenesis in *Drosophila* requires nonmuscle myosin heavy chain function, *Genes Dev.* **7**, 29–41.

Zamir, E., Katz, B.Z., Aota, S., Yamada, K.M., Geiger, B., Kam, Z. (1999) Molecular diversity of cell-matrix adhesions, *J. Cell Sci.* **112**, 1655–1669.

Zarnescu, D.C., Thomas, G.H. (1999) Apical spectrin is essential for epithelial morphogenesis but not apicobasal polarity in *Drosophila*, *J. Cell Biol.* **146**, 1075–1086.

Zelzer, E. (2000a) Cell fate choices in *Drosophila* tracheal morphogenesis, *BioEssays* **22**, 219–226.

Zelzer, E. (2000b) Interaction between the bHLH-PAS protein trachealess and the POU-domain protein drifter, specifies tracheal cell fates, *Mech. Dev.* **91**, 163–173.

Zhang, X., Bi, E., Novick, P., Du, L., Kozminski, K.G., Lipschutz, J.H., Guo, W. (2001) Cdc42 interacts with the exocyst and regulates polarized secretion, *J. Biol. Chem.* **276**, 46745–46750.

Zheng, L., Zhang, J., Carthew, R.W. (1995) *Frizzled* regulates mirror-symmetric pattern formation in the *Drosophila* eye, *Development* **121**, 3045–3055.

Zhurinsky, J., Shtutman, M., Ben-Ze'ev, A. (2000) Plakoglobin and beta-catenin: protein interactions, regulation and biological roles, *J. Cell Sci.* **113**, 3127–3139.

Exocytosis

Robert D. Burgoyne and Alan Morgan
The Physiological Laboratory, University of Liverpool, Crown Street, Liverpool, UK

1 Constitutive and Regulated Exocytosis 307

2 Stages in the Exocytotic Pathway 307
2.1 Priming Before Membrane Fusion 309
2.2 Fusion Pores 310
2.3 Full Fusion Versus Kiss-and-run Exocytosis 311

3 The Protein Machinery for Exocytosis 312
3.1 SNARE Proteins and Membrane Fusion 312
3.2 SNARE Regulators 314
3.2.1 NSF and α-SNAP 314
3.2.2 Sec1 Proteins 314
3.3 Rab GTPases and Their Effectors 315
3.4 A Model for the Events Leading to Membrane Fusion 317

4 Calcium Signals and Exocytosis 317
4.1 Calcium-dependency of Regulated Exocytosis 318
4.2 Calcium-dependency of Synaptic Vesicle Exocytosis 319
4.3 Spatial Relationship of Channels and Vesicles 319
4.4 Calcium and Exocytosis in Nonsecretory Cells 319

5 The Calcium Sensor for Regulated Exocytosis 320

6 Conclusions 322

Bibliography 322
Books and Reviews 322
Primary Literature 323

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Keywords

Ca^{2+} Sensor
Protein that detects changes in intracellular Ca^{2+} concentration and controls regulated exocytosis.

Constitutive Exocytosis
Exocytosis involving fusion of new secretory vesicles with the plasma membrane immediately after their synthesis.

Fusion Pore
Aqueous connection between the cell exterior and the inside of a secretory vesicle formed when membrane fusion occurs between the vesicle and plasma membranes.

Regulated Exocytosis
A form of exocytosis in which secretory vesicles are stored within the cells until the appropriate stimulus is generated to trigger their fusion with the plasma membrane.

Secretory Vesicle
Membrane-bounded vesicle containing secretory products or newly synthesized plasma membrane components.

SNARE Proteins
Proteins of the family of α-SNAP receptors that assemble to form a SNARE complex and function in vesicle docking and/or fusion.

Synapse
Junction between the axonal ending of a neuron and another neuron or target cell.

Synaptic Vesicle
Specialized neurotransmitter containing vesicle found in a nerve terminal.

Exocytosis is the process by which secretory vesicles fuse with the plasma membrane of cells to allow release of their contents and insertion of new membrane components. It is a mechanism that is essential for cell growth but also underlies the release of many different extracellularly-acting substances. All cells possess a pathway for constitutive exocytosis and many cell types also possess a pathway for regulated exocytosis that occurs only in response to an intracellular signal such as a rise in cytosolic calcium concentration. Regulated exocytosis is involved in the release of hormones and neurotransmitters and even in events involved in the fusion of the sperm and the egg. In all cases a common conserved machinery is resposible for membrane fusion during exocytosis.

1 Constitutive and Regulated Exocytosis

Exocytosis, the fusion of secretory vesicles with the plasma membrane, occurs in virtually all cells. This is required for the insertion of new plasma membrane lipids and integral membrane proteins such as receptors and ion channels. In addition, exocytosis is the major mechanism by which extracellular proteins are secreted from cells, important examples being antibodies and serum albumin. Such proteins are synthesized on the rough endoplasmic reticulum and delivered to the plasma membrane via the secretory pathway. This comprises a series of sequential membrane trafficking steps, each mediated by the selective targeting and fusion of membrane-bound carriers with appropriate organelle membranes (Fig. 1). As with membrane fusion events earlier in the secretory pathway, exocytosis can follow vesicle formation, without obvious delay or regulation. This ubiquitous process is known as *constitutive exocytosis*. In many cell types, however, a second pathway also exists in which exocytosis is tightly regulated to allow for the controlled release of vesicle contents, or triggered insertion of new membrane components, from preformed secretory vesicles. This process, known as *regulated exocytosis*, occurs only in response to a physiological signal, which in many cell types is a rise in cytosolic Ca^{2+} concentration. Various proteins are secreted by cells via regulated exocytosis, for example, digestive enzymes from exocrine cells and proteinaceous hormones from endocrine cells. In addition, small molecules can be secreted by regulated exocytosis; indeed, this is the mechanism by which neurotransmitters are released at the synapse (Fig. 2). Regulated exocytosis has other fundamental roles such as in the acrosome reaction in the sperm, which is essential for fertilization. The exocytosis of cortical granules in the egg forms the fertilization envelope and prevents further sperm fusion with the egg. The pathway for regulated exocytosis has been extensively studied in many different cell types from various organisms. For constitutive exocytosis, however, the bulk of current knowledge stems from the genetic analysis of this process in the yeast *Saccharomyces cerevisiae*.

Over the past 10 years or so, the use of biochemical, molecular biological, and genetic approaches has identified many of the proteins involved in constitutive exocytosis and in neurotransmission and other forms of regulated exocytosis. The interactions between these proteins and their relationship to membrane fusion have been studied in detail, although the full details of the exocytotic process still remain to be resolved. Similar molecular events appear to underlie all forms of exocytosis and use conserved components of a core vesicle docking/fusion machinery with specialized additions for Ca^{2+}-dependent regulation. We will describe current views on the mechanistic aspects of exocytosis and its triggering and regulation.

2 Stages in the Exocytotic Pathway

For the determination of the distinct stages of exocytosis, Ca^{2+}-regulated exocytosis has been studied in most detail as this most readily lends itself to analysis of the sequential steps involved. It has been dissected into a number of stages (Fig. 2). Regulated exocytosis involves fusion with the plasma membrane of a premade pool

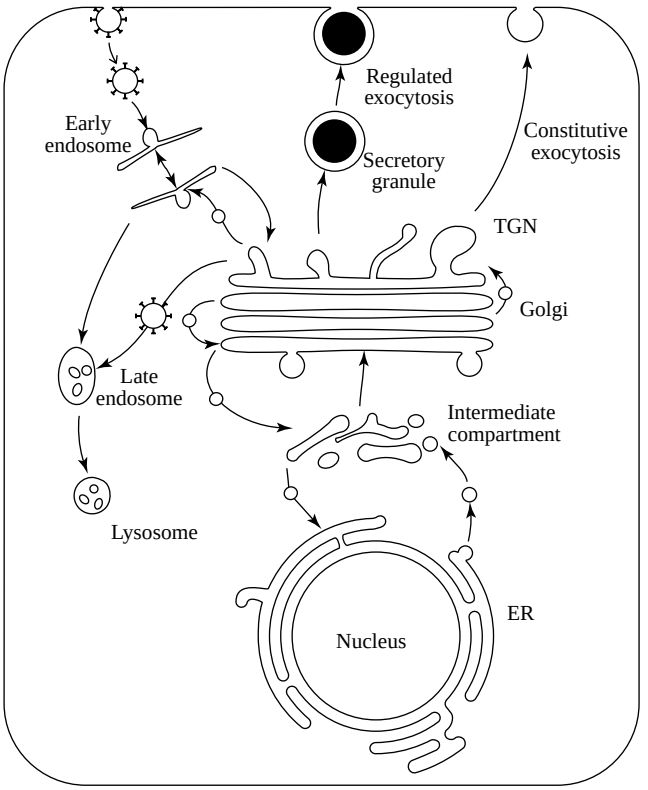

Fig. 1 Schematic representation of membrane trafficking pathways within mammalian cells. Proteins destined for the cell surface are inserted into the ER lumen and then transported forwards by vesicular traffic to the Golgi complex via the intermediate compartment. After sorting in the trans-Golgi network (TGN), proteins enter the secretory vesicle, which can then be directly transported to the cell surface for constitutive exocytosis. Alternatively, proteins are sorted into vesicles or secretory granules and stored until a Ca^{2+} signal triggers regulated exocytosis.

of vesicles. The pool of secretory vesicles may be fully releasable (as in mast cells, for example) but in many cell types (such as neuroendocrine, endocrine, and exocrine cells), a large reserve pool exists and only a small portion of the vesicles is initially releasable. In many cells, a so-called *ready-releasable* pool can undergo exocytosis within tens of milliseconds. These pools have been defined and studied through the use of permeabilized cell assays of exocytosis, single-cell electrophysiological approaches using patch-clamp capacitance measurement of cell surface area, amperometric measurement of the release of single vesicle contents, or by imaging exocytosis through the use of evanescent wave microscopy. The majority of work on synaptic vesicle exocytosis has relied on overall measurement of synaptic transmission based on a simultaneous pre- and postsynaptic recording. Alternatively,

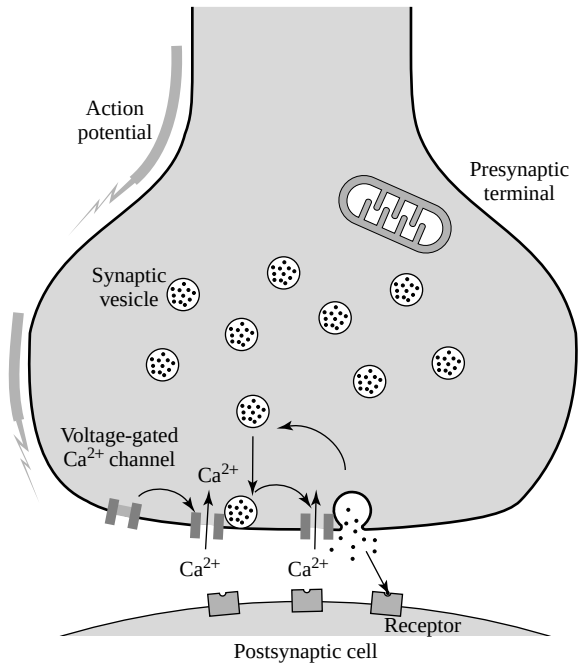

Fig. 2 Neurotransmitter release by exocytosis at the synapse. Voltage-gated Ca^{2+} channels are opened as an action potential reaches the presynaptic nerve terminal leading to Ca^{2+} entry. Synaptic vesicles docked in close proximity to the Ca^{2+} channels then undergo exocytosis within a submillisecond time course to release the stored neurotransmitter.

the synaptic vesicle cycle within synaptic terminals has been studied using uptake and release of the lipophilic FM dyes that report on plasma membrane area and changes following exocytosis and membrane retrieval.

The sequential stages within the exocytotic pathway have been shown to involve ATP-dependent priming steps, movement of vesicles to the vicinity of the plasma membrane, tethering, and then docking at release sites on the plasma membrane, conversion to a fully releasable state, triggered membrane fusion, release of vesicle contents, and finally, retrieval of the vesicle membrane (Fig. 3). In the case of Ca^{2+}-triggered exocytosis, Ca^{2+} may stimulate multiple early stages as well as fusion itself.

2.1 Priming Before Membrane Fusion

The term *priming* was used to describe an ATP-dependent process, demonstrated in permeabilized adrenal chromaffin cells that preceded Ca^{2+}-triggered fusion and increased the extent of exocytosis due to a Ca^{2+} trigger. Priming and membrane fusion were subsequently shown to be stimulated by distinct cytosolic proteins. ATP-dependent priming encompasses multiple events including reorganization of the actin cytoskeleton to allow recruitment

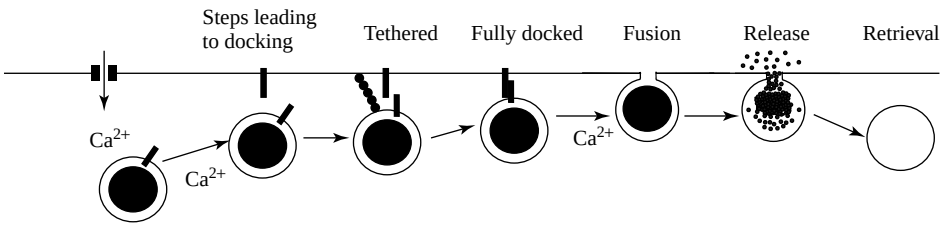

Fig. 3 Stages in the exocytotic pathway. Secretory vesicles are recruited to the plasma membrane where they undergo sequential tethering and then docking so that they become ready for exocytosis. Increased Ca^{2+} concentration can trigger this vesicle recruitment and also fusion of docked vesicles. After exocytosis and release of vesicle contents, vesicles are retrieved back into the cell.

of vesicles to the plasma membrane and modification of the so-called SNARE proteins by α-SNAP and the ATPase NSF. In addition, ATP is also required in a housekeeping role to maintain the levels of polyphosphoinositides required for continued exocytosis. It is well established that membrane fusion itself does not require ATP hydrolysis. The term priming has also been applied to an ill-defined step in synaptic vesicle maturation that may not be ATP-dependent. Within synapses, a pool of synaptic vesicles are described as *docked* as they seem to be very close to the presynaptic plasma membrane by electron microscopical examination. Only a small proportion (around 1–10%) of these vesicles are able to undergo exocytosis in response to Ca^{2+} elevation following a single action potential. This pool of docked vesicles can, however, be completely released through an unknown mechanism, by treatment with hypertonic sucrose. It has been suggested that the so-called *docked* vesicles have to undergo a maturation process to become releasable. This maturation step is not understood in molecular terms but is likely to involve the assembly of the fusion machinery and appears to involve the protein Munc-13.

2.2 Fusion Pores

It has been established through biophysical measurements, that membrane fusion proceeds through the formation of a fusion pore that can open transiently. Patch-clamp measurements of plasma membrane capacitance in mast cells were the first to allow the detection of changes in cell surface area due to single granule fusion events and also demonstrated a phenomenon known as *capacitance flicker*, which suggested that reversible fusion events were occurring. In these cells, Ca^{2+} is not the trigger for exocytosis, which instead can be activated by introduction of nonhydrolyzable GTP analogs. Later, electrophysiological measurements characterized the transient fusion events in mast cells and determined the current flow through the initial fusion pore. The fusion pore can subsequently reclose or alternatively lead to a phase of fusion pore expansion and thereby full exocytosis. The rate of fusion pore expansion was found to be a regulated phenomenon and was increased by elevating intracellular Ca^{2+} concentration in rat mast cells or by activating protein kinase C in eosinophils. The concept of the fusion pore has been validated using carbon-fiber amperometry,

which detected low levels of release of contents from mast cell granules during the initial phase after fusion (during capacitance flickering) and before the postulated fusion pore expansion had occurred.

2.3 Full Fusion Versus Kiss-and-run Exocytosis

Classical electron microscopical studies by Heuser and Reese established a pathway in which synaptic vesicles fused with the presynaptic membrane, moved away from the site of fusion, and were then recycled into the nerve terminal by clathrin-mediated endocytosis. An alternative pathway, later termed *kiss-and-run* fusion, was originally suggested for synaptic vesicle exocytosis and may be related to fusion pore formation and its reversibility. It was postulated that neurotransmitter release from a synaptic vesicle could occur very rapidly without the need for full fusion and flattening of the vesicle membrane into the plasma membrane, but instead by a reversible process that would rapidly retrieve the vesicle membrane for reuse after a transient fusion event. It has been predicted that the contents of a synaptic vesicle could be released on a submillisecond timescale through the type of fusion pore detected in mast cells and described subsequently in other secretory cell types including chromaffin cells, eosinophils, and neutrophils. It has become possible to resolve individual fusion events due to synaptic vesicle exocytosis in a nerve terminal by averaging many thousands of single vesicle fusion events. The data obtained showed that at low stimulus levels, exo/endocytosis was completed on a very short timescale with retrieval having an average time constant of 56 ms consistent with the existence of kiss-and-run exocytosis occurring under these conditions. The significance of kiss-and-run exocytosis for the regulation of neurotransmitter release is a controversial issue. Catecholamine release by kiss-and-run has been directly shown to occur in chromaffin cells using combined capacitance and amperometric recording and based on the trapping of extracellular markers in vesicles after exocytosis.

There is accumulating evidence to suggest that fusion pore closure may be a regulated process that would allow exocytosis to be switched from full fusion to a predominance of kiss-and-run fusion. While kiss-and-run fusion is likely to be sufficient for complete emptying of the synaptic vesicle, large secretory granules may be only partially emptied through a transient fusion process allowing regulation of the amount of release per fusion event. A variety of conditions were found to modify the amount released per granule from chromaffin cells and result in changes in the kinetics of single secretory events consistent with alterations in fusion pore dynamics. These changes could be a consequence of opening and rapid reclosure of a fusion pore, or more likely, reclosure after a period of fusion pore expansion to allow release of secretory vesicle contents. These effects have been seen after phobol ester treatment to activate protein kinase C, modification of intracellular Ca^{2+} levels, nitric oxide application, or following elevation of cAMP levels. Effects on exocytosis kinetics have also been observed in cells overexpressing cysteine string protein (Csp), synaptotagmin, a Munc18 mutant (R39C), a SNAP-25 mutant, or after introduction of anti-CAPS antibodies. In addition, changes in release kinetics, consistent with premature closure of the expanded fusion pore, have

been seen in cells overexpressing complexin, a protein that interacts only with the assembled SNARE complex. A mutation of a residue in complexin that is involved in this interaction prevented the effects on release kinetics. Further work will be needed to explain in detail how fusion pore dynamics are controlled during membrane fusion. A major mechanistic issue for fusion pore reversibility, and thereby, kiss-and-run exocytosis is how fusion pore closure occurs. There are two extreme possibilities. In the first model, the exocytotic machinery that produces fusion acts in a fully reversible fashion. In the second model, the fission machinery responsible for pinching off clathrin-coated endocytotic vesicles on the basis of the action of mechanoenzyme dynamin would be capable of functioning on a millisecond timescale to pinch off fused vesicles. It is also possible that both mechanisms could be operative.

3
The Protein Machinery for Exocytosis

A combination of experimental approaches has resulted in the identification of the protein machinery for exocytosis. From a convergence of biochemical assays and yeast genetics, it became apparent that exocytosis involves a core machinery that functions in both constitutive and regulated exocytosis and that this is highly conserved throughout evolution. The first protein identified using a biochemical approach was *N*-ethylmaleimide-sensitive fusion protein (NSF), which was purified on the basis of its ability to reconstitute intra-Golgi transport after blockade by *N*-ethylmaleimide (NEM). Sequencing revealed that NSF is a homolog of the yeast protein Sec18, which was already known to be essential for membrane traffic. NSF was subsequently found to be involved in regulated exocytosis. Knowledge of the mechanisms involved in exocytosis were advanced by the revelation that NSF and its cofactor, α-soluble NSF attachment protein (α-SNAP), interact with brain membranes via a SNAP receptor (SNARE) complex shown to comprise VAMP/synaptobrevin, syntaxin1, and SNAP-25 with homologs of these proteins being required for secretion in yeast. Genetic approaches have shown the essential nature of the homologs of these proteins in constitutive exocytosis in yeast. The synaptic SNARE proteins are proteolytic substrates for the clostridial neurotoxins, which are potent inhibitors of neurotransmission, thus suggesting that regulated exocytosis at the synapse occurs by the same basic process as constitutive membrane fusion. The idea of a universal fusion machinery is now generally accepted, although the functions of the individual components of the conserved core machinery and the precise mechanism of membrane fusion are still controversial.

3.1
SNARE Proteins and Membrane Fusion

The conserved core vesicle docking/fusion machinery consists of the SNARE proteins, which assemble to form the so-called SNARE complex (Fig. 4) between the vesicle and plasma membrane. Although originally defined biochemically as SNAP receptors, most SNAREs have been identified by possession of a characteristic sequence known as the SNARE motif. Searching for similar motifs in the human genome revealed 35 SNAREs, as compared to 22 in the yeast genome. Despite the large number of SNARE proteins, the synaptic SNARE complex remains the

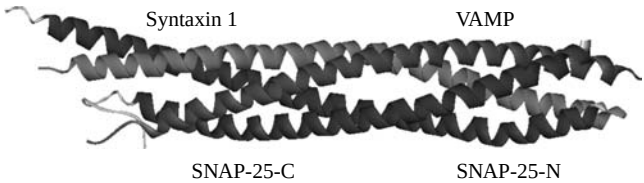

Fig. 4 The structure of the SNARE complex. The structure of the neuronal SNARE complex consisting of the cytoplasmic domains of syntaxin 1A, VAMP2, and SNAP-25 was determined by X-ray crystallography.

most intensively studied and most current ideas of SNARE function are based on this particular complex. It comprises three proteins: the integral membrane proteins, VAMP/synaptobrevin and syntaxin 1, and the membrane associated protein, SNAP-25. VAMP is a synaptic vesicle protein, whereas syntaxin and SNAP-25 are most enriched in the presynaptic plasma membrane. These localizations led to the classification of SNAREs as vesicle- (v-) or target- (t-)SNAREs, and the SNARE hypothesis suggested that the fidelity of vesicle docking and fusion is determined by the specificity of v-SNARE : t-SNARE interactions. More recently, a rival classification system has been proposed based on whether a conserved glutamine (Q) or arginine (R) is present in the SNARE motif in the so-called *zero layer* observed in the structure of the neuronal SNARE complex. The Q and R residues themselves, however, are not essential for membrane fusion but instead for the later disassembly of the SNARE complex by NSF and SNAPs.

The idea that SNARE complexes are essential for membrane fusion in living cells is well supported by genetic studies in yeast and higher organisms. Genetic ablation of synaptic SNAREs in *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Mus musculus* abolishes evoked neurotransmission. It is interesting to note, however, that the effect of inactivating syntaxins tends to be more dramatic than when other SNAREs are blocked. The use of clostridial neurotoxins has also provided important evidence for an essential role of SNAREs in exocytosis in various cell types. The key discovery in this regard was that tetanus and botulinum-B neurotoxin are metalloendoproteases that block neurotransmission presynaptically by selectively cleaving VAMP. It was then established that the remaining six botulinum neurotoxins cleaved either VAMP, syntaxin 1, or SNAP-25. Clostridial neurotoxins have been applied to many permeabilized secretory cells. Not all forms of exocytosis are blocked by clostridial neurotoxins but the most likely explanation for this is that exocytosis in such cells proceeds via toxin-resistant SNARE isoforms.

Several lines of evidence suggest that SNAREs participate in and may directly drive membrane fusion. First, tetanus toxin application results in the accumulation of docked vesicles at the synapse, suggesting that SNAREs function after morphological vesicle docking. Second, formation of the SNARE complex is energetically favorable (assembled SNARE complexes have extremely high thermal stability and require NSF and α-SNAP for disassembly). Third, the structure of the synaptic SNARE complex is such that formation of its coiled-coil bundle (Fig. 4)

would bring the transmembrane anchors of VAMP and syntaxin (and hence the vesicle and plasma membrane) into close apposition. Fourth, membrane fusion between artificial liposomes has been reconstituted using only recombinant SNAREs. These data have been taken as suggesting that the SNAREs form the core fusion machinery. This model, although popular, is not without its critics, however, as SNARE-mediated fusion is many orders of time slower than exocytosis. In addition, one controversial theory, based on studies of yeast vacuole fusion, contends that SNARE complex formation is an early event that acts as a signal to initiate an entirely different fusion mechanism that is ultimately mediated by vacuolar ATPase proteolipids. Clearly, further work is required for a full understanding of the events that lead to membrane fusion.

3.2 SNARE Regulators

3.2.1 NSF and α-SNAP

NSF and α-SNAP are soluble proteins that act to disassemble SNARE complexes. Due to the high stability of the SNARE complex, disassembly requires significant energy input in the form of ATP hydrolysis. This is provided by the ATPase NSF, and SNAP stimulation of the ATPase activity of NSF is essential for SNARE complex disassembly. It is thought that the conformational changes in NSF induced by ATP hydrolysis are transduced through α-SNAP to unwind the four helices of the SNARE complex.

The first evidence for a requirement of NSF and SNAPs in membrane fusion *in vivo* came from the yeast Sec17 and Sec18 mutants, which encode α-SNAP and NSF respectively, and are essential for all membrane trafficking steps. A role for NSF in regulated exocytosis *in vivo* was revealed by studies of the *comatose* (NSF) mutants in *D. melanogaster*. These mutants show a temperature-sensitive block in presynaptic neurotransmission. The mutant organisms accumulate SNARE complexes at the restrictive temperature, confirming that NSF is required to disassemble SNAREs in living cells. Questions remained, however, over the precise point in the membrane fusion process at which NSF and α-SNAP act. The use of stage-specific exocytosis assays in permeabilized secretory cells and high-resolution electrophysiological approaches have established that α-SNAP/NSF act in an early ATP-dependent priming reaction that precedes the later Ca^{2+}-dependent fusion step. Furthermore, this priming reaction is likely to be the disassembly of SNARE complexes, since mutant α-SNAPs unable to stimulate NSF ATPase activity and consequent SNARE disassembly cannot support exocytosis. It seems that α-SNAP/NSF act as molecular chaperones, continually disassembling SNARE complexes in order to release SNARE proteins to engage and function in further membrane fusion events.

3.2.2 Sec1 Proteins

Much attention has been focused on NSF, SNAPs, and SNAREs, but these are not the only evolutionarily conserved proteins required for intracellular membrane traffic. The first yeast secretory mutant to be identified harbored a mutation in Sec1. Genetic studies have established a general requirement for Sec1-related proteins in vesicle fusion in a wide variety of organisms. Sec1 family members have been shown to interact with syntaxin homologs, suggesting that the conserved function of Sec1 proteins in membrane fusion may be mediated via syntaxins. Indeed, the inhibition of neurotransmitter release in *Drosophila*

caused by overexpression of ROP is relieved by co-overexpression of syntaxin. Neuronal Sec1 (also known as Munc18) binds to syntaxin1a with high affinity *in vitro* to form a complex that precludes syntaxin binding to synaptic SNAREs. These observations have led to the suggestion that Sec1 proteins act as syntaxin chaperones, preventing SNARE complex assembly until signaled to release syntaxin. Although the precise nature of this signal is unknown, various observations suggest that it is linked to the completion of Rab-dependent vesicle tethering.

The proposed role of Sec1 proteins as negative regulators of syntaxin function is too simplistic as loss of Sec1 family members results in a complete block in membrane traffic, indicating an essential positive role. Studies of nSec1/Munc18 in PC12 cells suggest that, rather than merely preventing syntaxin entry into the core complex, the syntaxin–Sec1 interaction is an essential intermediate that enables activation of syntaxin for efficient incorporation into the SNARE complex. This notion is consistent with recent studies in chromaffin cells on the effects of a mutant form of nSec1/Munc18 (R39C) with a reduced affinity for syntaxin. Here, it was found that the mutant resulted in an increased rate of fusion pore expansion, which could result from the increased release of activated syntaxin into the fusion machinery. In contrast to neuronal Sec1, yeast Sec1will not bind efficiently to its cognate syntaxin (Sso1/2) in isolation, but only when it is assembled into a SNARE complex. This finding suggests that Sec1 functions during the membrane fusion process, acting at or after SNARE assembly consistent with a second effect of the Munc18 mutant to cause rapid closure of the fusion pore. In contrast to these models, studies of chromaffin cells from Munc18-1 knockout mice have revealed a major defect in granule docking, suggesting a much earlier role for this protein. Clearly, Sec1 proteins have essential and possibly multiple functions in physiological membrane fusion.

3.3 Rab GTPases and Their Effectors

Rab proteins comprise a large family of small GTPases. Individual Rab proteins are localized to distinct organelles, suggesting that each regulates a particular membrane traffic step. Rabs are thought to act as molecular switches determining how long vesicles remain competent for docking/fusion, before GTP hydrolysis. The initial findings implicating these GTPases in membrane traffic came from the identification of the Sec4 mutation which blocks constitutive exocytosis in yeast. The general consensus is that Rabs are major determinants of the specificity of membrane transport, and achieve this by selective interactions with distinct effector proteins. The best-characterized Rab effectors are involved in the initial tethering of vesicles to target membranes (Fig. 5). These include the exocyst complex (Sec4 effector in exocytosis in yeast) and EEA1 (Rab5 effector in endocytosis). Another class of Rab effectors is the motor proteins that act to control organelle movement within cells. For example, Rab5 controls movement of endosomes on microtubules and Rab6 binds to the kinesin-like protein, rabkinesin-6. In addition to microtubule-based motors, Rabs have also been implicated in actin-based motility by virtue of functional interactions with class V myosins. Finally, genetic (and in some cases, physical) interactions between various Sec1 homologs and Rabs suggest that

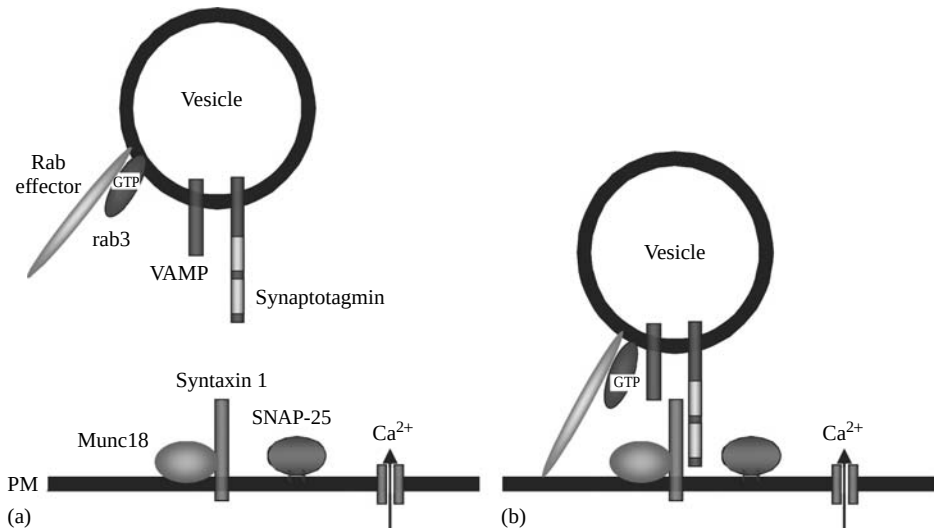

Fig. 5 Localization of proteins involved in exocytosis and role of Rab3 in tethering. The diagram shows the distribution of synaptic proteins on synaptic vesicles and the presynaptic membrane. When Rab3 becomes GTP-loaded (a), it recruits effectors that allow tethering of vesicles, (b) prior to engagement of the SNARE proteins.

the Sec1 family may represent a third distinct class of Rab effectors. However, a variety of other proteins that do not clearly segregate into these classes have also been proposed as Rab effectors, emphasizing the potential complexity of Rab function(s) in the cell.

Rab3 exists as four isoforms A–D in mammals and is the major Rab implicated in exocytosis in neurons and neuroendocrine cells. Rab3A is not an essential protein in mammals but may regulate neurosecretion by limiting the recruitment of synaptic vesicles into the releasable pool at the plasma membrane for subsequent fusion events. Rab3 mutants in *C. elegans* are also viable but have less synaptic vesicles close to the presynaptic membrane. Surprisingly, overexpression of active forms of Rab3A or other Rab3 isoforms has been found to inhibit exocytosis in various cell types. Several Rab3A effector proteins including RIM and Rabphilin 3A are involved in secretory vesicle exocytosis. Analysis of specific mutations that disrupt interaction of Rab3 with these effectors has excluded them, however, as the inhibitory effectors of Rab3. The inhibitory effects of Rab3 may instead be exerted via the effector Noc2, a protein expressed predominantly, if not exclusively, in neuroendocrine cells. Other studies have suggested that certain Rab3 isoforms may, in contrast, exert positive effects on exocytosis in some cell types. It is possible that Rab3 has multiple roles in exocytosis. Despite the suggestions that different Rab3 isoforms have distinct functions, a direct test of this possibility shows that the four Rab3A–D isoforms all had the same inhibitory effect after overexpression in PC12, chromaffin, and insulin-secreting cell lines. In addition, more recent work has established an important role for Rab27 in exocytosis in melanocytes and in cytotoxic T

lymphocytes and also in dense-core granule exocytosis in endocrine cells.

3.4 A Model for the Events Leading to Membrane Fusion

Figure 6 shows a model that has developed from the studies described above on the interactions of the SNARE proteins and their regulatory proteins. In this scheme, syntaxin is held in its closed conformation by Munc18. In some ill-defined way, possibly due to the action of a Rab/Rab effector, Munc18 is dissociated from syntaxin. Syntaxin then adopts its open confirmation, although Munc18 may remain present due to alternative interactions with the N-terminus of syntaxin. The SNARE helix of syntaxin, which is now exposed, acts as a nucleation site for interaction with the VAMP SNARE domain and the two helices of SNAP-25. Formation of the full complex then proceeds down its energy gradient, potentially releasing sufficient energy to overcome the hydration barrier and initiate bilayer fusion. Munc18 appears to be able to exert effects on the kinetics of the fusion process. After fusion, the SNARE complex is disassembled by α-SNAP and NSF, and Munc18 reforms its tight complex with the closed form of syntaxin.

4 Calcium Signals and Exocytosis

A number of different intracellular signals are involved in triggering regulated exocytosis. The most widespread and most

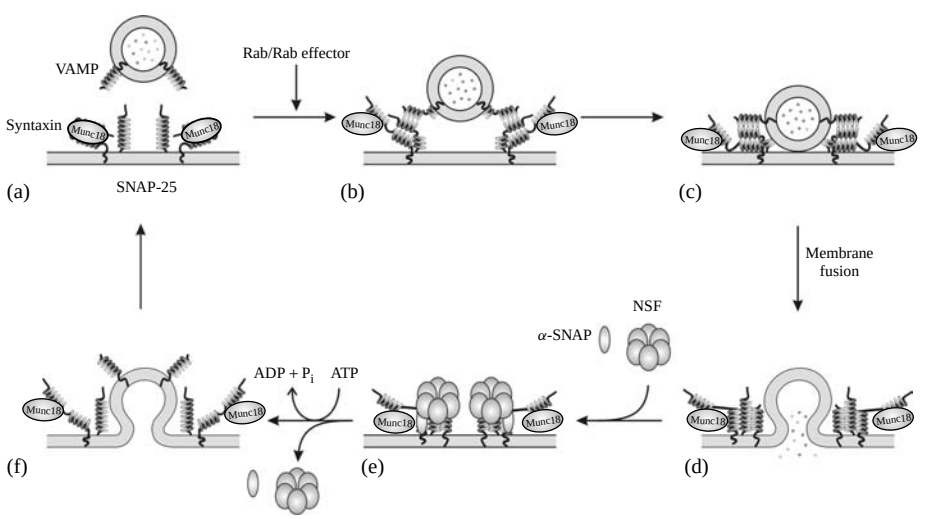

Fig. 6 Model of the protein–protein interactions to lead to and follow exocytosis. (a) Syntaxin is initially held in a closed conformation by Munc18, which prevents SNARE complex assembly. (b) Tethering due to the action of Rab/Rab effector leads to release of syntaxin by Munc18 into its open conformation and SNARE complex assembly begins. (c) Full assembly of the SNARE complex pulls the vesicle into proximity with the plasma membrane and (d) leads to membrane fusion. (e) After fusion, SNARE complexes bind α-SNAP and NSF and (f) these are disassembled following ATP hydrolysis by NSF. Munc18 can then interact with syntaxin to reform its closed conformation. Munc18 exerts additional effects during membrane fusion through as-yet uncharacterized interactions.

intensively studied is cytosolic Ca^{2+} concentration. In many cell types, regulated exocytosis is triggered by an elevation of cytosolic Ca^{2+} concentration from a resting level of 100 nM to micromolar concentrations due to Ca^{2+} entry at the plasma membrane or mobilization from intracellular stores. The limited diffusion of Ca^{2+} due to high levels of Ca^{2+} buffering within cells limits the distance over which Ca^{2+} can diffuse and exert its effects. Spatial distribution may result in only particular types of Ca^{2+} channels being associated with exocytosis and this may be enhanced by physical association of vesicles with Ca^{2+} channel proteins, as is believed to occur for P/Q- and N-type channels in synapses. The increased importance of Ca^{2+} stores in pancreatic exocrine cells and pituitary gonadotrophs appears to be determined by the spatial localization of ER Ca^{2+} stores in these cell types.

4.1 Calcium-dependency of Regulated Exocytosis

Considerable effort has been expended in the determination of the Ca^{2+} sensitivity of the exocytotic machinery in a wide range of cell types. Two major technical approaches have been used. Firstly, cell population measurements of granule release following cell permeabilization with direct addition of Ca^{2+}, and secondly, single-cell patch-clamp capacitance recording to follow the kinetics of exocytosis. Katz and Miledi established the importance of Ca^{2+} by showing a requirement for external Ca^{2+} for the release process at the neuromuscular junction. This was subsequently extended by Bill Douglas to chromaffin cells and posterior pituitary neurosecretory terminals. Assessment of the Ca^{2+}-dependency of exocytosis required, however, the means to control the internal Ca^{2+} concentration. This was first achieved by the use of electropermeabilization of sea urchin eggs and adrenal chromaffin cells and application of buffered Ca^{2+} solutions to the leaky cells. This revealed that granule exocytosis in these cells was half-maximally activated at 1 μM free Ca^{2+}. The similar Ca^{2+}-dependency of fusion in these two cell types suggested the possible existence of a common Ca^{2+} sensor mechanism in exocytosis. In the majority of cases, the use of permeabilized cells has shown a Ca^{2+} requirement for triggering of exocytosis in the low micromolar range. It later became apparent, however, that multiple Ca^{2+}-dependent steps exist in the exocytotic pathway.

A value for the Ca^{2+}-dependency of the final fusion step in regulated exocytosis in endocrine and other secretory cells was subsequently determined by patch-clamp measurement of plasma membrane capacitance. Membrane capacitance is directly proportional to the area of the lipid bilayer and so this technique allows quantitative assay of membrane insertion (exocytosis) or membrane removal (endocytosis) from the cell surface. In early studies, exocytosis was triggered by depolarization to open voltage-gated Ca^{2+} channels or by infusion of Ca^{2+}-buffers from the recording micropipette by intracellular dialysis. Accurate determination of the Ca^{2+} requirement of the final fusion step was achieved by using flash photolysis to break downcaged Ca^{2+} compounds and to give a homogenous elevation of free Ca^{2+} coupled with capacitance recording. This approach allowed an assessment of the relationship between free Ca^{2+} concentration and the initial rate of exocytosis from the ready-releasable pool. Such experiments, carried out in a range of endocrine

cell types, demonstrated rapid exocytosis with a Ca^{2+}-dependency in the 2 to 30 μM range.

4.2 Calcium-dependency of Synaptic Vesicle Exocytosis

Use of the flash photolysis/capacitance method on a specialized neuron, the retinal bipolar neuron, suggested that synaptic vesicle exocytosis involved a distinct low-affinity Ca^{2+} sensor. In this giant synapse, exocytosis was not detected until the Ca^{2+} concentration was >10–20 μM and was half-maximal at 190 μM free Ca^{2+}, a 10-fold higher Ca^{2+} concentration than for endocrine cell exocytosis. More recently, measurement of neurotransmission at a central mammalian synapse (the calyx of Held) has determined a Ca^{2+}-dependency in the low micromolar range (2–5 μM) comparable to that for endocrine cell exocytosis. These findings were interpreted as suggesting that the earlier value found in retinal bipolar neurons might not reflect the general situation in mammalian CNS synapses. Earlier data in support of a low-affinity receptor for synaptic vesicle exocytosis in the squid giant synapse included the demonstration of microdomains of very high Ca^{2+} during neurotransmitter release. In addition, exocytosis in this synapse was inhibited by the fast Ca^{2+} chelator BAPTA but not by the slower Ca^{2+} chelator EGTA suggesting that the sensing of Ca^{2+} by the fusion machinery must occur close to the Ca^{2+} channels. It is possible that the Held synaptic terminals are unusual in their Ca^{2+} requirements but blockade of transmission by the slow Ca^{2+} chelator EGTA in cortical synapses favors more widespread use of a low-affinity process.

4.3 Spatial Relationship of Channels and Vesicles

A further issue in understanding the relationship between Ca^{2+} and regulated exocytosis is the spatial distribution of secretory vesicles and the source of Ca^{2+}. In addition, much attention has been given to the question of how close fusion occurs to Ca^{2+} channels and the degree of coupling between Ca^{2+} entry and exocytosis. For synaptic vesicles, it is generally agreed that the rapid speed of neurotransmission, which can be as fast as 60 μs in central synapses, must mean that the vesicles that undergo fusion sense Ca^{2+} that has diffused no further than a few nanometers from the mouth of the open Ca^{2+} channel. This must also be the case if synaptic vesicle exocytosis involves a low-affinity Ca^{2+}-sensor. Molecular interactions that have been demonstrated between proteins of the exocytosis machinery and Ca^{2+} channel subunits suggest that synaptic voltage-gated Ca^{2+} channels involved in neurotransmission are physically part of the exocytotic site. The extent of coupling to Ca^{2+} channels has been examined for endocrine granule exocytosis in adrenal chromaffin cells and it has been suggested that a tight physical coupling between Ca^{2+} channels and secretory granules is unlikely to occur in these cells.

4.4 Calcium and Exocytosis in Nonsecretory Cells

The application of flash photolysis of caged Ca^{2+} compounds and patch-clamp capacitance recording to cultured fibroblastoid cell lines surprisingly showed that a Ca^{2+}-regulated pathway for exocytosis

existed even in these cells. Coupled with the demonstration that muscle myoblasts could take up and release acetylcholine in response to stimulation, this suggests that all cell types might possess a regulated exocytotic pathway. One possibility is that this pathway involves Ca^{2+}-dependent exocytosis of lysosomes as it has been demonstrated that even in nonsecretory cells, conventional lysosomes can be triggered to undergo exocytosis.

5 The Calcium Sensor for Regulated Exocytosis

The major role of Ca^{2+} as the trigger for regulated exocytosis predicts the existence of a conserved protein(s) capable of activating the SNARE-related fusion machinery upon binding Ca^{2+}. Several different candidate proteins have been put forward, but it now seems clear that synaptotagmin is the major Ca^{2+} sensor for regulated exocytosis. Originally discovered over 20 years ago as a synaptic vesicle membrane protein, termed p65, interest in this protein increased dramatically with the demonstration that it was a Ca^{2+}-dependent phospholipid-binding protein. This property of synaptotagmin is due to its two C_2 domains, C_2A and C_2B (Fig. 7), which are homologous to the C_2 domain in protein kinase C. Considerable variation in the Ca^{2+} dependence of phospholipid binding is exhibited by different synaptotagmin isoforms, with the plasma membrane synaptotagmins 3 and 7 displaying much higher Ca^{2+} affinity (EC_{50} of 1–2 μM) than the vesicular synaptotagmins 1 and 2 (EC_{50} of 10–20 μM). In addition, synaptotagmins have been shown to interact with a variety of proteins implicated in exocytosis, including Ca^{2+} channels, syntaxin 1, the SNARE complex, and β-SNAP. Synaptotagmin I can accelerate liposome fusion mediated by SNAREs but in a Ca^{2+}-independent manner. As with phospholipid binding, individual synaptotagmins display marked differences in the Ca^{2+}-dependency of syntaxin binding, with >200 μM Ca^{2+} required for maximal binding by synaptotagmins 1 and 2, in contrast with <10 μM Ca^{2+} for synaptotagmins 3 and 7. Recent work suggests that the key Ca^{2+}-dependent interaction for exocytosis involving synaptotagmin is with phospholipids.

Synaptotagmins may be the ubiquitous exocytotic Ca^{2+} sensors. Evidence to support this view comes from genetic studies in *Drosophila*, *C. elegans*, and mice, where synaptotagmin mutants display impaired synaptic transmission. In addition, mice with a point mutation (R233Q) reducing the affinity of synaptotagmin 1 for Ca^{2+}-dependent phospholipid binding display a decrease in the Ca^{2+} affinity of synaptic vesicle exocytosis. This strongly suggests that synaptotagmin 1 is a major Ca^{2+} sensor for synaptic vesicle exocytosis. A further aspect of synaptotagmin that has emerged is its ability to influence the kinetics of single granule release events in PC12 cells, suggesting that it may function during membrane fusion at the level of the fusion pore. The significance of Ca^{2+} sensing by either the C2A or C2B domains of synaptotagmin has been controversial. As noted above, the knock-in study on the R233Q mutant in mice suggested the importance of the C2A domain for the Ca^{2+} sensitivity of exocytosis. Detailed genetic analyses in *Drosophila* have shown, however, that Ca^{2+}-binding to the C2A domain is dispensable for neurotransmission and survival but that Ca^{2+}-binding to the C2B domain is crucial. Whilst it is clear that synaptotagmin plays a major role

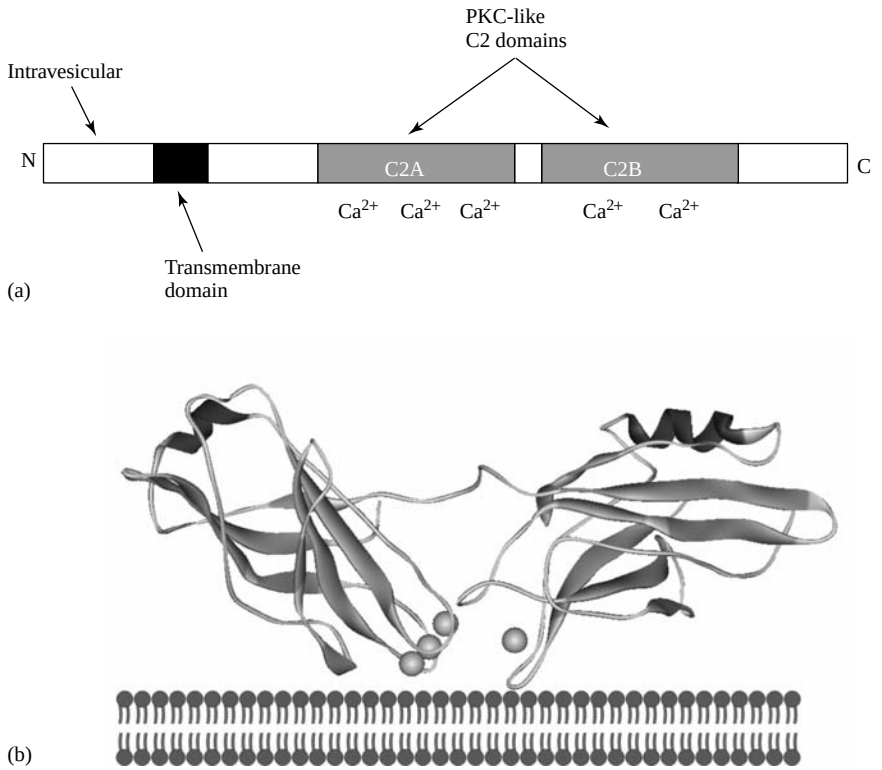

Fig. 7 Structure of synaptotagmins. (a) Domain organization of the synaptotagmins. (b) Structure of the cytoplasmic domain of synaptotagmin III. Synaptotagmin contains two C_2 domains with divalent cation binding sites occupied in the crystal by three (C2A, left) or one (C2B, right) Mg^{2+} ions. The side of the molecule that binds Ca^{2+} or Mg^{2+} also interacts with phospholipids.

in regulated exocytosis, how it contributes and when it acts during the steps leading to membrane fusion are unknown. In some models, synaptotagmin has a negative role in preventing SNARE-mediated fusion until the Ca^{2+} signal appears. In others, synaptotagmin has been proposed to play an active part in membrane fusion. Clearly, much is still to be learnt about this already extensively studied protein.

Other Ca^{2+} sensors may also contribute to the regulation of exocytosis. The C_2 domains present in synaptotagmin are also shared by protein kinase C, rabphilin 3A, Doc2, RIM, and munc13, all of which have been suggested to play a role in exocytosis. Ca^{2+} binding proteins that lack C_2 domains have also been implicated in granule exocytosis. Calmodulin binds Ca^{2+} via its four EF-hand domains and was one of the first proteins suggested to be involved in exocytosis. More recently, Ca^{2+}-calmodulin has been invoked as a universal requirement for the membrane fusion process, based on observations from several constitutive membrane fusion events. However, in some cell types, exocytosis is not blocked by calmodulin antagonists and so a regulatory rather than an essential role for this protein in exocytosis is likely.

6 Conclusions

The study of exocytosis for many years in a wide range of cell types has resulted in extensive characterization of the physiological regulation of secretion. For a long time, it was suspected that the mechanisms underlying regulated exocytosis would be different from cell to cell and, in particular, would be distinct in synapses where exocytosis is very rapid. The last few decades or so have seen an incredible explosion in our knowledge of the many proteins that form the machinery for vesicle docking and fusion in exocytosis and that regulate the process. One of the most striking aspects has been the discovery in the early 1990s that very similar proteins play the same roles in synaptic exocytosis and in constitutive exocytosis in yeast and thus form part of a highly conserved machinery. The expression of SNARE isoforms and their regulators has now been cataloged extensively in many types of secretory cells. It is apparent, therefore, that essentially the same core machinery with identical proteins (or closely related isoforms of them) functions in exocytosis in all cell types that have been examined and that these are universally essential proteins. A similar core machinery also functions in all other intracellular fusion events underlying membrane traffic within the cell. The universality of mechanisms underlying exocytosis has greatly simplified our thinking and resulted in a concentrated analysis of favored model secretory cells. A number of issues remain to be resolved for a full understanding of exocytosis. It is not yet completely clear whether the SNAREs are the proteins that drive membrane fusion or whether other proteins do this job alone or in conjunction with the SNAREs. The behavior of the lipid bilayers during membrane fusion, fusion pore formation, and pore expansion is an area about which little is known. The way in which synaptotagmin controls regulated exocytosis and the extent to which it participates in membrane fusion is unclear. After more than a decade of considerable advances in the study of exocytosis, we are now poised, however, to answer these final questions regarding the exact mechanisms involved in the important biological process of exocytosis.

See also Phospholipids.

Bibliography

Books and Reviews

Augustine, G. (2001) How does calcium trigger neurotransmitter release? *Curr. Opin. Neurobiol.* **11**, 320–326.

Brunger, A.T. (2000) Structural insights into the molecular mechanism of Ca^{2+}-dependent exocytosis, *Curr. Opin. Neurobiol.* **10**, 293–302.

Burgess, T.L., Kelly, R.B. (1987) Constitutive and regulated secretion of proteins, *Annu. Rev. Cell Biol.* **3**, 243–293.

Burgoyne, R.D., Morgan, A. (2003) Secretory granule exocytosis, *Physiol. Rev.* **83**, 581–632.

Chen, Y.A., Scheller, R.H. (2001) SNARE-mediated membrane fusion, *Nat. Rev. Mol. Cell Biol.* **2**, 98–106.

Fernandez-Chacon, R., Sudhof, T.C. (1999) Genetics of synaptic vesicle function: Toward the complete functional anatomy of an organelle, *Annu. Rev. Physiol.* **61**, 753–776.

Misura, K.M.S., May, A.P., Weis, W.I. (2000) Protein-protein interactions in intracellular membrane fusion, *Curr. Opin. Struct. Biol.* **10**, 662–671.

Richmond, J.E., Broadie, K.S. (2002) The synaptic vesicle cycle: exocytosis and endocytosis in Drosophila and C. elegans, *Curr. Opin. Neurobiol.* **12**, 499–507.

Sudhof, T.C. (1995) The synaptic vesicle cycle: a cascade of protein-protein interactions, *Nature* **375**, 645–653.

Sudhof, T.C. (2002) Synaptotagmins: why so many? *J. Biol. Chem.* **277**, 7629–7632.

Primary Literature

Albillos, A., Dernick, G., Horstmann, H., Almers, W., Alvarez de Toledo, G., Lindau, M. (1997) The exocytotic event in chromaffin cells revealed by patch amperometry, *Nature* **389**, 509–512.

Ales, E., Tabares, L., Poyato, J.M., Valero, V., Lindau, M., Alvarez de Toledo, G. (1999) High calcium concentrations shift the mode of exocytosis to the kiss-and-run mechanism, *Nat. Cell Biol.* **1**, 40–44.

Alvarez de Toledo, G., Fernandez-Chacon, R., Fernandez, J.M. (1993) Release of secretory products during transient vesicle fusion, *Nature* **363**, 554–558.

Archer, D.A., Graham, M.E., Burgoyne, R.D. (2002) Complexin regulates the closure of the fusion pore during regulated exocytosis, *J. Biol. Chem.* **277**, 18249–18252.

Baker, P.F., Knight, D.E. (1978) Calcium-dependent exocytosis in bovine adrenal medullary cells with leaky plasma membranes, *Nature* **276**, 620–622.

Baker, P.F., Whitaker, M.J. (1978) Influence of ATP and calcium on the cortical reaction in sea urchin eggs, *Nature* **276**, 513–515.

Barnard, R.J.O., Morgan, A., Burgoyne, R.D. (1997) Stimulation of NSF ATPase activity by α-SNAP is essential for SNARE complex disassembly and exocytosis, *J. Cell Biol.* **139**, 875–883.

Bennett, M.K., Calakos, N., Scheller, R.H. (1992) Syntaxin: a synaptic protein implicated in docking of synaptic vesicle at presynaptic active zones, *Science* **257**, 255–259.

Bezprovanny, I., Scheller, R.H., Tsien, R.W. (1995) Functional impact of syntaxin on gating of N-type and Q-type calcium channels, *Nature* **378**, 623–626.

Block, M.R., Glick, B.S., Wilcox, C.A., Wieland, F.T., Rothman, J.E. (1988) Purification of an N-ethylmaleimide-sensitive protein catalyzing vesicular transport, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7852–7856.

Bock, J.B., Matern, H.T., Peden, A.A., Scheller, R.H. (2001) A genomic perspective on membrane compartment organization, *Nature* **409**, 839–841.

Bollmann, J.H., Sakmann, B., Borst, J.G.J. (2000) Calcium sensitivity of glutamate release in a calyx-type terminal, *Science* **289**, 953–957.

Breckenridge, L.J., Almers, W. (1987) Currents through the fusion pore that forms during exocytosis of a secrtory vesicle, *Nature* **328**, 814–817.

Brenner, S. (1974) The genetics of Caenorhabditis elegans, *Genetics* **77**, 71–94.

Brose, N., Hofmann, K., Hata, Y., Sudhof, T.C. (1995) Mammalian homologs of Caenorhabditis elegans unc-13 gene define novel family of C2 domain proteins, *J. Biol. Chem.* **270**, 25273–25280.

Brose, N., Petrenko, A.G., Sudhof, T.C., Jahn, R. (1992) Synaptotagmin: a calcium sensor on the synaptic vesicle surface, *Science* **256**, 1021–1025.

Burgoyne, R.D., Barclay, J.W. (2002) Splitting the quantum: regulation of quantal release during vesicle fusion, *Trends Neurosci.* **26**, 176–178.

Carr, C.M., Grote, E., Munson, M., Hughson, F.M., Novick, P.J. (1999) Sec1p binds to SNARE complexes and concentrates at sites of secretion, *J. Cell. Biol.* **146**, 333–344.

Catterall, W.A. (2000) Structure and regulation of voltage-gated Ca^{2+} channels, *Annu. Rev. Cell. Dev. Biol.* **16**, 521–555.

Ceccarelli, B., Hurlburt, W.P., Mauro, A. (1973) Turnover of neurotransmitter and synaptic vesicles at the frog neuromuscular junction, *J. Cell Biol.* **57**, 499–524.

Chamberlain, L.H., Roth, D., Morgan, A., Burgoyne, R.D. (1995) Distinct effects of α-SNAP, 14-3-3 proteins and calmodulin on priming and triggering of regulated exocytosis, *J. Cell Biol.* **130**, 1063–1071.

Cheek, T.R., Burgoyne, R.D. (1986) Nicotine evoked disassembly of cortical actin filaments in bovine adrenal chromaffin cells, *FEBS Lett.* **207**, 110–113.

Chen, X., Tomchick, D.R., Kovrigin, E., Arac, D., Machius, M., Sudhof, T.C., Rizo, J. (2002) Three-dimensional structure of the complexin/SNARE complex, *Neuron* **33**, 397–409.

Christoforidis, S., McBride, H.M., Burgoyne, R.D., Zerial, M. (1999) The Rab5 effector EEA1 is a core component of endosome docking, *Nature* **397**, 621–625.

Chung, S.-H., Joberty, G., Gelino, E.A., Macara, I.G., Holz, R.W. (1999) Comparison of the effects on secretion in chromaffin and PC12 cells of Rab3 family members and mutants, *J. Biol. Chem.* **274**, 18113–18120.

Coorssen, J.R., Schmitt, H., Almers, W. (1996) Ca^{2+} triggers massive exocytosis in Chinese hamster ovary cells, *EMBO J.* **15**, 3787–3791.

Criado, M., Gil, A., Viniegra, S., Gutierrez, L.M. (1999) A single amino acid near the C-terminus of the synaptosome-associated protein of 25 kDa (SNAP-25) is essential for exocytosis in chromaffin cells, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7256–7261.

DiAntonio, A., Parfitt, K.D., Schwartz, T.L. (1993) Synaptic transmission persists in synaptotagmin mutants of drosophila, *Cell* **73**, 1281–1290.

Douglas, W.W. (1968) Stimulus-secretion coupling: the concept and clues from chromaffin and other cells, *Br. J. Pharmacol.* **34**, 451–474.

Dulubova, I., Sugita, S., Hill, S., Hosaka, M., Fernandez, I., Sudhof, T.C., Rizo, J. (1999) A conformational switch in syntaxin during exocytosis: role of munc18, *EMBO J.* **18**, 4372–4382.

Eberhard, D.A., Cooper, C.L., Low, M.G., Holz, R.W. (1990) Evidence that the inositol phospholipids are necessary for exocytosis, *Biochem. J.* **268**, 15–25.

Elhamdani, A., Martin, T.F.J., Kowalchyk, J.A., Artalejo, C.R. (1999) Ca^{2+}-dependent activator protein for secretion is critical for the fusion of dense-core vesicles with the membrane in calf adrenal chromaffin cells, *J. Neurosci.* **19**(17), 7275–7383.

Elhamdani, A., Palfrey, H.C., Artalejo, C.R. (2001). Quantal size is dependent on stimulation frequency and calcium entry in calf chromaffin cells, *Neuron* **31**, 819–830.

Fasshauer, D., Sutton, R.B., Brunger, A.L., Jahn, R. (1998) Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15781–15786.

Fernandez, J.M., Neher, E., Gomperts, B.D. (1984) Capacitance measurements reveal stepwise fusion events in degranulating mast cells, *Nature* **312**, 453–455.

Fernandez-Chacon, R., Alvarez de Toledo, G. (1995) Cytosolic calcium facilitates release of secretory products after exocytotic vesicle fusion, *FEBS Lett.* **363**, 221–225.

Fernandez-Chacon, R., Konigstorfer, A., Gerber, S.H., Garcia, J., Matos, M.F., Stevens, C.F., Brose, N., Rizo, J., Rosenmund, C., Sudhof, T.C. (2001) Synaptotagmin I functions as a calcium regulator of release probability, *Nature* **410**, 41–49.

Fesce, R., Grohovaz, F., Valtorta, F., Meldolesi, J. (1994) Neurotransmitter release: fusion or 'kiss-and-run', *Trends Cell Biol.* **4**, 1–4.

Fisher, R.J., Pevsner, J., Burgoyne, R.D. (2001) Control of fusion pore dynamics during exocytosis by Munc18, *Science* **291**, 875–878.

Geppert, M., Bolshakov, V.Y., Siegelbaum, S.A., Takei, K., De Camilli, P., Hammer, R.E., Sudhof, T.C. (1994a) The role of Rab3A in neurotransmitter release, *Nature* **369**, 493–497.

Geppert, M., Goda, Y., Hammer, R.E., Li, C., Rosahl, T.W., Stevens, C.F., Sudhof, T.C. (1994b) Synaptotagmin I: a major Ca^{2+} sensor for transmitter release at a central synapse, *Cell* **79**, 717–727.

Goda, Y., Sudhof, T.C. (1997) Calcium regulation of neurotransmitter release: reliably unreliable? *Curr. Opin. Cell Biol.* **9**, 513–518.

Graham, M.E., Burgoyne, R.D. (2000) Comparison of cysteine string protein (Csp) and mutant α-SNAP overexpression reveals a role for Csp in late steps of membrane fusion in dense-core granule exocytosis in adrenal chromaffin cells, *J. Neurosci.* **20**, 1281–1289.

Graham, M.E., Fisher, R.J., Burgoyne, R.D. (2000) Measurement of exocytosis by amperometry in adrenal chromaffin cells: effects of clostridial neurotoxins and activation of protein kinase C on fusion pore kinetics, *Biochimie* **82**, 469–479.

Graham, M.E., Washbourne, P., Wilson, M.C., Burgoyne, R.D. (2001) SNAP-25 with mutations in the zero layer supports normal membrane fusion kinetics, *J. Cell Sci.* **114**, 4397–4405.

Hartmann, J., Lindau, M. (1995) A novel Ca^{2+} dependent step in exocytosis subsequent to vesicle fusion, *FEBS Lett.* **363**, 217–220.

Hay, J.C., Fisette, P.L., Jenkins, G.H., Anderson, R.A., Fukami, K., Takenawa, T., Martin, T.F.J. (1995) ATP-dependent inositide phosphorylation required for Ca^{2+}-activated secretion, *Nature* **374**, 173–177.

Hay, J.C., Martin, T.F.J. (1992) Resolution of regulated secretion into sequential MgATP-dependent and calcium-dependent stages

mediated by distinct cytosolic proteins, *J. Cell Biol.* **119**, 139–151.

Hayashi, T., McMahon, H., Yamasaki, S., Binz, T., Hata, Y., Sudhof, T.C., Niemann, H. (1994) Synaptic vesicle membrane fusion complex: action of clostridial neurotoxins on assembly, *EMBO J.* **13**, 5051–5061.

Haynes, L.P., Evans, G.J.O., Morgan, A., Burgoyne, R.D. (2001) A direct inhibitory role for the Rab3 specific effector, Noc2, in Ca^{2+}-regulated exocytosis in neuroendocrine cells, *J. Biol. Chem.* **276**, 9726–9732.

Heidelberger, R., Heinemann, C., Neher, E., Matthews, G. (1994) Calcium dependence of the rate of exocytosis in a synaptic terminal, *Nature* **371**, 513–515.

Heinemann, C., Ruden, L.v., Chow, R.H., Neher, E. (1993) A two-step model of secretion control in neuroendocrine cells, *Pflugers Arch.* **424**, 105–112.

Henkel, A.W., Betz, W.J. (1995) Staurosporine blocks evoked release of FM1-43 but not acetylocholine from frog motor nerve terminal, *J. Neurosci.* **15**, 8246–8258.

Heuser, J.E., Reese, T.S. (1973) Evidence for recycling of synaptic vesicle membrane during transmitter release at the frog neuromuscular junction, *J. Cell Biol.* **57**, 315–344.

Holz, R.W., Bittner, M.A., Peppers, S.C., Senter, R.A., Eberhard, D.A. (1989) MgATP-independent and MgATP-dependent exocytosis. Evidence that MgATP primes adrenal chromaffin cells to undergo exocytosis, *J. Biol. Chem.* **264**, 5412–5419.

Iezzi, M., Escher, G., Meda, M., Charollais, A., Baldini, G., Darchen, F., Wollheim, C.B., Regazzi, R. (1999) Subcellular distribution and function of Rab3A, B, C and D isoforms in insulin-secreting cells, *Mol. Endocrinol.* **13**, 202–212.

Joberty, G., Stabila, P.F., Coppola, T., Macara, I.G., Regazzi, R. (1999) High affinity Rab3 binding is dispensable for Rabphilin-dependent potentiation of stimulated secretion, *J. Cell Sci.* **112**, 3579–3587.

Katz, B., Miledi, R. (1965) The effect of calcium on acetylcholine release from motor nerve terminals, *Proc. R. Soc. London, Ser. B* **161**, 496–503.

Klingauf, J., Kavalali, E.T., Tsien, R.W. (1998) Kinetics and regulation of fast endocytosis at hippocampal synapses, *Nature* **394**, 581–585.

Li, C., Ullrich, B., Zhang, J.Z., Anderson, R.G.W., Brose, N., Sudhof, T.C. (1995) Ca^{2+}-dependent and independent activities of neural and non-neural synaptotagmins, *Nature* **375**, 594–599.

Link, E., Edelmann, L., Chou, J.H., Binz, T., Yamasaki, S., Eisel, U., Baumert, M., Sudhof, T.C., Niemann, H., Jahn, R. (1992) Tetanus toxin action: inhibition of neurotransmitter release linked to synaptobrevin proteolysis, *Biochem. Biophys. Res. Commun.* **189**, 1017–1023.

Littleton, J.T., Chapman, E.R., Kreber, R., Garment, M.B., Carlson, S.D., Ganetzky, B. (1998) Temperature-sensitive paralytic mutations demonstrate that synaptic exocytosis requires SNARE complex assembly and disassembly, *Neuron* **21**, 401–413.

Littleton, J.T., Stern, M., Schulze, K., Perin, M., Bellen, H.J. (1993) Mutational analysis of Drosophila synaptotagmin demonstrates its essential role in Ca^{2+}-activated neurotransmitter release, *Cell* **74**, 1125–1134.

Llinas, R., Sugimori, M., Silver, R.B. (1992) Microdomains of high calcium concentration in a presynaptic terminal, *Science* **256**, 677–679.

Machado, J.D., Morales, A., Gomez, J.F., Borges, R. (2001) cAMP modulates exocytotic kinetics and increases quantal size in chromaffin cells, *Mol. Pharmacol.* **60**, 514–520.

Mackler, J.M., Drummond, J.A., Loewen, C.A., Robinson, I.M., Reist, N.E. (2002) The C2B Ca^{2+}-binding motif of synaptotagmin is required for synaptic transmission in vivo, *Nature* **418**, 340–344.

Mahal, L.K., Sequeira, S.M., Gureasko, J.M., Sollner, T.H. (2002) Calcium-independent stimulation of membrane fusion of SNAREpin formation by synaptotagmin I, *J. Cell Biol.* **158**, 273–282.

Matthew, W.D., Tsavaler, L., Reichardt, L.F. (1981) Identification of synaptic vesicle-specific membrane protein with a wide distribution in neuronal and neurosecretory tissue, *J. Cell Biol.* **91**, 257–269.

McMahon, H.T., Missler, M., Li, C., Sudhof, T.C. (1995) Complexins: cytosolic proteins that regulate SNAP receptor function, *Cell* **83**, 111–119.

McNew, J.A., Weber, T., Engelman, D.M., Sollner, T.H., Rothman, J.E. (1999) The length of the flexible SNAREpin juxtamembrane region is a critical determinant of SNARE-dependent fusion, *Mol. Cells* **4**, 415–421.

Monck, J.R., Alvarez de Toledo, G., Fernandez, J.M. (1990) Tension in secretory granule membranes causes extensive membrane transfer through the exocytotic fusion pore, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7804–7808.

Neher, E., Marty, A. (1982) Discrete change of cell membrane capacitance observed under conditions of enhanced secretion in bovine adrenal chromaffin cells, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6712–6716.

Neher, E., Zucker, S. (1993) Multiple calcium-dependent processes related to secretion in bovine chromaffin cell, *Neuron* **10**, 21–30.

Nonet, M.L., Grundahl, K., Meyer, B.J., Rand, J.B. (1993) Synaptic function is impaired but not eliminated in C.elegans mutants lacking synaptotagmin, *Cell* **73**, 1291–1305.

Nonet, M.L., Staunton, J.E., Kilgard, M.P., Fergestad, T., Hartwieg, E., Horvitz, H.R., Jorgensen, E.M., Meyer, B.J. (1997) Caenorhabditis elegans rab3 mutant synapses exhibit impaired function and are partially depleted of vesicles, *J. Neurosci.* **17**, 8061–8073.

Ohana, O., Sakmann, B. (1998) Transmitter release modulation in nerve terminals of rat neocortical pyramidal cells by intracellular calcium buffers, *J. Physiol.* **513**, 135–148.

Oyler, G.A., Higgins, G.A., Hart, R.A., Battenberg, E., Billingsley, M., Bloom, F.E., Wilson, M.C. (1989) The identification of a novel synaptosomal-associated protein SNAP-25, differentially expressed by neuronal subpopulations, *J. Cell Biol.* **109**, 3039–3052.

Pabst, S., Margittai, M., Vainus, D., Langen, R., Jahn, R., Fasshauer, D. (2002) Rapid and selective binding of the synaptic SNARE complex suggests a modulatory role of complexins in neuroexocytosis, *J. Biol. Chem.* **277**, 7838–7848.

Pallanck, L., Ordway, R.W., Ganetzky, B. (1995) A drosophila NSF mutant, *Nature* **376**, 25.

Parsons, T.D., Coorssen, J.R., Horstmann, H., Almers, W. (1995) Docked granules, the exocytic burst, and the need for ATP hydrolysis in endocrine cells, *Neuron* **15**, 1085–1096.

Perin, M.S., Fried, V.A., Mignery, G.A., Jahn, R., Sudhof, T.C. (1990) Phospholipid binding by a synaptic vesicle protein homologous to the regulatory region of protein kinase C, *Nature* **345**, 260–263.

Peters, C., Mayer, A. (1998) Ca^{2+}/calmodulin signals the completion of docking and triggers a late step of vacuole fusion, *Nature* **396**, 575–580.

Pevsner, J., Hsu, S.-C., Braun, J.E.A., Calakos, N., Ting, A.E., Bennett, M.K., Scheller, R.H. (1994) Specificity and regulation of a synaptic vesicle docking complex, *Neuron* **13**, 353–361.

Poirier, M.A., Xiao, W., Macosko, J.C., Chan, C., Shin, Y.-K., Bennett, M.K. (1998) The synaptic SNARE complex is a parallel four-stranded helical bundle, *Nat. Struct. Biol.* **5**, 765–769.

Robinson, I.M., Ranjan, R., Schwarz, T.L. (2002) Synaptotagmins I and IV promote transmitter release independently of Ca^{2+} binding in the C2A domain, *Nature* **418**, 336–340.

Rodriguez, A., Webster, P., Ortego, J., Andrews, N.W. (1997) Lysosomes behave as Ca^{2+}-regulated exocytic vesicles in fibroblasts and epithelial cells, *J. Cell Biol.* **137**, 93–104.

Rosenmund, C., Stevens, C.F. (1996) Definition of the ready releasable pool of vesicles at hippocampal synapses, *Neuron* **16**, 1197–1207.

Rybin, V., Ullrich, O., Rubino, M., Alexandrov, K., Simon, I., Seabra, M.C., Goody, R., Zerial, M. (1996) GTPase activity of Rab5 acts as a timer for endocytic membrane fusion, *Nature* **383**, 266–269.

Sabatini, B.L., Regehr, W.G. (1996) Timing of neurotransmission at fast synapses in the mammalian brain, *Nature* **384**, 170–172.

Salminen, A., Novick, P.J. (1987) A ras-like protein is required for a post-Golgi event in yeast secretion, *Cell* **49**, 527–538.

Scepek, S., Coorssen, J.R., Lindau, M. (1998) Fusion pore expansion in horse eosinophils is modulated by Ca^{2+} and protein kinase C via distinct mechanisms, *EMBO J.* **17**, 4340–4345.

Schiavo, G., Benfenati, F., Poulain, B., Rossetto, O., Polverino de Laureto, P., DasGupta, B.R., Montecucco, C. (1992) Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin, *Nature* **359**, 832–835.

Schneggenburger, R., Neher, E. (2000) Intracellular calcium dependence of transmitter release rates at a fast central synapse, *Nature* **406**, 889–893.

Sheng, Z.-H., Rettig, J., Cook, T., Catterall, W.A. (1996) Calcium-dependent interaction of N-type calcium channels with the synaptic core complex, *Nature* **379**, 452–454.

Siddiqi, O., Benzer, S. (1976) Neurophysiological defects in temperature-sensitive paralytic mutants of drosophila melanogaster, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3253–3257.

Søgaard, M., Tani, K., Ruby Ye, R., Geromanos, S., Tempst, P., Kirchhausen, T., Rothman, J.E., Söllner, T. (1994) A Rab protein is required for the assembly of SNARE complexes in the docking of transport vesicles, *Cell* **78**, 937–948.

Sollner, T., Bennett, M.K., Whiteheart, S.W., Scheller, R., Rothman, J. (1993a) A protein assembly-disassembly pathway *in vitro* that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion, *Cell* **75**, 409–418.

Sollner, T., Whiteheart, S.W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P., Rothman, J.E. (1993b) SNAP receptors implicated in vesicle targeting and fusion, *Nature* **362**, 318–324.

Steyer, J.A., Horstmann, H., Almers, W. (1997) Transport, docking and exocytosis of single secretory granules in live chromaffin cells, *Nature* **388**, 474–478.

Stinchcombe, J.C., Barral, D.C., Mules, E.H., Booth, S., Hume, A.N., Machesky, L.M., Seabra, M.C., Griffiths, G.M. (2001) Rab27a is required for regulated secretion in cytotoxic T lymphocytes, *J. Cell Biol.* **152**(4), 825–833.

Sugita, S., Shin, O.-h., Han, W., Lao, Y., Sudhof, T.C. (2002) Synaptotagmins form a hierarchy of exocytotic Ca^{2+} sensors with distinct Ca^{2+} affinities, *EMBO J.* **21**, 270–280.

Sun, J.-Y., Wu, X.-S., Wu, L.-G. (2002) Single and multiple vesicle fusion induce different rates of endocytosis at a central synapse, *Nature* **417**, 555–559.

Sutton, R.B., Fasshauer, D., Jahn, R., Brunger, A.T. (1998) Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4A resolution, *Nature* **395**, 347–353.

Terbush, D.R., Maurice, T., Roth, D., Novick, P. (1997) The exocyst is a multiprotein complex required for exocytosis in saccharomyces cerevisiae, *EMBO J.* **15**, 6483–6494.

Thomas, P., Wong, J.G., Almers, W. (1993) Millisecond studies of secretion in single rat pituitary cells stimulated by flash photolysis of caged Ca^{2+}, *EMBO J.* **12**, 303–306.

Tolar, L.A., Pallanck, L. (1998) NSF function in neurotransmitter release involves rearrangement of the SNARE complex downstream of synaptic vesicle docking, *J. Neurosci.* **18**, 10250–10256.

Trimble, W.S., Cowan, D.M., Scheller, R.H. (1988) VAMP-1: a synaptic vesicle-associated integral membrane protein, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4538–4542.

Vitale, M.L., Seward, E.P., Trifaro, J.-M. (1995) Chromaffin cell cortical actin network dynamics control the size of the release-ready vesicle pool and the initial rate of exocytosis, *Neuron* **14**, 353–363.

Voets, T., Toonen, R., Brian, E.C., de Wit, H., Moser, T., Rettig, J., Sudhof, T.C., Neher, E., Verhage, M. (2001) Munc-18 promotes large dense-core vesicle docking, *Neuron* **31**, 581–591.

von Gersdorff, H., Matthews, G. (1994) Dynamics of synaptic vesicle fusion and membrane retrieval in synaptic terminals, *Nature* **367**, 735–739.

Wang, C.-T., Grishanin, R., Earles, C.A., Chang, P.Y., Martin, T.F.J., Chapman, E.R., Jackson, M.B. (2001) Synaptotagmin modulation of fusion pore kinetics in regulated exocytosis of dense-core vesicles, *Science* **294**, 1111–1115.

Wang, Y., Okamoto, M., Schmitz, F., Hofmann, K., Sudhof, T.C. (1997) Rim is a putative Rab3 effector in regulating synaptic-vesicle fusion, *Nature* **388**, 593–598.

Weber, T., Zemelman, B.V., McNew, J.A., Westermann, B., Gmachl, M., Parlati, F., Sollner, T.H., Rothman, J.E. (1998) SNAREpins: minimal machinery for membrane fusion, *Cell* **92**, 759–772.

Wightman, R.M., Jankowski, J.A., Kennedy, R.T., Kawagoe, K.T., Schroeder, T.J., Leszczyszyn, D.J., Near, J.A., Diliberto, E.J. Jr., Viveros, O.H. (1991) Temporally resolved catecholamine spikes correspond to single vesicle release from individual chromaffin cells, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10754–10758.

Wu, M.N., Littleton, J.T., Bhat, M.A., Prokop, A., Bellen, H.J. (1998) ROP, the Drosophila sec1 homolog, interacts with syntaxin and regulates neurotransmitter release in a dosage-dependent manner, *EMBO J.* **17**, 127–139.

Xu, T., Ashery, U., Burgoyne, R.D., Neher, E. (1999) Early requirement for α-SNAP and NSF in the secretory cascade in chromaffin cells, *EMBO J.* **18**, 3293–3304.

Exploiting Model Organisms for Disease Intervention

Caretha L. Creasy, George P. Livi, and Christine Debouck
Division of Genomic & Proteomic Sciences, GlaxoSmithKline Pharmaceutical Research & Development, Collegeville, PA, USA

1 Exploiting Model Organisms for Disease Intervention 331

2 Model Organism Biology 332
2.1 Yeast and Invertebrate Organisms 332
2.2 Lower Vertebrates 334
2.3 Rodents 335

3 Development of Disease Models 337

4 Identification of Therapeutic Targets 340

5 Functional Validation of Targets 343

6 Identification of Pathways 344

7 Development of Functional Assays 345

8 Compound Identification and Progression 346

9 Summary and Future Perspective 348

Acknowledgments 349

Bibliography 349
Books and Reviews 349
Primary Literature 350

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Keywords

Complementation
The functional replacement of a mutated or deleted endogenous gene in a given organism by a gene introduced and expressed in that organism.

Enhancer Screen
A genetic screen to identify mutants that exhibit a more severely altered phenotype.

Epistasis
A situation in which a mutation in one gene (the epistatic gene) masks the phenotype of a mutation in another gene (the hypostatic gene); epistasis analysis refers to a genetic technique used to determine whether two proteins function in the same pathway, and if so, to order the proteins in the pathway.

High-throughput Screening
Screening of large numbers (100 000 to 1 000 000) of compounds (small molecules, natural products) using a format (biophysical, biochemical or functional assay) and a process (robotics, automation, miniaturization) that allow screening to be completed in a few days.

Lead Optimization
Chemistry devoted to improving the quality of a compound obtained from high-throughput screening, through optimization of the potency, efficacy, specificity, safety, pharmacokinetic and drug metabolism profiles.

Suppressor Screen
A genetic screen to identify mutants that result in reversion of an altered phenotype to a wild-type phenotype.

Synthetic Lethality
A situation in which two mutations (typically in different genes) cause lethality when present in combination, but do not give rise to a lethal phenotype when present individually.

Model organisms provide highly relevant and convenient platforms for characterizing novel protein functions and elucidating complex biological systems, including defining the roles of these proteins and systems in disease processes. This chapter describes various practical uses of model organisms (yeast, fly, worm, fish, and mouse) that contribute to the identification and validation of human therapeutic targets as well as impact on early drug development.

1
Exploiting Model Organisms for Disease Intervention

A major challenge in biomedical and pharmaceutical research is to understand the biological functions of proteins as they relate to human diseases, and to select the most promising targets for therapeutic intervention. A general roadmap of the drug discovery and development process is shown in Fig. 1.

The initial step of target identification is followed by target validation, which involves detailed studies to correlate a protein and its function with disease onset or progression. Once sufficient biological validation is available for a given target, a biophysical, biochemical or biological assay is developed and high-throughput screening is conducted to identify lead compounds (inhibitors, agonists, antagonists). This is followed by successive rounds of medicinal chemistry to optimize the lead compound, with the goal of improving its potency, efficacy, specificity, safety, pharmacokinetic and drug metabolism properties. Optimized compounds are then tested for safety and efficacy in animals and subsequently in human subjects in a series of clinical trials typically involving healthy volunteers (phase I) and patients (phases II and III).

The entire process of drug discovery and development requires an enormous expenditure of time and resources without the guarantee of bringing a drug to market, and only a small percentage of compounds successfully progress through preclinical and clinical testing and make it to the marketplace as new drugs. Hence, in an effort to reduce attrition of compounds during the late and costly phases of drug development, and to increase the overall probability of success, novel approaches are implemented upfront to significantly increase the biological validation of the therapeutic targets. The human genome sequence is actively being mined for new genes that play a central role in the onset or progression of pathophysiologies. The selection of successful therapeutic targets depends largely on the generation and integration of high-quality data from a variety of applied genomic, genetic, and biological technologies. A key element of genomics-based drug discovery today involves comparative and functional genomic studies in model organisms, including yeasts (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*), invertebrates (the soil nematode, *Caenorhabditis elegans*, and the fruit fly, *Drosophila melanogaster*), lower vertebrates (zebrafish, *Danio rerio*), and rodents (the mouse, *Mus musculus*, and the rat, *Rattus norvegicus*), with the aim of delineating gene functions in the context of a whole organism

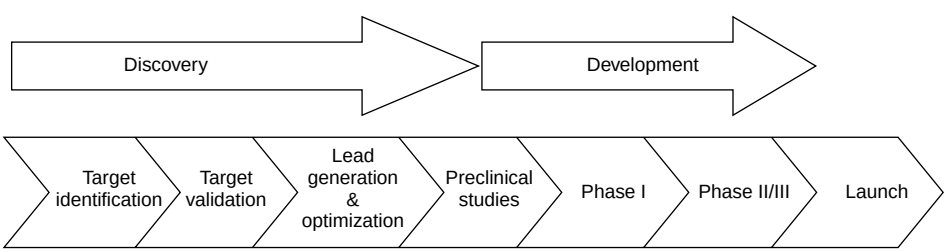

Fig. 1 Schematic representation of the typical drug discovery and development process.

that can be easily, rapidly and inexpensively manipulated. Gene-driven biology in these model organisms has dramatically advanced the understanding and exploitation of human proteins as therapeutic targets, and is contributing to the development of new medicines and new strategies for therapeutic intervention. The most important element for success with these highly diverse model organisms is knowing how to exploit their respective distinct advantages, and asking the right biological questions in the case of each right organism.

2 Model Organism Biology

The completion of the human genome sequence has served as a springboard for studying the functions of novel genes in *in vivo* systems, with special interest in model organisms that afford a high-throughput format and whose biology is as close to human biology as possible. Since the time and resources needed to generate transgenic rodents do not allow high-throughput studies in these organisms, scientists turned to simpler model organisms in which gene inactivation, pathway analysis and compound testing can be more easily performed, and in which results can be more easily interpreted. The data are then translated back to mammalian or, ideally, human systems where the proposed functions and processes can be verified. There are many advantages of working with nonmammalian model systems. The experiments are relatively fast, inexpensive, in some cases amenable to automation, and quite often highly informative, thereby contributing to our overall understanding of protein functions.

2.1 Yeast and Invertebrate Organisms

Many biological pathways relevant to humans are not only conserved, but were actually first uncovered in nonmammalian systems. The Nobel prize-winning work of Lee Hartwell, Paul Nurse, and Timothy Hunt to identify and analyze cell division cycle (CDC) mutants in yeast led to the current understanding of the human cell cycle and DNA repair and, ultimately, to a better understanding of cancer biology. Similarly, much of what we know about the critical players in Alzheimer's (presenilins, amyloid precursor protein (APP) and tau), Parkinson's (parkin) and Huntington's diseases were elucidated in the fly and in the worm.

In fact, a great deal of functional knowledge about mammalian proteins in general has been deduced through comparative bioinformatics and experimentally derived functions of orthologous proteins in yeast, worm, and fly. All of these species offer complete genome sequences with fully annotated core proteomes. Although there is considerable protein sequence conservation in many disease genes and protein classes, the key from a therapeutic perspective is to find disease relevance for members of protein families proven to be tractable ("druggable"), including G-protein coupled receptors (GPCRs), nuclear hormone receptors (NHRs), enzymes, and ion channels. These protein classes are collectively represented in the lower model organisms (Table 1). Much of the difference in the number of genes in each model organism as compared to humans, and therefore the number of orthologous relationships between humans and relatively simple model organisms can be attributed to the fact that the function of a single gene in a model

Tab. 1 Model organism orthologs[a] for major traditional drug target protein classes.

	C. elegans	*D. melanogaster*	*S. cerevisiae*	*S. pombe*
GPCRs	28	38	–	–
NRs	13	12	–	–
Kinases	143	150	48	57
Proteases[b]	110	121	50	59
Channels	54	60	6	6

[a] Strict orthologs were reciprocal-BLAST validated. They were obtained by BLAST analysis of each human target class protein to all proteins in each organism. The top model organism hit was then used for BLAST analysis versus a human protein database. If the top human hit from this second BLAST was the same as the human protein used in the first BLAST analysis, then a strict orthologous relationship was said to exist.
[b] Proteasome subunits were not included in the analysis.
Notes: GPCR: G-protein coupled receptor; NR: Nuclear Hormone Receptor.

organism is often fulfilled by two or three genes in humans, where each human gene has a highly specialized function. An example of this is provided by the *aurora* gene, a key mitotic regulator. In yeast, there is one gene (*IPL*), in *C. elegans* there are two genes (*air-1* and *air-2*), in *Drosophila* there are two genes (*aurora* and *ia1*), and in mammals there are three genes (*Aurora A*, *B*, and *C*).

Comparative genome sequence analyses have revealed that the *Drosophila* proteome (as compared to that of yeast and worm) is more similar to mammalian proteomes. For instance, although the fly has considerably fewer GPCRs than the worm (~160 and ~1100 GPCRs in the fly and worm respectively), the fly has a greater number of orthologous relationships with human GPCRs than the worm. In addition, despite the lesser number of NHRs in the fly compared to the worm, there are the same number of orthologous relationships between human and fly NHRs as there are between human and worm NHRs, and there is greater sequence conservation between fly and human NHRs than between human and worm NHRs. Also, it has been established that 60 to 80% of disease-linked human genes have fly orthologs while only 42 to 50% and ~30% of human disease-linked genes have orthologs in worms and yeast respectively. The high degree of sequence conservation between human and fly genes, the complex nature of the organism, and the availability of thousands of mutants has made *Drosophila* the most widely used invertebrate multicellular organism for genetic studies. However, the inability in maintaining frozen stocks, the difficulty in establishing truly high-throughput assays and the difficulty in treating flies with compounds are making this organism less attractive for focused drug discovery studies.

The soil nematode, *C. elegans*, was not used as a model organism until the mid 1960s when Sydney Brenner recognized its potential as a model system to study animal development. His pioneering work along with that of Robert Horvitz and John Sulston was recognized in 2002 with their receiving the Nobel Prize for Medicine. The success of *C. elegans* as a model

organism can be attributed to its ease of cultivation in the lab. This free-living soil nematode, approximately 1 mm in length, has a rapid life cycle (3 days), feeds on *Escherichia coli*, and can be grown in 96- and 384-well formats. Its transparent body allows one to readily observe physiological defects in live worms. In addition, the use of transgenic worm lines expressing transcriptional or translational fusions to green fluorescent protein (GFP) allows gene and protein expression patterns to be easily visualized in a live animal. Although the worm has a small number of cells (959 somatic cells in the adult), many cell types found in mammals are conserved in the worm, including nerve (302 neurons), muscle, excretory, and gut cells.

Two significant events occurred in 1998 that propelled *C. elegans* as a frontline model organism to impact drug discovery research: *C. elegans* became the first multicellular organism with a fully sequenced genome, and a phenomenon known as RNA interference (RNAi) was shown to work very efficiently in the worm. Indeed, Andy Fire's laboratory discovered that injection of double-stranded (ds) RNA molecules can block gene function via degradation of the targeted mRNA. With the entire worm genome sequence in hand and the ability to selectively inactivate any gene of interest by RNAi, rapid reverse genetic experiments became feasible. RNAi initially utilized injection of dsRNA into the worm gonad, but it was quickly discovered that worms can take up dsRNA via soaking or by feeding on *E. coli* engineered to express the dsRNA molecules. RNAi by feeding became a relatively inexpensive way to inactivate a gene of interest and, more importantly, opened the way for rapidly inactivating numerous genes (individually or simultaneously) for loss-of-function studies. Researchers have since created reusable RNAi libraries of bacterial strains. One such library currently covers ~86% of the entire worm genome and has been used in genome-wide RNAi screens to identify genes involved in fat metabolism, embryonic development, and aging.

As in *C. elegans*, RNAi has been an effective way to speed up functional analysis in *Drosophila*. While the primary method is to inject dsRNA into the embryo or to transfect *Drosophila* S2 cells in culture, researchers have developed methods combining P-element insertions and GAL4/UAS (upstream activating sequence) driver lines (see Sect. 3 on "Development of Disease Models" for a description of these lines) to generate flies that express dsRNA in a spatial or temporal manner. Thus, the impact of inactivating a developmentally essential gene in the adult or inactivating a gene in an individual tissue can now be studied. This technique opens the door for systematic RNAi analysis in *Drosophila*, although the need to maintain fly stocks and the difficulty in devising high-throughput screens currently prevent the same rapid whole-genome analyses that are possible with *C. elegans*.

The real power of working with lower model organisms is in combining the facile methods for reverse genetics (gene KOs, RNAi) and transgenics (genetic complementation, humanization), with classical genetics (suppressor/enhancer screens, epistasis analysis, synthetic lethality) and cell and whole-organism biology, as will be illustrated later in this chapter.

2.2 Lower Vertebrates

While yeast and invertebrates together share many of the features that are central to human biology (morphogenesis,

cell cycle, apoptosis, sexual reproduction, neurobiology, muscle development, metabolism, behavior, aging), zebrafish (*D. rerio*) expands the human connection to encompass vertebrate organogenesis, bone, heart, and blood development and physiology. Genome sequencing efforts are still ongoing for zebrafish, but it is already evident that large chromosomal regions are syntenic between zebrafish and humans, and it is projected that nearly all human proteins will have orthologs in zebrafish with 85% average protein homology.

Zebrafish has been used as a model organism since the 1970s when researchers at the University of Oregon began using it as a neuronal model. Before that time, it was primarily used in toxicology studies and in studies of development with little focus on it as an exploitable genetic system. However, unlike many fish, zebrafish exists as a diploid rather than as a triploid or tetraploid, and is therefore more readily amenable to genetic manipulation. Zebrafish are small (3 cm) in size, develop rapidly (freely swimming, fully formed fish in 4 days), and are fertilized externally, allowing synchronized studies. Like *C. elegans*, zebrafish embryos can grow in a 96-well format and are transparent, allowing the use of transcriptional/translational fluorescent reporters. However, unlike *C. elegans*, zebrafish are much more permeable to small molecules. They are also amenable to transient gene overexpression and gene knockdown studies in the embryo via injection of mRNA or antisense morpholino oligonucleotides respectively making reverse genetic studies very straightforward and avoiding the need to maintain actively growing stocks. While zebrafish offer many advantages for drug discovery research, there are some disadvantages: the generation time (3 months) is no shorter than that of the mouse, and gene targeting via homologous recombination is not possible at the moment. However, researchers have shown that cultured embryonic cells can undergo homologous recombination *in vitro*, hence, gene knockouts (KOs) should theoretically be possible.

2.3 Rodents

Moving up the evolutionary tree brings us to the mouse (*M. musculus*), which is regarded as the premier model organism for human therapeutic studies. It is, of course, a mammal with physiological processes (major organs, skeletal structure, and mode of reproduction) very similar to humans. The mouse genome shares high similarity with that of humans in terms of size and gene content, with a significant degree of gene synteny. In addition, it features a well-understood genetic system with many inbred mutant strains available and with facile genetic manipulation techniques. Biomedical and pharmaceutical researchers have relied on the mouse for many years as the workhorse for studying gene functions and for validating therapeutic targets *in vivo*. Gene deletion (also known as gene targeting) and gene addition technologies have been used extensively in the mouse to correlate a particular gene with the underlying pathophysiology of a disease. These types of techniques also allow one to predict whether modulation of a specific protein *in vivo* provides a desired therapeutic effect, and, consequently, to anticipate the effect of an agonist or antagonist compound, and recognize potential on-target or off-target side effects of such agents. In addition, transgenic mouse models (in addition to transgenic rats) are often used

to support the development of compounds in terms of understanding or predicting efficacy, toxicity, and drug metabolism.

The types of mouse models utilized in biomedical and drug discovery research include conventional gene KOs (complete loss of mouse gene function), knockins (KIs; replacement of the mouse gene with human cDNA or variant under control of the endogenous mouse promoter), transgenic overexpressors (expression of human cDNA or variant in a specific spatiotemporal manner, typically with a tissue-specific promoter; amenable to both mouse and rat models), and conditional KOs or conditional transgenics (offering complete spatiotemporal control of gene expression in the adult animal). Each has unique advantages and disadvantages, and the choice of model depends largely on the intended application and the questions being addressed. Another important factor is to engineer mice with the appropriate genetic background for phenotypic analysis. A few examples of how these different types of models can be used to impact various stages of drug discovery and development will be discussed later in this chapter.

While in recent years there has been a major focus on generating large numbers of conventional mouse KOs targeting genes of interest for understanding various diseases, it has become clear that there is a growing need for more sophisticated types of models. The ideal model is one that employs some form of gene switch technology, that is, a model in which one can control where and when a gene is switched on or off by administering a small-molecule inducer. These so-called "conditional KOs" and "inducible transgenic overexpressors" provide several distinct advantages over more conventional models: (1) they are more relevant to disease manifestation, offering inducibility and the potential for reversibility of a phenotype; (2) they allow one to assess gene function in an adult mouse, often in combination with other disease states (e.g. diet-induced obesity); (3) they are better models of drug action (i.e. a KO from the start of development may not simulate the effect of an antagonist administered to the adult); (4) they can overcome situations of embryonic or postnatal lethality (in which case a conventional model cannot be easily obtained); and (5) they can overcome the potential issue of functional compensation during development, in which case misleading phenotypic information can be generated.

A number of conditional methods are available. These include methods based on the tetracycline promoter/inducer (Tet-ON and Tet-OFF) system for regulated transgene expression, and methods that utilize recombinogenic loxP repeat sequences and regulated cre recombinase (cre–lox system) for conditional gene KOs. The cre–lox system has been used successfully to create null mutations in specific tissues and cell types. In this system, mice containing "floxed" alleles (i.e. alleles flanked with loxP sites) are crossed with mice containing cre expressed in the desired cell or tissue type, and the resulting gene KO reflects the expression characteristics of the promoter used to drive cre expression. Alternatively, cre can be expressed under the control of a promoter that is inactive in the absence of a small-molecule inducer, and a KO is induced wherever and whenever the inducer is administered. In addition to controlling cre expression at the transcriptional level, some systems involve the use of a ubiquitously expressed fusion protein, consisting of cre linked to a mutant form of the ligand (tamoxifen) binding domain

of the estrogen receptor (cre-ER). In the absence of tamoxifen, the cre-ER protein is sequestered in an inactive form bound to a heat shock protein, but it is activated posttranslationally by association with tamoxifen. Given the great advantages of conditional KO and inducible transgenic systems, many other variations on these basic themes are under development in several laboratories.

A potentially powerful approach for gene targeting in rodents is to exploit RNA interference. Gene knockdown has been achieved in adult mice utilizing vector-mediated or transgenically supplied siRNAs (short interfering RNAs) and shRNAs (short hairpin RNAs). Also, transgenically expressed siRNA has been used for gene silencing in rats, where embryonic stem (ES) cell technology has not yet been established to produce KO animals. While there are still considerable caveats and technical hurdles to overcome before adopting such methods for routine mouse model generation, RNAi combined with inducible transgenic technology has the potential to radically change the way rodents (and possibly other mammals) are exploited to understand gene function *in vivo*.

Another very powerful technique is to replace the targeted gene with a reporter gene (e.g. lacZ or luciferase) under control of the endogenous mouse promoter, allowing one to deduce the native expression of the targeted gene. Similarly, the reporter gene can be linked to an overexpressed transgene allowing one to confirm its appropriate expression in the desired cell type or tissue.

The impact of transgenic mice on target selection and validation is limited by the time and resources required to engineer, breed, and phenotypically evaluate large numbers of recombinant strains. There is the constant pressure of timeliness of delivering phenotypic data to support a decision about whether or not to launch a compound screen against a particular target. Certain advances in technology have helped to streamline and accelerate the rate limiting steps. These include (1) construction of gene targeting vectors by recombination (in some cases based on bacterial artificial chromosomes), (2) automated analysis of recombinant ES cells to identify clones containing targeted alleles, and (3) the use of tetraploid embryo complementation to generate mice fully derived from targeted ES cells eliminating the subsequent breeding steps required to obtain chimeras, and to screen for germ-line transmission of targeted alleles. However, each of these advances comes with factors that may limit their practical application to the drug discovery process, such as sacrificing a preferred genetic background for speed. The bottom line is that transgenic mouse model generation, genotyping, breeding and phenotyping require a major investment in specialized personnel, time, and money. However, the impact of these models is so significant that transgenic rodents play a major role in therapeutic target identification, drug candidate selection, and product development.

We will now concentrate on various aspects of drug discovery and development where model organisms are providing an impact, and cite a few interesting examples along the way.

3 Development of Disease Models

Genetic manipulation of model organisms is fundamental to the creation of many types of disease models that are used extensively in drug discovery. Transgenic technology in the mouse has given rise

to several models that recapitulate human diseases, and have been used to advance our understanding of pathophysiologies and to enable new methods of drug treatment to be explored.

One of the first models to be developed was a transgenic mouse expressing an activated v-Ha-ras oncogene under control of the mouse mammary tumor virus promoter (the famous ''OncoMouse''), which is prone to develop tumors. Since then, more sophisticated applications of transgenic technology have contributed to our understanding of the functions of oncogenes and tumor suppressor genes, as well as the mechanisms of metastasis, angiogenesis, and tumor maintenance and regression. For example, inducible expression of oncogenic Ras (an *H-RasV12G* transgene) in a mouse lacking the *INK4a* tumor suppressor gene results in an inducible melanoma model, providing genetic evidence that this allele plays a critical role in tumorigenesis and maintenance of solid tumors.

In the field of cardiovascular diseases, there are transgenic mouse models for hypertension, obesity, insulin resistance, lipoprotein regulation, and atherosclerosis. An example is the ApoE homozygous KO (−/−) mouse, which develops atherosclerotic lesions histologically similar to the lesions observed in humans. Further, ApoE −/− in combination with homozygous mutations in the low-density lipoprotein (LDL) receptor gene enhances the atherosclerotic phenotype.

Transgenic mouse models have been developed for Alzheimer's disease (AD), for example, mice engineered to overexpress human APP. These mice display AD-like neuropathologies (beta-amyloid peptide deposits, neuritic plaques, synaptic loss, neurodegeneration) and behavioral phenotypes (memory loss). They have been used to investigate new therapeutic approaches that would be impossible to pursue in humans due to the extreme difficulty in obtaining normal and diseased brain tissue during the course of disease progression. In addition, a PDAPP transgenic mouse model was created which expresses a disease-linked mutant form of human APP (which in humans results in increased production of the Abeta42 peptide found in amyloid plaques) and develops neuropathological signs of AD over time. Young PDAPP mice immunized with Abeta42 fail to develop signs of disease, whereas mice immunized at an older age show reduced disease progression, opening up the intriguing possibility of a vaccine approach for preventing or treating AD.

While transgenic technologies in the mouse allow one to closely incarnate the human condition, lower model organisms also play an important role in disease model development. However, drawing meaningful parallels with human biology can often be a challenge in these systems. Successfully recapitulating specific physiological aspects of human diseases in worm, fly, or fish has some clear advantages in terms of speed; plus these models are amenable to powerful platform genetic screens for further target identification and pathway analysis. Phenotypic suppression in specific disease models in worm, fly, or fish that are relevant to human biology, can reveal proteins whose human counterparts may turn out to be excellent targets for therapeutic intervention.

Advances in *Drosophila* genetics have enabled parallels between humans and flies to be exploited more effectively and more rapidly. One approach to generate disease models in the fly is to use the GAL4-UAS system, which allows the selective activation of any cloned gene

in a variety of tissue- and cell-specific patterns. This system utilizes a panel of "driver" fly lines, each of which expresses the yeast GAL4 transcriptional activator under a different tissue- or cell-specific promoter. A second transgenic fly line is generated in which the gene of interest is placed downstream of multiple DNA binding sites for *GAL4-UAS*. Using simple genetic crosses, transgenic flies can then express the target gene in the desired pattern by mating with driver fly stocks covering a range of tissues. This method has now been adapted to zebrafish, so one can easily engineer disease models in which overexpression of a gene induces disease in fly or fish (e.g. tumorigenesis, neurological defects).

The "FLP-out technique" in *Drosophila* (involving the FLP/FRT system of yeast) is another powerful way to achieve heritable loss of function in target tissues. The FLP/FRT system also provides a way to screen conditionally in target tissues for mutations that would normally be lethal or yield complex phenotypes. For example, the FLP/FRT system can be used to make the eye of the fly homozygous for a given mutation, while the rest of the animal has wild-type gene function. This technique allows genetic screens (e.g. mosaic screens) for mutations in genes involved in essential pathways, such as synaptic transmission or tumorigenesis. This approach is particularly useful when conducting suppressor/enhancer or gain-of-function screens involving phenotypes produced by tissue-specific overexpression of a human gene of interest. A tetracycline-inducible expression system has also been developed for transgenic research in the fly. Using tetracycline-inducible transactivators (Tet-ON), inclusion of low levels of the inducer drug doxycycline in normal fly food leads to rapid induction of target transgenes in adults, larvae, and embryos. Combining this with the GAL4-UAS system produces robust, spatially and temporally regulated transgene induction.

Drosophila is proving to be an excellent model for cancer research, with over 30 fly genes identified, which, when mutated, cause tumor growth or overproliferation; several have mammalian counterparts and functional conservation in mammals. One such gene, *Lats/Warts*, was identified using the FLP/FRT system in a screen for tumor suppressors in the fly, where somatic cell mutations result in tumor outgrowths. The phenotype of mutant *lats* flies can be rescued by the human *LATS1* gene indicating functional conservation. *Lats1*-deficient mice were shown to develop soft tissue sarcomas and ovarian tumors, and are sensitive to carcinogens. In addition, adenovirus-mediated overexpression of human *LATS* causes a decrease in proliferation in a variety of tumor cell lines. Subsequent work identified the *lats* gene product as a novel cell cycle mediator that inhibits CDK1 activity leading to G2/M arrest and promotes apoptosis through induction of Bax. Given that *Lats* is a tumor suppressor, it is not a drug target; however, *lats* mutants could be used in suppressor screens to identify targets that, when inactivated, prevent tumor outgrowth in the fly.

Given the critical need to identify more effective treatments for Alzheimer's (AD) and Parkinson's (PD) diseases, numerous laboratories have created models for these diseases. Mice, worms, flies, and zebrafish have been used to overexpress human tau, a protein that forms neurofibrillary tangles in the brains of AD patients and others with neurodegenerative diseases. In each organism, a clear phenotype

was observed that was consistent with what is known about tauopathies. For instance, in *Drosophila* and *C. elegans*, neuronal loss was observed, and tau accumulated in aggregates in the *C. elegans* model. In zebrafish, cytoskeletal disruptions that resembled neurofibrillary tangles were observed.

A Parkinson's disease model has been established in mice, *Drosophila* and *C. elegans* in which α-synuclein is overexpressed from a dopanergic promoter. In all three organisms, motor deficits and neuronal and dendritic loss are observed, similar to those observed in patients with PD, thus faithfully recapitulating the essential features of the human disorder. Genetic screens in worms or flies and microarray experiments with the mouse, fly, or worm models should be useful to better understand the disease and identify novel therapeutic targets for PD.

The ability to create sophisticated models in a simple vertebrate like zebrafish offers great potential. These models will in most cases mimic human diseases better than the worm or fly models will, and the ability to conduct genetic screens offers a clear advantage for zebrafish over the mouse. Besides the AD model mentioned above, cardiomyopathy, muscular dystrophy, bacterial pathogenesis, and cancer models have been created in zebrafish. One cardiomyopathy model utilizes a loss-of-function mutation in the TNNT2 gene which, when mutated in humans, causes familial cardiomyopathy. While cancer in zebrafish has been studied for many years, the first cancer model was recently established in zebrafish in which the mouse *c-Myc* transgene was expressed from a lymphocyte-specific promoter (Rag2) resulting in T-cell acute lymphoblastic leukemia.

4 Identification of Therapeutic Targets

Genetic screens in model organisms are often used to identify new therapeutic targets. Nonmammalian models are the systems mainly exploited for this purpose, although forward genetic screens in mice (*via* ENU [ethylnitrosourea] mutagenesis) or in ES cells have been successful in revealing genes/proteins of potential therapeutic interest. We will focus here on genetic screens in lower species. To be successful with this type of approach, one needs to begin with either a good disease model or initial knowledge that the molecular pathway underlying the disease is conserved in the organism of choice.

Forward genetic screens require no *a priori* knowledge of the genes involved and will identify genes coding for proteins of any class. In this approach, the model organism is subjected to chemical or physical mutagenesis (or P-element insertion in *Drosophila*) followed by a screen to identify mutants with the desired phenotype. Once identified, the mutations are mapped and the genes isolated. While forward genetic approaches are very powerful, considerable time and resources are needed to map the mutations and identify the genes in multicellular organisms despite recent advances in Single Nucleotide Polymorphism (SNP) mapping.

One recent phenotype-driven approach in the fly revealed an interesting protein, the product of a gene called *Indy*, whose human counterpart may be of interest as a drug target. Analysis of a collection of mutant *Drosophila* obtained by P-element insertions, identified five lines with >50% increase in life span. All P-element insertions were found to be in the same gene designated *Indy*, for "I'm Not Dead Yet." The maximum effect was

seen in heterozygous P/+ flies, less in homozygous P/P flies, and an opposite effect (shortened life span) was seen in P/deletion flies, suggesting that the level of *Indy* protein is critical for life span extension. *Indy* encodes a sodium-coupled citrate transporter (11 transmembrane protein) expressed in the *Drosophila* fat body that is involved in storage of fat, glycogen, and protein, equivalent to the mammalian liver. There are two human orthologs of *Indy*, and orthologs from mouse, rat, and worm have also been cloned. The hypothesis is that *Indy* mutations create a metabolic state that mimics moderate caloric restriction; therefore, modulation of the human protein(s) may be an effective therapeutic approach for tackling obesity.

Until fairly recently, the only approach to a genetic screen was through forward genetics. However, with gene sequences in hand, reverse genetic techniques can be used where the gene of interest is selectively inactivated by deletion in the genome (yeast, fly, and mouse), RNAi (worm and fly) or antisense oligonucleotides (zebrafish). Today, reverse genetics can be applied on a whole-genome scale in yeast using a library of yeast strains, each with a single gene deleted, or in *C. elegans* using a whole-genome RNAi library expressed in *E. coli*. The power of this approach is that once the screen is completed, the identity of the gene is known. This does not mean that traditional forward genetic approaches are no longer useful as gene inactivation methods cannot be used to identify gain-of-function mutations, and in the worm, the nervous system is somewhat refractory to RNAi even when RNAi-hypersensitive mutants are utilized.

A recent genome-wide RNAi analysis in *C. elegans* aimed at identifying fat regulatory genes, illustrates how the worm can be used to identify novel targets for tackling obesity. In this screen, the vital dye Nile Red was used to stain fat after RNAi of each *C. elegans* gene. Numerous genes were identified that increased or decreased fat storage, including genes in the insulin, serotonin, or tubby signaling pathways. By simply identifying the human orthologs of these genes, and selecting those that encode tractable drug targets (e.g. enzymes, transporters, ion channels, GPCRs, and nuclear receptors), a fairly short list of potential drug targets can be obtained. Literature searches with the human ortholog and all model organism orthologs to this protein can help to narrow the candidates further. Additional target validation studies can then be carried out with selected targets. A process for using genomic-scale RNAi in *C. elegans* to identify human proteins as potential drug targets is shown in Fig. 2.

Given that yeast is a single-cell organism, much of its application to target identification has focused on the cell cycle and conserved cellular pathways. *S. cerevisiae* clearly has its place in drug discovery, for example, as a platform for heterologous protein expression, functional cell-based assays, structure-function analysis of human proteins, and two-hybrid screening for protein–protein interactions. However, the fact that yeast lacks many of the biological processes involved in complex human diseases that require multicellularity as well as highly differentiated cell types (such as AD, schizophrenia, asthma, and osteoarthritis), limits its use in the identification and validation of most drug targets. Notwithstanding this, given that many molecular pathways are conserved from yeast to man, the sophisticated and ultrarapid molecular and genetic techniques available in yeast can be utilized to facilitate pathway dissection of those targets that have orthologs in yeast.

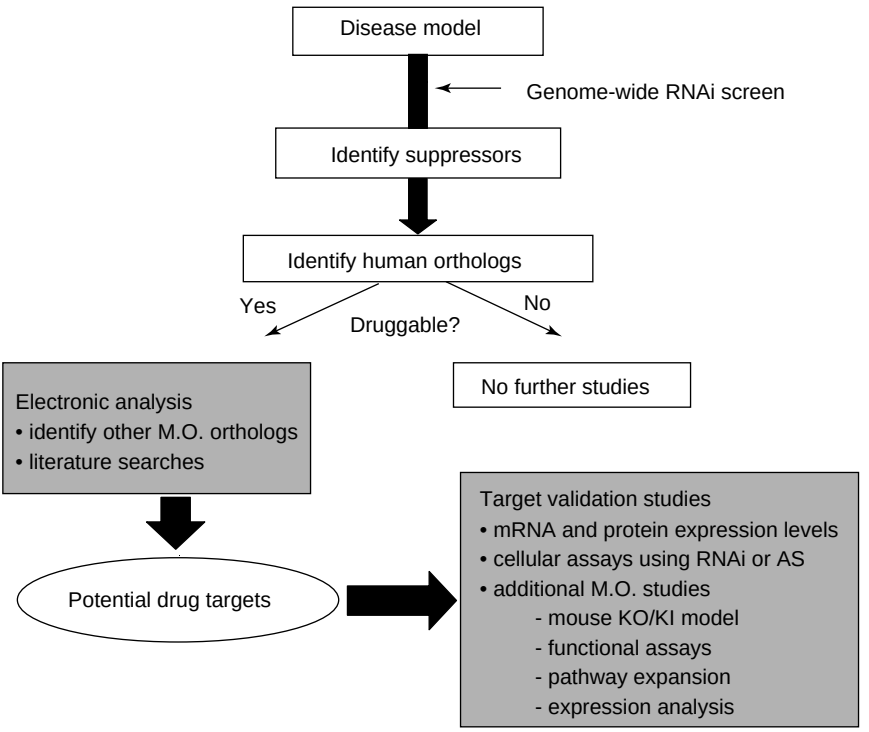

Fig. 2 Target identification process in *C. elegans* using genomic-scale RNAi. AS, antisense oligonucleotides; MO, model organism; KO, knockout; KI, knockin.

Another strategy for novel target discovery is to identify the molecular target of an existing drug or compound that has a desired physiological effect in animals or humans, but whose mechanism of action (MOA) is unknown (i.e. "orphan" compounds). Knowledge of the target of any compound is likely to accelerate the discovery of new medicines with a similar or related MOA. The purpose of a genetic-based MOA screen, typically implemented in yeast or worm is not just to understand the physiological effects of drug treatment in a worm or yeast cell, but rather, to use the model organisms as a platform for generating drug-resistant or drug-hypersensitive mutants that, through genetic characterization, may reveal proteins involved in mediating drug sensitivity. This approach offers the possibility of identifying the direct molecular target(s) of the drug, other proteins in the pathway leading to drug sensitivity, as well as proteins that mediate drug uptake, efflux, and metabolism. The first step, of course, is to determine whether the compound of interest induces a scorable phenotype (such as death) that can be used as the basis for screening or selecting mutants. In one such published example, fluoxetine (Prozac)-resistant mutants of *C. elegans* were isolated. The mutants define seven genes, two of which were genetically characterized and found to encode members of a family of novel 12-transmembrane proteins of unknown function. The functions of the orthologous human proteins may offer some insight into the MOA of fluoxetine.

Perhaps the best example of successful MOA studies that revealed proteins directly involved in drug binding and drug action, and in which the data obtained in a model organism were quickly translated to humans, are the studies conducted in yeast in the early 1990s that identified the direct targets and downstream signaling pathways for the immunosuppressive drugs rapamycin, FK506, and cyclosporin A. These drugs had been used to prevent organ rejection in transplant recipients, and the MOA studies in yeast took advantage of the fact that these drugs were originally isolated as natural products with antifungal activity. The hypothesis was that understanding the mechanism by which these compounds kill a yeast cell might reveal something about how they affected T-cell function in humans. Yeast mutants resistant to each drug were isolated, and the mutated genes were found to encode the proteins cyclophilin (shown to be a cyclosporin-binding protein) and FKBP12 (an FK506-binding protein that also binds rapamycin), collectively referred to as immunophilins. These proteins are enzymes that catalyze the folding of other proteins, and although drug binding specifically inhibits their enzymatic activity *in vitro*, this loss of function does not account for the cellular action of the drugs. Instead, the pharmacological effects of these drugs are derived from the formation of specific cyclophilin or FKBP12-drug complexes that interact with other downstream cellular proteins, namely, calcineurins and TORs, (targets of rapamycin). Thus, the immunophilins act as molecular chaperones for these drugs, delivering them to other sites of action in the cell, target proteins whose functions have now been well characterized. For example, the mammalian TORs are conserved PI-3-kinases that act as central regulators of cell growth, and because of this, the human counterparts are being investigated as targets for cancer. In all cases, the key proteins in the signaling pathways that govern the MOA of these drugs are highly conserved between yeast and man, and our current understanding of how these drugs work owes a great deal to some elegant genetic studies that originated in the simple baker's yeast.

These types of studies are now greatly facilitated by the availability of genomic-scale gene inactivation technologies. Having a library of >6000 yeast KO strains, and an RNAi library arrayed in *E. coli* covering all *C. elegans* genes, enables automated screening for drug-resistant or drug-hypersensitive mutants, and to then instantly know the identity of the gene, which, when inactivated, gives rise to the phenotype.

5
Functional Validation of Targets

Transgenic mouse models (i.e. gene KOs and overexpression transgenics) offer the greatest prospect for correlating a molecular target with underlying disease pathophysiology, and thus play an important role in therapeutic target selection and early validation. These models provide biological evidence in a mammal for or against the causal relationship between specific genes and disease. Having evidence of a link to disease, or to a phenotype (e.g. blood pressure), which, if modulated, could result in alleviation of disease symptoms, is usually a prerequisite to launching a high-throughput screening or subsequent optimization phase of a compound that involves a major investment in chemistry resources.

There are many examples in the literature, too many to list here, where

transgenic mouse models have been used to validate a particular human protein as a target for drug discovery. Recent published examples from our GlaxoSmithKline team include vanilloid receptor-1 (validated as a target for pain associated with inflammation), uncoupling protein-3 (validated as a target for obesity), the $GABA_{B1}$ receptor subtype (validated as a target for psychiatric and neurological disorders including epilepsy), and urotensin-II receptor (validated as a target for hypertension). There are also other excellent recent reviews on the subject that provide numerous examples.

Since a major challenge currently facing biomedical and pharmaceutical researchers is to extract from the human genome those targets with the greatest potential for therapeutic intervention, there are large-scale efforts underway to systematically KO the druggable genome in the mouse. This is being combined with comprehensive platform phenotyping (histopathology, serum biochemistry, clinical pathology, behavior, etc.) to reveal unexpected phenotypes relevant to human disease. A recent survey of the literature indicated that among those GPCRs that have been knocked out in the mouse, more than 50% had phenotypes that contributed (or would have contributed) to target selection and validation. This percentage increases significantly upon provocative challenge in these models. The next major hurdle is to interpret the avalanche of available *in vivo* biological data from hundreds of mouse KOs, and to focus screening and chemistry resources appropriately around newly validated and previously unexploited targets.

In support of this approach, a recent retrospective analysis of the targets for the top-selling one hundred drugs revealed that there is a strong correlation between mutant phenotypes in the mouse and known drug efficacy in man, that is, many mutant phenotypes have been informative in terms of understanding protein function and deciding pharmaceutical utility, and looking back, were predictive of the therapeutic effect of drugs developed against those particular targets. However, prospectively, not all conventional KOs were equally informative. Thus, validating a protein's role in disease often requires the generation and analysis of more advanced types of mouse models (e.g. tissue-specific conditional KOs and inducible transgenic overexpressors) as well as the use of other genomic approaches (e.g. microarray and proteomic platform technologies).

6
Identification of Pathways

Oftentimes, disease or disease susceptibility genes identified in human genetic studies encode proteins that are not chemically tractable, and/or the disease state is due to a loss-of-function mutation that cannot be reversed with a drug (e.g. ApoE, parkin, huntingtin, polycystic kidney disease (PKD1/2)). In these instances, experiments need to be undertaken to uncover the pathway in which the gene/protein functions with the goal of identifying a suitable drug target and obtaining a better understanding of the underlying cause and progression of the disease. Typically, yeast two-hybrid, proteomic, and transcriptomic studies are undertaken with the human gene to identify binding partners or genes that are up- or downregulated in the disease state. However, if an orthologous gene exists in a simple model organism, genetic studies can be undertaken to identify other genes, which, when inactivated, reverse a

particular phenotype observed in that organism. This type of pathway analysis can also be useful when the target is tractable, but there is poor understanding of the pathway. This may identify substrates or upstream activators, thus supporting high-throughput assay development.

Initially, experiments are conducted to determine if there is a phenotype associated with inactivating or overexpressing the disease gene. If a phenotype is observed, a genetic suppressor screen can be conducted to identify suppressors of the phenotype. The advantage of this approach is that a gene is identified that is able to reverse the disease state when inactivated. If a suppressor screen is not feasible, one can couple other genomic technologies (such as the yeast two-hybrid method or proteomic and transcriptomic expression analysis) with studies in model organisms, whereby genes identified by the genomic approaches are inactivated or activated in the disease model to identify those genes that are able to suppress the disease state.

Polycystic kidney disease, a common heritable disease affecting 1 in 1000 individuals, provides a good example. Mutations in either the PKD1 or PKD2 genes cause PKD in humans, but the function of these genes is largely unknown. Studies in *C. elegans* demonstrated that the two *C. elegans* orthologs, lov-1 and pkd-2, are required for proper male sensory behavior, and are localized to male sensory neurons. This information led researchers to examine the colocalization of the worm and mammalian PKD1 and PKD2 proteins in renal epithelial cells where they found that each localizes to focal adhesions indicating a common sensory function. Given that mutations in both lov-1 and pkd-2 exist, researchers are now poised to elucidate the pathway through the identification of suppressors of the male sensory behavioral phenotype in *C. elegans*.

7 Development of Functional Assays

The ability to complement a defect in a model organism gene through expression of a human gene (as cDNA), can often form the basis of a functional assay that can serve multiple purposes in drug discovery, including target validation, compound screening, and structure-function analyses. In the mouse, replacing a gene with its human counterpart is particularly useful when compounds have pronounced selectivity for the human protein. By generating such a ''humanized'' (KI) mouse model, one can use the model for target validation and for compound progression (see below in Sect. 8).

The advantages of creating ''humanized'' models in other model organisms are primarily speed and throughput. It takes a considerable amount of time to generate a humanized mouse model (12–18 months), but in a simpler organism, a similar model can take as little as 2 to 3 weeks in yeast, or 2 to 3 months in worm, fly, or zebrafish. Once the model is established, it can be used to test thousands of compounds compared to only a handful in the mouse.

The ease with which one can delete an entire gene using homologous recombination in yeast, makes yeast especially useful for complementation studies with human genes. Many human genes have been successfully studied in these types of experiments including several cell cycle genes (CDKs, PLK, CDC14, ATM and survivin), numerous members of the MAPK cascade (JNK1, p38, and PAK1), and various enzymes (LAG1, PDE4, PDE7,

FKBPs and cyclophilins, to name but a few). Yeast is also extremely useful for structure-function analyses, wherein individual codons in the human cDNA are altered and the ability of the mutants to functionally complement the yeast gene is readily determined.

The speed with which one can generate the mutants, obtain stably transformed yeast, and perform the cellular assays represents a very significant advantage compared to the use of mammalian cells and other model organisms. The human cDNA can be expressed at virtually any level in yeast using a variety of promoters and expression vectors, or via targeted integration of the human cDNA in the yeast genome (as a gene replacement) using homologous recombination. Coupled with a selective compound, mutagenesis studies can facilitate structure-activity relationship (SAR) studies particularly where a crystal structure is not available.

8
Compound Identification and Progression

With the exception of functional cell-based assays engineered in yeast cells, models organisms are not typically used as a platform for high-throughput drug screens. However, they can be particularly useful for secondary and tertiary screens to confirm that compounds function *in vivo*, as well as to reveal off-target side effects or toxicities.

Yeast has been successfully used in screening compounds with GPCRs where the integral membrane nature of the receptor makes purification impossible and where overexpression in mammalian cells sometimes causes endogenous GPCR signaling. In this system, the yeast mating pathway has been used to identify agonists or antagonists of human GPCRs. Heterothallic haploid *S. cerevisiae* cells exist as mitotically stable MAT**a** or MATα cells, where MAT**a** cells express **a** factor and the α-factor receptor (Ste2p), and MATα cells express α-factor and the **a** factor receptor (Ste3p). Upon appropriate mating factor binding, the receptor activates the heterotrimeric G-protein that in turn activates an MAPK signaling cascade, thereby leading to transcriptional activation of a variety of genes and to cell cycle arrest. When *STE2* or *STE3* are replaced with a human GPCR expressed from a plasmid, the pathway can be activated by an exogenously added ligand. Such assays also rely on deletion of the *FAR1* gene to prevent cell cycle arrest upon receptor activation, and utilization of an *FUS1-HIS3* reporter for growth selection in the absence of histidine when the mating pathway is activated. The existence of only two GPCRs controlling the mating pathway in yeast is a major advantage since deletion of STE2 or STE3 in a haploid cell results in a null receptor background, facilitating the assay readout. Besides serving as a platform to identify antagonists of a ligand-receptor interaction, this elegant system can also be used to screen for agonists (or surrogate agonists) of an orphan receptor when no ligand is available. While yeast has been a very effective tool to facilitate such GPCR drug discovery programs, the yeast cell wall renders the cells impermeable to some compounds. For this reason, yeast-based assays for intracellular targets are used only when no other option is available, and must then involve the use of hyper-permeable mutants deficient in ergosterol biosynthesis.

On the other hand, zebrafish holds the most promise for use in medium throughput compound screening. Zebrafish are relatively easy to grow in a 96-well format,

tolerate dimethyl sulphoxide (DMSO), and are highly permeable to compounds. In addition, the higher degree of conservation between zebrafish and human proteins may avoid the need to create humanized strains of zebrafish. In a recent paper, zebrafish were treated with 100 small molecules to assess whether they could be used to predict long QT syndrome, a serious cardiac clinical side effect. Zebrafish contain HERG-related K^+ channels, which, in mammals are known to mediate such cardiac liabilities. Using a simple optical videoimaging of the beating heart in young embryos to measure cardiotoxic effects, most of the compounds (22/23) that cause QT prolongation in humans were found to cause bradycardia and atrial ventricular block in zebrafish, indicating that zebrafish can be used to predict QT prolongation. These data suggest that if a suitable assay can be devised, zebrafish can be used for testing drugs for other toxic effects and thereby support the development of compounds.

Humanized mice have also been exploited as tools to support early stages of development of compounds, specifically in cases in which lead compounds display significant species selectivity toward the human target; that is, they exhibit a higher affinity for the human target than for the murine counterpart. A good example is the targeted replacement of the mouse chemokine receptor (CCR2) with its human ortholog (CCR2B). CCR2B has been considered a target for atherosclerosis supported by the phenotype of KO mice lacking either the receptor or its ligand (i.e. impaired monocyte recruitment and protection against development of atherosclerotic plaques). Humanized CCR2B KI mice are viable, appear normal, and can elicit a normal inflammatory response to intraperitoneal injection of thioglycollate (peritonitis model). Macrophages from the KI mice respond to the human MCP-1 ligand (monocyte chemoattractant protein-1) and to its murine counterpart (JE) in *in vitro* chemotaxis assays in a manner similar to human peripheral blood monocytes. Furthermore, human-selective CCR2B antagonists show an enhanced potency for blocking the cellular response to human MCP-1 versus the response to murine JE. The KI mouse provides the ability to test hCCR2B antagonists in a series of inflammatory disease models induced experimentally (such as the peritonitis model) or created genetically (such as an atherosclerosis model created by crossing onto an ApoE −/−background).

Mouse KOs, Kis, and transgenic overexpressors are also useful for investigating drug safety, revealing mechanisms of toxicity, and measuring drug metabolism and pharmacokinetic parameters. Compounds displaying toxicity in other animals can be tested in KO mice lacking the target (as well as in wild-type controls) to determine whether the observed toxicity is target-mediated or is due to lack of selectivity ("off-target" effects). Further, mice expressing reporter genes that are responsive to toxic insult or are part of known stress-induced pathways are useful in identifying mechanisms of toxicity. Mouse models sensitized to respond rapidly to carcinogens (e.g. p53 +/− mice) have also been extremely valuable in the assessment of the toxicity of compounds under development. Lastly, human drug metabolism genes have been introduced into mice in order to create models that more closely mimic the human drug response (e.g. human cytochrome P450 4B1, CYP4B1), and mutants lacking endogenous P450s have been used to study the role of these proteins in drug metabolism.

9
Summary and Future Perspective

The perfect model organism to support biomedical and pharmaceutical research would be a mammal with exquisite genetics and a completely sequenced genome, be amenable to genome-scale gene inactivation, have a rapid generation time, be transparent, grow in a 96-well format, and be drug permeable. However, since no model organism possesses all of these qualities, we must therefore continue to employ a repertoire of different organisms, each with specific advantages, and choose the most appropriate organism in which to investigate the key biological questions.

In the case of yeast, rapid growth and homologous recombination make it particularly useful for functional assay development to support pathway dissection and structure-function studies. While not a focus of this chapter, the rapid doubling time and the ability to scale up cultures using fermentors, has led to yeast being used extensively to overexpress human proteins for subsequent use as protein therapeutics or in biochemical assays and structural studies following protein purification. In addition, yeast has served as a workhorse to identify protein–protein interactions using the yeast two-hybrid assay.

In *Drosophila*, sophisticated genetic techniques like the GAL4-UAS and the FLP/FRT systems have made it particularly useful for establishing disease models to support target identification and pathway analysis using forward genetics. RNAi coupled to the GAL4-UAS system allows spatial and temporal gene inactivation studies in whole flies to support target validation studies.

Whole-genome RNAi, the ability to grow in a high-throughput format and the transparent nature of *C. elegans*, make it an excellent organism for disease model generation followed by rapid target identification or pathway expansion *via* reverse genetics.

Together, yeast, worms, and flies contain many of the fundamental genes and pathways relevant to human biology, but they lack the complex physiology inherent to higher mammals.

Zebrafish bridges the gap between invertebrates and mammals. Complex disease models coupled with both forward and reverse genetic techniques can be utilized to identify and validate targets, and elucidate pathways in fish. Transparent embryos, high drug permeability, and the ability to grow and treat embryos in microtiter plates allow this organism to be used for identification of compounds and for progression as well.

More relevant *in vivo* biological data in support of target validation typically come from studies in transgenic mice. Utilization of mouse disease models to determine lead compound efficacy provides additional target validation and can reveal toxicity issues or off-target effects. The key to successfully using simple model organisms (yeast, worm, fly, fish) to support target identification and validation is to have reliable mammalian cellular assays and, more importantly, predictive mammalian models to follow up on the findings. Hence, mouse models will continue to play a central role in biomedical and pharmaceutical research.

Future developments in model organism studies will inevitably evolve from additional technology advances leading to faster, cheaper, better, or novel methods to manipulate the organisms discussed in this chapter. As the genomes from more and more species are being fully sequenced, comparative sequence analysis will undergo dramatic progress in the

area of gene function prediction. For the most part, these organisms will not lend themselves to the types of genetic and genomic studies described herein, and will only contribute their genome sequences to furthering biological knowledge. This has been the case for the Japanese pufferfish *Fugu rubripes* whose compact genome has been completely sequenced and shown to comprise large regions of conserved synteny with the human genome. *Fugu*'s genome is being actively analyzed to identify disease-causing genes, to refine gene families into their subfamilies, such as the NHR family and its subgroups, and to validate *in silico* gene predictions from the human genome. However, it is unlikely that *Fugu* will be amenable to the types of manipulations currently possible with zebrafish.

Some organisms exploited during the pregenomic era may resurface as attractive models for gene function understanding, disease modeling or compound profiling. One such example might be the cellular slime mold, *Dictyostelium discoideum*, which has been used extensively to study cell differentiation, signal transduction, cell motility, and other unicellular or multicellular processes. The ongoing sequencing of *Dictyostelium*'s genome by an international consortium and the observation that RNAi gene silencing works in this organism are expected to revitalize its experimental attractiveness. In the meantime, scientists continue to use *Dictyostelium* for specific studies, such as in the recent identification of a gene responsible for lithium resistance that led to uncovering a role for inositol metabolism in bipolar affective disorder.

New biological knowledge and technology breakthroughs may lead to additional organisms joining the ranks of yeast, worm, fruit fly, zebrafish, and mouse as workhorses for the biomedical or pharmaceutical scientists. This is highly anticipated for the rat. Whereas certain genetic manipulations, such as gene KOs *via* homologous recombination in ES cells, are not currently possible, methods for production of KO rats using ENU mutagenesis have been developed. Rat KO models would be most useful since the rat is a fundamental model for pharmacological and toxicological research.

Finally, many new potential drug targets and pathways are predicted to come from genetic studies in man. Identification of disease-associated polymorphic variants in human genes will reveal many unexploitable proteins, and proteins of unknown function. Once again, it will be important to turn to the relatively simple model organisms to investigate protein functions, to reveal the physiological consequences of polymorphisms *in vivo*, to expand pathways, and ultimately to identify and help validate the most appropriate targets for pharmaceutical intervention.

Acknowledgments

We would like to thank Junping Jing for bioinformatic analysis and Marilyn Williams and Barbara Lewis for help with manuscript preparation.

Bibliography

Books and Reviews

Amatruda, J.F., Shepard, J.L., Stern, H.M., Zon, L.I. (2002) Zebrafish as a cancer model system, *Cancer Cell.* **1**, 229–231.

Barr, M.M. (2003) Super models, *Physiol. Genomics* **13**, 15–24.

Bosher, J.M., Labouesse, M. (2000) RNA interference: genetic wand and genetic watch dog, *Nat. Cell Biol.* **2**, E31–E36.

Carmeliet, P., Collen, D. (2000) Transgenic mouse models in angiogenesis and cardiovascular disease, *J. Pathol.* **190**, 387–405.

Culetto, E., Sattelle, D.B. (2000) A role for Caenorhabditis elegans in understanding the function and interactions of human disease genes, *Hum. Mol. Genet.* **9**, 869–877.

Debouck, C., Goodfellow, P. (1999) DNA microarrays in drug discovery and development, *Nat. Gen. Suppl.* **21**, 48–50.

Fortini, M.E., Bonini, N.M. (2000) Modeling human neurodegenerative disease in Drosophila, *Trends Genet.* **16**, 161–167.

Hannahan, D., Weinberg, R.A. (2000) The hallmarks of cancer, *Cell* **100**, 57–70.

Harris, S., Foord, S.M. (2000) Transgenic gene knock-outs: functional genomics and therapeutic target selection, *Pharmacogenomics* **1**, 433–443.

Hariharan, I.K., Haber, D.A. (2003) Yeast, flies, worms, and fish in the study of human disease, *N. Engl. J. Med.* **348**, 2457–2463.

Lewandowski, M. (2001) Conditional control of gene expression in the mouse, *Nat. Rev.* **2**, 743–755.

Pausch, M. (1997) G-protein coupled receptors in Saccharomyces cerevisiae: high-throughput screening assays for drug discovery, *Trends Biotechnol.* **15**, 487–494.

Potter, C.J., Turenchalk, G.S., Xu, T. (2000) Drosophila in cancer research: an expanding role, *Trends Genet.* **16**, 33–39.

Prosser, H., Rastan, S. (2003) Manipulation of the mouse genome: a multiple impact resource for drug discovery and development, *Trends in Biotechnol.* **21**, 224–232.

Rossant, J. (2003) Targeting mammalian genes – rats join in and mice move ahead, *Nat. Biotechnol.* **21**, 625–627.

Rubinstein, A.M. (2003) Zebrafish: from disease modeling to drug discovery, *Curr. Opin. Drug Discovery Dev.* **6**, 218–223.

Sharp, P.A. (2001) RNA interference – 2001, *Genes Dev.* **15**, 485–490.

Sluder, A.E., Maina, C.V. (2001) Nuclear receptors in nematodes: themes and variations, *Trends Genet.* **17**, 206–213.

Stern, H.M., Zon, L.I. (2003) Cancer genetics and drug discovery in the zebrafish, *Nat. Rev. Cancer* **3**, 533–539.

Thisse, C., Zon, L.I. (2002) Organogenesis – heart and blood formation from the zebrafish point of view, *Science* **295**, 457–462.

Tickoo, S., Russell, S. (2002) *Drosophila melanogaster* as a model system for drug discovery and pathway screening, *Curr. Opin. Pharmacol.* **2**, 555–560.

Tornell, J., Snaith, M. (2002) Transgenic systems in drug discovery: from target identification to humanized mice, *Drug Discovery Today* **7**, 461–470.

Udvadia, A.J., Linney, E. (2003) Windows into development: historic, current and future perspectives on transgenic zebrafish, *Dev. Biol.* **256**, 1–17.

Zambrowicz, B.P., Sands, A.T. (2003) Knockouts model the 100 best-selling drugs – will they model the next 100? *Nat. Rev.* **2**, 38–51.

Primary Literature

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.-M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., Sollewijn Gelpke, M.D., Roach, J., Oh, T., Ho, I.Y., Wong, M., Detter, C., Verhoef, F., Predki, P., Tay, A., Lucas, S., Richardson, P., Smith, S.F., Clark, M.S., Edwards, Y.J.K, Doggett, N., Zharkikh, A., Tavtigian, S.V., Pruss, D., Barnstead, M., Evans, C., Baden, H., Powell, J., Glusman, G., Rowen, L., Hood, L., Tan, Y.H., Elgar, G., Hawkins, T., Venkatesh, B., Rokhsar, D., Brenner, S. (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*, *Science* **297**, 1301–1310.

Ashrafi, K., Chang, F.Y., Watts, J.L., Fraser, A.G., Kamath, R.S., Ahringer, J., Ruvkun, G. (2003) Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes, *Nature* **421**, 268–272.

Barr, M.M., DeModena, J., Braun, D., Nguyen, C.Q., Hall, D.H., Sternberg, P.W. (2001) The *Caenorhabditis elegans* autosomal dominant polycystic kidney disease gene homologs lov-1 and pkd-2 act in the same pathway, *Curr. Biol.* **11**, 1341–1346.

Barr, M.M., Sternberg, P.W. (1999) A polycystic kidney-disease gene homologue required for male mating behaviour in *C. elegans*, *Nature* **401**, 386–389.

Behm, D.J., Harrison, S.M., Ao, Z., Maniscalco, K., Pickering, S.J., Grau, E.V., Woods, T.N., Coatney, R.W., Doe, C.P.,

Willette, R.N., Johns, D.G., Douglas, S.A. (2003) Deletion of the UT receptor gene results in the selective loss of urotensin-II contractile activity in aortae isolated from UT receptor knockout mice, *Br. J. Pharmacol.* **139**, 464–472.

Brown, A.J., Dyos S.L., Whiteway, M.S., White, J.H., Watson, M.A., Marzioch, M., Clare, J.J., Cousens, D.J., Paddon, C., Plumpton, C., Romanos, M.A., Dowell, S.J. (2000) Functional coupling of mammalian receptors to the yeast mating pathway using novel yeast/mammalian G-protein alpha-subunit chimeras, *Yeast* **16**, 11–22.

Bruckner, R.C., Cox, M.M. (1986) Specific contacts between the FLP protein of the yeast 2-micron plasmid and its recombination site, *J. Biol. Chem.* **261**, 11798–11807.

Cafferkey, R., Young, P.R., McLaughlin, M.M., Bergsma, D.J., Koltin, Y., Sathe, G.M., Faucette, L., Eng, W.K., Johnson, R.K., Livi, G.P. (1993) Dominant missense mutations in a novel yeast protein related to mammalian phosphatidylinositol 3-kinase and VPS34 abrogate rapamycin cytotoxicity, *Mol. Cell. Biol.* **13**, 6012–6023.

Carmell, M.A., Zhang, L., Conklin, D.S., Hannon, G.J., Rosenquist, T.A. (2003) Germline transmission of RNAi in mice, *Nat. Struct. Biol.* **10**, 91–92.

Chambers, J.K., Macdonald, L.E., Sarau, H.M., Ames, R.S., Freeman, K., Foley, J.J., Zhu, Y., McLaughlin, M.M., Murdock, P., McMillan, L., Trill, J., Swift A., Aiyar N., Taylor P., Vawter L., Naheed S., Szekeres P., Hervieu G., Scott, C., Watson, J.M., Murphy, A.J., Duzic, E., Klein, C., Bergsma, D.J., Wilson, S., Livi, G.P. (2000) A G protein-coupled receptor for UDP-glucose, *J. Biol. Chem.* **275**, 10767–10771.

Chin, L., Tam, A., Pomerantz, J., Wong, M., Holash, J., Bardeesy, N., Shen, Q., O'hagan, R., Pantginis, J., Zhou H., Horner, J.E., Cordon-cardo, C., Yancopoulos, G.D., DePinho, R.A. (1999) Essential role for oncogenic Ras in tumor maintenance, *Nature* **400**, 468–472.

Choy, R.K., Thomas, J.H. (1999) Fluoxetine-resistant mutants of *C. elegans* define a novel family of transmembrane proteins, *Mol. Cell.* **4**, 143–152.

Clapham, J.C., Arch, J.R., Chapman, H., Haynes, A., Lister, C., Moore, G.B., Piercy, V., Carter, S.A., Lehner, I., Smith, S.A., Beeley, L.J., Godden, R.J., Herrity, N., Skehel, M., Changani, K.K., Hockings, P.D., Reid, D.G., Squires, S.M., Hatcher, J., Trai, L.B., Latcham, J., Rastan, S., Harper, A.J., Cadenas, S., Buckingham, J.A., Brand, M.D., Abuin, A. (2000) Mice overexpressing human uncoupling protein-3 in skeletal muscle are hyperphagic and lean, *Nature* **406**, 415–418.

Crespo, J.L., Hall, M.N. (2002) Elucidating TOR signaling and rapamycin action: lessons from *Saccharomyces cerevisiae*, *Microbiol. Mol. Biol. Rev.* **66**, 579–591.

Davis, J.B., Gray, J., Gunthorpe, M.J., Hatcher, J.P., Davey, P.T., Overend, P., Harries, M.H., Latcham, J., Clapham, C., Atkinson, K., Hughes, S.A., Rance, K., Grau, E., Harper, A.J., Pugh, P.L., Rogers, D.C., Bingham, S., Randall, A., Sheardown, S.A. (2000) Vanilloid receptor-1 is essential for inflammatory thermal hyperalgesia, *Nature* **405**, 183–187.

Dodart, J.C., Mathis, C., Bales, K.R., Paul, S.M. (2002) Does my mouse have Alzheimer's disease? *Genes Brain Behav.* **1**, 142–155.

Eggan, K., Akutsu, H., Loring, J., Jackson-Grusby, L., Klemm, M., Rideout, W.M. III, Yanagimachi, R., Jaenisch, R. (2001) Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation, *Proc. Natl. Acad. Sci.* **98**, 6209–6214.

Eichinger, L., Noegel, A.A. (2003) Crawling into a new era – the *Dictyostelium* genome project, *EMBO J.* **22**, 1941–1946.

Feany, M.B., Bender, W.W. (2000) A Drosophila model of Parkinson's disease, *Nature* **404**, 394–398.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*, *Nature* **391**, 806–811.

Galcheva-Gargova, Z., Derijard, B., Wu, I.H., Davis, R.J. (1994) An osmosensing signal transduction pathway in mammalian cells, *Science* **265**, 806–808.

Gum, R.J., McLaughlin, M.M., Kumar, S., Wang, Z., Bower, M.J., Lee, J.C., Adams, J.L., Livi, G.P., oldsmith, E.J., Young, P.R. (1998) Acquisition of sensitivity of stress-activated protein kinases to the p38 inhibitor, SB 203580, by alteration of one or more amino acids within the ATP binding pocket, *J. Biol. Chem.* **273**, 15605–15610.

Hacker, U., Nystedt, S., Barmchi, M.B., Horn, C., Wimmer, E.A. (2003) *piggyBac*-based insertional mutagenesis in the presence of stably integrated *P* elements in *Drosophila*, *Proc. Natl. Acad. Sci.* **100**, 7720–7725.

Han, J., Lee, J.D., Bibbs, L., Ulevitch, R.J. (1994) A MAP kinase targeted by endotoxin and hyperosmolarity in mammalian cells, *Science* **265**, 808–811.

Hasuwa, H., Kaseda, K., Einarsdottir, T., Okabe, M. (2002) Small interfering RNA and gene silencing in transgenic mice and rats, *FEBS Lett.* **532**, 227–230.

Heitman, J., Movva, N.R., Hall, M.N. (1991) Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast, *Science* **253**, 905–909.

Heitman, J., Movva, N.R., Heistand, P.C., Hall, M.N. (1991) FK506-binding protein proline rotamase is a target for the immunosuppressive agent FK506 in *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci.* **88**, 1948–1952.

Imaoka, S., Hayashi, K., Hiroi, T., Yabusaki, Y., Kamataki, T., Funae, Y. (2001) A transgenic mouse expressing human CYP4B1 in the liver, *Biochem. Biophys. Res. Commun.* **284**, 757–762.

Ishibashi, S., Herz, J., Maeda, N., Goldstein, J.L., Brown, M.S. (1994) The two-receptor model of lipoprotein clearance: tests of the hypothesis in 'knockout' mice lacking the low density lipoprotein receptor, apolipoprotein E, or both proteins, *Proc. Natl. Acad. Sci.* **91**, 4431–4435.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., Linsley, P.S. (2003) Expression profiling reveals off-target gene regulation by RNAi, *Nat. Biotechnol.* **21**, 635–637.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D.P., Zipperlen, P., Ahringer, J. (2003) Systematic functional analysis of *the Caenorhabditis elegans* genome using RNAi, *Nature* **421**, 231–237.

Kazemi-Esfarjani P., Benzer, S. (2000) Genetic suppression of polyglutamine toxicity in Drosophila, *Science* **287**, 1837–1840.

Koltin, Y., Faucette, L., Bergsma, D.J., Levy, M.A., Cafferkey, R., Koser, P.L., Johnson, R.K., Livi, G.P. (1991) Rapamycin sensitivity in *Saccharomyces cerevisiae* is mediated by a peptidyl-prolyl cis-trans isomerase related to human FK506-binding protein, *Mol. Cell. Biol.* **11**, 1718–1723.

Kraemer, B.C., Zhang, B., Leverenz, J.B., Thomas, J.H., Trojanoswski, J.Q., Schellenberg, G.D. (2003) Neurodegeneration and defective neurotransmission in a Caenorhabditis elegans model of tauopathy, *Proc. Natl. Acad. Sci.* **100**, 9980–9985.

Lander, E.S., et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Langenau, D.M., Traver, D., Ferrando, A.A., Kutok, J.L., Aster, J.C., Kanki, J.P., Lin, S., Prochownik, E., Trede, N.S., Zon, L.I., Look, A.T. (2003) Myc-induced T cell leukemia in transgenic zebrafish, *Science* **299**, 887–890.

Langheinrich, U., Hennen, E., Stott, G., Vacun, G., (2002) Zebrafish as a model organism for the identification and characterization of drugs and genes affecting p53 signaling, *Curr. Biol.* **12**, 2023–2028.

Lasko, M., Vartiainen, S., Moilanen, A.-M., Sirvio, J., Thomas, J.H. (2003) Dopaminergic neuronal loss and motor deficits in *Caenorhabditis elegans* overexpressing human α-synuclein, *J. Neurochem.* **86**, 165–172.

Lee, Y.S., Carthew, R.W. (2003) Making a better RNAi vector for *Drosophila*: use of intron spacers, *Methods* **30**, 322–329.

Lee, S.S., Lee, R.Y., Fraser, A.G., Kamath, R.S., Ahringer, J., Ruvkun, G. (2003) A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity, *Nat. Genet.* **33**, 40–48.

Lewis, D.L., Hagstrom, J.E., Loomis, A.G., Wolff, J.A., Herweijer, H. (2002) Efficient delivery of siRNA for inhibition of gene expression in postnatal mice, *Nat. Genet.* **32**, 107–108.

Maglich, J.M., Caravella, J.A., Lambert, M.H., Willson, T.M., Moore, J.T., Ramamurthy, L. (2003) The first completed genome sequence from a teleost fish (Fugu rubripes) adds significant diversity to the nuclear receptor superfamily, *Nucleic Acids Res.* **31**, 4051–4058.

Martens, H., Novotny, J., Oberstrass, J., Steck, T.L., Postlethwait, P., Nellen, W. (2002) RNAi in *Dictyostelium*: the role of RNA-directed RNA polymerases and double-stranded Rnase, *Mol. Biol. Cell* **13**, 445–453.

McHale, M.M., Cieslinski, L.B., Eng, W.-K., Johnson, R.K., Torphy, T.J., Livi, G.P. (1991) Expression of human recombinant cAMP phosphodiesterase isozyme IV reverses growth arrest phenotypes in phosphodiesterase-deficient yeast, *Mol. Pharm.* **39**, 109–113.

Milan, D.J., Peterson, T.A., Ruskin, J.N., Peterson, R.T., MacRae, C.A. (2003) Drugs that induce repolarization abnormalities cause bradycardia in zebrafish, *Circulation* **107**, 1355–1358.

Munroe, R.J., Bergstrom, R.A., Zheng, Q.Y., Libby, B., Smith, R., John, S.W., Schimenti, K.J., Browning, V.L., Schimenti, J.C. (2000) Mouse mutants from chemically mutagenized embryonic stem cells, *Nat. Genet.* **24**, 318–321.

Ninomiya-Tsuji, J., Nomoto, S., Yasuda, H., Reed, S.I., Matsumoto, K. (1991) Cloning of a human cDNA encoding a CDC2-related kinase by complementation of a budding yeast cdc28 mutation, *Proc. Natl. Acad. Sci.* **15**, 9006–9010.

Pichler, F.B., Laurenson, S., Williams, L.C., Dodd, A., Copp, B.R., Love, D.R. (2003) Chemical discovery and global gene expression analysis in zebrafish, *Nat. Biotechnol.* **21**, 879–883.

Pothof, J., van Haaften, G., Thijssen, K., Kamath, R.S., Fraser, A.G., Ahringer, J., Plasterk, R.H., Tijsterman, M. (2003) Identification of genes that protect the *C. elegans* genome against mutations by genome-wide RNAi, *Genes Dev.* **17**, 443–448.

Powell-Braxton, L., Veniant, M., Latvala, R.D., Horano, K.I., Won, W.B., Ross, J., Dybal, N., Zlot, C.H., Young, S.G., Davidson, N.O. (1998) A mouse model of human familial hypercholesterolemia: markedly elevated low density lipoprotein cholesterol levels and severe atherosclerosis on a low-fat chow diet, *Nat. Med.* **4**, 934–938.

Prosser, H.M., Gill, C.H., Hirst, W.D., Grau, E., Robbins, M., Calver, A., Soffin, E.M., Farmer, C.E., Lanneau, C., Gray, J., Schenck, E., Warmerdam, B.S., Clapham, C., Reavill, C., Rogers, D.C., Stean, T., Upton, N., Humphreys, K., Randall, A., Geppert, M., Davies, C.H., Pangalos, M.N. (2001) Epileptogenesis and enhanced prepulse inhibition in GABAB1-deficient mice, *Mol. Cell. Neurosci.* **17**, 1059–1070.

Rogina, B., Reenan, R.A., Nilsen, S.P., Helfand, S.L. (2000) Extended life-span conferred by cotransporter gene mutations in *Drosophila*, *Science* **290**, 2137–2140.

Rohde, J., Heitman, J., Cardenas, M.E. (2001) The TOR kinases link nutrient sensing to cell growth, *J. Biol. Chem.* **276**, 9583–9586.

Rong, Y.S., Golic, K.G. (2000) Gene targeting by homologous recombination in *Drosophila*, *Science* **288**, 2013–2017.

Rorth, P., Szabo, K., Bailey, A., Laverty, T., Rehm, J., Rubin, G.M., Weigmann, K., Milan, M., Benes, V., Ansorge, W., Cohen, S.M. (1998) Systematic gain-of-function genetics in Drosophila, *Development* **125**, 1049–1057.

Rubin, G.M., Yandell, M.D., Wortman, J.R., Gabor, Miklos, G.L., Nelson, C.R., Hariharan, I.K., Fortini, M.E., Li, P.W., Apweiler, R., Fleischmann, W., Cherry, J.M., Henikoff, S., Skupski, M.P., Misra, S., Ashburner, M., Birney, E., Boguski, M.S., Brody, T., Brokstein, P., Celniker, S.E., Chervitz, S.A., Coates, D., Cravchik, A., Gabrielian, A., Galle, R.F., Gelbart, W.M., George, R.A., Goldstein, L.S., Gong, F., Guan, P., Harris, N.L., Hay, B.A., Hoskins, R.A., Li, J., Li, Z., Hynes, R.O., Jones, S.J., Kuehl, P.M., Lemaitre, B., Littleton, J.T., Morrison, D.K., Mungall, C., O'Farrell, P.H., Pickeral, O.K., Shue, C., Vosshall, L.B., Zhang, J., Zhao, Q., Zheng, X.H., Lewis, S. (2000) Comparative genomics of the eukaryotes, *Science* **287**, 2204–2215.

Scheer, N., Riedl, I., Warren, J.T., Kuwada, J.Y., Campos-Ortega, J.A. (2002) A quantitative analysis of the kinetics of Gal4 activator and effector gene expression in the zebrafish, *Mech. Dev.* **112**, 9–14.

Schenck, D., Barbour, R., Dunn, W., Gordon, G., Grajeda, H., Uido, T., Hu, K., Huang, J., Johnson-Wood, K., Khan, K., Kholodenko, D., Lee, M., Liao, Z., Lieberburg, I., Motter, R., Mutter, L., Soriano, F., Shopp, G., Vasquez, N., Vandevert, C., Walker, S., Wogulis, M., Yednock, T., Games, D., Seubert, P. (1999) Immunization with amyloid-beta attenuates Alzheimer-disease-like pathology in the PDAPP mouse. *Nature* **400**, 173–177.

Seibler, J., Zevnik, B., Kuter-Luks, B., Andreas, S., Kern, H., Hennek, T., Rode, A., Heimann, C., Faust, N., Kauselmann, G., Schoor, M., Jaenisch, R., Rajewsky, K., Kuhn, R., Schwenk, F. (2003) Rapid generation of inducible mouse mutants, *Nucleic Acids Res.* **31**, e12.

Shin, M.K., Levorse, J.M., Ingram, R.S., Tilghman, S.M. (1999) The temporal requirement for endothelin receptor-B signalling during neural crest development, *Nature* **402**, 496–501.

Simmer, F., Tijsterman, M., Parrish, S., Koushika, S.P., Nonet, M.L., Fire, A.,

Ahringer, J., Plasterk, R.H. (2002) Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi, *Curr. Biol.* **12**, 1317–1319

St. John, M.A., Tao, W., Fei, X., Fukumoto, R., Carcangiu, M.L., Brownstein, D.G., Parlow, A.F., McGrath, J., Xu, T. (1999) Mice deficient of Lats1 develop soft-tissue sarcomas, ovarian tumors and pituitary dysfunction, *Nat. Genet.* **21**, 182–186.

Stebbins, M.J., Urlinger, S., Byrne, G., Bello, B., Hillen, W., Yin, J.C. (2001) Tetracycline-inducible systems for *Drosophila*, *Proc. Natl. Acad. Sci.* **98**, 10775–10780.

Stowers, R.S., Schwarz, T.L. (1999) A genetic method for generating *Drosophila* eyes composed exclusively of mitotic clones of a single genotype, *Genetics* **152**, 1631–1639.

Struhl, G., Basler, K. (1993) Organizing activity of wingless protein in *Drosophila*, *Cell* **72**, 527–540.

Tavernarakis, N., Wang, S.L., Dorovkov, M., Ryazanov, A., Driscoll, M. (2000) Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes, *Nat. Genet.* **24**, 180–183.

Tao, W., Zhang, S., Turenchalk, G.S., Stewart, R.A., St. John, M.A., Chen, W., Xu, T. (1999) Human homologue of the *Drosophila melanogaster lats* tumor suppressor modulates *CDC2* activity, *Nat. Genet.* **21**, 177–181.

Tiscornia, G., Singer, O., Ikawa, M., Verma, I.M. (2003) A general method for gene knockdown in mice by using lentiviral vectors expressing small interfering RNA, *Proc. Natl. Acad. Sci.* **100**, 1844–1848.

Tomasiewicz, H.G., Flaherty, D.B., Soria, J.P., Wood, J.G. (2002) Transgenic zebrafish model of neurodegeneration, *J. Neurosci.* **70**, 734–745.

Valenzuela, D.M., Murphy, A.J., Frendewey, D., Gale, N.W., Economides, A.N., Auerbach, W., Poueymirou, W.T., Adams, N.C., Rojas, J., Yasenchak, J., Chernomorski, R., Boucher, M., Elsasser, A.L., Esau, L., Zheng, J., Griffiths, J.A., Wang, X., Su, H., Xue, Y., Dominguez, M.G., Noguera, I., Torres, R., McDonald, L.E., Stewart, A.F., DeChaira, T.M., Yancopoulos, G.D. (2003) High-throughput engineering of the mouse genome coupled with high-resolution expression analysis, *Nat. Biotechnol.* **21**, 1–8.

Warrick, J.M., Paulson, H.L., Gray-Board, G.L., Bui, Q.T., Fischbeck, K.H., Pittman, R.N., Bonini, N.M. (1998) Expanded polyglutamine protein forms nuclear inclusions and causes neural degeneration in *Drosophila*, *Cell* **93**, 939–949

Wianny, F., Zernicka-Goetz, M. (2000) Specific interference with gene function by double-stranded RNA in early mouse development, *Nat. Cell. Biol.* **2**, 70–75.

Williams, R.S.B., Cheng, L., Mudge, A.W., Harwood, A.J. (2002) A common mechanism of action for three mood-stabilizing drugs, *Nature* **417**, 292–295.

Wittenberg, C., Reed, S.I. (1989) Conservation of function and regulation within the Cdc28/cdc2 protein kinase family: characterization of the human Cdc2Hs protein kinase in *Saccharomyces cerevisiae*, *Mol. Cell. Biol.* **9**, 4064–4068.

Wittmann, C.W., Wszolek, M.F., Shulman, J.M., Salvaterra, P.M., Lewis, J., Hutton, M., Feany, M.B. (2001) Tauopathy in *Drosophila*: neurodegeneration without neurofibrillary tangles, *Science* **293**, 711–714.

Wong, A.K., Chin, L. (2000) An inducible melanoma model implicates a role for RAS in tumor maintenance and angiogenesis, *Cancer Metastasis Rev.* **19**, 121–129.

Xia, H., Mao, Q., Paulson, H.L., Davidson, B.L. (2002) siRNA-mediated gene silencing in vitro and in vivo, *Nat. Biotechnol.* **20**, 1006–1010.

Xu, T., Wang, W., Zhang, S., Stewart, R.A., Yu, W. (1995) Identifying tumor suppressors in genetic mosaics: the *Drosophila lats* gene encodes a putative protein kinase, *Development* **121**, 1053–1063.

Yang, X., Li, D.M., Chen, W., Xu, T. (2001) Human homologue of *Drosophila* lats, LATS1, negatively regulates growth by inducing G(2)/M arrest or apoptosis, *Oncogene* **20**, 6516–6523.

Zan, Y., Haag, J.D., Chen, K.-S., Shepel, L.A., Wigington, D., Wang, Y.-R., Hu, R., Lopez-Guajardo, C.C., Brose, H.L., Porter, K.I., Leonard, R.A., Hitt, A.A., Schommer, S.L., Elegbede, A.F., Gould, M.N. (2003) Production of knockout rats using ENU mutagenesis and a yeast-based screening assay, *Nat. Biotechnol.* **21**, 645–651.

Expression Systems for DNA Processes

Ka-Yiu San and George N. Bennett
Rice University, Houston, Texas, USA

1 Introduction to Host–Vector Systems 357

2 Components of Plasmid Vectors 358
2.1 Origin of Replication 358
2.2 Selectable Markers 359
2.3 Transcription and Translation Signals 359
2.4 Restriction Sites for Cloning 361

3 Choice of Expression System 363
3.1 *Escherichia coli* 363
3.2 Other Hosts 364
3.2.1 *Bacillus subtilis* 365
3.2.2 Yeast 365
3.2.3 Mammalian Cells 367
3.2.4 Insect Virus 369
3.2.5 Vaccinia Virus 369
3.2.6 Transgenic Plants and Animals 369

4 Process Considerations 370

5 Perspectives 371

Bibliography 372
Books and Reviews 372
Primary Literature 372

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Keywords

Antibiotic Resistance
The ability of the cell to grow in the presence of a chemical (antibiotic) that normally inhibits an essential cell function.

Bacteriophage
A virus that infects bacteria.

Coding Region
The nucleic acid segment that contains the linear arrangement of codons specifying the order of amino acids in the encoded protein.

Codon
A three-nucleotide unit in a molecule of mRNA, which specifies the particular amino acid or stop signal in the mRNA as it undergoes translation.

Fusion Protein
A genetic construct that attaches a coding region of a protein or a portion of a protein to another coding sequence such that the protein formed during translation is a linear combination of the two proteins.

Inducible Promoter
A promoter in which the initiation of transcription by RNA polymerase is controlled by a regulatory protein, which can be activated or inactivated by a chemical or environmental stimulus.

Partition
The usual segregation of the genetic material of the cell into the two daughter cells.

Posttranslational Processing
Changes to a protein made after it is formed by translation, which add other chemical groups (e.g. phosphate, sugar) or cleave the chain to generate the mature native protein.

Promoter
The DNA site at which the RNA polymerase specifically binds to initiate transcription.

Protease
An enzyme that cleaves a peptide bond between amino acids comprising the protein.

Ribosome Binding Site
The region of the mRNA near the site of translation initiation that is important for ribosome recognition.

RNA Secondary Structure
A structure formed within an RNA molecule by bending the molecule such that a series of specific base pairs can form between nearby complementary nucleotides, forming a region that has double-stranded character.

Secretion
The process of controlled release of a protein through the cell membrane.

Virus
An agent naturally able to infect a host cell and bearing genes that allow its complete reproduction within a specific host cell.

The production of a specific protein by recombinant DNA technology entails cloning the gene that encodes the protein from the desired organism, forming a suitable genetic construct such that the gene can be expressed in a host organism, and introduction and maintenance of the construct in the host to allow the production of an adequate yield of functional protein. The ability to take a gene from any organism and express it in adequate quantity in a system permitting the analysis of the structure and function of the protein encoded by the gene has great importance in the research laboratory. Physical and enzymatic features of the protein encoded by a newly discovered gene can be studied, and its functional location in cells or tissues can be determined, in an effort to better define its biological function. Industrial applications include the production of therapeutic and diagnostic proteins having great impact in the field of medicine, and the production of stable bulk enzymes for the food and specialty chemical industries.

1 Introduction to Host–Vector Systems

To produce a specific RNA or protein product in a particular host cell, a suitable DNA construct must be prepared. This section discusses the essential features of expression systems that are widely used for overproduction of proteins for a variety of purposes. The expression system is composed of an expression vector and a specific host cell. Not only is the selection of the ideal expression vector important but the choice of the appropriate host also can affect production efficiency considerably. First, the expression vectors are considered, and then the important attributes of the host are discussed. A number of organisms have been used as hosts for expression of foreign proteins. Since, however, the most complete picture is available for *Escherichia coli*, this prokaryotic organism serves as an example and is discussed in more detail.

Expression vectors usually consist of small, circular plasmids specifically designed with several key features that allow a foreign gene inserted into the plasmid to

be expressed in the host cell. Important elements of the plasmid include (1) an origin of replication, which allows the plasmid to be replicated in the host, (2) a selectable genetic marker, which allows cells bearing the plasmid to preferentially grow on a specifically composed medium, (3) transcription and translation signals recognized by the host cell, and (4) a suitable unique restriction or recombination site located appropriately with regard to the transcription and translation signals, at which the foreign DNA can be introduced. In Sections 2 and 3, the generalized structure typical of many expression vectors is illustrated, and the variation and importance of each of these features in the overall expression system is explained.

2 Components of Plasmid Vectors

2.1 Origin of Replication

The origin of replication of the plasmid is a specific DNA sequence that defines the beginning of the replication of the circular plasmid DNA molecule. Because of its essential function in this event, its structure also limits the frequency with which replication is initiated. Replication frequency defines the copy number of the plasmid (i.e. the number of individual plasmid molecules present in each cell). This parameter, in turn, is important because of the gene dosage effect. If each plasmid carries a copy of the gene, and each copy can give rise to a certain maximal amount of product per unit time, the more copies of the plasmid present in the cell, the greater will be the level of production that can be attained per cell. Of course, there is a limit to this effect, and above a certain level the cell machinery becomes saturated and the increased metabolic burden of maintaining the plasmid overcomes the advantage of gene dosage in the production process.

The most widely used expression vectors for *E. coli* are derived from the origin of replication of the ColEl plasmid and have copy numbers between 10 and 500. The properties of high gene dosage have been used advantageously for the construction of temperature-sensitive copy number control systems in which the copy number can be raised dramatically by increasing the temperature. A typical prokaryotic expression vector is represented in Fig. 1.

Another feature specified by the origin of replication is the compatibility of the plasmid. Since it is sometimes desirable to maintain two distinct plasmids within the same cell, two different origins of replication types would be employed on the two different plasmids. If the two plasmids bear the same type of origin, one eventually will dominate without special selection and the other will be lost from a portion of the cells as the cells grow and divide. This phenomenon is called *incompatibility*, and plasmids with the same origin type are defined as being in the same incompatibility group.

Although, strictly speaking, not part of the origin of replication but often located nearby, elements on some plasmids specify a stability function known as a *partition locus*. The exploitation of naturally occurring partition systems that maintain copy number and select against plasmid loss has been a successful approach to plasmid maintenance in long-term cultures. An example is the incorporation of the *parB* locus on vectors. This locus selects strongly against plasmid loss by killing the daughter cells that no longer carry the

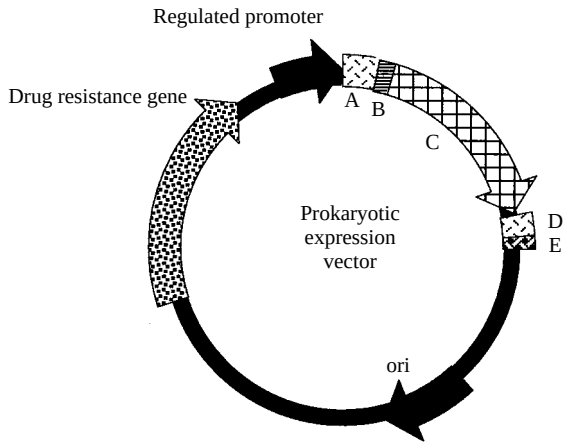

A = strong translation initiation signal
B = restriction sites for cloning
C = coding region of protein of interest
D = translation termination signal
E = strong transcription terminator
ori = origin of replication active in bacterial host

Fig. 1 A typical prokaryotic expression plasmid.

parB locus. This system does not require any specially constructed host strain.

2.2 Selectable Markers

The ability to select the vector and maintain the plasmid in the host cell is an important aspect of the expression system, especially in production regimens that involve a longer time course or continuous culture fermentations. The most common type of selectable marker incorporated on the plasmid is an antibiotic resistance element that allows selection with ampicillin, kanamycin, or tetracycline. Metabolic markers can also be used, for example, those allowing the cell to grow on a certain carbon source (e.g. sucrose) or those allowing the cell to grow without addition of an amino acid if the host usually requires it (e.g. a tryptophan requiring cell and a tryptophan marker).

Other approaches place an essential gene on the vector and use a special host strain in which that gene is deleted from the chromosome (e.g. the valine tRNA synthetase *valS* or the single-stranded DNA-binding protein *ssb*). If the plasmid is lost from the cell, the host cell alone is unable to grow. In such a system, addition of an antibiotic is not needed to select for the plasmid. This type of selection strategy is advantageous in reducing the cost of the media for large-scale production.

2.3 Transcription and Translation Signals

The transcription and translation signals allow the coding region of the gene to be efficiently recognized by the host cell machinery (i.e. RNA polymerase and ribosomes) that decodes the nucleotide sequence of the DNA of the gene into the appropriate sequence of amino acids in the protein. The signal specifying where

transcription starts is called the *promoter*. The more efficient the promoter, the greater the synthesis of messenger RNA of the gene under its control. Certain expression vectors employ a relatively strong promoter that is unregulated and is called a constitutive promoter. This approach is feasible if the protein to be made is very stable in the cell and does not cause any toxic effects or retard growth. More generally, however, a regulated promoter that can allow production of the desired protein at a suitable time in the production culture is used. This approach has two advantages: a short period of protein synthesis, and induction during the growth of the batch culture at the most opportune time (i.e. the period during which the optimal combination of cell density and specific protein production can be achieved).

The most widely used control systems for transcription initiation are those regulated by temperature or by the addition of a chemical agent or metabolite. A strong promoter from the bacteriophage lambda is normally prevented from binding to RNA polymerase by a repressor. If an altered form of the repressor is used, which is unstable at high temperature, an increase in temperature will release this negative control of the promoter, allowing large quantities of the desired messenger RNA to be synthesized. This promoter system can also be released from the control of the repressor by addition of DNA-damaging agents. This method is sometimes used instead of an increase in temperature.

Another widely used system is based on the repressor control of the lactose promoter. The DNA sequence bound by the lactose repressor can be combined with a variety of other stronger promoters (e.g. the tryptophan operon or lipoprotein gene promoters) by introducing the repressor binding site next to the promoter. In the new hybrid, the binding of the lac repressor blocks transcription, and the modified promoter then becomes controlled by the lactose repressor. The addition of an analogue of lactose, isopropylthiogalactoside, is most commonly used to induce the system.

A number of other regulation systems have been advocated for process control. Among these are phosphate, oxygen, acid, and arabinose. Each of these controlled systems can yield high expression under induced conditions, and particular conditions have been developed for each. Some are more appropriate for controlled levels of induction in smaller-scale laboratory experiments, and some have been specifically developed with the economics of large-scale culture in mind (e.g. oxygen). Some examples of regulatable promoters can incorporate a measure of feedback control into the host–vector circuit to balance expression to suit metabolic purposes. More elaborate combinations of regulatory genes controlled by different promoters can give rise to special responses or especially tight control of expression.

Some promoters are regulated by specific activators that stimulate transcription rather than by a negatively acting repressor system. An extension of the idea of regulating transcription initiation is the production of the desired messenger RNA through the use of a special RNA polymerase, not normally found in the cell, in combination with its unique promoter sequence. A system based on the bacteriophage T7 RNA polymerase and the T7 promoter it recognizes can yield very high specific production of the desired product. In this case, the T7 promoter would be located on the expression vector, and the T7 RNA polymerase gene is usually located on the chromosome under the control of a regulated system,

typically of the type just mentioned. Thus, induction in this system (e.g. by isopropylthiogalactoside addition to induce the lac repressor–controlled promoter) first produces the T7 RNA polymerase, which in turn synthesizes the specific messenger RNA corresponding to the gene of interest. Many variations of the T7 promoter expression system have been made and are commercially available.

Once a suitable means of producing the messenger RNA encoding the protein of interest has been defined, the next consideration is to translate the messenger RNA with optimal efficiency. Several factors affect this efficiency. One is the stability of the message. If the messenger RNA has a longer half-life in the cell, it can be translated proportionally more often before it is degraded. The presence of hairpin secondary structures, in which the nucleotide sequence allows intramolecular base pairing, particularly near the end of the message, seems to help stabilize the messenger RNA from degradation. Hairpin sequences act as part of the transcription termination signal and are normally included in the vector construct at the 3′ end of the message beyond the protein-coding sequence.

The major factor usually limiting translation is the ability of the ribosome to initiate translation on the messenger RNA. The ability of the ribosome to bind effectively to the messenger RNA can be impeded by secondary structures in the message, which block access to the translation initiation codon. The binding of the ribosome to the messenger RNA requires interaction between the ribosome, specifically including a segment of the ribosomal RNA, and the region of the messenger RNA about 10 nucleotides 5′ from the translation initiation codon. Optimal sequences for this ribosome binding region have been analyzed for a number of genes, and sequences corresponding to a highly efficient ribosome binding site are often included in the vector just preceding the position at which the protein-coding sequence is to be placed. Nucleotide sequences frequently placed in vectors for their efficient ribosome sites include chemically synthesized optimal consensus sequences, the T7 g10-L region, or those of other highly translated genes. While these standard translation initiation systems are suitable for adequate expression of most proteins, the sequence in this area often needs to be optimized to gain the highest level of production of a specific protein. Translational limitations can also occur as a result of the specific nature of the protein-coding region (e.g. unusual base composition, unique codon usage, or special secondary structures). These problems cannot be addressed at the expression vector level but sometimes can be addressed by selection of the host strain or by being optimized individually for that specific gene.

2.4 Restriction Sites for Cloning

Careful selection of the location of the restriction endonuclease cleavage site at which the protein-coding sequence is to be introduced into the plasmid can allow the expression of the protein in its most useful form. In many cases, it is desired to produce the complete intact natural protein, beginning with its N-terminal amino acid and ending with the native C-terminal amino acid. This is frequently the case when the exact structure of a mature, natural, pharmaceutically active protein is needed. To position the coding sequence correctly with respect to the ribosome binding site, a specific restriction endonuclease site incorporating an ATG

sequence can be used to place the translation initiation codon specifying methionine appropriately in the vector. Two commonly used restriction enzyme sites that contain an ATG and rarely exist in any given DNA sequence are *Nco* I and *Nde* I. A second unique restriction site is often used at which the C-terminal end of the gene is joined to the vector to orient the protein-coding segment during cloning. Such a construct would then enable translation of the complete cloned protein-coding region.

Although not all mature proteins have an N-terminal methionine, this is the primary product formed in bacterial systems. An extra N-terminal amino acid can be removed from the protein after synthesis by *in vitro* methods, or the desired protein can initially be made as part of a longer peptide chain with subsequent cleavage by a specific protease to generate the N-terminus of the mature protein product. A variety of N-terminal protein carrier elements have been used to make protein-fusion products (Fig. 2). In the formation of a fusion protein, the vector carries the appropriate signals for translation initiation as well as a coding region of the N-terminal component of the protein and a restriction endonuclease cleavage site at which the DNA fragment bearing the foreign gene can be placed in the proper reading frame. The properties of the N-terminal component of the protein fusion can be used to assist in purification of the fusion protein. For example, if this segment encodes an easily purified protein such as the maltose-binding protein or glutathione *S*-transferase, the combined fusion product can be readily isolated using affinity chromatography, which takes advantage of the binding properties of the attached protein component. The addition of traces of basic, acidic, or metal-binding amino acids (his-tag) in the carrier protein can also be used to aid in purification. This protein fusion concept can be utilized to allow for protein processing *in vivo*. For example, if the N-terminal segment encodes a localization signal, the protein can be directed to the periplasmic space of *E. coli* or, in some

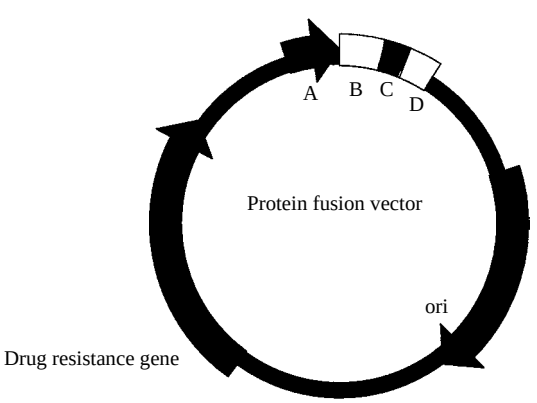

A = promoter
B = fusion tag for N-terminal fusion (e.g. glutathione *S*-transferase or oligohistidine tag)
C = encodes amino acid sequence for a cleavage site for a specific protease
D = restriction sites for placement of coding region of interest
ori = origin of replication

Fig. 2 A typical protein fusion vector.

cases, into the extracellular medium. This type of export can aid in the purification and stabilization of the protein. A vector employing protein A as the carrier portion is one of this type. C-terminal fusions can also be made where the added protein function is located at the 3′ end of the gene-coding sequence. There are also fusion-protein modules that can form a protease and cleave themselves out of the fusion protein to generate the proper linear protein product.

In some applications, the expression of the entire protein is not necessary. For raising antibodies or for the detection of an antigen by antibodies, only a short recognizable epitope is necessary; so in these cases, expression of only a short protein segment is sufficient. The cloned segment comprising the antigenic site or functional domain can be attached to a protein carrier segment that will allow effective exposure or activity of the foreign peptide. Placement of the foreign segment at the C-terminal end of the carrier protein has been used effectively for this purpose. In some cases, a suitable cell-surface protein can be used as the carrier portion to ensure that the foreign protein segment is localized to the surface of a cell or bacteriophage particle (phage display), thus presenting an immunologically active antigen or antibody species in a way that allows detection, isolation, or use of the living cell or virus particle. Such constructs are very useful for identifying protein segments that effectively bind specific molecules.

3 Choice of Expression System

Even considering the limited area of expression systems for *E. coli*, there is a wide choice of expression host–vector combinations. The system that is generally appropriate for a particular use is mentioned here. A few examples of specialized uses include the use of lambda vectors for ease in screening for expressed antigens in recombinant clones, phage-display vectors to look for specific epitopes or binding peptides, and vectors with simplified cloning systems for high-throughput analyses of many ORFs from genomic sequences. Specialized systems for examining protein–protein interactions in yeast or bacteria have been developed to identify protein domains that can bind to each other in functional complexes *in vivo*. *In vitro* expression systems can also be used, and these systems have improved and are available commercially to enable expression of protein from DNA in a small scale suitable for screening of a desired property of the protein if a suitably sensitive assay is available.

Families of related vectors are also available, which use a simplified recombination system to transfer the expressed gene among a variety of vectors suitable for different hosts or applications. This modular approach can allow a variety of hosts to be examined for protein production or to test the biological properties of the protein in different cells. These systems offered by various suppliers allow versatility in subsequent expression experiments without separate recloning steps.

3.1 *Escherichia coli*

The use of the correct host can have a significant impact on the final yield and quality of the desired protein product. In addition to the special genetic features that may be required for functioning of the plasmid vector system, there are other host features that are of general

importance. Factors generally useful for the host include the ability to grow rapidly and to a high density, the ability to be transformed in a manner that facilitates the introduction of the DNA construct, and the possession of a low recombination and mutagenic rate to ensure that the plasmid is not frequently lost, damaged, or otherwise inactivated. Stability to degradation is a major problem in the production of certain proteins. Therefore, host cells with reduced protease levels have been used to enhance the *in vivo* stability of the foreign protein of interest. Commonly used mutations in *E. coli* that have reduced protein degradation rates because of the inactivation of proteases are *lon, rpoH,* and *clp*. In some cases, the protein stability can be addressed at the expression vector level by using a fusion-protein construct. The presence of the longer carrier protein often will effectively stabilize a small foreign peptide segment. Other fusions (Nus) can be used to help solubilize the protein. Translation of large proteins may also be a limitation in many strains, but certain large proteins (e.g. polyketide synthases) have been produced in *E. coli*. The formation of insoluble inclusion bodies that contain an inactive form of the overexpressed protein has been a problem for some proteins. Strains with increased levels of chaperone proteins to help stabilize foreign proteins have been employed to attenuate these problems. Other efforts have included the use of thioredoxin, a redox protein, to improve proper folding of the protein via its use as a protein-fusion component.

3.2 Other Hosts

While *E. coli* continues to be the most widespread expression system, other host–vector systems have certain advantages. The strengths of the *E. coli* system are the variety of vectors and specifically altered hosts available, and the well-studied methods for genetically manipulating this organism. High levels of production can be attained. However, protease problems, formation of inactive inclusion bodies containing the product, and the lack of a eukaryotic glycosylation system limit production of a number of proteins from mammalian sources. Other bacteria have received some attention owing to their ability to grow on particular compounds, their potential for secretion of the protein, or their industrial potential.

Yeasts are a suitable production system for a number of processes. Not only are they well studied like *E. coli*, and amenable to scale-up, but also they are able to carry out some posttranslational processing of eukaryotic proteins such as specific cleavage or modification. Certain organisms used in the food industry are generally regarded as safe, and so less stringently purified material from such sources can be used in products for human consumption.

Several fungi are used to produce high yields of commercial proteins and are reasonably capable of glycosylation and secretion of proteins into the medium. Methods have not been so completely developed, however, and organisms of this class have not yet been as widely applied.

Viruses that infect insect cells (baculoviruses) have gained attention as a system for producing glycosylated proteins at reasonably high levels, but at less expense than is incurred using mammalian cells. While studies of the modification of proteins in various systems and the engineering of these posttranslational processing steps have made progress, this is still an area where significant improvement

can be made. The vaccinia virus has been used to express foreign antigens in whole animals, demonstrating its potential for use as a vaccine vector system. Although mammalian cell lines have been studied extensively and have the advantage of producing a truly identical processed mammalian protein, the costly, difficult scale-up acts as a limitation for commercial production with this system.

Finally, transgenic plants or animals can be used for protein production, and since the development time is long for these systems, they are suited for situations where long-term bulk production of a specific well-characterized product is warranted.

3.2.1 *Bacillus subtilis*

Bacillus subtilis has been considered to be a suitable host on the basis of several potentially advantageous properties. Considerable information is available concerning its molecular biology, physiology, and genetics. The details of translation, transcription, and plasmid systems are well studied. The employment of *B. subtilis* and related *Bacillus* species in other industrial production situations also has stimulated the development of *B. subtilis* for the production of recombinant proteins. Perhaps the most commercially interesting reason for using *B. subtilis* is the ability of this microorganism to secrete proteins into the medium. A simple system that would allow isolation of the protein from the supernatant broth directly, without the requirement for cell disruption, would simplify the downstream processing portion of the production process.

The expectation that *B. subtilis* would play a major role in the commercial production of recombinant proteins has been unrealized largely as a result of technical factors. Although regulated expression systems are available for *B. subtilis*, and secretion systems based on α-amylase, alkaline protease, or levansucrase have been reported, their use has been limited. The secretion system can be made to release some heterologous proteins into the medium, but not all of them are processed well. Another problem has been the presence of proteases that act to degrade the desired protein. This difficulty has been approached through the development of strains in which genes for several of the proteases are inactivated. Early observation of plasmid instability led to the development of integrated expression systems, in which the promoter and the gene to be expressed were recombined onto the bacterial chromosome. This strategy takes advantage of the active recombination system of *B. subtilis* to integrate the construct and thereby stabilize it. This approach to the formation of a stable overproducing strain may become more generally used in other organisms.

3.2.2 Yeast

Yeasts have many of the advantages of bacteria, for example, they are fast growing, high cell density is achieved, industrial-scale processes are established, and molecular biology and genetics are well developed; however, as eukaryotes they are capable of more sophisticated posttranslational protein processing than that found in bacteria. Yeasts have the transcription regulatory features and processing system of a eukaryotic cell. A general diagram of eukaryotic signals found in many expression vectors is presented in Fig. 3. The most commonly used yeast in laboratory research is *Saccharomyces cerevisiae*, an organism that can be cultured in the haploid state. Most industrial strains, however, have more than one copy of each chromosome per cell, which makes them more difficult to manipulate genetically.

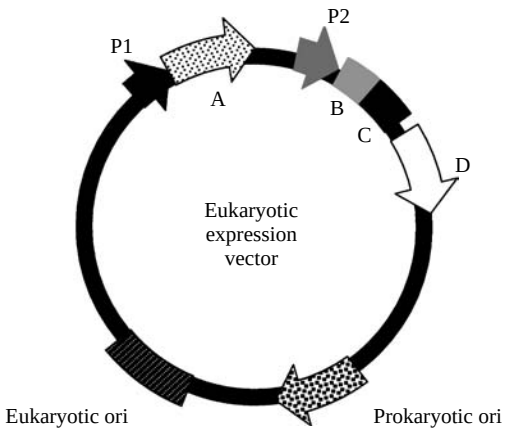

Fig. 3 A typical eukaryotic expression vector.

P1 = prokaryotic promoter
A = drug resistance for selection in prokaryotic systems
P2 = eukaryotic promoter
B = restriction sites for cloning
C = poly A feature
D = drug resistance element with expression features for selection in eukaryotic systems

Other yeasts, such as *Pichia pastoris* and *Hansenula polymorpha*, have also been investigated and may offer an advantage in the production of glycosylated proteins. *S. cerevisiae* vectors of either the plasmid or integrating type are available. The most common plasmid-based system derives from the 2 μm plasmid found in the nuclei at about 60 to 100 copies per cell. A large number of vectors have been derived from this plasmid, with the copy number varying depending on the size of the insert, the selective marker, and the growth conditions. Vectors employing promoters affected by the carbon source have been used to give high-level expression of heterologous proteins. Examples include various GAL4-regulated promoters, the alcohol dehydrogenase ADH2 promoter, the phosphoglycerate kinase promoter, the glyceraldehyde 3-phosphate dehydrogenase promoter, and temperature-regulated or copper-regulated systems.

The secretion and glycosylation abilities of yeast have received attention. The α-mating-factor secretion signal has been one of the most used, with the desired protein-coding sequence placed just after the α-factor segment specifying the signal for secretion and cleavage by protease. Yeast is capable of glycosylation, but the addition of sugar modifications to expressed human proteins is somewhat different from that found in the natural protein. For example, the protein produced in yeast often bears additional sugar residues at asparagine glycosylation sites. To reduce overglycosylation, *mnn* (mannan defective) or *alg* (asparagine-linked glycosylation) mutants of yeast have been studied and used as hosts for protein production. In efforts to improve the stability of proteins degraded by the vacuole proteinases, mutations that disrupt these protease genes have been introduced. The use of these special yeast strains has enhanced the ability

to isolate a number of proteins. Studies of the ubiquitin-mediated protein degradation process and the N-terminus removal or acetylation systems have led to the use of special strains or growth conditions to allow the desired N-terminus on the mature protein to be obtained. A major use of the yeast two-hybrid system for identifying interacting protein domains in functional genomics has been found, which can help elucidate possible functions for unknown proteins.

3.2.3 Mammalian Cells

Expression of proteins in mammalian systems affords production of the identical animal protein and thus eliminates any question regarding the effectiveness or natural character of the protein. Such factors as the expense of mammalian cell culture, low cell density, and low productivity have limited the use of this system to investigations in which the synthesis, processing, or role of the protein is being explored, or the protein cannot be satisfactorily expressed in a more amenable system.

While COS monkey cells are often used in combination with viral vectors for transient expression purposes for research experiments (Fig. 4), the development of a cell line capable of sustained production in a serum-free suspension culture is desired for long-term production. Suitable lines have been established in COS, CHO, mouse L cells, or other lines by linking the expression construct for the desired gene to a readily selected marker and integrating the entire DNA into a random chromosomal location (Fig. 5).

The expression construct usually contains an enhancer element, which allows the production of high-level expression (e.g. SV40, polyoma, or adenovirus enhancers; the heavy-metal-responsive metallothionein enhancer; glucocorticoid-responsive enhancers). An active promoter (e.g. the SV40 early promoter, the Rous sarcoma virus (RSV) promoter, adenovirus major late promoter) is commonly used, and appropriate efficient splicing

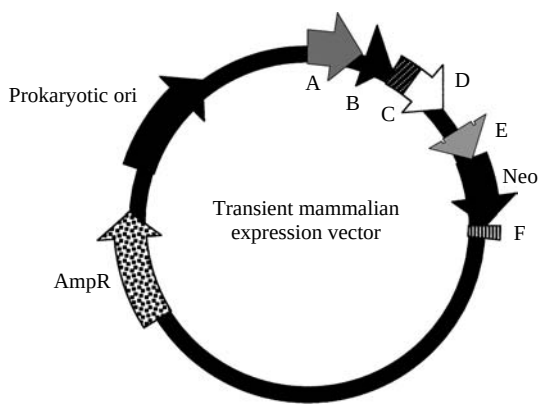

A = CMV enhancer/promoter
B = intron
C = restriction sites for cloning
D = poly (A) addition feature
E = SV40 enhancer/early promoter
F = poly (A) addition feature

Fig. 4 A transient expression vector for mammalian cells.

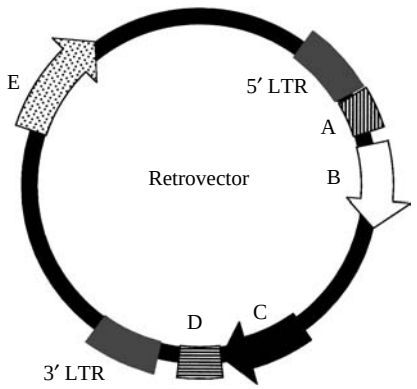

Fig. 5 A typical retrovector.

5′ LTR = long terminal repeat sequence which enables recombination for stable expression
A = retroviral packaging signal
B = drug resistance gene and associated expression features for selection in mammalian cells
C = inducible eukaryotic promoter element
D = restriction sites for placement of gene to be expressed
3′ LTR = long terminal repeat sequence which enables recombination for stable expression
E = drug resistance gene for selection in bacterial cells

signals derived from SV40, adenovirus, or immunoglobulin genes are also positioned on the expression vector along with polyadenylation signals.

To select for the presence of the integrated construct in the recipient cell line, a readily selectable gene is linked to the expression construct. A selectable marker that is frequently used is that specified by dihydrofolate reductase, which allows the cells to grow in the presence of methotrexate. Other selectable genes that can be used in a similar way are those for adenosine deaminase, ornithine decarboxylase, or asparagine synthetase, although in some cases a specific mutant cell line is required for selection. After the construct has been introduced into the appropriate cell line by calcium phosphate, DEAE-dextran, liposome-mediated transfection, or electroporation, the cells are grown under selection and with increasing levels of the selecting agent (e.g. methotrexate for dihydrofolate reductase gene amplification), and the genomic region at which the construct is integrated becomes amplified. This amplification provides a high gene dose in the genome of the host cell and increases the level of synthesis of the protein ten- to several hundredfold. Once formed, such cell lines can be stably maintained if care is taken to monitor for deletions. Improvements in the production level attained by single-copy integrated vectors and their adaptation to growth in serum-free suspension culture have simplified the use and prospects for mammalian cell production without modification.

For expression of a mutant form of a protein in a mammalian host, there are methods for homologous recombination, and these can be used to analyze the role of the protein in cell physiology.

3.2.4 Insect Virus

Insect virus expression systems for the production of proteins, particularly glycosylated mammalian proteins, have received attention because of the high level of expression (up to 50% of total cell protein) available with this technology. The baculovirus system employs strong transcription signals for expressing the polyhedron protein of the virus. These viruses replicate in the nucleus and form inclusion bodies, which are surrounded by a protective matrix containing a large amount of the polyhedron protein. The virus commonly used is *Autographa californica* nuclear polyhedrosis virus, and it is cultured on the cells of the fall armyworm, *Spodoperta frugiperda*. Silkworms can also be used as the host, providing a simple, low cost route for production.

This virus is so large (130 kb) that direct manipulation of the viral DNA is difficult, and the desired gene is usually introduced through *in vivo* recombination between the foreign gene construct and the viral genome. Newer methods for making the recombinant baculovirus at higher efficiency by incorporating a positive selection for the recombinants coupled with the introduction of reporter or detection systems have simplified the task of generating and identifying the desired viral expression constructs. Insect viruses have also been shown to infect a number of other cell types in the laboratory.

3.2.5 Vaccinia Virus

Vaccinia virus vectors have been used in the preparation of vaccines for animal diseases. The virus is a large poxvirus, which replicates in the cytoplasm of the host cell. While the virus is not a health hazard to humans, it is quite immunogenic and thus serves as a suitable carrier for the antigenic surface proteins of other viruses. The recombinant vaccinia virus is prepared by recombination *in vivo*, as is the case for the baculovirus system. A variety of markers are available for identification or selection of the recombinant viruses. These include β-galactosidase, thymidine kinase, neomycin resistance, and mycophenolic acid resistance. The introduced gene is placed under the control of a vaccinia virus promoter with appropriate signals, as in the case for expression in mammalian cells, and the host cells (e.g. CV-l cells) are coinfected with the vaccinia virus and the above-described construct bearing the desired gene, the expression signals, and the selective marker. After characterization, the recombinant vaccinia virus can be propagated on any suitable host. The antigenic components of several important disease-causing agents have been placed on the virus in this way (e.g. hepatitis B; rabies and herpes viruses). The recombinant vaccinia viruses seem to be able to produce an effective immune response to the foreign protein in whole animals.

3.2.6 Transgenic Plants and Animals

Whole plants or animals have been used as production systems. This strategy entails the preparation of the expression construct as for expression in mammalian cells and then introducing it into a cell that can be reproductively transmitted in a whole organism rather than into cultured cells. The integration event allows the altered chromosomal segment to exist in germ line cells. In this situation, the construct can then be transmitted to all the cells of the progeny of the original transgenic animal and on through succeeding generations.

The most widely used system for the production of a specific protein from transgenic animals takes advantage of the large amount of concentrated protein

secreted into milk by the mammary gland. The desired gene is attached to the synthesis and secretion signals used for natural production of the abundant milk protein, casein. This approach has been taken with sheep, goats, or cows, which are then able to secrete human interferon, α1-antitrypsin, or human factor IX in their milk. High levels of the protein can be produced (e.g. >10 g L^{-1} per liter of milk in some animals). The strategy enables long-term production and facilitates isolation of the protein in quantity without harming the animal. Once prepared and characterized, a herd of such animals would have a very low continuing production cost.

Transgenic plants can be created by the introduction of genes via infection by *Agrobacterium tumefaciens*, a bacterium that can pass the DNA of its Ti plasmid into plant cells, where this segment integrates into the plant genome. In nature, this transformation of plants by the bacterium generates a gall or tumorous callus on the plant at the site of the wound at which the infection occurred. By genetically engineering the Ti plasmid to remove the tumor-causing genes while retaining the DNA-transfer capability, useful vectors have been prepared. To effectively express a gene in the plant, suitable transcription and translation signals need to be incorporated surrounding the gene. Signals that have been used for this purpose are the nopaline synthetase promoter, the promoter from the gene for the small subunit of the carbon dioxide–fixing enzyme, ribulose bisphosphate carboxylase, and those derived from the cauliflower mosaic virus. In some instances, the gene can be expressed in specific locations in the plant where the metabolic consequences of the protein may be most desired.

The generation of a transgenic plant calls for the use of cells that are susceptible to infection and are competent in regenerating a plant through culturing techniques. These two criteria have limited the system for some plants. *Agrobacterium tumefaciens* infects dicotyledonous plants, a group that does not include most cereal plants. For the cereals and many other plants, DNA can be introduced into the cells by coating it onto an inert projectile and firing it into plant leaves or embryogenic calluses. Recombination followed by propagation from single cells under appropriate selection (usually kanamycin or phosphinothricin) can generate the transgenic plant. Fertile transgenic rice, corn, soybean and wheat plants have been formed. So far, the genes introduced into plants have been mainly intended to increase resistance to herbicides, insects, or viruses, and these are in wide commercial use (e.g. corn and cotton). Such properties as ripening, taste, or nutritional characteristics can also be altered.

The expression of industrially useful proteins in bulk in the seeds of plants has also been possible. This could have advantages for large-scale production of stable proteins and has been proposed as a method for generating antigens suitable as vaccines.

Viral vectors can also be used with certain plant species (e.g. cowpea mosaic virus), and they have been used to express animal virus surface antigens in whole plants. This approach is attractive in that the virus is easily isolated and stored and is very immunogenic but cannot harmfully infect animals.

4 Process Considerations

The synthesis of protein from cloned genes relies solely on the metabolic machinery

of the host cell. Consequently, any process factors that affect the host cell physiology will have a significant impact on the ultimate expression level. Factors such as temperature, impeller agitation speed, and medium formulation or feeding strategy in the case of a fed-batch fermentation are important variables to be considered in arriving at a production process. In industrial processes, where high volumetric productivity is the desired goal, extensive optimization studies are performed with these variables to achieve maximal overproduction of the specific protein in the shortest practical time.

Several general themes from the vast experimental observations of process conditions on *E. coli* protein production are mentioned here. The composition of the growth medium can affect both cell growth and recombinant protein yield. A complex medium generally yields a higher productivity. This is due to the presence of biosynthetic intermediates and cofactors, which the cell does not then need to synthesize, thus decreasing the metabolic burden. A balanced carbon-to-nitrogen ratio in the medium is also critical in achieving a high yield of recombinant protein. The temperature has a significant effect on cell growth and the waste products accumulated. The formation of acetate, a harmful waste product, is increased at high temperature. Lower temperature generally favors the formation of soluble protein, while higher temperature encourages the accumulation of the protein in inclusion bodies.

Another challenging task is to provide an adequate supply of oxygen to the culture, especially for dense cultures. The transfer of oxygen from the air stream can be increased, to a certain extent, by increasing the agitation speed. Oxygen supplementation is required in some cases, and genetic manipulation of the host by introduction of an oxygen-binding protein has also given positive results.

Waste product accumulation has a detrimental effect on recombinant protein production. Strategies to limit the effects of waste products have included lowering the temperature of the culture, carefully controlling the nutrient supply in a fed-batch system, and the use of genetically altered hosts that exhibit reduced levels of harmful metabolic products.

5 Perspectives

The ability to clone and express high levels of specific proteins in other organisms has led to great advances in the speed and detail with which biological systems can be analyzed. With the expansion of this technology to other organisms and the construction of more complicated and sophisticated derivatives of currently studied systems, the impact of this area will increase. Recombinant protein production has now become widely applicable with industrial microorganisms, the formation and use of transgenic plants and animals, and the analysis of unknown proteins associated with genetic disorders. The area of proteomics and the efforts to decipher the function of the coding regions of sequenced genomes will lead to an even greater use of expression systems and the development of more efficient specialized systems.

See also DNA Replication and Transcription; Gene Targeting.

Bibliography

Books and Reviews

Balbas, P. (2001) Understanding the art of producing protein and nonprotein molecules in *Escherichia coli*, *Mol. Biotechnol.* **19**, 251–267.

Baneyx, F. (1999) Recombinant protein expression in *Escherichia coli*, *Curr. Opin. Biotechnol.* **10**, 411–421.

Barr, P.J., Brake, A.J., Valenzuela, P. (Eds.) (1989) *Yeast Genetic Engineering*, Butterworths, Boston.

Curiel, D., Douglas, J.T. (Eds.) (2002) *Adenoviral Vectors for Gene Therapy*, Academic Press, New York.

Goeddel, D.V. (Ed.) (1990) *Gene Expression Technology*, *Methods in Enzymology*, Vol. 185, Academic Press, San Diego, CA.

Hauser, H., Wagner, R., Hauser, H. (Eds.) (1997) *Mammalian Cell Biotechnology in Protein Production*, de Gruyter, Berlin.

Higgins, S.J., Hames, B.D. (Eds.) (1999) *Protein Expression: A Practical Approach*, Oxford University Press, Oxford.

Hood, E.E., Howard, J.A. (Eds.) (2002) *Plants as Factories for Protein Production*, Kluwer, New York.

Makrides, S.C. (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*, *Microbiol. Rev.* **60**, 512–538.

Mattanovich, D., Weik, R., Thim, S., Kramer, W., Bayer, K., Katinger, H. (1996) Optimization of recombinant gene expression in *Escherichia coli*, *Ann. N.Y. Acad. Sci.* **782**, 182–190.

Merten, O.-W., Mattanovich, D., Lang, C., Larsson, G. (Eds.) (2001) *Recombinant Protein Production with Prokaryotic and Eukaryotic Cells: A Comparative View on Host Physiology*, Kluwer, New York.

O'Reilly, D.R., Luckow, V.A., Miller, L.K. (1992) *Baculovirus Expression Vectors–A Laboratory Manual*, Freeman, New York.

Reznikoff, W., Gold, L. (Eds.) (1986) *Maximizing Gene Expression*, Butterworths, Boston.

Swartz, J.R. (2001) Advances in *Escherichia coli* production of therapeutic proteins, *Curr. Opin. Biotechnol.* **12**, 195–201.

Thorner, J., Emr, S.D., Abelson, J.N. (Eds.) (2000) Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification, Methods in Enzymology, Vol. 326, Academic Press, New York.

Vaillancourt, P.E. (2002) *E. coli Gene Expression Protocols*, Humana Press, New York.

Weickert, M.J., Doherty, D.H., Best, E.A., Olins, P.O. (1996) Optimization of heterologous protein production in *Escherichia coli*, *Curr. Opin. Biotechnol.* **7**, 494–499.

Primary Literature

Abrahmsen, L., Moks, T., Nilsson, B., Uhlen, M. (1986) Secretion of heterologous gene products to the culture medium of *Escherichia coli*, *Nucleic Acids Res.* **14**, 7487–7500.

Amann, E., Ochs, B., Abel, K.J. (1988) Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*, *Gene* **69**, 301–315.

Bai, C., Elledge, S.J. (1996) Gene identification using the yeast two-hybrid system, *Methods Enzymol.* **273**, 331–347.

Baker, R.T. (1996) Protein expression using ubiquitin fusion and cleavage, *Curr. Opin. Biotechnol.* **7**, 541–546.

Bass, S.H., Yansura, D.G. (2000) Application of the *E. coli* trp promoter, *Mol. Biotechnol.* **16**, 253–260.

Bignell, C., Thomas, C.M. (2001) The bacterial ParA-ParB partitioning proteins, *J. Biotechnol.* **91**, 1–34.

Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E., LaBaer, J. (2002) Proteome-scale purification of human proteins from bacteria, *Proc. Natl. Acad. Sci. USA* **99**, 2654–2659.

Brizuela, L., Braun, P., LaBaer, J. (2001) FLEX gene repository: from sequenced genomes to gene repositories for high-throughput functional biology and proteomics, *Mol. Biochem. Parasitol.* **118**, 155–165.

Carrier, T., Jones, K.L., Keasling, J.D. (1998) mRNA stability and plasmid copy number effects on gene expression from an inducible promoter system, *Biotechnol. Bioeng.* **20**, 666–672.

Carrier, T.A., Keasling, J.D. (1999) Library of synthetic 5′ secondary structures to manipulate mRNA stability in *Escherichia coli*, *Biotechnol. Prog.* **15**, 58–64.

Carroll, M.W., Moss, B. (1997) Poxviruses as expression vectors, *Curr. Opin. Biotechnol.* **8**, 573–577.

Cereghino, J.L., Cregg, J.M. (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*, *FEMS Microbiol. Rev.* **24**, 45–66.

Chauthaiwale, V.M., Therwath, A., Deshpande, V.V. (1992) Bacteriophage lambda as a cloning vector, *Microbiol. Rev.* **56**, 577–591.

Cheng, X., Patterson, T.A. (1992) Construction and use of lambda PL promoter vectors for direct cloning and high level expression of PCR amplified DNA coding sequences, *Nucleic Acids Res.* **20**, 4591–4598.

Chevalet, L., Robert, A., Gueneau, F., Bonnefoy, J.Y., Nguyen, T. (2000) Recombinant protein production driven by the tryptophan promoter is tightly controlled in ICONE 2000, a new genetically engineered *E. coli* mutant, *Biotechnol. Bioeng.* **69**, 351–358.

Christensen, A.C. (2001) Bacteriophage lambda-based expression vectors, *Mol. Biotechnol.* **17**, 219–224.

Cramer, C.L., Boothe, J.G., Oishi, K.K. (1999) Transgenic plants for therapeutic proteins: linking upstream and downstream strategies, *Curr. Top. Microbiol. Immunol.* **240**, 95–118.

Cserjan-Puschmann, M., Kramer, W., Duerrschmid, E., Striedner, G., Bayer, K. (1999) Metabolic approaches for the optimization of recombinant fermentation processes, *Appl. Microbiol. Biotechnol.* **53**, 43–50.

Curry, K.A., Tomich, C.S. (1988) Effect of ribosome binding site on gene expression in *Escherichia coli*, *DNA* **7**, 173–179.

Daugherty, P.S., Olsen, M.J., Iverson, B.L., Georgiou, G. (1999) Development of an optimized expression system for the screening of antibody libraries displayed on the *Escherichia coli* surface, *Protein Eng.* **12**, 613–621.

de Smit, M.H., van Duin, J. (1990) Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis, *Proc. Natl. Acad. Sci. USA* **87**, 7668–7672.

De Sutter, K., Hostens, K., Vandekerckhove, J., Fiers, W. (1994) Production of enzymatically active rat protein disulfide isomerase in *Escherichia coli*, *Gene* **141**, 163–170.

Dieci, G., Bottarelli, L., Ballabeni, A., Ottonello, S. (2000) tRNA-assisted overproduction of eukaryotic ribosomal proteins, *Protein Expr. Purif.* **18**, 346–354.

Donovan, R.S., Robinson, C.W., Glick, B.R. (1996) Review: optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter, *Ind. Microbiol.* **16**, 145–154.

Drabner, B., Guzman, C.A. (2001) Elicitation of predictable immune responses by using live bacterial vectors, *Biomed. Eng.* **17**, 75–82.

Emory, S.A., Bouvet, P., Belasco, J.G. (1992) A 5′-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*, *Genes Dev.* **6**, 135–148.

Farmer, W.R., Liao, J.C. (2000) Improving lycopene production in *Escherichia coli* by engineering metabolic control, *Nat. Biotechnol.* **18**, 533–537.

Fields, S., Song, O. (1989) A novel genetic system to detect protein-protein interactions, *Nature* **340**, 245–246.

Friehs, K., Reardon, K.F. (1993) Parameters influencing the productivity of recombinant *E. coli* cultivations, *Adv. Biochem. Eng. Biotechnol.* **48**, 53–77.

Fuchs, E. (1999) The translation initiation signal in *E. coli* and its control, *Genet. Eng. (NY)* **21**, 15–35.

Fussenegger, M., Bailey, J.E., Hauser, H., Mueller, P.P. (1999) Genetic optimization of recombinant glycoproteins production by mammalian cells, *Trends Biotechnol.* **17**, 35–42.

Geisse, S., Gram, H., Kleuser, B., Kocher, H.P. (1996) Eukaryotic expression systems: a comparison, *Protein Expr. Purif.* **8**, 271–282.

Georgiou, G., Stathopoulos, C., Daugherty, P.S., Nayak, A.R., Iverson, B.L., Curtiss, R. III (1997) Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines, *Nat. Biotechnol.* **15**, 29–34.

Georgiou, G., Valax, P. (1996) Expression of correctly folded proteins in *Escherichia coli*, *Curr. Opin. Biotechnol.* **7**, 190–197.

Ghrayeb, J., Kimura, H., Takahara, M., Hsiung, H., Masui, Y., Inouye, M. (1984) Secretion cloning vectors in *Escherichia coli*, *EMBO J.* **3**, 2437–2442.

Giddings, G., Allison, G., Brooks, D., Carter, A. (2000) Transgenic plants as factories for biopharmaceuticals, *Nat. Biotechnol.* **18**, 1151–1155.

Grabherr, R., Nilsson, E., Striedner, G., Bayer, K. (2002) Stabilizing plasmid copy number to improve recombinant protein production, *Biotechnol. Bioeng.* **77**, 142–147.

Guan, K.L., Dixon, J.E. (1991) Eukaryotic proteins expressed in *Escherichia coli*: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase, *Anal. Biochem.* **192**, 262–267.

Guzman, L.M., Belin, D., Carson, M.J., Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter, *J. Bacteriol.* **17**, 4121–4130.

Han, S.J., Chang, H.N., DeMoss, J.A., Suh, E.J., Lee, J. (2000) Development and characterization of an oxygen-dependent inducible promoter system, the modified nar promoter in a mutant *Escherichia coli*, *Biotechnol. Bioeng.* **68**, 115–120.

Hartl, B., Wehrl, W., Wiegert, T., Homuth, G., Schumann, W. (2001) Development of a new integration site within the *Bacillus subtilis* chromosome and construction of compatible expression cassettes, *J. Bacteriol.* **183**, 2696–2699.

Hartley, J.L., Temple, G.F., Brasch, M.A. (2000) DNA cloning using in vitro site-specific recombination, *Genome Res.* **10**, 1788–1795.

Hensing, M.C., Rouwenhorst, R.J., Heijnen, J.J., van Dijken, J.P., Pronk, J.T. (1995) Physiological and technological aspects of large-scale heterologous-protein production with yeasts, *Antonie Van Leeuwenhoek* **67**, 261–279.

Hernan, R.A., Hui, H.L., Andracki, M.E., Noble, R.W., Sligar, S.G., Walder, J.A., Walder, R.Y. (1992) Human hemoglobin expression in *Escherichia coli*: importance of optimal codon usage, *Biochemistry* **31**, 8619–8628.

Houdebine, L.M. (2000) Transgenic animal bioreactors, *Trans. Res.* **9**, 305–320.

Hui, A.S., Eaton, D.H., de Boer, H.A. (1988) Mutagenesis at the mRNA decoding site in the 16S ribosomal RNa using the specialized ribosome system in *Escherichia coli*, *EMBO J.* **7**, 4383–4388.

Janknecht, R., de Martynoff, G., Lou, J., Hipskind, R.A., Nordheim, A., Stunnenberg, H.G. (1991) Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus, *Proc. Natl. Acad. Sci. USA* **88**, 8972–8976.

Kane, J.F. (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*, *Curr. Opin. Biotechnol.* **6**, 494–500.

Kellems, R.E. (1991) Gene amplification in mammalian cells: strategies for protein production, *Curr. Opin. Biotechnol.* **2**, 723–729.

Khlebnikov, A., Rosa, O., Skaug, T., Carrier, T.A., Keasling, J.D. (2000) Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture, *J. Bacteriol.* **182**, 7029–7034.

Khosla, C., Bailey, J.E. (1988) Heterologous expression of a bacterial haemoglobin improves the growth properties of recombinant *Escherichia coli*, *Nature* **331**, 633–635.

Khosla, C., Curtis, J.E., Bydalek, P., Swartz, J.R., Bailey, J.E. (1990) Expression of recombinant proteins in *Escherichia coli* using an oxygen-responsive promoter, *Biotechnology (NY)* **8**, 554–558.

Kirinaka, H., Kamihira, M., Iijima, S., Kobayashi, T. (1994) A new runaway type episomal vector for mammalian cells based on a temperature-sensitive simian virus 40 and inducible erythropoietin production, *Appl. Microbiol. Biotechnol.* **41**, 591–596.

Kjeldsen, T. (2000) Yeast secretory expression of insulin precursors, *Appl. Microbiol. Biotechnol.* **54**, 277–286.

Kleber-Janke, T., Becker, W.M. (2000) Use of modified BL21 (DE3) *Escherichia coli* cells for high-level expression of recombinant peanut allergens affected by poor codon usage, *Protein Expr. Purif.* **19**, 419–424.

Kohrer, C., Mayer, C., Grobner, P., Piendl, W. (1996) Use of T7 RNA polymerase in an optimized *Escherichia coli* coupled in vitro transcription-translation system. Application in regulatory studies and expression of long transcription units, *Eur. J. Biochem.* **236**, 234–239.

Kost, T.A., Condreay, J.P. (1999) Recombinant baculoviruses as expression vectors for insect and mammalian cells, *Curr. Opin. Biotechnol.* **10**, 428–433.

La Vallie, E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F., McCoy, J.M. (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm, *Biotechnology (NY)* **11**, 187–193.

Lam, K.H., Chow, K.C., Wong, W.K. (1998) Construction of an efficient *Bacillus subtilis* system for extracellular production of heterologous proteins, *J. Biotechnol.* **63**, 167–177.

Larrick, J.W., Yu, L., Naftzger, C., Jaiswal, S., Wycoff, K. (2001) Production of secretory IgA antibodies in plants, *Biomol. Eng.* **18**, 87–94.

Larsson, M., Brundell, E., Nordfors, L., Hoog, C., Uhlen, M., Stahl, S. (1996) A general bacterial expression system for functional analysis of cDNA-encoded proteins, *Protein Expr. Purif.* **7**, 447–457.

Latta, M., Philit, M., Maury, I., Soubrier, F., Denefle, P., Mayaux, J.F. (1990) Tryptophan promoter derivatives on multicopy plasmids: a comparative analysis of expression potentials in *Escherichia coli*, *DNA Cell Biol.* **9**, 129–137.

Lesley, S.A. (2001) High-throughput proteomics: protein expression and purification in the postgenomic world, *Protein Expr. Purif.* **22**, 159–164.

Liu, Q., Li, M.Z., Leibham, D., Cortez, D., Elledge, S.J. (1998) The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes, *Curr. Biol.* **8**, 1300–1309.

Luckow, V.A., Summer, M.D. (1989) High level expression of nonfused foreign genes with *Autographa californica* nuclear polyhedrosis virus expression vectors, *Virology* **170**, 31–39.

Maina, C.V., Riggs, P.D., Grandea, A.G. III, Slatko, B.E., Moran, L.S., Tagliamonte, J.A., McReynolds, L.A., Guan, C.D. (1988) An *Escherichia coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein, *Gene* **74**, 365–373.

Marchal, I., Jarvis, D.L., Cacan, R., Verbert, A. (2001) Glycoproteins from insect cells: sialylated or not? *Biol. Chem.* **382**, 151–159.

Martin, G.A., Kawaguchi, R., Lam, Y., DeGiovanni, A., Fukushima, M., Mutter, W. (2001) High-yield, in vitro protein expression using a continuous-exchange, coupled transcription/translation system, *Biotechniques* **31**, 948–950, 952–953.

Mattanovich, D., Kramer, W., Luttich, C., Weik, R., Bayer, K., Katinger, H. (1998) Rational design of an improved induction scheme for recombinant *Escherichia coli*, *Biotechnol. Bioeng.* **58**, 296–298.

Meerman, H.J., Georgiou, G. (1994) Construction and characterization of a set of *E. coli* strains deficient in all known loci affecting the proteolytic stability of secreted recombinant proteins, *Biotechnology* **12**, 1107–1110.

Mendoza-Vega, O., Sabatie, J., Brown, S.W. (1994) Industrial production of heterologous proteins by fed-batch cultures of the yeast *Saccharomyces cerevisiae*, *FEMS Microbiol. Rev.* **15**, 369–410.

Mikawa, Y.G., Maruyama, I.N., Brenner, S. (1996) Surface display of proteins on bacteriophage lambda heads, *J. Mol. Biol.* **262**, 21–30.

Min, K.T., Kim, M.H., Lee, D.S. (1988) Search for the optimal sequence of the ribosome binding site by random oligonucleotide-directed mutagenesis, *Nucleic Acids Res.* **16**, 5075–5088.

Miroux, B., Walker, J.E. (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels, *J. Mol. Biol.* **260**, 289–298.

Moss, B. (1996) Genetically engineered poxviruses for recombinant gene expression, vaccination, and safety, *Proc. Natl. Acad. Sci. USA* **93**, 11341–11348.

Murby, M., Uhlen, M., Stahl, S. (1996) Upstream strategies to minimize proteolytic degradation upon recombinant production in *Escherichia coli*, *Protein Expr. Purif.* **7**, 129–136.

Nakamura, K., Inouye, M. (1982) Construction of versatile expression cloning vehicles using the lipoprotein gene of *Escherichia coli*, *EMBO J.* **1**, 771–775.

Noren, K.A., Noren, C.J. (2001) Construction of high-complexity combinatorial phage display peptide libraries, *Methods* **23**, 169–178.

Novick, R.P. (1987) Plasmid incompatibility, *Microbiol. Rev.* **51**, 381–395.

O'Connor, C.D., Timmis, K.N. (1987) Highly repressible expression system for cloning genes that specify potentially toxic proteins, *J. Bacteriol.* **169**, 4457–4462.

Olins, P.O., Devine, C.S., Rangwala, S.H., Kavka, K.S. (1988) The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli*, *Gene* **73**, 227–235.

Parent, S.A., Fenimore, C.M., Bostian, K.A. (1985) Vector systems for the expression, analysis and cloning of DNA sequences in *S. cerevisiae*, *Yeast* **1**, 83–138.

Patnaik, R., Swartz, J.R. (1998) *E. coli*-based in vitro transcription/translation: in vivo-specific

synthesis rates and high yields in a batch system, *Biotechniques* **24**, 862–868.

Pilon, A., Yost, P., Chase, T.E., Lohnas, G., Burkett, T., Roberts, S., Bentley, W.E. (1997) Ubiquitin fusion technology: bioprocessing of peptides, *Biotechnol. Prog.* **13**, 374–349.

Pines, O., Inouye, M. (1999) Expression and secretion of proteins in *E. coli*, *Mol. Biotechnol.* **12**, 25–34.

Porter, R.D., Black, S., Pannuri, S., Carlson, A. (1990) Use of the *Escherichia coli* SSB gene to prevent bioreactor takeover by plasmidless cells, *Biotechnology (NY)* **8**, 47–51.

Possee, R.D. (1997) Baculoviruses as expression vectors, *Curr. Opin. Biotechnol.* **8**, 569–572.

Ramirez, D.M., Bentley, W.E. (1999) Characterization of stress and protein turnover from protein overexpression in fed-batch *E. coli* cultures, *J. Biotechnol.* **71**, 39–58.

San, K.Y., Bennett, G.N., Chou, C.H., Aristidou, A.A. (1994) An optimization study of a pH-inducible promoter system for high-level recombinant protein production in *Escherichia coli*, *Ann. N.Y. Acad. Sci.* **721**, 268–276.

Sears, I.B., O'Connor, J., Rossanese, O.W., Glick, B.S. (1998) A versatile set of vectors for constitutive and regulated gene expression in *Pichia pastoris*, *Yeast* **14**, 783–790.

Sharrocks, A.D. (1994) A T7 expression vector for producing N- and C-terminal fusion proteins with glutathione S-transferase, *Gene* **138**, 105–108.

Shuman, S. (1994) Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase, *J. Biol. Chem.* **269**, 32678–32684.

Skogman, S.G., Nilsson, J. (1984) Temperature-dependent retention of a tryptophan-operon-bearing plasmid in *Escherichia coli*, *Gene* **31**, 117–122.

Smajs, D., McKevitt, M., Wang, L., Howell, J.K., Norris, S.J., Palzkill, T., Weinstock, G.M. (2002) BAC library of *T. pallidum* DNA in *E. coli*, *Genome Res.* **12**, 515–522.

Smith, G.L., Symons, J.A., Alcami, A. (1999) Immune modulation by proteins secreted from cells infected by vaccinia virus, *Arch. Virol. Suppl.* **15**, 111–129.

Sobecky, P.A., Easter, C.L., Bear, P.D., Helinski, D.R. (1996) Characterization of the stable maintenance properties of the par region of broad-host-range plasmid RK2, *J. Bacteriol.* **178**, 2086–2093.

Su, T.Z., Schweizer, H., Oxender, D.L. (1990) A novel phosphate-regulated expression vector in *Escherichia coli*, *Gene* **90**, 129–133.

Tabor, S., Richardson, C.C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes, *Proc. Natl. Acad. Sci. USA* **82**, 1074–1078.

Thomas, J.G., Ayling, A., Baneyx, F. (1997) Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from *E. coli*. To fold or to refold, *Appl. Biochem. Biotechnol.* **66**, 197–238.

Tong, A.H., Drees, B., Nardelli, G., Bader, G.D., Brannetti, B., Castagnoli, L., Evangelista, M., Ferracuti, S., Nelson, B., Paoluzi, S., Quondam, M., Zucconi, A., Hogue, C.W., Fields, S., Boone, C., Cesareni, G. (2002) A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules, *Science* **295**, 321–324.

Uhlin, B.E., Molin, S., Gustafsson, P., Nordstrom, K. (1979) Plasmids with temperature-dependent copy number for amplification of cloned genes and their products, *Gene* **6** 91–106.

Vennaloweth, R.L., Rabinowitz, J.C. (1992) The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli* in vivo, *Mol. Microbiol.* **6**, 1105–1114.

Wang, H.Y., McConnell, D.J., O'Mahony, D.J. (1990) An efficient temperature-inducible vector incorporating the T7 gene 10 translation initiation leader region, *Nucleic Acids Res.* **18**, 1070.

Wang, L.-F., Doi, R.H. (1992) Heterologous Gene Expression in *Bacillus subtilis*, in: Doi, R.H., McGloughlin, M. (Eds.) *Biology of Bacilli: Applications to Industry*, Butterworth-Heinemann, Boston, pp. 63–104.

Weikert, C., Sauer, U., Bailey, J.E. (1998) An *Escherichia coli* host strain useful for efficient overproduction of secreted recombinant protein, *Biotechnol. Bioeng.* **59**, 386–391.

Westerlund-Wikstrom, B. (2000) Peptide display on bacterial flagella: principals and applications, *Int. J. Med. Microbiol.* **290**, 223–230.

Wong, S.L. (1995) Advances in the use of *Bacillus subtilis* for the expression and secretion of heterologous proteins, *Curr. Opin. Biotechnol.* **6**, 517–522.

Wurm, F., Bernard, A. (1999) Large-scale transient expression in mammalian cells for recombinant protein production, *Curr. Opin. Biotechnol.* **10**, 156–159.

Fat Cell: *see* Adipocytes

Female Reproduction System, Molecular Biology of

Guck T. Ooi and Jock K. Findlay
Prince Henry's Institute of Medical Research, Victoria, Australia

1 Principles: Regulation of Reproductive Processes in the Female 383
1.1 The Hypothalamic-pituitary-ovarian-uterine Axis 383
1.2 Levels of Regulation 384
1.3 Specificities of Actions 385

2 Examples of Regulatory Systems 385
2.1 The Steroidogenic System 385
2.1.1 Estrogen and its Receptors 385
2.1.2 Roles of Estrogen in Ovarian Functions 388
2.2 Gonadotropin System 388
2.2.1 Gonadotropins and their Receptors 388
2.2.2 Roles of Gonadotropins in Ovarian Functions 390
2.3 The Transforming Growth Factor β (TGF-β) Superfamily System 391
2.3.1 TGF-β Superfamily 391
2.3.2 TGF-β Superfamily Receptors 396
2.4 Insulin-like Growth Factor (IGF) System 399
2.4.1 IGFs, IGF-binding Proteins (IGFBPs) and IGF-receptors 399
2.4.2 Gene Structure and Expression in the Ovary 399
2.4.3 Roles of IGFs in Ovarian Function 403
2.5 Other Peptides/Enzymes 403
2.5.1 Endothelins 403
2.5.2 Oxytocin 404
2.5.3 Zona Pellucida Proteins 405
2.5.4 Leukemia Inhibitory Factor 405
2.5.5 Tissue Remodeling Enzymes 406

3 Pathophysiological Considerations 406
3.1 Genetic Defects 406
3.2 FSH Isoforms and Infertility 408
3.3 Reproductive Tumors and Inhibin 408

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

4 Perspectives 409

Bibliography 410
Books and Reviews 410
Primary Literature 411

Keywords

Betaglycan
A membrane, proteoglycan, originally discovered to be an accessory receptor for TGF-β2 (hence, also called TGF-β type III receptor), but now shown to bind inhibin with high affinity in a complex with activin type II receptor, thereby antagonizing activin actions.

BMPs
Bone morphogenetic proteins are bone-growth regulatory factors that belong to the TGF-β superfamily. They are synthesized as large precursor molecules, that are cleaved by proteolytic enzymes. The active form can consist of a dimer of two identical proteins or a heterodimer of two related bone morphogenetic proteins.

Endocrine Regulation
Regulation by a hormone made by an organ and secreted into the bloodstream, where it circulates and influences specific target organs.

Endothelins
Peptides of 21 amino acids with potent vasoconstrictor as well as mitogenic and hormone-releasing properties, made by a variety of epithelial and vascular smooth muscle cells.

Folliculogenesis
A developmental process in which follicles undergo a coordinated progression of recruitment, development (or atresia), selection, and ovulation.

FSH Isoforms
Forms of follicle-stimulating hormone that differ in the degree of glycosylation of the mature protein leading to differences in their biochemical properties and circulating half-lives in the blood.

G-Proteins
Coupling proteins, so-called because they bind GTP, which link receptors to effector molecules. They regulate a variety of enzymes and ion channels. The G-proteins form a heterotrimeric complex of α, β, and γ subunits. There are multiple members of each family of subunits, giving rise to a complexity of responses.

Gonadotropin-releasing Hormone
A decapeptide synthesized in hypothalamic neurones; it is released in a pulsatile fashion into the hypothalamic–pituitary portal blood system, and stimulates synthesis and release of gonadotropins by the pituitary gland.

Gonadotropins
Glycoprotein hormones, called *follicle-stimulating hormone*, *luteinizing hormone*, and *chorionic gonadotropin*, consisting of a common α subunit linked to a hormone-specific β subunit, which stimulate ovarian and placental cells.

GDF
Growth and differentiation factor, originally discovered in a search for extracellular factors essential for mammalian embryonic development. There are now numerous family members, some of which are also called BMPs due to their sequence homologies.

IGF
Originally called somatomedins, these are insulin-like polypeptides made by the liver and released into the blood when stimulated by somatotropin. Now known as insulin-like growth factors, of which there are IGF-I and IGF-II, they cause sulfate incorporation into collagen, RNA, and DNA synthesis, and are important for cell growth and differentiation.

IGFBP
A family of six of soluble proteins (termed IGFBP-1 through to -6) that bind insulin-like growth factors and modulate their biological actions at the cellular level.

Leukemia Inhibitory Factor
A pluripotent cytokine that inhibits differentiation of embryonic stem cells, is expressed in endometrial epithelial cells, and has been shown to be essential for implantation in mice.

Local Regulation
Autocrine or paracrine regulation of cells by factors made locally within a tissue.

Oxytocin
A nonapeptide with uterotonic, prostaglandin-releasing, and milk-letdown activities, made by hypothalamic neurones and ovarian cells.

Receptors
Membrane or intracellular polypeptides that bind hormones or other factors with a high degree of specificity and transduce the signal to the cell either as transcription factors regulating genes or via a second-messenger system.

Second-messenger Systems
Small cytoplasmic molecules, which, when activated, amplify the signal they receive (e.g. cyclic AMP and inositol phosphate).

Signal Transduction
The process by which a signal received at the cell surface is passed from one carrier to another to achieve a cellular response.

Steroidogenic Enzymes
A family of enzymes, many of which belong to the P450 gene family, that catalyze the synthesis and metabolism of steroid hormones such as estrogens, progestogens, and androgens.

Tissue Remodeling Enzymes
Enzymes belonging particularly to the matrix metalloproteinase family, such as MMPs 1, 2, 3, 7, and 9, and their associated tissue inhibitors, the TIMPS, which are responsible for remodeling the constituents of basement membranes and the extracellular matrix during growth, angiogenesis, tissue breakdown, and wound healing.

Transforming Growth Factor β-Superfamily
A family of pleiotropic factors that have high homology at the level of the gene and in their tertiary protein structure. Examples include TGF-β, inhibins, activins, and Müllerian-inhibiting substance.

Transmembrane Helices
G-protein-linked receptors are a family of receptors characterized by a structural motif of seven hydrophobic regions that span the cell membrane, forming a helical structure within the membrane.

Zona Pellucida Proteins
A family of highly glycosylated proteins unique to the surface coat of the oocyte, several of which serve as primary and secondary sperm binding proteins on the egg at the time of fertilization.

The female reproductive system consists of the hypothalamus, pituitary, ovaries, breasts, and uterus. These organs make up a finely tuned system that controls puberty, the menstrual cycle, pregnancy and birth, lactation, and menopause. With the tools available through molecular biological approaches, we are able to come to a greater understanding of the complex interactions that operate in this system. The hormones, growth factors, and enzymes that are able to exert unique and specific effects in this system do so either locally or via endocrine mechanisms. The elucidation of the molecules involved in this regulation and their mechanisms of action leads to a better understanding of the physiology and pathophysiology of reproduction, and provides the potential means for developing better methods to control fertility and infertility.

Outlined below are some of the major hormonal regulatory systems that are important in the female reproductive processes.

1 Principles: Regulation of Reproductive Processes in the Female

1.1 The Hypothalamic-pituitary-ovarian-uterine Axis

The regulation and coordination of all facets of the reproductive process in the female are controlled by the hypothalamic–pituitary–ovarian–uterine axis (Fig. 1). The activities of numerous hormones, growth factors, and cytokines, and their receptors and binding proteins, as well as enzymes, structural proteins, and other factors within this axis control the development of the reproductive system in fetal life and during puberty, as well as the cyclic production of hormones and eggs during reproductive life, the establishment and maintenance of pregnancy, the birth process, lactation, menopause, and the secondary sex characteristics.

An important process in female reproduction is ovarian folliculogenesis. This is a complex and dynamic process in which the follicles undergo a coordinated progression of recruitment, development or atresia, selection, and finally ovulation (Fig. 2). Peripheral endocrine hormones and local paracrine and autocrine factors contribute to this process.

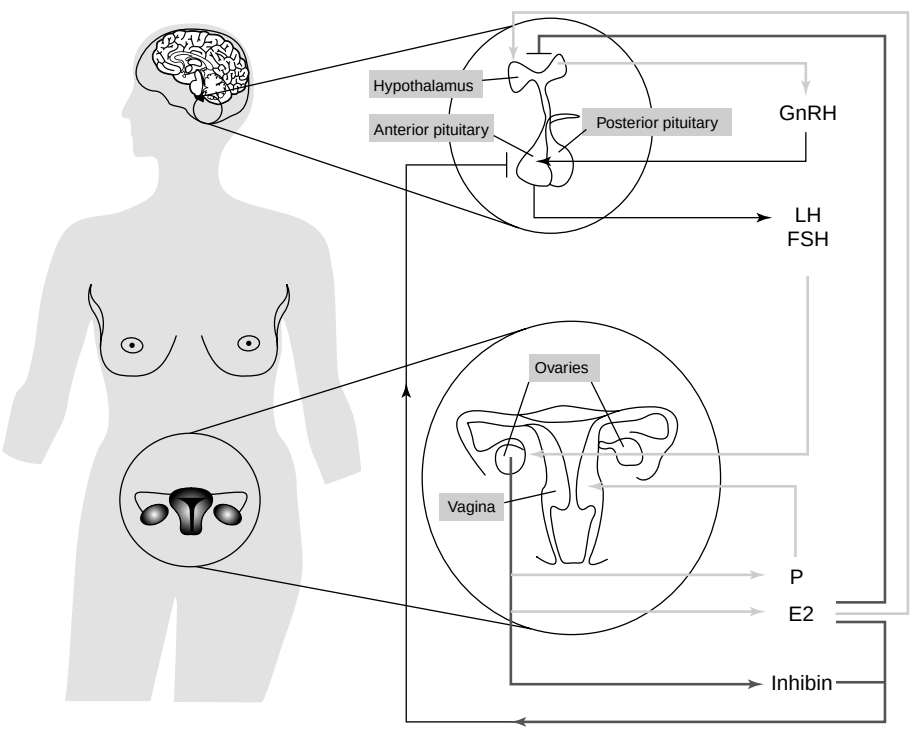

Fig. 1 Hormonal feedback control of the hypothalamic–pituitary–ovarian–uterine axis. Positive feedback is indicated by the blue lines, negative feedback by the red lines, and arrows indicate the direction of the controlling pathways. Estrogen has both positive and negative control on the hypothalamus, depending on the stages of follicular development. (See color plate p. xxi.)

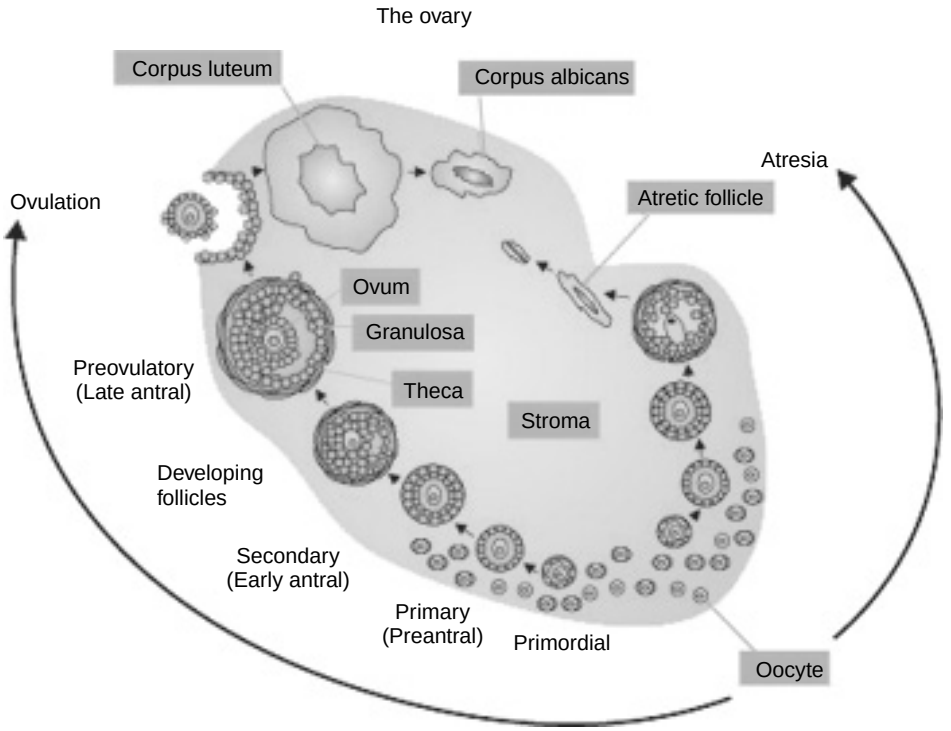

Fig. 2 Schematic representation of folliculogenesis. Follicles can be quiescent (i.e. primordial), committed to growth (preantral and antral), ovulatory, or atretic. Follicles are believed to leave the primordial pool in an ordered sequence, but at an unknown rate and become irreversibly committed to growth. The mechanisms responsible for the reinitiation of growth of primordial follicles remain unknown.

1.2 Levels of Regulation

The hypothalamic–pituitary–gonadal axis was characterized originally by the positive and negative feedback relationships of the endocrine hormones that had been shown to regulate the reproductive system (Fig. 1). For example, gonadotropin-releasing hormone (GnRH) released by the hypothalamus into the hypophyseal portal circulation stimulates both the synthesis and secretion of the luteinizing hormone (LH) and the follicle-stimulating hormone (FSH) by the gonadotropes in the anterior pituitary. FSH then stimulates production of estradiol-17β (E2) by the ovarian follicle; E2 in turn has a negative feedback effect to limit the production of FSH, and so an equilibrium is established that ultimately determines the extent of stimulation of egg and hormone production by the ovary.

In addition to the endocrine regulation of this axis, there is regulation at the organ and cellular level by locally produced growth factors and cytokines. Some of the endocrine hormones also act locally. For example, E2 is produced by the ovary and acts directly on the follicles; prostaglandins produced by the endometrium act within the uterus. The growth factors, cytokines, enzymes, and other factors identified more

recently within the reproductive axis are structurally identical to those produced in other organ systems. Nevertheless, they can be differentially regulated and can perform functions specific to the reproductive axis. Examples are the cyclic tissue remodeling by matrix metalloproteinases (MMPs) and angiogenic factors in the ovary and uterus, the role of activin in the functions of the pituitary, ovary, and placenta, and the actions of the insulin-like growth factors (IGFs) and their binding proteins in the ovary and endometrium.

1.3 Specificities of Actions

It has already been said that many of the hormones and other regulatory factors are not unique to the female reproductive system even though they perform specific functions. Regulation of hormones and cytokines occurs at their sites of production and at their sites of action by cell-specific expression of their receptors; the presence of binding proteins and metabolizing enzymes determines the extent and specificity of their actions.

The few substances that are unique to the reproductive system can generally be found in both the female and the male; they include GnRH and its receptor, FSH and its receptor, LH and its receptor, Müllerian-inhibiting substance (MIS) and its receptor, oxytocin, and relaxin. Substances unique to the female are the placental proteins such as the chorionic gonadotropins and lactogens (and a special class of interferons, ϕ, in ruminants), and the zona pellucida proteins (ZP1, 2 and 3), which coat the oocyte.

It is not possible, in the space available, to give a full account of the expression and actions of all the factors involved in regulating the female reproductive system. Therefore, we have briefly summarized selected examples of a number of hormonal regulatory systems to provide some insights into the subject. The reader is referred to the bibliography and to other references therein.

2 Examples of Regulatory Systems

2.1 The Steroidogenic System

2.1.1 Estrogen and its Receptors

Estrogens, androgens, and progestogens are the sex steroids regulating the female reproductive axis. They originate primarily from the ovary and the placenta, with important contributions by the uterus, adrenal gland, and other peripheral organs such as fat cells and skin. During the ovarian cycle and pregnancy, steroid secretion by the ovary and placenta occurs in a highly regulated and episodic fashion that is determined in part by the pituitary and placental gonadotropins. Changes in the levels and activities of the steroidogenic enzymes are major mechanisms determining these changes.

The synthesis of the steroids involves successive elimination of carbon atoms from the precursor, cholesterol, catalyzed by members of the cytochrome P450 gene family, namely, cholesterol side-chain cleavage cytochrome P450 (P450scc), 17α-hydroxylase (P450c17), and aromatase (P450arom) (Fig. 3). There are also 3α- and β-, 17α-, and 20α-hydroxysteroid oxidoreductases, which catalyze the interconversion of some of the steroids as well as isomerization reactions. Not only the cellular expression of these enzymes but also their electron donors have been extensively investigated in the female reproductive

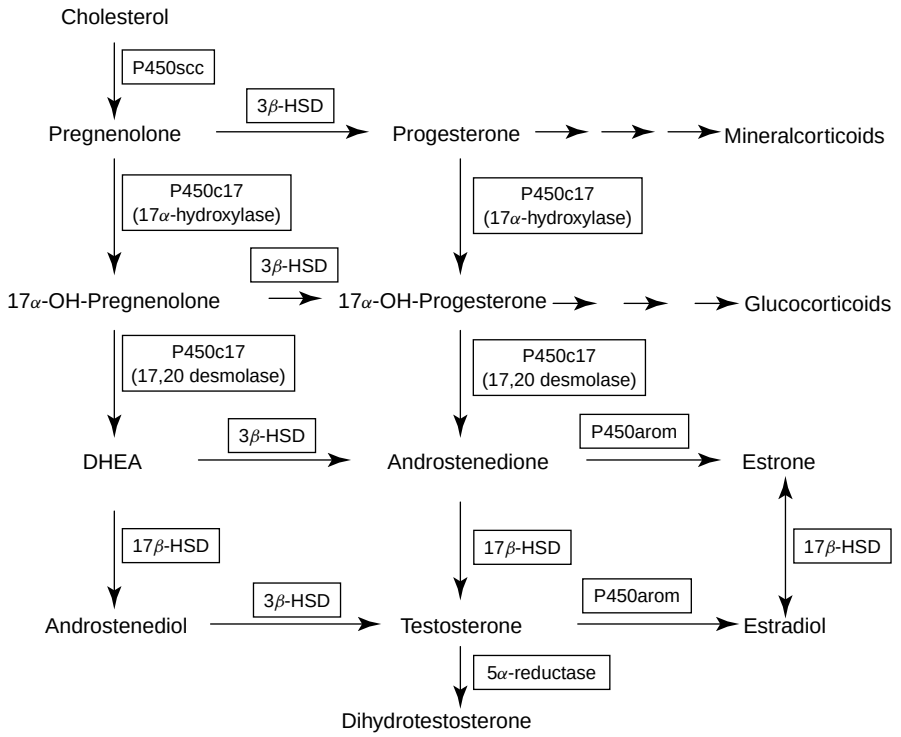

Fig. 3 Pathways involved in steroid hormone biosynthesis from cholesterol. Key enzymes involved are shown in boxes. P450c17 possesses both 17α-hydroxylase and C_{17-20} lyase (also known as 17,20 desmolase) activities. 3β-HSD is 3β-hydroxysteroid dehydrogenase, P450arom is aromatase, and 17β-HSD is also known as 17-ketoreductase.

system; studies have focused on gene expression, protein synthesis, and activity assays. The chromosomal localization of some steroidogenic enzymes is shown in Table 1.

Estrogen receptors (ERs) are ligand-inducible transcription factors. They are members of a superfamily of DNA-binding transcription factors that are divided according to their DNA-binding domains into steroid receptors, thyroid and retinoic acid receptors, and orphan receptors (lacking known ligands). The intracellular steroid receptors are generally associated with heat shock proteins, and in the presence of ligand, they dissociate from the heat shock proteins, become phosphorylated, undergo dimerization, and translocate to the nucleus.

Tab. 1 Chromosomal localization of human genes for estrogen receptors and some steroidogenic enzymes.

Components	*Chromosomal localization*
Erα	6q25.1
Erβ	14q
P450scc	15q23–24
P450c17	10q24.3
P450arom	15q21.1
5α-reductase type I	5p15
5α-reductase type II	2p23

ERs are of major importance in the female reproductive system. There are two types of ERs, called ERα and ERβ, each of which have six functional domains (Fig. 4). The N-terminal A/B domains contain the cell-specific and constitutive transactivation (TAF-1) functions and the isoform-specific sequences. The C domain or DNA-binding domain is the most conserved region, containing invariant cysteine repeats that form two "zinc fingers," which are crucial for the binding of the receptor to the DNA. The constitutive nuclear localization signal is found in the C to D region, which, together with the hormone binding E domain, is the site of interaction with the heat shock proteins. The E domain is also involved in hormone binding, nuclear translocalization, hormone-dependent transactivation, and dimerization. The F domain is found only in ER. ERα and ERβ have similar affinities for estradiol-17β, but exhibit differing affinities for other natural and synthetic estrogens, which may be important for their respective roles. The two types of ERs can form homo- or heterodimers, although the ERβ form may have an inhibitory influence on the ERα in a dimer.

As would be expected from the central role and multiple actions of estrogen in the female reproductive system, ERs have been localized in cells in the hypothalamus, pituitary, ovary, mammary gland, and uterus. ERα was found to be the predominate type in the uterus and mammary gland, whereas ERβ is more abundant in the ovary. The type of cell containing ER has been found to change according to the stage of differentiation or function, and in some cases such cells can be proximal to the cells eliciting the final response. For example, in the endometrium, epithelial–mesenchymal interactions involve an action of E2 on ER on the stromal cells, which then have a paracrine influence on the epithelial cells. Expression of ER is constitutive in some cases and can be up- and downregulated by estrogen itself, and by FSH in the ovary.

The classical genomic pathway for ER signaling involves binding of the ER dimer to estrogen response elements (ERE) in the promoter region of the target gene. Binding induces a conformational change in the ER that allows interactions with coactivators, leading to transcriptional stimulation of the target gene. These genes include

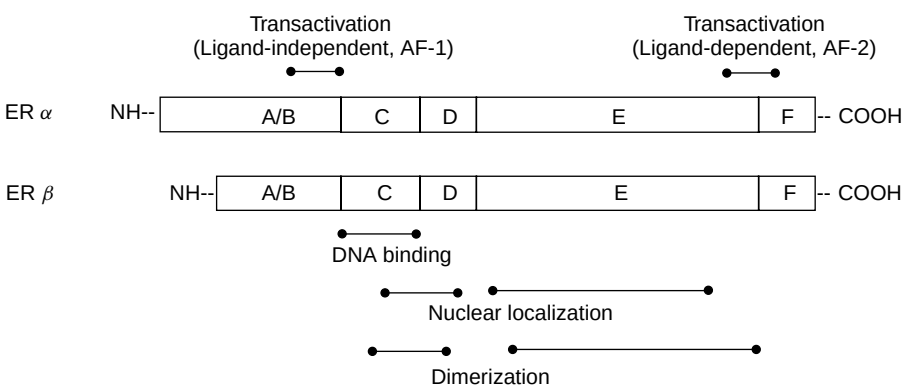

Fig. 4 The structural and functional domains of human ERα and ERβ. Structural domains are indicated as letters, and functional domains responsible for transactivation, DNA-binding, nuclear localization, and dimerization are indicated by horizontal bold lines.

some essential for reproduction such as prolactin and the progesterone receptor. ER can also alter gene transcription by a mechanism that does not involve binding of ER to the ERE. Instead, the ER interacts with other transcription factors such as Sp1, Jun, or nuclear factor kappa B, by a protein–protein interaction. Examples of genes regulated by this nonclassical pathway include collagenase and heat shock protein 27. There is evidence that this nonclassical pathway is of physiological relevance, and that it may subserve the agonist actions of selective estrogen receptor modulators or SERMs. There is also evidence for a nongenomic action of estrogen via membrane-bound receptors. This pathway is thought to mediate the fast actions of estrogen such as those associated with ion transport.

2.1.2 Roles of Estrogen in Ovarian Functions

Estrogens are essential for fertility in both males and females, but they are not necessary for survival after birth or for formation of the reproductive tract. The absence of estrogen or its actions in female rodents eventually leads to the appearance of testicular cells in the ovaries of mice with targeted disruption of the two ER genes and the *P450arom* gene. This supports a role for estrogen in somatic cell differentiation and shows that it is required for maintenance of the female somatic cell phenotype. It also underlines the plasticity of the ovary in terms of its hormonal milieu.

Estrogen has intrafollicular actions that are intimately related to the actions of FSH, the key driver of folliculogenesis. The capacity of follicles to make estrogen is limited by the thecal production of androgen, the substrate for estrogen, and the expression of P450arom in the granulosa cells, which is FSH-dependent. Although primary follicles express ER, they have little or no capacity to produce significant quantities of E2 because thecal androgen production is very low and the P450arom expression is low. FSH, together with insulin-like growth factor-1 (IGF-1) and activin, upregulates P450arom expression in the mural granulosa cells of secondary follicles. This combination of hormones, together with E2, itself stimulates proliferation and differentiation of the granulosa cells. As follicles grow, they produce more E2, the hallmark of a healthy follicle. Those tertiary or Graafian follicles selected for ovulation contain granulosa cells with the capacity to respond to LH and produce more E2.

The necessity of estrogen for normal folliculogenesis has been well established in rodent models, particularly those in which the ER and *P450arom* genes in mice have been targeted for disruption. However, it has been suggested that the ovarian roles for estrogen in polyovular species may not apply to monovular species, including women. Nevertheless, in those rare cases that document mutations in the human *P450arom* gene, folliculogenesis is not normal, and there is evidence that the follicles that are present in the ovaries are unhealthy. Further work is needed to resolve this question.

2.2 Gonadotropin System

2.2.1 Gonadotropins and their Receptors

LH and FSH are members of the glycoprotein hormone family, which also includes thyroid-stimulating hormone (TSH) and, in humans and horses, chorionic gonadotropin (CG). Each of these hormones consists of two different subunits, α and β, which are glycosylated, contain multiple disulfide linkages, and are noncovalently

linked to each other to form the active molecule. The α-subunit apoproteins are the same for all the glycoprotein hormones within a species and are encoded by a single gene comprising 4 exons, which, in the human is located on chromosome 6q12.21. The α-subunit protein contains 10 cysteines, which are involved in intrasubunit disulfide bridges and 2 N-linked glycosylation sites. The β-subunits are different for each hormone and determine the biological specificity of each hormone. The LHβ and FSHβ genes show similar genomic arrangements, suggesting a common ancestral origin. The human LHβ gene cluster is located on human chromosome 19q13.32; this cluster includes at least one LHβ gene and 6 hCGβ genes and pseudogenes. The FSHβ gene is located on human chromosome 11p13. The mature β-subunit protein contains 12 cysteine pairs that form 6 intrasubunit bridges, 2 N-linked glycosylation sites (one in LHβ) and a variable length from 111 to 145 amino acids. Addition of the carbohydrate moieties to the apoproteins is important for correct synthesis of bioactive gonadotropin dimers, and for achieving appropriate biological activity of the molecule. Regulatory mechanisms may operate through modifying the sugars on these hormones. The hCGβ protein has a C-terminal extension of 29 amino acids with 4 additional O-linked glycosylation sites that increases the half-life and biopotency of hCG over LH. It is of interest that the crystal structure of both gonadotropin subunits reveals the cystine knot structure, also seen in members of the transforming growth factor-β (TGF-β) family, nerve growth factor, and platelet derived–growth factor.

The anterior pituitary gland in a mature animal has several different types of endocrine cells, each of which secretes particular hormones. LH and FSH are synthesized in the gonadotrope cells of the anterior pituitary, with populations of cells synthesizing either LH or FSH, or both. This ability of some gonadotropes to produce more than one hormone makes them unusual in relation to the other endocrine cell types in the pituitary such as the corticotropes, thyrotropes, lactotropes, and somatotropes, each of which synthesizes and secretes largely a single tropic hormone in the adult. During development, however, it is likely that the different pituitary cell types arise from a common cell type under the control of cell-specific transcription factors that determine the temporal expression patterns of the hormonal genes in the pituitary.

LH and FSH are secreted in different patterns under different physiological circumstances, and the expression of the genes that encode them is also regulated differently under different conditions. GnRH provides the major stimulatory drive for expression of the β-subunit genes, and the steroid hormones, in general, have negative regulatory effects. The peptides, inhibin and follistatin, negatively regulate the expression of FSHβ, and activin stimulates expression of the *FSHβ* gene.

The LH/CG (LHR) and FSH receptors (FSHR) are members of the seven membrane–spanning domain, G-protein-linked receptor family. LH and hCG bind to the same receptor. Many of the receptors in this family have very short, extracellular amino termini, but the LHR and FSHR are unusual in that they have very long amino termini for hormone binding. The genes for LHR and FSHR are located on human chromosome 2p21 and 2p21-16, respectively. The LHR and FSHR genes are encoded by 11 and 10 exons respectively, with the last exon encoding the seven membrane–spanning domains and

the carboxy-terminal intracellular region. Exon 1 encodes the signal peptide to direct the newly synthesized protein to the luminal side of the endoplasmic reticulum and eventually to the cell membrane. Exons 2–9 of the FSHR and 2–10 of the LHR confer the hormone specificity and binding. These extracellular domains have a number of copies of an imperfectly repeated sequence of approximately 25 amino acids known as leucine-rich repeats. The functional significance of this structure is not known, but it may be involved in cell–cell or protein–protein interactions. There are also a number of potential N-linked glycosylation sites in the extracellular domains of these receptors but their role has not been elucidated. It has been suggested but not proven that to transmit its signal, the hormone bound to the extracellular domain interacts with the extracellular loops of the transmembrane segment. It is clear, however, that the third intracellular loop and the cytoplasmic tail are involved in G-protein coupling.

Relatively little is known about the transcriptional control of the gonadotropin receptor genes. For example, the FSHR gene of the rat and human has no TATA or CCAAT elements, no GC box motifs, and no CRE, AP-2, or methylation sites in its 5′-flanking region. A consensus E box sequence has been identified that binds upstream regulatory factors (USFs), members of a helix–loop–helix family that controls growth and differentiation. There is evidence that a site in the 3′-region for the E2F transcription factor may be important for promoter activity.

Both receptors are coupled to the G_s-protein that activates adenylyl cyclase leading to an increase in intracellular cAMP. However, there is accumulating evidence that both receptors can interact with G-proteins other than G_s such as $G_{q/11}$, G_{13}, and G_i. Both receptors can also activate other signal transduction pathways that involve phophoinositide turnover, intracellular calcium levels, and mitogen-activated kinases (MAPK).

After hormonal stimulation, the LHR and FSHR desensitize such that the efficiency of hormonal signaling is decreased. This occurs because of uncoupling of the receptor from the G_s-protein, internalization of the receptor resulting in a decrease in receptor number on the cell surface, and a decrease in receptor synthesis. Part of the mechanism of uncoupling involves binding of the receptor to inhibitory proteins called arrestins and targeting to clathrin-coated pits, but the details of this process are still under investigation.

2.2.2 Roles of Gonadotropins in Ovarian Functions

Both gonadotropin receptors are expressed in the ovary, and both can be detected as multiple transcripts that may arise as a result of the use of multiple transcription start sites or alternate polyadenylation signals. Both receptors have been shown to exhibit alternate splicing of the first 9 or 10 exons, and some of these transcripts would give rise to truncated peptides. It is not clear if these truncated forms of the receptor are translated, or if they have any functional significance, particularly during different phases of folliculogenesis.

FSHR are confined to the cell membranes of granulosa cells in the ovary, whereas LHR are found on the theca interna cells, the granulosa cells, and the luteal cells. In the case of granulosa cells, expression of the LHR occurs only at the late stages of follicular development during the antral/tertiary period, and is the result of stimulation of the cells by FSH, E2, and IGF-1. This cellular compartmentalization of gonadotropin receptors underlines the

"two cell, two gonadotropin" model of estrogen synthesis by the ovarian follicle. The follicle consists of an inner avascular granulosa cell layer within a basement membrane, surrounded by an outer vascular thecal cell layer. LH stimulates production of the precursor androgen by stimulating P450c17 in the theca interna. The granulosa cells, which lack the P450c17, do have the P450arom, which, under the influence of FSH, converts the thecal androgen to estrogen. Thus, both cell types and both gonadotropins are crucial for estrogen production during early folliculogenesis. Later in folliculogenesis, at the antral or tertiary stage, the granulosa cells acquire the capacity to respond to LH, which also regulates P450arom expression.

Both gonadotropins are essential for complete folliculogenesis, but in different ways and at different times. Neither gonadotropin is essential for the initiation of growth of resting primordial follicles into the follicular growth trajectory. FSH is not essential but may be facilitatory in the growth of primary and secondary preantral follicles. FSH is essential for the formation of healthy antral follicles, their selection as preovulatory follicles, and their survival from atresia. The mechanism by which FSH influences folliculogenesis includes stimulation of estrogen production, proliferation of granulosa cells by action on cyclin D2, and interacting with estrogen and IGF-1 to facilitate differentiation (see below). LH is necessary to stimulate the theca cells to produce androgen substrate for estrogen production, subsequently to drive the estrogen production in granulosa cells of selected antral follicles, and to cause ovulation and luteinization.

The mechanisms by which E2 and IGF-1 interact with FSH to enhance folliculogenesis are not clear, but probably involve an influence of FSH on the IGF-1 signal transduction pathway. The classic action of FSH is via the cAMP/protein kinase-A pathway to stimulate gene transcription. More recently, it has been shown that FSH can influence the MAPK pathway used by IGF-1 in granulosa cells by an action of cAMP through cAMP exchange proteins or GEFs on PI-3-kinase. Members of the forkhead (FOX) family of transcription factors are downstream targets of the IGF-1 pathway, and both FSH and E2 enhance expression of the gene for forkhead homolog of rhabdomyosarcoma (FKHR) and its protein in granulosa cells of developing follicles. Furthermore, E2 enhances expression of components of the IGF-1/PI3K/PDK1 pathway and expression of ERβ, indicating that E2 and IGF-1 may form an autocrine regulatory network in conjunction with FSH in developing follicles. FKHR has been linked to proliferation of granulosa cells, but the targets of FKHR are not known. The IGF-1/PDK1 pathway has been linked with cell survival. This may be the mechanism that FSH enhanced by E2, influences granulosa cell survival and thereby prevents atresia, whereas the action of FSH via the cAMP/PKA pathway facilitates differentiation of granulosa cells.

2.3 The Transforming Growth Factor β (TGF-β) Superfamily System

2.3.1 TGF-β Superfamily

TGF-β was described originally as a factor that induced normal rat kidney fibroblasts to form colonies in soft agar in the presence of epidermal growth factor (EGF). Subsequently it was found to belong to a family of multifunctional growth factors with diverse biological activities including cell growth, cell death

or apoptosis, cell differentiation, and extracellular matrix (ECM) synthesis. They play important actions in embryonic development, tissue remodeling, and wound healing, and the peptide and its receptors can be found in almost all cell types.

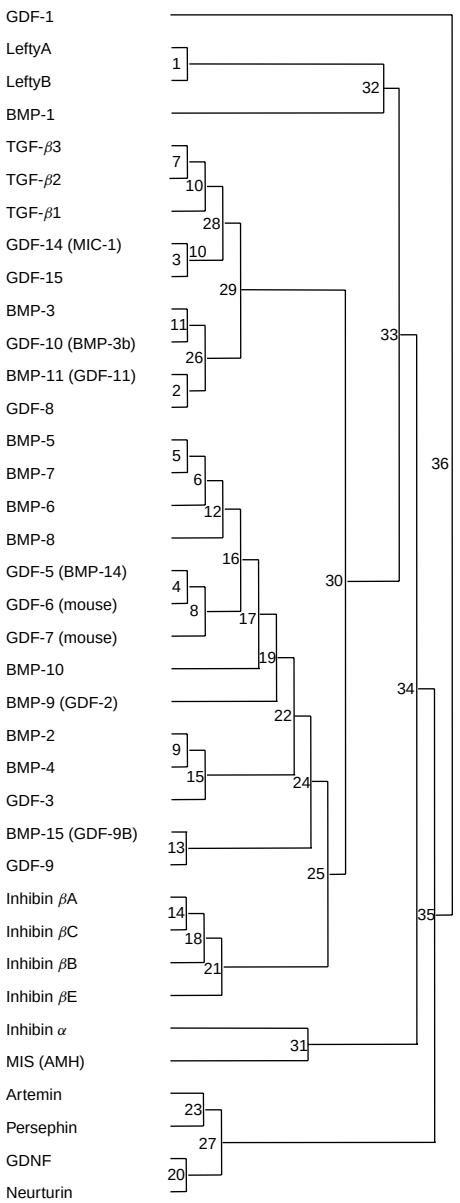

Five forms of TGF-β have been cloned, though only TGF-β_1, TGF-β_2, and TGF-β_3 have been found in the mammalian ovary. Each is a disulfide-linked homodimer with a molecular weight of 25 kDa in the unreduced form. Each monomer is 112 to 114 amino acids long and occupies the carboxy-terminal sequence of its precursor. There is a 70 to 80% sequence homology between the monomers, and there is complete conservation of the nine cysteine residues in each monomer. TGF-β generally is secreted in a biologically inactive, latent form, noncovalently associated with a dimer of the pro region of the precursor. The TGF-β superfamily of peptides is now used to describe a large group of extracellular, apparently disparate growth factors by virtue of their 30 to 40% amino acid homologies to TGF-βs (Fig. 5). They include the inhibins and activins, Müllerian-inhibiting substance, bone morphogenetic proteins, growth differentiation factors (GDFs), the decapentaplegic complex of *Drosophila*, *nodal* and *dorsalin-1* in mice, and *ceg-1* in *Caenorhabditis elegans* and *Drosophila*. Members of this superfamily share a number of similarities, including their synthesis as prepropeptide precursors with an N-terminal signal peptide, a middle prodomain followed by a mature peptide domain, as well as the formation of homo- or heterodimers

Fig. 5 TGF-β superfamily members. Mature peptide amino acid sequences of various human TGF-β superfamily members were aligned using GeneWorks program (IntelliGenetics, Inc.). Mouse GDF-6 and GDF-7 were also included in the analysis, as human GDF-6/7 was similar to GDF-5, also called cartilage-derived morphogenetic protein-1. Alternative names for some GDFs or BMPs were shown in parenthesis. Numbers on the hierarchical tree indicate the decreasing order of protein homologies.

through intermolecular disulfide bonds between conserved cysteine residues. Various members of each subfamily have been identified as having a regulatory role in the female reproductive system (Table 2).

In the ovary, TGF-β is expressed in follicular cells and has both proliferative and cytodifferentiation actions on granulosa cells. TGF-β can stimulate production of inhibin A and B by dispersed ovarian cells of postnatal rats, and like activin, TGF-β actions may be age-dependent.

The inhibin/activin family Inhibin was identified originally as a water-soluble testicular extract that prevented the hypertrophy of pituitary cells in response to castration. This definition was subsequently refined to designate an activity capable of selective suppression of FSH secretion, and on this basis, inhibin was isolated from ovarian follicular fluid and identified as a glycosylated heterodimer consisting of an α- and a β-chain joined by disulfide bonds. Like TGF-β, the mature dimer consists of the C-terminal sequences of a larger precursor of each chain, having a molecular weight of 31 to 32 kDa, although larger (e.g. 51, 58, 64, 105 kDa) biologically active forms exist. The α-chain precursor of 366 amino acids in the human is encoded by a single gene with only one intron, and the homology of the mature peptide (133 amino acids) is between 85 and 96% for human, bovine, ovine, and porcine specimens. The seven cysteine residues are totally conserved between species, but the human α-chain contains two potential N-linked glycosylation sites compared to one in the animal species. There are also other molecules such as αN, derived from the N-terminal end of the precursor of the α-chain, which do not have properly defined biological roles. There are five distinct forms of the β-chain (β_A, β_B, β_C, and β_E in mammals, and β_D in *Xenopus* sp.), and the β-chains in human are each coded by a single gene containing one intron. Only the β_A- and β_B-chains have been shown to dimerize with the α-chain, thereby giving rise to two isoforms of inhibin – inhibin A and inhibin B respectively. The human precursor peptides for β_A and β_B are 426 and 407 amino acids, respectively. The homologies of the mature β-chains (116 and 115 amino acids respectively) exceed 95%, and the nine cysteines are conserved. Indeed, the distribution of the cysteine residues in the α-, β_A-, and β_B-chains are very similar among the three peptides and other family members. Both inhibin A($\alpha\beta_A$) and inhibin B ($\alpha\beta_B$) have FSH-suppressing activity.

At the time when inhibin was being isolated from ovarian follicular fluid, an activity was found that stimulated FSH secretion. This activity, called activin, was identified as a nonglycosylated dimer of the β-chains with a molecular weight of 25 kDa. Activin exists as three isoforms,

Tab. 2 Members of the TGF-β superfamily that are involved in folliculogenesis.

TGF-β family:	TGF-β1
	TGF-β2
	TGF-β3
Inhibin/Activin family:	Inhibin A
	Inhibin B
	Activin A
	Activin B
	Activin AB
BMPs/GDFs family:	BMP-4
	BMP-5
	BMP-6
	BMP-7
	BMP-8b
	GDF-9
	GDF-9B (BMP-15)
Others:	MIS

activin A ($\beta_A\beta_A$), activin AB ($\beta_A\beta_B$), and activin B ($\beta_B\beta_B$), all of which appear to have similar FSH-stimulating activity. It is now known that activin is a pleiotropic regulator in many tissues including the developing embryo, where it acts as an inducer of the mesoderm, and in the bone marrow, where it is an erythroid-differentiating factor. Transgenic mice deficient for the inhibin β_B-subunit produced offspring, but the animals had developmental and reproductive disorders. There was a failure of eyelid fusion, and whereas the male offspring bred normally, the offspring of the females had high perinatal mortality. Furthermore, expression of the β_A-subunit was upregulated in the ovaries of the females. These data suggest that activin β_B plays a role in fetal development and is critical for female fecundity, but is not essential for mesoderm formation. In the ovary, activin is produced by the granulosa cells, and the pattern of expression of the activin subunits mRNA changes during folliculogenesis. Numerous functional studies support a role for activin in folliculogenesis. Activin can promote granulosa cell proliferation, potentiate FSH actions by increasing the FSH receptor expression on granulosa cells, and modulate steroidogenesis in both granulosa and theca cells. The actual function of activin in folliculogenesis is sometimes obscured by its age-specific effects. For example, it was shown that activin A stimulates growth of follicles from immature animals but blocks the FSH-induced growth of preantral follicles from adult animals. Nevertheless, a recent study suggested that activin from secondary follicles may cause small preantral follicles to remain dormant.

Follistatin or FSH-suppressing activity was also isolated from ovarian follicular fluid but was found to be unrelated structurally to inhibin. It exists in up to six isoforms (31–42 kDa), which arise as a result of a combination of alternate splicing of its mRNA and variable glycosylation of the mature peptide. It is encoded by a single gene, consisting of six exons, with the alternate splice site located in exon 5. All the follistatin genes and cDNAs from different mammalian species sequenced to date are more than 95% homologous at the amino acid level, with a homolog in *Xenopus* having 85% homology. The mature follistatin proteins consist of four contiguous domains, each encoded by a different exon; the last three domains are 52% homologous with each other and are also structurally related to EGF and to a pancreatic secretory inhibitor. Follistatin derives its FSH-suppressing activity from its ability to bind activin; it is expressed in the germ cells, embryo, pituitary, ovary, and placenta, as well as the brain, adrenal, bone marrow, and spleen, although not always coincidently with activin. It is also highly expressed in the kidney and pancreas, where the inhibin β-subunit mRNA has not been detected. Neither follistatin nor inhibin β-subunit is expressed in the liver. In the ovary, follistatin is expressed by the granulosa cells, and its basal expression increases with increased granulosa cell differentiation. Follistatin production can be upregulated by FSH and activin, but only in partially or totally undifferentiated granulosa cells, and not in fully differentiated cells. Since follistatin counteracts activin activities, its pattern of expression is consistent with the suggested hypothesis that there is a decreasing activin "tone" within follicles as they develop owing to higher levels of follistatin. This could have implications for activin access to its receptor signal transduction system as follicles mature, despite the presence of all the signaling system components. An analysis

of the activin and follistatin contents of individual follicles at defined stages of development is needed to support this hypothesis.

The BMP/GDF family BMPs were originally isolated by their ability to induce ectopic bone and cartilage formation *in vivo*, but it is now evident that BMPs are also multifunctional regulators of morphogenesis during development in vertebrates and invertebrates. GDF-1 was cloned initially from an 8.5-day mouse embryo in a search for extracellular factors essential for mammalian embryonic development. Subsequently, numerous family members have been discovered and have been classified into various subgroups on the basis of their sequence homologies. The BMPs/GDFs family now constitutes the single largest family within the TGF-β superfamily, and also includes BMP-like molecules from *Drosophila* sp. and *C. elegans and Xenopus* sp.

Recent studies on naturally occurring BMP/GDF mutations in sheep, as well as the creation of knockout mouse models have indicated the important role these factors play both in folliculogenesis and in embryonic development. BMP-4 and BMP-8b produced by the extraembryonic ectoderm appear to be essential for the generation of primordial germ cells in the mouse. BMP-4 mRNA, together with BMP-7 mRNA, is also expressed in the adult ovary, especially in thecal cells of healthy Graafian follicles. Both BMP-4 and BMP-7 increase the FSH-induced estradiol production by granulosa cells, whereas the FSH-induced progesterone production is decreased, leading to the possibility that these factors may function as luteinization inhibitors in the Graafian follicle. Genetic deletion in the mouse, of BMP-2, BMP-4, or BMP-7 caused embryonic (BMP-2, BMP-4) or perinatal (BMP-7) lethality.

GDF-9 is produced by the oocytes and targeted deletion of the *GDF-9* gene resulted in female infertility due to the failure of the theca cells to organize abnormal oocyte growth and decreased granulosa cell proliferation, which resulted in a block of folliculogenesis at the primary follicle stage. *In vitro* studies have shown that oocyte-derived GDF-9 stimulates the proliferation and suppresses the FSH-induced differentiation of rat granulosa cells *in vitro*.

BMP-15 (also called GDF-9b) is an oocyte-derived factor, which is expressed from the primary follicle phases onward. A natural X-linked mutation in sheep identified BMP-15 as an essential factor for female fertility. Intriguingly, BMP-15 has a dose-dependent effect since homozygous mutants have their follicular development arrested at the primary stage, whereas heterozygous animals have increased ovulation rates resulting in multiple pregnancies. Although the exact function of BMP-15 is still unknown, it has been shown to modulate FSH action in rat granulosa cells by suppressing FSH receptor expression.

BMP-6 is another TGF-β member expressed in the oocyte. Until recently, no information was available on its putative biological function in the ovary. *In vitro* studies using rat granulosa cells have demonstrated that BMP-6 has no effect on their proliferation, but inhibits FSH-induced progesterone production without affecting FSH-induced estradiol synthesis, which is reminiscent of a BMP-15 activity. However, in contrast to BMP-15, BMP-6 does not modulate the expression of the FSH receptor leading to the suggestion that, in addition to the theca-derived BMP-4 and BMP-7, the oocyte-derived

BMP-6 and BMP-15 could also function as inhibitors of luteinization. BMP-6 knockout mice, however, remain viable and are fertile.

MIS MIS, also called *anti-Müllerian* hormone (AMH) or Müllerian-inhibiting factor (MIF), is the testicular factor responsible for regression of the Müllerian primordia, the first step of male somatic differentiation. Native MIS is a high molecular weight (140–144 or 280–290 kDa) glycoprotein, formed by the assembly of dimers or tetramers of a 70 or a 72 kDa monomer. Like other members of the TGF-β superfamily, the MIS monomer is derived from a larger 546 amino acid precursor in which the mature form is at the C-terminal end of the molecule. The 140-kDa homodimer is further processed to a noncovalent complex of 110 and 25 kDa dimers, although it is not known whether this step is required for activation of the biological activity of MIS. The MIS gene, located on human chromosome 19, p13.3, is only 2.7 kb and is encoded by 5 exons. There are multiple transcription initiation sites, perhaps because the human MIS promoter lacks consensus TATA and CAAT boxes, similar to many so-called housekeeping genes.

In males, MIS is produced by the testicular Sertoli cells from the time of fetal sex differentiation to puberty, whereas in females, it is produced by the granulosa cells from the antral follicles of the ovary from the time of birth to the end of ovarian activity. Synthesis of MIS by granulosa cells is low compared with the Sertoli cell of the testis. In rodents, expression of MIS is highest in the granulosa cells of the follicles during the preantral and small antral stage, with atretic follicles having decreased or with no expression of MIS. It is completely absent in the corpus lutea, and is never found in the primordial follicles, theca cells, oocytes, and the interstitium. The physiological significance of expression of MIS in the female gonads, especially in the adult, is unclear, although the specific patterns of MIS expression suggest that MIS might play a role in follicle development and function. There is data to suggest that MIS might affect development of the fetal ovary. For example, transgenic female mice chronically overexpressing MIS undergo total Müllerian regression, and most lack ovaries; where there is ovarian tissue, it is severely depleted of germ cells and contains structures resembling seminiferous tubules. MIS has also been shown to inhibit aromatase activity in fetal ovaries. Further insights on the role of MIS in the ovary came from MIS-knockout mice. Female MIS-deficient animals are fertile and although they showed no obvious gross ovarian abnormalities, the ovaries of adult females showed premature loss of primordial follicles compared to that of wild-type animals. Further *in vitro* studies indicated that MIS acts as an inhibitory growth factor in early folliculogenesis to block the initiation of primordial follicle growth, probably through the paracrine influence of granulosa cells on the primordial follicles.

2.3.2 TGF-β Superfamily Receptors

Almost all of the TGF-β superfamily of ligands uses a conserved signaling pathway involving a family of structurally similar group of receptors, the TGF-β superfamily receptors. These cell surface, single-pass transmembrane receptors contain intracellular kinase domains that phosphorylate serine and threonine residues. There are two distinct receptor subtypes (type I and type II), which need to interact with one

another to produce a cellular response. Ligand binding induces type I and type II receptors in the cellular membrane to associate, and causes the type II to phosphorylate the type I receptor, thereby activating its kinase domain. The activated type I receptor then phosphorylates the SMAD family of intracellular mediators, of which there are 8 members (Smad 1–8) divided into 3 distinct classes based on their functional properties. The receptor-regulated SMADs (or R-Smads) comprising of Smad-1, -2, -3, -5 and -8 are anchored to the cell membrane through association with membrane-bound proteins, including SARA (SMAD anchor for receptor activation). R-Smads, when directly phosphorylated on two conserved serine residues at the C-terminus by the activated type I receptor kinase, then dissociate from the receptor complexes and SARA to form oligomeric complexes with a second class of SMADs called copartner SMADs (Co-Smads) of which Smad-4 is the only member. These heteromeric complexes translocate and accumulate in the nucleus where they associate with one of many DNA-binding proteins including various transcriptional coactivators and corepressors to regulated target gene transcriptions. A third class of SMADs, the inhibitory SMADs (I-Smads) of which Smad-6 and -7 are the two members, inhibit the signals from the serine/threonine kinase receptors to antagonize signal transduction. I-Smads constitute an autoinhibitory signaling pathway as their expression is induced by TGF-β superfamily members themselves.

There are 5 type II receptors in vertebrates, each of which can pair up with one of 7 type I receptors (also known as activin-like receptor kinases, or ALKs) to mediate signaling (Table 3). Recent studies have indicated that the active signaling receptor complex is made up of two type I and two type II receptors, forming a heterotetrameric complex binding two molecules of the ligand. While ligand-binding specificity is determined largely by the type II receptors (and sometimes in combination with the type I receptors), intracellular signaling to SMADs is entirely determined by the type I receptor. There are two

Tab. 3 Signaling components of the TGF-β superfamily members that are important in the female reproductive system. Receptor subtypes and R-SMADs utilized by each ligand are shown. Alternate names for some ALKs are shown in parenthesis.

Ligands	*Type II receptor*	*Type I receptor*	*R-SMADs*
TGF-β	*T*β RII	ALK-1 ALK-5 (*T*β RI)	Smad-2, Smad-3
Activin	*Act* RIIA *Act* RIIB	ALK-4 (*Act* RIB)	Smad-2, Smad-3
Nodal	*Act* RIIB	ALK-4 (*Act* RIB) ALK-7	Smad-2, Smad-3
BMPs	*BMP* RII *Act* RIIA *Act* RIIB	ALK-2 (*Act* RIA) ALK-3 (*BMP* RIA) ALK-6 (*BMP* RIB)	Smad-1, Smad-5, Smad-8
MIS	*MIS* RII	ALK-2 (*Act* RIA) ALK-3 (*BMP* RIA) ALK-6 (*BMP* RIB)	Smad-1, Smad-5, Smad-8

intracellular signaling pathways leading to SMADs activation; the ALK-4, -5 and -7 type I receptors phosphorylate Smad-2 and -3 which transduce TGF-β-like signals (like TGF-β, activin and nodal), and the ALK-1, -2, -3, -6 type I receptors phosphorylate Smad-1, -5 and -8 to mediate BMP-like signals (BMP-2, -4, -6, -7, and MIS).

The role of TGF-β receptors in the ovary has not been extensively studied. In the human ovary, receptors TβRI and TβRII were located in the granulosa, theca, and interstitial cells, with patterns of expression significantly different from each other.

Similarly, information on BMP signaling in the ovary is only now beginning to emerge. BMP receptor mRNAs are present in the ovary, with the strongest expression in granulosa cells and oocytes, and this is consistent with the observed BMP actions on granulosa cells. The importance of BMP signaling in the ovary is demonstrated by natural mutations and loss-of-function studies of the BMP receptors. BMPRIB (ALK-6) deficient mice showed that the absence of BMPRIB signaling causes infertility with several reproductive defects, including decreased levels of P450arom expression in granulosa cells. Studies on the Booroola sheep revealed a naturally occurring point mutation in the *BmprIB* gene, which is associated with increased ovulation rate; however, the consequence of this mutation on BMPRIB activity is currently unknown. This is especially relevant since the phenotype of the BMPRIB deficient mice is opposite to the phenotype of the Booroola sheep, and may suggest that the sheep model carries an activating mutation of the BMPRIB.

In the adult rat ovary, MISRII mRNA is found mainly in granulosa cells from preantral and small antral follicles, but primordial follicles seem to be deprived of MISRII, which is not consistent with the proposed role of MIS on the regulation of primordial follicle recruitment. This suggests that either MIS has a secondary effect on primordial follicles or that MIS signals through alternative pathways where MISRII is not involved.

In the ovary, activin is probably the most studied member of the TGF-β superfamily. Studies with radiolabeled activin revealed binding sites on rat granulosa cells during all phases of the ovarian cycle, as well as theca cells in developing follicles. Consistent with this, the expression of activin receptor mRNAs has been detected in the ovaries at various developmental stages (Table 4).

Tab. 4 Expression of activin receptor transcripts in the ovary.

Activin receptors	*Localization*	*Species*
Act RII	Oocytes	Rat, mouse
	Granulosa cells	Rat
	Granulosa-luteal cells	Human
Act RIIB	Oocytes	Mouse
	Granulosa cells	Human, rat
Act RIA (ALK-2)	Oocytes	Rat
	Antral and preantral follicles	Rat
	Granulosa-luteal cells	Human
Act RIB (ALK-4)	Oocytes and cumulus cells	Human, mouse
	Granulosa-luteal cells	Human

Although inhibin is a potent regulator of activin signaling, the presence of inhibin-specific receptors capable of activin-independent signaling has not been identified. Recent studies, however, have indicated that betaglycan, a membrane-bound proteoglycan originally identified as the type III receptor for TGF-β, can bind inhibin with high affinity and can mediate inhibin functional antagonism of activin signaling. Consistent with this, expression of betaglycan and its transcripts has been localized to different regions of the rat reproductive axis, in particular, to all cell types of the pituitary, being greatest in the gonadotropes, the target sites of inhibin in the endocrine control of FSH. Moderate levels of betaglycan mRNA were also detected in ovarian granulosa cells with lower expression in the thecal layer and the oocyte. In the testes, betaglycan mRNA was observed in the Leydig and tubule-specific germ cells.

2.4 Insulin-like Growth Factor (IGF) System

2.4.1 IGFs, IGF-binding Proteins (IGFBPs) and IGF-receptors

IGF-I and IGF-II are single-chain polypeptides, which are structurally related to insulin and play an important role in growth and development. Both IGFs play essential roles in developmental growth, as demonstrated by gene knockout experiments in mice. Deletion of the IGF-I gene caused growth retardation, and most animals died after birth. Those that survived are severely growth retarded with developmental defects including impaired ovarian follicular growth, suggesting that IGF-I is critical for prenatal and postnatal growth and development. In contrast, IGF-II null mice were born small but showed normal postnatal growth suggesting that IGF-II might be primarily responsible for intrauterine growth. IGF-I is synthesized by a variety of cell types, and plays a role in ovarian folliculogenesis by potentiating FSH-stimulated cAMP production (see Sect. 2.2.2), aromatase activity, and LH receptor expression by granulosa cells.

The IGF system is composed of three different component systems – the two IGF ligands (IGF-I and IGF-II), two receptors (IGF-I or type I receptor, and IGF-II or type II receptor), and six IGF-binding proteins (IGFBP-1 to -6). In addition, there are specific IGFBP proteases as well as IGFBP-related proteins (IGFBPrP). Altogether, these IGF system components determine and modulate the net actions of IGFs in target tissues.

2.4.2 Gene Structure and Expression in the Ovary

IGF-1 and IGF-II The human IGF-1 gene is located in chromosome 12 (Table 5), and has six exons spanning 80 kb. Owing to two alternative translation start sites in exons 1 and 2, different mRNA species

Tab. 5 Chromosomal localization of the human genes for the various IGF components.

IGF components	*Chromosomal localization*
IGF-I	12q22-q23
IGF-II	11p15.5
IGF-R1	15q25-q26
IGF-RII	6q26
IGFBP-1	7p13-p12
IGFBP-2	2q33-q34
IGFBP-3	7p13-p12
IGFBP-4	17q12-q21.1
IGFBP-5	2q33-q36
IGFBP-6	12q13
ALS	16p13.3

of 7.6, 1.3, and 1.1 kb are present in different proportions in different tissues. Although protein-coding sequences are found in exons 3 and 4, alternative splicing of exon 5 and 6 give rise to preproIGF-IA and IGF-IB respectively. IGF-I mRNA showed low and stage-specific expression in the ovary, being detectable only in the thecal cells of small antral follicles but not in the dominant follicle. IGF-I mRNA is absent is the granulosa cells of both antral and dominant follicles. Immunoreactive IGF-I, however, has been found in follicular fluids, but at levels lower than that present in peripheral plasma. It is likely that the follicular fluid IGF-I may be derived from the circulation suggesting that circulating IGF-I may regulate folliculogenesis in an endocrine manner, rather than intraovarian IGF-I acting locally.

The IGF-II gene is located in chromosome 11 contiguous to the insulin gene. It has a complex structure comprising of 4 promoters, 9 exons, and 2 polyadenylation sites, resulting in the expression of multiple IGF-II mRNA transcripts of 6, 5.3, 5, 4.8, 2.2, and 1.8 kb. In addition, expression of the IGF-II gene also showed transcriptional repression of the maternal allele due to genomic imprinting. Unlike IGF-I, IGF-II is the principal IGF in human ovarian follicles. IGF-II mRNA is expressed in the thecal cells of small antral but not dominant follicles. Granulosa cells from dominant but not small antral follicles expressed IGF-II mRNA. This indicated that as the follicles developed, the source of IGF-II was switched from thecal to granulosa cells. Expression of the IGF-II mRNA in the granulosa cells is regulated by gonadotropins and cAMP analogs. Follicular fluid contains immunoreactive IGF-II, at levels higher than that present in peripheral plasma, indicating the intraovarian source of IGF-II. Since there appears to be a coordinate regulation of IGF-II mRNA and the cholesterol side chain cleavage enzyme, P450scc, in human steroidogenic tissues, locally produced IGF-II may regulate steroidogenesis and the subsequent development of the dominant follicles.

IGF-receptors Although most of the biological functions of IGF-I and IGF-II are mediated by the type I IGF receptor, both type I and type II receptors are expressed in the ovary. The IGF-I receptor (type I) shows ligand-dependent tyrosine kinase activity and autophosphorylation, and has considerable structural and functional similarity to the insulin receptor but has higher affinity for IGF-I and IGF-II than for insulin. The signaling receptor is a heterotetramer comprising two α- and two β-subunits, and shows considerable overlap in signal transduction events to that of the insulin receptor including tyrosine phosphorylation of the β-subunits, interactions with docking proteins like IRS and Shc, and activation of PI-3 kinase and MAPK. Mice lacking the IGF-I receptor gene are small at birth and die shortly afterwards. Patients with an allelic deficiency in the type I IGF receptor gene show intrauterine and postnatal growth restriction, indicating that IGF-I is essential for embryonic growth. In the ovary, type I receptor mRNA is expressed in the granulosa cells of both small antral and dominant follicles, and is absent in thecal cells.

In contrast, the type II receptor (also known as the cation-independent mannose-6-phosphate receptor) is a single integral membrane protein, which can interact with a diverse group of ligands including peptide growth factors and lysosomal enzymes. It has a higher affinity for IGF-II than IGF-I and does not bind

insulin. Binding of IGF-II to the receptor results in the rapid internalization of IGF-II and its subsequent degradation by the lysosomes. This degradative process appears to be an important regulator of IGF-II activity, since mice deficient in the type II receptor show elevated IGF-II levels and die *in utero* due to somatic overgrowth and organ abnormalities suggestive of overstimulation from the type I receptor. Besides degrading IGF-II, recent studies have indicated that the type II receptor may have IGF-II-dependent signaling function through G protein–coupled receptor pathways.

Unlike the type I receptor mRNA whose expression is limited to the granulosa cells, the type II receptor mRNA is expressed in both the thecal and the granulosa cells from both antral and dominant follicles. The differential presence of both ligand and receptor in the same cells suggested that in small antral follicles, IGF-II could act in an autocrine manner in thecal cells, but not in granulosa cells. In the dominant follicles, granulosa-derived IGF-II can also act in an autocrine manner in granulosa cells.

IGFBPs In plasma, almost all of the IGFs circulate as high molecular weight complexes. The majority of the IGFs are present in a 150 kDa ternary complex comprising one molecule each of IGF, IGFBP-3 (the predominant IGFBP in serum) or IGFBP-5, and an 85 kDa glycoprotein, the acid-labile subunit (ALS). Plasma also contains lower molecular weight complexes of ~50 kDa, which are made up of several other IGFBP species (IGFBP-1, -2, -4 and -6) that are incompletely saturated with IGFs leaving virtually no free IGFs in circulation. This circulating source of IGFs, which has been regarded as contributing to the endocrine effects of IGFs, are largely determined by growth hormone status.

In extravascular biological fluids, however, most IGFs are bound to the IGFBPs as binary complexes since the ternary complex is unable to cross the endothelial barrier. IGFBPs function to determine the bioavailability of the IGFs, and therefore modulate its actions locally. The family of 6 IGFBPs (IGFBP-1 to -6) are structurally related to each other, and bind IGF-I and IGF-II with high affinity but share no homology with the IGF-receptors. These proteins are encoded by different genes, and have different biochemical and biological characteristics, which are due to differences in their primary structure.

All IGFBPs except IGFBP-6 have been found in the granulosa and thecal cells of the ovary (Table 6). Levels of IGFBPs in the follicular fluids vary depending on the developmental stage of the follicle.

IGFBP-1: Though found at very low concentrations in adult serum, it is the predominant IGFBP present in human amniotic fluid. It has equal affinity for both IGF-I and IGF-II, and is known to be serine-phosphorylated to varying extents depending on the tissue source. Dephosphorylated forms have a lower affinity for IGF-I and are thought to potentiate IGF actions, whereas phosphorylated forms may inhibit the mitogenic actions of IGFs.

The *IGFBP-1* gene is located on human chromosome 7p14-p12 where it is contiguous in a tail-to-tail configuration with that of *IGFBP-3*. The gene spans 5.2 kb and has 4 exons which contain all the coding regions of the gene. In the human ovary, IGFBP-1 is found only in the granulosa cells of dominant follicles, and is not present in the theca or small antral follicles. IGFBP-1 is also expressed in the corpus luteum.

Tab. 6 Expression of the various IGF components in the human female reproductive system. + indicates detectables as mRNA transcripts, − indicates no mRNA transcripts were detected.

IGF components	Endometrial cells		Small anthral follicles		Dominant follicles	
	Proliferative	Secretory	Thecal	Granulosa	Thecal	Granulosa
IGF-I	+	+	+	−	−	−
IGF-II	+	+	+	−	−	+
IGF-R1	+	+	−	+	−	+
IGF-RII	+	+	+	+	+	+
IGFBP-1	−	+	−	−	−	+
IGFBP-2	+	+	+	+	+	−
IGFBP-3	+	+	+	−	+	+
IGFBP-4	+	+	+	+	+	+
IGFBP-5	+	+	+	+	+	+
IGFBP-6	+	+	−	−	−	−

IGFBP-2: This is the most abundant IGFBP in human fetal serum during mid-gestation. The human *IGFBP-2* gene contains 4 exons, and maps to chromosome 2q33-q34. It spans ~32 kb of chromosomal DNA, with the first intron covering about 27 kb in size. IGFBP-2 is expressed in both the granulosa and thecal cells of small antral follicles. In dominant follicles, IGFBP-2 is expressed only in the thecal cells, suggesting a developmental loss of IGFBP-2 expression in the granulosa cells as the follicles become dominant.

IGFBP-3: IGFBP-3 is the most abundant circulating IGFBP in the postnatal animal. Serum levels increase with postnatal age to peak at puberty, reaching as high as $4\,\mu g\,mL^{-1}$. Most of the IGFBP-3 in circulation, when occupied by IGF-I or IGF-II, are complexed with ALS to form a 150 kDa ternary complex. IGFBP-3 expression is transcriptionally regulated by the growth hormone, and serum levels of IGFBP-3 are dependent on the growth hormone status. In the ovary, IGFBP-3 expression is found in the theca of all follicles, and the granulosa cells of dominant follicles. The gene for IGFBP-3 maps to chromosome 7p12-p14, about 20 kb from the *IGFBP-1* gene. The gene spans 8.9 kb and contains 4 coding exons.

IGFBP-4: IGFBP-4 is predominantly expressed in the liver, which represents the single largest source of IGFBP-4 in the adult rat. It is one of the most abundant IGFBP transcripts in the mid- and late-gestational embryo, particularly in mesodermally derived tissues like smooth and striated muscles, lungs, liver, and to a lesser extent, the heart. In the ovary, IGFBP-4 is expressed in both the thecal and granulosa cells of all follicles, and is a potent inhibitor of IGF-II and FSH-stimulated steroidogenesis in granulosa cells.

IGFBP-5: In humans, IGFBP-5 circulates at ~15% of IGFBP-3 levels, and like IGFBP-3, can also form ternary complexes with ALS when bound to IGF-I or IGF-II. IGFBP-5 is characterized by the presence of the 18 cysteine residues conserved among IGFBP-1 to -4, and lacks N-glycosylation site or RGD sequence. It contains a stretch of mostly basic amino acid residues that may mediate the binding

of IGFBP-5 to extracellular matrix and glycosaminoglycans. The human *IGFBP-5* gene maps to region 2q33–34 of chromosome 2, 20–40 kb from the *IGFBP-2* gene in a tail-to-tail orientation. Like IGFBP-4, IGFBP-5 is expressed in the theca and granulosa cells of all follicles.

IGFBP-6: In the endometrium, IGF-I expression is elevated during the estrogen-dominated proliferative phase, whereas IGF-II expression is elevated in secretory endometrium. Type I and type II IGF receptor mRNAs are both present in endometrial stroma, but are relatively more abundant in endometrial epithelium; neither show distinctive cyclical changes.

2.4.3 Roles of IGFs in Ovarian Function

IGF is an important growth factor that plays both an endocrine and a paracrine role in mammalian ovarian follicular development, growth of dominant follicles, and follicular atresia. IGF-I amplifies gonadotropin hormone actions through increasing steroidogenesis by granulosa cells and granulosa cell proliferation. Although in humans, IGF-II is the predominant endogenous IGF produced locally in the ovary, its actions appear to be similar to that of IGF-I. In granulosa cells, IGFs promote basal estrogen secretion through stimulation of aromatase activity and aromatase mRNA expression. IGFs can also augment the FSH-stimulated estrogen and progesterone secretion. In thecal cells, IGFs also stimulate DNA synthesis and can promote testosterone and androstenedione production.

The actions of IGFBP in the human ovary appear to be similar to that seen from studies using experimental animal models. These actions of IGFBPs are largely antagonistic toward IGFs, suggesting that the predominant role of IGFBPs is to sequester locally produced IGFs from binding to the IGF-receptors. In this way, IGFBPs can antagonize gonadotropin actions, and high intracellular IGFBP levels can lead to follicle atresia.

2.5 Other Peptides/Enzymes

2.5.1 Endothelins

Endothelin was described originally as a product of endothelial cells that had a potent vasoconstrictor effect on vessels by means of an action on the smooth muscle cells. Subsequently, its sites of production have expanded considerably to include many epithelial cells in the pituitary, ovary, endometrium, myometrium, and placenta. The endothelins are a family of three peptides (ET-1, ET-2, ET-3), encoded by three separate mammalian genes that are related to the gene for sarafotoxin in snakes. Each isoform is a 21-amino acid peptide, derived by posttranslational processing from a larger precursor peptide of around 200 amino acids. The endothelin-converting enzymes (designated ECE-1 and -2) that perform these cleavages have been cloned, but their specificities are still unclear. They may, together with neutral endopeptidase, be responsible for degradation of the ETs. The ETs act through one of three receptors (ETA, ETB, ETC), each of which has been cloned. Each receptor has a different affinity profile for the ETs, with ETA binding preferentially to ET-1 and predominating on smooth muscle cells of blood vessels, whereas ETB binds both ET-1 and ET-2 and has a wider cellular distribution. It has been suggested that the different receptor subtypes are coupled to different transduction systems, since such a relationship could explain the pluripotent actions of ETs, which now

include mitogenic and hormone-releasing activities as well as vasoconstriction.

2.5.2 Oxytocin

The uterotonic and milk-ejecting activities of the posterior pituitary and corpus luteum of the ovary were first described at the beginning of this century, but it was not until the early 1950s that the activity was identified as oxytocin, a cyclic nonapeptide. It was discovered that oxytocin is released by exocytosis of secretory granules, which also contain a carrier protein called neurophysin 1, which, with oxytocin forms part of a larger precursor molecule. The single-copy genes in all species are similar in structure and contain three exons, with exon A encoding the signal peptide, oxytocin and the amino-terminal region of neurophysin (amino acids 1–9) and exons B (amino acids 10–76) and C (amino acids 77–93) encoding the remainder of the neurophysin. Complex posttranslational processing of the precursor molecule by exo- and endopeptidases, a carboxypeptidase B-like enzyme, and an amidation enzyme gives rise to neurophysin and several pre forms of oxytocin, which associate with the neurophysin in the secretory granule before release from the cell.

The oxytocin gene is expressed in the cell bodies of the supraoptic and paraventricular nuclei of the hypothalamus of all species, and the precursor is processed en route along the axons of the median eminence and pituitary stalk to the storage granules of the posterior pituitary. Oxytocin released from this source is involved in the milk ejection reflex and the stimulation of uterine contractions.

The pregnant uterus is highly sensitive to the uterotonic effects of oxytocin, and is clinically used to induce labor. Expression of oxytocin is found in the human amnion, chorion, and decidua. Although there appears to be no apparent increase in oxytocin levels in the maternal plasma or intrauterine tissues during the onset of labor, there is a dramatic increase in oxytocin receptor expression in the myometrium, indicating that increased uterine sensitivity to oxytocin at this stage is due to changes in receptor levels.

Expression of the oxytocin gene in the ovary, particularly the corpus luteum, is well documented in ruminants, where it is believed to play a role in inducing the release of the uterine luteolysin prostaglandin F2α, which is responsible for regression of the corpus luteum of an infertile cycle. Clear understanding is lacking, however, of the level of expression and production of oxytocin in the ovary of other species, particularly humans, and its physiological role.

A single class of receptor sites for oxytocin has been characterized on the principal target cells, smooth muscle of the reproductive tract and myoepithelial cells of the mammary gland, as well as on the endometrium of sheep and humans. The human oxytocin receptor gene cDNA has been cloned. It encodes a 388 amino acid–polypeptide with seven-transmembrane domains typical of G protein–coupled receptors and is a member of the arginine–vasopressin/oxytocin family of receptors. The mRNA was found in two sizes, 3.6 kb in breast and 4.4 kb in ovary, uterine endometrium, and myometrium. The levels of oxytocin receptor mRNA are high in the myometrium, the maternally derived decidual cells, and the chorionic trophoblast cells at term, consistent with its role in human labor. The oxytocin receptor mRNA is also expressed in the sheep uterus, particularly in the endometrium, where oxytocin is responsible for releasing prostaglandin F2α. In all these cases, expression of the oxytocin receptor mRNA is

coincident with detection of the receptor protein and binding activity.

2.5.3 Zona Pellucida Proteins

The zona pellucida or egg surface coat is a transparent extracellular matrix surrounding mammalian oocytes, and is important for successful folliculogenesis, to ensure proper fertilization by the sperm, and to protect the embryo as it moves along the oviduct to the uterus before implantation. The zona pellucida consists almost entirely of three glycoproteins called ZP1, -2, and -3. Mouse ZP2 (120 kDa) and ZP3 (83 kDa) are present as heterodimers along the filaments that make up the extracellular coat of the egg, and the filaments are interconnected by ZP1 (200 kDa), a dimer of identical polypeptides held together by disulfide bonds. ZP3 has been identified as the primary sperm receptor on the mouse egg. There is a 22 amino acid–signal sequence, and the mature protein contains 402 amino acids rich in proline, serine, and threonine residues, as well as N- and O-linked glycosylation sites that make up the final molecular weight. ZP3 is encoded by a single-copy 8.6 kb gene containing eight exons; it yields a 1.5 kb mRNA, which is expressed only in the oocyte. Maximum expression occurs during oocyte growth, with levels falling and becoming undetectable after fertilization.

Sperm binding to mouse eggs is an example of carbohydrate-mediated cellular adhesion. Once bound, the sperm undergoes the acrosome reaction, which removes their egg-binding proteins; to remain bound, the inner acrosomal membrane of the sperm binds to ZP2, which acts as a secondary receptor and allows the sperm to penetrate the zona pellucida and fuse with the egg plasma membrane. Once fertilization has occurred, structural changes in the zona pellucida ensure that no more sperm bind and penetrate the matrix, thereby preventing polyspermic fertilization of the oocyte. The fertilized embryo, prior to implantation, continues to be surrounded by the zona pellucida, which is critical for their survival and development. Gene deletion studies in mice confirmed the importance of the zona pellucida proteins in folliculogenesis. ZP-1 deficient mice are fertile but showed reduced litter sizes due to early embryonic lethality. ZP-2 and ZP-3 deficient mice are infertile as early follicle development was impaired.

2.5.4 Leukemia Inhibitory Factor

Leukemia inhibitory factor (LIF) is a pleiotropic cytokine of the interleukin-6 family, and was initially identified by its ability to inhibit the proliferation of a myeloid leukemic cell line by inducing differentiation. It is now known that the 45 to 56 kDa glycoprotein has multiple activities including causing acute phase response in hepatocytes; regulation of differentiation and proliferation of hematopoietic cell lines; remodeling bone; and regulating growth; and inhibiting differentiation of embryonic stem cells, primordial germ cells, adipocytes, and endothelial cells. In adult mice, the highest expression of LIF mRNA is found in the endometrial epithelial cells, with the glandular expression of LIF being regulated throughout the reproductive cycle. Its expression is independent of the presence of a blastocyst and occurs transiently in response to a rise in E2 on day 4 of pregnancy, just prior to implantation. This transient expression of LIF was shown to be essential for implantation as shown by transgenic mice lacking LIF expression. These animals conceive, but the blastocysts fail to implant and develop, although the blastocysts remain viable and will develop to term if transferred to a

wild-type female. In human and primates, LIF is expressed in the endometrium in a menstrual cycle-dependent manner, with maximal expression observed at the time of implantation. Expression of LIF is localized to the glandular and luminal epithelial cells. LIF mRNA is absent or occurs at low levels during the proliferative or preovulatory phase. LIF is also expressed by the endothelial cells of the fallopian tube.

2.5.5 Tissue Remodeling Enzymes

The ovary and the uterus are unique among organs in having regular cyclic remodeling of the tissues. In the ovary, this involves the processes of ovulation (development of the follicles progressing through to oocyte extrusion), and the formation and regression of the corpus luteum (Fig. 2). In the uterus, there is the sloughing and regeneration of the endometrium at menstruation, tissue remodeling at implantation, preparation of the uterus and cervix for birth, and their reconstruction after birth. A major part of this process involves degradation of the extracellular matrix (ECM) by matrix metalloproteinases, which together with their associated inhibitors (serum macroglobulins, and TIMPS, tissue inhibitors of metalloproteinases) act coordinately to regulate the site and extent of connective tissue remodeling. The ECM is a protein network that consists of either an interstitial, collagen-rich, ground substance housing the stromal–interstitial cells, or the basement membrane underlying epithelial and stromal cells. Each has a characteristic composition that determines the type of MMP responsible for its degradation. For example, MMP-1, previously known as tissue collagenase, breaks down interstitial ECM, whereas MMP-2 and -9, which are gelatinases, are specific for the basement membrane ECM, which contains collagen IV, laminin, and fibronectin. The MMPs are neutral zinc proteinases that are secreted as latent zymogens requiring proteolysis for activation. The *in vivo* mechanisms of activation are not known but are thought to involve specific proteases, including plasmin and some of the MMPs themselves and an interaction with growth factors and cytokines such as tumor necrosis factor α (TNF-α), EGF, basic fibroblast growth factor (bFGF), TGF-β and interleukin-1 (IL-1), which are often found in association with the ECM.

The MMPs required for degrading interstitial and basement membrane ECM, and their inhibitors, can be found in the uterus and the ovary. Potential functions of MMPs in the uterus include angiogenesis, implantation, endometrial remodeling and menstruation, activation of growth factors such as FGF and the colony-stimulating factors (CSF) by releasing them from the ECM, preparation of the cervix for birth, and collagenolysis in the myometrium after birth. In the ovary, the process of ovulation requires breakdown by MMPs of both interstitial (by MMP-1) and basement membrane ECM (by MMP-2 and MMP-9) to allow release of the egg from the follicle. MMPs are also likely to be involved in the extensive angiogenesis and tissue remodeling that accompanies the formation of the corpus luteum from the recently ruptured follicle.

3 Pathophysiological Considerations

3.1 Genetic Defects

A number of genetic defects have been identified, which have helped to define the role of several hormones and pathways

in the female reproductive system. Some examples are given in this section.

Pulsatile GnRH from the hypothalamus acts on the GnRH receptor on the pituitary gonadotropes to regulate gonadotropin gene synthesis, hormone synthesis, and secretion. Any mutations that affect the proper functioning of this hypothalamic-pituitary axis will therefore impinge upon the reproductive physiology of the individuals concerned. Two known genetic diseases have the phenotype of loss of GnRH gene expression. The *hpg* mouse is infertile as a result of a lack of GnRH secretion from the hypothalamus due to the deletion of the GnRH gene. Transgenic insertion of a normal GnRH gene can restore reproductive function in these mice. In humans, however, individuals exhibiting a similar phenotype, hypogonadotropic hypogonadism, have impaired GnRH secretion despite having a normal GnRH gene. These individuals possess mutations that affect other genes that are important in the various levels of GnRH expression, synthesis, and actions. For example, individuals with Kallman's syndrome have infertility, which is associated with anosmia. The defect in Kallman's patients is due to an X-linked defect that arrests the migration of GnRH neurons from the olfactory placode to the hypothalamus during fetal development, such that in the adult there are no GnRH neurons in the hypothalamus. The *KAL* gene is located at Xp22.3 and encodes a neural adhesion molecule with structural similarities to neural cell adhesion molecule (NCAM). Mutations in the genes that are important for transcription factors (such as SF-1, Dax) and for gonadotrope development, GnRH biosynthesis (*PC-1* mutations causing abnormal processing of GnRH), or function (GnRHR mutation affecting GnRH action at the receptor level), cause varying degrees of hypogonadotropic hypogonadism. Interestingly, specific mutations of the human GnRH gene have not been found.

The gonadotropins, LH and FSH, are heterodimeric hormones containing a common α-subunit noncovalently bound to a hormone-specific β-subunit. Women with either homozygous or heterozygous mutations in the *FSHβ* gene have delayed puberty and primary amenorrhea. Although serum FSH was not detectable, serum LH was elevated, and treatment with exogenous FSH restores folliculogenesis, ovulation, and fertility, confirming the importance of FSH in antral development and granulosa cell steroidogenesis.

In familial male precocious puberty, an autosomal dominant, male-limited, gonadotropin-independent disorder, affected males begin puberty as young as age 4. The genetic defect leading to this syndrome has been shown to be due to a single base change in the LH receptor that results in an amino acid substitution in the sixth transmembrane domain, which in turn leads to constitutive activation of the receptor. Thus, cyclic AMP production is increased in the absence of LH, triggering Leydig cell activity and testosterone production. A similar syndrome is testotoxicosis, which also results in precocious puberty. The defect in this case is in the LH receptor linked G-protein Gsα, which activates adenylyl cyclase. A single base mutation again results in an amino acid substitution, causing constitutively activated Gsα, which elevates adenylyl cyclase and thus cAMP in the testis. This mutation, however, also results in resistance to hormones that activate cAMP (parathyroid hormone and TSH), and these two disparate effects are attributed to the mutated Gsα protein, which is unstable at 37 °C but retains function at the lower temperature of the testis. Both these syndromes

show a male-limited pattern of inheritance, probably because LH is sufficient to trigger production of testosterone in males, but both LH and FSH are necessary to trigger steroidogenesis in the ovary. Thus, constitutive activation of the LH receptor alone would not be sufficient to trigger precocious puberty in females.

McCune–Albright syndrome is characterized by polyostotic fibrous dysplasia, café au lait pigmentation of the skin, and multiple endocrinopathies, including precocious sexual development. The genetic defect in this syndrome is an activating mutation in the Gsα protein. This syndrome exhibits sporadic occurrence and is believed to be due to a dominant somatic mutation that occurs early in development, resulting in precocious puberty in both males and females, with females having luteinized follicular cells in the ovary at a very young age.

The importance of estrogen for normal development has been demonstrated in patients with genetic defects in the gene for P450arom, the key enzyme that catalyzes the conversion of testosterone to estradiol, androstenedione to estrone, and 16α-hydroxylated dehydroepiandrosterone to estriol. In females, absence of estrogen gives rise to ambiguous genitalia, and patients have hypergonadotropic hypogonadism at puberty, fail to develop secondary sexual characteristics, and exhibit progressive virilization. This demonstrates that estrogen is essential in the female for normal development of secondary sexual characteristics, including a pubertal growth spurt. Affected males have normal male sexual differentiation and pubertal maturation, but are tall with eunuchoid proportions. They exhibit continued linear growth into adulthood, and have unfused epiphysis, and osteoporosis due to estrogen deficiency. Although estrogen was shown to be essential for normal sperm production and function in mice, its role in fertility is not clear in men. These cases demonstrate that estrogen is important in both sexes for normal skeletal growth and development, and are supported by studies of transgenic mice lacking the *P450arom* gene.

3.2 FSH Isoforms and Infertility

There is now substantial data showing that pituitary and serum LH and FSH exist as structurally heterogeneous isoforms, and not as uniform structures. Variation in carbohydrate structure, particularly sialic acid content, accounts for most of this heterogeneity. Other factors (e.g. sulfation, phosphorylation, other modifications to the carbohydrate structure) are possible.

The population of isoforms can change according to the sex and endocrine status of the animal, involving mechanisms that are not understood. For example, human pituitary FSH isoforms are either larger and/or more acidic, with a longer half-life in the male than in the female. The physiological significance of changes in the relative proportions of FSH isoforms is more readily understood in terms of their differences in clearance rates than any difference in their intrinsic (in vitro) biological activity. It is not clear whether differences in the isoform profiles of gonadotropins in individuals are reflected in the subjects' fertility or infertility.

3.3 Reproductive Tumors and Inhibin

After the isolation and characterization of inhibin and development of radioimmunoassays, there have been several reports of elevated levels of inhibin

immunoactivity in the serum of patients with granulosa cell tumors of the ovary. Because inhibin is normally undetectable in the serum of postmenopausal women, the potential value of the peptide as a tumor marker is particularly significant in women without endogenous ovarian function. A number of studies have now indicated that in women with granulosa cell tumors, all forms of inhibin are increased, with the increase in either total serum inhibins or inhibin B being the most consistently elevated. While total serum immunoreactive inhibin is elevated in 100% of patients with granulosa cell tumors, inhibin B is elevated only in 89 to 100% of cases, and pro α-C in 90%, with inhibin A showing the greatest variability in their predictive value. Total serum immunoreactive inhibin is elevated in more than 80% of postmenopausal patients with mucinous carcinomas of the ovary, in 17% of patients with serous carcinomas, and in 17% of patients with clear cell carcinomas of the ovary. The nature of the inhibin product appears to vary with the type of tumor. Serum FSH levels vary inversely with inhibin levels in granulosa tumors, consistent with production of biologically active, dimeric inhibin. In mucinous tumors, on the other hand, FSH levels were generally high despite elevated inhibin, raising the possibility that in these tumors, the biologically inactive inhibin α-subunit precursor is the major product. Alternatively, the mucinous tumors may overproduce activin, a dimer of the β-subunits, which has FSH-stimulating activity. These data also suggest that expression of the three inhibin genes, α, β_A, and β_B, and the subsequent posttranslational processing and assembly of the subunits are regulated differently in the different tumors.

It is of interest in this regard that 70 of 71 transgenic mice with a deletion of the α-inhibin gene developed gonadal tumors, leading the authors to conclude that the inhibin α may be a tumor suppressor gene. An alternative hypothesis is that in the absence of inhibin α, activin is over-expressed and could be the agent responsible for tumor development. Whatever the explanation, the data are difficult to reconcile with the reports of elevated inhibin levels in patients with ovarian cancer, necessitating new concepts regarding the pathogenesis of ovarian tumors and their treatment.

Inhibin is also a product of the placenta, leading to the suggestion that it may be a useful marker of trophoblastic diseases, particularly hydatidiform mole, as an adjunct to the measurement of intact chorionic gonadotropin. Preliminary data support the potential of inhibin measurements for detecting molar pregnancies, where levels are elevated in comparison to those at the appropriate stage of normal pregnancy, but more data are needed.

4 Perspectives

Significant advances have been made in our understanding of the regulation of the female reproductive system in recent years, primarily in the areas of autocrine and paracrine regulation, and in the receptor and second-messenger systems, which subserve the actions of hormones and local regulators. Nevertheless, caution should be exercised in interpreting the available information. The existence of some of the putative local regulators is based on the detection of expression of the mRNA and protein *in situ* or *in vitro*. Their production and physiological role *in vivo* is less certain in many cases. Attempts to rectify this deficiency (e.g.

by using mice with gene deletions) have not always provided a clear indication of physiological roles, partly because of the possible substitution or compensation of one regulator for another in these circumstances.

The second-messenger systems that serve the pleiotropic actions of many of the hormones and regulators are not well understood, particularly where one hormone can have multiple effects on the same cell. There is still much to be learned about the transcription factors that link the hormone action to the gene and the mechanisms by which these factors regulate the genes. We now know from the example of the TGF-β receptors that association of different receptor subtypes can lead to different affinities for the ligands, which is another point at which regulation could occur. Until recently, there has been a tendency to overlook the importance of binding proteins and metabolizing enzymes in determining the local concentrations of *biologically active* hormones or factors; the molecular mechanisms controlling these substances are now under active investigation.

In addition to the reservations just noted, there are a number of key processes in the female reproductive system that we still do not understand. They include the following:

- The nature and regulation of the GnRH pulse generator
- The balance between endocrine and paracrine regulation of cells
- Control of reinitiation of growth and atresia of ovarian follicles
- Mechanisms controlling menstruation and implantation
- Immune responses to pregnancy.

Molecular approaches will play a major part in solving these inadequacies.

See also Genetics, Molecular Basis of; Male Reproductive System.

Bibliography

Books and Reviews

Britt, K.L., Findlay, J.K. (2002) Estrogen actions in the ovary revisited, *J. Endocrinol.* **175**, 269–276.

Chang, H., Brown, C.W., Matzuk, M.M. (2002) Genetic analysis of the mammalian transforming growth factor-beta superfamily, *Endocrinol. Rev.* **23**, 787–823.

Findlay, J.K., Drummond, A.E., Britt, K.L., Dyson, M., Wreford, N.G., Robertson, D.M., Groome, N.P., Jones, M.E., Simpson, E.R. (2000) The roles of activins, inhibins and estrogen in early committed follicles, *Mol. Cell. Endocrinol.* **163**, 81–87.

Findlay, J.K., Drummond, A.E., Dyson, M.L., Baillie, A.J., Robertson, D.M., Ethier, J.F. (2002) Recruitment and development of the follicle; the roles of the transforming growth factor-beta superfamily, *Mol. Cell. Endocrinol.* **191**, 35–43.

Giudice, L.C. (2001) Insulin-like growth factor family in Graafian follicle development and function, *J. Soc. Gynecol. Invest.* **8**, S26–S29.

Knobil, E., Neill, J.D. (Eds.) (1994) *The Physiology of Reproduction*, Vols. 1 and 2, 2nd edition, Raven Press, New York.

Ooi, G.T., Boisclair, Y.R. (1999) Molecular Biology of the IGFBPs, in: Rosenfeld, R., Roberts, C. Jr. (Eds.) *Contemporary Endocrinology: The IGF System*, Humana Press, Totowa, NJ, USA, pp. 111–139.

Palter, S.F., Tavares, A.B., Hourvitz, A., Veldhuis, J.D., Adashi, E.Y. (2001) Are estrogens of import to primate/human ovarian folliculogenesis? *Endocr. Rev.* **22**, 389–424.

Richards, J.S., Russell, D.L., Ochsner, S., Hsieh, M., Doyle, K.H., Falender, A.E., Lo, Y.K., Sharma, S.C. (2002) Novel signaling pathways that control ovarian follicular development, ovulation, and luteinization, *Recent Prog. Horm. Res.* **57**, 195–220.

Themmen, A.P.N., Huhtaniemi, I.T. (2000) Mutations of gonadotropins and gonadotropin receptors: elucidating the physiology and pathophysiology of pituitary-gonadal function, *Endocr. Rev.* **21**, 551–583.

Primary Literature

Berisha, B., Schams, D., Miyamoto, A. (2002) The expression of angiotensin and endothelin system members in bovine corpus luteum during estrous cycle and pregnancy, *Endocrine* **19**, 305–312.

Bondestam, J., Kaivo-oja, N., Kallio, J., Groome, N., Hyden-Granskog, C., Fujii, M., Moustakas, A., Jalanko, A., ten Dijke, P., Ritvos, O. (2002) Engagement of activin and bone morphogenetic protein signaling pathway Smad proteins in the induction of inhibin B production in ovarian granulosa cells, *Mol. Cell. Endocrinol.* **195**, 79–88.

Cataldo, N.A., Giudice, L.C. (1992) Insulin-like growth factor binding protein profiles in human ovarian follicular fluid correlate with follicular functional status, *J. Clin. Endocrinol. Metab.* **74**, 821–829.

Chang, S.Y., Tsai, M.Y., Huang, F.J., Kung, F.T. (2002) Expression of insulin-like growth factor (IGF), IGF receptor, and IGF-binding protein messenger ribonucleic acids in luteinized granulosa cells from different size follicles after controlled ovarian hyperstimulation, *J. Assist. Reprod. Genet.* **19**, 121–126.

Eimerl, S., Orly, J. (2002) Regulation of steroidogenic genes by insulin-like growth factor-1 and follicle-stimulating hormone: differential responses of cytochrome P450 side-chain cleavage, steroidogenic acute regulatory protein, and 3beta-hydroxysteroid dehydrogenase/isomerase in rat granulosa cells, *Biol. Reprod.* **67**, 900–910.

el-Roeiy, A., Chen, X., Roberts, V.J., LeRoith, D., Roberts, C.T. Jr., Yen, S.S. (1993) Expression of insulin-like growth factor-I (IGF-I) and IGF-II and the IGF-I, IGF-II, and insulin receptor genes and localization of the gene products in the human ovary, *J. Clin. Endocrinol. Metab.* **77**, 1411–1418.

el-Roeiy, A., Chen, X., Roberts, V.J., Shimasakai, S., Ling, N., LeRoith, D., Roberts, C.T. Jr., Yen, S.S. (1994) Expression of the genes encoding the insulin-like growth factors (IGF-I and II), the IGF and insulin receptors, and IGF-binding proteins-1-6 and the localization of their gene products in normal and polycystic ovary syndrome ovaries, *J. Clin. Endocrinol. Metab.* **78**, 1488–1496.

Ethier, J.F., Farnworth, P.G., Findlay, J.K., Ooi, G.T. (2002) Transforming growth factor-beta modulates inhibin A bioactivity in the LbetaT2 gonadotrope cell line by competing for binding to betaglycan, *Mol. Endocrinol.* **16**, 2754–2763.

Fabre, S., Pierre, A., Pisselet, C., Mulsant, P., Lecerf, F., Pohl, J., Monget, P., Monniaux, D. (2003) The Booroola mutation in sheep is associated with an alteration of the bone morphogenetic protein receptor-IB functionality, *J. Endocrinol.* **177**, 435–444.

Flores, J.A., Garmey, J.C., Lahav, M., Veldhuis, J.D. (1999) Mechanisms underlying endothelin's inhibition of FSH-stimulated progesterone production by ovarian granulosa cells, *Mol. Cell. Endocrinol.* **156**, 169–178.

Fried, G., Wramsby, H., Tally, M. (1998) Transforming growth factor-beta1, insulin-like growth factors, and insulin-like growth factor binding proteins in ovarian follicular fluid are differentially regulated by the type of ovarian hyperstimulation used for in vitro fertilization, *Fertil. Steril.* **70**, 129–134.

Geisthovel, F., Moretti-Rojas, I., Asch, R.H., Rojas, F.J. (1989) Expression of insulin-like growth factor-II (IGF-II) messenger ribonucleic acid (mRNA), but not IGF-I mRNA, in human preovulatory granulosa cells, *Hum. Reprod.* **4**, 899–902.

Giudice, L.C., van Dessel, H.J., Cataldo, N.A., Chandrasekher, Y.A., Yap, O.W., Fauser, B.C. (1995) Circulating and ovarian IGF binding proteins: potential roles in normo-ovulatory cycles and in polycystic ovarian syndrome, *Prog. Growth Factor Res.* **6**, 397–408.

Glister, C., Groome, N.P., Knight, P.G. (2003) Oocyte-mediated suppression of follicle-stimulating hormone- and insulin-like growth factor-induced secretion of steroids and inhibin-related proteins by bovine granulosa cells in vitro: possible role of transforming growth factor alpha, *Biol. Reprod.* **68**, 758–765.

Gougeon, A., Busso, D. (2000) Morphologic and functional determinants of primordial and primary follicles in the monkey ovary, *Mol. Cell. Endocrinol.* **163**, 33–42.

Hernandez, E.R., Roberts, C.T. Jr., Hurwitz, A., LeRoith, D., Adashi, E.Y. (1990) Rat ovarian insulin-like growth factor II

gene expression is theca-interstitial cell-exclusive: hormonal regulation and receptor distribution, *Endocrinology* **127**, 3249–3251.

Ignar-Trowbridge, D.M., Pimentel, M., Parker, M.G., McLachlan, J.A., Korach, K.S. (1996) Peptide growth factor cross-talk with the estrogen receptor requires the A/B domain and occurs independently of protein kinase C or estradiol, *Endocrinology* **137**, 1735–1744.

Jaatinen, R., Bondestam, J., Raivio, T., Hilden, K., Dunkel, L., Groome, N., Ritvos, O. (2002) Activation of the bone morphogenetic protein signaling pathway induces inhibin beta(B)-subunit mRNA and secreted inhibin B levels in cultured human granulosa-luteal cells, *J. Clin. Endocrinol. Metab.* **87**, 1254–1261.

Jabara, S., Christenson, L.K., Wang, C.Y., McAllister, J.M., Javitt, N.B., Dunaif, A., Strauss, J.F. (2003) Stromal cells of the human postmenopausal ovary display a distinctive biochemical and molecular phenotype, *J. Clin. Endocrinol. Metab.* **88**, 484–492.

Johnson, A.L., Bridgham, J.T., Swenson, J.A. (2001) Activation of the Akt/protein kinase B signaling pathway is associated with granulosa cell survival, *Biol. Reprod.* **64**, 1566–1574.

Kaivo-Oja, N., Bondestam, J., Kamarainen, M., Koskimies, J., Vitt, U., Cranfield, M., Vuojolainen, K., Kallio, J.P., Olkkonen, V.M., Hayashi, M., Moustakas, A., Groome, N.P., ten Dijke, P., Hsueh, A.J., Ritvos, O. (2003) Growth differentiation factor-9 induces Smad2 activation and inhibin B production in cultured human granulosa-luteal cells, *J. Clin. Endocrinol. Metab.* **88**, 755–762.

Klein, N.A., Battaglia, D.E., Woodruff, T.K., Padmanabhan, V., Giudice, L.C., Bremner, W.J., Soules, M.R. (2000) Ovarian follicular concentrations of activin, follistatin, inhibin, insulin-like growth factor I (IGF-I), IGF-II, IGF-binding protein-2 (IGFBP-2), IGFBP-3, and vascular endothelial growth factor in spontaneous menstrual cycles of normal women of advanced reproductive age, *J. Clin. Endocrinol. Metab.* **85**, 4520–4525.

Lanuza, G.M., Groome, N.P., Baranao, J.L., Campo, S. (1999) Dimeric inhibin A and B production are differentially regulated by hormones and local factors in rat granulosa cells, *Endocrinology* **140**, 2549–2554.

Lewis, K.A., Gray, P.C., Blount, A.L., MacConell, L.A., Wiater, E., Bilezikjian, L.M., Vale, W. (2000) Betaglycan binds inhibin and can mediate functional antagonism of activin signalling, *Nature* **404**, 411–414.

Mann, R.J., Keri, R.A., Nilson, J.H. (2003) Consequences of elevated luteinizing hormone on diverse physiological systems: use of the LHbetaCTP transgenic mouse as a model of ovarian hyperstimulation-induced pathophysiology, *Recent Prog. Horm. Res.* **58**, 343–375.

Mazerbourg, S., Bondy, C.A., Zhou, J., Monget, P. (2003) The insulin-like growth factor system: a key determinant role in the growth and selection of ovarian follicles? A comparative species study, *Reprod. Domest. Anim.* **38**, 247–258.

McGee, E.A., Smith, R., Spears, N., Nachtigal, M.W., Ingraham, H., Hsueh, A.J. (2001) Mullerian inhibitory substance induces growth of rat preantral ovarian follicles, *Biol. Reprod.* **64**, 293–298.

McMullen, M.L., Cho, B.N., Yates, C.J., Mayo, K.E. (2001) Gonadal pathologies in transgenic mice expressing the rat inhibin alpha-subunit, *Endocrinology* **142**, 5005–5014.

McWilliam, R., Leake, R.E., Coutts, J.R. (1995) Growth factors in human ovarian follicle fluid and growth factor receptors in granulosa-luteal cells, *Int. J. Biol. Markers* **10**, 216–220.

Minegishi, T., Tsuchiya, M., Hirakawa, T., Abe, K., Inoue, K., Mizutani, T., Miyamoto, K. (2000) Expression of steroidogenic acute regulatory protein (StAR) in rat granulosa cells, *Life Sci.* **67**, 1015–1024.

Mulsant, P., Lecerf, F., Fabre, S., Schibler, L., Monget, P., Lanneluc, I., Pisselet, C., Riquet, J., Monniaux, D., Callebaut, I., Cribiu, E., Thimonier, J., Teyssier, J., Bodin, L., Cognie, Y., Chitour, N., Elsen, J.M. (2001) Mutation in bone morphogenetic protein receptor-IB is associated with increased ovulation rate in Booroola Merino ewes, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5104–5109.

Nilsson, E.E., Skinner, M.K. (2003) Bone morphogenetic protein-4 acts as an ovarian follicle survival factor and promotes primordial follicle development, *Biol. Reprod.* **69**, 1265–1272.

Rankin, T.L., O'Brien, M., Lee, E., Wigglesworth, K., Eppig, J., Dean, J. (2001) Defective zonae pellucidae in Zp2-null mice disrupt folliculogenesis, fertility and development, *Development* **128**, 1119–1126.

Rankin, T., Talbot, P., Lee, E., Dean, J. (1999) Abnormal zonae pellucidae in mice lacking

ZP1 result in early embryonic loss, *Development* **126**, 3847–3855.

Rankin, T.L., Tong, Z.B., Castle, P.E., Lee, E., Gore-Langton, R., Nelson, L.M., Dean, J. (1998) Human ZP3 restores fertility in Zp3 null mice without affecting order-specific sperm binding, *Development* **125**, 2415–2424.

Richards, J.S., Sharma, S.C., Falender, A.E., Lo, Y.H. (2002) Expression of FKHR, FKHRL1, and AFX genes in the rodent ovary: evidence for regulation by IGF-I, estrogen, and the gonadotropins, *Mol. Endocrinol.* **16**, 580–599.

Samaras, S.E., Guthrie, H.D., Barber, J.A., Hammond, J.M. (1993) Expression of the mRNAs for the insulin-like growth factors and their binding proteins during development of porcine ovarian follicles, *Endocrinology* **133**, 2395–2398.

Schams, D., Berisha, B., Kosmann, M., Amselgruber, W.M. (2002) Expression and localization of IGF family members in bovine antral follicles during final growth and in luteal tissue during different stages of estrous cycle and pregnancy, *Domest. Anim. Endocrinol.* **22**, 51–72.

Schilling, B., Yeh, J. (1999) Expression of transforming growth factor (TGF)-beta1, TGF-beta2, and TGF-beta3 and of type I and II TGF-beta receptors during the development of the human fetal ovary, *Fertil. Steril.* **72**, 147–153.

Seto-Young, D., Zajac, J., Liu, H.C., Rosenwaks, Z., Poretsky, L. (2003) The role of mitogen-activated protein kinase in insulin and insulin-like growth factor I (IGF-I) signaling cascades for progesterone and IGF-binding protein-1 production in human granulosa cells, *J. Clin. Endocrinol. Metab.* **88**, 3385–3391.

Shull, M.M., Doetschman, T. (1994) Transforming growth factor-beta 1 in reproduction and development, *Mol. Reprod. Dev.* **39**, 239–246.

Sirotkin, A.V., Makarevich, A.V., Kwon, H.B., Kotwica, J., Bulla, J., Hetenyi, L. (2001) Do GH, IGF-I and oxytocin interact by regulating the secretory activity of porcine ovarian cells?, *J. Endocrinol.* **171**, 475–480.

Srivastava, M.D., Lippes, J., Srivastava, B.I. (1996) Cytokines of the human reproductive tract, *Am. J. Reprod. Immunol.* **36**, 157–166.

Teixeira, R.J., Silva, V.C., Gazolla, H.M., Cunha, S.B., Guimaraes, M.M. (2002) The relationship between ovarian structure and serum insulin, insulin-like growth factor-I (IGF-I) and its binding protein (IGFBP-1 and IGFBP-3) levels in premature pubarche, *J. Pediatr. Endocrinol. Metab.* **15**, 69–75.

Villavicencio, A., Iniguez, G., Johnson, M.C., Gabler, F., Palomino, A., Vega, M. (2002) Regulation of steroid synthesis and apoptosis by insulin-like growth factor I and insulin-like growth factor binding protein 3 in human corpus luteum during the midluteal phase, *Reproduction* **124**, 501–508.

Vitt, U.A., Hsueh, A.J. (2001) Stage-dependent role of growth differentiation factor-9 in ovarian follicle development, *Mol. Cell. Endocrinol.* **183**, 171–177.

Voutilainen, R., Franks, S., Mason, H.D., Martikainen, H. (1996) Expression of insulin-like growth factor (IGF), IGF-binding protein, and IGF receptor messenger ribonucleic acids in normal and polycystic ovaries, *J. Clin. Endocrinol. Metab.* **81**, 1003–1008.

Wilson, T., Wu, X.Y., Juengel, J.L., Ross, I.K., Lumsden, J.M., Lord, E.A., Dodds, K.G., Walling, G.A., McEwan, J.C., O'Connell, A.R., McNatty, K.P., Montgomery, G.W. (2001) Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of bone morphogenetic protein IB receptor (ALK-6) that is expressed in both oocytes and granulosa cells, *Biol. Reprod.* **64**, 1225–1235.

Wong, B.C., Boyd, C.A., Lanzendorf, S.E. (2003) Randomized controlled study of human zona pellucida dissection using the zona infrared laser optical system: evaluation of blastomere damage, embryo development, and subsequent hatching, *Fertil. Steril.* **80**, 1249–1254.

Zhou, J., Wang, J., Penny, D., Monget, P., Arraztoa, J.A., Fogelson, L.J., Bondy, C.A. (2003) Insulin-like growth factor binding protein 4 expression parallels luteinizing hormone receptor expression and follicular luteinization in the primate ovary, *Biol. Reprod.* **69**, 22–29.

Fingerprinting, DNA: *see* DNA Typing Analysis

Flow Cytometry

Michael G. Ormerod
34 Wray Park Road, Reigate, UK

1 Introduction 417

2 The Instrument 417
2.1 The Flow Chamber 417
2.2 Fluidics 418
2.3 Light Source 418
2.4 Optics 419
2.5 Signal Processing 419

3 Measurement 420
3.1 Light Scatter 420
3.2 Fluorescence 420
3.3 Data Processing 420
3.4 Cell Sorting 421

4 General Applications of Flow Cytometry 423
4.1 Immunofluorescence 423
4.2 Measurement of DNA 424
4.3 Other Flow Cytometric Methods 425
4.3.1 Measuring Enzyme Kinetics 425
4.3.2 Following Changes in Membrane Potential 425
4.3.3 Intracellular Calcium Ions 426
4.3.4 Intracellular pH 426
4.3.5 Intracellular Glutathione 426
4.3.6 Measuring Oxidative Species 426

5 Specific Fields of Application 427
5.1 Clinical 427
5.2 Long Term Labels for Cells 427

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

5.3 Measurement of Cell Proliferation 427
5.3.1 The DNA Histogram 428
5.3.2 Using BrdUrd/Anti-BrdUrd 428
5.3.3 The BrdUrd-Hoechst/PI Method 430
5.3.4 Measuring Cell Proliferation through the Division of Label between Daughter Cells 431
5.4 Studying Cell Death 431
5.4.1 Necrosis 431
5.4.2 Apoptosis 432
5.5 Analyzing and Sorting Chromosomes 434
5.6 Microbead Technology 435

6 Applications Other Than Mammalian Cell Biology 435
6.1 Microbiology 435
6.2 Marine Biology 435

Bibliography 435
Books and Reviews 435
Primary Literature 436
Websites 438
Some Companies Who Supply Flow Cytometers 438
Free Software 438
Other Useful Sites 438

Keywords

Flow Cytometry

Cell Sorting

A flow cytometer measures between one and thirteen or more parameters on single cells. The flow system is used to deliver cells to the point of measurement. In most instruments, the cells are interrogated by laser light. Measurements are made of scattered light and of fluorescence, either from the cells or, more commonly, from labelled probes. Sub-populations of cells may also be physically sorted for further study. The article briefly describes the main attributes of the technology and its applications in clinical areas and in mammalian biology.

1 Introduction

Flow cytometry is widely used in fields as diverse as immunology, hematology, cell biology, marine biology, and microbiology. As its name implies, it is a technique for measuring cells in a flow system. Its most important features are that it makes measurements on single cells, it records between 2 and 13 parameters per cell depending on the instrument, and it can measure thousands of cells a second. These features enable the analysis of subpopulations, in particular, subsets with a frequency as low as 1 in 10^4 can be enumerated. Additionally, some instruments are capable of sorting subpopulations of cells with a purity of 99%. In addition to cells, other particles, such as nuclei, chromosomes or small beads, can also be studied.

The disadvantage of the technique is that it requires a suspension of single cells, which can be difficult to obtain from solid tissues, and, while the flow cytometer can measure the amount of an entity in a cell, it cannot measure its intracellular distribution.

I have previously published a succinct description of flow cytometry and its applications. Other, longer, introductory books as well as books that describe protocols for many applications of flow cytometry are also available.

2 The Instrument

2.1 The Flow Chamber

The flow chamber lies at the heart of the instrument. In it, cells are injected into a stream of either water or saline solution (called the *sheath fluid*). The sample stream is thereby hydrodynamically focused so that single cells are delivered to a defined point at which a light beam is focused (Fig. 1). The light source is usually a laser, although some, mainly older instruments, use an arc lamp. As cells pass through the laser beam, they will scatter light and may also fluoresce. While strong autofluorescence may be observed in some cells, particularly marine algae, cells are usually labeled with one or more of a variety of fluorescent markers.

Most instruments will collect light scattered by the cells in a narrow angle in a forward direction (forward scattered light, FS) and also light scattered at right angles to the laser beam (side scatter, SS). Fluorescences are collected and separated by optical filters into different wavelengths. Typically, four fluorescent parameters can be acquired, although in some instruments, this can be as many as eleven.

There are two basic designs of flow chamber in general use; the cuvette flow chamber and the "jet-in-air" or "stream-in-air." The latter is only used in instruments that can physically sort cells (see Sect. 3.4 below).

In the cuvette chamber, the interrogation point, at which the laser beam is focused on the sample, is inside the chamber (Fig. 1). In an analyzer, the sheath and sample streams are taken to waste in a fully enclosed system; in a cell sorter, the stream emerges into the open air. In a "stream-in-air" chamber, the interrogation point is immediately outside the chamber. While this system has advantages for cell sorting, it has inferior optical properties. The airstream interface scatters light, which has to be partially blocked by an additional obscuration bar fitted in front of the lens used

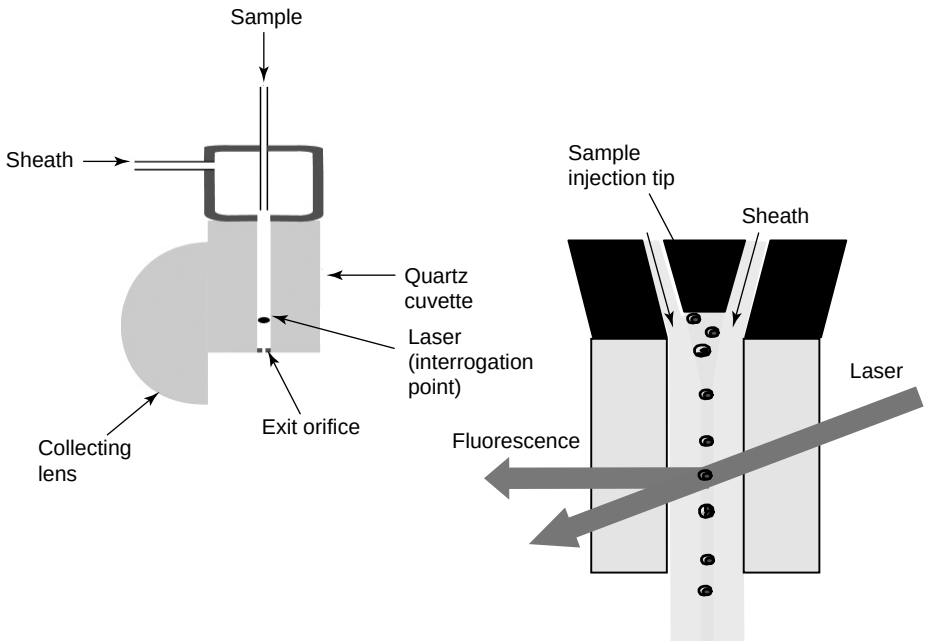

Fig. 1 Diagrammatic representation of a flow chamber. (See color plate p. xxiv.)

to collect fluorescent light at right angles to the laser beam. Because it is less efficient, it requires a laser of higher power.

2.2 Fluidics

The sheath fluid is normally forced through the flow chamber under pressure. The sample is usually injected into the sheath stream by applying a difference in pressure between sample and sheath fluid containers. The difference will control the volume flow rate of the sample. Alternatively, it may be injected into the stream using a motor connected to the piston of a syringe. The flow rate of the sample is then controlled by changing the speed of the motor. With this arrangement, a predefined volume of sample can be used, which enables absolute cell counting.

2.3 Light Source

Lasers are the preferred light source because they produce a single wavelength of high intensity. Nearly all instruments are fitted with an argon-ion laser, which gives blue light at 488 nm – a convenient wavelength for exciting the commonly used fluorochrome, fluorescein. Bench-top instruments are usually fitted with an air-cooled laser producing about 15-mW power. ''Stream-in-air'' sorters need more powerful lasers, frequently water-cooled, which have the advantage that they can, if required, be tuned to other wavelengths.

A second or third laser may be fitted so that cells can be excited at more than one wavelength. The choice depends on the applications; often a He–Ne laser giving red light at 630 nm is installed. A third laser will be selected to give UV, for

example, a He–Cd laser with an output at 325 nm. A wide range of solid state lasers are now available giving light at, for example, 405, 488, 530, or 635 nm.

2.4 Optics

The optical layout of typical flow cytometer is shown in Fig. 2.

Laser light is focused onto the point of measurement by a lens or lenses generally designed to give an elliptical laser spot. A blocker, or an obscuration bar is fitted behind the flow cell to block the laser beam. A light detector (such as a PIN diode) is positioned behind the obscuration bar to detect light scattered over a narrow angle in the forward direction.

Fluorescent and scattered light is collected at right angles to the laser beam and focused onto a series of photomultipliers (PMTs). Light of different wavelengths is selected using a series of dichroic and barrier (bandpass) optical filters. In the arrangement shown in Fig. 2, the dichroic filters are long pass filters, that is, they reflect light below a certain wavelength and pass light above that wavelength.

2.5 Signal Processing

Until recently, the signals from the PMTs were amplified and processed electronically before passing through analog-to-digital (ADC) converters, the digital signals being passed to a computer for data analysis. The trend now is to convert at an early stage and to manipulate the signals digitally. Using analog electronics, the signal is amplified in either a linear or logarithmic amplifier; using digital electronics, the amplified signal is presented on either a linear or a logarithmic scale. The signal may also be processed to output the width of the signal generated on the PMT, its height or the total integrated signal (area).

A threshold is set on the output voltage of one (or possibly two) of the measured parameters so that only signals above a certain limit are processed; this is sometimes referred to as the discriminator setting. The threshold, which is set by

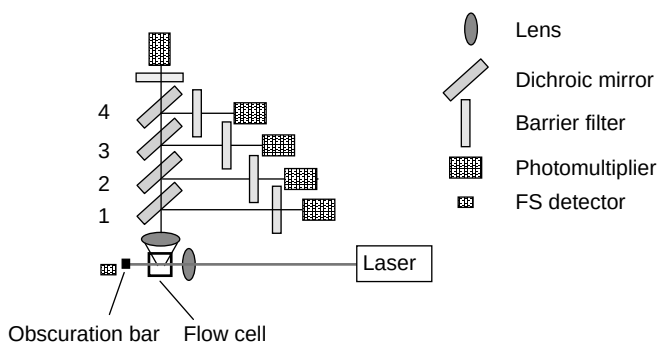

Fig. 2 The optical layout of a typical flow cytometer. Using an argon-ion laser emitting light at 488 nm, the optical components could be as follows: (1) Dichroic selecting light <500 nm; blue bandpass (488 nm). (2) Dichroic selecting light <540 nm; green bandpass (520 nm). (3) Dichroic selecting light <595 nm; orange bandpass (575 nm). (4) Dichroic selecting light <640 nm; red bandpass (620 nm) Final filter: long pass >650 nm. (See color plate p. xxiv.)

the operator, ensures that the instrument processes signals from cells (or other particles under study) and not from small debris or noise in the system.

3 Measurement

3.1 Light Scatter

Light scattered by particles as they pass through the laser beam is measured over a narrow angle in a forward direction (forward angle light scatter, FS) and also at right angles to the laser beam (side scatter, SS). The intensity of scatter is proportional to the size, shape and optical homogeneity of cells (or other particles); it is strongly dependent on the angle over which it is measured. Forward scatter may show variation between different flow cytometers reflecting the different geometry of the instruments.

Forward scatter tends to be more sensitive to the size and surface properties and can be used to distinguish live from dead cells while side scatter tends to be more sensitive to inclusions within cells and can be used to distinguish granulated from nongranulated cells.

3.2 Fluorescence

In instruments fitted with a laser giving blue light at 488 nm, the filter set will usually be selected to observe fluorescence from fluorescein (512 nm), phycoerythrin (PE) (560 nm) and either phycoerythrin-cyanine5 (PC5, 670 nm) or peridinin-chlorophyll (PerCP, 675 nm), these being the fluorochromes most frequently used for immunofluorescence work. In instruments with more than one laser and more detectors, up to 11 different fluorescences can be measured. There is a wide range of other fluorescent probes available (see, for example, the Molecular Probes catalogue). They are usually selected for use with one of the standard optical filters; for certain probes, specific filter sets may be required.

Most fluorochromes emit light over a wide spectral range. Consequently, light from one fluorochrome will be detected on more than one PMT. For example, light from fluorescein is selected using a green filter and from PE using an orange filter. However, the emission from fluorescein has an orange component, which will give a signal on the PE detector. This spectral overlap can confuse the interpretation of the data and has to be corrected. The compensation parameters can either be set up in the instrument, so that compensated data is recorded, or the data can be compensated post-analysis using software. While electronic compensation is adequate if up to four fluorescent parameters are being measured, compensation for larger numbers of fluorescent parameters is best handled digitally, either in real time or off-line.

3.3 Data Processing

Because of the large number of parameters collected on, perhaps, 20 000 cells, it is difficult to visualize correlations between the parameters. The strategy adopted for the analysis of multiparameter data is to use "gating." Two parameters are displayed in a correlated dot plot (often called a *cytogram*). A region is generated around a cluster of cells of interest; this region acts as a "gate." Another cytogram is generated from two other parameters and the computer is instructed to display

only those events that fall in the gate set on the first cytogram (see Fig. 3).

3.4 Cell Sorting

Some flow cytometers are able to sort selected subsets of cells. There are two types of cell sorters, piezo-electric and charged droplet.

The piezo-electric sorters separate cells mechanically. When a desired cell is detected, voltage is applied to a piezo-electric device, which deflects the cell into a collecting chamber. These sorters are relatively inexpensive, simple to operate

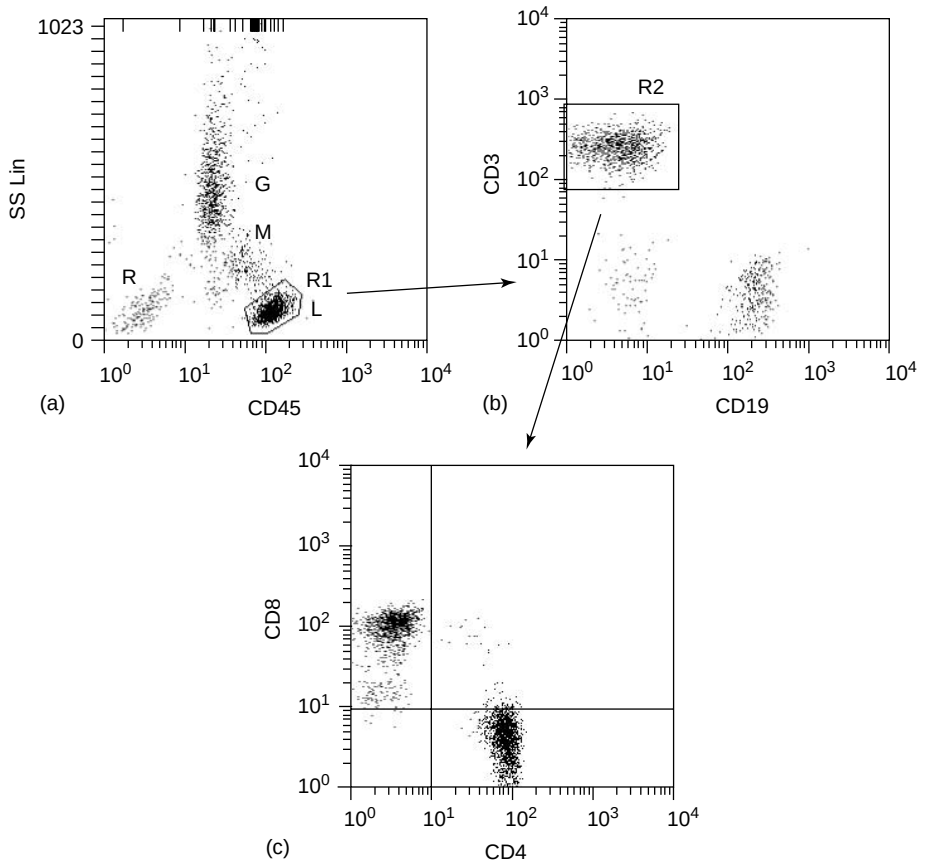

Fig. 3 Display of typical flow cytometric data. Peripheral blood leucocytes were labeled with antibodies to the cell surface antigens, CD3 (labeled with PE-cyanine5), CD4 (PE), CD8 (PE-Texas Red) and CD45 (fluorescein). (a) Initially, a cytogram of side scatter versus the pan leucocyte marker, CD45, was generated. Four clusters of cells are evident, cluster G is from granulocytes, M from monocytes, L from lymphocytes and R from residual red cells. A region (R1) was drawn around the lymphocytes. This was used as a gate set on subsequent cytograms. (b) Shows a cytogram of CD3 versus CD19 fluorescence of the lymphocytes only, identifying the T cells and B cells respectively. (c) Cytogram C has been gated on R1 and R2 and shows the T-cell subsets identified by CD4 and CD8. Data recorded on a Beckman Coulter XL using an argon-ion laser at 488 nm.

and slow (maximum sort rate is 300 cells s^{-1}). The sorted sample is always highly diluted, being mixed with a large amount of sheath fluid.

In charged droplet cell sorting, the stream from the flow cell emerges into the open air. The sheath fluid contains saline so that it is conductive. The flow cell is vibrated by a piezo-electric device, typically at 30 Hz, breaking the stream into well-defined droplets. When a cell to be sorted is in the next droplet to

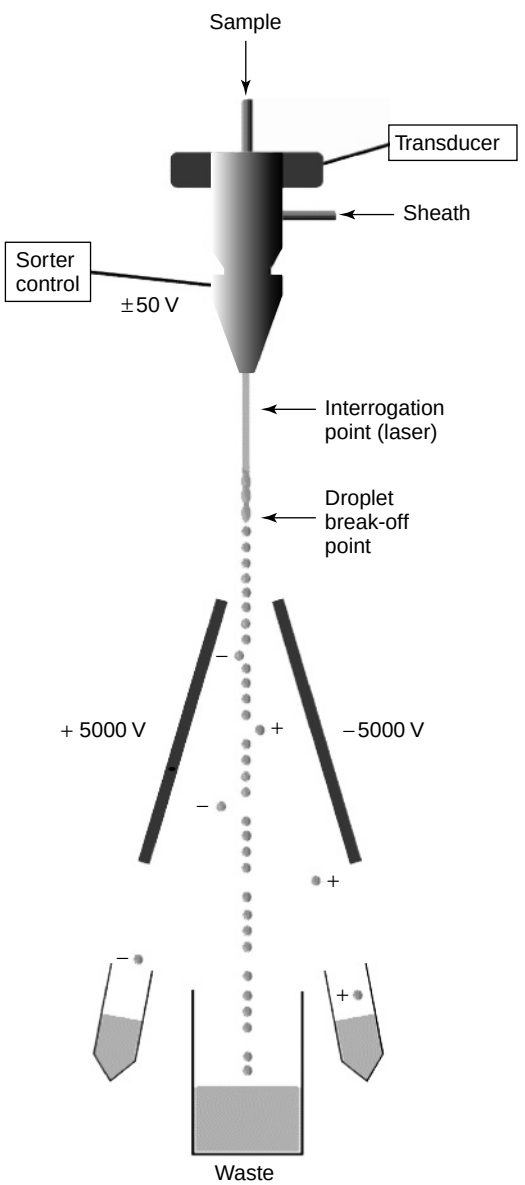

Fig. 4 A droplet deflection cell sorter.

be formed, the flow cell and stream is momentarily charged, thereby charging that droplet. The droplets pass through a pair of high voltage plates, charged droplets being deflected into the collecting vessels (Fig. 4).

The sorter has to be calibrated so that delay time between the cell passing through the laser beam (and being analyzed) and a droplet being formed is known to the computer, which will trigger the charge pulse at the appropriate moment. The conditions must be stable throughout the sort so that droplet delay time does not change. On modern machines, the instrument can monitor the position of the droplet break-off and automatically correct any minor variation.

The original charged droplet cell sorters could sort cells at speeds up to 5000 cells s^{-1}. The modern instruments sort at, at least, 15 000 cells s^{-1}; some machines can run at 50 000 cells s^{-1}. Despite the increase in speed, flow cytometric cell sorters give a low yield compared to bulk methods, such as immunomagnetic separation. If the concentration of the subpopulation to be sorted is 10%, at a flow rate of 10 000 cells s^{-1} and allowing for some losses through rejection of coincidences, the maximum number of cells collected will be about 3.10^6 h^{-1}. The major advantage is that the purity is high (>98%) and the sort decision can be based on a multiparametric analysis so that small subpopulations can be purified.

4 General Applications of Flow Cytometry

There are a wide variety of applications of flow cytometry in most aspects of cell biology. Table 1 lists the most important.

Tab. 1 Applications of flow cytometry.

Basic analysis
Immunofluorescence analysis
Cell surface markers
Intracellular and nuclear proteins
DNA content
Cell cycle analysis
Ploidy analysis
RNA content
Protein content
Measurement of functional parameters
Intracellular calcium ions
Intracellular pH
Intracellular glutathione
Membrane permeability
Membrane potential including mitochondrial membrane potential
Measuring oxidative burst
Specific topics
Cell sorting
Immunophenotyping
Analysis of cell proliferation
Analysis of cell death
Monitoring fusion of cells
Measurement of drug uptake
Analyzing and sorting chromosomes
Kinetic analysis of intracellular enzymes
Tracking cells *in vivo*
Observing binding and endocytosis of ligands
Monitoring electropermeabilization
Gene tracking using fluorescent proteins
Microbead assays

4.1 Immunofluorescence

This is the most widespread application of the technology. Any protein associated with a cell can be measured if a suitable antibody is available and if there are sufficient molecules present. The most commonly measured proteins, particularly in clinical laboratories, are those associated with leucocyte differentiation, the so-called, CD antigens, most of which are found on the surface of cells.

After fixing or permeabilizing cells, intracellular proteins can also be measured.

A variety of companies supply antibodies to most of the antigens of interest in routine clinical applications and for clinical and biological research. These antibodies are supplied labeled with different fluorochromes so that combinations can be selected to allow the simultaneous measurement of several proteins. Multicolor immunophenotypic analysis can reveal new, previously unrecognized, subsets of cells giving new insights into immunology and into the pathogenesis of diseases of the immune system, including cancer.

Fluorescence resonance energy transfer (FRET) is used in several applications in flow cytometry. If two fluorophores are sufficiently close to one another, energy may be transferred from one to the other. New fluorochromes for immunofluorescence have been made utilizing this principle. For example, if a conjugate of cyanine 5 and phycoerythrin (PE) is exposed to blue light, the PE will absorb the light, pass its energy to the cyanine, which fluoresces. The fluorescence is shifted from orange (PE alone) to red (cyanine 5). The principle can be applied to observe the interaction of two proteins on the surface of cells. Antibody to one protein is labeled with a donor dye, antibody to the other with an acceptor dye. When the proteins come into close contact, the fluorescence of the donor decreases while that of the acceptor increases.

4.2 Measurement of DNA

Measurement of the DNA content of cells gives information about the DNA ploidy and the cell cycle. There are several compounds whose fluorescence increases on binding to DNA and which bind stoichiometrically. The most commonly used of these is propidium iodide (PI), which is excited by blue light and fluoresces red.

Examples of DNA histograms are shown in Fig. 5. In order to estimate the percentage of cells in the G0/G1, S, and G2/M phases of the cell cycle, the overlap between G1 and early S and between G2 and late S has to be resolved. This is generally accomplished by a computer program that attempts to model the DNA histogram.

DNA measurement can be combined with immunofluorescence measurements (Fig. 6). If several immunofluorescent parameters are to be measured, PI may cause a problem because its fluorescence

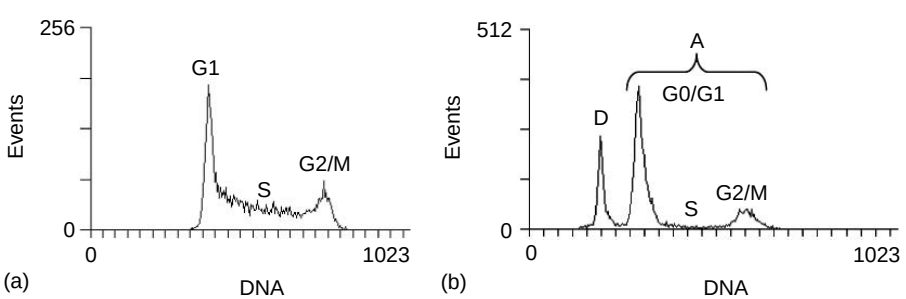

Fig. 5 DNA histograms. (a) Murine leukemic cell line. Cells were fixed in 70% ethanol, suspended in phosphate-buffered saline and stained with PI. The phases of the cell cycle are labeled. (b) Nuclei extracted from a formalin-fixed, paraffin-embedded biopsy of a breast carcinoma. D, diploid cells; A, aneuploid tumor cells, with the phases of the cell cycle labeled. Data recorded on a Beckman Coulter Elite using an argon-ion laser at 488 nm.

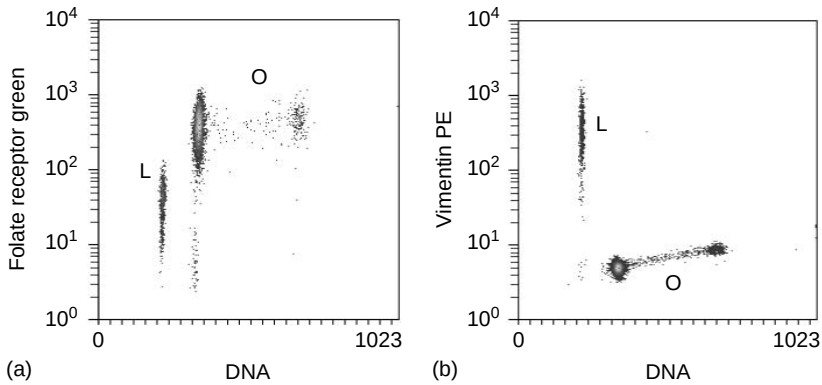

Fig. 6 Immunofluorescence and DNA measurement combined. Ascitic cells from a patient with ovarian carcinoma. The cells were labeled with anti-folate receptor-FITC, anti-vimentin-PE and PI. Note that there are both diploid and aneuploid cells present. (a) The diploid cells are leucocytes, which express vimentin (b) while the vimentin-negative ovarian carcinoma cells express folate receptor. Data, supplied by Willem Corver, University of Leiden, was recorded on a BD FACScan using an argon-ion laser at 488 nm.

spectrum overlaps strongly with commonly used fluorochromes, such as phycoerythrin and phycoerythrin-cyanine 5. Depending on which lasers are available, alternative dyes are often used; for example, a *bis*-benzimadazole, such as Hoechst 33342, which is excited by UV to give blue fluorescence, or the anthraquinone, DRAQ5, which fluoresces a deep red (700 nm).

4.3 Other Flow Cytometric Methods

4.3.1 Measuring Enzyme Kinetics

An enzyme substrate can be linked to a fluorescent molecule to give a nonfluorescent compound. The action of the enzyme will release the substrate leaving a fluorescent compound. Fluorophores, which have been used in this application, include fluorescein and rhodamine 110. Generally, this method will only give a crude estimate of enzyme activity since the kinetics of the reaction can be affected by the rate of diffusion of the complex into the cell, the rate of diffusion of the fluorophore out the cell and the released substrate out of the cell, and compartmentalization within the cell.

This approach has been adapted to enable charged compounds to be loaded into cells. For example, the calcium indicator, indo-1, is loaded into the cell by incubation with the acetoxymethyl ester. In the cell, esterases remove the acetoxymethyl groups leaving the fluorescent indo-1, which, being charged, remains trapped in the cell.

4.3.2 Following Changes in Membrane Potential

There are several lipophilic dyes that partition between the cell and the surrounding medium according to the plasma membrane potential. Of particular use are a series of oxonol dyes which, depending on their substituents, have different wavelengths of excitation and emission.

Many cyanine dyes, such as dicyanine, 3′,3′-dihexyloxacarbocyanine have a

positive charge and are taken up by the mitochondria and the strength of their fluorescence within the cell reflects the mitochondrial membrane potential. One of the cyanine dyes, abbreviated to JC-1, exists as a green fluorescent monomer at low concentrations and at low membrane potential. At high concentrations or higher membrane potential, it forms red-fluorescent aggregates. The dye can be excited at 488 nm and changes in mitochondrial membrane potential can be followed by measuring the ratio between green and red fluorescence.

4.3.3 Intracellular Calcium Ions

Ca^{2+} has an important role in cell signaling. The calcium indicator, indo-1, which is excited by UV, has a fluorescence emission maximum of about 490 nm in the absence of Ca^{2+}; in the presence of 1 mM Ca^{2+}, the emission maximum is about 410 nm. Changes in the concentration of intracellular Ca^{2+} are recorded by observing fluorescence at 400 and 520 nm and measuring the ratio between the two fluorescences.

Other dyes, which are sensitive to Ca^{2+} and can be excited by blue light, are fluo-3 and Fura Red. Ca^{2+} concentration does not affect the emission wavelength, just the intensity of emission so that these indicators cannot be used to measure changes in Ca^{2+} on a cell-by-cell basis; only the overall changes in a population of cells.

4.3.4 Intracellular pH

There are several compounds that can be used to measure intracellular pH; the most useful is SemiNaphthoFluorescein (SNARF-1), usually used as the carboxy-derivative. Under acidic conditions, carboxy-SNARF-1 fluoresces yellow-orange (580 nm), under alkaline conditions, it fluoresces red (640 nm). The pK_a is about 7.5 at room temperature and the pH is measured by recording the ratio between the fluorescences at 580 and 640 nm.

4.3.5 Intracellular Glutathione

Glutathione has an important role in the cell in reacting with potential damaging oxidative species. Increased levels of glutathione may also be associated with resistance to some cytotoxic drugs.

One assay for glutathione is based on the activity of the enzyme, glutathione-*S*-transferase. Cells are incubated with the dye, monochlorobimane (MClB), which crosses the plasma membrane. In the cytoplasm, the enzyme catalyzes the reaction with glutathione yielding a fluorescent conjugate, which is trapped in the cell. MClB-glutathione fluoresces blue when excited by UV. MClB gives excellent results with rodent cells. It is less effective in human cells, which lack the necessary variant of glutathione-*S*-transferase. As an alternative, monobromobimane, which reacts directly with sulphydryl groups, can be used.

4.3.6 Measuring Oxidative Species

There are several compounds which, in their reduced form, are nonfluorescent but which generate a fluorescent compound upon oxidation. Such compounds include dihydrorhodamine 123, dihydroethidium, and 2′,7′-dihydrodichlorofluorescein (dichlorofluorescin, DCFH). Dihydrorhodamine 123 and dihydroethidium diffuse freely into cells. Upon oxidation, rhodamine 123 is concentrated in the mitochondria and fluoresces green, while ethidium binds to DNA and fluoresces red.

DCFH can be loaded into cells as an ester. After conversion to DCFH by intracellular esterases, oxidation will yield a fluorescent product, 2′,7′-dichlorofluorescein.

These compounds have been used to observe oxidative burst in neutrophils, changes in the concentration of oxidative products during apoptosis and the production of intracellular oxidative species by ionizing radiation.

5 Specific Fields of Application

5.1 Clinical

Flow cytometry is universally used to classify leukemias. Immunophenotyping in combination with flow cytometry is also used to measure minimal residual disease in leukemias (Fig. 7), for stem cell enumeration in bone marrow transplant patients, monitoring HIV infection, quantifying feto-maternal hemorrhage, characterizing immunodeficiency diseases, analyzing paroxysmal nocturnal hemoglobinuria, in transplantation medicine, and for counting platelets and measuring platelet activation. Other routine clinical applications include the detection of auto-antibodies and the enumeration of reticulocytes in blood (Fig. 8).

Applications in the field of clinical research include the measurement of intracellular cytokines, the study of phagocyte biology and function (including phagocytosis and oxidative burst during activation), measurement of cell–cell interactions, the measurement of cell-mediated cytotoxicity and the study of viral infection.

5.2 Long Term Labels for Cells

There are some fluorescent labels that will survive in cells for days, if not weeks. If the cells divide, the label will be carried over into the daughter cells. One of the several compounds used is carboxyfluorescein diacetate, succinimidyl ester (CFDA SE). The compound will diffuse into cells where the diacetate groups are cleaved, generating a fluorescent compound; the succinimidyl groups react with thiols in the cytoplasm, retaining the dye in the cell.

Applications include studying cell fusion or clustering, tracking cells *in vivo*, identifying one partner in the interaction between two types of cell (for, example, a cytotoxicity reaction), and following cell division.

Genes for naturally fluorescent proteins (such as green fluorescent protein, GFP) have been introduced onto the DNA of cultured cells, thereby rendering the cells fluorescent. The GFP gene can be introduced into a plasmid for use as a reported molecule after transfection. Transfected cells can be purified by cell sorting on the flow cytometer.

5.3 Measurement of Cell Proliferation

The study of cell proliferation is an important and powerful application of flow cytometry. One can, for example, derive information about the proliferative status of tumors, observe the effects of cytotoxic drugs on the cell cycle or study the proliferation of subsets of lymphocytes undergoing immune stimulation.

There are several different approaches that can be used. One can measure the cell cycle phases, the rate of movement of cells through the cell cycle, the division of

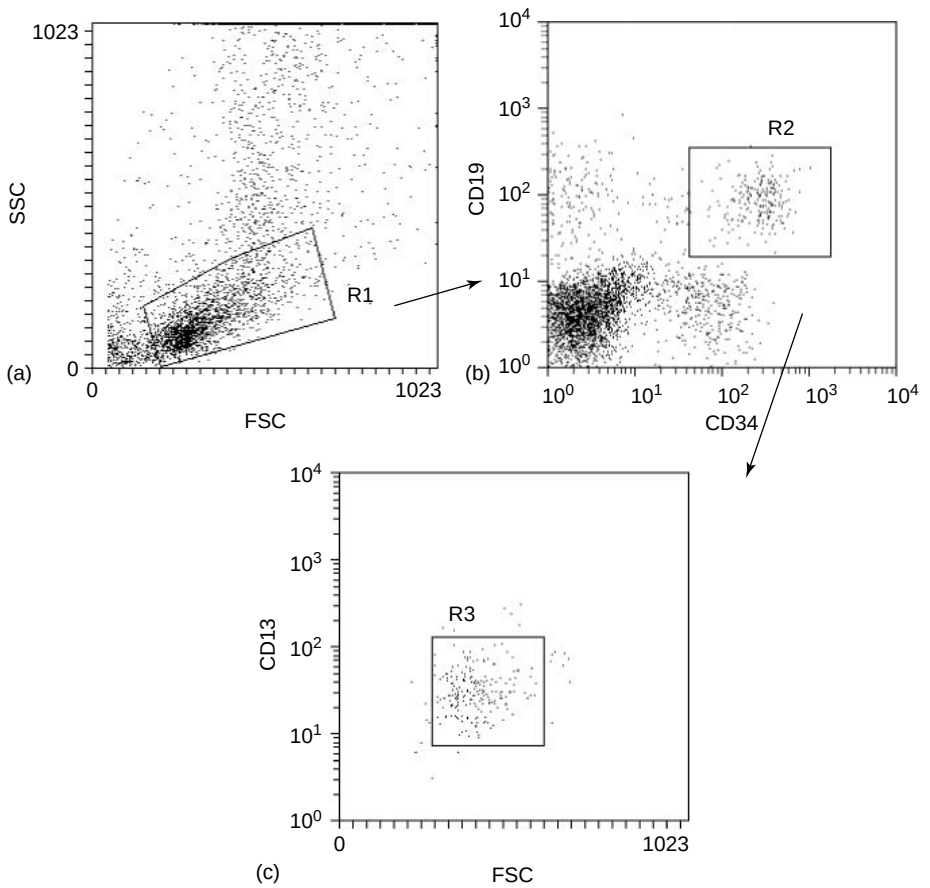

Fig. 7 Bone marrow sample obtained from a child with acute lymphoblastic leukemia undergoing treatment and considered to be in clinical and morphological remission. Mononucleated cells were separated on a density gradient and labeled with three antibodies associated with the phenotype of the leukemia - CD19/CD34/CD13. (a) Cytogram of light scatter with a region, R1, set on the lymphocytes and lymphoblastic cells. (b) Cytogram of CD19 versus CD34 expression, gated on R1. A region has been drawn around the positive cells. (c) Cytogram of CD13 versus forward scatter. Region, R3, has been drawn around the CD13 positive, tumor cells. There were 199 events in R3 out of a total of 10 000 events recorded. Data, supplied by Dario Campana, St. Jude's Hospital, Memphis, was recorded on a BD FACScan using an argon-ion laser at 488 nm.

label between daughter cells or observe the distribution of cell cycle–related proteins.

5.3.1 The DNA Histogram

A measurement of the DNA histogram should always form the basis of any study of changes in the proliferative state of cells. Some effects of chemotherapeutic drugs on the cell cycle are shown in Fig. 9.

5.3.2 Using BrdUrd/Anti-BrdUrd

If cells are pulse-labeled with 5-bromodeoxyuridine (BrdUrd), only cells in S-phase

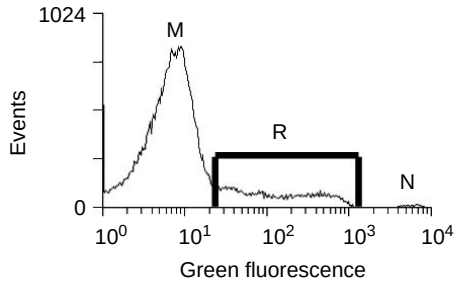

Fig. 8 Reticulocyte analysis. Histogram of green fluorescence from blood incubated with thiazole orange, which stains RNA and DNA. M, mature erythrocytes; R, reticulocytes; N, nucleated red cells. Data, supplied by Terry Hoy, University of Wales, Cardiff, was recorded on a BD FACScan using an argon-ion laser at 488 nm.

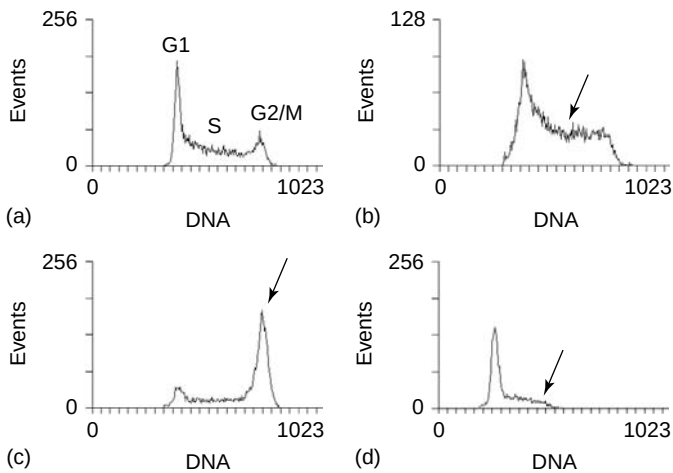

Fig. 9 DNA histograms showing the effects of drugs on the cell cycle. (a) L1210 murine leukemic cell line, no drug; the stages of the cell cycle are marked. (b) 5 h after a 2-h incubation with a Pt(IV) dicarboxylate at $3 \times IC_{50}$. There was an accumulation of cells in S-phase (arrowed) caused by a slowdown in the movement of cells through S-phase. (c) As (b) but recorded after 24 h. The drug has caused a block in G2 and cells have accumulated in that phase of the cell cycle (arrowed). (d) HL60 cells 4 h after the addition of 10 μM camptothecin. S-phase was completely blocked; cells originally in G2 have divided and, as there was no movement of cells from S to G2, there is an absence of cells in G2 (arrowed). Other details as in Fig. 5.

will contain BrdUrd. During subsequent incubation of the cells, the BrdUrd-labeled cells will progress around the cell cycle. The progression can be followed by fixing the cells at different time intervals and visualizing the BrdUrd with an anti-BrdUrd antibody; PI is added to show the phases of the cell cycle (Fig. 10). If cells are incubated with a drug, irradiated or a growth factor is withdrawn immediately after labeling, the effects of the treatment on progression around the cell cycle can be followed. Alternatively, pulse-labeling after treatment of the cells would reveal the state of the S-phase cells (see Fig. 10).

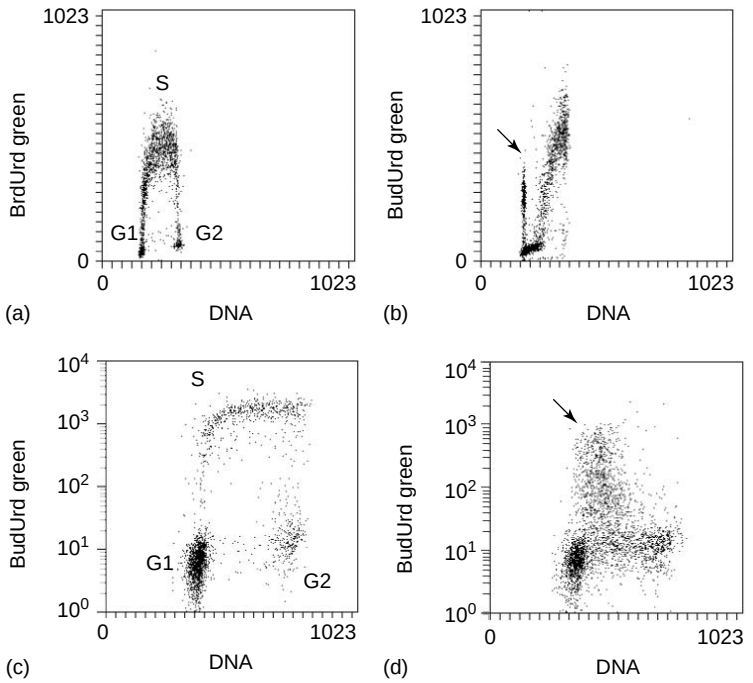

Fig. 10 Cytograms of green versus red (PI/DNA) fluorescence from cultured cells pulse-labeled with BrdUrd, fixed and labeled with anti-BrdUrd-FITC and PI. (a) & (b) Chinese hamster V79 cells, 0 (a) and 5 h (b) after pulse-labeling. At 5 h, the labeled cells had moved about halfway through the cell cycle. Cells in late S-phase at time 0 had divided and were in G1 (arrowed). Data supplied by George Wilson, Karmanos Cancer Institute, Detroit. (c) & (d) Neuroblastoma cell line 0 h after pulse-labeling. (c): no drug; (d): 24 h after a 2 h incubation with cisplatin. Note that in (d), there was a population of S-phase cells that were not synthesizing DNA. A further population were accumulated in early/mid S-phase and were synthesizing less DNA than the control (arrowed). All data recorded on BD FACScans using an argon-ion laser at 488 nm. Noted that a linear amplifier was used in (a) & (b) to record green fluorescence and a logarithmic amplifier for (c) & (d).

This method can be applied *in vivo*. BrdUrd has a half-life of about 20 min in an animal; a single incubation of BrdUrd is equivalent to a pulse label. Figure 11 shows nuclei from a rat mammary tumor. The rat was killed and the tumor removed 4 h after the injection of the BrdUrd. From the observed movement of the cells through S-phase during this time, the S-phase transit time can be calculated.

5.3.3 The BrdUrd-Hoechst/PI Method

The fluorescence of *bis*-benzimidazole dyes (Hoechst 33258 and Hoechst 33342) bound to DNA is quenched by BrdUrd. Consequently, continuous labeling with BrdUrd and subsequent staining of DNA with Hoechst 33258 separates cells according to the number of replications they underwent during the period of labeling. Addition of a DNA label unaffected by BrdUrd, such as PI, resolves the cell

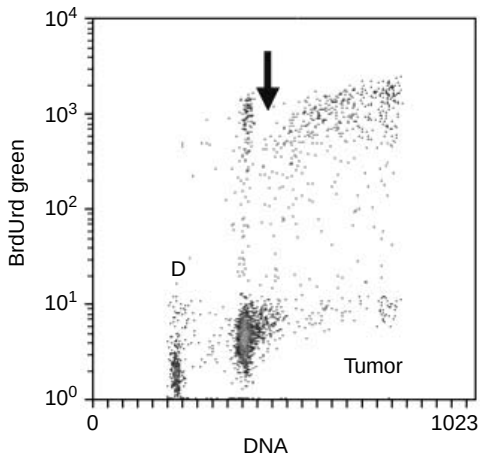

Fig. 11 Cytogram of green (anti-BrdUrd) fluorescence versus PI/DNA of a rat mammary tumor. The rat was injected with BrdUrd 4 h before the tumor was excised. The diploid cells are marked, D; the tumor was aneuploid. From the length of the "window" (arrowed) between G1 and the S-phase label, the time of S-phase can be calculated. Cells prepared by Tim Smith, Royal Marsden NHS Trust, Sutton, and the sample analyzed on a Coulter Elite by Jenny Titley with an argon-ion laser giving 200 mW at 488 nm.

cycle into the G0/G1, S and G2/M compartments. Hoechst 333258 is excited by UV, and the small number of machines equipped with a UV laser has limited the use of this technique. Now that two and three laser instruments with UV lasers are becoming more common, this method may find more widespread use.

5.3.4 Measuring Cell Proliferation through the Division of Label between Daughter Cells

Some labels will persist in cells for many days, if not weeks. Two compounds have been used for this application: carboxyfluorescein diacetate, succinimidyl ester (CFDA SE), and PKH26, an analog of acridine orange with an N-linked 26 carbon alkyl chain. CFDA SE diffuses into cells wherein it is converted to carboxyfluorescein, succinimidyl ester (CFSE); its reaction with amines stabilizes the fluorescent product in the cell. Because of its long alkyl side chain, PKH26 becomes anchored in the plasma membrane of cells.

To measure proliferation, the cells are labeled and their fluorescence recorded. After incubation under the appropriate experimental conditions, the fluorescence is rerecorded. The fluorescence from the cells that have divided once, twice, and so on, will have half, quarter the fluorescence of the undivided cells. Figure 12 shows data from murine T-cells in culture. After two days in the stimulated cultures, there are six peaks in the histogram of CFSE fluorescence; some of the cells had divided five times in 48 h, while other cells had not divided once.

5.4 Studying Cell Death

5.4.1 Necrosis

Assays for necrosis measure the integrity of the plasma membrane. There are several dyes that fluoresce on binding to DNA but are excluded from the cell by the plasma membrane, the most commonly used being PI. On addition of the dye, those cells with a damaged membrane show positive fluorescence.

An alternative approach is to load the cells with a fluorescent dye, for example, by incubation with carboxyfluorescein diacetate. When the plasma

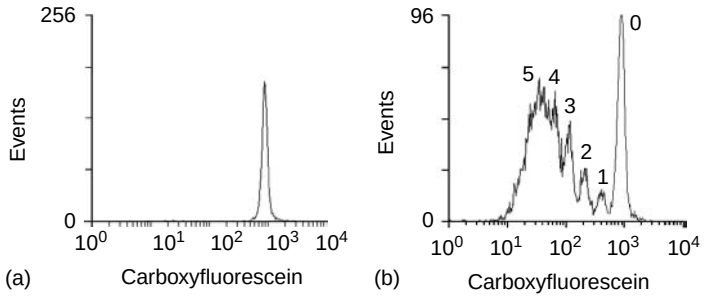

Fig. 12 Murine CD4+ cells in culture labeled with CFSE. The unstimulated cells (a) gave a single peak on the fluorescence histogram while the histograms from stimulated cells (b) exhibited multiple peaks due to dilution of the label during cell division. The numbers of cell divisions are marked on histogram (b). The cells were stimulated by incubating with antigen-presenting cells and an appropriate peptide. Data supplied by Maria Daly, GlaxoSmith Kline, and recorded on a Beckman Coulter XL with an argon-ion laser.

membrane is damaged, the dye is lost from the cell. The two assays can be combined.

5.4.2 Apoptosis

The application of flow cytometry in following the apoptotic cascade has been reviewed. The features that can be measured using flow cytometry include

- expression of proteins involved in the apoptotic cascade, including proteins generated during the cascade, such as activated caspases and cleaved poly (ADP-ribose) polymerase (PARP);
- activation of caspases;
- changes in the mitochondrial membrane potential;
- changes in other functional parameters, such as intracellular pH, calcium ions, glutathione, and oxidative species;
- changes in the plasma membrane;
- cell shrinkage;
- DNA degradation.

5.4.2.1 **Activation of Caspases** There are antibodies available, which are specific for the activated form of many caspases. Caspase activity may be observed directly by incubating cells with a compound made of a peptide, specific for one of the caspases, connected to a potentially fluorescent group. An active enzyme will release the fluorescent group. Typically, a compound such as DEVD-rhodamine 110 is used. An alternative form of the assay uses a fluorescently labeled enzyme inhibitor, which will be concentrated in those cells containing active enzyme.

5.4.2.2 **Mitochondrial Membrane Potential (MMP)** The collapse of the MMP is a central feature of the apoptotic cascade and flow cytometry is the main method used for monitoring the MMP. Several dyes, including rhodamine 123, 3,3′-dihexyloxacarbocyanine iodide and 5,5′,6,6′-tetrachloro-1,1′,3,3′-tetraethylbenzimidazolylcarbocyanine iodide (JC-1), are concentrated in mitochondria due to difference in the membrane potential of the

mitochondria in relation to the cytoplasm. When the mitochondrial membrane potential collapses, the overall fluorescence is reduced. In the case of JC-1, this is accompanied by a shift in the fluorescence from red to green.

5.4.2.3 **Changes in the Plasma Membrane** During apoptosis, the distribution and packing of the lipids in the plasma membrane changes. The permeability of the plasma membrane is also increased. Phosphatidyl serine (PS), which is normally exposed on the internal face of the plasma membrane, moves to the exterior of the cell during apoptosis. This change can be detected using the recombinant protein, annexin V, which binds to PS. Annexin V is labeled with one of the common fluorochromes, such as fluorescein or PE. Annexin V staining can be combined with one or more immunofluorescence stains to determine the phenotype of the apoptotic cells. Frequently, PI is added to distinguish cells in secondary necrosis.

There are several DNA binding dyes that are excluded by normal cells but are taken up by apoptotic cells; these include the *bis*-benzimidazole, Hoechst 33342, ethidium bromide, 7-aminoactinomycin D (7-AAD) and the dicyanine, YO-PRO-1. Figure 13 shows an example using 7-AAD.

5.4.2.4 **DNA Degradation** Cleavage of DNA at the linkers between the nucleosomes is a late event in apoptosis, resulting in DNA oligomers of about 200 bp. There are two methods by which this effect may be used to visualize apoptotic cells.

If apoptotic cells are fixed in ethanol and then resuspended in buffer, the small fragments of DNA are extracted from the cell. Consequently, the apoptotic cells will have a lower DNA content than normal cells. This manifests itself as a "sub-G1" peak in the DNA histogram (Fig. 14).

For the Tdt-mediated dUTP nick end-labeling (TUNEL) assay, the cells are fixed in paraformaldehyde followed by ethanol. All the fragments of DNA are retained within the cell and are available for labeling by the enzyme, terminal deoxynucleotidyl transferase (Tdt). A variety of labels may be incorporated into the DNA at the sites of the strand breaks, such as FITC-dUrd, biotin-dUrd, BrdUrd, by adding the triphosphate derivative to the incubation mix. Biotin-dUrd is visualized by the addition of streptavidin-FITC and BrdUrd by anti-BrdUrd-FITC. PI is finally added to

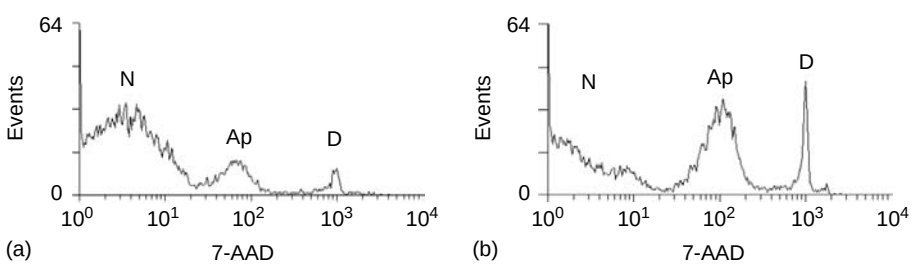

Fig. 13 Immature human thymocytes incubated with (a) dexamethasone for 0 h and (b) 24 h and then incubated with 7-AAD. The normal (N), apoptotic (Ap) and cells that had undergone secondary necrosis (D) are marked. Data, supplied by Ingrid Schmid, UCLA School of Medicine, was recorded on a BD FACScan using an argon-ion laser.

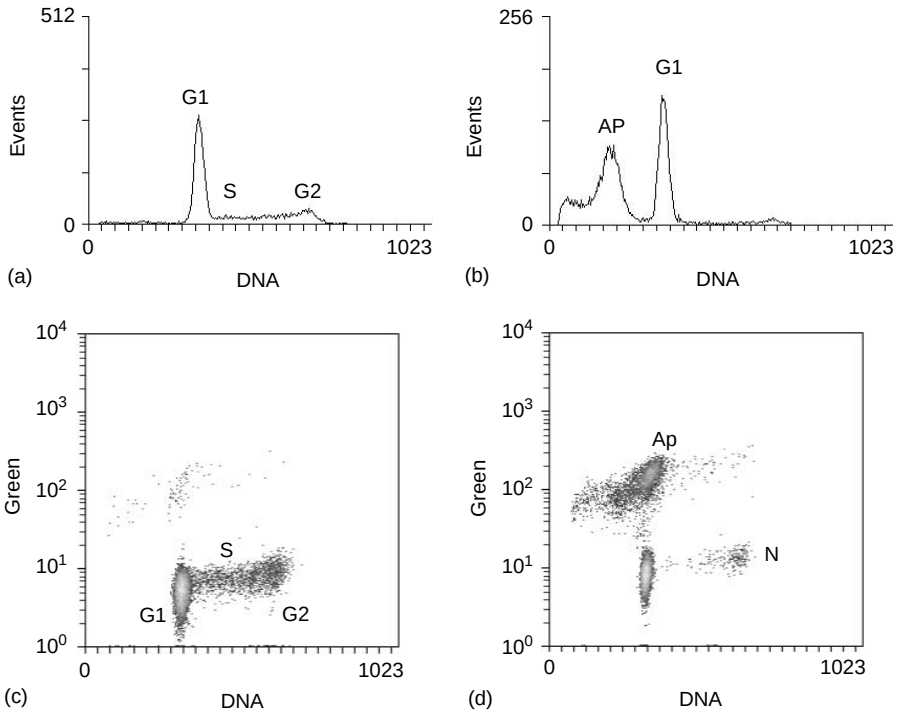

Fig. 14 Murine hemopoietic cell line that requires interleukin-3 (IL-3) for growth. (a) & (c): grown in the presence of IL-3; (b) & (d): after 16 h in the absence of IL-3. (a) & (b): DNA histograms. The sub-G1 peak from apoptotic cells is marked "Ap". (c) & (d): Cells labeled using the TUNEL assay. The cytograms of green fluorescence versus red (DNA/PI) fluorescence show the phases of the cell cycle and the negative, normal (N) and the positive, apoptotic cells (Ap). Simone Detre, Royal Marsden NHS Trust, London labeled cells for the TUNEL assay, Data recorded on a Beckman Coulter Elite using an argon-ion laser at 488 nm.

label the DNA. The position in the cell cycle from which apoptosis occurred is revealed as well as the percentage of apoptotic cells (Fig. 14).

5.5 Analyzing and Sorting Chromosomes

Sorting chromosomes requires a flow cytometer equipped with two large (5 W) argon-ion lasers and considerable expertise. A DNA stain of a suspension of chromosomes with a single dye (for example, PI) does not give enough resolution between the individual chromosomes. Two DNA stains are used – a *bis*-benzimidazole (Hoechst 33258) and chromomycin A3, each dye being excited by a separate laser. Hoechst 33258, excited by UV giving blue fluorescence, binds preferentially to AT-rich regions on the DNA; chromomycin A3, excited at 457 nm giving green fluorescence, binds preferentially to the GC-rich regions. The chromosomes can then be resolved on the basis of their total DNA content and their AT/GC ratio. All the individual chromosomes can be resolved with the exception of chromosomes 9–12, whose size and AT/GC ratios are not sufficiently separated.

5.6
Microbead Technology

Fluorescent microspheres (beads) offer a new emerging technology involving flow cytometry. A biomolecule can be attached to a bead and its interaction with other molecules studied; for example, the binding of two proteins; hybridization of DNA molecules; or the interaction of an enzyme with its substrate. Fluorescently labeled reporter molecules, such as antibodies, antigens, or nucleic acid probes can be used.

A particularly powerful application is the use of multiplexed arrays, for example, using two fluorescent dyes at eight different concentrations to give a set of 64 different beads. Each bead could carry a specific capture antibody; if the beads were fluorescent in the orange/red range, the detector antibodies might be labeled with fluorescein. Sixty-four assays could be carried out in the same tube simultaneously. The molecules measured might include hormones, cardiac markers, therapeutic drugs, drugs of abuse, and blood-borne viruses.

The same technology is being used for hybridization-based analyses for the detection of specific nucleic acid sequences.

Specific machines are being produced for these assays (for example, by Luminex Corporation).

6
Applications Other Than Mammalian Cell Biology

6.1
Microbiology

Generally, viruses are too small to be detected directly by flow cytometry but bacteria and other microorganizms, such as bacteria and yeasts, can be measured. Flow cytometric assays have been developed to determine features such as size, DNA content, surface receptors, membrane potential and intracellular pH.

One of the major applications has been in the study of antibiotic susceptibility. Flow cytometry is also being used to monitor water quality.

6.2
Marine Biology

The use of flow cytometry in monitoring the organisms in seawater is rapidly increasing. Many marine organizms are fluorescent and a combined measurement of light scatter and fluorescence at several wavelengths can yield a unique fingerprint for many organisms, such as the large variety of phytoplankton.

Instruments have been built to monitor seawater onboard a marine vessel and at fixed positions below sea level. Continuous monitoring has been achieved by mounting a small cytometer on a buoy and sending data to the mainland base by radio link (for example, Cytobuoy).

Bibliography

Books and Reviews

Darzynkiewicz, Z., Robinson, J.P., Crissman, H.A. (Eds.) (1999) Flow Cytometry, *Methods in Cell Biology*, 2nd edition, Vol. 41, Academic Press, San Diego.

Macey, M.G. (Ed.) (1994) *Flow Cytometry: Clinical Applications*, Blackwell Scientific, Oxford.

McCarthy, D.A., Macey, M.G. (Eds.) (2001) *Cytometric Analysis of Cell Phenotype and Function*, Cambridge University Press, Cambridge.

Ormerod, M.G. (1999) Flow Cytometry, *Royal Microscopical Society Handbook*, No. 44, Bios Scientific Publishers, Oxford.

Ormerod, M.G. (Ed.) (2000) *Flow Cytometry: A Practical Approach*, 3rd edition, IRL Press, Oxford University Press, Oxford.

Shapiro, H.M. (2003) Practical Flow Cytometry, 4th edition, Wiley-Liss, New York.

Primary Literature

Bedner, E., Smolewski, P., Amstad, P., Darzynkiewicz, Z. (2000) Activation of caspases measured in situ by binding of fluorochrome-labeled inhibitors of caspases (FLICA): correlation with DNA fragmentation, *Exp. Cell Res.* **259**, 308–313.

Begg, A.C., McNally, N.J., Shrieve, D.C., Karcher, H. (1985) A method to measure the duration of DNA synthesis and potential doubling time from a single sample, *Cytometry* **6**, 620–626.

Borowitz, M.J., Bray, R., Gascoyne, R., Melnick, S., Parker, J.W., Picker, L., Stetler-Stevenson, M. (1997) US-Canadian consensus recommendations on immunophenotypic analysis of hematologic neoplasia by flow cytometry: data analysis and interpretation, *Cytometry* **30**, 236–444.

Bright, G.R., Whitaker, J.E., Haugland, R.P., Taylor, D.L. (1989) Heterogeneity of the changes in cytoplasmic pH upon serum stimulation of quiescent fibroblasts, *J. Cell. Physiol.* **141**, 410–419.

Bromilow, I.M., Duguid, J.K. (1997) Measurement of feto-maternal haemorrhage: a comparative study of three Kleihauer techniques and tow flow cytometry methods, *Clin. Lab. Haematol.* **19**, 137–142.

Campana, D. (1997) Immunophenotypic analysis in the monitoring of minimal residual leukemia, *Rev. Clin. Exp. Hematol.* **1**, 42–56.

Carrano, A.V., Gray, J.W., Langlois, R.G., Burkhart-Schultz, K.J., Van Dilla, M.A. (1979) Measurement and purification of human chromosomes by flow cytometry and sorting, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1382–1384.

Davey, H.M. (2002) Flow cytometric techniques for the detection of microorganisms, *Methods Cell Sci.* **24**, 91–97.

Davis, B.H., Bigelow, N.C. (1994) Automated reticulocyte analysis. Clinical practice and associated new parameters, *Hematol. Oncol. Clin. North Am.* **8**, 617–630.

Darzynkiewicz, Z., Bruno, S., Del Bino, G., Gorczyca, W., Hotz, M.A., Lassota, P., Traganos, F. (1992) Features of apoptotic cells measured by flow cytometry, *Cytometry* **13**, 795–808.

Darzynkiewicz, Z., Gong, J., Juan, G., Ardelt, B., Traganos, F. (1996) Cytometry of cyclin proteins, *Cytometry* **25**, 1–13.

Darzynkiewicz, Z., Juan, G., Li, X., Gorczyca, W., Murakami, T., Traganos, F. (1997) Cytometry in cell necrobiology: analysis of apoptosis and accidental cell death (necrosis), *Cytometry* **27**, 1–20.

De Rosa, S.C., Brenchley, J.M., Roederer, M. (2003) Beyond six colours: a new era in flow cytometry, *Nat. Med.* **9**, 112–117.

van Eeden, S.F., Klut, M.E., Walker, B.A., Hogg, J.C. (1999) The use of flow cytometry to measure neutrophil function, *J. Immunol. Methods* **232**, 23–43.

van Engeland, M., Nieland, L.J.W., Ramaekers, F.C.S., Schutte, B., Reutlingsperger, C.P.M. (1998) Annexin V-affinity assay: a review on an apoptosis detection system based on phosphatidyl serine exposure, *Cytometry* **31**, 1–9.

Givan, A.L. (2001) Flow Cytometry, *First Principles*, 2nd edition, Wiley-Liss, New York.

Hampton, M.B., Winterbourn, C.C. (1999) Methods for quantifying phagocytosis and bacterial killing by human neutrophils, *J. Immunol. Methods* **232**, 15–22.

Harrison, P. (2000) Progress in the assessment of platelet function, *Br. J. Haematol.* **111**, 733–744.

Hedley, D.W., Chow, S. (1994) Evaluation of methods for measuring cellular glutathione content using flow cytometry, *Cytometry* **15**, 349–358.

Hug, H., Los, M., Hirt, W., Debatin, K.-M. (1999) Rhodamine 110-linked amino acids and peptides as substrates to measure caspase activity upon apoptosis induction in intact cells, *Biochemistry* **38**, 13906–13911.

Jacobberger, J.W. (2000) Flow cytometric analysis of intracellular protein epitopes, in: Stewart, C.C., Nicholson, J.K.A. (Eds.) *Immunophenotyping, Cytometric Cellular Analysis*, John Wiley & Sons, New York, pp. 361–406.

Jennings, C.D., Foon, K.A. (1997) Recent advances in flow cytometry: applications to the diagnosis of hematologic malignancy, *Blood* **90**, 2863–2892.

Jepras, R.I., Carter, J.H., Pearson, S.C., Paul, F.E., Wilkinson, M.J. (1995) Development of a robust flow cytometric assay for determining

viable numbers of bacteria, *Appl. Environ. Microbiol.* **61**, 2696–2701.

June, C.H., Rabinovitch, P.S. (1988) Flow cytometric measurement of cellular ionized calcium concentration, *Pathol. Immunopathol. Res.* **7**, 409–432.

Kallioniemi, O.P., Visakorpi, T., Holli, K., Isola, J.J., Rabinovitch, P.S. (1994) Automated peak detection and cell cycle analysis of flow cytometric DNA histograms, *Cytometry* **16**, 250–255.

Kellar, K.L., Iannone, M.A. (2002) Multiplexed microsphere-based flow cytometric assays, *Exp. Hematol.* **30**, 1227–1237.

Keren, D.F., McCoy, J.P., Carey, J.L. (Eds.) (2001) *Flow Cytometry in Clinical Diagnosis*, 3rd edition, American Society for Clinical Pathology Press, Chicago.

King, M.A. (2000) Detection of dead cells and measurement of cell killing by flow cytometry, *J. Immunol. Methods* **243**, 155–166.

Legendre, L., Courties, C., Troussellier, M. (2001) Flow cytometry in oceanography 1989–1999: environmental challenges and research trends, *Cytometry* **44**, 164–172.

Li, N., Goodall, A.H., Hjemdahl, P. (1999) Efficient flow cytometric assay for platelet-leucocyte aggregates in whole blood using fluorescence signal triggering, *Cytometry* **35**, 145–161.

Lyons, A.B. (2000) Analysing cell division in vivo and in vitro using flow cytometric measurement of CFSE dye dilution, *J. Immunol. Methods* **243**, 147–154.

McSharry, J.J. (2000) Analysis of virus-infected cells by flow cytometry, *Methods* **21**, 249–257.

Maino, V.C., Picker, L.J. (1998) Identification of functional subsets by flow cytometry: intracellular detection of cytokine expression, *Cytometry* **34**, 207–215.

Marchetti, P., Castedo, M., Susin, S.A., Zamzami, N., Hirsch, T., Macho, A., Haeffner, A., Hirsch, F., Geuskens, M., Kroemer, G. (1996) Mitochondrial permeability membrane transition is a central co – ordinating event in apoptosis, *J. Exp. Med.* **184**, 1155–1160.

Misteli, T., Spector, D.L. (1997) Applications of the green fluorescent protein in cell biology and biotechnology, *Nat. Biotechnol.* **15**, 961–964.

Nebe-von-Caron, G., Stephens, P.J., Hewitt, C.J., Powell, J.R., Badley, R.A. (2000) Analysis of bacterial function by multi-colour fluorescence flow cytometry and single cell sorting, *J. Microbiol. Methods* **42**, 97–114.

Nicholson, J.K.A. (1994) Immunophenotyping specimens from HIV-infected persons: laboratory guidelines from the Center for Disease Control and Prevention, *Cytometry* **18**, 55–59.

Nolan, J.P., Mandy, F.F. (2001) Suspension array technology: new tools for gene and protein analysis, *Cell Mol. Biol.* **47**, 1241–1256.

Ormerod, M.G. (1998) The study of apoptotic cells by flow cytometry, *Leukemia* **12**, 1013–1025.

Ormerod, M.G. (2001) Using flow cytometry to follow the apoptotic cascade, *Redox Rep.* **6**, 275–287.

Ormerod, M.G., Kubbies, M. (1992) Cell cycle analysis of asynchronous populations by flow cytometry using bromodeoxyuridine label and Hoechst-propidium iodide stain, *Cytometry* **13**, 678–685.

Petit, P.X., LeCoeur, H., Zorn, E., Dauguet, C., Mignotte, B., Gougeon, M.L. (1995) Alterations of mitochondrial structure and function are early events of dexamethasone-induced thymocyte apoptosis, *J. Cell Biol.* **130**, 157–165.

Prussin, K. (1997) Cytokine flow cytometry: understanding cytokine biology at the single-cell level, *J. Clin. Immunol.* **17**, 195–204.

Richards, S.J., Rawstron, A.C., Hillmen, P. (2000) Application of flow cytometry to the diagnosis of paroxysmal nocturnal hemoglobinuria, *Cytometry* **42**, 223–233.

Roederer, M. (2001) Spectral compensation for flow cytometry: visualization artifacts, limitations and caveats, *Cytometry* **45**, 194–205.

Rosenfeld, C.S., Bodensteiner, D.C. (1986) Detection of platelet alloantibodies by flow cytometry. Characterization and clinical significance, *Am. J. Clin. Pathol.* **85**, 207–212.

Shanahan, T. (1997) Application of flow cytometry in transplantation medicine, *Immunol. Invest.* **26**, 91–101.

Sintnicolaas, K., de Vries W., van der Linden, R., Gratama, J.W., Bolhuis, R.L. (1991) Simultaneous flow cytometric detection of antibodies against platelets, granulocytes and lymphocytes, *J. Immunol. Methods* **142**, 215–222.

Steen, H.B. (2000) Flow cytometry of bacteria: glimpses from the past with a view to the future, *J. Microbiol. Methods* **42**, 65–74.

Sutherland, D.R., Anderson, L., Keeney, M., Nayar, R., Chin-Yee, I. (1996) The ISHAGE guidelines for CD34+ cell determination by flow cytometry, *J. Hematother.* **3**, 213–226.

Szöllösi, J., Damjanovich, S., Matyus, L. (1998) Applications of fluorescence resonance energy transfer in the clinical laboratory: routine and research, *Cytometry* **34**, 159–179.

de Vries, E., Noordzij, J.G., Kuijpers, T.W., van Dongen, J.J. (2001) Flow cytometric immunophenotyping in the diagnosis and follow-up of immunodeficient children, *Eur. J. Pediatr.* **160**, 583–591.

Zamzami, N., Marchetti, P., Castedo, M., Decaudin, D., Macho, A., Hirsch, T., Susin, S.A., Petit, P.X., Mignotte, B., Kroemer, G. (1995) Sequential reduction of mitochondrial membrane potential and generation of reactive oxygen species in early programmed cell death, *J. Exp. Med.* **182**, 367–377.

Websites

Some Companies Who Supply Flow Cytometers

BD www.bd.com
Beckman Coulter www.beckmancoulter.com
Dakocytomation www.dakocytomation.com
Partec www.partec.de
Guava Technologies www.guavatechnologies.com
Luminex www.luminexcorp.com
Cytobuoy www.cytobuoy.com

Free Software

WinMDI http://facs.scripps.edu/software.html
Cylchred www.uwcm.ac.uk/uwcm/hg/hoy/software.html

Other Useful Sites

Molecular Probes www.probes.com
Information on flow cytometry and microbiology www.cyto.purdue.edu/flowcyt/research/micrflow/index.htm

Fluorescence Correlation Spectroscopy (FCS)

Aladdin Pramanik[1] *and Jerker Widengren*[2]*
[1] *Department of Biochemistry and Biophysics, Karolinska Institute, 106 91 Stockholm, Sweden*
[2] *Department of Physics, Albanova University Center, Royal Institute of Technology, 171 77 Stockholm, Sweden*

1 Introduction 463

2 Experimental Equipment – Aspects of the Standard Equipment 464

3 Theory of FCS 465
3.1 Fluorescence Autocorrelation Function for Translational Diffusion 465
3.2 Reaction Dynamics and Translational Diffusion 469
3.3 Reactions Leading to a Change in the Diffusion Properties 469
3.4 Reactions Leading to a Change in the Fluorescence Brightness 470
3.5 Cross-correlation Modalities of FCS in the Spectral and Spatial Mode 472
3.6 Image Correlation Spectroscopy – Correlations in the Spatial Domain 475
3.7 Rotational Diffusion 476
3.8 Higher Order FCS and Fluorescence Intensity Distribution Analysis (FIDA) 477

4 FCS Applications in Life Sciences, Biomedicine, and Drug Discovery 478
4.1 Protein–Protein Interactions 478
4.2 Ligand–Receptor (on Single Live Cell) and Intracellular Interactions 481
4.3 Vesicle–Peptide Interactions 486
4.4 Protein–DNA Interactions 489
4.5 DNA–DNA Interaction (DNA Sequencing) 489
4.6 Monitoring of Molecular Processes in the Fast Time Range ($<\tau_D$) (Chemical Reactions, Photophysics, Conformational Changes) 490
4.7 FCS and Related Techniques in Drug Screening 492

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

5 Additional Remarks and Conclusions 494

Bibliography 496
Books and Reviews 496
Primary Literature 496

Keywords

Fluorescence Correlation Spectroscopy (FCS)
Monitoring of thermodynamic fluctuations of molecules via ultrasensitive detection of their emitted fluorescence.

Brownian Motion
The spatial consequence of thermodynamic fluctuations of particles in a macroscopic system, regarded as the origin of diffusion.

Single Molecule Detection
The ultrasensitivity of an instrument, allowing a detection of a single molecule.

Molecular Interactions
Interactions between compounds such as protein–protein, protein–DNA, ligand–receptor, and DNA–DNA, here studied by FCS.

Drug Screening
Search for new substances with therapeutic activity with the aim to develop them into new drugs.

In Fluorescence Correlation Spectroscopy (FCS), fluctuations of the fluorescence intensity of fluorophore-labeled molecules, excited by a sharply focused laser beam, are observed. Owing to high sensitivity, detection at the level of single molecules is possible. Thermodynamic fluctuations can thus be revealed, which would be difficult to detect at a macroscopic level. The technique can, in principle, offer information about any molecular dynamic process in the nanosecond time range and longer, manifesting itself as a change in fluorescence intensity without the need for any perturbation or synchronization of the system studied. The specific features of FCS make it a versatile tool for biomolecular studies, including determination of translational and rotational diffusion, concentration and density of molecules, chemical kinetics, and binding reactions. In this review, an overview of the theory and the typical experimental setup for FCS measurements will be given, along with some examples of how FCS can be used for molecular dynamic studies.

1 Introduction

In macroscopic systems at thermodynamic equilibrium, the concentrations of molecules of different species, and other parameters of the system, do not by definition, change with time. Various processes may take place, but because of the normally large number of molecules present in the system, they tend to cancel each other out. Therefore, the relative changes caused by the molecular processes are of a negligible magnitude on the macroscopic scale. However, if the number of molecules in the system is small, the processes do not average out to the same extent, and relative fluctuations in concentration, or any other parameter, can become significant. Since the fluctuations are governed by the molecular processes that take place in the system, analysis of signal fluctuations can provide a powerful way of investigating these processes. The fluctuations of a parameter, measured in a system in thermodynamic equilibrium, can be described by the same mathematical formalism that would be used in a relaxation experiment. However, unlike relaxation methods, where a macroscopic perturbation is applied in the form of temperature, electric field, or pressure jump, no external disturbance of the system is required for the fluctuation measurement.

The concept of number fluctuation analysis, in which the fluctuating number of particles within a fixed volume is analyzed in terms of its time dependence, has been introduced a long time ago. An early experimental application was that of Svedberg, in which the diffusion coefficients of colloidal particles were measured on the basis of theoretical analysis of Smoluchowski. In his analysis, Smoluchowski used the concentration autocorrelation function:

$$G_C(\tau) \equiv \langle C(t)C(t+\tau)\rangle = \langle \delta C(t)\delta C(t+\tau)\rangle + \langle C\rangle^2, \quad (1)$$

where the brackets denote the time average. The time-dependent part of this function expresses the typical time course of a fluctuation and its magnitude.

Different techniques have been developed that provide indications of number fluctuations: the dynamic light scattering technique, also known under the names quasi-elastic light scattering or photon correlation spectroscopy, exploits the light scattering intensity from the particles of interest. The electrical conductance of a solution can also be used as an indicator since it reflects the number and charges of the particles contained in the solution. Fluctuation analysis of electrical currents over sections of cellular or artificial membranes, known as Voltage Clamp, has been extensively used in the field of neurophysiology (with the later developed patch clamp technique as its single-molecule counterpart).

The utilization of fluorescence intensity as the fluctuating quantity was originally introduced by Magde, Elson, and Webb under the name Fluorescence Correlation Spectroscopy (FCS), and in a series of papers, they presented the theory and the first experimental realization of the technique. In the first FCS experiments, the average number of fluorescent molecules in the observation volume was about 10 000. However, thanks to the remarkable averaging power of the autocorrelation function, it was still possible to extract the motion of individual molecules from the large uncorrelated bulk fluorescence. In this respect, FCS is perhaps the oldest discipline of single-molecule fluorescence spectroscopy.

Although FCS showed great potential in a number of fields in the 1970s, it had not found a more widespread use until the last five to ten years. The reasons for the limited success of FCS at the initial stage were mainly related to high background light levels and to low detection quantum yields.

The situation was dramatically changed by the introduction of extremely small observation volumes in FCS measurements. This, in combination with confocal epi-illumination, highly sensitive avalanche photodiodes for fluorescence detection, and very selective bandpass filters to discriminate the fluorescence from the background, made it possible to improve signal-to-background ratios in FCS measurements by several orders of magnitude. In a dynamic system, the relative fluctuations increase as the concentration of molecules decreases. Therefore, in order to obtain high enough relative fluctuations in fluorescence, the number of molecules residing in the detection volume at the same time should not be too high. With the improvements mentioned above, the mean number of molecules could be kept low, without compromising the signal level, and measurement times could be shortened drastically, enabling a more widespread use of the FCS technique. A sensitivity that permits the detection of single molecules was reached. Although FCS, in a strict sense, cannot be categorized as a single molecule spectroscopy technique, the capacity of the technique strongly benefits from this high sensitivity.

2 Experimental Equipment – Aspects of the Standard Equipment

In order to realize the FCS experiment, a small illuminated volume must be created from which the fluorescence light

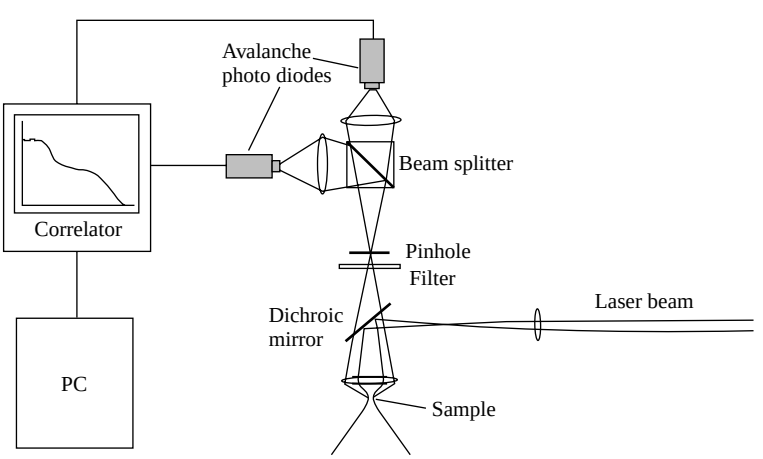

Fig. 1 Typical instrumentation for FCS experiments. A laser beam is directed via a dichroic mirror into a confocal microscope and focused by the objective into the sample. Emitted fluorescence from the sample is collected by the same objective, passes through the beam splitter, and is focused onto the image plane where the pinhole is located. The light passing the pinhole is spectrally filtered by highly selective bandpass emission filters and is focused onto the sensitive areas of the avalanche photodiodes. Standard TTL pulses from the diodes are fed to the PC-based correlator.

is collected and recorded. A typical FCS spectrometer is based on a confocal fluorescence microscope (Fig. 1). In contrast to confocal scanning, the confocal pinhole in FCS does not serve to enhance the spatial resolution. Its purpose is to reject the out-of-focus emission, and therefore it is larger than the point spread function of the objective. A typical range of diameters of the pinhole used in FCS is 30 to 100 μm. A laser is used as the light source for excitation. The main reason for using a laser is the possibility of focusing its well-collimated beam into a very small spot of about 0.5 μm in diameter. Additionally, the small spectral width of the laser light facilitates the spectral separation of the fluorescence emission from the excitation. The filter in front of the detector is usually a bandpass interference filter. At shorter wavelengths, it blocks the scattered laser light, and at longer wavelengths, it blocks the strongest band of Raman scattering from water. A typical sample is a dilute water-based solution of fluorescently labeled molecules. However, there is a relatively large flexibility in the choice of measurement environment. What is important is that the molecules are able to move in and out of the illuminated volume. This exchange of molecules is necessary for two purposes: (1) exchanging the photobleached molecules with the fresh ones or (2) measuring the kinetic parameters of the motion.

There are processes that can be studied by FCS on a spatially fixed set of molecules, or on just one fixed molecule, for example, intersystem crossing to the triplet state, photoisomerization, chemical reactions, or photon antibunching. However, at room temperatures, the fluorescent molecules photobleach after emitting typically 10^5 photons on the average. This is too few for a precise correlation analysis; therefore, the FCS experiments on fixed molecules have been performed almost only at cryogenic temperatures.

The experimental measurement of the fluorescence intensity $I(t)$ is accomplished by counting the number of photons, n, during time intervals of a predetermined width ΔT.

$$I(t) = I(i\Delta T) = \frac{n_i}{\Delta T}. \tag{2}$$

The time-dependent experimental correlation function is calculated as

$$G_{\exp}(\tau) = G_{\exp}(k\Delta T) = \frac{\dfrac{1}{M-k}\displaystyle\sum_{i=1}^{M-k} n_i n_{i+k}}{\left[\dfrac{1}{M-k}\displaystyle\sum_{i=1}^{M-k} n_i\right]^2}, \tag{3}$$

where M is the total number of counting intervals, so that the total duration of the measurement is $M\Delta T$.

3 Theory of FCS

3.1 Fluorescence Autocorrelation Function for Translational Diffusion

In FCS measurements, the macroscopic stochastic process of interest is the fluorescence intensity $I(t)$ detected from the observation volume. $I(t)$ depends on another stochastic process, namely, the concentration of fluorescent molecules, $C(\bar{r}, t)$, within the observation volume. Here, a brief derivation of the autocorrelation function for $I(t)$ will be given, restricted to the case in which translational diffusion of a single fluorescent solute into and out of the observation volume is the only process generating fluorescence fluctuations. This

case forms the basis for all kinds of dynamic systems that can be investigated with FCS.

In the majority of recent applications of FCS, a confocal microscope arrangement is used, where the observation volume is not physically restricted, such that molecules can freely diffuse into and out of it. It is defined by the dimensions of the laser beam focus, which excites the fluorescent molecules, and the collection-efficiency function of the microscope:

$$I(t) = Q\int_{-\infty}^{\infty} CEF(\bar{r})I_{\text{exc}}(\bar{r})C(\bar{r},t)\,\mathrm{d}V = Q\int_{-\infty}^{\infty} p(\bar{r})C(\bar{r},t)\,\mathrm{d}V \quad (4)$$

Here, $C(\bar{r}, t)$ is the concentration of fluorescent molecules at position $\bar{r}$ and time t. $Q = q\sigma_{\text{exc}}\Phi_f$, where σ_{exc} is the excitation cross-section of the fluorescent molecules under study, Φ_f is their fluorescence quantum yield, and q signifies the efficiency of detection of fluorescence, which is emitted from the center of the laser focus. The latter includes the solid angle of light collection, the transmission of the microscope optics and the spectral filters, as well as the detection quantum yield of the detector. $CEF(\bar{r})$ is the collection-efficiency function, describing the dependence of light losses on the exact coordinates $\bar{r}$ of the light source in the sample. $CEF(\bar{r})$ is mainly defined by the cooperation of the objective and the confocal pinhole. $I_{\text{exc}}(\bar{r})$ is the excitation intensity. The fluorescence emitted from a fluorophore molecule is normally assumed to be proportional to the excitation intensity. The product of $CEF(\bar{r})$ and $I_{\text{exc}}(\bar{r})$, which defines the spatial dependence of brightness at which a molecule is seen by the detector, denoted by $p(\bar{r})$, is approximated to be Gaussian distributed in the axial as well as in the radial directions:

$$p(\bar{r}) = I_0 \mathrm{e}^{-2(x^2+y^2)/\omega_1^2}\mathrm{e}^{-2z^2/\omega_2^2} \quad (5)$$

where I_0 is the peak excitation intensity in the center of the focused laser beam, and ω_1 and ω_2 denote the distance from this center in radial and axial dimensions respectively, at which the detected fluorescence has dropped by a factor, e^2 (see Fig. 2(a)). This simplification makes it possible to express the fluorescence autocorrelation function in a much simpler form and has been shown to be a reasonable approximation for most FCS instrumentations based on a confocal microscope arrangement. On the basis of this approximation, it is convenient to define a brightness profile, where $p(\bar{r})$ is normalized by the peak intensity I_0:

$$p_{\text{norm}}(\bar{r}) = \frac{p(\bar{r})}{I_0} = \mathrm{e}^{-2(x^2+y^2)/\omega_1^2}\mathrm{e}^{-2z^2/\omega_2^2} \quad (6)$$

The detected fluorescence intensity $I(t)$ is, for simplicity, assumed to be an ergodic and at least a wide-sense stationary stochastic process. Loosely speaking, this means that the time average of a process for, namely, one molecule over a longer time equals the ensemble average of the process for many molecules studied at a certain moment in time. Further, the correlation between two points in time for the process gets continuously smaller with increasing differences in time between the points. Since we can never get all possible realizations (outcomes) of $I(t)$ at our disposal, it would, without this assumption, be impossible to measure the true correlation function for $I(t)$ at some given conditions. However, with these assumptions, we can get a good estimation of the statistical properties of $I(t)$, such as the ensemble average and the autocorrelation function, if we measure

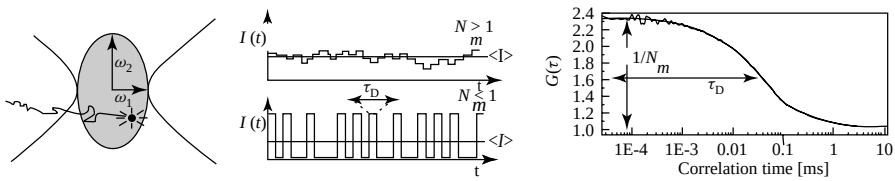

(a) (b) (c)

Fig. 2 (a) Schematic view of the observation volume (gray shaded area), given by the dimensions of the focused excitation laser beam (thick black line) and the collection-efficiency function, $CEF(\bar{r})$, of the confocal fluorescence microscope. Typically, the distribution of the detected fluorescence intensity can be approximated as a three-dimensional Gaussian distribution, as given by $p(\bar{r})$ in Eq. (5), where ω_1 and ω_2 are the distances from the center in the radial and axial directions at which $p(\bar{r})$ has dropped by a factor of e^2 compared to its peak value, $p(0)$. (b) The detected fluorescence intensity, $I(t)$, reflects the number of fluorescent molecules in the detection volume, which in FCS measurements, can be normally regarded as a Poisson process. For a Poisson process, the variance corresponds to the mean itself. For molecules freely diffusing into and out of the observation volume, the magnitude of the relative fluctuations (normalized with their mean, $\langle I \rangle$) increases with a lower mean number of molecules within the observation volume, N. (c) A typical FCS curve for a solute undergoing free diffusion in three dimensions into and out of an observation volume as that of Fig. 2(a). In FCS, the relative fluctuations of the detected fluorescence are quantified by the extrapolated autocovariance of $I(t)$ at $\tau = 0$, normalized by the mean intensity squared, $\langle I \rangle^2$. This corresponds to the amplitude of the time-dependent part of the normalized fluorescence autocorrelation function, given by Eq. (13), and equals $1/N$. The decay time, τ_D, of this function is related to the mean dwell time of the sample molecules in the observation volume, and can for translational diffusion be approximated as $\tau_D = 4D/\omega_1^2$, where D is the diffusion coefficient of the studied molecules.

one realization of $I(t)$ for a long enough time. Experimentally, the autocorrelation function of $I(t)$ is thus obtained from

$$R_{II}(\tau) = \frac{1}{T}\int_0^T I(t)I(t+\tau)\,\mathrm{d}t = \langle I(t)I(t+\tau)\rangle, \qquad (7)$$

where T is the measurement time. After normalization with the mean intensity squared and by assuming a wide-sense stationarity, the autocorrelation function of $I(t)$ takes the form

$$G(\tau) = \frac{R_{II}(\tau)}{\langle I\rangle^2} = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I\rangle^2} = \frac{\langle \delta I(0)\delta I(\tau)\rangle}{\langle I\rangle^2} + 1 \qquad (8)$$

Here, $\langle I \rangle$ is the mean collected fluorescence intensity and $\delta I(t) = I(t) - \langle I \rangle$ denotes the fluctuation. The first term in the two last expressions corresponds to the autocovariance function of $I(t)$ divided by the mean intensity squared.

In order to compute this correlation function, the behavior of concentrations must be put into Eq. (8), using Eq. (4):

$$G(\tau) = \frac{Q^2 \iint_{VV'} p(\bar{r})p(\bar{r}')g(\bar{r},\bar{r}',\tau)\,\mathrm{d}V\,\mathrm{d}V'}{\left(Q\overline{C}\int_V p(\bar{r})\,\mathrm{d}V\right)^2} + 1 = \frac{\iint_{VV'} p_{\mathrm{norm}}(\bar{r})p_{\mathrm{norm}}(\bar{r}')g(\bar{r},\bar{r}',\tau)\,\mathrm{d}V\,\mathrm{d}V'}{\left(\overline{C}\int_V p_{\mathrm{norm}}(\bar{r})\,\mathrm{d}V\right)^2} + 1, \qquad (9)$$

where $g(\bar{r}, \bar{r}', \tau) = \langle \delta C(\bar{r}, t) \delta C(\bar{r}', t + \tau) \rangle$ is the concentration autocovariance function and $\overline{C}$ is the equilibrium mean concentration. The concentration fluctuations can be expressed as $\delta C(\bar{r}, t) = C(\bar{r}, t) - \overline{C}$. The propagation of $\delta C(\bar{r}, t)$ in space and time is described by the diffusion equation:

$$\frac{\mathrm{d}\delta C(\bar{r}, t)}{\mathrm{d}t} = D\nabla^2 \delta C(\bar{r}, t) \qquad (10)$$

with the boundary condition $\delta C(\bar{r}, t) = 0$ at $|\bar{r}| = \infty$. Here, $\nabla^2 = (\partial^2/\partial^2 x, \partial^2/\partial^2 y, \partial^2/\partial^2 z)$ and D is the diffusion coefficient of the fluorescent molecules. In solution, the diffusion coefficient is given by the Stokes–Einstein equation: $D = kT/6\pi\eta R$, where k is the Boltzmann's constant, T is the absolute temperature, η is the viscosity, and R is the hydrodynamic radius of the diffusing molecule.

Normally, FCS measurements are performed in dilute solutions that can be considered ideal. This means that the fluorescent molecules do not interact with each other. At any instant of time, a fluctuation in the concentration at a position $\bar{r}$ is independent of the fluctuation at any other position $\bar{r}'$. In addition, the number of molecules within any subvolume is governed by Poisson statistics. The Poisson distribution describes the number of times an event has occurred as a function of time, where the events occur at random times, independent of each other, and with a constant average intensity (i.e. number of events per time unit). For a Poisson process, the probability that exactly N events have taken place within a time interval $t = \{0, t_1\}$ is given by

$$P(t_1, N) = \frac{\mathrm{e}^{-\lambda t_1}}{N!}(\lambda t_1)^N \qquad (11),$$

where λ is the intensity. The average value and the variance equal each other, and are given by

$$\sigma^2 = \langle N \rangle = \lambda t_1 \qquad (12)$$

Given these assumptions, Eq. (9) can finally be written as

$$G(\tau) = \frac{1}{N}\left(1 + \frac{4D\tau}{\omega_1^2}\right)^{-1}\left(1 + \frac{4D\tau}{\omega_2^2}\right)^{-1/2} + 1 = \frac{1}{N}G_{\mathrm{D}}(\tau) + 1. \qquad (13)$$

N is usually called the number of molecules in the observation volume. For simplicity, and for later reference, the normalized functional form of the correlation function for translational diffusion has been denoted as $G_{\mathrm{D}}(\tau)$. Since the spatial brightness function $p(\bar{r})$ has no abrupt limits, the meaning of the "observation volume" and its size deserves some attention. Owing to the gradual fall in brightness, there is no evident border allowing one to say whether a molecule is in or out of the observation volume. The contribution of a molecule to the fluorescence intensity simply decreases with its distance from $r = (0, 0, 0)$. A natural way of selecting the volume would be to put its boundary through points with some constant value of $p(\bar{r})$. While there are no well-defined limits and no defined shape of the observation volume, its absolute size is exactly defined. It corresponds to the volume containing N molecules, where N is defined in Eq. (13). As a Poissonian process, the mean of the number of molecules in the observation volume corresponds to its variance (Eq. 12). The value of N can be calculated to be

$$N = \frac{1}{G(0) - 1} = \overline{C}\frac{\left(\int_V p_{\mathrm{norm}}(\bar{r})\,\mathrm{d}V\right)^2}{\int_V p_{\mathrm{norm}}^2(\bar{r})\,\mathrm{d}V} \qquad (14)$$

Consequently, as we can see from Eq. (14), the measured number of molecules is determined exclusively by the concentration, the dimensions, and the shape of the normalized brightness profile $p_{\mathrm{norm}}(\bar{r})$ of the observation volume, and does not depend on the excitation intensity I_{exc} and the quantum yield (specific brightness) of the molecules, Q. This is an important feature of FCS.

For a sample that contains m different species with different fluorescence quantum yields and with different diffusion properties, Eq. (13) can be generalized to

$$G(\tau) = \frac{\sum_{i=1}^{m}\left[N_i Q_i^2 \left(1+\frac{\omega_1^2\tau}{4D_{\mathrm{i}}}\right)^{-1} \times \left(1+\frac{\omega_2^2\tau}{4D_i}\right)^{-1/2}\right]}{\left[\sum_{i=1}^{m}(N_i Q_i)\right]^2} + 1 = \frac{\sum_{i=1}^{m} N_i Q_i^2 G_{\mathrm{D}i}(\tau)}{\left[\sum_{i=1}^{m}(N_i Q_i)\right]^2} + 1, \qquad (15)$$

where $Q_i = (q\Phi_{\mathrm{f}}\sigma_{\mathrm{exc}})_i$ is the fluorescence brightness of the ith species, q is the detection efficiency, Φ_{f} is the fluorescence quantum yield, and σ_{exc} denotes the extinction coefficient. $G_{\mathrm{D}i}(\tau)$ is defined in Eq. (13). Note that the different fluorescent species contribute to the amplitude of $G(\tau)$ in proportion to the square of their relative fluorescence brightness, and it can be strongly misleading to interpret the "apparent" number of molecules N_{m} as the true mean number of molecules in the observation volume

$$N_{\mathrm{m}} = \frac{\left[\sum_{i=1}^{m} N_i Q_i\right]^2}{\sum_{i=1}^{m} N_i Q_i^2} \qquad (16)$$

3.2 Reaction Dynamics and Translational Diffusion

In general, there are two criteria, of which at least one needs to be fulfilled in order to gain information about a reaction process from fluctuation analysis of the detected fluorescence intensity and FCS:

1. The chemical reaction must lead to a change in the diffusion properties of the fluorescent species under investigation.
2. The chemical reaction must lead to a change in the fluorescence quantum yield or the excitation cross-section of the reactants.

In this way, the typical duration of the fluorescence bursts because of the passage of fluorescent molecules through the excitation volume or the magnitude of the fluorescence emission per molecule, reflected by the fluorescence brightness parameter, Q (Eq. 4), is changed.

3.3 Reactions Leading to a Change in the Diffusion Properties

In case the first criterion above is fulfilled, and if in addition, the relaxation time of the chemical reaction is slower than the average dwell times of the molecules in the observation volume, the reactant and product molecules can be considered as

separate types of noninteracting species on the time scale of the dwell times. The autocorrelation function of the fluorescence fluctuations can then be expressed in the form given in Eq. (15). In the most simple case, with a fluorescently marked ligand with fluorescence brightness Q_{free} and diffusion coefficient D_{free}, which binds to a receptor leading to a ligand–receptor complex with Q_{bound} and D_{bound}, (Eq. 15) can be written as

$$G_{\text{D}}(\tau) = \frac{N_{\text{free}} Q_{\text{free}}^2 \left(1 + \frac{\omega_1^2 \tau}{4 D_{\text{free}}}\right)^{-1} \left(1 + \frac{\omega_2^2 \tau}{4 D_{\text{free}}}\right)^{-1/2} + N_{\text{bound}} Q_{\text{bound}}^2 \left(1 + \frac{\omega_1^2 \tau}{4 D_{\text{bound}}}\right)^{-1} \left(1 + \frac{\omega_2^2 \tau}{4 D_{\text{bound}}}\right)^{-1/2}}{[N_{\text{free}} Q_{\text{free}} + N_{\text{bound}} Q_{\text{bound}}]^2} \quad (17)$$

This form of the autocorrelation function is applicable to a broad range of ligand–receptor interactions, either in solution or on/inside cells, monitored via FCS (Fig. 3).

3.4 Reactions Leading to a Change in the Fluorescence Brightness

In the FCS measurements, transient nonfluorescent states, or states with a different fluorescence brightness, present themselves as fluorescence fluctuations superimposed on those caused by concentration changes due to translational motion of the fluorophores in and out of the sample volume element (see Fig. 4). In the presence of a chemical reaction and translational diffusion, the concentration fluctuations of molecules of the ith species at position $\bar{r}$ and time t are determined by the following set of differential equations:

$$\frac{\mathrm{d}\Delta C_j(\bar{r}, t)}{\mathrm{d}t} = D_j \nabla^2 \Delta C_j(\bar{r}, t) - \sum_{k=1}^{M} T_{jk} \Delta C_j(\bar{r}, t), \quad j = 1 \text{ to } M \quad (18)$$

where M is the number of species participating in the chemical reaction, T_{jk} represents the matrix of the kinetic rate coefficients, and D_j is the diffusion coefficient of the jth component. In its general form, the normalized autocorrelation function of the detected fluorescence fluctuations will show a complex dependence on the eigenvalues and eigenvectors of a matrix derived from (Eq. 18), including the reaction rates as well as the coefficients of the translational diffusion, and cannot be expressed in an analytical form. Fortunately, simplifications are often possible.

Generally, it has been shown that if either of the two following conditions is fulfilled

1. diffusion is much slower than the chemical relaxation time(s) ($\tau_{\text{D}i} \gg \tau_{\text{chem}}$) and/or
2. the diffusion coefficients of all fluorescent species are equal then fluorescence correlation function can be separated into two factors; the first, $G_{\text{D}}(\tau)$, which depends on transport properties (diffusion or flow) and the second, $X(\tau)$,

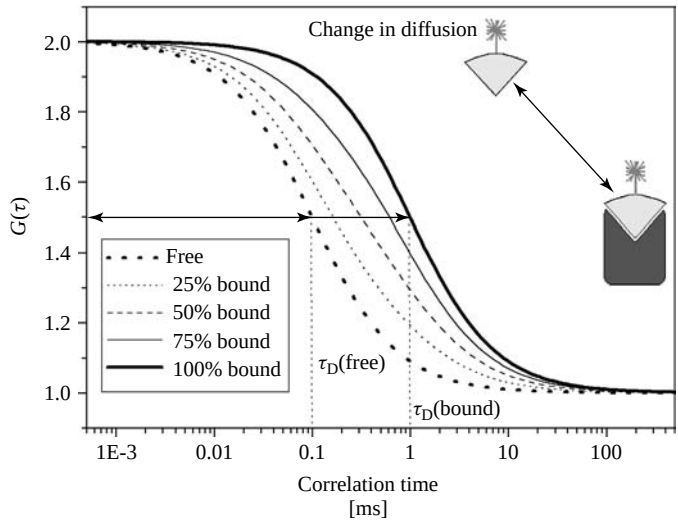

Fig. 3 Simulated FCS curves for a binding reaction involving a smaller, fluorescently marked molecule (ligand) binding to a larger molecule (receptor). In the simulation, the time range of the reaction is much longer than the passage times of the molecules through the observation volume. With a diffusion time of the ligand of 0.1 ms and that of the ligand–receptor complex of 1 ms, and with the fluorescence brightness of the molecules not being changed by the binding reaction, the fraction of bound ligands can be derived from the FCS curve as the fraction of molecules with a diffusion time of 1 ms (Eq. 17), with $Q_{\text{free}} = Q_{\text{bound}}$. The amplitude of $G(\tau)$ is not affected by the extent of binding.

which depends on the reaction rate constants:

$$G(\tau) = G_{\text{D}}(\tau)X(\tau) + 1 \qquad (19)$$

where

$$X(\tau) = \frac{\sum_{i,j=1}^{M} Q_i Q_j S_{ij}(\tau)}{\sum_{i=1}^{M} Q_i^2 \overline{C}_i} \qquad (20)$$

Here, Q_i denotes the fluorescence capacity of compound i and is defined in Eqs. 4 and 15. $S_{ij}(\tau)$ is the solution to the following set of differential equations and initial conditions:

$$\frac{\mathrm{d}S_{ij}(\tau)}{\mathrm{d}\tau} = \sum_{j=1}^{M} T_{ij} S_{jk}(\tau)$$

$$S_{ik}(0) = \overline{C}_i \Delta_{ik}, \qquad (21)$$

where T_{ij} corresponds to the reaction rate matrix of a reaction involving M different species, as given by Eq. (18). From Eqs. (20) and (21), one can see that $X(\tau)$ is a normalized sum of the terms $S_{ij}(\tau)$. Given an initial mean concentration, $\overline{C}_i$, of the compound i at $\tau = 0$, the terms $S_{ij}(\tau)$ state the average concentration of the state j at time τ, given that it was in state i at time 0.

It is, when possible, very convenient to be able to treat the kinetics of the

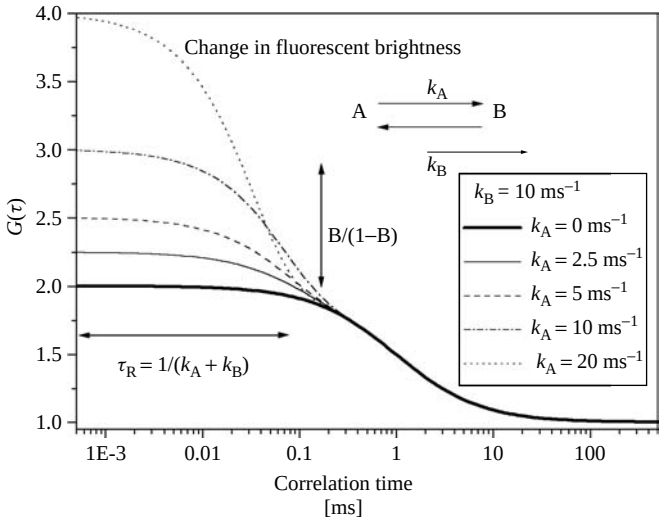

Fig. 4 Simulated FCS curves for a reaction, in which the fluorescence brightness of the participating molecules is changed. Here, a unimolecular reaction is considered, in which the compound B is nonfluorescent. The FCS curves can then be described by Eq. (19). The passage times of the molecules through the detection volume is for both A and B set to 1 ms, and the average total number of molecules in the observation volume $N = A + B$ is set to 1, such that A and B corresponds to the fractions of molecules in states A and B, respectively. The rate constant k_B has been given the value 10 ms^{-1}, and the values of k_A for the different curves are given in the inset. The relaxation time of the reaction, $\tau_{chem} = 1/(k_A + k_B)$, is much shorter than the passage times, such that the chemical reaction shows up as an additional relaxation process in the FCS curves. From Eqs. (19–21), it follows that the relaxation time of this process corresponds to τ_{chem}, and its amplitude is given by $B/(N(1 - B))$. The amplitude of the relaxation process due to diffusion corresponds to the total number of molecules participating in the reaction, $1/(A + B) = 1/N$, whereas the total amplitude represents $1/(AN)$, that is, the inverse number of fluorescently active molecules.

chemical reaction separately from the translational diffusion in the fluctuation analysis. As mentioned, a rather broad range of chemical reactions fulfills the above two criteria. In addition, for a reaction that does not fulfill the above criteria under standard conditions, it is sometimes possible to expand the observation volume, or to speed up the reactions under study, for instance, by using higher concentrations of unlabeled reactants, in order to make the relaxation time of the reaction faster than the dwell time (first criterion above).

3.5 Cross-correlation Modalities of FCS in the Spectral and Spatial Mode

By the use of two detector units, one will get two separate photocurrents. Following (Eq. 8), the normalized autocorrelation

function of the fluorescence intensities will then take the form

$$G(\tau) = \frac{\langle \delta I_1(0) \delta I_2(\tau) \rangle}{\langle I_1 \rangle \langle I_2 \rangle} + 1 \qquad (22)$$

where $\delta I_1(t)$ and $\delta I_2(t)$ signify the fluctuation terms in the detected fluorescence intensities in the first and the second detector, and $\langle I1 \rangle$ and $\langle I2 \rangle$ denote their corresponding mean values. One technical advantage of using two detectors in this way is that by cross-correlation, the shot noise inherent to photon detection can be eliminated. Of importance for measurements in the sub-microsecond time range is also the fact that the dead-time of the detectors can be circumvented. This makes it possible to follow fast processes by FCS down to the nanosecond time range. Another possibility is to let the two detectors cover different emission wavelength ranges, with the fluorescence intensity of the two emission ranges cross-correlated, the so-called dual-color FCS. In dual-color FCS, the observation volume is illuminated by two overlapping foci of two lasers with different wavelengths, in which one of the lasers selectively excites one of the two dye labels, and the other laser excites the other dye. However, under certain conditions, by the use of two-photon excitation or by the use of dyes with large differences in Stokes shifts, it is also possible to use only one laser as an excitation source for both dyes.

In a typical dual-color FCS experiment in which a bimolecular reaction is studied, two reaction partners are labeled with different dyes, for each of which the spectral range of the detectors is optimized. Upon binding, a new fluorescent species is formed, which is labeled with both of the dyes. In this way, the concentration and diffusion characteristics of two fluorescent species in solution as well as their reaction product can be followed in parallel (Fig. 5). For a bimolecular reaction, $A + B \leftrightarrow AB$, which is much slower than the translational diffusion of the molecules through the observation volume, three different experimental correlation functions are obtained. The normalized, time-dependent part of these functions can, in analogy to Eq. (16), ideally be described as

$$G_A(\tau) = \frac{1}{V} \times \frac{(K_A)^2 \overline{C}_A G_{D_A}(\tau) + (K_{Ab})^2 \overline{C}_{AB} G_{D_{AB}}(\tau)}{(K_A \overline{C}_A + K_{Ab} \overline{C}_{AB})^2} \qquad (23a)$$

$$G_B(\tau) = \frac{1}{V} \times \frac{(K_B)^2 \overline{C}_B G_{D_B}(\tau) + (K_{aB})^2 \overline{C}_{AB} G_{D_{AB}}(\tau)}{(K_B \overline{C}_B + K_{aB} \overline{C}_{AB})^2} \qquad (23b)$$

$$G_{AB}(\tau) = \frac{1}{V} \times \frac{(K_{Ab})(K_{aB}) \overline{C}_{AB} G_{D_{AB}}(\tau)}{(K_A \overline{C}_A + K_{Ab} \overline{C}_{AB})(K_B \overline{C}_B + K_{aB} \overline{C}_{AB})} \qquad (23c)$$

Here, V is the size of the detection volume, $\overline{C}_i$ denotes the mean concentration. $K_i = q_i \sigma_{abs(i)} \Phi_{f(i)} I_{exc(i)}$ is similar in meaning to Q_i and denotes the detected fluorescence per molecule, such that K_A is the fluorescence of species A registered at detector A, K_B is that of species B at detector B, K_{aB} is that of species AB at detector B (q_{aB}), and finally K_{Ab} is that of species AB at detector A. Likewise, q_i accounts for the detection quantum efficiency of the detectors and the transmittance of the fluorescence of species A to detector A (q_A), species B to detector B (q_B), species AB to detector B (q_{aB}), and species AB to detector A (q_{Ab}). $\sigma_{exc(i)}$ denotes the corresponding excitation cross-sections, I_{excA} and I_{excB} are the effective excitation intensities at

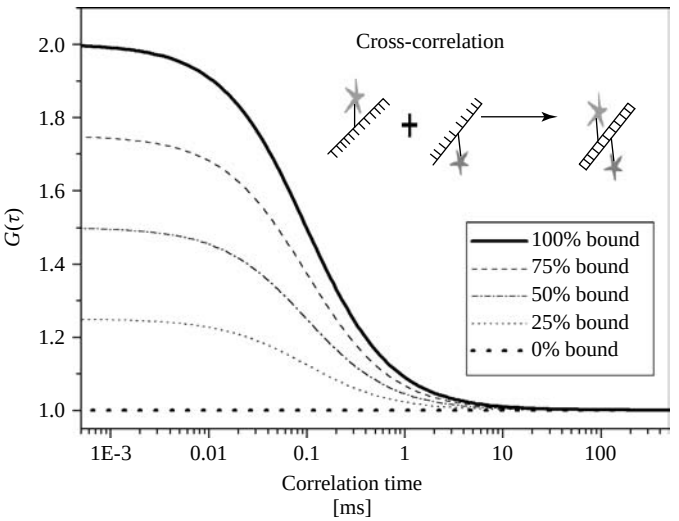

Fig. 5 Simulated experimental data of a dual-color cross-correlation experiment, where the association of two single-stranded, differently labeled, complementary DNA oligonucleotides is followed. The simulation is based on Eq. (23c), where the diffusion time of all species (the two single-stranded oligonucleotides, A and B, and the double-stranded product AB) is assumed to be 0.1 ms. The concentrations of the molecules are normalized so that $[\overline{C}_A + \overline{C}_{AB}]V = [\overline{C}_B + \overline{C}_{AB}]V = 1$, the fluorescence brightnesses are equal, $K_A = K_B = K_{Ab} = K_{aB}$, and the cross-talk in the fluorescence detection is neglected. As can be seen, the obtained information lies in the amplitude of the correlation curve.

the wavelengths where A and B are excited, and $\Phi_{f(i)}$ denotes the fluorescence quantum yields.

Depending on the choice of dyes and detection band pass filters, detection-unit cross talk, which generates additional terms to Eqs. (23a–c), has to be taken more or less into account. The cross-talk is related to the degree of overlap of the emission spectra of the dyes with that of the detection filters of the nondedicated channel.

The dual-color FCS approach can increase the selectivity and offers the possibility to follow the association/dissociation of a reaction product even if the reaction itself does not induce a change in either the diffusion properties or in the fluorescence capacity of the former reactants. The disadvantage is mainly the low signal-to-noise ratio that follows from the typically small amplitude of the cross-correlation term as well as from the fact that the fluorescence from all species will contribute to the denominator of Eq. (23c). Additionally, a less than perfect overlap of the volumes that the two detectors A and B observe can further reduce the amplitude of $G_{AB}(0)$. An imperfect overlap can be caused by the different extents of diffraction effects and aberrations in the optics at the two wavelengths.

The cross-correlation concept can also be applied in the spatial dimension. The fluorescence is then detected from more than one observation volume, which

makes it possible to measure velocities as well as directions of molecular transport processes.

3.6 Image Correlation Spectroscopy – Correlations in the Spatial Domain

In Scanning-FCS (S-FCS) experiments, a flow velocity is imposed by translating the sample at a uniform, known velocity through a laser beam. Theoretically, the FCS autocorrelation functions for uniform flow and for uniform sample translation are equivalent. For a uniform flow having a vectorial speed $\overline{V} = (v_x, 0, 0)$, directed perpendicular to the illumination of a Gaussian laser beam, where the diffusion is very slow, such that

$$\tau_1 = \frac{\omega_1^2}{4D} \gg \tau_f = \frac{\omega_1}{v} \tag{24}$$

the contribution due to translational diffusion can be neglected, and the normalized fluorescence autocorrelation function for one fluorescent species simplifies to

$$G(\tau) = \frac{1}{N} G_D(0) \times e^{-\left[\left(\frac{v_x \tau}{\omega_1}\right)^2\right]} + 1 \tag{25}$$

Image Correlation Spectroscopy (ICS) is an extension of S-FCS, which is based on quantitative analysis of confocal laser scanning microscopy images. As the scanning velocity is known, a fluorescence autocorrelation function in the spatial domain, which is equivalent to that in the time domain of Eq. (25), can be formed

$$G(\xi) = \frac{\langle \delta I(x) \delta I(x+\xi) \rangle}{\langle I \rangle^2} + 1 \tag{26}$$

where only the scanning velocity, v_x, as a scaling factor differs ξ from τ

$$\xi = v_x \tau \tag{27}$$

For a two-dimensional array of pixels, it is possible to extend (Eq. 26) to a two-dimensional spatial autocorrelation function with two independent variables, x and y:

$$G(\xi, \eta) = \frac{\langle \delta I(x, y) \delta I(x+\xi, y+\eta) \rangle}{\langle I \rangle^2} + 1, \tag{28}$$

which for an $N \times M$ pixelated image can be calculated and normalized as

$$G(\xi, \eta) = \frac{(1/NM) \sum_{k=1}^{N} \sum_{\ell=1}^{M} I(k, \ell) I(k+\xi, \ell+\eta)}{\left((1/NM) \sum_{k=1}^{N} \sum_{\ell=1}^{M} I(k, \ell) \right)^2} \tag{29}$$

where $M \gg \xi$ and $N \gg \eta$. From a computational point of view, it is more convenient to calculate the two-dimensional spatial autocorrelation function from the inverse Fourier transform of the power spectrum of the image data. The power spectrum is calculated as the product of the Fourier transform of $I(x, y)$ and its complex conjugate

$$G(\xi, \eta) = F^{-1}\{F[I(x, y)] \times F^*[I(x, y)]\} \tag{30}$$

In analogy to the expression for the autocorrelation function for S-FCS (Eq. 25), the autocorrelation function for ICS will take the form

$$G(\xi, \eta) = G(0) \times e^{-(\xi^2+\eta^2)/\omega_1^2} \tag{31}$$

In both S-FCS and ICS measurements, the temporal/spatial dependence of the fluorescence autocorrelation function gives information only about the radius of the excitation laser beam and the scanning velocity, and does not normally contain

information about the sample molecules themselves, unless the molecules or aggregates are larger than ω_1. Instead, the relevant information is to be found in the amplitude $G(0)$, from which the density of molecules or the density of aggregates and the average number of monomers per aggregate can be determined.

In S-FCS and ICS, the amplitudes of $G(0)$ of Eqs. (25 and 31) can be expressed in a similar way as in Eq. (15). However, for cases in which there is a distribution of aggregates, an additional contribution to the amplitude can be expected from the variation of the aggregate size. Assuming that the fluorescence brightness, Q, of each monomer contained in an aggregate is unaffected by the number of monomers, $\mu_1(1)$, in the aggregates, one can show that

$$G(0) = \frac{1}{N_{\text{mono}}^2} \sum_i N_i(\mu_1(i)^2 + \sigma_i^2) \quad (32)$$

where N_{mono} is the mean number of monomers within the observation volume, $\mu_1(i)$ and σ_i are the mean and the variance of the number of monomers within the ith population of clusters investigated, and N_i is the mean number of such aggregates.

Since confocal microscopic images can be generated within less than a second, ICS measurements are faster than those of S-FCS. The image is also more easily collected over a large cell area so that more sampling, which leads to a higher accuracy, is done. On the other hand, the sampling is performed with a lower resolution, leading to a lower precision than that for S-FCS. ICS has been applied to yield a quantitative measure of the degree of aggregation of fluorescently labeled molecules, primarily cell membrane proteins and lipids.

3.7 Rotational Diffusion

The processes of absorption and emission of light by fluorescent molecules are anisotropic. Therefore, in general, fluctuations in detected fluorescence also occur because of rotational movement of molecules.

These fluctuations can occur even if the excitation beam is unpolarized, or if fluorescence is detected from only one direction. If the absorption dipole moment of the molecule is turned parallel to the direction of propagation of the excitation beam, then there will be no absorption. Similarly, if the emission dipole moment points to the detector, the light will be emitted in all other directions, except for the axis connecting the molecule and the detector. The magnitude of these orientational fluctuations can be enhanced by the use of polarized excitation and detection.

Despite the general presence of orientational fluctuations in fluorescence, the experimental evidence is scarce. There are several reasons for this. First, if the rotation is very fast, like that of organic dye molecules in water, then the fluctuations are masked by a quantum effect called *photon antibunching*. Second, if the slowly rotating macromolecule carries many fluorescent labels, the dipole moments of different fluorophores can be oriented at different angles, resulting in a significantly reduced dependence of absorption and emission on the orientation of the macromolecule. Finally, when labeled to a macromolecule, a dye can typically undergo wobbling motion due to a nonrigid attachment of the dye to the molecule of interest, such that the rotational motion rather reflects the wobbling of the dye than the rotational motion of the molecule to which it is labeled.

On the other hand, FCS provides an interesting feature in the fact that the rotation measurement is decoupled from the excited state lifetime τ_e, since the orientation is probed by two consecutive excitation cycles of the molecule. This allows, in principle, the measurement of very slow rotational diffusion, which is inaccessible to macroscopic fluorescence methods.

3.8 Higher Order FCS and Fluorescence Intensity Distribution Analysis (FIDA)

Although it is an extremely sensitive and fast method, FCS does not utilize all the information that is present in the fluorescence fluctuations; for example, if we analyze a mixture of two or more fluorescent compounds, then the components are separated on the basis of different diffusion times. The contributions of the components to the amplitude of the correlation function are proportional not only to concentrations but also to the square of the brightness (i.e. quantum yield in the elementary theory) of the corresponding components (Eq. 16). Therefore, the determination of actual concentrations of components in the mixture requires additional information. It is also possible that some of the components have equal diffusion coefficients, and differ only by the brightness of fluorescence.

Alternative approaches to gain complementary information about formation and number distributions of aggregates in a solution or on a surface exist. Here, we will briefly describe two approaches: (1) the use of higher order autocorrelation functions and (2) how generation of photon counting histograms, or fluorescence intensity distribution analysis (FIDA) can be used to gain complementary information for FCS analysis.

Higher order autocorrelation functions for the fluctuations in the detected fluorescence can be defined as

$$G_{m,n}(\tau) = \frac{\langle(\delta I(t+\tau))^m(\delta I(t))^n\rangle - \langle(\delta I(t))^m\rangle\langle(\delta I(t))^n\rangle}{\langle I(t)\rangle^{m+n}} \tag{33}$$

where m and n are integers. For n, $m = 1$ (Eq. 33) reduces to the standard second-order autocorrelation function (Eq. 8). (The average fluctuations, $\langle\delta I(t)\rangle$, contained in the second term of the numerator equals zero.) For a second-order autocorrelation function, the amplitude is given by Eq. (14). However, for the case in which the sample contains several species, or one or several distributions of aggregates, more than one measured parameter, as that of $G(0)$, is needed to characterize the concentrations of the different species/aggregates in the sample (Eqs. 15 and 32). One way of obtaining more information from a temporal or spatial record of fluorescence fluctuations is to make use of higher order autocorrelation functions, as defined in Eq. (33).

Fluorescence Intensity Distribution Analysis (FIDA), or Photon Counting Histograming (PCH) extracts information about the actual concentrations and brightnesses of the molecules present in the fluorescence fluctuations. In FIDA, the intensity signal needs to be prepared and analyzed in a different way compared to FCS, but the experimental arrangement is nearly identical. In the processing of the recorded photon counts, a histogram of the photon count numbers is calculated, instead of computing the correlation function $G(\tau)$, as done in FCS. The fluorescence photons are counted in time intervals of, for example, $\Delta T = 40\,\mu s$ during the experiment with duration

$M \cdot \Delta T$. Dividing by the total number of intervals M transforms the histogram into a photocount number probability distribution $P(n)$.

The parameters extracted by FIDA are concentrations c_i and molecular brightnesses Q_j of the species i. No information about diffusion is obtained by FIDA. If the components should have equal brightness values Q, then FIDA could not resolve them. In this respect, FIDA and FCS are complementary methods, which extract different but unique data from the same signal.

It is important to note that although the FIDA theory ignores the diffusion, it is still vital for the method, acting as the source of molecule number fluctuations.

From the aspect of utilizing the fluctuation signal, it seems natural to combine the two complementary methods – FCS and FIDA. One example of such a combination has been realized in FIMDA (Fluorescence Intensity Multiple Distributions Analysis), which is an extension of FIDA. Here, several distributions are collected simultaneously, effectively using several counters in parallel, with different counting time intervals ΔT_i. In the resulting set of distributions, the same photocounts are used, but they are grouped using different time windows.

4 FCS Applications in Life Sciences, Biomedicine, and Drug Discovery

In this section, several examples will be given in order to illustrate some of the approaches of the FCS technique (Sect. 3). It is beyond the scope of this review to list all possible applications; rather, we give some examples in order to illustrate the potential of this technique. Fluorescence Correlation Spectroscopy is a powerful tool for examining molecular interactions of biological importance with very high specificity. The ability to study specific interactions by FCS and related techniques show considerable promise as these techniques could be used as tools in genomics and proteomics, and with respect to drug discovery/screening. In this chapter, we will exemplify how different categories of specific interactions can be monitored by FCS techniques in Sects. 4.1 to 4.5.

4.1 Protein–Protein Interactions

Proteins play a crucial role in virtually all biological processes. The first step in the action of a protein is its binding of another molecule. Proteins as a class of micromolecules are unique in being able to recognize and interact with highly diverse molecules. Proteins can also interact with themselves or with other proteins thereby performing a wide diversity of important functions. The spectrum of functions a protein can perform ranges from mechanic to protective functions. The blood protein fibrinogen can form insoluble fibers resulting in blood clotting. Important protective proteins such as antibodies or immune globulins bind with and thus neutralize foreign proteins and render them inactive, enabling the immune system to distinguish between self and nonself. Further, the hormone proteins bind to receptor proteins and thus carry messages throughout the body. Finally, molecular self-assembly can serve as a good example of protein–protein interaction that can play a fundamental role in biology. The assembly of identical molecules such as peptide and protein into aggregates in bulk at surfaces or

interacting with macromolecules is, to an increasing extent, believed to be a cause of diseases such as Alzheimer's disease, prion diseases, and others.

As an example, and to give a general idea of how FCS can be used to study protein–protein interactions, and in particular, aggregation processes of peptides and proteins, the aggregation of the Alzheimer amyloid β-peptide (Aβ) will be discussed. The aggregation process has been studied *in vitro* with different techniques: turbidimetry, light scattering, analytical ultra-centrifugation, circular dichroism spectroscopy, electron microscopy, atomic force microscopy, polyacrylamide gel electrophoresis, size exclusion chromatography, and quantitative fluorimetry. However, only the spectroscopic techniques allow the measurement of the aggregation process directly in solution. So far, quasi-elastic light scattering has provided the most detailed description of this process. Recently, the aggregation of Aβ and the protein–protein interaction in plant cells have been studied using FCS. In order to follow the polymerization process, it is necessary to label the peptide/protein with at least one fluorophore with an account on steric effects of the fluorophore molecules. Further, it is extremely important to pay attention to the buffer composition and its pH value used for the aggregation experiment. Data on the Aβ-aggregation process have been evaluated using the correlation function as given in Eqs. (15–17) (see Sect. 3.3). The aggregation of Aβ dissolved in Tris-buffered saline, pH 7.4, occurs above a critical concentration of 50 μM and proceeds from monomers/dimers into two discrete populations of large aggregates, without any detectable amounts of oligomers (Fig. 6). The polymerization of Aβ is a highly cooperative process in which the formation of very large aggregates precedes the formation of fibrils. The aggregation was reduced when Aβ was incubated in the presence of Aβ ligands, oligopeptides previously shown to inhibit fibril formation. The aggregation process occurs in two phases. In the early phase (1–120 min), diffuse networks and fibrils are formed. Previous studies, performed with other techniques, have provided limited information on the early (1–120 min) phase in the Aβ aggregation due to problems in measuring monomers, oligomers, and large aggregates simultaneously. The FCS technique is not limited by these problems, and measurements are made directly in solution without any further sample preparation thereby minimizing the risk for artifacts. In conclusion, FCS can be used as a highly sensitive and specific competition assay to identify potential inhibitors for the Aβ–aggregation process, and the appearance of aggregates in the cerebrospinal fluid of Alzheimer's patients using FCS is suggested as a diagnostic tool. FCS offers a powerful tool to study the aggregation process. It allows the polymerization process to be followed from the start point in real time, detecting monomers, oligomers, and large aggregates in solution simultaneously. In addition, FCS provides a fast and specific analysis of the aggregation process with high sensitivity. Further, it is reasonable to assume that analysis of the fluorescence raw data with FIDA/PCH (see Sect. 3.8), can extend the precision and accuracy further. FIDA/PCH offers complementary information to FCS, and makes it possible to distinguish aggregates from their different fluorescence brightness, which at least ideally scales linearly with the size of the aggregates. FCS, on the other hand, differentiates species based on their diffusive speed, which scales with the cubic root of the molecular size.

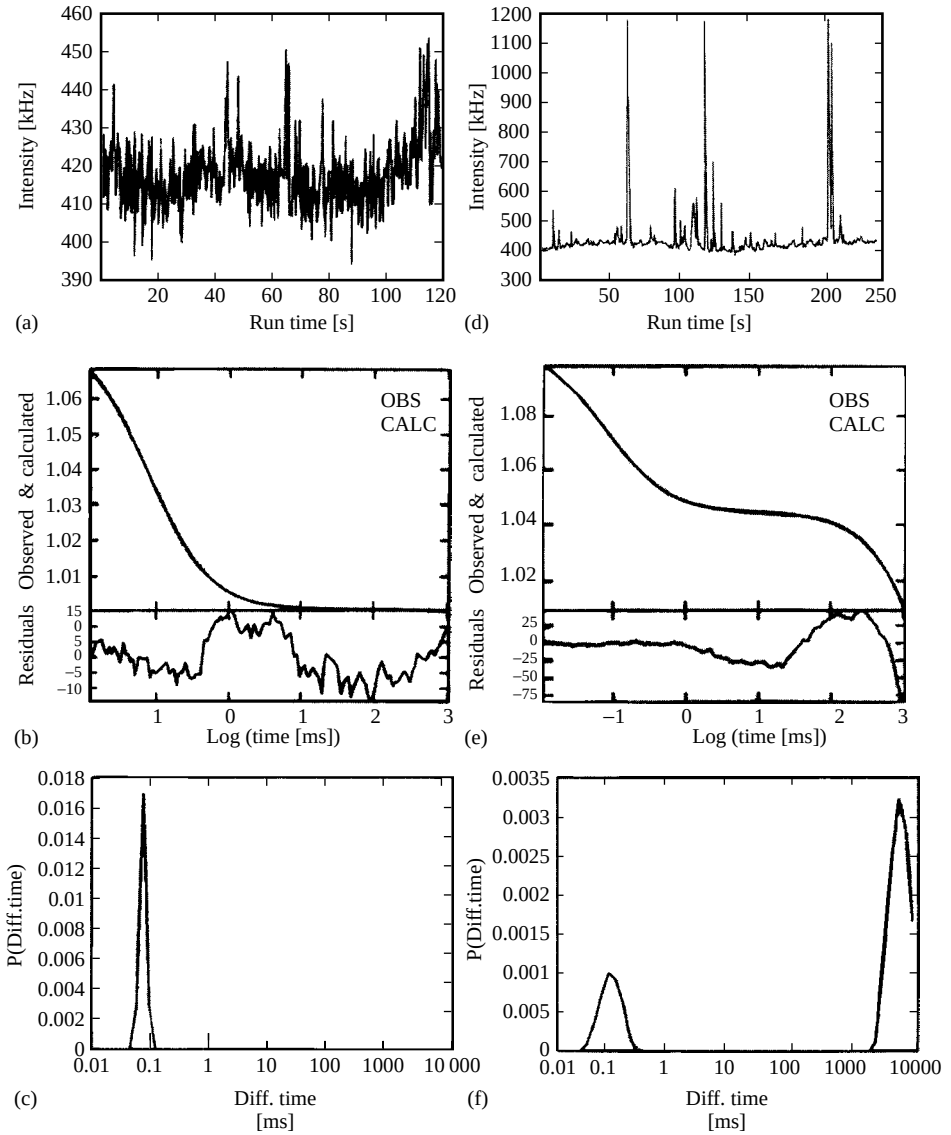

Fig. 6 The polymerization of Aβ studied by FCS using rhodamine-labeled Aβ as fluorescent probe (Eqs. 15 to 17, Sect. 3.3). Aβ (80 µM) was incubated with rhodamine-labeled Aβ (10 nM). The fluorescence intensity fluctuations were measured after (a) 2 min and (d) 40 min. The corresponding autocorrelation functions were calculated according to Eq. (34). (b) The monophasic form of the curve indicates that the diffusion times for the species present are similar. The diffusion time, τ_D, was found to be 90 µs. (e) After 40 min of incubation, the calculated autocorrelation function has a biphasic appearance. Species with a very high diffusion time ($\tau_D = 3$ s), corresponding to very large aggregates, as well as the initial monomers/dimers ($\tau_D = 90$ µs) can be observed. The distribution of diffusion times was calculated from the autocorrelation function using the CONTIN program (c and f).

According to estimations, a relative difference in diffusive times between two different molecular species of at least a factor of 1.6 is necessary, if the species have comparable quantum yields and high fluorescence signals.

4.2 Ligand–Receptor (on Single Live Cell) and Intracellular Interactions

The receptor concept is the primary theoretical basis for modern pharmacology. Drugs, hormones, neurotransmitters, toxin, and other biologically active substances can be referred to as ligands. Ligands exert their actions by way of interaction with receptors. The resulting receptor–ligand complexes produce alterations in physiological processes. For certain ligands (e.g. peptides, natural products), it is not possible to detect receptors with the conventional method radiography. If the numbers of receptors are few per cell, specific binding is difficult to detect because of the background. The radiography method includes several washing steps to remove unbound ligand. The half-life of the receptor–ligand complex is often shorter or similar to the time required for separations of free and bound ligands. Therefore, interactions between certain ligands and their different receptors are often overlooked by the conventional binding technique. The single-molecule detection sensitivity of FCS and its short measurement time can often provide a way to overcome these difficulties allowing receptor–ligand interactions to be followed in living cells, even if the receptors are present only in very sparse numbers.

The ligand–receptor interactions can be analyzed on the basis of the approach and the fluorescence intensity autocorrelation function as given in Eq. (17). In FCS measurements, the fluorescence intensity fluctuations are recorded from only those molecules that diffuse through the confocal laser volume element. The time required for the passage of fluorescent molecules through the volume element is determined by the diffusion coefficient, which is related to the size and shape of a molecule. Thus, diffusion times obtained from the analysis of fluorescence intensity fluctuations with autocorrelation functions allow differentiating faster diffusing and slower diffusing molecules, that is, in this context, to distinguish free ligands from ligand–receptor complexes (Sect. 3.3). In binding studies with the FCS technique, a mixture of several ligand–receptor complexes (components) with different molecular weights and corresponding different diffusion times can be analyzed without any need to physically separate unbound components from bound ones. Since in the FCS technique diffusion times of ligand molecules and ligand–receptor complexes are used as characteristic parameters for binding analysis, distribution of diffusion times can be obtained via a special algorithm that can reveal the appearance of several binding complexes. For the evaluations of the distribution functions, the CONTIN program using restrained regularization was applied. Instead of an evaluation of discrete diffusion processes with distinct species and characteristic diffusion times, the dynamic motion of receptor complexes in membranes is considered to be distributed with regard to their diffusion times. This program can thus be used to validate the presence of discrete ligand–receptor complexes in which the appearance of two binding complexes found by distribution analysis of diffusion times may suggest that these are representations of two different forms or subtypes of ligand–receptor

complexes. Ligand–receptor interactions in live cells using FCS have been performed in several laboratories. In what follows, an example will be given to illustrate how FCS can be used to investigate the interaction of ligands with their cognate receptors. In order to highlight ligand–receptor interactions on live cells (cell cultures) using FCS, the interaction between the ligand proinsulin C-peptide (CP) and its receptor will be demonstrated. It is very important to take background fluorescence into account for the judgment of specific ligand–receptor interaction when a ligand is tagged with a fluorescent dye. Background fluorescence usually originates from endogenous proteins if their excitation wavelength is close to that of the dye used. The endogenous proteins such as NADH and FAD produce autofluorescence in blue–green range. Therefore, sometimes it is advantageous to make excitation in near-IR range, where autofluorescence is less prominent. Fluorescence signals will be overestimated if they are not corrected for background fluorescence. In order to check background fluorescence, the autofluorescence was measured in cells without the addition of fluorophore labeled ligands. Background fluorescence differs very much depending on cell types. For example, at 514-nm excitation, very little background fluorescence was detected in human diploid fibroblasts, pancreatic insulinoma cells, and renal tubular cells, whereas huge background fluorescence was observed in endothelial cells from umbilical cord veins. CP was covalently labeled with the fluorophore rhodamine (Rh) and binding of Rh–CP was conducted in human renal tubular cells.

Fluorescence intensity fluctuations and autocorrelation functions of Rh–CP free in solution and bound to human renal tubular cell membranes are presented in Fig. 7. Examination of the Brownian motion of unbound Rh–CP in the buffer medium above the cell surface exhibits typical fluctuations (Fig. 7a) and a diffusion time (τ_{D1}) of 0.15 ms (Fig. 7b). When the volume element is positioned at the level of the cell membrane, the fluorescence intensity traces exhibited several prominent peaks (Fig. 7d). These peaks, with up to three times higher intensity than the baseline, are a clear representation of the Brownian motion of ligand–receptor complexes containing several rhodamine-labeled CP molecules. Correlation analysis of the intensity fluctuations then reveals a diffusion process of the cell-bound C-peptide with at least two components characterized by diffusion times of $\tau_{D2} = 1$ ms and $\tau_{D3} = 80$ ms, respectively (Fig. 7e). The results indicate the presence of specific binding of human proinsulin C-peptide to its membrane-bound receptors since bound Rh-CP is displaced following addition of 1000-fold molar excess of unlabeled C-peptide (Fig. 7f). Further evidence for the binding specificity of C-peptide is obtained from the fact that addition of excess unlabeled C-terminal pentapeptide was accompanied by displacement of bound Rh-CP, similar to what was observed after the addition of intact C-peptide. This suggests that the C-terminal segment is involved in the binding process (Fig. 7c).

Among other applications of FCS within this category, it has also been applied to precisely study ligand–receptor interactions for the purified acetylcholine receptors in solution. Using FCS, the C-peptide specific binding to macromolecules obtained from detergent-solubilized cell materials has also been shown. Further, combining imaging by confocal laser scanning microscopy and FCS (LSM-FCS) has made it possible to make images of ligand–receptor interaction and its dynamics

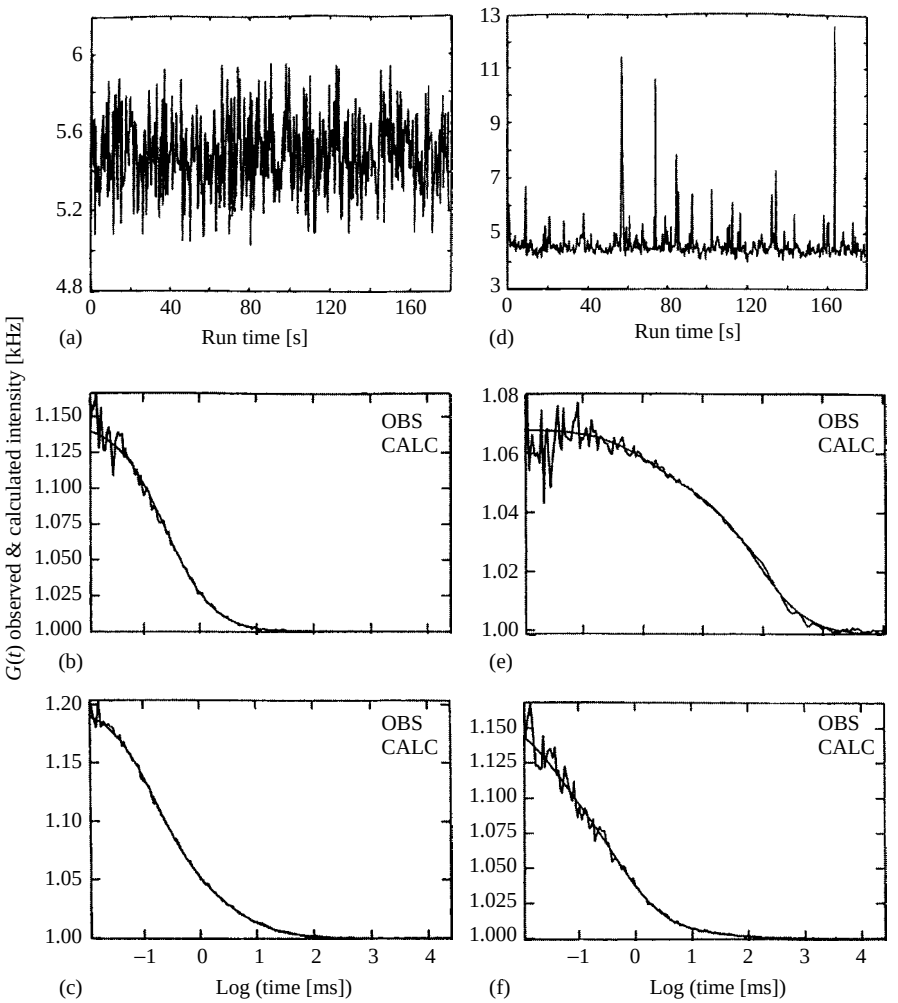

Fig. 7 C-peptide binding and displacement to the membranes of cultured renal tubular cells (Eq. (17), Sect. 3.3). (a) Fluorescence intensity fluctuations, (b) autocorrelation function for Rh-CP (5 nM) free in solution, $\tau_D = 0.15$ ms. (d) Fluorescence intensity fluctuations and (e) autocorrelation function for Rh-CP bound to membranes on the cell surfaces. Diffusion times (τ_D) and corresponding fractions (γ): $\tau_{D1} = 0.15$ ms, $\gamma_1 = 0.1$; $\tau_{D2} = 1$ ms, $\gamma_2 = 0.15$; $\tau_{D3} = 80$ ms, $\gamma_3 = 0.75$. Autocorrelation functions of displacement of membrane-bound Rh-CP by postincubation of a 1000-fold molar excess of (f) nonlabeled C-peptide and (c) nonlabeled C-terminal pentapeptide. In postincubations, renal tubular cells were first incubated with binding buffer containing 5 nM Rh-CP for 60 min and observed binding was treated as control. Then 5 μM nonlabeled CP or nonlabeled C-terminal pentapeptide was added to the same cells, incubated for three more hours, and checked for displacement.

in live cells. The LSM-FCS setup enables the determination of localization, topology, and transport of proteins in living cells, and the visualization and dynamic characterization of lipid bilayers. The LSM-FCS setup appears to be a very valuable technique to study ligand–receptor interactions in any part of living cells and/or cell compartments.

Having outlined the capacity of the FCS technique for cell surface studies, it is also important to emphasize that the measured parameters always have to be carefully evaluated with respect to artifacts. If not considered, photobleaching in the laser excitation beam and the limited reservoir of fluorescent molecules on the cell surface can lead to an underestimation of the density of receptor molecules on a cell surface. The adjustment of proper laser intensity requires great attention, both with respect to the dye and the cell survival. The extent of depletion of a fluorescent compound on a cell surface as measured with FCS depends on several parameters (e.g. diffusion speed, area of free diffusion, excitation intensity, and time of laser exposure). The depletion can thus not only lead to a reduced overall surface density but can also significantly skew a distribution of several diffusion species.

In addition to photobleaching, the incorporation of fluorophore into the ligand is an important factor in receptor binding studies with fluorescently labeled ligands. Ideally, the dye-labeled ligands should not contain unbound dye molecules as they may contribute to nonspecific interactions. Since dye–ligand complexes usually have a certain half-lifetime, the purity of dye-labeled ligand complexes should be checked from time to time (even 100% dye-labeled ligands lose their dye molecules (5–10%) after 2–3 months). A problem concerning ligand tagging is the labeling of low molecular weight ligands, which is related to the limitation of FCS to distinguish molecular sizes by translational diffusion. The increase in molecular weight of a labeled ligand by a factor of 8 only doubles its diffusion time (following from the Stokes–Einstein equation, see text following Eq. 10). It is thus sometimes difficult to distinguish the diffusion time of a dye-labeled ligand from that of the dye, particularly if the molecular weight of the ligand is less than that of the dye molecule (Sect. 4.1). However, Häberlein's group has been successful in labeling low molecular weight ligands with rhodamine when the ratio of molecular weights of the rhodamine-labeled kavain ligand and rhodamine was two. Further, in binding assays with dye-labeled ligands, there is always a risk that a dye can interfere with the affinity due to steric effects. This depends both on the structure of the dye and the position of the ligand to which the dye molecule is tagged. Therefore, it is important to check the activity of the dye-labeled ligands in functional studies. If the dye-labeled ligand fails to produce functional activity, it is advisable either to use another dye or to tag the dye to another position.

FCS analysis of ligand–receptor interactions in cell cultures demands a healthy state of cells. Nonliving cells are unsuitable for receptor binding studies with FCS. To this point, the lifetime of cells is a limiting factor for FCS measurements, where certain cells (e.g. fibroblasts) survive longer than others (e.g. neuronal cells). Membranes of dead cells lack the fluidity of the phospholipid environment, which leads to immobility of ligand–receptor complexes. Washing of the cells, a way to reduce background fluorescence – in this case, usually done 4 to 5 times with PBS – has to be done with care since this washing can also induce damages to cells. Low ionic strength

and temperature of the binding buffer can destroy cells. The pH of the binding buffer must be close to 7.4.

If the ligand–receptor interactions are followed on membranes of cultured cells, it is very critical that the focus of the laser beam is properly positioned on the membrane surface. It is a difficult task to be sure that the focus is exactly on the membrane. As a hint, we propose that if receptor binding is observed, a change of the focus 1 μm upward should lead to a reduction of the binding. One should take into consideration that if the receptor is membrane-bound, the binding observed in the cytoplasm is mostly nonspecific. In certain cases, if the focus is in the cytoplasm, one detects the internalization of receptors. If cells are cultured in eight-well Nunc chambers, it is useful to have them attached to the bottom of the chamber (Nunc chamber is usually used for the cell culture. In FCS technique, it is used as a sample chamber. It has a cover glass in the bottom, suitable for an inverted microscope). During the FCS measurement, cells should not move since a change of cell position during the measurement is likely to introduce errors.

Performing FCS measurements intracellularly confronts the practitioner with more challenges, in terms of photobleaching, autofluorescence, cell viability, and so on. For intracellular measurements, two-photon excitation has proved to be a useful mode of excitation. Two-photon excitation as a single quantum event was first predicted by Göppert-Mayer in 1931. It arises from the simultaneous absorption of two photons that combine their energies to cause the transition to the excited state of the fluorophore. The first experimental verification of this phenomenon could not be accomplished until the invention of lasers 30 years later. In 1990, two-photon excitation of fluorescence was introduced in laser-scanning microscopy as a new tool within biophysical research and given the name two-photon laser-scanning microscopy (TPLSM). Most of the properties that make TPLSM attractive derive from the quadratic dependence of the excitation probability on the excitation light intensity. In a strongly focused excitation beam, this quadratic dependence makes the excitation outside the focal region fall off with the fourth power of the distance from the focal plane in the axial direction. Since photobleaching can be considered to roughly scale with the excitation rate, out-of-focus photobleaching and attenuation of the excitation beam by out-of-focus absorption does not occur to any significant degree. Owing to the inherent confinement of the excitation volume in TPLSM, spatial filtering (pinhole) in the image plane is not necessary. This confinement also allows a greater flexibility in the detection procedure in the sense that the emitted light does not have to pass through the microscope objective, and the emitted light does not have to be focused, so that scattering of the fluorescence may occur without any loss of detection efficiency. Moreover, detection of two-photon excited fluorescence is not hampered by a high background owing to Raman and Rayleigh scattering of the excitation light. These features are attractive for FCS experiments, in particular, intracellular studies. FCS experiments with two-photon excitation in which the diffusion of fluorescently labeled beads were followed in the cytoplasm of live mouse fibroblast cells were first demonstrated in 1995. In further studies, intracellular dynamic processes have been followed by two-photon FCS, utilizing, for example, green fluorescent proteins (GFPs). Two-photon excitation can also be combined with one-photon excitation

such that two-color cross-correlation FCS can be performed with only one excitation wavelength.

4.3 Vesicle–Peptide Interactions

Phospholipid vesicles and artificial membrane systems have been studied intensively as a way of determining the properties of biological membranes. With vesicles it is easier to manipulate the various parameters of the membrane system and thus, to investigate membrane-bound enzymes and membrane transport systems without possible interfering reactions that occur in the cell membrane. One important use of vesicles is as vehicles in drug therapy. Vesicles, prepared with drugs entrapped inside, are being used as carriers for these drugs to target organs, thereby avoiding digestion and unwanted metabolic effects during delivery after intake. The interior of the vesicle is an aqueous environment, and with the extrusion technique, it is possible to prepare vesicles with different substances/dye molecules entrapped inside. Thus, the external and internal environment of phospholipid vesicles can be manipulated, and studies can be conducted on a variety of properties of these synthetic membranes, including their ability to incorporate or release molecules, and their interaction with various substances (peptides, proteins, nucleic acids).

In earlier studies, interactions between proteins and vesicles have been performed using different biochemical procedures, but very little has been done using FCS. However, recently some studies have been performed indicating the promise of FCS for these kinds of studies. The following example demonstrates how FCS has been applied to study molecular interactions between peptides/proteins and vesicle membranes based on the concept as given in Eqs. (15) to (17) (Sect. 3.3). Molecular interactions of melittin, a pore-forming peptide, and magainin, a membrane-breaking peptide, with phospholipid membranes were performed in rhodamine-entrapped vesicles (REV) and in rhodamine-labeled phospholipid vesicles (RLV). The REV entrapped free rhodamine inside could be released out of REV if pores/channels were formed into vesicles, or if they were broken. In contrast to REV, RLV did not entrap free rhodamine inside since rhodamine was covalently bound to the head groups of phospholipids. Therefore, the formation of channels into RLV could not result in rhodamine release in solution and only breakage of RLV could make free rhodamine appear freely in solution. Free rhodamine was used to calibrate the instrument as a reference to estimate diffusion times of the vesicles, and to justify whether the vesicles were intact or broken. The diffusion times of free rhodamine and vesicles were determined separately. The concentration ratio of peptide: phospholipid ranged from 1 : 1 to 1 : 40. At least a ratio of 1 : 20 (peptide : phospholipid) was necessary to obtain pronounced effects of peptides on vesicles.

Intensity autocorrelation functions $G(\tau)$ of free rhodamine and vesicles are shown in Fig. 8. Diffusion times (τ_D) of free rhodamine in solution (Fig. 8a), of REV vesicles with rhodamine entrapped inside as well as of RLV vesicles with rhodamine covalently bound to the head groups of phospholipids and not entrapped inside (Fig. 8b, c) are found to be 0.063 and 9 ± 1 ms respectively. Similar diffusion times are obtained when autocorrelation functions (Fig. 8a–c) are analyzed with the CONTIN program (see Sect. 4.2) with regard to the distribution of diffusion times

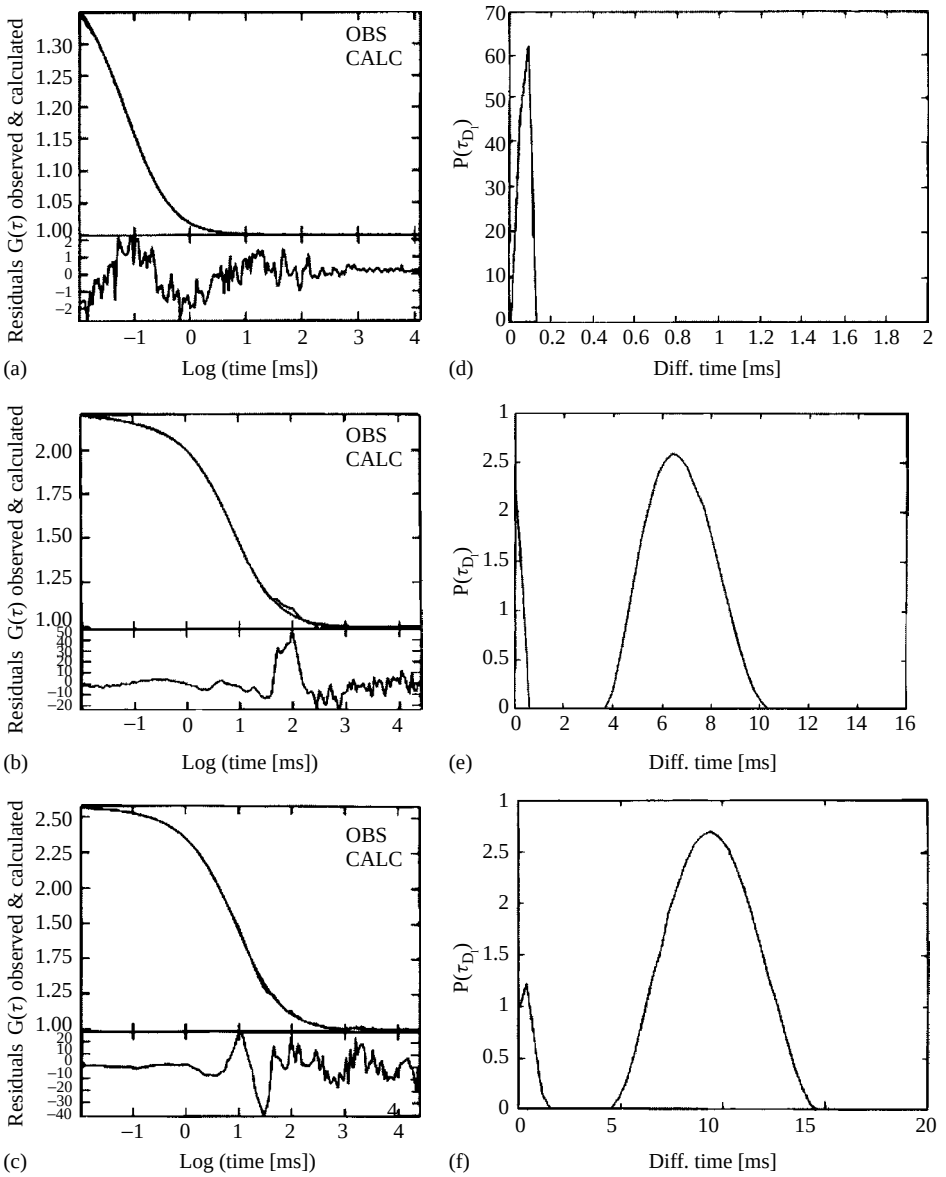

Fig. 8 The intensity autocorrelation function $G(\tau)$ of vesicles. (a) $G(\tau)$ of free rhodamine in solution; $\tau_D = 0.063$ ms. (b) $G(\tau)$ of rhodamine entrapped vesicles (REV), $\tau_D = 7.5$ ms; (c) $G(\tau)$ of rhodamine-labeled phospholipid vesicles (RLV), $\tau_D = 9.1$ ms; (d), (e), and (f) are the CONTIN distributions of diffusion times $P(\tau_{D_i})$ from (a), (b), and (c) respectively.

according to a model of multiple components (Fig. 8d–f). Owing to differences in diffusion times between free rhodamine and vesicles, it is possible to quantify the fractions of free rhodamine and vesicles; for example, in the ligand–receptor interaction studies, there is no need for a separation step for this quantification. The results demonstrate that melittin and magainin interact with phospholipid vesicle membranes in different ways. Melittin makes channels into vesicle membranes (Fig. 9, model I) as exposure of melittin to vesicles causes rhodamine to release only from REV but not from RLV. It is obvious that rhodamine cannot be released from RLV because the inside of RLV is free of dye molecules. In contrast, magainin breaks vesicle membranes (Fig. 9, model II) since addition of magainin to vesicles results in rhodamine release from both REV and RLV. As the inside of RLV is free of rhodamine, the appearance of rhodamine in solution confirms that these vesicles are broken into rhodamine-labeled phospholipid fragments after addition of magainin. This study illustrates several attractive features of the FCS technique for protein/peptide studies including vesicles. FCS is a sensitive and rapid technique to study molecular interactions of biologically important substances with vesicles. It

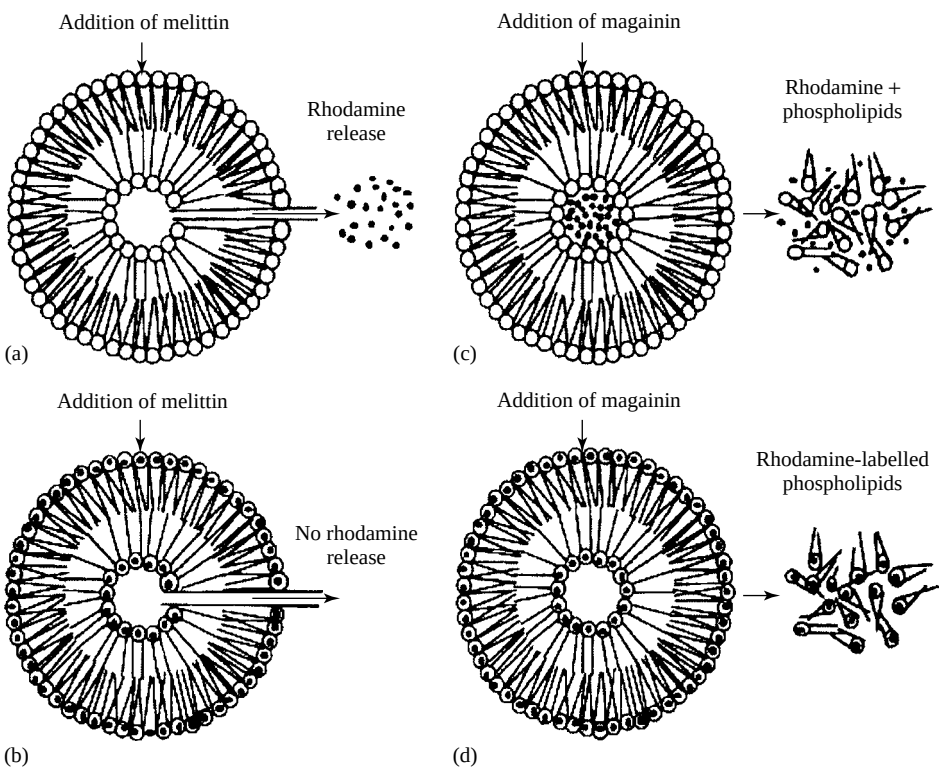

Fig. 9 Models for the peptide interactions with vesicles. Model I: Melittin makes channels into phospholipid vesicle membranes. (a) REV plus melittin; (b) RLV plus melittin; Model II: Magainin breaks phospholipid vesicle membranes; (c) REV plus magainin. (d) RLV plus magainin.

indicates that FCS can be used as a rapid protocol to test incorporation and release of drugs by vesicles, which is of pharmaceutical significance. It also provides a general idea of how FCS could be used to study molecular interactions of many other physiologically important peptides and proteins in vesicle membranes.

4.4 Protein–DNA Interactions

Binding of proteins (transcription factors) to DNA molecules plays key roles in assembling active transcription complexes. Generally, biochemical methods are used to study interactions between transcription factors and DNA molecules. Recently, protein–DNA interactions have been studied using biophysical methods such as NMR and FCS. It is important to mention that in certain cases, earlier data obtained by biochemical methods are inconsistent with those obtained by NMR and FCS with respect to the specificity and mechanisms of protein–DNA interactions. Here is an example of how FCS has been applied to study the interaction of the p53 protein (transcription factor) with DNA on the basis of the concept as given by Eqs. (15–17) (Sect. 3.3). The p53 protein is a sequence-specific transcription factor that plays a central role in cancer surveillance. p53 is a latent transcription factor that is activated by various forms of cellular stress including DNA damage. The p53 activation and consequently, p53-dependent transcription is controlled by the negative regulatory region in the CTD (C-terminal domain) of p53. Data on the p53 activation and its interactions with DNA as characterized by FCS demonstrate that a crucial mechanism by which p53 transcription function is regulated appears to operate through control of its DNA-binding properties. These FCS data present a two binding site mechanism, which postulates that p53 is not a latent DNA-binding factor but is involved in multiple interactions with DNA via its core and C-terminal domains. Thus, the earlier proposed allosteric and steric mechanisms of the p53 activation for binding to DNA are inconsistent with data obtained by FCS. Similar studies have been performed using capillary electrophoresis-single molecule FCS, and by fluorescence fluctuation spectroscopy. These examples provide a general idea of how FCS can be used to study specific interactions of various proteins with DNA. Regarding transcription factors, the FCS technique can offer a useful tool to study their interactions with DNA with reference to the specificity and mechanisms of interactions. In comparison with biochemical methods, FCS is not only fast and economic but also very specific. It is very likely that protein–DNA interaction with FCS can be performed in microchips. The most important perspective of protein–DNA assays with FCS is to study protein–DNA interactions at expression level in living cells.

4.5 DNA–DNA Interaction (DNA Sequencing)

Nucleic acid analysis plays a key role in understanding various biological processes. Generally, hybridization methods are used in order to detect specific DNA and RNA sequences, in which a specific probe (oligonucleotide, DNA, RNA) recognizes its complementary sequence in the target DNA and/or RNA. Traditionally, electrophoresis or liquid chromatography techniques are applied to analyze hybridization

reactions. However, these techniques have drawbacks as electrophoretic separations in agarose and polyacrylamide gel produce very poor resolution and possess lower sensitivity for long natural DNAs. With FCS, a higher sensitivity can be offered to detect DNA sequences using fluorescently labeled probes. For example, it has been described in detail how fluorescently labeled DNA oligonucleotides (19mer–23mer) have been used as probes for the HIV virus target RNA. The most important point is that FCS allows detecting and quantifying hybridization products without the need for bound–free physical separation by gel electrophoresis, filters, or column chromatography. As in the preceding examples, the separation by FCS is based on significant differences in diffusion times between the free and bound dye-oligonucleotides. The approach has also been described for multicomponent analysis using a combination of capillary electrophoresis and FCS. There are a number of publications concerning the fluorescence-based amplification at high template concentrations. As is known, the amplification of initially low copy numbers of templates is a general problem. Detection of DNA target sequences at the single-copy level without amplification can obviously overcome such limitations, which can result in nonspecific amplification at initially very low template numbers, for example, PCR (polymerase chain reaction). The major problem is that in nonspecific modes of detection, there is no guarantee that the specific target is available or that the correct sequence of the specific target amplification is produced in the amplification process. For DNA–DNA interactions of this kind, the two color cross-correlated FCS is a useful approach (Eqs. 22, 23(a–c), Sect. 3.5). In nucleic acid analysis by cross-correlated FCS, two oligonucleotide probes are fluorescently labeled in two different colors (each probe carries one color). These two fluorescently labeled oligonucleotide probes can be hybridized to a complementary target gene (DNA) such that both oligonucleotides bind to the same target at different sites. If the target is present in the sample, those probe molecules that bind to the same target are correlated to each other, in contrast to those probe molecules that do not bind to the same target. Even if there is a prominent excess of molecules in one color, the DNA carrying both probes on the same molecule can often be distinguished in a cross-correlation procedure. Two color cross-correlation FCS has been shown to be a specific mode of detection where the correct amplification product is identified and quantified. Two color cross-correlation FCS (and other cross-correlation/coincidence analyses based on similar approaches using ultrasensitive fluorescence techniques) can be useful for assays in microchips in order to screen for clinically relevant pathogens.

4.6 Monitoring of Molecular Processes in the Fast Time Range ($<\tau_D$) (Chemical Reactions, Photophysics, Conformational Changes)

For reactions that generate prominent fluorescence changes in the fluorescently marked molecules under study, and which are faster than the typical passage times (τ_D) of these molecules through the detection volume, the approach outlined in Sect. 3.4 is applicable.

Reactions that can be monitored include a range of chemical reactions, conformational kinetics, and transitions between different transient photophysical states of the fluorophores themselves.

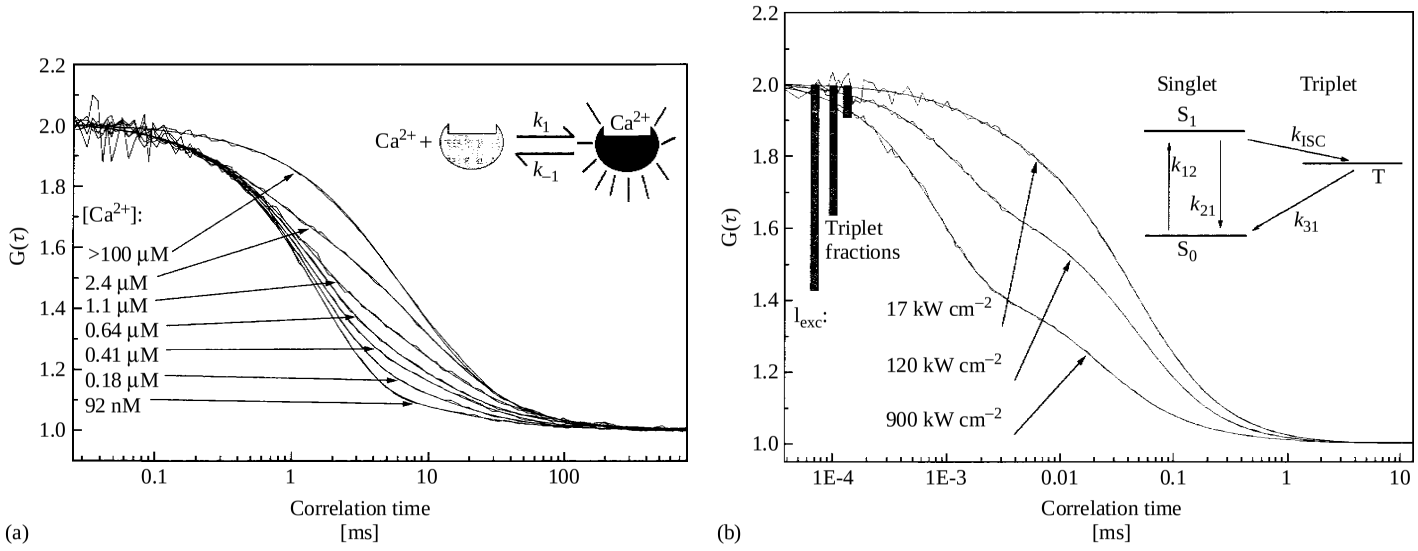

Fig. 10 (a) Experimental example of a reaction process that in an FCS experiment can be described by (Eqs. 18–21). Here, the association and dissociation of Ca^{2+} to the calcium-sensitive dye Rhod-II (Molecular Probes, Eugene, Oregon, USA) were followed. While almost nonfluorescent without calcium bound to it, the fluorescence brightness of the dye increases promptly upon binding to calcium. Here, the observation volume was deliberately expanded in order to make the passage times of the Rhod-II molecules long enough to exceed the relaxation time of the binding process to calcium. (b) Three correlation curves of the dye Rhodamine 6G in water with the total amplitudes normalized to 1. The FCS curves show two relaxation processes that are due to diffusion (slower) and singlet–triplet transitions (faster). With reference to (Eqs. 18–21), the processes correspond to $G_D(\tau)$ and $X(\tau)$, respectively, and the reaction matrix $\sum_{j,k=1}^{M} T_{jk}$ is given by the rate coefficients in the three-level reaction scheme enclosed in the graph with $M = 3$. With the normalization used, the amplitude of the fast process of the curves corresponds to the fraction of dye molecules in the detection volume that are in their triplet states. With increasing excitation intensities, this fraction is increased. The relaxation rate of the faster process corresponds to the relaxation time of the reaction scheme of the singlet–triplet transitions (inset), and increases from k_{31} (triplet deactivation rate) towards $k_{31} + k_{ISC}$ (sum of k_{31} and rate of intersystem crossing) with increasing excitation intensities.

A simple example of a chemical reaction monitored by FCS using this strategy is given in Fig. 10(a), where the fluorescence fluctuations in a calcium-sensitive fluorophore upon binding and dissociation of a calcium ion can be used to determine the calcium concentration in absolute terms as well as the local buffering properties. The same approach also applies to other ion-sensitive dyes, for instance, pH-sensitive dyes.

In Fig. 10(b), it is shown how FCS and the above approach provide a simple and convenient means to determine triplet state properties of fluorophores, offering a strong alternative to flash photolysis and other relaxation approaches. A wide range of photophysical states can be analyzed in a similar fashion, including charge transfer and trans–cis isomerization processes. The relative ease of measurement and the strong environmental sensitivity of these relatively long-lived photophysical states opens for microenvironmental monitoring applications where relatively small changes in the rates to or from these states owing to local environment influences can be followed.

Conformational kinetics of molecules can be followed via FCS if the fluorescence intensity of the fluorophore labeled to the molecule of interest clearly depends on its conformational state. The fluorescence can either be made sensitive to intramolecular distances, that is, between the fluorescent marker and some attached quencher molecule, or via fluorescence resonance energy transfer (FRET) between a donor and an acceptor fluorophore attached to the molecule of interest. Alternatively, the fluorescence is changed because of a conformation-dependent accessibility of quencher molecules to the labeling site of the fluorophore molecule.

4.7 FCS and Related Techniques in Drug Screening

A very important activity for generating new substances with therapeutic activity is the screening of compounds, which is based on their affinity to, or molecular interactions with certain target compounds. Functional genomics has resulted in the knowledge of a growing number of potential therapeutic targets. At the same time, combinatorial chemical synthesis techniques have enabled a rapid development of large compound libraries. Following this progress, the demand is stronger than ever for high throughput screening (HTS) technologies that can screen these potential therapeutic targets against large compound collections. This has stimulated technology development in the areas of assay automation, miniaturization, and detection technologies. Homogeneous assays, in which measurement can take place directly after mixture ("mix and measure"), without any filtration, separation, or washing steps in between, are well suited for HTS as they make the screening less time-consuming and labor-intense, and thus easier to automate. Miniaturization reduces the amount of biological and chemical reagents used per assay, and can facilitate higher throughput assay formats by enabling the use of parallel sample processing. Miniaturization, in turn, puts high demands on the sensitivity of the detection techniques used in the screening procedure.

Fluorescence-based detection methods can offer an outstanding sensitivity, and there is a strong trend towards using fluorescence as a readout mean in favor of, for example, radioactivity. Already some years ago, a strong interest from the pharmaceutical industry was focused on FCS

as a readout tool for HTS. This interest stemmed from several attractive features of the FCS technique as a HTS readout technique: (1) a sensitivity that enables detection at the level of single molecules; (2) small detection and sample volumes that strongly facilitate miniaturization; (3) lack of need for a physical separation of interacting from noninteracting molecules prior to measurement, which makes it possible to perform homogeneous assays; (4) reasonably fast readout times. Further, as indicated by previous studies (see Sect. 4.2), FCS can provide the possibility to identify and discover new target molecules on cell surfaces that has not been possible to detect by conventional methods. This flexibility in the choice of measurement medium thus makes it possible not only to perform HTS *in vitro* but also may open for HTS on a cellular basis.

The ability to distinguish two different particles in solution (e.g. bound and unbound ligands), which is essential in HTS, is typically based on differences in diffusion speeds in FCS. However, it should be pointed out that to resolve, for example, binding of a ligand to a receptor, their diffusion times through the detection volume need to differ quite substantially, since the diffusion constant is only approximately inversely proportional to the cubic root of the molecular weight (see also discussion in Sect. 4.1). The smaller the difference in diffusion speeds, the longer the measurement time needed to separate the species. Therefore, to reach short enough measurement times for HTS with FCS (of the order 1s), large differences in diffusion times between the free and bound species are required.

In view of this, FIDA/PCH (Sect. 3.8) offers a strong complementary analysis approach in HTS to time correlation analysis and FCS, providing a complementary set of readout parameters. In its basic version, the discrimination between species is based on differences in molecular brightness, and already a difference of 25% can allow two different species to be separated in the analysis, independent of the molecular mass, also allowing a faster readout. On the basis of FIDA/PCH assays can be designed in different ways such that the molecular brightness changes because of molecular interaction, either due to a change in environment or due to an energy transfer from one (donor) fluorophore to another (acceptor) fluorophore, or via changes in stoichiometry of fluorophore molecules present on the single particles.

HTS requires fast readout times and at the same time enough information from the sample to judge if, for example, a ligand has bound to a target molecule. Therefore, it is essential to maximize the number of fluorescence photons extracted from the molecules investigated during the readout time. Photophysical aspects are very important for practically all fluorescence spectroscopy techniques in which a high sensitivity or a fast readout is required. Since for HTS, both these properties are needed, this issue becomes of even greater importance. To optimize sensitivity and readout speed, additional strategies can be used. One useful strategy is to take advantage of several spectroscopic parameters, analyzing data from more than one output channel. In line with this strategy, the FIDA technique has been extended to also include lifetime, polarization, and emission wavelength ranges of the emitted fluorescence (Sect. 3.8). In this way, the global analysis of several signal parameters (from a combination of FCS and FIDA approaches) can significantly improve the probability for accurate molecular classification in compound screening. A further boost to the capacity would result from the

use of a parallel confocal optic readout. In combination with assay methods, proper liquid handling and robotics, as well as process flow and information management, ultrasensitive fluorescence detection techniques seem to be able to take HTS to new capacity levels. Assay systems that were capable of screening at their best thousands of compounds per day in the mid-1990s now transform into methods that are believed to be capable of 100 000 samples per day (so-called ultra-HTS (uHTS) methods).

5
Additional Remarks and Conclusions

Local fluctuations, reflecting intrinsic thermodynamic properties of a physical system in equilibrium, are usually a source of unavoidable noise in conventional ensemble experiments. FCS converts the thermodynamic noise into a very rich source of information about dynamic processes on the molecular scale. This is done by a combination of ultrasensitive fluorescence detection, providing the appropriate sensitive molecular indicator and a processing of the detected fluorescence data based on autocorrelation analysis. Autocorrelation analysis can, in a powerful way, select for correlated events on a molecular scale that reflects average properties of the molecules within an ensemble. In this chapter, we have illustrated how FCS and modalities thereof can be used to investigate a wide diversity of molecular phenomena taking place over a huge span of time-ranges.

FCS is excellently suited to retrieve temporal information about dynamic processes buried in a stochastic signal of an ensemble of molecules. Like in single molecule detection (SMD) experiments, where molecules are analyzed one by one, no synchronization by some external perturbation of the molecules is necessary to obtain this information. However, by detection and analysis of molecules one by one, and by the subsequent sorting of the molecules into frequency histograms based on their detected parameters, it is possible to reveal and characterize subpopulations to a greater extent than what is possible with FCS. On the other hand, the need in SMD for individual detection of the molecules makes the accumulation of a statistically sufficient amount of data slower than in FCS and related techniques. Moreover, at concentrations low enough for the measurement volume to harbor only one fluorescent molecule of interest at a time, many biological binding reactions and other forms of chemical reactions are very inefficient, if detectable at all. Owing to the lower concentrations of fluorescently marked molecules required, SMD approaches are also more sensitive to background light, which restricts the range of environments where they can be applied. Clearly, in terms of the conditions under which it can be applied, the information that can be derived, and the speed with which a statistically significant amount of data can be collected, FCS provides a bridge between conventional ensemble measurements and SMD-based approaches.

In FCS as well as in practically all SMD-approaches based on fluorescence detection, it is vital to maximize the number of detected fluorescence photons per molecule in order to obtain maximal information. Photobleaching and saturation of fluorescence set the ultimate limits. However, the information available from an FCS experiment can also be increased by taking advantage of the full information content of the detected fluorescence photons. In FCS, the temporal information of various dynamic processes can be used to determine concentrations of

reactants and reaction products, without the introduction of any physical separation steps. However, depending on the particular system under study, the determination of the concentration of the different reaction partners can sometimes be more easily determined by the use of an alternative way of data processing. In FIDA/PCH, the experimental conditions and the fluorescence raw data are identical, but the data is analyzed differently, with respect to intensity distributions, providing complementary information to FCS analysis. This illustrates the power and importance of an appropriate data processing of the fluorescence raw data, adapted to the system under study and the information sought. Yet another strategy is to take advantage of the additional properties of the detected fluorescence that can be registered in parallel with the fluorescence intensity (emission wavelength, polarization, fluorescence lifetime etc.). Spectral and spatial cross-correlation modalities of FCS, or the introduction of polarizers in front of the detectors to measure rotational diffusion, are examples thereof. When two or more measured parameter values, either of two different parameters in the same measurement or parameters from different measurements, can be expressed as a function of a common parameter, global analysis provides another data processing means to increase the accuracy of the parameter determination.

The drastic improvement in the sensitivity of the FCS instrumentation and the subsequent development of the FCS technique in the last decade illustrates how important the detected information content per molecule is for the overall performance of the method. FCS at its present stage has already found a wide application in biomedical research as a tool to investigate chemical reactions, in particular binding reactions in solutions and on cell surfaces. Clearly, the limits of the technique will be pushed further. The major challenges lie in extending the limits set by photophysical constraints, including the search for photochemically and photophysically more stable dyes, to take full advantage of the information content of the detected fluorescence and to optimally process the registered data, as well as to extend the possible environments where measurements can be performed. Of particular biomedical interest is to be able to more extensively investigate dynamic processes in intracellular environments. Here, FCS and related techniques are likely to be more successful than true SMD approaches because of the high background levels experienced inside cells, generated from light scattering and autofluorescence.

From the examples given, one can conclude that FCS offers a powerful tool for examining molecular interactions of biological importance with very high specificity, providing a fast and specific analysis for several important biological processes with high sensitivity. Thus, FCS can be used as a highly sensitive and specific assay to identify potential inhibitors for the aggregation process, involved in Alzheimer's and prion diseases, and to detect receptors in live cells, even if the receptors are present only in very sparse numbers. FCS can also be used as a rapid protocol to test incorporation and release of drugs by vesicles, which is of pharmaceutical significance. A very important perspective for the future of protein–DNA assays with FCS is to study protein–DNA interactions at expression level in living cells. Finally, two color cross-correlation FCS can be useful for assays in microchips in order to screen for clinically relevant pathogens.

Bibliography

Books and Reviews

Krichevsky, O., Bonnet, G. (2002) Fluorescence correlation spectroscopy: the technique and its applications, *Rep. Phys.* **65**, 251–297.

Muller, J.D., Chen, Y., Gratton, E. (2003) Fluorescence correlation spectroscopy, *Methods Enzymol.* **361**, 69–92.

Rigler, R., Elson, E.L. (Eds.) (2001) *Fluorescence Correlation Spectroscopy – Theory and Applications*, Springer-Verlag, Berlin, Heidelberg, New York.

Thompson, N.L., Lieto, A.M., Allen, N.W. (2002) Recent advances in fluorescence correlation spectroscopy, *Curr. Opin. Struct. Biol.* **12**, 634–641.

Widengren, J., Mets, Ü. (2002) Conceptual Basis of Fluorescence Correlation Spectroscopy and Related Techniques as Tools in Bioscience, in: Zander, C., Enderlein, J., Keller, R.A. (Eds.) *Single Molecule Detection in Solution*, Wiley-VCH Verlag, Berlin, pp. 69–120.

Primary Literature

Auer, M., Moore, K.J., Meyer-Almes, F.J., Guenther, R., Pope, A.J., Stoeckli, K.A. (1998) Fluorescence correlation spectroscopy: lead discovery by miniaturized HTS, *Drug Discovery Today* **3**, 457–465.

Ayed, A., Mulder, F.A., Yi, G.S., Lu, Y., Kay, L.E., Arrowsmith, C.H. (2001) Latent and active p53 are identical in conformation, *Nat. Struct. Biol.* **8**, 756–760.

Bacia, K., Schwille, P. (2003) A dynamic view of cellular processes by in vivo fluorescence auto- and cross-correlation spectroscopy, *Methods (Duluth)* **29**, 74–85.

Basche', T., Moerner, W.E., Orrit, M., Talon, H. (1992) Photon antibunching in the fluorescence of a single dye molecule trapped in a solid, *Phys. Rev. Lett.* **69**, 1516–1519.

Berland, K.M., So, P.T., Gratton, E. (1995) Two-photon fluorescence correlation spectroscopy: method and application to the intracellular environment, *Biophys. J.* **68**, 694–701.

Bernard, J., Fleury, L., Talon, H., Orrit, M. (1993) Photon bunching in the fluorescence from single molecules: A probe for intersystem crossing, *J. Chem. Phys.* **98**, 850–859.

Berne, B.J., Peccora, R. (1975) Dynamic Light Scattering, With Applications to Chemistry, *Biology and Physics*, Wiley, New York.

Björling, S., Kinjo, M., Földes-Papp, Z., Hagman, E., Thyberg, P., Rigler, R. (1998) Fluorescence correlation spectroscopy of enzymatic DNA polymerization, *Biochemistry* **37**, 12971–12978.

Boonen, G., Pramanik, A., Rigler, R., Häberlein, H. (2000) Evidence for specific interactions between a kavain derivative and human cortical neurons measured by fluorescence correlation spectroscopy, *Planta Med.* **66**, 7–10.

Bonnet, G., Krichevsky, O., Libchaber, A. (1998) Kinetics of conformational fluctuations in DNA hairpin-loops, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8602–8606.

Brinkmeier, M., Dörre, K., Stephan, J., Eigen, M. (1999) Two-beam cross-correlation: a method to characterize transport phenomena in micrometer-sized structures, *Anal. Chem.* **71**, 609–616.

Brock, R., Vamosi, G., Vereb, G., Jovin, T.M. (1999) Rapid characterization of green fluorescent protein fusion proteins on the molecular and cellular level by fluorescence correlation microscopy, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10123–10128.

Buehler, C., Stockli, K, Auer, M. (2001) The Integration of Single Molecule Detection Technologies into Miniaturized Drug Screening: Current Status and Future perspective, in: Valeur, B., Brochon, J.C. (Eds.) *New Trends in Fluorescence Spectroscopy, Applications to Chemical and Life Sciences*, Springer-Verlag, Berlin, Heidelberg, New York.

Carmichael, H.J., Walls, D.F. (1976) Proposal for the measurement of the resonant Stark effect by photon correlation techniques, *J. Phys.* **B9**, L43.

Chandrasekhar, S. (1943) *Rev. Mod. Phys.* **15**, 1.

Chattopaday, K., Saffarian, S., Elson, E., Frieden, C. (2002) Measurement of microsecond dynamic motion in the intestinal fatty acid binding protein by using fluorescence correlation spectroscopy, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14171–14176.

Chen, Y., Muller, J.D., Ruan, Q., Gratton, E. (2002) Molecular brightness characterization of EGFP in vivo by fluorescence fluctuation spectroscopy, *Biophys. J.* **82**, 133–144.

Chen, Y., Mueller, J., So, P.T.C., Gratton, E. (1999) The photon counting histogram in fluorescence fluctuation spectroscopy, *Biophys. J.* **77**, 553–567.

Denk, W., Stricler, J.H., Webb, W.W. (1990) Two-photon laser scanning microscopy, *Science* **248**, 73.

Dittrich, P., Malvezzi-Campeggi, F., Jahnz, M., Schwille, P. (2001) Accessing molecular dynamics in cells by fluorescence correlation spectroscopy, *Biol. Chem.* **382**, 491–494.

Ehrenberg, M., Rigler, R. (1974) Rotational brownian motion and fluorescence intensity fluctuations, *Chem. Phys.* **4**, 390–401.

Eigen, M., DeMayer, L. (1963) Relaxation Methods, in: Friess, S.L., Lewis, E.S., Weissberger, A. (Eds.) *Techniques of Organic Chemistry*, Vol. 8, Wiley, New York, pp. 895–1054.

Elson, E.L. (2001) Fluorescence correlation spectroscopy measures molecular transport in cells (review), *Traffic* **2**, 789–796.

Elson, E.L., Magde, D. (1974) Fluorescence correlation spectroscopy. 1. Conceptual basis and theory, *Biopolymers* **13**, 1–27.

Feher, G., Weissman, M. (1973) Fluctuation spectroscopy: determination of chemical reaction kinetics from the frequency spectrum of fluctuations, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 870–875.

Földes-Papp, Z., Thyberg, P., Björling, S., Holmgren, A., Rigler, R. (1997) Exonuclease degradation of DNA studied by fluorescence correlation spectroscopy, *Nucleosides Nucleotides* **16**, 781–787.

Göppert-Mayer, M. (1931) Ueber Elementarakte mit zwei Quantenspruengen, *Ann. Phys.* **19**, 273.

Halbsguth, C., Meissner, O., Haberlein, H.H. (2003) Positive cooperation of protoberberine type 2 alkaloids from Corydalis cava on the GABA(A) binding site, *Planta Med.* **69**, 305–309.

Haustein, E., Schwille, P. (2003) Ultrasensitive investigations of biological systems by fluorescence correlation spectroscopy, *Methods (Duluth)* **29**, 153–166.

Heinze, K.G., Koltermann, A., Schwille, P. (2000) Simultaneous two-photon excitation of distinct labels for dual-color fluorescence cross correlation analysis, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10377–10382.

Hertzberg, R.P., Pope, A.J. (2000) High-throughput screening: new technology for the 21st century, *Curr. Opin. Chem. Biol.* **4**, 445–451.

Hink, M.A., Bisseling, T., Visser, A.J.W.G. (2002) Imaging protein-protein interactions in living cells, *Plant Mol. Biol.* **50**, 871–883.

Hodgkin, A.L., Huxley, A.F., Katz, B. (1949) Measurements of ionic currents underlying activity in the giant axon of the squid, *Arch. Sci. Physiol. (Paris)* **3**, 129–150.

Kahya, N., Pecheur, E.I., de Boeij, W.P., Wiersma, D.A., Hoekstra, D. (2001) Reconstitution of membrane proteins into giant unilamellar vesicles via peptide-induced fusion, *Biophys. J.* **81**, 1464–1474.

Kask, P., Palo, K., Ullmann, D., Gall, K. (1999) Fluorescence intensity distribution analysis and its application in biomolecular detection technology, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13756–13761.

Kim, H.D., Nienhaus, G.U., Ha, T., Orr, J.W., Williamson, J.R., Chu, S. (2002) Mg2+-dependent conformational change of RNA studied by fluorescence correlation and FRET on immobilized single molecules, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4284–4289.

Kinjo, M. (1998) Detection of asymmetric PCR products in homogeneous solution by fluorescence correlation spectroscopy, *Biotechniques* **25**, 706–715.

Kinjo, M., Nishimura, G., Koyama, T., Mets, Ü., Rigler, R. (1998) Single-molecule analysis of restriction DNA fragments using fluorescence correlation spectroscopy, *Analyt. Biochem.* **260**, 166–172.

Kinjo, M., Rigler, R. (1995) Ultrasensitive hybridization analysis using fluorescence correlation spectroscopy, *Nucleic Acids Res.* **23**, 1795–1799.

Koppel, D.E., Axelrod, D., Schlessinger, J., Elson, E.L., Webb, W.W. (1976) Dynamics of fluorescence marker concentration as a probe of mobility, *Biophys. J.* **16**, 1315–1329.

Korlach, J., Schwille, P., Webb, W.W., Feigenson, G.W. (1999) Characterization of lipid bilayer phases by confocal microscopy and fluorescence correlation spectroscopy, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8461–8466.

LeCaptain, D.J., Michel, M.A., Van Orden, A. (2001) Characterization of DNA-protein complexes by capillary electrophoresis-single molecule fluorescence correlation spectroscopy, *Analyst* **126**, 1279–1284.

Lomakin, A., Chung, D.S., Benedek, G.B., Kirschner, D.A., Teplow, D.B. (1996) On the

nucleation and growth of amyloid β-protein fibrils: detection of nuclei and quantitation of rate constants, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1125–1129.

Magde, D., Elson, E.L., Webb, W.W. (1972) Thermodynamic fluctuations in a reacting system-measurement by fluorescence correlation spectroscopy, *Phys. Rev. Lett.* **29**, 705–711.

Magde, D., Elson, E.L., Webb, W.W. (1974) Fluorescence correlation spectroscopy. 2. An experimental realization, *Biopolymers* **13**, 29–61.

Magde, D., Webb, W.W., Elson, E.L. (1978) Fluorescence correlation spectroscopy. 3. Uniform translation and laminar flow, *Biopolymers* **17**, 361–376.

Meseth, U., Wohland, T., Rigler, R., Vogel, H. (1999) The resolution limit of fluorescence correlation spectroscopy for two-component, *Biophys. J.* **76**, 1619–1631.

Meissner, O., Haberlein, H. (2003) Lateral mobility and specific binding to GABA(A) receptors on hippocampal neurons monitored by fluorescence correlation spectroscopy, *Biochemistry* **42**, 1667–1672.

Mets, U., Rigler, R. (1994) Submillisecond detection of single rhodamine molecules in water, *J. Fluoresc.* **4**, 259–264.

Neher, E., Sakmann, B. (1976) Single-channel currents recorded from membrane of denervated frog muscle fibres, *Nature* **260**, 779–802.

Oren, M. (1999) Regulation of the p53 tumor suppressor protein, *J. Biol. Chem.* **274**, 36031–36034.

Palmer, A.G., Thompson, N.L. (1987) Theory of sample translation in fluorescence correlation spectroscopy, *Biophys. J.* **51**, 339–343.

Palmer, A.G., Thompson, N.L. (1989) Fluorescence correlation spectroscopy for detecting submicroscopic clusters of fluorescent molecules in membranes, *Chem. Phys. Lipids* **50**, 253–270.

Palo, K., Mets, U., Jäger, S., Kask, P., Gall, K. (2000) Fluorescence intensity multiple distributions analysis. Concurrent determination of diffusion times and molecular brightness, *Biophys. J.* **79**, 2858–2866.

Petersen, N.O. (1986) Scanning fluorescence correlation spectroscopy. I. Theory and simulation of aggregation measurements, *Biophys. J.* **49**, 809–815.

Petersen, N., Brown, C., Kaminski, A., Rocheleau, J., Srivastava, M., Wiseman, P.W. (1998) Analysis of membrane cluster densities and sizes in situ by image correlation spectroscopy, *Faraday Discuss.* **111**, 289–305.

Petersen, N.O., Hoddelius, P.L., Wiseman, P.W., Seger, O., Magnusson, K.-E. (1993) Quantitation of membrane receptor distributions by image correlation spectroscopy: concept and application, *Biophys. J.* **65**, 1135–1146.

Petersen, N.O., Johnson, D.C., Schlesinger, M.J. (1986) Scanning fluorescence correlation spectroscopy, *Biophys. J.* **49**, 817–820.

Pick, H., Preuss, A.K., Mayer, M., Wohland, T., Hovius, R., Vogel, H. (2003) Monitoring expression and clustering of the ionotropic 5HT(3) receptor in plasma membranes of live biological cells, *Biochemistry* **42**, 877–884.

Pitschke, M., Prior, R., Haupt, M., Riesner, D. (1998) Detection of single amyloid β-protein aggregates in the cerebrospinal fluid of Alzheimer's patients by fluorescence correlation spectroscopy, *Nat. Med.* **4**, 832–834.

Pramanik, A., Ekberg, K., Zhong, Z., Shabqat, J., Henriksson, M., Tibell, A., Tally, M., Wahren, J., Jörnvall, H., Rigler, R., Johansson, J. (2001) C-peptide binding to human cell membranes: importance of Glu27, *Biochem. Biophys. Res. Commun.* **284**, 94–98.

Pramanik, A., Juréus, A., Langel, Ü., Bartfai, T., Rigler, R. (1999) Galanin receptor binding in the membranes of cultured cells measured by fluorescence correlation spectroscopy, *Biomed. Chromatogr.* **13**, 119–121.

Pramanik, A., Juréus, A., Langel, Ü., Bartfai, T., Rigler, R. (2001) Fluorescence correlation spectroscopy detects galanin receptor diversity on insulinoma cells, *Biochemistry* **40**, 10839–10845.

Pramanik, A., Rigler, R. (2001) Ligand–receptor interactions in the membrane of cultured cells monitored by fluorescence correlation spectroscopy, *Biol. Chem.* **382**, 371–378.

Pramanik, A., Rigler, R. (2001) FCS-assay of Ligand-receptor Interactions in Living Cells, in: Rigler, R., Elson, E.L. (Eds.) *Fluorescence Correlation Spectroscopy (FCS). Theory and Applications*, Springer-Verlag, Berlin, Heidelberg, pp. 101–129.

Pramanik, A., Thyberg, P., Rigler, R. (2000) Molecular interactions of peptides with phospholipid vesicle membranes as studied by fluorescence correlation spectroscopy, *Chem. Phys. Lipids* **104**, 35–47.

Provencher, S.W. (1982) Contin: a general purpose constrained regulation program for inverting noisy linear algebraic and integral equations, *Comp. Phys. Commun.* **27**, 229–242.

Rauer, B., Neumann, E., Widengren, J., Rigler, R. (1996) Fluorescence correlation spectrometry of the interaction kinetics of tetramethylrhodamin alfa-bungarotoxin with Torpedo Californica acetylcholine receptor, *Biophys. Chem.* **58**, 3–12.

Rex, S. (1996) Pore formation induced by the peptide melittin in different lipid vesicle membranes, *Biophys. Chem.* **58**, 75–85.

Riesner, D. (2001) Protein Aggregation Associated with Alzheimer and Prion Diseases, in: Rigler, R., Elson, E.L. (Eds.) *Fluorescence Correlation Spectroscopy (FCS). Theory and Applications*, Springer-Verlag, Berlin, Heidelberg, pp. 225–247.

Rigler, R. (1995) Fluorescence correlations, single molecule detection and large number screening. Applications in biotechnology, *J. Biotech.* **41**, 177–186.

Rigler, R., Földes–Papp, Z., Mayer-Almes, F.J., Sammet, C., Völcker, M., Schnetz, A. (1998) Fluorescence cross correlation: a new concept for polymerase chain reaction, *J. Biotech.* **63**, 97–109.

Rigler, R., Mets, U. (1992) Diffusion of single molecules through a Gaussian laser beam, *SPIE* **1921**, 239–248.

Rigler, R., Mets, U., Widengren, J., Kask, P. (1993) Fluorescence correlation spectroscopy with high count rate and low background: analysis of translational diffusion, *Eur. Biophys. J.* **22**, 169–175.

Rigler, R., Pramanik, A., Jonasson, P., Kratz, G., Jansson, O.T., Nygren, P.-Å., Ståhl, S., Ekberg, K., Johansson, B.-L., Uhlén, S., Uhlén, M., Jörnvall, H., Wahren, J. (1999) Specific binding of proinsulin C-peptide to human cell membranes, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13318–13323.

Rigler, R., Widengren, J. (1990) Ultrasensitive Detection of Single Molecules by Fluorescence Correlation Spectroscopy, in: Klinge, B., Owman, C. (Eds.) *Bioscience*, Lund University Press, Lund, pp. 180–183.

Rigler, R., Widengren, J., Mets, U. (1992) Interactions and Kinetics of Single Molecules as Observed by Fluorescence Correlation Spectroscopy, in: Wolfbeis, O.S. (Ed.) *Fluorescence Spectroscopy*, Springer-Verlag, Berlin, pp. 13–24.

Ruan, Q.Q., Chen, Y., Gratton, E., Glaser, M., Mantulin, W.W. (2002) Cellular characterization of adenylate kinase and its isoform: two-photon excitation fluorescence imaging and fluorescence correlation spectroscopy, *Biophys. J.* **83**, 3177–3187.

Saito, K., Ito, E., Takakuwa, Y., Tamura, M., Kinjo, M. (2003) In situ observation of mobility and anchoring of PKC beta I in plasma membrane, *FEBS Lett.* **541**, 126–131.

Schaefer, D.W. (1973) Dynamics of number fluctuations: motile microorganisms, *Science* **180**, 1293.

Schwille, P. (2001) Fluorescence correlation spectroscopy and its potential for intracellular applications (review), *Cell Biochem. Biophys.* **34**, 383–408.

Schwille, P., Haupts, U., Maiti, S., Webb, W.W. (1999) Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation, *Biophys. J.* **77**, 2251–2265.

Schwille, P., Meyer-Almes, F.J., Rigler, R. (1997) Dual-color fluorescence cross-correlation spectroscopy for multicomponent diffusional analysis in solution, *Biophys. J.* **72**, 1878–1886.

Sundberg, S.A. (2000) High-throughput and ultra-high-throughput screening: solution and cell-based approaches, *Anal. Biotechnol.* **11**, 47–53.

Svedberg, T. (1911) *Z. Phys. Chem.* **77**, 147.

Tairi, A.-P., Hovius, R., Pick, H., Blasey, H., Bernard, A., Suprenant, A., Lundström, K., Vogel, H., (1998) Ligand binding to the serotonin 5HT3 receptor studied with a novel fluorescent ligand, *Biochemistry* **37**, 15850–15864.

Terada, S., Kinjo, M., Hirokawa, N. (2000) Oligomeric tubulin in large transporting complex is transport via kinesin in squid giant axons, *Cell* **103**, 141–155.

Tjernberg, L., Pramanik, A., Björling, S., Thyberg, P., Thyberg, J., Nordstedt, C., Terenius, L., Rigler, R. (1999) Amyloid β-peptide polymerization studied by fluorescence correlation spectroscopy, *Chem. Biol.* **6**, 53–62.

Vercammen, J., Maertens, G., Gerard, M., De Clercq, E., Debyser, Z., Engelborghs, Y. (2002) DNA-induced polymerization of HIV-1 integrase analyzed with fluorescence fluctuation spectroscopy, *J. Biol. Chem.* **277**, 38045–38052.

von Smoluchowski, M. (1914) *Wien Berichte* **123**, 2381.

Waizenegger, T., Fischer, R., Brock, R. (2002) Intracellular concentration measurements in adherent cells: a comparison of import efficiencies of cell-permeable peptides, *Biol. Chem.* **383**, 291–299.

Wennmalm, S., Rigler., R. (1999) On death numbers and survival times of single dye molecules, *J. Physiol. Chem.* **103**, 2516–2519.

Weiss, M., Hashimoto, H., Nilsson, T. (2003) Anomalous protein diffusion in living cells as seen by fluorescence correlation spectroscopy, *Biophys. J.* **84**, 4043–4052.

Widengren, J. (2001) Photophysical Aspects of Fluorescence Correlation Spectroscopy Measurements, in: Rigler, R., Elson, E.L. (Eds.) *Fluorescence Correlation Spectroscopy – Theory and Applications*, Springer-Verlag, Heidelberg.

Widengren, J., Mets, Ü., Rigler, R. (1995) Fluorescence correlation spectroscopy of triplet states in solution: a theoretical and experimental study, *J. Phys. Chem.* **99**, 13368–13379.

Widengren, J., Rigler, R., Mets, Ü. (1994) Triplet-state monitoring by fluorescence correlation spectroscopy, *J. Fluoresc.* **4**, 255–258.

Widengren, J., Rigler, R. (1996) Mechanisms of photobleaching investigated by fluorescence correlation spectroscopy, *Bioimaging* **4**, 149–157.

Widengren, J., Rigler, R. (1997) An alternative way of monitoring ion concentrations using fluorescence correlation spectroscopy, *J. Fluoresc.* **7**, 211–213.

Widengren, J., Rigler, R. (1998) Fluorescence correlation spectroscopy as a tool to investigate chemical reactions in solutions and on cell surfaces, *Cell. Mol. Biol.* **44**, 857–879.

Widengren, J., Terry, B., Rigler, R. (1999) Protonation kinetics of GFP and FITC investigated by FCS – aspects of the use of fluorescent indicators for measuring pH, *Chem. Phys.* **249**, 259–271.

Wohland, T., Friedrich, K., Hovius, R., Vogel, H. (1999) Study of ligand-receptor interactions by fluorescence correlation spectroscopy with different fluorophores: evidence that the homopentameric 5-hydroxytryptamine type 3As receptor binds only one ligand, *Biochemistry* **38**, 8671–8681.

Wölcke, J., Ullmann, D. (2001) Miniaturized HTS technologies – uHTS, *Drug Discovery Today* **6**, 637–646.

Yakovleva, T., Pramanik, A., Kawasaki, T., Tanno, K., Gileva, I., Lindegren, H., Langel, Ü., Ekström, T.J., Rigler, R., Terenius, L., Bakalkin, G. (2001) p53 latency: C-terminal domain prevents binding of p53 core to target but not to nonspecific DNA sequences, *J. Biol. Chem.* **276**, 15650–15658.

Yakovleva, T., Pramanik, A., Ekström, T., Terenius, L., Bakalkin, G. (2002) p53 latency – out of the blind alley, *Trends Biochem. Sci. (TiBS)* **27**, 612–618.

Yoshida, N., Kinjo, M., Tamura, M. (2001) Microenvironment of endosomal aqueous phase investigated by the mobility of microparticles using fluorescence correlation spectroscopy, *Biochem. Biophys. Res. Commun.* **280**, 312–318.

Zasloff, M. (1987) Magainin, a class of antimicrobial peptides from *Xenopus* skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5449–5453.

Zhong, Z., Pramanik, A., Ekberg, K., Kratz, G., Wahren, J., Rigler, R. (2001) Insulin binding monitored by fluorescence correlation spectroscopy, *Diabetologia* **44**, 1184–1188.

Zumbusch, A., Fleury, L., Brown, R., Bernard, J., Orrit, M. (1993) Probing individual two-level systems by correlation of single molecule fluorescence, *Phys. Rev. Lett.* **70**, 3584–3587.

Fluorescence In Situ Hybridization, Gene Mapping by: *see* Gene Mapping by Fluorescence In Situ Hybridization

Fluorescent 2D-difference Gel Electrophoresis: *see* Gel Electrophoresis, 2D-difference

Folding of Proteins: *see* Aggregation, Protein

Fluorescence

David M. Jameson
Department of Cell and Molecular Biology John A. Burns School of Medicine University of Hawaii 1960 East-West Rd. Honolulu, HI 96822

1 Introduction 441

2 Historical Overview 441

3 Basic Fluorescence Processes 442
3.1 The Fluorescence Emission Spectrum 444
3.2 The Excitation Spectrum of the Fluorescence 445
3.3 Quantum Yield 445
3.4 The Fluorescence Lifetime 446
3.5 The Polarization (Anisotropy) of the Emission 447

4 Multiphoton Fluorescence 450

5 Fluorescent Probes 452

6 Green Fluorescent Proteins 452

7 FRET 454

8 Flow Cytometry 455

9 FCS 456

10 FISH 456

Acknowledgments 456

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Bibliography 457
Books and Reviews 457
Primary Literature 457

Keywords

Excited State
The energy level a molecule reaches subsequent to the absorption of light, which is characterized by a different distribution of electron density compared to the ground (nonexcited) state of the molecule.

Fluorophore
A molecule that exhibits the property of fluorescence.

Stokes Shift
The difference in the wavelengths corresponding to the maxima of a fluorophore's absorption and emission.

Monochromator
An optical device that spatially separates electromagnetic radiation and hence allows for the isolation of a particular wavelength band.

Photomultiplier
An electronic device with a light-sensitive component that can convert light energy into an electrical signal and that can greatly amplify (multiply) the original photon energy.

Polarizer
An optical component that selectively transmits one direction of the electric field oscillation.

FRET
Fluorescence resonance energy transfer – the process by which the excited-state energy of a fluorophore (the donor) is transferred via a nonradiative process to another molecule (the acceptor).

FCS
Fluorescence correlation spectroscopy – a method that measures fluctuations in the fluorescence signal from a small sample volume and uses this data to gain information on diverse molecular parameters.

FISH
Fluorescence *in situ* hybridization – a method that utilizes fluorescently labeled nucleic acids sequences that hybridize to complementary sequences in cells, thus permitting their localization.

Fluorescence is the light emitted by an atom or molecule after a finite duration (typically on the order of nanoseconds) subsequent to the absorption of electromagnetic energy. Specifically, the emitted light arises from the transition of the excited species from its first excited electronic singlet level to its ground electronic level, and is usually (but not always) at longer wavelengths than the absorbed light. The development of highly sophisticated fluorescent probe chemistries, new laser and microscopy approaches, and site-directed mutagenesis has led to many novel applications of fluorescence in the chemical, physical, and life sciences. Fluorescence methodologies are now widely used in the biochemical and biophysical areas, in clinical chemistry and diagnostics, and in cell and molecular biology.

1 Introduction

Luminescent substances give rise to the emission of light after absorption of energy from an outside source. The most common source of this energy is light itself (photoexcited luminescence), although light emission following heating (thermal luminescence), chemical reactions *in vitro* (chemiluminescence) or *in vivo* (bioluminescence), or interaction with high-energy particles are also well known and have been extensively studied. In this article, we shall be concerned only with *fluorescence*, a process that occurs after photoexcitation. Moreover, we shall limit our discussion to fluorescence from molecules, as opposed to fluorescence from atoms.

2 Historical Overview

The first reported observation of fluorescence was that from a water infusion of the wood from a small Mexican tree, *Lignum nephriticum*. This wood was very popular in sixteenth to seventeenth century Europe, where it was known as *Lignum nephriticum* (kidney wood), because of its medicinal virtues for treating kidney affections. In 1565, Nicolás Monardes described a bluish opalescence from this infusion. These enigmatic optical properties eventually attracted the attention of several prominent scientists, including Athanasius Kircher, Isaac Newton, and Robert Boyle. The actual term *fluorescence* was coined in 1853 by Sir George Gabriel Stokes, whose seminal contributions helped systematize the study of light emission. Stokes clearly enunciated the fact that the fluorescent light was of longer wavelength than the exciting light – this wavelength difference between excitation and emission is now known as the *Stokes Shift*. Stokes' work led to a surge of qualitative studies on plant and animal extracts.

The theoretical foundations of modern quantitative fluorescence spectroscopy were established in the first half of the twentieth century by pioneers including Enrique Gaviola, Jean and Francis Perrin (father and son), Peter Pringsheim, Sergei Vavilov, F. Weigert, F. Dushinsky, Alexander Jabłoński, Theodor Förster, and more recently, Gregorio Weber. During this time, important concepts were introduced such as the excited-state lifetime, the polarization of the fluorescence, the quantum yield of

the emission process, and nonradiative energy transfer.

3
Basic Fluorescence Processes

Fluorescence occurs when a molecule (or atom) absorbs light and – after a finite interval usually on the order of nanoseconds – emits light, usually of lower energy. This phenomenon is illustrated in Fig. 1, which shows solutions of several different fluorophores illuminated by a UV hand lamp. Several aspects of the fluorescence phenomenon make it an extremely useful technique for investigations of molecular processes. Perhaps the most salient feature is the fact that the excited state persists for a duration in the range of nanoseconds, that is, some billionths of a second. Then, in essence, the excited fluorophore is like a molecular stopwatch that starts with the absorption of light and stops with the emission of light. The fluorophore thus provides a probe of events that occur on the nanosecond timescale. Observation of fluorescence requires an excitation source and a detector. In the simplest cases – as in the very early studies – these may correspond to sunlight and the eye. Presently, however, the availability of light sources providing radiation over a wide range of energies, encompassing the ultraviolet to the infrared, and of photodetectors with wide wavelength ranges, very high sensitivity (to the level of single photon detection), and ultrafast response times has greatly expanded the potential of fluorescence methods and has established their role in the chemical, physical, and life sciences.

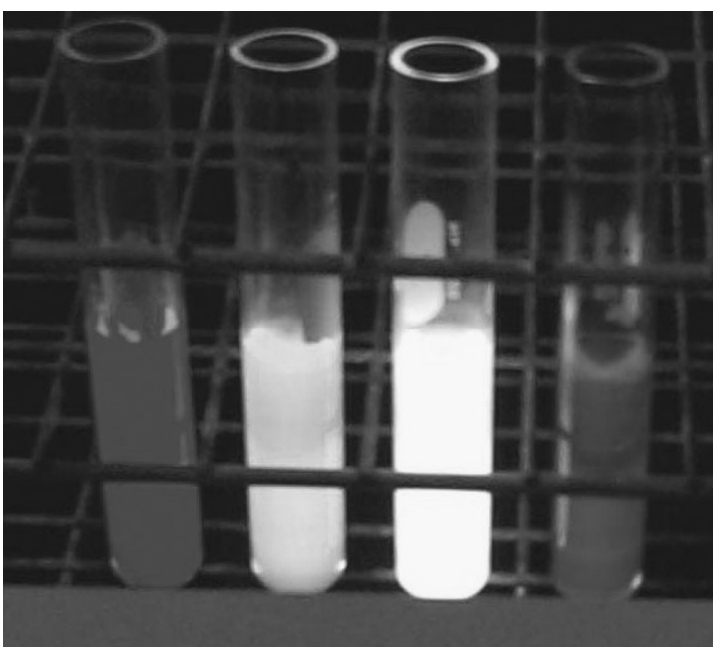

Fig. 1 Solutions of quinine sulfate, fluorescein, rhodamine, and ethidium bromide (from left to right) illuminated by the 365-nm line of a UV hand lamp.

Virtually all fluorescence data required for any research project will fall into one of the following categories:

1. fluorescence emission spectrum;
2. excitation spectrum of the fluorescence;
3. quantum yield;
4. fluorescence lifetime; and
5. polarization (anisotropy) of the emission.

Each of these categories will be discussed briefly since all of the special applications discussed later will utilize these parameters in one form or another. First, however, it is necessary to describe basic fluorescence instrumentation in more detail. To observe and record the fluorescence emission, we require an exciting source, as nearly monochromatic as possible, a monochromator to disperse the fluorescence emission, and a detector, usually a photomultiplier. The monochromatic excitation may be provided by a continuous light source such as a xenon arc (although laser sources that are intrinsically monochromatic may be used), and an additional monochromator. The plan of a typical spectrofluorimeter is shown in Fig. 2. Fluorescence measurements are almost

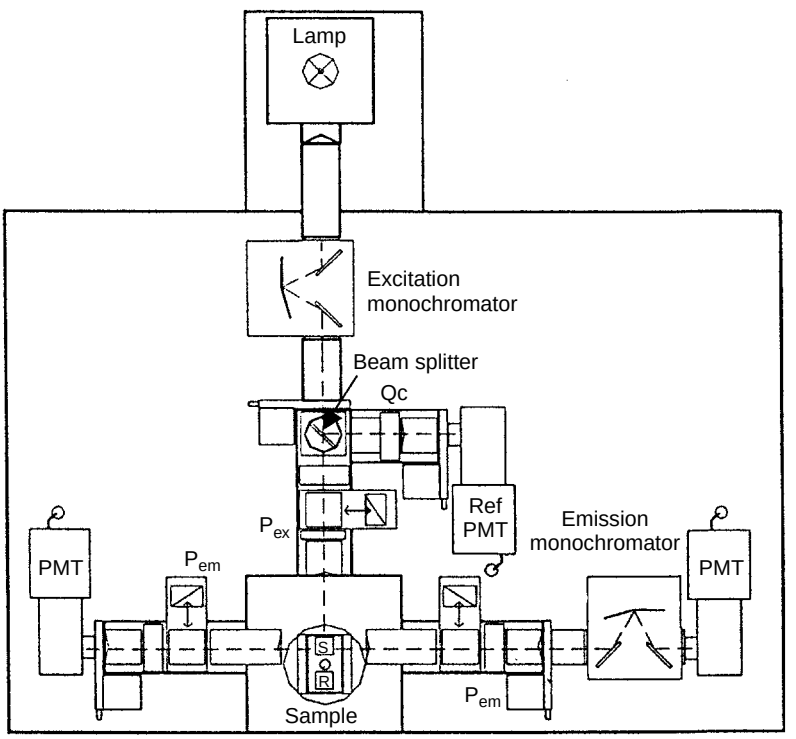

Fig. 2 Schematic representation diagram of a spectrofluorimeter indicating key components: revised with permission from commercial literature from ISS, Inc. The principal components include the xenon arc lamp, the excitation and emission monochromators, a quartz beam splitter, a quantum counting solution (Qc), three photomultiplier tubes (PMT) and excitation (P_{ex}) and emission (P_{em}) calcite prism polarizers.

always accomplished using a right-angle geometry, that is, the emission is detected in a direction 90° with respect to the excitation. This approach is used to reduce the level of exciting light reaching the detector side. All of the basic fluorescence measurements can also be made through a microscope and the quantitative use of fluorescence microscopy has increased tremendously in the last two decades. An in-depth discussion of the particular optics associated with fluorescence microscopes is beyond the scope of this article but the basic components of the spectrofluorimeter described above have counterparts in a fluorescence microscope.

3.1 The Fluorescence Emission Spectrum

In a typical emission spectrum, the excitation wavelength is fixed and the fluorescence intensity versus wavelength is obtained. A typical result is shown in Fig. 3(a) – which also depicts the absorption spectrum. The intensity axis of a fluorescence spectrum is usually reported in "arbitrary units." This nomenclature recognizes the fact that the observed intensity depends not only on the concentration of the fluorophore and on molecular factors such as the fluorophore's extinction coefficient (a measure of a molecule's ability to absorb light energy at a particular wavelength) and quantum yield (*vide infra*), but also on instrumental factors such as the monochromator slit widths, the photomultiplier voltage, and amplifier gains. Fluorescence spectra recorded in this manner are "technical" or "uncorrected" emission spectra. Uncorrected spectra do not take into account the wavelength-dependent response of the emission optics (principally the monochromator) and the photodetector (most commonly a photomultiplier tube, although devices such as *avalanche photodiodes* are also becoming more common).

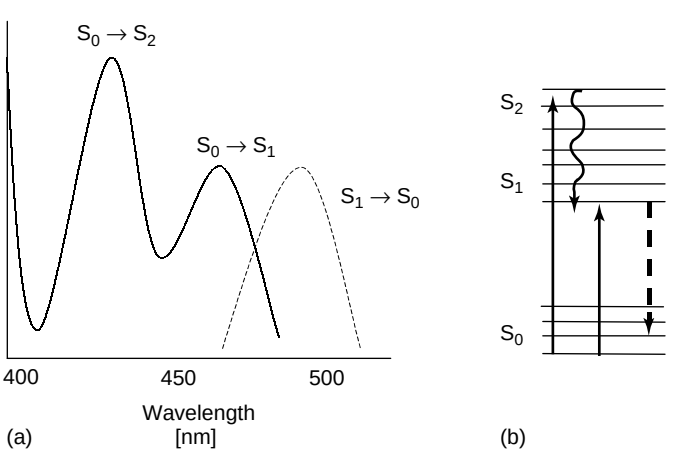

Fig. 3 (a) Schematic representation of the absorption (solid line) and emission (dotted line) spectrum of a fluorescent compound. (b) Perrin–Jabłoński Diagram. S_0 is the ground state while S_1 and S_2 are electronically excited states. The timescales of the absorption (solid up arrows), thermalization (thin squiggly down arrow) and emission (dotted down arrow) processes are typically in the range of $\sim 10^{-15}$, $\sim 10^{-12}$, and $\sim 10^{-9}$ s respectively.

Examination of a large number of emission spectra resulted in the formulation of certain general rules:

1. In a pure substance existing in solution in a unique form, the fluorescence spectrum is invariant, remaining the same independent of the excitation wavelength.
2. The fluorescence spectrum lies at longer wavelengths than the absorption (the Stokes shift).
3. The fluorescence spectrum is, to a good approximation, a mirror image of the absorption band of least frequency.

These general observations follow from consideration of the Perrin–Jabłoński diagram shown in Fig. 3(b). This diagram depicts the various electronic excited states that may be populated upon exposure of a fluorophore to light. Although the fluorophore may be excited into different singlet-state energy levels, which would correspond to excitation with different wavelengths, (e.g. S_1, S_2, etc.) rapid thermalization invariably occurs and emission almost always takes place from the lowest vibrational level of the first excited electronic state (S_1). This fact accounts for the independence of the emission spectrum from the excitation wavelength. The fact that ground-state fluorophores, at room temperature, are predominantly in the lowest vibrational level of the ground electronic state (as required from the Boltzmann's distribution law) accounts for the Stokes shift. Finally, the fact that the spacing of the energy levels in the vibrational manifolds of the ground state and first excited electronic states are usually similar accounts for the fact that the emission and absorption spectra (plotted in energy units such as reciprocal wave numbers) are approximately mirror images.

3.2 The Excitation Spectrum of the Fluorescence

The relative efficiencies of different wavelengths of incident light to excite fluorophores are represented by the excitation spectrum. In this case, the excitation is varied while the emission wavelength is kept constant. If the system is "well-behaved", that is, if the three general rules outlined above hold, one would expect that the excitation spectrum will match the absorption spectrum. Analogous to the "corrected" emission spectrum, the technical excitation spectrum must be corrected for the wavelength variation of the excitation source intensity.

3.3 Quantum Yield

The quantum yield (q) of a fluorophore is essentially a measure of how efficiently absorbed light is emitted. One way to view this quantity is $q =$ number of quanta emitted/number of quanta absorbed. The definition of the quantum yield and its first direct measurements were due to S. Vavilov in 1922, who discovered that fluorescein was an almost perfect emitter with a quantum yield close to unity. A more insightful expression relates the quantum yield to the relative rates of the radiative and nonradiative pathways, which deactivate the excited state. Hence,

$$q = \frac{k_r}{k_r + \sum k_{nr}} \tag{1}$$

where k_r and k_{nr} correspond to radiative and nonradiative processes respectively. In this equation, Σk_{nr} designates the sum of the rate constants for the various processes that compete with the emission

process. These processes include photochemical and dissociative processes in which the products are well-characterized chemical species (electron, proton, radical, molecular isomer) as well as less well-characterized changes that result in a return to the ground state with simultaneous dissipation of the energy of the excited state into heat (e.g. collisional quenching – *vide infra*). These latter processes are collectively called *nonradiative transitions* and two types have been clearly recognized: intersystem crossing and internal conversion. Intersystem crossing refers to the radiationless spin inversion of S_1 in the excited state that results in the isoenergetic, or almost isoenergetic, conversion into a triplet state. When spin inversion occurs in S_1, the resulting triplet reached is T_1, which is characterized by an energy intermediate between the ground state and the first excited singlet state. Another process that is nonradiative from the point of view of the original excited molecule is resonance energy transfer, which will be discussed later.

Molecules that decrease a fluorophore's quantum yield by direct interaction are often termed *quenchers*. Fluorescence quenching can result from interaction of the quenching molecule with the excited fluorophore – a process known as *collisional quenching* – or by formation of a ground-state complex between the quencher and the fluorophore. Both these quenching processes are usually reversible, as opposed to photobleaching of the excited fluorophore, which is usually (but not always) irreversible. Collisional quenching, which is also termed *dynamic quenching*, can be used to obtain information about the accessibility of the excited fluorophore to the solvent carrying the quencher. For example, the accessibilities of tryptophan residues in proteins can be accessed in this way; common quenchers used in such studies include iodide ions, acrylamide, and molecular oxygen. Dynamic quenching results in a decrease in both the fluorescence intensity and the excited-state lifetime. Static quenching, on the other hand, reduces the intensity but does not affect the excited-state lifetime, since the fluorophores that do not form a ground-state complex with the quencher will give rise to the same excited-state kinetics. The Stern–Volmer equation (below) is often used to describe simple quenching processes.

$$\frac{F_0}{F} = 1 + K_{SV}[Q] \qquad (2)$$

where F_0 and F are the fluorescence intensities observed in the absence and presence, respectively, of quencher, $[Q]$ is the quencher concentration and K_{SV} is the Stern–Volmer quenching constant. K_{SV} is equal to the product of k_q, the bimolecular quenching rate constant (proportional to the sum of the diffusion coefficients for fluorophore and quencher) and τ_0, the excited-state lifetime in the absence of quencher. In the case of purely dynamic quenching, $F_0/F = \tau_0/\tau$. If both static and dynamic quenching are occurring, then a modified Stern–Volmer equation may be applicable, for example,

$$\frac{F_0}{F} = (1 + k_q\tau_0[Q])(1 + K_a[Q]) \qquad (3)$$

where K_a is the association constant for the ground-state complex between quencher and fluorophore.

3.4 The Fluorescence Lifetime

The fluorescence lifetime is essentially a measure of the "average" duration of the excited state. Although we speak of

such properties as if they are studied in isolation, this is not usually the case. Emission processes are usually studied on *populations* of molecules and the properties of the supposed typical members of the population deduced from the macroscopic properties of the process (although recently, highly specialized instrumentation has been developed that permit, in favorable cases, the observation of emission processes from single molecules). Fluorescence lifetime measurements are traditionally realized using either the *impulse response* (time-domain) method, in which excitation is by a very brief pulse of light after which the direct decay of the fluorescence is observed, or the *harmonic response* (frequency-domain) method, in which the intensity of the excitation light is modulated sinusoidally and the phase shift of the fluorescence, relative to the excitation, is determined. Direct measurements of fluorescence lifetimes, using the harmonic response or phase method, were first successfully realized by the Argentinean, Enrique Gaviola in 1926. The first use of impulse methods to determine fluorescence lifetimes is generally credited to S.S. Brody in 1957. Since these early beginnings, both methods have advanced tremendously. Subnanosecond lifetimes can be routinely measured and these methods can also be applied through specially adapted fluorescence microscopes for *in vivo* cell studies. This method is known as *fluorescence lifetime imaging microscopy* (FLIM).

3.5 The Polarization (Anisotropy) of the Emission

Light can be considered as oscillations of an electromagnetic field perpendicular to the direction of propagation. Although both electric and magnetic components are present, we are usually concerned only with the electric field. In a typical polarization experiment, the sample is illuminated by monochromatic light through a polarizing device. A polarizer is an optical component, which selectively transmits one direction of the electric field oscillation. The most common polarizers used today are (1) dichroic devices, which operate by effectively absorbing one plane of polarization (e.g. Polaroid type-H sheets based on stretched polyvinyl alcohol impregnated with iodine); and (2) double refracting calcite ($CaCO_3$) crystal polarizers – which differentially disperse the two planes of polarization (examples of this class of polarizers are Nicol polarizers, Wollaston prisms, and Glan-type polarizers such as the Glan-Foucault, Glan-Thompson, and Glan-Taylor polarizers). If plane-polarized light of a particular electric vector orientation impinges on a sample, only those molecules that are properly oriented relative to this electric vector can absorb the light. Specifically, the probability of the absorption is proportional to the cosine squared of the angle θ between the exciting light and the molecule's absorption dipole ($\cos^2\theta$).

Hence, when we excite an ensemble of randomly oriented fluorophores with plane-polarized light, we are performing a *photoselection* process, creating a population of excited molecules that nominally have their excited dipoles lined up with the polarization direction of the excitation. If the fluorophore is excited into the lowest excited electronic state (the S_1 level shown in Fig. 3(b)), then the emitted light will usually be from a dipole oriented parallel to the absorption dipole.

As usually in fluorescence, the observation direction is at right angles to the excitation and the emitted light (suitably

isolated either with filters or a monochromator) is viewed through polarizers set to alternately pass the light, which is parallel or perpendicular to the laboratory axis (Fig. 4). Under these conditions, we can define the polarization function as shown in Fig. 4.

$$P = \frac{I_{||} - I_{\perp}}{I_{||} + I_{\perp}} \tag{4}$$

where P is the polarization and $I_{||}$ and $I_{\perp}$ are the parallel and perpendicular components, respectively, of the emission. If the fluorophore does not rotate during its excited-state lifetime (e.g. by being in a very viscous solvent or by being associated with a large macromolecule), then the emission will be polarized ($P > 0$), that is, the fluorescence intensity observed when the emission polarizer is oriented parallel to the vertical laboratory axis will exceed that observed when the emission polarizer is oriented perpendicular to the vertical laboratory axis. Note that an exception to this situation can occur when the fluorophore is excited into a higher electronic state (e.g. the S_2 level) if this energy level is characterized by an absorption dipole that is at a large angle relative to the emission dipole. In these cases, the observed polarization can be negative.

The analogous term anisotropy, usually denoted as r, is given by

$$r = \frac{I_{||} - I_{\perp}}{I_{||} + 2I_{\perp}} \tag{5}$$

In principle, the information content of the polarization and anisotropy functions are identical although there are practical circumstances that may argue in favor of one function over the other (e.g. the denominator in the anisotropy equation corresponds to the total intensity emitted from a sample excited with parallel polarized light and viewed at right angles). Polarization and anisotropy values can be readily converted via the equation

$$r = \frac{2P}{3 - P} \tag{6}$$

The essential aspect of fluorescence polarization, which makes it useful in the biomedical field, is that it provides information on the rotational mobility of a fluorescent molecule. In 1920, F. Weigert discovered that the fluorescence from solutions of dyes was polarized. He discovered that polarization increased with the size of

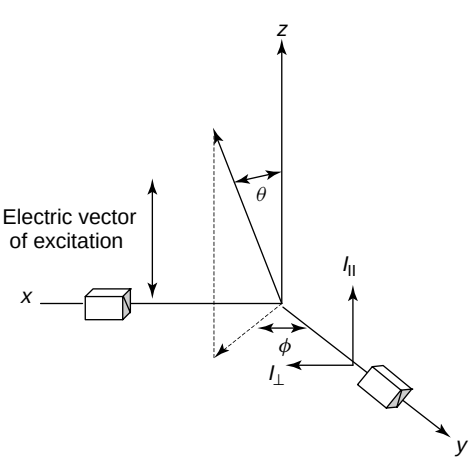

Fig. 4 Diagram depicting photoselection of an absorption dipole, making angles θ with the z-axis (in the x-z plane) and ϕ with the y-axis (in the x-y plane) by vertically polarized light and observation along the y-axis of the vertical (parallel) and horizontal (perpendicular) components of the emission through a polarizer. The rectangular boxes represent polarizers.

the dye molecule and the viscosity of the solvent, yet decreased as the temperature increased. He recognized that all of these considerations meant that fluorescence polarization increased as the mobility of the emitting species decreased. In 1925–1926, Francis Perrin (son of the famous French physicist, Jean Perrin), published several important papers describing a quantitative theory of fluorescence polarization: his famous equation, now known as the *Perrin equation*, contains most of the essential information that we use to this day and (in modern form) is given as

$$\frac{1}{p} - \frac{1}{3} = \left(\frac{1}{p_0} - \frac{1}{3}\right)\left(1 + \frac{RT}{V\eta}\tau\right) \quad (7)$$

where P is the observed polarization, P_0 is the limiting polarization (which is the polarization in the absence of rotation and which depends upon the relative orientation of the absorption and emission dipoles), R is the universal gas constant, T is the absolute temperature, V is the molar volume of the rotating unit, η is the solvent viscosity, and τ is the excited-state lifetime. This equation can be rewritten to have more direct application to the rotation of fluorescent molecules as

$$\frac{1}{p} - \frac{1}{3} = \left(\frac{1}{p_0} - \frac{1}{3}\right)\left(1 + \frac{3\tau}{\rho}\right) \quad (8)$$

where ρ is the Debye rotational relaxation time, which, for a sphere is given as

$$\rho_0 = \frac{3\eta V}{RT} \quad (9)$$

The limiting polarization, and hence the observable polarization, can vary between $+1/2$ and $-1/3$, depending upon the orientation of the emission dipole (the electric dipole on the nuclear framework, which exists during the excited state). Usually, though not always, excitation into the highest wavelength absorption band results in the highest possible limiting polarization. The rotational relaxation time of a molecule is the time it takes to rotate through an angle equal to the arccos of $1/e$ – which corresponds to $\sim 68.4°$. (A quick and very approximate estimation of the Debye rotational relaxation time, in nanoseconds, of a molecule can be obtained by dividing the molecule's molecular weight by 1000 – e.g. a spherical protein of molecular weight 50 000 Da will have a Debye rotational relaxation time in the range of 50 ns). Hence, polarization measurements give direct information on the rotational mobility of a fluorophore and on any molecule attached to the probe.

The Perrin equation can be written in terms of anisotropy, limiting anisotropy (r_0) and rotational correlation time (τ_c) as

$$\frac{r_0}{r} = \left(1 + \frac{\tau}{\tau_c}\right) \quad (10)$$

where τ_c, the rotational correlation time is one-third the Debye rotational relaxation time.

Since the mobility of a fluorescence molecule will change upon binding to another molecule, one can utilize fluorescence polarization to study and quantify biomolecular interactions. Consequently, fluorescence polarization has been used for many years to study biochemical systems, such as protein–protein and protein–ligand interactions, membrane dynamics, and in recent years, protein–DNA and DNA–DNA interactions. In the last two decades, fluorescence polarization methods have also been increasingly used in medical diagnostics and high-throughput screening. One of the most widely used techniques in clinical assays for drugs or metabolites is the fluorescence polarization immunoassay (FPIA). In this method, the target molecule, which

is often but not always a drug or metabolite, is chemically linked to a fluorophore (the probe fluorescein is often used). In the presence of an antibody specific for the target molecule, the fluorophore-linked target will be bound to the antibody and its rotational mobility will be slow (relative to the fluorescence lifetime) and hence the emission will have a high polarization. When the appropriate sample is introduced, derived, for example, from blood or other fluids, the target molecule, if present, will bind to the antibody, displacing the fluorephore-linked target. Free in solution, this fluorophore-target molecule will have an increased rotational mobility (lower Debye rotational relaxation time) and hence a lower polarization. Hence, it is relatively straightforward to setup a standard curve relating the observed polarization to the concentration of the target molecule in the sample. The diagnostic field, in general, uses the unit known as *millipolarization* or "mP", which is the normal polarization times 1000; for example, a polarization of 0.200 corresponds to 200 mP units. Fluorescence plate-readers, which can measure both intensity and polarization, are now routinely used for high-throughput screening. At the time of this writing, fluorescence polarization plate-readers for 1536 wells are available – but one can expect that number to increase in the near future.

4
Multiphoton Fluorescence

As long ago as 1931, the famous physicist Maria Goeppert-Mayer predicted the possibility of nonlinear, multiphoton absorption. It was not until the 1990s, however, that the modern use of multiphoton fluorescence microscopy developed. If an excitation light density is sufficiently high, which can be achieved using ultrashort pulsed (femtosecond), high peak power (kilowatt) laser sources such as a Titanium : Sapphire laser (which emits in the near-infrared, over a range of around 700–1000 nm), then two photons, each at approximately half the wavelength of the normal (i.e. one photon) absorption of the target chromophore, can be simultaneously absorbed. The shape and position of a fluorophore's two-photon absorption cross-section can differ somewhat from that expected from the one-photon absorption spectrum. Multiphoton excitation, coupled with a fluorescence microscope, offers several advantages over traditional one-photon excitation, namely, (1) a very large separation between the excitation and emission wavelengths (which greatly minimizes problems associated with scattered light); (2) an intrinsic confocal aspect; and (3) reduced overall photobleaching, since light outside the focal region is at a wavelength not usually absorbed by the sample. Single- and two-photon excitation processes are depicted in Fig. 5(a). The sketch on the left side of Fig. 5(b) attempts to depict a typical light density profile generated by a microscope objective – one notes that exciting light, although maximal at the focal point, persists throughout a larger area, which means that fluorophores both above and below the focal plane are excited. Fluorescence from the focal plane can be preferentially isolated using traditional confocal optics. The drawing on the right side of Fig. 5(b) attempts to depict the fact that the two-photon light density is at a far longer wavelength than the one-photon case, such that the sample is essentially transparent to this wavelength, and also that the light density at the focal point is sufficiently high to generate the two-photon effect so that fluorophore

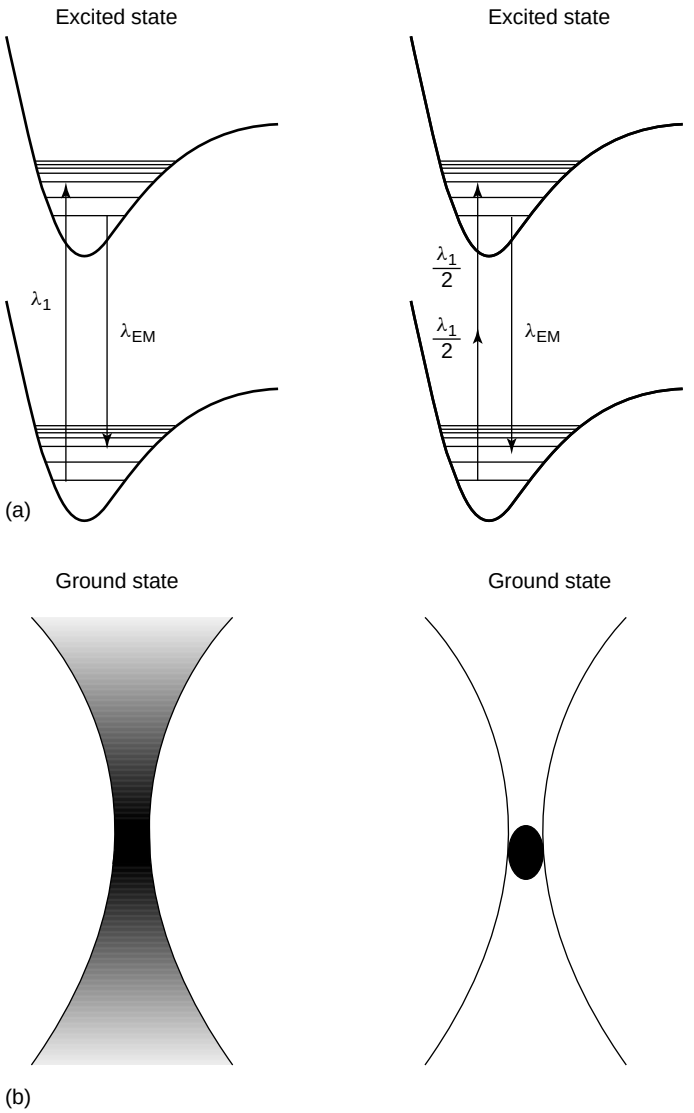

Fig. 5 Depiction of one- and two-photon absorption processes. (a) Energy level diagrams depicting (left) one-photon absorption and emission and (right) two-photon absorption and emission. (b) Depiction of effective excitation light energy in a volume element illuminated by a microscope objective for (left) one-photon and (right) two-photon processes.

excitation occurs only in a small volume near the focal point. Two-photon excitation is currently finding wide application in *in vivo* fluorescence microscopy.

5
Fluorescent Probes

Fluorescence probes can be broadly placed into two categories, namely, intrinsic probes and extrinsic probes. Intrinsic probes include any naturally occurring molecule that exhibits sufficient fluorescence to be of practical utility. Examples include the aromatic amino acids, especially tryptophan and tyrosine, NADH, FAD, FMN, some porphyrins, some fatty acids and lipids, some modified nucleic acids (such as the Wye base in some transfer RNAs), pyridoxal phosphate, chlorophylls, and pteridines. Extrinsic probes are those that can be introduced into the target system, either noncovalently or covalently. Up until the late 1970s, the choice of commercially available probes was quite limited – typically, researchers had to design and synthesize fluorophores for particular types of studies. The choice of fluorophore was dictated by the properties of the instruments (light sources and photomultiplier sensitivities) as well as by the lifetimes and spectral properties required. The first probe made for physicochemical studies of proteins was dimethylaminonaphthalene sulfonyl chloride (dansyl chloride), synthesized by Gregorio Weber around 1950 and designed to be conjugated to proteins and used for fluorescence polarization measurements. In the last two decades, however, the commercial availability of fluorescence has greatly expanded; thousands of fluorescence probes are now available. Examples of common noncovalent probes, which associate reversibly with various biomolecules are ANS, DPH, LAURDAN, ethidium bromide, Hoechst dyes, and DAPI, to name just a few. Covalent probes, on the other hand, are specifically functionalized with reactive groups such as sulfonyl chloride, isothiocyanate, succinimidyl ester, iodoacetamide, maleimide, hydrazine derivatives, azides, and others, to target a wide variety of functional groups in proteins, nucleic acids, carbohydrates, and lipids, including amine and carboxylic acid groups, thiol groups, hydroxyl groups, and so on. In addition to probes specifically designed to target biomolecules for solution studies, a wide range of probes exists, which are specialized for fluorescence microscopy studies on living cells. Such probes include a wide variety of ion-specific fluorophores that change one or more of their fluorescence properties (such as quantum yield, excitation or emission spectrum) upon binding ions including calcium, magnesium, zinc, sodium, potassium, or chloride, and hence can be used to follow alterations in the levels of such ions in living cells. Additionally, fluorescence studies on cells are often accomplished using fluorescently labeled antibodies targeted to specific cell constituents, chimeric proteins linked to green fluorescent proteins (GFP) (*vide infra*) or fluorophores that partition to specific cellular structures such as membranes, mitochondria, nuclei, and so on. An example of cells labeled with multiple fluorescent probes is illustrated in Fig. 6.

6
Green Fluorescent Proteins

The past decade has witnessed an explosion in the use of the family of naturally fluorescent proteins known as *green fluorescent proteins* or GFPs. GFP, a protein

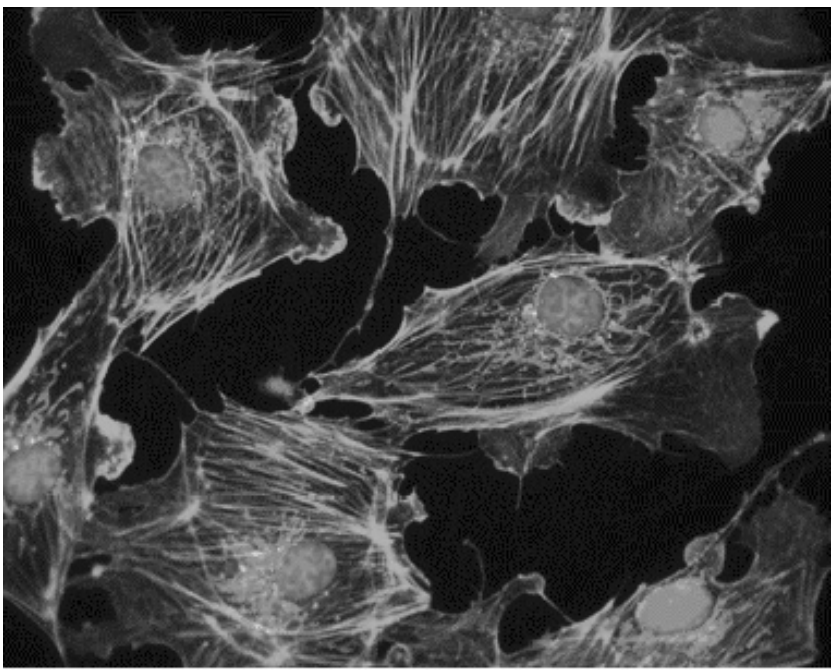

Fig. 6 Fluorescence microscope image (multiexposure) showing bovine pulmonary artery endothelial cells labeled with a red fluorescent dye that stains mitochondria in live cells, and then fixed, permeabilized, and stained with BODIPY FL phallacidin, a green fluorescent, high-affinity probe for F-actin, and DAPI, a probe that emits blue fluorescence upon binding to AT regions of DNA. Reproduced with permission from the Molecular Probes Image Gallery (copyright owned by Molecular Probes, Inc.) (see color plate p. xxv).

containing 128 amino acid residues, was originally isolated from the pacific northwest jellyfish *Aequorea victoria*. The fluorescence moiety in GFP, contained within a ß-sheet barrel, derives from autocatalytic and cyclization reactions involving three amino acid residues, namely, Ser65, Tyr66, and Gly67. This reaction is unusual and occurs only because of the unique environment offered by the overall GFP structure. The biological function of GFP is to absorb a photon emitted by Aequorin, a protein that catalyzes a chemiluminescence in the presence of calcium. By absorbing the blue photon (~470 nm) from the Aequorin complex and reemitting a green photon (~520 nm), GFP shifts the emission from the jellyfish to longer wavelengths, which presumably confers a biological advantage (for example, the green photon may be less prone to scattering processes as it traverses the ocean environment and hence conveys information over longer distances). In the 1990s, molecular biologists began to prepare *in vivo* chimeric proteins containing GFP (i.e. molecular biology techniques were used to link the DNA encoding for the protein of interest to the DNA encoding for GFP), which enabled observation of the chimeric proteins using fluorescent microscopy. It was soon found that alteration of the primary sequence of wild type GFP could result in GFP variants with diverse spectral properties

such as red- or blue-shifted emissions, higher quantum yields, and greatly improved photostability. These GFP mutants, sometimes denoted according to their colors such as yellow fluorescent protein (YFP), cyan fluorescent protein (CFP), or enhanced green fluorescent protein (EGFP), enable the tracking of multiple chimeric protein *in vivo* and can also serve as donors and acceptors in fluorescence resonance energy transfer (FRET) measurements (*vide infra*). Examples of mice carrying GFP-actin chimeras is shown in Fig. 7, which also demonstrates that the DNA encoding for the chimera is transmittable to new generations. More recently, other classes of intrinsically fluorescent proteins have been discovered in coral such as *Discosoma*, and in other organisms. A well-characterized example is the red fluorescent protein drFP593, more commonly known as DsRed. These proteins, like the GFP family, are able to generate a fluorescent moiety out of amino acids in the primary sequence. As the name suggests, the emission from these proteins is at longer wavelengths (red-shifted) relative to GFP.

7 FRET

Fluorescence resonance energy transfer (sometimes referred to as Förster resonance energy transfer in honor of Theodor Förster's contribution to the theory) is now widely utilized in both biochemistry and cell biology. The basic concept is that fluorescence molecules in their excited

Fig. 7 GFP mice (photograph reproduced with the permission of the University of Hawaii).

state can transfer the excited-state energy to a nearby acceptor molecule. Such FRET can be detected by a decrease in the quantum yield and lifetime of the donor fluorophore and an increase in the fluorescence of the acceptor (if the acceptor is a fluorophore). The rate of transfer, $k_T(r)$, is given by

$$k_T(r) = \frac{Q_D \kappa^2}{\tau_D r^6}\left(\frac{9000(\ln 10)}{128\pi^5 N n^4}\right) \times \int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda \quad (11)$$

where Q_D is the quantum yield of the donor, τ_D is the donor lifetime in the absence of acceptor, r is the distance between donor and acceptor, N is Avogadro's number, n is the refractive index of the medium between donor and acceptor, $F_D(\lambda)$ is the fluorescence intensity of the donor in the wavelength range λ to $\lambda + \Delta\lambda$, $\varepsilon_A(\lambda)$ is the extinction coefficient of the acceptor at wavelength λ, and κ^2 is the "orientation factor," which is assumed to be 2/3 in the case of dynamic random averaging between donor and acceptor. Since the efficiency of energy transfer depends upon the inverse sixth power of distance between donor and acceptor, FRET methods provide a "spectroscopic ruler" for molecular systems. The utility of FRET ultimately depends on the so-called *critical transfer distance* between donor and acceptor, which is the distance at which 50% transfer will occur. This value is typically in the range of 10 Å to 100 Å and depends upon the specific characteristics of the donor and acceptor molecules, such as the extent of overlap between the donor fluorescence and the acceptor absorption. Although FRET methods are normally applied *in vitro* and often form the basis of diagnostic assays, recent advances in using site-directed mutagenesis to alter the absorption and fluorescence properties of GFP has enabled the development of *in vivo* FRET between donor and acceptor GFP variants.

8
Flow Cytometry

In flow cytometry, individual cells flow past an illuminator (almost always a laser source) and a detector and a large number of quantitative determinations, based both on light scattering and fluorescence, can be carried out simultaneously. The cells may be rendered fluorescent by the use of specific fluorophores, such as fluorophore-linked antibodies or fluorophores, which can be targeted to specific biomolecules such as DNA or membranes. The fluorescence can be detected and a variety of information can be ascertained. In addition to simply the presence or absence of a signal, modern instrumentation may carry out multiple, sophisticated fluorescence determinations, including time-resolved measurements. Analysis rates as high as 100 000 cells per second have been achieved. A wide variety of biochemical, immunological, and clinical applications of flow cytometry has been established. Quantitative determination of cellular DNA is one important application of the method since one can obtain a distribution histogram of the DNA content of thousands of cells, which offers information on both the cell cycle stage as well as the presence of aneuploid cells. Immunophenotyping is another important application. In addition to simple fluorescence determinations, a flow cytometry may be used to sort and separate cells – in this case, the approach is often referred to as *fluorescence-activated cell*

sorting (*FACS*). Operationally, an FACS instrument is essentially a specialized fluorescence microscope, which utilizes electrostatic deflection of charged droplets to sort cells. Living cells are aspirated from a sample and injected individually into a stream that eventually breaks down into droplets, which enter an illumination chamber. Here, a laser source impinges upon the droplet and if the cell has the appropriate fluorescence properties (arising, for example, from fluorophore-tagged antibodies, fluorophores targeted to DNA- or GFP-related probes), a charge will be applied to the droplet causing it to be deflected by high-voltage plates to a collection area. In this way, cell populations can be sorted and quantified.

9
FCS

Fluorescence correlation spectroscopy (FCS – sometimes referred to as fluctuation correlation spectroscopy) takes advantage of the fact that – at very low concentrations and in very small observation volumes – the number of molecules resident at any given time in the observation volume will fluctuate as molecules enter or leave the volume. By analyzing the temporal properties of this fluctuation, one can deduce various molecular properties, for example, the diffusion coefficient, which in turn can be related to the size and shape of the molecule and the viscosity of the environment (e.g. cell membranes or cytoplasm). This information, in turn, can often be related to the state of aggregation of the target molecule, both as regards self-association and association with other biomolecules (e.g. proteins, membranes, or nucleic acids). In modern instruments, the sample consisting of a fluorescence probe associated with the target molecule is placed in a fluorescent confocal microscope and is illuminated with a CW (continuous wave) laser. In some cases, an ultrafast laser pulse characterized by a high repetition rate and high peak power is used to generate multiphoton excitation (*vide supra*), which allows excitation in a very small sample volume (typically on the order of femtoliters). FCS can be applied both *in vitro* and *in vivo*.

10
FISH

Fluorescence *in situ* hybridization (FISH) is a method used to localize and quantify specific nucleic acid sequences. Typically, fluorescence probes are incorporated into short sequences of single- or double-stranded DNA or RNA, either by direct labeling of individual nucleotides followed by polymerization or by direct labeling of the nucleic acid polymers. These probes can then be introduced into a cell where they will bind to complementary nucleic acid sequences (in the case of double-stranded probes, a denaturation and hybridization step may be included). The location of these sequences in the cell can then be determined by high-resolution fluorescence microscopy. In addition to probes for specific short nucleic acid sequences, one can utilize libraries of probes to "paint" an entire chromosome and hence use the method for spectral karyotyping.

Acknowledgments

The author thanks Dr. Ulises Acuna for information on Monardes' work.

Bibliography

Books and Reviews

Bernard, V. (Ed.) (2002) *Molecular Fluorescence*, Wiley-VCH, Weinheim, Germany.

Brand, L., Johnson, M.L. (Eds.) (1997) Methods in Enzymology, *Fluorescence Spectroscopy*, Vol. 278, Academic Press, San Diego.

Dewey, G. (Ed.) *Biophysical and Biochemical Aspects of Fluorescence Spectroscopy*, Plenum Press, New York.

Joseph, L. (Ed.) (1999) *Principles of Fluorescence Spectroscopy*, Kluwer Academic/Plenum Publishers.

Kraayenhof, R., Visser, A.J.W.G., Gerritsen, H.C. (Eds.) (2002) *Fluorescence Spectroscopy, Imaging and Probes: New Tools in Chemical, Physical and Life Sciences*, Springer-Verlag, Berlin.

Lakowicz, J. (Ed.) *Topics in Fluorescence Spectroscopy*, Vols. 1–6, Plenum Press, New York.

Marriott, G., Parker, I. (Eds.) (2002) Methods in Enzymology, *Biophotonics*, Vols. 360 & 361, Academic Press, San Diego.

Valeur, B., Brochon, J.-C. (Eds.) (2001) *New Trends in Fluorescence Spectroscopy: Applications to Chemical and Life Sciences*, Springer-Verlag, Berlin.

Van Der Meer, B.W., Coker, G., Chen, S.-Y.S. (Eds.) (1991) *Resonance Energy Transfer. Theory and Data*, Wiley-VCH, Weinheim, Germany.

Primary Literature

Bagatolli, L.A., Sanchez, S.A., Hazlett, T.L., Gratton, E. (2003) Giant vesicles, Laurdan, and two-photon fluorescence microscopy: evidence of lipid lateral separation in bilayers, *Methods Enzymol.* **360**, 481–500.

Barrio, J.R., Tolman, G.L., Leonard, N.J., Spencer, R.D., Weber, G. (1973) Flavin 1, N^6-ethenoadenine dinucleotide: dynamic and static quenching of fluorescence, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 941–943.

Belford, G.C., Belford, R.L., Weber, G. (1972) Dynamics of fluorescence polarization macromolecules, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1392–1393.

Berberan-Santos, M.N. (2001) Pioneering Contributions of Jean and Francis Perrin to Molecular Luminescence, in: Valeur, B., Brochon, J.-C. (Eds.) *New Trends in Fluorescence Spectroscopy*, Springer, Heidelberg, p. 7.

Brock, R., Jovin, T.M. (2003) Quantitative image analysis of cellular protein translocation induced by magnetic microspheres: application to the EGF receptor, *Cytometry* **52A**, 1–11.

Croney, J.C., Jameson, D.M., Learmonth, R. (2001) Teaching basic fluorescence principles with simple visual demonstrations, *J. Biochem. Ed.* **29**, 60–65.

Denk, W., Strickler, J.H., Webb, W.W. (1990) Two-photon laser scanning fluorescence microscopy, *Science* **248**, 73–76.

Eftink, M. (2000) Use of fluorescence spectroscopy as thermodynamics tool, *Methods Enzymol.* **323**, 459–473.

Ehrenberg, M., Rigler, R. (1972) Polarized fluorescence and rotational Brownian motion, *Chem. Phys. Lett.* **14**, 539–544.

French, T., So, P.T.C., Dong, C.Y., Berland, K.M., Gratton, E. (1997) Fluorescence Lifetime Imaging Techniques for Microscopy, in: Sluder, G., Wolf, D. (Eds.) *Methods in Cell Biology, Video Microscopy*, Vol. 56, Academic Press, New York, pp. 227–304.

Gadella, T.W.J., Clegg, R.M., Jovin, T.M. (1994) Fluorescence lifetime imaging microscopy: pixel by pixel analysis of phase-modulation data, *Bioimaging* **2**, 139–159.

Gautier, I., Tramier, M., Durieux, C., Coppey, J., Pansu, R.B., Nicolas, J.C., Kemnitz, K., Coppey-Moisan, M. (2001) Homo-FRET microscopy in living cells to measure monomer-dimer transition of GFT-tagged proteins, *Biophys. J.* **80**, 3000–3008.

Hanson, K.M., Behne, M.J., Barry, N.P., Mauro, T.M., Gratton, E., Clegg, R.M. (2002) Two-photon fluorescence lifetime imaging of the skin stratum corneum pH gradient, *Biophys. J.* **83**, 1682–1690.

Heikal, A.A., Hess, S.T., Baird, G.S., Tsien, R.Y., Webb, W.W. (2000) Molecular spectroscopy and dynamics of intrinsically fluorescent proteins: coral red (dsRed) and yellow (Citrine), *Proc. Natl. Acad. Sci. U.S.A.* **97**(22), 11996–12001.

Helms, M.K., Hazlett, T.L., Mizuguchi, H., Hasemann, C.A., Uyeda, K., Jameson, D.M. (1998) Site-directed mutants of rat testis fructose 6-phosphate, 2-kinase/fructose 2,6-bisphosphatase – localization of conformational alterations induced by ligand binding, *Biochemistry* **37**, 14057–14064.

Hess, S.T., Huang, S., Heikal, A.A., Webb, W.W. (2002) Biological and chemical applications of fluorescence correlation spectroscopy: a review, *Biochemistry* **41**, 697–705.

Hink, M.A., Bisselin, T., Visser, A.J. (2002) Imaging protein-protein interactions in living cells, *Plant Mol. Biol.* **50**, 871–883.

Hink, M.A., Borst, J.W., Visser, A.J. (2003) Fluorescence correlation spectroscopy of GFP fusion proteins in living plant cells, *Methods Enzymol.* **361**, 93–112.

Jameson, D.M. (2001) The Seminal Contributions of Gregorio Weber to Modern Fluorescence Spectroscopy, in: Valeur, B., Brochon, J.-C. (Eds.) *New Trends in Fluorescence Spectroscopy*, Springer, Heidelberg, p. 35.

Jameson, D.M., Croney, J.C. (2003) Fluorescence polarization: past, present and future, *Comb. Chem. High Throughput Screen.* **6**, 167–176.

Jameson, D.M., Croney, J.C., Moens, P.D. (2003) Fluorescence: basic concepts, practical aspects and some anecdotes, *Methods Enzymol.* **360**, 1–43.

Jameson, D.M., Eccliston, J.F. (1997) Fluorescent nucleotide analogs: synthesis and applications, *Methods Enzymol.* **278**, 363–390.

Jameson, D.M., Sawyer, W.H. (1995) Fluorescence anisotropy applied to biomolecular interactions, *Methods Enzymol.* **246**, 283–300.

Jameson, D.M., Seifried, S.E. (1999) Quantification of protein-protein interactions using fluorescence polarization, *Methods* **19**, 222–233.

Jasuja, R., Jameson, D.M., Nishijo, C.K., Larsen, R.W. (1997) Singlet excited state dynamics associated with nucleosides/nucleotides complexed with tetrakis(4-N-methyl)pyridyl)porphine, *J. Phys. Chem.* **101**, 1444–1450.

Lakowicz, J.R., Weber, G. (1973) Quenching of fluorescence by oxygen. A probe for structural fluctuations in macromolecules, *Biochemistry* **12**, 4161–4170.

Larsen, R.W., Jasuja, R., Hetzler, R.K., Muraoka, P.T., Andrada, V.G., Jameson, D.M. (1996) Spectroscopic and molecular modeling studies of caffeine complexes with DNA-intercalators, *Biophys. J.* **70**, 443–452.

Lillo, M.P., Beechem, J.M., Szpikowska, B.K., Sherman, M.A., Mas, M.T. (1997) Design and characterization of a multisite fluorescence energy-transfer system for protein folding studies: a steady-state and time-resolved study of yeast phosphoglycerate kinase, *Biochemistry* **36**, 11261–11272.

Lillo, M.P., Canadas, O., Dale, R.E., Acuna, A.U. (2002) Location and properties of the taxol binding center in microtubules: a picosecond laser study with fluorescent taxoids, *Biochemistry* **41**, 12436–12449.

Muller, J.D., Chen, Y., Gratton, E. (2000) Resolving heterogeneity on the single molecular level with the photon-counting histogram, *Biophys. J.* **78**, 474–486.

Muller, J.D., Chen, Y., Gratton, E. (2003) Fluorescence correlation spectroscopy, *Methods Enzymol.* **361**, 69–92.

Patton, W.F., Beechem, J.M. (2002) Rainbow's end: the quest for multiplexed fluorescence quantitative analysis in proteomics, *Curr. Opin. Chem. Biol.* **6**, 63–69.

Perrin, F. (1926) Polarisation de la lumière de fluorescence. Vie moyenne des molécules dans l'état excité, *Jour de Phys, VIe Série* **7**, 390–401.

Piston, D.W., Knobel, S.M. (1999) Quantitative imaging of metabolism by two-photon excitation microscopy, *Methods Enzymol.* **307**, 351–368.

Rocheleau, J.V., Edidin, M., Piston, D.W. (2003) Intrasequence GFP in Class I MHC molecules, a rigid probe for fluorescence anisotropy measurements of the membrane environment, *Biophys. J.* **8**, 4078–4086.

Subramaniam, V., Hanley, Q.S., Clayton, A.H., Jovin, T.M. (2003) Photophysics of green and red fluorescent proteins: implications for quantitative microscopy, *Methods Enzymol.* **360**, 178–201.

Weber, G. (1952) Polarization of the fluorescence of macromolecules. I. Theory and experimental method, *Biochem. J.* **51**, 145–155.

Weber, G. (1952) Polarization of the fluorescence of macromolecules. II. Fluorescent conjugates of ovalbumin and bovine serum albumin, *Biochem. J.* **51**, 155–167.

Williams, R.M., Zipfel, W.R., Webb, W.W. (2001) Multiphoton microscopy in biological research, *Curr. Opin. Chem. Biol.* **5**, 603–608.

Wu, M.M., Llopis, J., Adams, S.R., McCaffery, J.M., Teter, K., Kulomaa, M.S., Machen, T.E., Moore, H.P., Tsien, R.Y. (2000) Studying organelle physiology with fusion protein-targeted avidin and fluorescent

biotin conjugates, *Methods Enzymol.* **327**, 546–564.
Yan, Y., Marriott, G. (2003) Fluorescence resonance energy transfer imaging microscopy and fluorescence polarization imaging microscopy, *Methods Enzymol.* **360**, 561–580.
Zipfel, W.R., Williams, R.M., Christie, R., Nikitin, A.Y., Hyman, B.T., Webb, W.W. (2003) Live tissue intrinsic emission microscopy using multiphoton-excited native fluorescence and second harmonic generation, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7075–7080.

Fluorescence Activated Cell Sorting (FACS): *see* Flow Cytometry

Fluorescence Activated Chromosome Sorting: *see* Gene Mapping and Chromosome Evolution by Fluorescence–Activated Chromosome Sorting

Food Proteins: Interactions and Functionality

Charles Onwulata and Phoebe X. Qi
Eastern Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, Wyndmoor, PA, USA

1 Protein Structure and Conformation 506
1.1 Food Proteins 508
1.2 Enzymes 508
1.3 Energy and Nutritive Value 509

2 Protein Interactions 509
2.1 Protein–protein 510
2.2 Carbohydrates 510
2.3 Lipids 511
2.4 Hydration 512
2.5 Soluble Ions 512

3 Functional Protein Interactions 512
3.1 Gelation and Rheology 512
3.2 Emulsification 513
3.3 Foaming 514
3.4 Thermal Processing 514
3.5 Thermally Induced Mutagens 515

4 Food Protein Functionality 515
4.1 Partially Folded Proteins 515
4.2 Properties of the Molten Globule State 516
4.3 Functional Applications 516

Mention of brand or firm name does not constitute an endorsement by the U.S. Department of Agriculture above others of a similar nature not mentioned.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Bibliography 517
Primary Literature 517

Keywords

Emulsion
A thermodynamically unstable dispersion of micellar particles or globules in a liquid medium.

Foaming
A coarse dispersion of a gas in a protein matrix in which the bulk of the phase volume is the gas, with the matrix encasing the gas bubbles in thin sheets called *lamellae*, which are formed by agitating the protein matrix and gas together in the presence of a stabilizing agent.

Food Protein Interaction
The interaction of protein molecules and other compounds within their domain, which affects the proteins' behavior in food products.

Gelation
The formation of an ordered, continuous matrix entrapping another component, in which attractive and repulsive forces are balanced.

Hydrogen Bond
A highly directional bond in which a hydrogen atom is shared by two other atoms. One of the atoms is frequently covalently linked to that hydrogen atom.

Hydrophobic Interactions
Nonpolar molecules or groups tend to associate with one another in water and hold strong preference to nonpolar environments.

Phospholipids
Membrane lipids that possess polar groups linked to a diglyceride by a phosphodiester bridge.

Protein Micelle
Protein particles formed by colloidal association of protein monomers. The interaction is largely driven by the decrease in free energy resulting from hydrophobic attractions of nonpolar portions of the molecules.

Rheology
Normally refers to the flow and deformation of "nonclassical" materials such as rubber, molten plastics, polymer solutions, slurries and pastes, electrorheological fluids, blood,

muscle, composites, soils, and paints. These materials can exhibit varied and striking rheological properties that classical fluid mechanics and elasticity cannot describe.

Syneresis
Shrinking of the gel with expulsion of trapped liquid.

van der Waals Force
A nonspecific attractive force between atoms or molecules.

Proteins are the most abundant macromolecules found in living cells and approximate half of the cell's dry weight. They are required in the food of humans, fish, and most higher animals. Historically, food proteins have been selected for their nutritional value and can be obtained from a wide variety of naturally occurring sources. Proteins undergo a wide range of structural and conformational changes through a variety of complex interactions during processing and storage. Such changes can affect the principal purpose of dietary proteins, which is to supply nitrogen and amino acids for the synthesis of proteins in the body. It is through an understanding of these interactions and their effects on functionality that food proteins have played a major role in the food supply.

The macrostructure of a protein is primarily determined by its amino acid sequence. Amino acids are essential in basic nutrition, growth, and maintenance. Nine of the 20 identified amino acids, namely, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine, are considered essential. They are so classified because they cannot be synthesized by humans and must therefore be supplied in the diet. When essential amino acids are not supplied in sufficient quantity, they restrict the adequate utilization of other amino acids and are thus termed *limiting amino acids*.

The processing of food proteins is designed to reduce microbial and enzymatic spoilage, inactivate antinutritional substances, improve the availability of perishable foods, and enhance the sensory quality of the food. Chemical changes that occur during food processing can have a significant impact on the nutritional value, the sensory properties, as well as the general functionality of the food. Heat treatment of proteins under acidic conditions can result in denaturation or even total unfolding of proteins and inactivation of enzymes and antinutritional substances. More severe heating can cause proteins and/or enzymes to cross-link with one another. Process-induced chemical interaction can also occur between proteins and carbohydrates, which leads to nonoxidative, nonenzymatic browning (the Maillard reaction). Lipids, especially when unsaturated, are susceptible to oxidation. They can form lipid peroxides, which interact with proteins to yield lipid–protein complexes, thus decreasing the nutritional value of the food.

In addition to the above-mentioned purposes of food protein processing, proteins play an important role in improving the functionality of food ingredients. For instance, heat denaturation of whey proteins before spray-drying improves their

water-holding properties. Alkali treatment of soy proteins is used to improve their solubility and textural properties, and to obtain desirable rheological properties. Viscosity and solubility measurements are commonly used to obtain information about the functional behavior and physicochemical nature of proteins.

1 Protein Structure and Conformation

Proteins are natural compounds composed of amino acids organized at four different levels of structure: primary, secondary, tertiary, and quaternary. The primary structure consists of amino acids that are sequenced in a linear polypeptide chain and constitute the basic building blocks. Amino acids are joined together through an amide linkage or peptide bond, which is formed through the removal of a hydroxyl group from the carboxyl group of one amino acid and a hydrogen atom from the α-amino group of the adjoining amino acid. The C−N bond of the peptide linkage exhibits partial double-bond character and is not free to rotate, which imposes some constraints on the number of conformations that the polypeptide can assume (Fig. 1). There is, however, a large degree of rotational freedom around two single bonds $N-C_\alpha$ (φ) and $C_\alpha-C$ (ι) of the polypeptide.

The secondary structure of the polypeptide of the protein refers to the spatial arrangement of amino acid residues that are near one another in the amino acid sequence. It is frequently composed of α-helices, β-pleated sheets, and random coils, such as β-lactoglobulin (see Fig. 2), one of the major milk proteins. The α-helix structure exists in polypeptides that are tightly coiled into a rodlike structure with side chains extending outward from the helix. It is stabilized by strong hydrogen bonds between the NH and CO groups of the main chain. In contrast, a polypeptide chain in the β-pleated-sheet structure is almost fully extended. The β-pleated sheet is stabilized by hydrogen bonds between NH and CO groups in different peptide chains. The arrangement of member strands can be parallel, antiparallel, or a mixture of the two. Many proteins have various well-defined combinations of β-strands and α-helices. Some amino acids cannot form α-helices because of electrostatic repulsion or bulky side chains. As a result, the polypeptide assumes a random-coil structure having minimal electrostatic free energy. In general, regions of α-helices and β-sheets in the protein are well defined; random coils are not.

The tertiary structure refers to the steric relationships of amino acid residues that are far apart in the linear sequence and the overall architecture in three dimensions of the polypeptide chain whose folding brings into proximity parts of the molecule otherwise widely separated along the backbone. Proteins containing more than one polypeptide chain exhibit another level of organization. This quaternary structure describes the aggregation of the separate polypeptide chains of a protein, which is often stabilized by hydrogen bonds and van der Waals forces. Structural changes at this level have unique biological applications that are often mediated by enzymes. It is through systematic denaturation of the organized structures and forces of

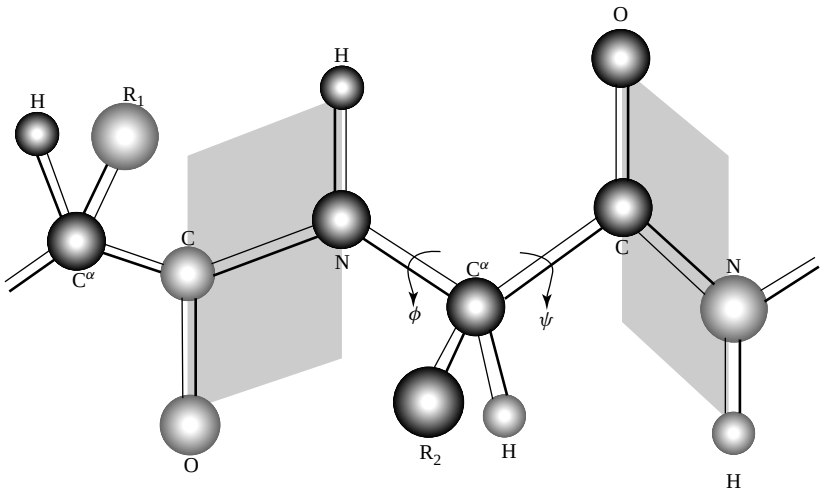

Fig. 1 Polypeptide chain linking two peptide units. The terms φ_ and ι refer to rotation about the α-carbon (C^{α}). The two peptide bonds adjacent to C^{α} are located in different planes.

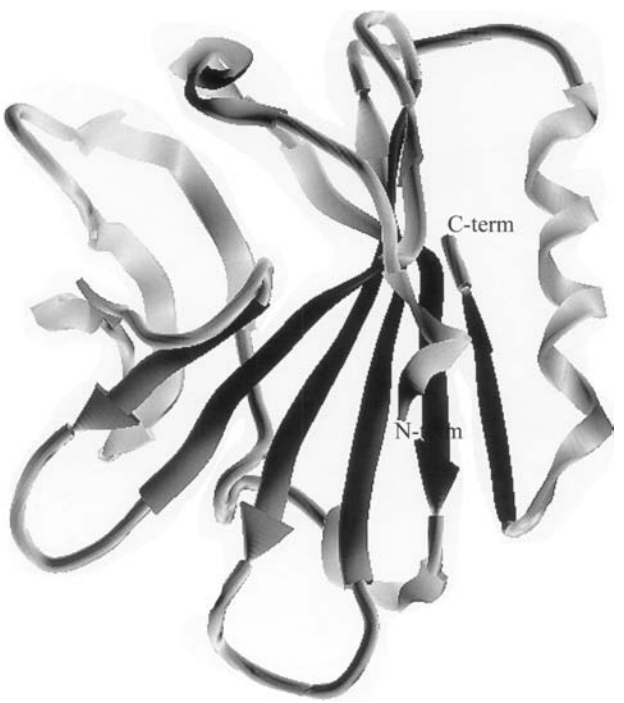

Fig. 2 Three-dimensional crystal structure of bovine β-lactoglobulin (PDB code 1B8E). Helices and β-sheet strands are shown in ribbons and random coil is shown in tubes. The arrows on the ribbons represent the direction of β-sheet strands.

food proteins that changes in functional behavior can be achieved.

1.1 Food Proteins

Food proteins are derived from a number of sources: plants, meat and fish, milk, eggs, and microbial proteins from unicellular or multicellular organisms, for example, bacteria, yeast, molds, and algae. Plant proteins are broadly classified according to their solubility, shape, prosthetic group, regulatory properties, as well as biological activity. They were first classified on the basis of their solubility as albumins, globulins, glutenins, and prolamines. Albumins are the most water-soluble globular proteins. The next most soluble proteins are globulins such as conglycinin (7S) and glycinin (11S) from soybean, which are soluble in dilute salt solutions at neutral pH. Glutenins such as wheat glutenins are soluble in dilute acid or alkali. Prolamines are corn or wheat storage proteins and are soluble in 70% ethanol.

Muscle or contractile proteins are derived directly from animal tissue and are the most conspicuous food proteins in the human diet. Muscle proteins are generally classified into three groups on the basis of their solubility in water: sarcoplasmic proteins, which are soluble in water or dilute salt; myofibrillar proteins, which are soluble in salt solution >0.6 M; and stromal proteins, which are the least soluble class of muscle proteins. Fibrous proteins, for example, myofibrillar proteins such as actin and myosin, or stromal proteins, such as collagen and elastin, have polypeptide chains arranged in long strands and serve a structural or protective role. Collagen, the major protein of connective tissue, occurs in several polymorphic forms consisting of three intertwined polypeptide chains.

Egg proteins are primarily globular proteins found in the albumen (egg white) and include ovalbumin, conalbumin, ovotransferrin, ovomucoid, and lysozyme. The egg yolk contains a variety of lipoproteins that act as emulsifying agents, as well as livetin (possibly derived from hen's blood) and phosvitin, an iron carrier.

Milk proteins consist of a colloidal dispersion of casein micelles and soluble whey proteins; their stability is of tremendous practical importance in the dairy industry. Caseins constitute nearly 80% of the total milk proteins. Casein micelles serve to fluidize proteins and solubilize the calcium and phosphate. They are extremely sensitive to changes in ionic environment and readily aggregate with increased concentration of calcium and magnesium ion or below the isoelectric point (pI). β-Lactoglobulin, α-lactalbumin, and lactoferrin are the three most abundant whey proteins in milk.

Other food proteins classified according to their prosthetic groups (tightly associated non-amino acid moieties) are lipoproteins, glycoproteins, and myoglobins (including hemoglobin) with attached fat or oil, sugar, or iron groups.

Proteins derived from unicellular organisms are grown on food processing by-products from which the protein is harvested and subsequently purified.

1.2 Enzymes

Enzymes are proteins that function as specific biological catalysts for chemical reactions in living systems. Enzymes are much more efficient and specific than other catalysts and operate in restricted conditions of temperature and pH. They

are frequently used in the food industry to modify the functional behavior of food proteins. For instance, proteases, which hydrolyze peptide bonds, can be recovered from plant sources such as papain and are used as commercial meat tenderizers. It is important, however, that the nutritive value of the product not be significantly reduced by the hydrolysis of the peptide bond of the protein catalyzed by the enzyme.

Enzymatic action on free amino acids is of significance in food spoilage and production of flavors in fermented foods. For example, enzymatic degradation of free amino acids by microorganisms of the genera *Staphylococcus, Pseudomonas,* and *Micrococcus* produces specific enzymes that are responsible for the degradation reactions. Enzymes are also used in the manufacture of beer, bread, cheese, coffee, vinegar, vitamins, and many other products.

1.3
Energy and Nutritive Value

The nutritional quality of proteins is essentially determined by their amino acid composition. The nutritional value of a protein for human consumption is based on the content of essential amino acids compared with human requirements and on the protein efficiency ratio (PER), which is determined by the weight gain of a rat divided by the protein intake. Nutrient availability and quality of food proteins depend on the processing treatments and interaction with other food components. Protein requirements for maintaining good health and growth are listed as the recommended dietary allowance (RDA) set by the US National Academy of Sciences. The RDA for a certain population varies depending on such factors as age, sex, level of activity, and individual metabolic factors. Protein requirements to meet growth and maintenance range from 2.0 g kg^{-1} body weight for infants to 0.8 g kg^{-1} body weight for adults. Energy needs increase during pregnancy, lactation, as well as during work, stress, sickness, and aging. Proteins differ in nutrient value because of differences in amino acid composition. Protein quality is measured by chemical scores and biological assays that assess such qualities as content of essential amino acids, nitrogen value, limiting amino acids, and efficiency ratio. Some sources of animal proteins include cheese, eggs, fish, meat, and milk. Vegetable proteins can be found in beans, grains, nuts, and a variety of vegetables.

2
Protein Interactions

Molecular forces governing protein interactions determine the relationship of the structure of individual proteins to their functional properties as well as the association of a protein with other compounds in the cell. These forces include covalent bonds, ionic interactions, hydrogen bonding, hydrophobic interactions, hydration, and steric repulsion. Covalent bonds include all the bonds of the primary structure of the protein as well as disulfide bonds, which are formed between cysteine residues and are dependent on the conformation and structure of the peptide chain. Of the covalent forces, disulfide bonding is the most important in protein interactions. Ionic interactions are between electrostatic charges and are the most fundamental noncovalent attractions in intermolecular forces. Noncovalent molecular forces are one to three orders of magnitude smaller than covalent bonding energy and include

hydrogen bonds, hydrophobic interaction, and repulsion forces. Protein denaturation involves the interruption of these noncovalent forces and results in changes in the secondary and tertiary structures. Proteolysis by appropriate enzymes or hydrolysis with strong acid or base is necessary to break covalently linked peptide bonds. Association and disassociation of the protein in secondary and tertiary structure maintain the integrity of food proteins.

2.1 Protein–protein

Protein–protein interactions often occur as a result of food processing that is designed to improve the functional properties of proteins for new product application. These interactions occur in two-stage processes consisting of unfolding of the native protein and exposing active sites, followed by association of the polypeptide chain through covalent and noncovalent forces. Figure 3 represents a molecular dynamics–simulated process of protein unfolding. Native bovine α-lactalbumin is a globular protein that undergoes gradual unfolding as a function of temperature. The final unfolded state maintains little secondary and tertiary structures with increased entropy relative to its native state. Stable protein interactions are established through cross-linking, ionic, hydrogen bonding, and hydrophobic interactions. Gelation is an association of proteins existing in a three-dimensional network with trapped water molecules. Protein cross-links are formed through formation of disulfide bonds by sulfhydryl groups or through hydrophobic interactions. To maintain a stable gel, a balance must exist between forces that promote formation of the network and those that oppose the formation.

2.2 Carbohydrates

Proteins and carbohydrates form irreversible complexes in nonoxidation reactions. Food protein heated in the presence of a reducing sugar results in the reaction

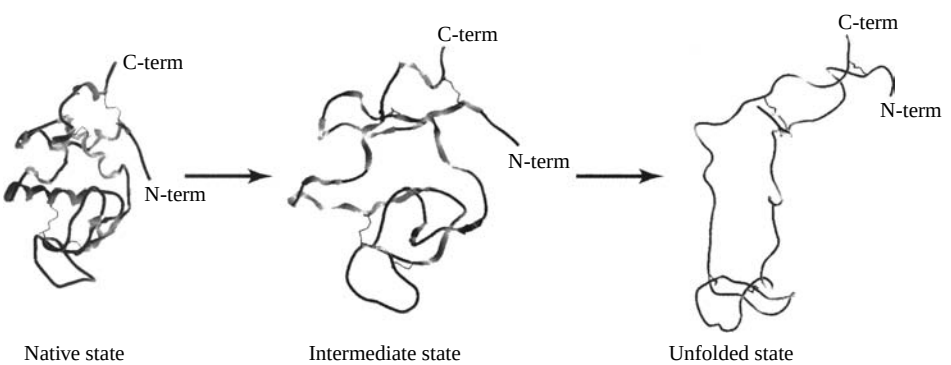

Fig. 3 Molecular dynamics simulation of the two-step unfolding process of bovine apo-α-lactalbumin in vacuum. The native state is taken from the X-ray crystal structure (PDB code 1HFZ) without Ca^{2+}, the intermediate state is a snapshot of a conformation during the simulation, and the unfolded state is achieved by heating the native molecule that reached equilibrium. Four disulfide bonds formed between the cysteine residues greatly stabilize the molecule.

of the carbonyl groups of the carbohydrate with the free amino group of the protein, followed by a cascade of reactions leading to polymers. These Maillard reactions are the major cause of protein damage in the drying of milk at moderate temperatures. In the early stages of the Maillard reaction (Fig. 4), the carbonyl of the reducing sugar undergoes nucleophilic attack by the amine, followed by loss of water and ring closure to form the glycosylamine. The Amadori rearrangement follows with the formation of 1-amino-2-keto sugar, which has been isolated and identified in browned dried fruit. However, more severe heating, required in the preparation of toasted breakfast cereals, bread, and biscuits, results in late Maillard-type damage and the formation of other undesirable by-products. One such by-product is acrylamide, a known animal carcinogen and human neurotoxin. Acrylamide is formed when the amino acid asparagine reacts with simple sugars at high temperature, as in fried or baked foods.

2.3 Lipids

Protein–lipid interactions in nature occur in food systems such as milk and eggs. These interactions occur most notably at the cellular and intracellular membrane level. Proteins that stabilize emulsions can adopt conformations that are capable of interacting with both lipid and aqueous phases by assuming the lowest free-energy form at the interface. The process can be represented in two steps: diffusion-controlled sorption at the interface, followed by protein unfolding. Emulsions are stabilized by hydrophobic interactions between the apolar region of the protein and the apolar aliphatic chain of the lipid. The importance of hydrophobic interactions has been convincingly demonstrated

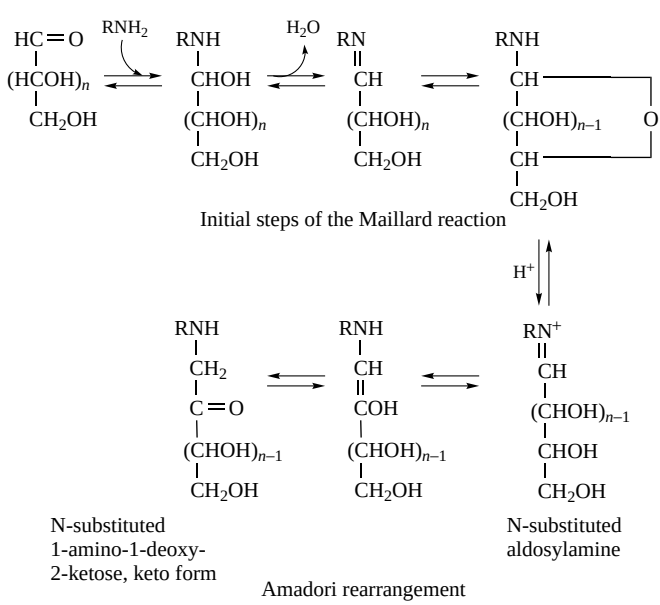

Fig. 4 Schematic representations of the Maillard reactions and the Amadori rearrangement.

in model systems in which the energy of protein–lipid interaction reaches a maximum around the pI of the protein. Protein–lipid interaction can be increased by high-pressure homogenization, which increases the number of lipid droplets and the interfacial surface area. Lipids protect proteins against thermal denaturation because of their high heat capacity and the absence of water. Strong interactions occur in flour–water mixtures, in which lipids can bind gluten proteins to form highly stable lipoglutenin complexes.

2.4 Hydration

Interactions of protein side chains with water determine the intrinsic properties of a protein, such as solubility, swelling, dispersibility, and wettability. Water absorption is considered as the most important step to imparting desired functional properties to proteins. Proteins interact with water through their side chains or backbone, and their solubility depends on factors such as hydrogen bonding, dipole–dipole and ionic interactions, pH, and temperature. Their ability to absorb and retain water plays a major role in the textural stability of food systems. Texturized proteins may be used to formulate meat or seafood analogues by taking advantage of their hydration properties.

2.5 Soluble Ions

Sodium, potassium, and calcium are the most important ionic components associated with solubility of proteins. The reactive nature of proteins allows for the manipulation of the ions to accomplish food preservation by, for example, controlled dehydration. Proteins are stable within a defined range of pH and may have either net positive or negative charges or may be neutral. At pH values above or below their pI, proteins interact more readily with water molecules to improve solubility. Divalent cations also act as the cofactors that are essential for some enzymes to be catalytically active. Salts used in food processing can denature proteins by charge neutralization. Severe denaturation results in irreversible realignment of the native structure of the protein.

3 Functional Protein Interactions

Physicochemical properties that enable proteins to affect the characteristics of foods during processing, storage, preparation, and consumption define the functional properties of proteins. Protein functions that affect food utilization are (1) water absorption – viscosity and gelation; (2) surface activity – gelation and foaming; and (3) chemical reactivity – textural properties. The functionality of a food product is determined experimentally because, while the study of protein structure provides information on physicochemical properties, the exact relationship between structure and function has not been fully understood. Changes in functionality are the result of protein interaction with water, other proteins, or the result of changes in the surface characteristics of the protein molecule.

3.1 Gelation and Rheology

Gelation is the formation of an extended network of denatured protein aggregates held together by intermolecular forces.

Gels can result from controlled protein denaturation and unfolding (egg white); controlled folding and realignment (collagen); controlled unfolding, disulfide interchange, and enhanced hydrophobic interactions (gluten); and enzyme hydrolysis and hydrophobic interactions (rennet milk gels). The stability of gels largely depends on the balance of hydrophobic interactions, hydrogen bonding, and electrostatic forces, which, in turn, are affected by pH and electrolyte content. Cross-linking of proteins in the matrix via disulfide bonds leads to the formation of heat-irreversible gels that are subsequently stabilized by hydrogen bonds. Higher–molecular weight proteins, with a larger percentage of hydrophobic amino acid residues, tend to form stronger gels.

Rheology describes how materials flow and deform during exposure to various force applications. Many factors affect the ability to flow, including concentration, temperature, and microstructure. Dairy-based foods, including chocolate milk, puddings, and processed cheeses, rely on protein–hydrocolloid interactions to impart stability and enhance the texture of the food system.

3.2 Emulsification

A stable protein emulsion is produced when proteins in a two-phase medium, accompanied by energy input (usually a combination of shear and heat), unfold slightly at the interface and align their nonpolar regions toward an oil phase with their hydrophilic regions toward the aqueous phase. The balance of the hydrophilic and hydrophobic forces in the protein maintains emulsion stability. Globular proteins with a highly ordered and stable tertiary structure, such as β-lactoglobulin, bovine serum albumin, and lysozyme are more likely to unfold and are considered to be good emulsifiers. Emulsification activity continues to increase with increasing degree of protein denaturation, provided solubility is not compromised. Surface hydrophobicity of proteins has been directly correlated with emulsifying activity. Emulsions are mechanically unstable and will separate over time. The breakdown in stability results in flocculation followed by coalescence of the particles by forces, as depicted in Fig. 5. Although emulsifiers can minimize interfacial surface energy and reduce

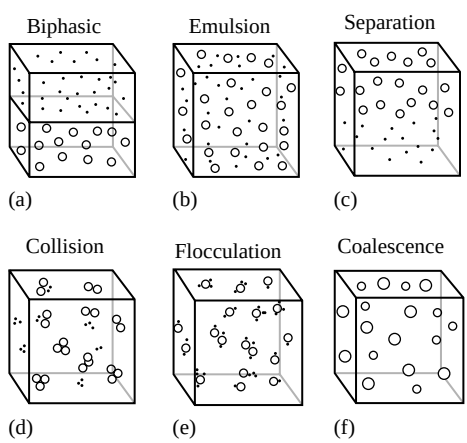

Fig. 5 Process of emulsion formation and breakdown. (a) Biphasic medium of liquid and semisolid emulsion, (b) thermodynamically unstable emulsion, (c) gradual migration of emulsion based on charge separation, (d) particle collision with increasing electrostatic charges, (e) clear aggregation and flocculation of collided particles, and (f) coalescence and syneresis of emulsion.

coalescence, emulsions gradually separate into two phases as the repulsive electrical charges decrease. Gradual increase in free energy leads to more frequent collisions of similarly charged particles. These collisions produce flocculated globules or clusters, which will continue to grow unless conditions change. Complete breakdown of the emulsion is the result of coalescence of these particles.

3.3 Foaming

Protein foams are emulsions of gas in a continuous aqueous protein phase containing various surfactants to prevent coalescence. Foods such as ice cream and whipped toppings are examples of stable foams formed by incorporation of a large volume of air during processing. Air is forced into the medium through mechanical mixing that increases the volume over 1,000%. Foam stability is maintained by surfactants that resist the influence of gravity, pressure, and temperature, which tend to rupture the stable emulsion. Controlled denaturation of the native protein by shear, temperature, or manipulation of electrostatic balance increases hydrophobicity and improves foaming. Foam-stabilizing conditions include high surface hydrophobicity, low net charge, minimal electrostatic repulsion, good solubility, and reasonable protein concentration. Foam expansion (overrun) is controlled by metering the mix and air into the equipment in proper proportions to form air cells of the desired size; overrun is computed as the percentage increase in the volume or the percentage decrease in the density of the mix. Foam stability may be controlled by addition of combinations of stabilizers and emulsifiers, such as salts, sugars, lipids, cations, proteins, and energy input (heat and shear) during processing. As foams age during storage, they shrink, leading to loss of air and separation.

3.4 Thermal Processing

Food proteins are thermally processed to enhance functionality, improve textural properties, or retard deterioration. In the preservation of fruits and vegetables (blanching), heat is used to denature enzymes. Processing may lead to a slight loss in nutritional quality. Most processes, however, improve quality by destroying antinutritional factors through inactivation of enzymes such as trypsin inhibitor or peroxidase. Complete denaturation of peroxidase (EC.1.11.1.7) is used as an indicator of adequate processing. Protein–protein interactions may be enhanced, leading to formation of new complexes. Factors such as temperature, the presence of salts and oxidizing/reducing agents may be used to produce desirable high-quality foods. Chilling or heating can denature proteins. Protein structure may be altered as a result of changes in the electric potential, pH, and so on, and such changes can be used to enhance functionality in food. Thermal denaturation of food proteins generally occurs between 45 and 85 °C, accompanied by exposure of hydrophobic groups that can lead to protein aggregation. More severe treatments can result in the splitting of the disulfide cross-links with release of hydrogen sulfide as well as alteration of amino acid residues with the formation of new intra- or intermolecular covalent cross-links. Low-temperature denaturation is mediated by a reduction of hydrophobic interactions

in conjunction with enhanced hydrogen bonding, leading to aggregation and precipitation of proteins.

3.5
Thermally Induced Mutagens

Mutagens are substances that cause heritable genetic changes (mutations). Most mutations are harmful and most mutagens are carcinogens and vice versa. Heterocyclic aromatic amines (HAAs) that are formed during broiling, grilling, roasting, or frying of meat and fish have been found to have a tumorigenic effect in various organs in mice, rats, and nonhuman primates, and many of them are therefore considered as possible human carcinogens. More than a dozen highly mutagenic HAAs have been identified; the most common forms are amino-imidazo-azaarenes (imidazoquinolines, imidazoquinoxalines, and imidazopyridines) and carbolines (pyridoindoles and pyridoimidazoles). The Millard reaction has been suggested to be a pathway for the formation of the imidazoquinoline or imidazoquinoxaline type of heterocyclic amines. Creatine and creatinine, free amino acids, and sugars have been shown as precursors. The same precursors may form the imidazopyridines. Aminocarbolines may be generated by the pyrolysis of amino acids and proteins. Precursors and tentative reaction routes for their formation have been identified from model experiments. It has been shown that when fried, different sorts of meat form different types and amounts of HAAs. Cooking conditions (temperature) and the amounts of different precursors may influence the formation and level of HAAs in food products. Extensive research has been carried out to develop methods to reduce or eliminate such compounds from the food supply, with moderate success.

4
Food Protein Functionality

Food proteins with special functional properties are manufactured by manipulating changes in the native secondary and tertiary conformations. Intermediate as well as stable forms of conformations may be produced at various stages of processing (Fig. 3). Food products with different functionality are created because of the large range of structures and their interaction with one another. This complex variety of conformations displays different properties such as water sorption, holding, and sorption, which are important functional attributes. The physical ability to entrap and hold water and lipid are essential for foaming. The ability of proteins to denature and form cohesive viscous matrices is what gives rise to their emulsifying and film-forming properties. The structure of the matrix formed depends on the nature of the protein; for instance, plasma protein gels form a "close" dense network and egg albumin gels form an "open" structure. Although the question of how a protein folds into a unique shape with specific functions is far from being solved, it is generally accepted that the primary amino acid sequence encodes its folding pathway, which in turn determines its functionality.

4.1
Partially Folded Proteins

The concept of the molten globule has been used to describe a partially folded equilibrium state between native and unfolded states. This intermediate has been found for many proteins in mild denaturing conditions, such as low or high pH, high temperature, or moderate concentrations of chemical denaturants (urea and guanidinium hydrochloride). The molten

globule is almost as compact as the native state and has a loosely packed hydrophobic core. It possesses pronounced secondary structure but little rigid tertiary structure and marked loss of enzymatic activity relative to its native state. It has long been established that changing the food protein structure can result in the alteration of food functionality. Thus, our basic understanding of the protein-folding process and the intermediates involved can directly lead to not only fundamental understanding of processing treatment but also the discovery and utilization of valuable food ingredients.

4.2 Properties of the Molten Globule State

The recently emerged energy-landscape theory and funnel concept for protein folding have provided us with a quantitative and pictorial view of the folding process. They revealed that multistep mechanisms that involve heterogeneous populations of molten globule–like intermediates may be more common than we previously realized. Many food proteins have been found to exist as the molten globule states under various conditions. α-Lactalbumin, at acidic pH, elevated temperature, or in the absence of calcium, is a paradigm of the classical molten globule state. Moreover, the number of proteins that has been found to possess structural features of the molten globule state as well as remain biologically functional is increasing drastically.

The typical molten globule state is more open than the native sate, but still represents a compact structure. The dynamic and flexible characteristics of the molten globule state allow water molecules to penetrate the protein more easily. The molten globule state is thus highly hydrated. In addition, the loosely packed side chains may emphasize the importance of solvent-separated hydrophobic interactions in the molten globule state.

In contrast with the association process in native proteins in which multimeric complexes are often formed, protein–protein interaction in the molten globule states may lead to aggregation. Recent research on human α-lactalbumin has suggested that polymeric forms of the protein may serve as antitumor and antibacterial agents.

4.3 Functional Applications

The knowledge that a single whey protein such as α-lactalbumin may have a myriad stable intermediate stages has become the basis for many new functional food ingredients derived from whey. Processing technologies have the potential to create many products from native, partially folded, or completely unfolded proteins.

Food proteins such as whey proteins can be modified using chemicals, heat, or shear in extrusion processes. Chemical treatment alone alters the reactive groups of the amino acids, resulting in not only selected changes in covalent bonding but also in the noncovalent forces that influence conformation such as van der Waals forces, electrostatic interactions, hydrophobic interactions, and hydrogen bonding. Heat and shear, on the other hand, alter the conformational structure of the protein through some degree of denaturation of the protein, exposing hydrophobic groups that are normally buried in the interior of the native protein. The changes caused by denaturation can be either reversible or irreversible. As the whey proteins denature, they become insoluble and show a great tendency to

aggregate. Heat-denatured whey proteins produce different fractions depending on the pH; for instance, whey proteins heated at 90 °C for 30 min at pH 2.5 remain soluble in aqueous solution but exhibit partial loss of solubility when the pH is raised to 4.5. There are two fractions at pH 4.5, one is the partially and irreversibly unfolded proteins (mainly α-lactalbumin and β-lactoglobulin) and the other is the nondenatured proteins, similar to the native proteins. An application of this temperature- and pH-dependent solubility behavior of whey proteins is whey gels at pH 7 after being heated for 120 min, exhibiting very low elasticity, firmness, and low solubility, an attribute that has been used in manufacturing soft, drinkable dairy products such as yogurts.

Fine-stranded particulate aggregates of heat-denatured whey proteins as visualized by atomic-force microscopy reveal a network structure in a heat-induced globular gel that is strongly dependent on the balance between attractive and repulsive forces among denatured protein molecules during aggregation. If the pH is far from the isoelectric point, and the ionic strength is sufficiently low, intermolecular electrostatic repulsion is dominant. Structural transitions are visible at the macroscopic level: fine-stranded gels are translucent, while particulate gels are opaque and syneresic. A fine-stranded gel formed at neutral pH is rubbery and deformable and can form a large strain without fracturing. The same gel, at acidic pH, however, is brittle because intramolecular disulfide bonding is unlikely to occur. Heat-induced aggregation of β-lactoglobulin follows a two-step model at neutral pH, in which small particles are formed first followed by aggregation into large clusters.

Twin-screw extrusion processing has been used by the food industries to impart unique properties and functionality to food proteins. The shear developed in extrusion processing, which can be controlled by varying the screw speed, is used to vary the degree of denaturation. By controlling the operating conditions, such as the amount of operating shear and the level of moisture, a variety of partially denatured products can be created with modified biological and nutritional functions. Furthermore, the combination of heat and shear in the extruder can accelerate the directed conformational changes considerably. If changes in globular-structured proteins occur at 80 °C with heat treatment alone, for example, significant denaturation or even unfolding can occur at 75 °C with the addition of shear in the extruder. At high shear and pH 8.0, whey proteins are texturized to form stringy products that are useful as meat extenders.

See also Bioorganic Chemistry.

Bibliography

Primary Literature

Alexandrescu, A.T., Evans, P.A., Pitkeathly, M., Baum, J., Dobson, C.M. (1993) Structure and dynamics of the acid-denatured molten globule state of α-lactalbumin: a two-dimensional NMR study, *Biochemistry* **32**, 1707–1718.

Alli, I.A. (2002) Protein–lipid interactions in food systems: a review, *Int. J. Food Sci. Nutr.* **53**, 249–261.

Alting, A.C., de Jongh, H.H., Visschers, R.W., Simons, J.W. (2002) Physical and chemical interactions in cold gelation of food proteins, *J. Agric. Food Chem.* **50**, 4682–4689.

Anema, S.G., Li, Y. (2003) Association of denatured whey proteins with casein micelles in heated reconstituted skim milk and its effect on casein micelle size, *J. Dairy Res.* **70**, 73–83.

Arai, M., Kuwajima, K. (2000) Role of the molten globule state in protein folding, *Adv. Protein Chem.* **53**, 209–282.

Brooks C.L. III, Gruebele, M., Onuchic, J.N., Wolynes, P.G. (1998) Chemical physics of protein folding, *Proc. Natl. Acad. Sci. USA* **95**, 11037–11038.

Caldini, G., Trotta, F., Cenci, G. (2002) Inhibition of 4-nitroquinoline-1-oxide genotoxicity by *Bacillus strains*, *Res. Microbiol.* **153**, 165–171.

Dickinson, E. (2001) Milk protein interfacial layers and the relationship to emulsion stability and rheology, *Colloids Surf., B, Biointerfaces* **20**, 197–210.

Dubreil, L., Compoint, J.P., Marion, D. (1997) Interaction of puroindolines with wheat flour polar lipids determines their foaming properties, *J. Agric. Food Chem.* **45**, 108–116.

Famelart, M.H., Tomazewski, J., Piot, M., Pezennec, S. (2003) Comparison of rheological properties of acid gels made from heated casein combined with beta-lactoglobulin or egg ovalbumin, *Int. Dairy J.* **13**, 123–134.

Farrell, H.M. Jr., Qi, P.X., Brown, E.M., Cooke, P.H., Tunick, M.H., Wickham, E.D., Unruh, J.J. (2002) Molten globule structures in milk proteins: Implications for potential new structure-function relationships, *J. Dairy Sci.* **85**, 459–471.

Ferguson, L.R. (2002) Natural and human-made mutagens and carcinogens in the human diet, *Toxicology* **181–182**, 79–82.

Fitzgerald, M.A., Martin, M., Ward, R.M., Park, W.D., Shead, H.J. (2003) Viscosity of rice flour: a rheological and biological study, *J. Agric. Food Chem.* **51**, 2295–2299.

Flanagan, J., Gunning, Y., FitzGerald, R.J. (2003) Effect of cross-linking with transglutaminase on the heat stability and some functional characteristics of sodium caseinate, *Food Res. Int.* **36**, 267–274.

Gerrard, J.A. (2002) Protein–protein crosslinking in food: methods, consequences, applications, *Trends Food Sci. Technol.* **13**, 391–399.

Goldman, R., Shields, P.G. (2003) Food mutagens, *J. Nutr.* **133**, 965S–973S.

Grinberg, V.Ya., Grinberg, N.V., Bikbov, T.M., Bronich, T.K., Mashkevich, A.Ya. (1992) Thermotropic gelation of food proteins, *Food Hydrocolloids* **6**, 69–96.

Gu, Y.S., Kim, I.S., Park, J.H., Lee, S.H., Park, D.C., Yeum, D.M., Ji, C.I., Kim, S.H. Wakabayashi, K., Kim, S.B. (2001) Effects of seasoning and heating device on mutagenicity and heterocyclic amines in cooked beef, *Biosci. Biotechnol. Biochem.* **65**, 2284–2287.

Hale, A.B., Carpenter, C.E., Walsh, M.K. (2002) Instrumental and consumer evaluation of beef patties extended with extrusion-textured whey proteins, *J. Food Sci.* **67**, 1267–1270.

Havea, P., Singh, H., Creamer, L.K. (2002) Heat-induced aggregation of whey proteins: comparison of cheese WPC with acid WPC and relevance of mineral composition, *J. Agric. Food Chem.* **50**, 4674–4681.

Jägerstad, M., Skog, S., Grivas, K., Olsson, K. (1991). Formation of heterocyclic amines using model systems, *Mutat. Res.* **259**, 219–233.

Jimenez-Guzman, J., Cruz-Guerrero, A.E., Rodriguez-Serrano, G., Lopez-Munguia, A., Gomez-Ruiz, L., Garcia-Garibay, M. (2002) Enhancement of lactase activity in milk by reactive sulfhydryl groups induced by heat treatment, *J. Dairy Sci.* **85**, 2497–2502.

Kikugawa, K., Hiramoto, K., Kato, T. (2000) Prevention of the formation of mutagenic and/or carcinogenic heterocyclic amines by food factors, *Biofactors* **12**, 123–127.

Kulmyrzaev, A., Bryant, C., McClements, D.J. (2000) Influence of sucrose on the thermal denaturation, gelation, and emulsion stabilization of whey proteins, *J. Agric. Food Chem.* **48**, 1593–1597.

Lauber, S., Krause, I., Klostermeyer, H., Henle, T. (2003) Microbial transglutaminase crosslinks β-casein and β-lactoglobulin to heterologous oligomers under high pressure, *Eur. Food Res. Technol.* **216**, 15–17.

Masuda, T., Ueno, Y., Kitabatake, N. (2001) Sweetness and enzymatic activity of lysozyme, *J. Agric. Food Chem.* **49**, 4937–4941.

Mine, Y. (1997) Effect of dry heat and mild alkaline treatment on functional properties of egg white proteins, *J. Agric. Food Chem.* **45**, 2924–2928.

Mosha, T.C., Gaga, H.E. (1999) Nutritive value and effect of blanching on the trypsin and chymotrypsin inhibitor activities of selected leafy vegetables, *Plant Foods Hum. Nutr.* **54**, 271–283.

Needs, E.C., Capellas, M., Bland, A.P., Manoj, P., MacDougal, D., Paul, G. (2000) Comparison of heat and pressure treatments of skim milk, fortified with whey protein concentrate, for set yogurt preparation: effects on milk proteins and gel structure, *J. Dairy Res.* **67**, 329–348.

Onwulata, C.I., Konstance, R.P., Smith, P.W., Holsinger, V.H. (1998) Physical properties of extruded products as affected by cheese whey, *J. Food Sci.* **63**, 814–818.

Onwulata, C.I., Smith, P.W., Konstance, R.P., Holsinger, V.H. (2001) Incorporation of whey products in extruded corn, potato or rice snacks, *Food Res. Int.* **34**, 679–687.

Pérez-Conesa, D., Ros, G., Periago, M.J. (2002) Protein nutritional quality of infant cereals during processing, *J. Cereal Sci.* **36**, 125–133.

Ptitsyn, O.B. (1994) Kinetic and equilibrium intermediates in protein folding, *Protein Eng.* **7**, 593–596.

Qi, P.X., Brown, E.M., Farrell H.M. Jr. (2001) 'New Views' on structure-function relationships in milk proteins, *Trends Food Sci. Technol.* **12**, 339–346.

Sanchez, C., Pouliot, M., Renard, D., Paquin, P. (1999) Uniaxial compression of thermal gels based on microfluidized blends of WPI and heat-denatured WPI, *J. Agric. Food Chem.* **47**, 1162–1167.

Schmitt, C., Sanchez, C., Desobry-Banon, S., Hardy, J. (1998) Structure and technofunctional properties of protein-polysaccharide complexes: a review, *Crit. Rev. Food Sci. Nutr.* **38**, 689–753.

Shah, A.J., Hansen, H.A., Jensen, B., Larsen, R.B. (1999) Fish crackers (keropok) produced by extrusion with addition of whey protein concentrate, *Food Aust.* **51**, 104–106.

Smales, C.M., Pepper, D.S., James, D.C. (2001) Protein modifications during antiviral heat bioprocessing and subsequent storage, *Biotechnol. Prog.* **17**, 974–978.

Stefancich, S., Delben, F. (1997) Interaction of food proteins with polysaccharides II. Properties upon freeze-drying of mixed solutions and thermal treatment of the lyophilizate, *J. Food Eng.* **31**, 347–363.

Stellwagen, E., Olivieri, E., Righetti, P.G. (2002) Salt-promoted protein folding, preferential binding, or electrostatic screening? *Proteins* **49**, 147–153.

Sugai, S., Ikeguchi, M. Conformational comparison between α-lactalbumin and lysozyme, *Adv. Biophys.* **30**, 37–84.

Svensson, M., Hakansson, A., Mossberg, A.K., Linse, S., Svanborg, C. (2000) Conversion of α-lactalbumin to a protein inducing apoptosis, *Proc. Natl. Acad. Sci. USA* **97**, 4221–4226.

Tedford, L.A., Kelly, S.M., Price, N.C., Schaschke, C.J. (1998) Combined effects of thermal effects of thermal and pressure processing on food protein structure, *Trans. Inst. Chem. Eng.: Part C, Food Bioproducts Process.* **72**, 80–86.

Urbano, G., Aranda, P., Gomez-Villalva, E., Frejnagel, S., Porres, J.M., Frias, J., Vidal-Valverde, C., Lopez-Jurado, M. (2003) Nutritional evaluation of pea (Pisum sativum L.) protein diets after mild hydrothermal treatment and with and without added phytase, *J. Agric. Food Chem.* **51**, 2415–2420.

Vasbinder, A.J., Alting, A.C., Visschers, R.W., de Kruif, C.G. (2003) Texture of acid milk gels: formation of disulfide cross-links during acidification, *Int. Dairy J.* **13**, 29–38.

Vassilenko, K.S., Uversky, V.N. (2002) Native-like secondary structure of molten globules, *Biochim. Biophys. Acta* **1594**, 168–177.

von Holy, A., Holzapfel, W.H. (1988) The influence of extrinsic factors on the microbiological spoilage pattern of ground beef, *Int. J. Food Microbiol.* **6**, 269–280.

Wagner, J.R., Sorgentini, D.A., Anon, M.C. (2000) Relation between solubility and surface hydrophobicity as an indicator of modifications during preparation processes of commercial and laboratory-prepared soy protein isolates, *J. Agric. Food Chem.* **48**, 3159–3165.

Weisburger, J.H., Veliath, E., Larios, E., Pittman, B., Zang, E., Hara, Y. (2002) Tea polyphenols inhibit the formation of mutagens during the cooking of meat, *Mutat. Res.* **516**, 19–22.

Were, L., Hettiarachchy, N.S., Kalapathy, U. (1997) Modified soy proteins with improved foaming and water hydration properties, *J. Food Sci.* **62**, 821–823.

Wolynes, P.G., Onuchic, J.N., Thirumalai, D. (1995). Navigating the folding routes, *Science* **267**, 1619–1620.

Yang, J., Dunker, A.K., Powers, J.R., Clark, S., Swanson, B.G. (2001) β-lactoglobulin molten globule induced by high pressure, *J. Agric. Food Chem.* **49**, 3236–3243.

Footprinting Methods to Examine the Structure and Dynamics of Nucleic Acids

Keiji Takamoto[1,3] *and Mark R. Chance*[1,2,3]
[1] *Departments Physiology and Biophysics, Albert Einstein College of Medicine, Morris Park Avenue, Bronx, New York, USA*
[2] *Biochemistry, Albert Einstein College of Medicine, Morris Park Avenue, Bronx, New York, USA*
[3] *Center for Synchrotron Biosciences, Albert Einstein College of Medicine Morris Park Avenue, Bronx, New York, USA*

1 Principles 523
1.1 Background and History of Nucleic Acid Footprinting 523
1.2 Fundamental Features of Method 526

2 Techniques 528
2.1 DNase Footprinting 528
2.2 Chemical Modification Method 528
2.3 Cleavage Using Hydroxyl Radicals 532
2.4 Quantitative Equilibrium and Time-resolved Footprinting 536
2.5 *In Vivo* Footprinting 539

3 Applications 539
3.1 Hydroxyl Radical Footprinting of RNA Folding 539
3.1.1 Thermodynamics 539
3.1.2 Kinetics 540

4 Future Prospects 542

Acknowledgment 543

Bibliography 543
Books and Reviews 543
Primary Literature 543

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Keywords

Footprinting
A technique used for probing macromolecular structures and interactions. It relies on the fact that some regions of the macromolecule are buried or become inaccessible as a consequence of their structures/interactions. Enzymes or chemical reagents are used as probes of the structure.

DNase I
An endonuclease (cleaves internally) for digestion of DNA. Has modest sequence specificity but generally cleaves at most sites in DNA.

Hydroxyl Radical
Highly reactive and unstable molecular species. Its small size and nondiscriminating high reactivity makes it an ideal probe for footprinting.

One-hit Kinetics
A characteristic of cleavage-based footprinting techniques that achieve limited cleavage of the population of macromolecules, such that, on average, each molecule cleaved is cleaved only once.

Fenton Chemistry
The most commonly used chemical method to generate hydroxyl radicals. The reaction is based on homolysis of hydrogen peroxide catalyzed by Fe(II) ion. By the reaction, Fe(II) ion is oxidized to Fe(III) ion and regenerated by the reductant (thiols, ascorbate) in the reagent.

Radiolysis
Chemical reaction driven by direct absorption of ionizing radiation energy. Particularly, aqueous solutions in the biological system generate the hydroxyl radicals through the absorption of ionizing radiation.

Reverse Transcriptase
The enzyme that makes the complementary DNA chain from RNA (reverse transcription). (Note: DNA polymerase makes DNA copy from DNA.)

Primer Extension
A technique commonly used for detecting chemical modification or cleavage of nucleic acids. Synthesized oligonucleotides are hybridized to specific sequence on the target nucleic acid and extended by the enzyme (DNA polymerase or reverse transcriptase). The enzyme stops at the modification or scission sites, and in electrophoretic separation, they appear as missing bands.

In Vivo Footprinting
Footprinting technique directly applied to living cells. It has a capability of probing interaction/structure *in situ* in a biological system.

Footprinting technology includes *in vitro* and *in vivo* methods of examining the average structure of macromolecules on the basis of the solvent accessibility and sensitivity of the macromolecule to covalent modification or cleavage. Analytical methods to separate, identify, and quantitate the unmodified and modified macromolecules are essential features of the method. In addition, a key characteristic of all footprinting assays is that the modification or cleavage reaction is of limited extent, such that the information obtained from the population of macromolecules represents the structure of the intact molecule, and thus is biologically relevant. For footprinting of nucleic acids, cleavage of accessible sites on the backbone or covalent modification of accessible bases is carried out in order to probe the structure. For cleavage-based methods, cleavage of the backbone is limited, such that on average, each molecule cleaved is cleaved only once. Sequencing gels are used to analyze the cleaved products; this allows structure probing at the monomer level due to the high resolution of the gels.

1 Principles

1.1 Background and History of Nucleic Acid Footprinting

Proteins that bind to specific regions of DNA play critical roles in the expression of genes and in the replication of the genome. Nucleic acid footprinting approaches were developed to understand specific interactions between DNA and proteins, and these investigations were quite active in the early to mid-1970s. The initial approaches in examining the specific interactions of protein and nucleic acids were of two types: the first involved isolation and characterization of a fragment of DNA that was bound to a protein of interest. The protected fragment isolation procedure involved digestion of DNA in the presence of the binding protein under study using DNase, and then subsequent chromatographic isolation of the remaining intact DNA followed by radioactive labeling and sequencing. At the time the protected fragment isolation methods were in use, Mirzabekov and coworkers and later Ptashne and coworkers recognized the value of methylation protection as a method to map protein–DNA interactions. In this case, the footprinting reagent dimethylsulfate (DMS) reacts with the N-7 of guanines and the N-3 of adenines to produce piperidine-labile sites. Specific protein contacts with these sites can block reactivity, resulting in reduced cleavage and thus identify the specific sequences involved in the binding.

Both these procedures fit the definition of "footprinting" as outlined for the purposes of this article, and they illustrate

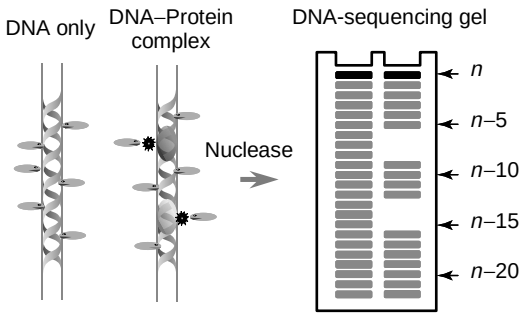

Fig. 1 Schematic representation of footprinting experiment. The nucleic acid reacts with the specific probe used to probe the macromolecule. The left-hand panel shows that all nucleotides are attacked equally, thus producing the uniform pattern in electrophoresis (right panel, left lane). If the nucleic acid interacts with proteins, the protein covers particular regions of the nucleic acid and prevents the attack by probe (middle panel), which results in missing bands in electrophoresis (right panel, right lane).

the two distinct approaches that are commonly used: that of cleavage versus modification of the nucleic acid to be probed. However, the term footprinting only came into common usage after the publication of the seminal articles by David Galas and Albert Schmitz. Since 1978, over 5000 publications using the term ''footprinting'' have been deposited in the PubMed database of the National Library of Medicine. (There are several references to footprinting in PubMed that precede 1978; these refer to ''footprinting'' of infants as a means of identification.)

The general experiment is outlined in Fig. 1. A DNA fragment is end labeled, typically with ^{32}P, although other labeling schemes can be used. Two conditions are illustrated, one in which the DNA binding protein is absent, the second in which two different proteins are added that bind to the DNA at discrete sites. The two samples are exposed to a limited dose of nuclease that, in the case illustrated, can cleave at each base position on the DNA with an equal probability. (Note that in the case of modification, the concept is similar, except that modification occurs only at the susceptible sites, cleavage is generated in a subsequent chemical work-up). In the case of the specific DNA binding sites, the nuclease is blocked from cleaving the DNA. The labeled fragments are isolated and run on a sequencing gel, separating the individual fragments. Since each position is cleaved with equal probability, one can, in the left-hand lane, see a complete set of fragments that represent DNA molecules of individual lengths ranging from the length of the uncleaved DNA, seen as the dark band at the top (length n), to fragments of length n-22, which appears at the bottom of the gel, with all the individual fragments in between. In the right-hand lane, the pattern is different – the bands from n-6 to n-8 and those from n-13 to n-15 are reduced in intensity. The limited dose provides a wide distribution of fragments. The binding of the protein reduces the reactivity of the nuclease at a number of distinct sites relative to the reactivity of those sites in the absence of protein. The data provides a map of the specific sites within the DNA in which the protein is indicated to bind.

Although these approaches using DNase I and DMS have been invaluable, additional footprinting assays have been developed using a wide range of enzymatic and chemical nucleases. In particular, some of the drawbacks of a number of chemical and enzymatic footprinting reagents include preference of a footprinting reagent for specific positions along the DNA and limiting the number of sites that can be

probed. For example, nucleases tend to cut only at a subset of the positions along the DNA backbone. Modification reagents like DMS result in cleavage at adenine- and guanine-containing sites that are accessible and thus susceptible to methylation; they cannot probe at cytosine or thymine. In addition, DNase I, since it is often of the same size as the protein that is bound to DNA, has a further disadvantage in that it can overestimate the size of the protected region, as well as be blind to any variations in reactivity within the binding site. Chemical nucleases, if they are small, can provide enhanced structural information compared to enzymatic nucleases. The most effective of these reagents and the one in widespread use is the hydroxyl radical (•OH), which is commonly generated by the venerable Fenton–Haber–Weiss reaction using Fe(II)-EDTA as a catalyst, which was developed by Tullius and coworkers. The •OH radical cleaves DNA by abstracting a hydrogen atom from the ribose/deoxyribose sugars along the DNA backbone; thus, there is virtually no sequence dependence and each position can be cleaved with equal probability. The size of the •OH radical is also advantageous; it has a Van der Walls surface area similar to that of water and thus its ability to react is highly correlated to the solvent accessibility of the reactive positions with the nucleic acid.

At the same time that new reagents were being developed for conducting footprinting experiments, protocols were developed for conducting quantitative equilibrium and kinetic footprinting experiments, initially using DNase I as the footprinting probe, but later incorporating hydroxyl radical–based methods because of their enhanced structural resolution. The development of quantitative footprinting required the development of statistical thermodynamic approaches to analyze multiple-site binding data and their interactions, and was aided by the advent of digital imaging technologies (PhosphorImagers) for the detection and quantitative analysis of radioactively labeled gels. Although the general approaches to analysis of individual sites, even at the single-nucleotide resolution, were prophesied by Galas & Schmitz, the use of explicit thermodynamics models in data analysis and imaging technologies in the digitization of data were significant advances.

Thus, experiments like those illustrated in Fig. 1 were transformed into multiple lanes where the protein concentration was titrated across a wide range of concentrations, the specific density within a binding site was quantitated from the digitized image, and the density changes as a function of protein concentration were analyzed in detail. This allowed the thermodynamic affinity of, and cooperativity among, individual sites involved in the DNA–protein interaction to be determined. The successes with DNase I equilibrium methods stimulated progress in the use of the •OH radical, with its enhanced structural resolution, in the probing of individual site binding isotherms. In addition, new scientific problems involving the structural rearrangements of catalytic RNA arose, and Cech and coworkers applied high-resolution •OH radical footprinting to the study of RNA structure and function allowing the monitoring of individual sites within the RNA as a function of ligand binding. This opened up a number of exciting opportunities for research in the RNA field and posed a challenge to footprinting technologies in that, for the nearly 400-nucleotide *Tetrahymena* ribozyme, structural information was, in principle, available at each and every nucleotide within the RNA. It was not until over 10 years

later that this potential structural resolution was harnessed, when single-band analysis of the folding of the L-21 RNA over its entire length was demonstrated.

Subsequent to the success in the analysis of equilibrium transitions, the method of DNase I footprinting was extended to kinetic analysis in the hundreds of millisecond range. In these experiments, protein and DNA were mixed in a stopped-flow device and given a pulse of nuclease to probe the location and extent of the macromolecular interaction as a function of time. This, for the first time, allowed individual site kinetic progress curves in the formation of specific DNA–protein contacts to be examined. Although kinetic studies using DNase I were developed by simply increasing the amount of DNase I used in the digestion and thus lowering the time required for sufficient cleavage, a similar approach to Fenton chemistry was not successful, increasing the concentration of reagents did not appreciably reduce the required exposure time to less than the minutes timescale. However, alternative approaches to generating •OH radicals as a probe for time-dependent studies were explored by Anderson and coworkers using peroxonitrous acid to generate •OH. The half-life of the disproportionation of peroxonitrite at neutral pH limited the timescale accessible using this reagent to a few seconds. Other methods of generating •OH radicals include radiolytic sources; however, the typical low-flux sources require exposure times on the order of hours. A breakthrough in the development of the •OH radical as a time-resolved probe of the nucleic acid structure was the application of synchrotron radiation as a radiolysis source; this has allowed time-resolved footprinting down to the tens of millisecond timescales and have been successful in the detailed examination of RNA folding and DNA–protein interactions.

As *in vitro* methods have progressed, *in vivo* footprinting has also had an impact, although the technological development has been slower. In this case, the method requires that the footprinting reagent enter the cell and react at accessible sites. Subsequently, a readout must be generated. This method was originally demonstrated to be successful by Schultz and coworkers on the basis of the genomic sequencing method by Church and Gilbert. Recently, the demonstration by Tullius and Ottinger that radiolysis could provide an effective *in vivo* footprinting tool may make the method more generally applicable.

Figure 2 lists the continuous evolution and remarkable progress of footprinting over the last 30 years; the methods of the prefootprinting era were originally codified and enhanced by Galas & Schmitz in 1978. Since that time, the introduction of high structural resolution footprinting probes, the introduction of quantitative equilibrium methods, the development of *in vivo* methods, and the development of fast time-resolved methods has resulted in footprinting becoming an extremely versatile tool in the understanding of nucleic acid structure and function.

1.2 Fundamental Features of Method

The original publication in *Nucleic Acids Research* was entitled "DNase footprinting: a simple method for the detection of protein–DNA binding specificity" and aside from introducing the term, it carefully laid out many of the basic principles of the method that are common to all successful footprinting approaches today. First was the principle of limited cleavage. The authors examined cleavage of labeled

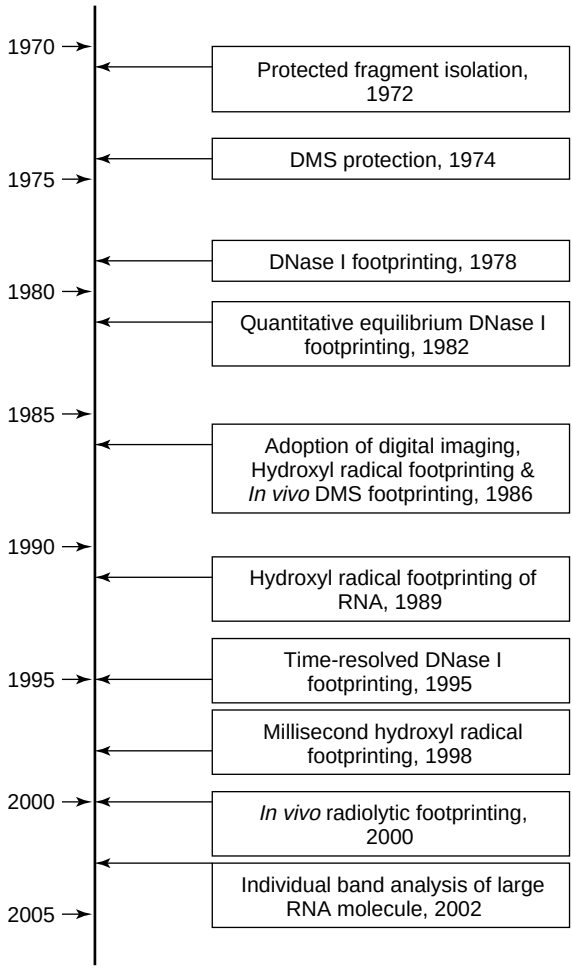

Fig. 2 Time table of development of footprinting techniques.

DNA fragments as a function of incubation time with DNase I, and upon examining the fragment distribution, determined that a limited incubation time was optimal. If the incubation time was too short, there was insufficient cleaved material to detect the products. If the incubation was carried out for an extended time, the population of smaller fragments was emphasized, skewing the distribution and depleting the number of larger fragments. This "one-hit" kinetic regime was "the condition in which those fragments produced have a high probability of coming from a single cut of the original piece of DNA." This assured a broad distribution of the fragments. It also provides for the probing of an intact, biochemically relevant population of molecules, as will be shown in this article. The second principle was that of quantitative analysis of the differential reactivity of the distinct sites within the DNA, ultimately suggesting the value of probing each and every site. It was presumed that chemical reaction kinetics would govern the production of cleaved fragments, and

that the differences in the extent of reaction of the nuclease with a number of specific sites as a function of added protein should be measured quantitatively. The ability to accurately measure the individual sites was critical to the analysis. Thus, this paper established the two principles of *limited dose* and *differential reactivity of individual sites* that are fundamental to all footprinting experiments today.

2 Techniques

2.1 DNase Footprinting

DNase I footprinting is one of the oldest techniques invented for nucleic acid structural analysis. The principle is simple. DNase cleaves DNA and generates fragments. Upon protein binding, parts of the DNA are covered by protein not accessible to enzyme and these regions cannot be cleaved. By electrophoretic separation, the end-labeled fragments can be independently examined. The inaccessible regions are visualized as missing bands in the gel as shown in Fig. 1. The technique also depends on the specificity of the enzyme used. DNase I does not have a particularly strong sequence specificity so that many sites within the DNA are probed.

The technique was first invented by Gilbert, Maniatis, and Ptashne as protected fragment isolation. The usefulness of this method was soon realized and it became a mainstream technique for examining DNA/protein interactions. The method was well codified by Galas and Schmitz and was subsequently applied to DNA/protein binding and induced loop formation and later to DNA/drug interactions. The various RNases have also been used for RNA/Protein or RNA folding analysis. The appearance of other footprinting methods (chemical modification, hydroxyl radical) has reduced its use but it is still a viable and important technique.

2.2 Chemical Modification Method

Chemical modification reagents are invaluable probes for the footprinting of both DNA and RNA. In this section, only the reactions that can be achieved under native conditions (e.g. the conditions where DNA/RNA retains its native structure) will be discussed. The reactions can be categorized in two types. The first category is backbone cleavable reactions (diethylpyrocarbonate (DEPC) for Adenine, dimethylsulfate (DMS) for Guanine) and the second category is primer extension that is used for detection of modifications (1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-*p*-toluene sulfonate (CMCT) and kethoxal [β-ethoxy-α-ketobutyraldehyde] for guanine and DMS for cytosine). Although the methylation by DMS occurs on adenine also, there is no established method to remove the methylated base; thus, it is placed in the first category. The methylation at N-7 on guanine does not stop the primer extension reaction; thus, this modification can only be detected by subsequent chemical cleavage. The methylated guanine and cytosine are unstable and subsequently their removal by piperidine for guanine and hydrazine for cytosine results in backbone scission. Similar results can be achieved with other acylation or alkylation reagents such as DEPC (piperidine treatment for scission). Hydroxylamine (cytosine), $KMnO_2$, and OsO_4 (thymine) treatments are likely to produce the pyrimidine glycols whose products are cleavable

by piperidine treatment. The chemistries and reaction products are summarized in Fig. 3.

Compared to backbone footprinting (nucleases and hydroxyl radical), the chemical modification methods require an additional step – either chemical cleavage or primer extension – to detect the products. This inconvenience is compensated by the fact that base-pair configuration can

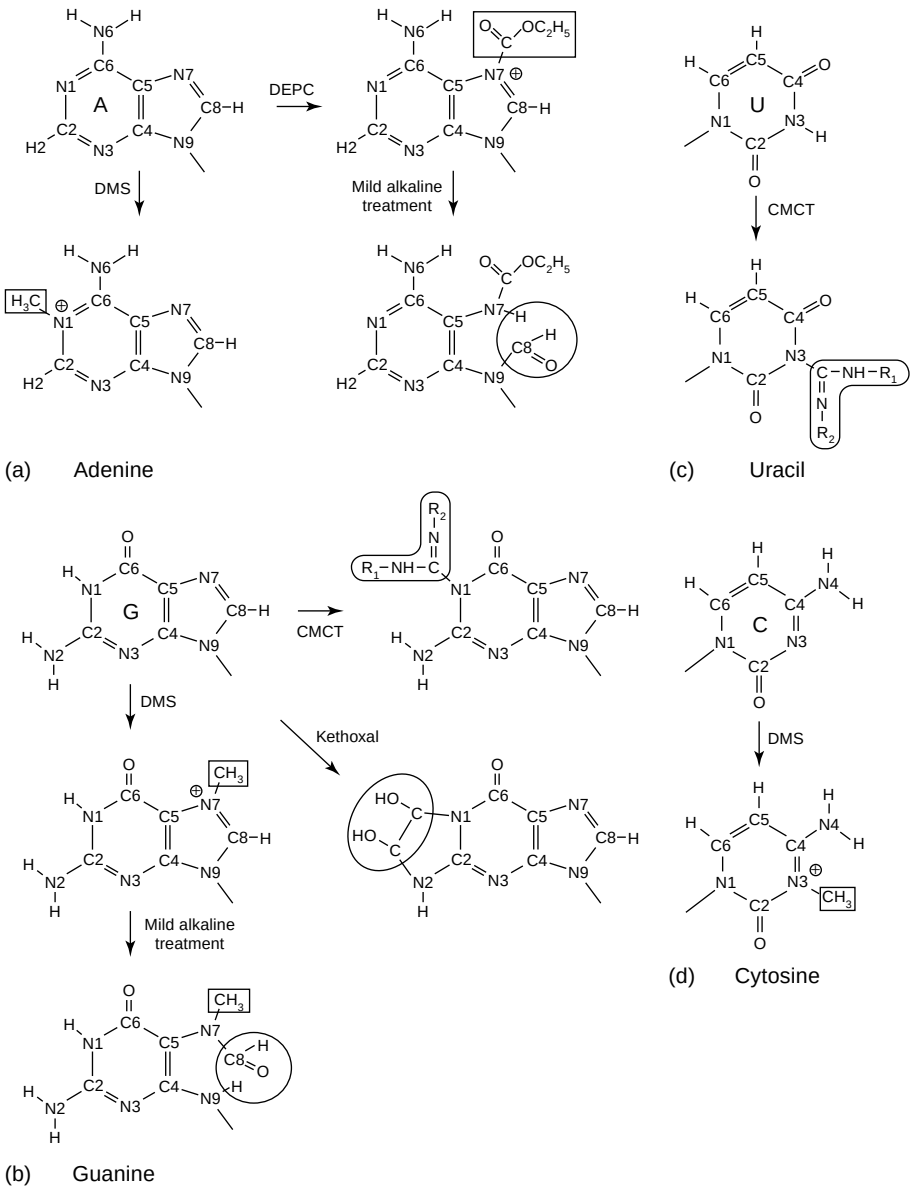

Fig. 3 The chemical modification reactions and reaction products.

be detected in many cases as shown in Fig. 4. The reagent DEPC modifies adenine N-7 that is accessible in the case of Watson–Crick base pair; however, if it is involved in the A-U (reverse) Hoogsteen base pair, it is protected from modification. In this latter base-pair configuration, N1 of adenine is accessible and thus modified by DMS. In the case of the Watson–Crick base pair, it is involved in hydrogen bonding and protected. These features are especially useful for the analysis of RNA structure that often involves noncanonical base pairs. Table 1 summarizes the chemical modifications used for DNA/RNA footprinting.

DEPC was first reported to be reactive with adenine (free and in nucleoside) in 1971, and was mentioned as a potential probe for the nucleic acid structure. Mirzabekov et al. first reported that DMS modifies DNA and later Maxam and Gilbert developed it as a chemical modification probe as well, using it in their famous chemical sequencing method. Peattie and Gilbert introduced DMS and DEPC for RNA tertiary structural analysis of tRNA. They also explored the possibility of using chemical probes for detecting RNA–protein interactions. In 1978, OsO_4 that is specific to thymine was introduced as a chemical modification reagent as an alternative to the Maxam–Gilbert chemistry. The reaction product is likely to be a thymine glycol and can be cleaved by piperidine treatment. $KMnO_4$ and hydroxylamine modifications were developed to replace some of the Maxam–Gilbert chemistries and later used as chemical probes. *N*-ethyl-*N*-nitrosourea was found to be an ethylating reagent for phosphate. It was soon realized that this reagent could be used for binding interference experiments. The modified phosphate prohibits protein binding and nitrocellulose separation of bound and unbound DNA was successfully used for determining promoter–RNA polymerase interactions. The contact site of initiator protein to the promoter region on DNA was determined by DMS. These probes (DEPC, DMS, hydroxylamine, OsO4) have been extensively used to probe DNA and RNA conformations.

The real power of chemical modification was demonstrated by combining the method with primer extension. Primer extension is a technique for detecting the modified bases and backbone scission using DNA polymerase (for DNA) or reverse transcriptase (for RNA). The specific primers for the regions of interest are chemically synthesized. The primers are annealed to the specific site in the sequence and extended by the enzyme. The enzyme stops its extension at the point at which a base is modified or the backbone is cleaved depending on the specific modifications and specificity of the enzyme. The extension reaction is generally carried out with ^{32}P-labeling allowing sensitive detection. The regions of the nucleic acid protected by protein or tertiary contacts within the molecule are neither modified nor cleaved by the reagents; thus, the extension reaction does not stop at those bases and missing bands appear in the electrophoresis results as shown in Fig. 1. Noller and coworkers used CMCT, DMS, and kethoxal as chemical probes and exploited the termination of extension at the modified base in combination with reverse transcription in order to probe the structure of the ribosome. The spectacular work of building the three-dimensional model of the 16S ribosome was done in Noller's laboratory by extensive chemical modification of protein protections on the rRNA in conjunction with the other methods. More

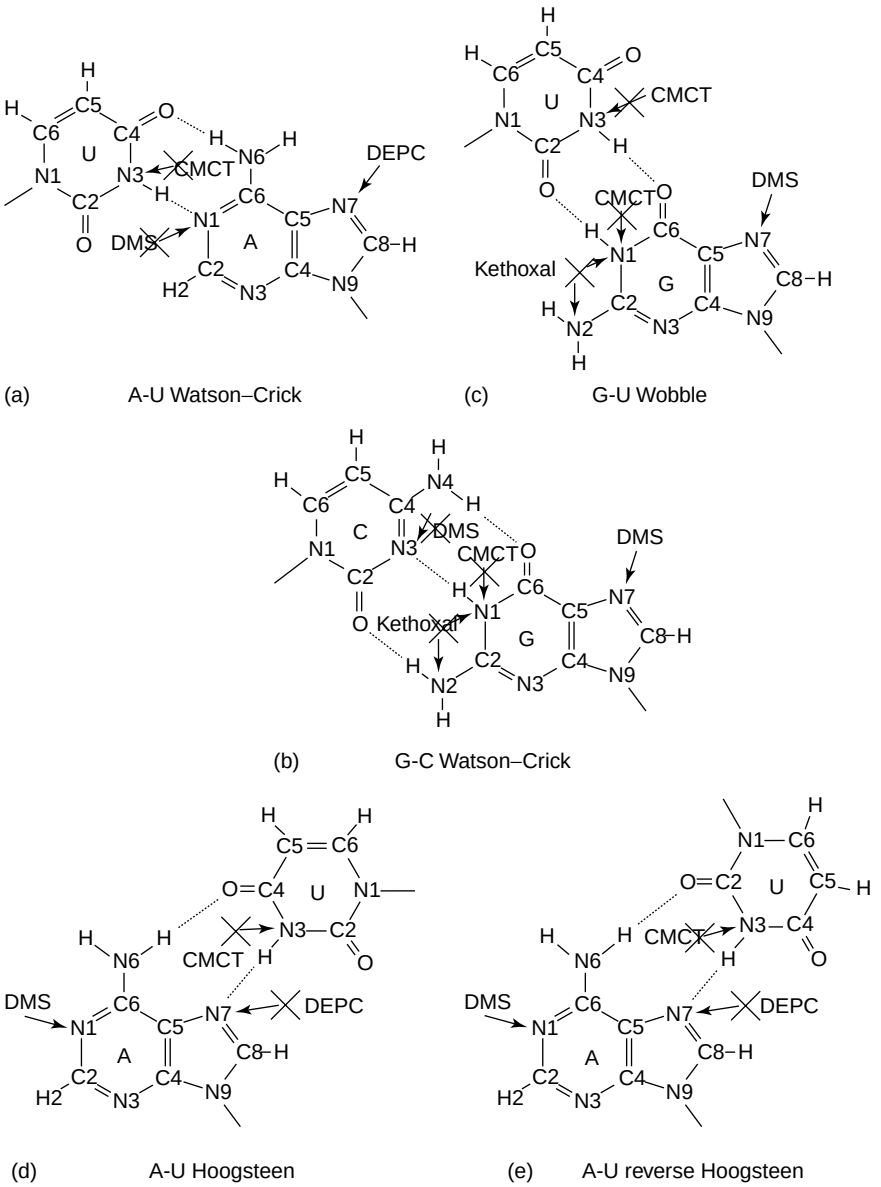

Fig. 4 The effect of base pairs on chemical modifications. The accessibility of modification sites depends on the environment surrounding the nucleotides. The formation of base pairs generally prohibits the chemical modification; however, it also depends on the base-pair configuration. Adenine N7 is open for DEPC in the Watson–Crick base pair, but is involved in a base pair in the Hoogsteen base pair, and thus protected. Adenine N1 has the opposite behavior; it is protected in the Watson–Crick base pair, but is accessible in the case of the Hoogsteen base pair. The careful analysis of chemical modification with multiple probes provides base configuration information in addition to secondary structure information.

Tab. 1 Summary of chemical modification.

Reagent	*Modification site*	**Chemical cleavage*	*Primer extension*	*Note*
CMCT	N1 of guanine	No	Yes	
	N3 of Uracil	No	Yes	
DEPC	N7 of adenine	Yes P	Yes	Detects Hoogsteen base pair
DMS	N1 of adenine	No	Yes	
	N3 of cytosine	Yes H	Yes	
	N7 of guanine	Yes P	No	No W-C detection
Hydroxylamine	C5 & C6 of cytosine	Yes P	–	Products: cytidine glycol (?)
Kethoxal	N1 & N2 of guanine	No	Yes	
$KMnO_4$	C5 & C6 of thymine	Yes P	Yes	Products: thymine glycol
OsO_4	C5 & C6 of thymine	Yes P	Yes	Products: Thymine glycol
Ethylnitrosourea	Backbone phosphate	Yes alkaline	–	Denaturing condition (50% EtOH)

*P stands for piperidine treatment, H stands for hydrazine treatment.

recently, chemical modification was used to prove the existence of alternative secondary structure in the group I ribozyme catalytic core.

2.3 Cleavage Using Hydroxyl Radicals

The hydroxyl radical is one of the most versatile and powerful probes for footprinting. It causes the scission of the polynucleotide backbone by hydrogen abstraction and subsequent strand cleavage. The hydroxyl radical's small size and high reactivity makes it possible to probe the solvent accessibility of the nucleic acid. The cleavage reaction is nonspecific (independent of base sequence) and reactivity strictly depends on the surface accessibility of the target atoms (ribose/deoxyribose 4′ and 5′-carbon).

Hydroxyl radicals can be generated by several methods. The most commonly used method is Fe(II)-EDTA Fenton chemistry. This method has the advantage that the necessary chemicals are easy to obtain, safe to handle, and cheap. A disadvantage is that the cleavage reaction requires exposure times on the order of minutes so that it is not suitable for most kinetic experiments. Peroxonitrous acid can also be used to generate hydroxyl radicals for kinetic analysis. The chemistry involves the homolytic cleavage of the molecule at neutral pH. The method gives a relatively fast burst of the hydroxyl radicals lasting 1 to 2 s. A third method for generating hydroxyl radicals is the radiolysis of water by ionizing radiation such as γ-rays or X-rays. One of the major advantages is that this method does not require added chemicals in order to generate the hydroxyl radicals so that the sample can be in solution conditions optimum for biological interactions. Also, since the source of hydroxyl radicals is water, there is no limit on the amount of hydroxyl radical generation. In the case of the γ-rays, extended time is necessary and hydrogen peroxide accumulation could be a potential problem. Conventional

X-ray sources for crystallography can be used for the hydroxyl radical generation. Hundreds of milliseconds of exposure is sufficient for the footprinting. The most powerful source for the hydroxyl radical generation so far has been high-flux X-ray synchrotron radiation. X-rays generated by a synchrotron light source can carry out footprinting on a millisecond timescale. Some of the chemistries involved in the generation of hydroxyl radicals are summarized in Fig. 5. The advantages and disadvantages of each method are summarized in Table 2.

The reactions involved in backbone scission are summarized in Fig. 6. Hydroxyl radicals abstract hydrogen from accessible and susceptible carbon atoms generating sugar radicals. 4′ and 5′ radicals generally lead to cleavage products; this has been elucidated by the reactivity changes induced by deuteration at each sugar carbon. The reaction paths are different under anaerobic versus aerobic conditions. In this section, we focus on the aerobic reaction. The yield of glycolic acid (4′-carbon attack product) and dialdehyde (5′-carbon attack product) are close to equal under aerobic conditions. Hydroxyl radical footprinting is especially suitable for nucleic acid structural analysis since the sugar-phosphate backbone faces the solvent. The accessibility of the backbone changes dramatically upon protein binding or formation of other tertiary contacts as shown in Fig. 7(a) that shows the loss of band intensity as the result of the formation of a DNA–protein complex as a function of added protein. Tertiary contact formation can also change backbone solvent accessibility significantly depending on the structure formed. Figure 7(b) shows the appearance of protections at specific regions in the Mg^{2+}-dependent folding of the *Tetrahymena* group I ribozyme.

Hydroxyl radical DNA footprinting was introduced by Van Dyke et al. using MPE-Fe(II) (Methidiumpropyl-EDTA-iron(II))

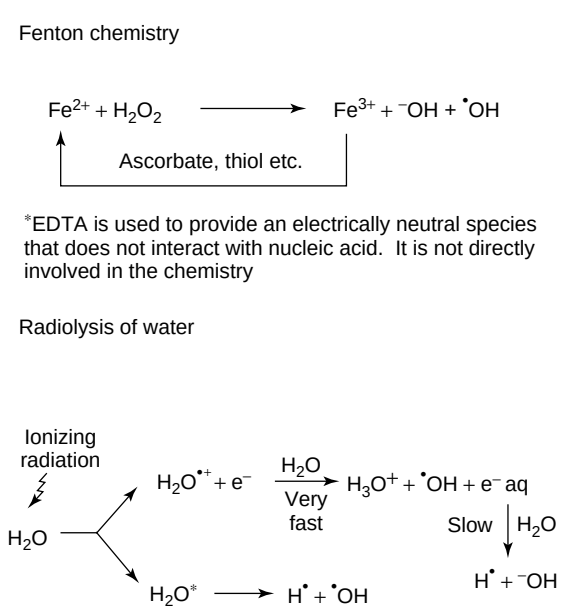

Fig. 5 Generation of hydroxyl radical. The generation of hydroxyl radicals can be achieved by several different ways. The upper panel summarizes the frequently used Fenton chemistry. The hydrogen peroxide undergoes homolysis and generates hydroxyl radical catalyzed by Fe(II) ion that is oxidized to Fe(III). In order to supply hydroxyl radicals continuously, the oxidized Fe(III) ions are reduced by added reductant. The EDTA is not directly involved in the chemistry. It is used for preventing the nonspecific or specific cation binding to the nucleic acid by chelating Fe(II) ions.

Tab. 2 Advantages and disadvantages of methods used for the hydroxyl radical generation.

Method	*Advantage*	*Disadvantage*	*Timescale*
Fenton chemistry	Easy to obtain chemicals, safe to handle, cheap	The scission takes a minute. Not fast enough for kinetics	Minute
Peroxonitrous acid	Fast enough for certain kinetic studies, no special equipment necessary	Reagent is not commercially available	Seconds
Synchrotron X-ray radiolysis	Fastest timescale	Synchrotron beamline is necessary, limited access	Milliseconds

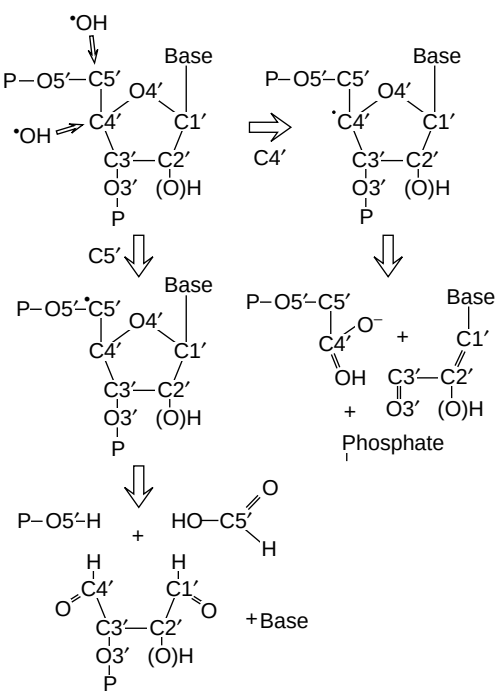

Fig. 6 Summary of backbone scission of nucleic acid by hydroxyl radicals under aerobic conditions. There are several pathways possible for backbone scission reactions induced by hydroxyl radicals. Hydroxyl radical abstracts hydrogens from the sugar carbons, specifically C4′ and C5′ position and producing sugar radicals. The subsequent reactions under aerobic condition lead to the backbone cleavage. The C4′ and C5′ reaction products are produced in almost equal amounts and governed by solvent accessibility of these atoms.

in 1982. This probe specifically intercalates into the double-helical region of DNA and preferentially cleaves the nearby backbones. Superoxide dismutase and catalase that scavenge superoxide were found to inhibit the reaction; the obvious involvement of hydroxyl radicals in the cleavage was pointed out. The versatile Fe(II)-EDTA approach was originally developed by Tullius and Dombroski as a DNA scission reagent and was soon used as a footprinting probe. The above-mentioned features of the hydroxyl radical in conjunction with the high-resolution gel electrophoretic separation of fragments enable the precise mapping of protein binding sites on DNA. Within a few years, it became a popular method for analyzing

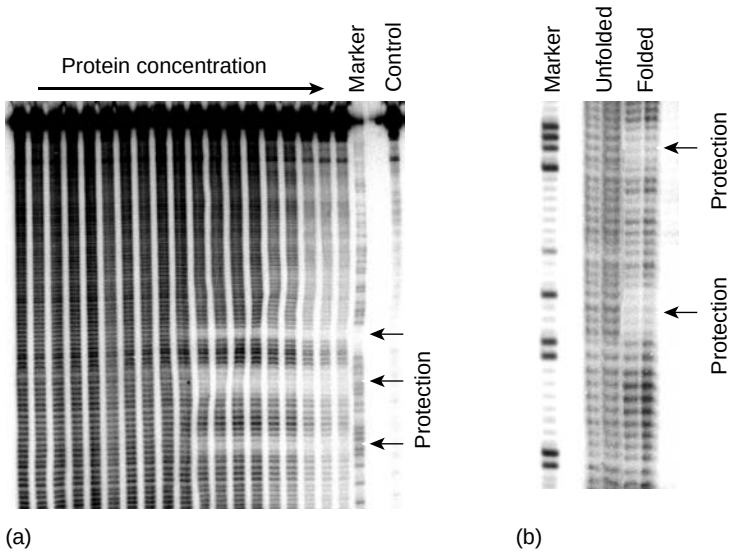

Fig. 7 Example of hydroxyl radical footprinting gel images. Panel (a): a protein titration binding assay. As protein concentration increases, the clear protections appear. Using quantitative analysis, thermodynamic (or kinetic, for progress curve) parameters can be calculated. Panel (b): RNA folding. RNA folds and discrete tertiary contacts form. The contacts may cover the backbone regions of RNA in the molecule and protect the molecule from hydroxyl radical attacks. The image shows modest protection of specific regions due to Mg^{2+}-dependent folding of the *Tetrahymena* group I ribozyme.

protein/DNA interactions. It was originally applied to RNA in the analysis of the pre-mRNA splicing complexes. The method was also applied to a wide range of protein/DNA complexes including nucleoprotein complexes at the replication origin, DNA/drug interactions, branched DNA structures, tRNA structure, and the folding of large RNAs. The method was then reinvented for the kinetic study of DNA–protein binding. The use of the synchrotron approach permits millisecond hydroxyl radical footprinting. The kinetics of *Tetrahymena* group I ribozyme folding were analyzed; the hierarchy of the folding pathway and the acceleration of folding by monovalent cations were revealed by this method.

The short half-life of the hydroxyl radical was exploited as a proximity probe by tethering Fe(II)-EDTA to the ribosomal protein S4. The Fe(II)-EDTA tethered to protein cleaves nucleotides up to 10 Å from the Fe(II) ion, allowing determination of the S4 binding site on the 16S ribosomal RNA. Tethered Fe(II)-EDTA was extensively used in conjunction with primer extension by Noller and coworkers for analysis of the ribosome structure. Recently, tethered Fe(II)-EDTA was utilized for the determination of spatial constraints of RNase P protein cofactor binding site and using these constraints, a model of the holoenzyme was built.

Other probes are available for backbone cleavage. 1,10-phenanthroline-copper

complex was introduced as a chemical nuclease in 1979. The chemistry is likely to be involved in copper-bound hydroxyl radical generation similar to Fenton chemistry. The uranyl(VI) ion (UO_2^{2+}) causes the backbone scission when it is excited by light. The method was applied to tRNA structural analysis and it was suggested that the metal binding site could be detected by this reagent. The iron(II) ion can also be used for probing divalent cation binding sites if it is used without EDTA chelator. Pb(II) can also cleave the backbone with lower specificity. There are also reports on the photochemical cleavage reagents based on acridine and photochemical homolytic cleavage of peroxides.

2.4 Quantitative Equilibrium and Time-resolved Footprinting

The detailed analysis of footprinting data can reveal structural changes in the molecule with single-nucleotide resolution. Since the degree of backbone scission by hydroxyl radicals reports the solvent accessibility of backbone and is independent of base sequence, it is possible to examine how the interaction forms quantitatively without worrying about reaction specificity or steric hindrance. For nucleic acid footprinting techniques, generally an end-labeled sample is used to visualize the reaction products by electrophoretic separation. After electrophoresis, the gel is dried and exposed to a PhosphorImager plate, then scanned (Fig. 8, Panel (a)). The intensity of each band in the digitized image represents the amount of fragment generated. The densitometric analysis of the bands transforms the bands in the gel image to a quantitative representation of the fragments (Fig. 8, Panel (b)). The change in the total density can be plotted and fitted using nonlinear regression algorithms (Fig. 8, Panel (c)). The procedure is summarized in Fig. 8.

Determining the "density" of produced fragments is conventionally performed by "box analysis." The bands of interest are inspected by the researcher using imaging software and manually "boxed-in" with a rectangle or polygon by manual action of the computer mouse. The intensities of each pixel are integrated and the total intensity of the boxed area is calculated (Fig. 8, Panel (b)). The bands in a row (belonging to a specific position in the sequence) in different columns (which are different condition of the experiment) are assembled and converted to fractional saturation $\overline{Y}$. Fractional saturation $\overline{Y}$ is the fractional extent of observed intensity normalized to the defined endpoints such as the intensity of the control lane with no protein is assigned as $\overline{Y} = 0$ and the intensity of a lane with protein at saturating concentrations is assigned as $\overline{Y} = 1$. Before the calculation of $\overline{Y}$, the data need to be "standardized." Standardization is the process that converts the arbitrary intensity to a relative value using a reference band in the same lane. The reference band should be chosen such that it should not change as a function of the experimental conditions. In Fig. 1, the reference band would be "outside" of the binding site. This standardization corrects for lane-to-lane variations due to differences in the amount of sample loaded (pipette error, precipitation yields etc.). In order to fit the plotted data set, isotherms or progress curves are generated using $\overline{Y}$. The calculated data set is fitted by the mathematical model that explains the phenomenon best. The Hill equation is commonly used for equilibrium ligand–binding studies. Two parameters are calculated in the Hill equation. The

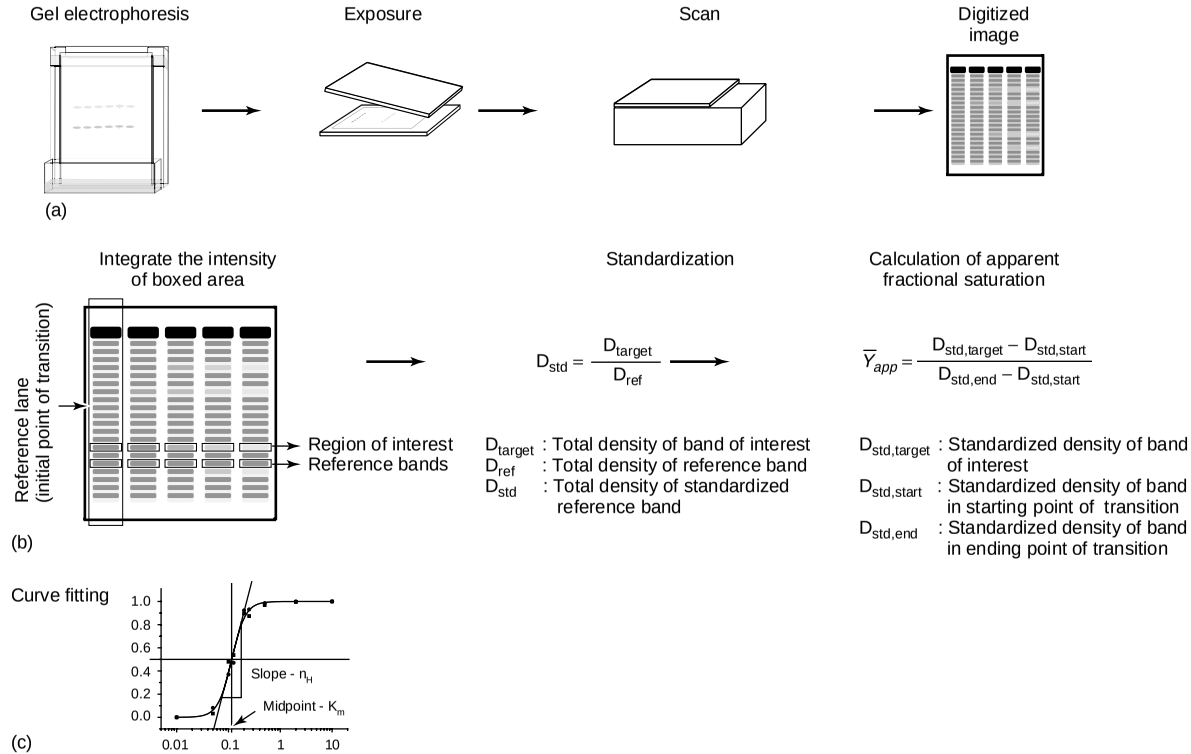

Fig. 8 Typical procedure for footprinting experiments. Panel (a): electrophoresis and scanning image. The ^{32}P-labeled sample is electrophoresed using sequencing gels (for better resolution), then dried and exposed to the imaging plate. After exposure, the imaging plate is scanned and the digitized gel image is generated. Panel (b): Data conversion: the digitized gel image is quantitated using densitometric software. The "Box analysis" or "single-band analysis" can be used for the quantification of the bands. By this procedure, the total intensities of the bands are converted to numerical data. The "standardization" within the lane is a process that eliminates the difference in loading and enables comparisons among the lanes. Fractional saturation is calculated using two endpoints that are the initial state and final state of the transitions. Panel (c): generated sets of fractional saturation values are fitted using the nonlinear regression curve-fitting algorithm, and thermodynamic or kinetic parameters of the transition are calculated.

midpoint reports the ligand concentration required for 50% response. The Hill coefficient indicates the cooperativity of the interaction. Detailed examinations of such thermodynamic and kinetic modeling approaches are provided elsewhere.

One common problem with the above approach that prevents maximal structural resolution from being extracted is that the bands may not be separated completely in the electrophoresis step. The box analysis cannot take into account the effect of overlapping areas and thus is potentially inaccurate. Although in the case of DNA gels, the background is generally low and separation is good, RNA gels can be more problematic. The box analysis can only be applied to very well separated bands or regions for accurate quantitative analysis. Figure 9, Panel (a) shows a typical gel image and the corresponding intensity profile. The minimum points on both sides of each band (peak) are not zero, indicating the overlapped area. To overcome these difficulties, the intensity profile can be fitted with a Lorentzian peak function

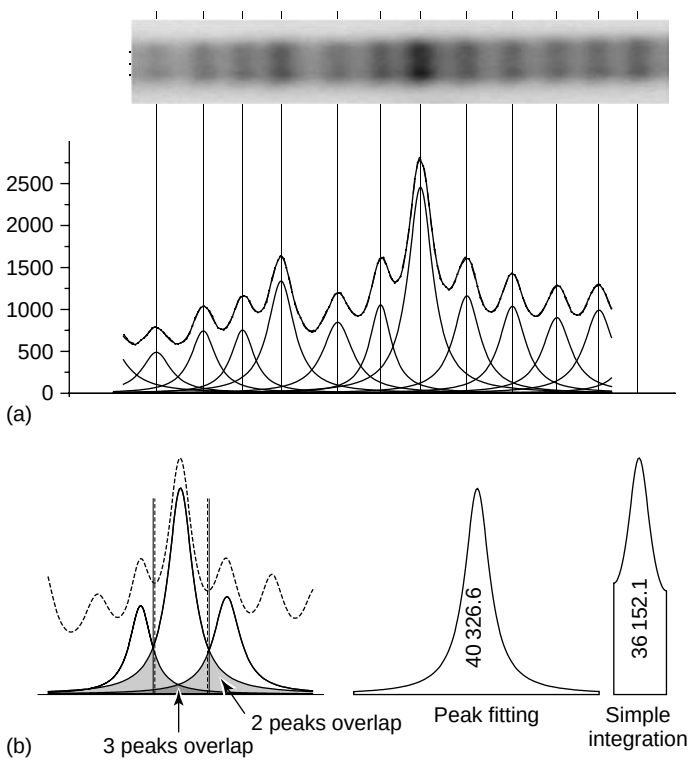

Fig. 9 The effect of overlapped area in total intensity calculations. Panel (a): typical electrophoresis gel image and its intensity profile. The intensity profile is fitted with a set of peak functions (usually Lorentzian). Panel (b): overlapped area and difference in two methods. The "box analysis" corresponds to the simple integration between the lowest points of peak. However, there is an overlap among the peaks (bands), which is contributed by bands from both sides. The integrated areas are different by 10% in this particular case.

and peak areas can be calculated (Panel (b)). The simple integration between the lowest positions in intensity gives a peak area of 36 152 while the fitted peak area is 40 326. In this case, box analysis has underestimated the intensity by 10%.

Tullius and coworkers have approached the problem using "single-band analysis." This was further developed for the structural analysis of the *Tetrahymena* group I ribozyme. In this case, the "single-band analysis" accurately provided single-nucleotide resolution data for more than 350 residues and gave more precise and detailed information about the tertiary structure formation involved in RNA folding.

Quantitative equilibrium analysis is the means to determine the thermodynamic parameters of binding, complex formation, and folding for nucleic acids. Quantitative DNase I footprinting was achieved and thermodynamic DNA–protein binding models were constructed by Ackers and coworkers. This framework was applied to hydroxyl radical footprinting in order to analyze rRNA structure and DNA conformational changes upon protein binding. Chemical denaturation methods have also been applied to RNA folding and the free-energy change and *m*-values revealed the stability and cooperativity of formation of tertiary contacts.

In addition to equilibrium footprinting, kinetic studies are also important for understanding macromolecular interactions. The first quantitative kinetic footprinting was done with peroxonitrous acid and kinetic DNase I footprinting was subsequently developed. The use of synchrotron X-rays has allowed progress curves for each tertiary contact in *Tetrahymena* group I ribozyme to be examined.

2.5 *In Vivo* Footprinting

In vivo footprinting is of great interest to molecular biologists. Many interactions cannot be reproduced *in vitro* because of the complexity of the cellular environment. DNA is packed in the chromatin and the conformations and environments are often quite different from *in vitro* conditions. The DMS chemical modification method is frequently chosen for *in vivo* footprinting, taking advantage of its small size and permeability to the interior of the cells without special treatments that could cause changes in cellular metabolism or response. $KMnO_4$ can be used as an *in vivo* footprinting reagent for the same reason as DMS. Although multicopy plasmids are relatively easy to detect by this method, detection of chromosomal DNA structure has several difficulties. The copy number is one per cell and there are complex structural arrangements (supercoiling, chromatin). A modified PCR protocol was developed for chromosomal DNA footprinting to enable detection. Recently, Tullius and coworkers explored the use of γ-ray for generating hydroxyl radicals *in vivo*. This is a potentially valuable advantage for *in vivo* footprinting since the radicals will be generated *in situ*.

3 Applications

3.1 Hydroxyl Radical Footprinting of RNA Folding

3.1.1 Thermodynamics

When RNA folds into discrete tertiary structures, regions of the backbone that become inaccessible to solvent are protected

from hydroxyl radical cleavage. Thus, unimolecular as well as bimolecular processes can yield "footprints." The folding of RNA can be monitored nucleotide by nucleotide when the extent of hydroxyl radical reactivity is followed as a function of a thermodynamic variable that shifts the folding equilibrium. This approach was first used to study the Mg^{2+}-induced folding of the *Tetrahymena* ribozyme to its catalytically active conformation.

When measurements are conducted under appropriate conditions, such transitions can yield folding isotherms allowing thermodynamic data to be derived locally within the molecule. For example, in experiments examining the Mg^{2+}-dependent folding of the *Tetrahymena* ribozyme, the isotherms describing the folding of the P4–P6 domain have higher apparent affinity for Mg^{2+} and are less concerted than the transitions within the catalytic core. The linkage of monovalent and divalent ion binding can also be examined with this method. In Mg^{2+}-dependent footprinting experiments on the isolated P4–P6 domain of the *Tetrahymena* ribozyme, direct competition exists between monovalent and divalent ions for sites within the A-bulge, that has been crystallographically shown to specifically coordinate Mg^{2+}, while other regions of the domain were unaffected as shown in Fig. 10.

Another method for examining the stability of RNA combines urea denaturation with footprinting. Approaches originally developed to determine the global stability of proteins have also been applied to RNA. When urea denaturation and footprinting are combined, site-specific unfolding isotherms are generated that can lead to an understanding of both stability changes as well as cooperative coupling within the molecule. For a two-state transition with respect to the specific protections examined, a linear dependence of folding-free energy on urea concentration is fit to the data to determine the folding-free energy in the absence of denaturant, ΔG°_{N-U}, and the sensitivity to urea denaturation, expressed as the *m*-value. This approach allows a quantitative determination of the free-energy differences for the formation of specific protections under different solution conditions or for different RNA sequences, including comparisons of mutant RNA molecules.

The combination of local structural information provided by hydroxyl radical footprinting with global measures of RNA conformation has proven valuable in understanding the processes of RNA folding. For example, hydroxyl radical footprinting of the bI5 group I intron ribozyme in concert with native gel electrophoresis demonstrated that the formation of a nonnative collapsed structure occurs at lower Mg^{2+} concentrations than native tertiary contacts.

3.1.2 Kinetics

The folding of many large RNA molecules is dominated by kinetically trapped intermediates in their folding pathways while other RNA molecules fold without kinetic traps. The development and application of an oligonucleotide hybridization assay presaged time-dependent hydroxyl radical footprinting as an effective probe of the time dependence of the Mg^{2+}-dependent folding of the group I *Tetrahymena* ribozyme. The hierarchy of this ribozyme's folding was observed to extend to the fastest folding events on the millisecond timescale using synchrotron X-ray hydroxyl radical footprinting. These

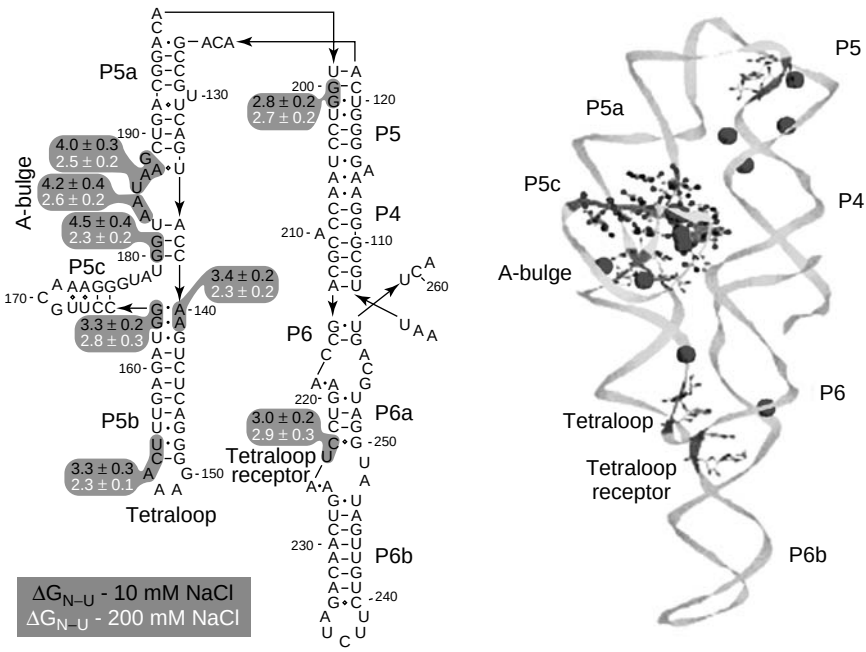

Fig. 10 ΔG°_{N-U} values of P4–P6 domain of *Tetrahymena* group I ribozyme. The hydroxyl radical footprinting experiments in conjunction with urea denaturation revealed the stability difference induced by differences in monovalent cation concentrations. The direct competition between Mg^{2+} and Na^{+} is observed in the region in which Mg^{2+}-specific binding sites are observed in the crystal structure (A-bulge). Modified from Uchida, T., He, Q., Ralston, C.Y., Brenowitz, M., Chance, M.R. (2002) Linkage of monovalent and divalent ion binding in the folding of the P4–P6 domain of the Tetrahymena ribozyme, *Biochemistry* **41** (18), 5799–5806.

studies showed that the native tertiary contacts of the P5c subdomain form first ($\sim 2\ s^{-1}$), followed by the P4–P6 domain ($\sim 1\ s^{-1}$), the peripheral helices ($0.2–0.4\ s^{-1}$), and lastly the catalytic core ($\sim 0.03\ s^{-1}$). These observations led to the novel conclusion that the organization of the catalytic core occurs within the constraints of the organized ring formed by the peripheral helices and P4–P6. Faster folding of the RNA induced by monovalent cation addition was also examined by synchrotron footprinting studies.

Progress curves obtained for the slow folding steps of the *Tetrahymena* ribozyme by time-resolved peroxonitrite footprinting are comparable to those obtained by the X-ray footprinting approach. This group of investigators used the same approach to show that folding of the M1 RNA occurs via the formation of two independently folding domains and that each domain subsequently proceeds through a discrete series of intermediates. Peroxonitrite footprinting was ideal for the kinetic analysis of 32 individual protections following the folding of a group II ribozyme derived from intron ai5γ. All the regions of this ribozyme fold slowly and synchronously on a timescale of several minutes with no evidence of kinetically

trapped intermediate species. These data suggest a folding mechanism in which the rate limiting step is early in the reaction pathway that is very different from the group I ribozymes that have been studied.

Peroxonitrite footprinting has also been used in thermodynamics and kinetics of the Mg^{2+}-dependent folding of the Varkud satellite ribozyme. This ribozyme folds rapidly within 2 s, that is, the first time point of this study while the acquisition of catalytic activity occurs more slowly appearing after 10 min. In contrast, the kinetics of folding of a minimal two-way helical junction hairpin ribozyme studied by Fe-EDTA are slow, requiring ~100 s to reach completion. Complementary time-resolved photo cross-linking studies demonstrated that a conformational change in a tertiary element (the "loop E-like" motif) of this ribozyme is coincidental with domain docking. These conformational changes are essential for catalysis.

In summary, chemical and radiolytic methods of hydroxyl radical generation are available that allow footprinting of transitions involving nucleic acid conformational change and ligand binding on timescales of milliseconds to minutes. This approach allows the time dependence of the transitions undergone by complex molecules to be quantitatively monitored with single-nucleotide resolution.

4 Future Prospects

The prospects for footprinting research have never been brighter. A technique that was once the exclusive province of molecular biologists is now a powerful probe for structural biology. Current techniques can monitor nucleic acid structure with monomer resolution and examine the dynamics of conformational changes on millisecond timescales. The growing availability of atomic resolution–structural information from crystallography and NMR studies is allowing footprinting data to be interpreted with increased accuracy and in the appropriate context of the structure. Increasingly, this allows specific structural hypotheses about conformational change and dynamics as a function of ligand binding to be tested using footprinting experiments.

A major advantage of footprinting studies is that the size of the macromolecule is not intrinsically limiting, such that intricate complexes can be examined at high resolution. Macromolecular assemblies control biological processes as fundamental as replication, transcription, translation, as well as many signal transduction pathways relevant to development and cell–cell communication. Cryo-electron microscopy and advanced mass spectrometry techniques are proving valuable in defining the size, shape, and components of such assemblies; it is clear that footprinting data will be used in conjunction with these methods to define the specific sites of important tertiary interactions. The possibility of footprinting *in vivo* provides an exciting prospect that such assemblies can be probed in the native cellular environment, although significant challenges to such experiments are obvious. However, considering the noteworthy progress of footprinting research over the last 30 years, it would not be a surprise to see such studies even 5 years hence. In any case, footprinting will remain a useful tool for structural and dynamics studies of nucleic acids for many years to come.

Acknowledgment

The authors wish to express their deep appreciation to Professor Michael Brenowitz for his many contributions to this work and to footprinting research in general. Financial support has been provided by the NIH through P41-EB-01979 and R01-GM-52348.

See also DNA–Protein Interactions; Footprinting Methods to Examine the Structure and Dynamics of Proteins; Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Brenowitz, M., Senear, D.F., Shea, M.A., Ackers, G.K. (1986) Quantitative DNase footprint titration: a method for studying protein-DNA interactions, *Methods Enzymol.* **130**, 132–181.

Brenowitz, M., Chance, M.R., Dhavan, G., Takamoto, K. (2002) Probing the structural dynamics of nucleic acids by quantitative time-resolved and equilibrium hydroxyl radical 'footprinting', *Curr. Opin. Struct. Biol.* **12**, 648–653.

Moine, H., Ehresmann, B., Ehresmann, C., Romby, P. (1997) Probing RNA Structure and Function in Solution, in: Simons, R.W., Grunberg-Manago, M. (Eds.) *RNA Structure and Function*, Cold Spring Harbor Laboratory Press, New York, pp. 77–115.

Revzin, A.E. (1993) *Footprinting Techniques for Studying Nucleic Acid-Protein Complexes (A volume of Separation, Isolation and Characterization of Biological Macromolecules)*, Academic Press, New York.

Stern, S., Moazed, D., Noller, H.F. (1988) Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension, *Methods Enzymol.* **164**, 481–489.

Sclavi, B., Woodson, S., Sullivan, M., Chance, M.R., Brenowitz, M. (1998) Following the folding of RNA with time-resolved synchrotron X-ray footprinting, *Methods Enzymol.* **295**, 379–402.

Sasse-Dwight, S., Gralla, J.D. (1991) Footprinting protein-DNA complexes *in vivo*, *Methods Enzymol.* **208**, 146–168.

Tullius, T.D., Dombroski, B.A., Churchill, M.E., Kam, L. (1987) Hydroxyl radical footprinting: a high-resolution method for mapping protein-DNA contacts, *Methods Enzymol.* **155**, 537–558.

Primary Literature

Ackers, G.K., Johnson, A.D., Shea, M.A. (1982) Quantitative model for gene regulation by lambda phage repressor, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1129–1133.

Ackers, G.K., Shea, M.A., Smith, F.R. (1983) Free energy coupling within macromolecules. The chemical work of ligand binding at the individual sites in co-operative systems, *J. Mol. Biol.* **170**, 223–242.

Balasubramanian, B., Pogozelski, W.K., Tullius, T.D. (1998) DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9738–9743.

Becker, P.B., Gloss, B., Schmid, C.W., Strahle, U., Schutz, G. (1986) *In vivo* protein-DNA interaction in a glucocorticoid response element require the presence of the hormone, *Nature* **324**, 686–688.

Becker, P.B., Ruppert, S., Schutz, G. (1987) Genomic footprinting reveals cell type-specific DNA binding of ubiquitous factors, *Cell* **51**, 435–443.

Berens, C., Streicher, B., Schroeder, R., Hillen, W. (1998) Visualizing metal-ion-binding sites in group I introns by iron(II)-mediated Fenton reactions, *Chem. Biol.* **5**, 163–175.

Blight, K.J., Rice, C.M. (1997) Secondary structure determination of the conserved 98-base sequence at the 3′ terminus of hepatitis C virus genome RNA, *J. Virol.* **71**, 7345–7352.

Breen, A.P., Murphy, J.A. (1995) Reactions of oxyl radicals with DNA, *Free Radic. Biol. Med.* **18**, 1033–1077.

Brenowitz, M., Senear, D.F., Shea, M.A., Ackers, G.K. (1986) Footprint titrations yield

valid thermodynamic isotherms, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8462–8466.

Brenowitz, M., Senear, D.F., Shea, M.A., Ackers, G.K. (1986) Quantitative DNase footprint titration: a method for studying protein-DNA interactions, *Methods Enzymol.* **130**, 132–181.

Buchmueller, K.L., Webb, A.E., Richardson, D.A., Weeks, K.M. (2000) A collapsed non-native RNA folding state, *Nat. Struct. Biol.* **7**, 362–366.

Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R., Doudna, J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing, *Science* **273**, 1678–1685.

Celander, D.W., Cech, T.R. (1991) Visualizing the higher order folding of a catalytic RNA molecule, *Science* **251**, 401–407.

Chance, M.R., Sclavi, B., Woodson, S.A., Brenowitz, M. (1997) Examining the conformational dynamics of macromolecules with time-resolved synchrotron X-ray 'footprinting', *Structure* **5**, 865–869.

Chaulk, S., MacMillan, A. (2000) Characterization of the Tetrahymena ribozyme folding pathway using the kinetic footprinting reagent peroxynitrous acid, *Biochemistry* **39**, 2–8.

Chaulk, S., Pezacki, J., MacMillan, A. (2000) Studies of RNA cleavage by photolysis of N-hydroxypyridine-2(1H)-thione. A new photochemical footprinting method, *Biochemistry* **39**, 10448–10453.

Church, G.M., Gilbert, W. (1984) Genomic sequencing, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1991–1995.

Churchill, M.E., Hayes, J.J., Tullius, T.D. (1990) Detection of drug binding to DNA by hydroxyl radical footprinting. Relationship of distamycin binding sites to DNA structure and positioned nucleosomes on 5S RNA genes of Xenopus, *Biochemistry* **29**, 6043–6050.

Culver, G.M., Cate, J.H., Yusupova, G.Z., Yusupov, M.M., Noller, H.F. (1999) Identification of an RNA-protein bridge spanning the ribosomal subunit interface, *Science* **285**, 2133–2136.

Deras, M.L., Brenowitz, M., Ralston, C.Y., Chance, M.R., Woodson, S.A. (2000) Folding mechanism of the Tetrahymena ribozyme P4-P6 domain, *Biochemistry* **39**, 10975–10985.

Dhavan, G.M., Chance, M.R., Brenowitz, M. (2003) Kinetics Analysis of DNA-Protein Interactions by Time Resolved Synchrotron X-ray Footprinting, in: Johnson, K.A. (Ed.) *Kinetic Analysis of Macromolecules: A Practical Approach*, Oxford, IRL Press, Oxford University Press.

Dhavan, G.M., Crothers, D.M., Chance, M.R., Brenowitz, M. (2002) Concerted binding and bending of DNA by Escherichia coli integration host factor, *J. Mol. Biol.* **315**, 1027–1037.

Dirheimer, G., Ebel, J.P., Bonnet, J., Gangloff, J., Keith, G., Krebs, B., Kuntzel, B., Roy, A., Weissenbach, J., Werner, C. (1972) Primary structure of transfer RNA (in French), *Biochimie* **54**, 127–144.

Doktycz, M.J., Larimer, F.W., Pastrnak, M., Stevens, A. (1998) Comparative analyses of the secondary structures of synthetic and intracellular yeast MFA2 mRNAs, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14614–14621.

Donis-Keller, H., Maxam, A.M., Gilbert, W. (1977) Mapping adenines, guanines, and pyrimidines in RNA, *Nucleic Acids Res.* **4**, 2527–2538.

Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.P., Ehresmann, B. (1987) Probing the structure of RNAs in solution, *Nucleic Acids Res.* **15**, 9109–9128.

Fang, X., Pan, T., Sosnick, T.R. (1999) A thermodynamic framework and cooperativity in the tertiary folding of a Mg2+-dependent ribozyme, *Biochemistry* **38**, 16840–16846.

Fang, X.W., Pan, T., Sosnick, T.R. (1999) Mg2+-dependent folding of a large ribozyme without kinetic traps, *Nat. Struct. Biol.* **6**, 1091–1095.

Franchet-Beuzit, J., Spotheim-Maurizot, M., Sabattier, R., Blazy-Baudras, B., Charlier, M. (1993) Radiolytic footprinting. Beta rays, gamma photons, and fast neutrons probe DNA-protein interactions, *Biochemistry* **32**, 2104–2110.

Friedmann, T., Brown, D.M. (1978) Base-specific reactions useful for DNA sequencing: methylene blue-sensitized photooxidation of guanine and osmium tetraoxide modification of thymine, *Nucleic Acids Res.* **5**, 615–622.

Galas, D.J., Schmitz, A. (1978) DNase footprinting: a simple method for the detection of protein-DNA binding specificity, *Nucleic Acids Res.* **5**, 3157–3170.

Germino, J., Bastia, D. (1983) Interaction of the plasmid R6K-encoded replication initiator protein with its binding sites on DNA, *Cell* **34**, 125–134.

Gilbert, W. (1972) The Lac Repressor and the Lac Operator, *CIBA Foundation Symp. 7: Polymerization in Biological Systems*, Elsevier Publishing Co., Amsterdam, 245.

Hampel, K.J., Burke, J.M. (2001) A conformational change in the "loop E-like" motif of the hairpin ribozyme is coincidental with domain docking and is essential for catalysis, *Biochemistry* **40**, 3723–3729.

Hampel, K.J., Burke, J.M. (2001) Time-resolved hydroxyl radical footprinting of RNA using Fe(II)-EDTA, *Methods* **23**, 233–239.

Hayes, J.J., Kam, L., Tullius, T.D. (1990) Footprinting protein-DNA complexes with gamma-rays, *Methods Enzymol.* **186**, 545–549.

Heilek, G.M., Noller, H.F. (1996) Site-directed hydroxyl radical probing of the rRNA neighborhood of ribosomal protein S5, *Science* **272**, 1659–1662.

Henn, A., Halfon, J., Kela, I., Orion, I., Sagi, I. (2001) Nucleic acid fragmentation on the millisecond timescale using a conventional X-ray rotating anode source: application to protein-DNA footprinting, *Nucleic Acids Res.* **29**, e122.

Hertzberg, R.P., Dervan, P.B. (1984) Cleavage of DNA with methidiumpropyl-EDTA-iron(II): reaction conditions and product analyses, *Biochemistry* **23**, 3934–3945.

Hiley, S., Collins, R. (2001) Rapid formation of a solvent-inaccessible core in the Neurospora Varkud satellite ribozyme, *Embo. J.* **20**, 5461–5469.

Hochschild, A. (1991) Detecting cooperative protein-DNA interaction and DNA loop formation by footprinting, *Methods Enzymol.* **208**, 343–361.

Hsieh, M., Brenowitz, M. (1996) Quantitative kinetics footprinting of protein-DNA association reactions, *Methods Enzymol.* **274**, 478–492.

Hsieh, M., Brenowitz, M. (1997) Comparison of the DNA association kinetics of the Lac repressor tetramer, its dimeric mutant $LacI^{adi}$ and the native dimeric Gal repressor, *J. Biol. Chem.* **272**, 22092–22096.

Humayun, Z., Kleid, D., Ptashne, M. (1977) Sites of contact between lambda operators and lambda repressor, *Nucleic Acids Res.* **4**, 1595–1607.

Inoue, T., Cech, T.R. (1985) Secondary structure of the circular form of the Tetrahymena rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 648–652.

Johnston, B.H., Rich, A. (1985) Chemical probes of DNA conformation: detection of Z-DNA at nucleotide resolution, *Cell* **42**, 713–724.

Joseph, S., Noller, H.F. (1998) EF-G-catalyzed translocation of anticodon stem-loop analogs of transfer RNA in the ribosome, *Embo. J.* **17**, 3478–3483.

Joseph, S., Weiser, B., Noller, H.F. (1997) Mapping the inside of the ribosome with an RNA helical ruler, *Science* **278**, 1093–1098.

Kent, O., Chaulk, S., MacMillan, A. (2000) Kinetic analysis of the M1 RNA folding pathway, *J. Mol. Biol.* **304**, 699–705.

Kimball, A., Guo, Q., Lu, M., Cunningham, R.P., Kallenbach, N.R., Seeman, N.C., Tullius, T.D. (1990) Construction and analysis of parallel and antiparallel holliday junctions, *J. Biol. Chem.* **265**, 6544–6547.

King, P.A., Jamison, E., Strahs, D., Anderson, V.E., Brenowitz, M. (1993) Footprinting' proteins on DNA with peroxonitrous acid, *Nucleic Acids Res.* **21**, 2473–2478.

Koblan, K.S., Bain, D.L., Beckett, D., Shea, M.A., Ackers, G.K. (1992) Analysis of site-specific interaction parameters in protein-DNA complexes, *Methods Enzymol.* **210**, 405–425.

Lancaster, L., Kiel, M.C., Kaji, A., Noller, H.F. (2002) Orientation of ribosome recycling factor in the ribosome from directed hydroxyl radical probing, *Cell* **111**, 129–140.

Latham, J.A., Cech, T.R. (1989) Defining the inside and outside of a catalytic RNA molecule, *Science* **245**, 276–282.

Leonard, N.J., McDonald, J.J., Henderson, R.E., Reichmann, M.E. (1971) Reaction of diethyl pyrocarbonate with nucleic acid components. Adenosine, *Biochemistry* **10**, 3335–3342.

Lu, M., Guo, Q., Mueller, J.E., Kemper, B., Studier, F.W., Seeman, N.C., Kallenbach, N.R. (1990) Characterization of a bimobile DNA junction, *J. Biol. Chem.* **265**, 16778–16785.

Maniatis, T., Ptashne, M. (1973) Multiple repressor binding at the operators in bacteriophage lambda, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1531–1535.

Maxam, A.M., Gilbert, W. (1977) A new method for sequencing DNA, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560–564.

Maxam, A.M., Gilbert, W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages, *Methods Enzymol.* **65**, 499–560.

Mirzabekov, A.D., Kolchinsky, A.M. (1974) Localization of some molecules within the grooves of DNA by modification of their complexes with dimethyl sulphate, *Mol. Biol. Rep.* **1**, 385–390.

Mirzabekov, A.D., Melnikova, A.F. (1974) Localization of chromatin proteins within DNA grooves by methylation of chromatin with dimethyl sulphate, *Mol. Biol. Rep.* **1**, 379–384.

Moazed, D., Stern, S., Noller, H.F. (1986) Rapid chemical probing of conformation in 16 S ribosomal RNA and 30 S ribosomal subunits using primer extension, *J. Mol. Biol.* **187**, 399–416.

Mortensen, U.H., Stevnsner, T., Krogh, S., Olesen, K., Westergaard, O., Bonven, B.J. (1990) Distamycin inhibition of topoisomerase I-DNA interaction: a mechanistic analysis, *Nucleic Acids Res.* **18**, 1983–1989.

Motoki, I., Yosinari, S., Watanabe, K., Nishikawa, K. (1991) Structural analyses on yeast tRNA(Tyr) and its complex with tyrosyl-tRNA synthetase by the use of hydroxyl radical 'footprinting', *Nucleic Acids Symp. Ser.* (25), 173–174.

Nielsen, P.E., Mollegaard, N.E. (1996) Sequence/structure selective thermal and photochemical cleavage of yeast-tRNA(Phe) by UO(2)2+, *J. Mol. Recognit.* **9**, 228–232.

Nielsen, P.E., Jeppesen, C., Buchardt, O. (1988) Uranyl salts as photochemical agents for cleavage of DNA and probing of protein-DNA contacts, *FEBS Lett.* **235**, 122–124.

Noller, H.F., Green, R., Heilek, G., Hoffarth, V., Huttenhofer, A., Joseph, S., Lee, I., Lieberman, K., Mankin, A., Merryman, C., et al. (1995) Structure and function of ribosomal RNA, *Biochem. Cell. Biol.* **73**, 997–1009.

Ottinger, L.M., Tullius, T.D. (2000) High-Resolution *in vivo* Footprinting of a Protein-DNA Complex Using gamma-Radiation, *J. Am. Chem. Soc.* **122**, 5901–5902.

Pan, J., Woodson, S.A. (1998) Folding intermediates of a self-splicing RNA: mispairing of the catalytic core, *J. Mol. Biol.* **280**, 597–609.

Peattie, D.A., Gilbert, W. (1980) Chemical probes for higher-order structure in RNA, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4679–4682.

Petri, V., Hsieh, M., Brenowitz, M. (1995) Thermodynamic and kinetic characterization of the binding of the TATA binding protein to the adenovirus E4 promoter, *Biochemistry* **34**, 9977–9984.

Petri, V., Hsieh, M., Jamison, E., Brenowitz, M. (1998) DNA sequence-specific recognition by the Saccharomyces cerevisiae "TATA" binding protein: promoter-dependent differences in the thermodynamics and kinetics of binding, *Biochemistry* **37**, 15842–15849.

Pogozelski, W.K., Tullius, T.D. (1998) Oxidative strand scission of nucleic acids: routes initiated by hydrogen abstraction from the sugar moiety, *Chem. Rev.* **98**, 1089–1108.

Qu, H.L., Michot, B., Bachellerie, J.P. (1983) Improved methods for structure probing in large RNAs: a rapid 'heterologous' sequencing approach is coupled to the direct mapping of nuclease accessible sites. Application to the 5′ terminal domain of eukaryotic 28S rRNA, *Nucleic Acids Res.* **11**, 5903–5920.

Ralston, C.Y., He, Q., Brenowitz, M., Chance, M.R. (2000) Stability and cooperativity of individual tertiary contacts in RNA revealed through chemical denaturation, *Nat. Struct. Biol.* **7**, 371–374.

Rangan, P., Masquida, B., Westhof, E., Woodson, S.A. (2003) Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1574–1579.

Rubin, C.M., Schmid, C.W. (1980) Pyrimidine-specific chemical reactions useful for DNA sequencing, *Nucleic Acids Res.* **8**, 4613–4619.

Russell, R., Zhuang, X., Babcock, H.P., Millett, I.S., Doniach, S., Chu, S., Herschlag, D. (2002) Exploring the folding landscape of a structured RNA, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 155–160.

Saito, I., Takayama, M., Matsuura, T. (1990) Phthalimide hydroperoxides as efficient photochemical hydroxyl radical generator: a novel DNA-cleaving agent. *J. Am. Chem. Soc.* **112**, 883–884.

Samaha, R.R., Joseph, S., O'Brien, B., O'Brien, T.W., Noller, H.F. (1999) Site-directed hydroxyl radical probing of 30S ribosomal subunits by using Fe(II) tethered to an interruption in the 16S rRNA chain, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 366–370.

Sampson, J.R., Sullivan, F.X., Behlen, L.S., DiRenzo, A.B., Uhlenbeck, O.C. (1987) Characterization of two RNA-catalyzed RNA cleavage reactions, *Cold Spring Harb. Symp. Quant. Biol.* **52**, 267–275.

Santoro, M.M., Bolen, D.W. (1992) A test of the linear extrapolation of unfolding free energy changes over an extended

denaturant concentration range, *Biochemistry* **31**, 4901–4907.

Sasse-Dwight, S., Gralla, J.D. (1991) Footprinting protein-DNA complexes *in vivo*, *Methods Enzymol.* **208**, 146–168.

Sawaki, Y., Ogata, Y. (1976) Photolysis of alpha-hydroperoxy ketones, *J. Am. Chem. Soc.* **98**, 7324–7327.

Scamrov, A.V., Beabealashvilli, R.S. (1983) Binding of actinomycin D to DNA revealed by DNase I footprinting, *FEBS Lett.* **164**, 97–101.

Schmitz, A., Galas, D.J. (1979) The interaction of RNA polymerase and lac repressor with the lac control region, *Nucleic Acids Res.* **6**, 111–137.

Schulte-Frohlinde, D., Bothe, E. (1984) Identification of a major pathway of strand break formation in poly U induced by OH radicals in presence of oxygen, *Z. Naturforsch.* **39c**, 315–319.

Sclavi, B., Sullivan, M., Chance, M.R., Brenowitz, M., Woodson, S.A. (1998) RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting, *Science* **279**, 1940–1943.

Sclavi, B., Woodson, S., Sullivan, M., Chance, M., Brenowitz, M. (1998) Following the folding of RNA with time-resolved synchrotron X-ray footprinting, *Methods Enzymol.* **295**, 379–402.

Sclavi, B., Woodson, S., Sullivan, M., Chance, M.R., Brenowitz, M. (1997) Time-resolved synchrotron X-ray "footprinting", a new approach to the study of nucleic acid structure and function: application to protein-DNA interactions and RNA folding, *J. Mol. Biol.* **266**, 144–159.

Senear, D.F., Bolen, D.W. (1992) Simultaneous analysis for testing of models and parameter estimation, *Methods Enzymol.* **210**, 463–481.

Senear, D.F., Brenowitz, M., Ackers, G.K. (1986) Energetics of cooperative protein-DNA interactions: Comparison between quantitative DNase footprinting titration and filterbinding, *Biochemistry* **25**, 7344–7354.

Serrano, M., Salas, M., Hermoso, J.M. (1990) A novel nucleoprotein complex at a replication origin, *Science* **248**, 1012–1016.

Shadle, S.E., Allen, D.F., Guo, H., Pogozelski, W.K., Bashkin, J.S., Tullius, T.D. (1997) Quantitative analysis of electrophoresis data: novel curve fitting methodology and its application to the determination of a protein-DNA binding constant, *Nucleic Acids Res.* **25**, 850–860.

Shelton, V.M., Sosnick, T.R., Pan, T. (1999) Applicability of urea in the thermodynamic analysis of secondary and tertiary RNA folding, *Biochemistry* **38**, 16831–16839.

Siebenlist, U., Gilbert, W. (1980) Contacts between Escherichia coli RNA polymerase and an early promoter of phage T7, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 122–126.

Sigman, D.S., Graham, D.R., D'Aurora, V., Stern, A.M. (1979) Oxygen-dependent cleavage of DNA by the 1,10-phenanthroline cuprous complex. Inhibition of escherichia coli DNA polymerase I, *J. Biol. Chem.* **254**, 12269–12272.

Singer, B., Sun, L., Fraenkel-Conrat, H. (1975) Effects of alkylation of phosphodiesters and of bases of infectivity and stability of tobacco mosaic virus RNA, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2232–2236.

Stern, S., Moazed, D., Noller, H.F. (1988) Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension, *Methods Enzymol.* **164**, 481–489.

Stern, S., Weiser, B., Noller, H.F. (1988) Model for the three-dimensional folding of 16 S ribosomal RNA, *J. Mol. Biol.* **204**, 447–481.

Strahs, D., Brenowitz, M. (1994) DNA conformational changes associated with the cooperative binding of cI-repressor of bacteriophage lambda to OR, *J. Mol. Biol.* **244**, 494–510.

Swisher, J., Su, L., Brenowitz, M., Anderson, V., Pyle, A. (2002) Productive folding to the native state by a group II intron ribozyme, *J. Mol. Biol.* **315**, 297–310.

Takamoto, K., He, Q., Morris, S., Chance, M.R., Brenowitz, M. (2002) Monovalent cations mediate formation of native tertiary structure of the tetrahymena thermophila ribozyme, *Nat. Struct. Biol.* **9**, 928–933.

Treiber, D.K., Williamson, J.R. (1999) Exposing the kinetic traps in RNA folding, *Curr. Opin. Struct. Biol.* **9**, 339–345.

Treiber, D.K., Williamson, J.R. (2001) Beyond kinetic traps in RNA folding, *Curr. Opin. Struct. Biol.* **11**, 309–314.

Tsai, H.Y., Masquida, B., Biswas, R., Westhof, E., Gopalan, V. (2003) Molecular modeling of the three-dimensional structure of the bacterial RNase pholoenzyme, *J. Mol. Biol.* **325**, 661–675.

Tullius, T.D., Dombroski, B.A. (1985) Iron(II) EDTA used to measure the helical twist along any DNA molecule, *Science* **230**, 679–681.

Tullius, T.D., Dombroski, B.A. (1986) Hydroxyl radical "footprinting": high-resolution

information about DNA-protein contacts and application to Lambda Repressor and Cro Protein, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5469–5473.

Uchida, T., He, Q., Ralston, C.Y., Brenowitz, M., Chance, M.R. (2002) Linkage of monovalent and divalent ion binding in the folding of the P4-P6 domain of the Tetrahymena ribozyme, *Biochemistry* **41**, 5799–5806.

Uchida, T., Takamoto, K., He, Q., Chance, M.R., Brenowitz, M. (2003) Multiple monovalent ion-dependent pathways for the folding of the L-21 *Tetrahymena thermophila* Ribozyme, *J. Mol. Biol.* **328**, 463–478.

Van Dyke, M.W., Hertzberg, R.P., Dervan, P.B. (1982) Map of distamycin, netropsin, and actinomycin binding sites on heterogeneous DNA: DNA cleavage-inhibition patterns with methidiumpropyl-EDTA.Fe(II), *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5470–5474.

Wang, X.D., Padgett, R. (1989) Hydroxyl radical ''footprinting'' of RNA: application to pre-mRNA splicing complexes, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7795–7799.

Webb, A.E., Weeks, K.M. (2001) A collapsed state functions to self-chaperone RNA folding into a native ribonucleoprotein complex, *Nat. Struct. Biol.* **8**, 135–140.

Wittberger, D., Berens, C., Hammann, C., Westhof, E., Schroeder, R. (2000) Evaluation of uranyl photocleavage as a probe to monitor ion binding and flexibility in RNAs, *J. Mol. Biol.* **300**, 339–352.

Zarrinkar, P.P., Williamson, J.R. (1994) Kinetic intermediates in RNA folding, *Science* **265**, 918–924.

Footprinting Methods to Examine the Structure and Dynamics of Proteins

Jing-Qu Guan[1,3] *and Mark R. Chance*[1,2,3]
[1] *Departments of Physiology & Biophysics, Albert Einstein College of Medicine, Bronx, New York, USA*
[2] *Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York, USA*
[3] *Center for Synchrotron Biosciences, Albert Einstein College of Medicine, Bronx, New York, USA*

1 Principle of Protein Footprinting 551
1.1 Background & History 551
1.2 Fundamental Features of Footprinting Methods 553

2 Techniques 554
2.1 Cleavage-based Protein Footprinting 554
2.1.1 End-labeling Procedures 555
2.1.2 Protein Cleavage 555
2.2 Modification-based Protein Footprinting 556

3 Applications 559
3.1 Quantitative Equilibrium Footprinting 559
3.2 Footprinting of Complexes 562

4 Future Prospects 564

Bibliography 565
Books and Reviews 565
Primary Literature 565
References 566

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Keywords

One-hit Kinetics
A characteristic of cleavage-based footprinting techniques that achieve limited cleavage of the population of macromolecules, such that on average, each molecule cleaved is cleaved only once.

Footprinting
A technique used for probing macromolecular structures and interactions. It relies on the fact that some regions of the macromolecule are buried or become inaccessible as a consequence of their structures/interactions. Enzymes or chemical reagents are used as probes of the structure.

Hydroxyl Radical
Highly reactive and unstable molecular species. Its small size and nondiscriminating reactivity make it an ideal probe for footprinting.

Fenton Chemistry
The most commonly used chemical method to generate hydroxyl radicals. It is based on homolysis of hydrogen peroxide catalyzed by Fe(II) ion. By the reaction, Fe(II) ion is oxidized to Fe(III) ion and regenerated by the reductant (thiols, ascorbate) in the reagent.

Radiolysis
Chemical reaction driven by direct absorption of ionizing radiation energy. Particularly, aqueous solutions in biological systems generate hydroxyl radicals and hydrated electrons through the absorption of ionizing radiation.

RNA Polymerase
Macromolecular complex involved in the gene transcription.

Mass Spectrometry
Method to study proteins or protein fragments using high-resolution mass identification.

Western Blot
Method of molecular identification using antibody recognition and size.

"Footprinting" refers to assays that examine ligand binding and conformational changes by determining the solvent accessibility of the backbone, bases, or side-chain structures of macromolecules through their sensitivity to chemical or enzymatic cleavage or modification reactions. Protein footprinting methods have been developed to examine protein structure and conformational changes by monitoring solvent accessibility using either modification or cleavage reactions.

Examples of such methods include acetylation, nonspecific limited proteolysis, or probing accessibility of a range of sites that can be specifically or nonspecifically cleaved or modified. The basis of these diverse chemical approaches is to monitor the change in accessibility of susceptible residues as a function of relevant conformational fluctuations. In addition, the amounts of material required for exploring the structure and function of proteins using footprinting is in the femtomole to picomole level for detailed analysis. Protein footprinting methods based on protease cleavage have been used to map protein structure, nucleic acid–protein interactions and protein folding intermediates. However, cleavage techniques using proteases suffer from limited structural resolution due to the large size of the probe. Hydroxyl radical methods of cleavage and modification have been pursued to overcome this drawback. In this article, we document the history of protein footprinting and the development of new technologies based on mass spectrometry to accurately locate and quantify protein structure and conformational changes upon ligand binding.

1 Principle of Protein Footprinting

1.1 Background & History

Understanding the details of protein structure and their interactions in macromolecular assemblies is one of the most important problems in biology. The interactions of proteins and their assembly into large complexes control processes as fundamental as replication, transcription, translation, metastasis, apoptosis, and many other signal transduction pathways relevant to development and cell–cell communication. Protein footprinting approaches, paralleling the successful development of nucleic acid footprinting approaches (*see also* Takamoto and Chance, Footprinting methods to examine the structure and dynamics of nucleic acids), have been developed and refined over the last 15 years in order to provide the molecular details of these interactions and their dynamics.

"Footprinting" refers to assays that examine ligand binding and conformational changes by determining the solvent accessibility of macromolecules through their sensitivity to chemical or enzymatic modification and cleavage reactions. A fundamental feature distinguishing footprinting from other techniques is the covalent modifications of the macromolecule that are essential to probing the solvent accessible sites. These irreversible changes to the macromolecule can negatively influence the experimental results unless sufficient care is taken in experimental design and analysis. Fortunately, during the 25 years of successful development of macromolecular footprinting approaches, a number of basic principles and control experiments have been established that have allowed the successful examination of structure and dynamics for a wide range of protein and nucleic acid systems, as illustrated in the thousands of footprinting publications that have appeared in the literature.

The initial approaches to examining the specific interactions of macromolecules from the standpoint of either proteins or the nucleic acids were of two types: the first

involved cleavage of the macromolecule followed by a readout of the cleaved fragments and the second, the covalent modification of the macromolecule and analysis of the modified products. In each case, the footprint represents the differential reactivity of accessible sites compared to those that are buried within the structure. For nucleic acids, the first cleavage-based methods that were developed were enzymatic, and involved isolation of a fragment of DNA that was "protected" from nuclease digestion due to its binding to a specific protein. Footprinting using chemical modification was developed shortly thereafter; the ligand-induced protection of nucleic acid from methylation was used to map protein–DNA interactions. These approaches evolved into the explicit implementation of (nucleic acid) cleavage-based footprinting that closely resembles the approaches in use today. These footprinting experiments rely on limited cleavage of the backbone (where each position along the backbone is sampled with equal probability) and the ability to uniquely identify the distinct cleavage products so that multiple sites could be individually measured.

Paralleling the approaches pioneered for nucleic acids, the first explicit implementation of protein footprinting used enzymatic cleavage of the macromolecular backbone to probe the contact interface between a protein and its ligand, in this case the interaction of an antigen–antibody complex. Monoclonal antibodies interact with specific epitopes and the antibody–antigen interactions were shown to protect these specific binding sites from protease cleavage. Different proteases were used to probe different sites within the antigen depending on the sequence specificity. In addition, the authors detected antibody binding–induced conformational changes in the antigen, demonstrating the power of footprinting techniques to, in principle, probe sites throughout the macromolecule. A drawback of this particular experiment was that [35-S] methionine labeling was used to label the antigen; this labeled methionines throughout the sequence and the cleavage products generated fragments that could not be uniquely identified by size. Refinements of this method were introduced. In these cases, the end-labeling allowed the unique identification of the fragments. Footprinting of proteins by chemical modification was refined. This method mapped the solvent accessibility of lysine residues by determining their reactivity to modification by acetylation. Although these two approaches incorporated fundamental features of footprinting, it should be noted that methods to modify proteins and probe their structure (originally using diazonium salts) are nearly 100 years old. In addition, specific methods to use modification reagents to map "buried" versus "free" histidine side chains go back almost 40 years. However, these modification methods primarily evolved into approaches to generate specific sites within proteins that were optically active so as to probe conformational changes using the covalent adducts.

Nevertheless, the modification and cleavage approaches outlined above fit the definition of "footprinting" for the purposes of this article, and they illustrate the two distinct approaches that are commonly used. Although the methods outlined above are valuable, additional footprinting assays have been developed using a number of additional approaches. In particular, some of the drawbacks of a number of chemical and enzymatic footprinting reagents include preference of a footprinting reagent for specific positions

within proteins, limiting the number of sites that can be probed. For example, proteases are often quite specific and cut only at a subset of the positions along the protein sequence. Also, the size of proteases, which often can be of the same size as the ligand, can result in an overestimation of the size of the protected region, as well as be blind to any variations in reactivity within the binding site. Modification methods in which lysine is acetylated are suitable for probing accessible lysine side chains, but cannot probe at other sites. Chemical proteases, if they are small, can provide enhanced structural information compared to enzymatic proteases. For example, CNBr can cleave at methionine sites and thus its cleavage products are sensitive to burial of methionine residues. However, the most effective chemical protease and the one in widespread use is the hydroxyl radical (.OH), which is commonly generated by the Fenton–Haber–Weiss reaction using Fe(II)-EDTA. It was shown that hydroxyl radicals could provide a relatively nonspecific probe of protein structure and map the sites of protein-DNA binding. The use of hydroxyl radical–based cleavage has the advantages of not only a small size of the probe but also a relatively nonspecific cleavage activity, maximizing the number of potential sites along the backbone that can be probed. However, the use of tethered sources of hydroxyl radicals, which generate the radicals at defined sites, has also proven to be a powerful approach in defining protein–protein interfaces and specific sites of interaction. Hydroxyl radicals have also been developed as side-chain modification reagents, and this approach has been coupled to proteolysis and mass spectrometry techniques to precisely map the modification sites.

1.2 Fundamental Features of Footprinting Methods

The key publication in *Nucleic Acids Research* that has critically influenced the field of footprinting was entitled "DNase footprinting: a simple method for the detection of protein–DNA binding specificity." Aside from popularizing the term *footprinting*, the paper carefully laid out many of the basic principles of the method that are common to all successful footprinting approaches today. The first to be examined was the principle of limited cleavage of the macromolecule. The authors examined the cleavage of labeled DNA fragments (the macromolecule) as a function of incubation time with DNase I (the cleavage agent), and upon examining the fragment distribution, determined that a limited incubation time was optimal. If the incubation time was too short, there was insufficient cleaved material to detect the products. If the incubation was carried out for an extended time, the population of smaller fragments was emphasized, skewing the distribution and depleting the number of larger fragments. This "one-hit" kinetic regime was "the condition in which those fragments produced have a high probability of coming from a single cut of the original piece of DNA." This assured a broad distribution of the fragments. For footprinting methods relying on the cleavage of proteins, these principles apply as well. The principle of limited cleavage also provides for the probing of an intact, biochemically relevant population of molecules. The second principle was that of quantitative analysis of the differential reactivity of the distinct sites within the macromolecule, ultimately suggesting the ability to probe, and the value of probing, each and every site. It was

presumed that chemical reaction kinetics would govern the production of cleaved fragments, and that the differences in the extent of reaction of the cleavage agent with a number of specific sites as a function of added protein should be measured quantitatively. The ability to quantitatively measure the individual sites was critical to the analysis. Thus, this paper established the two principles of *limited dose* and *differential reactivity of individual sites* that are fundamental to all footprinting experiments today.

2 Techniques

2.1 Cleavage-based Protein Footprinting

The demonstration of cleavage-based footprinting methods for nucleic acids inspired similar methods to be developed for the analysis of proteins. Initial approaches to examining protein structure and interactions utilized proteases or other cleavage-based reagents (like hydroxyl radicals) in order to assay the solvent-accessible sites. The experimental steps are as follows. Proteins are labeled at one end (see below) and then are subjected to the cleavage agent. The cleavage products are assayed for size by SDS-PAGE (Fig. 1). For nonspecific cleavage, solvent-accessible sites throughout the macromolecule would be sensitive to cleavage, in principle be probed, and would provide a wide range of fragments. One fundamental difference between footprinting studies of DNA and proteins is that for DNA, since the backbone is on the outside of the canonical helical structure, almost every site is potentially accessible. For proteins, the backbone may be buried on the inside of the molecule for much of the structure so the number

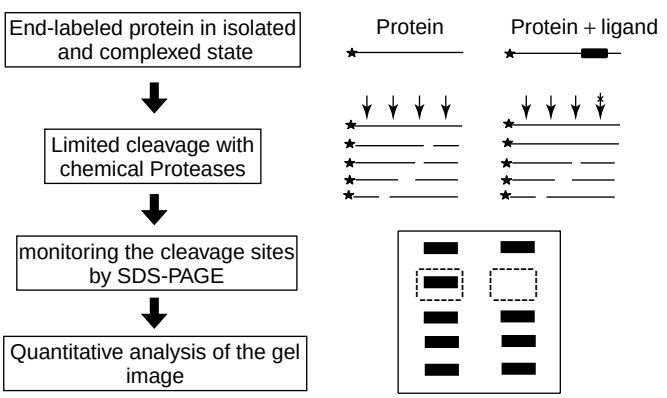

Fig. 1 Schematic representation of cleavage-based protein footprinting. A protein is specifically end-labeled and then subjected to limited cleavage under single-hit conditions in the presence and absence of the ligand. The cleavage products are then separated by size using high-resolution SDS-PAGE, followed by quantitative analysis of a digitized gel image. After correction for protein loading or cleavage-efficiency differences, intensities of corresponding bands in the lane with and without the ligand are compared to get information about the changes in susceptibility to cleavage as a result of ligand binding.

of potential cleavage sites are lessened. Thus, the resultant data will not cleave at each and every monomer position. However, sites on the "inside" are not likely candidates for the binding of a ligand such that the impact of this "disadvantage" is lessened. In addition, the variation in cleavage across the macromolecular sites interrogates the protein's tertiary structure. When a ligand binds to the protein, accessible cleavage sites on the macromolecule may experience reduced reactivity at the interface (Fig. 1). Many sites should remain unchanged, but allosteric interactions could increase or decrease the solvent accessibility at other sites throughout the macromolecule as well.

2.1.1 End-labeling Procedures

End-labeling is required if SDS-PAGE is to be used to separate the cleavage fragments for analysis. The most common methods of end-labeling are specific antibody binding to either terminus or phosphorylation. This end-labeling is advantageous for two reasons. First, it simplifies the identification of the fragments on the gel. For example, when a cleavage event occurs, it naturally creates two pieces. If end-labeling is used, only the fragment retaining the end-label is seen on the gel. One can then immediately identify the site of the cleavage based on the length of the observed fragment. Second, the end-labeling methods generally enhance the sensitivity of the assay, Western blotting (for antibody identification) or autoradiography (for P-32 end-labeling) have vastly increased sensitivity compared to gel staining methods. Use of antibodies requires either a specific antibody to either end of the protein to be available (or to be generated) or the addition of a specific epitope, for which a commercial antibody is available, to either end of the protein. This epitope can be engineered as part of a fusion protein purification strategy, for example, for His-tagged proteins commercial anti-His tag antibodies are available. Phosphorylation at either terminus can be used which is based on the sequence specificity of protein kinases. A phosphorylation sequence (like HMK or FLAG) that can be phosphorylated by a specific kinase is added to the end of the protein at the gene level.

2.1.2 Protein Cleavage

Protein cleavage is accomplished either by enzymatic or chemical means. Enzymes to cleave proteins have widespread and variable specificity. To achieve cleavage at a wide range of sites using a protease, a "cocktail" of several proteases is useful. However, the protease method has the ultimate drawback that the enzyme is large and may not be sufficiently sensitive to accessibility of the structural backbone to accurately probe the structure. As in the case of nucleic acids, the advent of the use of hydroxyl radicals enhanced the structural resolution of the technique considerably. Hydroxyl radicals have been used both free in solution, where they probe the entire molecular surface, and tethered to specific probe sites, or to sites throughout the macromolecule. A tethered source of hydroxyl radicals provides a cleavage agent with a defined radius of action, generally 10 to 15 Å. In the case of tethering to a wide distribution of sites, it has been found that if the tethers are linked to one molecule in a pair of interacting molecules, the target cleavage sites are localized close to the interacting surface of the binary partner. Thus, the modulation of reactivity "observed" in the hydroxyl radical assay can be dissected into the allosteric effects (which are observed

in addition to the effects of burial of the interface when "free" radicals are used) and the formation of an interface (which are observed in the experiment with tethered sources of radicals).

2.2 Modification-based Protein Footprinting

As outlined above, it was demonstrated that specific amino acids, like lysine, can act as the footprinting targets for protein footprinting. A protein tagged at one particular end, in both isolated and complexed states, is first treated with a reversible lysine reagent. At this stage only those lysines in a protein that are accessible to the reagent are modified. The protein is then unfolded and treated extensively with an irreversible lysine reagent to block those lysines that did not react previously, followed by reversing the first lysine modification. A lysine-specific endoproteinase is then used to cleave the tagged protein at the deblocked lysines. Separation of the proteolytic products by size and identification of the tagged fragments map the positions of these lysines. As above, gel-electrophoresis methods were used to map the susceptible sites. This provides a way to identify the lysines in a protein that react differently to the reversible reagent when the protein is free by itself and when it is present in a complex. This method takes advantage of the abundance of lysines in many proteins and the tendency of lysyl groups to be located on surfaces rather than in the interior, making lysyl groups attractive targets in probing protein–protein interactions. The lysines of a viral topoisomerase that were involved in intermolecular contacts with DNA were specifically mapped in this work. Hanai has also pointed out that many reagents available for the modification of proteins that were developed in the 1960s and 1970s could be "revived" to be used in footprinting assays. In fact, a number of novel approaches to modifying the backbone and specifically incorporated side-chain groups have been developed and successfully applied.

Another approach that has been recently developed is to use chemical modification reagents that are less specific to provide footprinting approaches. It has been well known for some time that hydroxyl radicals, besides generating backbone cleavage events, modify the side chains of proteins quite efficiently. Sulfur-containing, aromatic and aliphatic side chains provide targets that are more sensitive to attack than the protein backbone. Thus, these events are more frequent than the abstraction of the C-α carbon that is the first step in hydroxyl radical–based protein cleavage. The ability of side chains to be modified was combined with mass spectrometric detection to develop a protein footprinting technique. Hydroxyl radicals, as a reagent for footprinting, can be derived from Fenton reagents or from radiolysis sources. In the former case, the Fenton reagent induces a metal catalyzed reaction of hydrogen peroxide to produce hydroxyl radicals. In the latter case, ionization of water produces hydroxyl radicals directly.

The development of the hydroxyl radical as a modification reagent for footprinting and its application in conjunction with mass spectrometry was directly inspired by the development of deuterium exchange mass spectrometry methods. In these approaches, the protein of interest is subjected to a pulse of deuterium intended to label the structural regions that are solvent-accessible and to monitor changes in accessibility in response to the binding of a ligand. After solvent

labeling, the reaction is quenched, the protein is fragmented by proteolysis, the peptide fragments are separated by high-pressure liquid chromatography (HPLC), and mass spectrometry analysis is performed. Peptide fragments with increased mass, relative to control experiments without addition of deuterium, indicate *specific segments* that were solvent accessible and exchange competent during the deuterium pulse. The method used to minimize back exchange during the analytical steps is lowering the pH to ~2.5, thus only proteases with activity at acid pH (e.g. pepsin) can be used to fragment the protein. Although recent, column based, proteolysis methods have decreased the time required for this step, the back exchange is limiting.

Hydroxyl radical–mediated protein footprinting is similar. The overall method is outlined in Fig. 2. In this case, the protein solution is exposed to ionizing radiation and the hydroxyl radicals covalently react with surface-accessible residues, primarily side-chain groups. As in the deuterium exchange methods, the protein is subjected to proteolysis. However, in contrast to deuterium exchange, the production of stable modifications through hydroxyl radical exposure allows a wide range of samples as well as proteases to be used to fragment the protein under a wide range of solution conditions and pH values. Also, the stable modification of side chains allows a specific probe site to be identified using tandem mass spectrometry methods, while for deuterium exchange, typically the conformational change can only be localized to the specific peptide fragment. The drawback is that if a reactive side chain is not present in a particular peptide segment, there are no probes. However, the examination of side chains is complementary to the deuterium exchange method that examines backbone structure. Typically, 30% of the side-chain residues can be probed in a hydroxyl radical–mediated protein footprinting experiment (based on predicted sequence coverage of eight typically used residues – Met, Cys, Trp, Tyr, Phe, His, Pro, Leu). In addition, the use of mass spectrometry can uniquely identify the specific peptide fragments, such that the reactive sites can be easily identified.

To generate the limited dose required for footprinting and to quantitatively examine the reactivity of the specific peptides in question, a series of samples are exposed to variable doses, the samples digested and the individual peptides analyzed by HPLC and mass spectrometry (Fig. 2). Thus, a dose-response curve is generated for each peptide of interest; this generates a quantitative biophysical measure (based on the observed rate of modification) of the relative reactivity of the sites in the different peptides. Consistent and reliable quantitation, which is essential to footprinting, is provided by measuring the relative amounts of the modified and unmodified peptide products in the same experiment. Since the modifications are stable, it is relatively straightforward to use tandem mass spectrometric methods to specifically identify the amino acid positions of the protein that have been modified (bottom right of Fig. 2); these represent the probe sites for the analysis.

In the right-hand side of Fig. 2, where the reaction is carried out in the presence of a protein–protein complex, modification is suppressed at the site corresponding to the interacting surface, and unchanged at sites distal to the contact. The decrease of reactivity of peptides in the interface is quantitatively measured using the dose-response curves, a sample dose response for a peptide whose reactivity is suppressed

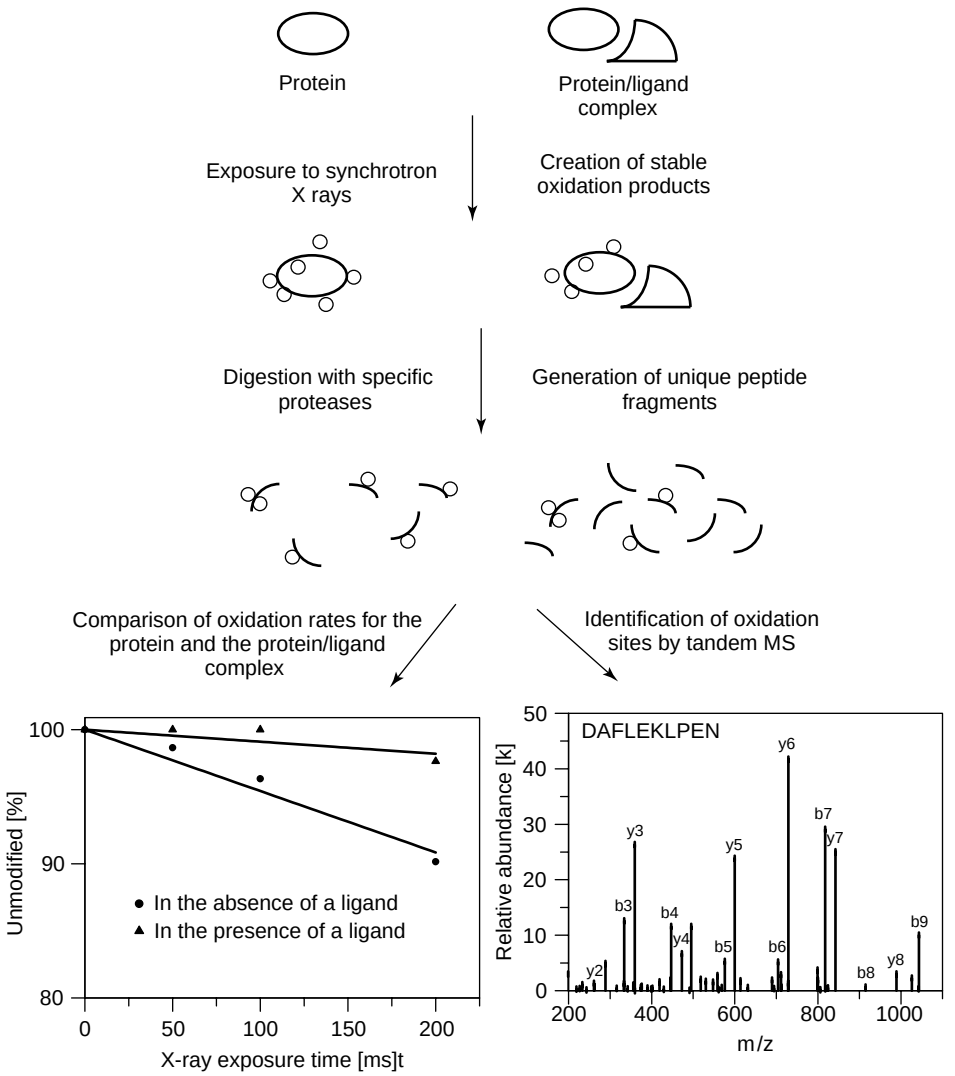

Fig. 2 A schematic representation of hydroxyl radical footprinting. A protein and its complex with a ligand are exposed to radiolysis or Fenton reagents, which produce hydroxyl radicals that modify the side chains of the protein. Subsequent to this X-ray exposure, the protein samples are digested with proteases, and MS is applied to quantitate the extent of modification. This quantitation provides information about the solvent accessibility of each peptide in both the isolated and complexed states. The particular modification sites are determined by tandem MS (MS/MS). In the dose-response example, a slower rate of modification is seen for the peptide when isolated from the complex compared to the case for the free protein. Thus, it is indicated that this peptide contains reactive side-chain residues influenced by the binding process. (Guan et al., 2002)

in the presence of the ligand is shown at the bottom left of the figure. It is critical for the method that the dose-response curves indicate a linear regime extrapolated to zero fraction modified – this assures that the reactivity of particular sites are not changing due to the oxidation process itself. This is a fundamental tenet of footprinting that the probed population is structurally intact. Also, the method of examining loss of the unmodified fraction emphasizes the interrogation of intact material. Although these particular refinements are unique to the hydroxyl radical–mediated protein footprinting approach, they are derived from a detailed knowledge and respect for safeguards that have evolved throughout the long history of development of "footprinting" research; such safeguards assure that the structural and biochemical information provided by footprinting methods are reliable. It must be emphasized, however, that footprinting provides only "local" information about the reactivity of the side-chain probes. Allosteric changes in conformation induced by ligand binding can also give rise to either protections (decreases in side-chain reactivity) or enhancements (increases in reactivity), depending on the induced conformational changes. This must be carefully borne in mind when interpreting data from these experiments, and examples of each of these are provided in the next section.

3 Applications

3.1 Quantitative Equilibrium Footprinting

Quantitative footprinting using protease cleavage has been developed to study calmodulin (CaM). Calmodulin is a small and highly acidic protein that consists of two homologous domains, which are connected by a flexible central linker. Each half-molecule (domain) contains a pair of calcium-binding sites in a helix–loop–helix or EF-hand motif. CaM binds calcium ions with positive cooperativity and generally activates target enzymes by binding to them in its calcium-saturated form and relieving their self-inhibition. The helix pair movements are accompanied by dramatic changes on the molecular surface of the protein, providing ideal targets for footprinting. Proteolytic footprinting has revealed the Ca^{2+} dependent changes of CaM and functional mutants. This work not only illustrates the susceptibility changes of different sites of CaM during the titration process but also provides direct information about the biological effects of mutation on CaM. For example, it was shown that the mutation E140Q reduces the binding affinity of sites at the C-domain, and sites in the N-domain saturate at a lower level of calcium than those in the C-domain in this mutant (Fig. 3(a)). The effect of this mutation is so severe that complete protection requires 100 μM calcium, 10 times higher than normally observed. Fluorescence studies support these data. By using the same approach, the Ca^{2+}-induced structural alternations of CaM in the presence of trifluoperazine (TFP), an antipsychotic drug that acts as a CaM antagonist, have also been probed. In the absence of TFP, chymotrypsin footprinting of a calcium titration of CaM shows several classes of cleavage probed by chymotrypsin: some positions are susceptible in the absence of calcium and become protected by calcium binding, whereas others are cleaved only at intermediate

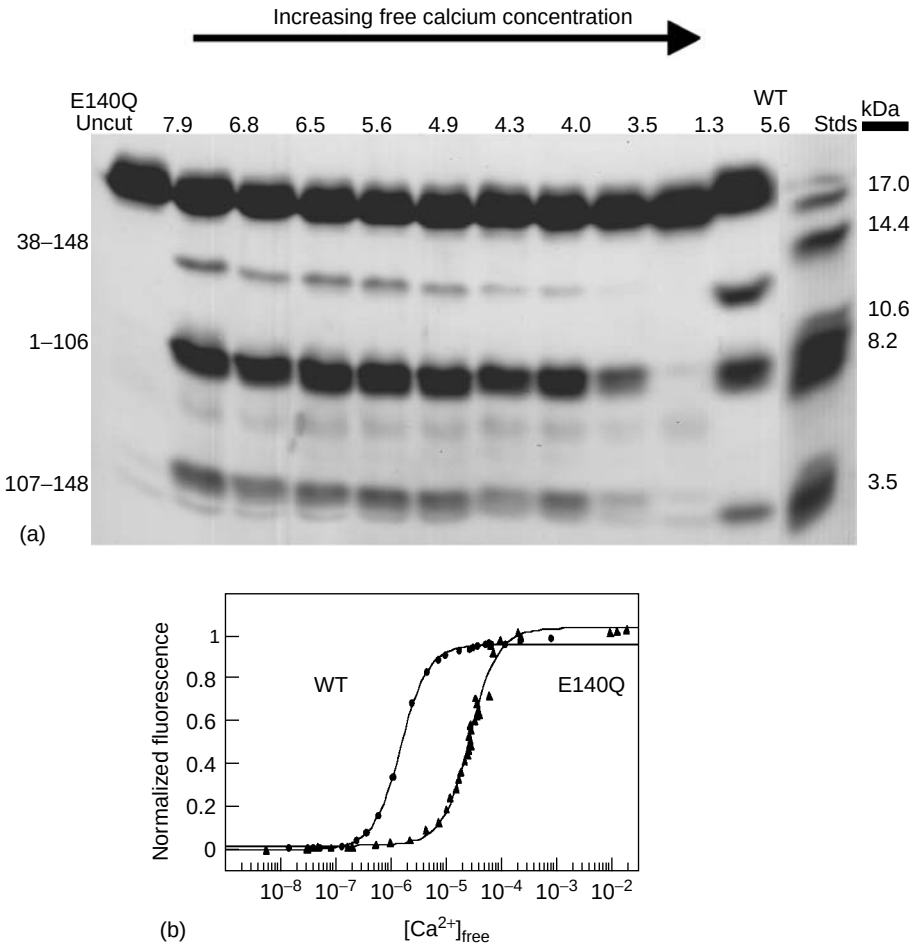

Fig. 3 Proteolytic footprinting can not only illustrate the susceptibility changes of different sites of a protein but can also provide direct information about the biological effects of mutation. Footprinting of CaM shows that a mutant E140 has significant site-specific effects on the binding of Ca^{2+} to CaM. This mutation dramatically reduces the binding affinity of sites at the C-domain, and the sites in the N-domain saturate at a lower level of calcium than those in the C-domain. (a) SDS-PAGE analysis of thrombin-cleaved E140Q-and WT-CaM in 50 mM HEPES, 100 mM NTA, 0.5 mM EGTA, pH 7.4, 25 °C. Lane 1: uncut CaM; Lane 11: products of thrombin footprinting of WT-CaM at pCa 5.6 for comparison of migration of fragments. Lane 12: MW standards. (b) Fluorescence analysis of calcium binding to WT- and E140Q-CaM in HEPES buffer (50 mM HEPES, 100 mM KCl, 0.5 uM NTA, 0.5 mM EGTA, pH7.4, 25 °C). (Shea et al., 2000)

levels of calcium binding. TFP protects CaM from proteolysis even under zero Ca^{2+} concentration, where the C-domains of CaM is normally susceptible. In the presence of TFP, a fragment that has a mass close to that of a half-molecule (a cleavage between the two domains) becomes protected as the concentration of calcium increases. It is necessary to conduct parallel proteolytic footprinting

titrations with multiple proteases to prove that any unusual susceptibility profiles reflect ligand-induced structure variations in the protein; thrombin and chymotrypsin are most frequently used as the proteases.

Modification based quantitative footprinting techniques also have been developed. Synchrotron X-ray protein footprinting has been used to monitor the equilibrium unfolding of proteins. After X-ray exposure, a quantitative map of the oxidized amino acids is generated to determine the relative solvent accessibility of the reactive residues. During folding, amino acid side chains of a protein undergo substantial changes in solvent accessibility that should attenuate the extent of their reaction with hydroxyl radicals in solution. By measuring the changes in reactivity and the degree of modification at amino acid side chains as a function of chemical denaturant concentration, the unfolding of apomyoglobin structure was monitored. It was shown that many of the residues within the hydrophobic core exhibit the expected, highly cooperative, two-state-unfolding behavior, while some residues within helix-G unfold locally and exhibit a lower free energy of stability. The results are fully consistent with those from fluorescence and NMR data.

This technique has been very recently extended to monitoring the structural alternations of gelsolin induced by Ca^{2+}. Gelsolin is a Ca^{2+}-dependent actin-regulatory protein composed of six homologous subdomains (S1–S6) that severs actin filaments and caps the fast-growing barbed end with high affinity. The inactive Ca^{2+}-free form of gelsolin exhibits a remarkably compact structure, in which the N-terminal (S1–S3) and C-terminal (S4–S6) halves are joined by a long linker and form an extensive interface (Fig. 4). Binding of Ca^{2+} to gelsolin and the Ca^{2+}-induced structural reorganization of gelsolin are associated with the actin-binding and actin-severing activity of gelsolin. However, only modest structural information is available regarding the Ca^{2+}-induced reorganization accompanying gelsolin activation. Synchrotron protein footprinting was applied to monitor the structural changes of gelsolin. Of the 81 tryptic peptides of gelsolin, 22 showed oxidation and 59 showed no oxidation upon radiolysis. Of the 22 oxidizable peptides, 5 exhibited a Ca^{2+}-dependent enhancement in oxidation rate, whereas 2 exhibited a Ca^{2+}-dependent protection. Ca^{2+}-titration isotherms monitoring the oxidation within residues 49–72 (within S1), 121–135 (S1), 162–166 (S2), and 722–748 (S6) indicate a three state activation process with an intermediate that was populated at a Ca^{2+} concentration of 1–5 μM that is competent for capping and severing activity. A second, structural transition with a midpoint of ~60–100 μM, where the accessibility of the above four peptides is further increased, is also observed (Fig. 5). The buried residues within the helical "latch" of S6 that contact an F-actin binding site on S2 and buried F-actin binding residues within S2 (Fig. 4), are unmasked in the sub-micromolar transition (Fig. 5). However, residues within S4 that are part of an extended β-sheet with S6 are revealed only in the subsequent transition at higher Ca^{2+} concentrations (Fig. 4); the disruption of this extended contact between S4 and S6 (and likely the analogous contact between S1 and S3) likely results in an extended structure permitting additional functions consistent with the fully activated gelsolin molecule. In summary, it has been demonstrated that protein footprinting is a quantitative biophysical

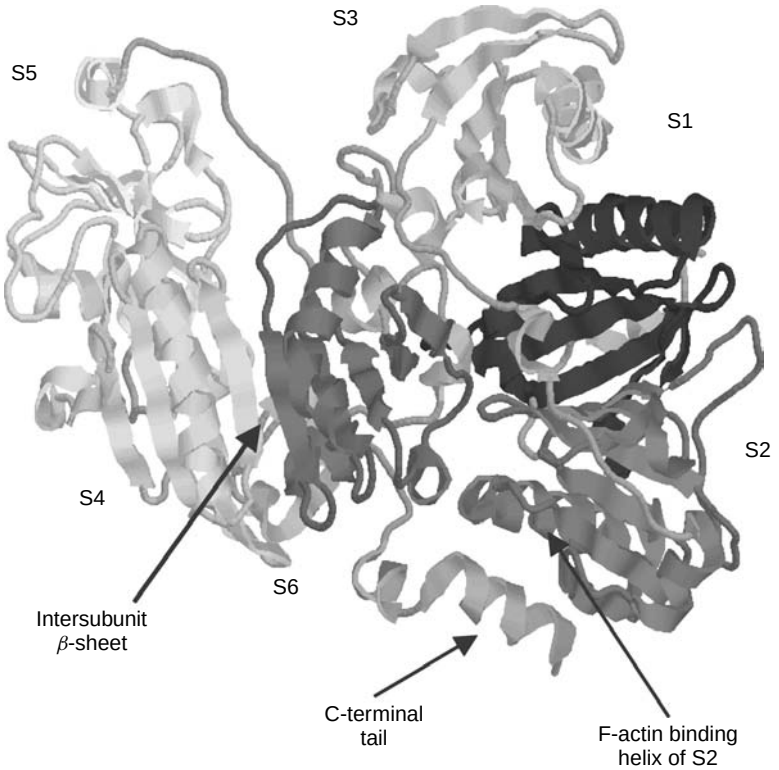

Fig. 4 Subunit structure of gelsolin (S1–S6). The inactive Ca^{2+}-free form of gelsolin exhibits a remarkably compact structure, in which the N-terminal (S1–S3) and C-terminal (S4–S6) halves are joined. The N-terminal half severs actin filaments through a mechanism that involves side binding through S2. The helical tail of S6 interacts with the F-actin-binding helix of S2 to make its actin-binding sites inaccessible in the absence of Ca^{2+}. Also, the intersubunit β-sheet between S4–S6 and S1–S3 must be sheared to provide for maximal Ca^{2+} loading. (Kiselar et al., 2003)

technique capable of probing differential responses of protein subdomains to ligand binding.

3.2
Footprinting of Complexes

Like gelsolin, cofilin is another actin-regulatory protein that binds to both monomeric and filamentous actin. Although crystal structures for the monomeric forms of both G-actin and cofilin have been described, the structure of the binary cofilin/G-actin is not available. Using synchrotron protein footprinting, the G-actin binding sites on cofilin has been probed. It was found that residues within peptides 4–20, 10–17, 83–96, 91–105, and 106–117 of cofilin exhibit strong protection upon G-actin binding. Specifically Leu13, Pro94, Met99, Leu108, and Leu112 define a binding surface consisting of helix α3 and N-terminal segments. These results are supported by previous

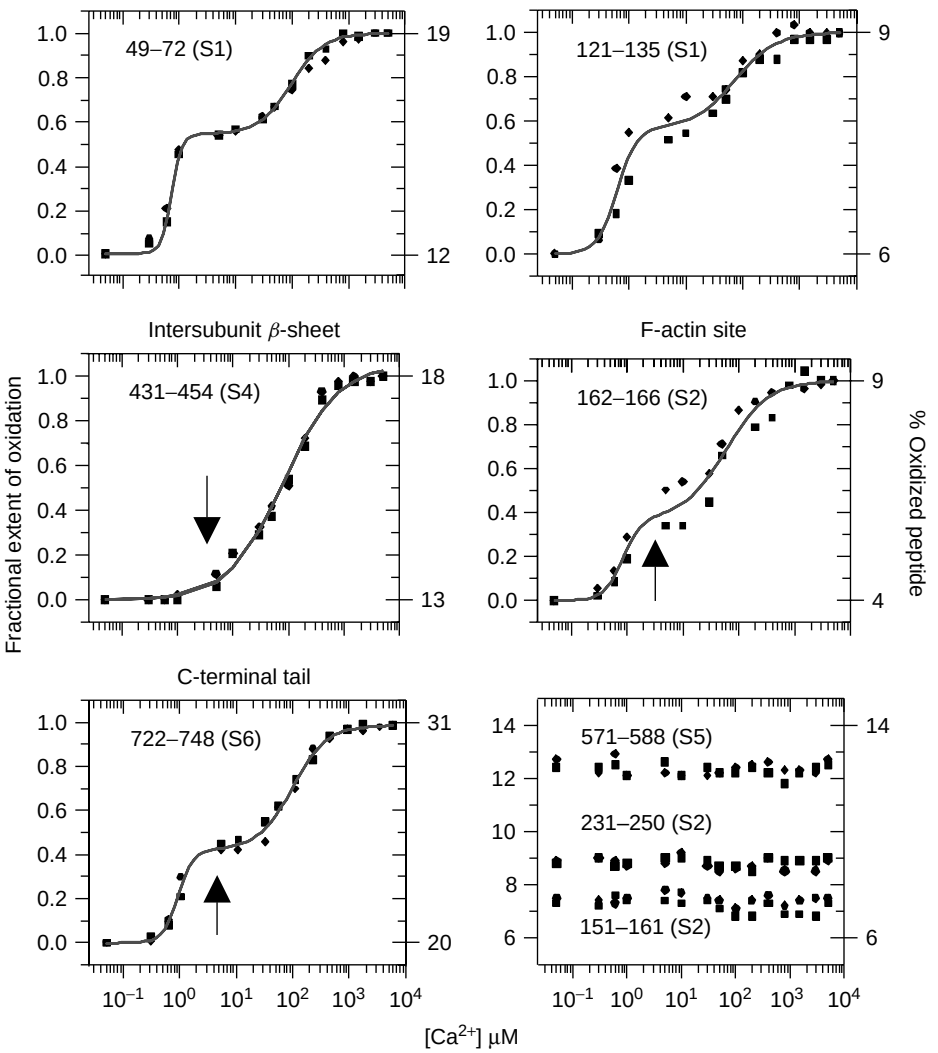

Fig. 5 Ca^{2+} titration isotherms indicating changes in oxidation extent for specific peptides after exposure of gelsolin to the X-ray beam for 80 ms. The percentage of oxidized peptide is shown on the right side of the *y*-axis; on the left side, the transition extent is indicated. The solid lines represent the fitting of the curves using appropriate one or two state Hill functions. Oxidized peptides within gelsolin show no changes in oxidation upon titration by Ca^{2+}. (Kiselar et al., 2003)

mutagenesis studies, which showed Arg96 and Lys98, Ser103 and Ser104 (corresponding to residues 119 and 120 from porcine cofilin) are important for both F- and G-actin binding. In addition, the footprinting studies indicate that peptide 106–117 is protected, consistent with recent work in which mutation of R109 to A and R110 to A were shown to reduce G-actin binding. Although the N-terminal

region of cofilin has been suggested to be essential for both G- and F-actin binding, including residues 1–4, the exact binding sites in the $\alpha 1$ helix have been unclear. It has been shown that the binding surface extends to Leu13, but does not include Leu21. In addition, the X-ray footprinting results suggest that the entire cofilin $\alpha 3$ segment contributes to the actin binding surface. Furthermore, the data suggests that the binding surface extends to Pro94, which is the residue prior to the start of the $\alpha 3$ helix.

Progress in footprinting protein complexes has also been made by designing unique sites in proteins at which a small reagent with cleavage activity can be tethered. First, a set of mutant proteins, each of which has a unique cysteine residue at a chosen location for conjugation to the cutting reagent, are prepared. The conjugated protein is then added to a target protein. Upon activation, the tethered cutting agent cleaves close-by segments of the target protein, producing two fragments. The cleavage fragments are separated by gel-electrophoresis, followed by Western blotting and visualized with antibodies against the termini of the target protein. An alternative to the single-cysteine tethered cutting approach involves tagging surface-accessible lysine side chains. These tethered cutting reagents have been used to study the binding sites of σ factors on RNA polymerase (RNAP). The RNAP of *Escherichia coli* is a member of a conserved family of multisubunit RNA polymerases. Core RNAP, which is a noncovalent complex of four proteins ($\alpha_2\beta\beta'$) with 380 kDa, binds to one of seven different σ proteins to form the holoenzyme ($\alpha_2\beta\beta'\sigma$) for RNA synthesis. Tethered cutting reagents attached to the engineered single-cysteine mutants of $\sigma 70$ have been used to identify sites on $\alpha_2\beta\beta'$ near each cutter. The cut sites on *E. coli* RNAP subunits β and β' made by reagents tethered to several different sites on $\sigma 70$ neatly outline a σ-core binding site and the orientation of $\sigma 70$ relative to the core RNAP subunits. Targeted protein footprinting by using lysine-conjugated $\sigma 70$ was shown to be consistent with that using cysteine-$\sigma 70$ mutants. The interaction between *E. coli* $\sigma 70$ and DNA in transcription complexes has also been studied by using tethered cutting reagents in Meares group. As discussed previously, the tethering process localizes the cleavage to the site of protein–protein interaction.

4 Future Prospects

Future development of protein footprinting technologies are focused on examining the structure of large macromolecular complexes, and probing the dynamics of macromolecules. Purification of biophysical quantities of stoichiomeric complexes has allowed the detailed examination of RNA transcription complexes of 380 kDa, and cytoskeletal assemblies like Arp 2/3 of 270 kDa using protein footprinting. Molecular cross-linking strategies will provide useful additional information as well. Other areas of development include using the flux of synchrotron radiation to provide time-resolved probes of protein dynamics, as has been seen in the studies of nucleic acids. An excellent example of the kinds of experiments that would be of interest would explore the timescales of activation of Ca^{+2} sensitive sites within gelsolin. In addition, detailed examination of side-chain chemistries will expand the sequence coverage of hydroxyl radical–protein footprinting. As availability

of new sequences for protein expression increases, structural and proteomic technologies like footprinting will continue to play a key role in understanding protein structure, function, and protein interactions.

See also DNA–Protein Interactions; Footprinting Methods to Examine the Structure and Dynamics of Nucleic Acids; Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Brenowitz, M., Chance, M.R., Dhavan, G., Takamoto, K. (2002) Probing the structural dynamics of nucleic acids by quantitative time-resolved and equilibrium hydroxyl radical 'footprinting', *Curr. Opin. Struct. Biol.* **12**, 648–653.

Datwyler, S.A, Meares, C.F. (2000) Protein-protein interactions mapped by artificial proteases: where σ factors binds to RNA polymerase, *TIBS* **25**, 408–414.

Datwyler, S.A., Meares, C.F. (2001) Artificial iron-dependent proteases, *Met. Ions Biol. Syst.* **38**, 213–254.

Heyduk, T., Baichoo, N., Heyduk, E. (2001a) Hydroxyl radical footprinting of proteins using metal ion complexes, *Met. Ions Biol. Syst.* **38**, 255–287.

Maleknia, S.D., Ralston, C.Y., Brenowitz, M.D., Doward, K.M., Chance, M.R. (2001) Determination of macromolecular folding and structure by synchrotron x-ray radiolysis techniques. *Anal. Biochem.* **289**, 103–115.

Shea, M.A., Sorensen, B.R., Pedigo, S., Verhoeven, A.S. (2000) Proteolytic footprinting titrations for estimating ligand-binding constants and detecting pathways of conformational switching of calmodulin, *Methods Enzymol.* **323**, 254–301.

Sheshberadaran, H., Payne, L.G. (1989) Protein footprinting method for studying antigen-antibody interactions and epitope mapping, *Methods Enzymol.* **178**, 746–765.

Primary Literature

Chance, M.R. (2001) Unfolding of apomyoglobin examined by synchrotron footprinting, *Biochem. Biophys. Res. Commun.* **287**, 614–621.

Chance, M.R., Sclavi, B., Woodson, S.A., Brenowitz, M. (1997) Examining the conformational dynamics of macromolecules with time-resolved synchrotron X-ray 'footprinting', *Structure* **5**, 865–869.

Galas, D.J., Schmitz, A. (1978) DNAse footprinting: a simple method for the detection of protein-DNA binding specificity, *Nucleic Acids Res.* **5**, 3157–3170.

Garrison, W.M. (1987) Reaction mechanisms in the radiolysis of peptides, polypeptides, and proteins, *Chem. Rev.* **87**, 381–398.

Hanai, R. (1996) Protein footprinting, *Tanpakushitsu-Kakusan-Kouso (Protein Acid Enzyme)* **41**, 929–933.

Hanai, R. (1999) Footprinting proteins, in: Creighton, T.E. (Ed.) *Encyclopedia of Molecular Biology*, John Wiley, New York, pp. 934–936.

Hanai, R., Wang, J.C. (1994) Protein footprinting by the combined use of reversible and irreversible lysine modifications, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11904–11908.

Heyduk, E., Baichoo, N., Heyduk, T. (2001b) Interaction of the alpha-subunit of Escherichia coli RNA polymerase with DNA: rigid body nature of the protein-DNA contact, *J. Biol. Chem.* **276**, 44598–445603.

Kiselar, J.G., Janmey, P.A., Almo, S.C., Chance, M.R. (2003) Visualizing the Ca^{2+}-dependent activation of gelsolin by using synchrotron footprinting, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3942–3947.

Kiselar, J.G., Maleknia, S.D., Chance, M.R. (2002) Hydroxyl radical probe of protein surfaces using synchrotron X-ray radiolysis and mass spectrometry, *Int. J. Radiat. Biol.* **78**, 101–114.

Maleknia, S.D., Brenowitz, M., Chance, M.R. (1999) Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry, *Anal. Chem.* **71**, 3965–3973.

Richards, F.M., Lamed, R., Wynn, R., Patel, D., Olack, G. (2000) Methylene as a possible universal footprinting reagent that will include hydrophobic surface areas: overview and feasibility: properties of diazirine as a precursor, *Protein Sci.* **9**, 2506–2517.

Takamoto, K., Chance, M.R. (2003) Footprinting methods to Examine the Structure and Dynamics of Nucleic Acids, in: Meyers, R. (Ed.) *Encyclopedia of Molecular Cell Biology and Molecular Medicine*, 2nd edition, Wiley, in press.

Traviglia, S.L, Datwyler, S.A, Meares, C.F. (1999) Mapping protein-protein interactions with a library of tethered cutting reagents: the binding site of σ70 on Escherichia coli rNA polymerases, *Biochemistry* **38**, 4259–4265.

Tullius, T.D., Dombroski, B.A., Churchill, M.E., Kam, L. (1987) Hydroxyl radical footprinting: a high-resolution method for mapping protein-DNA contacts, *Methods Enzymol.* **155**, 537–558.

Zhong, M., Lin, L., Kallenbach, N.R. (1995) A method for probing the topography and interactions of proteins: footprinting of myoglobin, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2111–2115.

References

Abbott, A. (2002) The society of proteins, *Nature* **417**, 894–896.

Boulton, S.J., Gartner, A., Reboul, J., Vaglio, P., Dyson, N., Hill, D.E., Vidal, M. (2002) Combined functional genomic maps of the C. elegans DNA damage response, *Science* **295**, 127–131.

Burtnick, L.D., Koepf, E.K., Grimes, J., Jones, E.Y., Stuart, D.I., McLauqhlin, P.J., Robinson, R.C. (1997) The crystal structure of plasma gelsolin: implications for actin severing, capping, and nucleation, *Cell* **90**, 661–670.

Dhavan, G.M., Crothers, D.M., Chance, M.R., Brenowitz, M. (2002) Concerted binding and bending of DNA by Escherichia coli integration host factor, *J. Mol. Biol.* **315**, 1027–1037.

Faga, L.A., Sorensen, B.R., VanScyoc, W.S., Shea, M.A. (2003) Basic interdomain boundary residues in calmodulin decrease calcium affinity of sites I and II by stabilizing helix-helix interactions, *Proteins* **50**, 381–391.

Garrison, W.M., Jayko, M.E., Bennett, W. (1962) Radiation induced oxidation of protein in aqueous solution, *Radiat. Res.* **16**, 483–502.

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., Superti-Furga, G. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes, *Nature* **415**, 141–147.

Gilbert, W. (1972) The lac repressor and the lac operator, *CIBA Foundation Symp. 7: Polymerization in Biological Systems*, Elsevier Publishing, Amsterdam, p. 245–259.

Guan, J.-Q., Vorobiev, S., Almo, S.C., Chance, M.R. (2002) Mapping the G-action binding surface of cofilin using synchrotron protein footprinting, *Biochemistry* **41**, 5765–5775.

Ha, J.H., Loh, S.N. (1998) Changes in side-chain packing during apomyoglobin folding characterized by pulsed thiol-disulfide exchange, *Nat. Struct. Biol.* **5**, 730–737.

Heyduk, E., Heyduk, T. (1994) Mapping protein domains involved in macromolecular interactions: a novel protein footprinting approach, *Biochemistry* **33**, 9643–9650.

Hoofnagle, A.N., Resing, K.A., Ahn, N.G. (2003) Protein analysis by hydrogen exchange mass spectrometry, *Annu. Rev. Biophys. Biomol. Struct.* **32**, 1–25.

Horinishi, H., Hachimori, Y., Kurihara, K., Shibata, K. (1964) *Biochem. Biophys. Acta* **86**, 477.

Humayun, Z., Kleid, D., Ptashne, M. (1977) Sites of contact between lambda operators and lambda repressor, *Nucleic Acids Res.* **4**, 1595–1607.

Janmey, P.A., Stossel, T.P. (1987) Modulation of gelsolin function by phosphatidylinositol 4,5-bisphosphate, *Nature* **325**, 362–364.

Jaren, O.R., Kranz, J.K., Sorensen, B.R., Wand, A.J., Shea, M.A. (2002) Calcium-induced conformational switching of Paramecium calmodulin provides evidence for domain coupling, *Biochemistry* **41**, 14158–14166.

Jay, D.G. (1984) A general procedure for the end labeling of proteins and positioning of

amino acids in the sequence, *J. Biol. Chem.* **259**, 15572–15578.

Kuboniwa, H., Tjandra, N., Grzesiek, S., Ren, H., Klee, C.B., Bax, A. (1995) Solution structure of calcium-free calmodulin, *Nat. Struct. Biol.* **2**, 768.

Lappalainen, P., Fedorov, E.V., Fedorov, A.A., Almo, S.C., Drubin, D.G. (1997) Essential functions and actin-binding surfaces of yeast cofilin revealed by systematic mutagenesis, *EMBO J.* **16**, 5520–5530.

Mirzabekov, A.D., Kolchinsky, A.M. (1974) Localization of some molecules within the grooves of DNA by modification of their complexes with dimethyl sulphate, *Mol. Biol. Rep.* **1**, 385–390.

Mirzabekov, A.D., Melnikova, A.F. (1974) Localization of chromatin proteins within DNA grooves by methylation of chromatin with dimethyl sulphate, *Mol. Biol. Rep.* **1**, 379–384.

Moriyama, K., Yonezawa, N., Sakai, H., Yahara, I., Nishida, E. (1992) Mutational analysis of an actin-binding site of cofilin and characterization of chimeric proteins between cofilin and destrin, *J. Biol. Chem.* **267**, 7240–7244.

Moriyama, K, Yahara, I. (1999) Two activities of cofilin, severing and accelerating directional depolymerization of actin filaments, are affected differentially by mutations around the actin-binding helix, *EMBO J.* **18**, 6752–6761.

Ojala, P.J., Paavilainen, V., Lappalainen, P. (2001) Identification of yeast cofilin residues specific for actin monomer and PIP2 binding, *Biochemistry* **40**, 15562–15569.

Owens, J.T., Miyake, R., Murakami, K., Chmura, A.J., Fujita, N., Ishihama, A., Meares, C.F. (1998a) Mapping the sigma70 subunit contact sites on Escherichia coli RNA polymerase with a sigma70-conjugated chemical protease, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6021–6026.

Owens, J.T., Chmura, A.J., Murakami, K., Fujita, N., Ishihama, A., Meares, C.F. (1998b) Mapping the promoter DNA sites proximal to conserved regions of σ70 in an Escherichia coli rNA polymerase-lacUV5 open promoter complex, *Biochemistry* **37**, 7670–7675.

Rashidzadeh, H., Khrapunov, S., Chance, M.R., Brenowitz, M. (2003) Solution structure and interdomain interactions of the Saccharomyces cerevisiae ''TATA Binding Protein'' (TBP) probed by radiolytic protein footprinting, *Biochemistry* **42**, 3655–3665.

Reboul, J., Vaglio, P., Rual, J.F., Lamesch, P., Martinez, M., Armstrong, C.M., Li, S., Jacotot, L., Bertin, N., Janky, R., Moore, T., Hudson, J.R. Jr., Hartley, J.L., Brasch, M.A., Vandenhaute, J., Boulton, S., Endress, G.A., Jenna, S., Chevet, E., Papasotiropoulos, V., Tolias, P.P., Ptacek, J., Snyder, M., Huang, R., Chance, M.R., Lee, H., Doucette-Stamm, L., Hill, D.E., Vidal, M. (2003) C. elegans ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression, *Nat. Genet.* **34**, 35–41.

Robinson, R.C., Turbedsky, K., Kaiser, D.A., Marchand, J.B., Higgs, H.N., Choe, S., Pollard, T.D. (2001) Crystal structure of Arp2/3 complex, *Science* **294**, 1679–1684.

Schmitz, A., Galas, D.J. (1979) The interaction of RNA polymerase and lac repressor with the lac control region, *Nucleic Acids Res.* **6**, 111–137.

Sclavi, B., Woodson, S., Sullivan, M., Chance, M.R., Brenowitz, M. (1998) Following the folding of RNA with time-resolved synchrotron X-ray footprinting, *Methods Enzymol.* **295**, 379–402.

Sharp, J.S., Becker, J.M., Hettich, R.L. (2003) Protein surface mapping by chemical oxidation: structural analysis by mass spectrometry, *Anal. Biochem.* **313**, 216–225.

Sheshberadaran, H., Payne, L.G. (1988) Protein antigen-monoclonal antibody contact sites investigated by limited proteolysis of monoclonal antibody-bound antigen: protein ''footprinting'', *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1–5.

Tjandra, N., Kuboniwa, H., Ren, H., Bax, A. (1995) Rotational dynamics of calcium-free calmodulin studied by 15N-NMR relaxation measurements, *Eur. J. Biochem.* **230**, 1014–1024.

Trester-Zedlitz, M., Kamada, K., Burley, S.K., Fenyo, D., Chait, B.T., Muir, T.W. (2003) A modular cross-linking approach for exploring protein interactions, *J. Am. Chem. Soc.* **125**, 2416–2425.

Uchida, T., Takamoto, K., He, Q., Chance, M.R., Brenowitz, M. (2003) Multiple monovalent ion-dependent pathways for the folding of the L-21 tetrahymena thermophila ribozyme, *J. Mol. Biol.* **328**, 463–478.

VanScyoc, W.S., Sorensen, B.R., Rusinova, E., Laws, W.R., Ross, J.B., Shea, M.A. (2002) Calcium binding to calmodulin mutants monitored by domain-specific intrinsic phenylalanine and tyrosine fluorescence, *Biophys. J.* **83**, 2767–2780.

Wang, L., Pan, H., Smith, D.L. (2002) Hydrogen exchange-mass spectrometry: optimization of digestion conditions, *Mol. Cell Proteomics* **1**, 132–138.

Yin, H.L., Stossel, T.P. (1979) Control of cytoplasmic actin gel-sol transformation by gelsolin, a calcium-dependent regulatory protein, *Nature* **281**, 583–586.

Zhang, M., Tanaka, T., Ikura, M. (1995) Calcium-induced conformation transition revealed by the solution structure of apo calmodulin, *Nat. Struct. Biol.* **2**, 758–767.

Fragile Sites

Grant R. Sutherland[1,3] *and Elizabeth Baker*[2,3]
[1] *Women's and Children's Hospital, Adelaide, Australia*
[2] *US Labs, Irvine, CA, USA*
[3] *University of Adelaide, Adelaide, Australia*

1 Definition 571

2 Nomenclature 571

3 Classification 571

4 Molecular Genetics 571
4.1 Group 1 571
4.2 Groups 2 and 3 577
4.3 Group 4 577

5 Cytogenetics 577

6 Induction of Fragile Sites 579
6.1 Group 1 579
6.2 Group 2 580
6.3 Group 3 580
6.4 Group 4 580
6.5 Group 5 580
6.6 Group 6 580

7 Inheritance 580

8 Clinical Significance 582
8.1 The Fragile X Syndromes 582

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

8.2 Other Fragile Sites 582
8.3 Cancer 582

Acknowledgments 583

Bibliography 583
Books and Reviews 583
Primary Literature 583

Keywords

Autosome
Any chromosome other than the sex (X and Y) chromosomes.

Dynamic Mutation
A multistep mutational process such that the chance of each step occurring is modified by the immediately preceding step.

Fragile X Syndrome
The most common form of familial mental retardation. Affected patients also have subtle dysmorphic features.

Jacobsen Syndrome
A mental retardation/congenital malformation syndrome caused by the loss of part of the long arm of one number 11 chromosome.

Polymorphism
The existence in a population of more than one version of the DNA sequence at any place in the genome such that the rarest version cannot be maintained by mutation alone.

Trinucleotide Repeat
A DNA sequence composed of three nucleotides repeated in tandem a number of times.

1 Definition

Fragile sites are specific points on chromosomes that show nonrandom gaps or breaks when the cells from which the chromosomes have been prepared have been exposed to a specific chemical agent or condition of tissue culture. Hence, a fragile site is an area of chromatin, which is not compacted when viewed in mitosis.

2 Nomenclature

There are two forms of nomenclature used for fragile sites. The cytogenetic nomenclature uses *fra* followed by the chromosome band in which the fragile site is present in parentheses. The rare fragile site in band 23.3 on the long (q) arm of chromosome 11 is thus fra(11)(q23.3). Each fragile site also has a gene symbol and this is *FRA*, the chromosome number, and an assigned letter of the alphabet. Fra(11)(q23.3) is *FRA11B*. The gene symbol is a locus name. Alleles at the locus can be specified following an *, the wild-type allele as N, and the mutant allele as R (for rare) or C (for common), followed by the chromosome band. Fra(11)(q23.3) in gene nomenclature would be *FRA11B*N* for the wild-type allele and *FRA11B*RQ233* for the mutant allele. In practice, the gene symbol is frequently used to specify the mutant allele and the wild-type allele is not mentioned.

3 Classification

Fragile sites are classified on the basis of their frequency in the population and the conditions of tissue culture under which they are expressed. Rare fragile sites are seen at a specific location on 1 in 40 chromosomes at most. Common fragile sites are probably present on all chromosomes and as such are normal components of chromosome structure.

Most fragile sites are not expressed spontaneously but must be induced. When they are expressed spontaneously, the proportion of metaphases in which they are seen can usually be greatly increased by appropriate conditions of induction. The classification, based upon population frequency and induction conditions, is shown in Table 1 and all known human fragile sites are listed in Table 2.

4 Molecular Genetics

4.1 Group 1

Only five of the group 1 fragile sites have been cloned and details of these are shown in Table 3.

The common features of these fragile sites, which are likely to be common to all the group 1 fragile sites include the following:

1. The fragile site is a dynamic mutation of a naturally occurring CCG trinucleotide repeat.
2. The CCG repeat is polymorphic in the normal population.
3. Increases in the number of copies of the repeating unit, above the normal range leads to a premutation. Individuals with the premutation do not express a fragile site. Premutation carriers of the fragile sites on the X chromosome associated with mental handicap do not suffer this problem. But premutation

Tab. 1 Classification of fragile sites.

Rare:	Group 1:	Folate sensitive fragile sites ($n = 22$)
	Group 2:	Distamycin A inducible a) Also inducible by BrdU ($n = 2$, *FRA16B*, *FRA17A*) b) Not inducible by BrdU, and only recorded in the Japanese ($n = 3$, *FRA8E*, *FRA11I*, *FRA16E*)
	Group 3:	BrdU requiring ($n = 2$, *FRA10B*, *FRA12C*)
Common:	Group 4:	Aphidicolin-inducible ($n = 76$)
	Group 5:	5-azacytidine inducible ($n = 4$)
	Group 6:	BrdU-inducible ($n = 7$)

Tab. 2 The 116 fragile sites in the human genome.

Gene symbol	*Chromosome band location*	*Type*	*Group*
FRA1A	1p36	Common, aphidicolin	4
FRA1B	1p32	Common, aphidicolin	4
FRA1C	1p31.2	Common, aphidicolin	4
FRA1L	1p31	Common, aphidicolin	4
FRA1D	1p22	Common, aphidicolin	4
FRA1M	**p21.3**	**Rare, folic acid**	**1**
FRA1E	1p21.2	Common, aphidicolin	4
FRA1J	1q12	Common, 5-azacytidine	5
FRA1F	1q21	Common, aphidicolin	4
FRA1G	1q25.1	Common, aphidicolin	4
FRA1K	1q31	Common, aphidicolin	4
FRA1H	1q42	Common, 5-azacytidine	5
FRA1I	1q44	Common, aphidicolin	4
FRA2C	2p24.2	Common, aphidicolin	4
FRA2D	2p16.2	Common, aphidicolin	4
FRA2E	2p13	Common, aphidicolin	4
FRA2A	**2q11.2**	**Rare, folic acid**	**1**
FRA2B	**2q13**	**Rare, folic acid**	**1**
FRA2F	2q21.3	Common, aphidicolin	4
FRA2K	**2q22.3**	**Rare, folic acid**	**1**
FRA2G	2q31	Common, aphidicolin	4
FRA2H	2q32.1	Common, aphidicolin	4
FRA2I	2q33	Common, aphidicolin	4
FRA2J	2q37.3	Common, aphidicolin	4
FRA3A	3p24.2	Common, aphidicolin	4
FRA3B	3p14.2	Common, aphidicolin	4
FRA3D	3q25	Common, aphidicolin	4
FRA3C	3q27	Common, aphidicolin	4

Tab. 2 (*continued*)

Gene symbol	*Chromosome band location*	*Type*	*Group*
FRA4A	4p16.1	Common, aphidicolin	4
FRA4D	4p15	Common, aphidicolin	4
FRA4B	4q12	Common, BrdU	6
FRA4C	4q31.1	Common, aphidicolin	4
FRA5E	5p14	Common, aphidicolin	4
FRA5A	5p13	Common, BrdU	6
FRA5B	5q15	Common, BrdU	6
FRA5D	5q15	Common, aphidicolin	4
FRA5F	5q21	Common, aphidicolin	4
FRA5C	5q31.1	Common, aphidicolin	4
FRA5G	**5q35**	**Rare, folic acid**	**1**
FRA6B	6p25.1	Common, aphidicolin	4
FRA6A	**6p23**	**Rare, folic acid**	**1**
FRA6C	6p22.2	Common, aphidicolin	4
FRA6D	6q13	Common, BrdU	6
FRA6G	6q15	Common, aphidicolin	4
FRA6F	6q21	Common, aphidicolin	4
FRA6E	6q26	Common, aphidicolin	4
FRA7B	7p22	Common, aphidicolin	4
FRA7C	7p14.2	Common, aphidicolin	4
FRA7D	7p13	Common, aphidicolin	4
FRA7A	**7p11.2**	**Rare, folic acid**	**1**
FRA7J	7q11	Common, aphidicolin	4
FRA7E	7q21.2	Common, aphidicolin	4
FRA7F	7q22	Common, aphidicolin	4
FRA7G	7q31.2	Common, aphidicolin	4
FRA7H	7q32.3	Common, aphidicolin	4
FRA7I	7q36	Common, aphidicolin	4
FRA8B	8q22.1	Common, aphidicolin	4
FRA8A	**8q22.3**	**Rare, folic acid**	**1**
FRA8C	8q24.1	Common, aphidicolin	4
FRA8E	**8q24.1**	**Rare, distamycin A**	**2b**
FRA8D	8q24.3	Common, aphidicolin	4
FRA9A	**9p21**	**Rare, folic acid**	**1**
FRA9C	9p21	Common, BrdU	6
FRA9F	9q12	Common, 5-azacytidine	5
FRA9D	9q22.1	Common, aphidicolin	4
FRA9B	**9q32**	**Rare, folic acid**	**1**
FRA9E	9q32	Common, aphidicolin	4
FRA10G	10q11.2	Common, aphidicolin	4
FRA10C	10q21	Common, BrdU	6
FRA10D	10q22.1	Common, aphidicolin	4
FRA10A	**10q23.3**	**Rare, folic acid**	**1**
FRA10B	**10q25.2**	**Rare, BrdU**	**3**
FRA10E	10q25.2	Common, aphidicolin	4

(*continued overleaf*)

Tab. 2 (*continued*)

Gene symbol	*Chromosome band location*	*Type*	*Group*
FRA10F	10q26.1	Common, aphidicolin	4
FRA11C	11p15.1	Common, aphidicolin	4
FRA11I	**11p15.1**	**Rare, distamycin A**	**2b**
FRA11D	11p14.2	Common, aphidicolin	4
FRA11E	11p13	Common, aphidicolin	4
FRA11H	11q13	Common, aphidicolin	4
FRA11A	**11q13.3**	**Rare, folic acid**	**1**
FRA11F	11q14.2	Common, aphidicolin	4
FRA11B	**11q23.3**	**Rare, folic acid**	**1**
FRA11G	11q23.3	Common, aphidicolin	4
FRA12A	**12q13.1**	**Rare, folic acid**	**1**
FRA12B	12q21.3	Common, aphidicolin	4
FRA12E	12q24	Common, aphidicolin	4
FRA12D	**12q24.13**	**Rare, folic acid**	**1**
FRA12C	**12q24.2**	**Rare, BrdU**	**3**
FRA13A	13q13.2	Common, aphidicolin	4
FRA13B	13q21	Common, BrdU	6
FRA13C	13q21.2	Common, aphidicolin	4
FRA13D	13q32	Common, aphidicolin	4
FRA14B	14q23	Common, aphidicolin	4
FRA14C	14q24.1	Common, aphidicolin	4
FRA15A	15q22	Common, aphidicolin	4
FRA16A	**16p13.11**	**Rare, folic acid**	**1**
FRA16E	**16p12.1**	**Rare, distamycin A**	**2b**
FRA16B	**16q22.1**	**Rare, distamycin A**	**2a**
FRA16C	16q22.1	Common, aphidicolin	4
FRA16D	16q23.2	Common, aphidicolin	4
FRA17A	**17p12**	**Rare, distamycin A**	**2a**
FRA17B	17q23.1	Common, aphidicolin	4
FRA18A	18q12.2	Common, aphidicolin	4
FRA18B	18q21.3	Common, aphidicolin	4
FRA19B	**19p13**	**Rare, folic acid**	**1**
FRA19A	19q13	Common, 5-azacytidine	5
FRA20B	20p12.2	Common, aphidicolin	4
FRA20A	**20p11.23**	**Rare, folic acid**	**1**
FRA22B	22q12.2	Common, aphidicolin	4
FRA22A	**22q13.1**	**Rare, folic acid**	**1**
FRAXB	Xp22.31	Common, aphidicolin	4
FRAXC	Xq22.1	Common, aphidicolin	4
FRAXD	Xq27.2	Common, aphidicolin	4
FRAXA	**Xq27.3**	**Rare, folic acid**	**1**
FRAXE	**Xq28**	**Rare, folic acid**	**1**
FRAXF	**Xq28**	**Rare, folic acid**	**1**

Source: Modified from Sutherland (1993, 2003) where detailed references to each fragile site can be found.

Tab. 3 Cloned rare fragile sites.

Site	*Repeat copy number*			*Associated disease*
	Normal alleles	*Premutation*	*Full mutation*	
Group 1				
FRAXA	6–55	55–230	230–2000	Fragile X syndrome
FRAXE	6–25	25–200	?>200	Mild mental retardation
FRAXF	6–29	?	?>200	None
FRA16A	17–50	100	?>200	None
FRA11B	11–17	85	?>200	Possible risk of offspring with Jacobsen syndrome
Group 2a				
FRA16B	7–12	?	450–2000	None
Group 3				
FRA10B	Complex	?	Up to ~500	None

carrier females are at increased risk of premature ovarian failure and males of a late onset (>50 years) tremor/ataxia neurodegenerative syndrome. As the number of copies of the repeating unit increases further, the DNA in the whole area is subjected to CpG methylation. Once this happens, the fragile site can be induced to appear and any gene in which the repeat is located ceases to be transcribed.

For *FRAXA*, the upper limit of copy number in normal chromosomes is about 55; from 55 to about 230 copies constitutes a premutation and beyond 230 is a full mutation. The number of copies changes from generation to generation, usually increasing when transmitted by women but not changing much when transmitted by males (Fig. 4).

The process by which the polymorphic trinucleotide repeat increases from within the normal range to give a fragile site is a multistep one known as *dynamic mutation*. The probability of a trinucleotide repeat changing in length is a function of the number of copies of the repeat that are present without interruption or imperfection. In the normal population, most alleles of the trinucleotide repeat have some imperfections. It is the alleles with the longest stretches of perfect repeat that form a reservoir of alleles in the normal population, which can proceed via dynamic mutation to fragile sites.

Since the length of repeat determines the rate at which further increases occur, the chance of a premutation converting to a full mutation increases with the size of the premutation alleles. This has been studied in most detail for *FRAXA*. For this fragile site, only women have offspring with full mutations. The upper limit of normal copy number is about 55. For women with premutations of less than 60 copies, the risk of having a child with a full mutation is close to zero. (The smallest alleles observed to go to a full mutation in a single generation have had 59 copies of

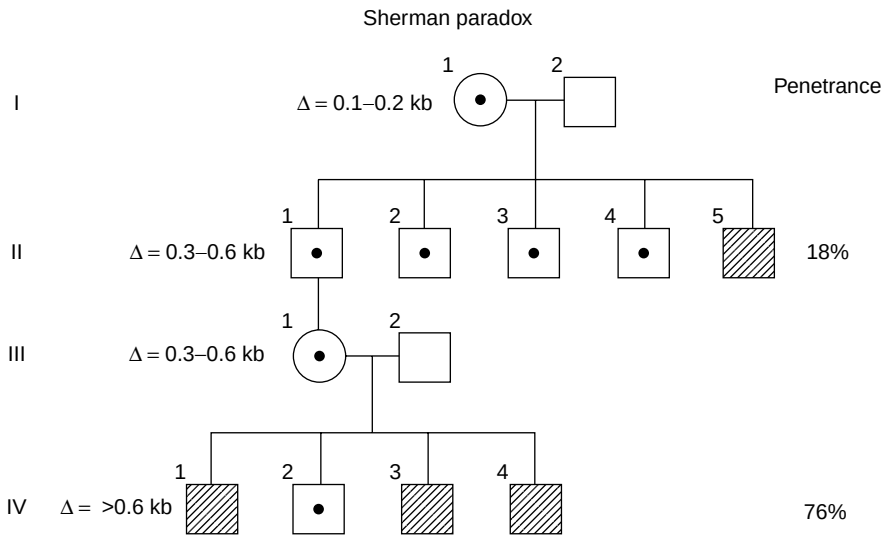

Fig. 1 The Sherman paradox. A hypothetical pedigree is used to illustrate the different levels of penetrance of the fragile X mutation depending upon the position of the carrier in the pedigree. The fragile X carrier I-1 has an 18% chance (approximately one in five risk) that her fragile X sons are affected, while the fragile X carrier III-1 has a 76% chance (three in four risk) that her fragile X sons are affected. The size of the amplification of the fragile X p(CCG)n repeat is given as Δ, that is, for a normal individual Δ = 0.0 kb. Values for Δ indicated are those expected hypothetically for individual 1 in each generation. Hatched squares indicate individual with fragile X syndrome, black dots indicate fragile X transmitters (from Richards and Sutherland 1992).

the repeat.) The chance of 60 to 80 copies converting to a full mutation is about 30%, for 80 to 90 copies it is about 80%, and for greater than 90 copies it is virtually 100%.

The limited numbers of normal alleles with long stretches of perfect repeat that form the reservoir of mutable alleles give rise to the phenomenon of linkage disequilibrium. The distribution of haplotypes around the fragile site locus is very different on normal and fragile site chromosomes.

The place and time of expansion of a full mutation from a premutation is uncertain. Most evidence points to expansion during oogenesis with a period of instability of the expanded repeat during early embryonic development during which period large repeats, in particular, break down to give different cell lineages with different numbers of copies of the repeat. This can manifest as multiple bands or smears on a Southern blot (e.g. Fig. 4, lane 6). As CpG methylation becomes established later in embryonic development, the repeat region is stabilized and further changes in copy number are rare.

The CCG repeats are adjacent to CpG islands and have been shown to be in the 5′-untranslated regions of the genes *FMR1* (at *FRAXA*), *CBL2* (at *FRA11B*), *FMR2* (at *FRAXE*) and *FAM11A* (at *FRAXF*). For *FRAXA*, a full mutation results in cessation of transcription of *FMR1*, and this is the molecular basis of fragile X syndrome; similar is the case for *FMR2* at *FRAXE* and mild mental retardation. *FRA11B* is in the 5′-untranslated region of the *CBL2*

oncogene; genes possibly associated with other cloned fragile sites have not yet been identified.

4.2 Groups 2 and 3

Two rare fragile sites, *FRA16B* from group 2 and *FRA10B* from group 3 have been characterized at the molecular level. Both are expansions of A/T (adenine/thymidine)-rich, minisatellite, DNA-repeat sequences. *FRA16B* is a massive expansion of a 33-bp sequence. Interestingly, this sequence has been reported to be 35-bp in length in a Japanese carrier of this fragile site. This fragile site does not appear to affect any gene and homozygotes have been described who are normal. FRA10B is due to an expansion of a 42-bp AT-rich minisatellite repeat that does not appear to affect any gene; again, homozygotes are normal.

4.3 Group 4

This is the largest group of fragile sites, and the only common ones that have been characterized at the molecular level. They are regions of fragility on the chromosome that extend over hundreds of kilobases of DNA, have no specific sequence features other than being AT-rich with many repetitive elements, and yet have genes within them (Table 4).

5 Cytogenetics

Fragile sites are seen by examining metaphase chromosomes under the light microscope. They appear as gaps, breaks, or unusual structures (Figs. 2 and 3) in a variable proportion of metaphases. The unusual chromosomal structures have arisen from breakage at the fragile site in a previous mitosis with the chromosomal material from either side of the fragile site ending up in different daughter cells.

Fragile sites are easier to see on plain-stained chromosomes than on banded chromosomes, especially when they occur in pale G-bands, as most do. It is often possible to see a fine strand of chromatin across the gap(s) at the fragile site.

The proportion of metaphases in which any particular fragile site is seen is highly variable. The common fragile sites, especially the aphidicolin-inducible ones are not often seen in more than 10% of metaphases but occasional individuals are

Tab. 4 Characterized group 4 common fragile sites.

Fragile site	*Approximate size*	*Gene content*
FRA2G	1 Mb	7 genes
FRA3B	900 kb	FHIT
FRA6F	1.2 Mb	5 genes
FRA7G	600 kb	CAV1, CAV2, TES, several more
FRA7H	160 kb	? none, contains SV40 integration site
FRA9E	9 Mb	PAPPA, at least 16 others
FRA16D	270 kb	WWOX1
FRAXB	500 kb	STS, GS1, TRL5A

Fig. 2 A selection of rare fragile sites. The chromosome on the left of each pair is plain-stained and the one on the right is G-banded.

encountered in whom one of these may be present in up to 30% of metaphases. The common fragile sites can be seen on both members of a chromosome pair in a single cell (homozygous expression). The rare autosomal fragile sites have never been seen in homozygotes, except for *FRA10B*, *FRA16B*, and *FRA17A* where this is without apparent phenotypic effect.

The rare fragile sites are not often seen in more than 50% of metaphases and are more typically present in 20 to 30% although some obligate carriers of a fragile site (from family studies) will not express it at all (presumed premutation carriers, see Sect. 4). An exception to these levels of cytogenetic expression is *FRA16B*; when induced with berenil, it can often be seen in more than 90% of metaphases.

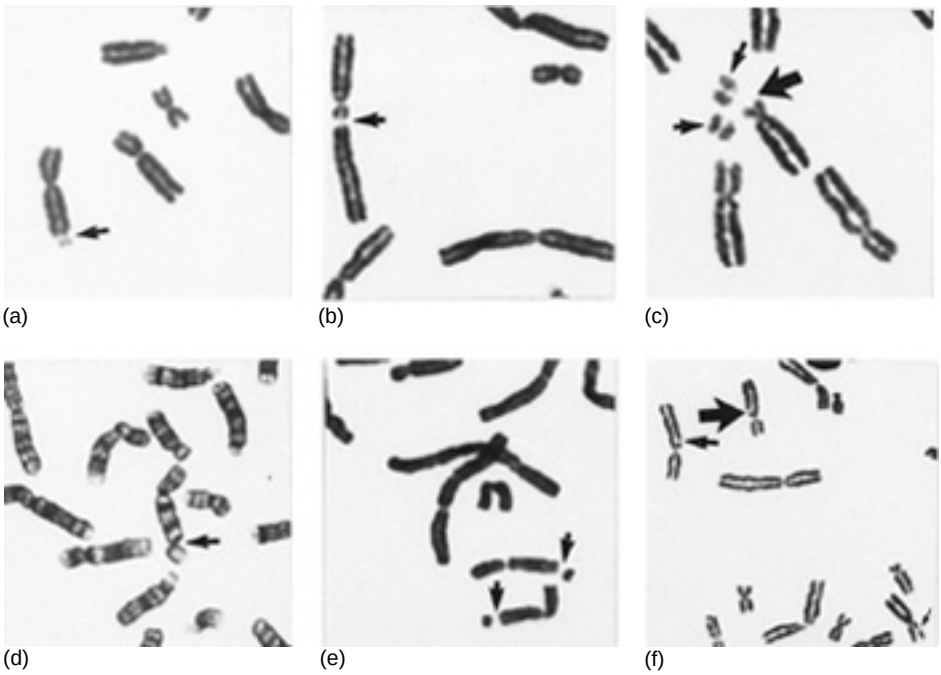

Fig. 3 Partial metaphases showing the appearance of various fragile sites (arrowed). (a) The fragile X (*FRAXA*); (b) expression of *FRA2B* as a chromosome gap; (c) *FRA9A* showing duplication of chromosome material (small arrows) distal to the fragile site (large arrow); (d) *FRA10B*; (e) homozygous expression of the common fragile site *FRA16D*; (f) BrdU-inducible fragile sites at *FRA6D* (medium arrow), *FRA9C* (small arrow), and *FRA10C* (large arrow).

6 Induction of Fragile Sites

Fragile sites have primarily been studied in chromosomes prepared from phytohemagglutinin stimulated human lymphocyte cultures. While some fragile sites can be induced in chromosomes from fibroblast or lymphoblastoid cultures and somatic cell hybrids, such induction is more difficult and the fragile sites are usually seen in lower proportions of metaphases than in lymphocyte cultures. Since the conditions required for induction of fragile sites differ for each of the groups shown in Table 1, each will be considered separately.

6.1 Group 1

These fragile sites require a relative deficiency of dCTP or dTTP at the time of DNA synthesis. A deficiency of dTTP can be achieved by using tissue culture medium free of folic acid and thymidine, by adding an inhibitor of folate metabolism such as methotrexate, aminopterin or trimethoprim, or specific inhibitors of thymidylate synthase such as fluorodeoxyuridine. A high concentration of thymidine in the culture medium will lead to a deficiency of dCTP resulting in fragile site expression. Induction conditions need to be in place for at

least 24 h before chromosome preparations are made.

6.2 Group 2

All these fragile sites are induced by distamycin A and a number of related compounds that bind to the minor groove of the DNA molecule in AT-rich regions. The most effective of these compounds is berenil. The two fragile sites in subgroup (a) are distinguished from those in subgroup (b) by also being inducible with BrdU, although this is not as effective an inducing agent as berenil.

6.3 Group 3

The two fragile sites are induced only by BrdU and the related compound BrdC.

6.4 Group 4

This large group of common fragile sites is seen in a small proportion of metaphases under the conditions that induce the Group 1 fragile sites. They can also be specifically induced by aphidicolin. Aphidicolin is added 24 h before chromosome harvest, and the addition of caffeine for the last 6 h of this period boosts the proportion of metaphases in which the fragile sites are seen.

6.5 Group 5

These fragile sites are induced by 5-azacytidine, which is added 8 h before chromosome harvest. If the addition is much before or after this timing, the fragile sites are not seen.

6.6 Group 6

BrdU induces these fragile sites and again the timing of addition of the compound is important. Maximum induction occurs when the addition is 4 to 6 h before chromosome preparation. If the BrdU is present for 24 h or more, none of these fragile sites will be seen.

7 Inheritance

The patterns of inheritance of some fragile sites are very unusual. The common fragile sites have not been studied and since they are on all chromosomes, no abnormalities in their inheritance would be expected. Of the rare fragile sites, the fragile X (*FRAXA*) associated with fragile X syndrome has been most studied. Before the molecular characterization of this fragile site, the inheritance patterns were particularly puzzling. Normal Mendelian segregation ratios were not seen, the penetrance of the fragile site and its associated syndrome depended upon position in the family (this anomaly was known as the *Sherman paradox*) (Fig. 1) and the intellectual status of the mother. There were normal males who had transmitted the fragile X, and fragile X syndrome was only apparent when the abnormal chromosome was inherited from the mother.

These anomalies in fragile XA inheritance were explained by the molecular characterization of the fragile site. The chance of a premutation expanding to a full mutation in a child is a function of the size of the mother's premutation. Full mutations only rarely regress to premutations when transmitted by women, but inevitably do so when transmitted by men.

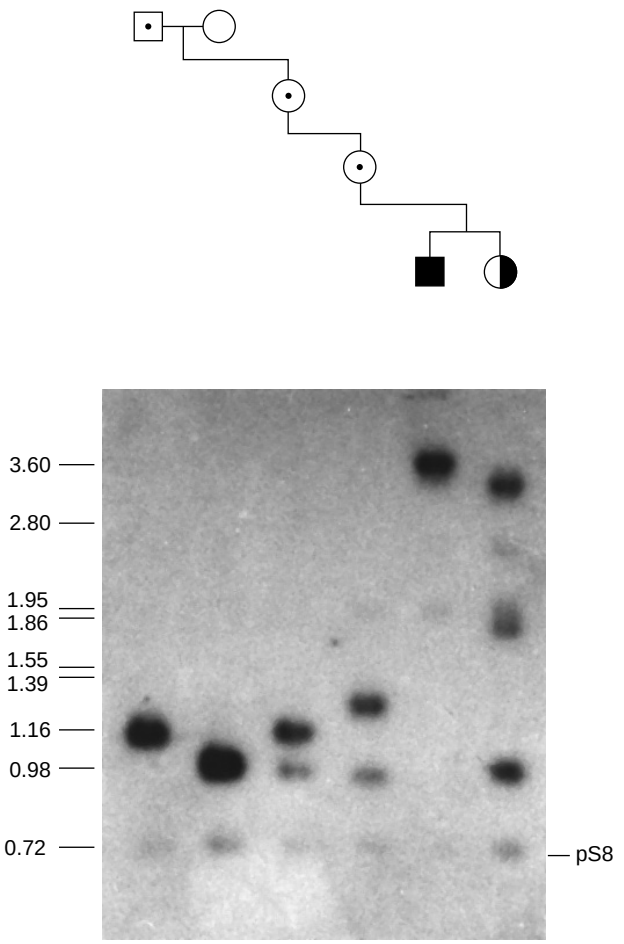

Fig. 4 Inheritance of the fragile X unstable element in a four-generation lineage from a large affected pedigree. Chromosomal DNA was digested with PstI and probed with pfxa3. Pedigree symbols: normal carrier male (dot in square); normal carrier female not expressing the fragile X (dot in circle); normal carrier female expressing the fragile X (half-shaded circle); affected fragile X syndrome male expressing the fragile X (shaded square); normal female (open circle). The carrier male in the first lane has no normal 1-kb band but a −1.15 kb band, his wife in the second lane has a 1-kb band. The daughter in lane 3 has a −1.15-kb band on her fragile X from her father and a 1-kb band from her mother. When this daughter has transmitted this band to her daughter, the fragile X band has increased to about 1.3-kb in size. This band has increased dramatically in size to about 3.5 kb in the fragile X syndrome boy, and appears as a somatically unstable smear in the carrier girl who expresses the fragile X (last lane). She has major bands of about 3.2-kb and 1.9-kb and a 1.0-kb band from her normal X chromosome.

Apart from the fragile sites on the X chromosome, there have been limited studies on the other Group 1 fragile sites. These are fully penetrant when transmitted by women but only 50% penetrant when transmitted by men. The reasons for this are unclear at this time. While the two autosomal Group 1 sites that have been cloned so far (*FRA16A* and *FRA11B*) have similar molecular mechanisms to the fragile X, very few family studies at the molecular level have been performed. While in at least the case of *FRA16A* and *FRAXE*, a male can transmit a full mutation, nothing is known about the likelihood of this happening.

The more frequently seen of the group 2 and 3 fragile sites (*FRA16B* and *FRA10B*) appear to have normal inheritance patterns.

8 Clinical Significance

8.1 The Fragile X Syndromes

One fragile site on the X chromosome (*FRAXA*) is of major medical significance in that it is a visible manifestation of the full mutation at this locus, which inactivates the *FMR1* gene. This gives rise to the *fragile X syndrome*, the most common familial form of intellectual handicap and one of the world's most common genetic diseases. This disease has been found in all population groups exposed to Western medicine but it is unknown whether it has the same frequency in all populations. Approximately 1 in every 4,000 children has fragile X syndrome.

Another fragile site on the X chromosome (*FRAXE*) is associated with a mild form of mental retardation. These males have no physical anomalies and the associated gene, *FMR2*, was the first to be isolated for nonspecific X-linked mental retardation. More than 50 families have been recorded in which mental retardation associated with this fragile site is present.

8.2 Other Fragile Sites

Some mothers of children with a mental retardation/malformation condition known as *Jacobsen syndrome* have *FRA11B*. Jacobsen syndrome is due to a deletion of part of the long arm of chromosome 11 and the deletion breakpoints in some cases are at or very close to the fragile site on their mother's chromosome. This is the only autosomal fragile site with possible clinical significance. Even here, very few women with *FRA11B* will have a child with Jacobsen syndrome and most Jacobsen syndrome patients have their chromosome 11 breakpoint distal to the fragile site.

The only rare autosomal fragile sites to have been recorded in homozygous form are *FRA10B*, *FRA16B*, and *FRA17A* where they are without apparent phenotypic effect. Group 1 autosomal fragile sites have never been seen in homozygotes. Any clinical effect in homozygotes would presumably depend upon the function of the gene in which the fragile site was located. It is likely that the presence of a fragile site detected cytogenetically means that a full mutation is present and any gene in which the fragile site is located is inactivated.

8.3 Cancer

There has been controversy for many years about a possible role for fragile sites in oncogenesis. The hypothesis is

that breakage at the fragile sites could give rise to chromosome rearrangements of the types seen in tumor tissue. Most studies had shown that when breaks in such rearrangements were carefully mapped in relation to fragile sites, the two did not coincide. However, *FRA11B* is within the oncogene *CLB2*, and a number of the common (group 4) fragile sites have within them putative tumor-suppressor genes. Nevertheless, there is no evidence that individuals with rare fragile sites (and everyone has all the common ones) are at any discernibly increased risk of developing malignant disease.

Acknowledgments

Support from the National Health and Medical Research Council of Australia is acknowledged.

See also Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Arlt, M.F., Casper, A.M., Glover, T.W. Common fragile sites. *Cytogenet. Genome Res.* (2003) **100**, 92–100.

Gecz, J. (2000) The FMR2 gene, FRAXE and non-specific X-linked mental retardation. *Ann. Hum. Genet.* **64**, 95–106.

Hagerman, R.J., Cronister, A. (1996) (Eds) *Fragile X syndrome*. 2nd Edition, Johns Hopkins University Press, Baltimore.

Mandel, J.L., Heitz, D. (1992) Molecular genetics of the fragile X syndrome: a novel type of unstable mutation. *Curr. Opin. Genet. Dev.* **2**, 422–430.

Richards, R.I., Sutherland, G.R. (1992) Dynamic mutations: A new class of mutations causing human disease. *Cell* **70**, 709–712.

Sutherland, G.R. (1993) Human Fragile Sites, in: *Genetic Maps, VI*. (S.J. O'Brien, Ed.). Cold Spring Harbor Laboratory Press, pp. 5.264–5.267.

Sutherland, G.R. (1991) The Detection of Fragile Sites on Human Chromosomes. in: Adolph, K.W. (Ed.) *Advanced Techniques in Chromosome Research*, Marcel Dekker, New York, pp. 203–222.

Sutherland, G.R. (2003) Rare fragile sites. *Cytogenet. Genome Res.* **100**, 77–84.

Sutherland, G.R., Baker, E. (2000) The clinical significance of fragile sites on human chromosomes. *Clin. Genet.* **58**, 157–161.

Sutherland, G.R., Hecht, F. (1985) *Fragile Sites on Human Chromosomes*. Oxford University Press, New York.

Primary Literature

Leehey, M.A., Munhoz, R.P., Lang, A.E., Brunberg, J.A., Grigsby, J., Greco, C., Jacquemont, S., Tassone, F., Lozano, A.M., Hagerman, P.J., Hagerman, R.J. (2003) The fragile X premutation presenting as essential tremor. *Arch Neurol.* **60**, 117–21.

Sherman, S.C., Jacobs, P.A., Morton, N.E., Froster-Iskenius, U., Howard-Peebles, P.N., Nielsen, K.B., Partington, M.W., Sutherland, G.R., Turner, G., Watson, M. (1985) Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum. Genet.* **69**, 289–299.

Fragile X-linked Mental Retardation

Ilse Gantois[1], *R. Frank Kooy*[1], *and Ben A. Oostra*[2]
[1] *University of Antwerp, Antwerp, Belgium*
[2] *Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands*

1 Introduction 588

2 Description of Fragile X Syndrome 588
2.1 Some Historical Aspects 588
2.2 Fragile Site 589
2.3 Unusual Pattern of Inheritance 589
2.4 Phenotype 590
2.5 Neuroanatomy 591

3 Identification of the Causative Gene 591
3.1 The Cloning of the Gene 591
3.2 Dynamic Mutation 592

4 *FMR1* Gene and Protein 593
4.1 Gene Structure and Expression Pattern 593
4.2 Structural Domains of FMRP 594
4.3 FMRP Associates with Ribosomes as Part of an mRNP 596
4.4 FMRP as a Possible Inhibitor of Translation 597
4.5 mRNA Targets of FMRP 597
4.6 Function of FMRP in Neurons 598

5 Paralog Genes FXR1 and FXR2 600

6 Amino-terminal Interacting Proteins 602

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

7 Animal Models 603
7.1 Comparison between Species 603
7.2 What Have We Learned from the Mouse Model? 604
7.2.1 Macroorchidism 604
7.2.2 Cognitive Testing 604
7.2.3 Motor Activity and Anxiety 604
7.2.4 Acoustic Startle Reflex 605
7.2.5 Audiogenic Seizures 606
7.2.6 Neuroanatomy and Physiology of the Knockout Brain 606
7.2.7 FMRP is Involved in Long-term Plasticity 607
7.2.8 FXR1/FXR2 Knockout Mice 607
7.3 What Have We Learned from Drosophila? 607
7.3.1 The Ortholog, dFmr1 607
7.3.2 dFmr1 is Involved in Neuronal Maturation 609
7.3.3 MAP1B may be Involved in the Fragile X Syndrome 610
7.3.4 dFmr1 is Involved in Circadian Rhythm Regulation 610
7.3.5 dFmr1 may be Involved in RNA Interference 611
7.3.6 dFmr1 is Modified After Translation 611

8 Phenotypic Effect of the Premutation 612

9 The FMR1 Mutation 613
9.1 Mechanism of CGG-repeat Expansion 613
9.2 Timing of CGG-repeat Expansion 615

10 Diagnostics 616
10.1 DNA Diagnostics 617
10.1.1 Southern Blotting 617
10.1.2 Example of DNA Diagnostics in a Fragile X Family 617
10.1.3 PCR Amplification Analysis 617
10.2 Protein Test 618

11 Therapeutic Options 619
11.1 CGG Methylation and Reactivation 619
11.2 AMPAkine Trial in Relationship with LTD Abnormality 620

12 Epilogue 620

Bibliography 621
Books and Reviews 621
Primary Literature 621

Keywords

Allele
Alternative form of a gene found at the same locus on homologous chromosomes.

CGG-repeat Expansion
Repeated-sequence CGG at the DNA level. When the number of repeats increases, this is referred to as an expansion.

FMR1
Gene that is shut down and causes fragile X mental retardation.

FMRP
The protein product of the *FMR1* gene.

Fragile Site
Can be induced in chromosomes by culturing cells in a folate-deficient medium.

Full Mutation
Is defined as a mutation in the *FMR1* gene with phenotypic symptoms.

Genotype
Genetic constitution of an individual.

Isoforms
Alternative splicing products of a gene.

Normal Transmitting Male
A phenotypically normal male who is a carrier of a premutation in the *FMR1* gene and may show cytogenetic expression of the fragile site.

Phenotypic
Observable characteristics.

Premutation
Is defined as a mutation in the *FMR1* gene without phenotypic symptoms.

Ribosome
A subcellular structure involved in protein production.

Fragile X syndrome is the most common form of inherited mental retardation. The disease is caused by methylation-induced transcriptional silencing of the fragile X mental retardation 1 gene (*FMR1*) as a result of the expansion of a CGG repeat in the 5′ UTR of the gene. This leads to the loss of fragile X mental retardation protein product (FMRP). FMRP is an RNA-binding protein that associates with translating polyribosomes as part of a large messenger ribonucleoprotein (mRNP) and modulates the translation of its RNA ligands. The *in vivo* targets of the FMRP are beginning to be elucidated and have been shown to include a subset of RNAs with a planar structure called *G-quartet*.

To study the physiological role of the FMR1 protein, animal models have been developed, including genetically modified mouse and *Drosophila*. The loss-of-function mouse model shows some features comparable to human patients including enlarged testes, a subtle behavioral phenotype, and discrete anomalies of dendritic spines. Studies in *Drosophila* indicate that dFmr1 in flies plays an important role in synaptogenesis and axonal arborization, which may underlie the observed deficits in flight ability and circadian rhythm in the dFmr1-mutant flies.

1 Introduction

Mental retardation can be defined as a failure to develop cognitive abilities and achieve a level of intelligence that would be appropriate for the age group. A percentage of the general population, variably estimated between 0.5 and 3%, is reported to be functioning two standard deviations below the average (i.e. to have an IQ of less than 70). Genetic determinants underlie approximately half of these conditions and an excess of affected males has often been reported, especially in the mild-to-moderate mental retardation range with an IQ between 70 and 35. Excess of males with mental retardation compared to females was eventually interpreted as being due to X-linked factors. One of the first described X-linked heritable mental retardation syndromes is the fragile X or Martin–Bell syndrome. After trisomy 21 or Down syndrome, it is the most frequent cause of mental retardation.

2 Description of Fragile X Syndrome

2.1 Some Historical Aspects

In 1943, Martin and Bell reported a large family in which 11 mentally retarded boys were born to mothers of normal intelligence and hypothesized that the mental defect of the boys was due to a sex-linked recessive gene, which was suppressed in the two grandfathers. More than a quarter of a century later, Lubs reported a unique cytogenetic marker at the end of the long arm of the X chromosome (Xq27.3) in cultured lymphocyte cells from four retarded male patients in a family with X-linked mental retardation. As more pedigrees were identified in which the marker on the X chromosome segregated with mental retardation, it became clear that the fragile site could be used as a diagnostic tool. In 1978, Turner observed macroorchidism (enlarged testicles) in a

large proportion of the patients with the fragile site. Reinvestigation of the pedigree described by Martin and Bell revealed a fragile site and the typical features of the syndrome in the affected patients. In recognition of the first published pedigree, the mental retardation syndrome associated with the fragile site on the X chromosome and macroorchidism was named *Martin–Bell syndrome*, though these days it is more usually called *fragile X syndrome*.

2.2 Fragile Site

Before the cloning of the responsible gene in 1991, the laboratory confirmation of the diagnosis of fragile X syndrome was dependent on cytogenetic detection of the fragile site at Xq27.3 in cultured lymphocytes, fibroblasts, or amniocytes (Fig. 1). This fragile site was called FRAXA (fragile site, X chromosome, A site), as it was the first fragile site identified on the X chromosome. On metaphase spreads, fragile sites are microscopically visible as a gap or a break. The fragile site is induced by culturing cells under folic acid deprivation before chromosome spreads are made. The percentage of cells that show fragile site expression may vary from individual to individual, a phenomenon for which there is no explanation. In fragile X males, the fragile site is usually seen in 2 to 60% of the cells, but the site can only be detected in 50% of obligate carrier females, limiting the utility of cytogenetic carrier detection in pedigrees.

2.3 Unusual Pattern of Inheritance

From the initial observations onward, it was clear that fragile X syndrome has an unusual pattern of inheritance (Fig. 2). In X-linked recessive diseases, female carriers are usually unaffected and transmit their carrier status to half of their daughters, while half of their sons may develop the disease. However, in the case of the fragile X syndrome, 30% of carrier females are mentally retarded, although often less severely than males. In addition, 20% of males with cytogenetic expression of the fragile site are not affected. Such males, called *normal transmitting males* (NTMs), pass the mutation on to their unaffected daughters. However, these carrier daughters have a 40% risk of transmitting the disorder to their sons. In comparison, sons of affected females have a recurrence risk of 50%. Only transmission through the female line can give rise to a full phenotype in the next generation. Other peculiarities include the

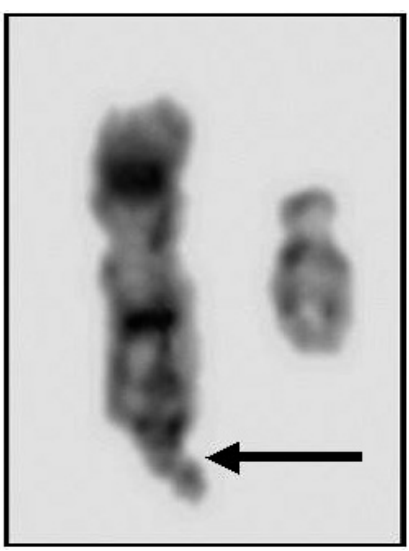

Fig. 1 Cytogenetic expression of the fragile site at Xq27.3 in a patient is clearly visible and indicated with an arrow.

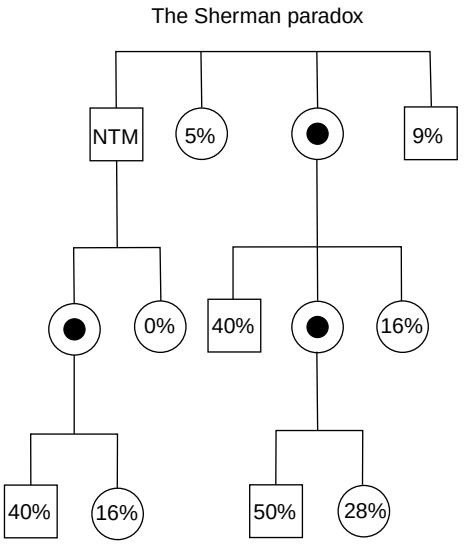

Fig. 2 A schematic representation of a fragile X family illustrates the Sherman paradox. The position within the pedigree affects the chance (indicated as a percentage) of being affected with the fragile X syndrome. NTM, normal transmitting male.

observation that more patients occur in the youngest generation of the pedigree. This increase in the penetrance of the disease with generations is called the *Sherman paradox*. The abnormal inheritance pattern could only be explained after the cloning of the gene and elucidation of the peculiar mutational mechanism.

2.4 Phenotype

Fragile X syndrome has an incidence of about 1 in 4000 males and 1 in 6000 females. Males suffer from mild to severe mental retardation with an IQ between 20 and 60, while females are usually less severely affected. Macroorchidism develops during or after puberty and is seen in more than 80% of all adult patients. Fragile X patients show physical characteristics such as a typical long face and prominent ears (Fig. 3), but these symptoms can be absent in younger children. Hyperextensible joints, hand calluses, and flat feet can also be observed and may point to connective tissue dysplasia. Associated behavioral problems may include hyperactivity, decreased attention span, repetitive behavior, poor eye contact, tactile defensiveness, and shyness. Speech and language skills are usually impaired in older boys. In 13 to 23% of the male patients, autistiform features have been observed, occurring mostly before adulthood. Notwithstanding this observation in a proportion of fragile X patients, there seems to be no general

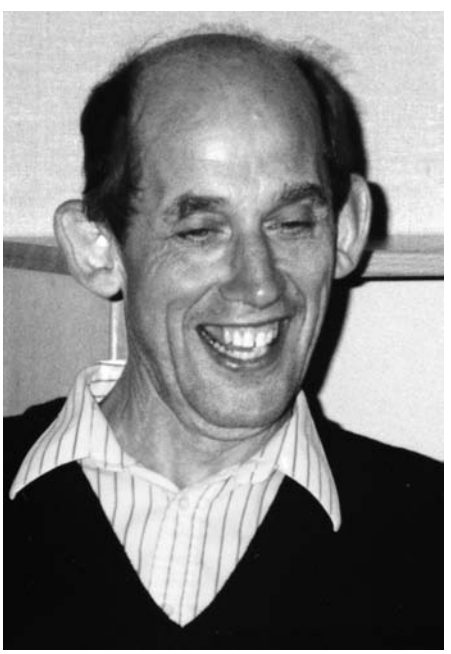

Fig. 3 Male fragile X patient displaying the typical facial characteristics of the syndrome with a long face and prominent ears.

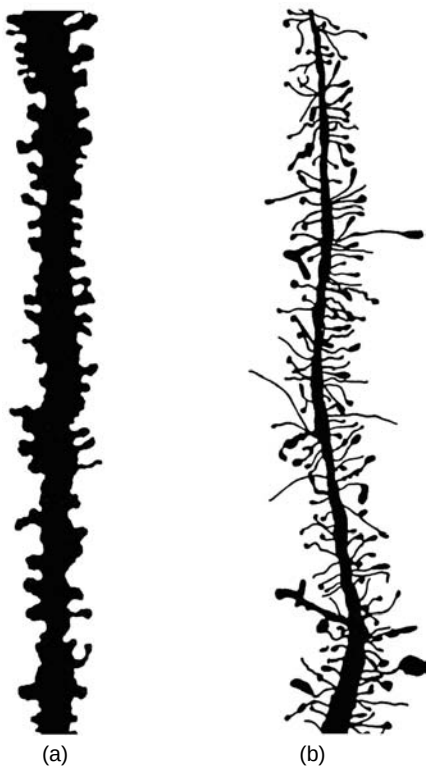

Fig. 4 Example of a typical spine morphology on dendrites from (a) a fragile X syndrome patient and (b) an unaffected control. Dendrites of the fragile X patient display more immature spines and a higher spine density. (Reprinted with permission from Fiala et al. (2002).)

association between fragile X syndrome and autism. In addition, fragile X boys suffer from a greater variability in total sleep time and sleep-rhythm maintenance when compared to controls. Fragile X patients have a normal life span.

2.5 Neuroanatomy

Brain abnormalities in fragile X patients have been determined using magnetic resonance imaging (NMR). A reduction in the size of the posterior cerebellar vermis and enlargement of the fourth ventricle compared to controls was reported. An increased volume of the hippocampus and the lateral ventricles in fragile X males was observed together with a significantly larger volume of the caudate nucleus. This is an important observation since the caudate nucleus, together with the lenticular and thalamic nucleus, forms the subcortical gray that allows the cerebral cortex to have an influence on behavior. In 3D reconstructions, fragile X brains appear to be larger than control brains. Volumetric changes were not confirmed by physical measurement of the brains of two deceased fragile X patients in a separate study.

An impaired neuronal development has been observed in fragile X patients. The spines on apical dendrites of pyramidal cells in the Golgi-impregnated visual and temporal cortex of fragile X patients are long, tortuous, immature, and thin, and fragile X patients exhibit an increased spine density on distal segments of apical dendrite shafts (Fig. 4). The higher spine density contrasts with other known mental retardation syndromes such as Down syndrome, in which a lower spine density is observed.

3 Identification of the Causative Gene

3.1 The Cloning of the Gene

Initially, the genetic defect underlying the fragile X syndrome was unknown and diagnosis was only based upon the observation of the fragile site at Xq27.3 in cultured cells. Positional cloning enabled

the isolation of the gene and helped understand the molecular mechanism leading to the disorder. Linkage analysis with polymorphic markers placed the fragile X locus between factor IX at Xq27.1 and factor VIII at Xq28, conforming the locus for the gene at or nearby the fragile site. Hybrid cell lines with translocation break points at the fragile site were very useful in developing and mapping new probes near the gene. Two cell lines, termed Micro21D and QIX, each containing a reciprocal part of the human X chromosome, were especially helpful. Micro21D contained the Xpter–Xq27.3 part of the human X chromosome and Q1X contains the Xq27.3–Xqter part of the human X chromosome.

Using markers linked to the fragile site, several YACs ranging from 200 to 950 kb could be isolated. Among these, YAC 209G4 contained 475 kb human DNA and was shown to cross the fragile site by *in situ* hybridization. A cDNA clone was isolated from a human fetal brain library using cosmids from this particular YAC and this gene was assigned the name *FMR1* (Fragile X Mental Retardation 1). Further analysis of this gene sequence showed that the translation breakpoints of Micro21D and Q1X lie in exon 1 of the gene. Surprisingly, exon 1 contained a trinucleotide repeat that shows length variation between control and fragile X patients. Expansion of the CGG repeat was shown to be involved in both the inactivation of the *FMR1* gene and the cytogenetic expression of the fragile site.

3.2 Dynamic Mutation

Studying the dynamic inheritance pattern of the CGG repeat resolved the complex inheritance of the fragile X syndrome. In more than 99% of all fragile X patients, the syndrome is caused by a dynamic expansion of a CGG repeat in the 5′ untranslated region (5′UTR) of the *FMR1* gene (Fig. 5). In the normal population, the CGG repeat is polymorphic with allele sizes between 5 and 50 repeat units, 30 units being the most frequent allele size. Alleles of this size range are inherited stably from generation to generation and the gene is transcribed and translated into protein. In patients, a trinucleotide CGG expansion of more than 200 to several thousands of CGG repeats, referred to as a *full mutation*, is observed. Because of the repeat expansion, the CGG repeat and the surrounding promoter region of the *FMR1* gene is methylated, inhibiting mRNA translation and causing absence of the protein product. Absence of the FMR1 protein product, FMRP, causes the characteristic features of the fragile X syndrome. The alleles with a repeat size between 50 and 200 CGG units are called *premutation alleles*. Premutations are unmethylated and produce FMR1 mRNA and FMRP. Premutation carriers are not affected with fragile X syndrome. However, repeats in the premutation size range may be instable upon transmission to the next generation. Expansion to a full mutation occurs only through maternal transmission, explaining why only progeny of female premutation carriers is at risk for the fragile X syndrome. The observation that the larger the premutation, the more likely it will expand to a full mutation in the next generation, resolved the Sherman paradox.

The number of CGG repeats itself does not always allow discrimination between the three categories of repeat sizes. Repeats of 200 to 230 CGG units can be either large premutations or small full mutations. However, discrimination between the two types of mutations is possible on

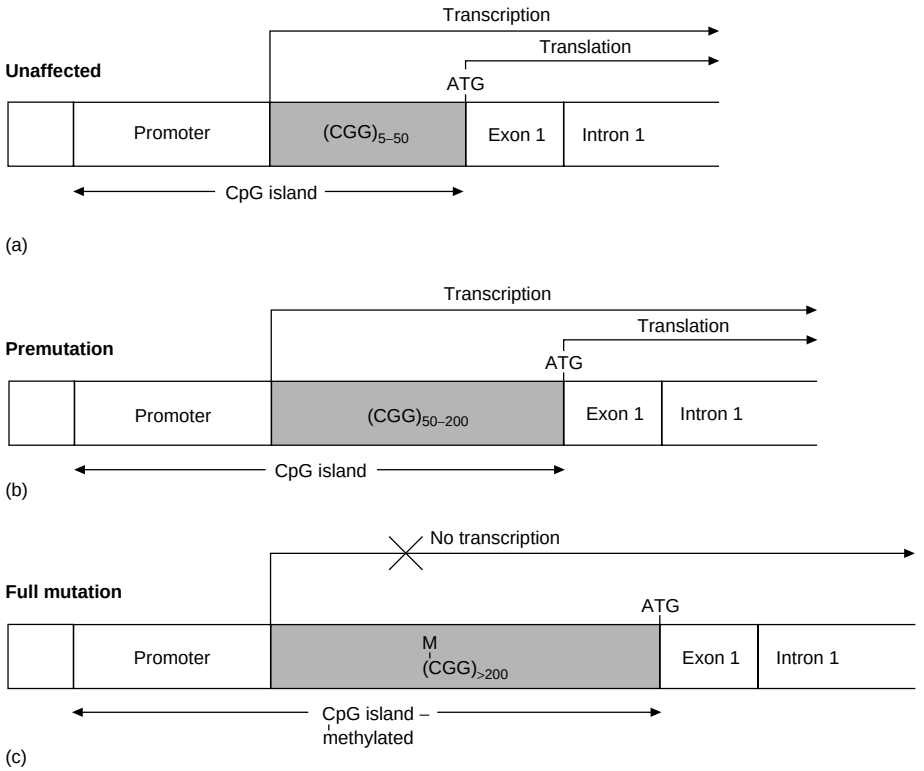

Fig. 5 Schematic representation of the transcription of (a) the *FMR1* gene in normal controls, (b) premutation carriers, and (c) full-mutation patients. The normal and the premutation size alleles (a and b) are transcribed and translated into FMRP. The full-mutation allele (c) is hypermethylated and therefore not transcribed.

the basis of the methylation status. Alleles comprising 45 to 60 CGG units are referred to as being of intermediate size. These may inherit stably, but may also increase or decrease by a few CGG units in length during transmission and become premutation alleles at risk for expansion to a full mutation in future generations. Factors influencing repeat stability include, but are limited to, the amount of AGG interspersions in the CGG repeat. In the *FMR1* gene, on average every 10th CGG repeat unit is interspersed with an AGG triplet. AGG repeats stabilize the repeat; fewer the AGG interspersions, the more prone it would be to intergenerational expansion. The instability threshold for FMR1 alleles is estimated to be about 34 to 38 uninterrupted CGG repeats, and the expansion of an interrupted CGG repeat of 59 units to a full mutation in the next generation has been observed.

4 *FMR1* Gene and Protein

4.1 Gene Structure and Expression Pattern

The *FMR1* gene has a genomic size of approximately 38 kb and consists of

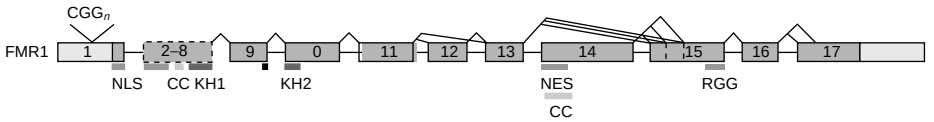

Fig. 6 The *FMR1* gene containing 17 exons and a CGG repeat in the 5′ UTR. The splice sites as well as the positions of the functional domains are indicated. NLS, nuclear localization signal; NES, nuclear export signal; CC, coiled coils; KH/RGG, KH, and RGG domains are important for RNA binding.

17 exons (Fig. 6). The transcript length is 4.4 kb, and all elements necessary for proper FMR1 expression *in vivo* are present in a 2.8 kb fragment immediately upstream of the ATG. The promoter sequence is GC-rich, as found in many housekeeping genes. Footprinting analyses of the promoter site revealed consensus-binding sites for ubiquitously expressed transcription factors such as Sp1, H4TF1/Sp1-like, and c-myc. In addition, an AP-2 binding site, specific for neuronal expression, was identified. Upstream of the CGG triplet repeat, a 20-nucleotide-long internal ribosome entry site (IRES) was identified. The function of this element in FMR1 is unclear, but IRES elements of other mRNAs play a role in neuronal regulation of translation in response to synaptic stimuli.

By Western blotting using antibodies directed against FMRP, different isoforms were identified with a molecular mass around 70 to 80 kDa and a maximal protein length of 631 amino acids. Excessive alternative splicing occurs especially in the 3′ terminal half of the gene, in exons 12, 14, 15, and 17 (Fig. 6). This can potentially give rise to 12 different mRNAs and protein isoforms. Skipping of exon 14 introduces a frameshift in the open reading frame, resulting in a different amino acid order in the carboxyl terminal half of the protein. The function of the different isoforms is still unknown and the different splice products are not tissue-specific. FMRP shows near-ubiquitous expression. It is highly expressed in ovary, thymus, eye, spleen, and esophageal epithelium with an abundant expression in brain and testis and a moderate expression in colon, uterus, thyroid, and liver, and little or no expression in the heart, aorta. or muscle. In the brain, the FMRP expression is restricted to differentiated neurons and Purkinje cells and is absent in nonneuronal cells. Neuronal FMRP is concentrated in the perikaryon and proximal dendrites and is also expressed in the synapses but is not detected in the axons. Especially high expression was found in Purkinje cells of the cerebellum and in neurons of the granular layer, the cortex, and the hippocampus. In the testis, the FMRP is expressed in the cytoplasm of primordial germ cells and in the early spermatogonia adjacent to the basal membrane of the tubuli seminiferi. Immunocytochemical studies revealed that FMRP is predominantly localized in the cytoplasm with high levels found near the rough endoplasmic reticulum, but some nuclear expression was also observed by both light and electron microscopy.

4.2 Structural Domains of FMRP

Two types of RNA binding domains have been identified in the protein, including two KH domains that show homology

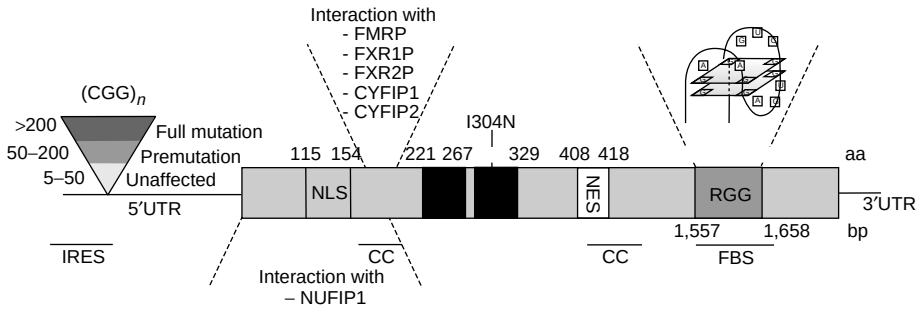

Fig. 7 Schematic representation of the FMR1 mRNA and protein. The known domains are indicated: IRES, internal ribosome entry site; NLS, nuclear localization signal; CC, coiled coils; KH, K-protein homology domain; NES, nuclear export signal; RGG box, Arg-Gly-Gly triplet; FBS, FMRP binding site. In addition, the Ile304Asn missense mutation found in a single, most severely affected patient is indicated. In the 5′ UTR, the different repeat classes are represented. The G-quartet structure present in the mRNA is depicted above FBS that overlaps with the RGG box. The interacting regions of FMRP with its paralog genes, *FXR1* and *FXR2*, and with other identified proteins, CYFIP proteins and NUFIP1, are indicated.

to hnRNP K domains and an RGG box containing a conserved Arg-Gly-Gly triplet. In addition, a nuclear localization signal (NLS), a nuclear export signal (NES), two coiled coils, and a G-quartet-binding structure have been identified (Fig. 7).

The KH domain consists of the consensus sequence, I/L/V-I-G-X_2-G-X_2-I. It is usually present in more copies in a protein and is able to bind ssRNA, mRNA, and rRNA. Using NMR (nuclear magnetic resonance), the KH domain was shown to consist of a stable three-stranded antiparallel β-sheet, orientated against three α-helices. The $\beta\alpha\alpha\beta\beta\alpha$ folds to a surface for RNA binding and a conserved Gly-Lys-X-Gly is centered on the loop between the first two helices. The importance of the KH domain for proper FMRP function is illustrated by a rare, unique patient with an Ile304Asn mutation located in the second KH domain. The mutated isoleucine is highly conserved in the KH domain sequence. The patient has a phenotype much more severe than any other known fragile X patient, including profound mental retardation and excessive macroorchidism. This patient still generates FMRP. The mutant FMRP has impaired RNA-binding capacity under elevated salt conditions. The mutation disrupts the normal folding of the KH domain. The FMRP–RNA binding can be disrupted by deletion of the C-terminus of the protein, containing the RGG box, suggesting that the RGG box plays an important role in FMRP–RNA binding. The RGG coding region of the FMR1 protein shows a high affinity for its own mRNA and other G-quartet-containing mRNAs, further described in Chapter 3.5. All these findings let us assume that FMRP has multiple binding regions to associate with RNA. *In vitro* translated FMRP binds directly to RNA homopolymers. It shows highest affinity for poly(G), a lower affinity for poly(U), and no affinity for poly(A) or poly(C). FMRP is able to bind selectively 4% of all brain mRNAs including its own mRNA, confirming that FMRP is an RNA binding protein. The creation of deletion mutants of FMRP, missing the N-terminal, the C-terminal, or

one of the KH domains, revealed that only the first KH domain interacts with RNA. This result agrees with the findings that the Ile304Asn mutant is still able to bind some RNAs since the mutation is situated in the second KH domain, although the capacity to bind poly(U) homopolymers is destroyed. Both the N-terminal and the C-terminal show RNA-binding abilities and the N-terminal binds poly(G) with a similar preference as the first KH domain.

FMRP contains both an NES and an NLS signal. The NES is approximately 10 amino acids long and resembles the NES motif of HIV1 Rev. It consists of a small stretch of hydrophilic amino acids containing four critically spaced, large, hydrophobic amino acids. The NES is located in exon 14 and absence of this exon results in trapping of the FMRP in the nucleus. The nuclear export mediated by the NES of FMRP is exportin1-dependent, since Leptomycin B treatment, an antibiotic acting specifically on exportin1, causes the accumulation of FMRP in the nucleoplasm. The NLS contains clusters of basic amino acids, but the sequence does not resemble other known nuclear localization signals. It is located between residues 115 and 154. The nuclear localization of FMRP mediated by the NLS is an energy-dependent process. The presence of both an NES and an NLS may suggest that the fragile X protein shuttles through the cell into and out of the nucleus. In accordance with this hypothesis, the protein has been observed in the nuclear pore during transfer between the nucleus and the cytoplasm.

FMRP contains two coiled coils involved in protein–protein interaction and located in exon 7 and in exons 11 to 14 respectively. The first coiled coil is potentially involved in binding of FMRP interacting proteins, while the second was suggested to play a role in FMRP–ribosome binding.

4.3 FMRP Associates with Ribosomes as Part of an mRNP

In human cell lines, FMRP colocalizes primarily with polyribosomes and rough endoplasmic reticulum. High levels of FMRP were found in perikarya, where it is concentrated in regions rich in ribosomes, particularly near or between rough endoplasmic reticulum cisternae. FMRP has also been detected in the granular component of nucleoli in COS cells. The nucleolus is a cellular component formed in response to the requirement for new ribosome synthesis. The FMRP–ribosome association is sensitive to 500 mM salt concentration and to low concentrations of RNase digestion that does not disturb the ribosome assembly *per se*. This suggests that FMRP is bound to the ribosomes via RNA. These findings are in contradiction with earlier experiments that suggested that FMRP binds to the 60S ribosomal subunit via part of exon 13 and 14, containing the second coiled coil domain. Whatever the role of the FMRP–60S ribosome binding, the RNA–protein interaction seems the most important as nearly all cytoplasmic FMRP can be cocaptured with poly(A) mRNA. Moreover, after EDTA treatment, the FMRP complex is released from the ribosomes.

In the cell, FMRP forms a messenger ribonucleoparticle, mRNP, with a sedimentation value of 60S and a size >660 kDa. The composition of the mRNP complex is not known, but some proteins involved in the mRNP particle include FXR1 and FXR2, paralogs of FMR1 (Chapter 4), NUFIP and CYFIP proteins (Chapter 5), nucleolin, a known protein in other mRNPs,

and an ortholog of the mouse protein 1/p50. It is not known whether all mRNP particles containing FMRP have the same composition. If this is not the case, the absence of FMRP may influence the functioning of the different types of mRNPs in a different way.

4.4 FMRP as a Possible Inhibitor of Translation

The association of FMRP with the polyribosomes suggested a potential controlling role in mRNA translation and stability. *In vitro*, FMRP is able to suppress translation of mRNA. This inhibition is dosage-dependent and is observed for all mRNAs tested at low FMRP concentrations in an *in vitro* synthesis system and in microinjected *Xenopus laevis* oocytes. This may suggest that cells with absent FMRP translate certain mRNAs in excess. Nevertheless, a study by Brown et al. (2001) (Chapter 3.6) showed that increased and decreased amounts of specific polysome-bounded mRNAs occur in fragile X patients. It was suggested that elevated levels of FMRP are sequestering mRNAs into incompetent granules causing the repression of reporter genes. Recently, it was reported that only a minor fraction of newly synthesized FMRP is associated with polyribosomes, while the majority of it is found in cytoplasmic granules that are resistant to nonionic detergents. Increasing the amount of FMRP resulted in trapping of the mRNAs in these granules together with several other mRNA-binding proteins. This suggested a potential dual role for FMRP both as a transporter of specific mRNAs and also as a repressor of translation in specific conditions.

FMRP containing the Ile304Asn mutation fails to repress translation despite the fact that the RNA-binding capacity of mutated FMRP is only mildly affected. The mutated protein appears as a part of a smaller, abnormal mRNP that is not associated with translating polyribosomes, perhaps because FMRP is also able to form dimers but the Ile304Asn protein fails to do so. This suggests that the loss of association to ribosomes and to form normal mRNP particles causes the severe phenotype of this patient.

4.5 mRNA Targets of FMRP

As FMRP has RNA-binding capacities, identifying the mRNAs bound by FMRP in the cell is a prerequisite for understanding its function. An interaction of FMRP with the 3′ UTR of the myelin basic protein mRNA was found using purified recombinant FMRP. Using a biotinylated-FMRP affinity resign, Sung et al. (2000) identified nine mRNAs from adult brain that are able to bind FMRP, including a neuronal NT2 EST and Tip60a, a Tat interactive protein. Also the *Xenopus* elongation factor 1A, xEF-1A, binds strongly to human FMRP. FMRP was demonstrated to inhibit mRNA translation of this gene, and in the absence of FMRP, the translation of human EF-1A is derepressed.

Recently, the C-terminal region of FMRP was identified as major site of specific mRNA interaction. A specific, high-affinity binding site for its own mRNA in the RGG box region of FMRP was identified. A 100-nucleotide mRNA fragment, situated in the region encoding the RGG box, was shown to contain a purine-quartet motif predicted to form a planar structure. This so-called G-quartet structure was shown to be involved in the binding of FMR1-mRNA by FMRP (Fig. 7). At the same time, Darnell et al. (2001) identified FMRP-bound mRNAs, by screening a

pool of 96-mer RNAs containing 52 bases of random sequence. A common motif preferentially bound by FMRP was identified as a G-quartet.

In vivo studies also revealed G-quartet-containing mRNAs as targets of FMRP. mRNAs associated with FMRP-mRNPs that interact with active translating ribosomes in mouse brain were isolated and hybridized on microarrays containing a significant proportion of murine genes. By subtracting the expression patterns of knockout and control mice, RNAs specifically bound to FMRP were identified. In a similar way, mRNA populations in the polyribosomal fractions of lymphoblasts of fragile X patients were compared with controls. The studies revealed a long list of FMRP-associated mRNAs. A total of 426 mRNAs were identified in both the human and the mouse studies. Moreover, 14 of the mRNAs were identified in both the experiments of Darnell et al. (2001) and Brown et al. (2001). Eight of them revealed a G-quartet structure including, for example, NAP22, which is present in axon terminals and dendritic spines and plays a role in maturation or maintenance of synapses. In addition, in both studies MAP1B was identified, which is involved in transport within neurons and is the human ortholog of the *Drosophila* futsch gene (Chapter 6).

Using the Fmr1 knockout mouse model, Zalfa et al. (2003) revealed that FMRP regulates translation of specific dendritic mRNAs. The translational efficiency, the percent of a given mRNA that is associated with polysomes (PMP = Percent of Messenger on Polysomes), was quantified and shown to be different for specific RNAs, including those that encode for cytoskeletal proteins such as Arc and MAP1B and signal transduction molecules such as α-CaMKII. Interestingly, interaction occurred through a ribonucleoprotein complex that includes the small dendritic noncoding RNA BC1. It was shown that BC1 directly binds to FMRP and is able to base-pair with regions of complementarity between BC1 and the regulated mRNAs. BC1 may be required for FMRP association with the targeted mRNAs and for determining the binding specificity.

In vivo binding of mRNAs to FMRP in hippocampal neurons was determined using antibody-positioned RNA amplification, a novel technique that allows identification of RNAs that are directly associated with FMRP in the cell. Approximately 83 genes believed to interact directly with FMRP were identified. The FMRP interacting mRNAs were mostly involved in diverse aspects of synaptic plasticity and neuronal maturation. G-quartets are present in a small percentage of the mRNAs indicating that other factors in addition to the G-quartet are important in FMRP–RNA binding. Interestingly, some of the proteins encoded by the RNA showed a discrete change in abundance and subcellular distribution. The identified mRNAs include the glucocorticoid receptor α (GRα). It is a low-affinity receptor for corticosteroid hormones, involved in the electrical activity in the hippocampus in response to mineralocorticoid receptor stimulation. The receptor showed a change in dendritical distribution in the hippocampus of the knockout mouse. Diminished responsiveness of the receptor is compatible with learning problems observed in fragile X patients.

4.6 Function of FMRP in Neurons

The ability of FMRP to bind RNA, its association with ribosomes, and the observation that the protein shuttles through the cell lead to the hypothesis that

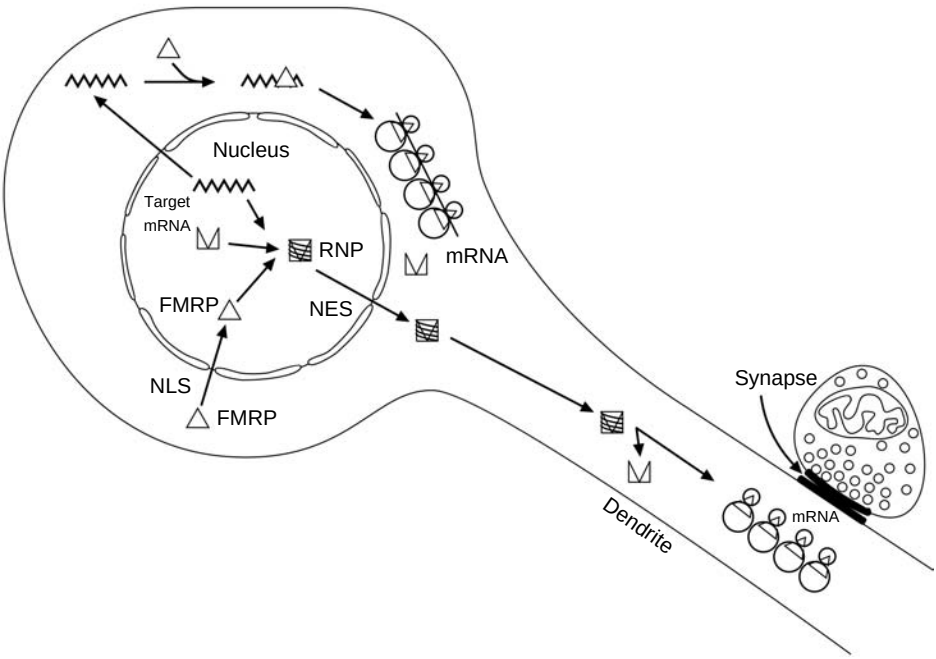

Fig. 8 A hypothetical model of the FMRP function in the neuron. FMRP is transported from the cytoplasm to the nucleus, where it assembles into an mRNP particle, containing specific mRNAs and other proteins. The mRNP particle is transported to the cytoplasm and mRNAs are presented to the polyribosomes in the perikaryon and the dendrites nearby the synapses.

FMRP may play a role in mRNA transport from the nucleus to specific destinations in the cell (Fig. 8). This hypothesis is supported by the observation that FMRP associates in the nucleus with mRNAs and proteins to form an mRNP particle that is transported through the cytoplasm and is presented to the ribosomes for translation. It is not clear what happens with FMRP after translation. FMRP may be released from the ribosomes as a free protein after translation termination or may be trapped in cytoplasmic granules.

Because of the putative role of FMRP in cellular mRNA transport, dendritic mRNA transport was investigated in an FMR1–EGFP stable transfected PC12 cell line with an inducible expression system (Tet-On) for regulated expression of the FMRP–GFP fusion protein. After Dox-induction, FMRP–GFP was localized in granules in the neurites of PC12 cells. Using time-lapse microscopy, the trafficking of FMRP–GFP granules into the neurites of living PC12 cells was demonstrated. Motile FMRP–GFP granules displayed two types of movements: oscillatory (bidirectional) and unidirectional anterograde. The average velocity of the granules was $0.19\ \mu m\ s^{-1}$ with a maximum speed of $0.71\ \mu m\ s^{-1}$. The movement of FMRP-labeled granules was microtubule-dependent. Colocalization studies showed that the FMRP–GFP granules also contained RNA, ribosomal subunits, kinesin heavy chain, and

FXR1P molecules as constituents. This observation stresses the putative role for FMRP–mRNP particles in dendritic mRNA transport to the synapse.

In the synapse, FMRP might regulate the translation of certain mRNAs. Interestingly, local protein synthesis plays an important role in neuronal processes, including learning and memory. FMRP can be detected in polysomes of synaptoneurosomes, neuronal preparations highly enriched in synapses. Synaptoneurosomes respond to stimulation by metabotropic glutamate agonists with fast-increasing polyribosome formation and accelerated protein synthesis. The synthesis of FMRP in the synaptoneurosomes increases after stimulation with the neurotransmitter glutamate, suggesting that synthesis of this protein in synaptoneurosomes is triggered by a class I glutamate receptor. On the other hand, different experiments reported that Fmr1 knockout mice have lowered expression levels of the glutamate receptor mGluR1 in the cerebral cortex and in the granule cell layer of the cerebellum, but not in the hippocampus. Thus, it has been suggested that this receptor alteration disturbs local protein synthesis and leads to impaired cortical synaptic plasticity. These findings suggest a specific function in the modulation of FMRP synthesis at postsynaptic sites and, in addition, that FMRP has a role in neurotransmitter-actuated polyribosome assembly. FMRP synthesis can also be affected by the brain-derived neurotrophic factor, BDNF, a known regulator of synaptic plasticity. BDNF downregulates the FMRP expression in cultured hippocampal neurons as well as in mouse brains. This downregulation by BDNF is a response to increased tyrosine kinase receptor signaling. The decreased FMR1-mRNA amounts could be correlated to a decreased amount of FMRP in the cell.

The *in vivo* trigger of FMRP synthesis is not known, but FMRP increases in the barrel cortex after whisker stimulation in rats, a model of experience-dependent plasticity. This increase was notably observed in subcellular fractions enriched for synaptoneurosomes and polyribosomes, suggesting a site-specific production of the protein. The altered level of FMRP most likely influences the translation of specific mRNAs in the synapse. Inhibition of the synthesis of its own mRNA, MAP1B, Arc, and α-CaMKII has been demonstrated, but it remains unclear whether all cellular RNAs that are bound by FMRP are translationally repressed.

Thus, FMRP plays an important role in translation of proteins in the dendrites after specific triggering. Abnormal synaptic protein synthesis in the absence of FMRP could underlie variable symptoms of the fragile X syndrome, including the presence of immature spines and impaired synaptic maturation.

5
Paralog Genes FXR1 and FXR2

FMR1 has two paralog genes called *FXR1* and *FXR2* (fragile X–related protein 1 and protein 2). FXR1 was identified by cross-hybridization with an FMR1 probe, while FXR2 was identified by yeast two-hybrid assays using whole FMRP as bait. The *FXR1* and *FXR2* genes have been localized to 3q28 and to 17p13.1 respectively. Overall similarity with FMR1 is high, especially in the amino-terminal half of the gene, and both FXR1 and FXR2 proteins contain identical functional domains as FMRP, including two KH domains, an RGG box, an NES, and an NLS (Fig. 9). Sequences

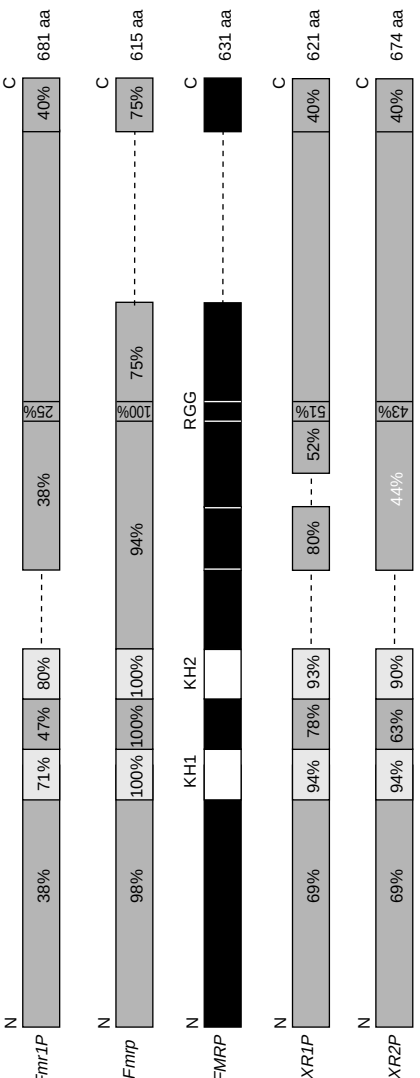

Fig. 9 Protein identity between human FMRP and human FXR1P and FXR2P, mouse Fmrp and Drosophila dFmr1P based on the alignment of Wan et al. (2000). The three RNA binding domains in the mouse Fmrp show 100% identity with the FMRP domains, whereas in the other proteins high identity is present in the KH domains and moderate identity in the RGG box region. Identity in the carboxyl end is not indicated because the low homology of the protein in this region does not allow an unambiguous alignment.

corresponding to exons 11 and 12 in FMRP are not identified in the two paralog genes. FMRP shows 86% and 70% homology to FXR1P and FXR2P respectively. Specific differences include the carboxyl end of FXR1P that is very different from that of FMRP and the FXR2P amino-terminus that contains an extra 10 amino acids compared to FXR1P and FMRP. The 5′ and 3′ UTR of FXR1 and FXR2 are up to five times shorter than that of FMR1. The 3′ UTR of FXR1 shows a high similarity with FMR1 and is highly conserved during evolution, while that of FXR2 shows no similarity with FMR1. No CGG repeat is present in the FXR1 sequence, while the 5′ UTR of FXR2 contains a short $(CGG)_4$ unit.

FXR1 has seven isoforms with the two major ones having a predominant mRNA, 2.2 to 2.4 kb in size. The protein FXR1P has a molecular mass of respectively 70 and 78 kDa. There is a tissue-specific alternative splicing for FXR1 pre-mRNA in the 3′ carboxyl part. The *FXR2* gene generates a 3 kb-sized mRNA and the protein FXR2P has a mass of 95 kDa.

The expression pattern of FXR1P and FXR2P shows some similarities but also important differences with that of FMRP. In the testis, FMRP is only expressed in the spermatogonia, whereas FXR1P is expressed in postmeiotic spermatides at high levels, especially in the microtubules of sperm tails, and FXR2P is expressed in the tubuli seminiferi. A specific isoform of FXR1P is highly expressed in the muscles and the heart, where FMRP expression is nearly absent. In the brain, both homolog proteins are expressed in neuronal cytoplasm. In fragile X patients, similar expression levels of FXR1 and FXR2 are noticed in all tissues compared to controls.

On the basis of the sequence similarity between FMR1, FXR1, and FXR2, it has been suggested that FXR1P and FXR2P may have a similar function in the cell as FMRP and can partially compensate for its loss in fragile X patients, explaining the relatively mild phenotype of patients. Similarities include the localization in the cytoplasm, the association with polyribosomes, and the capability to bind RNA. In addition, both the paralog proteins can form homodimers and heterodimers with each other and with FMRP (Fig. 7), though, like FMRP, FXR1P, and FXR2P are predominantly cytoplasmic proteins, with some expression observed in the nucleus. FXR1P is able to shuttle between the nucleus and the cytoplasm, while FXR2P and some of the isoforms of FXR1P shuttle between the nucleolus and the cytoplasm. However, though some functional redundancy seems likely, differences between the FXR proteins and FMRP are also apparent. Unlike FMRP, FXR1P and FXR2P do not inhibit translation and have a different localization in some tissues. Moreover, it is unclear whether FXR1P and FXR2P can bind G-quartet-containing mRNAs to the same extent as FMRP.

6 Amino-terminal Interacting Proteins

Using the N-terminal domain of FMR1 containing the NLS, Bardoni et al. (1999) identified a novel protein that interacts with FMRP by a yeast two-hybrid assay (Fig. 7). The interacting protein was called NUFIP (nuclear FMRP interacting protein). It shows no homology with known proteins in the databases. NUFIP maps to chromosome 13q14, while a pseudogene maps to chromosome 6q12.

NUFIP contains two C_2H_2 zinc finger motifs and an NLS. The protein is mainly localized in the nucleus in a dotlike distribution but is also observed in the cytoplasm, where it associates with ribosomes. NUFIP interacts with FMRP both via the N-terminal and the C-terminal half, suggesting the presence of two protein-interacting regions. Despite the fact that the 5′ region of the *FMRP* gene, used as bait to pick up NUFIP, is highly similar between FMRP, FXR1P, and FXR2P, no interaction between NUFIP and the other FXR proteins could be detected. The mouse NUFIP shows a 95% homology with the human NUFIP and has an expression pattern comparable with that of FMRP, with high expression in the neurons of the hippocampus and the cortex, and in the Purkinje cells of the cerebellum, but lower expression in the granular cell layer. No expression was detected in glia cells. NUFIP is also expressed in fibroblasts and lymphoblasts. Not only is the expression pattern of NUFIP similar to that of FMRP but, in addition, the RNA-binding capacities of NUFIP do resemble those of FMRP as NUFIP binds with high efficiency to poly(G), with less efficieny to poly(U), and shows no affinity to poly(A) or poly(C) homopolymers. The RNA binding domain of NUFIP has not been identified yet.

More recently, two additional FMRP interacting proteins, named CYFIP1 and CYFIP2 (cytoplasmatic FMRP interacting protein 1 and 2), were isolated. CYFIP1 and CYFIP2 are 88% identical on protein levels and both are conserved during evolution but do not show similarities with genes in the database. The CYFIP proteins are expressed in the synaptic terminals of neurons in the brains, where it colocalizes with FMRP to ribosomes in the cytoplasm. The first coiled coil region of FMRP, encoded by exon 7 and mediating FMRP–protein interaction, is involved in the interaction of FMRP with CYFIP proteins (Fig. 7). Despite the high sequence similarity between the CYFIP proteins, CYFIP1 interacts only with FMRP, whereas CYFIP2 interacts with FMRP, FXR1P, and FXR2P. Not much is known about the function of the CYFIP proteins, but CYFIP1 appeared to be an interactor of the Rac1 protein, a member of the Rho–GTPase family, which plays a role in maturation and maintenance of dendritic spines by controlling the actin cytoskeleton.

7 Animal Models

7.1 Comparison between Species

Orthologs of the *FMR1* gene have been identified in different species such as mouse, rat, chicken, *Xenopus laevis*, *Danio rerio*, monkey, and *Drosophila*. The protein is well conserved between species. The murine, frog, chicken, and fly proteins are 97, 86, 92, and 31% identical to the human FMRP respectively. Homology is nearly 100% within the KH domain, while a lower similarity in the RGG box region is observed. Outside the coding region, the 3′ UTR is well conserved among species. The 5′ UTR region is highly GC-rich in all species analyzed, and a CGG repeat has been detected in most mammals. Chicken contains a 99-unit-long CCT repeat in place of the CGG repeat, flanked on both sides by dinucleotide repeats. No instability upon generational transmission was observed in these repeats in species other than man and monkey.

7.2 What Have We Learned from the Mouse Model?

As in humans, the murine *FMR1* gene is located on the X chromosome. All functional domains present in the human protein, 2 KH domains, an RGG box, two coiled coils, an NLS, and an NES are well preserved (Fig. 9). In the 5′ UTR, a triplet repeat containing 9 to 11 CGG repeats has been identified, depending on the mouse strain, interspersed by 1 CGA triplet. No fragile site and no spontaneous CGG-repeat instability have been observed in the mouse. The expression pattern of the murine *FMR1* gene is comparable with that of the human gene with abundant expression in the testis and the hippocampus.

Approximately 10 years ago, a knockout mouse model was constructed. Using homologous recombination, the Fmr1 sequence was interrupted by a neomycin cassette in exon 5 of the gene. Although the knockout mutation in the mouse *FMR1* gene differs from the dynamic mutation in most fragile X patients, both mutations lead to the same result: absence of protein product. The fragile X knockout mouse is fertile, and visual inspection revealed no obvious abnormalities. It has been subjected to a wide range of pathological, physiological, and behavioral tests.

7.2.1 Macroorchidism

Macroorchidism is present in the great majority of fragile X patients after puberty. In the knockout mouse, macroorchidism can be observed from day 15 after birth, increasing to a difference of 30% in weight at half a year of age. Similar to observations in human patients, pathological observation revealed no structural abnormalities in the testis of the knockout mouse. It was concluded that increased Sertoli cell proliferation in the fragile X mice in a well-defined period shortly before and after birth leads to higher spermatogenic cell number and testicular weight.

7.2.2 Cognitive Testing

Spatial learning in mice can be measured using the Morris water-maze test. In this test, mice have to find a hidden platform in a basin filled with opaque water during 12 learning trials. In the reversal phase, the platform is moved to the opposite position in the basin, and mice have to find the platform in its new position during four additional trials. It is particularly in this reversal phase that mild learning deficit is observed in the knockout mice. Similar observations are reported when mice were tested using the plus-shaped water maze and the terrestrial radial maze. The nature of the spatial deficit is not clear. The deficit may point to hippocampal defects, but other brain areas can also play a role in the observed differences in this reversal test.

7.2.3 Motor Activity and Anxiety

Fragile X patients are reported to be hyperactive and anxious. Fragile X knockout mice are more active than their control littermates. However, as the difference in activity between groups is mild and variance within groups is high, hyperactivity may be too small to measure when registered during a single time period as illustrated in Fig. 10.

A seemingly contradictory observation is that Fmr1 knockout mice displayed increased anxiety-related responses in the exploratory behavior test, whereas lesser anxiety is reported when measured in the open field test. However, no difference in anxiety was registered in the elevated plus

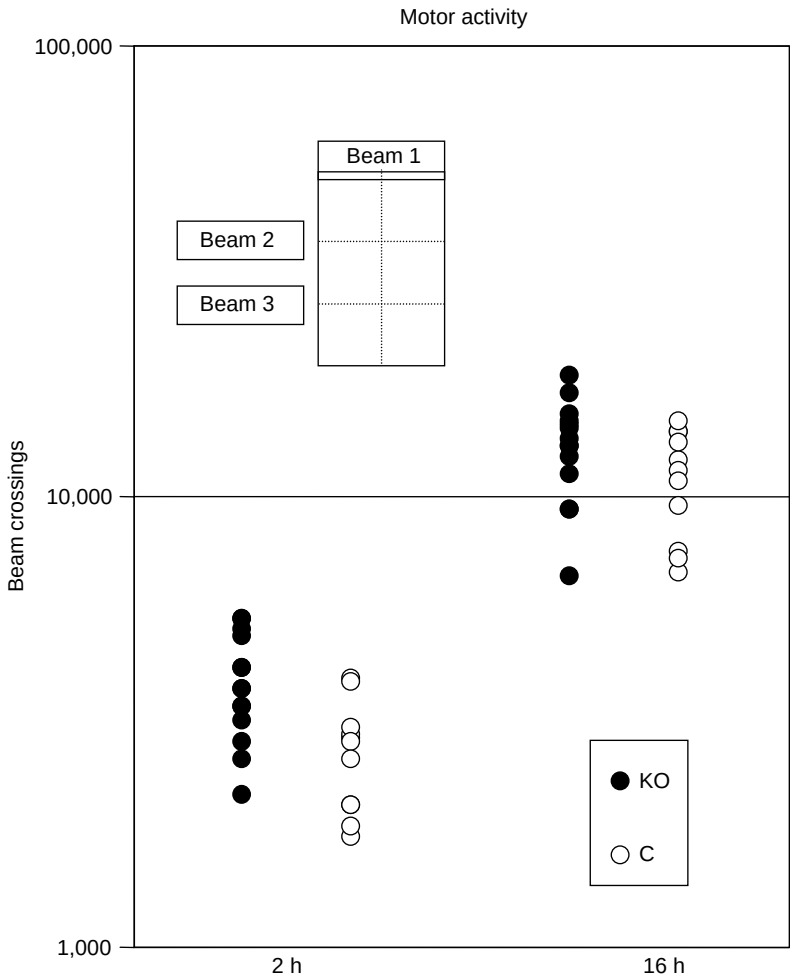

Fig. 10 Motor activity testing of the fragile X knockout mouse in comparison with control littermates. In one experiment, the total number of beam crossings was significantly elevated after 2 h of registration ($P < 0.005$), while after 16 h no statistical difference (N.S.) between both groups was observed. When the individual data were considered (black circles, knockout; white circle, control), it is noticed that a single outlier in the group of individual mice influenced the test result.

maze, considered to be the most suitable test to measure anxiety.

7.2.4 Acoustic Startle Reflex

The acoustic startle response is a behavioral tool to assess brain mechanisms of sensorimotor integration and is mediated by neurons in the lower brainstem. The knockout mice exhibited increased auditory startle-response amplitudes to low-intensity stimuli and decreased responses to high-intensity stimuli. The increased response to low-amplitude stimuli is compatible with the hyperarousal of fragile

X patients, while the decreased reactivity of the fragile X knockout mouse to high-amplitude stimuli is compatible with the decreased functioning of the neuronal connections in fragile X patients.

Prepulse inhibition was also affected in fragile X mice. In this variant of the acoustic startle response, a prepulse preceding the stimulus itself suppresses the response. Prepulse inhibition is dependent on a more complex forebrain circuitry. Inhibition was stronger in fragile X knockout mice compared to controls, suggesting involvement of the forebrain pedunculopontine nucleus pathway in fragile X syndrome.

7.2.5 Audiogenic Seizures

Spontaneous seizures are observed in more than 20% of the fragile X patients. These spontaneous seizures are never observed in the knockout mouse, but an increased sensitivity to epileptic seizures in response to auditory stimuli is detected in the knockout mouse. This may indicate that absence of Fmrp results in increased cortical excitability and is consistent with the sensory hypersensitivity of fragile X patients.

7.2.6 Neuroanatomy and Physiology of the Knockout Brain

Magnetic resonance studies of mouse brains did not reveal size differences in brain areas reported to be of abnormal size in human patients. As in human fragile X patients, impaired neuronal maturation is observed in the knockout mice. In cortical layer V of knockout mouse brains, pyramidal neurons have more elongated, thin, immature, and tortuous spines along the dendrites. In addition, a trend toward a higher spine density, characteristic for human fragile X patients, was observed. Since the knockout mice have an excess of immature spines, synaptic connectivity may be impaired. This may suggest that the fragile X protein is required in synaptic pruning and maturation processes in neurons, although at present it has not been established whether the neural abnormalities are the cause or the consequence of the mental retardation.

A detailed study of the hippocampus did reveal possible subtle neuroanatomic abnormalities. Sections through the hippocampus revealed excessive sprouting of the intra- and infra-pyramidal mossy fibers in subfield CA3 in knockout mice. This suggests increased axonal branching and synaptogenesis in the knockout mouse, perhaps related to the increased sensitivity to seizures. However, a decrease in size of the mossy fibers in the same hippocampal area was reported in an independent study. The reason for the discrepancy between both studies is not clear but could be related to strain differences. FVB mice with an unusually small CA3 subfield were used in the first-named study, whereas C57BL/6 mice with a relatively large CA3 subfield were used in the second study.

Detailed studies of dendritic processes of spiny stellate neurons in barrels, the cylindrical arrangements of neuronal cell bodies that define processes in the somatosensory cortex for the large facial whiskers, revealed additional subtle abnormalities. During normal development, dendritic processes extend in all directions from the cell body. Branches directed toward the interior of the barrel continue proliferating during development and those directed away from the interior are withdrawn. In knockout mice, no differences were observed for branches directed toward the barrel interior, while the branches

directed away from the interior remained more extensive compared to the controls.

Metabolic rates are also altered in the knockout mouse. Qin et al. (2002) measured the cerebral metabolic rates for glucose, CMR_{glc}, as an indicator of the level of functional activity in 38 brain regions of fragile X knockout mice and wild-type littermates. In 26 of the 38 regions, the CMR_{glc} level was significantly increased in the Fmr1 knockout mice, particularly in the regions of the limbic system and primary sensory and posterior parietal cortical areas. Interestingly, these areas are involved in motor activity and exploration, processes that are affected in the knockout mice. Thus, the increased level of the energy metabolism in cortical structures correlates with abnormal dendritic spine morphology and an increased excitatory state.

7.2.7 FMRP is Involved in Long-term Plasticity

Long-term potentiation (LTP), a long-lasting increase in synaptic efficiency, is believed to be involved in learning and memory. No altered LTP was measured in the hippocampus of the knockout mouse, but Li et al. observed severely reduced LTP when measured in the cortex. Reduced LTP could be correlated with a decreased quantity of GluR1 expression in the central synapses.

Long-term depression (LTD), a second form of hippocampal synaptic plasticity, is the opposite of LTP. It is triggered by the activation of group 1 metabotropic glutamate receptors. LTD appeared significantly enhanced in the hippocampus of the knockout mouse. While at first an unexpected finding, increase in LTD is consistent with the finding that Fmrp is a negative regulator of translation. As LTD is protein synthesis–dependent, absence of the translational inhibitor, FMRP, may inhibit the protein synthesis necessary for LTD (Fig. 11).

Since LTD and LTP are involved in learning and memory, it is tempting to speculate that abnormalities in synaptic connections in the knockout mouse may be related to the cognitive deficits in the fragile X syndrome.

7.2.8 FXR1/FXR2 Knockout Mice

It could be speculated that the abnormalities observed in the fragile X knockout mouse are relatively mild because the paralog genes *Fxr1* and *Fxr2* partially take over *Fmrp* function. Knockout mice for both paralog genes have been generated. Mice homozygous for the *Fxr1* knockout die shortly after birth because of respiratory problems (Siomi, personal communication). Mice heterozygous for the *Fxr1* mutation seem to be lighter and smaller than control littermates but not much is known about their phenotype. The *Fxr2* knockout mouse is viable and fertile and shows no obvious abnormalities. It has been subjected to a similar range of behavioral and pathological tests as the fragile X knockout mouse. Like the *Fmr1* knockout mouse, the *Fxr2* knockout mouse shows no obvious pathological abnormalities, is hyperactive, and performs poorly in the Morris water-maze test. Double or triple mutants of *Fmr1*/*Fxr* mice have not been described as yet, but such multiple knockouts may show us more of the possible redundancy in function of the Fxr proteins.

7.3 What Have We Learned from Drosophila?

7.3.1 The Ortholog, dFmr1

dFmr1, the *Drosophila melanogaster* ortholog of the *FMR1* gene, appears to

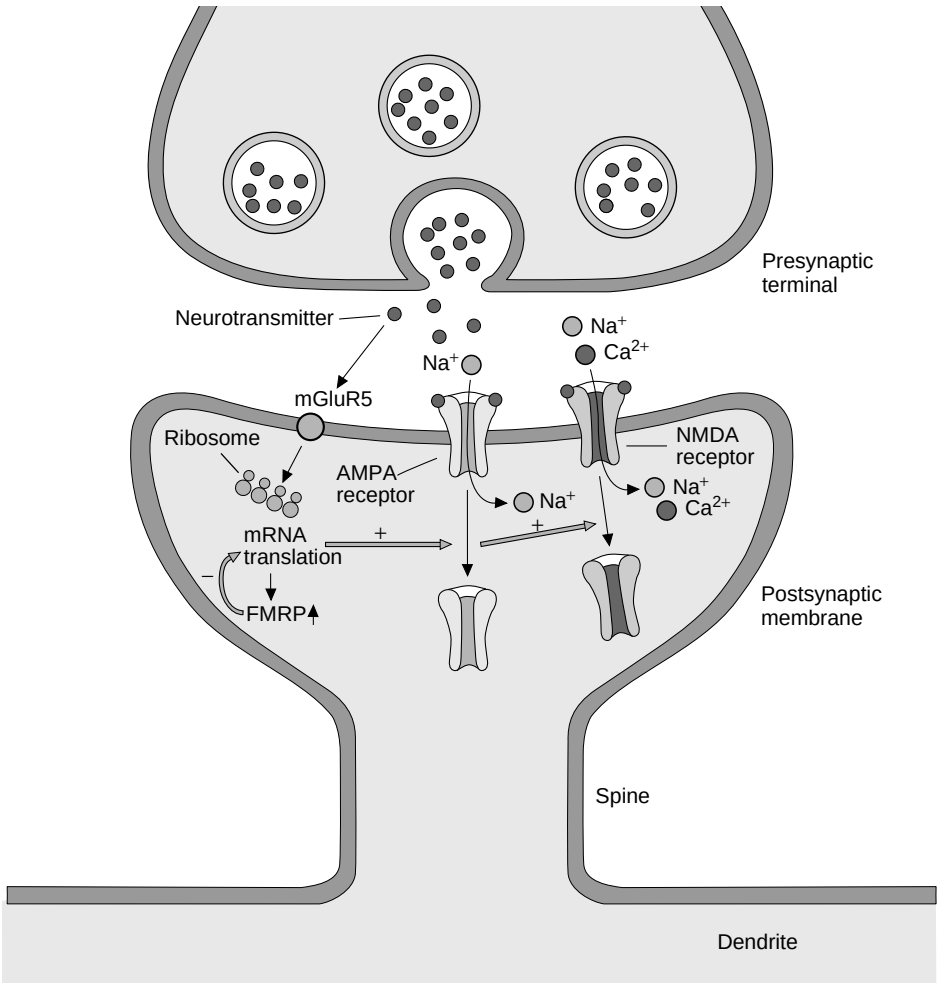

Fig. 11 Hypothetical model of the regulation of FMRP translation in a hippocampal synapse adapted from Huber et al. (2002). Stimulation of mGluR5 induces local mRNA synthesis. This results in novel protein synthesis that, in turn, stimulates the internalization of the AMPA and NMDA glutamate receptors, both essential for long-term plasticity. One of the proteins stimulated by mGluR5 activity is FMRP. As FMRP is a negative regulator of transcription, an increase in FMRP may result in a local decrease in mRNA translation. The resulting decrease in protein synthesis may slow down the internalization of the glutamate receptors, resulting in an increase in LTD.

be the only gene of this protein family in *Drosophila* with equal similarity to FMR1, FXR1, and FXR2 (Fig. 9). No CGG repeat is present in the 422-bp-long 5′ UTR region. Alternative splicing in dFmr1 occurs across the gene giving rise to different splice forms, but nothing is known about the expression of these splice products yet.

Like the human FMRP, the dFmr1 protein contains 2 KH domains, an RGG box with 50% identity to the human FMR1 RGG box, an NES, and an NLS. The dFmr1 NES shows little sequence conservation when compared to human FMRP and it remains to be determined whether it is a functional nuclear export signal. The region involved in the binding of FMRP to the 60S ribosomal subunit is also present in dFmr1.

The protein can be observed in the cytoplasm, but a nuclear localization has not been detected. dFmr1 is expressed in early embryo starting from gastrulation. It is highly expressed in neurons but not in glia. It is not enriched in synapses. dFmr1 is moderately expressed in muscles, in presynaptic motor neurons, and in postsynaptic muscles during embryonic development and in the larvae.

Using heteropolymer binding essays, dFmr1 shows high affinity for poly(G), weaker affinity for poly(U), and no affinity for poly(C) and poly(A), characteristics identical to the human protein. Point mutations have been introduced in the highly conserved isoleucine residues, corresponding to the human Ile304Asn mutation, within each of the KH domains (I244N and I307N). Like mutated human FMRP, mutated dFmr1 loses the capacity to bind poly(U); dFmr1 binds with high capacity to human FMR1, but shows only moderate binding capacities with FXR1 and FXR2 and with itself.

7.3.2 dFmr1 is Involved in Neuronal Maturation

Different mutant flies, including dFmr1 null mutants, dFmr1 overexpression mutants, and I244N and I307N KH point mutants, have been constructed to study the role of dFmr1 in *Drosophila*. The introduction of dFmr1 null mutants is not lethal, but some mutants fail to eclose properly, and only a few "escapers" survive. Overexpression of dFmr1 leads to severe rough-eye phenotype, whereas null mutants have no detectable effect in the structure of the eye. Photoreceptor development and function are not effected by loss of dFmr1, but overexpression of dFmr1 in the antennal disc of the eye causes apoptosis.

As observed in humans and mice, absence of dFmr1 and even different concentrations of dFmr1 compared to controls causes aberrant synaptic growth and plasticity. While overexpression of dFmr1 gives undergrowth of synapses and causes presynaptic enlargements of single synaptic boutons of twice the size, the null mutant of dFmr1 shows synaptic overgrowth and overelaboration of synaptic terminals but has no visible effect on the size of the synaptic boutons. Null mutants of dFmr1 exhibit aberrant small ectopic neurite branches, probably causing a misregulation of the intrinsic branching program. There is also a mistargeting of axons and looping fibers, potentially because of guidance defects. In dFmr1 null mutants, fiber extension abnormal morphology in the dorsal-cluster (DC) neurons is observed, as shown in Fig. 12. The level of dFmr1 determines the severity of the phenotype. The phenotype of

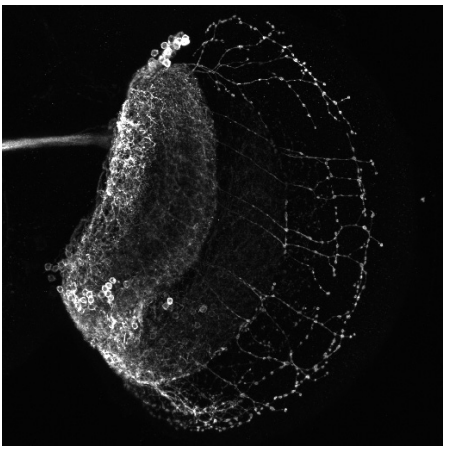

(a)

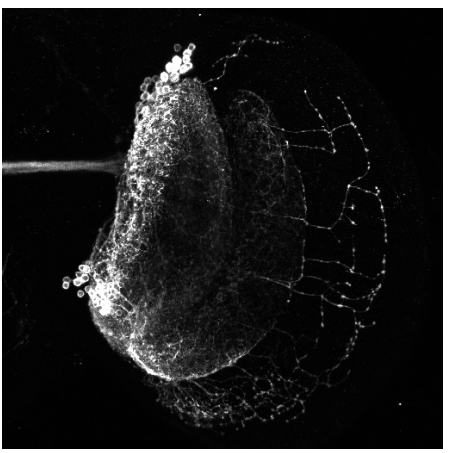

(b)

Fig. 12 Illustration of abnormal neuronal migration in *Drosophila* dFmr1 null mutant. Comparison of (a) heterozygous with (b) homozygous mutants shows a reduction in the number of dorsal-cluster neuron axons crossing from the lobula into the medulla. This finding suggests that in dorsal-cluster neurons, dFMR1 is required for proper extension and pathfinding of axons. (Photograph kindly provided by Katsu Okamura and Haru Siomi.)

the overexpression mutant is more severe than that of the loss-of-function mutant. dFmr1 overexpression caused a complete failure of axon extension from the lobula to the medulla. Overexpression of an I307N mutant dFmr1 causes defects similar to those observed in loss-of-function mutants but are less severe. These findings suggest that dFmr1 acts as an inhibitor of axon extension and as a regulator of axon growth and synapse formation in the cell.

7.3.3 MAP1B may be Involved in the Fragile X Syndrome

MAP1B is a protein that regulates the growth of dendrites, axons, and synapses. The *Drosophila* homolog for MAP1B has been identified as futsch. Using dFmr1-specific antibodies, futsch mRNA could be coprecipitated, suggesting that it may be part of the dFmr1–mRNP complex. MAP1B has also been identified as an *in vivo* target of FMRP and its mRNA contains a G-quartet. The level of futsch mRNA in *Drosophila* is inversely related to the level of dFmr1. Null mutants of dFmr1 display an increase in futsch mRNA, while overexpression mutants show a decrease in futsch-mRNA level. Overexpression of futsch in the fly causes neuromuscular junction overgrowth and increased bouton number, synaptic area, and branching. Introducing double futsch/dFmr1 null mutations rescued the structural and functional defects compared to dFmr1 mutant flies. This may suggest that dFmr1 regulates synaptic maturation and plasticity by functioning as a translational repressor of futsch mRNA.

7.3.4 dFmr1 is Involved in Circadian Rhythm Regulation

Most dFmr1 loss-of-function mutant flies develop until the pharate adult stage

but fail to eclose in the early morning. Moreover, while wild-type flies eclose during the early morning, dFmr1 null mutants may eclose at all times of the day. This suggests a defect in the circadian rhythm. This could be demonstrated by activity measurements. dFmr1 mutants show severely decreased activity levels without any day–night periodicity when the frequency with which flies cross an infrared beam was measured in the locomotor activity test, whereas wild-type flies exhibit a rhythmic pattern of rest and activity during a 23.5 h circadian cycle. The circadian defect was rescued by introducing the wild-type *dFmr1* gene as a transgene into the null mutant, while overexpression of dFmr1 lengthened the circadian pattern. Futsch mutants and dFmr1/futsch double mutants have no circadian defect and no change in locomotor activity or strength of rhythmicity.

Independently, Dockendorff et al. (2002) and Morales et al. (2002) concluded that clock genes are not affected in the dFmr1 mutant flies, suggesting that the dFmr1 mutations may have a strong effect on the pathway downstream from the endogenous clock components. Consistent with this hypothesis, the amplitude of the oscillations of the cAMP responsive-element binding protein, CREB, was found to be reduced in the mutant dFmr1 flies, suggesting that dFmr1 affects this known molecular output of the clock.

7.3.5 dFmr1 may be Involved in RNA Interference

dFmr1 was found to interact with L5 and L11, two large ribosomal subunit proteins. This demonstrates that, like the human fragile X protein, the dFmr1 protein associates with the 60S ribosomal subunit in *Drosophila*. The finding that not the dFmr1 region equivalent to the ribosomal binding domain in human FMRP but the 285–amino acid–long C-terminal end of the dFmr1 protein binds both ribosomal proteins was peculiar. The L5 protein is associated with 5S rRNA and forms a ternary complex by binding dFmr1 protein, suggesting that dFmr1 interaction with ribosomes takes place via the L11 protein and the L5/5S rRNA complex.

Two other proteins found associated with the dFmr1 protein in *Drosophila* are AGO2 and Dmp68. AGO2, a member of the Argonaute gene family, is an important component of the RNA-induced silencing complex (RISC), which recognizes and cleaves homologous mRNAs and mediates RNAi in Drosophila. Dicer, an enzyme that processes both miRNA and siRNA was also part of the complex containing AGO2 and dFmr1. p68, the human ortholog of Dmp68, unwinds short dsRNA in an ATP-dependent manner. Dmp68 is also required for efficient RNAi. Binding to both proteins suggests that dFmr1 may be involved in RNAi.

7.3.6 dFmr1 is Modified After Translation

CKII is a protein that phosphorylates nuclear and nonnuclear proteins and is highly conserved among species. Recent studies identified *Drosophila* casein kinase II as a protein phosphorylating dFmr1. The phosphorylated Ser406, located upstream of the RGG box in dFmr1, is a conserved amino acid between the fly and human FMR1 and FXR proteins. *In vitro*, replacement of the serine with alanine residues in dFmr1 protein and human FMRP abolishes CKII phosphorylation. Phosphorylation affects the efficiency of dFmr1 to form homomers and to bind other RNAs. Though phosphorylation was observed *in vivo*, the significance of this posttranslational modification remains unclear.

8
Phenotypic Effect of the Premutation

With the advent of a DNA-based diagnostic test, the study of premutation carriers has increased. Individuals with a premutation were initially thought to have no phenotypic manifestations. However, a number of studies have reported mild learning disabilities and emotional problems in a small subgroup of premutation carriers. Subsequent study demonstrated that these findings challenge the view that premutation carriers of FraX are unaffected. Premutation carriers have an increased risk for premature ovarian failure (POF) in females. Hundscheid et al. (2000) found that paternally inherited premutations were more likely to give rise to POF than maternally inherited premutations. This imprinting was not seen by the other groups, perhaps as a result of differences in the age of the study population or differences in inclusion criteria.

Recent studies in premutation carriers showed that the length of the CGG repeat has an inverse correlation with the FMRP levels, while it has a positive correlation with the FMR1 mRNA levels. Tassone et al. (2000) described premutation carriers with more than 100 CGG repeats having a reduced number of FMRP-positive cells but with at least fivefold elevated mRNA levels. These findings lead to the hypothesis of a compensatory mechanism. If the translation of FMR1 mRNA is less efficient or defective, low FMRP levels initiate via feedback induction FMR1 transcription. The resulting high FMR1 mRNA levels can only partially compensate the defect since normal FMRP levels are not achieved in cells with large premutation repeats. It is yet unknown whether FMRP levels itself are the crucial factor at work in this compensatory mechanism or whether it is coupled with the length of the CGG repeat. The expanded CGG repeat may lead to proportionally more open promoter conformation and consequently enhanced transcriptional activity. The normal FMR1 message levels in the mutant I304N cell line argues against a role for the lack of functional FMRP as a modulator of transcriptional activity of the *FMR1* gene. Another candidate for the upregulation of the FMR1 expression might be CGGBP1, a CGG-binding protein, that has shown the ability to regulate expression from the *FMR1* gene with a small CGG repeat. It can be hypothesized that in the case of an elongated CGG repeat, the CGGBP1 might be sequestered owing to extensive binding to the elongated repeat in the mRNA, resulting in a lowered active concentration of CGGBP1. A lower CGGBP1 concentration results in an increased FMR1 transcription.

Elevated FMR1 mRNA levels and/or reduced FMRP levels result in a phenotype recently described in older males carrying a premutation. These males may exhibit a unique neurodegenerative syndrome characterized by progressive intention tremor and ataxia. More advanced cases are accompanied by memory and executive function deficits, anxiety, and eventual dementia. MR imaging studies (T_2 signal) of the brain of symptomatic adult male premutation carriers showed a characteristic imaging, including hyperintensities of the middle cerebellar peduncle, cerebellar white matter lateral, superior, and inferior to the dentate nuclei, and volume loss involving the pons, mesencephalon, cerebellar cortex, cerebral cortex, white matter of the cerebral hemispheres, and corpus callosum. Neurohistological studies on the

brains of four symptomatic elderly premutation carriers demonstrated neuronal degeneration in the cerebellum and the presence of eosinophilic intranuclear inclusions in both neurons and astroglia. Furthermore, the inclusions showed a positive reaction with antiubiquitin antibodies, which suggests a link with the proteasome degradation pathway. The origin and constitution of the inclusions is poorly understood; however, elevated FMR1 mRNA levels have been proposed to be important for the formation of the inclusions.

A "knock-in" CGG triplet mouse has been generated that showed moderate CGG-repeat instability upon both maternal and paternal transmission. These mice showed an increased expression of FMR1 mRNA and a decreased expression of FMRP. The presence of ubiquitin-positive intranuclear inclusions was seen in neurons of the CGG expanded-repeat mouse. An increase in both the number and the size of the inclusions was observed during the course of life, which correlates with the progressive character of the cerebellar tremor/ataxia syndrome in humans. The observations in expanded-repeat mice suggest a correlation between the presence of intranuclear inclusions in distinct regions of the brain and the clinical features in symptomatic premutation carriers. The presence of inclusions in expanded-repeat mice supports a direct role of the *FMR1* gene, by either CGG expansion *per se* or mRNA level, in the formation of the inclusions, which in humans could have also be explained by a synergistic effect with another genetic locus.

These two clinical features are remarkable in that they are unique to premutation carriers, while full-mutation individuals are not affected. As the number of male premutation carriers is 10 times the number of males with a full mutation, the number of individuals who are at risk for the ataxia/tremor phenotype is large. It is not known whether females with a premutation suffer from this asimilar phenotype. If so, the age of onset can be expected to be later than in males.

9 The FMR1 Mutation

The FMR1 CGG repeat containing less than 50 CGGs is very stable. However, the CGG repeat above 50 repeat units is remarkably unstable. In the lower premutation range, the repeat can increase and decrease in size while staying in the premutation range. In the upper premutation range, the repeat can undergo massive expansions to more than several hundred repeats in full-mutation alleles. Two reports have suggested that full-mutation expansions are more likely to occur in male than in female offspring. However, a recent multicenter study found no significant differences in the proportion of full-mutation alleles in male or female fetuses.

9.1 Mechanism of CGG-repeat Expansion

Repeat instability is influenced by several factors. First, only transmission through the female germ line can give rise to a full mutation in the next generation. In contrast, transmission through the male germ line, irrespective of whether the male has either a full mutation or a premutation in his somatic cells, always reveals a premutation in the daughters. Second, the instability of premutation alleles in females is positively correlated

with the repeat size. They may expand to a different-sized premutation or a full mutation, larger repeats being more prone to expansion than smaller ones. Third, not only the length of the repeat is a risk factor for repeat instability but also the composition of the repeat. Cryptic AGG interruptions in the CGG repeat might provide stability to the repeat and instability might result from loss of AGG interruptions. In the normal population, the CGG repeat is interrupted by AGG trinucleotides, most often at positions 10 and 20. Loss of AGG repeats leads to alleles with longer perfect CGG-repeat tracts, which are prone to expansion, especially when the number of pure CGG units exceeds 30. Premutation alleles have either no or only one AGG repeat unit and are therefore unstable.

Loss of these AGG interruptions always occurs at the 3′ end of the repeat tract, and as a consequence, the longest pure CGG tract found within a premutation or full mutation is always located at the most 3′ end. The reason for this polar variation might involve the mechanisms underlying the repeat instability. The polarity of expansion of the CGG repeats in humans mimics the orientation-dependent instability of CGG repeats cloned into *Escherichia coli* and *Saccharomyces cerevisiae*. This finding suggests that a difference in leading- versus lagging-strand synthesis may be involved in repeat expansion.

Nonstructural, cis-acting factors may be related to a specific haplotype background as certain haplotypes of polymorphic markers in and around the *FMR1* gene are overabundant in fragile X chromosome. Dombrowski et al. (2002) found, in a population-based study, that gray zone alleles or protomutations (40–55 repeats) also are often associated with a specific fragile X syndrome-related haplotype. However, others, in a much smaller number of unstable events, did not find evidence for a specific haplotype as a risk factor for instability.

Studies in yeast and *E. coli* suggest that direction of replication, genetic background (including repair systems), transcription, and growth conditions have an influence on repeat instability too. Repeat instability is thought to be instigated by the formation of d(CGG) secondary structures. Several structures of the CGG repeat, such as hairpins and tetrahelical structures, have been suggested. The stability of the different structures seems to differ between both templates. Pausing of DNA polymerases at both hairpin and tetrahelical structures was observed *in vitro* as well as *in vivo*. This stalling of DNA polymerase might increase the probability of DNA polymerase slippage and this can result in expansion of the CGG repeat.

White et al. (1999) proposed the influence of RAD27, the yeast homolog of human FEN1, as a trans-acting factor on instability. RAD27 is necessary for correct processing of Okazaki fragments and its absence strongly increased CGG instability. Interestingly, White et al. (1999) did not observe any increase in the frequency of large expansion events when studying various CGG arrays in *mlh*1, *msh*2, *msh*3, and *msh*6 mutant backgrounds. In fact, it has been surprisingly concluded that repeat instability requires mismatch repair proteins.

It is likely that a multistep process is involved in repeat instability since it has never been observed that a normal repeat expands to a full-mutation repeat upon one transmission. The first step assumed to be involved in the initial expansion of trinucleotide repeats is slipped-strand

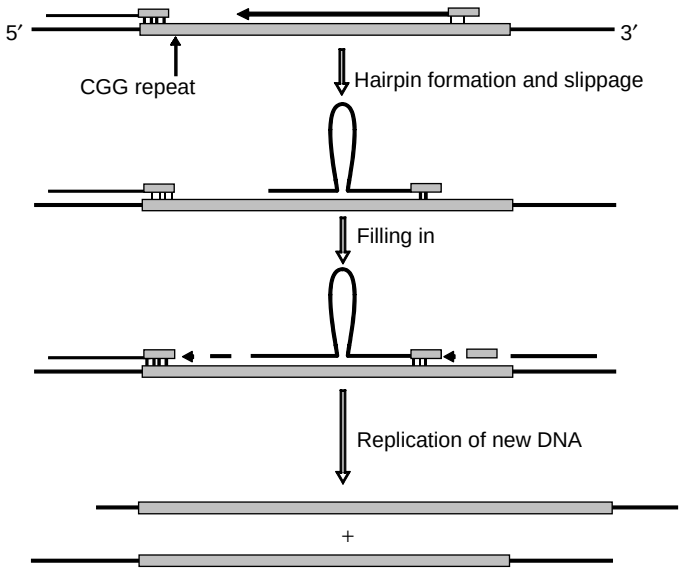

Fig. 13 Model for instability and expansion of the CGG repeat. During replication of the CGG repeat, a hairpin is formed followed by filling in of the gap, resulting in expansion of the repeat during multiple rounds of replication.

mispairing between DNA strands during DNA replication (Fig. 13) or possibly repair. Pausing of DNA polymerases allows the formation of secondary structures on the nascent lagging strand, which, if not correctly excised, should result in a small (up to 10 CGG-repeat units) increase. This slippage model can explain relatively small changes that occur within the normal and premutation alleles. The slippage model is the favored model, although other models, such as unequal meiotic sister chromatid exchange and/or gene conversion, are difficult to rule out.

For transition from premutation to full mutation, a different mechanism must be considered and it is thought that extremely unstable structures are formed in the nascent lagging strand owing to slippage on both ends of Okazaki fragments when these fragments contain solely CGG repeats. Incorrect repair will then lead to large expansions but might also result in contractions or deletions.

9.2 Timing of CGG-repeat Expansion

The timing of the repeat expansion and the difference observed in male and female transmission of the repeat is an intriguing feature of the mode of inheritance of the fragile X syndrome. How and when expansions from premutation to full mutation occur is not known.

Oocytes from full-mutation female fetuses have shown to contain a full mutation without a detectable premutation, suggesting a prezygotic model for expansion. In male fetuses with a full mutation in their somatic cells, the developing testes also show absence of FMRP in most cells. Testis tissue from a 13-week-old full-mutation fetus showed only full mutations

in the germ cells, while testis tissue from an older fetus showed evidence of both full and premutations in the germ cells. Adult full-mutation males show only a premutation in their sperm cells. The most likely model explaining these findings is the prezygotic model in which the full mutation is transmitted in the oocyte and the mutation in sperm-producing cells is regressed to a premutation size. While data by Malter et al. (1997) point to a prezygotic expansion, expansion occurring very early in embryogenesis (before day 3–5) cannot be ruled out.

The somatic mosaicism, often seen in patients, can be explained by instability that occurs postzygotically. Instability in the full-mutation range can occur, but also regression of the full mutation to a premutation can occur in some cells in somatic tissue. To study the exact timing of repeat expansion, an animal model will be required. The availability of an animal model to study the CGG-repeat instability will allow studying the repeat length at different time points during development.

10 Diagnostics

The identification of the *FMR1* gene and the mutational mechanism responsible for the fragile X syndrome introduced reliable DNA-based diagnostic tests, focused on the detection of the CGG-repeat amplification in patients and carriers, with or without associated hypermethylation. The two DNA methods to determine the CGG-repeat size are Southern blot hybridization and the polymerase chain reaction (PCR).

The availability of antibodies directed against FMRP allowed a diagnostic test to identify fragile X patients on the basis of the presence of FMRP in cells from unaffected individuals and its absence in cells from fragile X patients. Table 1 gives an overview of the different mutations in fragile X patients and carriers.

Tab. 1 Overview of the different mutations in fragile X patients and carriers.

Individual	Mutation	MR-associated CpG island	Number of repeats	Methylation of FMR-1	mRNA transcription	Risk of affected offspring
Normal	–	–	6–50	–	+	–
NTM	Pre-	–	50–200	–	+	–
Daughter of NTM/female carrier of a premutation	Pre-	–	50–200	–	+	+
Female carrier of a full mutation	Full[a]	70%[b]	>200	+[c]	+[d]	+
Male patient	Full[a]	+	>200	+	–	–

Note: MR: mental retardation; NTM: normal transmitting male.
[a]Somatic mosaicism is seen in about 40% of male patients (Nolin, 1994 #180). It is also observed in female carriers.
[b]In about 70% of females carrying a full mutation, mild-to-moderate mental retardation is found.
[c]Methylation is found on the inactive X and on the X chromosome carrying the full mutation.
[d]mRNA is only transcribed from the normal active X.

10.1 DNA Diagnostics

Very few fragile X cases have been reported without amplification of the CGG repeat and with either point mutations or deletions in other parts of the *FMR1* gene and DNA diagnostics is concentrating on detecting the repeat size.

The ability to detect both premutations and full mutations at the DNA level has made the diagnosis of the widespread disease available and has made genetic counseling possible through reliable detection of female and male carriers and patients, and can be used for prenatal diagnosis. Fragile X syndrome testing is based on the detection of CGG-repeat elongation in patients. This is based either on a Southern blot of leukocyte DNA digested with specific restriction endonucleases or on direct amplification of the CGG repeat using flanking primers.

10.1.1 Southern Blotting

The length of the CGG repeat can be determined by Southern blot analysis and mostly *Eco*RI or *Hind*III are used. Normal alleles will give a fragment of 5.2 kb. Longer fragments indicate the presence of a pre- or full mutation. Premutations are visible as fragments somewhat larger than 5.2 kb (up to 5.7 kb). Full mutations are visible as diffuse broad bands or "smears" owing to the presence of different repeat lengths (above 200 repeats) in different cells. Digestion with *Pst*I gives smaller fragments and, therefore, allows a better detection and sizing of premutations. By using an additional methylation-sensitive enzyme such as *Eag*I or *BssH*II, the length and the methylation status can be determined. This is applied to distinguish large premutations from small full mutations. For prenatal Southern blot diagnosis, digestion with a single enzyme (*Hind*III) is sufficient, as methylation has not occurred yet at the time (11 weeks) when chorionic samples were taken. DNA analysis is usually carried out using radioactive-labeled DNA probes; however, a good alternative is the use of nonradioactive (digoxigenin-labeled) probes followed by chemiluminescent detection.

10.1.2 Example of DNA Diagnostics in a Fragile X Family

Postnatal and prenatal diagnosis of fragile X syndrome can now be exercised with a high degree of accuracy by detecting the CGG-repeat amplification at the DNA level. Figure 14 shows a fragile X family in which premutations and full mutations are found. In the DNA of the grandmother (1) one normal 5.2 kb band (indicated by an arrow) is seen representing her 2 normal X chromosomes. The grandfather (2) is a normal transmitting male with a premutation of 70 repeats in the *FMR-1* gene; his DNA shows a band slightly larger than 5.2 kb. He passed his premutation on to his three daughters (4, 8, and 14) who are phenotypically normal carriers. Besides the 5.2 kb band of their normal chromosome, they have a premutation band of increased size (75 and 80 repeats). Daughter 4 has 1 affected daughter (6) who is mosaic for a premutation and a full mutation. Daughter 8 has 2 affected sons (12 and 13) with full mutations. Daughter 14 has one affected son (16) who is mosaic for a premutation and a full mutation.

10.1.3 PCR Amplification Analysis

Detection and accurate sizing of premutations can be done by PCR. The high CG content requires a specific PCR protocol. Initially, radioactivity was used to detect

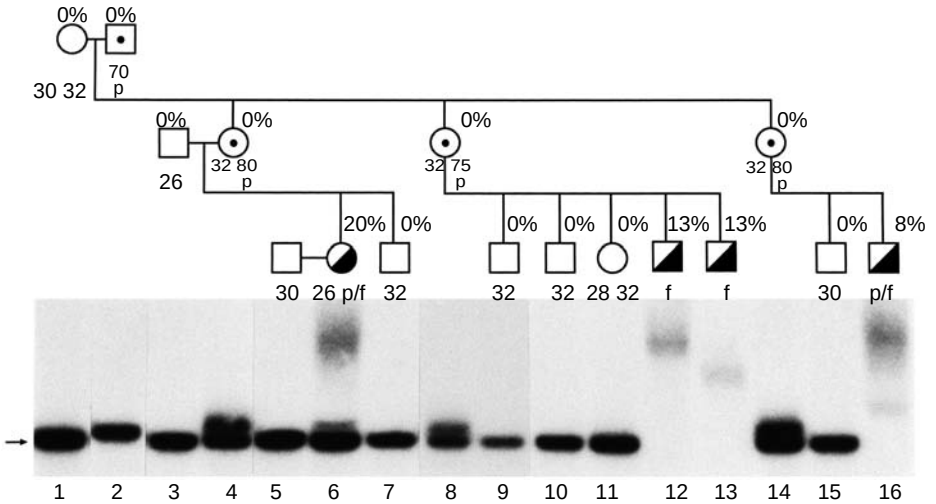

Fig. 14 Detection of the *FMR1* gene in a fragile X family by means of Southern blot analysis. {symbol individual 2} = normal transmitting male; {symbol individual 4} = normal female carrier; {symbol individual 6} = female carrier with cytogenetic expression; {symbol individual 12} = mentally retarded male; percentage of cytogenetic expression (if determined); number of CGG repeats for the different X chromosomes (determined by PCR) is indicated; p, premutation; f, full mutation. (Example of fragile X DNA diagnosis provided by Dr D. J. J. Halley and Dr A. M. W. van der Ouweland.)

the size of the PCR products. Alternatively, a nonradioactive method was developed using fluorescent-labeled precursors during the PCR amplification, and the size of the PCR product can be determined on an automated ABI sequencer. This last method gives a very accurate sizing of the PCR product in both males and females, starting from a very small amount of DNA.

Amplification of full mutations is very inefficient because it is difficult to amplify stretches of more than 100 to 200 CGG repeats, and it is therefore not a reliable method. Recently, an alternative PCR approach was described for the diagnosis of fragile X syndrome based on the methylation-sensitive conversion of C residues to U by bisulfite on single-strand DNA and subsequent amplification of the antisense strand with specific primers. A PCR with primers for methylated C residues will amplify the CpG dinucleotide region upstream of the CGG repeats exclusively in affected males. Moreover, the bisulfite treatment dramatically reduces the C + G component of the region; thus, the high T_m and the strong secondary structures are no longer obstacles for PCR amplification. This method can be used to identify males with a full mutation but cannot be applied for females with a full mutation owing to the presence of a nonmutated, methylated, inactive X chromosome.

10.2 Protein Test

Recently, alternative immunocytochemical tests have been described that are based on the direct detection of FMRP by a monoclonal antibody and the indirect visualization of antibody–antigen complexes

by alkaline phosphatase enzyme activity. The antibody test has the advantage of having the test result within 24 h, whereas routine DNA analysis by Southern blot generally takes more than 7 days.

A rapid method based on antibody detection of the FMR1 protein in cells of blood smears has been described and validated by Willemsen et al. FMRP in the cytoplasm of cells from unaffected individuals could be detected, whereas the cytoplasm of cells from affected males were not labeled for FMRP. For males, the diagnostic power of the test is high, because there is no overlap between the values of unaffected individuals and affected males with the full mutation. The diagnostic power of the test in identifying females with the full mutation is less specific compared to male patients. The X-inactivation process (lyonization) in cells from females with the full mutation may result in FMRP expression in a high percentage of their lymphocytes (up to 80%).

More recently, the protein test has been adapted for hair roots (Willemsen 1999), avoiding the need for blood drawing and allowing the analysis of the hair bulb cells that are of ectodermal origin and possibly less subject to selection effects in mosaic individuals than blood cells (of mesodermal origin). The diagnostic power of this new noninvasive test is high and it can be readily used to identify male patients. In addition, it was shown that the hair root test has the potential to serve as a prognostic indicator of the mental status in females with the full mutation.

11 Therapeutic Options

To study whether the fragile X syndrome is a potentially treatable disorder, rescue mice have been generated. Rescue mice are knockout mice in which an *FMR1* gene is introduced in the germ line by transgenesis. Introduction of FMRP as a cDNA under a CMV promoter in the Fmr1 knockout mouse resulted in expression of FMRP in the brain at roughly 50% of the control level. However, neither the phenotype, as measured by testicular weight, nor the cognitive capacities, as measured in the Morris water-maze, was rescued. Immunohistochemistry of the brain of the rescue mice showed FMRP expression in neurons, but also in nonneuronal cells, perhaps as a result of the CMV promoter used to drive FMR1 transcription.

Subsequently, Peier et al. (2000) created YAC transgenics, containing the whole human FMR1 with a large amount of flanking sequences, as a rescue model. Since the *FMR1* gene in the YAC is under the control of its own promoter, the gene is expressed in tissues similar to control mice. However, the overall level of FMRP seemed higher than that in control littermates. The testicular weight of the rescue mouse was restored, but the results of the behavioral test were controversial. In some tests, the rescue mice seemed "overcorrected"; for example, in the open field test, knockouts were significantly more active than controls, but the rescue mice were significantly less active than controls.

The generation of these two rescue mice suggests that concentration of FMRP as well as cell specificity and timing of expression are very important.

11.1 CGG Methylation and Reactivation

Rare individuals with expanded but unmethylated repeats have been described. Such individuals are generally unaffected

with the fragile X syndrome, demonstrating that methylation and not repeat elongation *per se* causes the typical features of the syndrome. Addition of the demethylating agent 5-azadeoxycytidine to cultured lymphoblasts of fragile X patients did enable the production of FMRP by these cells. However, the percentage of reactivated cells was disappointingly low. As methylation in the cell is followed by deacetylation, it was hypothesized that a combination of demethylating and acetylating drugs might enhance demethylation of the CGG repeat. Indeed, a synergistic effect of the two types of drugs was demonstrated to enhance the reactivation of the FMRP production approximately fivefold.

Though reactivation of protein expression *in vivo* is a promising result, it should be realized that demethylating agents are highly toxic, and as long as no drugs are developed that specifically target the FMR1-CGG repeat and that are less toxic, the use of demethylating and acetylating agents in patients remains elusive.

11.2 AMPAkine Trial in Relationship with LTD Abnormality

Synaptic plasticity is believed to be involved in learning and memory, and it is therefore tempting to speculate that the enhanced LTD observed in fragile X mice (see Chapter 6.2.7) is potentially correlated to the cognitive deficits observed in the fragile X syndrome. LTD is directly dependent on the AMPA receptor. The availability of a group of drugs called AMPAkines that directly targets the AMPA receptor prompted a trial with one of these agents called CX516 in a small group of fragile X patients. So far, no negative side effects have been reported, but the results on ameliorating the symptoms of the fragile X syndrome are awaited.

12 Epilogue

More than a decade of research following the discovery of the causative gene has revealed a wealth of information regarding the possible role of the lack of the *FMR1* gene in the abnormalities of the fragile X syndrome. Functional domains identified in the fragile X protein suggested a role in RNA transport through the cell. Subsequent experimental evidence revealed that FMRP transports specific mRNAs as an mRNP particle along the dendrites to the synapses in neurons and plays a role in translation after binding to polyribosomes. A large amount of possible mRNAs bound by FMRP whose translation may be regulated by FMRP has been determined, including a specific group of mRNAs that contain a planar G-quartet structure. A subset of the bound mRNAs identified by different research groups plays an important role in synapse maturation and plasticity. Though certain neurotransmitters, hormones or stress, are able to influence the FMRP levels, the relevant *in vivo* stimuli, if any, are not known at present. In addition, the effect of the variable levels of FMRP on individual bound mRNAs is not known. However, in general FMRP seems to inhibit translation. Further functional studies have to determine the importance of each individual gene whose translation is influenced by FMRP levels in fragile X syndrome. Thus, misregulation or mistrafficking of a subset of mRNAs as a consequence of the absence of FMRP could lead to impaired synaptic maturation and plasticity and lie at the basis of

the mental retardation and associated abnormalities in fragile X syndrome. Further elucidation of the role of FMRP in patients and animal models may help us understand why absence of the fragile X protein leads to the pathogenesis of the syndrome.

See also Brain Development; Genetics, Molecular Basis of; Medicinal Chemistry.

Bibliography

Books and Reviews

Bardoni, B., Mandel, J.L. (2002) Advances in understanding of fragile X pathogenesis and FMRP function, and in identification of X linked mental retardation genes, *Curr. Opin. Genet. Dev.* **12**, 284–293.

Chiurazzi, P., Oostra, B.A. (2000) Genetics of mental retardation, *Curr. Opin. Pediatr.* **12**, 529–535.

Gantois, I., Kooy, R.F. (2002) Targeting fragile X, *Genome Biol.* **3**, 1014.1–1014.5.

Hagerman, R.J., Hagerman, P.J. (Eds.) *Fragile X Syndrome: Diagnosis, Treatment, and Research*, Johns Hopkins University Press, Baltimore, MA, pp. 3–109.

Hagerman, R.J., Hagerman, P.J. (2001) Fragile X syndrome: a model of gene-brain-behavior relationships, *Mol. Genet. Metab.* **74**, 89–97.

Jin, P., Warren, S.T. (2003) New insights into fragile X syndrome: from molecules to neurobehaviors, *Trends Biochem. Sci.* **28**, 152–158.

Kooy, R.F. (2003) Of mice and the fragile X syndrome, *Trends Genet.* **19**, 148–154.

Kooy, R.F., Willemsen, R., Oostra, B.A. (2000) Fragile X syndrome at the turn of the century, *Mol. Med. Today* **6**, 194–199.

O'Donnell, W.T., Warren, S.T. (2002) A decade of molecular studies of fragile X syndrome, *Annu. Rev. Neurosci.* **25**, 315–338.

Oostra, B.A., Chiurazzi, P. (2001) The fragile X gene and its function, *Clin. Genet.* **60**, 399–408.

Willemsen, R., Oostra, B.A. (2000) FMRP detection assay for the diagnosis of the fragile X syndrome, *Am. J. Med. Genet.* **97**, 183–188.

Primary Literature

Abitbol, M., Menini, C., Delezoide, A.-L., Rhyner, T., Vekemans, M., Mallet, J. (1993) Nucleus basalis magnocellularis and hippocampus are the major sites of FMR-1 expression in the human fetal brain, *Nat. Genet.* **4**, 147–153.

Adinolfi, S., Bagni, C., Musco, G., Gibson, T., Mazzarella, L., Pastore, A. (1999) Dissecting FMR1, the protein responsible for fragile X syndrome, in its structural and functional domains, *RNA* **5**, 1248–1258.

Ashley, C.T., Sutcliffe, J.S., Kunst, C.B., Leiner, H.A., Eichler, E.E., Nelson, D.L., Warren, S.T. (1993a) Human and murine FMR1: alternative splicing and translational initiation downstream of the GGG repeat, *Nat. Genet.* **4**, 244–251.

Ashley, C.T., Wilkinson, K.D., Reines, D., Warren, S.T. (1993b) FMR-1 protein: conserved RNP family domains and selective RNA binding, *Science* **262**, 563–566.

Bächner, D., Mance, A., Steinbach, P., Wöhrle, D., Just, W., Vogel, W., Hameister, H., Poustka, A. (1993) Enhanced expression of the murine FMR1 gene during germ cell proliferation suggests a special function in both the male and the female gonad, *Hum. Mol. Genet.* **2**, 2043–2050.

Bakker, C.E., Kooy, R.F., D'Hooge, R., Tamanini, F., Willemsen, R., Nieuwenhuizen, I., de Vries, B.B.A., Reyniers, E., Hoogeveen, A.T., Willems, P.J., De Deyn, P.P., Oostra, B.A. (2000) Introduction of a *FMR1* transgene in the fragile X knockout mouse, *Neurosci. Res. Commun.* **26**, 265–277.

Bakker, C.E., Verheij, C., Willemsen, R., van der Helm, R., Oerlemans, F., Vermey, M., Bygrave, A., Hoogeveen, A.T., Oostra, B.A., Reyniers, E., De Boulle, K., D'Hooge, R., Cras, P., van Velzen, D., Nagels, G., Martin, J.-J., De Deyn, P.P., Darby, J.K., Willems, P.J. (1994) Fmr1 knockout mice: a model to study fragile X mental retardation, *Cell* **78**, 23–33.

Bardoni, B., Giglio, S., Schenck, A., Rocchi, M., Mandel, J.L. (2000) Assignment of NUFIP1 (nuclear FMRP interacting protein 1) gene to chromosome 13q14 and assignment of a pseudogene to chromosome 6q12, *Cytogenet. Cell. Genet.* **89**, 11–13.

Bardoni, B., Schenck, A., Mandel, J.L. (1999) A novel RNA-binding nuclear protein that

interacts with the fragile X mental retardation (FMR1) protein, *Hum. Mol. Genet.* **8**, 2557–2566.

Bardoni, B., Sittler, A., Shen, Y., Mandel, J.L. (1997) Analysis of domains affecting intracellular localization of the FMRP protein, *Neurobiol. Dis.* **4**, 329–336.

Bat, O., Kimmel, M., Axelrod, D.E. (1997) Computer simulation of expansions of DNA triplet repeats in the fragile X syndrome and Huntington's disease, *J. Theor. Biol.* **188**, 53–67.

Bliss, T.V.P., Collingridge, G.L. (1993) A synaptic model of memory: long-term potentiation in the hippocampus, *Nature* **361**, 31–39.

Bontekoe, C.J.M., Bakker, C.E., Nieuwenhuizen, I.M., van der Lans, H., de Lange, D., Hirst, M.C., Oostra, B.A. (2001) Instability of a $(CGG)_{98}$ repeat in the *Fmr1* promoter, *Hum. Mol. Genet.* **10**, 1693–1699.

Bontekoe, C.J.M., McIlwain, K.L., Nieuwenhuizen, I.M., Yuva-Paylor, L.A., Nellis, A., Willemsen, R., Fang, Z., Kirkpatrick, L., Bakker, C.E., McAninch, R., Cheng, N.C., Merriweather, M., Hoogeveen, A.T., Nelson, D., Paylor, R., Oostra, B.A. (2002) Knockout mouse model for Fxr2: a model for mental retardation, *Hum. Mol. Genet.* **11**, 487–498.

Brown, W.T., Houck, G.E., Jeziorowska, A., Levinson, F.N., Ding, X., Dobkin, C., Zhong, N., Henderson, J., Brooks, S.S., Jenkins, E.C. (1993) Rapid fragile X carrier screening and prenatal diagnosis using a nonradioactive PCR test, *JAMA* **270**, 1569–1575.

Brown, V., Jin, P., Ceman, S., Darnell, J., O'Donnell, W.T., Tenenbaum, S.A., Jin, X., Feng, Y., Wilkinson, K.D., Keene, J.D., Darnell, R.B., Warren, S.T. (2001) Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome, *Cell* **107**, 477–487.

Brown, V., Small, K., Lakkis, L., Feng, Y., Gunter, C., Wilkinson, K.D., Warren, S.T. (1998) Purified recombinant Fmrp exhibits selective RNA binding as an intrinsic property of the fragile X mental retardation protein, *J. Biol. Chem.* **273**, 15521–15527.

Brunberg, J.A., Jacquemont, S., Hagerman, R.J., Berry, K., Grigsby, J., Leehey, M.A., Tassone, F., Brown, W.T., Greco, C.M., Hagerman, P.J. (2002) Fragile X premutation carriers: characteristic MR imaging findings of adult male patients with progressive cerebellar and cognitive dysfunction, *AJNR Am. J. Neuroradiol.* **23**, 1757–1766.

Castren, M., Lampinen, K.E., Miettinen, R., Koponen, E., Sipola, I., Bakker, C.E., Oostra, B.A., Castren, E. (2002) BDNF regulates the expression of fragile X mental retardation protein mRNA in the hippocampus, *Neurobiol. Dis.* **11**, 221–229.

Ceman, S., Brown, V., Warren, S.T. (1999) Isolation of an FMRP-associated messenger ribonucleoprotein particle and identification of nucleolin and the fragile X-related proteins as components of the complex, *Mol. Cell. Biol.* **19**, 7925–7932.

Ceman, S., Nelson, R., Warren, S.T. (2000) Identification of mouse YB1/p50 as a component of the FMRP-associated mRNP particle, *Biochem. Biophys. Res. Commun.* **279**, 904–908.

Chen, L., Toth, M. (2001) Fragile X mice develop sensory hyperreactivity to auditory stimuli, *Neuroscience* **103**, 1043–1050.

Chiang, P.W., Carpenter, L.E., Hagerman, P.J. (2001) The 5′-untranslated region of the FMR1 message facilitates translation by internal ribosome entry, *J. Biol. Chem.* **276**, 37916–37921.

Chiurazzi, P., MacPherson, J., Sherman, S., Neri, G. (1996) Significance of linkage disequilibrium between the fragile X locus and its flanking markers, *Am. J. Med. Genet.* **64**, 203–208.

Chiurazzi, P., Pomponi, M.G., Pietrobono, R., Bakker, C.E., Neri, G., Oostra, B.A. (1999) Synergistic effect of histone hyperacetylation and DNA demethylation in the reactivation of the *FMR1* gene, *Hum. Mol. Genet.* **8**, 2317–2323.

Chiurazzi, P., Pomponi, M.G., Willemsen, R., Oostra, B.A., Neri, G. (1998) In vitro reactivation of the FMR1 gene involved in fragile X syndrome, *Hum. Mol. Genet.* **7**, 109–113.

Comery, T.A., Harris, J.B., Willems, P.J., Oostra, B.A., Irwin, S.A., Weiler, I.J., Greenough, W.T. (1997) Abnormal dendritic spines in fragile X knockout mice: maturation and pruning deficits, *Proc. Natl. Acad. Sci. USA* **94**, 5401–5404.

Corbin, F., Bouillon, M., Fortin, A., Morin, S., Rousseau, F., Khandjian, E.W. (1997) The fragile X mental retardation protein is associated with poly(A)+ mRNA in actively

translating polyribosomes, *Hum. Mol. Genet.* **6**, 1465–1472.

Coy, J.F., Sedlacek, Z., Bächner, D., Hameister, H., Joos, S., Lichter, P., Delius, H., Poustka, A. (1995) Highly conserved 3′ UTR and expression pattern of FXR1 points to a divergent gene regulation of FXR1 and FMR1, *Hum. Mol. Genet.* **4**, 2209–2218.

Crawford, D.C., Zhang, F., Wilson, B., Warren, S.T., Sherman, S.L. (2000) Fragile X CGG repeat structures among African-Americans: identification of a novel factor responsible for repeat instability, *Hum. Mol. Genet.* **9**, 1759–1769.

D'Hooge, R., Nagels, G., Franck, F., Bakker, C.E., Reyniers, E., Storm, K., Kooy, R.F., Oostra, B.A., Willems, P.J., De Deyn, P.P. (1997) Mildly impaired water maze performance in male *Fmr1* knockout mice, *Neuroscience* **76**, 367–376.

Darlow, J.M., Leach, D.R. (1998) Evidence for two preferred hairpin folding patterns in d(CGG).d(CCG) repeat tracts in vivo, *J. Mol. Biol.* **275**, 17–23.

Darnell, J., Jensen, K., Jin, P., Brown, V., Warren, S.T., Darnell, R.B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function, *Cell* **107**, 489–499.

Das, S., Kubota, T., Song, M., Daniel, R., Berry, K., Prior, T.W., Popovich, B., Rosser, L., Arinami, T., Ledbetter, D.H. (1997) Methylation analysis of the fragile X syndrome by PCR, *Genet. Test.* **1**, 151–155.

De Boulle, K., Verkerk, A.J.M.H., Reyniers, E., Vits, L., Hendrickx, J., van Roy, B., van den Bos, F., de Graaff, E., Oostra, B.A., Willems, P.J. (1993) A point mutation in the FMR1 gene associated with fragile X mental retardation, *Nat. Genet.* **3**, 31–35.

De Diego Otero, Y., Severijnen, L.A., van Cappellen, G., Schrier, M., Oostra, B., Willemsen, R. (2002) Transport of fragile X mental retardation protein via granules in neurites of PC12 cells, *Mol. Cell. Biol.* **22**, 8332–8341.

de Graaff, E., Rouillard, P., Willems, P.J., Smits, A.P.T., Rousseau, F., Oostra, B.A. (1995) Hotspot for deletions in the CGG repeat region of *FMR1* in fragile X patients, *Hum. Mol. Genet.* **4**, 45–49.

de Vries, B.B., Fryns, J.-P., Butler, M.G., Canziani, F., Wesby-van Swaay, E., van Hemel, J.O., Oostra, B.A., Halley, D.J.J., Niermeijer, M.F. (1993) Clinical and molecular studies in fragile X patients with a Prader-Willi-like phenotype, *J. Med. Genet.* **30**, 761–766.

de Vries, B.B.A., van den Ouweland, A.M.W., Mohkamsing, S., Duivenvoorden, H.J., Mol, E., van Gelsema, K.R., Halley, D.J.J., Sandkuijl, L.A., Oostra, B.A., Tibben, A., Niermeijer, M.F. (1997) Screening and diagnosis for the fragile X syndrome among the mentally retarded: an epidemiological and psychological survey, *Am. J. Hum. Genet.* **61**, 660–667.

de Vries, B.B.A., Wiegers, A.M., Smits, A.P.T., Mohkamsing, S., Duivenvoorden, H.J., Fryns, J.-P., Curfs, L.M.G., Halley, D.J.J., Oostra, B.A., van den Ouweland, A.M.W., Niermeijer, M.F. (1996) Mental status of females with an FMR1 gene full mutation, *Am. J. Hum. Genet.* **58**, 1025–1032.

Deelen, W., Bakker, C., Halley, D.J., Oostra, B.A. (1994) Conservation of CGG region in FMR1 gene in mammals, *Am. J. Med. Genet.* **51**, 513–516.

Devys, D., Lutz, Y., Rouyer, N., Bellocq, J.-P., Mandel, J.-L. (1993) The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation, *Nat. Genet.* **4**, 335–340.

Dobkin, C., Rabe, A., Dumas, R., El Idrissi, A., Haubenstock, H., Brown, W.T. (2000) Fmr1 knockout mouse has a distinctive strain-specific learning impairment, *Neuroscience* **100**, 423–429.

Dockendorff, T.C., Su, H.S., McBride, S.M., Yang, Z., Choi, C.H., Siwicki, K.K., Sehgal, A., Jongens, T.A. (2002) Drosophila lacking dfmr1 activity show defects in circadian output and fail to maintain courtship interest, *Neuron* **34**, 973–984.

Dombrowski, C., Levesque, S., Morel, M.L., Rouillard, P., Morgan, K., Rousseau, F. (2002) Premutation and intermediate-size FMR1 alleles in 10572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles, *Hum. Mol. Genet.* **11**, 371–378.

Drouin, R., Angers, M., Dallaire, N., Rose, T.M., Khandjian, W., Rousseau, F. (1997) Structural and functional characterization of the human FMR1 promoter reveals similarities with the hnRNP-A2 promoter region, *Hum. Mol. Genet.* **6**, 2051–2060.

Eberhart, D.E., Malter, H.E., Feng, Y., Warren, S.T. (1996) The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export signals, *Hum. Mol. Genet.* **5**, 1083–1091.

Eichler, E.E., Holden, J.J.A., Popovich, B.W., Reiss, A.L., Snow, K., Thibodeau, S.N., Richards, C.S., Ward, P.A., Nelson, D.L. (1995a) Length of uninterrupted CGG repeats determines instability in the *FMR1* gene, *Nat. Genet.* **8**, 88–94.

Eichler, E.E., Kunst, C.B., Lugenbeel, K.A., Ryder, O.A., Davison, D., Warren, S.T., Nelson, D.L. (1995b) Evolution of the cryptic *FMR1* CGG repeat, *Nat. Genet.* **11**, 301–307.

Eichler, E.E., Nelson, D.L. (1996) Genetic variation and evolutionary stability of the FMR1 CGG repeat in six closed human populations, *Am. J. Med. Genet.* **64**, 220–225.

Eichler, E.E., Richards, S., Gibbs, R.A., Nelson, D.L. (1993) Fine structure of the human FMR1 gene, *Hum. Mol. Genet.* **2**, 1147–1153.

Faust, C.J., Verkerk, A.J., Wilson, P.J., Morris, C.P., Hopwood, J.J., Oostra, B.A., Herman, G.E. (1992) Genetic mapping on the mouse X chromosome of human cDNA clones for the fragile X and Hunter syndromes, *Genomics* **12**, 814–817.

Feng, Y., Absher, D., Eberhart, D.E., Brown, V., Malter, H.E., Warren, S.T. (1997) FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe fragile X syndrome abolishes this association, *Mol. Cell* **1**, 109–118.

Feng, Y., Gutekunst, C.-A., Eberhart, D.E., Yi, H., Warren, S.T., Hersch, S.M. (1997) Fragile X mental retardation protein: nucleocytoplasmic shuttling and association with somatodendritic ribosomes, *J. Neurosci.* **17**, 1539–1547.

Feng, Y., Lakkis, L., Devys, D., Warren, S.T. (1995a) Quantitative comparison of FMR1 gene expression in normal and premutation alleles, *Am. J. Hum. Genet.* **56**, 106–113.

Feng, Y., Zhang, F., Lokey, L.K., Chastain, J.L., Lakkis, L., Eberhart, D., Warren, S.T. (1995b) Translational suppression by trinucleotide repeat expansion at FMR1, *Science* **268**, 731–734.

Fiala, J.C., Spacek, J., Harris, K.M. (2002) Dendritic spine pathology: cause or consequence of neurological disorders? *Brain. Res. Rev.* **39**, 29–54.

Fridell, R.A., Benson, R.E., Hua, J., Bogerd, H.P., Cullen, B.R. (1996) A nuclear role for the fragile X mental retardation protein, *EMBO J.* **15**, 5408–5414.

Fu, Y.-H., Kuhl, D.P.A., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J.M.H., Holden, J.J.A., Fenwick, R.G., Warren, S.T., Oostra, B.A., Nelson, D.L., Caskey, C.T. (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox, *Cell* **67**, 1047–1058.

Gantois, I., Bakker, C.E., Reyniers, E., Willemsen, R., D'Hooge, R., De Deyn, P.P., Oostra, B.A., Kooy, R.F. (2001) Restoring the phenotype of fragile X syndrome: insight from the mouse model, *Curr. Mol. Med.* **1**, 447–455.

Genc, B., Muller-Hartmann, H., Zeschnigk, M., Deissler, H., Schmitz, B., Majewski, F., Gontard, A., Doerfler, W. (2000) Methylation mosaicism of 5′-(CGG)(n)-3′ repeats in fragile X, premutation and normal individuals, *Nucleic Acids Res.* **28**, 2141–2152.

Godfraind, J.-M., Reyniers, E., De Boulle, K., D'Hooge, R., De Deyn, P.P., Bakker, C.E., Oostra, B.A., Kooy, R.F., Willems, P.J. (1996) Long-term potentiation in the hippocampus of fragile X knockout mice, *Am. J. Med. Genet.* **64**, 246–251.

Gold, B., Radu, D., Balanko, A., Chiang, C.S. (2000) Diagnosis of fragile X syndrome by Southern blot hybridization using a chemiluminescent probe: a laboratory protocol, *Mol. Diagn.* **5**, 169–178.

Gould, E.L., Loesch, D.Z., Martin, M.J., Hagerman, R.J., Armstrong, S.M., Huggins, R.M. (2000) Melatonin profiles and sleep characteristics in boys with fragile X syndrome: a preliminary study, *Am. J. Med. Genet.* **95**, 307–315.

Greco, C.M., Hagerman, R.J., Tassone, F., Chudley, A.E., Del, B., Jacquemont, S., Leehey, M., Hagerman, P.J. (2002) Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers, *Brain* **125**, 1760–1771.

Greenough, W.T., Klintsova, A.Y., Irwin, S.A., Galvez, R., Bates, K.E., Weiler, I.J. (2001) Synaptic regulation of protein synthesis and the fragile X protein, *Proc. Natl. Acad. Sci. USA* **98**, 7101–7106.

Gronskov, K., Hallberg, A., Brondum, N. (1998) Mutational analysis of the FMR1 gene in 118 mentally retarded males suspected of fragile

X syndrome: absence of prevalent mutations, *Hum. Genet.* **102**, 440–445.

Gunter, C., Paradee, W., Crawford, D.C., Meadows, K.A., Newman, J., Kunst, C.B., Nelson, D.L., Schwartz, C., Murray, A., Macpherson, J.N., Sherman, S.L., Warren, S.T. (1998) Reexamination of factors associated with expansion of CGG repeats using a single nucleotide polymorphism in FMR1, *Hum. Mol. Genet.* **7**, 1935–1946.

Hamel, B.C.J., Smits, A.P.T., de Graaff, E., Smeets, D.F.C.M., Schoute, F., Eussen, B.H.J., Knight, S.J.L., Davies, K.E., Assman-Hulsmans, C.F.C.H., Oostra, B.A. (1994) Segregation of FRAXE in a large family: clinical, psychometric, cytogenetic, and molecular data, *Am. J. Hum. Genet.* **55**, 923–931.

Hansen, R.S., Gartler, S.M., Scott, C.R., Chen, S.-H., Laird, C.D. (1992) Methylation analysis of CGG sites in the CpG island of the human FMR1 gene, *Hum. Mol. Genet.* **1**, 571–578.

Heitz, D., Devys, D., Imbert, G., Kretz, C., Mandel, J.L. (1992) Inheritance of the fragile X syndrome: size of the fragile X premutation is a major determinant of the transition to full mutation, *J. Med. Genet.* **29**, 794–801.

Hinds, H.L., Ashley, C.T., Sutcliffe, J.S., Nelson, D.L., Warren, S.T., Housman, D.E., Schalling, M. (1993) Tissue specific expression of FMR-1 provides evidence for a functional role in fragile X syndrome, *Nat. Genet.* **3**, 36–43.

Hinton, V.J., Brown, W.T., Wisniewski, K., Rudelli, R.D. (1991) Analysis of neocortex in three males with the fragile X syndrome, *Am. J. Med. Genet.* **41**, 289–294.

Hirst, M.C., White, P.J. (1998) Cloned human FMR1 trinucleotide repeats exhibit a length- and orientation-dependent instability suggestive of in vivo lagging strand secondary structure, *Nucleic Acids Res.* **26**, 2353–2358.

Holden, J.J.A., Percy, M., Allingham-Hawkins, D., Brown, W.T., Chiurazzi, P., Fisch, G., Gane, L., Gunter, C., Hagerman, R., Jenkins, E.C., Kooy, R.F., Lubs, H.A., Murray, A., Neri, G., Schwartz, C., Tranebjaerg, L., Villard, L., Willems, P.J. (1999) Eighth international workshop on the fragile X syndrome and X-linked mental retardation, August 16–22, 1997, *Am. J. Med. Genet.* **83**, 221–236.

Huber, K.M., Gallagher, S.M., Warren, S.T., Bear, M.F. (2002) Altered synaptic plasticity in a mouse model of fragile X mental retardation, *Proc. Natl. Acad. Sci. USA* **99**, 7746–7750.

Hull, C., Hagerman, R.J. (1993) A study of the physical, behavioral, and medical phenotype, including anthropometric measures, of females with fragile X syndrome, *Am. J. Dis. Child.* **147**, 1236–1241.

Hundscheid, R.D., Sistermans, E.A., Thomas, C.M., Braat, D.D., Straatman, H., Kiemeney, L.A., Oostra, B.A., Smits, A.P. (2000) Imprinting effect in premature ovarian failure confined to paternally inherited fragile X premutations, *Am. J. Hum. Genet.* **66**, 413–418.

Huot, M.E., Mazroui, R., Leclerc, P., Khandjian, E.W. (2001) Developmental expression of the fragile X-related 1 proteins in mouse testis: association with microtubule elements, *Hum. Mol. Genet.* **10**, 2803–2811.

Inoue, S., Shimoda, M., Nishinokubi, I., Siomi, M.C., Okamura, M., Nakamura, A., Kobayashi, S., Ishida, N., Siomi, H. (2002) A role for the Drosophila fragile X-related gene in circadian output, *Curr. Biol.* **12**, 1331–1335.

Irwin, S.A., Galvez, R., Greenough, W.T. (2000) Dendritic spine structural anomalies in fragile X mental retardation syndrome, *Cereb. Cortex* **10**, 1038–1044.

Irwin, S.A., Patel, B., Idupulapati, M., Harris, J.B., Crisostomo, R.A., Larsen, B.P., Kooy, F., Willems, P.J., Cras, P., Kozlowski, P.B., Swain, R.A., Weiler, I.J., Greenough, W.T. (2001) Abnormal dendritic spine characteristics in the temporal and visual cortices of patients with fragile X syndrome: a quantitative examination, *Am. J. Med. Genet.* **98**, 161–167.

Irwin, S.A., Swain, R.A., Christmon, C.A., Chakravarti, A., Weiler, I.J., Greenough, W.T. (2000) Evidence for altered Fragile-X mental retardation protein expression in response to behavioral stimulation, *Neurobiol. Learn. Mem.* **73**, 87–93.

Ishizuka, A., Siomi, M.C., Siomi, H. (2002) A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins, *Genes. Dev.* **16**, 2497–2508.

Ivanco, T.L., Greenough, W.T. (2002) Altered mossy fiber distributions in adult Fmr1 (FVB) knockout mice, *Hippocampus* **12**, 47–54.

Kang, S., Ohshima, K., Shimizu, M., Amirhaeri, S., Wells, R.D. (1995) Pausing of DNA synthesis in vitro at specific loci in CTG and CGG triplet repeats from human

hereditary disease genes, *J. Biol. Chem.* **270**, 27014–27021.

Kates, W.R., Abrams, M.T., Kaufmann, W.E., Breiter, S.N., Reiss, A.L. (1997) Reliability and validity of MRI measurement of the amygdala and hippocampus in children with fragile X syndrome, *Psychiatry. Res.* **75**, 31–48.

Kenneson, A., Zhang, F., Hagedorn, C.H., Warren, S.T. (2001) Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers, *Hum. Mol. Genet.* **10**, 1449–1454.

Khandjian, E.W., Bardoni, B., Corbin, F., Sittler, A., Giroux, S., Heitz, D., Tremblay, S., Pinset, C., Montarras, D., Rousseau, F., Mandel, J. (1998) Novel isoforms of the fragile X related protein FXR1P are expressed during myogenesis, *Hum. Mol. Genet.* **7**, 2121–2128.

Khandjian, E.W., Corbin, F., Woerly, S., Rousseau, F. (1996) The fragile X mental retardation protein is associated with ribosomes, *Nat. Genet.* **12**, 91–93.

Kiledjian, M., Dreyfuss, G. (1992) Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box, *EMBO. J.* **11**, 2655–2664.

Koch, M. (1999) The neurobiology of startle, *Prog. Neurobiol.* **59**, 107–128.

Kolehmainen, K. (1994) Population genetics of fragile X: a multiple allele model with variable risk of CGG repeat expansion, *Am. J. Med. Genet.* **51**, 428–435.

Kooy, R.F., D'Hooge, R., Reyniers, E., Bakker, C.E., Nagels, G., De Boulle, K., Storm, K., Clincke, G., De Deyn, P.P., Oostra, B.A., Willems, P.J. (1996) Transgenic mouse model for the fragile X syndrome, *Am. J. Med. Genet.* **64**, 241–245.

Kooy, R.F., Oostra, B.A., Willems, P.J. (1997) Molecular Detection of Dynamic Mutations, in: Adolph, K.W. (Ed.) *Human Genome Methods*, CRC Press, Boca Raton, FL, pp. 23–53.

Kooy, R.F., Reyniers, E., Verhoye, M., Sijbers, J., Bakker, C.E., Oostra, B.A., Willems, P.J., Van der Linden, A. (1999) Neuroanatomy of the fragile X knockout mouse brain studied using *in vivo* high resolution magnetic resonance imaging (MRI), *Eur. J. Hum. Genet.* **7**, 526–532.

Kovtun, I.V., Goellner, G., McMurray, C.T. (2001) Structural features of trinucleotide repeats associated with DNA expansion, *Biochem. Cell. Biol.* **79**, 325–336.

Kunst, C.B., Warren, S.T. (1994) Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles, *Cell* **77**, 853–861.

Laggerbauer, B., Ostareck, D., Keidel, E.M., Ostareck-Lederer, A., Fischer, U. (2001) Evidence that fragile X mental retardation protein is a negative regulator of translation, *Hum. Mol. Genet.* **10**, 329–338.

Li, J., Pelletier, M.R., Perez Velazquez, J.-L., Carlen, P.L. (2002) Reduced cortical synaptic plasticity and GluR1 expression associated with fragile X mental retardation protein deficiency, *Mol. Cell. Neurosci.* **19**, 138–151.

Li, Z., Zhang, Y., Ku, L., Wilkinson, K.D., Warren, S.T., Feng, Y. (2001) The fragile X mental retardation protein inhibits translation via interacting with mRNA, *Nucleic Acids Res.* **29**, 2276–2283.

Loesch, D.Z., Hay, D.A., Mulley, J. (1994) Transmitting males and carrier females in fragile X–revisited, *Am. J. Med. Genet.* **51**, 392–399.

Loesch, D.Z., Huggins, R., Petrovic, V., Slater, H. (1995) Expansion of the CGG repeat in fragile X in the FMR1 gene depends on the sex of the offspring, *Am. J. Hum. Genet.* **57**, 1408–1413.

Lubs, H.A. (1969) A marker X chromosome, *Am. J. Hum. Genet.* **21**, 231–244.

Lugenbeel, K.A., Peier, A.M., Carson, N.L., Chudley, A.E., Nelson, D.L. (1995) Intragenic loss of function mutations demonstrate the primary role of FMR1 in fragile X syndrome, *Nat. Genet.* **10**, 483–485.

Malter, H.E., Iber, J.C., Willemsen, R., de Graaff, E., Tarleton, J.C., Leisti, J., Warren, S.T., Oostra, B.A. (1997) Characterization of the full fragile X syndrome mutation in fetal gametes, *Nat. Genet.* **15**, 165–169.

Manley, K., Pugh, J., Messer, A. (1999) Instability of the CAG repeat in immortalized fibroblast cell cultures from Huntington's disease transgenic mice, *Brain. Res.* **835**, 74–79.

Martin, J.P., Bell, J. (1943) A pedigree of mental defect showing sex-linkage, *J. Neurol. Psychol.* **6**, 154–157.

Mazroui, R., Huot, M.E., Tremblay, S., Filion, C., Labelle, Y., Khandjian, E.W. (2002) Trapping of messenger RNA by fragile X mental retardation protein into cytoplasmic granules induces translation repression, *Hum. Mol. Genet.* **11**, 3007–3017.

Mineur, Y.S., Crusio, W.E. (2002) Behavioral and neuroanatomical characterization of FVB/N inbred mice, *Brain. Res. Bull.* **57**, 41–47.

Miyashiro, K.Y., Beckel, M., Purk, T.P., Becker, K.G., Barret, T., Liu, L., Carbonetto, S., Weiler, I.J., Greenough, W.T., Eberwine, J. (2003) RNA cargoes associating with FMRP reveal deficits in cellular functioning in Fmr1 null mice, *Neuron* **37**, 417–431.

Morales, J., Hiesinger, P.R., Schroeder, A.J., Kume, K., Verstreken, P., Jackson, F.R., Nelson, D.L., Hassan, B.A. (2002) Drosophila fragile X protein, DFXR, regulates neuronal morphology and function in the brain, *Neuron* **34**, 961–972.

Morris, R.G.M., Garrud, P., Rawlins, J.N.P., O'Keefe, J. (1982) Place navigation impaired in rats with hippocampal lesions, *Nature* **297**, 681–683.

Morton, N.E., MacPherson, J.N. (1992) Population genetics of the fragile-X syndrome: multiallelic model for the FMR1 locus, *Proc. Natl. Acad. Sci. USA* **89**, 4215–4217.

Mostofsky, S., Mazzocco, M.M.M., Aakalu, G., Warsofsky, L.S., Denckla, M.B., Reiss, A.L. (1998) Decreased cerebellar posterior vermis size in fragile X syndrome. Correlation with neurocognitive performance, *Neurology* **50**, 121–130.

Moutou, C., Vincent, M.-C., Biancalana, V., Mandel, J.-L. (1997) Transition from premutation to full mutation in fragile X syndrome is likely to be prezygotic, *Hum. Mol. Genet.* **6**, 971–979.

Muller-Hartmann, H., Deissler, H., Naumann, F., Schmitz, B., Schroer, J., Doerfler, W. (2000) The human 20-kDa 5′-(CGG)(n)-3′-binding protein is targeted to the nucleus and affects the activity of the FMR1 promoter, *J. Biol. Chem.* **275**, 6447–6452.

Murray, A., Webb, J., Grimley, S., Conway, G., Jacobs, P. (1998) Studies of FRAXA and FRAXE in women with premature ovarian failure, *J. Med. Genet.* **35**, 637–640.

Musco, G., Stier, G., Joseph, C., Castiglione Morelli, M.A., Nilges, M., Gibson, T.J., Pastore, A. (1996) Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome, *Cell* **85**, 237–245.

Musumeci, S.A., Bosco, P., Calabrese, G., Bakker, C., De Sarro, G.B., Elia, M., Ferri, R., Oostra, B.A. (2000) Audiogenic seizures susceptibility in transgenic mice with fragile X syndrome, *Epilepsia* **41**, 19–23.

Musumeci, S.A., Hagerman, R.J., Ferri, R., Bosco, P., Dalla, B., Tassinari, C.A., De, S., Elia, M. (1999) Epilepsy and EEG findings in males with fragile X syndrome, *Epilepsia* **40**, 1092–1099.

Nielsen, D.M., Derber, W.J., McClellan, D.A., Crnic, L.S. (2002) Alterations in the auditory startle response in Fmr1 targeted mutant mouse models of fragile X syndrome, *Brain. Res.* **927**, 8–17.

Nimchinski, E.A., Oberlander, A.M., Svoboda, K. (2001) Abnormal development of dendritic spines in *FMR1* knock-out mice, *J. Neurosci.* **21**, 5139–5146.

Nolin, S.L., Brown, W.T., Glicksman, A., Houck, G.E. Jr., Gargano, A.D., Sullivan, A., Biancalana, V., Bröndum-Nielsen, K., Hjalgrim, H., Holinski-Feder, E., Kooy, F., Longshore, J., MacPherson, J., Mandel, J.-L., Matthijs, G., Rousseau, F., Steinbach, P., Väisänen, M.-L., von Koskull, H., Sherman, S.L. (2003) Expansion of the fragile X CGG repeat in females with premutation or intermediate alleles, *Am. J. Hum. Genet.* **72**, 454–464.

Nolin, S.L., Houck, G.E., Gargano, A.D., Blumstein, H., Dobkin, C.S., Brown, W.T. (1999) FMR1 CGG-repeat instability in single sperm and lymphocytes of fragile-X premutation males, *Am. J. Hum. Genet.* **65**, 680–688.

Oberlé, I., Boué, J., Croquette, M.F., Voelckel, M.A., Mattei, M.G., Mandel, J.L. (1992) Three families with high expression of a fragile site at Xq27.3, lack of anomalies at the FMR-1 CpG island, and no clear phenotypic association, *Am. J. Med. Genet.* **43**, 224–231.

Oberlé, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boué, J., Bertheas, M.F., Mandel, J.-L. (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome, *Science* **252**, 1097–1102.

Oostra, B.A., Jacky, P.B., Brown, W.T., Rousseau, F. (1993) Guidelines for the diagnosis of fragile X syndrome, *J. Med. Genet.* **30**, 410–413.

Panagopoulos, I., Lassen, C., Kristoffersson, U., Aman, P. (1999) A methylation PCR approach for detection of fragile X syndrome, *Hum. Mutat.* **14**, 71–79.

Paradee, W., Melikian, H.E., Rasmussen, D.E., Kenneson, A., Conn, P.J., Warren, S.T. (1999) Fragile X mouse: strain effects of knockout

phenotype and evidence suggesting deficient amygdala function, *Neuroscience* **94**, 185–192.

Partington, M.W. (1984) The fragile X syndrome II: preliminary data on growth and development in males, *Am. J. Med. Genet.* **17**, 175–194.

Peier, A.M., McIlwain, K.L., Kenneson, A., Warren, S.T., Paylor, R., Nelson, D.L. (2000) (Over) correction of FMR1 deficiency with YAC transgenics: behavioral and physical features, *Hum. Mol. Genet.* **9**, 1145–1159.

Pieretti, M., Zhang, F., Fu, Y.-H., Warren, S.T., Oostra, B.A., Caskey, C.T., Nelson, D.L. (1991) Absence of expression of the FMR-1 gene in fragile X syndrome, *Cell* **66**, 817–822.

Price, D.K., Zhang, F., Ashley, C.T. Jr., Warren, S.T. (1996) The chicken FMR1 gene is highly conserved with a CCT 5′-untranslated repeat and encodes an RNA-binding protein, *Genomics* **31**, 3–12.

Qin, M., Kang, J., Smith, C.B. (2002) Increased rates of cerebral glucose metabolism in a mouse model of fragile X mental retardation, *Proc. Natl. Acad. Sci. USA* **99**, 15758–15763.

Reiss, A.L., Abrams, M.T., Greenlaw, R., Freund, L., Denckla, M.B. (1995) Neurodevelopmental effects of the FMR1 full mutation in humans, *Nat. Med.* **1**, 159–167.

Reiss, A.L., Freund, L., Tseng, J.E., Joshi, P.K. (1991) Neuroanatomy in fragile X females: the posterior fossa, *Am. J. Hum. Genet.* **49**, 279–288.

Reiss, A.L., Lee, J., Freund, L. (1994) Neuroanatomy of fragile X syndrome: the temporal lobe, *Neurology* **44**, 1317–1324.

Reyniers, E., Martin, J.-J., Cras, P., Van Marck, E., Handig, I., Jorens, H.Z.J., Oostra, B.A., Kooy, R.F., Willems, P.J. (1999) Post-mortem examination of two fragile X brothers with an FMR1 full mutation, *Am. J. Med. Genet.* **84**, 245–249.

Reyniers, E., Vits, L., De Boulle, K., van Roy, B., van Velzen, D., de Graaff, E., Verkerk, A.J.M.H., Jorens, H.Z.J., Darby, J.K., Oostra, B., Willems, P.J. (1993) The full mutation in the FMR-1 gene of male fragile X patients is absent in their sperm, *Nat. Genet.* **4**, 143–146.

Richards, R.I., Holman, K., Friend, K., Staples, A., Sutherland, G.R., Oudet, C., Biancalana, V., Mandel, J.L. (1994) FRAXAC2 instability, *Nat. Genet.* **7**, 123.

Richards, B.W., Sylvester, P.E., Brooker, C. (1981) Fragile X-linked mental retardation: the Martin-Bell syndrome, *J. Ment. Defic. Res.* **25**(Pt 4), 253–256.

Riddle, J.E., Cheema, A., Sobesky, W.E., Gardner, S.C., Taylor, A.K., Pennington, B.F., Hagerman, R.J. (1998) Phenotypic involvement in females with the FMR1 gene mutation, *Am. J. Ment. Retard.* **102**, 590–601.

Rousseau, F., Heitz, D., Biancalana, V., Blumenfeld, S., Kretz, C., Boué, J., Tommerup, N., Van Der Hagen, C., DeLozier-Blanchet, C., Croquette, M.-F., Gilgenkrantz, S., Jalbert, P., Voelckel, M.-A., Oberlé, I., Mandel, J.-L. (1991) Direct diagnosis by DNA analysis of the fragile X syndrome of mental retardation, *N. Engl. J. Med.* **325**, 1674–1681.

Rousseau, F., Heitz, D., Tarleton, J., MacPherson, J., Malmgren, H., Dahl, N., Barnicoat, A., Mathew, C., Mornet, E., Tejada, I., Maddalena, A., Spiegel, R., Schinzel, A., Marcos, J.A.G., Schorderet, D.F., Schaap, T., MacCioni, L., Russo, S., Jacobs, P.A., Schwartz, C., Mandel, J.L. (1994) A multicenter study on genotype-phenotype correlations in the fragile X syndrome, using direct diagnosis with probe StB12.3: the first 2,253 cases, *Am. J. Hum. Genet.* **55**, 225–237.

Rousseau, F., Rouillard, P., Morel, M.L., Khandjian, E.W., Morgan, K. (1995) Prevalence of carriers of premutation-size alleles of the FMRI gene–and implications for the population genetics of the fragile X syndrome, *Am. J. Hum. Genet.* **57**, 1006–1018.

Samadashwily, G.M., Raca, G., Mirkin, S.M. (1997) Trinucleotide repeats affect DNA replication in vivo, *Nat. Genet.* **17**, 298–304.

Schaeffer, C., Bardoni, B., Mandel, J.L., Ehresmann, B., Ehresmann, C., Moine, H. (2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif, *EMBO J.* **20**, 4803–4813.

Schenck, A., Bardoni, B., Moro, A., Bagni, C., Mandel, J.L. (2001) A highly conserved protein family interacting with the fragile X mental retardation protein (FMRP) and displaying selective interactions with FMRP-related proteins FXR1P and FXR2P, *Proc. Natl. Acad. Sci. USA* **98**, 8844–8849.

Schlotterer, C., Tautz, D. (1992) Slippage synthesis of simple sequence DNA, *Nucleic Acids Res.* **20**, 211–215.

Schwemmle, S., de Graaff, E., Deissler, H., Gläser, D., Wöhrle, D., Kennerknecht, I.,

Just, W., Oostra, B.A., Drabkin, H., Vogel, W., Steinbach, P. (1997) Characterization of FMR1 promoter elements by in vivo-footprinting analysis, *Am. J. Hum. Genet.* **60**, 1354–1362.

Sherman, S.L., Jacobs, P.A., Morton, N.E., Froster-Iskenius, U., Howard-Peebles, P.N., Nielsen, K.B., Partington, N.W., Sutherland, G.R., Turner, G., Watson, M. (1985) Further segregation analysis of the fragile X syndrome with special reference to transmitting males, *Hum. Genet.* **69**, 3289–3299.

Sherman, S.L., Morton, N.E., Jacobs, P.A., Turner, G. (1984) The marker (X) chromosome: a cytogenetic and genetic analysis, *Ann. Hum. Genet.* **48**, 21–37.

Shimizu, M., Gellibolian, R., Oostra, B.A., Wells, R.D. (1996) Cloning, characterization and properties of plasmids containing CGG triplet repeats from the FMR-1 gene, *J. Mol. Biol.* **258**, 614–626.

Sinden, R.R. (2001) Neurodegenerative diseases. Origins of instability, *Nature* **411**, 757, 758.

Siomi, H., Choi, M., Siomi, M.C., Nussbaum, R.L., Dreyfuss, G. (1994) Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome, *Cell* **77**, 33–39.

Siomi, H., Dreyfuss, G. (1995) A nuclear localization domain in the hnRNP A1 protein, *J. Cell. Biol.* **118**, 551–560.

Siomi, M.C., Higashijima, K., Ishizuka, A., Siomi, H. (2002) Casein kinase II phosphorylates the fragile X mental retardation protein and modulates its biological properties, *Mol. Cell. Biol.* **22**, 8438–8447.

Siomi, H., Matunis, M.J., Michael, W.M., Dreyfuss, G. (1993a) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif, *Nucleic Acids Res.* **21**, 1193–1198.

Siomi, H., Siomi, M.C., Nussbaum, R.L., Dreyfuss, G. (1993b) The protein product of the fragile X gene, FMR1, has characteristics of an RNA binding protein, *Cell* **74**, 291–298.

Siomi, M.C., Siomi, H., Sauer, W.H., Srinivasan, S., Nussbaum, R.L., Dreyfuss, G. (1995) FXR1, an autosomal homolog of the fragile X mental retardation gene, *EMBO J.* **14**, 2401–2408.

Siomi, M.C., Zhang, Y., Siomi, H., Dreyfuss, G. (1996) Specific sequences in the fragile X syndrome protein FMR1 and the FXR proteins mediate their binding to 60S ribosomal subunits and the interactions among them, *Mol. Cell. Biol.* **16**, 3825–3832.

Sittler, A., Devys, D., Weber, C., Mandel, J.-L. (1996) Alternative splicing of exon 14 determines nuclear or cytoplasmatic localisation of fmr1 protein isoforms, *Hum. Mol. Genet.* **5**, 95–102.

Slegtenhorst-Eegdeman, K.E., de Rooij, D.G., Verhoef-Post, M., Ruiz, A., Uilenbroek, T.J., van de Kant, H.J.G., Bakker, C.E., Oostra, B.A., Grootegoed, J.A., Themmen, A.P.N. (1998) Macro-orchidism in *Fmr1* knockout mice is caused by increased sertoli cell proliferation during normal testes development, *Endocrinology* **139**, 156–162.

Smeets, H.J.M., Smits, A.P.T., Verheij, C.E., Theelen, J.P.G., Willemsen, R., van de Burgt, I., Hoogeveen, A.T., Oosterwijk, J.C., Oostra, B.A. (1995) Normal phenotype in two brothers with a full FMR1 mutation, *Hum. Mol. Genet.* **4**, 2103–2108.

Sullivan, A.K., Crawford, D.C., Scott, E.H., Leslie, M.L., Sherman, S.L. (2002) Paternally transmitted FMR1 alleles are less stable than maternally transmitted alleles in the common and intermediate size range, *Am. J. Hum. Genet.* **70**, 1532–1544.

Sung, Y.J., Conti, J., Currie, J.R., Brown, W.T., Denman, R.B. (2000) RNAs that interact with the fragile X syndrome RNA binding protein FMRP, *Biochem. Biophys. Res. Commun.* **275**, 973–980.

Sung, Y.J., Dolzhanskaya, N., Nolin, S.L., Brown, T., Currie, J.R., Denman, R.B. (2003) The fragile X mental retardation protein FMRP binds elongation factor 1A mRNA and negatively regulates its translation in vivo, *J. Biol. Chem.* **278**, 15669–15678.

Sutcliffe, J.S., Nelson, D.L., Zhang, F., Pieretti, M., Caskey, C.T., Saxe, D., Warren, S.T. (1992) DNA methylation represses FMR-1 transcription in fragile x syndrome, *Hum. Mol. Genet.* **1**, 397–400.

Sutherland, G.R. (1977) Fragile sites on human chromosomes: demonstration of their dependence on the type of tissue culture medium, *Science* **197**, 265, 266.

Tamanini, F., Meijer, N., Verheij, C., Willems, P.J., Galjaard, H., Oostra, B.A., Hoogeveen, A.T. (1996) FMRP is associated to the ribosomes via RNA, *Hum. Mol. Genet.* **5**, 809–813.

Tamanini, F., van Unen, L., Bakker, C.E., Sacchi, N., Galjaard, H., Oostra, B.A.,

Hoogeveen, A.T. (1999) The FXR proteins exist as homo-multimers within messenger RNP particles, *Biochem. J.* 517–523.

Tamanini, F., Willemsen, R., van Unen, L., Bontekoe, C., Galjaard, H., Oostra, B.A., Hoogeveen, A.T. (1997) Differential expression of FMR1, FXR1 and FXR2 proteins in human brain and testis, *Hum. Mol. Genet.* **6**, 1315–1322.

Tassone, F., Hagerman, R.J., Chamberlain, W.D., Hagerman, P.J. (2000) Transcription of the FMR1 gene in individuals with fragile X syndrome, *Am. J. Med. Genet.* **97**, 195–203.

Tassone, F., Hagerman, R.J., Ikle, D.N., Dyer, P.N., Lampe, M., Willemsen, R., Oostra, B.A., Taylor, A.K. (1999) FMRP expression as a potential prognostic indicator in fragile X syndrome, *Am. J. Med. Genet.* **84**, 250–261.

Tassone, F., Hagerman, R.J., Loesch, D.Z., Lachiewicz, A., Taylor, A.K., Hagerman, P.J. (2000a) Fragile X males with unmethylated, full mutation trinucleotide repeat expansions have elevated levels of FMR1 messenger RNA, *Am. J. Med. Genet.* **94**, 232–236.

Tassone, F., Hagerman, R.J., Taylor, A.K., Gane, L.W., Godfrey, T.E., Hagerman, P.J. (2000b) Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome, *Am. J. Hum. Genet.* **66**, 6–15.

Tassone, F., Hagerman, R.J., Taylor, A.K., Mills, J.B., Harris, S.W., Gane, L.W., Hagerman, P.J. (2000c) Clinical involvement and protein expression in individuals with the FMR1 premutation, *Am. J. Med. Genet.* **91**, 144–152.

Taylor, A.K., Safanda, J.F., Fall, M.Z., Quince, C., Lang, K.A., Hull, C.E., Carpenter, I., Staley, L.W., Hagerman, R.J. (1994) Molecular predictors of cognitive involvement in female carriers of fragile X syndrome, *JAMA* **271**, 507–514.

Todd, P.K., Mack, K.J. (2000) Sensory stimulation increases cortical expression of the fragile X mental retardation protein in vivo, *Mol. Brain. Res.* **80**, 17–25.

Todd, P.K., Malter, J.S., Mack, K.J. (2003) Whisker stimulation-dependent translation of FMRP in the barrel cortex requires activation of type I metabotropic glutamate receptors, *Brain. Res. Mol. Brain. Res.* **110**, 267–278.

Turner, G., Till, R., Daniel, A. (1978) Marker X chromosomes, mental retardation and macro-orchidism, *N. Engl. J. Med.* **299**, 1472.

Turner, G., Turner, B. (1974) X-linked mental retardation, *J. Med. Genet.* **11**, 109–113.

Turner, G., Webb, T., Wake, S., Robinson, H. (1996) Prevalence of fragile X syndrome, *Am. J. Med. Genet.* **64**, 196, 197.

Usdin, K., Woodford, K.J. (1995) CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro, *Nucleic Acids Res.* **23**, 4202–4209.

Van Dam, D., D'Hooge, R., Hauben, U., Reyniers, E., Gantois, I., Bakker, C.E., Oostra, B.A., Kooy, R.F., De Deyn, P.P. (2000) Spatial learning, contextual fear conditioning and conditioned emotional response in *Fmr1* knockout mice, *Behav. Brain. Res.* **117**, 127–136.

van den Ouweland, A.M.W., Deelen, W.H., Kunst, C.B., Uzielli, M.L., Nelson, D.L., Warren, S.T., Oostra, B.A., Halley, D.J. (1994) Loss of mutation at the FMR1 locus through multiple exchanges between maternal X chromosomes, *Hum. Mol. Genet.* **3**, 1823–1827.

Verheij, C., Bakker, C.E., de Graaff, E., Keulemans, J., Willemsen, R., Verkerk, A.J.M.H., Galjaard, H., Reuser, A.J.J., Hoogeveen, A.T., Oostra, B.A. (1993) Characterization and localization of the FMR-1 gene product associated with fragile X syndrome, *Nature* **363**, 722–724.

Verheij, C., de Graaff, E., Bakker, C.E., Willemsen, R., Willems, P.J., Meijer, N., Galjaard, H., Reuser, A.J.J., Oostra, B.A., Hoogeveen, A.T. (1995) Characterization of FMR1 proteins isolated from different tissues, *Hum. Mol. Genet.* **4**, 895–901.

Verkerk, A.J.M.H., de Graaff, E., De Boulle, K., Eichler, E.E., Konecki, D.S., Reyniers, E., Manca, A., Poustka, A., Willems, P.J., Nelson, D.L., Oostra, B.A. (1993) Alternative splicing in the fragile X gene FMR1, *Hum. Mol. Genet.* **2**, 399–404.

Verkerk, A.J.M.H., Pieretti, M., Sutcliffe, J.S., Fu, Y.-H., Kuhl, D.P.A., Pizzutti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F., Eussen, B.E., van Ommen, G.-J.B., Blonden, L.A.J., Riggins, G.J., Chastain, J.L., Kunst, C.B., Galjaard, H., Caskey, C.T., Nelson, D.L., Oostra, B.A., Warren, S.T. (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome, *Cell* **65**, 905–914.

Wan, L., Dockendorff, T.C., Jongens, T.A., Dreyfuss, G. (2000) Characterization of dFMR1, a Drosophila melanogaster homolog of the fragile X mental retardation protein, *Mol. Cell. Biol.* **20**, 8536–8547.

Warren, S.T., Zhang, F., Licameli, G.R., Peters, J.F. (1987) The fragile X site in somatic cell hybrids: an approach for molecular cloning of fragile sites, *Science* **237**, 420–423.

Weiler, I.J., Greenough, W.T. (1993) Metabotropic glutamate receptors trigger postsynaptic protein synthesis, *Proc. Natl. Acad. Sci. USA* **90**, 7168–7171.

Weiler, I.J., Irwin, S.A., Klintsova, A.Y., Spencer, C.M., Brazelton, A.D., Miyashiro, K., Comery, T.A., Patel, B., Eberwine, J., Greenough, W.T. (1997) Fragile X mental retardation protein is translated near synapses in response to neurotransmitter activation, *Proc. Natl. Acad. Sci. USA* **94**, 5395–5400.

Wells, R.D. (1996) Molecular basis of genetic instability of triplet repeats, *J. Biol. Chem.* **271**, 2875–2878.

White, P.J., Borts, R.H., Hirst, M.C. (1999) Stability of the human fragile X (CGG)(n) triplet repeat array in Saccharomyces cerevisiae deficient in aspects of DNA metabolism, *Mol. Cell. Biol.* **19**, 5675–5684.

Willemsen, R., Bontekoe, C., Tamanini, F., Galjaard, H., Hoogeveen, A., Oostra, B.A. (1996) Association of FMRP with ribosomal precursor particles in the nucleolus, *Biochem. Biophys. Res. Commun.* **225**, 27–33.

Willemsen, R., Hoogeveen-Westerveld, M., Reis, S., Holstege, J., Severijnen, L., Nieuwenhuizen, I., Schrier, M., VanUnen, L., Tassone, F., Hoogeveen, A., Hagerman, P., Mientjes, E., Oostra, B. (2003) The FMR1 CGG repeat mouse displays ubiquitin-positive intranuclear neuronal inclusions; implications for the cerebellar tremor/ataxia syndrome, *Hum. Mol. Genet.* **12**, 949–959.

Willemsen, R., Los, F., Mohkamsing, S., van den Ouweland, A., Deelen, W., Galjaard, H., Oostra, B. (1997) Rapid antibody test for prenatal diagnosis of fragile X syndrome on amniotic fluid cells: a new appraisal, *J. Med. Genet.* **34**, 250, 251.

Willemsen, R., Mohkamsing, S., De Vries, B., Devys, D., van den Ouweland, A., Mandel, J.L., Galjaard, H., Oostra, B. (1995) Rapid antibody test for fragile X syndrome, *Lancet* **345**, 1147, 1148.

Willemsen, R., Smits, A., Mohkamsing, S., van Beerendonk, H., de Haan, A., De Vries, B., van den Ouweland, A., Sistermans, E., Galjaard, H., Oostra, B.A. (1997) Rapid antibody test for diagnosing fragile X syndrome: a validation of the technique, *Hum. Genet.* **99**, 308–311.

Winnepenninckx, B., Rooms, L., Kooy, R.F. (2003) Mental retardation: a review of the genetic causes, *Br. J. Dev. Disabil.* **49**, 29–44.

Wisniewski, K.E., Segan, S.M., Miezejeski, C.M., Sersen, E.A., Rudelli, R.D. (1991) The Fra(X) syndrome: neurological, electrophysiological, and neuropathological abnormalities, *Am. J. Med. Genet.* **38**, 476–480.

Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J.C., Warren, S.T., Schlessinger, D., Sutherland, G.R., Richards, R.I. (1991) Fragile X genotype characterized by an unstable region of DNA, *Science* **252**, 1179–1181.

Zalfa, F., Giorgi, M., Primerano, B., Moro, A., Di, P., Reis, S., Oostra, B., Bagni, C. (2003) The Fragile fragile X Syndrome syndrome Protein protein FMRP Associates associates with BC1 RNA and Regulates regulates the Translation translation of Specific specific mRNAs at synapses., *Cell* **112**, 317–327.

Zhang, Y.Q., Bailey, A.M., Matthies, H.-J.G., Renden, R.B., Smith, M.A., Speese, S.D., Rubin, G.M., Broadie, K. (2001) Drosophila fragile X-related gene regulates the MAP1B homolog futsch to control synaptic structure and function, *Cell* **107**, 591–603.

Zhang, Y., O'Connor, J.P., Siomi, M.C., Srinivasan, S., Dutra, A., Nussbaum, R.L., Dreyfuss, G. (1995) The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2, *EMBO J.* **14**, 5358–5366.

Free Radicals in Biochemistry and Medicine

Barry Halliwell
National University of Singapore, Singapore

1 What is a Radical? 635

2 Radicals *In Vivo* 635
2.1 Lipid Peroxidation 635
2.2 The Hydroxyl and Superoxide Radicals 636
2.3 Hydrogen Peroxide – A Nonradical 637

3 Radical Reactions 638

4 The Role of Transition Metal Ions 638

5 Antioxidant Defenses 639

6 Transgenic Animal Technology in the Study of Elevated SOD 641

7 Oxidative Stress and Repair Systems 641

8 Consequences of Oxidative Stress 642

9 Oxidative Stress and Human Disease 644

10 Assay Methodology 644
10.1 Aromatic Hydroxylation 645
10.2 Uric Acid Degradation 646
10.3 Assays for Oxygen-derived Species – ''Fingerprint'' Assays 646

11 Conclusion 647

Bibliography 647

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Books and Reviews 647
Primary Literature 647

Keywords

Antioxidant
A molecule that protects a biological target against oxidative damage.

Free Radical
Any species containing one or more unpaired electrons.

Oxidative Damage
The specific biomolecular damage caused by oxidative stress.

Oxidative Stress
An imbalance between the generation of reactive oxygen species and antioxidant protection, in favor of the former.

Polyunsaturated Fatty Acid
A fatty acid with two or more carbon–carbon double bonds in the side chain.

Reactive Oxygen Species
A collective name given to oxygen-containing radicals ($O_2^{\bullet -}$, $OH^{\bullet}$, peroxyl, alkoxyl) and some other nonradical derivatives of oxygen, such as H_2O_2.

Reactive Nitrogen Species
A collective name given to nitrogen-containing radicals ($NO^{\bullet}$, $NO_2^{\bullet}$) and some nonradical reactive derivatives, such as peroxynitrite ($ONOO^-$).

Free radicals and other "reactive" species are constantly generated *in vivo*, by "accidents of chemistry" and for specific metabolic purposes. The reactivity of different free radicals and "reactive species" varies, but some of them can cause severe damage to biological molecules, especially to DNA, lipids, and proteins. Antioxidant defense systems scavenge reactive species and minimize their formation, but are not 100% effective. Hence, repair systems exist to deal with molecules that have been oxidatively damaged. Damage to DNA by hydroxyl radicals appears to occur in all aerobic cells and might be a significant contributor to the age-dependent development of cancer. Similarly, oxidation of lipids by free radicals contributes to atherosclerosis and related vascular dysfunction, and to the vascular damage in diabetes. Free radicals may play a role in the aging process.

1 What is a Radical?

In the structure of atoms and molecules, electrons usually associate in pairs, each pair moving within a defined region of space around the nucleus. This space is referred to as an atomic or molecular *orbital*. One electron in each pair has a spin quantum number of +1/2, the other −1/2. A *free radical* is any species capable of independent existence (hence the term *free*), which contains one or more *unpaired electrons*, an unpaired electron being one that is alone in an orbital. The simplest free radical is an atom of the element hydrogen, with one proton and a single electron. Table 1 gives examples of other free radicals. The spectroscopic technique of *electron spin resonance* (ESR) is used to measure free radicals; it records the energy changes that occur as unpaired electrons align in response to a magnetic field. A superscript dot is used to denote free radical species (Table 1). Another widely used term is *reactive oxygen species*, which includes the oxygen radicals and certain oxidizing species that are not radicals, such as ozone (O_3) and singlet oxygen. A similar term *reactive nitrogen species* has been introduced to include radicals such as $NO^{\bullet}$ and $NO_2^{\bullet}$ (Table 1) and nonradicals such as peroxynitrite (see below).

2 Radicals *In Vivo*

The chemical reactivity of free radicals varies. One of the most reactive is the *hydroxyl radical* ($OH^{\bullet}$). Exposure of living organisms to ionizing radiation causes homolytic fission of O−H bonds in water (the major constituent of living cells)

$$H_2O \longrightarrow H^{\bullet} + OH^{\bullet} \qquad (1)$$

to give $H^{\bullet}$ and $OH^{\bullet}$. The hydroxyl radical reacts at a diffusion-controlled rate with almost all molecules in living cells. Hence, when $OH^{\bullet}$ is formed *in vivo*, it damages whatever it is generated next to – it cannot migrate any significant distance within the cell. The harmful effects on living organisms of excess exposure to ionizing radiation are thought to be largely initiated by attack of $OH^{\bullet}$ on proteins, carbohydrates, DNA, and lipids. For example, $OH^{\bullet}$ can abstract hydrogen atoms from fatty acid side chains in membrane lipids and initiate the process of lipid peroxidation.

2.1 Lipid Peroxidation

Initiation of peroxidation occurs by the attack of any species ($R^{\bullet}$) capable of abstracting hydrogen from a polyunsaturated fatty acid side chain in a membrane (such fatty acid side chains are more susceptible to free radical attack than are saturated or monounsaturated side chains).

$$-CH + R^{\bullet} \longrightarrow -C^{\bullet} + RH \qquad (2)$$

Species able to abstract hydrogen include $OH^{\bullet}$ and peroxyl radicals (Table 1). The carbon-centered radical that resulted from reaction (2) reacts rapidly with $O_2^{\bullet}$

$$-C^{\bullet} + O_2 \longrightarrow CO_2^{\bullet} \qquad (3)$$

and a fatty acid side chain *peroxyl radical* is formed. This can attack adjacent fatty acid side chains and *propagate* lipid peroxidation.

$$-CO_2^{\bullet} + -CH \longrightarrow -CO_2H + -C^{\bullet} \qquad (4)$$

Tab. 1 Examples of free radicals.

Name	*Formula*	*Comments*
Hydrogen atom	H^+	The simplest free radical
Trichloromethyl	$CCl_3^{\bullet}$	A carbon-centered radical (i.e. the unpaired electron resides on carbon. $CCl_3^{\bullet}$ is formed during metabolism of CCl_4 in the liver and contributes to the toxic effects of this solvent.
Superoxide	$O_2^{\bullet -}$	An oxygen-centered radical
Hydroxyl	$OH^{\bullet}$	An oxygen-centered radical. The most highly reactive oxygen radical known.
Thiyl	$RS^{\bullet}$	A group of radicals with an unpaired electron residing on sulfur.
Peroxyl, alkoxyl	$RO_2^{\bullet}$, $RO^{\bullet}$	Oxygen-centered radicals formed during the breakdown of organic peroxides.
Oxides of nitrogen	$NO^{\bullet}$, $NO_2^{\bullet}$	Both are free radicals. NO is formed *in vivo* from the amino acid L-arginine. NO_2 is found in polluted air and smoke from burning organic materials (e.g. cigarette smoke)

The chain reaction thus continues, and *lipid peroxides* ($-CO_2H$) accumulate in the membrane. Lipid peroxides destabilize membranes and make them "leaky" to ions. Peroxyl radicals can attack not only lipids but also membrane proteins (e.g. damaging enzymes, receptors, and signal transduction systems), and they can oxidize cholesterol.

2.2 The Hydroxyl and Superoxide Radicals

When $OH^{\bullet}$ is generated adjacent to DNA, it attacks both the deoxyribose and the purine and pyrimidine bases. Figure 1 shows the structures of some of the products generated by the attack of $OH^{\bullet}$ on the DNA bases: this wide range of products is characteristic of attack by $OH^{\bullet}$ and may be used to show that such attack has occurred *in vivo*. For example, if most or all of the compounds given in Fig. 1 are present in DNA that has been extracted from a tissue, this is strong evidence that the DNA has suffered attack by $OH^{\bullet}$. Such "$OH^{\bullet}$ fingerprint" experiments have been used to study the role of free radicals in DNA damage by radiation and toxic agents. It has also been found that there are larger amounts of these $OH^{\bullet}$-derived products in DNA from human cancerous tumors than from normal tissue. Whether this is due to increased $OH^{\bullet}$ formation or to decreased repair of the damage (see Section 7) remains to be evaluated.

Whereas $OH^{\bullet}$ is probably always harmful, other (less reactive) free radicals may often be useful *in vivo*. Free radicals are known to be produced metabolically in living organisms. For example, the free radical nitric oxide ($NO^{\bullet}$) is synthesized from the amino acid L-arginine by vascular endothelial cells, phagocytes, certain cells in the brain, and many other cell types. Nitric oxide is a vasodilator agent and an important neurotransmitter. It may also be involved in the killing of infectious organisms by macrophages in some mammalian species.

Superoxide radical ($O_2^{\bullet -}$) is the one-electron reduction product of oxygen. It is produced by phagocytic cells (neutrophils, monocytes, macrophages, eosinophils)

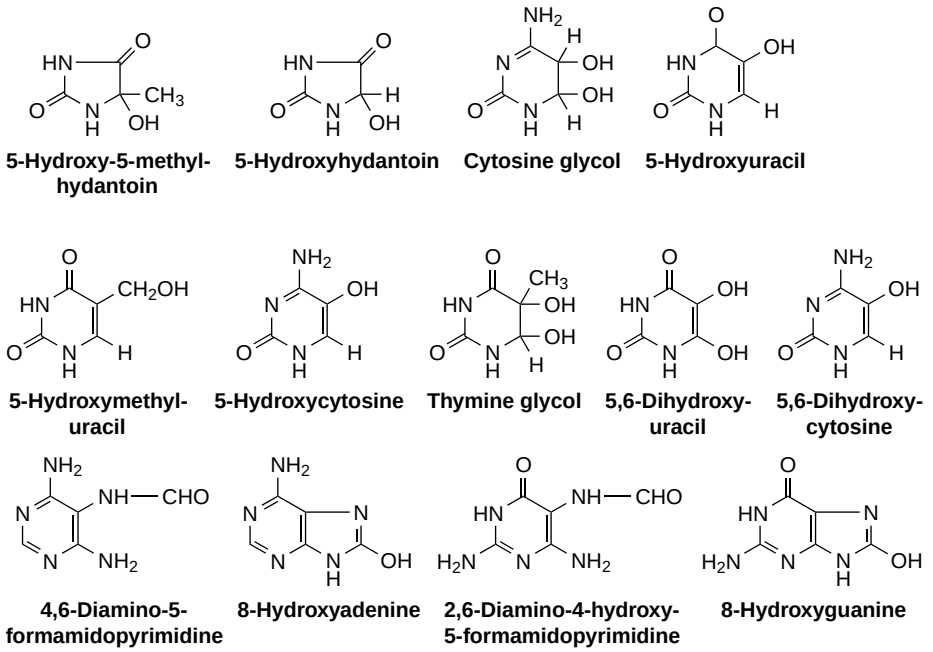

Fig. 1 Some products of hydroxyl radical attack on DNA bases.

and helps them to kill bacteria. Smaller amounts of extracellular $O_2^{\bullet-}$ can be generated, perhaps as an intercellular signal molecule, by several other cell types, including endothelial cells, lymphocytes, and fibroblasts. In addition to this deliberate metabolic production of $O_2^{\bullet-}$, some $O_2^{\bullet-}$ is produced within cells by mitochondria and endoplasmic reticulum. This is often thought to be an unavoidable consequence of "leakage" of electrons onto O_2 from their correct paths in electron transfer chains (i.e. an "accident of chemistry"), but others have argued that the H_2O_2 generated from $O_2^{\bullet-}$ signals the redox state of the mitochondria to the rest of the cell (see below).

Aerobic organisms also contain many "autoxidizable" compounds, compounds that react directly with oxygen to generate free radicals. For example, adrenaline (epinephrine) slowly reacts with O_2 to form $O_2^{\bullet-}$, which then oxidizes more epinephrine, setting up a chain reaction. These autoxidizable compounds presumably react with O_2 *in vivo*, providing another source of $O_2^{\bullet-}$.

2.3 Hydrogen Peroxide – A Nonradical

Most of the $O_2^{\bullet-}$ generated *in vivo* probably undergoes a dismutation reaction, represented by the overall equation:

$$2O_2^{\bullet-} + 2H^+ \longrightarrow H_2O_2 + O_2 \qquad (5)$$

Hydrogen peroxide (H_2O_2), a nonradical, resembles water in its molecular structure and is very diffusible within and between cells, probably crossing plasma membranes through the aquaporin channels. As well as arising from $O_2^{\bullet-}$, H_2O_2

is produced directly by the action of certain oxidase enzymes in cells, including amino acid oxidases and the enzyme xanthine oxidase. Xanthine oxidase catalyzes the oxidation of hypoxanthine to xanthine and of xanthine to uric acid. Oxygen is simultaneously reduced both to $O_2^{\bullet -}$ and to H_2O_2. Low levels of xanthine oxidase are present in many mammalian tissues, especially in the gastrointestinal tract. The levels of xanthine oxidase often increase when tissues are subjected to insult, such as trauma or deprivation of oxygen.

Like $O_2^{\bullet -}$, H_2O_2 has certain useful metabolic functions. For example, H_2O_2 is used by the enzyme *thyroid peroxidase* to help make thyroid hormones. H_2O_2 is involved in cell-signaling mechanisms. For example, in certain cell types it acts as a second messenger leading to displacement of the inhibitory subunit from the cytoplasmic transcription factor NF-κB. In lymphocyte cell lines, for example, H_2O_2 induces the expression of genes controlled by NF-κB. In some cellular systems, H_2O_2 mediates the effects of insulin and (transforming growth factor-β) TGFβ. How widespread and important these H_2O_2-dependent signaling systems are *in vivo* remains to be elucidated.

3 Radical Reactions

If two free radicals meet, they can join their unpaired electrons to form a covalent bond. Thus, for two hydrogen atoms,

$$H^{\bullet} + H^{\bullet} \longrightarrow H_2 \tag{6}$$

and for $NO^{\bullet}$ and $O_2^{\bullet -}$,

$$NO^{\bullet} + O_2^{\bullet -} \longrightarrow \underset{\text{Peroxynitrite}}{ONOO^-} \tag{7}$$

Peroxynitrite rapidly protonates to form peroxynitrous acid

$$ONOO^- + H^+ \longrightarrow \underset{\text{Peroxynitrous acid}}{ONOOH} \tag{8}$$

ONOOH and its products of decomposition and reaction with CO_2, which appear to include some $OH^{\bullet}$ and $NO_2^{\bullet}$, can cause severe damage to proteins, lipids, and DNA. Peroxynitrite is an example of a "reactive nitrogen species."

However, when a free radical reacts with a nonradical, a new radical results; that is, a *chain reaction* is set up. Since most biological molecules are nonradicals, the generation *in vivo* of reactive radicals such as $OH^{\bullet}$ usually sets off chain reactions. For example, attack of $OH^{\bullet}$ on fatty acid side chains in membranes and other biological lipoproteins can abstract hydrogen, leaving a carbon-centered radical and initiating the process of lipid peroxidation (Section 2.1). Attack of $OH^{\bullet}$ on the DNA bases produces radicals, which then undergo complex reactions to generate the end products shown in Fig. 1. For example, the initial radical produced when $OH^{\bullet}$ attacks guanine at position 8 can then form 8-hydroxyguanine by electron loss, or can break down to 2,6-diamino-4-hydroxy-5-formamidopyrimidine (Fig. 1).

4 The Role of Transition Metal Ions

Neither $O_2^{\bullet -}$ nor H_2O_2 is very reactive, although they do react with certain cellular targets. For example, $O_2^{\bullet -}$ can inactivate aconitase and interfere with mitochondrial energy production, whereas H_2O_2 can inhibit the caspase enzymes important in apoptosis. Nevertheless, the limited reactivity of $O_2^{\bullet -}$ and H_2O_2 allows them to be used for such purposes as signal

transduction: a similar comment applies to the generally poorly reactive radical $NO^{\bullet}$, which plays essential roles *in vivo*. However, if $O_2^{\bullet -}$ and H_2O_2 come into contact with iron or copper ions, the noxious $OH^{\bullet}$ can be formed.

$$O_2^{\bullet -} + H_2O_2 \xrightarrow{\text{Fe or Cu catalyst}} OH^{\bullet} + OH^{-} + O_2 \qquad (9)$$

However, the chemistry involved is much more complicated than is suggested by this simple summary equation. Hence, the presence, in a biological system, of transition metal ions able to catalyze reaction (9) can cause $O_2^{\bullet -}$ and H_2O_2 to be damaging. Because $OH^{\bullet}$ cannot migrate any significant distance, the damage will occur at the sites at which the "catalytic" metal ions are present. That is, it is *site-specific* free radical damage. For example, if such metal ions are bound to DNA, and $O_2^{\bullet -}$ plus H_2O_2 reach the DNA, $OH^{\bullet}$ will be generated and will damage the DNA (Fig. 1). However, if the metal ions were bound to membranes, then lipid peroxidation could result (Section 2.1).

Another way of causing damage by $O_2^{\bullet -}$ is via $ONOO^{-}$ formation (see above). At normal physiological levels of $O_2^{\bullet -}$ and $NO^{\bullet}$, $ONOO^{-}$ formation appears slow. At sites of tissue injury, more $O_2^{\bullet -}$ is formed as a result of phagocyte recruitment, and more $NO^{\bullet}$ is produced, usually owing to elevated expression of the genes encoding the inducible nitric oxide synthase enzyme, iNOS. At these higher levels of $O_2^{\bullet -}$ and $NO^{\bullet}$, more $ONOO^{-}$ is formed. One of the reactions of ONOOH with proteins leads to nitration of tyrosine residues, and the formation of 3-nitrotyrosine is often used as a "biomarker" of $ONOO^{-}$ formation *in vivo*.

5 Antioxidant Defenses

Living organisms have evolved antioxidant defenses to remove excess $O_2^{\bullet -}$ and H_2O_2. *Superoxide dismutase* enzymes (SODs) remove $O_2^{\bullet -}$ by accelerating its conversion to H_2O_2 (Eqn. 5) by about four orders of magnitude at pH 7.4. Mammalian cells have a SOD enzyme containing manganese at its active site (MnSOD) in the mitochondria. A SOD with copper and zinc at the active site (CuZnSOD) is also present, largely in the cytosol, but some CuZn SoD is also present in the space between the inner and outer mitochondrial membranes.

Because SOD enzymes generate H_2O_2, they work in collaboration with H_2O_2-removing enzymes. *Catalases* convert H_2O_2 to water and O_2.

$$2H_2O_2 \longrightarrow 2H_2O + O_2 \qquad (10)$$

Catalases are present in the peroxisomes of mammalian cells and probably serve to destroy H_2O_2 generated by oxidase enzymes located within these subcellular organelles. More important H_2O_2-removing enzymes in mammalian cells are the *glutathione peroxidases* (GSHPX), enzymes that require selenium (a selenocysteine residue residue, essential for enzyme activity, is present at the active site). GSHPX enzymes remove H_2O_2 by using it to oxidize reduced glutathione (GSH) to oxidized glutathione (GSSG). Glutathione reductase, a flavoprotein (FAD-containing) enzyme, regenerates GSH from GSSG, with NADPH as a source of reducing power (Fig. 2). Removal of H_2O_2 may also be achieved by thioredoxin-dependent systems. Thioredoxin is a small protein (Mr ~12 kDa) that can react with H_2O_2 to become oxidized in the presence of a peroxidase enzyme. A thioredoxin reductase, also a selenoprotein, regenerates the thioredoxin for reuse.

$$2GSH + H_2O_2 \longrightarrow GSSG + 2H_2O$$

$$2GSH + \text{fatty acid—OOH} \xrightarrow[\text{Glutathione peroxidase}]{} GSSG + \text{fatty acid—OH} + H_2O$$

$$GSSG + NADPH + H^+ \xrightarrow[\text{Glutathione reductase}]{} 2GSH + NADP^+$$

Fig. 2 The glutathione system. Reduced glutathione is a tripeptide, glutamic acid–cysteine–glycine. It is present at millimolar concentrations in most mammalian cells. In oxidized glutathione (GSSG), two tripeptides are linked by a disulfide bridge. Glutathione has several additional metabolic functions. Glutathione peroxidase can also destroy fatty acid (lipid–OOH) peroxides by converting them to hydroxy alcohols (lipid–OH). Mammalian cells additionally contain a phospholipid hydroperoxide glutathione peroxidase that can apparently perform the same reaction on lipid peroxides *within membranes;* how exactly it works is uncertain as yet.

The thioredoxin system additionally plays a key role in regulation of gene expression and cell function generally.

Another important antioxidant defense system consists of metal ion storage and transport proteins, which organisms have evolved to keep iron and copper safely protein-bound whenever possible, thus minimizing the occurrence of reaction (9).

Antioxidant defense enzymes are essential for healthy aerobic life. Because of impaired biosynthesis of certain amino acids, for example, SOD-negative mutants of *E. coli* will not grow aerobically unless given a rich growth medium. Even when so supplemented, SOD-negative *E. coli* cells grow slowly, suffer membrane damage, are abnormally sensitive to damage by H_2O_2 (perhaps because of reaction 9), and show a high mutation rate. In addition to antioxidant defense enzymes, some small-molecule free radical scavengers exist. GSH can scavenge various free radicals and ONOOH directly, as well as being a substrate for GSHPX enzymes. α-Tocopherol is the most important (but not the only) free radical scavenger within membranes and lipoproteins. α-Tocopherol inhibits lipid peroxidation by scavenging peroxyl radicals (Table 1), which are intermediates in the chain reaction described in Section 2.1:

$$\alpha - TH + LOO^{\bullet} \longrightarrow \alpha T^{\bullet} + LOOH \tag{11}$$

where $\alpha T^{\bullet}$ is the α-tocopherol radical and $LOO^{\bullet}$ is the peroxyl radical. Since $\alpha T^{\bullet}$ is less efficient at abstracting hydrogen than $LOO^{\bullet}$, the chain reaction of peroxidation is slowed down. Several biological mechanisms may exist for recycling $\alpha T^{\bullet}$ back to α-tocopherol, although none of them has yet been proved to operate *in vivo*. Likely mechanisms include the reaction of $\alpha T^{\bullet}$ with ascorbic acid at the surface of membranes and lipoproteins

$$\alpha T^{\bullet} + \text{ascorbate} \longrightarrow \alpha TH + \text{ascorbate}^{\bullet} \tag{12}$$

and/or with ubiquinol (reduced coenzyme Q) within membranes or lipoproteins:

$$CoQH_2 + \alpha T^{\bullet} \longrightarrow \alpha TH + CoQH^{\bullet} + H^+ \tag{13}$$

$CoQH_2$ can also directly scavenge $RO_2^{\bullet}$ radicals. Antioxidant defense enzymes

exist as a balanced coordinated system in mammals. Thus, although SOD is an important defense, an excess of SOD in relation to the activity of peroxide-metabolizing enzymes can be deleterious. This has been shown by transfecting cells with human cDNAs encoding SOD. The consequences of excess SOD activity may be relevant to the clinical condition known as *Down syndrome* (DS), a point being explored by the use of transgenic animals, as indicated in Section 6.

6 Transgenic Animal Technology in the Study of Elevated SOD

Down's syndrome is the most common human genetic disorder, occurring once in every 600 to 800 live births. Defects that may be suffered by DS individuals include short stature, malformations of the skin around the eyes, and mental retardation. Patients who survive beyond their thirties show an increased risk of developing Alzheimer's disease. DS is usually caused by the presence of three copies of chromosome 21 (trisomy 21), one of whose genes encodes Cu/Zn-SOD. Patients with DS show some evidence of increased oxidative damage.

Does the elevated level of Cu/Zn-SOD cause, or contribute to, DS? Transgenic mice that overexpress human Cu/Zn-SOD activity in all tissues are found to be more resistant than controls to O_2 toxicity, consistent with a role of $O_2^{\bullet-}$ in this process. They are also more resistant than controls to certain toxins, but they have abnormal neuromuscular junctions in the tongue and show some of the other neurological defects characteristic of Downs syndrome.

7 Oxidative Stress and Repair Systems

Normally, the production of $O_2^{\bullet-}$ and H_2O_2 is approximately balanced by the antioxidant defense systems; that is, the antioxidants are not present in great excess. One reason for this may be that production of some $O_2^{\bullet-}$ and H_2O_2 is useful *in vivo*, for example, for signaling purposes. Indeed, some free radical damage to DNA, proteins, and lipids occurs constantly even in health, and it has been argued that the aging process involves cumulative free radical damage over the lifetime of a species. However, if antioxidant levels fall or production of $O_2^{\bullet-}$ and H_2O_2 increases, an imbalance occurs and *oxidative stress* is said to result. This may result in increased levels of oxidative damage to biomolecules.

Oxidative stress can result from the following phenomena:

- Depletion of antioxidants – for example, through inadequate dietary intake of α-tocopherol, ascorbic acid (vitamin C), sulfur-containing amino acids (needed for GSH manufacture), or riboflavin (needed to make the FAD cofactor in glutathione reductase).
- Excess production of oxygen-derived species – for example, by exposure to elevated O_2 concentrations, by the presence of toxins that are metabolized to produce free radicals, or by excessive activation of "natural" radical-producing systems (e.g. inappropriate activation of phagocytic cells in chronic inflammatory diseases, such as rheumatoid arthritis and ulcerative colitis).

Cells can tolerate mild oxidative stress, which often results in upregulation of the synthesis of antioxidant defense enzymes in an attempt to restore the balance. For

example, exposure of *E. coli* to toxins that increase $O_2^{\bullet-}$ production accelerates the biosynthesis of at least 40 different proteins. Nine of these (including MnSOD) belong to the same regulon, controlled by the *sox* locus, which contains two adjacent genes, *soxR* and *soxS*. Expression of *soxS* is increased by $O_2^{\bullet-}$, and it then leads to expression of *soxR*. The excess $O_2^{\bullet-}$ forms an excess of H_2O_2, which oxidizes the protein oxyR, leading to activation of the transcription of another panel of genes, including genes encoding catalase and glutathione reductase enzymes (Fig. 3). Because antioxidant defenses do not scavenge 100% of the radicals generated *in vivo*, some oxidative damage occurs.

This damage is continuously repaired, however, with the result that baseline levels of damage are kept low. Table 2 summarizes some of the repair systems known to exist. Failure of repair systems may contribute to the aging process.

8 Consequences of Oxidative Stress

Although most tissues can adapt to mild oxidative stress, severe oxidative stress can cause cell damage and death, by a series of interacting mechanisms. Thus, oxidative stress appears to cause an increase in the levels of "free" Ca^{2+} within cells, as well as "free" iron, which can lead to $OH^{\bullet}$ generation. Some of this $OH^{\bullet}$ generation seems to occur within the nucleus, so that DNA is attacked. The hydroxyl radical attacks DNA in a multiplicity of ways (Fig. 1); one of the main products of $OH^{\bullet}$ attack on the purine bases is 8-hydroxyguanine, a mutagenic lesion that

Treat with H_2O_2: Upon treatment with low doses, *E. coli* adapts and becomes resistant to high levels of H_2O_2 that normally would kill it. One of the adaptation mechanisms is as follows:

Treatment with H_2O_2
↓
Oxidation of oxyR protein
↓
Activation of transcription of genes to increase synthesis of at least eight proteins, including catalase and glutathione reductase.

H_2O_2 also activates expression of other protective genes in *E. coli*, by mechanisms not involving *oxyR*.

Treat with $O_2^{\bullet-}$: The H_2O_2 response system above may be activated as follows:

Treatment with $O_2^{\bullet-}$

Formation of H_2O_2 and activation of the above response system

Activation of *soxR* and *soxS* genes, leading to increased synthesis of at least nine proteins, including MnSOD and a DNA repair enzyme

Fig. 3 Regulation of antioxidant defenses in *E. coli*.

Tab. 2 Repair of oxidative damage.

Substrate of damage	***Repair system***
DNA. All components of DNA can be attacked by $OH^{\bullet}$. Several other reactive oxygen species (ROS) attack guanine preferentially, as does ONOOH; H_2O_2 and $O_2^{\bullet -}$ do not attack DNA.	A wide range of enzymes exist that recognize abnormalities in DNA and remove them by excision, resynthesis, and rejoining of the DNA strand. There is some evidence that DNA repair capacity may decline in the elderly.
Proteins. Many ROS can oxidize – SH groups. $OH^{\bullet}$ attacks many amino acid residues. Proteins often bind transition metal ions, making them a target of attack by site-specific $OH^{\bullet}$ generation.	Oxidized methionine residues may be repaired by a methionine sulfoxide reductase enzyme. Other damaged proteins are destroyed by the cellular proteasome system. Proteasomal activity may decline in the elderly, and low activities have been found in brain tissues from subjects with Alzheimer's and Parkinson's diseases, where there is an accumulation of oxidized proteins.
Lipids. Some ROS (not including $O_2^{\bullet -}$ or H_2O_2) can initiate lipid peroxidation.	Chain-breaking antioxidants (especially α-tocopherol) remove chain-propagating peroxyl radicals. Phospholipid hydroperoxide glutathione peroxidase (Fig. 2) can remove peroxides from membranes. Normal membrane turnover can replace damaged lipids.

can lead to $G \rightarrow T$ transversions. An excessive rise in intracellular free Ca^{2+} can activate endonucleases and cause DNA fragmentation.

Several toxins impose oxidative stress during their metabolism. Carbon tetrachloride is one example (Table 1): it is metabolized *in vivo* by endoplasmic reticulum to produce a free radical (trichloromethyl peroxyl radical, $CCl_3O_2^{\bullet}$) that is efficient at initiating lipid peroxidation, by abstracting hydrogen from polyunsaturated fatty acid side chains (Section 2.1). Another is paraquat, a herbicide that causes lung damage in humans. Its metabolism within the lung leads to production of large amounts of $O_2^{\bullet -}$ and H_2O_2. Paraquat is reduced to a free radical by lung enzymes.

$$PQ^{2+} + e^- \longrightarrow PQ^{\bullet +} \qquad (14)$$

This radical then reacts very fast with O_2 in a nonenzymic reaction:

$$PQ^{\bullet +} + O_2 \longrightarrow PQ + O_2^{\bullet -} \qquad (15)$$

The paraquat is regenerated to repeat the cycle. Other examples of such "redox cycling" toxins are alloxan, diquat, and 6-hydroxydopamine. Cigarette smoking has many deleterious health effects, and oxidative stress is believed to be involved in several of these. Table 3 summarizes the mechanisms by which cigarette smoke can impose oxidative stress.

Oxidative stress can lead to increased levels of oxidative damage to biomolecules. For example, specific proteins undergoing oxidative damage in the brains of subjects with Alzheimer's disease have been identified. Hyperglycemia can cause

Tab. 3 Why cigarette smoking may impose oxidative stress.

Smoke contains many free radicals (both in the gas and tar phases), especially peroxyl radicals, that might attack biological molecules and deplete antioxidants, such as ascorbic acid and α-tocopherol. Smoke contains oxides of nitrogen, including the unpleasant nitrogen dioxide ($NO_2^{\bullet}$), a free radical that can initiate lipid peroxidation. The tar phase of smoke contains lipid-soluble hydroquinones, which may enter the tissues and the redox cycle. Some may also release iron from the iron storage protein ferritin, providing "free" iron to catalyze OH formation. • Smoking may irritate lung macrophages, encouraging them to make $O_2^{\bullet -}$. • Smokers' lungs contain more neutrophils than the lungs of nonsmokers, and smoke might activate these cells to make $O_2^{\bullet -}$. • Smokers who eat poorly and drink more alcohol than nonsmokers may have an insufficient dietary intake of nutrient antioxidants. • The effects of cigarette smoke on phagocytes are dose-related. Low levels may stimulate them, but high levels may poison phagocytes and so depress their activity.

severe oxidative stress, accounting for elevated levels of oxidative damage in diabetic patients.

9 Oxidative Stress and Human Disease

Oxidative stress is an inevitable accompaniment of tissue injury during human disease, for the reasons summarized in Fig. 4. For example, when phagocytes produce excess $O_2^{\bullet -}$, H_2O_2, and other species at sites of chronic inflammation, severe damage can result. This seems to happen in the inflamed joints of patients with rheumatoid arthritis and in the gut of patients with inflammatory bowel diseases, such as Crohn's disease and ulcerative colitis. Tissue injury can release metal ions from their storage sites within cells, leading to $OH^{\bullet}$ generation. This may be a particularly important mechanism in the brain, in that iron-dependent free radical reactions can occur after injury (e.g. by trauma or ischemia), spreading the injury to adjacent cells. Since oxidative stress occurs to some degree after every tissue injury, the main question to be asked in evaluating its role in human disease is not "Can we demonstrate oxidative stress?" but rather "Does the oxidative stress that occurs make a significant contribution to disease activity?" The answer to the latter question appears to be "yes" in at least some cases, including rheumatoid arthritis, Alzheimer's disease, diabetes, atherosclerosis, and ulcerative colitis. However, it may well be "no" in many others. Elucidating the precise role played by free radicals has not been easy because such species are difficult to measure, but the development of modern assay techniques is helping solve this problem.

10 Assay Methodology

The only technique that can detect free radicals directly is ESR spectroscopy, but only poorly reactive radicals can accumulate *in vivo* to levels that ESR techniques can detect. Hence, ESR studies of biological material detect fairly unreactive radicals, such as ascorbate radical. Highly reactive

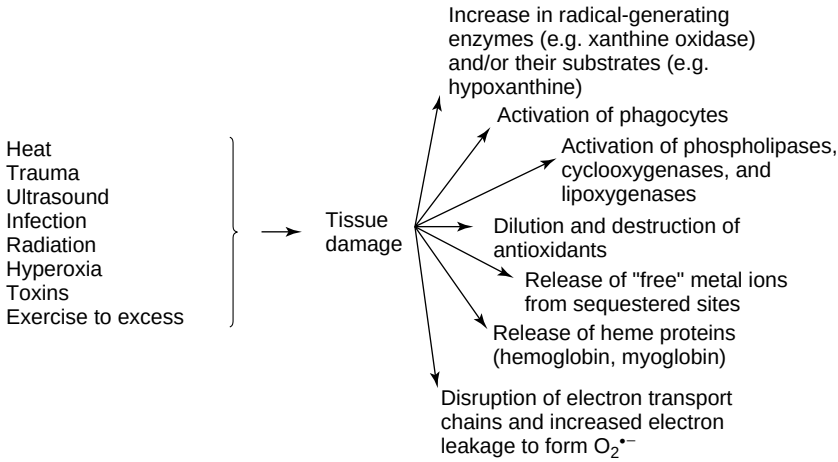

Fig. 4 How tissue damage can cause oxidative stress. [From Halliwell et al. (1992), with permission.]

radicals formed in biological systems can be identified by two general approaches. The first is trapping, and the second involves the measurement of the end products of free radical attack (Sections 10.2 and 10.3).

In trapping, the radical is allowed to react with a trap molecule to give one or more stable products, which are then measured. The most popular trapping method is spin trapping, in which the radical reacts with a "trap molecule" to form a more stable radical, which does accumulate to the level detectable by ESR. Spin traps such as α-phenyl-*tert*-butyl nitrone and 5,5-dimethylpyrroline-*N*-oxide have been useful in detecting certain free radicals *in vitro* and in whole animals, but they cannot be administered to humans.

10.1 Aromatic Hydroxylation

There are many trapping methods other than spin trapping. The technique of aromatic hydroxylation, for example, is based on the reaction of $OH^{\bullet}$ generated under physiological conditions with aromatic compounds at a diffusion-controlled rate, giving rise predominantly to hydroxylated end products. An aromatic hydroxylation assay was first applied to humans using salicylate (2-hydroxybenzoate) as a "trap" for $OH^{\bullet}$. Attack of $OH^{\bullet}$ on salicylate produces two major hydroxylated products: 2,3-dihydroxybenzoate and 2,5-dihydroxybenzoate. The latter product can be produced metabolically, whereas 2,3-dihydroxybenzoate apparently cannot. Hence, the formation of the latter product may be used to detect $OH^{\bullet}$ production *in vivo*.

Another aromatic trap for $OH^{\bullet}$ is the amino acid, phenylalanine. *In vivo*, the L-isomer of this amino acid is hydroxylated by phenylalanine hydroxylase at position 4 on the ring to give L-*p*-tyrosine. D-Phenylalanine is not recognized by this enzyme. By contrast, $OH^{\bullet}$ cannot distinguish between the two isomers: it acts on both L- and D-phenylalanine to produce a mixture of *o*-, *m*-, and *p*-tyrosines

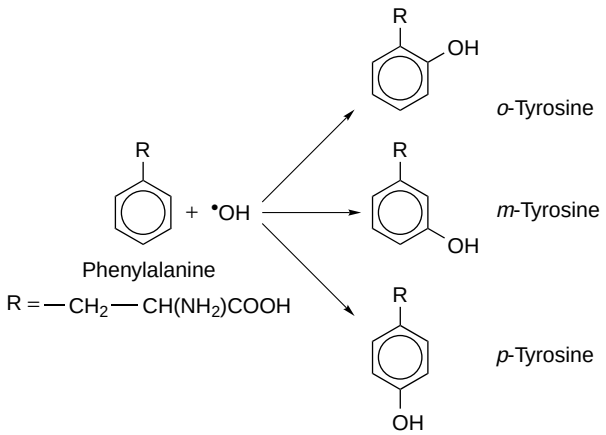

Fig. 5 Products resulting from the attack of hydroxyl radicals on the amino acid phenylalanine. Hydroxyl radicals add on to the aromatic ring to produce intermediate radicals, which are converted to hydroxylated products under physiological conditions.

(Fig. 5). Formation of these tyrosines has been used as an index of $OH^{\bullet}$ production by cells, by injured organs, and in food that has been irradiated.

10.2 Uric Acid Degradation

In humans and other primates, which lack a functional urate oxidase enzyme, uric acid is an end product of purine metabolism. (A gene resembling the urate oxidase gene of lower mammals is present in the human genome, but there is a stop codon in one of the exons.) It is widely believed that uric acid acts as an antioxidant *in vivo*. Hence, measurement of the products of attack of oxygen-derived species on uric acid is a potential marker of oxidative damage uniquely applicable to humans and to other primates. Uric acid is oxidized by several reactive oxygen species (including $OH^{\bullet}$, but not $O_2^{\bullet -}$ or H_2O_2). The major product of uric acid oxidation in all cases is allantoin, but other products formed are oxonic acid, oxaluric acid, cyanuric acid, and parabanic acid. The concentrations of all these products are increased in synovial fluid and serum from patients suffering from rheumatoid arthritis and in serum from patients with iron overload arising as a consequence of idiopathic hemochromatosis. The blood plasma of iron-overloaded hemochromatosis patients contains "free" iron, which can catalyze free radical reactions. By contrast, healthy subjects never contain "free" plasma iron or copper – it is safely bound to transport proteins.

10.3 Assays for Oxygen-derived Species – "Fingerprint" Assays

Instead of attempting to "trap" oxygen-derived species, one can sometimes implicate them as agents of tissue injury by examining the pattern of chemical change they produce upon reaction with certain biological molecules. Thus, end

products of lipid peroxidation (e.g. lipid hydroperoxides, Section 2.1) or of free radical damage to DNA (Fig. 1) and proteins can be measured. Attack on proteins of $OH^{\bullet}$ generated in site-specific metal ion–dependent reactions can generate carbonyl compounds on the proteins, and these "protein carbonyls" are often measured as a general index of such protein damage *in vivo*. Proteomic techniques are now being used to identify the specific proteins that are the target of oxidative damage in some human diseases.

11 Conclusion

Free radicals are a normal part of human metabolism: we cannot escape them. They can be both favorable and unfavorable, depending on the amount, the location, the chemical nature, and the levels of antioxidant defense.

See also Electron Spin Resonance of Biomolecules.

Bibliography

Books and Reviews

Beckman, J.S., Koppenol, W.H. (1996) Nitric oxide, superoxide, and peroxynitrite: the good, the bad, and ugly, *Am. J. Physiol.* **271**, C1424–C1437.

Brigelius-Flohe, R., Traber, M.G. (1999) Vitamin E: function and metabolism, *FASEB J.* **10**, 1145–1155.

Cadenas, E., Packer, L. (2001) *Handbook of Antioxidants*, 2nd edition, Marcel Dekker. New York.

Droge, W. (2002) Free radicals in the physiological control of cell function, *Physiol. Rev.* **82**, 47–95.

Fridovich, I. (1999) Fundamental aspects of reactive oxygen species, or what's the matter with oxygen? *Ann. N.Y. Acad. Sci.* **896**, 13–18.

Halliwell, B. (2001) Role of free radicals in the neurodegenerative diseases: therapeutic implications for antioxidant treatment, *Drugs Aging* **18**, 685–716.

Halliwell, B., Gutteridge, J.M.C. (1990) Role of free radicals and catalytic metal ions in human disease: an overview, *Methods Enzymol.* **186**, 1–85.

Halliwell, B., Gutteridge, J.M.C. (1999) *Free Radicals in Biology and Medicine*, 3rd edition, Clarendon Press, Oxford.

Halliwell, B., Gutteridge, J.M.C., Cross, C.E. (1992) Free radicals, antioxidants and human disease: where are we now? *J. Lab. Clin. Med.* **119**, 598–620.

Meister, A. (1995) Glutathione metabolism, *Methods Enzymol.* **251**, 3–7.

Rosen, G.M., Britigan, B.E., Halpern, H.J., Pou, S. (1999) *Free Radicals, Biology and Detection by Spin Trapping*, Oxford University Press, New York.

Schreck, R., Albermann, K.A.J., Baeuerle, P.A. (1992) Nuclear factor κB: An oxidative stress responsive transcription factor of eukaryotic cells (a review), *Free Radical Res. Commun.* **17**, 221–237.

Sies, H. (Ed.) (1991) *Oxidative Stress, Oxidants and Antioxidants*, Academic Press, London.

Sohal, R., Mockett, R., Orr, W. (2002) Mechanisms of aging: an appraisal of the oxidative stress hypothesis, *Free Radical Biol. Med.* **33**, 575.

Von Sonntag, C. (1987) *The Chemical Basis of Radiation Biology*, Taylor & Francis, London.

Primary Literature

Amstad, P., Peskin, A., Shah, G., Mirault, M.E., Moret, R., Zbinden, I., Cerutti, P. (1991) The balance between Cu,Zn-superoxide dismutase and catalase affects the sensitivity of mouse epidermal cells to oxidative stress, *Biochemistry* **30**, 9305–9313.

Babior, B.M., Lambeth, J.D., Nauseef, W. (2002) The neutrophil NADPH oxidase, *Arch. Biochem. Biophys.* **397**, 342–344.

Bredt, D.S. (1999) Endogenous nitric oxide synthesis: biological functions and pathophysiology, *Free Radical Res.* **31**, 577–596.

Brookes, P., Levonen, A., Shiva, S., Sarti, P., Darley-Usmar, V. (2002) Mitochondria: regulators of signal transduction by reactive oxygen and nitrogen species, *Free Radical Biol. Med.* **33**, 755.

Cadet, J., Bellon, S., Berger, M., Bourdat, A.G.M., Douki, T., Duarte, V., Frelon, S., Gasparutto, D., Muller, E., Ravanat, J.L., Sauvaigo, S. (2002) Recent aspects of oxidative DNA damage: guanine lesions, measurement and substrate specificity of DNA repair glycosylases, *Biol. Chem.* **383**, 933–943.

Castegna, A., Aksenov, M., Thongboonkerd, V., Klein, J.B., Pierce, W.M., Booze, R., Markesbery, W.R., Butterfield, D.A. (2002) Proteomic identification of oxidatively modified proteins in Alzheimer's disease brain. Part II: dihydropyrimidinase-related protein 2, alpha-enolase and heat shock cognate 71, *J. Neurochem.* **82**, 1524–1532.

Chen, D., Cao, G., Hastings, T., Feng, Y., Pei, W., O'Horo, C., Chen, J. (2002) Age-dependent decline of DNA repair activity for oxidative lesions in rat brain mitochondria, *J. Neurochem.* **81**, 1273–1284.

Chevion, M., Berenshtein, E., Stadtman, E.R. (2000) Human studies related to protein oxidation: protein carbonyl content as a marker of damage, *Free Radical Res. Commun.* **33**(Suppl 1), S99–108.

De Deken, X., Wang, D., Dumont, J.E., Miot, F. (2002) Characterization of ThOX proteins as components of the thyroid H_2O_2-generating system, *Exp. Cell Res.* **273**, 187–196.

Dean, R.T., Fu, S., Stocker, R., Davies, M.J. (1997) Biochemistry and pathology of radical-mediated protein oxidation, *Biochem. J.* **324**, 1–18.

Dizdaroglu, M., Jaruga, P., Birincioglu, M., Rodriguez, H. (2002) Free radical-induced damage to DNA: mechanisms and measurement, *Free Radical Biol. Med.* **32**, 1102–1115.

Gardner, P.R. (2002) Aconitase: sensitive target and measure of superoxide, *Methods Enzymol.* **349**, 9–23.

Giorgini, E., Tommasi, G., Stipa, P., Tosi, G., Littarru, G., Greci, L. (2001) Reactivity of ubiquinones and ubiquinols with free radicals, *Free Radical Res.* **35**, 63–72.

Halliwell, B. (2002) Hypothesis, proteasomal dysfunction: a primary event in neurogeneration that leads to nitrative and oxidative stress and subsequent cell death, *Ann. N.Y. Acad. Sci.* **962**, 182–194.

Halliwell, B., Clement, M.V., Long, L.H. (2000) Hydrogen peroxide in the human body, *FEBS Lett.* **486**, 10–13.

Halliwell, B., Kaur, H. (1997) Hydroxylation of salicylate and phenylalanine as assays for hydroxyl radicals: a cautionary note visited for the third time, *Free Radical Res.* **27**, 239–244.

Halliwell, B., Zhao, K., Whiteman, M. (1999) Nitric oxide and peroxynitrite. The ugly, the uglier and the not so good: a personal view of recent controversies, *Free Radical Res.* **31**, 651–669.

Ho, Y.S., Magnenat, J.L., Gargano, M., Cao, J. (1998) The nature of antioxidant defense mechanisms: a lesson from transgenic studies, *Environ Health Perspect.* **5**, 1219–1228.

Huang, T.T., Raineri, I., Eggerding, F., Epstein, C.J. (2002) Transgenic and mutant mice for oxygen free radical studies, *Methods Enzymol.* **349**, 191–213.

Kaur, H., Halliwell, B. (1990) Action of biologically-relevant oxidizing species upon uric acid. Identification of uric acid oxidation products, *Chem.-Biol. Interact.* **73**, 235–247.

Kuhn, H., Borchert, A. (2002) Regulation of enzymatic lipid peroxidation: the interplay of peroxidizing and peroxide reducing enzymes, *Free Radical Biol. Med.* **33**, 154–172.

Lee, M.H., Hyun, D.H., Jenner, P., Halliwell, B. (2001) Effect of overexpression of wild-type and mutant Cu/Zn-superoxide dismutases on oxidative damage and antioxidant defences: relevance to Down's syndrome and familial amyotrophic lateral sclerosis, *J. Neurochem.* **76**, 957–965.

Maier, C.M., Chan, P.H. (2002) Role of superoxide dismutases in oxidative damage and neurodegenerative disorders, *Neuroscientist* **8**, 323–334.

Meier, B. (2001) Superoxide generation of phagocytes and nonphagocytic cells, *Protoplasma* **217**, 117–124.

Melov, S., Schneider, J.A., Day, B.J., Hinderfeld, D., Coskun, P., Mirra, S.S., Crapo, J.D., Wallace, D.C. (1998) A novel neurological phenotype in mice lacking mitochondrial manganese superoxide dismutase, *Nat. Genet.* **18**, 99–100.

Meneshian, A., Bulkley, G.B. (2002) The physiology of endothelial xanthine oxidase: from urate catabolism to reperfusion

injury to inflammatory signal transduction, *Microcirculation* **9**, 161–175.

Nishikawa, T., Edelstein, D., Du, X.L., Yamagishi, S.-I., Matsumura, T., Kaneda, Y., Yorek, M.A., Beede, D., Oates, P.J., Hammes, H.-P., Giardino, I., Brownlee, M. (2000) Normalizing mitochondrial superoxide production blocks three pathways of hyperglycaemic damage, *Nature* **404**, 787–790.

Okado-Matsumoto, A., Fridovich, I. (2001) Subcellular distribution of superoxide dismutases (SOD) in rat liver: Cu, Zn-SOD in mitochondria, *J. Biol. Chem.* **276**, 38388–38393.

Pomposiello, P.J., Demple, B. (2001) Redox-operated genetic switches: the SoxR and OxyR transcription factors, *Trends Biotechnol.* **19**, 109–114.

Pratico, D., Iuliano, L., Amerio, G., Tang, L.X., Rokach, J., Sabatino, G., Violi, F. (2000) Down's syndrome is associated with increased 8,12-iso-iPF2alpha-VI levels: evidence for enhanced lipid peroxidation *in vivo*, *Ann. Neurol.* **48**, 795–798.

Sen, C.K., Packer, L. (1996) Antioxidant and redox regulation of gene transcription, *FASEB J.* **10**(7), 709–720.

Shringarpure, R., Davies, K.J. (2002) Protein turnover by the proteasome in aging and disease, *Free Radical Biol. Med.* **32**, 1084–1089.

Tanaka, T., Nakamura, H., Nishiyama, A., Hosoi, F., Masutani, H., Wada, H., Yodoi, J. (2001) Redox regulation by thioredoxin superfamily; protection against oxidative stress and aging, *Free Radical Res.* **33**, 851–855.

Tkeshelashvili, L.K., McBride, T., Spence, K., Loeb, L.A. (1991) Mutation spectrum of copper-induced DNA damage, *J. Biol. Chem.* **266**, 6401–6406.

Touati, D. (2002) Investigating phenotypes resulting from a lack of superoxide dismutase in bacterial null mutants, *Methods Enzymol.* **349**, 145–154.

Viera, L., Ye, Y.Z., Estevez, A.G., Beckman, J.S. (1999) Immunohistochemical methods to detect nitrotyrosine, *Methods Enzymol.* **301**, 373–381.

Fruit Fly (*Drosophila*) Genome: *see Drosophila* Genome

FTIR of Biomolecules

Barbara Stuart
University of Technology, Sydney, Australia

1 Experimental Methods 653
1.1 Instrumentation 653
1.2 Transmission Methods 653
1.3 Reflectance Methods 654
1.4 Microsampling Methods 656
1.5 Temperature Studies 657

2 Spectrum Interpretation 658
2.1 Mid-infrared Region 658
2.2 Near-infrared Region 660

3 Quantitative Analysis 661
3.1 Spectrum Manipulation 661
3.2 Concentration 663
3.3 Simple Analysis 664
3.4 Complex Analysis 664

4 Applications 665
4.1 Lipids 665
4.2 Proteins and Peptides 668
4.3 Nucleic Acids 674
4.4 Disease Diagnosis 675
4.5 Microbial Cells 678
4.6 Clinical Chemistry 681

Bibliography 681
Books and Reviews 681
Primary Literature 682

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30546-7

Keywords

Attenuated Total Reflectance Spectroscopy
Spectroscopic sampling technique utilizing a beam of radiation entering a crystal, which undergoes total internal reflection; the beam penetrates a fraction of a wavelength beyond the reflecting surface and when a material that selectively absorbs radiation is in close contact with the reflecting surface, the beam loses energy at the wavelength where the material absorbs.

Diffuse Reflectance Spectroscopy
Spectroscopic sampling technique utilizing the energy that penetrates one or more particles and is reflected in all directions.

Fourier Transform Infrared Spectroscopy
Spectroscopic technique based on the interference of radiation between two beams to yield an interferogram, a signal produced as a function of the change of pathlength between the two beams; the two domains of distance and frequency are interconvertible by the mathematical method of Fourier transformation.

Group Frequencies
Bands that appear in an infrared spectrum, which may be assigned to particular parts of a molecule.

Infrared Mapping
Spectroscopic technique that involves recording spectra at computer controlled positions at a micro level, generating contour or axonometric plots.

Mid-infrared
Region observed in infrared spectroscopy in the range 4000 to 400 cm^{-1} showing bands due to mainly fundamental modes.

Near-infrared
Region observed in infrared spectroscopy in the range 13 000 to 4000 cm^{-1} showing bands due to overtones and combination vibrations.

Infrared spectroscopy has proved to be a powerful tool for the study of biological molecules and the application of this technique to biological problems is continually expanding, particularly with the advent of increasingly sophisticated sampling techniques associated with Fourier transform infrared spectroscopy in recent decades. Biological systems, including lipids, proteins, peptides, biomembranes, nucleic acids, animal tissue, microbial cells, and clinical samples, have all been successfully studied using infrared spectroscopy. Infrared spectroscopy has been

employed for a number of decades for the characterization of isolated biological molecules, particularly proteins and lipids. However, the last decade has seen a rapid rise in the number of studies of more complex systems, such as diseased tissues. Microscopic techniques combined with sophisticated analytical methods allow for complex samples of micron size to be investigated. A review of the sampling methods used for biological molecules is provided here. Examples of how infrared spectroscopy may be applied to various biological fields are also provided.

1 Experimental Methods

Infrared spectroscopy is a technique based on the vibrations of the atoms of a molecule. An infrared spectrum is obtained by passing infrared radiation through a sample and determining what fraction of the incident radiation is absorbed at a particular energy. The energy at which a peak in an absorption spectrum appears corresponds to the frequency of a vibration of a part of a sample molecule, allowing for straightforward identification.

1.1 Instrumentation

In the past, dispersive instrumentation was used to obtain infrared spectra, but this approach has been almost entirely superseded by Fourier transform infrared (FTIR) spectroscopy. FTIR spectroscopy is based on the idea of the interference of radiation between two beams to yield an interferogram. An interferogram is a signal produced as a function of the change of pathlength between the two beams. The two domains of distance and frequency are interconvertible by the mathematical method of Fourier transformation. The basic components of an FTIR spectrometer are shown schematically in Fig. 1. The radiation emerging from the source is passed through an interferometer to the sample before reaching a detector. Upon amplification of the signal, in which high-frequency contributions have been eliminated by a filter, the data are converted to a digital form by an analog-to-digital converter and transferred to the computer for Fourier transformation.

1.2 Transmission Methods

Transmission spectroscopy is the oldest and most basic infrared method. The method is based upon the absorption of infrared radiation at specific wavelengths as it passes through a sample. There are several different types of transmission solution cells available. Fixed pathlength sealed cells are useful for volatile liquids, but cannot be taken apart for cleaning. Semipermanent cells are demountable

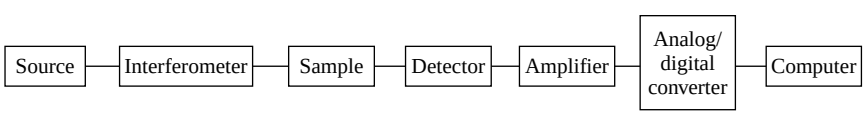

Fig. 1 Components of an FTIR spectrometer.

so that the windows can be cleaned. The spacer is usually made of polytetrafluoroethylene (PTFE or Teflon) and is available in a variety of thicknesses, allowing one cell to be used for various pathlengths. Variable pathlength cells incorporate a mechanism for continuously adjusting the pathlength and a vernier scale allows accurate adjustment. All these cell types are filled using a syringe and the syringe ports are sealed with PTFE plugs before sampling.

An important consideration in the choice of infrared cells is the type of window material. The material must be transparent to the incident infrared radiation, and normally, alkali halides are used in transmission methods. The cheapest and most widely used material is sodium chloride (NaCl). Where water is used as a solvent, particularly for biological samples, NaCl cannot be used as an infrared window material as it is very soluble in water, so calcium fluoride (CaF_2) or barium fluoride (BaF_2) are better choices. Small pathlengths (~0.010 mm) are available and help reduce the intensity of the very strong infrared modes produced in the water spectrum. The small pathlength also produces a small sample cavity, allowing samples in milligram quantities to be examined. The infrared modes of water are very intense and may overlap with the sample modes of interest, but this problem may be overcome by substituting water with deuterium oxide (D_2O). The infrared modes of D_2O occur at different frequencies to those observed for water because of the mass dependence of the vibrational frequency.

1.3 Reflectance Methods

Reflectance techniques may be used for samples that are difficult to analyze by the normal transmittance methods. Attenuated total reflectance spectroscopy (ATR) utilizes the phenomenon of total internal reflection (Fig. 2). A beam of radiation entering a crystal will undergo total internal reflection when the angle of incidence at the interface between the sample and crystal is greater than the critical angle. The critical angle is a function of the refractive indices of the two surfaces. The beam penetrates a fraction of a wavelength beyond the reflecting surface and when a material that selectively absorbs radiation is in close contact with the reflecting surface,

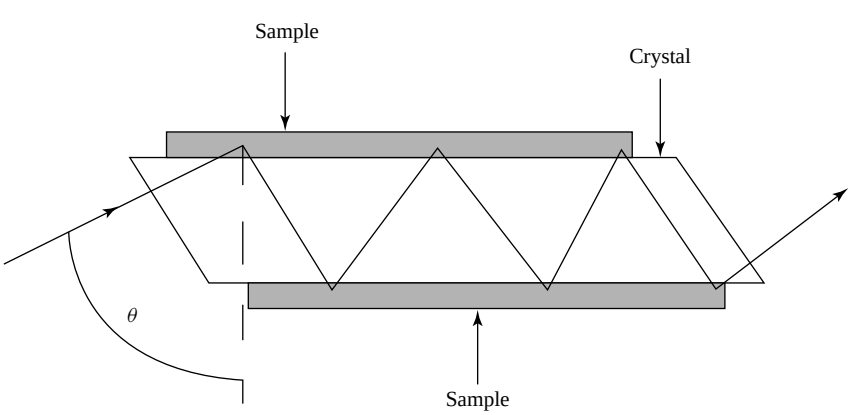

Fig. 2 Attenuated total reflectance spectroscopy.

the beam loses energy at the wavelength where the material absorbs. The resultant attenuated radiation is measured and plotted as a function of wavelength by the spectrometer and gives rise to the absorption spectral characteristics of the sample.

The depth of penetration in ATR is a function of wavelength (λ), the refractive index of the crystal and the angle of incident radiation (θ). The depth of penetration, d_p, for a nonabsorbing medium is given by the following formula:

$$d_p = \frac{\lambda / n_1}{2\pi[\sin\theta - (n_1/n_2)^2]^{1/2}} \quad (1)$$

where n_1 is the refractive index of the sample and n_2 is the refractive index of the ATR crystal. The crystals used in ATR cells are made from materials that have low solubility in water and are of very high refractive index, such as zinc selenide (ZnSe), germanium (Ge), and thallium/iodide (KRS-5). Different designs of ATR cells allow both liquid and solid samples to be examined. It is also possible to set up a flow-through ATR cell by including an inlet and outlet into the apparatus. This allows for the continuous flow of solutions through the cell and permits the spectral changes with time to be monitored. Multiple internal reflectance (MIR) and ATR are similar techniques, but MIR produces more intense spectra from multiple reflections. While a prism is usually used in ATR work, MIR uses specially shaped crystals that cause many internal reflections, typically 25 or more.

Specular reflectance occurs when the reflected angle of incident radiation equals the angle of incidence. The amount of light reflected depends on the angle of incidence, the refractive index, surface roughness, and absorption properties of the sample. For most materials, the reflected energy is only 5 to 10%, but in regions of strong absorptions, the reflected intensity is greater. The resultant data appears different from normal transmission spectra, as derivative-like bands result from the superposition of the normal extinction coefficient spectrum with the refractive index dispersion (based upon Fresnel's relations from physics). However, the reflectance spectrum can be corrected using a Kramers–Kronig transformation (or K–K transformation) and the corrected spectrum appears like the familiar transmission spectrum.

Increased pathlengths through thin coatings can be achieved using grazing angles of incidence (up to 85°). Grazing angle sampling accessories allow measurements to be made on samples over a wide range of angles of incidence. Solid samples, particularly coatings on reflective surfaces, are simply placed on a flat surface. The technique is also commonly used for liquid samples that can be poured into a Teflon trough. Oriented films on the liquid surface can be investigated using this material.

In external reflectance, the energy that penetrates one or more particles is reflected in all directions and this component is called *diffuse reflectance*. In the diffuse reflectance technique, commonly called *DRIFT*, a powdered sample is mixed with KBr powder. The DRIFT cell reflects radiation to the powder and collects the energy reflected back over a large angle. Diffusely scattered light can be collected directly from a sample or, alternatively, by using an abrasive sampling pad. DRIFT is particularly useful for sampling powders or fibers. Figure 3 shows diffuse reflectance from the surface of a sample. Kubelka and Munk developed a theory describing the diffuse reflectance process for powdered samples that relates the sample concentration to the scattered radiation intensity.

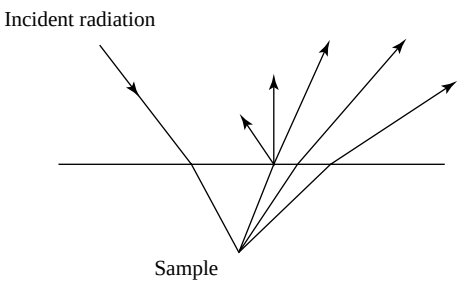

Fig. 3 Diffuse reflectance.

The Kubelka–Munk equation is

$$\frac{1-R^2}{2R} = \frac{c}{k} \tag{2}$$

where R is the absolute reflectance of the layer, c is the concentration and k is the molar absorption coefficient. An alternative relationship between the concentration and the reflected intensity is now widely used in near-infrared diffuse reflectance spectroscopy. The relationship is analogous to the Beer–Lambert Law:

$$\log\left(\frac{1}{R}\right) = k'c \tag{3}$$

where k' is a constant.

1.4 Microsampling Methods

It is possible to combine an infrared spectrometer with a microscope facility in order to study very small samples. In recent years, there have been considerable advances in FTIR microscopy with samples of the order of microns being characterized. In FTIR microscopy, the microscope sits above the FTIR sampling compartment. Infrared radiation from the spectrometer is focused onto a sample placed on a standard microscope $x-y$ stage. After passing through the sample, the infrared beam is collected by a Cassegrain objective that produces an image of the sample within the barrel of the microscope. A variable aperture is placed in this image plane. The radiation is then focused on a small area mercury cadmium telluride (MCT) detector by another Cassegrain condenser. The microscope also contains glass objectives to allow visual inspection of the sample. In addition, by switching mirrors in the optical train, the microscope can be converted from transmission mode to reflectance mode.

If a microscope facility is not available, there are other special sampling accessories available to allow examination of microgram or microliter amounts. This is accomplished using a beam condenser so that as much as possible of the beam passes through the sample. Microcells are available with volumes of around 4 μL and pathlengths up to 1 mm. A diamond anvil cell (DAC) uses two diamonds to compress a sample to a thickness suitable for measurement and increase the surface area. This technique can be used at normal atmospheric pressures, but it may also be applied to study samples under high pressures and improve the quality of the spectrum of trace samples. Alternatively, a multiple internal reflectance cell may also be used as this technique can produce stronger spectra.

Infrared imaging using FTIR microspectroscopic techniques has emerged as an effective approach to studying complex

or heterogeneous specimens. The technique can be used to produce a two- or three-dimensional picture of the properties of a sample. This is possible because, instead of reading the signal of only one detector as in conventional FTIR spectroscopy, a large number of detector elements are read during the acquisition of spectra. This is possible due to the development of focal plane array (FPA) detectors. Currently, a step-scanning approach is used, which means that the moving mirror does not move continuously during data acquisition, but waits for each detector readout to be completed before moving onto the next position. This allows thousands of interferograms to be collected simultaneously and then transformed into infrared spectra. In an FTIR imaging microspectrometer, the infrared beam from a Michelson interferometer is focused onto a sample with a reflective Cassegrain condenser. The light transmitted is collected by a Cassegrain objective and focused onto an FPA detector. The imaging process is illustrated by Fig. 4. The data are collected as interferograms with each pixel on the array having a response determined by its corresponding location on the sample. Each point of the interferogram represents a particular moving mirror position and the spectral data is obtained by performing a Fourier transform for each pixel on the array. Thus, each pixel (or spatial location) is represented by an infrared spectrum.

1.5 Temperature Studies

Variable temperature cells that are controlled to 0.1 °C in the range −180 to 250 °C can be obtained. An electrical heating system is used for temperatures above ambient, and liquid nitrogen with a heater is used for low temperatures. These cells may be used to study phase transitions and the kinetics of reactions. As well as transmission temperature cells, variable temperature ATR cells and cells for microsampling are available.

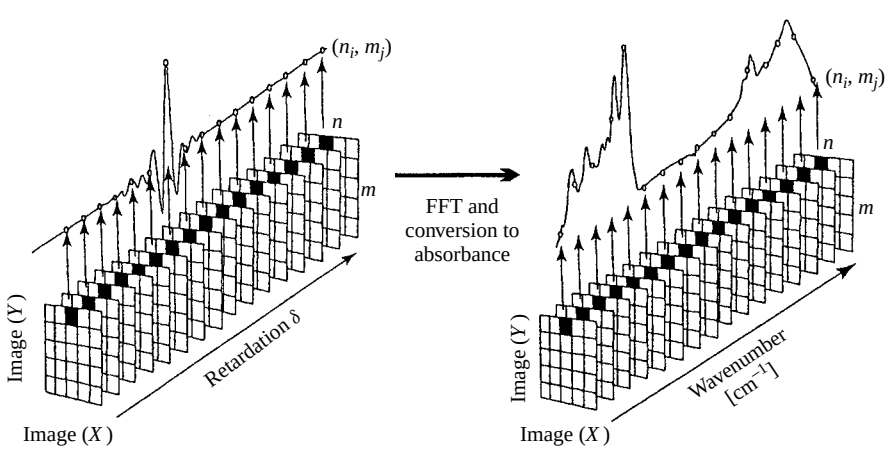

Fig. 4 FTIR imaging data set. (From Kidder, L.H., Levin, I.W., and Lewis, E.N. (1998) "Infrared Spectroscopic Imaging Microscopy: Applications to Biological Systems," in: de Haseth, J.A. (Ed.) *Proceedings of the 11th International Fourier Transform Spectroscopy Conference*, American Institute of Physics, p. 148.)

2 Spectrum Interpretation

The interpretation of an infrared spectrum is simplified by the fact that the bands that appear may be assigned to particular parts of a molecule, producing what are known as group frequencies. The infrared spectrum may be divided into three regions: the far-infrared (<400 cm^{-1}), the mid-infrared (4000–400 cm^{-1}) and the near-infrared (13 000–4000 cm^{-1}). The most useful parts of an infrared spectrum for the study of biological systems are the mid- and near-infrared regions.

2.1 Mid-infrared Region

The mid-infrared spectrum can be approximately divided into four regions and the nature of a group frequency may generally be determined by the region in which it is located. These regions are defined as follows: X−H stretching region (4000–2500 cm^{-1}); triple bond region (2500–2000 cm^{-1}); double-bond region (2000–1500 cm^{-1}); and fingerprint region (1500–600 cm^{-1}).

All the fundamental vibrations in the region 4000 to 2500 cm^{-1} may be attributed to X−H stretching. O−H stretching produces a broad band that occurs in the range 3700 to 3600 cm^{-1}. It should be noted that hydrogen bonding in a molecule has a significant effect on bond frequencies, such as those observed for the O−H bond. The O−H stretching vibration in a hydrogen-bonded dimer is observed in the range 3500 to 2500 cm^{-1}, rather than the usual range 3700 to 3600 cm^{-1}. N−H stretching is usually observed between 3400 and 3300 cm^{-1}. This absorption is generally much sharper than O−H stretching and may therefore be easily differentiated. C−H stretching bands from aliphatic compounds occur in the range 3000 to 2850 cm^{-1}. They are moderately broad and show medium intensity. It is also possible to resolve asymmetrical and symmetrical C−H stretching absorptions of a CH_3 group, which usually occur at about 2965 and 2880 cm^{-1} respectively, while the corresponding absorptions for a CH_2 group occur at 2930 and 2860 cm^{-1}. If the C−H bond is adjacent to a double bond or an aromatic ring, the C−H stretching frequency increases and absorbs between 3100 and 3000 cm^{-1}.

Triple bond stretching absorptions fall in the 2500 to 2000 cm^{-1} region because of the high force constants of the bonds. C≡C bonds absorb between 2300 and 2050 cm^{-1}, while the nitrile group (C≡N) occurs between 2300 and 2200 cm^{-1}. These groups may be distinguished since C≡C stretching is normally very weak, while C≡N stretching is of medium intensity. These are the most common absorptions in this region, but some X−H stretching absorptions, where X is a more massive atom like phosphorus or silicon may be observed. These absorptions usually occur near 2400 and 2200 cm^{-1} respectively.

The principal bands in the double-bond region are due to C=C and C=O stretching. Carbonyl stretching is one of the easiest absorptions to recognize in an infrared spectrum as it is usually the most intense band in the spectrum, and depending on the type of C=O bond, occurs in the region 1830 to 1650 cm^{-1}. Metal carbonyls may absorb above 2000 cm^{-1}. C=C stretching is much weaker and occurs at around 1650 cm^{-1}, but this band is often absent for symmetry or dipole moment reasons.

C=N stretching also occurs in this region and is usually stronger. The N−H bending vibration in amines occurs between 1630 and 1500 cm^{-1} and is usually strong.

At wave numbers of values greater than 1500 cm^{-1}, it is generally possible to assign each absorption band in an infrared spectrum, but this is not so for most absorptions observed below this value. This region is referred to as the fingerprint region, since quite similar molecules give distinctly different absorption patterns at these wave numbers. It has been assumed thus far that each band in an infrared spectrum can be assigned to a particular deformation of the molecule, the movement of a group of atoms or the bending or stretching of a particular bond. This is possible for many bands, particularly the stretching vibrations of multiple bonds that are well behaved. However, certain vibrations are not so well behaved and may vary by hundreds of wave numbers, even for similar molecules. This applies to most bending and skeletal vibrations, which absorb in the 1500 to 650 cm^{-1} region, for which small steric or electronic effects in the molecule lead to large shifts. A spectrum of a molecule may have a hundred or more absorption bands present; however, there is no need to assign the vast majority. The spectrum can be regarded as a fingerprint of the molecule.

Most single bonds absorb at similar frequencies and, hence, the vibrations couple, that is, the vibrations in the skeletons of molecules are not restricted to one or two bonds, but may involve a large part of the backbone. The observed pattern will depend on the carbon skeleton, and the resulting bands will originate from the oscillation of large parts of the skeleton, or the skeleton and the attached functional groups. C−C stretching frequencies may also couple with C−H bending vibrations. The C−O stretching frequency is one of the bands that can be useful for identification purposes. If no intense band appears in the fingerprint region, one may be sure that no C−O bonds are present. The frequency is rather variable, occurring anywhere between 1400 and 1000 cm^{-1}. Aromatic rings give rise to two bands at 1600 and 1500 cm^{-1}. They are usually sharp, but are of variable intensity, and occasionally, the band at 1600 cm^{-1} splits into a doublet. Aromatic rings and alkenes give rise to other bands that are perhaps the most useful in this region. They are out-of-plane C−H bending vibrations, which occur between 1000 and 700 cm^{-1}.

The information provided by group frequencies can be summarized in what are known as *correlation tables*. Figure 5 shows a correlation table for organic molecules. The lines in the correlation table represent the regions of the spectrum where bands appear for the relevant class of molecule. Advances in computer retrieval techniques have extended the range of information available from an infrared instrument by allowing comparison of an unknown spectrum with a bank of known compounds. This was always possible in the past by manually searching libraries of spectra. Programs widely available work by hunting through stored data to match intensities and frequencies of absorption bands. The computer will then output the best fits and the names of the compounds found. There is a wide range of spectral libraries of particular interest to life scientists. For instance, libraries for the following types of samples are currently available: proteins, peptides, enzymes, sugars, carbohydrates, fatty acids, and glycerides.

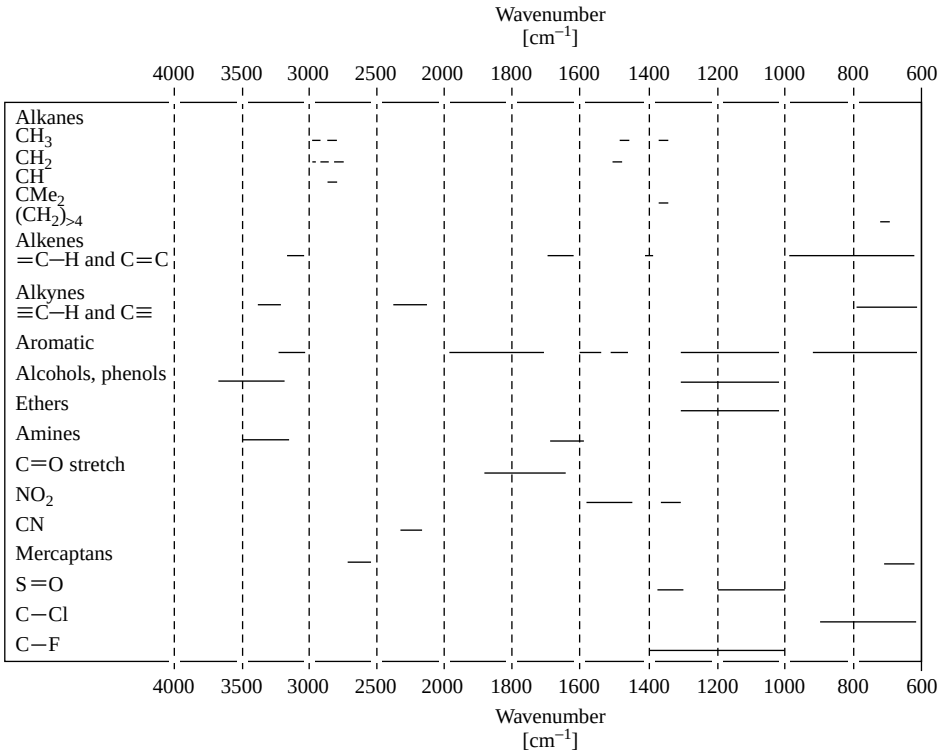

Fig. 5 Correlation table for organic molecules.

2.2
Near-infrared Region

The absorptions observed in the near-infrared region (13 000–4000 cm^{-1}) are overtones or combinations of the fundamental stretching bands that occur in the region 3000 to 1700 cm^{-1}. Overtone bands in an infrared spectrum are multiples of the fundamental absorption frequency. The energy required for the first overtone is twice the fundamental, assuming evenly spaced energy levels. Since the energy is proportional to the frequency absorbed and this is proportional to the wave number, the first overtone will appear in the spectrum at twice the wave number of the fundamental. Combination bands arise when two fundamental bands absorbing at $\overline{\nu}_1$ and $\overline{\nu}_2$ absorb energy simultaneously. The resulting band will appear at $(\overline{\nu}_1 + \overline{\nu}_2)$ wave numbers. The bands involved are usually due to C—H, N—H, O—H, P—H, or S—H stretching. Table 1 summarizes the near-infrared bands commonly found in biological molecules.

The resulting bands in the near-infrared are usually weak in intensity and the intensity generally decreases by a factor of 10 from one overtone to the next. The bands in the near-infrared are also often overlapped, making them less useful than the mid-infrared region for qualitative analysis. However, there are important differences between the near-infrared positions between functional groups and these differences can often be exploited

Tab. 1 Common near-infrared bands.

Wave number [nm]	*Assignment*
2200–2450	Combination C—H stretching
2000–2200	Combination N—H stretching, combination O—H stretching
1650–1800	First overtone C—H stretching
1400–1500	First overtone N—H stretching, first overtone O—H stretching
1300–1420	Combination C—H stretching
1100–1225	Second overtone C—H stretching
950–1100	Second overtone N—H stretching, second overtone O—H stretching
850–950	Third overtone C—H stretching
775–850	Third overtone N—H stretching

for quantitative analysis using multivariate analytical techniques.

3 Quantitative Analysis

3.1 Spectrum Manipulation

There are a number of techniques available to users of infrared spectrometers that help with both qualitative and quantitative interpretation of spectra. It is usual in quantitative infrared spectroscopy to use a baseline joining the points of lowest absorbance on the peak, preferably in reproducibly flat parts of the absorption line. The absorbance difference between the baseline and the top of the band is then used.

Noise, in a spectrum, may be diminished by smoothing. After a spectrum is smoothed, it becomes similar to the result of an experiment at a lower resolution. The features are blended into each other and the noise level decreases. A smoothing function is basically a convolution between the spectrum and a vector whose points are determined by the degree of smoothing that is applied.

A straightforward method of analysis for complex spectra is difference spectroscopy. Difference spectroscopy can be carried out by subtracting the infrared spectrum of one component of the system from the combined spectrum to leave the spectrum of the other component. If the interaction between components results in a change in the structural properties of either one or both of the components, the changes will be observed in the difference spectra. The changes may manifest themselves via the appearance of positive or negative peaks in the spectrum. Spectral subtraction may be applied to numerous applications and may be used for the interpretation of the data collected for solutions. In order to obtain the spectrum of a solution, it is necessary to record spectra of both the solution and the solvent alone. The solvent spectrum may then be subtracted from the solution spectrum. The concentration of the solvent alone is greater in the solvent in the solution and negative peaks may appear in the regions of solvent absorption. In certain circumstances, the spectrum due to the solvent may be very intense, making simple subtraction impossible. This situation can make it difficult to investigate the sample spectrum as solvent bands may overlap with the region under investigation. In such experiments, ATR provides a suitable substitute. The nature of the ATR technique produces a less intense solvent contribution to the overall infrared spectrum and so solvent spectra can be more readily subtracted from the sample spectrum of interest.

Spectra may also be differentiated. The benefits of the technique are two-fold: resolution is enhanced in the first derivative

since changes in gradient are observed; the second derivative gives a negative peak for each band and shoulder in the absorption spectrum. Differentiation can be used to resolve and locate peaks in an envelope. Sharp bands are enhanced at the expense of broad ones and this may allow selection of a peak, even when there is a broad band beneath. With modern FTIR spectrometers, it is possible to apply what is known as *Fourier derivation*. During this process, the spectrum is transformed to an interferogram. It is then multiplied by an appropriate weighting function and finally it is retransformed to give the derivative. This technique provides more sensitivity. A derivative spectrum of a peptide is illustrated in Fig. 6.

Deconvolution is the process of compensating for the intrinsic linewidths of modes in order to resolve overlapping bands. The technique yields spectra that have much narrower bands and is able to distinguish closely spaced features. The instrumental resolution is not increased, but the ability to differentiate spectral features can be significantly improved. This is illustrated by Fig. 6, which shows a broad band of a peptide before and after deconvolution has been applied. Peaks at quite close frequencies are now easily distinguished. The deconvolution technique generally involves several steps: computation of an interferogram of the sample by computing the inverse Fourier transform of the spectrum; multiplication of the interferogram

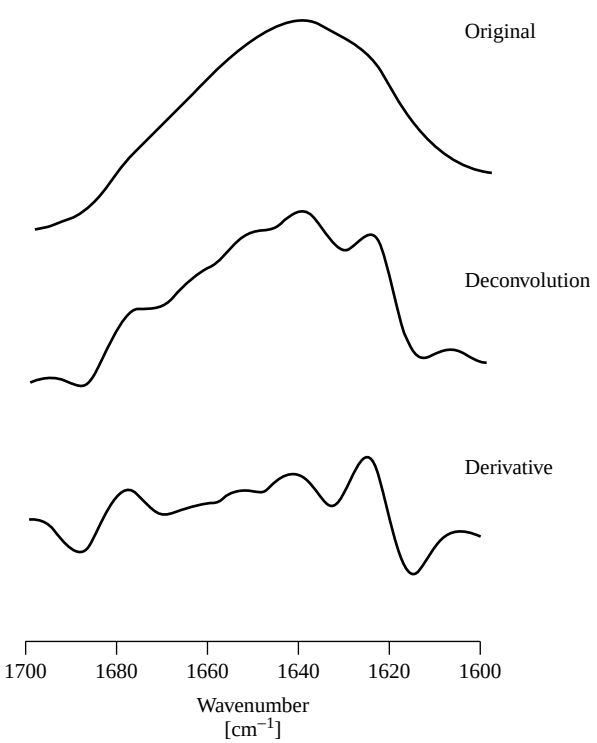

Fig. 6 Original, deconvolved, and derivative spectra of a peptide.

by a smoothing function and by a function consisting of a Gaussian–Lorentzian bandshape; and Fourier transformation of the modified interferogram. The deconvolution procedure is typically repeated iteratively for best results. At iteration, the lineshape is adjusted in an attempt to provide narrower bands without excessive distortion. There are three parameters that can be adjusted to tune the lineshape: the proportion of Gaussian and Lorentzian lineshape; the half-width of the line shape; and the narrowing function (degree of narrowing attempted on a scale 0–1).

Quantitative values for band areas of heavily overlapped bands can be achieved by using curve-fitting procedures. Many curve-fitting procedures are based on a least-squares minimization procedure. Generally, the procedure involves entering the values of the frequencies of the component bands (determined using derivatives and/or deconvolution) and then the program determines the best estimate of the parameters of the component curves. Apart from the obvious variables of peak height and width, the type of band shape needs to be considered. The class of band shape of an infrared spectrum depends on the type of sample. A choice of Gaussian, Lorentzian, or a combination of these band shapes is usually applied. Figure 7 provides an example of the curve-fitting process.

3.2 Concentration

The Beer–Lambert Law may be applied to relate light absorbance or transmittance to concentration in infrared spectroscopy. The absorbance of a solution is directly proportional to the thickness and the concentration of the sample according to the relationship:

$$A = \varepsilon c l \tag{4}$$

where A is the absorbance of the solution, c is the concentration, l the pathlength of the sample and ε is the molar absorptivity. The absorbance is equal to the difference between the logarithms of the intensity of the light entering the sample (I_0) and the intensity of the light transmitted (I) by the sample

$$A = \log_{10}\left(\frac{I_0}{I}\right) \tag{5}$$

Transmittance is defined as $T = I/I_0$, and % transmittance as $\% T = 100 \times T$.

The Beer–Lambert Law tells us that a plot of absorbance against concentration should be linear with a gradient of εl and pass through the origin. In theory, to analyze a solution of unknown

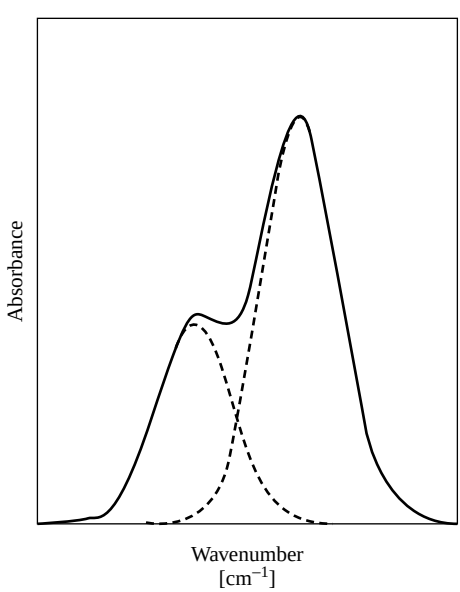

Fig. 7 Curve fitting of overlapping infrared bands.

concentration, solutions of known concentration should be prepared, a suitable peak chosen, the absorbance at this frequency measured, and a calibration graph plotted. The concentration of the compound in solution can be determined once its absorbance value is determined.

3.3 Simple Analysis

The quantitative analysis of a component in solution can be successfully carried out given that there is a suitable band in the spectrum of the component of interest. The band chosen for analysis should have a high molar absorptivity; not overlap with other peaks from other components in the mixture or solvent; be symmetrical; and give a linear calibration plot of absorbance versus concentration. Many quantitative infrared methods of analysis use the intensities of the C=O, N−H or O−H groups. The C=O stretching band is the most commonly used because it is a strong band in a spectral region relatively free of absorption by other functional groups and is not as susceptible as the O−H and N−H bands to chemical change or hydrogen bonding.

Solid mixtures can also be quantitatively analyzed using infrared spectroscopy. Solids are more susceptible to errors because of scattering problems. These analyses may be carried out with KBr discs or in mulls with an internal standard. The calibration curve is obtained by plotting the ratio of the absorbance of the analyte to that of the internal standard, against the concentration of the analyte. The absorbance of the internal standard varies linearly with the sample thickness.

3.4 Complex Analysis

The improvement in computer technology associated with spectroscopy has led to the expansion of quantitative infrared spectroscopy. The application of statistical methods to the analysis of experimental data is known as *chemometrics*. A detailed description of the subject is beyond the scope of this article, but several multivariate data analytical methods that are used for the analysis of FTIR data are summarized here, without detailing the mathematics associated with these methods. The most commonly used analytical methods in infrared spectroscopy are classical least squares (CLS); inverse least squares (ILS); partial least squares (PLS); and principal component regression (PCR). CLS (also known as *K-matrix methods*) and PLS (also known as *P-matrix methods*) are least-squares methods involving matrix operations. These methods can be limited when very complex mixtures are investigated and factor analysis methods such as PLS and PCR can be more useful. The factor analysis methods use functions to model the variance in a data set.

If it is not necessary to know the specific concentration of the species of interest, but to simply know whether the species is present or not in a complex sample, then a multivariate pattern recognition method, such as linear discriminant analysis (LDA) or artificial neural networks (ANNs), may be used to identify the spectral characteristics of the species of interest. Such methods are capable of comparing a large number of variables within a data set, such as intensity, frequency, and bandwidth. LDA and ANNs are known as *supervised methods* because *a priori* information is available about the data set. There are also unsupervised methods,

such as hierarchical clustering, which may be used to determine the components in a data set without any prior information about the data.

4 Applications

4.1 Lipids

Many lipids contain phosphorus and are classed as phospholipids. The infrared spectra of phospholipids can be divided into the spectral regions that originate from the molecular vibrations of the hydrocarbon tail, the interface region, and the head-group. The major infrared modes due to phospholipids are summarized in Table 2. The hydrocarbon tail gives rises to acyl chain modes. The most intense vibrations in the infrared spectra of lipid systems are the CH_2 stretching vibrations and these give rise to bands in the region 3100 to 2800 cm^{-1}. The CH_2 asymmetric and symmetric stretching modes at 2920 and 2851 cm^{-1} respectively, are generally the strongest bands in the spectra. The frequencies of these bands are conformation-sensitive and respond to changes of the trans/*gauche* ratio in the acyl chains. This is also the case for the vibrational modes due to the terminal CH_3 groups at 2956 cm^{-1} (asymmetric stretching) and 2873 cm^{-1} (symmetric stretching). The =C−H stretching bands due to unsaturated acyl chains are found at 3012 cm^{-1} and the bands due to methylene and methyl groups occur in the 1500 to 1350 cm^{-1} region. At around 1470 cm^{-1}, there are bands due to CH_2 bending and the number and frequency of these bands are dependent on acyl chain packing and conformation. While the asymmetric deformation modes of the CH_3 group are obscured by the scissoring bands, the symmetric deformation mode appears at 1378 cm^{-1}.

In certain phospholipid membranes that contain unsaturated acyl chains, the typical lamellar liquid crystalline phase converts to a micellar nonlamellar phase upon heating. Such a thermally induced transition involves a major structural rearrangement. Temperature studies of the infrared spectra of phospholipids provide a sensitive means of studying such transitions in lipids. Figure 8 shows the temperature dependence of the frequency of the symmetric CH_2 stretching band in the spectra of lipid membranes obtained from phosphatidylethanolamine. The increasing frequency with temperature indicates an increasing concentration of *gauche* bands

Tab. 2 Characteristic infrared bands of lipids.

Wave number [cm^{-1}]	*Assignment*
3010	=C−H stretching
2956	CH_3 asymmetric stretching
2920	CH_2 asymmetric stretching
2870	CH_3 symmetric stretching
2850	CH_2 symmetric stretching
1730	C=O stretching
1485	$(CH_3)_3N^+$ asymmetric bending
1473, 1472, 1468, 1463	CH_2 scissoring
1460	CH_3 asymmetric bending
1405	$(CH_3)_3N^+$ symmetric bending
1378	CH_3 symmetric bending
1400–1200	CH_2 wagging band progression
1228	PO_2^- asymmetric stretching
1170	CO−O−C asymmetric stretching
1085	PO_2^- symmetric stretching
1070	CO−O−C symmetric stretching
1047	C−O−P stretching
972	$(CH_3)_3N^+$ asymmetric stretching
820	P−O asymmetric stretching
730, 720, 718	CH_2 rocking

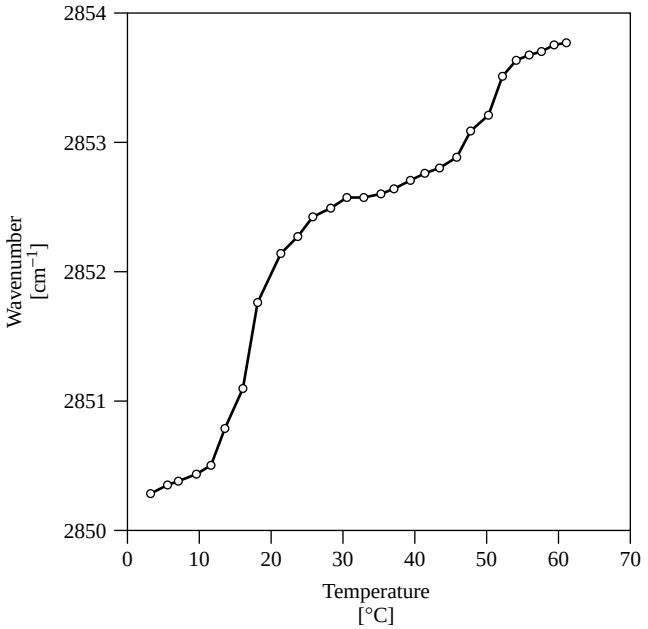

Fig. 8 Temperature dependence of the symmetric CH_2 stretching band of phosphatidylethanolamine.

in the acyl chains and this leads to the formation of the nonbilayer phase at higher temperatures. Figure 8 shows a frequency shift of about 2 cm^{-1} at 18 °C and this is associated with the gel to liquid crystal phase transition. An additional frequency shift of approximately 1 cm^{-1} at 50 °C is associated with a transition to the micellar phase. Both these transitions have been observed to be reversible.

Spectral modes arising from the headgroup and interfacial region also provide useful information. The most useful infrared bands for studying the interfacial region of lipid assemblies are the ester group vibrations, particularly the C=O stretching bands in the 1750 to 1700 cm^{-1} region. In diacyl lipids, this region consists of at least two bands originating from the two ester carbonyl groups. A band at 1742 cm^{-1} is assigned to the C=O mode of the first alkyl chain with a trans conformation in the carbon–carbon bond adjacent to the ester grouping, while the 1728-cm^{-1} C=O frequency of the second alkyl chain suggests the presence of a *gauche* band in that position. The frequency difference observed reflects the structural inequivalence of the chains, with the first alkyl chain initially extending in a direction perpendicular to the second alkyl chain and then developing a *gauche* bend in order to render the two chains parallel.

The choline, ethanolamine, and serine groups of lipids have characteristic modes. The CH_3 asymmetric stretching modes of the $(CH_3)_3N^+$ group are around 3040 cm^{-1} and are out of the range of the acyl chain CH_3 bands. The corresponding methyl bending vibrations of the choline group occur at higher frequencies, compared to those of the acyl chain methyl groups. The C–N stretching bands are found between 1040 and 800 cm^{-1}. There

are also several characteristic phosphate group vibrations: an asymmetric PO_2^- stretch around 1228 cm^{-1}; a symmetric PO_2^- stretch around 1085 cm^{-1}; P—O stretching modes in the region 900 to 800 cm^{-1}.

Adipocere is the waxy substance that can form from the tissue of dead bodies. It is known to consist mainly of fatty acids, as well as the calcium salts of certain fatty acids. During the formation of adipocere, triacylglycerols breakdown into fatty acids. Figure 9 illustrates the infrared spectrum of a sample of adipocere collected from grave soil. The spectrum shows the characteristic modes attributable to fatty acids and hydroxy fatty acids. Fatty acids show a distinctive carboxylic acid carbonyl stretching mode near 1700 cm^{-1}. Comparison with reference spectra of fatty acids indicates that the predominant fatty acid present in the sample is palmitic acid, although there is evidence that both myristic acids and stearic acid are also present in the sample. The fingerprint regions of each of the fatty acids can be used to differentiate these molecules. The carbonyl region 1800 to 1700 cm^{-1} of the sample shows several overlapping modes in the region where fatty acids are known to produce modes, supporting the observation that several fatty acids are present. The appearance of modes in the region 1600 to 1500 cm^{-1} is characteristic of calcium salts of fatty acids. These salts show carboxylate CO stretching modes near 1575 and 1540 cm^{-1}. In addition to overlapping modes due to fatty acids, there is a shoulder observed near 1740 cm^{-1}. This is a region associated with the presence of triacylglycerols. There is also a mode near 3010 cm^{-1}, which may be

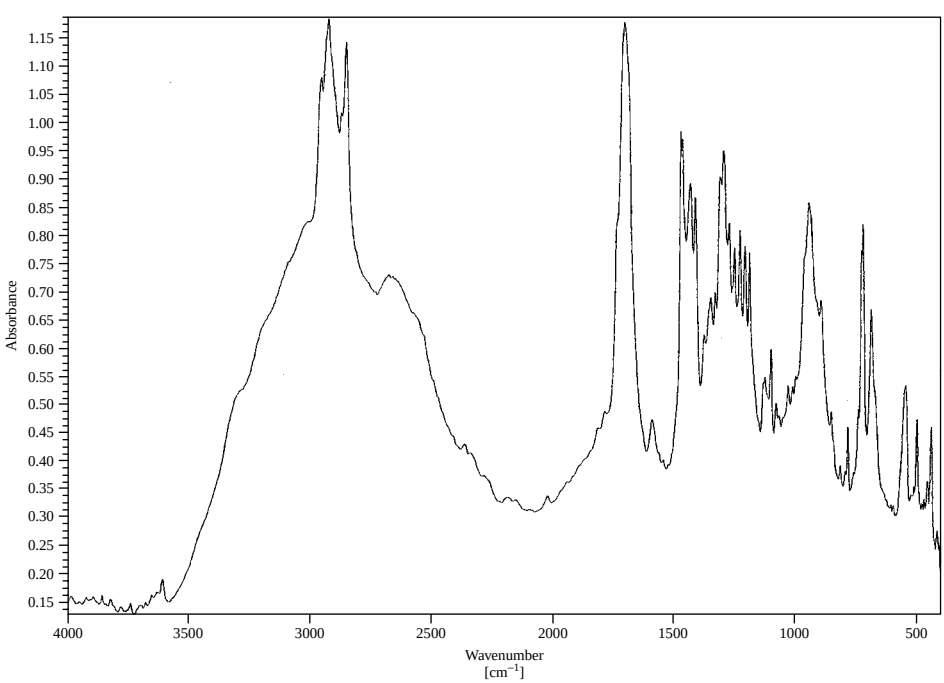

Fig. 9 Infrared spectrum of adipocere.

attributed to olefinic C—H stretching of triacylglycerols providing more evidence that triacylglycerols are still present in the adipocere. The major infrared modes of the adipocere sample are summarized in Table 3.

Quantitative infrared analysis can be carried out on blood serum to determine the relative amounts of lipid present. Triglycerides, phospholipids, and cholesteryl esters are the classes of lipid that occur in blood serum and these compounds occur naturally in concentrations that make infrared analysis attractive. These classes of compounds can be characterized by their carbonyl bands: the peak maxima appear at 1742 cm^{-1} for the triglycerides, at 1737 cm^{-1} for the phospholipid, and at 1723 cm^{-1} for the cholesteryl esters. However, the carbonyl peaks are heavily overlapped, but a least-squares method may be used to separate the components. The concentrations of these lipid components are usually in the range 0.03 to 0.3% in human blood, and standard solutions can be prepared in chloroform.

Tab. 3 Major infrared modes of adipocere.

Wave number [cm^{-1}]	*Assignment*
2950–2800	C—H stretching
2670–2660	Fatty acid O—H stretching
1720–1710	Triacylglycerol C=O stretching
1702	Fatty acid C=O stretching
1576	Fatty acid calcium salt carboxylate C—O stretching
1540	Fatty acid calcium salt carboxylate C—O stretching
1500–1400	Lipid CH_2 scissoring
1400–1200	Lipid CH_2 wagging
1200–1000	Lipid CO—O—C stretching
1000–800	C—H bending
800–700	Lipid CH_2 rocking

4.2 Proteins and Peptides

The infrared spectra of proteins and peptides exhibit absorption bands associated with their characteristic amide group. In-plane modes are due to C=O stretching, C—N stretching, N—H stretching, and O—C—N bending, while an out-of-plane mode is due to C—N torsion. The characteristic bands of the amide groups of protein chains are similar to the absorption bands exhibited by secondary amides in general, and are labeled *amide bands*. There are nine such bands, called amide A, amide B, and amide I-VII, in order of decreasing frequency and these bands are summarized in Table 4. Some of these bands are more useful for conformation studies than others and the amide I and amide II bands have been the most frequently used.

The amide II band mainly represents (60%) N—H bending, with some C—N stretching (40%). It is possible to split the amide II band into components depending on the secondary structure of the protein. The position of the amide II band is sensitive to deuteration, shifting from around 1550 cm^{-1} to a frequency of 1450 cm^{-1}. The amide II band of the deuterated protein overlaps with the H—O—D bending vibration, making it difficult to obtain information about the conformation of this band. However, the remainder of the amide II band at 1550 cm^{-1} may provide information about the accessibility of solvent to the peptide backbone. Hydrophobic environments or tightly ordered structures, such as α-helix or β-sheet, reduce the chance of exchange of the amide N—H proton.

The most useful infrared band for the analysis of the secondary structure of proteins in aqueous media is the amide

Tab. 4 Characteristic infrared amide bands of proteins.

Designation	*Wave number [cm^{-1}]*	*Assignment*
A	3300	N−H stretching in resonance with
B	3110	With overtone (2 × amide II)
I	1653	80% C=O stretching; 10% C−N stretching; 10% N−H bending
II	1567	60% N−H bending; 40% C−N stretching
III	1299	30% C−N stretching; 30% N−H bending; 10% C=O stretching; 10% O=C−N bending; 20% other
IV	627	40% O=C−N bending; 60% other
V	725	N−H bending
VI	600	C=O bending
VII	200	C−N torsion

I band, occurring between approximately 1700 and 1600 cm^{-1}. The amide I band represents 80% C=O stretching vibration of the amide group coupled to the in-plane N−H bending and C−N stretching modes. The exact frequency of this vibration depends on the nature of hydrogen bonding involving the C=O and N−H groups and this is determined by the particular secondary structure adopted by the protein. Proteins generally contain a variety of domains containing polypeptide fragments in different conformations. As a consequence, the observed amide I band is usually a complex composite, consisting of a number of overlapping component bands representing helices, β-structure, turns, and random structures.

In addition, the infrared contributions of the side chains of the amino acids that constitute the protein must also be considered. Amino acid side chains exhibit infrared modes that are often useful for investigating the local group in a protein. Fortunately, these contributions have been found to be small in D_2O compared to the contributions made by the amide I band. It is also important to be aware of the location of such modes as they may be confused with amide vibrations. The arginyl residue is the only residue that makes a significant contribution in the 1700 to 1600 cm^{-1} region, but even the bands at 1586 and 1608 cm^{-1} due to the arginyl residue do not contribute greatly compared to the amide I contributions. The characteristic side-chain infrared frequencies of amino acids are summarized in Table 5.

Resolution enhancement of the amide I band allows for the identification of various structures present in a protein or peptide. Derivatives and deconvolution can be used to obtain such information. However, simply measuring the peak heights of the resulting spectra does not provide an accurate measure of the proportion of the secondary structure. In Fourier deconvolved spectra, the bandshapes are distorted and the extent of this distortion depends on the widths of the component bands and the parameters used to compute the deconvolved spectra, so the resulting peak heights are complex functions of a number of parameters. Similarly, the peak heights in the derivative spectra cannot be used as a quantitative measure because they are functions of both intensities and the widths of the original bands. The best

Tab. 5 Characteristic frequencies of amino acid side chains.

Amino acid	*Wave number [cm^{-1}]*	*Assignment*
Alanine	1465	CH_2 bending
Valine	1450	CH_3 asymmetric bending
Leucine	1375	CH_3 symmetric bending
Serine	1350–1250	O–H deformation
Aspartic acid	1720	C=O stretching
Glutamic acid	1560	CO_2^- asymmetric stretching
	1415	CO_2^- symmetric stretching
Asparagine	1650	C=O stretching
Glutamine	1615	NH_2 bending
Lysine	1640–1610, 1550–1485	NH_3^+ deformation
	1160, 1100	NH_3^+ rocking
Phenylalanine	1602, 1450, 760, 700	Ring vibrations
Tyrosine	1600, 1450	Ring vibrations
Arginine	1608, 1586	Ring vibrations

method used for the estimation of protein secondary structure involves band-fitting the amide I band. The parameters required, the number of component bands and their positions, are obtained from the resolution enhanced spectra. The fractional areas of the fitted component bands are directly proportional to the relative proportions of structure that they represent. The percentages of helix, β-structure, and turns may be estimated by the addition of the areas of all the component bands assigned to each of these structures and expressing the sum as a fraction of the total amide I area. The assumption is made that the intrinsic absorptivities of the amide I bands corresponding to different structures are identical.

Infrared spectroscopy has been used in a large number of studies of proteins in a range of environments. Some examples of the proteins analyzed are listed in Table 6, which details the deconvolved amide I frequencies and secondary structure assignments made for a series of proteins in D_2O. These assignments are based on the fact that the secondary structures of these globular proteins have been very well characterized by X-ray crystallography. The characteristic frequencies of the secondary structures of proteins have also been estimated from normal coordinate calculations on model peptides and proteins of known structure. The frequencies are summarized in Table 7. These assignments are by no means exact and some bands may appear outside these ranges for some proteins, particularly when solvent interactions are considered.

An example of the approach to protein analysis is illustrated by Fig. 10, which shows the amide I band of the enzyme lysozyme in D_2O. The amide I band of this protein shows nine component bands and the relative areas of these components are listed in Table 8. The components may be assigned to the various types of secondary structures. The bands at 1623 and 1632 cm^{-1} are characteristic of β-structure, as is the band at 1675 cm^{-1}. The bands at 1667, 1684, and 1693 cm^{-1} occur at frequencies characteristic of turns and bends. The band at 1610 cm^{-1} is due to arginyl side-chain vibrations and

Tab. 6 Amide I band frequencies and secondary structure assignments for standard proteins.

Protein	*β-structure*					*Helix*		*Random*		*Turns and bends*				
Carbonic anhydrase		1636		1625	1679		1653		1645		1660	1668		1690	
Carboxypeptidase	1635	1628	1623	1679		1652		1642		1665	1671		1690		
Casein									1644			1671			
α-chymotrypsin	1637		1627	1674		1654		1647		1665		1681	1688		
Chymotrypsinogen		1637		1627	1674		1654		1646		1664		1682	1688	1696
Concanavalin A	1639	1634	1623	1671		1653		1646		1659		1684		1694	
Elastase	1639	1633	1620	1673		1657		1645		1665		1683			
Immunoglobulin G		1637		1624	1672		1651				1659			1688	
α-lactalbumin		1637		1627	1676		1652		1644		1665		1684		
β-lactalglobulin		1634	1623	1679		1654		1646		1685		1685		1692	
Lysozyme			1630		1673		1654		1641		1666		1682		
Papain		1640	1632	1621	1679		1654		1646		1662	1670		1687	
Ribonuclease A	1637	1628		1676		1655		1646		1665		1684	1688		
Trypsin		1638		1627	1673		1654		1646		1664		1681		
Trypsinogen		1636		1624	1675		1654		1646		1663		1684		1693
Trypsin inhibitor	1636			1675		1656		1646		1664		1685			
Mean		1637	1631	1624	1675		1654		1645		1663	1670	1683	1688	1694
Rms deviation		1.4	2.5	2.4	2.5		1.5		1.6		2.2	1.4	1.5	1.1	1.7
Maximum deviation		3	3	4	4		3		4		4	2	2	2	2

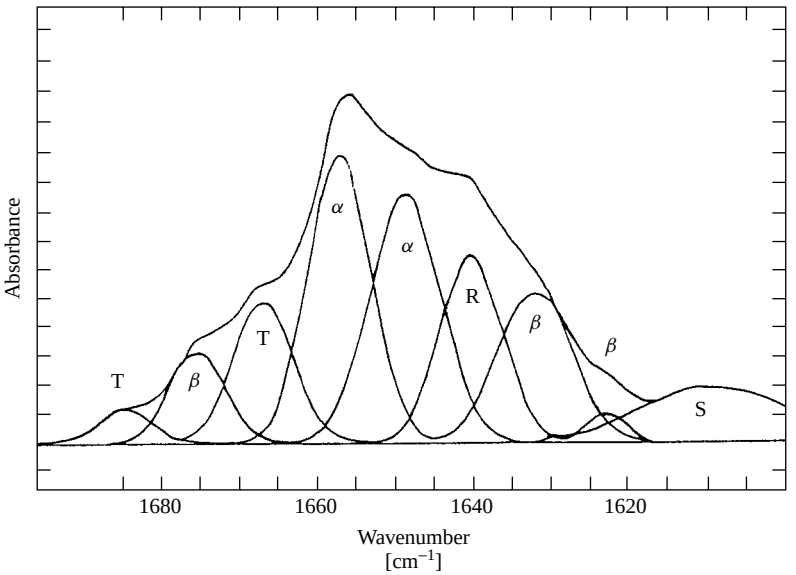

Fig. 10 Curve-fitted amide I band of lysozyme in D_2O.

Tab. 7 Characteristic amide I band frequencies of protein secondary structure.

Wave number [cm^{-1}]	*Assignment*
1695–1670	Intermolecular β-structure
1690–1680	Intramolecular β-structure
1666–1659	3-turn helix
1657–1648	α-helix
1645–1640	Random coil
1640–1630	Intramolecular β-structure
1625–1610	Intermolecular β-structure

Tab. 8 Analysis of the amide I band of lysozyme in D_2O.

Wave number [cm^{-1}]	*Relative area [%]*
1623	1
1632	15
1640	16
1648	24
1657	24
1667	11
1675	7
1684	2
1693	<1

the 1640-cm^{-1} band may be assigned to random coil. The two remaining bands at 1648 and 1657 cm^{-1} are due to the presence of α-helix. Usually, in protein infrared spectra, only one component due to α-helix is observed. However, X-ray data indicates the presence of two types of helix in lysozyme: α-helix and 3-turn helix. These different helices vibrate at different frequencies and the 1657-cm^{-1} band is assigned to α-helix, while the 1648-cm^{-1} band is due to 3-turn helix. These assignments give a quantitative estimate of 48% helix, 23% β-structure, 13% turns, and 16% random coil in lysozyme.

FTIR spectroscopy is particularly useful for probing the structure of membrane proteins. The technique can be used to study the secondary structure of proteins

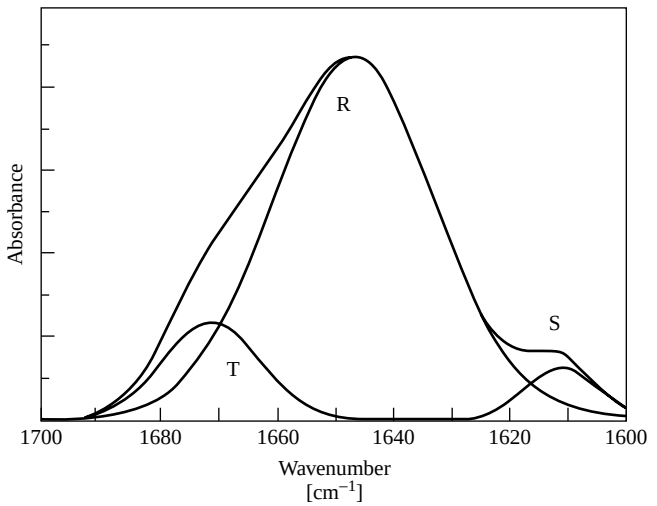

Fig. 11 Curve-fitted amide I band of MBP in D_2O.

Tab. 9 Analysis of the amide I band of MBP in D_2O and in reconstituted myelin.

D_2O wave number [cm^{-1}]	*Relative area [%]*	*Reconstituted myelin wave number [cm^{-1}]*	*Relative area [%]*
		1624	8
		1634	14
		1642	9
1647	86	1651	22
1671	14	1663	14
		1675	11
		1686	3

both in their native environment as well as after reconstitution into model membranes. Myelin basic protein (MBP) is a major protein of the nervous system and has been studied using FTIR spectroscopy in both aqueous solution and after reconstitution in myelin lipids. The amide I band of MBP in D_2O solution (deconvolved and curve-fit) is illustrated in Fig. 11. The amide band shows a small band at 1616 cm^{-1} that occurs at a frequency normally attributed to amino acid side-chain contributions. The largest component band observed at 1647 cm^{-1} is assigned to random coil and accounts for 86% of the total band area. The remaining component band at 1671 cm^{-1} is assigned to turns and bends in the protein. Table 9 lists the wave number and the fractional area of each component of the amide I band of MBP in D_2O. The FTIR spectrum of MBP protein after reconstitution in myelin, with all proteins removed from the myelin, shows that there are dramatic changes to the amide I band of MBP when the protein is complexed with myelin. The

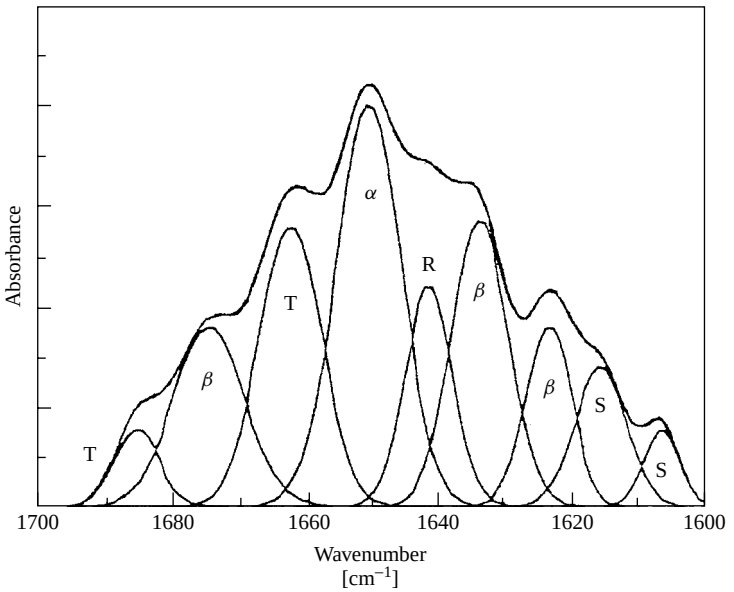

Fig. 12 Curve-fitted amide I band of MBP in myelin.

results of deconvolution and band-fitting the amide I band of MBP in myelin are shown in Fig. 12. The wave number and the fractional area of each component of the amide I band of MBP in myelin are also listed in Table 9. In Fig. 12, the two small components at 1606 and 1616 cm^{-1} occur in the frequency range attributed to side-chain contributions. The appearance of the band at 1624 cm^{-1} indicates the presence of water-bonded β-structure. Also, the component at 1634 cm^{-1} has been assigned to β-structure, with a high-frequency component due to β-structure observed at 1675 cm^{-1}. These bands contribute to 40% of the total band area. This estimation indicates that a notable amount of β-structure is formed when the results are compared to those obtained for MBP in D_2O, where no β-structure was observed. The component at 1651 cm^{-1}, contributing to 28% of the total amide I band area, has been assigned to α-helix in the protein. Thus, reconstitution of MBP in myelin produces a significant amount of α-helix in the protein in the predominantly lipid environment. Again, none of this type of structure was observed for the protein in an aqueous environment. The small component at 1642 cm^{-1} appears at a frequency normally attributed to random coil. The remainder of the components in Fig. 11 at 1663 and 1686 cm^{-1} have been assigned to turns and bends in the protein. The importance of the conditions in which biological molecules are examined in the infrared needs to be emphasized. The lipid environment more closely mimics the native environment, and thus presumably gives a far better indication of the native structure.

4.3 Nucleic Acids

Nucleic acids are vital molecules that carry the genetic code and are responsible for its expression by protein synthesis. The nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), may be

studied using infrared spectroscopy. The spectra of nucleic acids may be divided into the modes due to the constituent base, sugar, and phosphate groups. The bases (thymine, adenine, cytosine, guanine, and uracil) give rise to purinic and pyrimidinic vibrations in the range 1800 to 1500 cm^{-1} and these bands are sensitive markers for base pairing and base stacking effects. Bands in the region 1500 to 1250 cm^{-1} of nucleic acids are due to the vibrational coupling between a base and a sugar, and in the range 1250 to 1000 cm^{-1}, sugar–phosphate chain vibrations are observed. These bands provide information about backbone conformations. In the region 1000 to 800 cm^{-1}, sugar/sugar–phosphate vibrations are observed. The major infrared modes of nucleic acids are listed in Table 10.

Tab. 10 Major infrared bands of nucleic acids.

Wave number [cm^{-1}]	*Assignment*
2960–2850	CH_2 stretching
1705–1690	RNA C=O stretching
1660–1655	DNA C=O stretching, N–H bending
	RNA C=O stretching
1610	C=C imidazole ring stretching
1578	C=N imidazole ring stretching
1244	RNA PO_2^- asymmetric stretching
1230	DNA PO_2^- asymmetric stretching
1218	RNA C–H ring bending
1160, 1120	RNA ribose C–O stretching
1089	DNA PO_2^- symmetric stretching
1084	RNA PO_2^- symmetric stretching
1060, 1050	RNA and DNA Ribose C–O stretching
1038	RNA ribose C–O stretching
1015	DNA ribose C–O stretching
	RNA ribose C–O stretching
996	RNA uracil ring stretching, uracil ring bending
970, 916	DNA ribose-phosphate skeletal motions

4.4 Disease Diagnosis

In recent years, infrared spectroscopy has emerged as a powerful tool for characterizing tissue and disease diagnosis. High-pressure (pressure-tuning) infrared spectroscopy has been developed as a successful technique for the evaluation of malignancy in human tissues. The pressure dependence on parameters such as frequency, intensity, and band shape can provide further information about the structural changes associated with malignancy. For example, it is possible to detect cervical cancer arising from a premalignant state termed dysplasia by using infrared band differences among normal, dysplastic, and malignant cervical cells. A study of the pressure dependence of the frequencies of bands indicates that there are extensive changes in the degree of hydrogen bonding of phosphodiester groups of nucleic acids and the C–OH groups of proteins. Pressure studies also indicate changes in the degree of disorder of methylene chains of lipids in the malignant tissue. The infrared spectra of tissue samples with dysplasia demonstrate the same changes to a lesser degree than those observed for the cancer samples. Figure 13 shows the pressure dependence of the CH_2 bending mode frequency of the methylene chain of lipids for both normal and malignant tissue. Pressure increases this frequency because it induces ordering of the methylene chains in the lipid bilayers, thus increasing interchain interactions. In malignant cervical tissue, pressure induces a smaller shift in the frequency of this mode compared with normal tissue. The difference indicates that in cervical cancer, the methylene chains of

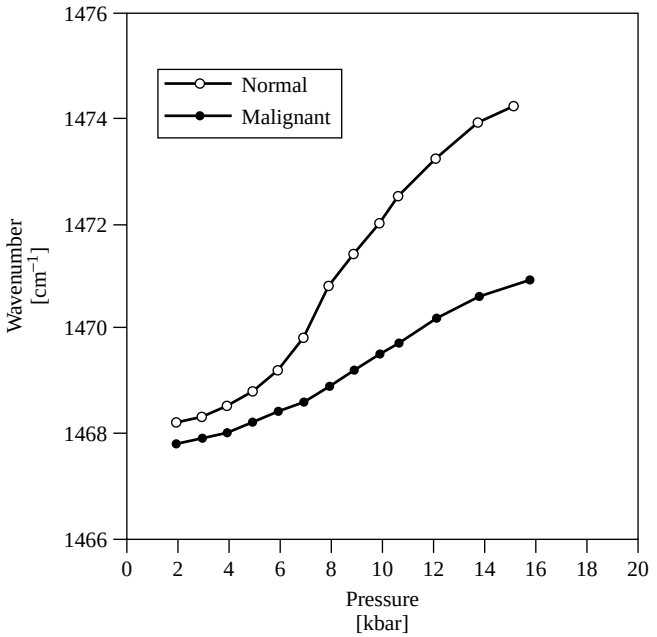

Fig. 13 Pressure dependence of CH_2 bending mode frequency of methylene chain of lipids for normal and malignant cervical tissues. (From Wong P.T.T. et al. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10988–10992.)

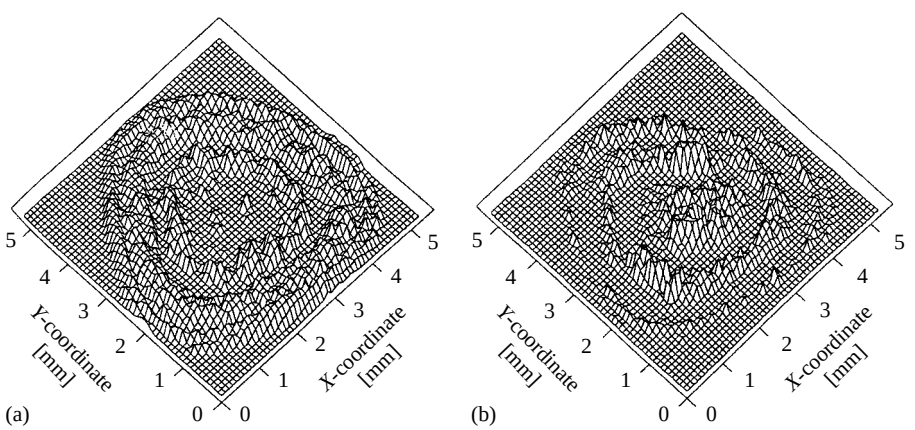

Fig. 14 Complementary functional group maps of a polished surface of sectional gallstone: (a) 875 cm^{-1} band of $CaCO_3$ and (b) 1050 cm^{-1} band of cholesterol. (From Wentrup-Byrne, E. et al. (1995) *Appl. Spectrosc.* **49**, 1028–1036.)

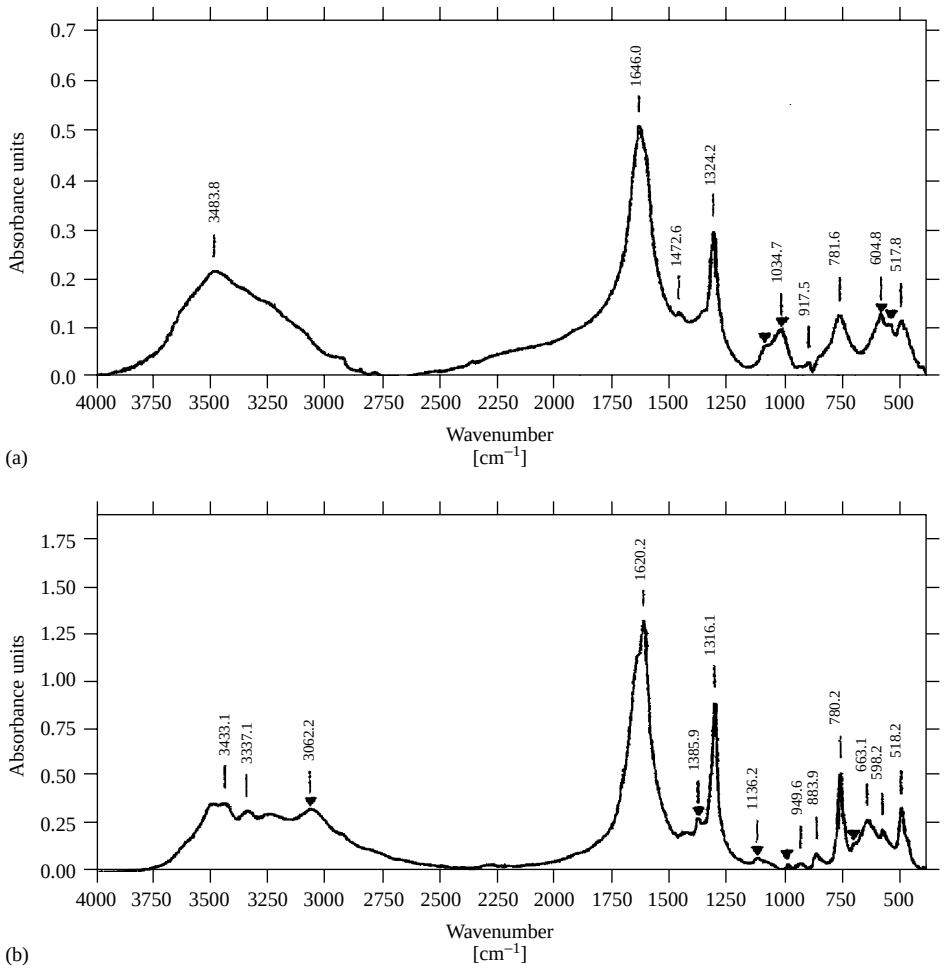

Fig. 15 Infrared spectra of kidney stones: (a) Stone mainly composed of calcium oxalate dihydrate and (b) stone mainly composed of calcium oxalate monohydrate. (From Estepa, L., Daudon, M. (1997) *Biospectroscopy* **3**, 347–369.)

lipids are more disordered than in normal cervical tissue.

Infrared mapping has the potential to be widely used to study diseased tissues. The identification of diseased regions of tissue is clearly complicated and the use of multivariate pattern recognition methods is appropriate. An example of where infrared microscopic mapping may be used to characterize disease is gallstones. Gallstone formation in the gallbladder is a common disease particularly in Western countries. The mechanism for the formation of gallstones is not thoroughly understood and knowledge of the distribution of the constituents within a stone may provide information about the mechanism of formation. A mixed stone comprised of

concentric bands has been polished and mapped. Using infrared spectroscopy, the concentric bands can be shown to consist of alternating regions of $CaCO_3$ and cholesterol. Figure 14 shows the complementary functional group maps of the polished surface of a sectioned gallstone: a map of the 875-cm^{-1} band of $CaCO_3$ and the 1050-cm^{-1} band of cholesterol.

Kidney disease is relatively common in humans and there are a number of causes of the disease. Knowledge of kidney stone composition and structure is important for establishing the etiology of the disease. FTIR spectroscopy has proved successful for the identification of the molecular structure and the crystalline composition of these complex stones. The main components of stones are calcium oxalates, calcium phosphates, and purines, but more than 85 mineral or organic compounds have been identified. There are different crystalline forms of the compounds. For example, calcium oxalate forms three crystalline forms in urine: calcium oxalate monohydrate, calcium oxalate dihydrate, and calcium oxalate trihydrate. One approach to diagnosis is to examine the form of calcium oxalate in the kidney stone. Examples of the infrared spectra produced for kidney stones of differing calcium oxalate forms are illustrated in Fig. 15. Figure 15(a) is the spectrum of a stone found to contain 95% calcium oxalate dihydrate, while Fig. 15(b) is the spectrum of a stone containing 85% calcium oxalate monohydrate. Quantitative analysis was carried out using an absorbance ratio of two peaks observed in the spectra of a binary mixture. The sample in Fig. 15(a) shows a high proportion of calcium oxalate dihydrate, which is linked to a high calcium/oxalate ratio and a high urine calcium content associated with renal hypercalciuria and hyperparathyroidism. In contrast, the predominance of calcium oxalate monohydrate for the sample in Fig. 15(b) indicates that there is a high oxalate content in the urine that is linked to primary hyperoxaluria, inflammatory bowel disease, or a high intake of oxalate-rich food.

4.5 Microbial Cells

Infrared spectroscopy has proved to be a valuable tool for characterizing and differentiating microbial cells. Given the

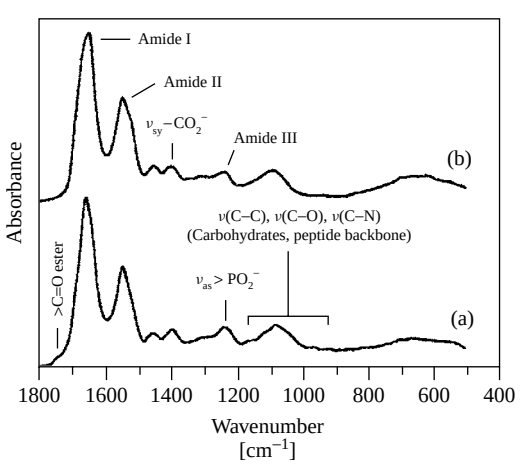

Fig. 16 Infrared spectra of *E. coli* (a) and ribonuclease A (b). (From Naumann, D. (2001) *Appl. Spectrosc. Rev.* **36**, 239–298.)

complex nature of such microorganisms, the spectra can contain a superposition of hundreds of infrared modes. However, the FTIR spectra of bacteria show modes predominantly due to the protein component. Although the cells also contain DNA and RNA structures, carbohydrates, and lipids, the various cell and membrane proteins form the major part of the cell mass. Figure 16 shows part of the infrared spectra of *Escherichia coli* and a typical protein, ribonuclease A, illustrating the similarity.

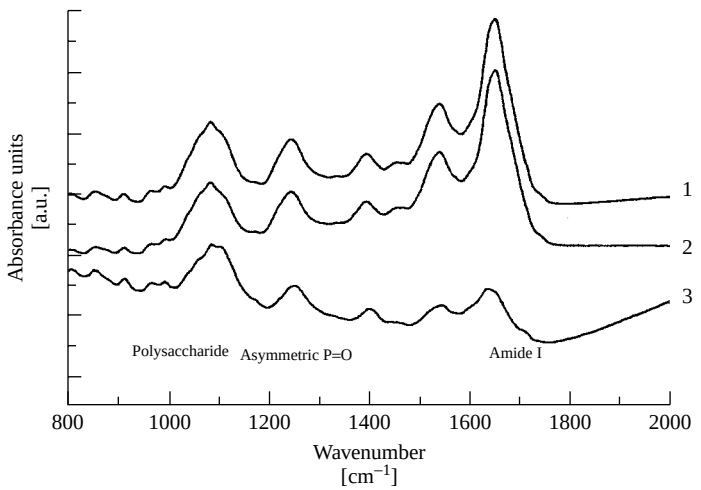

Fig. 17 Infrared spectra of an *E. coli* colony measured at center (1) and edge of colony (2) with corresponding difference spectrum (3). (From Choo-Smith, L.P. et al. (2001) *Appl. Environ. Microbiol.* **67**, 1461–1469.)

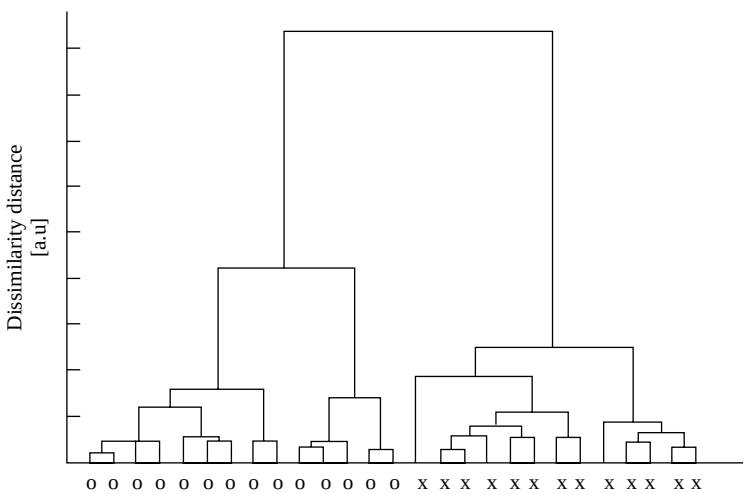

Fig. 18 Dendrogram from hierarchical clustering analysis of the infrared spectra of 7 h colonies of two *E. coli* strains. (From Choo-Smith, L.P. et al. (2001) *Appl. Environ. Microbiol.* **67**, 1461–1469.)

However, modes due to carbohydrates and lipids can also be observed in bacterial infrared spectra to a lesser extent.

FTIR spectroscopy provides a rapid method for identifying microorganisms responsible for infections. The technique has been used to monitor the biochemical heterogeneity of microcolonies of *E. coli*. By examining the individual spectra and calculating the difference spectra, it is possible to gain a better understanding of the source of the clustering. Figure 17 illustrates the infrared spectra obtained from the center and the edge of an *E. coli* colony. Despite the fact that the spectra appear very similar at first sight, difference spectra reveal variation. There are differences in the region near 1230 cm^{-1},

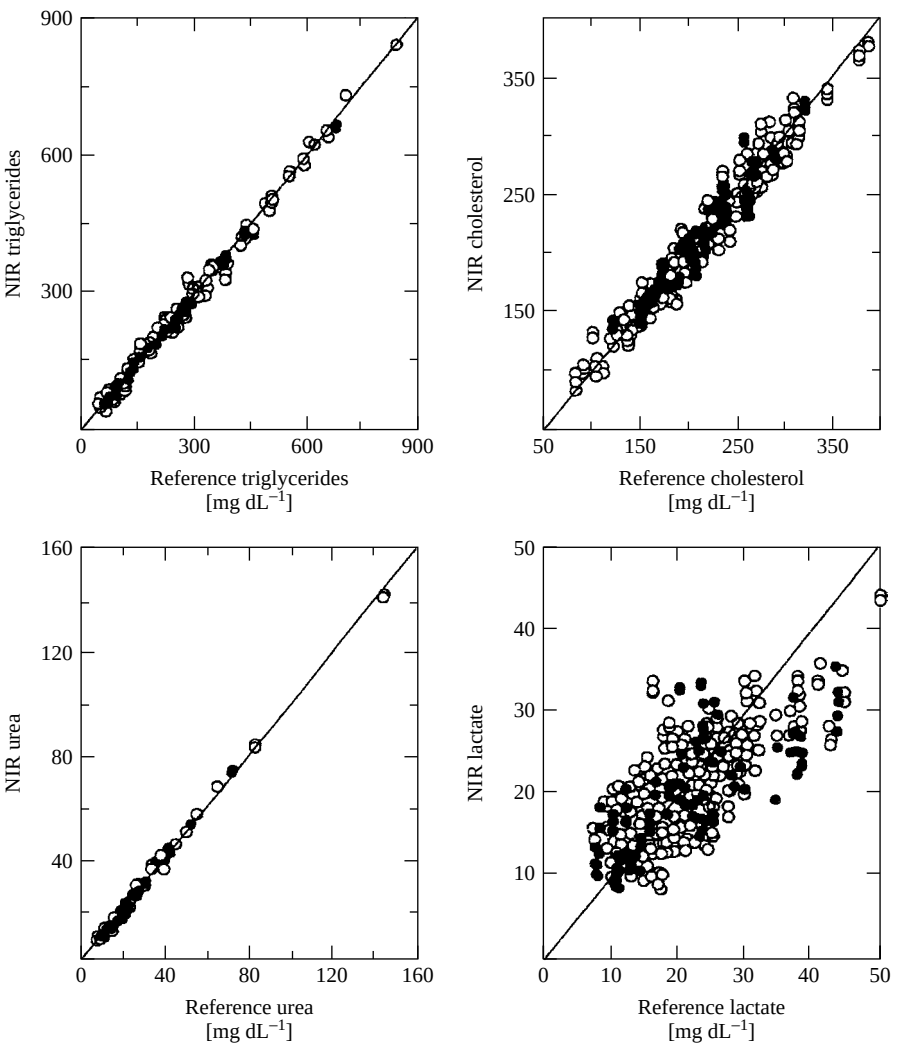

Fig. 19 Comparison of serum analyte levels determined using near-infrared spectroscopy to reference analytical results. (From Hazen, K.H. et al. (1998) *Anal. Chim. Acta* **371**, 255–267.)

which may be assigned to the phosphate double-bond asymmetric stretching vibration of the phosphodiester, free phosphate, and monoester phosphate functional groups. There are also differences observed for the protein amide I region (1670–1620 cm^{-1}), the symmetric stretching vibrations of COO^- functional groups (1400 cm^{-1}), and the carbohydrate region (1200–900 cm^{-1}). Hierarchical cluster analysis of the spectra of 7 h colonies of two *E. coli* strains were carried out and Fig. 18 illustrates the resulting dendrogram. This clearly shows the formation of two major clusters corresponding to the different strains.

4.6 Clinical Chemistry

Common infrared tests in clinical chemistry include glucose, blood, and urine analyses. The quantification of glucose in body fluids is an important aspect of clinical analysis, especially in the case of blood glucose measurements for diabetic patients. The methods of near-infrared (NIR) spectroscopy and ATR spectroscopy in the mid-infrared region lend themselves to the study of glucose in such aqueous environments, and PLS and PCR methods have been successfully applied to quantitative analysis of glucose samples. Biological fluids may also be investigated using NIR or ATR spectroscopy, both as native fluids or dry films. Figure 19 demonstrates the effectiveness of NIR spectroscopy as a technique for the quantitative analysis of native serum samples. This figure shows scatter plots that compare the concentration results obtained using NIR spectroscopy with the results obtained using standard reference clinical methods. In these experiments, the calibration procedure involved using a number of spectra and independent reference analyses and then optimizing PLS models for each constituent individually.

See also Detection and Characterization of Single Biomolecules.

Bibliography

Books and Reviews

Clark, R.J.H., Hester, R.E. (1996) *Biomedical Applications of Spectroscopy*, Wiley, Chichester.

Gremlich, H.U., Yan, B. (2000) *Infrared and Raman Spectroscopy of Biological Materials*, Marcel Dekker, New York.

Jackson, M., Sowa, M.G., Mantsch, H.H. (1997) Infrared spectroscopy: a new frontier in medicine, *Biophys. Chem.* **68**, 109–125.

Kalasinsky, V.F. (1996) Biomedical applications of infrared and Raman microscopy, *Appl. Spectrosc. Rev.* **31**, 193–249.

Mantsch, H.H., Chapman, D. (Eds.) (1996) *Infrared Spectroscopy of Biomolecules*, Wiley, New York.

Naumann, D. (2001) FT-infrared and FT-Raman spectroscopy in biomedical research, *Appl. Spectrosc. Rev.* **36**, 239–298.

Petrich, W. (2001) Mid-infrared and Raman spectroscopy for medical diagnostics, *Appl. Spectrosc. Rev.* **36**, 181–237.

Shaw, R.A., Mantsch, H.H. (1999) Vibrational biospectroscopy: from plants to animals to humans. A historical perspective, *J. Mol. Struct.* **480–481**, 1–13.

Shaw, R.A., Mantsch, H.H. (2000) Infrared spectroscopy in clinical and diagnostic analysis, in: Meyers, R.A. (Ed.) *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*, Wiley-Interscience, New York, pp. 83–102.

Stuart, B.H. (1997) *Biological Applications of Infrared Spectroscopy*, Wiley, Chichester.

Stuart, B.H. (2000) Infrared spectroscopy of biological applications, in: Meyers, R.A. (Ed.) *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*, Wiley-Interscience, New York, pp. 529–559.

Primary Literature

Arrondo, J.L.R., Goni, F.M. (1999) Structure and dynamics of membrane proteins as studied by infrared spectroscopy, *Prog. Biophys. Mol. Biol.* **72**, 367–405.

Arrondo, J.L.R., Muga, A., Castresana, J., Goni, F.M. (1993) Quantitative studies of the structure of proteins in solution by Fourier transform infrared spectroscopy, *Prog. Biophys. Mol. Biol.* **59**, 23–56.

Bandekar, J. (1992) Amide modes and protein conformation, *Biochim. Biophys. Acta* **1120**, 123–143.

Banyay, M., Sarkar, M., Grasland, A. (2003) A library of IR bands of nucleic acids in solution, *Biophys. Chem.* **104**, 477–488.

Barth, A. (2000) The infrared absorption of amino acid side chains, *Prog. Biophys. Mol. Biol.* **74**, 141–173.

Benedetti, E., Bramanti, E., Papineschi, F., Rossi, I., Benedetti, E. (1997) Determination of the relative amount of nucleic acids and proteins in leukemic and normal lymphocytes by means of Fourier transform infrared microspectroscopy, *Appl. Spectrosc.* **51**, 792–797.

Bhandara, P., Mendelson, Y., Stohr, E., Peura, R.A. (1994) Glucose determination in simulated blood serum solutions by Fourier transform infrared spectroscopy, *Appl. Spectrosc.* **48**, 271–273.

Brown, S.D. (2000) Chemometrics in: Meyers, R.A. (Ed.) *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*, Wiley-Interscience, New York, pp. 9669–9671.

Byler, D.M., Susi, H. (1986) Examination of the secondary structure of protein by deconvolved FTIR spectra, *Biopolymers* **25**, 469–487.

Carmona, P., Bellanato, J., Escolar, E. (1997) Infrared and Raman spectroscopy of urinary calculi: a review, *Biospectroscopy* **3**, 331–346.

Choo-Smith, L.P., Maquelin, K., van Vreeswijk, T., Bruining, H.A., Puppels, G.J., Ngo Thi, N.A., Kirschner, C., Naumann, D., Ami, D., Villa, A.M., Orsini, F., Doglia, S.M., Lamfarraj, H., Sockalingum, G.D., Manfait, M., Allouch, P., Endtz, H.P. (2001) Investigating microbial microcolony heterogeneity by vibrational spectroscopy, *Appl. Environ. Microbiol.* **67**, 1461–1469.

Ding, Q., Small, G.W., Arnold, M.A. (1998) Genetic algorithm-based wavelength selection for the near-infrared determination of glucose in biological matrices – initialization strategies and effects of spectral resolution, *Anal. Chem.* **70**, 4472–4479.

Estepa, L., Daudon, M. (1997) Contribution of Fourier transform infrared spectroscopy to the identification of urinary stones and kidney crystal deposits, *Biospectroscopy* **3**, 347–369.

Everitt, B.S. (1994) *Statistical Methods for Medical Investigations*, Wiley, Toronto.

Ferraro, J.K., Krishnan, K. (Eds.) (1990) *Practical Fourier Transform Infrared Spectroscopy*, Academic Press, San Diego.

Franke, J.E. (2000) Inverse least squares and classical least squares methods for quantitative vibrational spectroscopy, in: Chalmers, J.M., Griffiths, P.R. (Eds.) *Handbook of Vibrational Spectroscopy*, Wiley, New York, pp. 2276–2292.

Fuller, M.P., Griffiths, P.R. (1978) Diffuse reflectance measurements by Fourier transform infrared spectroscopy, *Anal. Chem.* **50**, 1906–1909.

Gunzler, H., Gremlich, H.U. (2002) *IR Spectroscopy: An Introduction*, Wiley-VCH, Weinheim.

Hasegawa, T. (2002) Principal component regression and partial least squares modeling, in: Chalmers, J.M., Griffiths, P.R. (Eds.) *Handbook of Vibrational Spectroscopy*, Wiley, New York, pp. 2293–2312.

Hazen, K.H., Arnold, M.A., Small, G.W. (1998) Measurement of glucose and other analytes in undiluted human serum with near-infrared transmission spectroscopy, *Anal. Chim. Acta* **371**, 255–267.

Heise, H.M. (2002) Glucose measurement by vibrational spectroscopy, in: Chalmers, J.M., Griffiths, P.R. (Eds.) *Handbook of Vibrational Spectroscopy*, Wiley, New York, pp. 3280–3294.

Humecki, H.J. (Ed.) (1999) *Practical Guide to Infrared Microspectroscopy*, Marcel Dekker, New York.

Jackson, M., Mantsch, H.H. (1993) Biomembrane structure from FTIR spectroscopy, *Spectrochim. Acta Rev.* **15**, 53–69.

Janatsch, G., Kruse-Jarres, J.D., Marbach, R., Heise, H.M. (1989) Multivariate calibration for assays in clinical chemistry using attenuated total reflection infrared spectra of human blood plasma, *Anal. Chem.* **61**, 2016–2023.

Kauppinen, J.K., Moffatt, D.J., Mantsch, H.H., Cameron, D.G. (1981) Fourier self-deconvolution: a method for resolving intrinsically overlapped bands, *Appl. Spectrosc.* **35**, 271–276.

Kidder, L.H., Haka, A.S., Lewis, E.N. (2002) Instrumentation for FT-IR imaging, in: Chalmers, J.M., Griffiths, P.R. (Eds.) *Handbook of Vibrational Spectroscopy*, Wiley, New York, pp. 1386–1404.

Liquier, J., Taillandier, E. (1996) Infrared spectroscopy of nucleic acids, in: Mantsch, H.H., Chapman, D. (Eds.) *Infrared Spectroscopy of Biomolecules*, Wiley, New York.

Mariey, L., Signolle, J.P., Amiel, C., Travert, J. (2001) Discrimination, classification, identification of microorganisms using FTIR spectroscopy and chemometrics, *Vib. Spectrosc.* **26**, 151–159.

Mark, H. (1996) *Principles and Practice of Spectroscopic Calibration*, Wiley, New York.

Messerschmidt, R.G., Harthcock, M.A. (Eds.) (1988) *Infrared Microspectroscopy: Theory and Applications*, Marcel Dekker, New York.

Mushayakarara, E., Levin, I.W. (1982) Determination of acyl chain conformation at the lipid interface region, *J. Phys. Chem.* **86**, 2324–2327.

Pevsner, A., Diem, M. (2001) Infrared spectroscopic studies of major cellular components. Part II: The effect of hydration on the spectra of nucleic acids, *Appl. Spectrosc.* **55**, 1502–1505.

Pouchert, C.J. (1981) *The Aldrich Library of Infrared Spectra*, Aldrich Chemical Company, Milwaukee.

Schmitt, J., Flemming, H.C. (1998) FTIR spectroscopy in microbial and material analysis, *Int. Biodeterior. Biodegrad.* **41**, 1–11.

Shaw, R.A., Kotowich, S., Leroux, M., Mantsch, H.H. (1998) Multivariate serum analysis using mid-infrared spectroscopy, *Clin. Biochem.* **35**, 634–632.

Shaw, R.A., Kotowich, S., Mantsch, H.H., Leroux, M. (1996) Quantitation of protein, creatine and urea in urine by near-infrared spectroscopy, *Clin. Biochem.* **29**, 11–19.

Smith, B.C. (2002) *Quantitative Spectroscopy: Theory and Practice*, Elsevier, Amsterdam.

Stuart, B.H. (1996) A Fourier transform infrared spectroscopic study of the secondary structure of myelin basic protein in reconstituted myelin, *Biochem. Mol. Biol. Int.* **38**, 839–845.

Stuart, B.H., Forbes, S., Dent, B.B., Hodgson, G. (2000) Studies of adipocere using diffuse reflectance infrared spectroscopy, *Vib. Spectrosc.* **24**, 233–242.

Taillander, E., Liquier, J. (2002) Vibrational spectroscopy of nucleic acids, in: Chalmers, J.M., Griffiths, P.R. (Eds.) *Handbook of Vibrational Spectroscopy*, Wiley, New York, pp. 3465–3480.

Watts, A., De Pont, J.J.H.H.M. (Eds.) (1986) *Progress in Protein-Lipid Interactions 2*, Elsevier, Amsterdam.

Wentrup-Byrne, E., Rintoul, L., Smith, J.L., Fredericks, P.M. (1995) Comparison of vibrational spectroscopic techniques for the characterization of human gallstones, *Appl. Spectrosc.* **49**, 1028–1036.

Wetzel, D.L., Levine, S. (1999) Microspectroscopy – imaging molecular chemistry with infrared microscopy, *Science* **285**, 1224–1225.

Wong, P.T.T., Wong, R.K., Caputo, T.A., Godwin, T.A., Rigas, B. (1991) Infrared spectroscopy of exfoliated human cervical cells: evidence of extensive structural changes during carcinogenesis, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10988–10992.

Fungal Biotechnology

Brian McNeil and Linda M. Harvey
Department of Bioscience, Glasgow, UK

1 Introduction 3
1.1 Physical Characteristics of Filamentous Fungi (Molds) 4
1.2 Metabolic Activities of Filamentous Fungi 4

2 Development of Fungal Biotechnology 4
2.1 Solid Substrate Fermentation (SSF) Systems 4
2.2 Submerged Liquid Fermentation (SLF) Systems 5
2.2.1 Penicillins 7
2.2.2 Enzyme Production by Fermentation 7
2.2.3 Other Liquid Fermentation Products 7

3 Production Modes 8
3.1 Batch Culture 8
3.2 Fed Batch Culture 8
3.3 Continuous Culture 8

4 Problems in Cultivation of Fungi in Fermentors 9

5 Filamentous Fungi and Heterologous Protein Production 9
5.1 Transcription 12
5.2 Translation 12
5.3 Posttranslational Modifications 12
5.4 Gene Fusion 12
5.5 Proteolysis 12
5.6 Conclusions 12

6 Biotechnology of Molds in Agriculture 13
6.1 Biocontrol Using Fungi 13
6.2 Mycoinsecticides 13

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

6.3 Production Technology 14
6.4 Other Fungal Biocontrol Agents 14
6.4.1 Future Prospects for Fungal Biocontrol 15
6.5 Ectomycorrhizal Fungi 16
6.6 Prospects 16

7 Fungal Biotransformations 16
7.1 Steroids from Fungi 17
7.2 Other Biopharmaceuticals from Fungi 17
7.3 Combinatorial Synthesis 17

8 Fungi and Toxic Materials 18

9 Genetic Improvement of Fungal Strains 18
9.1 Pathway Engineering in *Penicillium chrysogenum* 19

10 Perspective 20

Bibliography 20
Books and Reviews 20
Primary Literature 21

Keywords

Batch Culture
A closed culture system to which no major additions to nutrients are made after inoculation of the nutrient medium with the chosen microorganisms.

Continuous Culture (Chemostat)
A culture system in which the rate of addition of fresh medium balances the rate of removal of spent medium and cells, such that at steady state, the concentration of all reactants is constant.

Fed Batch Culture
A culture system to which nutrient is added at discrete time intervals or on a continuous basis, with the aim of avoiding either limitation or inhibition of O_2 owing to the nutrient.

Filamentous Fungi (Molds)
Eukaryotic microorganisms, characterized by growth at the apical tip, which form a branched network of vegetative filaments, a mycelium.

Solid Substrate Fermentation (SSF)
A low water fermentation process in which the fungus grows on or through the solid substrate, often a cereal grain or similar material.

Submerged Liquid Fermentation (SLF)
Describing liquid fermentation systems, within which the fungus is aerated and, often, mechanically agitated.

Molds (filamentous fungi) are used in the service of man to produce a wide range of valuable products, to improve feedstuffs, to carry out biotransformations, and to effect bioremediation. The technology used to cultivate and control mold activity on either solid or liquid media allows the production of high levels (kilos per cubic meter) of the desired products from selected microbial strains. Nevertheless, many problems in cultivating these microorganisms, due directly to their morphology, remain to be fully resolved. The biotechnological role of the filamentous fungi seems set to expand, with the development of effective transformation systems and increasing use of organisms such as *Aspergillus niger*, as efficient systems for the secretion of heterologous proteins. However, an understanding of the technological means of successfully cultivating these microorganisms will continue to be essential to the exploitation of their biotechnological potential.

1 Introduction

Biotechnology has been defined as "the use of whole cells or parts derived therefrom to catalyze the formation of useful products or to treat pollutants."

Within the broad field of biotechnology, the filamentous fungi (molds) have long occupied an important position. This diverse and metabolically versatile group of eukaryotic microorganisms is characterized by the formation of vegetative branching filaments (hyphae), which together form a mycelium. This growth pattern is a result of polarization of growth at the hyphal tip (apex).

Although most groups of molds have economically significant members, biotechnologically, the most important fungi are found within the Fungi Imperfecti, Ascomycetes, and Basidiomycetes.

Recent years have seen several major advances in our basic understanding of how the fungi work at the cellular level, including the publication of the full genome of the industrial workhorse, *A. niger* (by DSM). The publication of the genome of this organism potentially signals the start of a revolution in our exploitation of the fungi for mankind's benefit, since the relationship between genome, proteome, and cellular behavior in the fungi can now be approached on a more systematic and logical basis than hitherto. There has been great progress in the key areas of *functional genomics* (the allocation of function to each

identified gene) and *metabolic pathway engineering* (deliberate genetic alteration of a fungal strain to enhance synthesis of a desired metabolite). Advances in these areas will revolutionize our approach to understanding the physiology of the filamentous fungi.

Another recent development in our understanding of the basic function of the fungi is the discovery of "gene silencing" in the filamentous fungus, *Neurospora crassa*. It seems likely that both dsRNA interference and posttranscriptional gene silencing (PTGS) induced in fungi by the presence of transgenic DNA, involve sequence specific degradation of relevant messenger RNA, thus, effectively "silencing" the gene from which they were transcribed. This is a tremendously interesting new area with great potential for exploitation for mankind's benefit. Potentially novel methods of controlling gene activity in the fungi points the way to whole new strategies for controlling the growth of fungi in a range of environments, including medical, biotechnological, and ecological. At its simplest, this might imply the development of a completely new category of antifungal agents: given the woeful condition of our current antifungal armory and the increasing problems caused by fungi to the health of the human population, this would be immensely beneficial.

This is an exciting and rapidly developing area in the biotechnology of the fungi with revolutionary advances being made possible by our increased knowledge of fungal physiology.

1.1
Physical Characteristics of Filamentous Fungi (Molds)

The hyphal mode of growth confers a number of advantages on the filamentous fungi. These include the ability to translocate nutrients within the mycelial network, the ability to grow away from the initial growth point, such that hyphae at the edge of a colony constantly encounter fresh nutrients, and the ability to grow into solid materials (penetrative growth), thus allowing the uptake of nutrients whose availability would otherwise be restricted by diffusion.

In the growth process, fungi form branches, and branch frequency is closely related to nutrient availability.

1.2
Metabolic Activities of Filamentous Fungi

As a consequence of their mode of growth, the filamentous fungi possess a number of metabolic activities that are attractive to the biotechnologist. First among these is the synthesis and excretion of lytic enzymes (proteases, lipases, carbohydrases). Production of such hydrolytic enzymes aids fungal invasiveness. Under appropriate conditions, some fungi can secrete large quantities of proteins (50 gL^{-1} in some cases) into their environment.

The filamentous fungi are, as a group, prodigious producers of secondary metabolites. This structurally diverse group of products, formed in significant amounts during the stationary phase, includes compounds such as the penicillins and the gibberellins.

2
Development of Fungal Biotechnology

2.1
Solid Substrate Fermentation (SSF) Systems

Traditionally, molds have been used largely in the production and improvement

of foods. Natural colonization of moist grains or cereals by molds led to the discovery that certain molds could improve the flavor, texture, and nutritional value of largely carbohydrate materials.

The next steps involved control of the physicochemical conditions of the process and, later, isolation, selection, and use of particular strains. These processes, the classic examples of solid substrate fermentation (SSF) processes, are characterized by the overgrowth of the substrate, often a grain, by the fungus, which is then used as a source of enzymes to break down substrate components or for addition to other substrates. Such processes are often referred to by the Japanese term *koji*. Koji processes are still widely used in the Orient for formation of a wide range of products (Table 1). Most of these processes are now operated at the industrial scale, and many have been taken up by countries outside Asia. Tempeh, for example, in which the action of the fungi produces a material high in protein, flavorsome, and useful as a meat supplement, is produced in quantity in both Europe and the United States.

Europeans traditionally have used fungi such as *Penicillium roquefortii* and *P. camembertii* to grow on or in cheeses and to enhance flavor by production of lipases and proteases. Further use of SSF was in the industrial production of mushrooms, the most common being *Agaricus bisporus*.

All SSF systems are low water systems and generally have relatively low power inputs. Conversely, they are generally recognized as being difficult to monitor, control, and scale up. For example, O_2 transfer and the removal of CO_2 and heat can be difficult. Several SSF fermentor types are illustrated in Fig. 1.

2.2 Submerged Liquid Fermentation (SLF) Systems

Citric acid illustrates particularly well the development of the fermentation technology required for convenient formation of mold products in large quantities, at reasonable *rates* in liquid systems. Initially derived from lemon juice, citric acid was first produced industrially by a surface tray process using *A. niger*. The US firm of Pfizer used this relatively lengthy (6–12 days) labor-intensive process until 1991.

From the late 1940s, the use of the stirred-tank reactor (STR) for submerged liquid cultivation of *A. niger* resulted in citrate production processes lasting 3 to 5 days at 25 to 30 °C. A typical STR is shown in Fig. 1(d). Use of such fermentors

Tab. 1 Processes involving a koji stage.

Product	*Substrates*	*Mold involved*	*Location*
Shoyu (soy sauce)	Soy beans/wheat	*Aspergillus soyae* or *A. oryzae*	Japan (but similar products elsewhere)
Miso	Rice/barley soybeans	*A. soyae*	Japan
Sake (rice wine)	Rice	*A. oryzae*	Japan
Tempeh	Soybeans	*Rhizopus* sp. *R. oligosporus*	Indonesia

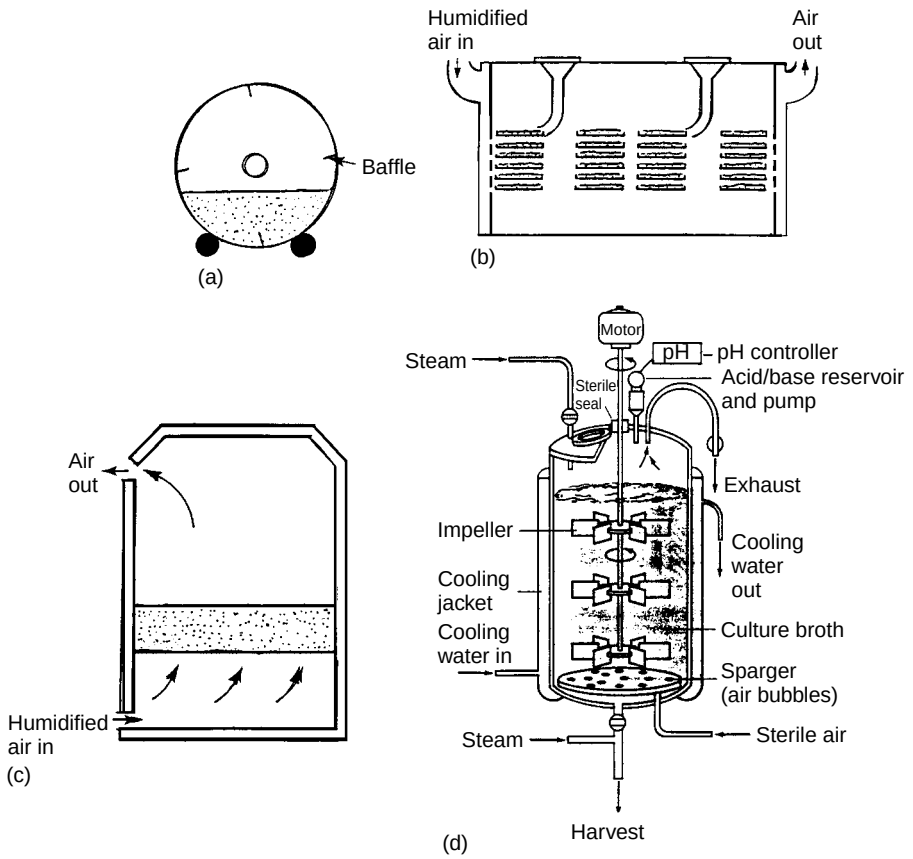

Fig. 1 Fermentor configurations showing three SSF types: (a) rotating drum, (b) tray system, and (c) formed air-flow system [from Whipps and Lumsden (1989)] and one SLF type: (d) stirred-tank bioreactor. [From Kinghorn and Turner (1992).]

allowed high oxygen transfer rates to be maintained, leading to high productivity with low manual labor needs. These fermentors were easier to monitor and control than SSF systems. The mold's physical environment could thus be more readily optimized. The STR is now the industry workhorse, a flexible reactor type used for the production of a wide range of products.

Some of the features developed in citrate fermentations are held in common with other STR mold fermentations, including the following:

- Use of highly bred (often relatively *genetically unstable*) pure cultures of production strains. (Production strains are *usually* stable these days!)
- Use of relatively crude raw materials such as molasses and corn steep liquor at the industrial scale
- Continuous aeration with sterile (filtered) air, and continuous stirring
- Maintenance of asepsis during fermentation (molds grow more slowly than bacteria and yeasts: thus any contamination by these microorganisms could lead to displacement of the fungus).

Current demand for citric acid is around 500 000 tons, though this amount is growing at 3 to 5% per year. Increasingly, the citric acid market is being concentrated in the hands of fewer, larger producers. Currently, only seven companies produce the vast bulk of citrate used.

2.2.1 Penicillins

At around the same time that submerged citrate processes developed, fermentations in STRs involving pure cultures of *Penicillium chrysogenum* for penicillin G production were also being developed. These processes are conceptually very similar to those used in citrate production. Initially, the carbon source was lactose, which permitted slow fungal growth and good excretion of the product, the secondary metabolite penicillin G. Later, fed batch processes using glucose feeds were developed. These are the current methods of production for penicillin G and V, which differ chemically only in the attached side chain (phenylacetic acid in the former, phenoxyacetic acid in the latter). Bulk penicillins G and V are increasingly used as raw materials for conversion into a wide range of semisynthetic penicillins.

Penicillins, like citrate, are mature biotechnological products produced in bulk and characterized by "survival of the fittest" with respect to producers.

2.2.2 Enzyme Production by Fermentation

Fungi are excellent sources of a wide range of lytic enzymes. Most of these are produced using the well-understood STR system described briefly in Sect. 2.2. In this area, the aspergilli are preeminent. The current world market for industrial enzymes, several of which are produced from aspergilli by SLF, is around $600 million per annum.

2.2.3 Other Liquid Fermentation Products

Table 2 lists some products of the liquid fermentation of fungi. In general, it is

Tab. 2 Some of the products of fungal biotechnology.

Organism	*Product(s)*	*Use*	*Mode of production*
Aspergillus niger	Citric acid	Acidulant	SLF (usually)
A. oryzae	Amyloglucosidase and other enzymes	Starch hydrolysis, lipolysis, proteolysis, etc.	SLF (fed batch)
Penicillium chrysogenum	Penicillins G and V	Antibacterial agents	SLF (fed batch)
Fusarium venenatum	Mycoprotein (Quorn)	Human foodstuff	SLF (continuous culture)
Gibberella fujikuroi	Gibberellins	Plant growth regulators	SLF
Tolypocladium (Beauveria)	Cyclosporin	Immunosuppressant in transplant surgery	SLF
Sclerotium glucanicum	Scleroglucan	Tertiary oil recovery; immunostimulant; antitumor agent	SLF (batch)
Agaricus campestris A. bisporus	Fruiting bodies (mushrooms)	Human foodstuff	SSF
Lentinus edodes	Fruiting body or cell extracts	Human foodstuff; immunostimulant	SLF

much easier to screen for antimicrobial activity of mold-derived compounds than for other activities. Indeed, some products such as cyclosporins were originally isolated for their antimicrobial characteristics (antifungal for cyclosporins), with their value as immunosuppressants being realized only later. It is not unreasonable to state that the use of cyclosporins have revolutionized transplant surgery.

Mucor species have been shown to accumulate large quantities of unsaturated fatty acids (e.g. linoleic). Although technically successful, SLF-based processes for production are currently not economic.

The production of mycoprotein (Quorn) by Marlow Foods (formerly ICI/RHM: Imperial Chemical Industries (now Zeneca) and Rank Hovis MacDougall) has been a considerable success. This is one of the few continuous-culture SLF processes, if not the only one, operated at the industrial scale. *Fusarium venenatum* is grown in STRs on a defined sugar–mineral salts medium, and the fungal biomass produced is harvested. By virtue of its filamentous nature, the biomass is textured, and this can be controlled or modified to simulate various meat products. The bland-tasting product readily accepts other flavors. Quorn appears to be an almost perfect foodstuff, having high protein quality, low fat content, zero cholesterol, and a reasonably high fiber content. Interest is also growing in the role of some fungal polysaccharides in modulating the immune systems of test animals. Such compounds are classed as biological response modifiers (BRMs) and include the β-glucan polysaccharides lentinan and scleroglucan. These compounds may stimulate the immune system by enhancement of macrophage activity.

3 Production Modes

3.1 Batch Culture

Although batch culture is the traditional mode of cultivation in fermentors, it is not suited to the formation of products where inhibition by substrate, product, oxygen limitation, or product precursors is a problem. Fungal citrate may be produced in batch culture.

3.2 Fed Batch Culture

In fed batch culture, the fermentation, often after an initial batch phase, is fed regularly or continuously with required nutrients or precursors. The feed is controlled to ensure that the level of the added compounds is never inhibitory to the culture, or, to avoid oxygen limitation, a rate is maintained that just balances the consumption of nutrients with oxygen consumption. Penicillins and amyloglucosidase are produced by this method, as are many other mold secondary metabolites.

3.3 Continuous Culture

The *chemostat* is the most common type of continuous-culture system encountered. Here the culture is continuously fed fresh nutrient medium, and spent medium, cells, and product are continuously drawn off. The system is not ideally suited to secondary metabolite production, but it is suitable for

biomass production, where productivities can be much higher than those in batch fermentations. This technique is also widely used on a small (lab) scale to study aspects of cellular biochemistry and physiology.

4 Problems in Cultivation of Fungi in Fermentors

Despite a long history of cultivation in both SSF and SLF systems, it is recognized that there are still some problems to be overcome. In relation to SSF systems, some of the problems have already been mentioned. In liquid systems, especially the STR, particular difficulties arise as a result of the particular morphology of the molds.

In these fermentor vessels, formation of a branched mycelium results in a rapid increase in the viscosity of the culture. The rheological character of the fermentation fluid may also change: a low viscosity Newtonian system may become highly viscous and non-Newtonian.

The general effect of this trend is to reduce mixing efficiency as well as oxygen and heat transfer in the fermentor. The result is heterogenicity within the fermentor, and, in many cases, fermentor productivity is limited.

Some producers of fungal metabolites accept these restrictions as an inevitable consequence of cultivation of filamentous fungi. Others, however, are actively seeking forms of the producing microorganisms that are less "filamentous" and have much shorter hyphae, to reduce the problems somewhat. It is only reasonable to state, however, that these problems are still far from being overcome.

5 Filamentous Fungi and Heterologous Protein Production

The filamentous fungi are showing particular promise in the area of production of heterologous proteins. Various fungi have been used for producing a wide range of proteins (Table 3). Initially, many of the systems were clearly aimed at examining

Tab. 3 Heterologous protein production in molds: some hosts, promoters, and products.

Host	*Promoter source*	*Origin of promoter*	*Product*
Aspergillus oryzae	Amylase	*A. oryzae*	Aspartyl protease
A. niger var. *awamori*	Glucoamylase	*A. niger*	Calf chymosin
Trichoderma reesei	Cellobiohydrolase I	*T. reesei*	Calf chymosin
A. nidulans	Alcohol dehydrogenase	*A. niger*	Human tissue plasminogen activator, human interferon, human growth hormone, human interleukin 6
Achlya bisexualis	SV40 promoter	Simian virus	Interferon

the feasibility of using fungi as expression systems for "foreign" genes, and these early studies often involved fungi such a *N. crassa* and *A. nidulans*. More recently, a number of companies have achieved worthwhile yields of recombinant products.

It is generally acknowledged that the filamentous fungi have a number of actual or potential advantages in this role, including the following:

1. Filamentous fungi are permissive in relation to the expression of foreign genes.
2. Some filamentous fungi can secrete into the extracellular environment copious quantities of homologous proteins. Under the correct physiological conditions, for example, *A. niger* can produce tens of grams of glucoamylase per liter of culture broth. One should not underestimate the attractions of this feature; here we have a group of organisms whose particular mode of existence requires them to possess efficient protein-exporting machinery; thus, given point 1, they would appear to be ideal secretion systems for heterologous proteins. This initial promise of copious secretion of heterologous proteins has yet to be fully realized, having proved to be rather more difficult than was first thought. Some of the reasons for this are discussed later.
3. Posttranslational patterns of processing in the filamentous fungi are much closer to those of higher eukaryotes. Since many of the proteins of commercial interest are of mammalian origin (e.g. tissue plasminogen activator (tPA), interferons); this suggests that filamentous fungi may, at first sight, be more promising hosts than well-known prokaryotic expression systems such as *Escherichia coli*. Specifically, patterns of glycosylation in fungal proteins and higher eukaryotic proteins have a degree of similarity.
4. Inclusion bodies (accumulations of denatured foreign proteins in the cytoplasm of bacteria) are not seen in fungi. Their absence is potentially advantageous, since recovery of such proteins involves cell disruption, separation, and renaturation, all of which imply increased production costs.
5. From a regulatory viewpoint, many fungi make attractive hosts for the production of proteins for food (e.g. chymosin) or therapeutic use (e.g. tPA) because the US Food and Drug Administration has given them GRAS (generally recognized as safe) status.
6. Relative to mammalian and plant cell cultures, fungal cultures are fast growing (doubling times of 3–4 h are not atypical), and the technology for cultivating molds is well characterized and readily available. Downstream processing methods are also well understood.
7. Mitotic stability of transformants is high.

Conversely, a number of potential drawbacks exist, including the following:

1. Transformation frequency is generally very low compared to transformation in bacteria and yeasts, and methods usually involve protoplast generation.
2. Filamentous fungi are a level of complexity above unicellular organisms such as bacteria and yeast, and their development of a complex three-dimensional network, with protein secretion localized to specific areas of the mycelium only, may add to the problems of operation. Despite a long industrial history, there is still a great

deal that we do not understand about the physiology, and the genetics, of industrially important fungi.

3. Many fungi produce other by-products, such as organic acids, which represent a diversion of substrate away from protein synthesis and may alter the culture pH. In addition, many fungal strains produce proteases (such as aspergillopepsin A), which could well degrade the potential product. In some cases, production of proteases can be circumvented. For example, Genencor deleted the gene for aspergillopepsin A from the production strain of *A. niger*, thus increasing the secretion of calf chymosin.
4. Because fungal cultures are highly viscous and pseudoplastic, mixing in fermentor vessels is poor. Thus, even if the conditions required to optimize heterologous protein production are known, it may not be possible to ensure that these conditions prevail throughout the fermentor vessel. If uniformly good conditions are not achieved, the overall result will be reduced productivity.
5. The relationship between copy number of the transforming gene and production of the gene product is not a simple one in some species. Such complexity may be the result of the rather random nature of the integration process.

Despite the difficulties just enumerated, there have been successes also. Notable among these is the production of calf chymosin (Genencor) and phytase (Novo Nordisk). As can be seen in Table 3, there is also considerable interest in the production of therapeutic proteins using molds. The requirements here are more rigorous than those for chymosin and phytase, which one might regard as industrial-scale products. In the case of chymosin, the aim is to maximize the yield of the product using essentially the same technology used for production of native enzymes (see Sect. 2.2). The companies involved, and more importantly, the regulatory authorities, are familiar with these, and provided evidence of the purity and authenticity of the product is available, progress toward industrial production is relatively swift. It is to be expected that the companies involved in homologous protein production will increasingly use the fungal expression technology developed over the past decade or so to improve production of many industrial enzyme processes. The GRAS status of many of the chosen host strains will be helpful in this endeavor.

Different criteria may apply with regard to the therapeutic proteins produced using fungal expression systems. Yields at the milligram- or even the microgram-per-liter level may be initially acceptable here, when the "retail" price is so high and society's need so great. Conversely, one might reasonably expect the path to regulatory approval to be somewhat longer, since the requirements for the demonstration of consistent purity and authenticity of products will be much more rigorous, and molds have no track record as sources of therapeutic properties.

One aspect that has become apparent is that secretion of heterologous proteins from fungi rarely reaches the levels achievable with homologous protein. For example, calf chymosin from *A. niger* reaches about 1.2 gL^{-1}, while *A. niger* glucoamylase can reach up to 20 to 30 gL^{-1}. This discrepancy has been the subject of great interest, and its resolution may go far in ensuring the success of the fungi as industrially important expression systems.

5.1 Transcription

In general, it would appear that the process of heterologous protein secretion is not restricted at the transcriptional level. (One interesting exception is the transcriptional limitation on chymosin expression in *Trichoderma reesei*). In view of the highly effective promoters in current use, this is not surprising. However, the site of integration (where the foreign DNA is introduced into the genome) may affect the expression level. This area certainly requires further study.

5.2 Translation

Despite some evidence that specific sequences upstream of the initiation code may influence translation, it is not yet clear whether protein production is markedly affected at this stage.

5.3 Posttranslational Modifications

Many eukaryotic proteins are glycosylated, and the specific pattern of glycosylation may influence protein folding (and thus activity) and aid in passage through the secretory system. If incorrectly glycosylated, they may be very rapidly degraded in mammalian systems. Unlike yeasts, fungi generally do not appear to hyperglycosylate proteins (a notable exception is expression of *T. reesei* cellulases in *A. nidulans*). Fungal patterns of protein glycosylation are closer to those of higher eukaryotes than other microorganisms.

At present, the role of glycoslylation in the secretion process is unclear and requires attention.

5.4 Gene Fusion

This strategy has improved secretion levels of a number of heterologous proteins, including chymosin and human interleukin 6. In gene fusion, the foreign gene is fused with a native gene, whose product is efficiency exported. For efficient secretion, in eukaryotes, proteins usually require an N-terminal hydrophobic secretion signal peptide; likewise, other sequences lacking the signal peptide may influence the protein's passage through the secretory system. Fusion of the calf chymosin gene with the promoter and signal peptide sequence of the glucoamylase gene led to improved production of calf chymosin in *A. nidulans*. Where product assay and screening procedures are relatively simple, a program of mutagenesis can lead to improvements in secretion levels, presumably through alterations in the sequences lacking the signal region.

5.5 Proteolysis

The problem of proteolysis induced by native proteases may also be of significance. As one might expect, most homologous proteins secreted by fungi are not degraded by native proteases, but heterologous proteins can be rapidly degraded. The deletion of the aspergillopepsin A gene has already been mentioned, but intracellular proteolysis may restrict secretion levels in many systems.

5.6 Conclusions

We selected the filamentous fungi as expression systems for foreign genes because these microorganisms are excellent

protein secretors and we believed we understood them well. What is now apparent is that we do not yet understand why certain fungi are effective protein exporters. In other words, we did not understand them as well as we thought. Fundamental research into posttranslational aspects of protein trafficking and the secretory system of fungi is required if we are to fully capitalize on the abilities of this group of microorganisms as expression systems.

6
Biotechnology of Molds in Agriculture

The filamentous fungi have a very close relationship with many plants, which can take the form of a mutually beneficial link between plant and mold (a mycorrhiza). Alternatively, many fungi are plant pathogens and can cause diseases in commercially important plant species. Other aspects of fungal activity that may impact on agricultural practice are the abilities of some fungi to attack insects, mites, nematodes, and other fungi. These latter activities indicate that fungi may have potential as biocontrol agents.

6.1
Biocontrol Using Fungi

The use of fungi to control growth of undesirable plant species (weeds), pests (e.g. insects, mites, nematodes), and other plant pathogenic fungi has a number of attractions. The first of these is the potential for specific control of the problem species by selection of a very narrow range pathogenic fungus. Other potential advantages include a possible reduction in chemical residues in or on foods or feeds, through reduced use of chemical insecticides, herbicides, pesticides, and fungicides, and the possibility of minimizing damage to the ecosystem (a consequence of the first two points listed). Thus, destruction of the problem species should not be accompanied by elimination of beneficial species or unacceptable amounts of damage to them. One final potential advantage of using living agents such as fungi in a biocontrol program is the possibility of sustained control of the pathogen or pest by permanent introduction of the biocontrol agent.

Conversely, many potential drawbacks to the use of fungi may exist, including the lack of complete control achieved by many biocontrol agents alone; the slow rate of kill of the pathogen, which may allow significant crop damage; difficulties in mass production, storage, and delivery of the infective agent itself; and reduction in efficacy due to nonideal environmental conditions. From a producer's commercial viewpoint, the possibility of persistence of the biocontrol fungus, which would reduce future sales, must also be considered. Biocontrol agents are usually more expensive to produce than chemical agents, and a greater degree of sophistication is required in their use.

6.2
Mycoinsecticides

Broadly speaking, two classes of fungi can be distinguished, biotrophic (i.e. those that can replicate only in a living host) and those capable of growth in fermentation processes. Although more than 100 fungal genera have been shown to attack insect pests, few species have actually been widely used in the field. Prominent among those that have been so used are the deuteromycete fungi

Beauveria bassiana, *Verticillium lecanii*, and *Metarhizium anisopliae*. All these fungi have been shown to produce conidia, the infective form of the organism *in vitro*. *Beauveria bassiana* is used in the former Soviet Union and in China to control pests such as Colorado potato beetle, pine caterpillars, and European corn borer.

Mycoinsecticides usually attack by direct penetration of the insect cuticle by means of enzyme action and, to a lesser extent, mechanical pressure. Since the first step in infection is the contact between the fungal spore and the insect cuticle, activity of the mycoinsecticide is heavily influenced by relative humidity and temperature. The ability of some molds to effectively secrete extracellular depolymerizing enzymes such as chitinase, proteases, and lipases is a vital feature in their ability to parasitize living insects. Thus, the potential for improvement by the use of modern molecular biological techniques clearly exists. The battery of enzymes produced corresponds to the major constituents of the insect cuticle.

In addition to enzymes, entomopathogenic fungi produce a number of toxic metabolites that may have a role in the lethal effects observed *in vivo*. *Beauveria bassiana* produces a cyclic peptide, beauvericin, and *M. anisopliae* produces a range of five related cyclic peptides, the destruxins. The synthetase gene for a very similar fungal cyclic peptide (enhiantin) has been isolated and characterized. The potential now exists for increasing the efficacy of mycoinsecticide antibiotic synthetases as probes for the isolation of synthetase genes and for the development of genetically improved strains. A better understanding of the mechanism of action and the genetics of pathogenicity should allow us to make significant improvements in these biocontrol agents by means of a molecular biological approach.

6.3 Production Technology

Ideally, the aim is a production process to turn out large numbers of conidia rapidly, cheaply, and reproducibly. Processes involved are submerged (liquid) fermentation, solid substrate, and combined.

In the West, much of the fermentation capacity is in the form of STRs, and these have been used to produce *B. bassiana* and *V. lecanii*. Many fungi, however, do not conidiate in these systems; instead, they form blastospores. These cells can be used in formulation of the mycoinsecticide, but they are considerably less stable than conidia. Surface (SSF) culture can produce large numbers of conidia, but tends to be labor-intensive. It has been successfully used in Brazil, Russia, China, Canada, and the United States for *V. lecanii* and *B. bassiana*. In many countries, it is an especially attractive, low technology, low cost local solution to local problems. Indeed, the desired fungus can be grown on nonsterile agricultural wastes. Care must be taken, however, for the danger of allergic reaction to inhalation of spores may be significant.

Combined processes usually involve STR production of mycelial cultures followed by a step that achieves induction of conidiation.

6.4 Other Fungal Biocontrol Agents

Many of the preceding comments regarding difficulties of production, formulation, application, and stability apply equally to the use of fungi in other biocontrol areas. However, fungi have been and are

being used to control weeds in many countries. Mycoherbicidal species such as *Colleotrichum gleosporoides* (active component of Collego), used to control jointvetch, and *Phytophthora palmivora* (active component of Devine) have achieved some success, and many other potential mycoherbicides are in development worldwide. Difficulties with potential mycoherbicides, such as having too broad a host range, and loss of virulence complicate development, but extensive research into the use of these agents continues. Both classical mutation and selection techniques, and recombinant DNA technology, could be used to produce stable, highly virulent strains, which defy heretofore resistant pests; these would be ideal candidates for use in integrated pest management schemes. So far, only natural isolates are in use, but as our ability to modify organisms genetically improves, progress will be made in the development of these promising nonchemical agents.

Fungi, of course, are significant plant pathogens. In the late 1980s, fungal crop diseases were estimated to cause crop damage costing several billion dollars annually in the United States alone. Legislation aimed at reducing fungicide usage inevitably means that alternatives must be found. Fortunately, many fungal species can attack and kill other species, including plant pathogenic types. Species such as *Peniophora gigantea* and *Trichoderma viride* have demonstrated antifungal activity. Again, despite considerable research, the use of such agents is still at an early stage. Before rapid progress using gene technology can be made, there must be fundamental studies into the biochemistry, genetics, and ecology of these microorganisms to improve on the "natural" strains in current use.

One interesting recent extension of biocontrol using fungi is the use of selected fungal species to attack "undesirable" plants, principally, *Papaver somniferum* (the Oriental poppy, source of opium and heroin). The misused products of this plant are a major source of social disruption, illness, and death in Western society. In a logical extension of fungal weed control systems, the use of several fungal species to attack the Oriental poppy plant was investigated. Two fungal species showed promise in attacking otherwise healthy plants.

The utilization of these lab/pilot findings in an actual field/natural context is distant, but there clearly is potential for attacking some "weed crops" using the fungi. Better understanding of the genetic basis of plant pathogenesis would permit a structured program to enhance the biocontrol potential of fungi against undesirable "crop" species. The ethical issues surrounding a program to develop a bioweapon targeting a known evil are highly complex and will potentially retard progress here more than scientific/technological challenges.

Fungal biocontrol agents have great promise as elements of integrated pest management systems, in which they may be used in conjunction with chemical agents at much reduced levels. Development of such agents, which are currently of questionable economic viability, will be driven forward by public demand, since biological control is perceived as ecologically more acceptable than chemical control.

6.4.1 Future Prospects for Fungal Biocontrol

Our increasing ability to sequence fungal genomes in a reasonable timescale and deduce the role of individual genes in

the fungal life cycle could revolutionize the use of fungi in biocontrol. Greater understanding of mechanisms of plant, insect, or nematode pathogenesis by fungi at the genetic level implies an ability to construct strains of enhanced pathogenicity. Conversely, such research potentially opens the door to manipulation of the genome of fungal species directly harmful to humans, such as *Coccidiodes immitis* or *Blastomyces dermatidis*.

6.5 Ectomycorrhizal Fungi

Ectomycorrhizas, mutually beneficial close associations between the root system of plants and a fungus (or fungi), occur in about 10% of world flora. The ectomycorrhizal fungi generally improve plant growth, increase the available area for nutrient uptake from soil, reduce the chance of infection by root pathogenic fungi, and increase tolerance to drought, toxins, and extremes of pH.

Ectomycorrhizal associations are particularly valuable in many economically important forest species from families such as Pinaceae (pine, larch, fir) Betulaceae (alder, birch), and Fagaceae (oak, chestnut, beech).

Deliberate introduction of mycorrhizal fungi into forest or nursery soils to stimulate growth has many attractions. Three methods have been used: addition of soil containing the fungi, incorporation of sporophore material, and, most reliably, inoculation of seedlings with pure cultures of vegetative mycelia. The main problem with the last approach is the difficulty of large-scale axenic cultivation of ectomycorrhizal fungi in STRs.

In addition to their widespread use in commercial forestry plantations (about 7–8 million trees annually are being ''tailored'' with specific ectomycorrhizal fungal symbionts in the United States alone), mycorrhizal fungi may also be of value in the reclamation of landfill sites. Revegetation of such sites may be hindered by presence of toxic leachates and gases, and also by absence of suitable fungal symbionts. It has been shown, however, that when fungal symbionts are present, host-plant tolerance to physical and chemical stresses improves. Ectomycorrhizal fungi may also be of help in afforestation of sites containing industrial wastes, such as anthracite waste, metallic mine tailings, and coal spoils.

6.6 Prospects

The future of ectomycorrhizal fungi in reforestation and afforestation programs looks very bright. Biotechnology is used to enhance natural processes in this instance. The remaining difficulties concerning consistent large-scale cultivation are being resolved. Further studies into the maintenance of the infectivity of ectomycorrhizal fungi after cultivation *in vitro* are still required.

7 Fungal Biotransformations

Certain fungi can carry out highly specific transformations of complex organic molecules to produce molecules of high pharmaceutical potency and high optical purity. Because of their oxidative character, fungi have great advantages in oxygen-mediated biotransformation, especially when stereospecificity is a requirement.

7.1 Steroids from Fungi

The advantages of an oxidative nature were demonstrated in the transformation of steroids by fungi such as *Curvularia, Rhizopus,* and *Penicillium raistrickii*. These fungi can carry out highly specific modifications to steroids – for example, hydroxylations, dehydrogenations, and side chain degradations – which would be extremely difficult to carry out chemically. Such modifications can alter the pharmacological activity considerably. Examples include conversion of progesterone to 11-β-OH-cortexolone by *Curvularia*.

7.2 Other Biopharmaceuticals from Fungi

The search for pharmaceutically active nonantibiotic drugs continues unceasingly, but screening for activities such as immunostimulation or immunosuppresion and antihypercholesteremia is difficult compared to antibiosis screens. It is perhaps unsurprising then that two very successful fungal products were discovered recently almost by accident. Cyclosporin A, for example, was found to be a potent immunosuppressant. However, initially, the producing organism (*Tolypocladium*) was being examined for antifungal activity and toxicity to mosquito larvae; its true potential was discovered later. The identification of the antihypercholesteremic agent mevinolin, originally isolated as an antifungal agent, was similarly serendipitous.

The discovery that cyclosporin A inhibits a cytosolic peptidyl-prolyl isomerase activity may point the way toward a more specific screen for immunosuppressive drugs based on this enzyme.

Edible fungi have long been reported to have many beneficial effects on consumption. Recent research has shown that compounds such as lentinan (from *Lentinan edodes*) and scleroglucan (from *Sclerotium glucanicum*) may be potent antitumor agents. Both compounds are high molecular weight branched glucans. Lentinan has also been shown to be capable of boosting a depressed immune system under trial conditions. Molds represent a tremendous (largely untapped) reservoir of biopharmaceuticals; the major hindrance to the biotechnological development of this resource is the current lack of directed and effective screens for desirable activities.

7.3 Combinatorial Synthesis

Although the natural synthetic versatility of the fungi is still remarkably underexploited for reasons discussed above, novel routes to libraries of potentially useful chemical entities are possible via combinatorial synthesis. In the context of fungal biotechnology, this implies taking an identified chemical species secreted/produced by a fungus, and chemically and/or enzymically creating a series of related compounds with potentially different pharmacological and chemical characteristics.

Such "compound libraries" are ideal for screening via high-throughput screens. Whilst the isolation of novel antiinfectious agents still dominates here, other areas of application are developing rapidly; for example, several potent inhibitors of breast cancer resistance protein (BCRP) were identified amongst chemically modified derivatives of the natural fungal BCRP inhibitor fumitremorgin. This illustrates the clear potential of such combinatorial synthesis in generating potent new compounds.

8
Fungi and Toxic Materials

White rot fungi, which have the ability to degrade a vast range of hazardous organic compounds, may be of use in systems designed to reduce the environmental impact of waste streams containing such materials or to reclaim land containing such compounds (bioremediation).

A great deal of work has been carried out on the lignin-degrading fungus *Phanerochaete chrysosporium*, and evidence suggests that lignin peroxidases are responsible for degradation of some xenobiotics.

The list of xenobiotics that can be degraded by fungi lengthens constantly. Presently included are insecticides such as DDT, lindane, and wood preservatives (e.g. pentachlorophenol), through to chlorinated biphenyls and dioxins. Despite the demonstrated potential of fungal systems in xenobiotics degradation and in the treatment of kraft pulp wastes (the toxicity of which is largely due to the presence of chlorinated phenols, catechols, and guaiacols), much work relating to how these organisms mineralize the organic materials remains to be done.

By virtue of their large surface area and the presence of charged molecules on their surfaces (polysaccharides, proteins), filamentous fungi have the ability to bind metal ions in significant amounts. Several potential uses can be made of this capacity, including the cleaning of metal-containing waste streams, the concentration of desirable metal ions such as Ni^{2+}, and the concentration of ions of uranium and thorium from process streams generated by the nuclear industry. The advantages of removing up to 90% of the metal ions from such process strains and concentrating them in the fungal biomass, which can be further processed to reduce its volume, are obvious.

9
Genetic Improvement of Fungal Strains

Once a fungus expressing a desirable trait, such as excretion of valuable metabolites or potentiality for biocontrol, has been recognized and isolated, there begins almost immediately a search to improve that characteristic. While much can be achieved in terms of metabolic or physiological control, the primary focus for strain improvement is at the genetic level.

In many fungi of commercial significance, one or more of the following methods have been adopted for strain improvement.

1. Selective screening and selection.
2. Mutagenesis, screening, and selection.
3. Use of selective "breeding," sometimes by the use of the parasexual cycle, to combine desirable characteristics of two strains to give a novel genotype.
4. In the past decade or so, the ability to directly transform fungi by the introduction of a desired gene or genes has been developed.

Methods 1 and 2 have served biotechnology well. In the case of penicillin G, for example, they have contributed to increases in titer from the early isolates to current production strains of several thousandfold. They do, however, have limitations. In particular, in some systems, mutagenesis to increase metabolite synthesis may adversely affect other desired characteristics (e.g. strain vigor and fitness) that are especially important for strains used for biocontrol or as bioinoculants.

More importantly, a point of diminishing returns will be reached with respect to

the conventional methods; indeed, it has been reached for organisms such as *P. chrysogenum*.

The parasexual cycle, first recognized in *A. nidulans*, has been widely exploited to improve a range of industrial fungi by combining characteristics from related strains. The usefulness of this technique is limited by the existence of fungal incompatibility systems that restrict the occurrence of anastamosis to closely related strains. This obvious restriction on selective breeding (it narrows the genetic pool from which one can select) can sometimes be overcome by the generation of protoplasts. Such systems have been successfully applied to a range of fungal types including *T. harzianum*, *B. bassiana*, and *M. anisopliae*. Interspecies transfer is also a possibility.

At best, such methods are somewhat random even in selective breeding programs. Ideally, there is a need to specifically transfer the gene or genes responsible for the desired activity from one organism to another. The potentialities of such direct transformation have been illustrated not only in heterologous protein production but also in improvement of β-lactam-producing fungi by Lilly Research Laboratories.

Since the early studies on transformation of *A. nidulans* and *N. crassa*, the number of selective markers has steadily increased, and now a range of nutritional and dominant selective marks is available. This second group has been applied in transformation studies of many economically important metabolite- or protein-producing fungi and some plant pathogens. They include bleomycin, hygromycin, and mutant β-tubulin genes conferring benonyl resistance.

Fungal transformation commences with cell wall digestion in the presence of osmotic stabilizers, using preparations such as Novozym 234, sometimes in conjunction with other enzymes (e.g. driselase and glucuronidase). This stage may critically affect transformation frequency.

Transforming DNA has been shown to integrate into the chromosomal DNA. In *A. nidulans*, it has been possible to study direct/indirect gene replacement, and gene disruption effects.

In the work by Lilly on β-lactam producers, dominant selectable markers such as hygromycin, phleomycin, and *amid S* (acetamidase gene of *A. nidulans*) have been used in the transformation systems of *Cephalosporium acremonium* and *P. chrysogenum*. Using these systems, the investigators made progress in three distinct areas. In the first application, targeted gene disruption was used to locate and clone a β-lactam synthetic gene (*pcbAB*) in *C. acremonium*. The second application involved use of gene dosage to improve cephalosporin C titer in *C. acremonium*. Addition of one extra copy of the *cefEF* gene (which encodes for the enzyme catalyzing the rate-limiting step in cephalosporin C synthesis) led to a 15% increase in titer at the pilot scale. By comparison, mutagenesis and screening of very large numbers of derivatives of the parent strain failed to lead to the identification of any that were capable of increased productivity. This work demonstrates the great potency of the new techniques in molecular biology in overcoming identified metabolic limitations in fungi.

9.1 Pathway Engineering in *Penicillium chrysogenum*

Penicillium chrysogenum has long been used industrially to synthesize "natural" penicillins, G and V, and a family

of semisynthetic derivatives of differing antimicrobial range and potency. Cephalosporins are a related family of β-lactams traditionally synthesized by *Acremonium chrysogenum*. However, the industry would prefer to produce the cephalosporins in the more traditional and better-understood *P. chrysogenum*. Additionally, *Acremonium* undergoes a complex morphological change from (pseudo) filamentous to yeast-like, which influences the process outcome and is difficult to control.

Since the pathways of both antibiotics involve the common intermediate, isopenicillin N, introduction of a ring-expanding enzyme (expandase) would confer the potential on *Penicillium* to synthesize cephalosporin-type molecules instead of penicillins.

Thus, the final application involved biosynthetic pathway engineering, that is, the alteration of the β-lactam synthesis pathway to give a new end product, which is either an antibiotic with altered capabilities or a useful intermediate for further synthesis. This project involved the expression of the *cefE* gene of *Streptomyces clavuligerus* in *P. chrysogenum*, which led to significant synthesis of a desired enzyme (deacetoxycephalosporin C synthetase).

Taken together, these studies illustrate the tremendous capabilities of new gene technology to alter, in a planned fashion, the capabilities of existing fungal strains. Currently, such alterations require a major directed effort, with highly specific aims, and cannot realistically be justified unless they result in the improvement of an organism or the introduction of a process of great economic significance. As the technology matures, its applicability will broaden, and, undoubtedly, the bioengineering approaches discussed here will be used in many fungi of current interest.

10 Perspective

The use of molds in man's service has a long history. Conversely, their industrial exploitation can be measured in decades. These microorganisms occupy a central role in the fermentation industry and, in the near future, will continue to do so.

However, the future of fungal biotechnology will be heavily dependent on new product development and the development of new areas in which molds can be applied usefully. There are particularly promising developments in the areas of biotransformations, bioremediation using fungi, and use of filamentous fungi as systems for the expression and production of heterologous proteins. The application of the new techniques of gene transfer to filamentous fungi will leave no current application area untouched and may revolutionize our use of these techniques.

See also Gene Targeting.

Bibliography

Books and Reviews

Carlisle, M.J., Watkinson, S.C., Gooday, G.W. (2001) The Fungi, 2nd edition, Academic Press, London, UK.

Demain, A.L., Davies, J.E. (1999) *Manual of Industrial Microbiology*, 2nd edition, Blackwell Scientific, Oxford, UK.

El-Mansi, E.M.T., Bryce, C.F.A. (1999) *Fermentation Microbiology and Biotechnology*, Taylor & Francis, London, UK.

Gibbs, P.A., Seviour, R.J., Schmid, F. (2000) Growth of filamentous fungi in submerged culture, *Crit. Rev. Biotechnol.* **20**, 17–48.

McNeil, B., Harvey, L.M. (1990) *Fermentation: A Practical Approach*, IRL Press at Oxford University Press, Oxford, UK.

McNeil, B., Harvey, L.M. (1993) Viscous fermentation products, *Crit. Rev. Biotechnol.* **13**, 275–304.

Nielsen, J. (2002) Combinatorial synthesis of natural products, *Curr. Opin. Cell Membr. Biol.* **6**(3), 297–305.

Ratledge, C., Kristiansen, B. (2000) *Basic Biotechnology*, 2nd edition, Cambridge University Press, UK.

Stephanopoulos, G.N., Aristodou, A.A., Nielsen, J. (1998) *Metabolic Engineering*, Academic Press, London, UK.

Straathof, A.J.J., Adlercreutz, P. (2000) *Applied Biocatalysis*, 2nd edition, Academic Press, Switzerland.

Strohl, W.R. (1997) *Biotechnology of Antibiotics*, 2nd edition. Marcel Dekker, New York.

Primary Literature

Amanullah, A., Blair, R., Nienow, A.W., Thomas, C.R. (1999) Effects of agitation intensity on mycelial morphology and protein production in chemostat cultures of recombinant *Aspergillus oryzae*, *Biotechnol. Bioeng.* **62**, 434–446.

Archer, D.B., Peberdy, J.F. (1997) The molecular biology of secreted enzyme production by fungi, *Crit. Rev. Biotechnol.* **17**, 273–306.

Barker, T.W., Worgan, J.T. (1981) Application of airlift fermenters to cultivation of filamentous fungi, *Eur. J. Appl. Microbiol. Biotechnol.* **13**, 77–83.

Bai, Z., Harvey, L.M., McNeil, B. (2003) Physiological responses of chemostat cultures of *Aspergillus niger* to simulated and actual oxidative stress, *Biotechnol. Bioeng.* **82**, 627–639.

Catalanatto, C., Azzalin, G., Macino, G., Cogoni, C. (2000) Gene silencing in worms and fungi, *Nature* **404**, 245.

Cox, P.W., Paul, G.C., Thomas, C.R. (1998) Image analysis of the morphology of filamentous microorganisms, *Microbiology* **144**, 817–827.

Forster, J., Gombert, A.K., Nielsen, J. (2002) A functional genomics approach using metabolomics and in silico pathway analysis, *Biotechnol. Bioeng.* **79**, 703–712.

Gibbs, P.A., Seviour, R.J. (1998) Production of exopolysaccharides by *Aureobasidium pullulans* in fermenters with low shear configurations, *Appl. Microbiol. Biotechnol.* **49**, 168–174.

Jorgensen, H., Nielsen, J., Villadsen, J., Mollgaard, H. (1995) Metabolic flux distribution in *Penicillium chrysogenum* during fed-batch cultivation, *Biotechnol. Bioeng.* **46**, 558–572.

Kim, E.Y., Yoo, Y.J. (1992) Analysis of broth rheology with cell morphology in *Cephalosporium chrysogenum*, *Biotechnol. Tech.* **6**, 501–506.

Luengo, J.M. (1995) Enzymatic synthesis of hydrophobic penicillins, *J. Antibiotic.* **48**, 1195–1121.

Nielsen, J., Olsson, L. (2002) An expanded role for microbial physiology in metabolic engineering and functional genomics: moving towards systems biology, *FEMS Yeast Res.* **2**(2), 175–181.

Piper, M.D.W., Daran-Lapujade, P., Bro, C., Regenberg, B., Knudsen, S., Nielsen, J., Pronk, J.T. (2002) Reproducibility of oligonucleotide microarray transcriptome analyses – an interlaboratory comparison using chemostat cultures of *Saccharomyces cerevisiae*, *J. Biol. Chem.* **277**, 37001–37008.

Serrano-Carreon, L., Flores, C., Galindo, E. (1997) γ-decalactone production by *Trichoderma harzianum* in stirred bioreactors, *Biotechnol. Prog.* **13**, 205–208.

Stephanopoulos, G., Vallino, J.J. (1991) Network rigidity and metabolic flux engineering in metabolite overproduction, *Science* **252**, 1675–1681.

Thykaer, J., Christensen, B., Nielsen, J. (2002) Metabolic network analysis of an adipoyl-7-ADCA-producing strain of *Penicillium chrysogenum*: elucidation of adipate degradation, *J. Metab. Eng.* **4**, 151–158.

Trinci, A.P.J. (1994) Evolution of the Quorn mycoprotein fungus *Fusarium graminearum* A3/5, *Microbiology* **140**, 2181–2188.

Trinh, L.B., Phue, J.N., Shiloach, J. (2003) Effect of methanol feeding strategies on production and yield of recombinant mouse endostatin from *Pichia pastoris*, *Biotechnol. Bioeng.* **82**, 438–444.

Van Winden, W.A., van Gulik, W.M., Schipper, D., Verheijen, P.J.T., Krabben, P., Vinke, J.L., Heijnene, J.J. (2003) Metabolic flux and metabolic network analysis of *Penicillium chrysogenum* using 2D COSY NMR measurements and cumulative bondomer simulation, *Biotechnol. Bioeng.* **83**, 75–92.

Wang, Y., McNeil, B. (1995) Production of the fungal exopolysaccharide scleroglucan by cultivation of *Sclerotium glucanicum* in an airlift

reactor with external loop, *J. Chem. Technol. Biotechnol.* **63**, 215–222.

www.dsm.com/en/g_aspergilus_en.htm (2001) DSM determines DNA sequence of *Aspergillus niger* genome.

Zaldivar, J., Borges, A., Johansson, B., Smiths, H.P.O., Villas-Boas, S.G., Nielsen, J., Olsson, L. (2002) Fermentation performance and intracellular metabolite patterns in laboratory and industrial xylose-fermenting *Saccharomyces cerevisiae*, *Appl. Microbiol. Biotechnol.* **59**, 436–442.

Zhou, W., Holzhauer-Rieger, K., Bayer, T., Schugerl, K. (1993) Cephalosporin C production by a highly productive *cephalosporium acremonium* strain in an airlift tower loop reactor with static mixers, *J. Biotechnol.* **28**, 165–177.

Gaucher Disease, Molecular Biology of

Ernest Beutler
The Scripps Research Institute, La Jolla, CA, USA

1 **Definition and History** 24

2 **The Glucocerebrosidase Gene** 24

3 **Polymorphisms and Deficiency Mutations** 24

4 **Population Genetics** 25

Bibliography 27
Books and Reviews 27
Primary Literature 28

Keywords

Glucocerebroside
A glycolipid comprised of ceramide and glucose (glucosylceramide). This is the substance that accumulates in Gaucher disease.

Glucocerebrosidase
The enzyme that breaks down glucocerebroside.

Pyruvate kinase
A glycolytic enzyme.

Polymorphism
A mutation that is found in at least 1% of the genes in a population.

Pseudogene
A gene that has undergone changes so that it no longer is able to function.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

1
Definition and History

Gaucher disease is the most common of the glycolipid storage disorders. First described in 1882 by Phillipe Charles Ernest Gaucher (1854–1918) in his doctoral thesis as an "epithelioma of the spleen," it is now recognized to be the consequence of the deposition of glucosyl ceramide (glucocerebroside) in the macrophages of the body. This glycolipid is a breakdown product of globosides and gangliosides, complex glycolipids that are constituents of the normal cell membrane. Normally, the enzyme glucocerebrosidase cleaves glucocerebroside into glucose and ceramide, but in Gaucher disease, a deficiency of glucocerebrosidase leads to accumulation of glucosyl ceramide.

Clinically, three major types of Gaucher disease have been delineated on the basis of the absence (type 1) or presence and severity (types 2 and 3) of primary central nervous system involvement. Type 1 disease is sometimes designated as the "adult" form of the disease, but since type 1 disease often becomes apparent in early childhood, this is a misleading term. Within each type, even within the same ethnic groups, the phenotypes and genotypes can be markedly heterogenous.

2
The Glucocerebrosidase Gene

The glucocerebrosidase gene (*GBA*) is located on chromosome 1 at q21. It is approximately 7 kb in length and contains 11 exons. It is in a region that contains several genes that are of medical interest, including the red cell and liver pyruvate kinase gene (*PKLR*), the thrombospondin gene, and the metaxin gene. A 5-kb pseudogene is located about 16 kb downstream from the glucocerebrosidase gene (Fig. 1). The pseudogene has maintained a high degree of homology with the functional gene. Although it is transcribed, an unusual property for a pseudogene to possess, it does not contain a long open reading frame and has a 55-bp deletion from what was once the coding region. Alu sequences that have been inserted into introns give the functional gene greater length than the pseudogene. The putative TATA- and CAAT-like boxes of the promoter have been identified about 260 bp upstream from the upstream ATG start codon.

The cDNA is about 2 kb in length. There are two ATGs at the 5′ end and both are utilized in translation. The relative importance or function of these two start sites is unknown. Messenger RNA of several different lengths has been detected, probably because of the existence of alternative polyadenylation sites, alternative splicing, or presence of pseudogene mRNA.

3
Polymorphisms and Deficiency Mutations

Eleven polymorphic sites are known to exist in the introns and flanking regions of the acid ß-glucosidase gene, but somewhat surprisingly only 4 haplotypes have been found. These sites and the haplotypes that have been identified so far are summarized in Table 1. A considerable number of point mutations that cause glucocerebrosidase deficiency, and hence Gaucher disease, have been identified. Some of the more common of these are summarized in Table 2. An up-to-date list can be found

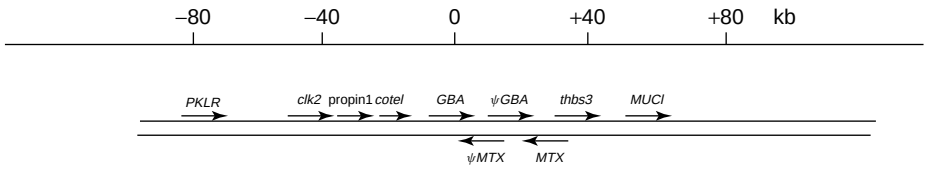

Fig. 1 A schematic representation of the genes surrounding the glucocerebrosidase (*GBA*) gene. The arrows show the direction of transcription.

Tab. 1 Nucleotides at 12 positions in the 4 glucocerebrosidase haplotypes.

Designation	*Frequency*	*Glucocerebrosidase haplotypes*											
		nt	*−802*	*−725*	*−614*	*2128*	*2834*	*3297*	*3747*	*3854*	*3931*	*4644*	*5135*	*6144*
1+	Common		A	C	C	A	C	G	G	T	G	del	C	G
2−	Common		G	T	T	G	G	A	G	C	A	A	A	A
3 African	Common African		G	T	T	G	A	A	G	C	A	A	A	A
4 Uncommon	Rare		A	C	C	A	C	G	A	T	G	del	C	G

at http://www.tau.ac.il/~racheli/genedis/gaucher/gaucher.html.

Some genes with the 1448C mutation have been found to contain other mutations, each corresponding to the sequence of the pseudogene. In some cases, this type of mutation has been shown to represent a crossover between the functional gene and the pseudogene, with loss of genetic material between the two. In other cases, the mechanism by which the abnormal mRNA has been formed has not been elucidated, and the alleles have merely been referred to as *complex alleles, pseudopattern, or rec* (for recombinant). The mechanism by which these are formed may all be identical to that demonstrated to occur in the XOVR mutation or may represent the result of gene conversion events.

Some of the mutations are relatively mild in their phenotypic effect and have been associated only with type 1 disease. Others are more severe and are also observed in type 2 or type 3 neuronopathic disease. In some instances, the number of cases that have been observed is so limited that the severity of the mutation is not known. A severity score has been devised that allows one to express the clinical severity of patients with Gaucher disease. As shown in Fig. 2, patients who are homozygous for the 1226G mutation tend to have relatively mild disease with a late onset, while patients who also carry one of the more drastic mutations have more severe disease with early onset. Homozygotes for the c.1342 G→C (D409H) present a unique clinical picture that includes calcification of cardiac valves and corneal opacities. The mechanism of this unusual disease phenotype, observed only with this mutation, is unknown.

4 Population Genetics

Of the many mutations that have now been documented, only three appear to approach polymorphic frequencies. The

Tab. 2 Gaucher disease mutations.

Point mutations that cause Gaucher disease						
cDNA[a]	***Amino acid***[a]	***Genomic***[a]	***Nucleotide substitution***	***Amino acid substitution***	***Population effect***	***Frequency***[b]
IVS2+1		1067	G→A[d]	–	?Severe	Uncommon
754	213	3548	T→A[c]	Phe→Ile	Severe	Uncommon
1226	370	5841	A→G	Asn→Ser	Mild	Common
1297	394	5912	G→T	Val→Leu	Severe	Uncommon
1342	409	5957	G→C[c]	Asp→His	Severe	Uncommon
1448	444	6433	T→C[c]	Leu→Pro	Severe	Common
1504	463	6489	C→T	Arg→Cys	Mild	Uncommon
1604	496	6683	G→A	Arg→His	Mild	Uncommon
Insertions and deletions that cause Gaucher disease						
84		1035	G→GG	–	Severe	Common
1263–1317 del		5879–5933[c] del		?	Severe	Uncommon

Recombination events that cause Gaucher disease			
Location of crossover event(s)[d]		***Effect***	***Population frequency***
>1343 <1388	>5957 <6272[e]	Severe	Uncommon
>1317 <1343	>5932 <5957	Severe	Uncommon
>1343 <1388	>5957 <6272	Severe	Uncommon

[a]Nucleotide or amino acid position.
[b]Common = high frequency in at least one population; Uncommon = found in a number of unrelated patients.
[c]Pseudogene sequence.
[d]Only approximate ranges can be given, since the pseudogene and functional gene contain long identical sequences.
[e]Physical fusion with loss of intergenic segment.

most common is the A→G transition, at nt 1226, which produces a protein with an Asn→Ser substitution at amino acid 370 of the mature protein. It is present in about 6% of the Ashkenazi Jewish population. This mutation is the principal cause of the high incidence of Gaucher disease in this ethnic group. An insertion of a guanine at nt 84 of the cDNA is the second common Jewish mutation. It is found in approximately 0.6% of the Jewish population. Because it produces a frameshift even before the amino terminus of the mature protein, this drastic mutation produces no enzyme protein. A T→C transition at nt 1448, producing a Leu→Pro substitution at amino acid 444 of the mature protein, is present at a relatively high frequency in the Norrbottnian population of Northern Sweden. The same mutation also occurs in other populations at low frequencies. It is

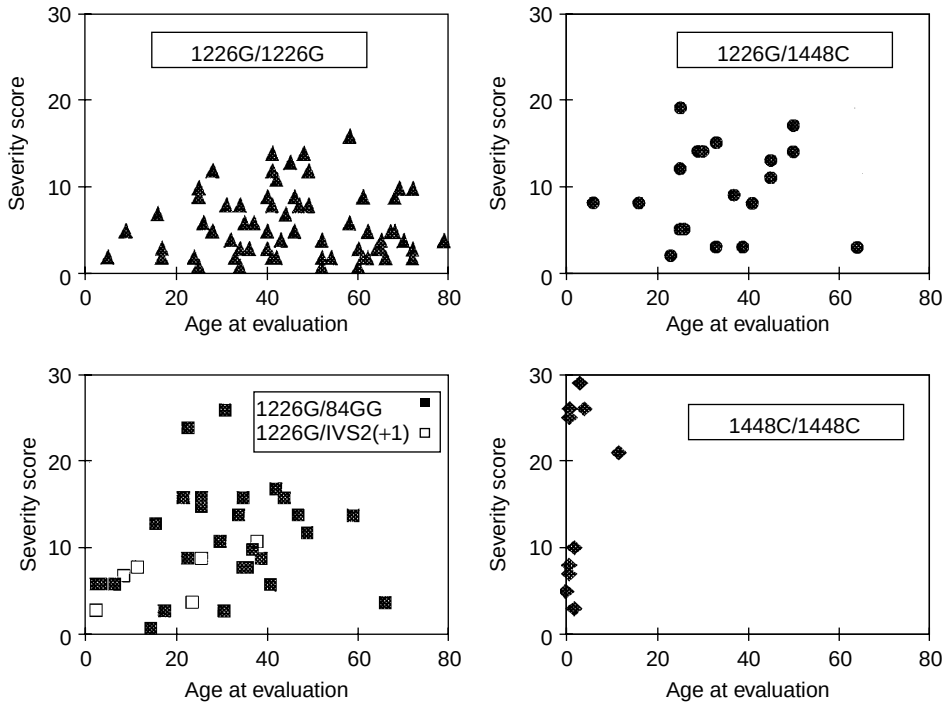

Fig. 2 The relationship between the age at the time of evaluation and the severity score of patients with four different Gaucher disease genotypes. The most severe disease is seen in the 1448C homozygotes, while many of the patients who were homozygous for the 1226G mutation are elderly and have very mild disease.

noteworthy that the homologous position in the pseudogene is occupied by a C, just as in the Gaucher disease–producing mutation.

The maintenance of a high degree of linkage disequilibrium between the 1226G mutation, common in the Jewish population, and the nearby *PKLR* gene implies that the 1226G mutation arose relatively recently, and estimates of its time of origin range from several hundred to several thousand years ago.

See also Genetics, Molecular Basis of; Medicinal Chemistry.

Bibliography

Books and Reviews

Balicki, D., Beutler, E. (1995) Gaucher disease, *Medicine (Baltimore)* **74**, 305–323.

Beutler, E. (1992) Gaucher disease: new molecular approaches to diagnosis and treatment, *Science* **256**, 794–799.

Beutler, E. (1993) Gaucher disease as a paradigm of current issues regarding single gene mutations of humans, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5384–5390.

Beutler, E. (1993) Modern diagnosis and treatment of Gaucher's disease, *Am. J. Dis. Child.* **147**, 1175–1183.

Beutler, E., Gelbart, T. (1996) Mutation update: glucocerebrosidase (Gaucher disease), *Hum. Mutat.* **8**, 207–213.

Beutler, E., Gelbart, T. (1998) Hematologically important mutations: Gaucher disease, *Blood Cells Mol. Dis.* **24**, 2–8.

Beutler, E., Grabowski, G. (2001) Gaucher Disease, in: Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D. (Eds.) *The Metabolic and Molecular Bases of Inherited Disease*, McGraw-Hill, New York, pp. 3635–3668.

Grabowski, G.A. (1993) Gaucher disease, enzymology, genetics, and treatment, *Adv. Hum. Genet.* **21**, 377–441.

Grabowski, G.A., Saal, H.M., Wenstrup, R.J., Barton, N.W. (1996) Gaucher disease: a prototype for molecular medicine, *Crit. Rev. Oncol. Hematol.* **23**, 25–55.

Primary Literature

Abrahamov, A., Elstein, D., Gross-Tsur, V., Farber, B., Glaser, Y., Hada-Halpern, I., Ronen, S., Tafakjdi, M., Horowitz, M., Zimran, A., Hadas-Halpern, I. (1995) Gaucher's disease variant characterised by progressive calcification of heart valves and unique genotype, *Lancet* **346**, 1000–1003.

Barton, N.W., Brady, R.O., Dambrosia, J.M., Di Bisceglie, A.M., Doppelt, S.H., Hill, S.C., Mankin, H.J., Murray, G.J., Parker, R.I., Argoff, C.E., Grewal, R.P., Yu, K.-T. (1991) Replacement therapy for inherited enzyme deficiency – macrophage-targeted glucocerebrosidase for Gaucher's disease, *N. Engl. J. Med.* **324**, 1464–1470.

Beutler, E. (1988) Gaucher disease, *Blood Rev.* **2**, 59–70.

Beutler, E., Dale, G.L., Guinto, E., Kuhl, W. (1977) Enzyme replacement therapy in Gaucher's disease: preliminary clinical trial of a new enzyme preparation, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4620–4623.

Beutler, E., Gelbart, T. (1993) Gaucher disease mutations in non-Jewish patients, *Br. J. Haematol.* **85**, 401–405.

Beutler, E., Gelbart, T., Kuhl, W., Sorge, J., West, C. (1991) Identification of the second common Jewish Gaucher disease mutation makes possible population based screening for the heterozygous state, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10544–10547.

Beutler, E., Gelbart, T., Kuhl, W., Zimran, A., West, C. (1992) Mutations in Jewish patients with Gaucher disease, *Blood* **79**, 1662–1666.

Beutler, E., Kay, A., Saven, A., Garver, P., Thurston, D., Dawson, A., Rosenbloom, B. (1991) Enzyme replacement therapy for Gaucher disease, *Blood* **78**, 1183–1189.

Beutler, E., Kuhl, W. (1986) Glucocerebrosidase processing in normal fibroblasts and in fibroblasts from patients with type I, type II, and type III Gaucher disease, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7472–7474.

Beutler, E., Kuhl, W., Sorge, J. (1984) Cross-reacting material in Gaucher disease fibroblasts. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6506–6510.

Beutler, E., Kuhl, W., Sorge, J. (1985) Glucocerebrosidase "processing" and gene expression in various forms of Gaucher disease, *Am. J. Hum. Genet.* **37**, 1062–1070.

Beutler, E., Nguyen, N.J., Henneberger, M.W., Smolec, J.M., McPherson, R.A., West, C., Gelbart, T. (1993) Gaucher disease: gene frequencies in the Ashkenazi Jewish population, *Am. J. Hum. Genet.* **52**, 85–88.

Beutler, E., West, C., Torbett, B.E., Deguchi, H. (2002) A chimeric mouse model of Gaucher disease, *Mol. Med.* **8**, 247–250.

Bornstein, P., McKinney, C.E., LaMarca, M.E., Winfield, S., Shingu, T., Devarayalu, S., Vos, H.L., Ginns, E.I. (1995) Metaxin, a gene contiguous to both thrombospondin 3 and glucocerebrosidase, is required for embryonic development in the mouse: Implications for Gaucher disease, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4547–4551.

Brady, R.O., Kanfer, J.N., Bradley, R.M., Shapiro, D. (1966) Demonstration of a deficiency of glucocerebroside-cleaving enzyme in Gaucher's disease, *J. Clin. Invest.* **45**, 1112–1115.

Brady, R.O., Kanfer, J.N., Shapiro, D. (1965) Metabolism of glucocerebrosides II. Evidence of an enzymatic deficiency in Gaucher's disease, *Biochem. Biophys. Res. Commun.* **18**, 221–225.

Brady, R.O., Pentchev, P.G., Gal, A.E., Hibbert, S.R., Dekaban, A.S. (1974) Replacement therapy for inherited enzyme deficiency. Use of purified glucocerebrosidase in Gaucher's disease, *N. Engl. J. Med.* **291**, 989–993.

Boas, F.E. (2000) Linkage to Gaucher mutations in the Ashkenazi population: effect of drift on decay of linkage disequilibrium and evidence for heterozygote selection, *Blood Cells Mol. Dis.* **26**, 348–359.

Chabas, A., Cormand, B., Grinberg, D., Burguera, J.M., Balcells, S., Merino, J.L., Mate, I.,

Sobrino, J.A., Gonzalez-Duarte, R., Vilageliu, L. (1995) Unusual expression of Gaucher's disease: cardiovascular calcifications in three sibs homozygous for the D409H mutation, *J. Med. Genet.* **32**, 740–742.

Colombo, R. (2000) Age estimate of the N370S mutation causing Gaucher disease in Ashkenazi Jews and European populations: a reappraisal of haplotype data, *Am. J. Hum. Genet.* **66**, 692–697.

Conradi, N.G., Sourander, P., Nilsson, O., Svennerholm, L., Erikson, A. (1984) Neuropathology of the Norrbottnian type of Gaucher disease: morphological and biochemical studies, *Acta Neuropathol. (Berl.)* **65**, 99–109.

Cox, T., Lachmann, R., Hollak, C., Aerts, J., van Weely, S., Hrebicek, M., Platt, F., Butters, T., Dwek, R., Moyses, C., Gow, I., Elstein, D., Zimran, A. (2000) Novel oral treatment of Gaucher's disease with N-butyldeoxynojirimycin (OGT 918) to decrease substrate biosynthesis, *Lancet* **355**, 1481–1485.

Dahl, N., Lagerström, M., Erikson, A., Pettersson, U. (1990) Gaucher disease type III (Norrbottnian type) is caused by a single mutation in exon 10 of the glucocerebrosidase gene, *Am. J. Hum. Genet.* **47**, 275–278.

Dale, G.L., Villacorte, D., Beutler, E. (1976) Solubilization of glucocerebrosidase from human placenta and demonstration of a phospholipid requirement for its catalytic activity, *Biochem. Biophys. Res. Commun.* **71**, 1048–1053.

Diaz, A., Montfort, M., Cormand, B., Zeng, B., Pastores, G.M., Chabas, A., Vilageliu, L., Grinberg, D. (1999) Gaucher disease: the N370S Mutation in Ashkenazi Jewish and Spanish patients has a common origin and arose several thousand years ago, *Am. J. Hum. Genet.* **64**, 1233–1238.

Elstein, D., Hadas-Halpern, I., Itzchaki, M., Lahad, A., Abrahamov, A., Zimran, A. (1996) Effect of low-dose enzyme replacement therapy on bones in Gaucher disease patients with severe skeletal involvement, *Blood Cells Mol. Dis.* **22**, 104–111.

Eyal, N., Wilder, S., Horowitz, M. (1990) Prevalent and rare mutations among Gaucher patients, *Gene* **96**, 277–283.

Grabowski, G.A., Gatt, S., Horowitz, M. (1990) Acid β-glucosidase: enzymology and molecular biology of Gaucher disease, *CRC Crit. Rev. Biochem. Mol. Biol.* **25**, 385–414.

Ginns, E.I., Choudary, P.V., Tsuji, S., Martin, B., Stubblefield, B., Sawyer, J., Hozier, J., Barranger, J.A. (1985) Gene mapping and leader polypeptide sequence of human glucocerebrosidase: implications for Gaucher disease, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7101–7105.

Glenn, D., Gelbart, T., Beutler, E. (1994) Tight linkage of pyruvate kinase (*PKLR*) and glucocerebrosidase (*GBA*) genes, *Hum. Genet.* **93**, 635–638.

Harinck, H.I.J., Bijvoet, O.L.M., van der Meer, J.W.H., Jones, B., Onvlee, G.J. (1984) Regression of bone lesions in Gaucher's disease during treatment with aminohydroxypropylidene bisphosphonate, *Lancet* **2**, 513.

Heitner, R., Elstein, D., Aerts, J., Van, W., Zimran, A. (2002) Low-dose N-butyldeoxynojirimycin (OGT 918) for type I Gaucher disease, *Blood Cells Mol. Dis.* **28**, 127–133.

Hobbs, J.R. (1985) Bone marrow transplantation for genetic diseases, *N. Engl. J. Med.* **312**, 1260.

Hollak, C.E.M., van Weely, S., van Oers, M.H.J., Aerts, J.M.F.G. (1994) Marked elevation of plasma chitotriosidase activity. A novel hallmark of Gaucher disease, *J. Clin. Invest.* **93**, 1288–1292.

Horowitz, M., Tzuri, G., Eyal, N., Berebi, A., Kolodny, E.H., Brady, R.O., Barton, N.W., Abrahamov, A., Zimran, A. (1993) Prevalence of nine mutations among Jewish and non-Jewish Gaucher disease patients, *Am. J. Hum. Genet.* **53**, 921–930.

Horowitz, M., Wilder, S., Horowitz, Z., Reiner, O., Gelbart, T., Beutler, E. (1989) The human glucocerebrosidase gene and pseudogene: structure and evolution, *Genomics* **4**, 87–96.

Hyun, J.C., Misra, R.S., Greenblatt, D., Radin, N.S. (1975) Synthetic inhibitors of glucocerebroside β-glucosidase, *Arch. Biochem. Biophys.* **166**, 382–389.

Kerem, E., Elstein, D., Abrahamov, A., Ziv, Y.B., Hadas-Halpern, I., Melzer, E., Cahan, C., Branski, D., Zimran, A. (1996) Pulmonary function abnormalities in type I Gaucher disease, *Eur. Respir. J.* **9**, 340–345.

Latham, T., Grabowski, G.A., Theophilus, B.D.M., Smith, F.I. (1990) Complex alleles of the acid β-glucosidase gene in Gaucher disease, *Am. J. Hum. Genet.* **47**, 79–86.

Lieberman, J., Beutler, E. (1976) Elevation of serum angiotensin-converting enzyme in

Gaucher's disease, *N. Engl. J. Med.* **294**, 1442–1444.

Liu, Y., Suzuki, K., Reed, J.D., Grinberg, A., Westphal, H., Hoffmann, A., Döring, T., Sandhoff, K., Proia, R.L. (1998) Mice with type 2 and 3 Gaucher disease point mutations generated by a single insertion mutagenesis procedure (SIMP), *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2503–2508.

Morimoto, S., Martin, B.M., Yamamoto, Y., Kretz, K.A., O'Brien, J.S., Kishimoto, Y. (1989) Saposin A: second cerebrosidase activator protein, *Proc. Natl. Acad. Sci, U.S.A.* **86**, 3389–3393.

Nolta, J.A., Sender, L.S., Barranger, J.A., Kohn, D.B. (1990) Expression of human glucocerebrosidase in murine long-term bone marrow cultures after retroviral vector-mediated transfer, *Blood* **75**, 787–797.

Pàmpols, T., Pineda, M., Girós, M.L., Ferrer, I., Cusi, V., Chabás, A., Sanmarti, F.X., Vanier, M.T., Christomanou, H. (1999) Neuronopathic juvenile glucosylceramidosis due to *sap*-C deficiency: clinical course, neuropathology and brain lipid composition in this Gaucher disease variant, *Acta Neuropathol. (Berl.)* **97**, 91–97.

Patrick, A.D. (1965) Short communications: a deficiency of glucocerebrosidase in Gaucher's disease, *Biochem. J.* **97**, 17C–18C.

Pennelli, N., Scaravilli, F., Zacchello, F. (1969) The morphogenesis of Gaucher cells investigated by electron microscopy, *Blood* **34**, 331–347.

Pentchev, P.G., Brady, R.O., Blair, H.E., Britton, D.E., Sorrell, S.H. (1978) Gaucher disease: Isolation and comparison of normal and mutant glucocerebrosidase from human spleen tissue, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3970–3973.

Ponce, E., Moskovitz, J., Grabowski, G. (1997) Enzyme therapy in Gaucher disease type I: effect of neutralizing antibodies to acid β-glucosidase enzyme therapy in Gaucher disease type 1: effect of neutralizing antibodies to acid beta-glucosidase, *Blood* **90**, 43–48.

Radin, N.S. (1996) Treatment of Gaucher disease with an enzyme inhibitor, *Glycoconjugate J.* **13**, 153–157.

Rappeport, J.M., Ginns, E.I. (1984) Bone-marrow transplantation in severe Gaucher disease, *N. Engl. J. Med.* **311**, 84–88.

Sano, A., Radin, N.S., Johnson, L.I., Tarr, G.E. (1988) The activator protein for glycosylceramide β-glucosidase from guinea pig liver: improved isolation method and complete amino acid sequence, *J. Biol. Chem.* **263**, 19597–19601.

Sato, Y., Beutler, E. (1993) Binding, internalization, and degradation of mannose-terminated glucocerebrosidase by macrophages, *J. Clin. Invest.* **91**, 1909–1917.

Sawkar, A.R., Cheng, W.-C., Beutler, E., Wong, C.-H., Balch, W.E., Kelly, J.W. (2002) Chemical chaperones increase the cellular activity of N370S β-glucosidase: a therapeutic strategy for Gaucher's disease. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15428–15433.

Shafit-Zagardo, B., Devine, E.A., Smith, M., Arredondo, G.A.F., Desnick, R.J. (1981) Assignment of the gene for acid beta-glucosidase to human chromosome 1, *Am. J. Hum. Genet.* **33**, 564–575.

Sibille, A., Eng, C.M., Kim, S.-J., Pastores, G., Grabowski, G.A. (1993) Phenotype/genotype correlations in Gaucher disease type I: clinical and therapeutic implications, *Am. J. Hum. Genet.* **52**, 1094–1101.

Sorge, J., Gelbart, T., West, C., Westwood, B., Beutler, E. (1985) Heterogeneity in type I Gaucher disease demonstrated by restriction mapping of the gene, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5442–5445.

Sorge, J., Gross, E., West, C., Beutler, E. (1990) High level transcription of the glucocerebrosidase pseudogene in normal subjects and patients with Gaucher disease, *J. Clin. Invest.* **86**, 1137–1141.

Sorge, J., West, C., Westwood, B., Beutler, E. (1985) Molecular cloning and nucleotide sequence of the human glucocerebrosidase gene, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7289–7293.

Tsuji, S., Martin, B.M., Barranger, J.A., Stubblefield, B.K., LaMarca, M.E., Ginns, E.I. (1988) Genetic heterogeneity in type 1 Gaucher disease: multiple genotypes in Ashkenazic and non-Ashkenazic individuals, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2349–2352, 5708.

Tybulewicz, V.L.J., Tremblay, M.L., LaMarca, M.E., Willemsen, R., Stubblefield, B.K., Winfield, S., Zablocka, B., Sidransky, E., Martin, B.M., Huang, S.P., Mintzer, K.A., Westphal, H., Mulligan, R.C., Ginns, E.I. (1992) Animal model of Gaucher's disease from

targeted disruption of the mouse glucocerebrosidase gene, *Nature* **357**, 407–410.

van Weely, S., Brandsma, M., Strijland, A., Tager, J.M., Aerts, J.M.F.G. (1993) Demonstration of the existence of a second, non-lysosomal glucocerebrosidase that is not deficient in Gaucher disease, *Biochim. Biophys. Acta* **1181**, 55–62.

Wigderson, M., Firon, N., Horowitz, Z., Wilder, S., Frishberg, Y., Reiner, O., Horowitz, M. (1989) Characterization of mutations in Gaucher patients by cDNA cloning, *Am. J. Hum. Genet.* **44**, 365–377.

Zimran, A., Gelbart, T., Beutler, E. (1990) Linkage of the Pvu II polymorphism with the common Jewish mutation for Gaucher disease, *Am. J. Hum. Genet.* **46**, 902–905.

Zimran, A., Sorge, J., Gross, E., Kubitz, M., West, C., Beutler, E. (1989) Prediction of severity of Gaucher's disease by identification of mutations at DNA level, *Lancet* **2**, 349–352.

Zimran, A., Sorge, J., Gross, E., Kubitz, M., West, C., Beutler, E. (1990) A glucocerebrosidase fusion gene in Gaucher disease. Implications for the molecular anatomy, pathogenesis and diagnosis of this disorder, *J. Clin. Invest.* **85**, 219–222.

Gel Electrophoresis, 2D-difference

John F. Timms
Ludwig Institute for Cancer Research, UCL Branch, London, WC1E 6BT, UK

1 Principles 34
1.1 Proteomics by 2-DE Differential Protein Expression Analysis 34
1.2 Methods of Protein Visualization: Problems and Pitfalls 35
1.3 Fluorescence Detection in 2-DE Expression Profiling 37

2 Fluorescence 2D-difference Gel Electrophoresis (2D-DIGE) 38
2.1 Background 38
2.2 Fluorescent Cyanine Dye Derivatives and Sample Labeling 38
2.3 Validation and Sensitivity of 2D-DIGE 40
2.4 Cy dye-Induced Molecular Weight Shifts, Poststaining and Spot Picking 42
2.5 Fluorescence Detection and Image Analysis 43
2.6 Applications 45

3 Perspectives 46

Bibliography 48
Books and Reviews 48
Primary Literature 48

Keywords

Proteomics
The characterization of biological systems, processes, or disease states in terms of quantitative and qualitative changes in proteins.

Two-dimensional Gel Electrophoresis (2-DE)
A polyacrylamide gel-based method for the separation of protein mixtures and/or differentially modified protein isoforms according to their charge and molecular weight.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

2-DE Differential Protein Expression Analysis
The comparison of protein expression levels between two or more biological states by 2-DE and postseparation protein visualization and quantitation.

Fluorescence 2D-difference Gel Electrophoresis (2D-DIGE)
A 2-DE-based differential protein expression analysis technique involving the derivatization of multiple protein samples with spectrally distinct fluorescent dyes, allowing separation and quantitation within the same 2-DE gel.

NHS-Cy2/Cy3/Cy5 Cyanine Dyes
A charge- and size-matched set of spectrally distinct fluorescent cyanine dye derivatives used in 2D-DIGE for the differential labeling of primary amine groups in multiple protein samples.

Fluorescence two-dimensional difference gel electrophoresis (2D-DIGE) is a recently developed 2D-gel-based proteomics technique that can be applied to provide a sensitive, rapid, and quantitative analysis of differential protein expression between two or more biological samples. This technique utilizes chemical derivatives of spectrally distinct fluorescent cyanine dyes (Cy2, Cy3, and Cy5) to covalently label different samples prior to mixing and separation on the same 2-DE gels. This differential labeling and mixing of samples helps avoid the gel-to-gel variation inherent in conventional 2-DE-based expression profiling and reduces the number of gels that need to be run. Since the technique is also compatible with downstream identification of protein spots from gels by mass spectroscopy, it is particularly useful for reproducible, high-throughput proteomics.

1 Principles

1.1 Proteomics by 2-DE Differential Protein Expression Analysis

Proteomics is the study of biological processes or disease states in terms of qualitative and quantitative changes in proteins. This includes the monitoring of changes in protein expression, activity, cellular or tissue distribution, modification, and interactions with other proteins, biomolecules, or drugs. Of these, the ability to determine significant alterations in protein expression that correlate with disease or occur consequent to induced changes in cells is fundamental to the exploitation of proteomics.

Historically, the most commonly applied method for this differential protein expression analysis is two-dimensional gel electrophoresis (2-DE), and despite its inherent detection and resolution limits (see Sect. 2), it remains the best and the most amenable technique for the global analysis of protein expression. This technique involves the physical separation of proteins

from biological samples using denaturing polyacrylamide gel electrophoresis (PAGE). Separation in the first dimension, or isoelectric focusing (IEF), is based upon the native charge of each protein. Proteins will migrate through a polyacrylamide gel with an applied pH gradient and electric field until they "focus" at the pH at which they are neutrally charged, their isoelectric point (pI). The first dimension is then applied to a second dimension polyacrylamide gel that is used to further resolve the proteins according to their relative molecular weight when complexed with the anionic detergent SDS, a technique known as sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The separation works on the principle that SDS binds to proteins with a constant stoichiometry, such that smaller SDS–protein complexes can migrate faster through the molecular "sieve" created by the polyacrylamide compared to the larger ones. Separated protein mixtures are then visualized directly in 2-DE gels, or on nitrocellulose or polyvinylidene difluoride (PVDF) membranes following electrophoretic transfer. Visualization can be achieved using various protein staining methods, immunodetection with specific antibodies, or by radioactive detection, when samples are first metabolically labeled with radioisotopes. Relative protein expression levels are then deduced by quantitation of protein staining (or labeling) using scanning densitometry or other imaging techniques. Significant differences in expression are then determined by comparing values across the different sample sets after computational superimposition of gel images and matching of protein spots or gel features.

In recent years, 2-DE expression profiling has been linked with the downstream identification of differentially expressed proteins directly from excised gel pieces. The analytical methods applied mainly rely on identification of peptides extracted from gel pieces following site-specific proteolysis. While it is possible to identify extracted peptides by direct Edman sequencing, mass spectrometric methods have come to the fore as the most sensitive and rapid means of identifying low-level proteins. The two main mass spectrometry (MS) techniques applied to protein identification from gels are peptide mass fingerprinting by matrix-assisted laser-desorption ionization mass spectrometry (MALDI-MS) linked to peptide mass database searching, and electrospray ionization-collision-induced dissociation tandem mass spectrometry (ESI-CID MS/MS) with fragment ion or sequence tag database searching. ESI-CID MS/MS usually incorporates front-end separation of the peptide digest by reverse-phase liquid chromatography, a technique abbreviated to LC-MS/MS. Since LC-MS/MS analysis provides sequence information on the sample, it allows the unambiguous identification of proteins as long as they are already represented in a database. This combination of 2-DE and MS provides the major driving force for high-throughput differential protein expression analysis in proteomics.

1.2 Methods of Protein Visualization: Problems and Pitfalls

Radiolabeling is the most sensitive method for protein visualization in 2-DE-based expression analysis and undoubtedly has the broadest dynamic range of all detection methods. Metabolic radiolabeling also has the advantage that it can provide temporal as well as quantitative information on protein synthesis, turnover, and posttranslational modifications. Despite

these advantages, metabolic radiolabeling is both hazardous and expensive and can result in unwanted DNA fragmentation, cell growth arrest, and even cellular apoptosis. For these reasons, many researchers have moved away from metabolic radiolabeling as a technique for expression profiling, and instead rely more on conventional protein staining strategies.

Over the years, many different kinds of noncovalent protein staining methods have been employed for visualization and quantitation of proteins in 2-DE gels or on membranes following electroblotting. The majority of these staining methods rely on dyes that interact directly with proteins, for example, Coomassie Brilliant Blue (CBB) R-250, colloidal CBB G-250, amido black, and the fluorescent dyes SYPRO Ruby and SYPRO Rose. Other sensitive direct staining methods include acidic silver nitrate and alkaline silver diamine staining of gels and colloidal gold staining of proteins electroblotted onto membranes. Indirect staining can also be achieved using fluorescent dyes such as Nile Red, SYPRO Orange, SYPRO Red, and SYPRO Tangerine, which interact with the SDS micelle around the protein. Alternatively, reverse staining of gels with zinc-imidazole, or potassium, copper, or zinc chloride–staining methods creates a semiopaque gel background, with proteins detected as transparent spots when viewed with appropriate back illumination.

Staining with CBB and silver are the most commonly used methods for comparative quantitation in 2-DE gels, and CBB staining still remains one of the most reliable and heavily used detection methods owing to its compatibility with downstream Edman sequencing and mass spectrometry. Although conventional CBB staining is relatively insensitive, colloidal CBB staining is capable of detecting as little as 20 ng of protein. Silver staining, although sensitive (2–4 ng), is not particularly compatible with subsequent microchemical analysis. This is largely due to the inclusion of aldehyde-based fixatives in staining procedures. These fixatives not only reduce the recovery of peptides from in-gel protein digests, but also result in protein alkylation that complicates Edman sequencing and peptide mass profiling by MALDI-MS. A solution to this problem has been the development of more MS-compatible silver-staining techniques that omit aldehydes and only superficially stain proteins at the gel surface. However, these modified staining protocols have decreased sensitivity and give higher levels of background staining. In addition, reduction of spectral intensity and inferior sequence coverage in peptide mass fingerprinting experiments have been reported using these protocols, although silver-destaining methods can be employed to improve mass spectrometric analysis.

The use of CBB, silver, and other nonfluorescent staining methods in differential protein expression analysis also suffers from two other major drawbacks. Firstly, these methods have relatively small linear dynamic ranges of detection, and thus are not ideal for the proteomic analysis of complex biological samples in which the dynamic range of protein abundance can be up to eight orders of magnitude. Secondly, staining by these methods can vary quite considerably from protein to protein and from gel to gel. This is particularly the case for silver staining in which temperature, the critical timing of the different steps in the staining protocols, and the lack of a staining end-point leads to appreciable variation in staining intensity from gel to gel.

1.3 Fluorescence Detection in 2-DE Expression Profiling

Improvements to proteomic protein visualization are now largely focused on fluorescence detection. Since measurement of light emission is intrinsically more sensitive than measurement of light absorbance, fluorescent protein stains have the potential to allow detection of very low levels of protein. In addition, fluorescent stains offer a much broader linear dynamic range of detection, making them more suitable for expression profiling of cell and tissue samples. Recently, the commercially available noncovalent fluorescent SYPRO stains from Molecular Probes, Inc. have been used for high sensitivity–protein visualization in gels and on membranes.

Of these stains, the ruthenium-based metal chelate fluorescent stain SYPRO Ruby appears to be the most sensitive for in-gel protein visualization. It is reported to detect down to 2 ng of protein (as good as the best silver-staining methods), and is an end-point stain, making it impossible to overstain gels. SYPRO Ruby binds to proteins in a manner similar to CBB, but has a linear dynamic range extending over three orders of magnitude, better than any nonfluorescent staining strategy. In addition, SYPRO Ruby staining has been reported to be compatible with downstream identification of proteins by MALDI-MS peptide mass profiling. These properties suggest that SYPRO Ruby staining is currently the best method for global protein detection and identification in 2-DE gels, although this is yet to be shown by the wider proteomics community.

Recently, several multiplexed fluorescence approaches have been reported in which general protein staining with SYPRO Ruby is combined in the same gel with an additional fluorescence detection method, using a fluorophore with a different emission maximum. These other staining methods are used to probe for particular functionalities within the sample and have included the identification of glycoproteins by Schiff's base staining with the green fluorescent hydrazide dye Pro-Q Emerald 300, monitoring of in-gel β-glucuronidase activity using a fluorogenic substrate, and the identification of penicillin-binding proteins using BODIPY derivatives of penicillin.

A number of different fluorophores have been used to covalently derivatize proteins within gels or on membranes. These include fluorescein isothiocyanate, 5-(4,6-dichlorotriazin-2-yl) aminofluorescein, monobromobimane, 2-methoxy-2,4-diphenyl-3(2H)furanone (MDPF), and dansyl chloride. However, none of these are routinely applied for differential expression profiling. More recently, several approaches have been used to covalently label proteins prior to 2-DE separation or following IEF but prior to SDS-PAGE. These methods have used monobromobimane and 5-iodoacetamidofluorescein to label the cysteine sulfhydryl groups of reduced proteins, or MDPF and 5-(4,6-dichlorotriazin-2-yl) aminofluorescein to label ε-amino groups of lysine residues and N-terminal α-amino groups. Perhaps the most promising recent development in this kind of strategy is the use of two or three spectrally distinct fluorescent cyanine dyes (Cy2, Cy3, and Cy5), which allows differential labeling of multiple samples for separation and comparison on the same 2-DE gel. This technique, known as *2D-difference gel electrophoresis*, is discussed in detail in the next section.

2 Fluorescence 2D-difference Gel Electrophoresis (2D-DIGE)

2.1 Background

Conventional 2-DE gel-based differential protein expression analysis requires comparison of protein levels across multiple samples (see Sect. 1.1). However, inherent variations in the way proteins resolve in one 2-DE gel versus another makes image superimposition and rigorous quantitation of differential expression a difficult and inaccurate process. Such gel-to-gel variations arise for a number of reasons, including heterogeneity across polyacrylamide gels, differential thermal fluctuations, electric and pH field differences and variable precipitation of samples in the first dimension. This inherent variation in the technique means that there is an absolute requirement for separation of multiple replicate samples and image analysis/spot matching across many gels in order to measure statistically significant alterations in spot intensities. This process makes conventional 2-DE expression profiling very labor-intensive, particularly when many different samples are to be compared. Moreover, as described above, some of the commonly used methods of protein detection give variable staining from gel to gel, and lack the sensitivity and dynamic linear range of detection for expression profiling of complex biological samples.

To overcome these problems, Unlu and coworkers devised a methodology for detecting differences in two protein extracts run on the same 2-DE gel. This technique, known as *fluorescence 2D-difference gel electrophoresis*, utilized chemical derivatives of spectrally distinct fluorescent cyanine dyes (Cy3 and Cy5) for labeling two different samples prior to separation on the same 2-DE gel. The resolved, labeled proteins are then detected at the appropriate excitation and emission wavelengths for Cy3 and Cy5 using a multiwavelength fluorescence detection device (see Sect. 2.5). In this way, the two samples are subjected to the same procedures and microenvironments during separation, thus raising the confidence with which protein changes between samples can be detected and quantified, as well as reducing the number of gels that must be run.

2.2 Fluorescent Cyanine Dye Derivatives and Sample Labeling

The cyanine dye derivatives used in 2D-DIGE are *N*-hydroxy-succinimidyl (NHS) esters of propyl-Cy3 and methyl-Cy5 (Fig. 1). These dye derivatives will undergo nucleophilic substitution reactions with primary amino groups (i.e. ε-amino groups of lysine residues and N-terminal α-amino groups) in a protein sample, resulting in covalent modification with the cyanine moiety and loss of the NHS leaving group. These reagents were designed so as not to perturb the relative electrophoretic mobility of differentially labeled proteins during 2-DE. Thus, the dyes are both mass and charge matched, and carry a single positive charge that matches that of the lysine residue to be modified. The addition of the relatively hydrophobic cyanine moiety to proteins was reported to cause sample precipitation that was dependent upon the extent of labeling. Thus, optimal conditions were established for "minimal" lysine labeling, where an estimated 1 to 2% of all lysine residues in a whole-cell lysate were labeled without detectable effect on protein solubility. This effectively means

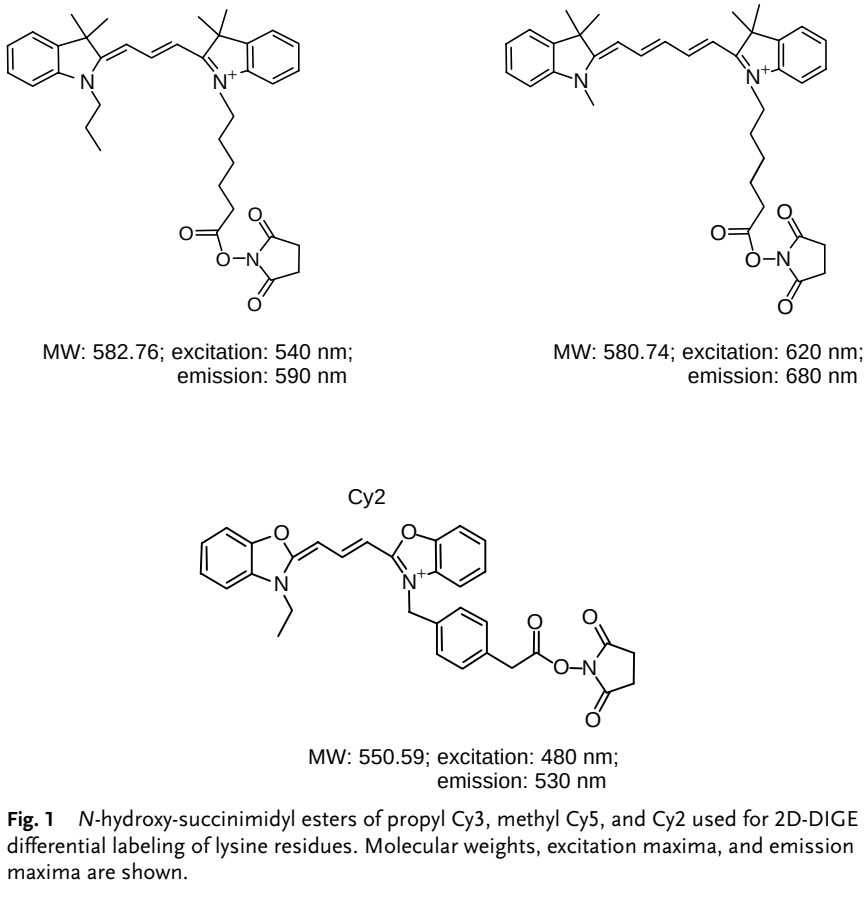

Fig. 1 *N*-hydroxy-succinimidyl esters of propyl Cy3, methyl Cy5, and Cy2 used for 2D-DIGE differential labeling of lysine residues. Molecular weights, excitation maxima, and emission maxima are shown.

that a minority of proteins is labeled on a single lysine residue, while the remainder is unlabeled. Reactions were carried out in an IEF-compatible lysis buffer (8 M urea, 2.5% CHAPS, 5 mM DTT and 5 mM HEPES) at pH 8.0 using 16 μmol dye per 1 μg total protein for 30 min on ice. Unreacted dye was inactivated by addition of an excess of 1,3-diamino-2-hydroxy-propane. Using this two-color 2D-DIGE methodology, the authors were able to successfully detect a single protein difference between two complex protein mixtures at the low nanogram level. This "minimal" labeling is also advantageous as it is unlikely to interfere with downstream in-gel tryptic digestion and MS identification of peptides.

The 2D-DIGE technology was acquired by Amersham Biosciences and is commercially available as the Ettan™ DIGE system. A third spectrally distinct cyanine dye, Cy2, has been introduced (Fig. 1), allowing simultaneous analysis of three samples on one gel. In an optimized protocol, labeling is performed using 4 to 8 pmol of dye per microgram of protein with an estimated dye: protein labeling ratio of <5%. Labeling is performed on ice in the dark for 30 min and any unreacted dye is inactivated by addition of a 50-fold molar excess

of free lysine for 10 min. The samples are then reduced with DTT and IEF sample buffer is added to the desired volume prior to 2-DE separation. Thus, labeling takes approximately 45 min, making it faster than any other staining method. In typical expression profiling experiments using the three-dye system, one of the dyes is reserved for labeling of an internal standard, usually comprising of a pool of equal amounts of all the test samples that are to be compared (Fig. 2). This internal normalization standard is then run on all gels against pairs of the test samples labeled with the other two dyes. Compared to the usual method of global normalization, normalization using the same standard on multiple gels allows direct comparison of the ratios of expression of a particular protein across the multiple samples. This greatly increases the ability to distinguish biological variation from gel-to-gel variation.

2.3 Validation and Sensitivity of 2D-DIGE

In a very rigorous validation of the method, Tonge et al. determined the quantitative variability in the 2D-DIGE process by applying it to mouse liver homogenates from control versus paracetamol-dosed animals. Using ImageMaster software, the authors established statistically valid thresholds for assigning quantitative changes between samples and found that these thresholds depended on the normalized spot volume, with lower volumes having significantly higher thresholds. This variation was correlated with the efficiency and sensitivity by which ImageMaster defines spot borders, which are less clear for low-intensity spots. Despite this, these thresholds were found to be unaffected by the Cy dye combination used or the individual operator, properties that demonstrate the robustness of the 2D-DIGE system. As for its application to expression profiling, Tonge et al. were able to identify significant changes in the expression of numerous proteins between control and paracetamol-treated mice, some of them novel, and also demonstrated that there was more variation between individual animals than was contributed by the error inherent to the 2D-DIGE protocol.

In another study using protein standards, it was reported that the linear dynamic range of detection for DIGE extended over five orders of magnitude, making it better than any other nonradioactive detection method. These authors also used "spiking" experiments to show that quantitation by DIGE was linear over at least three orders of protein abundance. The linearity and sensitivity of DIGE labeling was also tested by Amersham Biosciences on protein standards, and is shown to give a linear response from 1 to 10 000 ng (four orders of magnitude) and a detection limit of 0.25 to 0.95 ng, making it slightly more sensitive than silver or SYPRO Ruby staining. Despite this, we have reported that SYPRO Ruby detection on 2-DE gels reveals around 1.4 times the number of gel features than Cy dye labeling (Fig. 3 and. In this case, the Cy dye and SYPRO Ruby imaging was performed on the same imaging device using maximum exposure times without saturation of the most intense signals. However, by increasing exposure times it is possible to increase the overall sensitivity while saturating only a few of the signals from the more abundant protein spots.

A modified version of the original 2D-DIGE method has also been reported using the succinimidyl ester Alexa 488 and Alexa 532 fluorescent dye kits from Molecular Probes. Here, the degree of

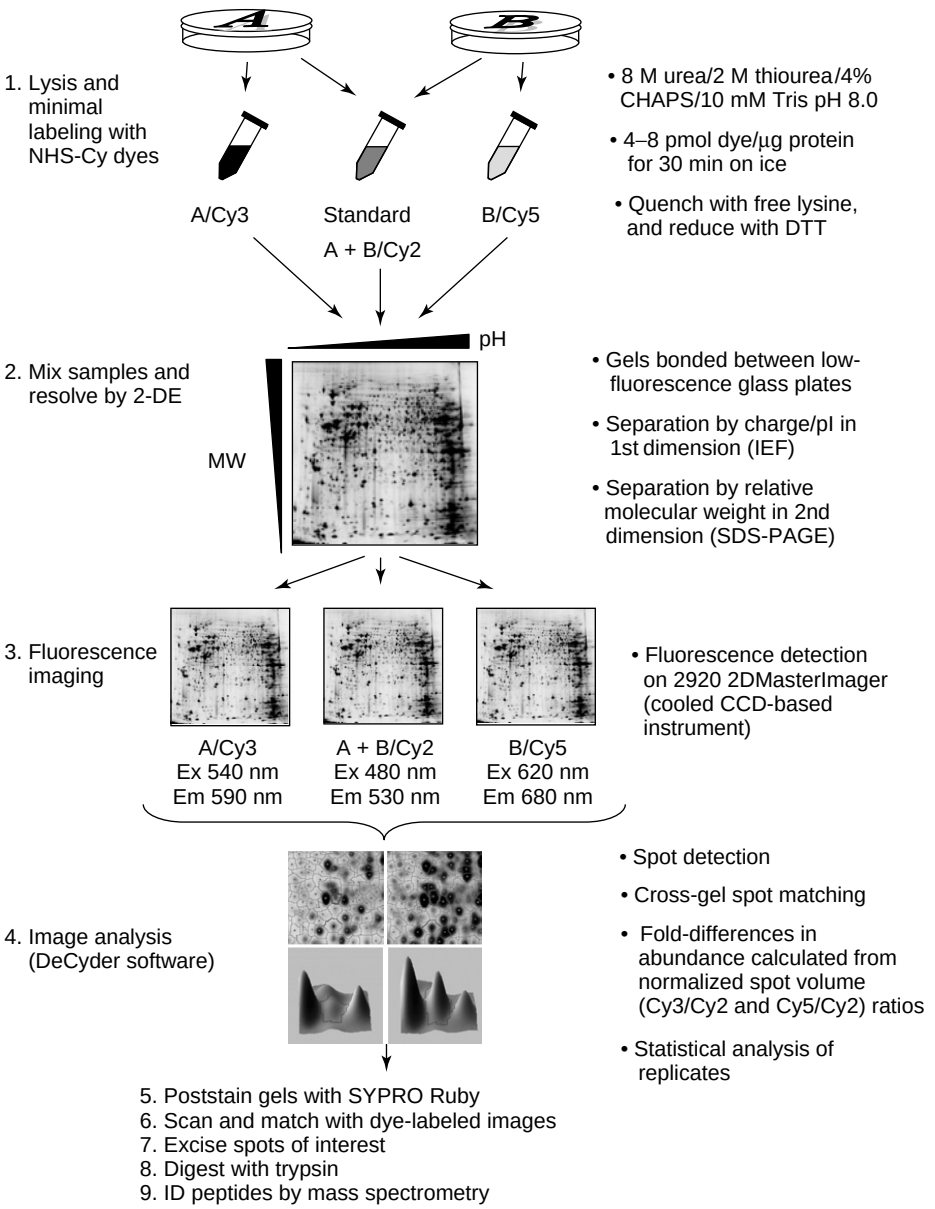

Fig. 2 Schematic representation of the 2D-DIGE protocol for minimal lysine labeling and using an internal standard for normalization.

labeling was calculated to be about 4-mol dye per mol protein and the sensitivity was reported to be similar to that obtained using a silver-staining kit (BioRad). A spiking experiment was used to show the linearity of the system and the ease with which differences could be identified using only simple image analysis software such

as Adobe Photoshop. Taken together, these studies demonstrate the robustness of the 2D-DIGE technique.

2.4 Cy dye-Induced Molecular Weight Shifts, Poststaining and Spot Picking

While minimal lysine labeling does not significantly alter the pI of proteins in 2-DE gels, it does cause a noticeable and somewhat variable shift in molecular weight because of the added mass of the dye (~460 Da). This mass shift, where the small percentage of labeled protein migrates more slowly than the corresponding majority of unlabeled protein, can be visualized by superimposition of the fluorescence 2-DE gel image with the corresponding poststained image (e.g. SYPRO Ruby stained). The shift is particularly noticeable in the lower molecular weight region of the gel. (Fig. 3). Although this molecular weight shift does not pose a problem for analytical 2D-DIGE comparisons of spot intensities (since samples are labeled to the same extent), accurate excision of a protein from preparatory gels may be compromised, particularly in the low molecular weight range. This problem can be overcome by cutting a larger region of interest from the gel, or better still, by poststaining gels to visualize the more abundant unlabeled protein spots. Thus, the sensitivity of 2D-DIGE is limited by the need to limit the modification to single lysine residues to reduce the changes in apparent molecular weight between modified

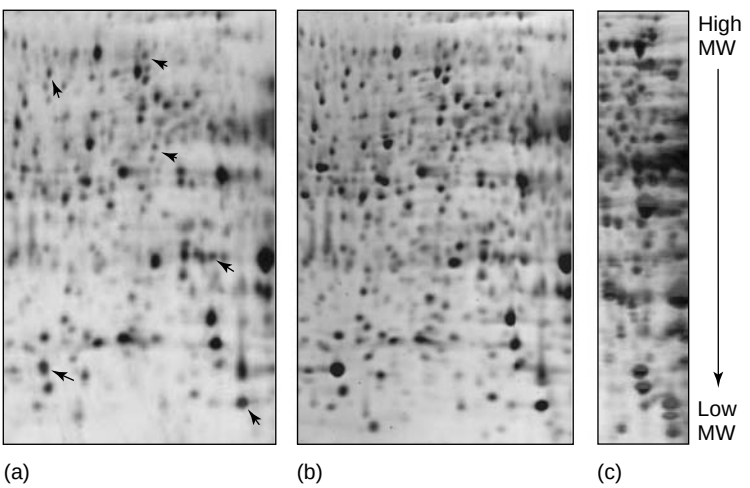

Fig. 3 (a) Merged 2-DE images of Cy3-labeled normal mammary luminal epithelial cell lysate (red) and Cy5-labeled ErbB-2 overexpressing luminal epithelial cell lysate (blue). Arrows indicate some of the obvious differences in expression between the two cell lines, (b) SYPRO Ruby poststaining of the same gel showing detection of a greater number of gel features. (c) Superimposition of a Cy dye gel image (blue) and SYPRO Ruby poststained gel image (red) revealing the shift in molecular weight due to dye labeling (figure adapted from Gharbi, S., Gaffney, P., Yang, A., Zvelebil, M.J., Cramer, R., Waterfield, M.D., Timms, J.F. (2002) Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system, *Mol. Cell. Proteomics* **1**, 91–98). (See color plate p. xxi.)

and unmodified proteins, so as not to compromise downstream identification of proteins by MS.

We, and others, have found that post-staining of dye-labeled proteins in 2D-DIGE gels with CBB, silver, or SYPRO Ruby is not compromised, and facilitates the spot-picking process. However, problems have been reported with silver or CBB poststaining of proteins labeled with Alexa dyes. This is likely to have been caused by masking of proteins by the bound dyes, which were used at a higher stoichiometry of labeling than the Cy dyes. SYPRO Ruby seems to be the best-suited method for poststaining since it is both sensitive and compatible with downstream MS identification of excised protein spots (see Sect. 1.3). It is also worth noting that bonding of gels to glass plates using bind-silane (see Sect. 2.5), severely reduces the sensitivity of MS-compatible silver staining, whereas the end-point SYPRO Ruby staining is unaffected. Moreover, SYPRO Ruby detection can be performed on the same fluorescence detection device as the Cy dye detection.

2.5 Fluorescence Detection and Image Analysis

Instrumentation for fluorescence detection of covalently derivatized or non-covalently stained proteins is currently based on either laser/photomultiplier tube (PMT) scanning or lamp/charge-coupled device (CCD)-based mechanisms. Laser/PMT-based instruments include the Molecular Dynamics™ Typhoon™ 9400 Variable Mode Imager and the Fuji Photo Film Company FLA-2000 Fluorescent Image Analyzer, while xenon arc lamp/CCD-based instruments include the Perkin Elmer Life Sciences Wallac-1442 ARTHUR™ Multiwavelength Fluoroimager and a variant of this, the Amersham Biosciences 2920 2DMasterImager™. A schematic of the CCD-based 2920 2DMasterImager™ is represented in Fig. 4. The CCD-based instruments tend to be more versatile since wavelength selection for different fluorophores can be customized using appropriate emission and excitation filters and do not rely on the output of a laser source. Although both types of imaging system can deliver resolutions as high as 50 μm, which is more than adequate for detection of 2-DE gel spots, some high-resolution CCD devices must work by collecting multiple images over the sample that are then stitched back together to form a composite of the whole image. This strategy increases the scanning time significantly, but can also create a patchwork effect of different background intensities across the image, particularly when the signal intensity is low.

In terms of efficiency of workflow, gels are best cast and run between low-fluorescence glass plates and then imaged between the plates directly after the second dimension run (Figs. 2 and 4). This prevents drying of gels during long imaging times and sets a fixed position for all images from a particular set of gels. As mentioned above, gels are best poststained prior to picking, because of the molecular weight shift caused by dye attachment. However, direct spot-picking without fixing and poststaining may increase the yield of recovered peptides from gel pieces for MS, as long as the spots can be accurately excised. We have also found that it is optimal to bond gels to one of the glass plates with bind-silane at the casting stage. This approach maintains the gels in virtually the same fixed position after poststaining, facilitating alignment of Cy dye and poststained images and allowing for safer

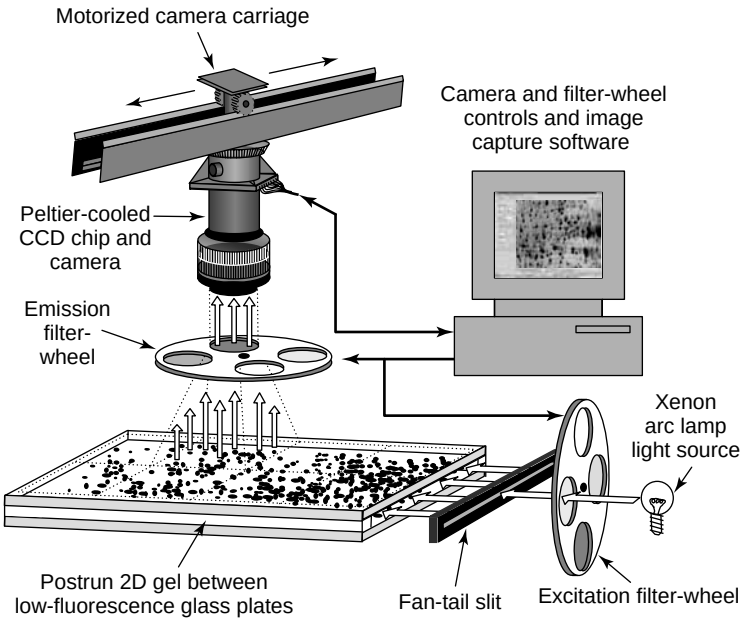

Fig. 4 Schematic representation of a xenon lamp/CCD-based fluorescence imager (based on the Amersham Biosciences 2920 2DMasterImager™). The gel is imaged *in situ* between low-fluorescence glass plates directly after the 2-DE run. Images are collected separately for the three Cy dyes using filter wheels for the appropriate excitation and emission wavelengths. SYPRO Ruby poststained gels are scanned on the same instrument, again using the appropriate light filters.

gel storage. Bonding is also an absolute necessity for most automated spot-picking platforms; however, it is important to note that milder fixing conditions must be employed when staining bonded gels in order to prevent overshrinkage and tearing of gels.

Numerous commercially available software packages have been used for spot detection and to match and evaluate spot intensity or volume differences across multiple gels. While all require some level of user intervention resulting in loss of reproducibility, their accuracy is still dependent on the inherent variability of 2-DE separations and protein staining. It is also important to be aware that there is variability in spot detection using different image analysis packages, so it is important for an investigator to determine prediction thresholds for their particular system (see Sect. 2.3). Although not available at the time of the original validation, 2D-DIGE-image analysis is most easily performed using DeCyder Differential Analysis Software (Amersham Biosciences). This software is split into two modules. The differential in-gel analysis (DIA) module is first used to detect overlapping, differentially labeled spots within the same gel, performing automatic background subtraction, quantitation, and normalization with the internal standard. The biological variation analysis (BVA) module then matches detected spots across gels and performs a statistical analysis (Student *t*-test

and 1-way and 2-way ANOVA analyses) to find consistent differences across replicate samples. This software package thus avoids the need for the investigator to set prediction limits, and is a far cry from a previous study in which a looped two-frame movie of Cy3 and Cy5 images was used to investigate proteomic changes in cardiac mitochondria associated with loss of creatine kinase. Although this was not a good way to quantify changes in protein abundance due to the lack of normalization, this method does demonstrate the simplicity of the DIGE process. A further advantage of the DeCyder software is that all matched spots are given the same spot identifier from gel to gel. This greatly assists in data analysis for multiple samples and in subsequent sample tracking, handling, and MS identification.

2.6 Applications

DeCyder software was used in a three-dye 2D-DIGE differential expression analysis studying the oncogenic effects of ErbB-2 overexpression and growth factor treatment in human mammary luminal epithelial cells. Protein expression over time (4 time points of growth factor treatment) and between two cell lines (normal versus ErbB-2 overexpressing) was simultaneously examined in triplicate in a single experiment. Thus, eight triplicate cell lysates (24 samples) were labeled with Cy3 or Cy5 and run against a Cy2 standard comprising a pool of all 24 samples (36 samples in total). After appropriate mixing, 12 gels were run with 300 μg of total protein, generating 36 images for DeCyder analysis. Although spot matching was not perfect across all gels for all features (i.e. less than 36 out of 36 matches for all spots), the analysis identified changes in the abundance (by spot volume) of numerous features with time and/or between cell lines that were significant within a 95% confidence limit ($p < 0.05$). Moreover, statistically valid differences could be detected for features that were only 1.3-fold up- or downregulated in ErbB-2 overexpressing cells versus normal. Spots of interest were excised from the same dye-labeled gels following SYPRO Ruby poststaining, imaging, and alignment, using a Syprot automated spot picker (Amersham Biosciences). Around 30 of the differentially regulated protein features were identified by MALDI-MS. While several of these proteins may contribute to ErbB-2-dependent transformation, the others are likely to be deregulated as a secondary consequence of the enhanced proliferation and morphological changes resulting from ErbB-2 overexpression.

2D-DIGE and DeCyder analysis were also applied to study the effects of benzoic acid treatment on the *Escherichia coli* proteome. Although only Cy3 and Cy5 were used, Cy3 was used for labeling of a pooled internal standard, and run against Cy5 labeled untreated or benzoic acid–treated cells. Using this internal standard method, a more accurate quantitative analysis was achieved compared to conventional 2-DE, and 179 differentially expressed protein spots were subsequently identified from gels by MALDI-MS. This study not only revealed the complexity of the response to benzoic acid treatment but also showed that membrane proteins and different protein isoforms could be analyzed using this method. In a slightly different application of the method, reciprocal labeling with Cy3 and Cy5 was used to identify reproducible and common compensatory changes in the whole brain tissue proteomes of five different genetically altered mouse models. With MALDI-MS identification, this analysis

revealed an isoform change in the heat shock 70 related protein, mortalin, and an alteration of the antioxidant protein 2, indicating that stress-responsive pathways may be altered artifactually in genetically modified mice. Finally, Zhou et al. have demonstrated the use of 2D-DIGE in quantifying differences in protein expression between laser capture microdissection–procured esophageal carcinoma cells and normal epithelial cells from the same human tumor-tissue sample. Analysis of Cy3- and Cy5-labeled lysates from approximately 2 50 000 cancer cells and normal cells (200 μg each) revealed 1038 and 1088 protein spots respectively. DeCyder DIA analysis showed that 58 spots were upregulated by > threefold and 107 were downregulated by > threefold in cancer versus normal cells, although only three proteins were subsequently identified by LC-MS/MS from SYPRO Ruby poststained gels. Normalization was performed using the total pixel volume of each image, rather than the internal standard method described above. Since replicate samples were not compared, the reproducibility and significance of the data is unclear, although this study does demonstrate the power of the 2D-DIGE technique for differential expression analysis of limited sample amounts.

In summary, these reports demonstrate that 2D-DIGE can be applied to provide sensitive and reproducible analysis of differential protein expression between two or more biological samples, be they cell lines, sorted cell types from tissues, whole tissues, or whole microorganisms. Since 2D-DIGE offers increased experimental confidence and accuracy in detecting quantitative protein changes while at the same time reducing experimental workload, it is a particularly advantageous technique for high throughput–expression proteomics. Moreover, the apparent compatibility of the technique with subsequent identification of protein spots from gels by MS further reinforces this notion.

3 Perspectives

The limitations of 2-DE-based approaches for differential expression analysis is worth mentioning. A standard broad pH-range 2-DE gel loaded with 200 μg of cell lysate may resolve between 2000 and 2500 protein features using the most sensitive detection methods. However, these features represent the most abundant and long-lived housekeeping proteins that can mask lower-abundance proteins, thus reducing the resolving power of 2-DE gels. Detection and identification of 1 ng (current detection limit of the best staining methods) of a 50-kDa protein at 100 copies per cell, would require separation of total protein from around 1.25×10^8 cells. Assuming 100% recovery of the protein of interest, this may represent up to 20 mg of total protein, a load that is too great for high-resolution 2-DE separation. A further problem is that hydrophobic, highly insoluble, very basic, or very small/large proteins are not well represented in 2-DE gel-based separations.

One approach to improving the detection of lower-abundance proteins has been to run larger amounts of material on narrow pH-range IEF "zoom" gels. Although this becomes very labor-intensive when a global expression analysis is required, DIGE labeling could easily be applied to this strategy, improving both efficiency and accuracy. Another approach has been the prefractionation of samples, for example, using affinity chromatography or subcellular fractionation prior to 2-DE

separation. Such methods work by reducing the dynamic range of protein expression within a sample as well as by reducing its complexity. However, prefractionation procedures should be kept as simple as possible to avoid sample-to-sample variation, they should target different molecular properties of proteins compared to 2-DE separation, and should be conducted under conditions that minimize sample loss, degradation, or introduction of artifactual protein modifications. Subcellular fractionation has already been combined with 2D-DIGE analysis to identify targets of paracetamol toxicity in liver mitochondrial subfractions and changes in murine cardiac mitochondrial proteins associated with loss of creatine kinase. In the first study, very rapid changes were seen in the levels of many proteins, indicating posttranslational modification rather than changes in gene regulation per se. This not only demonstrates the power of 2-DE over non-gel-based methods for resolving specific isoforms of proteins, but also shows the potential of 2D-DIGE for the accurate analysis of changes in posttranslational modifications.

In the move away from 2-DE-based methods, several gel-free proteomic strategies have been recently developed. Rather than relying on protein separation, these methods are based either on multidimensional chromatographic peptide separation linked to high-throughput MS, or more recently, direct protein recognition and selection on arrayed protein chips. Although protein microarray technologies could ultimately provide high-throughput screening of biological activities and even protein expression levels, they are as yet limited in terms of their availability and the number of targets. Another recently developed method known as *isotope-coded affinity tagging* (ICAT) combines multidimensional chromatography and differential isotope labeling for comparative quantitation of two samples by LC-MS/MS. Although the ICAT method circumvents the problems associated with 2-DE separations, the two methods have been combined, and like DIGE, this allowed the relative quantitation of proteins in two samples separated on the same 2-DE gel. Despite the emergence of these gel-free proteomic strategies, they have yet to prove themselves for the quantitative analysis of multiple samples.

Although in its infancy, 2D-DIGE will undoubtedly become more widespread as a protein expression profiling technique. It is as sensitive as existing 2-DE methodologies, but provides greater statistical accuracy for comparison across multiple samples while at the same time decreasing the workload. One can envisage the development of additional fluorescent protein tags that would allow more than three samples to be analyzed on the same 2-DE gel. Moreover, these tags could be used to target residues other than lysine, or to allow functional labeling of subsets of the proteome. For example, our lab and others are currently developing strategies for the labeling of reduced cysteine thiol groups with iodoacetamide or maleimide versions of the Cy dyes. Since cysteine residues occur at a much lower frequency than lysine residues, this strategy offers the potential for higher stoichiometry labeling and hence higher sensitivity, without adversely affecting the solubility and/or migration of proteins. The addition of affinity tags to these reagents would also allow enrichment of only the labeled subset of proteins, thereby reducing the complexity of biological samples and increasing the yield of low-abundance cysteine-containing proteins.

In terms of functional probes, cysteine labeling with iodoacetylated Cy dyes could

be applied to monitor changes in the oxidation state of certain proteins. The rationale behind this is that thiol groups are often targets of cellular oxidation, and an oxidized thiol group in a target protein would no longer be able to react with the iodoacetylated Cy dye. Indeed, we have preliminary evidence that high stoichiometry labeling of unreduced samples using these Cy dye derivatives can be used to monitor changes in cysteine reactivity in cells during oxidative stress induced by hydrogen peroxide treatment. Recently, BODIPY derivatives of a specific inhibitor have been used as chemically reactive fluorescent probes to covalently label the active site thiol group of cysteine protease family members. This novel functional proteomic approach allowed profiling and identification of active cysteine proteases in complex protein mixtures, and was used to monitor protease activity for the screening of small-molecule inhibitors directed against each enzyme family member. It is likely that other fluorescent probes for functionality will soon be designed and used for proteomic profiling of complex biological samples. With this in mind, the future of differential fluorescence labeling in proteomics looks bright.

Bibliography

Books and Reviews

Anderson, N.L., Anderson, N.G. (1998) Proteome and proteomics: new technologies, new concepts, and new words, *Electrophoresis* **19**, 1853–1861.

Görg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., Weiss, W. (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients, *Electrophoresis* **21**, 1037–1053.

Greenbaum, D., Baruch, A., Hayrapetian, L., Darula, Z., Burlingame, A., Medzihradszky, K.F., Bogyo, M. (2002) Chemical approaches for functionally probing the proteome, *Mol. Cell. Proteomics* **1**, 60–68.

Lilley, K.S., Razzaq, A., Dupree, P. (2002) Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation, *Curr. Opin. Chem. Biol.* **6**, 46–50.

Naaby-Hansen, S., Waterfield, M.D., Cramer, R. (2001) Proteomics–post-genomic cartography to understand gene function, *Trends Pharmacol. Sci.* **22**, 376–384.

Patton, W.F. (2000) Making blind robots see: the synergy between fluorescent dyes and imaging devices in automated proteomics, *Biotechniques* **28**, 944–948, 950–947.

Peng, J., Gygi, S.P. (2001) Proteomics: the move to mixtures, *J. Mass Spectrom.* **36**, 1083–1091.

Rabilloud, T. (2002) Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains, *Proteomics* **2**, 3–10.

Primary Literature

Adam, G.C., Sorensen, E.J., Cravatt, B.F. (2002) Chemical strategies for functional proteomics, *Mol. Cell. Proteomics* **1**, 781–790.

Baty, J.W., Hampton, M.B., Winterbourn, C.C. (2002) Detection of oxidant sensitive thiol proteins by fluorescence labeling and two-dimensional electrophoresis, *Proteomics* **2**, 1261–1266.

Berggren, K., Chernokalskaya, E., Steinberg, T.H., Kemper, C., Lopez, M.F., Diwu, Z., Haugland, R.P., Patton, W.F. (2000) Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex, *Electrophoresis* **21**, 2509–2521.

Berggren, K., Steinberg, T.H., Lauber, W.M., Carroll, J.A., Lopez, M.F., Chernokalskaya, E., Zieske L., Diwu Z., Haugland, R.P., Patton, W.F. (1999) A luminescent ruthenium complex for ultrasensitive detection of proteins immobilized on membrane supports, *Anal. Biochem.* **276**, 129–143.

Corthals, G.L., Wasinger, V.C., Hochstrasser, D.F., Sanchez, J.C. (2000) The dynamic range of protein expression: a challenge for proteomic research, *Electrophoresis* **21**, 1104–1115.

Daban, J.R., Bartolome, S., Samso, M. (1991) Use of the hydrophobic probe Nile red for

the fluorescent staining of protein bands in sodium dodecyl sulfate-polyacrylamide gels, *Anal. Biochem.* **199**, 169–174.

Diezel, W., Kopperschlager, G., Hofmann, E. (1972) An improved procedure for protein staining in polyacrylamide gels with a new type of Coomassie brilliant blue, *Anal. Biochem.* **48**, 617–620.

Dzandu, J.K., Johnson, J.F., Wise, G.E. (1988) Sodium dodecyl sulfate-gel electrophoresis: staining of polypeptides using heavy metal salts, *Anal. Biochem.* **174**, 157–167.

Gee, K.R., Kang, H.C., Meier, T.I., Zhao, G., Blaszcak, L.C. (2001) Fluorescent Bocillins: synthesis and application in the detection of penicillin-binding proteins, *Electrophoresis* **22**, 960–965.

Gharahdaghi, F., Weinberg, C.R., Meagher, D.A., Imai, B.S., Mische, S.M. (1999) Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: a method for the removal of silver ions to enhance sensitivity, *Electrophoresis* **20**, 601–605.

Gharbi, S., Gaffney, P., Yang, A., Zvelebil, M.J., Cramer, R., Waterfield, M.D., Timms, J.F. (2002) Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system, *Mol. Cell. Proteomics* **1**, 91–98.

Goldberg, R.L., Fuller, G.C. (1978) Quantitation of collagen in polyacrylamide gels by fluorescent scanning of MDPF-labeled proteins: an improvement over densitometric scanning of gels stained by Coomassie blue, *Anal. Biochem.* **90**, 69–80.

Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y., Aebersold, R. (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9390–9395.

Hoving, S., Gerrits, B., Voshol, H., Muller, D., Roberts, R.C., van Oostrum, J. (2002) Preparative two-dimensional gel electrophoresis at alkaline pH using narrow range immobilized pH gradients, *Proteomics* **2**, 127–134.

Jackson, P., Urwin, V.E., Mackay, C.D. (1988) Rapid imaging, using a cooled charge-coupled-device, of fluorescent two- dimensional polyacrylamide gels produced by labelling proteins in the first-dimensional isoelectric focusing gel with the fluorophore 2- methoxy-2,4-diphenyl-3(2H)furanone, *Electrophoresis* **9**, 330–339.

Kemper, C., Steinberg, T.H., Jones, L., Patton, W.F. (2001) Simultaneous, two-color fluorescence detection of total protein profiles and beta-glucuronidase activity in polyacrylamide gel, *Electrophoresis* **22**, 970–976.

Kernec, F., Unlu, M., Labeikovsky, W., Minden J.S., Koretsky, A.P. (2001) Changes in the mitochondrial proteome from mouse hearts deficient in creatine kinase, *Physiol. Genomics* **6**, 117–128.

Li, K.W., Geraerts, W.P., van Elk, R., Joosse, J. (1989) Quantification of proteins in the subnanogram and nanogram range: comparison of the AuroDye, FerriDye, and India ink staining methods, *Anal. Biochem.* **182**, 44–47.

Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., Yates, J.R.I. II (1999) Direct analysis of protein complexes using mass spectrometry, *Nat. Biotechnol.* **17**, 676–682.

MacBeath, G., Schreiber, S.L. (2000) Printing proteins as microarrays for high-throughput function determination, *Science* **289**, 1760–1763.

Meyer, T.S., Lamberts, B.L. (1965) Use of Coomassie brilliant blue R250 for the electrophoresis of microgram quantities of parotid saliva proteins on acrylamide-gel strips, *Biochim. Biophys. Acta* **107**, 144–145.

Neuhoff, V., Stamm, R., Pardowitz, I., Arold, N., Ehrhardt, W., Taube, D. (1990) Essential problems in quantification of proteins following colloidal staining with Coomassie brilliant blue dyes in polyacrylamide gels, and their solution, *Electrophoresis* **11**, 101–117.

Ortiz, M.L., Calero, M., Fernandez Patron, C., Patron, C.F., Castellanos, L., Mendez, E. (1992) Imidazole-SDS-Zn reverse staining of proteins in gels containing or not SDS and microsequence of individual unmodified electroblotted proteins, *FEBS Lett.* **296**, 300–304.

Patton, W.F. (2000) A thousand points of light: the application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics, *Electrophoresis* **21**, 1123–1144, [pii].

Quadroni, M., James, P. (1999) Proteomics and automation, *Electrophoresis* **20**, 664–677.

Rabilloud, T. (1999) Silver staining of 2-D electrophoresis gels, *Methods Mol. Biol.* **112**, 297–305.

Ruepp, S.U., Tonge, R.P., Shaw, J., Wallis, N., Pognan, F. (2002) Genomics and proteomics analysis of acetaminophen toxicity in mouse liver, *Toxicol. Sci.* **65**, 135–150.

Scheler, C., Lamer, S., Pan, Z., Li, X.P., Salnikow, J., Jungblut, P. (1998) Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS), *Electrophoresis* **19**, 918–927.

Shevchenko, A., Wilm, M., Vorm, O., Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels, *Anal. Chem.* **68**, 850–858.

Skynner, H.A., Rosahl, T.W., Knowles, M.R., Salim, K., Reid, L., Cothliff, R., McAllister, G., Guest, P.C. (2002) Alterations of stress related proteins in genetically altered mice revealed by two-dimensional differential in-gel electrophoresis analysis, *Proteomics* **2**, 1018–1025.

Smolka, M., Zhou, H., Aebersold, R. (2002) Quantitative protein profiling using two-dimensional gel electrophoresis, isotope-coded affinity tag labeling, and mass spectrometry, *Mol. Cell. Proteomics* **1**, 19–29.

Steinberg, T.H., Jones, L.J., Haugland, R.P., Singer, V.L. (1996) SYPRO orange and SYPRO red protein gel stains: one-step fluorescent staining of denaturing gels for detection of nanogram levels of protein, *Anal. Biochem.* **239**, 223–237.

Steinberg, T.H., Lauber, W.M., Berggren, K., Kemper, C., Yue, S., Patton, W.F. (2000) Fluorescence detection of proteins in sodium dodecyl sulfate- polyacrylamide gels using environmentally benign, nonfixative, saline solution, *Electrophoresis* **21**, 497–508.

Steinberg, T.H., Pretty On Top, K., Berggren, K.N., Kemper, C., Jones, L., Diwu, Z., Haugland, R.P., Patton, W.F. (2001) Rapid and simple single nanogram detection of glycoproteins in polyacrylamide gels and on electroblots, *Proteomics* **1**, 841–855.

Sumner, L.W., Wolf-Sumner, B., White, S.P., Asirvatham, V.S. (2002) Silver stain removal using H2O2 for enhanced peptide mass mapping by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Rapid Commun. Mass Spectrom.* **16**, 160–168.

Switzer, R.C.I. II, Merril, C.R., Shifrin, S. (1979) A highly sensitive silver stain for detecting proteins and peptides in polyacrylamide gels, *Anal. Biochem.* **98**, 231–237.

Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., Davison, M. (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology, *Proteomics* **1**, 377–396.

Unlu, M., Morgan, M.E., Minden, J.S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts, *Electrophoresis* **18**, 2071–2077.

Urwin, V.E., Jackson, P. (1991) A multiple high-resolution mini two-dimensional polyacrylamide gel electrophoresis system: imaging two-dimensional gels using a cooled charge-coupled device after staining with silver or labeling with fluorophore, *Anal. Biochem.* **195**, 30–37.

Urwin, V.E., Jackson, P. (1993) Two-dimensional polyacrylamide gel electrophoresis of proteins labeled with the fluorophore monobromobimane prior to first-dimensional isoelectric focusing: imaging of the fluorescent protein spot patterns using a cooled charge-coupled device, *Anal. Biochem.* **209**, 57–62.

Von Eggeling, F., Gawriljuk, A., Fiedler, W., Ernst, G., Claussen, U., Klose, J., Romer, I. (2001) Fluorescent dual colour 2D-protein gel electrophoresis for rapid detection of differences in protein pattern with standard image analysis software, *Int. J. Mol. Med.* **8**, 373–377.

Washburn, M.P., Wolters, D., Yates, J.R.I. II (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology, *Nat. Biotechnol.* **19**, 242–247.

Wilson, C.M. (1983) Staining of proteins on gels: comparisons of dyes and procedures, *Methods Enzymol.* **91**, 236–247.

Wolters, D.A., Washburn, M.P., Yates, J.R., (2001) An automated multidimensional protein identification technology for shotgun proteomics, *Anal. Chem.* **73**, 5683–5690.

Yan, J.X., Devenish, A.T., Wait, R., Stone, T., Lewis, S., Fowler, S. (2002) Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of Escherichia coli, *Proteomics* **2**, 1682–1698.

Yan, J.X., Wait, R., Berkelman, T., Harry, R.A., Westbrook, J.A., Wheeler, C.H., Dunn, M.J. (2000) A modified silver staining protocol

for visualization of proteins compatible with matrix-assisted laser desorption/ionization and electrospray ionization-mass spectrometry, *Electrophoresis* **21**, 3666–3672.

Yeargin, J., Haas, M. (1995) Elevated levels of wild-type p53 induced by radiolabeling of cells leads to apoptosis or sustained growth arrest, *Curr. Biol.* **5**, 423–431.

Zhou, G., Li, H., DeCamp, D., Chen, S., Shu, H., Gong, Y., Flaig, M., Gillespie, J.W., et al. (2002) 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers, *Mol. Cell. Proteomics* **1**, 117–124.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., et al. (2001) Global analysis of protein activities using proteome chips, *Science* **293**, 2101–2105.

Gene Distribution in the Human Genome

Yujing Zeng[1], *Javier Garcia-Frias*[1], *and Adam G. Marsh*[2]
[1] *Department of Electrical and Computer Engineering, University of Delaware, Newark, DE, USA*
[2] *Department of Marine Studies, University of Delaware, Lewes, DE, USA*

1 Macroscale Distribution 55
1.1 Gene Number Estimates 56
1.2 Gene Prediction 57
1.2.1 Public Consortium 58
1.2.2 Celera Effort 59
1.3 Gene Distribution 59

2 Meso-scale Organization 61
2.1 Gene Distribution versus Chromosomal Banding 62
2.2 Gene Distribution versus Sequence Composition 63
2.2.1 Uneven Distribution of GC Content Over the Genome 63
2.2.2 Isochores 64
2.2.3 Gene Properties versus GC Content 65
2.2.4 Evolutionary Hypothesis About Isochores 67
2.2.5 Recent Discussions About the Isochore Concept 68

3 Microscale Structure 69
3.1 Gene Model 69
3.2 CpG Islands 72
3.3 GGG Trinucleotide 75
3.4 Pseudogene Distributions 76
3.5 Transposable Elements 77

4 Metascale Integration 77
4.1 Associative Expression Networks 78
4.2 Functional Coordination and Physical Mapping 79

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

4.3 Physical Colocation and Dislocation 81
4.4 Dynamic Expression Linkages 83
4.5 Importance of Expression Controls 85

Bibliography 86
Books and Reviews 86
Primary Literature 86

Keywords

Chromosomal Bands
Transverse light and dark bands on the chromosomes produced by different staining techniques.

CpG Islands
Dinucleotide motif of a cytidine and guanidine (5′-C-p-G-3′), commonly unmethylated, within a GC-rich chromosomal domain.

EST (Expressed-Sequence Tag)
Short (300–500 nucleotide) single read DNA sequence derived from a randomly selected cDNA clone.

Exon
A coding DNA segment within a gene.

GC Content
The distribution of Guanine and Cytosine nucleotides among the genome, which is correlated with various structural and functional properties of the genome fragments.

Intron
A noncoding DNA segment within a gene.

Isochores
Long DNA segments, usually greater than 300 kb, which are characterized by a fair compositional homogeneity.

Promoter
DNA segment composed of the first few hundred nucleotides located ‘‘upstream’’ (on the 5′ side) of a gene, and which controls the transcription of such a gene.

Transcript
mRNA sequence produced by transcription from DNA.

Splicing
The process that removes all the introns to produce mature mRNA.

The human genome encodes the information about the development, physiology, and evolution of our own species. Among the whole DNA stretch with more than 3000 million nucleotides, the most interesting portions are those composed of genes, the DNA fragments that code for a defined biochemical function, such as the production of a particular protein. It is believed that genes are responsible for the major biological functions of the genome, and knowledge about them and their encoded proteins is crucial for basic biology, biomedical research, biotechnology, and health care.

In this contribution, we provide an overview of the ongoing research on the gene distribution in the human genome. It has been known for a long time that the distribution of protein-coding genes among human chromosomes is extremely uneven. DNA fragments with different gene densities show different compositional properties, which are believed to be correlated with gene composition, function, and evolution. Around 98% of the human genome does not code directly for proteins and therefore was once dismissed as "junk DNA." However, many studies have shown that these "junk" sections contain many important features essential for the biological function of a genome. The challenge to understanding how genomes "work" involves more than just the identification of protein-coding segments. We must understand the mechanisms by which all the components of a genome interact to produce a functional cellular system.

1
Macroscale Distribution

In order to perfectly define the human genome, it would be necessary to precisely locate each and every single gene. At present, this is not possible because gene identification from the human genome drafts is a complicated process given the multiple sequence elements that combine to ultimately produce (express) an operational gene. A gene is a locus of co-transcribed exons, which in the human genome tend to be short. Exons are separated by introns, which are sequences of long lengths (some of them exceeding 10 kb) that do not code for protein sequences. Transcription of a gene domain produces heteronuclear RNA (hnRNA), which is a linear copy of the coding DNA strand (except with complementary nucleotides). This hnRNA transcript contains all the exon and intron sequences. Post-transcriptional RNA processing splices out the intronic sequences and prepares the RNA transcript for translation as a mature messenger RNA molecule (mRNA). What is surprising is that these mature coding segments of mRNA (which are responsible for producing all the proteins in our bodies) represent only ~1.5% of our total genome DNA sequence. Thus, looking for

genes in our genome is like looking for needles in a haystack. In addition, a single gene may give rise to multiple transcripts by alternative splicing, and alternative transcription initiation and termination sites. Given the current limited knowledge about gene structure and function, none of the existing algorithms for gene prediction and annotation are perfect, and therefore we can only estimate the gene distribution based on our current understanding of how genes are structured and expressed.

1.1 Gene Number Estimates

Before the publication of the human genome drafts, estimates of the total number of genes in humans were proposed in the literature. However, different approaches produced different estimates, with some as high as 120 000, even when the methods utilized for the estimation were similar. One of the reasons for this variability was the necessity to use strong assumptions in order to obtain the estimate, which made it very difficult to assess the reliability of the predicted number. For instance, Ewing and Green estimated the number of genes by comparing a database of ESTs (expressed-sequence tags) with chromosome 22 and with a nonredundant set of mRNA sequences. The idea, explained in more detail in their work, is to estimate the total number of genes by counting the amount of overlap between two sets of genes obtained independently. Both comparisons (ESTs with chromosome 22 and ESTs with the mRNA sequences) produced an estimate in the order of 35 000 genes. This estimate is very different from the 60 000–70 000 genes predicted by Fields et al., who used a similar method to compare a set of ESTs with mRNA sequences. The difference is that Ewing and Green eliminate single unconfirmed ESTs in the counting process, thereby resulting in a more conservative estimate. Other methods predicted 80 000 genes by estimating the number of CpG islands in the genome and assuming that all islands are associated with genes and that 56% of the genes have an island. These assumptions were challenged by Ewing and Green, who suggest an estimate below 30 000. Other gene number estimates based on reassociation kinetics and genome comparisons with other species also produce very different results.

With the publication of the two drafts of the human genome sequence (The human genome sequence is available at http://genome.ucsc.edu/) in 2001, more comprehensive studies about gene distributions in the human genome have become possible. Although the genome sequence is still being updated, both the public and the private sequencing efforts have obtained similar estimates of around 30 000 genes for humans. These studies, however, are far from being complete and conclusive. There are several difficulties: first, the draft sequences do not provide a perfect coverage of the human genome and there are still errors in the sequence data. Therefore, even the best estimate from the sequence might not be close to the real one. Second, and more importantly, gene extraction from the genome is not a trivial task, and even the best gene prediction tools are not likely to be 100% reliable. Therefore, it is fair to say that our knowledge about the gene distribution in the human genome, although much better than before the publication of the draft sequences, is still far from complete. For instance, the work of Das et al., published after the human genome draft, suggests, by using experimental evidence, a higher estimate in the range between

41 000 and 45 000 genes is obtained. As we will see in the following sections, these estimates are still based on some assumptions. However, as an advantage over other methods that do not use direct sequence data, the availability of the genome draft allows the refinement of these assumptions. Notice also that the methods that are applied directly over the genome sequence to identify genes obtain not only estimates about the gene number but also about the gene distribution.

1.2 Gene Prediction

There are different methods to locate genes in the human genome. The simplest one is to perform alignments between known genes contained in databases such as RefSeq and the genome sequence. Although this approach is very powerful to locate genes that are already known, it does not allow the discovery of any new genes. Another related possibility is to make gene predictions by performing sequence comparison with previous known genes and proteins in other organisms. As before, the problem with this approach is that it cannot extract new genes, since, obviously, they are not included in the comparison set. A third method is based on homology studies between the genome sequence and libraries of ESTs. Obviously, the success of this method depends on the completeness and quality of the libraries, which should consist of relatively large ESTs matching the genes. Finally, *de novo* (*ab initio*) techniques try to recognize groups of exons directly from the genome sequence. In order to do so, these methods define gene models using *a priori* information about the gene structure (sequence composition, intron/exon structure and length, splice sites, etc.), and perform gene extraction by identifying regions in the genome sequence that correspond to the gene model.

Several models have been used in different *ab initio* methods, including neural networks, decision trees, and, more frequently, hidden Markov models (HMMs), which are the base of well-known tools such as Genscan and Genie. If our understanding of the cell mechanisms to identify genes were complete (or in other words, if *a priori* information about the gene structure were perfectly available), and provided that the DNA sequence were perfectly known, *ab initio* techniques would be able to extract and locate all genes contained in the genome. Unfortunately, this *a priori* knowledge is not available, and the lack of knowledge leads to several types of errors in gene prediction:

- False positives, resulting when sections of the genome sequence that do not correspond to genes are identified as such by the prediction method. This type of prediction error can occur for several reasons such as spurious predictions due to alignments produced by chance or by the presence of pseudogenes.
- Missing genes, resulting when sections of the genome corresponding to genes are not identified as such by the prediction method. This type of prediction error can occur for several reasons, including the inability of the method to either detect all gene components or to properly connect correctly detected components (fragmentation).
- Inaccurate genes, resulting when a gene is identified but the prediction is either incomplete or partially wrong. This type of prediction error occurs when the method misses some exons or incorporates wrongly predicted exons in the gene prediction.

The gene prediction methods just mentioned are not mutually exclusive. In fact, they are usually applied together with the objective to develop complex prediction tools capable of achieving the best possible sensitivity and specificity – that is, joint optimization to miss as few genes as possible with the least number of false positives.

1.2.1 Public Consortium

The first step in the effort from the public consortium to predict genes in the human genome draft is to utilize the Ensembl system. This system uses the *ab initio* tool Genscan to generate gene predictions that are confirmed by similarities to ESTs, mRNAs, and protein motifs. Each gene component (exon/intron) is confirmed independently, which increases the likelihood of fragmentation. In order to reduce this fragmentation, Genie, another *ab initio* technique, is utilized jointly with Ensembl. These Ensembl/Genie results are compared with known genes contained in RefSeq, SWISSPROT, and TrEMBL databases. At the end of this process, the total number of gene predictions is almost 32 000. From this prediction, approximately 15 000 of them correspond to known genes, 4000 come from the combination of Ensemble and Genie, and 13 000 from Ensembl alone.

In order to validate the quality of these predictions, different approaches are utilized to estimate the sensitivity and specificity of the prediction stage. The first approach used in the landmark paper of the public consortium is to compare these predictions with new independently discovered genes not used in the prediction process. Thirty-one genes newly discovered in chromosome X were considered in the aforementioned paper. From these genes, 28 are included in the draft, and 19 were successfully predicted, which gives an estimate for the sensitivity of the gene prediction process of 68%. Moreover, on average 79% of each gene was detected and the fragmentation can be estimated as 1.4 gene predictions per true gene. However, this approach should be taken with caution because the number of genes limits the accuracy of the statistical analysis. A higher sensitivity of 85% is obtained in the second approach, when the predicted genes are compared to a set of 15 294 mouse cDNAs generated by the RIKEN Genomic Sciences Center. Specifically, 81% of the genes in the mouse showed similarity to the human genome sequence and 69% were similar to the predicted human genes, resulting in a sensitivity of $69/81 = 85\%$. The final approach compares the predicted genes with the gene annotations in chromosome 22, which consisted of 539 confirmed genes and 133 pseudogenes. This comparison suggests a rate of false-positive or spurious predictions (including pseudogenes predictions) at around 20%.

The final estimate of the number of genes results from the development showed above: 17 000 gene predictions of unknown genes with a 20% false-positive rate and a fragmentation factor of 1.4 lead to an estimate of 9500 new unknown genes. If these gene predictions are assumed to contain 60% of all previously unknown human genes (following the validation results in the previous paragraph), the number of previously unknown human genes can be estimated as 16 000. Considering also the 15 000 known genes contained in the human genome draft, the total number of genes in the human genome can be estimated to be around 31 000. Notice that this number is highly dependent on the estimates of

specificity, sensitivity, and fragmentation of the gene prediction method utilized for the analysis.

1.2.2 Celera Effort

In order to perform gene prediction, Celera Corporation, led by Craig Venter, developed an approach called *Otto*. Otto considers different lines of evidence to locate genes. The first step in the process is to identity gene boundaries by examining matches of the genome sequence to EST and protein databases. These matches are combined into bins in such a way that a single gene is expected for each one of the bins.

Then, different procedures are applied to extract genes. A first set of genes is extracted from the genome by comparing the transcript database RefSeq with these bins. A total of ~6500 genes resulted from this process, which identified genes when some transcript in the RefSeq database matched the genome sequence for at least 50% of its length with more than 92% identity. These ~6500 genes are the ones in which Otto has a higher confidence. For the rest of the genome regions that have sequence similarity but do not present a clear match to known genes, Otto considers four lines of evidence (similarity to known proteins, similarity to human ESTs and cDNAs, similarity to rodent ESTs and cDNA, and conservation between mouse and human DNA) and tries to develop a gene model using Genscan. If the model is reasonable, and predicted exons are supported by at least one of the lines of evidence, a gene is considered identified. The number of additional genes obtained in this way by Otto was ~11 000. Considering the ~6500 genes identified before, the number of genes predicted by Otto is ~17 500 if only one line of evidence is required. The number decreases to ~17 000, ~15 500, and ~12 500 if two, three, or four lines of evidence are utilized.

It is interesting to note that Otto does not try to directly identify genes with gene prediction tools such as Genscan. Instead, the gene boundaries are predicted by first looking at homology evidence. The reason for this approach is that most gene prediction methods have problems performing boundary prediction. The specificity, sensitivity, and fragmentation rate of this method are superior to those of Genscan. However, predictions from Otto are very conservative, since they depend on ESTs and some gene transcripts are not included in the existing databases. Therefore, Otto's predictions are complemented with additional ones produced by three *de novo* techniques (GRAIL, Genscan, and FgenesH). The number of *de novo* predictions that did not overlap with known genes or Otto results was around 58 000. From this number ~21 500 are supported by one line of evidence, ~8600 by two, ~5000 by three, and ~1900 by four. Therefore, the total number of predicted genes obtained by combining these predictions with the ones proceeding from Otto, would be ~39 000, ~26 500, ~23 000, and ~20 000 depending if one, two, three, or four lines of evidence are required for the *de novo* methods (and assuming that one line of evidence is enough for the Otto predictions). Around 1000 additional genes can be predicted by identifying regions outside the original bins where there was a match between an EST and the genome sequence across a splice junction, and at least one line of evidence was present.

1.3 Gene Distribution

As the sequencing of larger and larger regions of the human genome was

completed, there was a series of increasingly comprehensive gene maps constructed and cross-referenced to the human genetic map. The gene map published in 1996 by Schuler et al. confirmed an uneven gene distribution among different chromosome bands. Two years later, a new gene map with higher coverage and accuracy was published, showing that the distribution of genes across individual chromosomes presents striking fluctuations. Significantly, higher than average gene densities are found in chromosomes 1, 11, 17, 19, and 22, while lower than average densities are found in chromosomes 4, 5, 8, 13, 18, and X.

The publication of the two draft sequences in 2001 represents a milestone in the study on human genomics. We are now closer than ever to understanding how our own genomes are structured in terms of the distribution and organization of genes within and between chromosomes. The first, full, genome structural analysis was completed by the International Human Genome Sequencing Consortium (IHGSC), which aimed at creating an initial integrated gene index (IGI) and an associated integrated protein index (IPI) for the human genome. Then the chromosomal distribution of the IGI gene set was examined. On average, the gene density in the human genome is 11.1 genes per Mb. However, this number presents significant variations among chromosomes, from 26.8 genes per Mb for chromosome 19 to 6.4 genes per Mb for chromosome Y. The number for chromosome Y could be even smaller, because the high number of pseudogenes contained in chromosome Y is likely to have generated numerous false-positives from the gene prediction studies.

Several months after the publication of the human genome, a draft physical map with annotations for a majority of the human transcripts was published by Zhuo et al. The map was generated by assembling the sequence clusters in the UNIGENE database into nonredundant sequence contigs, and then aligning these clusters to the human genome draft. As shown in Fig. 1, the estimated transcript densities show a significant difference among all the human chromosomes. The

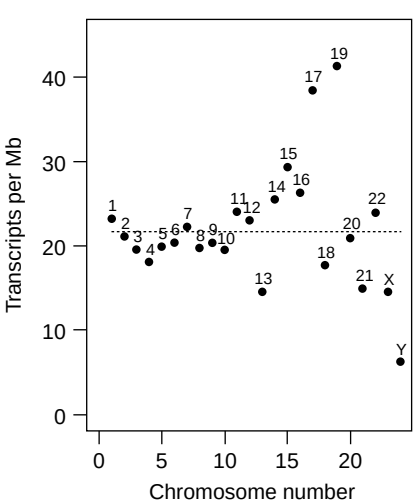

Fig. 1 Transcript densities on the different human chromosomes. The broken line indicates the average transcript density for the whole genome (from Zhuo, D. et al. (2001) Assembly, annotation, and integration of UNIGENE clusters into the human genome draft, *Genome Res.* **11**, 904–918).

highest density, found in chromosome 19, is above 40 transcripts per Mb, which is almost twice the value of the average density over the whole genome. The lowest gene densities were found in chromosomes X, Y, 21, and 13, which is consistent with the previous gene density estimates.

According to the record updated on January 25, 2003, there are 18 761 human genes reported in RefSeq that have their function and coding sequences known. Figure 2 shows the chromosomal distribution of these genes, excluding 225 genes that have not been located in the draft genome. The uneven gene distribution indicates that the human genome is organized in discrete regions, and that functional characteristics may get distributed in a related pattern.

2
Meso-scale Organization

With an estimate of around 30 000 genes and over 3 Gb nucleotides in the human genome sequence, the inevitable question that springs to mind is "where exactly are the genes located?" In other words, how are our 30 000 genes distributed over our 3 000 000 000 bp genome? Many contributions in the literature have tried to address this question during the last 30 years. As described in Sect. 1.3, genes are distributed unevenly among chromosomes. Moreover, different lines of evidence, which will be reviewed in this section, have led to the same conclusion: the gene distribution along each chromosome is conspicuously nonuniform. In this section, we move our observation from the

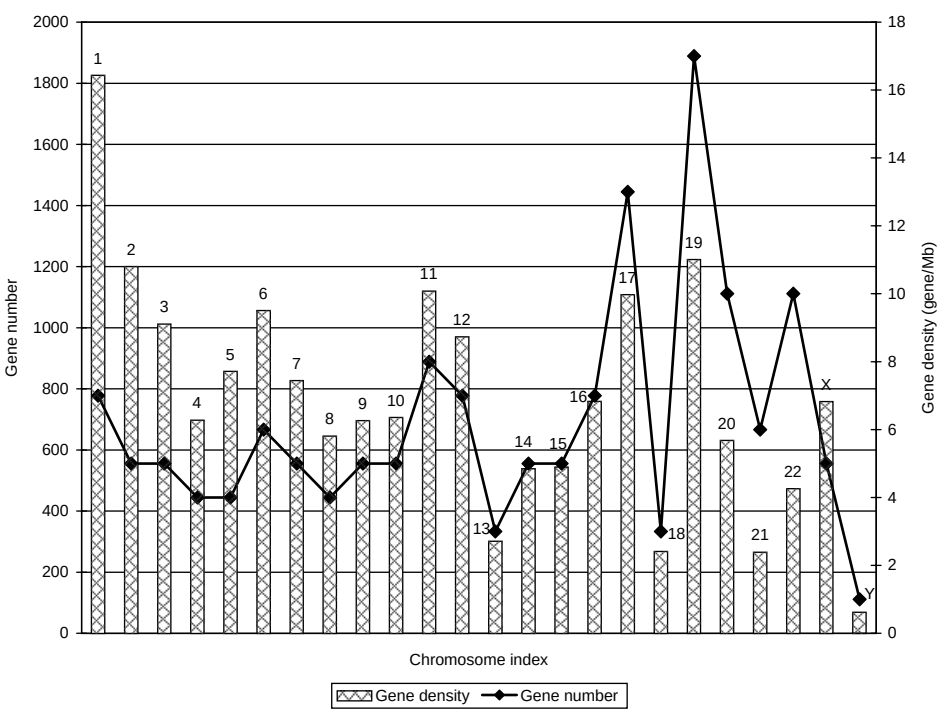

Fig. 2 Chromosomal distribution of the 18 761 genes reported in RefSeq as of January 25, 2003.

chromosome to a finer scale, and describe the gene distribution according to several biochemical features in the human genome.

2.1 Gene Distribution versus Chromosomal Banding

The uneven gene distribution in the human genome was hinted at long before the gene map on the whole genome was estimated, by the observation of the banding patterns produced when chromosomes are stained. More than 30 years ago, it was discovered that certain dyes selectively stain some regions of metaphase chromosomes more intensely than other regions, and that the resulting transverse banding patterns are specific for individual chromosomes.

Banding characteristics were readily applied for the identification of individual chromosomes. Shortly afterwards, other applications were discovered, with several studies showing that chromosomal bands might have functional and structural significance. Giemsa, for example, is a permanent DNA dye that has affinity for A + T rich DNA regions. After a pretreatment that denatures the DNA sequence with barium hydroxide, Giemsa darkens the regions of constitutive heterochromatin, DNA fragments that are always retained in a compact organization and contain almost no active genes. These dark bands are called C-bands and cover around 17 to 20% of the human chromosome complement. They are mostly distributed in pericentromeric regions. One exception is the Y chromosome that has a relatively large heterochromatic segment in its long arm, suggesting a low gene density on that chromosome.

The remaining 80% of the genome, the euchromatic component, is divided into G-, R-, and T-bands. G-bands are produced by staining the chromosomes with Giemsa, after subjecting them to a brief proteolytic treatment. R-banding (referred to as reverse Giemsa, since the resulting bands are roughly the reverse of G-bands) requires pretreatment of the chromosomes, prior to Giemsa staining, in a hot physiological alkaline solution. Under such conditions, the AT-richer DNA sections are denatured faster, reducing their affinity for Giemsa staining. Finally, T-bands are an extremely heat-resistant subset of R-bands, obtained by repeating the heat pretreatment. The distribution of all these bands along the human chromosomes is shown in the ideograms (Fig. 3), which are used broadly in cytogenetic analyses (e.g. for the detection of chromosome aberrations causing disease).

Chromosomal bands correspond to chromosome domains with different properties. For instance, it has been shown that early and late phases of mammalian DNA replication correlate with the banding pattern: R-bands replicate early, while G-bands are associated with a late replication time. This result has been confirmed by analysis of replication times across R/G boundaries. Moreover, the work of Matassi et al. has revealed that genes located closely on a chromosome have similar evolutionary rates, which may imply a relationship between evolutionary rates and chromosomal bands. One of the most interesting properties is the relationship between gene density and chromosome bands.

Chromosomal bands exhibit a clear correlation with gene density: Generally speaking, the gene density of R-bands is higher than that of G-bands.

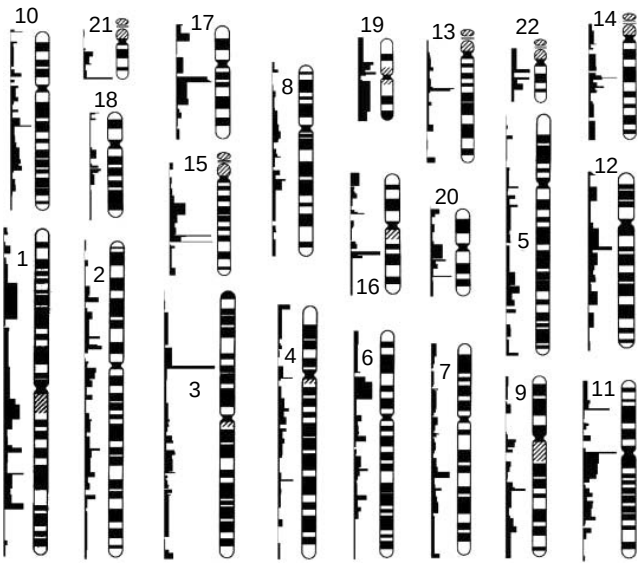

Fig. 3 Gene distributions on chromosomes 1–22. A plot of the average number of estimated gene *loci* across a moving window of 100 Kb is aligned to the ideograms of each chromosome. Genes are heterogeneously distributed both within and among chromosomes (see Caron, H., et al. (2001) The human transcriptome map: Clustering of highly expressed genes in chromosomal domains, *Science* **291**, 1289–1292).

Furthermore, over half of the genes are contained in T-bands, which represent only 15% of all bands. The combination of this result with the ideograms makes it is possible to obtain a rough picture of the gene distribution over each chromosome.

The base compositions of G-, R-, and T-bands are very different from each other: R-bands are richer in GC content than G-bands, and the 27% GC-richest R-bands correspond almost perfectly to the T-bands. In the context of the previous discussion, this fact suggests that the gene density may be related to the local compositional pattern of the DNA sequence, and establishes a connection between cytogenetic observations and molecular analysis.

2.2 Gene Distribution versus Sequence Composition

2.2.1 Uneven Distribution of GC Content Over the Genome

More than 30 years ago, ultracentrifugation experiments based on the use of sequence-specific ligands, such as silver ions, were performed to achieve high-resolution DNA fractionation. These experiments led to the discovery of the striking compositional heterogeneity in the nucleotide composition of human DNA. Specifically, it was found that the distribution of the GC content over the whole genome is far from uniform, since there is a substantial variation in average GC content among large DNA fragments.

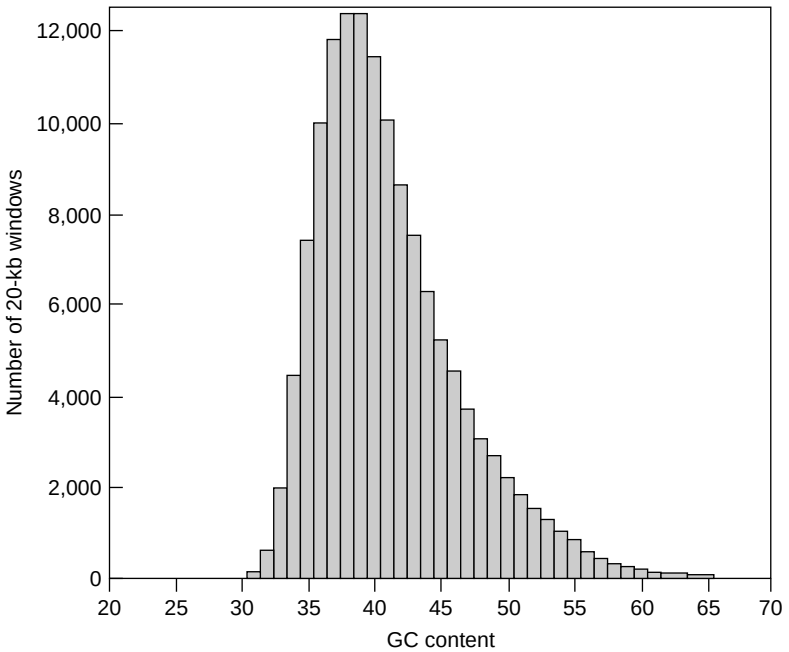

Fig. 4 GC content histogram obtained from the human genome draft (from IHGSC, *Nature* **409**, 2001).

The recent publication of the human genome confirms the uneven GC content distribution in a direct and global manner. Figure 4 shows the GC content histogram obtained from the human genome drafts using 20-kb windows. The figure confirms that local GC content undergoes long-range excursions from its genome-wide average of 41%. Recently, Pavlicek et al. obtained a compositional map of the human chromosomes by scanning the human genome sequence with a 100 Kb moving window. The map shows the large proportion of the genome characterized by long-stretch fragments with low GC content. The regions of GC content greater than 46% constitute only 12.5% of the whole-genome sequence. It is also obvious from the map that the GC content distribution is chromosome specific.

2.2.2 **Isochores**

Bernardi et al. proposed that the long-range variation in GC content reflects that the human genome is a mosaic of isochores, defined as long DNA segments (greater than 300 kb on average) that are characterized by a fair compositional homogeneity (above a 3-kb size level). Isochores in the human genome have been classified into five different families, depending on their percentage in GC compositions. Families L1 and L2 have a GC content lower than the human genome average, specifically in the range <37% for family L1 and 37 to 42% for family L2. The combination of both families makes up 62.5% of the human genome. A higher GC content than the average is observed in families H1, H2, and H3, which cover the ranges 42 to 46%, 46 to 52% and >52%, and constitute 25, 8.3, and

4.2% of the human genome respectively. Implicit in the isochore definition is the idea that the large-scale GC compositional heterogeneity is discrete or discontinuous, with abrupt changes from higher to lower (or vice versa) GC levels.

Different methods have been applied to determine the isochore families to which human genes belong. All these studies suggest a very nonuniform distribution of genes throughout isochore families: Almost 30% of all human genes reside in the roughly 4% of the genome that belongs to the GC-richest family, H3; in comparison, DNA fragments in families L1 and L2 have much lower gene density, which is less than one-sixth of that of H3.

The relationship between GC content and gene richness has been addressed quantitatively by Mouchiroud et al., who reported a direct positive linear relationship between the GC_3 values (GC levels of third codon positions) of coding sequences and the GC levels of the DNA regions in which they are embedded. This correlation permits the positioning of the distribution profile of coding sequences relative to that of DNA fragments. Consequently, it makes it possible to estimate the relative gene density with respect to different GC contents. In the work of Bernardi et al., this relative gene concentration was calculated by dividing the percentage of genes located in a given GC_3 interval by the percentage of DNA located in the corresponding GC range. With the availability of the draft human genome sequence, the gene density can be estimated directly. Figure 5 presents the gene density as a function of the GC content level, estimated by aligning 9315 known genes on the draft human genome. The results are similar to the earlier estimates, and show the high nonuniformity of the gene distribution in the human genome. Notice that the gene density increases from a very low average level in GC-poor regions to a much higher level in GC-rich regions.

2.2.3 Gene Properties versus GC Content

According to the slope change in gene concentration at the boundary between isochore families H1 and H2 (shown in Fig. 5), Bernardi et al. defined two "gene spaces" in the human genome: the "genome core" and the "empty space." The "genome core" consists of the DNA fragments belonging to isochore families H2 and H3. Although it only represents around 12% of the genome, it embeds more than half of the human genes. The other 88% of the genome has a very low gene density (one gene per 50–150 kb) and therefore is referred to as the "empty space" or the "empty quarter" (from the classical name of the Arabian desert). This division into core and empty spaces is also validated by other observations, such as (1) the similarity of the gene concentrations in each space; (2) the similarity of the heptanucleotide comprising the AUG initiation codon of human genes observed in L/H1 and H2/H3 isochores respectively, and (3) the identical chromosomal distribution of H2 and H3 isochores.

As indicated before, the compositional pattern of coding sequences is directly related to that of their embedding isochores. Therefore, it is expected that genes in the genome core should have a relative high GC content, while this content should be low for those genes in the empty space. Several studies have confirmed these expectations. Besides the compositional difference, the gene structures in the two spaces are significantly different. Interestingly, long genes are scarce in the core genome and more frequent in the empty quarter. A primary reason for this

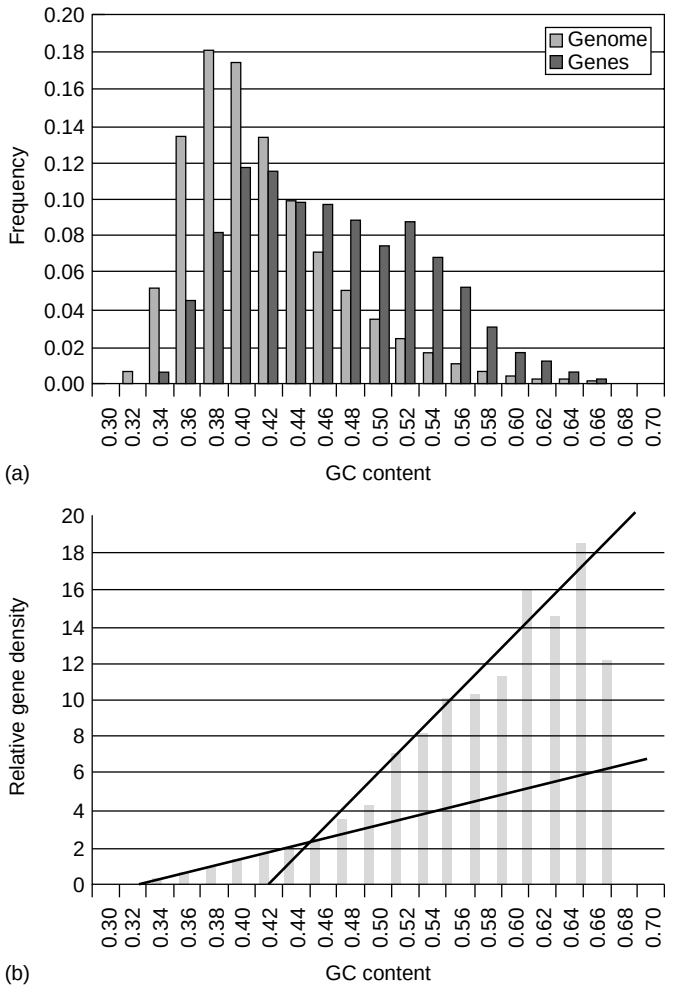

Fig. 5 (a) GC content distribution in 9315 known genes and in the whole genome. (b) Gene density as a function of GC content, obtained by taking the ratio of the data in (a). Values are less accurate at higher GC levels because the denominator is small. The two slopes represented here are associated to the gene core and empty space (from IHGSC, *Nature* **409**, 2001 and *Bernardi* et al, *Gene* **276**, 2001, see original papers for details).

length difference is the fluctuation in intron sizes, which, in the empty quarter are on average three times longer than in the genome core. In contrast, coding properties such as exon length or number of exons in a gene, are not so dependent on GC content. This is corroborated in Fig. 6, which shows the dependence of the mean values of exon and intron lengths with respect to the GC content. As shown in the figure, the intron size decreases as a function of local GC content, with a

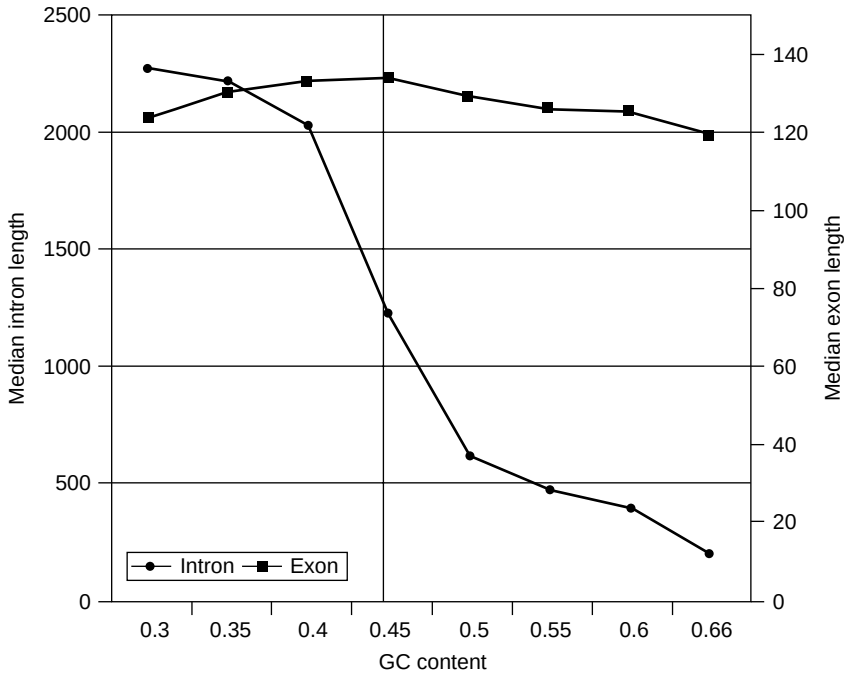

Fig. 6 Dependence of the exon and intron average length on the GC content (from IHGSC, *Nature* **409**, 860–921, 2001, see original paper for details).

sharp transition occurring at ~45% GC, in agreement with the boundary between the genome core and the empty quarter.

Different efforts have been carried out to localize DNA fragments with different GC-richness on human metaphase chromosomes. The results support that the two gene spaces are characterized by different structural and functional properties. GC-poor isochores are located in G-bands and in H3$^-$ R-bands (in which H3 isochores are not detected), which generally exhibit a closed chromatin structure. On the other hand, the chromosomal localization of the GC-richest isochore family (H3) corresponds to T-bands. Moreover, since isochore families are unevenly distributed over the chromosomes, some chromosomes (17, 19, 22) have a very high percentage of genome core fragments, whereas other ones (4, 13, 15, 18, X, and Y) have a very small percentage or no genome core segments at all. The explanation for such an uneven distribution is still unclear.

2.2.4 Evolutionary Hypothesis About Isochores

As explained in previous sections, a correlation is evident between %GC content and the gene density of a chromosomal domain. One of the questions raised by this distribution is which one is under the greatest selective pressure: does the %GC content of a chromosomal domain control the number of genes located within that domain or does the clustering of genes into local domains determine the %GC content? The basic question being asked is at what level does natural selection operate

to produce such clearly demonstrable isochore domains in the human genome. Although this is a simple question, its answer is buried under the complexity of separating gene function from genome structure. We know that in coding domains there is a codon bias toward GC dinucleotides in warm-blooded vertebrates that result in a higher proportion of hydrophobic and amphipathic amino acid residues in proteins. One of the consequences of maintaining protein structure at higher body temperatures (37 °C) may be an increased need for hydrophobic interactions, thus resulting in a strong selective pressure for greater %GC representation in coding domains. However, the regulation of gene expression events appears to be just as critically controlled by %GC composition. Large-scale alterations in chromosome structure related to GC hydrogen bonding in double-stranded DNA can alter the accessibility of expression regulators to promoter domains.

It is difficult to isolate the impact that nucleotide sequence has on the selective fitness of an organism because a gene sequence is operational at two levels: (1) gene level function, and (2) structural conformation of local chromosomal domains. The difficulty of dealing with the concept of isochores is that this phenomenological observation crosses these two functional boundaries. Identifying the most proximal cause of such sequence anomalies (% frequencies showing such large departures from a random frequency distribution) is difficult because of the interdependence of structure and function.

2.2.5 Recent Discussions About the Isochore Concept

In this chapter, we have presented the standard definition of the isochore concept. However, the plausibility of the "isochore" concept has been questioned in several recent papers. Most of the concerns focus on the definition of "compositional homogeneity" – which some recent papers interpret as relative to random sequences – on the capability of ultracentrifugation experiments to define intrasequence heterogeneity, and on the fraction of the human genome to which the isochore concept applies. Some of these concerns have been addressed by Bernardi and Clay. Other authors point to the fact that isochore families were not originally defined from sequence data, but from ultracentrifugation experiments that resulted in 5 (after a somewhat arbitrary reduction from 13) major overlapping Gaussian distributions of absorbance, each one associated with a distribution of GC content. Therefore, because of the overlapping, a DNA fragment with a given GC content could correspond to different isochore families, which is a drawback to the definition of isochores based only on the sequence data. In order to overcome this problem, the boundaries (in CG content) among isochore families can be defined in such a way that they do not overlap, as we directly did when we introduced isochores. However, this definition is different from the original one.

The concepts of homogeneity and heterogeneity are largely subjective, and therefore much more research is necessary to develop better methods to perform segmentation of the genome into regions of different GC content. However, the existence of GC-rich and GC-poor regions in the human genome has now been fully corroborated by the draft sequences. Moreover, many studies have indicated an interesting correlation between GC content domains and various biological properties, including transcriptional regulation, local DNA replication,

patterns of codon usage, and the posttranscriptional processing of mRNA. Although much more work is still necessary before reaching a clear conclusion, the study of the distribution of guanidine and cytidine nucleotides will be a primary route to a better understanding of the gene distribution in the human genome.

3 Microscale Structure

In Sect. 1, we saw that the most recent estimates predict that the human genome contains a number of genes in a 30 000 to 40 000 range, a surprisingly small amount if we consider that much less complex organisms such as *Arabidopsis* and *Caenorhabditis* are also supposed to have a similar number of genes. As a first attempt to explain this fact, it seems intuitive to think that we may need to look outside of the coding sequences for the mechanisms that generate the complexities inherent in the human development. This intuition has been confirmed by an increasing amount of recent work in the literature, which suggests that the explanation is buried in introns and intergenetic regions, once referred to as "junk sequence." An important piece of supporting evidence is the obvious correlation between the GC levels of coding sequences and their context DNA fragments in which they are embedded. The study of the patterns in these noncoding regions may provide hints on the gene distribution over the genome.

3.1 Gene Model

Human genes are characterized by long introns, which on average constitute more than 90% of a gene. The median of the intron sizes (~1000 bp) is about 8 times the median of internal exon sizes. Moreover, the distributions of gene size, intron size, and the number of introns in a gene present significant fluctuations, much higher than those of exons. This is shown in Fig. 7, where the exon and intron size distributions are shown for the human, the worm, and the fly. It is interesting to note that the exon distribution is very similar in the three species, while the intron distributions (and therefore the overall gene size distribution) presents a much higher difference among the three species, with the human distribution being the one with greater variability. Another interesting observation is the dependence between intron sizes and the local GC-level: on average, GC-rich regions show a scarcity of long introns, while the median intron size in GC-poor regions is about twice that encountered in the whole genome. This special pattern is believed to be related to the splicing mechanism. For example, Bernardi suggests that long introns in the gene-poor fragment may facilitate alternative splicing in tissue dependent genes.

Despite the controversy over the exact nature of isochores (whether phenomenological or functional), it is clear that the most prominent feature of protein-coding genes is their higher than average GC nucleotide (guanidine and cytidine) compositions. Majewski has compiled sequence analysis data for 10 858 identified human genes into a "model" gene composed of

- 2000 bp promoter region
- first exon (250 bp, mostly nontranslated)
- first intron (just 500 bp on the 5′ and 3′ ends)

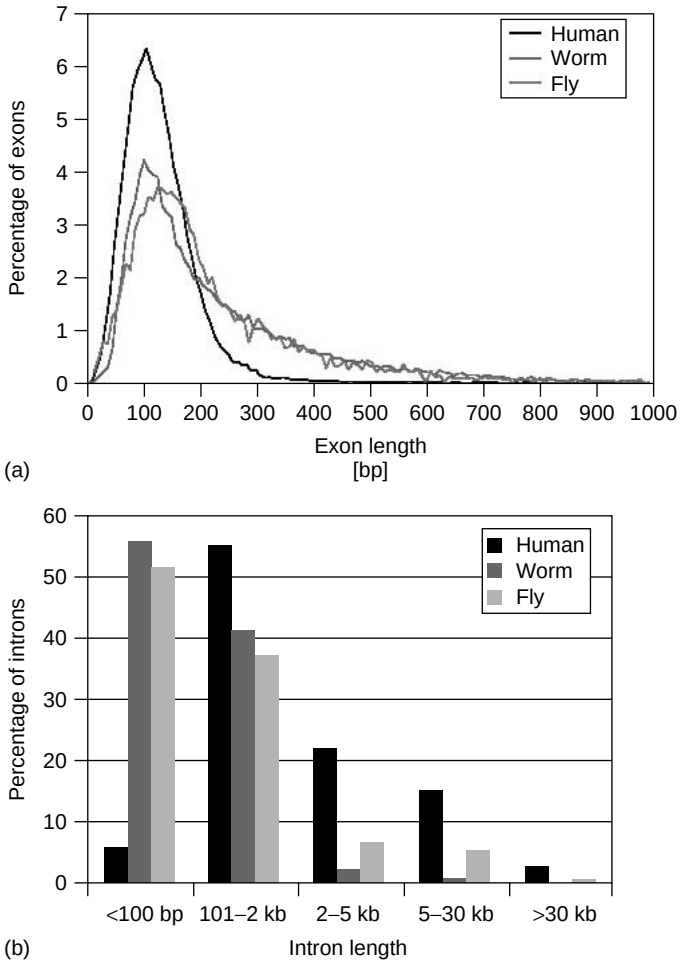

Fig. 7 Size distributions of exons, introns, and short introns for worm, fly, and human genomes (from IHGSC, *Nature* **409**, 860–921, 2001, see original paper for details).

- internal exon (500 bp)
- internal intron (just 500 bp on the 5′ and 3′ ends)
- terminal exon (500 bp)
- termination domain

In most vertebrates, exons and introns can be much larger than 1000 bp total. For this sequence analysis, the calculation of frequency distributions for each of these domains was executed by examining nucleotides at the same sequence positions relative to the splice sites and/or boundaries known to exist in the 10 000 + genes used for this study. Each of the essential gene elements categorized above have been condensed into a grand average in terms of nucleotide composition (Fig. 8).

Although genes have higher GC compositions, the distribution of G and C

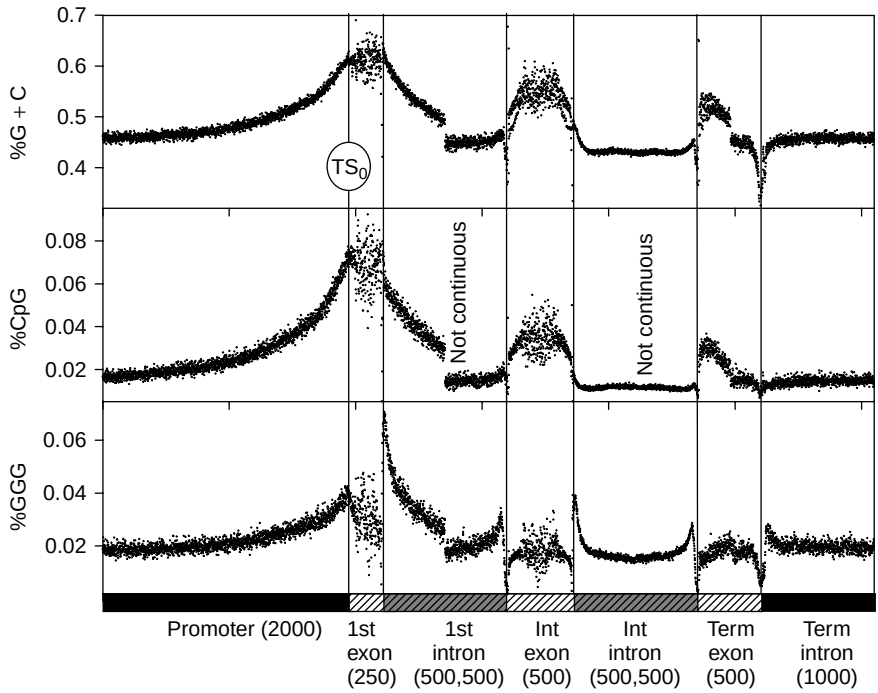

Fig. 8 The distribution of sequence elements within an archetypal model gene are shown for a composite of 10 858 currently known human genes (Majewski and Ott 2002). Gene positions were aligned relative to the transcriptional start site (TS_0) and nucleotide frequencies were then averaged by relative position to determine the frequency of (a) % composition of cytidine and guanidine nucleotides, (b) 'CpG' dinucleotide motifs, and (c) "GGG" trinucleotide repeats. The sequence composition evidences several striking patterns in terms of the distribution of G and C nucleotides across all described genes (int = internal; term = terminal). [These distributions are replotted from the original Fig. 5 of Majewski, J. and Ott, J. (2002) Distribution and characterization of regulatory elements in the human genome, *Genome Res.* **12**, 1827–1836 using the original data kindly provided by J. Majewski.]

nucleotide frequencies is heterogeneous, with distinct intragenic locales. Outside of a gene domain, the GC content is between 42 and 45%. As one moves (5′ to 3′) into the promoter domain (2000 bp from the first exon) the GC content steadily increases to upwards of 60% at the beginning of the first exon (transcription start site, TS_0). The first exon has an average GC composition of 55 to 60% without any spatial differentiation. Most interestingly, at the splice junction between the first exon and the first intron there is a sudden increase in the GC composition. The 5′ end of the first intron (the splice junction with the first exon) has a GC distribution of nearly 65%. This frequency then rapidly declines further into the intron, but exhibits another rise as one approaches the 3′ splice junction with the first internal exon, as shown in Fig. 8.

The internal exons exhibit an interesting feature with a peak in GC content between 100 to 150 bp away from the splice

junction on either end, and then a decline in GC frequency as the splice site is approached. The internal introns evidence an opposite distribution with peaks in GC frequencies right at the splice junctions, and then declines to average levels roughly 200 bp away from the junction. The presence of such significant spatial shifts in nucleotide frequencies (especially at boundary junctions) is suggestive of functional mechanisms for regulating gene transcription and RNA posttranscriptional processing. However, it is unlikely that the mere sequence composition of a gene element provides a specific enough mechanism for exerting fine-scale regulatory controls. Rather, spatial shifts in GC distributions are more likely to reflect changes in the distribution of specific regulatory motifs that have higher proportions of GC nucleotides.

The archetypical gene model provides a unique look at how human genes are structured. The most significant feature in terms of the potential regulation of gene expression rates is the distance, both upstream and downstream from the TS_0 site, for which significant biases for the presence of GC motifs can be found. There are several regulatory motifs in this wide domain whose mechanistic actions to control gene transcription events are determined by cytosine and guanidine nucleotides, such as CpG and GGG motif. The potential regulatory domain for direct transcriptional control over a single gene locus comprises a ~2000 bp window around the TS_0. This domain results in a large number of potential regulatory interactions among the special motif distributions; however, regulatory regions can exist much farther upstream. This range of potential control states is thus likely to establish the high recombinatorial repertoire in potential expression patterns that is essential for the evolution of complex network interactions.

3.2 CpG Islands

"CpG islands" (CGIs) are one of the most prominent regulatory motifs in the human genome. The name refers to a simple dinucleotide motif of a cytidine and guanidine (5′-C-p-G-3′) within a GC-rich chromosomal domain. The frequency of CpG dinucleotides in the human genome is much less than expected, as a result of a methylation/mutation process. Generally, about 60 to 90% of the CpG sequences within a genome are methylated. DNA methylation is a reaction, specific to cytosines in CpG dinucleotides, that transfers a methyl group from *S*-adenosylmethionine to the C5 position of cytosine. Methylated cytosines have a high mutation rate to thymines (CpG to TpG), which has been confirmed by many studies on DNA polymorphism or genetic diseases. This process, known as GC suppression, leads to an overall reduction in the frequency of GC content to about 41% of all nucleotides and a further reduction in the frequency of CpG dinucleotides to about a quarter of their expected frequency. CpG islands (those C–G motifs in GC-rich domains) are primarily unmethylated and are the exception to CpG underrepresentation in the genome.

CpG islands make up 1 to 2% of the DNA sequence. They are of great importance because the cytidine base can be methylated such that the methyl group extends upwards into the major groove of the DNA helix (Fig. 9). The intrusion of these polar methyl groups essentially disrupts the potential interactions between the local nucleotide bases and transcriptional regulators in this domain by altering the

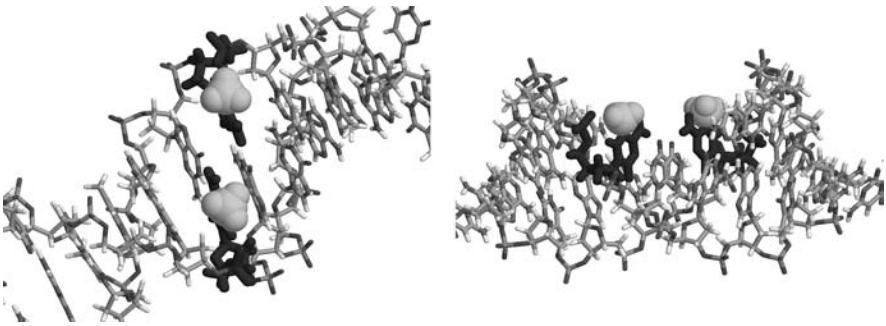

(a) (b)

Fig. 9 Three-dimensional structure models for the methylation site in a CpG motif; (a) top view perpendicular to the major groove, (b) side view parallel to the major groove. The 5-methyl-cytidine residues are drawn in the dark wireframe representation, the methyl group atoms are shown with Van der Walls atomic radii as gray spheres. The primary impact of the methylation appears to be a redistribution of charges around the C–G hydrogen bonding. The nonpolar methyl group is adjacent to the amino ($-NH_2$) group that would establish one of three hydrogen bonds with the complementary G-residue hydroxyl group (coordinate data from PDB ref# 1IG4).

local charge distribution. Usually, both the (+) strand motif and the complementary (−) strand motif cytidines are methylated (m5C) to increase the density of methyl groups within a small area (see Fig. 9):

$$5'-C5m-G-3'$$

$$3'-G-m5C-5'$$

This provides an effective blocking mechanism to "silence" regions of DNA from being transcribed and is one of the most active areas of current research into epigenetic mechanisms of gene regulation.

CGIs commonly occur in the promoter domains of functional genes. Overall, this pattern suggests that the majority of the methylated CpG motifs play a role in gene silencing (blocking transcription of noncoding regions, repetitive DNA elements, transposable elements, pseudogenes, etc.). But in active promoter domains, the methylation states of CpG motifs are not fixed such that changes in methylation can be used for up-and-down regulation of transcriptional activities at a specific gene locus.

An attractive property of CGIs is their significant structural difference from the rest of the genome. In the work of Gardiner–Garden and Frommer, CGIs were formally defined as sequences greater than 200 bp in length, with a GC content greater than 0.5, and with an observed to expected GC content ratio (CpG_{obs}/CpG_{exp}), greater than 0.6. This definition has been widely used in later studies. Recent work has suggested more specific structures for CGIs, especially for those potentially corresponding to promoter regions. This makes it practical to search for CGIs using computing tools. In the work from the International Human Genome Sequencing Consortium, a computer program was implemented to identify CpG islands and 28 890 CGIs were found in the draft human genome sequence with repeat regions masked. The count of 28 890 CGIs is reasonably close to previous estimates

of about 35 000, which were obtained by experimental methods. The convenience of CGIs searching has led to effective approaches for gene estimation. In 2000, Grunau et al. determined features that discriminate the promoter-associated and promoter-nonassociated CpG islands and used them for large-scale human promoter mapping.

As shown in Fig. 10, the density of CpG islands shows important variations in different chromosomes and correlates positively with estimates of the gene density. Chromosome Y is the one presenting the lowest density, with 2.9 islands per Mb. On the opposite extreme, chromosome 19 presents a density of 43 islands per Mb. The average for all the genomes is 10.5 islands per Mb, with most of the chromosomes having 5 to 15 islands per Mb. Similar trends are observed when the percentage of nucleotides contained in CpG islands is considered.

The coherent distribution of CpG islands within the model gene demonstrates the general feature that these motif elements comprise for all human genes (Fig. 8). There is a fourfold increase in CpG frequency in the promoter domain as one moves 5′ to 3′ from −2000 bp down to the first exon (0 bp at transcriptional start site). The high frequency occurrence of CpG islands is maintained through the first exon (+250 bp) and into the first intron. This distribution on either side of the first exon boundary suggests a potential regulatory influence of CpG methylation patterns extending from −1000 bp upstream of the TS_0, across the first exon (0–250), and then 500 bp into the first intron.

The position of CpG islands at the TS_0 appears to be a crucial mechanism for temporal and spatial regulation of gene expression rates. In Fig. 11, we have summarized gene expression data for 24 tissue types from both adults and embryos (orthologs in mouse embryos) on the basis of the presence/absence of a CpG island at the TS_0 site. In human adult tissues, there is a significant trend for genes exhibiting tissue-specific expression patterns to have a CpG island at the

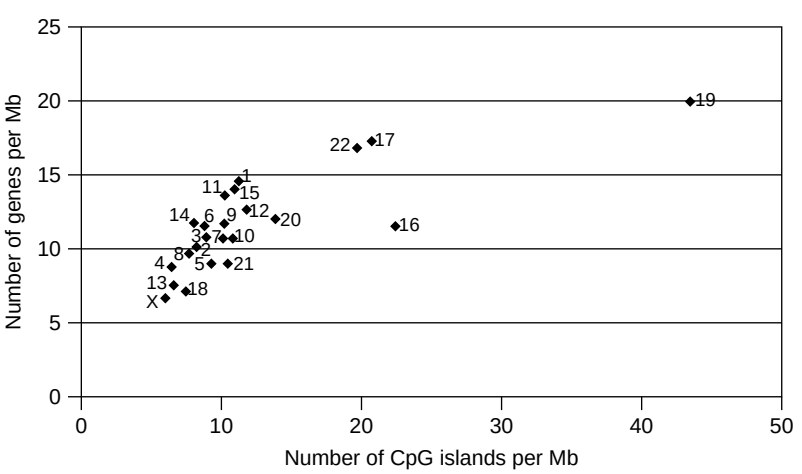

Fig. 10 Number of CpG islands per Mb versus number of genes per Mb (from IHGSC, *Nature* **409**, 860–921, 2001, see original paper for details).

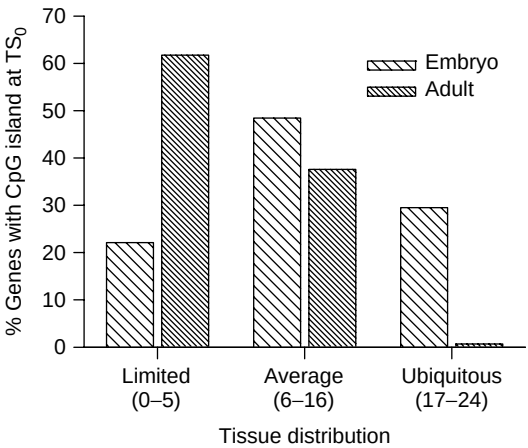

Fig. 11 Frequency of genes with CpG-island regulatory elements in their transcriptional start domains. Genes are separated into three categories depending the number of tissues that express those genes (collated expression data from 24 tissue types). Expression categories range from "limited" (expression in less than 5 tissue types) to ubiquitous (expression in more than 17 tissue types). Regulation of gene expression in adult tissues clearly demonstrates a frequency relationship between tissue-specificity and the occurrence of CpG islands at the TS_0 (data replotted from Ponger, L., et al. (2001) Determinants of CpG islands: Expression in early embryo and isochore structure, *Genome Res.* **11**, 1854–1860).

TS_0 (62%), in contrast to genes that are ubiquitously expressed (housekeeping genes) where only 0.7% exhibited a TS_0 CpG motif (Fig. 11). This pattern strongly indicates that CpG methylation may be one of the primary mechanisms by which tissue-specific expression patterns in adults are regulated. The pattern of CpG distributions in embryonic gene regulation does not appear to be correlated with tissue-specific expression.

3.3 GGG Trinucleotide

Another important regulatory motif is the "GGG" trinucleotide, which will obviously be more prevalent in GC-rich (isochore) domains. These motifs have been shown to be overrepresented within introns and have been implicated to function in regulating RNA splicing during posttranscriptional processing. In the archetypal model gene, the distribution of GGG trinucleotides shows a predominant location at the 5′ ends of intronic (noncoding) domains. This distribution supports the hypothesis that these motifs are critical control components of RNA posttranscriptional processing by the spliceosome protein complex. Also, it is quite significant that the GGG frequency is much lower than expected near the TS_0 and across the first exon. This deficiency suggests that the GGG motif does not commonly play a role in the regulation of gene expression rates; however, a few examples of GGG-mediated gene regulation do exist. It is important to keep in mind that the archetypal model just represents an overall average nucleotide distribution.

A related motif with less known function is the "CCC" trinucleotide. Splicing control elements (SCEs) can only process nucleotide motifs on the coding (transcribed) strand so that "GGG" and "CCC" distributions are not functionally reflexive (simple reverse complements of each other). The frequency of CCC motifs on the coding strand must present a distinct regulatory structure for SCE recognition, and although the representation of CCC motifs is not as significant in the archetypal model gene as are the GGG motifs, they evidence the same prominence in intron borders and scarcity in the promoter and

first exon domains. Consequently, the observed increase in %GC content flanking the TS_0 can primarily be attributed to an increase in the frequency of CpG islands, and not the distribution of trinucleotide repeats of cytidine or guanidine.

3.4 Pseudogene Distributions

Pseudogenes are a subclass of genes in which the coding domain is no longer subjected to selection forces acting on cellular fitness. There are two primary routes of pseudogene formation: (1) duplication of a chromosomal region, and (2) reverse transcription of an mRNA molecule into cDNA and consequent reintegration of that processed transcript back into the genome. In either case, the end result is that two copies of an operative gene now exist in the genome, one of which will continue to function under the selective pressures it has been adapted for, and the other is now independent of any selection pressures (i.e. it does not impact the cellular fitness of an organism if the copy is expressed or not expressed, because the original operative gene is continuing its function). The two cases are easily distinguished because with a gene domain duplication, all the noncoding regulatory regions are preserved in the copy (promoter, introns, terminal domain), while in the reverse transcription of mRNA (retropseudogene), the mRNA transcript has been processed such that the promoter, introns, and much of the terminal domain have been removed, or modified. For reverse-transcribed pseudogenes, they must become integrated back into the genome downstream from an active promoter for any possibility of transcriptional activity.

Pseudogenes can be actively transcribed and in that respect are truly "genes" other than the fact that there is no selective pressure on them to actually do anything useful. On chromosomes 21 and 22 (the most thoroughly studied to date), a total of 454 pseudogenes have been identified, and estimates based on the draft human genome to date suggest that there will be between 23 000 and 33 000 pseudogenes in total. (http://bioinfo.mbb.yale.edu/genome/pseudogene/) Thus, pseudogenes are likely to play an important role in evolution by allowing operative genes to obtain new functions without immediate detrimental effects to an organism. However, most pseudogenes appear to be destined for genomic oblivion once they are formed. Without a strong selection on function, random drift in the nucleotide sequence rapidly (on an evolutionary timescale) can introduce premature termination signals or frameshifts that result in nonfunctional mRNA transcripts. There are very few documented cases of pseudogenes evolving into new operative genes.

The documentation of pseudogenes in a genome is not easy. It is likely that there are numerous operative genes in the human genome that have a pseudogene origin. Because most genes themselves are located in GC isochore domains, they have higher than average %GC compositions. Once a pseudogene is formed, random mutational drift will tend to reduce the %GC composition toward the genome average. If the pseudogene adapts to provide a new function, its phylogenetic sequence affinity can be obscured by the shift in nucleotide composition away from its ancestral operative gene. Because only ~1.5% of the human genome is directly involved in coding for protein sequences, there is a large background of DNA for many recombinatorial possibilities to arise from rearranging DNA. A recent study has been able to identify the

incorporation of several mitochondrial genes into the human genome, suggesting a new route for pseudogene incorporation when eggs are fertilized by sperm. As our understanding of these processes increases, it is likely that we will be able to develop the computational tools to more fully identify operative-pseudogene relationships.

3.5
Transposable Elements

There are sequence elements in the genome that are very efficient at duplicating and moving their local DNA domains. These transposable elements (TEs) were originally described as parasitic DNA, because of their ability to self-replicate unchecked within a genome environment. However, our understanding of their role in genome structuring is just beginning to identify them as absolutely necessary or at least mutualistic elements. At present we know little about their molecular mechanism of operation, and are left to only speculate on the potential importance of their role in genome evolution. As discussed in the previous section on pseudogenes, the ability to create copies of genes is an important mechanism for either deriving novel functions from existing proteins, or to give existing proteins different patterns of expression. TEs appear to be the primary vehicle by which genomic material is shuffled around into alternative recombinations as a natural experiment resulting in new genes and expression patterns. "NEW" is not necessarily better and these alterations may or may not be selectively advantageous. However, such rearrangements can provide a greater molecular diversity within an organism for selection to then act upon. A large body of literature on TEs exists, particularly in terms of structure and function in *Drosophila*, and the reader should consult these resources.

4
Metascale Integration

The first pivotal study of functional genomics lead to Jacob & Monod's seminal description of the *Lac* operon in *Escherichia coli*. Their summary description of the "one gene equals one protein" organization of DNA set the stage for the next three decades of genomic research focusing on linear, direct models of gene interactions. This approach worked well for the single-target cloning strategies that were implemented in the 1970s to late 1980s. However, in the 1990s, as high-throughput technologies were developed, the collection of gigabytes of gene expression data opened up new avenues for approaching the functional organization of a genome as a three-dimensional structure, that is, one in which the genetic code was not just operative in a long, one-dimensional array of four nucleotides. It is now clearly evident that gene location within a chromosomal domain has a significant impact on transcriptional regulation.

The interaction between genes, their protein products and their chromosomal locations are now routinely considered in network organizations of "connectivity." Long gone (in terms of the volume of literature that has appeared in the last five years) is the focus of identifying linked companion genes in operon-like, downstream–upstream units. Even the paradigm of "one gene equals one protein" needs qualification as alternative splicing, bidirectional coding and post-transcriptional mRNA processing has lead us to realize that single coding domains

may produce several proteins, given the multiple mechanisms that are involved between the initiation of gene transcription and the folding of a final protein product.

Given the importance of describing how genes are "connected," there is considerable attention being focused on identifying the mechanistic relationships that may exist between coordinated patterns of gene expression (i.e. network associations) and the physical mapping of those transcriptionally linked genes. With the current estimate of only ~35 000 genes in the human genome, there is strong support for the idea that interactive mechanisms between chromosomal locations and mRNA processing are just as important in determining the physiological "fitness" of genes (e.g. their contribution to organismal survival) as are primary nucleotide sequences.

In this section, we will cover the essential relationships between expression connectivity and physical location, with the goal of describing the potential links between gene location, function, and evolution.

4.1 Associative Expression Networks

The field of proteomics has established the clear necessity of describing the coordination between gene expression rates and functional protein activities. This includes a focus on describing the multiple interactions between genes and/or proteins in an environmental context. Correlating single gene expression events (mRNA transcription) with protein-level functioning (specific activities) has been relatively unfruitful, with many variations in gene expression levels not reflected by corresponding changes in their protein pools. We now know that even for such "housekeeping" proteins as lactate dehydrogenase and Na^+- and K^+-ATPase, levels of gene expression and levels of functional protein activity are usually not directly correlated. The regulation of gene expression and protein activities appears to be coordinated at a higher level of physiological organization. We call this organizational structure a "network" to reflect the many parallel connections between gene expression events and protein activity levels.

Gene expression experiments, such as microarrays and gene chips, are widely used to identify genes that appear to function together because of their coordinated expression behavior. However, the best strategy for identifying such "coordination" in expression data is open for debate. The most general approach adopted has been one of paired correlation analysis to produce gene clusters associated by expression patterns. To extract finer-scale interactions between genetic components in terms of induction, repression, modulation, and coordination requires a more concerted analytical scheme to deduce network connectivities between gene or protein members. This level of analysis involves establishing an interactive structure or network around which the data can be quantitatively assessed. Both Boolean and Bayesian network models are currently employed for resolving finer features of interactions and dependencies in multivariate expression data.

One of the key components to describing complex network interactions is designing experimental manipulations of the expression system that will provide a wide range of expression data. The strategy is described as "extreme pathway" modeling or "perturbed expression" profiling. The idea is to generate as many potential variants of expression of different targets due to treatment stressors or transgenic manipulations (knockouts, overexpression).

Gene and protein interactions are then assessed across this diversity of expression patterns. In general, genetic systems now appear to be organized into "small-world" networks, which means they have discrete subnets or local communities that respond coordinately. This observation can simplify an interactive network analysis by limiting the size of the subnet units to be considered. Complete network datasets are cumbersome and suffer from statistical insensitivity because of the large numbers of component members. Isolating smaller subnet units in perturbed expression profiles can lead to the identification of statistically robust, local-community interactions.

An awareness of the adaptive versatility of subnet structuring is rapidly developing in environmental genomics. Simple connections between densely integrated subnets allow for (1) a large degree of flexibility in coordinating the regulation of biochemical and cellular activities, (2) provide a large recombinatorial mechanism for the *de novo* appearance of new functionalities, and (3) are robust in terms of their internal buffering capacity against external inputs. Subnet rewiring is potentially one of the dominant mechanisms of epigenetic interactions impacting species' environmental adaptations and evolution. Using high-order structure analyses to relate complex expression data (microarrays) to cellular function has a large potential to provide new insights into how genomic-level interactions are structured in adaptive network associations.

4.2 Functional Coordination and Physical Mapping

Unfortunately, one of the fundamental, guiding principles of expression network analyses is that coordinately transcribed genes will be functionally related. This assumption is based on the assumption that gene transcription rates are independent and that equivalent patterns of expression between two genes are thus indicative of parallel promoter activities. However, recent studies that have mapped transcriptional activities (from microarray and serial analysis of gene expression (SAGE) analyses) to the corresponding chromosomal gene locations have demonstrated that transcriptionally active domains are heterogeneously distributed throughout the genomes of humans, flies, nematodes, and yeasts. These domains are a very significant feature of mRNA expression dynamics because genes that are closely located on chromosomes are likely to evidence "correlated" expression patterns simply because of their physical location, rather than any functional association.

A Human Transcriptome Map (HTM) has been assembled by the integration of several independent databases from SAGE profile data, the draft human genome, and chromosomal maps. This HTM is accessible through a web portal based at the Amsterdam Medical Center, University of Amsterdam. (http://bioinfo.amc.uva.nl/HTM-bin/index.cgi/) Although the statistical determination of regions of increased gene expression (RIDGE) domains needs further exploration, the basic observation of heterogeneous distributions of transcriptional activity for 2.5×10^6 SAGE transcript tags across 12 tissues types is visually evident in the excellent array of web tools the HTM site provides for exploring specific expression locations. On an average, RIDGEs contain between 6 and 30 mapped genes per centiray (genes cR^{-1}) in contrast to the 1 to 2 genes cR^{-1} in domains that are low in transcriptional activity. In addition, RIDGE segment genes evidence average expression levels that are

sevenfold higher than the overall genome average. A rough calculation using gene targets and gene transcription rates would suggest that total transcriptase activity in RIDGEs is between 20 and 200-fold higher than in weakly expressed domains.

The selective mechanism (i.e. the advantage in molecular and biochemical "fitness") for the reason why genes are heterogeneously distributed across chromosomal domains and the reason why transcriptional activities evidence such large spatial differences remains unknown, but certainly not for a lack of speculation. Although there are many logical ideas on the table, the stumbling block is that biological systems are not logical; they do not arise from optimized design principles. Instead, chromosome structure and function have been determined purely by operational efficiency. Logically, the human genome could be as small as ~50 Mb, but operationally, functional competency has required 3.3 Gb, at least at this point in time in the ongoing evolution of the hominid genus (*Homo*).

With this in mind, there appeared to be preliminary support for a hypothesis that tissue-specific patterns of gene expression were coordinated by clustering genes into functional domains in order to unify transcriptional controls. This direct explanation could account for a large component of the variable distribution of genes into distinct, tissue-specific cassettes. However, in a summary comparison of the expression levels of over 11 000 genes from 14 different tissues, Lercher et al. demonstrated that those genes that are tissue-specific in their expression do not cluster together into local domains. To the contrary, their analysis shows that it is the ubiquitous housekeeping genes that do show strong positional clustering within a chromosome. This distribution suggests that the apparent clustering of genes that share high expression rates is primarily a consequence of the local clustering of housekeeping genes, or more generally, genes whose products are necessary in most human cell types.

Although the polytene banding patterns on chromosomes are direct evidence of large-scale spatial heterogeneity in chromosome structure, there does not appear to be any correlation between specific chromosomal structures and transcriptional activity. The expression domains in the fly genome reveal ~200 groups of adjacent genes that have similar patterns of expression across 80 experimental conditions. These groups have 10 to 30 nonrelated genes (in terms of cellular function) covering an average domain size of 100 Kb (range of 20–200 Kb). However, these domains do not evidence any spatial correlations to polytene structure.

The intrachromosome domain model of nuclear organization suggests that genes are more likely to be located at the surface of chromosome territories. In humans and mice, there is substantial evidence to indicate such a differential location of active genes at the surface of chromosomal territories. But the locations of both housekeeping and tissue-specific genes do not appear to be confined to just the peripheral domains of chromosomal territories, indicating that the tight and loose banding of chromatin may not restrict the transcriptional machinery from accessing different locations. There is clearly a near-neighbor effect on the transcription rate of most genes and it is likely that the intercorrelation of gene density and local transcriptional activity influences the gene organization of chromosomes, rather than the activity of individual promoter domains functioning in isolation.

4.3 Physical Colocation and Dislocation

For a better understanding of the distribution of functionally related genes, we have summarized the chromosomal distributions of the genes coding for enzymes involved in the glycolysis and the tricarboxylic acid pathways (TCA) using the Kyoto Encyclopedia of Genes and Genomes (Table 1; http://www.genome.ad.jp/kegg/kegg2.html). All human cells are dependent upon the oxidation of glucose via glycolysis and the TCA cycle for generating metabolic energy (ATP and reduced nicotinamide adenine dinucleotide ($NADH,H^+$)). Although these pathways are highly integrated in terms of the production (glycolysis) and utilization (TCA) of acetyl-coenzyme A, the constituent enzymes have very different intracellular locations: glycolysis = cytoplasm; TCA = mitochondria. Thus, one might expect that the transcriptional requirements for maintaining the protein pools in these different cellular locations could serve as a selective force to organize the genes into local domains (i.e. mitchondrial group, etc.). The physical map locations for these genes were obtained from the LocusLink database at the U.S. National Center for Biotechnology Information. (http://www.ncbi.nlm.nih.gov/LocusLink/)

Using the normalized chromosomal ideograms, Fig. 12 traces the sequential order of gene loci that are involved in the catabolism of glucose. Although these

Tab. 1 A serial list of the enzymes involved in glycolysis and the TCA cycle. Chromosomal locations were obtained from the LocusLink database and the loci codes used in Fig. 12 are indicated in the "map" column.

EC#	*Enzyme Name*	*Location*	*map*
2.7.1.1	Hexokinase	10q22	a
5.3.1.9	Glucose-P-isomerase	19q13.1	b
2.7.1.11	6-phosphofructokinase	21q22.3	c
4.1.2.13	Fructose-bis-P aldolase	16q22-q24	d
1.2.1.12	GAP-dehydrogenase	12p13	e
2.7.2.3	Phosphoglycerate kinase	xq13	f
5.4.2.1	Phosphoglycerate mutase	7q31-q34	g
4.2.1.11	Enolase	1p36.3-p36.2	h
2.7.1.40	Pyruvate kinase	1q21	I
1.1.1.27	Lactate dehydrogenase	11p15.4	j
1.2.1.51	Pyruvate dehydrogenase	xp22.2-p22.1	k
4.1.3.7	Citrate synthase	12p11-qter	l
4.2.1.3	Aconitase	9p22-p13	m
1.1.1.41	Isocitrate dehydrogenase	15q25.1	n
1.2.4.2	a-keto dehydrogenase	7p14-p13	o
2.3.1.61	S-succinyltransferase	14q24.3	p
6.2.1.4	Succinyl-CoA synthetase	2p11.2	q
1.3.5.1	Succinic dehydrogenase	5p15	r
4.2.1.2	Fumarase	1q42.1	s
1.1.1.37	Malate dehydrogenase	2p13.3	t
6.4.1.1	Pyruvate carboxylase	11q13.4	u
3.1.3.11	Fructose-bisphosphatase	9q22.3	v

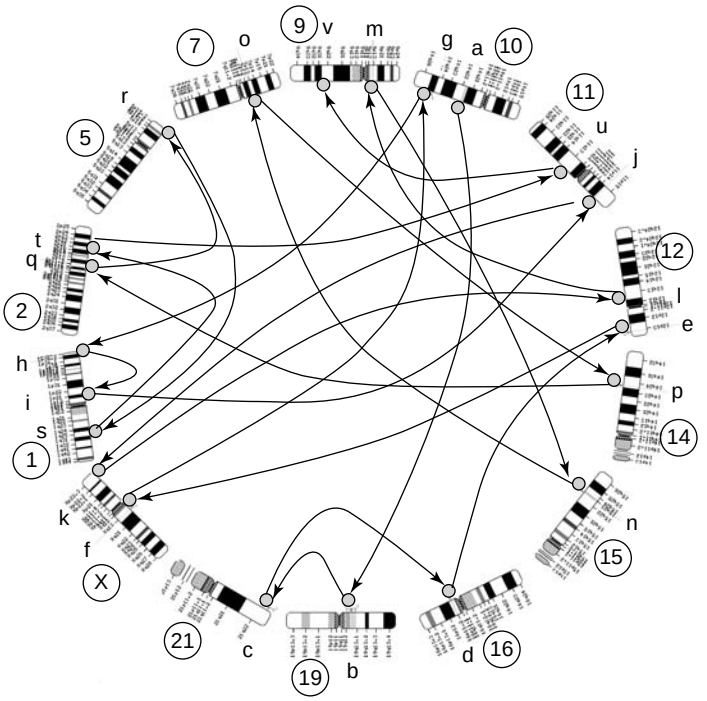

Fig. 12 The chromosomal locations of the gene loci involved in glycolysis and the TCA cycle. Although these metabolic pathways are dominant features in all human cells and are organized into two discrete cellular compartments (cytoplasm and mitochondria), there is no apparent colocalization of these genes within common chromosomal domains.

genes are as fundamental as one can get in terms of housekeeping genes in humans, they are not localized into operational groups of "glucose metabolism" or "mitochondrial locations" or "TCA cycle" cassettes in which their transcription rates could be coordinately controlled.

This result may appear to contradict the positional analysis of Lercher et al. where they conclude that it is the housekeeping genes that do show strong positional clustering within a chromosome. The above analysis of the glucose oxidation housekeeping enzymes (Fig. 12) is meant to illustrate that even though "housekeeping" genes appear to be coordinately located, "housekeeping" is only an operational definition that could easily apply to more than 80% of the genes expressed in any tissue. Consequently, with so many "housekeeping" genes in a genome, and those genes concentrated into domains, there is a high probability that neighboring genes in a location will be "housekeeping" even though they may share no functional relationship in terms of their biochemical activities.

A first glance at Fig. 12 begs the question "is this an organized pathway?" This is because we think of pathways as linear, chain reactions and anticipate a linear organization in the underlying genetic structure (such as in a *Lac* operon type model). However, in

most "housekeeping" pathways, component metabolites are often shared with multiple other metabolic pathways, creating an interactive network association of biochemical activities. Thus, there is no realistic mechanism for all the protein-coding loci of a pathway to be colocated if each gene member of that pathway is also a component in several other biochemical pathways. So the startling pattern evident in Fig. 12 is that the 22 primary genes involved in the glucose oxidation machinery are dislocated across 14 chromosomes. And in no instance are any two sequential protein genes colocated.

4.4 Dynamic Expression Linkages

In looking at the dislocation of genes coding for glycolytic proteins in Fig. 12, we struggle with an inability to understand the design by asking the question "why?" Why should functionally related genes be distributed randomly (without apparent pattern) across half of our chromosomes? However, a more direct route for us in terms of deciphering a design mechanism would be to start by trying to answer the question "why not?" From this perspective, a list of observations favoring a random distribution can be summarized as follows:

1. Natural selection is operative at the level of phenotype dysfunction. As long as any process (molecular, biochemical, cellular) does not diminish the overall fitness of the organism in its entirety, that trait is selectively neutral (the mechanism of natural selection eliminates individual genotypes with deleterious traits). In terms of gene organization, it makes no difference in selective fitness for an organism if metabolic pathway genes are colocated or not. As long as their expression levels are appropriately regulated, the way in which they are distributed across the chromosomes does not impact metabolic performance.
2. No chromosome is an island. Distributing functionally related genes across many chromosomes establishes an interdependency on chromosome composition. This could serve to maintain genome integrity over evolutionary timescales because the loss or gain of a chromosome is then catastrophic (i.e. severely compromises the fitness of the individual). Note that the glucose oxidation pathway alone "links" 14 of 24 chromosomes.
3. If two genes need to be coexpressed simultaneously, they do not have to reside next to each other. The only necessary constraint is that they need to be located downstream from promoter domains that respond to the same regulatory controls. The implication here is a very important point because it suggests that although gene location may not show distributional patterns, it is likely that promoter and regulatory sequences do.
4. Genomes are incredibly successful at adapting to the necessity of generating multiple expression profiles (development, tissue-specific, stress-response, environmental adaptation). This genotype to phenotype expression requires a great potential for recombinatorial diversity. Keeping pathway components separated allows for a large biological potential in expression diversity because sequential pathway reactions are not "hard-wired" (directly linked) to one another.

Of these four points, the first and second fall within the larger realm of

organismal evolution and are beyond the scope of this article. The third and fourth points represent the current frontiers of exploration in understanding eukaryotic genome organization.

The physical colocation of genes is not a necessary requirement of coordinated expression. In fact, it would appear to be the exception rather than the rule in terms of how genes are distributed across chromosomes. Given that genes are unequally distributed across chromosomes and that local transcriptional activities vary as a function of local gene densities (Sect. 4.2), a first hypothesis for a mechanism to establish parallel transcriptional controls would be for genes requiring similar expression patterns to be located in similar genetic neighborhoods (i.e. similar gene densities and transcriptional activities). To assess this potential for expression linkage, we have examined the local domains of the glycolytic enzymes (first nine enzymes in Table 1) looking at the number of sequence tags within a ±1 cR window surrounding a locus and the expression levels of those tagged transcripts. These cR measurements have been assessed using the GB-94 panel (GeneMap'99) where, 1 cR corresponds to ~280 kb of DNA. In Fig. 13, the number of SAGE tags are used as an index of the transcriptional activity surrounding each glycolytic gene locus, with the domain total equivalent to the total amount of transcription in the 2 cR window and the gene total being the

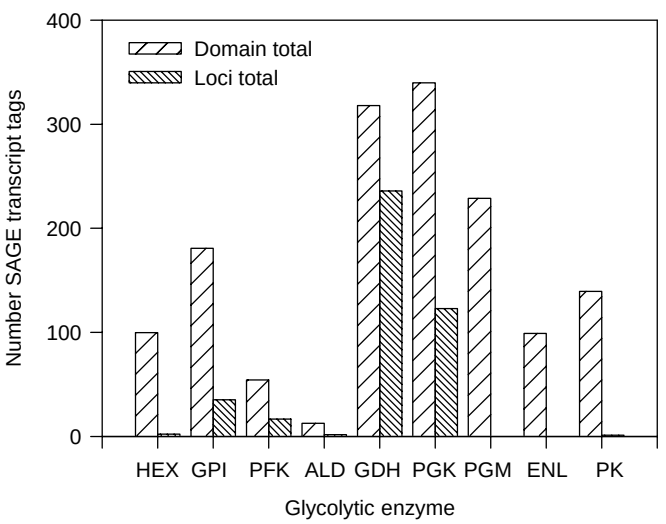

Fig. 13 Transcriptional activity in the loci domains for the glycolytic enzymes presented in Table 1. SAGE transcript tags for a ±1 cR region (GeneMap'99-GB4) around each gene loci were enumerated using the Human Transcriptome Map resources for average expression levels across 12 tissue types total. The histogram bars plot: (1) the total transcriptional activity over a 2 cR region (number of SAGE marker hits), and (2) the transcriptional activity of each gene locus (see chromosomal location data in Table 1). Despite the coordinated, biochemical activity of these genes for glycolysis, there appears to be no direct common transcriptional features regarding the chromosomal domains where these genes reside.

transcriptional activity of the individual locus. Here, there is no apparent relationship between the transcriptional domains of sequential glycolytic gene loci. The only two genes that come close to having similar domains are GDH and PGK, each with high domain totals and high individual levels of expression.

Much significance has been placed on the domain organization of genes into regions of differential expression activity; however, this appears to be a simple tautological observation, that is, that areas dense with genes show higher expression levels because they have more genes. This was one of the fundamental hypotheses of Bernardi in describing the phenomenological nature of isochore organization, which we have seen substantiated. Thus, future efforts to correlate gene expression levels across wide domains do not appear to be a fruitful avenue of research. The genetic complexity of a 500-Mb window is just too great to summarize with a single, average metric. Although it is a conceptually enticing approach in terms of design simplification, measurements of SAGE hits, %GC, GC_3 frequency, CpG islands, isochore boundaries, and so on, cannot describe the molecular mechanism underlying the dynamic expression linkages that are evident.

Intriguing new insights into the distribution of cis-regulatory elements across chromosomes is beginning to show distinctive spatial patterns that could potentially lead to an understanding of why genes are distributed the way they are. Characterizing sequence structures over small windows (~100 bp) has revealed a remarkable degree of consistent patterns in upstream regulatory sequences and exon/intron splicing boundaries. These patterns indicate a consistent spatial organization in regulatory domains that is independent of chromosomal location. Two big challenges to overcome in clearly identifying such cis-regulatory mechanisms is the difficulty of finding statistically significant sequence motifs in such short motifs and the development of quantitative models of regulatory interactions that can explain sequence structural organization.

4.5 Importance of Expression Controls

Another new frontier in assessing genomic organization is describing the mechanistic basis for the recombinatorial diversity of gene expression activities. In eukaryotes, gene promoter domains are rarely triggered by single activators or repressors. Instead, multiple interactions between several target sites within a promoter domain, and between several activating or repressing protein factors at each of those sites, results in the regulation of transcription rates. Consequently, the expression of a gene within a cell is more accurately described as a probabilistic function, rather than a deterministic one, that is, all cells within a tissue do not have the same expression profile at any instant in time. There is the potential for significant variation in expression levels, which is one of the key features of the successful organization of a genome in perpetuating a biological system.

We are now just beginning to see data quantifying levels of variance in gene expression rates. In a comprehensive study of expression patterns in yeast cells, only 7% of the statistically significant changes in gene expression levels following experimental stimuli were actually functional, that is, directly related to altering the cellular physiology of the yeast to accommodate the stimuli. The

remaining 93% of "significant" gene expression activity was essentially "random." In vertebrates, the significance of accounting for individual variance in gene expression levels has been documented in fish, and this approach is now being applied to screening human microarray expression data for disease detection by establishing a critical threshold of detection relative to differences in expression levels between different individuals.

Thirty years ago, scientists were amazed at how similar in composition the chromosomes of chimpanzees and humans appeared to be. This lead to the hypothesis that one of the primary forces in species evolution was not mutations in coding gene domains, but subtle alterations in the timing and expression levels of those genes. When it became clear that humans essentially had only three times the number of genes as in the fly *Drosophila*, we realized that coding gene domains were not the only evolutionary game in play. The large difference in biological system organization between humans and flies is likely impacted by the timing, magnitude, and coordination of gene expression events just as significantly as gene exon domains. Overall, the probabilistic (not deterministic) nature of gene expression controls will likely become one of the primary mechanisms determining how and where genes are distributed within the human genome.

See also Gene Mapping and Chromosome Evolution by Fluorescence–Activated Chromosome Sorting; Gene Mapping by Fluorescence In Situ Hybridization; Genetics, Molecular Basis of; Genomic DNA Libraries, Construction and Applications.

Bibliography

Books and Reviews

Dale, J., von Schantz, M. (2002) *From Genes to Genomes: Concepts and Applications of DNA Technology*, Wiley, Chichester.

Lander, E.S. et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Malcolm, S., Goodship, J. (Eds.) (2001) *Genotype to Phenotype*, BIOS Scientific, Oxford; Academic Press, San Diego, CA..

Marks, J. (2002) *What it means to be 98% Chimpanzee: Apes, People, and their Genes*, University of California Press, Berkeley.

Maroni, G. (2001) *Molecular and Genetic Analysis of Human Traits*, Blackwell Science, Malden, Mass.

Miesfeld, R.L. (1999) *Applied Molecular Genetics*, John Wiley, New York.

Saccone, C., Pesole, G. (2003) *Handbook of Comparative Genomics: Modern Methodology*, Wiley-Liss, New York.

Venter, J.C. et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Weissmann, G. (2002) *The Year of the Genome: A Diary of the Biological Revolution*, Times Books, New York.

Primary Literature

Adachi, N., Lieber, M.R. (2002) Bidirectional gene organization: a common architectural feature of the human genome, *Cell* **109**, 807–809.

Adams, M.D. et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project, *Science* **252**, 1651–1656.

Albert, R., Barabasi, A.L. (2000) Dynamics of complex systems: scaling laws for the period of Boolean networks, *Phys. Rev. Lett.* **84**, 5660–5663.

Albert, R. et al. (2000) Error and attack tolerance of complex networks, *Nature* **406**, 378–382.

Alon, U. et al. (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6745–6750.

Alvarez-Valin, F. et al. (2002) Isochores, GC(3) and mutation biases in the human genome, *Gene* **300**, 161–168.

Amaral, L.A.N. et al. (2000) Classes of small-world networks, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11149–11152.

Antequera, F., Bird, A. (1993) Number of CpG islands and genes in human and mouse, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11995–11999.

Antequera, F., Bird, A. (1999) CpG islands as genomic footprints of promoters that are associated with replication origins, *Curr. Biol.* **9**, 661–667.

Arkhipova, I.R. (2001) Transposable elements in the animal kingdom, *Mol. Biol.* **35**, 157–167.

Bailey, J.A. et al. (2002) Recent segmental duplications in the human genome, *Science* **297**, 1003–1007.

Barthelemy, M., Amaral, L.A.N. (1999) Small-world networks: evidence for a crossover picture, *Phys. Rev. Lett.* **82**, 5180–5180.

Bernardi, G. (1993) The isochore organization of the human genome and its evolutionary history – a review, *Gene* **135**, 57–66.

Bernardi, G. (1993) The vertebrate genome: isochores and evolution, *Mol. Biol. Evol.* **10**, 186–204.

Bernardi, G. (1995) The human genome: organization and evolutionary history, *Annu. Rev. Genet.* **29**, 445–476.

Bernardi, G. (2000) Isochores and the evolutionary genomics of vertebrates, *Gene* **241**, 3–17.

Bernardi, G. (2001) Isochores: myth or reality? *Med. Sci.* **17**, 1097.

Bernardi, G. (2001) Misunderstandings about isochores. Part 1, *Gene* **276**, 3–13.

Bernardi, G. et al. (1973) The specificity of deoxyribonucleases and their use in nucleotide sequence studies, *Nat. N. Biol.* **246**, 36–40.

Bernardi, G. et al. (1985) The mosaic genome of warm-blooded vertebrates, *Science* **228**, 953–958.

Bird, A.P. (1986) CpG-rich islands and the function of DNA methylation *Nature* **321**, 209–213.

Bolouri, H., Davidson, E.H. (2002) Modeling DNA sequence-based cis-regulatory gene networks, *Dev. Biol.* **246**, 2–13.

Bortoluzzi, S. et al. (1998) A comprehensive, high-resolution genomic transcript map of human skeletal muscle, *Genome Res.* **8**, 817–825.

Brazma, A., Vilo, J. (2000) Gene expression data analysis, *FEBS Lett.* **480**, 17–24.

Brudno, M. et al. (2001) Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing, *Nucleic Acids Res.* **29**, 2338–2348.

Burton, R.S. et al. (1999) Genetic architecture of physiological phenotypes: Empirical evidence for coadapted gene complexes, *Am. Zool.* **39**, 451–462.

Bustamante, C.D. et al. (2002) A maximum likelihood method for analyzing pseudogene evolution: Implications for silent site evolution in humans and rodents, *Mol. Biol. Evol.* **19**, 110–117.

Cargill, M. et al. (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes, *Nat. Genet.* **22**, 231–238.

Caron, H. et al. (2001) The human transcriptome map: Clustering of highly expressed genes in chromosomal domains, *Science* **291**, 1289–1292.

Clay, O., Bernardi, G. (2001) Compositional heterogeneity within and among isochores in mammalian genomes. II. Some general comments, *Gene* **276**, 25–31.

Cohen, B.A. et al. (2000) A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression, *Nat. Genet.* **26**, 183–186.

Corneo, G. et al. (1968) Isolation and characterization of mouse and guinea pig satellite deoxyribonucleic acids, *Biochemistry* **7**, 4373–4379.

Corthals, G.L. et al. (2000) The dynamic range of protein expression: a challenge for proteomic research, *Electrophoresis* **21**, 1104–1115.

Coulondre, C. et al. (1978) Molecular basis of base substitution hotspots in Escherichia coli, *Nature* **274**, 775–780.

Crawford, D.L. et al. (1999) Evolutionary physiology of closely related taxa: analyses of enzyme expression, *Am. Zool.* **39**, 389–400.

Crawford, D.L. et al. (1999) Evolutionary analysis of TATA-less proximal promoter function, *Mol. Biol. Evol.* **16**, 194–207.

Cross, S.H. et al. (2000) CpG island libraries from human chromosomes 18 and 22: landmarks for novel genes, *Mamm. Genome* **11**, 373–383.

Das, M. et al. (2001) Assessment of the total number of human transcription units, *Genomics* **77**, 71–78.

Davidson, E.H. et al. (2002) A genomic regulatory network for development, *Science* **295**, 1669–1678.

de Krom, M. et al. (2002) Stochastic patterns in globin gene expression are established prior to transcriptional activation and are clonally inherited, *Mol. Cell* **9**, 1319–1326.

de la Fuente, A. et al. (2002) Linking the genes: inferring quantitative gene networks from microarray data, *Trends Genet.* **18**, 395–398.

Deloukas, P. et al. (1998) A physical map of 30 000 human genes, *Science* **282**, 744–746.

DeRisi, J.L. et al. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale, *Science* **278**, 680–686.

D'Haeseleer, P. et al. (2000) Genetic network inference: from co-expression clustering to reverse engineering, *Bioinformatics* **16**, 707–726.

Dolstra, H. et al. (2002) Bi-directional allelic recognition of the human minor histocompatibility antigen HB-1 by cytotoxic T lymphocytes, *Eur. J. Immunol.* **32**, 2748–2758.

D'Onofrio, G. (2002) Expression patterns and gene distribution in the human genome, *Gene* **300**, 155–160.

D'Onofrio, G. et al. (1999) Evolutionary genomics of vertebrates and its implications, *Ann. N. Y. Acad. Sci.* **870**, 81–94.

Drouin, R. et al. (1994) High-resolution replication bands compared with morphologic G- and R-bands, *Adv. Hum. Genet.* **22**, 47–115.

Duret, L. et al. (1995) Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores, *J. Mol. Evol.* **40**, 308–317.

Dutrillaux, B. (1977) A theoretical model of structural chromosome rearrangement induction, *Ann. Genet.* **20**, 221–226.

Dutrillaux, B., Covic, M. (1974) Factors influencing the heat denaturation of chromosomes, *Exp. Cell Res.* **85**, 143–153.

Echols, N. et al. (2002) Comprehensive analysis of amino acid and nucleotide composition in eukaryotic genomes, comparing genes and pseudogenes, *Nucleic Acids Res.* **30**, 2515–2523.

Eisen, M.B. et al. (1998) Cluster analysis and display of genome-wide expression patterns, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863–14868.

Eisen, M.B. et al. (1999) Cluster analysis and display of genome-wide expression patterns, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10943.

Epstein, C., Butow, R. (2000) Microarray technology-enhanced versatility, persistent challenge, *Curr. Opin. Biotechnol.* **11**, 36–41.

Ewing, B., Green, P. (2000) Analysis of expressed sequence tags indicates 35 000 human genes, *Nat. Genet.* **25**, 232–234.

Fields, C. et al. (1994) How many genes in the human genome? *Nat. Genet.* **7**, 345–346.

Finnegan, E.J. (2002) Epialleles – a source of random variation in times of stress, *Curr. Opin. Plant Biol.* **5**, 101–106.

Forst, C. (2002) Network genomics – a novel approach for the analysis of biological systems in the post-genomic era, *Mol. Biol. Rep.* **29**, 265–280.

Forst, C.V., Schulten, K. (2001) Phylogenetic analysis of metabolic pathways, *J. Mol. Evol.* **52**, 471–489.

Friedman, N. et al. (2000) Using Bayesian networks to analyze expression data, *J. Comput. Biol.* **7**, 601–620.

Frith, M.C. et al. (2002) Statistical significance of clusters of motifs represented by position specific scoring matrices in nucleotide sequences, *Nucleic Acids Res.* **30**, 3214–3224.

Fujiyama, A. et al. (2002) Construction and analysis of a human-chimpanzee comparative clone map, *Science* **295**, 131–134.

Gallo, A. et al. (2002) Micro-processing events in mRNAs identified by DHPLC analysis, *Nucleic Acids Res.* **30**, 3945–3953.

Gardiner-Garden, M., Frommer, M. (1987) CpG islands in vertebrate genomes, *J. Mol. Biol.* **196**, 261–282.

Gasch, A.P. et al. (2000) Genomic expression programs in the response of yeast cells to environmental changes, *Mol. Biol. Cell* **11**, 4241–4257.

Ge, H. et al. (2001) Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae, *Nat. Genet.* **29**, 482–486.

Gerstein, M., Jansen, R. (2000) The current excitement in bioinformatics, analysis of whole-genome expression data: how does it relate to protein structure and function, *Curr. Opin. Struct. Biol.* **10**, 574–584.

Giaever, G., et al. (2002) Functional profiling of the Saccharomyces cerevisiae genome, *Nature* **418**, 387–391.

Gibson, G. (2002) Developmental evolution: getting robust about robustness, *Curr. Biol.* **12**, R347–R349.

Gibson, G. (2002) Microarrays in ecology and evolution: a preview, *Mol. Ecol.* **11**, 17–24.

Girvan, M., Newman, M.E.J. (2002) Community structure in social and biological networks, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821–7826.

Goncalves, I. et al. (2000) Nature and structure of human genes that generate retropseudogenes, *Genome Res.* **10**, 672–678.

Greenbaum, D. et al. (2002) Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts, *Bioinformatics* **18**, 585–596.

Grunau, C. et al. (2000) Large-scale methylation analysis of human genomic DNA reveals tissue-specific differences between the methylation profiles of genes and pseudogenes, *Hum. Mol. Genet.* **9**, 2651–2663.

Halushka, M.K. et al. (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis, *Nat. Genet.* **22**, 239–247.

Harrison, P.M., Gerstein, M. (2002) Studying genomes through the aeons: protein families, pseudogenes and proteome evolution, *J. Mol. Biol.* **318**, 1155–1174.

Harrison, P.M. et al. (2002) Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22, *Genome Res.* **12**, 272–280.

Hartemink, A.J. et al. (2001) Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks, *Pac. Symp. Biocomput.* 422–433.

Hatzimanikatis, V. et al. (1999) Proteomics: theoretical and experimental considerations, *Biotechnol. Prog.* **15**, 312–318.

Hochachka, P.W. et al. (1998) Integrating metabolic pathway fluxes with gene-to-enzyme expression rates, *Comp. Biochem. Physiol.* **120B**, 17–26.

Holmquist, G.P. (1992) Chromosome bands, their chromatin flavors, and their functional features, *Am. J. Hum. Genet.* **51**, 17–37.

Holter, N.S. et al. (2000) Fundamental patterns underlying gene expression profiles: simplicity from complexity, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8409–8414.

Huang, S. (1999) Gene expression profiling, genetic networks, and cellular states: an integrating concept for tumorigenesis and drug discovery, *J. Mol. Med.* **77**, 469–480.

Hubbard, T., Birney, E. (2000) Open annotation offers a democratic solution to genome sequencing, *Nature* **403**, 825.

Hurst, L.D., Eyre-Walker, A. (2000) Evolutionary genomics: reading the bands, *BioEssays* **22**, 105–107.

Ideker, T. et al. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network, *Science* **292**, 929–934.

Ioshikhes, I.P., Zhang, M.Q. (2000) Large-scale human promoter mapping using CpG islands, *Nat. Genet.* **26**, 61–63.

Jablonka, E. et al. (1998) 'Lamarckian' mechanisms in Darwinian evolution, *Trends Ecol. Evol.* **13**, 206–210.

Jackson, R.B. et al. (2002) Linking molecular insight and ecological research, *Trends Ecol. Evol.* **17**, 409–414.

Jacob, F., Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins, *J. Mol. Biol.* **3**, 318–356.

Jones, P.A., Takai, D. (2001) The role of DNA methylation in mammalian epigenetics, *Science* **293**, 1068–1070.

Kerr, M.K., Churchill, G.A. (2001) Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8961–8965.

Kholodenko, B.N. et al. (2002) Untangling the wires: a strategy to trace functional interactions in signaling and gene networks, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12841; **99**, 15245.

Kidd, D. (2001) Profiling serine hydrolase activities in complex proteomes, *Biochemistry* **40**, 4005–4015.

Kidwell, M.G., Lisch, D.R. (2001) Perspective: transposable elements, parasitic DNA, and genome evolution, *Evolution* **55**, 1–24.

Kikukawa, Y. et al. (2002) The 26S proteasome Rpn10 gene encoding splicing isoforms: Evolutional conservation of the genomic organization in vertebrates, *Biol. Chem.* **383**, 1257–1261.

King, M.C., Wilson, A.C. (1975) Evolution at two levels: molecular similarities and biological differences between humans and chimpanzees, *Science* **188**, 107–116.

Kohler, C., Grossniklaus, U. (2002) Epigenetics: the flowers that come in from the cold, *Curr. Biol.* **12**, R129–R131.

Kulp, D. et al. (1996) A generalized hidden Markov model for the recognition of human

genes in DNA, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **4**, 134–142.

Lander, E.S. et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Larsen, F., et al. (1992) CpG islands as gene markers in the human genome, *Genomics* **13**, 1095–1107.

Leibovitch, B.A. (2002) Chromatin structure, heterochromatin, and transposable genetic elements: Are these teammates? *Mol. Biol.* **36**, 189–195.

Lercher, M.J. et al. (2002) Clustering of housekeeping genes provides a unified model of gene order in the human genome, *Nat. Genet.* **31**, 180–183.

Li, W. (2001) Delineating relative homogeneous G + C domains in DNA sequences, *Gene* **276**, 57–72.

Liang, F. et al. (2000) Gene index analysis of the human genome estimates approximately 120 000 genes, *Nat. Genet.* **25**, 239, 240.

Long, M., Langely, C.H. (1993) Natural selection and the origin of *jingwei*, a chimeric processed functional gene in *Drosophila*, *Science* **260**, 91–95.

Lukashin, A.V., Fuchs, R. (2001) Analysis of temporal gene expression profiles: clustering by simulated annealing and determining the optimal number of clusters, *Bioinformatics* **17**, 405–414.

MacKenzie, S. et al. (2002) Post-transcriptional regulation of TNF-alpha during in vitro differentiation of human monocytes/macrophages in primary culture, *J. Leukocyte Biol.* **71**, 1026–1032.

Mahy, N.L. et al. (2002) Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH, *J. Cell Biol.* **159**, 753–763.

Mahy, N.L. et al. (2002) Spatial organization of active and inactive genes and noncoding DNA within chromosome territories, *J. Cell Biol.* **157**, 579–589.

Majewski, J., Ott, J. (2002) Distribution and characterization of regulatory elements in the human genome, *Genome Res.* **12**, 1827–1836.

Makalowski, W. (2000) Genomic scrap yard: how genomes utilize all that junk, *Gene* **259**, 61–67.

Marcotte, E.M. (2001) Measuring the dynamics of the proteome, *Genome Res.* **11**, 191–193.

Marsh, A.G. et al. (2000) Gene expression and enzyme activities of the sodium pump during sea urchin development: implications for indices of physiological state, *Biol. Bull.* **199**, 100–107.

Mattick, J.S., Gagen, M.J. (2001) The evolution of controlled multitasked gene networks: The role of introns and other noncoding RNAs in the development of complex organisms, *Mol. Biol. Evol.* **18**, 1611–1630.

McAdams, H.H., Arkin, A. (1999) It's a noisy business! Genetic regulation at the nanomolar scale, *Trends Genet.* **15**, 65–69.

McCullough, A.J., Berget, S.M. (1997) G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection, *Mol. Cell. Biol.* **17**, 4562–4571.

Miklos, G.L., John, B. (1979) Heterochromatin and satellite DNA in man: properties and prospects, *Am. J. Hum. Genet.* **31**, 264–280.

Mira, A., Ochman, H. (2002) Gene location and bacterial sequence divergence, *Mol. Biol. Evol.* **19**, 1350–1358.

Mouchiroud, D. et al. (1991) The distribution of genes in the human genome, *Gene* **100**, 181–187.

Musio, A. et al. (2002) Heterogeneous gene distribution reflects human genome complexity as detected at the cytogenetic level, *Can. Genet. Cytogenet.* **134**, 168–171.

Nakao, M. (2001) Epigenetics: interaction of DNA methylation and chromatin, *Gene* **278**, 25–31.

Newman, S.A., Muller, G.B. (2000) Epigenetic mechanisms of character origination, *J. Exp. Zool.* **288**, 304–317.

Nijhout, H.F. (2002) The nature of robustness in development, *BioEssays* **24**, 553–563.

Oleksiak, M.F. et al. (2001) Utility of natural populations for microarray analyses: isolation of genes necessary for functional genomic studies, *Marine Biotech.* **3**, S203–S211.

Oleksiak, M.F. et al. (2002) Variation in gene expression within and among natural populations, *Nat. Genet.* **32**, 261–266.

Oliver, J.L. et al. (2002) Isochore chromosome maps of the human genome, *Gene* **300**, 117–127.

Ozbudak, E.M. et al. (2002) Regulation of noise in the expression of a single gene, *Nat. Genet.* **31**, 69–73.

Pal, C., Miklos, I. (1999) Epigenetic inheritance, genetic assimilation and speciation, *J. Theor. Biol.* **200**, 19–37.

Papin, J.A. et al. (2002) The genome-scale metabolic extreme pathway structure in

Haemophilus influenzae shows significant network redundancy, *J. Theor. Biol.* **215**, 67–82.

Pavlicek, A. et al. (2002) A compact view of isochores in the draft human genome sequence, *FEBS Lett.* **511**, 165–169.

Pe'er, D. et al. (2001) Inferring subnetworks from perturbed expression profiles, *Bioinformatics* **17**, S215–S224.

Pesole, G., et al. (1999) Isochore specificity of AUG initiator context of human genes, *FEBS Lett.* **464**, 60–62.

Petrov, D.A., Hartl, D.L. (2000) Pseudogene evolution and natural selection for a compact genome, *J. Hered.* **91**, 221–227.

Pierce, V.A., Crawford, D.L. (1997) Phylogenetic analysis of thermal acclimation of the glycolytic enzymes in the genus Fundulus, *Physiol. Zool.* **70**, 597–609.

Plass, C. (2002) Cancer epigenomics, *Hum. Mol. Genet.* **11**, 2479–2488.

Ponger, L. et al. (2001) Determinants of CpG islands: expression in early embryo and isochore structure, *Genome Res.* **11**, 1854–1860.

Price, N.D. et al. (2002) Determination of redundancy and systems properties of the metabolic network of Helicobacter pylori using genome-scale extreme pathway analysis, *Genome Res.* **12**, 760–769.

Pruitt, K.D., Maglott, D.R. (2001) RefSeq and LocusLink: NCBI gene-centered resources, *Nucleic Acids Res.* **29**, 137–140.

Roest Crollius, H. et al. (2000) Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence, *Nat. Genet.* **25**, 235–238.

Roy, P.J. et al. (2002) Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans, *Nature* **418**, 975–979.

Saccone, S. et al. (1992) The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4913–4917.

Saccone, S. et al. (1996) Identification of the gene-richest bands in human chromosomes, *Gene* **174**, 85–94.

Saccone, S. et al. (1999) Identification of the gene-richest bands in human prometaphase chromosomes, *Chromosome Res.* **7**, 379–386.

Saccone, C. et al. (2002) Molecular strategies in Metazoan genomic evolution, *Gene* **300**, 195–201.

Schlosser, G. (2002) Modularity and the units of evolution, *Theory Biosci.* **121**, 1–80.

Schuler, G.D. et al. (1996) A gene map of the human genome, *Science* **274**, 540–546.

Shapiro, J.A. (2002) A 21(st) century view of evolution, *J. Biol. Phys.* **28**, 745–764.

Sherlock, G. (2000) Analysis of large-scale gene expression data, *Curr. Opin. Immun.* **12**, 201–205.

Smolke, C.D., Keasling, J.D. (2002) Effect of gene location, mRNA secondary structures, and RNase sites on expression of two genes in an engineered operon, *Biotechnol. Bioeng.* **80**, 762–776.

Somogyi, R., et al. (1997) The gene expression matrix: Towards the extraction of genetic network architectures, *Nonlinear Anal. Theory Methods Appl.* **30**, 1815–1824.

Spellman, P.T., Rubin, G. (2002) Evidence for large domains of similarly expressed genes in the *Drosophila* genome, *J. Biol.* **1**, 5.(article)

Strehl, S. et al. (1997) High-resolution analysis of DNA replication domain organization across an R/G-band boundary, *Mol. Cell Biol.* **17**, 6157–6166.

Sumner, A.T. (1982) The nature and mechanisms of chromosome banding, *Can. Genet. Cytogenet.* **6**, 59–87.

Thieffry, D., Romero, D. (1999) The modularity of biological regulatory networks, *Biosystems* **50**, 49–59.

Toh, H., Horimoto, K. (2002) Inference of a genetic network by a combined approach of cluster analysis and graphical Gaussian modeling, *Bioinformatics* **18**, 287–297.

Venter, J.C. et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Wagner, A. (2002) Estimating coarse gene network structure from large-scale gene perturbation data, *Genome Res.* **12**, 309–315.

Wagner, A., Fell, D.A. (2001) The small world inside large metabolic networks, *Proc. R. Soc. London, Ser. B-Biol. Sci.* **268**, 1803–1810.

Waterston, R. et al. (1992) A survey of expressed genes in Caenorhabditis elegans, *Nat. Genet.* **1**, 114–123.

Watts, D.J., Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks, *Nature* **393**, 440–442.

Wessels, L.F. et al. (2001) A comparison of genetic network models, *Pac. Symp. Biocomput.* 508–519.

Williams, E.J.B., Hurst, L.D. (2002) Clustering of tissue-specific genes underlies much of the

similarity in rates of protein evolution of linked genes, *J. Mol. Evol.* **54**, 511–518.

Woischnik, M., Moraes, C.T. (2002) Pattern of organization of human mitochondrial pseudogenes in the nuclear genome, *Genome Res.* **12**, 885–893.

Wolf, Y.I. et al. (2002) Scale-free networks in biology: new insights into the fundamentals of evolution? *BioEssays* **24**, 105–109.

Wray, G.A. (2002) Evolution of the gene network underlying wing polyphenism in ants, *Science* **297**, 249–252.

Yang, T.H., Somero, G.N. (1996) Activity of lactate dehydrogenase but not its concentration of messenger RNA increases with body size in barred sand bass, Paralabrax nebulifer (Teleostei), *Biol. Bull.* **191**, 155–158.

Yang, T.H., Somero, G.N. (1996) Fasting reduces protein and messenger RNA concentrations for lactate dehydrogenase but not for actin in white muscle of scorpion fish (Scorpaena guttata, Teleostei), *Mol. Marine Biol. Biotechnol.* **5**, 153–161.

Yeung, K.Y. et al. (2001) Validating clustering for gene expression data, *Bioinformatics* **17**, 309–318.

Zhuo, D. et al. (2001) Assembly, annotation, and integration of UNIGENE clusters into the human genome draft, *Genome Res.* **11**, 904–918.

Zoubak, S. et al. (1996) The gene distribution of the human genome, *Gene* **174**, 95–102.

Gene Expression and Replication of Picornaviruses

Christopher T. Cornell and Bert L. Semler
Department of Microbiology and Molecular Genetics, College of Medicine, University of California, Irvine, CA, USA

1 Introduction 95
1.1 Taxonomy of Picornaviruses 95
1.2 Disease Outcomes 95
1.3 Genome Organization 96

2 Virion Properties, Host Cell Binding, Entry, and Uncoating 97
2.1 Features of the Picornavirus Virion 97
2.2 Receptors and Virus Adsorption to the Host Cell 98
2.3 Virus Uncoating 99

3 Synthesis and Processing of Viral Polyprotein 99
3.1 Internal Ribosome Entry 99
3.2 Host Proteins Involved in IRES-mediated Translation 100
3.3 Overview of Viral Protein Functions 101
3.4 Viral-encoded Proteinases 101

4 Viral RNA Replication and Packaging 103
4.1 Modification of the Host Cell Environment 103
4.2 Viral Protein 2A 104
4.3 Viral Protein 2B 104
4.4 Viral Proteins 2C and 2BC 105
4.5 Viral Protein 3A 105
4.6 Viral Proteins 3B and 3AB 105
4.7 Viral Proteins 3C and 3CD 106
4.8 The Viral RNA-dependent RNA Polymerase 3D 107
4.9 VPg Uridylylation: Priming RNA Synthesis 107
4.10 Cis-acting Replication Elements 108

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

4.11 A Model for Picornavirus RNA Replication 109
4.12 RNA Encapsidation and Cell Lysis 112

Bibliography 114
Books and Reviews 114
Primary Literature 114

Keywords

Internal Ribosome Entry Site (IRES)
A region of RNA containing sequences and secondary structures that facilitate the cap-independent recruitment of ribosomes and initiation factors for translation of a downstream gene.

Noncoding Region (NCR)
A sequence of RNA that is not translated into protein.

RNA-dependent RNA Polymerase
An enzyme capable of elongating ribonucleic acid chains from an RNA template.

Proteinase
An enzyme that recognizes and cleaves specific peptide sequences within a protein or a subset of proteins.

Cap-independent Translation
A mechanism of translation that does not rely on the recognition of a 7-methyl-guanosine (m^7G) cap structure linked to the 5′ end of an mRNA by translation initiation factors.

Polyprotein
A large polypeptide containing the amino acid sequences of more than one mature gene product.

Picornaviruses are the agents responsible for numerous diseases of varying severities that affect the population of the world whose susceptible hosts include humans and other vertebrates. These viruses are presently classified into seven major genera and are responsible for diseases like paralytic poliomyelitis (poliovirus), myocarditis (coxsackievirus), the common cold (rhinovirus), hepatitis (hepatitis A virus), and foot-and-mouth disease (foot-and-mouth disease virus). Despite the ability of these viruses to cause a wide array of maladies, they all share common features among their replication and gene expression strategies, often allowing insights from studies on one virus to be applied to another. With the advent of powerful molecular biology techniques, picornaviruses have provided a wealth of information surrounding protein–protein and protein–RNA interactions that are critical for the process of viral replication and subversion of the host cell. Importantly, this has also furthered our understanding of eukaryotic gene expression and the regulation of other host cell processes.

1 Introduction

1.1 Taxonomy of Picornaviruses

The family *Picornaviridae* (*pico* means "small," i.e. small RNA virus) consists of seven major genera: enteroviruses, rhinoviruses, cardioviruses, aphthoviruses, hepatoviruses, parechoviruses, and teschoviruses. Parechoviruses and teschoviruses have been described only recently, as have two less-populated genera (kobuviruses and erboviruses). Table 1 shows a summary of these genera and representative virus(es) from each. These classifications are based primarily on virion properties (pH stability, sedimentation coefficients, buoyant densities), sequence homology, and differences in genome organization (discussed later in this article).

1.2 Disease Outcomes

Picornaviruses are responsible for significant diseases in a number of different hosts. The prototypic picornavirus, poliovirus (PV), causes poliomyelitis, a once feared disease that can cause a crippling form of paralysis. Interestingly, poliomyelitis is one of the earliest viral diseases ever to be documented, its clinical manifestations depicted on an Egyptian stele dating back to ~1400 B.C. Though largely considered a historic epidemic, there are regions of the world (primarily portions of Africa, India, and China) not yet certified "polio free" that have experienced outbreaks of paralytic poliovirus infection from unknown sources as recently as the year 2000. Much attention has also been focused on foot-and-mouth disease virus (FMDV), primarily due to a major outbreak in the United Kingdom and other areas of Europe in 2001 that resulted in the loss of substantial numbers of sheep and cattle, carrying with it dire economic consequences. Human rhinovirus (HRV) is responsible for causing the common cold, the single most prevalent reason for loss of workplace hours due to employee illness. Infection with coxsackievirus (CV) A16 can result in "hand, foot-and-mouth

Tab. 1 Genera of picornaviridae and representative species.

Genus	*Species*
Enteroviruses	Poliovirus (PV)
	Coxsackievirus (CV)
	Porcine enterovirus (PEV)
Rhinoviruses	Human rhinovirus (HRV)
	Bovine rhinovirus (BRV)
Cardioviruses	Mengovirus
	Encephalomyocarditis virus (EMCV)
	Theiler's murine encephalomyocarditis virus (TMEV)
Aphthoviruses	Foot-and-mouth disease virus (FMDV)
Hepatoviruses	Hepatitis A virus (HAV)
Parechoviruses	Human parechovirus (HPev)
Teschoviruses	Porcine teschnovirus (PTV)
Kobuviruses	Aichi virus (AiV)
Erboviruses	Equine rhinitis B virus (ERBV)

disease'' primarily in young children, whereas other strains of coxsackievirus are responsible for more rare symptoms including myocarditis, ocular conjunctivitis, and even paralysis. Parechoviruses (originally called *echo*viruses, for *e*nteric *c*ytopathic *h*uman *o*rphan viruses) are known to cause chronic meningoencephalitis, colds, neonatal carditis, and encephalitis. The cardioviruses are primarily mouse pathogens known to cause encephalomyelitis and demyelinating disorders, although there are rare instances of human cardioviral infections. And finally, hepatitis A virus (HAV), a hepatovirus, is a leading cause of liver disease that varies from isolated cases to widespread epidemics.

1.3 Genome Organization

Schematic representations of the genomes of several picornaviruses are depicted in Fig. 1. A typical picornavirus genomic RNA is ~7.0 kb to ~8.5 kb in length, single-stranded, positive-sense, and contains a 3′ poly(A) tract. Instead of a 7-methyl-guanosine cap normally found linked to the 5′ end of cellular mRNAs, picornavirus genomes possess a small viral protein, termed *VPg* (the 3B-gene product), covalently attached to their 5′ ends. VPg is uridylylated by the viral RNA polymerase to produce VPg-pU-pU, which acts as a protein primer for the initiation of viral RNA synthesis (discussed in Sect. 4). The genome of a typical picornavirus contains a large polyprotein coding region flanked by 5′ and 3′ noncoding regions (NCRs), which harbor several RNA sequence and structure elements necessary for RNA replication and translation. The 5′ noncoding region contains an IRES or internal ribosome entry site (discussed in Sect. 3), which allows the virus to initiate translation of its viral polyprotein via a cap-independent mechanism. Picornaviruses contain coding regions within their genomes that consist of a single open reading frame used to translate a large viral polyprotein of approximately 250 kDa, which is processed into both precursor and mature cleavage products by different viral-encoded proteinases. All picornaviruses (except for

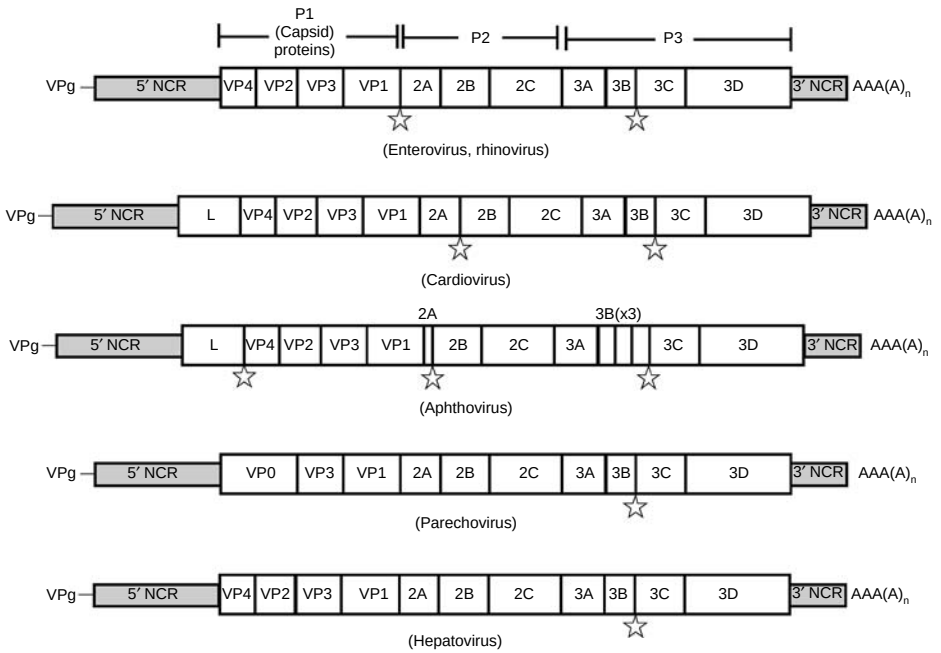

Fig. 1 Schematic of genomes from major genera of *Picornaviridae*. Shown are the genome organizations for entero-/rhinoviruses, cardioviruses, aphthoviruses, parechoviruses, and hepatoviruses. Above the top schematic are the regions of the polyprotein designated P1 (capsid proteins) and P2/P3 (nonstructural proteins). Covalently linked to the 5′ end of each genome is the small viral protein VPg (the 3B gene product) and shown in gray are the 5′ and 3′ noncoding regions, which contain extensive secondary structures (not depicted in the figure). The star(s) shown below each schematic indicate sites of primary cleavage within the viral polyprotein (occurring in cis) that are mediated by viral-encoded activities.

parechoviruses; see Fig. 1) generate four capsid proteins (VP1, VP2, VP3, and VP4), between one and three active proteinases, and six to eight replication proteins. The exact role(s) of the 3′ NCR as it relates to picornavirus replication is not completely understood, but it may contain determinants necessary for optimal RNA replication efficiency in various target tissue types. The 3′ poly(A) tract, absolutely required for viral infectivity, is thought to impart stability to the genomic RNA and is a likely site for replicase assembly involving both viral and host factors. It has recently been proposed that there is molecular "cross-talk" between the 3′ NCR, 3′ poly(A) tract, and 5′ NCR involving factors of viral and host origin that mediate RNA replication initiation (discussed in Sect. 4).

2 Virion Properties, Host Cell Binding, Entry, and Uncoating

2.1 Features of the Picornavirus Virion

The typical picornavirus virion is a nonenveloped, icosahedral structure derived from the assembly of 60 copies each of the capsid proteins VP1, VP2, VP3, and VP4.

Although the sequences of VP1, VP2, and VP3 can vary greatly from one picornavirus to another, these proteins always form an eight-stranded β-barrel structure that resembles a "jelly roll." Each virion contains a single copy of the RNA genome, and is void of any other viral proteins except the genome-linked protein VPg. Most picornavirus particles are resistant to harsh environmental conditions and cannot be inactivated with nonionic detergents (due to the absence of an envelope). Some are resistant to low pH. Such stability under low pH conditions allows enteroviruses, for example, to be resistant to degradation along the route of infection within the gastrointestinal tract.

2.2 Receptors and Virus Adsorption to the Host Cell

Picornaviruses utilize different molecules on the surfaces of target cells for adsorption and subsequent entry (see Table 2). These receptors include proteins, glycolipids, and carbohydrates, and are normally found only on a subset of tissues within certain organisms, resulting in stringent tissue specificity and host range. A majority of the receptors recognized by picornaviruses are members of the immunoglobulin-like superfamily. Three-dimensional studies of several picornaviruses have revealed the presence of a "canyon" that forms around the vertex of the fivefold axis of the virion particle and involves amino acids that are highly conserved between different viruses. For some rhinoviruses and enteroviruses, extensive structural studies of virus–receptor interactions have shown the canyon to be the site for binding to the immunoglobulin-like superfamily receptor molecule. This unique property differs from that of other picornaviruses, and viruses in general, that normally rely on molecules extending from the surface of the virion to participate in these interactions. The canyon morphology found in enteroviruses and rhinoviruses has been proposed to be immune-evasive by reducing the ability of soluble antibodies to recognize this conserved region of the capsid. Recently, a "pocket factor" (likely a fatty acid derived from the host) has been visualized in the canyon of rhinoviruses and enteroviruses,

Tab. 2 Cell surface receptors and coreceptors used by some picornaviruses.

Virus	*Receptor*
Poliovirus	Poliovirus receptor (PVR or CD 155)
Human rhinovirus	Intracellular adhesion molecule-1 (ICAM-1)
	Low density lipoprotein receptor (LDL-R)
Encephalomyocarditis virus	Vascular cell adhesion molecule-1 (VCAM-1)
	Sialic acid
Foot-and-mouth disease virus	Vitronectin receptor ($\alpha_v\beta_3$)
	Heparan sulfate
Coxsackievirus	Decay accelerating factor (CD155)
	ICAM-1
	Vitronectin receptor ($\alpha_v\beta_3$)
	Coxsackievirus-adenovirus receptor (CAR)
Parechovirus	Very late antigen-2 (VLA-2; $\alpha_2\beta_1$ integrin)
Hepatitis A virus	Hepatitis A virus cellular receptor-1 (HAVcr-1)

and it has been proposed that this factor somehow stabilizes the virus particle until it is replaced with the appropriate host receptor. Some antiviral compounds stably bind these pocket regions of the capsid, thereby preventing binding to the host cell. In contrast, it has recently been shown that viruses that do not bind immunoglobulin-like superfamily receptors utilize regions of their capsids not within, but just adjacent to this canyon, often involving loop structures on the surface of the virion.

2.3 Virus Uncoating

Independent of the cell surface molecule used, binding to the viral receptor triggers events that cause entry of the virus into the host cell and the subsequent uncoating of the virion to release the RNA genome into the cytoplasm. In many cases, the receptor need not be membrane-associated for this phenomenon to occur. For many picornaviruses, it has been shown that receptor-binding results in the loss of VP4 protein and the extrusion of VP1 to the virion surface. This "metastable" form of the virion particle may undergo further conformational changes that allow it to insert its RNA genome directly into the cytoplasm through the formation of a pore in the cellular membrane. Interestingly, studies have shown that adsorbed virus eluted from the cellular receptor after undergoing these conformation changes is no longer infectious.

Other picornaviruses, like rhinoviruses and foot-and-mouth disease virus, are known to enter the cell by receptor-mediated endocytosis. An acidification within the endosomal vesicle results in the disassembly of capsid proteins and release of the RNA into the cytoplasm. It is not clear if the uncoating of enterovirus particles (like poliovirus) is actually pH-dependent, since many enteroviruses demonstrate stability even at low pH.

3 Synthesis and Processing of Viral Polyprotein

3.1 Internal Ribosome Entry

Once inside the host cell, the input RNA genome is utilized as a template for the synthesis of structural (capsid) and nonstructural viral proteins. As previously mentioned, picornaviruses employ a cap-independent mechanism of translation initiation, utilizing several secondary structures and sequence elements contained with their 5′ NCRs to mediate a process known as *internal ribosome entry*. The first IRESs were described for poliovirus and encephalomyocarditis virus (EMCV, a cardiovirus) in 1988. Subsequently, it was shown that circular RNAs containing the EMCV 5′ NCR were capable of translation, confirming the fact that internal ribosome entry is an end-independent mechanism. Deletion analyses of several picornavirus IRES elements have revealed the presence of minimal "core" sequences necessary for initiation of translation with additional sequences and structures required for optimal translation activity.

The translation of capped cellular messenger RNAs is shut down early in a picornavirus infection (reviewed in Sect. 4), and translation factors that normally mediate host translation are utilized almost exclusively for the synthesis of viral proteins. The 5′ NCRs of all picornaviruses contain multiple AUG codons that are not utilized as authentic start sites for translation initiation, and the high degree of

secondary (and likely tertiary) structures found within this region of the genome prevents the ribosome from scanning to find the appropriate initiator codon. Therefore, picornaviruses use the combination of host-initiation factors and IRES elements to facilitate recruitment of the 43S ternary complex downstream of the nonauthentic AUG codons. Once there, ribosome scanning is thought to occur as it does in cap-dependent translation.

Primarily, on the basis of structural differences, two major types of IRES elements have been described for picornaviruses. For example, entero- and rhinoviruses possess type I IRES elements, whereas aphtho- and cardioviruses have a type II IRES. Initially, IRES structure was based on predictions by computer-generated folding and structure probing. More recently, visualization of IRES tertiary structure by transmission electron microscopy has been done with some picornaviruses (i.e. human rhinovirus, foot-and-mouth disease virus, hepatitis A virus) and has confirmed these initial predictions. Interestingly, IRES elements are not unique to picornaviruses, as other viruses (e.g. hepatitis C virus (a flavivirus), human herpesvirus (a DNA virus), and several retroviruses) utilize these elements for translation initiation. Furthermore, numerous eukaryotic mRNAs also possess IRESs and are capable of translation even during the early stages of a picornavirus infection (e.g. c-myc, fibroblast growth factor, Bip).

3.2
Host Proteins Involved in IRES-mediated Translation

All picornavirus internal ribosome entry sites investigated to date utilize several canonical host translation factors that recognize specific sequences and structures within the 5′ NCR, as well as other RNA binding factors of cellular origin. Evidence for this initially came from the fact that the picornavirus genome is translated immediately upon entering the cytoplasm of the host cell in the complete absence of any proteins of viral origin. Furthermore, several picornaviruses display cell-type and tissue-specific differences in IRES-mediated translation efficiencies, suggesting that these host factors play a critical role in virus tropism and tissue specificity within the host.

All picornavirus IRES elements utilize eukaryotic initiation factors to mediate translation. These include protein components of the eIF2 and eIF3 complexes, as well as some eIF4F (cap-binding complex) factors like eIF4A and 4B. Interestingly, several picornaviruses utilize a cleavage product of eIF4G generated from proteolysis mediated by the 2A or L proteinase (see Sect. 3.4), which simultaneously shuts down cap-dependent host protein synthesis. Consisting of the carboxy-terminal two-thirds of the protein, the cleavage product of eIF4G contains both the binding site for eIF3 and eIF4A, and can mediate IRES-driven translation initiation.

In addition to canonical eukaryotic translation factors, picornaviruses utilize several other host proteins for translation of their viral RNAs. Two proteins, La (lupus autoantigen) and PTB (polypyrimidine-tract binding protein), were the first described as having such a role. PTB has been shown to bind to multiple sites within the IRES of polio-, encephalomyocarditis- and foot-and-mouth disease virus, and different homologs of this protein have been shown to contribute to the ability of certain viruses (i.e. Theiler's murine encephalomyocarditis virus or TMEV) to replicate in particular cell types.

Furthermore, La is another instance of a host protein that retains activity in viral translation despite being cleaved by a viral proteinase, as is the case during a poliovirus infection. It has been proposed that proteins like La and PTB somehow stimulate the recruitment and assembly of translation initiation complexes by interacting with both the viral RNA and with various canonical host translation factors. In addition, other host proteins like PCBP (poly(rC)-binding protein) and unr (upstream of N-ras) have been shown to interact and stimulate translation of several picornavirus IRES elements.

3.3 Overview of Viral Protein Functions

Once translation begins, the viral polyprotein is synthesized. It is then cleaved by viral-encoded proteinases into several structural (capsid proteins encoded in the P1 region of the genome) and nonstructural (P2 and P3) proteins. Additionally, aphthoviruses and cardioviruses contain a leader protein (L) at the very amino-terminus of the polyprotein. This protein functions as a proteinase during aphthovirus infections; however, it does not function as a proteinase during cardioviral replication and its precise function is not yet known. The nonstructural proteins carry out several functions in the infected cell necessary to dramatically alter the intracellular environment to make it suitable for subsequent rounds of translation and RNA replication. This involves the use of not only mature viral proteins but several precursors to these proteins as well. Major functions of these proteins include a rearrangement of host membranes (2C and/or 2BC), changing overall membrane permeability (2B), altering nuclear trafficking patterns and Class I Major Histocompatibility Complex expression (3A), and the shutoff of host translation (i.e. L for aphthoviruses, 2A for enteroviruses, and rhinoviruses) and transcription (3C/3CD). Other P3-region proteins are responsible for the priming of RNA synthesis (VPg, the viral protein 3B), RNA elongation (3D), association of replication complexes with membranous vesicles (3AB), and the assembly of ribonucleoprotein complexes required for RNA synthesis initiation (i.e. 2BC, 3CD, and 3AB). The specific functions of several picornavirus nonstructural proteins will be described in greater detail in Sect. 4 of this article.

3.4 Viral-encoded Proteinases

All picornavirus genomes encode a single, large polyprotein, which is never actually observed during an infection. First suggested more than 35 years ago by Donald Summers and Jacob Maizel at the Albert Einstein College of Medicine, it is now known that viral proteinases are responsible for the co- and posttranslational processing of the viral polyprotein into its component structural and nonstructural proteins. This processing cascade is initiated by a primary cleavage event (occurring in cis i.e. monomolecular) that involves recognition of specific junction(s) within the polyprotein. The remainder of the cleavage events could also occur in cis but likely occur in trans (i.e. bimolecular). Since picornavirus proteinases contain sequences and structural features found in other well-known proteases, their study in the context of substrate recognition has furthered our knowledge regarding enzyme–substrate interactions in general. For picornaviruses, three major viral proteinase types have been

described: the L, 2A, and 3C-containing proteinases.

The L proteinase, located at the very amino-terminus of the aphthovirus polyprotein, is a thiol protease that adopts a fold, which resembles that of the enzyme papain. This protein cleaves itself from the viral polyprotein at its own carboxy-terminus (the L-P1 junction). The major function of the aphthovirus L proteinase is the cleavage of the eIF-4G component of eIF-4F, a host cell protein complex critical for cap-dependent translation initiation. This cleavage event mediates, in part, the host cell shutoff during an aphthovirus infection (described above).

The 2A proteinase is responsible for the primary cis cleavage event during enterovirus and rhinovirus polyprotein processing. This involves the cleavage of the scissile bond at the amino-terminus of 2A, liberating the P1 (capsid) polyprotein from P2-P3. The 2A protein of enteroviruses and rhinoviruses adopts a fold similar to that of chymotrypsin (a serine protease) but utilizes a cysteine as a nucleophile rather than a serine. Similar to the L proteinase of aphthoviruses, the major function of the 2A proteinase during enterovirus and rhinovirus infections is the cleavage of eIF-4G, which mediates the shutoff of host cell cap-dependent translation. Interestingly, it has also been suggested that the 2A protein of entero- and rhinoviruses plays a direct role in mediating cap-independent translation, and could also play other roles in RNA replication. For poliovirus, it appears that 2A contains sequences that are not required for proteolytic activity but are necessary for efficient RNA replication. The 2A protein of aphthoviruses, cardioviruses, and hepatoviruses is not proteolytically active. In hepatoviruses, 2A is cleaved from the polyprotein by the 3C proteinase, whereas, in the case of aphthovirus and cardioviruses, the primary cleavage at the carboxy-terminus of 2A occurs by an undefined mechanism that does not rely on the proteolytic activity of any viral proteinase (i.e. 2A, L, or 3C). For these viruses, it is thought that a conserved tetrapeptide motif (NPGP) could be responsible for the autocatalytic release of the 2A carboxy-terminus from the rest of the polyprotein via a breakage of the Gly–Pro bond. If not mediated by the tetrapeptide itself, this cleavage could be carried out by a cellular protease. A more intriguing possibility is the termination of translation just downstream of the conserved glycine residue, thereby preventing formation of the Gly–Pro peptide bond altogether.

The 3C proteinase is responsible for the majority of the cleavage events occurring within the picornavirus polyprotein that are not mediated by the L or 2A proteinases. This highly conserved proteinase adopts a fold similar to that of the serine protease chymotrypsin and is produced by all known picornaviruses. Extensive studies have shown that, for poliovirus, the immediate proteinase precursor (3CD) displays an increased ability to process polyprotein, specifically the capsid (P1) region. Furthermore, the structural and sequence context of the dipeptide, recognized by 3C proteinases (primarily Q-G), is critical for its cleavage activity, a likely reason why all additional Q-G dipeptide sequences present within the viral polyprotein are not recognized by 3C. Studies have also shown that 3C recognizes other substrates of nonviral origin. 3C is responsible for cleaving cellular proteins necessary for host transcription (e.g. TATA binding protein, Oct-1, and transcription factor IIIC), thereby mediating the shutoff of RNA synthesis catalyzed by all three types of cellular RNA polymerase. This presumably requires the ability of

3C to enter the nucleus of the infected cell, and it has been shown that a putative nuclear localization signal (KKKRD), conserved in all picornavirus polymerases, allows the 3CD proteinase precursor to shuttle into the nucleus. Additionally, it has been shown that the poliovirus 3C and/or 3CD proteinase cleaves poly(A) binding protein (PABP) at late times postinfection, which could serve to inhibit cap-dependent translation or destabilize cellular mRNAs.

4 Viral RNA Replication and Packaging

Picornaviruses utilize a relatively small number of viral-encoded gene products to dramatically alter the intracellular environment to allow the specific and rapid amplification of their RNA genomes. This involves extensive virus–host interactions at several intracellular locations and includes the formation of protein–protein and protein–RNA complexes that allow viral replication to occur with limited interference from host defense mechanisms. As outlined below, some of these macromolecular interactions are well-characterized while others are thus far a mystery. This section gives a summary of host cell modifications brought about by a picornavirus infection that are necessary for successful viral replication, followed by summaries of separate viral gene products and their precursors (where applicable) to which distinct replicative functions have been ascribed. It should be noted, however, that there is increasing evidence suggesting that RNA replication complexes assemble with whole precursor (i.e. P2-P3, P2, or P3) forms of these proteins. Proteolysis then occurs subsequent to the assembly of these complexes, generating the immediate precursor and mature polypeptides to carry out the functions described below.

4.1 Modification of the Host Cell Environment

Following the translation and processing of the viral polyprotein by viral-encoded proteinases, an abrupt decrease in the level of cellular cap-dependent translation occurs as the cap-binding complex, eIF-4F, no longer recognizes capped cellular mRNAs due to the proteolytic cleavage of one of its component polypeptides, eIF-4G (see, Sect. 3.4) or other mechanisms involving the sequestration of the cap-binding protein eIF-4E. This occurs within 2 h postinfection and results in the complete hijacking of the eukaryotic translation apparatus that is used exclusively for viral protein synthesis. The process of host cell shutoff was initially thought to involve one or more cellular proteinases that become "activated" by the picornavirus proteinase. However, 2A-mediated eIF-4G cleavage can be reconstituted with purified components *in vitro*, demonstrating that this process can occur in the absence of any cellular proteins. As previously mentioned, the products of eIF-4G cleavage actually participate in the process of IRES-mediated viral translation.

A hallmark of a typical picornavirus infection is the rearrangement of host intracellular membranes, mediated primarily by proteins encoded in the P2 region of the picornavirus genome, resulting in the formation of rosette-like structures consisting of endoplasmic reticulum and Golgi-derived membranes. This rearrangement compartmentalizes the intracellular environment, allowing certain

host and viral proteins to associate specifically with the genomic RNA while nonessential factors are excluded. The association of viral RNA with vesicles is maintained by proteins (described below) contained within the P2 and P3 regions of the genome that anchor RNA replication complexes. In addition to their physical arrangement, the protein-trafficking properties of these membranes also change. For example, a poliovirus infection inhibits ER-to-Golgi transport and increases membrane permeability. Interestingly, poliovirus may also utilize these membrane modifications to evade the host immune response by downregulating the transport of Class I Major Histocompatibility Complex molecules to the surface of the cell. While the transport of certain proteins to the plasma membrane is inhibited, others, like viral receptor molecules, are upregulated during an infection. Picornaviruses also modify nuclear-cytoplasmic trafficking within the infected cell by altering the nuclear pore complex, perhaps allowing the virus to utilize one or more host nuclear proteins for RNA replication occurring in the cytoplasm.

Many picornavirus infections also influence cellular signal transduction pathways by changing the phosphorylation patterns of their protein components. It was shown that infection of L929 (mouse fibroblast) cells by EMCV alters the phosphorylation state of one or more proteins in the MAP kinase pathway, activating them for use in IRES-mediated viral translation. It appears that other picornaviruses, including coxsackievirus (an enterovirus), also take advantage of these same pathways to mediate replication processes. Picornaviruses downmodulate several components of the apoptotic pathway (e.g. procaspases) to prevent premature cell death during an infection. Since picornaviruses likely replicate via the generation of a double-stranded RNA intermediate, they have also developed means by which to modulate the IFN α/β expression critical for the cellular antiviral response to dsRNA.

4.2
Viral Protein 2A

As previously described, the 2A protein of picornaviruses plays a major role in the shutoff of host protein synthesis (except for aphtho-, cardio-, and hepatoviruses in which 2A is not a proteinase). The active 2A protein catalyzes the primary cleavage event that liberates the P1 (capsid) region of the polyprotein from the P2-P3 (nonstructural) precursor polypeptide. Interestingly, the 2A protein of poliovirus may also play a role in viral RNA replication, involving carboxy-terminal sequences within the PV 2A protein that are not required for cis- or trans-proteolytic activity.

4.3
Viral Protein 2B

The 2B protein modifies intracellular membrane properties by increasing their permeability, inhibiting cellular exocytosis, and dissociating the Golgi complex to form vesicles upon which RNA replication occurs. It has been proposed that the increase in membrane permeability could be critical at late stages during an infection by facilitating the release of newly formed virions. Though its exact role has not been defined, the 2B protein is critical to the process of viral RNA replication since viruses harboring mutations in this gene have a small plaque phenotype and cannot be complemented in trans.

4.4 Viral Proteins 2C and 2BC

A precise function for the picornavirus 2C protein has not yet been described. However, 2C plays a role in the initiation of viral negative-strand RNA synthesis and contains NTPase motifs that are homologous to those found in DNA helicases. 2C helicase activity has not been observed directly, but its NTPase activity has been demonstrated. This NTPase activity is inhibited by guanidine-HCl, a chaotropic agent that prevents negative-strand RNA synthesis. Furthermore, guanidine-HCl resistant viruses contain mutations in their 2C coding sequences, lending support to the idea that this protein is critical to the process of RNA replication. Additionally, 2C may have a role in RNA encapsidation, as viruses resistant to encapsidation inhibitors (i.e. hydantoin), contain mutations in their 2C coding regions.

The 2C protein contains two known RNA binding motifs, and it is thought that this protein, along with its immediate precursor 2BC, binds sequences in the 3′ NCR of the negative-strand viral RNA to mediate positive-strand RNA synthesis initiation. In addition to functioning as RNA binding proteins, the 2C and 2BC polypeptides are critical for vesicle formation and directly associate with membranes in RNA replication centers. Indeed, the expression of these proteins in HeLa cells (in the absence of a viral infection) leads to the formation of membranous structures identical to those seen in a picornavirus infection. However, subsequent viral infection of the morphologically altered HeLa cells has shown that the preformed vesicles are not utilized by the incoming virus for RNA replication, demonstrating roles for these proteins in linking RNA replication to vesicle formation.

4.5 Viral Protein 3A

Protein 3A is a small, hydrophobic polypeptide that associates with intracellular membranes and is responsible for the inhibition of protein secretion and intracellular membrane transport. Protein 3A has been shown to be directly responsible for inhibiting the presentation of Class I Major Histocompatibility Complex molecules at the surface of the infected cell and has been shown to decrease the cellular secretion of antiviral cytokines such as IL-6 (interleukin-6), IL-8, and IFN-β. Protein 3A may contribute to the ability of picornaviruses to evade host immunity and could act in concert with viral protein 2C and/or 2BC in rearranging membranes into replication vesicles.

4.6 Viral Proteins 3B and 3AB

3B (also known as VPg) is a small, highly charged polypeptide that is covalently linked to the 5′ end of both positive- and negative-strand viral RNA. 3B (VPg) is a substrate for uridylylation (at a tyrosine conserved in all picornaviruses) by the viral RNA-dependent RNA polymerase, 3D. In addition to the presence of VPg covalently coupled to the 5′ end of negative- and positive-strand viral RNAs, the fact that all picornavirus RNAs begin with uridine nucleotides strongly suggests that VPg is the protein primer for RNA synthesis. Details surrounding RNA synthesis initiation and VPg are discussed later in this section.

Viral protein 3AB could act as an immediate precursor to the 3B (VPg) polypeptide

that assembles with other proteins in replication complexes. The 3AB polypeptide may provide a source of VPg localized to membranous replication complexes, since the 3A region of 3AB contains hydrophobic sequences known to mediate insertion of 3AB into vesicles. It has been shown that 3AB interacts with and stimulates the activity of the 3CD proteinase and 3D RNA-dependent RNA polymerase. In this regard, 3AB could recruit the viral polymerase to replication complexes. The 3AB polypeptide can also bind viral RNA sequences within the 3′ NCR of the genomic RNA; and via protein–protein contacts with the 3CD or 3D polypeptide, could localize the polymerase to the site of negative-strand synthesis initiation. The importance of the 3AB precursor protein is underscored by the fact that some picornavirus RNAs possessing optimized 3A/3B cleavage junctions are deficient in viral RNA synthesis but can be complemented with 3AB or 3ABC precursor polypeptides in trans.

4.7 Viral Proteins 3C and 3CD

The viral protein 3C is multifunctional in nature, acting as a chymotrypsin-like proteinase (described in Sect. 3) and binding viral RNA. The three-dimensional structures of the 3C proteinases from poliovirus, hepatitis A virus, and rhinovirus have been solved, and indicate the presence of RNA binding determinants that lie directly opposite the catalytic triad. The best characterized 3C molecule is that of poliovirus, whose immediate precursor (the 3CD polypeptide) is also an active proteinase and RNA binding protein. Although 3CD is also the precursor to the 3D RNA-dependent RNA polymerase, it does not possess measurable elongation activity *in vitro*. The presence of 3D polymerase sequences in the context of 3CD appears to enhance the ability of 3C to bind viral RNA and process polyprotein (specifically the P1 capsid precursor).

The 3CD precursor polypeptide plays central roles in viral RNA replication. 3CD forms a complex with host protein, PCBP and a viral RNA secondary structure ("cloverleaf") contained within the first ~100 nucleotides of the positive-strand RNA genome. Although the primary RNA binding determinants reside within the 3C portion of the 3CD polypeptide, it has been suggested that secondary, possibly lower affinity, binding determinants exist in the 3D polymerase domain of the molecule. The 3CD/PCBP/RNA cloverleaf complex is absolutely essential for RNA replication, with genetic evidence pointing to a role in negative-strand RNA synthesis. Since 3CD and PCBP participate in protein–protein contacts with PABP, it has been hypothesized that a long-range interaction could occur involving PABP bound to the 3′ poly(A) tract at the 3′ end of the viral RNA with 3CD and PCBP bound to the 5′ cloverleaf. Once assembled, a cis or trans cleavage of the 3CD molecule could occur to yield a 3D polymerase molecule that elongates from the VPg-primed template.

Finally, 3CD likely plays an important role in protein-primed viral RNA synthesis initiation. The process of VPg uridylylation, catalyzed by the 3D polymerase and responsible for generating the VPg-pU-pU RNA synthesis primer, is greatly stimulated (100 to 1000-fold) by the 3CD polypeptide. The mechanism behind this effect is unclear, but could involve direct binding of 3CD to a cis-acting RNA replication element utilized by picornaviruses as a template for this reaction. More details

regarding this process are given later in this section.

4.8 The Viral RNA-dependent RNA Polymerase 3D

The RNA-dependent RNA polymerases from several picornaviruses have been well-characterized. All are primer-dependent, can recognize and elongate from artificial (nonviral) templates, and are Mg^{2+}-dependent and actinomycin-D insensitive. Similar to those of other RNA viruses, the picornavirus RNA polymerase is highly error prone, misincorporating ~1 to 2 nucleotides per template copying event. Although this results in error prone replication, whereby a slight increase in mutation frequency could dramatically decrease viral fitness, it allows picornaviruses to evolve very rapidly. Within a given population of replicating viruses, there exist many subpopulations that have each generated unique sets of coding sequence mutations, many of which allow the virus to successfully evade the host antiviral immune response. Despite having an attenuated growth phenotype, the nonneurotropic Sabin vaccine strains of poliovirus are capable of regaining neurotropism via random nucleotide mutations generated by the RNA polymerase, resulting in rare cases of vaccine-derived poliomyelitis. Some drugs (e.g. ribavirin, a ribonucleoside analog) push this delicate equilibrium toward "error catastrophe" by increasing nucleotide misincorporation, producing strong antiviral effects. The development of optimized ribonucleoside analogs could yield antiviral drugs with broad-spectrum efficacies against many RNA viruses.

The three-dimensional structure of the poliovirus 3D polymerase has been solved. It adopts the conformation of a cupped "right-hand," a structural organization that is conserved among virtually all known polymerases. The poliovirus 3D polymerase is a highly cooperative enzyme, and utilizes extensive regions of polymerase–polymerase contacts to mediate polymerase oligomerization. This phenomenon was originally observed in the crystal lattice used to determine its three-dimensional structure, and has since been shown to be important for biological activity. "Interface I," involving contacts between the thumb subdomain of one polymerase molecule and the palm subdomain of an adjacent polymerase, is critical for binding primed nucleic acid template. "Interface II," involving interactions between the top of the thumb of one polymerase molecule and the base of the fingers of an adjacent polymerase, is necessary for the formation of catalytic sites necessary for efficient RNA chain elongation. Although polymerase–polymerase interactions have been observed in other RNA polymerases, the extensive interactions observed in the poliovirus polymerase may be unique.

4.9 VPg Uridylylation: Priming RNA Synthesis

The viral RNA-dependent RNA polymerase is the most highly conserved gene amongst members of the family *Picornaviridae*. Not only does it catalyze RNA chain elongation, it also catalyzes the generation of the protein primer itself. The addition of uridine nucleotides via a phosphodiester linkage at an absolutely conserved tyrosine of VPg (the 3B viral gene product) requires only a few reaction components, and can be recapitulated *in vitro* with synthetic VPg (a small protein of 22 amino acids), recombinant 3D polymerase, and a suitable

template. As noted above, the addition of polymerase precursor polypeptide (3CD) greatly stimulates this process *in vitro*. Early studies carried out with reaction components from poliovirus showed that a poly(A) tract alone can act as a template for the 3D-catalyzed synthesis of VPg-pU-pU, suggesting a common location for both VPg uridylylation and the priming of RNA synthesis initiation. However, nascently synthesized VPg-pU-pU may not be immediately used for RNA synthesis initiation, since free uridylylated VPg exists within the infected cell, possibly representing a pool from which VPg primers are taken when necessary.

The effects of a number of mutations within the picornavirus RNA polymerase have been documented in the literature. Work aimed at determining sequences within the polymerase critical for the process of VPg uridylylation has yielded some interesting correlations with previously described RNA synthesis phenotypes of viruses harboring mutations in these regions. For example, a mutation within the 3D RNA polymerase of the Sabin 1 strain of poliovirus was shown to confer a temperature-sensitive defect in VPg uridylylation, suggesting that its tissue-specific replication phenotype could be due, in part, to reduced polymerase activity.

4.10 Cis-acting Replication Elements

The mechanisms involved in the specific recognition of viral RNA in the presence of a vast excess of cellular messenger RNA involves structures and sequences within the picornavirus genome that interact specifically with host and viral proteins, and possibly with each other. In addition to sequences contained within the 5′ and 3′ noncoding regions of the viral RNA, several picornaviruses have been shown to possess structurally conserved RNA hairpins within their genomes that are essential for viral replication. First discovered in the genomic RNA of human rhinovirus 14 (HRV14), *cre* (*c*is-acting *r*eplication *e*lement) structures have also been described for enteroviruses, cardioviruses, and aphthoviruses. For a cre element to facilitate RNA replication, it must be present in a positive-strand context. Furthermore, its location within the genome is highly variable. In fact, the location of a cre element can be genetically manipulated without significantly disrupting RNA replication. A nonfunctional cre element can be complemented by a wild-type cre contained in a different region of the same RNA (i.e. in cis) or in trans from another viral RNA. For unknown reasons, evolutionary pressures have not resulted in a conservation of cre location, as closely related members of the same genus can possess cre elements in very different regions of their genomes. For example, the HRV2 cre element is located within the 2A proteinase gene, whereas the HRV14 cre lies within the VP1 capsid protein-coding region.

Analysis of cis-acting replication elements from a number of different picornaviruses show that they share a similar overall structure but vary in their primary sequences. A conserved AAACA motif has been found in all cre elements described thus far, suggesting only minimal primary sequence elements are required for cre function. It has been shown that the VPg uridylylation reaction is more efficient in the presence of the cre element rather than the 3′ poly(A) tract. For poliovirus, a strand-specific function for the cre element has been described, suggesting that the major role of the poliovirus

cre element is to mediate VPg uridylylation required for positive – rather than negative – strand RNA synthesis initiation. This suggests that poly(A) template-dependent VPg uridylylation could function primarily in negative-strand RNA synthesis initiation, and implicates the existence of strand-specific ribonucleoprotein complexes that form on the viral RNA for positive- versus negative-strand synthesis initiation.

4.11 A Model for Picornavirus RNA Replication

Figure 2 shows a schematic summary of the major steps thought to be involved in the process of picornavirus RNA replication, based largely on *in vitro* and cell culture data from experiments aimed at elucidating the mechanisms that underlie poliovirus (the prototypic picornavirus) RNA synthesis. The first step in the process of viral RNA replication is the synthesis of a negative-strand RNA utilizing the genomic positive-strand RNA as a template. As previously discussed, this utilizes a VPg-primed mechanism to initiate RNA synthesis, presumably near the terminus of the 3′ poly(A) tract. Several viral and/or host factors are thought to participate in complex formation at the 3′ end of the RNA, and the role of the 3′ NCR immediately upstream of the 3′ poly(A) tract is not exactly known. Whereas the 3′ poly(A) tract is absolutely required for picornavirus infectivity, it has been shown that for human rhinovirus and poliovirus, deleting the entire 3′ NCR (but leaving the poly(A) tract intact) does not completely inhibit viral replication. Although viruses harboring 3′ NCR deletions exhibit reduced RNA synthesis levels, they are capable of near wild-type levels of negative-strand RNA synthesis. This suggests that the poly(A) tract alone is sufficient for this process, and sequences and structures within the 3′ NCR may serve to optimize RNA replication initiation subsequent to synthesis of the negative strand.

The 3D RNA polymerase alone is capable of recognizing poly(A)-primer duplex RNA *in vitro*, but it is likely that the 3D polymerase participates in protein–protein interactions with other host and/or viral proteins necessary for efficient RNA synthesis *in vivo*. For example, it has been shown that the 3AB protein stimulates 3D polymerase activity *in vitro*, and 3AB-3D interactions have been confirmed by genetic analysis. Additionally, it is thought that the 3AB protein interacts with processed 3B (VPg) and the 3CD polypeptide, which could recruit the polymerase precursor(s) and VPg to the site of uridylylation at the 3′ poly(A) tract. The 3AB precursor may also anchor replication complexes associated with the viral RNA mediated by 3A amino acid sequences known to associate with membranous vesicles.

Once the negative-strand RNA is synthesized, this RNA is utilized as a template for multiple initiation events that produce large amounts of positive-strand RNAs. It is likely that a functional negative-strand RNA remains at least partially annealed to the positive strand from which it was generated, but this has not been directly demonstrated. Double-stranded RNA has been shown to be infectious in tissue culture, and a typical picornavirus infection results in a strong PKR-mediated dsRNA response in the infected cell, supporting the existence of a double-strand RNA (termed "RF" or *r*eplicative *f*orm) as a biologically relevant intermediate during genome amplification.

At one or more steps during the process of RNA replication, the viral RNA

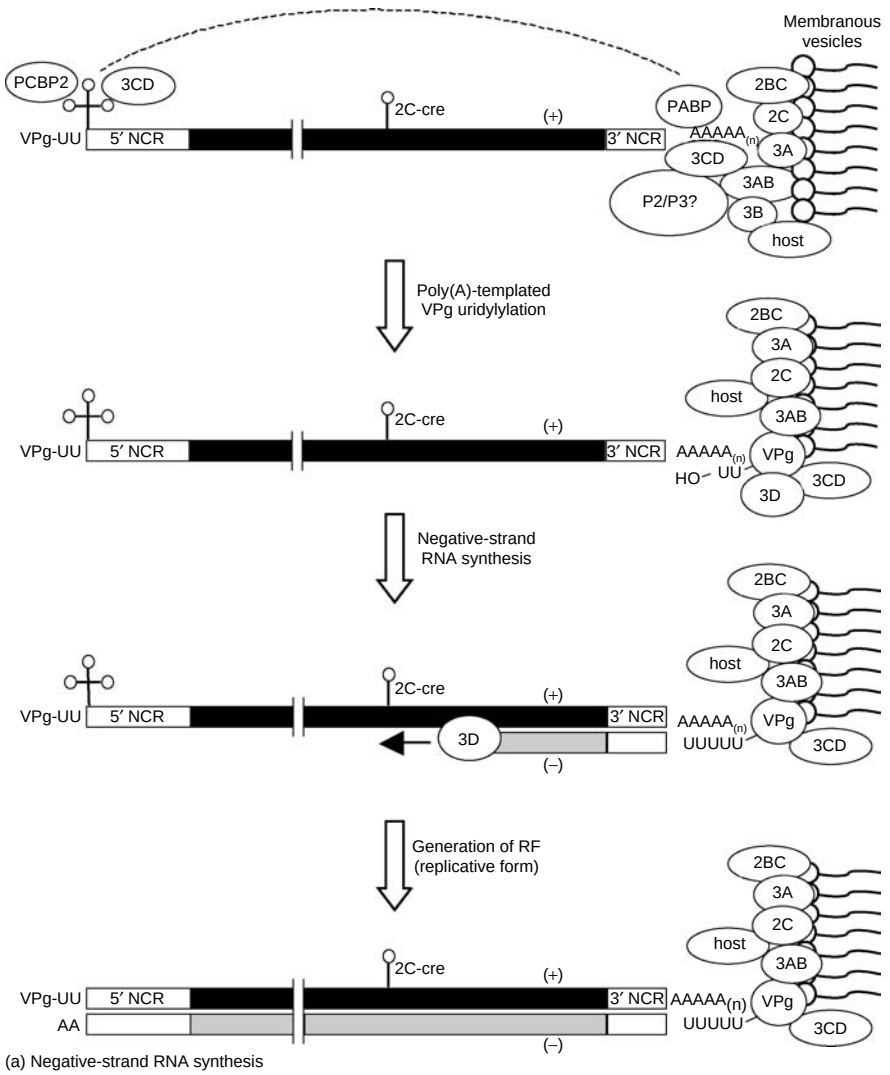

Fig. 2 A proposed mechanism for poliovirus RNA synthesis. (a) Negative-strand RNA synthesis. Precursor polypeptides (P2 and/or P3) or partially processed replication proteins (e.g. 2BC, 2C, 3AB, 3A) along with host factor(s) assemble at the 3′ end of the genomic positive-strand RNA in tight association with membranous vesicles (depicted by open circles and lines representing the polar headgroups of the lipid bilayer). Host protein PCBP2 and viral protein 3CD bind the 5′ RNA cloverleaf and may facilitate long-range contacts (dashed line) with PABP bound to the 3′ poly(A) tract. The 3′ poly(A) tract is utilized as a template for the 3D-catalyzed VPg uridylylation reaction (stimulated by the 3CD precursor polypeptide) to generate the VPg-pU-pU RNA synthesis primer. The 3D RNA polymerase remains associated with the uridylylated VPg or is provided from another source (possibly a 3CD precursor molecule) and catalyzes elongation of the negative-strand RNA, generating the RF (replicative form) dsRNA intermediate.

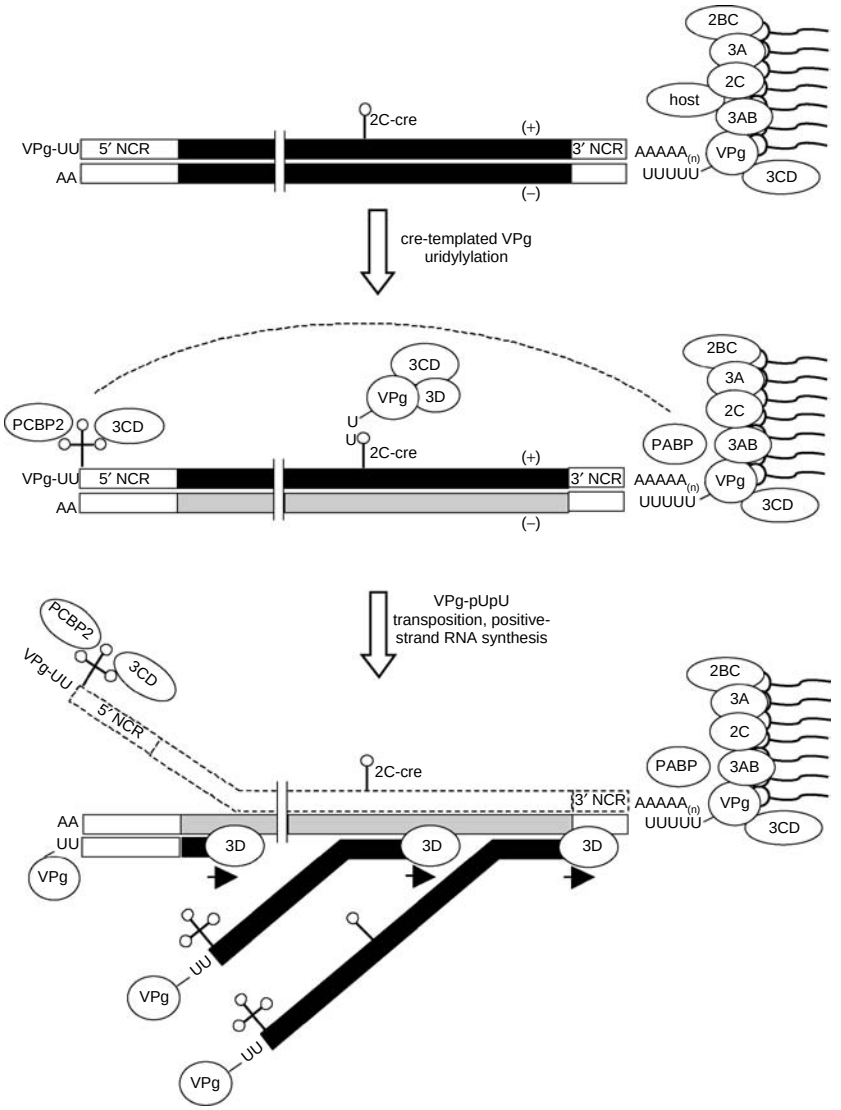

(b) Positive-strand RNA synthesis

Fig. 2 (*Continued*). (b) For positive-strand RNA synthesis, the cis-acting replication element (2C-cre) is utilized as a template for VPg uridylylation (instead of the 3′ poly(A) tract) that may also involve long-range 5′–3′ interactions (dashed line) similar to those described in panel A, facilitating translocation of the resulting VPg-pU-pU primer to the site of positive-strand RNA synthesis initiation. Once annealed to the 3′ end of the negative-strand RNA, the 3D polymerase synthesizes nascent positive-strand RNAs. Multiple RNA synthesis initiation events may occur in succession on the same replicative form dsRNA intermediate, consisting of a positive-strand RNA (indicated by dashed region) that is partially hybridized to a negative-strand RNA. It is not known if the parental positive-strand is displaced during this process by a progeny RNA molecule that remains duplexed with the negative-strand RNA.

could engage in long-range interactions facilitated by proteins bound to the 5′ and 3′ termini of the viral RNA. For poliovirus, PABP has been shown to interact with host protein, PCBP2 and the viral protein 3CD. With PCBP2 and 3CD bound to the cloverleaf element at the 5′ end of the genome, binding to PABP at the 3′ poly(A) tract could allow long-range interactions of the RNA and prepare it for negative- and/or positive-strand synthesis initiation, which could utilize a 3D polymerase liberated via the processing of 3CD within this ribonucleoprotein complex. Such a mechanism for RNA synthesis initiation is thought to occur primarily during negative-strand synthesis initiation, but it may occur for positive-strand synthesis as well. As mentioned, it has been shown that the poliovirus cis-acting replication element (2C-cre) is utilized primarily for the priming of positive-strand RNA synthesis. Given this, it is possible that a dsRNA intermediate (RF) maintains intrastrand contacts to facilitate positive-strand RNA synthesis initiation, bringing the cre element (located in the central portion of the coding sequence) in close proximity to the sites of initiation complex formation and VPg uridylylation.

4.12 RNA Encapsidation and Cell Lysis

The process of viral RNA synthesis results in the production of a vast excess of positive-strand RNA molecules from a single input genomic RNA. As they accumulate, these positive-strand RNAs are utilized as templates for additional rounds of viral translation or RNA synthesis, or they are encapsidated into virion particles. The mechanism by which viral RNA is selected for encapsidation is not clear, but it is likely that one or more signals within the RNA are recognized by viral and/or host factors that trigger encapsidation. On the basis of studies with poliovirus and other picornaviruses (e.g. Aichi virus), these signals may be contained within the IRES element found in the 5′ noncoding region of virion RNA. Furthermore, poliovirus replication carried out in the presence of hydantoin (which inhibits encapsidation) has revealed that the viral protein 2C may play a role in this process, since mutations in the 2C region of the genome confer resistance to this drug.

A schematic of the process of virion assembly is shown in Fig. 3. Once the 3C (or 3CD) proteinase processes the P1 (capsid) precursor polypeptide into VP0, VP3, and VP1, capsid assembly begins with the formation of a ''protomer,'' consisting of one copy of each of these proteins. Five protomers self-associate to form a pentamer, and twelve pentamers then arrange themselves around a newly synthesized RNA, forming the ''provirion'' structure. Alternatively, the pentamers may form an empty capsid structure (the 80S ''procapsid''; see Fig. 3) into which the RNA molecule is threaded. Independent of the assembly pathway, most picornavirus virions undergo a maturation event, which processes VP0 to VP2 and VP4, a cleavage mediated by an undefined proteinase that serves to stabilize the virion particle. This proteinase is not one of the known viral enzymes, and it has been hypothesized that the capsid proteins and/or the genomic RNA itself could carry out this reaction. Once a threshold number of virions are assembled within the infected cell, the cell undergoes lysis to release virus particles that go on to infect neighboring cells. The exact mechanism behind cell lysis is not known.

The release of progeny virions represents the successful culmination of a

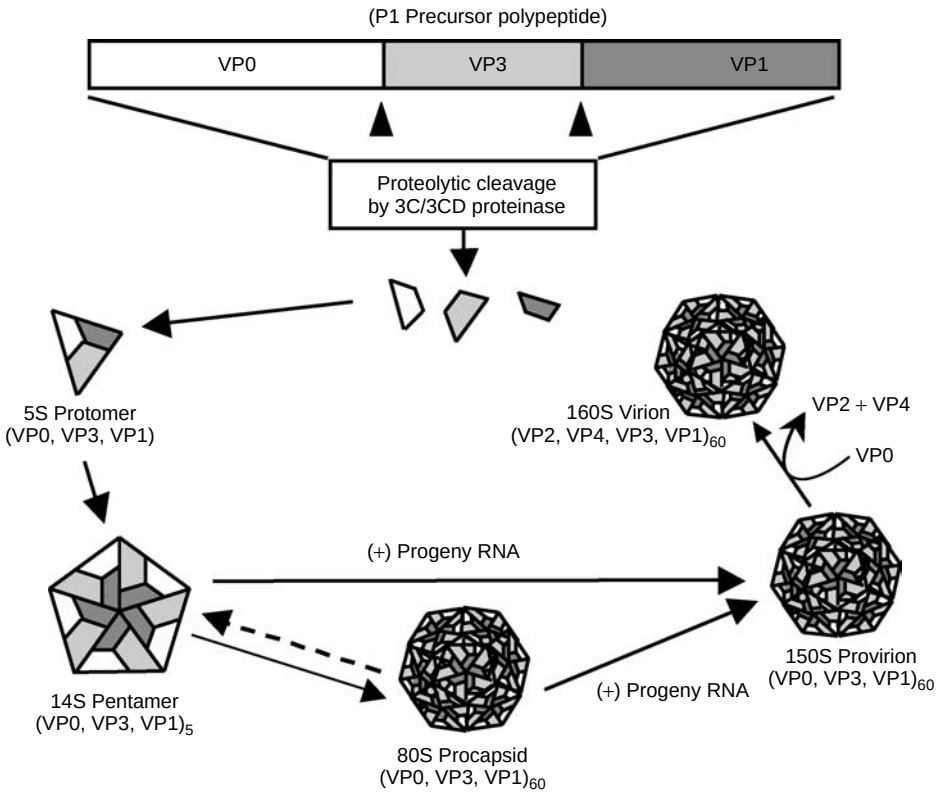

Fig. 3 Assembly of the picornavirus virion. The 3C/3CD proteinase cleaves the P1 capsid precursor to generate VP0, VP3, and VP1. These three proteins assemble into a 5S protomer. Five protomers then assemble into a 14S pentamer. Twelve 14S pentamers may reversibly assemble into the 80S procapsid (indicated by the solid and dashed lines) into which the progeny RNA is threaded, or may themselves assemble around the RNA molecule. In either case, this process results in the formation of the 150S provirion. The provirion undergoes a maturation event (except in parechoviruses), which cleaves VP0 into VP2 and VP4 by an undefined mechanism, forming the mature 160S virion particle.

series of complex protein–protein and protein–RNA interactions that allow a virus with a very limited coding capacity to effectively alter and subsequently utilize machinery within the host for its own translation and replication. Picornaviruses represent some of the most successful of all pathogens known to man, and the study of multilevel virus–host interplay has yielded important insights into not only viral processes but into cellular processes as well. Importantly, picornaviruses that cause very different diseases in their susceptible hosts appear to utilize similar replication strategies that could some day allow for the development of broad-range antiviral treatments that would be effective against several picornavirus infections. This article has attempted to provide a brief overview of some of the interesting aspects of picornavirus replication and gene expression. The reader is encouraged to explore the vast collection of primary research findings that are the basis of this

review, a small portion of which can be found in articles referenced in the bibliography appearing below.

See also RNA Virus Genome Packaging.

Bibliography

Books and Reviews

Andino, R., Boddeker, N., Silvera, D., Gamarnik, A.V. (1999) Intracellular determinants of picornavirus replication, *Trends Microbiol.* **7**, 76–82.

Belsham, G.J., Sonenberg, N. (2000) Picornavirus RNA translation: roles for cellular proteins, *Trends Microbiol.* **8**, 330–335.

Dougherty, W.G., Semler, B.L. (1993) Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes, *Microbiol. Rev.* **57**, 781–822.

Hogle, J.M. (2002) Poliovirus cell entry: common structural themes in viral cell entry pathways, *Annu. Rev. Microbiol.* **56**, 677–702.

Leong, L.E., Cornell, C.T., Semler, B.L. (2002) Processing Determinants and Functions of Cleavage Products of Picornavirus Polyproteins, in: Semler, B.L., Wimmer, E. (Eds.) *Molecular Biology of Picornaviruses*, ASM Press, Washington, D.C., U.S.A., pp. 187–197.

Rossmann, M.G., He, Y., Kuhn, R.J. (2002) Picornavirus-receptor interactions, *Trends Microbiol.* **10**, 324–331.

Semler, B.L., Wimmer, E. (Eds.) (2002) *Molecular Biology of Picornaviruses*, ASM Press, Washington, D.C., U.S.A.

Primary Literature

Aldabe, R., Barco, A., Carrasco, L. (1996) Membrane permeabilization by poliovirus proteins 2B and 2BC, *J. Biol. Chem.* **271**, 23134–23137.

Aldabe, R., Irurzun, A., Carrasco, L. (1997) Poliovirus protein 2BC increases cytosolic free calcium concentrations, *J. Virol.* **71**, 6214–6217.

Andino, R., Rieckhof, G.E., Achacoso, P.L., Baltimore, D. (1993) Poliovirus RNA synthesis utilizes an RNP complex formed around the 5′-end of viral RNA, *EMBO J.* **12**, 3587–3598.

Andino, R., Rieckhof, G.E., Baltimore, D. (1990) A functional ribonucleoprotein complex forms around the 5′ end of poliovirus RNA, *Cell* **63**, 369–380.

Banerjee, R., Tsai, W., Kim, W., Dasgupta, A. (2001) Interaction of poliovirus-encoded 2C/2BC polypeptides with the 3′ terminus negative-strand cloverleaf requires an intact stem-loop b, *Virology* **280**, 41–51.

Barco, A., Carrasco, L. (1995) A human virus protein, poliovirus protein 2BC, induces membrane proliferation and blocks the exocytic pathway in the yeast Saccharomyces cerevisiae, *EMBO J.* **14**, 3349–3364.

Barton, D.J., O'Donnell, B.J., Flanegan, J.B. (2001) 5′ cloverleaf in poliovirus RNA is a cis-acting replication element required for negative-strand synthesis, *EMBO J.* **20**, 1439–1448.

Bergmann, E.M., Mosimann, S.C., Chernaia, M.M., Malcolm, B.A., James, M.N. (1997) The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition, *J. Virol.* **71**, 2436–2448.

Bienz, K., Egger, D., Pasamontes, L. (1987) Association of polioviral proteins of the P2 genomic region with the viral replication complex and virus-induced membrane synthesis as visualized by electron microscopic immunocytochemistry and autoradiography, *Virology* **160**, 220–226.

Blyn, L.B., Towner, J.S., Semler, B.L., Ehrenfeld, E. (1997) Requirement of poly(rC) binding protein 2 for translation of poliovirus RNA, *J. Virol.* **71**, 6243–6246.

Chen, C.Y., Sarnow, P. (1995) Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs, *Science* **268**, 415–417.

Chen, H.H., Kong, W.P., Zhang, L., Ward, P.L., Roos, R.P. (1995) A picornaviral protein synthesized out of frame with the polyprotein plays a key role in a virus-induced immune-mediated demyelinating disease, *Nat. Med.* **1**, 927–931.

Cho, M.W., Teterina, N., Egger, D., Bienz, K., Ehrenfeld, E. (1994) Membrane rearrangement and vesicle induction by recombinant

poliovirus 2C and 2BC in human cells, *Virology* **202**, 129–145.

Cornell, C.T., Semler, B.L. (2002) Subdomain specific functions of the RNA polymerase region of poliovirus 3CD polypeptide, *Virology* **298**, 200–213.

Crotty, S., Andino, R. (2002) Implications of high RNA virus mutation rates: lethal mutagenesis and the antiviral drug ribavirin, *Microb. Infect.* **4**, 1301–1307.

Crotty, S., Cameron, C.E., Andino, R. (2001) RNA virus error catastrophe: direct molecular test by using ribavirin, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6895–6900.

Deitz, S.B., Dodd, D.A., Cooper, S., Parham, P., Kirkegaard, K. (2000) MHCI-dependent antigen presentation is inhibited by poliovirus protein 3A, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13790–13795.

Doedens, J.R., Kirkegaard, K. (1995) Inhibition of cellular protein secretion by poliovirus proteins 2B and 3A, *EMBO J.* **14**, 894–907.

Echeverri, A.C., Dasgupta, A. (1995) Amino terminal regions of poliovirus 2C protein mediate membrane binding, *Virology* **208**, 540–553.

Gamarnik, A.V., Andino, R. (1997). Two functional complexes formed by KH domain containing proteins with the 5′ noncoding region of poliovirus RNA, *RNA* **3**, 882–892.

Gamarnik, A.V., Andino, R. (2000) Interactions of viral protein 3CD and poly(rC) binding protein with the 5′ untranslated region of the poliovirus genome, *J. Virol.* **74**, 2219–2226.

Glass, M.J., Jia, X.Y., Summers, D.F. (1993) Identification of the hepatitis A virus internal ribosome entry site: in vivo and in vitro analysis of bicistronic RNAs containing the HAV 5′ noncoding region, *Virology* **193**, 842–852.

Goodfellow, I., Chaudhry, Y., Richardson, A., Meredith, J., Almond, J.W., Barclay, W., Evans, D.J. (2000) Identification of a cis-acting replication element within the poliovirus coding region, *J. Virol.* **74**, 4590–4600.

Graff, J., Cha, J., Blyn, L.B., Ehrenfeld, E. (1998) Interaction of poly(rC) binding protein 2 with the 5′ noncoding region of hepatitis A virus RNA and its effects on translation, *J. Virol.* **72**, 9668–9675.

Grubman, M.J., Zellner, M., Bablanian, G., Mason, P.W., Piccone, M.E. (1995) Identification of the active-site residues of the 3C proteinase of foot-and-mouth disease virus, *Virology* **213**, 581–589.

Haller, A.A., Semler, B.L. (1992) Linker scanning mutagenesis of the internal ribosome entry site of poliovirus RNA, *J. Virol.* **66**, 5075–5086.

Hansen, J.L., Long, A.M., Schultz, S.C. (1997) Structure of the RNA-dependent RNA polymerase of poliovirus, *Structure* **5**, 1109–1122.

Harris, K.S., Xiang, W., Alexander, L., Lane, W.S., Paul, A.V., Wimmer, E. (1994) Interaction of poliovirus polypeptide $3CD^{pro}$ with the 5′ and 3′ termini of the poliovirus genome. Identification of viral and cellular cofactors needed for efficient binding, *J. Biol. Chem.* **269**, 27004–27014.

Hellen, C.U., Wimmer, E. (1995) Translation of encephalomyocarditis virus RNA by internal ribosomal entry, *Curr. Top. Microbiol. Immunol.* **203**, 31–63.

Herold, J., Andino, R. (2001) Poliovirus RNA replication requires genome circularization through a protein-protein bridge, *Mol. Cell* **7**, 581–591.

Hobson, S.D., Rosenblum, E.S., Richards, O.C., Richmond, K., Kirkegaard, K., Schultz, S.C. (2001) Oligomeric structures of poliovirus polymerase are important for function, *EMBO J.* **20**, 1153–1163.

Hogle, J.M., Chow, M., Filman, D.J. (1985) Three-dimensional structure of poliovirus at 2.9 Å resolution, *Science* **229**, 1358–1365.

Hunt, S.L., Skern, T., Liebig, H.D., Kuechler, E., Jackson, R.J. (1999). Rhinovirus 2A proteinase mediated stimulation of rhinovirus RNA translation is additive to the stimulation effected by cellular RNA binding proteins, *Virus Res.* **62**, 119–128.

Jackson, R.J., Howell, M.T., Kaminski, A. (1990). The novel mechanism of initiation of picornavirus RNA translation, *Trends Biochem. Sci.* **15**, 477–483.

Jang, S.K., Krausslich, H.G., Nicklin, M.J., Duke, G.M., Palmenberg, A.C., Wimmer, E. (1988) A segment of the 5′ nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation, *J. Virol.* **62**, 2636–2643.

Kaminski, A., Howell, M.T., Jackson, R.J. (1990) Initiation of encephalomyocarditis virus RNA translation: the authentic initiation site is not selected by a scanning mechanism, *EMBO J.* **9**, 3753–3759.

Kolupaeva, V.G., Pestova, T.V., Hellen, C.U., Shatsky, I.N. (1998) Translation eukaryotic initiation factor 4G recognizes a specific

structural element within the internal ribosome entry site of encephalomyocarditis virus RNA, *J. Biol. Chem.* **273**, 18599–18604.

Le, S.Y., Siddiqui, A., Maizel, J.V.J. (1996) A common structural core in the internal ribosome entry sites of picornavirus, hepatitis C virus, and pestivirus, *Virus Genes* **12**, 135–147.

Li, X., Lu, H.H., Mueller, S., Wimmer, E. (2001) The C-terminal residues of poliovirus proteinase 2A(pro) are critical for viral RNA replication but not for cis- or trans-proteolytic cleavage, *J. Gen. Virol.* **82**, 397–408.

Lyle, J.M., Bullitt, E., Bienz, K., Kirkegaard, K. (2002) Visualization and functional analysis of RNA-dependent RNA polymerase lattices, *Science* **296**, 2218–2222.

McKnight, K.L., Lemon, S.M. (1998) The rhinovirus type 14 genome contains an internally located RNA structure that is required for viral replication, *RNA* **4**, 1569–1584.

Molla, A., Paul, A.V., Wimmer, E. (1991) Cell-free, de novo synthesis of poliovirus, *Science* **254**, 1647–1651.

Morasco, B.J., Sharma, N., Parilla, J., Flanegan, J.B. (2003) Poliovirus cre(2C)-dependent synthesis of VPgpUpU is required for positive- but not negative-strand RNA synthesis, *J. Virol.* **77**, 5136–5144.

Mosimann, S.C., Cherney, M.M., Sia, S., Plotch, S., James, M.N. (1997) Refined X-ray crystallographic structure of the poliovirus 3C gene product, *J. Mol. Biol.* **273**, 1032–1047.

Murray, K.E., Barton, D.J. (2003) Poliovirus CRE-dependent VPg uridylylation is required for positive-strand RNA synthesis but not for negative-strand RNA synthesis, *J. Virol.* **77**, 4739–4750.

Najarian, R., Caput, D., Gee, W., Potter, S.J., Renard, A., Merryweather, J., Van Nest, G., Dina, D. (1985) Primary structure and gene organization of human hepatitis A virus, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2627–2631.

Niepmann, M., Petersen, A., Meyer, K., Beck, E. (1997) Functional involvement of polypyrimidine tract-binding protein in translation initiation complexes with the internal ribosome entry site of foot-and-mouth disease virus, *J. Virol.* **71**, 8330–8339.

Nugent, C.I., Johnson, K.L., Sarnow, P., Kirkegaard, K. (1999) Functional coupling between replication and packaging of poliovirus replicon RNA, *J. Virol.* **73**, 427–435.

Ochs, K., Saleh, L., Bassili, G., Sonntag, V.H., Zeller, A., Niepmann, M. (2002) Interaction of translation initiation factor eIF4B with the poliovirus internal ribosome entry site, *J. Virol.* **76**, 2113–2122.

Parsley, T.B., Cornell, C.T., Semler, B.L. (1999) Modulation of the RNA binding and protein processing activities of poliovirus polypeptide 3CD by the viral RNA polymerase domain, *J. Biol. Chem.* **274**, 12867–12876.

Parsley, T.B., Towner, J.S., Blyn, L.B., Ehrenfeld, E., Semler, B.L. (1997) Poly (rC) binding protein 2 forms a ternary complex with the 5'-terminal sequences of poliovirus RNA and the viral 3CD proteinase, *RNA* **3**, 1124–1134.

Paul, A.V., Rieder, E., Kim, D.W., van Boom, J.H., Wimmer, E. (2000) Identification of an RNA hairpin in poliovirus RNA that serves as the primary template in the in vitro uridylylation of VPg, *J. Virol.* **74**, 10359–10370.

Paul, A.V., van Boom, J.H., Filippov, D., Wimmer, E. (1998) Protein-primed RNA synthesis by purified poliovirus RNA polymerase, *Nature* **393**, 280–284.

Pelletier, J., Sonenberg, N. (1988) Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA, *Nature* **334**, 320–325.

Pestova, T.V., Hellen, C.U.T., Shatsky, I.N. (1996a) Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry, *Mol. Cell Biol.* **16**, 6859–6869.

Pestova, T.V., Shatsky, I.N., Hellen, C.U.T. (1996b) Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes, *Mol. Cell Biol.* **16**, 6870–6878.

Pfister, T., Wimmer, E. (1999) Characterization of the nucleoside triphosphatase activity of poliovirus protein 2C reveals a mechanism by which guanidine inhibits poliovirus replication, *J. Biol. Chem.* **274**, 6992–7001.

Piccone, M.E., Zellner, M., Kumosinski, T.F., Mason, P.W., Grubman, M.J. (1995) Identification of the active-site residues of the L proteinase of foot-and-mouth disease virus, *J. Virol.* **69**, 4950–4956.

Pincus, S.E., Diamond, D.C., Emini, E.A., Wimmer, E. (1986) Guanidine-selected mutants of poliovirus: mapping of point mutations to polypeptide 2C, *J. Virol.* **57**, 638–646.

Rieder, E., Paul, A.V., Kim, D.W., van Boom, J.H., Wimmer, E. (2000) Genetic and biochemical studies of poliovirus cis-acting replication element cre in relation to VPg uridylylation, *J. Virol.* **74**, 10371–10380.

Rinehart, J., Gomez, R.M., Roos, R.P. (1997) Molecular determinants for virulence in Coxsackievirus B1 infection, *J. Virol.* **71**, 3986–3991.

Rossmann, M.G., Arnold, E., Erickson, J.W., Frankenberger, E.A., Griffith, J.P., Hecht, H.J., Johnson, J.E., Kamer, G., Luo, M., Mosser, A.G. (1985) Structure of a human common cold virus and functional relationship to other picornaviruses, *Nature* **317**, 145–153.

Saleh, L., Rust, R.C., Fullkrug, R., Beck, E., Bassili, G., Ochs, K., Niepmann, M. (2001) Functional interaction of translation initiation factor eIF4G with the foot-and-mouth disease virus internal ribosome entry site, *J. Gen. Virol.* **82**, 757–763.

Takeda, N., Kuhn, R.J., Yang, C.F., Takegami, T., Wimmer, E. (1986) Initiation of poliovirus plus-strand RNA synthesis in a membrane complex of infected HeLa cells, *J. Virol.* **60**, 43–53.

Teterina, N.L., Bienz, K., Egger, D., Gorbalenya, A.E., Ehrenfeld, E. (1997) Induction of intracellular membrane rearrangements by HAV proteins 2C and 2BC, *Virology* **237**, 66–77.

Towner, J.S., Ho, T.V., Semler, B.L. (1996) Determinants of membrane association for poliovirus protein 3AB, *J. Biol. Chem.* **271**, 26810–26818.

Van Kuppeveld, F.J., Hoenderop, J.G., Smeets, R.L., Willems, P.H., Dijkman, H.B., Galama, J.M., Melchers, W.J. (1997) Coxsackie virus protein 2B modifies endoplasmic reticulum membrane and plasma membrane permeability and facilitates virus release, *EMBO J.* **16**, 3519–3532.

Vance, L.M., Moscufo, N., Chow, M., Heinz, B.A. (1997). Poliovirus 2C region functions during encapsidation of viral RNA, *J. Virol.* **71**, 8759–8765.

Walter, B.L., Nguyen, J.H., Ehrenfeld, E., Semler, B.L. (1999) Differential utilization of poly(rC) binding protein 2 in translation directed by picornavirus IRES elements, *RNA* **5**, 1570–1585.

Witherell, G.W., Schultz-Witherell, C.S., Wimmer, E. (1995) Cis-acting elements of the encephalomyocarditis virus internal ribosomal entry site, *Virology* **214**, 660–663.

Xiang, W., Cuconati, A., Hope, D., Kirkegaard, K., Wimmer, E. (1998) Complete protein linkage map of poliovirus P3 proteins: interaction of polymerase $3D^{pol}$ with VPg and with genetic variants of 3AB, *J. Virol.* **72**, 6732–6741.

Gene Mapping and Chromosome Evolution by Fluorescence–Activated Chromosome Sorting

Malcolm A. Ferguson-Smith
Cambridge University Centre for Veterinary Science, Cambridge, UK

1 **Introduction 121**

2 **Fluorescence-activated Chromosome Sorting (FACS) 122**

3 **Chromosome Painting 124**

4 **Chromosome Microdissection 125**

5 **Gene Mapping 126**

6 **Chromosome Painting in the Diagnosis of Chromosome Aberrations 126**

7 **Karyotype Evolution 129**

8 **Conclusions 132**

Acknowledgements 132

Bibliography 132
Books and Reviews 132
Primary Literature 133

Keywords

BAC (Bacterial Artificial Chromosome)
A vector used in DNA cloning that is capable of carrying DNA inserts measuring 100 to 200 kb.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

Chromosome-specific Paint Probe
The product of PCR amplification of sorted chromosomes, using random PCR primers. When labeled with fluorochromes coupled to nucleotides in the amplification procedure, the paint probe can be hybridized by fluorescence *in situ* hybridization (FISH) techniques to chromosome preparations on microscope slides. This reveals an even distribution of FISH signals along the chromosomes from which the paint was derived.

Contig
A contiguous series of cloned DNA sequences with overlapping ends.

Cross-species Reciprocal Chromosome Painting
A technique used to identify blocks of chromosome homology shared between species. A chromosome-specific paint probe from species A is hybridized to the chromosomes of species B. Paint probes from species B are hybridized back to the chromosomes of species A. This reveals the chromosomal identity of each homologous region in both species.

Cosmid
A vector used in DNA cloning that is capable of carrying DNA inserts measuring 20 to 40 kb.

FISH (Fluorescence in situ Hybridization)
A method whereby fluorescence–labeled DNA probes are hybridized to their complementary sequences on denatured chromosomes fixed on microscope slides.

Flow Karyotype
Graphic representation of mitotic chromosomes separated by a dual laser flow cytometer according to size and base-pair ratio.

Interchromosomal Rearrangement
A structural chromosome change involving exchange between two or more nonhomologous chromosomes.

Intrachromosomal Rearrangement
A structural change involving inversion or insertion of material within a chromosome.

Plasmid Vector
A vector that can be used for cloning small DNA inserts of 3 to 8 kb.

PCR (Polymerase Chain Reaction)
A method for the primer-directed amplification of DNA sequences.

The ability to sort and collect chromosomes by flow cytometry provides studies on gene mapping and chromosome evolution with a unique source of chromosome-specific DNA. Chromosome-specific gene libraries made from this DNA have yielded genetic markers used in constructing ordered chromosome maps. This was one of the key steps in the project which led to sequencing the human genome. PCR amplification of sorted chromosomal DNA has been used to prepare fluorescence-labeled chromosome-specific probes. When hybridized *in situ* to chromosome preparations, these probes "paint" the entire chromosome. The judicious use of different combinations of fluorochromes allows each chromosome pair in the cell to be given a different color. This procedure is helpful in the diagnosis of chromosome rearrangements present in malignant cells or associated with handicapping syndromes. Paint probes from one species can be used to identify homologous chromosome regions in other species. Comparison of the patterns of homology produced by reciprocal cross-species painting adds information on the evolutionary history of a species by distinguishing ancient chromosome rearrangements from those that have occurred more recently. The study of comparative genomics and karyotype evolution owes much to the development of these molecular cytogenetic techniques.

1 Introduction

The fluorescence-activated cell sorter was first used successfully in 1975 to separate chromosomes from fluid suspensions. It was soon found that the method allowed the collection of substantial quantities of chromosomes of one type without cross-contamination with other chromosomes. The chromosome-specific DNA so produced was used to prepare chromosome-specific libraries of DNA in plasmid and other vectors. In the early days of the human genome project, these plasmid libraries were of great importance in cloning chromosome-specific markers, which were used to map individual chromosomes. Linkage of polymorphic markers in family studies allowed the order of markers along the chromosome to be determined, and the detection of linkage between markers and cloned sequences from known genes similarly allowed the assignment of disease and other genes to their location on chromosomes. In the late 1980s, mapping by genetic linkage was supplemented by fluorescence *in situ* hybridization (FISH) mapping in which suitably labeled DNA markers were hybridized *in situ* to their complementary sequences on human chromosomes. With the development of BAC, cosmid, and other vectors for DNA cloning, it became possible to build up series of contiguous sequences of DNA (contigs) along each chromosome on the basis of this preliminary marker map. This approach, coupled with large-scale sequencing of contigs, has been one of the key strategies used in sequencing the human genome. Now that a draft sequence, albeit with gaps, of the entire human chromosome complement has been achieved, chromosome sorting has less application in mapping human genes. However, it still has application

in the building of preliminary maps of unmapped species.

Chromosome-specific DNA prepared from sorted chromosomes may also be used to prepare chromosome-specific probes using the polymerase chain reaction (PCR) with random DNA primers to amplify and label the probes with fluorescent dyes. The resulting labeled DNA is hybridized by FISH to air-dried metaphases on microscope slides in such a way that the respective chromosomes are "painted" with fluorescence along the entire length of the chromosome. The technology has advanced to the extent that each chromosome can be recognized by painting with different combinations of fluorochromes (multicolor FISH) in one hybridization experiment. The main application of M-FISH is in the diagnosis of interchromosomal aberrations that cause physical and mental childhood handicap and in the cytogenetic analysis of cancer (see chapter on FISH).

Chromosome painting has played an important part in characterizing regions of genetic homology between chromosomes of different species. The first studies in the early 1990s used chromosome-specific plasmid libraries, but more recent work uses chromosome-specific painting probes amplified from sorted chromosomes. The first comparisons were between humans and other primates. It was confirmed that the karyotype of humans differed from that of the great apes by a tandem fusion of two chromosomes, which formed human chromosome 2 and which occurred after the divergence of the human lineage from the common ancestor with the chimpanzee. Numerous inversions have also occurred during the evolution of the great apes, but the only other interchromosomal rearrangement found was a reciprocal chromosome translocation between human chromosome equivalents 5 and 17 in the divergence of the gorilla. Interchromosomal rearrangements between human and other species have been observed since, some involving whole chromosome arms and some involving smaller segments. The remarkable feature has been the high level of conservation between genomes. The range in the number of conserved (syntenic) autosomal blocks between human and the mammalian species studied to date varies from 23 to 200, the majority in the order of 40 to 50 and the highest in the house mouse. Some rearrangements are more ancient than others and so are shared by a wider range of species. The pattern of rearrangements in a species reflects the history of the genome evolution of that species, and the comparison of patterns between species provides information about their relationship and their place on the evolutionary tree.

2 Fluorescence-activated Chromosome Sorting (FACS)

Chromosomes are sorted using a dual laser flow cytometer equipped with two 5 W argon ion lasers. One laser is tuned to emit 300 mW of light in the UV (351–364 nm) to excite Hoechst fluorescence, and the second laser is tuned to emit 300 mW at 458 nm to excite chromomycin A3 fluorescence.

Chromosome preparations are made by standard methods from short-term blood cultures, lymphoblastoid cell lines, and a variety of tissue culture cells including fibroblasts from biopsy material. Colcemid (0.1 $\mu g\ mL^{-1}$) is added to the actively dividing cell culture to arrest cells in metaphase over several hours. The mitotic cells are collected and

resuspended in 75 mM KCl for 15 min and are transferred to a buffer containing polyamines and Triton X-100. Chromosomes are released into suspension by rapid vortexing for 10 s and are stained with chromomycin A3 (final concentration 40 μg mL^{-1}) and Hoechst 33258 (final concentration 2 μg mL^{-1}) followed by 2 h of incubation at 4 °C. Fifteen minutes prior to flow sorting, sodium citrate (final concentration 10 mM) and sodium sulfate (final concentration 25 mM) are added to the sample. Excessive disruption of the mitotic cells should be avoided as this leads to chromosome fragmentation indicated by an unacceptable level of debris at the lower end of the flow karyotype (Fig. 1). Cultures that contain a high proportion of cells in the anaphase/telophase may also lead to reduced resolution owing to contamination from the inclusion of anaphase chromosomes.

The chromosomes suspended in polyamine buffer, and surrounded by sheath fluid, flow under pressure through the two laser beams sequentially to permit the fluorescence signal emitted from each chromosome to be collected separately and stored in the computer. After the chromosomes have passed through the two lasers, the fluid stream breaks into a series of droplets, some of which will contain a single chromosome. Sorting is achieved by applying an electrical charge to the droplets

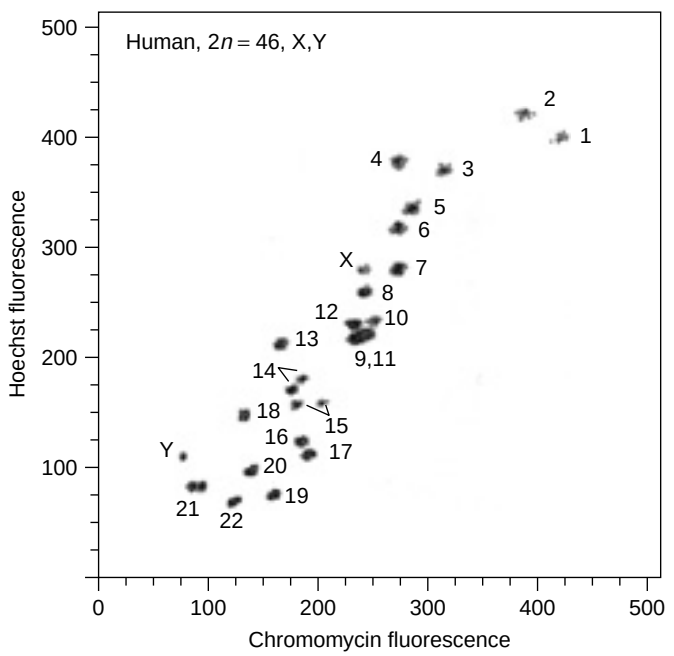

Fig. 1 Flow karyotype constructed from a fluid suspension of metaphase chromosomes from a normal male. The numbered peaks correspond to the numbers used in the classification of the human chromosome complement. Chromosomes are separated on the basis of DNA content and base-pair ratio. Note the separation of homologues of chromosomes 14, 15, and 21 due to heteromorphic blocks of repetitive DNA.

containing the chromosome of interest so that they can be deflected into a container as they pass between two high-voltage plates. Highly pure samples of two types of chromosome can be collected simultaneously. The fluorescence measurements from each chromosome are accumulated in large numbers in the computer and are used to construct a flow karyotype (Fig. 1), which reveals discrete clusters of signals, each cluster representing one or more chromosome types arranged in order of size and base-pair ratio. A–T–rich chromosomes sort above a diagonal line drawn through the middle of the chromosome clusters, while G–C–rich chromosomes sort below the line. The accumulation of signals can be observed on a video monitor and the chromosomes of interest selected for collection by a simple gating procedure. It takes less than 5 min to collect the 300 to 500 chromosomes required for PCR mapping and for the production of chromosome-specific paint probes. Larger samples of 2 to 3 million are required for preparing chromosome-specific DNA libraries, and this requires prolonged sorting.

3 Chromosome Painting

The DNA of 300 to 500 chromosomes sorted by FACS is amplified by degenerate oligonucleotide-primed PCR (DOP-PCR). After a primary round of amplification using the 6-MW primer (5′ CCGACT CGA GNN NNN NAT GTGG 3′) and unlabeled deoxynucleotide triphosphates, the DNA is subjected to a secondary round of amplification, this time incorporating a labeled nucleotide such as biotin-11-dUTP. The biotinylated probe is hybridized to denatured chromosomes in air-dried metaphase preparations using standard FISH methods and is detected using fluorochrome-conjugated streptavidin or anti-biotin antibodies. Many labeling techniques now use fluorescent dyes, such as Cy-dyes, coupled to the nucleotides used in the amplification procedure, and this avoids the need for an additional detection step.

The FISH signals are examined by fluorescence microscopy. An image processing system is helpful in analyzing preparations of low luminosity and in situations in which complex chromosome rearrangements are to be expected. Most digital fluorescence microscopy analysis systems employ a sensitive monochromatic CCD camera and excitation and emission filters appropriate to the fluorescence dye used as a label. When several paint probes, labeled with different combinations of dyes, are used, the gray-scale image of the fluorescence emitted by each fluorochrome is acquired sequentially and merged to produce a composite image in which each paint probe is assigned a separate false color on the computer monitor. Probe sets capable of assigning a different color combination to each pair of chromosomes in the complement are now available for human and mouse, using M-FISH and systems for classification and analysis of karyotypes with interchromosomal rearrangements (Fig. 2). The stored images acquired by these computerized systems are available for color adjustment, image enhancement, image reversal, and a number of other procedures, which assist in chromosome measurement and analysis. Spatial filters and contrast adjustment are particularly useful in enhancing the weak chromosome banding patterns obtained by DAPI counterstaining. Instead of the analysis of images of each individual

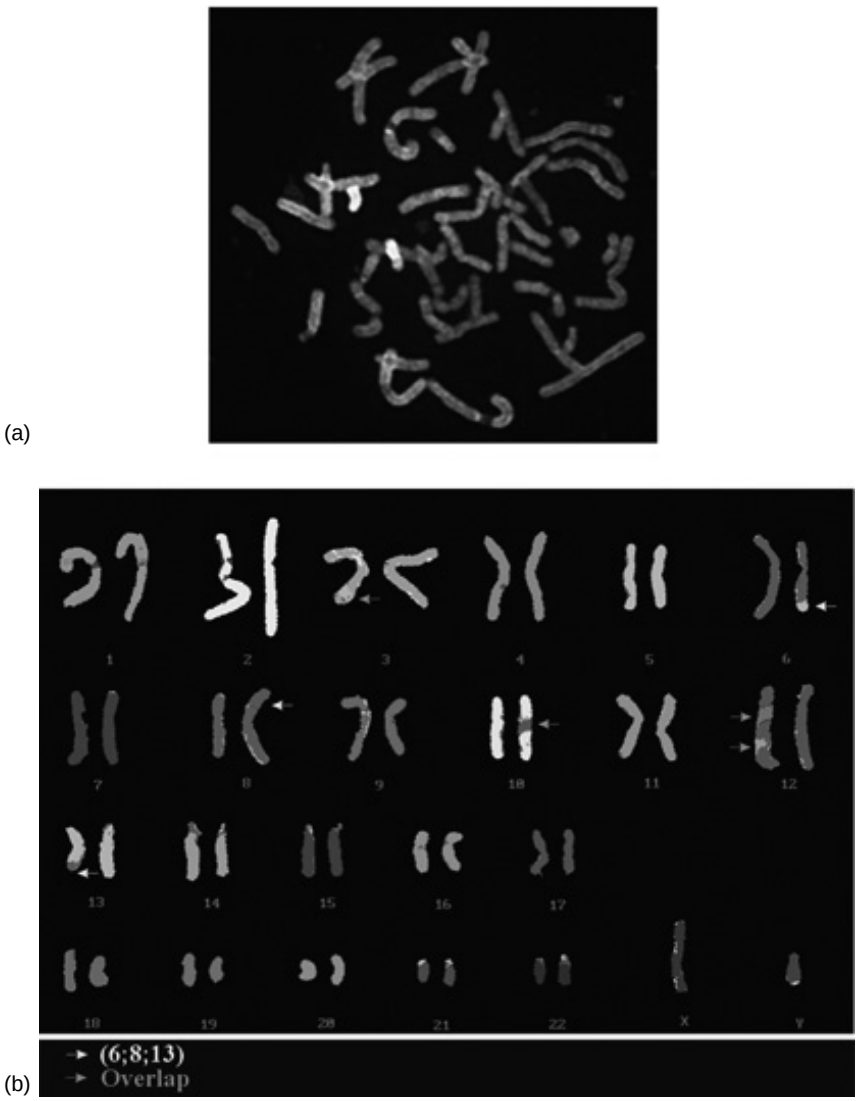

Fig. 2 Multicolor FISH. (a) Using a combination of five fluorochromes, each chromosome pair has a distinctive color. (b) This allows the identification of interchromosomal translocations between chromosomes 6, 8, and 13 in the karyotype. (See color plate p. xxii.)

color, the alternative spectral karyotyping (SKY) system analyzes the spectral characteristics of each pixel in one image using an interferometer. Both M-FISH and SKY provide similar resolution, each capable of detecting rearrangements as small as 2 Mb.

4 Chromosome Microdissection

Microdissection is an alternative method for preparing labeled chromosome-specific DNA. Glass microneedles are used under direct vision with an inverted microscope

and phase-contrast optics to scrape the chromosome of interest from air-dried chromosome preparations on a microscope slide. About 20 chromosomes are removed into a tube containing distilled water. DOP-PCR is used to make amplified primary products, which can then be used, in a further round of amplification with labeled nucleotides, to achieve the desired painting probe. The technique has been used with considerable success to generate single-chromosome-arm probes and region-specific probes. One application termed *multicolor banding* (MCB) is based on the microdissection of up to four to five regions along each chromosome. The resulting fragments are labeled with different combinations of fluorochromes so that hybridization to a metaphase produces a series of colored bands along the chromosome. This has been used to detect inversions and insertions in diagnostic cytogenetics. However, the technology is complex and is at present confined to the research laboratory.

5
Gene Mapping

Chromosome-specific DNA produced by flow sorting of chromosome suspensions can contribute to the assignment of genes to their respective chromosomes, provided that a DNA sequence from within the gene is known. In this method, PCR primers from both ends of the DNA sequence are used to amplify the gene in a panel made from primary amplification products of each chromosome, both in pools of five to six chromosomes and separately. It has been found that more than 85% of known genes can be mapped in this way. The method provides a rapid procedure for assigning linkage groups to the chromosomes of poorly mapped species, as has been shown recently in the construction of the genetic map of the dog.

Gene mapping in many species has been advanced by the construction of homology maps based on the results of reciprocal cross-species chromosome painting with chromosome-specific probes from the well-mapped human and mouse chromosomes (Fig. 3). While it is straightforward to assign linkage and radiation hybrid groups to their respective syntenic blocks in other species, their orientation and gene order has to be determined by FISH mapping on metaphase preparations with locus-specific probes (see chapter on FISH methods for the types of probe available). This is because intrachromosomal rearrangements, that is, inversions and insertions, are known to be relatively common events in species divergence. A higher level of resolution in obtaining gene order can be obtained from the longer chromosomes observed in interphase nuclei. However, the highest resolution by cytological methods is achieved using FISH on extended DNA fibers (Fig. 4). These fibers are produced by methods that release the DNA molecule from its associated proteins, especially histones, and allow spreading of the naked DNA molecules on microscope slides. Cloned DNA sequences of 1 kb or less can be hybridized successfully, and these can reveal such detail as the order and relative size of the exons and introns comprising a single gene. For further detail, DNA sequencing can be used, which is the ultimate level of resolution in gene maps.

6
Chromosome Painting in the Diagnosis of Chromosome Aberrations

Abnormal chromosomes can be sorted as well as normal ones provided they

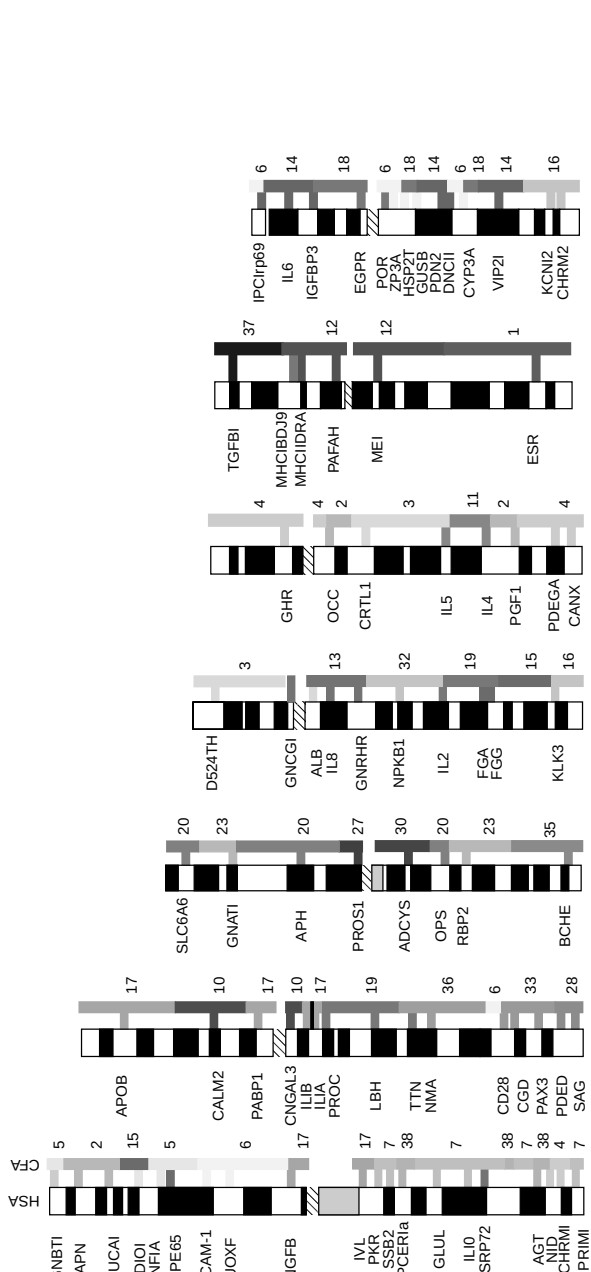

Fig. 3 Partial homology map based on reciprocal chromosome painting between domestic dog and human. The human karyotype is represented by a diagram of chromosomes 1 to 7 against which homology with segments of the dog chromosome complement are shown on the right, and selected gene markers are shown on the left.

Fig. 4 DNA fiber FISH showing order of three 130-kb BAC clones from the DMTR1 locus on human chromosome 9.

are separated from others in the flow karyotype. When paints are made from such aberrations and hybridized to normal metaphases (reverse painting), their origin from normal chromosomes can be determined rather precisely. Translocations involving two or more chromosomes are evident from the number of chromosome regions highlighted by the paint probe. Intrachromosomal duplications, which are among the most frequent aberrations, are evident from the observation that only one pair of chromosomes is painted. The duplicated region may be revealed by an increased dosage of label over the duplication.

Reverse painting can also be applied to probes prepared from the microdissection of abnormal chromosomes, especially small supernumerary chromosomes. However, the more usual approach is to use M-FISH or SKY systems in the analysis of complex aberrations. Where these are not available, single whole-chromosome paints (WCP) can be used in sequential rounds of denaturation and hybridization until the nature of the aberration is revealed.

As neither SKY nor M-FISH, nor WCP methods, can resolve inversions, color-banding methods have been developed to solve this problem. The published methods include bar code approaches using differentially labeled fragments of chromosome derived from radiation hybrids of human and rodent cells, each clone containing only a small number of human chromosome fragments. Different labels are used for each clone, so that the resulting paint probe set produces a series of color bands along each chromosome. The Harlequin approach (Cambio) uses M-FISH-labeled gibbon chromosomes, which when painted onto human chromosomes produce a color-banded appearance reflecting the 80 or more interchromosomal rearrangements that have occurred during the divergence of human and gibbon species. The MCB approach achieves a banding appearance by using several fragments removed by microdissection from each chromosome and distinguishing each fragment by a different combination of fluorochromes. All these methods can be used in the analysis of intrachromosomal rearrangements in both constitutional and neoplastic cytogenetics.

One of the aims of using multicolor painting probes in human and cancer cytogenetics is to identify the genes involved at the breakpoints of rearrangements. It has been found that these breakpoints may be associated with deletion or disruption of DNA followed by loss of gene function relevant to the phenotype. The strategy is first to identify the sites of the rearrangement and then to FISH-map the region using a series of BAC or other types of cloned DNA known, from the gene map, to be within the region of interest. Once

a clone that bridges the breakpoint has been found, subclones in plasmids can be used to narrow down the area to a point when DNA sequencing can be used to identify the pathological change. A more effective approach, based on comparative genome hybridization (CGH), has recently been introduced. Flow cytometry is used to isolate and collect the two derivative chromosomes involved in a rearrangement. Each is amplified, labelled with red and green respectively, and the mixture used to probe a microarray of genomic DNA reference BAC clones. Some 3000 reference clones (unlabeled), selected because they span the entire human genome at approximately 1-Mb intervals, are arrayed on glass slides for this type of analysis. The arrays are scanned to detect the ratio of red : green in the hybridization signals generated in each clone by the probe mixture. The information is examined in relation to the genetic map. This reveals the constitution of each derivative and the breakpoints of the rearrangement. A clone that gives an intermediate red : green ratio may be one that spans the breakpoint. The technique termed *array-CGH* has particular application in the identification of cancer genes and genes involved in dysmorphic and handicapping syndromes associated with constitutional chromosome aberrations. It is likely that the reference set of 3000 BAC clones will become an important part of the cytogeneticist's armamentarium in the future, both for array-CGH and for FISH mapping.

7
Karyotype Evolution

The extensive conservation of genetic information in the animal kingdom is evident from gene mapping studies, which show that the same groups of genes are often linked together in a wide range of species. In closely related species, this conservation often extends to the pattern of G-banding revealed by conventional karyotype analysis. The X-chromosome of placental mammals is a particularly striking example of conservation as it contains virtually all the same transcribed genes in every mammalian species. In marsupials, the homology with the human X is confined to the long arm of the human X. Human X short-arm genes are autosomal in marsupials. In birds, the Z-chromosome in all species studied has close homology with human chromosome 9, and the equivalent of the human X-linked genes are to be found mostly in two autosomes, for example, in chicken chromosomes 1 and 4.

The results from comparative G-banding studies are often difficult to interpret, and it has been found that cross-species chromosome painting gives much more accurate evidence of chromosome homology between species. As indicated in the Introduction, there are two main limitations of the painting method. The first is that intrachromosomal rearrangements cannot be detected unless interspersed with chromosomal exchanges (see Fig. 3). The second is that the efficiency of cross-species hybridization diminishes as the relationship between the two species becomes more distant. In general, reciprocal chromosome painting works well between placental mammals, but is difficult between mammals and marsupials and between mammals and birds. There are exceptions, and paints prepared from the tamar wallaby X hybridize, albeit weakly, with the long arm of the human X. In addition, paints prepared from human chromosome 4 have been reported to hybridize to chicken chromosome 4

in accordance with comparative mapping data; this result illustrates a remarkable degree of genomic conservation between species estimated to have diverged over 300 million years ago.

Chromosome rearrangements that have occurred during evolution can be used to study the degree of relationship between species and to speculate on the ancestral karyotype. When the same rearrangement is found in two species from separate families, but which is absent in all other species, it can be assumed that the shared rearrangement is a recent event in evolution. On the other hand, a rearrangement that is common to a wide range of divergent species (outgroup species) can be regarded as ancestral. It can usually be assumed that when large blocks of DNA homology identified by chromosome painting are retained intact in distantly related species, this arrangement represents the ancestral condition. When these blocks are divided between different chromosomes in a single species, or in a group of closely related species, the rearrangements responsible may be regarded as recent evolutionary events. Occasionally, an ancient rearrangement may be reconstituted by a more recent event, that is, there is a reversion to the ancestral state. Much can be learned about phylogenomic relationships and karyotype evolution from these chromosome-painting studies between closely related species and outgroup species. As a result, some conclusions can be drawn about the nature of the ancestral karyotype of all placental mammals.

A recent evolutionary event is the well-known fusion of ancestral chromosomes to form human chromosome 2. All extant great apes do not show this fusion, which must have arisen in the human lineage following divergence of the chimpanzee and human lineages from their common ancestor. In another example, which relates to the evolution of the primate chromosome destined to become the short arm of human chromosome 2, the gorilla and orangutan have a pericentric inversion, which is absent in both human and chimpanzee. At first it was thought that the arrangement in human and chimpanzee had occurred subsequent to the divergence of the gorilla and orangutan and therefore served to link chimpanzee and human. It was later found that the inversion was absent in Old World monkeys (outgroup species), indicating that the human arrangement in this case is the ancestral one. The inversion in the gorilla and orangutan must have occurred subsequent to the divergence of humans and chimpanzees from the lineages, leading to the other great apes.

Chromosome painting with human probes reveals a number of ancient syntenic blocks in outgroup species that are split between different human chromosomes (Table 1). For example, the homologues of human chromosomes 14 and 15 are fused in one or two blocks in most mammalian species with the exception of the apes, dog, and giant panda. In the latter two cases, independent fission events have occurred as comparatively recent evolutionary events. Other ancient syntenies are found to characterize orders and families and, sometimes, small groups of species. Their discovery can be helpful in assigning a species to its place in the phylogenetic tree. For example, it has been shown by comparative painting with dog paints that the common ancestor of the domestic dog, red fox, Arctic fox, and raccoon dog had a small chromosome number composed of 38 autosomal segments and that a number of specific

Tab. 1 Ancient syntenies revealed by the association of chromosome segments homologous to human chromosomes in the genomes of various mammals. Chromosome numbers refer to individual human chromosomes.

Species	***Ancient syntenies***								
	14/15	***3/21***	***12/22***	***16/19***	***4/8***	***7/16***	***3/19***	***1/2***	***5/19***
Great apes	−	−	−	−	−	−	−	−	−
Lesser apes	−	−	−	−	−	−	−	−	−
Old World monkeys	+	−	−	−	−	−	−	−	−
New World monkeys	±	±	−	−	−	−	−	−	−
Lemurs	+	+	+	−	−	−	−	−	−
Cattle	+	+	+	+	+	+	−	+	+
Sheep	+	+	+	+	+	+	−	+	+
Muntjac	+	+	+	+	−	+	−	+	+
Pig	+	+	+	+	−	−	−	−	+
Horse	+	+	+	+	+	+	−	−	−
Zebra	+	+	+	+	+	+	−	+	−
Dolphin	+	+	+	+	−	−	+	−	+
Seal	+	+	+	+	+	−	+	−	−
Cat	+	+	+	+	+	+	+	+	−
Mink	+	+	+	+	+	+	+	−	−
Dog	−	+	+	−	+	+	+	+	−
Panda	−	+	+	+	+	−	+	+	−
Rabbit	+	+	+	+	+	+	−	−	−
Shrew	+	+	−	+	−	−	−	−	−

interchromosomal rearrangements distinguish the karyotypes of each species. As the dog has a chromosome number of 78, the change from low numbers has been achieved by a series of fission events, with the formation of new centromeres. Chromosome fusion, on the other hand, has been the mechanism by which the three pairs of chromosomes of the Indian muntjac ($2n = 6, 7$) have evolved from an ancestral deer karyotype of $2n = 70$. This is illustrated by painting chromosome-specific probes of the Chinese muntjac ($2n = 46$) onto Indian muntjac chromosomes. Each Indian muntjac chromosome can be shown to be formed by the fusion of several apparently complete Chinese muntjac chromosomes (Fig. 5). Remnants of ancestral centromeres can be demonstrated at most of the fusion sites, indicating frequent centromere–telomere events, although this remains to be confirmed by molecular analysis.

Comparative studies of the genomes of mouse and human at the DNA sequence level now show that evolutionary breakpoints tend to occur within clustered gene families. It seems that the most likely mechanism is meiotic recombination between homologous sequences on nonhomologous chromosomes or chromosome segments, that is, a similar mechanism that occurs in the formation of pathological rearrangements both in somatic and in germ cells. Speciation can now be seen as resulting largely from the gain or loss of sequences within gene families, leading to the degeneration of some genes of known function and the generation of new genes of different function. It is believed that duplication events are the main source of evolutionary novelty.

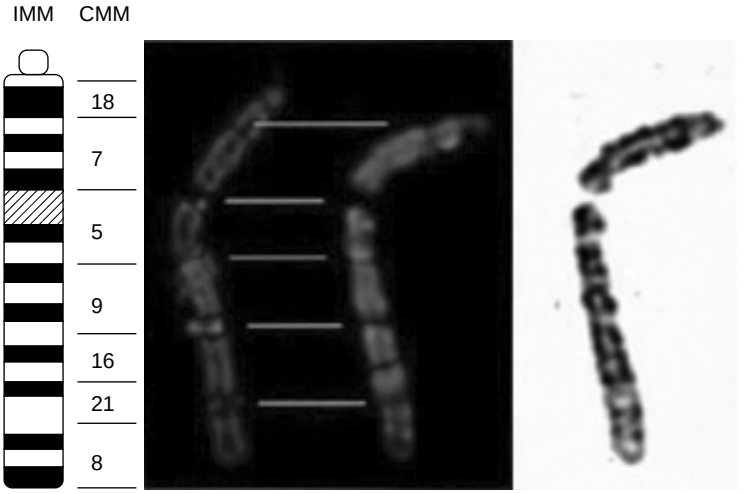

Fig. 5 Indian muntjac (IMM) chromosome 3 to which chromosome-specific paint probes of the Chinese muntjac (CMM) have been hybridized. This reveals that this chromosome has been formed by the tandem fusion of seven entire chromosomes of the Chinese muntjac. A centromere probe (red) indicates that remnants of ancestral centromeres are still present at the sites of fusion. (See color plate p. xxiii.)

8 Conclusions

The utility of chromosome painting extends from diagnostic cytogenetics to its application in comparative gene mapping and to studies of karyotype evolution. The development of highly effective techniques for sorting pure samples of chromosomes in fluid suspension has played a central role in theses advances. Future understanding of genome functions relating to gene regulation, chromosome structure and behavior, genome evolution, and the significance of noncoding conserved DNA will depend on the construction of comparative maps in a wide range of species. Only by such comparisons will it be possible to determine which parts of the genome are conserved in evolution because they retain vital functions. In these comparisons, chromosome sorting can be expected to continue to play a useful role.

Acknowledgements

The author is grateful to Dr Fengtang Yang, Dr Willem Rens, Dr Fumio Kasai, Mrs Patricia O'Brien, and other colleagues at the Cambridge Resource Centre for Comparative Genomics for the illustrations and for useful discussions in the preparation of this review.

See also Body Expression Map of Human Genome; Gene Distribution in the Human Genome; Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Dutrillaux, B. (1986) Evolution chromosomique chez les primates, les carnivores, et les rongeurs, *Mammalia* **50**, 1–203.

Ferguson-Smith, M.A. (1997) Genetic analysis by chromosome sorting and painting: phylogenetic and diagnostic applications, *Eur. J. Hum. Genet.* **5**, 253–265.

Ferguson-Smith, M.A., Smith, K. (2002) Cytogenetic analysis. Chapter 25, in: Rimoin, D.L., Connor, J.M., Pyeritz, R.E., Korf, B.R. (Eds.) *Emery & Rimoin's Principles and Practice of Medical Genetics*, Fourth Edition, Churchill Livingstone, London, pp. 690–722.

Gray, J.W., Carrano, A.V., Steinmetz, L.L., Van Dilla, M.A., Moore, D.H., Mayall, B.H., Mendelsohn, M.L. (1975) Chromosome measurement and sorting by flow systems, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1231–1234.

Meltzer, P.S., Guan, X.Y., Burgess, A., Trent, J.M. (1992) Rapid generation of region specific probes by chromosome microdissection and their application, *Nat. Genet.* **1**, 24–28.

Murphy, W.J., Stanyon, R., O'Brien, S. (2001) Evolution of mammalian genome organisation inferred from comparative gene mapping, *Gen. Biol.* **2**, 1–8.

Scherthan, H., Cremer, T., Arnason, U., Weier, H.-U., Lima-de-Faria, A., Fronicke, L. (1994) Comparative chromosome painting discloses homologous segments in distantly related mammals, *Nature Genetics* **6**, 342–347.

Schröck, E., du Manoir, S., Veldman, T., Schoell, B., Wienberg, J., Ferguson-Smith, M.A., Ning, Y., Lebetter, D., Bar-Am, I., Soenksen, D., Garini, Y., Ried, T. (1996) Multicolor spectral karyotyping of human chromosomes, *Science* **273**, 494–497.

Speicher, M.R., Ballard, S.G., Ward, D.C. (1996) Karyotyping human chromosomes by combinatorial multi-fluor FISH, *Nat. Genet.* **12**, 368–375.

Trask, B.J. (2002) Human cytogenetics: 46 chromosomes, 46 years and counting, *Nature Reviews, Genetics* **3**, 769–778.

Van Dilla, M.A., Deaven, L.L., Albright, K.L., Allen, N.A., Aubuchon, M., Bartoldi, M., Brown, N., Campbell, A., Carrano, A., Clark, L., Cram, L., Crawford, B., Fuscoe, J., Gray, J., Hildebrand, E., Jackson, P., Jett, J., Longmire, J., Lozes, C., Luedemann, M., Martin, J., McNinch, J., Meincke, L., Mendelson, M., Meyne, J., Moyzis, R., Munk, A., Perlman, J., Peters, D., Silvam, A., Trask, B. (1986) Human chromosome-specific DNA libraries: construction and availability, *Biotechnology* **4**, 537–552.

Wienberg, J., Stanyon, R. (1998) Comparative chromosome painting of primate genomes, *ILAR J.* **39**, 77–91.

Primary Literature

Carter, N.P., Ferguson-Smith, M.A., Perryman, M.T., Telenius, H., Pelmear, A.H., Leversha, M.A., Glancy, M.T., Wood, S.L.,, Cook, K., Dyson, H.M., Ferguson-Smith, M.E., Willatt, L.R. (1992) Reverse chromosome painting: a method for the rapid analysis of aberrant chromosomes in clinical cytogenetics, *J. Med. Genet.* **29**, 299–307.

Chowdhary, B.P., Raudsepp, T. (2000) HSA 4 and GGA 4: remarkable conservations despite 300 Myr divergence, *Genomics* **64**, 102–105.

Chudoba, I., Plesch, A., Lorch, T., Lemke, J., Claussen, U., Senger, G. (1999) High resolution multicolour-banding: a new technique for refined FISH analysis of human chromosomes, *Cytogenet. Cell Genet.* **84**, 156–160.

Davies, K.E., Young, B.D., Elles, R.G., Hill, M.E., Williamson, R.W. (1981) Cloning of a representative genomic library of the human X chromosome after sorting by flow cytometry, *Nature* **293**, 374–376.

Dehal, P., Predki, P., Olsen, A.S., Kobayashi, A., Folta, P., Lucas, S., Land, M., Terry, A., Zhou, C.L.E., Rash, S., Zhang, Q., Gordon, L., Kim, J., Elkin, C., Pollard, M.J., Richardson, P., Rokhsar, D., Uberbacher, E., Hawkins, T., Branscomb, E., Stubbs, L. (2001) Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution, *Science* **293**, 104–111.

Dutrillaux, B. (1979) Chromosomal evolution in primates: tentative phylogeny from Microcebus murinus (Prosimian) to man, *Human Genetics* **48**, 251–314.

Fiegler, H., Carr, P., Douglas, E.J., Burford, D.C., Hunt, S., Smith, J., Vetrie, D., Gorman, P., Tomlinson, I.P.M., Carter, N.P. (2003) DNA microarrays for comparative genomic hybridisation based on DOP-PCR amplification of BAC and PAC clones, *Genes, Chromosomes & Cancer* **36**, 361–374.

Florijn, R.J., Blonden, L.A.J., Vrolijk, J., Wiegant, J., Vaandrager, J.W., Baas, F., Den Dunnen, J.T., Tanke, H.J., Van Ommen, G.J.B.,

Raap, A.K. (1995) High resolution DNA fibre-FISH genomic DNA mapping and colour bar coding of large genes, *Hum. Mol. Genet.* **4**, 831–836.

Glas, R., Graves, J.A.M., Toder, R., Ferguson-Smith, M.A., O'Brien, P.C.M. (1999) Cross-species chromosome painting between human and marsupial demonstrates the ancient region of the mammalian X, *Mamm. Genome* **10**, 1115, 1116.

Graphodatsky, A.S., Yang, F., O'Brien, P.C.M., Perelman, P., Milne, B.S., Serdukova, N., Kawada, S.I., Ferguson-Smith, M.A. (2001) Phylogenetic implications of the 38 putative ancestral chromosome segments for four canid species, *Cytogenet. Cell Genet.* **92**, 243–247.

Graves, J.A.M., Watson, J.M. (1991) Mammalian sex chromosomes: evolution of organisation and function, *Chromosoma* **101**, 63–68.

Griffin, D.K., Sanoudou, D., Adamski, E., Mcgiffert, C., O'Brien, P., Wienberg, J., Ferguson-Smith, M.A. (1998) Chromosome specific comparative genome hybridisation for determining the origin of intrachromosomal duplications, *J. Med. Genet.* **35**, 37–41.

Harris, P., Boyd, E., Ferguson-Smith, M.A. (1985) Optimising human chromosome separation for the production of chromosome-specific DNA libraries by flow sorting, *Human Genetics* **70**, 59–65.

Jauch, A., Wienberg, J., Stanyon, R., Arnold, N., Tofanelli, S., Ishida, T., Cremer, T. (1992) Reconstruction of genomic rearrangements in great apes and gibbons by chromosome painting, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8611–8615.

John, H., Birnstiel, M., Jones, K. (1969) RNA : DNA hybrids at the cytological level, *Nature* **223**, 582–587.

Lichter, P., Cremer, C., Borden, J., Manuelidis, L., Ward, D.C. (1988) Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridisation using recombinant DNA libraries, *Human Genetics* **80**, 224–234.

Lichter, P., Tang, C.C., Call, K., Hermanson, G., Evans, G.A., Housman, D., Ward, D.C. (1990) High resolution mapping of human chromosome 11 by in situ hybridisation with cosmid clones, *Science* **247**, 64–69.

Mann, S.M., Burkin, D.J., Griffin, D.K., Ferguson-Smith, M.A. (1997) A fast, novel approach for DNA fibre fluorescence *in situ* hybridisation analysis, *Chromosome Res.* **5**, 145–147.

Muller, S., O'Brien, P.C.M., Ferguson-Smith, M.A., Wienberg, J. (1998) Cross species colour segmenting: a novel tool in human karyotype analysis, *Cytometry* **33**, 445–452.

Muller, S., Rocchi, M., Ferguson-Smith, M.A., Wienberg, J. (1997) Towards a multicolor chromosome bar code for the entire human karyotype by fluorescence *in situ* hybridisation, *Hum. Genet.* **100**, 271–278.

Pinkel, D., Straume, T., Gray, J.W. (1986) Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridisation, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 2934–2938.

Pinkel, D., Landegent, J., Collins, C., Fuscoe, J., Seagraves, R., Lucas, J., Gray, J. (1988) Fluorescence in situ hybridisation with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4, *Proc. Natl. Acad. Sci.* **85**, 9138–9142.

Rens, W., Yang, F., O'Brien, P.C.M., Solanky, N., Ferguson-Smith, M.A. (2001) A classification efficiency test of SKY and MFISH: identification of chromosomal homologies between *Homo sapiens* and *Hylobates leucogenys*, *Genes, Chromosomes Cancer* **31**, 65–74.

Ried, T., Baldini, A., Rand, T.C., Ward, D.C. (1992) Simultaneous visualisation of seven different DNA probes by in situ hybridisation using combinatorial fluorescence and digital imaging microscopy, *Proc. Natl. Acad. Sci.* **89**, 1388–1392.

Sargan, D.R., Yang, F., Squire, M., Milne, B.S., O'Brien, P.C.M., Ferguson-Smith, M.A. (2000) Use of flow-sorted canine chromosomes in the assignment of canine linkage, radiation, hybrid, and syntenic groups to chromosomes: refinement and verification of the comparative chromosome map for dog and human, *Genomics* **69**, 182–195.

Schmid, M., Nanda, I., Guttenbach, M., Steinlein, C., Hoehn, H., Schartl, M., Haaf, T., et al. (2000) First report on chicken genes and chromosomes 2000, *Cytogenet. Cell Genet.* **90**, 169–218.

Sillar, R., Young, B.D. (1981) A new method for the preparation of metaphase chromosomes for flow analysis, *J. Histochem. Cytochem.* **29**, 74–78.

Telenius, H., Pelmear, A.H., Tunnacliffe, A., Carter, N.P., Behmel, A., Ferguson-Smith,

M.A., Nordenskjold, M., Peragner, R., Ponder, B.A.J. (1992) Cytogenetic analysis by chromosome painting using DOP-PCR amplified flow-sorted chromosomes, *Genes, Chromosomes Cancer* **4**, 257–263.

Yang, F., Alkalaeva, E.Z., Perelman, P.L., Pardini, A.T., Harrison, W.R., O'Brien, P.C.M., Fu, B., Graphodatsky, A.S., Ferguson-Smith, M.A., Robinson, J.J. (2003) Reciprocal chromosome painting among human, aardvark, and elephant (superorder Afrotheria) reveals the likely eutherian ancestral karyotype, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1062–1066.

Yang, F., O'Brien, P.C.M., Milne, B.S., Graphodatsky, A.S., Solanky, N., Trifonov, V., Rens, W., Sargan, D., Ferguson-Smith, M.A. (1999) A complete comparative chromosomal map of the dog, red fox and human and its integration with canine genetic maps, *Genomics* **62**, 189–202.

Yang, F., O'Brien, P.C.M., Wienberg, J., Ferguson-Smith, M.A. (1997) A reappraisal of the tandem fusion theory of karyotype evolution in the Indian muntjak using chromosome painting, *Chromosome Res.* **5**, 109–117.

Gene Mapping by Fluorescence In Situ Hybridization

Barbara G. Beatty[1] *and Henry H.Q. Heng*[2]
[1] *Department of Pathology, University of Vermont College of Medicine, Burlington, VT, USA*
[2] *Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, Detroit, MI, USA*

1 Introduction 140

2 Probes for Gene Mapping by FISH 141
2.1 Identification of Probes from Genomic Databases 141
2.2 Probes for FISH Mapping 142
2.2.1 Genomic DNA Probes 142
2.2.2 Unique Sequence cDNA and EST Probes 143
2.2.3 Synthetic Oligonucleotides 143
2.3 Probe Labeling 144
2.3.1 Purification 144
2.3.2 Labeling Methods 144
2.3.3 Postlabeling Purification and Storage 146

3 DNA Targets for FISH Mapping 146
3.1 Metaphase Chromosomes 146
3.1.1 Metaphase Chromosome Preparation 147
3.2 Interphase Nuclei 148
3.2.1 Interphase Nuclei Preparation 149
3.3 Chromatin and DNA Fibers 149
3.3.1 Fiber FISH Target Preparation 150

4 Denaturation, Hybridization, and Detection 153
4.1 Pretreatment and Denaturation 153
4.2 Hybridization 154
4.2.1 Stringency Washes 154

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

4.3 Detection 154
4.3.1 Counterstaining and Banding 155

5 Microscopy and Image Analysis 156

6 Applications 157
6.1 Metaphase Mapping 160
6.1.1 Single-copy Genes 160
6.1.2 Segmental Duplications and FISH Mapping 161
6.1.3 Relational Mapping 161
6.1.4 FISH Mapping Combined with CGH and spectral karyotyping (SKY) 164
6.1.5 Comparative Mapping by ZOO-FISH 164
6.2 Interphase FISH Mapping 165
6.3 Chromatin and DNA Fiber Mapping 165

Bibliography 166
Books and Reviews 166
Primary Literature 166

Keywords

BAC
Bacterial artificial chromosome vector that is used to clone large genomic DNA inserts of 100 to 300 kb.

Chromosomal Bands
Light and dark staining patterns along the length of the chromosome that are unique for each chromosome and are produced by differential binding of dyes.

Chromatin Fibers
Strands of complexed DNA, and proteins (histones) organized into compact nucleosomes that are disk-like structures 10 nm in diameter and comprise 140 to 146 base pairs of DNA wound around histone octomers.

Released Chromatin Fibers
Linearized 30-nm interphase chromatin fibers experimentally released from interphase nuclei.

Released DNA Fiber
Linearized DNA strands experimentally released from nuclei or chromosomes serving as target for high-resolution FISH mapping.

FISH
Fluorescence *in situ* hybridization is the technique of localizing fluorescent nucleic acid probes (DNA, RNA) to complementary sequences of target DNA on a glass slide and visualizing them in a fluorescence microscope.

Interphase Nuclei
Nuclei that are between one mitosis and the next and contain decondensed chromatin.

Metaphase Chromosome
Fully condensed chromosome present at the mitotic stage (metaphase) in which chromosomes are attached to the mitotic spindle but have not yet segregated to the opposite spindle poles.

PAC
P1-derived artificial chromosome vector used to clone large genomic DNA inserts of 130 to 150 kb.

YAC
Yeast artificial chromosome vector used to clone large genomic DNA inserts of a few hundred kb to 1 Mb.

Gene mapping by fluorescence *in situ* hybridization (FISH) is the localization of a unique DNA sequence (probe) to a specific position (chromosomal band) within the genome (target). The DNA probe localization on the target is visualized using a fluorescent reporter molecule directly or indirectly bound to the probe. Although gene mapping by FISH usually refers to the localization of a gene or DNA sequence to a specific metaphase chromosome band, it can also include the relative position of genes along a chromatin or DNA fiber, or the identification of specific genomic BAC/PAC clones present in an array. FISH combines the specificity of recombinant DNA technology with basic cytogenetic techniques for target DNA preparation.

Valuable information regarding the biological and/or clinical significance of a gene can be obtained by identifying its subchromosomal location in both normal and disease conditions. Mapping a gene to certain critical chromosomal regions that may be deleted, amplified, or involved in translocation breakpoints can provide important information about the role of that gene in disease processes and may have significant clinical applications. Important information regarding gene function can also be obtained from mapping genes from individual tissues or disease states, identifying transgene insertion sites, studying interspecies synteny, and determining the order of and distance between two or more genes within a chromosome band.

1 Introduction

Fluorescence *in situ* hybridization (FISH) combines the chromosome preparation and banding techniques of classical cytogenetics with recombinant DNA technology. Specific DNA sequences (probes) when denatured will form a heteroduplex with denatured DNA on a glass microscope slide (target), and the resulting hybridization is visualized by a fluorescent reporter molecule (Fig. 1). The use of fluorescence over radioactivity as a mode of detection became feasible in the early 1980s when methods for labeling nucleic acids with haptens such as biotin were developed. This indirect approach to visualizing hybridized probe requires further detection steps with fluorescently labeled reporter molecules such as avidin or specific antibodies. Direct incorporation of fluorochromes into nucleic acids has now greatly simplified detection.

The first gene was mapped to human metaphase chromosomes using FISH in the mid-1980s, and now over 5000 articles utilizing gene mapping by FISH have been published. FISH is one of the few mapping techniques that does not require clones to overlap, to be positioned and ordered, and thus has broad applications in positional cloning studies. FISH now plays an important role in confirming *in silico* mapping information obtained through the human genome project, which has produced genome sequence databases for both human and other species.

FISH can localize a single gene to a subchromosomal band, can allow ordering of several genes within a chromosome band, and can aid in determining locus copy number, probe chimerism (YACs), pseudogenes, and location of gene family members. FISH mapping resolution is dependent on the level of chromatin condensation present in the target DNA and covers a wide range from high-resolution DNA fiber mapping (1–5 kb) to metaphase chromosome mapping at a resolution of 1 to 2 Mb, making it an exceptionally versatile mapping tool.

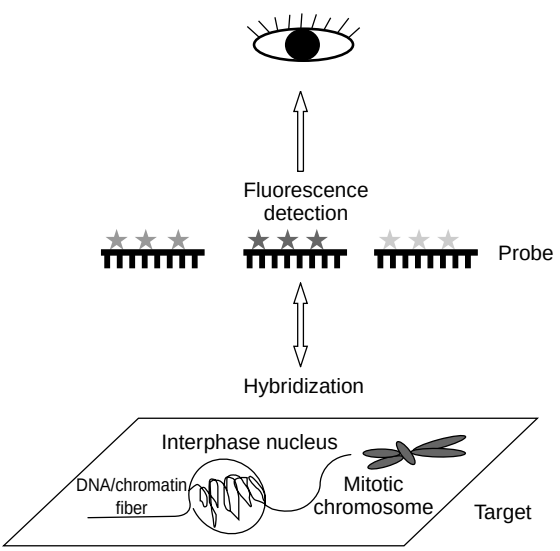

Fig. 1 Schematic representation of FISH. Basic FISH protocols consist of (a) preparation of target DNA in the form of metaphase chromosomes, interphase nuclei, or DNA/chromatin fibers on a slide; (b) labeling of probes directly with a fluor or indirectly with a hapten followed by a secondary fluor; (c) hybridization of the probe to the target; and (d) detection of the fluorescent hybridized probe under an epifluorescence microscope. (We acknowledge Dr Christine Ye and Miss Wei Lu for their help with the diagram.)

2
Probes for Gene Mapping by FISH

2.1
Identification of Probes from Genomic Databases

The human genome project has provided a wealth of data regarding DNA sequences suitable for gene mapping by FISH. Several databases containing information about expressed sequence tags (ESTs) and genes that have been sequenced or partially sequenced are available publicly on the World Wide Web. It may be necessary to search several websites in order to determine the relationship and ordering of these clones as no single database currently contains all this information. Although the quality of the clones is variable, use of these databases and *in silico* methods provide the best initial approach to identifying or localizing a particular DNA sequence to a chromosomal band or region. Validation of the localization can then be performed by FISH using a corresponding genomic DNA probe. Libraries containing bacterial artificial chromosome (BAC), P1 artificial chromosome (PAC), yeast artificial chromosome (YAC), and cosmid clones are now available for most of the human genome, and confirmation of the physical chromosomal location of these DNA sequences by FISH mapping is extremely useful.

There are several approaches to identifying clones (probes) suitable for human FISH mapping on the web. The first is to use the cytogenetic location. One can search for premapped FISH probes using the Genome Database http://www.gdb.org/hugo/ under "Chromosome Resources" and clones based on cytogenetic position can be obtained using one of the three following sites:

Genome database site: http://www.gdb.org/gdb/regionSearch.html,
GeneCards: http://bioinfo.weizmann.ac.il/cards/index.html, or
Genetic Location Database: http://cedar.genetics.soton.ac.uk/public_html/gmap.html.

A second approach is to obtain BAC, PAC, and cosmid clones specific for a particular gene or DNA marker by searching either the Genome Database (http://www.gdb.org) or the Locus link (http://www.ncbi.nlm.nih.gov/LocusLink/). In each case, the gene or marker name can be used to link to a DNA sequence database that can be screened to identify the clone name. For YACs, a similar approach is used at the Whitehead/MIT website (http://carbon.wi.mit.edu.8000/cgi-bin/conyig/phys_map) to identify both the clone and the contig containing overlapping clones along the chromosome.

To search for a corresponding genomic clone by gene or marker sequence, the sequence can be retrieved from a text search of http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=Nucleotide. The accession number or DNA sequence can then be copied and pasted into http://www.ncbi.nih.gov/blast/blast.cgi and databases NR (finished sequences), HTGS (rough draft sequence), or GSS (overlapping BAC ends) screened for those containing the corresponding genomic clones. If identification of nearby or overlapping clones is required, this can be achieved on the basis of knowing the BAC clone (1) name or sequence (http://www.TIGR.org/tdb/humgen/bac_end_search/bac_end_intro.html), (2) fingerprint patterns (http://genome.wustl.edu/gsc/human/human_database.shtml), or (3) DNA sequence alignment [(Ensembl (http://www.ensembl.org/), NCBI (http://

www.ncbi.nlm.nih.gov/cgi-bin/Entrez/hum_srch?chr=hum_chr.inf), or UCSF (http://genome.ucsc.edu/)].

Once the specific clones or sequences have been identified, there are several sources from which they can be obtained and the websites are listed as follows:

Cosmids, BACs, PACs, and YACS

- The Centre for Applied Genomics (http://tcag.bioinfo.sickkids.on.ca/) – Canada
- DHGP (http://www.rzpd.de/) – Germany
- HGMP Resource Centre (http://www.hgmp.mrc.ac.uk/) – UK

BACs, PACs, and YACs

- TIGEM (http://www.spr.it/iger/home.html) – Italy
- Research Genetics (http://www.resgen.com/) – USA
- Genome Systems (http://reagents.incyte.com/) – USA

BACs and PACs

- Sanger Center (http://www.sanger.ac.uk/HGP/) – UK
- Pieter de Jonge's lab (http://www.chori.org/bacpac/) – USA

Genome databases are also available on the web for other species such as: http://bos.cvm.tamu.edu/bovgbase.html (cattle), http://ratmap.gen.gu.se/DownloadGeneData.html (rat), and http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene (EST data for mouse, cow, zebra fish, clawed frog, fruit fly, and mosquito as well as wheat, barley, maize, and cress) as sources for genomic clones for FISH mapping.

2.2 Probes for FISH Mapping

The probes for FISH range from chemically synthesized oligonucleotides to total genomic DNA. Typically, DNA sequences used for FISH mapping should be of such a size and nature as to be readily visualized by a fluorescence reporter molecule under a 60 or 100× objective of an epifluorescence microscope. The two most common single-copy probes used in mapping experiments are cloned genomic sequences (YACs, BACs, PACs, cosmids, and plasmids) and enzymatically amplified unique sequences of cDNAs and expressed sequence tags. In all cases, the probe purity, type, and size are critical considerations for optimal mapping results. The degree of probe purity required is dependent on the labeling method to be used. The type and size of probe affects the labeling approach, prehybridization treatment, and detection method.

2.2.1 Genomic DNA Probes

Cloned single-copy genomic probes range in size from ~5 kb (plasmids) to ~1 Mb (YACs) and their signal strength is usually directly proportional to their size. Genomic probes can be labeled by nick translation or PCR and are visualized by direct or indirect labeling (see Sect. 2.3). Since genomic probes consist of contiguous genome sequences, they contain both unique sequences specific for the gene or marker, and commonly repeated sequences (SINEs/alu, LINEs) interspersed throughout the genome. In order to visualize the chromosomal location of the unique genomic sequences, with minimal nonspecific hybridization background, the common repeat sequences must be suppressed or blocked prior to hybridization.

This is achieved by preannealing the probe DNA with competitor total DNA or Cot-1 DNA of the same species. Cot-1 DNA is genomic DNA enriched in the commonly repeated genome sequences that is commercially available but only for human or mouse. This suppression hybridization approach is essential for mapping genomic probes to their unique location on target chromosomes.

Among the single-copy genomic probes, BACs and PACs are used most frequently for FISH mapping due to the availability of BAC and PAC libraries, their stability, and appropriate insert size (100 to 200 kb) for easy visualization. YACs must be checked for chimerism prior to use as they tend to be unstable and have a high rate of noncontiguous coligated insert sequences. Consequently, caution must be taken when using YACs for mapping purposes.

Other genomic probes important in FISH applications in molecular diagnostics, human cytogenetics, and genomics that also have mapping related functions include total genomic DNA used in comparative genomic hybridization (CGH) and chromosome painting probes. The application of these probes and their role in genome mapping will be discussed in Sect. 6. Satellite tandem repeat probes recognizing chromosome-specific tandem repeat sequences in the chromosomal centromeric and telomeric regions can be used in conventional and quantitative FISH mapping studies to identify specific chromosomes and measure telomeric repeats.

2.2.2 Unique Sequence cDNA and EST Probes

Unique sequence probes are smaller (<5 kb) than genomic probes and contain only expressed exonic sequences. These probes do not require blocking of common repeat sequences by Cot-1 DNA. As these probes contain noncontiguous genomic DNA sequences, their signal strength is related to the size and structure of the gene target and thus can be unpredictable. The transcribed (expressed) sequences within a gene may comprise a number of large contiguous exons or a number of short, widely separated exons. The former configuration would have a higher likelihood of producing a visible signal compared to that of the latter for the same sized cDNA probe. Thus, both the size and organization of the gene can determine the mapping results and should be tested on an individual basis. In general, cDNA or EST probes are best visualized by indirect labeling techniques and tyramide amplification detection methods (see Sects. 2.3.2 and 4.3)

2.2.3 Synthetic Oligonucleotides

Chemically synthesized oligonucleotides are usually too small to be visualized as unique sequence probes in mapping studies. However, they can be used for detection of high-frequency tandem repeat sequences such as those found in alpha satellite, telomeric, and Alu repeats present in centromeres, telomeres, and chromosome bands. These probes can be important in determining the chromosome-specific location of single-copy probes, and as an alternative to chromosome banding (see Sect. 4.3.1), especially in species other than human or mouse. Specialized artificial peptide nucleic acid (PNA) probes that have a polyaminoethyl glycine backbone replacing the pentose-phosphate backbone have the advantage that they bind DNA more stably at low salt concentrations. These probes have been useful in quantitative examination of telomeric repeat sequence lengths.

2.3 Probe Labeling

2.3.1 Purification

Prior to labeling, the starting probe DNA must be purified. The labeling method determines the extent of the required chemical purity of the probe. For preparations to be labeled by nick translation, DNA purification procedures are required for removal of RNA, which, if present, would result in significant nonspecific background. PCR labeling methods do not require the same purity of DNA. If cloned probes are being used, it is usually not necessary to remove probe DNA inserts from the host vector DNA prior to labeling since vector DNA does not reduce the labeling efficiency of insert DNA and will not hybridize to mammalian DNA targets. Cloned genomic probe preparations are most often isolated and purified using standard CsCl or alkaline lysis procedures, or commercially available kits (Qiagen, Roche Diagnostics).

2.3.2 Labeling Methods

The goal of labeling a probe for FISH is to both incorporate labeled nucleotides into the DNA and to produce short DNA fragments of 100 to 500 bp. This fragment size allows probe access to target chromosomal DNA while maintaining high specificity. Longer fragments can result in reduced target access, producing reduced specific signal and increased background. Shorter fragments usually require reduced stringency wash conditions and may result in significant nonspecific chromosomal background.

Nick translation, the preferred method of labeling for most genomic probes, both introduces a labeled nucleotide into the probe DNA, and cuts the DNA into appropriate fragments. Nick translation combines a DNase I enzyme, which randomly nicks single strands of the DNA template producing available 3′ends, with *Escherichia coli* DNA polymerase I, which both removes nucleotides by its 5′-3′ exonuclease activity and incorporates nucleotides at the available 3′ ends by its 5′-3′ polymerase activity. The labeling mixture consists of probe DNA, the enzymes DNAse 1 and DNA polymerase 1, a dNTP mixture and a fluor, aminoallyl, or hapten-coupled nucleotide (Table 1), usually in final volume of 100 μL or less. The reaction mixture is placed in a 15 to 16 °C water bath, for 1.5 to 2 h or longer depending on the labeled nucleotide being incorporate. Following incubation, the reaction is placed on ice and the size of the labeled fragments checked by electrophoresis on a 1.5 to 2% agarose minigel containing ethidium bromide. A standard sizing ladder should be included in the run. Samples can be heat-denatured prior to loading on the gel, but this is not necessary. If the fragment size is too large, the reaction can be continued with further incubation; if the fragment size is too small, the labeling procedure must be repeated for a shorter time. If the fragment size is within the correct range, the reaction is stopped by addition of 0.5 M ethylenediaminetetraacetic acid (EDTA), pH 8.0. High prime or random primer labeling is another less popular approach that may provide a higher efficiency of labeled nucleotide incorporation, but fragment size cannot be well controlled, and further digestion by DNase I is often required. PCR labeling using specific primers can be performed when amplifying DNA from any source, but requires some knowledge of the probe sequence. Alternatively,

Tab. 1 Common fluors and haptens for FISH.

	Peak wavelengths [nm]	
	Excitation	**Emission**
Fluor		
Fluorescein (FITC)	494	518
Alexa fluor 488	495	519
Cyanine Cy2	492	510
Oregon green 488	496	524
Spectrum green	497	524
Tetramethylrhodamine(TRITC)	550	570
Indocarbocyanine Cy3	550	570
Alexa 546	556	573
Spectrum orange	559	588
Texas red	596	620
Alexa 647	650	668
Indodicarbocyanine Cy5	650	670
DNA counterstain[a,b]		
4′-6-diamidino-2-phnylindole (DAPI)	365	450
Propidium iodide (PI)	540	615
Hoechst 33258	360	465
Chromomycin	445	570
Quinacrine	455	495
Haptens		
Biotin		
Digoxigenin (DIG)		
Dinitrophenyl aminohexanoic acid (DNP)		
Bromodeoxyuridine (BrdU)		

[a] Dye/DNA complex.
[b] Spectral properties are dependent on dye/DNA ratio, and ionic strength and pH of the dye solution.

using degenerate oligonucleotide primer (DOP)-PCR, complex probe libraries such as chromosome-specific painting probes can be prepared and labeled by incorporation of fluor, hapten, or amine group-derivatized nucleotides.

The type of labeling usually depends on the size of the probe and/or the target.

1. *Direct labeling*: Single-copy genomic probes such as PACs, BACs, or YACs that are >100 kb, and probes for large targets such as alpha satellite repeats, can usually be detected by the direct labeling approach. Fluorochrome-coupled or aminoallyl-derivatized nucleotides that bind amine reactive fluorescent dyes can be obtained commercially or prepared in the laboratory and are incorporated into probe DNA by nick translation, sequence-specific PRC, or DOP-PCR labeling. Fluorochrome-labeled DNA probes usually do not require further detection/amplification systems, however,

antibodies directed against some fluors are available if further amplification is required.

2. *Indirect labeling*: Unique sequence cDNA, EST, and smaller genomic probes (cosmids and plasmids) are too small to be visualized directly and require some form of amplification. These probes are best labeled with a hapten such as biotin or digoxigenin (DIG), which must be further reacted with a fluorescent reagent to be visualized. DNA probes labeled with biotin or DIG-conjugated nucleotides require secondary detection systems involving fluorochrome-labeled avidin or anti-DIG antibodies. In general, although direct labeling is usually a longer process then indirect labeling, detection is significantly expedited. The most common fluors and haptens used for FISH mapping are shown in Table 1.

2.3.3 Postlabeling Purification and Storage

Following the labeling process, the labeled DNA is purified by ethanol precipitation or with spin columns. Addition of carrier DNA (salmon or herring sperm DNA) or glycogen increases the yield of DNA during ethanol precipitation and reduces nonspecific background binding to both the target DNA and the glass slide. Genomic probes also require suppression hybridization with competitor Cot-1 or genomic DNA to block the common interspersed repeat sequences present (see Sect. 2.2.1). Following precipitation, the probe is centrifuged, washed with cold 70% ethanol, recentrifuged, and air-dried. The DNA pellet can be left as a dry pellet or resuspended in distilled water, Tris EDTA (TE) buff, or hybridization buffer and stored at −20 °C.

3 DNA Targets for FISH Mapping

3.1 Metaphase Chromosomes

Metaphase chromosomes have been widely used to assign a chromosomal map location to the many genes and DNA sequences (both human and other species) arising from the human genome project. Genomic DNA in the form of metaphase chromosomes is in its most condensed state and mapping resolution is limited to interprobe distances of >1 Mb. Localization of a single labeled probe on a chromosome can be determined by either the relative position of the signal along the length of the chromosome, or by its localization within a specific chromosomal band. Relative positioning is achieved by measuring the distance between the telomeric end of the p arm (pter) and the probe as a fraction of the total chromosome length (Flpter, or fractional length from the pter). Difficulties with this approach include nonuniform condensation along the chromosome and the presence of repeat sequence length polymorphisms in different individuals. This method has, for the most part, been replaced by band or subband localization.

Reproducible banding patterns unique for each chromosome have been developed using DNA staining procedures. Banding techniques that allow simultaneous visualization of the probe signal and the banding pattern provide an efficient and precise approach to metaphase gene mapping.

Multicolor probes can be used to visualize the order of several genes or markers along metaphase chromosomes. The labeled probes must be separated by at least 1 Mb to be distinguished, but since most chromosomal bands are several Mb

in width, ordering probes within a single band can be accomplished by this multicolor FISH. Mapping to telomeric and centromeric regions should be performed with caution and verified by other approaches, as chromatin folding in these regions tends to produce artifacts and order changes.

3.1.1 Metaphase Chromosome Preparation

The most common source of human and large mammal metaphase chromosomes suitable for gene mapping is mitogen-stimulated short-term cultures of peripheral blood lymphocytes. Amniocyte cultures, fibroblasts, lymphoblasts, and cell cultures can also produce good results. For mouse and rat, bone marrow, spleen, lymph nodes, and thymus are commonly used for chromosome preparation. For precise mapping, high quality, well-spread chromosomes with distinct banding patterns are essential. Preparation of mapping quality metaphase chromosomes is based on minor modifications of standard cytogenetic techniques. Cell synchronization may be required for some types of banding (see Sect. 4.3.1) and usually enhances the number of cells in metaphase, however, it is generally not necessary for most mapping situations.

There are four basic steps involved in the preparation of metaphase chromosomes from cultured cells:

1. Stimulation of peripheral T lymphocytes by the mitogen phytohaemagglutinin (PHA).
2. Inhibition of cell cycle by addition of a mitotic spindle inhibitor (colcemid) to arrest the cells in metaphase, that is, when the genomic DNA is condensed into chromosomes.
3. Hypotonic swelling of the cells to reduce cytoplasm and enhance chromosome spreading.
4. Fixation to inhibit further swelling, arrest cell function, and preserve good cell morphology.

The following comments pertain to human cells, but can easily be adapted to other species. Heparanized peripheral blood should be cultured within three to four days of collection in a medium containing 1% PHA to stimulate the division of T lymphocytes. When the cultured cells reach maximum mitotic activity (within 60–70 h), addition of colcemid for 20 to 30 min prior to harvest disrupts spindle formation and arrests the cells in metaphase. In this phase of the cell cycle, the nuclear membrane disintegrates, allowing the chromosomes to disperse within the cell. Production of good metaphase chromosome spreads depends on both the concentration of colcemid and the length of exposure. Too long a treatment results in short, condensed chromosomes and tight metaphases unsuitable for mapping purposes. Some protocols for preparation of murine chromosomes omit colcemid altogether. For adherent cells, subculturing 48 h and feeding 24 h prior to harvest at 50 to 80% confluence can enhance the mitotic index, and addition of colcemid for a longer time (4–6 h to overnight) is suggested for optimal metaphase chromosome production.

If cell synchronization is required, methotrexate (amethopterin) can be added to growing cells. Methotrexate inhibits incorporation of thymidine into DNA, thus arresting the cells in early S-phase. Following the removal of the S-phase block and addition of thymidine or its derivative bromodeoxyuridine (BrdU), the

cells resume DNA synthesis and proceed through the cell cycle in a synchronized manner. Substitution of BrdU is used when replication banding is required (see Sect. 4.3.1).

Harvested cells are swollen with a hypotonic solution of KCl or medium to increase fragility and lyse any remaining red blood cells. The loss of nuclear membrane results in the release of chromosomes throughout the enlarged cells facilitating their dispersion. The cells are slowly fixed in methanol:glacial acetic acid to enhance nucleoprotein cross-linking and gradually dehydrate the cells, preserving them in the swollen state. Proper fixation is critical for good chromosome morphology and attachment to the slides. Improperly fixed cells may result in cell contraction and poorly spread tight chromosomes. Overfixation can result in probe and reagent penetration problems, whereas underfixation may result in loss of chromatin and chromosome morphology. The fixed cells can be stored as pellets in tubes containing fixative at −20 °C for several months. Resuspension in fresh fixative prior to each use, and minimizing exposure to room temperature air prevents evaporation of methanol and fixative esterification due to uptake of moisture. The fixed cells are dropped onto clean, wet glass slides and allowed to run down the slide held at a 45° angle, and dried. Metaphase spreads should be aged at room temperature for at least two days before being used for FISH mapping. Slides can be stored at room temperature up to four weeks, or for several months at −20° in a sealed dry container.

Preparation of high-quality metaphase spreads with clear banding resolution suitable for mapping studies is dependent on many factors. Colcemid treatment and fixation, humidity, temperature, and drying time are all critical steps during slide preparation and affect chromosome spreading, morphology, and the degree of residual cytoplasm present on the slides. Residual cytoplasmic proteins and RNA can increase background and reduce visualization of signal. Under ideal conditions of temperature, humidity, and drying time, when viewed under a phase-contrast microscope, cytoplasm should not be visible and chromosomes should be dark gray with sharp borders. The use of commercially available controlled environment chambers has significantly improved the standardization of slide making.

3.2 Interphase Nuclei

Interphase nuclei mapping provides improved resolution for ordering probes situated too close to each other to be distinguished on metaphase chromosomes. Since the folding of DNA in somatic interphase nuclei is about 20× less than that found in metaphase DNA, the mapping resolution is increased to 50 to 100 kb, compared to 1 to 3 Mb in metaphase DNA. In certain regions of the genome, the distance between probe signals in interphase nuclei are proportional to their genomic distance, with a linear relationship existing between the mean square of the distance (≤1.5 μm) and the genetic distance (<1–2 Mb). Since probes separated by >1 to 2 Mb do not follow this relationship because of chromatin looping, mapping such probes may result in incorrect distance measurements.

Somatic interphase mapping does not provide information on chromosomal banding localization or location relative

to the centromere or telomere region. Probe order, however, can be determined by two approaches: in the first approach, three or more probes labeled with a single fluor can be ordered on the basis of distance measurements (mean-square distance) of pairwise hybridization experiments provided they are in the 50 kb to 1–2 Mb range. Use of probes with a high hybridization efficiency and evaluation of ≥100 nuclei are necessary for statistically accurate mapping results using this method. In the second approach, three probes, each labeled with a different fluor, or combination of fluors, can be hybridized simultaneously and the visualized order of signals can be examined in at least 50 nuclei.

The use of pronuclei, produced by the *in vitro* fusion of hamster eggs with hamster or human sperm, can further increase interphase mapping resolution by ~3 times as they have a diameter of ~50 μm compared to 10 to 20 μm for somatic interphase nuclei. Thus, a resolution of 20 to 25 kb can be achieved.

3.2.1 Interphase Nuclei Preparation

Interphase mapping is best performed on G0/G1(2 N) fibroblast nuclei grown to confluence to reduce the number of G2 (4 N) nuclei. The presence of G2 nuclei in uncultured preparations (peripheral blood lymphocytes) makes results more difficult to interpret. Fibroblast nuclei are large and flatten well following hypotonic treatment allowing visualization of signals on the same plane of focus. Fibroblasts grown under standard culture conditions until they reach confluence should be left for a further 6 to 7 days at 37 °C until no mitotic cells are observed. The cells are washed in a citrate saline solution, trypsinized, and collected by centrifugation. The pelleted cells are then treated with a hypotonic solution of KCl or medium, fixed with careful addition of 3 : 1 methanol : glacial acetic acid, and stored as a pellet at −20 °C as described for the preparation of metaphase chromosomes. The fixed cells are dropped onto clean glass slides and as the fixative evaporates, the cells flatten. The cell concentration should be such that there are sufficient nonoverlapping nuclei for adequate analysis. The air-dried slides can be stored at room temperature for up to four weeks or longer at −20 °C in a sealed container containing desiccant.

3.3 Chromatin and DNA Fibers

High-resolution fiber FISH methodology was developed as a result of the demand arising from the Human Genome Project. In order to group the increasing number of genes being mapped during the early 1990s, and to provide order within the same chromosomal region (or band), a high-resolution mapping approach was developed. High-resolution mapping has allowed for the construction of a finely integrated genetic and physical map, forming the foundation for the sequencing phase of the Human Genome Project.

As mentioned earlier, signal resolution using metaphase chromosomes is significantly limited. Chromatin fibers are highly compacted, and a single, small chromosomal band can contain up to a few Mb pairs of DNA, and include many genes. To take advantage of the less compacted status of chromatin, methodologies have been developed to produce released chromatin or DNA fibers for direct visualization. Originally named "free chromatin FISH," fiber FISH has demonstrated the feasibility of FISH signal

detection on extended chromatin fiber for chromosomal structure analysis and physical mapping. A variety of modifications to improve the original methodology, such as increasing the mapping resolution, simplifying the fiber preparation, reducing the length variation of released fiber, quantifying the measurements of FISH signals, performing multiple-color detection, and combining DNA-protein and codetection using released fibers quickly followed. Various names corresponding to the methods of fiber preparation include: extended chromatin or DNA fiber FISH, elongated chromatin FISH, DNA halo FISH, individual stretch DNA molecule FISH, direct visual hybridization (DIRVSH), quantitative DNA fiber mapping, and molecular combing. The original and alternative approaches are now collectively referred to as high-resolution fiber FISH.

3.3.1 Fiber FISH Target Preparation

The released fibers can be categorized as either chromatin fiber or DNA fiber according to the preparation protocols and the morphological features following DNA staining. However, in many cases, mixtures of both types of fibers are produced. Generally speaking, the harsher the releasing conditions, the more DNA fiber will be generated. Chromatin fiber can be further released into DNA fiber by changing the release conditions to strip chromatin proteins. Fibers that are generated from cloned DNA fragments or the molecular combing procedure will generate no chromatin fibers.

Each type of fiber has advantages and disadvantages in terms of resolution and coverage. The resolution of chromatin fiber FISH is around 20 kb with coverage up to a few megabases; the ideal range to study larger regions of the genome in detail. In contrast, resolution of DNA fiber FISH is a few kb with coverage around 200 to 500 kb, which is best suited to a small area requiring the highest resolution. Thus, it is necessary to choose the type of fibers according to the desired resolution and coverage. Often, a combination of both types of fiber is desirable. It should be noted that the use of chromatin fiber has a unique benefit based on the following two considerations. Chromatin fiber may be more useful than DNA fiber when a high-resolution approach is required to study DNA–protein complexes. In addition, when comparing pattern or size variations between homologous regions, it is easier to trace individual cells using partially released chromatin fiber than DNA fibers.

Chromatin fiber preparation Chromatin fiber preparation protocols are based on effective release of chromatin fibers without destroying the 30-nm chromatin fiber structure. Four typical approaches have been developed to achieve this purpose: (1) chemical treatment to interfere with chromosomal condensation; (2) alkaline buffer or extensive hypotonic treatments to lyse the nuclear envelope; (3) accumulation of late G2 or early G1 phase cells to generate free chromatin; and (4) physical stretching of the nuclei with a cell centrifuge to produce chromatin fibers.

The chemical treatment method was the original approach that was used to increase the frequency of free chromatin. Although the steps for this approach are tedious when compared to other methods, an advantage is that different drugs may be used to delineate the chromosome condensation process, an important strategy to study the high-order structure

of the chromosome. Preparation of chromatin fibers from cultured lymphocytes by drug treatment is accomplished by culturing lymphocytes for 48 to 52 h followed by short-term treatment with ethidium bromide (EtBr), BrdU, or mAMSA. Harvesting chromatin fiber uses the same standard protocols as for harvesting chromosomes. After checking the chromatin fiber density under phase-contrast microscopy, if it is determined that the fibers are too crowded on the slide, the concentration of the suspension can be adjusted by adding more fixation solution. Once a high-quality batch of slides is completed, they should be air-dried at room temperature for a day, and sealed in slide containers with parafilm. Slides may then be stored for several weeks at −20 °C. It should be noted that irreversible damage occurs to chromatin fiber slides that are overdehydrated. Fixed free chromatin suspensions can be stored at −20 °C for a much longer time.

A second approach to increase the amount of chromatin fiber uses protocols designed to open the nuclear envelope. The use of alkaline buffer to release chromatin fibers from cultured cells is based on the fact that nuclear lamins can be interrupted by high pH treatment, destabilizing the nuclear envelope, and rendering it susceptible to rupture by hypotonic treatment. Prolonged alkaline treatment can cause breakage of the nuclear envelope without hypotonic shock. Aliquots of a harvested fresh cell suspension are transferred into tubes containing alkaline buffer, and at various time intervals (3 to 5 min), the alkaline treatment is terminated by adding fixation solution followed by refixation. The fixed cell suspension is dropped onto slides and air-dried. Chromatin fiber density is checked using a phase-contrast microscope and the concentration of the suspension is adjusted accordingly. The optimal alkaline treatment for each particular cell line has to be obtained empirically using a brief screening test.

Generating chromatin fibers by this approach is affected by many factors including the length of treatment by the alkaline buffer and the pH (10–11.5). In general, a high pH over a longer period of time promotes the lysis of nuclei. The optimal combination for a given cell line can be found by systematically varying the conditions. Excessive treatment should be avoided since it reduces the number of useful chromatin fibers because of aggregation that follows release from the nuclei, and it can destroy the 300 Å structure leading to the production of the naked DNA fiber. Optimal results are best obtained by fixing the chromatin fibers quickly before they become aggregated following the release from the nuclei. The use of a small volume of alkaline buffer and a large volume of the prefixation solution is also a good way to avoid aggregation of chromatin fibers.

A cytospin also works well for generating chromatin fiber for many cell lines. Cells are resuspended in CSK buffer, incubated for a short time on ice, and then spun onto a microscope slide using the cytospin. The slides are then rinsed with fixation solution and examined under a phase-contrast microscope to determine the chromatin quality. Optimal cell concentrations vary for different types of cells and should be adjusted accordingly.

It should be noted that different protocols and even the same protocol will generate chromatin fibers with varied resolution, as it is difficult to control the releasing process that consists of multiple levels of DNA condensation. It is also common to obtain a mixture of 30-nm chromatin fibers and naked DNA fibers,

or fibers between chromatin and naked DNA including 10-nm chromatin fibers. An important step for fiber FISH detection is to include an internal marker probe of a known molecular size. By cohybridizing the probes in question with such a marker (using contrasting differential color detection), the size of the marker can be used to compare and measure the probes or regions in question. As long as an internal marker probe of known length is used, quantitative measurements can be performed, allowing the comparison of different chromatin fibers. For other types of analysis using fiber FISH where precise mapping information is not necessary, the use of a known marker for measuring purposes is not needed.

DNA fiber preparation Protocols for preparation of released DNA fibers fall into the following four categories: (1) complete release of DNA using detergent/alkaline treatment then linearizing fibers by gravity or by pulling with a coverslip; (2) generation of DNA fiber from a "Halo" preparation by protein extraction; (3) nuclear lysis and DNA release in a gel block then linearizing fibers by a mechanical or an electronic pulling force; and (4) linearizing DNA molecules by molecular combing. Two protocols that are very popular due to their simplicity and the quality of fibers produced will be described. The release of DNA fiber using alkali and linearization of the fiber by mechanical "pulling" is one of the simplest protocols to perform. Preparation of DNA fiber by pretreating cells within gel blocks and then mechanically "pulling" the fiber reduces interference from other nuclei. These two protocols should be useful as they exemplify most cases and can serve as a technical gate for additional protocols.

The first simple and popular method for preparing DNA fiber is the method of alkaline release and linearization of the DNA using mechanical pulling. Cells quickly fixed after hypotonic treatment are dropped onto a slide and treated with alkaline solution. One edge of a coverslip is placed on the slide and the coverslip is then pulled along the slide from one end to the other. The slides are ready for FISH after dehydration. The duration of alkaline solution treatment can be adjusted according to the desired release, with a longer treatment resulting in a greater amount of release from the nuclei. By adjusting the release conditions, this protocol can also be used to generate chromatin fiber. The alkaline solution can also be replaced with SDS-releasing buffer as described by others.

The second popular way of fiber preparation utilizes pretreatment of cells within gel blocks and mechanically linearizing the fiber. Lymphocytes are embedded in low-melting-point agarose and the resulting blocks of cells are treated with proteinase K and RNase A. A small piece of agarose-embedded DNA is placed at the end of a poly-L-lysine-coated slide. The slide is placed in a microwave oven, heated until the agarose melts, and the DNA is extended on the slide using a second slide that is pulled across the DNA fiber slide from one end to the other. The slide is then air-dried.

The fiber FISH molecular combing approach, that permits the stretching and aligning of deproteinized DNA on a solid surface, has recently drawn attention by demonstrating its effectiveness to high-resolution mapping. The advantage of this DNA fiber preparation method is the consistency of the length of prepared DNA fibers, an essential element of acceptable statistical analysis. Because of

the requirements of specific computer software and special instrumentation, this application has been limited to only a few research groups.

4 Denaturation, Hybridization, and Detection

4.1 Pretreatment and Denaturation

Prior to denaturation, pretreatment of slides with RNase and/or Proteinase K may be required to reduce signal background due to nonspecific binding, and to increase probe access to its target DNA. RNase pretreatment is usually required when using the counterstain propidium iodide, as it stains both DNA and RNA.

In order for probe DNA to bind to its homologous sequences on the target, both must be denatured, that is, made single-stranded. Denaturation of probe and target DNA can be performed separately or simultaneously. In either case, precipitated labeled probe DNA (see Sect. 2.3.3) is resuspended in a hybridization mixture containing formamide and dextran sulphate. Formamide destabilizes DNA such that it dissociates at a more convenient lower temperature than the normal "melting temperature." This is critical for maintaining target chromosome morphology. The concentration of formamide determines the stringency of the reaction, and a higher concentration should be used with repeat sequence probes to minimize cross-reaction. As oxidation of formamide can result in DNA depurination, it is critical that the formamide be of high quality, deionized, and have a pH of 7.0. Dextran sulphate changes the void volume increasing the "apparent" DNA concentration. If probe and target are to be denatured separately, the probe in hybridization buffer is placed in a 75 °C water bath for 5 min (genomic probes) or a 95 °C water bath for 10 min (unique and repeat sequence probes). Probes containing unique or tandem repeat sequences should be placed immediately on ice and added to the target as soon as possible. Probes requiring suppression hybridization must be denatured prior to the target DNA to allow for preannealing of the competitor DNA to take place. As discussed in Sect. 2.2.1, same species competitor (Cot-1 or genomic) DNA is added to any probe containing interspersed repeats to suppress the binding of these sequences to the homologous target repeat sequences. Slides containing the target DNA are denatured in 70% formamide for 2 min at 70 °C and immediately placed in ice cold 70% ethanol, dehydrated in 90 and 100% ethanol, and air dried. The denatured probe is then added to the slide, covered with a coverslip, sealed with rubber cement, and incubated at 37 °C overnight in a humidified chamber.

Alternatively, codenaturation of probe and target can be accomplished using a modified programmed slide warmer or PCR machine. The undenatured probe is added to the slide, covered with a coverslip, sealed with rubber cement, and placed in the slide warmer set to 70 to 95 °C. Both probe and target DNA are denatured simultaneously for 2 to 5 min, and the temperature is then ramped down to 37 °C for hybridization. This approach simplifies the denaturation procedure and minimizes handling of formamide, a toxic teratogen. Codenaturation is appropriate for probes not requiring the preannealing hybridization suppression step.

4.2 Hybridization

If probe and target DNA are denatured separately, the denatured probe is placed on the dried slide, covered with a coverslip, sealed with rubber cement, and placed in a humidified box in a 37 °C oven for hybridization. This temperature is critical for optimal hybridization. At temperatures below 37 °C, lower stringency hybridization can result, causing increased nonspecific chromosomal background signal. Conversely, higher temperatures may reduce signal intensity because of increased stringency conditions. If the slides dry out during this step, nonspecific background will be increased significantly. Although most probes are allowed to hybridize overnight, centromeric alpha satellite probes and some other larger genomic probes (BACs, YACs) may produce sufficient signal for visualization after 2 to 4 h. Hybridization timing should be determined empirically for individual probes.

4.2.1 Stringency Washes

Following hybridization, unbound DNA sequences are removed by washing with a solution of slightly higher stringency than that used for the hybridization. Stringency is determined by a combination of formamide concentration, temperature, and salt concentration. Varying any one of these will alter the stringency. Increasing the temperature and formamide concentration or lowering the salt concentration, all result in a higher stringency and require a more exact matching of probe and target sequence (lower % mismatched base pairing). Stringency conditions of hybridization and washing are key factors in optimizing the signal-to-noise ratio. If the stringency is too high, the background may be clean but may result in a weak signal; if the stringency is too low, high nonspecific background may obscure the true signal. Stringency conditions vary with the type of probe and target, and should be optimized for each mapping condition. The most common wash protocols use either a 50% formamide wash at a low temperature of 42 to 44 °C or a high temperature wash without formamide at 72 °C in 1 to 2x SSC (saline sodium citrate). Following the wash step, the slides can be counterstained and viewed immediately (direct labeled probes) or processed further for detection (indirect labeled probes).

4.3 Detection

Probes that have incorporated a fluorescently labeled nucleotide directly into the DNA sequence (direct labeling) can be detected immediately after the stringency wash and are thus the simplest to use. Biotin-labeled probes (indirect labeling) can be detected with reporter molecules such as avidin or streptavidin labeled with a fluorochrome. Signal amplification is achieved by addition of biotinylated anti-avidin/streptavidin antibody and another round of avidin/streptavidin labeled with fluorochrome. Similarly, DIG-labeled probes can be detected with an anti-DIG antibody followed by a DIG-labeled secondary antibody specific for the primary antibody and an anti-DIG antibody labeled with fluorochrome. Both systems work well for most probes. Although theoretically the number of rounds of amplification can be increased several times to enhance the signal, in reality, after one or two rounds, the nonspecific background increases substantially and lowers the signal-to-noise ratio. Instead, for weak signals or very small DNA probe sequences ($\leq$1 kb), another signal

amplification approach, termed *tyramide signal amplification* (TSA) or *catalyzed reporter deposition* (CARD) amplification, has been developed on the basis of the reaction of horseradish peroxidase and tyramine. A peroxidase-labeled hapten or antibody specific to the labeled probe is added to the bound probe, followed by biotinylated tyramine. The reaction of tyramine and peroxidase results in precipitation of a large number of biotin molecules at the site of probe hybridization. Biotin is then readily detected by fluorochrome-labeled avidin. This has resulted in successful localization of probes <1 kb in size.

Following the detection steps, slides can be counterstained directly or kept in the dark for 1 to 2 days at 4 °C in a 0.1 M phosphate buffer, pH 8.0 containing a detergent (0.05% NP-40). Slides should not be allowed to dry out before addition of the counterstain.

4.3.1 Counterstaining and Banding

The aim of physical mapping by metaphase FISH is to identify the specific chromosomal band to which the sequence is localized. This requires a method of visualizing both the gene or DNA sequence and the chromosomal bands simultaneously. Chromosomal bands range from 3 to 10 Mb in width and the number of bands per chromosomal spread can vary from 500 (low resolution) to 1000 (high resolution). Bands are named according to ISCN nomenclature and are numbered from the centromere to the telomere on each chromosome arm.

Several DNA binding dyes are available for staining chromosomes, some of which produce banding patterns similar to those produced in classical cytogenetics by the absorption dye Geimsa (G-banding). The two most common dyes (counterstains) used for chromosome banding following *in situ* hybridization are 4′,6′-diamidino-2-phenylindole (DAPI) and propidium iodide (PI). Other fluorescent DNA stains such as Hoechst 33258(QFH banding), quinacrine (Q banding), or a combination of chromomycin A3 and distamycin A (R banding) are infrequently used. Banding can also be achieved by cohybridization of PCR labeled long or short interspersed repetitive sequences (IRS) LINEs and SINEs (G-bands) or Alu (R-bands) repeat sequences.

DAPI binds to AT minor groove double-stranded DNA sequences and produces a reverse G-banding pattern with bright blue bands corresponding to AT rich, gene-poor regions and dark bands corresponding to GC and gene-rich DNA regions. Further treatment of DAPI-stained chromosomes with actinomycin D, or cell synchronization and incorporation of the thymidine analog BrdU into the late replicating DNA may enhance banding patterns in cases where the bands are not sufficiently distinct. Using software such as Adobe Photoshop to electronically invert grayscale images of DAPI-banded chromosome to the conventional G-banding pattern is also a convenient approach to chromosome identification. DAPI stain requires excitation by UV light and thus must be used with microscope systems that have a UV light source.

PI is a nonspecific double-stranded DNA and RNA intercalating (binding) dye that can produce an R (reverse)-banding pattern on metaphase chromosomes when cells in culture are synchronized by a thymidine block and then released from the block in the presence of BrdU. BrdU is then incorporated into the late replicating DNA (G positive, AT rich bands). Staining with Hoechst 33258 followed by UV irradiation reduces the PI staining of these bands, resulting in bright red staining

of the early replicating CG-rich R-bands and dark staining of late replicating G-bands. High-quality banding is dependent on appropriate synchronization, and the timing of the block release and cell harvest.

The choice of counterstain is in part determined by the fluorochromes used to label the probes being mapped. Probes labeled with red fluors such as TRITC, Texas red, spectrum Orange, Cy3 should not be used with target DNA counterstained with the red dye PI. DAPI can be used with both red and green fluorochromes and does not require cell synchronization for metaphase chromosome preparation. When using either counterstain, too strong a stain can obscure the banding pattern or result in a weakened probe signal. Both counterstains can be used on interphase nuclei and fiber FISH targets.

In order to preserve signal fluorescence, antifade chemicals such as ρ-phenylenediamine dihydrochloride have been developed to decrease fading caused by exposure to UV light. Antifade is usually combined with the counterstain and added to the slides after the detection step.

5
Microscopy and Image Analysis

Accurate mapping of genes by FISH is dependent on the spatial and spectral characteristics of the epifluorescence microscope, the sensitivity of the camera for capturing and recording images, and the digital-imaging resolution for analysis. High-quality epifluorescence microscopes equipped with a properly aligned high-intensity light source designed for high-light collection power (detection sensitivity) are essential for mapping analyses. The three commonly used light sources are mercury lamps, xenon lamps, and lasers. Mercury arc lamps are used most frequently as their emissions are similar to the excitation peaks of many commonly used fluors (254–646 nm). Xenon lamps tend to give a more even excitation across the 250- to 1000-nm range and are most useful with fluors excited above 700 nm. Both mercury and xenon lamps require filter sets that control the wavelength of light reaching the fluor. The most commonly used laser sources for FISH, argon ion lasers (emission 488 and 514 nm), and helium-neon lasers (emission 543, 594, and 633 nm), do not require excitation filters and only have a narrow range of emitted light, thus limiting their use to a restricted number of fluors.

The objective lenses must be of the maximum numerical aperture possible, of low self-fluorescing UV transmitting glass, and demonstrate minimum spherical and chromatic aberrations (apochromatic) to produce optimal sensitivity of fluorescent signal visualization. It is useful to have a range of objective magnifications – dry 10× and 20× for scanning metaphases and nuclei, and oil 60x and 100x for visualization and localization of the signal. A nonfluorescing oil must be used for the high-power objectives. Mixing oils from different sources may cause cloudiness, and use of oil stored for long periods of time can result in increased autofluorescence, both of which reduce signal visualization.

The interference filters should possess low autofluorescing properties and be properly aligned to reduce unwanted reflections. Filter sets for viewing fluorescence signals and counterstained chromosomes are determined by the fluorochromes that are being used and the type of mapping being performed. To map a single probe, narrowband pass

filters, which have optimized peak excitation and emission wavelengths for a single fluorochrome, are used to provide maximum signal intensity. Wideband pass filters are used in relational mapping to view two (dual-band pass) or three (triple-band pass) colors simultaneously. These filters result in a reduction of individual signal intensity due to "bleed through" of emission signals. The mechanism of switching excitation and emission filters to visualize the different fluors sequentially has been a critical issue for mapping with respect to proper alignment of signal with exact chromosomal location. Careful matching and mechanical adjustment, and use of internal reference microspheres, multiband pass filters, or automatic filter wheels have all been employed to minimize image shifts. Currently, automatic computer-controlled filter wheels and software packages that enable the capture of real-time images are used to minimize this registration problem.

Direct visualization of signals from small single-copy or unique DNA sequences is often difficult, even at high magnification (100x oil). Cooled charged-couple device (CCD) cameras are used for fine mapping studies as they provide excellent sensitivity, spatial resolution, and linearity with high dynamic range (12–16 bit). Cooling CCD cameras to minus 38 °C reduces dark current noise and background signals caused by the camera itself. Most high-resolution CCD cameras are cooled monochrome cameras that capture grayscale images based on light intensity. Grayscale images must be captured separately for each fluorochrome and transferred into an image analysis software program such as Adobe Photoshop for addition of color and merging. Modified color CCD cameras are also available, however, the spatial resolution is limited in the single-chip color CCD cameras with the result that some signals may be missed. The use of these cameras with real-time image capture and dual or triple filters eliminates the need for merging images. Further, image software has developed rapidly over the past decade and now facilitates quantitative FISH, automatic image focusing, acquisition at multiple wavelengths, visualization of multicolor probes, multicolor pseudocolor simultaneous imaging, enhancement of DAPI-banded chromosomes, chromosome classification, slide location of cells, fluorescence intensity, morphometry, image storage, and many forms of data tabulation making FISH mapping and FISH-based analyses more feasible in research and clinical laboratories.

6 Applications

The main function of gene mapping by FISH is the precise physical localization of genes of known biological activity. Although many genes and DNA sequences can now be positioned using *in silico* methods and the vast information available in the genome databases discussed in Sect. 2.1, visual verification is often required to finalize and verify the precise chromosomal location. Thus, FISH has played an important role in positional cloning as it is the most direct way to map and order DNA clones within a chromosomal band. To date, thousands of genes and markers obtained by standard cloning approaches have been mapped by FISH. FISH has also provided an effective approach for anchoring YAC and BAC contigs within a chromosomal band.

The resolution of FISH mapping depends on the degree of target DNA

Tab. 2 FISH mapping on DNA targets of increasing resolution.

Visual mapping of genes and DNA clones on		*Resolution*	*Application*	*Advantages (+) and disadvantages (−)*
Metaphase chromosome	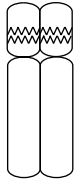	1–3 Mb	Gene mapping on chromosomal regions; detection of translocations, hemizygosity and regional amplification; detection of chimerism or homologous regions; whole genome mapping (SKY/M-FISH, CGH and ZOO-FISH)	+ Telomere–centromere orientation + Mapping to specific band + Whole genome approach + Highest coverage + Chromosome identity − Low resolution
Interphase Nucleus	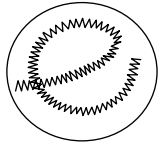	50–100 kb	Detection of translocation; visualization of chromatin domain interaction; monitor replication status; study gene amplification; ordering of clones	+ Distance measurement + Detect the level of gene amplification − Telomere-centromere orientation not possible − No chromosome identity

Released chromatin fiber	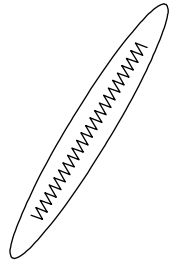	20–50 kb	Order of clones; visualization of the site of integration; mapping large chromosomal region with high resolution; DNA-protein codetection	+ High resolution + High coverage (20 kb–a few Mb) + Probe orientation with help of known markers – No chromosome morphological features
Released DNA fiber		1–5 kb	Ordering of clones; estimation of physical distance among probes; visualization of "gaps"; detection of size of deletion, amplification, and its orientation	+ Highest resolution + Probes orientation with help of known markers – Coverage is not high (1–a few hundred kb) – No chromosomal morphological feature

condensation, and spans several orders of magnitude ranging from 1 to 3 Mb on metaphase chromosomes, 50 to 100 kb in interphase nuclei, 20 to 50 kb on chromatin fiber, and 1 to 5 kb on DNA fibers. The type of target DNA used for FISH mapping will depend on the type of information required (see Table 2). Aspects of these three levels of gene mapping by FISH are presented in the following sections.

6.1 Metaphase Mapping

6.1.1 Single-copy Genes

Metaphase chromosomes present the most compacted form of DNA and thus provide the least-sensitive mapping resolution (1–3 Mb). Metaphase mapping has provided a subchromosomal banding localization for single-copy genes, viral insert sites, specific breakpoints, and regions of amplification or deletion. FISH mapping on metaphase chromosomes also orients genes relative to chromosome centromeres and telomeres. Key to metaphase FISH mapping is identification of true signal, the hallmark of which is a doublet signal on each chromosome (one per chromatid) of a homologous pair (see Fig. 2). The doublet signals must be visualized consistently at the same band location in each metaphase spread. The number of metaphase chromosomes needed to establish a chromosomal localization is determined by hybridization efficiency (% metaphases showing two pairs of consistent doublet signals). Hybridization efficiency is determined by the type and size of both probe and target DNA. Genomic BACs and PACs usually hybidize with >90% efficiency requiring analysis of 15 to 20 high-quality banded chromosomes. Smaller genomic probes (<10 kb) and cDNAs may show <20% hybridization efficiency and thus require analysis of more chromosomes (25–50).

Visualization of cDNAs can vary with the size of the target gene and its intron/exon

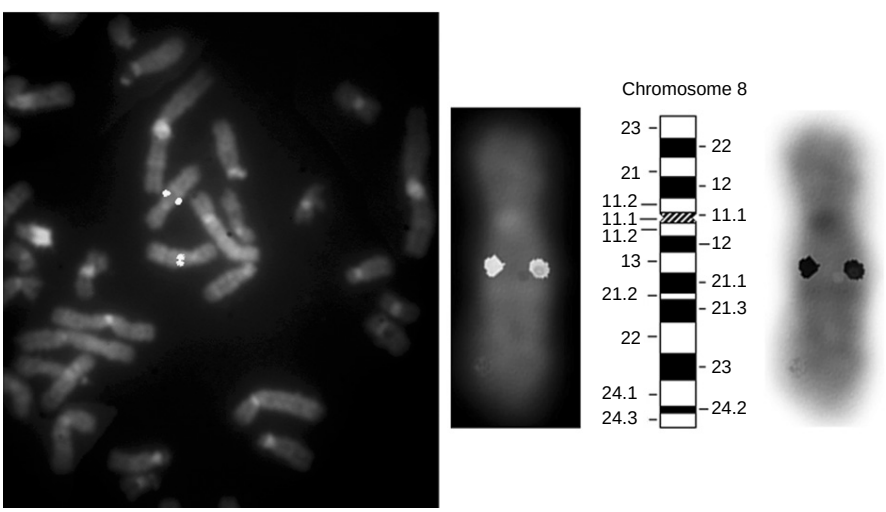

Fig. 2 Metaphase FISH mapping of a single-copy probe. Biotinylated DNA PAC probe (~100 kb) hybridized to normal human metaphase chromosomes counterstained with DAPI and detected with FITC. Doublet probe signals map to 8q13 on both chromosome 8 homologs.

structure, and usually requires amplification by approaches such as tyramide-based detection. Alternatively, cDNA screening of a PAC or BAC library can provide homologous genomic clones that are then easily mapped by FISH. Since one or more clones may be identified from such a screen, care must be taken in the interpretation of the results. Even if one genomic clone is identified, FISH mapping may result in either localization to a true mapping site on a single chromosome, or localization to two different chromosomes, suggesting the presence of a double or chimeric clone. In situations when several PAC/BAC clones are identified for one cDNA, PCR verification of the insert sequence should be performed prior to proceeding with FISH mapping. If two or more "verified" clones still result, localization to the same chromosomal site would indicate the true mapping site and redundancy within the library, whereas localization to different chromosome sites would suggest the presence of a gene family or a pseudogene. Since PAC/BAC libraries are publicly available (see Sect. 2.1), this approach has the advantage of providing both genomic clones and localization data for expressed sequences. FISH mapping on metaphase chromosomes is also a straightforward approach to the identification of the presence of chimerism within YAC clones, which are noted for their high rate of chimerism. Doublet hybridization signals at more than one chromosomal location indicate the presence of a chimeric clone and further analysis by somatic cell hybrids or other molecular techniques is advised.

Although the resolution of metaphase FISH is usually not sufficient for the fine mapping requirements of positional cloning, it provides an initial banding assignment that can be followed by more precise FISH mapping obtained with released chromatin or fiber FISH (Sect. 6.3). It should be noted that mapping in the subtelomeric and pericentromeric chromosomal regions is less accurate and should be avoided or used with caution as chromatin folding and duplications in these regions can result in inverted and/or erroneous ordering of probes.

6.1.2 Segmental Duplications and FISH Mapping

The presence of segmental duplications $\geq$1 kb with 90 to 98% identity has been reported to comprise 3.6% of the genome and has been the cause of some physical mapping ambiguities. These duplications show a strong pericentromeric and subtelomeric bias of approximately 5- to 10-fold, however, this clustering of duplications in the pericentromeric and subtelomeric regions is not uniformly distributed among all the chromosomes. BAC clones containing these duplicated sequences tend to be found more frequently in unordered and misassigned contigs as their chromosomal localization is difficult to assign. FISH mapping of BAC clones containing interchromosomal duplications has been used to determine their coverage and duplication patterns. Less than half the chromosomes positive by FISH have a corresponding chromosomal localization of these clones by BLAST sequence identity. FISH mapping data can thus be used as a standard with which to identify chromosomal localization of these duplicated sequences, and to compare the assembly completeness and accuracy of interchromosomal duplications.

6.1.3 Relational Mapping

Relational mapping is used when the order of two or more genes localized to

the same chromosomal band or region is important. A probe's relation to a cryptic deletion or amplification, a specific breakpoint (Fig. 3), or confirmation of probe order (Fig. 4) based on computational approaches, are situations in which FISH can provide valuable information. The condensed nature of metaphase DNA requires probe sequences to be separated by at least 1 Mb of DNA, and at least 25 metaphases should be scored for statistically accurate results. Multicolor probes are usually used for this type of analyses and the number of genes that can be mapped simultaneously will depend on fluorochrome availability and filter capabilities of the microscope. Mapping probes relative to a specific chromosomal breakpoint can provide information regarding their biological and/or diagnostic

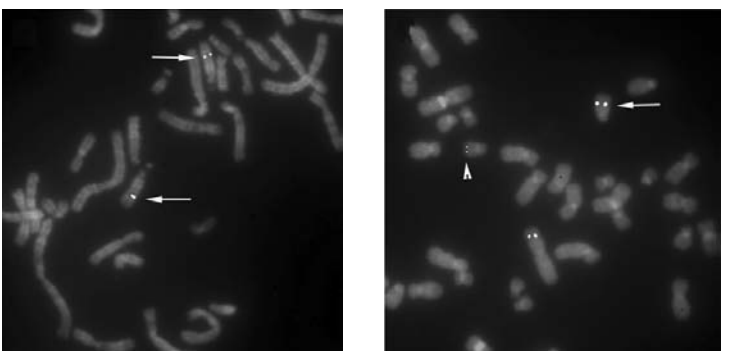

(a) (b)

Fig. 3 Mapping across a chromosome breakpoint. (a) A cosmid probe (14 kb) hybridized to normal human metaphase chromosomes showed doublet signals at 14q24 on both chromosome homologs. (b) Hybridization of the same cosmid probe to metaphase chromosomes from a patient with a 2 : 14 (p25;q24) translocation. The probe demonstrates doublet signals on both normal (arrow) and derivative (arrowhead) chromosome 14 as well a doublet signal on the short arm of derivative chromosome 2 (no arrow). A consistently stronger signal was noted on der2, indicating that the breakpoint is located closer to the proximal (centromeric) end of the cosmid. (From Barbara G. Beatty, Stephen W. Scherer. (2002), FISH: A Practical Approach reprinted by permission of Oxford University Press).

Fig. 4 Relational mapping on metaphase chromosomes using two fluorochromes. (a) Three probes: Pairwise mapping was used to order two YACs and one cosmid probe within a YAC contig, which mapped to 14q24.2–24.3. The two YACs were labeled with DIG or biotin and the cosmid with biotin. DIG was detected with rhodamine (red) and biotin with FITC (yellow). Left chromosome: DIG-YAC 1 plus cosmid; middle chromosome: DIG-YAC 2 plus cosmid; right chromosome: Biotin-YAC 1 plus DIG-YAC 2. Thus, the order shown is YAC 1 – cosmid – YAC 2. Note: YAC 2 often demonstrated double-hybridization signals in a lateral orientation suggesting two domains of strong hybridization or clustering of Alu sequences within the YAC (see right chromosome). (b) Two probes: Two PAC probes localized to 14q24-31 by FISH were labeled with different haptens (DIG-PAC A and biotin-PAC B) and hybridized simultaneously to normal human metaphase chromosomes counterstained with DAPI. PAC A (detected with rhodamine, red) localized telomeric to PAC B (detected with FITC, yellow). [From Barbara G. Beatty, Stephen W. Scherer. (2002), Fish: A Practical Approach, reprinted by permission of Oxford University Press.] (See color plate p. xxiv.)

significance. Genes that span a breakpoint can be identified if the sequence length on either side of the breakpoint is sufficient to produce a visible signal as shown in Fig. 3. Mapping DNA probes relative to other chromosomal abnormalities, regions of amplification or deletion, or to tandem repeats can be performed using metaphase chromosomes, provided probes are available for the abnormal region and the distance between them is >1 Mb. It is worth mentioning that quantitative FISH (Q-FISH) has demonstrated its ability to study the length of telomeric repeats for individual chromosomes. Using FITC-PNA probes and quantitative

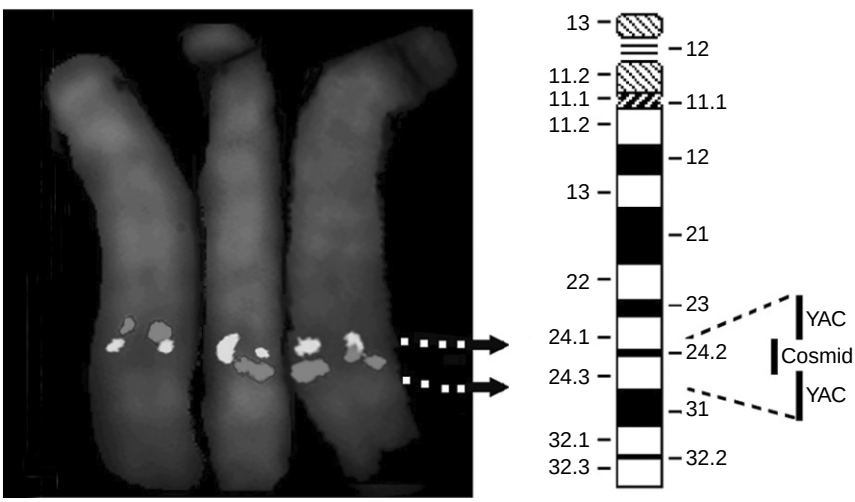

(a)

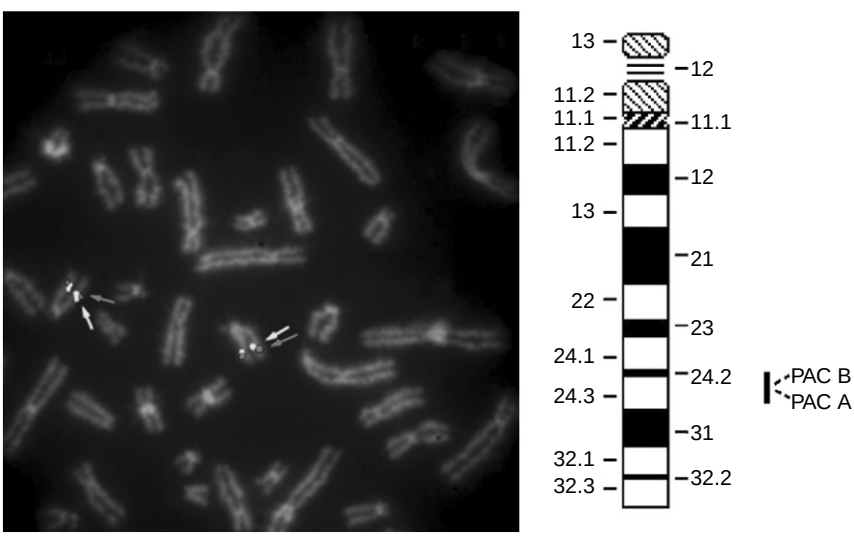

(b)

software, Q-FISH has been used to monitor changes in the telomeric repeats under different developmental, pathological, and disease conditions.

6.1.4 FISH Mapping Combined with CGH and spectral karyotyping (SKY)

With the development of the genome project, individual gene mapping has been expanded into genome mapping of chromosomal regions, whole chromosomes, and the entire genome. By combining individual gene mapping with entire genome or chromosomal FISH approaches such as CGH, which identifies regions of whole genome chromosome amplification and deletion, or SKY and M-FISH, which identify interchromosomal rearrangements, genes involved in these critical regions can be identified and localized.

CGH utilizes competitive hybridization of two probes (control and test) labeled with two different fluors onto normal metaphase chromosomes to reveal the presence and location of DNA copy number imbalances by comparing the relative signal intensities of each probe along the length of each chromosome. Special digital image analysis software programs developed to analyze and evaluate CGH data are available commercially. CGH has been particularly useful in identifying regions of DNA sequence copy number amplification and deletion in solid tumors. FISH mapping of single-copy or unique sequences within these regions can provide important information about the identification of candidate oncogenes and tumor suppressor genes respectively. Resolution of CGH on metaphase chromosomes is limited to several megabases. A significantly higher mapping resolution can be achieved using targets such as cloned DNA fragments (BACs, PACs, etc.) and cDNAs arrayed on glass slides. This high-resolution CGH, termed *matrix CGH*, can be used to map and study microdeletions and sequence overrepresentations.

Similarly, multicolor chromosome paint cocktails have now been developed for identifying interchromosomal rearrangements and mapping translocation breakpoints within the whole genome. Two commercially available approaches, SKY, and M-FISH utilize FISH probes derived from flow-sorted or microdissected chromosomes labeled by DOP-PCR (see Sect. 2.3.2). Each chromosome is identified by a unique spectral signature (SKY) or combination of fluors (M-FISH), permitting unambiguous identification of marker chromosomes and the different chromosome fragments involved in interchromosomal rearrangements. Sequential application of conventional G-banding and SKY on the same metaphase, as well as the combinatorial use of SKY and FISH, makes it possible to define the individual chromosomal bands involved in the structural rearrangements. Identification of putative genes or markers adjacent to or spanning specific breakpoints by *in silico* methods can then be verified and mapped by conventional FISH analysis.

6.1.5 Comparative Mapping by ZOO-FISH

Another application of gene mapping is the comparative mapping of map-rich species (human/mouse) with experimentally and economically important map-poor species (cattle, pig, horse, chicken). The location of conserved sequence segments between species can be visualized using composite chromosome-specific DNA probes (paints) from one species to hybridize to the chromosomes of another species (target). This cross-species painting approach is termed *ZOO-FISH* and has a resolution of approximately 7 Mb. The probes used are generally chromosome paints from

chromosome-specific libraries, but can be cDNAs or genomic probes (PACs/BACs) obtained from corresponding species-specific libraries. ZOO-FISH provides a mechanism for identifying major regions of conserved synteny between the two species being examined. This application of FISH has provided valuable information on regions of homology between the human genome and other map-poor species such as pig, chicken, and cattle, and between different members of a species including mammals, insects, and plants. It is expected that a better understanding of cross-species homologies will provide new models for the study of disease as well as new hypotheses of genome evolution.

6.2 Interphase FISH Mapping

FISH mapping using interphase nuclei provides a higher order of resolution than metaphase mapping and is used for clone ordering and generating physical maps with a resolution of <1 Mb. It is also useful for verifying distance measurements obtained from genome database contig information in the 50 kb to 1 Mb range. FISH mapping with pronuclei provides more precise ordering and estimation of genomic and physical distances and has been used in preparing high-resolution physical maps of chromosome 19. The concerns regarding pronuclei interphase mapping are that they are time consuming, laborious, and not suited to large-scale mapping approaches. Further, although the relationship between genomic and physical distance is considered linear from ~50 to 800 kb, the extent of DNA condensation within these nuclei cannot be controlled. Interphase mapping provides relative mapping distance measurements and should contain an internal control of markers a known distance apart. No information on chromosome location is provided by this approach.

6.3 Chromatin and DNA Fiber Mapping

Fiber FISH has demonstrated its advantages in a variety of applications. The majority of publications have involved physical mapping and chromosome structure studies of both plant and animal genomes.

Fig. 5 Sizing the "gap" in current physical map using fiber FISH: Two BAC probes were detected by two-color fiber FISH and the distance between two BAC probes represents the size of "gap".

These applications include: gap estimation for the current Human Genome Project as shown in Fig. 5; order and orientation for groups of genes/ESTs or DNA fragments; quantification of the size of amplified or duplicated fragments of special genes or chromosomal regions or integrated foreign inserts; identification or exclusion of genes or chromosomal regions defined by genetic markers; comparison of evolutionary conserved regions among various species; illustration of multiple repetitive sequences within chromosomal regions and direct visualization of genome organization; and length measurement of telomeric regions for the study of chromosomal packaging.

See also Body Expression Map of Human Genome; Gene Distribution in the Human Genome; Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Beatty, B., Mai, S., Squire, J. (Eds.) (2002) *FISH: A practical Approach*, Oxford University Press, London.

Heng, H.H.Q., Tsui, L.-C., Windle, B., Parra, I. (1995) High resolution FISH analysis in: Dracopoli, N., Haines, J., Korf, B., Moir, D., Morton, C., Seidman, C., Seidman, J., Smith, D. (Eds.) *Current Protocols in Human Genetics*, John Wiley and Sons, New York, pp. 4.5.1–4.5.25.

Heng, H.H.Q., Tsui, L.-C. (1994) FISH detection on DAPI-banded chromosomes, in: Choo, K.H.A. (Ed.) *Methods in Molecular Biology*, Vol. 33, Humana Press, Totowa, pp. 35–49.

Knuutila, S., Bjorkqvist, A.-M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larramendy, M.L., Tapper, J., Pere, H., El-Rifai, W.E., Hemmer, S., Wasenius, V.-M., Vidgren, V. Zhu, Y. (1998) DNA copy number amplifications in human neoplasms: Review of comparative genomic hybridization studies. *Am J Path* **152**, 1107–1123.

Lebo, R.V., Su, Y. (1994) Positional cloning and multicolor *in situ* hybridization, in: Choo, K.H.A. (Ed.) *Methods in Molecular Biology* Vol. 33, Humana Press, Totawa, pp. 409–438.

Leversha, M.A. (1997) Fluorescence *in situ* hybridization, in: Dear, P.H. (Ed.) *Genome mapping: A practical approach*, Oxford University Press, London, pp. 199–225.

Leversha, M.A. (2001) Mapping of genomic clones by fluorescence *in situ* hybridization. in: Starkey, M.P. and Elaswarapu, R. (Eds.) *Methods in Molecular Biology*, Vol. 175, Humana Press, Totowa, pp. 109–127.

Trask, B., (1999) Flourescence in situ hybridization, in: Birren, B., Green, E.D., Hieter, P., Klapholz, S., Myers, R.M., Riethman, H., Roskams, J. (Eds.) *Genome Analysis: A Laboratory Manual*, Volume 4, Cold Spring Harbor Laboratory Press, New York.

Primary Literature

Al-Bayati, H.K., Duscher, S., Kollers, S., Rettenberger, G., Fries, R., Brenig, B. (1999) Construction and characterization of a porcine P1-derived artificial chromosome (PAC) library covering 3.2 genome equivalents and cyntenic assignment of six type I and type II loci. *Mamm Genome* **10**(6), 569–572.

Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J., Eichler, E.E. (2001) Segmental duplications: Organization and impact within the current human genome project assembly. *Genome Research* **11**, 1005–1017.

Baldini, A., Ross, M., Nizetic, D., Vatcheva, R., Lindsay, E.A., Lehrach, H., Siniscalco, M. (1992) Chromosomal assignment of human YAC clones by fluorescence in situ hybridization: use of single-yeast-colony PCR and multiple labeling. *Genomics* **14**, 181–184.

Bensimon, A., Simon, A., Chiffaudel, A., Croquette, V., Heslot, F., Bensimon, D. (1994) Alignment and sensitive detection of DNA by a moving interface. *Science* **265**, 2096–2098.

Brandriff, B.F., Gordon, L.A., Trask, B.J. (1991) DNA sequence mapping by fluorescence in situ hybridization. *Environmental and Molecular Mutagenesis* **18**, 259–262.

Brandriff, B., Gordon, L., Trask, B. (1991) A new system for high-resolution DNA sequence mapping in interphase pronuclei. *Genomics* **10**, 75–82.

Burkholder, G.D. (1993) The basis of chromosome banding. *Applied Cytogenetics* **19**, 181–186.

Conti, C., Bensimon, A. (2002) A combinatorial approach for fast, high-resolution mapping, *Genomics* **80**, 135–137.

Du Manoir, S., Speicher, M.R., Joos, S., et al. (1993) Detection of complete and partial chromosome gains and losses by comparative genomic in situ hybridization. *Hum Genet* **90**, 590–610.

Dunham, I., Shimizu, N., Roe, B.A., Chissoe, S., Hunt, A.R., Collins, J.E., Bruskiewich, R., Beare, D.M., Clamp, M., Smink, L.J., Ainscough, R., Almeida, J.P., Babbage, A., Bagguley, C., Bailey, J., Barlow, K., Bates, K.N., Beasley, O., Bird, C.P., Blakey, S., Bridgeman, A.M., Buck, D., Burgess, J., Burrill, W.D., O.'Brien, K.P., et al. (1999) The DNA sequence of human chromosome 22, *Nature* **402**, 489–495.

Dutrillaux, B., Viegas-Pequignot, E. (1981) High resolution R- and G-banding on the same preparation. *Hum Genet* **57**, 93–95.

Eppig, J.T. (1996) Comparative maps:adding pieces to the mammalian jigsaw puzzle. *Curr Opin Genet Dev* **6**(6), 723–730.

Fan, Y.S., Davis, L., Shows, T. (1990) Mapping small DNA sequences by fluorescence in situ hybridization directly on banded metaphase chromosomes. *Proc Natl Acad Sci USA* **87**, 6223–6227.

Fauth, C., Speicher, M.R. (2001) Classifying by colors: FISH-based genome analysis. *Cytogenet. Cell Genet.* **93**, 1–10.

Fidlerova, H., Senger, G., Kost, M., Sanseau, P., Sheer, D. (1994) Two simple procedures for releasing chromatin from routinely fixed cells for fluorescence in situ hybridization. *Cytogenet Cell Genet* **65**, 203–205.

Florijn, R.J., Bonden, L.A.J., Vrolijk, H., et al. (1995) High resolution DNA fiber-FISH for genomic DNA mapping and color bar-cording of large genes. *Hum Mol Genet* **4**, 831–836.

Gervasini, C., Bentivegna, A., Venturin, M., Corrado, L., Larizza, L., Riva, P. (2002) Tandem duplication of the NF1 gene detected by high-resolution FISH in the 17q11.2 region, *Hum Genet.* **111**, 465–7.

Haaf, T., Ward, D.C. (1994) Structural analysis of α-satellite DNA and centromere proteins using extended chromatin and chromosomes, *Hum. Mol. Genet.* **3**, 697–709.

Haaf, T. (1996) High-resolution analysis of DNA replication in released chromatin fibers containing 5-bromodeoxyuridine, *Biotechniques* **21**, 1050–4.

Heiskanen, M., Hellsten, E., Kallioniemi, O.-P., et al. (1995) Visual mapping by fiber FISH. *Genomics* **30**, 31–36.

Heiskanen, M., Karhu, R., Hellsten, E., Peltonen, L., Kallioniemi, O.P., Palotie, A. (1994) High resolution mapping using fluorescence in situ hybridization to extended DNA fibers prepared from agarose-embedded cells, *BioTechniques* **17**, 928–934.

Heiskanen, M., Peltonen, L., Palotie, A. (1996) Visual mapping by high resolution FISH, *Trends Genet.* **12**, 379–382.

Heng, H.H.Q., Squire, J., Tsui, L.-C. (1992) High resolution mapping of mammalian genes by in situ hybridization to free chromatin, *Proc. Natl. Acad. Sci. USA* **89**, 9509–9513.

Heng, H.H.Q., Tsui, L.C., (1993) Modes of DAPI banding and simultaneous in situ hybridization. *Chromosoma* **102**, 325–332.

Heng, H.H.Q., Chamberlain, J.W., Shi, X.-M., Spyropoulos, B., TsuI, L.-C., Moens, P.B. (1996) Regulation of meiotic chromatin loop size by chromosomal position, *Proc. Natl. Acad. Sci. USA* **93**, 2795–2800.

Heng, H.H.Q., Spyropoulos, B., Moens, P.B. (1997) FISH technology in chromosome and genome research, *Bioessays* **19**, 75–84.

Heng, H.H.Q., Tsui, L.C. (1998) High resolution free chromatin/DNA fiber fluorescent in situ hybridization. *J of Chromatography* **806**, 219–229.

Heppell-Parton, A.C., Albertson, D.G., Fishpool, R., Rabbitts, P.H. (1994) Multicolour fluorescence in situ hybridization to order small, single-copy probes on metaphase chromosomes. *Cytogenet Cell Genet* **66**, 42–47.

Herrick, J., Xavier, M., Conti, C., Schurra, C., Bensimon, A. (2000) Quantifying single gene copy number by measuring fluorescent probe lengths on combed genomic DNA, *Proc. Natl. Acad. Sci. USA* **97**, 222–227.

Hirai, M., Suto, Y., Kanoh, M. (1994) A method for simultaneous detection of fluorescent G-bands and in situ hybridization signals. *Cytogenet Cell Genet.* **66**, 149–151.

Horelli-Kuitunen, N., Aaltonen, J., Yaspo, M.L., Eeva, M., Wessman, M., Peltonen, L., Palotie, A. (1999) Mapping ESTs by fiber-FISH. *Genome Res* **9**, 62–71.

Houseal, T.W., Dackowski, W.R., Landes, G.M., Klinger, K.W. (1994) High resolution mapping of overlapping cosmids by fluorescence in situ hybridization. *Cytometry* **15**, 193–198.

Inoue, K., Osaka, H., Imaizumi, K., Nezu, A., Takanashi. J., Arij, J., Murayama, K., Ono, J., Kikawa, Y., Mito, T., Shaffer, L., Lupski, J. (1999) Proteolipid protein gene duplications causing Pelizaeus-Merzbacher Disease: molecular mechanism and phenotypic manifestations, *Ann. Neurol.* **45**, 624–632.

Ioannou, P.A., Amemiya, C.T., Garnes, J., et al. (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet* **6**, 84–89.

Jiang, F., Lin, F., Price, R., Gu, J., Medeiros, L.J., Zhang, H.Z., Xie, S.S., Caraway, N.P., Katz, R.L. (2002) Rapid detection of IgH/BCL2 rearrangement in follicular lymphoma by interhase fluorescence in situ hybridization with bacterial artificial chromosome probes, *J. Mole. Diagn.* **4**, 144–149.

Joos, S., Fink, T.M., Ratsch, A., Lichter, P. (1994) Mapping and chromosome analysis: the potential of fluorescence in situ hybridization. *J Biotech* **35**, 135–153.

Kallioniemi, A., Kallioniemi, O.-P, Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F., Pinkel, D. (1992) Comparative genome hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258**, 818–821.

Kallioniemi, A., Kallioniemi, O.-P., Piper, J., et al. (1994) Detection and mapping of amplified DNA sequences in breast cancer by comparative genome hybridization. *Proc Natl Acad Sci USA* **91**, 2156–2160.

Karhu, R., Rummukainen, J, Lorch, T., Isola, J. (1999) Four-color CGH: A new method for quality control of comparative genomic hybridization. *Genes, Chromosomes & Cancer* **24**, 112–118.

Korenberg, J.R., Chen, X.-N. (1995) Human cDNA mapping using a high resolution R-banding technique and fluorescence in situ hybridization. *Cytogen Cell Genet.* **69**, 196–200.

Korenberg, J.R., Chen, X.N., Adams, M.D., Venter, J.C. (1995) Toward a cDNA map of the human genome. *Genomics* **29**, 364–370.

Korenberg, J.R., Yang-Feng, T., Schreck, R., Chen, X.N. (1992) Using fluorescence in situ hybridization (FISH) in genome mapping. *Trends Biotechnol* **10**, 27–32.

Langer, P.R., Waldrop, A.A., Ward, D.C. (1981) Enzymatic synthesis of biotin-labeled polynucleotides: novel nucleic acid affinity probes. *Proc Natl Acad Sci USA* **78**, 6633–6637.

Lawrence, J.B., Villnave, C.A., Singer, R.H. (1988) Interphase chromatin and chromosome gene mapping by fluorescence detection of in situ hybridization reveals the presence and orientation of two closely integrated copies of EBV in a human lymphoblastoid cell line. *Cell* **52**, 51–61.

Lawrence, J.B. (1990). A fluorescence in situ hybridization approach for gene mapping and the study of nuclear organization, *Genome Analysis* **1**, 1–38.

Lawrence, J.B., Singer, R., McNeil, J. (1990) Interphase and metaphase resolution of different distances within the human dystrophin gene. *Science* **249**, 928–932.

Lemieux, N., Malfoy, B., Fetni, R., Muleris, M., Vogt, N., Richer, C.L. (1994) In situ hybridization approach at infragenic level on metaphase chromosomes. *Cytogenet Cell Genet* **66**, 107–112.

Lestou, V.S., Strehl, S., Lion, T., Gadner, H., Ambros, P.F. (1996) High-resolution FISH of the entire integrated Epstein-Barr virus genome on extended human DNA, *Cytogenet. Cell Genet.* **74**, 211–217.

Lichter, P., Cremer, T., Borden, J., Manuelidis, L., Ward, D.C. (1988) Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum Genet* **80**, 224–234.

Lichter, P., Ledbetter, S.A., Ledbetter, D.H., Ward, D.C. (1990) Fluorescence in situ hybridization with Alu and L1 polymerase chain reaction probes for rapid characterization of human chromosomes in hybrid cell lines. *Proc Natl Acad Sci USA* **87**, 6634–6638.

Lichter, P., Tang, C.C., Call, K., Hermanson, G., Evans, G.A., Housman, D, Ward, D.C. (1990) High resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* **247**, 64–69.

Michalet, X., Ekong, R., Fougerousse, F., Rousseaux, S., Schurra, C., Hornigold, N., van Slegtenhorst, M., Wolfe, J., Povey, S., Bechmann, J.S., Bensimon, A. (1997) Dynamic molecular combing: stretching the whole

human genome for high resolution studies. *Science* **277**, 1518–1523.

Nath, J., Johnson, K.L. (1999) A review of fluorescence in situ hybridization (FISH): current status and future prospects. *Biotech & Histochem* **75**, 54–78.

Nederlof, P.M., Robinson, D., Abuknesha, R., Wiegant, J., Hopman, A.H.N., Tanke, H.J., Raap, A.K. (1989) Three-color fluorescence in situ hybridization for the simultaneous detection of multiple nucleic acid sequences. *Cytometry* **10**, 20–27.

Nederlof, P.M., van der Flier, S., Wiegant, J., Raap, A.K., Tanke, H.J., Ploem, J.S., van der Ploeg, M. (1990) Multiple fluorescence in situ hybridization. *Cytometry* **11**, 126–131.

Palotie, A., Heiskanen, M., Laan, M., Horelli-Kuitunen. (1996) High-resolution fluorescence in situ hybridization: a new approach in genome mapping. *Finn Med Soc Ann Med* **28**, 101–106.

Pandita, A., Godbout, R., Zielenska, M., Thorner, P., Bayani, J., Squire, J.A. (1997) Relational mapping of MYCN and DDX1 in band 2p24 and analysis of amplicon arrays in double minute chromosome and homogeneously staining regions by use of free chromatin FISH, *Genes Chromosomes Cancer* **20**, 243–252.

Parra, I., Windle, B. (1993) High resolution visual mapping of stretched DNA by fluorescent hybridization. *Nature Genet* **5**, 17–21.

Pinkel, D., Straume, T., Gray, J.W. (1986) Cytogenic analysis using quantitative, high sensitivity, fluorescence hybridization. *Proc Natl Aad Sci USA* **83**, 2934–2938.

Pinkel, D., Landegent, J., Collins, C., Fuscoe, J., Segraves, R., Lucas, J., Gray, J.W. (1988) Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proc Natl Acad Sci USA* **85**, 9138–9142.

Popescu, N.C., Zimonjic, D.B. (1997) Molecular cytogenetic characterization of cancer cell alterations. *Cancer Genet. Cytogenet.* **93**, 10–21.

Poulsen, T.S., Silahtaroglu, A.N., Gisselo, C.G., Gaarsdal, E., Rasmussen, T., Tommerup, N., Johnsen, H.E. (2001) Detection of illegitimate rearrangement within the immunoglobulin locus on 14q32.3 in B-cell malignancies using end-sequenced probes, *Genes Chromosomes Cancer* **32**, 265–274.

Raap, A.K., van de Corput, M.P., Vervenne, R.A., van Gijlswijk, R.P., Tanke, H.J., Wiegant, J. (1995) Ultra-sensitive FISH using peroxidase-mediated deposition of biotin- or fluorochrome tyramides. *Hum Mol Genet* **4**, 529–534.

Raderschall, E., Golub, E.I., Haaf, T. (1999) Nuclear foci of mammalian recombination proteins are located at single-stranded DNA regions formed after DNA damage, *Proc. Natl. Acad. Sci. USA* **96**, 1921–1926.

Rapp, A. (1998) Advances in fluorescence in situ hybridization, *Mutation Research* **400**, 287–298.

Raudsepp, T., Mariat, D., Guerin, G., Chowdhary, B.P. (2001) Comparative FISH mapping of 32 loci reveals new homologous regions between donkey and horse karyotypes. *Cytogenet Cell Genet* **94**, 180–185.

Rettenberger, G., Klett, C., Zechner, U., Kunz, J., Vogel, W., Hamesiter, H. (1995) Visualization of the conservation of synteny between humans and pigs by heterologous chromosomal painting. *Genomics* **26**, 372–378.

Reid, T., Arnold, N., Ward, D.C., Wienberg, J. (1993) Comparative high-resolution mapping of human and primate chromosomes by fluorescence in situ hybridization. *Genomics* **18**, 381–386.

Ried, T., Baldini, A., Rand, T.C., Ward, D.C. (1992) Simultaneous visualization of seven different DNA probes by in situ hybridization using combinatorial fluorescence and digital imaging microscopy. *Proc Natl Acad Sci USA* **89**, 1388–1392.

Riemersma, S.A., Jordanova, E.S., Schop, R.F., Philippo, K., Looijenga, L.H., Schuuring, E., Kluin, P.M. (2000) Extensive genetic alterations of the HLA region, including homozygous deletions of HLA class II genes in B-cell lymphomas arising in immune-privileged sites, *Blood* **96**, 3569–3577.

Rottger, S., Yen, P.H., Schempp, W. (2002) A fiber-FISH contig spanning the non-recombining region of the human Y chromosome. *Chromosome Res.* **10**, 621–35.

Scherthan, H., Cremer, T., Amason, U., Weier, H.-U., Lima-de-Faria, A., Fronicke, L. (1994) Comparative chromosome painting discloses homologous segments in distantly related mammals. *Nat Genet* **6**, 342–347.

Schriml, L.M., Padilla-Nash, H.M., Coleman, A., Moen, P., Nash, W.G., Menninger, J., Jones, G., Tied, T., Dean, M. (1999) Tyramide

signal amplification (TSA)-FISH applied to mapping PCR-labeled probes less than 1 kb in size. *BioTechniques* **27**, 608–613.

Schrock, E., duManoir, S., Veldman, T., et al. (1996) Multicolor spectral karyotyping of human chromosomes. *Science* **273**, 494–497.

Senger, G., Ragoussis, J., Trowsdale, J., Sheer, D. (1991) Fine mapping of the human MHC class II region within chromosome band 6p21 and evaluation of probe ordering using interphase fluorescence in situ hybridization. *Am J Hum Genet* **48**, 1–15.

Senger, G., Jones, T.A., Fidlerova, H., Sanseau, P., Trowsdale, J., Duff, M., Sheer, D. (1994) Released chromatin: linearized DNA for high resolution fluorescence in situ hybridization. *Hum Mol Genet* **3**, 1275–1280.

Shizuya, H., Birren, B., Kim, U.-U. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc Natl Acad Sci USA* **89**, 8794–8797.

Slijepcevic P. (2001) Telomere length measurement by Q-FISH. *Methods Cell Sci* **23**(1–3), 17–22.

Speel, E.J.M., Ramaekers, F.C.S., Hopman, A.H.N. (1997) Sensitive multicolor fluorescence in situ hybridization using catalyzed reporter deposition (CARD) amplification. *JHistochem Cytochem* **45**, 1439–1446.

Speicher, M.R., Ballard, S.G., Ward, D.C. (1996) Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature Genet* **12**, 368–375.

Speicher, M.R., duManoir, S., Schrock, E., et al. (1993) Molecular cytogenetic analysis of formalin-fixed, paraffin-embedded solid tumors by comparative genomic hybridization after universal DNA-amplification. *Hum Mol Genet* **2**, 1907–1914.

Takahashi, E., Koyama, K., Hirai, M., Itoh, H., Nakamura, Y. (1995) A high resolution cytogenetic map of human chromosome 2: localization of 434 cosmid markers by direct R-banding fluorescence in situ hybridization. *Cytogenet Cell Genet* **68**, 112–114.

Tanke, H.J., Florijn, R.J., Wiegant, J., Raap, A.K., Vrolijk, J. (1995) CCD microscopy and image analysis of cells and chromosomes stained by fluorescence in situ hybridization. *Histocheml J* **27**, 4–14.

Trask, B., Pinkel, D., van den Engh, G. (1989) The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. *Genomics* **5**, 710–717.

Trask, B.J. (1991) Gene mapping by in situ hybridization. *Genet and Dev* **1**, 82–87.

Trask, B.J., Massa, H., Kenwrick, S., Gitschier, J. (1991) Mapping of human chromosome Xq28 two-color fluorescence in situ hybridization of DNA sequences to interphase cell nuclei. *Am J Hum Genet* **48**, 1–15.

Trask, B.J., Massa, H., Brand-Arpon, V., Chan, K., Friedman, C., Nguyen, O.T., Eichler, E., van den Engh, G., Rouquier, S., Shizuya, H., Giorgi, D. (1998) Large multi-chromosomal duplications encompass many members of the olfactory receptor gene family in the human genome. *Hum Mol Genet* **7**, 2007–2020.

Trower, M.K., Orton, S.M., Purvis, I.J., Sanseau, P., Riley, J., Christodoulou, C., Burt, D., See, C.G., Elgar, G., Sherrington, R., Rogaev, E.I., St George-Hyslop, P., Brenner, S., Dykes, C.W. (1996) Conservation of synteny between the genome of the pufferfish (Fugu rubripes) and the region on human chromosome 14 (14q24.3) associated with familial Alzheimer disease (AD3 locus). *Proc Natl Acad Sci USA* **93**, 1366–1369.

Vaandrager, J.W., Schuuring, E., Rapp, T., Phillippo, J., Klelverda, K., Kluin, P. (2000) Interphase FISH detection of BCL2 rearrangement in follicular lymphoma using breakpoint-flanking probes. *Genes, Chromo & Cancer* **27**, 85–94.

Valdes, J., Tagle, D.A. (1997) Gene mapping goes from FISH to surfing the net. *Methods in Mol Biol* **68**, 1–10.

van den Engh, G., Sachs, R. and Trask, B. (1992) Estimating genomic distances from DNA sequence location in cell nuclei by a random walk model. *Science* **257**, 1410–1412.

van Ommen, G.J., Breuning, M.H., Raap, A.K. (1995) FISH in genome research and molecular diagnostics, *Curr Opin Genet Dev*. **5**, 304–308.

Vesa, J., Hellsten, E., Verkruyse, L.A., Camp, L.A., Rapola, J., Santavuori, P., Hofmann, S.L., Peltonen, L. (1995) Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. *Nature*. **376**, 584–587.

Viegas-Pequignot, E., Dutrillaux, B., Magdelenat, H., Coppey-Moisan, M. (1989) Mapping of single-copy DNA sequences on human chromosomes by in situ hybridization with

biotinylated probes: Enhancement of detection sensitivity by intensified-fluorescence digital-imaging microscopy. *Proc Natl Acad Sci USA* **86**, 582–586.

Volante, M., Pecchioni, C., Bussolati, G. (2000) Post-incubation heating significantly improves tyramide signal amplification. *The J of Histochem & Cytochem* **48**, 1583–1585.

Weier, H.U., Wang, M., Mullikin, J.C., Zhu, Y., Cheng, J.-F., Greulich, K.M., Bensimon, A., Gray, J.W. (1995) Quantitative DNA fiber mapping, *Hum. Mol. Genet.* **4**, 1903–1910.

Weier, H.U. DNA fiber mapping techniques for the assembly of high-resolution physical maps. (2001) *J Histochem & Cytochem* **49**, 939–948.

Wiegant, J., Ried, T., Nederlof, P., van der Ploeg, M., Tanke, H.J. and Raap, A.K. (1991) In situ hybridization with fluoresceinated DNA. *Nucleic Acids Res* **19**, 3237–3241.

Wiegant, J., Kalle, W., Mullenders, L., Brookes, S., Hoovers, J.M.N., Dauwerse, J.G., van Ommen, G.J.B. and Raap, A.K. (1992) High-resolution in situ hybridization using DNA halo preparations. *Hum Mol Genet* **1**, 587–592.

Ye, J., Lu, W., Moens, P., Liu, G., Bremer, S., Wang Y., Hughes, M., Krawetz, S., and Heng, H.H.Q. (2001) The combination of SKY and specific loci detection with FISH or Immunostaining. *Cytogenet Cell Genet* **93**, 195–202.

Yokota, H., Singer, M.J., van den Engh, G.J., Trask, B.J. (1997) Regional differences in the compaction of chromatin in human G0/G1 interphase nuclei. *Chromosome Res* **5**, 157–166.

Gene Targeting

Michael M. Seidman[1] *and John H. Wilson*[2]
[1] *National Institute on Aging, NIH, Baltimore, MD, USA*
[2] *Baylor College of Medicine, Houston, TX, USA*

1 Gene Targeting 175
1.1 The Problem 175

2 Homologous Recombination 175
2.1 Introduction 175
2.2 Repair of Broken Replication Forks 177
2.3 Targeted Recombination 179
2.4 Stimulation of Targeted Recombination 183

3 Polyamides 185
3.1 Netropsin and Distamycin 185
3.2 The Development of Sequence Specific Minor Groove Ligands 186
3.3 Affinity and Specificity 188
3.4 Biological Applications 190

4 Triple Helix–Forming Oligonucleotides 191
4.1 Triple Helical DNA 191
4.2 Triplex Formation under Physiological Conditions 192
4.3 Oligonucleotide Modifications Improve TFO Activity 193
4.4 Activity of TFOs *In Vivo* 194
4.5 TFOs Linked to DNA Reactive Compounds 194
4.6 Triplex Formation, by itself, can Promote Mutagenesis 195
4.7 Triplex-induced Recombination and Gene Correction 196
4.8 Expansion of the Triplex Binding Code 196

5 Mutation Repair by Small DNA Fragments 197
5.1 Small Fragment Homologous Replacement 197
5.2 Chimeric Oligonucleotides 198

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

6 Peptide Nucleic Acids 198

7 Summary 199

Bibliography 200
Books and Reviews 200
Primary Literature 201

Keywords

Gene Targeting
The recognition and binding of a specific sequence in the chromosome of living cells by a reagent constructed for that application. The purpose may be sequence manipulation or modulation of transcription.

Homologous Recombination
A process through which disrupted replication forks are restored, double-strand breaks are repaired, and genetic exchange occurs.

Polyamide
A short term for a synthetic molecule consisting of pyrrole and imidazole units joined by amide linkages. These bind in a sequence-restricted manner in the minor groove of an intact duplex.

DNA Triple Helix
A three-stranded structure in which a third strand of nucleic acid lies in the major groove of an intact polypurine:polypyrimidine duplex. Binding is sequence-specific, and stabilized by hydrogen bonds between the bases in the third strand and the purine strand of the duplex.

Peptide Nucleic Acid (PNA)
DNA analog in which the sugar phosphate backbone has been replaced by peptide linkages.

The recognition and binding of a specific sequence in the chromosome of living cells by a reagent constructed for that application. The purpose may be sequence manipulation or modulation of transcription.

1 Gene Targeting

Gene targeting reagents bind specific chromosomal sequences in living cells. They can, or would, be used to manipulate genomic sequences for gene inactivation, modification, or restoration of function; suppress or activate gene transcription; deliver site-specific DNA damage; and probe chromatin and chromosome structure. Practical applications include the construction of transgenic cell lines and animals for research and commercial purposes. They are also under development as pharmaceuticals for the treatment of infectious diseases, gene therapy of genetic disorders and so on. Current strategies are based on either recombinant DNA technology (to produce DNA or protein constructs) or synthetic molecules that are based on oligonucleotides or natural product antibiotics. While many technologies are the subject of active research, only one, homologous recombination (HR), is widely employed at this time. In this review, we will discuss the current status of gene targeting strategies and challenges to the development of gene targeting reagents for use in mammalian, including human, cells. As will become apparent, most of the literature describes the work in progress.

1.1 The Problem

Gene targeting reagents must recognize and bind a specific DNA sequence in living cells, and be in residence long enough to provoke the desired outcome. Thus, a successful reagent must be stable in a physiological environment; be deliverable to the nucleus of the target cell; be able to bind a target in the context of mammalian chromatin structure; have orders of magnitude preference for the intended target sequence over any other sequence; and must find the target with sufficient frequency to be of practical use. Normal DNA metabolism requires sequence recognition for HR and the regulation of transcription. The molecules involved in these processes, DNA, and the proteins that regulate transcription serve as explicit models for some of the reagents under development. One fundamental consideration is the nature of the recognition process (Fig. 1). Reagents that recognize sequence in an intact duplex must do so from the vantage point of the major groove, which is rich in recognition sites, or the minor groove, which is relatively poor. Those designed to ultimately pair with the target sequence must engage a pathway through which the complementary strand becomes accessible and available for hybridization.

2 Homologous Recombination

2.1 Introduction

Targeted gene modification was first demonstrated in mammalian cells nearly 20 years ago. These initial studies targeted chromosomal genes in established cell lines and in mouse embryonic stem (ES) cells. Collectively, they laid the foundations for subsequent studies of gene function in cells and mice, and for the creation of mouse models of human disease. They also raised the enticing possibility that genetic defects might be precisely corrected as the ultimate approach to gene therapy. These considerations have provided the impetus to understand the

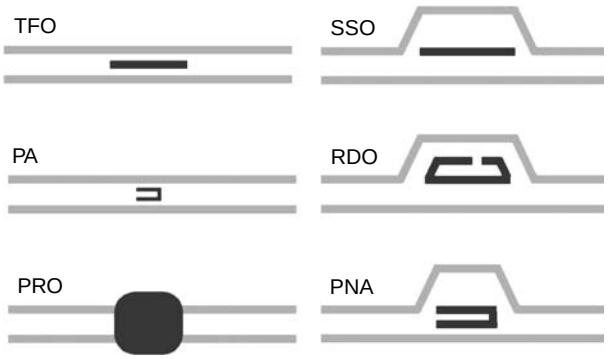

Fig. 1 Recognition of DNA sequences by various gene-targeting reagents. Targeting reagents either bind to duplex DNA or to the separated strands. Triplex forming oligonucleotides (TFOs) recognize their target DNA via interactions in the major groove. Polyamides (PAs) bind through the minor groove. Proteins (PROs) can make binding contacts in both grooves, although most commonly with moieties in the major groove. Single-stranded oligonucleotides (SSOs) pair with one of the separated strands and may mimic an intermediate in HR (see Fig. 3). It may be that small fragment homologous recombination (SFHR) proceeds through such an intermediate, but mechanistic studies have not been done. RNA:DNA oligonucleotides (RDOs) are envisioned to bind to both separated strands, but such binding has not been demonstrated. Peptide nucleic acids (PNAs) clamp onto one of the separated strands by Watson–Crick pairing and by forming Hoogsteen hydrogen bonds with moieties in the major groove of the PNA:DNA hybrid.

mechanism of gene targeting and to define the key experimental parameters for this technology.

Targeted gene modification can occur by HR and engages the cells own recombination machinery, which has been fine-tuned by evolution. For that reason, it possesses many of the attributes of an ideal system. HR occurs with very high specificity and selectivity, discriminating readily between genes that differ by only a few percent, and is extremely accurate, introducing very few errors. Two types of vectors – replacement (ends-out) and insertion (ends-in) – have been used for targeted recombination in mammalian cells (Fig. 2), with replacement vectors being far more common. The main drawback of targeted recombination in mammalian cells is low frequency; roughly 1 event per 10^5 to 10^7 treated cells under commonly used conditions. This presents a problem for therapeutic applications where targeted modification of a high fraction of a cell population is critical. Low frequency is less of a barrier for genome manipulation in proliferating cells in culture because the rare, properly modified cells can be identified and grown into a pure population. Nevertheless, a substantial increase in the frequency of targeted recombination would make genome modification less tedious and improve the chances for targeted gene

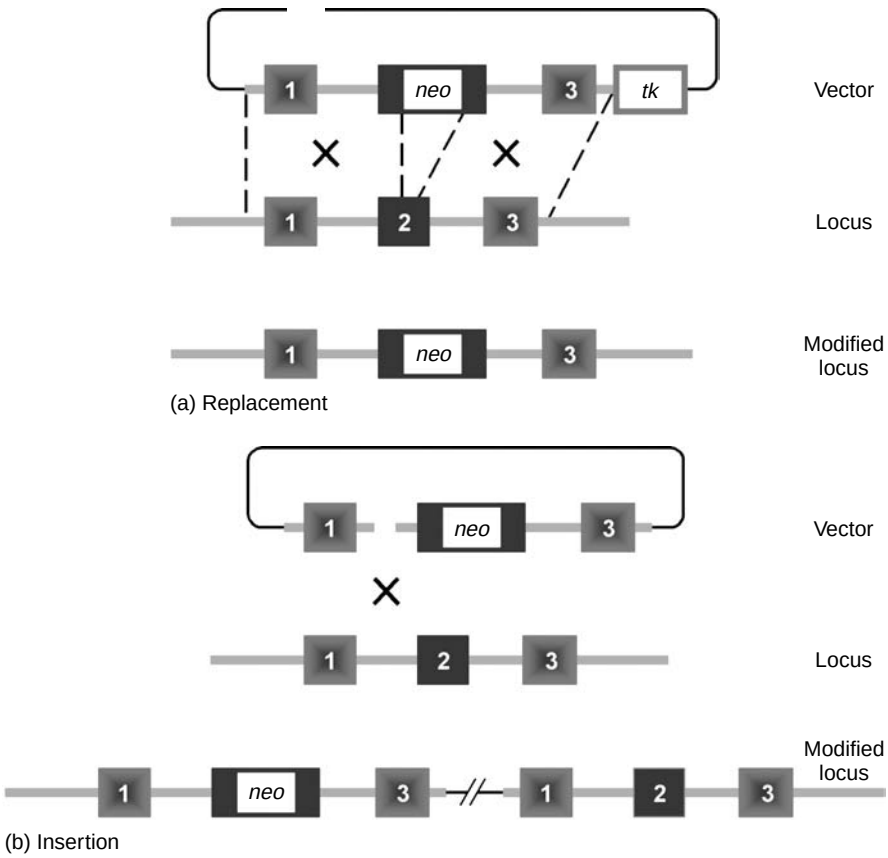

Fig. 2 Types of targeting vectors. (a) Replacement vector. Positive–negative selection is illustrated. Exon-2 is disrupted by incorporation of the gene for neomycin resistance (neo), which constitutes the positive selection marker. The herpesvirus *thymidine kinase* (*tk*) gene is included at one end of the homologous sequences and constitutes the negative selectable marker. The vector is linearized in the plasmid sequences (thin black line). HR in the homologous sequences on either side of the *neo* gene (indicated by Xs) replaces the chromosomal copy of exon-2 with the one from the plasmid vector. The *tk* gene is lost upon HR so that targeted cells survive selection against the *tk* gene. Random integrants (not shown) incorporate the *tk* gene and are killed by selection. (b) Insertion vector. The vector is linearized within the homologous DNA, in this instance between exons 1 and 2. HR at the double-strand ends inserts the vector into the locus, causing a duplication of the homologous sequences with the plasmid sequences located between the two copies. (After Muller, U. (1999) Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis, *Mech. Dev.* **82**, 3–21. Reprinted with permission of Elsevier Science Limited.)

therapy. Rational improvements depend on understanding the underlying process of HR. Clearly, genome manipulation cannot be the main function of HR in cells.

2.2 Repair of Broken Replication Forks

The initial demonstration that HR occurred in cultured mammalian cells was

somewhat surprising, as recombination had been thought about and studied mainly in the context of meiotic recombination during gamete formation in the germ line. Over time, it has become clear that HR in somatic cells is one of a number of cellular systems for dealing with DNA damage. It is specifically called into play when both strands of the DNA duplex have been damaged, as occurs, for example, with a double-strand break or when the strands have been cross-linked. Quantitatively, the most important role for HR is the recombinational repair of stalled or collapsed replication forks. In *Escherichia coli*, in the absence of the principal recombinase, RecA, about half the cells in a culture are dead, suggesting that a majority of replication forks are compromised in each replication cycle. In mammalian cells, with their much larger genomes, it is thought that replication forks stall more often, which is the likely reason why cells defective for the RecA homolog, Rad51, are inviable.

It is useful to consider, in overview, one of the ways by which HR can function to restart a broken replication fork because it provides an analog for the main reactions of targeted recombination. As shown in Fig. 3, when a replication fork suffers a break, or encounters one, it collapses, leaving a partially replicated duplex with a protruding single-stranded tail. With the aid of several accessory proteins, the Rad51 recombinase is loaded onto the single strand to form a helical filament of Rad51 with a central single strand of DNA. With the aid of additional accessory proteins, this structure coordinates the search for the corresponding sequence in the intact duplex. When the matching sequence is located, Rad51 catalyzes strand invasion, which pairs the invading single strand with

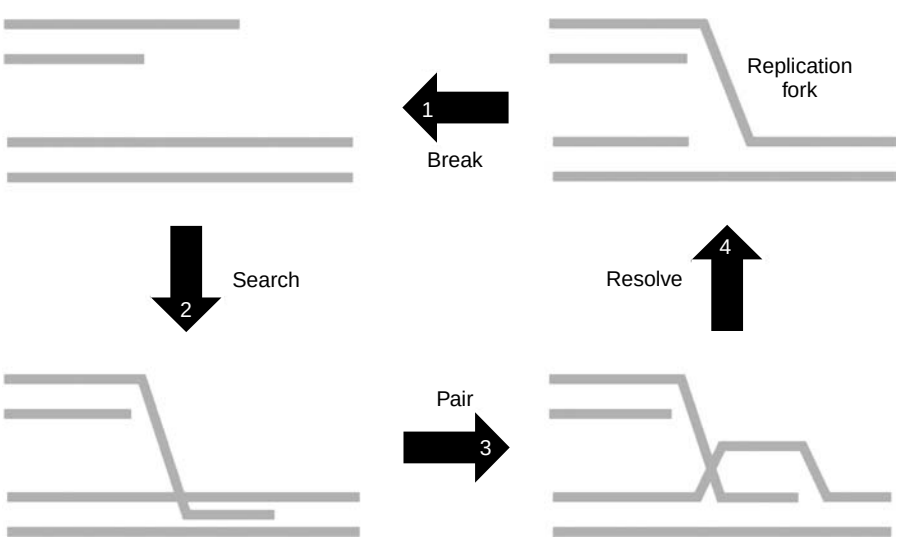

Fig. 3 Repair of a broken replication fork. The single-stranded tail at the broken end is coated with Rad51 recombinase, which then catalyzes the search for homology and the pairing of the single strand with its complement in the duplex. As shown in more detail in Fig. 4, this heteroduplex structure can be resolved by strand cleavage to regenerate a fork-like DNA structure. Several additional proteins are required to convert this branched structure into a functional replication fork.

its complement in the target duplex. This structure then serves as the substrate for a group of replication proteins that modify it to reconstitute the replication fork.

Two steps in this pathway – the search for homology and pairing with the complementary strand – provide structural analogies for the key steps in duplex recognition by all methods of gene targeting (Fig. 1). The search for homology and DNA pairing are thought to take place within the filament of Rad51 and single-stranded DNA, as originally proposed for RecA. In principle, the single-stranded DNA in the filament could approach the duplex DNA via the major groove or the minor groove. Studies using groove-specific DNA binding reagents, modified nucleotides, and cross-linkers, as well as structural analysis of the products, indicate that the initial attack most likely occurs through the minor groove. The search for homology is catalyzed by the recombinase, which binds to duplex DNA and transiently underwinds it, exposing the minor groove for interaction with the single strand in the Rad51 filament.

In the minor groove, there are insufficient hydrogen-bonding moieties to allow specific recognition of homology by base–base interactions. Instead, the recombinase apparently tests the single strand for homology with the duplex by catalyzing an exchange of pairing between bases in the duplex (mainly A and T) and those in the single strand. This trial pairing is accomplished by rotation of the bases through the minor groove in a plane perpendicular to the helix axis, breaking the Watson–Crick pairing in the duplex and reforming it, if possible, with the bases in the single strand. In regions of heterology, the imperfect pairing is rapidly reversed, allowing a new segment of duplex to be tested. When homology is encountered, the pairing between the single strand and its complement in the duplex is extended to form a heteroduplex (Fig. 3). The displaced single strand lies in the major groove of the newly formed heteroduplex. Although this is a triple-stranded structure, it is not the same as that in a triple helix, in which the duplex is intact (see below).

The search for homology and the formation of the heteroduplex intermediate are aided by Rad54 protein, which is a member of the Swi2/Snf2 family of DNA-dependent ATPases with helicase motifs. Rad51 and Rad54 interact *in vitro*. Rad54 acts upon the duplex target DNA to introduce negative and positive supercoiling. Domains of negative supercoiling around the target site would favor the unwinding of the duplex that occurs upon binding the Rad51 filament, which likely explains the ability of Rad54 to increase the efficiency of Rad51-mediated strand pairing. Both Rad51 and Rad54 are present in greater amounts in S-phase relative to the G1-phase, consistent with a role in the repair of replication forks.

2.3 Targeted Recombination

In its role in replication, HR must be nearly 100% effective in reconstituting broken or stalled replication forks. If the basic reactions of fork repair and targeted recombination are similar, then why are so few treated cells successfully modified? Examining some of the possibilities is illuminating; it records our step-by-step progress in understanding the nature of targeted recombination and points the way toward increasing efficiency.

One obvious difference between targeted recombination and fork repair is that the broken arm of a replication fork is already inside the nucleus, whereas the targeting

vector must be delivered from outside the cell. Two general methods – nuclear injection and mass delivery – have been used to introduce targeting vectors into cells. Early experiments using microinjection yielded about 1 targeted recombinant per 1000 injected cells. Mass delivery methods, which are vastly simpler and cheaper, typically yield targeted recombinants at frequencies 1000-fold less than microinjection. The basis for this difference is unclear. It does not appear to be a problem with transport to the nucleus, since 10 to 100% of treated cells can express genes transfected by the common methods of mass delivery such as electroporation, calcium phosphate precipitation, Fugene-6, and LipofectAmine. Thus, entry into the nucleus is not the fundamental barrier to efficient gene targeting.

A second, more subtle difference between targeted recombination and fork repair is the number of DNA ends. The broken arm of a replication fork has only one end, whereas a targeting vector has two. It was shown earlier on that linear vectors give several fold higher targeting frequencies than circular ones, as expected from targeting experiments in yeast. The presence of multiple ends, however, makes the transfected targeting vector a substrate for nonhomologous end joining (NHEJ), which along with HR is a major pathway for repair of double-strand breaks in mammalian cells. NHEJ operates throughout the cell cycle, whereas HR is prominent in mid to late S-phase and G2. A reasonable expectation is that NHEJ might compete with HR, effectively eliminating targeting substrates by circularization and multimerization. In addition, NHEJ is thought to be responsible for random integration, which occurs at roughly 1000-fold higher frequency than targeted recombination. The high background of random integration events can be minimized by any of the several tricks such as positive–negative selection that render most random integrants inviable (Fig. 2). Consistent with the idea that HR and NHEJ compete for substrates, repair of chromosomal double-strand breaks by HR is increased 2- to 25-fold when components of the NHEJ pathway are defective or inhibited. Surprisingly, however, neither the joining of transfected DNA ends nor random integration appears to be affected by the absence of NHEJ, and targeted recombination is not substantially increased. These observations and others suggest that there may be additional mechanisms for dealing with DNA ends. Although it is unclear to what extent NHEJ might affect recombination, the absence of a substantial effect in NHEJ-defective cells suggests that competition between NHEJ and HR is not the key to the low frequency of targeted recombination.

Another obvious difference between targeted recombination and fork repair is the proximity of the DNA to the target sequence. The DNA end at a broken replication fork is tethered in the general vicinity of the target sequence by virtue of its attachment to the oppositely directed replication fork. By contrast, the exogenous DNA must find its target among all the sequences in the genome: a daunting task. Nevertheless, the search for homology does not limit the frequency of targeted recombinants in mammalian cells. Varying the number of exogenous DNA molecules in the nucleus over a 1000-fold range (delivered by nuclear injection), or the number of targets in the genome by up to 400-fold had no significant effect on the frequency of targeted recombination. These counterintuitive results have been confirmed in competition experiments. Thus, the frequency of targeted

recombination is limited by something other than the search for homology. These results stand in stark contrast to those in yeast, where the process was found to be dependent on the number of target copies.

Targeted recombination and fork repair also differ significantly in the extent of homology, which, for the broken arm of a replication fork is as long as the replication bubble. For practical reasons, the length of homology in targeting vectors is limited to less than about 20 kb. Over this range, the frequency of targeting in mammalian cells shows a steep dependence on the length of homology, with a 2- to 3-fold change in length giving a 10-fold change in targeting frequency. Once again, these results are distinct from those in yeast, where gene targeting is linearly dependent on length of homology. They also stand in contrast to the linear length dependence of intrachromosomal HR in mammalian cells. The peculiar dependence of mammalian gene targeting on the length of homology suggests that the limiting step in the pathway is downstream of finding the target gene. In some undefined way, longer stretches of homology must exponentially improve the chances of a productive outcome.

Beyond the length of homology, there is also a potential difference in the quality of homology. A broken arm of a replication fork is identical to the intact arm (with the exception of very rare replication errors). The homologous segment of a recombination vector, however, is often isolated from nonisogenic sources and thus may carry nucleotide polymorphisms. Single-nucleotide differences between homologous sequences lead to mismatches in heteroduplex recombination intermediates. In bacteria, yeast, and mammalian cells, such mismatches reduce the frequency of HR in a way that is dependent on the mismatch-repair machinery and certain DNA helicases. In mammalian cells, use of nonisogenic DNA (up to about 1% different) can reduce the efficiency of targeting up to 20-fold. Although this can be a substantial effect, even isogenic DNA will function at very low absolute frequencies. In contrast to the effects of distributed mismatches, blocks of nonhomology at the ends or within the homologous regions (as shown in Fig. 2) have little or no effect.

Perhaps the most basic difference between fork repair and targeted recombination is the intended outcome. For reconstitution of the replication fork, the heteroduplex recombination intermediate must be resolved so that a fork is created (Fig. 4). For recombination, a further resolution is needed. Breakage of one strand at the fork can either release the exogenous DNA or integrate it into the chromosome to give a targeted replacement of the underlying DNA. In the parlance of recombination, release corresponds to a conversion event and integration corresponds to a crossover event. It is notable that repair of broken DNA in mammalian cells occurs with a very strong bias (up to 100-fold or more) against crossover events, perhaps as a protective mechanism against loss of heterozygosity. In mammalian cells, it seems that broken ends often invade a homologous duplex, prime DNA synthesis, and then disengage from the duplex. It is common to find randomly integrated targeting vectors that have been extended up to 10 kb or so at one or both ends by a prior homologous interaction with the target duplex. Thus, the inherent bias toward conversion-like resolution of targeting intermediates may constitute the principal barrier to gene targeting in mammalian cells.

The view that a preference for conversion limits the frequency of targeted recombination is supported by results in which

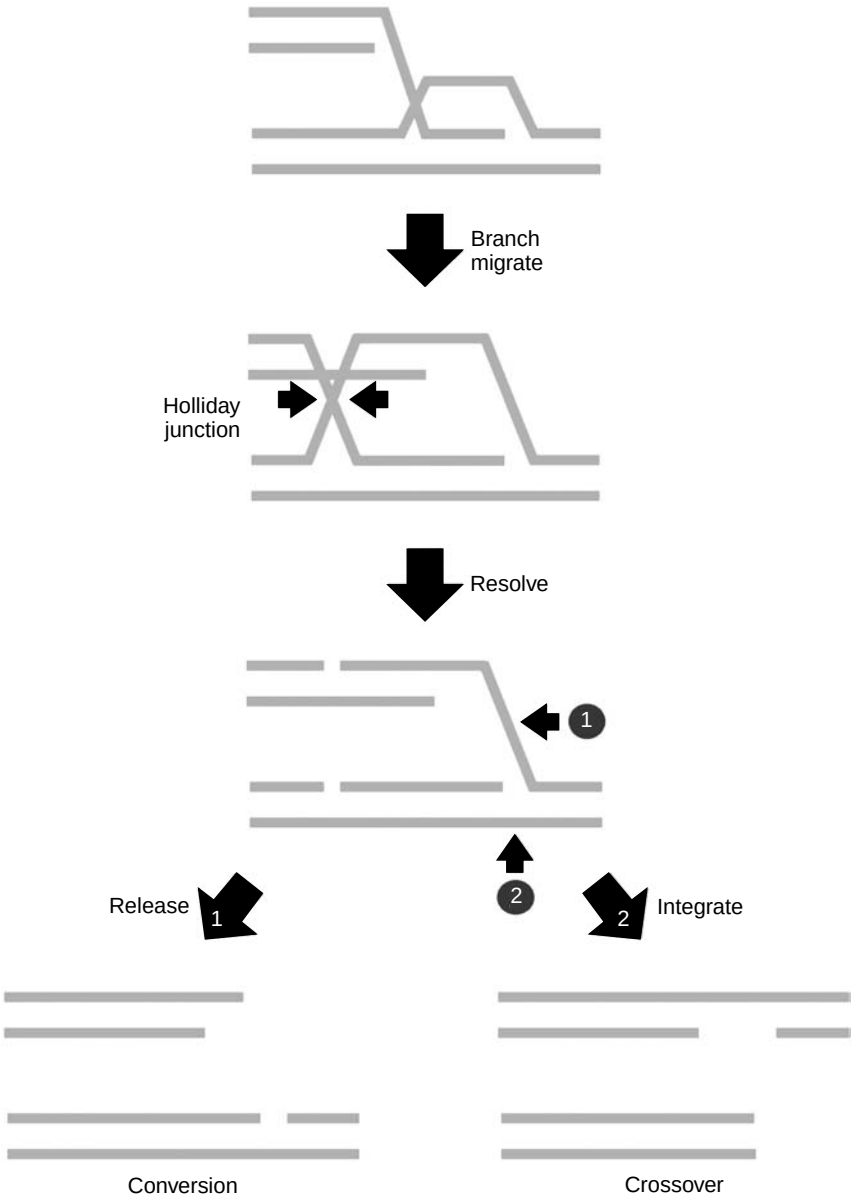

Fig. 4 Resolution of the initial heteroduplex intermediate. The initial paired structure can branch migrate to form a Holliday junction. Resolution of the Holliday junction by cleavage of the pair of strands indicated by the small arrows yields a branched structure equivalent to a replication fork. Cleavage of the branched structure at arrow 1 releases the targeting DNA, whereas cleavage at arrow 2 integrates the targeting DNA.

double-strand breaks are introduced into the target sequence. By embedding the recognition site for a rare-cutting endonuclease such as I-SceI in the chromosomal target, site-specific double-strand breaks can be introduced by transfection of an expression vector for I-SceI. When the targeting vector is included, targeted recombination can be stimulated up to 10 000-fold with the specific recombinants constituting several percent of treated cells. One or both of the chromosomal ends at the break are thought to invade the targeting duplex, extend themselves by DNA synthesis, and then disengage and pair with one another to reconstitute the duplex and eliminate the double-strand break. This mechanism, which is known as synthesis-dependent strand annealing (SDSA), provides a productive way of thinking about targeted recombination.

2.4 Stimulation of Targeted Recombination

An appreciation of the normal role HR plays in somatic cells, coupled with an understanding of the probable pathway for targeted recombination, allows one to devise rational strategies for increasing the frequency of the event. Three general approaches have been tried; altering levels of HR enzymes, introducing damage at the target site, and delivering DNA via viral infection. These approaches are not mutually exclusive, and significant improvements may require some combination of them. Below we discuss the rationale for each of those strategies and their current status.

In principle, it should be possible to increase frequencies by temporarily altering the cellular mix of HR enzymes in a way that favors targeted recombination. For example, one might try to overexpress enzymes that favor HR and inhibit those involved in competing pathways such as NHEJ. Thus far, such experiments have met with limited success. As mentioned above, inhibition of NHEJ by knocking out key components had no significant effect. Three HR enzymes – Rad51, Rad52, and Rad54 – have been tested by over expression. A fourfold overexpression of the Rad51 recombinase stimulated targeted recombination some two- to fourfold. By contrast, overexpression of Rad52 protein, which helps to load Rad51 onto single-stranded DNA, was inhibitory by about twofold. Overexpression of Rad51 and Rad52 together had little effect on targeting frequencies. Finally, overexpression of Rad54, which helps prepare the target duplex for invasion, stimulated recombination by about 10-fold (Z. Songyang, personal communication).

The positive results with Rad51 and Rad54 serve as proof of principle and encourage further experimentation along these lines, especially with enzymes in the more poorly defined, downstream portion of a pathway. The results with Rad52 raise two cautionary notes. First, it was shown that overexpression of Rad52 stimulated both intrachromosomal and extrachromosomal HR two- to fivefold. These assays measured effects on the single-strand annealing pathway of HR, which is independent of Rad51 and distinct from the likely pathway for targeted recombination. The effects of manipulating cellular enzyme levels should always be measured in targeted recombination assays. Second, stable overexpression of Rad52 was found to be toxic to cells, as was Rad51 in some cell lines. Consequently, careful adjustments in enzyme levels may be essential for optimal effects.

As mentioned above, the most dramatic effects on targeted recombination have been generated by I-SceI-mediated

cleavage of the target sequence on the chromosome. The requirement for prior modification, however, means that rare-cutting endonucleases such as I-SceI cannot be used as general tools to stimulate targeted HR, nor can restriction enzymes be used because their sites are all too common. However, zinc finger proteins can be engineered to recognize a diverse set of DNA sequences. By combining a zinc finger recognition domain with a nonspecific DNA-cleavage domain, zinc finger nucleases have been created that offer a flexible strategy for delivering site-specific double-strand breaks to the genome. In current designs, the zinc finger nuclease consists of a DNA binding domain composed of three zinc fingers linked to the nonspecific DNA-cleavage domain of FokI, a type IIS restriction enzyme. Each zinc finger recognizes a specific 3-nucleotide stretch of DNA so that the three fingers in a domain recognize a 9-nucleotide DNA segment. Because efficient DNA cleavage requires dimerization through the FokI nuclease domain, the effective recognition site is 18 nucleotides in length, long enough to be unique in mammalian genomes.

Zinc finger nucleases have been shown to stimulate intramolecular HR in 50% of plasmid molecules injected into *Xenopus* oocytes, and to cause site-specific somatic mutations in 50% of males, and germ-line mutations in 5% of males, when expressed in *Drosophila* larvae. They also stimulate targeted recombination in *Drosophila* by 100-fold, and in human cells by several 1000-fold. It should be noted that observations in *Drosophila* and in human cells indicate that high-level expression of some zinc finger nucleases can be toxic. If a broad range of zinc finger nucleases with high affinity and single-site specificity can be developed, it may be possible in the future for an investigator to select from a bank of such nucleases the one that will cleave a specific gene to stimulate localized gene targeting.

In the section on triple helix forming oligonucleotides (TFOs) in this review, we discuss two methods for using TFOs to induce HR and gene correction. First, by linking a DNA damaging agent such as psoralen to a TFO, it is possible to introduce site-specific damage to a genomic target specified by TFO binding. Introduction of a TFO into cells has been shown to stimulate up to several thousandfold intrachromosomal HR at a site containing the TFO target sequence. Second, it is possible to link a correcting segment of DNA to the TFO itself, using the binding specificity in the TFO to deliver the DNA to an adjacent target. Directed toward a plasmid target in mammalian cells, this approach stimulated homologous correction 5- to 10-fold above background levels. Improvements in design and delivery of such TFO reagents, as discussed later, will likely be required to enhance the effectiveness of this method for stimulating HR in cells.

The unexplained difference in targeting frequencies between nuclear injection and mass delivery methods raises the distinct possibility that alternative ways to introduce DNA into cells might yield positive benefits. One promising approach is the delivery of DNA by viral vectors. Although retroviral vectors have been shown to target at low frequencies, adeno-associated virus (AAV), which carries a single-stranded DNA genome, has been investigated most extensively. AAV viral vectors have been shown to target chromosomal sequences at frequencies ranging from 0.01 to 1% depending on the multiplicity of infection, and on the type of alteration introduced, with insertions into the target giving generally higher frequencies than deletions. Chromosomal targets carrying an I-SceI

site, which were corrected at low frequencies by the AAV vector alone, could be targeted at about 100-fold higher frequencies in the presence of I-SceI. The relatively high frequencies of targeted recombination by AAV may reflect the nuclear delivery of a high concentration of DNA by the viral preparations, or they may relate to the single-stranded nature of AAV DNA with its hairpin-protected ends. This appears to be a promising technology although it has had relatively little exposure.

3 Polyamides

3.1 Netropsin and Distamycin

Netropsin and distamycin are natural product antibiotics consisting of two or three pyrrole (Py) rings respectively connected by amide linkages. In addition, the molecules have a positively charged propylamidine at the end designated as the tail and either formyl (distamycin) or guanidine (netropsin) groups at the head (Fig. 5). The compounds have received attention for many years and it was established some time ago that they bound AT-rich sequences in the minor groove. Both drugs have a curvature that fits quite well with the floor of the minor groove. Crystallographic studies showed that a single netropsin molecule binds in the minor groove displacing the spine of hydration. There are nonspecific ionic interactions between the positively charged head and tail groups and the negatively charged phosphates of the duplex. Hydrogen bonds between the proton donors of netropsin (the amide NH) and the DNA bases position the drug, while the actual sequence recognition results from close van der Waals contacts between atoms on the drug and adenine residues. The poor binding to GC pairs is due to steric hindrance from the 2–amino group guanine, which interferes with occupancy of the groove. The structural studies detailed the interactions that determine the affinity and specificity

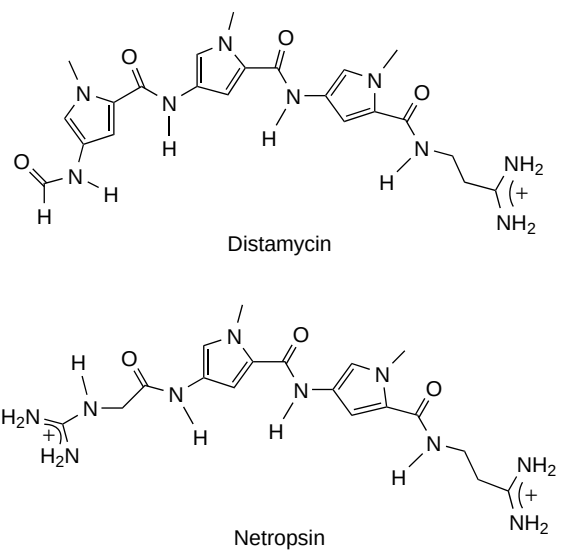

Fig. 5 The structures of distamycin and netropsin, two minor groove binding antibiotics. Each contains pyrrole residues joined by amide linkages. Note the positively charged tails and the sickle shape.

of the natural products and numerous synthetic derivatives. It was appreciated that these compounds might serve as leads for the development of sequence-specific DNA binders that could inhibit transcription of specific genes. If the suppressed genes were linked to diseases, such compounds could have pharmaceutical application. This has been a motivation for much of the work in the field.

3.2 The Development of Sequence Specific Minor Groove Ligands

A thorough discussion of the chemistry and structural properties of the numerous polyamide derivatives (sometimes called lexitropsins) is beyond the scope of this review. These features of the field have been reviewed in depth in several recent publications. We will summarize the important developments and then consider the current understanding of their biological activity and prospects for pharmacological utility.

In order to expand the binding options to GC pairs, Kopka et al. had suggested, and Lown had independently synthesized netropsin derivatives in which pyrrole was substituted with imidazole (Im). Replacement of the ring CH with N provided space for the exocyclic amine of guanine, and the N served as an acceptor of the H bond from the guanine donor. Although the new compounds recognized GC sites, they also continued to bind AT pairs, albeit with reduced affinity. These somewhat disappointing observations were rescued by one of the most important discoveries in the field. NMR studies of distamycin bound to AT sequences revealed that when the binding site contained five AT pairs, distamycin bound as an antiparallel, side-by-side, dimer, with the positively charged tails at either end (2 : 1 mode). Each drug molecule made contact with one of the duplex strands, with the N terminus at the 5′ end of the strand. In contrast, netropsin bound only in a 1 : 1 mode, because of the presence of positive charges at both ends of the molecule. The Dervan and Lown groups synthesized distamycin analogs with the pyrroles replaced by imidazole. The new compounds bound target sequences containing GC pairs, as predicted. However, further analysis by both groups determined that the optimal side-by-side arrangement for G:C binding was an imidazole:pyrrole pairing, with the imidazole interacting with the G and the pyrrole interacting with the C. Specificity was extended when the Dervan group introduced 3-hydroxypyrrole (Hp) which showed a preference for interaction with T. The Hp/Py pair recognized T:A and Py/Hp recognized A:T with discrimination greater than 10-fold depending on the orientation and target. The preference by Hp/Py for T:A is the result of shape selection by an asymmetric cleft in the minor grove and also hydrogen bonding between Hp and T. These results established pairing rules for the 2 : 1 binding motif: the pyrrole/imidazole pair (Py/Im) for C:G, Hp/Py for T:A, Py/Hp for A:T, while Py/Py is degenerate and recognizes both A:T and T:A (Fig. 6). The tails of the molecules (dimethylaminopropyl, in the compounds from the Dervan group) are placed at A:T pairs.

Additional affinity was gained when the two chains were linked (Fig. 7). Two methods of linkage have been employed. In one format, the chains are coupled via bonds between internal rings. While this did provide enhanced affinity, the more effective

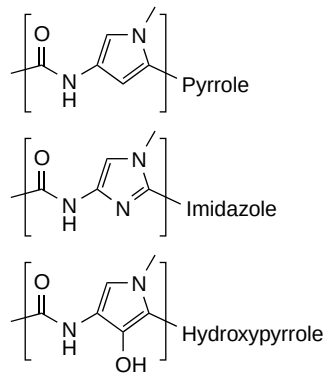

Fig. 6 The structures of the subunits of synthetic polyamide minor groove binders. When binding occurs in the 2 : 1 mode, pyrrole binds to A, T, or C, imidazole binds to G, and hydroxypyrrole binds to T. (From Goodsell, D. S. 2001, Goodsell, Sequence recognition of DNA by lexitropsins, *Curr. Med. Chem* **8**, 509–516. Reprinted with permission from Bentham Science Publishers Ltd.)

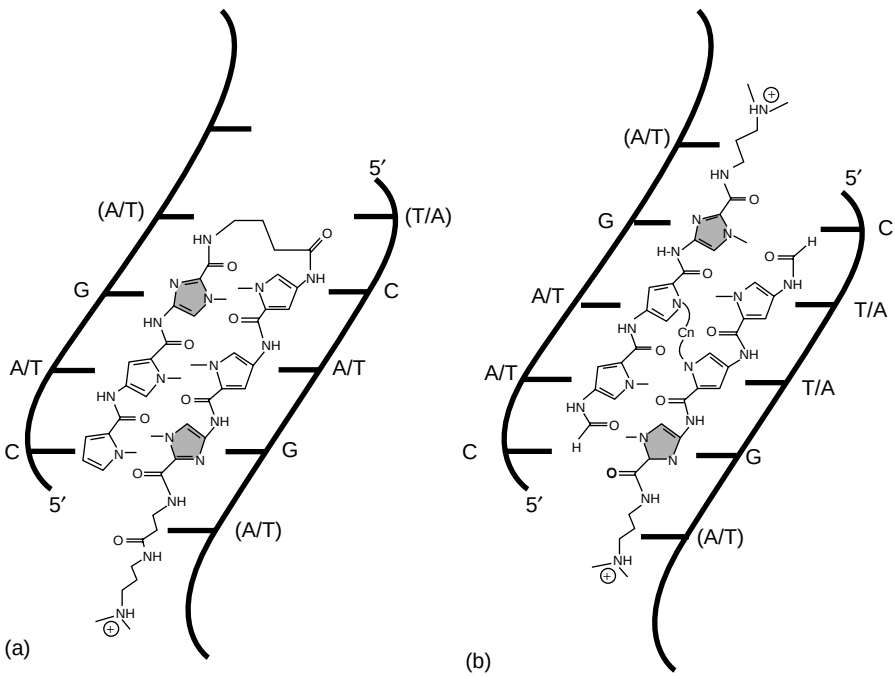

Fig. 7 (a) Hairpin and (b) ''stapled'' polyamides. Each contains 2 imidazole (stippled) residues and 4 pyrrole rings. The shifted phasing of the stapled version relative to the arrangement in the hairpin changes the target sequence. (A/T) or (T/A) base pairs at the ends of each sequence are recognized by the atoms in the linkers and tails. (Reprinted from Walker et al., 1998, PNAs **95**, 4315–4320 with permission from the National Academy of Sciences, USA.)

approach involved linkage via a head-to-tail format with the three-carbon linker, γ-aminobutyric acid. This has been employed in most subsequent syntheses. The hairpin locks the position of the subunits relative to one another, thus preventing slipped pairing schemes. This linker also binds at AT pairs, adding another

component to the recognition scheme. Although hairpin polyamides have been studied more extensively, the cross-linked polyamides continue to receive attention.

Another important development was the recognition that there was a limitation on the length of the polyamide chains. Sequence specificity and affinity were not improved beyond five rings. This was attributed to an overcurvature of the longer polyamide chain (as compared to the 3- and 4-ring compounds that match the minor groove curvature), resulting in a loss of register with the binding sites. A solution to this problem was provided by the placement of β-alanine residues after three or four polyamides. These insertions appear to act as "springs" to reset the phasing. When paired with each other in hairpin polyamides, or with Py, they interact with T:A and A:T pairs. When paired with Im, they can interact with C in G:C or C:G pairs, and may offer a solution to the otherwise "difficult" sequences. At this time, it is fair to say that considerable progress has been made in the development of the polyamide minor-grove binders. This reflects the persistent effort and ingenuity of several laboratories, the development of solid-phase methods of synthesis, and the allure of high affinity, sequence-specific ligands for ultimate pharmaceutical application.

3.3 Affinity and Specificity

The introduction of the hairpin polyamides advanced the affinity for a five-base target sequence by two orders of magnitude relative to the free three-ring compound. This brought the association constants into the nanomolar range, competitive with transcription factors. Subnanomolar affinities were reported for hairpin polyamides with four side-by-side rings binding to a six base target. These results clearly demonstrated high affinity binding by the polyamides, and suggested that they might be effective as inhibitors of transcription factor binding, and thus transcription.

The issue of specificity is, however, another matter. This question has been addressed in a number of publications from the Dervan laboratory. In one study, the binding of two eight-ring hairpin polyamides, designed to recognize sequences that contained one or two G:C pairs, was analyzed by DNAse footprinting. The constructs contained Py/Py pairs which do not distinguish A:T from T:A. Binding by one of the constructs to the specific site was subnanomolar, with two log poorer binding to two other sites with replacement of G:C by A:T or T:A. However, the other construct had a tenfold lower affinity for the specific target and only an eightfold discrimination between the specific target and the single mismatch sequence. In addition, weaker, but measurable binding (1 nM range) was seen with a double-mismatch site. In this example, both constructs bound the mismatched target with similar affinity. They differed substantially, however, in their affinity to the specific target.

Although replacement of Py/Py pairs with Hp/Py, does afford some discrimination between A:T and T:A pairs (in the range of 10 fold), it does so at the expense of affinity. Thus, the more effective binders are the compounds with no discrimination between A:T and T:A, and perhaps because of this, the Hp subunit has received little use in the recent work from the Dervan laboratory.

Binding to a single site in the human genome would require specific recognition of about 16 contiguous base pairs. In

the initial efforts to increase the target specificity of the polyamides the Dervan group constructed six- or eight-ring polyamides with internal β-alanine linkers. These bound via a "slipped pairing" mode in which there was a partial overlap of the individual polyamide chains in minor groove, with extension of the remaining unpaired chains along the target sequence. These did bind sequences as long as 16 bases. However, binding to mismatch sites was also observed with affinities relatively close (5–20-fold) to the perfect match targets. The Laemmli group has also prepared linked hairpin polyamides designed to bind insect telomeric repeat sequences. While the dimer bound with 10- to 20-fold higher affinity than the unlinked monomer, at higher concentrations it also bound almost all the sequences in the probe (termed "coating" by the Laemmli group). One unavoidable difficulty with all these studies is that analyses of specificity are limited by the sequences chosen for the binding experiments. Obviously, the array of potential binding sites in the genome cannot be represented with this strategy. However, even within this limitation it is apparent from visual inspection of the DNAse footprints from virtually all of the polyamide literature that binding occurs to many sites other than those of specific interest (Fig. 8).

A recent publication highlights some of these points. Colon cancer cells in culture were incubated with a polyamide designed to inhibit binding of the LEF-1 transcription factor to an eight-base

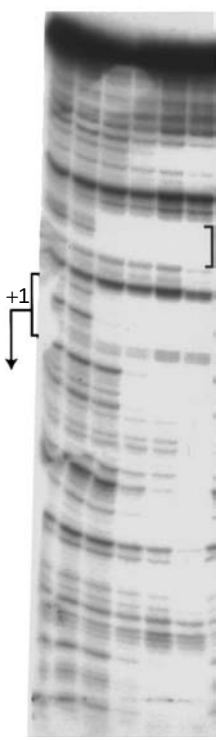

Fig. 8 DNAse footprint analysis of polyamide binding to the upstream promoter region of the *Xenopus TyrD* tRNA gene. A ^{32}P labeled restriction fragment was incubated with the polyamide at the indicated concentrations and then digested with DNAse. Protected regions are not digested and appear as blanks in the pattern. The brackets indicate the regions the polyamide was designed to protect. The arrow denotes the transcription start site. (From McBryant et al. (1999) *J. Mol. Biol.* **286**, 973–981. Reprinted with permission by Elsevier Science.)

sequence in the promoter of the cyclin D1 gene (CD1). The authors were careful to demonstrate nuclear localization of the polyamide. This was an important issue as another study suggested limitations in nuclear uptake of these compounds. The effect on message expression was examined by hybridization array analysis. Although the levels of a number of messages were affected, it was noteworthy that the levels of the targeted CD1 gene were unchanged. A computer search of the promoters of the 11 000 genes represented on the hybridization array yielded 37 380 candidate binding sites for the anti-CD1 polyamide.

3.4 Biological Applications

Early in the development of polyamides they were seen as having the potential to become rationally designed transcriptional inhibitors. Important support for this concept was provided by the demonstration that a hairpin polyamide could suppress transcription of a *Xenopus* 5S RNA gene. The compound was designed to bind a specific sequence in the binding site of the 5S RNA gene transcription factor, TFIIIA. The polyamide reduced 5S transcription in both cell extracts and live cells. In an experiment with therapeutic implications, combinations of appropriately designed polyamides were shown to inhibit HIV transcription in a cell-free assay, and to block HIV replication in peripheral blood lymphocytes. Since then there have been several studies describing biological applications of polyamides. They can inhibit transcription factor binding, and transcription *in vitro*. Most, but not all, sites on nucleosomal DNA are bound by polyamides, which block the progress of RNA polymerase through nucleosomes, apparently by inhibiting repositioning of nucleosomes during transcription. In an interesting new direction, polyamides were linked to transcriptional activation domain proteins. The polyamides provided sequence-specific DNA binding and the conjugates functioned as artificial transcription factors in transcription assays *in vitro*.

Perhaps the most spectacular publications are from Laemmli and colleagues, who synthesized polyamides targeted at the AT-rich, satellite scaffolding–associated regions of Drosophila. The satellites are found in condensed heterochromatin, and genes near these regions are silenced because of the chromatin structure. The satellite sequences are bound by the MATH (for multi-AT hooks) protein that interacts with AT tracts through minor groove contacts. The experimental rationale was based on the expectation that the polyamides would interfere with MATH dependent heterochromatin condensation. Treatment of nuclei resulted in extensive chromatin unfolding and selective staining of polytene chromosomes in the expected satellite region. Even more remarkable was the result of feeding the compounds to flies. The flies normally showed a strong variegation in eye (mottled red/white) color due to a juxtaposition of the *white* gene adjacent to the heterochromatin satellite region (resulting in variable silencing of the gene for the red color). Depending on the polyamide and the intended target, the flies showed an increase in the red color phenotype. This was interpreted as due to an opening of the chromatin resulting in relief of the suppression of the red color gene. Additional experiments with a different polyamide targeted to a different satellite at another locus also had pronounced effects on gene expression.

The results of both series of experiments suggest that extensive chromatin remodeling could result from polyamide treatment. While the effects on transcription were indirect, these results argued that polyamides could reach nuclei of living cells in whole organisms, with remarkable consequences for chromosome structure and gene expression.

Although considerable effort and intelligence has been brought to the development of polyamides, there are serious questions on the likelihood of their becoming successful drugs. Concerns about target specificity and effects on chromatin structure will have to be addressed before pharmacological applications can be considered.

4 Triple Helix–Forming Oligonucleotides

4.1 Triple Helical DNA

The DNA triple helix was discovered more than forty years ago and has been studied persistently ever since. The most stable triplexes form on polypurine:polypyrimidine tracts, which are abundant in the human genome, particularly in promoters and introns. The third strand of nucleic acid, which may consist of purines or pyrimidines (depending on the target sequence), lies in the major groove of an intact duplex. The complex is stabilized by two "Hoogsteen" hydrogen bonds between third strand bases and the bases in the purine strand of the duplex. These are formed without interfering with the hydrogen bonds between the bases of the duplex strands. Pyrimidine motif (Y.R:Y) third strands bind in a direction parallel to the purine strand in the duplex, with canonical base triplets of T.A:T and C.G:C (by convention the first base of the indicated base pair is the purine). Purine motif (R.R:Y) third strands bind antiparallel to the purine strand, with base triplets of A.A:T and G.G:C. Typically, the purine motif is used for G:C rich targets, and the pyrimidine motif for A:T rich targets (Fig. 9).

For most of the three decades following the initial discovery, triplex research was largely the province of DNA structural and physical chemists until methods were developed for automated, solid phase–based, oligonucleotide synthesis. The demonstration that synthetic oligonucleotides could form stable triplexes led to the explicit suggestion that TFOs could be developed as sequence-specific targeting reagents in living cells. Although largely conducted in the absence of biological assays, the many years of biochemical and biophysical research did define a number of obstacles that had to be, and still must be, overcome in order for TFOs to become a legitimate option for gene targeting. Consideration of biological applications raises four major issues: the activity, under physiologically relevant conditions, of TFOs directed against the classical homopurine:homopyrimidine targets in purified DNA; the activity against more general sequences; the activity against chromosomal targets, in the context of mammalian chromatin structure; and, the molecular consequences of successful triplex formation, that is, what happens after binding. Solutions to the first two have been, and will continue to be addressed by developments in oligonucleotide chemistry, while the third and fourth are the subject of active research from more biochemical and biological perspectives.

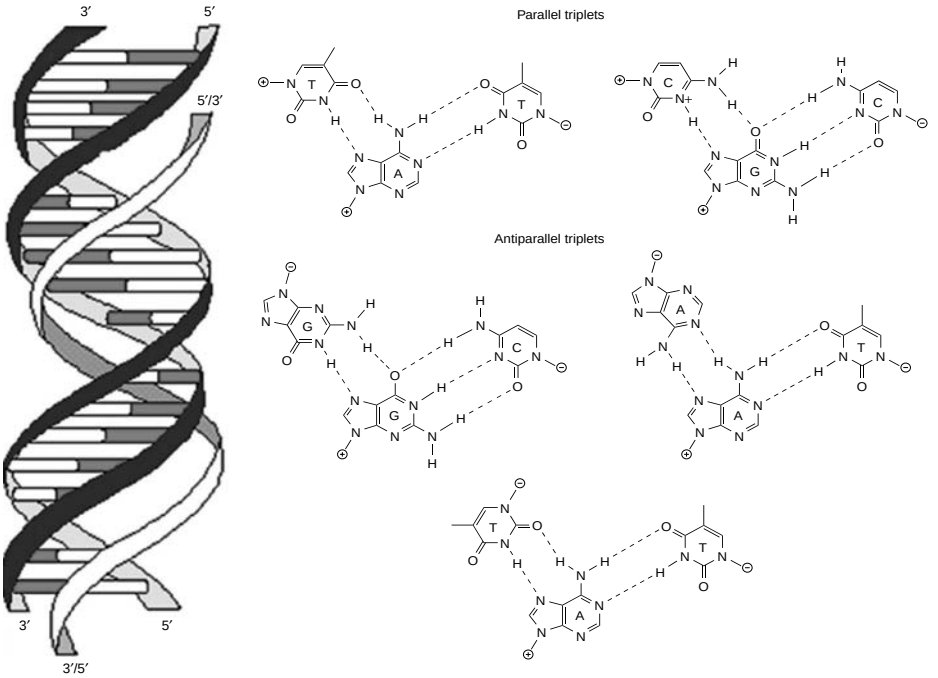

Fig. 9 The structure of the DNA triple helix. A third strand of nucleic acid lies in the major groove of an intact duplex. The target sequence consists of purines on one strand and pyrimidines on the other. Specificity and stability are provided by Hoogsteen hydrogen bonding between the bases in the third strand and the purine bases in the duplex. These hydrogen bonds do not perturb the Watson–Crick bonds of the duplex strands. Third strand orientation relative to the purine duplex strand is a function of the third strand composition – those in the pyrimidine motif bind with the same polarity as the duplex purine strand ("parallel"), while purine motif third strands bind in the opposite polarity (antiparallel). In pyrimidine, third strands T bind to A:T pairs, and C to G:C pairs. In purine, third strands G bind to G:C, A to A:T, and T can bind to A:T pairs. Note that the TA:T pairs in the two different motifs are not identical.

4.2 Triplex Formation under Physiological Conditions

Triplex formation is an inherent property of DNA without the requirement for enzymes or proteins. Nonetheless, the probability of complex formation is affected by fundamental biophysical considerations, which are further complicated by the constraints of a physiological environment. A negatively charged third strand must approach and bind a doubly negatively charged duplex, via a zipper type mechanism following an initial nucleation event involving three to five bases. In the test tube charge repulsion can be reduced by use of levels of Mg^{++} (5–10 mM) that are much higher than likely to be available as the free ion in cells. Pyrimidine motif triplexes do not form at physiological pH because of the requirement for cytosine protonation that occurs at relatively acidic pH ($pK_a = 4.5$). This is necessary for one of the hydrogen bonds (Fig. 9), and the resultant positive charge also

makes an important contribution to triplex stability. Purine motif third strands (which are G rich) can form G-tetrad structures in physiological levels of K^+, inhibiting triplex formation. Finally, conformational changes are required by both the third strand and the duplex target. All these factors contribute to the much slower kinetics of triplex formation relative to duplex formation, and reduce triplex stability (most triplexes, even under optimal conditions *in vitro*, are less stable than the underlying duplex).

4.3 Oligonucleotide Modifications Improve TFO Activity

Certain base and sugar modifications improve TFO activity under physiologically relevant conditions. For example, 5-methylcytosine partially alleviates the pH restriction of TFOs in the pyrimidine motif, due to the contribution of the methyl group to base stacking and/or the exclusion of water molecules from the major groove.

RNA third strands form more stable pyrimidine motif triplexes than the corresponding DNA strands, and that observation led to the use of RNA analog sugar residues such as 2′-*O*-Methyl (2′-OMe) (Fig. 10). Other analogs, such as the 2′, 4′ bridged ribose (Locked Nucleic Acid, LNA, or Bridged Nucleic Acid, BNA) also improve triplex stability. These modifications stabilize the C-3′ endo configuration of the sugar. This has the salutary effect of preorganizing the third strand in a conformation that is compatible with triplex formation, while also imposing minimal distortion on the underlying duplex. Many other modifications have been considered. These include intercalators linked to TFOs and the base analog propynyl-deoxyuridine, both of which stabilize triplexes. The Mg^{++} dependence of pyrimidine motif TFOs has also been reduced by modifications to the backbone, such as replacement of the phosphate linkage, or a bridging oxygen with a nitrogen. Substitution of a nonbridging oxygen in the backbone with a charged amine reduces G-tetrad formation by purine TFOs in physiological K^+, making more oligonucleotide available for triplex formation and reducing the charge repulsion.

A positive charge and an RNA-like sugar conformation have been combined in the 2′-*O*-(2-aminoethyl) (AE) ribose derivatives of Cuenoud and colleagues. TFOs carrying these substitutions can form triplexes at rates over 1000-fold faster than corresponding deoxy third strands. The stability of the resultant complex at physiological pH and low Mg^{++} concentration is also enhanced. NMR analysis indicates a specific interaction between the positively charged amines (at physiological pH) and phosphate groups in the purine strand of the duplex.

Although the biochemical data indicate that these modifications improve TFO binding, few have been tested in biological assays. In the following section, we will

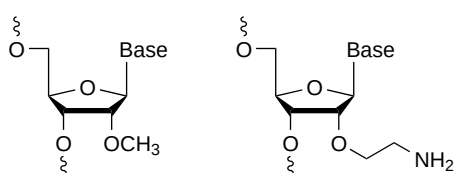

Fig. 10 The structure of modified sugars incorporated in triple helix forming oligonucleotides, 2′-O-methyl (2′-OMe) and 2′-O-aminoethyl (2′-AE).

discuss the biological activity of TFOs carrying some of these substitutions.

4.4 Activity of TFOs *In Vivo*

Any effort to develop TFO for gene targeting in living cells must be cognizant of the potential restriction to triplex formation by eukaryotic chromatin structure. It has been shown in biochemical experiments that target sequences in nucleosomes are poor substrates for triplex formation. However, chromatin structure is dynamic *in vivo*, and there are recent indications that manipulation of the biology of the cell can influence targeting efficiency (see below).

In the early biological studies, TFOs were envisioned as tools to inhibit gene expression by blocking transcription initiation or elongation (the antigene strategy). Triplex formation within promoter sites can block transcription factor access, and inhibit gene activation *in vitro*. Promoters of active genes are often nucleosome free or have nucleosomes whose interaction with the DNA has been relaxed by remodeling or histone modification. Results from a number of investigators suggest that TFOs designed to bind specific promoters can decrease expression of the targeted gene in mammalian cells. The results of these and similar studies suggest that TFOs can find chromosomal targets and produce a measurable biological endpoint. However, the evidence for target interaction is necessarily inferential and there is no direct demonstration of TFO binding at the target site. Furthermore, as is well documented in the antisense literature, oligonucleotides can have multiple effects on cells and transcript levels, via mechanisms unrelated to their original design.

In an alternative approach, TFOs have been used for genome modification, resulting in a change in target sequence. This has the advantage of introducing permanent changes in the target sequence, which simplifies interpretation of the experiments. Two strategies have been employed: in the first the TFO is linked to a DNA reactive compound, while in the second the binding of the TFO is sufficient to elicit a response.

4.5 TFOs Linked to DNA Reactive Compounds

TFOs linked to DNA reactive compounds have been used to introduce DNA adducts at specific target sites. TFOs conjugated to cleavage and alkylating reagents have been described, although these reagents have been largely restricted to biochemical experiments. The most useful, and most frequently used reagent is psoralen, a DNA cross-linker requiring photoactivation by long wave ultraviolet light. In the initial biological experiments, purine motif TFOs conjugated to psoralen were used to target cross-links to a specific site on mutation reporter plasmids or phage *in vitro*. Passage of the modified vectors through host cells resulted in mutagenesis of the target site. Psoralen-conjugated TFOs that transfected into cells were shown to induce base pair-specific mutations within a *supF* mutation reporter gene in a shuttle vector plasmid either cotransfected or previously transfected into cells. These results demonstrated intracellular targeting by a TFO, and also demonstrated that the mutagenesis assay could be used to relate the affinity of the TFO to the biological activity.

The shuttle vector experiments were followed by the demonstration of targeted mutagenesis of a chromosomal site by

a psoralen-linked pyrimidine TFO containing 2′-OMe sugars and a pyrene intercalator. The target sequence was in the endogenous hamster *HPRT* (hypoxanthine phosphoribosyltransferase) gene. Cells with inactivating mutations in *HPRT* can be quantitatively identified in a simple and unambiguous selection procedure. TFOs with uniform 2′-OMe substitutions were inactive, but those with the pyrene intercalator showed *HPRT* knockout activity at frequencies in the range of 10^{-4} to 10^{-3}. Sequence analysis of the mutant clones confirmed the localization of mutations to the target region. More recently, psoralen-linked TFOs containing several 2′-AE-residues as well as 2′-OMe substitutions have been prepared. These mixed substitution TFOs form triplexes at lower levels of Mg^{++} than required for the TFOs with only the 2′-OMe residues. Thermal melting experiments showed that the AE triplexes were more stable than the 2′-OMe only triplexes, and notably, were more stable than the underlying duplex at physiological pH. The AE TFOs were much more active in the *HPRT* knockout assay than the 2′-OMe TFOs. As discussed above, the enhanced biochemical and biological properties are due to the positive charge of the amine group. Recently it has been shown that psoralen-TFOs, carrying optimal levels of AE substitutions, can introduce mutations at the HPRT target site at frequencies in the range of 5%. This is sufficiently high to allow identification of cell clones with targeted mutations in direct screens without a requirement for selection.

Access to chromatin targets has been considered in studies with nuclei and in permeabilized cells and has been shown for psoralen-linked TFOs, as well as for a TFO alkylator conjugate. In these experiments, the frequency of cross-linking served as a measure of TFO binding at the time of photoactivation. Targeting activity was greatest in cells in which the target region was undergoing active transcription prior to permeabilization. In recent work, we have found that the frequency of targeted cross-links in living cells (and the resultant mutagenesis) is greatest in S-phase cells.

These results argue that transcription and replication can influence target access. However, the relationship between specific chromatin structures and target access remains to be determined.

4.6 Triplex Formation, by itself, can Promote Mutagenesis

Triplex structures are recognized by DNA repair proteins such as XPA/RPA. The possibility that this recognition could lead to mutagenesis was raised in experiments in cultured mammalian cells cotransfected with a plasmid containing a *supF* reporter gene with an embedded triplex target site and a purine motif TFO, designed to bind the target. Plasmids with mutations in the *supF* gene were recovered. These included deletions and point mutations distributed throughout the gene, and resembled those recovered from experiments with nicked plasmids. Mutagenesis was independent of psoralen, but did require functional DNA repair and transcription coupled repair functions. It seems likely that some triplex structures are recognized as lesions and trigger DNA repair with an error-prone outcome, as well as other events associated with DNA metabolism (see below).

Chromosomal targets in both cultured cells and transgenic mice have been mutagenized by treatment of cells or mice

with purine motif TFOs. Unconjugated purine motif TFOs were employed, and the mutations included frameshifts (in the homopolymer runs of the target sequence) and distributed point mutations. In the mouse experiment, all tissues tested showed TFO-induced mutagenesis except the brain, which had no mutagenesis over background, consistent with TFOs being unable to cross the blood–brain barrier. The mutations were consistent with error-prone filling of gaps or slipped pairing during gap filling. Thus TFO targeting is possible in whole animals.

4.7 Triplex-induced Recombination and Gene Correction

The suggestion that triplexes would be recognized as a substrate for DNA repair prompted the notion that they might also trigger recombination. Experiments with a TFO and shuttle vector plasmid carrying two (different) mutant copies of the *supF* gene with the TFO target site between them demonstrated that triplex formation, with or without psoralen cross-linking, could induce recombination to produce a functional *supF* gene to psoralen. Gene correction was dependent on a functional XPA protein, a DNA damage–recognition protein and a component of the nucleotide excision repair machinery.

This approach has been extended to a chromosomal target in mouse fibroblasts. This was prepared by integration into a single chromosomal location of a construct containing a triplex target sequence with different mutant *thymidine kinase* (*TK*) genes placed on either side. When the appropriate TFOs were microinjected into the nuclei of the cells, the yield of recombinants was 1 to 2%, more than 1000-fold over background. Analysis of the recombinant clones revealed that all the recombination events involved gene conversion. Similar studies using TFOs to stimulate recombination have also been described using a locus in CHO cells containing duplicated *APRT* genes as a target.

The observation that third strand binding could provoke DNA repair and stimulate recombination led to the development of a strategy to mediate targeted gene conversion. In this protocol, a TFO was linked to a short DNA fragment homologous to the target site (except for the base pair to be corrected). The conception of the approach is that the TFO domain mediates site-specific binding to target the molecule to the desired gene. This binding would also trigger repair to initiate recombination or gene conversion with the "tethered" donor sequence. This idea was tested with an oligomer containing a 40-mer donor element and a 30-mer TFO domain. Correction of a single base pair mutation in the *supF* reporter gene in a shuttle vector plasmid in mammalian cells was shown. *In vitro* studies in human cell-free extracts revealed a requirement for XPA and Rad51 in the recombination pathway.

4.8 Expansion of the Triplex Binding Code

Although research on the biological applications of TFOs must be considered to be in the early stages, it does appear that three of the important issues raised above have been addressed: it is possible to synthesize TFOs that are functional *in vivo*; chromosomal targets are accessible, and accessibility can be modulated by manipulating the cell biology; TFOs can be used to target mutagenesis and recombination. Thus, although there is a great deal to be done to extend these conclusions, expansion of

target options to mixed purine/pyrimidine sequences would appear to be the most important challenge facing the field. It is likely that solutions will come from synthetic chemistry via novel base analogs. Although there have been some interesting directions identified in the effort to expand the triplex binding code, it is fair to say that at this time there are no analogs that permit binding to general sequences. This remains an area of opportunity.

5 Mutation Repair by Small DNA Fragments

A seemingly attractive approach to introducing defined sequence changes into chromosomal loci would be based on conventional DNA fragments, homologous to the target region except for the intended change. The strategy is based on the notion that heteroduplex formation between the target and the donor sequence would be resolved, at least some of the time, in the experimentalist's favor. The method would take advantage of the four base recognition code, and relative ease and low cost of preparation of the correcting reagents. The formation of the critical heteroduplex would result from the action of the recombination apparatus in the host cells, or, (perhaps in the case of single-strand fragments) as a result of transient exposure of single-strand regions during replication, transcription, or repair. That such an approach might work is supported by the early demonstration of "marker rescue" by Edgall and Hutchison, in *E. coli*. Measurable correcting activity of single-strand oligonucleotides against a chromosomal target in yeast, an extrachromosomal target in mammalian cells, and, recently, the chromosomal *HPRT* gene in mammalian cells have also been reported. The frequency of correction in all these reports was too low to have practical utility. Short double-strand fragments were also shown to have low correcting activity in mammalian cells. In a recent work, oligonucleotides have been used to manipulate chromosomal gene sequences in *E. coli* at frequencies high enough to have practical utility. The much higher frequencies in the bacterial host may reflect fundamental differences in the exposure of single-strand regions in *E. coli* versus mammalian cells.

5.1 Small Fragment Homologous Replacement

Gruenert and colleagues have developed an approach based on double-strand fragments approximately 500-bp long (termed *small fragment homologous replacement, SFHR*). They have been largely concerned with the correction of mutations in the cystic fibrosis transmembrane conductance regulator (CFTR). They have introduced fragments designed to correct the common ΔF508 mutation (a deletion of three bases) and presented PCR-based evidence for conversion at the level of the chromosome and mRNA. They have argued that the conversion frequency was in the range of 1 to 10% of treated cells. Recently, this approach has been extended to the correction of a mutation in the dystrophin gene in mouse cells. Although evidence for correction was presented, expression of wild-type distrophin was not observed. While the SFHR method may prove useful, it has received attention from few laboratories and no mechanistic studies have been performed. One of the limitations in developing the methodology has been the focus on disease genes as targets. Although of obvious interest, they

do not lend themselves to straightforward quantitative measurement of correction events.

5.2 Chimeric Oligonucleotides

A review of oligonucleotide and DNA fragment–based gene correction technologies would not be complete without mention of chimeric oligonucleotides and the claims for their activity by Kmiec and coworkers. In the original description, these were foldback molecules containing a DNA double-strand region that contained a "correction" base pair flanked by a stretch of DNA: 2′-OMe hybrid (RNA–DNA oligonucleotide, RDO). Treatment of cells carrying a β-globin sickle-cell mutation with a "wild-type" chimeric oligonucleotide was reported to result in very high levels of correction (30–50%) after 6 h of treatment. However, there has been widespread failure to reproduce this and other claims from this group, and it is not clear that the original report is credible.

6 Peptide Nucleic Acids

Peptide nucleic acids (PNAs) are DNA analogs consisting of nucleobases attached to a peptide backbone of *N*-(2-aminoethyl)glycine residues (Fig. 11). The spacing of the bases along the backbone maintains the register relative to canonical DNA-permitting base pairing between PNA strands and DNA or RNA strands. The replacement of the phosphate backbone by a neutral linkage greatly reduces the charge repulsion between PNA and DNA or RNA strands. Thus, very stable PNA:DNA or PNA:RNA hybrids are formed. They can also form four stranded complexes with polypurine:polypyrimidine DNA duplex targets. Binding to duplex DNA involves triplex formation with one strand of PNA acting as a third strand paired via Hoogsteen hydrogen bonds to the purine strand of the duplex (as in a conventional triplex), while another PNA

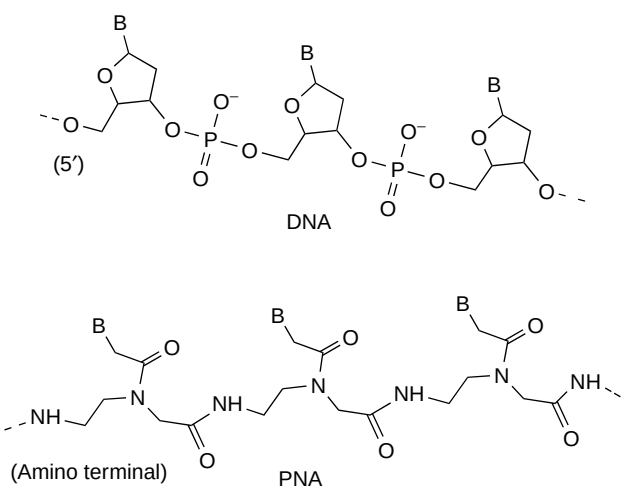

Fig. 11 The structure of a peptide nucleic acid backbone (PNA) as compared to DNA. (Reprinted from with permission from Elsevier Science).

strand forms Watson–Crick pairs with the same purine strand (PNAPNA:DNA). The latter interaction occurs by way of what has been termed *strand invasion* in which the duplex DNA partner is eventually displaced by the PNA strand. This results in a very stable "clamp" structure with the pyrimidine strand of the duplex looped out. Linkage of the two PNA strands to form bis-PNAs improves binding by reducing entropic barriers. Although PNAs bind with high affinity under optimal conditions, it is important to note that binding is dependent on breathing of the duplex. At physiological salt concentrations, binding is almost completely suppressed, suggesting that strand capture, rather than strand invasion, may be a better description of the process of complex formation between PNAs and duplex targets. Basic peptides have been attached to bis-PNAs in order to accelerate binding to linear duplex DNA. Recently, this approach was shown to partially ameliorate the restriction on binding by salt concentrations of 100 mM KCl.

The limitation on binding activity by physiological salt concentrations would appear to mitigate against gene targeting by PNAs *in vivo*. However, reports from the Glazer laboratory suggest that at least some binding occurs at chromosomal targets *in vivo*. PNAs were designed to form clamp complexes with polypurine sites in a *supF* gene, which was integrated into the genome of murine cells. Treatment of the cells with the PNAs resulted in mutagenesis of the *supF* gene, at frequencies 10-fold above the background. The mutations were single-base insertions and deletions, as well as single-base substitutions, within a run of guanines. The mutations could be rationalized by a replication slippage model. In another experiment, PNAs were targeted to polypurine sites upstream of the transcription start site for the human Gγ-globin gene. Treatment of K562 cells, which have low levels of γ-globin message, resulted in the enhanced expression of the gene from start sites at the PNA binding site as well as from the native promoter. The interpretation of both experiments is that the PNA clamp structure can form *in vivo* on a chromosomal target and the resultant "lesion" can stimulate error-prone DNA repair synthesis and/or transcription on the opened template strand. This structure also stimulated the recombination in cell-free extracts.

7 Summary

At the outset of this review, we noted that targeting reagents could be classified according to their approach to target recognition via the major or minor groove, or single strand. This is a useful way of thinking about the problem of sequence recognition and discrimination. At the end of the review, it is appropriate to consider the problem from another perspective and ask which strategies work in terms of selective sequence recognition. The simplest answer is that the successful approaches are those with the benefit of selection, during evolution, such as HR, or in the laboratory such as the zinc finger proteins reviewed elsewhere in this series. In contrast, those approaches based on chemistry supported by molecular modeling have been less effective. Thus, the considerable efforts to construct base analogs that expand the triplex binding code, or polyamides with high sequence selectivity, have had limited success. To a certain extent, this is the result of gaps in our understanding of the molecular architecture

of target sequences under physiological conditions. Although molecular models are often compelling and promise success, the results often expose the limitations of the models rather than confirm the efficacy of the constructions. It may be that substantial improvements of reagents, based on chemistry, will require the application of combinatorial, or quasi-combinatorial, strategies in which effective solutions are derived by evolution in the laboratory, rather than by reliance on models.

See also Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Bailly, C., Chaires, J.B. (1998) Sequence specific DNA minor groove binders. Design and synthesis of netropsin and distamycin analogues, *Bioconjugate Chemistry* **9**, 513–538.

Braasch, D.A., Corey, D.R. (2002) Novel antisense and peptide nucleic acid strategies for controlling gene expression, *Biochemistry* **41**, 4503–4510.

Choo, Y., Klug, A. (1995) Designing DNA-binding proteins on the surface of filamentous phage, *Curr. Opin. Biotechnol.* **6**, 431–436.

Cox, M.M. (2001) Recombinational DNA repair of damaged replication forks in Escherichia coli: questions, *Annu. Rev. Genet.* **35**, 53–82.

Dean, D.A. (2000) Peptide nucleic acids: versatile tools for gene therapy strategies, *Adv. Drug Deliv. Rev.* **44**, 81–95.

Frank-Kamenetskii, M.D., Mirkin, S.M. (1995) Triplex DNA structures, *Annu. Rev. Biochem.* **64**, 65–95.

Fry, C.J., Peterson, C.L. (2002) Transcription. Unlocking the gates to gene expression, *Science* **295**, 1847–1848.

Gilbert, D.E., Feigon, J. (1999) Multistranded DNA structures, *Curr. Opin. Struct. Biol.* **9**, 305–314.

Goodsell, D.S. (2001) Sequence recognition of DNA by lexitropsins, *Curr. Med. Chem.* **8**, 509–516.

Gowers, D.M., Fox, K.R. (1999) Towards mixed sequence recognition by triple helix formation, *Nucleic Acids Res.* **27**, 1569–1577.

Haber, J.E. (1999) DNA recombination: the replication connection, *Trends Biochem. Sci.* **24**, 271–275.

Luyten, I.A., Herdewijn, P. (2002) Hybridization properties of base-modified oligonucleotides within the double and triple helix motif, *Eur. J. Med. Chem.* **33**, 515–576.

Muller, U. (1999) Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis, *Mech. Dev.* **82**, 3–21.

Lusetti, S.L., Cox, M.M. (2002) The bacterial RecA protein and the recombinational DNA repair of stalled replication forks, *Annu. Rev. Biochem.* **71**, 71–100.

Nielsen, P.E. (2001) Targeting double stranded DNA with peptide nucleic acid (PNA), *Curr. Med. Chem.* **8**, 545–550.

Roth, D.B., Wilson, J.H. (1988) in: Kucherlapati, R., Smith, G. (Eds.) *Genetic Recombination*, American Society of Microbiology, Washington, DC, pp. 621–651.

Thuong, N.T., Helene, C. (1993) Sequence specific recognition and modification of double helical DNA by oligonucleotides, *Angewandte Chemie, Intl. Ed.* **32**, 666–690.

van der Weyden, L., Adams, D.J., Bradley, A. (2002) Tools for targeted manipulation of the mouse genome, *Physiol. Genomics* **11**, 133–164.

Vasquez, K.M., Marburger, K., Intody, Z., Wilson, J.H. (2001b) Manipulating the mammalian genome by homologous recombination, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8403–8410.

Wemmer, D.E. (1999) Ligands recognizing the minor groove of DNA: development and applications, *Biopolymers* **52**, 197–211.

Wemmer, D.E. (2000) Designed sequence-specific minor groove ligands, *Annu. Rev. Biophys. Biomol. Struct.* **29**, 439–461.

Wemmer, D.E., Dervan, P.B. (1997) Targeting the minor groove of DNA, *Curr. Opin. Struct. Biol.* **7**, 355–361.

Wolfe, S.A., Nekludova, L., Pabo, C.O. (2000) DNA recognition by Cys2His2 zinc finger proteins, *Annu. Rev. Biophys. Biomol. Struct.* **29**, 183–212.

Primary Literature

Adair, G.M., Nairn, R.S., Wilson, J.H., Seidman, M.M., Brotherman, K.A., MacKinnon, C., Scheerer, J.B. (1989) Targeted homologous recombination at the endogenous adenine phosphoribosyltransferase locus in Chinese hamster cells, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4574–4578.

Alberti, P., Arimondo, P.B., Mergny, J.L., Garestier, T., Helene, C., Sun, J.S. (2002) A directional nucleation-zipping mechanism for triple helix formation, *Nucleic Acids Res.* **30**, 5407–5415.

Aratani, Y., Okazaki, R., Koyama, H. (1992) End extension repair of introduced targeting vectors mediated by homologous recombination in mammalian cells, *Nucleic Acids Res.* **20**, 4795–4801.

Arimondo, P.B., Garestier, T., Helene, C., Sun, J.S. (2001) Detection of competing DNA structures by thermal gradient gel electrophoresis: from self-association to triple helix formation by (G,A)-containing oligonucleotides, *Nucleic Acids Res.* **29**, E15.

Arimondo, P.B., Moreau, P., Boutorine, A., Bailly, C., Prudhomme, M., Sun, J.S., Garestier, T., Helene, C. (2000) Recognition and cleavage of DNA by rebeccamycin- or benzopyridoquinoxaline conjugated of triple helix-forming oligonucleotides, *Bioorg. Med. Chem.* **8**, 777–784.

Arya, D.P., Bruice, T.C. (1999) Triple-helix formation of DNA oligomers with methylthiourea-linked nucleosides (DNmt): a kinetic and thermodynamic analysis, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4384–4389.

Asensio, J.L., Carr, R., Brown, T., Lane, A.N. (1999) Conformational and thermodynamic properties of parallel intramolecular triple helixes containing a DNA, RNA, or 2′-OMeDNA third strand, *J. Am. Chem. Soc.* **121**, 11063–11070.

Asensio, J.L., Lane, A.N., Dhesi, J., Bergqvist, S., Brown, T. (1998) The contribution of cytosine protonation to the stability of parallel DNA triple helices, *J. Mol. Biol.* **275**, 811–822.

Bailis, A.M., Rothstein, R. (1990) A defect in mismatch repair in Saccharomyces cerevisiae stimulates ectopic recombination between homologous genes by an excision repair dependent process, *Genetics* **126**, 535–547.

Baird, E.E., Dervan, P.B. (1996) Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids, *J. Am. Chem. Soc.* **118**, 6141–6146.

Baliga, R., Singleton, J.W., Dervan, P.B. (1995) RecA.oligonucleotide filaments bind in the minor groove of double-stranded DNA, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10393–10397.

Belitsky, J.M., Leslie, S.J., Arora, P.S., Beerman, T.A., Dervan, P.B. (2002) Cellular uptake of N-methylpyrrole/N-methylimidazole polyamide-dye conjugates, *Bioorg. Med. Chem.* **10**, 3313–3318.

Belmaaza, A., Wallenburg, J.C., Brouillette, S., Gusew, N., Chartrand, P. (1990) Genetic exchange between endogenous and exogenous LINE-1 repetitive elements in mouse cells, *Nucleic Acids Res.* **18**, 6385–6391.

Belousov, E.S., Afonina, I.A., Kutyavin, I.V., Gall, A.A., Reed, M.W., Gamper, H.B., Wydro, R.M., Meyer, R.B. (1998) Triplex targeting of a native gene in permeabilized intact cells: covalent modification of the gene for the chemokine receptor CCR5, *Nucleic Acids Res.* **26**, 1324–1328.

Bentin, T., Nielsen, P.E. (1996) Enhanced peptide nucleic acid binding to supercoiled DNA: possible implications for DNA "breathing" dynamics, *Biochemistry* **35**, 8863–8869.

Besch, R., Giovannangeli, C., Kammerbauer, C., Degitz, K. (2002) Specific inhibition of ICAM-1 expression mediated by gene targeting with Triplex-forming oligonucleotides, *J. Biol. Chem.* **277**, 32473–32479.

Bibikova, M., Breumer, K., Trautman, J.K., Carroll, D. (2003) Enhanced gene targeting by target cleavage with designed zinc-finger nucleases, *Science* **300**, 764.

Bibikova, M., Carroll, D., Segal, D.J., Trautman, J.K., Smith, J., Kim, Y.G., Chandrasegaran, S. (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases, *Mol. Cell Biol.* **21**, 289–297.

Bibikova, M., Golic, M., Golic, K.G., Carroll, D. (2002) Targeted chromosomal cleavage and mutagenesis in drosophila using zinc-finger nucleases, *Genetics* **161**, 1169–1175.

Blume, S.W., Lebowitz, J., Zacharias, W., Guarcello, V., Mayfield, C.A., Ebbinghaus, S.W., Bates, P., Jones D.E. Jr., Trent, J., Vigneswaran, N., Miller, D.M. (1999) The integral divalent cation within the intermolecular purine*purine. pyrimidine structure: a variable determinant of the potential for and

characteristics of the triple helical association, *Nucleic Acids Res.* **27**, 695–702.

Bollag, R.J., Liskay, R.M. (1988) Conservative intrachromosomal recombination between inverted repeats in mouse cells: association between reciprocal exchange and gene conversion, *Genetics* **119**, 161–169.

Brown, P.M., Fox, K.R. (1996) Nucleosome core particles inhibit DNA triple helix formation, *Biochem. J.* **319**, 607–611.

Bruscia, E., Sangiuolo, F., Sinibaldi, P., Goncz, K.K., Novelli, G., Gruenert, D.C. (2002) Isolation of CF cell lines corrected at DeltaF508-CFTR locus by SFHR- mediated targeting, *Gene Ther.* **9**, 683–685.

Campbell, C.R., Keown, W., Lowe, L., Kirschling, D., Kucherlapati, R. (1989) Homologous recombination involving small single-stranded oligonucleotides in human cells, *New Biol.* **1**, 223–227.

Capecchi, M.R. (1989) Altering the genome by homologous recombination, *Science* **244**, 1288–1292.

Carlomagno, T., Blommers, M.J., Meiler, J., Cuenoud, B., Griesinger, C. (2001) Determination of aliphatic side-chain conformation using cross-correlated relaxation: application to an extraordinarily stable 2′-aminoethoxy-modified oligonucleotide triplex, *J. Am. Chem. Soc.* **123**, 7364–7370.

Chan, P.P., Lin, M., Faruqi, A.F., Powell, J., Seidman, M.M., Glazer, P.M. (1999) Targeted correction of an episomal gene in mammalian cells by a short DNA fragment tethered to a triplex-forming oligonucleotide, *J. Biol. Chem.* **274**, 11541–11548.

Chandrasegaran, S., Smith, J. (1999) Chimeric restriction enzymes: what is next? *Biol. Chem.* **380**, 841–848.

Chen, Y.H., Yang, Y., Lown, J.W. (1996) Optimization of cross-linked lexitropsins, *J. Biomol. Struct. Dyn.* **14**, 341–355.

Chiang, S.Y., Burli, R.W., Benz, C.C., Gawron, L., Scott, G.K., Dervan, P.B., Beerman, T.A. (2000) Targeting the ets binding site of the HER2/neu promoter with pyrrole–imidazole polyamides, *J. Biol. Chem.* **275**, 24246–24254.

Cole-Strauss, A., Yoon, K., Xiang, Y., Byrne, B.C., Rice, M.C., Gryn, J., Holloman, W.K., Kmiec, E.B. (1996) Correction of the mutation responsible for sickle cell anemia by an RNA–DNA oligonucleotide, *Science* **273**, 1386–1389.

Cooney, M., Czernuszewicz, G., Postel, E.H., Flint, S.J., Hogan, M.E. (1988) Site-specific oligonucleotide binding represses transcription of the human c-myc gene in vitro, *Science* **241**, 456–459.

Cox, M.M., Goodman, M.F., Kreuzer, K.N., Sherratt, D.J., Sandler, S.J., Marians, K.J. (2000) The importance of repairing stalled replication forks, *Nature* **404**, 37–41.

Cuenoud, B., Casset, F., Husken, D., Natt, F., Wolf, R.M., Altmann, K.H., Martin, P., Moser, H.E. (1998) Dual recognition of double stranded DNA by 2′-aminoethoxy-modified oligonucleotides, *Angewandte Chemie, Intl. Ed.* **37**, 1288–1291.

Dagle, J.M. Weeks, D.L. (1996) Positively charged oligonucleotides overcome potassium-mediated inhibition of triplex DNA formation, *Nucleic Acids Res.* **24**, 2143–2149.

Datta, H.J., Chan, P.P., Vasquez, K.M., Gupta, R.C., Glazer, P.M. (2001) Triplex-induced recombination in human cell-free extracts: Dependence on XPA and HsRad51, *J. Biol. Chem.* **276**, 18018–18023.

Delacote, F., Han, M., Stamato, T.D., Jasin, M., Lopez, B.S. (2002) An xrcc4 defect or Wortmannin stimulates homologous recombination specifically induced by double-strand breaks in mammalian cells, *Nucleic Acids Res.* **30**, 3454–3463.

Deng, C., Capecchi, M.R. (1992) Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus, *Mol. Cell Biol.* **12**, 3365–3371.

Dickinson, L.A., Gulizia, R.J., Trauger, J.W., Baird, E.E., Mosier, D.E., Gottesfeld, J.M., Dervan, P.B. (1998) Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands [see comments], *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12890–12895.

Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S., Smithies, O. (1987) Targeted correction of a mutant HPRT gene in mouse embryonic stem cells, *Nature* **330**, 576–578.

Dwyer, T.J., Geierstanger, B., Bathini, Y., Lown, J.W., Wemmer, D. (1992) Design and binding of a distamycin A analog to d(CGCAAGTTGGC):d(GCCAACTTGCG): synthesis, NMR studies, and implications for the design of minor groove binding

oligopeptides, *J. Am. Chem. Soc.* **114**, 5911–5919.

Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S.M., Driver, D.A., Berg, R.H., Kim, S.K., Norden, B., Nielsen, P.E. (1993) PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules, *Nature* **365**, 566–568.

Egholm, M., Christensen, L., Dueholm, K.L., Buchardt, O., Coull, J., Nielsen, P.E. (1995) Efficient pH-independent sequence-specific DNA binding by pseudoisocytosine-containing bis-PNA, *Nucleic Acids Res.* **23**, 217–222.

Ellis, J., Bernstein, A. (1989) Gene targeting with retroviral vectors: recombination by gene conversion into regions of nonhomology, *Mol. Cell Biol.* **9**, 1621–1627.

Escude, C., Giovannangeli, C., Sun, J.S., Lloyd, D.H., Chen, J.K., Gryaznov, S.M., Garestier, T., Helene, C. (1996) Stable triple helices formed by oligonucleotide N3′ → P5′ phosphoramidates inhibit transcription elongation, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4365–4369.

Escude, C., Sun, J.S., Rougee, M., Garestier, T., Helene, C. (1992) Stable triple helices are formed upon binding of RNA oligonucleotides and their 2′-O-methyl derivatives to double-helical DNA, *C. R. Acad. Sci. III* **315**, 521–525.

Espinas, M.L., Jimenez-Garcia, E., Martinez-Balbas, A., Azorin, F. (1996) Formation of triple-stranded DNA at d(GA.TC)n sequences prevents nucleosome assembly and is hindered by nucleosomes, *J. Biol. Chem.* **271**, 31807–31812.

Essers, J., Hendriks, R.W., Wesoly, J., Beerens, C.E., Smit, B., Hoeijmakers, J.H., Wyman, C., Dronkert, M.L., Kanaar, R. (2002) Analysis of mouse Rad54 expression and its implications for homologous recombination, *DNA Repair (Amst)* **1**, 779–793.

Evans, E., Alani, E. (2000) Roles for mismatch repair factors in regulating genetic recombination, *Mol. Cell Biol.* **20**, 7839–7844.

Faruqi, A.F., Datta, H.J., Carroll, D., Seidman, M.M., Glazer, P.M. (2000) Triple-helix formation induces recombination in mammalian cells via a nucleotide excision repair-dependent pathway, *Mol. Cell Biol.* **20**, 990–1000.

Faruqi, A.F., Egholm, M., Glazer, P.M. (1998) Peptide nucleic acid-targeted mutagenesis of a chromosomal gene in mouse cells, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1398–1403.

Faruqi, A.F., Seidman, M.M., Segal, D.J., Carroll, D., Glazer, P.M. (1996) Recombination induced by triple-helix-targeted DNA damage in mammalian cells, *Mol. Cell Biol.* **16**, 6820–6828.

Felsenfeld, G., Davies, D.R., Rich, A. (1957) Formation of a three stranded polynucleotide molecule, *J. Am. Chem. Soc.* **79**, 2023–2024.

Ferguson, D.O., Alt, F.W. (2001) DNA double strand break repair and chromosomal translocation: lessons from animal models, *Oncogene* **20**, 5572–5579.

Flygare, J., Falt, S., Ottervald, J., Castro, J., Dackland, A.L., Hellgren, D., Wennborg, A. (2001) Effects of HsRad51 overexpression on cell proliferation, cell cycle progression, and apoptosis, *Exp. Cell Res.* **268**, 61–69.

Folger, K.R., Wong, E.A., Wahl, G., Capecchi, M.R. (1982) Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules, *Mol. Cell Biol.* **2**, 1372–1387.

Giovannangeli, C., Diviacco, S., Labrousse, V., Gryaznov, S., Charneau, P., Helene, C. (1997) Accessibility of nuclear DNA to triplex-forming oligonucleotides: the integrated HIV-1 provirus as a target, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 79–84.

Giovannangeli, C., Perrouault, L., Escude, C., Gryaznov, S., Helene, C. (1996) Efficient inhibition of transcription elongation in vitro by oligonucleotide phosphoramidates targeted to proviral HIV DNA, *J. Mol. Biol.* **261**, 386–398.

Golub, E.I., Kovalenko, O.V., Gupta, R.C., Ward, D.C., Radding, C.M. (1997) Interaction of human recombination proteins Rad51 and Rad54, *Nucleic Acids Res.* **25**, 4106–4110.

Gottesfeld, J.M., Belitsky, J.M., Melander, C., Dervan, P.B., Luger, K. (2002) Blocking transcription through a nucleosome with synthetic DNA ligands, *J. Mol. Biol.* **321**, 249–263.

Gottesfeld, J.M., Melander, C., Suto, R.K., Raviol, H., Luger, K., Dervan, P.B. (2001) Sequence-specific recognition of DNA in the nucleosome by pyrrole–imidazole polyamides, *J. Mol. Biol.* **309**, 615–629.

Gottesfeld, J.M., Neely, L., Trauger, J.W., Baird, E.E., Dervan, P.B. (1997) Regulation of gene expression by small molecules, *Nature* **387**, 202–205.

Gupta, R.C., Folta-Stogniew, E., O'Malley, S., Takahashi, M., Radding, C.M. (1999) Rapid exchange of A:T base pairs is essential for recognition of DNA homology by human Rad51 recombination protein, *Mol. Cell* **4**, 705–714.

Haber, J.E., Heyer, W.D. (2001) The fuss about Mus81, *Cell* **107**, 551–554.

Havre, P.A., Gunther, E.J., Gasparro, F.P., Glazer, P.M. (1993) Targeted mutagenesis of DNA using triple helix-forming oligonucleotides linked to psoralen, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7879–7883.

Helene, C. (1991) The antigene strategy: control of gene expression by triplex-forming-oligonucleotides, *Anti-Cancer Drug Des.* **6**, 569–584.

Helene, C., Thuong, N.T., Harel-Bellan, A. (1992) Control of gene expression by triple helix-forming oligonucleotides. The antigene strategy, *Ann. N. Y. Acad. Sci.* **660**, 27–36.

Hirata, R., Chamberlain, J., Dong, R., Russell, D.W. (2002) Targeted transgene insertion into human chromosomes by adeno-associated virus vectors, *Nat. Biotechnol.* **20**, 735–738.

Howard-Flanders, P., West, S.C., Stasiak, A. (1984) Role of RecA protein spiral filaments in genetic recombination, *Nature* **309**, 215–219.

Hunger-Bertling, K., Harrer, P., Bertling, W. (1990) Short DNA fragments induce site specific recombination in mammalian cells, *Mol. Cell Biochem.* **92**, 107–116.

Hutchison, C.A. III, Edgell, M.H. (1971) Genetic assay for small fragments of bacteriophage ϕ X174 deoxyribonucleic acid, *J. Virol.* **8**, 181–189.

Inoue, N., Hirata, R.K., Russell, D.W. (1999) High-fidelity correction of mutations at multiple chromosomal positions by adeno-associated virus vectors, *J. Virol.* **73**, 7376–7380.

Janssen, S., Cuvier, O., Muller, M., Laemmli, U.K. (2000a) Specific gain- and loss-of-function phenotypes induced by satellite-specific DNA-binding drugs fed to drosophila melanogaster, *Mol. Cell* **6**, 1013–1024.

Janssen, S., Durussel, T., Laemmli, U.K. (2000b) Chromatin opening of DNA satellites by targeted sequence-specific drugs, *Mol. Cell* **6**, 999–1011.

Jasin, M., Elledge, S.J., Davis, R.W., Berg, P. (1990) Gene targeting at the human CD4 locus by epitope addition, *Genes Dev.* **4**, 157–166.

Johnson, R.D., Jasin, M. (2000) Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells, *EMBO J.* **19**, 3398–3407.

Kabotyanski, E.B., Gomelsky, L., Han, J.O., Stamato, T.D., Roth, D.B. (1998) Double-strand break repair in Ku86- and XRCC4-deficient cells, *Nucleic Acids Res.* **26**, 5333–5342.

Kaihatsu, K., Braasch, D.A., Cansizoglu, A., Corey, D.R. (2002) Enhanced strand invasion by peptide nucleic acid-peptide conjugates, *Biochemistry* **41**, 11118–11125.

Kapsa, R., Quigley, A., Lynch, G.S., Steeper, K., Kornberg, A.J., Gregorevic, P., Austin, L., Byrne, E. (2001) In vivo and in vitro correction of the mdx dystrophin gene nonsense mutation by short-fragment homologous replacement, *Hum. Gene Ther.* **12**, 629–642.

Kelly, J.J., Baird, E.E., Dervan, P.B. (1996) Binding site size limit of the 2 : 1 pyrrole-imidazole polyamide-DNA motif, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6981–6985.

Kenner, O., Kneisel, A., Klingler, J., Bartelt, B., Speit, G., Vogel, W., Kaufmann, D. (2002) Targeted gene correction of hprt mutations by 45 base single-stranded oligonucleotides, *Biochem. Biophys. Res. Commun.* **299**, 787–792.

Khorlin, A.A., Krylov, A.S., Grokhovsky, S.L., Zhuze, A.L., Zasedatelev, A.S., Gursky, G.V., Gottikh, B.P. (1980) A new type of AT-specific ligand constructed of two netropsin-like molecules, *FEBS Lett.* **118**, 311–314.

Kim, H.G., Miller, D.M. (1998) A novel triplex-forming oligonucleotide targeted to human cyclin D1 (bcl-1, proto-oncogene) promoter inhibits transcription in HeLa cells, *Biochemistry* **37**, 2666–2672.

Kim, Y.G., Cha, J., Chandrasegaran, S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1156–1160.

Kochetkova, M., Iversen, P.O., Lopez, A.F., Shannon, M.F. (1997) Deoxyribonucleic acid triplex formation inhibits granulocyte macrophage colony-stimulating factor gene expression and suppresses growth in juvenile myelomonocytic leukemic cells, *J. Clin. Invest.* **99**, 3000–3008.

Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P., Dickerson, R.E. (1985) The molecular origin

of DNA-drug specificity in netropsin and distamycin, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1376–1380.

Kuhn, H., Demidov, V.V., Frank-Kamenetskii, M.D., Nielsen, P.E. (1998) Kinetic sequence discrimination of cationic bis-PNAs upon targeting of double-stranded DNA, *Nucleic Acids Res.* **26**, 582–587.

Kukreti, S., Sun, J.S., Garestier, T., Helene, C. (1997) Extension of the range of DNA sequences available for triple helix formation: stabilization of mismatched triplexes by acridine-containing oligonucleotides, *Nucleic Acids Res.* **25**, 4264–4270.

Kumar, K.A., Muniyappa, K. (1992) Use of structure-directed DNA ligands to probe the binding of recA protein to narrow and wide grooves of DNA and on its ability to promote homologous pairing, *J. Biol. Chem.* **267**, 24824–24832.

Lacroix, L., Lacoste, J., Reddoch, J.F., Mergny, J.L., Levy, D.D., Seidman, M.M., Matteucci, M.D., Glazer, P.M. (1999) Triplex formation by oligonucleotides containing 5-(1-propynyl)-2′- deoxyuridine: decreased magnesium dependence and improved intracellular gene targeting, *Biochemistry* **38**, 1893–1901.

Lee, J.S., Woodsworth, M.L., Latimer, L.J., Morgan, A.R. (1984) Poly(pyrimidine). poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH, *Nucleic Acids Res.* **12**, 6603–6614.

Letai, A.G., Palladino, M.A., Fromm, E., Rizzo, V., Fresco, J.R. (1988) Specificity in formation of triple-stranded nucleic acid helical complexes: studies with agarose-linked polyribonucleotide affinity columns, *Biochemistry* **27**, 9108–9112.

Liang, F., Han, M., Romanienko, P.J., Jasin, M. (1998) Homology-directed repair is a major double-strand break repair pathway in mammalian cells, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5172–5177.

Liang, F., Jasin, M. (1996) Ku80-deficient cells exhibit excess degradation of extrachromosomal DNA, *J. Biol. Chem.* **271**, 14405–14411.

Liang, F., Romanienko, P.J., Weaver, D.T., Jeggo, P.A., Jasin, M. (1996) Chromosomal double-strand break repair in Ku80-deficient cells, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8929–8933.

Liskay, R.M., Letsou, A., Stachelek, J.L. (1987) Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells, *Genetics* **115**, 161–167.

Lown, J.W., Krowicki, K., Bhat, U.G., Skorobogaty, A., Ward, B., Dabrowiak, J.C. (1986) Molecular recognition between oligopeptides and nucleic acids: novel imidazole-containing oligopeptides related to netropsin that exhibit altered DNA sequence specificity, *Biochemistry* **25**, 7408–7416.

Lukacsovich, T., Waldman, B.C., Waldman, A.S. (2001) Efficient recruitment of transfected DNA to a homologous chromosomal target in mammalian cells, *Biochim. Biophys. Acta* **1521**, 89–96.

Luo, Z., Macris, M.A., Faruqi, A.F., Glazer, P.M. (2000) High-frequency intrachromosomal gene conversion induced by triplex- forming oligonucleotides microinjected into mouse cells, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9003–9008.

Macris, M.A., Glazer, P.M. (2002) Transcription dependence of chromosomal gene targeting by triplex- forming oligonucleotides, *J. Biol. Chem.* **278**, 3357–3362.

Maeshima, K., Janssen, S., Laemmli, U.K. (2001) Specific targeting of insect and vertebrate telomeres with pyrrole and imidazole polyamides, *EMBO J.* **20**, 3218–3228.

Majumdar, A., Khorlin, A., Dyatkina, N., Lin, F.L., Powell, J., Liu, J., Fei, Z., Khripine, Y., Watanabe, K.A., George, J., Glazer, P.M., Seidman, M.M. (1998) Targeted gene knockout mediated by triple helix forming oligonucleotides, *Nat. Genet.* **20**, 212–214.

Majumdar, A., Puri, N., Cuenoud, B., Natt, F., Martin, P., Khorlin, A., Dyatkina, N., George, A.J., Miller, P.S., Seidman, M.M. (2003) Cell cycle modulation of gene targeting by a triple helix forming oligonucleotide, *J. Biol. Chem.* **278**, 11072–11077.

Manor, H., Rao, B.S., Martin, R.G. (1988) Abundance and degree of dispersion of genomic d(GA)n.d(TC)n sequences, *J. Mol. Evol.* **27**, 96–101.

Mansour, S.L., Thomas, K.R., Deng, C.X., Capecchi, M.R. (1990) Introduction of a lacZ reporter gene into the mouse int-2 locus by homologous recombination, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7688–7692.

Mapp, A.K., Ansari, A.Z., Ptashne, M., Dervan, P.B. (2000) Activation of gene expression by small molecule transcription factors, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3930–3935.

Mayfield, C., Ebbinghaus, S., Gee, J., Jones, D., Rodu, B., Squibb, M., Miller, D. (1994) Triplex formation by the human Ha-ras promoter inhibits Sp1 binding and in vitro transcription, *J. Biol. Chem.* **269**, 18232–18238.

Mazin, A.V., Bornarth, C.J., Solinger, J.A., Heyer, W.D., Kowalczykowski, S.C. (2000) Rad54 protein is targeted to pairing loci by the Rad51 nucleoprotein filament, *Mol. Cell* **6**, 583–592.

McCulloch, R.D., Read, L.R., Baker, M.D. (2003) Strand invasion and DNA synthesis from the two 3′ ends of a DSB, *Genetics* **163**, 1439–1447.

Melander, C., Herman, D.M., Dervan, P.B. (2000) Discrimination of A/T sequences in the minor groove of DNA within a cyclic polyamide motif, *Chemistry* **6**, 4487–4497.

Merrihew, R.V., Marburger, K., Pennington, S.L., Roth, D.B., Wilson, J.H. (1996) High-frequency illegitimate integration of transfected DNA at preintegrated target sites in a mammalian genome, *Mol. Cell Biol.* **16**, 10–18.

Miller, D.G., Petek, L.M., Russell, D.W. (2003) Human gene targeting by adeno-associated virus vectors is enhanced by DNA double strand breaks, *Mol. Cell Biol.* **23**, 3550–3557.

Moerschell, R.P., Das, G., Sherman, F. (1991) Transformation of yeast directly with synthetic oligonucleotides, *Methods Enzymol.* **194**, 362–369.

Mollegaard, N.E., Buchardt, O., Egholm, M., Nielsen, P.E. (1994) Peptide nucleic acid.DNA strand displacement loops as artificial transcription promoters, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3892–3895.

Moser, H.E., Dervan, P.B. (1987) Sequence-specific cleavage of double helical DNA by triple helix formation, *Science* **238**, 645–650.

Mrksich, M., Parks, M.E., Dervan, P.B. (1994) Hairpin peptide motif: A new class of oligopeptides for sequence specific recognition in the minor groove of double helical DNA, *J. Am. Chem. Soc.* **116**, 7983–7988.

Mrksich, M., Wade, W.S., Dwyer, T.J., Geierstanger, B.H., Wemmer, D.E., Dervan, P.B. (1992) Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7586–7590.

Myung, K., Datta, A., Chen, C., Kolodner, R.D. (2001) SGS1, the Saccharomyces cerevisiae homologue of BLM and WRN, suppresses genome instability and homeologous recombination, *Nat. Genet.* **27**, 113–116.

Nielsen, P.E., Egholm, M., Berg, R.H., Buchardt, O. (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide, *Science* **254**, 1497–1500.

Nielsen, P.E., Egholm, M., Buchardt, O. (1994) Evidence for (PNA)2/DNA triplex structure upon binding of PNA to dsDNA by strand displacement, *J. Mol. Recognit.* **7**, 165–170.

Nishinaka, T., Shinohara, A., Ito, Y., Yokoyama, S., Shibata, T. (1998) Base pair switching by interconversion of sugar puckers in DNA extended by proteins of RecA-family: a model for homology search in homologous genetic recombination, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11071–11076.

O'Hare, C.C., Mack, D., Tandon, M., Sharma, S.K., Lown, J.W., Kopka, M.L., Dickerson, R.E., Hartley, J.A. (2002) DNA sequence recognition in the minor groove by cross-linked polyamides: The effect of N-terminal head group and linker length on binding affinity and specificity, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 72–77.

Orr-Weaver, T.L., Szostak, J.W., Rothstein, R.J. (1981) Yeast transformation: a model system for the study of recombination, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6354–6358.

Orson, F.M., Thomas, D.W., McShan, W.M., Kessler, D.J., Hogan, M.E. (1991) Oligonucleotide inhibition of IL2R alpha mRNA transcription by promoter region collinear triplex formation in lymphocytes, *Nucleic Acids Res.* **19**, 3435–3441.

Paques, F., Haber, J.E. (1999) Multiple pathways of recombination induced by double-strand breaks in saccharomyces cerevisiae, *Microbiol. Mol. Biol. Rev.* **63**, 349–404.

Park, M.S. (1995) Expression of human RAD52 confers resistance to ionizing radiation in mammalian cells, *J. Biol. Chem.* **270**, 15467–15470.

Pazin, M.J., Kadonaga, J.T. (1997) SWI2/SNF2 and related proteins: ATP-driven motors that disrupt protein–DNA interactions? *Cell* **88**, 737–740.

Peffer, N.J., Hanvey, J.C., Bisi, J.E., Thomson, S.A., Hassman, C.F., Noble, S.A., Babiss, L.E. (1993) Strand-invasion of duplex DNA by peptide nucleic acid oligomers, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10648–10652.

Pelton, J.G., Wemmer, D.E. (1989) Structural characterization of a 2:1 distamycin A.d(CGCAAATTGGC) complex by two-dimensional NMR, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5723–5727.

Pennington, S.L., Wilson, J.H. (1991) Gene targeting in Chinese hamster ovary cells is conservative, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9498–9502.

Perkins, B.D., Wilson, J.H., Wensel, T.G., Vasquez, K.M. (1998) Triplex targets in the human rhodopsin gene, *Biochemistry* **37**, 11315–11322.

Pesco, J., Salmon, J.M., Vigo, J., Viallet, P. (2001) Mag-indo1 affinity for Ca(2+), compartmentalization and binding to proteins: the challenge of measuring Mg(2+) concentrations in living cells, *Anal. Biochem.* **290**, 221–231.

Petukhova, G., Stratton, S., Sung, P. (1998) Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins, *Nature* **393**, 91–94.

Petukhova, G., Van Komen, S., Vergano, S., Klein, H., Sung, P. (1999) Yeast Rad54 promotes Rad51-dependent homologous DNA pairing via ATP hydrolysis-driven change in DNA double helix conformation, *J. Biol. Chem.* **274**, 29453–29462.

Pierce, A.J., Hu, P., Han, M., Ellis, N., Jasin, M. (2001) Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells, *Genes Dev.* **15**, 3237–3242.

Podyminogin, M.A., Meyer, R.B., Gamper, H.B. (1995) Sequence-specific covalent modification of DNA by cross-linking oligonucleotides. Catalysis by RecA and implication for the mechanism of synaptic joint formation, *Biochemistry* **34**, 13098–13108.

Porteus, M.H., Baltimore, D. (2003) Chimeric nucleases stimulate gene targeting in human cells, *Science* **300**, 763.

Porteus, M.H., Cathomen, T., Weitzman, M.D., Baltimore, D. (2003) Efficient gene targeting mediated by AAV and DNA double strand breaks, *Mol. Cell Biol.* **23**, 3558–3565.

Porumb, H., Gousset, H., Letellier, R., Salle, V., Briane, D., Vassy, J., Amor-Gueret, M., Israel, L., Taillandier, E. (1996) Temporary ex vivo inhibition of the expression of the human oncogene HER2 (NEU) by a triple helix-forming oligonucleotide, *Cancer Res.* **56**, 515–522.

Puri, N., Majumdar, A., Cuenoud, B., Natt, F., Martin, P., Boyd, A., Miller, P.S., Seidman, M.M. (2002) Minimum number of 2′-O-(2-aminoethyl) residues required for gene knockout activity by triple helix forming oligonucleotides, *Biochemistry* **41**, 7716–7724.

Rapozzi, V., Cogoi, S., Spessotto, P., Risso, A., Bonora, G.M., Quadrifoglio, F., Xodo, L.E. (2002) Antigene effect in k562 cells of a peg-conjugated triplex-forming oligonucleotide targeted to the bcr/abl oncogene, *Biochemistry* **41**, 502–510.

Rayssiguier, C., Thaler, D.S., Radman, M. (1989) The barrier to recombination between Escherichia coli and Salmonella typhimurium is disrupted in mismatch-repair mutants, *Nature* **342**, 396–401.

Rice, K.P., Chaput, J.C., Cox, M.M., Switzer, C. (2000) RecA protein promotes strand exchange with DNA substrates containing isoguanine and 5-methyl isocytosine, *Biochemistry* **39**, 10177–10188.

Richardson, C., Moynahan, M.E., Jasin, M. (1998) Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations, *Genes Dev.* **12**, 3831–3842.

Roberts, R.W., Crothers, D.M. (1992) Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition, *Science* **258**, 1463–1466.

Rogers, F.A., Vasquez, K.M., Egholm, M., Glazer, P.M. (2002) Site-directed recombination via bifunctional PNA-DNA conjugates, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16695–16700.

Rommerskirch, W., Graeber, I., Grassmann, M., Grassmann, A. (1988) Homologous recombination of SV40 DNA in COS7 cells occurs with high frequency in a gene dose independent fashion, *Nucleic Acids Res.* **16**, 941–952.

Russell, D.W., Hirata, R.K. (1998) Human gene targeting by viral vectors, *Nat. Genet.* **18**, 325–330.

Sangiuolo, F., Bruscia, E., Serafino, A., Nardone, A., Bonifazi, E., Lais, M., Gruenert, D.C., Novelli, G. (2002) In vitro correction of cystic fibrosis epithelial cell lines by small fragment homologous replacement (SFHR) technique, *BMC. Med. Genet.* **3**, 8–21.

Sargent, R.G., Brenneman, M.A., Wilson, J.H. (1997) Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination, *Mol. Cell Biol.* **17**, 267–277.

Sargent, R.G., Wilson, J.H. (1998) Recombination and gene targeting in mammalian cells, *Curr. Res. Mol. Ther.* **1**, 584–692.

Scheerer, J.B., Adair, G.M. (1994) Homology dependence of targeted recombination at the Chinese hamster APRT locus, *Mol. Cell Biol.* **14**, 6663–6673.

Seeman, N.C., Rosenberg, J.M., Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 804–808.

Seidman, M.M., Bredberg, A., Seetharam, S., Kraemer, K.H. (1987) Multiple point mutations in a shuttle vector propagated in human cells: evidence for an error-prone DNA polymerase activity, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4944–4948.

Selva, E.M., New, L., Crouse, G.F., Lahue, R.S. (1995) Mismatch correction acts as a barrier to homeologous recombination in Saccharomyces cerevisiae, *Genetics* **139**, 1175–1188.

Shimizu, M., Konishi, A., Shimada, Y., Inoue, H., Ohtsuka, E. (1992) Oligo(2′-O-methyl)ribonucleotides. Effective probes for duplex DNA, *FEBS Lett.* **302**, 155–158.

Shulman, M.J., Nissen, L., Collins, C. (1990) Homologous recombination in hybridoma cells: dependence on time and fragment length, *Mol. Cell Biol.* **10**, 4466–4472.

Smih, F., Rouet, P., Romanienko, P.J., Jasin, M. (1995) Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells, *Nucleic Acids Res.* **23**, 5012–5019.

Smith, J., Bibikova, M., Whitby, F.G., Reddy, A.R., Chandrasegaran, S., Carroll, D. (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains, *Nucleic Acids Res.* **28**, 3361–3369.

Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A., Kucherlapati, R.S. (1985) Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination, *Nature* **317**, 230–234.

Smolik-Utlaut, S., Petes, T.D. (1983) Recombination of plasmids into the Saccharomyces cerevisiae chromosome is reduced by small amounts of sequence heterogeneity, *Mol. Cell Biol.* **3**, 1204–1211.

Song, K.Y., Schwartz, F., Maeda, N., Smithies, O., Kucherlapati, R. (1987) Accurate modification of a chromosomal plasmid by homologous recombination in human cells, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6820–6824.

Sonoda, E., Sasaki, M.S., Buerstedde, J.M., Bezzubova, O., Shinohara, A., Ogawa, H., Takata, M., Yamaguchi-Iwai, Y., Takeda, S. (1998) Rad51-deficient vertebrate cells accumulate chromosomal breaks prior to cell death, *EMBO J.* **17**, 598–608.

Stasiak, A., Di Capua, E., Koller, T. (1981) Elongation of duplex DNA by recA protein, *J. Mol. Biol.* **151**, 557–564.

Stein, C.A. (1997) Controversies in the cellular pharmacology of oligodeoxynucleotides, *Antisense Nucleic Acid Drug Dev.* **7**, 207–209.

Supekova, L., Pezacki, J.P., Su, A.I., Loweth, C.J., Riedl, R., Geierstanger, B., Schultz, P.G., Wemmer, D.E. (2002) Genomic effects of polyamide/DNA interactions on mRNA expression, *Chem. Biol.* **9**, 821–827.

Szostak, J.W., Wu, R. (1979) Insertion of a genetic marker into the ribosomal DNA of yeast, *Plasmid* **2**, 536–554.

Takasugi, M., Guendouz, A., Chassignol, M., Decout, J.L., Lhomme, J., Thuong, N.T., Helene, C. (1991) Sequence-specific photo-induced cross-linking of the two strands of double-helical DNA by a psoralen covalently linked to a triple helix-forming oligonucleotide, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 5602–5606.

Takata, M., Sasaki, M.S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A., Takeda, S. (1998) Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells, *EMBO J.* **17**, 5497–5508.

Taubes, G. (2002a) Gene therapy. Pioneering papers under the microscope, *Science* **298**, 2118–2119.

Taubes, G. (2002b) Gene therapy. The strange case of chimeraplasty, *Science* **298**, 2116–2120.

te Riele H., Maandag, E.R., Berns, A. (1992) Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5128–5132.

Thomas, K.R., Capecchi, M.R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells, *Cell* **51**, 503–512.

Thomas, K.R., Folger, K.R., Capecchi, M.R. (1986) High frequency targeting of genes to specific sites in the mammalian genome, *Cell* **44**, 419–428.

Torigoe, H., Hari, Y., Sekiguchi, M., Obika, S., Imanishi, T. (2001) 2′-O,4′-C-methylene bridged nucleic acid modification promotes pyrimidine motif triplex DNA formation at physiological pH: thermodynamic and kinetic studies, *J. Biol. Chem.* **276**, 2354–2360.

Trauger, J.W., Baird, E.E., Dervan, P.B. (1998) Recognition of 16 base pairs in the minor groove of DNA by a pyrrole–imidazole polyamide dimer. *J. Am. Chem. Soc.* **120**, 3534–3535.

Trauger, J.W., Baird, E.E., Mrksich, M., Dervan, P.B. (1996a) Extension of sequence specific recognition in the minor groove of DNA by pyrrole–imidazole polyamides to 9–13 base pairs, *J. Am. Chem. Soc.* **118**, 6160–6166.

Trauger, K.W., Baird, E.E., Dervan, P.B. (1996b) Recognition of DNA by designed ligands at subnanomolar concentrations, *Nature* **382**, 559–561.

Tzung, T.Y., Runger, T.M. (1998) Reduced joining of DNA double strand breaks with an abnormal mutation spectrum in rodent mutants of DNA-PKcs and Ku80, *Int. J. Radiat. Biol.* **73**, 469–474.

van der Steege, G., Schuilenga-Hut, P.H., Buys, C.H., Scheffer, H., Pas, H.H., Jonkman, M.F. (2001) Persistent failures in gene repair, *Nat. Biotechnol.* **19**, 305–306.

Vasquez, K.M., Christensen, J., Li, L., Finch, R.A., Glazer, P.M. (2002) Human XPA and RPA DNA repair proteins participate in specific recognition of triplex-induced helical distortions, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5848–5853.

Vasquez, K.M., Dagle, J.M., Weeks, D.L., Glazer, P.M. (2001a) Chromosome targeting at short polypurine sites by cationic triplex-forming oligonucleotides, *J. Biol. Chem.* **276**, 38536–38541.

Vasquez, K.M., Narayanan, L., Glazer, P.M. (2000) Specific mutations induced by triplex-forming oligonucleotides in mice, *Science* **290**, 530–533.

Vasquez, K.M., Wang, G., Havre, P.A., Glazer, P.M. (1999) Chromosomal mutations induced by triplex-forming oligonucleotides in mammalian cells, *Nucleic Acids Res.* **27**, 1176–1181.

Vasquez, K.M., Wensel, T.G., Hogan, M.E., Wilson, J.H. (1995) High-affinity triple helix formation by synthetic oligonucleotides at a site within a selectable mammalian gene, *Biochemistry* **34**, 7243–7251.

Vasquez, K.M., Wilson, J.H. (1998) Triplex-directed modification of genes and gene activity, *Trends Biochem. Sci.* **23**, 4–9.

Wake, C.T., Wilson, J.H. (1979) Simian virus 40 recombinants are produced at high frequency during infection with genetically mixed oligomeric DNA, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 2876–2880.

Wake, C.T., Wilson, J.H. (1980) Defined oligomeric SV40 DNA: a sensitive probe of general recombination in somatic cells, *Cell* **21**, 141–148.

Waldman, A.S., Liskay, R.M. (1988) Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology, *Mol. Cell Biol.* **8**, 5350–5357.

Wang, G., Levy, D.D., Seidman, M.M., Glazer, P.M. (1995) Targeted mutagenesis in mammalian cells mediated by intracellular triple helix formation, *Mol. Cell Biol.* **15**, 1759–1768.

Wang, G., Seidman, M.M., Glazer, P.M. (1996) Mutagenesis in mammalian cells induced by triple helix formation and transcription-coupled repair, *Science* **271**, 802–805.

Wang, G., Xu, X., Pace, B., Dean, D.A., Glazer, P.M., Chan, P., Goodman, S.R., Shokolenko, I. (1999) Peptide nucleic acid (PNA) binding-mediated induction of human gamma-globin gene expression, *Nucleic Acids Res.* **27**, 2806–2813.

Wartell, R.M., Larson, J.E., Wells, R.D. (1974) Netropsin. A specific probe for A-T regions of duplex deoxyribonucleic acid, *J. Biol. Chem.* **249**, 6719–6731.

White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E., Dervan, P.B. (1998) Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands [see comments], *Nature* **391**, 468–471.

Wilson, J.H. (1979) Nick-free formation of reciprocal heteroduplexes: a simple solution to the topological problem, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3641–3645.

Wilson, J.H., Leung, W.Y., Bosco, G., Dieu, D., Haber, J.E. (1994) The frequency of gene targeting in yeast depends on the number of target copies, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 177–181.

Wolfe, S.A., Greisman, H.A., Ramm, E.I., Pabo, C.O. (1999) Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code, *J. Mol. Biol.* **285**, 1917–1934.

Wurtz, N.R., Pomerantz, J.L., Baltimore, D., Dervan, P.B. (2002) Inhibition of DNA binding by NF-kappa B with pyrrole–imidazole polyamides, *Biochemistry* **41**, 7604–7609.

Xodo, L.E., Manzini, G., Quadrifoglio, F., van der Marel, G.A., van Boom, J.H. (1991) Effect of 5-methylcytosine on the stability of triple-stranded DNA–a thermodynamic study, *Nucleic Acids Res.* **19**, 5625–5631.

Yamamoto, A., Taki, T., Yagi, H., Habu, T., Yoshida, K., Yoshimura, Y., Yamamoto, K., Matsushiro, A., Nishimune, Y., Morita, T. (1996) Cell cycle-dependent expression of the mouse Rad51 gene in proliferating cells, *Mol. Gen. Genet.* **251**, 1–12.

Yanez, R.J., Porter, A.C. (1999) Gene targeting is enhanced in human cells overexpressing hRAD51, *Gene Ther.* **6**, 1282–1290.

Yanez, R.J., Porter, A.C. (2002) Differential effects of Rad52p overexpression on gene targeting and extrachromosomal homologous recombination in a human cell line, *Nucleic Acids Res.* **30**, 740–748.

Yu, D., Ellis, H.M., Lee, E.C., Jenkins, N.A., Copeland, N.G., Court D.L. (2000) An efficient recombination system for chromosome engineering in Escherichia coli, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5978–5983.

Zheng, H., Hasty, P., Brenneman, M.A., Grompe, M., Gibbs, R.A., Wilson, J.H., Bradley, A. (1991) Fidelity of targeted recombination in human fibroblasts and murine embryonic stem cells, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8067–8071.

Zheng, H., Wilson, J.H. (1990) Gene targeting in normal and amplified cell lines, *Nature* **344**, 170–173.

Zhou, X., Adzuma, K. (1997) DNA strand exchange mediated by the Escherichia coli RecA protein initiates in the minor groove of double-stranded DNA, *Biochemistry* **36**, 4650–4661.

Zimmer, A., Gruss, P. (1989) Production of chimaeric mice containing embryonic stem (ES) cells carrying a homeobox Hox 1.1 allele mutated by homologous recombination, *Nature* **338**, 150–153.

Gene Therapy and Cardiovascular Diseases

Michael E. Rosenfeld[1] *and Alan D. Attie*[2]
[1] *Department of Pathobiology, University of Washington, Seattle, WA, USA*
[2] *Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, USA*

1 Overview of Gene Therapy 217

2 Atherosclerosis: A Chronic Inflammatory Disease 218

3 Gene Therapy Targets for Cardiovascular Disease in the Lipoprotein Pathways 221
3.1 The Lipoprotein Pathways Provide Many Potential Gene Therapy Targets 221
3.2 Specific Genes are Critical for Dietary Lipid Absorption in the Intestine 222
3.3 The Intestine Exports Lipids on Chylomicron Particles 223
3.4 The Liver is a Major Producer of Plasma Lipoproteins 223
3.5 Lipoprotein Lipase as a Target for Gene Therapy 224
3.6 The LDL and VLDL Receptors are Attractive Targets for Gene Therapy 225
3.7 Increased Cholesterol and Phospholipid Efflux via ABCA1 Can Prevent Atherosclerosis 226
3.8 Some Apolipoproteins Can Prevent Atherosclerosis 227
3.9 The Liver SR-B1 Receptor Protects against Atherosclerosis 228

4 Vascular and Inflammatory Targets for Gene Therapy 229
4.1 Adhesion Molecules and Leukocyte Counter Receptors 229
4.2 Cytokines 230
4.3 Chemokines and Chemokine Receptors 230
4.4 Interleukins, the TNF-α family, and Interferon-γ 231
4.5 Growth Factors and Receptors 232

5 Pro- and Antioxidant Enzymes 233

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

6 Scavenger Receptors 234

7 Transcription Factors 235

8 Noninflammatory Gene Targets 236
8.1 Cell Death 236
8.2 Proteases 236
8.3 Regulators of Vascular Tone and Blood Pressure 237
8.4 Coagulation Related Genes 237
8.5 Osteogenic Proteins 238

Bibliography 238
Books and Reviews 238
Primary Literature 239

Keywords

ABCA1
A membrane lipid–transport protein involved in cholesterol and/or phospholipid efflux from cells. Its action is necessary for the extracellular assembly of lipoprotein particles.

Adhesion Molecules
These are proteins on the luminal surface of endothelial cells, bind to counter receptors on the membrane of leukocytes, and enable the leukocytes to attach tightly to the surface of the endothelial cells.

Angina Pectoris
Chest pain caused by reduced blood flow to the heart due to the blockage of coronary arteries by blood clots.

Angiogenesis
The process of formation of new blood vessels.

Angioplasty
An intervention to increase blood flow in a clogged artery. It involves insertion of a balloon catheter into the blocked artery and subsequent inflation of the balloon to break open or squash the atherosclerotic lesion and/or blood clot that is reducing blood flow.

Apical Membrane
In polarized cells (e.g. hepatocytes in the liver, enterocytes in the intestine), the membrane exposed to the outside world; in the hepatocytes, the bile canalicular membrane; in enterocytes, the membrane exposed to the intestinal lumen.

ApoA1
An apolipoprotein principally associated with HDL, an activator of lecithin cholesterol: acyltransferase. It interacts with the cells to mediate delivery of cholesterol ester from HDL particles.

ApoB100
An apolipoprotein associated with VLDL and LDL particles, synthesized in the liver, a ligand for the LDL receptor.

ApoB48
An apolipoprotein associated with chylomicrons, synthesized in the intestine; it is a truncated form of apoB100 and does not bind to the LDL receptor.

ApoE
An apolipoprotein principally associated with VLDL and chylomicrons, is responsible for the receptor-mediated clearance of IDL and chylomicron remnants, is a ligand for most members of the LDL-receptor superfamily. The apoE4 isoform is associated with increased risk of Alzheimer's Disease.

Apolipoproteins
The protein components of plasma lipoproteins.

Apoptosis
The programmed cell death of a cell or cellular suicide. It is a defined molecular pathway where proteolytic enzymes are activated to destroy key cellular proteins.

Atherosclerosis
An inflammatory disease in the arterial wall leading to the accumulation of cellular outgrowths that can become unstable, break off, and cause the formation of blood clots.

Basolateral Membrane
In polarized cells, the membrane exposed to the inside world; in hepatocytes, the sinusoidal membrane, which is exposed to venous circulation, in enterocytes, the basolateral membrane, which is exposed to the lymphatics and venous circulation.

Bile Acids
Detergent-like molecules formed from cholesterol. They are secreted by the liver and together with cholesterol and phospholipids, form bile. Bile forms micelles that emulsify lipids in the intestinal lumen aiding in their absorption.

Chemokines
A category of cytokines that function specifically to chemically attract leukocytes to sites of inflammation.

Chylomicron Remnants
Chylomicron particles that have been depleted of triacylglycerol after the lipoprotein lipase–mediated hydrolysis of their triacylglycerols.

Chylomicrons
Lipoprotein particles produced in the intestine to package and secrete dietary lipids. Chylomicrons are secreted into the mesenteric lymph.

Cytokines
Cell regulatory proteins secreted by many cell types that are key factors in the initiation and control of inflammation.

Familial Hypercholesterolemia
Elevation in LDL cholesterol due to mutations at the LDL-receptor locus.

Familial Hypoalphalipoproteinemia
Deficiency in HDL cholesterol due to mutations at the ABCA1 locus.

Fatty Streak
Earliest stage in the development of an atherosclerotic lesion. It consists of aggregates of macrophage-derived foam cells and lymphocytes that form underneath the endothelial lining of an artery.

Fibrous Cap
A thin layer of smooth muscle cells and connective tissue that encapsulates areas of dead cells and lipid in a developing atherosclerotic lesion. If the fibrous cap is too thin, it can break and enable formation of blood clots that can clog arteries.

Foam Cells
Primarily macrophages that have taken up large numbers of modified lipoprotein particles and have stored the excess lipid in non membrane bound cholesteryl-ester droplets and in increased numbers of secondary lysosomes.

Gallstones
Large crystals, usually composed of cholesterol, that form in the biliary tract and/or gall bladder when cholesterol levels in bile are too high relative to phospholipid and bile acids.

Growth Factors
Secreted by many cell types and by binding to specific receptors on the plasma membrane of adjacent cells, are potent regulators of cellular functions, including proliferation, migration, differentiation, and survival/apoptosis.

HDL
High-density lipoprotein, carries about 20% of plasma cholesterol. Its levels negatively correlate with risk of coronary heart disease. Is thought to mediate ‘‘reverse cholesterol transport.’’

Hemorrhage
Bleeding into a tissue. In atherosclerotic lesions, it occurs when the lesion forms fissures or ruptures and is an indication that the lesion is unstable.

IDL
An intermediate density lipoprotein, also called ''VLDL remnant.'' It is the VLDL particle that has been depleted of triacylglycerol through the action of lipoprotein lipase.

Intima
The inner layer of a muscular artery. It is the space immediately behind the endothelium where atherosclerotic lesions develop. In normal, nondiseased human arteries, there is a diffuse thickening of the intima that contains connective tissue and sparse numbers of macrophages and smooth muscle cells.

Ischemia
The inadequate perfusion of a tissue with blood leading to the reduced availability of oxygen and nutrients in the tissue.

LDL
Low-density lipoprotein carries approximately two-thirds of plasma cholesterol. Its levels are positively correlated with the risk of coronary heart disease and is formed in the circulation from the catabolism of VLDL.

LDL Receptor
Expressed in most tissues and is mainly necessary for normal clearance of LDL from the bloodstream.

Lipoprotein Lipase
Catalyzes the hydrolysis of VLDL and LDL triacylglycerols to free fatty acids and glycerol, and is present on the luminal surface of the capillary endothelium.

Lipoproteins
Particles that transport lipids in the bloodstream.

Media
The middle layer of a muscular artery that contains concentric rings of smooth muscle cells oriented perpendicular to blood flow. The orientation allows the contraction of the smooth muscle cells to constrict the artery. Atherosclerotic lesions do not initially form in the media but can expand and replace the media in very advanced stages of the disease.

Myocardial Infarction
The necrotic death of heart tissue due to a sudden blockage of blood flow to the affected tissue. It most frequently occurs following the formation of a blood clot on top of an atherosclerotic lesion that has ruptured in a main coronary artery.

Necrotic Zones
Areas within an atherosclerotic lesion that contain dead cells and debris, and lipid droplets formed from dead foam cells and aggregated lipoproteins. The expansion of necrotic zones into a *necrotic core* can make the lesion unstable and at risk of rupture.

Neovascularization
The formation of new, small blood vessels within an advanced atherosclerotic lesion. Neovessels are often formed deep in the lesion where there is limited diffusion of oxygen and nutrients, and can be a conduit for bringing more leukocytes into the advanced lesions.

Restenosis
The process of reformation of an arterial lesion that reduces blood flow following an intervention procedure such as angioplasty. A restenotic lesion is not the same as an atherosclerotic lesion as it forms primarily because of the migration and proliferation of smooth muscle cells following injury to the artery caused by the intervention procedure. It occurs in up to 40% of people who have angioplasty to increase blood flow to the heart.

Scavenger Receptors
Bind to a wide range of molecules, including modified forms of LDL. Are involved in the accumulation of cholesterol in macrophages and smooth muscle cells in the arterial wall.

Statins
Drugs that inhibit the synthesis of cholesterol. They increase the abundance of the LDL receptor and thus reduce plasma LDL levels.

Tangier Disease
A severe HDL deficiency syndrome caused by homozygous mutations at the ABCA1 locus.

Thrombosis
The formation of blood clots. A thrombus consists of variable numbers of aggregated platelets, red blood cells, and fibrin. In the heart, brain, or peripheral blood vessels, it can be an *occlusive thrombus* that blocks blood flow, or a smaller *mural thrombus* that is attached to an atherosclerotic lesion and does not block blood flow.

Vasospasm
The transient constriction of a blood vessel. Depending on the extent of constriction, blood flow can be reduced leading to tissue ischemia.

VLDL
Very low-density lipoprotein, a triacylglycerol-rich lipoprotein assembled in the endoplasmic reticulum and Golgi of hepatocytes and then secreted into the bloodstream. While in the circulation, its triacylglycerol core is depleted through the action of lipoprotein lipase and it gives rise to low-density lipoprotein.

1
Overview of Gene Therapy

Gene therapy refers to the expression of a recombinant gene in a patient. The recombinant gene can be delivered to the patient directly or it can be incorporated into cells *in vitro*, followed by delivery of the genetically modified cells to the patient. DNA delivery can occur in the patient as naked DNA or within a biological or artificial delivery vehicle. The biological delivery vehicle is usually a virus. A wide variety of artificial delivery vehicles have been developed. They usually consist of lipid vesicles (liposomes) or positively charged polymers that complex with DNA and have a high affinity for cell surfaces that are negatively charged.

Genetic therapy for the treatment of cardiovascular disease (CVD) is in its infant stage despite the large number of potential target genes associated with the development of atherosclerosis. Cardiovascular disease lends itself to genetic therapy because of the focal nature of atherosclerotic lesions and the availability of techniques for direct delivery of the genes to the affected areas. The desire to conduct gene therapy for atherosclerosis is in large part motivated by the large number of transgenic mouse studies in which expression or deletion of a gene has a profound therapeutic effect on the animal. For example, overexpression of the apolipoprotein-A1 gene protects animals from atherosclerosis. A logical extension is to mimic these studies with gene therapy.

Although the gene therapy field is now more than 20 years old, it has not fulfilled its initial promise. From the beginning, it has been impossible to achieve sustained expression of foreign genes at high levels. There is still an incomplete understanding of the fundamental processes by which DNA enters the cells, is transported to the nucleus, integrates into chromosomal DNA, and interacts with transcriptional machinery.

Regardless of the vehicle, DNA must be administered into the circulatory system in order to gain access to a high proportion of cells in a solid tissue. However, most tissues are not in direct contact with blood. They are shielded from blood by a tight layer of endothelial cells that comprise the luminal surface of arteries and veins. Particles as large as viruses or DNA/polymer aggregates cannot cross the small gaps (fenestrae) between endothelial cells. An important exception is the vasculature that perfuses the liver. These vessels have large fenestrae, allowing large particles to gain access to the hepatocytes in the liver. For this reason, it is relatively easy to target the liver and quite difficult to target extrahepatic tissues.

There are three predominant types of viral vectors used in gene therapy; adenoviruses, adeno-associated viruses, and lentiviruses. *Adenoviruses* belong to a family of human DNA tumor viruses that can cause noncancerous respiratory tract infections in humans. Upon infection, the viral DNA remains episomal (outside the host genome), thus making expression transient. Adenoviruses can infect both dividing and nondividing cells. Adenoviral infection causes an inflammatory response, which curtails the infection and can produce severe reactions in some individuals.

Adeno-associated virus is a small nonpathogenic single-stranded DNA virus. On its own, it cannot replicate. It depends on simultaneous expression of genes provided by adenovirus. The virus infects a broad range of dividing and nondividing cells. It can maintain its DNA as an

episome or can, at low frequency, integrate into the host cell's chromosomal DNA. Adeno-associated virus is potentially attractive because its expression tends to be long-lived and it does not elicit an inflammatory reaction the way adenoviruses do.

Lentiviruses are retroviruses, RNA viruses that replicate through a DNA intermediate. HIV is a lentivirus. Lentiviruses have been modified to infect a broad range of cells. They can infect nondividing cells. As is the case with adeno and adeno-associated virus, lentiviruses are produced that are replication-defective for gene therapy applications.

Injection of *naked DNA* into muscle results in transfection of the cells *in vivo*. Intravascular injection of naked DNA leads to uptake by hepatocytes in the liver. Some of the cells stably incorporate the foreign DNA and express its genes. Since animals are continually exposed to foreign DNA coming from the diet and from symbiotic and infectious microorganisms, they have evolved numerous mechanisms to protect their genomes from invasion by foreign DNA. These include degradation by nucleases, inefficient transport to the nucleus, cell death upon DNA uptake, and inactivation of promoters. Naked DNA is still very attractive because it avoids the complications of viral vectors. This justifies a continued effort to overcome the obstacles to efficient and prolonged expression of genes from the naked DNA.

Liposomes are vesicles formed from lipids, usually, phospholipids. Cationic lipids form stable complexes with DNA and can therefore be incorporated into liposomes. These complexes are routinely used by researchers to introduce DNA into cell lines *in vitro*. Their use *in vivo* is attractive because it does not raise the safety issues associated with viral vectors. However, they are far less efficient than viruses so their utility is still limited.

An alternative to direct injection of DNA or a vehicle into a patient is to obtain cells from a patient and incorporate a foreign gene into the cells *in vitro*. Cells that have been selected for stably incorporating the foreign gene can then be given back to the patient. The benefits of this step cannot be overstated; instability of transfected DNA is a major unsolved problem with *in vivo* gene therapy. This approach has been used in atherosclerosis research with cells derived from bone marrow. Since the bone marrow contains precursors of red blood cells and cells of the immune system, this is a practical way to specifically target such cells. After reintroduction to the bone marrow, cells containing a foreign gene enter into the circulation and repopulate tissues. For example, the bone marrow is a source of circulating phagocytic cells (''monocytes'' in the circulation, ''macrophages'' in most tissues, and ''Kupffer cells'' in the liver). Animal experiments have shown dramatic alterations in atherosclerosis susceptibility with the replacement of macrophages in the bone marrow. Partial hepatectomy stimulates a robust regenerative response in the liver. This makes possible the introduction of hepatocytes that have been genetically modified into a partially hepatectomized liver. The new cells respond to the proliferative signals and contribute to the new cell mass of the liver.

2 Atherosclerosis: A Chronic Inflammatory Disease

Atherosclerosis is thought to be the result of a chronic fibro-proliferative

inflammatory response. Animal studies suggest that this inflammatory disease is initiated following the deposition, modification, and retention of lipoproteins within the artery wall. The initial inflammatory response involves the upregulation of expression of *adhesion molecules* and *chemokines* by endothelial cells and smooth muscle cells, leading to the recruitment of monocytes and lymphocytes into the arterial *intima*. The monocytes rapidly convert into tissue macrophages that have the capacity to scavenge modified lipoproteins, secrete proinflammatory *cytokines* and *growth factors*, and further regulate immune functions via presentation of antigens. The *scavenger receptor* mediated uptake of modified lipoproteins transforms the macrophages into lipid-loaded *foam cells*, the accumulation of which is the hallmark of the fatty streak, the initial stage of atherosclerosis (Fig. 1).

The transition of the *fatty streak* into more advanced stages of the disease appears to involve the coalescence of deposited lipids coupled with the death of

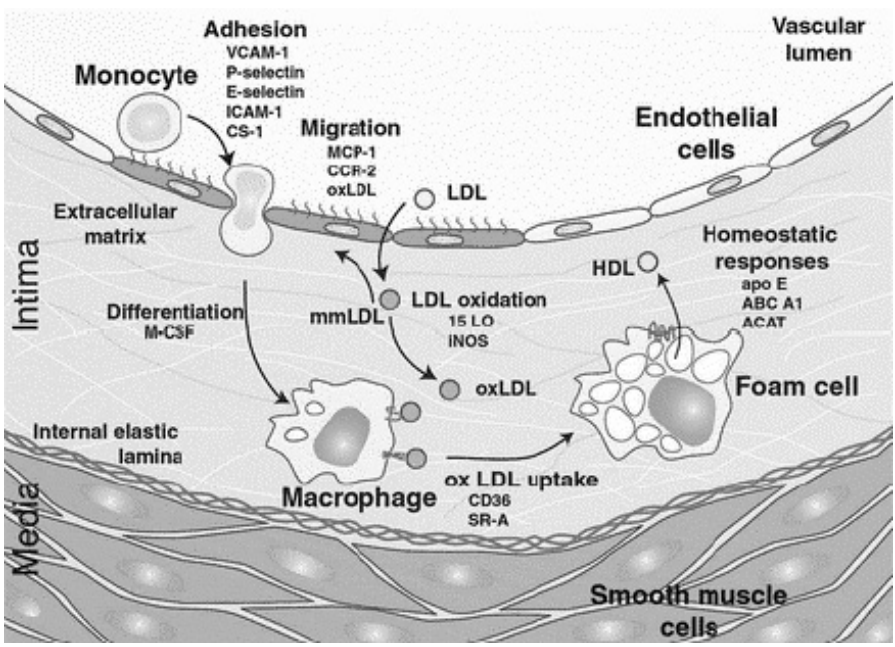

Fig. 1 Initiation of a fatty streak. Lipoproteins such as LDL become trapped within the extracellular matrix of the artery. The trapped lipoproteins are modified by processes such as oxidation or glycosylation (mmLDL, oxLDL) and become proinflammatory leading to the activation of endothelial cells to express adhesion molecules such as VCAM-1 and chemokines such as MCP-1. Monocytes are then recruited into the arterial intima and differentiate into macrophages most likely in response to increased expression of M-CSF. The modified lipoproteins are then taken up by scavenger receptors such as CD36 and the SRA1 expressed by the macrophages and this leads to foam cell formation. The excess cholesterol taken up by the foam cells is esterified by ACAT and stored in lipid droplets. It can be converted back to the more soluble free cholesterol and exported to extracellular HDL acceptors via cholesterol transporters, such as *ABC-A1*. (From Glass, C. K., Witztum, J. L. (2001) Atherosclerosis. The road ahead, *Cell* **104**(4), 503–516, included with permission of the Publisher.)

foam cells to form acellular *necrotic zones*. As in any classical inflammatory response, formation of the necrotic zones is accompanied by a wound-healing response where smooth muscle cells are recruited to encapsulate the necrotic zones by forming a *fibrous cap* around the lipid pools and necrotic debris (Fig. 2). With ongoing hyperlipidemia, however, there is a continuous influx, trapping, and modification of lipoproteins in the developing lesion that recruits additional inflammatory cells and leads to formation of new fatty streaks adjacent to or on top of the initial lesions. The combination of layering of new fatty streaks and erosion of the fibrous cap destabilizes the plaques allowing fissures, ruptures, and intraplaque *hemorrhage* to occur. While the exact mechanisms that cause plaques to rupture are currently unknown, the activation and/or death of macrophages with release of proteolytic enzymes such as *matrix metalloproteinases*, likely play a role. The resulting erosion and rupture exposes procoagulant proteins such as *tissue factor* to the blood and facilitates the formation of mural and *occlusive thrombi* (Fig. 3). The occlusion of the main coronary arteries causes the clinical outcomes of myocardial ischemia, *angina*, and ultimately *myocardial infarction*. Occlusion of cerebral vessels can lead to ischemic stroke while occlusion of peripheral blood vessels

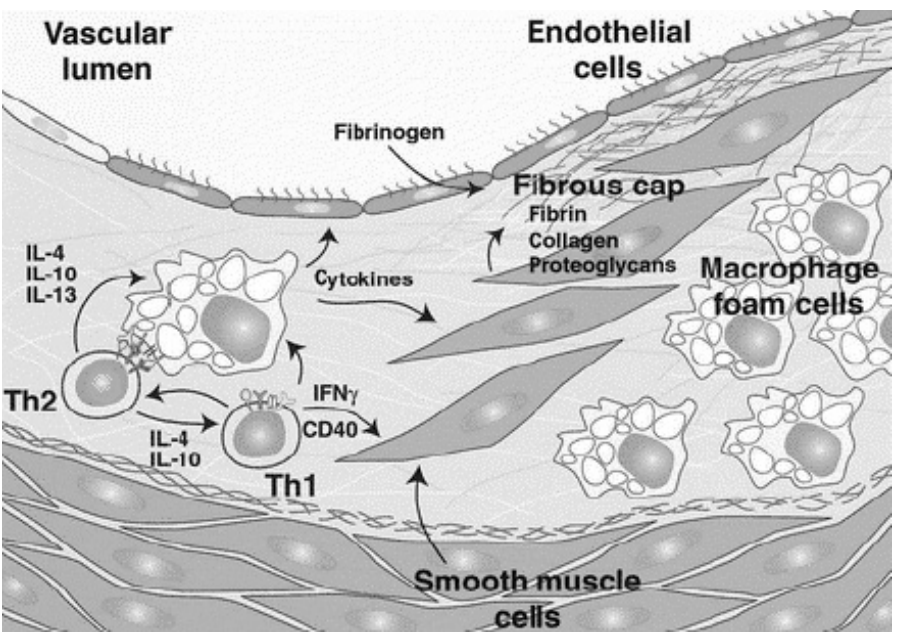

Fig. 2 Atherosclerotic lesion progression. Continued deposition of lipoproteins coupled with foam cell formation and interactions between foam cells, Th1 and Th2 T Helper cells and smooth muscle cells contribute to a chronic inflammatory response. Cytokines secreted by lymphocytes and macrophages exert both pro- and antiatherogenic effects. As part of a wound-healing response smooth muscle cells migrate from the medial portion of the arterial wall, proliferate, and secrete extracellular matrix proteins that form a fibrous cap. (From Glass, C.K., Witztum, J.L. (2001) Atherosclerosis. The road ahead, *Cell* **104**(4), 503–516, included with permission of the Publisher.)

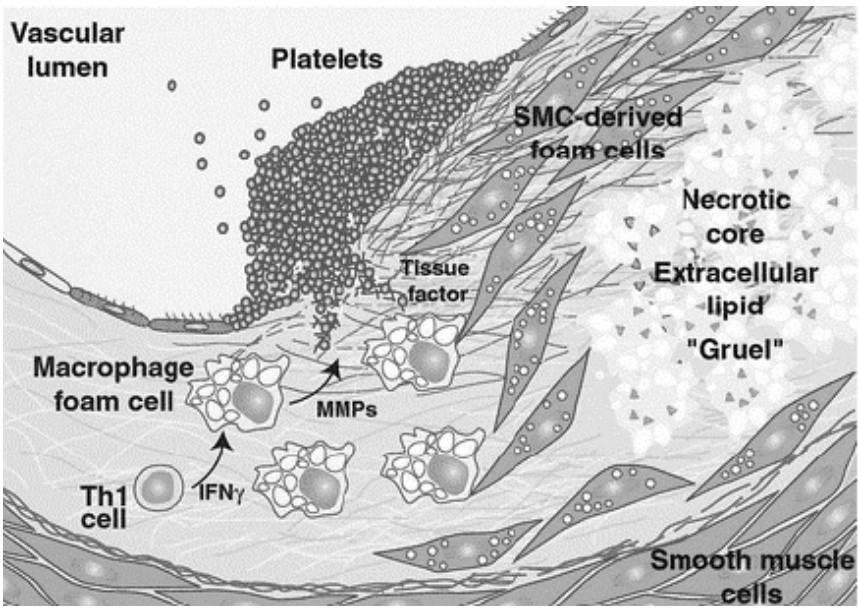

Fig. 3 Plaque rupture and thrombosis. The death of macrophage-derived foam cells and coalescence of trapped lipoproteins leads to the formation of an acellular necrotic zone and accumulation of extracellular cholesterol. Macrophage secretion of matrix metalloproteinases and neovascularization contribute to weakening of the fibrous cap. Plaque rupture exposes blood components to tissue factor, initiating coagulation, the recruitment of platelets, and the formation of a thrombus. (From Glass, CK and Witztum JL, Cell 2001; **104**(4): 503–516, included with permission of the Publisher.)

is associated with critical limb ischemia and gangrene.

3 Gene Therapy Targets for Cardiovascular Disease in the Lipoprotein Pathways

3.1 The Lipoprotein Pathways Provide Many Potential Gene Therapy Targets

A major risk factor for premature atherosclerosis is abnormalities in lipoprotein metabolism. In addition, people with diabetes mellitus frequently have lipoprotein abnormalities that elevate the risk of premature atherosclerosis that is already associated with diabetes.

Lipoprotein metabolism involves the interplay of the intestine, the liver, adipose tissue, and muscle. Endocrine organs also play a role, chiefly pancreatic islets, adrenals, the hypothalamus, and adipose tissue. However, the following discussion will emphasize the role of the liver because it is the most readily accessible target for gene therapy that is based upon the intravenous injection of DNA or DNA within a vector or carrier.

Animals transport large quantities of lipids through the bloodstream on carrier particles called lipoproteins. Lipoproteins consist of a shell of proteins called *apolipoproteins*, amphipathic lipids, primarily phospholipid and unesterified cholesterol, surrounding a core of

Tab. 1 Density, diameter, and composition of plasma lipoproteins.

Class	*Density [g mL^{-1}]*	*Diameter [nm]*	*Chol*	*PL*	*Protein*	*TG*	*CE*
Chylomicrons	0.93	75–1200	2	7	2	86	3
VLDL	0.93–1.006	30–80	7	18	8	55	12
IDL	1.006–1.019	25–35	9	19	19	23	29
LDL	1.019–1.063	18–25	8	22	22	6	42
HDL2	1.063–1.125	9–12	5	33	40	5	17
HDL3	1.125–1.210	5–9	4	35	55	3	13

Notes: Values represent percentage of dry mass. Chol = cholesterol; PL = phospholipids; TG = triacylglycerols; CE = cholesterol esters.

nonpolar lipid. The core composition consists of variable proportions of triacylglycerol and cholesterol ester.

When plasma is subjected to ultracentrifugation, most proteins sediment at the bottom of the centrifuge tube. Because they are associated with lipid and have a low buoyant density, lipoproteins float in the ultracentrifuge. The density at which they float is the basis for their classification (Table 1). In fasting plasma, most of the triacylglycerol is carried by very low-density lipoprotein (VLDL) particles. Cholesterol and cholesterol ester are carried by low-density lipoprotein (LDL) and high-density lipoprotein (HDL) particles, with LDL carrying about two-thirds of all cholesterol in human plasma.

3.2 Specific Genes are Critical for Dietary Lipid Absorption in the Intestine

It is important to remember that ordinarily, the major component of dietary fat is always triacylglycerol, not cholesterol. For example, milk and butter have very little cholesterol but are very high in triacylglycerol. Triacylglycerol (and other glycerolipids, such as phospholipids) is hydrolyzed in the intestinal lumen to yield monoglycerides and free fatty acids. These lipolysis products are then absorbed by the intestinal epithelial cells and resynthesized as triacylglycerols, phospholipids, and cholesterol esters. Thus, the intestine mediates both lipolysis (in the lumen) and reesterification (within the epithelial cells) of dietary lipids. The molecular mechanisms underlying intestinal absorption of fatty acids are still not understood. There is no consensus about whether a specific transporter is involved in fatty acid transport into the enterocytes.

Cholesterol absorption into the enterocyte is also not well understood. A specific inhibitor of cholesterol absorption, Ezetimibe, is in clinical use and is one line of evidence that there is a specific transporter for cholesterol at the membrane of the enterocyte that faces the intestinal lumen (the apical membrane).

The free fatty acids coming from intestinal lipolysis are solubilized into micelles within the intestinal lumen. These micelles consist of bile acids, cholesterol, and phosphatidylcholine. Bile acids are detergent-like molecules produced from cholesterol by the liver. Together with phospholipid and cholesterol, bile acids are secreted from the apical membrane of hepatocytes into the bile canaliculi, which drain into the bile duct and finally into the duodenum. As discussed below,

conversion of cholesterol to bile acids and secretion of biliary cholesterol are the principal routes of cholesterol elimination.

Two membrane cholesterol transporters have been identified in the intestine. One consists of a heterodimer of two proteins, ABCG5, and ABCG8. This pair is located at the apical membrane and transports cholesterol out of the enterocyte into the lumen. In addition, the transporter transports plant sterols and thus protects animals from accumulation of these nonphysiological sterol molecules. Mutations in ABCG5 or ABCG8 cause a rare disease called sitosterolemia, which involves elevated levels of cholesterol and plant sterols in the bloodstream. ABCA1 also transports cholesterol out of enterocytes. In contrast to ABCG5/ABCG8, this transporter appears to transport cholesterol across the basolateral membrane of the enterocyte, to the bloodstream and/or lymphatics. Overexpression of ABCG5/ABCG8 in transgenic mice reduces cholesterol absorption, making this an attractive target for therapeutic intervention.

3.3 The Intestine Exports Lipids on Chylomicron Particles

Since the intestine is primarily an absorptive organ, it must have the means of exporting newly absorbed lipids. The enterocyte reesterifies fatty acids and monoglycerides to form triacylglycerols and phospholipids. Absorbed cholesterol is esterified to form cholesterol esters. Triacylglycerols and cholesterol esters are then packaged into the core of lipoprotein particles unique to the intestine – *chylomicrons*. Rather than being secreted into the bloodstream, chylomicrons are secreted into the lymphatics. By secreting chylomicrons into the lymphatics, they gain entrance into the general circulation via the thoracic duct. This guarantees that extrahepatic tissues, principally adipose tissue and muscle, are the first to be exposed to the newly secreted chylomicrons – if chylomicrons were secreted directly into the bloodstream, they would first be delivered to the liver, via the portal vein.

After entry into the bloodstream, chylomicrons interact with the luminal surface of the capillary beds of adipose tissue and muscle. Here resides the enzyme *lipoprotein lipase*, which hydrolyzes the triacylglycerols all the way to free fatty acids and glycerol. (Cholesterol esters are not substrates for lipoprotein lipase.) The free fatty acids are then absorbed where they are reesterified and stored in triacylglycerol droplets (in adipocytes) or oxidized for energy (in muscle). The chylomicron depleted of its triacylglycerol (a chylomicron remnant) is rapidly cleared from the circulation by the liver. There is significant variation in the abundance and activation state of lipoprotein lipase. This affects the efficiency of clearance of triacylglycerol-rich lipoproteins and is therefore an attractive target for gene therapy.

3.4 The Liver is a Major Producer of Plasma Lipoproteins

Adipose tissue and liver are quite active in the conversion of carbohydrate into fat. Adipose tissue stores the fat in the form of triacylglycerol droplets. However, these droplets are quite dynamic; they are continuously being formed and turned over. The turnover of adipocyte droplets occurs through the action of hormone-sensitive lipase. The free fatty acids that are released from adipocytes can go to the liver where they can be used to make ketone bodies and/or are

reesterified to form triacylglycerol again. Whether derived from adipose tissue or from *de novo* synthesis, the liver does not generally store triacylglycerol. Triacylglycerol accumulation in the liver is generally regarded as a pathological state, termed *fatty liver* or *hepatic steatosis*.

3.5 Lipoprotein Lipase as a Target for Gene Therapy

In a healthy liver, triacylglycerol is packaged for secretion within the secretory pathway. This involves transfer of triacylglycerol to the endoplasmic reticulum lumen and interaction with apoB100 (Fig. 4). Within the secretory pathway, there are several steps in which triacylglycerol is incorporated into a growing VLDL particle. VLDL is then secreted into the bloodstream. Like chylomicrons, VLDL interacts with *lipoprotein lipase* while in the circulation. However, unlike chylomicrons, the depletion of triacylglycerol from VLDL yields a VLDL remnant, also termed *intermediate density lipoprotein* (IDL), which goes on to produce a stable lipoprotein,

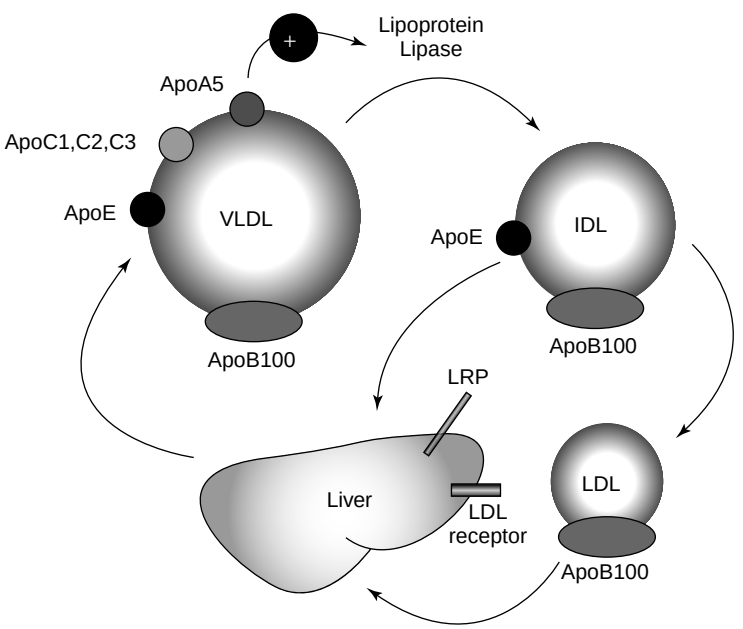

Fig. 4 Metabolic interrelationships of liver-derived plasma lipoproteins. The liver secretes VLDL, a particle enriched in triacylglycerol. It carries various apolipoproteins. ApoB100 is the only nonexchangeable protein on VLDL particles. ApoC2 and possibly, apoA5, are activators of lipoprotein lipase. This enzyme resides on the luminal surface of the capillary endothelium of adipose tissue and muscle and hydrolyzes the fatty acids from VLDL triacylglycerols. The resulting particle, IDL, can be cleared from the circulation by the liver or can become LDL. Clearance of IDL is mediated by apoE. Clearance of LDL is mediated by apoB100. The LDL receptor plays a critical role in the clearance of IDL and LDL; both apoE and apoB100 can bind to the LDL receptor. The LRP acts as a backup for IDL clearance; in the absence of the LDL receptor, it can clear IDL by binding to apoE.

LDL. Owing to the loss of most of the triacylglycerol, the predominant neutral lipids of LDL are cholesterol and cholesterol ester.

The catabolism of VLDL is dependent upon efficient hydrolysis of its triacylglycerol core by lipoprotein lipase. Lipoprotein lipase deficiency results in dramatic hypertriglyceridemia, liver enlargement (hepatomegaly), and inflammation of the pancreas (pancreatitis). However, elevated plasma VLDL can result from many other as-yet unidentified causes. It is very common in the human population and is associated with obesity, diabetes, and in some individuals, high-carbohydrate diets. Although these individuals have normal levels of lipoprotein lipase, increasing the level of this enzyme might still be a useful therapeutic strategy to lower plasma triacylglycerols. Since it is presently not feasible to target genes to the normal sites of expression of lipoprotein lipase (muscle and adipose tissue), expression in a nonphysiological site, the liver, has been tested in mice and in a naturally occurring feline model of lipoprotein lipase deficiency.

In addition to the expected role of lipoprotein lipase in catalyzing the hydrolysis of lipoprotein triacylglycerol, there is also a nonenzymatic property of potential therapeutic value. Lipoprotein lipase binds to lipoproteins and also binds to lipoprotein receptors in the LDL receptor family, most notably, the LDL receptor-related protein (LRP). In addition, lipoprotein lipase can bind to cell surface proteins that are enriched in acidic carbohydrate residues, the proteoglycans. The best evidence that these nonenzymatic functions are therapeutically relevant is the demonstration that a mutant form of lipoprotein lipase lacking enzymatic activity is still able to reduce plasma triacylglycerol by about 20 to 30%.

Lipoprotein lipase-mediated hydrolysis of triacylglycerol causes fatty acids to be taken up at the site of hydrolysis. In the liver, this leads to reesterification of the fatty acids and can result in an increase in cellular triacylglycerols. Through mechanisms not clearly understood, excessive triacylglycerol accumulation dampens the cell's responsiveness to insulin. In the liver, insulin normally suppresses glucose production. Thus, one potential side effect of expression of lipoprotein lipase in the liver is to reduce the ability of insulin to suppress glucose production, a potential problem for people who are already insulin-resistant or diabetic.

3.6 The LDL and VLDL Receptors are Attractive Targets for Gene Therapy

LDL is cleared from the circulation in large part through its interaction with the *LDL receptor*. Mutations affecting the expression or function of the LDL receptor are responsible for a common inherited disorder, *familial hypercholesterolemia*. Increased expression of the LDL receptor reduces LDL levels; thus, the LDL receptor has been used in gene therapy for hypercholesterolemia.

The VLDL receptor is a member of the LDL receptor family. It is expressed in a wide range of tissues, but not in the liver. However, adeno-associated virus was used to express this receptor in the livers of mice lacking the LDL receptor. These mice experienced a 40% reduction of cholesterol, due to enhanced clearance of both VLDL and LDL. Since patients who do not express the LDL receptor are immunologically naïve to this antigen, gene therapy with the VLDL receptor might be an effective treatment that would avoid the danger of immune rejection.

3.7 Increased Cholesterol and Phospholipid Efflux via ABCA1 Can Prevent Atherosclerosis

Epidemiological studies show an inverse relationship between HDL levels and risk of premature atherosclerosis. Unlike VLDL and chylomicrons, HDL is formed from its protein and lipid components in the bloodstream and interstitial fluids rather than within the secretory pathway of cells. The major apolipoproteins of HDL are apoA1 and apoA2. These proteins are secreted from hepatocytes and intestinal epithelial cells independently and also as minor components of VLDL and chylomicrons. ApoA1 and apoA2 bind to phospholipids. Phospholipids are available from the surface of VLDL after lipolysis. In addition, cells are able to efflux phospholipids and cholesterol through the action of ABCA1 (Fig. 5). Its crucial role in this process was established by the discovery that two types of severe inherited HDL deficiency syndromes are caused by mutations in ABCA1, Tangier Disease, and familial hypoalphalipoproteinemia (FHA).

Tangier Disease is a rare recessive disorder in which patients have almost no HDL. Cholesterol ester accumulates in macrophages and macrophage-rich tissues like spleen and liver. FHA is a very common dominant disorder in which people have low HDL (typically <30 mg dL^{-1}) and suffer from premature heart disease even without an elevation in LDL. Approximately 40% of patients with premature coronary heart disease have low HDL, making it the most common lipid disorder of heart disease patients.

Tangier Disease and FHA are caused by mutations in ABCA1. Tangier Disease patients are homozygous (or inherit two different mutant alleles). FHA patients are heterozygous for mutations at the ABCA1 locus. ABCA1 mutations lower HDL because they prevent phospholipid and/or cholesterol from effluxing and becoming associated with apoA1 and apoA2. In the absence of sufficient lipid, apoA1 is rapidly cleared by the kidneys.

Why do mutations in ABCA1 lead to premature heart disease? ABCA1 fulfills a rate-limiting step in the pathway by which cells get rid of cholesterol, and thus might be a critical protector against cholesterol overload. With the exception of hepatocytes, cells are unable to catabolize large quantities of cholesterol and must therefore protect themselves from cholesterol overload by expelling cholesterol to an appropriate extracellular carrier. It is interesting that ABCA1 is especially abundant in macrophages; macrophages can become engorged with cholesterol esters and form foam cells in the arterial wall. A mutation in ABCA1 might therefore predispose an individual to atherosclerosis by impeding cholesterol efflux (Fig. 1).

Proof-of-principle studies in mice nominate ABCA1 for the gene therapist's arsenal. Moderate overexpression of ABCA1 decreases atherosclerosis in mice, despite having a minimal effect on HDL levels. The most plausible explanation is that macrophage ABCA1 function is more important than HDL levels in protecting an animal from atherosclerosis. Indeed, if animals are transplanted with macrophages derived from mice lacking ABCA1, they show increased atherosclerosis. These animal studies suggest that increased expression of ABCA1 in macrophages might be a sensible gene therapy strategy for protection against or treatment of atherosclerosis.

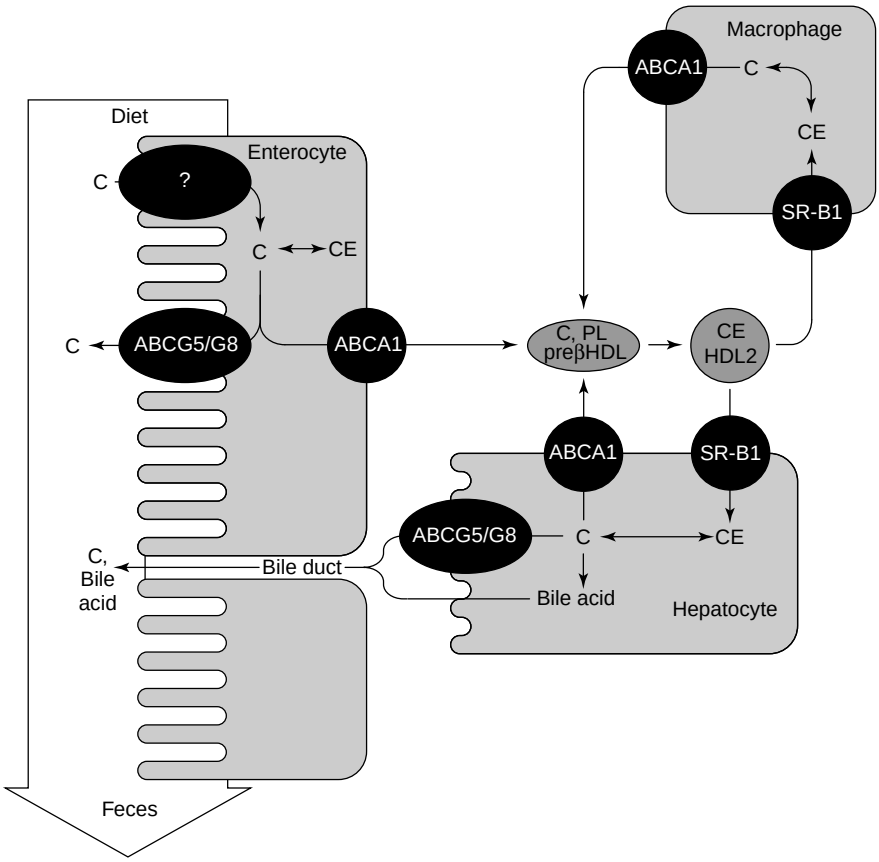

Fig. 5 ABC transporters involved in cholesterol transport. ABCG5 and ABCG8 function together as a heterodimer and mediate the transport of cholesterol from hepatocytes across the canalicular (apical) membrane into the bile. The ABCG5/G8 complex is also involved in cholesterol and plant sterol efflux from the enterocytes of the intestine to the intestinal lumen. This reduces the net absorption of cholesterol and prevents absorption of plant sterols. ABCA1 transports cholesterol and phospholipids on the sinusoidal (basolateral) surface of hepatocytes and enterocytes. In addition, ABCA1 is abundant in macrophages where it mediates cholesterol and phospholipid efflux. ApoA1 interacts with ABCA1 and mediates the formation of precursors to HDL (preβHDL). The esterification of the cholesterol in the preβHDL by the blood-borne enzyme, lecithin cholesterol acyl transferase (LCAT), produces mature HDL. Mature HDL can bind to the SR-B1 receptor and deliver cholesterol esters to hepatocytes. Much of the cholesterol ester is thus delivered and hydrolyzed by a cellular cholesterol esterase and secreted into bile as unesterified cholesterol.

3.8 Some Apolipoproteins Can Prevent Atherosclerosis

ApoA1 is a major protein constituent of HDL. It is an activator of an enzyme in the bloodstream, LCAT. This enzyme transfers a fatty acid from the *sn*-2 position of phosphatidylcholine to cholesterol to form a cholesterol ester. It enables HDL to "accept" cholesterol from cells or other lipoproteins and then to fill its

core with the cholesterol by esterifying it. ApoA1 also interacts with ABCA1 to mediate cholesterol and phospholipid efflux from cells.

Overexpression of apoA1 in transgenic mice or injection of large quantities of the apoA1 protein produces dramatic effects on atherosclerosis – both prevention and regression of atherosclerotic lesions. ApoA1 overexpression is a very attractive prospect for gene therapy. The biggest difficulty is that apoA1 levels in human plasma are already quite high, $\sim$1 mg mL^{-1}. Thus, with present technology, it would be quite difficult to achieve a significant increase in this steady state level.

ApoA5 is a newly discovered apolipoprotein that is associated with VLDL. Its deletion in mice results in elevated triacylglycerols and its overexpression in transgenic mice drives down triacylglycerol levels. It is likely that it plays a role in the ability of lipoprotein lipase to hydrolyze VLDL triacylglycerol. ApoA5 might be an attractive candidate for gene therapy of hypertriglyceridemia.

ApoE is a ligand for the LDL receptor and virtually all other members of the LDL receptor family. It is associated with VLDL and chylomicrons and to a varying extent, with HDL. Chylomicrons depend on apoE for their clearance from the circulation. In addition, although VLDL does have another competent LDL receptor ligand, apoB100, VLDL remnants are inefficiently cleared in the absence of apoE. There are three common allelic variants of apoE – apoE2, apoE3, and apoE4. ApoE2 has a greatly reduced affinity for its receptors and is therefore associated with impaired clearance of VLDL and chylomicron remnants. From these properties, one might predict that increasing apoE levels might be therapeutically desirable by promoting efficient removal from the circulation of remnant lipoproteins. However, studies in transgenic mice and rabbits revealed two unexpected consequences of apoE overexpression. First, increased apoE synthesis in the liver leads to increased VLDL production. Second, increased apoE in the VLDL particle slows the hydrolysis of triacylglycerol by lipoprotein lipase.

3.9 The Liver SR-B1 Receptor Protects against Atherosclerosis

The scavenger receptor-B1 (SR-B1) is highly expressed in tissues that actively convert cholesterol into steroid hormones; for example, adrenals, ovaries, and testis. It is also highly expressed in hepatocytes. SR-B1 binds to HDL and mediates the selective uptake of cholesterol ester from the HDL particle into the cell. Unlike the uptake of LDL particles, the particle itself is spared and can then recycle back to pick up another cargo of cholesterol. The overexpression of SR-B1 in transgenic mice causes a drastic drop in HDL levels and protects the animals from atherosclerosis. Conversely, disruption of the gene raises HDL and increases the susceptibility of animals to atherosclerosis. Interestingly, increased SR-B1 expression increases the secretion of cholesterol into the bile, indicating that uptake of cholesterol from HDL effectively targets the HDL toward biliary secretion.

The foregoing might seem to contradict the known inverse relationship between HDL levels and atherosclerosis. However, increased HDL can be due to increased HDL production or decreased HDL turnover. Increased HDL production is thought to be beneficial because it promotes cholesterol flow back to the liver. However, increased HDL, if it is due to an

impairment in its ability to deliver cholesterol to the liver, is not beneficial. The transgenic experiments in mice suggest that increased SR-B1 expression in the liver might be a viable approach to gene therapy for atherosclerosis.

4 Vascular and Inflammatory Targets for Gene Therapy

Small clinical trials of gene therapy for cardiovascular disease have been limited to induction of *angiogenesis* in the heart and legs for the treatment of *angina* and critical limb *ischemia* or to blocking of the proliferative response following injury induced by *angioplasty* and vascular grafting as a means of preventing *restenosis*. These trials have involved use of adenoviruses and DNA-containing liposomes for localized overexpression of angiogenic growth factors such as vascular endothelial growth factor (VEGF), fibroblast growth factor (FGF), hepatocyte growth factor (HGF) and antisense oligonucleotides or decoy oligonucleotides for inhibiting expression of cell cycle genes such as *c-Myc* or for preventing binding of transcription factors such as E2F to cis-acting response elements respectively. The localized overexpression of VEGF and other growth factors in the heart and limbs have had mixed results with some recorded successes in formation of collateral vessels, increased blood flow, and reduced pain. Large-scale clinical trials in the United States have been slowed because of the problems with virus-related inflammation and with VEGF-induced edema.

As discussed in the first section of this chapter, genes involved in lipid and lipoprotein metabolism are potential targets for reducing hyperlipidemia and the consequent deposition and retention of lipids and lipoproteins in the artery wall. However, there are still a considerable number of individuals who develop cardiovascular disease who do not have appreciably elevated lipids or are resistant to current therapies. Thus, other targets need to be considered. Genes involved in regulating leukocyte function and the inflammatory response within the artery wall are potential candidates.

4.1 Adhesion Molecules and Leukocyte Counter Receptors

The *adhesion molecules*, vascular cell adhesion molecule (VCAM-1) and intercellular adhesion molecule (ICAM-1), are members of the immunoglobulin superfamily of proteins. Together with their leukocyte counter receptors, the integrins VLA4 and MAC-1/CD11b:CD18, are potential targets for gene therapy. VCAM-1 and ICAM-1 are expressed by endothelial cells and support the tight binding of monocytes and lymphocytes to the endothelial cell plasma membrane. The selectins (E-, P-, and L-selectins) are also adhesion molecules that play a role in the initial tethering and rolling of leukocytes along the surface of the blood vessel. They are lectin-like molecules that recognize carbohydrate containing counterreceptors.

Knockouts of ICAM-1, E-, and P-selectins in mice all partially inhibit development of atherosclerosis. Thus, localized inhibition of expression of these genes might prevent initiation of lesions and limit progression of established lesions due to layering and the continuous influx of leukocytes. Adenovirus-mediated gene transfer to endothelial cells has been successful in both *in vitro* and *in vivo* studies and suggests that this approach could

be used for inhibiting expression of the adhesion molecules.

The transient nature of adenoviral-mediated gene transfer is a cause for concern. Given that the atherosclerotic lesions develop over several decades, effective adenoviral therapy would require repeated injections. The use of retroviruses or adeno-associated viruses to incorporate the gene into the cellular genome might be a better alternative. However, this approach is also limited by the small percentage of endothelial cells that proliferate under normal circumstances. Blocking expression of the leukocyte integrins is even more problematic. Currently, this would require transfection of bone marrow stem cells and would likely compromise the capacity to mount an inflammatory response at other sites. Overexpression of secreted forms of the adhesion molecules may be an effective alternate approach.

4.2 Cytokines

Cytokines are cell regulatory proteins that are key factors in the initiation and control of inflammation. They are classified as members of several different families of proteins that include: the tumor necrosis factor (TNF) family, interleukins, interferons, colony-stimulating factors, growth factors, and chemokines. There are significant overlaps in the functions of many of the cytokines. This overlap is in part due to the convergence of signal transduction pathways from the different cytokine specific receptors leading to activation of the same transcription factors. However, in other cases cytokines can have opposite effects and are frequently designated as being either proinflammatory or anti-inflammatory. Serum cytokine levels are elevated in people with established cardiovascular disease. Virtually, every major pro- and anti-inflammatory cytokine is expressed by cells within atherosclerotic lesions and all vascular cell types are capable of responding to these proteins. Thus, there is significant potential for the inhibition and treatment of CVD by manipulating the production of or response to cytokines using gene transfer approaches.

4.3 Chemokines and Chemokine Receptors

Monocyte chemotactic protein (MCP-1) and IL-8 are two potent leukocyte chemoattractants (*chemokines*) that play a role in recruiting leukocytes into the sites of inflammation. MCP-1 is expressed by all of the cell types within atherosclerotic lesions. Recent data from studies on atherosclerosis-prone mice with targeted deletion of MCP-1 or the chemokine receptors CCR-2 or CXCR-2, and studies of bone marrow transplantation of cells overexpressing MCP-1 or devoid of IL-8, strongly suggest that these molecules play a critical role in the initiation of fatty streaks.

Recent gene therapy studies with an N-terminal deletion mutant of the human MCP-1 gene transfected into the skeletal muscle of the atherosclerosis-prone apoE knockout mice have demonstrated the potential for inhibiting atherosclerosis by blocking MCP-1 activity. However, as with the adhesion molecules, there are caveats with regard to targeting the MCP-1, IL-8, and/or CCR-2 and CXCR-2 genes for CVD gene therapy. Because MCP-1 is expressed by all of the cells within the lesions and IL-8 primarily by macrophages, it is unclear whether adenoviral or retroviral approaches would be successful in inhibiting MCP-1 or IL-8 expression by macrophages.

Primary macrophages are extremely difficult to transfect, therefore, blocking their capacity for expressing MCP-1, IL-8, or the receptors likely would require repeatedly knocking out the genes in bone marrow cells. This again raises the question, to what extent would this compromise the inflammatory response at other sites. Thus, localized overexpression of mutated forms of these proteins such as the N-terminal deletion mutant sited above is the most logical approach to targeting chemokines for prevention and treatment of cardiovascular disease.

4.4 Interleukins, the TNF-α family, and Interferon-γ

Currently, *in vitro* and limited *in vivo* data suggest that IL-1, IL-2, TNF-α, IL-3, IL-6, IL-12, IL-15, and IL-18 appear to have proatherogenic properties, while the IL-1 receptor antagonist (IL-1ra), IL-9, IL-10, IL-11 and possibly IL-4, and IL-13 are likely to be antiatherogenic. Thus, localized inhibition of proatherogenic cytokines and overexpression of antiatherogenic cytokines with gene therapy has significant potential for prevention and treatment of cardiovascular disease.

TNF-α and IL-1-β are the proinflammatory cytokines that have received the most attention. *In vitro* studies have clearly demonstrated that these two cytokines potently stimulate the expression of adhesion molecules and chemokines and suggest that they play an important role in the atherogenic process. Surprisingly, targeted deletion of TNF-α or the p55 TNF-α receptor is not protective of atherosclerosis in mouse models. Furthermore, the chronic administration of a TNF-α binding protein reduces fatty streak formation only in female apoE knockout mice. The effects of deletion of IL-1 on atherosclerosis have not been reported. However, in knockouts of the IL-1ra, there was a trend toward increased foam-cell lesion area compared with the wild-type littermate controls. Successful gene therapy studies in animal models of cerebral, pancreatic, and articular inflammation with IL-1ra further support the potential for this approach to cardiovascular disease.

CD40 and CD40 ligand are members of the TNF receptor and protein families respectively. They are expressed by monocytes, macrophages, dendritic cells, lymphocytes, endothelial cells, and smooth muscle cells and play a role in many inflammatory processes including the expression of adhesion molecules, cytokines, chemokines, and matrix metalloproteinases. Recent data suggests that CD40/CD40 ligand play fundamental roles in the atherogenic process as CD40 ligand–apoE double knockout mice have significantly reduced atherosclerosis and lesions that contain reduced numbers of inflammatory cells, are connective tissue rich, and are likely resistant to plaque rupture. Thus, CD40 or CD40 ligand are both additional potential targets for localized CVD gene therapy.

Interferon gamma (IFN-γ) is an immuno-stimulatory cytokine that increases antigen presentation by macrophages and the activation of T lymphocytes. As macrophages and T lymphocytes are key cellular components of the atherosclerotic lesions at all stages of lesion development, gene therapy that targets IFN-γ expression could have a significant impact on the disease process. This is supported by several studies in mice that demonstrate a critical role for this cytokine. For example, atherosclerotic lesions are significantly reduced in IFN-γ-deficient mice crossed with either LDL-receptor

knockout or apoE knockout mice and are increased with chronic administration of recombinant IFN-γ. Interleukin-18 (IL-18) in part promotes inflammatory responses through the release of IFN-γ. Administration of recombinant IL-18 to apoE knockout mice, like the administration of IFN-γ, significantly increases lesion size. Not surprisingly, in IFN-γ/apoE double knockout mice, there are no effects of IL-18 on lesion development.

IL-2 is another interleukin with probable proatherogenic properties. It also functions to alter T-lymphocytes by affecting a shift to a Th1 T helper cell phenotype. It is expressed within atherosclerotic lesions and serum levels are increased in people with ischemic heart disease and unstable angina pectoris. However, to date, IL-2 has not been shown to play a direct role in the development of lesions in animal models. In contrast to IL-2, IL-4 and IL-10 in part function to cause a shift toward an antiatherogenic Th2, T helper cell phenotype. There is reduced atherosclerosis in LDL-receptor knockout mice transplanted with bone marrow from IL-4 deficient mice but no increase in atherosclerosis in cholesterol-fed, IL-4 deficient mice. On the other hand, fatty streak formation in IL-4-deficient mice immunized with HSP65 or Mycobacterium tuberculosis is significantly reduced when compared with lesions in wild-type C57BL/6 mice.

In addition to causing a Th2 shift, IL-10 inhibits differentiation of monocytes to macrophages, macrophage presentation of antigens, generation of reactive oxygen and nitrogen species, and endothelial cell expression of ICAM-1. There is accelerated formation of atherosclerotic lesions in IL-10 knockout mice and reduced formation of lesions in IL-10 transgenic mice. Thus, localized overexpression of IL-4 and IL-10 could provide some protection against cardiovascular disease. There is also recent evidence that IL-6 may play a role in the development of atherosclerosis. Treatment of fat-fed C57BL/6 or apoE knockout mice with recombinant IL-6 significantly increases fatty streak formation.

4.5 Growth Factors and Receptors

Growth factors are potent regulators of cellular functions including proliferation, migration, differentiation, and survival/apoptosis. Many growth factors and growth factor receptors are expressed or deposited in atherosclerotic lesions. These include all forms of platelet-derived growth factor (PDGF), basic fibroblast growth factor (bFGF), vascular endothelial growth factor (VEGF), insulin-like growth factors (IGF-1 and IGF-2), thrombin, endothelin-1, angiotensin-II, heparin binding-epidermal growth factor (Hb-EGF), several forms of transforming growth factor-beta (TGF-β), and the hematopoietic growth factors: macrophage colony-stimulating factor (M-CSF), granulocyte colony–stimulating factor (G-CSF), and granulocyte-macrophage colony-stimulating factor (GM-CSF).

The original "Response to Injury" hypothesis of Glomset and Ross posited that denuding injury to the endothelium led to the activation and attachment of platelets with release of PDGF. This was thought to lead to increased smooth muscle cell migration and proliferation and development of atherosclerotic lesions. Over the past three decades, the paradigm has shifted to account for the crucial role of inflammatory cells and the potential protective role of smooth muscle migration and proliferation in the formation of the fibrous cap (Figs. 1–3). Thus, for gene therapy approaches to CVD, localized overexpression

of growth factor genes could help stabilize lesions at risk of rupture and thrombosis. As proof of principle, transplant of fetal liver cells from mice deficient in PDGF-BB or chronic treatment with a PDGF-receptor antagonist or anti-PDGF receptor antibodies in apoE knockout mice leads to the formation of lesions that contain mostly macrophages, appear less mature, and have a reduced frequency of fibrous cap formation as compared with control mice. In contrast to primary atherosclerotic lesions, smooth muscle migration and proliferation contributes to *restenosis* following angioplasty and other interventions. Gene therapy approaches to inhibiting growth factor expression postinjury have been in development for several years.

Stimulating or inhibiting the expression of individual growth factors alone may not be sufficient to control vascular smooth muscle cell migration and proliferation. This is because both smooth muscle cell migration and proliferation are exceedingly complex processes and involve dissociation from the extracellular matrix through the activation and secretion of proteases, engagement of multiple growth factor receptors with activation of a variety of receptor associated and cytoplasmic kinases, activation of transcription factors, and production and activation of proteins involved in cell cycle traverse and DNA synthesis such as cyclins and cyclin-dependent kinases. Thus, there are potentially many targets other than growth factors or their specific receptors for gene therapy. This has been demonstrated by the ongoing clinical trials previously cited where antisense oligonucleotides or decoy oligonucleotides for inhibiting expression of cell cycle genes such as *c-Myc* or for preventing binding of transcription factors such as E2F are being tested as the means of preventing restenosis following angioplasty.

The hematopoietic growth factors M-CSF and GM-CSF play a role in the differentiation of monocytes and possibly in regulating macrophage proliferation and death within atherosclerotic lesions. Thus, these factors may also be good targets for gene therapy for both early and unstable lesions. This has been supported by data showing that M-CSF deficient mice crossed with apoE knockout mice have significantly reduced atherosclerosis. Furthermore, chronic administration of an antibody that blocks the M-CSF receptor (c-fms) also reduces early lesions in apoE knockout mice.

Angiogenic growth factors such as the VEGFs, HGF, and the angiopoietins have already been mentioned with regard to ongoing clinical trials for the induction of collateral circulation in people with occluded coronary and peripheral arteries. However, small *neovessels* also form in advanced atherosclerotic lesions and are additional conduits for recruitment of leukocytes into the established lesions. Blocking *neovascularization* could help reduce inflammation and stabilize the lesions. Thus, localized inhibition of expression of these factors or their respective receptors (FLT and FLK for VEGFs and Tie receptors for angiopoietins) are potential additional CVD targets.

5
Pro- and Antioxidant Enzymes

There is now ample evidence that polyunsaturated fatty acids, phospholipids, and cholesteryl-esters within lipoproteins that become trapped within the extracellular matrix of the artery wall can become oxidized. *In vitro* studies have demonstrated that macrophage scavenger receptors

(MSR) recognize oxidized lipoproteins and that foam cell formation is due in part to the accumulation of oxidized lipoproteins. The uptake of oxidized lipids can lead to intense oxidative stress within the foam cells and likely contributes to cell death.

It is currently unclear exactly how lipoproteins are oxidized within the artery wall. Several lines of evidence indicate that proinflammatory enzymes such as myeloperoxidase and the lipoxygenases can oxidize lipoproteins. Macrophages also express the bacteriocidal enzymes NADPH oxidase and the inducible form of nitric oxide synthase that generate the superoxide anion and nitric oxide respectively. These reactive oxygen and nitrogen species can also contribute to lipoprotein modification and oxidative stress within foam cells. Thus, it is possible that gene therapy could be used to inhibit expression of these enzymes within atherosclerotic lesions and thus reduce formation of the necrotic zones and help stabilize lesions that are susceptible to plaque rupture and thrombosis. This is supported by studies showing that mice deficient in 5-lipoxygenase or 12/15 lipoxygenase are protected from atherosclerosis and that mice transgenic for 12/15 lipoxygenase have accelerated formation of lesions.

In contrast to the lipoxygenases, knockout of myeloperoxidase and the NADPH oxidase subunits do not protect mice from atherosclerosis. As with other essential proinflammatory factors, a deficiency of these enzymes can have a significant effect on the capacity to mount an inflammatory response as is seen in people with a deficiency of NADPH oxidase and the resulting chronic granulomatous disease. Thus, future gene therapy to inhibit the expression or activity of these enzymes will require viral targeting to leukocytes only within the artery wall.

All cell types have protective mechanisms against the buildup of pro-oxidants that could disrupt cellular functions by oxidizing lipids, proteins, and DNA. These protective mechanisms include formation of the primary endogenous antioxidant glutathione by the enzyme complex glutamate cysteine ligase. Glutathione acts by enzymatically and chemically converting electrophilic centers to thioether bonds and as a substrate in the glutathione peroxidase mediated destruction of hydroperoxides. Complete glutathione redox systems are located in both the mitochondria and cytoplasm. Reduction of oxidized glutathione by glutathione reductase utilizes reducing equivalents supplied by NADPH and consumes NADPH at a higher rate than most other NADPH-dependent enzymes. Under conditions of oxidative stress with active glutathione redox cycling (as would occur in foam cells containing increased lipid peroxides), the cell's supply of glutathione and NADPH can be depleted, thus limiting many important redox sensitive and biosynthetic reactions and can lead to cell injury and death. Cells also contain superoxide dismutases that convert the superoxide anion to hydrogen peroxide and the enzyme catalase that converts hydrogen peroxide to water. Thus, targeted overexpression of the enzymes involved in glutathione metabolism – the superoxide dismutases and/or catalase could help reduce oxidative stress and foam cell death.

6
Scavenger Receptors

There are now known to be at least three families of broad specificity *scavenger receptors* that are expressed primarily by

macrophages within human atherosclerotic lesions. These include the type A family (subtypes I–III and the *ma*crophage *r*eceptor with *c*ollagenous structure or MARCO), type B receptors that include CD36, SR-B1/CLA-1 and the splice variant SR-B2, and an additional group of receptors such as the mucin-like receptor macrosalin/CD68 and LOX-1. All of these receptors bind oxidized LDL, but CD36 now appears to be the major receptor on macrophages responsible for accumulation of oxidized LDL.

Recent studies of hyperlipidemic mice with targeted deletions of SRA I/II or CD36 have clearly established that both types of receptors play a fundamental role in foam cell formation and lesion development *in vivo*. Macrophages within human lesions also express a specific receptor that recognizes advanced glycosylation end-product (AGE) proteins, the LDL receptor, VLDL receptor, and the LDL-receptor related protein/α_2-macroglobulin receptor (LRP). Because all of these receptors are expressed primarily by macrophages, this again brings up the problem that primary macrophages are extremely difficult to transfect and blocking their capacity for expressing scavenger receptors would likely require repeatedly knocking out the genes in bone marrow cells. Like adhesion molecules, an alternate approach that should be effective is the overexpression of soluble decoy scavenger receptors that would bind and clear modified lipoproteins and reduce lipoprotein trapping in the artery wall and the subsequent formation of foam cells. This is supported by a recent study showing that there is reduced atherosclerosis in LDL-receptor deficient mice treated with an adenoviral construct containing the human MSR AI extracellular domains.

7 Transcription Factors

Decoy oligonucleotides have been used to block binding of activated transcription factors to their consensus DNA response elements in the promoters of many genes. This may be a viable approach to reducing expression of many cytokines and other proinflammatory factors *in vivo* because activation of transcription factors such NF-κB, AP-1, and the STAT proteins are end points common to signal transduction pathways activated by various proinflammatory stimuli. Blocking the activation or expression of NF-κB could be a particularly effective therapy as there is activated NF-κB in all of the cell types within atherosclerotic lesions. In addition to blocking activation of NF-κB with decoy oligonucleotides, stimulating the activity of the natural inhibitors of NF-κB, the I-κB proteins, or inhibiting the expression or activity of the I-κB kinases that phosphorylate the I-κB proteins leading to the proteosomal degradation of the NF-κB inhibitors, may also be viable approaches. Targeting the expression of the NF-κB subunits p65 and p50 is another possible approach. Other transcription factors such as Egr-1, Nrf-1, and Nrf-2 may also be potential targets for cardiovascular disease. Egr-1 is an essential factor for regulating the expression of tissue factor, a key procoagulant protein expressed by cells in atherosclerotic lesions. Nrf-1 and Nrf-2 are important factors for regulating the expression of pro- and antioxidant enzymes and bind to consensus antioxidant response elements of these genes.

8
Noninflammatory Gene Targets

There are a large number of proteins that are not directly part of the inflammatory response that are potential targets for CVD gene therapy. These include proteins that regulate cell death, cellular proteases that may play a role in breaking down the arterial extracellular matrix, proteins that play a role in regulating blood pressure and vascular tone, pro- and anticoagulant proteins, and osteogenic proteins that may contribute to vascular calcification.

8.1
Cell Death

As noted, the death of macrophages likely plays a fundamental role in the formation of the necrotic zones of advanced lesions. There is ample evidence showing the presence of macrophages with fragmented DNA (a marker of cell death) located within or adjacent to these necrotic zones. The death of smooth muscle cells also plays a role in the thinning of the fibrous cap, and like the macrophages, smooth muscle cells with fragmented DNA have been documented within atherosclerotic lesions from both humans and experimental animals. Thus, strategies designed to reduce cell death within atherosclerotic lesions could help stabilize the plaques and prevent plaque rupture and formation of occlusive thrombi.

Although it is currently unclear what induces the death of either cell type *in vivo*, there is *in vitro* evidence that apoptosis (programmed cell death or cellular suicide) can be induced in both macrophages and smooth muscle cells by oxidized lipids and other proinflammatory factors. Thus, proteins known to play a role in apoptotic death are potential targets for gene therapy. These include death-promoting proteins such as Fas (a death receptor that is a member of the TNF-α receptor family) and the Fas ligand, the cysteine proteases (caspases 3, 8, 9), and other death pathway effector proteins such as Fas-associated death domain (FADD) (Fas-associated protein with death domain), TNF receptor associated death domain (TRADD) (TNF receptor associated death domain protein), TRAF (TNF receptor associated factor), and receptor interacting protein (RID). There are also a number of antiapoptotic proteins such as c-FLIP (Fas-associated death domain (FADD) protein (FADD)-like interleukin-1 beta-converting enzyme [FLICE (caspase-8)]-inhibitory protein), BCL-2 and BCL-X that could provide protection against cell death if locally overexpressed.

8.2
Proteases

Secretion of proteolytic enzymes such as the matrix metalloproteinases (MMPs) by cells within the blood vessel plays an essential role in enabling the cells to migrate and proliferate. However, excess secretion of these enzymes by activated leukocytes or release of lysosomal enzymes such as cathepsins following the death of cells likely contributes to the breakdown of the extracellular matrix and to the fissure and rupture of the plaques. This is supported by data showing the presence of activated MMPs in unstable human atherosclerotic plaques. Thus, strategies designed to inhibit the expression of these proteolytic enzymes or the activation of the zymogen forms by membrane-associated enzymes such as the ADAM family of proteases could help prevent plaque destabilization and formation of occlusive thrombi. Cells within atherosclerotic lesions also

express specific tissue inhibitors of metalloproteinases (TIMPs) and thus localized overexpression of these inhibitors could also be an effective gene therapy approach for cardiovascular disease.

8.3 Regulators of Vascular Tone and Blood Pressure

Elevated blood pressure is a known risk factor for cardiovascular disease. It is well established that as atherosclerotic lesions progress, the affected blood vessel loses its capacity to adequately vasodilate. Controlling acute changes in vascular tone that cause *vasospasm* may be even more fundamental to CVD as evidence suggests that vasospasm may precipitate plaque rupture and the formation of occlusive thrombi. There are currently many drugs available for chronically controlling high blood pressure but very few for managing vasospasm. Thus, localized gene therapy that targets the production of vasoactive substances by arterial wall cells or the response of smooth muscle cells to these vasoactive substances leading to increased vasodilation and reduced vasospasm may be effective approaches for the treatment of cardiovascular disease.

Nitric oxide is an extremely potent vasodilator and most endothelial cells express an endothelium specific form of nitric oxide synthase (e-NOS). However, e-NOS expression and NO production are reduced in atherosclerotic arteries. Thus, a localized increase in the expression of e-NOS could help alleviate the impaired vasodilatory properties of atherosclerotic arteries. Endothelial cells also produce the vasodilatory prostaglandin, prostacyclin (PGI_2). PGI_2 is produced from arachidonic acid by the action of the cyclooxygenases followed by PGI_2 synthase. Thus, increased expression of PGI_2 synthase could be an effective approach for increasing PGI_2 synthesis. Endothelial cells also produce a variety of vasoconstrictors such as endothelin-1 and endothelial cell hyperpolarizing factors. Thus, inhibiting expression of the proform of endothelin or the endothelin-converting enzyme that activates endothelin could help control vasoconstriction. There are also endogenous inhibitors of endothelin activation such as the endothelin-converting enzyme inhibitor that could also be locally overexpressed and play a beneficial role in regulating blood pressure. Similarly, endothelial cells express a form of the angiotensin-converting enzyme (ACE). Angiotensin-II is a potent vasoconstrictor. It is produced following the conversion of angiotensinogen to angiotensin-I by renin and from angiotensin-I to angiotensin-II by ACE. Thus, inhibition of endothelial expression of ACE could also be an effective therapy. Finally, blocking the response of smooth muscle cells to these various vasoconstrictors or increasing the response to the vasodilators by targeting expression of the specific smooth muscle cell receptors or the down stream signal transduction pathways may also be effective gene therapy targets.

8.4 Coagulation Related Genes

There are a variety of pharmacological agents that effectively inhibit *thrombosis*, and chronic administration of these drugs is associated with reduced frequencies of angina and myocardial infarction in people with established cardiovascular disease. However, because we still do not know what causes atherosclerotic plaques to rupture and because plaque rupture is

unpredictable, administration of anticoagulants throughout the three to four decades that lesions develop may not be practical for most people due to the accompanying bleeding disorders. Thus, specifically blocking the expression of procoagulant proteins such as tissue factor by cells within atherosclerotic lesions, may be an effective approach to preventing localized formation of occlusive thrombi following plaque rupture. Tissue factor participates in the extrinsic pathway of coagulation and binds to and activates factor VIIa, which in turn activates factor X enabling it to convert prothrombin to thrombin and thrombin to convert fibrinogen to fibrin, the primary protein component of blood clots. Cells within atherosclerotic lesions also express a specific tissue factor inhibitor; thus an additional possibility would be to stimulate expression of this inhibitor. Another potential approach for reducing formation of occlusive thrombi is to increase the production of thrombolytic factors such as plasmin by increasing the localized expression of tissue plasminogen activator or reducing the expression of the plasminogen activator inhibitors.

8.5 Osteogenic Proteins

Bone mineral (calcium-phosphate) is deposited in most advanced atherosclerotic plaques and in heart valves. However, it is still controversial as to whether calcification of the plaques is a good prognostic indicator of subsequent CVD events. Nevertheless, preventing vascular calcification is beneficial for interventions such as angioplasty and for reducing the need for heart valve replacement. Vascular calcification is now known to be an active cellular mediated process that is analogous to the process by which cartilage is converted to bone. It thus involves a balance between bone forming osteoblast type cells and bone removing osteoclast type cells and the expression of proteins that are both pro-osteogenic and antiosteogenic. These include the matrix gla proteins, osteopontin, osteoprotegrin, osteonectin, and the bone morphogenic proteins. Thus, localized gene therapy designed to either increase or inhibit expression of these proteins could have dramatic effects on reducing plaque calcification and complications resulting from "stiffening" of the blood vessels.

See also Medicinal Chemistry.

Bibliography

Books and Reviews

Freedman, S.B. (2002) Clinical trials of gene therapy for atherosclerotic cardiovascular disease, *Curr. Opin. Lipidol.* **13**, 653–661.

Glass, C.K., Witztum, J.L. (2001) Atherosclerosis. the road ahead, *Cell* **104**, 503–516.

Herweijer, H., Wolff, J.A. (2003) Progress and prospects: naked DNA gene transfer and therapy, *Gene Ther.* **10**, 453–458.

Khurana, R., Martin, J.F., Zachary, I. (2001) Gene therapy for cardiovascular disease: a case for cautious optimism, *Hypertension* **38**, 1210–1216.

Morishita, R. (2002) Recent progress in gene therapy for cardiovascular disease, *Circ. J.* **66**, 1077–1086.

Newby, A.C., Zaltsman, A.B. (2000) Molecular mechanisms in intimal hyperplasia, *J. Pathol.* **190**, 300–309.

Ross, R. (1999) Atherosclerosis – an inflammatory disease, *N. Engl. J. Med.* **340**, 115–126.

Springer, T.A., Cybulsky, M.I. (1996) Traffic signals on endothelium for leukocytes in health, inflammation, and atherosclerosis, in: Fuster, V., Ross, R., Topol, E.J. (Eds.) *Atherosclerosis and Coronary Artery Disease*, Lippincott-Raven, Philadelphia.

Tangirala, R.K., Tsukamoto, K., Chun, S.H., Usher, D., Pure, E., Rader, D.J. (1999) Regression of atherosclerosis induced by liver-directed gene transfer of apolipoprotein A-I in mice, *Circulation* **100**, 1816–1822.

Thomas, C.E., Ehrhardt, A., Kay, M.A. (2003) Progress and problems with the use of viral vectors for gene therapy, *Nat. Rev. Genet.* **4**, 346–358.

Von Der Thusen, J.H., Kuiper, J., Van Berkel, T.J., Biessen, E.A. (2003) Interleukins in atherosclerosis: molecular pathways and therapeutic potential, *Pharmacol. Rev.* **55**, 133–166.

Primary Literature

Aiello, R.J., Bourassa, P.A., Lindsey, S., Weng, W., Natoli, E., Rollins, B.J., Milos, P.M. (1999) Monocyte chemoattractant protein-1 accelerates atherosclerosis in apolipoprotein E-deficient mice, *Arterioscler. Thromb. Vasc. Biol.* **19**, 1518–1525.

Aiello, R.J., Brees, D., Bourassa, P.A., Royer, L., Lindsey, S., Coskran, T., Haghpassand, M., Francone, O.L. (2002) Increased atherosclerosis in hyperlipidemic mice with inactivation of ABCA1 in macrophages, *Arterioscler. Thromb. Vasc. Biol.* **22**, 630–637.

Babaev, V.R., Gleaves, L.A., Carter, K.J., Suzuki, H., Kodama, T., Fazio, S., Linton, M.F. (2000) Reduced atherosclerotic lesions in mice deficient for total or macrophage-specific expression of scavenger receptor-A, *Arterioscler. Thromb. Vasc. Biol.* **20**, 2593–2599.

Boisvert, W.A., Black, A.S., Curtiss, L.K. (1999) ApoA1 reduces free cholesterol accumulation in atherosclerotic lesions of ApoE-deficient mice transplanted with ApoE-expressing macrophages, *Arterioscler. Thromb. Vasc. Biol.* **19**, 525–530.

Boisvert, W.A., Curtiss, L.K. (1999) Elimination of macrophage-specific apolipoprotein E reduces diet-induced atherosclerosis in C57BL/6J male mice, *J. Lipid Res.* **40**, 806–813.

Boisvert, W.A., Santiago, R., Curtiss, L.K., Terkeltaub, R.A. (1998) A leukocyte homologue of the IL-8 receptor CXCR2 mediates the accumulation of macrophages in atherosclerotic lesions of LDL receptor-deficient mice, *J. Clin. Invest.* **101**, 353–363.

Boring, L., Gosling, J., Cleary, M., Charo, I. (1998) Decreased lesion formation in CCR2-/- mice reveals a role for chemokines in the initiation of atherosclerosis, *Nature* **394**, 894–897.

Brennan, M.L., Anderson, M.M., Shih, D.M., Qu, X.D., Wang, X., Mehta, A.C., Lim, L.L., Shi, W., Hazen, S.L., Jacob, J.S., Crowley, J.R., Heinecke, J.W., Lusis, A.J. (2001) Increased atherosclerosis in myeloperoxidase-deficient mice, *J. Clin. Invest.* **107**, 419–430.

Brown, D.L., Hibbs, M.S., Kearney, M., Isner, J.M. (1997) Differential expression of 92-kDa gelatinase in primary atherosclerotic versus restenotic coronary lesions, *Am. J. Cardiol.* **79**, 878–882.

Buono, C., Come, C.E., Stavrakis, G., Maguire, G.F., Connelly, P.W., Lichtman, A.H. (2003) Influence of interferon-gamma on the extent and phenotype of diet-induced atherosclerosis in the LDLR-deficient mouse, *Arterioscler. Thromb. Vasc. Biol.* **23**, 454–460.

Caligiuri, G., Rudling, M., Ollivier, V., Jacob, M.P., Michel, J.B., Hansson, G.K., Nicoletti, A. (2003) Interleukin-10 deficiency increases atherosclerosis, thrombosis, and low-density lipoproteins in apolipoprotein E knockout mice, *Mol. Med.* **9**, 10–17.

Chen, S.J., Rader, D.J., Tazelaar, J., Kawashiri, M., Gao, G., and Wilson, J.M. (2000) Prolonged correction of hyperlipidemia in mice with familial hypercholesterolemia using an adeno-associated viral vector expressing very-low-density lipoprotein receptor, *Mol. Ther.* **2**, 256–261.

Collins, R.G., Velji, R., Guevara, N.V., Hicks, M.J., Chan, L., Beaudet, A.L. (2000) P-Selectin or intercellular adhesion molecule (ICAM)-1 deficiency substantially protects against atherosclerosis in apolipoprotein E-deficient mice, *J. Exp. Med.* **191**, 189–194.

Cyrus, T., Pratico, D., Zhao, L., Witztum, J.L., Rader, D.J., Rokach, J., FitzGerald, G.A., Funk, C.D. (2001) Absence of 12/15-lipoxygenase expression decreases lipid peroxidation and atherogenesis in apolipoprotein e-deficient mice, *Circulation* **103**, 2277–2282.

Dawson, T.C., Kuziel, W.C., Osahar, T.A., Madea, N. (1999) Absence of CC chemokine receptor-2 reduces atherosclerosis in apolipoprotein E-deficient mice, *Atherosclerosis* **143**, 205–211.

Devlin, C.M., Kuriakose, G., Hirsch, E., Tabas, I. (2002) Genetic alterations of IL-1 receptor antagonist in mice affect plasma cholesterol level and foam cell lesion size, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6280–6285.

Elhage, R., Maret, A., Pieraggi, M.T., Thiers, J.C., Arnal, J.F., Bayard, F. (1998) Differential effects of interleukin-1 receptor antagonist and tumor necrosis factor binding protein on fatty-streak formation in apolipoprotein E-deficient mice, *Circulation* **97**, 242–244.

Fabunmi, R.P., Sukhova, G.K., Sugiyama, S., Libby, P. (1998) Expression of tissue inhibitor of metalloproteinases-3 in human atheroma and regulation in lesion-associated cells: a potential protective mechanism in plaque stability, *Circ. Res.* **83**, 270–278.

Fazio, S., Babaev, V.R., Burleigh, M.E., Major, A.S., Hasty, A.H., Linton, M.F. (2002) Physiological expression of macrophage apoE in the artery wall reduces atherosclerosis in severely hyperlipidemic mice, *J. Lipid Res.* **43**, 1602–1609.

Febbraio, M., Podrez, E.A. (2000) Targeted disruption of the class B scavenger receptor CD36 protects against atherosclerotic lesion development in mice, *J. Clin. Invest.* **105**, 1049–1056.

Forlow, S.B., Ley, K. (2001) Selectin-independent leukocyte rolling and adhesion in mice deficient in E-, P-, and L-selectin and ICAM-1, *Am. J. Physiol. Heart Circ. Physiol.* **280**, H634–H641.

Frenette, P.S., Wagner, D.D. (1997) Insights into selectin function from knockout mice, *Thromb. Haemost.* **78**, 60–64.

Galis, Z.S., Johnson, C., Godin, D., Magid, R., Shipley, J.M., Senior, R.M., Ivan, E. (2002) Targeted disruption of the matrix metalloproteinase-9 gene impairs smooth muscle cell migration and geometrical arterial remodeling, *Circ. Res.* **91**, 852–859.

George, J., Afek, A., Shaish, A., Levkovitz, H., Bloom, N., Cyrus, T., Zhao, L., Funk, C.D., Sigal, E., Harats, D. (2001) 12/15-Lipoxygenase gene disruption attenuates atherogenesis in LDL receptor-deficient mice, *Circulation* **104**, 1646–1650.

George, J., Mulkins, M., Shaish, A., Casey, S., Schatzman, R., Sigal, E., Harats, D. (2000) Interleukin (IL)-4 deficiency does not influence fatty streak formation in C57BL/6 mice, *Atherosclerosis* **153**, 403–411.

George, J., Shoenfeld, Y., Gilburd, B., Afek, A., Shaish, A., Harats, D. (2000) Requisite role for interleukin-4 in the acceleration of fatty streaks induced by heat shock protein 65 or mycobacterium tuberculosis, *Circ. Res.* **86**, 1203–1210.

Gupta, S., Pablo, A.M., Jiang, X., Wang, N., Tall, A.R., Schindler, C. (1997) IFN-gamma potentiates atherosclerosis in ApoE knock-out mice, *J. Clin. Invest.* **99**, 2752–2761.

Harats, D., Shaish, A., George, J., Mulkins, M., Kurihara, H., Levkovitz, H., Sigal, E. (2000) Overexpression of 15-lipoxygenase in vascular endothelium accelerates early atherosclerosis in LDL receptor-deficient mice, *Arterioscler. Thromb. Vasc. Biol.* **20**, 2100–2105.

Huber, S.A., Sakkinen, P., Conze, D., Hardin, N., Tracy, R. (1999) Interleukin-6 exacerbates early atherosclerosis in mice, *Arterioscler. Thromb. Vasc. Biol.* **19**, 2364–2367.

Inoue, S., Egashira, K., Ni, W., Kitamoto, S., Usui, M., Otani, K., Ishibashi, M., Hiasa, K., Nishida, K., Takeshita, A. (2002) Anti-monocyte chemoattractant protein-1 gene therapy limits progression and destabilization of established atherosclerosis in apolipoprotein E-knockout mice, *Circulation* **106**, 2700–2706.

Jalkanen, J., Leppanen, P., Narvanen, O., Greaves, D.R., Yla-Herttuala, S. (2003) Adenovirus-mediated gene transfer of a secreted decoy human macrophage scavenger receptor (SR-AI) in LDL receptor knock-out mice, *Atherosclerosis* **169**, 95–103.

Jong, M.C., van Dijk, K.W., Dahlmans, V.E., Van der Boom, H., Kobayashi, K., Oka, K., Siest, G., Chan, L., Hofker, M.H., Havekes, L.M. (1999) Reversal of hyperlipidaemia in apolipoprotein C1 transgenic mice by adenovirus-mediated gene delivery of the low-density-lipoprotein receptor, but not by the very-low-density-lipoprotein receptor, *Biochem. J.* **338**(Pt 2), 281–287.

Joyce, C.W., Amar, M.J., Lambert, G., Vaisman, B.L., Paigen, B., Najib-Fruchart, J., Hoyt, R.F., Jr., Neufeld, E.D., Remaley, A.T., Fredrickson, D.S., Brewer, H.B., Jr., Santamarina-Fojo, S. (2002) The ATP binding cassette transporter A1 (ABCA1) modulates the development of aortic atherosclerosis in C57BL/6 and apoE-knockout mice, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 407–412.

Kawashiri, M., Zhang, Y., Usher, D., Reilly, M., Pure, E., Rader, D.J. (2001) Effects of coexpression of the LDL receptor and apoE on cholesterol metabolism and atherosclerosis in LDL receptor-deficient mice, *J. Lipid Res.* **42**, 943–950.

Kim, I.H., Jozkowicz, A., Piedra, P.A., Oka, K., Chan, L. (2001) Lifetime correction of genetic

deficiency in mice with a single injection of helper-dependent adenoviral vector, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13282–13287.

King, V.L., Szilvassy, S.J., Daugherty, A. (2002) Interleukin-4 deficiency decreases atherosclerotic lesion formation in a site-specific manner in female LDL receptor−/− mice, *Arterioscler. Thromb. Vasc. Biol.* **22**, 456–461.

Kirk, E.A., Dinauer, M.C., Rosen, H., Chait, A., Heinecke, J.W., LeBoeuf, R.C. (2000) Impaired superoxide production due to a deficiency in phagocyte NADPH oxidase fails to inhibit atherosclerosis in mice, *Arterioscler. Thromb. Vasc. Biol.* **20**, 1529–1535.

Kozaki, K., Kaminski, W.E., Tang, J., Hollenbach, S., Lindahl, P., Sullivan, C., Yu, J.C., Abe, K., Martin, P.J., Ross, R., Betsholtz, C., Giese, N.A., Raines, E.W. (2002) Blockade of platelet-derived growth factor or its receptors transiently delays but does not prevent fibrous cap formation in ApoE null mice, *Am. J. Pathol.* **161**, 1395–1407.

Kozarsky, K.F., Donahee, M.H., Glick, J.M., Krieger, M., Rader, D.J. (2000) Gene transfer and hepatic overexpression of the HDL receptor SR-BI reduces atherosclerosis in the cholesterol-fed LDL receptor-deficient mouse, *Arterioscler. Thromb. Vasc. Biol.* **20**, 721–727.

Lemaitre, V., O'Byrne, T.K., Borczuk, A.C., Okada, Y., Tall, A.R., D'Armiento, J. (2001) ApoE knockout mice expressing human matrix metalloproteinase-1 in macrophages have less advanced atherosclerosis, *J. Clin. Invest.* **107**, 1227–1234.

Linton, M.F., Atkinson, J.B., Fazio, S. (1995) Prevention of atherosclerosis in apolipoprotein E-deficient mice by bone marrow transplantation, *Science* **267**, 1034–1037.

Luoma, J., Hiltunen, T. (1994) Expression of alpha 2-macroglobulin receptor/low density lipoprotein receptor-related protein and scavenger receptor in human atherosclerotic lesions, *J. Clin. Invest.* **93**, 2014–2021.

Lutgens, E., Daemen, M.J. (2002) CD40-CD40L interactions in atherosclerosis, *Trends Cardiovasc. Med.* **12**, 27–32.

Mallat, Z., Gojova, A., Marchiol-Fournigault, C., Esposito, B., Kamate, C., Merval, R., Fradelizi, D., Tedgui, A. (2001) Inhibition of transforming growth factor-beta signaling accelerates atherosclerosis and induces an unstable plaque phenotype in mice, *Circ. Res.* **89**, 930–934.

Mehrabian, M., Allayee, H., Wong, J., Shi, W., Wang, X.P., Shaposhnik, Z., Funk, C.D., Lusis, A.J., Shih, W. (2002) Identification of 5-lipoxygenase as a major gene contributing to atherosclerosis susceptibility in mice, *Circ. Res.* **91**, 120–126.

Murayama, T., Yokode, M., Kataoka, H., Imabayashi, T., Yoshida, H., Sano, H., Nishikawa, S., Nishikawa, S., Kita, T. (1999) Intraperitoneal administration of anti-c-fms monoclonal antibody prevents initial events of atherogenesis but does not reduce the size of advanced lesions in apolipoprotein E-deficient mice, *Circulation* **99**, 1740–1746.

Ni, W., Egashira, K., Kitamoto, S., Kataoka, C., Koyanagi, M., Inoue, S., Imaizumi, K., Akiyama, C., Nishida, K.I., Takeshita, A. (2001) New anti-monocyte chemoattractant protein-1 gene therapy attenuates atherosclerosis in apolipoprotein E-knockout mice, *Circulation* **103**, 2096–2101.

Pennacchio, L.A., Olivier, M., Hubacek, J.A., Cohen, J.C., Cox, D.R., Fruchart, J.C., Krauss, R.M., Rubin, E.M. (2001) An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing, *Science* **294**, 169–173.

Pinderski, L.J., Fischbein, M.P., Subbanagounder, G., Fishbein, M.C., Kubo, N., Cheroutre, H., Curtiss, L.K., Berliner, J.A., Boisvert, W.A. (2002) Overexpression of interleukin-10 by activated T lymphocytes inhibits atherosclerosis in LDL receptor-deficient mice by altering lymphocyte and macrophage phenotypes, *Circ. Res.* **90**, 1064–1071.

Ross, R., Glomset, J.A. (1973) Atherosclerosis and the arterial smooth muscle cell: proliferation of smooth muscle is a key event in the genesis of the lesions of atherosclerosis, *Science* **180**, 1332–1339.

Sano, H., Sudo, T., Yokode, M., Murayama, T., Kataoka, H., Takakura, N., Nishikawa, S., Nishikawa, S.I., Kita, T. (2001) Functional blockade of platelet-derived growth factor receptor-beta but not of receptor-alpha prevents vascular smooth muscle cell accumulation in fibrous cap lesions in apolipoprotein E-deficient mice, *Circulation* **103**, 2955–2960.

Savontaus, M.J., Sauter, B.V., Huang, T.G., Woo, S.L. (2002) Transcriptional targeting of conditionally replicating adenovirus to dividing endothelial cells, *Gene Ther.* **9**, 972–979.

Schreyer, S.A., Peschon, J.J., LeBoeuf, R.C. (1996) Accelerated atherosclerosis in mice lacking tumor necrosis factor receptor p55, *J. Biol. Chem.* **271**, 26174–26178.

Shichiri, M., Tanaka, A., Hirata, Y. (2003) Intravenous gene therapy for familial hypercholesterolemia using ligand-facilitated transfer of a liposome: LDL receptor gene complex, *Gene Ther.* **10**, 827–831.

Simons, M., Edelman, E.R. (1997) Antisense oligonucleotide inhibition of PDGFR-ß receptor subunit expression directs suppression of intimal thickening, *Circulation* **95**, 669–676.

Singaraja, R.R., Bocher, V., James, E.R., Clee, S.M., Zhang, L.H., Leavitt, B.R., Tan, B., Brooks-Wilson, A., Kwok, A., Bissada, N., Yang, Y.Z., Liu, G., Tafuri, S.R., Fievet, C., Wellington, C.L., Staels, B., Hayden, M.R. (2001) Human ABCA1 BAC transgenic mice show increased high density lipoprotein cholesterol and ApoAI-dependent efflux stimulated by an internal promoter containing liver X receptor response elements in intron 1, *J. Biol. Chem.* **276**, 33969–33979.

Smith, J.D., Trogan, E., Ginsberg, M., Grigaux, C., Tian, J., Miyata, M. (1995) Decreased atherosclerosis in mice deficient in both macrophage colony-stimulating factor (op) and apolipoprotein E, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8264–8268.

Sukhova, G.K., Schonbeck, U., Rabkin, E., Schoen, F.J., Poole, A.R., Billinghurst, R.C., Libby, P. (1999) Evidence for increased collagenolysis by interstitial collagenases-1 and -3 in vulnerable human atheromatous plaques, *Circulation* **99**, 2503–2509.

Suzuki, H., Kurihara, Y., Takeya, M., Kamada, N., Kataoka, M., Jishage, K., Sakaguchi, H., Kruijt, J.K., Higashi, T., Suzuki, T., van Berkel, T.J., Horiuchi, S., Takahashi, K., Yazaki, Y., Kodama, T. (1997) The multiple roles of macrophage scavenger receptors (MSR) in vivo: resistance to atherosclerosis and susceptibility to infection in MSR knockout mice, *J. Atheroscler. Thromb.* **4**, 1–11.

Thorngate, F.E., Rudel, L.L., Walzem, R.L., Williams, D.L. (2000) Low levels of extrahepatic nonmacrophage ApoE inhibit atherosclerosis without correcting hypercholesterolemia in ApoE-deficient mice, *Arterioscler. Thromb. Vasc. Biol.* **20**, 1939–1945.

Tsukamoto, K., Tangirala, R.K., Chun, S., Usher, D., Pure, E., Rader, D.J. (2000) Hepatic expression of apolipoprotein E inhibits progression of atherosclerosis without reducing cholesterol levels in LDL receptor-deficient mice, *Mol. Ther.* **1**, 189–194.

Whitman, S.C., Ravisankar, P., Daugherty, A. (2002) IFN-gamma deficiency exerts gender-specific effects on atherogenesis in apolipoprotein E−/− mice, *J. Interferon Cytokine Res.* **22**, 661–670.

Whitman, S.C., Ravisankar, P., Daugherty, A. (2002) Interleukin-18 enhances atherosclerosis in apolipoprotein E(−/−) mice through release of interferon-gamma, *Circ. Res.* **90**, E34–E38.

Whitman, S.C., Ravisankar, P., Elam, H., Daugherty, A. (2000) Exogenous interferon-gamma enhances atherosclerosis in apolipoprotein E−/− mice, *Am. J. Pathol.* **157**, 1819–1824.

Yoshimura, S., Morishita, R., Hayashi, K., Yamamoto, K., Nakagami, H., Kaneda, Y., Sakai, N., Ogihara, T. (2001) Inhibition of intimal hyperplasia after balloon injury in rat carotid artery model using cis-element 'decoy' of nuclear factor-kappaB binding site as a novel molecular strategy, *Gene Ther.* **8**, 1635–1642.

Yu, L., Li-Hawkins, J., Hammer, R.E., Berge, K.E., Horton, J.D., Cohen, J.C., Hobbs, H.H. (2002) Overexpression of ABCG5 and ABCG8 promotes biliary cholesterol secretion and reduces fractional absorption of dietary cholesterol, *J. Clin. Invest.* **110**, 671–680.

Zhao, L., Cuff, C.A., Moss, E., Wille, U., Cyrus, T., Klein, E.A., Pratico, D., Rader, D.J., Hunter, C.A., Pure, E., Funk, C.D. (2002) Selective interleukin-12 synthesis defect in 12/15-lipoxygenase-deficient macrophages associated with reduced atherosclerosis in a mouse model of familial hypercholesterolemia, *J. Biol. Chem.* **277**, 35350–35356.

Genetic Analysis of Populations

Rus A. Hoelzel
Durham University, Durham, UK

1 Introduction 244

2 Nature of Mutations and Mutation Rates 245

3 Dissemination of Genetic Change in Natural Populations 247
3.1 Natural Selection 247
3.2 Genetic Drift 248

4 Methods of Analysis 249
4.1 Sample Collection 249
4.2 Enzyme Electrophoresis 250
4.3 DNA Extraction 251
4.4 RFLP Analysis 252
4.5 DNA Sequencing 252
4.6 PCR Analysis 254
4.7 SSCP and Gradient Gel Analysis 254
4.8 RAPD Analysis 255

5 Choice of Genetic Marker and Interpretation 256
5.1 Allozymes 256
5.2 Nuclear Genes 256
5.3 Mitochondrial DNA 257
5.4 Chloroplast DNA 258
5.5 Minisatellite DNA 259
5.6 Microsatellite DNA 260

6 Statistical Analysis of Variation 261

7 Conclusions 263

Bibliography 263

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

Keywords

Demographics
The study of populations, especially growth rate and age structure.

Effective Population Size
The size of an ideal population that would show the same rate of decay in heterozygosity as the actual population.

Electrophoresis
The movement of molecules on a solid medium through an electric field. Various support mediums are used, including filter paper, starch gel, agarose, and polyacrylamide.

Molecular Clock
The rate of molecular genetic change at a given gene, often including the assumption of a constant rate.

Polymorphism
Two or more genetically distinct types in the same interbreeding population.

1 Introduction

The molecular genetic analysis of populations involves three main stages: sample collection, analysis, and interpretation. The sample must be representative of the population, and therefore collections need to be random and sufficiently large. The number of samples analyzed is especially important when comparisons are based on allele frequencies. In each case, the allele frequency determined by the sample will be an estimate, and the 95% confidence limits around that estimate depend on the sample size as follows:

$$p \pm 1.96\sqrt{\frac{p(1-p)}{n}}$$

where p is the allele frequency and n is the sample size. For example, for an estimated allele frequency of 55%, the 95% confidence limits around that estimate would range from 33 to 77% for a sample of 20 and from 45 to 65% for a sample of 100.

The method of molecular analysis and choice of genetic marker will depend on the specific question being addressed, although in general it is advantageous to assess diversity at a number of independent loci. Interpretation of results will depend in part on the method of analysis. DNA analytical methods offer great potential through the analysis of specific regions of the nuclear, mitochondrial, and chloroplast genomes, but numerous evolutionary mechanisms can effect change over time, and therefore interpretation based on time-dependent change must be done cautiously. An understanding of the genetic structure of populations facilitates the understanding of dispersal, reproductive, and social behavior, and provides essential data for the conservation of natural variation.

It has long been recognized that inherited characters vary and that the quantification of this variation can be used to distinguish local races. However, it was not until the advent of protein gel electrophoresis in the 1960s that the extent of genetic variation in natural populations began to be appreciated. Thousands of species have been investigated for protein polymorphisms, and most displayed extensive variation. Gel electrophoresis of allozymes (different electrophoretic forms of a given enzyme) became the standard method for the comparison of genetic variation within and between populations, and it is still in use today. However, most of the variation present in the genome cannot be detected by this method. The DNA sequence determines, through an RNA intermediary, the amino acid sequence in the polypeptides that combine to form proteins. Only a small proportion of the genome is translated into proteins, and the three-letter codon that determines the sequence of amino acids is degenerate, especially in the second and third positions (see Sect. 2). Therefore, even within the DNA sequence that encodes the protein, there can be considerable variation that is not expressed. Furthermore, only a small proportion of the changes that do affect the amino acid sequence will also affect the charge properties of the protein, which determine its migration through an electric current (and consequently only some changes can be detected by the electrophoretic method). Innovations over the last several decades have enabled the analysis of DNA directly, providing greater resolution and facilitating the interpretation of variation. This article concentrates on the use of these DNA analytical methods to interpret population-level genetic variation.

2 Nature of Mutations and Mutation Rates

Mutations that affect the sequence of nucleic acids can occur through the conversion of a single base (e.g. a point mutation changing A to T), through the loss by deletion of one or more bases, and through the addition by insertion of one or more bases. Change can also occur through the operation of a variety of DNA "turnover" mechanisms. For the purpose of this discussion, two mechanisms that generate variation in repetitive regions of DNA by changing the number of repeated elements will be emphasized. These are DNA slippage (Fig. 1a) and unequal crossing-over (Fig. 1b). In each case, a misalignment of DNA strands causes the molecule to be repaired in such a way that there is a gain or loss in an array of repeated elements. Repetitive DNA regions are very common in most vertebrate and plant genomes, and change in the copy number of repeats generally occurs at a much higher rate than change due to point mutations. For this reason, they provide high resolution to population genetic analyses and are often used.

Eukaryotic genes are divided into coding (exons) and noncoding regions (introns). Introns are transcribed into RNA, but not translated into proteins (Fig. 2), which means that these sequences are not phenotypically expressed. Point mutations accumulate at a higher rate in noncoding regions, in part because of a diminished level of selection. Within coding regions, the second or third position of the codon is often "synonymous"; that is, the nucleotide in those positions can change without changing the encoded amino acid. For the nuclear genome of eukaryotes, the nucleotide substitution rate (per base pair, per year) in introns

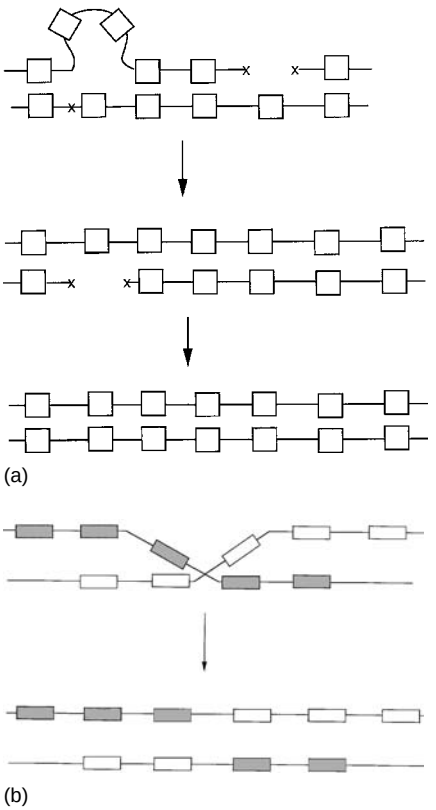

Fig. 1 (a) One of several possible DNA slippage mechanisms: the end result is an extension of the array within a single chromatid. Similar mechanisms lead to contraction of the array or slippage during replication. The process illustrated involves endonuclease-mediated breakage, slippage of the upper strand, and polymerase-mediated repair events. The boxes represent repeated elements. (b) Unequal crossing-over. Two strands, each with five repetitive elements, undergo an unequal crossing-over event, resulting in the gain of one element for the upper strand and a corresponding loss from the lower strand.

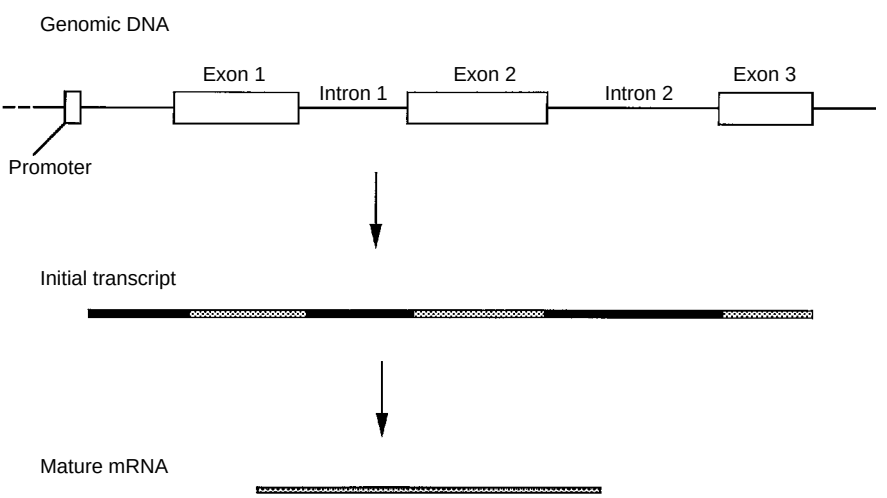

Fig. 2 Division of genes into coding (exons) and noncoding (introns) regions. The initial transcript is enzymatically modified to represent only the exon regions of the gene in the mature messenger RNA.

and synonymous sites ranges from about 1×10^{-9} to 10×10^{-9} with an average of about 5×10^{-9}. In exons, the rate varies from about 0.004×10^{-9} in histone genes to 2.8×10^{-9} in interferon A. Natural selection (see Sect. 3.1) usually prevents major disruption of coding sequences with repetitive DNA regions (though there are exceptions such as the FMR1 locus), but this is a common feature of noncoding regions. The rate of DNA turnover in these repetitive sequences depends on which mechanisms are affecting change and various other factors. The rate of change due to these processes can be very fast; rates up to 10^{-2} have been documented.

3 Dissemination of Genetic Change in Natural Populations

Natural selection and genetic drift are the two principal mechanisms by which genetic change is spread through a population. The correct interpretation of genetic variation depends on how the DNA region under analysis is affected by each of these mechanisms. Although it is beyond the scope of this article to provide a detailed discussion, the basic tenets are outlined below.

3.1 Natural Selection

The biochemical, morphological, physiological, or behavioral consequences of genetic makeup (*genotype*) are referred to as the *phenotype*. Natural selection is the differential reproductive success of different phenotypes. Generally, if a region of DNA is not expressed phenotypically (nor affected by or closely linked to a phenotypically expressed region), it is not exposed to natural selection. On the other hand, some genes are critical to survival or reproductive success and therefore are under strong selective control. A simple model for the change in allele frequency due to selection in a two-allele system for a diploid organism can be expressed as follows:

$$p' = \frac{p^2 w_{11} + pq w_{12}}{W}$$

where p' is the frequency of allele A_1 after selection, p is the frequency of A_1 before selection, q is the frequency of A_2 before selection, w_{11} is the viability of the homozygous genotype A_1A_1, w_{12} is the viability of the heterozygous genotype A_1A_2, and W is the average viability. From this relationship, it is possible to derive the expected change in allele frequency over time due to selection and make predictions about the dynamics of natural selection under different environmental conditions.

Genetic variation at a locus can be maintained in a population by balancing selection such as overdominance (when the heterozygote condition is selected for) and frequency-dependent selection, or limited by directional selection (when one allele is favored at the expense of the others). Adaptation to specific habitats and the general character of the local environment can also affect genetic diversity. For example, it has been proposed that habitat specialists might be expected to show lower levels of genetic diversity than habitat generalists, and species inhabiting relatively stable environments might be expected to show low variability. In each case, the prediction is based on the idea that a more complex or variable environment will support a greater degree of genetic variation within the population as a whole.

3.2 Genetic Drift

Genetic drift is the random changes in allele frequency over time in finite populations, resulting from variation in the number of offspring among individuals. According to the "neutral theory," most evolution occurs through the random fixation of selectively neutral (or slightly deleterious) mutations. This theory predicts that change accumulating in this way will do so at an approximately constant rate (a molecular clock). When this is true, genetic change can be used to interpret the demographic histories of populations. Figure 3 plots the expected relationship between effective population size, genetic diversity, and the time (in generations) required for lost variability to be recovered. The relationship between diversity (expressed as heterozygosity) and population size can be expressed as follows:

$$H_t = H_{t-1}\left(1 - \left(\frac{1}{2N_e}\right)\right)$$

where H_t is the heterozygosity in generation "t," H_{t-1} is the heterozygosity in the previous generation, and N_e is the effective population size. In general, diversity is lost more quickly in small populations owing to the greater likelihood of sexually recombining alleles that are identical by descent (inbreeding).

Stochastic events (such as a population bottleneck) can also reduce variation in a population. The result is a small population in which the effects of genetic drift are expected to reduce levels of variation, both as a consequence of the

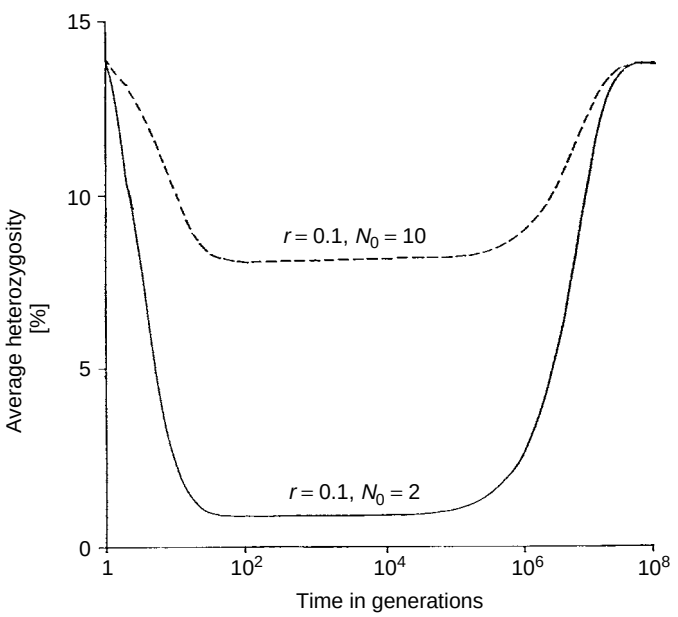

Fig. 3 The theoretical changes in average heterozygosity when a population goes through a bottleneck. Logistic growth is assumed, and the intrinsic rate of growth r is set at 0.1. In one case (solid line), the bottleneck effective population size is $N_0 = 2$; in the other (dashed line) $N_0 = 10$ [After Nei (1987)].

sampling of genotypes (and the resulting loss of allelic diversity) and by inbreeding (as described above). Once variation has been lost, the process of accumulating new mutations is very slow (Fig. 3). An equilibrium level of variation will take on the order of the inverse of the mutation rate to recover, which is in some cases longer than the expected lifespan of the species.

4 Methods of Analysis

There are numerous methods by which genetic variation can be directly or indirectly measured. This article concentrates on those methods that have been fundamental to population genetic studies, now and historically. The evolutionary history of a given gene can be affected by various processes including selection and drift, and provides just one record in the study population. For this reason, it is important to investigate variation at as many loci as practical, so that information on the history of the population can be accumulated, rather than simply the history of a gene in that population. At the same time, there can be some utility to choosing markers of a particular level of polymorphism to facilitate different types of analyses. For example, a marker that evolves relatively slowly but is thought to be selectively neutral would allow a relatively unambiguous assessment of the phylogenetic relationship between distantly related taxa. At the other extreme, hypervariable markers such as microsatellite or minisatellite arrays (see Sect. 5) can be used to assess paternity and close kinship relationships. For comparisons among populations, too much variation will obscure genetic differences (saturation will lead variation within a population to be as great as that between populations), and too little variation will not permit sufficient resolution. Repetitive DNA regions can vary in the rate of change over time over a very broad spectrum, owing to the nature of the processes involved. The comparison of these loci therefore needs to take this potential rate heterogeneity into account.

4.1 Sample Collection

Essentially, any cellular tissue is suitable for the extraction of DNA (see Sect. 4.3), though it some tissues (such as bone or wood) are harder to extract DNA from than from others (such as skin or leaves). The choice of tissue for enzyme electrophoresis (see Sect. 4.2) may be restricted by which enzyme systems are expressed in different tissues. The first step in any project is to define the objectives and plan what can be accomplished on the basis of time and budget. In most cases, it will be more important to include multiple loci than very large numbers of individuals; however, small sample sizes risk low power (see below) and distortion of allele frequencies due to sampling effects. The analysis of a large number of loci helps minimize the impact of locus-specific effects, such as selection, as indicated in the previous section. Some applications will require a choice among types of loci, for example, tests for biparental kinship require nuclear loci and high resolution, and the obvious choice becomes microsatellite loci (see Sect. 5.5).

There are two types of error in the relevant statistical comparisons among populations. A type I error occurs when a null hypothesis that should be accepted is rejected. In biological studies, the allowance for this type of error is often

set at 5%, though a more stringent criterion should be adopted if a large number of comparisons are done. A type II error occurs when a null hypothesis that should be rejected is accepted. This is usually expressed as the power of the test, which is 1 minus the type II error. For a comparison between two populations using just one marker, with a difference of 5% between allele frequencies, thousands of individuals would need to be screened to statistically distinguish between them, given a type I error of 5% and a power of 90%. The solution is to use more markers with an emphasis on those with a high level of variability. Typically, it will be sufficient to represent each population with a random sample of at least 50 individuals and compare putative populations at 10 to 20 polymorphic loci.

The collection of samples will depend to some extent on the biology of the study species. In mobile, sexually reproducing species, the dispersal behavior of males and females can affect the distribution of genotypes. For example, if females are phylopatric (reproducing rear where they were born), genetic variation between adult males and females in local samples may vary. Furthermore, in this case, the comparison of populations using a matrilineally inherited marker (such as mtDNA in mammals) will exaggerate the difference between populations. It is also important to take into account temporal and spatial patterns of movement in mobile species. A project that begins sampling in one location and finishes in another sometime later might simply follow the migration of a single reproductive population! With plant species, it is important to consider the dispersal of both pollen and seeds, and the possibility that seeds "bank" in the soil, providing a reservoir of genetic material from previous generations.

4.2 Enzyme Electrophoresis

The migration of proteins through an electric field in a gel matrix can be measured by enzyme electrophoresis. The charge properties of the protein determine the extent and direction of migration. The different electrophoretic forms of the enzyme are known as allozymes. Tissue homogenized in buffer can be used as the source material, and specific enzymes can then be visualized by immersing the gel into a medium containing the reaction conditions for that enzyme such that a colored dye is incorporated into the end product. The dye produces bands on the gel indicating the position of the enzyme. The number of bands will depend on the number of subunits in the enzyme and on whether the locus is homozygous or heterozygous. Gels can be prepared from a variety of media (starch, polyacrylamide, agar, etc.) and run by a variety of methods (horizontal, vertical, disc, etc.). For screening large numbers of individuals in a population survey, the most common method (also the simplest and least expensive) is horizontal starch gel electrophoresis. The apparatus for this method is schematized in Fig. 4. Gels running 10 to 30 samples can be cut horizontally 2 times (giving 3 slices) and each slice stained for a different enzyme. In most cases, the resolution of bands is sufficient, and poor resolution can often be improved by using a different gel medium or running method. In general, the different methods are designed to improve resolution and to permit the separation of more subtle differences between alleles.

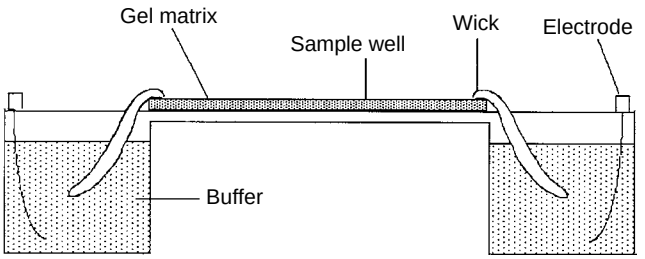

Fig. 4 Nondenaturing horizontal gel electrophoresis of proteins.

4.3 DNA Extraction

DNA is extracted from tissue by removing proteins, fats, and carbohydrates and then precipitating the DNA out of an aqueous salt solution. This procedure is usually straightforward for animal tissues, but it can be quite difficult with plant material because naturally occurring chemicals degrade the DNA or inhibit further analysis. The first step is to homogenize the tissue under an aqueous-based buffer that promotes cell lysis. By the most common method (Fig. 5), the homogenate is mixed with phenol and chloroform, which are immiscible with water. The proteins and other cellular compounds go into the organic (phenol) phase, while the DNA remains in the aqueous phase. After several repetitions, the DNA is precipitated with salt and ethanol. A careful preparation of DNA will include sheared fragments 20 to 100 kb long. Excessive agitation

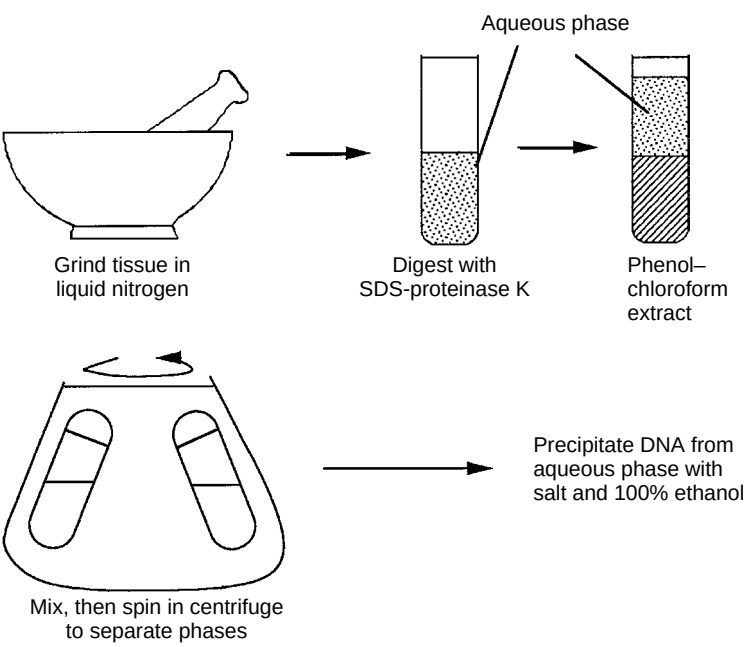

Fig. 5 Phenol–chloroform extraction of DNA from solid tissue.

further shears (degrades) the DNA. Since nucleases (which can be present on the fingers and particulate matter in the air) will also degrade DNA, the procedure must be done under sterile conditions.

4.4 RFLP Analysis

Restriction fragment length polymorphisms (RFLPs) can be analyzed when DNA is digested with a special enzyme called a *restriction endonuclease* (Fig. 6). These enzymes recognize specific short sequences of DNA (e.g. the enzyme *Eco*RI recognizes the sequence GAATTC) and cleave the DNA at every instance of that sequence. Mutations that create new sites or eliminate old ones can occur. The next step is to separate the DNA fragments by size in a gel matrix. This is done by electrophoresing the samples through an agarose or polyacrylamide gel under buffer. DNA naturally assorts by size (on a roughly logarithmic scale) by this method. The double-stranded DNA in the gel is then made single-stranded (usually with an alkaline solution) and is transferred by "Southern blotting" onto a solid membrane that binds DNA (Fig. 7). Then, a segment of DNA to be compared between individuals (e.g. the *Adh* gene or the mitochondrial genome) is labeled with a radioisotope (or some nonradioactive label) and hybridized (whereby complementary strands of DNA combine and re-form a double-stranded molecule) to the DNA bound to the membrane. Only the fragments of bound DNA that match this DNA probe will be made radioactive. After exposure to X-ray film, these relevant fragments can be visualized as black (exposed) bands on the film (Fig. 8). Now that DNA sequencing (see Sect. 4.5) has become routine and affordable, RFLP analysis is used less frequently for population studies.

4.5 DNA Sequencing

The procedure that determines the sequence of nucleic acid bases comprising

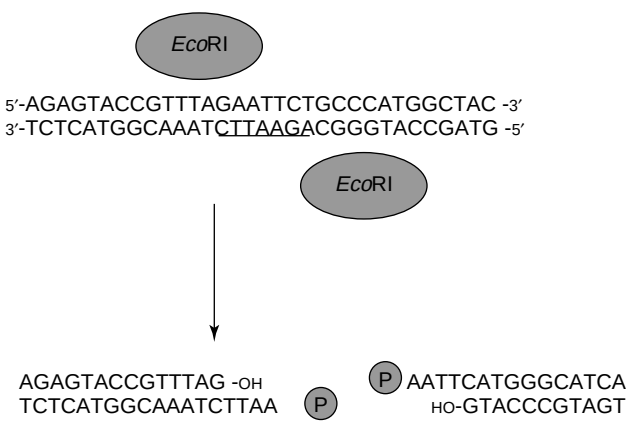

Fig. 6 Restriction digestion with the enzyme *Eco*RI. Digestion leaves a 3′ hydroxyl group and a 5′ phosphate group exposed, and either a 5′ overhang (shown), a 3′ overhang, or blunt ends, depending on the enzyme. *Eco*RI has a 6-bp recognition site. Many other common enzymes have 4-bp recognition sites.

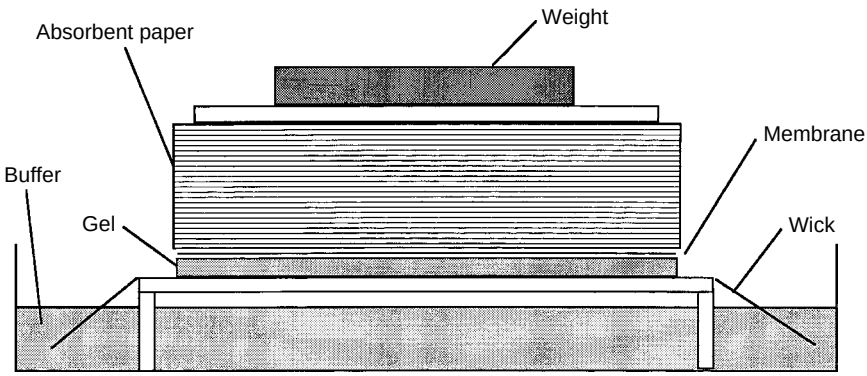

Fig. 7 Southern blotting to transfer DNA from gel to membrane.

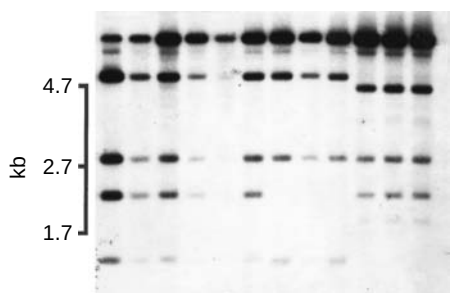

Fig. 8 Panel of 12 lions digested with *Ava*II and hybridized with a heterologous mtDNA probe. (Courtesy of Robyn Hottman and Steve O'Brien.)

a segment of DNA offers the greatest resolution for comparisons of genetic differentiation and permits the greatest scope for interpretation. There are several methods of DNA sequencing, but the one used most frequently is chain termination. By the chain termination method, a single-stranded DNA template is divided into four reaction mixes, one for each nucleotide base (A, T, C, G). A DNA polymerase (an enzyme that can synthesize DNA along a single-stranded template, starting at a double-stranded/single-stranded interphase), a short (15–25 bases) primer segment of DNA, and all four deoxynucleotide bases are included in each reaction (one of which is labeled either radioactively or with a fluorescent dye). In addition, each reaction will include a smaller quantity of one of the four dideoxynucleotides (A, T, C, or G). These bases lack the 3′-OH group (Fig. 6) necessary for DNA chain elongation and therefore terminate the reaction at every occurrence of that base. A comparison of the length of strands synthesized for each of the four reactions gives the sequence of bases (Fig. 9). Modern sequencing is facilitated by the polymerase chain reaction (see Sect. 4.6) and visualized by computer-generated profiles that display a series of peaks, each representing a different nucleotide labeled with a different fluorescent dye. This allows for very rapid analysis as the sequencing machine generates a computer file of the sequence as the DNA runs through the gel. It also permits the running of all four sequencing reactions (one for each nucleotide) in a single lane, since each is labeled with a distinct dye.

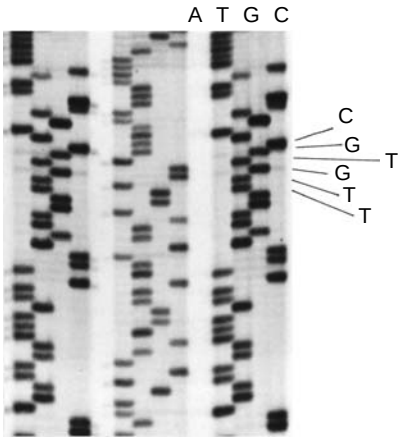

Fig. 9 A small portion of the nucleic acid sequence that can be "read" from a gel subjected to the chain termination method. The four reactions (one for each nucleic acid residue) are run side by side on a 6% polyacrylamide gel.

4.6 PCR Analysis

The polymerase chain reaction (PCR) permits the enzymatic amplification of a specific segment of DNA. This technique has greatly facilitated population-level screening of genetic variation by saving time and reducing expense. Two short primers of single-stranded DNA (usually 15–25 bases long) are designed flanking a region of interest. These are often based on published DNA sequences of particular genes or are developed from cloned DNA. A programmable heating block is then used to vary the reaction conditions. First the reaction mix is heated to denature the template DNA (to separate the double helix into its two complementary strands); then the temperature is lowered to allow the primers to anneal to the template, the reaction mix is heated to an optimal temperature for the polymerase, and the cycle is repeated 20 to 40 times (Fig. 10). The key to this technique is the use of a thermostable polymerase (most polymerases would be denatured and rendered inactive by the high temperatures necessary for the DNA denaturation step).

4.7 SSCP and Gradient Gel Analysis

Single-strand conformation polymorphisms (SSCPs) are based on the sequence-dependent mobilities of denatured (single-stranded) DNA when run through a nondenaturing gel. PCR-amplified DNA, usually 100 to 200 bp long, is denatured and loaded onto a gel in nondenaturing conditions. As the molecule migrates through the gel, it reanneals with itself, folding into a hairpin structure. The conformation and mobility of this structure depend on the sequence of DNA residues. As little as a single base change in 200 can alter the distance a fragment will migrate in the gel, although there is no simple correspondence between sequence composition and mobility. Therefore, differences can be detected quickly (permitting the screening of a large number of samples simultaneously), but they must be quantified by sequencing all products with distinct mobilities.

Gradient gel electrophoresis, an alternative method for detecting single base changes, is based on the melting (denaturing) properties of DNA in solu-

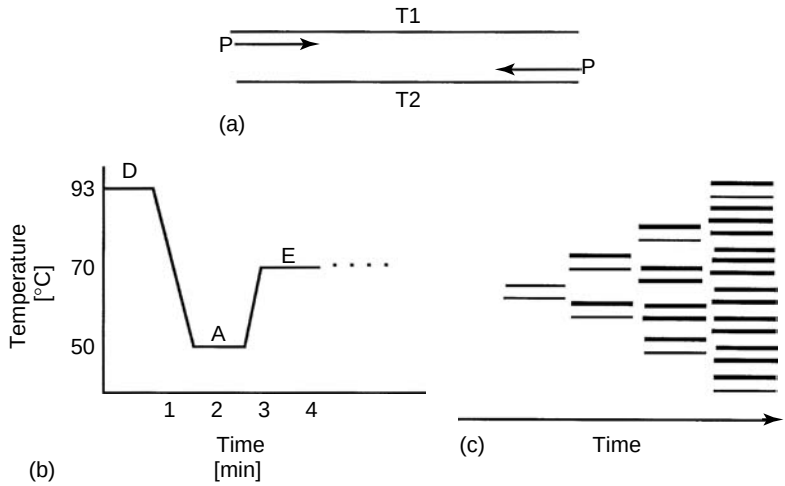

Fig. 10 The polymerase chain reaction (PCR). (a) Short oligonucleotide "primers" anneal to the denatured template DNA. (b) The reaction is cycled between denaturing, annealing, and extension temperatures. (c) Repeated cycling results in an exponential amplification of the target sequence.

tion. DNA melts in domains of two dozen to several hundred base pairs. Sequence changes within a domain are characterized by mobility changes. As the double-stranded molecule passes through the denaturing gradient, it becomes increasingly denatured, and its mobility reflects the sequence within each melting domain. Changes within the last domain would not normally be detected because sequence-dependent migration is lost upon completion of strand dissociation. To correct for this degradation, a 30- to 40-bp GC-rich sequence (a "GC-clamp") is added to the 5′ end of the PCR primers used to amplify the sequence being analyzed.

4.8 RAPD Analysis

Random amplified polymorphic DNA (RAPD) analysis uses just one short oligonucleotide (usually 10–12 bp) in a PCR reaction to amplify random segments of DNA. It works because a small number of fragments (5–10) will be amplified when the oligonucleotide anneals by chance on each template strand over a size range that can be amplified by PCR (usually about 3 kb). A 10-mer, for example, will anneal by chance once every million base pairs (about a thousand times in the human genome). There are two ways by which the amplified fragments can show polymorphisms. First, the amplified region may be repetitive and show length variation in the population (i.e. there will be a variable number of tandem repeats: VNTR). In this case, the relevant band will show a size variation, although very large alleles may not amplify (PCR preferentially amplifies smaller fragments). Second, a single base-pair polymorphism may interfere with the annealing of the primer in some individuals. In this case, the band will be present or absent, but if present, it is not expected to vary in size.

5 Choice of Genetic Marker and Interpretation

The nuclear genome is a mosaic of DNA regions under the influence of different mechanisms affecting change, and it is conserved by selection to different degrees. In addition, there are the mitochondrial genomes and, in plants, the chloroplast genomes, with their own unique properties. A population geneticist can use this variation in the mode and rate of evolution to advantage, but it also means that results must be interpreted cautiously. Six types of genetic marker are described here, each with different properties and, in some cases, best suited to particular levels of analysis.

5.1 Allozymes

Enzymes are a functional part of cell physiology and are therefore exposed to natural selection. The apparent rate of evolution in these markers varies greatly, from about 10^{-9} to 10^{-6} change per gene (given an average length of 1,000 bp) per year. In some cases – for example, in the histone genes – there is clearly strong selection, which would need to be considered when population patterns are interpreted. However, studies have shown that even in coding regions, the evolutionary rate is often relatively constant over time, and therefore allozyme polymorphism can be useful for assessing the differentiation of populations by drift. As described in Sect. 4.2, the analytical procedure (enzyme electrophoresis) is relatively fast and inexpensive; however, the level of resolution possible is relatively low, because only a small proportion of the total variation in a gene is expressed or detectable by electrophoresis (see Fig. 2 and Sect. 5.2).

The data generated by this method are for the most part Mendelian, and the patterns resulting from segregation at a single locus can usually be distinguished. This means that data can be interpreted with respect to theoretical predictions about the frequency of genotypes in a population, based on the frequency of alleles. Deviations from the expected frequency (as predicted by Hardy–Weinberg; see Sect. 6) can be quantified statistically and interpreted. Furthermore, knowing the function of the encoded protein can help with the interpretation of patterns of variation and deviations from expected frequencies. For example, an excess of heterozygotes could indicate selection for heterozygotic individuals (heterosis), while an excess of homozygotes could indicate inbreeding or directional selection. Knowledge of the gene product could help determine the likely factor. When the data are consistent with Hardy–Weinberg expectations, formulations to interpret genetic distance and relative levels of variation can be applied (see Sect. 6).

5.2 Nuclear Genes

Sequence data of nuclear genes, including those encoding enzymes, offer an immediate advantage over enzyme electrophoresis. This is because coding and noncoding regions can be differentiated, and within the coding region of the gene, synonymous and nonsynonymous sites can be distinguished. Thus, regions that are likely to be under the influence of natural selection can be separated from those that are more likely to be evolving at a neutral rate. It also becomes possible to compare sequences and to determine the

mode of mutation (point mutation, deletion, insertion, rearrangement, etc.), and in some cases, the mechanism. All these factors contribute to the accurate interpretation of the observed variation.

For comparisons between conspecific populations, it is necessary to identify both alleles for each individual, and this is typically more difficult in nuclear DNA sequencing than in enzyme electrophoresis because of the need to sequence a number of clones from each individual to ensure that both alleles have been sampled. One relatively easy way around such detailed sequencing is to screen the population using either the SSCP or gradient gel methods. These methods identify sequences that are distinct by as little as one base pair. In most cases, a single locus will show two bands when heterozygous and one when homozygous (although sometimes there are repeatable artifact bands related to additional stable conformations of the single strands). Combined with DNA sequencing to distinguish the different phenotypes indicated by SSCP or the gradient gel, this approach provides an efficient and very powerful protocol for the screening of variation at the population level. It shares most of the advantages of allozyme electrophoresis (apart from considerations of difficulty and expense) and provides considerably greater resolution.

Variation in nuclear genes can also be quantified using the RFLP method, which permits the rapid screening of a large number of samples. RFLP is limited, however, in that the precise character of the variable restriction sites is not known (unless the region has been sequenced and the sites can be mapped), and the allelic relationship at a locus cannot always be readily determined. Another consideration is the effect of methylation, to which some restriction enzymes are sensitive. Methylated sites are not cut by sensitive enzymes. This property originally evolved in bacteria as a defense against restriction enzymes produced by bacteriophage. Especially in the nuclear genome, some sites may be methylated, potentially creating within or between populations an RFLP that does not reflect a real sequence difference.

5.3 Mitochondrial DNA

In multicellular animals, the mitochondrial genome is a double-stranded, closed-circular molecule ranging in length from 15.7 to 19.5 kb. It evolves at a higher rate than comparable nuclear genomic regions (5–10 times faster; Fig. 11) and in most cases shows strict maternal transmission (no paternal contribution to the F1 generation). These two factors have made mitochondrial DNA (mtDNA) an attractive marker for the genetic comparison of animal populations. Higher levels of variation permit greater resolution, and the lack of sexual recombination facilitates interpretation. In plants the transmission is matrilineal, but the level of variation is generally very low.

Variation across the mitochondrial genome is heterogeneous, as a result of varying levels of sequence conservation, and in the "control region," DNA slippage. A common and very useful method of analysis is to sequence mtDNA regions that have been amplified by PCR. Sequences of up to 1,000 bp can now be routinely amplified and sequenced, and this is now the most common method of analysis.

Demographic considerations are important to the interpretation of mtDNA variation for two main reasons. First,

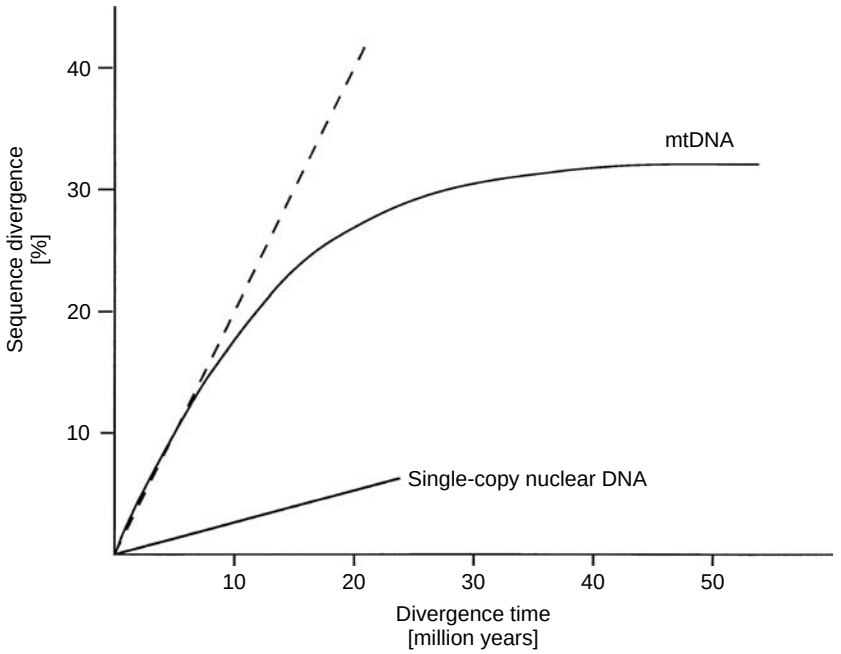

Fig. 11 Rates of nucleotide substitution in single-copy nuclear DNA and mtDNA, plotted with the theoretical rate without saturation effects in mtDNA (dashed line). [After Brown et al. (1979) in Hoelzel and Dover (1991).]

mtDNA is more labile to extreme fluctuations in population size than nuclear DNA, since a single mating pair will pass on four copies of the nuclear genome but only one copy of the mitochondrial genome. Also, the local extinction of matrilines can limit the diversity of mtDNA, especially in small populations. Second, the geographic distribution of mitochondrial haplotypes will depend on the dispersal behavior of females. If females are less likely to disperse (as is commonly the case for mammals), then mtDNA variation may show greater geographic structuring than nuclear DNA variation. Interpretation can also be complicated by the existence of more than one form of the mitochondrial genome in an individual, but this condition (heteroplasmy) is relatively rare.

5.4 Chloroplast DNA

Sequence variation in the chloroplast genome (cpDNA) is highly conserved, as is variation in the size of the genome for most taxa, though the range for all plant species is large (120–217 kb). Inheritance can be biparental, strictly maternal, or strictly paternal depending on the species. Most population studies have found very low levels of cpDNA variation, though there are exceptions. The utility of this genome as a population genetic marker will depend on the mode of transmission and level of variation for a given taxa. Strict paternal or maternal transmission may provide useful information even if the relative level of variation is low.

5.5 Minisatellite DNA

There is an abundance of repetitive "satellite" DNA in the nuclear genome of most complex organisms. These regions are classified according to their composition and structure. Minisatellite arrays are nested repeats of a highly conserved core sequence, 15 to 40 bp long, interspersed on most chromosomes. A minisatellite array can be up to 20 to 25 kb long. A high rate of unequal crossing-over generates a high level of length variation in these arrays. The "DNA fingerprinting" method took advantage of the conserved minisatellite core sequence and the large number of loci sharing that core repeat. A probe for the core sequence was used in a multilocus RFLP analysis, which reveals a ladder of hypervariable bands that could identify an individual or determine paternity with a very small chance of error (Fig. 12). While DNA fingerprinting is a powerful

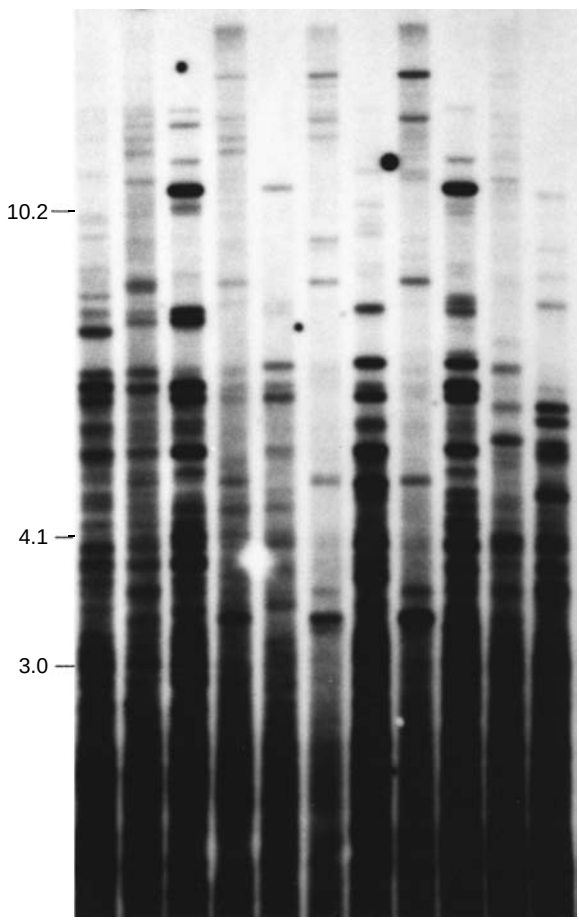

Fig. 12 Minisatellite DNA variation (DNA fingerprints). DNA from 11 southern elephant seals was digested with *Hae*III and probed with (CAC)*n*. Fragment sizes shown in kilobases.

tool for the assessment of close kinship, the interpretation of variation in this marker at the population level is not straightforward. One problem is that variation saturates such that variation within a population is often as great as variation between populations. This marker is typically more useful to a population geneticist as a tool to reveal reproductive strategy, which in turn has an important effect on the genetic structure of a population.

Exceptions to this general rule exist, especially when genetic variation for a population or species has been reduced. In this case, variation for the multilocus pattern must be statistically treated as a phenotype, since the allelic relationship between bands in the fingerprint cannot be determined in most cases. Typically, a relative measure of "band-sharing" is used as an index of genetic relatedness. An alternative is to clone an individual locus and use that DNA as a probe. Single-locus minisatellite length variation can be treated as a Mendelian character, although interpretation of levels and patterns of variation should be tempered by the understanding that various mechanisms (e.g. unequal crossing-over and slippage) may promote change at different rates. It should not be assumed that these rates are constant over the relevant time frames for population-level comparisons.

5.6 Microsatellite DNA

Microsatellites are small arrays (typically about 100 bp) of simple di- and trinucleotide repeats (longer repeat elements are less common). These arrays vary in length over time because of DNA slippage (Fig. 1). The rate of change varies considerably among loci, but some are hypervariable like minisatellite loci. They are visualized by designing oligonucleotide primers on either side of the array and amplification by PCR (see Sect. 4.6). A radioactive or fluorescent label is incorporated during the amplification, typically by labeling one of the primers. The amplification products are then run on a long polyacrylamide gel (to allow single base-pair resolution), and each individual should be represented by one (homozygote) or two (heterozygote) bands (Fig. 13). Polymerase slippage during amplification usually creates artifact bands. Screening for microsatellite DNA variation is often done using fluorescent labels and a DNA sequencing machine (see Sect. 4.5). This

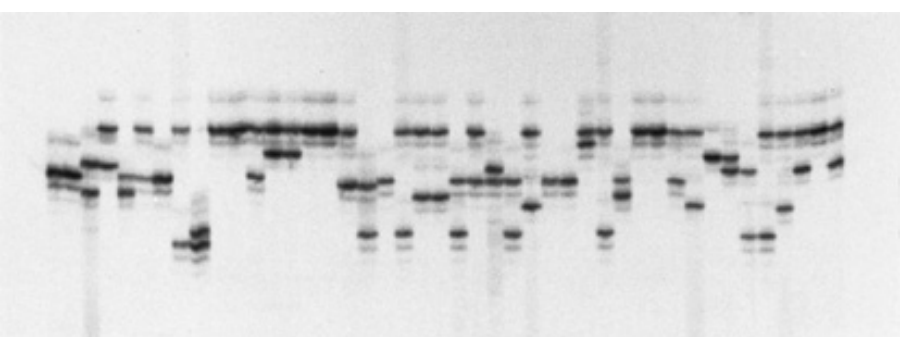

Fig. 13 Microsatellite DNA variation for 45 southern elephant seals.

allows the simultaneous running of multiple loci in a single lane (using multiple dyes) and computer-generated genotype data that can be precisely sized and compared using internal lane size standards.

Microsatellite DNA is especially useful for kinship assessment and paternity analyses, because a single locus is amplified at a time, and therefore Mendelian expectations and interpretation can be tested and applied (unlike multilocus DNA fingerprinting; see Sect. 5.5). Application at the population level is extremely common, though there are some important considerations for interpretation. These include recognition of the broad range of evolutionary rates among loci and the potential for rates to change over evolutionary time (e.g. with the accumulation of a mutation that interrupts the repetitive array and therefore the process of DNA slippage). However, the main consideration is the evolutionary model. Genetic distance measures are often assessed under the assumption of "infinite alleles," which means that a new mutation should be considered novel. Given sufficient evolutionary time, this assumption is violated for sequence data (and distance measures have been derived to account for this; see Sect. 6 and Fig. 11), since there are only four nucleotides in the genetic code. Over time "back mutations" occur, which can lead to the underestimation of evolutionary change (see Sect. 6). This is likely to happen more frequently with microsatellite DNA loci, both because of the high rate of change and because of the nature of the change. Different alleles represent different numbers of repeats in the array, and the process of slippage can either add or remove repeats over a finite range for a given locus. Therefore, back mutation is quite likely. Interpretation of genetic distance at this type of locus therefore requires a different evolutionary model, referred to as the *stepwise mutation model*, whereby evolution is assumed to occur in steps.

6 Statistical Analysis of Variation

A full mathematical description of the statistical methods for assessing and comparing levels of variation within and between populations is beyond the scope of this article (for reviews see Nei 1987 and Hartl 2000). However, I will briefly describe a few of the general ideas and some of the assumptions on which these analyses are based. For genetic changes that are selectively neutral, the accumulation of change can be interpreted to estimate time-dependent differentiation within and between populations. For the analysis of DNA, a statistical interpretation is generally based on the following assumptions: (1) that nucleotides are distributed randomly in the genome, (2) that variation arises solely by base substitution, (3) that substitution rates are the same for all nucleotides, and (4) that all relevant bands on a gel can be detected, and similar bands are not scored as identical.

The fourth condition can be controlled by careful methodology. The first three conditions do not strictly hold, but Nei (1987) has argued that small deviations will not significantly alter the results. Also, since there are only four bases in the genetic code, there is a certain probability that mutations will revert (e.g. from A to T then back to A). Formulations have been derived to compensate for this effect (see review in Nei 1987). Allozyme variation is measured phenotypically (as the electrophoretic mobility of proteins). Therefore, it is especially important to establish that these characters follow the Mendelian rules of segregation

during sexual recombination. This is usually done by conducting pedigrees or, when this is not practical, by testing for correspondence to the Hardy–Weinberg rule (which states that the gene and genotype frequencies will remain constant between generations in an infinitely large, random mating population). This rule is expressed as follows:

$$p^2 + 2pq + q^2 = 1$$

where p and q are allele frequencies in a two-allele system, and p^2, $2pq$ and q^2 represent the expected frequencies of the A_1A_1, A_1A_2, and A_2A_2 genotypes respectively. Deviations from the expected Hardy–Weinberg proportions of genotypes could indicate the action of selection or inbreeding.

At equilibrium, the amount of variation in a population depends on two factors: the effective population size (N_e) and the mutation rate (μ). This is typically summarized using the parameter θ:

$$\theta = 4N_e\mu$$

Estimates of θ can be derived from various measures of diversity including heterozygosity, nucleotide diversity, and gene diversity. Average heterozygosity is simply the number of heterozygous individuals (having two different alleles at a locus) divided by the total number of individuals. In Hardy–Weinberg equilibrium, the expected number of heterozygotes for a locus with two alleles is $2pq$ (see above). When there are more than two alleles (as is typically the case), we can calculate homozygosity (having the same two alleles at a locus) by summing the squared frequency for each allele, and subtract this total from 1 to derive expected heterozygosity (since 1 minus the homozygosity equals heterozygosity). Nucleotide diversity is a measure of the mean sequence divergence among alleles, while gene diversity is a measure of the number of different alleles in a population.

A common way to compare the level of diversity among and within populations is by using the inbreeding coefficients proposed by Sewall Wright in 1969, called *F-statistics*. The inbreeding coefficient can be defined as the probability that a pair of alleles in an inbred individual is identical by descent. Within a population, we can define F_{IS} as a measure of the reduction in heterozygosity (H) in an individual due to nonrandom mating ($F_{IS} = (H_S - H_I)/H_S$), where H_S is the expected heterozygosity for the population (based on Hardy – Weinberg expectations) and H_I is the average heterozygosity of all genes in an individual. We can use the same reasoning to assess genetic structure among populations. In this case, $F_{ST} = (H_T - H_S)/H_T$, where H_T is the expected heterozygosity for the entire sample (all populations combined) and H_S is the average H_S among all individual populations. It is essentially a measure of the proportion of genetic variance that is explained by differences among as opposed to within populations, and it varies over a range from 0 (no differentiation among populations) to 1 (completely isolated populations).

The application of these and similar methods allow estimates of gene flow, the time since populations diverged, the relative effective size among populations, and variation in mutation rates for a given gene among populations. The application of statistical methods such as Bayesian inference and autocorrelation can also help interpret patterns of population structure and mixing, though further detail is beyond the scope of this review. Additional

methods based on likelihood estimates and measures of genetic similarity can assign individuals to populations on the basis of their genotype (typically using microsatellite DNA loci) and can assess the dispersal behavior of different classes of individuals (e.g. males vs females).

Another important method reflects the way in which we think about the relationship among genotypes. Through phylogenetic reconstructions, we can explore possible historical relationships among taxa, and this has been very useful at the population level (often referred to as *phylogeography*). It also allows us to think about the ancestral relationship between pairs of genotypes. If we think backwards in time, we can consider the time when two evolutionary paths come together, at the point of coalescence. This represents the point at which these genes shared a most recent common ancestor and the time that has elapsed since then (the coalescence time). The pattern of coalescence is affected by the processes of demographic change and natural selection, and therefore allows inference about these processes. The utility of this approach is greatly facilitated by DNA-sequencing nonrecombining genomes (such as mitochondrial DNA), which allows us to trace lineages through the comparison of variation among homologous sequences at the level of the nucleotide.

7 Conclusions

Molecular methods have made it possible to test theoretical expectations about genetic diversity directly and have greatly facilitated our understanding of populations at the genetic level. Every year, there are significant advances that make this work ever more powerful and more routine. This has especially included advances in the computer-based screening of gels for DNA sequences and genotypes, and the software developed to analyze such data. There have also been exciting advances in the methods of interpretation, some of which are described above. In future, technology that allows us to further increase the number of sequences we analyze and the number of genotypes we can screen will allow greater insight into the evolution of population structure and diversity. These data allow us to understand evolutionary processes at a fundamental level and help us protect diversity in natural populations and thereby species for future generations.

See also Genetics, Molecular Basis of; Genetic Variation and Molecular Evolution.

Bibliography

Hoelzel, A.R., Dover, G.A. (1991) *Molecular Genetic Ecology*, Oxford University Press, Oxford.

Hartl, D.L. (2000) *A Primer of Population Genetics*, 3rd edition, Sinauer Associates, Sunderland, MA.

Hillis, D.M., Moritz, C., Mable, B.K. (Eds.) (1996) *Molecular Systematics*, 2nd edition, Sinauer Associates, Sunderland, MA.

Hoelzel, A.R. (Ed.) (1998) *Molecular Genetic Analysis of Populations, A Practical Approach*, 2nd edition, Oxford University Press, Oxford.

Maynard Smith, J. (1998) *Evolutionary Genetics*, 2nd edition, Oxford University Press, Oxford.

Nei, M. (1987) *Molecular Evolutionary Genetics*, Columbia University Press, New York.

Wright, S. (1969) *Evolution and the Genetics of Populations. The Theory of Gene Frequencies*, University of Chicago Press, Chicago.

Genetic Engineering of Antibody Molecules: *see* Antibody Molecules, Genetic Engineering of

Genetic Engineering of Hemoglobin

Jeremy Tame
Yokohama City University, Yokohama, Japan

1 Introduction 267

2 Expression Systems 270

3 Distal Residues 272
3.1 Oxygen Equilibrium Curves 273
3.2 Rates of Ligand Binding 274
3.3 Discrimination between Ligands 276

4 Proximal Histidine Mutants 277

5 C7 Mutants 278

6 N-terminal Mutants 278

7 Trp37(C3)β 279

8 Blood Substitutes 280
8.1 Nitric Oxide 281
8.2 96α Val$\rightarrow$Trp 283

9 Animal Hemoglobins 283
9.1 Geese 283
9.2 Fish 284
9.3 Crocodiles 285

10 Evolution 286

11 Nuclear Magnetic Resonance Spectroscopy (NMR) 286

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

12 **Summary 288**

Bibliography 288
Books and Reviews 288
Primary Literature 289

Keywords

Heme
Protoporphyrin IX, the ring molecule holding an iron atom to which oxygen can bind.

Distal
The side or face of a heme molecule within Hb, which is free to bind oxygen.

Proximal
The side or face of the heme that is covalently linked to a histidine side chain.

Allostery
The phenomenon of a protein changing shape on binding ligands or substrates.

Cooperativity
The phenomenon of multiple ligand binding showing an increase (positive cooperativity) or decrease (negative cooperativity) in ligand affinity with increasing ligation.

p_{50}
The partial pressure of oxygen, usually measured in mmHg, needed to half-saturate a sample of Hb.

Hemoglobin is the oxygen-carrying protein of the blood, binding oxygen at iron-containing heme groups within the protein. Its structure is well characterized in both the deoxy and oxy forms, and comparison of these structures has shown how the protein changes its conformation as the number of bound oxygen molecules increases. These conformations have different ligand affinities, and so the protein can regulate how tightly oxygen is bound. A detailed understanding of the protein mechanism and control of ligand affinity requires the study of mutant proteins. Many mutated hemoglobins have been detected clinically, but these naturally occurring mutants have generally been of limited use in analyzing hemoglobin function. A number of expression systems have therefore been devised to produce human hemoglobin from microorganisms. These systems make possible the alteration of

specific residues or groups of residues so that their functional role may be tested. Together with the wide variety of biophysical techniques employed in the study of hemoglobin, genetic engineering has proved to be a powerful tool for testing the function of different parts of the protein. Hemoglobin remains one of the most important systems for testing ideas concerning protein folding and function, protein interactions, and cooperativity. It is far from being completely understood, despite the widespread impression that this is so. The last decade has seen an explosion in hemoglobin research triggered by the ability to create artificial mutants of hemoglobin, and the increasing need for a safe, cheap blood substitute. This review attempts to highlight some of the advances made in recent years.

1 Introduction

Myoglobin (Mb) and hemoglobin (Hb) were the first proteins whose structures were determined by X-ray crystallography, the analysis of X-rays scattered from the protein in crystalline form. Both Mb and Hb form highly ordered crystals, which allow fine details of the protein to be resolved. Human carbonmonoxy hemoglobin has now been refined to 1.25-Å resolution (Protein Data Bank entry 1IRD (see the Protein Data Bank home page http://www.rcsb.org/pdb/)). Mb is an oxygen storage protein of the muscle. It consists of a polypeptide chain wrapped around a heme group, a ring molecule with an iron atom at its center, which gives the molecule its red color (Fig. 1). The polypeptide is coiled into eight α-helices (labeled alphabetically from the N-terminus) packed together to give a compact, roughly spherical shape. The heme is buried in a cleft in the protein between the E- and F-helices, and the iron atom is covalently attached to a histidine residue on the F-helix. Since this is the eighth residue from the N-terminus of the helix, it is known as *residue F8* as shown in Fig. 2. Oxygen binds to the opposite side of the heme, the "distal" side, in a ligand pocket. Two amino acid residues, histidine E7 and valine E11, touch the ligand in oxy-Mb. Both these residues are highly conserved throughout globin evolution. Histidine E7 is called the *distal histidine* to distinguish it from the "proximal histidine" F8. Mb binds to oxygen in a simple fashion with a single, fixed equilibrium constant.

As the oxygen-carrying protein of the blood, Hb has the conflicting task of

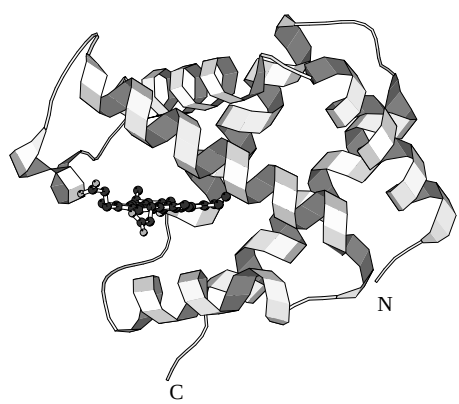

Fig. 1 A schematic diagram showing the highly helical tertiary structure of α globin. The polypeptide is shown as a ribbon trace of the Cα atoms, and the heme molecule is shown as individual atoms.

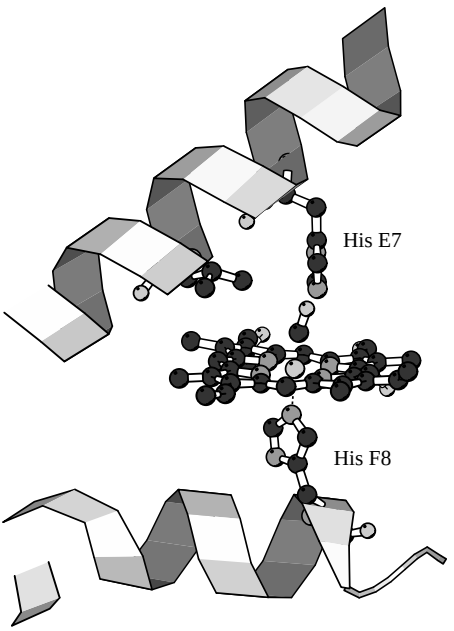

Fig. 2 A closer view of the heme region of β-globin, showing carbon monoxide bound to the heme iron. The heme is sandwiched between the E- and F-helices, and is covalently attached to the F-helix through the side chain of the histidine F8 residue. The E7 residue is also a histidine, which contacts ligands bound to the distal side of the heme.

binding oxygen tightly in the lungs and releasing it readily in the tissues. A simple oxygen-binding protein such as Mb would be extremely inefficient at transporting oxygen in this way. Hb changes its oxygen affinity with oxygen partial pressure, lowering it sufficiently to release oxygen around the body at relatively high oxygen concentrations. The means by which this is accomplished remained a puzzle until the structure was solved in both liganded and unliganded forms by Perutz and coworkers. The molecule consists of four Mb-like subunits, two α-type and two β-type, each containing a heme group. The four subunits form two $\alpha\beta$ dimers, which adopt slightly different positions relative to each other in the liganded and unliganded forms. Each $\alpha\beta$ dimer is relatively rigid. In the deoxy form, the protein constrains the iron atom at the center of each heme group, pulling it toward the proximal side. This form, which has more bonding interactions between the $\alpha\beta$ dimers than the oxy form and is energetically favored in the absence of ligand, is known as the *T*-(*tense*)*state*. Ligand binding at any heme serves to weaken the dimer–dimer interaction, and as more ligands bind to the protein, it favors the R-(relaxed)conformation more. In this form, the subunits have a higher oxygen affinity because the constraints on the movement of the iron atom are no longer there. Several allosteric effectors are known, which shift the equilibrium toward the T-state. These include chloride ions, hydrogen ions (above pH 6) and organic phosphates. Human red cells contain 2,3 diphosphoglycerate (DPG), which lowers the oxygen affinity of Hb substantially. Inositol hexaphosphate (IHP) is a nonphysiological ligand, which exerts an even stronger effect.

Researchers in the hemoglobin field are often asked why they work on a protein that is "solved." This is a natural question; it is now more than 30 years since Perutz put forward his famous "trigger" hypothesis that oxygen binding to the heme iron atoms pulls on the proximal histidine, breaking salt bridges holding the tetramer in the T-conformation and allowing the protein to relax to the R-state. In the Perutz model, the protein makes a concerted switch between two possible quaternary states. This idea has been subjected to numerous tests and is found to be in very good agreement with the experimental data. It is the classic example

of protein allostery found in biochemistry textbooks, and is unquestioningly accepted by generations of undergraduates. It cannot, however, be the whole truth since the intermediates on the pathway between the T- and R-states cannot have infinite energy (though at least one paper has suggested they might). The nature of these intermediates, and their energy relative to the deoxy T- and oxy R-states, has been a topic of heated debate for decades and is not yet fully resolved. The genetic engineering of hemoglobin opened up an enormous array of possible experiments through which the limits of the Perutz model could be tested and our understanding of hemoglobin expanded.

It should also be realized that the data with which Perutz worked 30 years ago were of a different standard when compared to X-ray data collected today. X-ray generators were very weak, and the film used to detect scattered X rays was not sensitive enough for accurate measurements of weak reflections. Today's synchrotron X-ray sources are many orders of magnitude brighter than simple laboratory generators, and image plate and CCD detectors allow the very rapid acquisition of accurate data to high resolution from small protein crystals. As crystallographic hardware and software have improved, it has become possible to refine Hb structures to higher resolution, even allowing individual atoms to be visualized in the best cases. The best current refinement of oxyHb was performed in 1983. Although at a nominal resolution of 2.1 Å, the geometry and temperature factors of the model clearly show that the data and software quality that existed 20 years ago were much inferior to those of today. Crystallographic laymen should not be fooled into thinking that the resolution or the crystal "R-factor" are uniformly reliable indices of model quality. Since the methods of data collection and refinement can also affect the final model, it is not too surprising to find that crystal structures of different Hbs solved by the same research group at the same time are more similar in some ways than models of native Hb from different groups.

Furthermore, it should be emphasized that crystallographic structures of even the highest quality cannot answer all questions concerning Hb binding to oxygen. Crystallography can tell us where atoms are, but not necessarily what they do, or how much energy is stored in a particular atomic interaction. In order to understand the relationship between the protein structure and its function, it is necessary to measure the affinity and kinetics of Hb for particular ligands, under different conditions. For example, it was noticed that the positions of the distal histidine and valine side chains are different in the α- and β-subunits, and appeared to have different functional roles. This observation provided the stimulus for the engineering of mutant proteins, but without the functional experiments the crystallographic hypotheses cannot be tested. Even when only clinically detected mutant Hbs were available, several of these played an important role in helping Perutz formulate his ideas. The ability to produce any desired mutant on demand allows us to investigate aspects of Hb function that are still incompletely understood, such as the Bohr effect, the change in oxygen affinity with pH. Functional analysis of such mutants complements the crystallographic work and provides a deeper insight into the protein mechanism. More recently, the production of Hb in microorganisms has allowed the protein to be labeled isotopically, greatly increasing the usefulness of nuclear magnetic resonance (NMR) in addressing structural and functional issues of such

a large protein. This review focuses on the genetic engineering of hemoglobin; metal-substituted hemoglobins, chemically modified Hbs, and native animal Hbs fall outside its scope. The reader should remember, however, that our understanding of this protein has come from a wide range of biophysical techniques applied over many years. Indeed, hemoglobin has been a principal protein used to hone many of those techniques, now widely applied elsewhere.

2 Expression Systems

The first requirement for a high yield of the protein of interest is a strong promoter that directs the transcription of the gene to form messenger RNA (mRNA). The promoters used are generally inducible. The first system successfully used to produce usable quantities of human globins in a microorganism utilized an *Escherichia coli* plasmid carrying the λ P_L promoter, which is switched off with λ-repressor. At that time, in the mid 1980s, this was the expression system of choice; a temperature-sensitive mutant of the λ-repressor was used to prevent transcription in the early growth phase. Gene expression was switched on by heat-shock at 42 °C to inactivate the repressor. The synthesis of large amounts of protein also depends on efficient translation of the mRNA. The expression of eukaryotic genes in *E. coli* is often difficult due to differences in codon usage and the mechanism of translation initiation. In *E. coli*, ribosome binding to mRNA involves base pairing of the 16S rRNA (part of the 30S subunit) to the Shine–Dalgarno region, about nine bases upstream of the initiator AUG codon. Secondary structure of the mRNA around the initiation site can have a marked inhibitory effect on translation. Indeed, if such a secondary structure is present, it may be removed only by altering the codons corresponding to the N-terminus of the protein to be expressed. Many eukaryotic genes, α- and β-globin among them, are very poorly expressed by simple expression systems. To overcome this problem, Nagai and Thøgersen developed the cleavable fusion protein expression vector using the blood clotting factor FXa. A leader sequence from a highly expressed protein is coupled to a short sequence encoding the tetrapeptide recognition site of FXa (Ile-Glu-Gly-Arg) and then the coding sequence of interest. The fusion protein produced in high yield can then be treated with FXa to release the C-terminal region corresponding to the desired polypeptide product. This method is now used routinely with histidine-tagged proteins, but β-globin was one of the first proteins to be expressed as a cleavable fusion protein. α-globin proved more difficult to express with this system, possibly due to the structure of the mRNA around the initiation codon. A key advantage of this system is that it produces mammalian globins with the correct N-terminal amino acids, with the initiator methionine removed. *Escherichia coli* is relatively inefficient at removing this residue if the second residue is valine or serine, and the extra methionine can have a marked effect on the functional properties of the protein. The main drawback, however, is that the fusion protein method produced inclusion bodies in *E. coli*, and requires the globin to be dissolved in urea or guanidine hydrochloride before cutting with FXa and refolding with heme. Overall, the cleavage/refolding procedure has an efficiency of about 25% in general, but can be lower if unstable mutants are made, so considerable effort is required to make

these mutants on the 50 to 100 mg scale. The method is, however, capable of synthesizing Hb, which is structurally and functionally identical to native human Hb. The report by Nagai, Poyart, and Perutz on the oxygen-binding properties of this protein is the first paper demonstrating the production of functional Hb using a bacterial host.

A completely different expression vector, described by Hoffman and colleagues in 1990, contains coding sequences for both α- and β-globin in a single operon, the two being linked by a "translational coupler." Ribosomes produce both types of globin from a single mRNA. Coexpression of α- and β-globin in this way within the cell results in the production of folded, fully functional tetrameric Hb. *Escherichia coli* increases production of heme to such an extent that the cells turn red. The globin genes were constructed from synthetic oligonucleotides to optimize codon usage, and transcription is driven by an IPTG (isopropylthiogalactoside) inducible promoter. Interestingly, the expression of β-globin alone gave mainly insoluble protein with no heme incorporated, so the coexpression of α- and β-globin appears to be necessary for the protein to fold. Hb produced in this way shows 4 hemes per tetramer and less than 5% oxidation of the hemes. The visible spectrum is identical to that of native HbA. However, amino acid sequencing shows that the β-subunits, and 90% of the α-subunits, retain the initiator methionine. Although the oxygen affinity and cooperativity of the protein are very close to those of native HbA, the Bohr effect is roughly 50% less, reflecting the role of the N-termini in mediating the proton dependence of the oxygen affinity. The binding of inositol hexaphosphate (IHP) to the protein also appears to be weakened by the extra methionine residues.

Hernan and colleagues investigated the importance of optimizing the codon usage in the expression of α- and β-genes in *E. coli*. Using a T7 expression system, they reported the first successful attempt to express β-globin from its cDNA as a nonfusion protein. In common with Hoffman et al., these researchers found that expression of β-globin alone led to the production of large amounts of insoluble protein within the cell. Attempts to express α-globin alone were unsuccessful, indicating that either the mRNA or the protein is less stable than its β-globin counterpart. The expression of β-globin to a high level (as an inclusion body) allowed the protein to be refolded with native α-globin using the procedure developed by Nagai, but without FXa cleavage. Amino acid sequencing showed retention by the recombinant protein of the deformylated N-terminal methionine residue. This system, therefore, produces Hb tetramers identical to HbA except that the β-subunits have an extra N-terminal residue.

Clearly, the principal problem in synthesizing Hb in *E. coli* is the inefficient removal of the N-terminal methionine from the protein. Two approaches have been used to overcome this difficulty. Firstly, Wagenbach and colleagues used the yeast *Saccharomyces cerevisiae* as the host organism. Coexpressing α- and β-globin genes in yeast yielded folded, tetrameric Hb with correctly processed N-termini, showing the different preferences of eukaryotic and prokaryotic methionyl aminopeptidases. Ho and colleagues used a different approach and chose to overexpress the native methionyl aminopeptidase together with globin genes. This method also produces authentic human Hb. The pHE2 plasmid made by the Ho group is now the most widely used expression system, and has led

to an enormous increase in the number of publications on artificial Hb mutants. Any system in which the host microorganism yields folded tetrameric Hb with heme incorporated is significantly easier to work with than the original Nagai system. Concerns have been raised, however, regarding the folded Hb produced in *E. coli*, which is reported by some authors to be slightly more prone to oxidation than HbA. Refolding the chains from denaturant produces protein with spectral and functional characteristics essentially identical to those of native HbA, but such systems cannot be used to make protein on a large scale without considerable effort.

Although strictly outside the scope of this review, there is also a considerable body of research on recombinant Mb. This is largely based on work from the Sligar group, which succeeded in producing folded, functional Mb within *E. coli*. With greater difficulty, the FXa fusion method can also be used to make mutant Mbs. Mb has served as a very useful model system, especially in testing ideas concerning the heme pocket. Since it is much simpler than Hb, having only a single subunit, it is rather easier to work with. Functional and crystallographic analyses of mutant Mbs, however, have played an important role in shaping ideas concerning ligand binding and selectivity of Hb.

3 Distal Residues

Two very highly conserved residues in the distal pocket lie in van der Waals contact with ligand molecules, a histidine (E7) and a valine (E11). In deoxy Hb, the distal valine of the β-subunit lies over the heme and obstructs oxygen binding. The equivalent residue in the α-subunit is further from the iron, and oxygen can seemingly be accommodated more readily. It was therefore predicted that ligands will bind to the α-subunits first. Kinetic experiments have clearly shown, however, that the very bulky *t*-butyl isocyanide binds to the β-subunits faster than the α-subunits, a result that underlines the difficulties in predicting function from structure. Further differences in the distal residues were seen in the high-resolution structure of oxy-Hb. In the α-subunits, a hydrogen bond was found between the oxygen ligand and the distal histidine. Since hydrogen atoms are very weak X-ray scatterers, hydrogen bonding in X-ray crystallographic models of proteins is generally inferred from the positions of other atoms. A similar hydrogen bond has been found in oxy-Mb, and demonstrated by neutron scattering. Strangely, no such hydrogen bond appeared to form in the β-subunit.

The interactions between the distal residues and bound ligands, proposed on the basis of crystallographic structures, could be tested only by obtaining suitable mutant Hbs in which the distal residues of the α- and β- subunits were replaced with other amino acids. To examine the proposed steric hindrance of the valine E11 residue, a number of different mutants are needed in which this amino acid is replaced with larger or smaller apolar residues. If the distal valine side chain does sterically hinder oxygen binding, these mutations should cause the oxygen affinity to change. Similarly, replacing the distal histidine residue with apolar residues will exclude the possibility of a hydrogen bond between the E7 residue and bound oxygen. If this bond were to occur in the native protein, the oxygen affinity should fall.

Aspartate, glutamate, and alanine have all been found at the E11β position in

naturally occurring Hbs. The first two of these mutants are unstable, since the introduction of a negative charge into the heme pocket greatly increases the rate of oxidation of the heme. Hb Sydney, $\alpha_2\beta$(E11 Ala)$_2$, is thermally unstable because the weakened heme–protein contact allows the heme to dissociate from the protein more easily. This mutant is potentially very interesting since the amino acid change is quite conservative. However, patients producing Hb Sydney also produce HbA, since humans carry two copies of the β-globin gene, and Hb Sydney cannot be separated from HbA since the two show identical electrophoretic behavior. Nevertheless, mixtures of Hb Sydney and HbA purified from hemolysate were shown to have increased oxygen affinity in the T-state. This experiment was one of the first in which it was shown that even conservative mutations can have important functional effects if they occur at key residues in the protein. To determine the behavior of the mutant more precisely, however, requires the synthesis of an artificial mutant.

3.1 Oxygen Equilibrium Curves

Since the principal function of Hb is to transport oxygen around the body, the majority of mutational studies of Hb have focused on the effects of amino replacements on the oxygen affinity of the protein. These measurements are carried out using an oxygen cell, ideally measuring spectral changes for both the oxygenation and deoxygenation curves to ensure that these are identical. These experiments show the well-known sigmoidal curves due to the increase in oxygen affinity with increasing oxygenation. The oxygenation was first studied in detail by Hill, who used a logarithmic plot (now named after him) to chart the oxygenation of the protein with increasing oxygen partial pressure. The slope of the curve is known as n, the cooperativity – the degree to which the ligation state of the hemes influences the ligand affinity of its neighbors. n_{max} is a very sensitive test of the quality of a sample of HbA. An important test for any expression system, therefore, is its ability to produce wild-type HbA with essentially identical oxygenation properties to those of HbA.

The Nagai and Thøgersen FXa fusion protein expression system led to the production of a number of mutant Hbs with altered distal residues in the β-subunit. Valine E11 was changed to alanine, leucine, and isoleucine, and histidine E7 was changed to glutamine. Upon comparison of the oxygen equilibrium curves of HbA and Hb (Val E11$\beta\rightarrow$Ala), it was found that the T-state oxygen affinity of the mutant protein is higher, but the R-state affinity is unaltered. Thus, the distal valine side chain does sterically hinder ligand binding, as suggested by the crystal structure, but only in the T-state. By determining the structure of deoxy Hb held in the R-state by a chemical cross-link, it has been shown that the β-chain distal valine moves only 0.3 Å from its position in T-state deoxy Hb, but it moves a further 0.7 Å in oxy-R state Hb. It seems that the smaller movement associated with the change of quaternary structure imposes greater strain on the ligand than the larger movement associated with ligation of the R-state.

The oxygen equilibrium curve of Hb Val E11$\beta\rightarrow$Leu is superimposable over that of HbA, the increase in size of the E11 residue having no effect on the oxygen affinity in either quaternary state. Hb (Val E11$\beta\rightarrow$Ile), however, shows a markedly reduced oxygen affinity throughout its

oxygen equilibrium curve but retains its cooperativity. The different properties of these proteins appear to be related to the flexibility of the E11 residue as much as to its size. Both valine and isoleucine are forked at the Cβ atom, which prevents rotation about the Cα–Cβ bond when these residues sit within an α-helix. In the crystal structure of Hb (Val E11$\beta \rightarrow$ Ile), the δ-methyl group of the isoleucine E11 sits over the heme iron atom, pushing histidine E7 away and repelling ligands. Leucine E11 does not cover the ligand binding site or press against the distal histidine. It was clearly demonstrated that nonbonded interactions with the distal valine could reduce ligand affinity, and that the size and conformation of the E11β residue are important determinants of its ability to hinder ligand binding. Three α-chain distal valine mutants (valine $\rightarrow$ alanine, leucine, and isoleucine) have been made with a modified form of the Nagai and Thøgersen system. Their oxygen equilibrium curves are superimposable on that of HbA. The contrast in behavior between the two subunit types is exactly that predicted from the crystallographic structure of native HbA; the E11 side chain of the α-chain clearly exerts no direct control over the oxygen affinity in the way the equivalent β-globin residue does.

In deoxy Hb, a water molecule is found within the heme pocket of the α-subunits but not the β-subunits. This water is hydrogen bonded to the distal histidine and lies close to the position the second oxygen atom occupies in oxy-HbA. It was not clear from the structural data, however, whether this water molecule reduces the affinity for heme ligands by competing with them for occupancy of the heme pocket. The structures alone suggested no significant functional role for the water molecule, but a different conclusion was reached from the association rates of a number of ligands binding to mutant Mbs – the distal histidine slows down the rate of ligand binding by holding a water molecule in the heme pocket with a hydrogen bond. The role of this water molecule was further complicated by the finding that in cryo-cooled crystals of deoxy human Hb, one α-subunit has the water molecule while the other does not.

Hbs having the distal histidine of either the α- or β-subunits replaced by glycine were too unstable to permit the accurate measurement of the oxygen equilibrium curves, the heme rapidly oxidizing to the ferric state. The histidine was also replaced by glutamine, an α-apolar amino acid capable of donating hydrogen bonds. This mutation is found in the α-subunits of opossum Hb, but all other vertebrate Hb sequences determined have histidine at the E7 position. In the case of the β-subunit mutant, the replacement led to a rise in the T-state oxygen affinity of the protein, but the R-state affinity was unchanged. The electron density map of this mutant in the deoxy form showed the E7 glutamine residue to be well localized, but the Nε and Oε atoms of the side chain could not be distinguished. The higher T-state oxygen affinity was attributed to the relief of steric hindrance due to the greater flexibility of the glutamine compared to histidine.

3.2
Rates of Ligand Binding

More detailed information regarding the function of the distal residues has been obtained from the kinetic studies of ligand binding in which the R-state rate parameters of the mutant Hbs binding to oxygen and carbon monoxide were determined. The group of Prof. J.S. Olson in Houston, in particular, has studied the functional

effects of mutations in Mb and Hb in great detail. For Hb, the largest effects were found when histidine E7α was replaced by glycine. Although oxygen equilibrium experiments could not be carried out on this unstable mutant, it was possible to measure the association and dissociation constants of oxygen binding to R-state protein and to calculate the oxygen affinity ($K = k_{on}/k_{off}$). In native globins, the distal histidine residue blocks ligand access to the heme from the solvent, and must swing away from the heme pocket to allow ligands to enter or leave. Removing the imidazole side chain creates a channel to the heme iron so that diatomic ligands can enter the heme pocket unhindered. In α-globin, this change allows both CO and oxygen to bind to the protein about seven times faster. The dissociation rate of CO is not greatly affected, but the dissociation of oxygen increased 60-fold. Replacing histidine E7α by glutamine gave similar but smaller effects, and the oxygen affinity is considerably decreased by either mutation. When the native protein is compared with Hb (His E7$\alpha \rightarrow$ Gly), the distal histidine appears to stabilize bound oxygen by about 1 kcal mol^{-1} and to destabilize bound CO by about the same amount.

The β-subunits of Hb showed quite different functional behavior when the distal histidine residue was changed to other amino acids. Mutating histidine E7β to glycine, glutamine, or phenylalanine has remarkably little effect on the rates of oxygen or carbon monoxide binding. Each of the mutant β-subunits binds to these ligands with rates indistinguishable from those of the native subunit, despite the considerable differences in the nature of the E7 residues. Histidine E7β is not, therefore, an effective barrier to ligand binding, unlike its counterpart in the α-subunits, which is consistent with the relatively high association rates observed for the native protein. The dissociation rates are also insensitive to changes in the E7 side chain. This agrees with the 1983 crystallographic structure of oxy-Hb, which shows histidine E7β to be too far away and in an unsuitable position to form a strong hydrogen bond with bound oxygen. Replacing the distal histidine of Mb with phenylalanine reduces the oxygen affinity 130-fold, yet the same mutation has no effect on the R-state oxygen affinity of β-globin. Nevertheless, NMR results from the Ho group suggest the distal histidine of both subunits in HbA can hydrogen bond to oxygen in the binding site.

From the oxygen equilibrium curves, it is clear that replacing the α-subunit distal valine with other aliphatic apolar residues has no effect on the oxygen affinity. When valine is changed to leucine or isoleucine, the R-state oxygen association and dissociation rates remain the same, but mutating the E11 residue to alanine causes both rates to increase roughly sevenfold. This mutation also increases the association rate of carbon monoxide 10-fold, but does not alter the dissociation rate. These increases in association rates show that Val E11α must be a kinetic barrier to diatomic ligands binding. α globin (Val E11$\alpha \rightarrow$ leucine) shows CO rate parameters indistinguishable from those of the native protein; α globin (Val E11$\alpha \rightarrow$ leucine) seems to bind CO only half as quickly, but the rate is the same as that of the native protein within error. The kinetics of oxygen and CO binding to β-globin are unaffected by mutating the distal valine residue to either alanine or leucine. This finding agrees with the oxygen equilibrium curves, which show that these mutations leave the R-state oxygen affinity of Hb unchanged. Replacing valine E11β with isoleucine

lowers the oxygen association rate (and affinity) by about 10-fold. None of the β-globin E11 mutants showed an increase in association rate for either ligand.

3.3 Discrimination between Ligands

Carbon monoxide is constantly released into the blood by heme breakdown and binds to free heme very tightly. Hb, therefore, has to weaken the binding of carbon monoxide to remain able to transport oxygen. Oxygen binds to simple hemes in a bent fashion, whereas CO assumes a linear conformation perpendicular to the heme plane. There is no room for CO to bind in this way to heme proteins without severe steric strain, which led to the suggestion that heme proteins can discriminate between oxygen and carbon monoxide by selectively sterically hindering CO binding. This type of steric discrimination is found in some model heme compounds, so distal steric hindrance can lower CO affinity relative to oxygen affinity, but heme proteins do not necessarily behave in the same way. Native α- and β-globins have M values ($M = K_{CO}/K_{O_2}$) of about 250, while that of Mb is much lower, about 20. The M value of β-globin is little altered by replacing the distal histidine with glycine, but the same mutation in α-globin and Mb raises the M value to 7000 and 2000 respectively. Similarly, replacing valine E11 with alanine increases the M-value of α-globin to 2700 but does not affect the β-globin value significantly. The distal residues of both α-globin and Mb, therefore, discriminate against CO binding, both by selective steric hindrance and by the formation of a hydrogen bond with oxygen. The distal residues of β-globin do not behave this way, at least in R-state Hb.

The precise geometry adopted by CO in globins, and the energy penalty incurred by the steric clash with the E7 and E11 residues, have been the subject of numerous papers from crystallographic and spectroscopic groups. The discussion was occasionally marked by a distinct clash of these two cultures. The structure of carbonmonoxy Hb has in fact been refined several times, beginning with the 2.7-Å refinement by Baldwin (PDB code 2HCO). At this resolution, the heme geometry is rather imprecise and so in the model the heme was fixed as a rigid plane, and the ligand geometry cannot be modeled reliably. Higher resolution (2.1 Å) was achieved in 1991 by the Dodson group, who found the ligand in both subunits in a position almost perpendicular to the heme. This contrasted with an earlier model of Mb in which the ligand was modeled in a strongly bent conformation into rather ambiguous electron density, albeit at 1.5-Å resolution. A more recent HbCO model refined to 2.2 Å also shows a strongly bent ligand. Spectroscopic evidence has in general favored a linear Fe-C-O conformation. The best current model of HbCO in the Protein Data Bank is refined to 1.25-Å resolution, allowing individual atoms of the ligand to be visualized, and shows the ligand of both subunits lying close to the heme normal. MbCO has now been re-refined to even higher (1.15 Å) resolution (PDB code 1BZR), showing the ligand position to be similar to that in HbCO. Earlier crystallographic models of carbonmonoxy-heme proteins may have suffered from ligand loss or heterogeneity, and problems with insufficient care in handling model restraints during refinement.

Resonance Raman spectroscopy is ideally suited to studies of heme proteins

since the well-characterized ligand bending and stretching frequencies are sensitive to any distortion caused by the protein. Different vibrational frequencies have been assigned to the stretching modes ν(Fe-CO) and ν(C-O). For carbonmonoxy HbA, these are 507 and 1951 cm^{-1} respectively, compared to 495 cm^{-1} and 1954 cm^{-1} for a simple model heme compound. HbCO gives a sharp line at 507 cm^{-1}, showing that both subunits have very similar environments for the ligand. Replacing the distal valine of either subunit with alanine, leucine, or isoleucine, or replacing valine E11β with methionine, produces no effect on any of these vibrational modes; the spectra of these seven mutant Hbs were shown by Lin and coworkers to be identical to that of HbA. This result is quite surprising, since it was known that the CO affinity of the mutant proteins is altered. The drop in CO affinity cannot be due to strong tilting or bending of the ligand away from the heme normal since this would bring about significant changes in the Raman spectrum. The ligand geometry therefore appears to remain unaltered in the different mutants and the steric strain is taken up by the protein moiety, not the ligand.

4
Proximal Histidine Mutants

Site-directed mutagenesis also permitted an ingenious test of Perutz's stereochemical model by deleting the bond between the proximal histidines and the heme iron atoms in the α- or β-subunits, or both together. Barrick and colleagues of the Ho group showed that Hb mutants in which the proximal histidine is replaced by glycine can mimic HbA in the presence of imidazole, which binds to the proximal face of the heme just like a histidine side chain. Removing the covalent link between the heme and the F-helix releases any tension between the two. These experiments showed clearly that the proximal as well as the distal histidines behave differently in the α- and β-subunits, confirming the earlier results of Fujii and colleagues of the Morimoto group, who worked extensively with metal hybrid Hbs. The rate of carbon monoxide binding was found to be controlled by proximal effects in the α-subunits rather than the distal effects, which are much more pronounced in the β-subunits. Proximal detachment significantly increased the ligand affinity and reduced cooperativity. It was found that, in the presence of CO bound to all four hemes, releasing tension in the α-subunits led to considerable T-structure, but releasing tension in the β-chains did not. Releasing tension in both led to an intermediate between the two. Barrick and colleagues showed that detachment of the α- and β-hemes has opposing effects, but their results can be interpreted as showing that the position of the β-heme iron is relatively insensitive to the quaternary state, and that distal effects are more important in this subunit. rHbCO (His87$\alpha \rightarrow$Gly) is held in the T-state by the α-chain freed of the pull of the heme iron; the additional mutation in rHbCO (His 87$\alpha \rightarrow$Gly/His92$\beta \rightarrow$Gly) allows the β-heme to move more in response to distal ligand, leading to a smaller value of L, the allosteric equilibrium constant, [T]/[R]. rHbCO (His 92$\beta \rightarrow$Gly) is fully in R-state since the α- subunit iron is pulled close to the heme, and the ligated heme in the β-subunit is freer to move away from Val E11. This mutant also showed a tendency to split into $\alpha\beta$ dimers. Although, the loss of heme-F helix coupling prevents quaternary structure switching, in accordance with the

Perutz model, residual cooperativity was found indicating additional heme–heme communication pathways that do not involve the Fe–His bond.

5
C7 Mutants

Of all the hydrogen bonds and salt bridges in the $\alpha_1\beta_2$ interface of deoxy Hb, the hydrogen bond between tyrosine C7(42)α_1 and asparagine G1(99)β_2 appears to be one of the strongest. Asp 99β is altered in several clinically detected Hbs that have high oxygen affinity and bind oxygen noncooperatively as a result of the destabilization of the T-state. Tyr42α has not been found to be altered in a naturally occurring mutant, however. This residue has been mutated to phenylalanine and histidine using the Nagai and Thøgersen system. Hb Tyr42(C7)$\alpha\rightarrow$Phe shows properties very similar to those of the natural Asp99β mutants, indicating that the altered oxygen binding by these proteins results from the loss of the hydrogen bond rather than from a change in the size or shape of the residues. Replacing C7α with histidine gives a protein with properties intermediate between HbA and Hb Tyr42(C7)$\alpha\rightarrow$Phe, showing that a weak hydrogen bond can form between the histidine and the asparagine, thus allowing the protein to bind oxygen cooperatively.

The β-subunit residue equivalent to tyrosine 42α is phenylalanine 41β. This has been mutated to tyrosine, in an attempt to see if the newly introduced hydroxyl group can hydrogen bond to Asp94α, stabilizing the T-state. This mutation has been discovered clinically (Hb Mequon), but its properties were studied only as a hemolysate. The R- and T-state oxygen affinities of the purified protein were similar to those of HbA. L_0, the allosteric equilibrium constant of deoxy Hb, was higher for the mutant protein, so the T-state is stabilized. The binding of chloride and 2,3 diphosphoglycerate (DPG) were normal. Molecular dynamics simulations have shown that the tyrosine residue is not in fact likely to interact with Asp94α; probably it hydrogen bonds to the carbonyl group of His97(FG4)β in the same β-chain. It is proposed that this interaction stabilizes the hydrogen bond between Asp99β and Tyr42α.

6
N-terminal Mutants

As described earlier, the synthesis of human α- and β-globin in *E. coli* as nonfusion products results in proteins carrying an extra methionine residue at the N-terminus. Three such proteins have been synthesized and described in detail, one carrying the extra residue on the β-chains ("$\beta + 1$Met"), and the mutants Val1(NA1)$\beta\rightarrow$Ala and Val1(NA1)$\rightarrow$Met. The last of these was made by simply deleting the first valine codon so that the methionine replaces valine1β. Changing the first codon to alanine allows *E. coli* to remove the initiator methionine. The β-chains were refolded with heme and α-globin after purification. These mutants show similar oxygen binding and identical CO binding kinetics to HbA in the absence of allosteric effectors. In the presence of IHP, the CO binding of $\beta + 1$Met is significantly altered, but the other mutants behave normally. None of the mutations appeared greatly to alter the dimer–tetramer equilibrium or cooperativity, though the mutants all showed lower oxygen affinity than HbA. The allosteric effects of DPG was abolished and the effect of IHP diminished in the case

of Hb $\beta + 1$Met. The Val1(NA1)$\beta \rightarrow$Met mutant was found to have a Bohr effect identical (within error) to that of HbA.

The crystal structures of the three mutants have been solved in the deoxy form and refined to 1.8-Å resolution to observe the perturbations caused by the changes. In deoxy HbA, valine 1β makes several contacts with the H-helix and EF corner of the same β-subunit. The hydrophobic side chain lies in a pocket between these two, and the α-amino group forms a hydrogen bond with the carbonyl oxygen atom of leucine 78β. An anion-binding site found between valine 1β and lysine 82(EF6)β binds phosphate and sulfate ions, but since the crystals were grown from ammonium sulfate, the ion is sulfate in this case.

In all three mutants, the structural changes are localized to the N-terminus of the β-subunits, but the $\beta + 1$Met and Val1$\beta \rightarrow$Ala mutants showed different changes from the Val1$\beta \rightarrow$Met. In the first two, the amino-terminus of the β-chain becomes more mobile and the affinity for anions drops. Nearly all of the interactions with the H-helix and EF corner are weakened or broken as the N-terminus moves away. This effect is greater in the $\beta + 1$ Met mutant, the extra methionine nearly abolishing anion binding. In Hb Val1$\beta \rightarrow$Ala, the β-subunit N-terminus moves toward the EF corner so that the α-amino group can bind to Asp79(EF3)β. This group then competes with anions for the amino group and so lowers anion affinity. Unlike the other two mutants, Val1$\beta \rightarrow$Met increases the affinity for anions and decreases the mobility of the N-terminus. The X-ray crystallographic difference map between this mutant and HbA clearly shows extra density for the methionine and the bound sulfate. The main chain coordinate shifts are smaller in this mutant than in the other two, although the more bulky methionine side chain causes some movement away from the H-helix. This mutant therefore shows the least perturbed structure, yet has the greatest difference in oxygen binding affinity compared to HbA.

How can this apparent anomaly be explained? The most likely explanation comes from the study of Hbs from ruminants such as cows, which have lower intrinsic oxygen affinity than other mammalian Hbs. Perutz and Imai noted that these Hbs have large hydrophobic residues at the 2(NA2)β position and proposed that such proteins interact more tightly with the H-helix and EF corner, pulling the A-helix with them, stabilizing the T-state in a manner analogous to DPG binding to HbA. This mechanism fits very well with the observed functional and structural properties of the Val1(NA1)$\beta \rightarrow$Met mutant. Increased anion binding to the N-terminus of the β-chain in this mutant also agree with the observation that bovine Hb shows increased oxygen-linked chloride binding.

7 Trp37(C3)β

Both from the earliest crystal structures and, more clearly, from the later X-ray studies of T-state Hb with varying degrees of ligation by Paoli and coworkers, it is evident that Trp37β at the $\alpha_1\beta_1$ interface is a key residue in Hb function. Four mutant β-globins in which this Trp is replaced by Tyr, Ala, Gly, or Glu have been expressed in *E. coli* and reconstituted into tetramers with normal α-globin and subjected to a variety of analyses including crystal structure determination. Kinetic measurements of CO binding to

the deoxygenated mutant Hbs and analytical gel chromatography indicate that the tetrameric form is destabilized, resulting in considerable dissociation of even the deoxygenated hemoglobins into $\alpha\beta$ dimers at micromolar protein concentrations. Even with IHP added to stabilize the T-state, the normal behavior of HbA was not restored. The substitutions at 37β showed the same order of tetramer destabilization, increase in oxygen affinity, decrease in cooperativity and apparent tension in the iron–histidine bond: Tyr < Ala < Gly < Glu. Both tertiary and quaternary structure changes were found in the X-ray models of the deoxy mutant Hbs. The tertiary changes centered around Asp94(G1)α–Pro95(G2)α and Asp99(G1)β–Asn 102(G4)β, with no significant tertiary changes seen within the C-helix carrying the mutated residue. The mobility of the last two residues of the α-subunits showed increased mobility, and, in the Gly and Glu mutants the Fe-His(F8) bonds in the α- and β-subunits were found to be shortened. These results indicate that the globin fold is less able to restrain the deoxy heme iron from moving toward the heme plane if Trp37β is altered.

A more detailed study of Hb dimerization was reported by Vallone and colleagues, who measured the dissociation constant for five single and three double mutants of HbCO by analytical ultracentrifugation. The free energy change of dimerization was found to vary linearly with the hydrophobic surface area buried. Each square Ångstrom of hydrophobic surface area was found to increase the binding energy by 15 cal mol^{-1}, in good agreement with the theoretical estimate by Eisenberg and MacLachlan. Other mutagenic studies by the group of Manning suggest that residues distant from the contact interface may also contribute significantly to the free energy of dimer–dimer association.

8
Blood Substitutes

The need for an artificial blood substitute has been highlighted in recent years by concern regarding blood-transmitted diseases, but another considerable problem is the failure of the supply of donated blood to meet the rising demand. To be useful, the desired blood substitute must be nontoxic and long-lasting, and it must have oxygen-binding characteristics close to those of whole blood. Hb itself is an obvious starting material to make a suitable oxygen-carrying solution, but there are substantial problems associated with the introduction of free Hb into the circulation. First the plasma protein haptoglobin binds to Hb to form a complex, which is rapidly removed from the bloodstream. Once the haptoglobin level has dropped, free Hb passes through the kidneys as $\alpha\beta$ dimers. Free Hb is also liable to irreversible oxidation to metHb, which does not bind oxygen. In any case, blood has an oxygen affinity substantially lower than free human Hb because of the DPG present inside the red cell.

Since even highly purified Hb derived from human blood is harmful when introduced into the bloodstream, the protein must be modified in some way before it can be used as a blood substitute. A great deal of effort has been devoted to finding suitable chemical modifications to stabilize the tetramer and lower its oxygen affinity. However, any substitute based on human blood still retains some risk of transfer of pathogens from the donor, and the production of Hb on a large scale in microorganisms seems a more likely route to a safe product. As mentioned earlier, ruminant Hbs have an intrinsically low oxygen affinity in the absence of DPG and it has been suggested

that bovine Hb may be more suitable for modification into a blood substitute. The use of transgenic animals might also be used to produce human Hb (or a variant) in an animal host to prevent antigenic responses in patients, but there are still concerns regarding pathogens in Hb from such sources.

Three problems remain though: the short lifetime of free Hb in the blood, the need to reduce oxygen affinity while maintaining cooperativity, and the reactivity of Hb with nitric oxide, NO. The first difficulty, Hb dissociation into dimers small enough to pass through the kidneys, can be overcome by cross-linking the tetramer. An ingenious means of covalently linking two $\alpha\beta$ dimers without using chemical methods is described by Looker and coworkers. These investigators, noticing that the N-terminus of one α-chain is close to the C-terminus of the other, reasoned that a tandem fusion of two α-globin chains would be able to fold with β-globin and function as a normal tetramer, but would prevent dissociation into $\alpha\beta$ dimers. The expression vector produced is very similar to that of Hoffman et al. but contain two copies of the α-globin gene fused by a glycine codon instead of a normal α-globin gene. The "di-α-globin" gene and the β-globin gene are organized in a single operon. Upon induction with IPTG, host *E. coli* cells produce β-globin and the di-α-globin, which fold and incorporate heme to give a molecule very similar in overall structure to HbA. The fusion of the α-subunits was not expected to alter the oxygen affinity of the protein. Naturally occurring mutants described clinically were therefore examined to find the one with suitable properties, namely, low oxygen affinity and high stability. Looker et al. identified Hb Presbyterian (Asn 108$\beta \rightarrow$ Lys) as a suitable mutation.

Four mutant Hbs were synthesized:

rHb 0.0: (Val1$\alpha \rightarrow$ Met), (Val1$\beta \rightarrow$ Met)
rHb 0.1: di-α Hb (Val1$\alpha \rightarrow$ Met), (Val1$\beta \rightarrow$ Met)
rHb 1.0: (Val1$\alpha \rightarrow$ Met), (Val1$\beta \rightarrow$ Met, Asn108$\beta \rightarrow$ Lys)
rHb 1.1: di-α, (Val1$\alpha \rightarrow$ Met), (Val1$\beta \rightarrow$ Met, Asn 108$\beta \rightarrow$ Lys)

All four proteins showed substantially reduced Bohr effects and lowered cooperativity. However, rHb 1.1 shows oxygen affinity comparable to that of human blood. This protein was crystallized in the deoxy form to determine its structure. The Asn108$\beta \rightarrow$ Lys mutation causes substantial shifts of the B- and G-helices, but the mechanism by which it reduces oxygen affinity was not clear. The crystal structure was solved a few years later.

The linking of the α-subunits has been shown to increase the intravascular half-life of Hb three- to fourfold in rats, and kidney function was not detectably impaired in dogs when 30% of the blood volume was replaced with an isotonic solution of rHb 1.1. Clinical trials with rHb 1.1 and other cross-linked Hbs derived from human blood showed a marked hypertensive effect due to the reaction of the free Hb with nitric oxide. This unforeseen difficulty has in fact proved the major stumbling block for all attempts to produce a safe blood substitute from Hb, natural or artificial.

8.1 Nitric Oxide

Nitric oxide (NO) acts as a chemical messenger in the control of many important processes in the human body, and a vast literature has grown up around its effects, in a remarkably short space of time. NO

activates guanyl cyclase, leading to a variety of tissue-specific responses. NO is highly unstable and was therefore believed to act only locally. The introduction of free Hb into the circulation can remove NO by two main reactions – firstly, the direct binding of NO to deoxy hemes, and secondly, the oxidation of oxy-Hb by NO to give metHb and nitrate. NO can also bind to metHb and reduce it to the ferrous form, but this reaction is probably too slow to have any clinical relevance. The rate of the second reaction can be lowered significantly by increasing the bulk of apolar residues in the heme pocket. It has been shown that the reactivity of both α- and β-globin to NO can be reduced in this way. A series of recombinant Hbs has been produced with a range of rate constants for NO-induced oxidation, and it has been demonstrated that the hypertensive effect of these mutants in rats correlates very well with this reaction rate. In contrast, the overall oxygen affinity and autooxidation rates of the rHbs did not correlate with the observed changes in blood pressure. The data also suggest that the entire hypertensive effect may be attributed to the reaction of NO with deoxy and oxy-Hb.

This last point was somewhat contentious in the 1990s, following reports suggesting that Hb might have a previously unrecognized function as a NO transporter. It was proposed by Stamler and coworkers that the single cysteine residue in human Hb, 93β, could bind NO when the protein is oxygenated and in the R-state, and then release it as the protein switched back to the T-state on losing oxygen in the tissues. Injecting free Hb locked in the R-state should therefore cause a greater hypertensive effect, since it would not release NO readily. Using a recombinant Hb with a destabilized T-state (and therefore high oxygen affinity), it could be shown that the hypertensive effect is in fact independent of p_{50} and cooperativity. Additional experiments showed that the thiol groups on the injected Hbs remained free. Other results have equally cast doubt on the hypothesis that Hb serves as a NO carrier; for example, the amount of *S*-nitrosohemoglobin found *in vivo* is very low, and escape from the red cell would be very difficult for NO due to the high concentration of reactive hemes present. Attention has now focused away from cysteine 93β. The recombinant Hb-based artificial oxygen carriers therefore seem to have a considerable advantage over formulations based on human or animal blood, since the NO reactivity may be significantly reduced while preserving the desired oxygen-binding characteristics. The Olson group has used recombinant Mb as a test system in order to find mutants around the heme, which can inhibit NO binding while allowing oxygen to bind freely. Similar work by the group of professors Brunori, Chiancone, and Vallone in Rome has also shown that residue B10 exerts control over ligand selectivity. By replacing Leu B10 with tyrosine and His E7 with glutamine, the Rome group showed that NO binding could be reduced significantly while allowing oxygen to bind.

Recent research from the group of Yonetani has indicated, however, that natural Hbs carrying NO on the α-subunits (which have higher NO affinity than the β-subunits) may overcome the problem of hypertension. A number of commercial enterprises are continuing to work hard since many years to find a chemical modification to convert natural Hb into a safe, sterile alternative to present-day blood transfusions. The problems with this technology are highlighted by at least one company that is investigating methods of removing A and B type

antigens from red blood cells to allow "typeless" transfusions. Stockholders in such a company are clearly betting that recombinant Hb research does not solve the clinical problems quickly.

8.2
96α Val→Trp

Another artificial Hb with low oxygen affinity but high cooperativity has been investigated by the Ho group in the hope of finding a clinically useful oxygen carrier. This is the 96αVal→Trp mutant. Molecular dynamics simulations suggested that the indole side chain of the tryptophan might interact with 99β Asp, but the crystal structure later showed that instead it points into the central cavity and forms hydrogen bonds with 101βGlu. (Hb seems to provide too stern a test for molecular dynamics at present). Combining this mutation with the Asn 108β→Lys mutation of Hb Presbyterian gives a recombinant Hb with even lower affinity, which retains cooperativity. Unfortunately, it also oxidizes very easily. A third (29αLeu→Phe) mutation was added in an attempt to slowdown oxidation, but had limited effect. Analysis of this triple mutant was complicated by the appearance of *bis*-histidine hemichrome, and overall the tetramer was significantly destabilized.

9
Animal Hemoglobins

Globins are found in an enormous variety of organisms including bacteria, yeast, plants, and vertebrates. Their evolution has been the subject of numerous discussions, and the protein family clearly has a very ancient lineage. In the course of vertebrate evolution, Hbs have appeared with different functional properties. Not all animal Hbs use DPG (as human Hb does) as an allosteric effector. Some fish Hbs use ATP, rhinoceros Hbs use GTP, and cat and sheep Hbs use no such effector at all. These different properties can be explained by mutations found at the DPG binding site of human Hb. Other properties of animal Hbs are less easily explained, and site-directed mutagenesis has been used to investigate them by mutating human HbA.

9.1
Geese

Animals living at high altitudes have developed Hbs with high oxygen affinity to extract oxygen from the thinner atmosphere. This is particularly true of high-flying birds such as bar-headed geese, which migrate over the Himalayas at altitudes of around 10 000 m. The closely related greylag goose has Hb with an oxygen affinity much closer to that of HbA. Sequence comparison between the two Hbs revealed only four amino acid changes. Perutz predicted that the increase in oxygen affinity would be found to be due to a single mutation, Pro 119(H2)α→Ala. This residue forms a contact with D6β (which is leucine in bar-headed goose but methionine in man) at the $\alpha_1\beta_1$ subunit interface. All mutations that disrupt such contacts destabilize the T-state relative to the R-state. The prediction was confirmed by engineering this mutation into human Hb using the Nagai and Thøgersen expression system. Jessen and coworkers demonstrated that the mutant Hb shows the same increase in oxygen affinity as bar-headed goose Hb compared to greylag goose. From the X-ray crystal structure, it is clear that the mutation has caused only localized effects and not a gross change

in the molecule. Interestingly, the Andean goose, which flies across a mountain chain in another continent, has evolved a Hb in which the D6β residue is changed to serine. This breaks the same contact and leads to the same increase in oxygen affinity.

The combination of the known molecular structure of human Hb and the sequences of the two highly related goose Hbs allowed Perutz to predict which was the functional mutation. Unfortunately, it has generally not proved possible to predict or explain so easily the unusual properties of a number of animal Hbs.

9.2 Fish

Within physiological conditions, human Hb shows a decrease in oxygen affinity as pH falls, a phenomenon called the *alkaline Bohr effect*. The cause of the Bohr effect has been the subject of numerous studies, but a recombinant Hb was used by the Nagai group to confirm the major role of His 146β. Many bony fish Hbs show an exceptionally strong alkaline Bohr effect, known as the Root effect after its discoverer. Oxygen affinity drops sharply and oxygen binding becomes noncooperative as the pH is lowered to about 6 or less. The effect has an important physiological function, namely, supplying oxygen to the swim bladder and the eye. Both these organs secrete lactic acid into the bloodstream, which then discharges oxygen. It was suggested by Perutz that this increased stabilization of the T-state relative to the R-state by hydrogen ions is due to a single mutation, Cys 93(F9)$\beta \rightarrow$ Ser. The serine hydroxyl group was thought to form a hydrogen bond with the terminal histidine, His 146β in the T-state. This histidine forms a salt bridge with Asp 94β in T-state, but not R-state, HbA.

Perutz's idea was tested by Luisi and Nagai in Cambridge, who made a series of human Hb mutants with the Nagai and Thøgersen system. The mutations Cys 93$\beta \rightarrow$ Ser and His 143$\beta \rightarrow$ Arg were introduced sequentially. Both reduced the Bohr effect. Additional changes, Asp 94$\beta \rightarrow$ Glu, His 144$\beta \rightarrow$ Gln, His 2$\beta \rightarrow$ Glu and Glu 90$\beta \rightarrow$ Ser, did lead to an increase in the Bohr effect, but the overall oxygen affinity of the final mutant is much higher than that of native fish Hbs.

The X-ray crystal structures of Hb from the Antarctic fish *Pagothenia bernachii* (now renamed *Trematomus bernachii*) have been solved in both the deoxy T-state and the carbonmonoxy R-state, but unfortunately these provide few clues to the groups responsible for the pH dependence of the allosteric equilibrium. A pair of close carboxyl groups was suggested to destabilize the R-state under acid conditions, but cannot alone be responsible for the very low oxygen affinity found at low pH. Furthermore, the deoxy structure shows that the C-terminal β-chain histidine residue does not form a salt bridge equivalent to the one found in human Hb. This was a considerable surprise since this salt bridge is a major contributor to the pH dependence of human Hb. It seems, however, that there are a number of different groups, which can cause the Bohr or Root effects.

The R-state structure of another Root effect Hb, from a fish called *spot*, has also been solved by Mylvaganam and coworkers from which it was suggested that the effect is in fact caused by destabilization of the R-state rather than stabilization of the T-state. However, the major Hb from *Trematomus newnesi*, a close relative of *T. bernachii* is only different at 14 residues from *T. bernachii* Hb but has almost no

pH dependence at all. It also has all the residues suggested to destabilize the R-state at low pH. Clearly, the Root effect could not be solved by simple examination of the models, so the Nagai group in Cambridge has produced more mutants attempting to transfer the Root effect from *T. bernachii* Hb to *T. newnesi* Hb, but so far without success. The final proof of any proposed mechanism will be its transplantation into human HbA by mutagenesis.

9.3 Crocodiles

Uniquely among vertebrate Hbs, the principal allosteric effector of crocodile Hb is not an organic phosphate but bicarbonate ions, which strongly stabilize the T-state. Two bicarbonate ions bind to each Hb tetramer, lowering the oxygen affinity as much as IHP does in HbA. This effect is possibly related to the ability of the crocodile to remain submerged for long periods. It was suggested by Perutz that the DPG binding site of human Hb has been modified by three or four mutations to create a pair of bicarbonate ion binding sites related by the molecular twofold axis. Site-directed mutagenic experiments based on this hypothesis have failed to transplant bicarbonate binding, however.

An alternative site was proposed at the β-subunit C-helices, which form a sliding contact with the α-FG corners. Crocodile Hbs show two more positive charges than HbA in this highly conserved region of the molecule. A number of experiments to identify the true bicarbonate binding site have been carried out using the expression vector described by Hoffman et al. This work involved mixing crocodile-like and normal human globins to create mixed tetramers, conclusively showing that the bicarbonate binding site is found in the β-subunits. It has now been demonstrated that the bicarbonate effect is introduced into human Hb by mutations in the C-helix, but additional mutations in the α-subunit are required to maintain cooperative oxygen binding. A total of 12 mutations were introduced into human hemoglobin to confer the bicarbonate effect. The resulting mutant, called Hb Scuba, shows the same bicarbonate effect as native crocodile Hb. Crocodile Hbs have one of the mutations mentioned earlier at the C7 position, Phe 41β→Tyr, and it is believed the bicarbonate hydrogen bonds to this residue (Fig. 3). Hb Scuba has now been crystallized and further research is underway to investigate its properties.

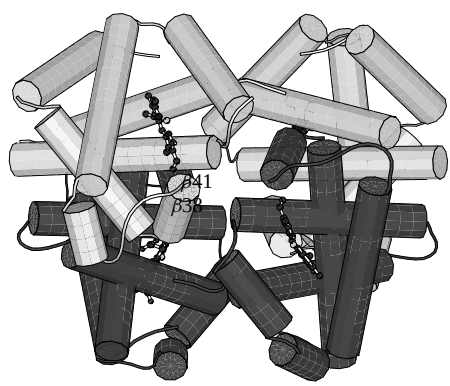

Fig. 3 A diagram of the Hb tetramer with the α-chains shown in dark gray and the β-chains shown in light gray. The 38β and 41β residues are implicated in the bicarbonate binding by crocodile Hbs, but a precise stereochemical model of this interaction has yet to be worked out.

So far, Hb Scuba is perhaps the most dramatic result from the Hb engineering, but it demonstrates that the evolution of new ligand binding sites may require more than the three or four mutations envisioned by Perutz or Kimura's "neutral hypothesis," and some of these may not have an obvious relevance from the molecular model of the protein.

10 Evolution

Genetic engineering has not only allowed us to probe the workings of the protein, but also allowed us to understand its evolution better. Hb evolution has attracted a great deal of interest because there is so much sequence data available for globins from many different species, and the structures of several of these proteins are known. The α- and β-globin genes appear to have diverged from a common ancestor around 500 million years ago. Although very similar, vertebrate α-globin has no equivalent of the D-helix in β-globin (Fig. 4). Human α-globin is therefore 5 amino acid residues shorter than β-globin. Two mutants were made by Komiyama and colleagues to test the functional role of the D-helix, one with the helix removed from the β-subunits and one with the same sequence inserted into the α-subunits. The Hoffman expression vector was used to synthesize the proteins as functional tetramers with a valine to methionine mutation at the N-terminus of each subunit. Both proteins, αD^+ and βD^-, folded and incorporated heme within the cell. Neither the insertion nor the deletion appreciably affected the oxygen affinity or cooperativity of oxygen binding – a remarkable demonstration of how robust the protein is to mutation. The Olson group has shown, however, that the loss of the D-helix does destabilize the βD^--subunit, which loses heme more readily than the native protein. Although neutral to ligand binding, removal of the D-helix from the β-subunits is an unfavorable mutation for the organism.

11 Nuclear Magnetic Resonance Spectroscopy (NMR)

In the last few years, both X-ray crystallography and macromolecular NMR have

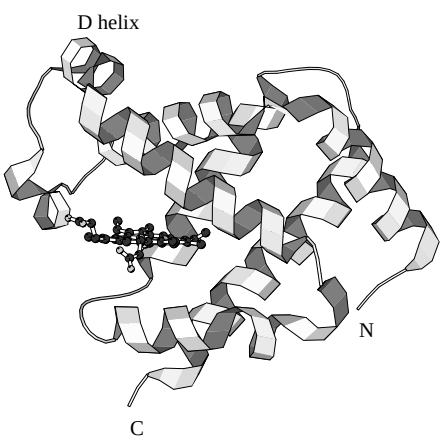

Fig. 4 The tertiary structure of β-globin, showing the extra D-helix not found in the α-chain. Apart from the D-helix, the overall structure of α- and β-globins is very similar. Deletion of the D-helix from the β-subunits and its insertion into the α-subunits still gives a functional Hb tetramer with cooperative oxygen binding.

been transformed by improvements in hardware and software. 800 MHz NMR magnets are now relatively common at protein research facilities, and 900 MHz machines are available, at a cost. Improvements in magnets and probes have now brought much larger proteins within the sphere of NMR, provided that they can be isotopically labeled appropriately. In the 1970s and 1980s, NMR was restricted to studies of the heme pocket or histidine residues of Hb. This work included some key work by the Ho group on the Bohr effect, which was a subject of considerable controversy.

In healthy humans, the pH of the blood is very tightly controlled within strict limits. Hb can, however, respond to small local changes in pH brought about by metabolism; decreases in pH lower the oxygen affinity, a phenomenon called the *alkaline Bohr effect* or, more usually, simply *the Bohr effect*. In fact, in 1904, Bohr discovered the drop in oxygen affinity with increased levels of carbon dioxide, five years before the concept of pH was put forward. The so-called *acid Bohr effect*, the decrease in oxygen affinity with rise in pH, is only seen at unphysiologically low pH for human Hb. The Bohr effect is expressed as the number of hydrogen ions released upon oxygenation (there is an absolute thermodynamic relationship between the change in oxygen affinity and the number of hydrogen ions taken up on oxygen binding). Histidine side chains have a pK close to neutral, and, therefore, are often able to bind or lose protons as their environment in protein changes. Proton NMR is extremely useful for studying the Bohr effect since the histidine resonances are often separate from other peaks in the spectrum. Furthermore, it is possible to find the pK of these histidines by titration, measuring the shift in peak positions with pH. One difficulty with this approach, however, is in assigning the resonances to the correct residues. Although many histidines could be assigned with confidence, marked disagreements also occurred. Using artificial Hb mutants, the Ho group has now completed the assignment of all surface histidine residues and measured their pKa values. These 24 residues account for 86% of the Bohr effect at pH 7.4 in 0.1 M chloride.

More recently, however, the Ho group has produced N^{15} isotopically labeled human HbA for analysis by NMR. This is an exciting development, which will allow us to analyze in much greater detail the protein in solution. For the majority of proteins, there is little doubt that crystal structures reflect the structure in solution quite faithfully – enzymes are still active in the crystal, and different crystal forms of the same protein but with different neighbor contacts show very similar structures. In the case of an allosteric protein such as Hb, there is the possibility that the crystallization might disturb the allosteric equilibrium, and select one allosteric form from a range of energetically similar conformations in equilibrium in solution. A new conformation of HbA was found in the early 1990s, called the *R2 state*. Within a few years, several papers appeared in which it was suggested, more or less strongly, that this might better reflect the "true" structure of HbA. In particular, this crystal form was grown from polyethylene glycol instead of ammonium sulfate, and the belief grew that the "high salt" conditions used by Perutz to make oxyHb and metHb crystals might inhibit Hb from undergoing the complete allosteric transition. A short article

in 1999 by the author critically reviewing the evidence each way suggested that obituaries of the R-state seemed premature. This drew a heated response from Mueser and colleagues, but crystallography alone is not able to test the solution state of a protein, and the case against the R-state remains weak at best. Recently, Lukin and colleagues have used isotopically labeled HbA to assess how well the NMR data agree with the R and R2 crystal structures. Under the conditions used, 0.1 M sodium phosphate at pH 7.0, it was found that a structure intermediate between the two fits the data best. NMR will no doubt play a large role in future Hb research.

12 Summary

Clearly, it has proved difficult to judge the functional roles of residues in Hb from the crystallographic structures alone, even where these are refined to high resolution in different ligation states. The functional properties of individual parts of the protein can be tested only by synthesizing mutants and analyzing the differences between them and HbA. Although the different effects on oxygen affinity of the valine E11 side chain were predicted correctly, the relative affinities of the α- and β-subunits for ligands were not. Despite many years of research, there is still much to learn about the nature of cooperative oxygen binding and the pathway by which deoxy Hb switches to oxy-Hb and back again. Site-directed mutagenesis continues to play an important role in this research; in fact, protein expression systems allowing folded, tetrameric Hb to be produced directly have led to a renaissance in the field.

While many clinically observed Hb mutations have shown that alteration of surface residues rarely affects too drastically the physiological properties of the protein, it is still remarkable that these changes produce so little effect. The tolerance of the globin fold to residue changes no doubt accounts for its evolutionary success. This property, fortunately, also gives us the opportunity to alter much of the sequence and still produce a functional protein, allowing many more mutants of interest to be made. While Hb is often presented in a rather ''cut-and-dried'' fashion in biochemistry textbooks, there remain a number of unsolved issues, and, no doubt, artificial mutants will be needed to resolve them.

Bibliography

Books and Reviews

Bunn, H.F., Forget, B.G. (1986) *Hemoglobin: Molecular, Genetic and Clinical Aspects*, Saunders, Philadelphia.

Dickerson, R.E., Geiss, I. (1983) *Hemoglobin*, Benjamin Cummings, Menlo Park, CA.

Fermi, G. & Perutz, M.F. (1984) *Atlas of Molecular Structure in Biology: Haemoglobin and Myoglobin*, Clarendon Press, Oxford.

Imai, K. (1982) *Allosteric Effects in Haemoglobin*, Cambridge University Press.

Perutz, M.F. (1978) Hemoglobin structure and respiratory transport, *Sci. Am.* **239**(6), 68–96.

Perutz, M.F. (1979) Regulation of oxygen affinity of hemoglobin. Influence of structure of the globin on the heme iron, *Annu. Rev. Biochem.* **48**, 327–386.

Perutz, M.F. (1983) Haemoglobin evolution. *Mol. Biol. Evol.* **1**, 1–28.

Perutz, M.F., et al. (1998) The stereochemical mechanism of the cooperative effects in hemoglobin revisited, *Annu. Rev. Biophys. Biomol. Struct.* **27**, 1–34.

Winslow, R.M., Vandegriff, K.D. (Eds.) (1997) *Advances in Blood Substitutes, Vol. 3: Industrial*

Opportunities and Medical Challenges, Springer, Berlin.

Primary Literature

Alayash, A.I. (1999) Hemoglobin-based blood substitutes: oxygen carriers, pressor agents, or oxidants? *Nat. Biotechnol.* **17**, 545–549.

Apostol, I. (1999) Assessing the relative stabilities of engineered hemoglobins using electrospray mass spectroscopy, *Anal. Biochem.* **272**, 8–18.

Barrick, D. et al. (1997) A test of the role of the proximal histidines in the Perutz model for cooperativity in haemoglobin, *Nat. Struct. Biol.* **4**, 78–83.

Barrick, D. et al. (2001) Distal ligand reactivity and quaternary structure studies of proximally detached hemoglobins, *Biochemistry* **40**, 3780–3795.

Bauer, C. et al. (1981) Analysis of bicarbonate binding to crocodilian hemoglobin, *J. Biol. Chem.* **256**, 8429–8435.

Chang, C.K. (2002) Effects of amino acid substitutions at β131 on the structure and properties of hemoglobin: evidence for communication between $\alpha_1\beta_1$ and $\alpha_1\beta_2$ subunit interfaces, *Biochemistry* **41**, 5644–5655.

Cheng, Y. et al. (2002) Ligand binding properties and structural studies of recombinant and chemically modified hemoglobins altered at β93 cysteine, *Biochemistry* **41**, 11901–11913.

Doherty, D.H. et al. (1998) Rate of reaction with nitric oxide determines the hypertensive effect of cell-free hemoglobin, *Nat. Biotechnol.* **16**, 672–676.

Dou, Y. et al. (2002) Myoglobin as a model system for designing protein based blood substitutes, *Biophys. Chem.* **98**, 127–148.

Dumoulin, A. et al. (1998) The N-terminal sequence affects distant helix interactions in hemoglobin, *J. Biol. Chem.* **273**, 35032–35038.

Eaton, W.A. et al. (1999) Is cooperative oxygen binding by hemoglobin really understood? *Nat. Struct. Biol.* **6**, 351–358.

Fang, T.Y. et al. (1999) Assessment of roles of surface histidyl residues in the molecular basis of the Bohr effect and of β143 histidine in the binding of 2,3 bisphosphoglycerate in human normal adult hemoglobin, *Biochemistry* **38**, 13423–13432.

Fronticelli, C. et al. (1991) Recombinant human hemoglobin: expression and refolding of β globin from *E. coli*, *J. Protein Chem.* **10**, 495–501.

Fujii, M. et al. (1993) The porphyrin-iron hybrid hemoglobins. Absence of the Fe-His bonds in one type of subunits favors a deoxy-like structure with low oxygen affinity, *J. Biol. Chem.* **268**, 15386–15393.

Gattoni, M. et al. (2001) The heme-globin and dimerization equilibria of recombinant human hemoglobins carrying site-specific β chain mutations, *Arch. Biochem. Biophys.* **15**, 172–178.

Hernan, R.A. et al. (1992) Human hemoglobin expression in *Escherichia coli*: Importance of optimal codon usage, *Biochemistry* **31**, 8619–8628.

Ho, C. et al. (1996) Roles of α114 and β87 amino acid residues in the polymerization of hemoglobin S: implications for gene therapy, *J. Mol. Biol.* **263**, 475–485.

Hoffman, S.J. et al. (1990) Expression of fully functional tetrameric human hemoglobin in *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8521–8525.

Imai, K. et al. (1989) Structural and functional consequences of amino acid substitutions in hemoglobin as manifested in natural and artificial mutants, *Protein Seq. Data Anal.* **2**, 81–86.

Imai, K. et al. (1991) Site-directed mutagenesis in hemoglobin – functional role of tyrosine α42(C7) at the $\alpha_1\beta_1$ interface, *J. Mol. Biol.* **218**, 769–778.

Imai, K., Tsuneshige, A., Yonetani, T. (2002) Description of hemoglobin oxygenation under universal solution conditions by a global allostery model with a single adjustable parameter, *Biophys. Chem.* **98**, 79–91.

Ishimori, K. et al. (1989) NMR study of human mutant hemoglobins synthesized in *E. coli* – consequences of tyrosine α42 substitutions. *J. Biol. Chem.* **264**, 14624–14626.

Jeong, S. et al. (1999) Recombinant hemoglobin (α29 Leu→Phe, α96 Val→Trp, β108, Asp→Lys) exhibits low oxygen affinity and high cooperativity combined with resistance to oxidation, *Biochemistry* **38**, 13433–13442.

Jessen, T.H. et al. (1991) Adaptation of bird hemoglobins to high altitudes – demonstration of molecular mechanism by protein engineering, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6519–6522.

Jessen, T.H. et al. (1994) Production of human hemoglobin in *Escherichia coli* using a

cleavable fusion protein expression vector, *Methods Enzymol.* **23**, 347–364.

Jia, L. et al. (1996) S-nitrosohaemoglobin: a dynamic activity of blood involved in vascular control, *Nature* **380**, 221–226.

Kavanaugh, J.S., Rogers, P.H., Arnone, A. (1992) High-resolution X-ray study of deoxy recombinant human hemoglobins synthesized from β-globins having mutated amino termini, *Biochemistry* **31**, 8640–8647.

Kavanaugh, J.S. et al. (1998) High-resolution crystal structures of human hemoglobin with mutations at tryptophan 37β: structural basis for a high-affinity T-state, *Biochemistry* **37**, 4358–4373.

Kiger L, et al. (1998) Thermodynamic studies on the equilibrium properties of a series of recombinant βW37 hemoglobin mutants, *Biochemistry* **37**, 4336–4345.

Kim, H.W. et al. (1995) A novel low oxygen affinity recombinant hemoglobin (α96 Val→Trp): switching quaternary structure without changing the ligation state, *J. Mol. Biol.* **248**, 867–882.

Kim, H.W. et al. (1994) Restoring allosterism with compensatory mutants in hemoglobin, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11547–11551.

Kimura, M. (1971) The neutral theory, *Sci. Am.* **241**, 94–104.

Komiyama, N. et al. (1991) Was the loss of the D helix in α globin a functionally neutral mutation? *Nature* **352**, 349–351.

Komiyama, N.H. et al. (1995) Transplanting a unique allosteric effect from crocodile into human haemoglobin, *Nature* **373**, 244–246.

Komiyama, N., Tame, J., Nagai, K. (1996) A hemoglobin-based blood substitute: transplanting a novel allosteric effect of crocodile Hb, *Biol. Chem.* **377**, 543–548.

Kroeger, K.S., Kundrot, C.E. (1997) Structures of a hemoglobin-based blood substitute: insights into the function of allosteric proteins, *Structure* **5**, 227–237.

Kwiatkowski, L.D. et al. (1998) Preparation and kinetic characterization of a series of βW37 variants of human hemoglobin A: evidence for high-affinity T quaternary structures, *Biochemistry* **37**, 4325–4335.

Larson, S.C. et al. (1999) A biochemical and biophysical characterization of recombinant mutants of fetal hemoglobin and their interaction with sickle cell hemoglobin, *Biochemistry* **38**, 9549–9555.

Lin, S.H. et al. (1990) The effect of the distal residues on the vibrational modes of the Fe-CO bond in hemoglobin studied by protein engineering, *Biochemistry* **29**, 5562–5566.

Looker, D. et al. (1992) A human recombinant haemoglobin designed for use as a blood substitute, *Nature* **356**, 258–260.

Looker, D. et al. (1994) Expression of recombinant human hemoglobin in *Escherichia coli*, *Methods Enzymol.* **231**, 364–374.

Lukin, J.A. et al. (2000) NMR reveals hydrogen bonds between oxygen and distal histidines in oxyhemoglobin, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10354–10358.

Lukin, J. et al. (2003) Quaternary structure of hemoglobin in solution, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 517–520.

Mathews, A.J. et al. (1991) The assignment of carbon monoxide association constants to the α and β subunits in native and mutant human deoxyhemoglobin tetramers, *J. Biol. Chem.* **266**, 21631–21639.

Mazzarella, L. et al. (1999) Crystal structure of *Trematomus newnesi* haemoglobin re-opens the root effect question, *J. Mol. Biol.* **287**, 897–906.

Miele, A.E. et al. (1999) Modulation of ligand binding in engineered human hemoglobin distal pocket, *J. Mol. Biol.* **290**, 515–524.

Miele, A.E. et al. (2001) Control of heme reactivity by diffusion: structural basis and functional characterization in hemoglobin mutants, *Biochemistry* **40**, 14449–14458.

Mylvaganam, S., et al. (1996) Structural basis for the root effect in haemoglobin, *Nat. Struct. Biol.* **3**, 275–283.

Nagai, K., Poyart, C., Perutz, M.F. (1985) Oxygen binding properties of human mutant hemoglobins synthesized in *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7252–7255.

Nagai, K. et al. (1987) Distal residues in the oxygen binding site of haemoglobin studied by protein engineering, *Nature* **329**, 858–860.

Nagai, K., Luisi B., Shih, D. (1988) Evolution of haemoglobin studied by protein engineering, *BioEssays* **8**, 79–82.

Olson, J.S. et al. (1988) The role of the distal histidine in myoglobin and hemoglobin, *Nature* **336**, 265–266.

Paoli, M. et al. (1996) Crystal structure of T state haemoglobin with oxygen bound at all four haems, *J. Mol. Biol.* **256**, 775–792.

Perutz, M.F., Imai, K. (1980) Regulation of oxygen affinity of mammalian haemoglobins, *J. Mol. Biol.* **136**, 183–191.

Peterson, E.S., Friedman, J.M. (1998) A possible allosteric communication pathway identified through a resonance Raman study of four β37 mutants of human hemoglobin A, *Biochemistry* **37**, 4346–4357.

Puius, Y. et al. (1998) Novel water-mediated hydrogen bonds as the structural basis for the low oxygen affinity of the blood substitute candidate rHb(α96 Val→Trp), *Biochemistry* **37**, 9258–9265.

Rohlfs, R.J. et al. (1998) Arterial blood pressure responses to cell-free hemoglobin solutions and the reaction with nitric oxide, *J. Biol. Chem.* **273**, 12128–12134.

Shen, T.J. et al. (1993) Production of unmodified human adult hemoglobin in *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8108–8112.

Shen, T.J. et al. (1997) Production of human normal adult and fetal hemoglobin a in *Escherichia coli*, *Protein Eng.* **10**, 1085–1097.

Simplaceau, V. et al. (2000) Chain-selective isotopic labeling for NMR studies of large multimeric proteins: application to hemoglobin, *Biophys. J.* **79**, 1146–1154.

Stamler, J.S. et al. (1997) Blood flow regulation by S-nitrosohemoglobin in the physiological oxygen gradient, *Science* **276**, 2034–2037.

Tame, J. et al. (1991) The functional role of the distal valine (E11) residues of α subunits in human hemoglobin, *J. Mol. Biol.* **218**, 761–767.

Tame, J.R.H. (1999) What is the true structure of liganded haemoglobin? *Trends Biochem. Sci.* **24**, 372–377.

Tame, J.R.H., Vallone, B. (2000) The structures of deoxy human haemoglobin and the mutant Hb Tyr α42 His at 120 K, *Acta Crystallogr., Sect D: Biol. Crystallogr.* **56**, 805–811.

Tsai, C.H. et al. (1999) Effects of substitutions of lysine and aspartic acid for asparagine at β108 and of tryptophan for valine at α96 on the structural and functional properties of human normal adult hemoglobin: roles of $\alpha_1\beta_1$ and $\alpha_1\beta_2$ subunit interfaces in the cooperative oxygenation process, *Biochemistry* **38**, 8751–8761.

Tsai, C.H. et al. (2001) Probing the importance of the amino terminal sequence of the β and γ chains to the properties of normal adult and fetal hemoglobins, *Biochemistry* **40**, 12169–12177.

Tsuneshige, A., Park, S., Yonetani, T. (2002) Heterotropic effectors control the hemoglobin function by interacting with its T and R states – a new view on the principle of allostery, *Biophys. Chem.* **98**, 49–63.

Tsuneshige, A., Yonetani, T. (2003) A novel blood transfusant candidate: intact human erythrocytes containing hemoglobin exclusively nitrosylated in the alpha-subunits, *Adv. Exp. Med. Biol.* **510**, 93–99.

Vallone, B. et al. (1998) Free energy of burying hydrophobic residues in the interface between protein subunits, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6103–6107.

Wagenbach, M. et al. (1991) Synthesis of wild type and mutant human hemoglobins in *Saccharomyces cerevisiae*, *Biotechnology* **9**, 57–61.

Weber, R.E. et al. (1993) Mutant hemoglobins α119-Ala and β55-Ser: functions related to high-altitude respiration in geese, *J. Appl. Physiol.* **75**, 2646–2655.

Whitaker, T.L. et al. (1995) The D-helix in Myoglobin and the β Subunit of hemoglobin is required for the retention of heme, *Biochemistry* **34**, 8221–8226.

Yuan, Y. et al. (2002) NMR investigation of the dynamics of tryptophan side chains in hemoglobins, *J. Mol. Biol.* **321**, 863–878.

Genetic Engineering of Vaccines

Tilahun D. Yilma, Leslie A. Jones, and Paulo H. Verardi
International Laboratory of Molecular Biology for Tropical Disease Agents, University of California, Davis, California

1 Introduction 295

2 Recombinant Vaccines 296
2.1 Vectors 296
2.2 Vaccinia Virus 296
2.3 Genetic Engineering 297

3 Vaccinia Virus Recombinant Vaccines 298
3.1 A Vaccinia Virus Recombinant Vaccine for Rinderpest 299
3.2 Vaccines for Peste-des-Petits Ruminants 302
3.3 A Vaccinia Virus Recombinant Vaccine for VSV 303
3.4 Recombinant Vaccinia Virus Vaccines for SIV 305
3.5 Safety and Efficacy of Live Recombinant Vaccines 306
3.6 Potential Recombinant Vaccines for HIV 307

4 Baculovirus Expression Vectors and Rapid Diagnostic Kits 307

5 Advantages of Vaccinia Virus Recombinants 308

Bibliography 309
Books and Reviews 309
Primary Literature 310

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

Keywords

Adjuvant Genes
Genes (often cytokine genes) that are used for the enhancement of the efficacy of recombinant vaccines.

Attenuating Genes
Genes that are used to enhance the safety of vaccines or vectors by reducing the virulence of agents.

Baculovirus
An insect virus widely used for abundant and inexpensive generation of recombinant proteins in insect cells or larvae.

Chimeric
Describing a fusion product of two or more genes or polypeptides often retaining the activities of both proteins.

Recombinant
An organism, nucleic acid, or proteins generated by genetic engineering.

Vaccinia Virus
An *orthopoxvirus* of uncertain origins causing no known disease that provides protective immunity against a number of *orthopoxvirus* infections, including smallpox.

Vector
A bacterium, virus, or a plasmid that is used for the introduction of nucleic acid (typically DNA) molecules for the generation of recombinants.

Using the techniques of molecular biology, effective modern vaccines are being developed by a variety of innovative methods that circumvent some of the problems associated with conventional vaccines. Live recombinant vaccines, which are prepared by isolating an immunogenic gene(s) from a disease organism and inserting it into a vector, are most promising. Such vaccines can provide immunity to agents that cannot be reliably attenuated or inactivated without destroying immunogenicity and allow one to distinguish vaccinated from infected animals.

1 Introduction

Over the past century, the gradual accumulation of knowledge concerning contagion and immunity has led to great advances in our ongoing struggle against infectious disease. In an effort to gain protection from the most serious infections, we have progressed from the earliest type of quarantine (the shunning of the obviously ill) to deliberate, though controlled and limited, exposure to an attenuated disease agent. One of the earliest examples of such exposure is seen in the battle against the dreaded scourge, smallpox. The dangerous and sometimes deadly practice of variolation – a deliberate exposure to the lesions of human cases in hopes of contracting a mild form of the disease – was practiced sporadically for centuries before the fortuitous discovery of the protection afforded by related, less virulent strains of virus. By 1796, Dr. Edward Jenner had begun to carefully introduce pox pustular material from cowpox lesions into his patients' skin – a process called vaccination – to protect against smallpox. The latter half of the nineteenth century brought an explosion of discoveries in immune protection, with specific vaccines produced for a wide variety of human and animal diseases. Over time, many of the most serious of human disease threats have retreated as a result of the advances of classical vaccine therapy.

Classical vaccines have generally been of three types: live attenuated agents, inactivated (killed) agents, or their modified toxins (toxoids). Vaccinia virus (VV), the poxvirus strain that protects human beings from smallpox, was attenuated naturally during its adaptation to different species. From the time of Pasteur and Koch, laboratories have worked to produce other attenuated disease agents (by varying their conditions of growth and maintenance or multiple passages through a different host species) or to find nonvirulent, protective strains in nature. To produce effective vaccines of some disease agents, it has been necessary to attempt inactivation of the agent by heat or chemicals, often destroying the nucleic acid while retaining the protein or carbohydrate components of the agent, thus eliminating its capacity for growth or replication while retaining the protective, immunity-inducing antigens.

The live attenuated and inactivated vaccines each have their advantages and disadvantages. Inactivated vaccines are usually easier to produce and are the only protection available against certain diseases. Protection against diseases caused by secreted toxins can be similarly accomplished by using heat or chemicals to convert these to toxoids, without toxicity but still immunogenic. However, the optimal dose of inactivated antigens must be provided for effectiveness, and even then these vaccines are poor inducers of cell-mediated immunity (CMI). Additionally, although thorough inactivation relieves safety concerns, there have been instances of diseases resulting from incomplete inactivation of an agent. Thus, inactivated vaccines often are less effective than live attenuated vaccines because live vaccines have the capacity to reproduce or replicate in the host organism, amplifying the antigen, and are excellent inducers of CMI. This amplification effect increases the stimulus to the immune system and may elicit a better immune response by providing optimal amounts of antigen. Maintaining live cultures and producing live vaccines, however, is much more demanding, requiring special laboratory and transport facilities. Many live vaccines

are heat labile and must be refrigerated, further complicating their use in remote areas. Other safety concerns also arise with live vaccines because attenuated strains may revert to virulence in the vertebrate host or an arthropod vector, new strains of the agent may arise from the vaccine strain, or the vaccine may actually induce the disease in pregnant or debilitated animals.

A further serious disadvantage of classical vaccines is that, whether live or killed, they contain whole disease agents that elicit a full range of immune responses. Thus, with very few exceptions, one cannot develop a serological test for distinguishing animals vaccinated with either live attenuated or inactivated vaccines from those recovered from natural infections. This is an important consideration in epidemiological studies, animal disease diagnosis and control, as well as in economics. Restrictions on the export/import of animals and their products because of positive antibody tests can result in substantial economic losses.

Advances in molecular biology in the last thirty years have given rise to a remarkable new generation of vaccines that includes subunit, synthetic peptide, live recombinant, and DNA vaccines. These advances include the identification of an agent's specific, immunity-inducing molecules, the synthesis of all or the active portions of these molecules, and the incorporation of the responsible immunogenic genes into the genome of a nonvirulent carrier or vector agent. A wide variety of vaccines are now being developed for use and may greatly improve our ability to protect against infectious diseases. Live recombinant vaccines, in particular, appear to hold great promise for continuing improvements in disease control.

2 Recombinant Vaccines

2.1 Vectors

Recombinant DNA technology now makes it possible to retain most of the advantages of live vaccines, while overcoming many of their inherent safety hazards and logistical disadvantages, by inserting and expressing heterologous genes in a variety of different viruses. Both DNA and RNA viruses are used extensively as expression vectors. Although the relatively small size of most of these viruses facilitates genetic engineering, this property also restricts the amount of DNA that can be inserted into the virus particle. In addition, many of these virus vectors have a limited host range that does not allow their replication in the vaccine's target species. They may be defective, requiring either a helper virus or special cell lines for their replication. Furthermore, many of these vectors are oncogenic and therefore may be unsuitable for use as vaccines in humans and domestic animals.

2.2 Vaccinia Virus

The disadvantages of the vector systems can be overcome by using VV as a vector and by creating a live recombinant vaccine with the immunogenic genes of disease agents incorporated into the VV genome. When the VV vector is introduced into an animal, the virus replicates in the host's cells and the foreign genes are expressed along with those of VV; thus, the animal becomes immune both to VV and to the agent from which the foreign gene was taken. The history of VV in human hosts attests to its safety: during this past century, the VV vaccination procedure has

protected hundreds of millions of people worldwide from smallpox, with minimal side effects and no adverse environmental impact, despite global release of the VV under varying hygienic conditions. Vaccination with this virus has been achieved by scarification (light scratching of the skin), by subcutaneous or intramuscular injection, and by intranasal or even oral administration. It can be grown outside the laboratory by scarification of the abdomen of a single calf to prepare more than 200 000 vaccine doses – a great advantage for use in developing countries. The virus is heat stable, particularly in the lyophilized state, and it can be reconstituted easily. Modern vaccine technology now gives us, through genetic engineering, the capability of using VV as a live vector to protect animals and people against other unrelated, disease agents.

A member of the poxvirus family, VV is a large, double-stranded DNA virus that packages a complete transcription system and replicates in the cytoplasm of infected cells. The virus has a wide host range that includes humans, cattle, horses, swine, sheep, goats, mice, and monkeys, making it useful for both human and a variety of animal vaccines. Its large genome can accept more than a dozen heterologous genes, opening the possibility of producing a multivalent vaccine that could protect against several diseases with one administration. The molecular biology of the virus is well documented and supports the engineering techniques necessary for producing, isolating, and identifying recombinants. The genome contains more than a dozen nonessential regions that can be used for incorporation of foreign antigen genes and reporter genes that assist in identifying the desired recombinants, in addition to the commonly used viral marker, the *TK* (thymidine kinase) gene. Using recombinant DNA technology, convenient marker rescue techniques and plasmid vectors containing homologous sequences to the virus DNA have been developed that allow recombination within infected cells.

2.3 Genetic Engineering

The application of recombinant DNA technology to constructing recombinant VVs utilizes convenient marker rescue techniques and plasmid vectors containing homologous sequences to the virus DNA, allowing recombination within infected cells. The strategy for using VV as a vector (as shown in Fig. 1) involves first constructing a plasmid carrying a foreign (heterologous) gene. This unit consists of a VV promoter fused to a foreign-protein-coding sequence that is flanked by DNA from a nonessential region of the VV genome. The chimeric gene is incorporated into the VV genome by homologous recombination in tissue culture cells that have been infected with wild-type VV and then transfected with the plasmid. Although any nonessential region of the VV genome can be used as the site of gene insertion, the thymidine kinase gene locus provides some advantages because recombinants are then TK^-. The TK^- phenotype serves to attenuate viral pathogenicity, and also can be easily distinguished from wild-type TK^+ virus by 5-bromo-2′-deoxyuridine (BrdU) selection. Rapid screening of recombinants of interest can also be facilitated with the co-expression of reporter genes, such as the *Escherichia coli* β-galactosidase (*lacZ*) gene. Expression of the *lacZ* gene results in a blue coloration in the presence of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal); one is thereby

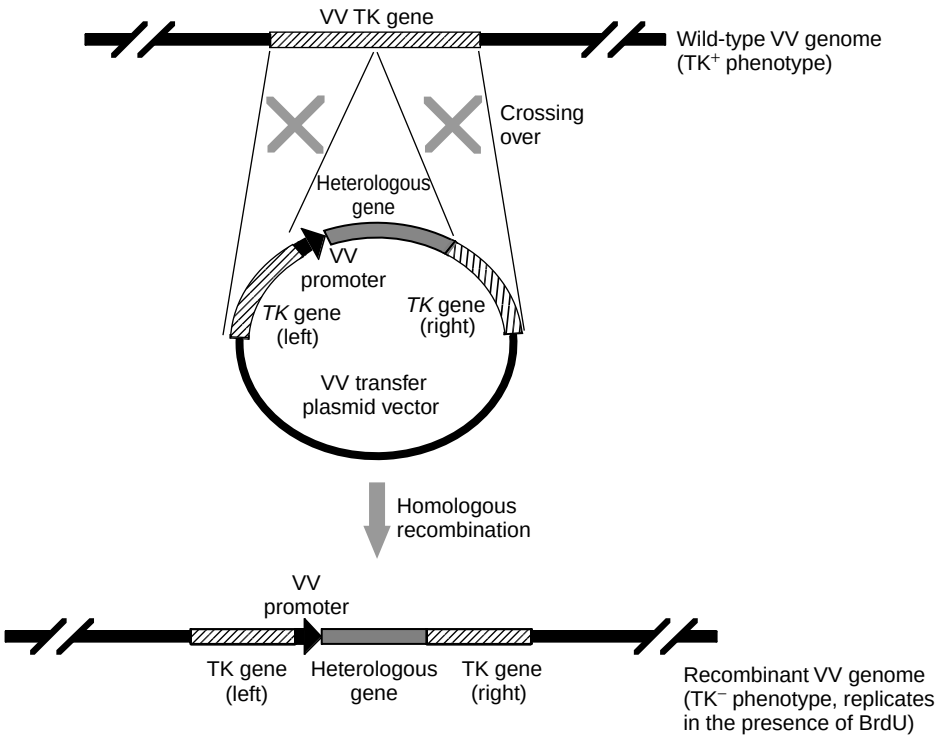

Fig. 1 Construction of recombinant VV vaccines by homologous recombination. The VV transfer vector contains the VV *TK* gene insertionally inactivated by a foreign (heterologous) gene under the control of a VV promoter. Upon infection of cells in tissue culture with the wild-type VV (TK^+ phenotype) and transfection of the same cells with the transfer vector, a double crossing-over event may occur between the homologous sequences in the transfer vector and the VV genome. This results in the transfer of the heterologous gene under the VV promoter to the VV genome, and the insertional inactivation of the *TK* gene. Recombinant VVs are selected from wild-type viruses by serial plaque assays in TK^- cell lines in the presence of BrdU. TK^- cells infected with wild-type VV (with TK activity) are able to phosphorylate BrdU and lethally incorporate this thymidine analog into the VV genome. TK^- cells infected with recombinant VVs (without TK activity) are unable to incorporate BrdU into the VV genome and as a result, progeny viruses are viable.

able to distinguish recombinants from wild-type virus by their blue plaque phenotype.

3 Vaccinia Virus Recombinant Vaccines

A description of three VV recombinants that we have constructed and used with varying degrees of success as vaccines will illustrate the potential of these agents: a vaccine for rinderpest that induced 100% protection against a challenge greater than 1000 times the normally lethal dose, a vaccine for vesicular stomatitis that was moderately successful in protecting against the disease, and a vaccine for simian immunodeficiency virus (SIV) that did not prevent infection but significantly

reduced the virus load from control levels and delayed the onset of clinical disease.

3.1
A Vaccinia Virus Recombinant Vaccine for Rinderpest

Our most successful VV recombinant vaccine is one developed for rinderpest virus (RPV), the causative agent of the single most important viral disease of livestock in developing countries. An acute, febrile, highly contagious viral disease of ruminants (particularly cattle and buffalo) with a rapid course and a mortality rate exceeding 90%, rinderpest is characterized by inflammation, hemorrhaging, necrosis, and erosion of the gastrointestinal tract accompanied by bloody diarrhea, wasting, and death. This disease was controlled in Africa in the 1970s by an international effort called the Joint Project/15 (JP/15), which vaccinated 124 million cattle, but recent resurgences have necessitated a second Pan-African Rinderpest Campaign, currently in progress.

Continuing outbreaks of rinderpest have pointed out the logistical difficulties of reliance on live vaccines in the regions where this disease is prevalent, but the vaccine currently in use, the Plowright tissue culture rinderpest vaccine (TCRV) for rinderpest, is an attenuated, live vaccine with the limitations common to such preparations. It is extremely heat labile and requires maintenance of a cold chain; in most instances, the vaccine must be used within 30 minutes of reconstitution to ensure potency. TCRV is effective, but the requirement of skilled personnel for vaccine preparation and administration and a need for refrigeration facilities – often lacking in the hot and isolated areas in which the disease is prevalent – are obstacles to its use.

A VV recombinant vaccine for rinderpest, many felt, would provide an alternative and circumvent many of the obstacles limiting the use of TCRV. The heat-stable lyophilized form of VV is easily produced, stored, and transported; and it can be administered by a number of routes including scarification, intradermal, intramuscular, subcutaneous, and oral. It is possible, as was shown in the smallpox eradication campaign, for each independent area to carry on vaccine production by collecting and emulsifying scabs from vaccinated animals for use as needed. The self-sufficiency thus afforded to local animal handlers is a rare and valuable empowerment for developing countries that frees them from reliance on international assistance. For these reasons, a VV recombinant vaccine for rinderpest is a valuable addition to the arsenal of defense and raises the possibility of eradicating this disease. Our efforts to construct such a vaccine have had a highly successful conclusion: we have produced one of the two most effective recombinant vaccines to date. Our rinderpest vaccine, which was designated as a prototype for genetically engineered vaccines for export, has been approved for field trials in Kenya and Ethiopia by the U.S. Department of Agriculture/Animal and Plant Health Inspection Service and the U.S. Agency for International Development. The other successful agent is a recombinant VV vaccine for rabies, currently being extensively used to eradicate sylvatic rabies in Europe and North America.

The causative virus of rinderpest is enveloped and has a single-stranded RNA genome with a minus polarity. It is in the family *Paramyxoviridae* and is a member of the morbillivirus genus along with measles virus of human beings, distemper virus of dogs, and peste-des-petits-ruminants virus

(PPRV) of goats and sheep. We propagated the highly virulent Kabete "O" strain of RPV in primary bovine kidney cells and characterized eight viral proteins. We made cDNA copies of the phosphoprotein (P), hemagglutinin (H), fusion (F), and nucleoprotein (*N*) genes and determined the complete nucleotide sequence of these genes. In all paramyxoviruses, the H and F surface proteins elicit protective immunity. Standard procedures were used to construct VV recombinants expressing the H gene (vRVH) or the F gene (vRVF) of RPV.

Using these VV recombinants (vRVH and vRVF), protective immune response studies in cattle were conducted in the high containment facility at the Plum Island Animal Disease Center. Each animal was shown to be seronegative to RPV prior to vaccination, and 10^8 plaque-forming units (PFU) of the VV recombinant vaccine was administered in a single vaccination by intradermal inoculation and scarification. Groups of five animals received either one recombinant (vRVH or vRVF) or a mixture of both recombinants (vRVH + vRVF). Two animals were vaccinated with TCRV as positive controls, and two were left unvaccinated as negative controls. For a VV control, two animals were vaccinated with v50, a VV recombinant expressing the glycoprotein gene of the vesicular stomatitis virus (VSV).

Pock lesions developed as early as four days in all animals vaccinated with the recombinants, but were limited to the site of inoculation and healed completely by two weeks. Humoral responses to the vaccines were assessed by serum neutralization (SN) assay. All animals vaccinated with the recombinants or TCRV produced SN antibodies to RPV. As expected, control animals had no detectable antibody titer.

To evaluate protection, all animals were challenged subcutaneously with a heavy dose (10^3 $TCID_{50}$–tissue culture infectious dose for 50% of subjects) of RPV in the prescapular lymph node region on day 35 following primary immunization. As low as 1 $TCID_{50}$ induced clinical rinderpest with 100% mortality. Cattle vaccinated with the recombinants or TCRV were completely protected when challenged, even though the inoculation was greater than 1,000 times a lethal dose of RPV. Some of the cattle, most of these vaccinated with the F recombinant (vRVF), had strong anamnestic responses to RPV after challenge inoculation; this indicated replication of the challenge virus. No anamnestic response could be demonstrated in the groups vaccinated with the mixture of both recombinants and TCRV, indicating little or no challenge virus replication.

Viral infection and spread are mediated by receptor binding (to H) and membrane fusion (by F), so it is reasonable to expect a vaccine preparation containing both antigens to be superior to one containing the H- or the F-glycoprotein alone. Vaccinating with the vRVF + vRVH mixture in the field, however, is cumbersome and expensive, and particularly burdensome for developing countries. For this reason as well as to obtain greater standardization and more reliable results, we decided to include both genes in a single recombinant. Using techniques similar to those described earlier, we developed a VV (Wyeth strain) double recombinant that includes both the *F* and the *H* genes (vRVFH) of RPV; this recombinant expresses the authentic RPV F- and H-proteins. The *F* and *H* genes were inserted into the *VV TK* and *H* gene regions, respectively, thus inactivating these *VV* genes and giving an even more highly attenuated VV recombinant vaccine. Indeed, cattle vaccinated with this

double recombinant (vRVFH) had no pock lesions at the site of inoculation, indicating a high degree of attenuation.

Cattle were vaccinated with 10^8 PFU of vRVFH by intradermal inoculation in the neck region. Two groups of cattle were vaccinated with VV recombinants and kept in separate isolation rooms. In the first group, five animals were vaccinated with vRVFH. In the second group, four animals were vaccinated with a mixture of vRVF and vRVH. In addition, two unvaccinated animals were housed with each group, to assess the transmissibility of VV recombinants from vaccinated to contact animals. Unlike animals vaccinated with the mixture of vRVH and vRVF, those vaccinated with vRVFH developed no detectable pock lesions; a thorough examination failed to demonstrate pock lesions in the contact animals in both groups. Further, serum samples taken from contact animals on the days of vaccination (day 0) and challenge (day 28) were negative to VV by SN and plaque reduction assays. All cattle vaccinated with the recombinants produced SN antibody to RPV as early as eight days after vaccination. However, all contact and control animals lacked detectable SN antibody to RPV during the course of the experiment.

One month postvaccination, all animals in both groups were challenge-inoculated with 10^3 $TCID_{50}$ of the pathogenic Kabete ''O'' strain of RPV. Cattle vaccinated with VV recombinants (both groups) were completely protected from rinderpest, exhibiting no detectable illness and a normal temperature of 38 °C. The four unvaccinated contacts developed high fever (42 °C) by day 2 and died by day 6 after challenge, showing lesions typical of severe rinderpest. Daily monitoring for two weeks showed no detectable clinical disease in vaccinated animals, and the experiment was terminated. In a confined field trial in DebreZeit, Ethiopia, the vaccine was shown to be effective for at least one year postvaccination.

We have also tested vRVFH and the parental Wyeth strain for pathogenicity by inoculating immunodeficient nude mice intraperitoneally, and no clinical disease was detected in 12 test mice. In addition, we assessed the effects of vRVFH and the mixture of single recombinants (vRVF + vRVH) on the immune systems of rhesus macaques infected with the simian immunodeficiency virus (SIV). These monkeys suffer from an immunodeficiency disease virtually identical to the human acquired immunodeficiency syndrome (AIDS). Hence, SIV is a legitimate model for HIV. Four normal macaques and four SIV-infected macaques were vaccinated with vRVFH. Apparently, the lack of development of pock lesions in cattle vaccinated with vRVFH is host specific; both groups of monkeys developed small pock lesions at the site of inoculation, but these healed completely by two weeks postvaccination. No clinical disease or dissemination of VV was observed in either normal or SIV-infected macaques during a one-month observation following vaccination.

More recently, we have developed a second-generation recombinant VV vaccine (v2RVFH) that induces both humoral immune responses and protective immunity comparable to those elicited by TCRV. We enhanced efficacy by using strong synthetic promoters to increase the expression of the *F* and *H* genes of RPV. Moreover, we added the *lacZ* gene to serve as an easily identifiable marker to aid in recombinant VV selection during the development of the vaccine, and more importantly, to distinguish the vaccine virus (v2RVFH) from

other naturally occurring poxviruses during vaccination programs.

Compared to vRVFH or the single recombinants expressing F or H, v2RVFH displays a distinct plaque morphology in cell culture that is characterized by the formation of massive syncytia due to higher levels of expression of both glycoproteins. Radioimmunoprecipitation studies showed that the band intensities for the F- and H- proteins expressed by v2RVFH were at least threefold higher than those generated by vRVFH or the single recombinants. We conducted short- and long-term safety and efficacy studies in African cattle, and demonstrated that as little as 10^2 PFU of v2RVFH, given intramuscularly, protects 50% of the vaccinated cattle from virulent challenge with RPV. Unlike the intradermal route, the intramuscular route facilitates mass vaccination of cattle under field conditions. We also showed that a dose of 10^8 PFU of v2RVFH induced levels of SN antibodies to RPV comparable to those elicited by TCRV, and that this dose provides sterilizing immunity to cattle for at least two years postvaccination. This level of protection suggests that v2RVFH confers long-term immunity equivalent to TCRV. Lastly, v2RVFH is not only highly efficacious but also safe, since it is attenuated by the insertional inactivation of the *TK* gene. As a result, no pock lesions were detected after vaccination, and there was no lateral transmission to contact animals.

Rinderpest appears to be an excellent candidate for eradication using the VV recombinant vaccines. There is only one serotype of RPV, although there are different strains manifesting different degrees of pathogenicity in the field. A vaccine against one strain will immunize against all, including peste-des-petits ruminants virus (PPRV) of sheep and goats. Use of the rinderpest recombinant VV vaccines in areas of the world where PPRV is endemic would also aid in the control and eradication of PPR.

3.2 Vaccines for Peste-des-Petits Ruminants

Peste-des-Petits Ruminants (PPR) is a disease of small ruminants caused by a virus (PPRV) closely related to RPV of cattle. The disease is characterized by necrotizing and erosive stomatitis, enteritis, and pneumonia. Morbidity and mortality are very high, and the virus is a serious health problem in small ruminants in Africa and the Middle East. Although cattle can be infected with PPRV, they do not exhibit clinical disease. Vaccination of goats with vRVFH provided complete protection against a challenge-inoculation with 10^3 $TCID_{50}$ of PPRV.

All goats used in these experiments were determined to be seronegative to PPRV and RPV at the start of the study. Eight goats were vaccinated with 10^8 PFU of vRVFH by intradermal inoculation and scarification. For controls, six goats were vaccinated with the parental VV (Wyeth strain). On day 35 postvaccination, goats were challenge-inoculated with 10^3 $TCID_{50}$ of PPRV, strain Mieliq-Sudan, by subcutaneous injection just anterior to the prescapular lymph node. Serum was collected on days 0, 35, 42, and 56, and titers to RPV and PPRV were determined by SN and by ELISA (enzyme-linked immunosorbent assay). Antibody directed to the F-protein was measured by ELISA only, because the glycoprotein is not the antireceptor of the virus and does not usually induce SN antibody. Goats vaccinated with vRVFH had SN (32–512) and ELISA (2–16) titers to RPV and undetectable SN titers to PPRV by

day 35 postvaccination. Anti-PPRV ELISA titers of vaccinated goats were undetectable in four goats and low (2, 2, 2, and 8) in the remaining four goats on the day of challenge.

All control animals lacked detectable SN and ELISA titers to either PPRV or RPV on the day of challenge. After challenge inoculation, however, there was a strong anamnestic response, as indicated by high SN and ELISA titers. In contrast, the postchallenge ELISA titers to PPRV and RPV of goats vaccinated with vRVFH ranged from 64 to 2048 and from 64 to 1024, respectively. Although SN titers to PPRV could not be detected in the animals vaccinated with vRVFH on the day of challenge inoculation, all vaccinees were completely protected against PPRV, and no clinical signs (e.g. fever or diarrhea) were exhibited over the following period. In contrast, all the control animals vaccinated with the parental Wyeth strain of VV were febrile (>39.7 °C) and suffered from severe PPR within 4 days postchallenge. Three animals died; the rest were euthanized.

Clearly, this vaccine – a VV double recombinant vaccine expressing the *F* and *H* genes of RPV – completely protects goats against PPR, even though no neutralizing antibody to PPRV could be detected before challenge inoculation. Although no anti-PPRV ELISA titers were found in four of the vaccinated animals, and the remaining four had very low titers, all vaccinated goats were equally protected. It has been reported that in cattle administered classic live viral vaccines to RPV, any detectable SN titer is associated with complete protection; this was, however, the first report of protection in the morbillivirus group in the complete absence of SN antibody to the homologous virus. In this regard, we have recently provided evidence that CMI may be the major determinant in protection against morbilliviruses. Cattle vaccinated with subunit antigens (F and H) of RPV expressed in baculovirus produced viral neutralizing antibody but were not protected.

Control of PPR is of great economic importance in countries engaged in large-scale rearing of small ruminants. Until a recombinant vaccine is developed expressing the homologous *F* and *H* genes of PPRV, vRVFH may be used for the control of PPR. Replication of VV is unaffected by antibody to PPRV, and vRVFH can be used to immunize young animals that still have passive (maternal) antibody to PPRV. Also, one can employ a serological test to distinguish vRVFH-vaccinated goats from those exposed to the whole virus, either by vaccination or infection. We have developed such rapid, inexpensive diagnostic kits on the basis of the *N* gene of RPV and PPRV (see Sect. 4).

3.3 A Vaccinia Virus Recombinant Vaccine for VSV

One of the earliest VV recombinants prepared and tested in our laboratory was a vaccine for vesicular stomatitis (VS), a contagious disease of horses, cattle, and pigs characterized by vesicular lesions on the tongue and other areas of the oral mucosa. Interestingly, the virus can be used to infect mice in a model system in which it causes neuropathy and death in 100% of the subjects. A serious cause of economic loss throughout North and South America, this disease is notable for the confusion of diagnosis between VS and foot-and-mouth disease, a pathological epizootic condition with a 100% morbidity rate.

Experimental attenuated and inactivated VSV vaccines have been developed, but

the U.S. Department of Agriculture generally prohibits their use for vaccination because vaccinees cannot be distinguished from naturally infected animals. It is possible to distinguish animals vaccinated with a subunit preparation by serology, however, and this is an essential consideration for epidemiological studies, disease control, and establishment of quarantine programs. Advantages of subunit and live vaccines could theoretically be combined in a VV recombinant expressing specific VSV proteins, and such recombinants were constructed by the methods described for rinderpest.

Of all the antibodies to various proteins of VSV, only those directed to G, the external glycoprotein, have been shown to be protective in mice. When DNA copies of mRNAs for the G- or N- proteins of VSV were linked to a VV promoter and inserted into the genome of VV, the recombinants retained infectivity and synthesized VSV polypeptides. We demonstrated that the polypeptides expressed by these recombinants are similar or identical in size and antigenicity to the natural VSV proteins, and that the G-glycoprotein is transported normally to the cell surface (perhaps important for antigenicity). After intradermal vaccination with live recombinant virus expressing the G-protein (v50), mice produced VSV-neutralizing antibodies. VSV SN titers were detected by day 14 and increased over a 42-day period. Half the mice were given a booster vaccination on day 28, which resulted in a seven- to eightfold increase in SN titers over that obtained with a single vaccination.

Challenge was carried out by intravenous injection of VSV, which produces lethal encephalitis in mice within 6–12 days. For control purposes, a group of mice was vaccinated with vHBs4, a VV recombinant that expresses the hepatitis B virus surface antigen gene. When this group was challenged, 7 of 11 animals died of encephalitis. By contrast, only 1 of 15 mice that received a primary vaccination with v50 died, and none of 16 that received the two vaccinations died.

To further evaluate the immunogenicity of the recombinant virus, cattle were vaccinated intradermally and later given intralingual challenge injections of virulent VSV; the degree of protection correlated directly with the neutralizing antibody produced following vaccination. After vaccination, typical pox lesions appeared within four days. SN titers to VSV were detected on day 7 and reached values of 80–480 by day 29. At that time a second vaccination was given, following which the titers increased several fold. On day 44, the cattle were challenged by intralingual injection of 10^2 and 10^3 PFU of VSV. Preliminary experiments indicated that most unvaccinated cattle challenged in this manner developed vesicular lesions within two days. All controls (unvaccinated cows and cows vaccinated with vHBs4) developed typical vesicular lesions at the 10^2 and 10^3 PFU VSV injection sites. In contrast, only two of six vaccinated cows developed vesicular lesions at the lower challenge dose of VSV. The remaining four cows had no lesions at the 10^2 PFU VSV injection sites, although all animals developed lesions at the 10^3 PFU sites.

Good correlation was observed between antibody titers and protection against challenge with 10^2 PFU of VSV. All animals that were protected had SN antibody titers of 1280 or greater, whereas those that were sensitive had titers of 640 or lower. The degree of protection obtained is particularly noteworthy in view of the severity of the experimental challenge, compared to that of a natural infection.

The VV recombinants protected mice against a VSV challenge that would normally kill all mice within 6 to 12 days. In cattle, two successive vaccinations with the VV recombinant vaccine for VSV provided protection for two-thirds of the animals against a challenge of 100 PFU; this protection was closely correlated with neutralizing antibody titers. The lack of complete protection for all vaccinees may indicate a need for CMI, which is poorly induced by vaccines expressing only the G-protein. It also may be possible to construct recombinants synthesizing considerably more G-glycoprotein, which would provide more effective protection.

3.4 Recombinant Vaccinia Virus Vaccines for SIV

Experimental infection of rhesus macaques with SIV provides a valuable animal model for AIDS vaccine research. Several lines of inquiry recently have brought out the following information:

1. Inactivated whole-virus vaccines have provided limited protective immunity against SIV in macaques challenged with virus propagated in human but not rhesus peripheral blood mononuclear cells.
2. Complete protection from challenge with pathogenic, rhesus-grown SIV was demonstrated reproducibly only in monkeys vaccinated with non-pathogenic (attenuated) *nef*-deleted SIV mutants.

However, whether a live retroviral vaccine could ever meet the safety requirements for human use is highly questionable; this virus integrates into the host genome, induces disease in some animals long after vaccination, and is pathogenic in neonates.

Thus, as an alternative to inactivated and live attenuated classical vaccines, we and others have evaluated the potential of priming rhesus macaques with a live recombinant VV vaccine expressing the glycoproteins of SIV, and then boosting with the same recombinant subunit antigen expressed in mammalian cells, Chinese hamster cells (CHO), or a baculovirus expression vector (see Sect. 4). Protection in these animals was evaluated by intravenous challenge with pathogenic SIV isolates. Although binding antibodies were detected in all animals following immunization with these recombinants, neutralizing antibodies were undetectable one week prior to virus challenge – results that differed from those for macaques vaccinated with whole, inactivated SIV. However, all animals became infected after intravenous inoculation with 1–10 AID_{50} (animal-infectious dose for 50% of subjects) of challenge virus.

There was no protection against infection afforded by any of our recombinant vaccines, even though *in vitro* suppression of virus was demonstrated, and cell-free infectious virus loads in plasma (plasma viremia) of macaques primed and boosted with certain preparations of SIV surface glycoprotein were considerably lower than in unimmunized controls at 2 weeks postchallenge (typically the peak of viremia during acute infection). These studies suggest that immunization with surface glycoprotein may not necessarily provide protective immunity against immunodeficiency virus infection, even though the relationship between reduction in virus load and delay in the onset of disease remains to be evaluated. One protocol did have notable success in heightening the immune response: animals primed with a VV recombinant vaccine and then

boosted with a baculovirus subunit product showed greatly enhanced antibody responses compared to those boosted with the original vaccine.

3.5 Safety and Efficacy of Live Recombinant Vaccines

With the threat of untoward effects always in mind, we have investigated possibilities of further attenuation of our VV recombinant vaccines by the insertion of lymphokine genes like interferon-gamma (IFN-γ) or interleukin-2 (IL-2). This gives us a new approach to enhance safety that should be applicable to other live recombinant vaccines. We constructed fusion proteins of IFN-γ (human and murine species) with various immunogens and demonstrated reduced pathogenicity for immunodeficient (athymic nude) mice with several VV recombinants expressing these proteins. Immunodeficient mice were able to clear the recombinant VV at doses as high as 10^8 PFU, while as few as 10^2 PFU of VV is normally 100% lethal for nude mice. Mixtures of recombinant VVs were always lethal to the mice if one of the recombinant viruses did not express the *IFN-γ* gene, showing that IFN-γ-induced attenuation in VV is only operational in a "cis" fashion.

We have further provided evidence that the antiviral and attenuating activities of IFN-γ have different mechanisms of action. We demonstrated that, unlike antiviral activity, the attenuating activity of IFN-γ is not species specific. Moreover, we have shown attenuating activity of IFN-γ in the absence of any detectable antiviral activity. Although human IFN-γ did not have antiviral activity on murine cells, expression of this cytokine did increase the attenuation of recombinant VVs for athymic mice. It is interesting to note that the anti-inflammatory genes of VV (e.g., its secreted IFN-γ receptor homolog) do not appear to block the attenuating activity of IFN-γ.

The development of an adjuvant system for VV recombinant vaccines may be crucial if the technology is to be considered an effective alternative to classical forms of vaccination, such as inactivated or live attenuated vaccines. We demonstrated that IFN-γ mixed with G-glycoprotein of VSV reproduces the adjuvant effect of complete Freund's adjuvant and further increases the anamnestic response on booster administration. Others demonstrated the adjuvant effects of IL-2.

The use of immunoregulatory genes like *IFN-γ* and *IL-2* as adjuvant genes in multivalent VV recombinant vaccines must be further evaluated. If these lymphokine genes work as adjuvants to augment the immune response, the concept of vaccination will be revolutionized. Such natural products could eliminate the need for harsher adjuvants (e.g., Freund's) and thus avoid the untoward effects presently observed. The human *IFN-γ* gene could be incorporated into VV vectors for human pathogens, and similar vectors could be constructed for other domestic species.

Another method to increase the safety and efficacy of recombinant VVs is to inactivate or delete virulence genes. The NYVAC VV was constructed by deleting 18 *VV* genes from the extremities of the virus in which the virulence and species-specificity genes are encoded. This virus is similar to MVA, causing an aborted infection in mammalian cells but replicating in chicken embryo cells. These viruses generally require multiple boosts to generate protective immunity; thus, we have been exploring the effect of deleting VV virulence genes individually. VV is one of many

viruses with genes that code for specific proteins to suppress host immunity. These immunomodulating genes are usually not essential for virus growth *in vitro*. They include homologs of cytokine receptors, complement control proteins, and serine protease inhibitors (serpins). In particular, VV contains a gene (B8R) that codes for a secreted protein with sequence similarity to the extracellular domain of the IFN-γ receptor. The B8R protein neutralizes the antiviral activities of IFN-γ from several species. Serpins play important roles in the regulation of immune and inflammatory responses as well as cell death. The B13R (SPI-2) gene codes for a serpin homolog that interferes with host inflammatory responses by inhibiting the proteolytic activity of caspase-1, also known as IL-1β converting enzyme, and granzyme B. Similarly, the (SPI-1) B22R gene plays a role in reducing the host's immune responses to the virus. We hypothesized that deleting these immunomodulating genes would increase the efficacy and safety of recombinant VVs. While deletion of these genes had no effect on virus replication *in vitro*, recombinant VVs lacking these genes were attenuated for mice. There was a significant decrease in weight loss and mortality in normal mice, and immunodeficient nude mice survived significantly longer than controls inoculated with parental virus. Finally, the deletion of these genes did not reduce humoral immune responses. Mice and rats vaccinated with the recombinant VVs showed identical humoral responses to both homologous and heterologous genes expressed by VV.

3.6
Potential Recombinant Vaccines for HIV

Several laboratories are investigating VV recombinants as vaccines for human immunodeficiency virus (HIV) infection. It has been shown that the HIV antigens expressed by some of these recombinants are correctly processed and transported to the appropriate cellular compartments. Moreover, experimental animals as well as human volunteers that were vaccinated were shown to develop detectable CMI to the expressed HIV proteins. There is considerable interest and some optimism that an effective vaccine might be prepared using VV as a vector.

Innovative solutions to the safety problems occasioned by putting such a vaccine into general use are gaining attention, especially following the report of a disseminated VV infection caused by smallpox vaccination of an army recruit with subclinical HIV infection. It has been demonstrated that VV is attenuated by insertional inactivation of its *TK* gene. Also, the expression of IL-2 and murine IFN-γ has prevented disseminated VV infection in immunodeficient nude mice.

4
Baculovirus Expression Vectors and Rapid Diagnostic Kits

The VV recombinant vaccines prepared for rinderpest, PPR, VS, and SIV, all lend themselves to employment of a new rapid diagnostic technique that our laboratory has prepared in kit form. The baculovirus expression vectors used in our laboratory give us the capacity to produce large quantities of different virus proteins that may be used as coating antigens for ELISAs. When an internal protein gene such as the *N* gene of RPV or VSV, or the gag gene of SIV, is incorporated into the genome of this insect virus strain before infection of a larval host, the virus polyhedrin promoter drives a very

high level of expression of recombinant protein (up to 67% of the weight of the larva). The protein from one infected and homogenized larva (0.2–0.3 g) provides antigen sufficient for coating 200 standard 96-well ELISA plates, which can be used to diagnose up to 10 000 serum samples in duplicate. Purification of the N or gag protein appears to be unnecessary; there was no interference or loss of sensitivity using the crude larval lysate for coating ELISA plates.

Used in conjunction with a recombinant vaccine incorporating a gene for a major immunogen of the disease agent, these ELISAs allow distinction of vaccinated animals from those that have been naturally infected. Vaccination with the recombinant produces antibody to the incorporated gene's product only, while natural infection brings about responses to many different antigens. This test differentiates animals showing responses to a variety of proteins (naturally infected by the disease agent) from those that respond only to the vaccine's foreign gene product (vaccinees).

We have also introduced a novel approach for the production of monoclonal antibodies to one specific protein of a virus or other agent consisting of several proteins, without the use of purified antigen in either the immunization or screening phase of the procedure. This approach utilizes a VV recombinant expressing the gene of the target protein for the immunization of mice and the whole virus for screening of the hybridoma to identify the desired monoclonal antibodies. This system has been effective in producing monoclonal antibodies to rare antigens like Alzheimer's protein, or to those – the VSV nucleocapsid (N) protein, for example – that cannot be easily screened for biological activities such as hemagglutination or neutralization.

5 Advantages of Vaccinia Virus Recombinants

Experience with mass production, storage, and administration of smallpox vaccine suggests that VV or related poxvirus recombinants could be the basis for veterinary and human vaccines offering advantages over current preparations. Lyophilized VV is heat stable and can be effectively administered by a variety of methods. During the global eradication of smallpox, scarification or jet gun administration was used to vaccinate millions of people. We have vaccinated animals by intradermal inoculation and scarification, and we plan to investigate the effectiveness of oral administration for wildlife and cattle.

It is possible to make a polyvalent vaccine using VV as the vector, by expressing several genes that represent a number of unrelated agents. While the manipulations for including several genes are increasingly difficult with each addition, new methods are being developed rapidly. The possibility of developing a polyvalent vaccine, with a single VV recombinant expressing a number of heterologous genes, is especially important for nomadic regions of Africa and Asia. Theoretically, one stock of VV could carry genes from most of the important livestock diseases in a region, and one inoculation could vaccinate an animal against half-a-dozen or more diseases.

The VV recombinant vaccine contains very few antigens in common with disease agents, and so we have found it possible to develop diagnostic kits for rinderpest, vesicular stomatitis, and simian AIDS that can distinguish vaccinated animals from those exposed to the whole agent either by infection or vaccination. Currently, we are expanding this technology to provide

rapid diagnosis for a variety of livestock diseases.

We recognize the potential safety problems with widespread use of VV as a vector for different vaccines. However, the strains used for vaccine construction were used worldwide without serious problems in the successful eradication of smallpox. Further, it has been demonstrated that insertional inactivation of the *TK* gene attenuates the virus. In constructing the first rinderpest double recombinant vaccine (vRVFH), the *TK* gene of VV was inactivated by insertion of the RPV *H* gene; insertional inactivation of the VV HA region by incorporating the *F* gene of RPV has led to further attenuation of the recombinant virus. The absence of detectable pock lesions at the site of vaccination by the double recombinant is an indication of its greater attenuation, even though it is still completely protective. Moreover, vRVFH is highly attenuated and is apathogenic for both athymic nude mice and SIV-infected rhesus macaques.

To improve efficacy, we constructed a second generation VV double recombinant vaccine for rinderpest (v2RVFH) that expresses both the *F* and *H* genes of RPV under strong synthetic VV promoters at the TK site. v2RVFH-infected cells express higher levels of the F and H glycoproteins than vRVFH and show extensive syncytium formation.

It is important that safety considerations be thoroughly addressed before live recombinant viruses are introduced into the environment. The rapidly expanding epidemic of HIV, particularly in Africa, demands that live vaccines do not pose a threat to indigenous populations. The safety of VV recombinant vaccines has been confirmed by the absence of transmission of VV from vaccinated to contact animals. No pock lesions or antibody to VV, as measured by SN and plaque reduction assays, could be demonstrated in control animals housed in contact with cattle vaccinated with vRVFH or v2RVFH. We believe that these outstanding safety features of these vaccines make them suitable for field use. We believe also that these vaccines are only the first to bring the benefits of modern molecular biology and biotechnology to disease control. As knowledge and expertise in these fields expand, recombinant vaccines will be constructed for many of the more stubborn afflictions of human beings and animals, providing creative solutions to disease problems, old and new.

See also Gene Targeting; Medicinal Chemistry.

Bibliography

Books and Reviews

Brown, F. (Ed.) (1994) *Recombinant Vectors in Vaccine Development: Developments in Biological Standardization*, Karger, New York.

Chanock, R.M., Giensburg, H., Lerner, R., Brown, F. (Eds.) (1987) *Vaccines 87. Modern Approaches to New Vaccines Including Prevention of AIDS*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

National Academy of Sciences & National Research Council (1988) *Science & Technology for Development: Prospects Entering the 21st Century*, NAS/NRC, Washington, DC.

Quinnan, G.V. (Ed.) (1985) *Vaccinia Viruses as Vectors for Vaccine Antigens*, Elsevier, North Holland, New York.

Talwar, G., Rao, K., Chauhan, V. (Eds.) (1994) *Recombinant and Synthetic Vaccines*, Narosa Publishing House, New Delhi.

Yilma, T. (1987) Adjuvant genes for infectious virus recombinant vaccines: use of lymphokines as adjuvants, *Biotechnol. U.S.A.* 238–246.

Yilma, T. (1989) Prospects for the total eradication of rinderpest, *Vaccine* **7**, 484–485.

Yilma, T. (1990) A modern vaccine for an ancient plague: rinderpest, *Biotechnology* **8**, 1007–1009.

Yilma, T. (1991) The Role of Biotechnology in Tropical Diseases, in: Williams, J., Kocan, K., Gibbs, E.P. (Eds.) *Tropical Veterinary Medicine: Current Issues and Perspectives, Ann. N.Y. Acad. Sci.*, 653, R.D.S. Branckaert, Rome, Italy, pp. 1–6.

Yilma, T. (1991) A vaccinia virus recombinant vaccine for rinderpest, *World Animal Rev.* (a Quarterly Journal of the FAO/UN on Animal Health, Production, and Products) **69**, 2–6.

Primary Literature

Ada, G. (1991) Vaccine development. Real and imagined dangers, *Nature* **349**, 369.

Ahmad, S., Bassiri, M., Banerjee, A.K., Yilma, T. (1993) Immunological characterization of the VSV nucleocapsid (N) protein expressed by recombinant baculovirus in *Spodoptera exigua* larva: use in differential diagnosis between vaccinated and infected animals, *Virology* **192**, 207–216.

Ahmad, S., El-Amad, Z., Giavedoni, L., Marthas, M., Haigwood, N.L., Kilpatrick, J., Scandella, C.J., Gardner, M.B., Luciw, P.A., Yilma, T. (1993) Immune Response of Rhesus Macaques to Recombinant SIV gp160 Does Not Protect From Challenge Infection, in: Ginsberg, H.S. (Ed.) *Vaccines 93: Modern Approaches to New Vaccines Including Prevention of AIDS*, Cold Spring Harbor Laboratory Press, Plainview, New York, pp. 57–62.

Ahmad, S., Lohman, B., Marthas, M., Giavedoni, L., El-Amad, Z., Haigwood, N.L., Scandella, C.J., Gardner, M.B., Luciw, P.A., Yilma, T. (1994) Reduced virus load in rhesus macaques immunized with recombinant gp160 and challenged with simian immunodeficiency virus, *AIDS Res. Hum. Retroviruses* **10**, 195–204.

Anderson, K.P., Fennie, E.H., Yilma, T. (1988) Enhancement of a secondary antibody response to vesicular stomatitis virus "G" protein by IFN-γ treatment at primary immunization, *J. Immunol.* **140**, 3599–3604.

Barrett, T., Rossiter, P.B. (1999) Rinderpest: the disease and its impact on humans and animals, *Adv. Virus Res.* **53**, 89–110.

Bassiri, M., Ahmad, S., Giavedoni, L., Jones, L., Saliki, J.T., Mebus, C., Yilma, T. (1993) Immunological responses of mice and cattle to baculovirus-expressed F and H proteins of rinderpest virus: lack of protection in the presence of neutralizing antibody, *J. Virol.* **67**, 1255–1261.

Carroll, M.W., Moss, B. (1997) Poxviruses as expression vectors, *Curr. Opin. Biotechnol.* **8**, 573–577.

Giavedoni, L., Ahmad, S., Jones, L., Yilma, T. (1997) Expression of gamma interferon by simian immunodeficiency virus increases attenuation and reduces postchallenge virus load in vaccinated rhesus macaques, *J. Virol.* **71**, 866–872.

Giavedoni, L.D., Jones, L., Gardner, M.B., Gibson, H.L., Ng, C.T., Barr, P.J., Yilma, T. (1992) Vaccinia virus recombinants expressing chimeric proteins of human immunodeficiency virus and γ interferon are attenuated for nude mice, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 3409–3413.

Giavedoni, L., Jones, L., Mebus, C., Yilma, T. (1991) A vaccinia virus double recombinant expressing the F and H genes of rinderpest virus protects cattle against rinderpest and causes no pock lesions, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8011–8015.

Giavedoni, L.D., Planelles, V., Haigwood, N.L., Ahmad, S., Kluge, J.D., Marthas, M.L., Gardner, M.B., Luciw, P.A., Yilma, T.D. (1993) Immune response of rhesus macaques to recombinant simian immunodeficiency virus gp130 does not protect from challenge infection, *J. Virol.* **67**, 577–583.

Giavedoni, L.D., Yilma, T. (1996) Construction and characterization of replication-competent simian immunodeficiency virus vectors that express gamma interferon, *J. Virol.* **70**, 2247–2251.

Grubman, M.J., Mebus, C., Dale, B., Yamanaka, M., Yilma, T. (1988) Analysis of the polypeptides synthesized in rinderpest virus-infected cells, *Virology* **163**, 261–267.

Hsu, D., Yamanaka, M., Miller, J., Dale, B., Grubman, M., Yilma, T. (1988) Cloning of the fusion gene of rinderpest virus: comparative sequence analysis with other morbilliviruses, *Virology* **166**, 149–153.

Ismail, T., Ahmad, S., D'Souza-Ault, M., Bassiri, M., Saliki, J., Mebus, C., Yilma, T. (1994) Cloning and expression of the nucleocapsid gene of virulent Kabete O strain of

rinderpest virus in baculovirus: use in differential diagnosis between vaccinated and infected animals, *Virology* **198**, 138–147.

Ismail, T.M., Yamanaka, M.K., Saliki, J.T., el-Kholy, A., Mebus, C., Yilma, T. (1995) Cloning and expression of the nucleoprotein of peste des petits ruminants virus in baculovirus for use in serological diagnosis, *Virology* **208**, 776–778.

Jones, L., Ahmad, S., Chan, K., Verardi, P., Morton, W.R., Grant, R., Yilma, T. (2000) Enhanced safety and efficacy of live attenuated SIV vaccines by prevaccination with recombinant vaccines, *J. Med. Primatol.* **29**, 231–239.

Jones, L., Giavedoni, L., Saliki, J.T., Brown, C., Mebus, C., Yilma, T. (1993) Protection of goats against peste des petits ruminants with a vaccinia virus double recombinant expressing the F and H genes of rinderpest virus. *Vaccine* **11**, 961–964.

Jones, L., Ristow, S., Yilma, T., Moss, B. (1986) Accidental human vaccination with vaccinia virus expressing nucleoprotein gene, *Nature* **319**, 543.

Jones, L., Tenorio, E., Gorham, J., Yilma, T. (1997) Protective vaccination of ferrets against canine distemper with recombinant pox virus vaccines expressing the H or F genes of rinderpest virus, *Am. J. Vet. Res.* **58**, 590–593.

Kieny, M.P., Lathe, R., Drillien, R., Spehner, D., Skory, S., Schmitt, D., Wiktor, T., Koprowski, H., Lecocq, J.P. (1984) Expression of rabies virus glycoprotein from a recombinant vaccinia virus, *Nature* **312**, 163–166.

Mackett, M., Yilma, T., Rose, J.K., Moss, B. (1985) Vaccinia virus recombinants: expression of VSV genes and protective immunization of mice and cattle, *Science* **227**, 433–435.

Moss, B. (1996) Genetically engineered poxviruses for recombinant gene expression, vaccination, and safety, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 11341–11348.

Pastoret, P.P., Brochier, B. (1996) The development and use of a vaccinia-rabies recombinant oral vaccine for the control of wildlife rabies; a link between Jenner and Pasteur, *Epidemiol. Infect.* **116**, 235–240.

Seet, B.T., Johnston, J.B., Brunetti, C.R., Barrett, J.W., Everett, H., Cameron, C., Sypula, J., Nazarian, S.H., Lucas, A., McFadden, G. (2003) Poxviruses and immune evasion, *Annu. Rev. Immunol.* **21**, 377–423.

Verardi, P.H., Aziz, F.H., Ahmad, S., Jones, L.A., Beyene, B., Ngotho, R.N., Wamwayi, H.M., Yesus, M.G., Egziabher, B.G., Yilma, T.D. (2002) Long-term sterilizing immunity to rinderpest in cattle vaccinated with a recombinant vaccinia virus expressing high levels of the fusion and hemagglutinin glycoproteins, *J. Virol.* **76**, 484–491.

Verardi, P.H., Jones, L.A., Aziz, F.H., Ahmad, S., Yilma, T.D. (2001) Vaccinia virus vectors with an inactivated gamma interferon receptor homolog gene (B8R) are attenuated in vivo without a concomitant reduction in immunogenicity, *J. Virol.* **75**, 11–18.

Warren, J. (2002) Preclinical AIDS vaccine research: survey of SIV, SHIV, and HIV challenge studies in vaccinated nonhuman primates, *J. Med. Primatol.* **31**, 237–256.

Wiktor, T.J., Macfarlan, R.I., Reagan, K.J., Dietzschold, B., Curtis, P.J., Wunner, W.H., Kieny, M.P., Lathe, R., Lecocq, J.P., Mackett, M., Moss, B., Koprowski, H. (1984) Protection from rabies by a vaccinia virus recombinant containing the rabies virus glycoprotein gene, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7194–7198.

Yamanaka, M., Dale, B., Crisp, T., Cordell, B., Grubman, M., Yilma, T. (1992) Sequence analysis and editing of the phosphoprotein (P) gene of rinderpest virus, *Virology* **190**, 553–556.

Yamanaka, M., Hsu, D., Crisp, T., Dale, B., Grubman, M., Yilma, T. (1988) Cloning and sequence analysis of the hemagglutinin gene of the virulent strain of rinderpest virus, *Virology* **166**, 251–253.

Yilma, T., Anderson, K., Brechling, K., Moss, B. (1987) Expression of an Adjuvant Gene (Interferon-γ) in Infectious Vaccinia Virus Recombinants, in: Chanock, R.M. (Ed.) *Vaccine 87: Modern Approaches to New Vaccines: Prevention of AIDS and Other Viral, Bacterial, and Parasitic Diseases*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 393–396.

Yilma, T., Hsu, D., Jones, L., Owens, S., Grubman, M., Mebus, C., Yamanaka, M., Dale, B. (1988) Protection of cattle against rinderpest with vaccinia virus recombinants expressing the HA or F gene, *Science* **242**, 1058–1061.

Yilma, T., Mackett, M., Rose, J.K., Moss, B. (1985) Vaccinia Virus Recombinants Expressing Vesicular Stomatitis Genes Immunize Mice and Cattle, in: Quinnan, G.V. (Ed.) *Vaccinia Viruses as Vectors for Vaccine Antigens: Proceedings of the Workshop on Vaccinia Viruses as Vectors for Vaccine Antigens*, Chevy Chase,

Maryland, USA, Nov. 13–14, 1984, Elsevier, New York, pp. 187–200.

Yilma, T., Owens, S., Fennie, E.H., Anderson, K.P. (1989) Enhancement of primary and secondary immune responses by interferon-gamma, *Adv. Exp. Med. Biol.* **251**, 145–152.

Yilma, T., Perryman, L.E., McGuire, T.C. (1982) Deficiency of interferon-γ but not interferon-β in Arabian foals with severe combined immunodeficiency, *J. Immunol.* **129**, 931–933.

Yilma, T., Ristow, S.S., Moss, B., Jones, L. (1987) A novel approach for the production of monoclonal antibodies using infectious vaccinia virus recombinants, *Hybridoma* **6**, 329–335.

Genetic Intelligence, Evolution of

David S. Thaler and Bradley T. Messmer
Rockefeller University, New York, NY, USA

1 **Historical and Intellectual Context 316**

2 **Cairns and Precursors 318**

3 **Critical Reading of the Adaptive Mutation Literature 320**

4 **Mechanisms Proposed for Adaptive Mutagenesis 321**

5 **How Might Genetic Intelligence Have Evolved? 322**

6 **Genetic Algorithms 324**

Acknowledgments 329

Bibliography 329

Keywords

DNA Polymerase
Enzymes that make DNA using DNA as a template. There are several different DNA polymerases in most cells and they have somewhat different properties with respect to their error rate and ability to use noncanonical templates (examples of noncanonical templates are ribonucleotides in DNA, 8-oxy-Guanine and other chemical lesions in bases or backbone).

Evolution
Change with time or in the series (generations) of a lineage in the same direction as the flow of time.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

Genetics
Mechanisms by which information is transferred between generations. The term ''genetics'' has its origins in biology and ''generations'' refers to progeny cells, viruses, or multicellular organisms. The genetic properties of descent with modification are present in other contexts, such as computer programs known as *genetic algorithms*.

Genetic Evolution
Change in the hereditary information with time, especially if transferred to subsequent generations.

Genetic Algorithms
A class of computer algorithms with certain characteristics. GAs consist of bit strings that give rise to progeny bit strings. Some progeny are identical to the parent and some are different. The population of bit strings compete with each other for survival and reproduction according to selection criteria mediated by the program environment.

Genetic Change
Alterations of hereditary information. Typically – but not always – genetic change occurs concomitantly with replication of the organism.

Mutation
A genetic change in which preexisting hereditary information is not transferred to a new context.

Recombination
A genetic change in which preexisting hereditary information is transferred to a new context.

Recombination, Homologous, Legitimate, Generalized
Recombination in which the similarity of two sequences is a requisite for their ability to recombine. To a first approximation, the homologous recombination depends on the similarity not on the specific sequence.

Stress
A differentiated physiological state in response to nonoptimality perceived by the organism. SOS in response to the stress of DNA damage was the first characterized stress response. Others include responses to osmotic stress, the presence of unfolded proteins in the cytoplasm, and various nutritional starvations. Various stress responses tend to partially overlap.

Intelligence
Tools and mechanisms for receiving information from the environment and adjusting responses; problem solving – arguments over types and measures of intelligence abound.

Genetic Intelligence
Hereditary information responsible for the mechanisms that generate genetic change.

Environment
Systems are often defined as having an ''inside'' and an ''outside.'' The environment of a system consists of anything ''outside.'' The outside of an organism is defined by its skin; the outside of a cell is defined by its membrane. The outside of a chromosome is the cellular milieu and the outside of a DNA base is all of the above as well as its sequence context.

Random Mutation
A mutation that is no more likely to occur in environments that select for it than in environments that do not. In the context of genetics and mutation, ''random'' does not mean that every possible change is equally likely to occur. Even in the narrow context of point mutations, there are inherent biases in all examined systems.

Adaptive, Directed, or Cairnsian Mutation
A mutation that is more likely to occur in circumstances in which the mutation is advantageous, that is, nonrandom mutation. ''Strong'' adaptive mutations are those that do not appear to be generalized stress-responses and that are highly specific.

Lethal Environmental Selection
Environments that kill unfit organisms and kill them quickly, that is, before they have the ability to mount a stress response that might prolong their agony, or, allow them to consider ways to escape the problematic situation of the selection.

Nonlethal Environmental Selection
Environments that prevent, stop, or slow the growth of unfit organisms without killing them. These situations provide the organism with problems that it has the opportunity to solve over time. Often a stress response may be involved in an organism's attempt to solve the problem of nonlethal or slowly lethal selections.

Fixation of Genetic Change in an Individual Organism
An irrevocable commitment that a new allele will be inherited by progeny.

The purpose of this article is fourfold: (1) To frame the intellectual context of ''adaptive'', ''directed'', or ''Cairnsian'' mutation. This area of research is – and will likely remain – controversial. (2) To review the points to be considered when critically reading in the field; (3) to consider the question of how mechanisms for an ''intelligent'' generation of diversity could evolve and, in this context; (4) to suggest connections between the fields of genetic algorithms in computer science and those of biological evolution proper (carbon – rather than silicon – based life forms).

1
Historical and Intellectual Context

Lamarckian interpretations persisted longer in microbiology than in other domains of biology. In part, this was due to the plethora of traits and organisms being observed and interpreted with a limited set of conceptual tools. Different traits and mechanisms turned out to be involved, which are now realized to include at least five classes: operon induction, mutation, transposition, chromosome rearrangement, especially duplication, and phase variation such as those mediated by the site-specific recombination systems, lysogenic transformations, and other extragenic elements. The difficulty in inferring genotype from phenotype for these quite different mechanisms contributed to the confusion. The polemic reached high levels and resolution of the controversy – albeit temporary – had a major effect on the development of modern molecular biology.

Luria and Delbruck made an enormous dent in the debate with their classic experiments that were designed to address one aspect of the Generation of Diversity (GOD) question: "Do mutations arise in response to the conditions that select for them or are mutagenic variants preexisting in the population but unseen in the absence of selection?" The Luria and Delbruck experiments are of the following form. A number of replica liquid cultures were each inoculated with a small number of bacteria. "Small" means few enough cells such that no T1 phage–resistant cells are present to begin with. The cultures were then allowed to grow to saturation and the number of mutant cells in a sample was determined. Mutant cells were enumerated by spreading samples of the cultures onto petri plates that had been seeded with several million T1 phages. Only cells resistant to T1 phage could give rise to colonies. The question is, "Did the T1-resistant phage exist in the liquid culture before there existed any selection for them or was the resistance to T1 phage induced by the selective condition itself?" In the former case, the role of the environment is restricted to selection for preexisting mutants; in the latter case, the environment also plays a role in generating the mutants it subsequently selects. Luria and Delbruck made the quantitative argument that resistance to T1 phage arose in the absence of selection. The argument was based on two elements: (1) If each cell has a given probability (e.g. one in a million) of mutating in response to the challenge, when a large number (typically one hundred million cells present in one-tenth of a milliliter of a saturated overnight culture of the bacteria) of bacteria are spread onto a plate, the number of resistant mutants found on each plate should represent a Poisson distribution around the arithmetic average (e.g. 100 in the numerical example given). (2) On the other hand, suppose that mutants arise during the nonselective overnight growth at the same rate of one in a million. Mutant cells, arising before the culture is saturated, will give rise to mutant progeny, and cultures will therefore contain clones of mutants rather than single mutants. Luria and Delbruck's model in which mutants arise independently of their subsequent selection predicted a wide (i.e. wider than Poisson distribution) variation in the number of mutants in a series of cultures. The quantitative analysis of Luria and Delbruck assumes that mutation occurs at a constant probability per cell division.

The respreading experiments of Newcomb, as well as the replica plating of

Lederberg and Lederberg were both advanced as confirmation of the conclusion reached by Luria and Delbruck. Replica plating used sib-selection to achieve a direct demonstration that clones of resistant mutants (to the drug streptomycin) existed prior to and independent of selection.

The formulation of Luria and Delbruck states that mutation is a function of the number of cell generations with a fixed probability of mutation at each generation. This way of thinking about mutation resonated well with a molecular mechanism that involved misinsertion of bases during normal replication. It also fits with a model of gradualistic evolution that occurs via the misinsertion of bases during normal replication. Molecular clocks based on a constant rate of random-base substitution made tidy icing on the hostess twinkee of the neo-Darwinian synthesis.

Many interpretations of these three classic experiments erroneously conflate the elements of clonality with the time that a mutant arises. The incorrect assumption being that if mutants arise independent of the selection, then mutants will give rise to clones before the assay. On the other hand, if mutants arise after the selection is imposed, then each mutation event will result in a separate enumerable colony. In the former case, enumerating the number of mutants is an indirect measure of the mutation rate; in the latter case, the rate of mutation is given simply by counting the number of mutant cells and dividing it by the total number of cells tested.

A shrill stridency associated with the view that all mutations arise without regard to their subsequent selection is well illustrated by Monod in his 1970 book *Chance and Necessity*:

> In modern biological research, some of the most outstanding work, both as to methodology and to significance, bears upon the area known as molecular genetics (Benzer, Yanofsky, Brenner and Crick). In particular, this work has made it possible to analyze the different types of discrete accidental alterations a DNA sequence may suffer. Various mutations have been identified as due to
>
> 1. The substitution of a single pair of nucleotides for another pair;
> 2. The deletion or addition of one or several pairs of nucleotides; and
> 3. Various types of "scrambling" of the genetic text by inversion, duplication, displacement, or fusion of more or less extended segments.
>
> We call these events accidental; we say that they are random occurrences. And since they constitute the *only* possible source of modifications in the genetic text, itself the *sole* repository of the organism's hereditary structures, it necessarily follows that chance *alone* is at the source of every innovation, of all creation in the biosphere. Pure chance, absolutely free but blind, at the very root of the stupendous edifice of evolution: this central concept of modern biology is no longer one among other possible or even conceivable hypotheses. It is today the *sole* conceivable hypothesis, the only one that squares with observed and tested fact. And nothing warrants the supposition – or the hope – that on this score our position is likely ever to be revised (pp. 112–113 Italics are in the original)

Monod goes on to say:

> There is no scientific concept, in any of the sciences, more destructive of anthropocentrism than this one, and no other so rouses an instinctive protest from the intensely telonomic creatures that we are.

Monod is correct in the consequences of his last statement but he is wrong in his dogmatic emphasis on the impossibility of contradiction. The authors of the present article believe that "modern biological research" including molecular genetics does not preclude the opportunity for

theory and experiment in which mind and nature are not *a priori* at odds.

2
Cairns and Precursors

In 1988, a paper that continues to stimulate intense work and controversy among molecular biologists, geneticists, evolutionists, and even philosophers was published. The results suggest that some mutations arise more frequently when they are being selected for, in apparent contradiction to the idea that mutations arise without regard for their utility. The role of the environment in selecting among preexisting variants is not challenged by these approaches; rather the environment is endowed with roles in generating the variation upon which selection acts.

An essential aspect of protocols that search for adaptive mutagenesis is the use of nonlethal selection, for example, lactose prototrophy. Luria and Delbruck's selection was for resistance to phage T1. A cell becomes resistant to T1 by mutating its gene for the phage receptor. *Escherichia coli* has about 200 phage receptors on its surface. Even if a bacterium was "smart" enough to immediately mutate the gene for the receptor as soon as it "saw" itself threatened by phage, it would still be dead by virtue of an already existing protein. Phage receptors are passed along with the wall at cell division. Vulnerability to phage lysis continues for several generations even after a mutation prevents the gene from encoding a new phage receptor. For this reason, the experimental protocol of phage resistance was only capable of documenting mutations that had occurred several generations before selection was applied. This protocol was only able to address the point that mutants can and do arise in the absence of selection. It was never capable of addressing the question of whether selection can induce mutation. A similar limitation is inherent in the Lederberg and Lederberg experiments because of the nature of the streptomycin-resistance assay. In order for the resistance phenotype to be expressed in a cell, sensitive ribosomes must be diluted out by growth after the mutation has occurred.

Cairn's 1988 paper and the questions raised by it were not without precursors. Some of these precursors will be mentioned with an eye to what was different in the formulation of Cairns.

Max Delbruck may have been the first to recognize the limitations that could be rigorously inferred from his paper with Luria. In the 1946 Cold Spring Harbor Symposium (published in 1947), Delbruck says "... In view of our ignorance of the causes and mechanisms of mutations, one should keep in mind the possible occurrence of specifically adaptive mutations...or should exclude their occurrence in a variety of specific cases."

The conclusion that mutation occurs as a consequence of errors arising during replication is incomplete on several counts. Frances Ryan, in the late 1950s and early 1960s, demonstrated that prototrophic revertants of *E. coli* histidine auxotrophs could accumulate while the cells were not growing. Meiosis is another case in which genetic change is massively increased during a single replicative generation. Interallelic recombination per cell division in yeast meiosis is increased between 10^3 and 10^4 when compared to the mitotic recombination rate between the same alleles. Radiation and chemical stimulation of genetic change certainly make it clear that the environment can

create the lesions in DNA, which are subsequently processed to fix mutations. Premutagenic lesions of many sorts in many sorts of cells are induced by chemicals or radiation in nongrowing cells and require subsequent growth for enumeration. That they may also require growth for their fixation has the makings for adaptive mutagenesis.

Mutagenic mechanisms can alter replication fidelity even at nonlesioned sites. One of the products of the SOS DNA damage-repair response regulon acts as a negative regulator of the $5' > 3'$ exonuclease of DNA polymerase. When the $5' > 3'$ exonuclease functions, the polymerase is between 10^3 and 10^4 times more accurate. This nuclease activity is known as *editing* or *proofreading* because it preferentially excises incorrectly, that is, non-Watson–Crick-paired terminal nucleotides. During the SOS response, overall DNA replication is decreased in fidelity. The qualitative nature of genetic change is also altered during the SOS response, that is, some types of mutations arise relatively more frequently than others. Nontargeted mutagenesis during the SOS response was the subject of considerable polemic around ideas of ''inducible evolution'' that were the immediate precursors of controversy around specifically adaptive mutation.

Barbara McClintock considered genomic change to be an important mechanism for both differentiation and evolution. These ideas were formed before the rise of the operon theory, partly in the context of somatic adaptation. McClintock remained concerned with the possibilities and properties of adaptive genomic change. In her 1983 Nobel prize lecture, she said ''A goal for the future would be to determine the extent of knowledge the cell has of itself and how it utilizes this knowledge in a ''thoughtful'' manner when challenged.'' James Shapiro, in the discussion of his 1984 paper, pointed out that the Luria and Delbruck, Newcomb and Lederberg and Lederberg experiments involved immediately lethal selection and could only detect mutants that had originated in the absence of selection. He ended this important paper with the statement: ''Indeed, now that we know about mobile genetic elements, inducible mutator systems and multiple biochemical activities that reorganize DNA molecules, the most pertinent questions in studies of hereditary change must be questions of control and regulation.''

The special contribution of Cairns and colleagues is that the distinction between generalized stress responses and ''thoughtful'' reactions of the organism was for the first time explicitly incorporated into experimental protocols. The adequacy of the methods used has been the subject of much criticism and refinement, but the field of adaptive mutation, while not getting calmer, has – on average – been reaching higher intellectual levels ever since.

If Ryan had included nonselected genes in his experiments, adaptive mutation as presently discussed would have been discovered before 1960. Shapiro's 1984 paper does not include nonselected genes; when the experimental design was extended to include tests inherent in Cairns et al. 1988, the results of Mittler and Lensky indicated a more generalized stress response. On the other hand, more recent work by Shapiro in 1994 indicated that the sequence of mutations was different depending on whether the mutations arose under selective or nonselective conditions.

3
Critical Reading of the Adaptive Mutation Literature

This field of science is likely to remain populated by mixed herds of sheep, goats, and unicorns. To aid critical reading, some hints and questions to keep in mind while examining the literature will be enumerated.

1. If mutants occur prior to selection but they grow slowly, then the appearance of Luria–Delbruck "jackpots" will be suppressed. Slow growers also may come up later under selection giving a false clue that the mutation arose after the selection was imposed.
2. Always consider the possibility of, and the amount of, growth under conditions where growth is not supposed to happen. For example, in cases where a lactose negative cell is asked to mutate such that it can grow on medium with lactose as the sole carbon source, might there be a smidgen of glucose in the lactose medium? Even if cell numbers do not increase, there may be cell turnover via cannibalism. Is there a chance for cross-feeding between cell types that are present, or could cross-feeding arise under the selective conditions? In the absence of growth, are other processes such as phage induction, conjugation, or meiosis still possible and how might they contribute to the results observed?
3. Is the type of change occurring under selection the same type that occurs during nonselective growth? If the spectrum of mutations recovered is different, then the evidence is much stronger that a different mechanism is operating under one condition than another. One must be on the lookout for unexpected sources of selection and for reconstructions to address them.
4. Are the genetic dependencies of mutation the same under selective and nonselective conditions? Different genetic dependencies are bona fide evidence of different pathways, although one must always be cautious of the "kicking 'em when they're down" syndrome, that is, some mutations may result in a cell that is weaker and not able to grow well under challenged conditions.
5. Is the selection completely lethal or nonlethal? If nonlethal, how much of a stress is it? Is there a general stress response pathway – such as SOS, carbon starvation (cAMP), or amino acid starvation (stringent response), which could be invoked under the experimental conditions? In what ways does the experimental design distinguish between known stress responses and keep in mind the possibility of new ones?
6. What is the evidence for specificity? Are there control genes that are not under selection but whose mutation or recombination rate can be measured? What is the relation of control genes and their assays to the experimental ones? The best experimental designs have symmetry, that is, selections for one or the other gene with measurement of the rate of change in both. Are the results reported as both absolute quantities of change (e.g. mutation rate per cell) and as a ratio of different types of change? Is the evidence for "adaptive" mutation based only on the time of mutants arising or is it also based on ratios and the spectrum of mutation? For example, it has been noted that some replicons give a higher mutation rate under some physiological conditions. Again, this speaks about the consideration of experimental design and the necessity to be both rigorous and open-minded in considering both

what the results "make you believe" and what they "allow you to believe."

4
Mechanisms Proposed for Adaptive Mutagenesis

Many mechanisms have been proposed for the apparently Lamarckian inheritance that occurs under some regimes of nonlethal selection. In this section, several such mechanisms will be mentioned. Some of the following classifications are orthogonals. This list should be considered rather like a deck of cards of ideas, which the thinker can draw upon in interpreting results and controversies.

1. Mechanisms in which the mutation rate is increased over a small portion (say 1%) of the genome in each cell such that in a population all regions will be highly mutagenized. This category was proposed by Barry Hall who subsequently provided counterevidence to the corollary that sequence changes should be clustered.
2. Mechanisms in which the global generation of mutations is increased in response to properly perceived stress (PPS). This class of events is to generalize most directly from SOS-induced mutagenesis and from the facts of meiosis, that is, that there are physiological circumstances of PPS in which the overall mutation rate is altered. Slightly more focused and specific are models in which the mutation rate for a subset of the total genes in the organism is altered as a function of some condition. Candidate mechanisms to focus mutation onto a subset of genes include genes and processes that regulate operon and regulon expression. The induction of specialized DNA polymerases and/or the inhibition of polymerase editing can enter via this route.
3. Mechanisms that invoke error-prone DNA synthesis, either over small regions of the genome such as gaps created during conditions of starvation or over an entire round of replication. Specialized polymerases can contribute to region, lesion, or time-restricted mutation.
4. Homologous recombination is part of several proposals. Rosenberg has reviewed recombination models of several sorts including those in which error-prone synthesis is concomitant with recombination and others in which the degree of sequence similarity required for homologous recombination is varied under conditions of properly perceived stress.
5. Most intriguing are mechanisms in which the generation of mutation and its fixation are separated. This separation is accomplished by making fixation dependent on growth at the critical moment. This class of mechanisms was first suggested in the seminal 1988 article by Cairns et al. They proposed a specific mechanism in which error-prone transcription occasionally results in a wild-type transcript being produced from a mutant locus. This mRNA will allow the cell to grow momentarily. The model further proposes that the change in physiology attendant to momentary growth results in a burst of reverse transcriptase activity that acts on mRNA resulting in DNA that homologously recombines with the chromosomal locus to fix the mutation.

FW Stahl proposed a mechanism in which gaps in the DNA arise during starvation and are filled in by error-prone

DNA synthesis. New DNA synthesis is normally undermethylated and subject to mismatch correction in favor of the old chain. However, if the cell starts to grow – for example, owing to error-prone synthesis resulting in a wild-type allele – then the mismatch will be resolved by replication, and the mutation will be fixed in one of the progeny cells. Shapiro and colleagues have proposed and presented supporting evidence that gene fusions mediated by phage mu are modified by local transcription and replication.

Models of the type proposed by Cairns and by Stahl have often been discounted on the basis of evidence of double mutants among the products of single selections. This finding is taken to be evidence of the cells having been through a general state of high mutation rate spread throughout the genome. Although consistent with a temporary state of generalized high mutation, the presence of double mutants among the products of selection for single mutants is not definitive evidence. It has not been generally appreciated that the same result would be obtained if reversible intermediates were present at several places around the genome at the same time. If the integrated effect of reversible intermediates is positive for cell replication, then they will all be rescued via replication in a sort of "group hitchhiking."

5 How Might Genetic Intelligence Have Evolved?

The generation of genetic variation is, in large part, a function of the genes and physiology of DNA metabolism. Genes responsible for the generation of particular alleles will tend to be coinherited with the alleles that they played a role in generating. The components exist for feedback between the generators of genetic diversity and the environment that selects among variants. The utilization of this capability to generate, among the range of all possible variations that subset that is "anticipated" to be of use, is admittedly a radical proposition but we contend that it should be taken seriously. Figure 1 illustrates the conventional view of evolution. In this simple view, the sole act of the environment is to select among variants; there is no role for the

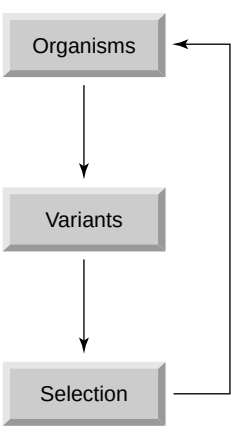

Fig. 1 A conventional view of evolution. The environment functions only at the selection step.

environment to play in the generation of diversity. In Fig. 2, a more complex and realistic view of evolution is shown. The environment acts both on the generation of variation and as a selective agent among variants.

The action of the environment on the generation of variation takes several forms and goes via several routes. In direct action on DNA, radiation or chemicals alter the chemistry leading to a change in base sequence. These agents can also act less directly by modifying proteins involved in DNA metabolism. This is the route by which topoisomerase inhibitors or agents that alter nucleotide pools can promote mutation and recombination. Even the most direct action of the outside environment on DNA requires processing by DNA repair and replication proteins in order to become fixed as a mutation. The "environment" of a DNA sequence means the cell's internal milieu and state as well as that which is outside an organism's skin or membrane. Even its sequence context is properly considered part of the "environment" of a particular base.

At no stage of evolution could the generation of diversity ever be purely "random," if by "random" it is meant that every sort of change is equally probable. For the initial discussion, let us limit the case to substitution mutations in which the polymerase allows the pairing and ligation of a non-Watson–Crick base pair, and uses that mispair as the primer for further extension. Until another round of replication, the mutation is not fixed; it remains vulnerable to mismatch correction. A purely random model for the generation of missense mutations would suppose that all the substitutions are equally common and that the rate of substitution is invariant. Both these points are not accurate statements. The rate of each possible substitution mutation is different, both because the polymerase has a different propensity to make each sort of mistake and because the mismatch correction system has different affinities for each of the possible mismatched base pairs. The rate of substitution mutation is strongly affected by the nucleotide pools; such pools are altered with cell physiology. The proofreading of polymerase is specifically inhibited by the SOS-induced gene *umu*D. Transcriptional focusing could allow regulation of genetic change to be as tightly coupled to the organism's perception of its environment, as is gene expression.

Much of what is inherited is not due to a direct effect on survival; rather it is the product of neutral selection. The information mediated by perception that modulates physiology may or may not be congruent with that which alters genetic metabolism. In Fig. 2, what is the relationship of information flowing through arrows 3 and 4, and what modulates that relationship via physiology and evolution? We are not claiming that evolution moves inexorably toward anticipatory mutation. We simply point out that the tools and feedback mechanisms exist such that life *could* evolve the ability to focus the generation of variation where an evolved mechanism "guesses" or "anticipates." Mutation rate is sometimes portrayed as a compromise of fidelity, economy, and the occasional need to generate variation. Differential variation with respect to genes and time offers a way out of compromise, particularly if sensitive to the environment.

Rapid bacterial evolution to antibiotic resistance and clonal evolution in oncogenesis both involve concerted genetic changes that may include aspects of selection-promoted mutation. The immune system of mammals has several aspects of focused

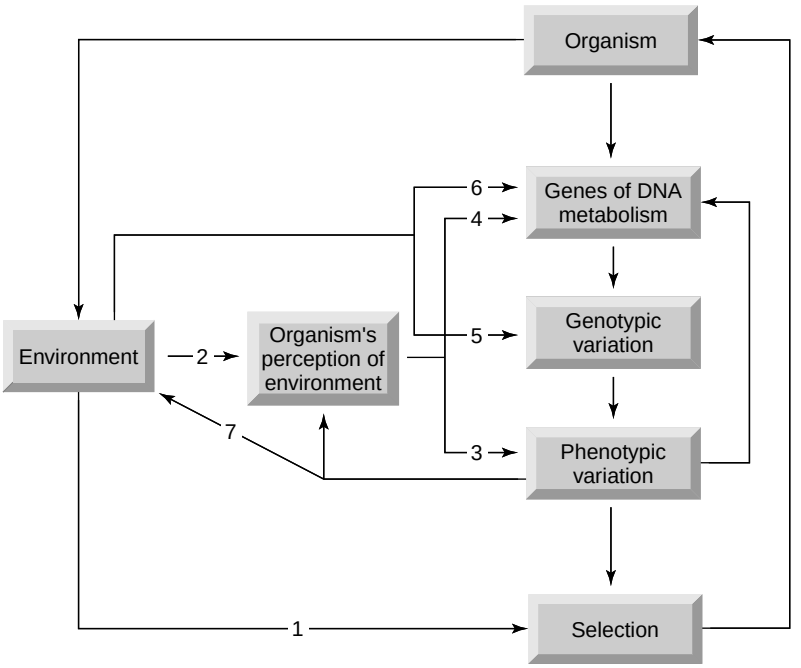

Fig. 2 A more complete view of evolution. The environment functions at three stages. (1) The environment is the proximate agent of selection. (2) The environment is perceived by the organism. (3) Organisms use their perception of the environment to modify their physiology, as in operon induction. (4) Organisms use their perception of the environment to modify their genetic metabolism, as in the SOS and p53 pathways. (5) The environment directly impinges on the DNA directly via such agents as radiation and chemical mutagens. (6) The environment can interact indirectly on DNA via the genes of DNA metabolism, as in topoisomerase inhibitors in chemotherapy. (7) The organism modifies environmental interaction with the genome as in metabolic activation, or, detoxification.

and adaptive mutation, including somatic mutagenesis at the VDJ junction and receptor editing. Study of these processes implies a further feedback loop.

6 Genetic Algorithms

Genetic algorithms are computer codes inspired by analogies to biology and the chromosome, and capable of evolving. In this section, the means of generating diversity in genetic algorithms and the possibility of using genetic algorithms to evolve a pseudogenetic intelligence are discussed. Computer algorithms are not constrained to mimic the biological systems that inspired them. However, in terms of mechanisms to generate diversity, there is a richness of carbon-based genetics, which has not yet found its way into silicon.

The evolution of a computer began by constructing analogies for natural components. The binary code sequence of 0s and 1s is analogous to the nucleotide sequences of G, A, T, and C. The

next analogy involves natural selection. Selection is typically taken to be a process whereby individuals of greater fitness are more likely to survive and propagate. The fitness evaluation of genetic algorithms is not directly related to the generation of diversity and will not be discussed here further.

The basic outline of a genetic algorithm is as follows:

1. Create a population of chromosomes
2. Evaluate the fitness of all chromosomes
3. Select parents from among fitter chromosomes
4. Produce offspring from parents while employing diversity generating operators
5. Apply selective pressure to remove less fit chromosomes
6. Return to step 2.

Each of these steps requires variables that may be chosen arbitrarily. In particular, the variable operators in step 4 could be developed to give genetic algorithms greater potential for generating diversity in manners similar to natural systems.

Genetic algorithms utilize bit strings analogous to chromosomes. The length of the chromosome is arbitrary for any given algorithm. The figure below is an example of a binary chromosome of length 40.

1010010010000101111001100101010101010101

The reproductive aspect of genetic algorithms varies widely in method, but there are some common features that are essential to any genetic algorithm. The most important of these features is the generation of diversity among the offspring of the parent chromosomes.

It is the generation of diversity and the subsequent selection for the fittest individuals that allows genetic algorithms to display evolutionary behavior. Natural systems have evolved a wide variety of techniques for altering the arrangement of their chromosomes and analogous techniques can be instituted in genetic algorithms. Doing so might have the double advantage of allowing genetic algorithms to adapt to more complex situations while also providing a tool for observing evolutionary behavior of the algorithm.

The simplest means of generating diversity in a genetic algorithm is random mutation concomitant with replication. A bit flip (from 0 to 1 or *vice versa*) occurs at an identical and fixed probability for each position at each replication. This is analogous to random-base substitutions in the nucleotide sequence of an organism. The probability with which mutation occurs is set by the algorithm design. We doubt that random mutations alone can provide the optimal generation of diversity for genetic algorithms.

Insertion and deletion of single bases are simple extensions of base substitutions. Polymerase replication errors are most commonly cited as the cause for these mutations in natural systems. They are manifested in a genetic algorithm in the same manner as random bit flips, at some given probability in each reproduced chromosome. Below are examples from both nucleotide and binary sequences.

Random Mutation

	Natural chromosome
Parent	GATCCGTACGATCGGGGGCTACATCCTGAGA
Offspring	GATCCGTACGATCGG**A**GGCTACATCCTGAGA

	Binary chromosome
Parent	0100101010100101110010101010001
Offspring	0100101010100100110010101010001
Random Insertion	
	Natural chromosome
Parent	GATCCGTACGATCGCGGGCTACATCCTGAGA
Offspring	GATCCGTACGATCGC**G**GGGCTACATCCTGAGA
	Binary chromosome
Parent	0100101010100101110010101010001
Offspring	0100101010100101111001010101001
Random Deletion	
	Natural chromosome
Parent	GATCCGTACGATCG**C**GGGCTACATCCTGAGA
Offspring	GATCCGTACGATCGGGGCTACATCCTGAGA
	Binary chromosome
Parent	0100101010100101110010101010001
Offspring	0100101010100101100101010001

It is important to note that these three mutations do not always occur at the same rate at every location of a natural chromosome. They frequently depend upon sequence context. For example, inserted bases resulting from a looping-out error on a newly synthesized DNA strand occur more often in stretches of repeated bases (i.e. GACCT**AAAAA**TCGAT). Therefore, the genetic algorithm that seeks to model this phenomenon must vary the probability of the mutation occurring in specific sequence environments. These types of mutations can provide diversity from parent to offspring but do not allow any exchange of information between existing parents. For that to occur, mating, and some form of genetic recombination must take place.

Genetic recombination is a common occurrence in natural systems and forms the basis of sexual reproduction. The ability to exchange and reorder DNA is fundamental to the evolution of sexual organisms; it is also important to genetic algorithms. Currently, genetic algorithms recombine in very limited, but powerful, ways. The figure illustrates a common form of recombination used. In this technique, the two parent chromosomes are first passed through a crossover probability function. This determines the probability that the two parents will recombine. Once passed, a point on the chromosome is randomly chosen (in this case the tenth position) and the codes are swapped from parent to parent.

Parent 1	1111111111111111111111111111111111111111
Parent 2	0000000000000000000000000000000000000000
Offspring 1	1111111111000000000000000000000000000000
Offspring 2	0000000000111111111111111111111111111111

The power of this procedure derives from the fact that it allows the exchange of large blocks of code; blocks that could contain useful groups of code. Studies have shown that blocks of code that tend to increase the fitness of the chromosome increase their frequency in the population over many generations. These blocks

have been termed schemata and they are roughly analogous to genes. Other extensions of this procedure are sometimes used, typically allowing for more than one crossover point. However, despite its power, this technique is not an accurate representation of natural systems. Biological organisms do not experience completely random recombinations. Instead, there are a number of regulatory conditions that affect the probabilities of where and when recombination occurs. In homologous (= generalized = legitimate) recombination, such as that which occurs during meiosis, recombination is dependent on similarity between two sequences. The exact degree of similarity required is itself a genetically regulated parameter. Similarity matching in genetic algorithms would require an operator that compares lengths and similarity, and if satisfied, recombines after an appropriate probability function. For example, the two parent chromosomes in the figure below will only recombine if there is similarity in 9 bits out of 10, and then only 50% of the time.

Parent 1	**10101001***00111011010***100101011010010101011**
Parent 2	00010110*00101011011*111100110101101011110**0**
Offspring 1	**1010100100111**0110111110011010110101111**00**
Offspring 2	00010110001010**1101010010101101001010101 1**

This form of recombination has several advantages. It possesses a greater number of variables that can be altered to study their effect. The size of the region of similarity, the degree of similarity required, where and how often within the region of similarity the recombination will occur, and the probability of recombination then occurring are all arbitrary values. The analogies of these values are poorly understood in natural systems and genetic algorithms can provide insight into their interactions. Also, unlike recombinational operators that predetermine the number of crossover points, this method has the inherent potential for multiple crossovers. In biological systems, the probability of multiple crossovers is often different from the simple product of individual events. This genetic property called *interference* has not yet been incorporated into genetic algorithms.

The recombinational model just described allows for the creation of recombinational hotspots and coldspots. Hotspots are regions where generalized recombination occurs with increased frequency; they are a common phenomenon in biological organisms. Hotspots would be invoked at the point of the probability check. The probability that recombination occurs once homology has been established could then be directly influenced by the sequence of the similarity. In addition to specific sequences influencing the probability of recombination, the probability could be made a function of the ratio of 1s to 0s, or it could depend upon the prevalence of repeated bits (analogous to the proposal that GC regions or perfect palindromes are hotspots in biological systems).

Homologous recombination is not the only means by which natural organisms recombine. Site-specific recombination occurs not because of sequence similarity but because of the recognition of certain sequences by specific proteins that bind to those sequences. The chemical pathways are not significant for genetic algorithms, but their results can be simulated to good effect. Typically, site-specific recombination removes, inserts, or copies

regions of DNA, as is the case with many transposons. Including an analogous function in genetic algorithms would allow another means of generating diversity among binary chromosomes. For example, in the chromosome below, the sequence 0000011111 is a transposable element that, at some probability, will move itself into the center of any sequence of 011001.

Parent 1 1101010011**00000111**1001101011*011001*0101011
Offspring 1 1101010011001101011011**00000111**10010101011

Occasionally, in natural systems transposons will incorporate and relocate pieces of chromosomal DNA. Genetic algorithms can also model this behavior. For example, the previous transposable element could be made to recognize only the repeat 0s and 1s of the sequence, regardless of the code in between them.

Parent 1 1101010011**00000**101011**1111**1001101011*011001*0
Offspring 1 1101010011001101011011**00000**101011**1111**10010

Such a system allows useful groups of code (i.e. genes) to be moved around the chromosome. Alternately, they could be deleted once excised, copied into the new location without being removed from their original site, inverted, or a combination of the latter two.

As with homologous recombination, site-specific recombinational techniques require many arbitrary values. The sequences recognized, the size of the transposable unit, the probability of a recombinational event occurring, and the rate at which each type of recombinational event occurs all must be given to the algorithm.

Most cells contain a great deal more DNA than that required for encoding their genes. Eukaryotic chromosomes are estimated to be up to 98% irrelevant DNA. These vast quantities of unused DNA are considered to be a sort of blackboard upon which evolution can scribble without affecting those regions of the chromosome encoding essential genes. Therefore, some expansion of the bit sequences available must occur. One means of doing this is to elongate the chromosomes with regions of "junk" sequence that are not included in the fitness evaluation.

Biological evolution is horizontal as well as vertical. Intergeneric recombination moves DNA between widely separated lineages. Plasmids and viruses move DNA from cell to cell with great promiscuity. Many prokaryotics are metabolically

Parent 1
1101010011**00000**101011**1111**1001101011*011001*0

Deleted
1101010011001101011*011001*0

Copied
1101010011**00000**101011**1111**1001101011*011***00000**101011**1111***001*0

Inverted
1101010011**1111**11101**00000**001101011*011001*0

Copied and Inverted
1101010011**00000**101011**1111**1001101011*011***1111**11101**00000***001*0

adapted for the uptake of DNA they encounter in the environment. Given that natural organisms have a great deal more DNA than just their chromosomes, genetic algorithms could profit from analogous forms of binary code.

Regions of code that do not initially encode useful genes must at some point become recruited into the fitness evaluation if they are to ever be more than "junk" code. Regulation of which regions of the chromosome are to be used in the fitness evaluation is under the control of the chromosome in natural systems. Aspects of gene regulation and expression can be incorporated into genetic algorithms to provide similarity to natural systems. An analogy with promoter sequences can be used to determine which region of the chromosome is to be evaluated, thus defining the genes. They also make it possible to hand over some controls to the chromosome itself. For example, the probability that homologous recombination will occur could be given to the chromosome by encoding the probability value on the chromosome after a specific promoter. As a result, one of the variables that influences the generation of diversity is itself subject to genetic change. Any value usually specified in the structure of a genetic algorithm could be relocated to the chromosome in this way, in effect giving the chromosome the ability to alter its own means of changing. It is even possible to allow the chromosomes to evolve promoters by predefining both default values for the variables and the sequence of a promoter, initially absent from the population, which would encode the value of that variable. In this way, we hope to evolve genetic intelligence as the binary chromosome takes control of its own means of changing. New insights are almost certain to arise through the simultaneous study of carbon- and silicon-based life forms.

Acknowledgments

David S. Thaler is a Senior Scholar of the Ellison Medical Foundation. Fiona Doetsch and Joshua Lederberg gave encouragement and insight. The authors are thankful to the Sloan Foundation for their support.

See also Genetics, Molecular Basis of.

Bibliography

Batson, G., Batson, M.C. (1986) *Angels Fear. Towards an Epistomology of the Sacred*, Bantam Books, New York.

Brock, T.D. (1990) *The Emergence of Bacterial Genetics*, Cold Spring Harbor Press, Cold Spring Harbor, New York.

Cairns, J., Overbaugh, J., Miller, S. (1988) The origin of mutants, *Nature* **335**, 142–145.

Caporale, L.H. (2002) *Darwin in the Genome. Molecular Strategies in Biological Evolution*, McGraw-Hill, New York.

Delbruck, M. (1986) *Mind From Matter? An Essay on Evolutionary Epistemology*, Blackwell Scientific Publications, Palo Alto, CA.

Foster, P.L. (1993) Adaptive mutation: the uses of adversity, *Annu. Rev. Microbiol.* **47**, 467–504.

Friedberg, E.C., Wagner, R., Radman, M. (2002) Molecular biology – specialized DNA polymerases, cellular survival, and the genesis of mutations, *Science* **296**, 1627–1630.

Holland, J.H. (1992) *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and AI*, Bradford Series in Complex Adaptive Systems, MIT Press, Cambridge, MA.

Illinois Genetic Algorithm Laboratory. http://www-illigal.ge.uiuc.edu/.

Kauffman, S.A. (1993) *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, Oxford.

Lenski, R.E., Mittler, J.E. (1992) The directed mutation controversy and neo-Darwinism, *Science* **259**, 188–194.

McClintock, B. (1983) The significance of responses of the genome to challenge, *Science* **226**, 792–801.

Rosenberg, S.M. (1994) In pursuit of a molecular mechanism for adaptive mutation, *Genome* **137**, 893–899.

Shapiro, J.A. (1993) Natural genetic engineering of the bacterial genome, *Curr. Opin. Genet. Dev.* **3**(6), 845–848.

Thaler, D.S. (1994) The evolution of genetic intelligence, *Science* **264**, 224–225; Correspondence and reply, *Science* **265**, 1994–1996.

Thaler, D.S. (1994) Sex is for sisters: intragenomic recombination and homology-dependent mutation as sources of evolutionary variation, *Trends Ecol. Evol.* **9**, 108–110.

Thaler, D.S. (1999) Hereditary stability and variation in evolution and development, *Evol. Dev.* **1**, 113–122.

Genetic Variation and Molecular Evolution

Werner Arber
Biozentrum, University of Basel, Klingelbergstrasse 70, Basel, Switzerland

1 Introduction 333

2 Principles of Molecular Evolution 334
2.1 Evolutionary Roles of Genetic Variation, Natural Selection, and Isolation 334
2.2 Molecular Mechanisms of the Generation of Genetic Variation 336

3 Genetic Variation in Bacteria 337

4 Local Changes in the DNA Sequences 338

5 Intragenomic DNA Rearrangements 339
5.1 Site-specific DNA Inversion at Secondary Crossover Sites 340
5.2 Transposition of Mobile Genetic Elements 341

6 DNA Acquisition 342

7 The Three Natural Strategies Generating Genetic Variations Contribute Differently to the Evolutionary Process 343

8 Evolution Genes and Their Own Second-order Selection 345

9 Arguments for a General Relevance of the Theory of Molecular Evolution for All Living Organisms 346

10 Conceptual Aspects of the Theory of Molecular Evolution 347
10.1 Pertinent Scientific Questions 347

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

10.2 Philosophical Values of the Knowledge on Molecular Evolution 348
10.3 Aspects Relating to Practical Applications of Scientific Knowledge on Molecular Evolution 350

Bibliography 351
Books and Reviews 351
Primary Literature 351

Keywords

Biological Evolution
A nondirected, dynamic process of diversification resulting from the steady interplay between spontaneous mutagenesis and natural selection.

DNA Rearrangement
Results from mostly enzyme-mediated recombination processes, which can be intra- or intermolecular.

Evolution Gene
Its protein product acts as a generator of genetic variations and/or as a modulator of the frequency of genetic variations.

Gene Acquisition
Results from horizontal transfer of genetic information from a donor cell to a receptor cell. With bacteria, this can occur in transformation, conjugation, or phage-mediated transduction.

Natural Selection
Results from the capacity of living organisms to cope with the encountered physicochemical and biological environments. Largely depending on its genetic setup and physiological phenotype, each organism may have either a selective advantage or a selective disadvantage as compared to the other organisms present in the same ecosystem.

Spontaneous Mutation
Defined here as any alteration of nucleotide sequences occurring to DNA without the intended intervention of an investigator. The term mutation is used here as a synonym of genetic variation.

Transposition
DNA rearrangement mediated by a mobile genetic element such as a bacterial insertion sequence (IS) element or a transposon.

Variation Generator
An enzyme or enzyme system whose mutagenic activity in the generation of genetic variation has been documented.

The comparison of DNA sequences of genes and entire genomes offers interesting insights into the possible evolutionary relatedness of genetic information of living organisms. Together with a relatively rich database from experimental microbial genetics, conclusions can be drawn on the molecular mechanisms by which genetic variations are spontaneously generated. A number of different specific mechanisms contribute to the overall mutagenesis. These mechanisms are here grouped into three natural strategies of the spontaneous generation of genetic variations: local changes of DNA sequences, intragenomic rearrangement of DNA segments, and acquisition of foreign DNA by horizontal gene transfer. These three strategies have different qualities with regard to their contributions to the evolutionary process. As a general rule, none of the known mechanisms producing genetic variants is clearly directed. Rather, the resulting alterations in the inherited genomes are more random. In addition, usually only a minority of resulting variants provide a selective advantage. Interestingly, in most of the molecular mechanisms involved, the products of so-called evolution genes are involved as generators of genetic variation and/or as modulators of the frequencies of genetic variation. Products of evolution genes work in tight collaboration with nongenetic factors such as structural flexibilities and chemical instabilities of molecules, chemical and physical mutagens, and random encounter. All of these aspects contributing to the spontaneous generation of genetic variations together form the core of the theory of molecular evolution. This theory brings neo-Darwinism to the molecular level. In view of the increasing evidence coming particularly from microbial genetics, knowledge of molecular evolution can be seen as a confirmation of Darwinism at the level of biologically active molecules, in particular, nucleic acids and proteins. Philosophical and practical implications of this knowledge will be briefly discussed.

1 Introduction

Evolutionary biology has traditionally devoted its major attention to the comparison of phenotypical traits of higher organisms, both of those actually living and of those that have been extinct (paleontological fossils). The resulting theory of descent is, together with other criteria, at the basis of the systematic classification of living organisms. Darwin's theory of natural selection brought a new element into the understanding of the long-term development of forms of life. Natural selection is the result of organisms coping with the encountered living conditions that are dependent on both the environmental physicochemical conditions and the activities of all living forms in a particular ecological niche. The Darwinian theory of evolution also postulated that intrinsic

properties of life are not entirely stable and principally identical for all organisms of a given species. In the so-called modern synthesis, in which evolutionary biology and genetics became integrated, transmissible phenotypic variations representing the substrate of natural selection were explained as due to genetic variations (or mutations). Shortly thereafter, deoxyribonucleic acid (DNA) was identified as the carrier of genetic information. DNA is thus also the target for mutagenesis. Within the last few decades a rapid development of molecular genetics with novel research strategies leading to genomics, sequencing of entire genomes, and functional studies of genes and their products paved the way for hitherto inaccessible knowledge on the basis of life and its multiple manifestations. This also relates to the process of molecular evolution. A synoptical insight into the various molecular mechanisms contributing to the generation of genetic variations represents a molecular synthesis between the neo-Darwinian theory and molecular genetics. This synthesis can confirm the Darwinian evolution at the molecular level.

2 Principles of Molecular Evolution

The principles of molecular evolution to be outlined here are founded on

1. the neo-Darwinian theory of evolution with its three pillars of genetic variation, natural selection, and isolation;
2. the solidly established microbial genetics database;
3. DNA sequence comparisons with bioinformatic tools;
4. physicochemical knowledge on the reactivity, conformational flexibility, and chemical stability of biologically active molecules.

2.1 Evolutionary Roles of Genetic Variation, Natural Selection, and Isolation

The long-term maintenance of any form of life requires a relatively high stability of its genetic information. However, rare occasional genetic variations occur in all organisms. This gives rise to mixed populations of organisms with the parental genome and organisms having one or more alterations in their genome. These populations are steadily submitted to natural selection. The experience shows that, in general, favorable genetic variations are considerably less frequent than unfavorable ones. The latter provide a selective disadvantage. Indeed, genetic variants with unfavorable genetic alterations will sooner or later get eliminated from propagating populations, which will become enriched for organisms carrying favorable genetic variations. It should be noted that by far not all alterations in the nucleotide sequences of a genome will lead to a change in the phenotype of the organism. However, such silent and neutral mutations may later become physiologically relevant in conjunction with still other, upcoming DNA sequence changes.

Natural selection is by no means a constant element. It varies both in time and in space. This is due to variations in the physicochemical environmental conditions as well as to variations of the life activities of all the different organisms present in a particular ecological niche and forming an ecosystem. Since a genetic variation may also affect the influence that the organism exerts on the other organisms present in the same ecosystem (e.g. think of weeds and pathogenicity

effects, but also of beneficial, synergistic effects), any novel mutation may not only influence the life of the concerned organism itself but also the lives of other cohabitants of the same ecological niche.

The third pillar of biological evolution – besides genetic variation and natural selection – is isolation. Evolutionary biologists define two different aspects of isolation. One of these is *geographic isolation*, which may seriously reduce the number of potential habitats for an organism. The other type of isolation is called *reproductive isolation*. For example, two distantly related diploid organisms may not be fertile in sexual reproduction. But reproductive isolation can also be seen in a wider definition to seriously limit the possibility of horizontal transfer of segments of genetic information between two different kinds of organisms.

As summarized in Fig. 1, genetic variation drives biological evolution. Complete genetic stability would render any

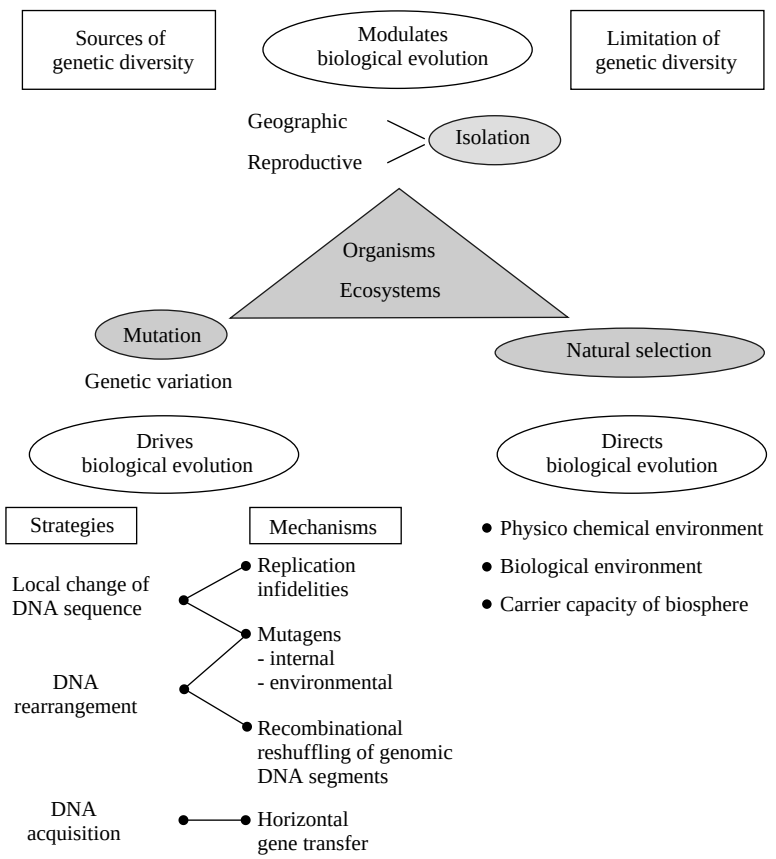

Fig. 1 Synoptical presentation of major elements of the theory of molecular evolution. A number of specific mechanisms, each with its own characteristics, contribute to the four groups of mechanisms of genetic variation listed. Each of the specific mechanisms follows one and sometimes more than one of the three principal, qualitatively different strategies of genetic variation.

evolutionary process impossible. A very high frequency of genetic variation would rapidly lead to the extinction of the concerned organisms because of the stated prevalence of unfavorable mutations in the spontaneous generation of genetic variations. It is natural selection together with the available sets of genetic variants that determines the direction of biological evolution, or in other words, the directions in which the branches of the tree of evolution grow. Finally, the geographic and reproductive isolations modulate the evolutionary process.

2.2 Molecular Mechanisms of the Generation of Genetic Variation

The concept to be presented here requires the reader to question some long-established textbook knowledge, such as the claim that spontaneous mutations would largely result from errors, accidents, and illegitimate processes. We defend here the alternative view that living nature actively cares for biological evolution

1. by making use of intrinsic properties of matter such as a certain degree of chemical instability and of structural flexibility of molecules, and
2. by having developed genetically encoded systems, the products of which are involved in the generation of genetic variations and in modulating the frequencies of genetic variation.

It might be relevant to mention here that the term mutation is differently defined in classical genetics and in molecular genetics. In classical genetics, a mutant is any variant of a parental form, showing in its phenotypic properties some alteration that becomes transmitted to the progeny. In contrast, it has become a habit in molecular genetics to call any alteration in the parental nucleotide sequence of the genome a mutation, whether it has phenotypic consequences or not. There is good reason to assume that in most spontaneously occurring mutagenesis events, the specific mechanism involved will not pay attention to whether the sequence alteration at the involved target site will cause a phenotypic alteration or not. Therefore, for studies on mechanisms and on the statistics of their occurrence, it is indicated to follow the molecular genetics definition of the term mutation, which we use here as a synonym of genetic variation. We will use the term *spontaneous mutation* to label any type of DNA sequence alteration unintended by the investigator. This definition says nothing about whether a mutation relates to a phenotypic change.

At present, the research on genetic variation mainly follows two strategies. First, the increasing availability of entire and partial genome sequences offers excellent means to compare regulatory sequence elements, specific domains of genes, entire genes, groups of genes, and entire genomes with regard to DNA sequence homologies and genome organization. Within a given species, this can reveal a genetic polymorphism. Between more or less related species, it can give a reliable measure for evolutionary relatedness. For example, the molecular clock – an indicator of evolutionary relatedness – is based on single nucleotide alterations. Sequence comparisons can often suggest how sequence alterations could have occurred in the course of past evolution. Secondly, a more reliable insight into the generation of genetic variations can be gained by the observation of individual processes of mutagenesis. In view of the large size of genomes and of the rare and random

occurrence of spontaneous DNA sequence alterations, this approach is relatively difficult. However, a rich database is already available from microbial genetics, particularly from bacteria and their viruses and plasmids. Their relatively small genomes are haploid so that phenotypic effects caused by genetic variation normally become rapidly manifested. With appropriate selection and screening techniques, this allows one to identify occasional, functionally relevant mutations in populations. On the other hand, investigations on structural alterations in the genomes of individual bacterial colonies, for example, by the study of restriction fragment length polymorphism, can reveal when and where on the genome a DNA rearrangement must have occurred.

A quite solid, general result of this type of experimental investigations reveals that in the spontaneous generation of genetic variations, it is not just a single mechanism at work. Rather, a number of mechanistically different processes contribute to the overall mutagenesis. We will discuss selected examples below. Interestingly, a critical evaluation of the situation shows that the specific mechanisms of mutagenesis often depend both on nongenetic elements and on activities of specific enzymes, the products of the so-called evolution genes. The multitude of thus identified, distinct mechanisms contributing to the formation of genetic variations can be grouped into three qualitatively different natural strategies (Fig. 1). These are

1. local changes in the DNA sequences,
2. rearrangement of DNA segments within the genome,
3. the acquisition by horizontal transfer of a DNA segment originating in another kind of organism.

Selected examples for each of these strategies will be discussed in Sects. 4, 5, and 6.

3 Genetic Variation in Bacteria

Several seminal discoveries, largely based on work carried out with microorganisms between 1943 and 1953, were essential for the later development of molecular genetics.

1. It was realized that bacteria and bacteriophages have genes that can mutate, and that spontaneous mutations in microorganisms normally arise independently of the presence of selective agents. It was also learned that the genetic information of bacteria and of some bacteriophages is carried in DNA molecules rather than in other biological macromolecules such as proteins.
2. The newly discovered phenomena of DNA transformation, bacterial conjugation, and bacteriophage-mediated transduction revealed natural means of horizontal gene transfer between different bacterial cells.
3. It was seen that horizontal gene transfer has natural limits, including systems of host-controlled modification, which are today known as *DNA restriction–modification systems*.
4. Mobile genetic elements were identified as sources of genetic instability and were seen to represent mediators of genetic rearrangements. While such rearrangements are often caused by transposition, they can also result from the integration of a bacteriophage genome into the genome of its bacterial host strain, which is thereby rendered lysogenic.

It is at the end of this fruitful period of discoveries, in 1953, that structural analysis of DNA molecules led to the double-helix model. The filamentous structure of DNA molecules made it clear as to how genetic information could be contained in the linear sequences of nucleotides. The double helical nature of the model also offered an understanding of semiconservative DNA replication at the molecular level and thus of information transfer into progeny.

Many classical microbial genetic investigations were carried out with *Escherichia coli* K-12. Its genome is a single circular DNA molecule (chromosome) of about 4.6×10^6 bp. In periods of growth, the rate of spontaneous mutagenesis is about 10^{-9} per bp and per generation. This represents one new mutation in every few hundred cells in each generation. *E. coli* has several well-studied bacteriophages and plasmids. This material facilitates investigations on life processes in these bacteria.

Under good growth conditions, the generation time of *E. coli*, measured between one cell division and the next, is very short, on the order of 30 min. Upon exponential growth, this leads to a multiplication factor of 1000 every 5 h. Thus, a population of 10^9 cells representing 30 generations is reached from an inoculum of a single cell in only 15 h. This rapid growth rate greatly facilitates population genetic studies and thus facilitates investigations on the evolutionary process.

On the filiform DNA molecules of *E. coli* and its bacteriophages and plasmids, the genetic information is relatively densely stored as linear sequences of nucleotides or base pairs. Genes depend on the presence of continuous sequences of base pairs (reading frames) that encode specific gene products, usually proteins, and of expression control signals that ensure the occurrence of gene expression at the relevant time with the needed efficiency. Mutations can affect reading frames as well as control signals, both of which represent specific DNA sequences. In addition, some specific DNA sequences relate to the control of the metabolism of the DNA molecules, in particular, their replication.

In the following Sects. 4, 5, and 6, we will describe selected examples of mechanisms contributing each in its specific manner to the formation of genetic variants. We will group them into the already mentioned three major natural strategies (Fig. 1), although, as we will later see, some of the specific mechanisms involve more than one of these strategies.

4
Local Changes in the DNA Sequences

The process of DNA replication is one of the important sources of genetic variation, which may depend both on structural features of the substrate DNA and on functional characteristics of the replication fork.

Some of the "infidelities" of DNA replication are likely to depend on tautomeric forms of nucleotides, that is, a structural flexibility inherent in these organic compounds. Base pairing depends on specific structural forms. Conformational changes of nucleotides can result in a mispairing if short-living, unstable tautomeric forms are "correctly" used in the synthesis of the new complementary strand upon DNA replication. For this reason, we do not consider mutations resulting in this process as errors and call them *infidelities*. DNA replication is indeed one of the sources of nucleotide substitution, and this plays an

important role in the evolutionary development of biological functions.

An inherent low degree of chemical instability of nucleotides represents another source of mutagenesis. For example, cytosine can undergo oxidative deamination to become uracil. Upon replication, this gives rise to an altered base pairing and results also in nucleotide substitution.

Other replication infidelities that may relate to slippage in the replication fork can result in either the deletion or the insertion of one or a few nucleotides in the newly synthesized DNA strand. If such mutations occur within reading frames for protein synthesis, the phenotypic effect may be drastic. This is the case when, in the protein synthesized from a gene affected by a frameshift mutation, the amino acid sequence downstream of the site of mutation strongly differs from that of the nonmutated product. In addition, the size of such a mutated protein is usually altered depending on the chance occurrence of an appropriate stop codon in the new reading frame.

Proofreading devices and other enzymatic repair systems prevent replication infidelities from producing mutations at intolerably high rates. Generally, they act rapidly after replication by screening for imperfect base pairing in the double helix. Successful repair thereby requires specific means to distinguish the newly replicated DNA strand from its template, the complementary parental strand. Because of these correction activities, many primary mispairings are removed before they have the opportunity to become fixed as mutations. DNA repair systems modulate the frequencies of mutagenesis.

Genetic information of some viruses, and sometimes also segments of genetic information of chromosomal origin, may pass through RNA molecules, which may later become retrotranscribed into DNA. No efficient repair systems are known to act at the level of RNA. Indeed, RNA replication shows a higher degree of infidelity than DNA replication. In consequence, genetic information that becomes replicated as RNA molecules generally suffers increased mutation rates.

A relatively large number of internal and environmental chemicals exert mutagenic effects by means of molecular mechanisms that in many cases are well understood and often cause local DNA sequence changes. Some intermediate products of the normal metabolic activities of a cell may be mutagenic and may thus contribute to spontaneous local mutagenesis. The mutagens of the environment include not only a multitude of chemical compounds but also ultraviolet radiation and some physicochemical constraints such as elevated temperature, which influence the chemical stability of nucleotides. Each of these mutagens and mutagenic conditions contributes in a specific way to the generation of genetic variations.

Again, many of the potential sequence alterations brought about by internal and environmental mutagens are efficiently repaired by enzymatic systems. However, since the efficiency of such repair is rarely 100%, evolutionarily relevant mutations persist.

5 Intragenomic DNA Rearrangements

Various recombination processes are well known to mediate DNA rearrangements, which often result in new nucleotide sequences.

While in haploid organisms general recombination is not essential for propagation, it influences genetic stability at

the population level in various ways as a generator of new sequence varieties. For example, it can bring about sequence duplications and deletions by acting at segments of homology that are carried at different locations in a genome. General recombination is also known to act in the reparation of damage caused to DNA by ionizing radiation. In this case, an intact genome can become assembled from undamaged segments of sister copies of the chromosome by homologous recombination. The best-known contribution of general recombination to genetic diversity is meiotic recombination bringing about random recombinants between the paternal and the maternal chromosomes in diploid organisms.

Two other widely spread types of recombination systems are dealt with in more detail in Sects. 5.1 and 5.2: site-specific recombination and transposition. Both are known to contribute to genetic variation. Still other recombination processes, such as the one mediated by DNA gyrase can, for the time being, perhaps best be grouped as illegitimate recombinations. This group may contain several different molecular mechanisms that act with very low efficiency and have remained at least in part unexplained.

5.1 Site-specific DNA Inversion at Secondary Crossover Sites

Genetic fusions represent the results of joining together segments of two genes (gene fusions) or of two operons (operon fusions) that are not normally together. An *operon* is a set of often functionally related genes that are copied into messenger RNA (i.e. transcribed) as a single unit. As a result of this organization, those genes are coordinately regulated; that is, they are turned on or off at the same time. Therefore, in an operon fusion, one or more genes are put under a different transcription control, but the genes *per se* remain unchanged. In contrast, gene fusion results in a hybrid gene composed of sequence motifs and often of functional domains originating in different genes.

In site-specific DNA inversion, a DNA segment bordered by specific DNA sequences acting as sites of crossing-over becomes periodically inverted by the action of the enzyme DNA invertase. Depending on the location of the crossover sites, DNA inversion can give rise to gene fusion or to operon fusion. The underlying flip-flop system can result in microbial populations composed of organisms with different phenotypic appearances: if, for example, the DNA inversion affects the specificity of phage tail fibers, as is the case with phages P1 and μ of *E. coli*, phage populations with two different host ranges will result.

Occasionally, a DNA sequence that deviates considerably from the efficiently used crossover site, a so-called *secondary crossover site*, can serve in DNA inversion, which thus involves a normal crossover site and a secondary crossover site. This process results in novel DNA arrangements, many of which may not be maintained because of lethal consequences or reduced fitness; but if a few new sequences are beneficial to the life of the organism, these may be selectively favored. This DNA rearrangement activity can thus be looked at as evolutionarily important. Since many different DNA sequences can serve in this process as secondary crossover sites, although at quite low frequencies, site-specific DNA inversion systems act as variation generators in large populations of microorganisms. I have thus postulated that this evolutionary role of DNA inversion systems may be more important than

their much more efficient flip-flop mechanism, which can at most help a microbial population to more readily adapt to two different, frequently encountered environmental conditions. As a matter of fact, other strategies could be used as well for the latter purpose.

Computer-aided comparison of DNA sequences quite often reveals that independent genes may consist of a particular domain with high homology and of other DNA sequences showing no significant signs of relatedness. DNA inversion using secondary sites of crossing-over is a potential source of such mosaic genes. DNA inversion can span over relatively large distances in DNA molecules and has the advantage of not loosing any DNA sequences located between the two sites of crossing-over. Deletion formation represents another source for gene fusion, but it has the disadvantage that the DNA sequences between the sites of crossing-over are usually eliminated.

5.2 Transposition of Mobile Genetic Elements

Nine different mobile genetic elements have been found to reside, often in several copies, in the chromosome of *E. coli* K-12 derivatives. This adds up to occupation of about 1% of the chromosomal length by such insertion sequences, also called *IS elements*. At rates on the order of 10^{-6} per individual IS element and per cell generation, these mobile genetic elements undergo transpositional DNA rearrangements. These include simple transposition of an element and more complex DNA rearrangements such as DNA inversion, deletion formation, and the cointegration of two DNA molecules. Because of different degrees of specificity in the target selection upon transposition, the IS-mediated DNA rearrangements are neither strictly reproducible nor fully random. Transposition activities thus also act as variation generators. In addition to DNA rearrangements mediated by the enzyme transposase, which is usually encoded by the mobile DNA element itself, other DNA rearrangements just take advantage of extended segments of DNA homologies at the sites of residence of identical IS elements at which general recombination can act. Altogether, IS elements represent a major source of genetic plasticity of microorganisms.

Transposition occurs not only in growing populations of bacteria but also in prolonged phases of rest. This is readily seen with bacterial cultures stored at room temperature in stabs (little vials containing a small volume of growth medium in agar). Stabs are inoculated with a drop of a bacterial culture taken up with a platinum loop, which is inserted ("stabbed") from the top to the bottom of the agar. After overnight incubation, the stab is tightly sealed and stored at room temperature. Most strains of *E. coli* are viable in stabs during several decades of storage. That IS elements exert transpositional activities under these storage conditions is easily seen as follows.

A stab can be opened at any time, a small portion of the bacterial culture removed, and the bacteria well suspended in liquid medium. After appropriate dilution, bacteria are spread on solid medium. Individual colonies grown upon overnight incubation are then isolated. DNA from such subclones is extracted and fragmented with a restriction enzyme. The DNA fragments are separated by gel electrophoresis. Southern hybridization with appropriate hybridization probes can then show whether different subclones reveal restriction fragment length polymorphisms

(RFLPs), which are indicative of the occurrence of mutations during storage.

If this method is applied to subclones isolated from old stab cultures, and if DNA sequences from residential IS elements serve as hybridization probes, an extensive polymorphism is revealed. None or only little polymorphism is seen with hybridization probes from unique chromosomal genes. Good evidence is available that transposition represents a major source of this genomic plasticity observed in stabs, which at most allow for a very residual growth at the expense of dead cells. One can conclude that the enzymes promoting transposition are steadily present in the stored stabs. Indeed, the IS-related polymorphism increases linearly with time of storage for periods as long as 30 years. In a culture of *E. coli* strain W3110 analyzed after 30 years of storage, each surviving subclone had suffered on the average about a dozen RFLP changes as identified with hybridization probes from eight different residential IS elements, of which IS5 was the most active. Lethal mutations could of course not be identified in this study.

Lethal mutations that affect essential genes for bacteriophage reproduction can be accumulated in the prophage state of the phage genome in its lysogenic host. Such mutants can be screened for their inability to produce infective phage particles upon induction of phage reproduction. Experiments were carried out with *E. coli* lysogenic for a phage P1 derivative grown in batch cultures at 30 °C for about 100 generations allowing for alternative periods of growth and rest. Most of the independent lethal mutants could thereby be identified to be caused by the transposition of an IS element from the host chromosome into the P1 prophage that is maintained in its host as a plasmid. In these experiments, IS2 was the most active element and it mainly inserted into a few preferred regions of the P1 genome, but each time at another site. The used insertion targets did not show any detectable homology or similarity with each other. This is another good example for an enzymatically mediated variation generator. The experiment as such identifies IS transposition as a major source of lethal mutagenesis.

There is no evidence available that bacterial mobile genetic elements would play an essential role in the bacterial life span extending from one cell division to the next. However, these elements are major players in the evolution of bacterial populations. As we have seen here, they contribute to intragenomic DNA rearrangements. Depending on the target sequences involved, the resulting mutations may often be lethal by interrupting essential reading frames or expression control regions. Favorable mutations may be relatively rare, but these can contribute to evolutionarily advantageous developments of the genome. In Sect. 6, we will see that mobile genetic elements also play important roles in the natural strategy of DNA acquisition.

6 DNA Acquisition

While the mutagenesis mechanisms belonging to the strategies of local changes in the DNA sequences (Sect. 4) and of intragenomic DNA rearrangements (Sect. 5) are exerted within the genome and can affect any part of the genome, an additional strategy of spontaneous sequence alterations depends on an external source of genetic information. In DNA or gene acquisition, genetic information indeed originates from an organism other than

the one undergoing mutagenesis. In bacteria, DNA acquisition can occur by means of transformation, conjugation, or virus-mediated transduction. In the latter two strategies of horizontal gene transfer, a plasmid or a virus, respectively, acts as natural gene vector.

The association and dissociation of chromosomal genes with natural gene vectors often arises from transpositional activities and from general recombination acting at IS elements that are at different chromosomal locations. These mechanisms have been well studied with conjugative plasmids and with bacteriophage genomes serving in specialized transduction. For example, composite transposons, which are defined as two identical IS elements flanking a segment of genomic DNA (often with more than one gene unrelated to the transposition process), are known to occasionally transpose into a natural gene vector and, after their transfer into a receptor cell, to transpose again into the receptor chromosome. Hence, together with other mechanisms, such as site-specific and general recombination, transposition represents an important promoter of horizontal gene transfer.

Several natural factors seriously limit gene acquisition. Transformation, conjugation, and transduction depend on surface compatibilities of the bacteria involved. Furthermore, upon penetration of donor DNA into receptor cells, the DNA is very often confronted with restriction endonucleases. These enzymes cause a fragmentation of the invading foreign DNA, which is subsequently completely degraded. Before fragments become degraded, however, they are recombinogenic and may find a chance to incorporate all or part of their genetic information into the host genome. Therefore, we interpret the role of restriction systems as follows: they keep the rate of DNA acquisition low, and at the same time they stimulate the fixation of relatively small segments of acquired DNA to the receptor genome. This strategy of acquisition in small steps can best offer microbial populations the chance to occasionally extend their biological capacities without extensive risk of disturbing the functional harmony of the receptor cell by acquiring too many different functions at once. These considerations have their relevance at the level of selection exerted on the hybrids resulting from horizontal DNA transfer. This selection represents one of the last steps in the acquisition process.

DNA acquisition by horizontal gene transfer is a particularly interesting source of new genetic information for the receptor bacterium because the chance that the acquired DNA exerts useful biological functions is quite high – most likely, it has already assumed the same functions in the donor bacterium.

An interesting hypothesis links the universality of the genetic code with the important role played by horizontal gene transfer in the evolutionary development of the living world. According to this view, those organisms using the most common genetic language would, in the long term, be able to profit best from the increasing worldwide pool of genetic functions under the pressure of adapting to changing living conditions.

7
The Three Natural Strategies Generating Genetic Variations Contribute Differently to the Evolutionary Process

Biological evolution is a systemic process. As outlined above, many different specific mechanisms contribute to generate genetic diversity that represents at any time

the substrate for natural selection. The building up of functional complexity is a stepwise process, in which many random attempts of genetic alterations become rapidly rejected, while relatively few novel sequences are approved as favorable by natural selection and are maintained and amplified. The genome can thus be seen as a cabinet in which key information from favorable historical developments is stored. Stepwise, additional favorable information is added. In the context of changing selective conditions, stored information, having lost its functionally beneficial relevance, may be deleted or favorably altered. As we have seen, a multitude of different mechanisms are behind this dynamic process. For a better understanding of the events, we have grouped the identified mechanisms into three major natural strategies of genetic variation: local changes in the DNA sequences, intragenomic DNA rearrangements, and DNA acquisition. These three strategies have different qualities with regard to their contributions to biological evolution.

The local DNA sequence change is probably the most frequently involved strategy of genetic variation. Indeed, its frequency, which depends primarily on intrinsic properties of matter, chemical instability and conformational flexibility, would be intolerably high if it would not be modulated by efficient enzymatic systems of DNA repair. Local sequence changes bring about nucleotide substitution, the deletion and the insertion of one or a few base pairs, or a local scrambling of a few base pairs. These sequence changes can contribute to a stepwise improvement of a biological function. It must be kept in mind that the functional test for such improvement is carried out by natural selection. In principle, a long series of stepwise local sequence changes could also be expected to bring about a novel biological function. However, this kind of long-term process can gain efficiency only once natural selection starts to be exerted on such upcoming function.

In contrast, the reshuffling of DNA segments within the genome can be considered as a tinkering with existing elements, whereby favorable gene fusions and operon fusions may occasionally result. DNA rearrangement can also be the source of gene duplication and higher amplification, which are widely recognized contributions to the evolutionary progress.

In Sect. 6, we have already pointed to the evolutionarily high efficiency of horizontal gene transfer. As a matter of fact, DNA acquisition allows the recipient organism to share in the success of evolutionary developments made by others. In drawing the evolutionary tree of bacteria, DNA acquisition should be accounted for by more or less randomly adding temporal horizontal shunts between individual branches. It must be kept in mind that usually only small DNA segments flow through such shunts in horizontal gene transfer.

Several of the specific mechanisms of genetic variation employ, strictly speaking, more than one of the three strategies shown in Fig. 1. In transposition of IS elements, for example, a chromosomal DNA segment consisting of the mobile genetic element can undergo a translocation and thereby become inserted at a new target site. As a rule, the target sequence thereby gets duplicated, which usually involves a few nucleotides. Thus, this transposition event will consist of both a DNA rearrangement and a local sequence change.

As far as we know, most of the well-studied microbial strains use in parallel each of the three natural strategies for the generation of genetic variations. In

addition, bacteria very often use not only one, but several different specific mechanisms for mutagenesis by each of the strategies. Dissimilar specific mechanisms often work with different efficiencies as reflected by their contribution to the overall mutation rate. For any given strategy, it might be less relevant which specific mechanism is at work than the fact that the particular strategy finds its application with an evolutionarily useful efficiency. In other words, specific mechanisms may substitute for each other within a strategy, at least to some degree. This rule does not apply between the strategies because of the difference in the qualities of their evolutionary contributions.

The efficiency displayed by a given specific mechanism of spontaneous mutagenesis may depend on both internal (e.g. availability of enzymes that mediate mutagenesis) and external factors (e.g. environmental stress). It is also to be noted that some mechanisms may act more or less randomly along a DNA molecule, while other mechanisms may show regional or site preferences for their activities. In view of these considerations, we tend to assume that an evolutionarily fit (or well prepared) organism should best be able to use a few specific mutagenesis mechanisms for each of the three strategies to generate genetic variations. In Sect. 8, we will explain what we mean by evolutionary fitness.

8 Evolution Genes and Their Own Second-order Selection

The attentive reader will have seen in the description of some specific mechanisms contributing to the spontaneous mutagenesis that besides a number of nongenetic factors, specific products of genes are very often at work. These gene products can belong to systems for repair of DNA damage and will, in this case, modulate the frequency of mutagenesis. Similarly, restriction enzymes seriously reduce both the chance of DNA acquisition and the size of a DNA segment that may eventually be acquired by the recipient cell. Other gene products such as transposases and other mediators of DNA recombination act as generators of genetic variations. Since variation generators and modulators of the frequencies of genetic variation are evolutionary functions, we call the underlying genetic information *evolution genes*. In the microbial world, these evolution genes generally play no essential role in the physiology of individual lives going by cell division from one generation to the next. Under laboratory conditions, neither restriction–modification systems nor enzymes for DNA rearrangements are needed for the propagation of bacteria. The role of such enzyme systems is primarily evolutionary and becomes manifest at the level of populations.

We assume that evolutionary genes are themselves submitted to selective pressure. However, such selection cannot follow the rules of direct selection for improvements of essential functions such as those of housekeeping genes. Rather, the selection for the presence and improvement of a variation generator will be exerted at the level of populations. Clearly, it will also be an individual that may one day undergo a mutation, which improves an evolutionary function. This function will also be exerted in its progeny in which appropriate genetic variants of genes for directly selected products will be either more or less abundant. Any genetic alteration that affects an evolution gene and proves in the long term to be of higher evolutionary

value will be maintained and will provide an evolutionary benefit to the carrier of the involved gene. In the long run, this will lead to fine-tuning of the evolutionary functions of both variation generators and of modulators of the frequency of genetic variation. The underlying indirect selection based on the cells ability to provide genetic variants at a well-balanced level is called *second-order selection*.

We must be aware that some gene products may exert their essential functions for the benefit of both the life of the individual and the evolutionary progress of the population. In these cases, we assume that evolutionary selection is exerted for both kinds of functions and will eventually bring the gene to a fine-tuned state to optionally carry out its functions for each of the different purposes. However, as we have already mentioned, a number of gene products involved in genetic variation are inessential for the lives of individual bacterial cells. Similarly, the products of many housekeeping genes are inessential for biological evolution.

9 Arguments for a General Relevance of the Theory of Molecular Evolution for All Living Organisms

Largely on the basis of evidence from microbial genetics, we have so far postulated that the products of a number of evolution genes contribute, each in its specific way, to the generation of genetic variants at evolutionarily useful frequencies. Thereby, the sources of mutagenesis may relate either to the activity of the evolution gene product itself (e.g. a transposase) or to a nongenetic factor (e.g. a chemical mutagen or an intrinsic structural flexibility of a nucleotide). In many cases, nongenetic factors and products of evolution genes cooperate in the formation of genetic variants at physiologically tolerable and evolutionarily beneficial levels. This is, for example, the case in spontaneous mutagenesis by an environmental mutagen when some of the primary damage on the DNA gets successfully repaired while some other damage leads to a fixed mutation.

The theory of evolution postulates that life on Earth started almost four billion years ago with primitive, unicellular microorganisms. It is in the first two billion years that microbes must have developed the basis for the actual setup of evolutionary strategies and the underlying evolution genes. One can postulate that the acquired evolutionary capacities could have allowed some microbial populations to undergo a division of labor in more and more stable associations of cells. This development might later have led to multicellular organisms. In this kind of development, the evolutionary fitness of the involved organisms might have been an important precondition. At still later stages of further evolutionary development, the three natural strategies for generating genetic variations (Sects. 4 to 7) must have continued exerting their evolutionary influence together with some other factors such as the formation of symbiotic associations. As a matter of fact, we attach considerable evolutionary relevance to endosymbiosis of higher organisms with bacteria. Such situations of cohabitation may form an ideal condition for occasional horizontal gene transfer between the close associates.

A scientifically justified quest for further experimental proof of the postulates of the theory of molecular evolution remains quite difficult to be answered. Clearly, there is a need for research on

the spontaneous generation of genetic variation in higher organisms, ideally at the level of the genomes. While this is already quite difficult in microorganisms, it is of increased perplexity with the much larger genomes of higher organisms. However, sequence comparisons now offer fruitful ways to search for sequence homologies, sequence similarities, single nucleotide polymorphisms, as well as for traces of intragenomic DNA rearrangements and of horizontal transfer of genetic information. Data so far available are in support of the principles of molecular evolution outlined in Sect. 2, which are likely to be valid for any kind of living organism.

Some of the general evolutionary strategies developed in microorganisms must have also turned out to be useful for the developmental and physiological processes at somatic levels of higher organisms. The generation of antibody diversity in the immune systems of vertebrates by genetic rearrangements and so-called *somatic mutagenesis* is a good example. Another example is the enzymatic repair of DNA damage caused in somatic cells by external mutagens such as UV irradiation.

These considerations illustrate that whatever gene function may prove to be useful for whatever particular purpose, it may be evolutionarily retained and in the course of time further fine-tuned. We have already encountered this principle in the microbial world, where we have postulated multifunctional enzymes (such as working both for the physiology of the cells and for an evolutionary task) to become evolutionarily improved both by direct and by second-order selection for the various functions. This may also be the case in higher organisms.

10 Conceptual Aspects of the Theory of Molecular Evolution

With reference to Sect. 2.2, it is fair to again explicitly state that for the time being, evolution genes and evolution functions are a concept rather than a fully proven fact. This concept is based on a particular way to interpret numerous available experimental data. We will briefly analyze the difficulties to clarify the situation in a scientific debate. This will be followed by pointing to philosophical and more practical values of a deeper understanding of the molecular processes that drive biological evolution.

10.1 Pertinent Scientific Questions

In the history of scientific investigations, biologists have often searched for evidence that living organisms would be able to specifically modify, or adapt, their genetic information in order to better cope with upcoming changes in the living conditions. Most of these attempts have failed to give the expected response. In other cases, where a certain degree of specific adaptation could be observed, specific causal explanations have sometimes been found upon deeper investigation. However, there is at present no good scientific evidence for a general rule that genetic alterations would always be directed toward a specific goal. This situation favors the view that spontaneous mutations affect DNA more or less randomly, which is in line with the general observation that only a minority of spontaneous mutants prove to be favorable under the encountered living conditions.

The postulate of evolution genes that act as generators of genetic variations may be a surprise in this context. This has to do with a widely followed concept of genetic

information as a strict program for the fulfillment of a specific task. This definition does apply to many housekeeping genes, the products of which efficiently catalyze a reaction that reliably always yields the same product. This is not how a variation generator that does not work efficiently and that yields a different product from case to case functions. A good example is the transposition of mobile genetic elements. However, not all scientists see the primary function of a mobile genetic element in genetic variation. Rather, some colleagues interpret IS translocation, which often goes along with the replication of the element, as a selfish activity. This interpretation considers mobile genetic elements as parasites with the argument that their activity would often harm their host cell.

This discussion shows that the concept of evolution genes cannot easily be defended by referring to scientific evidence. Rather, the concept reflects an attitude of the observer of natural events. According to the view defended in this article, nature actively cares for biological evolution. The products of evolution genes are actively involved in generating different kinds of genetic variants at frequencies insuring both a certain genetic stability required for maintaining the concerned form of life and a low frequency of genetic variations as the driving force of evolution. This interpretation recognizes biological evolution as an essential principle of self-organization of life on Earth.

Another pertinent question that cannot find an easy scientific answer relates to the evolutionary function of viruses. At least some viruses are clearly identified to sometimes act as gene vectors in horizontal gene transfer. Some of them also temporarily integrate their genome into the host genome. This relates to the lysogenic state of bacteria as well as to endogenous viruses such as retroviruses that reside in many higher organisms. Again, one may wonder if these viruses primarily fulfill evolutionary functions for the evolutionary development of their hosts or whether they should rather be looked at as parasites that may carry out some evolutionary function by accident.

While prokaryotic organisms have genomes that are relatively densely packed with functional genes, many higher organisms have extended segments of intergenic DNA sequences, some of which are highly repetitive. Some of these noncoding sequences are highly homogeneous with regard to their nucleotide composition. While the biological roles played by noncoding regions are still not well understood, it has been postulated that compositional constraints may influence natural selection. These aspects have not been covered in this article and they may more specifically relate to the molecular evolution of higher organisms, in addition to the principles outlined here largely on the basis of evidence from microbial genetics.

10.2 Philosophical Values of the Knowledge on Molecular Evolution

One of the central questions of human curiosity is to know where life – and more specifically human life – comes from. The Darwinian theory opposed to the idea of a specific act of creation for each particular form of life, the alternative explanation of a steady evolutionary development implying the descent of the actual species from common ancestors. The directions of evolution are thereby given by natural selection acting steadily on all available forms of life

including all present variants. Until recently, the sciences could not specifically explain how genetic variants are generated. With recently developed research strategies, molecular genetics can now fill this gap. The branch of science called *molecular evolution* explains that there is not just a single source for genetic variants. Rather, many different specific mechanisms contribute to the generation of genetic variants at low frequencies. These mechanisms follow one and sometimes more than one basic strategies of evolutionary development. These are local changes in the DNA sequences, rearrangements of DNA segments within the genome, and acquisition of segments of foreign DNA by horizontal gene transfer. As a general rule, spontaneous mutagenesis is not specifically directed; it is at least to some extent random, so that only a minor fraction of spontaneous genetic variants turn out to be favorable for the concerned organism and thus provide it with a selective advantage. Nevertheless, new knowledge on the precise molecular mechanisms of the generation of genetic variations provides strong evidence that, in many cases, specific enzymes are involved – the products of evolution genes. These products work in tight collaboration with nongenetic factors that can be intrinsic properties of matter or environmental conditions. This view of the evolutionary process represents the core of a theory of molecular evolution and it can be seen as an extension of neo-Darwinism at the molecular level.

It should be clearly stated that the theory of molecular evolution does not explain the origin of life. It can, however, explain that biological evolution exerted in all living beings is a steady, dynamic process that is actively promoted not only by intrinsic properties of matter but also by the intervention of products of evolution genes or more generally of evolutionary functions of many different gene products. A recently published book devoted to these exciting insights is entitled *Darwin in the Genome*.

The high philosophical value of this extension of our worldview is obvious and merits to be widely discussed and evaluated in its various cultural dimensions. One interesting aspect is the implied duality of the genome. Indeed, evolution genes are located together with all the other genes on the genome and on accessory DNA molecules such as plasmids and viral genomes. While probably a major part of gene products carries out functions to benefit the cell and the individual, often a multicellular organism, probably a minority of gene products works for the biological evolution of the concerned population. Note, however, that the generation of a novel genetic variant obviously also occurs in an individual cell. But this act of creation has only a small chance to bring to the concerned cell and to its descendants a selective advantage. More often, the mutation is unfavorable and renders the life of the concerned organism more troublesome. As long as the spontaneous rate of mutagenesis remains low (mostly thanks to the intervention of evolution genes), unfavorable mutations are tolerable at the level of propagating populations.

Incidentally, the situation described here offers a possible explanation to the quite difficult theodicean question: why does God, despite His love for the human creature, admit that physically evil events such as a mutation causing an inheritable disease can occur to individuals? As a matter of fact, Genesis describes creation as a stepwise process, which implies the permanent evolutionary expansion of the diversity of life forms. Genesis also states that God evaluated this system as good.

Hence, biological evolution occurs according to God's intention to amplify diversity of life on our planet. Occasional unfavorable mutations affecting rare individuals in populations is a sacrifice brought to the creative force residing in the system of molecular evolution.

In brief, the genetic information contained in each genome – of bacteria as well as of all higher organisms – represents an internal duality. It serves individuals for the fulfillment of their individual lives and it serves populations for a slow but steady expansion of life forms and thus of biodiversity.

10.3 Aspects Relating to Practical Applications of Scientific Knowledge on Molecular Evolution

Living organisms today occupy an amazing variety of ecological niches on our planet Earth. These niches include extreme physicochemical conditions such as elevated temperatures, high pressure, and quite unusual compositions of chemical elements. However, despite the intrinsic potential of the living world to evolutionarily expand, one can estimate the carrying capacity of the planet for life to be in the order of 10^{30} living cells. Although this is a very large number, it seriously limits free expansion of life in its various forms.

The following reflections should help illustrate this statement. An adult human being carries in the order of 10^{13} cells. The human population today, thus occupies a share of about 10^{23} cells of the available 10^{30}. Incidentally, this happens to be close to the average available for each of the estimated 10^{7} different species of organisms on the planet. Bacteria propagate by cell division as outlined in Sect. 3. Extending the reflections made there, one can conclude that from an inoculum with one single bacterial cell, one can theoretically expect to obtain 10^{30} cells within only 50 h. In reality, growth will be stopped much earlier by lack of nutrition, but this reflection illustrates well the enormous internal forces for expansion of a given form of life.

Similarly, a high potentiality for evolutionary expansion toward more diversity resides in the mechanisms of molecular evolution that are described in this article. These mechanisms can serve us as a basis to better understand both the origin and the steady replenishment of biodiversity as well as the internal limits set to evolutionary expansion. This knowledge can and should increasingly be used as a background for any measures taken toward the protection of biodiversity and of habitats for diverse forms of life. Last but not the least, a better understanding of the evolutionary process can be of help to render the development of agricultural and related practices more sustainable.

Genetic engineering offers ample new possibilities for the sustainable production of medical drugs, to obtain food of higher quality, and to reduce the nocent impact of the human civilization on the environment. The serious reservations made by large parts of the human population impede many of the proposed biotechnological applications. A part of these concerns refer to unpredictable long-term effects of genetically modified organisms (GMO) that are released into the environment, such as in agricultural applications. Scientific assessments of long-term and, in particular, evolutionary effects of such applications are thus required. Knowledge of the natural strategies of molecular evolution can provide a good basis for such studies. As a matter of fact, in genetic engineering, DNA sequence alterations are

brought about within the genome by site-directed mutagenesis in studies of the biological functions of specific genes. In addition, well-defined segments of DNA are introduced into other organisms either in view of amplifying the particular DNA segment or in view of harvesting a specific gene product. GMO can also directly serve in applications as GM food and for bioremediation by microorganisms. A candid comparison of these practices involving genetic engineering with the natural strategies of generation of genetic variations reveals a high degree of similarities. The amounts of nucleotides or lengths of DNA sequences involved in these genetic modifications, both in genetic engineering and in the natural genetic variation, are of the same order of magnitude. Depending on the strategy involved, they may concern one to a few base pairs, or in other instances, a DNA segment containing a sequence domain or one to a few genes, both in intragenomic DNA rearrangements and in the horizontal transfer of DNA between two different organisms. Thus, one can principally expect that long-term evolutionary risks of GMO compare with the biohazard intrinsic to the natural process of biological evolution. Similar risks may also be inherent in classical breeding techniques.

These considerations ask for a more integral, holistic, and critical evaluation of the impact of past and present human activities on the natural process of biological evolution. Such assessments should address any human impact on genetic variation, natural selection, and isolation. As far as we know from the long-term history, the foundations of life and its evolutionary development on Earth are relatively robust. This is good news for us human beings, but it should not exempt us from a responsible and well-reflected use of our scientific knowledge in any attempt to render our own lives more easy and comfortable.

See also Bacterial Growth and Division; Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Caporale, L.H. (Ed.) (1999) *Molecular Strategies in Biological Evolution*, Vol. 870, Annals of the New York Academy of Sciences, New York, NY.

Caporale, L.H. (2003) *Darwin in the Genome: Molecular Strategies in Biological Evolution*, McGraw-Hill, New York.

Kucherlapati, R., Smith, G.R. (Eds.) (1988) *Genetic Recombination*, American Society for Microbiology, Washington, DC.

Moses, R.E., Summers, W.C. (Eds.) (1988) *DNA Replication and Mutagenesis*, American Society for Microbiology, Washington, DC.

Shapiro, J.A. (1983) *Mobile Genetic Elements*, Academic Press, New York.

Primary Literature

Arber, W. (1991) Elements in microbial evolution, *J. Mol. Evol.* **33**, 4–12.

Arber, W. (1993) Evolution of prokaryotic genomes, *Gene* **135**, 49–56.

Arber, W. (1995) The generation of variation in bacterial genomes, *J. Mol. Evol.* **40**, 7–12.

Arber, W. (2000) Genetic variation: molecular mechanisms and impact on microbial evolution, *FEMS Microbiol. Rev.* **24**, 1–7.

Arber, W. (2002) Evolution of prokaryotic genomes, *Curr. Top. Microbiol. Immunol.* **264/I**, 1–14.

Arber, W. (2002) Molecular evolution: comparison of natural and engineered variations, *The Pontifical Academy of Sciences. Scripta Varia* **103**, 90–101.

Arber, W. (2003) Elements for a theory of molecular evolution, *Gene.* **317**, 3–11.

Arber, W. (2003) Traditional Wisdom and Recently Acquired Knowledge in Biological

Evolution, *Proceedings of UNESCO Conference on* "Science and the Quest for Meaning"; in press.

Arber, W., Hümbelin, P., Caspers, P., Reif, H.J., Iida, S., Meyer, J. (1981) Spontaneous mutations in the *Escherichia coli* prophage P1 and IS-mediated processes, *Cold Spring Harbor Symp. Quant. Biol.* **45**, 38–40.

Arber, W., Naas, T., Blot, M. (1994) Generation of genetic diversity by DNA rearrangements in resting bacteria, *FEMS Microbiol. Ecol.* **15**, 5–14.

Bernardi, G. (2000) Isochores and the evolutionary genomics of vertebrates, *Gene* **241**, 3–17.

Bernardi, G. (2000) The compositional evolution of vertebrate genomes, *Gene* **259**, 31–43.

Drake, J.W. (1991) Spontaneous mutation, *Annu. Rev. Genet.* **25**, 125–146.

Echols, H., Goodman, M.F. (1991) Fidelity mechanisms in DNA replication, *Annu. Rev. Biochem.* **60**, 477–511.

Galas, D.J., Chandler, M. (1989) Bacterial Insertion Sequences, in: Berg, D.E., Howe, M.M. (Eds.) *Mobile DNA*, American Society for Microbiology, Washington, DC, pp. 109–162.

Glasgow, A.C., Hughes, K.T., Simon, M.I. (1989) Bacterial DNA Inversion Systems, in: Berg, D.E., Howe, M.M. (Eds.) *Mobile DNA*, American Society for Microbiology, Washington, DC, pp. 637–659.

Iida, S., Hiestand-Nauer, R. (1987) Role of the central dinucleotide at the crossover sites for the selection of quasi sites in DNA inversion mediated by the site-specific Cin recombinase of phage P1, *Mol. Gen. Genet.* **208**, 464–468.

Lorenz, M.G., Wackernagel, W. (1994) Bacterial gene transfer by natural genetic transformation in the environment, *Microbiol. Rev.* **58**, 563–602.

Naas, T., Blot, M., Fitch, W.M., Arber, W. (1994) Insertion sequence-related genetic variation in resting *Escherichia coli* K-12, *Genetics* **136**, 721–730.

Naas, T., Blot, M., Fitch, W.M., Arber, W. (1995) Dynamics of IS-related genetic rearrangements in resting *Escherichia coli* K-12, *Mol. Biol. Evol.* **12**, 198–207.

Sandmeier, H. (1994) Acquisition and rearrangement of sequence motifs in the evolution of bacteriophage tail fibers, *Mol. Microbiol.* **12**, 343–350.

Sengstag, C., Arber, W. (1983) IS2 insertion is a major cause of spontaneous mutagenesis of the bacteriophage P1: non-random distribution of target sites, *EMBO J.* **2**, 67–71.

Weber, M. (1996) Evolutionary plasticity in prokaryotes: a panglossian view, *Biol. Philos.* **11**, 67–88.

West, S.C. (1992) Enzymes and molecular mechanisms of genetic recombination, *Annu. Rev. Biochem.* **61**, 603–640.

Woese, C.R. (1987) Bacterial evolution, *Microbiol. Rev.* **51**, 221–271.

Genetics and Gene Therapy in Cardiovascular Diseases: *see* Gene Therapy and Cardiovascular Diseases

Genetics and Molecular Biology of Lung Cancer

Jack A. Roth
The University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

1 Familial Predisposition to Lung Cancer 355

2 Oncogene Activation 356
2.1 *myc* Family 356
2.2 *ras* Family 358
2.3 *ras* in Carcinogen-induced Tumors 358
2.4 Human Studies 358

3 Growth Factors and Autocrine–Paracrine Growth Control 360
3.1 c-*erb B-2* 361
3.2 TGF-α 361

4 Tumor Suppressor Genes 362

5 Premalignancy and Second Primary Cancers 364

6 Conclusions and Future Prospects 365

Bibliography 366
Primary Literature 366

Keywords

Non–small-cell Lung Cancers
A group of lung cancers, including adenocarcinoma, squamous carcinoma, and large-cell undifferentiated carcinoma, which together comprise 80% of lung cancers.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

Oncogene
A gene that has a dominant transforming effect on normal host cells frequently mediated by growth factors, growth factor receptors, or intermediaries in the signal transduction process.

Small-cell Lung Cancer
A specific lung cancer histology in which the cells frequently contain neurosecretory granules. The cancers, often disseminated at diagnosis, initially respond well to chemotherapy but recur rapidly.

Tumor Suppressor Gene
A gene whose homozygous deletion or inactivation contributes to transformation of the host cell.

Lung cancer is a paradigm of carcinogen-induced human cancer. It is estimated that for every 3 million cigarettes purchased, a lung cancer follows 35 years later. Although the majority of lung cancers occurs in men, the incidence in women is rising rapidly in some regions.

Epidemiologic studies have irrefutably linked cigarette smoking and human lung cancer as well as other etiologic agents such as asbestos and radon. However, the molecular events within the cell that culminate in the development and progression of human lung cancer remain poorly understood. This is an important area for future research for several reasons. Elucidation of the molecular events in the genesis of lung cancer may have relevance to mechanisms for other carcinogen-induced tumors. Prevention, early diagnosis, and treatment strategies will be more rationally designed if they are based on molecular mechanisms.

The observation that genes responsible for carcinogenesis were altered forms of genes normally present in eukaryotic cells initiated many of the advances in molecular biology that have increased our understanding of lung carcinogenesis at the molecular level. Many of these genes have been implicated in the development of human cancer. An understanding of the molecular basis for lung carcinogenesis is essential for the development of improved methods of diagnosis, staging, treatment, and prevention for lung cancer.

The cell of origin for lung cancers is controversial. All histologies for non–small-cell lung cancer (NSCLC) have phenotypic features of the differentiated cell types in normal or injured bronchial epithelium. Small-cell lung cancer (SCLC) cells have neuroendocrine markers including high levels of the polypeptide hormones gastrin-releasing peptide and calcitonin, creatine kinase isoenzyme BB, L-dopa decarboxylase, and neuron-specific enolase. Endocrine cells can be found in normal bronchial mucosa. Thus, one possibility is that each of the four major histologies arises from alterations in its preexisting normal counterpart.

An alternative hypothesis is that the four types of lung cancer arise from a common stem cell and are related through a common differentiation pathway of the normal bronchial epithelium. This is supported by the clinical observation that SCLC tumors

can contain mixtures of SCLC and non-SCLC histologies. These transitions have been observed *in vitro* following insertion of the appropriate oncogene. For example, insertion of a mutated H-*ras* oncogene in SCLC cells with overexpression of c-*myc* causes transition to the large-cell undifferentiated phenotype.

1 Familial Predisposition to Lung Cancer

That the vast majority of cigarette smokers, including heavy smokers, do not develop lung cancer suggests the existence of an inherited predisposition or cofactors such as additional carcinogens. Studies have compared risk factors in individuals with histologically verified lung cancer and individuals with other smoking-related cancers. There was lack of increased risk of developing lung cancer when only lung cancer in relatives was considered. However, there was a significant excess of cancers at all sites for relatives of the lung cancer patients. This suggests a heritable variation in response to carcinogens. Respiratory diseases also predispose to development of lung cancer. Lung cancer families have been studied. The development of lung cancer in young individuals (50 or less) was compatible with Mendelian codominant inheritance of a rare autosomal gene. This gene was not involved for older persons, reflecting the medical history of noncarriers having long-term exposure to tobacco. The aryl hydrocarbon hydroxylase gene product can metabolize promutagenic and procarcinogenic compounds in cigarette smoke. The aromatic hydrocarbon-inducible cytochrome P4501 A I gene is highly expressed at the RNA level in normal lung from active cigarette smokers, but not in normal lung from nonsmokers. The ability to metabolize debrisoquine is genetically determined and is associated with susceptibility to lung cancer. Studies have shown that the ability to metabolize debrisoquine is independently associated with susceptibility to lung cancer in a case–control study.

The rapid development of technology in molecular biology has resulted in the identification of genetic alterations in human lung cancer. However, the plethora of genetic abnormalities detected has created some confusion as to their significance in the genesis and progression of lung cancer. Small-cell lung cancer has been extensively studied, partly because of the availability of many established SCLC cell lines that can be grown in serum-free, chemically defined media. However, only 20% of lung cancers are of the small-cell type. Furthermore, the biologic behavior of SCLC is quite distinct from NSCLC. SCLC disseminates early in its course and exhibits a marked sensitivity to chemotherapy followed by early recurrence. Thus, it is possible that genetic events in SCLC and NSCLC may differ. Recent studies summarized here indicate differences. Thus, molecular mechanisms may not be generalizable among the different forms of lung cancer. The relative importance and sequence dependence of these genetic alterations in the genesis and progression of lung cancer is unknown.

This article summarizes areas of recent investigation of molecular events in human lung cancer including oncogene

Tab. 1 Oncogenes and "tumor suppressor genes" altered in lung cancer oncogenes.

c-*myc*, L-*myc*, N-*myc*, c-*myb*, K-*ras*, N-*ras*, H-*ras*, c-*raf*, c-*fur*, c-*fes*, c-*fms*, c-*erb B-1* (EGFR), c-*erb B-2* (HER2, *neu*), c-*sis*, c-*rlf, bcl*
Tumor Suppressor Genes *p53*, *RB*, *APC*, *CDH13*, *RARβ*, *FHIT*, *RASSF1A*, *TIMP3*, *p16*, *MGMT*, *DAPK*, *FUS1*

Source: Minna J.D., Roth J.A., Gazdar A.F. (2002) Focus on lung cancer, *Cancer Cell* **1**, 49–52.

activation and inactivation of tumor suppressor genes. Table 1 lists the genes that have been implicated in lung carcinogenesis.

2 Oncogene Activation

2.1 *myc* Family

One of the first specific genetic changes associated with lung cancer was the observation that the c-*myc* oncogene was amplified in a subgroup of SCLC cell lines. The SCLC cell lines having an amplified c-*love* gene are morphological and biochemical variants of SCLC (SCLC-V). SCLC-V cell lines have a rapid doubling time, higher cloning efficiency, increased tumorigenicity, and increased resistance to X-rays compared to SCLC. In addition, SCLC-V cell lines do not express L-dopa decarboxylase or peptide hormones, which are typically expressed on cells of neuroendocrine origin. They are distinguished from NSCLC, however, by having elevated levels of the BB isoenzyme or creatine kinase and neuron-specific enolase. It has been reported that SCLC-V cell lines showed high levels of c-*myc* amplification and c-*myc* mRNA levels. Only one NSCLC cell line out of five in the report showed c-*myc* amplification. The c-*myc* gene was transfected into the H209 classic SCLC cell line. One of the transfectants expressing high levels of c-*myc* had an increase in doubling time and increased cloning efficiency, but L-dopa decarboxylase levels and bombesin-like immunoreactivity were unchanged.

Amplification of c-*myc* was observed in both classic and variant SCLC cell lines. However, c-*myc* mRNA levels were more elevated in the variant cell lines. Three classic lines had amplification of N-*myc* RNA and one variant line had amplification of N-*myc* and *myb*. Three SCLC-V cell lines showed high levels of a *v-fms*-related transcript related to but not identical with the *CSF 1I* receptor. Subpopulations of SCLC cells with intermediate neuroendocrine differentiation have been identified. The cell lines expressed some neuroendocrine markers such as L-dopa decarboxylase, but not others such as bombesin and neurotensin. These cell lines also had high levels of c-myc protein. Expression of c-myb protein was seen for two cell lines in which c-myc expression was low or not detectable. Expression of c-rafl protein was low in 11 of the 12 cell lines. Four of five NSCLC cell lines expressed c-myc protein at high levels, and these were all of the large-cell undifferentiated morphology. In another study, 3 of 27 NSCLC DNA samples from primary tumors had amplification of the c-*myc* gene. Others have identified high levels of c-*myc* mRNA in an NSCLC cell line in the absence of DNA amplification. Run-on transcription studies for this cell line showed that the transcriptional rate *for* c-*myc* was high.

Analysis of SCLC cell lines for c-*myc* amplification revealed additional EcoRl restriction fragments suggesting *myc*-related

genes. A third gene in the *myc* family, L-*myc*, was cloned and showed homology to c-*myc* and N-*myc*. Four SCLC cell lines had amplified L-*myc* genes. The L-*myc* gene has been cloned and sequenced and consists of 3 exons and 2 introns spanning 6.6 kb of human DNA. There is homology with discrete regions of N-*myc* and c-*myc*. The L-*myc* gene encodes a series of nuclear phosphoproteins that arise by alternative mRNA processing. L-*myc* can cooperate with an activated c-Ha-ras to transform primary rat-embryo fibroblasts. However, the transforming efficiency was 1 to 10% of that seen with c-*myc*. One study found a correlation between restriction fragment length polymorphisms (RFLPs) of the L-*myc* gene and lymph node metastases in NSCLC. The presence of either the S band (6 kb) or the S and L (10 kb) bands was associated with lymph node metastases.

Amplification and increased expression of the N-*myc* gene occurs in SCLC and NSCLC. Expression of N-*myc* in SCLC biopsies was detected by *in situ* hybridization. Increased expression was associated with poor response to chemotherapy and short survival. Amplification of N-*myc* gene sequences ranging from 5- to 170-fold was observed in SCLC cell lines. Both c-*myc* and N-*myc* were amplified, but only one member of the myc family was amplified in any one cell line. Amplification of N-*myc* has been reported in an adenocarcinoma of the lung. Others have found amplification of one of the *myc* family in 2 of 12 SCLC cell lines. Again, only one member of the family was amplified in each cell line. All cell lines had deletions of chromosome 3. When fresh tumor specimens were analyzed, amplification and rearrangement of *myc* genes was heterogeneous. N-*myc* or L-*myc* amplification was noted in 4 of 17 small-cell cancers. Amplification of c-*myc* was seen in 3 of 12 non–small-cell lung cancers. In some cases, amplification was seen in the primary tumor but not metastases. In two cases, amplification was seen only in cell lines, not in the original tumors. Expression of *myc* family genes was demonstrated in SCLC cell lines and nude-mouse xenografts using *in situ* hybridization techniques. The molecular mechanisms regulating the expression of each of the *myc* family genes are complex. Both c-*myc* and L-*myc* mRNA showed loss of transcriptional attenuation, which correlated with overexpression seen in cell lines without gene amplification. Regulation of N-*myc* expression correlated with promoter activity and gene amplification.

An interesting association between responsiveness to bombesin and *myc* family expression has been noted by one investigator. SCLC cell lines responsive to bombesin showed constitutive expression of L-*myc*. Nonresponsive cell lines expressed N-*myc* or c-*rnyc*.

The significance of increased expression of *myc* family genes remains uncertain. Initially, c-*myc* amplification was described in SCLC cell lines with variant morphology. This variant morphology is also called *small-cell/large-cell carcinoma* and is thought to indicate an unfavorable prognosis. Cell lines with the variant morphology have relatively more resistance to chemotherapy and radiation therapy. However, review of pathology specimens of patients with extensive-disease SCLC showed that the variant cell type was rare, occurring in only 4.4% of 550 specimens. There were no significant differences in response rates to chemotherapy or prognosis for patients with "classic" compared with variant morphology. Amplification of the c-*myc* gene was more frequent in cell lines from SCLC patients with tumor relapse than in untreated patients. Amplification of c-*myc* was associated with

shorter survival in relapsed patients. The role of increased expression of *myc* family genes in the genesis of lung cancer requires further definition. It is likely that increased c-*myc* expression leads to progression of SCLC; it appears unlikely to be a primary event because it is detected in a minority of tumors. The association of increased c-*myc* expression with the variant cell type and the significance of this cell type require additional clarification. Increased expression may occur by several mechanisms and is not always associated with gene amplification. Alterations in *myc* expression in NSCLC have not been extensively studied, but in one case several NSCLCs showed increased expression of c-*myc*.

2.2
ras Family

Oncogenes of the ras family (homologous to the rat sarcoma vines) have three primary members (H-*ras*, K-*ras*, and N-*ras*) and are among the most common activated oncogenes found in human cancer. The *ras* genes code for a protein that is located on the inner surface of the plasma membrane, have GTPase activity, and may participate in signal transduction. *ras* oncogenes are activated by point nucleotide mutations that alter the amino acid sequence of p21.

2.3
ras in Carcinogen-induced Tumors

Studies in mice with carcinogen-induced lung cancers implicate genes of the ras family in the carcinogenesis process. Mouse lung tumors induced by tetranitromethane contained mutated K-*ras* genes. Mice harboring the mutated H-*ras* transgene developed tumors exclusively in the lungs within weeks of birth. Lung tumors can be induced in mice with the tobacco-specific nitrosamines 4-(*N*-methyl-*N*-nitrosamino)-1-(3-pyridyl)-1-butanone (NNK) or nitrosodimethylamine (NDMA). Ninety percent of these tumors had transforming genes in the NIH3T3 assay, and in all lung tumors, this was K-*ras*. The mutations were generally GC-to-AT transitions, indicating that DNA methylation is the most likely pathway to the induction of neoplasia by these carcinogens.

2.4
Human Studies

Amplification of *ras* oncogenes is uncommon in lung cancer. Heighway and Hasleton found no amplification of K-*ras* in 25 primary specimens. However, in one study a lymph node metastasis showed 30-fold amplification of K-*ras*. Expression of p21 as measured by binding of monoclonal antibody rp-35 increases with increasing size of the primary tumor.

Activation of the K-*ras* oncogene by point mutation occurs in lung cancer cell lines. A mutation in codon 12 (glycine to cysteine) occurs in *Calu* 1 and in codon 61 in *PR310* (glutamine to histidine). RFLP analysis for a codon 12 mutation (glycine to arginine) did not show any mutations in 24 primary NSCLC lung cancers. Other mutations have been detected in the 12th codon using a highly sensitive technique based on amplification with the polymerase chain reaction and detection with a panel of oligonucleotide probes. K-ras mutations in most studies were confined to adenocarcinomas of the lung and occurred in 30% of tumors. Mutations were not observed in adenocarcinomas from nonsmokers. K-ras mutations appear to be an independent prognostic factor, indicating

a poor prognosis, and are unrelated to conventional staging criteria such as tumor size or lymph node metastases. In all patients studied, mutations occurred in a single allele.

ras mutations can also influence the differentiation of tumor cells. For example, infection of SCLC cell lines with the Harvey murine sarcoma virus alters the phenotype of variant cells but not that of classic cells. Following infection, the variant SCLC cell line developed features of a large-cell undifferentiated lung carcinoma including increased carcinoembryonic antigen (CEA) and keratin expression.

Studies done to date favor the interpretation of *ras* activation as a progression factor in lung cancer. It apparently is activated in about one-third of adenocarcinomas arising in patients with a heavy smoking history. Antisense (AS) technology has been used to find the effects of eliminating expression of a mutant K-*ras* oncogene in NSCLC cells. A homozygous mutation at codon 61 was detected in the NCI-H460a large-cell undifferentiated NSCLC cell line clone with a normal glutamine residue (CAA) substituted by histidine (CAT), using hybridization with specific oligonucleotide probes. An AS K-ras RNA construct selectively blocked the production of mutant p21, allowing investigators to study the contribution of the mutated p21 protein to the malignant phenotype. A recombinant plasmid clone was constructed using a wild-type, 2-kb K-ras genomic DNA segment carrying second and third exons with flanking intron sequences subcloned into an Apr-I-neo expression vector in AS orientation. The intron sequence used has a low degree of homology with other *ras* genomic sequences, leading to specific inhibition of K-*ras* with preservation of H-*ras* and N-*ras* expression. Previous studies on uptake of *ras* AS oligonucleotides by cancer cells resulted in cell death instead of regulated growth. This is probably because functional p21 protein is necessary for cell viability, and the oligonucleotides unselectively blocked *p21* expression produced by all three members of the *ras* gene family. Unselective blockade of oncogene expression can therefore be toxic to both normal and cancer cells. The 2 kb DNA insert was stably integrated into H460a cells as shown by Southern hybridization. Northern blot analysis detected expression of AS RNA. Western blot analysis showed 95% reduction in K-ras p21 protein synthesis in the clones expressing the AS RNA, while H460a cells and sense K-*ras* clones showed unchanged levels of K-*ras* p21 protein. Total p21 detected with a pan-ras monoclonal antibody showed only a slight decrease in the AS clones, suggesting that other ras genes were not affected. AS transfectants showed a threefold reduction in growth compared to sense transfectants and parental H460a cells but continued to grow in culture. Expression of AS K-*ras* RNA reduced the growth rate of H460a tumors in nu/nu mice. Tumorigenicity of cell lines expressing AS RNA was assessed by subcutaneous injection of 10^5 cells in nu/nu mice. Unmodified H460a cells formed tumors in all mice in 15 days. No tumor developed in mice injected with H460a AS cells during 120 days of observation, while H460a cells transfected with *Apr-I-neo S* plasmid formed tumors similar to H460a cells. These experiments show that in H460a cells engineered to synthesize AS K-*ras* RNA, the levels of K-ras mRNA and K-ras p21 protein are dramatically reduced. Thus, a recombinant construct can be made that distinguishes among members of the ras family. AS RNA generated from the genomic DNA

of the K-*ras* gene can thus specifically inhibit K-ras expression. Inhibition of K-ras reduced the growth rate of H460a cells but did not alter cell viability or continued growth in culture. This suggests that redundancy in p21 expression may compensate for absence of expression by one member of this family so that functions essential for maintenance of cell viability are preserved. This technique provides an opportunity to determine the effects of selective inhibition of oncogenic protein expression on the malignant phenotype. This study and subsequent confirmatory ones showed that reversal of a single genetic abnormality was sufficient to prevent human cancer cells from forming tumors. Thus, reversing all the genetic lesions in a cancer cell is not necessary for a therapeutic effect. These observations raise the intriguing possibility of specific gene therapy for cancer. Sequences could be delivered to tumor cells via viral vectors that specifically inhibit expression of the oncogenes activated in the cancer cell. Such constructs would be relatively nontoxic because, as in the preceding example, they could target a single gene whose function might be subsumed by other redundant genes of the same family. Thus, repeated infusions of immunologically distinct vectors could be performed. Retroviral vectors have the added advantage of being incorporated in the genome of only cells that replicate, thus favoring integration in cancer cell DNA.

Sporadic alterations in a variety of other oncogenes have been described. Amplification of c-*erb B-1*, c-*myc*, and c-*myb* as well as deletions in c-H-*ras* and c-*myb* have been reported. Expression of multiple protooncogenes in SCLC was determined in one study using Northern analysis. Increased expression of *myc* family genes was confirmed. ras family and c-*raf I* were expressed in all cell lines. Other oncogenes including c-*fes*, c-*fos*, c-*erb B-1*, c-*mos*, c-*sis*, c-*erb A*, c-src, and c-*abl* were expressed weakly or not expressed at all. Components of telomerase including the RNA (hTR) and enzyme activity (hTERT) are expressed at high levels in lung cancers and may provide a mechanism for cellular immortality.

3 Growth Factors and Autocrine–Paracrine Growth Control

Tumor cells that produce a growth factor and express its receptor may show self-stimulatory or autocrine growth. Cells that are regulated by an autocrine loop have several features. They secrete a biologically active growth factor and demonstrate increased proliferation to that factor. Antibodies that bind specifically to the growth factor will inhibit cell growth. Growth factors may act to stimulate growth in adjacent cells in a paracrine manner. Interaction of ligand and receptor in the cytoplasm of the cell may result in the formation of an internal autocrine loop.

Autocrine growth factors have been implicated in the stimulation of SCLC cell growth. The majority of SCLC cell lines produce bombesin. This 14–amino acid peptide is identical to a carboxyterminal heptapeptide sequence of a mammalian analogue, gastrin-releasing peptide. SCLC cell lines express a single class of high-affinity, saturable binding receptors for bombesin. Bombesin is also a potent stimulator of clonal growth for human NSCLC. Bombesin receptors have not been demonstrated on NSCLC. Oncogene products related to growth factor receptors of the tyrosine kinase family are implicated in retrovirally induced neoplasia. For

example, *v-erb B*, the transforming gene of avian erythroblastosis virus, codes for a truncated version of the epidermal growth factor receptor (EGFR). Activation of the overexpressed normal receptor gene appears to be sufficient for transformation of NIH 3T3 cells. NSCLC cells express high levels of functional EGFR. This is an uncommon finding for SCLC. NSCLC cells also show amplification of the *EGFR* gene. This suggests that growth factors and their receptors may play an important role in the development and/or maintenance of the malignant phenotype. In contrast to SCLC, autocrine growth control mechanisms have not been clearly defined for NSCLC. Studies have suggested that insulin-like growth factor I (IGF-I) may participate in autocrine growth stimulation.

3.1
c-*erb B-2*

c-*erb B-2* is a member of the EGFR family, which comprises three genes with a receptor-like structure: an external Iigand binding domain, a transmembrane domain, and a cytoplasmic tyrosine kinase domain. The *neu*/c-*erb B-2* oncogene was first identified in an ethylnitrosourea-induced rat neuroblastoma. The rat oncogene is activated by a point mutation, but the mechanism in human cells appears to be overexpression. The gene has homology to the *EGFR* gene, and the gene product (185 kDa) is a tyrosine kinase. The structure of this protein is receptor-like and several ligands have been identified. The *erb 13-2* gene is activated in NSCLC. In one study, amplification of the c-*erb B-2* occurred in only 10% of 60 paired samples of NSCLC and normal lung. However, adenocarcinomas showed high mRNA levels of c-*erb B-2*. In contrast, *SCLC* cells did not express c-*erb B-2*. A study of freshly excised surgical specimens confirmed and extended these findings. NSCLC specimens showed high levels of c-*erb B-2* RNA expression in 6 of 16 samples compared to paired normal lung samples. Increased expression occurred in both early and advanced stages. Fresh SCLC showed minimal or no expression. In a study using immunohistochemical techniques, expression of the c-*erb B-2* gene product, p185, occurred at higher levels in the tumor than in bronchiolar epithelium. The investigators found that c-*erb B-2* expression in adenocarcinomas is independently correlated with diminished survival.

3.2
TGF-α

Transforming growth factor type alpha (TGF-α) is a single polypeptide of 50 amino acids that is derived from a 160–amino acid transmembrane precursor by proteolytic cleavage. TGF-α is structurally and functionally related to the epidermal growth factor (EGF) and it binds to the EGFR. TGF-α alone does not transform normal rat kidney (NRK) cells. Anchorage-independent growth is seen when TGF-α and TGF-β are added together. TGF-α is a candidate autocrine growth factor for NSCLC. Earlier studies indicated that medium conditioned from A549 lung carcinoma cells can promote the growth of human NSCLC cells in culture. This conditioned medium is known to contain TGF-α, and exogenous TGF-α added to cultures increased colony formation.

It appears that it is not necessary for TGF-α to be cleaved from its conserved integral membrane glycoprotein to have biologic activity. When baby hamster kidney (BHK) cells were transfected with an expression vector containing altered sequences, the cells did not secrete TGF-α but expressed the pro-TGF-α on the cell

surface. The membrane-bound pro-TGF-α bound to EGFR on A431 cells and induced receptor autophosphorylation. Solubilized pro-TGF-α induced tyrosine autophosphorylation of EGFR in intact receptor-expressing cells and stimulated anchorage-independent growth of NRK fibroblasts. Thus, both pro-TGF-α and TGF-α could function as autocrine growth factors. Possibilities include interaction of pro-TGF-α with a cytoplasmic form of the EGFR as well as with EGFR expressed on adjacent cells.

The presence of EGFR on lung cancer cells suggests that an autocrine/paracrine growth mechanism may be operative. This possibility was investigated *in vitro* with cloned NSCLC cell lines. None of the cell lines expressed EGF by Northern analysis. All cell lines expressed TGF-α mRNA. TGF-α binds to the EGFR and therefore the investigators studied the biologic response to and production of TGF-α by these cell lines. Each cell line expressed EGFR by [I^{125}]EGF competitive binding and Scatchard analysis and by phosphorylation. The receptors were functionally active as determined in immune-complex kinase assays. The cloned NSCLC cell lines H322a, H226b, H596b, and H460a showed stimulated uptake of tritiated thymidine in response to both TGF-α and EGF. Exogenously added TGF-α increased colony formation in soft agar for three of the four cell lines in media containing serum. All cell lines expressed some TGF-α mRNA, although to differing degrees. Cell lysates and spent media competed for EGFR binding with EGF, thus showing production of TGF-α-like activity.

There was heterogeneity in the mechanism of autocrine growth stimulation among the four cell lines. H322a and H226b cells were specifically growth-inhibited by the anti-TGF-α monoclonal antibody AB-3 at low cell density, suggesting that the antibody blocks an autocrine growth loop. This inhibition was dose-dependent and did not occur following addition of the isotope identical to monoclonal antibody MOPC21. However, the AB-3 antibody did not alter the growth of H460a and H596b cells even though these cells express EGFR and secrete TGF-α. Suramin, which blocks binding of ligand to receptor in other autocrine systems, inhibited the growth of both cell lines. Addition of TGF-α specifically reversed inhibition by suramin. Addition of platelet-derived growth factor did not reverse suramin inhibition. This suggests that autocrine activation for this cell line occurred in the intracellular compartment between unprocessed receptor and unsecreted ligand. TGF-α appears to be an important autocrine growth factor for NSCLC cells of both squamous and adenocarcinoma histologies.

4
Tumor Suppressor Genes

The inactivation of certain genes may contribute to tumor growth. In one scenario, both copies of the gene must be eliminated or inactivated to eradicate the growth-suppressive function of the gene. Because both copies must be eliminated, the tumor suppressor gene is called *recessive*. The retinoblastoma *(Rb)* gene supports this model. Patients with familial predisposition have a germ line inactivation of one copy of the *Rb* gene. The tumor develops when the wild-type allele is either inactivated or deleted. This model has spawned multiple studies searching for consistent chromosomal deletions in human tumors.

One of the earliest and most consistent chromosomal deletions in lung cancer

occurs in chromosome 3 (p14–p23) for SCLC. Cytogenetic studies of fresh tumors confirmed observations on cell lines. Allelic loss in this region was documented with polymorphic DNA probes and occurred at a high frequency. Loss of heterozygosity for alleles on chromosomes 3, 11, 13, and 17 occurs in NSCLC as well. The high frequency of chromosomal deletions for both SCLC and NSCLC has suggested that loss of specific gene function may be a critical step in the development of lung cancer. Two candidate suppressor genes are the nuclear oncogenes *p53* and *Rb*.

Loss of heterozygosity on chromosome 13q suggests that the *Rb* locus, located at I3q14, may be deleted. Studies show that up to 60% of SCLC and 75% of carcinoid cell lines do not express *Rb* mRNA. However, 90% of NSCLC cell lines expressed *Rb*.

The *p53* gene encodes a 375–amino acid phosphoprotein that can form complexes with viral proteins such as large T-antigen and E 113. Missense mutations are common for the *p53* gene and are essential for the transforming ability of the oncogene. The mechanism of p53 transformation is controversial. The *p53* gene appears multifunctional, with major domains that can transactivate, bind proteins, bind sequence-specific DNA, and oligomerize with p53. Abnormalities in one or more of these functions could contribute to abrogation of the tumor suppressor function of *p53*. Certain mutations also have a dominant transforming capability. The wild-type *p53* gene may suppress genes that contribute to uncontrolled cell growth and proliferation or activate genes that suppress uncontrolled cell growth. Thus, absence or inactivation of wild-type p53 may contribute to transformation. However, some studies indicate that the presence of the mutant p53 may be necessary for full expression of the transforming potential of the gene.

Mutations of *p53* are common in a wide spectrum of tumors. These mutations occur in both NSCLC and SCLC cell lines and fresh tumors. The precise role of these mutations in oncogenesis and the mechanisms involved are subjects of active investigation. Mutations in *p53* positively correlate with lifetime cigarette consumption. Radon exposure, which increases lung cancer risk, is also associated with *p53* mutations, although the mutation types differ from those seen in tobacco-associated lung cancer. Insertion of a wild-type *p53* in lung cancer cell lines with a mutant or deleted *p53* can suppress their growth, even though the cells have multiple genetic lesions. The p53 protein may be overexpressed in lung cancer cells, although this effect is not always associated with the presence of a mutant *p53* gene. However, overexpression of the p53 protein has correlated positively with a poor prognosis. Deletions in 171, *p53* mutations, and 3p deletions have been detected in preneoplastic lesions of the lung such as severe dysplasia. Mutations in *p53* and *ras* appear to be independent events.

A common theme with many of the genetic lesions occurring in cancers is their relationship to regulation of progression through the cell cycle. Premature entry of cells into the next phase of the cell cycle may prevent completion of important molecular events such as DNA repair and may lead to genetic instability. The cell cycle may pause at both the G1 and G checkpoints. Phosphorylation of the Rb protein will release the transcription factor *E2F* and mediate transition across the G_1 checkpoint. Inactivation of *Rb* by mutation or deletion could therefore cause a state of activation. A gene encoding an inhibitor (called variously p21, WAF-1, Cip1, or

Sdi1) of the cyclins – the proteins that mediate cell cycle progression – is one of the downstream effectors of *p53* and mediates G1 arrest. Another inhibitor of the cyclin system is *p16*. This gene is frequently inactivated in many cancers including lung cancer.

Tumor suppressor genes that less frequently inactivate in lung cancers include *PTEN, hOGG1* (DNA repair), and *BAP1* (ubiquitination). The primary mechanism of inactivation of expression of the majority of tumor suppressor genes identified to date is promoter hypermethylation. Frequent inactivation of expression by promoter methylation has been observed for *APC, CDH13, RARβ, FHIT, RASSF1A, TIMP3, p16, MGMT*, and *DAPK*. Methylated DNA sequences can also be detected in the bronchial epithelium from smokers without lung cancer.

5 Premalignancy and Second Primary Cancers

Many of these molecular alterations have been identified in lung epithelium of current and former smokers without lung cancer. Such changes are almost never found in true lifetime nonsmokers (less than 100 cigarettes smoked in a lifetime). Many of these molecular lesions are clonal. Such clonal changes can persist for decades after smoking cessation, suggesting a proliferation in cell population. 3p allele loss at several 3p sites appears to be the earliest change followed by 9p (p16 locus), which can be identified in histologically appearing normal epithelium in smokers.

Patients who have had a primary epithelial cancer of the upper aerodigestive tract (head and neck, esophagus, and lungs) have a higher risk of developing a simultaneous or subsequent second primary cancer. It is hypothesized that the epithelial surface of the upper aerodigestive tract may share a common carcinogen exposure and increased risk of cancer development. It is thought that epithelial cancers can arise independently as separate primary cancers following prolonged carcinogen exposure. This effect is called *field cancerization*. Such field effects have also been proposed for breast and colorectal cancers. This hypothesis is testable in that it predicts that if a common carcinogen is involved, second primary cancers will arise from similar but independent events.

To be informative, a molecular marker for determining the independent origin of cancers must have several characteristics. The marker should be associated with the development of the cancer. Alterations in the marker such as mutations should occur early in the development of the cancer. Finally, these mutations should be clonally preserved. The *p53* gene fulfills these criteria for the three most common primary and second primary cancers of the aerodigestive tract: head and neck, lung, and esophagus. Many studies have shown that inactivation of the *p53* gene by mutation or deletion results in cell transformation. Mutations in the *p53* gene occur early in the development of upper aerodigestive tract cancers and are detected frequently in fresh tumor specimens. Mutations in the *p53* gene occur in premalignant lesions from aerodigestive tract cancers. Mutations of the *p53* gene have been frequently detected in early-stage cancers of the head and neck, as well as the lung and esophagus, and the incidence of these mutations does not increase with advancing stage of disease, as would be expected if *p53* mutations were associated with cancer progression. Mutations of the *p53* gene show clonal fidelity

for recurrent cancers and metastases. Differences in *p53* mutations between the primary cancer and second primary cancers would provide evidence of independent origin for these multifocal tumors.

Thirty-one patients with primary cancers of the head and neck and associated second primary cancers were studied by single-strand conformation polymorphism analysis and DNA sequencing. The overall frequencies of *p53* mutations among primary tumors and second primary cancers were 42% (13/31) and 37%(13/35) respectively. Mutations were found in 19 of 52 head and neck squamous cell carcinomas (36%), in 4 of 7 squamous cell carcinomas of the lung (57%), in 2 of 5 adenocarcinomas of the lung, and in 1 of 2 squamous cell carcinomas of the esophagus. Twelve samples had p53 mutations on exon 5, nine on exon 7, and one on exon 8. Six samples had mutations in more than one exon. Twenty-one of 31 patients (68%) had *p53* mutations in one or more specimens. In all 21 cases, the genetic lesions were discordant, such that the presence or location of the mutations differed from those of the second and third primary cancers. In the five patients with *p53* mutations in the initial primary cancer and second primary cancer, the mutations occurred in different regions of the *p53* gene. In the other 16 patients, *a p53* mutation was found in one primary tumor but not in the other. In eight of these cases, a mutation was found in the first primary but not the subsequent primary cancer, and in the other eight cases, a *p53* mutation was not detected in the initial primary cancer but was detected in subsequent primary cancers.

The discordant *p53* mutations in second primary cancers arising in patients with primary epithelial cancer of the upper aerodigestive tract suggest that these cancers arise as independent events. These observations provide the first indication of a molecular basis for field cancerization effects in cancers of the upper aerodigestive tract.

6 Conclusions and Future Prospects

Elimination of tobacco consumption is the most effective way to prevent lung cancer. However, this goal remains elusive. Even if it is achieved, lung cancer will develop well into the twenty-first century in a large number of persons with long exposure histories. Thus, early detection and new therapeutic options are needed to improve the dismal survival for lung cancer. Progress in defining the molecular events involved in the genesis of lung cancer cells has been rapid over the last five years. Potential benefits for patients may be substantial. Identification of persons at high risk for development of lung cancer based on molecular profiles may enable cost-effective implementation of prevention and early detection programs. Conventional therapy has emphasized killing of the last cancer cell by surgical excision or destruction by radiation or chemotherapy. Alteration of oncogene expression by recombinant gene constructs may result in reversal of the transformed phenotype.

Successful therapy and prevention interventions that reverse genetic lesions may be possible. Genetic constructs could specifically inhibit expression of mutant proteins by dominant oncogenes and could replace the function of deleted or mutated tumor suppressor genes if they could be delivered with high efficiency to tumor cells *in vivo*. Viral vectors have the potential for this. The aerodigestive tract is suited to this approach because high concentrations of these relatively nontoxic agents could

be achieved with local installation, thus avoiding the dilutive effects of intravenous injection. If these agents are efficacious, their lack of toxicity may provide a sufficiently high therapeutic index, permitting their use as an adjuvant to surgery to treat patients with earlier stages of cancer, or as preventive therapy for individuals with documented genetic abnormalities in premalignant lesions.

See also Genetics, Molecular Basis of; Medicinal Chemistry.

Bibliography

Cross, M., Dexter, T.M. (1991) Growth factors in development, transformation, and tumorigenesis, *Cell* **64**, 271–280.

Hanahan, D., Weinberg, R.A. (2000) The hallmarks of cancer, *Cell* **100**, 57–70.

Hong, W.K., Mulshine, J.L. (Eds.) (1992) *Biology of and Novel Therapeutic Approaches for Epithelial Cancers of the Aerodigestive Tract*, Monograph No. 13, Journal of the National Cancer Institute, Bethesda, MD.

Malkinson, A.M. (1989) The genetic basis of susceptibility to lung tumors in mice, *Toxicology* **54**, 241–271.

Minna, J.D., Roth, J.A., Gazdar, A.F. (2002) Focus on lung cancer, *Cancer Cell* **1**, 49–52.

Proctor, R.N. (2001) Tobacco and the global lung cancer epidemic, *Nat. Rev. Cancer* **1**, 82–86.

Roth, J.A. (1992) New approaches to treating early lung cancer, *Cancer Res.* **52**, S2652–S2657.

Primary Literature

Aisner, S.C., Finkelstein, D.M., Ettinger, D.S., et al. (1990) The clinical significance of variant-morphology small-cell carcinoma of the lung, *J. Clin. Oncol.* **8**, 402–408.

Bradley, E.C., Schecter, G.P., Matthews, M.J., et al. (1982) Erythroleukemia and other hemologic complications of intensive therapy in long-term survivors of small cell lung cancer, *Cancer* **49**, 221–223.

Boice, J.D. Jr., Fraumeni, J.F. Jr. (1985) Second cancer following cancer of the respiratory system in Connecticut, 1935–1982, *NCI Monograms* **68**, 83–98.

Cajot, J.F., Anderson, M.J., Lehman, T.A., et al. (1992) Growth suppression mediated by transfection of p53 in Hut292DM human lung cancer cells expressing endogenous wild-type p53 protein, *Cancer Res.* **52**, 6956–6960.

Christensen, P.H., Joergensen, K., Munk, J., et al. (1987) Hyperfrequency of pulmonary cancer in a population of 415 patients treated for laryngeal cancer, *Laryngoscope* **97**, 612–614.

Chung, K.Y., Mukhopadhyay, T., Kim, J., et al. (1993) Discordant *p53* gene mutations in primary head and neck cancers and corresponding second primary cancers of the upper aerodigestive tract, *Cancer Res.* **53**, 1676–1683.

de Vries, N., Snow, G.B. (1986) Multiple primary tumors in laryngeal carcinoma. *J. Laryngol. Otol.* **100**, 915–918.

Fujiwara, T., Grimm, E.A., Mukhopadhyay, T., et al. (1994) Induction of chemosensitivity in human lung cancer cells *in vivo* by adenoviral-mediated transfer of the wild-type p53 gene, *Cancer Res.* **54**, 2287–2291.

Funa, K., Steinholtz, L., Nou, E., et al. (1987) Increased expression of N-*myc* in human small cell lung cancer biopsies predicts lack of response to chemotherapy and poor prognosis, *Am. J. Clin. Pathol.* **88**, 216–220.

Georges, R.N., Mukhopadhyay, T., Zhang, Y., et al. (1993) Prevention of orthotopic human lung cancer growth by intratracheal instillation of a retroviral antisense K-ras construct, *Cancer Res.* **53**, 1743–1746.

Hollstein, M., Sidransky, D., Vogelstein, B., et al. (1991) *p53* mutations in human cancers, *Science* **253**, 49–53.

Horio, Y., Takahashi, T., Kuroishi, T., et al. (1993) Prognostic significance of p53 mutations and 3p deletions in primary resected non-small cell lung cancer, *Cancer Res.* **53**, 1–4.

Ji, L., Fang, B., Yen, N., et al. (1999) Induction of apoptosis and inhibition of tumorigenicity and tumor growth by adenovirus vector-mediated fragile histidine triad (FHIT) gene overexpression, *Cancer Res.* **59**, 3333–3339.

Ji, L., Nishizaki, M., Gao, B., et al. (2002) Expression of several genes in the human chromosome 3p21.3 homozygous deletion region by an adenovirus vector results in tumor

suppressor activities *in vitro* and *in vivo*, *Cancer Res.* **62**, 2715–2720.

Little, C.D., Nau, M.M., Carney, D.N., et al. (1983) Amplification and expression of the *c-myc* oncogene in human lung cancer cell lines, *Nature* **306**, 194–196.

Mao, L., Lee, J.S., Kurie, J.M., et al. (1997) Clonal genetic alterations in the lungs of current and former smokers [see comments], *J. Natl. Cancer Inst.* **89**, 857–862.

Martini, N., Bains, M.S., Burt, M.E., et al. (1985) Incidence of local recurrence in second primary tumors in resected stage I lung cancer, *J. Thorac. Cardiovasc. Surg.* **89**, 836–841.

Mathisen, D.J., Jensik, R.J., Feber, L.P., et al. (1984) Survival following resection for second and third primary lung cancers, *J. Thorac. Cardiovasc. Surg.* **88**, 502–510.

Mukhopadhyay, T., Roth, J.A. (1993) A codon 248 p53 mutation retains tumor suppressor function as shown by enhancement of tumor growth by antisense *p53*, *Cancer Res.* **53**, 4362–4366.

Mukhopadhyay, T., Tainsky, M., Cavender, A.C., et al. (1991) Specific inhibition of K-ras expression and tumorigenicity of lung cancer cells by antisense RNA, *Cancer Res.* **51**, 1744–1748.

Naylor, S.L., Johnson, B.E., Minna, J.D., et al. (1987) Loss of heterozygosity of chromosome 3p markers in small-cell lung cancer, *Nature* **329**, 451–454.

Quinlan, D.C., Davidson, A.G., Summers, C.L., et al. (1992) Accumulation of p53 protein correlates with a poor prognosis in human lung cancer, *Cancer Res.* **52**, 4828–4831.

Rodenuhuis, S., Slebos, R.J.C., Evers, S.G., et al. (1991) The ras oncogene family and lung cancer, *J. Cell Biochem.* Suppl. **15F**, 3.

Rosengart, T.K., Martini, N., Ghosn, P., et al. (1991) Multiple primary lung carcinomas: prognosis and treatment, *Ann. Thorac. Surg.* **52**, 773–779.

Roth, J.A., Nguyen, D., Lawrence, D.D., et al. (1996) Retrovirus-mediated wild-type p53 gene transfer to tumors of patients with lung cancer, *Nat. Med.* **2**, 985–991.

Schneider, P.M., Hung, M.C., Chiocca, S.M., et al. (1989) Differential expression of the c-*erb*B-2 gene in human small cell and non-small cell lung cancer, *Cancer Res.* **49**, 4968–4971.

Sellers, T.A., Bailey-Wilson J.E., Elston R.C., Wilson, A.F., Ooi, W.L., Elston, G.Z., Rothschild, H. (1990) Evidence for Mendelian inheritance on the pathogenesis of lung cancer. Increased familial risk for, *J. Natl. Cancer Inst.* **82**, 1272–1279.

Sozzi, G., Pastorino, U., Moiraghi, L., et al. (1998) Loss of FHIT function in lung cancer and preinvasive bronchial lesions, *Cancer Res.* **58**, 5032–5037.

Sozzi, G., Tornielli, S., Tagliabue, E., et al. (1997) Absence of Fhit protein in primary lung tumors and cell lines with FHIT gene abnormalities, *Cancer Res.* **57**, 5207–5212.

Sozzi, G., Veronese, M.L., Negrini, M., et al. (1997) The FHIT gene 3p14.2 is abnormal in lung cancer, *Cell* **85**, 17–26.

Suzuki, H., Takahashi, T., Kuroishi, T., et al. (1992) p53 mutations of non-small cell lung cancer in Japan: association between mutations and smoking, *Cancer Res.* **52**, 734–736.

Swisher, S.G., Roth J.A., Nemunaitis, J., et al. (1999) Adenovirus-mediated p53 gene transfer in advanced non-small cell lung cancer, *J. Natl. Cancer Inst.* **91**, 763–771.

Thomas, P.A., Holmes, E.C., Piantadosi, S., et al. (1993) Malignant disease appearing late after operation for T1 N0 non- small-cell lung cancer, *J. Thorac. Cardiovasc. Surg.* **106**, 1053–1058.

Thomas, P., Rubinstein, L., Lung Cancer Study Group (1990) Cancer recurrence after resection: T1 N0 non-small cell lung cancer, *Ann. Thorac. Surg.* **49**, 242–247.

Tomizawa, Y., Sekido, Y., Kondo, M., et al. (2001) Inhibition of lung cancer cell growth and induction of apoptosis after reexpression of 3p21.3 candidate tumor suppressor gene SEMA3B, *Proc. Natl. Acad. Sci. U.S.A* **98**, 13954–13959.

Vahakangas, K.H., Samet, J.M., Metcalf, R.A., et al. (1992) Mutations of p53 and ras genes in radon-associated lung cancer from uranium miners, *Lancet* **339**, 576–580.

Weinberg, R.A. (1991) Tumor suppressor genes, *Science* **254**, 1138–1146.

Weiner, D.B., Norderg, J., Robinson, R., et al. (1990) Expression of the *neu* gene-encoded protein (P185neu) in human non-small cell carcinomas of the lung, *Cancer Res.* **50**, 421–425.

Wistuba I., Behrens, C., Milchgrub, S., et al. (1999) Sequential molecular abnormalities are

involved in the multistage development of squamous cell lung carcinoma, *Oncogene* **18**, 643–650.

Yokota, J., Wada, M., Yoshida, T., et al. (1988) Heterogeneity of lung cancer cells with respect to the amplification and rearrangement of myc family oncogenes, *Oncogene* **2**, 607–611.

Zhang, W.W., Fang, X., Mazur, W., et al. (1994) High-efficiency gene transfer and high-level expression of wild-type p53 in human lung cancer cells mediated by recombinant adenovirus, *Cancer Gene Ther.* **1**, 5–13.

Zochbauer-Muller, S., Fong, K.M., Virmani, A.K., et al. (2001) Aberrant promoter methylation of multiple genes in non-small cell lung cancers, *Cancer Res.* **61**, 249–255.

Genetics, Molecular Basis of

D. Peter Snustad
University of Minnesota, Minnesota, USA

1 Genetic Information 374
1.1 Four-letter Alphabet 374
1.2 The Gene, the Basic Unit of Function 374
1.3 Genes are Located on Chromosomes 375
1.4 Autosomal and Sex-linked Inheritance 377
1.5 Mendelian versus Quantitative Traits 380
1.6 Genomics: Analyses of Entire Genomes 380

2 DNA Replication: The Genotypic Function 381
2.1 Semiconservative Replication 381
2.2 Proofreading 381
2.3 The Complex Replisome 382

3 Gene Expression: The Phenotypic Function 385
3.1 Colinearity between Gene and Polypeptide 386
3.2 Transcription 386
3.3 Introns and RNA Splicing 387
3.4 RNA Editing 388
3.5 Translation 389
3.6 Inteins 391
3.7 Some Complex Gene/Protein Relationships 391
3.8 Pathways of Gene Expression are Often Complex 392
3.9 Pleiotropy and Epistasis 394
3.10 Penetrance and Expressivity 394
3.11 Intragenic Complementation 395
3.12 Intergenic Noncomplementation 395
3.13 Regulation of Gene Expression 395
3.14 RNA-mediated Gene Silencing 397

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

4 Mutation: The Evolutionary Function 398
4.1 Dominance, Partial Dominance, and Codominance 398
4.2 Transitions, Transversions, and Frameshift Mutations 399
4.3 Gain-of-function and Loss-of-function Mutations 400
4.4 Suppressor and Enhancer Mutations 400
4.5 Chromosome Aberrations 401
4.6 Polyploidy 401
4.7 Nondisjunction and Aneuploidy 401
4.8 Expanding Trinucleotide Repeats in Humans 402

5 Recombination: New Combinations of Genes to be Acted on by Natural or Artificial Selection 402
5.1 Segregation 402
5.2 Independent Assortment 403
5.3 Linkage, Crossing-over, and Genetic Maps of Chromosomes 404
5.4 Transposable Genetic Elements 405
5.5 Physical Maps of Chromosomes 406
5.6 Map-position-based Identification of Genes 407
Acknowledgements 409

Bibliography 409
Books and Reviews 409
Primary Literature 409

Keywords

Aneuploidy
The presence of an extra or missing chromosome (or chromosomes) in a cell or an organism. Any condition in which the chromosomes are not all present in equal number.

Attenuation
A mechanism by which gene expression in prokaryotes is regulated by the premature termination of transcription.

Autosomal Inheritance
The pattern of inheritance observed for genes located on chromosomes other than the sex chromosomes (chromosomes carrying sex-determining genes).

CentiMorgan (cM)
The measure of the distance between markers on genetic maps based on the average number of crossover events that occur during meiosis. A distance of 1 cM indicates

that 1 recombinant chromosome is present among every 100 chromosomes produced during meiosis.

Codon
The unit of three contiguous nucleotides in mRNA specifying the incorporation of one amino acid in the polypeptide produced by translating that mRNA on the polyribosomes.

Complementation Test
The introduction of two recessive mutations into the same cell but on different chromosomes (a *trans* heterozygote) to determine whether the mutations are both in the same gene or are in two different genes. If the mutations are both in the same gene, the $m1+/+m2$ heterozygote will exhibit a mutant phenotype, whereas if they are in two different genes, the *trans* heterozygote will exhibit the wild-type phenotype.

Crossing-over
A recombination process through which new combinations of genes on specific chromosomes are produced by the breakage and reunion of their DNA molecules.

Dominance
The control of the phenotype of an organism by one allele of a gene correlated with the exclusion of any effect of the other allele.

Epistasis
The interaction of nonalleles. Any condition in which an allele of one gene masks the expression of one or more alleles of a different gene (a nonallele).

Exon
A nucleotide sequence of a gene that corresponds to a sequence that is present in the final processed RNA product of the gene.

Frameshift Mutation
A mutation that alters the codon reading frame of a gene, by either inserting or deleting one or more nucleotide pairs in multiples other than three.

Gain-of-function Mutation
A mutation that produces a gene product with a new function.

Genetic Map
A diagram of a chromosome with distances between markers based on recombination frequencies (centiMorgans).

Heritability
The proportion of the total phenotypic variability present for a trait that results from genetic factors rather than environmental effects.

Independent Assortment
During meiosis, each pair of homologous chromosomes lines up at the metaphase plate independently of every other pair. As a result, the alleles of a gene on one chromosome segregate independently of the alleles of a gene on any nonhomologous chromosome.

Intron
A sequence of nucleotide pairs in a gene that is not present in the mature RNA because it is excised from the primary transcript during processing.

Loss-of-function Mutation
A mutation that impairs or abolishes the expression of a gene or renders its product less active or nonfunctional.

Meiosis
The process by which the chromosome number in reproductive cells is reduced to half the number present in other (somatic) cells of the body. Chromosomes duplicate once and cells duplicate twice to produce haploid gametes.

Mitosis
The separation of the daughter chromatids produced by chromosome replication and the division of the cytoplasm to produce two identical progeny cells.

Mutation
A heritable change in the structure of the genetic material of an organism. When used in the broad sense, mutations include both ''point mutations,'' involving changes in the structure of individual genes, and gross changes in chromosome structure (chromosome aberrations). In the narrow sense, mutations include only ''point mutations.'' The term *mutation* is used to refer to (1) the *process* by which the change occurs and (2) the *result* of the process, the *alteration* in the gene or genetic material.

Nondisjunction
The failure of homologous chromosomes or sister chromatids to separate or disjoin from each other during meiosis or mitosis.

Polyploidy
The presence of three or more copies of each chromosome in a cell or an organism.

Recombination
The generation of new combinations of genes in progeny, which were not present together in either of the parents, either (1) by independent assortment of nonhomologous chromosomes during meiosis or (2) by crossing-over (breakage and exchange of parts) of homologous chromosomes during meiosis or mitosis.

Segregation
The separation of the maternal and paternal chromosomes, and thus the alleles of genes in heterozygotes, from each other during the reductional division of meiosis.

Sex-linked Inheritance
A pattern of inheritance that occurs when the gene controlling a trait is located on a chromosome carrying genes that determine the sex phenotype of the organism.

Suppressor Mutation
A mutation that partially or completely eliminates the phenotypic effect of another mutation.

Transition
A mutation resulting from the replacement of one purine with the other purine and/or one pyrimidine with the other pyrimidine.

Transposable Genetic Element
A DNA unit that can move from one location in a genome to another location or even to a different genome.

Transversion
A mutation resulting from the replacement of a purine with a pyrimidine and/or a pyrimidine with a purine.

The phenotype of a living organism is controlled by its genotype, the summation of its genetic information, acting within the constraints imposed by the environment in which the organism exists. Much of the genetic material of an organism is organized into basic functional units called genes, which specify RNA and/or protein products. Some genes encode one primary gene product, either an RNA molecule or polypeptide. Other genes produce two or more related polypeptides by RNA editing, differential transcript splicing, or the assembly of genes from gene segments during development. The genetic information of all living organisms, whether viruses, bacteria, corn plants, or humans, is stored in the sequence of bases (purines and pyrimidines) or base pairs in the deoxyribonucleic acid (DNA) present in their chromosomes. In some viruses, the genetic information is stored in the sequence of bases in ribonucleic acid (RNA). The genetic information is encoded using a four-letter alphabet: the four bases adenine (A), guanine (G), cytosine (C), and thymine (T). In RNA, uracil (U) replaces the thymine present in DNA. In the double-stranded DNA present in most cellular organisms, adenine and thymine form one base pair (A:T) and guanine and cytosine form a second base pair (G:C).

The genetic material of an organism must carry out three essential functions: (1) the genotypic function, transmission of the genetic information from generation to generation; (2) the phenotypic function, directing the growth and development of

the offspring into mature, reproductive adults; and (3) the evolutionary function, mutation, allowing organisms to evolve in response to changes in the environment. Mutation produces new genetic variability, which provides the raw material for evolution. Recombination of genetic material occurs by the independent assortment of nonhomologous chromosomes and by crossing-over between homologous chromosomes. This recombination provides new combinations of genes and thus new phenotypes on which natural selection acts during the process of evolution.

1 Genetic Information

The genetic information of living organisms is stored in large macromoleules called *nucleic acids*. These nucleic acids are of two types: *DNA* contains the sugar 2′-deoxyribose and *RNA* contains the sugar ribose. In all eukaryotic organisms, the genetic information is stored in giant DNA molecules located in one to many chromosomes, the number depending on the species. In some viruses that contain no DNA, the genetic information is stored in RNA.

1.1 Four-letter Alphabet

The genetic information is stored in nucleic acids using a four-letter alphabet: the four bases *adenine* (A), *guanine* (G), *cytosine* (C), and *thymine* (T) in DNA or *uracil* (U) in RNA. In DNA, which has a double-stranded structure in cellular organisms, the bases are present in pairs: A with T and G with C (Fig. 1). Although a four-letter alphabet may seem too simple to store enough information to produce the vast phenotypic variability observed in living organisms, recall that the Morse code is based on just two symbols – dots and dashes. Moreover, computers perform their amazing feats using a binary code composed of 0s and 1s. Even with just four letters, a vast amount of genetic information can be stored in the large nucleic acids present in living cells. Consider, for example, that one complete copy of the human genome (all the genetic information in one complete set of human chromosomes) contains three-billion (3×10^9) base pairs of DNA. Since the number of different sequences of 4 letters used n at a time is 4^n, one can see that the human genome has the capacity to store a huge amount of information with $n = 3 \times 10^9$.

1.2 The Gene, the Basic Unit of Function

The basic functional unit of genetic information is the *gene*, defined operationally by the complementation test and most commonly specifying the amino acid sequence of one polypeptide chain or the nucleotide sequence of one RNA molecule. Different forms of a given gene are called alleles. The wild-type alleles of a gene are those that exist at relatively high frequencies in natural populations and yield wild-type or "normal" phenotypes; they are usually symbolized by a + or a symbol with a + superscript (e.g. w^+ for the allele that yields wild-type red eyes in fruit flies). Alleles of a gene that result in abnormal or non-wild-type phenotypes are called

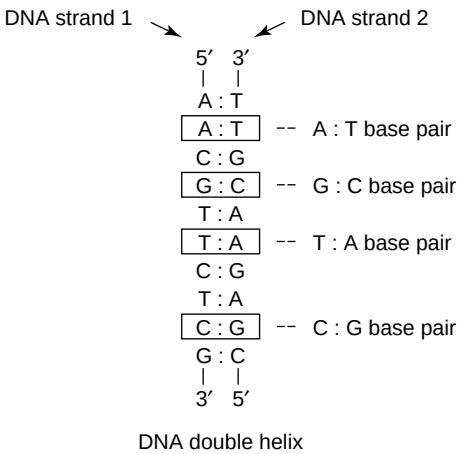

Fig. 1 Two-dimensional view of a segment of double-stranded DNA showing the four base pairs used to store genetic information.

mutant alleles. They are usually symbolized by one to three letters written in italics (e.g. w and w^{ap} for the alleles that cause white and apricot eye color respectively in fruit flies). Many eukaryotes such as corn plants, fruit flies, and humans contain two copies of their genome in most cells, two copies of each of their chromosomes; such eukaryotes are called diploids. Thus, diploid organisms may contain two different alleles of any given gene, in which case they are heterozygous (e.g. w/w^{+}, w^{ap}/w^{+}, w/w^{ap}) or two identical copies of a given gene, in which case they are homozygous (e.g. w^{+}/w^{+}, w/w, w^{ap}/w^{ap}). A w^{ap}/w^{+} heterozygous fruit fly has wild-type red eyes. The w^{+} allele is expressed in this heterozygous fly; w^{+} is thus called the dominant allele. The w^{ap} allele is not expressed in this heterozygous fly; it is said to be recessive because its effect on the phenotype is masked by the w^{+} allele.

The *complementation test* is performed by producing cells or organisms that contain two recessive mutant genes located on two different chromosomes, that is, *trans* heterozygotes, and by determining whether these cells or organisms have mutant or wild-type phenotypes. If the two mutant genes are allelic, that is, the defects or mutations are in the same gene, the *trans* heterozygote will have a mutant phenotype. If the two mutant genes are not allelic, that is, the mutations are in two different genes, the *trans* heterozygote will have the wild-type phenotype. The rationale behind the complementation test is illustrated in Fig. 2.

1.3 Genes are Located on Chromosomes

One of the important discoveries in biology was that the genetic information of organisms is present in structures called chromosomes, because the transmission of these organelles could be followed during cell division and reproduction. This research culminated in the "Chromosome Theory of Inheritance," which is the core of modern genetics. The DNA molecules that carry the genetic information are packaged into chromosomes with the aid of proteins and RNA molecules. In a eukaryotic cell, the chromosomes are present in a membrane-bounded compartment called the nucleus. The prokaryotic equivalent is the nucleoid, which is not surrounded

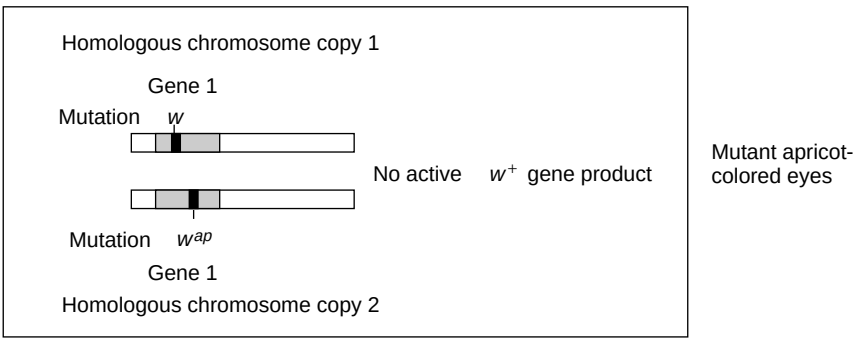

(a) Two mutations in the same gene.

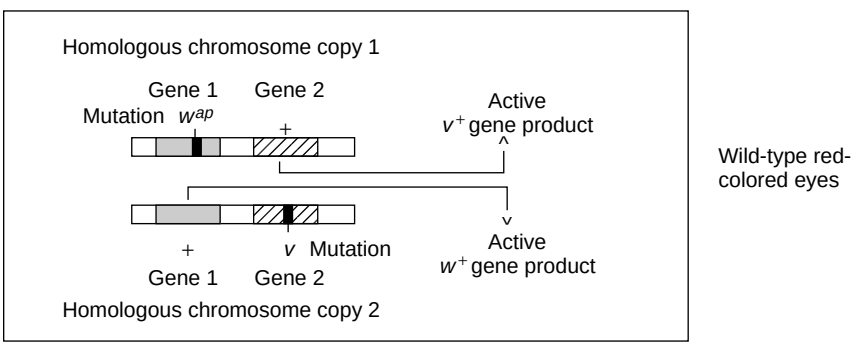

(b) Two mutations in two different genes.

Fig. 2 Illustration of the complementation test used to operationally define the gene, the basic unit of function of genetic material. The operation is to place the two recessive mutations in question in the same cell or cells of a multicellular organism on two separate chromosomes, that is, to construct a *trans* heterozygote, and to determine whether this cell or organism has a mutant or a wild-type phenotype. If the phenotype is mutant, the two mutations are in the same gene; this is illustrated for the w (white eyes) and w^{ap} (apricot eyes) mutations of *Drosophila* in (a). If the phenotype of the *trans* heterozygote is wild type, the two mutations are in two different genes and the two mutations are said to complement each other. Complementation between the w^{ap} and v (vermilion eye color) mutations of *Drosophila* is illustrated in (b); note that active (wild-type) products of both genes (w^+ and v^+) are present in the *trans* heterozygote shown in (b) – thus, the wild-type phenotype.

by a membrane. An important function of the chromosomes is to ensure the proper distribution of the genetic material to daughter cells during cell division.

Each chromosome contains a single large DNA molecule packaged in a matrix of RNA and protein. The DNA is highly compacted by coiling and supercoiling. If the DNA in a single human chromosome were uncoiled so that it was a perfectly linear molecule, it would measure between 2.5 and 8.5 cm in length. During cell division, this DNA molecule is present in a chromosome that is only about 0.5 μm in diameter and 3 to 10 μm in length. Eukaryotic chromosomes have several distinct structural features that are visible under the light microscope (Fig. 3). After replication, each chromosome is composed of

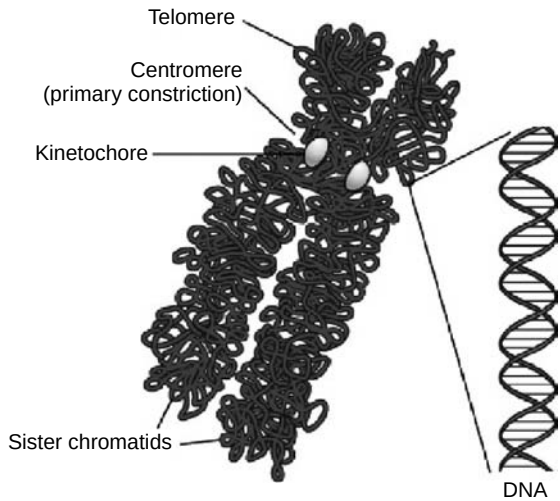

Fig. 3 The structure of a highly condensed replicated eukaryotic chromosome. (Reproduced from Snustad, D.P., Simmons, M.J., *Principles of Genetics*, 3rd edition, Copyright © 2003 by John Wiley & Sons, Inc., Hoboken, NJ, USA. This material is used by permission of John Wiley & Sons, Inc.)

two sister chromatids joined together at a constriction called the centromere The kinetochore, a protein structure present in each centromere, plays an important role in the separation of daughter chromosomes during cell division. The position of the centromere varies from the middle to the end of the chromosome. The ends of the chromosomes are called telomeres; they contain unique molecular structures that enhance chromosome stability.

1.4 Autosomal and Sex-linked Inheritance

In many eukaryotes, sex phenotype is controlled by key regulatory genes that are present on chromosomes that are not present as morphologically identical or nearly identical pairs of homologous chromosomes. In humans, for example, an important male-determining gene, *SRY* (for *s*ex-determining *r*egion of the *Y*), which encodes the testis-determining factor (TDF), is located on a small chromosome called the *Y chromosome*. This chromosome is normally present only in males (there are rare exceptions). During reproductive divisions, it pairs with a much larger chromosome called the *X chromosome*, which is present in two copies in females. The X and Y chromosomes share common terminal regions, allowing them to pair during meiosis. This chromosomal mechanism of sex determination is called the XX-XY mechanism. It occurs in humans and other mammals, the fruit fly *Drosophila melanogaster*, and a number of other species. In some species, the Y chromosome is completely absent and males have one less chromosome than females (the XX-XO mechanism of sex determination). The X and Y chromosomes are referred to as sex chromosomes, and all the other chromosomes (present in morphologically identical pairs) are called *autosomes*. Humans, for example, contain the 2 sex chromosomes and 44 autosomes (22 homologous pairs).

The human Y chromosome plays a major role in sex determination; however, it is small and contains very few genes that affect other traits. In contrast, the X chromosome is large and contains

a large number of important genes. The absence of an X chromosome is lethal in humans. Given that females contain two X chromosomes and males contain one X and one Y, whereas both females and males contain two copies of each autosome, traits controlled by genes located on sex chromosomes will exhibit different patterns of inheritance than traits controlled by genes on autosomes

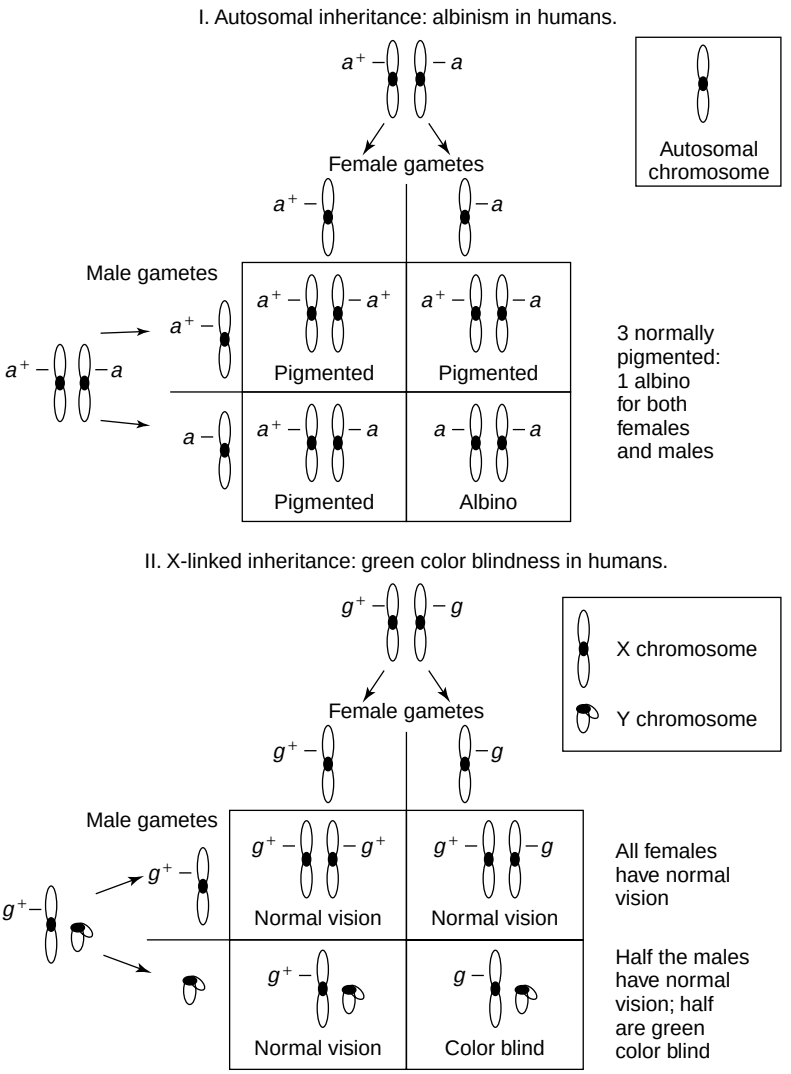

Fig. 4 Autosomal versus sex-linked (X-linked) inheritance. (a) Autosomal inheritance: the expected results are shown for a mating between two individuals who are both heterozygous for a recessive allele causing albinism. (b) X-linked inheritance: the predicted results are shown for a mating between a woman who is heterozygous for a recessive allele causing green color blindness and a male with normal vision.

(Fig. 4). In humans, albinism is caused by a recessive mutation in any of several autosomal genes, whereas green color blindness is caused by a recessive mutant allele of a gene on the X chromosome.

Note that the X chromosome of a male is always passed to his daughters, because the offspring would be male if it received a Y chromosome. Given that the X chromosome contains many essential genes, a son has to get an X chromosome from his mother. Because a male has only one X chromosome, recessive traits such as color blindness and hemophilia are expressed much more frequently in males. Males are *hemizygous* (''half that of a zygote'') for X-linked genes; if a recessive X-linked allele is present, it will be expressed. Females contain two X chromosomes; they will have to be homozygous for a recessive X-linked allele to express the trait. Thus, if an X-linked recessive allele is present in a population with a frequency of q, and if there is random mating in the population with respect to this trait, the recessive trait will occur with a frequency of q in males and a frequency of q^2 in females. To express the trait, females will have to obtain one X chromosome carrying the mutant allele from the mother via the egg and a second X chromosome carrying the mutant allele from the father through the sperm. These are independent events, and the probability that two independent events will both occur is equal to the product of the probabilities of the individual events ($q \times q = q^2$). Thus, it is easy to see why males express X-linked traits more frequently than females. In contrast, traits controlled by genes located on autosomes are usually expressed with equal frequency in the two sexes.

Females who receive a recessive X-linked mutant allele from their father often do not express it because they receive the wild-type allele from their mother. However, they will transmit the mutant allele to half of their hemizygous sons, who will express the trait. These heterozygous women are referred to as ''carriers'' of the trait; they carry the mutation, but do not show any effect of its presence.

Females have two copies of every gene on the X chromosome; males have only one. In humans and other mammals, this difference in gene dosage is corrected by inactivating one of the two X chromosomes in each cell in females. This inactivation occurs at random, so half of a woman's cells express her maternal X chromosome and the other half express her paternal X chromosome. Women are, therefore, mosaics for the expression of genes on their X chromosomes. The inactivation occurs early during development and results in a highly condensed chromosome that is visible under the light microscope as a densely staining ''Barr body.'' The inactivated X chromosome undergoes reactivation during oogenesis so that all eggs contain active X chromosomes.

Sex is determined by many different mechanisms in other species. In birds, males contain two identical sex chromosomes and females contain two distinct sex chromosomes. In ants, bees, wasps, and other Hymenoptera, males have only one copy of each chromosome, whereas females carry two copies. Males are ''haploid''; they develop from unfertilized eggs. Females are diploid, developing from fertilized eggs as in most other eukaryotes. In still other species, special sex chromosomes are not present, and environmental factors play key roles in sex determination. Thus, when considering sex-linked inheritance, one must not extrapolate from humans and fruit flies to other species.

1.5 Mendelian versus Quantitative Traits

Mendel studied pea plants that differed in specific phenotypic characteristics such as red flowers versus white flowers, tall versus dwarf, round seeds versus wrinkled seeds, and so on. As a result, he was able to classify the progeny of his crosses into distinct phenotypic classes and calculate the frequency of each class. Such traits are now referred to as *Mendelian traits*, because they yield the predicted monohybrid, dihybrid, and so on, segregation ratios in genetic crosses. Many other traits show continuous variation such that the progeny of crosses cannot be placed into discreet phenotypic classes. In humans, height and weight exhibit continuous variation from the shortest or smallest individual to the tallest or heaviest individual. The inheritance of such traits must be studied using quantitative measurements; therefore, the traits are commonly called *quantitative traits*. The genes that control quantitative traits are no different than the genes that control Mendelian traits. The difference is that quantitative traits are influenced by a large number of genes; they are multifactorial, being influenced by many factors, both genetic and environmental. Geneticists have developed statistical tools that can be used to estimate the number of genes affecting a trait, the proportion of the phenotypic variability that is genetic (caused by genes), and the proportion that is environmental (due to environmental factors). For a given trait, the proportion of the total phenotypic variability that is caused by genetic factors is the trait's *heritability*. Estimates of heritability for traits such as yield in grains and growth rate in domestic livestock have played an important role in the enhanced agricultural productivity realized in developed countries of the world during the last half century.

1.6 Genomics: Analyses of Entire Genomes

The science of genetics began with Mendel's monohybrid, dihybrid, and trihybrid crosses in garden peas at the monastery in Brünn (now Brno in the Czech Republic). Most of the early studies focused on the effects of one or a few genes. As the science matured, slot-blot and dot-blot hybridization technology allowed geneticists to simultaneously examine the expression of many genes. These blot hybridization procedures culminated in the development of microarray technologies that facilitated studies of hundreds to thousands of genes at once. Then, in 1995, the complete nucleotide sequence of the genome of the bacterium *Haemophilus influenzae* was published. The nucleotide sequences of many other bacteria were subsequently reported, along with the sequence of the genome of the yeast *Saccharomyces cerevisiae*. The nearly complete sequences of several model systems – the worm *Caenorhabditis elegans*, the fruit fly *D. melanogaster*, and the plant *Arabidopsis thaliana* – followed. Then, in early 2001, two drafts of the sequence of the human genome were published, followed in 2002 by drafts of the sequences of the genomes of two subspecies of rice.

Sophisticated computer programs were developed that allowed scientists to scan these sequences and identify open reading frames (ORFs), sequences with no "stop" signals that would prevent the synthesis of a protein product in at least one of the three reading frames.

These tools could be used to predict the presence and location of all the genes in a genome with reasonable accuracy. Gene-specific hybridization probes were then synthesized on the basis of the gene sequences and were used to prepare whole-genome microarrays. New technologies were used to array thousands of oligonucleotide hybridization probes on silicon wafers only a few square centimeters in size. These microarrays, commonly called "gene chips," allow geneticists to simultaneously study the expression of all the genes in an organism. All these developments led to a new subdiscipline of genetics – *genomics* – focused on the structure and function of entire genomes.

2 DNA Replication: The Genotypic Function

The genetic information of an organism must be transmitted from cell to cell during development and from generation to generation during reproduction. This transfer of genes from parents to offspring – the *genotypic function* – occurs by the accurate replication of DNA, that is, by the production of two progeny DNA molecules that are identical to the parental DNA molecule.

2.1 Semiconservative Replication

When Watson and Crick worked out the double-helix structure of DNA in 1953, they recognized that the complementary nature of the two strands – A paired with T and G paired with C – might play an important role in its replication. If the two strands of a parental double helix of DNA separated, the base sequence of each parental strand could serve as a template for the synthesis of a new complementary strand, producing two identical progeny double helices. This process is called *semiconservative replication* because the parental double helix is half conserved, each parental single strand remaining intact (Fig. 5). Meselson and Stahl documented the semiconservative replication of DNA in *E. coli* in 1958.

2.2 Proofreading

DNA replication is amazingly accurate with only about one error for every billion bases incorporated. This accuracy is necessary to keep the mutation load at a tolerable level, especially in large genomes such as those of mammals, which contain 3×10^9 nucleotide pairs. On the basis

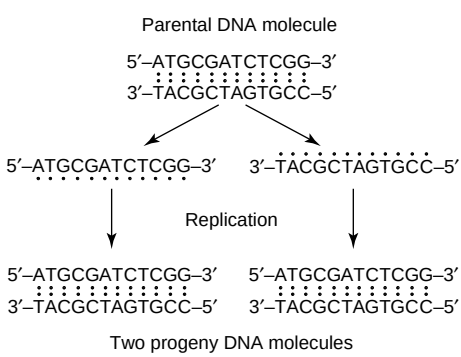

Fig. 5 Semiconservative DNA replication. The single strands of the parental double helix are separated and each strand serves as a template for the synthesis of a complementary strand of DNA. The process results in the production of two progeny double helices that are identical to the parental double helix.

of the dynamic structures of the four nucleotides in DNA, the observed fidelity of DNA replication is much higher than expected. The thermodynamic changes in nucleotides that allow the formation of hydrogen-bonded base pairs other than A : T and G : C predict error rates of 10^{-5} to 10^{-4} or 1 error per 10 000 to 100 000 incorporated nucleotides. The predicted error rate of 10 000 times the observed error rate raises the question of how this high fidelity of DNA replication is achieved. The answer is that a mechanism for *proofreading* the nascent DNA chain as it is being synthesized has evolved in living organisms.

DNA proofreading involves scanning the termini of nascent DNA chains for errors and correcting them before continuing chain extension. This process is carried out by a 3′ → 5′ exonuclease activity that is built into DNA polymerases. When a template–primer DNA has a terminal mismatch (an unpaired or incorrectly paired base or sequence of bases at the 3′ end of the primer), the 3′ → 5′ exonuclease activity of the DNA polymerase clips off the unpaired base or bases (Fig. 6). When an appropriately base-paired terminus is produced, the 5′ → 3′ polymerase activity of the enzyme begins resynthesis by adding nucleotides to the 3′ end of the primer strand. In mutant organisms that lack the proofreading activity of polymerases, the mutation rate is orders of magnitude higher than that in organisms with normal proofreading activity. Thus, proofreading is an important component of the semiconservative replication of DNA.

2.3 The Complex Replisome

DNA replication is complex, requiring the participation of a large number of proteins, and only a few of the most

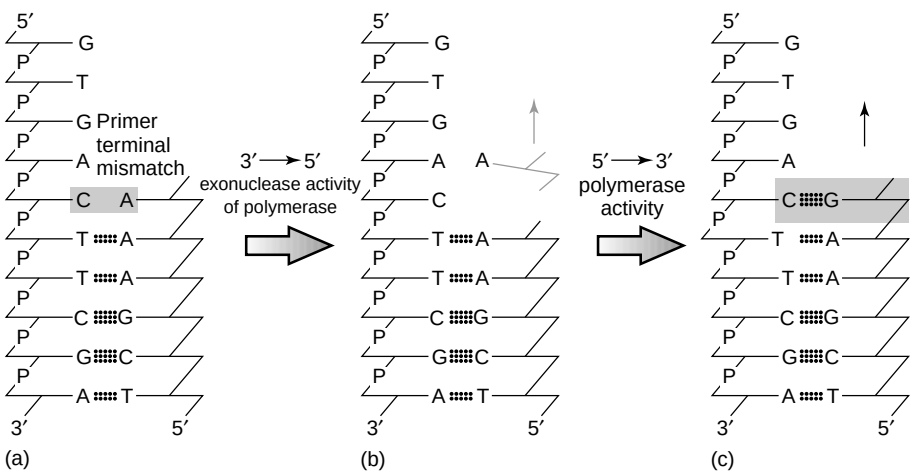

Fig. 6 Proofreading by the 3′ → 5′ exonuclease activity of DNA polymerases during DNA replication. If DNA polymerase is presented with a template and primer containing a 3′ primer terminal mismatch (a), the 3′ → 5′ exonuclease activity will cleave off the mismatched terminal nucleotide (b). Then, given a correctly base-paired primer terminus, DNA polymerase will catalyze a 5′ → 3′ covalent extension of the primer strand (c). (Reproduced from Snustad, D.P., Simmons, M.J., *Principles of Genetics*, 3rd edition, Copyright © 2003 by John Wiley & Sons, Inc., Hoboken, NJ, USA. This material is used by permission of John Wiley & Sons, Inc.)

important features of the process will be considered here. A nucleic acid chain has a chemical polarity based on the phosphodiester bonds that link the 5′ and 3′ carbons of adjacent nucleotides; that is, each chain will have a 5′ carbon at one end and a 3′ carbon at the other end. The complementary strands of a DNA double helix have opposite chemical polarity, with one strand $5' \rightarrow 3'$ and the other strand $3' \rightarrow 5'$, moving unidirectionally along the molecule. Because both nascent strands are extended as a replication fork moves along a parental double helix, one strand is extended at the macromolecular level in the $5' \rightarrow 3'$ direction and the other strand in the $3' \rightarrow 5'$ direction. However, DNA polymerases only catalyze $5' \rightarrow 3'$ synthesis; at the molecular level, all synthesis is $5' \rightarrow 3'$.

DNA synthesis is continuous on the progeny strand that is being extended in the overall $5' \rightarrow 3'$ direction but is discontinuous on the strand growing in the overall $3' \rightarrow 5'$ direction (Fig. 7). Discontinuous replication occurs by the synthesis of short DNA strands (1000 to 2000 nucleotides long in bacteria and 100 to 200 nucleotides long in eukaryotes). The short DNA strands are called "Okazaki fragments" after the scientists who discovered them.

The Okazaki fragments are initiated by short RNA primers synthesized by DNA primase. The RNA primers are subsequently replaced by DNA sequences by the combined $5' \rightarrow 3'$ exonuclease and polymerase activities of a repair DNA polymerase (DNA polymerase I in *Escherichia coli*), and the Okazaki fragments are then joined by DNA ligase.

DNA replication involves many additional enzymes and other proteins. Replication requires that the two strands of a parental DNA molecule be separated during the synthesis of new complementary strands. Given that each gyre, or turn, of DNA is about 10 nucleotide pairs long, a DNA molecule must be rotated 360° once for every 10 replicated base pairs. In *E. coli*, DNA replicates at a rate of about 30 000 nucleotides per minute. Thus, a replicating DNA molecule must spin at 3000 revolutions per minute to facilitate the unwinding of the parental DNA strands. The unwinding process is catalyzed by enzymes called DNA helicases. The unwound strands are prevented from re-pairing by becoming coated with single-strand DNA-binding proteins (SSB proteins). The binding of SSB proteins to single-stranded DNA is cooperative; that is, the binding of the first SSB monomer stimulates the binding of additional monomers at contiguous sites on the DNA chain. Because of the cooperativity of SSB protein binding, an entire single-stranded region of DNA is rapidly coated with SSB protein.

Bacterial chromosomes contain circular molecules of DNA. With DNA spinning at 3000 revolutions per minute during unwinding of the parental strands, a swivel or axis of rotation is required to prevent tangles of supercoils from forming ahead of the replication fork. The required axes of rotation are provided by enzymes called DNA topoisomerases. The topoisomerases catalyze transient breaks in DNA molecules but use covalent linkages to themselves to hold on to the cleaved molecules, allowing subsequent reformation of the cleaved bonds. The transient single-strand break produced by the activity of topoisomerase I provides an axis of rotation that allows the segments of DNA on opposite sides of the break to spin independently, with the phosphodiester bond in the intact strand serving as

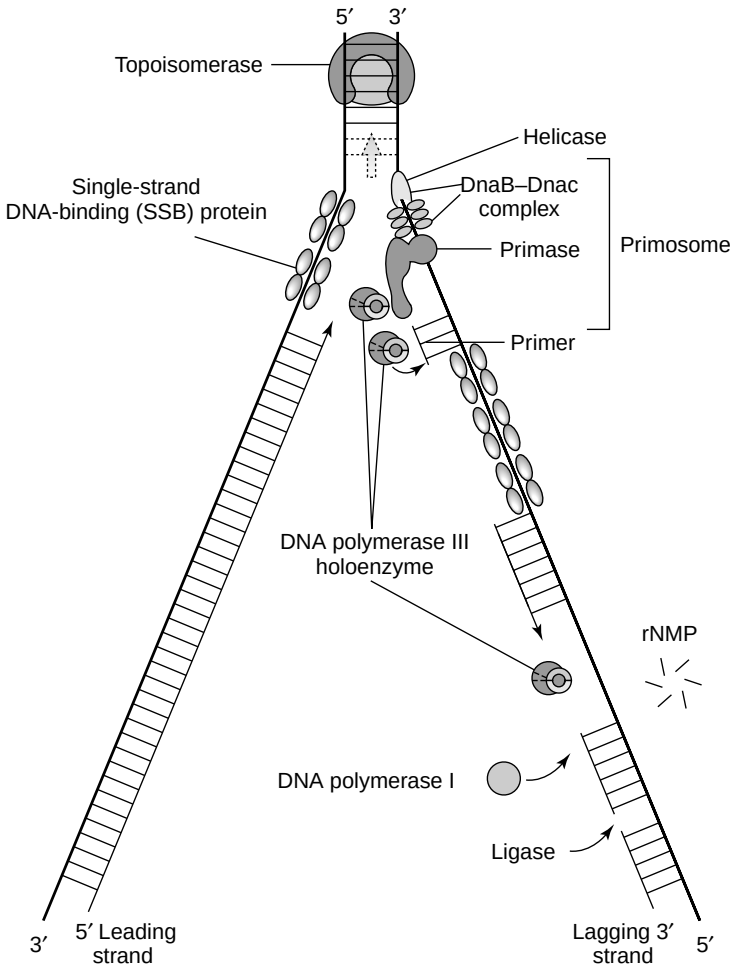

Fig. 7 Diagram of a replication fork in *E. coli* showing the major components of the replisome (rNMP, ribonucleoside monophosphates). (Reproduced from Snustad, D.P., Simmons, M.J., *Principles of Genetics*, 3rd edition, Copyright © 2003 by John Wiley & Sons, Inc., Hoboken, NJ, USA. This material is used by permission of John Wiley & Sons, Inc.)

a swivel. Thus, during DNA replication, only a short segment of DNA in front of the replication fork needs to spin – the segment up to the closest transient nick by topoisomerase I. Another *E. coli* topoisomerase called DNA gyrase is also required for DNA replication. DNA gyrase uses energy from ATP and introduces and removes negative (left-handed) supercoils in DNA.

DNA replication also requires a number of different proteins that are involved in the initiation or priming of synthesis at specific origins of replication. These proteins bind to specific nucleotide sequences at the origin and

induce localized strand separation or "bubbles" in which the synthesis of nascent strands begins. All the enzymes and DNA-binding proteins involved in replication assemble into a *replisome* at each replication fork and act in concert as the fork moves along the parental DNA molecule (Fig. 7).

3
Gene Expression: The Phenotypic Function

The genetic information controls the growth and development of the organism, be it a virus, a bacterium, a plant, or an animal. This genetic information must be expressed accurately – the *phenotypic function* – both spatially and temporally to produce the appropriate three-dimensional form of the organism. In multicellular organisms, the genetic information must control the growth and differentiation of the organism from the single-celled zygote to the mature adult. To accomplish this phenotypic function, each gene of an organism must be expressed at the proper time and in the proper cells during development. The initial steps in the pathways of gene expression, transcription and translation, are quite well elucidated; these steps are illustrated for the expression of the human β-globin gene in Fig. 8. In contrast,

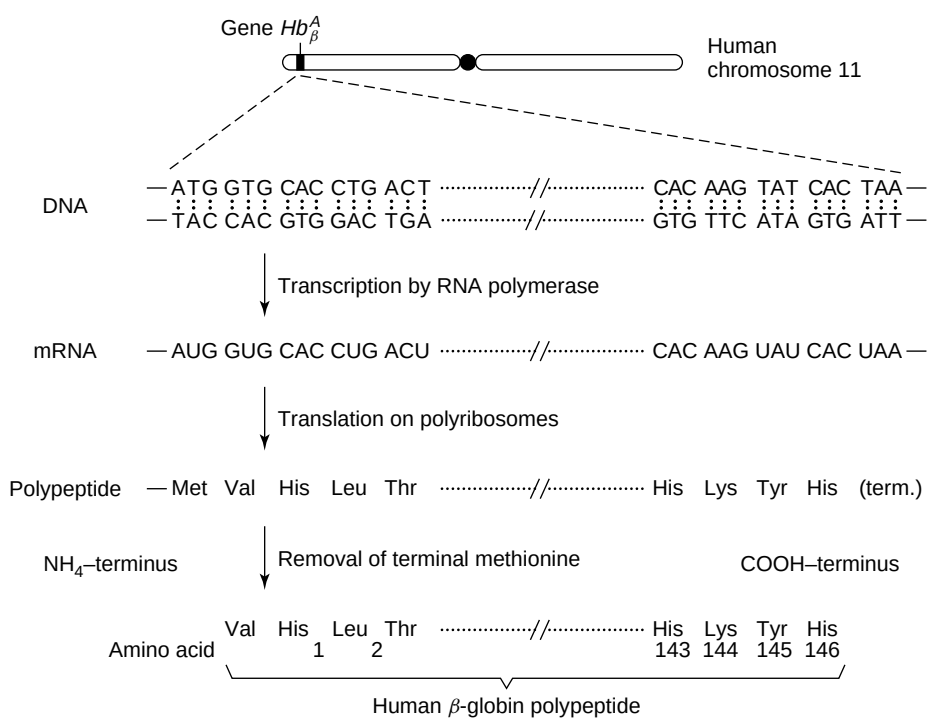

Fig. 8 Schematic diagram showing the first steps in the expression of the human gene (Hb^A_β) encoding β-globin: transcription, translation, and the proteolytic removal of the amino-terminal methionine residue from the primary translation product. For simplicity, only the terminal portions of the coding sequence and the polypeptide product are shown.

we are just beginning to understand morphogenesis at the cell, tissue, and organ levels.

3.1 Colinearity between Gene and Polypeptide

The genetic information is stored in linear sequences of nucleotide pairs in DNA (or nucleotides in RNA, in some cases). Transcription and translation convert this genetic information into *colinear* sequences of amino acids in polypeptides, which function as the key intermediaries in the genetic control of the phenotype. The first three base pairs of the coding sequence of a gene specify the first amino acid of the polypeptide, the next three base pairs (four to six) specify the second amino acid, and so on, in a colinear fashion (Fig. 9). Although the coding regions of most of the genes in higher eukaryotes are interrupted by noncoding sequences called introns, their presence does not invalidate the concept of colinearity. The presence of introns in genes simply means that there is no direct correlation in physical distances between the positions of base-pair coding triplets in a gene and the positions of amino acids in the polypeptide specified by that gene. The coding sequences and the polypeptides that they encode are still colinear.

3.2 Transcription

The first step in gene expression, *transcription*, involves converting genetic information stored in the form of base pairs in double-stranded DNA into the sequence of bases in a single-stranded molecule of messenger RNA (mRNA). This process is catalyzed by enzymes called RNA polymerases and occurs when one strand of the DNA is used as a template to synthesize a complementary strand of RNA using the same base-pairing rules as for DNA replication, except that uracil is incorporated into RNA at positions where thymine would be present in DNA (see Fig. 8, top). Like replication, transcription is a complex process involving numerous proteins called transcription factors. Transcription is initiated at specific sequences within regions called promoters located adjacent to the gene. Transcription initiation factors bind to these sequences and induce localized unwinding of the DNA molecule. RNA synthesis occurs within these locally unwound "transcription bubbles" as RNA polymerase moves along the DNA template strand. Chain extension occurs in the 5′ → 3′ direction by a mechanism very similar to DNA synthesis. The termination of transcription also occurs at specific nucleotide sequences,

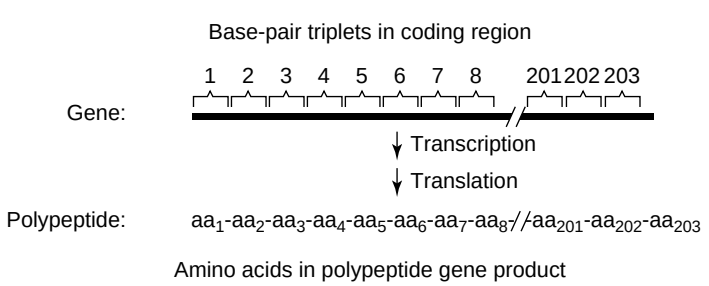

Fig. 9 Colinearity between the base-pair triplets in the coding region of a gene and the amino acid sequence in the polypeptide product of the gene.

sometimes with the aid of other termination proteins.

3.3 Introns and RNA Splicing

Most of the genes of eukaryotes are interrupted by sequences that are not represented in the mature RNA products of these genes. These sequences are called *introns* (for *in*tervening sequences). The coding sequences and other sequences that are present in the final RNA products of these interrupted genes are called *exons* (for *ex*pressed sequences). The structure of a typical eukaryotic gene is shown in Fig. 10.

Not all eukaryotic genes contain introns, so they are not required for expression. When present, the number of introns per gene varies from one to over 50, and the intron size varies from about 50 nucleotide pairs to thousands of nucleotide pairs. The human *DMD* gene, which is responsible for Duchenne muscular dystropy when nonfunctional, is one of the largest known. The *DMD* gene contains 78 introns and is over 2.5-million nucleotide pairs in length. Rare genes of Archaea and of a few viruses of prokaryotes also contain introns. In the case of these "split" genes, the primary transcript contains the entire sequence of the gene and the intron sequences are excised during RNA processing.

For genes that encode proteins, the splicing mechanism must be precise; it must join exon sequences with accuracy to the single nucleotide to assure that codons in exons distal to introns are read correctly. Accuracy to this degree would seem to require precise splicing signals, presumably nucleotide sequences within introns and

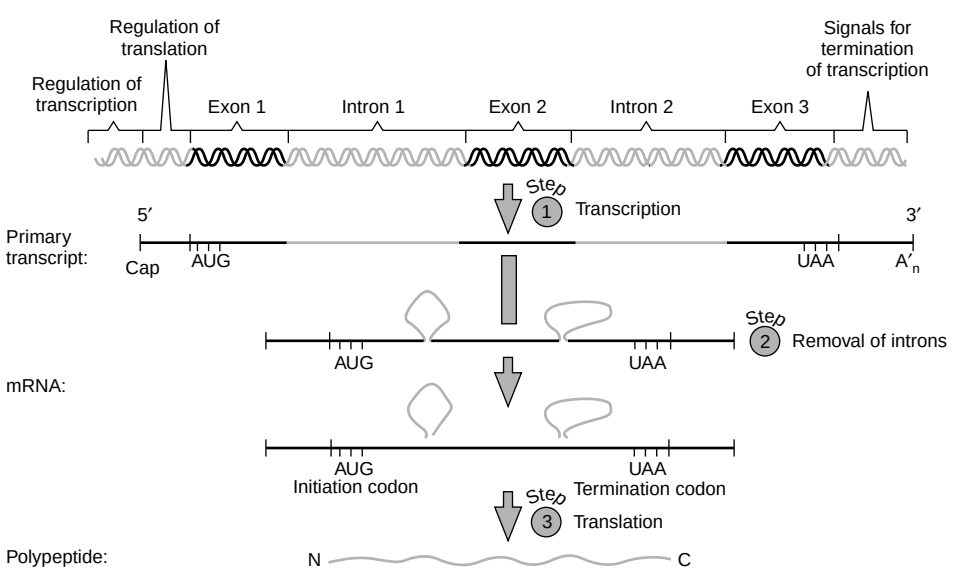

Fig. 10 Structure of a typical eukaryotic gene. Note that the intron sequences are present in the primary transcript but are removed during nuclear processing to produce the mature mRNA prior to its export to the cytoplasm for translation. (Reproduced from Snustad, D.P., Simmons, M.J., *Principles of Genetics*, 3rd edition, Copyright © 2003 by John Wiley & Sons, Inc., Hoboken, NJ, USA. This material is used by permission of John Wiley & Sons, Inc.)

at the exon–intron junctions. However, in the primary transcripts of nuclear genes, the only completely conserved sequences of different introns are the dinucleotide sequences at the ends of introns. In primary transcripts, all introns begin with GU and end with AG. For nuclear genes, there is also one somewhat conserved sequence located about 30 nucleotides from the 3′ splice site. The introns of genes of mitochondria and chloroplasts also contain conserved sequences, but they are different from those of nuclear genes.

There are three major classes of intron excision from RNA transcripts. The introns of tRNA precursors are excised by precise endonucleolytic cleavage and ligation reactions catalyzed by special splicing endonuclease and ligase activities. The introns of some rRNA precursors are removed autocatalytically in a unique reaction mediated by the RNA molecule itself. No protein enzymatic activity is involved. The introns of nuclear pre-mRNA transcripts are excised in two-step reactions carried out by complex ribonucleoprotein particles called spliceosomes. For details about the splicing mechanisms, see [Lewin (2000) or Alberts et al. (2002)].

3.4 RNA Editing

According to the central dogma of molecular biology, genetic information flows from DNA to RNA to protein during gene expression. Normally, the genetic information is not altered in the mRNA intermediary. However, the discovery of *RNA editing* has shown that exceptions do occur. RNA editing processes alter the information content of gene transcripts in two ways: (1) by changing the structures of individual bases and (2) by inserting or deleting uridine monophosphate residues. The first type of RNA editing, which results in the substitution of one base for another base, is rare. This type of editing was discovered in studies of the apolipoprotein-B *(apoB)* genes and mRNAs in rabbits and humans. Apolipoproteins are blood proteins that transport certain types of fat molecules in the circulatory system. In the liver, the *apoB* mRNA encodes a large protein 4563 amino acids long. In the intestine, the *apoB* mRNA directs the synthesis of a protein only 2153 amino acids long. Here, a C residue in the pre-mRNA is converted to a U, generating an internal UAA translation–termination codon, which results in the truncated apolipoprotein. The C → U conversion is catalyzed by a sequence-specific RNA-binding protein with an activity that removes amino groups from cytosine residues. In some transcripts present in plant mitochondria, most of the Cs are converted to U residues.

A second, more complex type of RNA editing occurs in the mitochondria of trypanosomes. In this case, uridine monophosphate residues are inserted (occasionally deleted) into gene transcripts, causing major changes in the polypeptides specified by the mRNA molecules. This RNA editing process is mediated by guide RNAs transcribed from distinct mitochondrial genes. The guide RNAs contain sequences that are partially complementary to the pre-mRNAs to be edited. Pairing between the guide RNAs and the pre-mRNAs results in gaps with unpaired A residues in the guide RNAs. The guide RNAs serve as templates for editing, as Us are inserted in the gaps in pre-mRNA molecules opposite the As in the guide RNAs. In some cases, two or more different guide RNAs participate in the editing of a single pre-mRNA. For some unknown reason, RNA editing plays a major role in

the expression of genes in the mitochondria of trypanosomes and plants.

3.5 Translation

During *translation*, the sequence of bases in the mRNA molecule is converted ("translated") into the specified sequence of amino acids in the polypeptide gene product according to the rules of the genetic code (Fig. 8, center). Each amino acid is specified by one or more codons, and each codon contains three nucleotides. Of the 64 possible nucleotide triplets, 61 specify amino acids and 3 specify polypeptide chain termination. Translation occurs on ribosomes, which are complex macromolecular structures located in the cytoplasm (Fig. 11). Translation involves three types of RNA, all of which are transcribed from DNA templates (chromosomal genes). In addition to mRNAs, 3 to 5 RNA molecules (rRNA molecules) are present as part of the structure of each ribosome, and 40 to 60 RNA molecules (tRNA molecules) function as adaptors by mediating the incorporation of the proper amino acids into polypeptides in response to specific nucleotide sequences in mRNAs. The amino acids are attached to the correct tRNA molecules by a set of

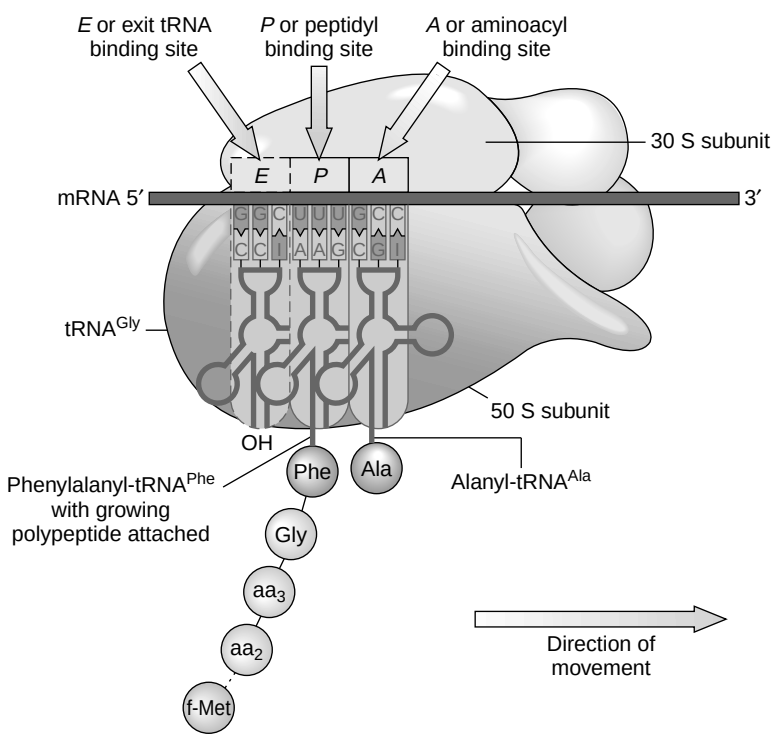

Fig. 11 Diagram of a bacterial 70S ribosome showing the three tRNA binding sites and the base pairing between the codons of the mRNA and the anticodons of the tRNAs. (Reproduced from Snustad, D.P., Simmons, M.J., *Principles of Genetics*, 3rd edition, Copyright © 2003 by John Wiley & Sons, Inc., Hoboken, NJ, USA. This material is used by permission of John Wiley & Sons, Inc.)

activating enzymes called aminoacyl tRNA synthetases. The tRNA molecules contain nucleotide triplets called anticodons, which base pair with the codons in mRNA during the translation process.

The ribosomes may be thought of as workbenches, complete with machines and tools needed to make a polypeptide. They are nonspecific in the sense that they can synthesize any polypeptide (any amino acid sequence) encoded by a particular mRNA molecule, even an mRNA from a different species. Each mRNA molecule is simultaneously translated by several ribosomes, resulting in the formation of polyribosomes. The translation of the sequence of nucleotides in an mRNA molecule into the sequence of amino acids in its polypeptide product can be divided into three stages: (1) polypeptide chain initiation; (2) chain elongation; and (3) chain termination.

In *E. coli*, the translation initiation process involves the 30S subunit of the ribosome, a special initiator tRNA, an mRNA molecule, three soluble protein initiation factors IF-1, IF-2, and IF-3, and one molecule of GTP. In the first stage of the initiation of translation, a free 30S subunit interacts with an mRNA molecule and the initiation factors. The 50S subunit joins the complex to form the 70S ribosome in the final step of the initiation process. The addition of the 50S ribosomal subunit to the complex positions the initiator tRNA, methionyl-$tRNA_f^{Met}$, in the peptidyl (*P*) site of the ribosome with the anticodon of the tRNA aligned with the AUG initiation codon of the mRNA. Methionyl-$tRNA_f^{Met}$ is the only aminoacyl tRNA that can enter the *P* site directly, without first passing through the aminoacyl (*A*) site. With the initiator AUG positioned in the *P* site, the second codon of the mRNA is in register with the *A* site, dictating the aminoacyl tRNA binding specificity at that site and setting the stage for the second phase in polypeptide synthesis, chain elongation.

The process of polypeptide chain elongation is basically the same in both prokaryotes and eukaryotes. The addition of each amino acid to the growing polypeptide occurs in three steps: (1) binding of an aminoacyl tRNA to the *A* site of the ribosome; (2) transfer of the growing polypeptide chain from the tRNA in the *P* site to the tRNA in the *A* site by the formation of a new peptide bond; and (3) translocation of the ribosome along the mRNA to position the next codon in the *A* site. During step 3, the nascent polypeptide-tRNA and the uncharged tRNA are translocated from the *A* and *P* sites to the *P* and *E* sites respectively. These three steps are repeated in a cyclic manner throughout the elongation process. Polypeptide chain elongation proceeds rapidly. In *E. coli*, all three steps required for the addition of one amino acid to the growing polypeptide chain to occur in about 0.05 s. Thus, the synthesis of a polypeptide containing 300 amino acids takes only about 15 s.

Polypeptide chain elongation undergoes termination when any of the three chain-termination codons (UAA, UAG, or UGA) enters the *A* site on the ribosome. These three stop codons are recognized by soluble proteins called release factors (RFs). In *E. coli*, there are two release factors, RF-1 and RF-2. RF-1 recognizes termination codons UAA and UAG; RF-2 recognizes UAA and UGA. In eukaryotes, a single release factor (eRF) recognizes all three termination codons. Termination is completed by the release of the mRNA molecule from the ribosome and the dissociation of the ribosome into its subunits.

3.6 Inteins

Occasionally, the primary translation product of a gene contains one or more short amino acid sequences, called *inteins*, that excise themselves from the nascent polypeptide. Inteins occur in both eukaryotic and prokaryotic polypeptides; one of the first inteins discovered is in the RecA protein (required for recombination) in *Mycobacterium tuberculosis*, the bacterium that causes tuberculosis. The ability to carry out intein excision is a function of the structure of the primary translation product and is thus encoded in the gene just like any other amino acid sequence.

3.7 Some Complex Gene/Protein Relationships

As discussed in Section 3.3, most eukaryotic genes are split into expressed sequences (exons) and intervening sequences (introns). In some cases, transcripts of split genes may undergo several different types of splicing, making the relationships between genes and proteins more complex than the usual one gene–one polypeptide. In other cases, expressed genes are assembled from "gene pieces" during the development of the specialized cells in which they are expressed.

When the transcripts of an interrupted gene undergo alternate pathways of transcript splicing, different exons are joined to produce a related set of mRNAs that encode a family of closely related polypeptides. These interrelated polypeptides are called *protein isoforms*. Alternate splicing pathways are often tissue-specific, producing related proteins that carry out similar, but not necessarily identical, functions in different types of cells. The mammalian tropomyosin genes produce complex families of protein isoforms. Tropomyosins regulate muscle contraction in animals. One mouse tropomyosin gene produces at least 10 different polypeptides by alternate pathways of transcript splicing. Genes of this type do not fit the one gene–one polypeptide concept very well. Such genes can be defined as DNA sequences that are single units of transcription and encode a set of protein isoforms.

Genetic information is not always organized into genes of the type described in Section 1.2. In a few cases, genes are assembled from a storehouse of gene segments during the development of an organism. The immune system of vertebrate animals depends on the synthesis of proteins called antibodies to provide protection against infections by viruses, bacteria, toxins, and other foreign substances. Each antibody contains four polypeptides, two identical heavy chains and two identical light chains. The light chains are of two types: kappa and lambda. Each antibody chain contains a variable region, which exhibits extensive diversity from antibody to antibody, and a constant region, which is largely the same in all antibodies. In germ-line chromosomes, the DNA sequences encoding these antibody chains are present in gene segments, and the gene segments are joined together to produce genes during the differentiation of the antibody-producing B-lymphocytes from progenitor cells. The B-lymphocytes subsequently differentiate into antibody-secreting plasma cells.

This process of *gene assembly during development* is illustrated in Fig. 12. A kappa light chain gene is assembled from three gene segments V_k (V for variable region), J_k (J for joining segment), and

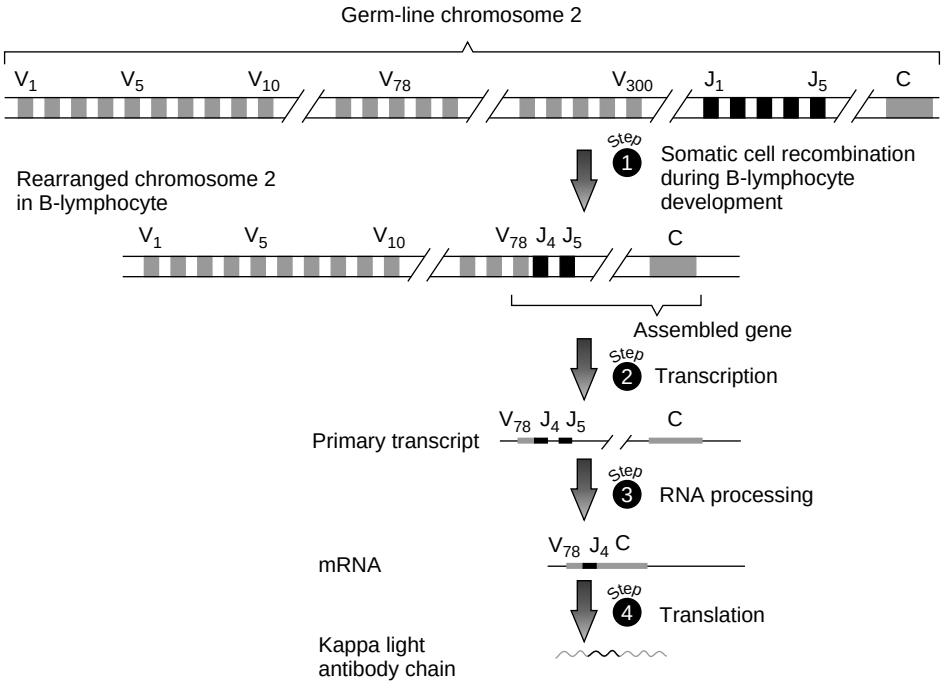

Fig. 12 Assembly of a gene encoding an antibody kappa light chain from gene segments during B-lymphocyte differentiation in humans. (Reproduced from Snustad, D.P., Simmons, M.J., *Principles of Genetics*, 3rd edition, Copyright © 2003 by John Wiley & Sons, Inc., Hoboken, NJ, USA. This material is used by permission of John Wiley & Sons, Inc.)

C_k (C for constant region) during B-lymphocyte development. Together, the V_k and J_k gene segments encode the variable region of the kappa light chain, whereas the C_k gene segment encodes the constant region. No functional V_k–J_k–C_k kappa light chain gene is present in any human germ-line chromosome. Instead, human chromosome 2 contains a cluster of about 300 V_k gene segments, another cluster of 5 J_k gene segments, and a single C_k gene segment (Fig. 12). During the differentiation of each B-lymphocyte, recombination joins one of the V_k gene segments to one of the J_k gene segments. Any J_k segments remaining between the newly formed V_k–J_k exon and the C_k gene segment become part of an intron that is removed during the processing of the primary transcript. Similar somatic recombination events are involved in the assembly of the genes encoding antibody heavy chains, lambda light chains, and T-lymphocyte receptor proteins.

3.8 Pathways of Gene Expression are Often Complex

The pathway through which a gene exerts its effect on the phenotype of the organism is often long and complex, especially in multicellular eukaryotes (Fig. 13). Pathways of gene action frequently involve protein–protein and other macromolecular interactions, cell–cell interactions and

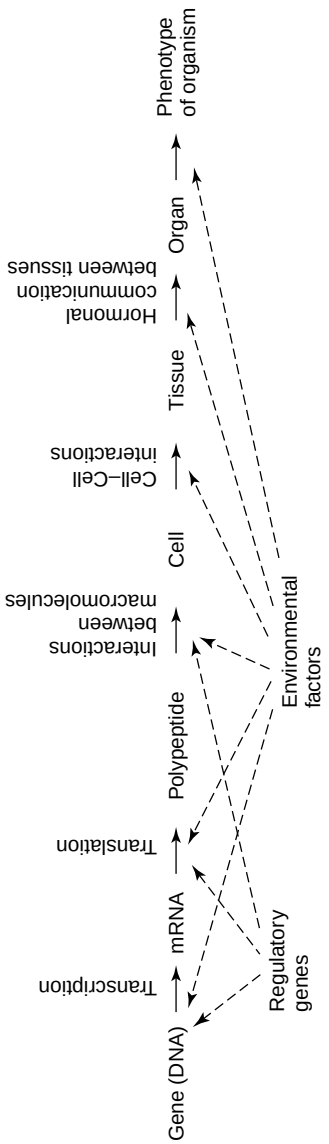

Fig. 13 A typical pathway of gene expression showing some of the factors that can influence the effect of a gene on the phenotype of an organism.

intercellular communication by hormones and other signal molecules, tissue and organ interactions, and restrictions imposed by environmental factors. Note that, although not shown here, each gene also has an effect on the phenotype of the population in which the organism lives (population genetics) and ultimately on the phenotype of the biosphere (ecological genetics).

3.9 Pleiotropy and Epistasis

Sometimes, a single gene influences many aspects of an organism's phenotype. When this occurs, the gene is said to be pleiotropic. The mutant gene that causes sickle-cell anemia in humans provides a classic example of *pleiotropy*. This inherited disorder results from a mutation in the β-globin gene that changes the sixth amino acid from a glutamic acid in the wild-type β-globin to a valine in the sickle-cell polypeptide. This amino acid substitution alters the conformation of the polypeptide, causing aggregation of hemoglobin molecules and the development of grossly deformed red blood cells. These sickle-shaped cells have reduced capacity to transport oxygen and, therefore, cause the anemia. However, the sickle cell allele not only causes hemolytic anemia in the homozygous state but also has many other effects such as enlarged spleen, impaired growth, recurrent pain, and increased susceptibility to microbial and viral infections. Perhaps the allele's most unexpected pleiotropic effect is the enhanced resistance to *Paramecium falciparum* malaria, which it provides when present in the heterozygous state.

When two or more genes influence a trait, an allele of one of them may mask the effects of alleles of the other gene on the phenotype. When an allele has such an overriding effect, it is said to be epistatic to the other genes. The occurrence of such interactions between different genes (nonalleles) is called *epistasis*. There are many examples of epistasis. In *Drosophila*, a recessive mutation in the cinnabar gene causes the eyes of the fly to be bright red. A recessive mutation in another gene results in white eyes. Flies that are homozygous for both these mutations have white eyes. The white mutation is, therefore, epistatic to the cinnabar mutation. In humans and other mammals, recessive mutant alleles that cause albinism are usually epistatic to other genes affecting eye and hair color.

3.10 Penetrance and Expressivity

Sometimes, a mutant gene can be present in an organism without having any effect on its phenotype. Such a gene is said to exhibit *incomplete penetrance*. In humans, polydactyly, the presence of extra fingers and toes, provides an example of incomplete penetrance. This condition is caused by a dominant mutation that results in polydactyly in some, but not all, individuals who are known to carry it. Penetrance is measured by determining what proportion of the individuals who carry a gene exhibit the trait made possible by its presence. Incomplete penetrance can be a serious problem in pedigree analysis, sometimes resulting in the assignment of incorrect genotypes.

In other cases, the presence of a gene may result in a range of phenotypic effects from mild to severe. Such a gene is said to exhibit *variable expressivity*. The dominant Lobe eye mutation in *Drosophila* provides an example of variable expressivity; its phenotypic effect varies from tiny compound eyes, through a

full range of intermediates, to large, lobulated eyes.

3.11 Intragenic Complementation

The results of complementation tests are usually unambiguous when recessive mutations that result in the synthesis of no gene product, partial gene products, or totally defective gene products are used – for example, deletions of segments of genes or polypeptide chain–terminating mutations. When mutations that cause amino acid substitutions are used, the results are sometimes ambiguous because of the occurrence of a phenomenon called *intragenic complementation*.

The functional forms of some proteins are dimers or higher multimers consisting of two or more polypeptides. These polypeptides may be either the products of a single gene or the products of two or more different genes. When the active form of the protein contains two or more homologous polypeptides, intragenic complementation sometimes occurs. *Inter*genic complementation (discussed in Section 1.2) and *intra*genic complementation are distinct phenomena.

As a simple example of intragenic complementation, consider an enzyme that functions as a homodimer, that is, a protein containing two copies of a specific gene product. An organism that is heterozygous for two different mutations in the gene will produce some dimers that contain the two different mutant polypeptides. We call these heterodimers. Such heterodimers may have partial or complete (wild-type) function; when this occurs, intragenic complementation has occurred. In such cases, the *trans* heterozygote has a wild-type phenotype or a phenotype intermediate between mutant and wild type. Why such heteromultimers should be active when the two corresponding homomultimers are inactive is not clear. Apparently, the wild-type sequence of amino acids in the nonmutant segment of one mutant polypeptide somehow compensates for the mutant segment of the polypeptide encoded by the second mutant allele, and vice versa. In the case of noncomplementing mutations in a gene encoding a multimeric protein, the heteromultimers are nonfunctional, just like the mutant homomultimers.

3.12 Intergenic Noncomplementation

In some cases, mutations known to be located in two different genes fail to complement each other. This phenomenon is called *intergenic noncomplementation*, and it has become a powerful tool for identifying genes encoding products that interact. If two different polypeptides are part of a macromolecular complex, mutations in the two genes encoding them may not complement each other. The presence of mutant polypeptides in the complex may render the entire complex nonfunctional, even if wild-type products of both genes are present. Thus, the occurrence of intergenic noncomplementation provides evidence that the two gene products interact in some way. Intergenic noncomplementation is allele-specific; not all mutations in the genes involved will exhibit the phenomenon.

3.13 Regulation of Gene Expression

In all organisms, gene expression is highly regulated so that energy is used to synthesize gene products only when those products are needed for the growth and

differentiation of the organism. In higher eukaryotes, only a small proportion of the genes in any genome are expressed in a given cell type. Thus, gene expression is highly programmed such that genes needed to make neurons are only turned on in developing nerve cells, genes needed to make red blood cells are only expressed in progenitors of erythrocytes, and so on. Most microorganisms exhibit a striking ability to regulate the expression of specific genes in response to environmental signals. The expression of particular genes is turned on when the products of these genes are needed for growth. Their expression is turned off when the gene products are no longer needed. Gene expression is regulated at several different levels: transcription, mRNA processing, mRNA turnover, translation, and posttranslation (Fig. 14). However, the regulatory mechanisms with the largest effects on phenotype act at the level of transcription.

Most regulatory mechanisms fit into two general categories: (1) the rapid turn-on or turn-off of gene expression in response to environmental changes (especially important in microorganisms) and (2) preprogrammed circuits of gene expression (important in all organisms).

Certain "housekeeping" gene products – tRNA molecules, rRNA molecules, ribosomal proteins, and the like – are essential components of all living cells. These genes are continually being expressed in most cells; they are referred to as *constitutive genes*. Other gene products are needed only in the presence of specific metabolites. They are not expressed in the absence of the metabolite but are turned on when the metabolite is present. This process is called *induction*. Genes whose expression is regulated in this manner are called *inducible genes*; their products, if enzymes, are called *inducible enzymes*. Enzymes that are involved in catabolic pathways are often inducible. Other genes are turned on and they synthesize their gene products unless the metabolite synthesized by those products is present in the environment. Then, they are turned off. This process is called *repression*, and genes that are regulated in this manner are called *repressible genes*. Enzymes that are components of anabolic pathways are often repressible. Both induction and repression occur at the level of transcription.

The regulation of gene expression, induction or repression, can be accomplished by both positive and negative control mechanisms. Both mechanisms involve the participation of *regulator genes* – genes whose products regulate the expression of other genes. In *positive control mechanisms*, the product of the regulator gene functions by turning on the expression of one or more genes, whereas in *negative control mechanisms*, the product of the regulator gene is involved in shutting off the expression of genes. Positive and negative regulation mechanisms can both mediate either inducible or repressible gene expression.

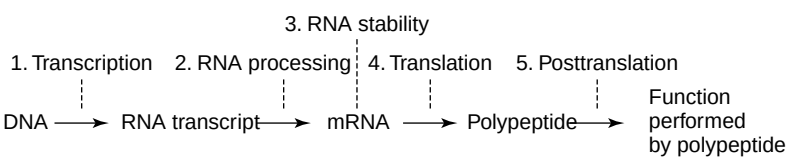

Fig. 14 Pathway of gene expression showing five stages at which gene expression is regulated.

A given gene is expressed when RNA polymerase binds to its promoter and synthesizes an RNA transcript that contains the coding region of the gene. The product of the regulator gene acts by binding to a site called the regulator protein binding site (*RBS*) adjacent to the promoter of the structural gene (or genes). When the product of the regulator gene is bound to *RBS*, transcription of the structural gene(s) is turned on in a positive control system or is turned off in a negative control system. The regulator gene products are called *activators* in positive control systems and *repressors* in negative control systems. Whether a regulator protein can bind to the *RBS* depends on the presence or absence of *effector molecules* in the cell. The effectors are usually small molecules such as amino acids, sugars, and similar metabolites. The effector molecules involved in induction of gene expression are called *inducers*; those involved in repression of gene expression are called *corepressors*.

The effector molecules (inducers and corepressors) bind to regulator gene products (activators and repressors) and cause changes in the three-dimensional structures of these proteins. Such changes in protein conformation are called *allosteric transitions*. In the case of activators and repressors, the allosteric transitions alter their ability to bind to regulator protein binding sites adjacent to the genes that they control.

In microorganisms, *operons* – negatively and coordinately regulated units of gene expression – play important regulatory roles. Each operon contains one to many structural genes, adjacent to promoter and operator sequences. The operator is the binding site for the repressor (inducible system) or the repressor/corepressor complex (repressible system). When the repressor or repressor/corepressor complex is bound to the operator, it prevents RNA polymerase from transcribing the structural genes of the operon. When the repressor is not bound to the operator, RNA polymerase binds to the promoter and transcribes the genes in the operon. Another important regulatory mechanism in microorganisms – attenuation – occurs by the control of premature termination of transcription.

Regulation of gene expression in eukaryotes usually involves the binding of several proteins called *transcription factors* to upstream promoters and to regulatory sequences called *enhancers* and *silencers*, which can be located either upstream of, downstream from, or within the genes that they regulate. As the names suggest, enhancers and silencers increase and decrease, respectively, the levels of gene expression.

3.14 RNA-mediated Gene Silencing

RNA-mediated gene silencing (*RMGS*) was first studied in plants where it was called *posttranscriptional gene silencing* (*PTGS*). Similar RNA-induced gene silencing processes were subsequently discovered in animals and fungi and were called *RNA interference* (*RNAi*) and *quelling*, respectively. RMGS is triggered by the presence of double-stranded RNA (dsRNA). The dsRNA may be composed of two complementary or partially complementary single strands or a self-complementary RNA molecule containing complementary regions such that it folds back on itself and forms a hairpin-like structure. The dsRNAs are degraded to 21- to 22-nucleotide-long dsRNA fragments called *small interfering RNAs* (*siRNAs*) by the

ribonuclease "Dicer." The siRNA fragments subsequently become associated with an endonuclease-containing complex called RISC for "RNAi silencing complex." This complex binds to and degrades mRNA molecules that contain sequences complementary to the sequence of the siRNAs. RMGS requires several other gene products, including an RNA-dependent RNA polymerase, an RNA helicase, and the proteins in RISC. Additional components of the RMGS pathways are currently being investigated in both plants and animals.

RMGS is believed to be a defense mechanism protecting the host organism from viral infections, runaway transposons, and other invading DNAs and dsRNAs. In addition, the recent identification of families of naturally occurring *microRNAs* with self-complementary sequences that form "hairpin" structures and give rise to siRNA-like fragments suggests that RMGS may play important roles in regulating growth and development and other important natural processes. Finally, the mechanisms by which PTGS, RNAi, and quelling occur in various species are probably similar but not necessarily identical. Indeed, there are probably multiple pathways by which RMGS can occur, even within a species.

In addition to their normal biological function(s), RMGS has become an important tool for reverse genetic analyses, allowing researchers to shut off genes in a controlled manner. In *C. elegans*, RNAi was used to systematically "knock out" the expression of each of the genes on chromosomes I and III respectively. In plants, chimeric gene cassettes that direct the synthesis of hairpin RNAs and induce PTGS have proven to be effective in producing gene knockouts. Thus, RMGS is an important tool in the field of functional genomics.

4 Mutation: The Evolutionary Function

Although the genetic information must be transmitted from generation to generation with considerable accuracy, it is not static, but it undergoes occasional change or *mutation – the evolutionary function* – to produce new genetic variability that provides the raw material for ongoing evolution. The new variant genes produced by mutation are called mutant alleles and often result in abnormal or mutant phenotypes. When used in the narrow sense, mutation refers only to changes in the structures of individual genes. However, in the broad sense, mutation refers to any heritable change in the genetic material and includes gross changes in chromosome structure or chromosome aberrations. In addition, the word mutation refers not only to the change in the genetic material but also to the process during which the change occurs.

4.1 Dominance, Partial Dominance, and Codominance

When mutations occur producing new mutant alleles, these alleles can exhibit a range of interactions with the original or wild-type allele. The phenomenon of interactions between various alleles of a given gene is called *dominance*. If an organism that is heterozygous for a wild-type allele and a mutant allele has the same phenotype as an organism that is homozygous for the mutant allele, the mutant allele is *dominant* and the wild-type allele is *recessive*. If the heterozygote has the wild-type phenotype, the reverse is true: the mutant allele is recessive and the wild-type allele is dominant. If the phenotype of the heterozygote is intermediate between the

phenotypes of the respective homozygotes, the mutant allele exhibits *partial dominance* (sometimes called semidominance) or no dominance (if the phenotype is precisely intermediate to the phenotypes of the two homozygotes). If both alleles produce their phenotypic effects in heterozygotes, the alleles are said to be *codominant*. Humans with the AB blood type (genotype $I^A I^B$), for example, have both type A antigens and type B antigens on their red blood cells. Thus, the I^A and I^B alleles are codominant.

4.2 Transitions, Transversions, and Frameshift Mutations

Point mutations within individual genes may be either base-pair substitutions or the insertion or deletion of one or a few contiguous base pairs. Base-pair substitutions usually result in the substitution of a single amino acid in the mutant polypeptide gene product. Base-pair substitutions are of two types: *transitions* and *transversions*. Transitions occur when one purine is substituted for the other purine and the corresponding pyrimidine for pyrimidine substitutions take place in the complementary strand of DNA. Transversions involve purine for pyrimidine and pyrimidine for purine substitutions. Of the 12 different base-pair substitutions, 4 are transitions and 8 are transversions (Fig. 15).

Sickle-cell anemia in humans is the result of a single base-pair substitution, a transversion, in the adult β-globin gene. This disorder occurs in individuals who are homozygous for the altered β-globin gene. This single base-pair substitution in the Hb_β^S gene changes the sixth amino acid of the β-globin polypeptide from glutamic acid in Hb_β^A homozygotes to valine in Hb_β^S homozygotes (Fig. 16). This one amino acid change in the human β-globin chain results in sickle-shaped red blood cells and in sickle-cell anemia in individuals homozygous for the Hb_β^S allele. Thus, a single base-pair substitution in DNA can have a very large effect on the phenotype of the organism harboring the mutation.

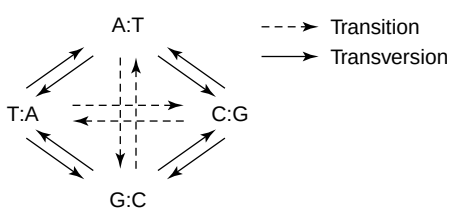

Fig. 15 Base-pair substitutions in DNA.

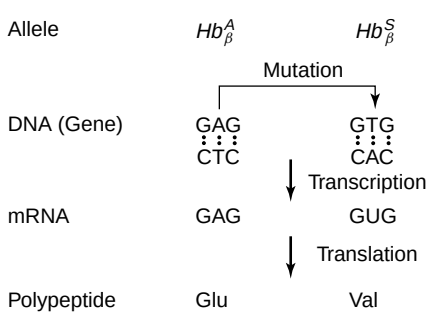

Fig. 16 The mutational origin of sickle-cell anemia in humans.

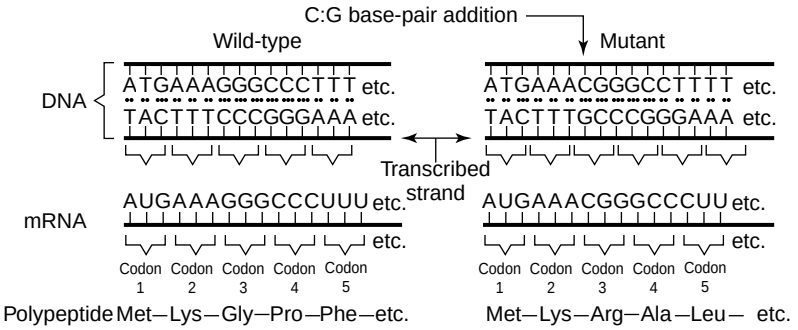

Insertions or deletions of one or two base pairs alter the reading frame of the gene distal to the site of the mutation.

Fig. 17 The effect of a frameshift mutation, in this case a C:G base-pair insertion. (Reproduced from Snustad, D.P., Simmons, M.J., *Principles of Genetics*, 3rd edition, Copyright © 2003 by John Wiley & Sons, Inc., Hoboken, NJ, USA. This material is used by permission of John Wiley & Sons, Inc.)

The insertion or deletion of one or two base pairs within the coding sequence of a gene alters the codon reading frame in the mRNA (Fig. 17); thus, such mutations are referred to as *frameshift mutations*. Frameshift mutations usually result in totally nonfunctional gene products.

4.3 Gain-of-function and Loss-of-function Mutations

Mutations may also be classified on the basis of their effect on the function of the altered gene products. Mutations that diminish or eliminate the activity of a gene product are called *loss-of-function* mutations. Most recessive mutations are simple loss-of-function mutations. However, some loss-of-function mutations are dominant (for example, when the mutant product interferes with the activity of the wild-type gene product) or partially dominant (for example, when the threshold level of gene product activity requires two functional copies of the gene). Many dominant mutations are *gain-of-function* mutations; they produce mutant alleles that encode products with new functions. Some gain-of-function mutations result in the synthesis of altered polypeptides with novel activity. Others result in the synthesis of the gene products in cells or tissues in which they are not normally synthesized.

4.4 Suppressor and Enhancer Mutations

Two classes of mutations have proven to be especially valuable to researchers investigating the genetic control of a specific biological process. Given an organism with a mutant phenotype caused by a mutation in one gene that affects a trait, researchers can often identify other genes affecting the trait by screening for *suppressor mutations* or *enhancer mutations*. Suppressor mutations partially or completely cancel the effect of the original mutation, whereas enhancer mutations cause the mutant phenotype to be more extreme. In both cases, the mutations usually occur in genes that encode products involved in the same pathway or process, acting either upstream of

or downstream from the product of the gene altered by the original mutation.

4.5 Chromosome Aberrations

There are four types of gross chromosome rearrangements or *chromosome aberrations*: duplications, deletions, inversions, and translocations. A *duplication* is the occurrence of a segment of a chromosome in two or more copies per genome. A *deletion* or *deficiency* results from the loss of a segment of a chromosome. An *inversion* occurs when a segment of a chromosome is turned end-for-end relative to its orientation in a normal chromosome. A *translocation* results when a segment of a chromosome is broken off and becomes reattached to another chromosome. All four types of chromosome aberrations can have major effects on the phenotype and fertility of an organism.

4.6 Polyploidy

Numerical changes in chromosome number are referred to as changes in *ploidy*. *Euploidy* involves variations in the number of copies of the genome or complete sets of chromosomes in a cell or an organism. *Aneuploidy* occurs when there are missing or extra copies of one or a few chromosomes.

All the other sections of this article have focussed on the genetics of *diploid* ($2n$) organisms such as humans and fruit flies. Diploid organisms contain two copies of the genome in their germline cells and most, but not all, somatic cells. However, not all organisms are diploids; some animals, for example, certain fish and salamanders, and many plants are *polyploid*, that is, contain more than two copies of the genome. Many important crop species are polyploids: seedless watermelons are triploid ($3n$), cotton is a tetraploid ($4n$), bread wheat is a hexaploid ($6n$), and strawberries are octaploids ($8n$). Polyploidy has played an important role in the evolution of plants.

4.7 Nondisjunction and Aneuploidy

During mitosis, each chromosome duplicates and the two daughter chromosomes separate, with one going to each progeny cell. Similarly, during meiosis, the two homologous chromosomes separate during the first division, and then the two daughter chromosomes (previously chromatids) move to separate cells during the second division. The result of a normal mitosis or meiosis is that all the progeny cells receive complete sets of chromosomes. Occasionally, however, homologous chromosomes and daughter chromosomes fail to separate or disjoin properly during mitosis or meiosis. The failure of chromosomes to separate properly during mitosis or meiosis is called *nondisjunction*. Nondisjunction produces cells or gametes with extra or missing chromosomes; it is the primary cause of aneuploidy.

Whereas euploidy is not detrimental *per se*, aneuploidy is, especially in animals. In humans, aneuploidy is almost always lethal. Excluding individuals who are mosaics containing cells with two or more different karyotypes (chromosome sets), only seven types of aneuploidy exist in humans. *Monosomy* occurs when one member of a pair of chromosomes is missing; *trisomy* occurs when three copies of a chromosome are present. In humans, only one monosomic condition, Turner syndrome (X0, the presence of a

single X chromosome), is viable. Monosomy for any of the autosomes is lethal. Five trisomic conditions are viable in humans: Klinefelter syndrome (XXY), Down syndrome (trisomy 21), Patau syndrome (trisomy 13), Edward syndrome (trisomy 18), triplo-X (XXX), and XYY. Trisomy for any of the other autosomes is lethal. The message is clear; the genes in our genome have evolved as a coadapted set of instructions. Each of the genes in our genome must be expressed at the proper time and in the proper place during development to produce a normal human being. Anything that disrupts the delicate balance of preprogrammed gene expression required for normal development is highly deleterious.

4.8
Expanding Trinucleotide Repeats in Humans

Another class of mutations – *expanding trinucleotide repeats* – are responsible for several inherited neurological diseases in humans. Tandemly repeated sequences of one to six nucleotide pairs are known as simple tandem repeats. Such repeats are dispersed throughout genomes, including the human genome. Repeats of three nucleotide pairs, trinucleotide repeats, are known to increase in copy number and give rise to mutant alleles. In humans, several trinucleotides undergo such increases in copy number. Expanded CGG trinucleotide repeats on the X chromosome are responsible for fragile X syndrome, the most common form of inherited mental retardation. Normal X chromosomes contain from 6 to about 50 copies of the CGG repeat at the *FRAXA* site. Mutant X chromosomes contain up to 1000 copies of the tandem CGG repeat at this site. CAG and CTG trinucleotide repeats are involved in several inherited neurological diseases, including Huntington disease, myotonic dystrophy, Kennedy disease, dentatorubral pallidoluysian atrophy, Machado Joseph disease, and spinocerebellar ataxia. In all these neurological disorders, the severity of the disease is correlated with trinucleotide copy number – the higher the copy number, the more severe the disease symptoms. In addition, the expanded trinucleotides associated with these diseases are unstable in somatic cells and between generations. This instability gives rise to the phenomenon of "anticipation," which is the increasing severity of the disease or earlier age of onset that occurs in successive generations as the trinucleotide copy number increases.

5
Recombination: New Combinations of Genes to be Acted on by Natural or Artificial Selection

Mutation produces new genetic variability, but the resulting mutant genes must be placed in new combinations with previously existing genes so that natural selection (or artificial selection in the case of plant and animal breeding) can preserve those combinations that produce the organisms best adapted to specific environments (or desired by the breeder). These new combinations are produced by recombination mechanisms that are essential to the process of evolution.

5.1
Segregation

In an organism that is heterozygous for two different alleles of a gene, such as tall (*D*) and dwarf (*d*) in Mendel's peas, the two alleles segregate from each other during the formation of gametes. The

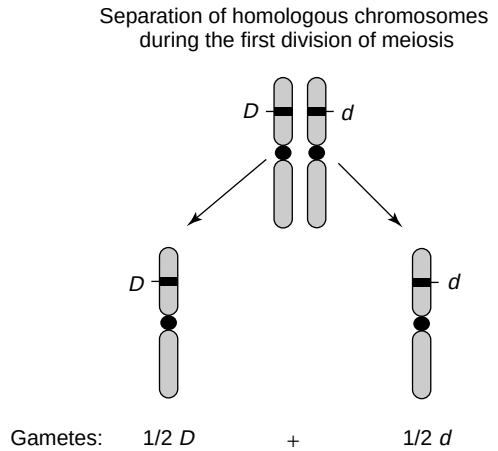

Fig. 18 Segregation of alleles during the first division of meiosis.

biological basis of this segregation is the pairing and subsequent separation of homologous chromosomes during the first (reductional) division of meiosis (Fig. 18).

5.2 Independent Assortment

New combinations of genes on nonhomologous chromosomes are produced by the independent assortment of chromosomes during the first division of meiosis (Fig. 19). Consider a cross between two double homozygotes, for example, *AA BB* and *aa bb*, where genes *A* and *B* are on different chromosomes. The F_1 progeny (F_1 for first filial generation) will be double heterozygotes (*Aa Bb*). Half of the gametes produced by F_1 progeny will have the same combination of alleles as the

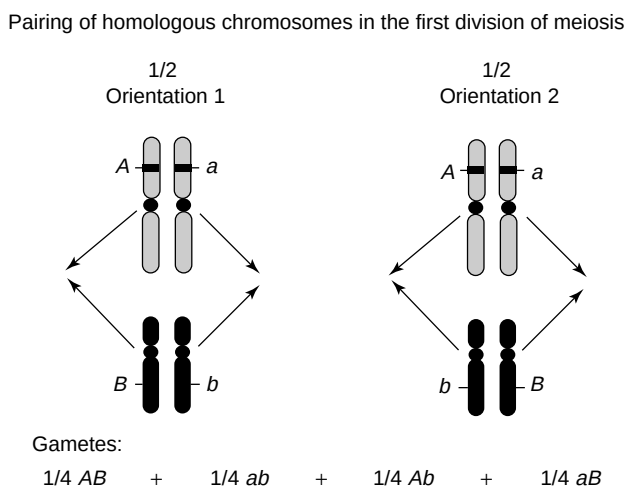

Fig. 19 The generation of new combinations of genes by the independent assortment of nonhomologous chromosomes during meiosis.

parents (1/4 AB + 1/4 ab). The other half will have new (recombinant) combinations of the alleles (1/4 Ab and 1/4 aB). The recombination of pairs of alleles on different chromosomes is the result of the *independent assortment* of the homologous chromosomes during the first division of meiosis. Each pair of homologous chromosomes lines up at the metaphase plate independently of every other pair. Both maternal chromosomes may be on the same side of the metaphase plate, with the two paternal chromosomes on the other side, or one maternal chromosome may be on the same side of the metaphase plate as the paternal member of the other chromosome pair, and vice versa. The result is that the segregation of the alleles of a gene on one chromosome occurs independently of the segregation of the alleles of any gene on a nonhomologous chromosome.

5.3 Linkage, Crossing-over, and Genetic Maps of Chromosomes

Because organisms contain far more genes than chromosomes, not all pairs of alleles can assort independently of one another. Humans, for example, contain 23 pairs of chromosomes and approximately 35 000 genes. Thus, each chromosome contains many genes. Genes that are relatively close together on the same chromosome tend to stay together; if less than 50% of the gametes produced during meiosis in a double heterozygote have a recombinant combination of the alleles (see Section 5.2), the genes are said to be *linked*. New combinations of the alleles of genes on the same chromosome are produced by *crossing-over* (breakage and exchange of parts) between homologous chromosomes during meiosis and mitosis (Fig. 20). During an early stage (prophase) of meiosis, the homologous chromosomes pair (a process called *synapsis*) side by side. This pairing facilitates crossing over, which occurs by cutting DNA strands with enzymes called endonucleases, exchanging single strands of DNA from homologous chromosomes with the aid of recombination proteins, and rejoining the strands of DNA in new combinations with the assistance of DNA polymerases and DNA ligases. DNA ligases seal single-strand breaks in DNA molecules; they play important roles in replication, recombination, and DNA repair processes.

The frequency of crossing over and thus the frequency at which recombinant combinations of linked genes occur are approximately proportional to the distance between genes on the chromosome.

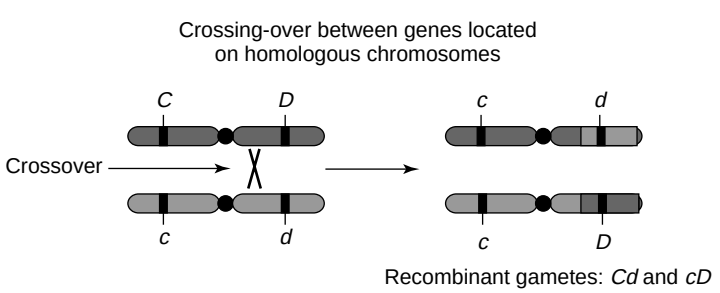

Fig. 20 The generation of new combinations of genes by crossing-over between genes located on homologous chromosomes during meiosis or mitosis.

This relationship is used to prepare *genetic maps* of chromosomes, which show the linear order and the relative positions of genes on a chromosome. Crosses are performed between organisms with different combinations of the alleles of genes on a specific chromosome, and the frequency of gametes carrying recombinant combinations of the alleles of the genes is determined. As an interesting historical sidelight, this genetic mapping procedure was developed by Alfred H. Sturtevant while he was still an undergraduate working in Thomas Hunt Morgan's laboratory at Columbia University in 1911. When genes are relatively far apart on a chromosome, multiple crossovers may occur between them. When two crossovers (or an even number) occur between two genes, the second cancels out the first, and all the resulting gametes will have parental combinations of the alleles of the genes. Thus, map distances are based on the average number of crossovers between genes. One map unit (one centiMorgan, cM) is the genetic distance in which an average of 1 chromatid out of a 100 will have undergone a crossover event. For distances less than 20 map units, the map distance is approximately equal to the proportion of recombinant chromosomes produced in percent, or 1 cM equals the distance yielding 1% recombinant gametes. Figure 21 contains an abbreviated genetic map of the X chromosome in *D. melanogaster*.

5.4 Transposable Genetic Elements

Living organisms contain remarkable DNA elements that can jump from one site in the genome to another site. These mobile elements are called *transposable genetic elements* or *transposons*. The insertion of one of these transposons into a gene will often render the gene nonfunctional, producing a mutant allele. Indeed, the wrinkled allele that Mendel studied in the garden pea was caused by the insertion of a transposable element.

Transposable elements are present in both prokaryotes and eukaryotes and

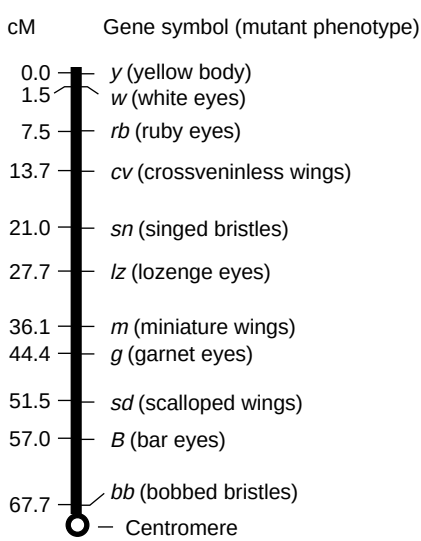

Fig. 21 Abbreviated genetic map of the X chromosome of *Drosophila*. Distances are given in centiMorgans (approximately equal to the percentage of recombinant chromosomes).

exhibit a variety of structures and mechanisms of transposition. However, most of them can be grouped into one of three classes. *Cut-and-paste transposons* move by excision from one site and insertion into another site. *Replicative transposons* move during replication, with the old copy remaining at the initial site and the nascent copy being inserted at a new site. Thus, the copy number increases with each transposition. *Retrotransposons* move via the synthesis of DNA copies of the RNA transcribed from the element ("retro" for reverse flow of genetic information, RNA → DNA, rather than DNA → RNA as in transcription). Some of the retrotransposons are closely related to retroviruses, or RNA tumor viruses, that use this RNA → DNA lifestyle.

Transposable genetic elements are major components of genomes, making up more than 40% of the DNA in the human genome and perhaps as much as 80% of the maize genome. They are responsible for a large number of spontaneously occurring mutations and play important roles in genome evolution. Crossing-over between two copies of a transposon in a chromosome will produce deletions or inversions depending on their orientations with respect to each other. Two transposons can also move the segment of DNA between them from one location to another if they transpose simultaneously. Indeed, transposable elements are in part responsible for the rapid spread of antibiotic and drug resistance in bacteria, a phenomenon of major concern to the medical community.

5.5 Physical Maps of Chromosomes

Whereas genetic maps are based on recombination frequencies, *physical maps* of chromosomes are based on the molecular distances – base pairs (bp), kilobases (kb, 1000 bp), and megabases (Mb, 1-million bp) – separating genes or molecular markers. Physical distances do not correlate directly with genetic map distances because recombination frequencies are not always proportional to molecular distances. However, the two are often reasonably well correlated in euchromatic regions of chromosomes. In humans, 1 cM is equivalent, on average, to about 1 Mb of DNA.

When mutations change the nucleotide sequences in restriction enzyme cleavage sites, the enzymes no longer recognize them. Other mutations may create new restriction sites. These mutations result in variations in the lengths of the DNA fragments produced by digestion with various restriction enzymes. Such *restriction fragment length polymorphisms*, or *RFLPs*, have proven invaluable in constructing detailed genetic and physical maps of chromosomes. The RFLPs can be mapped just like other genetic markers; they segregate in crosses as though they are codominant alleles. Correlations between genetic and physical maps can be established by positioning clones of genetically mapped genes or RFLPs on the physical map. Markers that are mapped both genetically and physically are called *anchor markers*. They anchor the physical map to the genetic map and vice versa. *Sequence-tagged sites* (*STSs*) – unique genomic DNA sequences 200 to 500 base pairs long – and *expressed-sequence tags* (*ESTs*) – DNA copies of mRNAs – are often used as hybridization probes to anchor physical maps to genetic maps.

In humans, the most useful RFLPs involve short sequences that are present as tandem repeats. The number of copies of each sequence present at a given site on a chromosome is highly variable. These sites, called *variable number*

tandem repeats (*VNTRs*) are highly polymorphic. VNTRs vary in fragment length because of differences in the number of copies of the repeated sequence between the restriction sites. *Microsatellites* are another class of polymorphisms that have proven extremely valuable in constructing high-density maps of eukaryotic chromosomes. Microsatellites are polymorphic tandem repeats of sequences only two to five nucleotide pairs long. They are called microsatellites because they are a subset of the satellite sequences present in the highly repetitive DNA of eukaryotes. When two or more segments of a chromosome have been identified, mapped physically, and shown to overlap, they are said to form a *contig*, and their combined physical maps form a *contig map*.

5.6 Map-position-based Identification of Genes

The availability of detailed genetic and physical maps of chromosomes permits researchers to isolate genes on the basis of their location in the genome. This approach, called *positional cloning*, can be used to identify any gene regardless of its function. The steps in positional cloning are illustrated in Fig. 22. The gene is first mapped to a specific region of a chromosome by genetic crosses or, in the case of humans, by pedigree analysis. The gene is next localized on the physical map of this region of the chromosome.

Positional cloning involves "walking" or "jumping" along a chromosome until the desired gene is reached. Chromosome walks are initiated by the selection of a molecular marker (e.g. RFLP or gene clone) close to the gene and the use of this clone as a hybridization probe to screen a genomic library for overlapping sequences. Physical maps are constructed for the overlapping clones identified in the library screen, and the restriction fragment farthest from the original probe is used to screen a second genomic library constructed by using a different restriction enzyme or prepared from a partial digest of genomic DNA. Repeating this procedure several times and isolating a series of overlapping genomic clones allow a scientist to walk the required distance along a chromosome to the desired gene.

When the distance from the closest molecular marker to the gene of interest is large, a technique called chromosome jumping can be used to speed up an otherwise laborious walk. Each jump can cover a distance of 100 kb or more. Like a walk, a jump is initiated by using a molecular probe such as an RFLP as a starting point. However, with chromosome jumps, large DNA fragments are prepared by partial digestion of genomic DNA with a restriction endonuclease. The large genomic fragments are then circularized with DNA ligase. A second restriction endonuclease is used to excise the junction fragment from the circular molecule. This junction fragment will contain both ends of the long fragment; it can be identified by hybridizing DNA fragments separated by gel electrophoresis to the initial molecular probe. A restriction map of the junction fragment is prepared, and a restriction fragment that corresponds to the distal end of the long genomic fragment is cloned and used to initiate a chromosome walk or a second chromosome jump. Chromosome jumping has proven especially useful in work with large genomes such as the human genome.

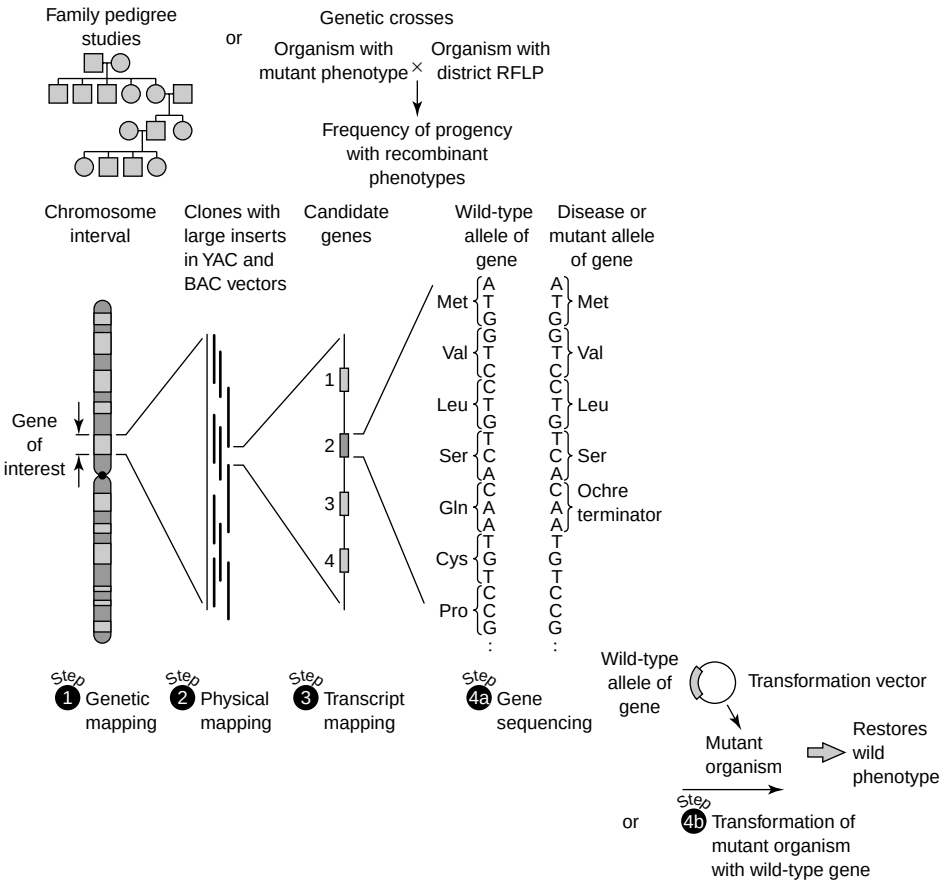

Fig. 22 Steps involved in cloning of genes on the basis of their location in the genome (positional cloning). The gene is first mapped to as small a region of a chromosome as possible. In humans, mapping must be done using family pedigree data, and candidate genes must be screened by comparing the sequences of genes from "normal" individuals with the sequences from individuals with the trait or genetic disorder. In other species, mapping is done using data from genetic crosses, and candidate genes are screened by transforming mutant organisms with the wild-type allele of the gene and seeing whether it restores the wild-type phenotype. (Reproduced from Snustad, D.P., Simmons, M.J., *Principles of Genetics*, 3rd edition, Copyright © 2003 by John Wiley & Sons, Inc., Hoboken, NJ, USA. This material is used by permission of John Wiley & Sons, Inc.)

Verification that a clone of the desired gene has been isolated is accomplished in various ways. In organisms such as *Drosophila*, verification is achieved by introducing the wild-type allele of the gene into a mutant organism and showing that it restores the wild-type phenotype. In humans, verification commonly involves determining the nucleotide sequences of the wild-type gene and several mutant alleles and showing that the coding sequences of the mutant genes are defective and unable to produce functional gene products.

Acknowledgements

Figures 3, 6, 7, 10, 11, 12, 17, and 22 have been reproduced from Snustad, D.P., Simmons, M.J., *Principles of Genetics*, 3rd edition, Copyright © 2003 by John Wiley & Sons, Inc., Hoboken, NJ, USA. This material is used by permission of John Wiley & Sons, Inc.

See also Alternatively Spliced Genes; Brain Development.

Bibliography

Books and Reviews

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2002) *Molecular Biology of the Cell*, 4th edition, Garland Publishing, New York.

Bateson, W. (1909) *Mendel's Principles of Heredity*, Cambridge University Press, Cambridge, UK.

Beckwith, J.R., Zipser, D. (1970) *The Lactose Operon*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Bell, S.P., Dutta, A. (2002) DNA replication in eukaryotic cells, *Annu. Rev. Biochem.* **71**, 333–374.

Benkovic, S.J., Valentine, A.M., Salinas, F. (2001) Replisome-mediated DNA replication, *Annu. Rev. Biochem.* **70**, 181–208.

Carlson, E.A. (1966) *The Gene: A Critical History*, W. B. Saunders, Philadelphia, PA.

Crow, J.F. (2000) Two centuries of genetics: a view from half-time, *Annu. Rev. Genomics Hum. Genet.* **1**, 21–40.

Crow, J.F., Dove, W.F. (Eds.) (2000) *Perspectives on Genetics, Anecdotal, Historical, and Critical Commentaries, 1987–1998*, University of Wisconsin Press, Madison, WI.

Davies, K. (2001) *Cracking the Genome, Inside the Race to Unlock Human DNA*, Free Press, New York.

Drake, J.W. (1970) *The Molecular Basis of Mutation*, Holden-Day, San Francisco, CA.

Echols, H. (2001) *Operators and Promoters: The Story of Molecular Biology and Its Creators*, University of California Press, Berkeley, CA.

Karp, G. (2001) *Cell and Molecular Biology: Concepts and Experiments*, 3rd edition, John Wiley & Sons, New York.

Keegan, L.P., Gallo, A., O'Connell, M.A. (2001) The many roles of an RNA editor, *Nat. Rev. Genet.* **2**, 869–878.

Kornberg, A., Baker, T.A. (1992) *DNA Replication*, 2nd edition, W. H. Freeman, San Francisco, CA.

Lee, T.I., Young, R.A. (2000) Transcription of eukaryotic protein-coding genes, *Annu. Rev. Genet.* **34**, 77–137.

Lewin, B. (2000) *Genes VII*, Oxford University Press, New York.

Lewis, R. (2003) *Human Genetics, Concepts and Applications*, 5th edition, WCB/McGraw-Hill, Dubuque, IA.

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., Darnell, J. (2000) *Molecular Cell Biology*, 4th edition, W. H. Freeman, San Francisco, CA.

Mount, D.W. (2001) *Bioinformatics, Sequence and Genome Analysis*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Neidhardt, F.C., Curtis, R. III, Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanek, B., Reznikoff, W.S., Riley, M., Schaechter, M., Umbarger, H.E. (Eds.) (1996) Escherichia coli *and* Salmonella typhimurium, *Cellular and Molecular Biology*, 2nd edition, ASM Press, Washington, DC.

Peters, J.A. (Ed.) (1959) *Classic Papers in Genetics*, Prentice Hall, Englewood Cliffs, NJ.

Ptashne, M., Gann, A. (2002) *Genes & Signals*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Snustad, D.P., Simmons, M.J. (2003) *Principles of Genetics*, 3rd edition, John Wiley & Sons, Hoboken, NJ.

Stahl, F.W. (1979) *Genetic Recombination*, W. H. Freeman, San Francisco, CA.

Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A., Weiner, A.M. (1987) *Molecular Biology of the Gene*, 4th edition, Vols. I and II, Benjamin/Cummings, Menlo Park, CA.

Primary Literature

Adams, M.D., and 194 co-authors (2000) The genome sequence of *Drosophila melanogaster*, *Science* **287**, 2185–2195.

Avery, O.T., MacLeod, C.M., McCarty, M. (1944) Studies on the chemical nature of the substance inducing transformation in pneumococcal types, *J. Exp. Med.* **79**, 137–148.

Beadle, G.W., Tatum, E.L. (1941) Genetic control of biochemical reactions in neurospora, *Proc. Natl. Acad. Sci. U.S.A.* **27**, 499–506.

Benzer, S. (1961) On the topography of the genetic fine structure, *Proc. Natl. Acad. Sci. U.S.A.* **47**, 403–415.

Brenner, S., Jacob, F., Meselson, M. (1961) An unstable intermediate carrying information from genes to ribosomes for protein synthesis, *Nature* **190**, 576–581.

Brenner, S., Stretton, A.O.W., Kaplan, S. (1965) Genetic code: the 'nonsense' triplets for chain termination and their suppression, *Nature* **206**, 994–998.

Brody, S., Yanofsky, C. (1963) Suppressor gene alteration of protein primary structure, *Proc. Natl. Acad. Sci. U.S.A.* **50**, 9–16.

Cairns, J. (1963) The bacterial chromosome and its manner of replication as seen by autoradiography, *J. Mol. Biol.* **6**, 208–213.

Creighton, H.B., McClintock, B. (1931) A correlation of cytological and genetical crossing over in *Zea mays*, *Proc. Natl. Acad. Sci. U.S.A.* **17**, 492–497.

Crick, F.H.C., Barnett, L., Brenner, S., Watts-Tobin, R.J. (1962) General nature of the genetic code for proteins, *Nature* **192**, 1227–1232.

Delbrück, M., Bailey, W.T. Jr. (1946) Induced mutations in bacterial viruses, *Cold Spring Harbor Symp. Quant. Biol.* **11**, 33–37.

Fleischmann, R.D., and 39 co-authors (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd, *Science* **269**, 496–512.

Fraenkel-Conrat, H., Singer, B. (1957) Virus reconstitution II. Combination of protein and nucleic acid from different strains. *Biochem. Biophys. Acta* **24**, 540–548.

Fraser, A.G., Kamath, R.S., Zipperien, P., Martinez-Campos, M., Sohrmann, M., Ahringer, J. (2000) Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference, *Nature* **408**, 325–330.

Freese, E. (1959) The specific mutagenic effect of base analogues on phage T4, *J. Mol. Biol.* **1**, 87–105.

Garen, A., Garen, S. (1963) Complementation *in vivo* between structural mutants of alkaline phosphatase from *E. coli*, *J. Mol. Biol.* **7**, 13–22.

Gönczy, P., Echeverri, C., Oegema, K., Coulson, A., Jones, S.J.M., Copley, R.R., Duperon, J., Oegema, J., Brehm, M., Cassin, E., Hannak, E., Kirkham, M., Pichler, S., Flohrs, K., Goessen, A., Leidel, S., Alleaume, A.-M., Martin, C., Özlu, N., Bork, P., Hyman, A.A. (2000) Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III, *Nature* **408**, 331–336.

Hershey, A.D., Chase, M. (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage, *J. Gen. Physiol.* **36**, 39–56.

Huberman, J.A., Riggs, A.D. (1968) On the mechanism of DNA replication in mammalian chromosomes, *J. Mol. Biol.* **32**, 327–341.

International Human Genome Sequencing Consortium, Lander, E.S., and 248 co-authors (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Jacob, F., Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins, *J. Mol. Biol.* **3**, 318–356.

Kornberg, A. (1960) Biologic synthesis of deoxyribonucleic acid, *Science* **131**, 1503–1508.

Lederberg, J., Tatum, E.L. (1946) Novel genotypes in mixed culture of biochemical mutants of bacteria, *Cold Spring Harbor Symp. Quant. Biol.* **11**, 113, 114.

Lederberg, J. (1947) Gene recombination and linked segregations in *Escherichia coli*, *Genetics* **32**, 505–525.

McClintock, B. (1956) Controlling elements and the gene, *Cold Spring Harbor Symp. Quant. Biol.* **21**, 197–216.

McManus, M.T., Sharp, P.A. (2002) Gene Silencing in mammals by small interfering RNAs, *Nat. Rev. Genet.* **3**, 737–747.

Mendel, G. (1866) Experiments in Plant Hybridization, in: Peters, J.A. (Ed.) *Classic Papers in Genetics*, pp. 1–20, Prentice Hall, Englewood Cliffs, NJ. (Reprint of William Bateson's translation of Mendel's paper Versuche über Pflanzenhybriden. *Verh. Naturforsch. Ver Brünn* 4, 3–47.)

Meselson, M.S., Stahl, F.W. (1958) The replication of DNA in *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.* **44**, 671–682.

Muller, H.J. (1927) Artificial transmutation of the gene, *Science* **66**, 84–87.

Nirenberg, M.W., Matthaei, J.H. (1961) The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic

polyribonucleotides, *Proc. Natl. Acad. Sci. U.S.A.* **47**, 1588–1602.

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., Kainuma, R., Sugino, A., Iwatsuki, N. (1968) *In vivo* mechanism of DNA chain growth, *Cold Spring Harbor Symp. Quant. Biol.* **33**, 129–143.

Ptashne, M. (1967) Specific binding of the λ phage repressor to λ DNA, *Nature* **214**, 232–234.

Schlesinger, M.J., Levinthal, C. (1963) Hybrid protein formation of *E. coli* alkaline phosphatase leading to *in vitro* complementation, *J. Mol. Biol.* **7**, 1–12.

Speyer, J.F., Lengyel, P., Basilio, C., Ochoa, S. (1962) Synthetic polynucleotides and the amino acid code, IV, *Proc. Natl. Acad. Sci. U.S.A.* **48**, 441–448.

Sturtevant, A.H. (1913) The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association, *J. Exp. Zool.* **14**, 43–59.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, *Nature* **408**, 796–815.

The *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology, *Science* **282**, 2012–2018.

Taylor, J.H., Woods, P.S., Hughes, W.L. (1957) The organization and duplication of chromosomes as revealed by autoradiographic studies using tritium-labeled thymidine, *Proc. Natl. Acad. Sci. U.S.A.* **43**, 122–128.

Tettelin, H., and 38 co-authors (2001) Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*, *Science* **293**, 498–506.

Venter, J.C., and 272 co-authors (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Waterhouse, P.M., Helliwell, C.A. (2003) Exploring plant genomes by RNA-induced gene silencing, *Nat. Rev. Genet.* **4**, 29–38.

Watson, J.D., Crick, F.H.C. (1953) Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid, *Nature* **171**, 737, 738.

Watson, J.D., Crick, F.H.C. (1953) Genetical implications of the structure of deoxyribonucleic acid, *Nature* **171**, 964–967.

Yanofsky, C., Carlton, B.C., Guest, J.R., Helinski, D.R., Henning, U. (1964) On the colinearity of gene structure and protein structure, *Proc. Natl. Acad. Sci. U.S.A.* **51**, 266–272.

Genome Scanning Method; Restriction Landmark Genomic Scanning (RLGS)

Hisato Okuizumi[1], *Tomoki Matsuyama*[2], *and Yoshihide Hayashizaki*[2,3]
[1] *National Institute of Agrobiological Sciences, Ibaraki, Japan*
[2] *RIKEN Yokohama Institute, Kanagawa, Japan*
[3] *RIKEN Wako Main Campus, Saitama, Japan*

1 Concept of "Genome Scanning" 415

2 Principle of RLGS 416
2.1 Breakthroughs Enabling the Development of RLGS 416
2.2 The Principle and Procedure of RLGS 417

3 Techniques 419
3.1 Preparation of Genomic DNA 420
3.1.1 Materials and Stock Buffers 420
3.1.2 DNA Preparation 420
3.2 Blocking and Direct DNA Labeling at Restriction Landmark 420
3.2.1 Materials and Stock Buffers 420
3.2.2 Procedures of Blocking, Restriction Enzyme Digestions, and Labeling 421
3.3 Two-dimensional Separation with Electrophoresis 422
3.3.1 Materials, Apparatus, and Stock Solution 422
3.3.2 Gel Casting for Vertical 1st-D Disc Gel 423
3.3.3 Running 1st-D Gel 423
3.3.4 Extruding the Gel 424
3.3.5 *In Situ* Restriction Enzyme Digestion 426
3.3.6 Preparation of 2nd-D Gel 426
3.3.7 Running 2nd-D Gel 427
3.3.8 Gel Drying and Autoradiography 428

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

4 RLGS Pattern and Its Advantages 430

5 Screening for DNA Methylation Using RLGS-M 430

6 Virtual RLGS 432

7 Perspective 436

Bibliography 437
Books and Reviews 437
Primary Literature 437
URLs 439

Keywords

Genome Scanning
High-speed survey of the presence or absence of landmarks throughout a genome and the measurement of their copy number in each locus.

Restriction Landmark
Use of restriction enzyme sites as guideposts throughout the genome.

Blocking (in RLGS)
Reduction of background by interfering with the incorporation of labeling reagents into nonspecific-damaged sites of DNA with pretreatment using enzyme reaction(s).

The restriction landmark genomic scanning (RLGS) technology is a method for high-speed survey of the presence or absence of restriction landmarks throughout a genome and the measurement of their copy number based on the new concept that restriction enzyme sites are used as landmarks. RLGS employs (1) direct end-labeling of the genomic DNA digested with a restriction enzyme and (2) high-resolution, two-dimensional electrophoresis. RLGS has the following advantages. (1) High speed–scanning ability. Thousands of restriction landmarks can be scanned simultaneously. (2) Scanning field can be extended by the use of different kinds of landmarks in an additional series of electrophoresis; (3) This method can be applied to any organism because direct labeling of restriction enzyme sites, and not the hybridization procedure, is employed as a detection system. (4) Spot intensity reflects the copy number of the restriction landmark. Thus, haploid and diploid genomic DNAs can be distinguished. (5) Using a methylation-sensitive enzyme, the methylated state of genomic DNA can be screened. Thus, RLGS is a very useful system not only for genome mapping but also for various studies in biological and medical fields, such as research for cancer development, aging, and so on.

1
Concept of "Genome Scanning"

Genome scanning is defined as high-speed survey of the presence or absence of landmarks throughout a genome and measurement of their copy number in each locus. Originally, the concept of "genome scanning" arose from the idea of overall detection of the physical condition of the whole genomic DNA. From this standpoint, it would be simplest to detect all fragments of genomic DNA generated by restriction enzyme cleavage after electrophoresis followed by ethidium bromide staining. Initially, efforts were made to visualize whole genomic DNA fragments according to this approach using the *Escherichia coli* genome. However, this approach was limited to small-sized genomes. As the complexity of the genome increases, the copy number of DNA molecules of the haploid genome equivalent decreases in the specific amount of genomic DNA and it becomes more difficult to separate and detect the DNA fragments. In the case of the human genome, which is 3×10^9 bp, a single copy locus per haploid genome of 1 μg human genomic DNA produces only 0.5 attomoles (5×10^{-19} mol) fragments.

To overcome this problem, scanning only landmark information was tried, ignoring the other genomic DNA regions. A landmark would be a guidepost on the genome, which can be visualized as signals through the assay method (scanning method), representing a one-to-one correspondence of a signal to its locus. The ideal landmark and scanning method should offer the following: (1) robust landmark signals that can be detected reproducibly, (2) high speed–scanning ability (multiplex scanning), (3) measurement of landmark copy number from its signal strength, (4) simple procedure and robotization, (5) expandability of scanning field, (6) applicability to any organism, (7) exportability of the information of landmark to any organism, and (8) exportability of the technique and the information to other laboratories.

As the most classic method, Southern hybridization initially provided a quantitative and qualitative method for visualizing landmarks directly from genomic DNA using a probe with a unique sequence. In this system, a signal specific to each locus can be detected because characteristic molecule recognition of DNA hybridization occurs in a sequence-specific manner. Next, a polymerase chain reaction (PCR) was developed as an alternative method for detecting landmarks. PCR is a technique for amplifying a region of several hundred base pairs, flanked by a pair of oligonucleotide primers. In the PCR technique, the specificity of signals depends on the specificity of hybridization of the primers to a unique genomic locus.

To facilitate the screening of genome DNA, two approaches can be considered for identifying landmarks on complex genomes such as mammals. One approach is the simplification of the procedure for detecting loci by establishing robotized systems. In Southern hybridization, using a unique probe and in PCR, for amplifying a unique sequence, a robust but a single landmark can be identified. To establish hybridization- or PCR-based genome scanning, the development of a robotized system is essential, because only one locus can be assayed in one procedure. The other approach is the development of a multiplex method by which multiple loci can be screened in one procedure. Typical techniques of hybridization-based and PCR-based multiplex methods are DNA fingerprinting using repeating sequences

(VNTR, IAP etc.) as probes and the Alu-PCR method respectively. Trials to identify mutant loci and construct genome mapping have proved successful in some cases using these conventional approaches based on the concept of genome scanning. However, the type of available repeating sequences limits the scanning field and scanning speed, because the basic principle of both approaches to produce signal specificity is sequence-to-sequence hybridization.

Our group has introduced a new concept termed *restriction landmark*, in which each restriction enzyme recognition site can be used as a landmark. On the basis of this concept, we developed a restriction landmark genomic scanning (RLGS) method, which employs (1) direct end-labeling of the genomic DNA digested with a restriction enzyme and (2) high-resolution, two-dimensional electrophoresis. This method provides an alternative multiplex approach to genome analysis. Because of the strict specificity of sequence recognition of the restriction enzyme, this technique enables us to scan multiple and robust landmarks. Here, we present the concept and method of RLGS, which allows restriction landmarks to be surveyed throughout a genome.

In 2003, the completion of human genome sequence was announced. The draft genome sequences of many other model organisms have also been reported. The database provides us with a substantial amount of information that can influence medicine, drug design, and industrial and agricultural production. The genome sequence is just a part of genetic information, however, because DNA modification, mainly methylation, is essential genetic and epigenetic information.

2 Principle of RLGS

2.1 Breakthroughs Enabling the Development of RLGS

Direct end-labeling followed by electrophoretic separation had been very difficult with genomic DNA of high complexity, such as those of mammals. However, three significant breakthroughs paved the way for the development of RLGS. First was the discovery of effective enzymes. Conventionally used 6-bp cutters produce too many DNA fragments that cannot be separated as discrete signals. Various 8-bp cutters or rare cutters were discovered, which could produce an appropriate number of DNA fragments for analysis. Second was the establishment of a high-resolution electrophoretic system. Methods of fine disk agarose gel electrophoreses for the first dimension (1st-D) have been developed to enable high-resolution electrophoretic separation. This enabled us to subject the 1st-D agarose samples to high-resolution polyacrylamide gel electrophoresis in the second dimension (2nd-D), because the thickness of the second polyacrylamide gel depends on the thickness of the first agarose gel. Third was the development of the technique for blocking nonspecific radioactive incorporation. As the genome size increases, the copy number of DNA fragments generated from the specific amount of DNA decreases. However, the background radioactivity produced by the incorporation into nonspecific-damaged sites of genomic DNA does not change. Therefore, the signal/noise ratio decreases, depending on the genome size. Generally, when the genome size increases to more than 1×10^8 bp, it becomes very difficult to

detect the signals, because they become buried in the background. To overcome this problem, we developed a blocking technique to reduce the nonspecific incorporation of radioactivity. The principle of the blocking technique is based on the incorporation of the dideoxynucleotide analog into the damaged sites, such as nicks, gaps, and/or double-strand breaks. After the blocking reaction, the radioactive nucleotides cannot be elongated by the polymerase reaction from the sites in which these nucleotide analogs are incorporated. Moreover, if a sulfur atom of the analog is connected to the phosphorus atom at a position in place of an oxygen atom in the molecules, such as dideoxy-[α-thio] nucleotide analogs, it is very difficult to eliminate it by the 3′ exonuclease activity of polymerase in the labeling step (Fig. 1).

2.2 The Principle and Procedure of RLGS

RLGS is a technique for scanning restriction landmarks at intervals of less than 1 Mbp. As mentioned above, the random cleavage caused during the preparation of sample DNAs affects the RLGS pattern. To produce a clear pattern with low background, DNA of high quality is essential. The critical factors are not only the average length and the extent of damage (nicks and gaps) to it but also the content of the DNA population, which is less than 10 kbp in length, because such small DNA fragments increase the background. In the DNA preparation procedure, more care should be taken to avoid enzymatic degradation (with endonuclease leaking from the lysosome and other enzymes in the lysis step), rather than mechanical shearing. Therefore, when DNA is prepared by proteinase K and the phenol extraction method, the samples should be treated for less than 10 min, with an excess of Proteinase K.

Figure 2 shows the entire procedure of the RLGS method. In comparison with conventional DNA electrophoresis, this procedure that consists of eight steps is specially designed to reduce the background generated by the incorporation of radioactivity in the blocking step (a) and to separate a large number of DNA fragments very precisely (high-resolution, two-dimensional electrophoresis, (b–f)).

(a) *Blocking*: The principle of the blocking procedure is mentioned above. This step is done to reduce the background generated by the incorporation of radioactivity into the nonspecifically damaged sites.

(b) *Landmark cleavage (with restriction enzyme A)*: This step is for cleavage of genomic DNA at the restriction landmarks. The type of restriction enzymes used should be chosen by the frequency of sites in the genome and the sensitivity of DNA methylation. The restriction enzyme B should be selected on the basis of the frequency of the restriction sites that have an appropriate number of spots that appear on the RLGS pattern.

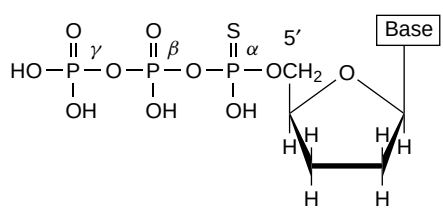

Fig. 1 Dideoxynucleotide analog for blocking method (dideoxynucleoside [α-thio] triphosphate).

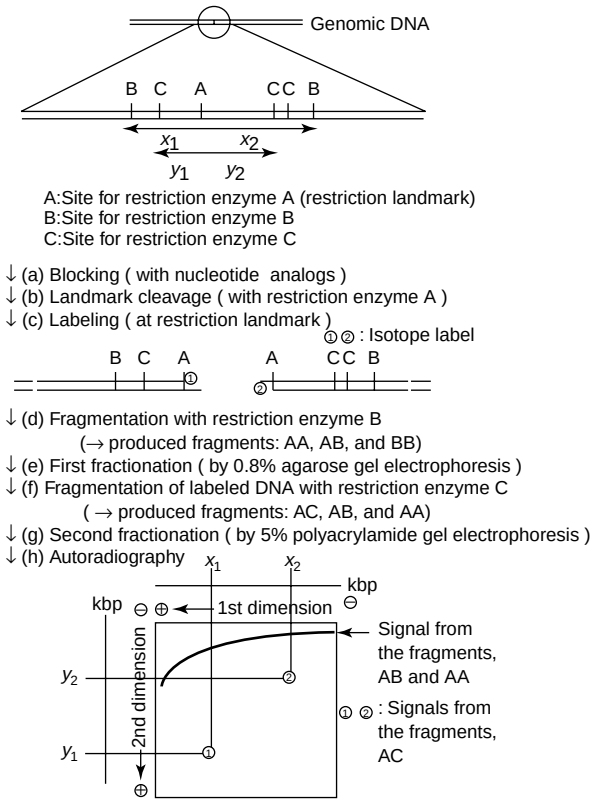

Fig. 2 Procedure for genome scanning by two-dimensional gel electrophoresis. x_1 and x_2 represent the distance from a restriction landmark to the neighboring site for restriction enzyme A or B. y_1 or y_2 represent the distance from the site of A to C. 1 and 2 indicate the labeled sites at the end of the landmark.

The number of analyzable spots is limited to less than 3000 spots in one gel by the resolution of the present technique. DNA methylation affects the RLGS pattern when methylation-sensitive enzyme is used as restriction enzyme A. In other words, the change of DNA methylation throughout the genome can be scanned, using a methylation-sensitive enzyme. Further discussion is presented below (see Sect. 5).

(c) *Labeling*: The labeling method depends on the shape of the end of the restriction fragment. In the case of the 5′-protruding end, a filling reaction of [α-^{32}P] deoxynucleotide with DNA polymerase is used. And for labeling at the 3′-protruding ends (or blunt end), a reaction of addition of [α-^{32}P] dideoxynucleotide with terminal deoxynucleotidyl transferase (terminal transferase) at the 3′ end of the DNA fragments is used (Hayashizaki, Y. et al. (1995) VCH Publishers, pp. 813–817).

(d) *Fragmentation of labeled DNA with restriction enzyme B*: This step is for further cleavage of the labeled genomic

DNA at the restriction landmark sites in a test tube. Because the total length of human or other mammalian genomic DNA is approximately 3×10^9 bp, more than 3000 DNA fragments with an average length of 1 Mbp will be produced by digestion with a restriction enzyme, even if *Not*I, the most infrequent 8-bp recognition enzyme, is used. However, such long DNA fragments cannot be prepared from DNA resources, such as cancer tissues, and, even if these DNA fragments were to be separated, pulse-field gel electrophoresis is used although its resolution is not enough to analyze so many DNA fragments. To solve this problem, we employed the digestion reaction of labeled fragments with restriction enzyme B, which appears more frequently in the genome than restriction landmarks. Owing to reduction in the length of the DNA fragments, the high-resolution disc agarose gel can be used, in which the entire resolution range is expanded to more than 40 cm.

In the case in which restriction enzyme A is used and the number of restriction sites are more than approximately 6000 in the genome (average length of this enzyme is less than 500 kbp), the treatment of restriction enzyme B is sometimes skipped. At the present level of the technique, too many such spots cannot be separated in one gel. However, if the fragmentation with restriction enzyme B is omitted, the average length of DNA fragments of almost all 8-bp cutters and rare cutters will be more than 30 kbp and these fragments will be trapped at the top of the agarose gel, resulting in reduction of the spot number. The final RLGS profile is the screening of an appropriate number of spots dispersed over most of a two-dimensional field.

(e) *First fractionation by agarose gel electrophoresis*: To obtain a good RLGS profile in the 1st-D agarose gel electrophoresis, the gel should be as thin and fine as possible, and the samples should be electrophoresed as long as possible to precisely separate more than 3000 signals. The prototype procedure using a horizontal gel of this step is described in other literature.

(f) *Fragmentation of labeled DNA with restriction enzyme C*: To resolve so many signals, these fragments should be subjected to further separation. The agarose disc gel should be treated with reaction mixture of restriction enzyme C. Usually, an enzyme whose restriction sites are more frequent than those of restriction enzyme B, such as a 4-bp cutter, should be used as the restriction enzyme C. Owing to this cleavage reaction, the DNA fragments can be electrophoresed depending on the distance from the site of restriction enzyme A to C (γ_1 and γ_2 in Fig. 2).

(g) *Second fractionation*: After the connection of the 1st-D agarose gel to the 2nd-D polyacrylamide gel, the DNA fragments are subjected to further 2nd-D gel electrophoresis. The sizes of the DNA fragments range from 2 kbp to 70 bp. Though technically vertical gels are used, original protocol has been described as the horizontal electrophoresis system.

(h) *Autoradiography*: The final gel samples are dried and autoradiographed.

3 Techniques

In this section, we describe the most recent protocol, which uses a vertical electrophoresis system. Use of an alternative

horizontal electrophoresis system has been described in other literature.

3.1 Preparation of Genomic DNA

3.1.1 Materials and Stock Buffers

Materials Proteinase K (Nakalai), RNase A (Nakalai), phenol (the commercially available phenol should be distilled), chloroform, isoamyl alcohol, sodium sarkosylate, 8-hydroxyquinoline, 50-mL Falcon tube, mortar and pestle, and dialysis tube.

Stock buffers

- TE: 10 mM Tris-Cl pH 7.5, 1 mM EDTA.
- PCI (phenol-chloroform-isoamyl-0.1% 8-hydroxyquinoline): Distilled phenol, chloroform, and isoamyl alcohol are mixed in a ratio of 50 : 49 : 1, and 8-hydroxyquinoline is adjusted to 0.1% in the final mixed reagent. PCI should be buffered with TE.
- Lysis buffer: 1 mg mL^{-1} Proteinase K, 150 mM EDTA, 1% sarkosyl, and 10 mM Tris-Cl pH 8.0.

3.1.2 DNA Preparation

A sample of 0.5 g of tissue wrapped in aluminum foil is frozen with liquid nitrogen. The sample is crushed with a hammer, and then powdered in a mortar that has been chilled in liquid nitrogen. For the study of cell lines, a cell pellet washed with phosphate-buffered saline (PBS) can be used in place of the powdered tissue. The sample is transferred into a 50-mL Falcon tube, followed by the addition of 20 mL lysis buffer. Stirring with a spatula gives a viscous sample. The mixture is incubated at 55 °C for 5 min and then 20 mL of PCI is added immediately, followed by gentle stirring until the border between the aqueous and phenol phases disappears. The sample is centrifuged at 3000 rpm for 30 min with RL101 swing rotor (Tomy). The aqueous phase is transferred into a new 50-mL Falcon tube. This phenol extraction should be repeated more than twice. The sample is dialyzed overnight against TE that has been autoclaved. After recovery from dialysis tube, 20 μL of 1 mg mL^{-1} boiled RNase A is added to digest RNA and is incubated for 2 h at 37 °C. The elimination of RNA is confirmed by 0.8% agarose electrophoresis using 1 μL of the reaction mixture. Next, PCI extraction is performed to inactivate RNase A. Ethanol, 40 mL, is added with gentle stirring. The precipitate is transferred to an Eppendorf sample tube and suspended in 500 μL of TE.

3.2 Blocking and Direct DNA Labeling at Restriction Landmark

3.2.1 Materials and Stock Buffers

Materials DNA polymerase I (Toyobo), deoxynucleoside [α-thio] triphosphates (dGTP[α]S, dATP[α]S, dCTP[α]S, dTTP[α]S, (Toyobo)), dideoxynucleoside [α-thio] triphosphates (ddGTP[α]S, ddATP[α]S, ddCTP[α]S, ddTTP[α]S, (Toyobo)), dideoxynucleoside triphosphates (ddGTP, ddATP, ddCTP, ddTTP, (Toyobo)), [α-^{32}P] dGTP (3000 Ci $mmol^{-1}$) (Amersham, UK), [α-^{32}P] dCTP (6000 Ci $mmol^{-1}$) (Amersham), *Not*I (Takara), *Eco*RV (Toyobo), *Mbo*I (Takara), Sequenase Ver. 2.0 (Amersham), terminal deoxynucleotidyl transferase (Takara), bromophenol blue (BPB), xylene cyanol (XC), and dithiothreitol (DTT).

Stock buffers

- 10 × high buffer (HB): 500 mM Tris-Cl (pH 7.4), 100 mM MgCl2, 1 M NaCl, and 100 mM DTT.

- 10 × medium buffer (MB): 500 mM Tris-Cl (pH 7.4), 100 mM MgCl2, 500 mM NaCl, and 100 mM DTT.
- 4 × blocking buffer (BB): 4 × MB, 1.6 μM dGTP[αS], 0.8 μM dCTP[αS], 16.0 μM ddTTP[αS] (or ddTTP), and 16.0 μM ddATP[αS] (or ddATP).
- TE buffer: 10 mM Tris-HCl (pH 7.4), 1 mM EDTA.
- PCI; refer to Sect. 3.1.1.

3.2.2 Procedures of Blocking, Restriction Enzyme Digestions, and Labeling

Blocking The genomic DNA is treated with 1.75 units of DNA polymerase I in 10 μL of 1 × high buffer in the presence of 0.4 μM dGTP[αS], 0.2 μM dCTP[αS], 4.0 μM ddTTP[αS] (or ddTTP), and 4.0 μM ddATP[αS] (or ddATP). ddTTP[αS] and ddATP[αS] (Fig. 1) rather than ddTTP and ddATP are recommended to obtain a good pattern. This combination of nucleotide analog is applied to the labeling of *Not*I landmarks. In this blocking reaction, the deoxynucleoside [α-thio] triphosphate(s) (dXTP[αS]), which is incorporated into the site of landmark cleavage (dGTP[αS] and dCTP[αS] in this case), should be chosen as 0.4 μM or 0.2 μM of dXTP[αS] (the concentration depends on the specific activity of the radioactive compound used in the labeling step) and the dideoxynucleoside [α-thio] triphosphate(s) (ddYTP[αS]), which cannot be incorporated into the site of landmark cleavage (ddTTP[αS] and ddATP[αS]), should be chosen as 4.0 μM of ddYTP[αS]. This reaction mixture is incubated at 37 °C for 30 min. DNA polymerase I is inactivated by heating at 65 °C for 30 min.

Protocol 1

7.0	μL	sample DNA
2.5	μL	4 × blocking buffer
0.5	μL	DNA polymerase I [3.5U/L]
10	μL	total

↓ Mix with pipetting
↓ Incubate at 37 °C for 30 min
↓ Incubate at 65 °C for 30 min.

Landmark cleavage DNA is digested with enough units of landmark restriction enzyme A, such as *Not*I. The reaction mixture is incubated at 37 °C for 2 h.

Protocol 2

10.0	μL	sample DNA
2.0	μL	10 × HB
2.0	μL	10 × HB
2.0	μL	*Not*I [10 U μL^{-1}]
6.0	μL	H_2O
20.0	μL	total

↓ Mix with pipetting
↓ Incubate at 37 °C for 2 h.

Labeling The cleavage ends are filled in with 1.3 units of Sequenase Ver. 2.0 in the presence of 0.165 μM [α-^{32}P] dGTP (3000 Ci $mmol^{-1}$) and 0.165 μM [α-^{32}P] dCTP (6000 Ci $mmol^{-1}$) at 37 °C for 30 min. The polymerase reaction is stopped by heating up to 65 °C for 30 min.

Protocol 3

20.0	μL	sample
0.3	μL	1M DTT
1.0	μL	[α-^{32}P]dGTP
1.0	μL	[α-^{32}P]dCTP
0.1	μL	Sequenase Ver. 2.0 [13 U μL^{-1}]
22.4	μL	total

↓ Mix with pipetting
↓ Incubate at 37 °C for 30 min
↓ Incubate at 65 °C for 30 min.

Cleavage with restriction enzyme B The second digestion is performed with restriction enzyme B, such as *Eco*RV, in 30 μL of 1 × high buffer. The final reaction mixture is as shown in Protocol 4. The reaction mixture is incubated at 37 °C for 1 h.

Protocol 4

22.4 μL	sample DNA
1.0 μL	10 × HB
2.0 μL	*Eco*RV [10 U μL^{-1}]
4.6 μL	H_2O
30.0 μL	total

↓ Mix with pipetting
↓ Incubate at 37 °C for 1 h.

3.3 Two-dimensional Separation with Electrophoresis

3.3.1 Materials, Apparatus, and Stock Solution

Materials

- SeaKem GTG agarose (FMC); acrylamide (Nakalai); methylenebisacrylamide (Nakalai); ammonium persulfate (Nakalai); bovine serum albumin (BSA) (Sigma, fraction V); disposable plastic syringes (1 mL, 5 mL); needles, three-way stopcocks for syringes.
- 10 × 1st-D buffer (for 1st-D): Dissolve 242 g Tris, 109 g sodium acetate trihydrate, 42 g NaCl, 23.4 g EDTANa2 in 1.8 L of H_2O and adjust the pH to 8.15 with acetic acid. Adjust the volume to 2 L and autoclave it. The pH of the 1 × buffer is 8.0.
- 10× tris-borate, EDTA (TBE) buffer (for 2nd-D): Dissolve 1210 g Tris, 770 g boric acid, and 75 g EDTANa2 in H_2O and adjust to 20 L. Final concentration for electrophoresis is 1 × TBE.
- 6 × dye solution: 0.25% BPB, 0.25% × C in (30% glycerol) in H_2O.
- 10% ammonium persulfate (Nakalai).
- N,N,N′,N′-Tetramethylethylenediamine (TEMED) (Nakalai)

Gel

- 1st-D agarose gel: 0.8% agarose (SeaKem GTG) in 1 × 1st-D buffer with 5% sucrose. (Adjust 0.24 g agarose to 30 mL with 1 × 1st-D buffer containing 5% sucrose.)
- Connecting gel: 0.8% agarose (SeaKem GTG) in 1 × TBE with 5% sucrose. (Adjust 0.24 g agarose to 30 mL with 1 × TBE containing 0.2 mL 6 × dye without glycerol.)(note) Polyacrylamide gel should be prepared just before the use for 2nd-D gel electrophoresis.

Electrophoretic apparatus

- 3.0 mm i.d., 3.6 mm o.d., 10 m (for *in situ* digestion) (#9), Teflon tubing, (Nichias).
- 3 mm i.d., 5 mm o.d., 1 m (No. 7078BR), Silicon tubing (for 1st-D), (Sanplatec).
- 1st-D gel holder (composed of 4 mm i.d. and 60-cm long glass tube with a 2.4 mm i.d. and 63-cm long Teflon tubing inside. Silicon stopper #4 with a hole is set at 4 cm below the top edge).
- Electrophoretic tank for 1st-D (Fig. 3).
- Electrophoretic tank for 2nd-D (Figs. 4 and 5).

Note: 1st-D gel holder and electrophoretic tanks for 1st-D and 2nd-D are manufactured and supplied by Biocraft Inc., Tokyo, Japan: URL, http://www.biocraft.co.jp; E-mail, info@bio-craft.co.jp.

Preparation of the gel holder for the vertical 1st-D disc gel The prototype 1st-D disc gel method with Teflon tubing was originally developed by Asakawa. In this protocol, we

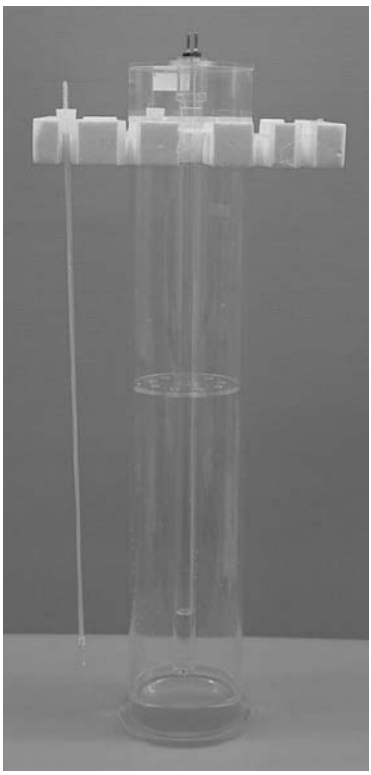

Fig. 3 1st-D electrophoretic tank with 1st-D gel holder on the hanging parts.

introduce an extensively modified Teflon tubing method that was adapted from Okazaki et al.

3.3.2 Gel Casting for Vertical 1st-D Disc Gel

- Melt the agarose gel solution in a microwave oven.
- Cool the solution down to 55–65 °C (Fig. 6).
- Connect a 5-mL plastic syringe fitted with a three-way stopcock to the top edge of the 1st-D gel holder with a 3-cm long silicon tubing.
- Suck up the gel solution (Fig. 6) gradually to reach the line (1 cm below the top).
- Close the stopcock and hold it on the support stand (Figs. 7 and 3).
- Wait for 15 min for the gel to solidify.
- Open the stopcock to the air and disconnect it.
- Add 350 mL of 1 × 1st-D buffer to the cathodal bottom tank.
- Fit each gel on the anodal top tank at the silicon stopper.
- Place the anodal tank on the cathodal tank and fill the anodal tank with 350 mL of 1st-D buffer.

3.3.3 Running 1st-D Gel

- Apply an aliquot of sample containing 1.5 μg DNA.
- Electrophoresis at 100 V for 1 h and 230 V for about 23 h until the BPB migrates 50 cm.

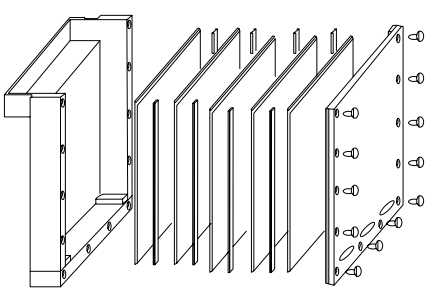

Fig. 4 Assembly of glass plates in the 2nd-D electrophoretic apparatus.

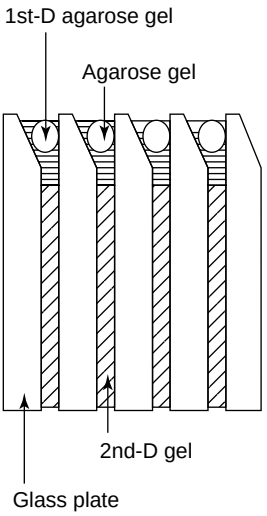

Fig. 5 Partial cross-section of a 2nd-D electrophoretic apparatus after the 1st-D gel connection.

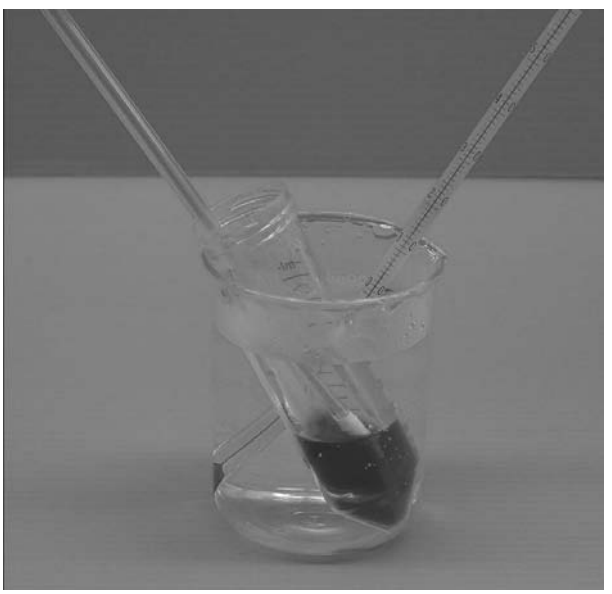

Fig. 6 Casting of 1st-D gel.

3.3.4 Extruding the Gel

- After electrophoresis, remove the cathodal top buffer with an aspirator and take out each gel holder.
- Expel the gel slowly using a cut-off yellow tip of a 1-mL syringe (Fig. 8).
- Cut off the gel 4 cm from the bottom edge (Fig. 9).
- Soak the gel in 50-mL Falcon tube containing 40 mL of 1 × HB (Fig. 10).
- Sway the tubes gently and equilibrate the buffer system of the gel for 10 min.

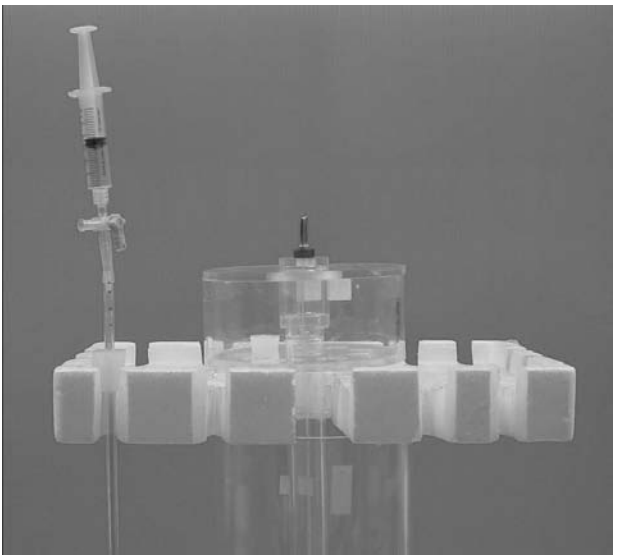

Fig. 7 Magnified view of the hanging parts and 1st-D gel holder.

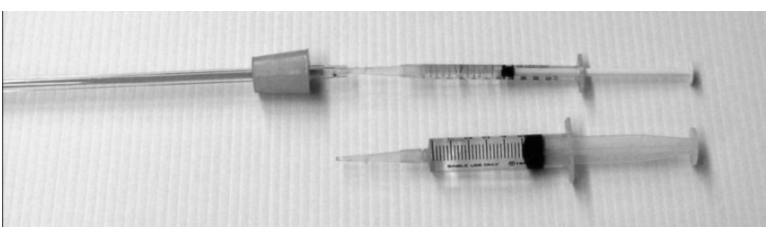

Fig. 8 Syringes to expel the 1st-D gel.

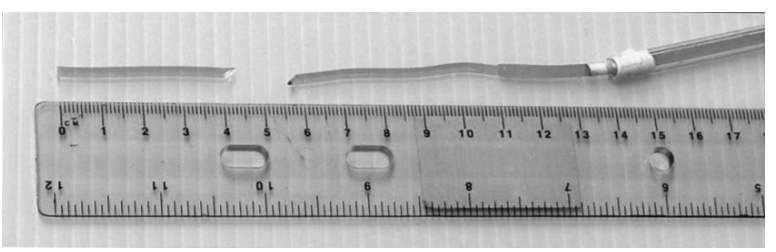

Fig. 9 Cutting the 1st-D gel.

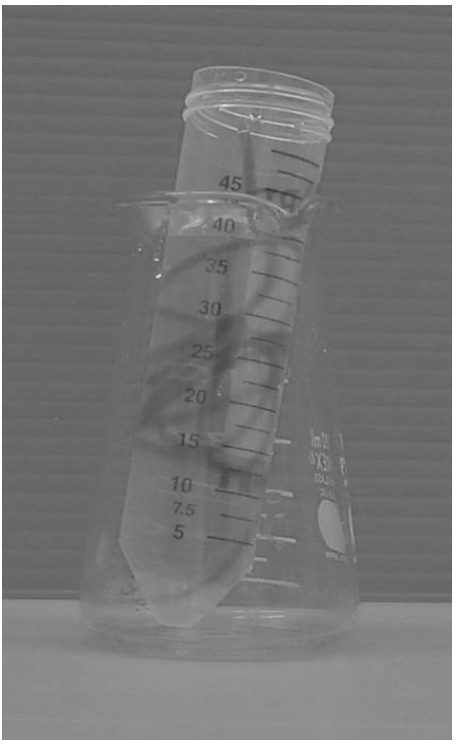

Fig. 10 Changing the buffer system for the 1st-D gel.

- Change the buffer again and equilibrate completely for 10 min.

3.3.5 *In Situ* Restriction Enzyme Digestion

Prepare Teflon tubing with an inside diameter that is slightly larger than that of the gel. Mount the 10-mL plastic syringe and connect it to the 3.0 mm i.d. and 45-cm long Teflon tubing (#9) with 3-cm silicon tubing.

- Pour the equilibrated gel into a tray containing 20 mL of 1 × HB (Fig. 11).
- Gently suck up the gel into the Teflon tubing (Fig. 11).
- Remove the buffer inside the Teflon tubing.
- Fill the tubing with 1.5 mL of enzyme solution, 1500 units of restriction enzyme (*Mbo*I), and 0.01% BSA by sucking up slowly (Fig. 12).
- Remove the syringe and make loop the tubing by connecting one end to the other.
- Put the looped tubing into a plastic bag (Fig. 13).
- Seal the bag and incubate at 37 °C for 2 h.

3.3.6 Preparation of 2nd-D Gel

Preparation of acrylamide of gel solution The polyacrylamide gel solution should be prepared just before the 2nd-D gel electrophoresis. Usually, 5% polyacrylamide gel is employed. Dissolve 145 g acrylamide, 5 g methylenebisacrylamide, 2-g ammonium persulfate in 3 L of 1× TBE buffer. Add 0.8 mL TEMED just before solidifying the gel.

2nd-D gel assembly Figures 4 and 5 show the recent type apparatus for the vertical 2nd-D electrophoretic system. This is composed of upper and lower electrophoretic tank and a box containing 5 glass plates with 1-mm thick spacers between the glass plates, which enables the preparation of 4 gels by pouring acrylamide solution into the box in one procedure. It is not necessary to reset the glass plates onto the electrophoretic apparatus. This apparatus can be also used as an electrophoretic tank itself.

The glass plates for 2nd-D are 480 (W) × 410 (H) × 5 (thick) mm in size. The gel is cast into a gel box carrying 4 gel spaces, which are made by 5 glass plates neatly placed side by side with 1-mm gel spacers in between. After the gel plates and spacers are assembled, acrylamide gel solution is poured. The gel surface is adjusted

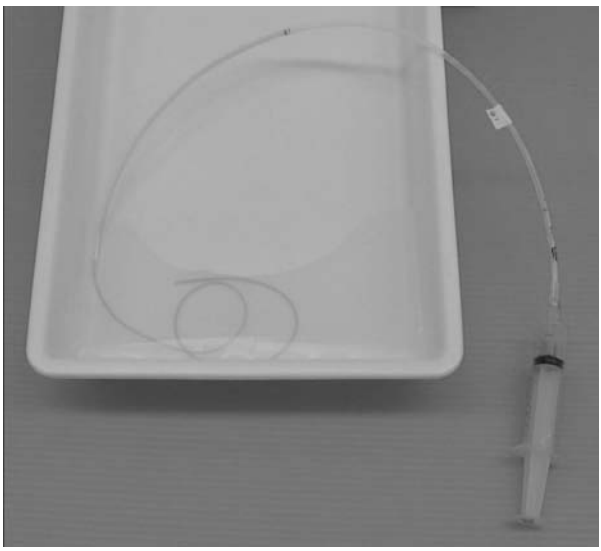

Fig. 11 Using a syringe to pull the 1st-D gel up into the Teflon tubing.

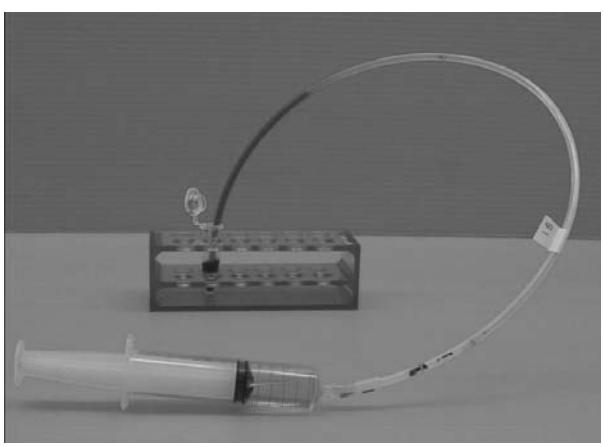

Fig. 12 Filling the tubing with enzyme C solution.

5 mm below the tapering edge of the glass plate. 1 mL of 2-butanol is laid gently on each acrylamide gel solution. Wait for 1 h for the gel to solidify.

3.3.7 Running 2nd-D Gel

- Rinse the top of the gel with 1 × TBE to remove 2-butanol and wipe the surface with a sheet of paper (Kimwipe) to remove the TBE completely.
- After 2 h, digestion with *Mbo*I expels the 1st-D gel from the tubing into a 50 mL-Falcon tube containing 40 mL of 1 × TBE.
- Shake the tubes gently and equilibrate the buffer system of the gel for 10 min.

Fig. 13 Tubing loop in the plastic bag.

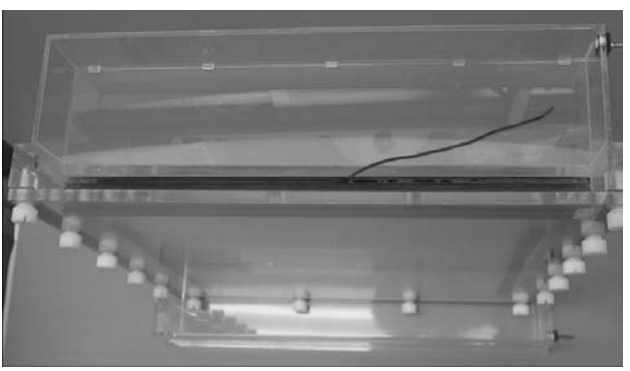

Fig. 14 Placing the 1st-D gel on the 2nd-D gel (top view).

- Discard the buffer and place the 1st-D gel on the top of 2nd-D acrylamide gel (Figs. 14 and 15).
- Connect 1st-D gel and 2nd-D gel with 2 mL of connecting gel using a 5-mL syringe with 22 g needle.
- Pour 0.2 mL of 6 × dye solution on the gel.
- Electrophorese at 100 V for 1 h and 150 V for about 23 h till the BPB reaches the bottom.

3.3.8 Gel Drying and Autoradiography

- Dry up the gel at 80 °C.
- Autoradiography using XAR5 (Kodak) in the cassette with intensified screen, Quanta Rapid™ (Sterling, see supplier, Knight Imaging Products at URL, http://www.ssl.gb.com/knight/intensifying_screens.htm).
- 24 h for short exposure.
- 7 to 10 days for long exposure (Fig. 16).

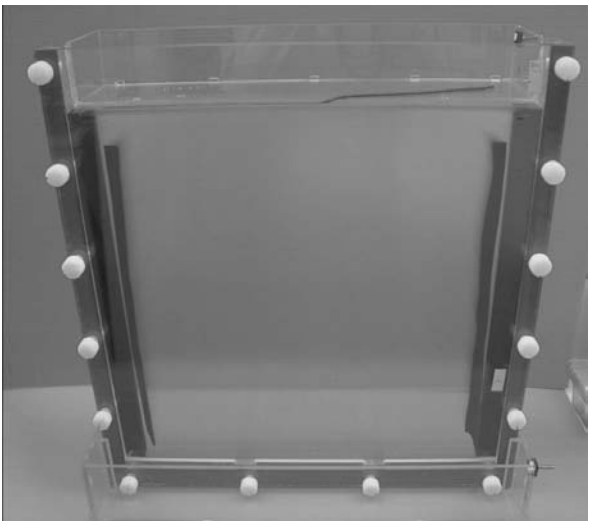

Fig. 15 Placing the 1st-D gel on the 2nd-D gel (side view).

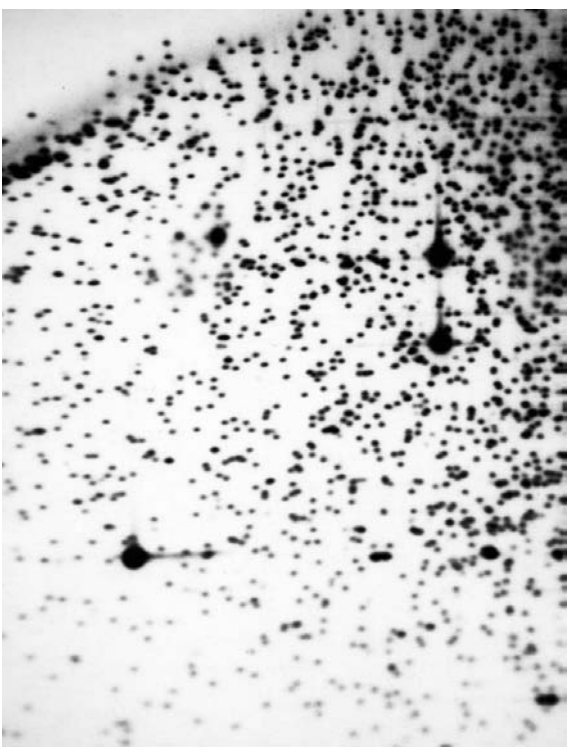

Fig. 16 RLGS profile of hamster genome DNA. *Not*I, *Eco*RV, and *Mbo*I were used as restriction enzymes A, B, and C respectively.

4 RLGS Pattern and Its Advantages

Figure 16 shows the resulting RLGS profile using *Not*I, *Eco*RV, and *Mbo*I as restriction enzyme A (restriction landmark), B, and C respectively. As shown in Fig. 2, the x, y coordinates $((x_1, y_1), (x_2, y_2))$ of each spot indicate the distances from the site of restriction enzyme A through B, and from A through C, respectively. In spite of the large size of mammalian or plant genomes, it is inconceivable that two sets of identical valued parameters (x, y) were to be produced from different landmarks on the genome. Therefore, one spot on the RLGS profile corresponds to one locus. In the mice RLGS profile, more than 2500 spots appeared in one gel, indicating that the information corresponds to more than 2500 times that from Southern hybridization or PCR with unique probes (primers). The high resolution of this system results from the two-dimensional system of combining agarose gel and polyacrylamide gel, and almost all the regions of the two-dimensional area are used efficiently to separate a large amount of spots.

RLGS has the following advantages for genome scanning. (1) It has high speed–scanning ability. Thousands of restriction landmarks can be scanned simultaneously. (2) The scanning field can be extended by using different kinds of landmarks in an additional electrophoresis series. More than 10 kinds of restriction enzymes can be used (Table 1). (3) This method can be applied to any organism because direct labeling of restriction enzyme sites, and not hybridization procedure, is employed as a detection system. (4) Spot intensity reflects the copy number of the restriction landmark. In Fig. 16, several enhanced spots can be seen. These spots are derived from ribosomal DNA, which is the typical repetitive sequence in mammalian genome. The fact that haploid and diploid genomic DNAs can be distinguished in RLGS will be used for linkage analysis. (5) Using methylation-sensitive enzyme, the methylated state of genomic DNA can be screened (see Sect. 5). (6) When CpG-rich enzymes, such as *Not*I, *Bss*HII and so on are used, CpG islands near the genes are preferentially screened. In the case of *Not*I, 89% of all *Not*I spots are located on or near the transcripts. Thus, RLGS is very advantageous compared with hybridization-based and PCR-based genome scanning methods.

Tab. 1 Landmark enzymes A (both vertebrate methylation-sensitive and -insensitive), B and C.

5MeCG-sensitive enzyme A
*Not*I, *Sfi*I, *Fse*I, *Csp*I, *Mlu*I, *Asc*I, *Bss*HII, *Cla*I, *Xho*I, *Eco*52I, *Sal*I, *Nar*I, *Nru*I, *Rsr*II
5MeCG-insensitive enzyme A
*Pac*I, *Swa*I, *Sse*8387I, *Pme*I, *Eag*I
Enzyme B (also used as enzyme C)
*Bam*HI, *Dra*I, *Eco*RI, *Eco*RV, *Hind*III, *Pst*I, *Pve*II, *Sph*I
Enzyme C
*Mbo*I, *Hin*fI

5 Screening for DNA Methylation Using RLGS-M

The RLGS method can be used in two ways. One is for genome scanning of the copy number of each locus corresponding to a spot using a methylation-insensitive enzyme as a restriction landmark. The other is for the detection of DNA methylation throughout the genome using a methylation-sensitive enzyme. Table 1

lists both types of restriction enzymes. Generally, the only known site for DNA methylation is the 5-position of the cytosine residue (5MeC), which is preferentially located on the palindromic dinucleotide 5′CG3′/5′GC3′. The 5MeCpG methylation-insensitive enzyme cleavage is not affected by DNA methylation in the vertebrate genome. The recognition sites of *Pac*I and *Swa*I do not include the C residue, and the site for Sse8387I also does not include the CG sequence, although the population of 5MeC located on CG sequence is not small. Therefore, when using these enzymes on the RLGS profile, the intensity of the spots depends entirely on the copy number of the corresponding locus.

As shown in Fig. 17 (RLGS-M, restriction landmark genomic scanning for DNA methylation), the intensity of the RLGS spots changes depending on DNA methylation when a 5MeCpG-sensitive enzyme is used as restriction enzyme A. If both alleles of the restriction landmark for a methylation-sensitive restriction enzyme, such as *Not*I, *Bss*HII and so on are demethylated, the corresponding spot displays full intensity. However, if one or both of the alleles are methylated, the spot intensity is reduced by half or disappears (Fig. 17). Thus, the degree of methylation can be screened by taking advantage of methylation-sensitive enzymes. Examples of this have been reported in other literature.

The RLGS method described here can be applied to many biological fields such as, cancer research (detection of amplification, loss of heterozygosity, methylation change), high-speed genome mapping (a large number of analyzable loci), isolation of imprinting genes, identification of individuals, and phylogenic analysis of species. Thus, the RLGS method is very useful for scanning not only the change of copy number of each locus but also the change in DNA methylation.

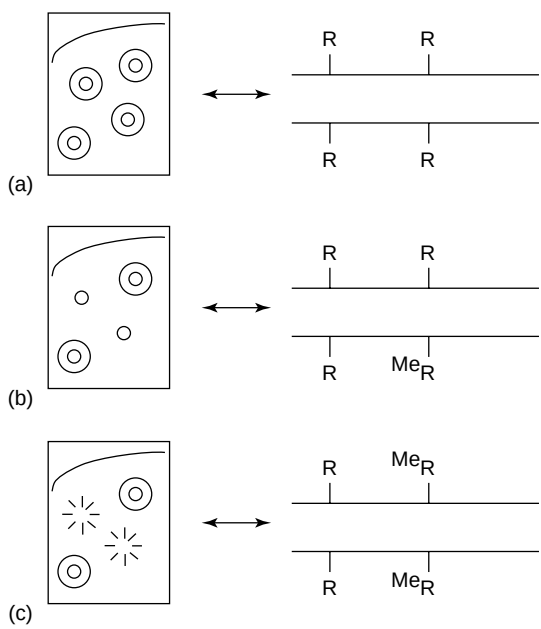

Fig. 17 Schematic representation of alteration in spot intensity due to genomic DNA methylation. (a) when both alleles are demethylated, the spots show diploid intensity; (b) when one allele is methylated, the corresponding spot shows half intensity; (c) when both alleles are methylated, the spots disappear. R, restriction landmark; MeR, the methylated restriction site.

6
Virtual RLGS

To develop analyses using RLGS profiles, the most important step is the identification of polymorphic spots. However, the amount of DNA available for cloning in a single spot is very small. In addition, the isolated-out acrylamide gel fragment contains two thousand times as many nonlabeled DNA fragments as the labeled DNA fragments. Therefore, it sometimes has the difficulty to clone the target DNA fragment. To overcome this problem for organisms where the entire genomic DNA sequence is known, we have developed a novel and efficient computer simulation to identify the spots, which we call RLGS pattern-designated virtual imaging restriction landmark genome scanning (Vi-RLGS). We used the *Arabidopsis* genome sequence to construct a Vi-RLGS system because its entire nuclear genomic DNA sequence has recently been reported, although there are a few small gaps.

First, a computer simulation for the RLGS procedure was written, following the flowchart shown in Fig. 2. In this simulation program, the landmark restriction enzymes and the recognition sites are indicated by A. Similarly, B and C indicate the second and third restriction enzymes, respectively. Therefore, "AA," "AB," "AC," "BB," "BC," and "CC" indicate DNA fragments with the respective sites at either end. In step (b), the DNA is digested with enzyme A. Next, it is digested with enzyme B, and "AA" and "AB" fragments are used in the following step. In step (f), in-gel digestion with enzyme C following 1st-D electrophoresis is performed, and step (g) involves 2D-electrophoresis in the computer simulation. Presently, there are some gaps in the *Arabidopsis* genomic sequence, and the data used in this program have sequence ends at either end. The landmarks located in the regions of flanking gaps and the end sequences of inserts are irregular spots that affect the RLGS analysis. Therefore, the irregular spots are simulated manually and added to the resulting simulation patterns. Finally, the "AC," "AB," and "AA" fragments are presented as Vi-RLGS profiles containing sequence information as well as their chromosome location.

To compute the size of each DNA fragment, we used Southern's method in which the reciprocal of mobility (M) plotted against fragment length (L) is linear ($L = 1/M$). This relationship is linear over a wider range than the conventional semilogarithmic plot. In addition, the following equation was adapted to high voltage gradients and the scanning range used in the RLGS method:

$$M = \frac{k_1}{L - k_2} + M_0$$

Where

$$M_0 = \frac{M_3 - M_1 A}{1 - A}$$

$$A = \left\{\frac{L_1 - L_2}{L_2 - L_3}\right\}\left\{\frac{M_3 - M_2}{M_2 - M_1}\right\}$$

$$k_1 = \frac{L_1 - L_2}{1/(M_1 - M_0) - 1/(M_2 - M_0)}$$

$$k_2 = L_1 - \left\{\frac{k_1}{M_1 - M_0}\right\}.$$

M_0 is a factor calculated to give the best fit to a straight line with the above equation. To calculate M_0, three points, defined by molecular weight standards L_1, L_2, and L_3 and their corresponding mobilities (M_1, M_2, and M_3) were chosen. To provide the precise simulation pattern, three data points were obtained from actual electrophoresis patterns using a

1-kbp ladder marker in a 1st-D 0.8% agarose gel and a 100-bp ladder marker in a 2nd-D 5% polyacrylamide gel, and were substituted for L_{1-3} and M_{1-3} in the above equation. In a preliminary test of this program, a simulated pattern was generated using the entire chloroplast genome sequence for *Arabidopsis thaliana*. *Not*I is used as a universal restriction landmark enzyme in animal and plant RLGS analysis. Here, we performed a *Not*I-*Arabidopsis* simulation profile using our Vi-RLGS system, and *Eco*RV and *Mbo*I were used as enzymes B and C respectively (*Not*I-*Eco*RV-*Mbo*I).

The Vi-RLGS patterns in the range 0.6 to 7.5 kbp in 1st-D and 50 to 300 bp in the 2nd-D are shown in Fig. 18(a), and the real RLGS pattern in the range corresponding to Fig. 18(a) is shown in Fig. 18(b). The two spots indicated by arrows in Fig. 18(b) were cut from the gel and cloned using the previous method. The sequences confirmed that their patterns concurred with the theoretical Vi-RLGS. There is a difference in the number and pattern of spots in Figs. 18(a and b). There are two possible reasons for this difference: (1) there are some gaps in the *Arabidopsis* genomic DNA sequence; specifically, highly repeated sequences, such as centromeres, telomeres, the ribosomal RNA gene clusters, and their flanking regions have not yet been reported; and (2) DNA modification. In plant DNA, 5′-methylcytosine (5MeC) occurs at cytosine residues in symmetrical sequences, such as CpG and CpNpG. In *Arabidopsis* genomic DNA, the 5MeC content (5MeC/5MeC+C) is approximately 5.2%. *Not*I is sensitive to 5MeC, while *Eco*RV and *Mbo*I are insensitive to this methylation. Black arrowheads in Fig. 18(b) indicate spots that are present in Fig. 18(a) and absent or faint in Fig. 18(b). To confirm the influence of 5MeC in *Not*I landmark detection, the spots indicated by arrows in Fig. 18(b) were subjected to a PCR assay. Primers to amplify an approximately 1-kbp sequence containing *Not*I sites were synthesized and *Not*I-digested genomic DNA was amplified. The presence of amplification products indicates that the *Not*I sites are methylated and that the RLGS spot cannot be detected in the real RLGS pattern. Figure 18(c) shows the PCR assay of six spots: 4T16H5, 4F13C5, 2F26H6, 3T28A8, 2F12C20, and 4T24H24. Spots present in the real RLGS pattern, 4T16H5 and 4F13C5 were not amplified, and the masked spots 2F26H6, 3T28A8, 2F12C20, and 4T24H24 were amplified. The amplified DNA fragments were cloned and confirmed to contain *Not*I sites in the target DNA region. In addition, the RLGS pattern of approximately 22%-reduced 5MeC hypomethylated *Arabidopsis* plants generated using AzaC was produced. As a result, some spots present in Fig. 18(a) and absent in Fig. 18(b) were present in RLGS pattern from hypomethylated plants. This means that the masked spots mainly result from DNA methylation. To find spots present in virtual RLGS files and absent in real RLGS patterns requires exhaustive detection of methylated regions in *Arabidopsis* genomic DNA. Owing to the analysis of the difference between Vi-RLGS and the experimentally produced RLGS pattern, the dynamic change of methylation status of genomic DNA can be screened. On the other hand, the yellow arrowheads in Fig. 18(b) are absent or have a slight difference of mobility in the Vi-RLGS profiles. We speculate that they result from gaps in the sequence or from sliding due to methylation in the flanking regions. Nevertheless, the comparison of virtual and real RLGS patterns identified approximately 80% of the spots in the real RLGS pattern based on their

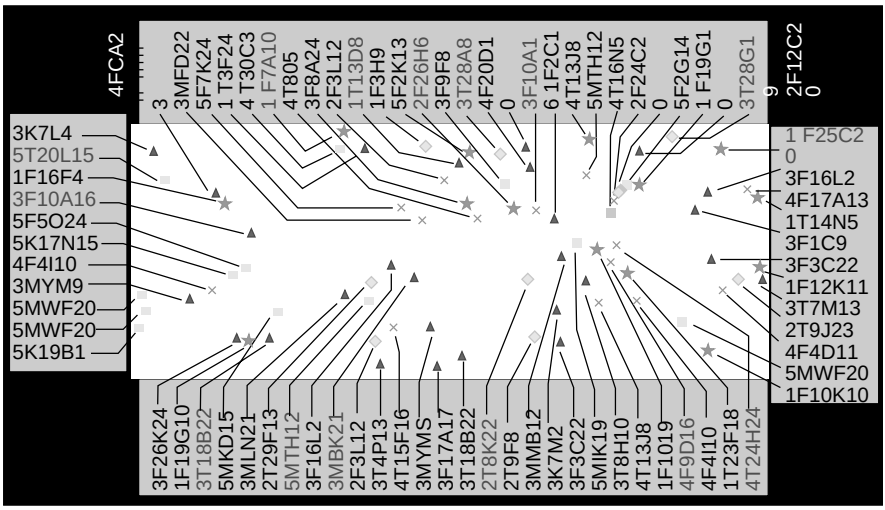

(a)

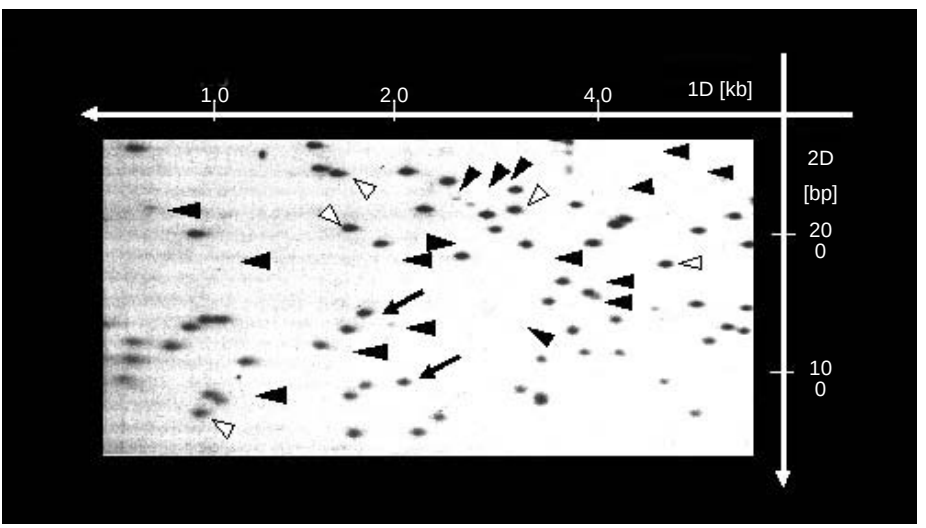

(b)

Fig. 18 Comparison of the Vi-RLGS profile using nucleic DNA sequences of *A. thaliana* and an experimental RLGS pattern (*Not*I-*Eco*RV-*Mbo*I). (a) Vi-RLGS profile of nuclear genomic DNA. Each spot is given the clone name available at the Web site of The *Arabidopsis* Information Resource (TAIR). Head numbers indicate the chromosome number. Circle (violet): chromosome 1; rhombus (yellow): chromosome 2; triangle (red): chromosome 3; cross (blue): chromosome 4; square (green): chromosome 5. (b) The real RLGS pattern corresponding to the region shown in (a). Black arrowheads indicate spots present only in the Vi-RLGS profile. White arrowheads indicate spots that are absent in the Vi-RLGS profile. (c) PCR assay of seven spots comparing uncut genomic DNA and *Not*I-digested genomic DNA. The marker is *Hinc*II-digested ø174. (d) Vi-RLGS analysis of the mutant. (e) Southern hybridization pattern of control (C) and the mutant (M) digested with *Not*I and *Eco*RV using a clone of the deleted region as the probe. (See color plate p. xxiii.)

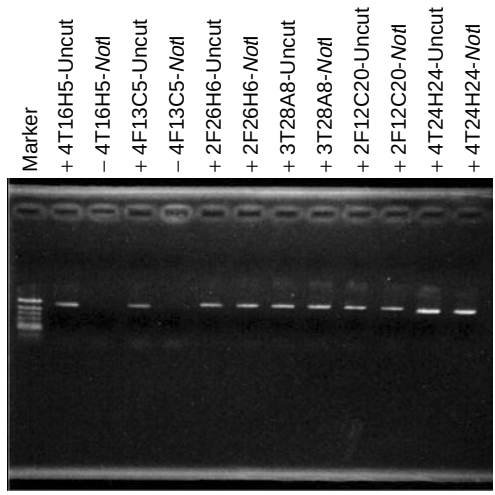

(c)

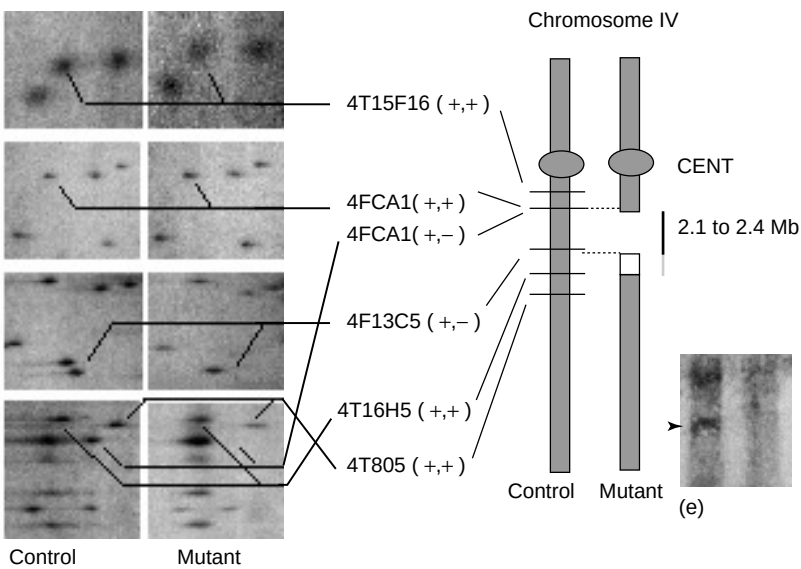

Fig. 18 (*Continued*).

pattern and molecular weight as determined by 1st-D and 2nd-D mobility.

Mutations can be induced using physical mutagens, such as X rays, gamma rays, fast neutrons, and ion beams. Since we expect the Vi-RLGS method to be useful for detecting genetic mutations, such as deletions and translocations, we analyzed the RLGS patterns of an X-ray-induced mutant. Figure 18(d) shows that spots 4FCA1 and 4F13C5 are absent in the mutant RLGS pattern. These spots are

linkage markers localized on chromosome IV. In addition, one 4FCA1 spot (centromere side) is present and another is absent in the mutant pattern. These results indicate that a deletion occurred just downstream from 4FCA1. Since spots 4T16H5 and 4T8O5, which are close to 4F13C5, are present in the mutant pattern, this deletion is approximately 2.1 to 2.4 Mbp in size. To confirm the deletion, a clone of the region between 4FCA1 and 4F15C5 was isolated and used as a probe in Southern hybridization (Fig. 18e). The hybridization pattern showed that this region was absent in the mutant genome.

Presently, there are 177 known *Not*I sites in the *Arabidopsis* genome. The Vi-RLGS system revealed that if enzyme B is replaced with another restriction enzyme (excluding *Eco*RV) such as *Dra*I, approximately 88.7% (157/177 spots) of the *Not*I sites will be present in the resolvable scanning field of the *Not*I-*Eco*RV-*Mbo*I and *Not*I-*Dra*I-*Mbo*I patterns. In addition, the average size of a *Not*I fragment of *Arabidopsis* genomic DNA is approximately 706 kbp (125 Mbp = *Arabidopsis* genome size/177). Although some spots are masked by DNA methylation, our *Not*I-*Eco*RV-*Mbo*I profile covers 79.5% of all *Not*I sites, and an Mbp-order deletion is readily detected by the *Not*I-Vi-RLGS profiles. Further Vi-RLGS analysis using various other landmark enzymes will increase the number of effective landmarks and allow more detailed genomic scanning.

Vi-RLGS saves considerable time and labor, and provides a simple procedure for identifying polymorphic spots in RLGS analysis as well as for screening suitable restriction enzymes. In theory, this method is useful for the genomic analysis of organisms with large genomes, even without sequence information for each chromosome. Therefore, the Vi-RLGS system offers a different spectrum of mutation detection than microarrays, and will be a useful tool for detecting mutations in DNA length and modification in forthcoming postgenome sequencing research.

7 Perspective

One of the advantages of RLGS when compared with protein 2-dimensional gels is that the spots of interest can be cloned. The spot cloning technique will expand the application field of RLGS. In fact, a gene whose mRNA transcription is regulated in imprinted manner was cloned using this technique.

The concept of genome scanning arose from the need to survey the genome in the biochemical approach to DNA. Three classic methods that have been previously developed are hybridization-based method, PCR-based method and staining (ethidium bromide)-based method. RLGS is a fourth type based on the novel concept that restriction enzyme sites can be used as landmarks by direct end-labeling, resulting in a variety of specific features and advantages.

The RLGS method had previously been applied to several major research fields, including (1) screening polymorphism, genetic linkage analysis, and positional cloning; (2) cloning of imprinted genes; (3) methylation and genome structure analyses in a tumor or a cloned animal; and (4) discrimination of agricultural crop cultivars.

Analysis of genome modification, such as methylation, is the most important target phenomenon. When the location on

the genome is specified, microarray or PCR techniques can be used. In particular, microarray analysis has enormous capacity to screen many points throughout the genome. However, the whole genome sequence data is the essential prerequisite for the design of the microarray. On the other hand, to screen the genome whose entire sequence has not been analyzed, RLGS is used as it is clear that it is the most rapid system to discover new landmarks because the microarray system can only screen known sites that are printed in a microarray. Even in the screening of a completely sequenced genome, RLGS is very rapid, economical, and convenient for screening the methylation status of genomic DNA because, unlike microarrays, it does not require any special equipment. Thus, RLGS will be a useful method to compliment microarray technology for analyzing whole genome sequence data.

Bibliography

Books and Reviews

Costello, J.F., Smiraglia, D.J., Plass, C. (2002) Restriction landmark genome scanning, *Methods* **27**, 144–149.

Hayashizaki, Y., Hirotsune, S., Okazaki, Y., Muramatsu, M., Asakawa, J. (1995) Restriction Landmark Genomic Scanning Method, *Molecular Biology and Biotechnology*, VCH Publishers, pp. 813–817.

Hayashizaki, Y., Watanabe, S. (Eds) (1997) *Restriction Landmark Genomic Scanning (RLGS)*, Springer-Verlag, Tokyo.

Primary Literature

Brilliant, M.H., Gondo, Y., Eicher, E.M. (1991) Direct molecular identification of the mouse pink-eyed unstable mutation by genome scanning, *Science* **252**, 566–569.

Chapman, V.M., Hirotsune, S., Okazaki, Y., Hatada, I., Mukai, T., Kawai, J., Hirasawa, T., Nishitani, Y., Watanaba, S., Shiroishi, T., Moriwaki, K., Matsuda, Y., Manly, K., Elliott, R., Hayashizaki, Y. (1992) Mouse Genome Mapping in An Interspecific Cross Using Restriction Landmark Genomic scanning (RLGS), *CSH Meeting on Genome Mapping and Sequencing Abs*, p. 247.

Dietrich, W., Katz, H., Lincoln, S.E., Shin, H.-S., Friedman, J., Dracopoli, N.C., Lander, E.S. (1992) A genetic map of the mouse suitable for typing intraspecific crosses, *Genetics* **131**, 423–447.

Fisher, S.G., Lerman, L.S. (1979) Length-independent separation of DNA restriction fragments in two-dimensional gel electrophoresis, *Cell* **16**, 191–200.

Hatada, I., Hayashizaki, Y., Hirotsune, S., Komatsubara, H., Mukai, T. (1991) A genomic scanning method of higher organisms using restriction sites as landmarks, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9523–9527.

Hayashizaki, Y., Hatada, I., Hirotsune, S., Okazaki, Y., Komatsubara, H., Mukai, T., Kawai, J., Hirasawa, T., Nishitani, Y., Watanabe, S., Manly, K., Elliott, R., Taylor, B., Chapman, V. (1992) A Genomic Scanning Method for Higher Organisms Using Restriction Sites as Landmarks (Restriction Landmark Genomic Scanning: RLGS) and its Application for Mouse Genome Mapping with Recombinant Inbred Strains, *CSH Meeting on Genome Mapping and Sequencing Abs*, p. 120.

Hayashizaki, Y., Hirotsune, S., Hatada, I., Tamatsukuri, S., Miyamoto, C., Furuichi, Y., Mukai, T. (1992) A new method for constructing *Not*I linking and boundary libraries using a restriction trapper, *Genomics* **14**, 733–739.

Hayashizaki, Y., Hirotsune, S., Okazaki, Y., Hatada, I., Shibata, H., Kawai, J., Hirose, K., Watanabe, S., Fushiki, S., Wada, S., Sugimoto, T., Kobayakawa, K., Kawara, T., Sibuya, T., Mukai, T. (1993) Restriction landmark genomic scanning method and its various applications, *Electrophoresis* **14**, 251–258.

Hayashizaki, Y., Hirotsune, S., Okazaki, Y., Shibata, H., Akasako, A., Muramatsu, M., Kawai, J., Hayashizaki, Y., Shibata, H., Hirotsune, S., Sugino, H., Okazaki, Y., Sasaki, N.,

Hirose, K., Imoto, H., Okuizumi, H., Muramatsu, M., Komatsubara, H., Shiroishi, T., Moriwaki, K., Katsuki, M., Hatano, N., Sasaki, H., Ueda, T., Mise, N.,Takagi, N., Plass, C., Chapman, V.M. (1994) Identification of an imprinted U2af binding protein related sequence on mouse chromosome 11 using the RLGS method, *Nat. Gen.* **6**, 33–40.

Hirasawa, T., Watanabe, S., Shiroishi, T., Moriwaki, K., Taylor, B., Matsuda, Y., Elliott, R., Manly, K., Chapman, V.M. (1994) A genetic linkage map of the mouse using restriction landmark genomic scanning (RLGS), *Genetics* **138**, 1207–1238.

Hirotsune, S., Hatada, I., Komatsubara, H., Nagai, H., Kuma, K., Kobayakawa, K., Kawara, T., Nakagawara, A., Fujii, K., Mukai, T., Hayashizaki, Y. (1992) New approach for detection of amplification in cancer DNA using restriction landmark genomic scanning, *Cancer Res.* **52**, 3642–3647.

Kawai, J., Hirotsune, S., Hirose, K., Fushiki, S., Watanabe, S., Hayashizaki, Y. (1993) Methylation profiles of genomic DNA ofmouse developmental brain detected by restriction landmark genomic scanning (RLGS) method, *Nucleic Acid Res.* **21**, 5604–5608.

Lindsay, S., Bird, A.P., (1987) Use of restriction enzymes to detect potential gene sequences in mammalian DNA, *Nature* **327**, 336–338.

Matsuyama, T., Kimura, M.T., Koike, K., Abe, T., Nakao, T., Asami, T., Ebisuzaki, T., Held, W.A., Yoshida, S., Nagase, H. (2003) Global methylation screening in the *Arabidopsis thaliana* and *Mus musculus* genome: applications of virtual image restriction landmark genomic scanning (Vi-RLGS), *Nucleic Acids Res.* **31**, 4490–4496.

Nagai, H., Hirotsune, S., Komatubara, H., Hatada, I., Mukai, T., Hayashizaki, Y., Matsubara, K. (1993) Genomic analysis of human hepatocellular carcinomas using restriction landmark genomic scanning, *Cancer Detection Prev.* **17**, 399–404.

Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, M.F., Ramirez-Slis, R., Webster, T.D., Ledbetter, D.H., Caskey, C.T. (1989) Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6686–6690.

Ohgane, J., Wakayama, T., Kogo, Y., Senda, S., Hattori, N., Tanaka, S., Yanagimachi, R., Shiota, K., (2001) DNA methylation variation in cloned mice, *Genesis* **30**, 45–50.

Okazaki, Y., Hirose, K., Hirotsune, S., Okuizumi, H., Sasaki, N., Ohsumi, T., Yoshiki, A., Kusakabe, M., Muramatsu, M., Kawai, J., Watanabe S., Plass, C., Chapman, V.M., Nakao, K., Katsuki, M., Hayashizaki, Y. (1995) Direct detection and isolation of RLGS spot DNA marker tightly linked to a specific trait using RLGS spot-bombing method, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5610–5614.

Okazaki, Y., Okuizumi, H., Ohsumi, T., Nomura, O., Takada, S., Kamiya, M., Sasaki, N., Matsuda, Y., Nishimura, M., Tagaya, O., Muramatsu, M., Hayashizaki, Y., (1996) A genetic linkage map of the Syrian hamster and localization of cardiomyopathy locus on chromosome 9qa2.1-b1 using RLGS spot-mapping, *Nat. Genet.* **13**, 87–90.

Okazaki, Y., Okuizumi H., Sasaki, N., Ohsumi, T., Kuromitsu, J., Kataoka, H., Muramatsu, M., Iwadate, A., Hirota, N., Kitajima, M., Plass, C., Chapman, V.M., Hayashizaki, Y. (1994) A genetic linkage map of the mouse using an expanded production system of restriction landmark genomic scanning (RLGS Ver.1.8), *Biochem. Biophys. Res. Commun.*, **205**, 1922–1929.

Saiki, R.S., Scharf, S., Faloona, F., Mullis, K.B., Horn, H., Erlich, H.A., Arnheim, N. (1985) Enzymatic amplification of b-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia, *Science* **230**, 1350–1354.

Saito, A., Iimure, K., Okuizumi, H., (2002) Development of DNA polymorphic marker for cultivar discrimination in mat rush (*Juncus effusus* L.), *Brain Technol. News* **94**, 10–13 (in Japanese).

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis, *J. Mol. Biol.* **98**, 503–517.

Southern, E.M. (1979) Measurement of DNA length by gel electrophoresis, *Anal. Biochem.* **100**, 319–323.

The Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, *Nature* **408**, 796–815.

Uitterlinden, A.G., Slagboom, P.E., Knook, D.L., Vijg, J. (1989) Two dimensional DNA fingerprinting of human individuals, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2742–2746.

URLs

Amersham	http://www.jp.amershambiosciences.com/i.e._index.asp
Biocraft	http://www.bio-craft.co.jp/
FMC	http://www.fmc.com/
Kodak	http://wwwjp.kodak.com/
Nakalai	http://www.nacalai.co.jp/
Nichias	http://www.nichias.co.jp/
Sanplatec	http://www.sanplatec.co.jp/index2.html
Sigma	http://www.sigma-aldrich.co.jp/index2.html
Sterling	http://www.ssl.gb.com/knight/mainindex.htm
TAIR	http://www.arabidopsis.org/
TaKaRa	http://www.takara-bio.co.jp/
Tomy	http://bio.tomys.co.jp/
Toyobo	http://www.toyobo.co.jp/

Genomic DNA Libraries, Construction and Applications

Eugene R. Zabarovsky
Microbiology and Tumor Biology Center, Karolinska Institute, Stockholm, Sweden

1 Principles 443

2 Techniques 445
2.1 General Characteristics of λ-based Vectors Used for Construction of Genomic Libraries 445
2.2 Construction of General Genomic Libraries 447
2.3 Construction of Jumping and Linking Libraries. Use of Linking and Jumping Clones to Construct a Physical Chromosome Map 451

3 Applications and Perspectives 455
3.1 Cloning DNA Markers Specific for a Particular Chromosome 455
3.2 CpG Islands as Powerful Markers for Genome Mapping; CpG Islands and Functional Genes 457
3.3 Alu-PCR and Subtractive Procedures to Clone CpG Islands from Defined Regions of Chromosomes 458
3.4 IBD (Identical-by-descent) Fragments for Identification of Disease Genes 460
3.5 Strategies to Map and Sequence Genomes; Hierarchical, Whole-genome, and Slalom Sequencing Approaches 460
3.6 Restriction-site-tagged Microarrays to Study CpG-Island Methylation 462
3.7 Restriction-site-tagged Sequences to Study Biodiversity 464

4 Summary 465

Bibliography 465
Books and Reviews 465
Primary Literature 465

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

Keywords

Blue–white Selection
Not really selection but color identification. Vectors carrying the β-galactosidase (*lacZ*) gene (or part of it) produce blue plaques in the presence of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal). If this gene is located in a stuffer fragment, then all recombinants will form white plaques and parental vectors will produce blue plaques in the presence of X-gal.

Genetic Selection
Usually, in cloning, selection against parental, nonrecombinant molecules in favor of the recombinant. For λ-based vectors used for construction of genomic libraries, the two most commonly used types of selection are *Spi* and *supF*. *Spi*+ phages carrying *red* and *gam* genes cannot grow in *E. coli* lysogens carrying prophage P2; since, however, the majority of the vectors contain these genes in a stuffer fragment, only recombinant phages can grow in such *E. coli* strains. Selection for *supF* exploits λ vectors carrying amber mutations. These vectors cannot replicate without the *supF* gene, which must be present either in the host or in the cloned insert. If the insert carries the *supF* gene, only recombinant phages will be able to replicate in an *E. coli* host without the suppressor gene.

Polylinker
A short DNA fragment (in the vector) that contains recognition sites for many restriction enzymes, which can be used for cloning DNA fragments into this vector.

Restriction Enzyme
An enzyme that recognizes a specific sequence in DNA and can cut at or near this sequence. In cloning procedures, the most commonly used enzymes produce specific protruding (sticky) ends at the ends of the DNA molecule. Each enzyme produces unique sticky ends. The DNA molecules possessing the same sticky ends can be efficiently joined with the aid of DNA ligase (see ligation).

(STS) Sequence-tagged Site
A short (200–500 bp) sequenced fragment of genomic DNA that can be specifically amplified using PCR. STS represents or is linked to some kind of marker (i.e. it is mapped to a specific locus on a chromosome).

By virtue of the powerful technology developed in molecular biology, it is possible to isolate any DNA fragment in the genome of an organism and, after reverse transcription, any transcribed gene in the form of a complementary DNA. The isolation (cloning) procedure involves the insertion of the DNA fragment into a vector, capable of replication in a microorganism, which allows production of large quantities of the DNA fragment for physical or biological analysis. Upon determination of the location in the genome from which the particular DNA

fragment was derived, that fragment acquires the property of a DNA marker. Such DNA markers are a prerequisite for physical and genetic mapping of the genome of the organism. DNA markers are also of importance for the diagnosis of genetic diseases. DNA markers can be divided into several different classes depending on the way in which the markers were selected among the fragments of genomic DNA. Examples of such classes are anonymous, micro- and minisatellites, restriction fragment length polymorphism (RFLP) markers, and *Not*I linking clones.

Vectors and clone libraries of different types can be used to clone markers. Lambda-based vectors and genomic libraries of different kinds are commonly used for this purpose. Many different variants of λ-based vectors that combine features of different cloning vehicles (plasmids, M13 and P1 phages) have been created for this purpose. The use of each vector is usually limited to a specific task: the construction of general genomic libraries (which contain all genomic DNA fragments) or special genomic libraries (which contain only a particular subset of genomic DNA fragments). Among these special libraries, *Not*I linking and jumping libraries have particular value for physical/genetic mapping and sequencing of the human genome. Shotgun and slalom libraries are usually used for sequencing purpose and comparative genomics.

1 Principles

In molecular biology, "cloning" is the insertion of DNA with interesting information into a specific vector that allows replication and transfer of the cloned DNA from one host to another. The vector containing the inserted DNA is called a *recombinant vector* to distinguish it from its parental vector, which does not contain any foreign DNA. Usually, "interesting information" is a piece of DNA obtained from any target organism; it can be a gene (or part of a gene) or simply anonymous DNA sequences for which no function is yet known. It can originate directly from DNA or can be obtained from reverse transcription of RNA molecules. The main idea of cloning is to obtain the interesting piece of DNA in a quantity large enough for analysis and further experiments. Now, the vectors and strategies used for cloning come in many different types. This chapter concentrates on the widely used λ-based vectors and the construction of genomic libraries, which played an important role in the Human Genome Project.

A genomic library is a collection of recombinant vectors; it contains DNA fragments representing the genome of a particular organism. Genomic libraries can be either general, containing DNA fragments covering the whole genome, or special, containing only specific genomic fragments that differ in certain parameters. Some are CG rich whereas others contain only particular size fragments of DNA obtained after digestion with a particular restriction enzyme, contain specific repeats, and so on. Important special genomic libraries are the jumping and linking types (see Sect. 2.3).

Cloned DNA fragments can be located to a specific site of a chromosome, after which they can serve as markers for physical and genetic mapping. Different types of markers are used. The so-called

anonymous markers represent randomly cloned DNA fragments whose functions or specific features are not known. Other DNA markers can possess specific features. They can contain a known gene or expressed sequences with unknown function, CpG islands (also associated with genes, see Sect. 3.2), or recognition sites for rare cutting restriction enzymes convenient for long-range mapping. Such markers can be polymorphic, that is, they have different structures in different individuals (they are usually distinguished on the basis of different mobility in gel electrophoresis). Such markers are extremely important in mapping and cloning human disease genes and for construction of genetic maps. Three types of polymorphic markers are commonly used (Fig. 1).

Single-nucleotide polymorphism (SNP) markers are DNA fragments that have a point mutation in some individuals of a population. The advantages of SNPs are their abundant numbers ($>10^6$) and the fact that they can be detected by nonelectrophoretic methods, for example, using microarrays. However, usually SNP has only two alleles. A subtype of SNPs are RFLP markers that recognize genomic fragments containing polymorphic recognition sites for a particular restriction endonuclease (e.g. *Taq*I, *Msp*I). The same chromosomal regions in different individuals contain or lack this recognition site.

A second form of DNA polymorphism results from variation in the number of tandemly repeated (VNTR) DNA sequences in a particular locus. Usually, they are divided in two types – mini- and microsatellites. Minisatellites are DNA fragments 0.1 to 20 kb long that contain many copies (from 3 to more than 40) of

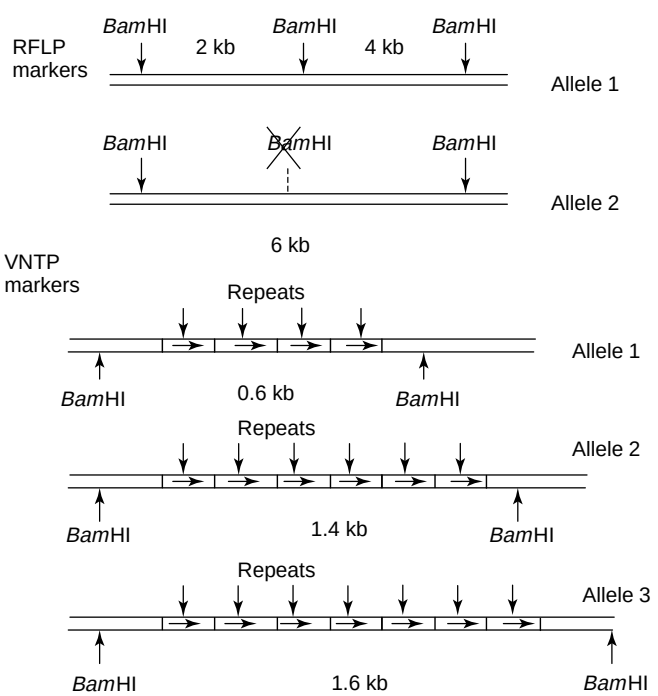

Fig. 1 The difference between RFLP and VNTR polymorphism.

6 to 60 bp repeats. All these repeats share a 10 to 15 bp core sequence similar to the generalized recombination signal (*chi*) of *E. coli*. When DNA from different individuals is digested with a restriction enzyme that does not cut inside these repeats, the length of the fragments produced will depend on the number of repeats at the locus. Since the minisatellites and repeats constitute a relatively large fragment, it is possible to discriminate between different alleles using ordinary nondenaturing gel electrophoresis and Southern blot analysis. Many different alleles (usually more than five) can be distinguished at a locus containing minisatellites. Minisatellites cluster around the distal ends of human chromosomes and, sometimes, are located near the genes.

Microsatellites are relatively short DNA fragments (usually <100 bp) with repeat units from 1 to 5 bp (such as A, AC, AG, AAC, AAAG, etc.). They are numerous (5×10^5) and uniformly distributed throughout the human genome with an estimated average spacing of about 6 kb.

Microsatellites can be very polymorphic (more than 10 alleles at the same locus), and polymorphism usually increases with the number of repeats. Microsatellites with fewer than 10 copies are usually not polymorphic. Since microsatellites are short, they can be analyzed quickly by using the polymerase chain reaction (PCR) with primers flanking each locus. Different alleles can be resolved using denaturing gel electrophoresis. Very frequently, microsatellites are associated with Alu repeats (see Sect. 3.1) and this creates problems for their use with PCR, since Alu repeats are conserved in the human genome. Thus, flanking primers for the microsatellite located in the Alu repeat are unlikely to be useful because they prime from many Alu repeats and, instead of discrete bands, give a smear after electrophoretic separation of the PCR products.

An optimal marker should have the features of all the different types of markers just discussed. The ideal marker should contain (1) a gene, (2) a CpG island that has been shown to be very conserved in the genome and can be used for comparative gene mapping in different species, (3) a rare cutting restriction site useful for physical mapping, and (4) polymorphic sequences (e.g. microsatellites). One of the best candidates for having all these features together is *Not*I linking clones, that is, recombinant clones containing the *Not*I restriction site.

2 Techniques

2.1 General Characteristics of λ-based Vectors Used for Construction of Genomic Libraries

Among other more modern vectors used for construction of genomic libraries (yeast, bacterial and P1 artificial chromosomes; YAC, BAC, and PAC respectively), phage λ-based vectors are still very popular. The reason is that the genetics and features of both λ phage and *E. coli* (the host for λ phage) are well known. The size of the phage DNA that can be packaged into viable phage particles is limited between 37.7 and 52.9 kb. This means that it is possible to biologically regulate the size range of the cloned DNA fragment. There exist extremely efficient *in vitro* systems for packaging such DNA into λ phage particles (10^9 plaque-forming units per μg of DNA) to produce viable phages. To combine the features of different vector systems, extensive modifications of λ phage vectors were developed (Fig. 2).

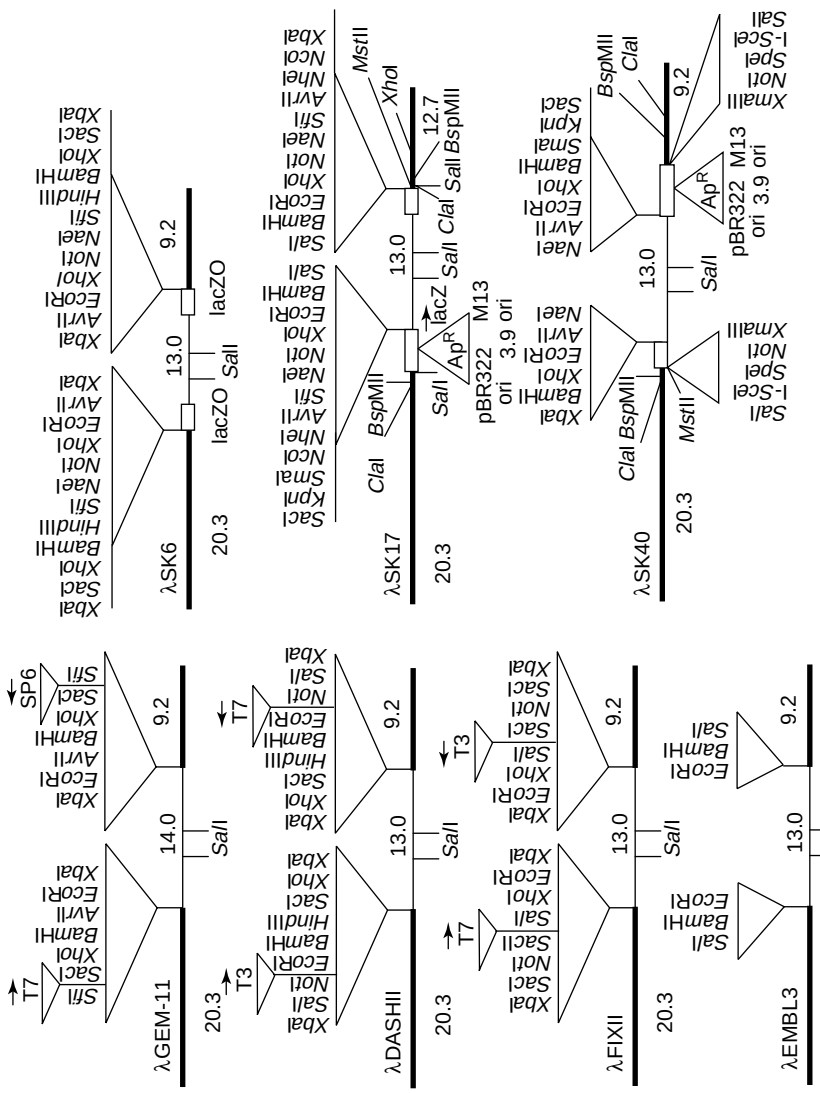

Fig. 2 Examples of λ-based vectors: standard λ vectors (λGEM11, λFIXII, λDASH, λEMBL3, λSK6) and diphasmid vectors (λSK17 and λSK40). Sizes are in kilobases. Not all restriction sites are shown. Heavy lines represent vector arms, thin lines denote the stuffer fragment; open boxes mark plasmid and M13 sequences; *lacZO* is the *lac* operator sequence. T3, T7, SP6 – promoters for T3, T7, and SP6 RNA polymerases.

Cosmids are essentially plasmids that contain the *cos* region of phage λ responsible for packaging of DNA into the phage particle. The advantages of cosmids are easy handling (as with plasmids) and large cloning capacity. Since the plasmid body is usually small (3–6 kb), large DNA molecules (46–49 kb) can be cloned in these vectors.

Phasmids are λ phages that have an inserted plasmid. They have the same basic features as λ phage vectors, but the inserted foreign DNA fragment can be separated from the body of the phage DNA and converted into the plasmid form. After the conversion, the cloned DNA fragment will exist as a recombinant plasmid.

Hyphages represent another type of λ-based vectors. They were constructed from M13 vectors with a built-in *cos* site of λ. Since these vectors have the main features of M13 vectors, they can be obtained in single-stranded form. Their distinctive feature is the capability to be packaged with high efficiency into λ-phage-like particles. This decreases the chance of recovering nonrecombinant vectors and opens the possibility of constructing a representative genomic library in single-stranded form.

Diphasmids are vectors, which offer the opportunity to combine the advantages of phages (λ and M13) and plasmids. Diphasmids can be divided into two classes: those that can replicate as phage λ (an improvement over phasmids) and those that are incapable of replicating as phage λ (i.e. a cosmid capable of being packaged into phage M13 particles).

In some cases, it is more convenient to work with a genomic library in plasmid than in λ phage form. The construction of a representative genomic library directly in a plasmid vector has several drawbacks and difficulties. However, all these problems can be easily solved with the help of phasmid and diphasmid vectors. Often a genomic library is constructed in a λ phage and then converted in its entirety to plasmid form.

2.2 Construction of General Genomic Libraries

Representativity is one of the most important features of a genomic library. In a "representative" genomic library, every genomic DNA fragment will be present in at least one of the recombinant phages of the library. In practice, however, this is difficult to achieve. Some genomic fragments are not clonable because of the strategy used for construction of the genomic library. For example, if the maximal cloning capacity of the vector is 18 kb and *Eco*RI digestion is used to construct the library, no genomic *Eco*RI fragments bigger than 18 kb will be present. In some cases, genomic DNA fragments can suppress the growth of the vector or the host cell, with the result that its cloning can be restricted to specific vector systems.

The important reason for decreased representativity is the different replication potential of different recombinant phages. Most researchers work with amplified libraries. The ligated DNA molecules are packaged into the λ phage particles and plated on a lawn of *E. coli* cells (usually many petri dishes are used for such plating). Then all λ phage particles are eluted from the plaques, and the liquid eluted from all the petri dishes is mixed. Glycerol or dimethyl sulfoxide is added to the eluate, and the aliquots are kept at – 76 °C. This procedure is called *amplification* of the library.

With amplification, a library can be kept for many years and can be used for many experiments by many researchers. On the other hand, each recombinant phage present in the library gives a single plaque

at the first plating, and since different recombinant phages differ in their growth potential, there may be differences of 100 times in the abundance of clones after amplification. This means that in the amplified library, some of the phages are present 100 times more often than others. In this case, to recover all recombinant phages obtained after packaging into λ phage particles, one needs to plate 100 times more phages than that obtained after the original plating. Since such quantities are difficult to achieve in practice, some recombinant phages are virtually lost from the library after amplification.

How is the representativity of a library estimated? A library is considered to be representative if after the first plating (before amplification) it contains a number of recombinant clones together, containing genomic DNA fragments equal to 7 to 10 genome equivalents. For example, human genomic DNA contains approximately 3×10^9 bp. If the vector contains on an average 15 kb inserts, the representative library should contain 1.4 to 2.0×10^6 recombinant clones.

The way in which the genomic DNA fragments are produced for cloning is also important. The more randomly the genomic DNA is broken, the more representative a library can be obtained. Clearly, the *Eco*RI enzyme (6 bp recognition site) will cut genomic DNA less randomly than *Sau*3AI (4 bp recognition). Probably, the shearing of DNA molecules using physical methods (e.g. syringe, sonication) is the most reliable way to obtain randomly broken DNA molecules.

An important characteristic feature of a library is the percentage of recombinants. For most purposes, if a library contains more than 80% of recombinant phages, it is better to omit the genetic selection procedure because it usually decreases the representativity of the library. To calculate the percentage of recombinants, one can use genetic (*Spi*) selection as in the case of λEMBL vectors. Another approach relies on blue–white color identification (e.g. λ-Charon series). A third class of vectors has both genetic selection and blue–white color identification (λSK4, λSK6). There are three commonly used ways to construct genomic libraries (Table 1, Fig. 3).

The original ("classical") method includes generation of sheared genomic DNA fragments using physical or enzymatic manipulations followed by the physical separation of fragments of a particular size using, for example, ultracentrifugation or gel electrophoresis. The vector DNA is digested with two (or even three) restriction enzymes whose recognition sites are located in the polylinker and are separated by a few base pairs. The arms and the stuffer piece are purified from the small oligonucleotide molecules released after the digestion by, for example, precipitation with polyethylene glycol (PEG) 6000. The stuffer piece and both arms will now have different sticky ends, preventing re-creation of the original vector molecules during subsequent ligation with genomic DNA fragments.

In the second "dephosphorylation" approach, the phage arms are prepared by simultaneous digestion with two restriction enzymes as shown earlier. Genomic DNA is partially digested to the extent that DNA fragments with sizes in the range of 15 to 20 kb will represent the majority of the products. These DNA fragments are dephosphorylated to prevent their ligation to each other and are then ligated to the vector arms. If too big or too small genomic DNA fragments are ligated to the phage arms, size limitations will make it impossible for these recombinant molecules to yield viable phages. Compared to the preceding

Tab. 1 Comparison of three basic methods used in construction of representative genomic libraries.

Method	*DNA [µg] needed to construct a representative genomic library*	*DNA fractionation*	*Self-ligation of*		*Effectiveness of packaging*		*Number of packaging reactions to get representative library at maximal efficiency per microgram of genomic DNA*	*Genetic selection necessary to remove nonrecombinants*
			vector DNA	*genomic DNA*	*Per microgram of vector DNA*	*Per microgram of genomic DNA*		
Classical	100–1000	Yes	Yes	Yes	10^5–10^7	10^5–10^7	1	Yes
Dephosphorylation	5–10	No	Yes	No	10^4–10^5	10^5–10^6	3–5	Yes
Partial filling in	5–10	No	No	No	10^5–10^7	10^5–10^7	1	No

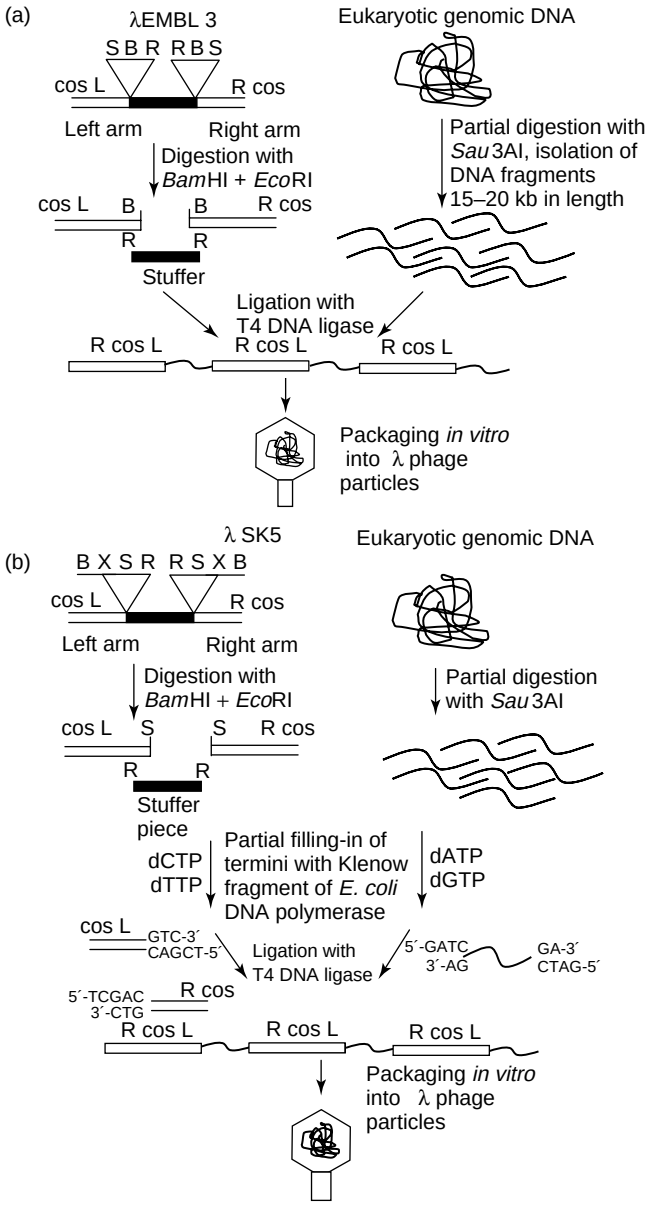

Fig. 3 Two approaches for constructing genomic libraries: (a) classical method and (b) partial filling-in method. Lcos and Rcos – left and right parts of the *cos* site correspondingly; B, *Bam*HI; R, *Eco*RI; S, *Sal*I; X, *Xma*III.

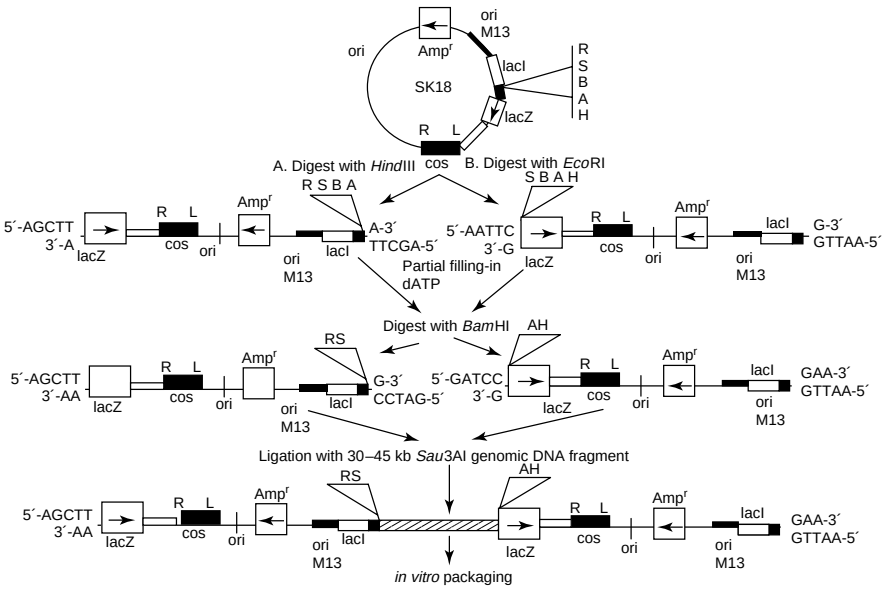

Fig. 4 One approach to the construction of genomic libraries in cosmid vectors (not all restriction sites are shown): A, *Acc*I; B, *Bam*HI; H, *Hind*III; R, *Eco*RI; S, *Sma*I.

method, this procedure is quick, and representative libraries can be obtained from a small quantity of genomic DNA.

The third "partial filling-in" method, also avoids fractionation steps (Fig. 3). Phage arms are prepared as described before (in this particular case *Sal*I and *Eco*RI are shown, but many other combinations can be used), and the sticky ends produced are partially filled in with the Klenow fragment of DNA polymerase I (or other DNA polymerase) in the presence of dTTP and dCTP. Genomic DNA partially digested with *Sau*3AI is also partially filled in, but in the presence of dATP and dGTP. Under such conditions, self-ligation of vector arms or genomic DNA is impossible.

Genomic libraries can be constructed in cosmids using the same approaches just described. The absence of selection against nonrecombinant vector and the possibility of packaging into phage particles concatemers composed solely of cosmid fragments make it even more important to prevent self-ligation of vector fragments. Many similar approaches have been suggested and one of them is shown in (Fig. 4) where the formation of vector-concatemers is prevented by partial filling in.

A similar effect can be achieved by dephosphorylation or by digestion at the first step with *Acc*I and *Sma*I instead of *Eco*RI and *Hind*III. *Sma*I produces blunt ends and *Acc*I gives sticky ends with only two protruding base pairs. The ligation of these ends will be far less effective than that for *Bam*HI and *Sau*3AI sticky ends (four protruding base pairs).

2.3 Construction of Jumping and Linking Libraries. Use of Linking and Jumping Clones to Construct a Physical Chromosome Map

For long-range mapping and cloning of large stretches of genomic DNA, the

two most widely used methods are construction of overlapping DNA sequences (contigs) using chromosome walking (e.g. BAC cloning) and chromosome jumping. The technique of BAC cloning is now used by many laboratories. Still, this approach is not devoid of problems and drawbacks. These problems could be diminished, however, by using jumping/linking libraries. Moreover, jumping and linking libraries can be used independently for construction of a long-range restriction map using pulsed field gel electrophoresis (PFGE). Jumping clones contain DNA sequences adjacent to neighboring *Not*I sites, and linking clones contain DNA sequences surrounding the same restriction site.

The two best-known kinds of jumping libraries are the *Not*I jumping library and the "general" jumping (hopping) library. The basic principle of both methods is to clone only the ends of large DNA fragments rather than continuous DNA segments, as in BAC clones. Internal DNA is deleted by controlled biochemical techniques. The main difference is that in the first type of library, complete digestion with a rare cutting enzyme (*Not*I is the most popular) is used, and the second is based on a partial digestion with a frequently cutting enzyme, followed by isolation of DNA fragments of desired size. Using the first type of library, it is possible to jump over long distances (>1000 kb), but only from certain starting points (i.e. those containing the recognition site for the rare cutting enzyme). Using the hopping library, it is possible to start jumping from practically any point and to cover a defined but shorter distance (<150 kb). Only the first type of jumping libraries can be used in conjunction with linking libraries to create genomic maps, as described next.

There are two main approaches for the construction of *Not*I jumping and linking libraries. According to the first method (Fig. 5a), jumping libraries are constructed as follows: DNA of high molecular mass, isolated in low-melting agarose, is completely digested with *Not*I.

The DNA is ligated at very low concentration, in the presence of a dephosphorylated plasmid containing a marker (*supF* gene), and is then circularized, trapping the *supF* gene, which acts like a marker to select clones that contain the ends of a long fragment. Another enzyme, one that has no recognition site in the plasmid, is used to digest the large circular molecules into small fragments, each of which is cloned in a vector phage carrying amber mutations. Recombinant phages containing the plasmid with the two terminal fragments are selected in an *E. coli* strain lacking the suppressor gene.

The linking library can be constructed in different ways. In the original protocol, the genomic DNA is partially digested with *Sau*3A and size selected to obtain 10 to 20 kb fragments. The DNA is then diluted and circularized in the presence of *supF* marker plasmid. The circular products are digested with *Not*I, ligated into a *Not*I-digested suppressor-dependent vector (NotEMBL3A), and plated on a suppressor-negative host.

In another approach, DNA from a total genomic library in a circular form (e.g. cosmid) is digested with *Not*I, and a selectable marker (e.g. resistance to the antibiotic) is inserted into recombinants containing this site. Then these recombinants are selected by their resistance to the antibiotic.

The most important drawback is that all these methods used to construct linking libraries exploit strategies and vectors different from those used to construct jumping libraries. Thus, some fragments

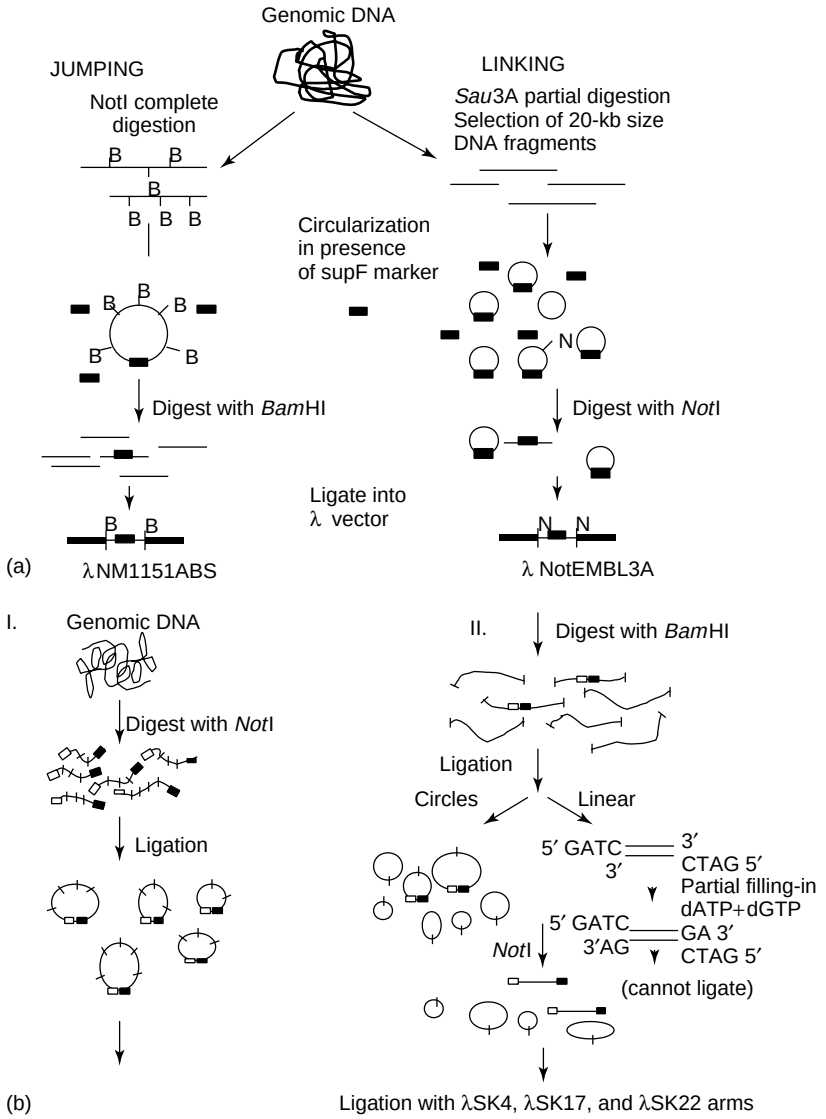

Fig. 5 Two approaches for the construction of jumping and linking libraries: (a) using *supF* marker and (b) using partial filling in. (a) Black boxes, *supF* marker; B, *Bam*HI. (b) Black and white bars denote *Not*I sites and vertical slashes represent *Bam*HI sites; (b I), (b II), construction of the jumping library; (b II), construction of the linking library. In this case, digestion of the genomic DNA with *Bam*HI is the first step.

present in one library (e.g. jumping) will be absent in another (e.g. linking), which creates serious problems for the use of these libraries in mapping.

An integrated approach for construction of jumping and linking libraries is outlined in Fig. 5(b). The most important feature here is that the same vectors and

protocol are used for construction of both libraries.

For the linking library, genomic DNA is completely digested with *Bam*HI. Subsequently, the DNA is self-ligated at a very low concentration (without a *supF* marker) to yield circular molecules as the main product. To eliminate any linear molecules, the sticky ends are partly filled in with the Klenow fragment in the presence of dATP and dGTP. Since the Klenow fragment also has exonuclease activity, all the *Bam*HI sticky ends are neutralized and nearly all ends generated upon random DNA breakage become unavailable for ligation. Subsequently, the DNA is cut with *Not*I and cloned in λSK4, λSK17, and λSK22 vectors.

The same strategy is applied in the construction of a *Not*I jumping library. One initial step is added: genomic DNA is fully digested with *Not*I and self-ligated at a very low concentration. The subsequent steps are the same as for the linking library. An obvious difference between this approach and the preceding ones is that the procedure combines a biochemical selection for *Not*I jumping fragments with improved ligation kinetics during the preparation of the libraries (see Table 2 for a comparison of two methods).

In theory, a linking library is sufficient to construct a physical chromosomal map. When linking clones are used to probe a PFGE genomic blot (*Not*I-digested genomic DNA), each clone should reveal two DNA fragments, which are adjacent in the genome. Thus, in principle, one should be able to order the rare cutting sites with just a single library and one digest, although it will generally not be possible to distinguish between two fragments of the same size. To resolve such ambiguities, it is important to use several different libraries, each for a particular enzyme, and to overlap the resulting patterns just as in ordinary restriction

Tab. 2 Comparison of two main methods for construction of *Not* I jumping libraries.

Method I (with supF marker)	***Method II (with partial filling in)***
Materials required for construction of jumping library	
2 μg genomic DNA (10 mL ligation volume)	1 μg genomic DNA (5 mL ligation volume)
40 μg vector arms	1 μg vector arms
500 μL-sonicated extract (SE) for *in vitro* packaging	15 μL SE
2000 μL freeze-thaw extract for *in vitro* packaging (FTL)	10 μL FTL
Cloning capacity	
0–12 kb	0.2–23 kb
Expected yield	
$1-5 \times 10^4$/μg genomic DNA	$1-5 \times 10^5$/μg genomic DNA
Percentage of recombinants before genetic selection	
<1%	45–60%
Maximum sizes of jumps	
450 kb	>1000 kb

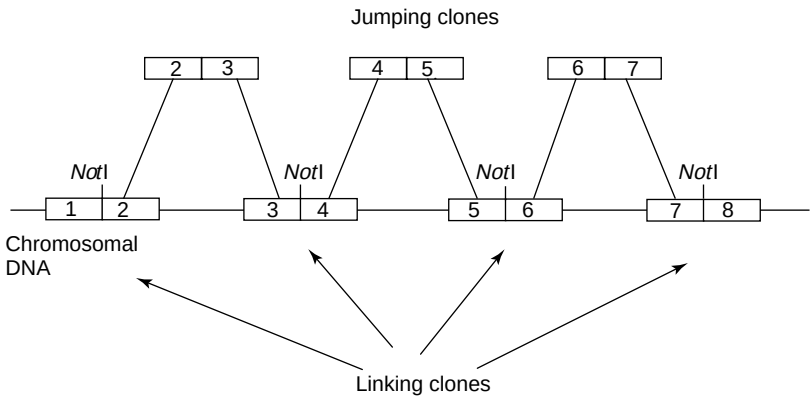

Fig. 6 Long-range mapping using jumping and linking libraries.

fragment analysis. To accomplish this for the whole human genome will be very laborious. However, for small stretches of the genome containing 5 to 10 *Not*I sites, this approach can be efficient. The use of jumping and linking libraries in a complementary fashion simplifies this approach (Fig. 6).

Moreover, by cross-screening the two libraries, it should be possible in principle to jump from clone to clone (–jumping–linking–jumping, etc.), to generate an ordered map without using PFGE techniques at all. One of the main problems with this approach is that the presence of very CG-rich regions and repeats in the human DNA may result in cross-hybridization between different clones.

In the shotgun sequencing approach for long-range genome mapping, the hybridization technique is replaced by sequencing. Jumping and linking clones are sequenced from the ends and, subsequently, the linear order of the *Not*I clones on a chromosome can be established using a computer program. Even a 20 bp sequence is likely to uniquely identify a sequence in the human genome. The sequence data provide a means to discriminate even between different instances of the same class of repeats.

3 Applications and Perspectives

3.1 Cloning DNA Markers Specific for a Particular Chromosome

It is important, for many purposes, to clone individual DNA markers for a specific chromosome. The most straightforward approach for such purposes is to prepare special libraries that contain recombinant clones from a particular chromosome. One approach is based on the use of the fluorescence-activated cell sorter (FACS) system. FACS operates on the principle of rapid analysis of suspended particles, single cells, or even chromosomes. A suspension of chromosomes stained with fluorescent dyes (usually Hoechst 33258 and chromomycin A3) passes through the focus of a laser beam that excites the DNA-bound fluorescent dyes. Equal-sized droplets are formed by ultrasonic dispersion, and the droplets containing the desired chromosome as indicated by the

fluorescence measurement are deflected from the main stream by an electric field and collected in a tube. DNA isolated from sorted chromosomes can be used for construction of genomic libraries.

Another approach is based on the use of hybrid cell lines. To obtain such somatic cell hybrids, human cells are fused with rodent cells (e.g. mouse or hamster). When the resulting hybrid cells are grown in culture, there is a progressive loss of human chromosomes until only one or a few of them are left. At this step, some of the segregant hybrid cells can be quite stable. In a modification of this technique, a human cell line is transfected with a plasmid vector or infected with a retroviral vector that contains a selectable marker (e.g. *gpt* or *neo*). Transfected clones are screened for those that contain only a single integrated plasmid per cell (and thus containing a single marked chromosome). These transformants are micronucleated by prolonged colcemid treatment and the microcells, containing only one chromosome, are produced using special techniques. The microcells are fused to mouse cells, and the resulting human–mouse microcell hybrids, containing the single marked chromosome, are isolated by growth in selective medium. Hybrid cells containing only fragments of human chromosomes can be produced using human chromosomes with translocations and deletions, or the fragments can be produced experimentally by X-irradiation (radiation hybrids). Such hybrid cell lines are very useful not only for constructing genomic libraries but also for physical mapping of already isolated DNA markers and genes. Another advantage of somatic hybrid cells is that DNA is available in large amounts and can be used for different purposes. Human-specific clones can be isolated from the library by hybridization to total human DNA.

At least 40% of the human genome consists of repetitive sequences of different kinds. Among them, about 50% are repeats randomly distributed in the human genome – different kinds of short (SINE) and long (LINE) interspersed (repeating) elements. The most abundant among them is the Alu repeat family. Alu repeats are present at $>1.0 \times 10^6$ copies per genome with an estimated average spacing of about 3 kb. Alu repeats have a length of about 300 bp and consist of 2 homologous units. Related repeats also exist in other mammals. Different members of this family from the same species usually have homology of 80 to 90% but are only about 50% identical in different species. These repeats have conserved and variable regions. It is possible to find conserved sequences that are species-specific. These conserved sequences can be used as primers for PCR, to specifically amplify human sequences in the presence of nonhuman DNA. These features are the basis for using Alu repeats for isolation of human chromosome–specific sequences from hybrid cell lines containing human and nonhuman DNA sequences. Moreover, if a hybrid cell line contains only a short piece of human chromosome, the Alu-PCR approach can be used for isolation of markers specific for a defined region of the chromosome. The principles of the approach are shown in Fig. 7.

In the case of two hybrid cell lines, one containing a complete human chromosome (HCL1) and the other carrying the same chromosome but with a deletion (HCL2), this method offers the possibility of obtaining markers specific for the deletion. This variant can be called the differential Alu-PCR approach for obtaining

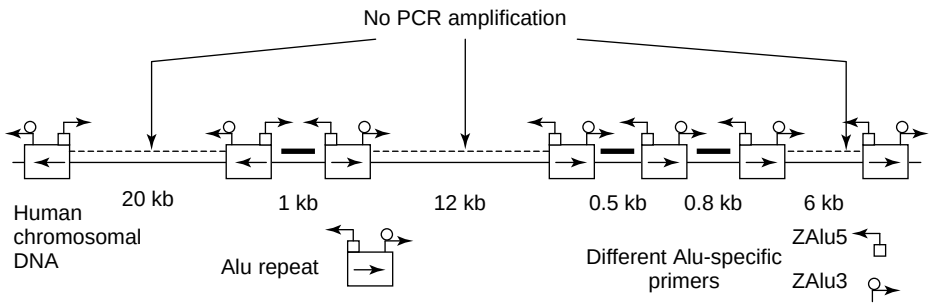

Fig. 7 General scheme for Alu-PCR.

DNA markers. The approach is mainly used in two variants.

In one variant, Alu-PCR is done using DNA from both cell lines, and the products of the reactions are separated by agarose gel electrophoresis. Some bands present in the products from HCL1 will be absent among the products from HCL2. These bands can be excised and cloned, giving markers localized in the deletion. The disadvantage of this approach is that usually such Alu-PCR results in a large number of products that have a very complex pattern and look like a smear on the gel. Among the solutions to this problem that have been suggested is the use of more specific primers (only for a subset of the Alu repeats), or genomic DNA digested with restriction enzyme. Another solution is to use hybrid cell lines that contain only small pieces of human chromosomes.

The second variant of the Alu-PCR approach is mainly used in connection with sources that contain only a limited amount of human material: YAC clones and radiation hybrid cell lines containing small pieces of the human chromosomes. The YACs can, for example, be used for Alu-PCR and the total products of the PCR reaction can be used as a probe to screen genomic libraries (e.g. in cosmids). The hybridization pattern reveals which cosmids are present in one YAC, which in other YACs, and which are present in one but absent in another. Such an approach is also useful for mapping.

Another approach to obtaining region-specific libraries is to use chromosome microdissection to physically remove the chromosomal region of interest; the minute quantities of microdissected DNA can be subjected to a microcloning procedure. Spreads of human chromosomes are made and stained using standard cytogenetic techniques. DNA from an individual band is then cut from the chromosome using ultrafine glass needles or is isolated with the help of laser equipment. In the latter case, all other chromosomes are destroyed by the laser, and intact DNA is present only in the chromosome of interest. DNA obtained from only a few (2–20) chromosomes is enough for constructing a region-specific library. This DNA is amplified using PCR and cloned in plasmid or λ vectors.

3.2 CpG Islands as Powerful Markers for Genome Mapping; CpG Islands and Functional Genes

Although human DNA is highly methylated, stably unmethylated sequences (about 1% of the genome) have been

observed in human chromosomal DNA. Such sequences occur as discrete "islands," usually 1 to 2 kb long, that are dispersed in the genome. They are usually called *CpG (rich) islands* because they contain more than 50% of CG (human genome contains, on average, about 40% of CG contents). Their distinctive feature is the presence of CpG pairs at a predicted frequency, whereas elsewhere in the genome, it is present at a frequency less than 25%. Altogether, there are about 30 000 islands in the haploid genome (the average spacing is about 1 per 100 kb). It is now clear that the majority (if not all) of CpG islands are associated with genes. It has been shown that recognition sites for many of the rare cutting enzymes are closely associated with CpG islands. For example, at least 82% of all *Not*I and 76% of all *Xma*III sites are located in the CpG islands. More than 20% of CpG-island-containing genes have at least one *Not*I site in their sequence, while about 65% of these genes have *Xma*III site(s).

3.3
Alu-PCR and Subtractive Procedures to Clone CpG Islands from Defined Regions of Chromosomes

The Alu-PCR approach is used successfully for cloning DNA markers, but it does result in cloning small DNA fragments (500 bp) between Alu sequences. Alu sequences are distributed in a random fashion and are not linked with genes or other markers. An obvious suggestion for making Alu-PCR more useful for mapping is to use not simply genomic DNA from different sources but linking libraries constructed from these sources. This modification has certain advantages: using isolated probes, it is easy to clone a parental linking clone (e.g. *Not*I), which is a natural marker on the chromosome, convenient for linkage with other markers. Furthermore, linking clones are located in CpG-rich islands that are associated with genes. According to this scheme, linking libraries are constructed from different hybrid cell lines containing either whole or deleted human chromosomes. Then, total DNA isolated from these libraries can be used for Alu-PCR in the manner described in Sect. 3.1. However, in this case every PCR product (either discrete bands or total product) is used as a probe to isolate linking clones from the defined region of the chromosomes.

Genomic subtractive methods represent potentially powerful tools for identification of deleted sequences and cloning region–specific markers. This approach has given rewarding results in less-complex systems such as yeast or cDNA libraries, but the great complexity of the human genome has generated serious problems. These problems can be overcome by reducing the complexity of the human genomic sequences. Two approaches have been suggested to achieve this aim. In one (representational difference analysis), only a subset of genomic sequences (e.g. *Bam*HI fragments less then 1 kb) is used for subtractive procedures; this approach will result in cloning of random sequences. In the other, *Not*I linking libraries are used instead of whole genomic DNA. Intermediate products, that is, circles after the first ligation step can also be successfully utilized for subtraction (Fig. 8).

The *Not*I linking library is at least 100 times lower in complexity than the whole human genome. It is approximately equal in complexity to the yeast genome. Since this approach is not linked with Alu repeats, it offers the possibility of isolating *Not*I linking clones that are unavailable for cloning using Alu-PCR.

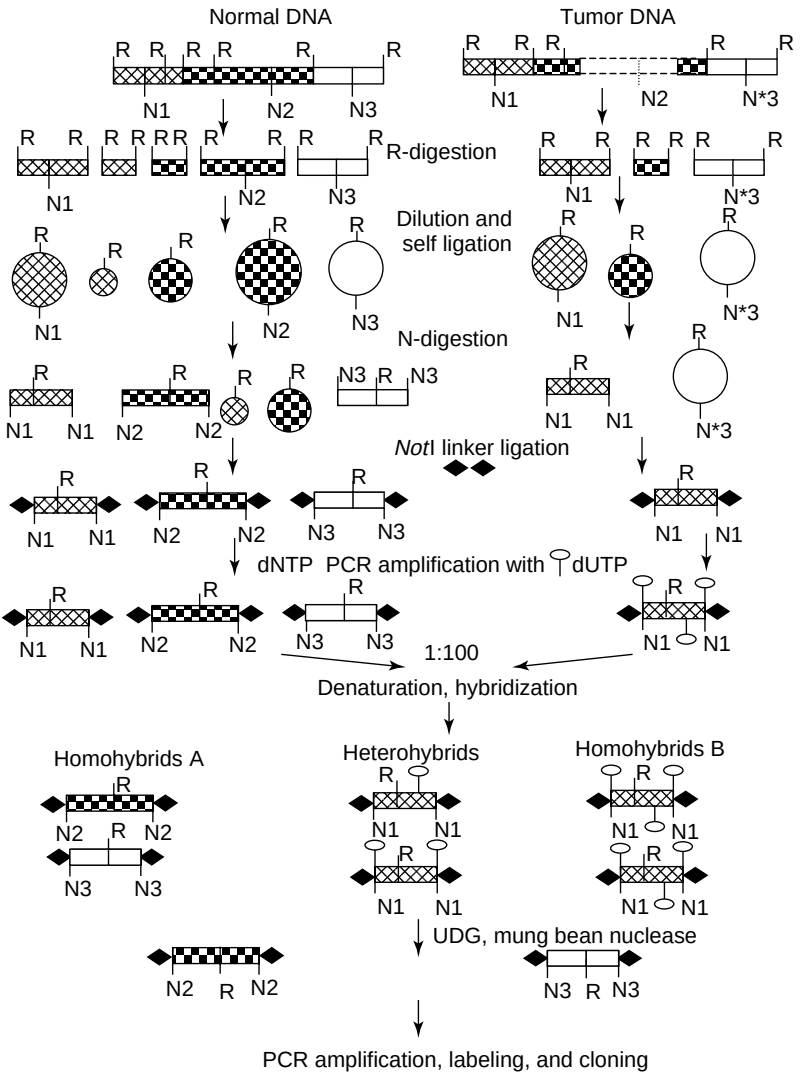

Fig. 8 General scheme for the *Not*I-CODE (Cloning *O*f *DE*leted_ sequences) procedure. N, *Not*I sites; R, restriction enzymes recognizing a 4 to 6 bp sequence. Methylated *Not*I site is indicated by an asterisk. UDG, uracil-DNA glycosylase destroys all DNA molecules containing dUTP. Mung bean nuclease digests double-stranded DNA with mismatches and all single-stranded molecules. Circles digested with *Not*I and PCR-amplified are called NR, *Not*I representation, because only sequences surrounding *Not*I sites are present in this amplification product. The N2 site is deleted and the N*3 site is methylated in tumor DNA.

3.4 IBD (Identical-by-descent) Fragments for Identification of Disease Genes

Development of the methods permitting cloning of identical sequences (CIS) between two sources of DNA can be very useful for many purposes, including isolation of disease genes. Identical-by-descent(IBD) sequences refer to segments of the human genome shared by two individuals because they are inherited from a common ancestor. Regions that are IBD between individuals affected with a disease conceivably can contain the disease gene(s). Two approaches were suggested to clone such IBD sequences.

In GMS (genomic mismatch scanning), each DNA preparation is digested with *Pst*I to yield fragments with protruding 3′ ends. The 3′ protruding ends are protected from digestion by exonuclease III (*Exo*III) in later steps. One of the DNA preparations is fully methylated at all GATC sites with *E. coli Dam* methylase (DAM+). The other DNA preparation remains unmethylated. The two DNA pools are then mixed in equal ratios, denatured, and allowed to reanneal. Digestion of the reannealed DNA with both *Dpn*I and *Mbo*I, which cut at fully methylated and unmethylated GATC sites respectively, results in digestion of the homohybrids. The heterohybrids are resistant to both *Dpn*I and *Mbo*I digestion and survive this treatment. Discrimination between perfect, mismatch-free heterohybrids and those with mismatches is done by the MutHLS enzyme. Only perfect duplexes will escape nicking during this step. All DNA molecules, except mismatch-free ones, are degraded further with *Exo*III. Thus, the full-length, unaltered heterohybrids are purified from the other DNA fragments.

Another method was called *CIS* (cloning of identical sequences). The scheme of the CIS-procedure is shown in Fig. 9.

DNA A and B is digested with *Bam*HI and ligated to special linkers containing two recognition sites for *Mvn*I. DNA A is PCR-amplified in the presence of dUTP and m5dCTP; thus, all cytosines will be methylated. DNA B is PCR-amplified in the presence of normal dCTP and biotinylated primers. The two DNA preparations are mixed in equal ratios, denatured, and hybridized. Subsequently, the DNA is digested with *Mvn*I. This enzyme can digest only dsDNA molecules without methylcytosine and will digest all homohybrids B (they contain at least four sites for *Mvn*I).

The DNA mixture is next treated with mung bean nuclease. This nuclease destroys all imperfect hybrids and ssDNAs. Thus, after this treatment we will have only perfect homohybrids A and perfect (without any mismatches) heterohybrids. The DNA mixture is then treated with UDG (uracil-DNA glycosylase). This enzyme removes the uracil base from the DNA and thus destroys all DNA from individual A. As a result, there will be only ssDNA from individual B, which is identical to the DNA in individual A.

3.5 Strategies to Map and Sequence Genomes; Hierarchical, Whole-genome, and Slalom Sequencing Approaches

During the last few years, impressive progress has been made in mapping and sequencing whole genomes of various organisms. Two basic strategies have so far been employed for genome sequencing.

According to one scheme (hierarchical approach), the whole genome is mapped using different types of markers, and a minimal set of large-insert clones, such

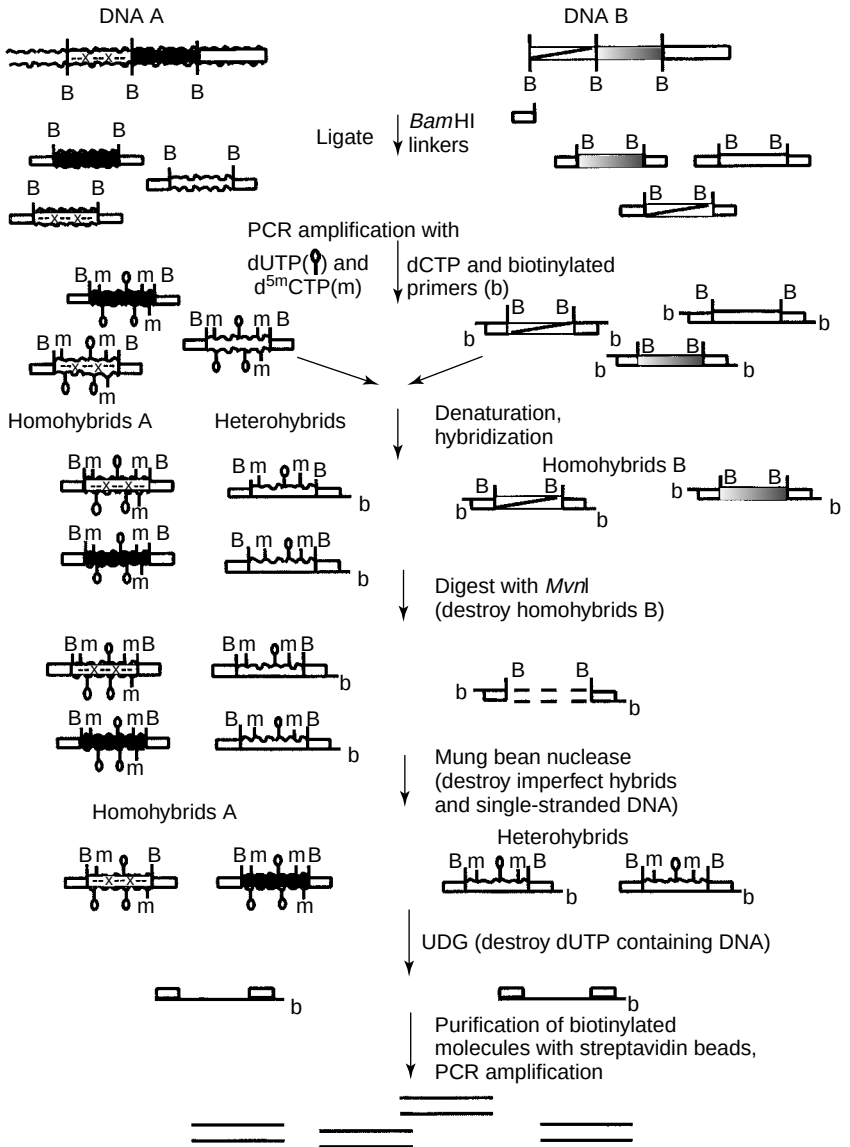

Fig. 9 General scheme of the CIS, cloning of identical sequences, procedure. The same enzymes are used as in the CODE method but in a different order, and the result is opposite.

as cosmids, PAC, or BAC clones, is established. Subsequently, these large-insert clones are sequenced using a shotgun sequencing strategy: small-insert libraries, containing randomly sheared fragments of the large-insert clones, are constructed and sequenced.

A variant of this strategy, the whole-genome shotgun (WGS) sequencing strategy, was developed later and has proved

valuable. This method involves end sequencing of large- (PAC, BAC, cosmids) and small-insert (2 and 10 kb) clones. DNA fragments in the small-insert clones are generated by physical shearing of whole genomic DNA. All resulting reads are joined in one sequence with special computer programs. The WGS method requires the generation of sequences covering the whole genome 10 to 15 times. If sequence coverage is less, then the contig assembly process cannot be completed, and sequences and clones will just represent islands, without connection or order. Therefore, despite impressive technological progress, mapping and sequencing of even small bacterial genomes is expensive and laborious.

After completion of the genomic sequence from one organism, there will be a great demand, in many cases, for comparison with the genomes of other individuals, related species, pathogenic and nonpathogenic strains, and so on, in the growing field of comparative genomics. Such comparisons are highly relevant to our understanding of human and animal health, evolution, and ecology.

An efficient strategy for simultaneous genome mapping and sequencing was recently developed. The approach is based on slalom libraries, which combine features of general genomic, jumping, and linking libraries. First experiments demonstrated the feasibility of the approach, and showed that the efficiency (cost-effectiveness and speed) of existing mapping/sequencing methods can be improved at least 5- to 10-fold. The slalom allows the establishment of a physical map, with minimal sets of overlapping clones, which will pinpoint differences in genome organization between organisms. At the same time, considerable sequence coverage of the genome (about 50%) will be achieved. This will make it possible to locate virtually every gene in a genome, for more detailed (comparative) study. Furthermore, since the efficiency of contig assembly in the slalom approach is virtually independent of sequence read length, even short sequences, as produced by rapid high-throughput sequencing techniques suffice to complete a physical map and sequence scan of a small genome. A combination of these new sequencing techniques with the slalom approach increases the power of the method 10 to 50 times more and makes it an efficient tool for comparative genomics. The main principle of the slalom libraries is shown in Fig. 10.

Two standard genomic *Eco*RI- and *Bam*HI-digested libraries are constructed and they will completely cover the whole genome. The problem is how to put *Eco*RI and *Bam*HI fragments in the correct order. It can be solved using the connecting library. The connecting library can be constructed as follows: DNA isolated "en masse" from an *Eco*RI library is digested with *Bam*HI, circularized in the presence of the Kan^r gene, and plated on agar with kanamycin. The clones isolated in this manner will be identical in structure to the clones from an *Eco*RI jumping library prepared in the classical way. By comparing end sequences in these three slalom libraries, all clones can be positioned relative to each other and a minimal contig of overlapping clones may be established.

3.6 Restriction-site-tagged Microarrays to Study CpG-Island Methylation

Methylation, deletions, and amplifications of cancer genes constitute important mechanisms in carcinogenesis, and CGI (CpG-island-containing) microarrays were

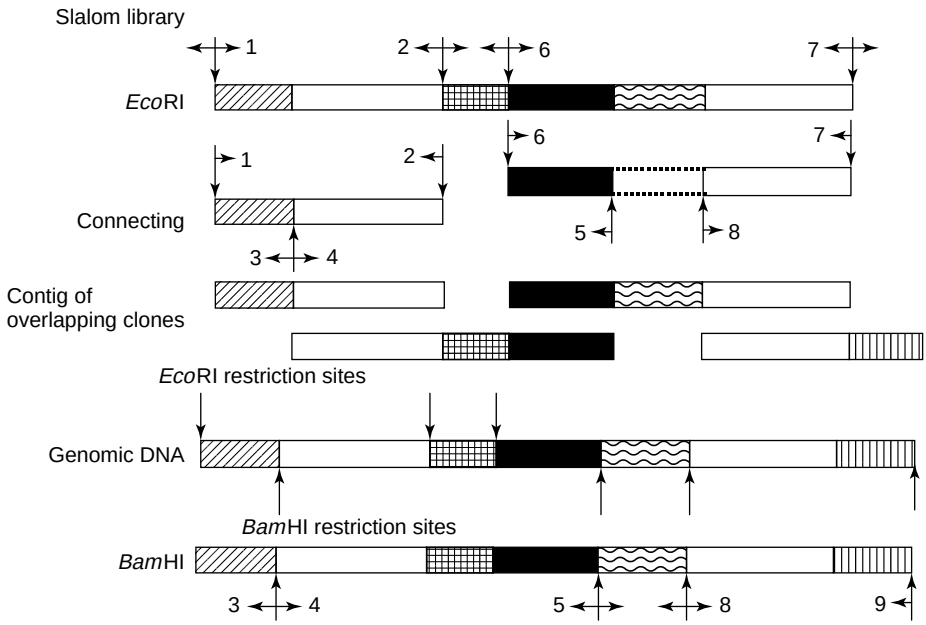

Fig. 10 Simplified slalom library approach scheme. Numbers designate identical end sequences in different libraries that can be joined by a computer program in a contig of overlapping clones. Dashed lines show genomic DNA sequences deleted in the connecting library.

suggested to study hypermethylation in cancer cells. These microarrays can detect methylation changes in tumor DNA. However, it is unclear whether these microarrays can be used to detect hemizygous methylation or copy-number changes. As the whole human genome DNA is used for labeling and the clones are small (0.2–2 kb), this creates a serious problem. Oligonucleotide-based microarrays also can be used to study methylation changes in cancer cells; however, only a limited number of genes can be tested in such experiments.

A rough estimation is that the human genome contains 15 000 to 20 000 *Not*I sites. Therefore, thousands of genes could be tested with *Not*I clone microarrays. The fundamental problems for genome-wide screening using *Not*I clones are (1) the size and complexity of the human genome, (2) the number of repeat sequences, and (3) the comparatively small sizes of the inserts in *Not*I clones (on average 6–8 kb). A special procedure was developed to amplify only regions surrounding *Not*I sites, the so-called *Not*I representations (NRs, see Fig. 8). Other DNA fragments were not amplified. Therefore, only 0.1 to 0.5% of the total DNA is labeled. Interestingly, sequences surrounding *Not*I sites contain 10-fold fewer repetitive sequences than the human genome on average, and therefore, these microarrays are not as sensitive as other methods to the background hybridization caused by repeats. The main idea of this application is clear from Fig. 8 (tumor DNA). If a particular *Not*I site is present in the DNA, then the circle will be opened with *Not*I and labeled. However, if

this *Not*I site is deleted or methylated, then the NR will not contain the corresponding DNA sequences. The *Not*I microarrays can simultaneously detect copy-number changes and methylation and, therefore, they allow the simultaneous study of genetic and epigenetic factors.

The technique underlying the preparation and use of *Not*I microarrays is applicable to any restriction enzyme and represents a new type of microarray, referred to as *restriction-site-tagged* (RST) microarrays. Such RST microarrays can be used for different purposes, for example, to study species composition of complex microbial systems.

3.7 Restriction-site-tagged Sequences to Study Biodiversity

There is still much to learn about the human normal microflora. The human gut contains approximately 1 to 2 kg of bacterial cells. The number of these cells in the intestine is 10 to 100 times larger than the number of cells in the human body, but at best 10 to 15% of the microbial species are known. To be able to analyze complex microbial mixtures is of great importance for many applications. For instance, differences between individual compositions of the normal flora will be instrumental for future analysis of the effects on the normal flora composition of diet, foods, geographical location, and medication. Conversely, the effects of gut microflora on aging, autoimmunity, and colonic cancer risk can be studied.

Analysis of human *Not*I flanking sequences (see Sect. 2.3) have demonstrated that even short sequences surrounding *Not*I sites can yield information sufficient to isolate new genes or uniquely describe eukaryotic or prokaryotic genomes.

These results led to the realization that it would be possible to use short sequences surrounding *Not*I sites or, in general, restriction-site-tagged sequences (RSTS) for the analysis of complex microbial mixtures. The collection of *Not*I tags represents the *Not*I sequence passport or in short *Not*I passport. *Not*I passporting means the process of creating *Not*I tags/passports.

The general design of the experiment is as follows. Genomic DNA is digested with *Not*I and ligated to a linker with *Not*I sticky ends. This linker contains the *Bpm*I recognition sites. This restriction nuclease cuts 16/14 bp outside of the recognition site. The ligation mixture is PCR-amplified with special primers and, finally, 19 bp tags flanking *Not*I sites are generated. DNA from, for instance, fecal samples and surgical specimens is digested with *Not*I, and a *Not*I passport for the particular specimen is generated. A comparison of such passports from different individuals or from the same individual before and after drug treatment will reveal the differences between them.

Analysis of tags for *Not*I, PmeI, and SbfI for 70 completely sequenced bacteria revealed that more than 95% of tags are species-specific and even different strains of the same species can be distinguished. None of these tags matched human or rodent sequences. Therefore, the approach allows analysis of complex microbial mixtures such as those in the human gut and identification with high sensitivity of a particular bacterial strain on a quantitative and qualitative basis.

A similar approach can be used for eukaryotic cells, for example, for analysis of cancer cells.

RSTS-passporting and RST-microarray approaches are mutually complementary. These two approaches are based on completely different biochemical techniques but aim to solve the same problems.

4 Summary

While several different strategies are available to obtain and use DNA markers for identifying and mapping DNA sequences in complex organisms, no single system is likely to suffice for obtaining a complete and accurate map and sequence of the human genome. Rather, a combination of different approaches and vector systems is needed to corroborate data from different sources.

See also Body Expression Map of Human Genome; Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Ausubel, F.M., Kingston, R.E., Brent, R., Moore, D.D., Sedman, J.G., Struhl, K., Smith, J.A. (1987–2003) *Current Topics in Molecular Biology*, Wiley, New York.

Bird, A. (2002) DNA methylation patterns and epigenetic memory, *Genes Dev.* **16**, 6–21.

Brown, T.A. (1999) *Genomes*, Wiley, New York co-published with BIOS Scientific Publishers, Oxford.

Collins, F.S. (1988) Chromosome Jumping, in: Davis, K.E. (Ed.) *Genome Analysis: A Practical Approach*, IRL Press, Oxford, pp. 73–94.

Mueller, R.F., Young, I.D. (2001) *Emery's Elements of Medical Genetics*, Churchill Livingstone, Edinburgh.

Poustka, A., Lehrach, H. (1988) *Chromosome Jumping: A Long Range Cloning Technique*, in: Setlow, J.K. (Ed.) Genetic Engineering Principles and Methods, Vol. 10, Brookhaven National Laboratory, Upton, New York and Plenum Press, New York, London, pp. 169–193.

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edition, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York.

Strachan, T., Read, A. (1999) *Human Molecular Genetics*, Wiley, New York co-published with BIOS Scientific Publishers, Oxford.

Zabarovsky, E.R., Kashuba, V.I., Gizatullin, R.Z., Winberg, G., Zabarovska, V.I., Erlandsson, R., Domninsky, D.A., Bannikov, V.M., Pokrovskaya, E., Kholodnyuk, I., Petrov, N., Zakharyev, V.M., Kisselev, L.L., Klein, G. (1996) *Not*I jumping and linking clones as a tool for genome mapping and analysis of chromosome rearrangements in different tumors, *Cancer Detect. Prev.* **20**, 1–10.

Zabarovsky, E.R., Winberg, G., Klein, G. (1993) The SK-diphasmids – vectors for genomic, jumping and cDNA libraries, *Gene* **127**, 1–14.

Primary Literature

Adorjan, P., Distler, J., Lipscher, E., Model, F., Muller, J., Pelet, C., Braun, A., Florl, A.R., Gutig, D., Grabs, G., Howe, A., Kursar, M., Lesche, R., Leu, E., Lewin, A., Maier, S., Muller, V., Otto, T., Scholz, C., Schulz, W.A., Seifert, H.H., Schwope, I., Ziebarth, H., Berlin, K., Piepenbrock, C., Olek, A. (2002) Tumour class prediction and discovery by microarray-based DNA methylation analysis, *Nucleic Acids Res.* **30**, e21, 1–9.

Allikmets, R.L., Kashuba, V.I., Bannikov, V.M., et al. (1994) *Not*I linking clones as tools to join physical and genetic mapping of the human genome, *Genomics*, **19**, 303–309.

Bicknell, D.C., Markie, D., Spurr, N.K., Bodmer, W.F. (1991) The human chromosome content in human x rodent somatic cell hybrids analyzed by a screening technique using Alu PCR, *Genomics* **10**, 186–192.

Bird, A., Taggard, M., Frommer, M., Miller, O.J., Macleod, D. (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA, *Cell* **40**, 91–99.

Brenner, S., Johnson, M., Bridgham, J., et al. (2000) Gene expression analysis by massively

parallel signature sequencing (MPSS) on microbead arrays, *Na. Biotechnol.* **18**, 630–634.

Brookes, A.J., Porteous, D.J. (1991) Coincident sequence cloning, *Nucleic Acids Res.* **19**, 2609–2613.

Brown, P.O., Botstein, D. (1999) Exploring the new world of the genome with DNA microarrays, *Nat. Genet.* **21**(Suppl. 1), 33–37.

Broder, S., Venter, J.C. (2000) Sequencing the entire genomes of free-living organisms: the foundation of pharmacology in the new millennium, *Annu. Rev. Pharmacol. Toxicol.* **40**, 97–132.

Carninci, P., Shibata, Y., Hayatsu, N., et al. (2001) Balanced-size and long-size cloning of full-length, cap-trapped cDNAs into vectors of the novel lambda-FLC family allows enhanced gene discovery rate and functional analysis, *Genomics* **77**, 79–90.

Cheung, V.G., Gregg, J.P., Gogolin-Ewens, K.J., et al. (1998) Linkage-disequilibrium mapping without genotyping, *Nat. Genet.* **18**, 225–230.

Collins, F.S., Weissman, S.M. (1984) Directional cloning of DNA fragments at a large distance from an initial probe: a circularization method, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6812–6816.

Collins, F.S., Drumm, M.L., Cole, J.L., et al. (1987) Construction of a general human chromosome jumping library, with application to cystic fibrosis, *Science* **235**, 1046–1049.

Costello, J.F., Fruhwald, M.C., Smiraglia, D.J., et al. (2000) Aberrant CpG-island methylation has non-random and tumour-type-specific patterns, *Nat. Genet.* **24**, 132–138.

Cross, S.H., Charlton, J.A., Nan, X., Bird, A.P. (1994) Purification of CpG islands using a methylated DNA binding column, *Nat. Genet.* **6**, 236–244.

Eads, C.A., Danenberg, K.D., Kawakami, K., et al. (2000) MethyLight: a high-throughput assay to measure DNA methylation, *Nucleic Acids Res.* **28**, e32, 1–8.

Galm, O., Rountree, M.R., Bachman, K.E., et al. (2002) Enzymatic regional methylation assay: a novel method to quantify regional CpG methylation density, *Genome Res.* **12**, 153–157.

Gonzalgo, M.L., Liang, G., Spruck, C.H., et al. (1997) Identification and characterization of differentially methylated regions of genomic DNA by methylation-sensitive arbitrarily primed PCR, *Cancer Res.* **57**, 594–599.

Hayashizaki, Y., Hirotsune, S., Okazaki, Y. (1993) Restriction landmark genomic scanning method and its various applications, *Electrophoresis* **14**, 251–258.

Kunkel, L.M., Monaco, A.P., Middlesworth, W., Ochs, H.D., Latt, S.A. (1985) Specific cloning of DNA fragments absent from the DNA of a male patient with an X chromosome deletion, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4778–4782.

Kutsenko, A., Gizatullin, R., Al-Amin, A.N., et al. (2002) *Not*I flanking sequences: a tool for gene discovery and verification of the human genome, *Nucleic Acids Res.* **30**, 3163–3170.

Lamar, E.E., Palmer, E. (1984) Y-encoded, species-specific DNA in mice: evidence that the Y chromosome exists in two polymorphic forms in inbred strains, *Cell* **37**, 171–177.

Lander, E.S., Linton, L.M., Birren, B., et al., International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Larsen, F., Gundersen, G., Prydz, H. (1992) Choice of enzymes for mapping based on CpG islands in the human genome, *Genet. Anal. Tech. Appl.* **9**, 80–85.

Li, J., Protopopov, A., Wang, F., et al. (2002) *Not*I subtraction and *Not*I-specific microarrays to detect copy number and methylation changes in whole genomes, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10724–10729.

Li, J., Wang, F., Kashuba, V., et al. (2001) Cloning of deleted sequences (CODE): a genomic subtraction method for enriching and cloning deleted sequences, *Biotechniques* **31**, 788–793.

Li, J., Wang, F., Zabarovska, V.I., et al. (2000) COP – a new procedure for cloning single nucleotide polymorphisms, *Nucleic Acids Res.* **28**, e1,1–5.

Lindblad-Toh, K., Tanenbaum, D.M., Daly, M.J., et al. (2000) Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays, *Nat. Biotechnol.* **18**, 1001–1005.

Lisitsyn, N., Lisitsyn, N., Wigler, M. (1993) Cloning the differences between two complex genomes, *Science* **259**, 946–951.

Lucito, R., West, J., Reiner, A. (2000) Detecting gene copy number fluctuations in tumor cells by microarray analysis of genomic representations, *Genome Res.* **10**, 1726–1736.

Mirzayans, F., Mears, A.J., Guo, S.W., Pearce, W.G., Walter, M.A. (1998) Identification of

the human chromosomal region containing the iridogoniodysgenesis anomaly locus by genomic-mismatch scanning, *Am. J. Hum. Genet.* **61**, 111–119.

Myers, E.W., Sutton, G.G., Delcher, A.L., et al. (2000) A whole-genome assembly of Drosophila, *Science* **287**, 2196–2204.

Nelson, S.F. (1995) Genomic mismatch scanning: current progress and potential applications, *Electrophoresis* **16**, 279–285.

Nelson, S.F., McCusker, J.H., Sander, M.A. (1993) Genomic mismatch scanning: a new approach to genetic linkage mapping, *Nat. Genet.* **4**, 11–18.

Nussbaum, R.L., Lesko, J.G., Lewis, R.A., Ledbetter, S.A., Ledbetter, D.H. (1987) Isolation of anonymous DNA sequences from within a submicroscopic X chromosomal deletion in a patient with choroideremia, deafness, and mental retardation, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6521–6525.

Palmisano, W.A., Divine, K.K., Saccomanno, G., et al. (2000) Predicting lung cancer by detecting aberrant promoter methylation in sputum, *Cancer Res.* **60**, 5954–5958.

Pinkel, D., Segraves, R., Sudar, D., et al. (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays, *Nat. Genet.* **20**, 207–211.

Poustka, A., Pohl, T.M., Barlow, D.P., Frischauf, A.M., Lehrach, H. (1987) Construction and use of human chromosome jumping libraries from NotI-digested DNA, *Nature* **325**, 353–355.

Protopopov, A., Kashuba, V., Zabarovska, V.I., et al. (2003) An integrated physical and gene map of the 3.5-Mb chromosome 3p21.3 (AP20) region implicated in major human epithelial malignancies, *Cancer Res.* **63**, 404–412.

Ronaghi, M., Pettersson, B., Uhlen, M., Nyren, P. (1998) A sequencing method based on real-time pyrophosphate, *Science* **281**, 363–365.

Rosenberg, M., Przybylska, M., Straus, D. (1994) 'RFLP subtraction': a method for making libraries of polymorphic markers, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6113–6117.

Shi, H., Maier, S., Nimmrich, I., et al. (2003) Oligonucleotide-based microarray for DNA methylation analysis: principles and applications, *J. Cell Biochem.* **88**, 138–143.

Smith, C.L., Lawrance, S.K., Gillespie, G.A., et al. (1987) Strategies for mapping and cloning macroregions of mammalian genomes, *Methods Enzymol.* **151**, 461–489.

Snijders, A.M., Nowak, N., Segraves, R., et al. (2001) Assembly of microarrays for genome-wide measurement of DNA copy number, *Nat. Genet.* **29**, 263, 264.

Sugimura, T., Ushijima, T. (2000) Genetic and epigenetic alterations in carcinogenesis, *Mutat. Res.* **462**, 235–246.

Ushijima, T., Morimura, K., Hosoya, Y., et al. (1997) Establishment of methylation-sensitive-representational difference analysis and isolation of hypo- and hypermethylated genomic fragments in mouse liver tumors, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2284–2289.

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W. (1995) Serial analysis of gene expression, *Science* **270**, 484–487.

Venter, J.C., Adams, M.D., Myers, E.W., et al. (2001) The Sequence of the Human Genome, *Science* **291**, 1304–1351.

Waterston, R.H., Lindblad-Toh, K., Birney, E., et al. (2002) Initial sequencing and comparative analysis of the mouse genome, *Nature* **420**, 520–562.

Worm, J., Aggerholm, A., Guldberg, P. (2001) In-tube DNA methylation profiling by fluorescence melting curve analysis, *Clin. Chem.* **47**, 1183–1189.

Yan, P.S., Chen, C.M., Shi, H., et al. (2001) Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays, *Cancer Res.* **61**, 8375–8380.

Zabarovska, V.I., Gizatullin, R.G., Al-Amin, A.N., et al. (2002) Slalom libraries: a new approach to genome mapping and sequencing, *Nucleic Acids Res.* **30**, e6, 1–8.

Zabarovska, V., Kutsenko, A., Petrenko, L., et al. (2003) *Not*I passporting to identify species composition of complex microbial systems, *Nucleic Acids Res.* **31**, e5, 1–10.

Zabarovska, V., Li, J., Fedorova, L., et al. (2000) CIS – cloning of identical sequences between two complex genomes, *Chromosome Res.* **8**, 77–84.

Zabarovsky, E.R., Allikmets, R.L. (1986) An improved technique for the efficient construction of gene library by partial filling-in of cohesive ends, *Gene* **42**, 119–123.

Zabarovsky, E.R., Boldog, F., Thompson, T., et al. (1990) Construction of a human chromosome

3 specific *Not*I linking library using a novel cloning procedure, *Nucleic Acids Res.* **18**, 6319–6324.

Zabarovsky, E.R., Boldog, F., Erlandsson, R., et al. (1991) A new strategy for mapping the human genome based on a novel procedure for constructing jumping libraries, *Genomics* **11**, 1030–1039.

Zabarovsky, E.R., Kashuba, V.I., Zakharyev, V.M., et al. (1994) Shot-gun sequencing strategy for long range genome mapping: first results, *Genomics* **21**, 495–500.

Zabarovsky, E.R., Winberg, G., Klein, G. (1993) The SK-diphasmids – vectors for genomic, jumping and cDNA libraries, *Gene* **127**, 1–14.

Zardo, G., Tiirikainen, M.I., Hong, C., et al. (2002) Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors, *Nat. Genet.* **32**, 453–458.

Zoubak, S., Clay, O., Bernardi, G. (1996) The gene distribution of the human genome, *Gene* **174**, 95–102.

Genomic Imprinting, Molecular Genetics of

Robert Feil, Yuji Goto, and David Umlauf
Centre National de la Recherche Scientifique, Montpellier, France

1 Genomic Imprinting 471
1.1 Embryological Evidence 472
1.2 Imprinted Chromosomal Domains 474

2 Imprinted Genes 476

3 Molecular Mechanisms 478
3.1 Imprinting-control Regions 478
3.2 Reading the Imprint 480

4 Imprinting and Disease 481

5 Evolution of Imprinting 483

Bibliography 484
Books and Reviews 484
Primary Literature 484

Keywords

Androgenetic Embryo
An embryo with two paternal genomes, and no maternal genome, produced by nuclear transplantation.

Chromatin
DNA packaged around nucleosomes. The degree of packaging differs between active (euchromatic) and inactive (heterochromatic) chromosomal regions.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

DNA Methylation
Attachment of methyl (CH_3) groups to the bases of DNA. In mammals, DNA methylation occurs at cytosines that are followed by guanines (at CpG dinucleotides).

Epigenetic Modification
Any heritable, but reversible, alteration of DNA or associated nucleosomes above the level of the DNA sequence. This additional layer of information may indicate the parental origin of the chromosome.

Genomic Imprinting
A parent-of-origin-dependent mechanism whereby certain gene loci become expressed from only the maternal or only the paternal chromosome.

Histone Modification
The histones in nucleosomes can be altered by covalent modifications. At imprinting-control regions, these modifications are different between the parental alleles.

Imprinting and Behavior
Some imprinted domains are associated with behavioral phenotypes, and genetic disruption of certain imprinted genes gives aberrant behavior.

Imprinting and Cancer
The epigenetic maintenance of imprinting is frequently deregulated in cancer. Since imprinted genes are important in cell proliferation and differentiation, such deregulation is probably involved in the process of tumorigenesis.

Imprinting and Growth
Many imprinted genes influence fetal growth and development. Imprinted genes that enhance growth are mostly expressed from the paternal allele. Several other imprinted genes, which reduce growth, are expressed from the maternal allele.

Imprinting-control Regions
DNA sequence elements that are essential for imprinted gene expression. They are modified by DNA methylation and epigenetic modifications on the chromatin.

Nucleosome
The basic structural unit of chromatin, consisting of ~150 bp of DNA wrapped around an octamer of histone proteins (two each of four different histones).

Nutrient Transfer
Imprinted genes are important for the development of the extraembryonic membranes. These are essential for nutrient transfer to the developing embryo.

Parthenogenesis
The derivation of offspring from eggs only. Parthenogenesis is viable in some animal groups, such as in bird species, but is embryonic lethal in mammals because of the functional nonequivalence of the maternal and the paternal genome.

Uniparental Disomy
Inheritance of a particular chromosome in two copies from one parent, with absence of the chromosome from the other parent.

Genomic imprinting is a developmental mechanism in mammals and other organisms leading to repression or expression of genes depending on whether they are inherited from the mother or the father. The imprinted expression of genes is regulated by various epigenetic alterations, including DNA methylation and covalent modifications at histones. A large number of imprinted genes have been identified in placental mammals. Mostly clustered in the genome, these play important roles in embryonic and extraembryonic development, and in behavior. In humans, genetic and epigenetic alterations at imprinted genes are involved in different disease syndromes and in cancer.

1 Genomic Imprinting

Gene expression is not determined solely by the DNA code itself. It depends also on different epigenetic features. The term *epigenetic* is used to refer to mechanisms that do not involve changes in the DNA sequence and that are heritable from one cell generation to the next. Unlike heritable changes due to mutation or directed gene rearrangement (such as in the immunoglobulin genes), epigenetic modifications are reversible and can be removed from genes and chromosomes without leaving behind any permanent change to the genetic material. The main epigenetic modifications by which gene expression can be altered are DNA methylation and modifications to the chromatin. A well-known epigenetic mechanism is X-chromosome inactivation. In mammalian X-chromosome inactivation, sequential epigenetic modifications lead to the (random) transcriptional repression of one of the two X-chromosomes in all the somatic cells of females.

In this article, we consider a particular class of epigenetic imprints, those that mark the parental origin of genomes, chromosomes, and genes. Genes regulated by such "genomic imprinting" are expressed depending on whether they are on the maternally or on the paternally derived chromosome. Some imprinted genes are expressed only from the paternal chromosome, whereas others are exclusively expressed from the maternal chromosome. During the last fifteen to twenty years, imprinting has evolved from the initial observations in mouse embryos to a rapidly expanding field with importance

for mammalian development and genetics, and human disease. A large number of imprinted genes have now been identified. In addition, molecular studies have unraveled the underlying molecular mechanisms. Imprinting is not unique to mammals but is known to occur in seed plants and invertebrate species as well. This article, however, focuses on the regulation and role of autosomal imprinted genes in mammals.

Following the discovery of genomic imprinting, and the identification of the first imprinted genes in mammals, researchers in the field hypothesized that the epigenetic marks that regulate parent-of-origin-dependent expression are established in either the female or the male germ line and (after fertilization) are maintained throughout development. This epigenetic information needs to be removed upon passage of the imprinted gene through the germ line in the developing fetus, however, so that new imprints can be established. Recent studies on DNA methylation and other epigenetic modifications showed that, indeed, there are three distinct phases in imprinting: establishment of the imprint in the male or female germ line, somatic maintenance of the imprint after fertilization, and its erasure upon (re-)passage through the germ line (Fig. 1).

1.1 Embryological Evidence

Embryological studies in the mouse provided the first evidence that, in mammals, both a maternal and a paternal genome are required for the production of viable offspring. Significantly, it was found that monoparental embryos, carrying either two maternal or two paternal genomes, cannot develop to term. Such monoparental embryos were obtained by nuclear transplantation, immediately following the fertilization of the egg by the sperm. By replacing the female pronucleus (female genome) with a male pronucleus (male genome), for example, it was possible to produce androgenetic embryos (which have two paternal genomes). Conversely, embryos with two maternal genomes (gynogenotes) were made by replacing the male pronucleus with a female pronucleus. Embryos with two maternal genomes were derived by activation of unfertilized eggs (parthenogenesis) as well. Intriguingly, both gynogenetic (parthenogenetic) and androgenetic embryos survived only for a few days after implantation in the uterus and were found to have major developmental abnormalities.

Gynogenetic (parthenogenetic) and androgenetic embryos have different developmental phenotypes (Fig. 2). After

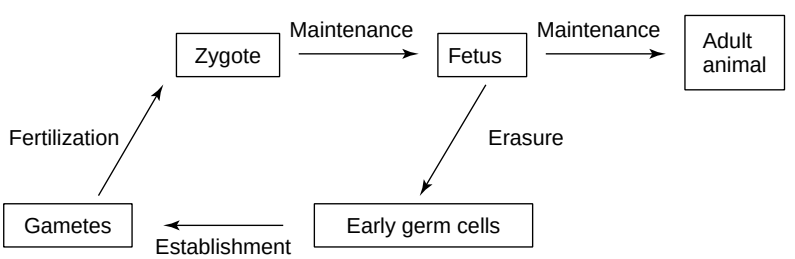

Fig. 1 Ontogeny of genomic imprinting: germ line establishment, somatic maintenance, and erasure.

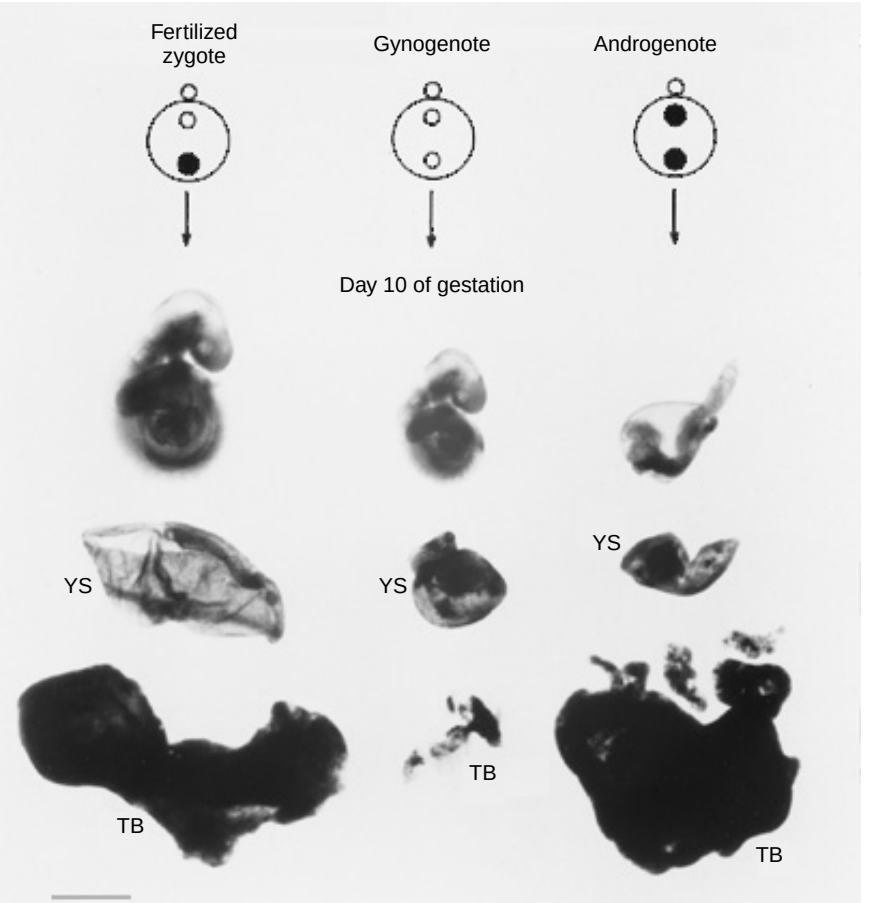

Fig. 2 Normal, androgenetic, and gynogenetic mouse embryos and their extraembryonic membranes at day 10 of gestation. Shown are the embryo, the yolk sac (YS), and the trophoblast (TB).

implantation into recipient females, parthenogenetic conceptuses develop to only about day 10 of gestation, with an apparently normal but small embryo. Development of the extraembryonic membranes (yolk sac and trophoblast), in contrast, is severely deficient, and these are the tissues that are important for nutrient transfer to the embryo. The phenotype of androgenetic conceptuses is opposite to that of parthenogenetic conceptuses. Whereas the extraembryonic tissues are normal in the androgenetic conceptuses, the embryo proper is retarded and progresses rarely beyond the four- to six-somite stage. The investigations on monoparental embryos established that both the parental genomes are required for normal mammalian development. They also provided evidence for the existence of genetic loci at which expression depends on the parental origin of the gene. In parthenogenetic and androgenetic embryos, individual imprinted genes are either expressed from both the

gene copies (double gene dose) or are not expressed at all. Cumulatively, the aberrant levels of expression of imprinted genes are responsible for the striking phenotypes of the two types of monoparental embryos.

1.2 Imprinted Chromosomal Domains

The embryological evidence from the monoparental embryos was reinforced by genetic studies demonstrating that specific chromosomal domains are subject to imprinting. Particularly, mice that were heterozygous for chromosomal translocations were intercrossed to obtain embryos and offspring with uniparental disomy for individual chromosomes (or chromosomal regions). Since during meiosis there is sometimes nondisjunction at the chromosome with the translocation, some of the resulting gametes comprise two copies of the translocated chromosome, whereas others contain none. Embryos that arise from two of such opposite gametes will have two copies of all the chromosomes, but for the translocated chromosome, both the copies will be paternal or maternal. By using different translocation lines, such uniparental disomic embryos were generated for almost all autosomal chromosomes. Phenotypic analyses unraveled the role of subsets of imprinted genes that reside in two paternal or two maternal copies in the different uniparental disomies. These studies also revealed that the maternal and paternal copies of individual chromosomal regions have frequently opposite roles in development and after birth (Fig. 3).

One of the imprinted domains is on the distal portion of mouse chromosome 7. When present in two maternal copies (maternal disomy), it leads to reduced growth and fetal death, whereas paternal disomy of this distal region is associated with enhanced growth and embryonic death. Some 10 imprinted genes have been mapped to this region. Several of these are part of the insulin-like growth factor/insulin signaling pathway (IGF/INS pathway). Being key players in the regulation of fetal growth and development, they contribute to the opposite growth phenotypes in the maternal and paternal distal-7 disomies. The corresponding chromosomal region in humans, chromosome 11p15.5, is involved in the Beckwith–Wiedemann syndrome (BWS), a human growth disorder that can be caused by paternal disomy of this imprinted region.

Another chromosomal domain with opposite phenotypes in paternal and maternal disomies is the proximal portion of chromosome 11. Mice with paternal disomy of this region are larger than their normal littermates, whereas maternal disomy mice are smaller. This indicates that there are imprinted genes in this region, of which aberrant levels of expression in the maternal and paternal disomies cause their abnormal growth. So far, two imprinted genes have been identified in this domain, *U2af1-rs1* and *Grb10*. The latter could be responsible for the phenotypes of the maternal and paternal disomies. Its main embryonic transcript is expressed from the maternal allele only, and it encodes a protein with a negative effect on the growth-regulating IGF/INS pathway.

Disomy phenotypes at a few other imprinted domains involve abnormal postnatal behavior. Paternal disomy for distal mouse chromosome 2, for instance, gives offspring that are hyperactive, whereas maternal disomy is associated with reduced activity after birth. Such behavioral phenotypes emphasize that imprinted genes can affect behavior. A small number of

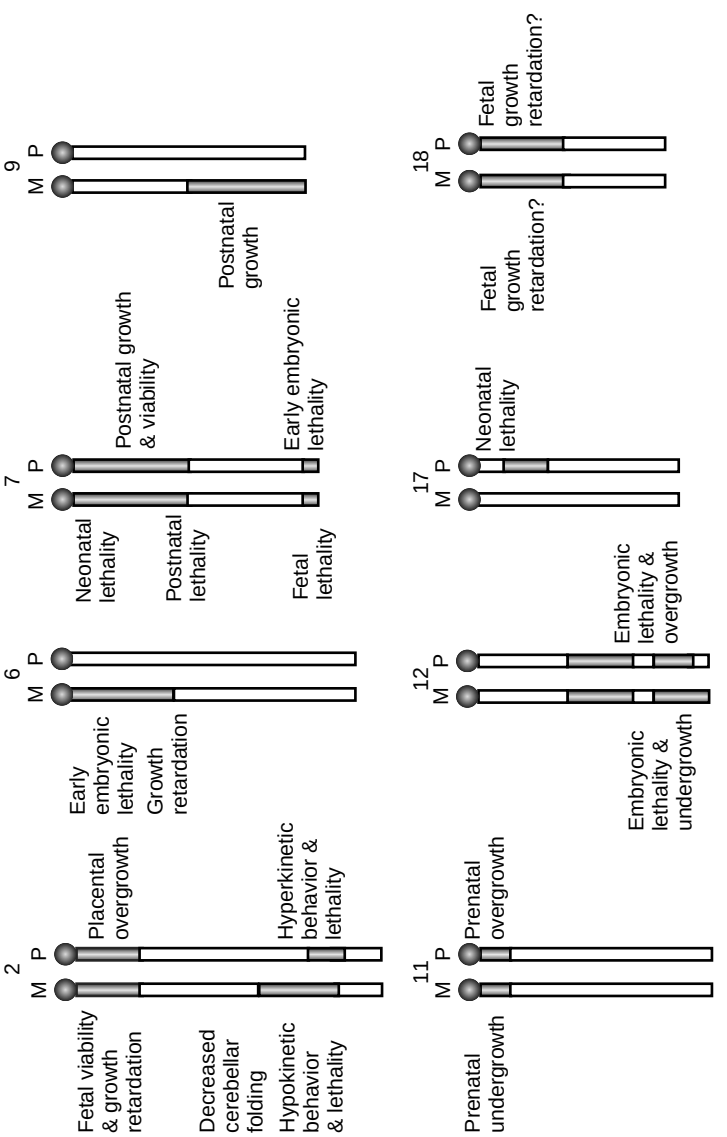

Fig. 3 Imprinted chromosomal domains in the mouse with the associated developmental phenotypes in maternal (indicated to the left) and paternal (to the right) disomies.

imprinted genes were discovered on distal mouse chromosome 2. Two of these have neuroendocrine functions (*Gnas* and *Gnas-xl*) and are involved in the behavioral phenotypes of the maternal and paternal disomy mice.

In total, 12 chromosomal regions with imprinting phenotypes have been identified on 8 different autosomal chromosomes (Fig. 3). The large majority of the known imprinted genes maps to these chromosomal regions. Probably, the remainder of the genome comprises few imprinted genes or contains imprinted genes that give rise to minor phenotypes only when present in two maternal or two paternal copies.

2 Imprinted Genes

It is unknown which proportion of mammalian genes is imprinted and estimates vary between about 100 and a 1000 genes. To date, however, some 70 imprinted genes have been detected in the mouse and most of these are imprinted in humans as well. A consistent feature of imprinted genes is that they are organized in clusters in the genome. These clusters are hundreds to several thousands of kilobases in size and are similarly organized in humans and mice. We selected several imprinted clusters as examples, and we describe their roles in development and behavior. A comprehensive presentation of imprinted genes is given elsewhere.

A well-known imprinted cluster is on distal mouse chromosome 7 (Fig. 4) and on the corresponding chromosome 11p15.5 in humans. This cluster comprises 11 imprinted genes. Several of these genes play key roles in fetal growth and development. The insulin-like growth factor-2 gene (*Igf2*), at the proximal side of the cluster, is expressed from the paternal allele only. Transgenic mice inheriting a null *Igf2* allele from the father are much smaller than their littermates; maternal inheritance of the targeted allele does not alter the phenotype. This strong paternal effect on fetal growth is primarily due to the loss of IGF2 in the extraembryonic membranes, which decreases nutrient transfer to the developing fetus. The neighboring insulin-2 gene (*Ins2*), also of the IGF/INS pathway, is located at about 20 kb from *Igf2*. In the yolk sac, it is the paternal chromosome that expresses *Ins2*, whereas the maternal chromosome is repressed. The paternal expression of *Igf2* and *Ins2* is regulated by an "imprinting-control region" downstream of *Igf2*, close to a maternally expressed imprinted gene (*H19*) that produces a noncoding RNA. At the distal side of the cluster, the *Cdkn1c* gene (also called *p57Kip2*) codes for a cyclin-dependent kinase inhibitor. This imprinted gene is expressed from the maternal allele only. When *Cdkn1c* expression is ablated by gene targeting in the mouse, offspring are enhanced in size and also display other similarities to the Beckwith–Wiedemann syndrome in humans. Interestingly, a similar growth phenotype arises as a consequence of *Igf2* overexpression in mice. One role of CDKN1C could therefore be to inhibit the growth-promoting action of IGF2. Thus, several genes in the imprinted gene cluster are involved in the regulation of fetal growth and seem part of the same signaling pathway.

Four of the other imprinted genes at the distal 7 cluster display allelic expression in the extraembryonic tissues, in which they are expressed from the maternal allele. One of these, *Ascl2* (also named *Mash2*) encodes a transcription factor that is essential for placental development. Genetic studies

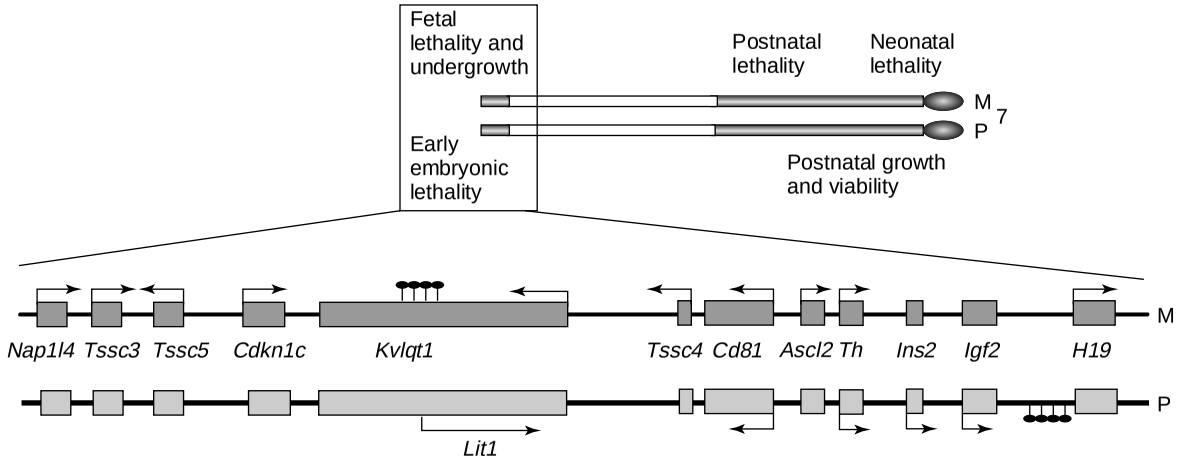

Gene	Product	Function
Tssc3	Pleckstrin homology-domain protein	Placental development
Cdkn1c	Cyclin-dependent kinase inhibitor	Tumor suppressor gene
Kvlqt1	Voltage-dependent potassium channel	Heart and ear function
Lit1	Noncoding RNA	Imprinting regulator?
Ascl2	Transcription factor	Placental development
Ins2	Growth factor	Embryonic growth
Igf2	Growth factor	Embryonic growth
H19	Noncoding RNA	Translational control?

Fig. 4 The imprinted domain on distal mouse chromosome 7 (human chromosome 11q15). Shown are the imprinted genes and their known functions. Lollypops indicate the allele-specific DNA methylation at the two imprinting-control regions.

show that the imprinting of the four extraembryonic genes, and that of *Cdkn1c*, is regulated by a second "imprinting-control region," which is located in the central portion of the cluster.

Amongst the imprinted genes that influence the IGF/INS pathway, there is also the IGF2-receptor gene (*Igf2r*) on mouse chromosome 17. *Igf2r* is expressed exclusively from the maternal allele and exerts a negative effect on growth by reducing the levels of active IGF2. Whereas most imprinted mouse genes are imprinted in humans as well, *Igf2r* is one of the exceptions. In humans, this gene is expressed from both the parental alleles.

Some chromosomal domains comprise imprinted genes that are expressed predominantly in the brain. One of these clusters maps to the central portion of mouse chromosome 7 (and to human chromosome 15q11–q13) and comprises a large number of genes that are all expressed from the paternal chromosome only (Fig. 5c). In humans, loss of expression at these genes (*SNRPN, ZNF127, NDN*, and others) leads to the Prader–Willi syndrome, a variable disorder that is partly due to a hypothalamic defect (see below). Biallelic expression of the genes and loss of expression of a neighboring imprinted gene (*UBE3A*) is associated with the clinically distinct Angelman syndrome (AS). The regulation of imprinting in this domain is complex and involves at least two distinct genetic elements.

A minority of imprinted genes are not part of an imprinted gene cluster. One of these is the U2af1-related sequence-1 gene (*U2af1-rs1*) on proximal mouse chromosome 11. This intronless gene is repressed on the maternal chromosome and encodes a brain-specific RNA splicing factor homologous to the splicing factor U2AF. The imprinted *U2af1-rs1* gene has arisen via a retrotransposition event in rodents, and in humans there is no equivalent imprinted gene.

3 Molecular Mechanisms

3.1 Imprinting-control Regions

The expression of imprinted genes is regulated by epigenetic modifications that mark the parental alleles to be active or repressed. These epigenetic modifications are put onto key regulatory elements, depending on the parental origin of the allele, and lead to the allelic gene expression.

At all imprinted loci, there are sequence elements at which DNA methylation is present on one of the two parental alleles only. At many of these "differentially methylated regions" (DMRs), the DNA methylation originates from either the egg or the sperm. After fertilization, the allelic methylation is maintained in the somatic cells. Regions with such a germ line methylation mark are essential in the control of imprinting. They are referred to as *imprinting-control regions*. Most imprinting-control regions are rich in CpG dinucleotides and correspond to CpG islands.

At the imprinted *U2af1-rs1* gene, DNA methylation is present exclusively on the repressed maternal allele (Fig. 5a). This differential DNA methylation becomes established during oogenesis along its CpG island, located at the 5′ side of the gene, and spreads throughout the entire maternal gene during early embryonic development.

The maternally expressed *Igf2r* gene has an imprinting-control region within the second intron that is methylated on the

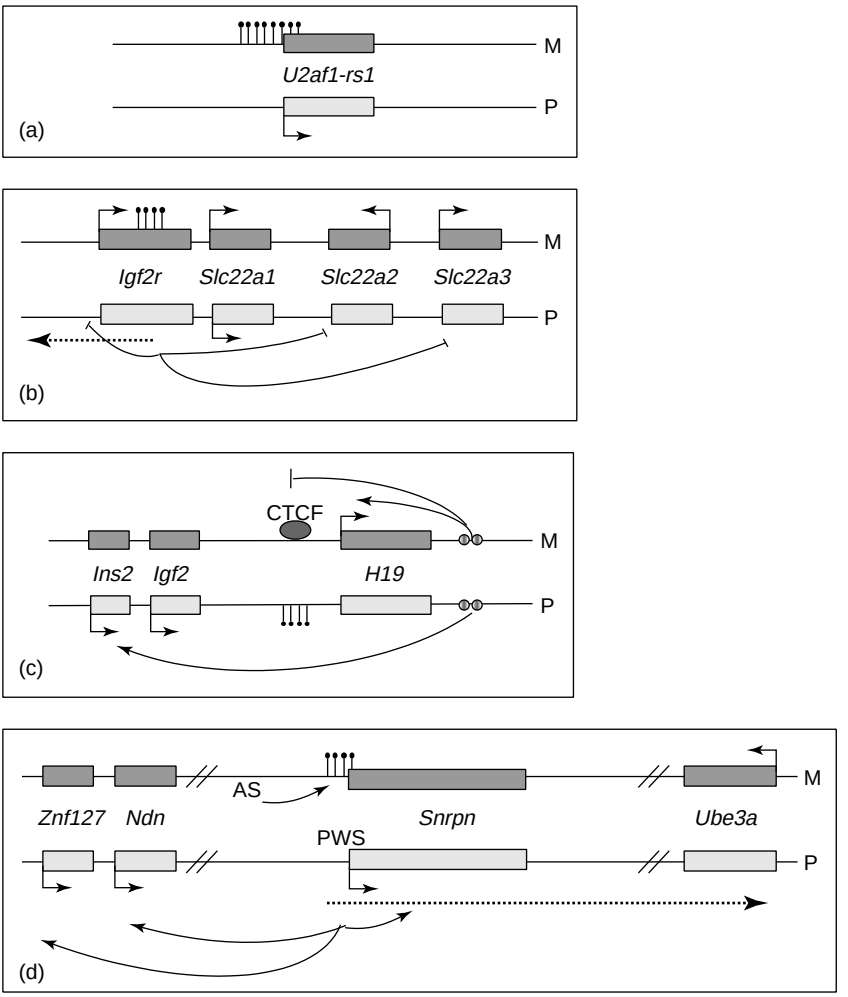

Fig. 5 Reading the imprint. Imprinting-control regions confer allelic gene expression at the (a) *U2af1-rs1* gene, (b) the Igf2r locus, (c) the imprinted cluster containing the *Igf2* gene, and (d) at the PWS/AS region in the mouse. Lollypops indicate the allele-specific DNA methylation at the imprinting-control regions. Antisense transcripts are shown as interrupted lines; circles indicate transcriptional enhancers.

maternal allele (Fig. 5b). This maternal methylation is established during oogenesis and is maintained in all the somatic lineages. The intronic imprinting-control region is essential for the allelic repression at the locus: removal by gene targeting leads to expression from both the parental alleles.

Most imprinting-control regions are methylated on the maternal allele. However, in some, the DNA methylation is found at the paternal allele, and it is the maternal allele that is unmethylated. One of these paternal methylation marks controls the allelic expression of the *Igf2* and *Ins2* genes on distal mouse chromosome 7

(Fig. 5c). This region, a CpG island located upstream of the close-by *H19* gene, acquires its DNA methylation during spermatogenesis. After fertilization, this paternal mark is maintained in all the somatic tissues. Deletion of the control region gives rise to biallelic expression of *Igf2* and *Ins2*.

It is unclear why imprinting-control regions attract DNA methylation in either the female or the male germ line. Several studies suggest, however, that close-by tandemly-repeated sequences might be essential in this choice process.

More is known about the DNA methyltransferases (DNMT) that are involved in the germ line establishment of the methylation marks. The methyltransferases DNMT3A and DNMT3B are essential in this process. In addition, a DNMT-like protein (DNMT3L) is required for the establishment of methylation imprints as well, particularly in the female germ line. Once established, allelic patterns of DNA methylation need to be maintained in the developing embryo. The maintenance methyltransferase, DNMT1, plays an important role in this process and differential chromatin features are likely to be involved as well. At imprinting-control regions, pronounced differences in histone modifications have been detected between the parental alleles. Levels of histone acetylation are low on the allele that comprises methylated DNA, whereas high levels of acetylation are present on the chromatin of the opposite, unmethylated, allele. In addition, there are strong allelic differences in histone methylation at specific lysine residues on histone H3. Whereas methylation of lysine residue 9 of H3 is detected on the parental allele that has DNA methylation, it is on the opposite parental allele (without DNA methylation) that there are high levels of H3 lysine-4 methylation. At several imprinted loci, there is also evidence for allele-specific chromatin compaction, occurring in association with the differential histone modifications.

It is not yet understood how DNA methylation, histone acetylation, and histone methylation are mechanistically linked at imprinting-control regions. However, at several imprinting-control regions, (allelic) DNA methylation was found to be linked to histone deacetylation (the removal of the acetyl group from the histones). This link is brought about by proteins that bind the methylated DNA and attract large protein complexes that comprise histone deacetylases. It is to be explored also to which extent the differential histone modifications are important in the somatic maintenance (and germ line establishment) of the allelic patterns of DNA methylation at imprinted loci. Nonhistone proteins, binding to the unmethylated allele of many imprinting-control regions, are likely to be involved in the maintenance of the allelic DNA methylation as well.

3.2 Reading the Imprint

Imprinting-control regions are comparable in that they all have allele-specific DNA methylation and differential chromatin organization. The way in which this gives rise to imprinted gene expression differs between loci. The simplest scenario, whereby differential methylation and associated chromatin features lead to imprinted gene expression, is that of the *U2af1-rs1* gene on mouse chromosome 11 (Fig. 5a). Here, methylated DNA and compacted chromatin are present across the promoter on the maternal allele. As a consequence, the gene can be transcribed from the paternal allele only.

The imprinting-control region of *Igf2r* regulates allelic expression in a rather different way (Fig. 5b). Here, the maternal methylation covers the promoter of an antisense transcript. As a consequence, this antisense transcript (named *Air*) is produced from the (unmethylated) paternal allele only. Via a yet-unclear mechanism, this paternal antisense transcript represses the paternal *Igf2r* gene and two flanking ion-transporter genes (*Slc22a2* and *Slc22a3*). A similar antisense transcript is produced at the imprinting-control region that regulates the extraembryonic tissue-specific imprinted genes on distal mouse chromosome 7 (Fig. 3).

Another example of how a germ line mark conveys imprinted expression is provided by the *Ins2-Igf2-H19* locus on distal mouse chromosome 7 (Fig. 5c). Here, the imprinting-control region located upstream of the noncoding *H19* gene, is methylated on the paternal allele and acts as a chromatin boundary on the unmethylated maternal allele. In fact, this upstream element has multiple recognition sites for a zinc finger protein called CTCF. The binding of CTCF is prevented by methylation. This chromatin protein is therefore associated with the unmethylated maternal allele only, at which it forms a specialized chromatin structure. This unusual structure insulates the *Igf2* and *Ins2* promoters from their enhancers (located downstream of *H19*). As a consequence, *Igf2* and *Ins2* are repressed on the maternal chromosome. This maternal repression is not exclusively at the transcriptional level but influences posttranscriptional processes as well.

The central portion of mouse chromosome 7 (Fig. 5d) corresponds to the Prader–Willi syndrome (PWS) region and Angelman syndrome (AS) region on human chromosome 15q11–13. The key regulatory element in this domain is the 5′ portion of the *SNRPN* gene and it is methylated on the maternal chromosome. This imprinting-control region is essential for the paternal expression of *SNRPN* and its flanking genes (including *Znf127* and *Ndn*). When the element is deleted on the paternal chromosome, these brain-specific genes are all no longer expressed. Upstream of the *SNRPN* gene, a paternal RNA of several hundreds of kilobases in size is produced as well. This transcript is in antisense orientation to a gene at the far extremity of the imprinted domain. This gene, *UBE3A*, is the only one in the cluster that is repressed on the paternal chromosome. The *SNRPN* imprinting-control region itself is regulated by a second control region, which is located further upstream, and is essential for the acquisition of the allelic DNA methylation at *SNRPN*. Precisely how the allelic expression and repression is brought about along thousands of kilobases remains to be unraveled. It has been observed, however, that there is differential timing of replication in the S-phase between the parental chromosomes along the entire domain. Such a differential replication timing has been detected at other imprinted loci as well. Future work should investigate the role of the differential replication timing and whether it reflects differential chromatin organization along entire imprinted domains.

4
Imprinting and Disease

In many genetic diseases, the clinical manifestations depend on whether the mutation is inherited from the mother or the father. Although imprinting is suspected to be involved, causal genes and molecular mechanisms are yet to

be identified for most of these disorders. Clinical phenotypes can be associated with uniparental disomies as well, similarly as in the mouse. Additionally, imprinting can become deregulated during embryonic development, by epigenetic alterations or by somatic mutations, resulting in loss or biallelic expression of imprinted genes. Such somatic loss of imprinting can result in specific disease phenotypes as well.

Beckwith–Wiedemann syndrome (BWS) is a fetal overgrowth syndrome with a high incidence of embryonal tumors, including Wilms' tumor of the kidney and rhabdomyosarcoma. Genetically, the syndrome is linked to the cluster of growth-related imprinted genes comprising *IGF2* (see Fig. 4). Paternal disomy of this domain is responsible for a proportion of cases and leads to a double dose of *IGF2* expression and loss of expression of *CDKN1C* and other maternally expressed genes in the cluster. BWS can also be caused by genetic mutations at *CDKN1C* and by alterations at the *KVLQT1* gene, where there is one of the two imprinting-control centers of the cluster. The finding that the growth syndrome can be caused by mutations at different places in the imprinted domain supports the idea that its genes are coregulated and involved in the same biological functions. The majority of the BWS cases are sporadic, however, and apparently without genetic mutations. These are mostly caused by epigenetic alterations in the developing embryo. In some of the sporadic cases, for instance, there is aberrant, biallelic methylation at the imprinting-control region at the *H19* gene (Fig. 5c). This results in expression of *IGF2* from both the parental chromosomes during development and therefore in fetal overgrowth.

The neurobehavioral Angelman syndrome (AS) includes mental retardation, ataxia, and hyperactivity and arises from maternal deletion or paternal disomy of the imprinted domain on chromosome 15q11–13. Prader–Willi syndrome (PWS), on the other hand, arises from paternal deletion or maternal disomy of this imprinted domain. This opposite syndrome involves mild mental retardation, obesity due to hyperphagia, and hypogonadism. Cases with small genetic deletions have been identified, and analysis of these patients has revealed that the PWS and AS are caused by distinct regions in the large imprinted domain (Fig. 5d). The smallest identified deletions in PWS remove the imprinting element at the 5′ portion of *SNRPN*. This gives loss of expression of *SNRPN, NDN, ZNF127*, and several other genes in the cluster. The smallest deletions in AS removes the control region that is essential for the establishment of the epigenetic imprint at *SNRPN*. Consequently, there is expression of *SNRPN, NDN*, and *ZNF127* from both the parental alleles and loss of expression of the *UBE3A* gene located at the end of the cluster. The latter seems to be the main cause of the clinical phenotype of AS.

Amongst other imprinting disorders are Albright Hereditary Osteodystrophy (AHO), linked to a cluster of imprinted neuroendocrinal genes on human chromosome 20q, and transient neonatal diabetes mellitus (TNDM), linked to chromosome 6q24–25. The latter is mostly sporadic and is caused by aberrant expression of the imprinted gene *ZAC*. This zinc finger protein–encoding gene has a CpG island with maternal DNA methylation. It was discovered recently that in cases of TNDM without genetic defects, this imprinting-control region has lost its methylation.

Epigenetic alterations at imprinting-control regions occur frequently in tumors

as well. This has been observed in Wilms' tumor of the kidney, but also in lung cancer, breast cancer, and various other cancers. In particular, *IGF2* was found to be expressed from both the parental alleles during tumorigenesis and this could confer a proliferative advantage to the cells. In many cases, the biallelic *IGF2* expression is caused by acquisition of DNA methylation at the imprinting-control region upstream of *H19*, similarly as in BWS. This epigenetic alteration occurs early in tumor formation and could be linked to the pathological tendency of tumorigenic cells to acquire methylation at CpG islands.

When early embryos are taken from their natural environment and put into a culture dish, it can lead to aberrant imprinting as well. This was observed in the mouse and in domestic animals. It is unclear, at present, whether loss of imprinting due to embryo culture is mechanistically comparable to that in human imprinting disorders or in tumors. However, culture of embryos and early embryonic cells can also induce aberrant DNA methylation at imprinting-control regions. This results in biallelic, or loss of, imprinted gene expression and can have pronounced phenotypic consequences at later developmental stages. An important issue to be investigated is whether there are culture conditions that do not affect imprinting and would thus be best suitable for *in vitro* culture and manipulation procedures in animals and humans.

5 Evolution of Imprinting

There is a lot of interest in how broadly imprinting is conserved amongst mammals. Also, in species other than the mouse, both the parental genomes are essential for normal development. Parthenogenesis, for instance, leads to embryonic lethality in humans, pigs, and sheep. In the latter (ruminant) species, parthenogenetic conceptuses die shortly after implantation, due to deficient development and functioning of the extraembryonic membranes. These studies indicate that imprinting is conserved amongst different groups of mammals. Indeed, most of the known imprinted mouse genes are also imprinted in humans and, as far as this has been analyzed, in other placental mammals as well.

Evolutionary biologists have proposed several hypotheses to explain why imprinting has arisen in placental mammals and to account for the different imprinting-related phenotypes. In placental mammals, there is continuous transfer of nutrients from the mother animal to the developing offspring, and this determines their development and growth. Possibly, the most attractive theory of imprinting says that paternally inherited genes tend to increase nutrient transfer and thereby the growth of the developing fetus. This would enhance their chances of being propagated to future generations. Maternally derived genes, however, would be best propagated by limiting the growth of the developing fetus. This is because too high a burden of nutrient transfer compromises the reproductive success of the mother animal and hence of all its offspring. During the evolution of placental mammals, there would therefore have been a 'parental tug-of-war' between these opposing maternal and paternal strategies, leading to balanced combinations of expression levels of maternally and paternally derived genes. As outlined with different examples, imprinted genes such as *Igf2, Ins2*, and *Igf2r* indeed play important roles in nutrient

transfer and growth, for instance, by promoting or reducing the development of the extraembryonic tissues. Other imprinted genes are important in determining the activity of the newborn animals, which, again, could have an impact on nutrient transfer, but now after birth.

See also Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Beechey, C.V., Cattanach, B.M.C., Selley, R.L. (2002) MRC Mammalian Genetics Unit, Harwell, United Kingdom. World Wide Web site – Mouse imprinting data and references (http://www.mgu.har.mrc.ac.uk/imprinting/imptables.html).

Feil, R., Khosla, S. (1999) Genomic imprinting in mammals: an interplay between chromatin and DNA methylation? *Trends Genet.* **15**, 431–435.

Lee, J.T. (2003) Molecular links between X-inactivation and autosomal imprinting: X inactivation as the driving force for the evolution of imprinting? *Curr. Biol.* **13**, R242–R254.

Ohlsson, R., Tycko, B., Sapienza, C. (1998) Mono-allelic expression: 'there can only be one', *Trends Genet.* **14**, 435–438.

Reik, W., Walter, J. (2001) Genomic imprinting: parental influence on the genome, *Nat. Rev. Genet.* **2**, 21–32.

Sleutels, F., Barlow, D.P., Lyle, R. (2000) The uniqueness of the imprinting mechanism, *Curr. Opin. Genet. Dev.* **10**, 229–233.

Surani, M.A. (1998) Imprinting and the initiation of gene silencing in the germ line, *Cell* **93**, 309–312.

Tilghman, S.M. (1999) The sins of the fathers and mothers: genomic imprinting in mammalian development, *Cell* **96**, 185–193.

Wilkins, J.F., Haig, D. (2003) What good is genomic imprinting? *Nat. Rev. Genet.* **4**, 359–368.

Primary Literature

Barlow, D.P., Stöger, R., Herrmann, B.G., Saito, K., Schweifer, N. (1991) The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the *Tme* locus, *Nature* **349**, 84–87.

Bartolomei, M.S., Zemel, S., Tilghman, S.M. (1991) Parental imprinting of the mouse *H19* gene, *Nature* **351**, 153–155.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., Bestor, T.H. (2001). *Dnmt3l* and the establishment of maternal genomic imprints, *Science* **294**, 2536–2539.

Brandeis, M., Kafri, T., Ariel, M., Chaillet, J.R., McCarrey, J., Razin, A., Cedar, H. (1993) The ontogeny of allele-specific methylation associated with imprinted genes in the mouse, *EMBO J.* **12**, 3669–3677.

Buiting, K., Saitoh, S., Gross, S., Dittrich, B., Schwartz, S., Nicholls, R.D., Horsthemke, B. (1995) Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15, *Nat. Genet.* **9**, 395–400.

Cattanach, B.M., Kirk, M. (1985) Differential activity of maternally and paternal derived chromosome regions in mice, *Nature* **315**, 496–498.

Cavaillé, J., Seitz, H., Paulsen, M., Ferguson-Smith, A.C., Bachellerie, J.P. (2002) Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region, *Hum. Mol. Genet.* **11**, 1527–1538.

Chaillet, J.R., Vogt, T.F., Beier, D.R., Leder, P. (1991) Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis, *Cell* **66**, 77–83.

Charlier, C., Segers, K., Karim, L., Shay, T., Gyapay, G., Cockett, N., Georges, M. (2001) The callipyge mutation enhances the expression of coregulated imprinted genes in cis without affecting their imprinting status, *Nat. Genet.* **27**, 367–369.

Cockett, N.E., Jackson, S.P., Shay, T.L., Farnir, F., Berghmans, S., Showder, G.D., Nielsen, D.M., Georges, M. (1996) Polar overdominance at the ovine callipyge locus, *Science* **273**, 236–238.

Constancia, M., Hemberger, M., Hughes, J., Dean, W., Ferguson-Smith, A., Fundele, R.,

Stewart, F., Kelsey, G., Fowden, A., Sibley, C., Reik, W. (2002) Placental-specific IGF-II is a major modulator of placental and fetal growth, *Nature* **417**, 945–948.

Cui, H., Cruz-Correa, M., Giardello, F.M., Hutcheon, D.F., Kafenok, D.R., Brandenburg, S., Wu, Y., He, X., Powe, N.R., Feinberg, A.P. (2003) Loss of IGF2 imprinting: a potential marker of colorectal cancer risk, *Science* **299**, 1753–1755.

Dean, W., Bowden, L., Aitchison, A., Klose, J., Moore, T., Meneses, J.J., Reik, W., Feil, R. (1998) Altered imprinted gene methylation and expression in completely ES cell-derived mouse fetuses: association with aberrant phenotypes, *Development* **125**, 2273–2282.

Debaun, M.R., Niemitz, E.L., Feinberg, A.P. (2003) Association of in vitro fertilization and Beckwith-Wiedemann syndrome and epigenetic alterations of LIT1 and H19, *Am. J. Hum. Genet.* **72**, 156–160.

DeChiara, T.M., Robertson, E.J., Efstratiadis, A. (1991) Parental imprinting of the mouse insulin-like growth factor II, *Cell* **64**, 849–859.

Doherty, A.S., Mann, M.R., Tremblay, K.D., Bartolomei, M.S., Schultz, R.M. (2000) Differential effects of culture on imprinted *H19* expression in the pre-implantation embryo, *Biol. Reprod.* **62**, 1526–1535.

Feil, R., Walter, J., Allen, N.D., Reik, W. (1994) Developmental control of allelic methylation in the imprinted mouse *Igf2* and *H19* genes, *Development* **120**, 2933–2943.

Feil, R., Boyano, M.D., Allen, N.D., Kelsey, G. (1997) Parental chromosome-specific chromatin conformation in the imprinted *U2af1-rs1* gene in the mouse, *J. Biol. Chem.* **272**, 20893–20900.

Ferguson-Smith, A.C., Cattanach, B.M., Barton, S.C., Beechey, C.V., Surani, M.A. (1991) Embryological and molecular investigations of parental imprinting on mouse chromosome 7, *Nature* **352**, 609–610.

Ferguson-Smith, A.C., Sasaki, H., Cattanach, B.M., Surani, M.A. (1993) Parent-origin-specific epigenetic modification of the mouse *H19* gene, *Nature* **362**, 751–755.

Fitzpatrick, G.V., Soloway, P.D., Higgins, M.J. (2002) The brain on microarrays, *Nat. Genet.* **32**, 426–431.

Fournier, C., Goto, Y., Ballestar, E., Delaval, K., Hever, A.M., Esteller, M., Feil, R. (2002) Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes, *EMBO J.* **23**, 6560–6570.

Gregory, R.I., Randall, T.E., Johnson, C.A., Khosla, S., Hatada, I., O'Neill, L.P., Turner, B.M., Feil, R. (2001) DNA methylation is linked to deacetylation of histone H3, but not H4, on the imprinted genes *Snrpn* and *U2af1-rs1*, *Mol. Cell. Biol.* **21**, 5426–5436.

Gunaratne, P.H., Nakao, M., Ledbetter, D.H., Sutcliffe, J.S., Chinault, A.C. (1995) Tissue-specific and allele-specific replication timing control in the imprinted human Prader-Willi syndrome region, *Genes Dev.* **9**, 808–820.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El Maarri, O., Reik, W., Walter, J., Surani, A. (2002) Epigenetic reprogramming in mouse primordial germ cells, *Mech. Dev.* **117**, 15–23.

Hata, K., Okano, M., Lei, H., Li, E. (2002) Dnmt-3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice, *Development* **129**, 1983–1993.

Hatada, I., Nabetani, A., Arai, Y., Ohishi, S., Suzuki, M., Miyabara, S., Nishimune, Y., Mukai, T. (1997) Aberrant methylation of an imprinted gene *U2af1-rs1* (SP2) caused by its own transgene, *J. Biol. Chem.* **272**, 9120–9122.

Howell, C.Y., Bestor, T.H., Ding, F., Latham, K.E., Mertineit, C., Trasler, J.M., Chaillet, J.R. (2001) Genomic imprinting disrupted by a maternal effect mutation in the *Dnmt1* gene, *Cell* **104**, 829–838.

Humpherys, D., Eggan, K., Akutsu, H., Hochedlinger, K., Rideout, W.M., Biniszkiewcz, D., Yanagimachi, R., Jaenisch, R. (2001) Epigenetic instability in ES cells and cloned mice, *Science* **293**, 95–97.

Inoue, K., Kohda, T., Lee, J., Ogonuki, N., Mochida, K., Noguchi, Y., Tanemura, K., Kaneko-Ishino, T., Ishino, F., Ogura, A. (2002) Faithful expression of imprinted genes in cloned mice, *Science* **295**, 297.

Izumikawa, Y., Naritoma, K., Hariyama, K. (1991) Replication asynchrony between homologs 15q11.2: cytogenetic evidence for genomic imprinting, *Hum. Genet.* **87**, 1–5.

Jouvenot, Y., Poirier, F., Jami, J., Paldi, A. (1999) Bi-allelic transcription of *Igf2* and *H19* in individual cells suggests a post-transcriptional contribution to genomic imprinting, *Curr. Biol.* **9**, 1199–1202.

Judson, H., Hayward, B.E., Sheridan, E., Bonthron, D.T. (2002) A global disorder of imprinting in the human female germ line, *Nature* **416**, 539–542.

Khosla, S., Aitchison, A., Gregory, R., Feil, R. (1999) Parental allele-specific chromatin configuration in an insulator/imprinting control element upstream of the mouse *H19* gene, *Mol. Cell. Biol.* **19**, 2556–2566.

Khosla, S., Dean, W., Brown, D., Reik, W., Feil, R. (2001) Culture of pre-implantation mouse embryos affects fetal development and the expression of imprinted genes, *Biol. Reprod.* **64**, 918–926.

Killian, J.K., Byrd, J.C., Jirtle, J.V., Munday, B.L., Stoskopf, M.K., MacDonald, R.G., Jirtle, R.L. (2000) M6P/IGF2R imprinting evolution in mammals, *Mol. Cells* **5**, 707–716.

Kitsberg, D., Selig, S., Brandeis, M., Simon, I., Keshet, I., Driscoll, D.J., Nicholls, R.D., Cedar, H. (1993) Allele-specific replication timing of imprinted gene regions, *Nature* **364**, 459–463.

Knoll, J.H., Cheng, S.D., Lalande, M. (1994) Allele specificity of DNA replication timing in the Angelman/Prader-Willi syndrome imprinted chromosomal region, *Nat. Genet.* **6**, 41–46.

Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., Ogur, S., Ishino, F. (2002) Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells, *Development* **129**, 1807–1817.

Li, E., Beard, C., Jaenisch, R. (1993) Role for DNA methylation in genomic imprinting, *Nature* **366**, 362–365.

Mager, J., Montgomery, N.D., de Villena, F.P., Magnuson, T. (2003) Genome imprinting regulated by the mouse Polycomb group protein Eed, *Nat. Genet.* **33**, 502–507.

McGrath, J., Solter, D. (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes, *Cell* **37**, 179–183.

Milligan, L., Forné, T., Antoine, E., Weber, M., Hémonnot, B., Dandolo, L., Brunel, C., Cathala, G. (2002) Turnover of primary transcripts is a major step in the regulation of mouse *H19* gene expression, *EMBO Rep.* **3**, 774–779.

Muscatelli, F., Abrous, D.N., Massacrier, A., Boccaccio, I., Le Moal, M., Cau, P., Cremer, H. (2000) Disruption of the mouse *Necdin* gene results in hypothalamic and behavioral alterations reminiscent of the human Prader-Willi syndrome, *Hum. Mol. Genet.* **12**, 3101–3110.

Nabetani, A., Hatada, I., Morisaki, H., Oshimura, M., Mukai, T. (1997) Mouse *U2af1-rs1* is a neomorphic imprinted gene, *Mol. Cell. Biol.* **17**, 789–798.

Paulsen, M., El-Maarri, O., Engemann, S., Strodicke, M., Franck, O., Davies, K., Reinhardt, R., Reik, W., Walter, J. (2000) Sequence conservation and variability of imprinting in the Beckwith-Wiedemann syndrome cluster in human and mouse, *Hum. Mol. Genet.* **9**, 1829–1841.

Pant, V., Mariano, P., Kanduri, C., Mattson, A., Lobanenkov, V., Heuchel, R., Ohlsson, R. (2003) The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the *H19* imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains, *Genes Dev.* **17**, 586–590.

Perk, J., Makedonski, K., Lande, L., Cedar, H., Razin, A. (2002) The imprinting mechanism of the Prader-Willi/Angelman regional control center, *EMBO J.* **21**, 5807–5814.

Rainier, S., Johnson, L.A., Dobry, C.J., Ping, A.J., Grundy, P.E., Feinberg, A.P. (1993) Relaxation of imprinted genes in human cancer, *Nature* **362**, 747–749.

Reik, W., Brown, K.W., Schneid, H., Le Bouc, Y., Bickmore, W., Maher, E.R. (1995) Imprinting mutations in the Beckwith-Wiedemann syndrome suggested my altered imprinting patterns in the IGF2-H19 domain, *Hum Mol. Genet.* **4**, 2379–2385.

Reik, W., Collick, A., Norris, M.L., Barton, S.C., Surani, M.A. (1987) Genomic imprinting determines methylation of parental alleles in transgenic mice, *Nature* **328**, 248–251.

Ripoche, M.A., Kress, C., Poirier, F., Dandolo, L. (1997) Deletion of the *H19* transcription unit reveals the existence of a putative imprinting control element, *Genes Dev.* **11**, 1596–1604.

Sasaki, H., Jones, P.A., Chaillet, J.R., Ferguson-Smith, A.C., Barton, S.C., Reik, W., Surani, M.A. (1992) Parental imprinting: potentially active chromatin of the repressed maternal allele of the mouse insulin-like growth factor II (Igf2) gene, *Genes Dev.* **6**, 1843–1856.

Schoenherr, C.J., Levorse, J.M., Tilghman, S.M. (2002). CTCF maintains differential methylation at the *Igf2/H19* locus, *Nat. Genet.* **33**, 66–69.

Simon, I., Tenzen, T., Reubinoff, B.E., Hillman, D., McCarrey, J.R., Cedar, H. (1999) Asynchronous replication of imprinted genes is established in the gametes and maintained during development, *Nature* **401**, 929–932.

Sleutels, F., Zwart, R., Barlow, D.P. (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes, *Nature* **415**, 810–813.

Smith, R.J., Dean, W., Konfortova, G., Kelsey, G. (2003) Identification of novel imprinted genes in a genome-wide screen for maternal methylation, *Genome Res.* **13**, 558–569.

Stöger, R., Kubicka, P., Liu, C.G., Kafri, T., Razin, A., Cedar, A., Barlow, D. (1993). Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal, *Cell* **73**, 61–71.

Strain, L., Warner, J.P., Johnston, T., Bonthron, D.T. (1995) A human parthenogenetic chimaera, *Nat. Genet.* **11**, 111–113.

Surani, M.A.H., Barton, S.C., Norris, M.L. (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis, *Nature* **308**, 548–550.

Szabo, P., Tang, S.H., Rentsendorj, A., Pfeifer, G.P., Mann, J.R. (2000) Maternal-specific footprints at putative CTCF sites in the *H19* imprinting-control region give evidence for insulator function, *Curr. Biol.* **10**, 607–610.

Swart, R., Sleutels, F., Wutz, A., Schinkel, A.H., Barlow, D.P. (2001) Bidirectional action of the *Igf2r* imprint control element on upstream and downstream imprinted genes, *Genes Dev.* **15**, 2361–2366.

Takagi, N., Sasaki, M. (1975) Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse, *Nature* **256**, 640–642.

Tucker K.L., Beard C., Bausmann J., Jackson-Grusby L., Laird, P.W., Lei, H., Li, E., Jaenisch, R. (1996) Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not nonimprinted genes, *Genes Dev.* **10**, 1008–1020.

Varrault, A., Bilanges, B., Mackay, D.J., Basyuk, E., Ahr, B., Fernandez, C., Robinson, D.O., Bockaert, J., Journot, L. (2001) Characterization of the methylation-sensitive promoter of the imprinted ZAC gene supports its role in transient neonatal diabetes mellitus, *J. Biol. Chem.* **276**, 18653–18656.

Webber, A.L., Ingham, R.S., Levorse, J.M., Tilghman, S.M. (1998) Location of enhancers is essential for the imprinting of H19 and Igf2 genes, *Nature* **391**, 711–715.

Weksberg, R., Shen, D.R., Fei, Y.L., Song, Q.L., Squire, J. (1993) Disruption of insulin-like growth factor 2 imprinting in Beckwith-Wiedemann syndrome, *Nat. Genet.* **5**, 143–150.

Xin, Z., Allis, C.D., Wagstaff, J. (2001) Parent-specific complementary patterns of histone H3 lysine 9 and H3 lysine 4 methylation at the Prader-Willi syndrome imprinting center, *Am. J. Hum. Genet.* **69**, 1389–1394.

Yokomine, T., Kuroiwa, A., Tanaka, K., Tsudzuki, M., Matsuda, Y., Sasaki, H. (2001) Sequence polymorphisms, allelic expression status and chromosome localisation of the chicken IGF2 and *MPR1* genes, *Cytogenet. Cell Genet.* **93**, 109–113.

Yoon, B.J., Herman, H., Sikora, A., Smith, L.T., Plass, C., Soloway, P.D. (2002) Regulation of DNA methylation of *Rasgrf1*, *Nat. Genet.* **30**, 92–96.

Young, L.E., Fernandes, K., McEvoy, T.G., Butterwith, S.C., Gutierrez, C.G., Carolan, C., Broadbent, P.J., Robinson, J.J., Wilmut, I., Sinclair, K.D. (2001) Epigenetic change in *IGF2R* is associated with fetal overgrowth after sheep embryo culture, *Nat. Genet.* **27**, 153–154.

Genomic Responses to Oxidative Stress

John G. Scandalios
Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695, USA

1 Introduction 491

2 Gene Responses 492

3 Oxygen and Oxygenation of the Earth 493

4 Reactive Oxygen and Oxygen Toxicity 494

5 Defenses against Reactive Oxygen 495
5.1 Superoxide Dismutase (SOD) 496
5.2 Catalase (CAT) 497

6 ROS, Telomeres, and Aging 498

7 Oxidative Stress 500

8 ROS Can Serve Necessary and Useful Functions 501

9 Oxidative Stress and Gene Expression 503

10 Genomic Scale ROS-Responsive Gene Expression 505

11 Approaching an Integrated View of Stress Responses 507

12 The Paradox 509

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

13 Genomic Fluidity 509

14 Conclusions 511

Bibliography 512
Books and Reviews 512
Primary Literature 512

Keywords

Oxidative Stress
Excess accumulation of reactive oxygen species (ROS) resulting from metabolism, exposure to environmental oxidants, toxicants, radiation, or numerous biotic and abiotic stresses, which perturb cellular redox balance to a more oxidized state and disrupt normal biological functions, often leading to cell death.

Reactive Oxygen Species (ROS; Including "Free Radicals")
Highly reactive oxygen intermediates, which, at certain cellular concentrations are highly toxic to cells often leading to cell death. These include hydrogen peroxide, superoxide anion, hydroxyl radicals, and singlet oxygen. At cellular equilibrium, ROS may play key signaling roles in gene expression.

Signal Transduction
A mechanism by which cells or organisms perceive a signal and transmit it via the proper pathways to elicit a genomic response.

Genome
The totality of a cell's genetic information, including genes and other DNA sequences.

Gene
A segment of the genome (DNA) that codes for a functional product.

Promoter
A sequence of nucleotides on the DNA that is required for the initiation of transcription by RNA polymerase.

Genomic Fluidity
The capacity of the genome to reorganize rapidly in response to a given stimulus or signal.

The environment in which they exist affects all living organisms. Differences among individual organisms in response to environmental stresses are common, whether the stress factors are natural environmental variations, infectious agents, environmental chemicals, or any other natural or anthropogenic environmental variables. During their evolution, organisms have evolved a variety of ways of adapting to environmental changes. However, the underlying mechanisms by which cells or organisms perceive environmental adversity and mobilize their defenses to it are just beginning to be understood. Such an understanding is essential in any future attempts to engineer organisms for greater tolerance or resistance to more frequent and rapid environmental changes. As a consequence of recent developments in molecular biosciences and elucidation of the human and other genomes and genomic analyses, a deeper appreciation of the mechanisms by which genes may perceive environmental signals and start a cascade of biochemical events to effect a response to a given signal is now emerging. The most fundamental stress encountered by all aerobic organisms is "oxidative stress," which is a consequence of the aerobic lifestyle on our planet. A broad range of environmental stresses affecting prokaryotes and eukaryotes are directly or indirectly related to oxygen toxicity. Organisms have evolved sophisticated ways of coping with life in an oxygen-requiring environment. This oxygen paradox is dealt with below.

1 Introduction

Every living organism is affected by its environment. The environment, whether internal or external to the organism, is continually changing, and the organism must adapt if it is to survive. However, an organism apparently well adapted to the environment at any one time may be poorly adapted only a short time later if it cannot modify its physiology or behavior in response to changing conditions. Organisms that can adjust to changes in the environment are likely to exhibit a greater degree of adaptiveness than those that cannot. Environmental changes, irrespective of its source, cause a variety of "stresses" or "shocks" that a cell must face repeatedly, and to which its genome must respond in a programmed manner for the organism to survive. Thus, a dominant theme in modern biology is gene expression. It is desirable to determine how genes are expressed in a selective manner in response to an external or internal signal or stimulus. Examples are responses to light, oxidative stress, pathogenicity, wounding, anaerobiosis, thermal shock, and the "SOS" response in microorganisms. For cells of any organism to respond to external cues, they must be able to perceive these cues or signals and transduce such perceptions into the appropriate response. Some sensing mechanism(s) must be present to alert the cell to imminent danger and to trigger the orderly sequence of events that will mitigate this danger. In addition, there are genomic responses to unanticipated, unprogrammed challenges for which the genome is unprepared, but to which it responds in discernible though initially unforeseen and unpredictable ways. Many, though not all, signals are perceived at

the cell surface by plasma membrane receptors. Activation of such receptors by mechanisms such as ligand binding may lead to alterations in other cellular components, ultimately resulting in alterations in cell shape, ion conductivity, gene activity, and other cellular functions. Identification and isolation of mutants that are unable to respond, or that respond abnormally to a particular signal may provide ways to decipher the mechanisms by which a particular signal is transduced into a given response. Long before humans began manipulating and altering their environment, organisms from the simplest to the most complex began evolving methods to cope with stressful stimuli. Consequently, most living cells possess an amazing capacity to cope with a wide diversity of environmental challenges, including natural and synthetic toxins, pathogens, extreme temperatures, high metal levels, and radiation. Many studies in the past have demonstrated clear "cause-and-effect" relationships upon exposure of a given organism or cell to a particular environmental factor or stressor. However, only recently have certain environmental insults been shown to elicit specific genomic responses. At present, little is known of the underlying molecular mechanisms by which the genome perceives environmental signals and mobilizes the organism to respond. Such information is not only interesting in and of itself, but is also essential in any future attempts to engineer organisms for increased tolerance to environmental adversity. Recent dramatic advances in molecular biology and genomics have made it possible to investigate the underlying mechanisms utilized by organisms to cope with environmental stresses. Investigations of genomic responses to challenge are shedding light on unique DNA sequences capable of perceiving stress signals, thus allowing the cell or organism to mobilize its defenses. The general picture emerging from recent studies involves the sensing of a signal and the transduction of the signal to the transcription apparatus to catalyze transcription initiation. The steps involved in such a process may be summarized as follows:

- A signal, normally the intracellular or extracellular concentration of a small molecule, is perceived by a sensor.
- The signal is then transmitted to the regulatory "activator" protein.
- The signal transduction changes the conformation of the activator protein.
- The altered activator protein binds to a specific DNA site.
- The DNA–protein interaction catalyzes the binding or activity of RNA polymerase to facilitate transcription initiation.

2
Gene Responses

Terminally differentiated cells express an array of genes required for their stable functioning and precise metabolic roles. A genome can respond in a rapid and specific manner by selectively decreasing or increasing the expression of specific genes. Genes whose expression is increased during times of stress presumably are critical to the organism's survival under adverse conditions. Examination and study of such "stress-responsive" genes has implications for human health and well being, for agricultural productivity, and for furthering basic biological knowledge. In addition to aiding the organism under stress, genomes that are modified by

stress can be utilized to study the molecular events that occur during periods of increased or decreased gene expression. The mechanisms by which an organism recognizes a signal to alter gene expression and responds to fill that need are important physiologically and render possible the examination of gene regulation under various environmental regimens. The mechanisms of induction of stress response in genes are similar among various organisms examined. Similarities in stress-induced changes in gene expression have been observed for a variety of stressors. Some that have been studied in detail include radiation, thermal shock, pathogenic infections, anaerobiosis, trauma, photostress, physical wounding, oxidative stress, water stress, and heavy metals. In all cases, specific changes in transcript and/or protein expression have been observed in various organisms subjected to such challenges.

Recently, numerous studies by several laboratories using a variety of organisms indicate that oxidative stress is a common denominator underlying many diseases and environmental insults, which can lead to cell death in virtually all aerobes. It is also becoming clear that a variety of different biotic and abiotic stresses cause their deleterious effects, directly or indirectly, via reactive oxygen species (ROS) generation.

For example, numerous toxic environmental chemicals such as xenobiotics, pesticides, herbicides, fungicides, ozone, cigarette smoke, and radiation, cause harmful environmental effects via generation of free radicals and other reactive oxygen species.

The focus of this report will be on oxidative stress, its causes and consequences, and mechanisms employed by organisms to cope with it.

3 Oxygen and Oxygenation of the Earth

Of all the planets in our solar system, only Earth is known to contain molecular oxygen (O_2) in its atmosphere and to support aerobic life. However, when Earth was formed about 4.5 billion years ago, its atmosphere was unlike the present, being primarily reducing and essentially free of oxygen. The earliest living organisms were most likely anaerobic heterotrophs living in the primitive ocean depths, shielded from the damaging effects of solar ionizing radiation. The earliest relatively low levels of oxygen were likely the result of photolytic dissociation of water by the sun's ionizing radiation. The bulk of Earth's present oxygen concentration (21% O_2) is derived from the photosynthetic activities of cyanobacteria and plants. It has been estimated that the earth contains about 410×10^3 Emol (1 Erda mole $= 10^8$ mol) of oxygen, and of this, 38.4×10^3 Emol is in the hydrosphere as water. Molecular oxygen is present in the atmosphere (37 Emol) and in the hydrosphere (0.4 Emol) and undergoes continuous turnover, with the total oxygen exchange estimated at $\sim 15 \times 10^3$ Emol per 10^6 years. Aerobic life is responsible for the major portion of oxygen turnover, with photosynthesis being the main input into the oxygen reservoir, and respiration, the main output. The two processes are in approximate equilibrium, and fossil fuel combustion is the major source of oxygen loss from the reservoir.

The accumulation of dioxygen (O_2) in the earth's atmosphere allowed for the evolution of the enormous variety of aerobic organisms that use O_2 as the electron acceptor, thus providing a higher yield of energy compared with fermentation and anaerobic respiration.

4
Reactive Oxygen and Oxygen Toxicity

In its ground state (its normal configuration, O_2), molecular oxygen is relatively unreactive. However, during normal metabolic activity and as a consequence of various environmental perturbations (e.g. extreme temperatures, radiation, xenobiotics, toxins, air pollutants, various biotic and abiotic stresses, and diseases), O_2 is capable of giving rise to frightfully reactive excited states as free radicals and derivatives.

The oxidation powers of molecular oxygen are restricted because electrons can only be absorbed from another species whose electron spin is antiparallel to the two unpaired, parallel-spin electrons in diatomic oxygen. This spin restriction renders ground state molecular oxygen sufficiently unreactive, so that it cannot abstract electrons from other species. However, removal of the spin restriction by adding a single e^-, or upon transfer of energy to oxygen from a photosensitizer (e.g. chlorophyll; flavin-containing compounds), increases the reactivity of oxygen. Photosensitizers can harvest light and energize O_2 to form singlet oxygen ($^1\Delta_g O_2$), which can interact directly with another molecule, transferring the additional energy to the target molecule.

The complete reduction of O_2 to water requires four electrons. O_2 has a preference for a stepwise univalent pathway of reduction resulting in partially reduced intermediates (Fig. 1). The reactive species of reduced dioxygen include the superoxide radical ($O_2^{\bullet-}$), hydrogen peroxide (H_2O_2), and the hydroxyl radical ($OH^\bullet$). The latter can also be generated by the interaction of $O_2^{\bullet-}$ and H_2O_2 in the presence of metal ions. Both $O_2^{\bullet-}$ and $OH^\bullet$ are extremely reactive and can cause molecular damage leading to cell death. The hydroxyl radical reacts with virtually anything, inflicting indiscriminate and extensive intracellular damage. The $O_2^{\bullet-}$ is the conjugate base of a weak acid, the perhydroxyl radical (HO_2), whose pK_a is 4.69 ± 0.08. Thus, under acidic conditions, the very reactive perhydroxyl radical may predominate following a one-electron reduction of dioxygen, while at higher pH values the $O_2^{\bullet-}$ is predominant. These and the physically energized form of dioxygen, singlet oxygen ($^1O_2$) are the biologically most important ROS (Table 1). An activation energy of ~22 kcal mol^{-1} is required to raise molecular oxygen (O_2) from its ground state to its first singlet state. In higher plants, this energy is readily obtained from light quanta via such transfer molecules as chlorophyll. Unless abated, all of these intermediate oxygen species are extremely reactive and cytotoxic in all organisms. ROS can interact with proteins, lipids, and nucleic acids to cause severe molecular damage (Table 2). Thus, oxygen provides a paradox in that it is essential for aerobic life and yet, in its reduced forms, it is one of the most toxic substances with which life on earth must

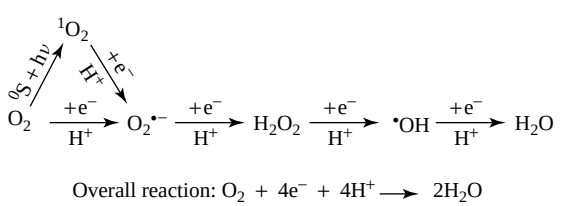

Fig. 1 Pathways in the univalent reduction of oxygen to water leading to generation of various intermediate ROS.

Tab. 1 Reactive oxygen species (ROS) of interest in oxidative stress.

Name	***Notation***	***Some comments & basic sources***
Molecular oxygen (triplet ground state)	O_2; $^3\Sigma$	Common form of dioxygen gas
Singlet oxygen (1st excited singlet state)	$^1O_2$; $^1\Delta$	Photoinhibition; UV irradiation; PS II e^- transfer Rx (chloroplasts)
Superoxide anion	$O_2^{\bullet-}$	Formed in many photooxidation Rx (flavoprotein, redox cycling); Mehler Rx in chloroplasts; mitochondrial e^- transfer Rx; glyoxysomal photorespiration; peroxisomal activity; nitrogen fixation; Rx of O_3 and OH^- in apoplastic space; defense against pathogens; oxidation of xenobiotics
Hydrogen peroxide	H_2O_2	Formed from $O_2^{\bullet-}$ by dismutation; photorespiration; β-oxidation; proton-induced decomposition of $O_2^{\bullet-}$; defense against pathogens
Hydroxyl radical	$OH^\bullet$	Decomposition of O_3 in apoplastic space; defense against pathogens; Rx of H_2O_2 with $O_2^{\bullet-}$ (Haber-Weiss); Rx of H_2O_2 with Fe^{++} (Fenton); highly reactive with all macromolecules
Perhydroxyl radical	$O_2H^\bullet$	Protonated form of $O_2^{\bullet-}$; Rx of O_3 and OH^- in apoplastic space
Ozone	O_3	UV radiation or electrical discharge in stratosphere; Rx involving combustion products of fossil fuels and UV radiation in troposphere

cope. ROS are found in virtually all intracellular organelles or compartments as a consequence of normal metabolic activity. Each organelle or compartment has potential targets for oxidative damage, as well as mechanisms for the elimination of excess ROS accumulation.

5 Defenses against Reactive Oxygen

To minimize the damaging effects of ROS, aerobic organisms evolved both nonenzymatic and enzymatic antioxidant defenses (Table 3). Nonenzymatic defenses include compounds of intrinsic antioxidant properties, such as vitamins C and E, glutathione, and β-carotene. Purely enzymatic defenses, such as superoxide dismutases (SOD), catalases (CAT), and peroxidases (Px) protect by directly scavenging superoxide radicals and hydrogen peroxide, converting them to less reactive species. Superoxide dismutases catalyze the dismutation of $O_2^{\bullet-}$ to H_2O_2, and catalases and peroxidases reduce H_2O_2 to $2H_2O$. The similarity between the SOD and CAT reactions is that each is an oxidation–reduction in which the substrate, $O_2^{\bullet-}$ for SOD and H_2O_2 for CAT, is both reductant and oxidant,

Tab. 2 Examples of ROS damage to lipids, proteins, and DNA.

Oxidative Damage to Lipids:
- Occurs via several mechanisms of ROS reacting with fatty acids in the membrane lipid bilayer leading to membrane leakage and cell death.
- In foods, lipid peroxidation causes rancidity and development of undesirable odors and flavors.

Oxidative Damage to Proteins:
- Site-specific amino acid modifications (specific AA differ in their susceptibility to ROS attack)
- Fragmentation of the peptide chain
- Aggregation of cross-linked reaction products
- Alteration in electrical charge
- Increase in susceptibility to proteolysis
- Oxidation of Fe-S centers by $O_2^{\bullet -}$ destroys enzymatic function
- Oxidation of specific AA "marks" proteins for degradation by specific proteases
- Oxidation of specific AA (e.g. Try) leads to cross-linking

Oxidative Damage to DNA:
- DNA deletions, mutations, translocations
- Base degradation, single-strand breakage
- Cross-linking of DNA to proteins

whereas different reductants are required for the peroxidases, depending upon their specificities. Under some conditions, CAT can act as an efficient peroxidase. SODs deal with the first product of the univalent reduction of O_2, converting it to H_2O_2, which must then be destroyed by catalases and/or peroxidases. Thus, the SODs and CATs serve, in tandem, as frontline antioxidant defenses:

$$O_2^{\bullet -} + O_2^{\bullet -} + 2H^+ \xrightarrow{\text{SOD}} O_2 + H_2O_2 (K_2 = 2.4 \times 10^9\ M^{-1}\ s^{-1})$$

$$H_2O_2 + H_2O_2 \xrightarrow{\text{CAT}} 2H_2O + O_2 (K_1 = 1.7 \times 10^7\ M^{-1}\ s^{-1})$$

$$H_2O_2 + R(OH)_2 \xrightarrow{\text{Px}} 2H_2O + R(O)_2 (K_4 = 0.2\text{–}1 \times 10^3\ M^{-1}\ s^{-1})$$

5.1 Superoxide Dismutase (SOD)

SODs have been isolated and characterized from a wide variety of organisms. One class consists of SODs with Cu(II) plus Zn(II) at the active site (Cu/ZnSOD), another with Mn(III) (MnSOD), a third with Fe(III) (FeSOD), and a fourth with Ni(II/III) (NiSOD). Cu/ZnSODs are generally found in the cytosol of eukaryotic cells, in chloroplasts, and in some prokaryotes; MnSODs are found in prokaryotes and in mitochondria; FeSODs are generally found in prokaryotes, in algae, and in some higher plant chloroplasts; NiSODs have been found in *Streptomyces*. Unlike most other organisms that have only one of each type of SOD in the various cellular compartments, plants have multiple forms of each type encoded by more than one gene, indicative that plants have far more complex antioxidant defenses. Plants also produce a large variety of small nonenzymatic antioxidant compounds as second tier defenses, such as glutathione, ascorbate, tocopherols, flavonoids, alkaloids, and carotenoids in high concentrations that are capable of quenching ROS. The dismutation of $O_2^{\bullet -}$ to $O_2 + H_2O_2$ by SOD is hardly a bargain, as the resulting H_2O_2 can react with metal ions, giving rise to the highly toxic $OH^{\bullet}$. Fortunately, catalases come to the rescue by degrading H_2O_2 to O_2 and H_2O. Most aerobes, including mammals, possess at least one form of homotetrameric catalase with ferriheme at the active sites.

Tab. 3 Some natural antioxidants.

(A) Nonenzymatic antioxidant molecules	
Antioxidant molecule	***Subcellular location***
Ascorbate (vitamin C)	Plastid; apoplast; cytosol, vacuole
β-carotene	Plastid
Glutathione, reduced (GSH)	Plastid; mitochondrion; cytosol
Polyamines (e.g. putrescine, spermine)	Nucleus; plastid; mitochondrion; cytosol
α-tocopherol (vitamin E)	Cell & plastid membranes
Zeaxanthin	Chloroplast

(B) Antioxidant Enzymes		
Enzyme	***EC Number***	***Subcellular Location***
Ascorbate peroxidase (APx)	1.11.1.11	Plastid stroma & membranes; cytosol
Peroxidases (nonspecific)	1.11.1.7	Cytosol; cell wall-bound; apoplast
Catalase (CAT)	1.11.1.6	Glyoxysome; peroxisome; cytosol; mitochondria
Superoxide dismutases (SOD)	1.15.1.1	Cytosol (*Cu/ZnSOD*); plastid (*Cu/ZnSOD;FeSOD*); mitochondrion (*MnSOD*); peroxisome (*Cu/ZnSOD*)
Dehydroascorbate reductase (DHAR)	1.8.5.1	Cytosol; plastid stroma
Glutathione reductase (GR)	1.6.4.2	Mitochondrion; cytosol; plastid stroma
Monodehydroascorbate reductase (MDHAR)	1.6.5.4	Plastid stroma
Glutathione *S*-transferases (GST)	2.5.1.18	Cytosol; Microsomal

5.2 Catalase (CAT)

CAT is largely, but not exclusively, localized in peroxisomes, wherein many H_2O_2-producing enzymes reside. Thus, catalase, which exhibits a high K_m for H_2O_2, can act upon the H_2O_2 produced before it diffuses to other parts of the cell. Catalase is a tetrameric heme-containing enzyme that is found in all aerobic organisms. Because of its wide distribution, evolutionary conservation, and capacity to rapidly degrade hydrogen peroxide, it has been proposed that CAT plays an important role in systems that have been evolved to allow organisms to live in aerobic environments.

Catalase is one of the most active catalysts produced by nature. It decomposes hydrogen peroxide at an extremely rapid rate, corresponding to a catalytic center activity of about 10^7 min^{-1}. Depending upon the concentration of H_2O_2, it exerts a dual function. At low concentrations ($<10^{-6}$ M) of H_2O_2, it acts "peroxidatically," where a variety of hydrogen donors (e.g. ethanol, ascorbic acid) can be oxidized in the following manner:

$$RH_2 + H_2O_2 \longrightarrow R + 2H_2O$$

At high concentrations of substrate, catalase decomposes toxic hydrogen peroxide at an extremely rapid rate using the "catalatic" reaction in which H_2O_2 acts as both an acceptor and donor of hydrogen molecules: $2H_2O_2 \rightarrow 2H_2O + O_2$.

Catalase is unique among H_2O_2 degrading enzymes in that it degrades H_2O_2 without consuming cellular reducing equivalents. Hence, catalase provides the cell with a very energy-efficient mechanism to remove hydrogen peroxide. Therefore, when cells are stressed for energy and are rapidly generating H_2O_2 through "emergency" catabolic processes, H_2O_2 is degraded by catalase in an energy-efficient manner. This results in a net gain of reducing equivalents and, therefore, of cellular energy. It has been proposed that catalase may be uniquely suited to regulate the homeostasis of H_2O_2 in the cell. In the catalatic mode, catalase has a very high apparent Michaelis constant and, therefore, is not easily saturated with substrate. Thus, the enzyme activity increases linearly over a wide range of H_2O_2 concentrations and, thereby, maintains a controlled intracellular H_2O_2 concentration. In mammalian systems, organs with high concentrations of catalase (i.e. liver and kidney) have low levels of endogenous H_2O_2 and organs with low concentrations of catalase (i.e. lung and heart) have high endogenous levels of H_2O_2. Further, if catalase activity is inhibited, H_2O_2 concentrations rise in the liver. As in the case of SOD, multiple catalases (isozymes), encoded by specific genes are found in plants, whereas animals exhibit one form of catalase. Both *Cat* and *Sod* genes respond differentially to various stresses known to generate ROS (Fig. 2).

6
ROS, Telomeres, and Aging

Many hypotheses have been proposed to explain the root cause of aging. One broad-based hypothesis is that generalized homeostatic failure leads to age-related decline. Another is the "free-radical theory of aging" suggesting that endogenous ROS continually damage cellular macromolecules, including DNA. Incomplete repair of such damage would lead to its accumulation over time resulting in age-related deterioration.

The evidence implicating the generation of ROS as key factors in determining longevity is accumulating. Much of the earlier evidence was correlative. However, recent evidence has identified longevity-influencing genes responsive to ROS. It has further been shown that continual induction of DNA damage during normal aging results in genomic instability due to persistent DNA lesions, mutations, stalled repair, and transcription interference. The involvement of ROS in limiting lifespan has been suggested by analyses of transgenic *Drosophila*, which systematically overexpress both *Sod* and *Cat* genes. Some strains of these transgenic flies live up to 30% longer than their natural counterparts, whereas flies carrying only one of these constitutively expressed transgenes do not live longer, suggesting that ROS toxicity plays an important role in longevity.

Much attention has also recently been given to the role of telomere shortening as a causative factor in aging/senescence. Telomeres are stretches of repetitive DNA of tandem short sequence repeats (TTAGGG) that "cap" the ends of eukaryotic chromosomes to protect against degradation. Telomeres, in most human cells, shorten with each round of DNA replication because they lack the

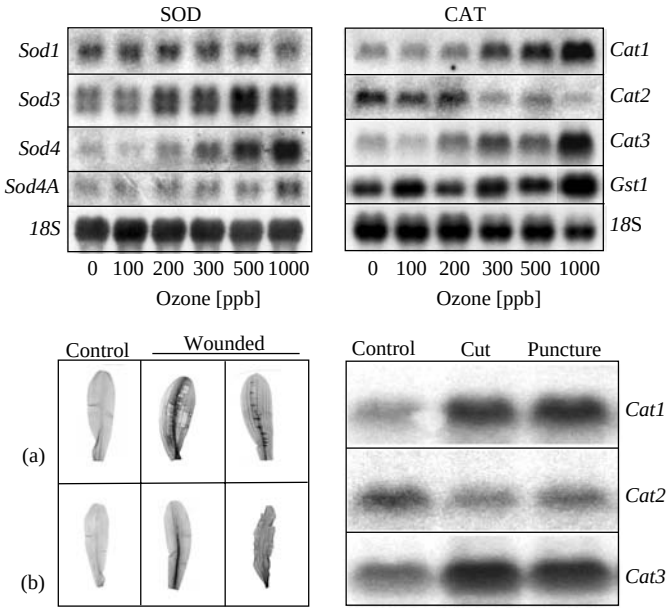

Fig. 2 Response to acute ozone exposure of the *Sod* (top-left) and *Cat* (top-right) genes in maize leaves after 6 h O_3 fumigation. As noted, some of the transcripts are upregulated (e.g. *Sod3, Sod4, Cat1, Cat3*), while others are downregulated (e.g. *Sod1, Cat2*). The 18S rRNA is a loading control. The production of H_2O_2 in maize leaves (bottom-left) in response to wounding. Seven-day-old plants were excised at the base of the stems and supplied with DAB (diaminobenzidine) for 6 h. The plants then were wounded and continuously supplied with DAB for 4 h. The production of H_2O_2 can be visualized by the deposition of the brown-red (dark areas) color products in the leaves. (a) Control and wounded leaves (wounding was conducted near the main vein), (b) control and wounded leaves (wounding was conducted at the main vein and cut at the leaf edges). *Cat* transcript accumulation in response to wounding (bottom-right). RNA was isolated from similarly treated leaves, indicative that differential transcript accumulation is effected by induction of H_2O_2. (See color plate p. xxv.)

enzyme telomerase. Telomerase is a specialized reverse transcriptase, which helps to replicate the telomere ends of chromosomes. However, telomerase is normally expressed only in germline cells and is derepressed in tumor cells in which telomeres are stabilized. It has recently been demonstrated that oxidative stress is the main reason for telomere shortening. It has been observed that H_2O_2 plus Cu(II) induced 8-oxo-7, 8-dihydro-2′-deoxyguanosine (8-oxodG) formation in the telomere sequences more efficiently than in nontelomere sequences. Oxidative damage is not repaired as well in telomeric DNA as elsewhere in the chromosome. Oxidative stress accelerates telomere loss, while antioxidants decelerate it. Thus, oxidative stress is a critical modulator of telomere loss and telomere-driven replicative senescence is primarily a stress response.

Given the available information, aging is likely a multifactorial process; whether ROS are peripheral targets that correlate with longevity or with central regulators of human aging remain to be resolved.

7 Oxidative Stress

Reactive oxygen species such as superoxide ($O_2^{\bullet -}$), hydrogen peroxide (H_2O_2), and hydroxyl radical ($OH^{\bullet}$) are produced in all aerobic organisms and normally exist in the cell in balance with antioxidant molecules. Oxidative stress occurs when this critical balance is disrupted because of depletion of antioxidants or excess accumulation of ROS, or both (Fig. 3). That is, when antioxidants are depleted and/or if the formation of ROS increases beyond the ability of the defenses to cope, then oxidative stress and its noxious consequences ensue. Such stress can happen when severely adverse environments or physiological conditions overwhelm biological systems. One rapid and clear indicator of oxidative stress is the induction of antioxidant defenses and/or increases in endogenous ROS levels. The formation of ROS can be accelerated as a consequence of various environmental stress conditions, including UV radiation, high light intensities, exposure to herbicides, extreme temperatures, toxins such as cercosporin and aflatoxin, air pollutants, metals, wounding, and xenobiotics. Many inducers of oxidative stress are known; carcinogens, mutagens, and toxins. It has become evident that ROS production is a common denominator in many diseases and environmental insults and can lead to

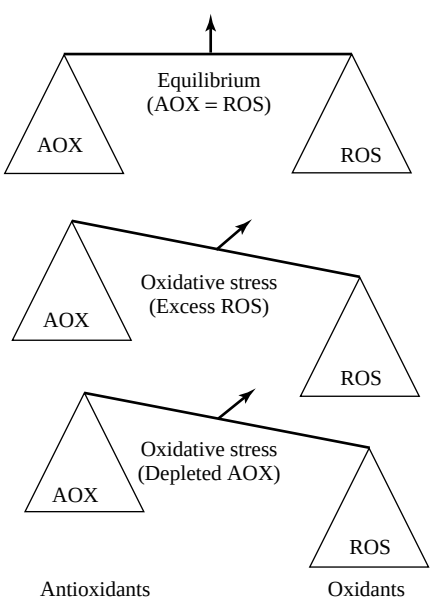

Fig. 3 Oxidative stress results from imbalance between the levels of antioxidants (AOX) and reactive oxygen species (ROS). Cells are normally able to balance the production of oxidants and antioxidants to maintain redox equilibrium. Oxidative stress occurs when this equilibrium is upset by excess levels of ROS, or depletion of antioxidant defenses.

severe cellular damage leading to physiological dysfunction and cell death in virtually all aerobes (Fig. 4).

When oxidative stress occurs, cells function to counteract the oxidant effects and to restore redox balance by resetting critical homeostatic parameters. Such cellular activity leads to activation or silencing of genes encoding defensive enzymes, transcription factors, and structural proteins.

8 ROS Can Serve Necessary and Useful Functions

It has recently become clear that ROS are not always harmful metabolic by-products as generally believed. When tightly regulated, ROS perform critical functions in the cell. In fact, a significant body of evidence indicates that ROS, particularly $O_2^{\bullet-}$ and H_2O_2, act as intracellular signaling molecules. In bacteria, the transcription factor OxyR activates a number of genes inducible by H_2O_2, while the transcription factors SoxR/SoxS mediate responses to $O_2^{\bullet-}$. Yeast also has two distinct adaptive stress responses, one directed toward H_2O_2 and one toward $O_2^{\bullet-}$. In higher eukaryotes, both animals and plants, oxidative stress responses are more complex and are modulated by several regulators. ROS-dependent redox cycling of cysteinyl thiols is critical for establishing protein–protein and protein–DNA interactions that determine many aspects of signal transduction pathways by regulating the activity of many transcription factors. For example, activation of nuclear factor κB (NFκB) and activator protein-1 (AP-1), known to have critical roles in proliferation, differentiation, and morphogenesis, can result from stimulation by diverse agents proceeding through a common pathway involving ROS generation. The pathway leading to H_2O_2 production and subsequent redox activation of NFκB has been shown to

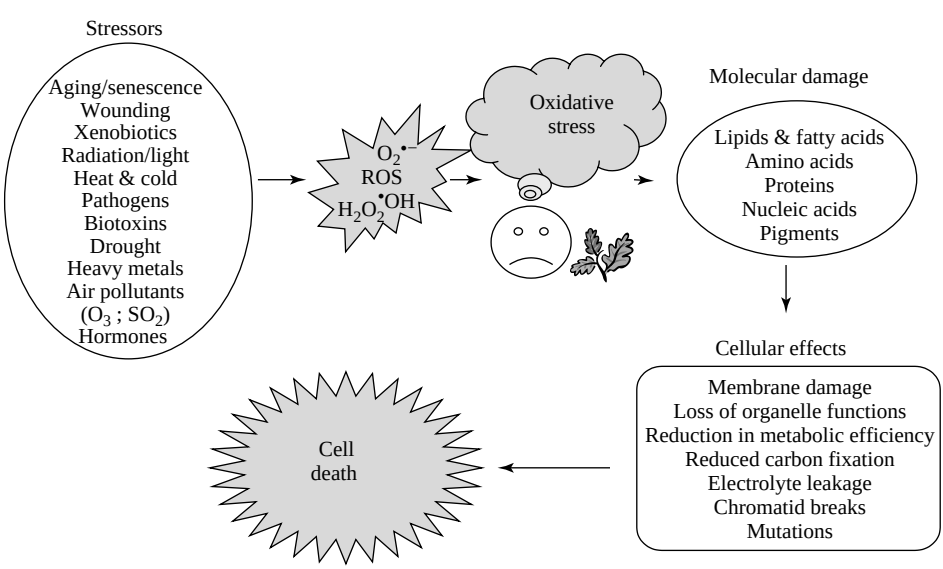

Fig. 4 Scheme showing some of the initiators (stressors) of ROS and the biological consequences leading to a variety of physiological dysfunctions that can lead to cell death.

involve the Rho family of small GTP-binding proteins.

Both intracellular and extracellular sources of ROS are capable of modulating gene expression. Low doses of H_2O_2 (<20 μM) can elicit changes in phosphorylation of specific regulatory proteins including protein kinase B (AKT). Direct signaling action of H_2O_2 in the differential regulation of antioxidant genes in plants likely occurs via protein–DNA interactions in the region of the antioxidant responsive element (ARE; TGACTCA), NFκB, and ABA responsive element (ABRE-2; ACGT) in the promoters of these genes. In addition to induction of defense gene expression, other roles of ROS in plants include direct killing of pathogens, involvement in cell wall structure, and promotion of programmed cell death. In yeast and animal cells, ROS have been shown to arrest cell division, and cell cycle progression is under negative ROS control. A clear and powerful example of how ROS are put to constructive use was the observation that $O_2^{\bullet -}$ plays an important role against invading microbes, in effect serving as a broad-spectrum antibiotic. In response to invasion by pathogens, plants also mount a broad range of defense responses, including a rapid and transient production of large amounts of ROS ("oxidative burst").

Thus, ROS play different, or even opposing roles, during different cellular processes (Fig. 5). For example, under physiological conditions, H_2O_2 may play an important role in signal transduction pathways and in activation of the transcription factor NFκB, while under pathologic (stress) conditions, H_2O_2 can lead to apoptosis or necrosis. Consequently, the steady state level of ROS within cells is critical and is determined by the interplay between ROS-generating mechanisms and the subtle modulating roles played by cellular antioxidants. In plants, H_2O_2 is generated under a diverse range of biotic and abiotic conditions, and its accumulation in specific tissues at specific developmental times, in the appropriate quantities,

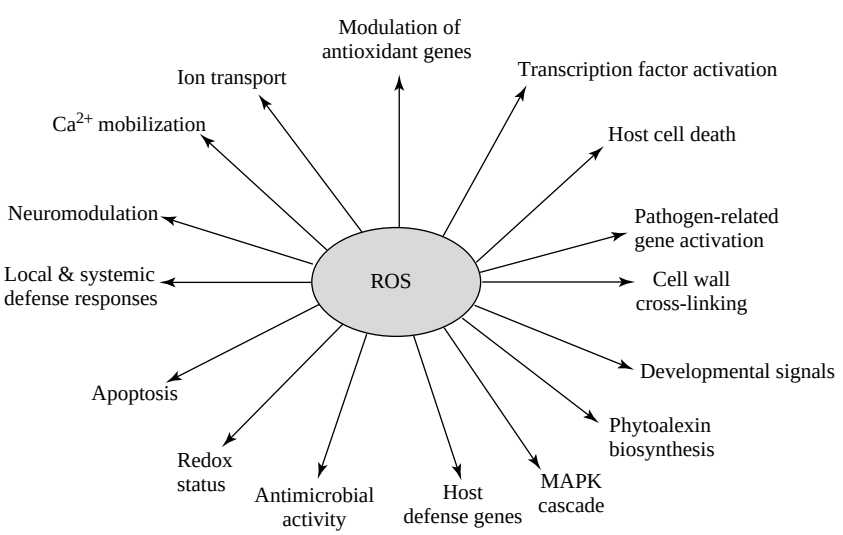

Fig. 5 Scheme showing some of the useful pleiotropic roles of ROS known to occur in most higher organisms, indicative of the fact that ROS are not always harmful to cells.

benefits plants and can mediate cross-tolerance to other stresses. H_2O_2 affects gene expression and activates MAP kinases, which in turn function as regulators of transcription.

9 Oxidative Stress and Gene Expression

Although observations have led to the suggestion that cells have the means to sense ROS and to induce specific responses, the underlying mechanisms are still not fully understood. The transcriptional network that responds to ROS in eukaryotes is currently being deciphered, whereas the prokaryotic system is better understood.

Nearly two decades ago, it was shown that the expression of ~30 proteins were induced by H_2O_2 in bacteria. Of these 30 proteins, 12 were maximally induced within 10 min and 18 between 10 and 30 min. The OxyR regulatory protein was subsequently shown to regulate expression of 9 of the 12 rapidly induced proteins. The tetrameric OxyR protein is a member of the LysR family of transcription activators and exists in two forms, reduced and oxidized; only the oxidized form is able to activate transcription. Further studies led to the identification of a number of OxyR-activated genes. Similarly, the SoxRS regulatory proteins were found to regulate expression of $O_2^{\bullet -}$-responsive proteins in bacteria. Regulation of the *SoxRS* regulon occurs by conversion of SoxR to an active form that enhances *soxS* transcription. The enhanced levels of SoxS in turn activate expression of the regulon. In addition to SoxR and OxyR, several other transcriptional regulators modulate the expression of antioxidant genes in bacteria, indicative of the complexity and connectivity of overlapping regulatory networks. No apparent homologs of OxyR, SoxR, or SoxS have been found in eukaryotes, but a number of other transcription factors have been found to play a role in regulating the expression of antioxidant genes in eukaryotes. In yeast, transcription regulators of antioxidant genes include ACE1, MAC1, YAP1, YAP2, HAP1, and HAP2/3/4. In higher eukaryotes, oxidative stress responses are more complex and are modulated by several different regulators. In mammalian systems, NFκB and AP-1 are involved in regulating the oxidative stress response. The "antioxidant responsive element" (ARE), present in the promoter region of mammalian glutathione *S*-transferase (GST), metalothionein-I, and Mn*Sod* genes, causes induction of these genes in response to oxidants. NFκB, AP-1, and ARE have also been found in promoters of antioxidant genes in higher plants (Fig. 6). The role of these factors is not unique to activation of antioxidant genes, as they are known (particularly NFκB) to play central roles in regulating cellular responses to other stresses as well as regulating normal growth and metabolism.

There is substantial evidence that a variety of biotic and abiotic stresses induce ROS, which serves as a common factor in regulating various signaling pathways (Fig. 7). Similar stresses also activate mitogen-activated protein kinases (MAPK)s with kinetics that either precede or parallel H_2O_2 production, suggesting that MAPKs may be one of several converging points in the defense-signaling network. In addition, exogenous application of several plant hormones and toxins has been shown to induce $O_2^{\bullet -}$ and H_2O_2 synthesis, leading to differential induction of some antioxidant genes and isozymes

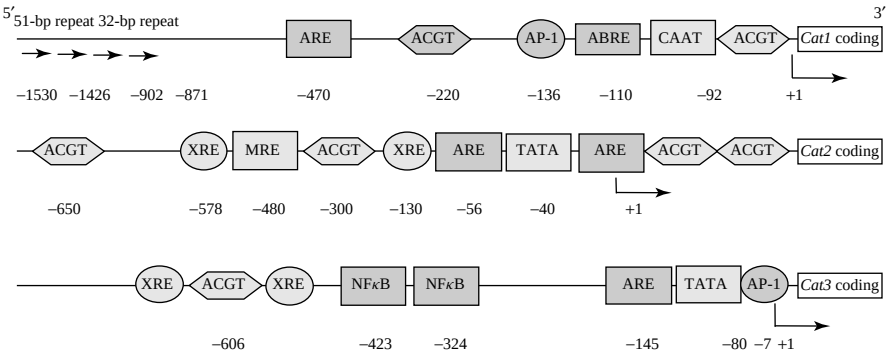

Fig. 6 Schematic representation of the promoter region of each of the three catalase (*cat*) genes of maize indicating the locations of the NFκB, ARE, AP-1, and ABRE motifs relative to the transcription start site (+1) of each gene.

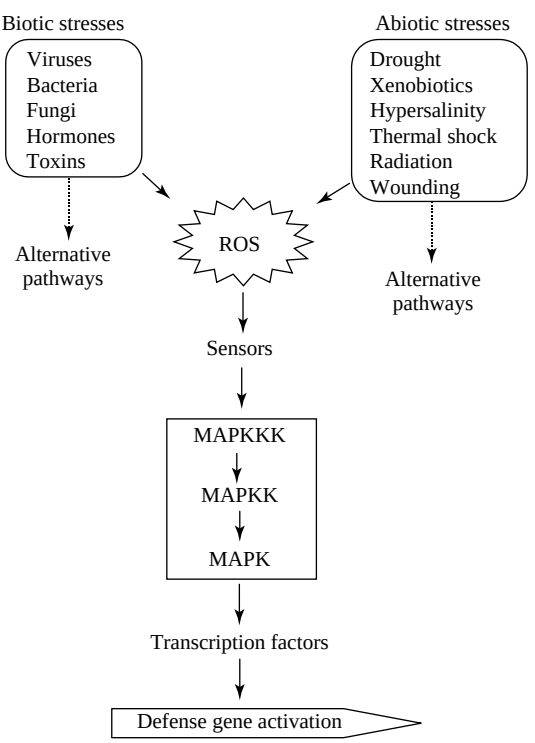

Fig. 7 Convergence of various biotic and abiotic stress stimuli onto MAPK pathways via ROS as a common factor, leading to activation of antioxidant defense genes. Alternative pathways have also been implicated.

(Fig. 8). Thus, the identification of all genes and proteins regulated by H_2O_2 is an important step toward treatments that might confer tolerance to multiple, but interrelated, stresses. In addition to induction/repression of antioxidant defense genes, ROS are known to similarly affect expression of a variety of other genes involved in different signaling pathways in microbes, yeast, plants, and animals.

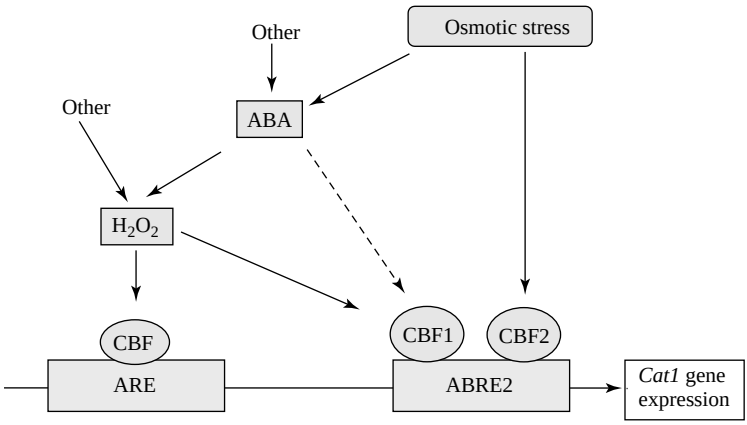

Fig. 8 Schematic representation of the possible signal transduction pathways of *cat1* gene expression in response to the plant hormone abscisic acid (ABA) and osmotic stress in maize. ARE: antioxidant responsive element; ABRE2: ABA responsive element; CBF1: *cat1* binding factor-1; CBF2: *cat1* binding factor-2.

10 Genomic Scale ROS-Responsive Gene Expression

The advent of microarray expression analysis makes possible the assessment of gene expression on a genomic scale, rendering tens of thousands of genes assayable in a single experiment. Thus, identification of ROS-responsive genes on a global scale is now tenable. DNA microarrays can comprehensively examine gene expression networks during oxidative stress. There is now significant progress being made in surveying gene expression in response to H_2O_2 in *Escherichia coli*, yeast, and higher plants.

A genome-wide transcription profile of *E. coli* cells exposed to H_2O_2 was examined with a DNA microarray composed of 4169 *E. coli* open reading frames. Gene expression was measured in isogenic wild type and *oxyR* deletion mutants (Δ*oxyR*) to confirm that the H_2O_2-response regulator OxyR activates most of the H_2O_2-inducible genes. A very rapid and strong induction was observed of a set of OxyR-regulated genes in the wild type but not in the Δ*oxyR*, providing internal validation of the experiment and confirmation for the induction of the oxidative stress genes identified earlier by other means. Several new H_2O_2-inducible genes were also identified: some were members of the OxyR regulon and some were induced by an OxyR-independent mechanism suggestive of other H_2O_2 sensors and regulators in *E.coli*. Several genes repressed by OxyR were highly expressed in the Δ*oxyR* mutant. Overall, the mRNA of 140 genes in the wild type, and 167 genes in the Δ*oxyR* were significantly induced after H_2O_2 treatment. It was also found that *soxS* was induced by H_2O_2, indicating an overlap with other regulatory pathways. Two genes, *Fpr* and *sodA*, known to be members of the SoxRS regulon were also highly induced by H_2O_2 in both wild type and Δ*oxyR*. The microarray data also showed an overlap between oxidative stress, heat shock, and SOS responses. The results from the *E. coli* microarrays clearly indicate that the activities of transcription

factors in addition to OxyR and SoxRS are likely modulated by oxidative stress.

Whole-genome expression patterns in *Saccharomyces cerevisiae* cells exposed to H_2O_2, in addition to other stresses, indicated that about two-thirds of the genome is involved in the response to environmental changes, and the global set of genes induced/repressed by each environmental signal were identified. The response to oxidative stress involves about one-third of the yeast genome and the maximal effects on gene expression occur slightly later, relative to other stresses examined during similar time courses, with most of the transcriptome returning to prestress levels within 2 h following exposure to H_2O_2. Genes that are repressed for ~60 min after exposure to H_2O_2 are only transiently repressed in other stress time courses. Thus, genes encoding the translation apparatus and its regulators are remarkably coordinated for each environmental change, although the dynamics of each response are different. The expression programs following H_2O_2 or $O_2^{\bullet-}$ treatment were essentially identical despite the fact that different ROS are involved. There was strong induction of genes known to be involved in detoxification of both H_2O_2 and $O_2^{\bullet-}$, such as catalase, superoxide dismutase, and glutathione peroxidase, as well as genes involved in oxidative and reductive reactions (e.g. thioredoxin, glutathione reductase, glutaredoxin). The genes most strongly induced in response to H_2O_2 and $O_2^{\bullet-}$ were dependent on the transcription factor Yap1p for their induction. Genes moderately induced by ROS and other signals are regulated by different transcription factors, depending on the conditions, and different upstream signaling pathways may govern their response. It has also been demonstrated that in *Schizosaccharomyces pombe*, H_2O_2 activates the Sty1 (stress-activated mitogen-activated protein kinase, MAPK) pathway in a dose-dependent manner via two sensing mechanisms. At low H_2O_2 levels, a two-component signaling pathway, which feeds into either of the two (Wak1 or Win1) stress-activated mitogen-activated protein kinase kinase kinases (MAPKKKs) regulates Sty1. At high H_2O_2 levels, however, Sty1 activation is controlled mainly by an independent two-component mechanism, which requires the function of both Wak1 and Win1. In addition, the individual bZip transcription factors, Pap1 and Atf1, were found to function within a limited range of $[H_2O_2]$: Pap1 activates target genes at low $[H_2O_2]$, whereas Atf1 controls transcriptional responses to high $[H_2O_2]$, with some minor overlap. Some apparent cross talk among Sty1, Atf1, and Pap1 has been detected. Thus, *S. pombe* deploys a combination of stress-responsive regulatory proteins to gauge and trigger the appropriate transcriptional response to increasing H_2O_2 concentrations. This organism mounts two separate responses to oxidative stress – an adaptive response to low-level H_2O_2 exposure that protects it from subsequent exposures to higher $[H_2O_2]$, and an acute response that allows the cell to survive a sudden, potentially lethal dose of H_2O_2.

Large-scale cDNA microarray analysis of the *Arabidopsis* transcriptome during oxidative stress identified 175 nonredundant expressed-sequence tags (ESTs) from a sample of 11 000, which are regulated by H_2O_2. Of these, 62 are repressed and 113 are induced. In addition, RNA blots showed that some of the H_2O_2-regulated genes are also modulated by other signals known to involve oxidative stress. Furthermore, a substantial portion of these genes have predicted functions in defense response, cell rescue and signaling, and transcription, underscoring the

pleiotropic effects of H_2O_2 in the response of plants to stress. Overall, the microarray used was estimated to represent only ~30% of the *Arabidopsis* genome, depending on redundancy, and 1 to 2% of the genes represented in the array are affected by H_2O_2-imposed oxidative stress, being comparable to the situation in yeast. Of the 175 genes identified as H_2O_2-responsive, most have no obvious direct role in oxidative stress but may be linked to oxidative stress indirectly, as a consequence of other biotic and abiotic stresses, explaining their sensitivity to H_2O_2. Among the genes induced by H_2O_2 were genes encoding transcription factors, suggesting that they may mediate downstream H_2O_2 responses. As in other organisms, expression of the MAPKs in *Arabidopsis* is induced by oxidative stress, which in turn can mediate the induction of oxidative stress-responsive genes.

11 Approaching an Integrated View of Stress Responses

The traditional view of ROS as mere indiscriminate, reactive by-products of cellular metabolism has recently undergone a metamorphosis. This view came about by the discovery that ROS may act as signal-transducing molecules and that activation of intracellular transcription factors such as OxyR, SoxRS, NFκB, and AP-1 occur via interaction with ROS, leading to gene transcription (Fig. 9).

Genome sequencing and expression profiling using DNA or oligonucleotide microarrays and related technologies, have been used effectively in the study of global gene expression patterns in response to different growth and environmental conditions to which organisms are exposed. Subsequent hierarchical clustering methods allow for the allocation of genes, coregulated temporally or in response to a given signal, into specific expression groups or regulons. The numbers of genes that can be detected by these methods in response to any given environmental or developmental signal far exceed the limited number that could have been detected only a few years ago. With the current methods, the expression of tens of thousands of genes can be detected in a single experiment in response to ROS, or to any given signal. Since transcription of genes into mRNA is governed by transcription factors that bind to cis-regulatory regions of the DNA in the vicinity of the target gene, the question arises as to whether large coregulated groups of genes share cis-regulatory elements that bind to common transcription factors. The data available with respect to oxidative stress seem to suggest this. Cis-acting elements within the promoters of ROS-activated genes are being defined as well as their cognate trans-acting factors. A comparative analysis of promoter sequences of genes with similar expression profiles should provide a basis for unraveling common regulatory sequences and overlapping gene expression networks modulating ROS-responsive genes. The antioxidant enzymes catalase and superoxide dismutase play key roles in modulating the levels of endogenous H_2O_2 and $O_2^{\bullet -}$, which at specific concentrations act to, in turn, modulate the expression of other ROS-responsive genes.

The use of microarrays and future derivations thereof, to examine global gene responses to ROS is assured. It has become clear that there are far more genes and gene clusters responding to ROS than previously thought, and that ROS likely play far more key roles in cellular activities than anticipated. As global gene

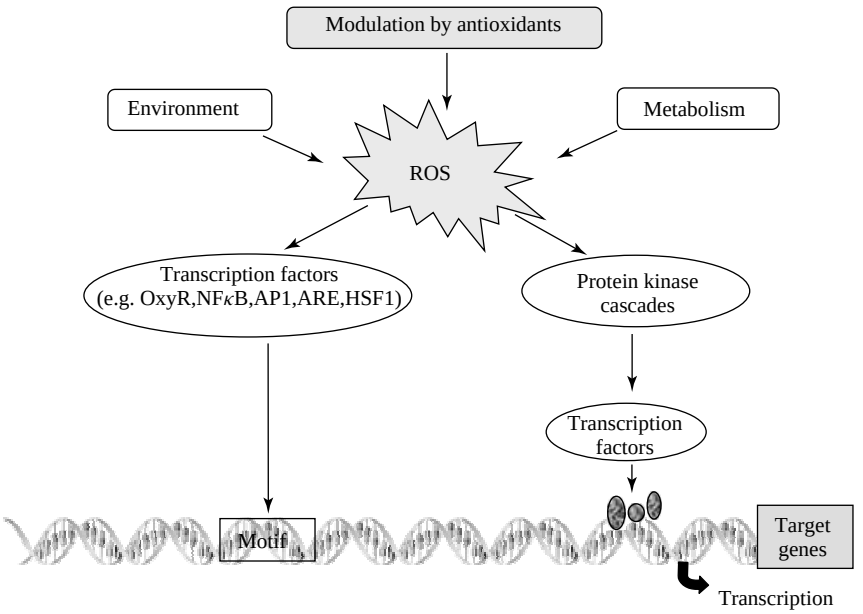

Fig. 9 The major signaling pathways activated in response to oxidative stress. ROS originating from environmental signals or from metabolic activity are modulated by antioxidants to nontoxic levels at which point they serve as signaling molecules. ROS can activate gene transcription in two ways: (1) via transcription factors, such as NFκB, AP-1, and ARE-binding proteins (ARE-BP) that can interact directly with specific DNA motifs on promoters of target genes, or (2) via activation of MAPK cascades, which in turn activate transcription factors that trigger target gene transcription. The degree to which a given pathway is activated depends on the nature and duration of the stress, as well as on cell type and developmental stage.

responses to ROS are examined temporally and spatially along with different time courses and varied oxidant concentrations, even more complex regulons and regulatory cascades will emerge. It will also be instructive to examine gene expression profiles in specific mutants to help delineate the roles of specific regulators and of ROS. The identification of alternate pathways of ROS-dependent gene expression and characterization of the redox-sensing mechanisms involved should point to new insights and directions. Some genes have been identified whose transcription is responsive to a variety of stresses in addition to oxystress, while others appear to be responsive only to ROS or other specific signals. Although some regulatory systems have been implicated in modulating such responses, the complete network of regulators of ROS responses that activate such genes remains unclear.

Microarray results from different organisms clearly underscore the fluidity of genomes to reorganize and respond to changes in the cellular and extracellular environment. However, characterization of global gene expression programs at the transcript level is only the first step toward defining the role or function of each ROS-responsive gene. It is not unlikely that small changes in gene expression could

lead to large alterations in protein levels. Consequently, proteomic analyses are essential to correlate mRNA changes with protein levels. There is, in fact, ample data demonstrating that ROS can alter the activity of cellular proteins.

A large and growing number of sequenced genomes and emerging technologies present us with enormous opportunities to advance biological science and our knowledge of how genomes perceive signals to respond to variable environments. Oxidative stress and the responses to it seem to be central to many of the key biological questions. However, as the euphoria of new sequencing and technological advances begin to fade, we must recognize the magnitude of the problems that still need to be solved. Knowing the sequences of tens of thousands of ROS-responsive genes only reminds us that we still do not know the many proteins they encode, nor the biochemical or biological function of the great majority of such proteins. How such proteins interact with ROS to drive the various physiological processes in aerobic organisms remain a great puzzle. For the future, the fundamental challenge will be to integrate the information now being obtained on gene expression patterns with structural, topological, and functional parameters and interactions of the various proteins encoded by ROS-responsive regulons, and to view the cell in which they function holistically.

12 The Paradox

The oxygen paradox is indeed the paradox of evolution itself. Evolutionary pressures have made the best of a bad situation by generating mechanisms to curtail the undesirable toxic effects of ROS, which are an unavoidable consequence of the aerobic lifestyle, and to put them to constructive uses. Indeed, evolution has coopted ROS to serve necessary and useful purposes in the maintenance of cellular homeostasis and in the communication of cells with the external environment. The focus must now be placed toward a more thorough understanding of how ROS-mediated signals are perceived, transduced, and interpreted by the cell's genetic machinery. Perhaps the most noteworthy observation to date concerning oxidative stress and its negative and positive roles is its universality among aerobic organisms and the similarities emerging in the regulatory mechanisms underlying their roles in all species.

The classical view of ROS as villains that indiscriminately destroy biomolecules has undergone a shift in which positive biological roles are considered as well. It is now accepted that ROS, particularly H_2O_2 and $O_2^{\cdot -}$, are carefully regulated metabolites capable of signaling and communicating critical information to the cell's genetic machinery. Redox regulation of gene expression by oxidants and antioxidants is emerging as a vital mechanism in the health of all eukaryotes, including man.

It would indeed be interesting and challenging to identify all the changes in gene expression regulated by oxidative stress, and to determine their commonality among diverse species. Such a global analysis of the effects of H_2O_2 and $O_2^{\cdot -}$ on the transcriptome of eukaryotes has not yet been attained, but with the emergence of postgenomic technologies it will likely be not far off.

13 Genomic Fluidity

There are programmed responses to stress that are initiated within the genome itself,

which can lead to new and irreversible genomic modifications. Thus, in many organisms, the genome may reorganize itself on facing an adversity for which it is unprepared in order to ensure the organism's survival. For example, cells are able to sense the presence in their nuclei of ruptured ends of chromosomes and then to activate mechanisms that will bring together and then unite these ends, one with another. This is the most revealing example of the sensitivity of cells to what is occurring within them.

The capacity of the genome to reorganize and respond rapidly ("genomic fluidity") was solidified by the pioneering work of Barbara McClintock, who demonstrated that discrete genetic loci could transpose in the genome of maize, and was subsequently shown to occur universally among all organisms. Genomic flexibility is an extraordinary adaptability of organisms to their environment. In the course of evolution, those organisms that had the capacity to adapt to new environments survived. This capacity was due to the ability of each organism or species that survived, to possess in its biological repertoire the means to cope with environmental adversity. In situations in which multiple environments existed, those organisms that possessed motility (e.g. animals) could exchange one environment for another; those with a stationary lifestyle (e.g. plants) had to rely on alternate defenses; a case in point being the far more complex, genetically and biochemically, antioxidant defenses in plants versus their animal counterparts. However, in adverse environments (natural or anthropogenic) from which organisms cannot escape, they must rely on diverse and unique sets of responses encoded within, or resulting from the flexibility of their genome. Environmental stress itself can instigate genome modifications by mobilizing cell mechanisms that can restructure genomes in various ways. A novel example is the restructuring of the host genome by plant galls. In this case, one organism is forced to produce a wholly new structure for the benefit of another. The exact structural organization of a given gall gives it a uniqueness that begins with an initial stimulus characteristic of each invasive species. The galls on legume roots associated with nitrogen-fixing bacteria provide a clear example of reprogramming of the plant (legume) genome by a stimulus received from a foreign organism (bacteria). In other examples, a single plant may have several distinctly different galls on its leaves, each catering to a different insect species. The stimulus for placing the insect egg into the leaf initiates reprogramming of the plant's genome, forcing it to make a unique structure adapted to the needs of the developing insect. Such genomic reprogramming induced by a variety of organisms, including bacteria, fungi, and insects, are not a requisite response of the plant host genome during its normal developmental cycle. Interestingly, gall formation involves upregulated levels of ROS and of antioxidant defense genes. Such genomic reorganization allows the organism to cope with stress, and in the long term may provide the bases for evolution of new species.

As the genome of various organisms is being investigated with state-of-the-art molecular techniques, more knowledge is being gained on the structural components involved in the response of specific genes to environmental insult. However, we are just beginning to learn about how the cell perceives danger and initiates the truly remarkable observed responses to it. Once such

information is available, and a better understanding of the defense response is attained, the engineering of organisms for resistance or sustained tolerance to various adverse environmental stresses will be possible.

A signal transduction pathway contains elements that enable a signal to be transmitted within and between cells and to be translated into an appropriate response. Cells can respond to a variety of environmental, physical, and chemical stimuli using a diverse range of transduction and response mechanisms. The essential features of a signal transduction pathway comprise a *receptor* (recognition element) capable of detecting a stimulus, *second messengers* (transmission elements) such as calcium or phosphorylation cascades, and *response elements* (e.g. gene transcription). Such signaling networks are now amenable to study and dissection by biochemical and genetic approaches that may elucidate the underlying mechanisms and lead to the identification of the molecules responsible. A thorough understanding of how organisms perceive, respond, and adapt to changing environmental and developmental stimuli is now attainable.

14 Conclusions

The survival of organisms on earth depends upon the interactions of their genomes with the environments in which they exist. In the course of evolution, organisms evolved a complex array of mechanisms for adapting to both minor and major fluctuations in the environment. The emergence of oxygenic photosynthesis presented early life forms with the greatest environmental challenge and an opportunity. The challenge was to develop antioxidant defenses in order to survive; the opportunity was to exploit the reactivity of oxygen for energy yielding and biosynthetic reactions. The opportunity led to the highly diversified life forms that evolved sufficient defenses and managed to exploit the aerobic lifestyle. Oxygen toxicity likely led to massive extinctions of those organisms unable to cope with it, unless they took refuge in isolated anaerobic niches. Thus, oxygen is a "double-edged sword" in that it makes life on earth possible, but in its reduced forms (reactive oxygen species, ROS), it is highly toxic and lethal. Oxidative stress can arise from an imbalance between generation and elimination of ROS leading to excess ROS levels inflicting indiscriminate damage to virtually all biomolecules leading to various diseases and cell death. The notion that ROS are merely toxic by-products of O_2 metabolism has recently been altered by experimental evidence indicating that ROS are carefully regulated metabolites capable of signaling and communicating information to the cell's genetic machinery. Redox regulation of gene expression by oxidants and antioxidants is emerging as a vital mechanism in the health and disease of all organisms. Irrespective of how or when ROS are generated, an increase in intracellular oxidants results in two critical effects: damage to various cell components and activation of specific signaling pathways, influencing various cellular processes leading to proper cell functions or to cell death. Genomic tools are accelerating the discovery of ROS-responsive genes on a global scale and are expanding our understanding of the oxidative stress response and the pleiotropic roles of ROS in signaling and gene expression.

See also Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Beckman, K.B., Ames, B.N. (1998) The free radical theory of aging matures, *Physiol. Rev.* **78**, 547–581.

Dalton, T.P., Shertzer, H.G., Puga, A. (1999) Regulation of gene expression by reactive oxygen, *Annu. Rev. Pharmacol. Toxicol.* **39**, 67–101.

Finkel, T. (2003) Oxidant signals and oxidative stress, *Curr. Opin. Cell Biol.* **15**, 247–254.

Finkel, T., Holbrook, N.J. (2000) Oxidants, oxidative stress and the biology of aging, *Nature* **408**, 239–247.

Fridovich, I. (1995) Superoxide radical and superoxide dismutases, *Annu. Rev. Biochem.* **64**, 87–112.

McClintock, B. (1984) The significance of responses of the genome to challenge, *Science* **226**, 792–801.

Scandalios, J.G. (Ed.) (1990) *Genomic Responses to Environmental Stress*, Academic Press, San Diego, CA.

Scandalios, J.G. (Ed.) (1992) *Molecular Biology of Free Radical Scavenging Systems*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Scandalios, J.G. (Ed.) (1997) *Oxidative Stress and the Molecular Biology of Antioxidant Defenses*, Cold Spring Harbor Laboratory Press, Plainview, New York.

Scandalios, J.G. (2001) Molecular Responses to Oxidative Stress, in: Hawkesford, M.J., Buchner, P. (Eds.) *Molecular Analysis of Plant Adaptation to the Environment*, Kluwer Academic Publishers, Netherlands, pp. 181–208.

Scandalios, J.G. (2002) Oxidative stress responses: what have genome-scale studies taught us? *Genome Biol.* **3**, 1019.1–1019.6.

Scandalios, J.G. (2002) The rise of ROS, *Trends Biochem. Sci.* **27**, 483–486.

Scandalios, J.G., Guan, L.M. (2001) Transcription Factors Regulating Antioxidant Gene Expression in Response to Biotic and Abiotic Signals, *Gene Families: Studies of DNA, RNA, Enzymes and Proteins*, World Scientific, Singapore, pp. 287–301.

Primary Literature

Babior, B.M. (1984) The respiratory burst of phagocytes, *J. Clin. Invest.* **73**, 599–601.

Dismukes, G.C., Klimov, V.V., Baranov, S.V., Kozlov, Y.N., DasGupta, J., Tyryshkin, A. (2001) The origin of atmospheric oxygen on earth: the innovation of oxygenic photosynthesis, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2170–2175.

Fink, R.C., Scandalios, J.G. (2002) Molecular evolution and structure function relationships of the superoxide dismutase gene families in angiosperms, and their relationship to other eukaryotic and prokaryotic superoxide dismutases, *Arch. Biochem. Biophys.* **399**, 19–35.

Gasch, A., Spellman, P., Kao, C., Harel, O., Eisen, M., Storz, G., Botstein, D., Brown, P. (2000) Genomic expression programs in the response of yeast cells to environmental changes, *Mol. Biol. Cell* **11**, 4241–4257.

Guan, L., Scandalios, J.G. (1996) Molecular evolution of maize catalases and their relationship to other eukaryotic and prokaryotic catalases, *J. Mol. Evol.* **42**, 570–579.

Guan, L.M., Scandalios, J.G. (2000) Hydrogen peroxide mediated catalase gene expression in response to wounding, *Free Radical Biol. Med.* **28**, 1182–1190.

Guan, L.M., Zhao J., Scandalios, J.G. (2000) *Cis*-elements and *trans*-acting factors that regulate expression of the maize *Cat1* antioxidant gene in response to ABA and osmotic stress: H_2O_2 is the likely intermediary signaling molecule for the response, *Plant J.* **22**, 87–95.

Ruzsa, S., Scandalios, J.G. (2003) Altered Cu metabolism and differential transcription of Cu/Zn *Sod* genes in a Cu/Zn SOD-deficient mutant of maize: evidence for a Cu-responsive transcription factor, *Biochemistry* **42**, 1508–1516.

Scandalios, J.G. (1965) Subunit dissociation and recombination of catalase isozymes, *Proc. Natl. Acad. Sci. U.S.A.* **53**, 1035–1040.

Von Zglincki, T. (2002) Oxidative stress shortens telomeres, *Trends Biochem. Sci.* **27**, 339–344.

Genomic Sequencing (Core Article)

Todd Charles Wood[1] *and Jeffrey P. Tomkins*[2]
[1] *Bryan College, Dayton, TN, USA*
[2] *Clemson University, Clemson, SC, USA*

1 Introduction to Genome Sequencing 515

2 The Sequencing Technique 516

3 Sequencing Bacterial and Archaeal Genomes 524

4 Sequencing Eukaryotic Genomes 526

5 Interpreting the Sequence 532

6 Summary and Principles 533

Bibliography 534
Books and Reviews 534
Primary Literature 534

Keywords

Annotation
Discovery or prediction of biologically important features (e.g. genes, repeat sequences, etc.) in DNA sequence.

Assembly
Recovery of a single consensus sequence from random shotgun reads.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

Bacterial Artificial Chromosome (BAC)
A single-copy plasmid cloning system for large DNA molecules (100–300 kb).

Base-calling
Derivation of nucleotide assignments from sequencing chromatograms.

Fingerprinting
Enzymatic restriction of large-insert clones electrophoretically separated and viewed on an agarose gel.

Genetic Map
A set of genetic markers ordered by their recombination frequencies.

Genome
The complete complement of DNA for an organism consisting of all genes and organized on one or more chromosomes.

Physical Map
An ordered set of large-insert clones providing a physical coverage of a genome.

Production or Shotgun Sequencing
A generation of redundant, raw sequence data from random subclones of a longer target clone.

Beginning with the landmark work of Watson and Crick and continuing with the technological breakthroughs of the 1970s, 1980s, and the 1990s, molecular genetics and genomics has revolutionized all aspects of biology. Today, scientists are able to decipher the precise order of every nucleotide of an organism's complete genome, using a conceptually simple approach called *shotgun sequencing*. Briefly described, shotgun sequencing involves randomly breaking the target genome into smaller pieces, which are sequenced and reassembled into a complete sequence using computer software. Genome sequences provide a basis for further research into gene regulation, pathology, evolution, and metabolism. In order to appreciate a complete genome sequence for the experimental data that it is, it is necessary to review the steps taken to generate genomic sequence. This article will review both the technical aspects of genome sequencing as well as the various strategies employed for sequencing large, eukaryotic genomes.

1
Introduction to Genome Sequencing

The history of modern genomics developed from a fusion of genetics, cytogenetics, biochemistry, and molecular biology. Each of these fields has contributed to our understanding of the material basis of inheritance and how genes code for organismal phenotypes. We may briefly define a genome as the complete set of DNA for a given organism, organized into chromosomes and containing all the genes, which code for the organism's traits (Fig. 1). Thus, historically separate fields of study find union in genomics, the study of organismal genomes. Genome sequencing represents one of the most fundamental steps in genomic analysis. To sequence a genome is to decipher the precise order of nucleotides (the sequence) of that genome. Genome sequencing serves as a foundation for future analyses of gene regulation, chromosome structure, genetic pathologies, and evolution.

With the historic deciphering of the chemical structure of DNA by Watson and Crick in 1953, the modern era of molecular biology and genomics began in earnest. Efficient methods of DNA sequencing were introduced in the 1970s, and the complete 16 569 bp human mitochondrial genome sequence was published in 1980. By the middle of the 1980s, researchers were already beginning to discuss the feasibility of deciphering the sequence of the entire human genome, and the official Human Genome Project began in 1990. The successful application of shotgun sequencing to bacterial genomes by the Institute for Genomic Research in the mid-1990s ushered in a period of rapid expansion of genome sequencing. Today, the complete genomes of more than 100 bacteria and archaea have been sequenced, together with several eukaryotes. Two different drafts of the human genome sequence were published in 2001, and the human genome sequence was completed in 2003.

Despite the enormous changes wrought on modern biology by genomics, the basic techniques of DNA sequencing have not fundamentally changed in the 30 years since their development. This article will begin with an overview of Sanger's sequencing technique and the shotgun strategy utilized in modern genome-sequencing projects. We will then discuss strategies for obtaining the sequence of large genomes, including the hierarchical mapping/sequencing method and the whole-genome shotgun (WGS) that has recently become popular for sequencing eukaryotic genomes. The article will conclude with a brief summary of methods for identifying biological features in raw DNA sequence data.

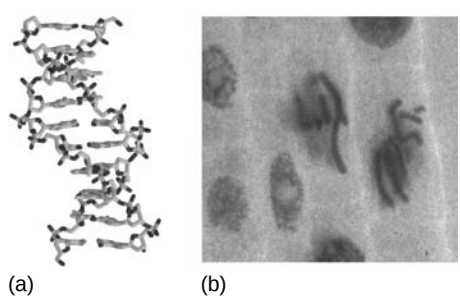

Fig. 1 An organism's genome is the complete complement of DNA, (a) consisting of all genes and (b) organized into chromosomes.

2
The Sequencing Technique

The basic sequencing technique developed in the 1970s begins with the polymerization of synthetic strands of DNA from a template strand. The template strand is a cloned piece of DNA that originates from the organism of interest (see below). The polymerization of the synthetic strand of DNA takes place in a mixture of normal deoxynucleoside triphosphates (dNTPs) and a dideoxynucleoside triphosphate (ddNTP). ddNTPs lack a 3′ hydroxyl but in all other ways are identical to normal dNTPs. They can substitute for normal dNTPs in DNA polymerization reactions, but the missing 3′ hydroxyl prevents elongation of any DNA strand when a ddNTP is incorporated (Fig. 2). As a result, the chain terminates at ddNTP incorporation, giving ddNTPs their colloquial name "chain terminators."

Chain terminators are used in sequencing by including a small amount of them in a synthetic DNA polymerization reaction. The result is random incorporation of chain terminators into growing DNA strands, yielding a complete set of fragments that begin at the same 5′ end but terminate at 3′ positions where a chain terminator has been incorporated (Fig. 2). For example, when ddATP is included in synthetic DNA polymerization, the resulting strands will vary in length but all will terminate at the 3′ end with an adenine. Using the other three ddNTPs, fragments can be generated that terminate with a known base at the 3′ end of the fragment (Fig. 3).

Chain-terminated fragments are then size-fractionated by electrophoresis on a polyacrylamide slab gel. Because the gel has a very fine pore size, fragments differing in length by a single nucleotide can be separated. This allows each chain-terminated fragment to be distinguished on the polyacrylamide gel. Traditionally, fragments in the gel were visualized by incorporating radioactively labeled dATP in the sequencing reactions and exposing the gel to X-ray film after electrophoresis.

One of the major innovations that rendered genome sequencing practical is the introduction of fluorescent labels, which alleviated the need for radioactive dATP and X-ray film. In fluorescent sequencing, the DNA fragments corresponding to the four different ddNTP chain terminators are labeled with a different fluorescent tag. These tags are excited by a laser at the end of the gel and the fluorescent light is detected at the different wavelengths of each of the four dyes. Detection of the fluorescent tags can take place as the electrophoresis is still occurring, eliminating the need to stop the electrophoresis and expose the gel to X-ray film. For fluorescent tags attached to the sequencing primer (dye-primer), the four sequencing reactions corresponding to the four bases must occur in isolation. Alternatively, fluorescent tags can be attached directly to the ddNTP (dye terminator), which allows for a single sequencing reaction in a single container for each template to be sequenced. Generally, dye-terminators reduce the frequency of anomalous mobility during electrophoresis but are poorly recognized by the DNA polymerase, presumably because of the added bulk of the attached fluorescent dye. Dye primers, while requiring isolated preps and an extra mixing stage prior to electrophoresis, provide a consistently readable signal.

A record of fluorescent tags as they pass by the detector during electrophoresis is called a *sequence trace*. A trace can be converted to a visual representation called a *chromatogram* (Fig. 4) for translation

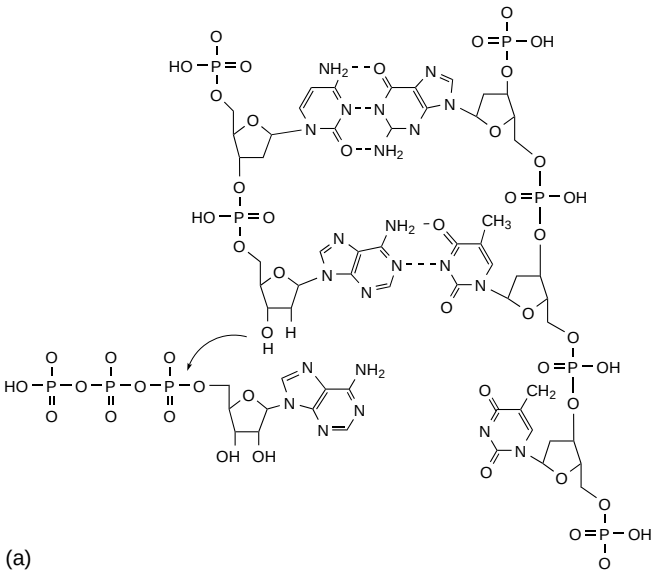

(a)

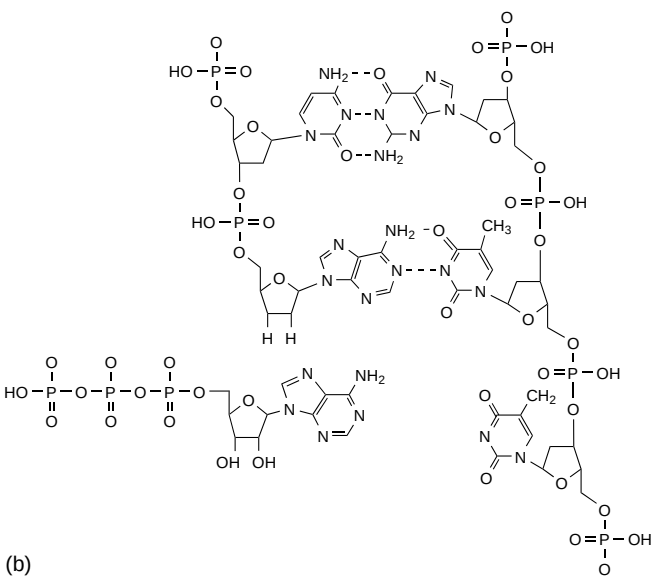

(b)

Fig. 2 (a) During normal DNA polymerization, the 3′ hydroxyl is necessary for formation of the phosphate bond that elongates the nascent DNA strand. (b) If the 3′ hydroxyl is missing, as when a ddNTP (chain terminator) is incorporated, the nascent strand cannot be elongated further.

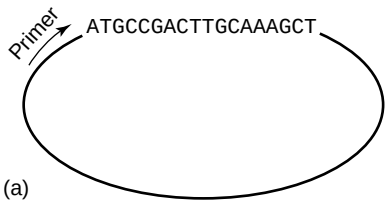

(a)

ddATP	ddCTP	ddGTP	ddTTP
ATGCCGACTTGCAAA ATGCCGACTTGCAA ATGCCGACTTGCA ATGCCGA A	ATGCCGACTTGCAAAGC ATGCCGACTTGC ATGCCGAC ATGCC ATGC	ATGCCGACTTGCAAAG ATGCCGACTTG ATGCCG ATG	ATGCCGACTTGCAAAGCT ATGCCGACTT ATGCCGT AT

(b)

Sequence	A	C	G	T
ATGCCGACTTGCAAAGC**T**				—
ATGCCGACTTGCAAAG**C**		—		
ATGCCGACTTGCAAA**G**			—	
ATGCCGACTTGCAA**A**	—			
ATGCCGACTTGCA**A**	—			
ATGCCGACTTGC**A**	—			
ATGCCGACTTG**C**		—		
ATGCCGACTT**G**			—	
ATGCCGACT**T**				—
ATGCCGAC**T**				—
ATGCCGA**C**		—		
ATGCCG**A**	—			
ATGCC**G**			—	
ATGC**C**		—		
ATG**C**		—		
AT**G**			—	
A**T**				—
A	—			

(c)

Fig. 3 (a) The sequencing reaction generates fragments from the template, which is originally cloned into a suitable sequencing vector. (b) For each of the four chain terminators, all possible fragments that terminate at the 3′ end with that nucleotide are generated in the sequencing reaction. (c) When electrophoresed on a polyacrylamide gel, the fragments can be visualized, and the sequence can be read directly from the gel.

into a nucleotide sequence. Translation of a chromatogram into a sequence is called *base calling*. In the chromatogram, nucleotides correspond to peaks of fluorescence, which are displayed with different colors representing different nucleotides. Because of the obvious need for rapid base calling in large-scale genome projects, software has been developed to automate the process. There are several base-calling

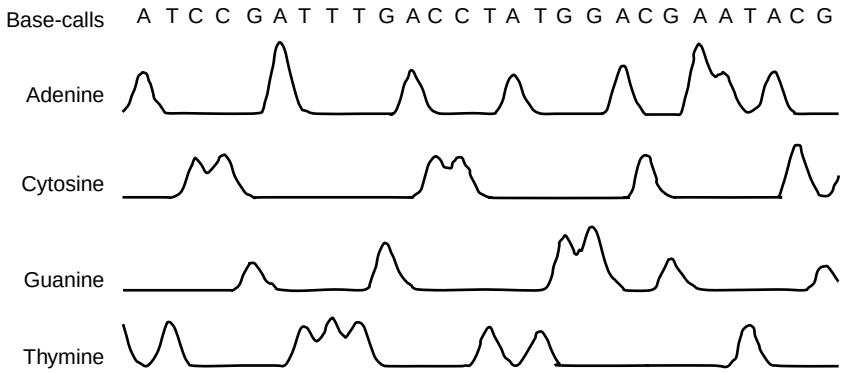

Fig. 4 A sequence chromatogram generated by an automatic sequencer and visualized as fluorescent intensity (*y*-axis) and electrophoretic mobility (*x*-axis).

programs available, including commercial software provided by manufacturers of automated DNA sequencers. The Phred program, developed by Phil Green at the University of Washington, is the most commonly used program today.

Phred base-calls in a four-step process. First, peak locations are predicted using Fourier methods. Prediction of peak location is necessary for regions where nucleotides are unreadable or ambiguous. In the next two steps, the program identifies observed peaks and matches them to the predicted peak locations. As observed peaks are matched to the predicted peak locations, the program is able to identify and resolve ambiguities in the sequence. Finally, any unambiguous observed peak that was not matched to a predicted peak is called and inserted into the sequence.

Owing to problems inherent in the sequencing reaction and in electrophoresis, sequence traces rarely yield a perfect sequence. Aberrant mobility of short sequences, unreacted sequencing reagents, DNA hairpin structures, and even simple diffusion can all produce regions of ambiguity in sequence traces. Because of these problems, Phred assigns an error probability to each base in the sequence. In calculating the error probability, Phred considers peak spacing, the ratio of the largest uncalled peak to the called peaks, and the peak resolution for the short sequence surrounding each nucleotide. Phred converts the error probabilities into "quality values," which are 10 times the negative logarithm of the error probability. Thus, a quality value of 30 corresponds to an error probability of 1 in 1000. Higher Phred quality values are to be desired.

A second technical advance that facilitated the rate of genome sequencing is the advent of practical capillary electrophoresis. Originally, for each limited set of sequencing reactions, a separate polyacrylamide gel needed to be created for the electrophoresis step, and the slab gel had to be exposed to X-ray film to visualize the radioactively labeled DNA fragments. With fluorescent labels, the sequence can be read by a fluorescent detector during electrophoresis. In either case, the step of creating the slab gel was time consuming, and electrophoresis in a slab gel separates the sequencing fragments slowly. With capillary sequencers, the DNA fragments are fractionated by electrophoresis through a gel embedded in a capillary tube with an internal diameter

of 100 μ. Capillary sequencers require fewer reagents for the gel and completely eliminate the time-consuming process of manually preparing, loading, and editing the electrophoretic gel.

The addition of robotics to DNA sequencing processes provided a means to automate the more repetitive processes. When combined with fluorescent detection and capillary electrophoresis, highly automated sequencing devices such as the ABI PRISM 3730 or 3100 DNA Analyzers and the Amersham Pharmacia MegaBACE 4000 automated sequencing system can sequence hundreds of thousands of nucleotides everyday (Fig. 5). Special plates containing the results of sequence reactions are merely loaded into the machines. Robotics within the machine load each sample into a capillary, and detectors record the fluorescence as the fractionated sequencing fragments approach the end of the capillary.

Owing to limitations of the electrophoresis technique, approximately 500 to 1000 nucleotides can be reliably read from each template, even when using a capillary sequencer. To sequence even a small genome like the 0.5 Mb genome of *Mycoplasma genitalium*, the DNA molecule must be broken into small pieces that are able to be sequenced. The general strategy employed for recovering the complete sequence of a large DNA molecule is called *shotgun sequencing*. First proposed in the early 1980s, shotgun sequencing requires random fragmentation of the longer DNA molecule, cloning the fragments into a sequencing vector, and sequencing the fragments (Fig. 6). The sequence of the original DNA molecule is then assembled from the fragments using computational algorithms.

The first step in shotgun sequencing is the subcloning of the DNA sequencing target, also called *library construction*. In this step, the target DNA is broken into smaller pieces, which are then cloned into vectors that can be used in sequencing reactions. Any collection of clones that represent complete coverage of a larger piece of DNA is called a *library*. To create a sequence-ready shotgun library, the DNA of the target molecule is a fragment enzymatically by a partial digestion with a nuclease or restriction enzyme, or the DNA can be physically sheared, for example, by sonication. Physical shearing generally produces a more randomized sample of the target sequence and is more suitable for high throughput sequencing centers.

Two types of cloning vectors are available for shotgun sequencing, each with its advantages and disadvantages. The

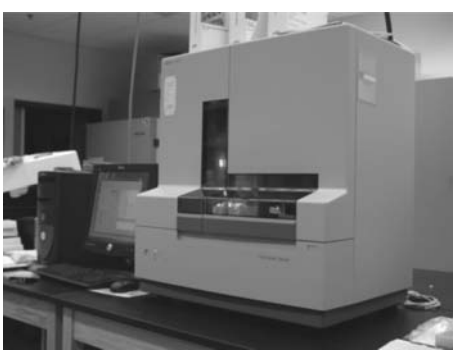

Fig. 5 A modern capillary sequencer, the ABI Prism 3100 Genetic Analyzer.

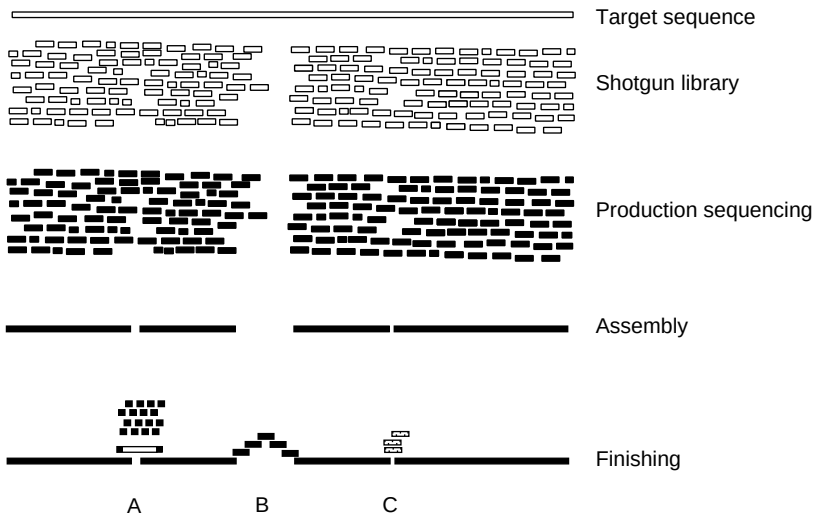

Fig. 6 In the shotgun-sequencing strategy, the target sequence is fragmented into random pieces, which are cloned in a sequencing vector to form the shotgun library. Production sequencing generates sequence data for the shotgun clones to at least a sixfold coverage. Computer programs assemble the shotgun reads into contiguous sequences. During finishing, various strategies are used to close the remaining gaps, including identification by end sequences and shotgun sequencing of a clone that spans the gap A, primer walking B, and alternative chemistry sequencing C.

first is bacteriophage M13, which naturally exists as single-stranded DNA. Since DNA sequencing templates must be single-stranded DNA, M13 naturally provides sequence-ready DNA templates. M13 clones exhibit a bias, in that repetitive sequences are poorly represented in M13 libraries. Alternatively, double-stranded plasmid clones can also serve as the vector for subcloning shotgun libraries. Plasmids do not exhibit the same cloning bias as M13 vectors, and thus represent the target DNA sequence more evenly. The disadvantage of plasmids is that they are double-stranded and therefore require additional steps to purify single-stranded DNA sequencing templates. A large genome-sequencing center will often use a combination of M13 and plasmid subclones for different sequencing needs.

Having created a suitable shotgun library, the next step is typically referred to as *production sequencing*, because it resembles the factory production typical of an assembly line. A random selection of shotgun clones is prepared and sequenced using the automated sequencing techniques discussed above. The number of clones necessary to reconstruct the original target sequence depends on the average length of sequence obtained from a single shotgun clone, the length of the target sequence, and the desired accuracy of the completed sequence. For production sequencing, genome-sequencing centers typically strive for a six- or eightfold coverage of the target sequence; that is, each nucleotide of the target is sequenced six (or eight) different times. This level of redundancy assures that most of the target sequence will be covered by the randomly

positioned shotgun clones, and whatever gaps or ambiguities remain after the shotgun sequencing can be quickly resolved.

After generating a sixfold redundant shotgun sequence, the sequence traces are then processed computationally. First, the traces are base-called using Phred or a comparable program. Second, the sequences are "assembled" into contiguous sequences called *contigs*. A contig is composed of two or more sequence reads that originate from an overlapping region of the target sequence. Contig assembly begins with an automated step, and concludes with a manual editing stage. The automated step is carried out by an assembler program such as Phrap, the TIGR Assembler, or CAP3. The assembler examines the sequence reads for regions of near identity (allowing for errors in the sequence read) and attempts to reconstruct the original target sequence. If the sequence reads were a perfect and redundant random representation of the target sequence, the assembler could theoretically reconstruct the entire target sequence from the shotgun sequence. In reality, variations in the quality of sequence reads, regions of the target with low shotgun representation, and repetitive sequences can result in assembly of many different contigs (depending on the size of the target sequence and the redundancy of the shotgun clones).

All assembly programs work in very similar ways. First, regions of overlap are identified by doing standard sequence comparisons between the sequence reads. Unlike a normal sequence comparison program, however, assemblers also consider the quality values of the sequence reads in calculating similarity scores between reads. From the overlap information, the assembler then reconstructs sequence contigs and some also calculate quality values of each nucleotide in the contig, based on the number of reads available at that position. To be certain that the contig comes only from the target sequence, sequence reads are typically filtered before assembly in order to remove contaminating DNA, such as vector sequence, that often appears at the 5′ end of a sequence read.

Having completed the automatic assembly, the sequence rarely resolves into a single contig of uniform quality. Reasons for obtaining multiple contigs include failure of the assembly program, cloning bias, regions that are difficult to sequence, and regions that were poorly represented in shotgun sequence reads by chance. To resolve these problems and to produce a single contig of uniform quality, manual inspection and editing of the contigs is necessary, together with additional sequencing reactions where appropriate. This manual phase is called either *gap closure* or more commonly *finishing*.

Finishing is aided by the use of a graphical contig editor such as Consed. For each assembly, Consed displays a list of the contigs and sequence reads. After selection of a contig, Consed displays the consensus sequence together with each sequence read that was assigned to that region by the assembler. The consensus sequence is displayed at the top of the assembly window, and the sequence reads appear in an alignment format just below the consensus (Fig. 7). Important features such as base quality are color-coded for easy identification of low-quality regions. Other windows in Consed display the actual chromatograms of any desired sequence reads.

The additional work necessary to close a gap in the sequence assembly depends on the cause of the gap. Regions of low shotgun representation can often be resolved by sequencing the opposite end

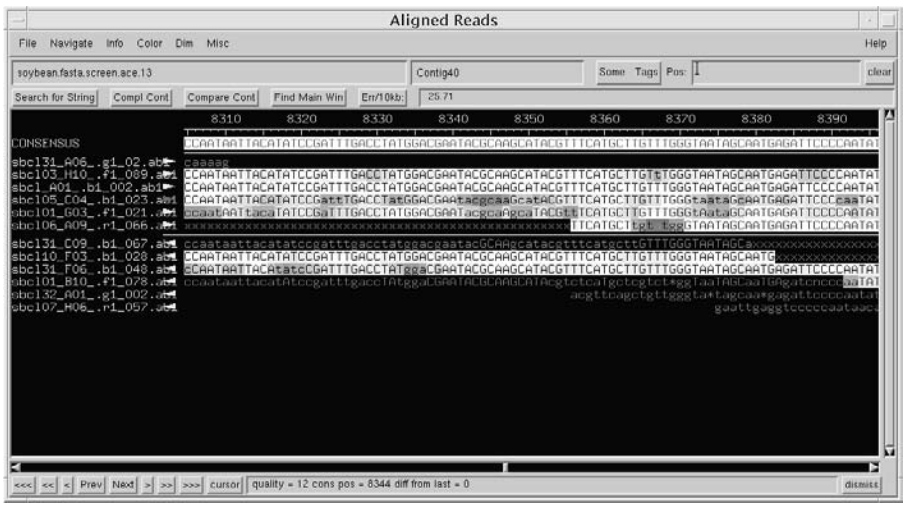

Fig. 7 A sequence assembly shown in a window of Consed. The consensus contig sequence is displayed at the top of the window, with individual sequence reads aligned and displayed beneath. The background color indicates the relative quality of base-calls. Uppercase letters with white backgrounds are of high quality; lowercase letters with black background are of poor quality. Nucleotides represented as *x* come from contaminating vector sequences, which have been removed prior to assembly.

of the clone insert for reads adjacent to the low-quality region. This will generate sequence reads for both ends of the clone insert and will usually provide sufficient sequence coverage to allow the gap to be closed. If the gap is caused by vector bias, it will be necessary to clone the missing region into a different vector. For example, as we discussed above, M13 does not effectively clone regions of repetitive DNA. Utilization of both M13 and plasmid clones can resolve problems caused by this cloning bias. In cases in which the assembly algorithm has failed to derive the correct consensus sequence, the statistical parameters of the assembly algorithm can be manipulated to generate the desired assembly. For example, assembly algorithms assemble regions with tandem repeats poorly. These errors are easily identified because of the abnormally large number of sequence reads assigned to a single region.

The most difficult class of problems encountered in finishing are regions that are intrinsically difficult to sequence, such as simple sequence repeats, homopolymeric regions, or regions with secondary structure. These types of gaps are easily identified because of the uniformity with which the quality drops off. In other types of gaps, the low quality may be observed in only a few of the shotgun reads, but in areas that are difficult to sequence, all shotgun reads will exhibit the same low quality at the same point in the sequence.

Resolution of sequence gaps requires a number of different strategies, each designed to deal with a particular sequencing problem. For example, specialized chemistry designed for simple sequence repeats can produce high-quality reads of repetitive regions. Larger repeats or secondary

structures could require a targeted subcloning strategy designed to break up the difficult region prior to sequencing. In another finishing strategy, called *primer walking*, sequencing primers are designed from the sequences at the boundaries of gaps. The sequencing reaction is then repeated, generating new sequence. This process of primer development and sequencing is repeated until the gap is closed.

After *finishing* has produced a single contig, the consensus sequence is validated in a number of ways. First, the length of the sequence is compared to the expected length of the target sequence. The expected length is typically measured by a restriction digest of the target sequence. Second, the length of the restriction fragments observed for the target sequence are compared to the length of the restriction fragments predicted for the consensus sequence. Whereas matching the length to the expected length confirms that a sequence of the correct size has been generated, matching the predicted and observed restriction fragment lengths ensures that the general order of the sequence assembly is good. This procedure can be repeated with multiple restriction enzymes to verify the assembly even further.

3
Sequencing Bacterial and Archaeal Genomes

As outlined in the previous section, the basic technique of shotgun sequencing is suitable for contiguous sequences of 10 Mb or less with few repeat sequences. For the small genomes of many bacteria and archaea, a straightforward shotgun strategy works well. The size and repetitive nature of eukaryotic genomes prevent the ready application of a shotgun-sequencing technique, necessitating specialized strategies for sequencing these genomes. Depending on the level of completion required, eukaryotic genomes are sequenced by a mapping strategy (for high-quality sequences) or by a whole-genome shotgun (WGS) strategy (for rough draft sequences), or by a hybrid of both. Strategies currently in development attempt to target the genes of large eukaryotic genomes by specialized cloning techniques. In this section, we will discuss the shotgun strategy used to generate the first bacterial genome sequence, *Haemophilus influenzae*. The following section will focus on strategies used in eukaryotic genome sequencing.

The first organismal genome completely sequenced by the shotgun method was that of *H. influenzae*, sequenced by a team from the Institute for Genomic Research (TIGR) and published in 1995. As a bacterial genome, the repetitive DNA content was much lower than most eukaryotes, and the size of the complete chromosome was only 1.8 million nucleotides. Despite being the first organismal genome, the issues considered and strategies employed in sequencing are still representative of modern genome projects.

Because the success of the shotgun method depends on a random selection of sequence templates, the TIGR team expended much care on template library construction. Two types of libraries were made for the *H. influenzae* project: a short-insert (1.6–2 kb) plasmid library and two long-insert (15–20 kb) λ phage libraries. The plasmid library was used as the primary sequencing template, while the λ libraries were reserved for finishing and validation (we will discuss the utilization of the λ libraries below). For all libraries, the genomic DNA was

mechanically sheared rather than digested by restriction enzymes. As mentioned previously, shearing the source DNA ensures a more random library sample than restriction digest.

On the basis of simple statistical calculations, we know that the probability that a nucleotide will be unsequenced in a shotgun-sequencing project is $P_o = e^{-m}$, where m is the genome coverage. Sequencing random clones sufficient to cover the genome five times (fivefold coverage) results in a probability of 0.0067 that a nucleotide will be unsequenced. For the 1.8 million-nucleotide *H. influenzae* genome, a fivefold shotgun coverage should result in approximately 12 000 unsequenced nucleotides, distributed randomly throughout the genome in sequence gaps between the contigs.

Given an average high-quality read-length of 460 nucleotides (the limits of the sequencing technology at that time), approximately 19 000 sequence reads would be necessary to cover the *H. influenzae* five times. The TIGR team sequenced 19 687 short-insert templates using dye-primer sequencing chemistry (*forward reads*). To supplement these reads and to provide contig assembly information, 9297 templates were re-sequenced at the opposite end of the insert (known as *reverse reads*), also using dye-primer chemistry. The result is a pair of 460-nucleotide sequence reads that are known to be 700–1100 nucleotides apart, a significant advantage during finishing.

Automatic assembly of forward and reverse reads yielded 210 contigs. Because the parameters used in assembly are optimized for general use, local variations can occur due to differences in repeat- or GC-content. As a result, manual inspection of potential overlaps resulted in a reduction to only 140 contigs, which could not be further combined without additional sequencing reactions. The remaining gaps could be categorized into two types depending on the orientation of paired sequence reads. If the forward reads at the end of one contig matched the corresponding reverse reads from the same template at the end of a second contig, the TIGR team called the gap a *sequence gap*. Because they are spanned by a single, small-insert plasmid, sequence gaps are known to be small (less than 1500 nucleotides), and known templates (the plasmid inserts spanning the gaps) are available for immediate, additional sequencing. For the *H. influenzae* genome, 98 gaps were sequence gaps.

The second type of gap occurred when forward/reverse reads of the same template did not span the gap. The TIGR team labeled these gaps *physical gaps*, because a clone insert that spanned the gap was not immediately available. Additional strategies were devised to identify and prepare suitable sequencing templates that spanned the physical gaps. For two gaps, protein sequences could be used to orient the adjacent contigs. This was possible because the gap occurred within a protein-coding gene, the 5′ end of which was sequenced on one contig and the 3′ end on another. By comparing the contig sequences to known protein sequences, these overlaps could be detected and PCR used to generate suitable sequencing templates.

The majority of the gaps were closed using one of two strategies. First, DNA fingerprinting and hybridization to oligonucleotides prepared from the ends of contigs revealed possible overlapping DNA segments, from which templates could be prepared. For example, if oligonucleotide probes prepared from the ends of two

contigs hybridized to the same restriction fragment of the genomic DNA, it is likely that the hybridizing fragment contains the sequence spanning the gap between the two contigs. The other successful gap closure method was the paired forward/reverse sequences from the λ libraries. Small gaps are more likely to be spanned by the large inserts (15–20 kb) of the λ libraries than the small inserts (~2 kb) of the plasmid library. The remaining physical gaps were closed by simple combinatorial PCR, systematically using oligonucleotide primers from each possible contig pair.

These strategies resulted in a single contig with a consensus sequence of 1 830 137 nucleotides. The assembly of the consensus sequence was validated using additional λ library forward/reverse sequence reads and restriction fragments and restriction site locations. Since the λ library has a known insert size of 15–20 kb, paired reads that deviate significantly from that distance would indicate a misassembly. Sizes of restriction fragments generated by three different enzymes could also be matched to the predicted sizes of the consensus sequence. Restriction site locations from restriction mapping could also be matched to the restriction sites found on the consensus sequence. For all of these validations, the *H. influenzae* consensus sequence matched the observed characteristics of the genome.

Despite being the very first bacterial genome to be sequenced by the shotgun method, much of the techniques and strategies developed are the same as those used today for bacterial and archaeal genome projects. Library construction remains an important step to assure a random coverage of the genome. Sequencing both ends of the inserts and using different sequencing chemistries are both common strategies for gap closure and finishing. The major difference between a modern genome project and the *H. influenzae* project is the rate. Facilitated by capillary sequencers and more powerful computers for assembly and finishing, a bacterial genome of the same size as *H. influenzae* can be completed in a fraction of the time and at a fraction of the cost.

4 Sequencing Eukaryotic Genomes

Eukaryotic genomes differ from bacterial and archaeal genomes in several important ways that make genome-sequencing projects more complex. Although several bacteria and archaea are known to have multiple chromosomes or *megaplasmids*, a typical prokaryotic genome is a single, circular chromosome. In contrast, eukaryotic genomes are divided into several (sometimes many) linear chromosomes. Eukaryotic genomes are also much larger, on average: the largest bacterial genome sequenced to date is the 9 Mb chromosome from *Bradyrhizobium japonicum*. At 3200 Mb, the human genome is approximately 355 times larger (see also Fig. 8). The size and multiple chromosomes alone would challenge the assembly algorithms and finishers, but the sequence of eukaryotic genomes also contain a high repeat content, rendering cloning, sequencing, and assembly of particular regions extremely difficult.

Because of these limitations, extra presequencing strategies must be utilized to make the genome amenable to sequencing, and the definition of "finished genome" is often project-specific. The most common strategy employed for eukaryotic genome sequencing is a "map-then-sequence" approach that seeks to

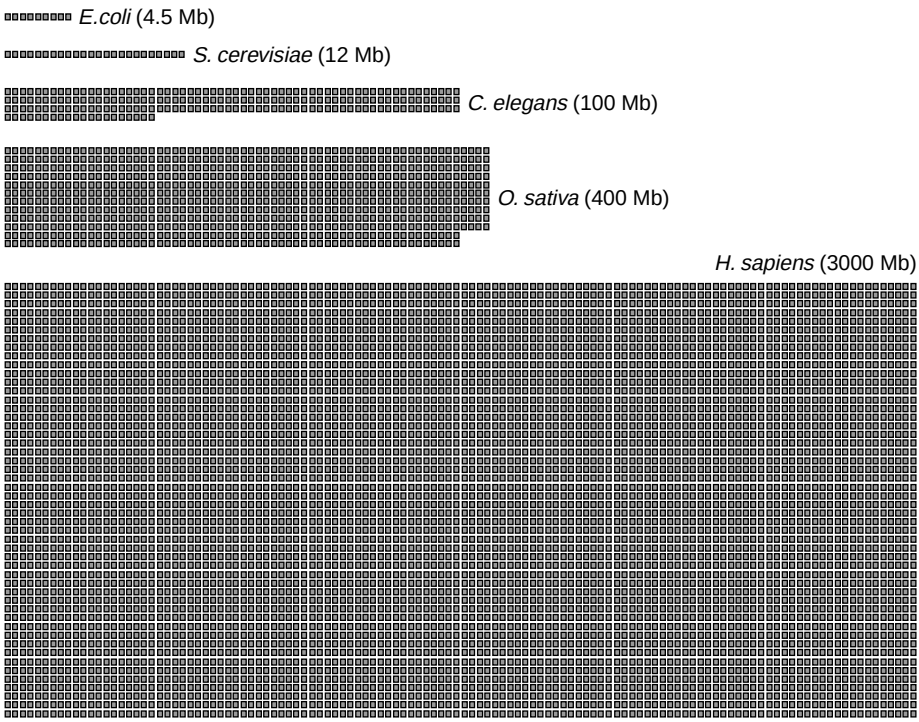

Fig. 8 Relative sizes of sequenced genomes. Each square represents 500 000 nucleotides, approximately the size of the complete *M. genitalium* genome. Genomes shown represent the K12 strain of *E. coli*, baker's yeast *S. cerevisiae*, the nematode *C. elegans*, rice *O. sativa*, and the human genome *H. sapiens*.

obtain a physical map of the genome prior to the shotgun-sequencing phase. The physical map is composed of ordered, overlapping large-insert clones anchored to markers in the genome, identified by genetic or cytogenetic mapping. On the basis of the order and orientation of the map, individual clones are selected and subjected to the standard shotgun sequencing and finishing, as described above (Fig. 9). There are several advantages to this strategy, including an overall reduction in the number of repeats that need to be resolved during any particular finishing phase and the identification of problem regions before sequencing even begins.

Ideally, the physical map and finished genome sequence should cover the entire genome, but this level of accuracy is rarely achieved. Regions of concentrated repeats, such as telomeres and centromeres, do not clone well and are extremely difficult to sequence. Regions of high repeat content are called *heterochromatin*. Euchromatin has a much lower repeat concentration and is believed to contain the majority of the genes. As a result, most eukaryotic sequencing projects aim to completely sequence the euchromatic regions of the genome. For example, approximately one-third of the fruitfly (*Drosophila melanogaster*) genome was heterochromatic and consequently not sequenced. Even when dealing with just

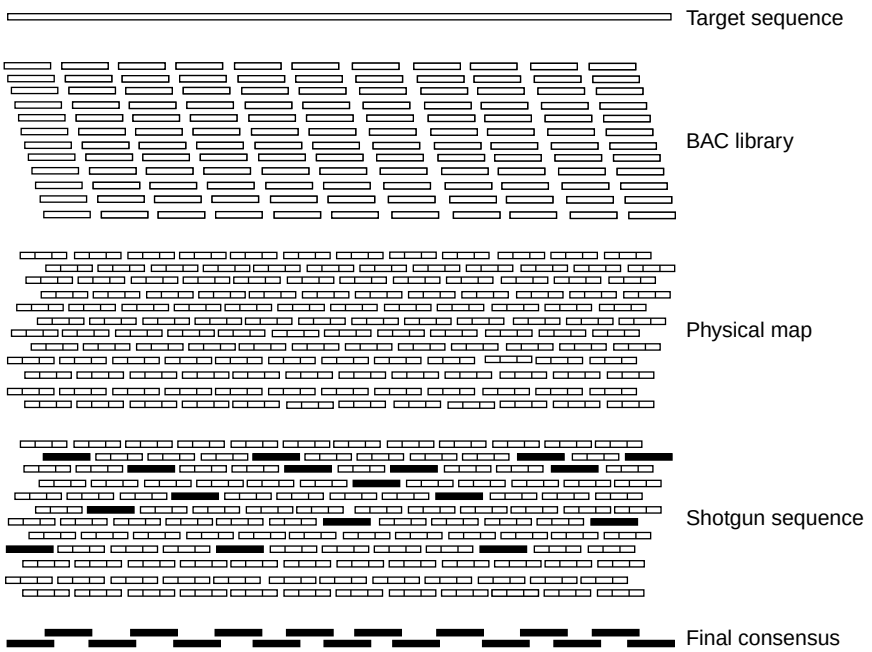

Fig. 9 The map-then-sequence strategy used for large, eukaryotic genomes. The target sequence is fragmented by partial restriction digest and cloned into a large-insert BAC library. The BAC inserts are fingerprinted and overlaps detected to generate a physical map. On the basis of the physical map, clones are selected for shotgun sequencing (see Fig. 6), and the final consensus consists of a series of overlapping clones.

euchromatin, small regions that are difficult to sequence can still occur, and most eukaryotic genome projects allow for a certain number of gaps in the "finished" product, provided the gap length is known. The public Human Genome Project allows an unspecified number of gaps, as long as their length is less than 150 kb.

Because physical mapping is beyond the scope of this article, we will here provide only a cursory overview of the map and sequence strategy. A physical map begins with the construction of large-insert libraries. Currently, the vector of choice for mapping is the bacterial artificial chromosome (BAC). BACs can hold approximately 100–300 kb of insert and are not prone to recombination. Unlike shotgun libraries, the source DNA for a BAC library is typically fragmented by partial restriction digest. Protocols for BAC cloning randomly sheared DNA are still in development. Because of the inherent biases introduced by partial restriction digests, multiple libraries constructed with different restriction enzymes are often necessary to get complete, random coverage of the genome.

After the construction of the genomic library, BAC inserts are ordered and aligned by "fingerprint" analysis. Each clone is completely digested with the same restriction enzyme, and the fragments are fractionated and viewed by agarose gel electrophoresis (Fig. 10). A new technology called *high information*

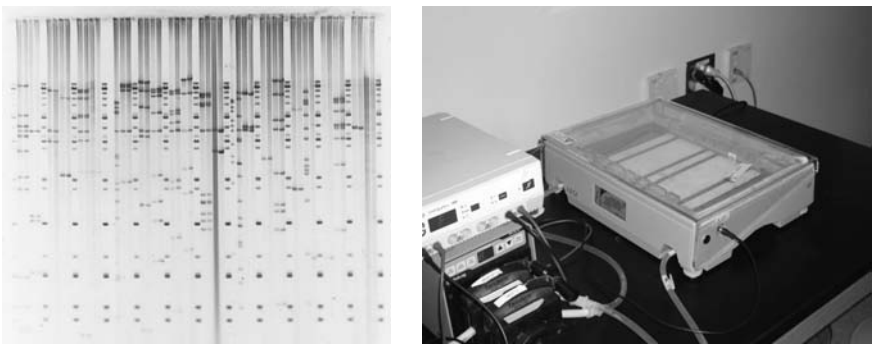

(a) (b)

Fig. 10 (a) A fingerprint gel generated from a large-insert BAC library of rice (b) and the electrophoresis apparatus used for fingerprinting.

content fingerprinting (HICF) utilizes up to five different restriction enzymes followed by dye labeling and electrophoresis on a DNA capillary sequencer rather than a standard agarose gel. HICF reduces labor, increases automation and the number of fragments per BAC, and allows automatic measurement of fragment mobility from chromatograms. Owing to the increased fragment number, more robust contigs can be generated from the fingerprint data.

The relative mobilities of each clone's restriction fragments are loaded into a computer program (such as FPC) that determines overlapping regions between clones based on shared fragment mobilities. The result is a set of physical map contigs (Fig. 11). Because restriction fingerprints provide less data than actual sequencing, the assembly of physical map contigs is less precise than sequence contigs and consequently requires a deeper coverage of the genome. Many physical maps cover the genome 10 to 15 times. As with shotgun sequencing, the coverage is uneven, and many difficult-to-clone regions of the genome may be missing from the physical map contigs. Closure of gaps in physical maps takes place through both computational and additional laboratory experiments, using some strategies analogous to gap closure in sequence finishing.

The goal of the physical mapping effort is to produce a set of minimally overlapping clones that cover the entire euchromatic portion of the genome, also called a *sequence-ready* map. Because the coverage can be uneven, the physical map must be "anchored" by hybridization with marker sequences of a known, unique genomic position. These markers can be identified by a number of mapping methods, including traditional genetic mapping. Once the physical map is anchored, any remaining gaps in the euchromatin coverage can be identified, and strategies to fill those gaps can be devised and implemented.

As an aid to the creation of the sequence-ready map, the ends of the BAC clone inserts can be sequenced. Although some of these "BAC-end sequences" or "sequence-tagged connectors" (STCs) will undoubtedly come from repetitive sequence, many will be unique sequences that represent the ends of only one clone insert. STCs of clones bordering gaps in the physical map can be used to generate novel probes to identify clones and possibly contigs that span the gap. During the

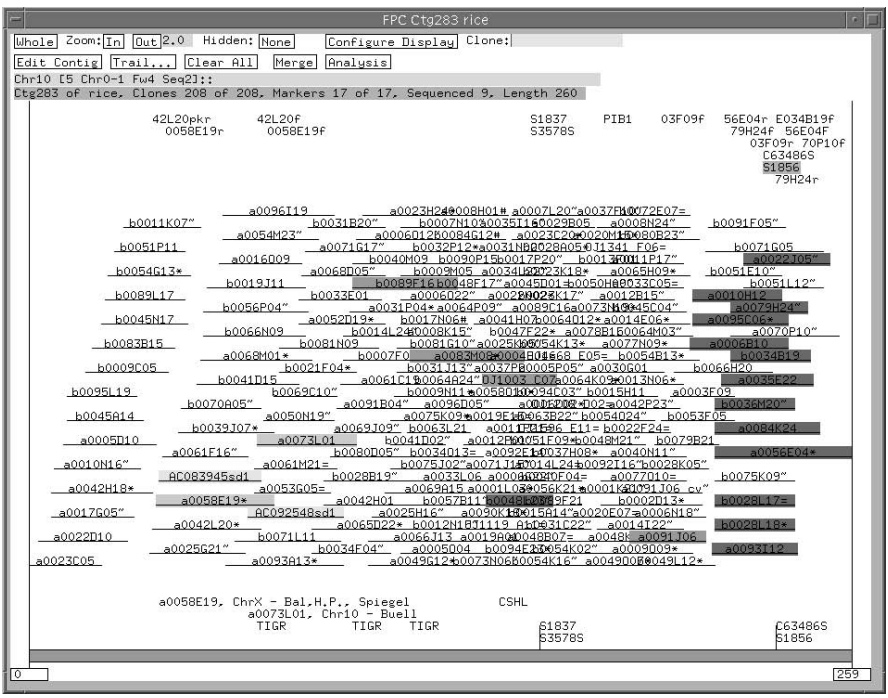

Fig. 11 A physical map contig from the rice genome, as displayed by the physical mapping software FPC.

sequencing phase, STCs can be used to choose the next minimally overlapping clone with a great deal of precision. STCs can also be a valuable dataset in their own right by providing a preliminary survey of the contents of a eukaryotic genome.

After (or even during) construction of the physical map, sequencing can commence using the standard shotgun method. A sequence-ready clone is chosen, shotgun libraries are prepared, and the clone is sequenced. Additional clones are selected and sequenced with minimal overlap with the existing sequence. Finishing is limited to the regions that do not overlap with an existing, finished sequence.

With a physical map, another advantage becomes immediately apparent. Unlike a pure shotgun approach for small genomes, the sequencing of a mapped genome can be divided between more than one sequencing center. The genome can be divided into regions or even whole chromosomes, and sequencing can take place in many different labs simultaneously. Additionally, even at a single sequencing center, sequencing can commence at different positions in the genome. As new sequence-ready clones are identified by physical mapping, shotgun sequencing and finishing can commence on other clones. Thus, the mapping and sequencing at a single lab can proceed simultaneously.

The map-then-sequence strategy has been successfully applied to several prominent eukaryotic genome projects. Genome sequences of the nematode *Caenorhabditis elegans* and the plant *Arabidopsis thaliana* were both completed in this way, and the public Human Genome Project also

used this method. The alternative strategy is the whole-genome shotgun (WGS) or a mixture of mapping and WGS. WGS sequences are rarely finished to the same quality as a mapped-and-sequenced genome, but for certain genome projects, WGS provides a reasonable and affordable alternative to a completed sequence. Because of the unfinished quality of WGS sequences, they are often referred to as *draft sequences*.

Though the vast majority of genome sequence is obtained during the shotgun-sequencing phase, the majority of effort (and money) is invested in the finishing phase. With the complete human genome sequence, finished genome sequences from other vertebrates may be an unnecessary expense for comparatively little information. Because genomes of mammals and vertebrates have conserved gene content and order, the human genome sequence can serve as a reference template for other genome sequences. A WGS project could generate four- or fivefold coverage of the genome and still theoretically achieve 98% coverage of the euchromatic region of the genome. A WGS for a vertebrate would produce a very large number of sequence contigs that could be ordered by reference to the human genome.

As with a standard shotgun-sequencing project, a eukaryotic WGS begins with careful construction of genomic clone libraries. Celera's attempt to sequence the human genome by WGS showed that multiple clone libraries with differing insert sizes are essential to the assembly of sequence contigs. For each library, both ends of the clone inserts must be sequenced to provide pairs of sequences of a known distance apart on the chromosome. Paired-end sequences are necessary to order and orient sequence contigs into "*scaffolds*." Finally, the actual assembly of eukaryotic WGS sequence reads usually proceeds in a multistep process, even with advanced computational hardware. Typically, the repeat sequences will be masked in the early assembly steps to prevent misassembly due to highly conserved repetitive sequences. Once contigs and scaffolds of unique sequences have been assembled, the repeat sequences can be added back to the assembly with the unique sequences acting as an assembly framework.

A third application of WGS is to use a mixed approach, with elements of both map-then-sequence and WGS. For example, a low-coverage WGS can be generated rapidly at a large genome-sequencing center while finishing could be completed on a clone-by-clone basis at smaller sequencing centers. This hybrid approach was used for the *Drosophila* genome project. Celera provided a nearly 15-fold shotgun coverage of the *Drosophila* genome, and finishing was completed in a number of publicly funded laboratories.

Despite the advantages of WGS for some eukaryotes, the drawbacks are also significant. Experience with the rice genome has shown that contigs produced by WGS may be a poor substrate for *ab initio* gene prediction (see below). In a detailed analysis of the complete sequence of rice chromosome 10, researchers predicted twice as many genes as had been predicted by previous WGS sequencing efforts of the same chromosome. Additionally, genes predicted from the WGS contigs were on average one-third shorter than genes predicted from the finished sequence.

Several new techniques have been proposed for specifically targeting gene-rich sequences during the initial library construction, prior to shotgun sequencing. In one method, methylated DNA is eliminated prior to library construction. In

eukaryotic genomes, transposons are often methylated at CpG dinucleotides as a means of downregulating their transcription. By excluding methylated DNA from library construction, only unmethylated genes should be included in the library. The primary drawback of this method is that most organisms use methylation as a means of transcriptional control of cellular genes as well as transposons. The second method of gene-targeted cloning, called *cot-based cloning and sequencing* (CBCS), relies on the fact that repetitive DNA reanneals more quickly than unique DNA after denaturation. By isolating the DNA that reanneals the slowest, nonrepetitive DNA can be cloned without eliminating methylated genes. Both these techniques are still in the development phase and have not yet been widely applied to genome sequencing.

5
Interpreting the Sequence

Once the genome sequence has been completed to a previously defined set of quality standards, the biologically important features of the sequence can be identified in a process called *annotation*. Biologically important features identified in a typical genome project include genes and transposable elements. For bacterial or archaeal genomes, gene identification is relatively straightforward, while gene identification in eukaryotic genomes is much more difficult due to the presence of introns. Three basic methods are used for protein-coding gene identification: *ab initio* gene prediction, sequence similarity searches, and expressed-sequence tag (EST) sequencing. All genome projects use both gene prediction and homology searching, while EST sequencing is primarily a technique for eukaryotic gene identification.

The goal of *ab initio* gene prediction is to correctly identify the start and stop codons (and intron/exon boundaries, if relevant) of a gene given only the DNA sequence and statistical parameters derived from known genes. For the simple structure of a bacterial or archaeal genome, genes can be identified from open reading frames that match a known set of codon preferences. Glimmer is a popular program for prokaryotic gene prediction. Statistical parameters considered for eukaryotic genes can include codon preferences, frequencies of longer combinations of codons, and attributes of intron/exon boundaries. These statistical parameters are derived from analyses of a large number of experimentally identified and sequenced genes. Because eukaryotic gene-prediction algorithms must correctly identify all possible exons, some of which can be very short, gene prediction in eukaryotic DNA is much less accurate than in prokaryotic DNA. Genscan and FGenesH are commonly used eukaryotic gene-prediction programs.

Genes can also be identified by their similarity to known sequences. Programs such as BLAST or FASTA compare a query sequence to a database of sequences. For each sequence in the database, a similarity score is calculated, and the statistical significance of each similarity score can be estimated with a high degree of accuracy. If the query sequence matches a sequence in the database with a statistically significant similarity score, we can infer with confidence that the sequences are related either through a common biochemical or physiological function. Because the statistical properties of sequence similarity scores are well understood, sequence similarity searching is the most

reliable computational method of identifying protein-coding genes, RNA genes, pseudogenes, and transposable elements. Because sequence similarity programs do not identify start and stop codons or intron/exon boundaries, additional manual editing is necessary to derive a gene prediction from the significant similarity to a known sequence.

Because eukaryotic gene-prediction algorithms do not produce completely reliable results, experimental evidence is often sought to supplement computational gene identification efforts. The most common experimental method of gene identification is to isolate messenger RNAs, clone their cDNAs into a sequencing vector, and generate a single sequence read from each clone. Although the single read is insufficient to sequence the entire cDNA to high quality, it is generally enough to generate 500–700 nucleotides of the cDNA. This sequence information serves as a tag that can mark the location of a gene in genomic sequence, hence the name expressed-sequence tag. EST sequences are collected into databases, which are then compared to genomic sequence using sequence comparison programs (BLAST and FASTA). Gene identification by EST is limited by the isolation and preparation of pure mRNA sequences, because rRNA and other stable RNA can contaminate the cDNA library. Furthermore, any one tissue only transcribes a fraction of potential genes at any one time. cDNA libraries should be constructed from multiple tissues at multiple growth stages.

A typical eukaryotic genome project will utilize every gene identification method during annotation. Several different *ab initio* gene-prediction programs will be run on the genomic sequence. Additionally, sequence similarity searches will be performed on several different databases, including previously characterized protein sequences, known transposable elements, and ESTs from the target organism (and closely related species where available). The results of these computational analyses are then manually edited into a final annotation by a trained genome researcher. The annotation and the sequence can be deposited in one of the public DNA sequence databases, such as GenBank.

6
Summary and Principles

As consumers of genomic information, the majority of biologists are not directly involved in the generation of genome sequence data. Consequently, it is important for biologists to understand how genome sequence data is obtained in order to appreciate the advantages and limitations of using complete genomes. As we have explained, there is no single definition of "complete genome" to which all genome projects conform. Different levels of completion provide their own strengths and weaknesses. Sequences completed to the quality of most bacterial genomes and early eukaryotic genomes (yeast, *Caenorhabditis elegans*, *Arabidopsis*) give excellent insight into both the content and organization of the genome. For eukaryotes, such high-quality genome sequences require significant investment of time and funding. Draft sequences created by WGS are excellent for comparative genomics with close relatives (especially if the close relative has a high-quality completed sequence) and can provide insight into gene content. Draft sequences are probably unsuitable for comparative genomics with distantly related organisms and may not be adequate for *ab initio* gene prediction.

It is helpful to remember that the completed genome sequence is experimental data and subject to experimental error. Even high-quality, finished genome sequences will have errors. Most genomic sequence that is subject to finishing is completed to an accuracy of 1 error in 10 000 nucleotides. With three billion nucleotides, we should expect 300 000 single-nucleotide errors in the human genome sequence. For most researchers, this level of error will hardly ever be noticeable, but occasionally researchers may stumble across one of these errors.

Finally, researchers should also keep in mind that efforts to annotate genomic sequence, and in particular to identify protein-coding genes, are subject to much higher and less quantifiable error rates than actually obtaining the sequence itself. For important research projects on particular genes or gene families, predicted genes should be subjected to experimental verification, based on the evidence used to predict the gene. Genes predicted from a combination of EST similarity, *ab initio* predictions, and significant similarity to a closely related protein sequence will require less verification than those predicted from *ab initio* gene-prediction software alone.

Presently, genome technology and sequencing continues to grow at an amazing rate. Many of the strategies and limitations discussed in this article could be resolved at some time in the future. As genomics continues to grow, our understanding of the chemical basis of life will also grow, providing new platforms and methods for understanding disease, inheritance, and evolution. The benefits of investing in genome sequencing will continue to be discovered for years to come.

See also Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Gibson, G., Muse, S.V. (2002) *A Primer of Genome Science*, Sinauer, Sunderland, MA.

Green, E.D. (2001) Strategies for the systematic sequencing of complex genomes, *Nat. Rev. Genet.* **2**, 573–583.

Green, E.D., Birren, B., Klapholz, S., Myers, R.M., Hieter, P. (Eds.) (1999) *Genome Analysis: A Laboratory Manual*, 4 vols, CSHL Press, Cold Spring Harbor, New York.

Mount, D.W. (2001) *Bioinformatics: Sequence and Genome Analysis*, CSHL Press, Cold Spring Harbor, New York.

Primary Literature

Adams, M.D., et al. (2000) The genome sequence of *Drosophila melanogaster*, *Science* **287**, 2185–2195.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* **25**, 3389–3402.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R., Young, I.G. (1981) Sequence and organization of the human mitochondrial genome, *Nature* **290**, 457–465.

Balding, D.J., Bishop, M., Cannings, C. (Eds.) (2001) *Handbook of Statistical Genetics*, Wiley, New York.

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S. and Venter, J.C. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*, *Science* **273**, 1058–1073.

Burge, C., Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA, *J. Mol. Biol.* **268**, 78–94.

Chen, M., Presting, G., Barbazuk, W.B., Goicoechea, J.L., Blackmon, B., Fang, G., Kim, H., Frisch, D., Yu, Y., Sun, S., Higingbottom, S., Phimphilai, J., Phimphilai, D., Thurmond, S., Gaudette, B., Li, P., Liu, J., Hatfield, J., Main, D., Farrar, K., Henderson, C., Barnett, L., Costa, R., Williams, B., Walser, S., Atkins, M., Hall, C., Budiman, M.A., Tomkins, J.P., Luo, M., Bancroft, I., Salse, J., Regad, F., Mohapatra, T., Singh, N.K., Tyagi, A.K., Soderlund, C., Dean, R.A., Wing, R.A. (2002) An integrated physical and genetic map of the rice genome, *Plant Cell* **14**, 537–545.

Delcher, A.L., Harmon, D., Kasif, S., White, O., Salzberg, S.L. (1999) Improved microbial gene identification with GLIMMER, *Nucleic Acids Res.* **27**, 4636–4641.

Ewing, B., Green, P. (1998) Base-calling of automated sequencer traces using *Phred*. II. Error probabilities, *Genome Res.* **8**, 186–194.

Ewing, B., Hillier, L., Wendl, M.C., Green, P. (1998) Base-calling of automated sequencer traces using *Phred*. I. Accuracy assessment, *Genome Res.* **8**, 175–185.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd, *Science* **269**, 496–512.

Goff, S.A., Ricke, D., Lan, T., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., Briggs, S. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*), *Science* **296**, 92–100.

Gordon, D., Abajian, C., Green, P. (1998) *Consed*: a graphical tool for sequence finishing, *Genome Res.* **8**, 195–202.

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Meg, J., Trevaskis, E., Underwood, K., Wohldman, P., Waterston, R., Wilson, R., Marra, M. (1996) Generation and analysis of 280 000 human expressed sequence tags, *Genome Res.* **6**, 807–828.

Huang, X., Madan, A. (1999) CAP3: a DNA sequence assembly program, *Genome Res.* **9**, 868–877.

International Human Genome Mapping Consortium. (2001) A physical map of the human genome, *Nature* **409**, 934–941.

International Human Genome Sequencing Consortium. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Marra, M.A., Kucaba, T.A., Dietrich, N.L., Green, E.D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D., Waterston, R.H. (1997) High throughput fingerprint analysis of large-insert clones, *Genome Res.* **7**, 1072–1084.

McPherson, J.D. (1997) Sequence ready – or not? *Genome Res.* **7**, 1111–1113.

Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H., Remington, K.A., Anson, E.L., Bolanos, R.A., Chou, H.H., Jordan, C.M., Halpern, A.L., Lonardi, S., Beasley, E.M., Brandon, R.C., Chen, L., Dunn, P.J., Lai, Z., Liang, Y., Nusskern, D.R., Zhan, M., Zhang, Q., Zheng, X., Rubin, G.M., Adams, M.D., Venter, J.C. (2000) A whole-genome assembly of *Drosophila*, *Science* **287**, 2196–2204.

Pearson, W.R. (2000) Flexible sequence similarity searching with the FASTA3 program package, *Methods Mol. Biol.* **132**, 185–219.

Rabinowicz, P.D., McCombie, W.R., Martienssen, R.A. (2003) Gene enrichment in plant genomic shotgun libraries, *Curr. Opin. Plant Biol.* **6**, 150–156.

Rabinowicz, P.D., Schutz, K., Dedhia, N., Yordan, C., Parnell, L.D., Stein, L., McCombie, W.R., Martienssen, R.A. (1999) Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome, *Nat. Genet.* **23**, 305–308.

Rice Chromosome 10 Sequencing Consortium. (2003) In-depth view of structure, activity, and evolution of rice chromosome 10, *Science* **300**, 1566–1569.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A., Slocombe, P.M., Smith, M. (1977a) Nucleotide sequence of bacteriophage ΦX174 DNA, *Nature* **265**, 687–695.

Sanger, F., Nicklen, S., Coulson, A.R. (1977b) DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467.

Venter, J.C., et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Yu, J., et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*), *Science* **296**, 79–92.

Glycobiology

Akira Kobata
Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan

1 Introduction 539

2 Principles of Studying the Sugar Chains of Glycoconjugates 540
2.1 Release of Sugar Chains of Glycoconjugates as Oligosaccharides 540
2.2 Fractionation of Oligosaccharides 544
2.3 Structural Analysis of Oligosaccharides 545

3 Rules Detected in the Sugar Chain Structures of Glycoconjugates 548

4 Functional Roles of the Sugar Chains of Glycoconjugates 551
4.1 Glycoprotein Hormones 553
4.2 Immunological Systems 554
4.3 Fertilization 556
4.4 Recombinant Glycoproteins 556
4.5 Unusual Sugar Chains 557
4.6 Blood Coagulation 559
4.7 Cartilage Formation 560
4.8 Sugar Chains of Glycolipids 560

5 Cloning of the Glycosyltransferases and Other Enzymes Related to the Formation of Sugar Chains of Glycoconjugates 561

Bibliography 563
Books and Reviews 563
Primary Literature 563

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

Keywords

Anomers
The two isomeric structures, formed because the aldose C-1 atom becomes asymmetric in making the sugar chains, are called the α- and the β-anomers.

Fc Receptor
A cell-surface receptor that binds the hinge region of immunoglobulin G.

Furanose
The isomeric ring structure formed by four carbons and one oxygen atom of aldoses and ketoses.

Glycoconjugates
The generic name of carbohydrates that are linked to various biomolecules. They can be classified into glycoproteins, glycolipids, and proteoglycans.

Glycohormones
Peptide hormones containing sugar chains.

Glycon and Aglycon
Glycosidases, which cleave the sugar chains, have two kinds of specificity: one for glycon and one for aglycon. Glycon specificity is directed to the sugar moiety released from the portion of a sugar chain containing a reducing terminal, and its anomeric configuration. The aglycon specificity is directed to the structure of the portion to which the released sugars are linked. This specificity varies by enzymes. For example, diplococcal β-galactosidase cleaves the Galβ1$\rightarrow$4GlcNAc linkage but not the Galβ1$\rightarrow$3GlcNAc and the Galβ1$\rightarrow$6GlcNAc linkages. In contrast, the enzyme from *Streptococcus* 6646K hydrolyzes all three linkages.

Lectin
The general name for proteins other than antibodies that bind specifically to particular sugar chain structures.

Oligosaccharide Alcohols
Oligosaccharides, the reducing termini of which are converted to alditols.

Protein C
A serine protease zymogen; becomes a potent inhibitor of blood coagulation upon activation by thrombin.

Pyranose
The isomeric ring structure formed by five carbons and one oxygen atom of aldoses and ketoses.

Many proteins produced by multicellular organisms contain covalently linked sugar chains and are called *glycoproteins*. Because of the difficulties associated with the structural study of the sugar chains of glycoproteins, however, the functional aspects of the sugar moieties of glycoproteins were ignored during the long history of protein research.

In addition to the structural multiplicity, another factor makes the structural study of the sugar chains of glycoproteins difficult. Many glycoproteins contain more than one sugar chain in one molecule. Even in the case of a glycoprotein with only one sugar chain, there is widespread microheterogeneity of sugar chain structure because the absence of a template in the biosynthetic machinery of sugar chains makes possible the formation of incomplete chains. Therefore, each sugar chain must be separated before it can be subjected to structural study.

Interest in glycoconjugate research was stimulated early in the 1960s by elucidation of the antigenic determinants of human blood types and the molecular basis of antigenic conversion of bacteria. This research area has further attracted the interest of biologists, because many studies on cell biology have suggested the possibility that the sugar chains of glycoproteins and glycolipids play an important role as signals of cell-to-cell recognition, which are crucial in multicellular organisms.

Development of gene technology in recent years has also accelerated the functional study of glycoprotein sugar chains. This technology has opened a way to obtaining substantial amounts of bioactive proteins, which are useful but occur in small amounts in the animal body. However, many proteins produced by animal cells occur as glycoproteins. Since many nonglycosylated recombinant proteins, obtained by using bacterial hosts, do not express the expected biological activities, importance of the study of their sugar moieties has been noticed. "Glycobiology" is a scientific field established on the basis of the possibility of elucidating the biological information in the sugar chains of glycoconjugates and using these data to increase our knowledge of biology.

1 Introduction

In contrast to nucleic acids and proteins, sugar chains of many different structures can be formed by using a small number of monosaccharide units. Let us consider the smallest unit of chains: A–B. In the case of nucleic acids, only one structure is made by assigning adenylic acid to A and guanylic acid to B. In the case of protein also, only one structure is made when A and B are assigned to valine and serine respectively. In the case of sugar chains, however, many isomeric structures are formed. Suppose that A and B are assigned to galactose and mannose respectively. As shown in Fig. 1, galactose can be linked at the four hydroxyl groups of mannose: C-2, C-3, C-4, and C-6. Accordingly, four isomeric structures can be formed. By virtue of the ability of the galactose residue to take two anomeric configurations, α (solid line in Fig. 1) and β (dotted line in Fig. 1), the number of possible isomeric structures becomes eight. Furthermore, galactose residues can occur

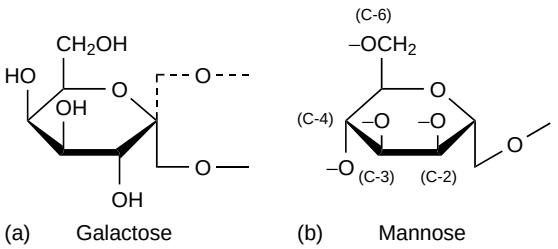

Fig. 1 Construction of sugar chains.

in the furanose form as well as in the pyranose form shown in Fig. 1. Thus, 16 isomeric structures are possible for the disaccharide Gal→Man. When the number of units increases (to three, four, etc.), only one structure can be formed by assigning a particular unit at each position in the case of nucleic acids and proteins, because they are linear constructs. In contrast, the number of isomeric sugar chains increases by geometrical progression, because branching can be formed in sugar chains larger than a disaccharide. This means that sugar chains, but not nucleic acids and proteins, have the following characteristic feature: they can form multiple structures with a small number of units.

2 Principles of Studying the Sugar Chains of Glycoconjugates

2.1 Release of Sugar Chains of Glycoconjugates as Oligosaccharides

As already described, each sugar chain of a glycoconjugate must be isolated for structural study. The sugar chains of glycoproteins can be classified into two groups, *O*-linked and *N*-linked. An *O*-linked or mucin-type sugar chain contains at its reducing terminus an *N*-acetylgalactosamine residue that is linked to the hydroxyl group of either a serine or a threonine residue in a polypeptide chain. Such sugar chains can be released as oligosaccharide alcohols by heating the glycoprotein at 48 °C in a solution of 1 M $NaBH_4$ and 0.05 N NaOH for 16 h. The use of hydrazinolysis to remove acyl groups from the *N*-acetylaminosugar residues of oligosaccharide alcohols, followed by re-*N*-acetylation with ^{14}C- or ^{3}H-labeled acetic anhydride, will produce isotope-labeled oligosaccharide alcohols, which are useful for fractionation and structural analyses.

The other group of sugar chains of glycoproteins, the *N*-linked or asparagine-linked sugar chains, are linked to the amide group of an asparagine polypeptide residue; each one contains an *N*-acetylglucosamine residue at its reducing terminus. These sugar chains are quantitatively released as oligosaccharides by heating thoroughly dried glycoproteins in anhydrous hydrazine at 80 to 100 °C for 8 to 10 h. Although some of the acetyl groups of the *N*-acetylaminosugars of oligosaccharides are removed by hydrazinolysis, they can be recovered by *N*-acetylation of the oligosaccharide fraction. The oligosaccharides can be converted quantitatively to ^{3}H-labeled oligosaccharide alcohols by reduction with $[^{3}H]NaBH_4$. Although the mechanism is not known, it was reported that *O*-linked sugar chains are cleaved by milder hydrazinolysis (60 °C for 5 h). This treatment can be effectively used for converting *O*-linked sugar chains to oligosaccharides with reducing terminal.

N-Glycanases (glycopeptidases) from several biological sources are commercially available. These enzymes hydrolytically cleave the GlcNAc→Asn linkages of glycoproteins and leave aspartic acid residues on the polypeptide portion. The glycosylamine residues of the released oligosaccharides are spontaneously hydrolyzed into oligosaccharides and ammonia. With this catalytic action, the enzymes release all *N*-linked sugar chains of glycoproteins without destroying the polypeptide moieties. However, one must take care for the limits of application of these enzymes, because they can have substrate specificities including the steric hindrance of the polypeptide portion.

Several endo-β-*N*-acetylglucosaminidases (hereafter referred to as *endos*), which act on *N*-linked sugar chains, were reported. These enzymes hydrolytically cleave the *N*,*N*′-diacetylchitobiose moieties located at the reducing termini of *N*-linked sugar chains. Accordingly, the enzymes release most of the sugar chains of glycoproteins intact by leaving either the *N*-acetylglucosamine residue or the Fucα1→6GlcNAc group on the polypeptides. Some of these enzymes are now commercially available and each one has a unique substrate specificity (Table 1). Endo D requires Manα1→3Manβ1→4GlcNAc for its glycon specificity. The α-mannosyl residue of the trisaccharide group is the most important and should not be linked by other sugars except at its C-4 position. Endo H requires for its specific glycon a tetrasaccharide structure: Manα1→3Manα1→6Manβ1→4GlcNAc. Again, the most important sugar residue is the α-mannosyl residue at the nonreducing terminal. Addition of sugars at the hydroxyl group except for the C-2 position of this residue will make the oligosaccharide resistant to the enzyme action. Endo C_{II} requires for its specific glycon a branched pentasaccharide: Manα1→3Manα1→6(Manα1→3)Manβ1→4GlcNAc. The two terminal α-mannosyl residues must be either nonsubstituted or replaced by other sugars only at their C-2 positions. These enzymes have rather broad aglycon specificities, since they cleave the sugar chains linked to *N*-acetylglucosamine, the Fucα1→6GlcNAc, the GlcNAc→Asn (peptides), and the Fucα1→6GlcNAc→Asn (peptides) groups. The enzymes except for endo H can also cleave the sugar chains linked to *N*-acetylglucosaminitol and the Fucα1→ 6*N*-acetylglucosaminitol groups, but the rate of hydrolysis is much lower than others. Recently, another endo-β-*N*-acetylglucosaminidase called *endo M*, which can cleave most of the *N*-linked sugar chains, were found in the culture medium of *Mucor hiemalis*. This enzyme may be useful for the future development of glycotechnology because it also has transglycosylation activity. An interesting evidence is that the enzyme can transglycosylate *N*-linked sugar chains to the GlcNAcβ1-Gln (peptides) groups as well as the GlcNAcβ1-Asn (peptides) groups.

Several other endoglycosidases shown in Table 1 were also found and have been used for the study of sugar chains. Endo-α-*N*-acetylgalactosaminidase hydrolytically cleaves the GalNAcα1→Ser (or Thr) linkage of the Galβ1→3GalNAcα1→Ser (or Thr) peptides and is commercially available by the name of *O*-glycanase. Endo-β-galactosidases with different substrate specificities have been obtained from various bacteria. The endo-β-galactosidase from *Streptococcus pneumoniae* releases the GalNAcα1→3(Fucα1→2)Gal and the Galα1→3(Fucα1→2)Gal groups from

Tab. 1 Structural requirements of endoglycosidases.

Enzyme Name	***Substrate Structure***
Endo-β-*N*-acetylglucosaminidase	
Endo D	R | 6 R-4Manα1-3Manβ1-4GlcNAcβ1 ¦ 4GlcNAc-Asn R | 6
Endo H	R R | | 2 6 Manα1-3Manα1 \ 6 R-4Manβ1-4GlcNAcβ1 ¦ 4GlcNAc-Asn 3 R R | 6
Endo C_{II}	R R | | 2 6 Manα1-3Manα1 \ 6 R-4Manβ1-4GlcNAcβ1 ¦ 4GlcNAc-Asn 3 R-2Manα1
Endo-α-*N*-acetylgalactosaminidase	Galβ1-3GalNAcα1 ¦ Ser(Thr)
Endo-β-galactosidase	
S. pneumoniae	GalNAcα1-3Galβ1 ¦ 4GlcNAcβ1- ___ (Gal) 2 (Glc) \| Fucα1
E. freundii *B. Fragilis*	GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1- ¦ 4GlcNAcβ1-___ 3 \| Fucα1
C. perfringens	Galα1-3Galβ1 ¦ 4GlcNAcβ1-___
Endosialidase	Neu5Acα2-8Neu5Acα2-8Neu5Acα2 ¦ 8Neu5Acα2- ¦ 8Neu5Acα2-___ (Gc) (Gc) (Gc) (Gc) (Gc)

Vertical dashed lines, positions of hydrolysis; R, either hydrogen or sugars.

Type II (see Sect. 5) blood group A and B active sugar chains respectively. The enzyme has been successfully used to characterize the blood group substances. Endo-β-galactosidases from *Escherichia freundii* and from *Bacteroides fragilis* hydrolytically cleave the β-galactosidic linkages in the $(\text{Gal}\beta1\rightarrow4\text{GlcNAc}\beta1\rightarrow3)_n$

groups (*N*-acetyllactosaminoglycan) included in the sugar chains of glycoconjugates. Although the enzyme from *E. freundii* cleaves most of the β-galactosyl linkages in both linear and branched *N*-acetyllactosaminoglycan, the enzyme from *B. fragilis* cannot hydrolyze the β-galactosyl linkage of the $\rightarrow$GlcNAcβ1$\rightarrow$6($\rightarrow$GlcNAcβ1$\rightarrow$3)Galβ1$\rightarrow$4GlcNAc group. Another endo-β-galactosidase, purified from the culture medium of *Clostridium perfringens*, releases Galα1$\rightarrow$3Gal from the Galα1$\rightarrow$3Galβ1$\rightarrow$4GlcNAc group located at the nonreducing termini of the sugar chains of glycoconjugates. The glycon specificity of the enzyme is very strict and blood group B determinant cannot be hydrolyzed.

Endosialidase is purified from a bacteriophage K1. It hydrolyzes both *N*-acetylneuraminic acid polymers and *N*-glycolylneuraminic acid polymers in which sialic acid residues are linked by an α2$\rightarrow$8 linkage. The enzyme cannot hydrolyze oligomers smaller than tetrasaccharide and *O*-acetylated *N*-glycolylneuraminic acid polymers.

With the specificities shown in Table 1, endoglycosidases were quite useful for the structural study of sugar chains. For example, endo H cleaves all high mannose–type sugar chains but not complex-type sugar chains, which will be described later. Accordingly, this enzyme is effective in discriminating between the two subgroups of *N*-linked sugar chains. Endo D is effective in discriminating between two isomeric sugar chains:

$$\begin{aligned} &\mathrm{R} \rightarrow \mathrm{GlcNAc}\beta1 \rightarrow 2\mathrm{Man}\alpha1 \\ &\quad \rightarrow 6(\mathrm{Man}\alpha1 \rightarrow 3)\mathrm{Man}\beta1 \\ &\quad \rightarrow 4\mathrm{GlcNAc}\beta1 \rightarrow 4\mathrm{GlcNAc} \end{aligned}$$

and

$$\begin{aligned} &\mathrm{Man}\alpha1 \rightarrow 6(\mathrm{R} \rightarrow \mathrm{GlcNAc}\beta1 \\ &\quad \rightarrow 2\mathrm{Man}\alpha1 \rightarrow 3)\mathrm{Man}\beta1 \\ &\quad \rightarrow 4\mathrm{GlcNAc}\beta1 \rightarrow 4\mathrm{GlcNAc} \end{aligned}$$

in which R represents sugars, because the former is cleaved by the enzyme but the latter is not.

Glycosphingolipids are the major glycolipids found in animal tissues. Since all glycosphingolipids contain just one sugar chain in one molecule, isolation of a single glycosphingolipid is enough for the structural study of that glycolipid's sugar chain. A glycosphingolipid with a single sugar chain is still a mixture of different molecules, however, as a result of the structural heterogeneity characterizing the ceramide portion of the molecule. To alleviate this complicated situation, glycoceramidases are used to release the sugar chains of glycosphingolipids intact as oligosaccharides. The oligosaccharides can then be labeled as described for the oligosaccharides released from glycoproteins and fractionated by using appropriate methods (see Sect. 2.2).

As described later, the sugar chains of proteoglycans are unique and are called *glycosaminoglycans*. Many glycosaminoglycan chains in proteoglycans are linked to the serine residues, constructing the Ser-Gly-X-Gly sequence of core proteins through the Galβ1$\rightarrow$3Galβ1$\rightarrow$4Xyl group located at their reducing termini. Accordingly, these sugar chains can be almost quantitatively released as polysaccharides by alkaline borohydride treatment as described for *O*-linked sugar chains of glycoproteins. Structures of the released polysaccharides are then analyzed by using appropriate endoglycosidases, which cleave particular glycosidic linkages in sugar chains. Among them, several eliminases listed in

Tab. 2 Bacterial eliminases.

Enzymes	*Source*	*Main substrates*[a]
Hyaluronidases		
Hyaluronidase	*Streptomyces hyalurolyticus*	HA
Hyaluronidase SD	*Streptococcus dysgalactiae*	HA
Chondroitinases		
Chondroitinase ABC	*Proteus vulgaris*	CS, DS, HA
Chondroitinase AC	*Flavobacterium heparinum*	CS, HA
Chondroitinase ACII	*Arthrobacter aurescens*	HA, CS
Chondroitinase B	*Flavobacterium heparinum*	DS
Chondroitinase C	*Flavobacterium sp.* Hp 102	CS
Heparitinases		
Heparitinase I	*Flavobacterium heparinum*	HS
Heparitinase II	*Flavobacterium heparinum*	HS, Hep
Heparinase	*Flavobacterium heparinum*	Hep

[a]HA, hyaluronic acid; CS, chondroitin sulfate; DS, dermatan sulfate; HS, heparan sulfate; Hep, heparin.

Table 2 have been most effectively used to analyze the detailed structures of proteoglycans. As shown in Fig. 2, all these enzymes cleave the amino sugar–uronic acid linkages of glycosaminoglycan and produce various disaccharides, summarized in Table 3. The enzymes in Table 2 differ in substrate specificity. For example, chondroitinase ABC cleaves both the sulfated GalNAcβ1$\rightarrow$3GlcA and the sulfated GalNAcβ1$\rightarrow$3IdoA linkages, while chondroitinase AC cannot cleave the sulfated GalNAcβ1$\rightarrow$3IdoA linkage. On the contrary, chondroitinase B cleaves only the sulfated GalNAcβ1$\rightarrow$3IdoA linkage. Chondroitinase C cleaves the $SO_4 \rightarrow$6GalNAc1$\rightarrow$3GlcA linkage but not the $SO_4 \rightarrow$4GalNAcβ1$\rightarrow$3GlcA and $SO_4 \rightarrow$6($SO_4 \rightarrow$ 4)GalNAc β 1 $\rightarrow$ 3GlcA linkages. Since the disaccharides in Table 3 can be mutually separated by high-performance liquid chromatography (HPLC) and quantitatively analyzed by using approximately 0.5 nmol of sample, molecular constructions of chondroitin sulfates, hyaluronic acids, heparins, and heparan sulfates can be analyzed in detail, including their structural heterogeneities.

Recently, endo-β-xylosidase and endo-β-galactosidase, which cleave the sugar–polypeptide linkage region of proteoglycans, were isolated from the midgut gland of a mollusk, *Patinopecten*. Since the enzymes cleave all glycosaminoglycans containing the Galβ1$\rightarrow$3Galβ1$\rightarrow$4Xyl group at their reducing termini, and the released polysaccharides contain reducing termini that can be tagged by ^{3}H or fluorescent groups, they will become useful tools for the study of proteoglycans.

2.2 Fractionation of Oligosaccharides

Acidic oligosaccharides containing sialic acids, phosphate, or sulfate can be fractionated by either paper electrophoresis or ion-exchange column chromatography. Gel permeation chromatography

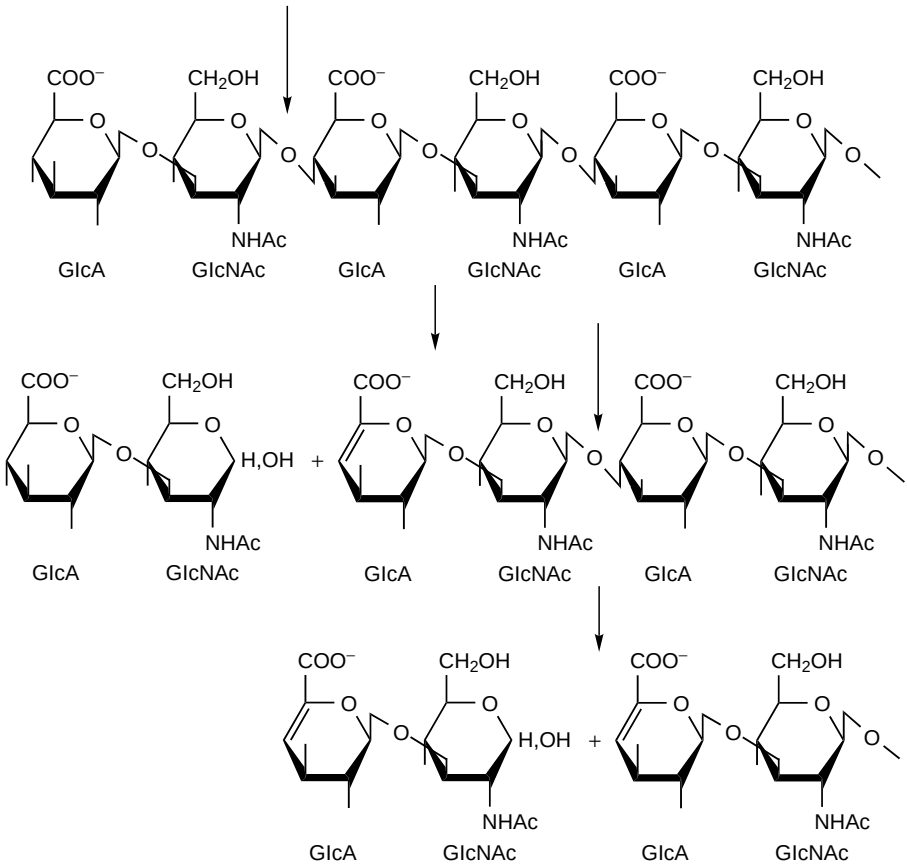

Fig. 2 Hydrolysis of hyaluronic acid by streptococcal hyaluronidase.

supplemented with ultrafine Bio-Gel P-4 and HPLC were also widely used. In addition, affinity chromatography with use of immobilized lectin columns offers a unique and effective method of fractionating oligosaccharides. Table 4 summarizes the binding specificities of immobilized lectin columns. Since the behavior of an oligosaccharide is mainly determined by the structural requirement of each immobilized lectin column and is not affected by additional sugar chain moieties, a mixture of many oligosaccharides can be separated into its components by means of an appropriately chosen series of lectin columns.

Two-dimensional mapping of known oligosaccharides, released from *N*-linked sugar chains by either glycopeptidase or hydrazinolysis, was reported. This method will be useful for obtaining an oligosaccharide pattern of *N*-linked sugar chains of a glycoprotein.

2.3 Structural Analysis of Oligosaccharides

The monosaccharide sequence of an oligosaccharide can be determined by sequential exoglycosidase digestion. Since an exoglycosidase releases a particular

Tab. 3 Unsaturated disaccharides released from glycosaminoglycans by digestion with eliminases.

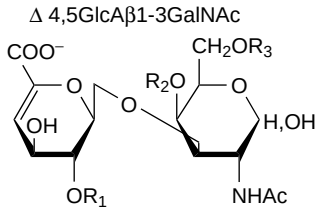

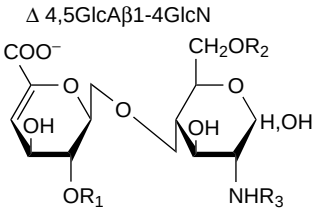

	R_1	R_2	R_3		R_1	R_2	R_3
ΔDi-OS	H	H	H	ΔDiHS-OS	H	H	Ac
ΔDi-UA2S	SO_3^-	H	H	ΔDiHS-UA2S	SO_3^-	H	Ac
ΔDi-4S	H	SO_3^-	H	ΔDiHS-NS	H	H	SO_3^-
ΔDi-6S	H	H	SO_3^-	ΔDiHS-6S	H	SO_3^-	Ac
ΔDi-diS_B	SO_3^-	SO_3^-	H	ΔDiHS-diS_1	H	SO_3^-	SO_3^-
ΔDi-diS_D	SO_3^-	H	SO_3^-	ΔDiHS-diS_2	SO_3^-	H	SO_3^-
ΔDi-diS_E	SO_3^-	SO_3^-	H	ΔDiHS-diS_3	SO_3^-	SO_3^-	Ac
ΔDi-triS	SO_3^-	SO_3^-	SO_3^-	ΔDiHS-triS	SO_3^-	SO_3^-	SO_3^-

monosaccharide from the nonreducing terminal of an oligosaccharide, the radioactivity or fluorescent groups incorporated at the reducing terminal of an oligosaccharide remains in the oligosaccharide portion until it has been completely digested to the tritium-labeled monosaccharide alcohol or fluorescent-labeled monosaccharide located at the reducing end. Because many exoglycosidases have narrow aglycon specificities, not only the sequence but also the linkage position of each monosaccharide can be determined by using such enzymes. For example, β-galactosidase purified from the culture medium of *Streptococcus pneumoniae* can cleave the Galβ1→4GlcNAc linkage but not the Galβ1→3GlcNAc and the Galβ1→6GlcNAc linkages. Therefore, the enzyme is an effective reagent for the determination of the Galβ1→4GlcNAc structure, which is widely distributed in many glycoconjugates. α-L-Fucosidase from *Bacillus fulminans* cleaves only the Fucα1-2*N*-acetyllactosamine group. In contrast, the enzyme from *Almond emulsin* cleaves Le^x and Le^a determinants, but cannot cleave other α-fucosyl linkages. The enzyme from the hepatopancreas of marine gastropods can cleave all α-fucosyl linkages. Accordingly, one can determine the position at which the α-fucosyl residue is linked by making proper use of these α-L-fucosidases.

Two α-mannosidases with different aglycon specificities were found in the mycellium of *Aspergillus saitoi*. α-Mannosidase I cleaves the Manα1→2Man linkage only. Accordingly, the enzyme is an effective reagent for identifying a series of high mannose–type sugar chains (see Sect. 3), which are all converted to Manα1→6(Manα1→3)Manα1→6(Manα1→3)Manβ1→4GlcNAcβ1→4GlcNAc by incubation with the enzyme. Another enzyme of *A. saitoi*, α-mannosidase II, is also useful because it releases one mannose residue from the R→Manα1→6(Manα1→3) Manβ1→ group but not from the Manα1→6(R→Manα1→3)Manβ1→ group. Therefore, the enzyme is an effective reagent when it is desired to assign

Tab. 4 Binding specificities of immobilized lectin columns.

Lectin	***Structure Necessary for Binding***
Concanavalin A	R-2Manα1 \ 6 Manβ1-R R-2Manα1 / 3 R-2Manα1-2Manα1-R
Phytohemagglutinin E_4	
Strong binding (retarded at 20°C and 2°C)	±Fucα1 \| 6 R-3Galβ1-4GlcNAcβ1-2Manα1 \ 6 GlcNAcβ1-4Manβ1-4GlcNAcβ1-4GlcNAc R \ 4 Manα1 / 3 R-4GlcNAcβ1 / 2
Weak binding (retarded at 2°C only)	±Fucα1 \| 6 R-3Galβ1-4GlcNAcβ1-2Manα1 \ 6 3 Manβ1-4GlcNAcβ1-4GlcNAc R /
Datura stramonium agglutinin	
Strong binding (bound)	R-3Galβ1-4GlcNAcβ1 \ 6 Manα1-R R-3Galβ1-4GlcNAcβ1 / 2 R-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-R
Weak binding (retarded)	R-3Galβ1-4GlcNAcβ1 \ 4 Manα1-R R-3Galβ1-4GlcNAcβ1 / 2
Aleuria aurantia lectin	Fucα1 | 6 R-4GlcNAc
Allomyrina dichotoma lectin	Neu5Acα2-6Galβ1-4GlcNAcβ1-R

R represents either hydrogen or sugar chains.

positions to the outer chains in a complex-type sugar chain (see Sect. 3).

In addition to the analytical techniques described so far, methylation analysis is essential for the complete structural determination of oligosaccharides. Smith degradation, which analyzes the periodate oxidation products of oligosaccharides, is also useful for the study of many oligosaccharides. Physical methods such

as nuclear magnetic resonance and fast atom bombardment mass spectrometry are also becoming useful techniques for the structural analysis of sugar chains.

3
Rules Detected in the Sugar Chain Structures of Glycoconjugates

Since the *N*-linked sugar chains found in glycoproteins usually contain more than 10 monosaccharides, the structural multiplicity represented is theoretically enormous. Indeed, it might be impossible to elucidate the biological information in sugar chains if we had to handle such a large number of isomers. Fortunately, studies of the sugar chain structures of various glycoproteins have revealed that a series of structural rules exists in them, and variable regions are limited to parts of their structures.

The *N*-linked sugar chains of glycoproteins contain as a common core the pentasaccharide Manα1→6(Manα1→3) Manβ1→4GlcNAcβ1→4GlcNAc, which we shall call the "trimannosyl core". On the basis of the structures and locations of the sugar residues added to the trimannosyl core, *N*-linked sugar chains are classified into three subgroups (Fig. 3). Complex-type sugar chains contain no mannose residues other than those in the trimannosyl core. Outer chains with an *N*-acetylglucosamine residue at their reducing termini are linked to the two α-mannosyl residues of the trimannosyl

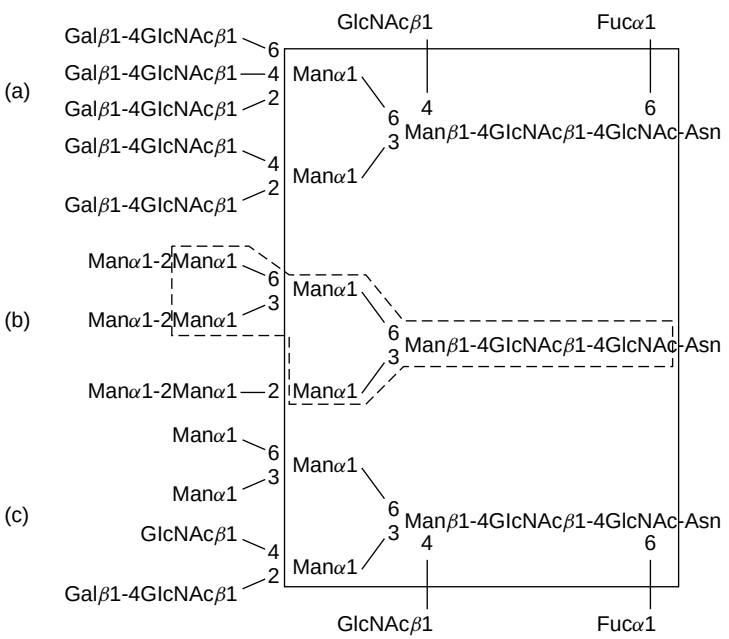

Fig. 3 Three subgroups of *N*-linked sugar chains: (a) complex-type; (b) high mannose–type; (c) hybrid-type. Within solid lines are the trimannosyl core structures common to all *N*-linked sugar chains. The structure enclosed by dashed lines is the common heptasaccharide core of high mannose–type sugar chains; for structures outside, variation in sugar chains is possible.

core. High mannose–type sugar chains contain only α-mannosyl residues in addition to the trimannosyl core. Manα1→6(Manα1→3)Manα1→6(Manα1→3)Manβ1→4GlcNAcβ1→4GlcNAc is commonly included in this type of sugar chain (enclosed by a dashed line in Fig. 3). Hybrid-type sugar chains were so named because they have the characteristic features of both complex- and high mannose–type sugar chains. One or two α-mannosyl residues are linked to the Manα1→6 arm of the trimannosyl core, as in the case of the high mannose–type, and the outer chains found in complex-type sugar chains are linked to the Manα1→3 arm of the core. Presence or absence of the α-fucosyl residue linked to the C-6 position of the proximal *N*-acetylglucosamine residue and the β-*N*-acetylglucosamine residue linked to the C-4 position of the β-mannosyl residue of the trimannosyl core (bisecting GlcNAc) contributes to the structural variation of complex-type and hybrid-type sugar chains.

Among the three subgroups of *N*-linked sugar chains, the largest structural variation resides in the complex-type sugar chains. This variation is formed by two structural factors. As shown in Fig. 4(a), mono-, bi-, tri-, tetra-, and pentaantennary sugar chains are formed by adding one to five *N*-acetylglucosamine residues to the trimannosyl core. Two isomeric triantennary sugar chains, containing either the GlcNAcβ1→4(GlcNAcβ1→2)Manα1→3 group or the GlcNAcβ1→6(GlcNAcβ1→2)Manα1→6 group, are found. These isomeric sugar chains are called 2,4-branched and 2,6-branched triantennary sugar chains respectively. On these *N*-acetylglucosamine residues, various outer chains are formed (Fig. 4b) by the concerted action of many glycosyltransferases and sulfotransferases. It is important to note that bisecting GlcNAc is never elongated by the action of glycosyltransferases. A combination of the antennary and the various outer chains will form a large number of different complex-type sugar chains. In most glycoproteins, the α-fucosyl residue linked at the C-6 position of the proximal *N*-acetylglucosamine residue of the trimannosyl core is not elongated by other sugars. However, it was found that the fucosyl residue of the hybrid-type sugar chain of octopus rhodopsin is substituted by a β-galactosyl residue at its C-4 position.

In contrast to *N*-linked sugar chains, *O*-linked sugar chains have fewer structural rules. At present, the latter chains are grouped into the six classes by their core structures, as shown in Fig. 5. *O*-Linked sugar chains with the cores 5 and 6 are found in a limited number of glycoproteins.

More than 200 glycosphingolipids, which are different in their sugar moieties, have been isolated from mammalian tissues. Most of these glycosphingolipids fall into five groups by virtue of their tetrasaccharide core structures (Fig. 6). In addition to these major groups, there are minor groups containing trisaccharides, disaccharides, and galactose as common cores. These glycosphingolipids are formed by a complicated biosynthetic network. Figure 7 summarizes the biosynthetic pathway of gangliosides, the generic name for glycosphingolipids containing a sialic acid residue. Although the structures of the gangliosides listed in Fig. 7 are well elucidated, the biosynthetic pathways leading to some of them have not been established.

Although proteoglycans should also be considered to be glycoproteins inasmuch as sugar chains are linked to protein cores, their sugar chains, called glycosaminoglycans, are much longer

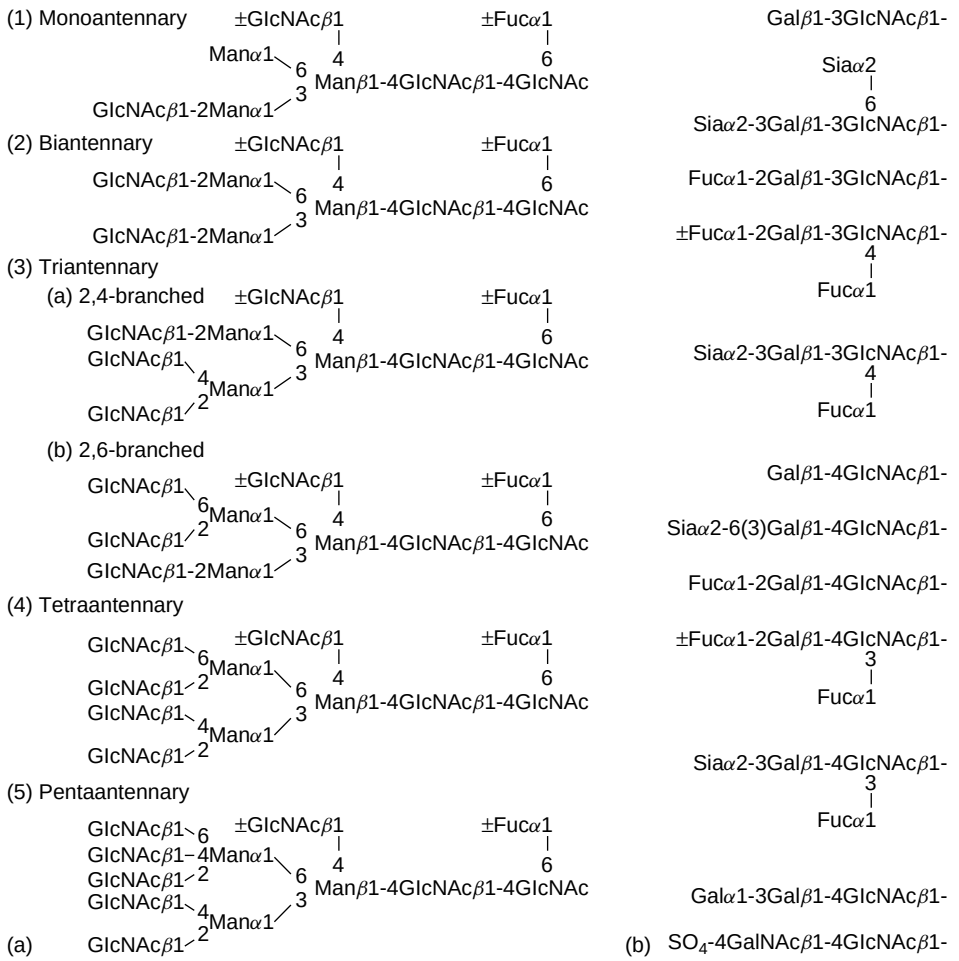

Fig. 4 Two major elements in the formation of the various structures of complex-type sugar chains: (a) branching of complex-type sugar chains and (b) various outer-chain structures found in complex-type sugar chains.

(100–200 monosaccharide residues) than regular *N*- and *O*-linked sugar chains and contain many anionic residues, such as uronic acid and sulfate. Furthermore, their structures are unique in that their basic construction is disaccharide repeats. On the basis of the disaccharide structures, glycosaminoglycans have long been classified into six groups: hyaluronic acids, chondroitin 4-sulfates, chondroitin 6-sulfates, dermatan sulfates, heparan sulfates, and keratan sulfates. However, detailed study of the structures of glycosaminoglycans has revealed that many microheterogeneities occur in the sugar chains, and it is rather hard to discriminate chondroitin 4-sulfates, chondroitin 6-sulfates, and dermatan sulfates because of the occurrence of their hybrid structures. Accordingly, glycosaminoglycans are currently classified into four groups as shown in Fig. 8.

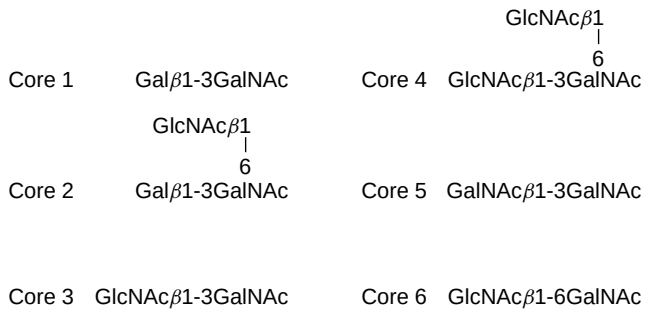

Fig. 5 Core structures found in *O*-linked sugar chains.

Fig. 6 Structures of major tetrasaccharide cores found in animal glycosphingolipids.

Gangliotetraose (ganglio-series)
Galβ1-3GalNAcβ1-4Galβ1-4Glc

Globotetraose (globo-series)
GalNAcβ1-3Galα1-4Galβ1-4Glc

Isoglobotetraose (isoglobo-series)
GalNAcβ1-3Galα1-3Galβ1-4Glc

Lacto-*N*-tetraose (lacto-series)
Galβ1-3GlcNAcβ1-3Galβ1-4Glc

Lacto-*N*-neotetraose (neolacto-series)
Galβ1-4GlcNAcβ1-3Galβ1-4Glc

4 Functional Roles of the Sugar Chains of Glycoconjugates

Accumulation of information about the structural characteristics of glycoconjugate sugar chains has enabled us to consider their functional roles in molecular terms. Glycoconjugates are distributed widely inside and outside the cells of multicellular organisms. Furthermore, the molecular construction of a single glycoconjugate group is quite variable. Figure 9 presents the molecular structures of three typical glycoproteins. Two different types of glycoproteins are found as extracellular components. Many serum glycoproteins have a limited number of sugar chains in one molecule (as in Fig. 9a). Many epithelial cells lining the alimentary tract and the respiratory tract secrete viscous glycoproteins called *mucins*. These glycoproteins work to protect the surface of mucous membrane from invasion of various pathogenic bacteria and viruses. As shown in Fig. 9(c), mucins have many short sugar chains, which are distributed all over the polypeptide moiety or are in clusters. Most of the cell-surface receptors are plasma membrane-integrated glycoproteins. They are anchored in the phospholipid bilayer by their hydrophobic amino acid cluster portion, and the sugar chains are linked to the polypeptide portion extended outside the cells (Fig. 9b). Although the functional roles of the sugar chains of glycoproteins are not simple because of their structural multiplicity, it is possible to classify them

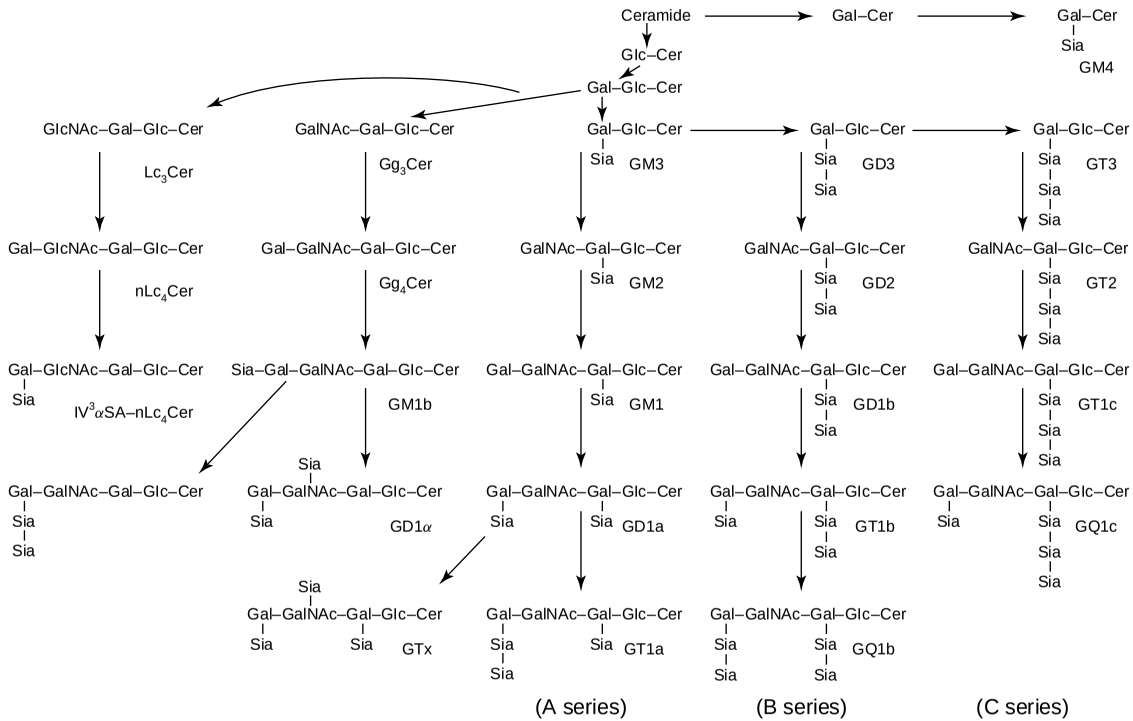

Fig. 7 Biosynthetic pathway of gangliosides.

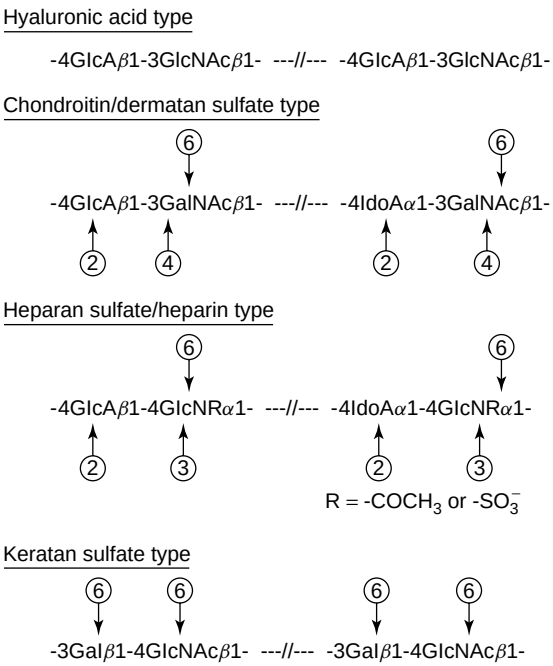

Fig. 8 Classification of glycosaminoglycans and their microheterogeneities. Except for hyaluronic acid, epimerization at the C-5 position of uronic acids and *N*- and *O*-sulfation (locations indicated by circled numbers) are the sources of microheterogeneities.

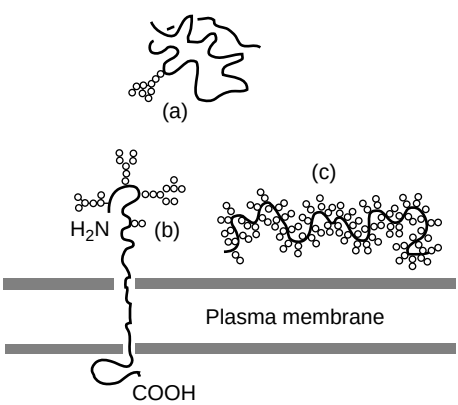

Fig. 9 Structural multiplicity of glycoproteins: (a) a serum glycoprotein, (b) a plasma membrane glycoprotein, and (c) a mucin. Open circles and black lines represent monosaccharides and polypeptide chains, respectively.

roughly into two groups: one that acts to confer physicochemical properties on proteins and another that acts as signals of various recognition phenomena.

Some of the representative functional roles of the sugar chains of glycoconjugates as recognition signals are introduced in the following parts of this section.

4.1 Glycoprotein Hormones

Four glycohormones have been known to occur in a variety of mammals. Three of them are produced in the same organ: lutropin and follitropin are produced by gonadotrophs, while thyrotropin is made

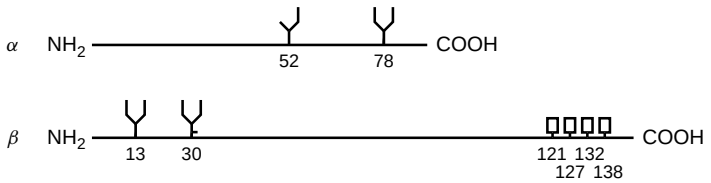

Fig. 10 Structure of hCG. Horizontal lines are polypeptide portions, trees are the *N*-linked oligosaccharides, and the boxes are the *O*-linked sugar chains.

by thyrotrophs in the anterior pituitary. Chorionic gonadotropin (Fig. 10), however, is produced by placental trophoblasts. All these hormones are composed of two noncovalently linked subunits of different sizes, designated α and β. Since the α-subunits of all four glycohormones have an identical amino acid sequence within an animal species, the specificity of each hormone to bind to its target cells had been considered to reside in its β-subunit. However, elucidation of the structures of *N*-linked sugar chains of the four glycohormones revealed that the α-subunits of these hormones are not in fact the same.

Furthermore, deglycosylated human chorionic gonadotropins (hCGs), obtained by enzymatic digestion or chemical means, were found to express no hormonal activity even though they bind to the target cells more strongly than their natural counterparts. Therefore, the *N*-linked sugar chains of glycohormones might play crucial roles in the biological functions of these substances. Actually, comparative study of the biological activities of hCGs, in which only one of the *N*-glycosylation sites was eliminated by recombinant DNA technology, revealed that *N*-glycosylation at the Asn52 of the α-subunit is essential for the expression of hormonal activity, while removal of the *N*-linked sugar chains at either Asn13 or Asn30 of the β-subunit or Asn78 of the α-subunit had no effect. Another important evidence shown by this line of study is that removal of Asn78 glycosylation of the α-subunit markedly reduced its assembly with the β-subunit. Glycosylation of the two *N*-glycosylation sites of the β-subunit is not essential for its assembly with the α-subunit, but elimination of Asn30 glycosylation inhibits the secretion of uncombined β-subunit. These results indicated that the *N*-linked sugar chains of hCG are important for constructing the correct conformation of each subunit. That the presence of at least one *N*-acetylglucosamine residue at each of the four *N*-glycosylation sites of hCG is enough to keep the correct folding of the two subunits was shown two decades ago by investigating the effects of digestion with exo- and endoglycosidases on the folding and assembly of the two subunits.

4.2 Immunological Systems

It is known from recent studies in immunology that both cellular and humoral immunological systems are controlled by a complicated network connecting the interactions of immunocompetent cells. In response to soluble immunological and inflammatory factors, leukocytes adhere to each other and to cells of other types (e.g. platelets, vascular endothelial cells, etc.). These interactions are characteristic in that they quickly form a strong and temporary contact between cells within a relatively short period compared to

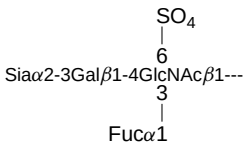

Fig. 11 Structure of the ligand of L-selectin.

the permanent cell adhesion observed in tissue-forming cells. Therefore, such interaction should be mediated by the special adhesive molecules that appear on the surface of activated cells.

Selectins were found by the studies to elucidate the mechanism that facilitates the recirculation of lymphoid cells from the intravascular compartment to lymph nodes. L-selectin, which is expressed on the surface of leucocytes, is considered to play an essential role in the homing of peripheral lymphocytes. The structure of its ligand was recently elucidated as sialyl-6-sulfo-Le^x (Fig. 11). P-selectin, which is expressed on the surfaces of activated platelets, megakaryocytes, and vascular endothelial cells, plays an important role in the early phases of leukocyte recruitment and adherence of activated platelets to neutrophils, monocytes, natural killer cells, and some subsets of T-lymphocytes. The ligands of P-selectin are rather complicated. They recognize both tyrosine sulfate residues of the polypeptide moiety and the sialyl Le^x residues of the sugar chains of the target cell-surface glycoproteins. A membrane-integrated glycoprotein that mediates the binding of monocytes and neutrophils was found on the surface of vascular endothelial cells. This glycoprotein called E-selectin is not constitutively expressed on the endothelial cells but is rapidly induced on the cells activated by the action of cytokines. Accordingly, the glycoprotein plays an important role in accumulating monocytes and neutrophils at the site of inflammation in an animal body. Adhesion of cells to E-selectin correlated with the presence on the cell surface of the Neu5Acα2→3Galβ1→3 or 4(Fucα1→4 or 3)GlcNAcβ1→ group.

Many intercellular mediators of glycoprotein nature were found to play roles in controlling the T-cells and a B-cell network. That the sugar chains work as ligands of these mediators and start their signal transductions was found. It was reported that a recombinant human interleukin-1β recognizes glycans of glycosylphosphatidylinositol anchor. That interleukin 2 recognizes high mannose–type sugar chains was also reported.

Immunoglobulin G (IgG) plays a major role in humoral immunity. It is unique among serum glycoproteins in that it contains sugar chains of extremely high microheterogeneity produced by the presence or absence of the two galactoses, the bisecting GlcNAc and the fucose residue underlined in the biantennary sugar chain in Fig. 12. In spite of this complex nature,

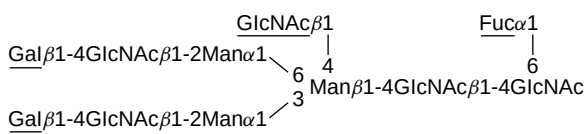

Fig. 12 The largest desialylated *N*-linked sugar chain of human IgG. Variation occurs in the presence or absence of the underlined sugar residues.

oligosaccharides of IgG samples purified from healthy human sera gave almost identical molar ratios. An interesting finding is that the IgG samples purified from sera of patients with rheumatoid arthritis are prominently lacking in galactose residues. This galactose deletion phenomenon in rheumatoid arthritis patients is limited to the sugar chains of IgG among serum glycoproteins and is induced by lower affinity of the β-galactosyltransferase in B-cells to UDP-galactose. The degalactosylated human IgG sample, obtained by *Streptococcus* 6646K β-galactosidase digestion, binds less effectively to the subcomponent C1q of the first component of the human complement and the Fc receptor. No decrease in binding to protein A was observed in the degalactosylated IgG. These results indicated that the function of IgG molecules can be modified by different degrees of maturation of their sugar moieties.

The data so far introduced indicated that glycobiology is indispensable for the future development of immunology.

4.3 Fertilization

Fertilization of mammals starts when a capacitized sperm meets a matured egg within an oviduct. Mammalian eggs are covered with a layer called *zona pellucida* (ZP), which is synthesized and secreted from the oocytes in developing follicles. It has been proposed that the ZP plays important roles in fertilization, including facilitation of species-specific recognition by sperm, prevention of polyspermy, and protection of fertilized eggs until the blastocyst stage.

In mice, the ZP is reported to be composed of three glycoproteins: ZP1, ZP2, and ZP3. It was reported that *O*-linked sugar chains, containing α-galactosyl residues at their nonreducing termini, of ZP3 work as sperm receptors. It was also reported that the β-galactosyltransferase on the sperm plasma membrane mediates sperm–egg binding by interacting with the sugar chains of ZP glycoproteins. The rationale of the latter hypothesis is partly proven by the finding in porcine ZP of the *N*-linked and *O*-linked sugar chains enriched in nonreducing terminal *N*-acetylglucosamine residues, which will interact with β-galactosyltransferase.

Occurrence of *N*-linked sugar chains with *N*-acetylglucosamines at their nonreducing termini was also found to occur in mouse testis. Targeted disruption of *Man2a2*, a gene encoding α-mannosidase IIx that is responsible for the processing of hybrid-type sugar chains to the Manα1$\rightarrow$6(GlcNAcβ1$\rightarrow$2Manα1$\rightarrow$3) Manβ1$\rightarrow$4GlcNAcβ1$\rightarrow$4GlcNAc group: the core portion of complex-type sugar chains, resulted in male infertility. The infertility is induced because spermatogenic cells fail to adhere to Sertoli cells and are prematurely released from the testis to the epididymis. By investigating the structures of glycopeptides that inhibited germ cell/Sertoli cell adhesion, it was found that the nonasaccharide of spermatogenic cells shown in Fig. 13 is responsible for the attachment of the two cell lines. The receptor on Sertoli cells, which specifically binds to the nonasaccharide, remains to be investigated.

4.4 Recombinant Glycoproteins

As already described, many recombinant glycoproteins have been obtained by using various animal cell lines as hosts.

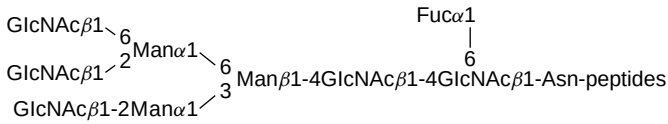

Fig. 13 Structure of the glycopeptide that inhibited adhesion of germ cells and Sertoli cells.

However, as evidenced through comparative study of the sugar chains of γ-glutamyltranspeptidases purified from the kidneys and livers of various mammals, both organ- and species-specific differences occur in the sugar chains of glycoproteins (Fig. 14). In addition, an altered glycosylation phenomenon is reflected in the sugar chains of recombinant glycoproteins because many of the cell lines used are somewhat malignant cells. Accordingly, the sugar chains of recombinant glycoproteins may display structural variations according to the type of cells used, even though the polypeptide structures are the same. Actually, comparative study of the *N*-linked sugar chains of natural human interferon-β1 (IFN-β1) and three recombinant IFN-β1s, produced by different cell lines transfected with the gene coding for human IFN-β1, revealed that their sugar patterns were different in spite of containing the same number of complex-type sugar chains. The differences occur both in the antennary structures and in the structures of outer-chain moieties. This phenomenon affords us a way to elucidate the function of the sugar chains by investigating the biological activities of several preparations of recombinant glycoproteins. Comparative study of the sugar patterns and the *in vivo* activities of several preparations of recombinant human erythropoietin revealed that the activity was proportional to the ratio of tetraantennary to biantennary oligosaccharides.

4.5 Unusual Sugar Chains

Sugar chains that link to the polypeptide moiety through Man-Ser and Thr linkages are found widely in yeast but are extremely rare in mammals. Investigation of the molecular mechanism of α-dystroglycan and laminin binding revealed that the sugar chains of α-dystroglycan shown in Fig. 15(a) are responsible for the binding. Therefore, finding of such unusual sugar chains as an important signal in the function of peripheral nerve dystroglycan may indicate that various unusual sugar chains occur as minor components in mammals and play important roles in particular tissues.

Notch signaling is an evolutionary conserved mechanism that is used by metazoans to control cell fates through local cell interactions. It is controlled by two ligand protein families called *Delta* and *Serrate/Jagged*. The binding affinities of Notch to the two ligands were further controlled by a protein called *Fringe*. It was found that Fringe is an *N*-acetylglucosaminyltransferase, which adds an *N*-acetylglucosamine residue to the α-fucosyl residues, linked to the serine residues of the epidermal growth factor (EGF) repeat of Notch. Addition of this *N*-acetylglucosamine residue will lead to further elongation of the sugar chains to the tetrasaccharide shown in Fig. 15(b). Fringe shows quite a high requirement of fucose residues and cannot add an

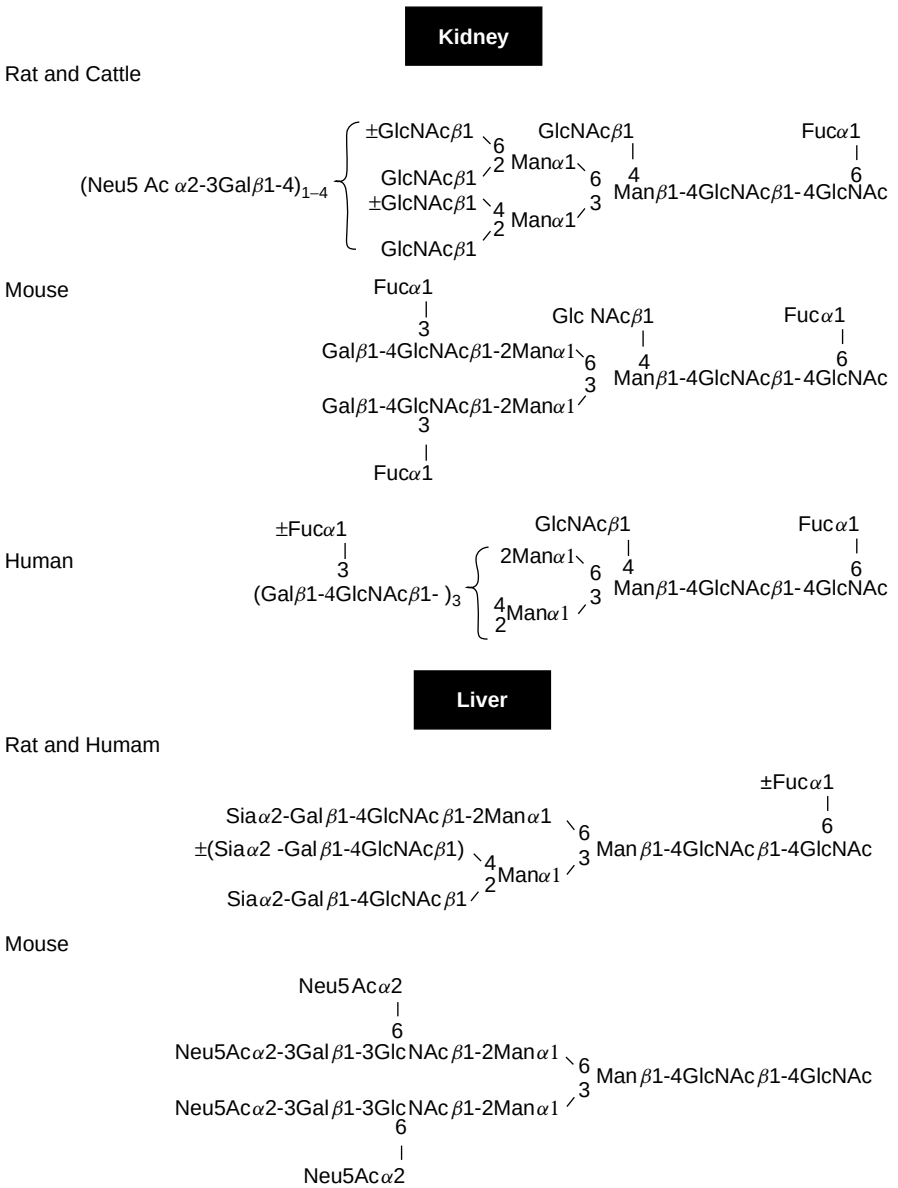

Fig. 14 Major sugar chain structures of γ-glutamyltranspeptidases purified from the kidneys and the livers of various mammalian species.

N-acetylglucosamine residue to any other sugar. It was recently found that the *N*-acetylglucosaminyltransferase, which adds the *N*-acetylglucosamine residue of the sugar chains of dystroglycan, acts only on the mannose residue linked to α-dysroglycan and cannot act on other α-mannosyl derivatives.

Fig. 15 Structures of the sugar chains of (a) α-dystroglycan and (b) Notch.

Neu5Acα2-3Galβ1-4GlcNAcβ1-2Manα1-Ser(or Thr)
(a)

Neu5Acα2-3Galβ1-4GlcNAcβ1-3Fucα1-Ser
(b)

These results indicated that the same trisaccharide residues are formed on different sugar chains by the action of various key isozymes and are used at different sites of differentiation of multicellular organisms. The trisaccharide residue is also distributed at the nonreducing termini of both *N*-linked and *O*-linked sugar chains. Therefore, structural elucidation of the receptors recognizing the trisaccharide residue is quite an interesting target for future study.

4.6 Blood Coagulation

Elucidation of the detailed structures of glycosaminoglycans is also opening a new age in proteoglycan study. Heparin was identified more than 70 years ago as an anti–blood coagulation material. A recent finding that a pentasaccharide structure in the heparin molecule, shown in Fig. 16, binds specifically to antithrombin led to the elucidation of the molecular mechanism of the anticoagulant activity of this glycosaminoglycan. With this finding as a turning point, many bioactive segments of glycosaminoglycans have been determined. Since blood clots are not formed on healthy blood vessel walls, some anticoagulation material may be included on the wall surfaces. On the basis of the observation that the introduction of heparinase solution into the bloodstream prominently decreases anticoagulation activity, occurrence of heparin-like sugar chains in the endoderm has been suspected. Actually, a heparan sulfate proteoglycan, which has a core protein different from that of heparin proteoglycan, is synthesized by the cultured endoderm obtained from bovine and rat artery. Another candidate anticoagulation factor is thrombomodulin found on the surface of endoderm. This protein binds to thrombin and suppresses its fibrin-clotting activity. A thrombomodulin–thrombin complex also enhances degradation of factors Va and VIIa by activating protein C. Accordingly, thrombomodulin protects the organism from thrombin overproduction. Upon chondroitinase ABC digestion, thrombomodulin was revealed to be a proteoglycan containing chondroitin sulfate. An interesting finding is that the activity of thrombin-dependent suppression of fibrin clotting is prominently decreased by

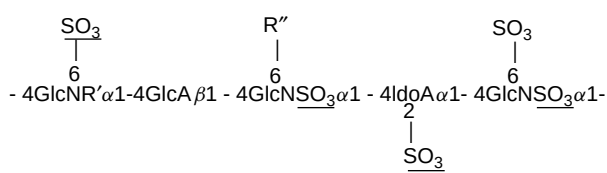

Fig. 16 Smallest unit of heparin necessary for binding to antithrombin III. R′ = either sulfate or acetyl residue. R″ = either hydrogen or sulfate residue. The four sulfate residues underlined in the figure are essential for the binding.

treating thrombomodulin with chondroitinase ABC, while thrombin-dependent protein C activation is not affected. This information will be effective in developing new oligosaccharide drugs in the near future.

4.7 Cartilage Formation

Cartilage is a tissue rich in extracellular matrix. Approximately 50% of the dry weight of cartilage is occupied by a proteoglycan called *aggrecan*. The structure of the core protein of aggrecan was elucidated by successful cloning of its cDNA. A lethal recessive mouse mutant, which is cartilage matrix–deficient (*cmd/cmd*), is an effective model that can be used in the investigation of the functional role of aggrecan. The skeleton of the embryo of this mutant is strikingly different from that of age-paired normal mouse embryo, and the limbs are drastically shortened and deformed. The mutant mouse was found to be genetically deficient in the biosynthesis of the core protein of aggrecan. When the mesenchymal cells obtained from normal mouse are cultured, they form nodules rich in extracellular matrix molecules. The cells within these nodules are round. Cultures of the mesenchymal cells of a *cmd* mouse also form nodules. However, these nodules are devoid of extracellular matrix, being simple clusters of cells of various types and sizes. When aggrecan is added to the culture medium, the cells from a *cmd* mouse form nodules similar to those observed in the normal cells' culture. Immunohistochemical observation of the apparently normalized nodules with use of antiaggrecan and anti-type II collagen antibodies revealed that these two molecules coexist in the extracellular matrix as in the case of nodules formed by normal cells. The level of fibronectin mRNA in the cultured *cmd* cells was four to eight times higher than that in the cultured normal cells, and abnormal accumulation of fibronectin was observed in the nodule. This abnormal formation of fibronectin in *cmd* cells was also suppressed to normal level by the addition of aggrecan. Treatment of aggrecan with chondroitinase ABC affected the observed activities. These data indicated that proteoglycan plays an essential role in the histogenesis of cartilage.

4.8 Sugar Chains of Glycolipids

It was found in 1971 that brain gangliosides bind cholera toxin and block its toxic effect on cells. Later, GM1 was found to be the major molecule participating in the expression of the inhibitory effect. Actually, GM1 in the plasma membrane of cells works as a receptor of cholera toxin, which will lead to penetration of the A-subunit of the toxin into the cells. The A-subunit then expresses the physiological effect of cholera toxin by triggering the activation of adenylate cyclase. In the decades that followed this discovery, several other bacterial toxins were reported to interact with particular gangliosides. However, the binding specificities of these toxins have not been well established, as in the case of cholera toxin.

Glycosphingolipids can be separated by thin-layer chromatography. Upon overlaying the plate with a homogeneous suspension of bacteria or viruses labeled with ^{125}I and washing the plate with buffer, one can use radioautography to detect glycosphingolipids that bind to the bacteria or viruses. This technique permitted the elucidation of the sugar chain structures required for binding to more than 12 bacteria and 14

viruses. Interestingly, some of the structures recognized by bacteria are located not in the nonreducing termini but within the sugar chains.

Functional aspects of the sugar chains of glycosphingolipids in mammalian cell differentiation have been mainly investigated on gangliosides. It was found that the ganglioside profile of HL-60 cells, a human promyelocytic leukemia cell line, changes differently according to the cells' differentiation to granulocytes and monocytes induced by dimethyl sulfoxide and a phorbol ester, respectively. Ganglioside GM3 increases remarkably by macrophage-like cell differentiation, while neolacto-series gangliosides increase by granulocytic differentiation. An important set of observations suggests that GM3 is a highly potent inducer, able to lead the monocytic differentiation of HL-60 cells, while neolacto-series ganglioside induces granulocytic differentiation of the cells.

5 Cloning of the Glycosyltransferases and Other Enzymes Related to the Formation of Sugar Chains of Glycoconjugates

Sugar chains of glycoconjugates represent the secondary gene products formed by the concerted action of glycosyltransferases. Regulation of the biosynthesis of sugar chains is under the control of the expression of various glycosyltransferases, their substrate specificities, and their location in specific tissue sites.

The importance of substrate specificities of glycosyltransferases was first evidenced by the study of an *N*-acetylgalactosaminyltransferase, which is responsible for the formation of blood group A determinant. This enzyme requires the presence of the Fucα1$\rightarrow$2Gal group as its acceptor. Accordingly, A-enzyme cannot transfer an *N*-acetylgalactosamine residue to the Galβ1$\rightarrow$3GlcNAc group (Type I chain) or the Galβ1$\rightarrow$4GlcNAc group (Type II chain), which widely occurs in the sugar chains of glycoproteins and glycolipids. However, once a Fucα1$\rightarrow$2 residue is added to the galactose moiety of these groups, they can be acceptors of the A-enzyme.

Le^b determinant in Fig. 17, however, cannot become an acceptor of the A-enzyme in spite of having the essential Fucα1$\rightarrow$2Gal group. Therefore, the A·Le^b structure shown in Fig. 17 can only be made by the sequential action of H-, A-, and Le-enzymes.

In this biosynthetic scheme, A-enzyme and Le-enzyme compete for the same substrate sugar chain (H in Fig. 17). Accordingly, the oligosaccharide pattern that will be finally formed is controlled by the relative activities of the two enzymes expressed in a particular tissue.

Development of molecular biology opened the door of the gene-cloning age. Over 110 glycosyltransferases from various animal species were cloned by the year 2001. It had long been believed that each of the monosaccharide linkages including the anomeric structure in the sugar chains is formed by a particular enzyme. However, data obtained so far indicated that there are many cases in which a monosaccharide linkage is formed by multiple different glycosyltransferases. For example, at least six different α-fucosyltransferases were found to produce the Galβ1$\rightarrow$4(Fucα1$\rightarrow$3)GlcNAc group (Le^x determinant) from the Type II chain. This finding led to a conclusion that the Le^x determinants, which are known to widely

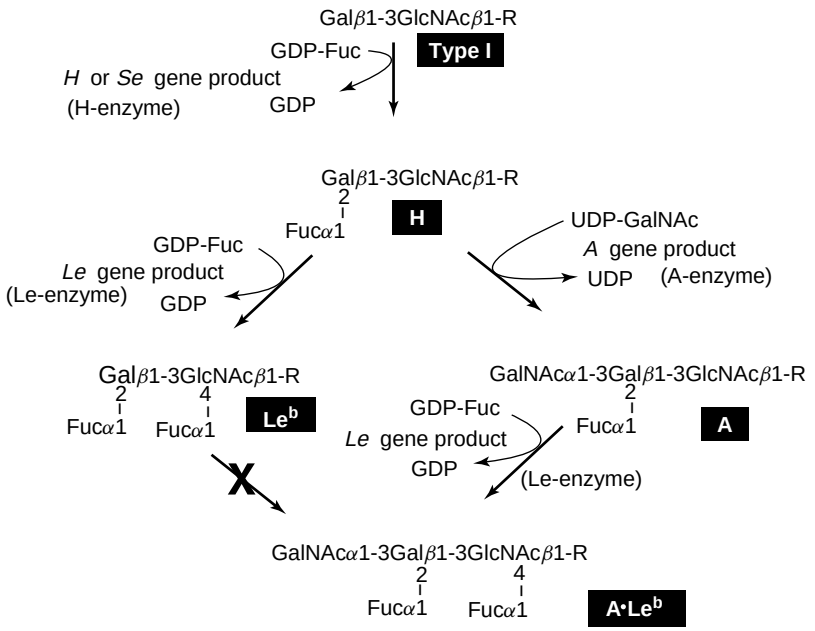

Fig. 17 Biosynthetic pathway of the A·Le^b blood group determinant.

occur in an animal body, are produced by different α-fucosyltransferases in different organs. A similar situation is found in the case of β-galactosyltransferases, which form the Galβ1→4GlcNAc group. At least seven different β-galactosyltransferases were found to catalyze the formation of the disaccharide group.

Another case of interest is found in the formation of *O*-linked sugar chains. As is evident from the structures of various *O*-linked sugar chains reported so far, formation of the GalNAcα1→Ser or Thr group in the polypeptide chains is the start to form various cores of *O*-linked sugar chains. At least nine distinct isoforms of the polypeptide α-*N*-acetylgalactosaminyltransferases (ppGalNAcTases-T1∼T9) were found to occur in mammals by the year 2001. Analysis of EST (expressed sequence tags) and genome databases suggested that additional forms of the enzyme remain to be characterized. By Northern blot analysis, it was confirmed that transcripts encoding ppGalNAcTase-T1 and -T2 are expressed in most rodents and in human tissues. ppGalNAcTase-T3 is found predominantly in human and murine testes, kidney, digestive, and reproductive tracts. The ppGalNAcTase-T4 transcript is highly expressed in rodent sublingual gland, stomach, and colon. ppGalNAcTase-T7 expression is similar to that of ppGalNAcTase-T4. ppGalNAcTase-T6 is highly expressed in human placenta and trachea. The ppGalNAcTase-T5 transcript is most abundant in murine sublingual gland and colon. The ppGalNAcTase-T9 transcript is found mainly in the murine sublingual gland, testes, and digestive and urogenital tracts. It was also found that the family of ppGalNAcTases is expressed in a unique spatial and temporal manner during murine development.

Only a subset of hydroxyamino acids are substituted with GalNAc in the mammalian glycoproteins. Thus, it was important to elucidate the enzymatic basis to determine which serines and threonines will become decorated with *O*-glycans. Studies of the kinetic properties of the expressed ppGalNAcTases from their cloned genes demonstrated that many of them have different substrate specificities toward polypeptide portions, which are partly overlapping. In addition, the expression patterns of various ppGalNAcTases in various tissues are different. Therefore, the initiation of mucin-type *O*-glycosylation appears to have a much more complex regulation than previously expected.

On the basis of the evidences so far described, organ-specific expression of a glycosyltransferase is now becoming an important target as a basis of the formation of the oligosaccharide pattern of a glycoprotein in the organ. In addition to this, elucidation of the mechanism of the different expression of one glycosyltransferase in different organs will be an important target, because of the presence of organ-specific glycosylation of a protein as evidenced by the study of γ-glutamyltranspeptidases (see Sect. 4.4 and Fig. 14).

A detailed study of the substrate specificities of glycosyltransferases including the structures of the polypeptide portions of the glycoprotein acceptors is also considered an important target in glycobiology, because there are many cases in which two glycoproteins produced by a single cell contain different sets of oligosaccharides. The phenomenon of site-directed glycosylation as evidenced in hCG may be explained by such studies.

See also Bioorganic Chemistry.

Bibliography

Books and Reviews

Allen, H.J., Kisailus, E.C. (Eds.) (1992) *Glycoconjugates: Composition, Structure, and Function*, Marcel Dekker, New York.

Conrad, H.E. (Ed.) (1998) *Heparin-Binding Proteins*, Academic Press, San Diego.

Fukuda, M., Kobata, A. (Eds.) (1993) *Glycobiology–A Practical Approach*, IRL Press, Oxford, UK.

Kobata, A. (1992) *Structures and Functions of the Sugar Chains of Glycoproteins*, *Eur. J. Biochem.* **209**, 483–501.

Montreuil, J., Vliegenthart, J.F.G., Schachter, H. (Eds.) (1995) *Glycoproteins I*, Elsevier, Amsterdam.

Montreuil, J., Vliegenthart, J.F.G., Schachter, H. (Eds.) (1997) *Glycoproteins II*, Elsevier, Amsterdam.

Montreuil, J., Vliegenthart, J.F.G., Schachter, H. (Eds.) (1996) *Glycoproteins and Disease*, Elsevier, Amsterdam.

Takahashi, N., Muramatsu, T. (Eds.) (1992) *CRC Handbook of Endoglycosidases and Glycoamidases*, CRC Press, Boca Raton.

Taniguchi, N., Honke, K., Fukuda, M. (Eds.) (2002) *Handbook of Glycosyltransferases and Related Genes*, Springer-Verlag, Berlin.

Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., Marth, J. (Eds.) (1999) *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, New York.

Primary Literature

Akarna, T.O., Nakagawa, H., Sugihara, K., Narisawa, S., Ohyama, C., Nishimura, S., O'Brien, D.A., Moremen, K., Millán, J.L., Fukuda, M.N. (2002) Germ cell survival through carbohydrate-mediated interaction with sertoli cells, *Science* **295**, 124–127.

Amano, J., Straehl, P., Berger, E.G., Kochibe, N., Kobata, A. (1991) Structures of mucin-type sugar chains of the galactosyltransferase purified from human milk. Occurrence of the ABO and Lewis blood group determinants, *J. Biol. Chem.* **266**, 11461–11477.

Ashikari, S., Habuchi, H., Kimata, K. (1995) Characterization of heparan sulfate oligosaccharides that bind to hepatocyte growth factor, *J. Biol. Chem.* **270**, 29586–29593.

Bai, X.M., Bame, K.J., Habuchi, H., Kimata, K., Esko, J.D. (1997) Turnover of heparan sulfate depends on 2-*O*-sulfation of uronic acids, *J. Biol. Chem.* **272**, 23172–23179.

Blair, S.S. (2000) Notch signaling: fringe really is a glycosyltransferase, *Curr. Biol.* **10**, R608–R612.

Carlson, D.M. (1968) Structures and immunochemical properties of oligosaccharides isolated from pig submaxillary mucins, *J. Biol. Chem.* **243**, 616–626.

Cebo, C., Dambrouck, T., Maes, E., Laden, C., Strecker, G., Michalski, J.C., Zanetta, J.P. (2001) Recombinant human interleukins IL-Ialpha, IL-Ibeta, IL-4, IL-6, and IL-7 show different and specific calcium dependent carbohydrate-binding properties, *J. Biol. Chem.* **276**, 5685–5691.

Chiba, A., Matsumura, K., Yamada, H., Inazu, T., Shimizu, T., Kusunoki, S., Kanazawa, I., Kobata, A., Endo, T. (1997) Structures of sialylated *O*-linked oligosaccharides of bovine peripheral nerve α-dystroglycan: the role of a novel *O*-mannosyl-type oligosaccharide in the binding of α-dystroglycan with laminin, *J. Biol. Chem.* **272**, 2156–2162.

Chiba, Y., Suzuki, M., Yoshida, S., Yoshida, A., Ikenaga, H., Takeuchi, M., Jigami, Y., Ichishima, E. (1998) Production of human compatible high mannose-type ($Man_5GlcNAc_2$) sugar chains in *Saccharomyces cerevisiae*, *J. Biol. Chem.* **273**, 26298–26304.

Coetzee, T., Fujita, N., Dupree, J., Shi, R., Blight, A., Suzuki, K., Suzuki, K., Popko, B. (1996) Myelination in the absence of galactocerebroside and sulfatide: normal structure with abnormal function and regional instability, *Cell* **86**, 209–219.

Collins, B.E., Ito, H., Sawada, N., Ishida, H., Kiso, M., Schnaar, R.L. (1999) Enhanced binding of the neural siglecs, myelin-associated glycoprotein and schwann cell myelin protein, to Chol-1 (alpha-series) gangliosides and novel sulfated Chol-1 analogs, *J. Biol. Chem.* **274**, 37637–37643.

Dairaku, K., Spiro, R.G. (1997) Phylogenic survey of endomannosidase indicates late evolutionary appearance of this N-linked oligosaccharide processing enzyme, *Glycobiology* **7**, 579–586.

Day, P.J., Owens, S.R., Wesche, J., Olsnes, S., Roberts, L.M., Lord, J.M. (2001) An interaction between ricin and calreticulin that may have implications for toxin trafficking, *J. Biol. Chem.* **276**, 7202–7208.

Dell, A., Morris, H.R. (2001) Glycoprotein structure determination by mass spectrometry, *Science* **291**, 2351–2356.

Delorme, E., Lorenzini, T., Griffin, J., Martin, F., Jacobsen, F., Boone, T., Elliott, S. (1992) Role of glycosylation on the secretion and biological activity of erythropoietin, *Biochemistry* **31**, 9871–9876.

Doering, T.L., Masterson, W.J., Hart, G.W., Englund, P.T. (1990) Biosynthesis of the glycosyl phosphatidylinositol membrane anchors, *J. Biol. Chem.* **265**, 611–614.

Dwir, O., Kansas, G.S., Alon, R. (2001) Cytoplasmic anchorage of L-selectin controls leukocyte capture and rolling by increasing the mechanical stability of the selectin tether, *J. Biol. Chem.* **155**, 145–156.

Ellies, L.G., Tsuboi, S., Petryniak, B., Lowe, J.B., Fukuda, M., Marth, J.D. (1998) Core 2 oligosaccharide biosynthesis distinguishes between selectin ligands essential for leukocyte homing and inflammation, *Immunity* **9**, 881–890.

Endo, Y., Yamashita, K., Tachibana, Y., Tojo, S., Kobata, A. (1979) Structures of the asparagine-linked sugar chains of human chorionic gonadotropin, *J. Biochem.* **85**, 669–679.

Epperson, T.K., Patel, K.D., McEver, R.P., Cummings, R.D. (2000) Noncovalent association of P-selectin glycoprotein ligand-1 and minimal determinants for binding to P-selectin, *J. Biol. Chem.* **275**, 7839–7853.

Ferguson, M.A., Homans, S.W., Dwek, R.A., Rademacher, T.W. (1988) Glycosyl-phosphatidylinositol moiety that anchors *Trypanosoma brucei* variant surface glycoprotein to the membrane, *Science* **239**, 753–759.

Fiete, D.J., Beranek, M.C., Baenziger, J.U. (1998) A cystein-rich domain of the mannose receptor mediates GalNAc-4-SO_4 binding, *Proc. Natl. Acad. Sci. USA* **95**, 2089–2093.

Fukuda, M.N., Matsumura, G. (1976) Endo-β-galactosidase from *Escherichia freundii*: purification and endoglycosidic action on keratan sulfates, oligosaccharides, and blood group active glycoprotein, *J. Biol. Chem.* **251**, 6218–6225.

Fukushima, K., Hara-Kuge, S., Ohkura, T., Seko, A., Ideo, H., Inazu, T., Yamashita, K. (1997) Lectin-like characteristics of recombinant human interleukin-Ibeta recognizing

glycans of the glycosylphosphatidylinositol anchor, *J. Biol. Chem.* **272**, 10579–10584.

Gao, Y., Wells, L., Comer, F.I., Parker, G.J., Hart, G. (2001) Dynamic *O*-glycosylation and cytosolic proteins: cloning and characterization of a neutral cytosolic β-*N*-acetylglucosaminidase from human brain, *J. Biol. Chem.* **276**, 9838–9845.

Green, E.D., Morishima, C., Boime, I., Baenziger, J.U. (1985) Structural requirements for sulfation of asparagine-linked oligosaccharides of lutropin, *Proc. Natl. Acad. Sci. USA* **82**, 7850–7854.

Habuchi, H., Kobayashi, M., Kimata, K. (1998) Molecular characterization and expression of heparan-sulfate 6-sulfotransferase, *J. Biol. Chem.* **273**, 9208–9213.

Hakomori, S. (1964) A rapid permethylation of glycolipid and polysaccharide catalyzed by methylsulfinyl carbanion in dimethyl sulfoxide, *J. Biochem.* **55**, 205–208.

Haltiwnger, R.S., Blomberg, M.A., Hart, G.W. (1992) Glycosylation of nuclear and cytoplasmic proteins. Purification and characterization of a uridine diphospho-*N*-acetylglucosamine: polypeptide beta-*N*-acetylglucosaminyltransferase, *J. Biol. Chem.* **267**, 9005–9013.

Hammond, C., Braakman, I., Helenius, A. (1994) Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control, *Proc. Natl. Acad. Sci. USA* **91**, 913–917.

Hanisch, F.-G., Reis, C.A., Clausen, H., Paulsen, H. (2001) Evidence for glycosylation-dependent activities of polypeptide *N*-acetylgalactosaminyltransferases rGalNAc-T2 and -T4 on mucin glycopeptides, *Glycobiology* **11**, 731–740.

Hart, G.W., Kreppel, L.K., Comer, F.I., Arnold, C.S., Snow, D.M., Ye, Z.Y., Cheng, X.G., DellaManna, D., Caine, D.S., Earles, B.J., Akimoto, Y., Cole, R.N., Hayes, B.K. (1996) O-GlcNAcylation of key nuclear and cytoskeletal proteins: reciprocity with O-phosphorylation and putative roles in protein multimerization, *Glycobiology* **6**, 711–716.

Hase, S., Ikenaka, T., Matsushima, Y. (1978) Structure analyses of oligosaccharides by tagging of the reducing end sugars with fluorescent compound, *Biochem. Biophys. Res. Commun.* **85**, 257–263.

Heikkinen, J., Risteli, M., Wang, C., Latvala, J., Rossi, M., Valtavaara, M., Myllyla, R. (2000) Lysyl hydroxylase 3 is a multifunctional protein processing collagen glucosyltransferase activity, *J. Biol. Chem.* **275**, 36158–36163.

Hu, Y., Szente, B., Kiely, J.-M., Gimbrone, M.A. (2001) Molecular events in transmembrane signaling via E-selectin. SHP2 association, adapter protein complex formation and ERK1/2 activation, *J. Biol. Chem.* **276**, 48549–48553.

Imai, Y., Singer, M.S., Fennie, C., Lasky, L.A., Rosen, S.D. (1991) Identification of a carbohydrate based endothelial ligand for a lymphocyte homing receptor, *J. Cell Biol.* **113**, 1213–1221.

Inoue, S., Kitajima, K., Inoue, Y. (1996) Identification of KDN (deaminoneuraminic acid) residues in mammalian tissues and human lung carcinoma cells. Evidence of the occurrence of KDN-glycoconjugates in mammals, *J. Biol. Chem.* **271**, 24341–24344.

Ito, M., Yamagata, T. (1986) A novel glycosphingolipid-degrading enzyme cleaves the linkage between the oligosaccharide and ceramide of neutral and acidic glycosphingolipids, *J. Biol. Chem.* **261**, 14278–14282.

Ito, S., Yamashita, K., Spiro, R.G., Kobata, A. (1977) Structure of a carbohydrate moiety of a unit A glycopeptide of calf thyroglobulin, *J. Biochem.* **81**, 1621–1631.

Kagawa, Y., Takasaki, S., Utsumi, J., Hosoi, K., Shimizu, H., Kochibe, N., Kobata, A. (1988) Comparative study of the asparagine-linked sugar chains of natural human Interferon-β1 and recombinant human Interferon-β1 produced by three different mammalian cells, *J. Biol. Chem.* **263**, 17508–17515.

Kessler, M.J., Reddy, M.S., Shah, R.H., Bahl, O.P. (1979) Structures of N-glycosidic carbohydrate units of human chorionic gonadotropin, *J. Biol. Chem.* **254**, 7901–7908.

Kobata, A. (1979) Use of endo- and exoglycosidases for structural studies of glycoconjugates, *Anal. Biochem.* **100**, 1–14.

Koide, N., Muramatsu, T. (1974) Endo-β-*N*-acetylglucosaminidase acting on carbohydrate moieties of glycoproteins: purification and properties of the enzyme from *Diplococcus pneumoniae*, *J. Biol Chem.* **249**, 4897–4904.

Kreuger, J., Salmivirta, M., Sturiale, L., Gimenez-Gallego, G., Lindhal, U. (2001) Sequence analysis of heparan sulfate epitopes with graded affinities for fibroblast growth factors 1 and 2, *J. Biol. Chem.* **276**, 30744–30752.

Krieg, J., Glasner, W., Vicentini, A., Doucey, M.A., Loffler, A., Hess, D., Hofsteenge, J. (1997) C-mannosylation of human RNase 2 is an intracellular process performed by a variety of cultured cells, *J. Biol. Chem.* **272**, 26687–26692.

Lee, S.J., Evers, S., Roeder, D., Parlow, A.F., Risteli, J., Lee, Y.C., Feizi, T., Langen, H., Nussenzweig, M.C. (2002) Mannose receptor-mediated regulation of serum glycoprotein homeostasis, *Science* **295**, 1898–1901.

Lowe, J.B., Ward, P.A. (1997) Therapeutic inhibition of carbohydrate-protein interactions in vivo, *J. Clin. Invest.* **99**, 822–826.

Lübke, T., Marquardt, T., Etzioni, A., Hartmann, E., von Figura, K., Körner, C. (2001) Complementation cloning identifies CDG-IIc, a new type of congenital disorder of glycosylation, as a GDP-fucose transporter deficiency, *Nat. Genet.* **28**, 73–76.

Malhotra, R., Wormald, M.R., Rudd, P.M., Fischer, P.B., Dwek, R.A., Sim, R.B. (1995) Glycosylation changes of IgG associated with rheumatoid arthritis can activate complement via the mannose-binding protein, *Nat. Med.* **1**, 237–243.

Manzella, S.M., Hooper, L.V., Baenziger, J.U. (1996) Oligosaccharides containing β1,4-linked N-acetylgalactosamine, a paradigm for protein-specific glycosylation, *J. Biol. Chem.* **271**, 12117–12120.

Mellquist, J.L., Kasturi, L., Spitalnik, S.L., Shakin-Eshleman, S.H. (1998) The amino acid following an Asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency, *Biochemistry* **37**, 6833–6837.

Mizuochi, T., Kobata, A. (1980) Different asparagine-linked sugar chains on the two polypeptide chains of human chorionic gonadotropin, *Biochem. Biophys. Res. Commun.* **97**, 772–778.

Mori, E., Takasaki, S., Hedrick, J.L., Wardrip, N.J., Mori, T., Kobata, A. (1991) Neutral oligosaccharide structures linked to asparagine of porcine zona pellucida glycoproteins, *Biochemistry* **30**, 2078–2087.

Nishikawa, Y., Pegg, W., Paulsen, H., Schachter, H. (1988) Control of glycoprotein synthesis: purification and characterization of rabbit liver UDP-*N*-acetylglucosamine:α-3-D-mannoside β-1,2-*N*-acetylglucosaminyltransferase I, *J. Biol. Chem.* **263**, 8270–8281.

Nitz, M., Ling, C.-C., Otter, A., Cutler, J.E., Bundle, D.R. (2002) The unique solution structure and immunochemistry of the *Candida albicans* β-1,2-mannopyranan cell wall antigens, *J. Biol. Chem.* **277**, 3440–3446.

Ogata, S., Muramatsu, T., Kobata, A. (1975) Fractionation of glycopeptides by affinity column chromatography on concanavalin A-Sepharose, *J. Biochem.* **78**, 687–696.

Oike, Y., Kimata, K., Shinomura, T., Suzuki, S., Takahashi, N., Tanabe, K. (1982) A mapping technique for probing the structure of proteoglycan core molecules, *J. Biol. Chem.* **257**, 9751–9758.

Ohyama, Y., Kasai, K., Nomoto, H., Inoue, Y. (1985) Frontal affinity chromatography of ovalbumin glycoasparagine on a concanavalin-A Sepharose column. A quantitative study of the binding specificity of the lectin, *J. Biol. Chem.* **260**, 6882–6887.

Parekh, R.B., Dwek, R.A., Sutton, B.J., Fernandes, D.L., Leung, A., Stanworth, D., Rademacher, T.W., Mizuochi, T., Taniguchi, T., Matsuta, K., Takeuchi, F., Nagano, Y., Miyamoto, T., Kobata, A. (1985) Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG, *Nature* **316**, 452–457.

Parsons, T.F., Pierce, J.G. (1984) Free alpha-like material from bovine pituitaries. Removal of its O-linked oligosaccharide permits combination with lutropin-beta, *J. Biol. Chem.* **259**, 2662–2666.

Perrimon, N., Bernfield, M. (2000) Specificities of heparan sulfate proteoglycans I developmental processes, *Nature* **404**, 725–728.

Purohit, S., Shao, K., Balasubramanian, S.V., Bahl, O.P. (1997) Mutants of human chorionic gonadotropin lacking N-glycosyl chains in the α-subunit: mechanism for the differential action of the N-linked carbohydrates, *Biochemistry* **36**, 12355–12363.

Reitman, M.L., Kornfeld, S. (1981) Lysosomal enzyme targeting: *N*-Acetylglucosaminylphosphotransferase selectively phosphorylate native lysosomal enzymes, *J. Biol. Chem.* **256**, 11977–11980.

Rice, K.G., Weisz, O.A., Barthel, T., Lee, R.T., Lee, Y.C. (1990) Defined stoichiometry of binding between triantennary glycopeptide and the asialoglycoprotein receptor of rat hepatocytes, *J. Biol. Chem.* **265**, 18429–18434.

Robbins, P.W., Hubbard, S.C., Turco, S.L., Wirth, D.F. (1977) Proposal for a common oligosaccharide intermediate in the synthesis of membrane glycoproteins, *Cell* **12**, 893–900.

Sargiacomo, M., Sudol, M., Tang, Z., Lisanti, M.P. (1993) Signal transducing molecules and glycosyl-phosphatidylinositol-linked proteins form a caveolin-rich insoluble complex in MDCK cells, *J. Cell Biol.* **122**, 789–807.

Schrag, J.D., Bergerson, J.J., Li, Y., Borisova, S., Hahn, M., Thomas, D.Y., Cygler, M. (2001) The structure of calnexin, an ER chaperone involved in quality control of protein folding, *Mol. Cell* **8**, 633–644.

Skelton, T.P., Hooper, L.U., Srivastava, V.V., Hindsgaul, O., Baenziger, J.U. (1991) Characterization of a sulfotransferase responsible for the 4-O-sulfation of terminal beta-N-acetyl-D-galactosamine on asparagine-linked oligosaccharides of glycoprotein hormones, *J. Biol. Chem.* **266**, 17142–17150.

Srikrishna, G., Wang, L., Freeze, H.H. (1998) Fucose β-1-P-Ser is a new type of glycosylation: using antibodies to identify a novel structure in *Dictyostelium discoideum* and study multiple types of fucosylation during growth and development, *Glycobiology* **8**, 799–811.

Stanley, P., Ioffe, E. (1995) Glycosyltransferase mutants: key to new insights in glycobiology, *FASEB J.* **9**, 1436–1444.

Steeves, R.M., Denton, M.E., Barnard, F.C., Henry, A., Lambert, J.M. (1999) Identification of three oligosaccharide binding sites in ricin, *Biochemistry* **38**, 11677–11685.

Sturla, L., Etzioni, A., Bisso, A., Zanardi, D., De Flora, G., Silengo, L., De Flora, A., Tonetti, M. (1998) GDP-D-mannose-4,6-dehydratase in leukocyte adhesion deficiency type II syndrome, *FEBS Lett.* **429**, 274–278.

Tabas, I., Schlesinger, S., Kornfeld, S. (1978) Processing of high mannose oligosaccharides to form complex type oligosaccharides on the newly synthesized polypeptides of the vesicular stomatitis virus G protein and the IgG heavy chain, *J. Biol. Chem.* **253**, 716–722.

Tachibana, Y., Yamashita, K., Kobata, A. (1982) Substrate specificity of mammalian endo-β-*N*-acetylglucosaminidase: study with the enzyme of rat liver, *Arch. Biochem. Biophys.* **214**, 199–210.

Tai, T., Yamashita, K., Kobata, A. (1975) Synthesis and mass-fragmentographic analysis of partially *O*-methylated 2-*N*-methylglucosamines, *J. Biochem.* **78**, 679–686.

Takahashi, N. (1977) Demonstration of a new amidase acting on glycopeptides, *Biochem. Biophys. Res. Commun.* **76**, 1194–1201.

Takasaki, S., Kobata, A. (1976) Purification and characterization of an endo-β-galactosidase produced by *Diplococcus pneumoniae*, *J. Biol. Chem.* **251**, 3603–3609.

Takasaki, S., Mizuochi, T., Kobata, A. (1982) Hydrazinolysis of asparagine-linked sugar chains to produce free oligosaccharides, *Methods Enzymol.* **83**, 263–268.

Takatsuki, A., Arima, K., Tamura, G. (1971) Tunicamycin, a new antibiotic I. Isolation and characterization of tunicamycin, *J. Antibiot.* **24**, 215–223.

Takeuchi, M., Inoue, N., Strickland, T.W., Kubota, M., Wada, M., Shimizu, R., Hoshi, S., Kozutsumi, H., Takasaki, S., Kobata, A. (1989) Relationship between sugar chain structure and biological activity of recombinant human erythropoietin produced in Chinese hamster ovary cells, *Proc. Natl. Acad. Sci. USA* **86**, 7819–7822.

Tarentino, A.L., Maley, F. (1974) Purification and properties of an endo-β-*N*-acetylglucosaminedase from *Streptomyces griseus*, *J. Biol. Chem.* **249**, 811–817.

Terayama, K., Oka, S., Seiki, T., Miki, Y., Nakamura, A., Kozutsumi, Y., Takio, K., Kawasaki, T. (1997) Cloning and functional expression of a novel glucuronyltransferase involved in the biosynthesis of the carbohydrate epitope HNK-I, *Proc. Natl. Acad. Sci. USA* **94**, 6093–6098.

Tomiya, N., Awaya, J., Kurono, M., Endo, S., Arata, Y., Takahashi, N. (1988) Analysis of N-linked oligosaccharides using a two-dimensional mapping technique, *Anal. Biochem.* **171**, 73–90.

Tozawa, R., Ishibashi, S., Osuga, J., Yamamoto, K., Yagyu, H., Ohashi, K., Tamura, Y., Yahagi, N., Iizuka, Y., Okazaki, H., Harada, K., Gotoda, T., Shimano, H., Kimura, S., Nagai, R., Yamada, N. (2001) Asialoglycoprotein receptor deficiency in mice lacking the major receptor subunit: its obligate requirement for the stable expression of oligomeric receptor, *J. Biol. Chem.* **276**, 12624–12628.

Tsuchiya, N., Endo, T., Matsuta, K., Yoshinoya, S., Aikawa, T., Kosuge, E., Takeuchi, F., Miyamoto, T., Kobata, A. (1989) Effects of galactose depletion from oligosaccharide chains on immunological activities of human IgG, *J. Rheumatol.* **16**, 285–290.

Uchida, T., Robbins, P.W. (1963) Analysis of serologic determinant groups of the Salmonella E-group O-antigen, *Biochemistry* **2**, 663–668.

Uchimura, K., Muramatsu, H., Kadomatsu, K., Fan, Q.W., Kurosawa, N., Mitsuoka, C., Kannagi, R., Habuchi, O., Muramatsu, T. (1998) Molecular cloning and characterization of an *N*-acetylglucosamine-6-*O*-sulfotransferase, *J. Biol. Chem.* **273**, 22577–22583.

Wandall, H.H., Hassan, H., Mirgorodskaya, E., Kristensen, A.K., Roepstorff, P., Bennett, E.P., Nielsen, P.A., Hollingsworth, M.A., Burchell, J., Taylor-Papadimitriou, J., Clausen, H. (1997) Substrate specificities of three members of the human UDP-*N*-acetyl-a-D-galactosamine: polypeptide *N*-acetylgalactosaminyltransferase family, GalNAc-T1, -T2, and -T3, *J. Biol. Chem.* **272**, 23503–23514.

Wallis, R., Dodd, R.B. (1998) Interaction of mannose-binding protein with associated serine proteases: effects of naturally occurring mutations, *J. Biol. Chem.* **275**, 30962–30969.

Weisshaar, G., Hiyama, J., Renwick, R.G.C. (1991) Site-specific N-glycosylation of human chorionic gonadotrophin – structural analysis of glycopeptides by one- and two-dimensional ^{1}H NMR spectroscopy, *Glycobiology* **1**, 393–404.

Wilkins, P.P., Moore, K.L., McEver, R.P., Cummings, R.D. (1995) Tyrosine sulfation of P-selectin glycoprotein ligand-1 is required for high affinity binding to P-selectin, *J. Biol. Chem.* **270**, 22677–22680.

Wyss, D.F., Choi, J.S., Li, J., Knoppers, M.H., Willis, K.J., Arulanandam, A.R.N., Smolyar, A., Reinherz, E.L., Wagner, G. (1995) Conformation and function of the N-linked glycan in the cell adhesion domain of human CD2, *Science* **269**, 1273–1278.

Yanagishita, M., Hascall, V. (1992) Cell surface heparan sulfate proteoglycans, *J. Biol. Chem.* **267**, 9451–9454.

Yamagata, T., Saito, H., Habuchi, O., Suzuki, S. (1968) Purification and properties of bacterial chondroitinases and chondrosulfatases, *J. Biol. Chem.* **243**, 1523–1535.

Yoshida, A., Kobayashi, K., Manya, H., Taniguchi, K., Kano, H., Mizuno, M., Inazu, T., Mitsuhashi, H., Takahashi, S., Takeuchi, M., Hermann, R., Straub, V., Talim, B., Voit, T., Topalogu, H., Toda, T., Endo, T. (2001) Muscular dystrophy and neuronal migration disorder caused by mutations in a glycosyltransferase, POMGnT1, *Dev. Cell* **1**, 717–724.

Yu, S., Kitajima, K., Inoue, S., Inoue, Y. (1991) Isolation and structural elucidation of novel type of ganglioside, deaminated neuraminic acid (KDN)-containing glycosphingolipid, from rainbow trout sperm. The first example of the natural occurrence of KDN-ganglioside, (KDN)GM_3, *J. Biol. Chem.* **266**, 21929–21935.

Zhang, Y., Iwasa, T., Tsuda, M., Kobata, A., Takasaki, S. (1997) A novel monoantennary complex-type sugar chain found in octopus rhodopsin: occurrence of the Galβ1→4Fuc group linked to the proximal *N*-acetylglucosamine residue of the trimannosyl core, *Glycobiology* **7**, 1153–1158.

Glycoproteins: Roles in Cellular Homeostasis and Disease

Michael William King
Indiana University School of Medicine, Center for Regenerative Biology and Medicine, Holmstedt Hall, Terre Haute, IN, USA

1 Introduction 572

2 Mammalian Glycopeptide Linkages 574

3 Glycoprotein Synthesis 577
3.1 Synthesis of N-linked Glycopeptides 577
3.2 Synthesis of O-linked Glycopeptides 579

4 Role of N-glycosylation in Protein Folding 581

5 Lysosomal Targeting by Specific Glycosylation 581

6 Glycoproteins: Recognition and Homeostasis 582
6.1 Animal Lectins 582
6.1.1 C-type Lectins 583
6.1.2 S-type Lectins: Galectins 585
6.1.3 I-type Lectins: Siglecs 586
6.1.4 Pentraxins 587
6.2 Lectin Activities of the Cytokines 588
6.3 Sugar Alterations that Affect Immunity 589

7 Clinical Significances of Glycoproteins 590
7.1 Degradation Defects of GlcNAc-Asn Glycoproteins 590
7.2 Congenital Disorders of Glycosylation, CDG 593
7.2.1 Leukocyte Adhesion Deficiency Syndrome II, LAD II 595
7.3 Exploitation of Host Glycosylation 596

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

8 Medical Applications of Carbohydrates and Glycoproteins 598

Bibliography 600
Books and Reviews 600
Primary Literature 600

Keywords

Glycoconjugate
Structure composed of at least one sugar residue linked to another compound such as protein or lipid.

Glycosidase
Any of the class of enzymes that hydrolyze glycosidic bonds.

Glycosyltransferase
Any of the class of enzymes that synthesize glycosidic bonds.

Glycosidic
Referring to the covalent attachment of, or to, sugar residues.

Glycan
Complex molecule composed in part of carbohydrate.

Oligosaccharide
Compound composed of several sugars residues linked together.

Lectin
Compound with affinity for binding sugar residues.

Interleukin
Term for a group of protein factors that affect primary cells and are derived from macrophages and T cells that have been stimulated by antigens or mitogens.

Cytokine
Substances that are produced by cells of the immune system and can affect the immune response.

Chemokine
Proinflammatory activation-inducible proteins that are mainly chemotactic for different cell types.

Proteosome
Multiprotein complex involved in the degradation of cellular proteins.

N-glycosidic
Referring to the linkage of sugars to nitrogen residues.

O-glycosidic
Referring to the linkage of sugars to hydroxyl residues.

Abbreviations

Dol, Dol-P, Dol-P-P:	dolichol, dolichol phosphate, dolichol pyrophosphate
Gal:	galactose
Glc:	glucose
GPI:	glycophosphatidylethanola
GalNAc:	N-acetylgalactosamine
GlcNAc:	N-acetylglucosamine
Man:	mannose
Fruc:	fructose
Fuc:	fucose
Sia, NeuAc, NANA:	sialic acid
GDP:	guanosine diphosphate
UDP:	uridine diphosphate
Asn:	asparagine

Modification of proteins through the addition of carbohydrates is one of the most significant alterations that occurs to these biological molecules and serves multiple purposes. Quite striking is the fact the protein glycosylation is controlled by factors that do not depend upon the information encoded in our genes. Sugar attachment is necessary in many proteins to ensure that they attain the proper folded state (i.e. structure) and thus, can carry out their intended functional roles. Equally important is the role glycosylation plays in the context of information. Cell surface glycoproteins are necessary for accurate and efficient cell–cell adhesion and cell–cell communication. The ability of certain families of proteins (lectins) to properly recognize the context of glycosylation on other structures (primarily proteins) is also critically important in cellular ''information exchange.'' In order to control the regulatory effects of glycosylation, the process can be in a constant state of flux, changing as cells differentiate. Probably, most telling is that many examples are known where disruptions in the pattern of glycosylation, as well as in the degradation of glycosylated proteins, lead to pathological conditions.

1 Introduction

The attachment of carbohydrate to proteins and lipids generates a class of molecules referred to as the glycoconjugates. Glycoconjugates are found on the cell surface, within the cytoplasm and nucleus, and circulate in the plasma. These molecules serve as important regulators of cell–cell communication, cell–cell adhesion and cell–cell recognition, modulation of immune responses, and regulation of many other biological processes such as coagulation. Attachment of sugar residues to proteins generates the specialized class of molecules called glycoproteins. Sugar attachment takes place both co- and post-translationally and is controlled by factors that do not depend upon the information encoded in our genes. At least 13 different carbohydrates have been identified attached to protein. These glycoprotein linkages include, N-glycosylation, O-glycosylation, glypiation (GPI bridge), C-mannosylation, and phosphoglycation. The most frequently occurring glycoprotein modifications are the N- and O-glycosylations that take place within the endoplasmic reticulum, Golgi, and cytosol.

The first description of carbohydrate linkage to protein was that of the N-glycosylation of the egg protein ovalbumin. Subsequently, numerous proteins have been shown to be modified by carbohydrate attachment and most of the commonly occurring monosaccharides have been found in these attachments. In fact, glycosylation of proteins as a functionally significant modification spans the spectrum of organisms from bacteria to plants to animals. The enzymes necessary for carrying out the attachment of sugars to proteins are collectively called glycosyltransferases and at least 16 of this class of enzymes have been identified. Enzymes are also necessary for the removal of sugars in the context of both carbohydrate remodeling and glycoprotein degradation. These latter enzymes are called glycosidases.

Membrane-associated carbohydrate is exclusively in the form of oligosaccharides covalently attached to proteins forming glycoproteins, and to a lesser extent, covalently attached to lipid forming the glycolipids. The predominant sugars found in glycoproteins are glucose (Glc), galactose (Gal), mannose (Man), fucose (Fuc), fructose (Fruc), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc), xylose (Xyl), and *N*-acetylneuraminic acid (NANA or NeuAc); the latter being more commonly referred to as sialic acid (Sia). Figure 1 shows the symbols used to designate these sugars in the other figures. The distinction between proteoglycans and glycoproteins resides in the level and types of carbohydrate modification. The carbohydrate modifications found in glycoproteins are rarely complex: carbohydrates are linked to the protein component through either O-glycosidic or N-glycosidic bonds. The N-glycosidic linkage is through the amide group of asparagine. The O-glycosidic linkage is to the hydroxyl of serine (Ser), threonine (Thr), or hydroxylysine (Hyl) and in some cases, hydroxyproline (Hyp). The linkage of carbohydrate to hydroxylysine and hydroxyproline is generally found only in the collagens. The linkage of carbohydrate to 5-hydroxylysine is either the single sugar galactose or the disaccharide glucosylgalactose. In Ser- and Thr-type O-linked glycoproteins, the carbohydrate directly attached to the protein is most frequently *N*-acetylgalactosamine (GalNAc). The sugar directly attached to asparagines in N-glycosylations is most frequently *N*-acetylglucosamine (GlcNAc).

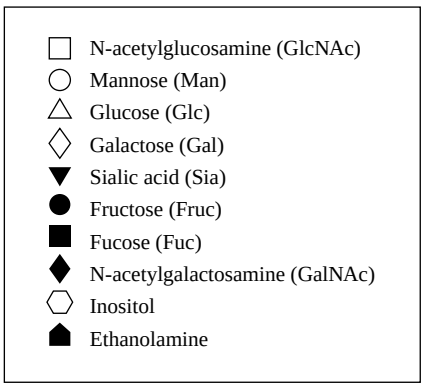

Fig. 1 Symbols used to depict the various carbohydrates, as well as other associated compounds, present in the structures described in Figs. 2 through 6.

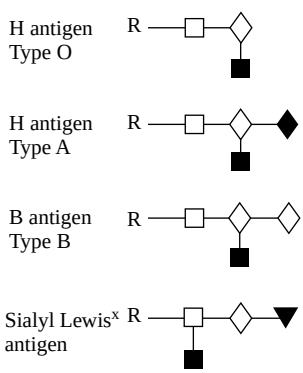

Fig. 2 Structures of the carbohydrates that constitute the ABO and sialyl Lewis^x antigens. "R" refers to either the protein or ceramide backbone of the ABO antigens. The ABO antigens are linked to proteins when they are secreted; they are linked to ceramides when present on plasma membranes. The "R" indicates the protein backbone where the carbohydrate antigen is directly added to asparagine residues in the typical GlcNAc-β-Asn linkage. When the ABO antigens are added to lipid, "R" represents the ceramide backbone of the membrane-associated glycosphingolipids.

That protein glycosylation is indeed an important component of cellular homeostasis is evidenced by the frequency of its occurrence, its requirement for protein structure and function, and the high degree of association between disease and abnormal glycoprotein synthesis and degradation. Further proof of the importance of protein glycosylation stems from an analysis of the degree of its occurrence. It has been estimated that when combining O-glycosylation and N-glycosylation of proteins, half of all proteins will eventually be identified as being carbohydrate modified.

In many cases, carbohydrates that are attached to proteins, and the structures they form when attached, are also found attached to lipids. A clinically and forensically relevant example of this overlap of modification between proteins and lipids is reflective of the ABO blood group antigens. The ABO blood group antigens are the carbohydrate moieties of glycolipids on the surface of cells as well as the carbohydrate portion of serum glycoproteins (Fig. 2). When present on the surface of cells, the ABO carbohydrates are linked to sphingolipid and are therefore of the glycosphingolipid class. When the ABO carbohydrates are associated with protein in the form of glycoproteins they are found in the serum and are referred to as the secreted forms of these antigens. Some individuals produce the glycoprotein forms

of the ABO antigens while others do not. This fact distinguishes the property of secretors from nonsecretors and this characteristic is the result of the action of the secretor (*Se*) locus.

The aim of this chapter is to review the types of protein glycosylation, the mechanisms of synthesis of the glycopeptide linkages, and the mechanisms involved in glycopeptide remodeling and degradation. In addition, the importance of carbohydrate recognition, through the actions of the animal lectins, will be discussed. An emphasis will be placed on the function of lectin activity in innate and acquired immunity reactions. In order to demonstrate the broad importance of protein glycosylation, numerous examples of the clinical relevance of carbohydrate synthesis and degradation will be presented. In addition, the therapeutic potential for inhibitors of glycosylation as well as the use of sugars themselves as treatment for certain conditions will be discussed. Recent advances in the field of glycobiology have demonstrated the relevance of the "sugar code" to medicine and indeed, this field is recognized as a major new discipline of medical significance.

2 Mammalian Glycopeptide Linkages

Most proteins that are secreted or bound to the plasma membrane are modified by carbohydrate attachment. The part that is modified, in plasma membrane–bound proteins, is the extracellular portion. Intracellular proteins are less frequently modified by carbohydrate attachment.

N-glycosidic bonds are the most widely found carbohydrate–protein linkages in animals. The β-glycosylamine linkage of *N*-acetylglucosamine (GlcNAc) to asparagine (Asn) in proteins represents the prominent form and is the site to which further attachment of a variety of complex and polymannose oligosaccharides occurs. The GlcNAc-β-Asn linkage is observed in a vast array of animal proteins including numerous hormones and plasma proteins, immunoglobulins, cell surface receptors, and the lectins. Although others sugars besides GlcNAc can be found attached to asparagine residues in prokaryotes, the only other non-GlcNAc N-glycosidic bond that has been found in mammalian proteins is the β-glucose linkage (Glc-β-Asn) observed in laminin.

O-glycosidic bonds represent the next most commonly occurring sugar-peptide bond found in animal proteins. The variety of proteins in which O-glycosylation is found is much greater than that for the N-glycosylated proteins. All hydroxylated amino acids (Ser, Thr, Hyl, and Hyp) can accept a sugar residue to become O-glycosylated. However, the most commonly occurring O-linkage is to serine and threonine. It should be noted as well that since hydroxyproline and hydroxylysine are prevalent in the collagens, these residues are frequently modified with sugar in the various functional collagen proteins. All vertebrate collagens contain galactose (Gal) attached to hydroxylysine in the Gal-β-Hyl linkage. The α-galactosamine linkage of *N*-acetylgalactosamine (GalNAc) to serine or threonine represents the predominant form of O-glycosylation in mammalian proteins. This linkage is considered the hallmark of the mucins that are glycoproteins and comprise the major portion of mucosal secretions.

The GlcNAc-β-Ser/Thr linkage is unique in that it is present in proteins found in the nucleus and cytoskeleton. Indeed, it

was the first sugar–protein linkage found to exist outside of the ER and Golgi synthesis and secretory machinery and this modification regulates protein function (see Sect. 3.2). Whereas, many other sugars can and are attached to the GalNAc-α-Ser/Thr structure, in the GalNAc-β-Ser/Thr setting, no other carbohydrates have been shown to be linked.

Several proteins of the coagulation cascade, as well as proteins that harbor a conserved domain first identified in epidermal growth factor (EGF), have been shown to be modified by either fucose (Fuc) or glucose (Glc) addition. Coagulation factors VII, IX, and XII each harbor the Fuc-α-Ser/Thr linkage. In these proteins, fucose is attached alone or serves as the branch for the attachment of a short oligosaccharide unit. The coagulation factors VII and IX also contain glucose attached in a β-linkage to serine. The Glc-β-Ser linkage is also found in another regulator of the coagulation cascade, thrombospondin.

Mannose (Man) is prevalently found linked to cell wall proteins in yeast and in bacteria. However, few examples of the Man-Ser/Thr linkage in mammalian proteins exist. The linkage of mannose has been shown in several proteoglycans and glycoproteins of neural origin. In addition, the α-dystroglycans of peripheral nerves have been shown to be O-mannosylated. These observations, limited though they may be, seem to suggest that O-mannosylation is a unique feature of glycopeptides of the nervous system. Mannose attachment through a carbon-carbon linkage to the C-2 of tryptophan (Trp) represents a novel glycopeptide linkage that has been observed in interleukin-12 and ribonuclease 2.

Many membrane-associated eukaryotic proteins are anchored in place by linkage to glycosylphosphatidylinositol (GPI). The components of the GPI linkage (Fig. 3) are quite variable and utilize both glycopeptide and glycophospholipid interactions to tether proteins to the cell surface. Although variable in structure, the GPI linkage does contain a core composed of phosphoethanolamine-6-Man-α1,2-Man-α1,6-Man-α1,4-GlcNAc-α1,6-*myo*-inositol-1-phospholipid implicating a conserved biosynthetic pathway in eukaryotes. The GPI linkage is most abundant in parasitic protozoa; however, several important mammalian proteins have been shown to be cell surface associated via the GPI linkage. These include lymphoid cell

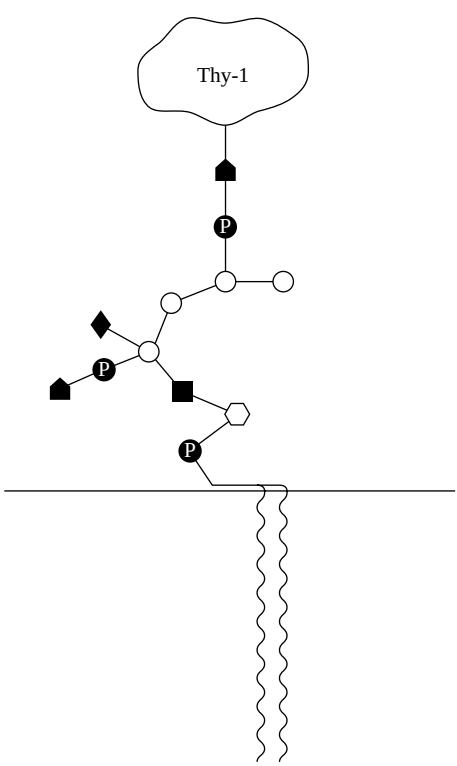

Fig. 3 Structure of the linkage of Thy-1 (CD90) to the GPI anchor. Definitions of symbols used are defined in Fig. 1.

antigens such as thymocyte antigen-1, Thy-1 (also identified as CD90), and lymphocyte function associated antigen-3, LFA-3; cell adhesion molecules such as neural cell adhesion molecule, *N*-CAM; and hydrolytic enzymes such as acetylcholinesterase and 5'-nucleotidase.

Of clinical significance is the erythrocyte GPI-anchored protein, CD55 (also termed *decay accelerating factor*, (DAF). The naming of proteins with the designation CD followed by a number originated from the use of antibodies to cell surface markers to phenotype or sort cells. These cell surface markers were termed *cluster of determination* (CD) *antigens*. Loss of CD55 leads to increased red cell lysis (hemolysis) with a concomitant increase in urinary excretion of hemoglobin. This disorder is termed *paroxysmal nocturnal hemoglobinuria* (PNH).

The most complex peptide-sugar linked molecules in the body are the proteoglycans. Proteoglycans are composed of a core protein (numerous types), a linker protein and glycosaminoglycans (GAGs). GAGs are long unbranched polysaccharides containing a repeating disaccharide unit. The disaccharide units contain either of two modified sugars, GalNAc or GlcNAc, and a uronic acid such as glucuronate or iduronate. The GAGs include hyaluronates, chondroitin sulfates, dermatan sulfates, keratin sulfates, heparin sulfates, and heparins. The core proteins of cartilage are known as aggrecans. Owing to the high negative charge of GAGs and their extended conformation, these molecules impart high viscosity to the environment in which they reside. Proteoglycans are located primarily on the surface of cells or in the extracellular matrix (ECM). The primary GAGs present in proteoglycans are hyaluronates, keratin sulfates and chondroitin sulfates. The GAGs themselves are attached to a carbohydrate linker that is O-glycosylated at serine or threonine residues. Some forms of the keratin sulfates are linked to the protein core through an *N*-asparaginyl bond. The protein cores of proteoglycans are rich in serine and threonine residues that allow for multiple GAG attachments.

There are numerous enzymes involved in the generation of peptide–glycosyl bonds as well as the additional carbohydrate–carbohydrate linkages that occur following formation of a glycopeptide linkage. These enzymes are collectively referred to as glycosyltransferases. At least 16 enzymes have been identified in various organisms that are responsible for the attachment of sugar to protein (8 mammalian enzymes are

Tab. 1 Mammalian glycosyltransferases involved in the synthesis of carbohydrate to protein linkages.

Enzyme	*Donor*	*Linkage*
Oligosaccharyl-transferase	$Glc_3Man_9GlcNAc_2$-PP-Dol	GlcNAc-β-Asn
GalNAc-transferase	UDP-GalNAc	GalNAc-α-Ser/Thr
GlcNAc-transferase	UDP-GlcNAc	GlcNAc-β-Ser/Thr
Fuc-transferase	GDP-fucose	Fuc-α-Ser/Thr
Xyl-transferase	UDP-xylose	Xyl-β-Ser
Gal-transferase	UDP-galactose	Gal-β-Hyl
Man-transferase	GDP-Man	Man-α-Trp
Glycogenin	UDP-glucose	Glc-α-Tyr

listed in Table 1). In almost all situations, the sugar–protein linkage is formed by the direct enzymatic transfer of an activated monosaccharide to a specific amino acid residue. Exceptions to this include the GPI linkage and formation of the GlcNAc-β-Asn linkage (see Sect. 3.2). Oligosaccharides are then generated through the sequential attachment of additional sugars via the activities of various glycosyltransferases.

3 Glycoprotein Synthesis

3.1 Synthesis of N-linked Glycopeptides

The predominant carbohydrate attachment in glycoproteins of mammalian cells is via addition of GlcNAc to asparagine residues forming the N-glycosidic linkage. The site of carbohydrate attachment in N-linked glycoproteins is found within a strict consensus sequence of amino acids, Asn-X-Ser/Thr, where X is any amino acid except proline. In addition, the presence of glutamic acid or tryptophan residues at the X position of the consensus reduces or eliminates glycosylation. When an analysis of proteins in the public databases is carried out, it can be shown that approximately 65% of all the proteins contain at least one occurrence of the Asn-X-Ser/Thr consensus. Of course, the presence of the consensus does not ensure that the native protein will indeed be glycosylated.

The formation of the GlcNAc-β-Asn linkage occurs in the ER through co-translational addition of a preassembled carbohydrate core structure that is delivered via a carbohydrate–dolichol lipid intermediate. The preassembled carbohydrate core structure comprises three terminal residues of glucose attached to a branched cluster of nine mannose residues that are in turn attached to two GlcNAc residues attached to the isoprenoid-like lipid, dolicholpyrophosphate. The structure is abbreviated $Glc_3Man_9GlcNAc_2$-P-P-dolichol. This structure is commonly referred to as the *lipid-linked oligosaccharide* (LLO), whereas the oligosaccharide structure itself ($Glc_3Man_9GlcNAc_2$) is termed the *en bloc* oligosaccharide. In mammalian cells, the importance of the terminal glucose residues is evident from the fact that transfer of $Man_9GlcNAc_2$-P-P-dolichol is some 25 times less efficient than the complete structure. In addition, structures that contain three terminal glucose residues, but not the complete $Man_9GlcNAc_2$ structure, are efficiently transferred to protein by oligosaccharyl-transferase.

Synthesis of the *en bloc* dolichol-P-P-oligosaccharide unit begins on the cytoplasmic face of the ER membrane and prior to transfer to the protein, the structure "flips" to the luminal side (Fig. 4). Sugars used for glycoprotein synthesis (both N-linked and O-linked) are activated by coupling to nucleotides. Glucose and GlcNAc are coupled to uridine diphosphate (UDP), and mannose is coupled to guanosine diphosphate (GDP). The first reaction in dolichol-P-P-oligosaccharide synthesis involves the formation of GlcNAc-P-P-dolichol with the release of uridine monophosphate (UMP) from the nucleotide-activated sugar, UDP-GlcNAc. The second GlcNAc and the Man-transferase reactions proceed via sugar transfer from the nucleotide-activated sugar directly to GlcNAc-P-P-dolichol. From this point, additional mannose residues are added to the $Man_1GlcNAc_2$-P-P-dolichol by transfer from Man-P-dolichol, which itself is formed from dolichol phosphate and GDP-Man.

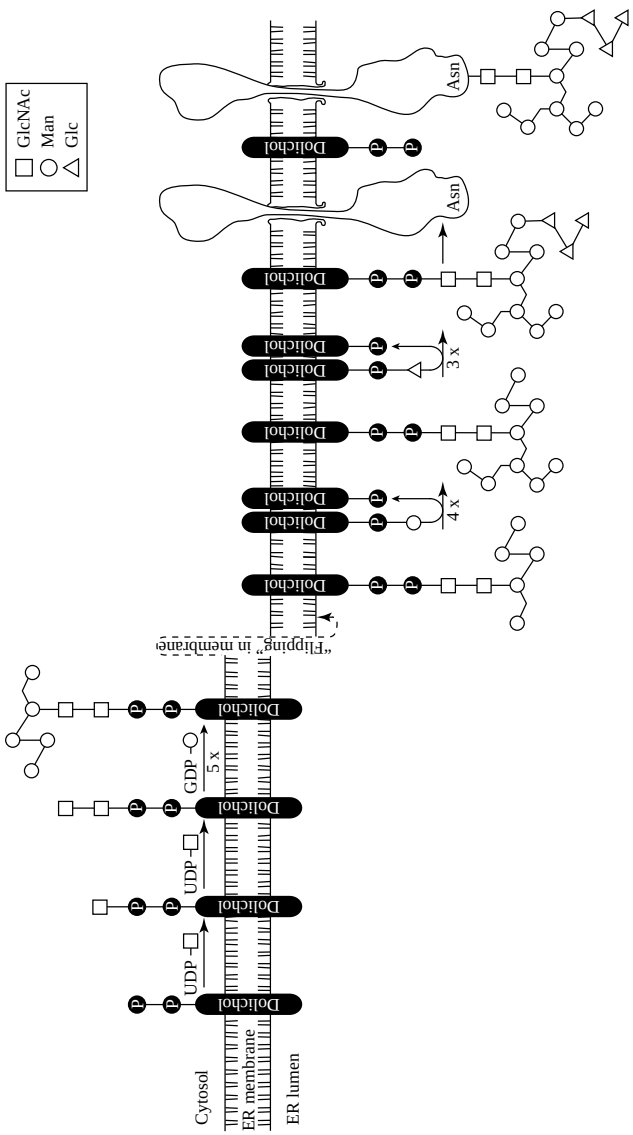

Fig. 4 Pathway for the synthesis and transfer of the *en bloc* oligosaccharide structure during N-glycosylation in the ER.

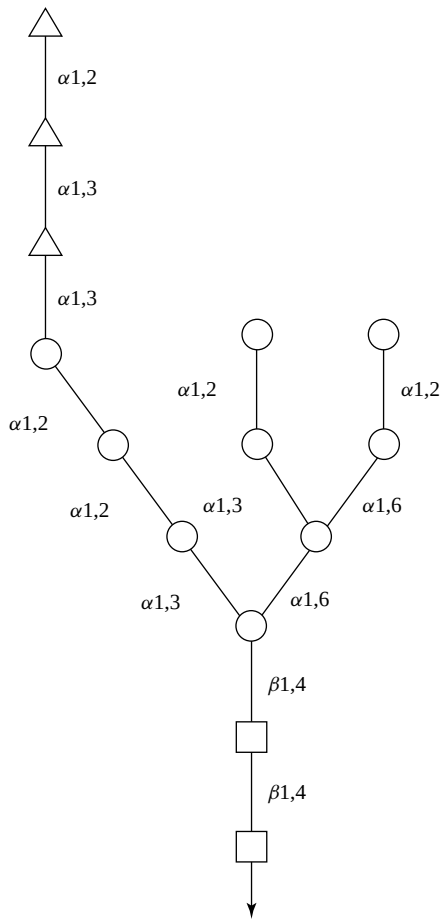

Fig. 5 Structure of the *en bloc* oligossacharide (also referred to as the *lipid-linked oligosaccharide, LLO*) attached in all N-linked glycosylation reactions. Refer to Fig. 1 for the identification of symbols used for sugar residues.

Immediately following transfer of the *en bloc* oligosaccharide (Fig. 5) unit to the protein, processing and alteration of the composition of the oligosaccharide ensues and continues as the protein passes through the ER then into and through the Golgi apparatus. Initially, the terminal glucose is removed through the action of glucosidase I (GI), a membrane-bound enzyme recognizing α1,2-linked glucose. The remaining two glucose residues are then removed by glucosidase II (GII), a soluble enzyme recognizing α1,3-inked glucose. After removal of the glucose residues, the action of α-mannosidases removes several mannose residues as the protein progresses to the Golgi.

The action of the various glucosidases and mannosidases leaves N-linked glycoproteins containing a common core of carbohydrate consisting of three mannose residues and two GlcNAc. Through the action of a wide range of glycosyltransferases and glycosidases, a variety of other sugars are attached to this core as the protein progresses through the Golgi. These latter reactions generate the three major families of N-linked glycoproteins: high-mannose type, complex type, and hybrid type (Fig. 6). High-mannose type contains varying amounts of mannose residues in addition to those contained in the core. Hybrid type contains various sugars and amino sugars. The complex type is similar to the hybrid type, but in addition, contains sialic acids in varying amounts.

3.2 Synthesis of O-linked Glycopeptides

The synthesis of O-linked glycoproteins occurs via the stepwise addition of nucleotide-activated sugars directly onto the polypeptide as opposed to the addition of *en bloc* synthesized oligosaccharide to asparagine residues in N-linked glycoproteins. The addition of sugar through linkage to serine or threonine residues occurs via the action of glycosyltransferases similar to those used in the synthesis of the N-linked glycoproteins. Attachment of sugars in O-linked glycoproteins

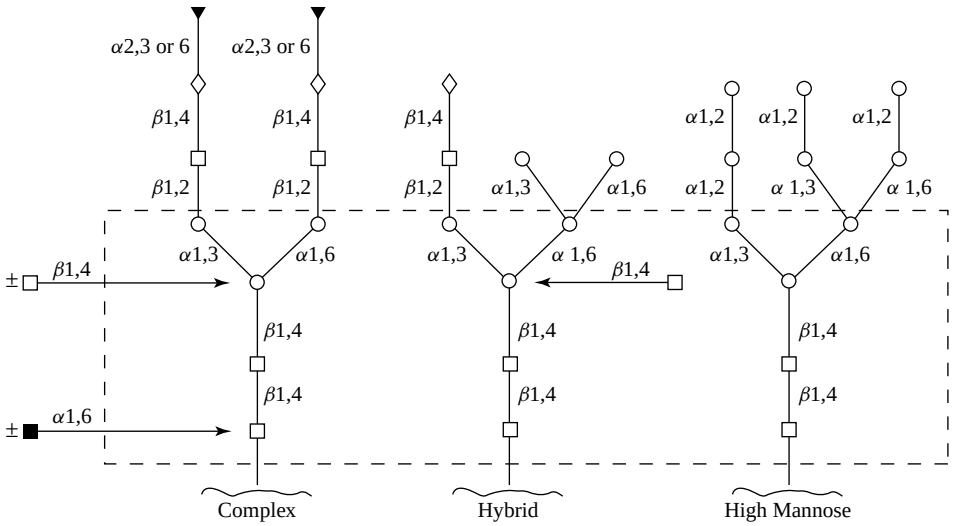

Fig. 6 Structures of the three main classes of N-glycans. Class is identified below each structure. Refer to Fig. 1 for identification of the symbols used for sugar residues. The core pentasaccharide structure common to all N-glycans is enclosed in the dashed rectangle.

occurs posttranslationally as the proteins pass through the Golgi. As for the sugars attached to N-glycans, O-linked sugars are nucleotide-activated prior to addition.

As discussed above, there are a wide variety of sugar linkages to amino acids harboring a hydroxyl moiety. However, the most commonly occurring linkage in mammalian O-glycans is GalNAc-α-Ser/Thr. There are at least nine GalNAc-transferases that have been identified and all are generally homologous to each other. Despite intensive investigation, no clear consensus sequence has been identified that directs the incorporation of GalNAc onto serine or threonine residues. A common theme does exist in that a preference for incorporation occurs near the amino terminus where there are clusters of serine/threonine residues within a β-turn and at a distance from charged amino acids. In addition, threonine residues appear to be the preferred substrates for GalNAc-transferases.

The formation of the GlcNAc-β-Ser/Thr linkage, catalyzed by *O*-GlcNAc transferase (OGT), is unique in several respects. First, this O-linkage is observed to occur in proteins that reside within the cytosol or the nucleus. Secondly, the enzyme glycosylates serine/threonine residues that are also sites for phosphorylation indicating that there may be some type of relationship or cyclical regulation between these two distinct types of protein modifications. Many transcription factors and RNA polymerase II are modified by *O*-GlcNAc linkage and when so modified the activity of these factors is repressed. More recently it was shown that OGT is recruited to transcriptional complexes by the transcriptional corepressor Sin3. The resultant *O*-GlcNAc modification of transcription factors in the complex leads to gene silencing.

4
Role of N-glycosylation in Protein Folding

Proteins that are destined to enter the secretory pathway of the cell must first attain a proper tertiary or quaternary structure. Failure to properly fold or to form the correct multi-subunit structure results in those peptides being retained in the ER and leads to their eventual degradation. As described above, following the addition of the *en bloc* synthesized oligosaccharide ($Glc_3Man_9GlcNAc_2$) to asparagine residues, the structure is the target for glucosidase (GI and GII) action. Removal of glucose from the *en bloc* oligosaccharide after it is attached to protein is a signal to the protein sorting apparatus that the protein is appropriately folded. In an apparent competition with the action of the glucosidases is the action of the glucosyltransferase, UDP-Glc:glycoprotein glucosyltransferase (GT). This enzyme adds glucose residues to the terminal mannose of $Man_{7-9}GlcNAc_2$ structures (i.e. the products of GI and GII activity as well as α-mannosidases). Quite strikingly, the activity of GT is higher toward misfolded glycoproteins than it is toward properly folded ones.

Many enzymes involved in protein folding, as well as several molecular chaperones, have been identified within the ER. Of particular import are the chaperones calnexin and calreticulin. Calnexin and calreticulin are homologous proteins with calnexin being membrane bound and calreticulin being soluble. Both of these chaperones recognize and bind to glycoproteins in the ER and are thus members of the larger class of proteins that recognize carbohydrate moieties termed *lectins*. Calnexin and calreticulin specifically bind to monoglucosylated oligosaccharides where the ideal substrate is $Glc_1Man_9GlcNAc_2$. After chaperone binding, the action of GII will lead to the liberation of calnexin and calreticulin. The oligosaccharide can then be reglucosylated by GT allowing the glycoprotein to once again be recognized by the lectin chaperones. This cyclic process of deglucosylation by GII followed by glucosylation by GT will continue until the protein achieves the proper folded conformation. Although important, it should be noted that the calnexin/calreticulin glycoprotein interaction system is not the only mechanism by which cells recognize and dispose of misfolded glycoproteins in the ER. The principal site for the degradation of misfolded soluble and membrane-associated glycoproteins is the proteasome.

5
Lysosomal Targeting by Specific Glycosylation

Enzymes that are destined for the lysosomes (lysosomal enzymes) are directed there by a specific carbohydrate modification. During transit through the Golgi apparatus a residue of GlcNAc-1-phosphate (GlcNAc-1-P) is added to the carbon-6 hydroxyl group of one or more specific mannose residues that have been added to these enzymes. The GlcNAc is activated by coupling to UDP and is transferred by UDP-GlcNAc:lysosomal enzyme GlcNAc-1-phosphotransferase (GlcNAc-phosphotransferase), yielding a phosphodiester intermediate: GlcNAc-1-P-6-Man-protein. A second reaction (catalyzed by GlcNAc 1-phosphodiester-*N*-acetylglucosaminidase) removes the GlcNAc leaving mannose residues phosphorylated in the 6 position: Man-6-P-protein. A specific Man-6-P receptor (MPR) is present in the membranes of the Golgi apparatus. Binding of

Man-6-P to this receptor targets proteins to the lysosomes.

Two distinct MPRs have been identified and both are members of the P-type lectin family. Both are type I integral membrane glycoproteins that contain an N-terminal extracellular domain, a single transmembrane domain, and a C-terminal cytoplasmic domain. One receptor is large with a molecular weight of approximately 300 kDa, the other receptor is smaller with a molecular weight of approximately 46 kDa. Structural similarities between these two receptors indicate that they are derived from a single ancestral gene with the larger receptor arising through multiple gene duplications. The extracellular portion of the larger receptor contains 15 repeating elements, each of which is highly similar to the extracellular domain of the smaller receptor. Both receptors exist as dimers embedded in the membrane.

The large receptor binds two moles of Man-6-P and the smaller binds 1 mol of Man-6-P per subunit, thus binding 4 and 2 mol of Man-6-P per dimer, respectively. The bovine and murine versions of the smaller receptors require divalent cations for ligand binding and thus the receptor has been termed the *cation-dependent Man-6-P receptor* (CD-MPR). However, the human counterpart may not require cations for its activity. The larger receptor does not require divalent cations for ligand binding and is therefore, commonly referred to as the *cation-independent Man-6-P receptor* (CI-MPR). However, the CI-MPR has been shown to bind the nonglycosylated polypeptide hormone, insulin-like growth factor II (IGF-II) and as such the larger MPR is more frequently identified as IGF-II/MPR. The IGF-II/MPR is available at the cell surface and its role in binding IGF-II is to target this hormone for degradation in the lysosomes. In addition to IGF-II, the IGF-II/MPR has been shown to bind a diverse array of Man-6-P-containing proteins as well as several nonglucosylated proteins. Although IGF-II/MPR and CD-MPR exhibit distinct activities, both receptors function to target newly synthesized lysosomal enzymes to the lysosomes.

The clinical significance of mannose phosphorylation for the proper targeting of enzymes to the lysosomes is evident from the identification of the molecular defect leading to the development of I-cell disease (mucolipidosis II, ML-II) and pseudo-Hurler polydystrophy (mucolipidosis III, ML-III). The characteristic physiological consequences of I-cell disease are severe psychomotor retardation and craniofacial abnormalities. The disease acquired its name due to the presence of large phase-dense inclusion bodies in the fibroblasts of afflicted individuals. Inheritance of defective GlcNAc-phosphotransferase, the first of the two enzymes required for Man-6-P synthesis, results in I-cell disease and ML-III.

6 Glycoproteins: Recognition and Homeostasis

6.1 Animal Lectins

The lectins are proteins that recognize and bind to specific types of carbohydrate structures displayed on other proteins and/or lipids. This class of protein does not, however, include the immunoglobulins, which are by themselves a specialized class of proteins that recognize carbohydrate. The basic definition of an animal lectin, therefore, describes nonimmunoglobulins that

are capable of differential agglutinization of erythrocytes. However, the lectins do not display enzymatic activity toward their respective substrates.

The contexts in which carbohydrate recognition occurs are quite diverse and as such, the lectins represent a broad group of proteins consisting of many unrelated families (Table 2). As the term lectin refers to proteins that can bind to carbohydrate, they all contain at least one primary carbohydrate recognition domain (CRD). Lectins often contain additional CRDs that allow for an increase in binding affinities of the primary CRD. In addition, several types of lectins attain carbohydrate-binding capabilities through oligomerization generating multivalent complexes.

Since carbohydrate recognition can be observed in a wide variety of biological contexts, the lectins constitute an extremely diverse group of proteins, many of which have no activity relationships other than the ability to bind carbohydrate. It is clear, therefore, that the biological activities associated with lectin binding are many (Table 3) and extremely important for numerous aspects of cellular homeostasis. Perhaps, the most important homeostatic function of the animal lectins is their role in the regulation of immune responses and autoimmunity. This latter fact will be the focus of the following review of several of the lectin families.

6.1.1 C-type Lectins

C-type lectins were initially defined as those proteins having a CRD of approximately 120 amino acids with a dependence on Ca^{2+} ions for binding activity. The dependence on Ca^{2+} for carbohydrate binding is demonstrated by the fact that

Tab. 2 Lectin families and their carbohydrate specificities.

Family	*Characteristics*	*Specificities*
C-type lectins (7 subfamilies)	Require Ca^{2+} for activity	Variable
Collectins (C-type subfamily III)	C-type CRD with collagen-like domains	Variable, mannose
Galectins (S-type lectins)	Sulfydryl-dependent or β-galactosidase binding	Strict for β-galactosides
Siglecs (I-type lectins)	Contain immunoglobulin-like domains	Sialic acid
Phosphomannosyl receptors (P-type lectins)	Recognize mannose-6-phosphate	Mannose-6-phosphate
Pentraxins	Quaternary structure composed of 5 identical polypeptides that form a ring with a central hole	Variable and can be noncarbohydrate
Calnexin and calreticulin	Recognition of properly folded glycoproteins in the ER	Glucose
Ficolins	Contain fibrinogen-like homology domain	Variable
Tachylectins	Distinct CRD but also contain fibrinogen-like domain	GlcNAc/GalNAc
ERGIC-53	Legume lectin-like fold	Mannose

Tab. 3 Functions associated with animal lectins discussed in the chapter.

Functions	*Lectin examples*
Cell–matrix interactions	Galectins
Cell–cell interactions	Selectins, galectins, I-type lectins
Cell migration, cell routing	Selectins, galectins, I-type lectins
Molecular chaperones in endoplasmic reticulum during glycoprotein synthesis	Calnexin, calreticulin
Recognition of foreign glycans, self/nonself recognition	Collectins, pentraxins
Intracellular trafficking of glycoprotein	P-type lectins
Regulation of cytokine release	Galectins, selectins
Anti-inflammatory actions	Galectins
Scavenging debris	Pentraxins

loss of the ion results in a conformational change in the CRD and subsequent loss of carbohydrate binding. The C-type lectin family can be further subdivided into at least seven subfamilies dependent upon the nature of other noncarbohydrate recognition domains as well as overall gene structure. C-type lectins recognize carbohydrate structures on many different forms of protein, including cell surface proteins and extracellular matrix (ECM) proteins and other ECM structures containing carbohydrate modifications such as proteoglycans, glycosaminoglycans, and glycolipids. C-type lectins also exhibit broad specificity in the types of ligands they bind including Fuc, Gal, Glc, GlcNAc, maltose, and Man and *N*-acetylmannosamine (ManNAc). One of the most important functions for carbohydrate recognition by C-type lectins is the role that function plays in pathogen recognition by the innate immune system.

The collectins are subtype III C-type lectins defined by having an additional collagen-like domain. The presence of the collagen-like domain allows the collectins to form a triple helical structure that creates a multivalent CRD, which lies at the carboxy-terminus of the protein. Mannan-binding lectin, MBL (also termed *mannose-binding protein*) is the best characterized human collectin. Although termed mannose-binding lectin, MBL actually exhibits higher affinity for GlcNAc than Man. Four additional collectins have been well characterized including lung surfactant proteins A (SP-A) and D (SP-D), collectin-43, and bovine conglutinin. The major function of the collectins is recognition of carbohydrate structures on microbial surfaces leading to activation of the complement cascade and phagocytosis.

The complement system is a host-defense mechanism that is engaged by both the innate and antibody-mediated (acquired) immune systems. Three pathways lead to complement activation; classic, alternative, and lectin. The classic pathway is activated by antibody–antigen complexes reflective of the function of complement in acquired immunity, whereas, the other two function in innate immunity. More than 30 plasma and membrane-bound proteins (identified by the letter C followed by a number) participate in the complement pathway. The classic pathway requires activation of C1q (a subcomponent of C1), whereas the lectin pathway involves activation of C2 and C4. MBL is synthesized by

the liver and secreted into the circulation. It is considered one of the acute-phase proteins synthesized by the liver. Expression levels of MBL rise in response to inflammatory signals. Upon binding to carbohydrate structures on microbial surfaces, MBL activates the complement system in a C4/C2-dependent manner, independent of C1q. The mechanism of MBL activation of complement requires the interaction of MBL with a family of serine proteases termed *mannose-binding lectin-associated serine proteases* (MASPs). The action of MBL-MASP in the lectin-activated complement cascade is similar to the C1q-antibody complex that activates the serine proteases C1r and C1s.

The selectins are subtype IV C-type lectins defined by the presence of the C-type lectin CRD, an epidermal growth factor–like domain and a variable number of complement regulatory protein-like repeat domains. The selectins are also members of the cell adhesion family of proteins. Each selectin is tethered to the plasma membrane via a single transmembrane domain. There are three characterized selectins that were originally identified as the adhesion molecules: endothelial–leukocyte adhesion molecule-1 (ELAM-1), murine lymph node homing receptor, and platelet granule membrane protein-140. These three proteins are now referred to as E-selectin, L-selectin, and P-selectin respectively.

A major function of the selectins is the recruitment of neutrophils to sites of inflammation. P-selectin is constitutively stored in secretory granules of platelets and endothelial cells. When stimulated by the appropriate response (e.g. IL-1 and TNF-α) these cells present P-selectin to the cell surface via fusion of the intracellular vesicles with the plasma membrane. Through this mechanism, P-selectin is involved in mediating very early responses of leukocytes to inflammatory signals. E-selectin is also synthesized by endothelial cells but is not constitutively present. Expression of E-selectin is regulated by cytokines such as IL-1 and TNF-α. L-selectin is expressed by most leukocytes and plays a role in the recruitment of these cell types into lymphatic tissue. Constitutively expressed L-selectin is rapidly released from the surface of activated leukocytes. Activation of L-selectin signaling, via ligand-induced cross-linking, results in potentiation of the responses of neutrophils to other stimuli. These effects are mediated by intracellular activation of both tyrosine and serine/threonine phosphorylations, the latter mediated through MAP kinases.

6.1.2 S-type Lectins: Galectins

The S-type lectins were originally so designated because the original family members exhibited a sulfhydryl-dependence on carbohydrate binding. The S-type lectins are now commonly referred to as the *galectins*. The galectins are all defined by a canonical CRD having an affinity for β-galactosides. To date, 14 galectins have been identified in mammals. The basic carbohydrate structures recognized by the galectins are an N-acetylated lactose (*N*-acetyllactosamine, LacNAc with the structure Galβ1,4GlcNAc). The galectins can be further divided into three subfamilies based upon their structures. The prototype subfamily (galectin-1, -2, -5, -7, -10, -11, -13, and -14) is comprised of a single CRD, the tandem-repeat type (galectin-4, -6, -8, -9, and -12) contains two homologous CRDs, and the chimeric type contains a single CRD combined with a unique amino or carboxy-terminus. The only mammalian chimeric type galectin thus far identified is galectin-3. The prototype galectins can

form noncovalent homodimers resulting in a functionally bivalent molecule, typified by mammalian galectin-1. Owing to their multivalent nature, galectins are capable of cross-linking glycoconjugates containing LacNAc structures. It is the effects of cross-linking that result in the initiation of biological responses to galectins. Galectin activities are observed both at the cell surface and intracellularly. All galectins are synthesized on soluble ribosomes and following synthesis, are targeted to specific locations including intracellular compartments and membranes or membrane-bound vesicles. The exact mechanisms by which galectins are secreted from cells remain unknown as none are synthesized with a canonical secretory signal sequence.

One of the most crucial functions of the galectins is in regulation of the inflammatory response. Galectin-1 is expressed by antigen-stimulated T cells, activated B cells, and by macrophages. The action of galectin-1 in these systems is to stimulate apoptosis resulting in cell death. Negative growth effects of galectin-1 are also seen in breast and prostate cancer cell lines. In addition to galectin-1, galectin-7, -8, -9, and -12 have been shown to exhibit proapoptotic activity. In contrast, galectin-3 exhibits antiapoptotic activities in similar types of cells. Whereas the proapoptotic effects of galectin-1 can be observed by adding the protein to activated T cells, the antiapoptotic effects of galectin-3 appears to be the result of intracellular activities associated with the protein. Interestingly, galectin-3 has some striking sequence similarities to the antiapoptotic Bcl-2 protein family.

Galectin-1 is a potent inhibitor of activated T-cell-mediated inflammatory responses. Using experimentally induced models of several autoimmune disorders (e.g. myasthenia gravis, multiple sclerosis, and rheumatoid arthritis) it has been shown that administration of recombinant galectin-1 can significantly ameliorate the resultant clinical symptoms. The potency of galectin-1 in these studies suggests that future advances will see clinical application of galectin-1 for the treatment of activated T-cell-mediated autoimmune diseases. Conversely, the presence of autoantibodies to galectin-1 has been identified in several neurodegenerative disorders. Likewise, the levels of autoantibodies to galectin-3 are inversely associated with severity of Crohn's disease (inflammatory bowel disease, IBD).

Expression of galectins and/or galectin-binding sites has also been correlated with various types of cancers as well as their severities. Elevated expression of galectin-1 has been shown to be correlated with the malignant potential of various cancers. Equally telling is the observation that survival potential from gliomas declines as the level of galectin-1 protein rises. Galectin-3 expression levels have also been shown to be elevated in gastric and liver cancers. In contrast, expression of galectin-8 declines during tumiorogenic progression.

It is clear from the data that the galectins exhibit activities ranging from cell adhesion, cell migration, proliferation, and differentiation, to roles in apoptosis and cancer. Emerging data indicates that galectins or galectin antagonists will find therapeutic use as anti-inflammatory mediators, immunosuppressants, or anti-cancer agents.

6.1.3 I-type Lectins: Siglecs

I-type lectin is the term collectively given to the family of carbohydrate-binding proteins that belong to the immunoglobulin (Ig) superfamily of proteins. The Ig superfamily of proteins all have at least

one domain composed of two β-sheets that are disulfide bonded and stacked together. There are at least three subtypes of Ig domain (V-set, C1-set, and C2-set) that are defined by the number and arrangement of the β-strands in the domain.

All of the I-type lectins are membrane anchored with their CRDs on the outside of the cell. The I-type lectins are divided into two main subfamilies. One subfamily includes the cell adhesion molecules, L1, *N*-CAM, and intercellular adhesion molecule-1 (ICAM-1) as well as the cell surface glycoproteins CD83 and CD2. The other subfamily is composed of the siglecs, which are so named because of their binding specificity for sialic acid residues. Siglecs are also characterized by the presence of a characteristic V-set Ig domain at their amino termini followed by varying numbers of C2-set domains.

The siglecs are primarily expressed on cells of the hematopoietic system. A macrophage lectin-like adhesion molecule termed *sialoadhesin* and CD22, and another member of the Ig superfamily that shows restricted B-cell expression were the first two members of the siglec group to be characterized. Both sialoadhesin and CD22 were shown to mediate cell–cell interactions through their abilities to recognize sialylated glycoconjugates. Sialoadhesin is now referred to as siglec-1 and CD22 as siglec-2 in reference to their being the first and second member characterized respectively. Although structurally different from the siglecs, the I-type lectins CD83 and cell adhesion molecule L1 also have been shown to bind to sialic acid residues. To date, at least 11 siglecs have been characterized. Siglec-3 was originally defined as CD33, a marker for cells of the myelomonocytic lineage. Several siglecs have been identified that are structurally very similar to siglec-3 and as such, they have all been classified as members of the CD33/siglec-3 subfamily of siglecs. This subfamily includes siglec-3, -5, -6, -7, -8, -9, -10 and -11.

Like the pentraxins (see Sect. 6.1.4) the siglecs, principally the siglec-3 family, are involved in regulatory aspects of innate immunity reactions. All of the siglec-3 family members contain a motif in their intracellular tails termed an *immunoreceptor tyrosine-based inhibitory motif* (ITIM) indicating that these proteins act as inhibitory receptors on hematopoietic cells. Associated with the ITIM is a tyrosine residue that becomes phosphorylated in response to activation of activating receptors. The function of the ITIM is to recruit phosphatases to activated receptor complexes, thereby deactivating the activating receptors by phosphate removal. Siglec-3 has been shown to interact with two inhibitory tyrosine phosphatases, SHP-1 and SHP-2, and this interaction requires the presence of phosphotyrosine and the ITIM. The presence of phosphotyrosine in the cytoplasmic tail of siglec-3 also enhances the subsequent interaction with sialylated ligands suggesting that tyrosine phosphorylation plays a regulatory role in ligand binding to siglec-3.

6.1.4 Pentraxins

The pentraxin family of lectins is defined by having primary sequence similarities to major acute-phase reactants, C-reactive protein (CRP) and serum amyloid P component (SAP). The pentraxins assume a characteristic multi-subunit complex consisting of 5 to 10 identical subunits that form a pentameric ring structure with a central open core, hence the term pentraxin. The original identity applied to CRP stems from the observation that it was an antibody-like molecule associated with the C polysaccharide component

of the *pneumococcal* bacterium. CRP was originally shown to be a complement-activating precipitin with a high affinity for phosphocholine. Although originally not identified as a carbohydrate-binding protein, CRP was found to exhibit affinity for galactans (galactose-containing oligosaccharides) and phosphogalactose using a binding site distinct from that involved in its precipitin activity. Similarly, SAP not only has carbohydrate-binding specificity it also has an affinity for phosphoethanolamine. In general, the pentraxins represent a class of molecules that exhibit calcium-dependent binding to a number of substances bearing phosphate esters and carbohydrates. Classes of pentraxins have been divided into those with binding similarities to CRP (i.e. preference for phosphocholines) or SAP (i.e. preference for phosphoethanolamines).

Another defining feature associated with pentraxins is the observation that their expression is markedly increased as a consequence of the activities of immunomodulatory molecules such as IL-1 and TNF-α as well as in response to tissue injury and necrosis. Both IL-1 and IL-6 can dramatically induce the acute-phase response of the liver leading to increased CRP synthesis and secretion. That these factors are important in the induction of CRP synthesis is demonstrated by the fact that IL-1 receptor antagonists or anti-IL-6 antibodies can both block CRP synthesis in cultured hepatic cells. Reciprocally, addition of CRP to human alveolar macrophages in culture induces the synthesis of IL-1 and TNF-α suggesting a feedback amplification of the response to proinflammatory cytokines.

A role for the pentraxins in innate immunity functions is suggested by their reactivities with the complement system as well as with phagocytic leukocytes. CRP is consistently shown to activate the classic complement pathway and binds to the site of complement-induced cell injury. In this system, it has been shown that CRP interacts with complement protein C1q and SAP with C4-binding protein, C4bp.

6.2 Lectin Activities of the Cytokines

Cytokines are a family of soluble proteins that function in both immune- and non-immune systems to mediate cell–cell communication. The classic model of cytokine action is that these proteins bind to cell surface receptors leading to the activation of intracellular signaling cascades. Recent evidence indicates that inhibitors of glycosylation in turn inhibit cytokine-receptor interactions and subsequent signal transduction. Many cytokines have been shown to have specific lectin activity including IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, and IL-12 as well as TNF-α. As these cytokines do not, as yet, belong to one of the well-defined lectin families, they have been identified as orphan lectins.

Discussion of the lectin-like activities of IL-1, IL-2, and IL-6 serve as interesting examples that includes carbohydrate recognition, receptor activation, and regulation of immune system function. IL-1 constitutes a family of proinflammatory proteins consisting of IL-1α, IL-1β, and IL-1 receptor antagonist, IL-1RA. The carbohydrate structures recognized by IL-1α and IL-1β are quite distinct. IL-1α recognizes a very rare disialylated di-antennary N-glycan whereas, IL-1β interacts with the carbohydrate structures found on neural cell glycolipids of the GM4 type. IL-2 is produced by activated T cells and exhibits biological activities that include additional stimulation of T cells, activation of B cells, and activation of other leukocytes that harbor IL-2 receptors on their surfaces. The

mechanism of IL-2-mediated activation of signal transduction events involves two distinct interactions. One interaction involves binding to the IL-2 receptor itself. The other interaction has been shown to require the presence of N-glycans containing five or six mannose residues with a structure $Man_{5-6}GlcNAc_2$ on the CD3 molecule associated with the T-cell receptor. Interaction of IL-2 with the glycan moiety of CD3 does not require calcium, thus, IL-2 is not a member of the C-type lectin family. IL-6, which functions in hematopoiesis, immunity, and inflammation, exhibits bifunctional binding characteristics like that of IL-2. IL-6 binds to its receptor, which in turn leads to association with an additional protein identified as gp130 as it is a glycoprotein of 130 kDa. As a lectin, IL-6 binds to both O- and N-glycans that are sulfated. Highest affinity is to the HNK-1 epitope (sulfo-GlcNAcGal), which is the terminal glycosylation of a glycoprotein, first identified on the surface of eggs of the frog *Rana temporaria*. Recently, it was demonstrated that the carbohydrate structure on gp130, to which IL-6 binds, is an HNK-1 epitope.

6.3 Sugar Alterations that Affect Immunity

Well over 200 proteins that are expressed on the surfaces of leukocytes are known to be glycosylated. In addition, many proteins involved in the processes of signal transduction that are responsible for the activation and differentiation of B and T cells are O- and N-glycosylated. Adding to the complexity of all these glycosylated proteins is the fact that many individual leukocyte proteins are additionally linked to glycosaminoglycan carbohydrate structures. Because the carbohydrate structures present on leukocyte membrane proteins are nearly equivalent in size to the typical immunoglobulin domain, these structures themselves have the capability to regulate the properties of the cell surface proteins.

As discussed above, animal lectins play critical roles in immunity and disease. The effects of some lectins are inhibitory to cell growth, while other effects promote growth and indeed can even promote tumorigenesis. It would, therefore, seem appropriate for cells to devise mechanisms to regulate the activities of lectins. Many cells activate the expression of lectins in response to a variety of stimuli. However, cellular responses to lectins can also be effected by regulating the events of cell surface presentation of the glycan ligands recognized by the lectins. This latter phenomenon has been best characterized in the cell of the immune system. One of the most important carbohydrate modifications, with respect to leukocyte function, is sialylation. Many of the ligand-receptor interactions discussed above are regulated by the presence or absence of sialic acid residues. The siglecs and selectins require sialylation of their respective ligands in order to bind. In contrast, the binding of many galectins is inhibited by the presence of sialic acid.

Many changes in glycosylation occur during the course of T-cell differentiation as a consequence of changes in the expression of numerous glycosyltransferases and glycosidases. One important glycosylation related to T-cells differentiation is an increase in the level of sialylation. Differences in sialylated structures also distinguish different subsets of T cells from one another. When activated T cells convert to memory T cells, the level of sialylation decreases. Similarly, $CD8^+$ T cells have much less sialic acid on their surfaces than do $CD4^+$ cells. As expected from these observations, the differentiation of T

cells leads to changes in the expression of glycosyltransferases and glycosidases. The glycosidase, sialidase (Neu-1) is induced following stimulation of T cells, whereas, the O-glycan sialyltransferase, ST3Gal-I is induced as T cells mature.

A clear demonstration of the necessity for specific types of glycosylation in the control of immune responses stems from studies with transgenic "knockout" mice. Murine *N*-acetylglucosaminyltransferase V (MGAT5) catalyzes the addition of *N*-acetylglucosamine to α6-linked mannose residues of tri- and tetraantennary branched N-glycans. Mice lacking Mgat5 exhibit an exaggerated cellular-immune response to antigen presentation and develop glomerulonephritis, which is characteristic of autoimmune syndromes. These effects are the result of an increase in T-cell receptor (TCR) clustering on the cell surface. The increased clustering leads to a reduced requirement for ligand interaction as evidenced *in vitro* by enhanced TCR signaling with reduced amounts of anti-CD3 antibodies, relative to control cells. It is likely that the *in vivo* response is due to enhanced recruitment of the TCR to antigen-MHC complexes. An additional experiment is aimed at discerning the function of a specific glycosidase, α-mannosidase II (MII) in immune responses. MII is required to remove terminal mannose residues during the conversion of high-mannose type glycans to the complex type. Mice deficient in this enzyme were shown to develop dyserythropoietic anemia (a deficiency in cells of the erythroid lineage of hematopoietic cells), an autoimmune disorder characteristic of immune complex glomerulonephritis (similar to that observed in MGAT5$^{-/-}$ mice) and have antinuclear antibodies similar to those observed in persons with systemic lupus erythematosus.

7 Clinical Significances of Glycoproteins

7.1 Degradation Defects of GlcNAc-Asn Glycoproteins

The proper degradation of glycoproteins has important medical relevance. Degradation of glycoproteins occurs within the lysosomes and requires specific lysosomal hydrolases of the glycosidase family. Exoglycosidases remove sugars sequentially from the nonreducing end and exhibit restricted substrate specificities. In contrast, endoglycosidases cleave carbohydrate linkages from within and exhibit broader substrate specificities.

Several inherited disorders involving the abnormal storage of glycoprotein degradation products have been identified in humans (Table 4). These disorders result from defects in the genes encoding specific glycosidases, leading to incomplete degradation and subsequent overaccumulation of partially degraded glycoproteins. As a general class, such disorders are known as lysosomal storage diseases and include the diseases known as mucolipidoses that result from incomplete degradation of the carbohydrate portions of glycolipids whose structures are the same as those found on glycoproteins.

Aspartylglucosaminuria (AGU) results from a deficiency in the lysosomal glycosidase, *N*-aspartyl-β-glycosaminidase (AGA) that is responsible for hydrolyzing the GlcNAc-β-Asn bond of glycoproteins. Glycoproteins are completely degraded within the lysosome prior to GlcNAC-β-Asn attack by AGA. The enzyme catalyzes bond hydrolysis independent of the complexity of the associated carbohydrate structure. Owing to the defect, glycoasparagines, primarily *N*-acetylglucosaminyl-asparagine,

Tab. 4 Disorders of glycoprotein degradation.

Disease	*Enzyme defect*	*Gene name*	*Location*
Aspartylglycosaminuria	*N*-aspartyl-β-glycosaminidase	*AGA*	4q32-q33
α-Mannosidosis	α-mannosidase	*MANB*	19p13.3-q12
β-Mannosidosis	β-mannosidase	*MANBA*	4q22-q25
α-Fucosidosis	α-fucosidase	*FUCA*	1p34
Galacotosialidosis	Neuraminidase and β-galactosidase-2 deficiency secondary to protective protein/cathepsin A (PPCA) defects	*PPCA (PPGB)*	20q13.1
Sialidosis (Mucolipidosis I)	Neuraminidase (sialidase)	*NEU1*	6p21.3
Schindler disease	α-*N*-acetylgalactosaminidase	*NAGA*	22q11
G_{M1} Gangliosidosis	β-galactosidase-1	*GLB1*	3p21.33
G_{M2} Gangliosidosis (Sandhoff disease)	hexosaminidase B	*HEXB*	5q13
Mucolipidosis II: I-cell disease and Mucolipidosis III Pseudo-Hurler	*N*-acetylglucosamine-1-phosphotransferase	*GNPTA*	4q21-q23

accumulate in lysosomes leading to cell death. The defect results in a slow progressive global delay in psychomotor development in AGU patients. A delay in speech, accompanied with clumsiness, is often preceded by recurrent upper respiratory infections. Patients will only attain the developmental capacity of a 5 to 6-year-old with a progression to severe mental retardation in adulthood. Several mutations in the AGA gene have been identified in AGU patients, spanning insertions, deletions, and point mutations.

Human α-mannosidosis comprises disorders that are associated with abnormal levels and excretion of mannose-rich oligosaccharides. The infantile form (type I), with onset occurring in the first 12 months after birth, causes rapid progressive mental retardation, hepatosplenomegaly, and severe skeletal abnormalities. Afflicted individuals usually die between 3 and 12 years of age. The juvenile-adult form (type II), with onset between 1 and 4 years of age, is less severe with milder symptoms and slower progression. Patients also suffer recurrent infections that may be related to resultant defects in leukocyte chemotaxis. These disorders result from deficiencies in the lysosomal form of the enzyme, α-mannosidase. This enzyme hydrolyses Man-α1,3-Man and Man-α1,6-Man linkages.

Human β-mannosidosis is associated with increased storage and excretion of Man-β1,4-GlcNAc caused by a deficiency in β-mannosidase activity. The enzyme catalyzes the hydrolysis of Man-β1,4-GlcNAc linkages, hence the accumulation of these structures in its absence. Like α-mannosidosis patients, patients lacking β-mannosidase exhibit symptoms with a range of severities. The most frequent symptoms are respiratory infections, hearing loss with associated speech impairment, and mental retardation.

Fucosidosis is caused by a deficiency in the lysosomal hydrolase α-fucosidase. Originally, two classifications of fucosidosis were defined. Type I was the more

severe with infantile onset and type II was the milder form. However, it has become apparent that individuals suffering from fucosidosis exhibit a continuum of symptoms with a wide spectrum of severity. The most severely afflicted individuals have psychomotor retardation, skeletal abnormalities, coarse facial features, and growth retardation. The enzyme defect results in the increased accumulation and excretion of a wide variety of fucose-containing glycoproteins, glycolipids, and oligosaccharides. There are multiple forms of α-fucosidase, all of which have a common subunit encoded by the *FUCA1* gene. At least 23 different mutations in the *FUCA1* gene have been identified in fucosidosis patients.

Sialidosis is a disorder that results from a deficiency in the neuraminidase (sialidase) responsible for cleaving terminal α2,3- and α2,6-sialyl linkages present in oligosaccharides and glycoproteins. Type I sialidosis is the milder of the two forms of the disease. Symptoms are generally benign in type I patients with the most notable sign being the presence of a ''cherry-red spot'' in the fundus of the eyes. Indeed, type I sialidosis is commonly referred to as ''cherry-red spot-myoclonus'' syndrome. Type II sialidosis can be divided into infantile and juvenile forms. Type II individuals suffer from heptosplenomegaly, proteinuria, hydrops, skeletal abnormalities, and coarse facial features. The primary accumulating structures in type I sialidosis are sialylated mono-, di-, and triantennary N-glycosidic *N*-acetyllactosamine oligosaccharides harboring Man-β1,4-GlcNAc at the reducing end.

Neuraminidase (the enzyme deficient in sialidosis) is just one enzyme of a multienzyme complex that includes neuraminidase, β-galactosidase-2, and a protective protein/cathepsin A (PPCA). PPCA has some sequence similarities to carboxypeptidase and deamidase. A disorder that is related to sialidosis, identified as galactosialidosis, is caused by deficiencies in both neuraminidase and β-galactosidase-2 that are secondary to defects in PPCA. Patients afflicted with galactosialidosis have coarse facial features, cherry-red spots, foam cells in the marrow, vacuolated lymphocytes, and vertebral defects. There are three phenotypic variants of galactosialidosis. The early infantile form is associated with hydrops, visceromegaly, edema, skeletal dysplasia, and early death. The late infantile form results in growth retardation, hepatosplenomegaly, and heart abnormalities. The juvenile/adult form causes angiokeratoma, mental retardation, ataxia, and neurologic deterioration. Numerous different sialylated mono-, di-, tri-, and tetraantennary oligosaccharides accumulate in the lysosomes and are excreted into the body fluids. When the PPCA gene was knocked out in transgenic mice, the animals developed a disorder with striking similarities to the severe-form human galactosialidosis strongly implicating that defects in this gene are the cause of galactosialidosis.

Indicative of the overlap in diseases associated with defects in the lysosomal degradation of both lipid and protein structures harboring similar oligosaccharide moieties are G_{M1} gangliosidosis, G_{M2} gangliosidosis, Schindler disease, and mucolipidosis II and III. These diseases are considered within the context of the family of disorders referred to as lysosomal storage or glycosphingolipid storage diseases. Because the enzymes whose deficiencies lead to these diseases are responsible for the degradation of similar carbohydrate structures on glycolipids and glycoproteins, they can be considered in

a discussion of either topic. Because the two gangliosidotic diseases are more appropriately discussed in the context of glycolipid degradation defects, they will not be discussed here. A brief account of mucolipidosis II (I-cell disease) and mucolipidosis III (pseudo-Hurler polydystrophy) was given earlier.

Schindler disease (often classified as a glycosphingolipid storage disease) is caused by a deficiency in the lysosomal hydrolase α-*N*-acetylgalactosaminidase. This enzyme was formerly called α-galactosidase B. Owing to the enzyme defect sialylated and asialoglycopeptides with terminal α-*N*-acetylgalactosaminyl residues accumulate in the lysosomes. In addition, glycolipids and oligosaccharides harboring terminal α-*N*-acetylgalactosaminyl residues also accumulate. There are three forms of Schindler disease. Type I individuals appear normal for their first 9 to 12 months then a developmental delay sets in followed by increasing neurologic impairment and profound psychomotor retardation. The type II disease is an adult onset disorder characterized by mild intellectual impairment and angiokeratoma. Type III Schindler comprises an intermediate and a variable form whose symptoms range from speech and language delay with autistic presentation to seizures and moderate psychomotor retardation.

7.2 Congenital Disorders of Glycosylation, CDG

Congenital disorders of glycosylation (CDGs) represent a constellation of diseases that result from defects in the N-linked glycosylation pathway. To date, there are 10 CDGs that have been classified (Table 5) and in 9 of these the enzyme defect(s) has been identified and characterized. These diseases are clustered into two broad categories. Group I CDG diseases are defined by alterations/deficiencies in the synthesis and/or transfer of the Dol-PP-oligosaccharide precursor to asparagine residues in substrate proteins. Group II CDG diseases are defined as those that result from defects in subsequent N-linked glycan processing. Each of the 10 CDGs present in the clinic with a similar constellation of symptoms including hypotonia and hepatomegaly. All but CDG-Ib result in psychomotor retardation of varying degrees. Because these disorders all result from defective N-glycosylation, individuals who present with several of the typical clinical symptoms can be tested for protein

Tab. 5 Congenital disorders in glycosylation.

CDG type	*Defect*	*Gene name*	*Location*
Ia	Phosphomannomutase 2	*PMM2*	16p13.3-p13.2
Ib	Phosphomannose isomerase	*PMI*	15q22-qter
Ic	$Man_9GlcNAc_2$-PP-Dol glucosyltransferase	*ALG6*	1p22.3
Id	$Man_5GlcNAc_2$-PP-Dol mannosyltransferase	*ALG3*	3
Ie	Dol-P-Man synthase 1	*DPM1*	20q13
If	Dol-P-Man utilization defect 1	*MPDU1*	17p12-13
Ix	Like Ia but normal *PMM2* and *PMI*		
IIa	*N*-acetylglucosaminyltransferase II	*MGAT2*	14q21
IIb	Glucosidase I	*GSC1*	2p12-13
IIc (LAD II)	GDP-fucose transporter	*FUCT1*	11

misglycosylation by gel electrophoresis. Commonly, serum transferring is the protein of choice for this analysis. It is important to note that these disorders are only reflective of deficiencies of N-glycosylation and that diseases/disorders are known to result from deficiencies in O-glycosylation, GPI linkage and the biosynthesis of proteoglycans, all of which involve carbohydrate addition and remodeling in the context of a protein backbone.

CDG-Ia is the most commonly occurring CDG, with appearance in individuals of European ancestry being highest. Although there is considerable variability in the clinical phenotypes observed in CDG-Ia patients, there is always some level of psychomotor retardation. In addition, children are ataxic and have skeletal abnormalities consisting of long limbs and short torsos. Owing to defective synthesis of coagulation factors by the liver (primarily factor XI, antithrombin III, protein C and protein S), patients have severe coagulation defects. Adding to the situation is hepatomegaly with consequent liver dysfunction. CDG-Ia results from mutations in phosphomannomutase 2 (PMM2), the enzyme that is required to convert Man-6-P to Man-1-P used in the generation of GDP-Man. Over 60 mutations in PMM2 have been identified that either decrease enzyme activity or stability.

CDG-Ib results from defects in the isomerase (phosphomannose isomerase, PMI) necessary to convert fructose-6-P to Man-6-P. Deficiencies in this enzyme cause loss of plasma proteins via the intestines (protein losing enteropathy, PLE), diarrhea, and cyclical vomiting. Liver dysfunction is caused by hepatomegaly and results in coagulation defects and hypoglycemia. Thankfully, CDG-Ib patients are spared from mental defects. Because free mannose is not found at any appreciable level in foods, dietary supplementation with mannose can ameliorate some of the symptoms of CDG-Ib.

CDG-Ic results from defects in α1,3 glucosyltransferase (ALG6) and eight mutations have been identified in the gene in patients with this disorder. ALG6 catalyzes the addition of the first glucose residue to the *en bloc* oligosaccharide that is subsequently N-linked to proteins. As a consequence of the defective ALG6 enzyme, the oligosaccharide unit is not transferred from dolichol to proteins. Symptoms of CDG-Ic are similar to those of CDG-Ia but much less severe. Patients have frequent seizures and mild psychomotor retardation. Intestinal symptoms of CDG-Ic are markedly exacerbated by intestinal viral infections due to a decline of glycosylated structures (both proteins and proteoglycans) on the basolateral surfaces of intestinal enterocytes.

Few cases of CDG-Id, CDG-Ie, and CDG-If have been reported in the literature. Each of these three disorders are related via enzymes that synthesize or utilize Dol-P-mannose structures. CDG-Id results from deficiencies in Dol-P-Man:Dol-PP-$Man_6GlcNAc_2$ α1,3-mannosyltransferase (α-1,2-Man-T). This enzyme transfers Man from Dol-P-Man to Dol-PP-$Man_5GlcNAc_2$ of the growing *en bloc* oligosaccharide. CDG-Ie results from defects in the catalytic subunit of Dol-P-Man synthase (GDP-Man:Dol-P mannosyltransferase, *DPM*1). Dol-P-Man is required for mannose elongation reactions necessary to produce the *en bloc* oligosaccharide of N-glycans, GPI-anchored proteins, and many other Man-containing glycoconjugates. Patients with CDG-If have defects in an enzyme, MPUD1, required for the utilization of Man-P-Dol and Glc-P-Dol. CDG-Id and CDG-Ie individuals suffer severe neurological impairment including

psychomotor retardation and intractable seizures. Both disorders result in eye abnormalities, optic atrophy in CDG-Id, and cortical blindness in CDG-Ie. Persons with CDG-If have clinical symptoms including hypotonia, frequent seizures, severe psychomotor retardation, and cerebral atrophy. Additionally, these individuals fail to thrive after birth.

CDG-IIa results from loss of the enzyme UDP-GlcNAc:α-6-D-mannosidase β1,2-*N*-acetylglucosaminyltransferase II (GlcNAc-TII), which is encoded by the *MGAT2* gene. The results of this defect are significant loss of complex sialylated structures. Following transfer of the *en bloc* oligosaccharide to asparagine residues, only monosialylated structures are made. Symptoms of CDG-IIa include severe psychomotor retardation, hypotonia, coarse facies, and frequent infections. In experiments in mice, it was found that Mgat2 null animals all died within four weeks after birth and a significant number die *in utero* from gastrointestinal blockage. From these results, it is speculated that the true incidence of human Mgat2 defects may go undetected because of spontaneous fetal abortion or death shortly after birth.

7.2.1 Leukocyte Adhesion Deficiency Syndrome II, LAD II

CDG-IIb is more commonly referred to as leukocyte adhesion deficiency syndrome II (LAD II). LAD II belongs to the class of disorders referred to as *primary immunodeficiency syndromes* as the symptoms of the disease manifest because of defects in leukocyte function. Symptoms of LAD II are characterized by unique facial features, recurrent infections, persistent leukocytosis, defective neutrophil chemotaxis, and severe growth and mental retardation. The genetic defect resulting in LAD II is in the pathway of fucose utilization leading to loss of fucosylated glycans on the cell surface. An additional feature of LAD II is that individuals harbor the rare Bombay (hh) blood type at the ABO locus as well as lack the Lewis blood group antigens (see Fig. 2). The Bombay blood type is characterized by a deficiency in the H (referred to as the O-type), A and B antigens due to loss of the fucose residue. Each of these blood group antigens contains a Fuc-α1,2-Gal modification that is the final carbohydrate addition to these antigens. These fucosylation reactions are catalyzed by α1,2-fucosyltransferase, which is encoded by the *H* and *Se* loci. The defective neutrophil chemotaxis is due to the loss of a selectin ligand on these cells. This ligand is the sialylated Lewis^x antigen, another blood group antigen.

The recurrent infections seen in LAD II patients are the result of the defective neutrophil function. Neutrophils are involved in innate immunity responses to bacterial infection. To carry out their role in host-defense mechanisms, neutrophils must adhere to the surface of the endothelium at the site of inflammation, which is an event mediated by cell surface–adhesion molecules. The selectin family (E-, L-, and P-selectins) of animal lectins are necessary to mediate the initial process of neutrophil adherence to the endothelium. The selectins recognize sialylated fucosylated lactosamines typified by the Lewis^x antigen (see Fig. 2). Once neutrophils adhere and roll along the surface of the endothelium (due to vascular flow), the integrin family of adhesion molecules allow for firm adherence followed by tissue penetration. A related disorder, termed LAD I, is caused by the absence of CD18, which is the β2 subunit of the leukocyte integrin found on the surface of neutrophils and monocytes.

Because there is widespread loss of fucosylated antigens in LAD II patients, each of which can be formed through the actions of several fucosyltransferases, the role of these enzymes in the disease could be ruled out. In addition, normal levels of the α1,2-, α1,3- and α1,4-fucosyltransferases were observed in the serum of LAD II individuals. These observations indicated that the pathways to GDP-fucose synthesis or utilization by the fucosyltransferases in the Golgi must be deficient in LAD II. Fucose can be converted to GDP-fucose by salvage of free fucose (exogenous or derived through glycoconjugate degradation) or by epimerization of GDP-mannose. In order for GDP-fucose to be utilized by Golgi fucosyltransferases it must first be transported into the Golgi from the cytosol where it is synthesized. Examinations of the enzymes involved in the synthesis and transport of GDP-fucose have been undertaken. While the activity of one of the enzymes of GDP-mannose epimerization (GDP-D-mannose 4,6-dehydratase, GMD) has been shown to be reduced in LAD II patients, it has been determined that the major defect causing LAD II is an impairment in the transport of GDP-fucose into the Golgi.

7.3 Exploitation of Host Glycosylation

Many viruses, bacteria, and parasites exploit the presence of cell surface carbohydrates, principally associated with protein, using interaction with them as the means of entry into the cell (Table 6).

Tab. 6 Cell surface glycoproteins that function as receptors or coreceptors for pathogen entry.

Cell surface protein	*Pathogen*
ICAM-1	Rhinoviruses
ICAM-1	Coxsackieviruses
N-CAM	Rabies virus
Dystroglycan	Lymphocytic choriomeningitis virus
Dystroglycan	Lassa fever virus
Dystroglycan	Oliveros and Mobala viruses
Dystroglycan	*Mycobacterium leprae*
Chemokine receptors: CCR1, CCR5, and CXCR4	Poxviruses
CD4	Human immunodeficiency virus
CD46	Human herpes virus 6
CD46	Measles virus
CD55	Coxsackieviruses
CD55	Echoviruses
Pvr	Poliovirus
HveB and C	Herpes simplex viruses 1 and 2
Integrin $\alpha_v\beta_3$	Foot-and-mouth disease virus
Integrin $\alpha_v\beta_5$	Cocksackieviruses
Integrin $\alpha_v\beta_3$ and $\alpha_v\beta_5$	Adenovirus
VLA-2	Echoviruses
Erythrocyte IL-8 receptor (Duffy blood group antigen)	*Plasmodium vivax*
Glycophorin	*Plasmodium falciparum*
Glycophorin	Influenza virus
Lewis^b antigen	*Helicobacter pylori*

Additionally, the surface glycoproteins of many pathogens are necessary for interaction with host cell surface antigens. Several clinically important examples will be presented here.

Human immunodeficiency virus (HIV), the causative agent of acquired immunodeficiency syndrome (AIDS), is a retrovirus that is a member of the family of lentiviruses that cause degenerative diseases. All lentiviruses establish a persistent infection of cells of the macrophage/monocyte lineage. Infective HIV contains a coat surface glycoprotein that is designated gp120. It is interaction of HIV gp120 with the T-cell surface glycoprotein, CD4, that is the major interaction required to initiate the infection process. CD4 is a ligand for major histocompatibility complex (MHC) class II molecules and is a member of the Ig superfamily of proteins. Macrophages and dendritic cells express barely detectable levels of CD4, thus, HIV infection is thought to occur through interactions with other glycoproteins or lectins on the surfaces of these cell types. Indeed, a protein on dendritic cells identified as DC-SIGN, has been demonstrated to have high affinity for gp120. Additionally, it is known that the host proteins are incorporated into HIV virions and consequently aid in HIV attachment to cells. Examples include ICAM-1, which when present in the virion surface allow HIV to attach to cells expressing the ICAM-1 ligand LFA-1 (GPI-anchored protein).

Although CD4 has been clearly demonstrated to be the major receptor for HIV infection, evidence has shown that CD4 alone is insufficient for fusion of virions with cells to which they are attached. Virus entry requires interaction with additional cell surface glycoproteins. These proteins have been termed HIV coreceptors. The strongest coreceptors for HIV infection are proteins that belong to a class of cellular receptors known as the chemokine receptors. Chemokines are chemoattractant proteins having roles in inflammation and immune responses as well as roles in early development. The most potent chemokine receptors involved in HIV entry into cells are CXCR4 and CCR5.

Human herpesvirus 6 (HHV-6) infection occurs in virtually all persons within the first two years of life and persists the entire lifetime. In immunocompromised patients, HHV-6 causes opportunistic infections and is the causative agent of *exanthema subitum*. HHV-6 has been linked to multiple sclerosis and to the progression of AIDS. The cellular receptor for HHV-6 is the cell surface type I glycoprotein, CD46.

Dystroglycan (DG) is a component of the dystrophin–glycoprotein complex. It is a laminin receptor encoded by a single gene and cleaved by postranslational processing into two proteins, peripheral membrane α-DG, and transmembrane β-DG. Alpha-DG interacts with laminin-2 in the basal lamina and β-DG binds to dystrophin-containing cytoskeletal proteins in muscle and peripheral nerves. DG is involved in agrin- and laminin-induced acetylcholine receptor clustering at neuromuscular junctions, morphogenesis, early development, and the pathogenesis of muscular dystrophies. Evidence has demonstrated that α-DG present on Schwann cell membranes is the receptor for *Mycobacterium leprae* and also serves as the receptor for the arenavirus class of pathogens. Arenaviruses cause hemorrhagic fever in humans. Lymphocytic choriomeningitis virus (LCMV), Lassa fever virus (LFV), Oliveros and Mobala (all members of the arenavirus family) all bind to α-DG. The specificity of this interaction was demonstrated by the resistance to LCMV infection of cells harboring a null mutation in DG.

The MN blood group system is a well-characterized set of erythrocyte surface antigens that represent the variable carbohydrate modifications of the transmembrane glycoprotein, glycophorin. Glycophorin is the cellular receptor for influenza virus as well as the receptor for erythrocyte invasion by the malarial parasite *P. falciparum*. The related malarial parasite, *P. vivax*, interacts with the erythrocyte chemokine receptor for the cytokine IL-8 (which is in fact a chemokine). This erythrocyte surface glycoprotein is also known as the Duffy blood group antigen.

*H. pylori*is a gram-negative bacterium that is found in the stomachs of over 50% of the human population. However, even with this wide distribution, significant pathological consequences to infection are only seen in a subpopulation. In this subpopulation, the bacterium is responsible for chronic active gastritis and gastric and duodenal ulcers and has also been shown to be the causative agent for one of the most common forms of cancer in humans, gastric adenocarcinoma. Indeed, in 1994, *H. pylori* was designated as a human carcinogen. *H. pylori* resides mainly in the mucous layer of colonized individuals. The bacterium can, however, directly interact with gastric epithelial cells expressing the Lewis^b (Le^b) blood group antigen on their surfaces. The Lewis^b antigen is very similar to the B antigen (see Fig. 2) except that the terminal residue in the Lewis^b carbohydrate structure is Fuc, whereas on the B antigen it is Gal. Although binding to the Lewis^b antigen allows *H. pylori* to directly interact with the gastric epithelium, it is not this interaction that is important in the clinical course of colonization. Progression in the development of gastric cancer involves loss of stomach parietal and pepsinogen-producing epithelial cells. The loss of these cells types is associated with an increase in the exposure of progenitor cells and in the synthesis of sialylated glycans. *H. pylori* also binds to sialic acid containing glycans with the terminal structure, Sia-α2,3-Gal-β. This is the same structure present in the sialylated Lewis^x antigen (see Fig. 2). Persistent colonization of *H. pylori* at sites with increased progenitor cell proliferation is conducive to the further transformation of these cells.

8
Medical Applications of Carbohydrates and Glycoproteins

Given that several biotechnology and pharmaceutical companies are testing the utility of certain carbohydrate-related compounds in the treatment of disease (Table 7 here), it is clear that oral delivery of sugars has benefits far beyond simply serving as sweeteners. In addition to using sugars as therapeutics, the use of inhibitors of carbohydrate recognition and glycosylation inhibitory compounds has clinical significance for treatments ranging from cancer to diabetes.

The congenital disorder in glycosylation type Ib (CDG-Ib) is caused by a deficiency in the ability of cells to convert fructose to mannose (see Sect. 7.2). As indicated earlier, the administration of mannose to afflicted individuals has shown striking promise in the amelioration of the severe intestinal symptoms of this disease. Leukocyte adhesion deficiency II (LAD II) results from a defect in the intracellular transport of GDP-fucose into the Golgi (see Sect. 7.2.1). Oral supplementation of fucose has been shown to induce the expression of fucosylated selectin ligands on neutrophils and core fucosylation of serum glycoproteins. The result of these supplementations showed

Tab. 7 Carbohydrate compounds or inhibitors in the clinic.

Company	*Carbohydrate/inhibitor*	*Clinical indication*
Austin Research Inst.	Mannan/mucin formulation	Breast cancer
Biomira	Sialyl Tn antigen conjugate	Metastatic cancer
GLYCODesign	Swainsonine analog	Renal cell carcinoma
Oxford glycosciences	Imino sugar analog	Gaucher's disease
Progen Industries	Sulfated oligosaccharide	Cancer
Progenics	G_{M2}/G_{D2}/KLH conjugate	Cancer
Progenics	G_{D2}/KLH conjugate	Melanoma
SafeScience	Pectin-based oligosaccharide	Cancer
Synsorb Biotech	Oligosaccharide conjugate	*Clostridium difficile* infection
Synsorb Biotech	Oligosaccharide conjugate	*Escherichia coli* O157:H7 infection

that neutrophil counts returned to normal reducing infection in rats and fever as well as an improvement in psychomotor capabilities.

The selectin class of lectins is critically important in the normal function of leukocytes in processes of inflammation. Studies in mice showed that blocking the targets of the selectins can reduce inflammatory responses to blood vessel damage. Blocking L-selectin recognition of the Lewis^x antigen showed particular promise. However, human clinical trials have been disappointing. Currently, the drug company, Wyeth, is testing a P-selectin inhibitor in the treatment of ischemic reperfusion injury with initial success. Inhibition of P-selectin binding to its ligand may also hold promise as an anticancer regimen. Metastatic cancer cells adhere to P-selectin on platelets, which in turn protect them from being recognized by the immune system.

Golgi α-mannosidase II is a key enzyme in the processing of N-linked glycoproteins and has been shown to be a potential target for the development of antiviral and anticancer therapies. The glycosylation inhibitors swainsonine and deoxymannojirimycin block the action of α-mannosidase II. A deoxymannojirimicin analog, *N*-nonyldeoxynojirimicin (NN-DNJ) has been shown to block the infectivity of human hepatitis B and C viruses. The ER located α-mannosidase I is responsible for trimming of a single mannose residue and as a consequence acts as a signal to target misfolded glycoproteins for degradation by the proteasome. The potent α-mannosidase I inhibitors, kifunensine and 1-deoxymannojirimycin, may prove useful in the design of novel inhibitors that prevent degradation of misfolded proteins in genetic diseases.

The action of the intestinal enzyme, α-glucosidase, is to aid in the digestion and subsequent absorption of dietary saccharides. Two sugar-related α-glucosidase inhibitors, BAY m-1099 (Miglitol) and BAY g-5421 (Acarbose), have been targeted in the clinic as antidiabetic drugs. The action of these inhibitors prevents complete digestion of dietary carbohydrate, thus reducing glucose uptake. Although targeted for the same condition, the effects of these inhibitors are somewhat different owing to the fact that only Miglitol shows appreciable absorption into the blood stream. A secondary benefit to the uptake of Miglitol is that studies have shown that not only does the drug impair glucose absorption but it also enhances insulin

responsiveness and decreases glucagon-induced glycogen breakdown.

See also Bioorganic Chemistry.

Bibliography

Books and Reviews

Aula, P., Jalanko, A., Peltonen, L. (2001) Aspartylglucosaminuria, in: Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D. (Eds.) *The Metabolic & Molecular Basis of Inherited Disease*, 8th edition, McGraw-Hill, New York.

Cummings, R.D. (1992) Synthesis of Asparagine-Linked Oliogsaccharides. Pathways, Genetics and Metabolic Regulation, in: Allen, H.J., Kisalius, E.C. (Eds.) *Glycoconjugates: Composition, Structure and Function*, Marcel Dekker, New York, pp. 333–360.

Desnick, R.J., Schindler, D. (2001) α-*N*-acetylgalactosaminidase Deficiency: Schindler Disease, in: Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D. (Eds.) *The Metabolic & Molecular Basis of Inherited Disease*, 8th edition, McGraw-Hill, New York.

Hart, G.W. (1997) Dynamic *O*-linked glycosylation of nuclear and cytoskeletal proteins, *Annu. Rev. Biochem.* **66**, 315–335.

Hurtley, S.M., Helenius, A. (1989) Protein oligomerization in the endoplasmic reticulum, *Annu. Rev. Cell. Biol.* **5**, 277–307.

Kilpatrick, D.C. (2002) Animal lectins: a historical introduction and overview, *Biochim. Biophys. Acta* **1572**, 187–197.

Parodi, A.J. (2000) Protein glucosylation and its role in protein folding, *Annu. Rev. Biochem.* **69**, 69–93.

Ruoslathi, E. (1988) Structure and biology of proteoglycans, *Annu. Rev. Cell Biol.* **4**, 229–255.

Thomas, G.H. (2001) Disorders of Glycoprotein Degradation: α-Mannosidosis, β-Mannosidosis, Fucosidosis and Sialidosis, in: Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D. (Eds.) *The Metabolic & Molecular Basis of Inherited Disease*, 8th edition, McGraw-Hill, New York.

van den Steen, P., Rudd, P.M., Dwek, R.A., Opdenakker, G. (1998) Concepts and principles of *O*-linked glycosylation, *Crit. Rev. Biochim. Mol. Biol.* **33**, 151–208.

Williams, A.F., Barclay, A.N. (1988) The immunoglobulin superfamily – domains for cell surface recognition, *Annu. Rev. Immunol.* **6**, 381–405.

Primary Literature

Akahani, S., Nangia-Makker, P., Inohara, H., Kim, H.R., Raz, A. (1997) Galectin-3, anti-apoptotic molecule with a functional HH1 (NWGR) domain of Bcl-2 family, *Cancer Res.* **57**, 5272–5276.

Alkhatib, G., Combadiere, C., Broder, C.C., Feng, Y., Kennedy, P.E., Murphy, P.M., Berger, E.A. (1996) CC CKR5: a RANTES, MIP-1α, MIP-1β receptor as a fusion cofactor for macrophage-tropic HIV-1, *Science* **272**, 1955–1958.

Angata, T., Brinkman-van der Linden, E.C.M. (2002) I-type lectins, *Biochim. Biophys. Acta* **1572**, 294–316.

Apweiler, R., Hermjakob, H., Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database, *Biochim. Biophys. Acta* **1473**, 4–8.

Becker, D.J., Lowe, J.B. (1999) Leukocyte adhesion deficiency type II, *Biochim. Biophys. Acta* **1455**, 193–204.

Bjoern, S., Foster, D.C., Thim, L., Wiberg, F.C., Christensen, M., Komiyama, Y., Pederson, A.H., Kisiel, W. (1991) Human plasma and recombinant factor VII. Characterization of *O*-glycosylation of serine residues 52 and 60 and effect of site-directed mutagenesis of serine 52 to alanine, *J. Biol. Chem.* **266**, 11051–11057.

Borén, T., Falk, P., Roth, K.A., Larson, G., Normark, S. (1993) Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens, *Science* **262**, 1892–1895.

Burda, P., Aebi, M. (1999) The dolichol pathway of *N*-linked glycosylation, *Biochim. Biophys. Acta* **426**, 239–257.

Cao, W., Henry, M.D., Borrow, P., Yamada, H., Elder, J.H., Ravkov, E.V., Nichol, S.T., Compans, R.W., Campbell, K.P., Oldstone, M.B. (1998) Identification of alpha-dystroglycan as a receptor for lymphocytic choriomeningitis virus and Lassa fever virus, *Science* **282**, 2079–2081.

Cebo, C., Dambrouck, T., Maes, E., Laden, C., Strecker, G., Michalski, J.C., Zanetta, J.P.

(2001) Recombinant human interleukins IL-1alpha, IL-1beta, IL-4, IL-6 and IL-7 show different and specific calcium independent carbohydrate-binding properties, *J. Biol. Chem.* **276**, 5685–5691.

Cebo, C., Durier, V., Lagant, P., Maes, E., Florea, D., Lefebvre, T., Strecker, G., Vergoten, G., Zanetta, J.P. (2002b) Function and molecular modeling of the interaction between human interleukin 6 and its HNK-1 oliogosaccharide ligands, *J. Biol. Chem.* **277**, 12246–12252.

Cebo, C., Vergoten, G., Zanetta, J.-P. (2002a) Lectin activities of cytokines: functions and putative carbohydrate-recognition domains, *Biochim. Biophys. Acta* **1572**, 422–434.

Cheshenko, N., Herold, B.C. (2002) Glycoprotein B plays a predominant role in mediating herpes simplex virus type 2 attachment and is required for entry and cell-to-cell spread, *J. Gen. Virol.* **83**, 2247–2255.

Choudhury, P., Liu, Y., Bick, R.J., Sifers, R.N. (1997) Intracellular association between UDP-glucose: glycoprotein glucosyltransferase and an incompletely folded variant of alpha1-antitrypsin, *J. Biol. Chem.* **272**, 13446–13451.

Chui, D., Sellakumar, G., Green, R.S., Sutto-Smith, M., McQuistan, T., Marek, K.W., Morris, H.F., Dell, A., Martin, J.D. (2001) Genetic remodeling of protein glycosylation *in vivo* induces autoimmune disease, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1142–1147.

Clapham, P.R., McKnight, A. (2002) Cell surface receptors, virus entry and tropism of primate lentiviruses, *J. Gen. Virol.* **83**, 1809–1829.

Clausen, H., Bennett, E.P. (1996) A family of UDP-GalNAc: polypeptide *N*-acetylgalactosaminyl-transferases control the initiation of mucin-type *O*-glycosylation, *Glycobiology* **6**, 635–646.

Comer, F.I., Hart, G.W. (2000) *O*-glycosylation of nuclear and cytosolic protein. Dynamic interplay between *O*-GlcNAc and *O*-phosphate, *J. Biol. Chem.* **275**, 29179–29182.

Crocker, P.R., Gordon, S. (1986) Properties and distribution of a lectin-like hemagglutinin differentially expressed by murine stromal tissue macrophages, *J. Exp. Med.* **164**, 1862–1875.

Crocker, P.R., Varki, A. (2001) Siglecs, sialic acids and innate immunity, *Trends Immunol.* **22**, 337–342.

Dahms, N.M., Hancock, M.K. (2002) P-type lectins, *Biochim. Biophys. Acta* **1572**, 317–340.

Damtew, B., Rzewnicki, D., Lozanski, G., Kushner, I. (1993) IL-1 receptor antagonist affects the plasma protein response of Hep3B cells to conditioned medium from lipopolysaccharide-stimulated monocytes, *J. Immunol.* **150**, 4001–4007.

Danguy, A., Camby, I., Kiss, R. (2002) Galectins and cancer, *Biochim. Biophys. Acta* **1572**, 285–293.

Daniels, M.A., Hogquist, K.A., Jameson, S.C. (2002) Sweet 'n' sour: the impact of differential glycosylation on T cell responses, *Nat. Immunol.* **3**, 903–910.

de Beer, T., Vliegenthart, J.F., Loffler, A., Hofsteenge, J. (1995) The hexapyranosyl residue that is that is *C*-glycosidically linked to the side chain of tryptophan-7 in human RNAse U_S, is a-mannopyranose, *Biochemistry* **34**, 11785–11789.

Demetriou, M., Granovsky, M., Quaggin, S., Dennis, J.W. (2001) Negative regulation of T-cell activation and autoimmunity by Mgat5*N*-glycosylation, *Nature* **409**, 733–737.

Dhume, S.T., Adams-Burton, C.R., Shumak, K.H., Laine, R.A. (1994) Polylactosamines are not obligate receptors for invasion of Plasmodium falciparum malaria as shown in HAMPAS carian II-gal erythrocytes, *Glycobiology* **4**, 903–908.

Dinarello, C.A. (1994) The interleukin-1 family: 10 years of discovery, *FASEB J.* **8**, 1314–1325.

Ding, D.X., Vera, J.C., Heany, M.L., Golde, D.W. (1995) N-glycosylation of the human granulocyte-macrophage colony-stimulating factor receptor alpha subunit is essential for ligand binding and signal transduction, *J. Biol. Chem.* **270**, 24580–24584.

Doucey, M.-A., Hess, D., Blommers, M.J.J., Hofsteenge, J. (1999) Recombinant human interleukin-12 is the second example of a C-mannosylated protein, *Glycobiology* **9**, 435–441.

Doucey, M.-A., Hess, D., Cacan, R., Hofsteenge, J. (1998) Protein *C*-mannosylation is enzyme-catalyzed and uses dolichyl-phosphate-mannose as a precursor, *Mol. Biol. Cell* **9**, 291–300.

Dragic, T., Litwin, V., Allaway, G.P., Martin, S.R., Huang, Y., Nagashima, K.A., Cayanan, C., Maddon, P.J., Koup, R.A., Moore, J.P., Paxton, W.A. (1996) HIV-1 entry into $CD4^+$ cells is mediated by the chemokine receptor CC-CKr-5, *Nature* **381**, 667–673.

Drent, M.L., Tollefsen, A.T., van Heusden, F.H., Hoenderdos, E.B., Jonker, J.J., van der Veen, E.A. (2002) Dose-dependent efficacy of miglitol, an alpha-glucosidase inhibitor, in type 2 diabetic patients on diet alone: results of a 24-week double-blind placebo-controlled study, *Diabetes Nutr. Metab.* **15**, 152–159.

Drickamer, K. (1999) C-type lectin domains, *Curr. Opin. Struct. Biol.* **9**, 585–590.

Elhammer, A.P., Poorman, R.A., Brown, E., Maggiora, L.L., Hoogerheide, J.G., Kezdy, F.J. (1993) The specificity of UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase as inferred from a data base of *in vivo* substrates and from *in vitro* glycosylation of proteins and peptides, *J. Biol. Chem.* **268**, 10029–10038.

Endo, T. (1999) *O*-mannosyl glycans in mammals, *Biochim. Biophys. Acta* **1473**, 237–246.

Englund, P.T. (1993) The structure and biosynthesis of glycosyl phosphatidylinositol protein anchors, *Annu. Rev. Biochem.* **62**, 121–138.

Etzioni, A., Tonetti, M. (2000) Leukocyte adhesion deficiency II – from A to almost Z, *Immunol. Rev.* **178**, 138–147.

Feng, Y., Broder, C.C., Kennedy, P.E., Berger, E.A. (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor, *Science* **272**, 872–877.

Finne, J., Krusius, T., Margolis, R.K., Margolis, R.U. (1979) Novel mannitol-containing oligosaccharides obtained by mild alkaline borohydride treatment of chondroitin sulfate proteoglycan from brain, *J. Biol. Chem.* **254**, 10295–10300.

Fortin, J.F., Barbeau, B., Hedman, H., Lundgren, E., Tremblay, M.J. (1999) Role of the leukocyte function antigen-1 conformational state in the process of human immunodeficiency virus typr-1-mediated syncytium formation and virus infection, *Virology* **257**, 228–238.

Freeze, H.H. (2001) Update and perspectives on congenital disorders of glycosylation, *Glycobiology* **11**, 129R–143R.

Fujita, T. (2002) Evolution of the lectin-complement pathway and its role in innate immunity, *Nat. Rev. Immunol.* **2**, 346–353.

Gabius, H.-J., André, S., Kaltner, H., Siebert, H.-C. (2002) The sugar code: functional lectinomics, *Biochim. Biophys. Acta* **1572**, 165–177.

Galve de Rochemonteix, B., Wiktorowicz, K., Kushner, I. (1993) C-reactive protein increases production of IL-1α, IL-1β, and TNF-α, and expression of mRNA by human alveolar macrophages, *J. Leukocyte Biol.* **53**, 439–445.

Gewurz, H., Zhang, X.-H., Lint, T.F. (1995) Structure and function of the pentraxins, *Curr. Opin. Immunol.* **7**, 54–64.

Gillespie, W., Paulson, J., Kelm, S., Pang, M., Baum, L.G. (1993) Regulation of α2,3-sialyltransferase expression correlates with conversion of peanut agglutinin (PNA)$^+$ to PNA-phenotype in developing thymocytes, *J. Biol. Chem.* **268**, 3801–3804.

Harris, R.J., Ling, V.T., Spellman, M.W. (1992) *O*-linked fucose is present in the first epidermal growth factor domain of factor XII but not protein C, *J. Biol. Chem.* **267**, 5102–5107.

Harris, R.J., Spellman, M.W. (1993) *O*-linked fucose and other posttranslational modifications unique to EGF modules, *Glycobiology* **3**, 219–224.

Helenius, A., Aebi, M. (2001) Intracellular functions of N-linked glycans, *Science* **291**, 2364–2369.

Hille-Rehfeld, A. (1995) Mannose-6-phosphate receptors in sorting and transport of lysosomal enzymes, *Biochim. Biophys. Acta* **1241**, 177–194.

Hiller, M.M., Finger, A., Schweiger, M., Wolf, D.H. (1996) ER degradation of a misfolded luminal protein by the cytosolic ubiquitin-proteasome pathway, *Science* **273**, 1725–1728.

Hooper, L.V., Gordon, J.L. (2001) Glycans as legislators of host-microbial interactions: spanning the spectrum from symbiosis to pathogenicity, *Glycobiology* **11**, 1R–10R.

Horuk, R., Chitnis, C.E., Darbonne, W.C., Colby, T.J., Rybicki, A., Hadley, T.J., Miller, L.H. (1993) A receptor for the malarial parasite *Plasmodium vivax*: the erythrocyte chemokine receptor, *Science* **261**, 1182–1184.

Hughes, R.C. (1997) The galectin family of mammalian carbohydrate-binding molecules, *Biochem. Soc. Trans.* **25**, 1194–1198.

Jacob, G.S. (1995) Glycosylation inhibitors in biology and medicine, *Curr. Opin. Struct. Biol.* **5**, 605–611.

Johansen, P.G., Marshall, R.D., Neuberger, A. (1961) Carbohydrates in protein. The preparation and some of the properties of a glycopeptide from hen egg ovalbumin, *Biochem. J.* **78**, 518–527.

Junghans, U., Waheed, A., von Figura, K. (1988) The "cation-dependent" mannose-6-phosphate receptor binds ligands in the absence of divalent cations, *FEBS Lett.* **237**, 81–84.

Karlsson, K.-A. (1995) Microbial recognition of target-cell glycoconjugates, *Curr. Opin. Struct. Biol.* **5**, 622–635.

Korner, C., Linnebank, M., Koch, H.G., Harms, E., von Figura, K., Marquardt, T. (1999) Decreased availability of GDP-L-fucose in a patient with LAD II with normal GDP-D-mannose dehydratase and FX protein activities, *J. Leukocyte Biol.* **66**, 95–98.

Kornfeld, S. (1992) Structure and function of the mannose-6-phosphate/insulinlike growth factor II receptors, *Annu. Rev. Biochem.* **61**, 307–330.

Kunz, S., Borrow, P., Oldstone, M.B.A. (2002) Receptor structure, binding, and cell entry of arenaviruses, *Curr. Top. Microbiol. Immunol.* **262**, 111–137.

Kushner, I. (1993) Regulation of the acute phase response by cytokines, *Perspect. Biol. Med.* **36**, 611–622.

Lalani, A.S., Masters, J., Zeng, W., Barrett, J., Pannu, R., Everett, H., Arendt, C.W., McFadden, G. (1999) Use of chemokine receptors by poxviruses, *Science* **286**, 1968–1971.

Lee, G.W., Lee, T.H., Vilcek, J. (1993) TSG-14, a tumor necrosis factor- and IL-1-inducible protein, is a novel member of the pentraxin family of acute phase proteins, *J. Immunol.* **150**, 1804–1812.

Loris, R. (2002) Principles of structures of animal and plant lectins, *Biochim. Biophys. Acta* **1572**, 198–208.

Lowe, J.B. (2001) Glycosylation, immunity, and autoimmunity, *Cell* **104**, 809–812.

Lu, J., The, C., Kishore, U., Reid, K.B.M. (2002) Collectins and ficolins: sugar pattern recognition molecules of the mammalian innate immune system, *Biochim. Biophys. Acta* **1572**, 387–400.

Marquardt, T., Luhn, K., Srikrishna, G., Freeze, H.H., Harms, E., Vestweber, D. (1999) Correction of leukocyte adhesion deficiency type II with oral fucose, *Blood* **94**, 3976–3985.

Martin, L.T., Marth, J.D., Varki, A., Varki, N.M. (2002) Genetically altered mice with different sialyltransferase deficiencies show tissue specific alterations in sialylation and sialic acid 9-*O*-acetylation, *J. Biol. Chem.* **277**, 32930–32938.

McCracken, A.A., Brodsky, J.L. (1996) Assembly of ER-associated protein degradation *in vitro*: dependence on cytosol, calnexin, and ATP, *J. Cell Biol.* **132**, 291–298.

Moremen, K.W., Trimble, R.B., Herscovics, A. (1994) Glycosidases of the asparagines-linked processing pathway, *Glycobiology* **4**, 113–125.

Paquette, J.S., Fortin, J.F., Blanchard, L., Tremblay, M.J. (1998) Level of ICAM-1 surface expression on virus producer cells influences both the amount of virion-bound host ICAM-1 and human immunodeficiency virus type-1 infectivity, *J. Virol.* **72**, 9329–9336.

Parker, C.G., Fessler, L.I., Nelson, R.E., Fessler, J.H. (1995) Drosophila UDP-glucose:glycoprotein glucosyltransferase: sequence and characterization of an enzyme that distinguishes between denatured and native proteins, *EMBO J.* **14**, 1294–1303.

Patel, K.D., Cuvelier, S.L., Wiehler, S. (2002) Selectins: critical mediators of leukocyte recruitment, *Semin. Immunol.* **14**, 73–81.

Paul, S.P., Taylor, L.S., Stansbury, D.W., McVica, D.W. (2000) Myeloid specific human CD33 is an inhibitory receptor with differential ITIM function in recruiting the phosphatases SHP-1 and SHP-2, *Blood* **96**, 483–490.

Perillo, N.L., Marcus, M.E., Baum, L.G. (1998) Galectins: versatile modulators of cell adhesion, cell proliferation, and cell death, *J. Mol. Med.* **76**, 402–412.

Rabinovich, G.A., Daly, G., Dreja, H., Tailor, H., Riera, C.M., Hirabayshi, J., Chernajovsky, Y. (1999) Recombinant galectin-1 and its genetic delivery suppress collagen-induced arthritis via T cell apoptosis, *J. Exp. Med.* **190**, 385–398.

Rabinovich, G.A., Rubinstein, N., Fainboim, L. (2002a) Unlocking the secrets of galectins: a challenge at the frontier of glycol-immunology, *J. Leukocyte Biol.* **71**, 741–752.

Rabinovich, G.A., Rubenstein, N., Toscano, M.A. (2002b) Role of galectins in inflammatory and immunomodulatory processes, *Biochim. Biophys. Acta* **1572**, 274–284.

Rambukkana, A., Yamada, H., Zanazzi, G., Mathus, T., Salzer, J.L., Yurchenco, P.D., Campbell, K.P., Fischetti, V.A. (1998) Role of alpha-dystroglycan as a Schwann cell receptor for mycobacterium leprae, *Science* **282**, 2076–2079.

Rini, J.M., Lobsanov, Y.D. (1999) New animal lectin structures, *Curr. Opin. Struct. Biol.* **9**, 578–684.

Robinson, D., Phillips, N.C., Winchester, B. (1975) Affinity chromatography of human liver α-D-mannosidase, *FEBS Lett.* **53**, 110–112.

Rorive, S., Belot, N., Decaestecker, C., Lefranc, F., Gordower, L., Micik, S., Salmon, I., Kiss, R., Camby, I. (2001) Galectin-1 is highly expressed in human gliomas with relevance for modulation of invasion of tumor astrocytes into brain parenchyma, *Glia* **33**, 241–255.

Rudd, P.M., Wormald, M.R., Stanfield, R.L., Huang, M., Mattson, N., Speir, J.A., DiGennaro, J.A., Fetrow, J.S., Dwek, R.A., Wilson, I.A. (1999) Roles for glycosylation of cell surface receptors involved in cellular immune recognition, *J. Mol. Biol.* **293**, 351–366.

Santoro, F., Kennedy, P.E., Locatelli, G., Malnati, M.S., Berger, E.A., Lusso, P. (1999) CD46 is a cellular receptor for human herpesvirus 6, *Cell* **99**, 817–827.

Schenk, B., Fernandez, F., Waechter, C.J. (2001) The ins(ide) and outs(ide) of dolichyl phosphate biosynthesis and recycling in the endoplasmic reticulum, *Glycobiology* **11**, 61R–70R.

Schreiner, R., Schnabel, E., Weiland, F. (1904) Novel *N*-glycosylation in eukaryotes: laminin contains the linkage unit β-glucosylasparagine, *J. Cell Biol.* **124**, 1071–1081.

Schwartz, N.B., Pirok, E.W., Mensch, J.R., Domowicz, M.S. (1999) Domain organization, genomic structure, evolution, and regulation of expression of the aggrecan gene family, *Prog. Nucleic Acid Res. Mol. Biol.* **62**, 177–225.

Sherblom, A.P., Sathyamoorthy, N., Decker, J.M., Muchmore, A.V. (1989) IL-2, a lectin with specificity for high mannose glycopeptides, *J. Immunol.* **143**, 939–944.

Sousa, M.C., Ferrero-Garcia, M.A., Parodi, A.J. (1992) Recognition of the oligosaccharide and protein moieties of glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase, *Biochemistry* **31**, 97–105.

Spiro, R.G. (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds, *Glycobiology* **12**, 43R–56R.

Spiro, R.G., Spiro, M.J. (1971) Studies on the biosynthesis of the hydroxylysine-linked disaccharide unit of basement membranes and collagens III. Tissue and subcellular distribution of glycosyltransferases and the effects of various conditions on the enzyme levels, *J. Biol. Chem.* **246**, 4919–4925.

Stamenkovi, I., Seed, B. (1990) The B-cell antigen CD22 mediates monocyte and erythrocyte adhesion, *Nature* **345**, 74–77.

Steel, D.M., Whitehead, A.S. (1994) The major acute phase reactants: C-reactive protein, serum amyloid P component and serum amyloid A protein, *Immunol. Today* **15**, 81–88.

Sturla, L., Etzioni, A., Bisso, A., Zanardi, D., De Flora, G., Silengo, L., De Flora, A., Tonetti, M. (1998) Defective intracellular activity of GDP-D-mannose-4,6-dehydratase in leukocyte adhesion deficiency type II syndrome, *FEBS Lett.* **429**, 274–278.

Taira, S., Nariuchi, H. (1988) Possible role of neuraminidase in activated T cells in the recognition of allogeneic Ia, *J. Immunol.* **141**, 440–446.

Takeda, J., Kinoshita, T. (1993) GPI-anchor biosynthesis, *TIBS* **20**, 367–371.

Takeda, J., Miyata, T., Kawagoe, K., Iida, Y., Endo, Y., Fujita, T., Takahashi, M., Kitani, T., Kinoshita, T. (1993) Deficiency of the GPI anchor caused by a somatic mutation of the PIG-A gene in paroxysmal nocturnal hemoglobinuria, *Cell* **73**, 703–711.

Ten Hagen, K.G., Bedi, G.S., Tetaert, D., Kingsley, P.D., Hagen, F.K., Balys, M.M., Beres, T.M., Degand, P., Tabak, L.A. (2001) Cloning and characterization of a ninth member of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family, pp-GalNAcTase-T9, *J. Biol. Chem.* **276**, 17395–17404.

Uhlenbruck, G., Sölter, J., Janssen, E., Haupt, H. (1981) Two new, additional "combining sites" of C-reactive protein: lectin specificity of the anti-galactan type and anti-haemocyanin reactivity, *Hoppe Seyler's Z. Physiol. Chem.* **362**, 1167–1169.

Vallee, F., Karaveg, K., Herscovics, A., Moremen, K.W., Howell, P.L. (2000) Structural basis for catalysis and inhibition of N-glycan processing class I alpha 1,2-mannosidases, *J. Biol. Chem.* **275**, 41287–41298.

van den Elsen, J.M., Kuntz, D.A., Rose, D.R. (2001) Structure of Golgi alpha-mannosidase II: a target for inhibition of growth and metastasis of cancer cells, *EMBO J.* **20**, 3008–3017.

Walter, S., Fassbender, K., Gulbins, E., Liu, Y., Rieschel, M., Herten, M., Bertsch, T., Engelhardt, B. (2002) Glycosylation processing inhibition by castanospermine prevents experimental autoimmune encephalomyelitis by

interference with IL-2 receptor signal transduction, *J. Immunol.* **132**, 1–10.

Wells, L., Vosseller, K., Hart, G.W. (2001) Glycosylation of nucleocytoplasmic proteins: signal transduction and *O*-GlcNAc, *Science* **291**, 2376–2378.

Yang, X., Su, K., Roos, M.D., Chang, Q., Paterson, A.J., Kudlow, J.E. (2001) *O*-linkage of *N*-acetylglucosamine to Sp1 activation domain inhibits its transcriptional capability, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6611–6616.

Yang, X., Zhang, F., Kudlow, J.E. (2002) Recruitment of *O*-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein *O*-GlcNAcylation to transcriptional repression, *Cell* **110**, 69–80.

Ying, S.C., Gewurz, A.T., Jiang, H., Gewurz, H. (1993) Human serum amyloid P component oligomers bind and activate the classical complement pathway via residues 14–26 and 76–92 of the A chain collegen-like region of C1q, *J. Immunol.* **150**, 169–176.

Yuen, C.-T., Chai, W., Loveless, R.W., Lawson, A.M., Margolis, R.U., Feizi, T. (1997) Brain contains HNK-1 immunoreactive *O*-glycans of the sulfoglucuronyl lactosamine series that terminate in 2-linked or 2,6-linked hexose (mannose), *J. Biol. Chem.* **272**, 8924–8931.

Zúñiga, E.I., Rabinovich, G.A., Iglesias, M.M., Gruppi, A. (2001) Regulated expression of galectin-1 during B cell activation and implications for T-cell apoptosis, *J. Leukocyte Biol.* **70**, 73–79.

Zwickl, P., Seemuller, E., Kapelari, B., Baumeister, W. (2001) The proteasome: a supramolecular assembly designed for controlled proteolysis, *Adv. Protein Chem.* **59**, 187–222.

Gonad Differentiation Genetics

Stephen S. Wachtel
Division of Reproductive Genetics, Department of Obstetrics and Gynecology, University of Tennessee, Memphis, TN, USA

1 Introduction 609

2 Search for the Testis-determining Gene 610
2.1 The H-Y Antigen Gene (*HYA*) 610
2.2 Bkm Satellite DNA 611
2.3 Zinc Finger Y Gene (*ZFY*) 612
2.4 Sex-determining Region Y Gene (*SRY*) 615
2.4.1 *SRY* in the Human 615
2.4.2 *SRY* in the Mouse 617
2.4.3 *SRY* in Other Species 618
2.4.4 The SRY Protein 619

3 X-linked Sex-determining Genes 620
3.1 Scandinavian Wood Lemming 620
3.2 Genealogical Evidence in the Human 621
3.3 Mapping Studies: The *DAX1* Gene 621

4 Autosomal Sex-determining Genes 624
4.1 Genealogical Evidence in the Human 624
4.2 Mapping Studies: The *SOX9* Gene 624
4.3 Other Autosomal Sex-determining Genes 625
4.3.1 The *SF1* Gene 625
4.3.2 The *WT1* Gene 625
4.3.3 The *WNT*-4 Gene 625

5 XX Sex Reversal 626
5.1 XX Males 626
5.1.1 In the Human 626
5.1.2 In the Mouse 627

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

5.1.3 In the Goat 627
5.1.4 In the Dog 628
5.2 XX True Hermaphrodites 628
5.2.1 In the Human 628

6 XY Sex Reversal 629
6.1 XY Females 629
6.1.1 In the Human 629
6.1.2 In the Mouse 630
6.2 XY True Hermaphrodites in the Human 632

7 Summary and Conclusions 632

Bibliography 633
Books and Reviews 633
Primary Literature 633

Keywords

Sex Determination
The process by which the sexually undifferentiated embryo is induced to become male or female.

Gonad
The sex gland, testis or ovary, which produces the haploid gametes, and secretes the hormones that determine appearance of the secondary sex traits.

Testis-determining Gene
The Y-chromosomal gene that governs male sex determination.

Sex Reversal
Any of several conditions in which gonadal sex is at variance with chromosomal sex or body sex.

The manner in which the indifferent embryonic gonad is induced to become testis or ovary is a prime question in developmental biology, and may have more general relevance for ontogeny as a whole. Identification of a pathway of sex-determining genes and their products will provide useful insights into the nature of sex differentiation and the various syndromes of abnormal gonadal development, as well as new approaches to the diagnosis and study of these conditions. Differentiation of the gonad is the key event leading to development of the sex phenotype. In males, this event is governed by the Y chromosome testis-determining gene, and other, recently discovered, X-linked and autosomal genes. In females, in the absence of the Y chromosome, alternative pathways are set in motion.

1 Introduction

In human embryos, as in mammalian embryos in general, there are two sets of genital ducts: (1) the Wolffian ducts, which give rise to the epididymides, vasa deferentia, and seminal vesicles, and (2) the Müllerian ducts which give rise to the fallopian tubes, uterus, and cephalad vagina. In males, the Wolffian ducts develop and the Müllerian ducts regress, and in females, the Müllerian ducts develop and the Wolffian ducts regress. Because the internal male ducts appear and the Müllerian ducts regress just after differentiation of the testis, the question was asked whether the testis might play some role in those events.

In the early 1940s, a young French scientist named Alfred Jost set out to answer that question. And just after World War II, he described a series of experiments that mark a turning point in the biology of sex differentiation. When Jost castrated rabbit fetuses at day 19 *in utero*, they went on to develop as females regardless of the sex of the gonad that he had removed. He inferred that the inherent tendency of the mammalian embryo is to develop as a female, and that maleness is imposed by the testis against that tendency.

To clarify the role of the male gonad in the growth of the male genital ducts, Jost grafted a fetal testis next to a fetal ovary in a 20-day rabbit embryo. On inspecting the embryo at 28 days (just before birth), he noted inhibition of the Müllerian duct on the side of the grafted testis and "partial maintenance" of the Wolffian duct. When a small crystal of testosterone propionate was substituted for the implanted testis, he again noted the maintenance of the Wolffian duct, but this time there was no inhibition of the Müllerian duct.

It was thus demonstrated that the fetal testis produces two factors – one that induces development of the Wolffian derivatives, and one that blocks development of the Müllerian derivatives. Today, we know that the former is testosterone and the latter is the anti-Müllerian hormone (AMH) (also called *Müllerian inhibiting substance* or *MIS*). In normal male fetuses, some of the testosterone is converted to the more potent androgen, dihydrotesterone (DHT), which mediates the timely appearance of the male external genitalia.

On the basis of his observations, Jost summarized the fundamental mechanism of mammalian sex differentiation in this paradigm. *Sex differentiation involves a series of orderly processes: establishment of genetic sex at conception, translation of genetic sex into gonadal sex, and translation of gonadal sex into body sex.*

The gene that sets these processes in motion is located on the Y chromosome (Fig. 1). In the presence of the Y, the gonad becomes a testis; the testis secretes testosterone and AMH, and further development is male. In the absence of the Y, the gonad becomes an ovary; testosterone and AMH are not secreted, and further development is female.

In this article, we describe the search for the Y-situated testis-determining gene, with emphasis on the molecular methods leading to its isolation and characterization in 1990. We continue with a discussion of autosomal and X-linked sex-determining genes, and we conclude with a survey of the extraordinary conditions that can result when any of the sex-determining genes are mutated or expressed abnormally.

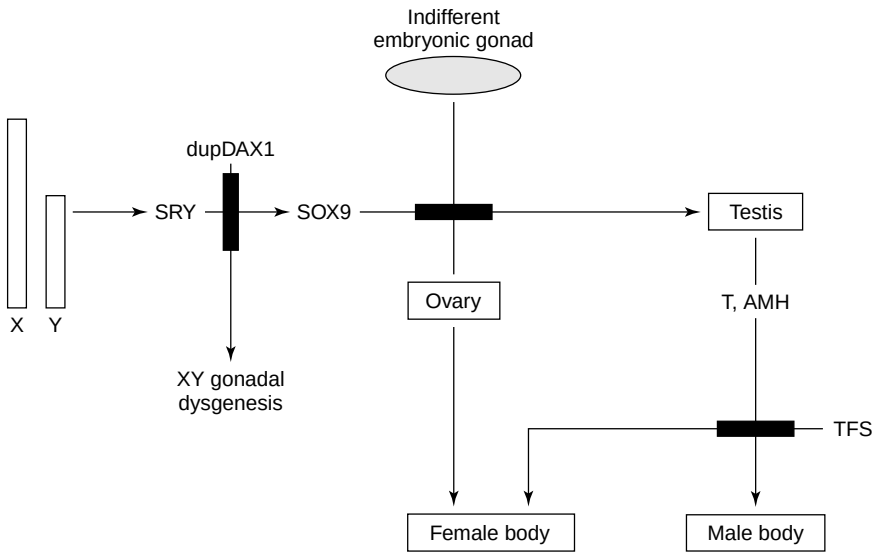

Fig. 1 Sex determination in the human. In XX embryos, the indifferent embryonic gonad becomes an ovary and the body develops along female lines. In the presence of the Y-chromosome-situated *SRY* gene, the indifferent gonad becomes a testis. The testis secretes testosterone (T), which induces the wolffian ducts, and anti-Müllerian hormone (AMH), which blocks formation of the Müllerian ducts; in this case, the body develops along male lines. Response to testosterone and its metabolite, dihydrotestosterone (DHT), is mediated by the androgen receptor, encoded by a gene on the X chromosome. When this gene is mutated or absent, the embryo develops as a female despite the presence of Y chromosome and testes. The resulting condition is called *testicular feminization syndrome* (TFS). Duplication of another X-linked gene, *DAX1*, blocks the testis-determining function of *SRY*, and the result is XY gonadal dysgenesis (see Fig. 7). The *SOX9* gene, on chromosome 17, is a downstream target of *SRY*. Duplication of *SOX9* in the absence of the Y chromosome causes development of XX testes (not indicated in the figure).

2 Search for the Testis-determining Gene

2.1 The H-Y Antigen Gene (*HYA*)

It is a dictum of biology that genes of fundamental significance are conserved and thus relatively unchanged from species to species. On this basis, a number of genes have been proposed as candidates for the Y-linked testis-determining gene. One such candidate was the gene for the male-specific H-Y antigen.

In inbred strains of the mouse, male grafts are rejected by female recipients, whereas grafts exchanged among the other sex combinations are accepted. This happens because male tissues contain a cell-surface antigen, H-Y, encoded by a Y-chromosomal gene that has no counterpart in females. When challenged with male grafts, female mice also produce antibodies that can react with male cells from a broad spectrum of mammalian species. In view of this phylogenetic stability, it was proposed in 1975 that H-Y antigen was the product of the testis-determining gene.

To test the idea, H-Y antigen was studied in cases of abnormal sex differentiation – especially in cases of sex reversal

such as XX male syndrome, XY gonadal dysgenesis, and true hermaphroditism. If H-Y were the product of the testis-determining gene, it should always be present in association with testicular tissue, regardless of chromosomal sex or body sex. For nearly a decade, the proposal was supported experimentally, but in 1984, a line of XX male mice lacking the H-Y antigen was discovered, and the search for the testis-determining gene was redirected.

[In 1995, the gene for human H-Y antigen (*HYA*) was mapped to the long arm of the Y chromosome, and the gene for mouse H-Y *(Hya)*, to the *short* arm of the mouse Y, between *Zfy-1* and *Zfy-2* (see Sect. 2.3). But the function of H-Y remains to be ascertained.]

2.2 Bkm Satellite DNA

A satellite is a repetitive sequence of chromosomal DNA in which the nucleotide pairs G:C and A:T are represented unequally. This gives the satellite a density that is different from the density of sequences in which the nucleotide pairs are equally represented (Fig. 2). And this enables geneticists to identify and isolate satellite DNA from the main body of genomic DNA by centrifugation along a density gradient. A common repetitive sequence in satellite DNA is the tetranucleotide, GATA, which is conserved in evolution.

Satellites occur in the heterochromatic regions of the chromosomes, which are in general genetically inert. They are common, for example, in the long arm of human Y chromosome. In snakes, the female is the heterogametic sex – that is, the female carries the ZW chromosome pair, corresponding to the XY pair in mammals. Like the mammalian Y, the snake W contains heterochromatic regions rich in satellites.

In fact, several satellites have been isolated from the W chromosome of the poisonous snake, *Bungarus fasciata*, the banded krait. One such

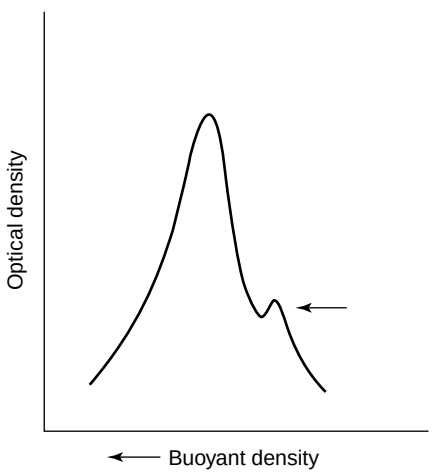

Fig. 2 Satellite DNA. The buoyant density of double-stranded DNA – determined by centrifugation through a cesium chloride density gradient – is a function of G–C content. The larger peak above includes about 90% of the genome, centered about a value of 1.70 g cm^{-3} corresponding to a mean G–C content of 42%, which is typical of mammals. The smaller peak (arrow), representing about 10% of the genome, is a satellite DNA with a G–C content of some 30% and a buoyant density of 1.69 g cm^{-3}. Satellites may have densities greater or smaller than the density of the main band. (From Wachtel, S.S., Tiersch, T.R. (1994) The Search for the Male-determining Gene, in: Wachtel, S.S. (Ed.) *Molecular Genetics of Sex Determination*, Academic Press, San Diego, CA, pp. 1–22).

satellite – consisting mostly of GATA repeats – was called *banded krait minor satellite DNA*, or *Bkm*, for short. The *Bkm* repeat sequence was found to be a regular component of the W chromosome in a wide variety of ophidian species.

Bkm was not limited to snakes. It was found in other reptiles, and in birds, mammals, and even insects. Although present in autosomes, and hence in both sexes of the mouse, the GATA repeats were concentrated in the sex-determining region of the Y chromosome. This prompted researchers to ask whether *Bkm* might be involved in mammalian sex determination.

In addressing that question, Lalji Singh and Kenneth Jones of the University of Edinburgh used *Bkm* as a probe for Y chromosome sequences in XX*Sxr* males. The *Sxr* "mutation" causes XX embryos of the mouse to develop as phenotypic males with small sterile testes. It is transmitted by fertile XY*Sxr* carrier males to half their XX offspring (Sect. 5.1.2). In 1982, it was generally assumed that male development in XX*Sxr* mice was due to a mutant autosomal gene, or alternatively, due to a submicroscopic Y-autosome translocation. But Singh and Jones found a concentration of *Bkm* in one of the X chromosomes of the sex-reversed males, similar to the cluster that is normally found in the Y. They showed that *Sxr* sex reversal is due to a nonreciprocal recurrent Y-X translocation of the sex-determining gene in the germ line of the carrier male. Because the translocation was so small as to have escaped detection by cytogenetic analysis, *Bkm* and the testis-determining gene appeared to be closely linked, and perhaps identical.

This was a pivotal work, because it resolved a fundamental mechanism of sex reversal in the mouse, and led the way to similar studies in XX males of the human (see Sect. 5.1.1). Yet, later studies showed that *Bkm* was not likely to be the testis-determining gene in man, because GATA sequences were present throughout the human genome and there were no concentrations of *Bkm* in the human Y. Once again the search for the testis-determining gene was redirected, but with the work on *Bkm*, the study of vertebrate sex determination had passed from the purview of immunogenetics into the realm of molecular biology.

2.3 Zinc Finger Y Gene (*ZFY*)

With respect to sex-reversed conditions such as XX male syndrome and XX true hermaphroditism (Sect. 5), the question arises why one XX embryo should develop unambiguous testes (an XX male), and another, ovaries and testes in some combination – for example, ovotestes or ovaries and testes (an XX true hermaphrodite). In 1966, Malcom Ferguson-Smith proposed that both conditions could arise as a result of X-Y translocation involving transfer of the testis-determining gene from the Y chromosome to the X (Fig. 3).

According to that proposal, the sex of an XX embryo carrying the X;Y translocation would depend on variable patterns of X chromosome inactivation *(lyonization)*. If the X;Y translocation chromosome (the one bearing the testis-determining gene) were preferentially activated in the developing gonads, the embryo would become a male. If it were activated in one gonad and inactivated in the other, the embryo would become a true hermaphrodite with testis versus ovary. An ovotestis would develop if the X;Y chromosome were activated in one part and inactivated in another part of a

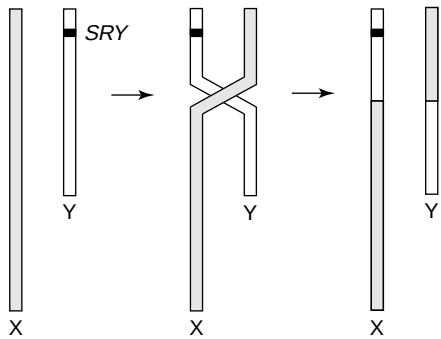

Fig. 3 Translocation of the male-determining gene resulting from crossover of the X and Y chromosomes during meiosis. If the translocated fragment contains the testis-determining gene (*SRY*), and if an egg is fertilized by a sperm cell carrying the affected X chromosome, the resultant embryo becomes an XX male. X;Y translocations have been identified in about 70% of human XX males.

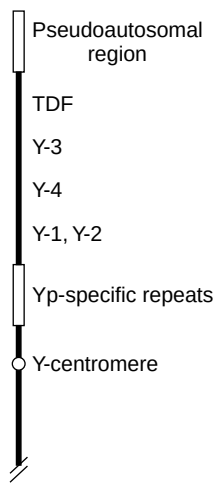

Fig. 4 Deletion map of the Y chromosome. Order of hypothetical Y-fragments based on their frequency of occurrence in XX males; see text.

single gonad – and so on. In support of the proposal of Ferguson-Smith, X-Y translocations have been observed in about 70% of XX males that have been studied.

The fact that these translocations involved unequal amounts of Y chromosome material enabled construction of Y chromosome deletion maps. The closer to the testis-determining gene a particular fragment is, the more likely that it will be included in an X;Y translocation. Thus, translocated fragments closer to the testis-determining gene should be found in a greater proportion of XX males than should fragments further from that gene.

After the enrichment and separation of populations of Y chromosomes in a flow cytometer, it is possible to digest the DNA to obtain Y-fragments of variable size. These can be used as probes to detect corresponding Y-fragments in XX males. Some of the fragments could be found more often than others in DNA from different XX males. Consider four hypothetical fragments: Y-1, Y-2, Y-3, and Y-4. Say, for example, that Y-1 and Y-2 are present in about half of the XX males surveyed, whereas Y-4 is present in three-fourths and Y-3 is present in all. It could be inferred that among these sequences, Y-3 is closest to the testis-determining gene, Y-4 is next in line, and Y-1 and Y-2 are the farthest. When sequences are localized in this way, a deletion map can be constructed (Fig. 4).

In 1987, David Page and colleagues at the Whitehead Institute used this method to produce a map of the short arm of the human Y chromosome (Yp), containing 13 deletion intervals. One of the intervals – denoted 1A2 and consisting of only 140 kb (about 0.2% of the Y chromosome) – was present in an XX male who possessed only 300 kb of Y-DNA. It was absent in a woman with a Y,22 translocation who possessed all but 160 kb of the Y. Because she was a woman, the testis-determining gene (called *TDF* in those days) was most probably situated in the part of the Y that she lacked. It was inferred that interval 1A2 contains "an essential portion if not all" of *TDF*, and that *TDF* could be further localized by looking for highly conserved sequences within 1A2.

A series of Y-probes was developed, representing collectively the entire 1A2 deletion interval. The probes were tested against samples of DNA from a wide variety of species in Southern blots *(Noah's Ark blots)*. Many of the probes identified corresponding fragments from humans and great apes, but under conditions of greater stringency, they did not hybridize with DNA from more distant species.

One probe, however, designated pDP1007, and consisting of a 1.3 kb *Hind*III fragment, was hybridized with a single-copy sequence in males from every mammalian species tested, including gorilla, rhesus monkey, owl monkey, rabbit, dog, goat, horse, and cattle. This indicated that a fragment corresponding to pDP1007 was present in the Y chromosome of each of these species. In addition to the Y-specific fragment, pDP1007 identified a fragment that was present in males and females. Because the hybridization signal of the second fragment was stronger (more dense) in females than it was in males, that fragment was assumed to be on the X chromosome.

Next, the DNA sequence of the Y-fragment was determined, and the order of amino acids in the corresponding protein was derived. When compared to other amino acid sequences, the sequence of the protein encoded by pDP1007 was found to resemble closely the sequences of proteins with tandemly repeated "finger" domains. Finger proteins are regulatory. They are known to bind DNA and thereby regulate transcription–conversion of the DNA sequence into a messenger RNA sequence.

Because pDP1007 recognized a sequence in a tiny fragment of Y-DNA present in an XX male and absent in an "XY" female, because the pDP1007 sequence was phylogenetically conserved and male-specific, and because the corresponding protein was likely to govern transcription, it was suggested that the protein identified by pDP1007 was the product of the testis-determining gene, *TDF*. According to that suggestion, the protein might bind DNA in a sequence-specific manner, and thereby regulate transcription of another gene in the pathway leading to testicular differentiation.

The gene recognized by pDP1007 was called *ZFY* (for *zinc finger Y*) because zinc ions establish complexes with its finger domains reminiscent of the complexes formed between iron porphyrin and the subunits of hemoglobin. The X-linked homolog was called *ZFX*.

Once *ZFY* was described, geneticists began to look for similar genes in species more distant than those already studied. When the pDP1007 probe was used to search for *ZFY* sequences in reptiles, positive hybridization was obtained, but the results were the same in male and female. Was there *expression* of

ZFY in both sexes? To answer that question, the pDP1007 probe was used to screen RNA extracted from embryos of the snapping turtle, in which sex is determined by the temperature of incubation (embryos that develop at higher temperatures become females; those that develop at lower temperatures become males). The probe hybridized with RNA from embryos maintained at female-determining *and* at male-determining temperatures, indicating that *ZFY* or a *ZFY*-like molecule is transcribed in both sexes in that species.

In the mouse, two *ZFY* genes were identified on the Y chromosome *(Zfy-1* and *Zfy-2)*, and one on the X *(Zfx)*. The *Zfy-1* and *Zfy-2* genes were both found in sex-reversed XX*Sxr* male mice but only *Zfy-1* was found in XXSxr^b males. Because XXSxr^b mice carry a deleted, if still male-determining, *Sxr* region, it could be inferred that *Zfy-2* is not required for differentiation of the mouse testis. Both genes were expressed during development of the mouse embryonic testis, but neither was expressed in mutant embryonic testes lacking germ cells. These findings seemed to imply that neither *Zfy-1* nor *Zfy-2* played a critical role in the development of the mouse testis, but they did not rule out a possible role for *Zfy-1* in the ontogeny of the male germ cell.

In marsupials, the pDP1007 probe hybridized with DNA from males and females, and *in situ* hybridization revealed *ZFY*-homologous sequences on an autosome. This was unexpected because in the marsupials, as in the eutherian mammals, male sex is determined by the Y chromosome. In light of these findings, it was concluded that *ZFY* could not be the determinant of male sex in marsupials. This meant that the mechanism of sex determination must be different in marsupials and eutherians, or if it were the same, that *ZFY* could not be the testis-determining gene in man.

The findings in mouse and marsupial seemed to argue against a testis-determining role for *ZFY*. Yet, there were other reasons for believing that *ZFY* could not be *TDF*. In their original paper, David Page and coworkers reported that the fragment defined by pDP1007 was absent in all of the XX true hermaphrodites that they studied and in some of the XX males. Male differentiation in these cases seemed to be due to constitutive mutation of autosomal or X-linked testis-determining genes downstream of *TDF*.

The argument for mutation of downstream genes in some cases of XX sex reversal was negated by the discovery of Y-DNA in XX males who lacked any trace of *ZFY*. One would not expect to find an XX male with an X-Y crossover *and* mutation in a downstream gene. So the discovery of four such males – each ZFY-negative and each carrying Y-specific sequences adjacent to the pseudoautosomal boundary – now seemed to eliminate *ZFY* as a candidate for *TDF*, and to outline a new region of the Y chromosome in which *TDF* must lie. A search of that region, comprising 35 kb adjacent to the pseudoautosomal boundary, yielded a "new" gene, Y-specific, and like *ZFY*, conserved across a broad spectrum of mammalian species. The new gene was called *SRY (sex-determining region Y)*.

2.4 Sex-determining Region Y Gene (*SRY*)

2.4.1 *SRY* in the Human

The *SRY* gene was discovered in 1990 by Andrew Sinclair, while he was working in the laboratory of Peter Goodfellow in London. Initially, Sinclair digested

DNA from the 35-kb region near the Y chromosome pseudoautosomal boundary, and subcloned a number of fragments, some 4 kb each. These 4-kb fragments were then cut into smaller pieces, each about 500 to 1000 bp in length (0.5–1 kb). Fifty of the fragments produced in this way were used as probes to test for corresponding fragments in DNA from males and females of the human, mouse, and bovine, and from human–hamster somatic cell hybrids containing the human X or Y chromosome. Seven of the probes identified single-copy, male-specific sequences in *Eco*R1-digested human DNA, but only one of the probes could also identify male sequences in mouse and bovine. That one was called *pY53.3*.

Because it hybridized most conspicuously with Y-DNA, a 0.9 kb *Hinc*II subclone of pY53.3 was used to assess the role of the *SRY* gene (Fig. 5). Initially, the subclone was used to probe a series of Noah's Ark blots containing DNA from male and female of several diverse species, including the human, chimpanzee, rabbit, pig, horse, bovine, and tiger. In each case, a unique male-specific sequence was identified; at reduced stringencies, additional *SRY*-like fragments were identified in male and female, indicating the presence of X-linked or autosomal *SRY*-like sequences.

In humans, the pY53.3 subclone identified a 2.1 kb male-specific fragment. When the nucleotide sequence of the human male-specific fragment was evaluated, two open reading frames were discovered. The longer of the two was found to encode 223 amino acids with a "striking similarity" to part of the Mc mating-type protein of the fission yeast, *Schizosaccharomyces pombe*, and to a conserved DNA-binding motif common among the nuclear high-mobility-group proteins, HMG1 and HMG2. Earlier, we noted that genes of fundamental significance are conserved in evolution (Sect. 2.1). The extreme evolutionary conservation of the *SRY* gene was thus consistent with its proposed critical function.

The *SRY* conserved motif – spanning 80 amino acids – was now compared with the corresponding motif of the rabbit. Sixty-four (80%) of the 80 conserved amino acids were also present in the rabbit, but only 54% similarity was revealed outside the conserved motif. In other studies, the 0.9-kb fragment of probe pY53.3 was used to test for corresponding RNA transcripts in human ovary and testis, and in male lung and kidney. A 1.1-kb transcript was

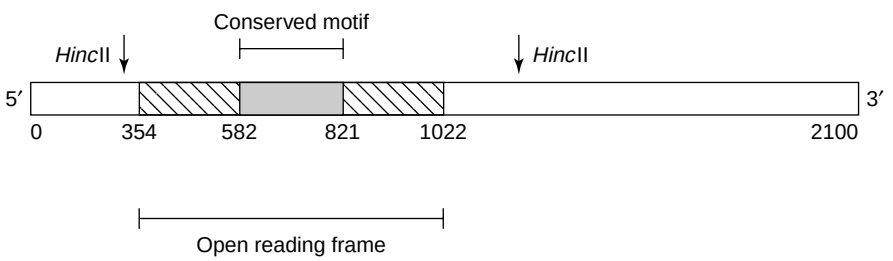

Fig. 5 Graphic representation of the 2.1 kb probe, pY53.3, which identifies *SRY*. Numbers indicate base pairs from the 5′ end; see text. [Modified after Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovell-Badge, R., Goodfellow, P.N. (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif, *Nature* **346**, 240–244].

identified in the testis, but not in the other male tissues, in Northern blots.

The foregoing observations – male specificity, extreme phylogenetic conservation, and testis-specific expression – seemed to confirm the proposition that *SRY* and *TDF* were the same. Yet there was room for other genes in the 35-kb pseudoautosomal fragment containing *SRY*, and it remained to be demonstrated unambiguously that *SRY* was the Y-situated gene that governed testicular differentiation. This could be done in either of the two ways: by showing that SRY was absent or mutated in sex-reversed XY women, or by showing that transgenic insertion of *SRY* could induce testicular differentiation in genetically female (XX) embryos, in the mouse for example. In fact, both conditions were met (see Sect. 6.1 and Sect. 2.4.2).

2.4.2 *SRY* in the Mouse

The mouse homolog of *SRY* was described by John Gubbay and colleagues in a paper that appeared in July 1990, in *Nature*, back to back with the paper of Sinclair et al. When the human pY53.3 was used to probe Southern blots containing DNA from male and female mice, a 3.5 kb male-specific fragment was found. This strongly hybridizing fragment was found not only in normal XY males but also in sex-reversed XX*Sxr* and XXSxr^b males.

Sxr^b, a deleted *Sxr* (Sxr^a), was the smallest fragment of the mouse Y chromosome known to contain the murine testis-determining gene corresponding to *TDF* of the human. At the time (1990), only *Zfy-1* and repetitive sequences such as *Bkm* had been identified within Sxr^b, in addition to the pY53.3-hybridizing sequence. Thus, pY53.3 had now defined a novel single-copy sequence corresponding to *SRY*, in the smallest region of the mouse Y chromosome known to be male-determining.

Further analysis revealed a mouse clone, p4.2.2 containing a 380 bp mouse fragment that hybridized at high stringency only to the Y-specific 3.5 kb band. This indicated that clone p4.2.2 contained the mouse homolog of pY53.3. When the clone was sequenced in the region of homology with pY53.3, a 62% similarity was observed, and a long open reading frame corresponding to one of the reading frames in *SRY* (and to another open reading frame in the rabbit homolog of *SRY*). These observations signaled the presence of a new gene, which was called *Sry*.

When *Sry* was evaluated further, a 237-bp fragment was found within the open reading frame, which had an 80% sequence similarity to the corresponding fragment within pY53.3. As with its human homolog, the mouse *Sry* gene exhibited a remarkable similarity to the conserved 80-amino acid motif of the fission yeast *S. pombe*, an extraordinary finding given the phylogenetic distance between yeast and mammals. And like its human counterpart, *Sry* was shown to encode a protein with a DNA-binding motif. Moreover, *Sry* RNA transcripts were detected in the urogenital ridge of 10.5-day-old male embryos – just before differentiation of the testis – but not in the corresponding tissue of 10.5-day-old female embryos.

Compelling evidence for a sex-determining role of *Sry* came a year later, when the gene was found to induce testicular differentiation if introduced as a "transgene" into XX embryos of the mouse. A 14-kb fragment containing *Sry* was injected into fertilized eggs and the eggs were transferred into the oviducts of pseudopregnant females. The resulting embryos were recovered at 14 days and the

gonads scored histologically. Most of the 158 embryos obtained in that way were normal XY males or XX females, but 8 of the XX embryos were transgenic, and of those, 2 had testes. Both lacked *Zfy-1*; both carried "many copies of *Sry*".

In other experiments, injected embryos were allowed to develop to term. Among the 49 males and 44 females that were born, Southern blotting revealed 5 that were transgenic. Two were XY males and two were XX females (later found to be fertile). The last was a sex-reversed XX male. In May 1991, that mouse was pictured on the cover of *Nature*.

Thus, the search for the Y-situated testis-determining gene had come to an end. In summary, about 25% of the transgenic mice – those in which *Sry* had been incorporated – developed as males. This was due to *Sry*, because sequencing revealed no genes other than *Sry* in the injected 14-kb fragment. Human *SRY* did not function as a sex determinant when it was injected into the mouse embryos. But this was not surprising given the degree of sequence divergence (23 of 79 amino acids) within the DNA-binding motif of the human and mouse homologs. Yet, *SRY* was shown to be the human testis-determining gene, because mutation within the *SRY* conserved motif was found in cases of human XY sex reversal (see discussion in Sect. 6).

2.4.3 *SRY* in Other Species

Location of *SRY* in marsupials is of interest because, as noted earlier, the *ZFY* gene is autosomal in marsupials. The marsupial mammals (Metatheria) diverged from placental mammals (Eutheria) some 100 to 150 million years ago. Since the Y chromosome is sex-determining in both groups, it would seem likely that the testis-determining gene of each is derived from a common ancestral gene. Conservation of a common Y-specific sequence might thus be expected.

On this basis, a group of Australian and British geneticists set out to identify the marsupial homolog of SRY by using the 0.9 kb subclone of the human pY53.3 probe. Southern blots were prepared from the DNA of two marsupial species representing major orders that diverged some 50 million years ago: *Macropus eugenii*, the tammar wallaby, and *Sminthopsis macroura*, the stripe-faced dunnart, a small mouse-like carnivore. Male-specific fragments were detected in each of the two species, and in addition, a number of fragments shared by male and female.

Next, the pY53.3 probe was used to screen a cDNA library constructed from testis RNA of *S. macroura*. A positive clone was obtained and used to probe Southern blots for *SRY* homologs. It hybridized at high stringency to DNA from males but not from females, indicating that it had been derived from a Y-specific transcript.

Further study pointed to the existence of a Y-specific marsupial gene having a region of homology with the human *SRY* gene, and an open reading frame encoding an HMG type box of 79 amino acids (82 and 86% similarity with human *SRY* within the conserved domains of *S. macroura* and *M. eugenii*, respectively). Thus, unlike *ZFY*, the marsupial *SRY* gene appeared to be on the Y chromosome; this was later confirmed by *in situ* hybridization. These findings provided evidence for a monophyletic origin of the mammalian Y chromosome, and supported the notion that the sex-determining gene is the same, or at least similar, in eutherian and marsupial mammals.

SRY is not limited to the mammals. Using a probe from the conserved motif

of the human *SRY* gene, geneticists at the University of Tennessee in Memphis screened genomic DNA from 23 different species representing 5 vertebrate classes, 400 million years of vertebrate evolution, and several alternative modes of sex determination. Among the species studied were the chicken, in which the female is the heterogametic sex (ZW), the alligator, in which sex is determined by temperature, and the sex-changing wreckfish, *Anthias squamipinnis*, in which all fish begin life as females (Table 1). In Southern blots, positive hybridization was observed in every case – in males and females. Although there was no sex-specific reaction outside of the mammals, and the function of *SRY*-like sequences in the nonmammals remained to be determined, these findings underscored the importance of the *SRY* sequence.

2.4.4 The SRY Protein

In the adult human testis, the predominant *SRY* transcript derives from a single exon. The transcript includes the open reading frame for a protein of 204 amino acids. Initial evaluation of the *SRY* nucleotide sequence revealed that it could encode a protein with DNA-binding properties. As noted in Sect. 2.4.2, the 80-amino acid conserved motif within the SRY protein is common among certain proteins known to have DNA-binding activity – namely, the high-mobility-group proteins. These proteins bind to DNA in a nonspecific manner, although variants can function as sequence-specific transcription factors.

The ability of the SRY/Sry protein to bind DNA has been demonstrated in several studies. For example, IRE-ABP is a rat protein that complexes to the IRE-A insulin response element

Tab. 1 Phylogenetic conservation of *ZFY* and *SRY*.

Class	*Species*	*Mechanism of sex determination*	*ZFY hybridization in*	*SRY hybridization in*
Mammalia	Human	XX/XY	Male (female)[c]	Male
	Mouse	XX/XY	Male (female)[c]	Male
	Kangaroo	XX/XY	Male, female	Male
	Wallaby	XX/XY	Male, female	Male
Aves	Chicken	ZZ/ZW	Male, female	Male, female
	Cockatiel	ZZ/ZW	Male, female	NT
Reptilia	Leopard gecko	TSD	Male, female	NT
	Alligator	TSD	Male (female?)[d]	Male, female
	Cottonmouth	ZZ/ZW	Male, female	NT
	Crotalus atrox[a]	ZZ/ZW	NT	Male, female
Osteichthyes	Catfish	XX/XY	Male, female	Male, female
	Wreckfish[b]	Sex changing	Male, female	Male, female
Agnatha	Lamprey	?	NT	Male, female

[a]Western diamondback rattlesnake.
[b]*Anthias squamipinnis:* These fish begin life as females; some change to males.
[c]The related sequence, *ZFX (Zfx)*, is found on the X chromosome in these species.
[d]The female was not tested.
Notes: TSD: Temperature-dependent sex determination; NT: Not tested.
Source: After Wachtel, S.S., Tiersch, T.R. (1994) The Search for the Male-determining Gene, in: Wachtel, S.S. (Ed.) *Molecular Genetics of Sex Determination*, Academic Press, San Diego, CA, pp. 1–22.

of the glyceraldehyde-3-phosphate dehydrogenase gene. The IRE-ABP protein contains a single HMG box with 67% similarity to the corresponding HMG box in mouse *Sry* and 98% similarity to the HMG box in mouse Sry-like protein, a4. Fusion proteins incorporating the HMG of Sry or IRE-ABP, or both, bind to the IRE-A target site, 5′-TTCAAAC-3′; binding is inhibited by mutations within the Sry fusion protein.

The human SRY protein resembles two T-cell proteins: TCF-1 and TCF-1α. The former binds to the sequence 5′-AACAAAG-3′, which occurs in several T-cell genes, including the CD3ε enhancer (the AACAAAG sequence also occurs upstream of the *SRY* and *Sry*γ genes). Geneticists in England therefore induced production of SRY protein in *Escherichia coli* and Sf9 insect cells, and tested the protein for ability to bind the TCF-1 target sequence, AACAAAG, and variants of the sequence. SRY protein was incubated with double-stranded oligonucleotides, and the resulting complexes were studied in an electrophoresis mobility shift assay. Wild-type SRY formed complexes with oligonucleotides containing AACAAAG, but binding was negligible when that sequence was replaced with CCGCGGT or when mutant SRY from XY women was substituted. Binding did not occur when single-stranded DNA was used.

Subsequently, it was found that SRY protein, like HMG1, recognizes four-way junctions regardless of their sequence. Four-way junctions, rare structures induced in DNA by recombination or by intrastrand base pairing of inverted repeat sequences, resemble the four arms of an X in a plane, with two angles of about 120° and two angles of about 60°. Moreover, a sharp bend is produced in linear DNA when SRY binds to the AACAAAG sequence. Thus, the interaction between HMG boxes and DNA seems to involve specific structures.

In December 1994, a group headed by Patricia Donahoe and Michael Weiss reported that SRY protein could activate transcription of the *AMH* gene. Transcription was found to be secondary to partial side-chain insertion of the *SRY* conserved motif in a widened minor groove. This implied that expression of *AMH* was dependent on a normal function of *SRY*, or in other words, that *AMH* was a downstream component of the *SRY* pathway.

Given the phylogenetic conservation of *SRY* and the apparently identical function of *SRY* and *Sry*γ genes in man and mouse, it is remarkable that human and mouse SRY proteins exhibit disparate binding properties. Whereas the HMG domains of human and mouse SRY (hSRY and mSRY) both recognize the sequence 5′-AACAAAG-3′, mSRY recognizes another sequence, 5′-AACAATG-3′, with significantly greater affinity. Moreover, in situations in which hSRY and mSRY react with the same nucleotide sequences, the mouse and human HMG domains may exhibit different flexing (bending) properties. And on binding to DNA, mSRY develops prominent major groove contacts compared to hSRY. These findings and sequence differences outside the HMG box could explain why human SRY does not induce testicular differentiation in transgenic XX mice.

3 X-linked Sex-determining Genes

3.1 Scandinavian Wood Lemming

The wood lemming, *Myopis schisticolor*, is a small mouse-like rodent that inhabits the

mossy grasslands of Scandinavia. Unlike its cousin, *Lemmus lemmus*, notorious for its frenetic migrations, this quiet species is remarkable for a 4 : 1 sex ratio, favoring the female. There are two kinds of female: one has two X chromosomes, and the other has an X and a Y. The XY females are fertile and anatomically indistinguishable from their XX sisters.

Cytogenetic analysis reveals a normal Y chromosome in the XY females. In fact, it is a modified X that causes sex reversal. There are two kinds of X chromosome in the wood lemming, identifiable by differential Giemsa-banding in the short arm (Xp). One, denoted X, is found in XY males; the other, denoted X*, is found in X*Y females. Since XO rodents become fertile females, it is not surprising that X*Y females of the wood lemmings are fertile.

These observations point to the occurrence in wood lemmings of an X-linked gene, perturbation of which can block the function of the Y chromosome. But according to Ohno's Law, a gene that is on the X chromosome in one mammalian species will be on the X chromosome in other mammalian species. This implies occurrence of an X-linked sex-reversing gene in species such as the human.

3.2 Genealogical Evidence in the Human

In Sect. 6.1.1, we describe the syndrome of *XY gonadal dysgenesis*, which results from a failure of the testis-determining pathway in XY fetuses. Affected fetuses develop ovaries initially, but in humans, two normal X chromosomes are needed to sustain ovarian differentiation. In 45,X and 46,XY human embryos – those lacking a second X chromosome – the ovaries degenerate and are represented at around the time of birth by "streak gonads," which are functionally inert. Women with XY gonadal dysgenesis are accordingly sterile and, in general, eunuchoid.

As might be expected, some of these cases are due to mutation within the *SRY* gene, but familial clusters are known in which the condition is transmitted as an X-linked trait (Fig. 6). This seems to vindicate the requirement of Ohno's Law, yet there is even more compelling evidence for an X-linked sex-reversing gene.

3.3 Mapping Studies: The *DAX1* Gene

In 1980, an anomalous additional band was found in the short arm of the X chromosome in a profoundly retarded girl with cleft palate, skeletal malformations, and various other congenital anomalies. The girl had an XY chromosome pair. The karyotype was designated 46,dup(X)(p21→pter),Y, signifying a duplication extending from Xp21 to the Xp terminus. When the girl died at 5 years of age, postmortem studies revealed small uterus and fallopian tubes, and ovarian-like stroma containing a few degenerative follicles. There was no sign of testicular tissue. And the external genitalia were those of a normal female.

The abnormal X chromosome was inherited from the mother (Fig. 7). When the mother became pregnant again, amniocentesis revealed a fetal XY karyotype with the abnormal X chromosome. The pregnancy was terminated at 20 weeks. The fetus was a female with multiple congenital defects similar to those of the propositus. External and internal genitalia resembled those of a normal female fetus of that age; the gonads were fetal ovaries.

Other cases of XY sex reversal involving Xp duplication have been reported. In 1994, a group of American geneticists

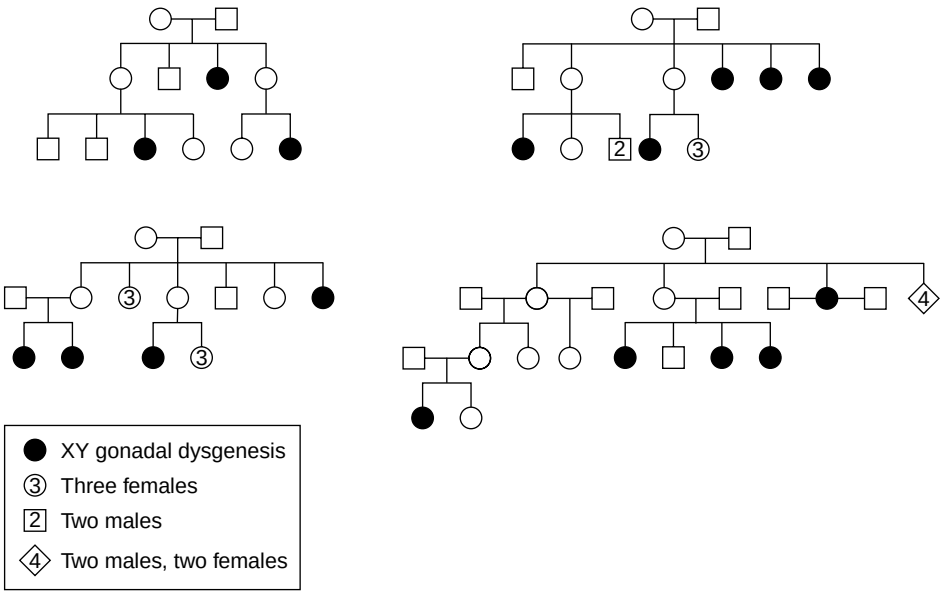

Fig. 6 Familial aggregates of XY gonadal dysgenesis consistent with transmission of an X-linked gene. (From Wachtel, S.S., Tiersch, T.R. (1994) The Search for the Male-determining Gene, in: Wachtel, S.S. (Ed.) *Molecular Genetics of Sex Determination*, Academic Press, San Diego, CA, pp. 1–22).

headed by Pamela Arn described maternal half siblings with ambiguous external genitalia and a small duplication in Xp: in both children, the karyotype was 46,X,dup(X)(p21.2→p22.11),Y, signifying duplication of the region Xp21.2→p22.11. Except for undifferentiated gonads and external genitalia, these children manifested none of the abnormalities found in other sex-reversed patients with duplications in the Xp21 region. This could be explained by the limited size of the duplication in these cases.

In light of these findings, a "new" sex-reversing gene, *SRVX (sex reversal X)* was assigned to the region Xp21.2→p22.11, comprising a few megabases distal to the ornithine transcarbamalase (*OTC*) locus. It was inferred (i) that perturbation of this gene, as in cases of Xp21 duplication, could block male differentiation of the gonad, despite presence of an intact *SRY* sequence; (ii) that some cases of XY gonadal dysgenesis could be due to *submicroscopic* duplication involving the *SRVX* locus; and (iii) that the new gene was part of a pathway of testis-determining genes that includes *SRY*.

By evaluation of overlapping duplications, Italian geneticists narrowed the locus of the new gene to a 160-kb region of Xp21 adjacent to the adrenal hypoplasia congenita locus and within roughly 100 kb of *DXS319*. They called the gene *DSS* for *dosage sensitive sex reversal*, and confirmed that having two active copies of the gene blocks male development in XY embryos. Since the gene was found to be absent in certain males, it could not be required for differentiation of testes, and thus could not be a regular part of the male-determining pathway.

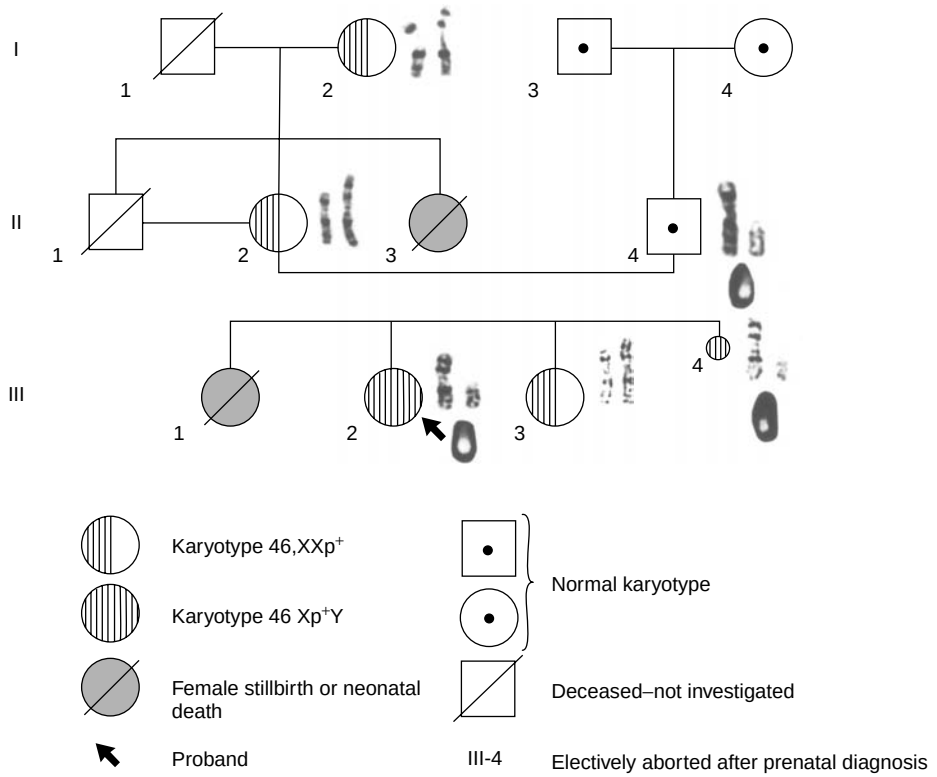

Fig. 7 XY sex reversal caused by inheritance of duplicated X chromosome. The duplicated band was found in the short arm of the X chromosome in the maternal grandmother (I-2), mother (II-2), and normal sister (III-3) of the proband, and in the proband (III-2, arrow) and her sex-reversed fetal sibling (III-4). The father's chromosomes (II-4) were normal. Giemsa-banded chromosomes with quinacrine-stained inserts. [From Bernstein, R., Koo, G.C., Wachtel, S.S. (1980) Abnormality of the X chromosome in human 46,XY female siblings with dysgenetic ovaries, *Science* **207**, 768–769].

Today, the relevant gene is called *DAX1* for *dosage sensitive sex reversal-adrenal hypoplasia congenita-critical region of the X chromosome, gene 1*. The *DAX1* gene is one of the nuclear hormone receptor superfamily. It encodes a nuclear protein that binds DNA and regulates transcription mediated by the retinoic acid receptor. Mutations in *DAX1* cause adrenal hypoplasia congenita but do not affect sex determination in human males. In mice, XX embryos lacking both copies of the *Dax1* gene become females (evidently *DAX1* is not an essential component of the sex-determining pathway in males or females).

In mouse embryos, *Dax1* is expressed in the somatic cells of the indifferent gonad regardless of karyotype, but it is downregulated in the developing testis. These observations suggest that, in addition to its endocrinological function, *DAX1* acts as an antagonist to *SRY*. Thus, duplication of *DAX1* blocks *SRY*-mediated differentiation of the embryonic testis.

4
Autosomal Sex-determining Genes

4.1
Genealogical Evidence in the Human

The syndrome of XY gonadal dysgenesis can occur in familial clusters, and can be caused by transmission of an X-linked gene. In some families, however, the frequency of affected relatives is consistent with the frequency that would be expected if the condition were inherited as a male-limited autosomal trait. It seems reasonable to assume that autosomal genes are involved in the etiology of at least some women with XY gonadal dysgenesis, because most cases are sporadic, and *SRY* mutations are found in a minority of XY women (20% or less). Moreover, if *SRY* encodes a transcription factor, it should activate other genes in the testis-determining pathway (Table 2). And *SRY/Sry* itself may be activated by other genes (see Sect. 7).

Tab. 2 Genes of the human male sex-determining cascade.

Gene	*Locus*	*Presumptive function(s)*
SRY	Yp11.3	Initiator of the testis-determining pathway
SOX9	17q24.1-q25.1	Sertoli cell differentiation,[a] positive regulation of AMH
WT1	11p13	Tumor suppressor; formation of indifferent gonad, testis (?)
SF1	9p33	Formation of indifferent gonad, testis (?) regulation of AMH
DAX1	Xp21.2	Antagonist of *SRY*
WNT-4	1p35	Formation of indifferent gonad

[a] Evidently, *SOX9* is an immediate downstream target of *SRY*.

4.2
Mapping Studies: The *SOX9* Gene

Campomelic dysplasia is a human syndrome characterized by bowing of the long bones, dwarfism, cleft palate, and various other congenital malformations. Owing to the severity of the condition, affected children usually die within a few weeks or months of birth. Patients with campomelic dysplasia can be XX or XY, but some two-thirds of those with XY are sex reversed. A typical case might involve a stillborn XY female with undifferentiated gonads, slightly enlarged clitoris, and normal uterus and fallopian tubes.

Study of *de novo* reciprocal translocations of chromosome 17 in three different patients with this condition allowed geneticists to assign the genes for campomelic dysplasia and associated sex reversal to the chromosome region 17q24.3→q25.1. Fluorescence *in situ* hybridization with chromosome painting probes showed that the translocated segment comprised about 15% of chromosome 17 in each of the subjects. This was consistent with occurrence of breakpoints at the interphase between 17q24-q25, distal to the GH locus (growth hormone) and proximal to *TK1* (thymidine kinase). Two of the patients were XY females; one was an XX female. The karyotypes were

46, XY,t(7; 17)(q32; q24.2) or

46, XY,t(7; 17)(q34; q25.1)

46, XY,t(13; 17)(q22; q25.1)

46, XX,t(1; 17)(q42.13; q24.3 or q25.1).

The foregoing observations signaled the presence of a sex-determining gene (initially called *SRA1*) on chromosome 17 (association of campomelic dysplasia and sex reversal could be due to a single mutation affecting two developmental pathways, or to a deletion involving contiguous genes). Today the gene is called *SOX9*.

We have indicated that *SRY* belongs to a group of genes that code for HMG proteins. In aggregate, these constitute the *SOX (SRY-like HMG box)* family of genes. It is remarkable that duplication of *SOX9* can cause XX sex reversal – that is, it can induce the formation of XX testes – in the absence of *SRY*, and that SOX9 is the first marker to appear in newly differentiated Sertoli cells. These observations suggest that *SOX9* is an early downstream target of *SRY* (Fig. 1). In fact, mouse *Sox9* is expressed on day 10.5 in the indifferent embryonic gonads of male and female. But when *Sry* is expressed on day 11.5 – the time of differentiation of the mouse testis – transcription of *Sox9* is increased in the male and decreased in the female.

4.3 Other Autosomal Sex-determining Genes

There is evidence for autosomal sex-determining genes other than *SOX9*. Thus, for example, cases of XY sex reversal have been reported in patients with deletions in chromosome 9p. In one case, the subject was an XY female with a *de novo* 9p deletion and streak gonads but no other dysmorphic features. The *SRY* gene was intact.

4.3.1 The *SF1* Gene

In fact, the *SF1 (steroidogenic factor)* gene on chromosome 9 has a significant role in the embryogenesis of the adrenal gland and gonad in humans, and in mammals generally. In humans, mutation of the *SF1* gene leads to adrenal failure and XY gonadal dysgenesis. In mice with homozygous deletions of *SF1*, the gonads and adrenals both fail to develop, and the affected young die shortly after birth. Since these organs initiate development but do not finish it, *SF1* may participate in the development of the indifferent gonad, but not directly in testicular organogenesis. The *SF1* gene is on human chromosome 9p33, and contains a conservative DNA-binding motif consisting of two zinc fingers.

4.3.2 The *WT1* Gene

Wilms tumor of the kidney accounts for 15% of cancers in children. It may be found together with aniridia, urogenital malformations, and mental retardation (WAGR syndrome). The tumor develops after mutation or loss of the *WT1* tumor suppressor gene on chromosome 11p13. Affected XY children have gonadal dysgenesis, suggesting a role for the *WT1* gene in development of the testis. The gene contains 10 exons, which generate 24 isoforms, some of which can activate *SRY* and *DAX1* promoters *in vitro*.

4.3.3 The *WNT*-4 Gene

The *WNT-4* gene is on human chromosome 1p35. The WNT-4 protein is a member of a family of locally acting cellular growth factors. The gene is expressed in the mesonephros and in the adrenal gland in both sexes. In females, WNT-4 signaling is required for development of the Müllerian ducts and for differentiation of oocytes and interstitial cells. *WNT-4* is downregulated in the developing testis. In mice, deletion of *Wnt-4* causes XX sex reversal. And in humans, duplication

of the WNT-4 region on chromosome 1 causes XY sex reversal. Moreover, *Wnt-4* upregulates *Dax1*, so duplication of *Wnt-4* would be expected to suppress expression of *Sry*. These observations suggest a key role for *Wnt-4* signaling in gonadal sex determination.

5
XX Sex Reversal

5.1
XX Males

5.1.1 In the Human

Males with female sex chromosomes (46,XX) were first described in 1964. A chief feature of the XX male syndrome is the small, sterile testis, although in general, XX males are normal-appearing. The testicular cords are narrower than those in normal men and consist of Sertoli cells only (Fig. 8). Female-like breasts develop in about one-third of XX males, and hypospadias is present in some cases. The syndrome occurs with a frequency of about 1/20 000 male births.

Testicular development in the absence of a Y chromosome could be explained by constitutive activation of downstream testis-determining genes, or alternatively, by Y-X translocation. These conditions – constitutive mutation and Y-X translocation – would be represented by two classes of XX male, one lacking the *SRY* gene (Y−) and one having it (Y+). So it is remarkable that two such classes have now been identified, and that the Y(−) XX males are notable for female-like breasts (gynecomastia), cryptorchid testes, and abnormal external genitalia – traits uncharacteristic of Y(+) XX males.

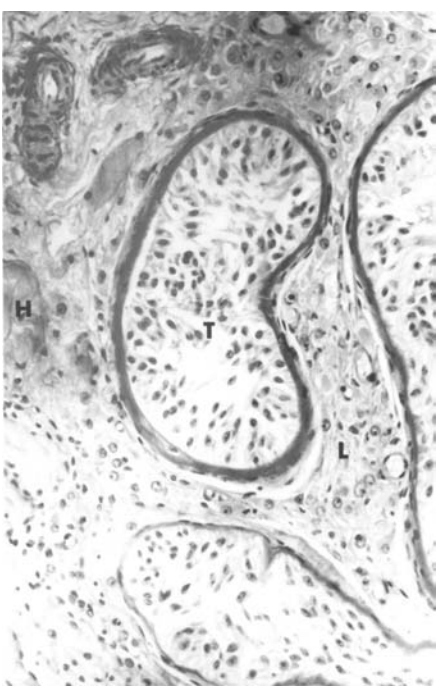

Fig. 8 The XX testis. Section of testicular biopsy from a 20-year-old 46,XX male. Some of the tubules are hyalinized (H). Some (T) are notable for thickening of the basement membrane, and are lined with Sertoli cells only; germ cells are absent. Leydig cells (L) are present in the interstitium. This photomicrograph was the generous gift of Professor Albert de la Chapelle. (From Wachtel, S.S., Bard, J. (1981) The XX Testis, in: Laron, Z. (Ed.) *The Intersex Child*, Karger, Basel, pp. 116–132).

In summary, then, most XX males are *SRY*-positive, due to X;Y translocation, but some are *SRY*-negative. Absence of *SRY* in XX males – implying mutation of downstream genes, or duplication, as in the case with *SOX9* – is associated with gynecomastia and abnormality of the external genitalia. Occasionally, XX males occur in kindreds with other XX males (and sometimes in kindreds with XX true hermaphrodites). These XX males usually lack Y-DNA (Sect. 5.2).

5.1.2 In the Mouse

XX males of the mouse were first described in 1971. Sex reversal in these animals was attributed originally to mutation of an autosomal gene, *Sxr*. But as noted in Sect. 2.2, the XX*Sxr* males were found to be the product of a recurrent nonreciprocal Y-X crossover involving the sex-determining region of the Y – namely, the *Sxr* region. Since XX*Sxr* males are sterile, the condition is transmitted by carrier XY*Sxr* males. The carriers produce four kinds of sperm – X, X*Sxr*, Y, Y*Sxr* – and thus four kinds of offspring: fertile XX females, XX*Sxr* sex-reversed males, XY males, and carrier XY*Sxr* males (Table 3). In 1984, certain XX*Sxr* males were found to lack the male-specific H-Y antigen (thereby excluding *Hya* as a candidate for *TDF*). This was due to a deletion within the *Sxr* region. The deletion was called Sxr^b to distinguish it from the nondeleted *Sxr*, called Sxr^a (Sect. 2.4.2).

Analysis of the *Sxr* region has revealed the presence of several genes: *Sry*, the testis-determining gene; *Hya*, the gene controlling expression of the H-Y antigen; *Spy*, a gene involved in spermatogenesis; *Zfy-1* and *Zfy-2*, homologs of human *ZFY*; and *Ube1-γ1* and *Ube1-γ2*, genes involved in ubiquitin-mediated protein metabolism. Of these, *Hya*, *Spy*, and *Zfy-2* are not found in the Sxr^b deletion, which comprises at most, 320 kb of DNA. The breakpoints of the deletion fall within the *Zfy* genes; Sxr^b thus includes a fusion gene whereby the 5′ promoter of *Zfy-2* is joined to the coding region of *Zfy-1*.

Tab. 3 Four kinds of offspring produced by XY*Sxr* carrier males.

Sperm	*Progeny*	*Sex phenotype*	*Bkm foci*[a]
X	XX	Fertile female	None
Y	XY	Noncarrier male	Y
X*Sxr*	XX*Sxr*	Sex-reversed male[b]	X
Y*Sxr*	XY*Sxr*	Carrier male	Y

[a] Rich clusters of *Bkm* normally associated with the sex-determining region of the mouse Y chromosome.
[b] XX males are sterile.
Source: After Wachtel, S.S., Tiersch, T.R. (1994) The Search for the Male-determining Gene, in: Wachtel, S.S. (Ed.) *Molecular Genetics of Sex Determination*, Academic Press, San Diego, CA, pp. 1–22.

5.1.3 In the Goat

Hornlessness in goats is caused by a dominant autosomal gene initially called *P (Polled)*. Goats heterozygous for the polled gene ($P/+$) are therefore born without horns, but they are otherwise unaffected. Goats homozygous for polled (P/P) are also born without horns but XX$P/P$ goats have testes or ovotestes and exhibit a range of sexual phenotypes varying from nearly normal male to hermaphroditic female. So *P* acts as an autosomal recessive testis-determinant.

In December 2001, geneticists in France reported that polled sex reversal is due to deletion from chromosome 1q43 of an 11.7-kb DNA fragment consisting mostly of repetitive sequences. Evidently, the deletion impedes transcription of two linked genes, one of which, *PISRT1*, may affect transcription of *SOX9*. According to one scheme, *PISRT1* normally inhibits the testis-inducing function of *SOX9*, and when *PISRT1* is deleted, as in XX*P*/*P* goats, *SOX9* induces the testicular cascade, even in the absence of *SRY*.

5.1.4 In the Dog

XX sex reversal is not uncommon among certain breeds of the dog, notably the cocker spaniel and Weimaraner. In 1978, veterinarians at the University of Pennsylvania described an XX male pup with undescended right testis, small penis with hypospadias, and hypoplastic uterus. Taken together, these findings were not surprising, as XX males had already been described in the cocker spaniel. But the subsequent finding was extraordinary.

The mother of the XX male was grossly female in appearance, but surgical exploration revealed that she had bilateral ovotestes, each containing prominent ovarian follicles and well-developed testicular tubules. She was a true hermaphrodite. The XX male pup had inherited his testicles from his mother.

These observations support the view that XX male syndrome and XX true hermaphroditism are alternative faces of the same coin. Other XX males and XX true hermaphrodites have been discovered in the extended pedigree, findings consistent with transmission of an autosomal or X-linked gene.

5.2 XX True Hermaphrodites

5.2.1 In the Human

There are numerous references to hermaphrodites in the ancient literature, for example, in Plato's *Symposium* and Ovid's *Metamorphoses*. The earliest reference appears in the cuneiform tablets unearthed at the site of the Royal Library at Nineveh (7th century BCE), and an account of the religious and social obligations of hermaphrodites appears in the *Mishnah*, *Bikkurim 4:1–5*, representing an oral tradition committed to writing nearly 2000 years ago.

According to our modern designation, a true hermaphrodite is a person with ovarian and testicular tissue – in particular, ovarian follicles and testicular tubules – in separate gonads or in a single gonad called the ovotestis. Simultaneous occurrence of male and female gonadal tissue is often associated with pronounced ambiguity of the secondary sex traits. The condition may be found in patients with a variety of karyotypes such as 46,XX or 46,XY, and mosaic complements such as 46,XX/46,XY. But 60% of human true hermaphrodites are 46,XX, raising the question, as in the case of XX males, how testicular differentiation can occur in the absence of the Y chromosome (i.e. in the absence of *SRY*).

Whereas X;Y translocation is common among XX males, it is uncommon among XX true hermaphrodites. When 35 XX true hermaphrodites were screened for Y-DNA sequences including *SRY*, the *SRY* sequence was found in only 4. This indicated that gene mutation is a likely cause of XX true hermaphroditism. And that was borne out in genealogical studies. Earlier (Sect. 5.1.1), we noted that XX males and XX true hermaphrodites sometimes occur in the same kindred. In

one such instance, the proband was an XX true hermaphrodite with prominent gynecomastia, small penis, hypospadias, and scrotal ovotestes. A brother of the proband was unaffected, but he sired an XX true hermaphrodite with ambiguous genitalia and bilateral ovaries with an appended right testis. Another unaffected brother sired two XX males: one had a small phallus with severe chordee, hypospadias, and hypoplastic testes, but no ovarian tissue; the other had similar traits. An autosomal sex-reversing gene was indicated in the pedigree (Fig. 9). In another similar family, an XX male and his XX true hermaphrodite sibling had a cousin who was an XX true hermaphrodite. *SRY* was not present in any of the affected family members.

Absence of Y-DNA in XX true hermaphrodites and some XX males points to genes other than *SRY* in the development of the normal testis. Although cryptic Y-bearing cell lines in the gonad cannot be ruled out in sporadic cases of XX sex reversal, occurrence of XX males and XX true hermaphrodites in families is consistent with transmission of autosomal sex-reversing genes (discussed in Sect. 4).

6 XY Sex Reversal

6.1 XY Females

6.1.1 In the Human

As noted in Sect. 3.2, two X chromosomes are required for differentiation of the normal human ovary. In 45,X human embryos, the ovary develops normally during the first trimester, but the germ cells die and the follicles become atretic. The gonads are represented later by fibrous streaks of connective tissue devoid of endocrinologic activity. The resulting condition is called *Turner syndrome*.

According to Susumu Ohno, the X and Y chromosomes evolved from a pair of homologous autosomes; the Y lost most of its genetic material to become a specialized testis inducer while the X remained invariant. Accordingly, when the sex-determining function of the Y chromosome is lost or compromised, the XY gonad develops in the same way as the XO gonad. The resulting condition, called *XY gonadal dysgenesis*, or *Swyer syndrome*, is characterized by primary amenorrhea and delayed puberty, and occurs in about one of

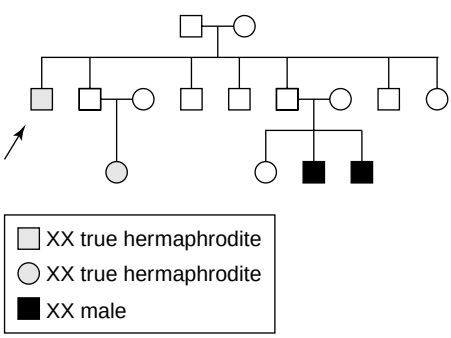

Fig. 9 Family with XX true hermaphrodites and two XX males. The propositus (arrow) was a masculinized true hermaphrodite with scrotal ovotestes. The genealogy is consistent with transmission of an autosomal sex-reversing gene. [On the basis of the study by Skordis, N., Stetka, D.G., MacGillivray, M.H., Greenfield, S.P. (1987) Familial 46,XX males coexisting with familial 46,XX true hermaphrodites, *J. Pediatr.* **110**, 244–248 (1987). From Wachtel, S.S. (1994) XX Sex Reversal in the Human, in: Wachtel, S.S. (Ed.) *Molecular Genetics of Sex Determination*, Academic Press, San Diego, CA, pp. 267–285].

20 000 female births. In addition to having nonfunctional streak gonads, women with XY gonadal dysgenesis have a significant risk of developing germ cell tumors such as gonadoblastoma and dysgerminoma (25–35%).

When the *SRY* gene was cloned, geneticists tested its function by evaluating the *SRY* sequence in women with XY gonadal dysgenesis. Mutation or loss of the *SRY* gene would explain failure of male development in these women, and would help to confirm its testis-determining role. Testing was accomplished by amplifying the *SRY* conserved domain in the polymerase chain reaction (PCR), then evaluating the nucleotide sequence of the PCR product, and comparing it to the sequence in normal XY males. By the end of 1993, more than a hundred cases of XY gonadal dysgenesis had been studied in this way. *De novo* mutations within the *SRY* conserved domain (the HMG box) were identified in about 15% of those cases.

Other mutations were detected in at least five of the affected women: they had male relatives with the same mutation. In one kindred, a T→C transition caused a phenylalanine-to-serine change at amino acid position 109 in the propositus and in the father, two brothers, and an uncle. No decrease in DNA-binding activity was observed when the mutant SRY protein was compared with wild-type SRY against the 5′-AACAAAG-3′ target site (see Sect. 2.4.4). In another family, a G→C substitution causing a valine-to-leucine change was identified at position 588 of pY53.3 in two XY female siblings, and in the father, a paternal uncle, and a paternal aunt (also an XY female). These familial variants might have been no more than benign polymorphisms (which would explain their presence in normal males). Alternatively, the same mutation might have had different effects in different individuals, depending on the genetic background.

Mutation in the *SRY* conservative domain in cases of XY gonadal dysgenesis was consistent with the putative sex-determining role of the *SRY* gene, but the low frequency of *de novo* mutations suggested other causes of the syndrome, and in particular, other genes in the sex-determining pathway. Mutation outside the HMG box could also account for a proportion of cases. A precedent was described in 1992, in a case involving a deletion 1.7 kb upstream of the *SRY* transcription initiation site. The deletion comprised some 25 to 50 kb of DNA, 5′ to *SRY* in the Y chromosome short arm.

Other forms of XY gonadal dysgenesis, namely those caused by perturbation of X-linked or autosomal genes, are discussed in Sects. 3 and 4.

6.1.2 In the Mouse

XY sex reversal XY females of the laboratory mouse, *Mus musculus*, were originally sired by a chimeric male produced with embryonic stem cells infected with a retrovirus. Because sex reversal was inherited with the Y chromosome, the mutation was designated $Y^{Tdy.m1}$, or Y for short. Later, when the mouse *Sry* gene was isolated, geneticists learned that it was deleted in XY females. Mapping studies showed that the deletion comprised nearly 11 kb of DNA, but did not include the *Zfy1* and *Zfy2* genes. Proviral sequences were not detected. The XY females were fertile but their reproductive lifespans were foreshortened, and they produced small litters.

XYd sex reversal When certain males of the mouse subspecies, *M. m. domesticus*,

are crossed with females of the inbred strain C57BL/6 (B6), the F_1 males are normal. But in subsequent backcrosses (B6 females × N) (where N = the backcross generation number), XY females and XY hermaphrodites are produced. This became known as XYd sex reversal (d = *domesticus*). On cytogenetic analysis, the Y chromosomes of the sex-reversed mice were indistinguishable from those of normal XY males. In Southern blots, hybridization patterns were also indistinguishable when DNA samples from XY males and XYd females were tested with Y-specific probes including *Sry*.

Because these findings indicated that XYd sex reversal was not due to a mutation of the Y chromosome, geneticists proposed that the testis-determining gene of the *M. m. domesticus* subspecies (Sry^{Dom}) could not interact functionally with an autosomal gene in the B6 strain, which they called *tda-1* (testis-determining autosomal-1). Hence, Sry^{Dom}, *tda-a/tda-1* mice could become females or hermaphrodites, but not males. The absence of sex-reversed XY females in the F_1 generation could be explained by the presence of the *M. domesticus* dominant *Tda1* allele in all of the XY (heterozygous) offspring. In subsequent backcross generations, however, the probability of inheriting the dominant *Tda1* allele would be reduced. In the first backcross generation, 25% of all autosomal loci would be homozygous for B6 alleles. In the sixth backcross generation (N6), nearly 100% of the autosomal genes would be of B6 origin, and in fact, all XY progeny in N6 have become females or hermaphrodites.

Several inbred mouse strains were found to have Y chromosomes of *M. m. domesticus* origin (e.g. Posch-1, AKR and SJL) although sex reversal was unreported in these strains. When the Posch-1 Y chromosome was introduced by backcrossing into the B6 genetic background, the results were the same as those just described. When the AKR Y chromosome was introduced, ovotestes were observed in 14- to 16-day-old embryos, but the ovotestes became virilized, and there was no evidence of sex reversal after birth. When the SJL Y chromosome was introduced, normal testes developed; ovotestes were found in only 1 of 87 fetuses. Thus, not all Y^{Dom} chromosomes could generate XYd sex reversal in the B6 strain.

Later, in Southern blots of DNA from AKR and SJL, alternative hybridization patterns were identified with Y-specific probes. Genetic heterogeneity was confirmed when sequencing revealed that *Sry* is polymorphic in the different *M. m. domesticus* Y chromosomes. Thus, although the precise mechanism of XYd sex reversal remained to be determined, *Sry* polymorphism could contribute to the different testis-inducing capacities of the Y^{Dom} chromosome in B6 mice.

T-associated sex reversal The *H-2*-linked T-locus mutants of the mouse block normal development of the notochord and generate phenotypes that can vary from loss of the entire notochord to occurrence of minor abnormalities of the tail. When the T^{hp} (T hairpin tail) mutation was being introduced into the B6 genetic background, geneticists noted that at the N3 and N4 backcross generations, XY fetuses with T^{hp} were females or hermaphrodites. This required backcrossing to the B6 inbred strain, and presence of an AKR Y^{Dom} chromosome; sex reversal did not occur when the T^{hp} mutation was backcrossed into the C3H strain, or when a non-AKR Y was present.

Since T^{hp} is a deletion of proximal chromosome 17, it was inferred that sex reversal in mice inheriting the mutation

was caused by hemizygous expression of a gene on the undeleted copy of chromosome 17. The gene was called *Tas*, for *T-associated sex reversal*. Since XYd and *Tas*-mediated XY sex reversal both occur when an *M. m. domesticus* Y chromosome is introduced into the B6 genetic background, it was asked whether they might be related conditions.

Other forms of XY sex reversal XY females have also been generated by deletion of *Sx1* repeat sequences in the mouse Y chromosome. In those cases, *Sry* is intact but its expression may be altered, suggesting the possibility of long-range position effects.

6.2 XY True Hermaphrodites in the Human

Among the 195 true hermaphrodites surveyed by Willem van Niekerk, 24 (12.3%) had the 46,XY karyotype. Of those, 13 had an ovary on one side of the body and a testis on the other; the other 11 had at least one ovotestis. The mosaic phenotype of the true hermaphrodite (testis versus ovary, in different gonads or in the same gonad), implies mosaic expression of the testis inducer, and this has been borne out by molecular analysis of the *SRY* gene.

The subject of one study was a newborn baby with ambiguous external genitalia and 46,XY karyotype. The gonads were described as ovotestes that contained tubules, Sertoli cells, spermatogonia, and oocytes. There were no Leydig cells and no intact follicles. DNA was extracted from white blood cells and gonadal tissue, and the *SRY* conserved motif was amplified in the PCR. When the PCR products were sequenced, no mutation was found in *SRY* from the white blood cells, but two mutations were found in gonadal tissue. One, a ''silent'' mutation, resulted in a T→C transition in the third nucleotide of codon 22, but there was no change in the amino acid (alanine). The other, a T→A transversion in the second nucleotide of codon 44, induced an amino acid substitution of histidine for leucine. Neither mutation was identified in DNA from the patient's father.

The two mutations that were not identified in all samples of gonadal tissue of the patient serve to confirm mosaic expression of the *SRY* testis inducer, and to explain development of the abnormal gonad in this case. Because the two mutations occurred together in the same cells, it was inferred that they had been induced simultaneously by a mutagen. In fact the mother had worked with organic solvents during early pregnancy.

7 Summary and Conclusions

The function of the human Y chromosome was debatable as recently as 1959, when a number of publications appeared showing that it determines sex. By 1990, the Y-chromosomal testis-determining gene had been cloned, and soon thereafter a comprehensive Y-deletion map containing 132 loci and 43 ordered intervals was produced; each interval was defined by chromosomal breakpoints and contained less than 800 kb of DNA.

By the year 2000, several additional sex-determining genes were identified: the X-linked sex-reversing *DAX1* gene, and the SRY-related *SOX9* gene on chromosome 17q. Another sex-reversing gene *(SF1)* was indicated on chromosome 9. And in keeping with its presumptive role as a transcription factor, the SRY protein was

found to recognize and bind a particular DNA sequence, 5′-AACAAAG-3′.

These findings are consistent with the notion that maleness is the end result of a cascade of events governed by a pathway of sex-determining genes. It is generally held that *SRY* is the initiator of the cascade. But *Sry* of the mouse is transcribed between 9.5 and 12.5 days of gestation (the mouse testis appears at 11.5 days). This implies that *Sry* itself is regulated or "turned on" by another gene or group of genes.

In any case, *SOX9* is activated early in the developing testis, indicating that *SOX9* is an early, and perhaps the first, downstream target of *SRY*. Besides *SOX9*, other genes in the pathway have been identified and are being analyzed – notably *DAX1*, *SF1*, *WT1*, and *WNT-4* (discussed in Sects. 3.3, 4.2, 4.3). *SF1* may participate in the development of the indifferent embryonic gonad (which first appears as an outcropping in the coelomic epithelium). The precise sex-determining roles of *WT1* and *WNT-4* remain to be elucidated.

Yet significant advances are being made in our understanding of the heretofore enigmatic conditions of sex reversal. For example, duplication of *SOX9* can "energize" the sex-determining cascade in the absence of *SRY*, giving rise to XX testes, and duplication of *DAX1* can block the *SRY* gene, causing aberrant ovarian development in women with XY gonadal dysgenesis. As for the future, we can look forward to a time when the genetics of sex determination will be more thoroughly understood, when many of the syndromes of human sex reversal will be identified and perhaps even corrected prenatally, and when the gender of important domestic species will be placed under absolute control.

Bibliography

Books and Reviews

Gubbay, J., Lovell-Badge, R. (1994) The Mouse Y Chromosome, in: Wachtel, S.S. (Ed.) *Molecular Genetics of Sex Determination*, Academic Press, San Diego, CA, pp. 43–67.

Hampikian, G.H., Cooper, D.W., Marshall Graves, J.A. (1994) Sex Determination in Marsupials and Monotremes, in: Wachtel, S.S. (Ed.) *Molecular Genetics of Sex Determination*, Academic Press, San Diego, CA, pp. 143–170.

Jost, A. (1970) Hormonal factors in the sex differentiation of the mammalian foetus, *Philos. Trans. R. Soc. London, B* **259**, 119–130.

McLaren, A. (1990) What makes a man a man, *Nature* **346**, 216–217.

Morrish, B.C., Sinclair, A.H. (2002) Vertebrate sex determination: means to an end, *Reproduction* **124**, 447–457.

Nagamine, C.M. (1994) XY Sex Reversal in the Mouse, in: Wachtel, S.S. (Ed.) *Molecular Genetics of Sex Determination*, Academic Press, San Diego, CA, pp. 83–105.

Ohno, S. (1967) *Sex Chromosomes and Sex-linked Genes*, Springer-Verlag, New York.

Roberts, L.M., Shen, J., Ingraham, A. (1999) New solutions to an ancient riddle: defining the differences between Adam and Eve, *Am. J. Hum. Genet.* **65**, 933–942.

Sinclair, A.H. (1994) The Cloning of *SRY*, in: Wachtel, S.S. (Ed.) *Molecular Genetics of Sex Determination*, Academic Press, San Diego, CA, pp. 23–41.

Wachtel, S.S., Simpson, J.L. (1994) XY sex reversal in the human, in: Wachtel, S.S. (Ed.) *Molecular Genetics of Sex Determination*, Academic Press, San Diego, CA, pp. 287–309.

Primary Literature

Acherman, J.C., Ito, M., Hindmarsh, P.C., Jameson, J.L. (1999) A mutation in the gene encoding steroidogenic factor-1 causes XY sex reversal and adrenal failure in humans, *Nat. Genet.* **22**, 125–126.

Affara, A.A., Chalmers, I., Ferguson-Smith, M.A. (1993) Analysis of the *SRY* gene in 22 sex-reversed XY females identifies four new point mutations in the conserved DNA binding domain, *Hum. Mol. Genet.* **2**, 785–789.

Arn, P., Chen, H., Tuck-Muller, C.M., Mankinen, C., Wachtel, G., Li, S., Shen, C.-C., Wachtel, S.S. (1994) SRVX, a sex reversing locus in Xp21.2-Xp22.11, *Hum. Genet.* **93**, 389–393.

Bardoni, E., Zanaria, E., Guioli, S., Floridia, G., Worley, K.C., Tonini, G., Ferrante, E., Chiumello, G., McCabe, E.R.B., Fraccaro, M., Zuffardi, O., Camarino, G. (1994) A dosage sensitive locus at chromosome Xp21 is involved in male to female sex reversal, *Nat. Genet.* **7**, 497–501.

Behringer, R.R., Finegold, M., Cate, R.L. (1994) Müllerian-inhibiting substance function during mammalian sexual development, *Cell* **79**, 415–425.

Bernstein, R., Koo, G.C., Wachtel, S.S. (1980) Abnormality of the X chromosome in human 46,XY female siblings with dysgenetic ovaries, *Science* **207**, 768–769.

Capel, B., Albrecht, K.H., Washburn, L.L., Eicher, E.M. (1999) Migration of mesonephric cells into the mammalian gonad depends on *Sry*, *Mech. Dev.* **84**, 127–131.

Cattanach, B.M., Pollard, C.E., Hawkes, S.G. (1971) Sex-reversed mice: XX and XO males, *Cytogenetics* **10**, 318–337.

Crawford, P.A., Dorn, C., Sadovsky, Y., Milbrandt, J. (1998) Nuclear receptor DAX-1 recruits nuclear receptor corepressor N-CoR to steroidogenic factor 1, *Mol. Cell. Biol.* **18**, 2949–2956.

de la Chapelle, A., Hortling, H., Niemi, M., Wennstrom, J. (1964) XX sex chromosomes in a human male, first case, *Acta Med. Scand. Suppl.* **412**, 25–38.

de la Chapelle, A., Tippett, P.A., Wetterstrand, G., Page, D. (1984) Genetic evidence of X-Y interchange in a human XX male, *Nature* **307**, 170–171.

Eicher, E.M., Washburn, L.L., Schork, N.J., Lee, B.K., Shown, E.P., Xu, X., Dredge, R.D., Pringle, M.J., Page, D.C. (1996) Sex-determining genes on mouse autosomes identified by linkage analysis of C57BL/6J-YPOS sex reversal, *Nat. Genet.* **14**, 206–209.

Eicher, E.M., Washburn, L.L., Whitney, J.B. III, Morrow, K.E. (1982) Mus poschiavinus Y chromosome in the C57BL/6J murine genome causes sex reversal, *Science* **217**, 535–537.

Ferguson-Smith, M.A. (1966) X-Y chromosomal interchange in the aetiology of true hermaphroditism and of XX Klinefelter's syndrome, *Lancet* **2**, 475–476.

Fredga, K., Gropp, A., Winking, H., Frank, F. (1976) Fertile XX- and XY-type females in the wood lemming Myopus schisticolor, *Nature* **261**, 225–227.

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P., Lovell-Badge, R. (1990) A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryologically expressed genes, *Nature* **346**, 245–250.

Hacker, A., Capel, B., Goodfellow, P., Lovell-Badge, R. (1995) Expression of *Sry*, the mouse sex determining gene, *Development* **121**, 1603–1614.

Hamerton, J.L., Dickson, J.M., Pollard, C.E., Grieves, S.A., Short, R.V. (1969) Genetic intersexuality in goats, *J. Reprod. Fertil.* (Suppl. 7), 25–51.

Haqq, C.M., King, C.-Y., Ukiyama, E., Falsafi, S., Haqq, T.N., Donahoe, P.K., Weiss, M.A. (1994) Molecular basis of mammalian sexual determination: activation of Müllerian inhibiting substance gene expression by SRY, *Science* **266**, 1494–1500.

Huang, B., Wang, S., Ning, Y., Lamb, A.N., Bartley, J. (1999) Autosomal XX sex reversal caused by duplication of *SOX9*, *Am. J. Med. Genet.* **87**, 349–353.

Ito, M., Yu, R., Jameson, J.L. (1997) DAX-1 inhibits SF-1-mediated transactivation via a carboxy-terminal domain that is deleted in adrenal hypoplasia congenital, *Mol. Cell. Biol.* **17**, 1476–1483.

Jost, A. (1947) Recherches sur la différenciation sexuelle de l'embryon de Lapin. III. Role des gonades foetales dans la différenciation somatique, *Arch. Anat. Microb. Morphol. Exp.* **36**, 271–316.

Karl, J., Capel, B. (1998) Sertoli cells of the mouse originate from the coelomic epithelium, *Dev. Biol.* **203**, 323–333.

Kent, J., Wheatley, S.C., Andrews, J.E., Sinclair, A.H., Koopman, P. (1996) A male-specific role for SOX9 in vertebrate sex determination, *Development* **122**, 2813–2822.

Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., Lovell-Badge, R. (1991) Male development of chromosomally female mice transgenic for *Sry*, *Nature* **351**, 117–121.

Kwok, C., Weller, P.A., Guioli, S., Foster, J.W., Mansour, S., Zuffardi, O., Punnett, H.H., Dominguez-Steglich, M.A., Brook, J.D., Young, I.D., Goodfellow, P., Schafer, A.J. (1995) Mutations in *SOX9*, the gene responsible for campomelic dysplasia and autosomal sex reversal, *Am. J. Hum. Genet.* **57**, 1028–1036.

McDonald, M.T., Flejter, W., Sheldon, S., Putzi, M.J., Gorski, J.L. (1997) XY sex reversal and gonadal dysgenesis due to 9p24 monosomy, *Am. J. Med. Genet.* **73**, 321–326.

Meyers-Wallen, V.N., Bowman, L., Acland, G.M., Palmer, V.L., Schlafer, D., Fajt, V. (1995) Sry-negative XX sex reversal in the German shorthaired pointer dog, *J. Hered.* **86**, 369–374.

Nachtigal, M.W., Hirokawa, Y., Enyeart-VanHouten, D.L., Flanagan, J.N., Hammer, G.D., Ingraham, H.A. (1998) Wilms' tumor 1 and Dax-1 modulate the orphan receptor SF-1 in sex-specific gene expression, *Cell* **93**, 445–454.

Nagamine, C.M., Chan, K., Kozak, C.A., Lau, Y.-F. (1989) Chromosome mapping and expression of a putative testis-determining gene in mouse, *Science* **243**, 80–83.

Page, D.C., Mosher, R., Simpson, E.M., Fisher, E.M.C., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A., Brown, L. (1987) The sex-determining region of the human Y chromosome encodes a finger protein, *Cell* **51**, 1091–1104.

Pailhoux, E., Vigier, B., Chaffaux, S., Servel, N., Taourit, S., Furet, J.-P., Fellous, M., Grosclaude, F., Cribiu, E.P., Cotinot, C., Vaiman, D. (2001) A 11.7-kb deletion triggers intersexuality and polledness in goats, *Nat. Genet.* **29**, 453–458.

Pivnick, E., Wachtel, S., Woods, D., Simpson, J.L., Bishop, C.E. (1992) Mutations in the conserved domain of *SRY* are uncommon in XY gonadal dysgenesis, *Hum. Genet.* **90**, 308–310.

Rosenberg, H.S., Clayton, G.W., Hsu, T.C. (1963) Familial true hermaphroditism, *J. Clin. Endocrinol.* **23**, 203–206.

de Santa Barbara, P., Bonnard, N., Boizet, B., Desclkozeaux, M., Moniot, B., Sudbeck, P., Scherer, G., Poulat, E., Berta, P. (1998) Direct interaction of SRY-related protein SOX9 and steroidogenic factor 1 regulates transcription of the human anti-Müllerian hormone gene, *Mol. Cell. Biol.* **18**, 6653–6665.

Selden, J.R., Moorhead, P.S., Koo, G.C., Wachtel, S.S., Haskins, M.E., Patterson, D.F. (1984) Inherited XX sex reversal in the cocker spaniel dog, *Hum. Genet.* **67**, 62–69.

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.-M., Lovell-Badge, R., Goodfellow, P.N. (1990) A gene from the human sex determining region encodes a protein with homology to a conserved DNA binding motif, *Nature* **346**, 240–244.

Singh, L., Phillips, C., Jones, K.W. (1984) The conserved nucleotide sequences of Bkm, which define Sxr in the mouse, are transcribed, *Cell* **36**, 111–120.

Smith, C.A., McClive, P.J., Western, P.S., Reed, K.J., Sinclair, A.H. (1999) Conservation of a sex-determining gene, *Nature* **402**, 601–602.

Swain, A., Lovell-Badge, R. (1999) Mammalian sex determination: a molecular drama, *Genes Dev.* **13**, 755–767.

Swain, A., Narvaez, V., Burgoyne, P., Camerino, G., Lovell-Badge, R. (1998) Dax1 antagonizes Sry action in mammalian sex determination, *Nature* **391**, 761–767.

Swyer, G.I.M. (1955) Male pseudohermaphroditism: a hitherto undescribed form, *Br. Med. J.* **2**, 709–712.

Therkelsen, A.J. (1964) Sterile male with the chromosome constitution 46,XX, *Cytogenetics* **3**, 207–218.

Thomas, L., Tiepolo, L., Zuffardi, O., Rodewald, A. (1977) Nullisomy for the distal portion of Xp in a male child with a Y/X translocation, *Hum. Genet.* **39**, 277–281.

Tiersch, T.R., Mitchell, M.J., Wachtel, S.S. (1991) Phylogenetic conservation of the *SRY* gene, *Hum. Genet.* **87**, 571–573.

Turner, H.H. (1938) A syndrome of infantilism, congenital webbed neck and cubitus valgus, *Endocrinology* **23**, 566–574.

van Niekerk, W.A., Retief, A.E. (1981) The gonads of human true hermaphrodites, *Hum. Genet.* **58**, 117–122.

Verga, V., Erickson, R.P. (1989) An extended long-range restriction map of the human sex-determining region on Yp, including ZFY, finds marked homology on Xp and no detectable Y sequences in an XX male, *Am. J. Hum. Genet.* **44**, 756–765.

Wachtel, S.S. (1998) X-linked sex-reversing genes, *Cytogenet. Cell. Genet.* **80**, 222–225.

Wachtel, S.S., Ohno, S., Koo, G.C., Boyse, E.A. (1975) Possible role for H-Y antigen in the primary determination of sex, *Nature* **257**, 235–236.

Whitfield, L.S., Lovell-Badge, R., Goodfellow, P. (1993) Rapid sequence evolution of the mammalian sex-determining gene *SRY*, *Nature* **364**, 713–715.

Growth Factors and Oncogenes in Gastrointestinal Cancers

Eiichi Tahara
Hiroshima University, Hiroshima, Japan

1 Introduction 3

2 Genetic and Epigenetic Alterations and Abnormal Growth Factor/Cytokine Network in Esophageal Cancer 4
2.1 Genetic and Epigenetic Alterations in Esophageal SCC 4
2.2 Abnormal Growth Factor/Cytokine Network in Esophageal SCC 6

3 Genetic and Epigenetic Alterations and Abnormal Growth Factor/Cytokine Network in Gastric Cancer 8
3.1 Genetic and Epigenetic Alterations in Gastric Cancer 8
3.2 Factors Associated with Increased Incidence of Gastric Cancer 13
3.3 Abnormal Growth Factor/Cytokine Network in Gastric Cancer 16

4 Genetic and Epigenetic Alterations and Abnormal Growth Factor/Cytokine Network in Colorectal Cancer 18
4.1 Genetic and Epigenetic Alterations in Colorectal Cancer 19
4.2 Factors Associated with Increased Incidence of Colorectal Cancer 21
4.3 Abnormal Growth Factor/Cytokine Network in Colorectal Cancer 22

5 Conclusion 23

Bibliography 24
Books and Reviews 24
Primary Literature 25

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

Keywords

Cell Adhesion Molecules
Surface ligands, usually glycoproteins, that mediate cell to cell adhesion. Their functions include the assembly and interconnection of various vertebrate systems, as well as maintenance of tissue integration, wound healing, morphogenic movements, cellular migration, and metastasis.

Cell Cycle Regulators
Proteins that regulate the cell division cycle. This family of proteins involves a wide variety of classes, including cyclin-dependent kinases, mitogen-activated kinases, cyclins, and phosphoprotein phosphatases as well as their putative substrates, such as chromatine-associated proteins, cytoskeletal proteins, and transcription factors.

Cytokines
Polypeptides secreted by inflammatory leukocytes, macrophages and lymphocytes in response to microbes and other antigens that mediate and regulate immune and inflammatory reactions. They generally act locally in a paracrine or an autocrine manner rather than in an endocrine manner.

Growth Factors
Signal molecules that act to control cell growth and differentiation in the receptor-dependent fashion. The alterations of these proteins lead to transformation and the accompanying loss in growth control. Some of the growth factors and their receptors are involved in the products of oncogens.

Oncogenes
Genes that can convert cells to cancerous growth by attacking crucial cellular machinery. They encode for growth factors, growth-factor receptors, protein kinases, signal transducers, nuclear phosphoproteins, and transcription factors. These genes are constitutively expressed after structural and /or regulatory changes, resulting in uncontrolled cell proliferation. They can be classified into viral oncogenes (v-oncogenes) and cellular oncogenes (proto-oncogenes).

Telomerase
Essential ribonucleoprotein reverse transcriptase that adds telomeric DNA to the ends of eukaryotic chromosomes. Telomerase is expressed in the testis and ovary, but repressed in normal human somatic tissues. Telomerase activity is seen in more than 90% of human cancers.

Tumor-suppressor Genes
Genes inhibit expression of the tumorigenic phenotype. They are normally involved in holding cellular growth in check. When tumor-suppressor genes are inactivated or lost, a barrier to normal proliferation is removed, leading to unregulated growth.

A large number of molecular events are involved in the development and progression of gastrointestinal carcinomas. Among them, common and distinct events of genetic and epigenetic alterations in oncogenes, tumor-suppressor genes, cell adhesion molecules, DNA repair genes, and genetic instability as well as telomerase activation are observed in esophageal, gastric, and colorectal cancers. In gastric cancer, the pattern of genetic and epigenetic alterations also differs depending on the two histological types, intestinal type or well-differentiated type and diffuse type or poorly differentiated type, indicating that there are two distinct major genetic pathways for gastric carcinogenesis.

In addition to these events, gastrointestinal cancer cells express a broad spectrum of the growth factor/cytokine receptor systems that organize complex cancer-stromal interaction, which confer cell growth, apoptosis, morphogenesis, angiogenesis, progression, and metastasis. However, these abnormal growth factor/cytokine networks also are different among esophageal, gastric, and colorectal cancers, respectively. Importantly, NF-κ**B** activation induced by inflammation may act as a key player for induction of growth factor/cytokine network in gastrointestinal cancers.

1 Introduction

Multiple genetic and epigenetic alterations in oncogenes, tumor-suppressor genes, cell cycle regulations, cell adhesion molecules, DNA repair genes and genetic instability as well as telomerase activation are responsible for the multistep process of human gastrointestinal carcinogenesis. However, a scenario or particular combination of these alterations differs in esophageal, gastric, and colorectal cancers. Namely, common and distinct molecular events are observed in esophageal, gastric, and colorectal cancers, respectively. Moreover, two types of gastric cancer, well-differentiated or intestinal type, and poorly differentiated or diffuse-type carcinomas also exhibit a distinct pattern of genetic pathways.

Besides these genetic and epigenetic events, gastrointestinal cancer cells express a broad spectrum of growth factors, cytokine or both, including epidermal growth factor (EGF) family, transforming growth factor (TGF)-β, heparin binding (HB)-EGF, PDGF, IGF, basic fibroblast growth factor (FGF), interleukin (IL)-1α, IL-6, IL-8 and osteopontin (OPN). These growth factors and cytokines act as autocrine, paracrine, and juxtacrine modulators of the growth of cancer cells, and then organize complex interplay between cancer cells and stromal cells, which plays an important role in cell growth, apoptosis, morphogenesis, angiogenesis, progression and metastasis. Interestingly, the expression of these growth factors, cytokines or both by cancer cells is also different among esophageal, gastric, and colorectal cancers.

This article will provide an overview of the molecular machinery that underlies gastrointestinal carcinogenesis and focuses on abnormal growth factor/cytokine network in gastrointestinal cancers.

2 Genetic and Epigenetic Alterations and Abnormal Growth Factor/Cytokine Network in Esophageal Cancer

Esophageal cancer is the third most frequent gastrointestinal cancer in the world. The most recent estimates are that esophageal cancer is the sixth most common cancer in men (212 600 new cases, 4.9% of all cancers) and the ninth most common in women (103 200 new cases, 2.7% of all cancers). The two main histological types of esophageal cancer are squamous cell carcinoma (SCC) and adenocarcinoma, but SCC is the more prevalent type worldwide. The development of esophageal SCC exhibits a multistep, progressive process. An early indicator of this process is an increased proliferation of esophageal epithelial cells including basal cell hyperplasia, dysplasia, and carcinoma *in situ*. This multistep process requires the accumulation of multiple genetic and epigenetic alterations and overexpression of growth factors/cytokine receptor systems, leading to the evolution of clonal cell populations that possess growth advantages over other cells as demonstrated in the progression model of head and neck cancer. This paragraph thus will describe recent advances in molecular dissection of multistep tumorigenesis of esophageal SCC and abnormal growth factor/cytokine network that contributes to the development and progression of esophageal SCC.

2.1 Genetic and Epigenetic Alterations in Esophageal SCC

Numerous genetic and epigenetic alterations are implicated in the development and progression of esophageal SCC (Table 1). This cancer is frequently associated with loss of heterozygosity (LOH) at multiple chromosomal loci including 3p, 5q, 9p, 9q, 13q, 17p, 17q, and 18q. No significant differences have been found in the prevalence of LOH at various loci in SCC and adenocarcinoma of the esophagus.

Tab. 1 Genetic and epigenetic alterations found in esophageal SCC.

Genetic and epigenetic alterations	*Incidence [%]*
Tumor suppressors	
P53 LOH, mutation	40–60
APC LOH	60–70
DCC LOH	20–40
Rb LOH	40–50
BRCA1 LOH	60
3p LOH	40–100
5p LOH	62
9p LOH	45–76
9q LOH	60
13q LOH	57
14q LOH	65
17p LOH	43–65
17q LOH	62
RARβ2 loss	40–50
Cell cycle regulators	
p16 loss, mutation	45–76
Cyclin D_1 amplification	40–50
Oncogenes	
EGFR amplification	10–15
Telomerase activity	86
TERT expression	96

Among these alterations, LOH and mutation of the *p53* gene at chromosome 17p13 occur at an early stage of esophageal carcinogenesis, such as dysplasia and carcinoma *in situ*. About 50% of esophageal SCC harbor mutations of the *Tp53* gene, most of which are missense mutations leading to amino acid changes within exons 5–8, which encode the entire DNA binding domain of the p53 molecule and the flanking splice sites. Considering the

base substitution spectrum, G:C to T:A transversion is common in esophageal carcinoma, similar to that in carcinomas of the lung and liver. This situation is different from the finding that colorectal carcinomas frequently contain G:C to A:T transitions at CpG dinucleotides. This evidence suggests that different environmental and intrinsic factors may affect the tumorigenesis of esophageal and colorectal carcinomas. It is of interest that LOH of the *APC, DCC*, and *Rb* genes shows high frequency but these genes are very rarely or never mutated in esophageal SCC.

The retinoic acid receptor (RAR) β gene is a putative tumor-suppressor gene on chromosome 3p24, where a high frequency of LOH is found in many human cancers, including esophageal cancer. The human RARβ has three isoforms (β1, β2, β4). Overexpression of RARβ2 induces inhibition of tumor cell growth and apoptosis in human cancer cell lines including esophageal cancer cells. Moreover, Induction of RARβ2 suppresses cyclooxygenase-2 (COX2) expression in esophageal cancer cells. More importantly, DNA methylation of RARβ2 promoter CpG sites has been reported to cause the loss of RARβ2 expression in many human cancers including lung, breast, prostate, stomach, head and neck, and esophageal cancers. RARβ is expressed in 90% of normal esophageal mucosa, while it is detected in only 60% of dysplastic lesions and in 50% of SCC. These findings indicate that loss of RARβ, or more specifically, the isoformβ2, is an early event associated with esophageal carcinogenesis and the status of squamous differentiation.

p16, an inhibitor of cyclin D1/cyclin-dependent kinase, is located on chromosome 9p21. It is inactivated by 9p21LOH with *de novo* p16 promoter hypermethylation in the majority of esophageal SCC. Recent molecular analysis of precancerous laryngeal lesions suggests that loss of p16 protein is an early step toward malignant transformation in head and neck tissues. This protein forms binary complexes with CDK4 and CDK6, inhibiting their ability to phosphorylate the Rb protein. Loss of the p16 protein may bring about increased Rb phosphorylation and allow cells to enter into S-phase. In fact, we have confirmed that homozygous deletion of the *p16* gene is closely correlated with the increased expression of cyclin D1, CDK4 and phosphorylated Rb protein in esophageal SCC cell lines.

In 1989, we discovered the coamplification of *hst-1* and *int-2*, both of which are located on chromosome 11q13, in about 50% of primary tumors and in 100% of metastases of esophageal SCC. Gene amplification, however, was not accompanied by overexpression of the two genes. Subsequently, Jiang et al. found that the cyclin D1, which is located on the same locus as *hst-1* and *int-2* genes, was amplified in 32% of SCC, associated with overexpression. The amplification of the cyclin D1 is closely correlated with tumor staging, depth of tumor invasion, and metastasis. In the esophagus, 71% of SCC and 64% of adenocarcinoma are positive for increased cyclin D1 nuclear staining, indicating that overexpression of cyclin D1 is common in both types of cancer. Cyclin D1 binds to Rb protein and stimulates its phosphorylation. Hyperphosphorylation of Rb in response to overexpressed cyclin D1 may lead to uncontrolled cell cycling and increased cell proliferation.

As for oncogene activation, amplification of the EGF receptor (*EGFR*) gene occurs in 10–15% of advanced cases of

esophageal SCC, accompanied by overexpression of EGFR. The frequency of K-ras mutation is very low in esophageal SCC, whereas it takes place in 50% of sporadic colorectal carcinoma. c-*erb*B2 is amplified in esophageal adenocarcinoma but not in esophageal SCC. Recently, Inazawa's group reported that *ZASC1* encoding a Krüppel-like zinc finger protein is involved in the pathogenesis of esophageal SCC as one of the targets for 3q26 amplification. *CIAP1*, a number of the *IAP* (antiapoptotic) gene family, may also be a target for 11q22 amplification.

Telomerase, a ribonucleoprotein enzyme, is necessary for cancer cells to maintain their telomere and to become immortal. Results of a 1998 study on cell immortalization show, however, that activation of telomerase alone is not enough to immortalize certain epithelial cells, and that inactivation of the p16/Rb pathway is needed. More than 80% of gastrointestinal carcinomas exhibit high level of telomerase activity and overexpression of human telomerase reverse transcriptase (hTERT). The expression of hTERT is closely associated with activation of telomerase *in vitro* and *in vivo*. It is of interest to note that telomerase activity as well as hTERT expression is detected in about 45% of dysplasia and in 90% of SCC of the esophagus. Telomerase activation may also play a critical role in early stage of esophageal SCC.

Recently, Chen et al. reported that LOH at 13q 33–34 including *ING1*, a candidate tumor-suppressor gene, was observed in about 60% of esophageal SCC, associated with mutation as well as loss of ING1 protein. ING1, a novel growth inhibitor, cooperates directly with p53 in growth regulation by modulating the ability of p53 to act as a transcriptional activator. Genetic or epigenetic alterations in ING1 may be also involved with esophageal SCC. Sonoda et al. reported that loss of *LRP1B* (low density lipoprotein receptor–related protein 1*B*) often occurs in esophageal SCC.

These results overall indicate that accumulation of the above-mentioned genetic and epigenetic alterations is involved in the multistep carcinogenesis and progression of esophageal SCC. Inactivation of tumor-suppressor genes on 3p (ex. RARβ2) and p53, and telomerase activity may be important for converting normal stratified squamous epithelium to dysplasia. Because p16 inactivation and 9q LOH are found occasionally in mild dysplasia, but frequently in severe dysplasia and in carcinoma *in situ*, these alterations may have implications for transformation to malignancy. Amplification of cyclin D1and *EGFR* genes, inactivation of tumor-suppressor genes on 5q, 13q, and 18q, and abnormal expression of growth factor/cytokine receptor system may confer progression and metastasis of esophageal SCC. The genetic progression model of esophageal SCC (Fig. 1) is quite similar to that of head and neck SCC

2.2 Abnormal Growth Factor/Cytokine Network in Esophageal SCC

Esophageal SCC cells express a variety of growth factor/ receptor and cytokines including the EGF family, PDGF, transforming growth factor β (TGFβ), interleukin (IL)-1α and IL-6. Among them, the EGF/TGFα receptor system plays a major role in the cell growth and progression of esophageal SCC through signaling of receptor-linked tyrosine kinases.

In many cancer cells, both EGF and TGFα act as autocrine growth factors through EGFR which is encoded by the proto-oncogene *c-erbB1*.

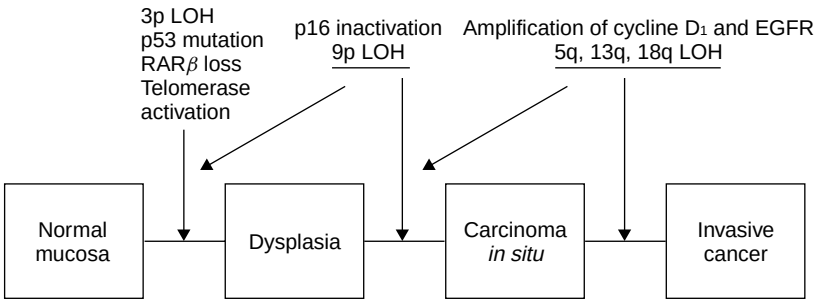

Fig. 1 Genetic progression model of esophageal SCC.

Membrane-bound TGFα binds to EGFR on the surface of contiguous cells and induces receptor autophosphorylation, leading to signal transduction, which is regarded as juxtacrine mitogenetic stimulation. In addition, EGF or TGFα induce the expression of TGFα and EGFR, thus indicating the presence of an autocrine loop of EGF/TGFα/receptor system. We have found that EGF is expressed in 30% of esophageal SCC while TGFα and EGFR are expressed in 77 and 89% of esophageal SCC, respectively. Coexpression of TGFα and EGFR is closely correlated with dysplastic progression and high grade of malignancy. The number of EGFR receptors is about 10 times higher than that in gastric carcinomas, which might be one of the reasons for the rapid growth of esophageal SCC compared to that of other carcinomas.

Esophageal SCC expresses PDGF A- and B-chains and PDGF B-receptor, suggesting the existence of a multiautocrine loop in the growth and progression of tumor cells. What is interesting is that EGF or TGFα increase the expression of PDGF A- and B-chain in esophageal SCC cell lines.

Grb7, a ligand for both EGFR and c-*erb*B2, is also implicated in esophageal SCC. Tanaka et al. found overexpression of Grb7in 44 % of the tumors and coexpression of Grb7 with EGFR or c-*erb*B-2 in 32% of the advanced cases, suggesting that Grb7 functions as an extracellular stimulus for progression of esophageal SCC.

EGF or TGFα produced by the tumor cells can make not only an autocrine loop of EGF/TGFα receptor system for tumor growth stimulation but also an induction of matrix metalloproteinases such as intestinal collagenase and stromelysin to evoke a cascade of cellular events that are involved in extracellular matrix degradation and tumor invasion. Moreover, one of the important substrates of EGFR kinase is β-catenin, a regulatory protein for cadherin. Upon phosphorylation of β-catenin, the cells are dissociated by loss of cadherin function. A phosphorylated form of β-catenin is detected in tumor cells of esophageal SCC and adenocarcinomas. β-catenin is also involved as a downstream transcriptional activator of the Wnt signaling pathway. Recently, β-catenin has been shown to induce cyclin D1 expression, suggesting that free β-catenin may be implicated in an increase in cell cycling.

In addition to phosphorylation of β-catenin, downregulation of E-cadherin and α-catenin occurs in both esophageal SCC and adenocarcinomas, while upregulation of P-cadherin takes place in

esophageal SCC. The reduced expression of α-catenin correlates more with invasion and nodal metastasis than E-cadherin reduction in esophageal SCC.

Above all, esophageal SCC exhibit multiple autocrine growth factor-receptor loops that may participate not only in cell growth but also in tumor invasion and metastasis, associated with reduction of cell adhesion molecules. Understanding the biology of esophageal cancer is indispensable to precise diagnosis and proper cancer treatment.

3 Genetic and Epigenetic Alterations and Abnormal Growth Factor/Cytokine Network in Gastric Cancer

Gastric cancer is the most common cancer worldwide and is second only to lung cancer as a cause of cancer mortality. Most recent world estimates indicate that 798 000 new cases are diagnosed and 628 000 deaths occur annually from gastric cancer. The highest incidence is still observed in Japan because of the remarkable increase in the aged population over 60 years old. Most gastric cancers arise distally from the antrum and pylorus, but about 20% involve the cardia and fundus and approximately 10% involve the stomach diffusely.

There are several histological classifications of gastric cancer. Lauren divided gastric cancer into two types, intestinal and diffuse, and the Japan Research Society for Gastric Cancer classified it into five common types. The JRSGC classification is similar to that of the World Health Organization. This article will use a two-type classification: the intestinal or well-differentiated type and the diffuse or poorly differentiated type.

The genetic and epigenetic changes found in gastric carcinoma differ depending on the histological type of gastric cancer, indicating that different carcinogenetic pathways exist for the intestinal and diffuse types of carcinomas. In addition, cancer-stromal interaction through the growth factor/cytokine receptor system, which plays a pivotal role in morphogenesis, cancer progression, and metastasis, is also different between the two types of gastric carcinomas.

Meta-analysis of epidemiological studies and animal models show that both intestinal and diffuse types of gastric cancer are equally associated with *Helicobacter pylori* (*H. pylori*) infection. However, *H. pylori* infection may play a role only in the initial steps of gastric carcinogenesis. Differences in *H. pylori* strain, patient age, exogenous or endogenous carcinogens and genetic factors such as DNA polymorphism and genetic instability may be implicated in two distinct major genetic pathways for gastric carcinogenesis.

The next paragraph will describe the genetic pathways of the two types of gastric cancer and the abnormal growth factor/cytokine network that organizes remarkably complex interplay between cancer cells and stroma cells in gastric cancer.

3.1 Genetic and Epigenetic Alterations in Gastric Cancer

Genetic and epigenetic alterations in oncogenes, tumor-suppressor genes, cell-cycle regulators, cell adhesion molecules, DNA repair genes, and genetic instability as well as telomerase activation, are responsible for tumor genesis and progression of gastric cancer (Tables 2 and 3; Figs. 2 and 3). Among them, inactivation of various genes including *p16*, *hMLH1*, cadherin1(*CDH1*),

Tab. 2 Genetic and epigenetic alterations found in intestinal and diffuse types of gastric cancer (1).

Genetic and epigenetic alterations	***Incidence of cases with indicated alterations [%]***	
	Intestinal type	***Diffuse type***
Tumor suppressors		
p53 LOH, mutation	60	75
p73 LOH	53[a]	24
APC LOH, mutation	40–60	0
DCC LOH	50	0
LOH of Chr. 1q	44	0
LOH of Chr. 7q	53	33
LOH of Chr. 17q	0	40[b]
Loss of pS2 expression	49	31
Loss of RARβ2	50	73[c]
Loss of RUNX3	37	40
Cell cycle regulators		
Cyclin E amplification	33	7
Cyclin E overexpression	26	27
CDC25B overexpression	33	73
Loss of p16 expression	50[d]	10
Loss of p27 expression	46	69

[a] Preferentially found in foveolar-type adenocarcinoma.
[b] Preferentially found in patients younger than 35 years of age.
[c] $p = 0.0335$; Fisher's exact test.
[d] $p = 0.0023$; Fisher's exact test.

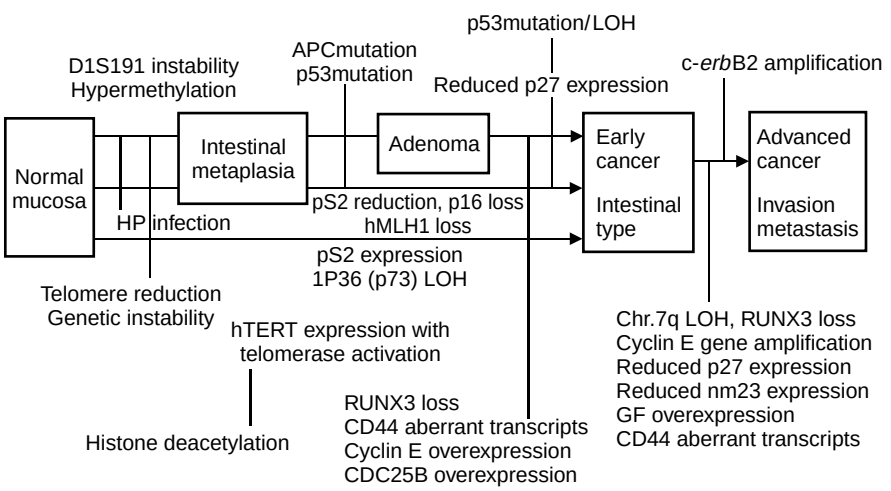

Fig. 2 Multiple genetic and epigenetic alterations during human gastric carcinogenesis (intestinal type).

Tab. 3 Genetic and epigenetic alterations found in intestinal and diffuse types of gastric cancer (2).

Genetic and epigenetic alterations	***Incidence of cases with indicated alterations [%]***	
	Intestinal type	***Diffuse type***
Oncogenes		
K-ras mutation	10	0
c-met amplification	19	39
K-sam amplification	0	33
c-*erb*B2 amplification	20	0
Adhesion molecules		
E-cadherin, mutation	0	50
Loss of CDH1	55	79[b]
CD44 aberrant transcript	100	100
Microsatellite instability	20–40	20–70[a]
(hMLH1 methylation)	(5–20)	(0)
Histone deacetylation	61	82
Telomere/Telomerase		
Telomere reduction	62	53
Telomerase activity	100	90
TERT expression	100	86

[a] Preferentially found in patients younger than 35 years of age.
[b] $p = 0.0175$; Fisher's exact test.

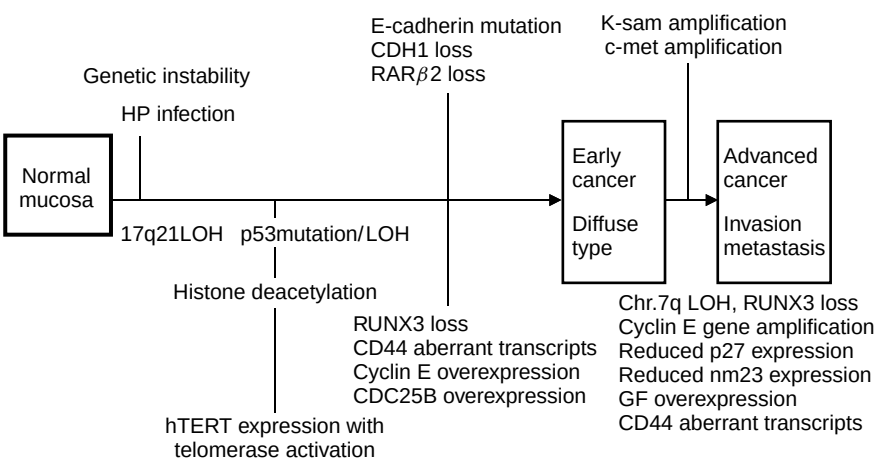

Fig. 3 Multiple genetic and epigenetic alterations during human gastric carcinogenesis (diffuse type).

RARβ2, *pS2* and *RUNX3* by DNA methylation is involved in two distinct major genetic pathways of gastric cancer. Hypermethylation of the p16 and of hMLH1 promoters is preferentially associated with intestinal type, whereas concordant hypermethylation of the CDH1 and RARβ2 promoters predominantly occurs in diffuse type gastric carcinoma. Loss of RUNX3 and pS2 expression by promoter methylation is a common event in both types of gastric carcinoma. The scenario of these epigenetic alterations indicates that there are at least two types of CpG island methylator phenotypes in intestinal and diffuse types of gastric cancer. Recently, Cho et al. reported that promoter hypomethylation of a novel cancer/testis antigen gene CAGE was found in 35% of chronic gastritis and in 78% of gastric cancer.

In addition to promoter hypermethylation or hypomethylation, acetylated histone H4 is obviously reduced in the majority of gastric carcinoma. Histone H4 is progressively deacetylated from the early stage (precancerous lesions) to the late stage (invasion and metastasis) in gastric carcinogenesis. Since there is no difference in the level of acetylated histone H4 between the intestinal and diffuse types of gastric cancer, histone H4 deacetylation is a common event in both types of gastric cancer.

In the multistep process of intestinal type carcinogenesis, the genetic pathways can be divided into three subways: an intestinal metaplasia → adenoma → carcinoma sequence, an intestinal metaplasia → carcinoma sequence and *de novo*. Infection with *H. pylori* may be a strong trigger for hyperplasia of hTERT-positive "stem cell" in intestinal metaplasia. Genetic instability and hyperplasia of hTERT positive stem cells precede replication error at the D1S191, DNA hypermethylation at the D17S5 locus, pS2 loss, RARβ2 loss, RUNX3 loss, CD44 abnormal transcripts and p53 mutation, all of which accumulate in at least 30% of incomplete intestinal metaplasia. All of these epigenetic and genetic changes are common events in intestinal type gastric cancer. An adenoma → carcinoma sequence is found in about 20% of gastric adenoma with APC mutations. In addition to these events, p53 mutation and LOH, RUNX3 loss, reduced p27 expression, cyclin E overexpression and abnormal transcript of c-met allow malignant transformation from the above precancerous lesions to intestinal type gastric carcinoma (Fig. 2). DCC loss, APC mutations, 1qLOH, p27 loss, reduced TGFβ receptor, reduced *nm23* and *c-erbB2* gene amplification are frequently associated with an advanced stage of intestinal-type gastric carcinoma. The *de novo* gastric carcinoma involves LOH and abnormal expression of *p73* gene that is responsible for the development of foveolar-type gastric cancers with pS2 expression.

On the other hand, LOH at chromosome 17p, mutation or LOH of p53, LUNX3 loss, and mutation or loss of E-cadherin are preferentially involved in the development of diffuse type gastric carcinoma. These events may occur simultaneously or in the relatively short term in superficial gastritis induced by *H. pylori* infection. In addition to these alterations, gene amplification of *K-sam* and *c-met*, RUNX3 loss, 7q LOH, cyclin E gene amplification, p27 loss as well as reduced nm23 confer progression, and metastasis, frequently associated with productive fibrosis. Mixed gastric carcinomas composed of intestinal and diffuse components exhibit some but not all of the molecular events described

so far for each of the two types of gastric cancer (Fig. 3).

Several proto-oncogenes, including *c-met, K-sam,* and *c-erbB2,* are frequently activated in gastric cancer. The amplification of the *c-met* gene encoding a receptor for hepatocyte growth factor/scatter factor (HGF/SF) is found in 19% of intestinal type and 39% of diffuse type gastric cancers. Most of gastric carcinomas express two different c-met transcripts, one of 7.0 kb and another of 6.0 kb. Expression of the 6.0 kb c-met transcript correlates well with tumor staging, lymph node metastasis, and depth of tumor invasion. The *K-sam* gene has at least four transcriptional variants. Type II transcript, which is expressed only in carcinoma cells, encodes a receptor for keratinocyte growth factor (KGF). *K-sam* is preferentially amplified in 33% of advanced diffuse or scirrhous type gastric cancer. But *K-sam* gene amplification is never seen in esophageal or colorectal carcinomas. In contrast to *K-sam, c-erbB2* is preferentially amplified in 20% of intestinal type gastric cancer but not in diffuse type gastric cancer. Overexpression of *c-erbB2* associated with gene amplification confers a poor prognosis and liver metastasis. The amplification of *c-erbB1* (EGFR) and *c-erbB3* is seen in 3 and 0%, respectively, of gastric cancer.

RUNX3, a Runt domain transcription factor involved in TGFβ signaling, is a candidate tumor-suppressor gene localized in 1p36, a region commonly deleted in a variety of human cancers, including gastric cancer. *RUNX* gene family is composed of three members, *RUNX1/AML1, RUNX2* and *RUNX3*, and encodes the DNA binding(α) subunits of the Runt domain transcription factor polyomavirus enhancer-binding protein 2 (PEBP2)/core-binding factor (CBF), which is a heterodimeric transcription factor. All three *RUNXs* play important roles in both normal developmental processes and carcinogenesis. *RUNX1,* which is required for definitive hematopoiesis, is the target of chromosome translocations in leukemia. *RUNX2,* which is essential for osteogenesis, is mutated in the human disease cleidocranial dysplasia. *RUNX3* is necessary for the suppression of cell proliferation of gastric epithelium, neurogenesis of the dorsal root ganglia, and T-cell differentiation. The gastric epithelium of RUNX3 knockout mice shows hyperplasia, reduced rate of apoptosis, and reduced sensitivity to TGFβ1, suggesting that the tumor-suppressor activity of RUNX3 operates downstream of the TGFβ signaling pathway.

Recent studies on RUNX3 methylation in human cancers demonstrated that loss of RUNX3 by hypermethylation of the promoter CpG island was detected not only in 64% of gastric cancer but also in 73% of hepatocellular carcinoma, 70% of bile duct cancer, 75% of pancreatic cancer, 62% of laryngeal cancer, 46% of lung cancer, 25% of breast cancer, 23% of prostate cancer, 12% of endometrial cancer, 2.5% of uterine cervical cancer, and 5% of colon cancer. The RUNX3 methylation is especially frequent in cancers from tissues of a foregut origin. Interestingly, the RUNX3 methylation is found in 8% of chronic gastritis, 28% of intestinal metaplasia, and 27% of gastric adenoma, but not in chronic hepatitis B. These findings suggest that RUNX3 is a target for epigenetic gene silencing in gastric carcinogenesis.

Cell adhesion molecules may also work as tumor suppressors. Mutations in the E-cadherin gene occur preferentially in 50% of diffuse type gastric carcinoma. E-cadherin mutations affecting exons 8 or 9 induce the scattered morphology, decreased cellular adhesion and increased

cellular motility of diffuse gastric cancers. The mutations are even detected in intramucosal carcinoma. E-cadherin germline mutations responsible for hereditary diffuse gastric cancer (HDGC) have been reported since 1998, but their frequency is extremely rare. The β- and γ-catenin mutations but not E-cadherin mutations bring about constitutive Tcf transcriptional activity in gastric and pancreatic cancer cells. As mentioned in esophageal SCC, cross-talk between β-catenin and receptor tyrosine kinases including c-met, EGFR and c-*erb*B2 is found in gastric cancer cells, leading to diffuse spreading or scattering of gastric cancer cells through enhanced Wnt signaling pathway. These results indicate that genetic and epigenetic alterations in E-cadherin and catenins are involved in the development and progression of diffuse and scirrhous-type gastric cancer.

As to alterations in cell-cycle regulators, the cyclin E gene is amplified in 15 to 20% of gastric carcinomas that are associated with overexpression. Gene amplification or overexpression of cycline E, or both cause aggressiveness and lymph node metastasis. Cyclin D1 gene amplification, on the other hand, is exceptional in gastric cancer but frequently occurs in esophageal SCC. p27, a member of the cip/kip family of CDK inhibitors, binds to a wide variety of cyclin/CDK complexes and inhibits kinase activity. We have found that growth suppression of interferon-β is associated with the induction of p27 in the gastric cancer cell line TMK-1. More importantly, reduction in p27 expression is frequently observed in advanced gastric cancers while it is well preserved in 90% of gastric adenomas and 85% of early cancers. Reduced expression of p27 significantly correlates with tumor invasion and nodal metastasis. The expression of p27 in gastric cancer is inversely correlated with the expression of cyclin E. Loss of p27 expression and gain of cyclin E promotes progression and metastasis of gastric cancer.

Reduction in p27 occurs at posttranslational levels, resulting from ubiquitin mediated proteosomal degradation rather than genetic abnormalities.

An important downstream target of cyclin/CDKs at G1/S transition is a family of E2F transcription factors. Gene amplification of E1F-1 is seen in 4% of gastric cancers and in 25% of colorectal cancers. Overexpression of *E2F* is detected in 40% of primary gastric carcinomas. In addition, E2F and cyclin E tend to be coexpressed in gastric cancer, whereas 70% of gastric cancers show lower levels of E2G-3 expression than corresponding normal mucosa. These results overall suggest that gene amplification and abnormal expression of the *E2F* gene may permit the development of gastric cancer.

3.2 Factors Associated with Increased Incidence of Gastric Cancer

Three major factors, including environmental factors, host factors and genetic factors, cooperatively affect the genesis of gastric cancer. Of these, environmental factors are the most important, as diet and cigarette smoking are primary offenders; in particular, the presence of carcinogens such as *N*-nitroso compounds and benzo[α]pyrene is directly linked to carcinogenesis. The mutation spectrum of the *p53* gene is different between intestinal type and diffuse type gastric cancers, as p53 mutation at A:T sites are common in intestinal type carcinoma whereas GC $\rightarrow$ AT transitions are predominant in diffuse type carcinoma. Carcinogenic *N*-nitrosoamines, which cause mainly GC $\rightarrow$ AT base substitutions, are found in many

foods and can also be produced from amines and nitrates in the acidic environment of the stomach.

As for host factors, meta-analysis of relationship between *H. pylori* infection and gastric cancer has indicated that *H. pylori* infection is associated with a twofold increased risk of gastric cancer. Younger *H. pylori* infected patients have a higher relative risk of gastric cancer than older patients. *H. pylori* infection is equally associated with intestinal-type and diffuse-type gastric cancers. In fact, the observations in a Mongolian gerbil model of stomach carcinogenesis show that *H. pylori* infection promotes stomach carcinogenesis induced by chemical carcinogens, and that histological types of gastric carcinoma may depend on the concentration of chemical carcinogens rather than *H. pylori* infection. Eradication of the bacteria evidently decreases the incidence of gastric cancer in the Mongolia gerbil model.

H. pylori infection produces reactive oxygen and nitrogen species that cause DNA damage, followed by chronic gastritis and intestinal metaplasia. Goto et al. reported that the expression of inducible nitric oxide synthase (iNOS) and nitrotyrosine in the gastric mucosa was significantly high in *H. pylori* infected patients who developed gastric cancer at least two years after the initial biopsies. These findings suggest that high production of iNOS and nitrotyrosine may participate in gastric carcinogenesis.

Not only can *H. pylori* activate NF-κB in gastric epithelial cells, but activated NF-κB activates the transcription of IL-1, IL-6, IL-8, TNF-α, and cyclooxygenase-2 (Cox-2). Successful eradication of *H. pylori* leads to downregulation of Cox-2 in the epithelial and stromal cells. High expression of Cox-2 mRNA, protein, and enzymatic activity is observed in the tumor cells of intestinal type gastric cancer. Loss of Cox-2 promoter methylation may enhance Cox-2 expression and promote gastric carcinogenesis associated with *H. pylori*.

Strain of *H. pylori* and genetic factors play a critical role in susceptibility to stomach carcinogenesis. Prinz et al. reported that CagA+/VacAs1+ strains of *H. pylori* that are blood-group antigen-binding adhesion (BabA2)-positive are associated with activity or chronicity of gastritis. Adherence of *H. pylori* via BabA2 may play a key role for efficient delivery of VacA and CagA. Moreover, Hatakeyama's group has recently shown that CagA binds an Src homology 2 (SH2)-containing tyrosine phosphatase SHP-2 in a tyrosine phosphorylation-dependent manner and stimulates the phosphatase activity of SHP-2. In addition, they found that prevalent CagA protein in East Asian countries are significantly more potent in binding SHP-2 and in inducing cellular morphological changes than are CagA proteins of Western isolates. Differences in the biological activity of Western and East Asian CagA protein, which are determined by variation in the tyrosine phosphorylation sites, may underlie the different incidences of gastric cancer in these two geographic areas. Regarding the development of gastric mucosa-associated lymphoid tissue (MALT) lymphoma, heat-shock protein 60 kDa (hsp60) of *H. pylori* is an important antigen in the pathogenesis of MALT lymphoma.

In addition to *H. pylori* strains, DNA polymorphism including HLA, MUC1, T-cell helper 1 and IL-1β has been reported to be associated with an increased risk of both atrophic gastritis induced by *H. pylori* and gastric cancer. In addition, El-Omar et al. reported that proinflammatory genotypes of tumor necrosis factor α and IL-10 are associated with increased risk

of gastric cancer. More excitingly, Magnusson et al. found that distinct HLA class IIDQ and DR alleles are associated with the development of gastric cancer and infection with *H. pylori*. The DQA1*0102 is associated with protection from *H. pylori* infection, whereas the DRB*1601 is associated with cancer development, particularly, *H. pylori*-negative diffuse type gastric cancer. These host factors as well as *H. pylori* strain may determine why some individual infected with *H.pylori* develop gastric cancer while others do not.

The most important factor implicated in gastric carcinogenesis is genetic instability including microsatellite instability (MSI) and chromosomal instability. MSI due to epigenetic inactivation of the hMLH1 is found in 15 to 39% of sporadic intestinal type gastric cancer, of which 70% are associated with loss of hMLH1 by hypermethylation of the hMLH1 promoter. Intestinal type gastric cancers with MSI often occur in patients over 73 years of age and often occur in the antrum pylori. They are also associated with abundant lymphocyte infiltration, a putative favorable prognosis, and multiple tumors. In addition, MSI at the locus D1S191 is found in 26% of intestinal metaplasia and 46% of intestinal type gastric cancer. An identical pattern of MSI at the locus D1S191 is detected in both intestinal type cancer and the adjacent intestinal metaplasia, suggesting the sequential development of intestinal adenocarcinoma from intestinal metaplasia.

On the other hand, diffuse type gastric cancers with MSI occur mostly in patients under 35 years of age, and are frequently accompanied by scirrhous type carcinoma. This type of cancer, however, harbors no germline mutation of hMLH1 and hMSH2 and no alteration at BAT-RII, but is frequently associated with LOH on chromosome 17q21 including the *BRCA1* gene. Loss of the *BRCA1* by promoter methylation may have implications for the genesis of diffuse type gastric cancer.

Chromosomal instability leading to DNA aneuploidy is also an underlying factor in cancer. Telomere length is necessary for maintaining chromosomal stability. Recent evidence indicates that in the absence of telomerase, telomere shortening can bring about telomere dysfunction that causes both DNA breaks and chromosome gain or loss. Conversely, telomerase can inhibit chromosomal instability. Most intestinal type gastric cancers have remarkably shortened telomere length, associated with high levels of telomerase activity and significant expression of human telomerase reverse transcriptase (hTERT). More importantly, over 50% of intestinal metaplasia, as well as adenoma, express low levels of telomerase activity. We have found that *H. pylori* infection is a strong trigger for hyperplasia of hTERT positive cells in intestinal metaplasia, followed by increased telomerase activity and telomere reduction. Therefore, telomere reduction and telomerase activation play the most critical roles in an initial step of gastric carcinogenesis.

Mutations in the *APC* gene participate in chromosomal instability. Recently Kaplan et al. reported that mutation in *APC* may be responsible for chromosomal instability in colon cancer. It remains to be examined whether gastric cancer cells carrying a truncated *APC* gene are defective in chromosome segregation. APC protein directly binds to a kinetochore protein and is an avid *in vitro* substrate of the mitotic check-point protein Bub1. There

is no mutation in the *hBub1* gene in gastric cancer.

3.3 Abnormal Growth Factor/Cytokine Network in Gastric Cancer

Gastric cancer cells express a broad spectrum or growth factors, cytokines or both, including TGF-α, TGF-β, EGF, amphiregulin (AR), cripto, heparin binding (HB)-EGF, PDGF, IGF II, basic fibroblast growth factor (bFGF), IL-1α, IL-6, IL-8, and OPN. These growth factors and cytokines function as autocrine, paracrine, and juxtacrine modulators for the growth of cancer cells, and they organize the complex interaction between cancer cells and stromal cells which play a key role in morphogenesis, invasion, neovascularization, and metastasis (Fig. 4). Interestingly, the expression pattern of these growth factors and cytokines by cancer cells differs in the two histological types of gastric carcinomas. The EGF family including EGF, TGFα, cripto and AR are commonly overexpressed in intestinal type carcinoma, whereas TGFβ, IGF II, and bFGF are predominantly overexpressed in diffuse type carcinoma. Coexpression of EGF/TGF-α, EGFR and cripto correlates well with the biological malignancy, as these factors induce metalloproteinases. Overexpression

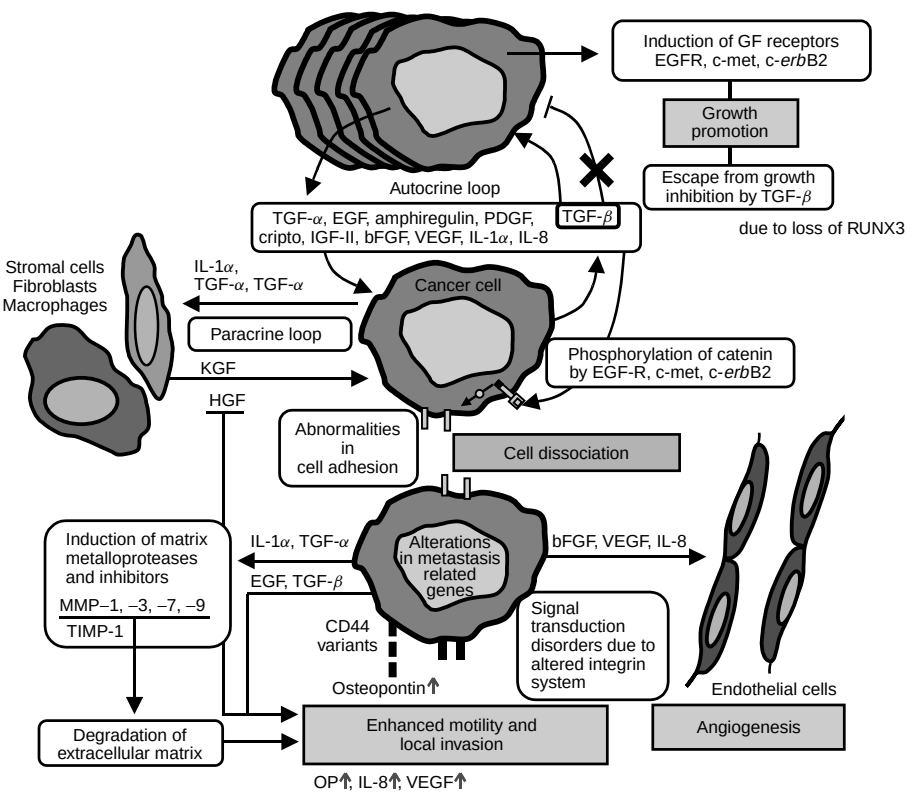

Fig. 4 Cancer-stromal interaction in gastric cancer through growth factors and cytokines.

of cripto is frequently associated with intestinal metaplasia and gastric adenoma. Akagi et al. have recently shown that gastric cancer cells express neutrophilin-1(NRP-1), which acts as a coreceptor for VEGF-165 and increases its affinity for VEGF receptor 2 endothelial cells. EGF induces both NRP-1 and VEGF expression, suggesting that regulation of NRP-1 expression in gastric cancer is intimately associated with EGF/EGFR system.

IL-1α is a cytokine mainly produced by activated macrophages through NF-κB activation and mediates many of the local and systemic responses to infection and inflammation. Gastric cancer cells also produce it. We have found that IL-1α acts as an autocrine growth factor for oral and gastric carcinoma cells and plays a pivotal role as a trigger for induction of EGF and EGFR expression. The expression of IL-1α by tumor cells is induced by either IL-1α, EGF, or TGF-α, while IL-α upregulates the expression of TGF-α and EGFR by tumor cells themselves, indicating that an intimate interplay between IL-1α and EGF/receptor system stimulates the growth of gastric cancer cells. In addition to IL-1α, IL-6 is also an autocrine growth stimulator for gastric cancer cells. The expression of IL-1α by tumor cells is induced by IL-6, while IL-1α increases the expression of IL-6 by tumor cells themselves. Currently, Fukayama's group reported that IL-1β may act as an autocrine growth factor in a human Epstein-Barr virus-associated gastric carcinoma.

IL-8, a member of the CXC chemokine family, induces haptotactic migration and proliferation of melanoma cells and angiogenesis. More importantly, gastric cancer cell lines express mRNA and protein for IL-8 and IL-8 receptors (IL-8RA and IL-8RB). More than 80% of primary tumors coexpress IL-8 and IL-8 receptor; this coexpression correlates directly with tumor vascularity and tumor progression. IL-8 enhances the expression of EGFR, type IV collagenase (MMP-9), VEGF, and IL-8 itself by tumor cells, while IL-8 decreases expression of E-cadherin. Moreover, IL-8 increases MMP-9 activity and the ability of gastric cancer cells to invade through Matrigel. IL-8 may play an important role in the growth and progression of gastric carcinoma by autocrine and paracrine mechanisms.

In addition to IL-8, VEGF and bFGF participate mainly in neovascularization in gastric cancer. We have shown that eight gastric cancer cell lines secrete VEGF into conditioned media. EGF or IL-1α upregulates VEGF expression by tumor cells, whereas interferon-γ downregulates it. VEGF promotes angiogenesis and the progression of gastric carcinoma, especially intestinal type. VEGF-C produced by tumor cells participates in the development of lymph node metastasis. On the other hand, bFGF produced by tumor cells is frequently associated with angiogenesis and extensive fibrosis in diffuse type carcinoma, particularly those of the scirrhous type. Interestingly, Nakazawa et al. reported that keratinocyte growth factor (KGF) produced by gastric fibroblasts specifically binds to K-sam on tumor cells and then stimulates proliferation of cancer cells, resulting in the development of the scirrhous type of gastric cancer. KGF from gastric fibroblasts may underline the remarkable proliferation of scirrhous gastric cancer cells in a paracrine manner.

Stromal cells, especially fibroblasts stimulated by growth factors or cytokines such as IL-1α, TGF-α, and TGFβ, secrete HGF/SF, which can function in a paracrine manner as a morphogen or mitogen of tumor cells. For example, in the case

of a cancer cell clone maintaining expression of cell adhesion molecules, HGF/SF promotes tubular formation of tumor cells, resulting in intestinal type gastric cancer. Conversely, in the case of a clone with reduced expression of cell adhesion molecules, HGF/SF can act as a mitogen and induce scattering of tumor cells, resulting in diffuse type gastric cancer.

OPN, also termed Eta-1 (early T-lymphocyte activation-1), which is a reported protein ligand of CD44, is overexpressed in 73% of gastric cancer. The coexpression of OPN and CD44 v9 in tumor cells correlates with the nodal metastasis in diffuse type gastric cancer. The *CD44* gene contains at least 20 exons, 12 of which can be alternatively spliced to make up a wide variety of molecular variants. All gastric cancer cell lines and primary tumors show overexpression of abnormal CD44 transcripts containing the intron 9 sequence. The intestinal metaplasia also expresses these variants but normal gastric mucosa does not express them. Currently, Medico et al. reported that OPN is an autocrine mediator of HGF induced invasive growth.

4 Genetic and Epigenetic Alterations and Abnormal Growth Factor/Cytokine Network in Colorectal Cancer

Cancer of the colon and rectum is the fourth most common cancer in the world. In 1996, an estimated 875 000 new cases were diagnosed worldwide, accounting for 8.5% of all new cancers. Approximately 98% of malignant colorectal tumors are adenocarcinoma. Rectal tumors account for about 27%, while almost 50% occur proximal to the splenic flexure.

The accumulation of multiple genetic and epigenetic alterations in tumor-suppressor genes, oncogenes and DNA mismatch repair genes takes place in the multistep process of colorectal carcinogenesis. Inactivation of APC, p53, and DCC and K-ras mutations are involved in a major genetic pathway for colorectal tumorigenesis showing the course of malignant progression from normal mucosal cells through adenomas (adenoma–carcinoma sequence) (Fig. 5). This section will make an overview of multiple

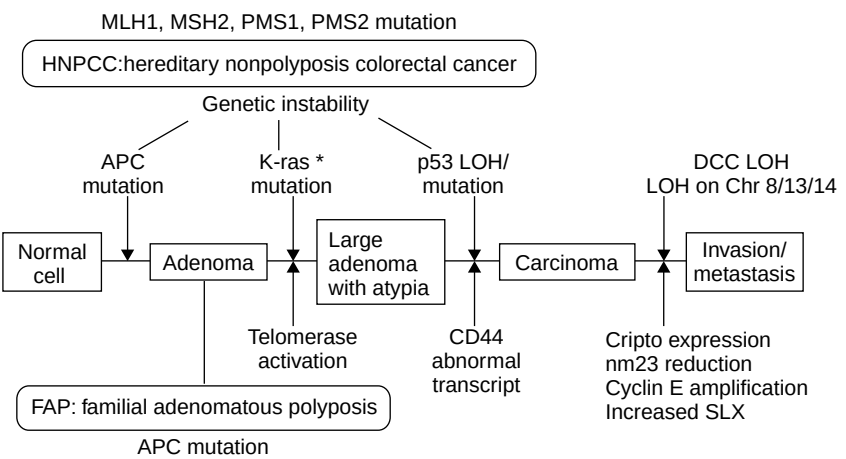

Fig. 5 A major genetic pathway for colorectal carcinogenesis. (*K-ras mutation is infrequent in flat adenoma)

genetic and epigenetic alterations responsible for colorectal carcinogenesis and abnormal growth factor/cytokine network, which is implicated in the progression and metastasis of colorectal cancer.

4.1 Genetic and Epigenetic Alterations in Colorectal Cancer

The *APC* gene, first isolated as a causative gene for familial adenomatous polyposis (FAP), encodes a large protein of 2843 amino acids, which forms a complex with α-and β-catenins and may mediate cell adhesion, cytoskeletal anchoring, and signal transduction. The *APC* gene is abnormal in the germline of FAP patients. LOH and mutations of the *APC* gene occur in 60% of sporadic colorectal adenomas and adenocarcinomas. These tumors harbor loss of the *APC* gene in one allele and mutation of the gene in the remaining allele, supporting Knudson's two-hit theory. The characteristic of the mutation is a base substitution that leads to a stop codon (nonsense mutation), which occurs in about 70% of FAP kindred. The APC alterations are found even in small adenomas with mild atypia. The tumor-suppressor function of APC has also been demonstrated in mouse models. These results and animal models show that inactivation of APC is an initiating genetic event for colorectal tumorigenesis.

The major tumor-suppression activity of APC is regulation of β-catenin. APC's association with β-catenin and promotion of the degradation of β-catenin are most relevant to its tumor-suppressor function. APC binds not only β-catenin but also glycogen synthase kinase-3β (GSK3β) and Axin. APC and Axin serve as a scaffold to facilitate the phosphorylation of β-catenin by GSK3β. The phosphorylated β-catenin is ubiquitinated by β-Trcp and the ubiquitinated β-catenin is then degraded by proteasome. Mutant APC proteins cannot regulate the degradation of β-catenin, resulting in excessive β-catenin that interacts with T-cell factor (TCF)-4 and translocates into the nucleus. The β-catenin/TCF-4 complex then activates the expression of many genes including *c-myc*, matrix metalloproteinase-7 (*MMP-7*), peroxisome proliferator-activated receptor δ (*PPARδ*) and cyclin D1, leading to promotion of colon tumorigenesis. In fact, overexpression of *c-myc, PPARδ* and cyclin D1 is observed in colorectal cancer. A recent study of PPARδ deficient (Ppard−/−) mice has shown that colon tumor formation is significantly greater in mice nullizygous for PPAR, suggesting that PPARδ attenuates colon carcinogenesis. It is contrary to previous reports suggesting that activation of PPARδ is causally associated with colon polyp formation. Further work is necessary to clarify the role of PPARδ in colon carcinogenesis.

On the other hand, somatic mutations of β-catenin are detected in both human and rodent colorectal tumors that do not have the APC mutation. All the β-catenin mutations found in colorectal cancers occur at the critical region for phosphorylation of β-catenin by GSK3β. The mutant β-catenin is resistant to APC-mediated degradation.

Mutation in the *K-ras* oncogene is involved in the progression from small adenoma with mild atypia to large adenoma with severe atypia. About 40 to 50% of large adenomas with severe atypia and adenocarcinomas contain K-ras point mutations at codon 12 or 13, compared to only 10% of small adenomas with mild atypia. Conversely, the frequency of K-ras mutation is lower (<10%) in superficial-type or flat adenomas even in the presence

of significant atypia. Much evidence indicates that mutation of K-ras alone can bring about a hyperplastic lesion that has a limited potential to progress to larger tumors. Mutant K-ras can, however, promote tumor progression in lesions initiated by APC mutation.

LOH of the *p53* gene locus is detected in about 80% of colorectal adenocarcinoma, and most harbor inactivation of the *p53* gene in both alleles. Because only 5 to 20% of the adenomas have p53 inactivation, it must play a crucial key in the transition from adenoma to adenocarcinoma. There are hot spots for point mutations in the highly conserved regions such as codon 175, 248, and 273, where G:C to A:T transitions occur. Abnormal accumulation of p53 protein detected by immunohistochemistry is frequently associated with deeply invasive carcinomas and carcinomas with metastasis. Almost all mutant p53 proteins derived from cancers have altered sequence-specific DNA binding and transcription activities.

Overexpression of *c-erbB2* has been reported in 80 to 100% of colorectal cancers. The amplification of *c-erbB2*, which is common in intestinal type gastric cancer, occurs in about 10% of colorectal carcinomas. Moreover, the cyclin E gene, a positive regulator of cell cycle progression, is amplified in about 10% of colorectal carcinomas. The overexpression of cyclin E is detected in 5% of adenomas and in 20% of adenocarcinomas. Among adenomas, a significant correlation is observed between cyclin E expression and the grade of atypia. Overexpression of cyclin E is prominent in carcinoma invading the submucosa or deeper compared to those limited to the mucosal layer. Cyclin E expression is thus a candidates molecular biomarker for predicting malignant progression of colorectal as well as gastric cancers. Reduction in p27 expression participates in progression and poor prognosis of colorectal cancer as well.

LOH of 18q including *DCC, Smad4,* and *Smad2* genes is frequently associated with advanced colorectal cancer. Among them, *DCC* and *Smad4* play important roles in colon cancer progression. LOH of the *DCC* gene is rare in adenoma but frequent (about 70%) in adenocarcinoma. LOH of *DCC* increases as the tumor invades deeply, and almost all the metastatic liver tumors show this LOH. Moreover, reduced expression of *DCC* in colorectal cancer is correlated with a poor prognosis. *DCC* encodes a receptor for netrin-1, but its function in normal colon epithelial cells remains unclear. *Smad4* encodes a protein that plays a critical role in the TGF-β signal transduction pathway. Although *Smad4* was isolated as a tumor-suppressor gene for pancreatic cancer, somatic mutations of *Smad4* frequently take place in advanced colon cancer, suggesting that Smad4 inactivation confers progression of colorectal cancer. In addition, *Smad4* germline mutations are responsible for juvenile polyposis, an autosomal disease that has high susceptibility for hamartomatous polyposis and gastrointestinal cancer.

In addition to these molecular events, and as mentioned for gastric cancer, telomere reduction may result in chromosomal instability and telomerase reactivation. The colorectal adenomas and adenocarcinomas share shorter telomeres than those in normal tissues. We have found that more than 90% of colorectal adenocarcinomas express extremely high levels of telomerase activity regardless of tumor staging and histological differentiation. All the adenomas also exhibit considerable levels of telomerase activity. Telomerase activity and stabilization of telomeres occur concomitantly with the acquisition of

immortality, contributing to an early stage of colorectal carcinogenesis.

Beside the major genetic pathway (adenoma–carcinoma sequence), an alternative genetic pathway exists for colorectal carcinogenesis. The so-called *de novo* carcinogenesis exhibits no adenoma–carcinoma sequence but develops directly from the colorectal epithelial cells that share p53 inactivation, followed by APC inactivation. K-ras mutation is not detected in *de novo* carcinogenesis. This type of colorectal carcinogenesis is frequently found in Japanese patients.

Tab. 4 Differences between MSI positive and MSI negative colorectal cancers.

	MSI +	*MSI −*
Location	Proximal	Distal
Ploidy	Near diploid	Aneuploid
Chromosomal instability	Rare	Common
Mutation Frequency		
P53	Low	High
APC	Low	High
TGFβ RII	High	Low
BAX	High	Low
Methylation	High	Low
Survival	Better	Worse
Hereditary syndrome	HNPCC	FAP

4.2 Factors Associated with Increased Incidence of Colorectal Cancer

Genetic and epigenetic alterations in DNA mismatch repair genes including hMHL1 and hMSH2 raise MSI that has implications for predisposition to colorectal cancer. The MSI occurs in 10 to 15% of sporadic colorectal cancer. These sporadic, mostly right-side colon cancers with MSI are associated with hypermethylation of the CpG islands in the hMLH1 promoter, resulting in loss of hMLH1, loss of function of other genes such as *p16*, defective mismatch repair, and widespread MSI. Hereditary nonpolyposis colorectal cancer (HNPCC), caused by inherited germline mutations in hMLH1 and hMLH2, accounts for 3 to 10% of colorectal cancer. Genes coding for TGFβ type II receptor, insulin-like growth factor (IGF)2 receptor, proapoptotic protein BAX, cell cycle regulator E2F-4, and mismatch repair proteins MSH3 and MSH6 are mutated in HNPCC or MSI sporadic colon cancer.

Some differences between MSI positive and MSI negative colorectal cancers are presented in Table 4. Tumor location, ploidy, mutation frequency, methylation, and survival are different between MSI positive and MSI negative tumors, although there are overlaps between the two.

In addition, sporadic colon cancer with MSI is also associated with altered expression of IGF2, namely, loss of imprinting (LOI) of IGF2. Importantly, the normal colonic mucosa exhibits aberrant hypermethylation and LOI of IGF2 as a sign of a field defect. However, Feinberg et al. have recently reported that hypomethylation of H19 and IGF2 is a mechanism for LOI and is found in both colorectal cancers and normal mucosa from the same patients. Moreover, they reported that LOI of IGF2 provides a potential heritable biomarker for colon cancer predisposition.

NF-κB activation is also associated with chronic inflammatory bowel diseases (IBD) and colorectal cancer. As mentioned in Sect. 3.2, NF-κB activation leads to production of enzymes such as iNOS and COX2 and enhanced expression of growth factors and cytokines. IBD including ulcerative colitis and Crohn's disease

induce persistent NF-κB activation in tissue macrophages and epithelial cells of the colonic mucosa. Both inflammatory bowel diseases are well known to increase the risk of colorectal cancer. The link between COX2 and colorectal cancer is supported strongly by epidemiological and experimental evidence. COX2 is overexpressed in adenomas and carcinomas of the colon. COX2-null mice are resistant to colorectal carcinogenesis. Long-term consumption of aspirin or other COX inhibitors has been reported to reduce the relative risk of colorectal cancer. These results indicate that COX2 contributes to colorectal tumorigenesis.

4.3 Abnormal Growth Factor/Cytokine Network in Colorectal Cancer

Colorectal carcinomas express multiple growth factors, such as EGF and TGFα and their receptors thus creating autocrine loops. TGFα and EGF are overexpressed in colorectal adenomas and the majority of colorectal carcinomas. Coexpression of TGFα, EGF or both, and EGFR is well correlated with high grade of malignancy and metastasis.

As described in gastric cancer, NRP-1 induced by EGF is expressed in all of colorectal cancer tissues and cell lines but not in the adjacent nonmalignant colonic mucosa. A recent study of NRP-1 in colon cancer suggests that NRP-1 may contribute to colon cancer angiogenesis and that EGF and mitogen-activated protein kinase signaling may play an important role in NRP-1 regulation in colon cancer cells.

The cripto gene was originally identified in undifferentiated human embryonal carcinoma cells and encodes a 37 amino acid region that shares structural homology with other members of the EGF family. However, cripto does not bind to the EGFR and its receptor has not been identified. Strong expression of mRNA and protein for cripto is found in 60 to 80% of colorectal cancers but not in normal colorectal mucosa. It is detected in 40% of tubular adenomas and 86% of tubulovillous adenomas, respectively. These findings suggest that cripto expression may be involved in the early stages of malignant transformation. Amphiregulin (AR) is another member of the EGF family that utilizes EGFR. About half of colorectal carcinomas as well as 60% of adenomas express AR. It has been confirmed that cripto and AR act as autocrine growth stimulators for colorectal cancer cell lines. Because AR is also expressed in normal colorectal epithelium, AR may participate in the regulation of growth of normal as well as colorectal cancer cells.

TGFβ-1 is expressed in over half of colorectal cancers. Interestingly, high levels of TGFβ-1 expression in tumor cells and elevated plasma levels of mRNA are associated with advanced Dukes' stage, suggesting that there is a correlation between TGFβ overexpression and tumor progression. Moreover, circulating TGFβ-1 may serve as a predictor of liver metastasis after resection of colorectal cancer. In addition, TGFβ produced by cancer cells stimulates angiogenesis by inducing thymidine phosphorylase, regulation of extracellular matrix adhesion molecules such as carcinoembryonic antigen (CEA), and by the enhanced secretion of gelatinase B, a matrix degrading enzyme.

VEGF and bFGF are also expressed strongly in colorectal cancer in contrast to normal colorectal epithelium and adenoma. The expression of bFGF is higher in Dukes stage D than in Dukes stage B colorectal cancer. Moreover, bFGF and

VEGR are elevated in the serum of patients with aggressive advanced colorectal cancer. These circulating growth factors as well as TGFβ may be useful biomarkers for understanding angiogenesis and malignancy. On the other hand, the expression of VEGF-C and VEGF-D correlates with lymph-node metastasis in colorectal carcinoma, and these expressions are heterogeneous and elevated at the invasive edge of tumors.

Activation of the pp60src protein kinase activity occurs during colorectal tumorigenesis. The kinase activity of pp60src is highly regulated and is induced by many growth factors. Recently, pp60src has been reported to be essential for the induction of VEGF by hypoxia. The specific activity of pp60src is higher in colorectal polyp than in normal mucosa, and is further increased in colorectal carcinoma and in metastatic colon tumor in liver. Further study is needed to clarify the mechanism for increased pp60src activity during colorectal carcinogenesis.

The abnormal transcripts of the *CD44* gene are also expressed in all of colorectal cancers. As mentioned in gastric cancer, the *CD44* gene consists of at least 20 exons, of which 10 are alternatively spliced to make up variants. Among several *CD44* variants, aberrant transcripts with retention of intron 9 are best for distinguishing carcinoma tissues from normal tissues in the colorectum. However, the variants do not correlate with nodal or distant metastasis.

A candidate suppressor gene of tumor metastasis, *nm23*, encodes nucleotide diphosphate kinase and c-myc transcription factor (PuF). Although most of colorectal cancers express nm23 at higher levels than the corresponding normal mucosa, an inverse correlation is observed between nm23 expression and tumor staging. Moreover, reduced expression of nm23 is associated with distant metastasis.

Another candidate for a molecular marker that indicates metastatic potential is cell surface carbohydrate, sialyl-dimeric Le antigens. Both sialyl Lex (SLX) and sialyl Lea (SLA or Ca 19-9) as ligands bind to E-selectin, also known as ELAM-1, one of the adhesion molecules on activated endothelial cells. SLX and SLA may participate in distant metastasis through interaction between cancer cells and endothelial cells. The expression of SLX in colorectal carcinomas shows significant correlation with liver metastasis and poor prognosis.

5 Conclusion

A large number of molecular events are involved in the development and progression of gastrointestinal carcinomas. Among them, common and distinct events of genetic and epigenetic alterations are observed in esophageal, gastric, and colorectal cancers. MSI confers the initial step of gastric and colorectal carcinomas, while it is less involved in esophageal SCC. Chromosomal instability (telomere reduction) and telomerase activation participate commonly in the very early stage of gastrointestinal carcinogenesis. p53 inactivation and RUNX3 loss by promoter hypermethylation are also common events, although RUNX3 loss is less in colorectal cancer. APC LOH and DCC LOH are commonly detected in the majority of the three gastrointestinal cancers, although APC mutations occur mainly in colorectal cancer. K-ras mutation is often found in colorectal cancer, whereas it is extremely rare in esophageal SCC and gastric cancer. Amplification of the cyclin D1 gene

is preferentially found in esophageal SCC, while the gene amplification of cycline E is frequently associated with both gastric and colorectal adenocarcinomas. Reduced expression of the CDK inhibitors such as p16 and p27 is often found in gastrointestinal cancers. In gastric cancer, the pattern of genetic and epigenetic alterations also differs depending on the two histological types, intestinal or well-differentiated type and diffuse or poorly differentiated type. The amplification of *c-met* and *K-sam* genes and the mutation/loss of the E-cadherin gene as well as *RARβ2* loss occur preferentially in diffuse type, whereas the amplification of *c-erbB2* gene, pS2 reduction, p16 loss and hMLH1 loss as well as APC mutation is predominantly found in intestinal type.

In addition to these events, gastrointestinal cancer cells express a broad spectrum of the growth factor/cytokine receptor systems that organize complex interactions between cancer cells and stromal cells, which confer cell growth, apoptosis, morphogenesis, angiogenesis, progression and metastasis. However, these abnormal growth factor/cytokine networks are also evidently different among esophageal, gastric, and colorectal cancers, respectively. Importantly, NF-κB activation induced by inflammation may act as a key player for induction of growth factor/cytokine networks in gastrointestinal cancers.

Overall, the observations on the molecular events involving growth factors and oncogenes in gastrointestinal cancers will no doubt provide a deeper understanding of prevention, molecular diagnosis, and therapeutics of these cancers. In fact, by applying these molecular events of gastrointestinal cancers to routine clinical practice, we have implemented molecular pathological diagnosis of gastrointestinal cancer in collaboration with Hiroshima City Medical Clinical Laboratory since 1993. We have analyzed more than 10 000 cases of gastrointestinal biopsy and surgery and then obtained additional information on differential diagnosis, biological malignancy, and tumor multiplicity. We believe this approach will better serve science, but more importantly, patient care.

See also Growth Factors; Oncology, Molecular.

Bibliography

Books and Reviews

Correa, P., Miller, M., Mannick, E.E. (1997) Oxidative Damage During the Gastric Precancerous Process, in: Tahara, E. (Ed.) *Molecular Pathology of Gastroenterological Cancer*, Springer-Verlag, Tokyo, Japan, pp. 23–29.

Fenoglio-preiser, C., Noffsinger, A., Stemmermann, G. (2004) Pathology and Natural History of Gastric Cancer, in: Abbruzzese, J.L., Evans, D.B., Willett, C.G., Fenoglio-Preiser, C. (Eds.) *Gastrointestinal Oncology*, Oxford University Press, New York, pp. 281–293.

Fidler, I.J. (1997) Critical Determinants of Human Colon Cancer Metastasis, in: Tahara, E. (Ed.) *Molecular Pathology of Gastroenterological Cancer*, Springer-Verlag, Tokyo, Japan, pp. 15–21.

Grizzle, W.E., Shibata, D., Manne, U., Myers, R.B., Frost, A.R. (1999) Molecular and Histopathologic Changes in the Development of Colorectal Neoplasia, in: Srivastava, S. Henson, D.E. Gazdar, A. (Eds.) *Molecular Pathology of Early Cancer*, IOS Press, Amsterdam, Netherlands, pp. 135–170.

Mao, L., El-Naggar, A.K. (1999) Molecular Changes in the Multistage Pathogenesis of Head and Neck Cancer, Srivastava, S., et al. (Eds.) *Molecular Pathology of Early Cancer*, IOS Press, Amsterdam, Netherlands, pp. 189–205.

Nakamura, Y., Tokino, T., Isomura, M., Inazawa, J., Aoki, T., Mori, T., Shimada, M.,

Miura, K., Suzuki, K. (1997) Multistep Carcinogenesis of Esophageal Carcinoma, in: Tahara, E. (Ed.) *Molecular Pathology of Gastroenterological Cancer*, Springer-Verlag, Tokyo, Japan, pp. 15–21.

Su, L.K. (2004) Molecular Biology of Colorectal Cancer, in: Abbruzzese, J.L., et al. (Eds.) *Gastrointestinal Oncology*, Oxford University Press, New York, pp. 635–644.

Tahara, E. (1997) Cell Growth Regulation and Cancer: Stromal Interaction, in: Sugimura, T. Sasaki, M. (Eds.) *Gastric Cancer*, Oxford University Press, Tokyo, Japan, pp. 100–108.

Tahara, E. (2004) Genetic Pathways of Two Types of Gastric Cancer, in: Buffler, P., et al. (Eds.) *Mechanisms of Carcinogenesis: Contributions of Molecular epidemiology*, IARC Scientific Publications No. 157, International agency for Research on Cancer, Lyon, France, pp. 327–349.

Tahara, E., Yasui, W., Yokozaki, H. (1996) Abnormal Growth Factor Networks in Neoplasia, in: Pusztai, L., et al. (Eds.) *Cell Proliferation in Cancer: Regulatory Mechanisms of Neoplastic Cell Growth*, Oxford Medical Publications, New York, pp. 131–153.

Tahara, E., Yokozaki, H., Yasui, W. (1993) Growth Factors in Gastric Cancer, in: Nishi, M. Ichikawa, H. Nakajima, T. Maruyama, K. Tahara, E. (Eds.) *Gastric Cancer*, Springer-Verlag, Tokyo, Japan, pp. 209–217.

Tahara, E., Yokozaki, H., Yasui, W. (1999) Stomach–Genetic and Epigenetic Alterations of Preneoplastic and Neoplastic Lesions, in: Srivastava, S. Henson, D.E. Gazdar, A. (Eds.) *Molecular Pathology of Early Cancer*, IOS Press, Amsterdam, Netherlands, pp. 341–361.

Tahara, H., Tahara, E. Jr., Tahara, E., Ide, T. (1997) Telomeres and Telomerase in Gastrointestinal Cancer, in: Tahara, E. (Ed.) *Molecular Pathology of Gastroenterological Cancer*, Springer-Verlag, Tokyo, Japan, pp. 245–259.

Tselepis, C., Jankowski, J. (2004) Molecular Biology of Esophageal Cancer, in: Abbruzzese, J.L., et al. (Eds.) *Gastrointestinal Oncology*, Oxford University Press, New York, pp. 155–166.

Yasui, W., Yokozaki, H., Tahara, E. (1997) Molecular Diagnosis of Gastrointestinal Cancer, in: Tahara, E. (Ed.) *Molecular Pathology of Gastroenterological Cancer*, Springer-Verlag, Tokyo, Japan, pp. 187–207.

Yokozaki, H., Yasui, W., Tahara, E. (2001) Genetic and Epigenetic Changes in Stomach Cancer, in: Jeon, K.W. (Ed.) *International Review of Cytology Vol.20: A Survey of Cell Biology*, Academic Press, New York, pp. 49–95.

Yoshida, K., Yasui, W., Kagawa, Y., Tahara, E. (1997) Multiple Genetic Alterations and Abnormal Growth Factor Network in Human Esophageal Carcinomas, in: Tahara, E. (Ed.) *Molecular Pathology of Gastroenterological Cancer*, Springer-Verlag, Tokyo, Japan, pp. 31–41.

Primary Literature

Akagi, M., Kawaguchi, M., Liu, W., McCarty, M.F., Takeda, A., Fan, F., Stoeltzing, O., Parikh, A.A., Jung, Y.D., Bucana, C.D., Mansfield, P.F., Hicklin, D.J., Ellis, L.M. (2003) Induction of neuropilin-1 and vascular endothelial growth factor by epidermal growth factor in human gastric cancer cells, *Br. J. Cancer* **88**, 796–802.

Akhtar, M., Cheng, Y., Magno, R.M., Ashktorab, H., Smoot, D.T., Meltszer, S.J., Wilson, K.T. (2001) Promoter methylation regulates *Helicobacter pylori*-stimulated cyclooxygenase-2 expression in gastric epithelial cells, *Cancer Res.* **61**, 2399–2403.

Akiyama, Y., Maesawa, C., Ogasawara, S., Terashima, M., Masuda, T. (2003) Cell-type-specific repression of the maspin gene is disrupted frequently by demethylation at the promoter region in gastric intestinal metaplasia and cancer cells, *Am. J. Pathol.* **163**, 1911–1919.

Aoki, T., Takeda, S., Yanagisawa, A. (1994) APC and p53 mutations in *de novo* colorectal adenocarcinomas, *Hum. Mutat.* **3**, 342–346.

Arber, N., Hibshoosh, H., Moss, S.F., Sutter, T., Zhang, Y., Begg, M., Wang, S., Weinstein, I.B., Holt, P.R. (1996) Increased expression of cyclin D1 is an early event in multistage colorectal carcinogenesis, *Gastroenterology* **110**, 669–674.

Arber, N., Gammon, M.D., Hibshoosh, H., Britton, J.A., Zhang, Y., Schonberg, J.B., Roterdam, H., Fabian, I., Holt, P.R., Weinstein, I.B. (1999) Overexpression of cyclin D1 occurs in both squamous carcinomas and adenocarcinomas of the esophagus and in adenocarcinomas of the stomach, *Hum. Pathol.* **30**, 1087–1092.

Bodmer, W.F., Bailey, C.J., Bodmer, J. et al. (1987) Localization of the gene for familial

adenomatous polyposis on chromosome 5, *Nature* **328**, 614–616.

Bronner, C.E., Baker, S.M., Morrison, P.T., Warren, G., Smith, L.G., Lescoe, M.K., Kane, M., Earabino, C., Lipford, J., Lindblom, A. et al. (1994) Mutation in the DNA mismatch repair gene homologue *hMLH1* is associated with hereditary nonpolyposis colon cancer, *Nature* **368**, 258–261.

Carvalho, F., Seruca, R., David, L., Amorin, A., Seixas, M., Bennett, E., Clausen, H., Sobrinho-Simoes, M. (1997) MUC1 gene polymorphism and gastric cancer–an epidemiological study, *Glycoconjugate J.* **14**, 107–111.

Chen, L., Matsubara, N., Yoshino, T., Nagasaka, T., Hoshizima, N., Shirakawa, Y., Naomoto, Y., Isozaki, H., Riabowol, K., Tanaka, N. (2001) Genetic alterations of candidate tumor suppressor ING1 in human esophageal squamous cell cancer, *Cancer Res.* **61**, 4345–4349.

Cho, B., Lee, H., Jeong, S.W., Bang, Y.J., Lee, H.J., Hwang, K.S., Kim, H.Y., Lee, Y.S., Kang, G.H., Jeoung, D.I. (2003) Promoter hypomethylation of a novel cancer/testis antigen gene CAGE is correlated with its aberrant expression and is seen in premalignant stage of gastric carcinoma, *BBRC* **307**, 52–63.

Chong, J.M., Sakuma, K., Sudo, M., Osawa, T., Ohara, E., Uozaki, H., Shibahara, J., Kuroiwa, K., Tominaga, S., Hipp, Y., Aburatani, H., Funata, N., Jukayama, M. (2002) Interleukin-1β expression in human gastric carcinoma with Epstein-Barr Virus Infection, *J. Virol.* **76**, 6825–6831.

Crawford, H.C., Fingleton, B.M., Rudolph-Wwen, L.A., Goss, K.J., Rubinfeld, B., Polakis, P., Matrisian, L.M. (1999) The metalloproteinase matrilysin is a target of β-catenin transactivation in intestinal tumors, *Oncogene* **18**, 2883–2891.

Cruz-Correa, M., Cui, H., Giardiello, F.M., Powe, N.R., Hylind, L., Robinson, A., Hutcheon, D.F., Kafonek, D.R., Brandenburg, S., Wu, Y., He, X., Feinberg, A.P. (2004) Loss of imprinting of insulin growth factor II gene: a potential heritable biomarker for colon neoplasia predisposition, *Gastroenterology* **126**, 964–970.

Cui, H., Onyango, P., Brandenburg, S., Wu, Y., Hsieh, C.L., Feinberg, A.P. (2002) Loss of imprinting in colorectal cancer linked to hypomethylation of *H19* and *IGF2*, *Cancer Res.* **62**, 6442–6446.

Cui, H., Cruz-Correa, M., Giardiello, F.M., Hutcheon, D.F., Kafonek, D.R., Brandenburg, S., Wu, Y., He, X., Powe, N.R., Feinberg, A.P. (2003) Loss of *IGF2* imprinting: a potential marker of colorectal cancer risk, *Science* **299**, 1753–1755.

Cunningham, J.M., Christensen, E.R., Tester, D.J., Kim, C.Y., Roche, P.C., Burgart, L.J., Thibodeau, S.N. (1998) Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability, *Cancer Res.* **58**, 3455–3460.

Eden, A., Gaudet, F., Waghmare, A., Jaenisch, R. (2003) Chromosomal instability and tumors promoted by DNA hypomethylation, *Science* **300**, 455.

El-Omar, E.M., Rabkin, C.S., Gammon, Md., Vaughan, T.L., Risch, H.A., Schonberg, J.B., Stanford, J.L., Mayne, S.T., Goedert, J., Blot, W.J., Fraumeni, J.F. Jr., Chow, W.H. (2000) Increased risk of noncardia gastric cancer associated with proinflammatory cytokine gene polymorphisms, *Gastroenterology* **124**, 1193–1201.

El-Omar, E.M., Carrington, M., Chow, W.H., McColl, K.E., Bream, J.H., Young, H.A., Herrera, J., Lissowska, J., Yuan, C.C., Rothman, N., Lanyon, G., Martin, M., Fraumeni, J.F. Jr., Rabkin, C.S. (2000) Interleukin-1 polymorphisms associated with increased risk of gastric cancer, *Nature* **404**, 398–402.

Eppert, K., Scherer, S.W., Ozcelik, H., Pirone, R., Hoodless, P., Kim, H., Tsui, L.C., Bapat, B., Gallinger, S., Andrulis, I.L., Thomsen, G.H., Wrana, J.L., Attisano, L. (1996) *MADR2* maps to 18q21 and encodes a TGFβ-regulated MAD-related protein that is functionally mutated in colorectal carcinoma, *Cell* **86**, 543–552.

Fearon, E.R., Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis, *Cell* **61**, 759–767.

Feilchenfeldt, J., Brundler, M.A., Soravia, C., Totsch, M., Meier, C.A. (2004) Peroxisome proliferator-activated receptors (PPARs) and associated transcription factors in colon cancer: reduced expression of PPARgammacoactivator 1 (PGC-1), *Cancer Lett.* **203**, 25–33.

Feinberg, A., Tycho, B. (2004) The history of cancer epigenetics, *Nat. Rev. Cancer* **4**, 143–153.

Goto, T., Haruma, K., Kitadai, Y., Ito, M., Yoshimura, M., Sumii, K., Hayakawa, N.,

Kajiyama, G. (1999) Enhanced expression of inducible nitric oxide synthase and nitrotyrosine in gastric mucosa of gastric cancer patients, *Clin. Cancer Res.* **5**, 1411–1415.

Guilford, P.J., Hopkins, J.B., Grady, W.M., Markowitz, S.D., Willis, J., Lynch, H., Rajput, A., Wiesner, G.L., Lindor, N.M., Burgart, L.J., Toro, T.T., Lee, D., Limacher, J.M., Shaw, D.W., Findlay, M.P., Reeve, A.E. (1999) E-cadherin germline mutations define an inherited cancer syndrome dominated by diffuse gastric cancer, *Hum. Mutat.* **14**, 249–255.

Harman, F.S., Nicol, C.J., Marin, H.E., Ward, J.M., Gonzalez, F.J., Peters, J.M. (2004) Peroxisome proliferator-activated receptor-delta attenuates colon carcinogenesis, *Nat. Med.* **10**, 481–483.

Hayashi, K., Goodison, S., Urquidi, V., Tarin, D., Lotan, R., Tahara, E. (2003) Differential effects of retinoic acid on the growth of isogenic metastatic and non-metastatic breast cancer cell lines and their association with distinct expression of retinoic acid receptor β isoforms 2 and 4, *Int. J. Oncol.* **22**, 62–629.

Hayashi, K., Yokozaki, H., Naka, K., Yasui, W., Lotan, R., Tahara, E. (2001) Overexpression of retinoic acid receptor β induces growth arrest and apoptosis in oral cancer cell lines, *Jpn. J. Cancer Res.* **92**, 42–50.

Hayashi, K., Yokozaki, H., Goodison, S., Oue, N., Suzuki, T., Lotan, R., Yasui, W., Tahara, E. (2001) Inactivation of retinoic acid receptor β by promoter CpG hypermethylation in gastric cancer, *Differentiation* **68**, 13–21.

He, T.C., Chan, T.A., Vogelstein, B., Kinzler, K.W. (1999) PPARδ is an APC-regulated target nonsteroidal anti-inflammatory drugs, *Cell* **99**, 335–345.

Higashi, H., Tsutsumi, R., Fujita, A., Yamazaki, S., Asaka, M., Azuma, T., Hatakeyama, M. (2002) Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylations site, *PNAS* **99**, 14428–14433.

Hollstein, M., Sidransky, D., Vogelstein, B., Harris, C.C. (1991) *p53* mutations in human cancers, *Science* **253**, 49–53.

Hollstein, M.C., Metcalf, R.A., Welsh, J.A., Montesano, R., Harris, C.C. (1990) Frequent mutation of the *p53* gene in human esophageal cancer, *Proc. Natl. Acad. Sci. USA* **87**, 9958–9961.

Huang, J.Q., Sridhar, S., Chen, Y., Hunt, R.H. (1998) Meta-analysis of the relationship between *Helicobacter pylori* seropositivity and gastric cancer, *Gastroenterology* **114**, 1169–1179.

Imoto, I., Yuki, Y., Sonoda, I., Ito, T., Shimada, Y., Imamura, M., Inazawa, J. (2003) Identification of *ZASC1* encoding a Krüppel-like zinc finger protein as a novel target for 3q26 amplification in esophageal squamous cell carcinomas, *Cancer Res.* **63**, 5691–5696.

Imoto, I., Yang, Z.Q., Pimkhaokham, A., Tsuda, H., Shimada, Y., Imamura, M., Ohki, M., Inazawa, J. (2001) Identification of *cIAP1* as a candidate target gene within an amplicon at 11q22 in esophageal squamous cell carcinomas, *Cancer Res.* **61**, 6629–6634.

Ionov, Y.M., Peinado, A., Malkhosyan, S., Shibata, D., Perucho, M. (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis, *Nature* **363**, 558–561.

Irby, R.B., Mao, W., Coppola, D., Kang, J., Loubeau, J.M., Trudeau, W., Karl, R., Fujita, D.J., Jove, R., Yeatman, T.J. (1999) Activating SRC mutation in a subset of advanced human colon cancers, *Nat. Genet.* **21**, 187–190.

Ito, Y. (2004) Oncogenic potential of the RUNX gene family: 'overview', *Oncogene* **23**, 4198–4208.

Jiang, W., Zhang, Y.J., Kahn, S.M., Hollstein, M.C., Santella, R.M., Lu, S.H., Harris, C.C., Montesano, R., Weinstein, I.B. (1993) Altered expression of the cyclin D1 and retinoblastoma genes in human esophageal cancer, *Proc. Natl. Acad. Sci. USA* **90**, 9026–9030.

Kaplan, K.B., Burds, A.A., Swedlow, J.R., Bekir, S.S., Sorger, P.K., Nathke, I.S. (2001) A role for the adenomatous polyposis coli protein chromosome segregation, *Nat. Cell. Biol.*. **3**, 429–432.

Karin, M., Cao, Y., Greten, F.R., Li, Z.W. (2002) NF-κB in cancer: from innocent bystander to major culprit, *Nat. Rev. Cancer* **2**, 301–310.

Kato, J.Y., Matsushime, H., Heibert, S.W., Ewen, M.E., Sherr, C.J. (1993) Direct binding of cyclin D to the retinoblastoma gene product (pRb) and pRb phosphorylation by the cyclin D-dependent kinase, CDK4, *Gene Dev.* **7**, 331–342.

Kim, T.Y., Lee, H.J., Hwang, K.S., Lee, M., Kim, J.W., Bang, Y.J., Kang, G.H. (2004) Methylation of RUNX3 in various types of

human cancers and premalignant stages of gastric carcinoma, *Lab. Invest.* **84**, 479–484.

Kinzler, K.W., Vogelstein, B. (1996) Lessons from hereditary colorectal cancer, *Cell* **87**, 159–170.

Kitahara, K., Yasui, W., Yokozaki, H., Semba, S., Hamamoto, T., Hisatsugu, T., Tahara, E. (1996) Expression of cyclin D1, CDK4 and $p27^{KIP1}$ is associated with the $p16^{MTS1}$ gene status in human esophageal carcinoma cell lines, *J. Exp. Ther. Oncol.* **1**, 7–12.

Kuniyasu, H., Yoshida, K., Yokozaki, H., Yasui, W., Ito, H., Toge, T., Ciardiello, F., Persico, M.G., Saeki, T., Salomon, D.S., Tahara, E. (1991) Expression of *cripto*, a novel gene of the epidermal growth factor family, in human gastrointestinal carcinomas, *Jpn. J. Cancer Res.* **82**, 969–973.

Levine, A.J. (1997) p53, the cellular gatekeeper for growth and division, *Cell* **88**, 323–331.

Li, M., Song, S., Lippman, S.M., Zhang, X.K., Liu, X., Lotan, R., Xu, X.C. (2002) Induction of retinoic acid receptor-beta suppresses cyclooxygenase-2 expression in esophageal cancer cells, *Oncogene* **21**, 411–418.

Li, Q.L., Ito, K., Sakakura, C., Fukamachi, H., Inoue, K., Chi, X.Z., Lee, K.Y., Nomura, S., Lee, C.W., Han, S.B., Kim, H.M., Kim, W.J., Yamamoto, H., Yamashita, N., Yano, T., Ikeda, T., Itohara, S., Inazawa, J., Abe, T., Hagiwara, A., Yamagishi, H., Ooe, A., Kaneda, A., Sugimura, T., Ushijima, T., Bae, S.C., Ito, Y. (2002) Causal relationship between the loss of RUNX3 expression and gastric cancer, *Cell* **109**, 113–124.

Liu, C., Kato, Y., Zhang, Z., Do, V.M., Yankner, B.A., He, X. (1999) β-Trcp couples β-catenin phosphorylation-degradation and regulates *Xenopus* axis formation, *Proc. Natl. Acad. Sci. USA* **96**, 6273–6278.

Lotan, R., Xu, X.C., Lippman, S.M., Ro, J.Y., Lee, J.S., Lee, J.J., Hong, W.K. (1995) Suppression of retinoic acid receptor-beta in premalignant oral lesions and its up-regulation by isotretinoin, *N. Engl. J. Med.* **332**, 1405–1410.

Lynch, H.T., Grady, W., Lynch, J.F., Tsuchiya, K.D., Wiesner, G., Markowitz, S.D. (2000) E-cadherin mutations-based genetic counseling and hereditary diffuse gastric carcinoma, *Cancer Genet. Cytogenet.* **122**, 1–6.

Lynch, H.T., Smyrk, T., McGinn, T., Lanspa, S., Cavalieri, J., Lynch, J., Slominski-Castor, S., Cayouette, M.C., Priluck, I., Luce, M.C. (1995) Attenuated familial adenomatous polyposis (AFAP). A phenotypically and genotypically distinctive variant of FAP, *Cancer* **76**, 2427–2433.

Magnusson, P.K.E., Enroth, H., Eriksson, I., Held, M., Nyren, O., Engstrand, L., Hansson, L.E., Gyllenstem, U.B. (2001) Gastric cancer and human leukocyte antigen, *Cancer Res.* **61**, 2684–2689.

Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L., Lutterbaugh, J., Fan, R.S., Zborowska, E., Kinzler, K.W., Vogelstein, B., Brattain, M., Willson, J.K.V. (1995) Inactivation of the type II TGF-β receptor in colon cancer cells with microsatellite instability, *Science* **268**, 1336–1338.

Medico, E., Gentile, A., Celso, C.L., Williams, T.A., Gambarotta, G., Trusolino, L., Comoglio, P.M. (2001) Osteopontin is an autocrine mediator of hepatocyte growth, *Cancer Res.* **61**, 5861–5868.

Montesano, R., Hollstein, M., Hainaut, P. (1996) Genetic alterations in esophageal cancer and their relevance to etiology and pathogenesis: a review, *Int. J. Cancer* **69**, 225–235.

Mukhopadhyay, D., Tsiokas, L., Zhou, X.M., Foster, D., Brugge, J.S., Sukhatme, V.P. (1995) Hypoxic induction of human vascular endothelial growth factor expression through c-Src activation, *Nature* **375**, 577–581.

Myeroff, L.L., Parsons, R., Kim, S.J., Hedrick, L., Cho, K.R., Orth, K., Mathis, M., Kinzler, K.W., Lutterbaugh, J., Park, K., Bang, Y.J., Lee, H.Y., Park, J.G., Lynch, H.T., Roberts, A.B., Vogelstein, B., Markowitz, S.D. (1995) A transforming growth factor β receptor type II gene mutation common in colon and gastric but rare in endometrial cancers with micro-satellite instability, *Cancer Res.* **55**, 5545–5547.

Nakagawa, H., Chadwick, R.B., Peltomaki, P., Plass, C., Nakamura, Y., de la Chapelle, A. (2001) Loss of imprinting of the insulin-like growth factor II gene occurs by biallelic methylation in a core region of H19-associated CTCF-binding sites in colorectal cancer, *PNAS* **98**, 591–596.

Nakamura, T., Watanabe, M., Yamanaka, M., Hirokawa, Y., Suzuki, H., Ito, H., Yatani, R., Shiraishi, T. (2001) The role of epigenetic modifications in retinoic acid receptor $\beta 2$ gene expression in human prostate cancers, *Lab. Invest.* **81**, 1049–1057.

Nakazawa, K., Yashiro, M., Hirakawa, K. (2003) Keratinocyte growth factor produced by gastric

fibroblasts specifically stimulates proliferation of cancer cells from scirrhous gastric carcinoma, *Cancer Res.* **63**, 8848–8852.

Olschwang, S., Hamelin, R., Laurent-Puig, P., Thuille, B., Rycke, Y.D., Li, Y.J., Muzeau, F., Girodet, J., Salmon, R.J., Thomas, G. (1997) Alternative genetic pathways in colorectal carcinogenesis, *Proc. Natl. Acad. Sci. USA* **94**, 12122–12127.

Parikh, A.A., Fan, F., Liu, W.B., Ahmad, S.A., Stoeltzing, O., Reinmuth, N., Bielenberg, D., Bucana, C.D., Klagsbrun, M., Ellis, L.M. (2004) Neuropilin-1 in human colon cancer: expression, regulation, and role in induction of angiogenesis, *Am. J. Pathol.* **164**, 2139–2151.

Parkin, D.M., Pisani, P., Ferlay, J. (1999) Estimates of the worldwide incidence of 25 major cancers in 1990, *Int. J. Cancer* **80**, 827–841.

Pellman, D. (2001) Cancer. A CINtillating new job for the APC tumor suppressor, *Science* **291**, 2555–2556.

Picard, E., Seguin, C., Monhoven, N., Rochette-Egly, C., Siat, J., Borrelly, J., Martinet, Y., Martinet, N., Vignaud, J.M. (1999) Expression of retinoid receptor genes and proteins in non-small-cell lung cancer, *J. Natl. Cancer Inst.* **91**, 1059–1066.

Prinz, C., Schoniger, M., Rad, R., Becker, I., Keiditsch, E., Wagenpfeil, S., Classen, M., Rosch, T., Schepp, W., Gerhard, M. (2001) Key importance of the Helicobacter pylori adherence factor blood group antigen binding adhesion during chronic gastric inflammation, *Cancer Res.* **61**, 1903–1909.

Qiu, H., Lotan, R., Lippman, S.M., Xu, X.C. (2000) Lack of correlation between expression of retinoic acid receptor-beta and loss of heterozygosity on chromosome band 3p24 in esophageal cancer, *Genes Chromosomes Cancer* **28**, 196–202.

Qiu, H., Zhang, W., El-Naggar, A.K., Lippman, S.M., Lin, P., Lotan, R., Xu, X.C. (1999) Loss of retinoic acid receptor-beta expression is an eary event during esophageal carcinogenesis, *Am. J. Pathol.* **155**, 1519–1523.

Sano, H., Kawahito, Y., Wilder, R.L., Hashiramoto, A., Mukai, S., Asai, K., Kimura, S., Kato, H., Kondo, M., Hla, T. (1995) Expression of cyclooxygenase-1 and -2 in human colorectal cancer, *Cancer Res.* **55**, 3785–3789.

Sato, N., Maitra, A., Fukushima, N., van Heek, N.T., Matsubayashi, H., Iacobuzio-Donahue, C.A., Rosty, C., Goggins, M. (2003) Frequent hypomethylation of multiple genes overexpressed in pancreatic ductal adenocarcinoma, *Cancer Res.* **63**, 4158–4166.

Schlessinger, J. (2000) New roles for Src kinases in control of cell survival and angiogenesis, *Cell* **100**, 293–296.

Serrano, M., Hannon, G.J., Beach, D. (1993) A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4, *Nature* **366**, 704–707.

Shibata, D., Peinado, M.A., Ionov, Y., Malkhosyan, S., Perucho, M. (1994) Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation, *Nat. Genet.* **6**, 273–281.

Shibata, D., Reale, M.A., Lavin, P., Silverman, M., Fearon, E.R., Steele, G. Jr., Jessup, J.M., Loda, M., Summerhayes, I.C. (1996) The DCC protein and prognosis in colorectal cancer, *N. Engl. J. Med.* **335**, 1727–1732.

Shibata, H., Toyama, K., Shioya, H., Ito, M., Hirota, M., Hasegawa, S., Matsumoto, H., Takano, H., Akiyama, T., Toyoshima, K., Kanamaru, R., Kanegae, Y., Saito, I., Nakamura, Y., Shiba, K., Noda, T. (1997) Rapid colorectal adenoma formation initiated by conditional targeting of the *Apc* gene, *Science* **278**, 120–123.

Shimizu, N., Inada, K., Nakanishi, H., Tsukamoto, T., Ikehara, Y., Kaminishi, M., Kuramoto, S., Sugiyama, A., Katsuyama, T., Tatematsu, M. (1999) *Helicobacter pylori* infection enhances glandular stomach carcinogenesis in Mongolian gerbils treated with a chemical carcinogen, *Carcinogenesis* **20**, 669–676.

Shimizu, N., Ikehara, Y., Inada, K., Nakanishi, H., Tsukamoto, T., Nozaki, K., Kaminishi, M., Kuramoto, S., Sugiyama, A., Katsuyama, T., Tatematsu, M. (2000) Eradication diminishes enhancing effects of *Helicobacter pylori* infection on glandular stomach carcinogenesis in Mongolian gerbils, *Cancer Res.* **60**, 1512–1514.

Shinmura, K., Kohno, T., Takahashi, M., Sasaki, A., Ochiai, A., Guilford, P., Hunter, A., Reeve, A.E., Sugimura, H., Yamaguchi, N., Yokota, J. (1999) Familial gastric cancer: clinicopathological characteristics, RER phenotype and germline p53 and E-cadherin mutations, *Carcinogenesis* **20**, 1127–1131.

Shtutman, M., Zhurinsky, J., Sincha, I., Albanese, C., D'Amico, M., Pestell, R., Ben-Ze'ev, A. (1999) The cyclin D1 gene is a target

of the β-catenin/LEF-1 pathway, *Proc. Natl. Acad. Sci. USA* **96**, 5522–5527.

Sonoda, I., Imoto, I., Inoue, J., Shibata, T., Shimada, Y., Chin, K., Imamura, M., Amagasa, T., Gray, J.W., Hirohashi, S., Inazawa, J. (2004) Frequent silencing of *low density lipoprotein receptor-related protein 1B (LRP1B)* expression by genetic and epigenetic mechanisms in esophageal squamous, *Cancer Res.* **64**, 3741–3747.

Souza, R.F. (2002) Molecular and biologic basis of upper gastrointestinal malignancy–esophageal carcinoma, *Surg. Oncol. Clin. N. Am.* **11**, 257–272.

Su, L.K., Vogelstein, B., Kinzler, K.W. (1993) Association of the APC tumor suppressor protein with catenins, *Science* **262**, 1734–1737.

Tahara, E. (1990) Growth factors and oncogenes in human gastrointestinal carcinomas, *J. Cancer Res. Clin. Oncol.* **116**, 121–131.

Tahara, H., Kuniyasu, H., Yokozaki, H., Yasui, W., Shay, J.W., Ide, T., Tahara, E. (1995) Telomerase activity in preneoplastic and neoplastic gastric and colorectal lesions, *Clin. Cancer Res.* **1**, 1245–1251.

Takaguchi, M., Achanzar, W.E., Qu, W., Li, G., Waalkes, M.P. (2003) Effects of cadmium on DNA-(Cytosine-5) methyltransferase activity and DNA methylation status during cadmium-induced cellular transformation, *Exp. Cell Res.* **286**, 355–365.

Takubo, K., Nakamura, K., Izumiyama, N., Mafune, K., Tanaka, Y., Miyashita, M., Sasajima, K., Kato, M., Oshimura, M. (1997) Telomerase activity in esophageal carcinoma, *J. Surg. Oncol.* **66**, 88–92.

Tanaka, S., Mori, M., Akiyoshi, T., Tanaka, Y., Mafune, K., Wands, J.R., Sugimachi, K. (1997) Coexpression of Grb7 with epidermal growth factor receptor or Her2/erbB2 in human advanced esophageal carcinoma, *Cancer Res.* **57**, 28–31.

Termuhlem, P.M., Curley, S.A., Talamonti, M.S., Saboorian, M.H., Gallick, G.E. (1993) Site-specific differences in pp60c-src activity in human colorectal metastases, *J. Surg. Res.* **54**, 293–298.

Thibodeau, S.N., Bren, G., Schaid, D. (1993) Microsatellite instability in cancer of the proximal colon, *Science* **260**, 816–819.

Toyota, M., Ohe-Toyota, M., Ahuja, N., Issa, J.P. (2000) Distinct genetic profiles in colorectal tumors with or without the CpG island methylator phenotype, *Proc. Natl. Acad. Sci. USA* **97**, 710–715.

Tsuda, T., Tahara, E., Kajiyama, G., Sakamoto, H., Terada, M., Sugimura, T. (1989) High incidence of coamplification of *hst*-1 and *int*-2 genes in human esophageal carcinomas, *Cancer Res.* **49**, 5505–5508.

Tsushima, H., Ito, N., Tamura, S., Matsuda, Y., Inada, M., Yabuuchi, I., Imai, Y., Nagashima, R., Misawa, H., Takeda, H., Matsuzawa, Y., Kawata, S. (2001) Circulating transforming growth factor β1 as a predictor of liver metastasis after resection in colorectal cancer, *Clin. Cancer Res.* **7**, 1258–1262.

Ueda, T., Sasaki, H., Kuwahara, Y., Nezu, M., Shibuya, T., Sakamoto, H., Ishii, H., Yanagihara, K., Mafune, K., Makuuchi, M., Terada, M. (1999) Deletion of the carboxylterminal exons of K-sam/FGFR2 by short homology-mediated recombination, generating preferential expression of specific messenger RNAs, *Cancer Res.* **59**, 6080–6086.

Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Nakamura, Y., White, R., Smits, A.M., Bos, J.L. (1988) Genetic alterations during colorectal-tumor development, *N. Engl. J. Med.* **319**, 525–532.

Wan, H., Oridate, N., Lotan, D., Hong, W.K., Lotan, R. (1999) Overexpression of retinoic acid receptor β in head and neck squamous cell carcinoma cells increases their sensitivity to retinoid-induced suppression of squamous differentiation by retinoids, *Cancer Res.* **59**, 3518–3526.

Weinberg, R.A. (1998) Bumps on the road to immortality, *Nature.* **396**, 23–24.

Xu, M., Jin, Y.L., Fu, J., Huang, H., Chen, S.Z., Qu, P., Tian, H.M., Liu, Z.Y., Zhang, W. (2002) The abnormal expression of retinoic acid receptor-beta, p53 and Ki67 protein in normal, premalignant and malignant esophageal tissues, *World J. Gastroenterol.* **8**, 200–202.

Xu, X.C., Sneige, N., Liu, X., Nandagiri, R., Lee, J.J., Lukmanji, F., Hortobagyi, G., Lippman, S.M., Dhingra, K., Lotan, R. (1997) Progressive decrease in nuclear retinoic acid receptor bet messenger RNA level during breast carcinogenesis, *Cancer Res.* **57**, 4992–4996.

Yamasaki, R., Yokota, K., Okada, H., Hayashi, S., Mizuno, M., Yoshino, T., Hirai, Y., Saitou, D.,

Akagi, T., Oguma, K. (2004) Immune response in *Helicobacter pylori*-induced low-grade gastric-mucosa-associated lymphoid tissue (MALT) lymphoma, *J. Med. Microbiol.* **53**, 21–29.

Yasui, W., Yokozaki, H., Shimamoto, F., Tahara, H., Tahara, E. (1999) Molecular-pathological diagnosis of gastrointestinal tissues and its contribution to cancer histopathology, *Pathol. Int.* **49**, 763–774.

Yasui, W., Kuniyasu, H., Yokozaki, H., Semba, S., Shimamoto, F., Tahara, E. (1996) Expression of cyclin E in colorectal adenomas and adenocarcinomas: correlation with expression of Ki-67 and p53 protein, *Virchows Arch.* **429**, 13–19.

Yonemura, Y., Endo, Y., Fujita, H., Fushida, S., Ninomiya, I., Bandou, E., Taniguchi, K., Miwa, K., Ohoyama, S., Sugiyama, K., Sasaki, T. (1999) Role of vascular endothelial growth factor C expression in the development of lymph node metastasis in gastric cancer, *Clin. Cancer Res.* **5**, 1823–1829.

Yu, H.P., Wu, S.Q., Lu, W.H., Li, Y.Y., Li, F., Wang, X.L., Su, Y.H. (2004) Telomerase activity and expression of telomerase genes in squamous dysplasia and squamous cell carcinoma of the esophagus, *J. Surg. Oncol.* **86**, 99–104.

Harvesting Chemical Energy: *see* Metabolic Basis of Cellular Energy

Growth Factors

Antony W. Burgess
Ludwig Institute for Cancer Research, Australia

1 Introduction 640

2 Discovery 642

3 Growth Factor Families 645
3.1 Insulin-like Growth Factors 645
3.2 Nerve Growth Factors/Neurotrophins 647
3.3 Epidermal Growth Factor Family 647
3.4 Transforming Growth Factor (TGF-β) Family 648
3.5 Platelet-derived Growth Factors (PDGFs) and Vascular Endothelial Growth Factors (VEGFs) 649
3.6 Fibroblast Growth Factors (FGFs) 650
3.7 Hepatocyte Growth Factor (HGF) 651
3.8 Hematopoietic Growth Factors 651

4 Perspectives 654

Bibliography 654
Books and Reviews 654
Primary Literature 654

Keywords

Angiogenesis
The production of blood vessels.

Apoptosis
Programmed cell death resulting from genetic damage and/or absence of growth regulators.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30547-5

Autocrine
Describing the actions of a growth factor/cytokine on the same cell that produces the growth factor.

Cancer
The cell mass that results from the uncontrolled production of genetically altered cells in an animal.

Cell Production
The process of hierarchical proliferation and maturation of tissue-specific stem cells, which leads to the generation of mature cells in a specific tissue.

Chemokines
Structurally related smaller cytokines (8–10 kDa) that act on inflammatory cells.

Colony stimulating Factors
Proteins that stimulate the proliferation and maturation of myeloid precursor cells.

Cytokines
Proteins released from one cell that modulate the proliferation, differentiation, and/or function of cells in a specific lineage.

Differentiation
The alteration in gene expression as precursor cells divide and progress toward their functional forms.

Endocrine
Describing the effects of a growth factor/cytokine/hormone produced in one organ on the cells of another organ.

Growth Factors
Proteins that influence the proliferation and maturation of tissue-specific precursor cells.

Hormones
Molecular signals released from one cell that modulate the function or production of other cells. In the broadest sense, the term *hormone* includes steroids, peptides, and lipids. In traditional physiology, the term was used to describe the molecules released by one organ that acted on the cells of another organ.

Interleukins
Proteins released from stromal or hematopoietic cells that influence the proliferation, self-renewal, commitment, and/or maturation of hematopoietic cells.

Lymphangiogenesis
The production of lymphatic vessels.

Lymphokines
Proteins that modulate the proliferation, maturation, or function of lymphoid cells.

Paracrine
Describing the actions of a growth factor/cytokine on cells in close vicinity to the cells producing the growth factor.

Receptors
Cell surface proteins that are stimulated by exogenous ligands (e.g. growth factors) to transfer a signal to the cytoplasmic compartment of a cell.

Signal Transduction
Transfer of extracellular (membrane) interactions to intracellular (cytoplasmic and/or nuclear) responses such as enzyme activation and/or modulation of gene transcription.

Stem Cells
The cells responsible for the renewal of cells in specific lineages. Stem cells have the capability of dividing to renew themselves or dividing to produce differentiated progeny that are committed to the production of mature cells.

Cell production in all tissues of multicellular organisms is under the control of a network of tissue-specific protein regulators called *growth factors*. Although the growth factor network may be modulated by circulating hormones, these regulators are usually produced in the tissue in which they function. Growth factors have been identified in the epithelial, neural, lymphoid, myeloid, and hepatic systems; in many cases, a number of closely related growth factors are produced in each tissue. As well as controlling the normal production of mature cells, growth factors are important for regulating inflammatory, wound-healing, and antiviral responses. Many diseases appear to be associated with inappropriate growth factor production or responses. Overproduction of growth factors can lead to accumulation of cellular deposits and infiltration of tissues by inflammatory cells. Although cancers arise as the result of numerous genetic lesions, many metastatic cancers appear to be driven by the autocrine action of growth factors or mutations, which lead to activation of the receptor.

New therapeutic approaches to cancer involve the use of specific growth factors to improve hematopoietic recovery from cytotoxic or radiation therapy. Furthermore, growth factor antagonists and growth factor receptor inhibitors are being developed to suppress the proliferation and metastasis of the tumor cells. An understanding of the mechanism of action of growth factors and the range of action of these proteins is also important for developing effective therapies for autoimmune diseases, such as, rheumatoid arthritis, skin diseases, and diabetes.

1 Introduction

Multicellular organisms must control the production, maturation, and function of cells in each tissue or organ. It is now clear that locally secreted proteins modulate the cell physiology of each tissue. These regulatory proteins have been given a variety of names, including growth factors, cytokines, and interleukins (ILs). It is well known that the endocrine systems consist of a network of circulating regulatory molecules (hormones) that includes proteins, peptides, and steroids; growth factors can be considered to be an extension of the endocrine system. While there is a direct overlap between the growth factor and endocrine systems, it is often helpful to consider that the major actions of endocrine hormones occur in tissues remote from their tissue of origin, whereas growth factors appear to have a major function within the tissues that produce them.

In the main, there is no functional or structural characteristic that distinguishes different classes of growth factors – most appear to be capable of stimulating multiple biological responses that depend critically on the differentiation state of their target cells. For example, one of the hematopoietic growth factors, granulocyte colony stimulating factor (G-CSF), stimulates the proliferation of immature bone marrow cells as well as activating bacterial killing by mature neutrophils. For most growth factors, so many distinct biological effects have been observed *in vitro* that it is not yet possible to define their specific role in particular tissue systems. However, both pharmacological and genetic studies have demonstrated that growth factors can alter cell production, organogenesis, and disease susceptibility in animals. It should be noted that the action of a particular growth factor will usually depend on the presence of other growth factors, the availability of target cells, and the display of the appropriate cell surface receptors (Fig. 1).

Growth factors stimulate cells by binding to specific cell surface proteins called *receptors*. Upon binding the growth factor, many receptors dimerize and activate intracellular kinase networks. Some receptors, such as the epidermal growth factor receptor (EGFR), are ligand-dependent tyrosine kinases; others, such as the IL-2 receptor, associate with and activate particular membrane-associated tyrosine kinases. Activation of these enzyme systems can have effects on cell metabolism, membrane turnover, cytoskeletal organization, cell movement, and cell division.

Cell production occurs continuously throughout embryonic and adult life. During the earliest stages of development, cytokines such as leukemia inhibitory factor (LIF) appear to be necessary to facilitate implantation, to stimulate cell division, and to prevent differentiation (i.e. to maintain pluripotentiality). At the later stages of blastogenesis and gastrulation, cells display a bewildering array of growth factor/cytokine receptors, and the availability of particular ligands can then determine the motility, connectivity, and proliferative state of the cells. Often organs are remodeled through the absence of a tissue-specific growth factor. Late in gestation, the vitreous humor of the eye is full of hyalocytes – macrophage-like cells that depend on granulocyte/macrophage colony stimulating factor (GM-CSF) for their survival (and perhaps function). Normally, these hyalocytes persist until a few days after birth, when GM-CSF levels decrease and hyalocytes undergo rapid, apoptotic death. In mice engineered to overexpress GM-CSF, the hyalocytes persist in the vitreous humor.

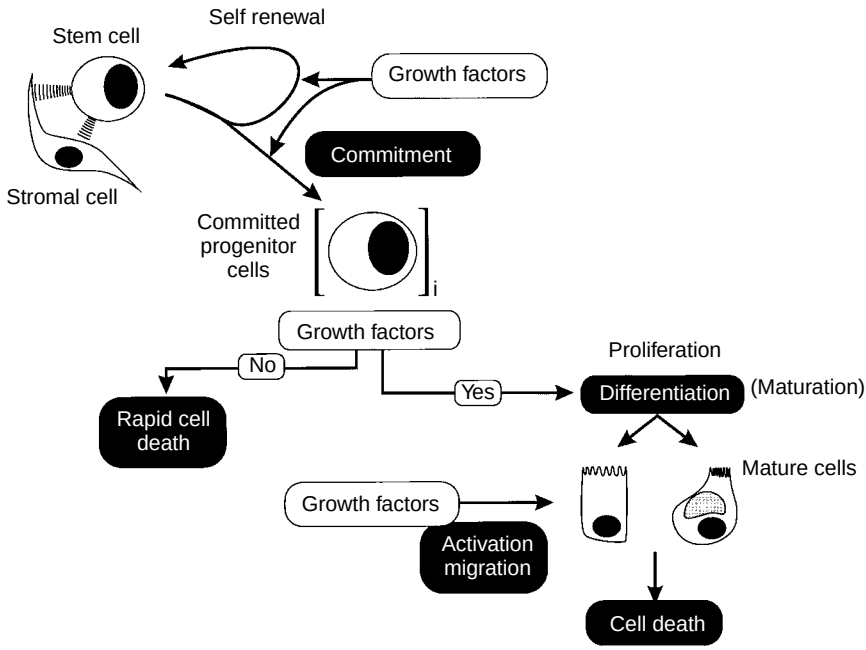

Fig. 1 There are several sites of action of growth factors during the production of cells. The action of the growth factor is likely to be quite different at the different stages of maturation.

The trafficking of mature cells between organs, and in particular from the blood into tissue spaces and even the mucosal layers, appears to be controlled by growth factors/cytokines and/or chemokines. The presence of several cytokines and membrane proteins from foreign organisms can trigger a strong proliferative response or the activation of effector functions. The action of bacterial cell wall proteins and lipids synergizes with regulators such as G-CSF or GM-CSF to induce intense antibacterial responses.

Recently, our traditional concepts of growth factor physiology have been challenged by the existence of genetically deficient animals. Often, the physiological actions of specific growth factors have been inferred from the properties of these molecules *in vitro*. For example, G-CSF stimulates the production of neutrophils in bone marrow cultures. Mice have been produced that have a genetic disruption of the G-CSF locus. Despite a profound neutrophil deficiency, these mice still produce some neutrophils. Thus, G-CSF cannot be the only growth factor controlling neutrophil production. Similarly, mice that lack the ability to produce transforming growth factor (TGF-α) have normal skin, esophagus, stomach, and colon mucosas. Although there are several growth factors closely related to TGF-α, it is not clear whether mucosal cell production continues because of the availability of other growth factors or whether the actual function of TGF is quite different from our current hypotheses.

It is not only the growth factors/cytokines that occur as families but there are also families of receptors for specific growth factor families. There are at least

four members of the EGFR family: *EGFR*, *erb B2 (HER 2)*, *erb B3 (HER 3)*, and *erb B4*. These receptors form both homodimer and heterodimer signaling complexes, all of which are capable of stimulating specific sets of intracellular signaling pathways.

Although there is a vast array of growth factors and cytokines, the receptor systems are built up from sets of domain structures. Thus, the hematopoietin receptor families have homologous hematopoietin, fibronectin, and immunoglobulin domains, which form the extracellular binding site. The ligand-binding domains are usually attached to a short intracellular tail. The nerve growth factor (NGF)/tumor necrosis factor receptor family consists of four to eight "six-cysteine" domains in the extracellular region, and these are attached to intracellular domains associated with apoptosis or CD-40-like domains. The other major class of receptors encodes ligand-activated intracellular tyrosine kinases. Again, the extracellular regions have immunoglobulin and fibronectin-like clusters, which are attached to the intracellular kinase.

Receptor activation appears to require ligand-induced aggregation, which triggers either the receptor kinase or associated intracellular kinases. Once activated, the receptor enzymes stimulate several intracellular signaling cascades. While the relative importance of the different signaling cascades has not been determined in detail, it appears that many tyrosine kinase receptors activate the *rasraf-map* kinase pathways, and many of the cytokine (hematopoietin) receptors activate the *JAK/STAT* signaling system. These signaling pathways are dependent on continuous occupancy of the receptor by ligand and even then, receptor downmodulation and phosphatase activation tend to dampen the signaling processes.

Clearly, the normal growth and differentiation of cells are controlled by the growth factor/cytokine network. However, the complex events triggered by the cytokine network are also the major targets for tumorigenic lesions (e.g. autocrine growth factor production, constitutively activated receptors, or intracellular signaling proteins forced into the active state by mutation(s)). If two or more of these lesions occur in cells that have damaged tumor suppressor genes, there is a considerable likelihood that the cell will proliferate and differentiate uncontrollably – that is, become cancerous.

2 Discovery

The incredible potency and diverse biological activities of growth factors have led to multiple, independent discoveries. A typical example is the identification of interferon 2 as a member of a family of proteins capable of inducing antiviral responses and its discovery as a growth factor in the lymphoid system, where it was called *interleukin-6* (IL-6). Some growth factors, such as fibroblast growth factor (FGF), have been identified independently in as many as 20 different biological systems. While these studies emphasize the multiple actions of growth factors, some confusion has developed with regard to their nomenclature and their likely physiological function(s). It should not be assumed that a particular action of a growth factor *in vitro* is automatically related to its role in the whole animal.

The first growth factor identified in 1906 was erythropoietin (epo). This discovery arose from physiological observations on the humoral control of production of red blood cells. While NGF and epidermal

growth factor (EGF) were discovered in the early 1950s, as a result of observations on the innervation of transplanted tumors, the discovery of most other growth factors has been due to their activity in laboratory culture systems. More recently, growth factors such as stem cell factor (also called the c-*kit* ligand) and thrombopoietin have been identified as a consequence of the characterization of their respective cell surface receptors. Table 1 lists the chromosomal locations of many growth factors for mice and humans.

The discovery of growth factors is, in many ways, a reflection of the serendipity and excitement of science. Time and again they turned up in the most unexpected way. When Stanley Cohen was analyzing the nature of the molecule responsible for nerve innovation of tumors, he planned some simple experiments to test whether it was a protein, a lipid, or a nucleic acid. One of these tests required the use of a phosphodiesterase (a hydrolytic enzyme, extracted from snake venom, which breaks down nucleic acids). Instead of destroying the growth factor, the enzyme appeared to increase its activity: in fact, the enzyme preparation was contaminated with nerve growth factor. Cohen recognized the connection between snake venom and the mouse salivary gland – a unique tissue that contains almost 1 mg g^{-1} wet weight of both NGF and EGF. Eventually, both NGF and EGF were isolated from the mouse salivary gland. Curiously, only the male mouse salivary glands have this enormous store of growth factors, and to this day it is not clear why they are there.

The human equivalent of EGF was first recognized by an activity (urogastrone) in urine, which inhibited gastric acid secretion (and acid-induced ulceration) when injected into dogs. Harold Gregory designed the purification procedure for urogastrone and after many years obtained sufficient quantities to determine its amino acid sequence. This substance was homologous to the mouse salivary gland EGF and consequently, urogastrone was renamed as *human EGF*.

The growth factors/cytokines that control the production of white blood cells were discovered at the same time, but independently, by scientists in Australia and Israel. It was several years before these groups realized they were working on the same set of molecules. The two Australians, Ray Bradley and Donald Metcalf, had been trying to grow leukemic cells; the two Israelis, Dov Pluznik and Leo Sachs, thought that they had grown tissue mast cells in the laboratory. When Bradley finally "succeeded" in devising a method to use normal cells to stimulate the growth of small colonies of leukemic cells, he rushed the cultures over to Metcalf, who realized that some of the normal cells were in fact producing the colonies. The leukemic cells were producing a "colony stimulating factor" (CSF) that allowed the proliferation and differentiation of blood cell precursors to form thousands of mature neutrophils and macrophages. Pluznik and Sachs also realized that their colonies were made up of mature white blood cells.

At first, only one CSF was recognized, but it gradually became clear that there were several distinct CSFs: macrophage CSF (called CSF-1 or M-CSF), granulocyte/macrophage (GM)-CSF, granulocyte (G)CSF, and even a molecule that stimulates cell production in almost all of the hematopoietic lineages (initially named *multi-CSF*, now known as *interleukin 3* (*IL-3*)). These CSFs were initially recognized by their activities in laboratory culture systems; indeed, many scientists believed the CSFs were artifacts and were simply

Tab. 1 Chromosomal locations of growth factors and their receptors.

Growth factor	***Chromosome location***	
	Human	***Mouse***
Insulin	11p15	
IGF-I	12	10
IGF-II	11	7
Nerve growth factor (NGF)	1p22	3
NGF receptor	17q–12–22	11
EGF	49–25–27	3
TGF-α	2p13	
EGF receptor	7p14/12	11
neu	17	11
TGF-β1	19q13.1–13.3	7
TGF-β2		1
TGF-β3	14q–23–24	12
Inhibin α	2q33	1
Inhibin β	2q33	1
Müllerian inhibitory substance	19	10
TNF-α/β	6p	17
TNF-receptor (type 1)	1p36	4 distal
TNF-receptor (type 2)	12p13	6 distal
PDGF A-chain	7	
PDGF B-chain	22	15
PDGF receptor α	4q–11–12	
PDGF receptor β	5q31–32	18
a-FGF	5	
b-FGF	4	
wnt-2	11	
*hst/ks*3		17
Hepatocyte growth factor (HGF)	7q11.2–11.21	
HGF receptor (c-*met*)	7q	6
IL–1α/β: (α-chains)	2q–13–21	2
IL-2	4q–26–28	3B–C
IL-3	5q–23–31	11A5–B1
IL-4	5q31	11A5–B1
IL-5	5q31	11A5–b1
IL-6	7P15–P21	5 proximal
IL-7	8q12–13	3
IL-9	5q22–35	13
IL-11	19q13.3–13.4	
IL-12A (p35)	3p12–3q13.2	
IL-12B (p40)	5q31–33	
VEGF-A	6p21.1	17
VEGF-B	11q13.1	19
VEGF-C	4q34.3	8
VEGF-D	Xp22.13	x

Tab. 1 (*continued*)

Growth factor	*Chromosome location*	
	Human	*Mouse*
Receptors		
IL-1 (type 1)	2q12	1 centro
IL-2α	10p14–15	2A2–A3
IL-2β	22q11.2–q12	
IL-4	16p11.2–12.1	7 distal
IL-5α		6 distal
IL-6	1	
IL-7		15 proximal
M-CSF (also called CSF-1)	1p13–p21	3F3
M-CSF receptor (c-*fms*)	5q33.2–33.3	18D
GM-CSF	5q23–31	11A5–B1
GM-CSF receptor α	αxp21-pter, Ypter-p11.2	
GM-CSF receptor β	229–12.3–13.1	
G-CSF	17q11.2–21	11D–E1
G-CSF receptor	1p34.2–35.1	4 distal
Stem cell factor	12q–22–24	10
SCF receptor (c-*kit*)	4q11–12	5
Erythropoietin	7q11–22	5G
Epo receptor	19pter-q12	9
Leukemia inhibitory factor	22q12.1–12.2	11A1–A2

nutrients missing from imperfect tissue culture media. Although our knowledge of the physiology of the CSFs is still rudimentary, genetic studies and clinical trials have shown how powerful these regulators are *in vivo*. Many cancer patients are now treated with G-CSF or GM-CSF to improve the rate of hematopoietic recovery after cytotoxic chemotherapy.

3 Growth Factor Families (see Figure 2)

3.1 Insulin-like Growth Factors

In 1922, Banting and Best discovered insulin, the regulator of glucose metabolism, in pancreatic extracts. Insulin was the first protein to have its amino acid sequence determined. It is now known that insulin is synthesized as a single precursor protein, which is processed and secreted as a disulfide-linked heterodimer. As well as controlling blood glucose levels, insulin stimulates cell proliferation; however, two other closely related proteins, insulin-like growth factors I and II (IGF-1 and IGF-II), are more potent stimulators of cell proliferation. IGF-I was discovered as the protein in plasma responsible for non-suppressible insulin-like activity. IGF-I, which was known for some time as a somatomedin, also mediates the action of growth hormone. IGF-II was the first growth factor to be associated with the aberrant growth of tumor cells in the laboratory. While the members of this family circulate in the serum, their availability is controlled by specific binding proteins (BPs). The cell surface receptors for insulin

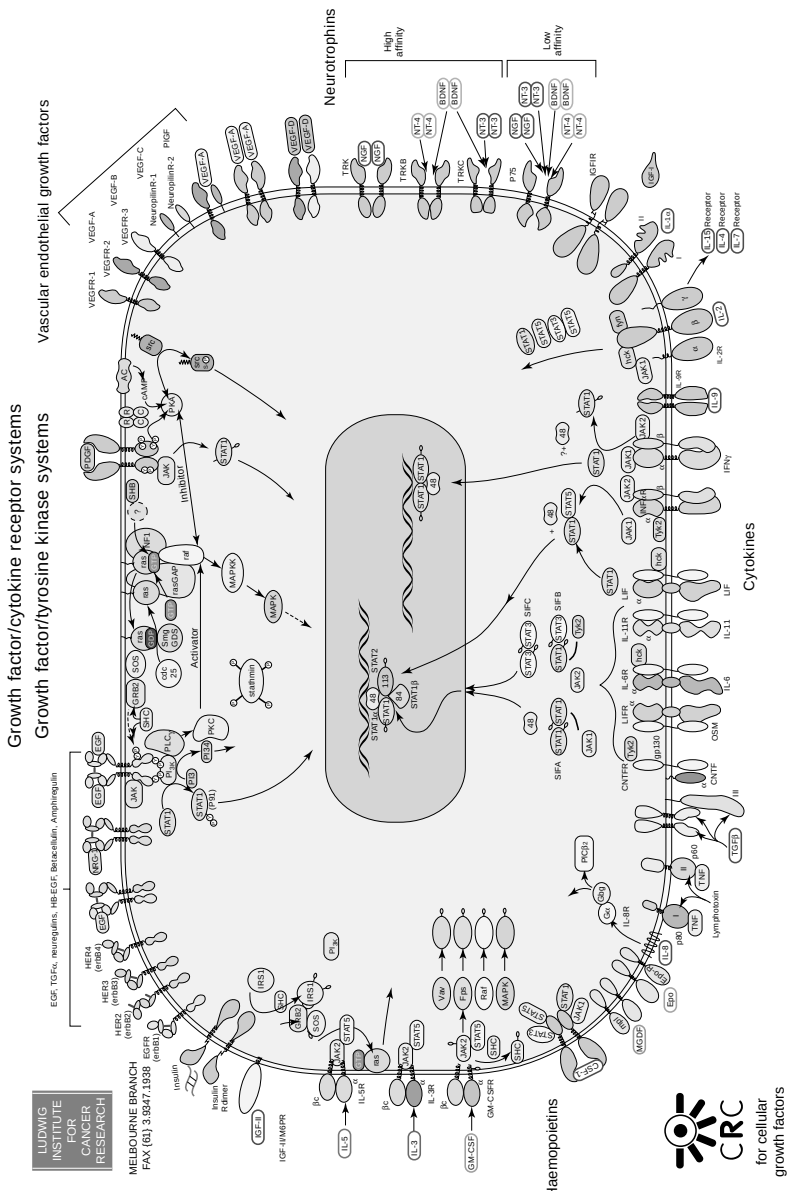

Fig. 2 Growth factors/cytokines and their receptor signaling systems.

and IGF-I are composed of four chains ($\alpha 2$, $\beta 2$). The β-chains are transmembrane, ligand-dependent tyrosine kinases, and the α-chains bind the ligand. The IGF-II receptor is a large (2000 amino acids) single-chain structure with no obvious intracellular catalytic domain. The IGF-II receptor also functions as the receptor for mannose-6-phosphate, which targets acid hydrolases to liposomes. It has been suggested that these two binding specificities might regulate complementary processes during tissue remodeling.

The actions of the IGFs are modulated by a complex set of circulating and tissue-specific binding proteins. The IGFs circulate in a complex with IGFBP-3 (this prevents degradation and excretion). Upon release from the IGFBP-3 complex, the IGFs complex with two other IGFBPs to be transported into the tissue compartments. Again, the BPs appear to control the release and action of the IGFs in their target tissues. This complex set of interactions ensures that the effective concentrations of free IGFs are kept under control and maintained for longer times than many other growth factors.

3.2 Nerve Growth Factors/Neurotrophins

Nerve growth factor, a humoral substance produced by some tumors, is responsible for innervation of these tumors. The discovery of NGF in the 1940s initiated the search for other tissue-specific growth factors and led quickly to the identification of EGF (see Sect. 3.3). Initially NGF was isolated from the salivary glands of male mice as a complex of three proteins; however, the active component is a single protein of molecular weight 30 000. The three-dimensional structure of NGF has revealed a separate, novel arrangement of three extended segments of twisted antiparallel β-sheet. Several other neurotrophins [brain-derived neurotrophic factor (BDNF), NT-3, and NT-4/5] have been discovered. BDNF stimulates peripheral sensory ganglia as well as cholinergic and dopaminergic neurons. More recently, there have been reports of a glial growth factor that is a differentially spliced form of a ligand for the *erb B3* or *erb B4* receptors.

Detailed structure–function studies have identified many of the residues required for the binding of NGF and BDNF to the low affinity (1–5 nM) neurotrophin receptor p75. However, the role of this receptor in NGF signaling is still unclear. The other class of neurotrophin receptors, *trk A, trk B,* and *trk C,* are ligand-activated tyrosine kinases. *trk A* binds NGF, *trk B* binds BDNF or NT-4, and *trk C* recognizes NT-3.

During embryological development, the neurotrophins regulate the survival of specific classes of neurones. NGF appears to function on peripheral neurones of the sympathetic system, whereas BDNF has an effect on neural crest cells and NT-3 acts on oligodendrocyte precursors.

3.3 Epidermal Growth Factor Family

It is now 40 years since Stanley Cohen described the discovery of the factor that induces in mice both premature opening of eyelids and tooth eruption. A small protein (53 amino acids) with three disulfide bonds and two distinct folding domains, EGF stimulates lung maturation and the formation of gastrointestinal mucosa, and inhibits the secretion of acid from the gastric mucosa. EGF binds to and stimulates a single-chain tyrosine kinase receptor.

There are several growth factors related to EGF – namely, TGF-α, amphiregulin, cripto, heparin binding-EGF (HBEGF), heregulin, and β-cellulin. TGF-α was discovered as an autocrine growth factor induced by some tumor viruses. EGF and TGF-α are expressed as precursor molecules. The TGF-α precursor is expressed on the surface of cells, and the mature protein is released by specific proteolysis. Amphiregulin and the neuregulins bind to distinct receptors within the EGF receptor family.

There are four closely related tyrosine kinase receptors in the EGFR family: *EGFR, HER 2, HER 3,* and *HER 4*. The sequence homology in the ligand-binding domain is between 40 and 50%, and in the kinase domain up to 80% similarity. However, it is clear that the ligand-binding specificity for each receptor is distinct, with the *EGFR* having a preference for EGF and TGF-α, whereas *HER 3* and *HER 4* bind to the neuregulins. There is now considerable evidence that heterodimers between *EGFR* and *HER 2, HER 2* and *HER 3,* and so on, bind appropriate ligands with high affinity and activate distinct signaling pathways.

Activation of the *EGFR* kinase involves ligand-induced conformational change and oligomerization. It is still not clear how many proteins are required to form the mitogenic signaling complex, but the receptor becomes autophosphorylated as well as adding phosphate to an adaptor protein called *SHC*. Phosphorylated SHC appears to influence the subcellular location of a complex between grb2 and SOS, the exchange protein that adds GTP to *ras*. Formation of the GTP–*ras* complex initiates a signaling cascade required for mitogenic responses to EGF. While the *ras*-activated pathway may be needed for mitogenesis, *EGFR* kinase must also activate other essential pathways for mitogenesis to occur.

The inappropriate expression of the *EGFR* and *HER 2* has been associated with several tumors including gliomas, where all high grade tumors overexpress the *EGFR*. *HER 2* and *HER 3* are expressed in many breast and pancreatic tumors. It also appears that β-cellulin, *HER 3,* and activated *ras* are in part responsible for many pancreatic carcinomas. By blocking autocrine growth factor production or the *EGFR,* it is possible to sensitize colonic tumors to chemotherapeutic drugs such as *cis*-platinum. The EGF receptor family are appropriate targets for anticancer signaling therapeutics.

3.4 Transforming Growth Factor (TGF-β) Family

TGF-β is released after tumor virus infection of fibroblasts. Indeed, TGF-β synergizes with TGF-β to induce large colony formation by normal rat kidney (NRK) cells. TGF-β was also isolated independently because of its ability to inhibit the proliferation of epithelial cells. TGF-β also inhibits the proliferation of both hematopoietic and lymphoid cells. TGF-β induces the differentiation of bronchial epithelial cells and prechondrocytes, but inhibits the differentiation of adipocytes and myocytes. A latent form of TGF-β is released from platelets as a complex with a specific binding protein. Active TGF-β can be released from the complex either by proteolysis or by acid treatment. Several other cytokines belong to the TGF-β superfamily: two forms of TGF-β have been isolated from platelets and three other related sequences have been detected by screening cDNA libraries. Inhibin, the Müllerian

inhibitory substances (MIS), the bone morphogenic proteins (BMPs), and activin are all related to TGF-β. The availability of the three-dimensional structure for TGF-β will allow models to be developed for the other family members.

TGF-β induces the synthesis of extracellular matrix (ECM) components *in vivo* and *in vitro*. TGF-β is also a powerful enhancer of monocyte function and a suppressor of lymphopoiesis. Mice that lack a functional *TGF-β1* gene have been produced; while they are born as apparently normal animals, within two weeks they develop severe autoimmune and hematopoietic defects.

TGF-β appears to control the production and action of ECM proteins by regulating secretion, increasing the level of integrin receptors, and decreasing the breakdown of ECM proteins. Taken together, these actions appear to have the capacity to control organogenesis during embryological development. In adults, TGF-β acts as a profound immunosuppressant. Indeed, some patients with brain tumors secrete so much TGF-β that their immune systems fail to function effectively. TGF-β may enhance the granulation phase of wound-healing, but the presence is presumably timed carefully, inasmuch as excess levels inhibit epithelialization. Interestingly, both TGF-α and TGF-β are produced by many tumors. It has been postulated that the tumors often downregulate the TGF-β receptors and are not responsive to the inhibitory actions of this growth factor, while the action of the TGF-β on surrounding cells leads to the secretion of extra ECM and thus, to a more favorable environment for tumor proliferation.

Three receptors have been identified for TGF-β: types I, II, and III. Types I and II bind TGF-β with high affinity (K_D5–25 pM). The type III TGF-β receptor is a large transmembrane proteoglycan with no obvious cytoplasmic signaling motif. The TGF-β type II receptor has ligand-dependent Ser/Thr kinase activity, which appears to signal growth inhibition. The type II receptor requires the type I receptor for ligand binding.

3.5 Platelet-derived Growth Factors (PDGFs) and Vascular Endothelial Growth Factors (VEGFs)

For more than 70 years it has been known that serum, but not plasma, stimulates cells to proliferate in culture. Not until the early 1970s, however, was it discovered that the growth factors in serum are actually released from platelets. It took almost 10 years of protein chemistry to purify and sequence the first platelet-derived growth factor (PDGF) – a disulfide-linked heterodimer of two related polypeptides (A and B), each containing approximately 100 amino acids. At the same time as the amino acid sequence for PDGF was determined, the nucleotide sequence for the oncogene (v-*sis*) encoded by the simian sarcoma virus was characterized. The protein sequence predicted from the v-*sis* gene is closely related to the PDGF-β chain.

Different cell types are known to secrete either PDGF-AA or PDGF-BB homodimers, both of which act as growth factors for fibroblasts, smooth muscle cells, and glial cells. There appear to be two forms of the PDGF receptor monomer (α and β), and these can combine to produce several, functional receptor dimers: $\alpha\alpha$, $\alpha\beta$, and $\beta\beta$. Whereas the $\alpha\alpha$ receptor responds to all forms of PDGF (i.e. AA, AB, and BB) the $\beta\beta$ receptor responds only to PDGF-BB. The PDGF receptors are typical ligand-dependent tyrosine kinases, but the cytoplasmic domain contains an

insert that modulates the intracellular signaling processes. PDGF increases the rate of wound-healing in rodents, and initial reports indicate that it may be a valuable agent for treating cutaneous ulcers in diabetic patients.

Both the PDGF-A and PDGF-B chains are synthesized in a wide variety of cells – endothelium, muscle, and macrophages. The A-chain is induced by fibroblasts in response to other growth stimuli – that is, a normal autocrine loop appears to control the proliferation of some cell types. This autocrine function can be disturbed in several diseases, such as sarcomas, atherosclerosis, and rheumatoid arthritis.

Although the PDGF family is small, there is a related set of growth factors – the vascular endothelial growth factor (VEGF) family that controls the production of endothelial cells. Several isoforms of VEGF are produced from a single gene by differential splicing of its mRNA precursor. The smallest form, $VEGF_{121}$, is a weak acidic protein that does not bind to heparin and is secreted from cells in a freely diffusible form. The longer forms of VEGF bind to heparin and appear to be released in response to plasmin.

VEGF stimulates the production of endothelial cells from both small and large blood vessels. As well as promoting angiogenesis, VEGF stimulates monocyte chemotaxis. VEGF is produced by many neoplastic cells. Interestingly, antibodies against VEGF can inhibit the production of blood vessels *in vivo*, and as a consequence, inhibit the growth of human tumor cell lines. Similarly, negative dominant VEGF receptors can inhibit the vascularization of experimental tumors.

Actually, there are four closely related growth factors in the VEGF family: VEGF, VEGF-B, VEGF-C, and VEGF-D. The first two family members stimulate the formation of blood vessels, whereas VEGF-C and VEGF-D stimulate the production of the lymphatic vessels.

3.6 Fibroblast Growth Factors (FGFs)

In the 1930s, it was discovered that brain and pituitary extracts stimulate the proliferation of fibroblasts in culture. By the mid-1980s, two forms of FGFs had been identified, purified, and sequenced. Both the acidic (a-FGF) and basic (b-FGF) forms are single-chain proteins of approximately 140 amino acids, and 55% of their amino acid sequences are identical. There are several other members of the FGF family: *wnt*-2, *hst/k*53, *FGF*-5, *FGF*-6, *FGF*-7 (keratinocyte growth factor, KGF), *FGF*-8, and *FGF*-9 (glial activating factor), and there appears to be a distant homology (20–25%) to ILs 1α and β. Both a- and b- FGFs stimulate the proliferation of cells, including colonic and breast epithelia, endothelial cells, and muscle cells. Although a-FGF and b-FGF are associated with the ECM, neither protein has a classical signal peptide that would direct their secretion. It is still not known how these proteins are released from cells.

Although a-FGF and b-FGF bind to the same set of cell surface receptors, b-FGF has the higher affinity for its receptor. While the high affinity of the FGFs for heparin-sulfate is useful for their purification, this property often interferes with binding studies to the cell surface. FGFs are potent angiogenic agents and are also capable of accelerating the healing of cutaneous wounds and damaged nerves. b-FGF is mitogenic for oligodendrocytes, astrocytes, and Schwann cells, and it acts to prolong the survival of neuronal cells in culture. In pharmacological doses, b-FGF

can enhance the remyelination of neuronal sheaths and can prevent the death of neurons in the dorsal root ganglion.

FGFs have also been purported to play a role in tumor blood vessel formation. A retrovirus encoding *FGF-4* (also called K-FGF) induces tumors in mice. Furthermore, the *FGF-4* gene is amplified in a number of breast cancers. *FGF-7* (KGF) and its receptor (*FGFR-2*) are both expressed in prostate cancer cells. A variant of *FGFR-2* (K-sam) has also been detected in stomach cancer cells and appears to be associated with the tumor phenotype.

3.7 Hepatocyte Growth Factor (HGF)

Originally discovered as a factor that increases the motility and spreading of cells (called *scatter factor*), hepatocyte growth factor (HGF) stimulates the proliferation of primary hepatocytes and increases the invasiveness of endothelial and epithelial cells. HGF appears to be involved in liver regeneration, tumor progression, and several embryological processes. HGF is a 92-kDa disulfide-linked heterodimer consisting of a light chain (33 kDa) and a heavy chain (62 kDa), which are produced from a single-chain precursor.

The HGF receptor (c-*met*) is a 190-kDa heterodimer in which the α-chain (50 kDa) appears to be extracellular and is linked to the β-chain (145 kDa), which is a membrane-spanning protein with a cytoplasmic tyrosine kinase domain, via a disulfide bond. It is expressed in a number of epithelial tissues, including liver and colon. Interestingly, a fragment of HGF containing two Kringle domains binds to c-*met* and stimulates the motility response and the receptor tyrosine kinase activity, but does not stimulate a mitogenic response in primary rat hepatocytes. Mutations in c-*met* have been associated with a number of cancer types. There are two other receptors related to c-*met*: c-*ron* and c-*sea*.

3.8 Hematopoietic Growth Factors

At least 20 cytokines or growth factors have been identified by their action on the production or function of blood cells. The major classes of hematopoietic regulators are ILs, lymphokines, CSFs, erythropoietin, and stem cell factor (SCF). Table 2 lists some of the biological actions of these molecules; however, it must be emphasized that as with the other growth factors, each one has multiple activities, which are not necessarily limited to the hematopoietic system. Several hematopoietic growth factors are being used clinically: in particular, IL-2 in conjunction with lymphokine activated killer LAK cells for cancer therapy; granulocyte/macrophage CSF and G-CSF to accelerate neutrophil recovery after chemotherapy or bone marrow transplantation; epo to stimulate red blood cell production in kidney dialysis or cancer patients; and thrombopoietin (tpo) to improve the rate of platelet recovery in thrombocytopenic patients. Three-dimensional structures have been determined for a number of hematopoietic growth factors – many consist of four helical bundles packed to form a central cylinder (Fig. 3). The structure–function relationships of this class of growth factors are being examined in detail, and already several potent antagonists (e.g. for IL-1 and IL-4) have been produced. As might be expected from such a large number of hematopoietic growth factors, several classes of cell surface receptors have been identified: tyrosine kinase (e.g.

Tab. 2 Biological actions of the hematopoietic growth factors.

Name	*Action*
IL−1α/β	Stimulates thymocyte proliferation Increases expression of IL-2 receptor on T-lymphocytes
IL-2	Stimulates proliferation of T-lymphocytes
IL-3 (also called multi-CSF)	Stimulates production of cells in all myeloid lineages
IL-4	Stimulates proliferation of B-lymphocytes Stimulates production of IgM and potentiates mast cell proliferation in response to IL-3
IL-5	Induces B-cell proliferation and differentiation in the mouse only Stimulates production and activation of eosinophils
IL-6 (also called interferon β2)	Stimulates hematopoietic progenitor cells, B-lymphoid, and myeloma cells; induces acute phase responses
IL-7	Stimulates proliferation of thymocytes, T-lymphocytes (including cytotoxic T-cells), and early B-cell precursors
IL-8	Stimulates neutrophil chemotaxis and activates bacterial killing Inhibits IgE production by B-lymphocytes
IL-9	Stimulates proliferation of T-lymphoid cell lines Potentiates proliferative effect of IL-2 on fetal thymocytes
IL-10	Inhibits cytokine production by T-lymphocytes
IL-11	Stimulates production of IgG generating B-cells Induces acute phase response Augments ability of IL-3 to stimulate megakaryocyte colonies
IL-12	Stimulates cytokine production by NK cells and T-cells
IL-13	Inhibits cytokine production by monocytes and is a costimulator of B-cell lymphocyte proliferation
IL-14	Induces proliferation of activated B-lymphocytes
IL-15	Stimulates the proliferation of cytotoxic T-lymphocytes
M-CSF (also called CSF-1)	Stimulates production, activation, and proliferation of monocytes and macrophages
GM-CSF	Stimulates production and activation of neutrophils, eosinophils, and macrophages
G-CSF	Stimulates production and activation of neutrophils
Stem cell factor	Stimulates production of all hematopoietic progenitor cells
Erythropoietin	Stimulates production of red blood cells
Leukemia inhibitory factor	Inhibits differentiation of embryonal stem cells Induces acute phase responses Stimulates megakaryocyte and platelet production Stimulates differentiation of Ml leukemic cell line
Thrombopoietin	Stimulates production of platelets

M-CSF, c-kit), multichain receptors (e.g. IL-2, IL-3, IL-6, and GM-CSF), and single-chain receptors (e.g. G-CSF). Several of the hematopoietic growth factor receptors share adaptor subunits (e.g. GM-CSF, IL-3, and IL-5 share a subunit, and these, along with the LIF oncostatin M, all appear to signal via complexes with the cell surface glycoprotein gp130).

Cytokine receptor binding has proved to be a complex and interesting process. IL-2 binds to either of two receptor chains (α or β) with relatively low affinity: K_D3 and 70 nM, respectively. This low affinity

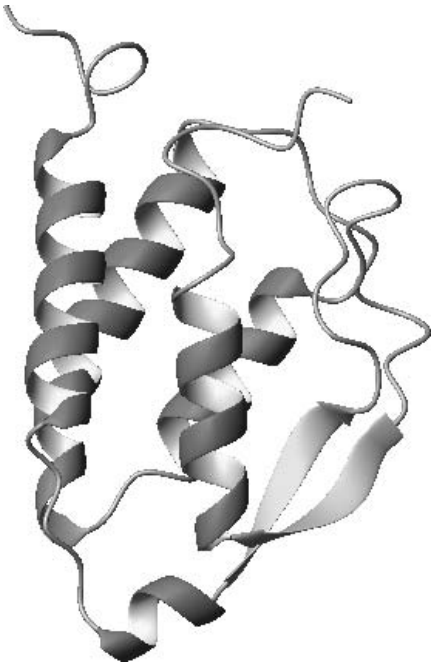

Fig. 3 The three-dimensional structure of GM-CSF illustrates the typical four-helix bundle found in many of the hematopoietic regulators.

binding does not stimulate cells. However, if both the α- and β-chains are present as a complex, the IL-2 binds with high affinity (K_D5 pM) and stimulates T-cell proliferation. The β-chain also associates with another cell surface protein called the γ-chain, which can increase the affinity of the β-chain for IL-2. The highest affinity state of the IL-2 receptor is achieved when the $\alpha\beta\gamma$ complex forms (K_D2 pM). The β- and γ-chains also complex with the IL-15 receptor to form the functional complex for IL-15 signaling, and the γ-chain complexes with the IL-4 receptor to form the high affinity IL-4 binding complex.

The IL-2 receptor binding subunits do not encode intracellular kinases; however, upon binding of IL-2, cytoplasmic tyrosine kinases in the *src* family are activated. In particular, *lck* associates with the IL-2 receptor complex, but it is still not clear whether the physical association is required for the signaling process. The γ-chain cytoplasmic region appears to be essential for activating the *src* family kinases and for stimulating the pathways leading to c-*fos* and c-*jun* expression.

Interleukin 6 binds to a specific receptor on the cell surface (IL-6Rα), but this interaction is insufficient for signaling. Once formed, the IL-6:IL-6Rα complex binds to another cell surface protein, gp130, and the trimer self-associates to form the hexameric signaling complex. gp130 also acts as the signal transduction receptor for ciliary neurotrophic factor (CNTF), LIF, and oncostatin M.

The initial signaling events stimulated by the (IL-6:IL-6Rα: gp130)$_2$ complex involve the activation of the *JAK* kinase enzymes. It is not clear whether *JAK* is associated with gp130 before the signaling complex forms, but after binding of the IL-6:IL-6Rα complex and aggregation, the *JAK* kinase is activated and phosphorylates the receptor, as well as *JAK* and *STAT* (p91). Phosphorylation of the *STAT* occurs via the *JAK* kinase after the *STAT* binds to the *JAK*-phosphorylated gp130. Once phosphorylated, *STAT* is released from the receptor and associates with a smaller protein (p48) and translocates to the nucleus, where it acts as a transcriptional regulator. Similar activation systems operate in response to many of the cytokines (e.g. growth hormone, epo, interferons, IL-3, GM-CSF, and G-CSF).

Interestingly, the cytokine/*JAK*/*STAT* signaling system is also regulated negatively. A phosphatase competes for binding to specific *JAK* phosphorylation sites on the receptor. Once bound, the phosphatase removes the phosphate from the *JAK*

kinase and deactivates the enzyme. A mutant form of the epo receptor that lacks the hematopoietic cell phosphatase-binding site has been discovered. In these people, the *JAK* kinase remains activated and associated with the epo receptor, leading to an increase in red blood cell production.

4 Perspectives

The biological action of a particular growth factor will be dependent on the differentiation state of the target cell, the presence of other growth factors, and/or the existence of cell–cell contacts (see Fig. 2). Cell–cell contacts and/or multiple growth factors acting early in many cell production pathways appear to be required to induce self-renewal of the tissue-specific stem cells. In a number of tissue systems, the presence of particular growth factors is essential to maintain cell viability at all stages of the differentiation process. If the growth factor concentration decreases, many of the immature cells will die. Mature cells usually have a definite lifetime before disintegrating; however, the functional activity of many mature cells can be activated by growth factors; for example, GM-CSF and G-CSF will prime mature neutrophils to kill bacteria more effectively.

A number of pathological responses involve excess stimulation of growth factor systems: psoraisis, inflammation, arthritis, neurodegenerative diseases, and cancer. As our understanding of these growth factor/receptor systems increases, it is becoming possible to design signaling therapeutics to treat these diseases.

See also Bioorganic Chemistry.

Bibliography

Books and Reviews

Bradshaw, R.A., Blundell, T.L., Lapatto, R., McDonald, N.Q., Murray-Rust, J. (1993) Nerve growth factor revisited, *Trends Biochem. Sci.* **18**, 48–52.

Burgess, A.W., Tran, T.T. (1994) GROCYT, available from World Wide Web site. URL: http://www.ludwig.edu.au/.

Callard, R.E., Gearing, A.J.H. (Eds.) (1994) *The Cytokine Facts Book*, Academic Press, London.

de Vos, A.M., Ultsch, M., Kossiakoff, A.A. (1992) Human growth hormone and extracellular domain of its receptor: crystal structure of the complex, *Science* **255**, 306–312.

Heath, J.K. (1993) *Growth Factors*, Oxford University Press, Oxford.

Lieschke, G.J., Burgess, A.W. (1992) Granulocyte colony-stimulating factor and granulocyte–macrophage colony-stimulating factor, *N. Engl. J. Med.* **327**, 28–35.

Meagher, A. (1990) *Cytokines*, Open University Press, Milton Keynes.

Miyazono, K., Ochijo, H., Heldin, C.-H. (1993) Enhanced bFGF expression in response to transforming growth factor-stimulation of AkR-23 cells, *Growth Factors* **8**, 11–22.

Nicola, N.A. (Ed.) (1994) *Guidebook to Cytokines and their Receptors*, Oxford University Press, Melbourne.

Sporn, M.B., Roberts, A.B. (Eds.) (1991) *Peptide Growth Factors and their Receptors*, Vols I and II, Springer-Verlag, Berlin.

Primary Literature

Akai, J., Storey, K. (2003) Brain or brawn: how FGF signaling gives us both, *Cell* **115**(5), 510–512.

Bagley, C.J., Woodcock, J.M., Guthridge, M.A., Stomski, F.C., Lopez, A.F. (2001) Structural and functional hot spots in cytokine receptors, *Int. J. Hematol.* **73**(3), 299–307.

Baldwin, M.E., Stacker, S.A., Achen, M.G. (2002) Molecular control of lymphangiogenesis, *BioEssays* **24**(11), 1030–1040.

Bunn, R.C., Fowlkes, J.L. (2003) Insulin-like growth factor binding protein proteolysis, *Trends Endocrinol. Metab.* **14**(4), 176–181.

Burgess, A.W., Cho, H.S., Eigenbrot, C., Ferguson, K.M., Garrett, T.P., Leahy, D.J., Lemmon, M.A., Sliwkowski, M.X., Ward, C.W., Yokoyama, S. (2003) An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors, *Mol. Cells* **12**(3), 541–552.

Calo, V., Migliavacca, M., Bazan, V., Macaluso, M., Buscemi, M., Gebbia, N., Russo, A. (2003) STAT proteins: from normal control of cellular events to tumorigenesis, *J. Cell Physiol.* **197**(2), 157–168.

Cao, R., Eriksson, A., Kubo, H., Alitalo, K., Cao, Y., Thyberg, J. (2004) Comparative Evaluation of FGF-2-, VEGF-A-, and VEGF-C-induced angiogenesis, lymphangiogenesis, vascular fenestrations, and permeability, *Circ. Res.*

Cheung, J.Y., Miller, B.A. (2001) Molecular mechanisms of erythropoietin signaling, *Nephron* **87**(3), 215–222.

Clemmons, D.R., Maile, L.A. (2003) Minireview: Integral membrane proteins that function coordinately with the insulin-like growth factor I receptor to regulate intracellular signaling, *Endocrinology* **144**(5), 1664–1670.

Cox, A.D., Der, C.J. (2002) Ras family signaling: therapeutic targeting, *Cancer Biol. Ther.* **1**(6), 599–606.

Dunne, A., O'Neill, L.A. (2003) The interleukin-1 receptor/Toll-like receptor superfamily: signal transduction during inflammation and host defense, *Sci. STKE* (171), re3.

Falls, D.L. (2003) Neuregulins: functions, forms, and signaling strategies, *Exp. Cell Res.* **284**(1), 14–30.

Ferrara, N., Gerber, H.P., LeCouter, J. (2003) The biology of VEGF and its receptors, *Nat. Med.* **9**(6), 669–676.

Gaffen, S.L. (2001) Signaling domains of the interleukin 2 receptor, *Cytokine* **14**(2), 63–77.

Gessner, A., Rollinghoff, M. (2000) Biologic functions and signaling of the interleukin-4 receptor complexes, *Immunobiology* **201**(3–4), 285–307.

Goldfarb, M. (2001) Signaling by fibroblast growth factors: the inside story, *Sci. STKE* **2001**(106), E37.

Harris, R.C., Chung, E., Coffey, R.J. (2003) EGF receptor ligands, *Exp. Cell Res.* **284**(1), 2–13.

Heldin, C.H., Eriksson, U., Ostman, A. (2002) New members of the platelet-derived growth factor family of mitogens, *Arch. Biochem. Biophys.* **398**(2), 284–290.

Huber, A.B., Kolodkin, A.L., Ginty, D.D., Cloutier, J.F. (2003) Signaling at the growth cone: ligand-receptor complexes and the control of axon growth and guidance, *Annu. Rev. Neurosci.* **26**, 509–563.

Itoh, H., Kataoka, H. (2002) Roles of hepatocyte growth factor activator (HGFA) and its inhibitor HAI-1 in the regeneration of injured gastrointestinal mucosa, *J. Gastroenterol.* **37**(Suppl. 14), 15–21.

Jorissen, R.N., Walker, F., Pouliot, N., Garrett, T.P., Ward, C.W., Burgess, A.W. (2003) Epidermal growth factor receptor: mechanisms of activation and signalling, *Exp. Cell Res.* **284**(1), 31–53.

Kaushansky, K. (2003) Thrombopoietin: a tool for understanding thrombopoiesis, *J. Thromb. Haemost.* **1**(7), 1587–1592.

Krebs, D.L., Hilton, D.J. (2000) SOCS: physiological suppressors of cytokine signaling, *J. Cell Sci.* **113**(Pt 16), 2813–2819.

LeRoith, D., Roberts, C.T. Jr. (2003) The insulin-like growth factor system and cancer, *Cancer Lett.* **195**(2), 127–137.

Li, X., Eriksson, U. (2003) Novel PDGF family members: PDGF-C and PDGF-D, *Cytokine Growth Factor Rev.* **14**(2), 91–98.

Liu, F. (2003) Receptor-regulated Smads in TGF-beta signaling, *Front Biosci.* **8**, S1280–S1303.

Ma, P.C., Maulik, G., Christensen, J., Salgia, R. (2003) c-Met: structure, functions and potential for therapeutic inhibition, *Cancer Metastasis Rev.* **22**(4), 309–325.

Martinez-Moczygemba, M., Huston, D.P. (2003) Biology of common beta receptor-signaling cytokines: IL-3, IL-5, and GM-CSF, *J. Allergy Clin. Immunol.* **112**(4), 653–665.

Mercer, K.E., Pritchard, C.A. (2003) Raf proteins and cancer: B-Raf is identified as a mutational target, *Biochim. Biophys. Acta* **1653**(1), 25–40.

Metcalf, D. (2003) The unsolved enigmas of leukemia inhibitory factor, *Stem Cells* **21**(1), 5–14.

Naka, T., Nishimoto, N., Kishimoto, T. (2002) The paradigm of IL-6: from basic science to medicine, *Arthritis Res.* **4**(Suppl. 3), S233–S242.

Niwa, H. (2001) Molecular mechanism to maintain stem cell renewal of ES cells, *Cell Struct. Funct.* **26**(3), 137–148.

Reddy, E.P., Korapati, A., Chaturvedi, P., Rane, S. (2000) IL-3 signaling and the role of Src

kinases, JAKs and STATs: a covert liaison unveiled, *Oncogene* **19**(21), 2532–2547.

Scott, C.L., Robb, L., Papaevangeliou, B., Mansfield, Nicola, N.A., Begley, C.G. (2000) Reassessment of interactions between hematopoietic receptors using common beta-chain and interleukin-3-specific receptor beta-chain-null cells: no evidence of functional interactions with receptors for erythropoietin, granulocyte colony-stimulating factor, or stem cell factor, *Blood* **96**(4), 1588–1590.

Simon, H.U. (2003) Neutrophil apoptosis pathways and their modifications in inflammation, *Immunol. Rev.* **193**, 101–110.

Verma, A., Kambhampati, S., Parmar, S., Platanias, L.C. (2003) Jak family of kinases in cancer, *Cancer Metastasis Rev.* **22**(4), 423–434.

Yeung, Y.G., Stanley, E.R. (2003) Proteomic approaches to the analysis of early events in colony-stimulating factor-1 signal transduction, *Mol. Cell. Proteomics* **2**(11), 1143–1155.

Heme Enzymes

Thomas L. Poulos
University of California, Irvine, CA, USA

1 Introduction 35

2 Peroxidases 36
2.1 Introduction 36
2.2 Catalytic Cycle 37
2.3 Crystal Structures 38
2.4 Active Site Structures 39
2.5 Catalytic Mechanism 39
2.6 Enzyme Substrate Complexes 41
2.7 Di-heme Peroxidases 41
2.8 Other Heme Peroxidases 44
2.9 Catalase 44

3 Cytochrome P450 45
3.1 Introduction 45
3.2 Structure and Mechanism 46
3.3 Electron Transfer 48

4 Chloroperoxidase 49
4.1 Introduction 49
4.2 Compound I Formation 50
4.3 Reactions Catalyzed by Chloroperoxidase 50

5 Nitric Oxide Synthase (NOS) 51
5.1 Introduction 51
5.2 Overall Structure 53
5.3 Active Site Structure 54

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

6 Heme Oxygenase 54
6.1 Introduction 54
6.2 Structure 55
6.3 Mechanism 56

7 Cytochrome *c* Oxidase (CCO) 57
7.1 Introduction 57
7.2 Structure 57
7.3 Electron Transfer 58
7.4 O_2 Activation 58
7.5 Proton Pumping 60

8 Cyclooxygenase 61
8.1 Introduction 61
8.2 Structure and Mechanism 62

9 Hydroxylamine Oxidoreductase (HAO) 63

Bibliography 64
Books and Reviews 64
Primary Literature 64

Keywords

Catalases
Heme enzymes that both oxidize and reduce H_2O_2.

Cytochrome c Oxidase or CCO
The terminal oxidase in mitochondrial respiration. Electrons are transferred through the mitochrondrial electron transfer system and the final acceptor is CCO. CCO receives electrons from cytochrome *c* and uses these electrons to reduce oxygen to water. The energy released by this process is used to pump protons across the mitochondrial membrane, which drives the synthesis of ATP.

Cytochrome P450
A large group of heme enzymes that use an O_2-derived O atom to hydroxylate a variety of different compounds. These enzymes provide one of the primary means for detoxification of various foreign molecules that are concentrated in the liver. Most drugs are metabolized by the P450 system.

Heme
One of the most abundant and widely used prosthetic groups in biological systems. The heme iron can serve to bind and store O_2 as in the globins, can transfer electrons

between proteins by cycling between the Fe^{2+} and Fe^{3+} redox states, and can participate in breaking the O—O bond and storing peroxide or O_2-derived oxidizing equivalents in the heme group and surrounding proteins for the purpose of carrying out redox transformations of various substrates.

Heme Oxygenase
The enzyme that initiates the oxidation and breakdown of heme.

Heterolysis and Homolysis
This refers to the way the peroxide O—O bond breaks. Homolysis generates two hydroxyl radicals, $HO^{\cdot}$. Heterolysis generates two ions, HO^{-} and HO^{+}. Most heme enzymes use heterolysis.

Nitric Oxide Synthase or NOS
Nitric Oxide Synthase catalyzes the oxidation of the common amino acid, L–arginine, to nitric oxide (NO) and L–citrulline. NO is one of the most potent signaling molecules in biology. NO exerts its effects on the cardiovascular, immune, and neuronal systems.

Peroxidases
Heme enzymes that utilize peroxides, usually H_2O_2, to oxidize other molecules. The most well-known peroxidases oxidize small aromatic alcohols.

Enzymes that contain the heme prosthetic group include a large number of proteins found in nearly all biological systems. As used in this article, heme enzymes are defined as enzymes that catalyze the oxidative chemical transformation of a substrate. A good deal is now known about structure–function relationships in heme enzymes, primarily because the heme prosthetic group exhibits easily detectable spectroscopic properties that enable various intermediates to be studied by a variety of spectral methods. In addition, there now are a large number of known crystal structures. Perhaps the most intriguing aspect of heme enzymes is how the same heme group can adopt such different functions, which clearly relates to how the protein controls the chemistry of the heme group. As a result, heme enzymes have provided particularly rich systems for detailed structure-function studies.

1 Introduction

Heme (Fig. 1) is one of the most diverse prosthetic groups in biology. Some key biological functions of heme bound to proteins include (1) reversible oxygen binding and oxygen delivery to tissues; (2) electron transfer between proteins; and (3) enzyme reactions that primarily involve peroxide- or oxygen-dependent oxidation reactions. The function the heme serves

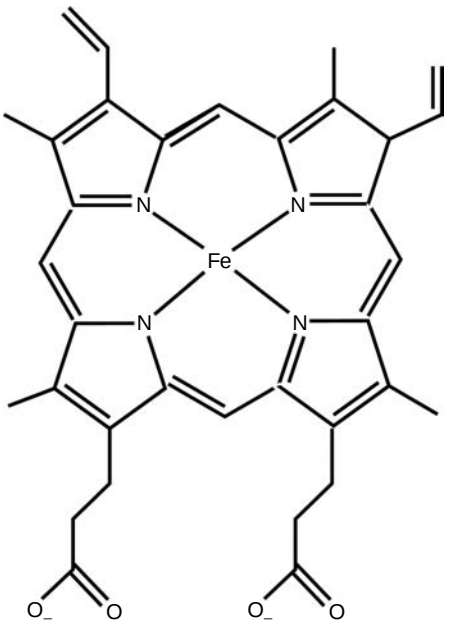

Fig. 1 The heme prosthetic group. The four-pyrrole nitrogens lie in the same plane and coordinate the heme iron. This leaves the two axial coordination positions (above and below the heme plane) available for coordinating protein side chains. In most heme enzymes, the protein contributes one axial ligand and the remaining one is either vacant or occupied by an easily dissociated solvent molecule.

depends on the surrounding polypeptide environment. The most well-known heme proteins are the globins, which both store and deliver oxygen. The heme bound to cytochromes serves primarily as an electron carrier and transfer agent where the heme iron cycles between the Fe^{2+} and Fe^{3+} oxidation states. Perhaps the most diverse use of heme is in enzyme-catalyzed reactions. Here, the heme can serve to mediate the reduction or oxidation of substrates. Although some heme enzymes can catalyze reductive reactions, the most well-known heme-containing enzymes are those that catalyze oxidation reactions and use either oxygen or peroxide as the oxidizing agent. In either case, one role of the heme is to coordinate oxygen or peroxide to the heme iron that helps to promote cleavage of the O—O bond (Fig. 2). The oxidizing power of the oxygen or peroxide ligand is thus stored at the heme site where it can be utilized to carry out specific hydroxylation or oxidation reactions. A small subset of heme enzymes utilize inorganic molecules such as hydroxylamine as oxidants. In this article, the focus is on the heme enzymes for which crystal structures have been determined.

2 Peroxidases

2.1 Introduction

Heme peroxidases occur throughout the biosphere and catalyze the oxidation of various substrates at the expense of peroxide. Peroxidases are involved in plant growth hormone metabolism, the production of thyroxine, neutrophil-mediated detoxification reactions, prostaglandin biosynthesis, and lignin degradation. While peroxidases are found in plants, animals, and microorganisms, the most thoroughly understood are nonmammalian heme peroxidases. On the basis of sequence alignments, and now crystal structures, the nonmammalian peroxidases have been divided into three classes: intercellular (Class I), extracellular fungal (Class II), and extracellular plant peroxidases (Class III). The extracellular enzymes are glycosylated, and in the case of Class III enzymes, there often are many isozymes with varying

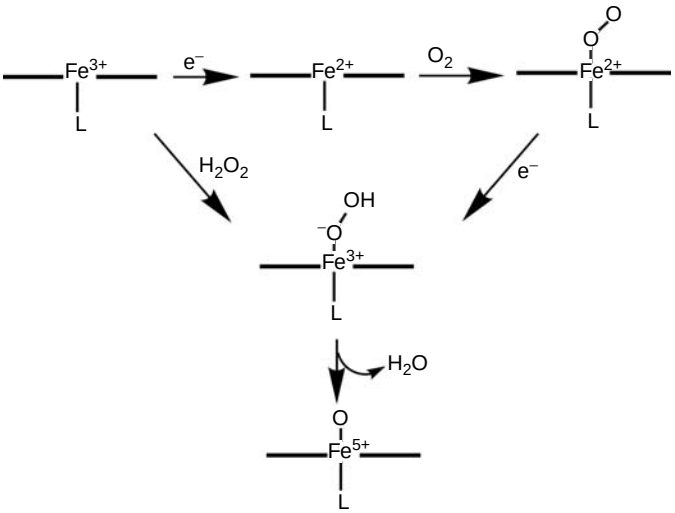

Fig. 2 A generic heme enzyme mechanism. The heme iron must first be reduced from Fe^{3+} to Fe^{2+} since molecular oxygen, O_2, binds only to the Fe^{2+} form of heme iron. Once the $Fe^{2+}-O_2$ complex is formed, a second electron transfer effectively leads to reduction of O_2 to the peroxide level. In most cases, this accompanied by protonation of the Fe-linked peroxide. At this stage, the peroxide O−O bond is cleaved, leaving behind an oxygen with only six valence electrons, a potent oxidizing agent. Exactly how the oxidizing equivalents are stored differs from one enzyme to the next, but effectively, the system behaves as if the iron is now in the Fe^{5+} oxidation state. The oxidizing power of the $Fe^{5+}-O$ center is utilized to carry out the required oxidative transformations.

degrees of glycosylation. The nonmammalian heme peroxidases are mostly single polypeptide chains in the range of 30 000–40 000 Da and contain a single ferric protoporphyrin IX as the prosthetic group. The extracellular peroxidases contain two calcium-binding sites and at least four disulfide bonds, presumably for the additional stability required of secreted enzymes.

There is a wealth of structural information available on peroxidases because several crystal structures have been solved. This, coupled with the cloning and expression of several peroxidases in recombinant systems, has opened the way for the use of protein-engineering methods to study structure–function relationships.

2.2 Catalytic Cycle

The overall peroxidase cycle occurs in three distinct steps.

$$Fe^{3+}P + H_2O_2 \longrightarrow \underset{\text{Compound I}}{Fe^{4+}{-}O\,P^{\bullet}} + H_2O \quad (1)$$

$$Fe^{4+}{-}O\,P^{\bullet} + S \longrightarrow \underset{\text{Compound II}}{Fe^{4+}{-}O\,P} + S^{\bullet} \quad (2)$$

$$Fe^{4+}{-}O\,P + S \longrightarrow Fe^{3+}P + H_2O + S \quad (3)$$

In step 1, the peroxide removes two electrons from the enzyme to give what is called Compound I. In doing so, the peroxide O—O bond is broken, giving water, and the second peroxide-derived oxygen atom remains coordinated to the iron. The iron-linked oxygen contains only six valence electrons and, hence, needs two more to give the complete complement of eight valence electrons. As a result, this oxygen is a potent oxidizing agent. One electron is removed from the iron to give the oxy–ferryl intermediate (Fe^{4+}—O), and a second electron is removed from the porphyrin (P in the above scheme) to give a porphyrin π cation radical. In some peroxidases, an amino acid side chain is oxidized to a radical rather than the porphyrin. The formation of Compound I is normally quite easy to follow since the resting enzyme is red/brown in color, while Compound I is green. In step 2, the porphyrin radical is reduced, giving a substrate radical $S^{\cdot}$. Owing to porphyrin reduction, Compound II is no longer green but red. Finally, in step 3, Compound II is reduced by a second substrate molecule. In plant peroxidases, the substrates normally are small aromatic molecules, and once substrate radicals form, they dimerize or disproportionate in nonezymatic reactions.

2.3 Crystal Structures

Several peroxidase structures are now known. All classes of peroxidases exhibit a similar three-dimensional structure, and a representative example is shown in Fig. 3. Although the sequence identity can be low (≈15%) between peroxidases, the crystal structures show a very similar fold independent of sequence homology

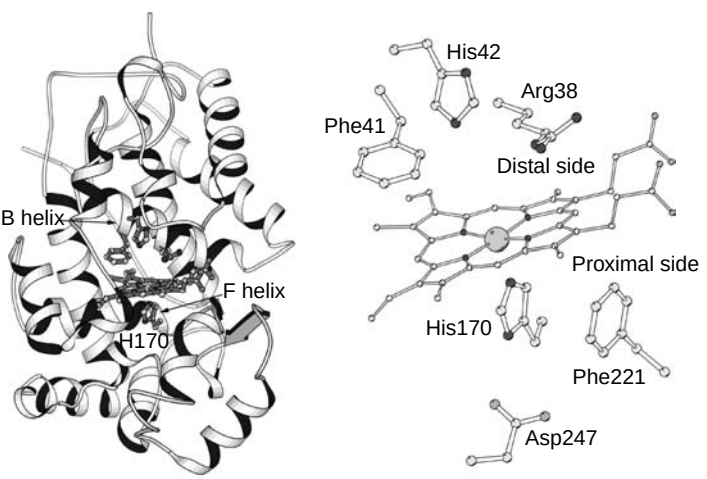

Fig. 3 The crystal structure of horseradish peroxidase or HRP taken after Gajhede et al. (1997). The heme is embedded between two helices, F and B. The F helix provides one axial heme ligand, His170. The region surrounding His170 is called the proximal pocket. The opposite side of the heme is called the distal pocket and is where the substrate, peroxide, binds. The distal helix B provides conserved side chains, His42 and Arg38, that are essential for catalysis. His42 and Arg38 are called the distal histidine and distal arginine.

or phylogenetic origin. This underscores an important "rule" in structural biology: tertiary structure is far more conserved than is primary structure. All the various peroxidases consist of a 10-helix core, two of which, the B and F helices, sandwich the heme in place. The various peroxidases differ in structure primarily on the surface. For example, Class III peroxidases have two additional helices on the surface (F′ and F″) where cytochrome *c* peroxidase (CCP) has a beta structure. It is thought that these additional helices in the Class III enzymes may play a role in substrate binding.

2.4 Active Site Structures

Not surprisingly, the active site of heme peroxidases is highly conserved (Fig. 3). The proximal pocket contains the histidine ligand coordinated to the heme iron. The histidine ligand donates an H bond to the conserved aspartic acid. This interaction is thought to impart a greater imidazolate character to the histidine ligand resulting in a partial negative charge on the histidine that can stabilize the Fe^{3+} state of the heme. This is one reason why peroxidases exhibit a lower redox potential in the Fe^{3+}/Fe^{2+} couple than the globins, even though both the globins and peroxidases use a histidine ligand. Directly adjacent to the histidine ligand, most peroxidases have an aromatic residue stacked parallel to and in contact with the histidine ligand. This residue is phenylalanine in most peroxidases but tryptophan in the two known Class I peroxidase structures, CCP and ascorbate peroxidase (APX). For the class I peroxidases, the tryptophan indole ring N atom donates an H bond to the conserved proximal aspartic acid.

On the opposite, distal side of the heme, the B helix provides key catalytic residues. The distal histidine and arginine are thought to operate in concert to help cleave the peroxide O−O bond. Stacked parallel to the distal heme surface is a conserved aromatic residue that is phenylalanine in known crystal structures except for the Class I peroxidases, which use tryptophan.

2.5 Catalytic Mechanism

Most of the effort in understanding the peroxidase mechanism has centered on the formation of Compound I. The distal histidine operates as an acid/base catalyst by removing a proton from the peroxide O atom directly linked to the iron (Fig. 4) and delivering it to the distal peroxide O atom. The net result is heterolytic fission of the peroxide O−O bond. This leaves behind an "oxene" O atom that contains only six valence electrons – a potent oxidant. The oxene O atom removes one electron from the iron to give Fe^{4+} and one from a neighboring organic group, normally the porphyrin. The crystal structure of Compound I shows that the distal arginine swings in to H-bond with the iron-linked oxygen atom. This may account in part for the stability of the Fe^{4+}−O center in Compound I.

The formation of Compound I is fast, with a second-order rate between 10^7–10^8 $M^{-1}s^{-1}$ for most heme peroxidases. Site-directed mutagenesis has been used to probe this mechanism by making amino acid substitutions of key catalytic groups in the active site. Changing the distal histidine to leucine lowers the rate of Compound I formation by 10^5, demonstrating that the distal histidine is, indeed, a critical residue. Somewhat surprisingly, the nature of the proximal ligand was

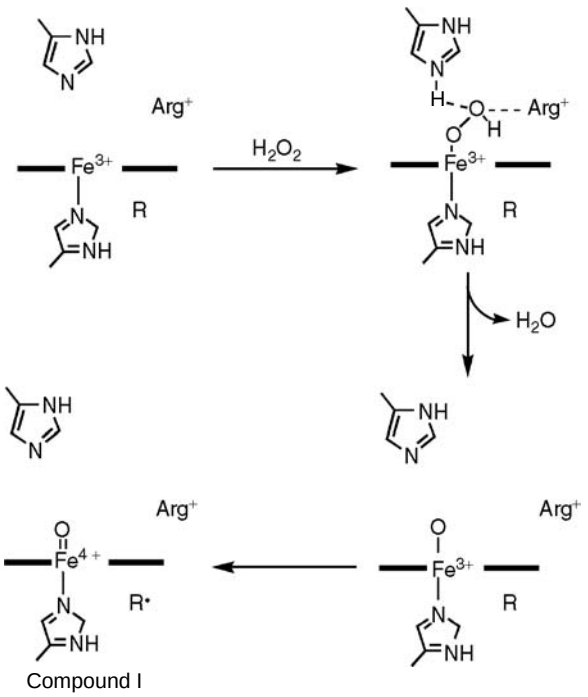

Fig. 4 The mechanism of Compound I formation. The peroxide binds to the heme iron, and one peroxide proton is transferred to the distal histidine and delivered to the leaving peroxide O atom. Precise transfer of the peroxide proton by the distal histidine ensures rapid and specific heterolytic cleavage of the O−O bond. Fe^{3+}−O is only a hypothetical intermediate and rapidly picks up two electrons, one from the iron to give Fe^{4+} and another from either the heme or a nearby amino acid side chain to give the R• radical. Reduction of R• by a substrate molecule leaves behind $Fe^{4+} = O$, which is called Compound II.

found not to be important for Compound I formation since a glutamine replacing the proximal histidine affected only the stability of Compound I and not the rate of formation.

An intriguing feature of peroxidases is how the protein controls where oxidizing equivalents are stored. In all peroxidases, one peroxide-oxidizing equivalent is stored on the iron, thus converting Fe^{3+} to Fe^{4+}. In nearly all peroxidases, the second oxidizing equivalent is stored on the heme as a porphyrin cation radical. The one exception is CCP, where Trp191 stores the second peroxide-oxidizing equivalent as a cationic radical. Trp191 occupies the equivalent position of Phe221 in horseradish peroxide (HRP) (Fig. 3). At first, it was thought that the reason CCP has a tryptophan radical is that tryptophan is easier to oxidize than the corresponding residue in most other heme peroxidases, phenylalanine. However, the structure of APX showed that APX has a tryptophan

equivalent to Trp191 in CCP, yet APX forms the traditional porphyrin cationic radical. This has lead to detailed studies using mutagenesis, crystallography, spectroscopy, and theoretical calculations to try and understand how CCP preferentially stabilizes a tryptophan radical rather than a porphyrin radical. The answer appears to be a combination of subtle structural differences resulting in an electrostatic stabilization of the extra positive charge on the tryptophan cation radical. For example, the electronegative S atoms of two mionine residues near the tryptophan radical site in CCP appear to help stabilize the tryptophan radical.

2.6 Enzyme Substrate Complexes

All the known heme peroxidase structures show that one heme edge is accessible to small aromatic molecules that are typical peroxidase substrates. It generally is thought that small substrates interact at this site and deliver electrons directly to the heme, resulting in reduction of the Compound I porphyrin radical or Compound II $Fe^{4+}-O$ center. Nevertheless, there is no direct structural evidence for this view. To date, the structure of only a limited number of heme peroxidase–substrate complexes are known. Manganese peroxidase oxidizes Mn^{2+} to Mn^{3+}, and in a complex with dicarboxylic acids, the Mn^{3+} operates as a diffusible oxidant of lignin. The crystal structure of manganese peroxidase shows that Mn^{2+} coordinates with protein carboxylates and one heme propionate. In APX, the ascorbate binds near the same heme edge very near to where Mn^{2+} binds in manganese peroxidase (Fig. 5) but is closer to the surface and does not come directly into contact with the heme propionate. Cytochrome *c* peroxidase is unusual amongst the known heme peroxidase structures since its substrate is another protein, cytochrome *c*. The crystal structure of the three substrate–peroxidase complexes are shown in Fig. 5.

2.7 Di-heme Peroxidases

Various bacteria produce peroxidases that contain two hemes and oxidize cytochrome *c* as follows:

$$2\text{cyt.c}(Fe^{2+}) + H_2O_2 + 2H^+ \longrightarrow 2\text{cyt.c}(Fe^{3+}) + 2H_2O \qquad (4)$$

The two hemes (Fig. 6) have been designated low-potential (LP) and high-potential (HP). The HP heme is hexacoordinate, with histidine and metionine residues serving as the axial heme ligands. This heme is the entry point for electrons from the reducing substrate. The most well-understood di-heme peroxidase is from *Pseudomonas aeruginosa*, and the proposed mechanism is shown below.

$$\underset{\text{high}}{Fe^{2+}}\ \underset{\text{low}}{Fe^{3+}} + H_2O_2 \longrightarrow \underset{\text{high}}{Fe^{3+}}\ \underset{\text{low}}{Fe^{4+}-O} + H_2$$

Compound I

$$\underset{\text{high}}{Fe^{3+}}\ \underset{\text{low}}{Fe^{4+}-O} + e^- \longrightarrow \underset{\text{high}}{Fe^{3+}}\ \underset{\text{low}}{Fe^{3+}} + H_2O$$

Compound II

$$\underset{\text{high}}{Fe^{3+}}\ \underset{\text{low}}{Fe^{3+}} + e^- \longrightarrow \underset{\text{high}}{Fe^{2+}}\ \underset{\text{low}}{Fe^{3+}} \qquad (5)$$

Compound II

One of the more interesting features of this mechanism is that the enzyme is inactive unless the HP heme iron is first

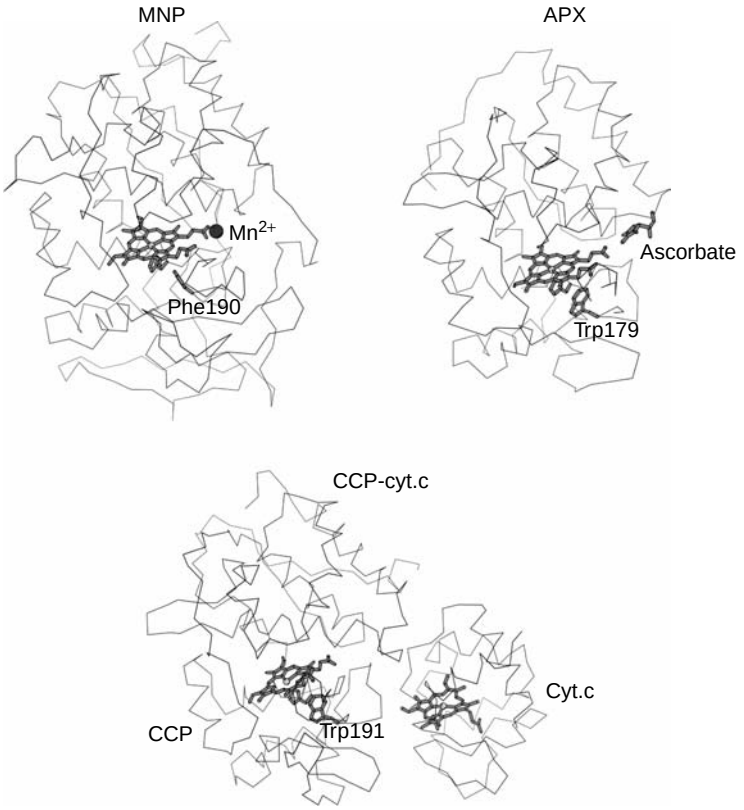

Fig. 5 Crystal structure of three peroxidase–substrate complexes. In MNP-manganese peroxidase, the substrate, Mn^{2+}, is directly coordinated to one of the heme propionate carboxyl groups. Once Compound I is formed, the electron from Mn^{2+} very likely transfers to the heme via the heme propionate. In APX, the ascorbate binds (taken from Sharp et al. (2003)) in a position very similar to Mn^{2+} in MNP except the ascorbate does not interact directly with the heme propionate. Instead, the negatively charge ascorbate interacts with surface positive charges. The structure of the noncovalent complex formed between CCP and cytochrome *c* taken from Pellitier & Kraut also is shown. The cytochrome *c* heme edge directly contacts a section of the polypeptide that contains Trp191. Trp191 forms a radical in Compound I and is known to be essential for activity. The prevailing view is that the electron from cytochrome *c* is delivered to the Trp191 radical.

reduced from Fe^{3+} to Fe^{2+}. However, a closely related di-heme peroxidase from another bacteria, *Nitrosomonas europaea*, does not require reduction of the HP heme. That this is actually a rather intriguing difference became evident once the crystal structures were solved (Fig 6).

As expected from sequence homology, the structures are very similar. Each enzyme contains approximately 300 residues folded into two distinct domains. These are designated as the HP and LP domains in Fig. 6. The most interesting difference is in the LP domain around the LP heme.

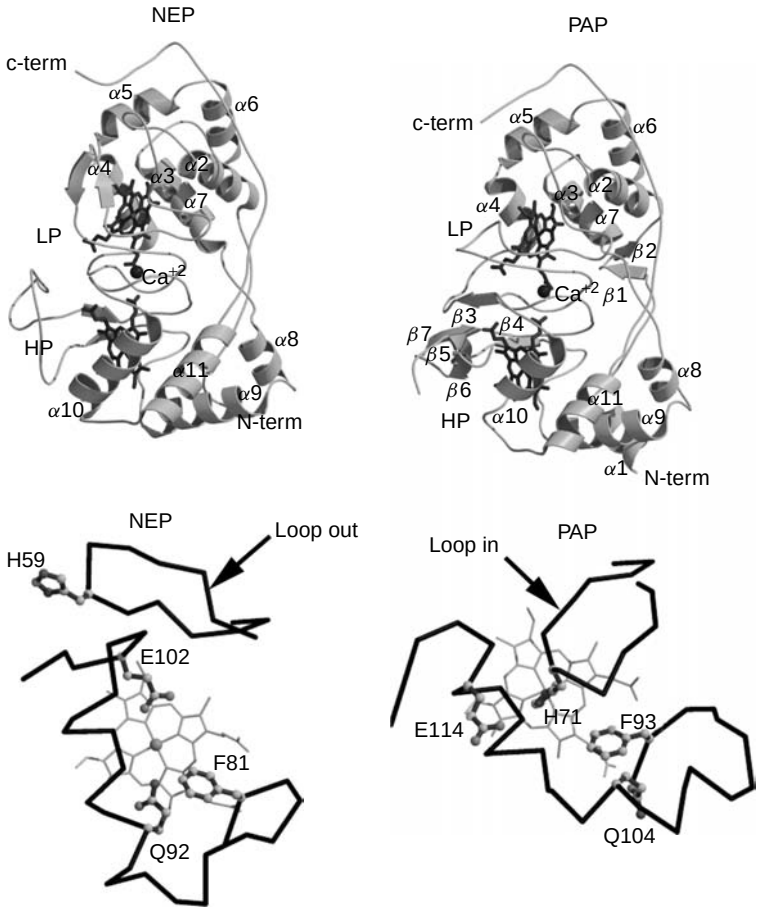

Fig. 6 Structure of two di-heme peroxidase, one from *Pseudomonas* (PAP) and one from *Nitrosomonas* (NEP). The region around the LP is shown in detail. Note than in NEP, the axial coordination position of the LP heme is available for reaction with peroxide. However, in PAP the axial coordination position is tied up with His71. Therefore, His71 must first dissociate from the heme before peroxide and react with the LP heme. This change is triggered by reduction of the HP heme by an unknown mechanism.

As shown in Fig. 6, the distal pocket where peroxide will interact with the heme iron has His71 coordinated to the heme iron in the *Pseudomonas* enzyme. However, in the *Nitrosomonas* enzyme the corresponding Histidine residue, His59, is pointing out and away from the heme pocket. Both structures are in the fully oxidized state, suggesting that in the oxidized state the *Nitrosomonas* enzyme is active, while the *Pseudomonas* enzyme is inactive. This is supported from solution biochemistry studies since the *Nitrosomonas* enzyme is fully active when both the LP and HP heme irons are in the Fe^{3+} state. However, in the *Pseudomonas* enzyme, the HP heme iron must first be reduced to the Fe^{2+} state before the LP can react with peroxides. It

therefore appears that reduction of the HP heme leads to dissociation of the His71 ligand in the LP heme, which makes available the LP heme distal pocket for reaction with peroxide. Presumably, the active forms of the LP heme in the *Pseudomonas* enzyme now look like the LP distal pocket in the *Nitrosomonas* enzyme. The remarkable aspect of this conformational change is that the introduction of a single electron into the HP heme results in the rupture of a histidine–Fe bond in the LP heme and a large rearrangement of the entire distal pocket. How this happens remains unknown.

Another interesting feature of these structures is that the peroxide-binding pocket does not contain a distal catalytic histidine as do other peroxidases. Instead, a glutamic acid residue (Glu102 in the *Nitrosomonas* and Glu114 in *Pseudomonas*, Fig. 6) is situated in a position to serve as an acid–base catalyst.

2.8 Other Heme Peroxidases

While the nonmammalian heme peroxidases have received most of the attention, the diversity of heme-containing peroxidases is very large. Mammals produce well-known peroxidases like myeloperoxidase, lactoperoxidase, and thyroid peroxidase. While these peroxidases contain heme and have active sites similar to the nonmammalian heme peroxidases, these enzymes are larger and have different three-dimensional structures. The most well understood is myeloperoxidase. The function of myeloperoxidase in white blood cells and the related eosinophil peroxidase is to oxidize halides and thiocyanate ions to produce HOCl, HOBr, and HOSCN – potent antipathogen oxidizing agents. The marine worm *Amphitrite ornata* produces a dehalogenating heme-containing peroxidase. *Mycobacterium tuberculosis*, the causative agent of tuberculosis, produces a heme catalase/peroxidase that can metabolize the most widely used antituberculoses drug, isoniazid. Strains resistant to isoniazid have a point mutation in the catalase/peroxidase, which implicates the peroxidase as a drug activator.

2.9 Catalase

Catalases catalyze the breakdown of toxic H_2O_2 to water and molecular oxygen.

$$2H_2O_2 \longrightarrow H_2O + O_2 \qquad (6)$$

Catalases are tetrameric proteins with one heme per monomer. Tyrosine rather than histidine serves as an axial heme ligand (Fig. 7). Like peroxidases, catalases contain a catalytically important distal histidine, but in catalases, this histidine lies parallel to the heme plane rather than perpendicular to it as seen in peroxidases. Because catalases both oxidize and reduce H_2O_2, it is not possible to study intermediates using H_2O_2. However, it has been possible to generate the Compound I intermediate using hydroperoxides, ROOH, since hydroperoxides cannot reduce Compound I.

The general acid–base catalytic mechanism proposed for catalase (Fig. 7) is quite similar to that proposed for peroxidases. Again, the distal histidine (His75, Fig. 7) serves to transfer a proton from the iron-linked O atom to the distal O atom. The function of Asn148 appears to be primarily in helping to orient the peroxide as well as to donate an H bond. Catalases differ from peroxidases in that a second peroxide molecule reduces Compound I, thus generating molecular oxygen and water.

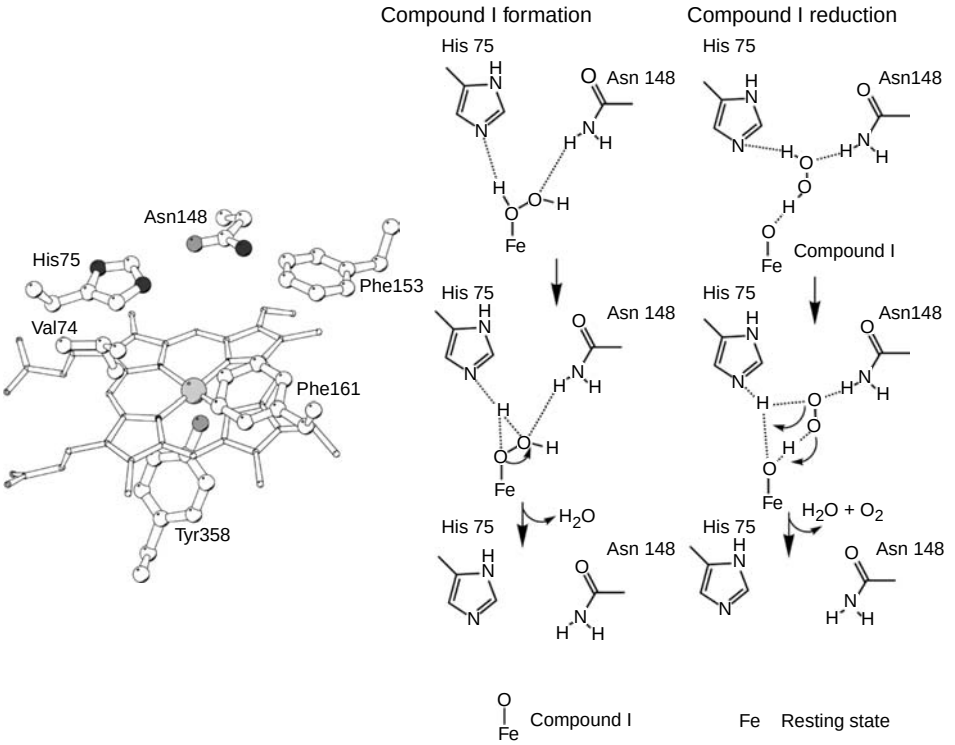

Fig. 7 The catalase active site structure and proposed mechanism. Catalase uses a tyrosine as an axial ligand rather than histidine as in peroxidases. The distal pocket contains histidine but unlike peroxidases, the histidine lies parallel to the heme, not perpendicular. Even so, the histidine can serve the same acid–base catalytic function as proposed for peroxidases. The unique feature of catalases is that peroxide is both oxidized and reduced. The first peroxide molecule is reduced to give Compound I, but the second peroxide is oxidized to give O_2.

3 Cytochrome P450

3.1 Introduction

Cytochromes P450 catalyze the hydroxylation of a vast array of molecules in the following reaction

$$\mathrm{R{-}}H + \mathrm{O_2} + \mathrm{NADPH/NADH} + 2\mathrm{H^+} \Longrightarrow \mathrm{R{-}OH} + \mathrm{H_2O} + \mathrm{NADP^+/NAD^+}. \quad (7)$$

Cytochromes P450 occur throughout the biosphere, and even prior to the current genome era, a large number of *P450* genes had been identified, making this one of the most widespread and diverse biocatalysts. Since then, the various genome projects have revealed an astonishingly large number of *P450* genes. Humans have approximately 57 expressed P450s, while *Arabadopsis* has 273 P450s. Given such a large number of P450s, it is not too surprising that P450s participate in a wide array of biological functions where hydroxylation of organic compounds are

required. For example, P450s participate in the metabolism of sex hormones, vitamin D, and bile acids in mammals; ecdysones in insects; and terpenes in plants. P450 monooxygenase systems also participate in the metabolism of a variety of drugs and other xenobiotic substances. That is, foreign substances, such as drugs, are concentrated in the liver where a variety of P450s serve to hdyroxylate these potentially toxic molecules. Hydroxylation renders such molecules more soluble for easier excretion from the body. Thus, P450s provide one of the primary means of eliminating toxic substances. P450s also participate in the conversion of aromatic hydrocarbons into potent carcinogens. This normally involves the oxidation of an aromatic heterocycle by a P450 to an epoxide. Epoxides provide potent electrophilic centers capable of chemically modifying biological nucleophiles like DNA. P450s also are the target of certain classes of commercially useful antifungal, argochemical, and therapeutic agents.

Eukaryotic P450s are membrane–bound, and liver P450s are associated with microsomal membranes. The P450 is anchored in place by an N–terminal hydrophobic peptide and a limited number of nonpolar patches on the molecular surface. Thus, P450s can be viewed as "floating" on the surface of the membrane. P450s also are found in mitochondrial membranes where they normally participate in the production of essential steroids. In sharp contrast, bacterial P450s are soluble and do not have the N–terminal membrane anchor. Because the bacterial P450s are soluble and easier to produce in bacterial recombinant expression systems, much of our detailed understanding of P450s derives from studies with bacterial enzymes. Of these, the camphor monooxygenase system (P450cam) from *Pseudomonas putida* is the most well known and best understood.

3.2 Structure and Mechanism

Figure 8 outlines the P450 catalytic cycle. The enzyme binds substrate, RH, which displaces a hydroxide ion coordinated to

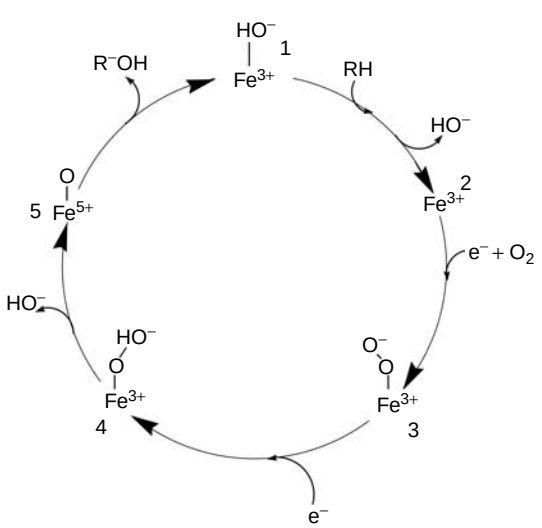

Fig. 8 The overall P450 catalytic cycle. (1) The resting enzyme; (2) substrate binds and displaces the hydroxide coordinated to the heme iron; (3) reduction of the heme iron leads to the oxy-complex; (4) protonation and reduction of the oxy-complex give the hydroperoxy complex; (5) cleavage of the peroxide O—O bond give the hypothetical Fe^{5+}—O intermediate that hydroxylates the substrate.

the heme iron. This results in an easily detected spectral shift owing to the change of the iron spin state from hexacoordinate low-spin to pentacoordinate high-spin. Reduction of the heme iron followed by oxygenation and a second electron transfer step leads to O−O bond cleavage and formation of the hypothetical Fe^{5+}−O intermediate. The Fe^{5+}−O intermediate, or its electronic equivalent, removes an H atom from the substrate followed by formation of the R−OH bond. The electron transfer reactions involve complex formation between an electron donor protein and the P450.

P450s are single polypeptide chains on the order of 40 to 50 kDa and contain a single heme (Fig. 9). Unlike the peroxidases and globins, the amino acid providing one axial iron ligand is the sulfur atom of a cysteine residue. This cysteine–heme ligation is responsible for how P450s were named since the S−Fe bond imparts unique spectral features to P450s. Early on, it was known that when a sample of liver microsomes is treated with a reducing agent in the presence of CO, the sample exhibits a strong absorption band at 450 nm, while most heme proteins maximally absorb at 420 nm. The name of the substance giving rise to this unusual spectrum was P450 (for "pigment 450") since, at the time, it was not known that the 450-nm material was a protein and not only a colored pigment. Even after the function of P450 was worked out, the name 450 was retained.

Several P450 crystal structures are now known and the overall fold is basically the same, even for the limited number of membrane-bound P450 structures that are known. Nevertheless, the various P450s exhibit greater structural variability than do the peroxidases. The main reason is that different P450s must bind and hydroxylate quite different substrates. Hence, the active site architecture changes quite substantially from one P450 to the next depending on substrate-binding requirements. In addition, for some P450s,

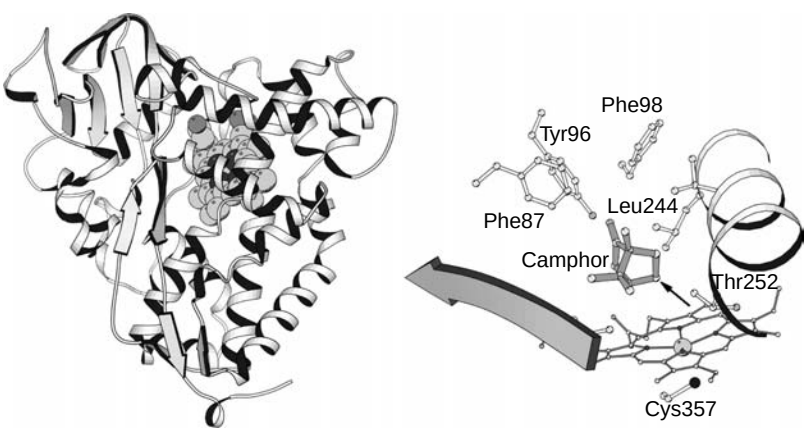

Fig. 9 The structure of the most well-known P450, P450cam from *Pseudomonas putida*. P450cam oxidizes camphor and is one step in the oxidative assimilation of camphor as a carbon source. The camphor is held in place by specific protein-substrate contacts such that the atom to be hydroxylated (indicated by the arrow) is positioned correctly for stereo- and regioselective interactions with the activated O atom linked to the iron.

there are substantial changes between the substrate-free and – bound structures presumably because the active site must open and close in order to allow substrates to enter and products to leave. Another possible reason for sealing off the active site is the need to protect the oxygen-activation chemistry at the heme site from bulk solvents. The plasticity of P450s may also be of special significance in those P450s that are required to hydroxylate a variety of molecules with seemingly little in common. From the limited knowledge now in hand, it appears that helices and loops near the entry of the active site can adapt to the size and shape of the substrate.

3.3
Electron Transfer

Most heme enzyme catalytic cycles require the input of electrons. In order to activate molecular oxygen, the heme iron must first be reduced from Fe^{3+} to Fe^{2+} since O_2 binds only to the reduced heme iron. The source of electrons is normally another protein that contains a redox active cofactor. P450s use electron donors containing either flavins or (Fe_2S_2) centers. Microsomal P450s utilize P450 reductase, which contains FMN and FAD (Fig. 10). Electrons are transferred to the FMN from NADPH *via* FAD. The path of electron flow is thus NADPH-FAD-FMN-heme. The reductase and P450 must form a complex that enables the FMN to be correctly positioned relative to the P450 heme for efficient electron transfer. There are basically two problems in understanding such reactions. First, how does the complex form? What regions of the surfaces interact and what are the dominant forces that hold the complex together? Secondly, how does the electron transfer from donor to acceptor? There is active debate on this topic. One view holds that it is sufficient to get the two redox active cofactors close enough such that

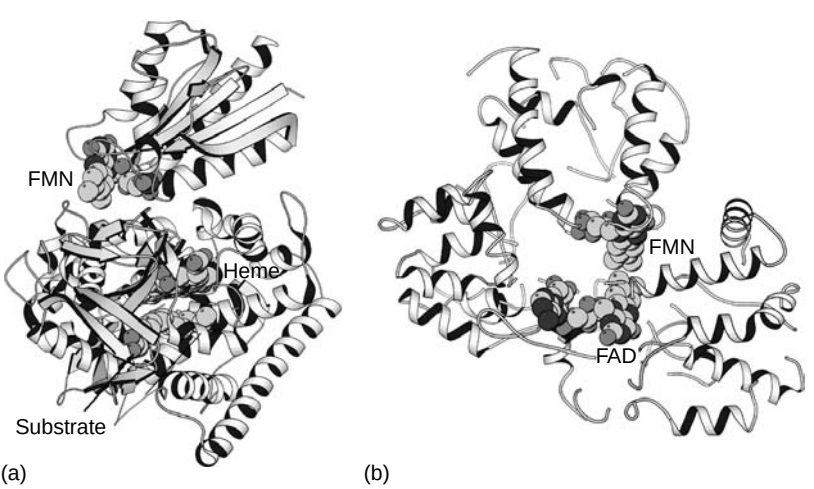

Fig. 10 The crystal structure of the FMN domain of P450BM3 complexed to the P450BM3 heme domain is shown in (a). The structure of cytochrome P450 reductase is shown in (b). The flow of electrons is NADPH-to-FAD-FMN-heme. Therefore, the FMN domain of the P450 reductase must form a complex with P450 for intramolecular electron transfer.

the electron tunnels through the protein matrix without need for a specific electron transfer path. An alternate view holds that the electron transfer proteins are "wired" with specific paths along the polypeptide backbone and, possibly, side chains.

To probe these questions in any detail requires crystal structures, and thus far, there is only one known structure of a P450 electron transfer complex (Fig. 10). The structure shown in Fig. 10 is that of P450BM3. This is a bacterial P450 that is the closest known bacterial P450 homologue to microsomal P450s. Like microsomal P450s, P450BM3 utilizes a FAD/FMN reductase. However, in P450BM3, the heme and flavin reductase are fused together into a single polypeptide chain giving the following modular architecture: heme-FMN-FAD. The structure of the complex shown in Fig 10 is that of P450BM3 with the FAD domain removed. The FMN is oriented down toward the P450, which allows the FMN to directly contact the P450. The FMN to heme distance is about 18 Å. In comparing this structure with that of P450 reductase (Fig. 10), an interesting puzzle arises. Note that in P450 reductase, the FMN and FAD cofactors are directly contacting one another. This structure is not compatible with the structure of the P450BM3 heme–FMN complex. In order for the FMN domain of the entire reductase to form the same type of complex observed in the P450BM3 structure, the FMN and FAD domains must move apart. The required motion must be large enough so that the FAD domain swings out of the way to allow the exposed edge of the FMN to dock on to the P450 surface. One possible scenario is that FAD delivers an electron to the FMN *via* direct contact between the flavins as seen in the crystal structure, but then the FAD rotates out of the way thus freeing the FMN to dock with the P450 and deliver its electron to the heme. How such molecular gymnastics are controlled and the precise mechanisms involved remain a mystery. Even so, the tools are in hand to probe these problems in molecular-level detail.

4 Chloroperoxidase

4.1 Introduction

The crystal structure of one heme halide peroxidase is known (Fig. 11). Chloroperoxdidase (CPO) is secreted by the mold *Caldariomyces fumago*, where it functions to chlorinate 1,3-cyclopentanedione to the natural product 2,2-dichloro-1,3-cyclopentanediol (caldariomycin). The catalytic cycle of CPO is very similar to other heme peroxidases. However, in CPO, the oxidizing power of Compound I is utilized to oxidize Cl^- ions to HOCl, and the HOCl then operates as an organic chlorinating agent. Traditional heme peroxidases cannot catalyze these reactions probably because the redox potential of Cl^- is too high. However, CPO operates at very low pHs ($\approx$3) where the redox potential and hence reactivity of Compound I may be significantly increased over other peroxidases that operate closer to the physiological pH.

CPO consists of a single 30 000-Da polypeptide, one heme, and one disulfide bond. The crystal structure shows at least 14 different sites of glycosylation. Unlike other heme peroxidases, the proximal ligand is cysteine and not histidine. Since CPO has the same cysteine–heme ligand environment as P450s, CPO exhibits spectral properties very similar to P450s. The most significant is the 450-nm

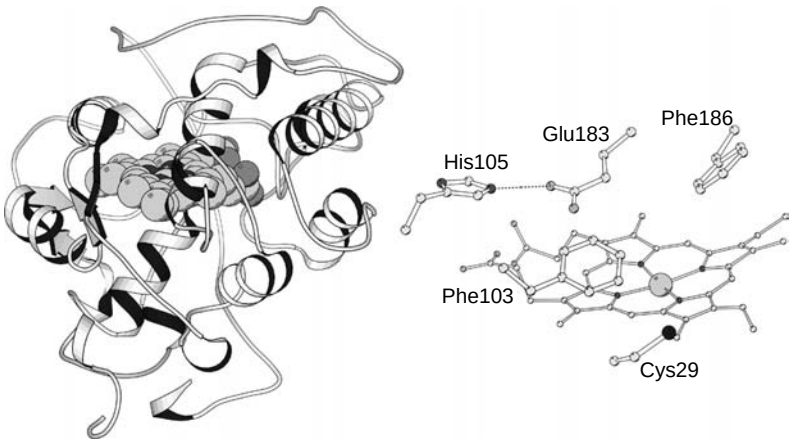

Fig. 11 The structure of chloroperoxidase (CPO). CPO combines features of P450s and peroxidase. Like P450s, the proximal axial heme ligand is a cysteine residue, and like peroxidases, the distal pocket contains polar residues that can serve a catalytic function. With CPO Glu183 serves the same function as the distal histidine in peroxidases.

band observed in the visible absorption spectrum when the enzyme is reduced in the presence of carbon monoxide to give the $Fe^{2+}-CO$ complex.

4.2 Compound I Formation

Like other heme peroxidases, CPO also reacts with peroxides to give Compound I, which contains the Fe(IV)−O center and a porphyrin π cation radical. However, unlike other heme peroxidases, CPO does not utilize a histidine residue as an acid–base catalyst. It appears instead that CPO employs Glu183 (Figs. 11 and 12) for this purpose. Glu183 is close enough to the peroxide-binding site such that the glutamic acid side chain could operate like the distal histidine in other heme peroxidases (Fig. 12). CPO thus shares properties with both P450 and peroxidases. Like P450s, CPO uses cysteine as a proximal heme ligand, but unlike P450, the distal pocket contains a potential catalytic group, Glu183. P450s instead appear to use water as the source of protons required for cleavage of the O−O bond. The more polar nature of the distal pocket in CPO due primarily to the Glu183–His104 H-bonded pair is more similar to the polar distal pocket of traditional heme peroxidases. Therefore, CPO can be said to have a P450-like proximal cysteine ligand environment while the distal peroxide-binding site is more peroxidase-like.

4.3 Reactions Catalyzed by Chloroperoxidase

Since CPO is structurally a P450-peroxidase hybrid, it is not too surprising that CPO is one of the most diverse of the known heme enzymes. In addition to operating as a chloroperoxidase, CPO exhibits activities characteristic of traditional heme peroxidase, catalase, and P450s.

Halogenation (normal function)

$$RH + H_2O_2 + Cl^- + H^+ \longrightarrow R{-}Cl + 2H_2O$$

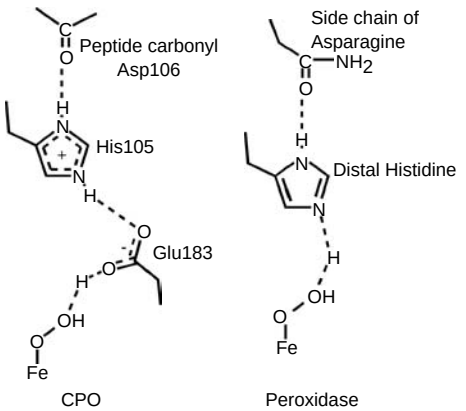

Fig. 12 The catalytic mechanism of CPO compared with traditional heme peroxidases. The glutamic acid in CPO is considered to serve the same role as the distal histidine in peroxidases.

$$\text{Dehydrogenation } 2RH + H_2O_2 \xrightarrow[\text{(heme peroxidases)}]{} R{-}R + 2H_2O$$

$$H_2O_2 \text{ decomposition } 2H_2O_2 \xrightarrow[\text{(catalase)}]{} O_2 + 2H_2O$$

$$\text{Oxygen insertion } R + H_2O_2 \xrightarrow[\text{(P450)}]{} R{-}O + H_2O \qquad (8)$$

Sulfoxidation and epoxidation reactions catalyzed by CPO can proceed with a high degree of enantioselectivity. Even the typical peroxidase dehydrogenation reactions can proceed with enantioselectivity, something other peroxidases cannot do. The CPO crystal structure provides some insights into these unusual functional properties. Although CPO is a helical protein, the overall fold is unique and does not have the typical peroxidase fold. In addition, the heme is accessible from the top distal surface and not the heme edge. Bracketing the distal heme pocket near the $Fe^{4+}{-}O$ center are two key phenylalanine residues, Phe103 and Phe186, and other hydrophobic groups that form a shallow pocket that is the most probable site for substrate binding (Fig. 11). The asymmetric nature of this pocket can account for the observed enantioselectivity.

5 Nitric Oxide Synthase (NOS)

5.1 Introduction

The biological significance of NO as a major signaling molecule became evident in the late 1980s, and since then, progress has been remarkable. NO was recognized as molecule of the year by *Science Magazine* in 1992, and the discovery of NO as the molecule responsible for smooth muscle relaxation was awarded the Nobel Prize in 1998. Initially, it was recognized that NO plays a key role in three separate physiological systems: cardiovascular, immune, and nervous. The NO for each system is produced by a separate isoform of nitric oxide synthase (NOS): eNOS (e = endothelial; cardiovascular system), iNOS (I = inducible; immune system), and nNOS (n = neuronal). The growing number of genomes shows that NOS is not confined to mammals but is also present in insects and bacteria. Hence, NO undoubtedly plays an important but as yet unknown

role in a large number of organisms and is perhaps a much "older" enzyme than initially thought. From the perspective of structural biology, that there are only three human NOS isoforms makes structure-based isoform-selective drug design a tractable problem.

In addition to its important role as a signaling molecule, NO is known to be involved in a number of pathological states including hypertension, septic shock, stroke, and degenerative neural diseases. In the immune system, nitric oxide was found to be linked to the ability of macrophages to kill tumor cells and fungi. Nitric oxide is a reactive free radical and either NO itself or further metabolized products such as peroxynitrite ($^{-}$ ONOO) are responsible for the killing ability of NO. In the cardiovascular system, the discovery of NO as a messenger molecule was based on the effect of NO as a metabolic breakdown product of nitroglycerin and other organic nitrates on vasodilation. In the nervous system, the effect of excitatory amino acids on the production of NO lead to the discovery that NO is involved in neural transmission systems. Recently, it has become apparent that NO acts as a neurotoxic agent in Alzheimer's disease.

Owing to the instability of NO and its potency as a messenger molecule, it is necessary to tightly regulate rapid NO production at the enzyme level for constitutively expressed NOSs, eNOS, and nNOS. Both enzymes bind Ca^{2+}–calmodulin (Fig. 13),

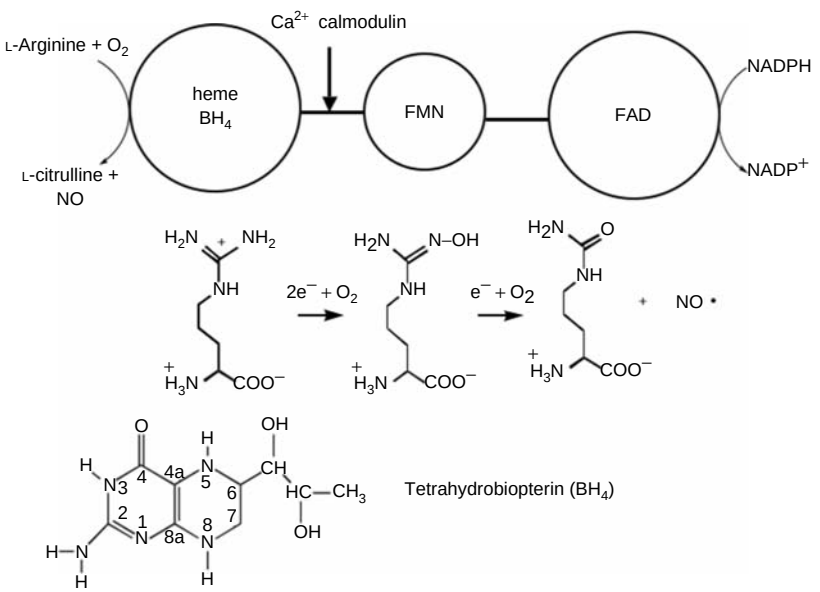

Fig. 13 NOS is a complex multidomain enzyme. Similar to P450s, electrons from NADPH-to-FAD-to-FMN-to-heme, except in NOS, the FMD/FAD reductase and heme domains are linked together in a single polypeptide chain. The linker connecting the heme and flavin domains binds the well-known regulatory molecule, calmodulin. The binding of calmodulin activates NOS by stimulating electron transfer from the flavin domain to heme by an as yet unknown mechanism. The net result is oxidation of the common amino acid, L–arginine, to citrulline and NO. A unique feature of NOS is its requirement for the cofactor, tertrahydrobiopterin (BH4).

which stimulates NOS activity and, hence, NO formation. For example, glutamate acting on the brain NMDA receptor causes the influx of Ca^{2+} that binds to calmodulin that, in turn, stimulates NO production by nNOS. In contrast, iNOS is not sensitive to Ca^{2+} even though iNOS has calmodulin tightly bound as an additional subunit. Thus, the production of NO in the immune system appears not to be controlled at the enzyme level as with eNOS and nNOS but rather in the biosynthesis of iNOS itself at the level of transcription. Why iNOS retains the ability to bind calmodulin yet remains insensitive to Ca^{2+} is unknown.

Nitric oxide exerts its effects in the vascular and neural systems by stimulating guanylate cyclase. The binding of NO to the guanylate cyclase heme iron results in cyclic cGMP production, which, in turn, results in a number of physiological effects. One such effect is neural transmission. The cardiovascular effects exhibited by NO also are mediated *via* guanylate cyclase.

5.2 Overall Structure

The modular architecture of NOS presents an especially challenging problem to structural biologists. It appears likely that NOS can adopt many conformational states depending on how the various structural modules are oriented relative to one another. The availability of various domain orientations also appears to be an essential part of NOS function. The regulatory properties of calmodulin reside in the control of electron transfer from the reductase domain to the heme domain. The most probable mechanism is a reorientation of the FMN domain such that it docks properly to the heme domain for electron transfer. While conformational heterogeneity may provide a means for controlling enzyme activity, it presents a particularly vexing problem for crystallization. As a result, crystallographers have focused on structures of the heme domain alone, which contains the active site and BH4-binding sites.

The structure of the eNOS heme domain and active site is shown in Fig. 14. The

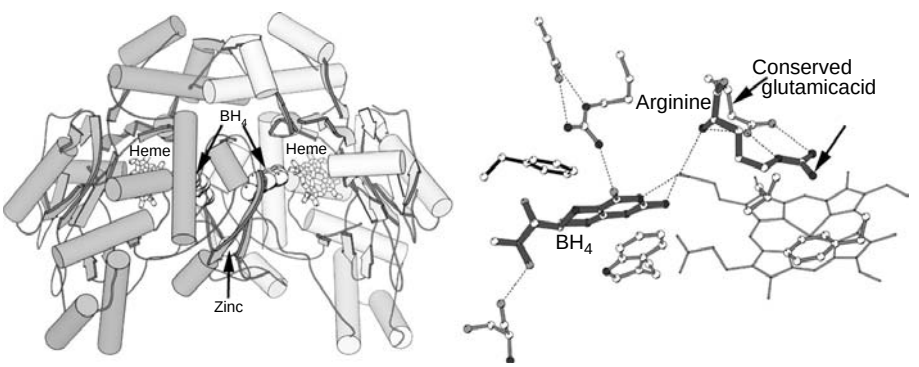

Fig. 14 Structure of the NOS heme domain. The mammalian NOS isoforms function as heme domain dimers as shown. The substrate, L–arginine, is held in place by specific H-bonding interactions. The BH4 cofactor H-bonds with the same heme propionate that H-bonds with the L–arginine substrate, which helps to understand the observed cooperativity in L–arginine and BH4 binding. It now appears that BH4 serves as a source of an electron required in activation of the iron-bound dioxygen.

structure of iNOS and nNOS also are known, but since these structures are all virtually the same with nearly identical active sites, only eNOS will be discussed in detail. NOS functions as a homodimer with the heme domain providing the dimer interface (Fig. 14). A Zn^{2+} ion lies along the dyad axis of symmetry where it is tetrahedrally coordinated to symmetrically related pairs of cysteine residues. The function of the redox inactive Zn^{2+} ion in proteins is normally structural or as a Lewis acid in hydrolytic reactions. In NOS, it appears that the main role of Zn^{2+} is to help stabilize the dimer interface. Since both BH4 molecules are bound at the dimer interface, it appears that the Zn^{2+} also helps to stabilize the BH4 site.

5.3 Active Site Structure

Also shown in Fig. 14 is the active site with the substrate, L–arginine, bound. The hydrogen-bonding environment surrounding the α-carbon position explains the selectivity for L–arginine as opposed to D–arginine. An especially important set of interactions are the hydrogen bonds between the conserved glutamic acid and L–arginine guanidinium group. This interaction helps to hold the substrate in place such that the atom to be hydroxylated sits just above the heme iron. In this regard, NOS operates in exactly the same way as P450s; that is, specific substrate-protein interactions hold the substrate in place close to the heme iron for region- and stereoselective hydroxylation. Where NOS differs most is in the strict requirement for BH4. Note (Fig. 14) that both BH4 and L–arginine hydrogen bond with the same heme propionate. This helps to explain the well-known synergy between L–arginine and BH4 binding. However, the main function of BH4 now appears to be as an electron donor to the heme iron atom during the oxygen activation reaction. With P450s, electrons derive from the flavins of P450 reductase. With NOS, however, one of the two electrons required for oxygen activation derives from BH4 giving a BH4 radical. Ultimately, the BH4 radical is re-reduced by the NOS flavins. Precisely why NOS requires this additional level of complexity in the electron transfer process compared with P450s remains unknown.

6 Heme Oxygenase

6.1 Introduction

The degradation of heme proteins results in the release of heme. Hemoglobin is the main source of heme and approximately 6 to 8 g of hemoglobin are degraded daily, thus releasing about 300 mg of heme per day. Free heme is not reutilized but instead is degraded while the iron released is stored for further use. Heme oxygenase or HO is the enzyme responsible for the breakdown of heme. HO oxidizes heme to biliverdin and CO and is the most important reaction in biological systems to generate CO (Fig. 15). Biliverdin is reduced by biliverdin reductase to bilirubin, which is then conjugated with glucuronic acid and excreted. Heme is red, biliverdin is blue–green, and bilirubin is yellow–red. This is the reason that a hematoma, or bruise, from damaged tissue changes color with time. The second product of the HO reaction, CO, also is of growing interest. To some extent, CO can mimic NO in its ability to stimulate guanyl cyclase. This, together with the cellular localization of the HO isoforms, has lead to the proposal that

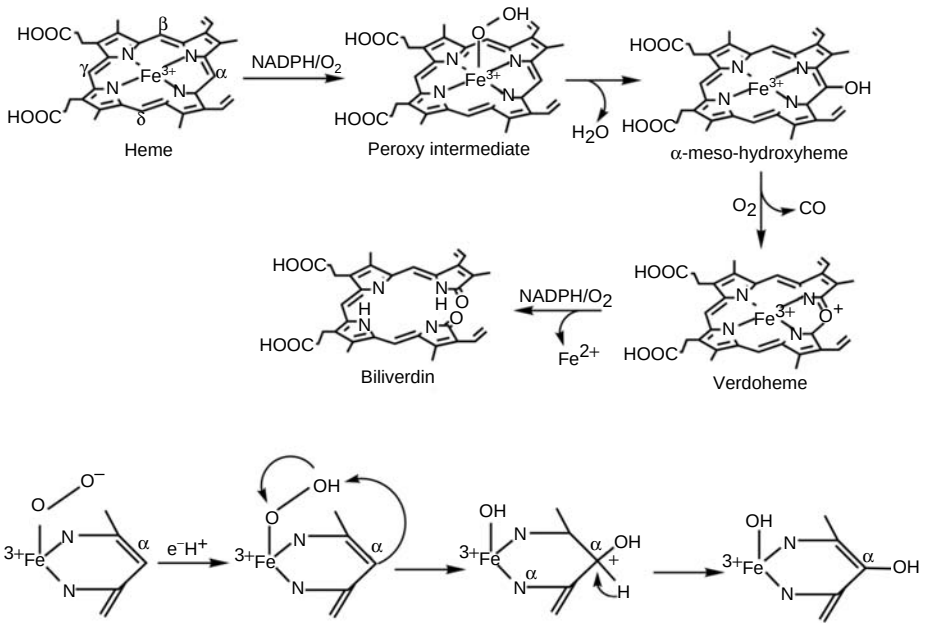

Fig. 15 The catalytic cycle of heme oxygenase. Reduction of the oxy-complex gives the peroxy-complex similar to other heme enzymes. However, unlike other heme enzymes the peroxide O—O bond is not cleaved prior to substrate oxidation. Instead, the peroxy intermediate itself serves as the hdyroxylating agent. The HO active site is structured such that only the heme *α-meso* carbon can react with the peroxy intermediate. This is followed by a series of reactions leading to the formation of biliverdin and release of CO. Details of the critical hydroxylation step are outlined.

CO may serve as an important signaling. This remains an intriguing yet controversial topic.

Humans have two HO isoforms designated HO1 and HO2. They differ primarily in tissue location and regulation. HO1 levels increase in response to oxidative stress and, like iNOS, is induced during the inflammatory response. HO1 was the first to be cloned and expressed in a bacterial recombinant expression system. Since HO1 contains a C–terminal membrane anchor segment, it first was necessary to remove the C–terminal peptide in order to generate a soluble version of HO1. Fortunately, the truncated soluble version of HO1 is fully active. The availability of large quantities of soluble recombinant HO1 has provided the basis for much of our current understanding of structure–function relationships in HO1.

6.2 Structure

At the time of this writing, the structures of three HOs have been determined: human and rat HO1 and the bacterial HO from *Neisseria meningitidis*. The structure of human HO1 is shown in Fig. 16. The enzyme is comprised of only helices and random coil and represents a novel helical fold. The heme is sandwiched between the distal and proximal helices. The proximal His25 provides one axial heme ligand. Unlike peroxidases and P450s, the heme

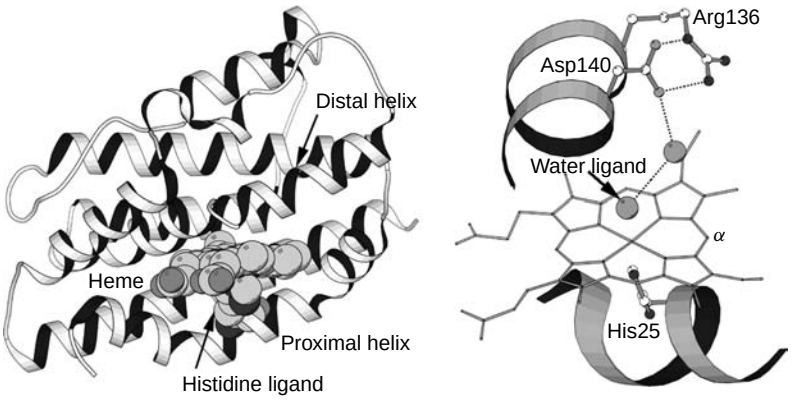

Fig. 16 The structure of human heme oxygenase-1 or HO1. The heme is relatively exposed in HO compared to other heme enzymes. This is required since heme is the substrate and hence, efficient catalysis requires both binding of substrate and ready release of product. The active site structure shows a network of H bonds involving ordered water molecules important for catalysis.

is relatively exposed in HO1. The negatively charged heme propionates are held in place by a series of positively charged amino acids.

The distal helix is of particular interest owing to several conserved glycine residues and the resulting flexibility of the heme pocket. In the crystals of human HO1 used to determine the structure, there are two molecules in the asymmetric unit. The distal helix adopts slightly different conformations in each. In one, the distal helix is further from the heme, thus giving the "open" conformer, while in the other molecule, the distal helix is situated closer to the heme, giving the "closed" conformer. This type of flexibility is required by HO since heme is a substrate; so, active site motion is required for substrate binding and product release.

6.3 Mechanism

Figure 15 provides an outline of the HO mechanism. Each of the intermediates has been identified and characterized using a variety of spectroscopic and kinetic techniques. The source of electrons required to drive the reaction are from NADPH. Reducing equivalents from NADPH are delivered to the HO heme *via* P450 reductase. Hence, like P450s, HO must form a complex with the reductase. HO differs markedly from P450 and peroxidases in that peroxide is the active hydroxylating agent. In both P450s and peroxidases, the peroxide O—O bond is cleaved, giving Fe^{3+}—O or its electronic equivalent. With these enzymes, it is the iron-linked oxygen that serves as the hydroxylating agent. As shown in Fig. 15, it is the distal peroxide O atom that forms a bond with the α-meso carbon of heme.

The first step of the reaction, hydroxylation of heme to give α-*meso*-hydroxyheme, has received the most attention. A key question is how the enzymes ensure that only the α-*meso* heme carbon is hydroxylated. The structure supports the view that the primary factor is steric. In the closed conformation, the only heme *meso*

carbon that remains available for reaction is the *α-meso* carbon. The active site provides an open pocket just above the *α-meso* carbon that is partially filled with water molecules (Fig. 16). Oxygen binding to heme prefers a bent conformation and it appears most likely that the oxygen bends toward the *α-meso* carbon, thus placing the distal O atom close enough to the *α-meso* for reaction. The bacterial HO structure also supports this view.

Another key factor is that HO uses peroxide as the hydroxylating agent rather than the Fe−O center found in other the heme enzymes. Recent mutagenesis work has shed considerable light on this issue. Changing Asp140 (Fig. 16) to other residues converts HO into a peroxidase. The mutant HO now cleaves the peroxide O−O bond, thus generating the Fe(IV)−O center, which can then oxidize traditional peroxidase aromatic phenol substrates. It appears that Asp140 and its surrounding H-bonding environment (Fig. 16) are essential for preventing cleavage of the O−O bond. One can view this as a balance between O−O bond cleavage and *α-meso* hydroxylation, with the local H-bonding environment favoring *α-meso* hydroxylation. Disruption of this network as in the Asp140 mutants tips the balance in favor of O−O bond cleavage. In addition, it appears that the Asp140-water hydrogen-bonding network helps to steer the iron-linked dioxygen toward the *α-meso* carbon. An interesting feature of the HO mechanism shared by P450s is the lack of amino acid side chains that directly interact with the iron-linked dioxygen. In both P450 and HO, water plays a key role in catalysis by serving the acid–base catalytic functions fulfilled by histidine in peroxidases.

7 Cytochrome *c* Oxidase (CCO)

7.1 Introduction

CCO is the terminal electron acceptor in eukaryotic and some bacterial respiratory systems. The enzyme catalyzes the reduction of O_2 to water by utilizing electrons from ferrocytochrome *c*. In the process, CCO pumps protons across the mitochondrial membrane thereby generating the proton motive force required for the biosynthesis of ATP. Needless to say, this is a central enzyme in bioenergetics and has been the topic of extensive investigations. CCO is a complex transmembrane enzyme consisting of several subunits. However, only subunits I and II are highly conserved in both prokaryotes and eurkaryotes and also contain the redox active centers required for the catalytic cycle. Subunit II contains the so-called Cu_A site, which is the entry point of electrons from ferrocytochrome *c*. Subunit I contains the hexacoordinate heme a, pentacoordinate heme a_3, and Cu_B. Precisely how the transfer of electrons from ferrocytochrome *c* through the various redox centers in CCO is coupled to the pumping of oc protons through CCO and across the membrane continues to be a challenging problem under active investigation.

7.2 Structure

Research on CCO took a major step forward in 1995 with the structure determination of both eukaryotic and prokaryotic CCO structures. These structures represent one of the major achievements in structural biology and have opened the way toward a molecular-level understanding of

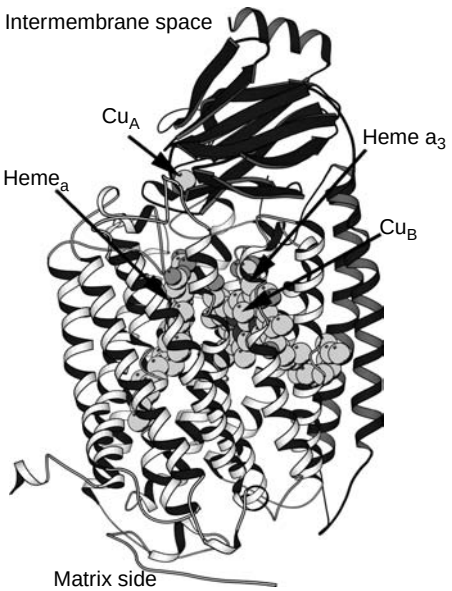

Fig. 17 The structure of mammalian cytochrome *c* oxidase showing only subunits I (light shading) and subunit II (dark shading).

one of Nature's most complex and critical biochemical processes. The structure of eukaryotic CCO is shown in Fig. 17, while Fig. 18 is a more schematic diagram illustrating the relationship between redox centers. Subunits I and II contain the redox active centers. The smaller subunit II consisting of 227 residues has a primarily β-sheet domain sitting over the top of subunit I on the surface of the membrane. The Cu_A center consists of two Cu atoms coordinated to histidine, metionine, and cysteine residues located in the subunit II β-sheet domain. Subunit II also contains two transmembrane helices that interact with helices in subunit I. The larger subunit I has 514 residues and is composed of 10 transmembrane helices. Heme a utilizes two histidine residues as axial ligands. Heme a_3 is coordinated by only one histidine. The remaining axial coordination position remains open for binding of O_2. Finally, Cu_B is coordinated by histidine residues and is situated only 4.7Å from the iron of heme a_3 (Fig. 17). A unique feature of the Cu_B site revealed by the x-ray structures is that one Cu_B histidine ligand is covalently attached to a neighboring tyrosine residue (Fig. 19).

7.3 Electron Transfer

Since electrons are delivered to CCO from ferrocytochrome *c*, it is necessary for cytochrome *c* first to form a complex with CCO. Because cytochrome *c* is a soluble protein in the intermembrane mitochondrial space, the complex must form between regions of CCO situated on the membrane surface. It is well known that electrostatic interactions dominate complex formation with positive charges on cytochrome *c* interacting with negative charges on CCO, most likely with residues in subunit II. The entry point of an electron from ferrocytochrome *c* is thus Cu_A. The rate of electron transfer is fast, on the order of 70 000 sec^{-1}. Next, the electron is transferred from Cu_A to heme A at a rate of $\approx$20 000 sec^{-1}. Finally, the electron is transferred from heme a to the site of O_2 activation, heme a_3. Since heme a and heme a_3 are close, the electron transfer very likely occurs *via* heme edges of closest approach.

7.4 O_2 Activation

A variety of sophisticated spectral methods have been utilized in an attempt to understand how O_2 is activated at the heme a_3/Cu_B center and what intermediates are formed during this process. Time-resolved

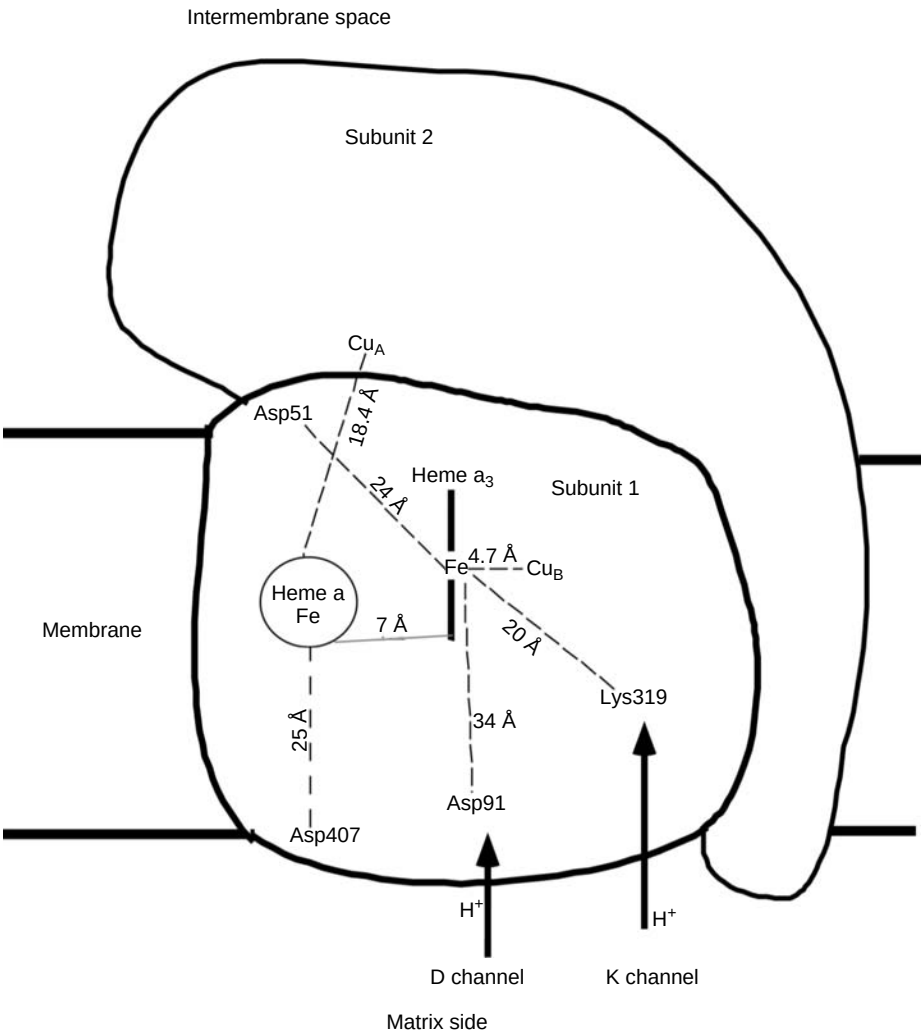

Fig. 18 A schematic diagram of cytochrome *c* oxidase. The various distances between heme and copper centers are indicated and the possible proton-pumping channels highlighted.

resonance Raman is uniquely suited to providing such structural information. In the Raman studies, the time course of various Raman bands is first resolved and then structures are assigned to these transient species. The results of these studies are summarized in Fig. 19. The a_3 heme iron must first be reduced to Fe^{2+} in order to bind O_2. The oxy–complex in Fig. 19 is shown as the Fe^{3+}–superoxide complex, which is likely the favored resonance structure. Cu_B delivers an electron to the superoxide intermediate, thus generating the peroxy-complex. As in peroxidases, the peroxide O−O bond cleaves heterolytically giving the Fe^{3+}−O hypothetical species in which the O atom has only six valence electrons and is a potent oxidizing agent

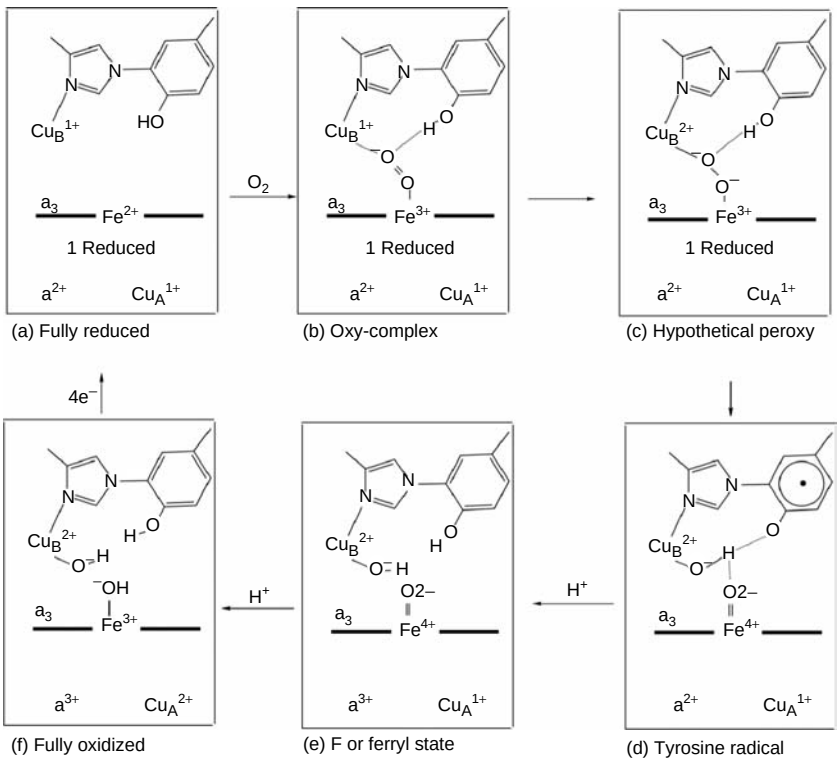

Fig. 19 Schematic representation of the cytochrome *c* oxidase (CCO) O_2 activation process taken after Proshlyakov et al. CCO has three redox active metal centers. Cytochrome *c*, the soluble electron transfer protein, delivers electrons to the two-copper Cu_A center and then to heme a. From heme a, electrons move to the binuclear heme a_3/Cu_B center where O_2 activation occurs. The fully reduced enzyme in (a) binds molecular oxygen to give the oxy-complex in (b), which is equivalent to the Fe^{3+}–superoxide complex as shown in (b). In (c), Cu_B has transferred an electron to dioxygen to give the peroxy-complex. The step from (c) to (d) results in heterolytic cleavage of the O—O bond leaving hydroxide bound to Cu_B and an oxygen atom linked to the iron. The iron-linked O atom contains only six valence electrons and is a potent oxidant. As a result, one electron is removed from the iron to give Fe^{4+} and, one from another group, which now is thought to be tyrosine, giving the structure in (d). Intermediate (d) is similar to peroxidase Compound I. Steps (d) to (e) and (e) to (f) involve intramolecular electron transfer, which gives back the resting state oxidized enzyme. The cycle begins again by reduction of (f) by ferrocytochrome *c*.

requiring two more electrons. One electron is transferred from the iron to give Fe^{4+} and the second from the tyrosine residue covalently linked to a Cu_B histidine ligand to give a tyrosine radical. Tyrosine thus serves a similar role to the porphyrin radical in peroxidase Compound I.

7.5 Proton Pumping

The mechanism just described shares similar features to peroxidases and P450s. The unique feature of CCO, however, is that the electron transfer and O_2 activation

processes are coupled to the movement of protons across the membrane. It is generally thought that CCO must have a channel that utilizes protein groups to move protons along from one proton acceptor to the next. Two primary channels, designated D and K after Asp91 and Lys319, have been the primary regions thought to provide proton channels. Site-directed mutants in channel D in bacterial CCOs results in the enzyme being trapped at intermediate four (Fig. 18).

8 Cyclooxygenase

8.1 Introduction

Prostaglandins are long chain fatty acids derived from arachidonic acid (Fig. 20). Arachidonic acid is first released from glycerolphospholipids followed by the oxidative reaction shown in Fig. 20. The ability of prostaglandins to stimulate smooth muscle contraction and regulate blood pressure was the first indication of the hormonal action of prostaglandins and related compounds. The initial and committed step in the conversion of arachidonic acid to prostaglandins is catalyzed by cyclooxygenases, also called prostaglandin enderperoxidase synthase or PGHS. Owing to the potent regulatory properties of prostaglandins, PGHS is the target for some well-known drugs. Anti-inflammatory drugs such as aspirin and ibuprofen as well the new generation of COX-2 inhibitors exhibit their effects by blocking PGHS. PGHS is a membrane-bound peroxidase, but unlike traditional

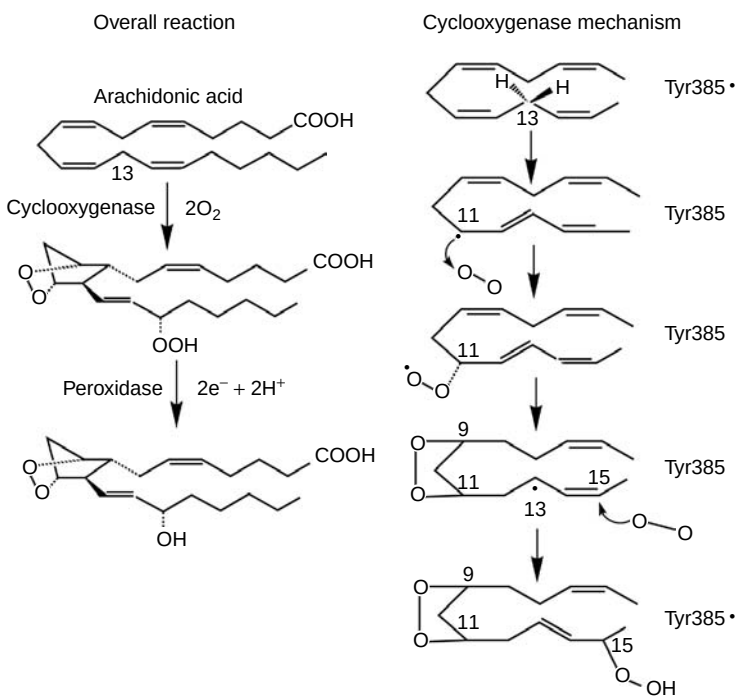

Fig. 20 The overall reaction and mechanism of cyclooxygenase. The reaction is initiated by oxidation of the substrate by a nearby tyrosine radical.

heme peroxidases, PGHS catalyzes two distinct reactions at two different sites. The first is the cyclooxygenase reaction (Fig. 20) followed by a peroxidase reaction that cleaves the O—O bond of the peroxide attached to C15 (Fig. 20). Both known isoforms, PGHS-1 and PGHS-2, are very similar and differ primarily in regulatory properties. PGHS-1 is constitutively expressed while PGHS-2 is induced.

8.2 Structure and Mechanism

The structure of PGHS-1 in a complex with arachidonic acid is shown in Fig. 21. Consisting of 576 residues PGHS-1 is much larger than the other heme peroxidases considered in this article. This increased complexity is due in part to the addition of a membrane-binding domain and the relatively large size of the substrate, and there is additional complexity in the catalytic cycle, which requires two activities. Despite these differences, the peroxidase active site is very similar to other peroxidases. Histidine serves as an axial heme ligand (Fig. 21), while in the distal pocket, His207 serves as the acid–base catalytic group required for heterolytic fission of the peroxide O—O bond. However, PGHS-1 has Gln203 in place of the distal arginine found in other peroxidases. This same type of histidine/glutamine distal site pocket is found in another mammalian peroxidase, myeloperoxidase.

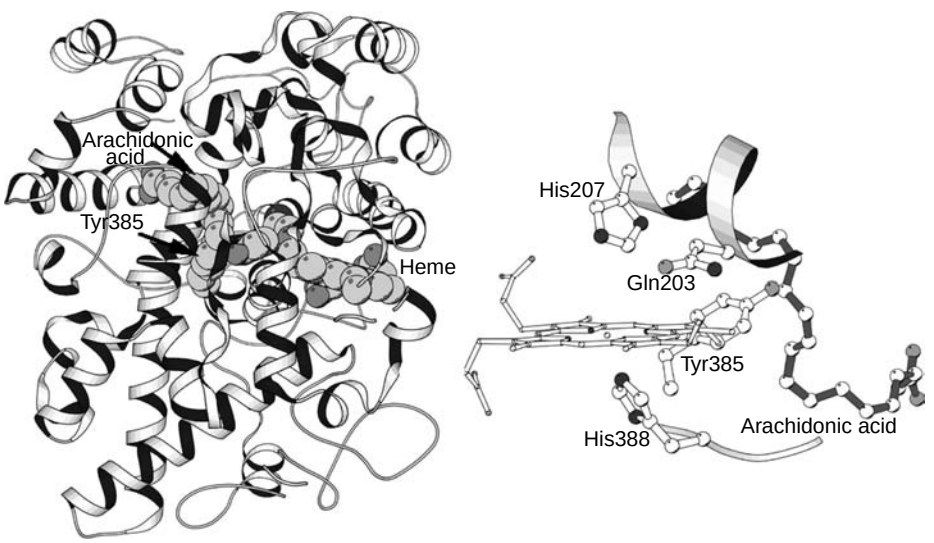

Fig. 21 The structure of PGHS-1. Although PGHS-1 is membrane bound, it is not an integral membrane protein like cytochrome *c* oxidase. Instead, PGHS contains a small membrane-binding domain, which results in most of the enzyme "floating" on the surface of the membrane. The enzyme has two active site regions. One is the heme site where peroxidase catalysis takes place. The heme active site architecture is very similar to traditional peroxidase. Note the distal His207, which operates as an acid–base catalytic group as in other peroxidases. The second site where fatty acid substrates bind is near Tyr385. The heme-site peroxidase activity results in oxidation of Tyr385 to a radical that then supports the cyclooxygenase activity as outlined in Fig. 20.

The mechanism shown in Fig. 20 requires that Tyr385 first be oxidized to a radical. This reaction is achieved *via* the traditional peroxidase mechanism.

$$Fe^{3+}P\ Tyr385 + ROOH \longrightarrow \underset{\text{Compound I}}{Fe^{4+}{-}O\ P^{\bullet}\ Tyr385} + ROH \quad (9)$$

$$Fe^{4+}{-}O\ P^{\bullet}\ Tyr385 \longrightarrow Fe^{4+}{-}O\ P^{\bullet}\ Tyr385. \quad (10)$$

$$\underset{\text{Compound II}}{Fe^{4+}{-}O\ P\ Tyr385^{\bullet}} + AA \longrightarrow Fe^{4+}{-}O\ PTyr385 + AA^{\bullet} \quad (11)$$

$$Fe^{4+}{-}O\ P\ Tyr385 + AA^{\bullet} + 2O_2 \longrightarrow^{\bullet} Fe^{4+}{-}O\ P\ Tyr385 + O_2\ AAOO^{\bullet} \quad (12)$$

$$Fe^{4+}{-}O\ P\ Tyr385 + O_2AAOO^{\bullet} \longrightarrow^{\bullet} Fe^{4+}{-}O\ P\ Tyr385^{\bullet} + O_2\ AAOOH \quad (13)$$

In Step 1, the usual Compound I is formed *via* reaction with a hydroperoxide. Tyr385 is less than 5 Å from the heme; so the oxidation of Tyr385 to a radical in Step 2 by the porphyrin radical can readily occur. As shown in Fig. 20, C13 of arachidonic acid (AA) directly contacts Tyr385, thus enabling regio- and stereoselective abstraction of the *pro*S hydrogen to give the arachidonic acid radical, AA• After the two oxygenation reactions in Steps 4 and 5, the AA peroxy radical in Step 5 oxidizes Tyr385 to regenerate the Try385 radical. Details of the cyclooxygenase reaction in Steps 4 and 5 are outlined in Fig. 20. Reduction of O_2AAOOH to O_2AAOH proceeds *via* the traditional peroxidase mechanism to give Compound I.

9 Hydroxylamine Oxidoreductase (HAO)

HAO catalyzes the four-electron oxidation of hydroxylamine to nitrite

$$NH_2OH + H_2O \longrightarrow NO_2^{-} + 4e^{-} + 5H^{+} \quad (14)$$

The electron acceptor from HAO is thought to be a *c*–type cytochrome. HAO is found primarily in microorganisms that oxidize ammonia. Little is known about the details of the HAO mechanism, but it is included in this article as an example of the fascinating complexity of some heme enzymes. The HAO monomer consists of 570 residues with a remarkable eight-heme-per-monomer structure (Fig. 22). Since functional HAO is a homo-trimer, the active enzyme contains a total of 24 hemes. One of the hemes exhibits unusual spectroscopic properties with an absorption maximum near 463 nm. The so-called P460 heme is covalently cross-linked to a tyrosine residue from a neighboring subunit. In addition, the P460 heme is also thought to be the site of substrate oxidation. The P460 heme site does share some properties with other heme enzymes. One axial ligand is histidine, while the remaining coordination position is free to interact with substrates. The P460 distal pocket contains a histidine residue that could serve the same acid–base catalytic function found in peroxidases. The remaining seven hemes most likely function to transfer electrons from the P460 site. The requirement for so many hemes may be based on the chemical requirements of converting hydroxylamine to nitrite. This is a two-step process:

$$NH_2OH \longrightarrow HNO + 2H^{+} + 2e^{-}\ \text{dehydrogenation} \quad (15)$$

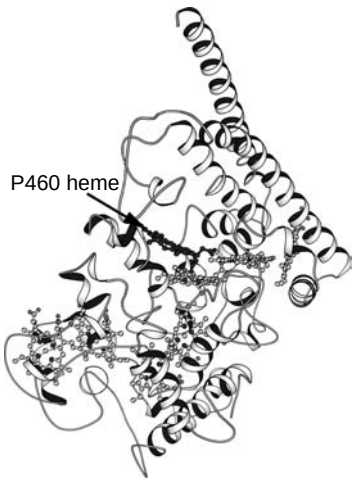

Fig. 22 The structure hydroxylamine oxidoreductase or HAO. Although HAO functions as a trimer, only the monomer is shown. Highlighted is one of the eight hemes, P460, which is the site of substrate oxidation.

$$HNO + H_2O \longrightarrow HNO_2 + 2H^+ + 2e^- \text{ oxygenation.} \quad (16)$$

HNO remains bound as an enzyme intermediate. If continuous electron flow is not maintained, the HAO generates N_2O rather than NO_2^-. Thus, two electrons must be removed, and since hemes are only one-electron carriers, it appears that the additional hemes are required to ensure the off-loading of electrons such that the HNO intermediate does not short-circuit to generate other products like N_2O or NO.

See also Hemoglobin; Metalloenzymes.

Bibliography

Books and Reviews

Dunford, H.B. (1999) *Heme Peroxidases*, Wiley-VCH, New York.

Ludwig, B., Bender, E., Arnold, S., Huttemann, M., Lee, I., Kadenbach, B. (2001) *Chem. biochem.* **2**, 392–403.

Messerschmidt, A., Huber, R., Poulos, T, Wieghardt, K. (Eds.) (2001) *Handbook of Metalloproteins* Volume 1 Wiley, Chichester, New York, Weinheim, Brisbane, Singapore, Toronto.

Ortiz de Montellano, P.R (Ed.) (1995) *Cytochrome P450: Structure, Mechanism, and Biochemistry* Plenum Press, New York.

Poulos, T.L., Fenna, R. (1994) *Met. Biol.* **30**, 25–75.

Poulos, T.L., Li, H., Raman, C.S., Schuller, D.J. (2001) *Adv. Inorg. Chem.* **51**, 243–293.

Primary Literature

Allain, E.J., Hager, L.P., Deng, L., Jacobsen, E.N. (1993) *J. Am. Chem. Soc.* **115**, 4415–4416.

Brezinski, P., Adelroth, P. (1998) *J. Bioenerg. and Biomembr.* **30**, 99–107.

Chance, B. (1943) *J. Biol. Chem.* **151**, 553–577.

Chouchane, S., Lippai, I., Magliozzo, R.S. (2000) *Biochemistry* **39**, 9975–9983.

Colonna, S., Gaggero, N., Manfredi, A., Casella, L., Gullotti, M., Pasta, P. (1990) *Biochemistry* **29**, 10465–10468.

Crane, B.R., Arvai, A.S., Ghosh, D.K., Wu, C., Getzoff, E.D., Stuehr, D.J., Tainer, J.A. (1998) *Science* **279**, 2121–2126.

Cupp-Vickery, J.R., Poulos, T.L. (1995) *Nat. Struct. Biol.* **2**, 144–153.

Edwards, S.L., Raag, R., Wariishi, H., Gold, M.H., Poulos, T.L. (1993) *Proc. Natl. Acad. Sci. U S A* **90**, 4429–4440.

Erman, J.E., Vitello, L.B., Miller, M.A., Shaw, A., Brown, K.A., Kraut, J. (1993) *Biochemistry* **32**, 9798–9806.

Fergusson-Miller, S., Babbock, G. (1996) *Chem. Rev.* **96**, 2889–2907.

Fischmann, T.O., Hruza, A., Niu, X.D., Fossetta, J.D., Lunn, C.A., Dolphin, E., Prongay, A.J., Reichert, P., Lundell, D.J., Narula, S.K., Patricia, C.W. (1999) *Nat. Struct. Biol.* **6**, 233–242.

Fulop, V., Ridout, C.J., Greenwood, C., Hajdu, J. (1995) *Structure* **3**, 1225–1233.

Gajhede, M., Schuller, D.J., Henriksen, A., Smith, A.T., Poulos, T.L. (1997) *Nat. Struct. Biol.* **4**, 1032–1038.

Goodin, D., McRee, D.E. (1993) *Biochemistry* **32**, 3313–3324.

Hasemann, C.A., Kurumbail, R.G., Boddupalli, S.S., Peterson, J.A., Deisenhofer, J. (1995) *Structure* **3**, 41–62.

Huyett, J.E., Doan, P.E., Gurbriel, R., Houseman, A.L.P., Sivaraja, M., Goodin, D.B., Hoffman, B.M. (1995) *J. Am. Chem. Soc.* **117**, 9033–9041.

Igarashi, N., Moriyama, H., Fujiwara, T., Fukumori, Y., Tanaka, N. (1997) *Nat. Struct. Biol.* **4**, 276–284.

Iwata, S., Ostermeier, C., Ludwig, B., Michel, H. (1995) *Nature* **376**, 660–669.

Jensen, G.M., Bunte, S.W., Warshel, A., Goodin, D.B. (1998) *J. Phys. Chem. B* **102**, 8221–8228.

Kannt, A., Lancaster, R., Michel, H. (1998) *Biophys. J.* **74**, 708–721.

Kunishima, N., Fukuyama, K., Matsubara, H., Hatanaka, H., Shibano, Y., Amachi, T. (1994) *J. Mol. Biol.* **235**, 331–344.

Kurumbail, R., Stevens, A., Gierse, J., McDonald, J., Stegeman, R., Pak, J., Gildenhaus, D., Miyashiro, J., Penning, T., Seibert, K., Isakson, P.C. Stallings, W. (1996). *Nature* **384**, 644–688.

Lad, L., Schuller, D.J., Shimizu, H., Friedman, J., Li, H., Ortiz de Montellano, P.R., Poulos, T.L. (2003) *J. Biol. Chem.* **278**, 7834–7843.

Libby, R.D., Thomas, J.A., Kaiser, L.W., Hager, L.P. (1981) *J. Biol. Chem.* **257**, 5030–5037.

Luong, C., Miller, A., Barnett, J., Chow, J., Ramesha, C., Browner, M. (1996) *Nat. Struct. Biol.* **384**, 927–933.

Malkowski, M.G., Ginell, S.L., Smth, W.L., Garavito, R.M. (2000) *Science* **289**, 933–1937.

Michel, H., Behr, J., Harrenga, A., Kannt, A. (1998) *Annu. Rev. Biophys. Biomol. Struct.* **27**, 329–356.

Murthy, M.R., Reid, T.J.d., Sicignano, A., Tanaka, N., Rossmann, M.G. (1981) *J. Mol. Biol.* **152**, 465–499.

Ortiz de Montellano, P.R. (2000) *Curr. Opin. Chem. Biol.* **4**, 221–227.

Ostermeier, C., Harrenga, A., Ermler, U., Michel, H. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 10547–10553.

Pelletier, H., Kraut, J. (1992) *Science* **258**, 1748–1755.

Pfitzner, U., Odenwald, A., Ostermann, T., Weingard, L., Ludwig, B., Richter, O. (1998) *J. Bioenerg. Biomembr.* **30**, 89–97.

Picot, D., Loll, P.J., Garavito, R.M. (1994) *Nature* **367**, 243–249.

Podust, L.M., Poulos, T.L., Waterman, M.R. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 3068–3073.

Podust, L., Kim, Y., Arase, M., Neely, B., Beck, B., Bach, H., Sherman, D., Lamb, D., Kelly, S., Waterman, M. (2003) *J. Biol. Chem.* **278**, 12214–12221.

Poulos, T.L., Kraut, J. (1980a) *J. Biol. Chem.* **255**, 10322–10330.

Poulos, T.L., Kraut, J. (1980b) *J. Biol. Chem.* **255**, 8199–8205.

Poulos, T.L., Finzel, B.C., Howard, A.J. (1987) *J. Mol. Biol.* **195**, 687–700.

Proshlyakov, D., Pressler, M., Babcock, G. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 8020–8025.

Raman, C.S., Li, H., Martasek, P., Kral, V., Masters, B.S., Poulos, T.L. (1998) *Cell* **95**, 939–950.

Ravichandran, K.G., Boddupalli, S.S., Hasermann, C.A., Peterson, J.A., Deisenhofer, J. (1993) *Science* **261**, 731–736.

Schuller, D.J., Wilks, A., Ortiz de Montellano, P.R., Poulos, T.L. (1999) *Nat. Struct. Biol.* **6**, 860–867.

Schuller, D.J., Zhu, W., Stojiljkovic, I., Wilks, A., Poulos, T.L. (2001) *Biochemistry* **40**, 11552–11558.

Sevrioukova, I.F., Li, H., Zhang, H., Peterson, J.A., Poulos, T.L. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 1863–1868.

Sharp, K.H., Mewies, M., Moody, P.C., Raven, E.L. (2003) *Nat. Struct. Biol.* **10**, 303–307.

Shimizu, H., Schuller, D.J., Lanzilotta, W.N., Sundaramoorthy, S., Arciero, D.M., Hooper, A.B., Poulos, T.L. (2001) *Biochemistry* **40**, 13483–13490.

Sivaraja, M., Goodin, D.B., Smith, M., Hoffman, B.M. (1989) *Science* **245**, 738–740.

Smith, W.L., DeWitt, D.L., Garavito, R.M. (2000) *Annu. Rev. Biochem.* **69**, 145–182.

Sucheta, A., Georgiadis, K., Einarsdottir, O. (1997) *Biochemistry* **36**, 554–565.

Sugishima, M., Omato, Y., Kakuta, Y., Sakamoto, H., Noguchi, M., Fukuyama, K. (2000) *FEBS Lett.* **471**, 61–66.

Sundaramoorthy, M., Terner, J., Poulos, T.L. (1995) *Structure* **3**, 1367–1377.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., Yoshikawa, S. (1995) *Science* **269**, 1069–1074.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., Yoshikawa, S. (1996) *Science* **272**, 1136–1144.

Wang, M., Roberts, D.L., Paschke, R., Shea, T.M., Masters, B.S., Kim, J.-J. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 8411–8416.

Weyand, M., Hect, H., Keiss, M., Liaud, M., Vilter, H., Schomburg, D. (1999) *J. Mol. Biol.* **293**, 595–611.

Williams, P.A., Cosme, J., Sridhar, V., Johnson, E.F., McRee, D.E. (2000) *Mol. Cell* **5**, 121–132.

Yoshikawa, Y. (1997) *Curr. Opin. Struct. Biol.* **7**, 574–579.

Yoshikawa, S., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., Yamashita, E., Inoue, N., Yao, M., Fei, M., Libeu, C., Mizushima, T., Yamaguchi, H., Tomizaki, T., Tsukihara, T. (1998) *Science* **280**, 1723–1729.

Hemoglobin

Gino Amiconi and Maurizio Brunori
University La Sapienza, Rome, Italy

1 General Aspects 70

2 Structural Features and Basic Terminology 71

3 Hemoglobin Components in Erythrocytes 76

4 Derivatives with Heme Ligands 77
4.1 Ferrous Derivatives 77
4.2 Ferric Derivatives 77

5 The Reaction with Heme Ligands 77
5.1 Functional Properties at Equilibrium 77
5.2 Kinetic Aspects 80

6 Mechanisms of Cooperative Binding and Allostery 81
6.1 Concerted Model of Allostery 83
6.2 Structural Changes That Accompany Oxygen Binding 84
6.3 Ligand-binding Pathways 85

7 Assembly of Globin Monomers 86
7.1 Variations Across Species 87
7.2 Subunit Interaction and Cooperativity 88

8 Reactions with Nitric Oxide 89
8.1 Reactions of NO with Various Hb Derivatives 90
8.2 Physiological Implications 91

9 The Globin Gene and its Biosynthesis 92

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

10 Evolutionary Considerations 97

11 Hemoglobin Variants 101
11.1 Nature of Mutants 101
11.2 Molecular Basis of Hemoglobin Diseases 101

12 Hemoglobin Solutions as Blood Substitutes 103

Bibliography 105
Books and Reviews 105
Primary Literature 106
Early References 106
Recent References 106

Keywords

Allosteric Protein
A protein whose active site reactivity can be altered by the binding of a ligand (the allosteric effector) at a nonoverlapping site.

Blood Substitutes
Any material (e.g. human plasma, solutions of substances such as serum albumin or dextrans) used for transfusion, as a part of treatment of hemorrhage and/or circulatory shock. Since many of these agents provide only oxygen transport and volume replacement (e.g. those based on chemically modified or genetically engineered hemoglobins), they are more correctly termed (hemoglobin-based) *oxygen carriers*. On the other hand, plasma expander is a synonym used for solutions aimed exclusively at the prevention of hypovolemic (low blood volume) shock.

Bohr Effect
The influence of pH on the oxygen affinity of hemoglobin.

Cooperativity
Any process in which the initial event (e.g. the binding of one ligand, such as O_2, to the heme iron) acts to facilitate subsequent similar events (e.g. the binding of other identical ligands) at other sites of the same polymeric protein (e.g. one tetrameric hemoglobin molecule). Cooperative processes are initially phenomenological, that is, they usually begin with the observation of sigmoid (S-shaped) curves (quite different from the rectangular hyperbola, observed in the combination of O_2 with myoglobin) obtained in the plot of ligand-saturation fraction versus ligand concentration, and later develop into a mechanistic explanation.

Geminate Recombination
The rebinding of a ligand molecule to its original site (the heme iron) after photolysis and before escaping to the solvent to mingle with other ligand molecules. The kinetics of this reaction, which typically is a picosecond to nanosecond process, is independent of bulk ligand concentration.

Kinetics and Mechanism
The study of reaction rates (called *chemical kinetics*) is concerned not only with the speed of conversion of the reactants to products but also with the sequence of physical and chemical events (whose series is called the *reaction mechanism*) by which this conversion occurs. Thus, by noting how the rate of a reaction is affected by changing experimental conditions, the chronology of discrete steps within the overall process can be determined, and details of what is happening at the molecular level in each step can sometimes be learned.

Molecular Diseases
A sickness in which the clinical manifestations arise from alterations in the structure and function of some specific biomolecule (such as an enzyme).

Molecular Evolution
The process of cumulative changes in the structure of a protein (e.g. a hemoglobin subunit) because of the stochastic element of random mutagenesis. It is generally accepted that nature evolves protein for novel function by redesigning existing protein frameworks. Only major changes in the sequence of existing genes (such as insertions, tandem duplication, and circular permutations) provide opportunities for activating latent functionalities or for producing new functions.

Hemoglobin (Hb) is the generic name for a vital protein, basic to oxygen (O_2) metabolism of all vertebrates, some invertebrates, and plants that perform nitrogen fixation. The only vertebrates that survive without Hb in their blood are antarctic fishes (such as *Chaenocephalus aceratus Lonnberg*) that bear the Hb genes but do not express the proteins. Structural and functional properties of Hb vary widely in the different species, according to physiological requirements. The multifaceted behavior of Hb has challenged the interest of many scientists with different backgrounds (biochemists, physiologists, geneticists, and biophysicists), who have investigated this protein from extremely diverse points of view (from the quantum chemistry of the heme iron to the unloading of O_2 into the swim bladder of fish, thus regulating the creature's buoyancy). Increasing concern over viral contamination of blood is spurring the development of a blood substitute; solutions of chemically modified hemoglobin represent one option.

Moreover, Hb exhibits biochemical and biophysical phenomena similar to those of many multisubunit enzymes in their interactions with substrates and cofactors. Hb, therefore, has been the prototype used to do such fundamental work in

biochemistry as determining the mechanisms by which different parts of proteins communicate (i.e. the structural basis of cooperativity and allostery). It is not, therefore, surprising that tetrameric Hb has been called the *hydrogen molecule of biochemistry* because understanding its functions is so basic to protein chemistry at large; the hydrogen atom has been considered as myoglobin (Mb), the monomeric hemoprotein contained in red muscles.

1 General Aspects

Hemoglobin is a protein able to undergo a reversible reaction with O_2 by virtue of containing heme (Fe-protoporphyrin IX) as the prosthetic group. Classically, hemoproteins present in the blood or the hemolymph of vertebrates and invertebrates are named *hemoglobin* (Hbs), while those present in the striated or smooth muscles of all animals are called *myoglobin* (Mbs). However, Hbs dissolved in the blood of invertebrates are often called *erythrocruorins*. Very recently, simple hemoproteins similar to Hbs but different from Mbs have been discovered in tissues of humans, such as liver (named *cytoglobin*) and brain (named *neuroglobin*). Proteins having characteristics corresponding to the foregoing definition are also found in tissues of invertebrates other than blood and muscle (e.g. in the cells of the tracheal wall in *Gastrophylus* larvae or in the nerve cells of marine mollusks), in plants (e.g. lupin and soybean), and in single-cell organisms (e.g. protozoa, fungi, and bacteria): in all these cases, the term Hb has been adopted. A number of microorganisms (such as *Mycobacterium tuberculosis*) were recently shown to express Hbs, which are important components of pathogenicity. Examination of their three-dimensional structure clearly shows that these new hemoproteins belong to the extended globin superfamily, in spite of the fact that their polypeptidic chain is approximately 30% shorter than classical Hbs; this is why they have been designated *truncated Hbs*. On the other hand, the O_2 carrier present in the blood of several polychaete worms is named *chlorocruorin* since its prosthetic group is a chloroporphyrin, characterized by a formyl group substituted for the vinyl group in position 2 of the protoporphyrin IX.

In this article, attention is mainly focused on human Hb, the intraerythrocyte protein involved not only in the transport of O_2 from lungs to tissues but also in the reverse transport of carbon dioxide. Hb concentration in normal human erythrocytes is extremely high (340 g dm^{-3}) and corresponds to a 5.2 mM solution, taking a molecular mass of 65 000 g mol^{-1}. One erythrocyte contains approximately 300 million Hb tetramers. Even though the erythrocyte Hb molecules (which are spheroids of dimension $6.4 \times 5.5 \times 5$ nm) are on average only 1 nm apart, they can rotate and flow past one another without hindrance. All general ideas that originated from studies on human Hb have been successfully transferred to other vertebrate and invertebrate Hbs, and in general, to more complex respiratory proteins, which are sometimes very large and extracellular.

2
Structural Features and Basic Terminology

Hemoglobin consists of four polypeptide chains (globins). These apoproteins carry as a prosthetic group the heme, an Fe(II) complex of protoporphyrin IX (synthesized in the bone marrow from glycine and acetate). In the free heme, Fe(II) is rapidly oxidized to Fe(III) by O_2 and water, whereas in Hb, Fe(II) combines reversibly with O_2 and remains in the ferrous state in both the oxygenated and deoxygenated derivatives. Reversibility in binding with O_2 is achieved in Hb by the folding of the polypeptide chain around the heme group, which is enclosed in a hydrophobic pocket. When O_2 is bound at the Fe(II) (in a 1 : 1 stoichiometry), an electron rearrangement occurs, with the Fe(II) changing from high spin to low spin and the adduct acquiring the $Fe(III)O_2^-$ character.

Hemoglobin subunits are identical in pairs: each polypeptide chain of the two subunits designated α^1 and α^2 (where indices are used to denote relative location in the tetramer) has 141 amino acid residues; each of the β^1 and β^2 subunits has 146 amino acid residues. The tetrameric molecule of adult normal Hb is usually indicated as $\alpha_2\beta_2$. The α- and β-subunits have different amino acid sequences (see Table 1), but they fold up spontaneously into a similar three-dimensional structure (see Fig. 1), the so-called *globin fold*. The α- and β subunits, respectively, are made

Tab. 1 Primary structure of human globin subunits.

Helix	*α*	*ζ*	*β*	*δ*	*γ*	*ε*	*Myoglobin*
NA1			1 Val	Val	Gly	Val	1 Val
NA2	1 Val	Ser	2 His	His	His	His	2 Leu
NA3	2 Leu	Leu	3 Leu	Leu	Phe	Phe	–
A1	3 Ser	Thr	4 Thr	Thr	Thr	Thr	3 Ser
A2	4 Pro	Lys	5 Pro	Pro	Glu	Ala	4 Glu
A3	5 Ala	Thr	6 Glu	Glu	Glu	Glu	5 Gly
A4	6 Asp	Glu	7 Glu	Glu	Asp	Glu	6 Glu
A5	7 Lys	Arg	8 Lys	Lys	Lys	Lys	7 Trp
A6	8 Thr	Thr	9 Ser	Thr	Ala	Ala	8 Gln
A7	9 Asn	Ile	10 Ala	Ala	Thr	Ala	9 Leu
A8	10 Val	Ile	11 Val	Val	Ile	Val	10 Val
A9	11 Lys	Val	12 Thr	Asn	Thr	Thr	11 Leu
A10	12 Ala	Ser	13 Ala	Ala	Ser	Ser	12 His
A11	13 Ala	Met	14 Leu	Leu	Leu	Leu	13 Val
A12	14 Trp	Trp	15 Trp	Trp	Trp	Trp	14 Trp
A13	15 Gly	Ala	16 Gly	Gly	Gly	Ser	15 Ala
A14	16 Lys	Lys	17 Lys	Lys	Lys	Lys	16 Lys
A15	17 Val	Ile	18 Val	Val	Val	Met	17 Val
A16	18 Gly	Ser	–				18 Glu
AB1	19 Ala	Thr	–				19 Ala
B1	20 His	Gln	19 Asn	Asn	Asn	Asn	20 Asp
B2	21 Ala	Ala	20 Val	Val	Val	Val	21 Val
B3	22 Gly	Asp	21 Asp	Asp	Glu	Glu	22 Ala
B4	23 Glu	Thr	22 Glu	Ala	Asp	Glu	23 Gly

(*continued overleaf*)

Tab. 1 (*continued*)

Helix	***α***	***ζ***	***β***	***δ***	***γ***	***ε***	***Myoglobin***
B5	24 Thr	Ile	23 Val	Val	Ala	Ala	24 His
B6	25 Gly	Gly	24 Gly	Gly	Gly	Gly	25 Gly
B7	26 Ala	Thr	25 Gly	Gly	Gly	Gly	26 Gln
B8	27 Glu	Glu	26 Glu	Glu	Glu	Glu	27 Asp
B9	28 Ala	Thr	27 Ala	Ala	Thr	Ala	28 Ile
B10	29 Leu	Leu	28 Leu	Leu	Leu	Leu	29 Leu
B11	30 Glu	Glu	29 Gly	Gly	Gly	Gly	30 Ile
B12	31 Arg	Arg	30 Arg	Arg	Arg	Arg	31 Arg
B13	32 Met	Leu	31 Leu	Leu	Leu	Leu	32 Leu
B14	33 Phe	Phe	32 Leu	Leu	Leu	Leu	33 Phe
B15	34 Leu	Leu	33 Val	Val	Val	Val	34 Lys
B16	35 Ser	Ser	34 Val	Val	Val	Val	35 Ser
C1	36 Phe	His	35 Tyr	Tyr	Tyr	Tyr	36 His
C2	37 Pro	Pro	36 Pro	Pro	Pro	Pro	37 Pro
C3	38 Thr	Gln	37 Trp	Trp	Trp	Trp	38 Glu
C4	39 Thr	Thr	38 Thr	Thr	Thr	Thr	39 Thr
C5	40 Lys	Lys	39 Gln	Gln	Gln	Gln	40 Leu
C6	41 Thr	Thr	40 Arg	Arg	Arg	Arg	41 Glu
C7	42 Tyr	Tyr	41 Phe	Phe	Phe	Phe	42 Lys
CD1	43 Phe	Phe	42 Phe	Phe	Phe	Phe	43 Phe
CD2	44 Pro	Pro	43 Glu	Glu	Asp	Asp	44 Asp
CD3	45 His	His	44 Ser	Ser	Ser	Ser	45 Arg
CD4	46 Phe	Phe	45 Phe	Phe	Phe	Phe	46 Phe
CD5	47 Asp	Asp	46 Gly	Gly	Gly	Gly	47 Lys
CD6	–		47 Asp	Asp	Asn	Asn	48 His
CD7	48 Leu	Leu	48 Leu	Leu	Leu	Leu	49 Leu
CD8	49 Ser	His	49 Ser	Ser	Ser	Ser	50 Lys
D1	–		50 Thr	Ser	Ser	Ser	51 Thr
D2	–		51 Pro	Pro	Ala	Pro	52 Glu
D3	–		52 Asp	Asp	Ser	Ser	53 Ala
D4	–		53 Ala	Ala	Ala	Ala	54 Glu
D5	–		54 Val	Val	Ile	Ile	55 Met
D6	50 His	Pro	55 Met	Met	Met	Leu	56 Lys
D7	51 Gly	Gly	56 Gly	Gly	Gly	Gly	57 Ala
E1	52 Ser	Ser	57 Asn	Asn	Asn	Asn	58 Ser
E2	53 Ala	Ala	58 Pro	Pro	Pro	Pro	59 Glu
E3	54 Gln	Gln	59 Lys	Lys	Lys	Lys	60 Asp
E4	55 Val	Leu	60 Val	Val	Val	Val	61 Asp
E5	56 Lys	Arg	61 Lys	Lys	Lys	Lys	62 Lys
E6	57 Gly	Ala	62 Ala	Ala	Ala	Ala	63 Lys
E7	58 His	His	63 His	His	His	His	64 His
E8	59 Gly	Gly	64 Gly	Gly	Gly	Gly	65 Gly
E9	60 Lys	Ser	65 Lys	Lys	Lys	Lys	66 Val
E10	61 Lys	Lys	66 Lys	Lys	Lys	Lys	67 Thr
E11	62 Val	Val	67 Val	Val	Val	Val	68 Val
E12	63 Ala	Val	68 Leu	Leu	Leu	Leu	69 Leu
E13	64 Asp	Ala	69 Gly	Gly	Thr	Thr	70 Thr

Tab. 1 (*continued*)

Helix	***α***	***ζ***	***β***	***δ***	***γ***	***ε***	***Myoglobin***
E14	65 Ala	Ala	70 Ala	Ala	Ser	Ser	71 Ala
E15	66 Leu	Val	71 Phe	Phe	Leu	Phe	72 Leu
E16	67 Thr	Gly	72 Ser	Ser	Gly	Gly	73 Gly
E17	68 Asn	Asp	73 Asp	Asp	Asp	Asp	74 Ala
E18	69 Ala	Ala	74 Gly	Gly	Ala	Ala	75 Ile
E19	70 Val	Val	75 Leu	Leu	Ile, Thr	Ile	76 Leu
E20	71 Ala	Lys	76 Ala	Ala	Lys	Lys	77 Lys
EF1	72 His	Ser	77 His	His	His	Asn	78 Lys
EF2	73 Val	Ile	78 Leu	Leu	Leu	Met	79 Lys
EF3	74 Asp	Asp	79 Asp	Asp	Asp	Asp	80 Gly
EF4	75 Asp	Asp	80 Asn	Asn	Asn	Asp	81 His
EF5	76 Met	Ile	81 Leu	Leu	Leu	Leu	82 His
EF6	77 Pro	Gly	82 Lys	Lys	Lys	Lys	83 Glu
EF7	78 Asn	Gly	83 Gly	Gly	Gly	Pro	84 Ala
EF8	79 Ala	Ala	84 Thr	Thr	Thr	Ala	85 Glu
F1	80 Leu	Leu	85 Phe	Phe	Phe	Phe	86 Leu
F2	81 Ser	Ser	86 Ala	Ser	Ala	Ala	87 Lys
F3	82 Ala	Lys	87 Thr	Gln	Gln	Lys	88 Pro
F4	83 Leu	Leu	88 Leu	Leu	Leu	Leu	89 Leu
F5	84 Ser	Ser	89 Ser	Ser	Ser	Ser	90 Ala
F6	85 Asp	Glu	90 Glu	Glu	Glu	Glu	91 Gln
F7	86 Leu	Leu	91 Leu	Leu	Leu	Leu	92 Ser
F8	87 His	His	92 His	His	His	His	93 His
F9	88 Ala	Ala	93 Cys	Cys	Cys	Cys	94 Ala
F10	–		–				95 Thr
FG1	89 His	Tyr	94 Asp	Asp	Asp	Asp	96 Lys
FG2	90 Lys	Ile	95 Lys	Lys	Lys	Lys	97 His
FG3	91 Leu	Leu	96 Leu	Leu	Leu	Leu	98 Lys
FG4	92 Arg	Arg	97 His	His	His	His	99 Ile
FG5	93 Val	Val	98 Val	Val	Val	Val	–
G1	94 Asp	Asp	99 Asp	Asp	Asp	Asp	100 Pro
G2	95 Pro	Pro	100 Pro	Pro	Pro	Pro	101 Ile
G3	96 Val	Val	101 Glu	Glu	Glu	Glu	102 Lys
G4	97 Asn	Asn	102 Asn	Asn	Asn	Asn	103 Tyr
G5	98 Phe	Phe	103 Phe	Phe	Phe	Phe	104 Leu
G6	99 Lys	Lys	104 Arg	Arg	Lys	Lys	105 Glu
G7	100 Leu	Leu	105 Leu	Leu	Leu	Leu	106 Phe
G8	101 Leu	Leu	106 Leu	Leu	Leu	Leu	107 Leu
G9	102 Ser	Ser	107 Gly	Gly	Gly	Gly	108 Ser
G10	103 His	His	108 Asn	Asn	Asn	Asn	109 Glu
G11	104 Cys	Cys	109 Val	Val	Val	Val	110 Ala
G12	105 Leu	Leu	110 Leu	Leu	Leu	Met	111 Ile
G13	106 Leu	Leu	111 Val	Val	Val	Val	112 Ile
G14	107 Val	Val	112 Cys	Cys	Thr	Ile	113 His
G15	108 Thr	Thr	113 Val	Val	Val	Ile	114 Val
G16	109 Leu	Leu	114 Leu	Leu	Leu	Leu	115 Leu

(*continued overleaf*)

Tab. 1 (*continued*)

Helix	*α*	*ζ*	*β*	*δ*	*γ*	*ε*	*Myoglobin*
G17	110 Ala	Ala	115 Ala	Ala	Ala	Ala	116 His
G18	111 Ala	Ala	116 His	Arg	Ile	Thr	117 Ser
G19	112 His	Arg	117 His	Asn	His	His	118 Arg
GH1	113 Leu	Phe	118 Phe	Phe	Phe	Phe	119 His
GH2	114 Pro	Pro	119 Gly	Gly	Gly	Gly	120 Pro
GH3	115 Ala	Ala	120 Lys	Lys	Lys	Lys	121 Gly
GH4	116 Glu	Asp	121 Glu	Glu	Glu	Glu	122 Asp
GH5	117 Phe	Phe	122 Phe	Phe	Phe	Phe	123 Phe
H1	118 Thr	Thr	123 Thr	Thr	Thr	Thr	124 Gly
H2	119 Pro	Ala	124 Pro	Pro	Pro	Pro	125 Ala
H3	120 Ala	Glu	125 Pro	Gln	Glu	Glu	126 Asp
H4	121 Val	Ala	126 Val	Met	Val	Val	127 Ala
H5	122 His	His	127 Gln	Gln	Gln	Gln	128 Gln
H6	123 Ala	Ala	128 Ala	Ala	Ala	Ala	129 Gly
H7	124 Ser	Ala	129 Ala	Ala	Ser	Ala	130 Ala
H8	125 Leu	Trp	130 Tyr	Tyr	Trp	Trp	131 Met
H9	126 Asp	Asp	131 Gln	Gln	Gln	Gln	132 Asn
H10	127 Lys	Lys	132 Lys	Lys	Lys	Lys	133 Lys
H11	128 Phe	Phe	133 Val	Val	Met	Leu	134 Ala
H12	129 Leu	Leu	134 Val	Val	Val	Val	135 Leu
H13	130 Ala	Ser	135 Ala	Ala	Thr	Ser	136 Glu
H14	131 Ser	Val	136 Gly	Gly	Gly, Ala	Ala	137 Leu
H15	132 Val	Val	137 Val	Val	Val	Val	138 Phe
H16	133 Ser	Ser	138 Ala	Ala	Ala	Ala	139 Arg
H17	134 Thr	Ser	139 Asn	Asn	Ser	Ile	140 Lys
H18	135 Val	Val	140 Ala	Ala	Ala	Ala	141 Asp
H19	136 Leu	Leu	141 Leu	Leu	Leu	Leu	142 Ile
H20	137 Thr	Thr	142 Ala	Ala	Ser	Ala	143 Ala
H21	138 Ser	Glu	143 His	His	Ser	His	144 Ala
H22	–		–				145 Lys
H23	–		–				146 Tyr
H24	–		–				147 Lys
H25	–		–				148 Glu
H26	–		–				149 Leu
HC1	139 Lys	Lys	144 Lys	Lys	Arg	Lys	150 Gly
HC2	140 Tyr	Tyr	145 Tyr	Tyr	Tyr	Tyr	151 Tyr
HC3	141 Arg	His	146 His	His	His	His	152 Gln
HC4	–		–				153 Gly

up of seven and eight α-helical segments, interrupted by nonhelical sections. The helices are named from A to H, and the nonhelical segments lying in between are lettered AB, BC, and so on; the nonhelical portions at the subunit ends are named NA (at the amino terminus) and HC (at the carboxyl terminus). Residues within each segment (helical as well as nonhelical) are numbered from the amino end (e.g. A1–A16, EF1–EF8, etc.). Therefore, each amino acid residue (usually represented with the three-letter code) is uniquely identified by its topology, its position in the

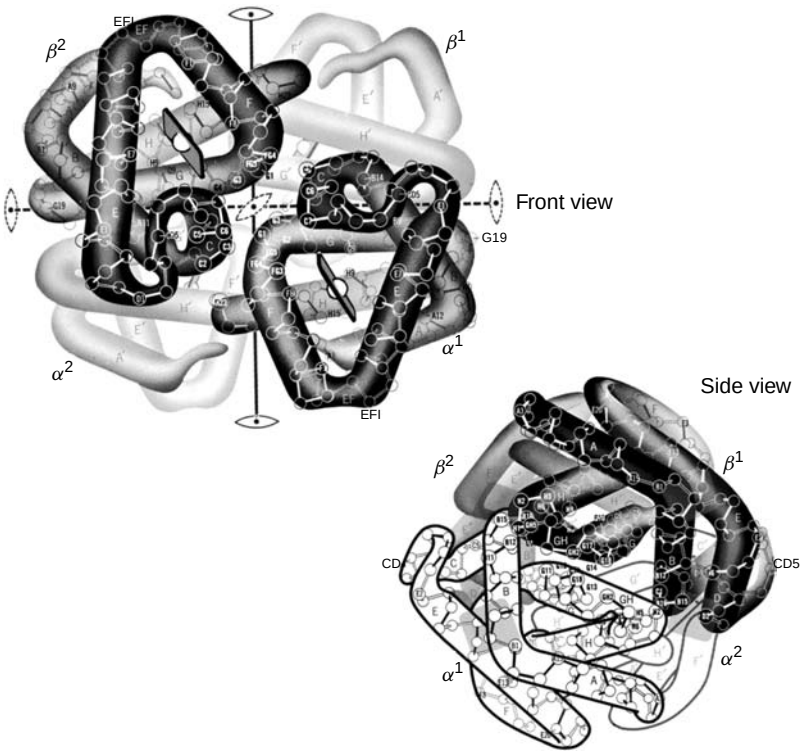

Fig. 1 Overview of ferric Hb, which is similar to the oxygenated derivative (HbO_2). Only the α-carbons of the main chains are shown. Those residues whose side chains are involved in contacts between subunits are circled and given boldface numbers. There are two kinds of unlike subunit contact: $\alpha^1\beta^1$ (or $\alpha^2\beta^2$) and $\alpha^1\beta^2$ (or $\alpha^2\beta^1$), the first interface being more extensive than the second (see also Table 2); the $\alpha^1\beta^1$ contact involves the B, C, and H helices and the GH corner, whereas the $\alpha^1\beta^2$ contact concerns mainly helices C and G and the FG corner. [Reproduced with permission from Irving Geis and R. E. Dickerson (1983) *Hemoglobin: Structure, Function, Evolution and Pathology*, Benjamin/Cummings, Menlo Park, CA.]

sequence (in parentheses), and the corresponding subunit [e.g. PheCD1(42)β].

The heme is bound to the globin by a number of weak hydrophobic interactions and by the mainly covalent bond between Fe(II) and the imidazole-N_ε of histidine F8, known as *proximal*.

In the β subunits, such a residue is followed by a reactive cysteine [CysF9(93)β], which faces the surface of the molecule: most, but not all, globins also have a histidine (HisE7, known as *distal*) on the O_2-combining side of the heme. This residue, however, is too far (>0.4 nm) from the iron to be expected to directly coordinate with the metal, even though under certain conditions, it does bind with the metal (see Sect. 4.2); usually, the N_ε atom of HisE7 makes a hydrogen bond with bound O_2 (see Fig. 2). The packing of subunits into the Hb tetramer is such that there are close interlocking connections of side groups between unlike subunits (i.e.

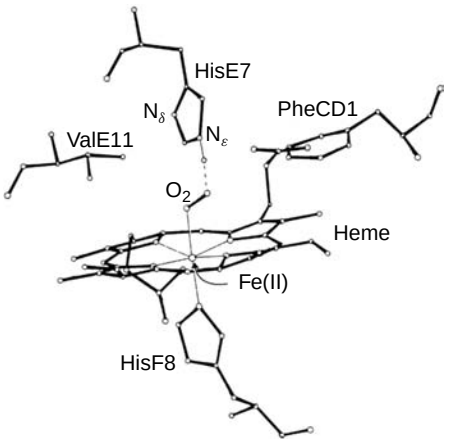

Fig. 2 Schematic representation of the heme pocket. Phenylalanine CD1 and the proximal histidine F8 are present in all hemoglobins. In vertebrates, valine E11 and the distal histidine E7 are involved in the reversible O_2 binding to the heme Fe(II). Note the hydrogen bond between the N_ε of the distal HisE7 and the bound O_2 molecule. HisE7 has a pK_a of about 5.5. Neutron diffraction has shown that in MbO_2, this amino acid is protonated only on N_ε, whereas in the carbon monoxy derivative, protonation occurs only on N_δ, which faces the bulk water.

Tab. 2 Surface area (nm^2) buried between contact interfaces in deoxy and oxy human hemoglobin.

Nature of surface area	*Deoxy Hb*	*Oxy Hb*
Buried in contact		
$\alpha^1\beta^1$ or $\alpha^2\beta^2$	18.76	20.48
$\alpha^1\beta^2$ or $\alpha^2\beta^1$	14.72	11.60
$\alpha^1\alpha^2$	7.91	9.66
$\beta^1\beta^2$	0	7.06
Accessible to solvent	72.55	80.32

α and β) but little contact between those that are alike (see Table 2).

3 Hemoglobin Components in Erythrocytes

Besides $\alpha_2\beta_2$ (also called HbA or HbA_0), human erythrocytes contain other Hb components, some as products of additional globin genes and others as posttranslational modifications of all these components. Among the hemoproteins belonging to the first group are HbA_2 and HbF (or fetal Hb), constituted from normal α-chains and combined with subunits whose structure differs significantly from that of normal β-chains; these non-α subunits are designated δ-chains (HbA_2 or $\alpha_2\delta_2$) and γ-chains (HbF or $\alpha_2\gamma_2$). Humans have also two embryonic Hbs using ζ- (α-like) and ε- (β-like) subunits: Hb Gower-1 of composition $\zeta_2\varepsilon_2$, Hb Gower-2 or $\alpha_2\varepsilon_2$, and Hb Portland or $\zeta_2\gamma_2$. Amino acid sequences of human chains are reported in Table 1.

Many small molecules (e.g. hexoses, cyanate ions derived from urea, and penicillin) are capable of forming covalent adducts with Hb and may reflect metabolic perturbation. Of significance for their clinical implications in diabetes are glycated Hbs (called Hb A_{Ia1}, A_{Ia2}, A_{Ib}, and A_{Ic}, according to their order of chromatographic elution) in which the amino terminus of the β-chains is bound to a molecule of glucose, glucose 6-phosphate, or fructose 1,6-bisphosphate; glucose can also form a ketoamine linkage with the amino-terminal group of α-chains as well as the ε-amino group of several lysyl residues of both α- and β-subunits. HbA_{Ic} is the most abundant minor (posttranslationally modified) Hb component in human erythrocytes, reaching about 4% of the total.

4 Derivatives with Heme Ligands

4.1 Ferrous Derivatives

The form of ferrous Hb with no ligand attached to the heme iron has been named unligated Hb (or deoxy Hb). It may be prepared by removal of the heme ligand (such as O_2, CO, NO, alkylisocyanides, and nitroso aromatic compounds, with NO being the strongest and bulky alkylisocyanides being the weakest ligand) from the corresponding ferrous ligated form. Usually, the starting derivative is oxygenated Hb (oxy Hb or HbO_2) and deoxygenation is achieved when such a derivative is exposed to vacuum or is equilibrated with an inert gas (e.g. N_2, Ar) or by addition of sodium dithionite ($Na_2S_2O_4$), which removes O_2 from solution. Deoxy Hb may also be prepared from the ferric derivatives by reduction of the Fe(III) by borohydride, dithionite, or ascorbate.

4.2 Ferric Derivatives

In the presence of O_2, the Fe(II) is subject to slow spontaneous oxidation, yielding ferric Hb or met Hb (which cannot combine with O_2). In blood, the rate of autoxidation is about 3% per day, but the oxidized iron-porphyrin is reduced to its active form by reductase systems. *In vitro* ferrous derivatives (such as oxy and deoxy Hb) are oxidized rapidly by ferricyanide or by nitrite. The water molecule bound to the Fe(III) in met Hb can be replaced by F^-, OCN^-, SCN^-, N_3^-, imidazole, or CN^-, where F^- is the weakest and CN^- the strongest ligand.

An endogenous ligand, such as the imidazole of distal HisE7, may also bind to Fe(III). This is a state that is either actually part of the distribution of substates populated in native Hb (named *bis-histidine complex* since the fifth and sixth axial positions of the Hb–Fe(III) are coordinated by proximal and distal histidines respectively), or may result from a large alteration of the geometry within the heme pocket; in this case, the low-spin derivative is called a *hemichrome*, a term connoting a denatured state. A more pronounced denaturation of the protein may cause the formation of other hemichromes; in one of these, one ligand of the Fe(III) is the sulfur of CysF9(93)β. Addition of dithionite to the hemichromes results in the reduction of heme Fe(III) to Fe(II), with the production of low-spin compounds called *hemochromes*.

5 The Reaction with Heme Ligands

The reaction of Hb with O_2 or other ligands of the ferrous heme is associated with large modifications of the spectral (see Fig. 3) and magnetochemical properties, which change in proportion with the amount of ligand bound.

5.1 Functional Properties at Equilibrium

The fundamental representation of the reversible combination of Hb with O_2 is the ligand-binding curve wherein the fraction of O_2-bearing heme groups (Y) is measured as a function of the O_2 partial pressure (p). The sigmoid shape of the curve is in marked contrast to the rectangular hyperbola found for the isolated α- or β-subunits (and also for

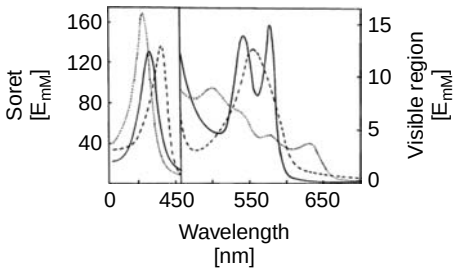

Fig. 3 Optical absorption spectra of deoxy (dashes), oxy (solid curves), and met (dots) hemoglobins in the visible and near-ultraviolet (the so-called Soret band) regions at pH 7.4: E_{mM}, millimolar extinction coefficient in heme (i.e. referred to an equivalent molecular mass of 16 250 g mol^{-1}).

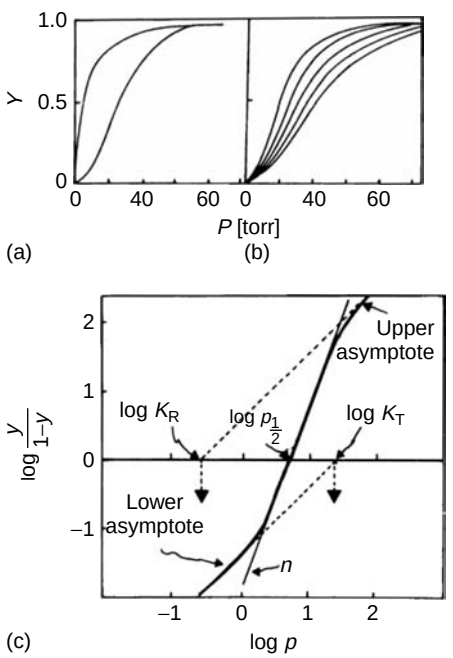

Fig. 4 Two ways of plotting the O_2 equilibrium curves of hemoglobin: (a) and (b) fractional saturation with O_2 (Y), reported as a function of O_2 partial pressure (p); (c) the Hill plot – see Eq. (5). (a) The typical shape of the O_2 equilibrium curve for Hb is sigmoid (right), while the curve for isolated β-subunits is hyperbolic (left). Both proteins are tetrameric ($\alpha_2\beta_2$ and β_4). In the case of the β-subunits, the four O_2-binding sites are equivalent and independent; on the other hand, the shape of the Hb oxygenation curve implies that the affinity of the free sites depends on the fraction of hemes already bound to O_2; (b) O_2-binding curves for Hb at various pH values (i.e. the Bohr effect): left to right, pH 7.6, 7.4, 7.2, 7.0, and 6.8; (c) the Hill plot for Hb oxygenation begins and ends with straight lines, called *asymptotes*, with a positive slope of unity ($n = 1$). The intercepts of these asymptotes, with the line parallel to the abscissa when $\log[Y/(1-Y)] = 0$ (i.e. when oxy and deoxy Hb are equal in concentration) give the values of the O_2 equilibrium constants for the T state (K_T) and the R state (K_R). Often the asymptotes cannot be determined accurately; in such cases, the oxygenation process is described in terms of the maximum slope of the curve n (taken as a measure of the degree of cooperativity) and the overall O_2 affinity in logarithmic terms (i.e. $\log p_{1/2}$), which corresponds to the value of $\log p$ when $\log[Y/(1-Y)] = 0$ (see Table 3).

Mb, see Fig. 4a). The sigmoidal curve reflects a phenomenon called *cooperativity*, a complicated process whereby initial O_2 binding favors more binding (as would occur in the lungs), or, from the opposite perspective, some O_2 release favors more release (as would occur in the peripheral tissues). Cooperativity characterizes all vertebrate Hbs, the only exception so far being the Hbs from *Sphenodon punctatus*, a reptilian relict from the Triassic period.

A useful way of analyzing the experimental data is based on the following scheme:

$$\mathrm{Hb} + n\mathrm{O_2} \rightleftharpoons \mathrm{Hb(O_2)}_n \qquad (1)$$

Assuming that the ratio of activities of Hb and $Hb(O_2)_n$ is given by their concentration ratio, the association equilibrium constant, K, can be written as follows:

$$K = \frac{[Hb(O_2)_n]}{[Hb]p^n} \quad (2)$$

where the activity of O_2 is expressed in terms of its partial pressure p. The average fraction of hemes oxygenated is

$$Y = \frac{[Hb(O_2)_n]}{[Hb] + [Hb(O_2)_n]} \quad (3)$$

Therefore, replacing $[Hb(O_2)_n]$ by Eq. (2), Eq. (3) gives

$$\frac{Y}{1 - Y} = Kp^n \quad (4)$$

or in logarithmic terms (the so-called *Hill equation*; see Fig. 4c),

$$\log \frac{Y}{1 - Y} = \log K + n \log p \quad (5)$$

Equation (5) has no direct physical meaning but, with a suitable choice of K and n, it is able to empirically fit the experimental results with only two parameters: (1) $p_{1/2}$, that is, the value of O_2 pressure required to yield half saturation of hemes (p value at $Y = 0.5$), which indicates the overall affinity of Hb for O_2 (the higher the affinity, the lower the $p_{1/2}$); and (2) n, the slope in the Hill plot (see Fig. 4c), which is related to the shape of ligand-binding curve and therefore to cooperativity (the higher its value above 1, the higher the cooperativity). A value of n equal to 1 (as occurring in tetrameric β-subunits) corresponds to a hyperbolic binding curve, indicating equivalence of hemes and absence of functional interactions. Observation of n less than 1 does not necessarily imply negative cooperativity (i.e. that binding of a ligand diminishes the affinity for subsequent binding) in that it can also arise from various kinds of heterogeneity: intramolecular (e.g. large difference in affinity for O_2, between α- and β-subunits in the tetramer) or intermolecular (e.g. mixture of various Hb species or conformations not at equilibrium). The O_2 affinity of Hb is lowered by H^+, Cl^-, CO_2, and 2,3-D-bisphosphoglycerate (BPG), all known as *heterotropic ligands*. In particular, the O_2-binding modulation by H^+ is known as the *Bohr effect* and reflects the release of protons upon Hb oxygenation at physiological pH (the alkaline Bohr effect), whereas protons are taken up upon oxygenation at pH values below ~6 (the acid or reverse Bohr effect). An unusually large Bohr effect, typical of some Hbs from many teleost fishes, is called the *Root effect*. By increasing the concentration of any one of the heterotropic ligands, the O_2 equilibrium curve is shifted to the right (see Fig. 4b) – that is, toward lower affinity – thus facilitating O_2 release. Table 3 reports the O_2-binding parameters at equilibrium of human Hb under different solvent conditions.

From a phenomenological viewpoint, mammalian Hbs can broadly be divided into two groups: those with an intrinsically high O_2 affinity, which is lowered in the erythrocytes by BPG, and those with an intrinsically low O_2 affinity, which is lowered little, or not at all, by BPG. Rodents, pigs, dogs, camels, horses, marsupials, and most primates belong to the first category; cats, sheep, goats, deer, cows, and one primate, the lemur, belong to the second. Typically, Hbs with high intrinsic O_2 affinity have $p_{1/2}$ values between 4 and 6 torr, while those with low affinity have $p_{1/2}$ values between 10 and 20 torr (measured at 20 to 25 °C in 0.05 M bis-Tris, with Cl^- concentration of 0.1 M, pH 6.5 to 7.5 as shown in Table 3).

Tab. 3 Values of parameters describing the functional properties of human hemoglobin in 0.05 M Tris, pH 7.4; Cl^- concentration 0.10 M, and 20 °C.

Conditions	*Oxygen equilibria in the presence and in the absence of BPG*					
	$p_{1/2}$[torr]	n	$K_T[\mu M^{-1}]$	$K_R[\mu M^{-1}]$	L_0	$\Delta G^c[kJ\ mol^{-1}]$
	5.3	2.8	0.018	3.9	7.3×10^5	13.0
+2 mM BPG	14.0	3.1	0.008	3.0	3.0×10^6	16.3

Ligand	*Kinetics of binding of various ligands*			
	$k'_{overall}[M^{-1}\ s^{-1}]$	$k_{overall}[s^{-1}]$	k'_4/k'_1	k_1/k_4
O_2	4.7×10^6	35	5	150
CO	2.0×10^5	0.015	40	10
NO	1.5×10^7	0.00005	1	100

5.2 Kinetic Aspects

Hemoglobin kinetics with ferrous heme ligands approximates second-order behavior in the combination direction and first-order behavior in the dissociation direction. Among the gaseous ligands that bind to the ferrous derivatives, at neutral pH and room temperature, NO binds fastest, O_2 not quite as fast, and CO significantly more slowly. The dissociation reactions follow a different order (see Table 3). When the rate constants of binding (k') or dissociation (k) of the first and fourth gaseous ligands are compared, the kinetic manifestation of cooperativity is observed (see Table 3). In this connection, it has been shown that cooperative binding in kinetic terms results from a significant increase in the association rate constants for CO and a decrease in the dissociation rates for O_2 and NO; accordingly, the transition state, as compared to reactants and products along the reaction path to heme, is expected to be reactant-like for O_2 and NO, and product-like for CO.

Since light breaks the bond between heme–Fe(II) and all heme ligands, the kinetic behavior of rebinding can be investigated after rapid photodissociation by illuminating with a laser pulse. The relaxation kinetics over all chemically meaningful timescales (from picosecond to second) demonstrates the existence of different rate-limiting steps for the various heme ligands and, therefore, the existence of intermediates in the reaction pathway. A four-state scheme is required:

$$\begin{array}{ccccc} \text{HbX} & & \text{P} & & \text{M} \\ \text{ligand} & \rightleftharpoons & \text{heme} & \rightleftharpoons & \text{protein} \\ \text{bound} & & \text{pocket} & & \text{matrix} \\ & \rightleftharpoons & \text{Hb} + \text{X} & & \\ & & \text{bulk} & & \end{array} \qquad (6)$$

where: HbX represents the bound state, Hb + X is an unligated molecule (i.e. the two entirely separated free species, Hb and X), P and M are intermediate states in which the heme Fe has no sixth ligand, but X remains trapped either in the heme pocket P or in the protein matrix M. From the latter states, the ligand may either rebind (to yield HbX) in a *geminate process* or escape to the solvent (to yield Hb + X).

The structural features of states P and M have been investigated by X-ray diffraction in single crystals of carbonylated myoglobin (MbCO). The short-lived photolyzed state was shown to be structurally complex in so far as the photolyzed CO migrates inside the protein matrix while in state M, before escaping into the bulk or eventually recombining in what has been called *geminate rebinding*. Moreover, room-temperature, time-resolved Laue crystallography has pointed out that during its pathway inside the protein matrix, CO momentarily occupies small internal cavities, previously shown to be able to bind one atom of xenon.

If the data are forced into the four-step scheme, as in Eq. (6), the biochemical system (e.g. HbCO) starts in the potential-well HbX, in which the ligand (e.g. CO) is bound to the heme Fe. When the laser flash breaks the Fe–ligand bond, the system rapidly moves (relaxation time about 3 ps) to well P where the ligand molecule is in the heme pocket. At this point, the three gaseous ligands (O_2, CO, NO) behave differently. On a picosecond to nanosecond timescale, all ligands except CO show a substantial fraction of geminate rebinding; thus, at approximately 100 ps, HbX is the predominant form after photolysis of NO, while M predominates after photolysis of CO, and O_2 is more evenly distributed between HbX and M. Further evolution of species M is orders of magnitude slower (nanoseconds). Although the overall rates are much the same for O_2, CO, and NO, the amplitude of this second phase is different for the three ligands, the value being smallest for NO (since for this ligand, the picosecond rebinding process is dominant and, accordingly, little ligand is left to react in the nanosecond geminate process). Without discussing the chemistry yielding the difference between NO and CO, we note that the rate-limiting step in bimolecular NO binding is the diffusion of ligand into and within the protein since reactivity with the heme is so high that almost every NO molecule that approaches will bind; on the other hand, assuming a similar rate of diffusion, only a small fraction of the photodissociated CO rebinds in a geminate mode.

6 Mechanisms of Cooperative Binding and Allostery

Cooperative binding of O_2 by Hb is the classical example of what is referred to as an *allosteric phenomenon*, which defines the concept that the binding of a ligand at one site on a macromolecule affects the binding of a second ligand at another site via a conformational change or an allosteric transition. In allosteric binding, therefore, the uptake of one ligand (e.g. one O_2) by an oligomeric protein (e.g. tetrameric Hb) influences the ligand affinities of the remaining unfilled binding sites; in human Hb, the fourth O_2 molecule binds with a 200-fold higher affinity than the first one. Likewise, binding of oxygen at the heme is affected by the association of other ligands (e.g. H^+ or BPG) at different nonheme sites on the protein.

The detailed structural bases of cooperativity continue to elicit contrasting views. A number of models have been proposed that in various ways hybridize the two classical theories developed during the 1960s by J. Monod, J. Wyman, and J. P. Changeux (often referred to as the *MWC model*) and by D. E. Koshland, G. Nemethy, and D. Filmer (often referred to as the *KNF model*). Actually, the MWC and KNF models are the limiting cases of a more general

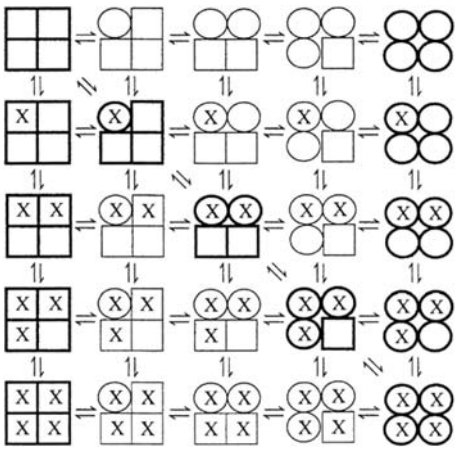

Fig. 5 General representation of ligand (X) binding and ligand-linked conformational changes, illustrating the contrasting ways of generating cooperativity according to the sequential or KNF model (indicated by the diagonal in boldface) and the concerted or MWC model (represented by the equilibria between the two extreme columns, all intermediates being absent or negligible or undetectable). Each subunit within the tetramer can have only two distinct tertiary conformations, one (circle) with high affinity for heme ligand or R state and the other (square) with low affinity or T state. In the sequential model, heme–ligand binding induces an isomerization in each ligated subunit, which assumes the high-affinity conformation; the unligated subunits retain their low-affinity structure, even though the contacts with the ligated neighbor(s) are altered, thereby increasing the tendency to switch to the high-affinity state. An important property of the sequential model is that the extent of binding and structural changes determined at different ligand concentrations must be identical. In the concerted model, the tetramer is postulated to exist in only two quaternary states (R and T) in which all four protomers are in either one of two conformations (squares or circles). The quaternary states are in equilibrium with each other at every degree of saturation: the fully ligated tetramer is predominantly in the R state, whereas the T state is favored in the fully unligated species. The conformational transition of the subunits (between square and circle) is concerted, involving all protomers within the tetramer, with the result that the molecule conserves structural symmetry. This corresponds to an extreme case of cooperativity among the protomers, characterized by a virtual absence of the hybrid states (i.e. coexistence of circles and squares in one tetramer), which, on the contrary, are present in the sequential model. The sequential model can account for both positive and negative cooperativity, whereas the concerted model can accommodate only positive cooperativity.

scheme that can be depicted as a square array (Fig. 5) such that the number of allosteric conformers (in a row) equals the number of binding species in a given column, including the unligated or reference species. The entire system of chemical species (macromolecule and small ligand, X) is considered to be in equilibrium. The reactions within the array are described by appropriate mass action laws and involve the addition of iX ligands and a change of the cth molecular conformation, as follows:

$$\mathrm{Hb}_{00} + i\mathrm{X} \rightarrow \mathrm{Hb}_c\mathrm{X}_i \qquad (7)$$

For tetrameric human Hb (and, in general, for all vertebrate Hbs), each conformation can bind a maximum number of four X ligands, and, accordingly, there are 4 + 1 conformations in the square array (i.e. 4 ligated and 1 unligated). The MWC two-state model postulates that the two extreme allosteric states (first and fifth columns of Fig. 5) are the dominant ones; the KNF model envisages the species along one of the diagonals to be preferentially populated (see Fig. 5). Since high cooperativity is always associated with reduction of the population of the intermediate species (partially oxygenated Hb intermediates never exceed a few percent of total Hb), these two alternative mechanisms are, to all intents and purposes, indistinguishable by analysis of O_2 binding data, which, in fact, are often described by Eq. (5), the empirical Hill equation.

6.1 Concerted Model of Allostery

Among the various approaches emphasizing the role of subunit–subunit interactions, the concerted (or MWC) model has a very simple algebraic description. Its basic idea is that the intrinsic O_2 affinity at the heme is controlled by the quaternary structure of the protein rather than by the number of ligand molecules already bound to the tetramer. This model assumes that there are only two different quaternary structures: one with fewer and weaker interactions among the subunits (fully ligated Hb or relaxed R) and the other with more and stronger bonds between the subunits (deoxy Hb or tense T). In the R or high-affinity state, Hb displays the ability to bind O_2 tightly; this functional property is damped but not abolished in the T or low-affinity state. On the other hand, the T state interacts more strongly than the R state with heterotropic ligands (H^+, Cl^-, BPG, and CO_2). The transition from T to R is an all-or-none process: the symmetry of the molecule is conserved (i.e. hybrid RT states cannot exist) and the O_2 affinity of all subunits in a tetramer would be either equally low or equally high. Oxygenation of the subunits within each quaternary state is associated with small local structural changes such that the T state, with one O_2 molecule bound, and the R state, with three O_2 molecules bound, will have conformations that are locally different from those of the unligated T and fully ligated R respectively. The assumption of the concerted model in its original formulation is that these local structural differences are not transmitted to neighboring subunits and thus do not alter their O_2 affinity; rather, they affect the relative stability of the R and T quaternary states. Therefore, although binding to the four hemes in each of the T and R states is itself noncooperative, the shift in the T ↔ R equilibrium gives rise to a sigmoid binding curve.

The MWC model allows one to describe the ligand binding by Hb at equilibrium, with only three independent parameters: the two microscopic dissociation constants of O_2, with the protein in low-affinity (K_T) and high-affinity (K_R) states, often reported in terms of a ratio $c(= K_T/K_R)$, and the concentration ratio of the two quaternary states in the absence of ligand $L_0 = [T]_0/[R]_0$. The total free energy of cooperativity, δG^0, can be calculated from the binding of the last and first ligand molecules, which is approximately equal to $\delta G^0 = 2.303$ RT $\log(K_R/K_T)$ (i.e. the difference in free energy of binding O_2 to the R state and the T state).

Values of parameters illustrating the interpretation of the cooperative effects in terms of the MWC model are reported in Table 3. The effect of heterotropic ligands (H^+, Cl^-, BPG, and CO_2) is also described in terms of the concerted allosteric model as follows: addition of BPG (which binds to the Hb tetramer in a 1 : 1 ratio) shifts the equilibrium in favor of the T structure, with an increase of L_0; the same effect is obtained by lowering pH.

The original MWC model assumed K_T and K_R to be intrinsic parameters independent of heterotropic ligands. However, detailed functional O_2-binding data, coupled with the structural knowledge gained from X-ray crystallography and spectroscopic studies indicate that, in human HbA_0, K_T varies over a wide range as a function of H^+, Cl^-, CO_2, and BPG concentration, whereas K_R is little affected by heterotropic ligands (except by protons, and only below pH 7). Thus, the original MWC model had to be modified.

In general, each quaternary state (T or R) may differ in tertiary conformation of the subunits (t or r), depending on the lower or higher intrinsic heme–ligand affinity of the chains (squares and circles in Fig. 5), and therefore, the system may be described in terms of (at least) four alternative states denoted as follows: Rr and Rt, for conformations at high and low affinity for heme ligands of the subunits within the R structure, and Tt and Tr, for low and high-affinity conformations of chains within the T structure.

6.2 Structural Changes That Accompany Oxygen Binding

The T state is assumed to have quaternary interactions between subunits typical of deoxy Hb, while the R state is believed to have those of oxy Hb (as determined by crystal X-ray diffraction data). In deoxy Hb, the Fe is displaced from the plane of the porphyrin nitrogens toward the proximal histidine by 0.06 nm. There are two reasons for this: (1) the larger radius of the Fe in its high spin state; and (2) steric repulsions between the N_ε of the proximal histidine and the porphyrin nitrogens. Moreover, the heme group is domed, that is, the Fe-bound nitrogens of the porphyrins are out of the plane of the carbon atoms of the heme (see Fig. 2), by 0.016 nm in the α-subunit and by 0.01 nm in the β-subunit. Upon single-subunit oxygenation, the Fe atom barely moves toward the porphyrin, which remains domed within the T state; this prevents the heme from adopting the optimal geometry with the bound oxygen (i.e. planar with the Fe in "the plane"), thus producing strain. Relief of strain at the active site requires the Fe to move toward the porphyrin plane, accompanied by a displacement of the proximal histidine together with the helix F and the corner FG (the so-called *allosteric core*). In particular, in both subunits, there is a shift of helix F across the heme plane by about 0.1 nm, accompanied by a rotation; the overall effect is to permit a reorientation of the proximal histidine so that the Fe can move toward the heme plane.

To understand the connection between the foregoing tertiary modifications and quaternary structural changes, the Hb tetramer can be thought of as a pair of dimers, $\alpha^1\beta^1$ and $\alpha^2\beta^2$, whose relative orientation has to alter; in fact, the contacts $\alpha^1\beta^1$ (and $\alpha^2\beta^2$) remain rigid. The tertiary changes make these dimers a misfit in the quaternary T structure, with consequent promotion of the T-to-R transition. The transition consists of a 12 to 15° rotation of the dimer $\alpha^1\beta^1$ relative to the dimer $\alpha^2\beta^2$ and a translation of one dimer relative to the other by 0.08 nm (Fig. 6). Helix C β_2, which is in contact with the corner FG α_1, forms a ball-and-socket joint around which the dimers turn, while helix C α_1 slides relative to FG β_2. Together, the two contacts form a two-way switch that ensures that the $\alpha\beta$ dimers click back and forth between the two stable quaternary positions (T and R).

All heterotropic ligands at physiological pH lower the O_2 affinity by forming additional hydrogen bonds that specifically stabilize the T structure. BPG binds in a cavity on the dyad axis of the tetramer, limited by the amino termini and helices H of the β-subunits; it forms hydrogen bonds with ValNA1(1), HisNA2(2), LysEF6(82), and HisH21(143). The functionally relevant protons (i.e. those released upon oxygen uptake and on the basis of the Bohr effect; see also Fig. 4b) are discharged in deionized solution only

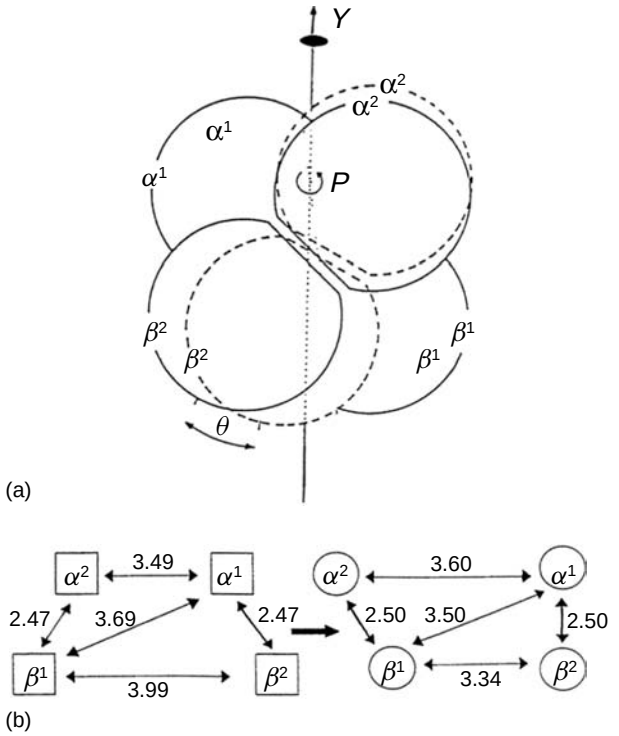

Fig. 6 Schematic diagram illustrating the change in quaternary structure (a) and in distances (nm) between heme groups (b) that accompanies ligand binding to Hb. Bold symbols refer to the deoxy derivative and light symbols to ligated Hb. Both structures have a dyad axis (*Y*) relating the $\alpha^1\beta^1$ dimer to the $\alpha^2\beta^2$ dimer, and the molecular contacts between α^1 and β^1 (as well as those between α^2 and β^2) change very little; accordingly, the positions of ligated and unligated $\alpha^1\beta^1$ dimers have been superimposed. The main differences between the two conformational states, in going from the deoxy to the ligated derivative, are (1) a rotation of about 13.5° (θ) about a pivot (*P*) of the $\alpha^2\beta^2$ dimer relative to the $\alpha^1\beta^1$ dimer so that the two β chains are about 0.6 nm further apart in the former than in the latter; and (2) a shifting of the $\alpha^2\beta^2$ dimer along the *P* axis by about 0.1 nm into the page. Major conformational change in the T-to-R transition involves the $\alpha^1\beta^2$ (and $\alpha^2\beta^1$) subunit interfaces.

from HisHC3(146)β, a residue that forms a hydrogen bond with AspFG1(94)β in the T structure. In the presence of Cl^- and/or BPG, other residues [e.g. ValNA1(1)β and LysEF6(82)β] contribute additional Bohr protons.

6.3 Ligand-binding Pathways

Hemoglobins have a binding pocket at the heme that protects Fe(II) from the solvent. This feature has biological importance

because, in water, the heme Fe(II) can oxidize into the inactive ferric state. The stability of the Fe(II)O_2 complex in Hb is controlled by the presence of a hydrogen bond with the distal histidine (located at E7; see Table 1), as in Fig. 2. This crucial interaction, in fact, slows the rate of O_2 dissociation from microseconds (which has been measured in sperm-whale Mb mutant with valine at E7) to many milliseconds, and simultaneously increases the O_2 affinity to physiologically meaningful values; moreover, it reduces the rate of autoxidation of the heme Fe(II), which, otherwise, would occur much too rapidly since oxidation happens in the fraction of Hb that is deoxygenated at any one moment (the larger the fraction of molecules, the faster the Hb autoxidizes).

Examination of the X-ray structures of many hemoproteins shows that the protein environment of the heme, if rigid, would prevent the entrance and the exit of even small gaseous ligands (e.g. O_2, CO). Empirical energy function calculations, based on the crystallographic coordinates, have shown that possible pathways in the neighborhood of the heme pocket would have barriers of the order of 400 kJ mol^{-1} in a rigid protein; such energy barriers would lead to escape times (of ligand from the heme pocket) of the order of 10^{60} seconds at room temperature, infinitely long on any timescale. Thus, protein fluctuations are required for physiological function. The dynamics of the heme pocket residues are revealed by the results of NMR and other studies, showing that PheCD1(43) and PheCD4(46), which wedge the prosthetic group into its pocket and are packed tightly between the heme and the distal helix E, flip over at rates faster than 10^4 s^{-1}; this is possible only if the entire heme pocket breathes fast.

Dihedral rotations of key side chains, explored by molecular dynamics, indicate that energy barriers ($\approx$40 kJ mol^{-1}) would be present in a fluctuating protein; these barriers are compatible with those estimated from rebinding studies for HbCO after photolysis. This dilemma has been addressed by (1) examining X-ray structures with small (e.g. CO) and bulky (e.g. alkylisocyanides) ligands bound to the heme Fe; (2) temperature-dependent X-ray diffraction; (3) molecular dynamics simulation; (4) time-resolved crystallographic experiment on photolyzed Mb crystals; and (5)site-directed and random mutagenesis. Most of these investigations were carried out initially on Mb. Nine clusters of residues have been suggested as possible likely pathways for ligand escape from the interior of the protein. The most direct pathway is constituted by residues ArgCD3(45), HisE7(64), ThrE10(67), and ValE11(68), between the CD loop, the helix E, and the heme; recent analysis of kinetic data on approximately 100 Mb mutants indicates that the ligand escape route involves, by and large, the so-called histidine gate.

7
Assembly of Globin Monomers

The basic globin fold, which has been conserved throughout evolution, leads to a number of possible quaternary structures, from dimers to large assemblies. Vertebrate Hbs are tetrameric, but in the animal kingdom, globin chains can assemble into a myriad of different quaternary structures from stable dimers to very large macromolecular aggregates, like the extracellular invertebrate Hbs with a molecular mass of the order of 10^6 g mol^{-1} or higher.

Subunit interactions in assembly occur in stages and have been described as a series of coupled equilibria that are often ligand dependent. In the case of tetrameric Hb, dissociation is described in terms of the scheme:

$$\text{tetramer} \rightleftharpoons \text{dimer}$$

$$\text{dimer} \rightleftharpoons \text{monomer}$$

When dissociation is limited to the dimer stage, only symmetrical splitting is observed (i.e. $\alpha\beta$ dimers and not α_2 and β_2 dimers), reflecting the nature of molecular contacts: namely, the dissociation between like chains ($\alpha\alpha$ or $\beta\beta$) are considerably less extensive than those between unlike chains (see Table 2), of which there are two kinds. Cleavage occurs only along the $\alpha^1\beta^2$ interface (yielding $\alpha^1\beta^1$ and $\alpha^2\beta^2$ dimers) and is different for the two quaternary states, that is, it is ligand linked. In fact, the tetramer–dimer equilibrium constant of human Hb, at physiological ionic strength, is about 10^{-6} M (in terms of tetramer) for the oxy derivative and about 10^{-11} M for deoxy Hb. A variability of the tetramer–dimer dissociation constant has been reported from comparative studies of a number of Hbs from different vertebrate species. Further dissociation of the dimer to produce α- and β-chains is negligible under physiological conditions. The isolated subunit can exist as dimers (in the case of the α-chains) and even tetramers (in the case of the β-chains). It is generally assumed that dissociation of these oligomeric subunits into monomers must occur before they can combine to form $\alpha\beta$ dimers, which then aggregate to produce the fully tetrameric Hb. The overall scheme describing formation of cooperative Hb tetramer is as follows:

$$\begin{array}{ccccc} & & 2\alpha_2 & \beta_4 & \\ & & \Downarrow\Uparrow & \Downarrow\Uparrow & \\ 4\alpha + \beta_4 & \longrightarrow & 4(\alpha\beta) & \rightleftharpoons & 2(\alpha_2\beta_2) \end{array} \qquad (8)$$

Data on the association–dissociation reaction for human oxy Hb are summarized in Table 4.

7.1 Variations Across Species

Although Hbs from closely related species behave alike, those of distantly related ones may show marked differences. For example, fish Hbs appear to be very stable toward dissociation, even in the ligated form. Association of tetramers upon deoxygenation appears to be widespread among birds, amphibians, and some reptiles, in which species the association seems to have considerable physiological importance. Such aggregation, in fact, contributes to the Bohr effect to the extent

Tab. 4 Subunit dissociation constants (K, in M/heme) for the isolated α and β chains, and for the $\alpha\beta$ dimer and the $\alpha_2\beta_2$ tetramer, in all cases as oxygenated derivatives. The table also reports the kinetic association (k_{ass}) and dissociation (k_{diss}) rate constants. All data in 0.1 M Tris-HCl$^+$ 0.1 M NaCl$^+$ 1 mM Na_2 EDTA, pH 7.3 to 7.5 at 21.5 to 25 °C.

Reaction	*K*	k_{ass} (M^{-1} s^{-1})	k_{diss} (s^{-1})
$\alpha_2 \rightleftharpoons 2\alpha$	1.1×10^{-4} M	9.1×10^5	–
$\beta_4 \rightleftharpoons 4\beta$	4.8×10^{-16} M^3	–	0.05
$\alpha\beta \rightleftharpoons \alpha + \beta$	4.4×10^{-13} M	2.0×10^6	8.9×10^{-7}
$\alpha_2\beta_2 \rightleftharpoons 2(\alpha\beta)$	7.1×10^{-7} M	2.8×10^7	20

that tetramer–tetramer interaction is pH dependent. The best-studied system is that of the adult bullfrog *Rana catesbeiana*, in which two tetrameric components, B and C, associate upon deoxygenation to form aggregates that appear to be trimers of tetramers (of the form BC_2), showing an extremely low oxygen affinity, a Hill coefficient that reaches values exceeding 4, and a large contribution to the overall alkaline Bohr effect. A variant of human Hb (named Hb Olympia), in which ValB2(20) β is substituted by a methionyl residue, has oxygenation properties somewhat like those of the frog Hb (i.e. self-association in the deoxy state, cooperativity values as high as 4, and aggregation altering the Bohr effect).

In contrast to vertebrate Hbs, the invertebrate heme proteins exist as monomers, dimers, tetramers, and higher polymeric forms, and their aggregation state often depends on the heme ligand. Thus, the oligomeric state of Hbs from the sea cucumber *Caudina arenicola* varies from monomer (met Hb) to dimers (oxy and cyanomet Hb) to tetramers (deoxy Hb). The extracellular Hb of the earthworm *Lumbricus terrestris* is much larger, forming a complex of more than 200 globin monomers. An additional difference between vertebrate and invertebrate Hbs is in the quaternary structure. Thus, some tetrameric invertebrate Hbs (e.g. those of the arcid clam *Scapharca inaequivalvis* and the marine worm *Urechis caupo*) are assembled inside out. In fact, both *Scapharca* and *Urechis* Hbs have intersubunit contacts that involve the helices E, while the helices G and H (essential intersubunit contacts in vertebrate Hbs) face the solvent in that they are exposed on the outside of the molecule. However, the helix F of *Scapharca* Hb forms an important intersubunit interface, while the helix F of *Urechis* Hb is not involved in any intersubunit contact. Accordingly, *Scapharca* Hb shows cooperativity, while that from *Urechis* does not.

7.2 Subunit Interaction and Cooperativity

When a heme ligand is bound by Hb, the binding site becomes subjected to forces not present before. The strains thus set are propagated to the boundary between subunits – note that tetramer dissociation appears to take place far more readily with ligated Hb, relative to the deoxy state. This phenomenon of ligand-linked dissociation is extremely marked with Hbs from animal species, such as the lamprey (*Lampetra planerii*). In the latter case, cooperativity in the reaction with O_2 (as indicated by a value of Hill coefficient higher than 1) is exclusively dependent on oxygen-linked reversible dissociation into subunits. In the absence of O_2, lamprey Hb forms dimers (and tetramers) with low oxygen affinity; uptake of oxygen induces dissociation into monomers with higher ligand (e.g. O_2) affinity. Thus, the fraction of dimers decreases with rising oxygen saturation and, conversely, the oxygen affinity drops with increasing Hb concentration. From these observations, it follows that changes in the free energy of monomer association with heme-ligand binding must be expected. The major manifestations of the relation between cooperativity among the hemes and subunit interactions within tetrameric Hbs are the following: (1) dissociated $\alpha\beta$ dimers bind oxygen with high affinity, which is almost equal to the average affinity of the R-state Hb and the isolated α- and β-subunits, but without any detectable cooperativity; and (2) assembly of deoxy dimers into tetramers produces a large reduction in overall affinity for O_2. Since the tetramer binds the four heme ligands with

successively increasing affinity constants, the number of ligated hemes must control the interaction energy at the dimer–dimer interface (i.e. the extent to which free energy is transduced, at each binding step, into structural and energy alterations at the level of interdimer contacts). The reaction scheme that defines the free-energy changes for the O_2 binding and dimerization processes can be depicted as follows:

$$
\begin{tikzcd}
 & (\alpha\beta) \arrow[d, "\Delta G_{21}"'] & + & (\alpha\beta) \arrow[rr, "{}^{0}\Delta G_2"] & & (\alpha_2\beta_2) \arrow[d] & \\
 & (\alpha\beta)X \arrow[d, "\Delta G_{21}"'] & + & (\alpha\beta) \arrow[rr, "{}^{1}\Delta G_2"] & & (\alpha_2\beta_2)X \arrow[d] & \\
\Delta G_2 & (\alpha\beta)X \arrow[d, "\Delta G_{22}"'] & + & (\alpha\beta)X \arrow[rr, "{}^{2}\Delta G_2"] & & (\alpha_2\beta_2)X_2 \arrow[d] & \Delta G_4 \\
 & (\alpha\beta)X_2 \arrow[d, "\Delta G_{22}"'] & + & (\alpha\beta)X \arrow[rr, "{}^{3}\Delta G_2"] & & (\alpha_2\beta_2)X_3 \arrow[d] & \\
 & (\alpha\beta)X_2 & + & (\alpha\beta)X_2 \arrow[rr, "{}^{4}\Delta G_2"] & & (\alpha_2\beta_2)X_4 \arrow[uu, bend right, "\Delta G_{41}"', "\Delta G_{42}", "\Delta G_{43}", "\Delta G_{44}"] &
\end{tikzcd}
\qquad (9)
$$

The values of the free-energy changes of these coupled reactions are reported in Table 5. The difference in O_2-binding free energy between the tetramer and the dimer is defined as the *total regulatory energy*, which is gained in native tetramer. From Eq. (9), it is seen that this difference $\delta G_4 - \delta G_2$ is equivalent to the difference in association energy of the ligated and unligated dimer forms, ${}^{4}\Delta G_2 - {}^{0}\Delta G_2$. Experimentally, these latter terms can be evaluated by careful association-state measurements of deoxy and oxy Hb. The regulatory energy of human Hb, calculated from data of Table 5, corresponds to 26.7 kJ mol^{-1}, that is,

$$(2\Delta G_{21} + 2\Delta G_{22}) - (\Delta G_{41} + \Delta G_{42} + \Delta G_{43} + \Delta G_{44}) \qquad (10)$$

8 Reactions with Nitric Oxide

Nitric oxide (NO) has been widely used as a ligand of the ferrous heme iron, similarly to O_2 and CO; over the years, comparison between the three gases has provided a wealth of information on the structure and function of Hb. However, the chemistry of NO is much more complex than that of other ligands of the ferrous heme iron, and interest has risen considerably since it was discovered that NO is a universal second messenger regulating circulation,

Tab. 5 Intrinsic free-energy changes (kJ mol^{-1}) in the cooperative mechanism of human hemoglobin: symbols based on Eq. (9).

Binding of heme ligand to		*Dimer–tetramer assembly reaction*
Dimers	*Tetramers*	
$\Delta G_{21} - 35.1$	$\Delta G_{41} - 22.8$	${}^{0}\Delta G_2 - 69.2$
$\Delta G_{22} - 35.1$	$\Delta G_{42} - 22.1$	${}^{1}\Delta G_2 - 47.1$
	$\Delta G_{43} - 32.6$	${}^{2}\Delta G_2 - 35.1$
	$\Delta G_{44} - 36.2$	${}^{3}\Delta G_2 - 30.9$
		${}^{4}\Delta G_2 - 33.2$

brain alert and defense against invaders (see *Nitric Oxide* in the *Encyclopedia*).

8.1 Reactions of NO with Various Hb Derivatives

Nitric oxide can bind to the ferrous heme iron of all known hemoproteins, yielding a simple adduct characterized by a very high stability constant ($K \sim 10^9\ M^{-1}$). However, NO can react not only with deoxy Hb (or deoxy Mb) but also with the oxygenated and the met derivatives, yielding a complex network of reactions that are summarized in Fig. 7.

The reaction of NO with oxy Hb, forming met Hb and nitrate, is a very rapid process ($k_{on} \sim 6–9 \times 10^7\ M^{-1}\ s^{-1}$ at 20 °C) competing with the straight binding to deoxy Hb ($k_{on} \sim 2 \times 10^7\ M^{-1}\ s^{-1}$ at 20 °C). The reaction between NO and HbO_2 with formation of nitrate is considered the main route to quench NO that is continuously produced in circulating blood through the physiological metabolic pathway (see Sect. 8.2). Moreover, the same reaction occurring in striated muscles and in the heart with oxy myoglobin (MbO_2) has been considered an important defense mechanism against the inhibitory effects of NO on cytochrome *c* oxidase (and thus on cellular respiration). This reaction of MbO_2 with NO is intrinsically very rapid and is made efficient by the high concentration of the pigment in the red muscles (0.3–0.6 mM). The cycle reproduced in Fig. 7 also includes the reduction of the met Hb (or met Mb), catalyzed by specific reductases.

However, contrary to O_2 and CO, NO also binds the oxidized form of Hb to produce an iron–nitrosyl complex HbFe(III)NO ($k_{on} \sim 4 \times 10^3\ M^{-1}\ s^{-1}$ and $k_{off} \sim 1 \times 10^3\ s^{-1}$), not indicated in Fig. 7. In the presence of an excess of NO, this complex undergoes reductive nitrosylation to finally generate the reduced nitrosyl adduct and nitrite:

$$\begin{aligned} &\mathrm{HbFe(III)} + \mathrm{NO} \\ &\rightleftharpoons [\mathrm{HbFe(III)NO} \Leftrightarrow \mathrm{HbFe(II)NO^+}] \\ &\xrightarrow[\mathrm{H_2O}]{\mathrm{NO}} \mathrm{HbFe(II)NO} + \mathrm{NO_2^-} + 2\mathrm{H^+} \end{aligned} \tag{11}$$

It should be recalled that the inorganic reaction of NO with free O_2 in solution ($k \sim 6\ M^{-1}\ s^{-1}$) is too slow to compete

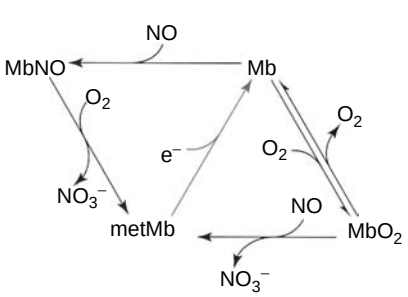

Fig. 7 The reactions of myoglobin with O_2 and NO. The primary physiological function involves the reversible binding of O_2 to deoxymyoglobin (Mb) to yield oxymyoglobin (MbO_2), which facilitates the transport of O_2 from the periphery of the cell to the mitochondria for use in respiration. MbO_2 reacts rapidly (and irreversibly) with nitric oxide (NO) to yield nitrate and ferric myoglobin (met Mb), thereby quenching free NO that might otherwise inhibit cytochrome *c* oxidase. Met Mb is reduced to Mb by met Mb reductase (e^- arrow); also, Mb binds rapidly with NO, yielding nitrosylmyoglobin (MbNO). This pathway might be of greater significance in cellular compartments with low O_2 concentrations, for example, in the immediate environment of mitochondria. In the presence of O_2, MbNO can be converted back to met Mb much more rapidly than thermal dissociation of NO, yielding nitrate again.

with the direct bimolecular reactions of NO with deoxy or oxy Hb. However, alternative reactions may come into play under different conditions, such as the paraphysiological ones where NO concentration is 10^3 to 10^6 times lower than that of O_2 and Hb. In particular, when oxygenated hemes in the quaternary R state of Hb are around $\geq$95% and when NO is present at concentrations lower than the level of vacant hemes, the unoccupied binding sites will be extremely reactive toward NO. Under these conditions, a minute amount of NO bound to the heme iron may follow an alternative chemistry, as outlined below.

It has been proposed that NO reversibly reacts with CysF9(93)β only when Hb is in the quaternary R state (e.g. oxy Hb) but not in the quaternary deoxy state; in other words, this reaction of NO to Hb should be incorporated into the MWC two-state model. If this is the case, NO would represent a third gaseous molecule (in addition to O_2 and CO_2) carried by Hb under normal conditions; this view, however, is controversial.

8.2 Physiological Implications

Nitric oxide is continuously produced in the endothelial cells through the enzymatic degradation of L-arginine by the enzyme NO synthase. This gas acts as a signal transduction molecule that mediates a variety of physiological responses, including dilatation of blood vessels, achieved by activation of soluble guanylate cyclase (a heme-containing enzyme), with an increase in the intracellular cyclic guanosine-monophosphate concentration and thus, relaxation of vascular smooth muscles. The supposedly rapid conversion of NO to biologically inactive metabolites in blood (e.g. nitrate) formed the rationale for inhalation NO therapy, because the short half-life (~2 ms) of NO in blood should confine its effects to the pulmonary circulation. Erythrocytes may be considered to be a major sink for NO by virtue of the quick oxidation reaction of NO with oxy Hb (see Sect. 8.1) to produce met Hb and nitrate, due to (1) the speed of this bimolecular reaction (half-life about 1 μs, according to extrapolation of the *in vitro* data); and (2) the high concentration of oxy Hb in blood, which is in the millimolar range (2 mM in tetramers), to be compared with endothelium-derived local-level of NO that reaches only nanomolar concentrations.

Nitric oxide was also shown to be transported bound to sulfhydryls groups through a process called *nitrosation*, occurring by reaction of the nitrosonium cation and the deprotonated sulfhydryls:

$$R{-}S^- + NO^+ \rightleftharpoons R-SNO \qquad (12)$$

This reaction occurs quite easily not only with reduced but also with oxidized glutathione, which is at high concentration in the erythrocytes (~2.5 mM). The same reaction was shown to occur with the CysF9(93)β of human HbA, the rate of formation and dissociation of β93-SNO being dependent on the quaternary state of Hb (whether R or T). According to this hypothesis, NO binds cooperatively to the minor (~1%) population of deoxygenated heme in oxygenated erythrocytes present in the arterial circulation, forming iron–nitrosyl–Hb; the NO group is then transferred (through a one-electron oxidation) to CysF9(93)β to produce its S-nitro derivative (<50 nM *in vivo*), which dissociates NO upon deoxygenation in the peripheral tissues and elicits vasorelaxation *via* transnitrosation reactions to low molecular

mass thiols (such as glutathione, yielding *S*-nitroglutathione) or high molecular mass sulfhydryls (specifically, the anion exchanger or band 3 in the erythrocyte membrane).

This hypothesis is opposed on the basis of the argument that two phenomena would combine to make the reaction of NO with Hb 10^6 times slower *in vivo* than *in vitro*: (1) adjacent to the endothelium, the blood flow creates an erythrocyte-free layer, whose dimension (proportional to the vessel diameter) modulates the amount of NO that diffuses into smooth muscle layer; and (2) the rate of NO scavenging by erythrocyte is 500- to 1000-fold smaller than that of an equivalent concentration of cell-free Hb, possibly due to the existence of multiple diffusional barriers for NO surrounding red blood cells (unstirred plasma layer, laminar-flowing blood, and erythrocyte membrane matrix). Therefore, the local concentration of NO (~400 nM) produced by endothelial cells is in large excess of that required to activate guanylate cyclase (the level for its half-maximum activity being ~20 nM), thus ensuring that sufficient NO reaches its target enzyme and induces smooth muscle relaxation.

Support and dissent to both models mostly depend on the wide spectrum of analytical approaches employed to measure NO and nitrosothiols, methods that often yield divergent values and thus lead to conclusions with many aspects being disputed. In particular, the main dissent rests on the similarity or difference in levels of *S*-nitrosothiols and HbFe(II)NO, between arterial and venous blood: thus, in contrast with proponents of the model assigning an essential physiological role to – SNO transport, opponents report no artery-to-vein gradient.

9
The Globin Gene and its Biosynthesis

In humans, the initial ζ (α-like) and ε (β-like) embryonic subunits, synthesized in the yolk sac during the third through eighth week of gestation, are gradually replaced by α- and γ-chains so that the original Hb Gower-1 is replaced, first by Hb Gower-2 and Hb Portland and finally by HbF (see Sect. 3). At about seven weeks of gestation, when liver and spleen become the major sites of erythropoiesis, HbF accounts for about 50% of total Hb. Synthesis of β-chains, minor at first, begins to accelerate around the thirtieth to fortieth weeks of pregnancy (see Fig. 8a). Within eight weeks after birth, when erythrocyte production has shifted from liver and spleen to bone marrow, mainly α- and β-chains are synthesized. At 6 months after birth, the adult $\alpha_2\beta_2$ form has essentially taken over, along with the minor $\alpha_2\delta_2$ component (Fig. 8B).

Each of these Hbs, synthesized during the development from embryo to fetus to infant (Fig. 8b), consists of two α-type and two β-type subunits. The α-type (α, ζ) and β-type (β, γ, ε, δ) families of human genes are clustered, respectively, near the telomere of the short arm of chromosome 16 and on the short arm of chromosome 11. Each globin gene in a cluster, like most other eukaryotic genes, possesses a mosaic structure in that it contains three coding blocks (exons) separated by two intervening sequences (introns) of DNA (Fig. 9). It is worth noting that the gene encoding neuroglobin (located in chromosome 14), as well as that encoding cytoglobin (located on chromosome 17), have three introns like the gene of plant Hb (see also Sect. 1). Exons, introns, and the flanking regions (whose directions are indicated by 5′ or upstream and 3′ or downstream, relative

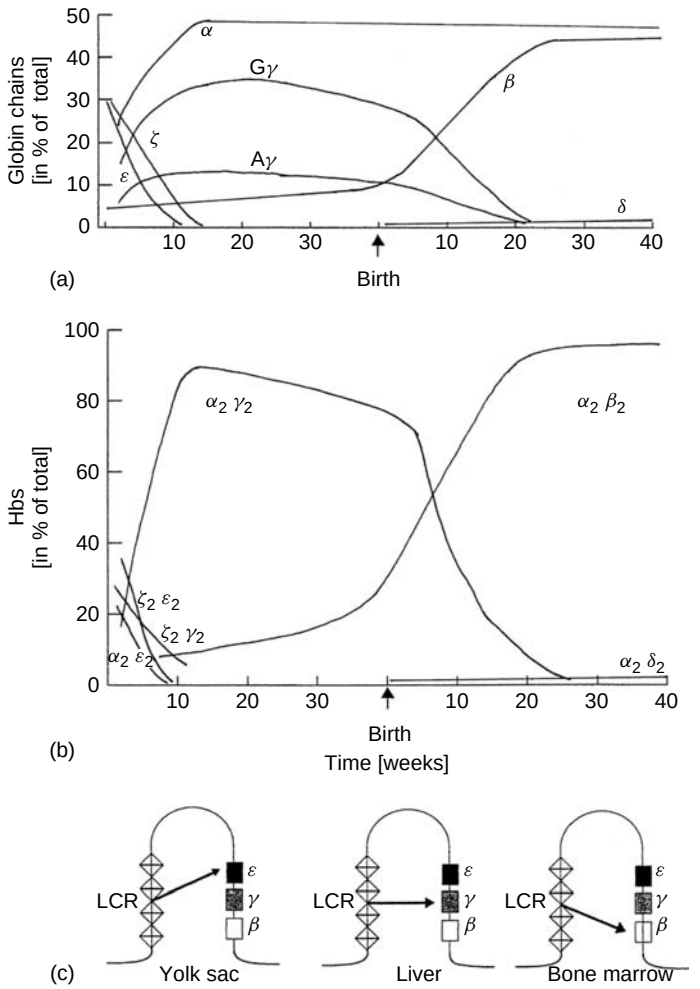

Fig. 8 Relative amounts of globin subunits (a) and Hb tetramers (b) at different stages in the development of the human embryo, fetus, and infant. Modifications in the pattern of globin gene expression are controlled by a regulatory element, located at the 5′ flanking end and called *locus control region* or LCR. In the absence of an interaction with LCR, individual globin genes are either inactive or very poorly expressed. In a series of interactions of LCR with downstream genes of the β-type gene cluster (c), the mutually exclusive binding at each development stage is indicated by the arrows.

to transcription start site of each gene) are transcribed into nuclear mRNA precursors (pre-mRNA). While pre-mRNA is still in the cell nucleus, the intervening sequences are excised and the terminals of the exonic sequences religated to yield the mature mRNA (a process termed *splicing*). Lastly, mRNA is translated into globin. Conserved nucleotide sequences, important for gene function and regulation, are found just

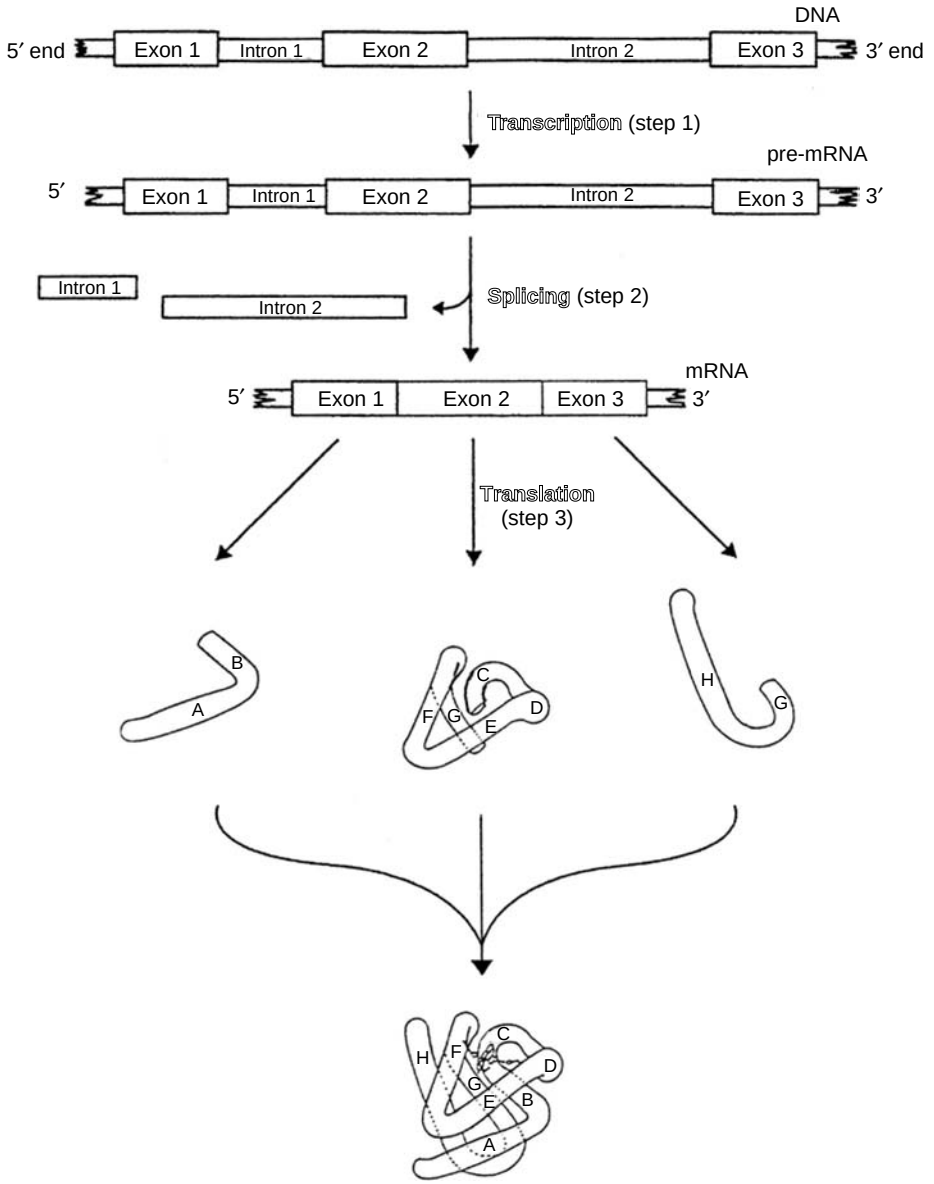

Fig. 9 The mosaic or exon–intron structure of a globin gene encoding a typical polypeptide chain of mammalian Hb or Mb. From the gene, a primary transcript or pre-mRNA is synthesized (step 1); the intervening sequences or introns are then excised (step 2) by splicing to yield the mature mRNA; this is translated (step 3) into the polypeptide chain, which spontaneously adopts the typical globin fold. Three globin "domains" corresponding to the three exonic transcripts are schematically illustrated.

upstream of the mRNA coding blocks (i.e. at the 5′ end), at the exon–intron boundaries, and at the 3′ end of the mRNA coding sequences. Moreover, expression of individual genes in a cluster, as development from embryo to newborn proceeds, is dependent on cis-acting elements of gene switching as well as trans-acting proteins.

Cis-linked regulatory elements include promoters, enhancers, silencers, insulators, boundary elements, and the locus control region or LCR (see Fig. 8c). The human β-globin LCR – a region of DNA located from about 6 to 22 kb upstream of ε-globin gene – is composed of five domains (called *hypersensitive sites* or HS) of nucleosome disruption that exhibit extremely high sensitivity to DNase I digestion, thereby rendering the region accessible to transcription and chromatin remodeling factors: four of these HS are erythroid specific, whereas one is ubiquitous. The entire β-globin locus remains in an inactive DNase-I-resistant chromatin conformation in cells in which the globin genes are not expressed; in erythroid cells, the entire locus shows a higher degree of sensitivity to DNase I, indicating that it is in a more open and accessible chromatin configuration. In particular, in embryonic cells, the susceptibility to the digestion attack is higher in the region containing the ε-globin and γ-globin genes, whereas it is the zone where the δ-globin and β-globin genes are located that is more sensitive in the adult erythroid cells.

However, efficient transcription of genes requires a high local concentration of relevant trans-acting factors. Nuclear compartmentalization provides an effective means to locally enhance the level of rapidly moving trans-acting proteins, a process that can be achieved by spatial clustering of chromatin-associated binding sites for cognate transcriptional factors.

Some erythroid-specific factors are targets of histone acetyltransferases and some exist as phosphoproteins, indicating that their activities are regulated by posttranslational modification and that they may be involved in chromatin remodeling of the β-globin locus.

Three main models have been proposed to elucidate the regulation mechanism of the LCR: (1) in the looping model, the LCR acts as an integral unit, forming interactions with transcription factors and the core-flanking sequences; this structural unit loops in such a way as to directly associate with gene-proximal promoter and enhancer elements, thereby activating the appropriate gene expression (Fig. 8c); (2) according to the linking model, the LCR promotes the formation of a chain of transcriptional facilitator proteins that extends along the chromatin fiber between the LCR and the globin gene family to activate the individual gene promoter; and (3) in the tracking or scanning model, erythroid-specific and ubiquitous transcription factors and cofactors bind recognition sequences in the LCR, forming an activation complex that migrates – or tracks – linearly along the DNA helix of the locus; when such a complex encounters the basal transcription machinery located at the correct (according to the developmental stage) promoter, the complete transcriptional apparatus is assembled and transcription of gene initiates.

Chromosome organization is such that the members of each globin family are arrayed from amino-terminal coding (5′) to carboxy-terminal coding (3′) sequences (see Fig. 9). In the α-family genes, the three exons code for residues 1–31, 32–99, and 100–141 respectively; in the β-family genes, the corresponding positions are 1–30, 31–104, and 105–146. Even though the breaks occur in all chains at identical

peptide bonds along the helical segments (i.e. between residues B12 and B13 and between G6 and G7), they do not mark the boundaries of clearly defined structural domains; they do, however, correspond to smaller compact modules. Thus, the central exon (exon 2 in Fig. 9) encodes a module that provides a competent heme-binding site (helices E and F) and the $\alpha^1\beta^2$ sliding contact (helix C and FG corner); such a module binds heme tightly and specifically, but ligand-linked structural fluctuations (as observed in the protein fragment encoded by the central exon of the present-day Mb gene) are large enough to yield an unstable O_2 complex. Addition of the two terminal polypeptide modules (coded by exon 1, which corresponds to helices A and B and constitutes a scaffolding for the heme pocket, and exon 3, which provides helices G and H and contributes most of the $\alpha^1\beta^1$ packing contacts; see Fig. 9) reduces the structural degrees of freedom of the central module; the ensuing damping effect in the amplitude of the conformational fluctuations stabilizes the efficiency of the holoprotein as a unit of biochemical machinery.

The α-type gene cluster contains three functional genes (ζ, $\alpha1$, and $\alpha2$) and three pseudogenes indicated by the prefix ψ ($\psi_{\alpha1}$, $\psi_{\alpha2}$, and ψ_{ζ}). Pseudogenes have sequence homology to the functional genes but are defective in some essential coding and/or regulatory regions and thus cannot be expressed in a globin chain. The β-type gene cluster has five active genes (ε, G_γ, A_γ, δ, and β) and one pseudogene (ψ_β). The two fully functional and nearly identical α ($\alpha1$ and $\alpha2$) and γ (G_γ and A_γ) genes possibly reflect their duplication during evolution.

In these gene families, all vertebrate globins are encoded in the same order in which they are expressed during the development: $\varepsilon \rightarrow \gamma \rightarrow \delta, \beta$ in the β-cluster, and $\zeta \rightarrow \alpha$ in the α-cluster.

Throughout development from embryo to infant, a balance in expression between α-like and β-like globin genes is maintained, yielding fully functional Hb tetramers (Fig. 8b). However, the α- and β-globin genes are regulated in distinctly different ways. In particular, the α-globin gene does not require an added enhancer for expression in transfected erythroid and nonerythroid cells; in contrast, the β-globin gene requires the presence of a viral enhancer in cis for transient expression in nonerythroid cells. Hb synthesis is coordinated to minimize the accumulation of free subunits that form cytotoxic precipitates. A chaperone, known as *α-Hb Stabilizing Protein*, prevents *in vivo* precipitation of free α-chains, suggesting that α-globin levels and their incorporation into Hb tetramers is controlled posttranslationally. This is in contrast to β-globin concentrations, which are largely governed via regulation of transcription.

However, the regulatory mechanisms do not take into account whether the genes, which are yet to be activated, are functional or not. Consequently, when the programmed developmental time for the activation of globin gene is reached, the transcription machinery will be positioned even when the gene is defective and unable to produce any protein.

Alteration of the balanced synthesis of globins in the two clusters leads to a very important type of molecular disease called *thalassemia*. In β-thalassemia, a partial or complete deficiency of β-chains is associated with a prolonged biosynthesis of the fetal γ-chains (yielding HbF or $\alpha_2\gamma_2$) and possibly with the accumulation of free α-chains, which associate in the insoluble aggregates, with toxic effects on the development of erythroblasts. In α-thalassemia,

a relative deficiency in α-subunit biosynthesis results in an excess of γ- or β-chains, which assemble into γ_4 (Hb Bart's) and β_4 (HbH) homotetramers respectively.

10 Evolutionary Considerations

The essential features of the globin fold (i.e. the overall three-dimensional structure of the protein, with its ensemble of α-helical and nonhelical stretches; see Fig. 1) from very diverse organisms – ranging from mammals to insect and plant root nodules – are highly conserved, even when the amino acid sequences have diverged until similarity is virtually undetectable. This finding suggested to M. F. Perutz that the spontaneous folding of the globin chain implies stereochemical similarities in sequence, which satisfy the overall structural configuration in spite of large differences in amino acid composition; this led to the hypothesis that the globin family has evolved by divergence from a common ancestral gene (see Fig. 10). Moreover, the discovery of Hbs in virtually all kingdoms of organisms has suggested that the ancestral gene is ancient. Various estimates place the time of duplication of the globin genes from their single ancestor during early vertebrate evolution at approximately 800 million years ago (see Fig. 10). Although the resolution of the phylogenetic tree is poor, evolutionary analysis suggests that neuroglobins do not group with

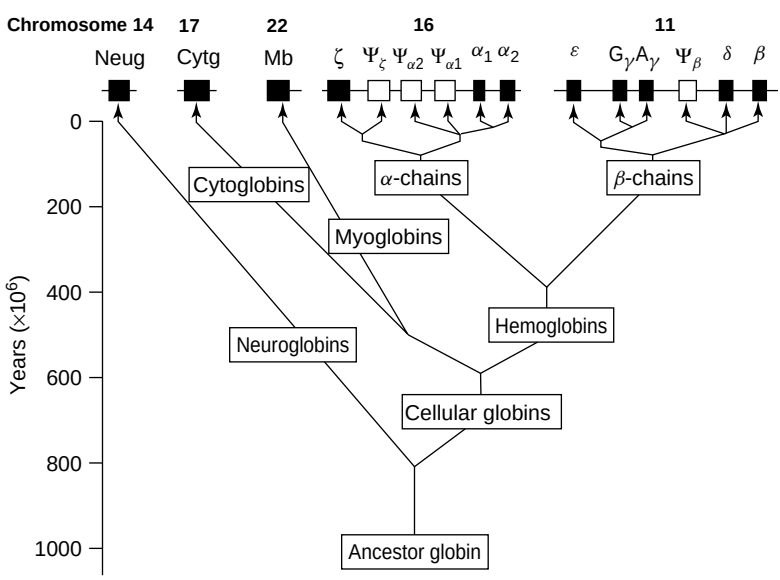

Fig. 10 Possible evolution of the globin genes, as deduced from DNA and amino acid sequence differences. The arrangement of the globin genes and pseudogenes (denoted by the prefix ψ) in humans is shown at the top, together with the indication of the chromosome where the gene is located. Estimated times of divergence in millions of years are only approximate. The high similarity between two α-genes (α_1 and α_2) and between genes G_γ and A_γ indicates that their duplication must have occurred quite recently in evolutionary time. The G_γ and A_γ genes produce chains with glycine or alanine, respectively, at position H14(136) (see also Table 1).

vertebrate Hbs and Mbs (thought to have diverged 500–600 million years ago) and must be older. A distinct position of neuroglobins is also suggested by exon–intron structure (see Sect. 9). The very early ancestor to vertebrates contained a single ancestral gene. On the basis of the antiquity of neuroglobin, it has been proposed that the last common ancestor to all vertebrates most likely possessed two globin loci. Thus, duplication of the ancestral locus resulted in two loci, developing into two different types of globin: neuroglobin and cellular globin (see, Fig. 10). A further duplication of cellular globin locus allowed the development of the myoglobin and cytoglobin loci, on one hand, and of Hb locus, on the other. The Hbs obtained their function in the circulatory system of the jawed vertebrates (fish, amphibians, and reptiles) after their divergence from the lineage leading to the myoglobins and cytoglobins, an event probably related to an increase in body size and to the development of an efficient circulatory apparatus. Myoglobins and cytoglobins separated later. From the phylogenetic data, it is unclear whether this occurred before the jawed vertebrate diverged from the jawless vertebrates or agnathans (lampreys and hagfish). In the Hb locus, gene duplication gave rise to a cluster encoding several monomeric Hbs, as found in agnathans. This allowed the specialization of individual genes in α-type or β-type Hbs that are found closely linked in the same locus (as in poikilothermic jawed vertebrates). It is thought that a common ancestral globin gene gave rise to distinct α- and β-globin genes through duplication events, followed by the separation of the α- and β-genes to different chromosomes, as they are in homeothermic vertebrates (birds and mammals). The original α-gene in turn subdivided into α and ζ somewhere in the reptilian line; on the other hand, the original β-gene differentiated into adult β and fetal γ during the rise of mammals. The original γ-gene itself later produced both γ and ε, whereas primates generated a δ-gene from the β-gene. During the long evolution of globin, only two positions (F8 and CD1) have remained invariant (i.e. have been found in all Hbs): the proximal histidine F8 and phenylalanine CD1, which wedges the heme into its pocket (see Fig. 2). This means that to ensure the satisfaction of similar stereochemical requirements, the vast majority of amino acid replacements have been conservative, that is, at a given position in the sequence, a residue has been replaced by many others (depending on the species) of the same general class. Since these mutations do not appear to be correlated with any fundamental change in the functional properties of Hb, they are called "neutral" and simply reflect accumulation of random nucleotide substitutions that occur at low frequency during DNA replication. It is expected, of course, that highly deleterious mutations will have been weeded out by natural selection.

Analysis of available sequences (>300) shows that evolutionary divergence has been constrained primarily by an almost absolute conservation of the hydrophobicity of the residues buried in the helix-to-helix and helix-to-heme contacts. Survival of mutant proteins in the globin family has been restricted to those that maintain the basic globin fold, tolerating variations not only in residues occupying surface positions but also in those buried in the molecule. Changes in surface residues are sometimes reflected in radically different quaternary structures with novel subunit contacts (see Sect. 7), typical of some invertebrate Hbs, which maintain reversible oxygen binding and cooperativity.

As for other proteins, the more distant two globins are in the phylogenetic tree, the more dissimilar the amino acid sequences of shared proteins will be. Thus, if the comparison between members of the globin family is restricted to vertebrates, the number of common residues is relatively high. Indeed, precisely 27 amino acids in the α-subunits and 18 amino acids in the β-subunits have remained invariant (based on the amino acid sequence of 32 species), and most of them are located at sequence positions responsible for efficient oxygen delivery: 14 are heme contacts, 16 are intersubunit contacts essential for allosteric mechanism, 8 are involved in hydrogen bonds or salt bridges within the subunits or at the $\alpha^1\beta^1$ interface, 6 are involved in internal nonpolar contacts, and a single proline serves to turn the BC corner in the β-subunits.

It has been estimated that changes in the Hb genes accumulated at an average rate of one residue per α- or β-subunit every 2 to 3 million years. This notion is in line with the general ideas of evolution, illustrated by the observation that, in the β-chains, there are eight differences between man (*Homo sapiens*) and rhesus monkey (*Macaca* spp.), 24 differences between man and cow, 45 differences between man and chicken, and 91 differences between man and shark.

However, a close scrutiny of available data for calculating evolutionary rates from the distribution of mutations on the genealogical tree (all corrected with assumptions about the supposed nature of molecular evolution) does not help in discriminating unequivocally between two possible conclusions: (1) the average rate of accumulation of mutations in globin sequences has remained approximately constant from the ancestral to the present-day Hb genes; or (2) genetic diversification occurred in various phases with discontinuous evolutionary rates.

The first conclusion is easier to explain by the neutralist theory, according to which most evolutionary changes at the molecular level and most of the variability within a species are caused not by selection but by random drift of mutant genes that are selectively equivalent. In other words, evolutionary rates are essentially determined by the structure and function of the molecules and not by environmental conditions or population size. The approximate constancy of the evolutionary rate for globin genes is supported by comparison of the divergence between human α- and β-chains and the divergence between the α-chain of carp and the β-chain of man (see Table 6). The two α-chains and the human β-chain differ to roughly the same extent. Since the sequence of the α-chain of man differs from that of carp in about half the positions, the α-subunits of the two distinct lineages (one leading to carp and the other to man) appear to have accumulated mutations independently and at approximately the same rate over a span of 500 million years (i.e. since gene duplication between α- and β-chains arose).

Although the biological-clock hypothesis proved to be a useful tool for the quantitative investigation of the time evolution of biological processes at the level of proteins, the clock itself does not appear to be as regular as a simple Poisson process. For example, the clock seems to have changed its rate during evolution while (or when) a tetrameric Hb was first assembled, starting from an ancestral monomeric globin. In fact, analysis of the reconstructed genealogical tree of globin chains (see Fig. 11) shows an acceleration–deceleration pattern of molecular evolution between ancestor and descendants, evidence favoring positive selection

Tab. 6 Number of differences between the amino acid sequences of human α- and β-chains compared with the number of differences between the sequences of α-chain from carp and human β-chain.

Type of change in the genetic code	*Human α- versus human β-chain*	*Carp α- versus human β-chain*
No change	62	61
One nucleotide	55	49
Two nucleotides	21	29
Addition or deletion of a residue	9	10
Total	147	149

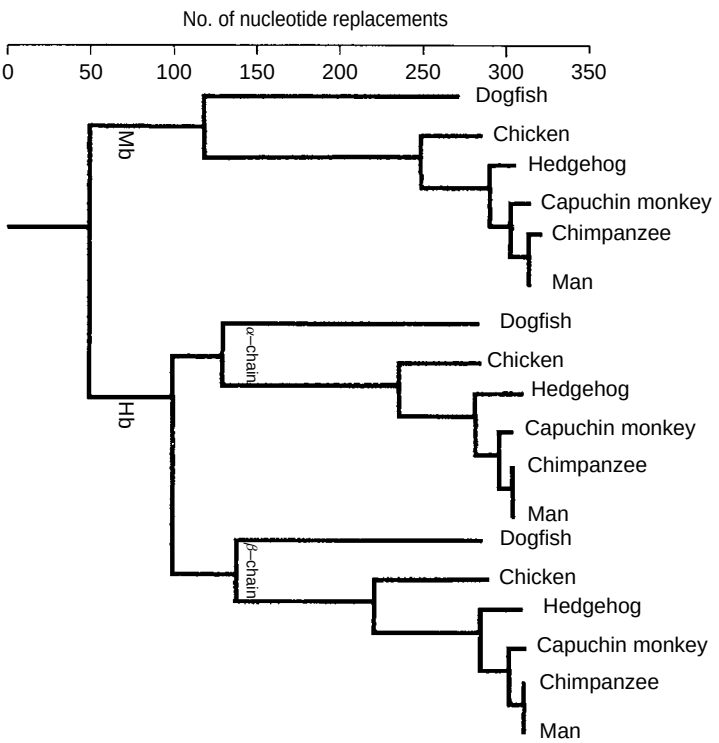

Fig. 11 Acceleration–deceleration pattern in rates of globin evolution during vertebrate phylogeny. The portions of the genealogical tree depict the phyletic distance (i.e. the number of sequence substitutions that have occurred as estimated from sequence data of 159 globins) from the early vertebrates to the present.

rather than neutral fixation. Thus, the selectionist or Darwinian view maintains that for a mutant allele to spread through a species, it must carry some selective advantage; it is also presumed that at each stage in the evolution of the tetramer, the portion of the molecule just being developed accepted nucleotide changes

more rapidly than the regions that had already been evolved or had yet to evolve. As stated by R. E. Dickerson, in the evolution of globin, there seem to be periods when sequence changes were predominantly a consequence of natural selection; on the other hand, after satisfactory selection of a globin suited to its physiological role, the rate of accumulation of sequence changes appears to have settled down to a nearly constant value – a behavior that is hard to reconcile with anything but random fixation of near-neutral mutations. In conclusion, neither positive selection nor neutral drift is the sole force directing changes in globin during evolution, but each model seems applicable and relevant at different time spans during the evolution of globin.

11 Hemoglobin Variants

11.1 Nature of Mutants

Several hundred mutant Hbs (882 up to the year 2004) within the human population have been isolated and chemically characterized: variants of α-subunits relative to those of β-subunits are approximately in a ratio of 1 : 2. In most cases, the abnormality consists of the replacement of a single amino acid residue per $\alpha\beta$ dimer; all these modifications are consistent with single-base substitution in DNA coding for the globin subunits. Some abnormal Hbs have residues deleted or inserted (168 variants); in others, the subunits are cut short or elongated (12 variants); others are constituted by fusion of two different subunits (9 variants); and yet others show more than one point mutation in the same polypeptide chain (15 variants). Of the large number of Hb mutants, a significant fraction has deleterious effects on health (see Table 7). Various databases of Hb variants have been developed, with increased capacity for sophisticated queries and prompt updating (e.g. http://globin.cse.psu.edu).

11.2 Molecular Basis of Hemoglobin Diseases

Pathological symptoms in the carriers of Hb diseases have been associated with altered oxygen affinity (causing polycythemia if it is raised or cyanosis if it is lowered), decreased stability of the protein (determining hemolysis and clumps of denatured Hb called *Heinz bodies*), an increased tendency to form ferric Hb (mainly producing cyanosis), and abnormal intracellular polymerization (inducing a contorted banana- or sickle-shaped erythrocyte). In many cases, these biochemical abnormalities have been interpreted with the help of the three-dimensional structure.

Many of the structural abnormalities giving rise to clinical symptoms are clustered around the heme pockets or are in the vicinity of the $\alpha^1\beta^2$ interface, so important in allosteric transitions. Some of the typical structural changes leading to abnormal Hbs are as follows:

1. A common structural pattern causing an increase in O_2 affinity is the loss of hydrogen bonds, salt bridges, or nonpolar interactions that stabilize the Hb tetramer and the T state in particular; moreover, some mutations increase O_2 affinity because they impair BPG interaction by modifying its binding site.
2. Unstable Hbs (128 variants) usually involve (1) an opening of heme pocket, loss of heme, and subsequent precipitation of globin; (2) insertion of a proline into a helix, where a kink is introduced

Tab. 7 Selected list of point mutations determining abnormal human Hbs.

Residue	*Replacement*	*Common name*	*Comments*
		Change in oxygen affinity	
CE3(45)α	His → Arg	Fort de France	Oxygen affinity ↑; stabilization of R state
G2(95)α	Pro → Ser	Rampa	Oxygen affinity ↑; tetramer dissociation
HC3(141)α	Arg → His	Suresnes	Oxygen affinity ↑; destabilization of T state
E18(74)β	Gly → Asp	Shepherds bush	Oxygen affinity ↑; decrease of BPG association
G4(102)β	Asn → Thr	Kansas	Oxygen affinity ↓; stabilization of T state
		Heme loss and/or denaturation	
H19(136)α	Leu → Pro	Bibba	Unstable; proline disrupts helix H
CD1(42)β	Phe → Ser	Hammersmith	Unstable, loss of heme; water attracted into heme
B14(32)β	Leu → Arg	Castilla	Unstable; positive charge in molecular interior
		Molecules with iron permanently in ferric oxidation state	
E7(58)α	His → Tyr	M Boston	Fe (III) ligated to Tyr instead of His
E14(70)β	Ala → Asp	Seattle	Stabilization of Fe (III) by negative charge of Asp
		Sickling	
A3(6)β	Glu → Val	S	Sickling; fit of Val into EF cleft of another tetramer
GH4(121)β	Glu → Lys	O Arab	Enhancement of sickling in S/O heterozygote; residue 121 β lies close to residue 6 β and Lys increases interaction between tetramers

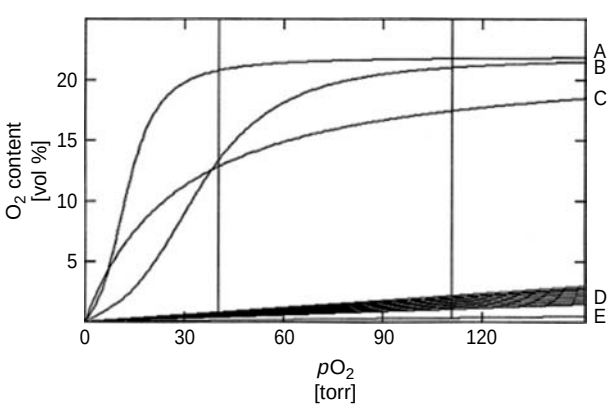

Fig. 12 The O_2 content of blood and some O_2 carriers as a function of O_2 partial pressure under physiological conditions (pH 7.4, 37 °C): A, native Hb stripped from BPG; B, whole blood; C, pyridoxylated, glutaraldehyde-polymerized Hb; E, plasma. Area D comprises the range of commonly employed perfluorocarbons (compounds forming fine and stable emulsions with water). The vertical lines indicate venous (40 torr) and arterial (105 torr) oxygen pressure. It can be noted that release of O_2 by a solution of stripped, native Hb is very small, while for polymerized pyridoxylated Hb, it is approximately half that of blood.

that usually disrupts the tertiary structure; or (3) unfavorable removal of a nonpolar side chain on the molecular surface (which excludes the interior of the protein from the surrounding water), introduction of a charged group inside the molecule, or deletion of single residues or segments of polypeptide chains. All these events can disrupt the basic folding of Hb.

3. The M (for met Hb) mutants are the class of Hbs in which the Fe atom of one type of subunit is permanently oxidized as a result of the replacement of a heme contact residue; the most common mutations are the substitution of either the proximal histidine (HisE7) or the distal histidine (HisF8) by a tyrosine in the α-, the β-, or the γ-subunit.
4. In general, the mutations that lead to abnormal Hb behavior are not located on the exterior of the tetramer; yet, the replacement of a surface glutamate at position A3(6)β with a hydrophobic valine in HbS or sickle-cell mutant (the most prevalent Hb variant worldwide) has deleterious effects. Upon deoxygenation of a concentrated Hb solution, HbS tetramers aggregate into a gel, distorting and rigidifying the erythrocytes into a variety of bizarre shapes. As a consequence, the red cells may not be able to transverse the microcirculation vessels, with transient to permanent blockage of local oxygenation. When this happens, the resulting organ damage is a major cause of the morbidity and mortality of sickle-cell anemia. The key to erythrocyte sickling in HbS is the presence of a hydrophobic pocket located between helices E and F and made up of PheF1(85)β and LeuF4(88)β, which binds ValA3(6)β; upon oxygenation, the helix F motion of the β-subunits pulls the PheF1(85)β of one tetramer away from potential contact with the ValA3(6)β side chain of a second tetramer, leading to dissociation of the fibrils into oxygenated HbS tetramers.

12 Hemoglobin Solutions as Blood Substitutes

Many blood substitutes should be more properly called O_2 carriers since they provide O_2 transport and volume replacement but do not perform other functions of blood (e.g. facilitating coagulation and immune responses). They are life-support systems able to dissolve (and thereby transport) large amounts of O_2; this mode of operation is at variance with simple plasma expanders, which contain dissolved O_2 at low concentration ($\approx$0.3 mM). Among potential O_2 carriers, attention has been focused on perfluorocarbons (synthetic compounds that reversibly dissolve O_2) and a variety of Hb preparations, either free in solution or encapsulated into lipid vesicles. To be efficient, perfluorocarbons have to be used under high O_2 pressure (see Fig. 12), otherwise, the amount of O_2 transported is much too low; moreover, they are immunotoxic.

Unprocessed solutions of Hb are unsuitable as O_2 carriers for several reasons. First of all, Hb quickly disappears from plasma, partly through formation of a complex with the plasma-protein haptoglobin (mainly cleared by the liver with a half-life of 10–30 min) but mostly through direct filtering off by the kidney, with a half-life of 1 h. The direct elimination though the kidney is possible because in dilute solutions, tetrameric Hb easily dissociates into free $\alpha\beta$ dimers; the kidney is permeable to proteins of molecular mass $<40\,000$ g mol^{-1}, unless very highly charged.

Second, O_2 release to tissues is poor as a result of the high affinity of Hb dissolved in plasma, which is stripped from BPG, the allosteric effector that modulates the efficiency of oxygenation. At 37 °C and pH 7.4, $p_{1/2}$ would decrease from the physiologically desirable level of about 26 torr to about 10 torr, which is too low for effective delivery (see Fig. 12).

Finally, unprocessed solutions of Hb are not suitable as O_2 carriers because the isosmotic concentration of Hb dissolved in plasma (70 g dm^{-3}) is about half that of normal blood (150 g dm^{-3}). To overcome these drawbacks, several covalent modifications of Hb have been described to produce Hb-based blood substitutes, with the idea of capitalizing on their potential advantages, such as, the Hb solutions can be used immediately without the need for blood group typing, they can be stored in a ready-for-use formulation for years and heat-treated, thus eliminating the risk of viral transmission.

Thus, the first generation of Hb-based O_2 carriers were designed to prevent dissociation into dimers and, therefore, to prolong the half-life of Hb within circulation. The molecular mass of chemically engineered O_2 carriers has been increased by (1) intra- and inter-tetrameric cross-linking with glutaraldehyde or glycolaldehyde; (2) conjugation to dextran and polyethylene glycol derivatives; and (3) protein engineering by fusion of the two α-subunits. Moreover, to decrease the O_2 affinity of Hb and, therefore, to improve unloading to the tissues, covalent attachment to the α-subunits of pyridoxal 5′-phosphate (which is known to mimic the functional effect of BPG) or specific chemical changes [e.g. Asn G10(108)$\beta \rightarrow$ Lys] obtained by site-directed mutagenesis, were realized. Other compounds, such as bis(3,5-dibromosalicyl)fumarate and *nor*-2-formylpyridoxal-5′-phosphate, reduce the O_2 affinity and, at the same time, stabilize the tetramer.

Large-scale polymerization (such as that obtained by glutaraldehyde treatment) not only prolonged the vascular retention of the Hb molecules but also lowered the osmotic activity by decreasing the number of particles per unit volume. In particular, the isosmotic concentration in plasma was increased from 70 g dm^{-3} (of free Hb) up to 150 g dm^{-3} (i.e. the Hb level in blood of healthy subjects). The latter system showed some disadvantages, however, since the solution viscosity was much higher than that of blood. Under chosen conditions, glutaraldehyde-treated Hb forms polymers containing two to eight tetramers, having an isosmotic concentration of 100 g dm^{-3} and a viscosity lower than that of blood; the vascular retention of this compound in rabbits was about 20 h, a time sufficient for several clinical applications. In fact, when the Hb derivative is used solely for the purpose of O_2 transport, as in the prevention of cardiac ischemia (e.g. during balloon angioplasty), a relatively short half-life (5–10 h) may actually be useful.

Identifying alternative sources of Hb has been an additional, although not secondary problem, since supply of the human protein is limited. Therefore, the use of Hb from other mammals (such as ox, pig, and mouse) has been considered since the Hb molecule is scarcely antigenic; however, the possibility of anaphylactic reactions after repeated applications is not unlikely, even though the antigenic response in humans can be minimized by modification of Hb with polyethylene glycol.

However, severely adverse outcomes were noted in clinical trials with first-generation Hb-based O_2 carriers, which

were therefore dropped in favor of developing second-generation products. In fact, outside the erythrocyte, Hb appears to be inherently toxic, being able to generate highly reactive O_2 species (mainly, superoxide and peroxide) and to interfere with the modulation of blood pressure. Thus, met Hb levels of polymerized Hb can increase in 24 h from an initial 3 to 40%. Harmful reactive O_2 species, which are by-products of Hb auto-oxidation, are known to produce cellular damage and constitute an important concern in the use of modified Hb as reperfusion agents. Moreover, the reaction of met Hb with H_2O_2, produced during the process of auto-oxidation, is known to proceed via the formation of the highly reactive ferryl Hb (i.e. HbFe(IV)), a short-lived chemical species that peroxidizes lipids, degrades carbohydrates, and modifies proteins as well as nucleic acids. Moreover, extracellular oxy Hb – in contrast to that encapsulated in erythrocytes (see Sect. 8) – is highly vasoconstrictive and the resulted hypertension is a significant side effect of most Hb-based blood substitutes. All these effects can explain some of the unique toxicological effects associated with the use of the first-generation O_2 carriers in the clinic, on the basis of chemically modified Hb.

Strategies aimed at solving the above-reported toxicities form the basis for the second-generation Hb prototypes of blood substitutes being able to safely carry O_2. Thus, intraerythrocyte antioxidant enzymes (such as superoxide dismutase and catalase) or water-soluble vitamin E analogs, have been cross-linked to Hb in the attempt to control Hb oxidative side reactions, obtaining amelioration of free-radical-mediated injury to cellular systems. Moreover, the active sites of Hb subunits have been engineered (e.g. LeuB10 of both chain types replaced with large aromatic amino acids) in order to reduce the rate of NO binding (by $\geq$30-fold), and therefore the NO scavenging activity, while still retaining efficient O_2 transport.

See also Heme Enzymes; Metalloenzymes.

Bibliography

Books and Reviews

Antonini, E., Brunori, M. (1971) *Hemoglobin and Myoglobin in Their Reactions with Ligands*, North-Holland, Amsterdam, The Netherlands.

Antonini, E., Rossi Bernardi, L., Chiancone, E., (Eds.) (1981) *Methods in Enzymology*, Vol. 76, *Hemoglobins*. Academic Press, New York.

Dickerson, R.E., Geis, I. (1983) *Hemoglobin: Structure, Function, Evolution and Pathology*, Benjamin/Cummings, Menlo Park, CA.

Eaton, W.A., Hofrichter, J. (1990) Sickle cell hemoglobin polymerization, *Adv. Protein. Chem.* **40**, 63–279.

Everse, J., Vandegriff, K.D., Winslow, R.M. (Eds.) (1994) *Methods in Enzymology*, Vols. 230, 231, and 232, *Hemoglobins*. Academic Press, New York.

Fermi, G., Perutz, M.F. (1981) *Atlas of Molecular Structures in Biology: Hemoglobin and Myoglobin*, Clarendon Press, Oxford, UK.

Forget, B.G., Higgs, D.R., Steinberg, M., Nagel, R.L. (2001) *Disorders of Hemoglobin: Genetics, Pathophysiology and Clinical Managements*, Cambridge University Press, Cambridge, UK.

Imai, K. (1982) *Allosteric Effects in Hemoglobin*, Cambridge University Press, Cambridge, UK.

Perutz, M.F. (1990) Mechanisms regulating the reactions of human hemoglobin with oxygen and carbon monoxide, *Annu. Rev. Physiol.* **52**, 1–25.

Wyman, J., Gill, S.L. (1990) *Binding and Linkage. Functional Chemistry of Biological Macromolecules*, University Science Books, Mill Valley, CA.

Primary Literature

Early References

Adair, G.S. (1925) A critical study of the direct method of measuring the osmotic pressure of hemoglobin, *Proc. Roy. Soc.* **108A**, 627–637.

Agmon, N., Hopfield, J.J. (1983) CO binding to heme proteins: a model for barrier height distributions and slow conformational changes, *J. Chem. Phys.* **79**, 2042–2053.

Antonini, E., Wyman, J., Brunori, M., Fronticelli, C., Bucci, E., Rossi-Fanelli, A. (1965) Studies on the relations between molecular and functional properties of hemoglobin. V. The influence of temperature on the Bohr effect in human and horse hemoglobin, *J. Biol. Chem.* **240**, 1096–1103.

Austin, R.H., Beeson, K.W., Eisenstein, L., Frauenfelder, H., Gunsalus, I.C. (1975) Dynamics of ligand binding to myoglobin, *Biochemistry* **14**, 5355–5373.

Brunori, M., Bonaventura, J., Bonaventura, C., Antonini, E., Wyman, J. (1972) Carbon monoxide binding by hemoglobin and myoglobin under photodissociating conditions, *Proc. Natl. Acad. Sci. USA* **69**, 868–871.

Bucci, E., Fronticelli, C. (1965) A new method for the preparation of α and β subunits of human hemoglobin, *J. Biol. Chem.* **240**, 551–552.

Chiancone, E., Norne, J.E., Forsen, S., Antonini, E., Wyman, J. (1972) Nuclear magnetic resonance quadruple relaxation studies of chloride binding to human oxy- and deoxyhaemoglobin, *J. Mol. Biol.* **70**, 675–688.

Edelstein, S.J. (1971) Extensions of the allosteric model to haemoglobin, *Nature* **230**, 224–227.

German, B., Wyman, J. (1937) The titration curves of oxygenated and reduced hemoglobin, *J. Biol Chem.* **177**, 533–550.

Gibson, Q.H. (1959) The photochemical formation of a quickly reacting form of haemoglobin, *Biochem. J.* **71**, 293–303.

Hopfield, J.J., Shulman, R.G., Ogawa, S. (1971) An allosteric model of hemoglobin: I, kinetics, *J. Mol. Biol.* **61**, 425–443.

Imai, K. (1983) The Monod-Wyman-Changeux allosteric model describes haemoglobin oxygenation with only one adjustable parameter, *J. Mol. Biol.* **167**, 741–749.

Koshland, D.e. Jr., Nemethy, G., Filmer, D. (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits., *Biochemistry* **5**, 365–385.

Monod, J., Wyman, J., Changeux, J.-P. (1965) On the nature of allosteric transitions: a plausible model, *J. Mol. Biol.* **12**, 88–18.

Perutz, M.F. (1970) Stereochemistry of cooperative effects in hemoglobin, *Nature* **228**, 726–738.

Rossi-Fanelli, A., Antonini, E., Caputo, A. (1961) Studies on the relations between molecular and functional properties of hemoglobin, *J. Biol. Chem.* **236**, 391–400.

Roughton, F.J.W., Otis, A.B., Lyster, R.L.J. (1955) The determination of the individual equilibrium constants of the four intermediate reaction between oxygen and sheep hemoglobin, *Proc. R. Soc. London* **144B**, 29–54.

Sawicki, C.A., Gibson, Q.H. (1976) Quaternary conformational changes in human hemoglobin studied by laser photolysis of carboxyhemoglobin, *J. Biol. Chem.* **251**, 1533–1542.

Shulman, R.G., Hopfield, J.J., Ogawa, S. (1975) Allosteric interpretation of haemoglobin properties, *Q. Rev. Biophys.* **8**, 325–420.

Szabo, A., Karplus, M. (1972) A mathematical model for structure-function relations in hemoglobin, *J. Mol. Biol.* **72**, 163–197.

Wyman, J. (1964) Linked functions and reciprocal effects in hemoglobin: a second look, *Adv. Protein Chem.* **19**, 233–286.

Wyman, J., Allen, D.W. (1951) The problem of the heme interactions in hemoglobin and the Bohr effect, *J. Polym. Sci.* **7**, 499–518.

Zito, R., Antonini, E., Wyman, J. (1964) The effect of oxygenation on the rate of digestion of human hemoglobins by carboxypeptidases, *J. Biol. Chem.* **239**, 1804–1808.

Recent References

Ackers, G.K. (1998) Deciphering the molecular code of hemoglobin allostery, *Adv. Protein Chem.* **51**, 185–253.

Adachi, S., Park, S.Y., Tame, J.R.H., Shiro, Y., Shibayama, N. (2003) Direct observation of photolysis-induced tertiary structural changes in hemoglobin, *Proc. Natl. Acad. Sci. USA* **100**, 7039–7044.

Alayash, A.I. (2004) Oxygen therapeutics: Can we tame haemoglobin? *Nat. Rev.* **3**, 152–159.

Amiconi, G., Ascoli, F., Barra, D., Bertollini, A., Matarese, R.M., Verzili, D., Brunori, M. (1989) Selective oxidation of methionine β(55)D6 at the $\alpha_1\beta_1$ interface in hemoglobin completely

destabilizes in T-state, *J. Biol. Chem.* **264**, 17745–17749.

Barrick, D., Ho, N.T., Simplaceanu, V., Dahlquist, F.W., Ho, C. (1997) A test of the role of the proximal histidines in the Perutz model for cooperativity in haemoglobin, *Nat. Struct. Biol.* **4**, 78–83.

Bettati, S., Mozzarelli, A., Perutz, M.F. (1998) Allosteric mechanism of haemoglobin: rupture of salt-bridges raises the oxygen affinity of the T-structure, *J. Mol. Biol.* **281**, 581–585.

Bourgeois, D., Vallone, B., Schotte, F., Arcovito, A., Miele, A.E., Sciara, G., Wulff, M., Anfinrud, P., Brunori, M. (2003) Complex landscape of protein structural dynamics unveiled by nanosecond Laue crystallography, *Proc. Natl. Acad. Sci. USA* **100**, 8615–8617.

Brunori, M. (2000) Structural dynamics of myoglobin, *Biophys. Chem.* **86**, 221–230.

Brunori, M. (2001) Nitric oxide, cytochrome-c oxidase and myoglobin, *Trends Biochem. Sci.* **26**, 21–23.

Brunori, M., Gibson, Q.H. (2001) Cavities and packing defects in the structural dynamics of myoglobin, *EMBO Rep.* **2**, 674–679.

Burmester, T., Weich, B., Reinhardt, S., Hankeln, T. (2000) A vertebrate globin expressed in the brain, *Nature* **478**, 461–462.

Eaton, W.A., Henry, E.R., Hofrichter, J. (1991) Application of linear free energy relations to protein conformational changes: the quaternary structural change of hemoglobin, *Proc. Natl. Acad. Sci. USA* **88**, 4472–4475.

Eaton, W.A., Henry, E.R., Hofrichter, J., Mozzarelli, A. (1999) Is cooperative oxygen binding by hemoglobin really understood?, *Nat. Struct. Biol.* **6**, 351–358.

Franzen, S., Bohn, B., Poyart, C., Martin, J.-L. (1995) Evidence for sub-picosecond heme doming in hemoglobin and myoglobin: a time-resolved resonance Raman comparison of carbonmonoxy and deoxy species, *Biochemistry* **34**, 1224–1237.

Frauenfelder, H., Wolynes, P.G. (1985) Rate theories and puzzles of hemeproteins kinetics, *Science* **229**, 337–345.

Frauenfelder, H., Sligar, S.G., Wolynes, P.G. (1991) The energy landscapes and motions of proteins, *Science* **254**, 1598–1603.

Gibson, Q.H. (1989) Hemoproteins, ligands and quanta, *J. Biol. Chem.* **264**, 20155–20158.

Gibson, Q.H. (1999) Kinetics of oxygen binding to hemoglobin A, *Biochemistry* **38**, 5191–5199.

Gill, S.J., Robert, C.H., Coletta, M., Di Cera, E., Brunori, M. (1986) Cooperative free energies for nested allosteric models as applied to human hemoglobin, *Biophys. J.* **50**, 747–752.

Gladwin, M.T., Schechter, A.N. (2004) NO contest: nitrite versus S-nitroso-hemoglobin, *Circ. Res.* **94**, 851–855.

Gladwin, M.T., Lancaster, J.R., Freeman, B.A., Schechter, A.N. (2003) Nitric oxide's reactions with hemoglobin: a view through SNO-storm, *Nat. Med.* **9**, 496–500.

Hardison, R. (1998) Hemoglobins from bacteria to man: evolution of different patterns of gene expression, *J. Exp. Biol.* **201**, 1099–1117.

Harju, S., McQuenn, K.J., Peterson, K.R. (2002) Chromatin structure and control of beta-like globin gene switching, *Exp. Biol. Med.* **227**, 683–700.

Jia, L., Bonaventura, C., Bonaventura, J., Stamler, J.S. (1996) S-nitrosohaemoglobin: a dynamic activity of blood involved in vascular control, *Nature* **21**, 221–226.

Kihm, A.J., Kong, Y., Hong, W., Russell, J.E., Rouda, S., Adachi, K., Simon, M.C., Blobel, G.A., Weiss, M.J. (2002) An abundant erythroid protein that stabilizes free alpha-haemoglobin, *Nature* **417**, 758–763.

Lepeshkevich, S.V., Karpiuk, J., Sazanovich, I.V., Dzhagarov, B.M. (2004) A kinetic description of dioxygen motion within α- and β-subunits of human hemoglobin in the R-state: geminate and bimolecular stages of the oxygenation reaction, *Biochemistry* **43**, 1675–1684.

Navas, P.A., Li, Q., Peterson, K.R., Swank, R.A., Rohde, A., Roy, J., Stamatoyannopoulos, G. (2002) Activation of the beta-like globin genes in transgenic mice is dependent on the presence of the beta-locus control region, *Hum. Mol. Genet.* **11**, 893–903.

Paoli, M., Liddington, R., Tame, J., Wilkinson, A., Dodson, G. (1996) Crystal structure of T state haemoglobin with oxygen bound at all four haems, *J. Mol. Biol.* **256**, 775–792.

Pawloski, J.R., Stamler, J.S. (2002) Nitric oxide in RBCs, *Transfusion* **42**, 1603–1609.

Perutz, M.F. (1989) Mechanisms of cooperativity and allosteric regulation in proteins, *Q. Rev. Biophys.* **22**, 139–236.

Perutz, M.F., Wilkinson, A.J., Paoli, M., Dobson, G.G. (1998) The stereochemistry of cooperative effects in hemoglobin revisited, *Annu. Rev. Biophys. Biomol. Struct.* **27**, 1–34.

Peterson, K.R. (2003) Hemoglobin switching: new insights, *Curr. Opin. Hematol.* **10**, 123–129.

Riggs, A.F. (1998) Self-association, cooperativity and supercooperativity of oxygen binding to hemoglobins, *J. Exp. Biol.* **201**, 1073–1084.

Rivetti, C., Mozzarelli, A., Rossi, G.L., Henry, E.R., Eaton, W.A. (1993) Oxygen binding by single crystals of hemoglobin, *Biochemistry* **32**, 2888–2906.

Rousseau, D.L., Tan, S.L., Ondrias, M.R., Noble, R.W. (1984) Absence of cooperative energy at the heme in liganded hemoglobins, *Biochemistry* **23**, 2857–2865.

Royer, W.E., Jr., Knapp, J.E., Strand, K., Heaslet, H.A. (2001) Cooperative hemoglobins: conserved fold, diverse quaternary assemblies and allosteric mechanisms, *Trends Biochem. Sci.* **26**, 297–304.

Sawado, T., Halow, J., Bender, M.A., Groudine, M. (2003) The beta-globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation, *Genes Dev.* **17**, 1009–1018.

Scerri, A.C. (2003) What are the molecular switches for the various globin genes (embryonal, foetal and adult) and what triggers them? *IUBMB Life* **55**, 167–169.

Schotte, F., Lim, M., Jackson, T.A., Smirnov, A.V., Soman, J., Olson, J.S., Phillips, G.N. Jr., Wulff, M., Anfinrud, P.A. (2003) Watching a protein as it functions with 150-ps time-resolved x-ray crystallography, *Science* **300**, 1944–1947.

Scott, E.E., Gibson, Q.H., Olson, J.S. (2001) Mapping the pathways for O_2 entry into and exit from myoglobin, *J. Biol. Chem.* **276**, 5177–5188.

Sun, D.Z.P., Zou, M., Ho, N.T., Ho, C. (1997) The contribution of surface histidyl residues in the alpha-chain to the Bohr effect of human adult normal hemoglobin: roles of global electrostatic effects, *Biochemistry* **36**, 6663–6673.

Tame, J.R.H. (1999) What is the true structure of liganded hemoglobin? *Trends Biochem. Sci.* **24**, 372–377.

Unzai, S., Eich, R., Shibayama, N., Olson, J.S., Morimoto, H. (1998) Rate constants for O_2 and CO binding to the alpha and beta subunits within the R and T states of human hemoglobin, *J. Biol. Chem.* **273**, 23150–23159.

Wittemberg, J.B., Bolognesi, M., Wittenberg, B.A., Guertin, M. (2002) Truncated hemoglobins: a new family of hemoglobins widely distributed in bacteria, unicellular eukaryotes, and plants, *J. Biol. Chem.* **277**, 871–874.

Yonetani, T., Park, S.I., Tsuneshige, A., Imai, K., Kanaori, K. (2002) Global allostery model of hemoglobin. Modulation of O_2 affinity, cooperativity, and Bohr effect by heterotropic allosteric effectors, *J. Biol. Chem.* **277**, 34508–34520.

Zhang, H.B., Liu, D.P., Liang, C.C. (2002) The control of expression of the alpha-globin gene cluster, *Int. J. Hematol.* **76**, 420–426.

Hemophilia, Molecular Genetics of

Francesco Giannelli and Peter M. Green
Department of Medical and Molecular Genetics, Division of Genetics and Development, Guy's, King's and St. Thomas' School of Medicine, London, UK

1 Genetics of the Hemophilias 111
1.1 The Factor VIII Gene and its Product 113
1.2 The Factor IX Gene and its Product 115

2 Hemophilia A Mutations 117
2.1 Gross Sequence Changes 118
2.2 Small Sequence Changes 119
2.2.1 Changes Affecting RNA Processing 119
2.2.2 Changes Affecting mRNA Translation 120
2.2.3 Mutations Causing Alterations of the Fine Structure of Factor VIII 121
2.3 Functional Interpretation of Observed Sequence Changes and Correlations between Genotype and Phenotype 123

3 Hemophilia B Mutations 124
3.1 Gross Sequence Changes 125
3.2 Small Sequence Changes 125
3.2.1 Mutations Affecting Transcription, RNA Processing, and Translation 126
3.2.2 Missense Mutations and Single Amino Acid Deletions 127
3.3 Functional Interpretation of Observed Sequence Changes and Correlations between Genotype and Phenotype 128

4 Considerations Arising from the Study of Mutations in Hemophilia A and B 129

5 Mutational Heterogeneity and Genetic Counseling in the Hemophilias 130

6 Progress in the Treatment of Hemophilia 131

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

Bibliography 132
Books and Reviews 132
Primary Literature 132

Keywords

CpG and CpG island
CpG is a dinucleotide formed by C followed by G, and CpG islands are genomic regions unusually rich in CpGs and found before genes and especially genes with ubiquitous expression patterns.

Founder Mutation
A mutation traceable to a common ancestor and found relatively frequently in some populations.

Frameshift
Mutation of a coding sequence due to loss or addition of one or two nucleotides or multiples thereof. This alters the amino acid sequence from the point of mutation onward and frequently leads to premature termination of the encoded polypeptide.

Male-driven Evolution
A situation where the male, by having a higher mutation rate than the female, contributes a greater proportion of the variation acted on by natural selection.

Missense
Mutation resulting in the substitution of one amino acid for another.

Mutational Load
The sum of the deleterious mutations present in a zygote.

Nonsense
Mutation converting a codon specifying an amino acid into one determining the end of a polypeptide sequence.

Uninemic
Of events affecting a single molecule of double-stranded DNA as that present in a chromatid or a prereplication chromosome.

Zygote
The cell formed by the fusion of a male and female gamete and representing the origin of a new potential individual.

The two X-linked hemorrhagic diseases hemophilia A and hemophilia B have made an important contribution to the understanding of human genetics. The first historical record of the inheritance of a disease and of "genetics"-motivated advice relates to these diseases, as do the first description of the X-linked pattern of inheritance, the appreciation of equilibria between selection and mutation, and the first attempt to estimate mutation rates. In modern times, the hemophilias have clearly illustrated the successes and problems of replacement therapy and have provided some insight into the causes of immunological complication associated with the therapeutic administration of the missing gene products. It is therefore not surprising that nowadays they attract the attention of scientists and clinicians involved in the development of gene therapy. The hemophilias, which affect 1/5000 males in the population, are members of a large group of diseases characterized by very high mutational heterogeneity. This group is of particular importance because, on the one hand, the wealth of natural mutants that each member possesses allows accumulation of detailed information on the structural features that are important to the functions of the gene and gene product involved in the disease; on the other, this creates the least favorable situation for the development of fully efficient strategies for the provision of the carrier and prenatal diagnoses that are necessary for genetic counseling. Surprisingly, mutation analysis in hemophilia A has revealed that homologous intranemic recombination between duplicated sequences can represent an important cause of recurring gene inactivation.

1
Genetics of the Hemophilias

A clear reference to the X-linked diseases, now called *hemophilias*, is found in the Babylonian Talmud of the fifth century A.D., wherein the Rabbi advises mothers who had lost two sons as a result of uncontrolled bleeding after circumcision not to expose other sons to such a risk. Furthermore, the X-linked recessive pattern of inheritance of these diseases, characterized by affected males and unaffected females who transmit the disorder, was lucidly described in 1803, more than a century before the discovery of sex-linked inheritance and 50 years ahead of Mendel's work. In 1935, it was realized that the loss of hemophilia genes due to the affected males' low probability of reproduction would have led to the elimination of the disease if an equilibrium had not existed between such loss and the gain of hemophilia genes by new mutations. Under such an equilibrium, the rate of renewal of hemophilia genes in the population is $1/3\ (1-f)$ per generation, where f is the effective fertility or genetic fitness of the affected male and 1/3 expresses the well-known fact that males account for one-third of the X-linked genes in the population, as they have only one X chromosome, whereas females have two. In 1947, investigation of the linkage between hemophilia and color blindness (another X-linked trait) offered the first hint that hemophilia was not a single disease, because while most families showed close linkage between hemophilia and color blindness, one did not. Five years later, however, the existence of two forms

of hemophilia was clearly demonstrated by complementation of the coagulation defect in a mixture of plasma from different hemophiliacs. Now, we know that deficiency of coagulation factor VIII (FVIII) causes hemophilia A, while deficiency of factor IX (FIX) causes hemophilia B and that the identical clinical features of these diseases are entirely to be expected, since FIX and FVIII are respectively the enzyme and cofactor of a functional unit that contributes to the proteolytic cascade, which converts fibrinogen into fibrin by activating coagulation factor X (Fig. 1).

Hemophilia A is 5 to 6 times more frequent than hemophilia B and, currently, in the United Kingdom, it has an incidence of approximately 1/5000 males. Until the introduction of modern therapy, the effective fertility of patients with hemophilia A or B was half that of normal individuals and, therefore, the rate of renewal of the pool of hemophilia A and B genes in the population was 1/6 per generation. This is in keeping with the clinical heterogeneity of the diseases. Since hemophilia A is 5 to 6 times more frequent than hemophilia B, and the rate of renewal of hemophilia A and B genes is similar, it follows that the rate of mutation for hemophilia A is about 5 to 6 times greater than that for hemophilia B. Indirect estimates of mutation rates in hemophilia have been based on the equilibrium postulated above, but recently, a direct estimate independent of the postulate was made in hemophilia B, through the study of the UK population, and this gave a value of 7.7×10^{-6} mutations per gene per generation, with a 95% confidence interval of 6.2 to 9.1. However, the mutation rate is higher in the male than in the female germline as the estimate of the former is 18.8×10^{-6} and that of the latter is 2.18×10^{-6} so that the ratio

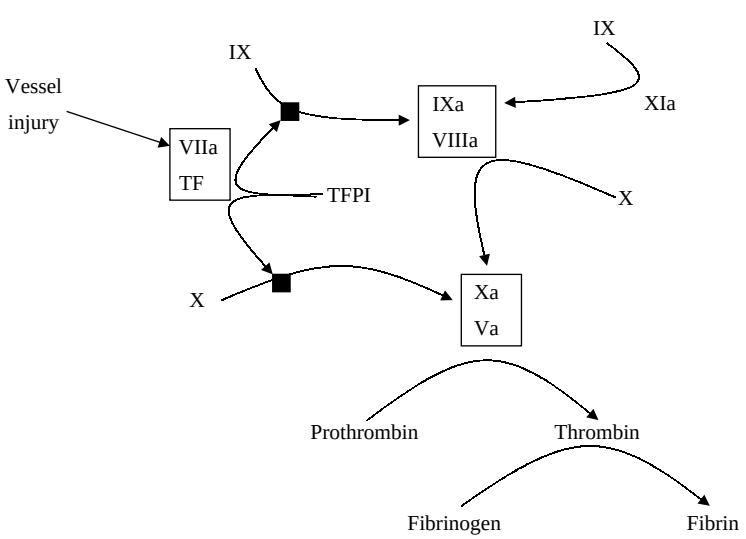

Fig. 1 The role of FVIII and FIX in blood coagulation. After tissue damage, activated FVII (FVIIa) in complex with tissue factor (TF) activates both FIX and FX, and, in turn, activated FX (FXa) and thrombin activate FVIII (not shown). Inhibition of the FVIIa–TF complex by TFPI makes further conversion of FIX to its active form dependent on FXIa and activation of FX on the FVIIIa–FIXa complex.

of male to female mutation rates is 8.64, with a 95% confidence interval of 5.46 to 14.5. This high ratio is in keeping with the theory of male-driven evolution and may be explained by the different development and physiology of the male and female gonads. Extrapolation from these data led to the suggestion that the deleterious mutation rate in the human genome is greater than 1 and possibly as great as 4 per zygote per generation. This is important because such a high mutational load in turn suggests that the long-term survival of humans depends on genetic recombination and, hence, sexual reproduction, which allows the elimination of chromosome regions carrying deleterious mutations by a process of truncating selection. This is a process that allows the elimination of mutations in bunches by preferentially eliminating zygotes with multiple deleterious mutations. In this way, it can prevent the decline in genetic fitness from generation to generation that could follow deleterious mutation rates as high as suggested above.

1.1 The Factor VIII Gene and its Product

The FVIII gene (*F8*) is 186-kb long (Fig. 2) and contains 26 exons ranging in size from 69 to 3106 bp. The latter is an internal exon (number 14) and as such is of exceptional size. The introns also vary considerably in size, from 200 bp (intron 17) to 32.4 kb (intron 22). The latter intron is of particular interest because it contains a CpG island associated with two additional RNA transcripts, one of opposite polarity to the FVIII transcript, entirely contained within intron 22 and defining an intronless gene, *F8A*, of about 2 kb and one of the same polarity as FVIII, *F8B*. This contains a first exon derived from intron 22 that contributes eight

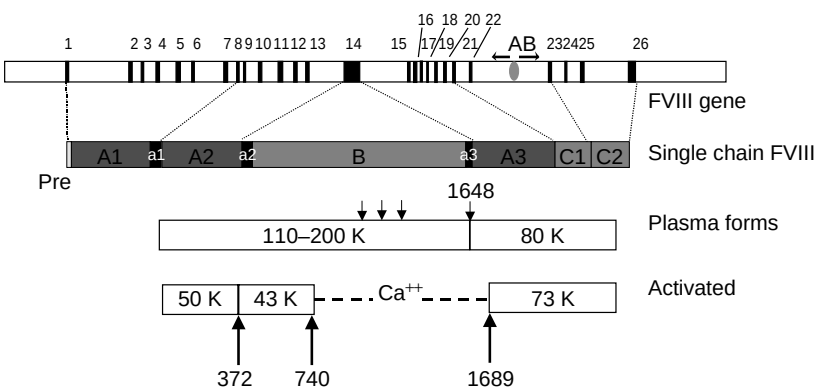

Fig. 2 Factor VIII gene (*F8*) and protein. Diagram of *F8* gene (top bar) and FVIII (lower bars). Exons are indicated by black boxes and numbered (1–26). Arrows in opposite directions over intron 22 indicate the presence of CpG island (green box) associated with the start of two transcripts of opposite orientation (A, B). Dotted lines indicate exons coding for the different protein domains (see text). Arrows in plasma and activated forms indicate site of proteolytic cleavages (see text) and numbers associated with arrows indicate position of amino acid at the amino terminal end of cleavage sites. Numbers in bars are weight in Daltons of individual protein segments. (See color plate p. xxiv.)

codons to a message comprising exons 23 to 26 of the *F8* gene. The functional significance of these two transcripts is still uncertain. A protein encoded by *F8B* has not been found yet, while the protein encoded by *F8A* has been found associated to Huntingtin. The above CpG island in intron 22 and the *F8A* gene are part of a 9.5-kb sequence called *int22h* that is repeated 500 and 600 kb more telomerically.

The FVIII mRNA is thought to be 9028-nt long with a 170 leader sequence followed by 2351 codons and an 1805-nt untranslated 3′ tail. The primary translation product starts with a prepeptide of 19 residues needed for intracellular transport and secretion. The remainder of the protein consists of six domains and three small acidic peptides (a_1 to a_3) arranged as follows: $A_1a_1A_2a_2Ba_3A_3C_1C_2$. The A domains consist of approximately 330 amino acids (aa) and are homologous to ceruloplasmin. The B domain (encoded by exon 14) contains 983 aa and is unique. The C domains consist of approximately 150 aa and are homologous to a recently described milk fat globule binding protein and the slime mold's discoidin. Factor V, the cofactor of factor X that is responsible for the activation of prothrombin, is a homolog of FVIII and has a similar protein domain structure; however, it lacks the three acidic peptides (a_1, a_2, and a_3) and has a B domain with no homology to that of FVIII.

The FVIII polypeptide undergoes extensive intracellular modification including glysosylation, sulfation, and proteolytic cleavage. Glycosylation is both O and N linked, and FVIII contains 26 glycosylatable asparagines of which 19 are in the B domain. Sulfation is known to occur at the following tyrosines: Tyr_{346}, Tyr_{718}, Tyr_{719}, Tyr_{723}, Tyr_{1664}, and Tyr_{1680}. Secretion of FVIII requires not only post translational glycosylation of the asparagines in the B domain but also processing of the glycosidic chains to fully glucose-trimmed mannose 9 structures. The transfer of FVIII from the endoplasmic reticulum to the Golgi apparatus is achieved with the intervention of trafficking factors, as revealed by the fact that absence of endoplasmic reticulum Golgi intermediate compartment -53, or ERGIC-53 for short, causes combined deficiency of both FVIII and FV.

Prior to secretion, FVIII is cleaved not only to release the prepeptide but also at the B/a_3 boundary (aa 1648–1649) to form a molecule with a light chain consisting of $a_3A_3C_1C_2$ (aa 1649–2232) and a heavy chain consisting of $A_1a_1A_2a_2B$. This chain, however, is of variable length since the B domain is also cleaved at variable internal positions.

Circulating FVIII is bound to and protected by the von Willebrand factor (vWF). Two FVIII regions are important for this binding: the a_3 peptide and in particular the sulfated tyrosine at position 1680 within this peptide and a region at the carboxyl end of the C_2 domain comprising residues 2303–2332. Conversion of FVIII to its active form entails cleavage at the a_1/A_2 and the a_3/A_3 boundaries (aa Arg_{372}-Ser_{373} and Arg_{1689}-Ser_{1690}). The former cleavage transforms factor VIII into a three-chain molecule and the latter releases factor VIII from von Willebrand factor. Activation is also accompanied by cleavage at the a_2/B boundary. The activated FVIII is relatively unstable and, in particular, the dissociation of the A_2 domain from the active heterotrimer results in loss of activity. Activated protein C, an important regulator of coagulation, cleaves FVIII at the A_1/a_1 boundary (Arg_{336}-Met_{337}) and causes its inactivation, but this protein also cleaves at Arg_{562}

and this may represent an important target for inactivation by protein C. In addition, work on the clearance of FVIII from the circulation has suggested that in the absence of vWF binding FVIII is rapidly cleared by a process that involves a member of the family of endocytic receptors: the low-density lipoprotein receptor–related protein also called the α_2-*macroglobulin receptor*.

The large size and multiple forms of the FVIII protein in circulation have hindered crystallographic investigation, but a model of the tridimensional structure of the A domains of FVIII has been derived, by analogy, from the crystal structure of ceruloplasmin. Crystals of the C_2 domain of FVIII have provided data on the tridimensional structure of this module and, finally, two-dimensional analysis of crystals formed by activated FVIII in association with phospholipidic membranes has yielded a picture of the arrangement of the activated FVIII domains relative to each other, the phospholipidic membrane and activated FIX.

The C_2 domain essentially consists of an eight-stranded antiparallel β-barrel formed by two β-sheets tightly packed against each other. Loops projecting from one end of the barrel have been shown to be important for the binding to phospholipidic membranes. Thus, it appears that three hydrophobic feet formed by the side chains of Met_{2199} and Phe_{2220}, Val_{2223} and Leu_{2251} plus Leu_{2252} embed into the membrane, while the four basic residues Arg_{2215}, Arg_{2220}, Lys_{2227}, and Lys_{2249} interact with the negative charges of the phospholipids.

The C_1 domain of FVIII can be modeled on the C_2 domain, especially in the core regions that are more conserved. The model of the three A domains of the activated FVIII suggests that these domains are arranged in a triangular structure, while the two-dimensional crystals of activated FVIII prepared on monolayers of negatively charged phospholipids indicate that the C_2 domain penetrates the membrane with four loops. The C_1 domain is almost at a right angle to C_2 and has its major axis parallel to the membrane. The A_3 domain is in close association with the C_1 and C_2 domains near the membrane, and the heterotrimer of A domains is tilted approximately 65° relative to the membrane plane, with the A_2 domain protruding partially between the A_1 and A_3 domains.

1.2 The Factor IX Gene and its Product

The FIX gene (*F9*) is approximately 33-kb long (Fig. 3) and contains eight exons varying in size from 25 to 1935 bp. The introns also vary in size from 188 to 9473 bp. The gene promoter has been partly characterized and binding sites for a number of transcription factors (NF1-L, androgen receptor, LF-A1/HNF4, DBP, and members of the C/EBP family) have been found close to the transcription start site. The FIX mRNA is 2802-nt long and includes a leader sequence 29- to 50-nt long, depending on the true position of the translation start, and 1390 nt of 3′ untranslated tail. There is a close correspondence between *F9* exons and protein domains. These are modules shared by large families of proteins, and the structure of the *F9* gene is consistent with the exon-shuffling hypothesis of gene evolution. Three other proteins of blood coagulation are coded by genes with the same structure as *F9*, that is, factors VII, X, and protein C. These genes are probably derived from the same ancestral gene as FIX. More remotely related to *F9* gene is the prothrombin gene, as these two genes

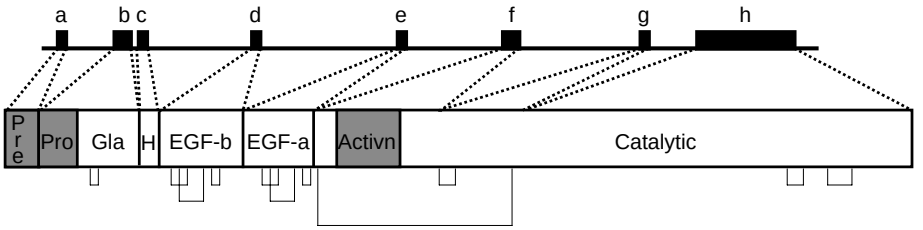

Fig. 3 Factor IX gene (*F9*) and protein. Diagram of *F9* gene (upper line) and FIX (lower bar). Exons are indicated by black boxes and labeled (a–h). Dotted lines show domains encoded by individual exons. Green areas are protein domains cleaved prior to secretion (pre, pro) or at activation: pre = prepeptide; pro = propeptide; gla = gla region; H = hydrophobic stack; EGF-b = first epidermal growth factor domain with Ca^{++} binding site; EGF-a = second epidermal growth factor domain; activn = activation peptide flanked by carboxyl end of light chain of activated factor IX and amino end of the catalytic region. Disulfide bridges in factor IX are shown below the protein domains. (See color plate p. xxiv.)

show identical arrangements only for the first three exons and are homologous only in this region and in that encoding the catalytic domain.

The first exon of *F9* codes for the predomain of the leader peptide that is involved in the transport and secretion of FIX. The second exon codes for the prodomain of the leader peptide and the gla region. These two domains are functionally related since the propeptide binds the enzyme responsible for the γ-carboxylation of 11 glutamic acid residues that give the name to the gla region (γ-carboxyglutamic acid = gla). The third exon codes for a small hydrophobic region that contains the 12th gla residue of FIX. The gla region confers on FIX affinity for phospholipidic membranes. The fourth and fifth exons code for domains homologous to epidermal growth factor (EGF domains). The first of these domains is called *type B* and contains a high-affinity Ca^{++} binding site contributed by Asp_{47}, Asp_{49}, Gln_{50}, and Asp_{64}. The latter residue also undergoes β-hydroxylation in a proportion of the FIX molecules while Ser_{53} and Ser_{61} are glycosylated. The second EGF domain that lacks the Ca^{++} binding site and β-hydroxylated residue is called *type A*. The EGF domains are involved in protein–protein interaction. Thus, the gla plus the EGF1 domain of FIX seems to interact with the activated FVII and tissue factor complex that activates FIX and the interface between the two EGF domains with residues 1811–1818 of FVIII. The sixth exon codes for the activation domain that comprises a peptide of 35 aa (residues 146 to 180) that is cleaved off during activation. This contains two glycosylated asparagines (Asn_{157} and Asn_{167}). Finally, the last two exons of the *F9* gene code for a catalytic domain that has all the features of serine proteases (e.g. trypsin). Thus, His_{221}, Asp_{269}, and Ser_{365} are considered the cardinal residues of the active center and Asp_{359} the residue conferring affinity for basic residues. Residues 301–303 and 333–339 of FIX appear to interact with residues 558–565 and 698–712 of the A_2 domain of activated FVIII.

FIX is synthesized in the liver as a precursor molecule with a leader peptide of 46, 41, or 39 aa; it is not known which of these is correct, but 39 is probable. This peptide is cleaved prior to secretion in two independent steps for the pre- and prodomains. Circulating FIX is a

monomer of 415 aa that upon activation by FXI or FVII plus tissue factor is converted into a two-chain molecule.

Crystallographic and other structural studies have shown that the gla domain that contains the phospholipids binding site has at the N terminus a tightly hydrogen-bonded complex involving two calcium ions and a pseudoplanar network of six to seven calcium ions that are tightly associated with gla residues. A helix then connects the calcium-associated fragment to the EGF1 domain. The calcium ion bound to the EGF1 domain is presumed to fit in the interface between the gla and EGF1 domain. The principal axes of the oblong EGF domains form an angle of 100° stabilized by an interdomain salt bridge. The catalytic region is localized above the EGF2 domain so that the activated FIX molecule shows a curved configuration with the catalytic center about 80 Å above the phospholipidic membrane surface. The concave surface of activated FIX contains interacting sites for FVIII and sites affected by hemophilia B–causing mutations.

2 Hemophilia A Mutations

The large size and complexity of the *F8* gene has meant that adequate screening of this gene for hemophilia A mutations has had to wait for the development of novel, efficient, and fast methods. Of particular value in this respect was the use of traces of FVIII mRNA in the RNA prepared from peripheral lymphocytes. This, in particular, has revealed that 20% of all cases and 50% of those with severe hemophilia are due to gross unexpected disruptions of the *F8* gene. The other hemophilia A mutations are mostly small changes that involve single base substitutions or the insertion and deletion of a few nucleotides, while gross deletions or duplications are relatively rare (Fig. 4).

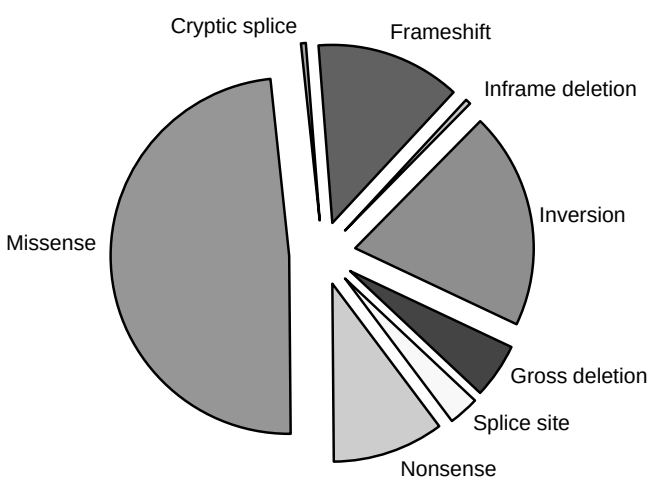

Fig. 4 Pie chart showing the different types of mutations reported in hemophilia A. Cryptic splice refers to mutations generating novel splice signals. In-frame deletion refers to mutations causing loss of one or more codons.

2.1 Gross Sequence Changes

In the United Kingdom, the study of 25% of the hemophilia A population has shown that 1.2% of patients have gross gene deletions. The world database of hemophilia A mutations (HAMSTeRS at http://europium.csc.mrc.ac.uk) records 116 gross deletions. These may involve the whole or a part of the gene (e.g. a single exon), and are usually associated with severe disease. In this series, the four deletions that are accompanied by moderate hemophilia affect exons 5 and 6; exon 22; exons 23 and 24; and exon 25. The mRNA in these cases is expected to maintain the normal reading frame and, presumably, the deletion simply leads to the loss of the aa encoded by the missing exons.

Large insertions are quite rare, but two interesting patients have been reported with insertion of the repetitive LINE sequence in AT-rich regions of exon 14. These have been shown to be recent transposition events from an active LINE sequence in chromosome 22.

The major cause of gross *F8* gene rearrangements is inversions. These, as mentioned above, cause 50% of the instances of severe hemophilia A or about 20% of all cases of the disease. Two main types of inversion exist: one that breaks the *F8* gene in intron 22 and one that breaks intron 1. The former accounts for 45% and the latter for 5% of patients with severe hemophilia A.

The inversions breaking intron 22 (Fig. 5) were the first to be discovered, and formal proof was obtained that the inversions result from intranemic homologous recombination between repeats of a 9.5- kb sequence called *int22h*. One copy of this sequence (*int22h-1*) is part of intron 22 of the *F8* gene, while a second (*int22h-2*) and third copy (*int22h-3*) are respectively located 500 and 600 kb more telomerically and are both in inverted orientation relative to *int22h-1*. These repeats are more than

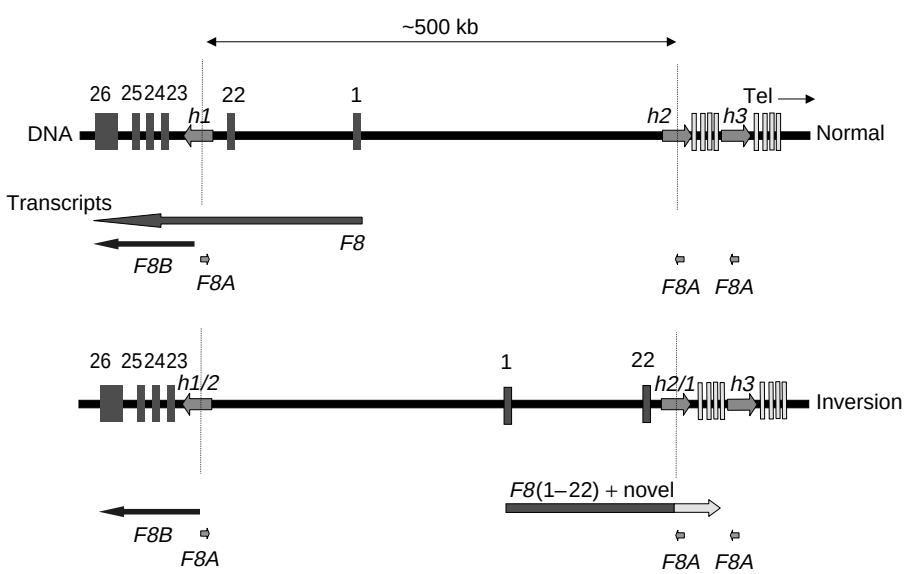

99.9% identical. Inversions due to homologous recombination between *int22h-1* and *int22h-3* are 5 to 6 times more common than those involving *int22h-1* and *int22h-2*, but the reason for this bias is unknown. It also appears that these inversions arise predominantly during male gametogenesis. Since the above inversions cause severe disease and hence are under strong negative or purifying selection, their high frequency is explained by a high rate of mutation. Estimates of 4 and 7×10^{-6} *int22h*-related inversions per gamete per generation have been offered. The *int22h*-related inversions are the first mutations sought in patients with severe hemophilia A, and this is done either by Southern blotting or by a more rapid PCR-based procedure.

The inversions breaking intron 1 of *F8* (Fig. 6) arise by intranemic homologous recombination like the *int22h*-related inversions. The sequence involved in this recombination is a 1041-bp segment of intron 1 called *int1h-1* that is duplicated in inverted orientation 140 kb more telomerically to yield *int1h-2*. These two repeats are more than 99.9% identical.

The *int1h*-related inversions are about ninefold rarer than the *int22h*-related inversions, and this reflects the size difference between the *int1h* and *int22h* repeats, in keeping with the idea that the frequency of homologous recombination is related to the length and degree of identity of the sequences involved.

2.2 Small Sequence Changes

A variety of small sequence changes has been reported, but none so far affecting the putative *F8* promoter. The known mutations, therefore, may primarily affect RNA processing, translation, or the fine structure of FVIII.

2.2.1 Changes Affecting RNA Processing

The presence of traces of FVIII mRNA in peripheral lymphocytes offers the opportunity to detect mutations altering the structure of mRNA irrespective of where they are located within the *F8* gene or, indeed, to confirm that mutations postulated to alter mRNA processing actually do so. Many of these mutations can be expected to alter the donor or acceptor splice signal with consequent skipping of the affected exon. In some patients, however, alteration of the normal splice signal is accompanied by the activation of cryptic splice signals in the affected exon or in the adjacent intron, thus respectively causing partial exon exclusion and exon extension. Finally, more rarely, patients are found to

Fig. 5 Diagram of *int22h*-related inversion, showing its transcriptional consequences. Top: normal structure of DNA region and *F8* gene transcripts (not drawn to scale). *F8* = transcript encoding coagulant factor, *F8A* and *F8B* = transcripts arising from CpG island in *int22h* repeat. Green arrows in genomic DNA represent *int22h* repeats (*h*1 = *int22h*-1, *h*2 = *int22h*-2, *h* = *int22h*-3). Red bars indicate some numbered exons of the *F8* gene, yellow bars are exons unrelated to the *F8* gene. Vertical dotted lines are sites where intranemic homologous recombination occurs and leads to the inversion. Tel → shows direction to telomere. Arrow point in transcripts indicates 3′ end. Bottom: structure of DNA region with inversion resulting from recombination of *int22h*-1 with *int22h*-2. *H1/2* and *h2/1* are the recombined *int22h* repeats. The transcripts show that the *F8* transcript contains only exons 1 to 22 of *F8* attached to the normally unrelated yellow exons and that exons 23 to 26 are present only in the *F8B* transcript. NB inversions involving *int22h*-3 yield analogous transcripts. (See color plate p. xxiii.)

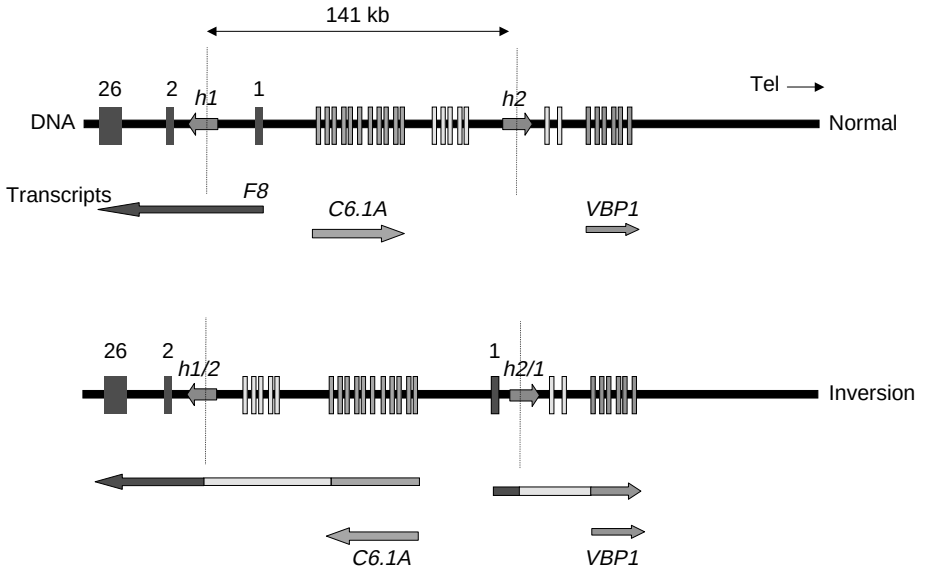

Fig. 6 Diagram of *int1h*-related inversion, showing its transcriptional consequences. Top: normal structure of DNA region and relevant transcripts (not drawn to scale). Red bars indicate numbered exons of *F8*. Light green and blue bars are exons of the *C6.1A* and *VBP1* gene respectively; yellow bars are exons unrelated to the above three genes. Green arrows in DNA are *int1h*-1 (*h1*) and *int1h*-2 (*h2*). Vertical dotted lines indicate sites where intranemic homologous recombination occurs and leads to inversion. Arrow point in transcripts indicates 3′ end. Bottom: structure of DNA region with inversion and relevant transcripts. *H1/2* and *h2/1* are recombined *int1h* repeats. Transcripts show that *F8* contributes to two mRNAs: one containing all but the last exon of *C6.1A*, some novel exons and exons 2 to 26 of *F8* and one containing exon 1 of *F8* plus novel exons and all but the first exon of *VBP1*. The patients' cells also produce normal *C6.1A* and *VBP1* mRNA. (See color plate p. xxv.)

produce mRNA with an additional exon because of single base substitutions generating a novel splice signal in a region of an intron close enough to a cryptic splice signal appropriate to the definition of a novel exon.

The severity of the disease in patients with the above mutations appears to be related to whether a fraction of the pre-mRNA can be spliced into wild-type molecules or whether the abnormal mRNA has retained a normal reading frame. When the normal reading frame has been lost and no normally spliced mRNA is produced, the patient tends to have severe hemophilia A.

2.2.2 Changes Affecting mRNA Translation

HAMSTeRS lists nonsense mutations at 95 different codons, and at 12 of these sites, the mutations were transitions at a CpG dinucleotide. These are usually observed in multiple patients because of the relatively high rate of transition mutations at this type of dinucleotide.

Nonsense mutations should result in premature termination of translation and hence in truncated FVIII, which may or may not be stable enough to be secreted. However, a phenomenon called *nonsense mediated mRNA degradation* may lead to mRNA instability and hence

prevent or grossly reduce FVIII synthesis. Conversely, some nonsense mutations may also impair exon recognition and splicing. In this case, exon skipping may restore the reading frame in at least some mRNA molecules, which can then be expected to produce FVIII lacking the amino acids encoded by the exon that has been skipped.

Nonsense mutations are fairly evenly distributed along the FVIII coding sequence and are generally associated with severe hemophilia A.

HAMSTeRS lists small deletions causing frameshifts at 138 sites and small insertions also causing frameshifts at 48 sites of the FVIII coding sequence. At most of these sites, the mutations appear to be unique events but there are some sites that are repeatedly affected and thus represent hotspots of deletion or insertion. The most spectacular hotspots are represented by regions of single base repeats such as strings of A that are present at various locations in exon 14. These appear to undergo slippage during DNA replication, thus acquiring or losing a base. High mutation rates at mononucleotide repeats could lead to somatic mosaicism in the patient with reversion of the mutation, and this would have beneficial effects. Alternatively, errors occurring during transcription could result in some mRNA molecules with a restored reading frame. It is interesting, therefore, to note that while patients with frameshift mutations usually have severe disease, a few with mild or moderate disease have been listed and these bear frameshifts affecting polyA strings. In the absence of the above corrections, occurring either in some cells at DNA replication or in some transcripts during gene expression, frameshifts tend to cause functional defects similar, by and large, to gross deletions and nonsense mutations.

2.2.3 Mutations Causing Alterations of the Fine Structure of Factor VIII

HAMSTeRS contains two mutations resulting in amino acid addition: Ile at codon 613 in one case and a string of six amino acids after residue 1888 in the other. Conversely, the database contains eight amino acid deletions, of which five affect a single residue, two affect two residues, and one affects four residues. However, the mutations that affect the fine structure of FVIII are mostly amino acid substitutions. The above database, after exclusion of the UK patients studied in the course of establishing a confidential national database of hemophilia A mutations and pedigrees, contains 437 different missense mutations, 324 of which are unique events while 113 have been observed two or more times. The above UK study has examined 767 unrelated patients and can be contrasted with the data summarized in HAMSTeRS in order to estimate the total number of FVIII missense mutations that can cause hemophilia A. The 437 missense mutations in HAMSTeRS represent the group of missense mutations causing hemophilia A that is already known. The size of the group formed by hemophilia A missense mutations that are still unknown can be estimated if the missense mutations found in the UK sample are divided into those that are listed in HAMSTeRS and those that are not. The ratio of the latter to the former is proportional to the number of different missense mutations that can cause hemophilia A and are not found in HAMSTeRS. In fact, 102 of 202 missense mutations found in the UK study are novel relative to HAMSTeRS and, hence, the missense mutations causing hemophilia A and yet not in this collection should be 221 (i.e. $102/202 \times 437$), so that the total number of missense mutations capable of causing

hemophilia A can be assumed to be about $221 + 437 = 658$. This is not a particularly high figure if one considers that the *F8* gene encodes 2351 amino acids and that 15,588 missense mutations can occur in this gene.

The distribution of missense mutations causing hemophilia A is far from being homogeneous (Fig. 7). Thus, the B domain is bereft of such mutations while it has a proper share of nonsense mutations and even an excess of frameshift mutations. This is explained by the fact that the B domain does not contribute to the coagulant activity and hence to the antihemophilic function of FVIII.

The three A domains of the patients' FVIII show similar numbers of missense mutations and these preferentially appear to occur at residues that are identical in the three A domains, as the known mutations affect 44% of these residues while they involve only 28% and 24% of those that are, respectively, identical in only two domains or different in all three domains. Similarly, the known missense mutations in the C domains affect 42% of the residues that are identical in all the C domains of both factor VIII and factor V, while they affect only 22% of the residues that are identical in the C domains of factor VIII and 11% of the residues that are not identical in the C domain of factor VIII.

These data indicate that substitution of residues important for the basic structure of the A and C domains are much more likely to cause hemophilia A than substitution of other amino acids.

A number of missense mutations affect residues of known functional importance. Thus, Arg has been found to be substituted for Ser_1 at the cleavage site of the prepeptide. Pro, His, or Cys, and Pro or Leu, respectively, have been found substituted for Arg_{372} and Ser_{373}, which form an activation cleavage site. Similarly, Cys or His, and His or Arg, respectively, have been found substituted for Arg_{1689} and Ser_{1690}, which also form an activation cleavage site.

Conversely, some missense mutations have indicated residues important to the stability of the FVIIIa heterotrimer because they affect the interfaces between the A_1/A_2, A_1/A_3, and A_2/A_3 domains. This group of mutations includes the following: R284E, R284P, S289L, R527W, R531H, R531G, Q694I, R698L, R698W, L1932F, M1947I, G1948D, and H1954L. Interestingly, coagulant assays on patients with these mutations show higher values when they include FVIII activation in

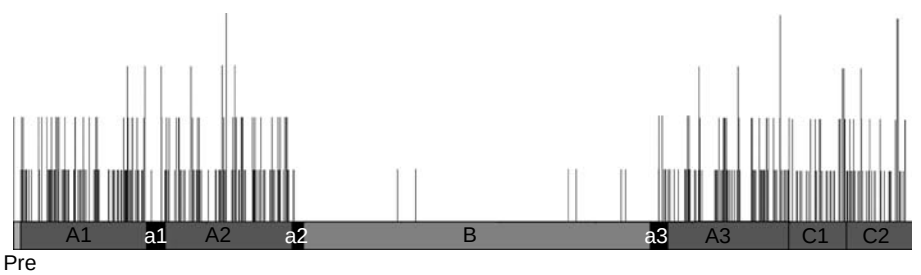

Fig. 7 Missense mutations of *F8* associated with hemophilia A. The horizontal bar shows the domains of the FVIII polypeptide. The vertical lines show the positions of missense mutations, and their heights are proportional to the number of different substitutions (from 1 to 4) found at each residue. Any line thicker than the minimum thickness indicates mutations at two or more residues close to one another but not necessarily adjacent in FVIII (data from HAMSTeRS).

the measuring phase (1-stage assays) than when FVIII activation is allowed to precede the measuring phase (2-stage assays) as this is likely to allow greater FVIII inactivation, prior to measurement, if the FVIIIa heterotrimer is unstable. *Ad hoc* experiments for some of these mutations have demonstrated abnormally rapid dissociation of the A_2 domain from the heterotrimer.

Missense mutations in hemophilia A may also offer some information on the importance of sites of protein interaction. Three regions of FVIII are considered binding sites for FIX: residues 558–565, 698–712, and 1811–1818. The first two interact respectively with the region of FIX residues 333–339 and 301–303; the third appears to interact with the interface between the two EGF domains of FIX. In HAMSTeRS, eight missense mutations are listed for the first site, nine for the second, and none for the third. With regard to the interaction of FVIII with vWF, the Y1680C and Y1680F mutations found in hemophilia A patients bear witness to the importance of this sulfated tyrosine for the binding to vWF. Fifteen different missense mutations affecting the binding of FVIII to vWF have been found in the C domains.

2.3 Functional Interpretation of Observed Sequence Changes and Correlations between Genotype and Phenotype

A proportion of the observed FVIII mutations can undoubtedly be considered detrimental, that is, gross deletions or other gross rearrangements involving the essential sequences of the gene, such as the *int22h*- and *int1h*-related inversions, alteration of absolutely conserved elements of the splice site consensuses, frameshifts, and nonsense mutations. However, the accurate prediction of the functional consequences of mutations detected in genomic DNA may demand mRNA analysis.

Many mutations are not overtly detrimental and do not lead to gross mRNA aberrations. In order to determine the detrimental nature of such mutations, it is important to have accurate diagnoses and to perform a very thorough mutation screen on the *F8* gene. In particular, one should exclude diseases that can mimic hemophilia A. At present, we know of two such conditions: the combined deficiency of FV and FVIII associated with defects in secretion pathways and the Normandy-type von Willebrand disease due to mutations that impair the ability of vWF to bind FVIII. Mutation screening of the *F8* gene should ensure levels of detection as close as possible to 100%. Currently, this can be done best by procedures that include the use of mRNA. Under these conditions, mutations that represent the only change detected in the *F8* gene are likely to be the cause of the disease. In some cases, family studies can show that the mutation found in a patient is of recent origin and this strongly supports its deleterious nature. Also, the independent occurrence of the same mutation in two or more patients strongly supports the idea that the mutation is the cause of the disease. Less useful indicators of the disease-causing role of observed sequence changes are considerations on the degree of conservation of the mutated residues of FVIII or the chemical difference between the normal and the mutant FVIII residues. The accumulation of thorough hemophilia A mutation data should also help identify neutral polymorphisms, as these may appear as second changes common to patients with different disease-causing mutations.

With regard to the correlation between phenotype and genotype, missense mutations are frequently associated with mild or moderate disease.

Variability in the severity of the disease has been reported among hemophilia A patients with the same mutation, and this may be due to polymorphic variation in genes that encode factors important for the maturation, secretion, transport, clearance, and modulation of the activities of FVIII.

The most worrying phenotypic feature of patients with hemophilia A is the propensity to develop antibodies against therapeutic FVIII. The risk of developing such antibodies or inhibitors is strongly correlated with the type of mutation causing the disease. Thus, a risk of about 30% is associated with the *int22h*-related inversions, gross deletions, and nonsense mutations, while frameshifts show a lower risk of about 20% and missense mutations an overall risk of 8%. HAMSTeRS shows 39 missense mutations associated with the presence of inhibitors and for 20 of these mutations, there are two or more patients per mutations with information on the inhibitor status, forming a total of 161 patients, 48 of which (29.8%) had inhibitors. This suggests that most of the risk of inhibitors associated with missense mutations is due to a small group of residues and their substituents that possibly create novel antigenic epitopes. Some evidence that this might be the case has been obtained for inhibitors associated with the mutations R593C and R2150H. With regard to the first mutation, a patient has been reported with inhibitors directed exclusively against the wild-type FVIII, while, with regard to the second mutation, two patients with high titer inhibitors were reported who still displayed significant levels of FVIII activity, thus suggesting that the patients' mutant FVIII was not recognized and hence not blocked by the patients' inhibitors. The high risk of inhibitors associated with gross rearrangements and nonsense and frameshift mutations is best explained by the inability of patients with these mutations to produce FVIII. This could result in lack of immunological tolerance to FVIII, which would not be recognized as "self." As nonsense and frameshift mutations are expected to cause similar effects on FVIII synthesis, the reasons why frameshifts are associated with a somewhat lower risk of inhibitors should be considered. The insertion or deletion mutants in polyA regions ($\geq$6A residues) appear to be mainly responsible as they show a very low risk of inhibitors. Thus, only five (6.5%) of 76 patients with these frameshifts are reported as having inhibitors in HAMSTeRS. Possibly, this is due to errors of replication or transcription of the mutated polyA regions, which, as mentioned above, may allow production of traces of FVIII that may serve to develop immune tolerance to FVIII.

3 Hemophilia B Mutations

The vast majority of hemophilia B mutations are due to single base-pair substitution, or deletion and insertion of a few base pairs, while gross deletions and other rearrangements are rare (Fig. 8). So far, 2847 mutations involving small sequence changes and 90 gross deletions have been reported in the world database (http://www.kcl.ac.uk/ip/petergreen/haemBdatabase.html). However, large population samples provide more precise information on the relative frequency of gross and small sequence changes.

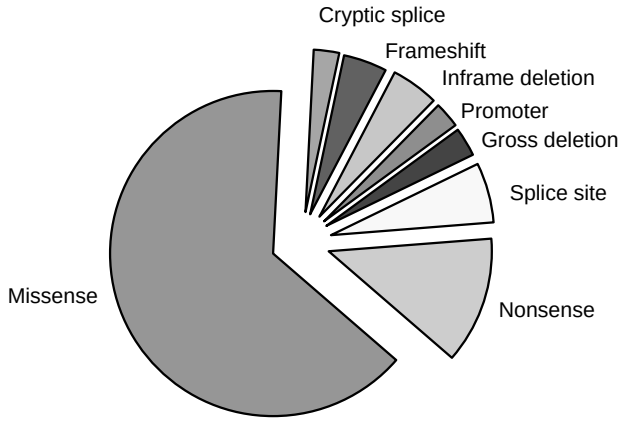

Fig. 8 Pie chart showing the different types of mutations reported in hemophilia B. Promoter refers to mutations affecting the *F9* promoter.

3.1 Gross Sequence Changes

A fairly unbiased sample representing >70% of the UK hemophilia B population has shown that gross sequence changes account for about 1.5% of the patients. Most gross mutations are deletions that may involve the whole or part of the *F9* gene. They usually cause severe disease with no measurable FIX protein in circulation. However, in one patient, with deletion of the fourth exon (exon d), significant amounts of circulating FIX have been reported. Such a deletion should not alter the reading frame of the mRNA. Large insertions are very rare and among those of particular interest is the insertion of an Alu repeat, as it raises the issue of ongoing retrotranspositions in the human genome.

3.2 Small Sequence Changes

The 2847 hemophilia B mutations due to small sequence changes (<20 bp) consist of 2646 single base substitutions, 160 deletions, 38 insertions, and 3 combined deletions and insertions. Furthermore, 28 patients showed two, and one patient three *F9* gene mutations, probably, and in some cases certainly, representing the association of a detrimental mutation with additional "private" neutral sequence changes. Excluding such additional changes, 916 unique molecular events have been identified, while the remainder are repeats of previously observed changes. Many of these are CG > TG or CA changes that define mutational hotspots but, in some instances, frequent repeated observation may be due to founder mutations. This probably applies to some of the −6G > A promoter mutations, to the three missense mutations G60S, T296M, and I397T, and to the 17810A > G base substitution in intron 5 that creates a novel donor splice signal. This fits the consensus better than the normal site, is located 13-nt downstream of exon e and causes a mild hemophilia, accounting for 5.5% of hemophilia B patients in the United Kingdom.

3.2.1 Mutations Affecting Transcription, RNA Processing, and Translation

Detrimental promoter mutations should, by definition, impair transcription. Twenty-one different promoter mutations have been reported (Fig. 9). They affect binding sites for different transcription factors (LF.A1/HFN4, DBP, C/EBP), and usually cause a disease characterized by a clinical and hematological age-related improvement (especially postpubertally) that may lead to complete loss of symptoms. However, no age-related improvement has been reported in patients with substitution of residue −26 (G > C, G > A or G > T), which is part of a weak androgen response element (nt −36 to −22) that in its 3′ region overlaps with an LF.A1/HNF4 binding site. It seems probable that the improvement of patients with promoter mutations is due to the greater dependence of the promoter on developmentally related stimulatory events such as increases in androgen levels.

Very high amplification of nucleic acid sequences has allowed examination of tissue-specific mRNAs in tissues that do not express these mRNAs as traces of the relevant mRNA are nevertheless found. However, the *F9* gene, mainly expressed in the liver, is unusual because while trace levels of FIX mRNA are present in lung, skeletal muscle, and heart, no traces of full-length FIX mRNA are found in peripheral lymphocytes, lymphocyte-derived cell lines, or pancreas. Instead, a novel mRNA is found that contains only the last two exons of the *F9* gene. This has prevented thorough analysis of the *F9* mutations that affect mRNA processing. Nevertheless, several mutations that can be expected to affect mRNA splicing have been reported as probable causes of hemophilia B. These include base substitution or deletion of the very highly conserved dinucleotides at the start and end of each intron, alteration of less conserved elements of the splice signal and also the generation of novel splice signals such as that described in the previous section as causing 5.5% of all UK hemophilia B and a new donor splice in the 3′ untranslated tail of the message. The latter has been observed repeatedly in patients with severe disease and appears likely to disrupt the processing of the 3′ end of the message.

Mutations affecting translation, that is, frameshifts and nonsense codons, are scattered throughout the FIX coding sequence. Interestingly, these mutations, irrespective of their location, are usually associated with absence of FIX protein in circulation. Thus, even translation stops very close to the end of the protein (e.g. codon 411) have been found in patients

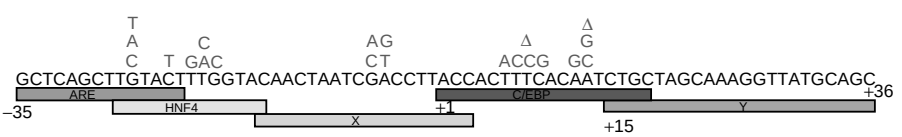

Fig. 9 Hemophilia B mutations affecting the promoter of *F9*. Mutant residues are in red, Δ indicates single base deletion. Bars below promoter sequence indicate mutation-affected regions known to bind transactivating factors. ARE = androgen responsive element, HNF4 = liver-specific transcription factor binding site, C/EBP = binding site for members of the C/EBP family of transactivating factors, X = region binding HNF4 and members of the steroid hormone receptor super family (ARP1, Coop/Ear3) that appear to exercise repressor activity on *F9* promoter. (See color plate p. xxiv.)

with no circulating FIX. This suggests that even moderately truncated FIX and/or FIX mRNA containing premature stop codons are unstable.

3.2.2 Missense Mutations and Single Amino Acid Deletions

Twelve different single aa deletions and 549 substitutions have been reported so far plus five aa substitutions that are probably or certainly neutral R-44H, I-40F, A147T, F178L, and V328I. The above 549 missense mutations affect 225 residues of FIX (Fig. 10) plus seven codons that contribute to splice consensuses. It is difficult to know whether these seven mutations mainly act by changing an amino acid in FIX or by impairing mRNA formation.

The leader peptide is relatively lightly affected and its prodomain shows mutations essentially in a region important for its cleavage from the mature protein. The Val-$_{10}$, which is important for the binding of the γ-carboxylase that modifies the gla residues, has shown two substitutions (A-10T and A-10V) that do not cause hemophilia B but are associated with hypersensitivity to warfarin, an antagonist of vitamin K, which is the cofactor of γ-carboxylase. Several residues of the gla domain are affected by mutation, including gla residues (except those at positions 15, 36, and 40) and the Cys_{18} and Cys_{23}, which form a disulfide bridge. All together 55 missense mutations have been found in this domain, which contains 38 amino acids. The EGF type B domain shows 66 different missense mutations affecting, among others, residues involved in the high affinity Ca^{++} binding site and the cysteines of the three disulfide bridges characteristic of these domains. The EGF type A domain shows a lower number of missense mutations (43), many of which affect the six cysteines involved in disulfide bridges. The activation domain shows relatively few mutations and these are concentrated in regions of obvious functional importance such as Cys_{132}, which is part of a disulfide bridge that holds together the light and heavy chain of activated FIX and its immediate neighbors; the residues of the cleavage sites necessary for the excision of the activation peptide; and the first residues of the heavy chain of activated FIX that are part of the catalytic domain. Three hundred and nine different amino acid substitutions are distributed throughout the catalytic domain, and bear witness to the fairly highly constrained nature of

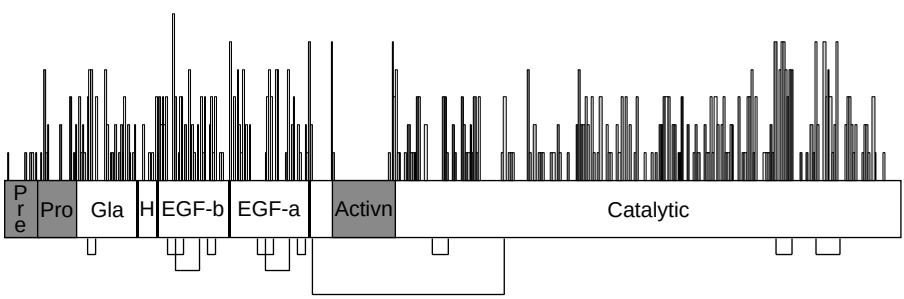

Fig. 10 Missense mutations of *F9* associated with hemophilia B. The horizontal bar shows the domains of the FIX polypeptide with its disulfide bridges indicated below. The vertical lines mark the positions of missense mutations and their heights are proportional to the number of different substitutions (from 1 to 6) found at each residue (data from http://www/kcl.ac.uk/ip/petergreen/haemBdatabase.html).

this domain. The highest concentration of missense mutations is found in the region comprising Ser_{365} of the active center and also in a region comprising residues that form the entrance to the activated FIX specificity pocket. The α-helix comprising residues 330 to 338 is also rich in missense mutations. A number of loops border the substrate-binding groove. Of these, two: (residues 199–204 and 223–229) seem mutation-poor, while the others (235–245, 256–268, 312–322, and 340–347) show average to high numbers of mutations.

In general, the missense mutations found in hemophilia B show a strong preference for residues that are conserved in the homologs of FIX: FVII, FX, and protein C. Thus, 291 different substitutions have been found at the 107 residues that are absolutely conserved, 113 at the 73 amino acids that show only one or two conservative changes, 58 at the 45 residues that show one nonconservative change, and 90 at the 229 less-conserved amino acids.

The current list of hemophilia B mutations also identifies residues that are mandatory for the structural and/or functional integrity of FIX because different types of substitutions are associated with hemophilia B. Thus, 53 residues show two different substitutions, 48 show three, 32 show four, 8 show five, and 5 show six. The last two groups comprise Glu_{33}, Pro_{55}, Cys_{88}, Cys_{95}, Cys_{132}, Arg_{145}, Arg_{180}, Cys_{222}, Cys_{361}, Asp_{364}, Ser_{365}, Thr_{380}, and Ser_{384}. The abundance of FIX residues that are known to have suffered multiple substitutions also suggests that the data available on hemophilia B mutations already form a fairly good representation of the mutational spectrum of hemophilia B. We have also calculated the target for single base substitutions causing hemophilia B. This comprises 825 substitutions of which 189 are expected to cause nonsense and 636 missense mutations. The current list of observed hemophilia B mutations in fact contains 73 different nonsense and 554 different missense mutations that represent 38.6% and 87%, respectively, of the total number expected.

3.3 Functional Interpretation of Observed Sequence Changes and Correlations between Genotype and Phenotype

The criteria that can be used to decide whether an *F9* sequence change observed in a hemophilia B patient is the cause of the disease are essentially those already mentioned for mutations of the *F8* gene and hemophilia A. The lack of a convenient source of FIX mRNA of course prevents the analysis of most mRNA processing defects and the clear identification of intronic mutations capable of causing hemophilia. Fortunately, the systematic analysis of hemophilia B populations indicates that no more than 1–2% of the *F9* mutations are likely to be in regions other than the proximal promoter, exons, and normal RNA processing signals. Thus, the lack of mRNA precludes the rapid detection of only a very small proportion of the hemophilia B mutations.

As in hemophilia A, gross deletions, frameshifts, and nonsense mutations are usually associated with a severe phenotype and predisposition to the inhibitor complication. The latter was recognized as early as 1983 in hemophilia B. Patients with hemophilia B bearing identical mutations appear to show less phenotypic variation than analogous groups of patients with hemophilia A. Comparison of FIX coagulant activity and protein concentration in blood may indicate whether mutations

merely affect the function or also the secretion and/or stability of Factor IX. Thus, the prepeptide mutation I-30N, which alters one of the essential features of the prepeptide (its hydrophobic core) and may thus affect secretion, and also substitutions at cysteines involved in disulfide bridges, which may affect stability, are associated with low protein levels. Conversely, mutations affecting key functional residues such as Asp_{369} and Ser_{365} show normal protein levels.

4 Considerations Arising from the Study of Mutations in Hemophilia A and B

It is instructive to compare some features of the molecular genetics of hemophilia A and B. These two diseases are clinically indistinguishable and their differential diagnosis requires specific coagulant assays. Nevertheless, once hemophilic patients are separated into the A and B groups, some differences emerge. In hemophilia A, a much larger proportion of patients is severely affected. This is explained by the *F8* gene proclivity to disruptive inversions involving the *int1h* and *int22h* sequences. This difference in the hemophilias' mutational spectra (see Figs. 4 and 8) also largely explains why 20% of hemophilia A and only 3% of hemophilia B patients develop inhibitors during the course of replacement therapy.

The amino acid sequence of FIX appears much more constrained than that of FVIII because while 4.2% of all *F8* missense mutations are thought to be able to cause hemophilia A, that is so for 21.2% of the *F9* missense mutations. Since the B domain of FVIII represents 39.3% of the FVIII polypeptide and does not participate in coagulation, the above difference is largely but by no means entirely due to this domain. Thus, the apparent correspondence between the ratio of the sizes of *F8* and *F9* and the ratio of the incidence of hemophilia A and B appears to result from three main factors: the predisposition of the *F8* gene to disruption by inversion, the presence in FVIII of the large B domain that tolerates amino acid substitutions, and the more constrained structure of FIX.

The study of an unbiased sample of hemophilia B patients representing 72% of the UK population has allowed a direct estimate of the hemophilia B mutation rate and of the rates of different types of mutations. Thus, transitions at CpG sites appear to occur at the rate of 97×10^{-9} per nucleotide per generation, transitions at other sites at the rate of 7.3×10^{-9}, transversion at CpG sites at 5.4×10^{-9}, and transversions at other sites at 7.0×10^{-9}. Hence, the rate of transitions at CpG sites is 13-fold greater than other base substitutions. While this explains the deficit of CpG doublets observed in the human genome, it is still insufficient to explain the proportion of hemophilia B (34.8%) or hemophilia A (32.3%) mutations due to transitions at CpG sites. A full explanation must consider natural selection at essential sites of the coding regions of *F8* and *F9*. At these sites, selection will oppose the tendency to lose CpG doublets. Consequently, CpG sites are relatively more frequent at critical sites of gene coding regions, and their mutations are, therefore, particularly likely to cause disease. In keeping with this idea, we observe that of the 96 base substitutions that can occur at CpG sites in the coding sequence of the *F9* gene, 58, or 60%, are expected to cause hemophilia B, while only 767, or 24% of the 3179 base

substitutions at other sites are expected to cause the disease.

5
Mutational Heterogeneity and Genetic Counseling in the Hemophilias

The great mutational heterogeneity of hemophilia A and B is a problem with regard to the carrier and prenatal diagnoses that are required for precise genetic counseling.

In the hemophilias, coagulant assays often do not provide definite carrier diagnoses because the vagaries of X chromosome inactivation result in broad variation of FVIII or FIX values in carriers and considerable overlap between carrier and normal females.

The analysis of the intrafamilial segregation of intragenic polymorphic markers may provide definite carrier diagnoses as well as first trimester prenatal diagnoses but it inevitably fails in a large proportion of cases. The diagnostic failures are due to the limited informativity of available polymorphic markers in the population of interest, to the logistic problem of obtaining samples from all required individuals, and especially to the large proportion of families that have a single affected individual. For example, in the United Kingdom and Sweden, approximately 50% of the families with hemophilia A or B have a single affected individual and the polymorphic marker approach is of very modest value in these families.

Only the direct detection of the gene defect can offer maximum diagnostic success. Therefore, we have argued that ideal strategies to optimize genetic counseling in the field of hemophilia should be based on the construction of national confidential databases of mutation and pedigree information. This can be achieved by characterizing the mutation of an index individual in each family and collecting his/her pedigree. Such national databases are now being constructed in the United Kingdom and other developed countries. They, of course, provide information on the molecular biology of the disease and allow very rapid carrier and prenatal diagnoses on the blood relatives of the index individuals because such diagnoses can be based simply on the analysis of the region of the gene that is defective in the index person.

A further concern in genetic counseling is the issue of somatic or gonadal mosaicism for the hemophilia A or B mutations. Since mutations in the *F8* or *F9* gene may arise postzygotically, individuals who have suffered such a mutational event may have a mixture of cells that either have or have not a hemophilia mutation. If this mixture is present in somatic tissues, the individual is a somatic mosaic, while if the mixture is confined to germline tissue he/she is a gonadal mosaic. Anecdotal evidence for somatic mosaicism is available for both hemophilia A and B, and evidence for gonadal mosaicism has been presented for Duchenne muscular dystrophy. In general, mosaic status will be found in an individual who is the first in the family to carry the relevant mutation, but exceptionally a mosaic may result from a zygote with two or more X chromosomes such that the X chromosome carrying a preexisting hemophilia mutation is lost from some of the cells during postzygotic development.

Obviously, it is impossible to predict the risk of a mosaic mother transmitting the mutant gene to her offspring as this will depend on the proportion of eggs carrying the mutant gene and their chance of maturation. Genetic counseling

will therefore generally have to rely on estimation of average empiric risks. This can be obtained from careful, unbiased population studies. Thus, in the UK population, we investigated how many hemophilia B patients or carriers could be attributed to gonadal mosaicism in mothers designated as noncarriers when their leucocyte DNA was tested. Since we observed no patients or carriers from 47 informative births in the UK hemophilia B population, we concluded that the empiric risk of a "noncarrier" mother of a hemophilia B patient manifesting as a gonadal mosaic by transmitting the mutation to a second child was less than 0.062.

In the case of detectable somatic mosaicism, the upper limit of risk given should be that appropriate to carrier mothers, and, if requested, prenatal diagnosis should be undertaken.

Of course, the chance of detecting somatic mosaics depends on how efficient the laboratory test is in identifying the mutant DNA mixed with that of wild type. For this purpose, mutation detection methods based on chemical mismatch cleavage or denaturing gradient high-performance liquid chromatography are preferable to sequencing.

6
Progress in the Treatment of Hemophilia

In the second half of the last century, therapies based on concentrates of FVIII and FIX prepared from human plasma increased the patients' life expectancy to near-normal values. However, in the 1980s, HIV-infected blood products created a cohort of patients suffering from this lethal disease. Similarly, high rates of hepatitis C infection have marred the progress achieved with the introduction of blood-derived coagulant concentrates. Moreover, recently, concern has arisen about the transmission through blood products of the Creutzfeldt–Jakob prion disease.

Fortunately, the early cloning of the *F9* and *F8* genes has allowed the development of industrial procedures for the production of both FIX and FVIII from mammalian cell cultures created to express these factors *in vitro* so as to minimize risks from the above-mentioned viruses and prions.

In the last 18–20 years, gene therapy has been an important treatment aim. The hemophilias appear well suited to this type of approach because there is no need to ensure tight regulation of the expression of the corrective gene, and any site of expression allowing the coagulant factor to reach the blood circulation should be suitable. Moreover, the enhancement of coagulant level required for the success of this approach is relatively modest because even an increase of 1–2% of the normal level can produce important clinical effects such as transforming severe disease into moderate disease and reducing the number of spontaneous bleeding episodes experienced by the patient. A substantial body of work has been done to develop suitable expression constructs, using retroviral vectors or vectors derived from DNA viruses such as adenovirus or adenovirus-associated virus, and systems of delivery for both *in vivo* or *ex vivo* approaches to treatment (i.e. the injection of expression constructs *in vivo* or the modification with the expression constructs of the patient's cells *ex vivo* followed by culture and expansion *in vitro* and then reinjection into the patient).

Model animals were then used to test various therapy schemes and some

experiments were positive enough to allow initiation of clinical trials, mainly in the United States. Thus, 13 hemophilia A patients, known to have HIV and hepatitis C, and expected to be at very low risk of developing inhibitors, were infused with amphotrophic retroviral vectors for FVIII expression at doses between 3×10^7 and 9×10^8 viral particles (vp) per kilogram; no significant benefit ensued.

Eight hemophilia B patients were injected intramuscularly with an associated adenovirus vector of serological type 2 (AAV2) for FIX expression and one patient receiving the lowest dose was reported to show 1–2% FIX levels for 40 months, but at higher viral doses, no significant FIX level was observed and in all patients the development of antiviral antibodies precluded readministration of the vector.

Administration of factor IX constructs in AAV2 vectors via the hepatic artery in two patients at a dose of 10^{12} vp/kg resulted in transient FIX levels of 10–12%.

In a further trial, autologous skin fibroblasts from six patients were transformed by electroporation with an FVIII expressing plasmid and expanded *in vitro* so as to inject $1–4 \times 10^8$ cells in the patients' greater omentum. Four patients showed FVIII levels above pretreatment values but these returned to baseline values in all patients after 12 months.

Thus, clinical trials have not yet led to satisfactory treatment schedules but since they have not revealed significant toxicity in the patients treated so far, they leave the door open for future advances. These, of course, will always have to consider patients' safety as paramount because established replacement therapy is reasonably effective in both hemophilias and any new treatment must achieve better results with no increase in risk.

See also Heterochromatin and Eurochromatin – Organization, Packaging, and Gene Regulation.

Bibliography

Books and Reviews

Bloom, A.L., Forbes, C.D., Thomas, D.P., Tuddenham, E.G.D. (1994) *Haemostasis and Thrombosis*, 3rd edition, Churchill Livingstone, Edinburgh, Scotland.

Lee, C.A. (1996) *Haemophilia*, Vol. 2, Baillière's Clinical Haematology 9, Baillière Tindall, London, UK.

Mannucci, P.M., Tuddenham, E.G. (2001) The haemophilias – from royal genes to gene therapy, *N. Engl. J. Med.* **344**, 1773–1779.

Oldenburg, J., Brackmann, H.H., Hanfland, P., Schwaab, R. (2000) Molecular genetics of haemophilia A, *Vox Sanguinis* **78**(Suppl. 2), 33–38.

Tuddenham, E.G.D., Cooper, D.N. (1994) *The Molecular Genetics of Haemostasis and its Inherited Disorders*, Oxford Monographs on Medical Genetics 25, Oxford University Press, Oxford, UK.

Primary Literature

Bagnall, R.D., Waseem, N., Green, P.M., Giannelli, F. (2002) Recurrent inversion breaking intron 1 of the factor VIII gene is a frequent cause of severe hemophilia A, *Blood* **99**, 168–174.

Bagnall, R.D., Waseem, N., Green, P.M., Colvin, B., Lee, C., Giannelli, F. (1999) Creation of a novel donor splice site in intron 1 of the factor VIII gene leads to activation of a 191 bp cryptic exon in two haemophilia A patients, *Br. J. Haematol.* **107**, 766–771.

Brandstetter, H., Bauer, M., Huber, R., Lollar, P., Bode, W. (1995) X-ray structure of clotting factor IXa: active site and module structure related to Xase activity and

hemophilia B, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9796–9800.

Chen, S-W.W., Pellequer, J-L., Schved, J-F., Giansily-Blaizot, M. (2002) Model of a ternary complex between activated factor VII, tissue factor and factor IX, *Thromb. Haemost.* **88**, 74–82.

Crow, J.F. (1999) The odds of losing at genetic roulette, *Nature* **397**, 293–294.

Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F., Kazazian, H.H. Jr. (1991) Isolation of an active human transposable element, *Science* **254**, 1805–1808.

Evatt, B. (2000) Creutzfeldt-Jakob disease and haemophilia: assessment of risk, *Haemophilia* **6**(Suppl. 1), 94–99.

Fijnvandraat, K., Turenhout, E.A., van den Brink, E.N., ten Cate, J.W., van Mourik, J.A., Peters, M., Voorberg, J. (1997) The missense mutation Arg593 > Cys is related to antibody formation in a patient with mild hemophilia A, *Blood* **89**, 4371–4377.

Fressinaud, E., Mazurier, C., Meyer, D. (2002) Molecular genetics of type 2 von Willebrand disease, *Int. J. Hematol.* **75**, 9–18.

Giannelli, F., Green, P.M. (2000) The X chromosome and the rate of deleterious mutations in humans, *Am. J. Hum. Genet.* **67**, 515–517.

Giannelli, F., Anagnostopoulos, T., Green, P.M. (1999) Mutation rates in humans. II. Sporadic mutation-specific rates and rate of detrimental human mutations inferred from haemophilia B, *Am. J. Hum. Genet.* **65**, 1580–1587.

Giannelli, F., Choo, K.H., Rees, D.J., Boyd, Y., Rizza, C.R., Brownlee, G.G. (1983) Gene deletions in patients with haemophilia B and anti-factor IX antibodies, *Nature* **303**, 181–182.

Gitschier, J., Wood, W.I., Goralka, T.M., Wion, K.L., Chen, E.Y., Eaton, D.H., Vehar, G.A., Capon, D.J. (1984) Characterisation of the human factor VIII gene, *Nature* **312**, 326–330.

Green, P.M., Rowley, G., Giannelli, F. (2003) Unusual expression of the *F9* gene in peripheral lymphocytes hinders investigation of *F9* mRNA in hemophilia B patients, *J. Thromb. Haemost.* **1**, 2675–2676.

Green, P.M., Saad, S., Lewis, C.M., Giannelli, F. (1999) Mutation rates in humans. I. Overall and sex-specific rates obtained from a population study of haemophilia B, *Am. J. Hum. Genet.* **65**, 1572–1579.

Green, P.M., Waseem, N.H., Bagnall, R.D., Giannelli, F. (1998) Mutation analysis and genetic service: the construction and use of national confidential databases of mutations and pedigrees, *Genet. Test.* **1**, 181–188.

Hakeos, W.H., Miao, H., Sirachainan, N., Kemball-Cook, G., Saenko, E.L., Kaufman, R.J., Pipe, S.W. (2002) Hemophilia A mutations in the factor VIII A2-A3 subunit interphase destabilize factor VIIIa and cause one-stage/two-stage activity discrepancy, *Thromb. Haemost.* **88**, 781–787.

High, K.A. (2001) Gene transfer as an approach to treating haemophilia, *Circ. Res.* **88**, 137–144.

Lakich, D., Kazazian, H.H., Antonarakis, S.E., Gitschier, J. (1993) Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A, *Nat. Genet.* **5**, 226–241.

Lenting, P.J., Neels, J.G., van den Berg, B.M.M., Clijsters, P.P.F.M., Meijerman, D.W.E., Pannekock, H., van Mourik, J.A., Mertens, K., van Zonneveld, A.J. (1999) The light chain of factor VIII comprises a binding site for low density lipoprotein receptor-related protein, *J. Biol. Chem.* **274**, 23734–23739.

Levinson, B., Kenwrick, S., Gamel, P., Fisher, K., Gitschier, J. (1992) Evidence for a third transcript from the human factor VIII gene, *Genomics* **14**, 585–589.

Levinson, B., Kenwrick, S., Lakich, D., Hammonds, G. Jr., Gitschier, J. (1990) A transcribed gene in an intron of the human factor VIII gene, *Genomics* **7**, 1–11.

Li, X., Scaringe, W.A., Hill, K.A., Roberts, S., Mengos, A., Careri, D., Pinto, M.T., Kasper, C.K., Sommer, S.S. (2001) Frequency of recent retrotransposition events in the human factor IX gene, *Hum. Mutat.* **17**, 511–519.

Miyata, T., Hayashida, H., Kuma, K., Mitsuyasa, K., Yasunaga, T. (1987) Male-driven molecular evolution: a model and nucleotide sequence analysis, *Cold Spring Harbor Symp. Quant. Biol.* **58**, 863–867.

Moussalli, M., Pipe, S.W., Hauri, H.P., Nichols, W.C., Ginsburg, D., Kaufman, R.J. (1999) Mannose-dependent endoplasmic reticulum (ER)-Golgi intermediate compartment-53-mediated ER to Golgi trafficking of coagulation factors V and VIII, *J. Biol. Chem.* **274**, 32539–32542.

Naka, H., Brownlee, G.G. (1996) Transcriptional regulation of the human factor IX promoter by the orphan receptor superfamily factors, HNF4, ARP1 and COUP/Ear3, *Br. J. Haematol.* **92**, 231–240.

Naylor, J.A., Green, P.M., Rizza, C.R., Giannelli, F. (1992) Factor VIII gene explains all cases of haemophilia A, *Lancet* **340**, 1066–1067.

Naylor, J.A., Brinke, A., Hassock, S., Green, P.M., Giannelli, F. (1993) Characteristic mRNA abnormality found in half the patients with severe haemophilia A is due to large DNA inversions, *Hum. Mol. Genet.* **2**, 1773–1778.

Naylor, J.A., Green, P.M., Montandon, J.A., Rizza, C.R., Giannelli, F. (1991) Detection of three novel mutations in two haemophilia A patients by rapidly screening whole essential regions of the factor VIII gene, *Lancet* **337**, 635–639.

Naylor, J.A., Buck, D., Green, P.M., Williamson, H., Bentley, D.R., Giannelli, F. (1995) Investigation of the factor VIII intron 22 repeated region (*int22h*) and the associated inversion functions, *Hum. Mol. Genet.* **4**, 329–333.

Nichols, W.C., Seligsohn, U., Zivelin, A., Terry, V.H., Colette, H.E., Wheatley, M.A., Moussalli, M.J., Hauri, H.P., Ciaravella, N., Kaufman, R.J., Ginsburg, D. (1998) Mutations in the ER-Golgi intermediate compartment protein ERGIC-53 cause combined deficiency of coagulation factors V and VIII, *Cell* **93**, 61–70.

O'Donovan, M.C., Oefner, P.J., Roberts, S.C., Austin, J., Hoogendoorn, B., Guy, C., Speight, G., Upadhyaya, M., Sommer, S.S., McGuffin, P. (1998) Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection, *Genomics* **52**, 44–49.

Oldenburg, J., Quenzel, E.M., Harbrecht, U., Fregin, A., Kress, W., Muller, C.R., Hertfelder, H.J., Schwaab, R., Brackmann, H.H., Hanfland, P. (1997) Missense mutations at ALA-10 in the factor IX propeptide: an insignificant variant in normal life but a decisive cause of bleeding during oral anticoagulant therapy, *Br. J. Haematol.* **98**, 240–244.

Peerlinck, K., Jacquemin, M.G., Arnout, J., Hoylaerts, M.F., Gilles, J.G., Lavend'homme, R., Johnson, K.M., Freson, K., Scandella, D., Saint-Remy, J.M., Vermylen, J. (1999) Antifactor VIII antibody inhibiting allogeneic but not autologous factor VIII in patients with mild hemophilia A, *Blood* **93**, 2267–2273.

Perera, L., Darden, T.A., Pedersen, L.G. (2001) Modelling human zymogen factor IX, *Thromb. Haemost.* **85**, 596–603.

Peters, M.F., Ross, C.A. (2001) Isolation of a 40-kDa Huntingtin-associated protein, *J. Biol. Chem.* **276**, 3188–3194.

Pittman, D.D., Wang, J.H., Kaufman, R.J. (1992) Identification and functional importance of tyrosine sulfate residues within recombinant factor VIII, *Biochemistry* **31**, 3315–3325.

Pratt, K.P., Shen, B.W., Takeshima, K., Davie, E.W., Fujikawa, K., Stoddard, B.L. (1999) Structure of the C2 domain of human factor VIII at 1.5 Å resolution, *Nature* **402**, 439–442.

Ragni, M.V. (2002) Safe passage: a plea for safety in hemophilia gene therapy, *Mol. Therapy* **6**, 436–440.

Rizza, C.R., Spooner, R.J.D. (1983) Treatment of haemophilia and related disorders in Britain and Northern Ireland during 1976–1980: report on behalf of the directors of haemophilia centres in the United Kingdom, *Br. J. Haematol.* **286**, 929–933.

Rowley, G., Saad, S., Giannelli, F., Green, P.M. (1995) Ultrarapid mutation detection by multiple, solid-phase chemical cleavage, *Genomics* **30**, 574–582.

Schwaab, R., Brackmann, H.H., Meyer, C., Seehafer, J., Kirchgesser, M., Haack, A., Olek, K., Tuddenham, E.G.D., Oldenburg, J. (1995) Haemophilia A: mutation type determines risk of inhibitor formation, *Thromb. Haemost.* **74**, 1402–1406.

Stoilova-McPhie, S., Villoutreix, B.O., Mertens, K., Kemball-Cook, G., Holzenburg, A. (2002) 3-Dimensional structure of membrane-bound coagulation factor VIII: modeling of the factor VIII heterodimer within a 3-dimensional density map derived by electron crystallography, *Blood* **99**, 1215–1223.

Stoylova, S.S., Lenting, P.J., Kemball-Cook, G., Holzenburg, A. (1999) Electron crystallography of human blood coagulation factor VIII bound to phospholipids monolayers, *J. Biol. Chem.* **274**, 36573–36578.

Toole, J.J., Knopf, J.L., Wozney, J.M., Sultzman, L.A., Brecker, J.L., Pittman, D.D., Kaufman, R.J. (1984) Molecular cloning of a cDNA encoding human antihaemophilic factor, *Nature* **312**, 342–347.

Vielhaber, E., Jacobson, D.P., Ketterling, R.P., Liu, J.Z., Sommer, S.S. (1993) A mutation in the 3′ untranslated region of the factor IX gene in four families with hemophilia B, *Hum. Mol. Genet.* **2**, 1309–1310.

Walsh, C.E. (2003) Gene therapy progress and prospects: gene therapy for the haemophilias, *Gene Ther*. **10**, 999–1003.

Yoshitake, S., Schach, B.G., Foster, D.C., Davie, E.W., Kurachi, K. (1985) Nucleotide sequence of the gene for human factor IX (antihemophilic factor B), *Biochemistry* **24**, 3736–3750.

Heterochromatin and Eurochromatin – Organization, Packaging, and Gene Regulation

Boris A. Leibovitch and Sarah C. R. Elgin
Washington University, St. Louis, MO, USA

1 Introduction 139

2 Cytological Observations 140

3 Biochemical Characteristics of Heterochromatin 143
3.1 DNA in Heterochromatin 143
3.2 Modification of DNA in Heterochromatin 145
3.3 Histone Modifications and Chromatin Structure 146
3.4 Specific Heterochromatin Proteins 147

4 Heterochromatin Behavior 149
4.1 Self-association 149
4.2 Late Replication 150
4.3 Recombination 150

5 Conclusions: Euchromatin/Heterochromatin Relationships 150

Bibliography 152
Books and Reviews 152
Primary Literature 152

Keywords

DNA Modification
Enzymatic addition of chemical groups to DNA occurring after DNA synthesis; in eukaryotes, most commonly methylation of cytosine.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

Epigenetics
A mode of gene regulation, based not on the DNA sequence itself but on stable packaging, heritable during cell division.

Heterochromatin
The silenced domains of chromatin, characterized as maintaining a condensed form, defined by a set of distinctive biochemical properties and cytological behaviors.

Heterochromatin Protein 1 (HP1)
A conserved protein predominantly associated with pericentric heterochromatin, capable of interacting both with a specifically modified histone and with the enzyme that generates the modification.

Histones
A family of low molecular weight proteins, very basic and evolutionarily conserved, which form the core structure for nucleosome assembly in chromatin formation.

Nucleosomes
The basic repeating subunit of chromatin consisting of a tetramer of histones H3 and H4 plus two dimers of histones H2A and H2B, with ca. 170 bp of DNA wrapped around in two left-handed turns; each particle connects to the next by a short stretch of linker DNA, which interacts with histone H1.

Posttranslational Modification of Proteins
An enzymatic addition or removal of various chemical groups to proteins (including histones) after synthesis, including acetylation, methylation, phosphorylation, and so on.

Repetitive DNA
The DNA sequences that are represented by multiple copies in the genome; the copies may be in tandem arrays or scattered as individual copies.

Transposable Elements
Various families of moderately repetitive DNA sequences originally capable of transposition; their distribution and numbers vary between different individuals and strains.

The genomes of eukaryotes are packaged in chromatin, a complex of DNA with histones in a nucleosomal array, and folded further through interactions with other proteins into higher order domains. The different levels of packaging help define whether genes are available for expression, as in euchromatin, or generally silenced, as in heterochromatin. The condensed heterochromatin domains are characterized

by a group of features, including low gene density, high levels of repetitious sequences, location at the nuclear periphery, late replication, and little recombination during meiosis, apparently reflecting a particular pattern of posttranslational modifications of the histones, association of heterochromatin-specific proteins, and regular spacing of nucleosomes.

1 Introduction

In 1928, Heitz observed that the staining of liverwort *Pellia epiphylla* interphase nuclei is not uniform, suggesting the presence of two fractions of chromatin as manifested by weakly stained areas and areas that maintain the intense staining of metaphase chromosomes. Weakly stained areas were designated as *euchromatin*, while the dark positively heteropycnotic areas were labeled *heterochromatin*. The fate of these regions was traced from interphase to metaphase and back to interphase; it was observed that these regions remained differentially stained in almost all phases of the cell cycle, showing a common level of condensation only in metaphase. The most prominent heterochromatic regions are located mainly in or close to the centromeres and telomeres of chromosomes. The development of better microscopes and application of various genetic, biochemical, and molecular approaches, coupled with the examination of different species in similar studies, have provided a much deeper understanding of the underlying mechanisms generating this different packaging. Table 1 describes the major differences between these two chromatin fractions, giving both cytological and biochemical characteristics. The reader should bear in mind that the differences described reflect the sum of many observations made in a wide range of different organisms, cell types, and so on. Heterochromatic regions may manifest all or only some of these features. It seems likely that a similar condensed chromatin organization can be achieved in different ways, using redundant molecular mechanisms. Recent data also indicate that some of the features characteristic of heterochromatin may be utilized to downregulate the expression of genes in euchromatic domains. This implies the existence in the nucleus of a continuum of chromatin states, potentially generated by various combinations of the features listed in Table 1.

Some of the properties of heterochromatin are reversible; several must be transiently altered or re-created following replication, indicating the presence of various mechanisms to accomplish this with fidelity. This, in turn, implies that the formation of heterochromatin (or heterochromatin-like structures) may be used as a mechanism of epigenetic inheritance, maintaining particular states (silent or activatable) of gene expression. The heterochromatin located around centromeres, telomeres, and some other regions maintained in that form in all cell types is usually referred as *constitutive heterochromatin*. Regions that show a cell-type-specific (or otherwise selected) heritable silencing of particular genes, chromosome domains or whole chromosomes are referred to as *facultative heterochromatin*. Many questions remain as to how constitutive and

Tab. 1 Distinctions between euchromatic and constitutive heterochromatic regions of complex genomes.

Feature	*Euchromatin*	*Heterochromatin*
Proportion of genome	50–80%	20–50%
Interphase appearance	Decondensed (lightly stained)	Condensed (densely stained – heteropycnotic)
Chromosomal location	Distal arms	Usually pericentromeric, telomeric
Nuclear location	General	Often clumped (nuclear periphery, around nucleolus)
Sequence complexity	High; unique DNA, interspersed middle repetitious sequences	Low; repetitious DNA: satellite DNA, blocks of middle repetitious sequences
Gene density	High	Low
DNA methylation	CpG islands hypomethylated	Fully methylated (in mammals and plants)
Histone modification;	High histone acetylation; specific amino acids are methylated	Low histone acetylation; specific amino acids are methylated
Nucleosome spacing	Irregular	Regular
Nuclease accessibility	Variable	Low
Mechanism of gene silencing	Individual genes	Large domains
Meiotic recombination	Normal	Absent or significantly decreased
Replication timing	Throughout S phase	Late S phase

facultative domains are targeted for packaging in a heterochromatic form, a process sometimes referred to as *heterochromatinization*. This article will summarize our current knowledge concerning the differentiation of chromatin into these two fractions, and the possible role of heterochromatin in genome organization and regulation of gene expression.

2 Cytological Observations

Observations using light microscopy showed that the dark heteropycnotic chromatin visible in interphase nuclei corresponds to the strongly stained domains in metaphase chromosomes surrounding centromeres and telomeres. Small, intensely stained islands within the chromosome arms, referred to as *interstitial* or *intercalary heterochromatin*, have been observed in chromosomes of some organisms. In interphase, the strongly stained domains lie at the periphery of the nucleus, close to the nuclear membrane. These observations were confirmed subsequently by immunostaining with antibodies to proteins specifically enriched in heterochromatin (see Sects. 3.3 and 3.4), and by *in situ* hybridization with DNA probes complementary to the repetitive DNA fractions that are concentrated around centromeres. Analyses using electron microscopy (EM) to examine interphase nuclei have also demonstrated the presence of two chromatin fractions, with the electronically dense, tightly packaged fraction located mainly at the nuclear periphery and around the nucleolus.

The patterns seen on EM photographs, by light microscopy after staining with dyes or immunostaining, or with *in situ*

hybridization of repetitious sequences, all indicate that heterochromatin in the interphase nucleus is organized in several large blocks, which tend to localize at the nuclear membrane. This suggests that the heterochromatic regions of different chromosomes in the interphase nucleus interact with each other, as the number of spots observed is smaller than the total number of centromeres and other heterochromatic regions. Careful analysis of the spatial organization of chromosomes within the nucleus of some organisms has shown that the distribution of heterochromatin is not random. The centromeres and surrounding pericentromeric heterochromatin tend to concentrate on one side of the nucleus, while telomeric regions of the same chromosomes are clustered on the opposite side. This phenomenon is often referred to as *Rabl orientation*, honoring Carl Rabl, who described it first in 1885. The positioning of the heterochromatic telomeres and centromeres implies that the euchromatic regions in between also assume defined positions within the nuclei. Direct observations have confirmed the existence of the so-called chromosomal territories with limited fixed positions for different chromosomal segments in specific areas of the nucleus.

While constitutive heterochromatin (e.g. pericentromeric and telomeric regions) is present in an invariant pattern at given sites and in given chromosomes, facultative heterochromatin is represented by chromosomal regions or whole chromosomes that are capable of existing in one state in some cells and in the other state in other cells. In some organisms, such as the mealy bug *Planococcus citri*, the full haploid set of paternal chromosomes is found in the heterochromatic state. Mammals demonstrate the well-known example of X-chromosome inactivation in female nuclei. The inactivated X-chromosome forms the so-called Barr body, usually identified as a condensed mass at the periphery of female nuclei. The number of Barr bodies correlates well with the number of inactivated X-chromosomes in cases in which the individual carries more than two X-chromosomes. Identification of the inactive X-chromosome as the condensed Barr body led to the suggestion that the formation of heterochromatin might be a mechanism for downregulating gene expression. In the normal situation, this inactivation serves to compensate for the differences in dosage of X-chromosomal genes between females, with two X-chromosomes, and males, with single X-chromosome. X-chromosome inactivation serves to diminish the negative effect of additional X-chromosomes, and such aneuploids survive, while most cases of human aneuploidy (extra copies of a chromosome) are lethal. While the decision of which X chromosome to inactivate appears to be random in somatic cells, the decision once made is stably inherited through multiple rounds of mitosis, a critical observation indicating that changes in chromatin structure play a role in epigenetic inheritance. In these cases, gene activity is determined not by the DNA sequence *per se*, but by the particular packaging of the gene in chromatin.

A second phenomenon demonstrating the link between gene inactivation and the location and packaging of a gene in a heterochromatic form is position effect variegation (PEV), described in Drosophila, yeast, and mammals. In these cases, a gene normally located in a euchromatic environment is positioned adjacent to or within heterochromatin by chromosomal rearrangement or transposition (Fig. 1).

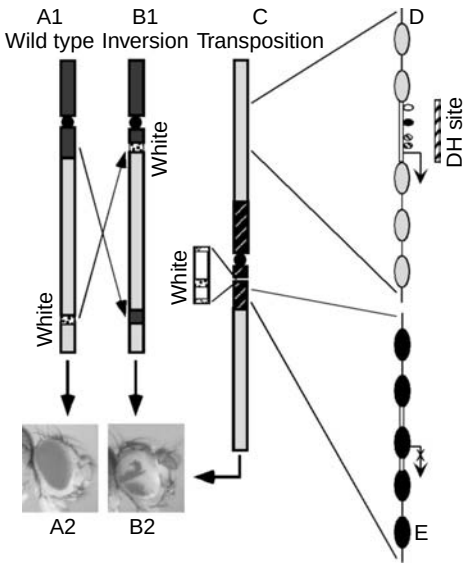

Fig. 1 Position effect variegation in Drosophila. (A1, A2) The wild type *white* gene (spotted rectangle) is required to generate red fly eyes; it is localized in a euchromatic segment (light grey) of the X chromosome. (B1, B2) Inversions of the chromosome segment with one breakpoint in pericentromeric heterochromatin (black) and one breakpoint adjacent to *white* position the gene adjacent to heterochromatin, leading to inactivation of the gene in some eye progenitor cells, resulting in a variegating phenotype. (C, B2) The same result is observed when *white* is transposed into heterochromatin as a part of a transposon. (D) Genes in euchromatic domains are generally packaged in an irregular pattern of nucleosomes (grey ovals); one observes accessible (open, nucleosome free) sites at or near the 5′ end of genes, allowing interaction with transcription factors (small circles) that can regulate initiation of transcription from the start site (bent arrow). Such open sites are often called DH sites (DNase I hypersensitive sites) or HS sites (hypersensitive sites) because they are readily cleaved by nucleases. (E) In heterochromatic domains, the nucleosomes (black ovals) have a characteristic, different pattern of histone modification, and interact with a subset of chromosomal proteins, some specific to these domains. Within heterochromatic regions, the nucleosomes show very regular spacing, and appear to mask promoters and other regulatory sites, inhibiting transcription (crossed bent arrow).

In both cases, one observes that the gene is stochastically inactivated in some cells, but remains active in others, resulting in a variegating phenotype; the active or inactive status, once set, is mitotically inherited in the progeny of the cells, while the phenotype, at least in yeast, is stably maintained over many generations, that is, through meiotic divisions. The variegating gene in such situations is now found located close to the nuclear periphery, with heterochromatin, at a higher frequency. Inactivation of some genes during normal development may occur concomitantly with their spatial relocation to heterochromatin domains. Genetic analysis has shown that a screen can identify mutations in other loci that suppress or enhance the pattern

of variegation, observed as a loss in the reporter gene silencing or an increase in reporter gene silencing. Such second-site modifiers have frequently been found to be genes that encode chromosomal proteins, or proteins required for chromatin assembly and function. The study of such suppressors and enhancers of variegation [*Su(var)*s and *E(var)*s] has greatly facilitated studies of the biochemical properties of heterochromatin and the resultant epigenetic marks.

3 Biochemical Characteristics of Heterochromatin

3.1 DNA in Heterochromatin

Kinetic analysis of the renaturation of genomic DNA has shown that the genomes of higher eukaryotes are made up not only of unique sequences (present in one copy per haploid genome), which renature slowly, but also of repetitious sequences (present in multiple copies per haploid genome), which renature more quickly. Constitutive heterochromatin is enriched in repetitive DNA and has relatively little unique DNA. Most of the genes encoding proteins are present in a single copy. Genetic analysis has identified only a few loci that map within the pericentric heterochromatin. Sequencing of several eukaryotic genomes has confirmed the conclusion that heterochromatin is gene-poor, and provided more precise knowledge about the distribution and features of DNA in both chromatin fractions. Interestingly, analysis of the DNA sequences from Drosophila has predicted the presence of at least ~100 genes embedded in the pericentric heterochromatin; thus, while it is gene-poor, the region is not devoid of genes.

Repetitive DNA may be roughly divided into two classes based on the copy number – middle repetitive and highly repetitive DNA. Genes encoding the large ribosomal RNAs, 5S RNA, and histones fall into the middle repetitive class in most eukaryotic organisms. However, the majority of the middle repetitive DNA consists of intact and/or damaged transposable elements. In some organisms, including humans, up to ca. 50% of the DNA is derived from transposable elements, representing different families, including pro-retroviruses. These repeats can be scattered as single copies, found at apparently random sites within genomic DNA, or may be grouped in tandem or inverse arrays. Sometimes different transposable elements are found nested together, one inside of another. Just as unique DNA sequences are characteristic of euchromatin, a high frequency of clusters and/or nested copies of moderate repeats/transposable elements is also a feature of heterochromatin. A significant fraction of the transposable elements, particularly those copies in heterochromatin, carry various mutations that render them incapable of transposing.

Several investigations have suggested that the presence of repetitious elements may serve as a signal for targeting formation of heterochromatin. For example, the *Drosophila melanogaster* gene *white* is normally found at a euchromatic site and is fully expressed, giving flies a red eye; however, a tandem array of *white* genes, generated from a transgene, results in inactivation, regardless of the site of insertion within the chromosomes, leading to a variegated eye phenotype. This variegation, similar to PEV, is the result of stochastic inactivation of the cluster

of *white* genes in some progenitor cells of the eye in which this gene is normally active. However, the presence of repetitive DNA is not sufficient to trigger the formation of heterochromatin in all cases. For example, the histone genes, repetitive in most eukaryotes, are not packaged in cytologically visible heterochromatin, nor do they manifest other biochemical characteristics of heterochromatin. However, the repetitious rRNA genes, found in the nucleolus, can manifest heterochromatic properties, including the silencing of marker genes inserted into this domain.

The second class of repeats, present almost exclusively in heterochromatin, is highly repetitive DNA, also known as *satellite* DNA. Satellite DNA consists of short sequences (from ca. 10 bp to a few hundred bp) repeated in tandem arrays up to a million times. Some satellite DNA fractions are chromosome-specific, while others occur on all chromosomes, though in varying amounts. In some species, including the kangaroo rat *Dipodomys ordii* and some beetles from the coleopteran family *Tenebrionidae*, satellite DNA accounts for the majority of the genomic DNA. If the base composition of a given repeat is GC- or AT-rich, that DNA may be separated from the bulk of the DNA in a cesium chloride (CsCl) density gradient, giving a separate "satellite" peak, heavier or lighter than the bulk DNA, respectively. Single copies or clusters of transposable elements often interrupt such arrays. Cloning and sequencing of the DNA in the boundary regions between heterochromatin and euchromatin (often called transition zones) has shown that this part of the genome consists of a considerable mix of both types of repeats, with an intermediate density of genes.

The pericentromeric location of the most highly repeated sequences raises the question of their possible participation in the formation of the centromere (the constriction that serves as a platform for assembly of kinetochore proteins) and the kinetochore (the complex of proteins to which the nuclear spindle fibers (microtubules) attach to accomplish chromosome segregation). The matter is controversial, with several competing points of view. Some evidence has suggested that satellite DNA is both necessary and sufficient to form a centromere, while other experiments indicate that there is no requirement for a particular DNA sequence. In the latter view, the centromere is a specialized chromatin structure maintained in an epigenetic pattern. This structure may arise in different ways, for example, with the help of specific proteins such as the centromere-specific histone CENP-A, a histone H3 variant.

In addition to pericentromeric heterochromatin, differentially stained regions that have many characteristics of heterochromatin are found near or at telomeres. In most organisms, the telomere consists of $(TG)_n$ repeats that interact with a group of specialized proteins, creating a specific chromatin structure. The system enables telomeres to support DNA replication without any loss of DNA due to unidirectional copying, to prevent the fusion of chromosome ends, and to direct the chromosome ends to sites close to the nuclear membrane. Localization is thought to involve interactions between telomere proteins and the nuclear lamina, a structure just below the inner nuclear membrane. Drosophila differs in having clusters of particular transposable elements at the ends of the chromosomes, flanked downstream by different satellite-like repeats – telomere-associated sequence (TAS). It is unclear

which feature – DNA sequence and/or proteins – makes telomeric regions in all organisms heterochromatic. Studies of genes affecting PEV of visible markers transposed into or close to telomeres have shown that while many of the same genes affect silencing in telomeres and mating-type loci in *Saccharomyces cerevisiae*, different groups of genes influence telomeric and pericentric PEV in Drosophila. This implies a different chromatin structure in telomeric and pericentromeric heterochromatin in this organism.

3.2 Modification of DNA in Heterochromatin

In mammals and plants (but not in Drosophila or yeast), there is a strong correlation between heterochromatic regions of chromosomes and the degree of DNA methylation. Historically, the role of DNA methylation in gene silencing and formation of heterochromatin was first demonstrated by the observation that DNA of the inactive X-chromosome is hypermethylated. Methylation in mammals usually occurs behind the replication fork at position C5 of cytosine bases in CpG dinucleotides. C-residues in other sequences have been found to be modified, but at a lower frequency. This modification does not occur exclusively within heterochromatic domains. Many euchromatic genes have CpG-rich elements (so-called "CpG islands") in their 5′ upstream region, which are important for the regulation of the gene through interaction with the transcriptional regulatory protein Sp1. Methylation of these islands is usually an indicator of gene silencing; presumably, methylation blocks binding by Sp1 to its target sequence. Removal of 5′mC (for example by replicating DNA in the presence of 5-azacytidine, a form of C that cannot be methylated) can reactivate these genes. The same drug can activate silenced genes on the inactive X-chromosome in female mammalian nuclei.

DNA methylation is a reversible process. Only the parental strand of any newly replicated DNA double helix will contain the methylated cytosine. In the usual situation, this remaining 5′mC will interact with 5′mC-binding proteins that direct site-specific enzymes, the maintenance DNA methyltransferases, to methylate the cytosine of the CpG on the opposite DNA strand. If this interaction is blocked, or if this interaction and subsequent modification fails for any reason, the DNA of half of the granddaughter cells will lose 5′mC at this site. However, a second group of DNA methyltransferases has been identified that are capable of *de novo* methylation of cytosines on the nonmodified DNA template. These DNA methylases restore methylation of some genes and in heterochromatin after the massive demethylation of DNA that occurs as part of normal early development in mammals. We do not know, however, how these enzymes are targeted to the sequences that need to be so modified. Some sequences that one might expect to be silenced are not demethylated during this general event, among them proviral sequences of IAP particles (which resemble transposable elements in the mouse genome) and imprinted genes. Imprinted genes are expressed specifically from the chromosome inherited from the mother or the chromosome inherited from the father, with the allele from the other parent being silenced. Not surprisingly, these genes in mammals exhibit differential DNA methylation; the silenced genes show additional features of facultative heterochromatin.

3.3 Histone Modifications and Chromatin Structure

The different cytological appearance of heterochromatin has led to the suggestion that the molecular organization and DNA packaging in the underlying chromatin must be different. Data from the last few years has shed significant light on this subject. Chromatin is a chain of repeating subunits (nucleosomes) of ca. 200 bp, with approximately 146 bp of DNA folded around a histone octamer (core) consisting of a tetramer of $(H3 + H4)_2$ and two dimers of (H2A + H2B). These core particles are joined together by short (20–60 bp) linkers, which are generally associated with histone H1, other linker histones such as H5, or HMG proteins. Nucleosome location can generally be regarded as sequence-independent, although DNA characteristics such as bendability can influence nucleosome positioning. Histones interact with each other to form the octamer and bind DNA primarily through their histone-fold domains, leaving their N-terminal tails available for other interactions. Presumably these interactions stabilize the higher-order folding of the basic chain into 30 nm fibers and higher order packaging arrays. Posttranslational modification of the histones plays a major role in determining their interactions with other chromatin components (DNA, other histones, and non-histone proteins), and hence in stabilizing or destabilizing higher order structures.

The most common *in vivo* modifications are acetylation, methylation, phosphorylation, ubiquitination, or polyADP-ribosylation. Most of the sites of modification are clustered in the N-terminal regions, changing the total charge and hydrophobicity of a given histone tail. In addition, several sites of modifications have been described in the central histone-fold motifs. Histone modifications are catalyzed by enzymes that introduce or remove certain groups at one or more positions. The enzymes are often present as part of large multimeric chromatin-remodeling complexes, targeted by associated gene-specific activators and repressors. The very complex pattern of histone modifications has led to the idea of a histone "code," defining different packaging states. In particular, modification of histone H3 by methylation at lysine 9 (giving H3-mK9) has been associated with packaging into heterochromatin, while acetylation in general, and acetylation of H4 at lysine 5 and lysine 8 specifically, are associated with packaging in euchromatic form.

A second type of complex, the ATP-dependent chromatin remodeling complex, defines nucleosome formation behind the replication fork and/or destabilizes nucleosomes in an already mature nucleosomal array. All remodeling complexes of this type include a subunit possessing a DNA- and/or nucleosome-inducible ATPase activity. Various versions of these complexes are simultaneously present in a given cell, modified by the temporary addition or removal of different subunits. Such complexes are involved in both activation and repression of many genes. The exact composition of the complexes, their targets for action, and their mechanism(s) for localization at the appropriate targets are all currently under investigation. Complexes of both types, capable of histone modification and of ATP-dependent remodeling, are essential for gene activation, as even within euchromatic domains nucleosome arrays render many enhancer

and promoter sites inaccessible to regulatory proteins and/or general transcription factors and RNA polymerase. In addition, nucleosomes suppress RNA elongation during transcription.

Nucleosome arrays in heterochromatin are structurally similar but not identical to those in euchromatin. Digestion of chromatin with micrococcal nuclease reveals a more regular spacing in heterochromatic arrays, with a narrower cleavage site in the linker. Heterochromatic regions, overall, are less accessible to both general and site-specific nucleases, and to most other reagents that interact with DNA. These properties suggest that some characteristic features of nucleosome structure, positioning, and mobility are different in heterochromatin and euchromatin, presumably reflecting differences in internucleosomal interactions as well as differences in the associated chromosomal proteins. These changes may be due to the various histone modifications mentioned above. Certainly the average pattern of histone modifications differs between heterochromatin and euchromatin, creating a distinct signature for each domain using the histone code. However, the histone modification pattern is likely to vary for different genes in a given domain (e.g. for euchromatic genes that are actively being transcribed vs. not being transcribed; for inducible vs. uninducible genes, etc.), and for different types of heterochromatin, creating a continuum of chromatin states. The discovery of the patterns of posttranslational histone modifications, with the realization that enzymes responsible for the modifications had in some cases been previously identified as activators or repressors of gene expression, was one of major recent breakthroughs in the chromatin field.

The differences in heterochromatic nucleosome arrays, specifically the more regular spacing, have suggested that specific chromatin-remodeling complexes might be involved in assembling heterochromatin. *In vitro* assays have demonstrated that two Drosophila chromatin remodeling complexes, CHRAC and ACF, can generate arrays with heterochromatin-like regular spacing. The human WSTF–ISWI and ACF complexes appear to be the mammalian homologs of CHRAC and ACF, based both on their composition and by their ability to space nucleosomes regularly on DNA. These complexes are concentrated in heterochromatic regions in late S-period, the time when the bulk of the heterochromatin is replicated.

3.4 Specific Heterochromatin Proteins

Heterochromatin is concentrated in the pericentric and telomeric regions, and is rich in satellite DNA and other repetitious DNA. Several different proteins that specifically interact with satellite DNA have been found in vertebrates and in Drosophila. Most of these proteins interact as well with homologous sequences scattered in euchromatic regions, and some also function in the regulation or chromatin organization of euchromatic genes. Most of these proteins interact with other proteins to form multimeric complexes. For example, Drosophila GAGA factor, the protein specifically interacting with $(GA)_n/(CT)_n$ short repeats in the promoter regions of many genes, regulating their expression, is found also to interact with $(AAGAG)_n$ and $(AAGAGAG)_n$ satellite DNA fractions, concentrated in the pericentromeric heterochromatin of several chromosomes. This protein has been found in several multimeric complexes, and interacts

with the chromatin-remodeling complex NURF. Interestingly, mutations in GAGA factor enhance PEV of genes transferred to heterochromatin, suggesting that its dominant role is in support of gene expression. Whether GAGA factor association has an impact on the chromatin structure of the satellites where it binds is yet to be demonstrated.

The first protein specifically characterized as a heterochromatin protein, Drosophila Heterochromatin Protein 1 (HP1), binds poorly to DNA, with little specificity. HP1 is thought to play a structural role in heterochromatin, since an antipodal dosage dependence is observed: PEV is suppressed by mutations in the gene for HP1 that result in a loss of functional HP1, and is enhanced by an increase in the amount of HP1, for example from a duplication or a transgene. This type of action is also observed for a few other chromosomal proteins, *SU(VAR)3–9* (an H3 methyl-transferase), *SU(VAR)3–7* (a zinc finger protein), and HP2 (a potential structural protein). Homologs of HP1 have been found in the yeast *Schizosaccharomyces pombe*, worms, other insects, amphibians, mammals, and plants. Full inactivation of HP1 through loss or missense mutations, or by the expression of intracellular antibodies, is lethal. From one to three HP1 homologs have been identified in different organisms, but the proteins are not identical, nor can they substitute for each other in a genetic analysis; all are required. Immunocytological analyses indicate that the homologs have different patterns of nuclear distribution, with only one of them restricted to pericentric heterochromatin. The three proteins in mammals can physically interact with each other. In transgenic Drosophila, expression of the genes for the HP1 human homologs also enhances PEV, suggesting a conserved functional interaction with the Drosophila machinery that establishes and maintains heterochromatin, including the originally identified HP1, encoded by *Su(var)205*.

The HP1's are small proteins, ca. 200 to 250 amino acids, with two conserved domains joined by a more variable hinge region. The N-terminal chromo domain was first recognized by homology with a conserved domain of Polycomb, a protein required for developmentally maintained silencing of homeotic genes in the euchromatin. The chromo domain has since been identified in several other chromosomal proteins. HP1 can dimerize through interactions involving the chromo shadow domain, and this structure appears to form a platform for interactions with many other chromosomal proteins, including *SU(VAR)3–9* and HP2.

The binding of HP1 to nucleosomes is facilitated by methylation of lysine 9 in histone H3, a histone modification characteristically enriched in heterochromatin. The modified histone tail fits nicely within the fold generated by the HP1 chromo domain. Missense mutations that lead to a loss of *Su(var)* activity (loss of silencing) are mutations that result in a loss of this stable interaction. This H3 modification is driven by one (or perhaps more) specific histone methyltransferase encoded by homologous genes found in mammals, yeast (*S. pombe*) and Drosophila. The founding member in Drosophila is *SU(VAR) 3–9*. As the name implies, mutations of the gene encoding this protein suppress the PEV in Drosophila. This enzyme in mammals and flies forms complexes with HP1 and often colocalizes with HP1 in heterochromatic regions of chromosomes. The ability of HP1 to interact with both the modified histone (H3-mK9) and the enzyme responsible for the modification

suggests a mechanism for epigenetic inheritance of the modification pattern, and for the spreading of a heterochromatic domain once initiated.

HP1 can also interact with various proteins that are associated with repressor activity in euchromatic domains. The best-studied cases are in mammals. Gene repression is generally accompanied by histone deacetylation and, at least in some cases, by methylation of histone 3 at lysine 9. For example, a complex of mammalian HP1, SUV39H1, and the transcriptional repressor Rb has recently been described. Retinoblastoma protein (Rb) plays an important role in the induction of the G1-S transition during the cell cycle; mutations of Rb have been associated with several tumors. It has been suggested that Rb acts as a negative regulator by binding SUV39H1 (the human homolog of *SU(VAR)3–9*), resulting in local methylation of lysine 9 of histone H3, thereby recruiting HP1 to "fix" the gene in an inactive state. The process in this case is highly specific, involving only one nucleosome located on the promoter element of the cyclin D gene, repressed by Rb *in vivo*. Rb has also been found in complexes with histone deacetylase. It is unclear whether both silencing complexes are essential for complete inactivation, or whether cells and genes may choose between these modes of inactivation, one (deacetylation) being easily reversible, the other (methylation) being stable through generations. Since SUV39H1 cannot methylate acetylated lysine 9 in histone H3, the order of events seems obvious in this case. However, it is unclear whether the process can stop after the first step (deacetylation) without progressing to the next step (H3 methylation and binding of HP1). These data (and more) suggest that heterochromatin proteins play roles beyond the packaging of heterochromatic domains, being involved in regulating the activity of targeted euchromatic genes. It is unclear, however, how these processes are managed to limit the silenced domain to one or a few nucleosomes, whereas heterochromatic domains appear able to spread over large distances (at least 10 kb, observed in *S. pombe*).

4 Heterochromatin Behavior

4.1 Self-association

Many of the proteins enriched in heterochromatin, including HP1, are able to form homomeric and heteromeric complexes. This no doubt contributes to the association of heterochromatic regions in a small number of blocks, easily visible on a cytological level. This property of heterochromatic proteins may be utilized as a mechanism for gene inactivation. The best studied case involves the Ikaros family proteins, recently analyzed in mammalian pre-B cells. Several homologous proteins of this family have two domains, one participating in protein–protein interactions and the other in DNA binding. The Ikaros proteins show sequence-specific binding to DNA and form homo- and heteromeric complexes with each other. Most of the target sequences are multiple short repeats comprising a satellite DNA fraction concentrated in pericentric heterochromatin. However, identical short repeats are present in the regulatory regions of many genes. Ikaros suppresses the activity of the terminal deoxynucleotidyl transferase gene at the appropriate stage of lymphocyte differentiation by competing with a gene-specific activator for a partly overlapping binding site. Reduced chromatin accessibility for restriction enzymes and DNA

methylation are observed at the promoter of the repressed gene. Importantly, the inactivated gene is placed within or next to a heterochromatin block, most likely due to the formation of multimeric complexes between the Ikaros protein bound at the promoter and that bound in the pericentric heterochromatin. Whether or not silencing is dependent on this relocalization remains to be seen. Similar observations of relocalization have been made for some other genes in this cell lineage.

4.2 Late Replication

Early observations on incorporation of labeled precursors during DNA synthesis showed that the diffuse weakly staining euchromatin was replicated primarily during the first half of S-phase, while the heteropycnotic heterochromatic regions showed incorporation of label primarily during the second half of S-phase. These observations were confirmed by direct analysis of replication times for different DNA sequences: those enriched in constitutive heterochromatin were replicated at the end of S-phase. The same observations have held true for chromosomes and for separate chromosomal domains that showed the characteristics of facultative heterochromatin. For example, the inactive X-chromosome replicates late in S-phase, while the active X replicates early. It is still unclear what drives this replication schedule. There are many possibilities, ranging from the idea that the densely packaged heterochromatin may be less accessible to the replication machinery, to the suggestion that replication of heterochromatin requires proper amounts of specific chromatin-remodeling complexes such as WSTF–ISWI and ACF.

4.3 Recombination

Careful genetic mapping in several eukaryotes has shown that the frequency of recombination is usually significantly decreased toward centromeres and telomeres. Crossing-over is very low near heterochromatin domains in general. Similar observations have been made for regions located within or close to the silenced mating-type loci in yeast (*S. cerevisiae*), domains that appear heterochromatic by the criteria discussed above (silencing of transposed reporter genes, low accessibility to nuclease digestion, lack of histone acetylation, etc.). We do not know why heterochromatic regions are less susceptible to recombination. The ability of heterochromatic regions to "stick together" should facilitate the exchange of genetic material. In fact, spontaneous or induced chromosomal rearrangements with breakpoints in heterochromatin are seen more frequently then those with breakpoints in euchromatin. Of course, this may be explained simply as a consequence of the more frequent survival of the resulting progeny, due to the low gene density in regions of constitutive heterochromatin. A plausible explanation might lie in the inaccessibility of densely packed chromatin in this region to the recombinational machinery.

5 Conclusions: Euchromatin/Heterochromatin Relationships

The above discussion has provided many examples that point to the differences between heterochromatin and euchromatin, while at the same time emphasizing that no one single feature provides

a unique identifier of heterochromatin. A constellation of several characteristics is required to define a region as heterochromatic, and this combination may be different in different heterochromatic domains. For example, HP1 is commonly associated with pericentric heterochromatin, but is not associated with the inactive X-chromosome in mammals. Furthermore, different organisms may employ different strategies to generate a heterochromatic domain – one that has a regular chromatin structure, is inaccessible to a variety of probes, and is lacking in meiotic recombination. For example, while many plants and animals (notably mammals) utilize DNA methylation as a heritable mark of heterochromatin formation and gene silencing, Drosophila, yeast, and many other organisms do not appear to exhibit significant levels of DNA methylation. While organisms from the yeast *S. pombe* to Drosophila and humans utilize a strategy of H3-K9 methylation and HP1 association to form heterochromatin, the yeast *S. cerevisiae* does not. It is clear that higher eukaryotes have a number of histone methyltransferases that can modify H3-K9, and that these play both overlapping and distinct roles. Histone H3 and H4 hypoacetylation does seem to be a common strategy, but one used in different ways in different organisms. There has been some speculation that the range of biochemical mechanisms used by an organism to maintain heterochromatin structure, and concomitant silencing, may reflect the size of its genome. This may be correlated with the load of retroviruses and transposable elements – sequences that need to be silenced – commonly carried in the genomes of organisms with higher C values.

While heterochromatin is gene-poor, it is not devoid of genes. Models of heterochromatin structure must be able to accommodate the observation that the activity of these genes depends on the presence of normal heterochromatin structure; inactivation of HP1 leads to a loss of function of these genes. At the same time that we have noted specific biochemical features associated with heterochromatin formation, we have found that the same features can be found within euchromatic domains. For example, HP1 and H3-mK9 have been found not only in pericentric heterochromatin, but are also utilized in repressive regulation of some normal eukaryotic genes. Thus, one definition for *heterochromatic genes* may be those that are dependent on HP1 for optimal expression, whereas *euchromatic genes* might be defined as those exhibiting a loss of expression on association with HP1. As more genes are discovered that map to heterochromatic domains, it will be of interest to see if this pattern holds for other biochemical marks as well.

A critical question now under intense study is that of targeting heterochromatin formation: how is the pattern of differential chromatin packaging established? Recent data indicate that small interfering RNAs may play a role here, in addition to the regulation achieved by the RNA interference system by targeting transcription products for degradation (posttranscriptional silencing). A similar small RNA fraction could interact either with DNA in heterochromatin, for example, repeated sequences or transposable elements, or with proteins enriched in heterochromatin. Some chromodomain-containing proteins, notably MOF, can interact with RNA. Most compelling, mutations in several known genes participating in the formation of such small RNAs affect heterochromatin

silencing and centromeric function in the yeast *S. pombe*. One prerequisite for the formation of such small RNAs is the formation of long double stranded RNA, for example by bidirectional transcription. Interestingly, many transposable elements (commonly overrepresented in heterochromatin domains) are capable of bidirectional transcription. Thus heterochromatin formation may have derived from the need to silence invading transposons and retroviruses, to be adapted later in evolution to help create structures such as the centromere, and to help regulate gene expression throughout the genome. Experiments during the next several years should clarify many of the issues raised above.

See also Cell Nucleus Biogenesis, Structure and Function; Chromosome Organization within the Nucleus; DNA–Protein Interactions; Electron Microscopy in Cell Biology; Gene Mapping and Chromosome Evolution by Fluorescence–Activated Chromosome Sorting; Gene Mapping by Fluorescence In Situ Hybridization; Hemophilia, Molecular Genetics of; Preimplantation Mammalian Development, Regulation of Gene Expression in.

Bibliography

Books and Reviews

Dernburg, A.F., Karpen, G.H. (2002) A chromosome RNAissance, *Cell* **111**, 159–162.

Eissenberg, J.C., Elgin, S.C.R. (2000) The HP1 protein family: getting a grip on chromatin, *Curr. Opin. Genet. Dev.* **10**, 204–210.

Georgopoulos, K. (2002) Haematopoietic cell-fate decisions, chromatin regulation and Ikaros, *Nat. Rev. Immunol.* **2**, 162–174.

Grewal, S.I., Elgin, S.C.R. (2002) Heterochromatin: new possibilities for the inheritance of structure, *Curr. Opin. Genet. Dev.* **12**, 178–187.

Huang, Y. (2002) Transcriptional silencing in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, *Nucleic Acids Res.* **30**, 1465–1482.

Jones, P.A., Takai, D. (2001) The role of DNA methylation in mammalian epigenetics, *Science* **293**, 1068–1070.

Lamb, J.C., Birchler, J.A. (2003) The role of DNA sequence in centromere formation, *Genome Biol.* **4**, 214.1–214.4.

Lyon, M.F. (2002) X-chromosome inactivation and human genetic disease, *Acta Paediatr. Suppl.* **91**, 107–112.

Neely, K.E., Workman, J.L. (2002) The complexity of chromatin remodeling and its links to cancer, *Biochim. Biophys. Acta.* **1603**, 19–29.

Parada, L., Misteli, T. (2002) Chromosome positioning in the interphase nucleus, *Trends Cell Biol.* **12**, 425–432.

Richards, E.J., Elgin, S.C.R. (2002) Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects, *Cell* **108**, 489–500.

Turner, B.M. (2002) Cellular memory and the histone code, *Cell* **111**, 285–291.

Primary Literature

Akhtar, A., Zink, D., Becker, P.B. (2000) Chromodomains are protein-RNA interaction modules, *Nature* **407**, 405–409.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., Kouzarides, T. (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain, *Nature* **410**, 120–124.

Bongiorni, S., Mazzuoli, M., Masci, S., Prantera, G. (2001) Facultative heterochromatization in parahaploid male mealybugs: involvement of a heterochromatin-associated protein, *Development* **128**, 3809–3817.

Bozhenok, L., Wade, P.A., Varga-Weisz, P. (2002) WSTF-ISWI chromatin remodeling

complex targets heterochromatic replication foci, *EMBO J*. **21**, 2231–2241.

Bruvo, B., Pons, J., Ugarkovic, D., Juan, C., Petitpierre, E., Plohl, M. (2003) Evolution of low-copy number and major satellite DNA sequences coexisting in two Pimelia species-groups (Coleoptera), *Gene* **312**, 85–94.

Cheutin, T., McNairn, A.J., Jenuwein, T., Gilbert, D.M., Singh, P.B., Misteli, T. (2003) Maintenance of stable heterochromatin domains by dynamic HP1 binding, *Science* **299**, 721–725.

Collins, N., Poot, R.A., Kukimoto, I., GarcÌa-JimÈnez, C., Dellaire, G., Varga-Weisz, P.D. (2002) An ACF1/ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin, *Nat. Genet.* **32**, 627–632.

Cryderman, D.E., Tang, H., Bell, C., Gilmour, D.S., Wallrath, L.L. (1999) Heterochromatic silencing of Drosophila heat shock genes acts at the level of promoter potentiation, *Nucleic Acids Res.* **27**, 3364–3370.

Czermin, B., Schotta, G., Hulsmann, B.B., Brehm, A., Becker, P.B., Reuter, G., Imhof, A. (2001) Physical and functional association of SU(VAR)3–9 and HDAC1 in Drosophila, *EMBO Rep.* **2**, 915–919.

Farkas, G., Gausz, J., Galloni, M., Reuter, G., Gyurkovics, H., Karch, F. (1994) The *Trithorax-like* gene encodes the Drosophila GAGA factor, *Nature* **371**, 806–808.

Filesi, I., Cardinale, A., van der Sar, S., Cowell, I.G., Singh, P.B., Biocca, S. (2002) Loss of heterochromatin protein 1 (HP1) chromodomain function in mammalian cells by intracellular antibodies causes cell death, *J. Cell Sci.* **115**, 1803–1813.

Hatch, F.T., Mazrimas, J.A. (1974) Fractionation and characterization of satellite DNAs of the kangaroo rat (*Dipodomys ordii*), *Nucleic Acids Res.* **1**, 559–575.

Hediger, F., Neumann, F.R., Van Houwe, G., Dubrana, K., Gasser, S.M. (2002) Live imaging of telomeres. yKu and Sir proteins define redundant telomere-anchoring pathways in yeast, *Curr. Biol.* **12**, 2076–2089.

Heitz, E. (1928) Das heterochromatin der Moose, *Jehrb. Wiss. Botanik* **69**, 762–818.

Hilliker, A.J. (1985) Assaying chromosome arrangement in embryonic interphase nuclei of *Drosophila melanogaster* by radiation induced interchanges, *Genet. Res.* **47**, 13–18.

Hoskins, R.A., Smith, C.D., Carlson, J.W., Carvalho, A.B., Halpern, A., Kaminker, J.S., Kennedy, C., Mungall, C.J., Sullivan, B.A., Sutton, G.G., Yasuhara, J.C., Wakimoto, B.T., Myers, E.W., Celniker, S.E., Rubin, G.M., Karpen, G.H. (2002) Heterochromatic sequences in a Drosophila whole-genome shotgun assembly, *Genome Biol.* **3**, 85.1–85.16.

Hwang, K.K., Eissenberg, J.C., Worman, H.J. (2001) Transcriptional repression of euchromatic genes by Drosophila heterochromatin protein 1 and histone modifiers, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11 423–11 427.

Ishii, K., Arib, G., Lin, C., Van Houwe, G., Laemmli, U.K. (2002) Chromatin boundaries in budding yeast: the nuclear pore connection, *Cell* **109**, 551–562.

James, T.C., Elgin, S.C.R. (1986) Identification of a nonhistone chromosomal protein associated with heterochromatin in *Drosophila melanogaster* and its gene, *Mol. Cell Biol.* **6**, 3862–3872.

James, T.C., Eissenberg, J.C., Craig, C., Dietrich, V., Hobson, A., Elgin, S.C.R. (1989) Distribution patterns of HP1, a heterochromatin-associated nonhistone chromosomal protein of Drosophila, *Eur. J. Cell Biol.* **50**, 170–180.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., Jenuwein, T. (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins, *Nature* **410**, 116–120.

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., Reik, W. (2003) Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse, *Genesis* **35**, 88–93.

Lu, B.Y., Emtage, P.C., Duyf, B.J., Hilliker, A.J., Eissenberg, J.C. (2000) Heterochromatin protein 1 is required for the normal expression of two heterochromatin genes in Drosophila, *Genetics* **155**, 699–708.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., Grewal, S.I.S. (2001) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly, *Science* **292**, 110–113.

Ng, H.H., Ciccone, D.N., Morshead, K.B., Oettinger, M.A., Struhl, K. (2003) Lysine-79 of histone H3 is hypomethylated at silenced loci in yeast and mammalian cells: a potential mechanism for position-effect variegation, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1820–1825.

Nicolas, E., Roumillac, C., Trouche, D. (2003) Balance between acetylation and methylation of histone H3 lysine 9 on the E2F-responsive dihydrofolate reductase promoter, *Mol. Cell Biol.* **23**, 1614–1622.

Nielsen, A.L., Oulad-Abdelghani, M., Ortiz, J.A., Remboutsika, E., Chambon, P., Losson, R. (2001) Heterochromatin formation in mammalian cells: interaction between histones and HP1 proteins, *Mol. Cell* **7**, 729–739.

Noma, K., Allis, C.D., Grewal, S.I.S. (2001) Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries, *Science* **293**, 1150–1155.

Rabl, C. (1885) Uber Zelltheilung, *Morphol. Jahrb.* **10**, 214–330.

Saccani, S., Natoli, G. (2002) Dynamic changes in histone H3 Lys 9 methylation occurring at tightly regulated inducible inflammatory genes, *Genes Dev.* **16**, 2219–2224.

Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., Reuter, G. (2002) Central role of Drosophila SU(VAR)3–9 in histone H3-K9 methylation and heterochromatic gene silencing, *EMBO J.* **21**, 1121–1131.

Shaffer, C.D., Stephens, G.E., Thompson, B.A., Funches, L., Bernat, J.A., Craig, C.A., Elgin, S.C.R. (2002) Heterochromatin protein 2 (HP2), a partner of HP1 in Drosophila heterochromatin, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14 332–14 337.

Smothers, J.F., Henikoff, S. (2001) The hinge and chromo shadow domain impart distinct targeting of HP1-like proteins, *Mol. Cell Biol.* **21**, 2555–2569.

Sun, F.L., Cuaycong, M.H., Elgin, S.C.R. (2001) Long-range nucleosome ordering is associated with gene silencing in *Drosophila melanogaster* pericentric heterochromatin, *Mol. Cell Biol.* **21**, 2867–2879.

Sun, X., Le, H.D., Wahlstrom, J.M., Karpen, G.H. (2003) Sequence analysis of a functional Drosophila centromere, *Genome Res.* **13**, 182–194.

Tamaru, H., Selker, E.U. (2001) A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*, *Nature* **414**, 277–283.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., Martienssen, R.A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi, *Science* **297**, 1833–1837.

Xiao, H., Sandaltzopoulos, R., Wang, H.M., Hamiche, A., Ranallo, R., Lee, K.M., Fu, D., Wu, C. (2001) Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions, *Mol. Cell* **8**, 531–543.

Ye, Q., Callebaut, I., Pezhman, A., Courvalin, J.C., Worman, H.J. (1997) Domain-specific interactions of human HP1-type chromodomain proteins and inner nuclear membrane protein LBR, *J. Biol. Chem.* **272**, 14 983–14 989.

Zhao, T., Heyduk, T., Allis, C.D., Eissenberg, J.C. (2000) Heterochromatin protein 1 binds to nucleosomes and DNA in vitro, *J. Biol. Chem.* **275**, 28 332–28 338.

HIV Therapeutics, Biochemistry of

Raveendra Dayam and Nouri Neamati
University of Southern California, Los Angeles, CA

1 Reverse Transcriptase Inhibitors 157

2 Nucleoside Reverse Transcriptase Inhibitors (NRTIs) and Nucleotide Reverse Transcriptase Inhibitors (NRTIs) 157

3 Nonnucleoside Reverse Transcriptase Inhibitors (NNRTIs) 162

4 Protease Inhibitors 163

5 Integrase Inhibitors 167

6 HIV Entry Inhibitors 168
6.1 Inhibitors of GP120-CD4 Interactions 168
6.2 Chemokine Receptor Antagonists 170
6.3 Viral Fusion Inhibitors 170

7 Other Potential Drug Targets 171
7.1 HIV-1 Nucleocapsid Protein as a Drug Target 171
7.2 Targeting HIV-1 Proteins Vif, Tat, and Rev 171

Bibliography 172
Books and Reviews 172
Primary Literature 172

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

Keywords

HAART
Highly Active Antiretroviral Therapy – an aggressive anti-HIV treatment usually including a combination of three or more drugs that are targeted at different steps in the HIV life cycle.

Multidrug-resistant HIV Strain
A strain of HIV resistance to at least two or more different classes of drugs.

Allosteric Binding Site
A site or a cavity on an enzyme other than its substrate binding site or active site.

Peptidomimetic
A small chemical compound that is designed to mimic a peptide in order to be recognized by an enzyme as its natural substrate.

Nucleoside
Nucleoside is an organic molecule consisting of a heterocyclic base (a purine or a pyrimidine) and a pentose sugar.

Nucleotide
A nucleotide is an organic molecule consisting of a heterocyclic base (a purine or a pyrimidine), a pentose sugar, and a phosphate group.

Transition State
An activated state of a molecule that has partly undergone a chemical reaction. Enzymes are thought to bind and stabilize the transition state, thus lowering the energy of activation needed to drive the reaction to completion.

Lipodystrophy
A disorder of adipose (fatty) tissue characterized by a selective loss of body fat. Patients with lipodystrophy have a tendency to develop insulin resistance, diabetes, a high triglyceride level (hypertriglyceridemia), and fatty liver.

Cytochrome P-450
A superfamily of hundreds of closely related heme proteins found throughout the phylogenetic spectrum, from animals, plants, fungi, to bacteria. In animals, these P-450 enzymes serve two major functions: (1) biosynthesis of steroids, fatty acids, and bile acids and (2) metabolism of endogenous and a wide variety of exogenous substrates, such as toxins and drugs (biotransformation).

Although significant progress has been made in the treatment of HIV infection, we are still far from wining the battle. The combination of highly active antiretroviral therapy (HAART) in compliant patients comes close to stopping virus evolution; however, eradication of the infection has not been achieved because of the persistence of latent HIV-1 in resting memory CD4 + T cells. Moreover, several factors including the emergence of multidrug-resistant HIV strains, drug toxicity, the patient's ability to adhere to the prescribed therapy, and expensive medication have necessitated a reason to develop novel drugs and novel approaches to combat HIV infection.

Currently, the HAART regimen consists of nucleoside and nonnucleoside reverse transcriptase (RT) combined with protease (PR) inhibitors. This article summarizes all currently approved drugs for HIV-1 infection as well as drugs that have made significant progress in clinical trials but have not yet been approved. Where appropriate, their mechanism of action, mechanism of resistance, and toxicity are discussed

1 Reverse Transcriptase Inhibitors

HIV-1 RT is a heterodimeric enzyme of 1000 amino acid residues. RT (reverse transcriptase) has been an attractive and highly explored drug target for AIDS chemotherapy. Of the 19 currently available FDA approved drugs, 11 are targeted at RT. This enzyme plays a vital role in the HIV replication cycle by catalyzing the conversion of viral RNA into a double-stranded, complementary DNA through at least three different enzymatic reactions. In the first step, RT catalyzes the conversion of the single-stranded viral RNA to a double strand of RNA:DNA hybrid using its RNA-dependent DNA polymerase activity. In the second step, RT generates a single-stranded DNA chain complementary to the viral RNA by the hydrolysis of a previously formed strand of RNA:DNA hybrid through its RNase H activity. Finally, using its DNA-dependent DNA polymerase activity, RT generates a second DNA chain to complete the formation of a double-stranded viral DNA. RT inhibitors prevent the formation of double-stranded viral DNA (proviral DNA) from the single-stranded viral RNA.

2 Nucleoside Reverse Transcriptase Inhibitors (NRTIs) and Nucleotide Reverse Transcriptase Inhibitors (NRTIs)

Both NRTIs (nucleoside reverse transcriptase inhibitors) and NtRTIs (nucleotide reverse transcriptase inhibitors) inhibit the catalytic activity of RT by directly interacting with the deoxynucleoside triphosphate (dNTP) substrate binding site of the RT after metabolic conversion to their corresponding 5′-phosphate derivatives (for NRTIs) and diphosphate derivatives (for NtRTIs). NRTIs and NtRTIs are not antiviral as such. They must be phosphorylated by host cellular kinases to form the 5′-triphosphates and are referred to as *prodrugs*. Currently, 7 NRTIs and an NtRTI (Tenofovir) are in clinical use (Figs. 1 and 2). The clinically used NRTIs include zidovudine, stavudine, zalcitabine,

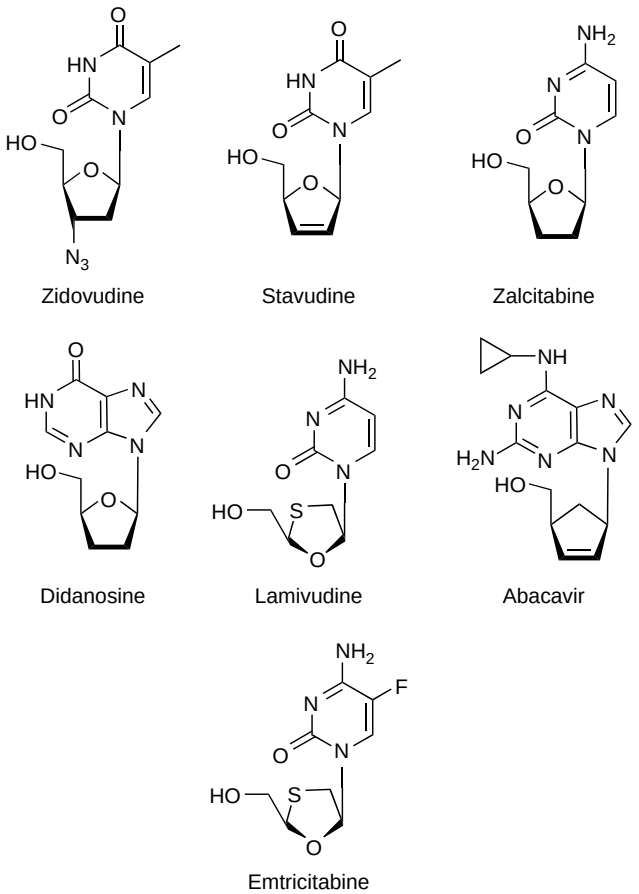

Fig. 1 Structures of FDA approved NRTIs.

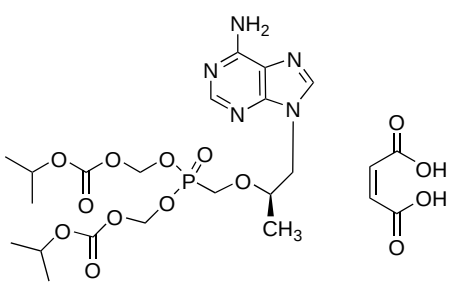

Fig. 2 Structure of Tenofovir disoproxil fumarate.

didanosine, lamivudine, abacavir, and emtricitabine (Table 1). Owing to considerable differences in their structures, NRTIs and the NtRTI utilize different enzymes for intracellular conversion to their active drug forms, and subsequently follow different metabolic pathways to be excreted. For example, thymidine analog, zidovudine (AZT) is converted to its 5′-monophosphate (AZTMP) form by

Tab. 1 Clinically used NRTIs, NtRTIs, and NNRTIs.

Name or brand name	*Marketing company*	*Mechanism of action*	*Activity spectrum*
Zidovudine (AZT), Retrovir®	GlaxoSmithKline	DNA chain termination	HIV-1 and HIV-2
Didanosine (ddI), Videx®	Bristol Myers-Squibb	DNA chain termination	HIV-1 and HIV-2
Zalcitabine (ddC), Hivid®	Hoffmann-La Roche	DNA chain termination	HIV-1 and HIV-2
Stavudine (d4T), Zerit®	Bristol Myers-Squibb	DNA chain termination	HIV-1 and HIV-2
Lamivudine (3TC), Zeffix®	GlaxoSmithKline	DNA chain termination	HIV-1 and HIV-2, HBV
Abacavir (ABC), Ziagen®	GlaxoSmithKline	DNA chain termination	HIV-1 and HIV-2
Emitricitabine (FTC), Emtriva®	Gilead Sciences	DNA chain termination	HIV and HBV
Tenofovir disoproxil fumarate (bis(POC)PMPA), Viread®	Gilead Sciences	DNA chain termination	HIV-1 and HIV-2, various other retroviruses and HBV
Nevirapine, Viramune®	Boehringer Ingelheim	Binding to an "allosteric" pocket, nonsubstrate binding site of HIV-1 RT	HIV-1
Delavirdine, Rescriptor®	Pfizer	Binding to an "allosteric" pocket, nonsubstrate binding site of HIV-1 RT	HIV-1
Efavirenz, Sustiva®, Stocrin®	Bristol Myers-Squibb	Binding to an "allosteric" pocket, nonsubstrate binding site of HIV-1 RT	HIV-1

thymidine kinase. The phosphorylation of AZTMP to AZTDP is catalyzed by cellular thymidylate (dTMP) kinase. Finally, AZTDP is converted to AZTTP by cellular nucleoside 5′-diphosphate (NDP) kinase.

In general, clinically used RT inhibitors are analogs of natural nucleotides (dNTPs) substrates. After metabolic conversion to their corresponding phosphate forms, NRTIs, and NtRTIs compete with natural dNTP substrates for incorporation into the growing DNA chain. As a result of the incorporation into the growing DNA chain by RT, the growth of the DNA chain is terminated. This is due to the lack of a 3′-hydroxyl group on the sugar moiety of NRTIs and NtRTI. The drugs act as DNA chain terminators and inhibit the catalytic activity of RT. All NRTIs and the NtRTI are believed to proceed with a similar mechanism to inhibit RT, and thereby suppress the replication of HIV.

A major clinical problem with the current NRTIs and NtRTI is the emergence of resistant viral strains. The imposed drug pressure and the inherent lack of a proofreading mechanism by RT are the driving forces for emergence of drug-resistant HIV strains. The resultant RT mutations are specific to each NRTI and therefore show decreased sensitivity to the corresponding NRTI. The RT mutations that cause resistance are diverse. Therefore, an RT mutant resistant to a particular NRTI can be inhibited by another NRTI. This condition is strategically exploited by designing combination regimens to effectively bring down viral loads in patients with resistant HIV viral strains. However, cross-resistant RT mutants have also been isolated for a number of NRTIs and NtRTI combinations. The emergence of drug-resistant HIV RT strains is attributed to two different kinds of mechanisms of drug resistance. Specific mutations in the RT substrate binding site can selectively prevent incorporation of the NRTI into the growing DNA chain, while the enzyme still retains the ability to incorporate the natural dNTP substrates. HIV RT achieves this discriminative ability through specific mutations in the substrate binding site that cause steric hindrance to the NRTI, and therefore selectively interfere with its incorporation. The M184V and M184I mutation of RT both confer high-level clinical resistance to lamivudine, and represent a classical example of RT mutation (Table 2). The second known mechanism of drug resistance is the excision of NRTI from the growing DNA chain. The drug resistant RT efficiently incorporates the NRTI into the DNA chain, which temporarily blocks DNA growth. However, the NRTI is subsequently removed (excised) from the blocked DNA chain allowing elongation to proceed. The best-known example of this kind of resistance is the excision of AZTMP, which causes AZT resistance.

A common problem with NRTIs and NtRTI-based treatment is long-term drug related toxicity. Long-term toxicities associated with these drugs are mostly related to the intracellular phosphorylation of the drugs. The treatment with NRTIs and NtRTI is associated with a range of toxicities thought to result from the inhibition of mitochondrial polymerase-γ, leading to impaired synthesis of the mitochondrial enzymes that generate ATP through oxidative phosphorylation (Table 2). All the NRTIs are associated with hyperlactatemia and lactic acidosis. Patients taking stavudine and didanosine seem to be at greater risk than patients taking other NRTIs. The NRTIs and NtRTI mitochondrial toxicity is organ and drug specific, possibly as a result of variability in intracellular drug penetration and the metabolism to the triphosphate form.

Tab. 2 Major drug resistant mutations, toxic side effects, and drug interactions of clinically used NRTIs and NtRTI.

Name	*Mutations associated with resistance*	*Toxic side effects*	*Drug interactions*
Zidovudine	M41L, E44D, D67N, K70R, V118I, L210W, T215Y/F, K219Q/E	Anemia, granulocytopenia, myopathy, hepatic transaminase elevations, headaches, fatigue, nausea	Should not be combined with Stavudine (Zerit®)
Didanosine	K65R, L74V	Peripheral neuropathy, pancreatitis, lactic acidosis, diarrhea, headaches, vomiting, rash, headaches, high blood pressure	Should not be combined with Zalcitabine (Hivid®), Stavudine (Zerit®)
Zalcitabine	K65R, T69D, L74V, M184V	Peripheral neuropathy, pancreatitis, rash, chest pain, fever, nausea, changes in liver function, and mouth sores, headaches, high blood pressure	Should not be combined with Stavudine (Zerit®), Lamivudine (Zeffix®), Didanosine (Videx®)
Stavudine	M41L, E44D, K65R, D67N, K70R, V118I, L210W, T215Y/F, K219Q/E	Lipodystrophy, peripheral neuropathy, lactic acidosis, diarrhea, fever, fatigue, headaches, high blood pressure	Should not be combined with Zidovudine (Retrovir®), Zalcitabine (Hivid®), Didanosine (Videx®)
Lamivudine	E44D, K65R,V118I, M184V/I	Nausea, diarrhea, abdominal pain, headaches, fatigue, neutropenia, nausea, vomiting, sleeping disorders	Should not be combined with Zalcitabine (Hivid®),
Abacavir	K65R, L74V, Y115F, M184V	Hypersensitivity, gastrointestinal disturbance, rash, fever, fatigue, headache, nausea, and vomiting	
Emitricitabine	K65R, M184V/I	Lactic acidosis, fatty liver, skin color changes (rarely), headache, diarrhea, nausea, rash	Should not be combined with Lamivudine (Zeffix®),
Tenofovir disoproxil	K65R	Increased levels of creatine and transaminases, reduced bone-mineral density, nausea, vomiting, and loss of appetite, headaches, high blood pressure	Should not be combined with Lamivudine (Zeffix®), Didanosine (Videx®), Abacavir (Ziagen®)

3
Nonnucleoside Reverse Transcriptase Inhibitors (NNRTIs)

The NNRTIs (nonnucleoside reverse transcriptase inhibitors) prevent HIV replication by binding to a pocket (nonsubstrate or allosteric binding site) that is located in close proximity to the natural dNTP substrate binding site of RT. They do not compete with the natural dNTP substrates for the binding site. The binding of NNRTIs to this pocket induces significant distortion in the substrate binding site of RT. This in turn decreases the catalytic activity of the enzyme. NNRTIs slow down the polymerase activity of RT, and thereby suppress the HIV replication by binding to a nonsubstrate binding pocket on RT. Hence, NNRTIs do not act as DNA chain terminators, and they do not need to be incorporated into the growing DNA chain to inhibit the catalytic activity of RT.

The three NNRTIs approved by the FDA for clinical use are nevirapine, delavirdine, and efavirenz (Fig. 3). Despite their structural diversity, all the NNRTIs show high specificity for HIV-1 RT, over that of HIV-2 RT and other DNA polymerases. The NNRTIs are very different from NRTIs and NtRTIs in many aspects; as (1) they do not bind to the substrate binding site, (2) they do not need to be converted to an active form, (3) they are very selective to HIV-1 RT, and (4) they are less toxic *in vivo*. The critical problem with NNRTIs is the rapid onset of viral resistance when administered alone because of the high mutation rate of RT. Mutations specifically occur at amino acid residues surrounding the NNRTI binding pocket. The most common amino acid mutations that confer resistance to NNRTIs are those that directly interfere with RT drug binding (Fig. 4). For this reason, NNRTIs are clinically used only in combination with other drugs, most often with NRTIs or NtRTIs.

Nevirapine and efavirenz are metabolized by the liver cytochrome P450 family of enzymes. Delavirdine happens to be an inhibitor of the cytochrome P450. Nevirapine and efavirenz are broken down by the liver and can subsequently interact with other drugs metabolized by the liver. The

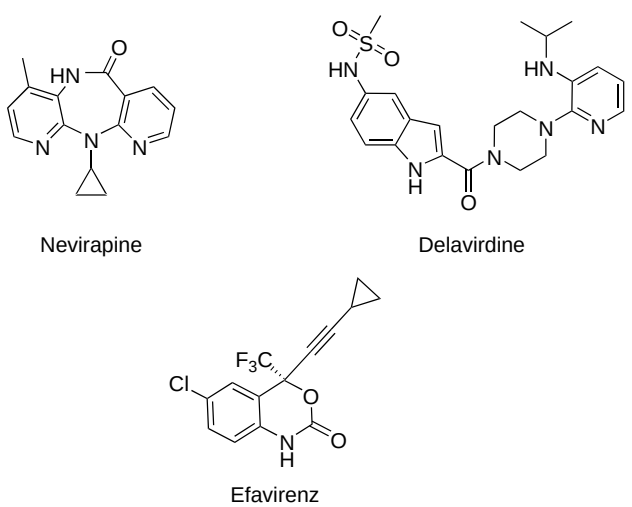

Fig. 3 Structures of FDA approved NNRTIs.

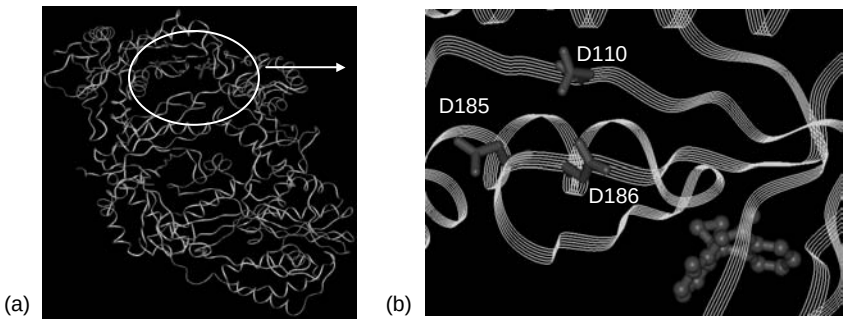

Fig. 4 The crystal structure of HIV-1 RT complexed with nevirapine (PDB1S1X). Nevirapine is a nonnucleoside inhibitor and binds to an allosteric (nonsubstrate) binding pocket. The ribbon model shows RT enzyme (a) while three aspartate residues at polymerase active site of RT and the adjacent nevirapine binding pocket are enlarged. (b) Catalytically important three aspartate residues (D110, D185, and D186) are represented as stick models. The ball and stick model represents the binding mode of nevirapine.

Tab. 3 Observed drug resistant mutations and side effects of NNRTIs.

NNRTIs	*Mutations associated with resistance*	*Toxic side effects*
Nevirapine	L100I, K103N, V106A/M, V108I, Y181C/I, Y188C/L/H, G190A	Skin rash, Stevens-Johnson syndrome (rare), diarrhea, fever, depression, headache
Delavirdine	K103N, V106M, Y181C, Y188L, P236L	Maculopapular rash, Stevens-Johnson syndrome (rare), headache, gastrointestinal disorders
Efavirenz	L100I, K103N, V106M, V108I, Y181C/I, Y188L, G190S/A, P225H	Central nervous system stimulation, dizziness

serious side effects caused by the clinical use of these drugs are liver damage and other closely related problems. Other common side effects associated with NNRTI treatment are maculopapular rash, hepatic transaminase elevations, and central nervous system complaints such as dizziness (Table 3).

4 Protease Inhibitors

HIV-1 PR is an aspartic protease with two catalytically essential aspartic acid residues in its active site (from each monomer). The catalytically active PR (protease) exists as a homodimer. The two identical subunits of 99 amino acid residues (monomer) are noncovalently associated in a twofold C_2 symmetric fashion (Fig. 5). PR proteolytically cleaves the polyproteins that are encoded by the HIV *gag* and *pol* genes in the final stage of the viral life cycle. Inactivation of PR by either mutational or chemical inhibition leads to the production of immature, noninfectious viral particles. Therefore, the function of PR is essential for proper virion assembly and maturation.

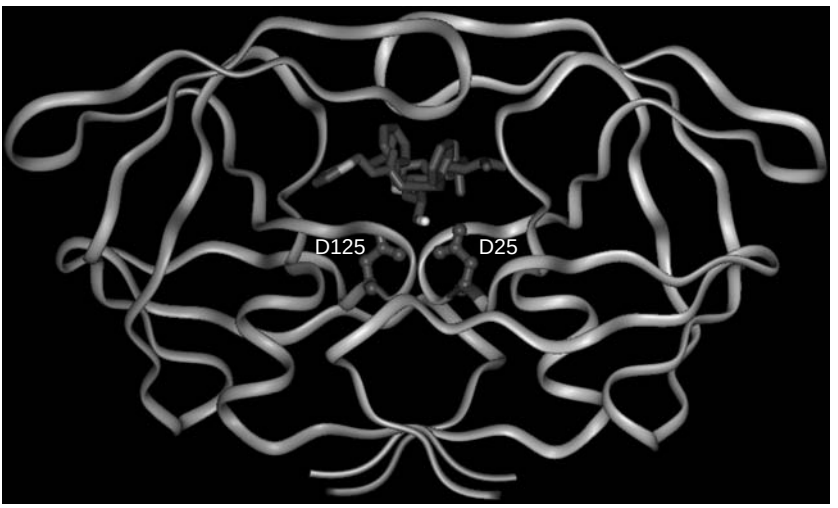

Fig. 5 Crystal structure of HIV-1 protease (PR) complexed with ritonavir (PDB1HXW). The two monomers of the PR are represented as ribbon models (yellow and magenta). The two catalytic aspartate residues (D25 and D125) are rendered as ball and stick models while the ritonavir is represented by a stick model. The hydroxyl group of the ritonavir occupies the space between two aspartate residues. (See color plate p. xxvi).

PR inhibitors (PIs) prevent HIV replication by competitively binding to and inhibiting the proteolytic activity of PR.

The seven peptidomimetic PIs approved by the FDA for clinical use are saquinavir, ritonovir, indinavir, nelfinavir, amprenavir, lopinavir, and atazanavir (Fig. 6). The rationale behind the development of the PIs came from the architecture of the PR active site and the tetrahedral transition state of the bound substrate. Despite their structural diversity, all the clinically used PIs possess a pseudosymmetric diaminoalcohol core unit (Fig. 6). All PIs currently used in therapy contain a hydroxyl group that is correctly positioned between the two catalytic aspartate residues in order to both donate and accept a hydrogen bond (Table 4). This closely resembles the tetrahedral transition state of substrate cleavage (Fig. 5). All PIs are reversible inhibitors that compete with natural substrate. They have a higher binding affinity to PR than peptide substrates because PR binds to the transition state to a greater degree than the substrate itself.

Although PIs belong to a potent class of antiviral drugs, their clinical effectiveness has been hampered by the emergence of drug resistant and cross-resistant PR mutants. Specific mutations at PR active site that affect PI binding are responsible for the emergence of drug-resistant HIV strains. Continued replication under the pressure of PI, and the high transcriptional error rate by HIV RT are attributed to the emergence of drug-resistant HIV mutants (Table 5).

The primary hepatic enzymes responsible for the metabolism of all the clinically used protease inhibitors are the cytochrome P450 isoenzymes such as CYP 3A4, CYP2C9, CYP2D6. Some PIs are potent inhibitors or inducers of the cytochrome P450 isoenzymes. As a result, PIs affect the metabolic processes and

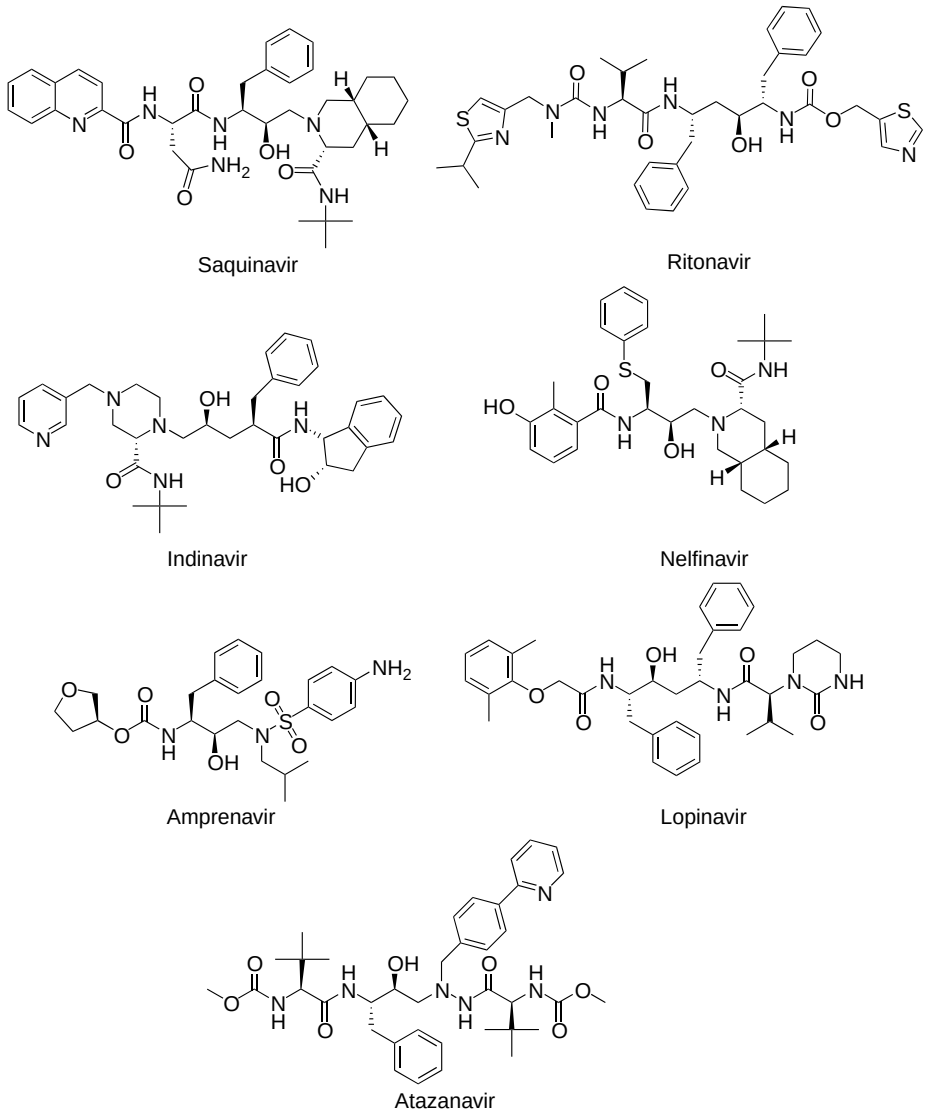

Fig. 6 Structures of FDA approved PIs.

blood levels of other drugs that are also metabolized by the cytochrome P450 isoenzymes. Consequently, the metabolism and blood level of PIs are also affected by drugs that inhibit the cytochrome P450 isoenzymes. For example, ritonovir, a potent inhibitor of the cytochrome P450 3A4, slows down the body's processing of lopinavir, and boosts blood levels of lopinavir. This coformulation of lopinavir and ritonavir, known as *Kaletra*, was approved by the FDA and is highly effective in bringing down the viral loads in HIV-1 infected patients. This ritonovir boosted-PI is

Tab. 4 Clinically used HIV-1 protease inhibitors.

Name or brand name	*Marketing company*	*Mechanism of action*	*Activity spectrum*
Saquinavir, Invirase®, Fortovase®	Hoffmann-La Roche	Transition state, peptidomimetic inhibitor	HIV-1 and HIV-2
Ritonavir, Norvir®	Abbott Laboratories	Transition state, peptidomimetic inhibitor	HIV-1 and HIV-2
Indinavir, Crixian®	Merck	Transition state, peptidomimetic inhibitor	HIV-1 and HIV-2
Nelfinavir, Viracept®	Agouron Pharmaceuticals	Transition state, peptidomimetic inhibitor	HIV-1 and HIV-2
Amprenavir, Agenerase®, Prozei®	GlaxoSmithKline	Transition state, peptidomimetic inhibitor	HIV-1 and HIV-2
Lopinavir, Kaletra®[a]	Abbott Laboratories	Transition state, peptidomimetic inhibitor	HIV-1 and HIV-2
Atazanavir, Reyataz®	Bristol-Myers Squibb	Transition state, peptidomimetic inhibitor	HIV-1 and HIV-2

[a]Kalatra® is the US FDA approved combination drug containing ritonavir and lopinavir.

Tab. 5 Known drug-resistant mutations and side effects of PIs.

PRIs	*Mutations associated with resistance*	*Toxic side effects*
Saquinavir	L10I/R/V, G48V, I54V/L, A71V/T, C73S, V77I, V82A, I84V, L90M	Altered sense of taste, upset stomach, heartburn
Riton	L10F/I/R/V, K20M/R, V32I, L33F, M36I, M46I/L, I54V/L, A71V/T, V77I, V82A/F/T/S, I84V, L90M	Perioral paresthesias, pancreatitis, flushing
Indinavir	L10I/R/V, K20M/R, L24I, V32I, M36I, M46I/L, I54V, V82A/F/T, I84V, L90M	Altered sense of taste, chelitis, retinoid effects, hyperbilirubinemia, nephrolithiasis, paronychia, rash, neutropenia
Nelfinavir	L10I/F, D30N, M36I, M46I/L, A71V/T, V77I, V82A/F/T/S, I84V, N88D/S, L90M	Most severe diarrhea of the PRIs, nephrolithiasis
Amprenavir	L10F/I/R/V, V32I, M46I/L, I47V, I50V, I54L/V/M, G73S, I84V, L90M	Lactic acidosis, perioral paresthesias, peripheral paresthesias, Stevenes-Johnson syndrome
Lopinavir	L10F/I/R/V, K20M/R, L24I, V32I, L33F, M46I/L, I47V/A, I50V, F53L, I54V/L/A/M/T/S, L63P, A71V/T, G73S, V82A/F/T/S, I84V, L90M	Pancreatitis, allergic reactions, rash
Atazanavir	L10I/F/V, K20R/M/I, L24I, V32I, L33I/F/V, M36I/L/V, M46I, G48V, I50L, I54L, A71V, G73C/S/T/A, V82A, I84V, N88S, L90M	Hyperbilirubinemia, lactic acidosis

a powerful anti-HIV drug and is currently favored as a first-line therapy. The serious side effects that are associated with clinical use of PIs are liver toxicity, lipodystrophy, hyperglycemia, hyperlipidemia, and sexual dysfunction (Table 5). The other common side effects caused by PIs are nausea, diarrhea, upset stomach, and heartburn. Lipodystrophy is fat redistribution with accumulation of neck and abdominal fat that is usually associated with loss of facial, buttocks, and extremity fat. The probable reason for this condition is suppression of breakdown of the nuclear sterol regulatory element binding proteins (nSERBP) by PIs. Suppression of the breakdown of nSERBP results in increased fatty acid and cholesterol biosynthesis in liver, which leads to lipodystrophy and insulin resistance in adipose tissue.

5 Integrase Inhibitors

HIV-1 integrase (IN) catalyzes two crucial steps required for the integration of viral DNA into the host chromosome. In the first step, while in the cytoplasm of an infected cell, IN selectively cleaves two nucleotides (GT) from the 3′ terminal end of the viral cDNA, in a reaction known as 3′-*processing*. Immediately after translocation to the nucleus as a component of the preintegration complex, IN randomly inserts the viral cDNA into the host genome. This reaction is referred to as *strand transfer or integration*. Hence, the interactions of IN with viral cDNA are sequence specific, but that with the host DNA are sequence independent. Consequently, their respective binding sites are expected to be geometrically and spatially different within the catalytic site of IN.

Currently, there are no FDA approved drugs targeting IN on the market. IN is an important and attractive antiretroviral drug target because there are no cellular homologs to IN, and the reactions catalyzed by IN are unique. A major concern with AIDS treatment is the emergence of drug-resistant viral strains, and the highly active antiretroviral therapy fails to eradicate the virus in HIV-infected patients. Therefore, addition of novel drugs, targeting alternative steps in the viral replication processes such as integration, to existing multidrug regimens would improve the outcome of treatment. Previously, several structurally diverse compounds were reported as IN inhibitors. Among all reported IN inhibitors, the β-diketoacid class of compounds has proved to be of most promise. Two of them, S-1360 and L-870, 810 have entered into clinical studies (Fig. 7). These compounds are believed to inhibit IN by interacting

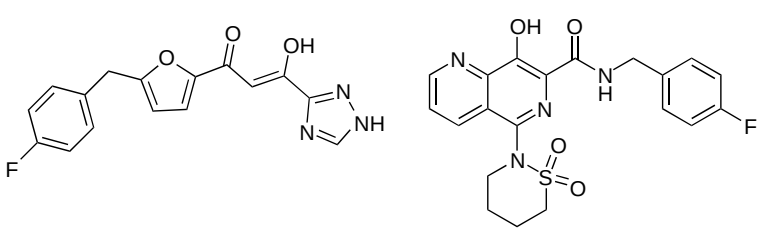

Fig. 7 Structures of integrase inhibitors.

Tab. 6 HIV-1 integrase inhibitors in clinical trials.

Name	*Pharmaceutical company*	*Mechanism of action*	*Resistant mutations*	*Status*
S-1360	Shionogi/GSK	Selective inhibition of strand transfer reaction of HIV-1 integrase	Q148K, I151L, N155S	Phase II
L-870, 810	Merck	Selective inhibition of strand transfer reaction of HIV-1 integrase	F121Y, T125K, V72I, V151I	Phase I

with the active site Mg^{2+} ion, and are selective inhibitors of the strand transfer reaction. These compounds showed antiviral activity against a variety of HIV strains expressing significant resistance to all of the currently available FDA approved drugs for HIV/AIDS, including NRTIs, NNRTIs, and PIs. After passing preliminary safety assessment studies, S-1360 and L-870, 810 compounds entered into clinical trials (Table 6).

6 HIV Entry Inhibitors

The process of HIV-1 entry into cells is conveniently divided into three main stages. First is the attachment of the virus to the host cell surface, mediated by the interaction of the viral envelope glycoprotein gp120 with the host T-lymphocyte cell surface protein CD4. This is an obligatory step for efficient infection of cells by HIV-1. Second is the interaction of the virus with chemokine coreceptors mediated by exposed domains of gp120. This is the result of induced conformational changes in gp120 upon its initial interactions with the CD4 receptor. Third, is the fusion of the virus and host cell membranes mediated by gp41. The above-mentioned steps are complex molecular events, and have been considered for the design of HIV-1 entry inhibitors (Table 7).

6.1 Inhibitors of GP120-CD4 Interactions

The HIV envelope glycoprotein gp120 interaction with the host cell surface CD4 protein, the principal receptor for HIV, is an indispensable step for HIV entry into cells. The CD4 protein belongs to the superfamily of immunoglobulin (Ig) like proteins, and is expressed primarily on T lymphocytes and macrophages. Binding of CD4 to gp120 is mediated by the interaction of N-terminal domain of CD4 with a cavity in gp120, located at the junction of both the inner and outer domains as well as bridging sheet of gp120 (Fig. 8). Detailed understanding of this docking process was elucidated through the recently solved crystal structure of a molecular complex consisting of an HIV-1 gp120 core, a two-domain fragment of CD4 and the Fab fragment of 17b, a neutralizing monoclonal antibody (Fig. 8). Currently, there are no FDA approved drugs targeting the gp120-CD4 interaction available for clinical use. However, two drugs, FP-21399 and BMS-488043, designed to interfere with gp120-CD4 interaction, have recently entered into clinical studies. An attachment

Tab. 7 The US FDA approved or experimental HIV-1 entry inhibitors.

Name or experimental code	***Pharmaceutical company***	***Mechanism of action***	***Status***
Enfuvirtide (T-20) Fuzeon®	Trimeris and Hoffmann-La Roche	gp41 inhibitor	FDA approved (2003)
T-1249	Trimeris	gp41 inhibitor	Phase I/II
BMS-488043	Bristol-Mayers Squibb	gp120-CD4 binding inhibitor or Attachment inhibitor	
PRO-542	Progenics Pharmaceuticals	gp120-CD4 binding inhibitor or Attachment inhibitor	Phase II
TNX-355	Tanox and Bigen Idec	gp120-CD4 binding inhibitor or Attachment inhibitor	Phase I
GW-873, 140	GlaxoSmithKline	CCR5 inhibitor	Phase I
SCH-D	Schering-Plough Corporation	CCR5 inhibitor	Phase I
UK-427, 857	Pfizer	CCR5 inhibitor	Phase II
AMD070	AnorMED Inc	CXCR4 inhibitor	Preclinical
KRH-1636	Kureha Chemical	CXCR4 inhibitor	Preclinical

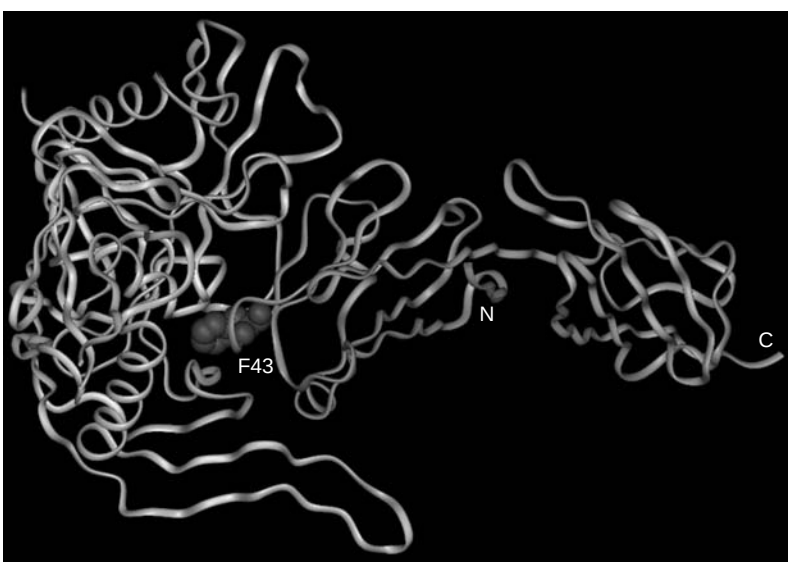

Fig. 8 The ribbon diagram of crystal structure shows binding of gp120 (magenta) to CD4 receptor (yellow). The F43 (shown as CPK model) of CD4 protrudes into a hydrophobic cavity on gp120. In this orientation, the CD4 binding cavity on gp120 is clearly visible. (See color plate p. xxvi).

inhibitor, PRO-542, a recombinant tetravalent CD4-IgG2 fusion protein, is currently under Phase II clinical trials. Another recently reported small-molecule attachment inhibitor BMS-378806, also targets viral entry by inhibiting the docking of the gp120 to the CD4 receptor through a specific and competitive mechanism. In fact, BMS-378806 binds to gp120 at or near the CD4 binding pocket. TNX-355, a monoclonal anti-CD4 antibody is also currently in Phase II clinical trials. This monoclonal antibody inhibits HIV replication by blocking the interaction between gp120 and CD4.

6.2 Chemokine Receptor Antagonists

Apart from the CD4 receptor, HIV needs a coreceptor for entry into host cells. The chemokine receptors CXCR4 and CCR5 have been identified as the principal coreceptors for HIV-1. The CCR5 receptor was found to mediate HIV-1 entry into macrophages and primary T cells (M or R5 tropic), whereas CXCR4 was shown to allow HIV-1 entry into primary T cells and T cell lines (T or X4 tropic). Both CCR5 and CXCR4 are members of the superfamily of guanine-nucleotide-binding protein-coupled receptors (GPCR). The importance of CCR5 for HIV-1 transmission and pathogenesis was established by reports demonstrating that individuals homozygous for a defective CCR5 allele are highly resistant to HIV-1 infection, whereas heterozygotes progress slower toward AIDS, probably due to a reduced surface expression of CCR5. However, targeting these cellular receptors for antiviral therapy has potential for serious side effects due to perturbations in the natural function of these receptors. Interestingly, knockout of the CCR5 gene in mice does not induce significant pathology. On the other hand, the deletion of CXCR4 in mice resulted in an embryonic lethal phenotype, suggesting that CXCR4 function is essential in mammals. Currently, there are no FDA approved drugs targeting either the CCR5 or the CXCR4 binding to gp120. The CCR5 antagonists under clinical studies are TAK-220, SCH-D, UK-427857, and GW873140 (AK-602 or ONO-4128). The small-molecule CXCR4 antagonists currently in clinical trials are KRH-1636, AMD3100, and AMD070. AMD3100, a bis-tetraazamacrocycle, is a potent inhibitor of HIV entry that specifically inhibits infection of X4 HIV-1 and HIV-2 strains but not R5 HIV-1 isolates. This molecule is a highly specific CXCR4 antagonist, and is currently in Phase II clinical trials. It is yet unclear whether AMD3100 competes directly for the gp120-binding site on CXCR4, or whether the compound prevents gp120 binding by inducing conformational changes in CXCR4 through an allosteric mechanism.

6.3 Viral Fusion Inhibitors

The binding of gp120 to CD4 and coreceptors triggers conformational changes in gp41, another HIV envelope glycoprotein, which leads to fusion of viral and host cell membrane. The exact mechanism of the conformational changes in gp41 that drive membrane fusion is not clearly understood. The most widely accepted model proposes that upon CD4 and coreceptor engagement, gp41 forms a prehairpin intermediate, in which the fusion peptide is inserted into the cellular membrane. This transient structure can be targeted by peptide inhibitors to disrupt the fusion process. Drugs targeting structural intermediates required for fusion can efficiently

block viral infection. Enfuvirtide (T-20) is a peptide based HIV-1 fusion inhibitor approved by the FDA for the treatment of HIV infection in treatment-experienced patients. Enfuvirtide, a synthetic peptide of 36 amino acids, binds to the HIV gp41 and blocks the fusion of viral and cellular membranes. Enfuvirtide competitively binds to the first heptad repeat (HR1) region of gp41, preventing the formation of a six-helix bundle, which is critical for membrane fusion to occur. Enfuvirtide exhibited synergistic activity when combined with individual antiretroviral agents, including zidovudine, lamivudine, nelfinavir, and indinavir. The most frequent adverse effects associated with the use of enfuvirtide are reactions at the injection site. Mutations in the HR1 region of gp41 envelope gene that are associated with resistance to Enfuvirtide include G36D/S, I37V, V38A/M, Q39R, N42T, and N43D. An enfuvirtide related peptide, T-1249, also binds to the N-terminal heptad repeat region (N-HR) of gp41 and blocks HIV-1 replication. T-1249 is currently in Phase I/II clinical studies.

7
Other Potential Drug Targets

7.1
HIV-1 Nucleocapsid Protein as a Drug Target

HIV-1 nucleocapsid (HIV-1 NC) protein is a small, highly basic protein with two conserved zinc (Zn^{2+}) finger domains. HIV-1 NC is generated during protease processing of the *gag* polyprotein that occurs during retrovirus maturation. HIV-1 NC is essential for the viral life cycle. This is due to its role in recognizing viral RNA from cellular RNA for dimerization and packaging, promoting binding of the essential t-RNA primer to the primer site, stimulating reverse transcription, and protecting viral RNA from nucleases. Mutations in HIV-1 NC are not tolerated, and a strong relationship has been established between the three-dimensional organization of NC and its biological functions. Mutations in the zinc finger domains of NC lead to a complete loss of infectivity. Hence, HIV-1 NC is a potential antiretroviral-drug target. Compounds that interfere with Zn^{2+} binding (destabilization of Zn^{2+} tetrahedral geometry), or eject the Zn^{2+} ion from the zinc finger domains of HIV-1 NC cause the denaturation of protein. This is lethal for the virus, and leads to formation of noninfective virions. Over the years several compounds have been reported that selectively interact with viral zinc finger proteins, but none of them have successfully reached advanced clinical studies.

7.2
Targeting HIV-1 Proteins Vif, Tat, and Rev

HIV accessory and regulatory proteins such as Vif, Tat, and Rev share a common function as adapter molecules to recruit cellular factors for various steps in the viral replication cycle. It is known that Vif-defective viruses are capable of entering target cells but encounter an early block in virus replication before integration of the viral DNA into the host genome. Currently, none of the inhibitors of Vif, Tat, and Rev have advanced to clinical trials.

See also AIDS/HIV, Molecular and Cell Biology; Pharmacogenomics and Drug Design.

Bibliography

Books and Reviews

Carr, A. Toxicity of antiretroviral therapy and implications for drug development, *Nat. Rev. Drug Discov.* 2003, **2**, 624–634.

De Clercq, E. (2002) Strategies in the design of antiviral drugs, *Nat. Rev. Drug Discov.* **1**, 13–25.

Flexner, C. (1998) HIV-protease inhibitors, *N. Engl. J. Med.* **338**, 1281–1292.

Frankel, A.D., Young, J.A. (1998) HIV-1: fifteen proteins and an RNA, *Annu. Rev. Biochem.* **67**, 1–25.

Katz, R.A., Skalka, A.M. (1994) The retroviral enzymes, *Annu. Rev. Biochem.* **63**, 133–173.

Pierson, T.C., Doms, R.W., Pohlmann, S. (2004) Prospects of HIV-1 entry inhibitors as novel therapeutics, *Rev. Med. Virol.* **14**, 255–270.

Richman, D.D. (2001) HIV chemotherapy, *Nature* **410**, 995–1001.

Sulkowski, M.S. (2003) Hepatotoxicity associated with antiretroviral therapy containing HIV-1 protease inhibitors, *Semin. Liver Dis.* **23**, 183–194.

Wynn, G.H., Zapor, M.J., Smith, B.H., Wortmann, G., Oesterheld, J.R., et al. (2004) Antiretrovirals, part 1: overview, history, and focus on protease inhibitors, *Psychosomatics* **45**, 262–270.

Primary Literature

Allaway, G.P., Davis-Bruno, K.L., Beaudry, G.A., Garcia, E.B., Wong, E.L., Ryder, A.M., Hasel, K.W., Gauduin, M.C., Koup, R.A., McDougal, J.S., et al. (1995) Expression and characterization of CD4-IgG2, a novel heterotetramer that neutralizes primary HIV type 1 isolates, *AIDS Res. Hum. Retroviruses* **11**, 533–539.

Arion, D., Kaushik, N., McCormick, S., Borkow, G., Parniak, M.A. (1998) Phenotypic mechanism of HIV-1 resistance to 3′-azido-3′-deoxythymidine (AZT): increased polymerization processivity and enhanced sensitivity to pyrophosphate of the mutant viral reverse transcriptase, *Biochemistry* **37**, 15908–15917.

Balzarini, J., Herdewijn, P., De Clercq, E. (1989) Differential patterns of intracellular metabolism of 2′,3′-didehydro-2′,3′-dideoxythymidine and 3′-azido-2′,3′-dideoxythymidine, two potent anti-human immunodeficiency virus compounds, *J. Biol. Chem.* **264**, 6127–6133.

Barney, S., Guthrie, K., Davis, D., et al. (1998) Pentafuside (T20), a novel inhibitor of HIV-1 fusion and infection, is synergistic when used in combination with reverse transcription and protease inhibitors in vitro, *Antiviral Res.* **37**, A54.

Bold, G., Fassler, A., Capraro, H.G., Cozens, R., Klimkait, T., Lazdins, J., Mestan, J., Poncioni, B., Rosel, J., Stover, D., Tintelnot-Blomley, M., Acemoglu, F., Beck, W., Boss, E., Eschbach, M., Hurlimann, T., Masso, E., Roussel, S., Ucci-Stoll, K., Wyss, D., Lang, M. (1998) New aza-dipeptide analogues as potent and orally absorbed HIV-1 protease inhibitors: candidates for clinical development, *J. Med. Chem.* **41**, 3387–3401.

Bourdais, J., Biondi, R., Sarfati, S., Guerreiro, C., Lascu, I., Janin, J., Veron, M. (1996) Cellular phosphorylation of anti-HIV nucleosides. Role of nucleoside diphosphate kinase, *J. Biol. Chem.* **271**, 7887–7890.

Caron, M., Auclair, M., Vigouroux, C., Glorian, M., Forest, C., Capeau, J. (2001) The HIV protease inhibitor indinavir impairs sterol regulatory element-binding protein-1 intranuclear localization, inhibits preadipocyte differentiation, and induces insulin resistance, *Diabetes* **50**, 1378–1388.

Carr, A., Miller, J., Law, M., Cooper, D.A. (2000) A syndrome of lipoatrophy, lactic acidaemia and liver dysfunction associated with HIV nucleoside analogue therapy: contribution to protease inhibitor-related lipodystrophy syndrome, *AIDS Res. Hum. Retroviruses* **14**, F25–F32.

Carr, A., Samaras, K., Thorisdottir, A., Kaufmann, G.R., Chisholm, D.J., Cooper, D.A. (1999) Diagnosis, prediction, and natural course of HIV-1 protease-inhibitor-associated lipodystrophy, hyperlipidaemia, and diabetes mellitus: a cohort study, *Lancet* **353**, 2093–2099.

Carr, A., Samaras, K., Burton, S., Law, M., Freund, J., Chisholm, D.J., Cooper, D.A. (1998) A syndrome of peripheral lipodystrophy, hyperlipidaemia and insulin resistance in patients receiving HIV protease inhibitors, *AIDS* **12**, F51–F58.

Cheng, Y.C., Dutschman, G.E., Bastow, K.F., Sarngadharan, M.G., Ting, R.Y. (1987) Human immunodeficiency virus reverse transcriptase. General properties and its interactions with nucleoside triphosphate analogs, *J. Biol. Chem.* **262**, 2187–2189.

Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rollins, B., Ponath, P.D., Wu, L., Mackay, C.R., LaRosa, G., Newman, W., Gerard, N., Gerard, C., Sodroski, J. (1996) The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates, *Cell* **85**, 1135–1148.

Cocchi, F., DeVico, A.L., Garzino-Demo, A., Arya, S.K., Gallo, R.C., Lusso, P. (1995) Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells, *Science* **270**, 1811–1815.

Darke, P.L., Nutt, R.F., Brady, S.F., Garsky, V.M., Ciccarone, T.M., Leu, C.T., Lumma, P.K., Freidinger, R.M., Veber, D.F., Sigal, I.S. (1988) HIV-1 protease specificity of peptide cleavage is sufficient for processing of gag and pol polyproteins, *Biochem. Biophys. Res. Commun.* **156**, 297–303.

Demene, H., Dong, C.Z., Ottmann, M., Rouyez, M.C., Jullian, N., Morellet, N., Mely, Y., Darlix, J.L., Fournie-Zaluski, M.C., Saragosti, S., et al. (1994) 1H NMR structure and biological studies of the His23–¿Cys mutant nucleocapsid protein of HIV-1 indicate that the conformation of the first zinc finger is critical for virus infectivity, *Biochemistry* **33**, 11707–11716.

Dezube, B.J., Dahl, T.A., Wong, T.K., Chapman, B., Ono, M., Yamaguchi, N., Gillies, S.D., Chen, L.B., Crumpacker, C.S. (2000) A fusion inhibitor (FP-21399) for the treatment of human immunodeficiency virus infection: a phase I study, *J. Infect. Dis.* **182**, 607–610.

Doms, R.W., Moore, J.P. (2000) HIV-1 membrane fusion: targets of opportunity, *J. Cell Biol.* **151**, F9–14.

Donzella, G.A., Schols, D., Lin, S.W., Este, J.A., Nagashima, K.A., Maddon, P.J., Allaway, G.P., Sakmar, T.P., Henson, G., De Clercq, E., Moore, J.P. (1998) AMD3100, a small molecule inhibitor of HIV-1 entry via the CXCR4 coreceptor, *Nat. Med.* **4**, 72–77.

Doranz, B.J., Grovit-Ferbas, K., Sharron, M.P., Mao, S.H., Goetz, M.B., Daar, E.S., Doms, R.W., O'Brien, W.A. (1997) A small-molecule inhibitor directed against the chemokine receptor CXCR4 prevents its use as an HIV-1 coreceptor, *J. Exp. Med.* **186**, 1395–1400.

Doranz, B.J., Rucker, J., Yi, Y., Smyth, R.J., Samson, M., Peiper, S.C., Parmentier, M., Collman, R.G., Doms, R.W. (1996) A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors, *Cell* **85**, 1149–1158.

Engelman, A., Mizuuchi, K., Craigie, R. (1991) HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer, *Cell* **67**, 1211–1221.

Esnouf, R., Ren, J., Ross, C., Jones, Y., Stammers, D., Stuart, D. (1995) Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors, *Nat. Struct. Biol.* **2**, 303–308.

Feng, Y., Broder, C.C., Kennedy, P.E., Berger, E.A. (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor, *Science* **272**, 872–877.

Furman, P.A., Fyfe, J.A., St Clair, M.H., Weinhold, K., Rideout, J.L., Freeman, G.A., Lehrman, S.N., Bolognesi, D.P., Broder, S., Mitsuya, H., et al. (1986) Phosphorylation of 3′-azido-3′-deoxythymidine and selective interaction of the 5′-triphosphate with human immunodeficiency virus reverse transcriptase, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8333–8337.

Gilboa, E., Mitra, S.W., Goff, S., Baltimore, D. (1979) A detailed model of reverse transcription and tests of crucial aspects, *Cell* **18**, 93–100.

Gorelick, R.J., Fu, W., Gagliardi, T.D., Bosche, W.J., Rein, A., Henderson, L.E., Arthur, L.O. (1999) Characterization of the block in replication of nucleocapsid protein zinc finger mutants from moloney murine leukemia virus, *J. Virol.* **73**, 8185–8195.

Gorelick, R.J., Gagliardi, T.D., Bosche, W.J., Wiltrout, T.A., Coren, L.V., Chabot, D.J., Lifson, J.D., Henderson, L.E., Arthur, L.O. (1999) Strict conservation of the retroviral nucleocapsid protein zinc finger is strongly influenced by its role in viral infection processes: characterization of HIV-1 particles containing mutant nucleocapsid zinc-coordinating sequences, *Virology* **256**, 92–104.

Gorelick, R.J., Benveniste, R.E., Gagliardi, T.D., Wiltrout, T.A., Busch, L.K., Bosche, W.J., Coren, L.V., Lifson, J.D., Bradley, P.J., Henderson, L.E., Arthur, L.O. (1999) Nucleocapsid

protein zinc-finger mutants of simian immunodeficiency virus strain mne produce virions that are replication defective in vitro and in vivo, *Virology* **253**, 259–270.

Graves, M.C., Lim, J.J., Heimer, E.P., Kramer, R.A. (1988) An 11-kDa form of human immunodeficiency virus protease expressed in Escherichia coli is sufficient for enzymatic activity, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2449–2453.

Guo, J., Wu, T., Kane, B.F., Johnson, D.G., Henderson, L.E., Gorelick, R.J., Levin, J.G. (2002) Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein, *J. Virol.* **76**, 4370–4378.

Guo, Q., Ho, H.T., Dicker, I., Fan, L., Zhou, N., Friborg, J., Wang, T., McAuliffe, B.V., Wang, H.G., Rose, R.E., Fang, H., Scarnati, H.T., Langley, D.R., Meanwell, N.A., Abraham, R., Colonno, R.J., Lin, P.F. (2003) Biochemical and genetic characterizations of a novel human immunodeficiency virus type 1 inhibitor that blocks gp120-CD4 interactions, *J. Virol.* **77**, 10528–10536.

Hatse, S., Princen, K., Bridger, G., De Clercq, E., Schols, D. (2002) Chemokine receptor inhibition by AMD3100 is strictly confined to CXCR4, *FEBS Lett.* **527**, 255–262.

Hatse, S., Princen, K., Vermeire, K., Gerlach, L.O., Rosenkilde, M.M., Schwartz, T.W., Bridger, G., De Clercq, E., Schols, D. (2003) Mutations at the CXCR4 interaction sites for AMD3100 influence anti-CXCR4 antibody binding and HIV-1 entry, *FEBS Lett.* **546**, 300–306.

Hazuda, D.J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J.A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C., Miller, M.D. (2000) Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells, *Science* **287**, 646–650.

He, Y., Vassell, R., Zaitseva, M., Nguyen, N., Yang, Z., Weng, Y., Weiss, C.D. (2003) Peptides trap the human immunodeficiency virus type 1 envelope glycoprotein fusion intermediate at two sites, *J. Virol.* **77**, 1666–1671.

Huang, M., Maynard, A., Turpin, J.A., Graham, L., Janini, G.M., Covell, D.G., Rice, W.G. (1998) Anti-HIV agents that selectively target retroviral nucleocapsid protein zinc fingers without affecting cellular zinc finger proteins, *J. Med. Chem.* **41**, 1371–1381.

Huffnagle, G.B., McNeil, L.K., McDonald, R.A., Murphy, J.W., Toews, G.B., Maeda, N., Kuziel, W.A. (1999) Cutting edge: role of C-C chemokine receptor 5 in organ-specific and innate immunity to Cryptococcus neoformans, *J. Immunol.* **163**, 4642–4646.

Ichiyama, K., Yokoyama-Kumakura, S., Tanaka, Y., Tanaka, R., Hirose, K., Bannai, K., Edamatsu, T., Yanaka, M., Niitani, Y., Miyano-Kurosaki, N., Takaku, H., Koyanagi, Y., Yamamoto, N. (2003) A duodenally absorbable CXC chemokine receptor 4 antagonist, KRH-1636, exhibits a potent and selective anti-HIV-1 activity, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4185–4190.

Jacobson, J.M., Lowy, I., Fletcher, C.V., O'Neill, T.J., Tran, D.N., Ketas, T.J., Trkola, A., Klotman, M.E., Maddon, P.J., Olson, W.C., Israel, R.J. (2000) Single-dose safety, pharmacology, and antiviral activity of the human immunodeficiency virus (HIV) type 1 entry inhibitor PRO 542 in HIV-infected adults, *J. Infect. Dis.* **182**, 326–329.

Joly, V., Yeni, P. (2000) Non-nucleoside reverse transcriptase inhibitors, *Ann. Med. Interne (Paris)* **151**, 260–267.

Joly, V., Flandre, P., Meiffredy, V., Leturque, N., Harel, M., Aboulker, J.P., Yeni, P. (2002) Increased risk of lipoatrophy under stavudine in HIV-1-infected patients: results of a substudy from a comparative trial, *AIDS* **16**, 2447–2454.

Kaldor, S.W., Kalish, V.J., Davies, J.F. II, Shetty, B.V., Fritz, J.E., Appelt, K., Burgess, J.A., Campanale, K.M., Chirgadze, N.Y., Clawson, D.K., Dressman, B.A., Hatch, S.D., Khalil, D.A., Kosa, M.B., Lubbehusen, P.P., Muesing, M.A., Patick, A.K., Reich, S.H., Su, K.S., Tatlock, J.H. (1997) Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease, *J. Med. Chem.* **40**, 3979–3985.

Kempf, D.J., Marsh, K.C., Kumar, G., Rodrigues, A.D., Denissen, J.F., McDonald, E., Kukulka, M.J., Hsu, A., Granneman, G.R., Baroldi, P.A., Sun, E., Pizzuti, D., Plattner, J.J., Norbeck, D.W., Leonard, J.M. (1997) Pharmacokinetic enhancement of inhibitors of the human immunodeficiency virus protease by coadministration with ritonavir, *Antimicrob. Agents Chemother.* **41**, 654–660.

Kempf, D.J., Sham, H.L., Marsh, K.C., Flentge, C.A., Betebenner, D., Green, B.E., McDonald, E., Vasavanonda, S., Saldivar, A., Wideburg, N.E., Kati, W.M., Ruiz, L., Zhao, C., Fino, L., Patterson, J., Molla, A., Plattner, J.J., Norbeck, D.W. (1998) Discovery of ritonavir, a potent inhibitor of HIV protease with high oral bioavailability and clinical efficacy, *J. Med. Chem.* **41**, 602–617.

Kempf, D.J., Marsh, K.C., Denissen, J.F., McDonald, E., Vasavanonda, S., Flentge, C.A., Green, B.E., Fino, L., Park, C.H., Kong, X.P., et al. (1995) ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2484–2488.

Kempf, D.J., Norbeck, D.W., Codacovi, L., Wang, X.C., Kohlbrenner, W.E., Wideburg, N.E., Paul, D.A., Knigge, M.F., Vasavanonda, S., Craig-Kennard, A., et al. (1990) Structure-based, C2 symmetric inhibitors of HIV protease, *J. Med. Chem.* **33**, 2687–2689.

Kilby, J.M., Lalezari, J.P., Eron, J.J., Carlson, M., Cohen, C., Arduino, R.C., Goodgame, J.C., Gallant, J.E., Volberding, P., Murphy, R.L., Valentine, F., Saag, M.S., Nelson, E.L., Sista, P.R., Dusek, A. (2002) The safety, plasma pharmacokinetics, and antiviral activity of subcutaneous enfuvirtide (T-20), a peptide inhibitor of gp41-mediated virus fusion, in HIV-infected adults, *AIDS Res. Hum. Retroviruses* **18**, 685–693.

Kohl, N.E., Emini, E.A., Schleif, W.A., Davis, L.J., Heimbach, J.C., Dixon, R.A., Scolnick, E.M., Sigal, I.S. (1988) Active human immunodeficiency virus protease is required for viral infectivity, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4686–4690.

Kronenberg, A., Riehle, H.M., Gunthard, H.F. (2001) Liver failure after long-term nucleoside antiretroviral therapy, *Lancet* **358**, 759–760.

Kuritzkes, D.R., Jacobson, J., Powderly, W.G., Godofsky, E., DeJesus, E., Haas, F., Reimann, K.A., Larson, J.L., Yarbough, P.O., Curt, V., Shanahan, W.R. Jr. (2004) Antiretroviral activity of the anti-CD4 monoclonal antibody TNX-355 in patients infected with HIV type 1, *J. Infect. Dis.* **189**, 286–291.

Kushmann, S.E., Moore, J.P. (2004) HIV-1 entry inhibitor entrance, *Trends Pharmacol. Sci.* **25**, 117–120.

Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., Hendrickson, W.A. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody, *Nature* **393**, 648–659.

Lallemand, F., Salhi, Y., Linard, F., Giami, A., Rozenbaum, W. (2002) Sexual dysfunction in 156 ambulatory HIV-infected men receiving highly active antiretroviral therapy combinations with and without protease inhibitors, *J. Acquir. Immune. Defic. Syndr.* **30**, 187–190.

Lam, P.Y., Jadhav, P.K., Eyermann, C.J., Hodge, C.N., Ru, Y., Bacheler, L.T., Meek, J.L., Otto, M.J., Rayner, M.M., Wong, Y.N., et al. (1994) Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors, *Science* **263**, 380–384.

Lana, R., Nunez, M., Mendoza, J.L., Soriano, V. (2001) Rate and risk factors of liver toxicity in patients receiving antiretroviral therapy, *Med. Clin. (Barc)* **117**, 607–610.

Lapham, C.K., Ouyang, J., Chandrasekhar, B., Nguyen, N.Y., Dimitrov, D.S., Golding, H. (1996) Evidence for cell-surface association between fusin and the CD4-gp120 complex in human cell lines, *Science* **274**, 602–605.

Larder, B.A., Kemp, S.D. (1989) Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT), *Science* **246**, 1155–1158.

Lewis, W., Dalakas, M.C. (1995) Mitochondrial toxicity of antiviral drugs, *Nat. Med.* **1**, 417–422.

Lin, P.F., Blair, W., Wang, T., Spicer, T., Guo, Q., Zhou, N., Gong, Y.F., Wang, H.G., Rose, R., Yamanaka, G., Robinson, B., Li, C.B., Fridell, R., Deminie, C., Demers, G., Yang, Z., Zadjura, L., Meanwell, N., Colonno, R. (2003) A small molecule HIV-1 inhibitor that targets the HIV-1 envelope and inhibits CD4 receptor binding, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11013–11018.

Liu, R., Paxton, W.A., Choe, S., Ceradini, D., Martin, S.R., Horuk, R., MacDonald, M.E., Stuhlmann, H., Koup, R.A., Landau, N.R. (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection, *Cell* **86**, 367–377.

McQuade, T.J., Tomasselli, A.G., Liu, L., Karacostas, V., Moss, B., Sawyer, T.K., Heinrikson, R.L., Tarpley, W.G. (1990) A synthetic HIV-1 protease inhibitor with antiviral activity arrests HIV-like particle maturation, *Science* **247**, 454–456.

Mely, Y., Cornille, F., Fournie-Zaluski, M.C., Darlix, J.L., Roques, B.P., Gerard, D. (1991) Investigation of zinc-binding affinities of Moloney murine leukemia virus nucleocapsid protein and its related zinc finger and modified peptides, *Biopolymers* **31**, 899–906.

Meyer, P.R., Matsuura, S.E., Mian, A.M., So, A.G., Scott, W.A. (1999) A mechanism of AZT resistance: an increase in nucleotide-dependent primer unblocking by mutant HIV-1 reverse transcriptase, *Mol. Cell.* **4**, 35–43.

Meyer, P.R., Matsuura, S.E., Zonarich, D., Chopra, R.R., Pendarvis, E., Bazmi, H.Z., Mellors, J.W., Scott, W.A. (2003) Relationship between 3′-azido-3′-deoxythymidine resistance and primer unblocking activity in foscarnet-resistant mutants of human immunodeficiency virus type 1 reverse transcriptase, *J. Virol.* **77**, 6127–6137.

Michael, N.L., Chang, G., Louie, L.G., Mascola, J.R., Dondero, D., Birx, D.L., Sheppard, H.W. (1997) The role of viral phenotype and CCR-5 gene defects in HIV-1 transmission and disease progression, *Nat. Med.* **3**, 338–340.

Miller, K.D., Jones, E., Yanovski, J.A., Shankar, R., Feuerstein, I., Falloon, J. (1998) Visceral abdominal-fat accumulation associated with use of indinavir, *Lancet* **351**, 871–875.

Nagasawa, T., Hirota, S., Tachibana, K., Takakura, N., Nishikawa, S., Kitamura, Y., Yoshida, N., Kikutani, H., Kishimoto, T. (1996) Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1, *Nature* **382**, 635–638.

Navia, M.A., Fitzgerald, P.M., McKeever, B.M., Leu, C.T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L., Springer, J.P. (1989) Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1, *Nature* **337**, 615–620.

Priestle, J.P., Fassler, A., Rosel, J., Tintelnot-Blomley, M., Strop, P., Grutter, M.G. (1995) Comparative analysis of the X-ray structures of HIV-1 and HIV-2 proteases in complex with CGP 53820, a novel pseudosymmetric inhibitor, *Structure* **3**, 381–389.

Ramboarina, S., Morellet, N., Fournie-Zaluski, M.C., Roques, B.P., Moreller, N. (1999) Structural investigation on the requirement of CCHH zinc finger type in nucleocapsid protein of human immunodeficiency virus 1, *Biochemistry* **38**, 9600–9607.

Reeves, J.D., Gallo, S.A., Ahmad, N., Miamidian, J.L., Harvey, P.E., Sharron, M., Pohlmann, S., Sfakianos, J.N., Derdeyn, C.A., Blumenthal, R., Hunter, E., Doms, R.W. (2002) Sensitivity of HIV-1 to entry inhibitors correlates with envelope/coreceptor affinity, receptor density, and fusion kinetics, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16249–16254.

Ren, J., Esnouf, R.M., Hopkins, A.L., Warren, J., Balzarini, J., Stuart, D.I., Stammers, D.K. (1998) Crystal structures of HIV-1 reverse transcriptase in complex with carboxanilide derivatives, *Biochemistry* **37**, 14394–14403.

Ren, J., Esnouf, R., Garman, E., Somers, D., Ross, C., Kirby, I., Keeling, J., Darby, G., Jones, Y., Stuart, D., et al. (1995) High resolution structures of HIV-1 RT from four RT-inhibitor complexes, *Nat. Struct. Biol.* **2**, 293–302.

Rice, W.G., Schaeffer, C.A., Harten, B., Villinger, F., South, T.L., Summers, M.F., Henderson, L.E., Bess, J.W. Jr., Arthur, L.O., McDougal, J.S., et al. (1993) Inhibition of HIV-1 infectivity by zinc-ejecting aromatic C-nitroso compounds, *Nature* **361**, 473–475.

Rice, W.G., Supko, J.G., Malspeis, L., Buckheit, R.W. Jr., Clanton, D., Bu, M., Graham, L., Schaeffer, C.A., Turpin, J.A., Domagala, J., et al. (1995) Inhibitors of HIV nucleocapsid protein zinc fingers as candidates for the treatment of AIDS, *Science* **270**, 1194–1197.

Roberts, J.D., Bebenek, K., Kunkel, T.A. (1988) The accuracy of reverse transcriptase from HIV-1, *Science* **242**, 1171–1173.

Roberts, N.A., Martin, J.A., Kinchington, D., Broadhurst, A.V., Craig, J.C., Duncan, I.B., Galpin, S.A., Handa, B.K., Kay, J., Krohn, A., et al. (1990) Rational design of peptide-based HIV proteinase inhibitors, *Science* **248**, 358–361.

Sarafianos, S.G., Das, K., Clark, A.D. Jr., Ding, J., Boyer, P.L., Hughes, S.H., Arnold, E. (1999) Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with beta-branched amino acids, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10027–10032.

Seelmeier, S., Schmidt, H., Turk, V., von der Helm, K. (1988) Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6612–6616.

Sham, H.L., Kempf, D.J., Molla, A., Marsh, K.C., Kumar, G.N., Chen, C.M., Kati, W., Stewart, K., Lal, R., Hsu, A., Betebenner, D., Korneyeva, M., Vasavanonda, S., McDonald, E., Saldivar, A., Wideburg, N., Chen, X., Niu, P., Park, C., Jayanti, V., Grabowski, B., Granneman, G.R., Sun, E., Japour, A.J., Norbeck, D.W., et al. (1998) ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease, *Antimicrob. Agents Chemother.* **42**, 3218–3224.

Shearer, M.H., Timanus, D.K., Benton, P.A., Lee, D.R., Kennedy, R.C. (1998) Cross-clade inhibition of human immunodeficiency virus type 1 primary isolates by monoclonal anti-CD4, *J. Infect. Dis.* **177**, 1727–1729.

Sluis-Cremer, N., Arion, D., Parniak, M.A. (2000) Molecular mechanisms of HIV-1 resistance to nucleoside reverse transcriptase inhibitors (NRTIs), *Cell. Mol. Life Sci.* **57**, 1408–1422.

Spence, R.A., Kati, W.M., Anderson, K.S., Johnson, K.A. (1995) Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors, *Science* **267**, 988–993.

Ugolini, S., Moulard, M., Mondor, I., Barois, N., Demandolx, D., Hoxie, J., Brelot, A., Alizon, M., Davoust, J., Sattentau, Q.J. (1997) HIV-1 gp120 induces an association between CD4 and the chemokine receptor CXCR4, *J. Immunol.* **159**, 3000–3008.

Vacca, J.P., Dorsey, B.D., Schleif, W.A., Levin, R.B., McDaniel, S.L., Darke, P.L., Zugay, J., Quintero, J.C., Blahy, O.M., Roth, E., et al. (1994) L-735,524: an orally bioavailable human immunodeficiency virus type 1 protease inhibitor, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4096–4100.

Vandercam, B., Moreau, M., Horsmans, C., Gala, J.L. (1998) Acute hepatitis in a patient treated with saquinavir and ritonavir: absence of cross-toxicity with indinavir, *Infection* **26**, 313.

von Moltke, L.L., Durol, A.L., Duan, S.X., Greenblatt, D.J. (2000) Potent mechanism-based inhibition of human CYP3A in vitro by amprenavir and ritonavir: comparison with ketoconazole, *Eur. J. Clin. Pharmacol.* **56**, 259–261.

Warren, K.J., Boxwell, D.E., Kim, N.Y., Drolet, B.A. (1998) Nevirapine-associated Stevens-Johnson syndrome, *Lancet* **351**, 567.

Wild, C., Oas, T., McDanal, C., Bolognesi, D., Matthews, T. (1992) A synthetic peptide inhibitor of human immunodeficiency virus replication: correlation between solution structure and viral inhibition, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10537–10541.

Williams, M.C., Gorelick, R.J., Musier-Forsyth, K. (2002) Specific zinc-finger architecture required for HIV-1 nucleocapsid protein's nucleic acid chaperone function, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8614–8619.

Witvrouw, M., Pannecouque, C., Van Laethem, K., Desmyter, J., De Clercq, E., Vandamme, A.M. (1999) Activity of non-nucleoside reverse transcriptase inhibitors against HIV-2 and SIV, *AIDS* **13**, 1477–1483.

Witvrouw, M., Balzarini, J., Pannecouque, C., Jhaumeer-Laulloo, S., Este, J.A., Schols, D., Cherepanov, P., Schmit, J.C., Debyser, Z., Vandamme, A.M., Desmyter, J., Ramadas, S.R., de Clercq, E. (1997) SRR-SB3, a disulfide-containing macrolide that inhibits a late stage of the replicative cycle of human immunodeficiency virus, *Antimicrob. Agents Chemother.* **41**, 262–268.

Wlodawer, A., Erickson, J.W. (1993) Structure-based inhibitors of HIV-1 protease, *Annu. Rev. Biochem.* **62**, 543–585.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J., Kent, S.B. (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease, *Science* **245**, 616–621.

Wyatt, R., Kwong, P.D., Desjardins, E., Sweet, R.W., Robinson, J., Hendrickson, W.A., Sodroski, J.G. (1998) The antigenic structure of the HIV gp120 envelope glycoprotein, *Nature* **393**, 705–711.

Yarasheski, K.E., Tebas, P., Sigmund, C., Dagogo-Jack, S., Bohrer, A., Turk, J., Halban, P.A., Cryer, P.E., Powderly, W.G. (1999) Insulin resistance in HIV protease inhibitor-associated diabetes, *J. Acquir. Immune. Defic. Syndr.* **21**, 209–216.

Zhang, J.L., Choe, H., Dezube, B.J., Farzan, M., Sharma, P.L., Zhou, X.C., Chen, L.B., Ono, M., Gillies, S., Wu, Y., Sodroski, J.G., Crumpacker, C.S. (1998) The bis-azo compound FP-21399 inhibits HIV-1 replication by preventing viral entry, *Virology* **244**, 530–541.

Zhou, Y., Kurihara, T., Ryseck, R.P., Yang, Y., Ryan, C., Loy, J., Warr, G., Bravo, R. (1998) Impaired macrophage function and enhanced T cell-dependent immune response in mice lacking CCR5, the mouse homologue of the

major HIV-1 coreceptor, *J. Immunol.* **160**, 4018–4025.

Zou, Y.R., Kottmann, A.H., Kuroda, M., Taniuchi, I., Littman, D.R. (1998) Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development, *Nature* **393**, 595–599.

Homeodomain Proteins

Thomas R. Bürglin
Department of Biosciences at Novum and CGB, Karolinska Institutet, Huddinge, Sweden

1 What is a Homeodomain? 182
1.1 Introduction 182
1.2 The Homeodomain Sequence 182
1.3 Structure of the Homeodomain 185
1.4 DNA-binding Properties 185
1.5 Classification of Homeodomains 186
1.6 Atypical Homeodomains 187

2 The Antennapedia Superclass 189
2.1 The HOX Cluster 189
2.1.1 Labial (lab) Family 192
2.1.2 Proboscipedia (pb) Family 192
2.1.3 PG-3 Family 192
2.1.4 Deformed (Dfd) Family 192
2.1.5 Sex-combs-reduced (Scr) Family 193
2.1.6 Antennapedia (Antp) Family 194
2.1.7 Abdominal-B (Abd-B) Family 196
2.1.8 Even-skipped (eve) Family 198
2.1.9 Mox Family 199
2.2 ParaHox Cluster 199
2.2.1 Gsx Family 201
2.2.2 Xlox Family 201
2.2.3 Caudal (cad) Family 201
2.3 The NK Cluster 201
2.3.1 NK-1 Family 201
2.3.2 Tlx Family 201
2.3.3 Lbx Family 201
2.3.4 NK-3 Family 202

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

2.3.5 NK-4 Family 202
2.3.6 Msx Family 202
2.4 Noncluster Antennapedia Superclass Genes 202
2.4.1 NK-2 Family 202
2.4.2 Distal-less (Dll) Family 202
2.4.3 Empty-spiracles (ems) Family 202
2.4.4 Engrailed (en) Family 203
2.4.5 Hlx Family 203
2.4.6 Hex Family 203
2.4.7 Other Families 204
2.5 Antennapedia Superclass Related Genes in Lower Animals 204

3 Paired and prd-like Classes 204
3.1 Paired (prd) Class 204
3.1.1 Family I 206
3.1.2 Family II 206
3.1.3 Family III 206
3.1.4 Family IV 206
3.1.5 Family V 206
3.1.6 Family VI 206
3.2 Paired-like (prd-like) Class 206

4 POU Class and HNF Classes 207
4.1 POU Class 207
4.1.1 Family I 207
4.1.2 Family II 207
4.1.3 Family III 207
4.1.4 Family IV 207
4.1.5 Family V 208
4.1.6 Family VI 208
4.1.7 Orphans 208
4.2 HNF Class 208

5 Cut Superclass 208
5.1 CUX Class 209
5.2 ONECUT Class 210
5.3 SATB Class 210
5.4 COMPASS (CMP) Class 210

6 ZF and LIM Classes 210
6.1 ZF Class 211
6.2 LIM Class 211

7 Six/so Class 212

8 Prospero (Pros) Class 213

9 TALE Superclass 214
9.1 PBC Class 214
9.2 MEIS Class 214
9.3 IRO Class 215
9.4 TGIF Class 215

10 Evolution of Homeobox Genes 216
10.1 Fungal Homeobox Genes 216
10.2 Plant Homeobox Genes 216
10.3 Origin and Diversification of Homeobox Genes 217

Bibliography 218
Books and Reviews 218
Primary Literature 218

Keywords

Homeobox
A conserved sequence of about 180 nucleotides that is found in many developmental control genes.

Homeodomain
The protein domain encoded by the homeobox. It is a DNA-binding domain.

Homeotic Gene
Mutations in a homeotic gene cause transformations – homeotic transformations – of groups of cells or body segments. For example, the cells of a segment are transformed into cell types of a different body segment. A large number of homeotic genes are homeobox genes, but not all. And not all homeobox genes are homeotic genes.

HOX Cluster
A gene cluster in animals made up of Antennapedia-type homeobox genes (Hox genes). The genes in this cluster function primarily in the pattern formation along the anterior–posterior body axis.

Hox Genes
The homeobox genes that reside in the Hox cluster(s). This designation should be reserved for use in vertebrates.

POU Genes
A special well-known class of homeobox genes that encode two domains, a POU-specific domain and a POU homeodomain. The POU-specific domain binds DNA also.

TALE Homeobox Genes
A special subgroup of homeobox genes, whose homeodomain is normally 63 amino acids long, rather than the standard 60, because of three extra amino acids between helix 1 and helix 2. Highly conserved from plants to animals.

Homeodomain proteins are DNA-binding proteins that play major roles in many developmental processes of animals as well as fungi and plants. The homeodomain is a protein domain of about 60 amino acids that binds DNA. Proteins containing a homeodomain can regulate the expression of other genes in development and differentiation. Hundreds of homeobox genes have been identified, and they can be grouped into many different classes. Often, other conserved protein domains are found linked to a homeodomain. Several particular types of homeobox genes lie next to each other on the chromosome in clusters. The best-known cluster, the HOX cluster, is found in all bilateral animals. Vertebrates contain four HOX clusters that have arisen through duplication in early vertebrate evolution; its genes are called *Hox* genes. Lower chordates, insects, and nematodes have only one HOX cluster. Of particular interest is that many of the HOX cluster genes function in the process of pattern formation along the anterior–posterior body axis. However, a lot of homeobox genes are not found in clusters and do not function in pattern formation, but play roles in the determination of cell fates and cell differentiation. Homeobox genes thus perform key roles for all aspects of the development of an organism.

1 What is a Homeodomain?

1.1 Introduction

Since their discovery in 1983, homeobox genes, and the homeodomain proteins they encode, have turned out to play important roles in the developmental processes of all multicellular organisms. While certainly not the only developmental control genes, they have been shown to play crucial roles from the earliest steps in embryogenesis – such as setting up an anterior–posterior gradient in the egg of the fruit fly *Drosophila melanogaster* – to the very latest steps in cell differentiation – such as the differentiation of neurons in the nematode *Caenorhabditis elegans*. They are found in plants, fungi, and all animal phyla. Since their original discovery, hundreds of homeobox genes have been described, and this article, which is based on previous compilations, attempts to give an overview of these genes.

1.2 The Homeodomain Sequence

The homeobox was originally described as a conserved DNA motif of about 180 bp. The protein domain encoded by the homeobox, the homeodomain, is, accordingly, about 60 amino acids long. The first genes found to encode homeodomain proteins were the *Drosophila* developmental control

genes, in particular, homeotic genes, from which the name "homeo" box was derived. However, many homeobox genes are not homeotic genes; the homeobox is a sequence motif, while "homeotic" is a functional description for genes that cause homeotic transformations when mutated. Further, not all homeotic genes are homeobox genes.

As more and more homeobox-containing genes were isolated, it became clear that the homeodomains of closely related genes had sequence similarity that extended upstream and/or downstream of the standard core of 60 amino acids. In contrast, when highly divergent homeodomain sequences are compared, the sequence similarity does not extend over the full 60 residues. Figure 1(a) shows a consensus sequence that is based on a compilation of 346 homeodomain sequences. The numbering scheme beneath this consensus is in accordance with that used in the publications that describe the homeodomain structure.

An analysis of the frequency of each amino acid at each position of the homeodomain indicates that certain positions are significantly more conserved than others. A summary of these conserved amino acids is given in Fig. 1(b). Under each consensus (i.e. most common) amino acid, other amino acids that occur at that position are listed in decreasing order of frequency. Amino acids found in fewer than five sequences have been omitted. The gray background bars highlight those residues in each position, which can be found in at least 85% or more of the homeodomain sequences. For example, at position 47, the two hydrophobic residues isoleucine (I) and valine (V) are found in 333 of the 346 sequences (96%), while a threonine (T) is found in only 9 sequences (2.6%). In addition, two sequences have an asparagine (N) residue and two have a histidine (H) and are thus not even shown in Fig. 1(b). In such a way, one can identify positions where amino acids with particular properties, such as hydrophobic residues (e.g. position 47), are highly conserved.

The conserved positions in Fig. 1(b) have been arbitrarily divided into three groups, depending on how many different amino acids occur at a particular position in this compilation. There are seven very highly conserved positions (¤). However, only two positions are almost totally invariant, the tryptophan (W) residue at position 48 and the asparagine (N) residue at position 51. Occasionally, the tryptophan (W) can be substituted by a phenylalanine (F) residue. Three other positions predominantly contain one amino acid but occasionally permit substitutions: position 16, normally a leucine (L), can accommodate a methionine(M) or other hydrophobic residues; positions 20 and 49, normally phenylalanine (F) residues, can be substituted by tyrosine (Y) or other hydrophobic residues. Position 53 is almost invariably an arginine (R), though lysine (K) is possible. Position 57 accommodates either an arginine (R) or a lysine (K) residue, although other residues are possible.

A second group of highly conserved positions (▼) has fewer than six different amino acids at a particular position. Six such positions are found: 5, 26, 40, 45, 47, and 57. Thirteen positions fall into the third group of moderately conserved positions (*), which can have up to nine different amino acids at a particular position. However, those positions of the homeodomain that are not especially marked by any of the three previous symbols are not completely devoid of information. Often, while many different amino acids can occur at a particular

```
                Helix 1              Helix 2              Helix 3/4
               H  H   H      H        HH HH H    H  HH
  m m#m#                   #  #  #          ###   B# BB #B# #
RRRKRTAYTRYQLLELEKEFHFNRYLTRRRRIELAHSLNLTERQVKIWFQNRRMKWKKEN
....|....|....|....|....|....|....|....|....|....|....|....|
1   5   10   15   20   25   30   35   40   45   50   55   60
```

(a)

```
....|....|....|....|....|....|....|....|....|....|....|....|
     ▼   *   **   ¤*        *▼     *  **   *  ▼    *▼*▼¤¤*¤*¤¤   ▼
RRRKRTAYTRYQLLELEKEFHFNRYLTRRRRIELAHSLNLTERQVKIWFQNRRMKWKKEN
KKKPGQTFSKEVTAA KRAYLQSKKPSAAEIEQISAATGMSDTVIRV  C K ARERRQS
PNSG CPIETSAVER RNK QKQPFIDIPKKARV NEIC NKAR QT  S H Q L QDT
GSGS VSLKSAIIRT  QH ERTQNVNKQDLVSM RTVS EPET T   K A T Q  MQ
SDAA AV DAH KVV  AR AEKMR CVET LHF QL Q PMS          Y M  LK
N  R KN  DQ RGS  SS RYCH  ASKM TA  SR D KQK          V A  IE
Q  T RI  VK  TI  TY KTEN  VEDA QN  EK H DTD          S I  RH
T  V  L  NF  QQ  HF NADS   GHQ KK  DM K R            N H  AV
E  L  R  QT  K   EC FSHV   PL  RD  KN R                Y  VR
D  I     PN  D      SI A   LY  S   LQ E                D  SL
A  Y     GL  F      GV E    S  Y   TV A                F  KA
L  N     YD  Y      DL      T  F    I                 K  N
V  Q     EP  H      YN      F  C    D                 T
   H      R         MH         M
   M
No. of different amino acids encountered:    ¤1-2; ▼3-5; *6-9
```

(b)

Fig. 1 (a) A homeodomain consensus sequence (presented in single-letter code, as are all subsequent sequence figures) that is based on 346 homeodomain sequences. The standard numbering scheme is shown beneath the consensus. The three α-helices (a composite derived from the structures of the Antp, en, and MATα2 homeodomains) are shown schematically as cylinders above the consensus. Special symbols mark amino acids positions that play a role for intramolecular or protein–DNA contacts of the homeodomain; they are shown above the consensus and have the following meanings: amino acids designated ''H'' contribute to the hydrophobic core that is responsible for the tertiary structure of the homeodomain; residues designated ''B'' contact bases in the major groove and are responsible for sequence-specific DNA contacts; residues designated ''m'' make contacts in the minor groove; the pound sign (#) indicates residues that contact the sugar-phosphate backbone of the DNA. (b) Amino acids encountered at a given position in the homeodomain (based on 346 sequences). For each position, the amino acid most frequently encountered is listed at the top, while other amino acids are listed beneath in decreasing order of their frequency of occurrence. Residues in the shaded group are found in at least 85% of the 346 homeodomain sequences. Amino acids occurring fewer than five times are not shown. The symbols on top indicate highly conserved positions: ¤, most highly conserved positions, with only one or two different amino acids found at that position; ▼, highly conserved position, with three to five different amino acids at that position; ∗, conserved positions, with six to nine different residues at that position. *Note*: the 346 selected sequences for this compilation are biased toward animal, in particular, Antp-like homeodomain sequences.

position, some amino acids appear to be excluded. For example, at position 1, no tyrosine (Y) or tryptophan (W) has been found in the 346 sequences examined.

These 60 positions, especially the more conserved ones, define the typical homeodomain. Residues in the DNA-binding region of helix 3 (see Sect. 1.2) are especially conserved and constitute a "trademark" for the homeodomain. The sequence compilation shown here is mainly composed of animal homeodomains, with a bias toward Antennapedia-like sequences. Nevertheless, even plant homeodomain sequences show a similar pattern of conserved residues, although amino acid frequencies at particular positions may vary from the one presented here.

1.3 Structure of the Homeodomain

The underlying reason for the conservation of particular amino acids is the structure of the homeodomain. Several structures have been determined using either NMR or X-ray analysis, for example, fly Antennapedia (Antp), fushi tarazu (ftz), engrailed (en), or mammalian PBX1, Oct1, HNF1α, or yeast MATα2. The core of the homeodomain consists essentially of three α-helices. The NMR studies of the Antp homeodomain identified a kink in the third helix, at position 52/53, such that this helix is considered to consist of two separate helices, helix 3 and helix 4. The three α-helices are represented as shaded cylinders above the consensus sequence in Fig. 1(a).

The various symbols (H, B, #, m) above the helices in Fig. 1(a) give a schematic summary of various contacts made by the amino acids of the homeodomain, as deduced from structural and genetic data. Not every homeodomain makes every interaction shown; in particular, DNA contacts can be slightly different in different homeodomains. The three helices are held together by a hydrophobic core. The residues marked by "H" take part in the formation of this hydrophobic core. Helix 2 and helix 3 are connected by a tight turn called a *helix-turn-helix motif* that is also found in prokaryotic gene regulatory proteins. Helix 1 is connected to helix 2 by a loop, which contains three extra amino acids in TALE (three amino acid loop extension) homeodomains (see below). Helix 1 lies approximately parallel to helix 2, and also crosses over helix 3. At the crossover area, the highly conserved residues in helix 1 interact with the highly conserved residues tryptophan and phenylalanine of helix 3. Figure 2 is a schematic representation of how the three helices pack against one another and how they lie with respect to the DNA.

1.4 DNA-binding Properties

NMR and X-ray analysis of DNA–protein complexes for several homeodomain proteins have allowed the identification of the residues critical for sequence-specific contacts and contacts to the DNA backbone. These studies have been complemented by *in vivo* and *in vitro* studies on the DNA-binding properties of homeodomains and confirm the importance of some of the DNA–amino acid contacts that have been deduced by the structural studies.

Helix 3, also termed the *recognition helix*, lies in the major groove of the DNA and plays the most important role in making sequence-specific contacts (Fig. 2). Those residues that make sequence-specific contacts are indicated with a "B" in Fig. 1(a) above the consensus. Many other residues make nonsequence-specific contacts to the

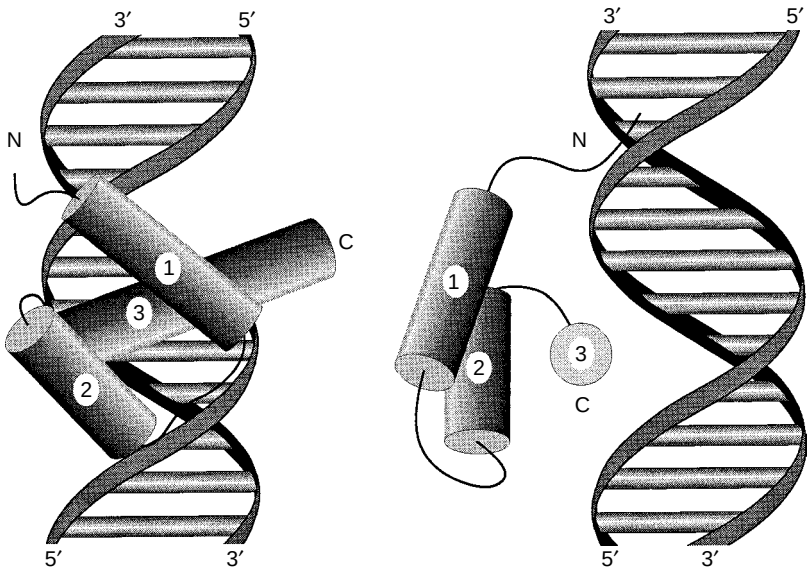

Fig. 2 Two views of a simplified schematic representation of the three-dimensional structure of the homeodomain and how it binds to DNA. The three numbered cylinders show helices 1, 2, and 3; N marks the amino terminus of the homeodomain and C the carboxyl terminus. The double helix represents the DNA. The left panel is rotated by about 90° with respect to the right panel.

DNA backbone (indicated by "#"). As can be seen, these residues are not only found in helix 3 but also in helix 2 and at the amino terminus of the homeodomain. The amino terminus reaches into the minor groove of the DNA, where the residues marked "m" can contact the DNA in the minor groove. A very critical residue that determines sequence specificity is found at position 9 of helix 3 (i.e. position 50). Exchanging this amino acid can cause drastic changes in the type of DNA sequence that is recognized by a particular homeodomain. Several sequences that are bound by homeodomains have been identified. A binding site for Antp is AGCCATTAGA, with the core being ATTA (TAAT on the other strand). This core-binding site is too small for providing sufficient specificity to activate only the intended target genes. Different types of homeodomain proteins employ different strategies to solve this problem. Some combine several DNA-binding domains in a single protein, some form homodimers, and many form heteromeric complexes with homeodomain or other types of transcription factors.

1.5 Classification of Homeodomains

The purpose of comparing and classifying the sequences of homeobox-containing genes is to determine evolutionary relationships between the different genes to find out which genes are orthologous between different species and which ones merely belong to the same group or class. With the advent of many completed genomes, this is now a much easier task. The resulting classification allows one to

examine how structural conservation in evolution correlates with functional conservation. In order to arrange the homeodomain sequences into logical groups, the terms *superclass*, *class*, and *family* are used here. Their use reflects the hierarchy of relationships such that a superclass encompasses several classes and a class can be further subdivided into families.

Figure 3 shows a comparative tree that has been derived from a much larger tree; this tree presents selected genes of different classes and families. Such a tree provides an easy means of grouping homeodomain sequences for classification. All superclasses and classes as well as most families presented here are well conserved in evolution, and members can be found in deuterostomes as well as protostomes. In fact, a very useful criterion for the identification of families is that they should be conserved over a long time period. The ideal definition for a family would be that its members were derived from a single gene in the ancestor of protostomes and deuterostomes; classes would comprise several families with common features. If a homeobox gene, whose homeodomain sequence is very divergent from others, is only found within related species of a particular phylum, it indicates that it is probably derived from another homeobox gene, for example, by duplication, and has undergone rapid evolutionary changes. Generally, classes of homeobox genes have less than 50 to 55% sequence identity within their homeodomains when compared to other classes. Within a particular family, the homeodomain can be 80 to 90% identical from flies to vertebrates. The genes within the Antp superclass, particularly those of the HOX cluster, are somewhat of an exception since they tend to be highly conserved, especially those in the center of the cluster.

Many of the homeodomain classes have additional conserved protein domains or motifs outside of the homeodomain. Such sequence motifs, which sometimes can be even more conserved than the homeodomain itself, provide extra criteria to differentiate the various classes of homeobox genes. Figure 4 shows a schematic diagram of homeobox gene families and classes that encode special conserved sequence motifs outside of the homeodomain. Homeodomain proteins of several families of the Antp superclass contain a short "Hexapeptide" motif upstream of the homeodomain. Conserved domains found in other classes can be 100 amino acids or longer, and for many, the structure has been determined either by X-ray crystallography or NMR analysis. Some of these domains can even be repeated within the same homeodomain protein, and some homeodomain proteins have several homeodomains.

The sections that follow introduce the various classes of homeobox genes. Figures 5 and 6 present an overview of homeodomain sequences of selected genes from the various classes and families. Detailed descriptions of expression patterns and gene functions are beyond the scope of this article, and the reader is referred to the literature and PubMed (http://www.ncbi.nlm.nih.gov/).

1.6 Atypical Homeodomains

Quite a number of homeobox genes have now been found whose homeodomain is different in length from the typical 60 amino acids. In order to get optimal alignments for these homeodomains, extra amino acids have to be looped out in various places. This was first noticed in yeast MATα2, where a better alignment

of the helix 1 region to the consensus could be achieved by looping out three extra amino acids. As already pointed out, X-ray crystallography of the MATα2 homeodomain revealed that its structure is indeed very similar to en and Antp, and that three amino acids are in fact looped out between helix 1 and helix 2. Many genes have now been isolated that have three extra amino acids between helix 1 and helix 2 from a variety of organisms (Fig. 6, TALE superclass, see below).

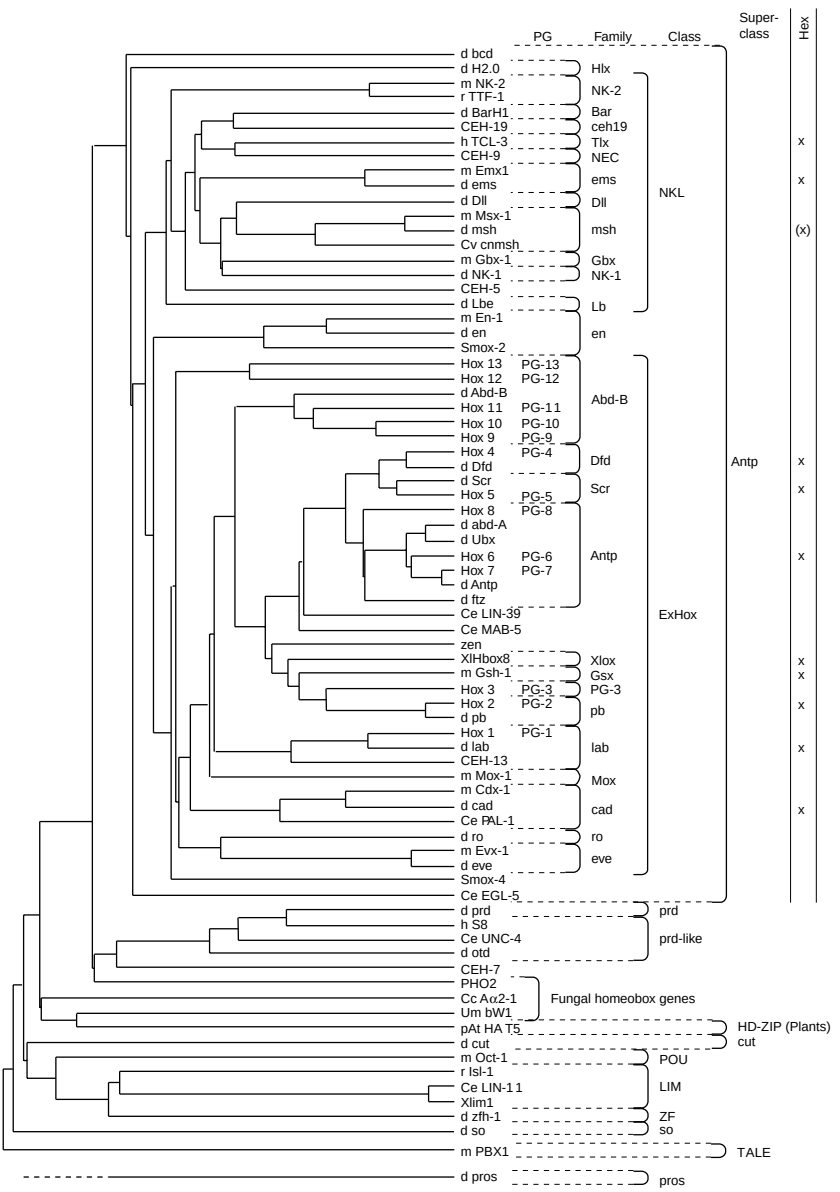

Extra amino acids have also been found in several genes between helix 2 and helix 3, for example, in the liver transcription factor LFB1(HNF1α), in the prospero protein, in fly defective proventriculus (dve), or in plant HD-ZIP III proteins (see 4.2, 5.4, 10.2). The homeodomains of the fungal proteins bE and AαZ have additional amino acids both between helix 1 and helix 2 and between helix 2 and helix 3. Homeodomain sequences that require such insertions and/or deletions are referred to as *atypical homeodomains* to contrast them from the typical homeodomain with 60 amino acids. However, "atypical" by itself is not a useful descriptor for the classification of homeobox genes. In fact, insertions and deletions in the homeodomain have occurred many times independently in evolution. Even within typical classes or families, such divergence is possible. For example, unlike other sine oculis (so) class homeobox genes, the *C. elegans* so class protein CEH-35 has one extra residue between helix 1 and helix 2.

2 The Antennapedia Superclass

This is a large group of genes that are related to one of the first homeobox genes discovered, Antennapedia. A number of these genes are organized into gene clusters, that is, the HOX cluster, the ParaHox cluster, which is a "sister" of the HOX cluster, and the NK cluster. However, many homeobox families within the Antp superclasses do not reside in a cluster and are referred to as *dispersed*. Common features of the Antp superclass genes are that they do not encode large conserved domains outside of the homeodomain but only small motifs. The Hexapeptide motif is found in genes of the HOX, ParaHox, and NK cluster, as well as in dispersed genes, such as the ems family (Fig. 4). The Hexapeptide is usually separated by an intron from the homeodomain. Other small motifs are found in NK-2 homeodomain proteins or in en homeodomain proteins (Fig. 4). Within the Antp superclass, two major groups can be distinguished, the Extended Hox (ExHox) class of genes with the Hox, Mox, Evx, and ParaHox genes and the NKL class of genes with the NK cluster genes and gene families such as NK-2, Dll, ems, NEC, Bar, Xnot, Gtx, Gbx, and Mnx (Fig. 3).

2.1 The HOX Cluster

The first, and also best-known, homeobox genes are the homeotic selector genes of *D. melanogaster* that are organized into

Fig. 3 Comparative tree showing different families and classes of homeobox genes. The tree is based on the aligned homeodomain sequences. The more similar the two homeodomain sequences are to each other, the shorter the horizontal distance is from the branch point to the endpoint, which represents the homeodomain of a particular gene. Conversely, the longer a branch of the tree is, the less related that branch is to other homeodomains. At the endpoints of individual branches, examples of particular homeobox genes for selected families and classes are given. The paralogue groups of the vertebrate HOX clusters are indicated by PG, and parentheses mark the families, classes, and superclasses, which are indicated on the right. Those genes that are not assigned to classes are either divergent members of existing classes (e.g. LIN-39, MAB-5, zen) or are orphans for which the origin has not been determined (e.g. CEH-7, CEH-5). The column at the right indicates families having a Hexapeptide upstream of the homeodomain.

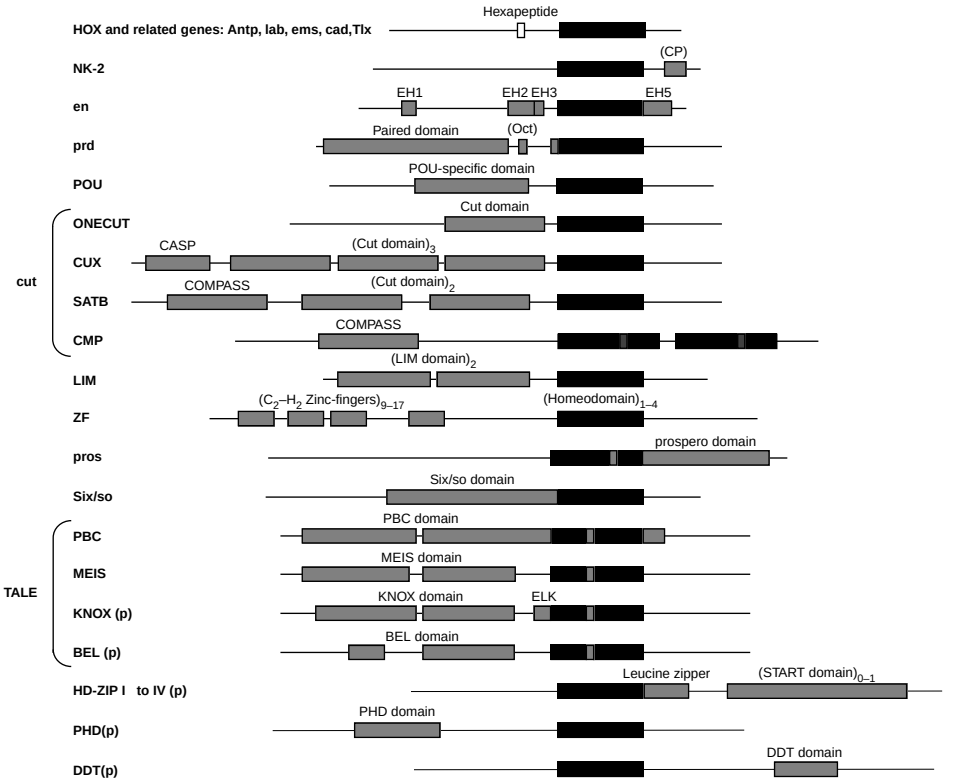

Fig. 4 Schematic representation of families and classes of homeobox genes encoding conserved motifs outside of the homeodomain. On the left are the names of the different selected families and classes. The black box represents the homeodomain, and insertions within the homeodomain, either between helix 1 and 2 or helix 2 and 3, are indicated by differently patterned boxes. The other boxes with different patterns represent conserved sequence motifs specific to individual classes or superclasses. The length of the boxes is approximately proportional to the size of the domains. The connecting linker regions (black lines) can be substantially different between genes and do not reflect true length. The Hexapeptide is found in several different families within the Antp superclass, mainly in *Hox* genes. In several instances, the number of domains can vary. The CP motif in the NK-2 family and the Oct motif in the paired class do not occur in all members of the class/family. In the ZF class, the number of zinc fingers as well as the number of homeodomains (one to four) can vary substantially, and the zinc fingers can be interspersed with the homeodomains. The lower part of the panel shows plant homeodomain classes (p). The plant HD-ZIP III and IV classes have a START domain, HD-ZIP I and II do not.

two complexes, the Antennapedia complex (ANT-C) and the Bithorax complex (BX-C), summarily referred to as the *homeotic gene complex* (HOM-C). In *Drosophila*, only one such cluster is found, while higher vertebrates have four homologous clusters, each on a different chromosome (Fig. 7). These clusters have been termed *HOX clusters* in vertebrates, and the usage of naming these clusters "HOX clusters" has now been expanded to include other animal phyla, including *Drosophila*. The

four vertebrate HOX clusters originally arose through duplications of a single cluster at some point in early vertebrate evolution. Indeed, in the cephalochordate amphioxus (lancelet), only a single cluster has been found. On the other hand, one has to note that in teleost fish, extra HOX clusters (as well as extra homeobox genes and extra copies of many other genes) are found since it is thought that the ancestor of this group of fish underwent an extra genome duplication event. HOM-C related clusters of homeobox genes are now commonly referred to as *HOX cluster* in all species where they are found. However, the use of the word Hox to name individual genes in the HOX cluster should be restricted to vertebrates.

Figure 7 shows the *C. elegans, Drosophila,* and vertebrate HOX clusters. The genes in the four vertebrate clusters (termed *HOXA, HOXB, HOXC, HOXD*) can be aligned such that 13 groups are formed that are highly related based on their sequence. These groups are referred to as *paralogue groups 1 to 13* (PG-1 to PG-13), and individual genes in mice are named *HoxA-1, HoxA-2,* and so on. Each cluster has missing genes for some of the paralogue groups. This suggests that the vertebrate cluster duplication took place after all 13 groups had formed and that during or after the duplication events, some genes were lost. Overall, there are 39 Hox genes in mouse and human. In vertebrates, but not in flies, the Evx genes are closely linked to the HOX cluster, indicating that these genes are part of the ancestral HOX cluster, although this gene family has a separate name. Likewise, the Mox genes are also associated with the HOX cluster, albeit at the opposite end (not shown). Both Evx and Mox genes are referred to as *extended HOX cluster genes.*

The lines and brackets in Fig. 7 between the fly and vertebrate clusters indicate how the Hox genes are related to the fly genes. For example, the *Drosophila* gene *labial* is the orthologue of the PG-1 genes *HoxA-1, HoxB-1,* and *HoxD-1* in mice. Some genes have no direct one-to-one correspondence. For example, the fly gene *Abd-B* is homologous to PG-9 through PG-13, probably due to a separate expansion of the Abd-B family genes in the deuterostome lineage. Further, the central genes in the cluster (PG-6 to PG-8, and ftz to abd-A) may have arisen independently by gene duplication within each phyla. A degenerate HOX cluster containing seven homeobox genes has been found in the nematode *C. elegans*. Two Abd-B family genes have split from the cluster, while the remaining five homeobox genes have also been separated into two subgroups by a series of other genes (Fig. 7). Recently, it has become apparent that an ancestor of *C. elegans* seems to have lost several homeobox genes since in several other nematode species, additional homeobox genes are present in the HOX cluster.

The genes of the HOX cluster have received particular attention not only because they are clustered on the chromosome but also because the way they are physically clustered correlates with the way they are expressed along the anterior–posterior body axis (Fig. 8). Thus, genes at one end of the cluster are expressed and function in the anterior body region, while, as one progresses along the chromosome, the genes are expressed and function further and further toward the posterior of the animal. This colinearity of genes and expression pattern is even more striking when one considers that the Hox genes are all transcribed in the same direction. Thus, the genes of PG-13 are usually referred to as the genes of the 5′ end of

the cluster and are expressed posteriorly, while those of PG-1 are at the 3′ end of the cluster and are expressed anteriorly.

The Hexapeptide, also known as *Pentapeptide*, was first described in HOX cluster genes. All genes in the HOX cluster, apart from the Abd-B family of genes encode this short conserved peptide motif upstream of the homeodomain. The Hexapeptide has a conserved core sequence of six amino acids (Fig. 9). Apart from the Hexapeptide, HOX cluster gene families often possess additional sequence conservation outside of the homeodomain. In particular, short stretches of the amino-terminal region of several families are conserved; in some cases, the similarity exists even between different families.

2.1.1 Labial (lab) Family

The labial family of genes is located at the 3′ end of the cluster. The *Drosophila* lab homeodomain is 80 to 85% identical to the PG-1 group genes, and the homeodomains of the lab family are 55 to 67% identical to other HOX cluster genes. The lab family genes encode distinct Hexapeptide sequences upstream of the homeodomain (Fig. 9).

2.1.2 Proboscipedia (pb) Family

The vertebrate paralogue group PG-2 is closely related to the *Drosophila* gene *pb*. Members of this family seem to be lacking in *C. elegans*, suggesting that homeobox genes can be lost in evolution.

2.1.3 PG-3 Family

The PG-3 group of genes seemed to have no corresponding gene in *Drosophila*. However, analysis of HOX cluster genes in other arthropods suggests that the *Drosophila* genes *zen* and *zen2* are the homologs of the PG-3 genes, and that they diverged substantially in flies. They have taken on novel roles during development, that is, dorsal–ventral patterning, rather than anterior–posterior patterning. Another *Drosophila* gene not involved in anterior–posterior patterning is *bicoid* (*bcd*). *bcd* is the most divergent gene in the *Drosophila* HOX cluster (Fig. 3). Its functional role is as a maternal morphogen in the early *Drosophila* embryo, where the protein forms a gradient in the egg. Recent evidence indicates that *bcd* has arisen by a duplication from a PG-3 gene during insect evolution and has undergone rapid sequence divergence as an adaptation to its novel function.

2.1.4 Deformed (Dfd) Family

This family comprises the *Drosophila Deformed* gene and the vertebrate PG-4 group genes. This family also has small

Fig. 5 Examples of homeodomain sequences of different groups. Only the homeodomain sequences are shown and flanking sequences have been omitted. The homeodomain sequences have been grouped on the basis of classes and families, and major groups have been boxed and labeled on the right. In the case of the SATB class sequence, one residue has been removed in the loop between helix 1 and 2 to maintain alignment. In this figure and all subsequent sequence alignments, the following conventions apply: A consensus (based on the most frequently encountered amino acid at a particular position) is shown at the top and dots indicate identities to the consensus. Dashes indicate a gap introduced for optimal alignment. Where sequences are incomplete or unavailable, empty spaces have been left; an asterisk (*) marks a stop codon. Some of the species codes used in this and subsequent figures are d, *D. melanogaster*; Ce, *C. elegans*; m, mouse; r, rat; c, chicken; h, human; zf, zebra fish; x, *Xenopus laevis*.

```
....|...10....|...20....|...30....|...40....|...50....|...60
RRRKRTAYTRYQLLELEKEFHFNRYLTRRRRIELAHSLNLTERQVKIWFQNRRMKWKKEN   Consensus

HOX cluster homeobox families

NNSG..NF.NK..T.............A....I.NT.Q.N.T...........Q..RV   d lab          lab
PNAV..NF.TK..T.........K....A..V.I.A..Q.N.T...........Q..RE   m HoxA1
PGGL..NF.TR..T.........K..S.A..V.I.AT.E.N.T...........Q..RE   m HoxB1
PSAI..NFSTK..T.........K....A....I.NC.Q.NDT...........Q..RE   m HoxD1
NGTN..NF.TH..T.......TAK.VN.T..T.I.SN.K.Q.A...........E..RE   Ce CEH-13

P..L.....NT............K..C.P....I.A..D.......V........H.RQT   d pb           pb
S..L.....NT............K..C.P..V.I.AL.D.......V........H.RQT   m HoxA2

LK.S...F.SV..V...N..KS.M..Y.T....I.QR.S.C.............F..DI   d zen          PG-3
SK.G......P..V............M.P..V.M.NL.......I..........Y..DQ   m HoxA3

P..T..TF.SS.IA...QH.LQG....AP.LAD.SAK.A.GTA......K...RRH.IQS   d bcd

PK.Q......H.I........Y...........I..T.V.S...I.............D.   d Dfd          Dfd
PK.S......Q.V...................I..T.C.S................DH   m HoxA4

EK.Q......N.V........THK....K....V....M................H....   Ce LIN-39

TK.Q..S.....T...................I..A.C.....I..........L...H   d Scr          Scr
GK.A........T...................I..A.C.S...I.............D.   m HoxA5

.K.G.QT.....T...................I..A.C.....I...............   d Antp         Antp
...G.QT...F.T..........H........I..A.C.....I..........L...L   d abd-A
...G.QT.....T........T.H........M..A.C.....I..........L...I   d Ubx
SK.T.QT.....T............I......DI.NA.S.S...I..........S..DR   d ftz
G..G.QT.....T...................I.NA.C.....I...............   m HoxA6
.K.G.QT.....T...................I..A.C.....I.............H   m HoxA7
...G.QT.S...T.......L..P....K....VS.A.G....................   m HoxB8
QK.T.QT.....T..........K........I..T.T.....I......            Sm Smox-1
SK.T.QT.S.S.T........YHK....K..Q.ISET.H................H...A   Ce MAB-5

SKKG.QT.Q...TSV..AK.QQSS.VSKKQ.E..RLQTQ..D..I..........A...K   Ce EGL-5
SK.I.....SI.........QN....S.L...QI.AI.D...K..........V....DK   Cv cnox2
     CSFGHRKII...R..KY......D..L.F.RN.D.S.S.I.V........Q...Q   Cv cnox1
H.KA..I.GTT.TQQ..DM.KGQM.VVGAE.EN..QR.G.SPS..R.......S.HRRKQ   Ce CEH-23

V.K..KP.SKF.T.......L..A.VSKQK.W...RN.Q................N..NS   d Abd-B        Abd-B
T.K..CP..KH.T.......L..M....D..Y.V.RL..................M..I.   m HoxA9
G.K..CP..KH.T.......L..M....E..L.ISR.VH..D............L..M.   m HoxA10
S.K..CP..K..IR...R..F..V.INKEK.LQ.SRM....D............E..L.   m HoxD11
A.K..KP..KQ.IA...N..LV.EFIN.QK.K..SNR...SDQ............K.RVV   m HoxD12
G.K..VP..KL..K...N.YAI.KFINKDK.RRISAAT..S....T.......V.D..IV   h HoxD13
GKK..QP.KKD.ISR..Y.YSV.Q...NK..S..SAQ.M.D.K...V...VCYRQYSAES   Ce NOB-1

V..Y...F..D..GR.....YKEN.VS.P..C...AQ...P.STI.V........D.RQR   d eve            eve
M..Y...F..E.IAR.....YREN.VS.P..C...AA...P.TTI.V........D.RQR   m Evx-1

A.KE...F.KE..R...A..AHHN....L..Y.I.VN.D.S.....V..........RVK   mox-1           MOX

ParaHox cluster homeobox families

SK.M...F.ST......R..AS.M..S.L....I.TY...S.K..........V.H...G   m Gsh-1         Gsx

NK.T......A.........L..K.IS.P..V...VM......HI..............E   x Xlhbox8       Xlox
NK.T....S.A.........HYDK.IS.P..L...A......CHI...........E.FE   Hm Lox3a
NK.T....S.S..F........DK.IS.P..V...S.......HI.............ME   Htr-A2
```

(a)

regions of conservation in the amino-terminal part of the protein.

2.1.5 Sex-combs-reduced (Scr) Family

The *Drosophila Scr* genes and the PG-5 group genes are very similar to the Antp family genes, but probably form their own family since the Hexapeptide sequences of Scr and the PG-5 proteins are more similar to each other than to Hexapeptides of other Hox cluster proteins. The homeodomain of the *C.*

```
....|...10....|...20....|...30....|...40....|...50....|...60
RRRKRTAYTRYQLLELEKEFHFNRYLTRRRRIELAHSLNLTERQVKIWFQNRRMKWKKEN   Consensus

KDKY.VV..DF.R......YCTS..I.I..KS...QT.S.S...........A.ERTS.   d cad            cad
KDKS.VV..DH.R........YS..I.I..KS...AN.G.............A.ER.V.   m Cdx-1
KDKY.VV..DQ.R........YS..I.I..KA...VN.G.S.T.........A.ER.I.   x Xcad1
ADKY.MV.SD..R........TSPFI.SD.KSQ.STM.S.....I........A.DRRDK   Ce PAL-1

NK cluster homeobox families

P..A...F.YE..VS..NK.KTT...SVCE.LN..L..S...T..........T....Q.   d slou/NK-1       NK-1
P..A...F.YE..VA..NK.RAT...SVCE.LN..L..S...T..........T....QH   c Sax1
M..A...F.YE..VA..NK.KTS...SVVE.LN..IQ.Q.S.T..........T....H.   Ce CEH-1
...A...F.YE..VT..NK.QST...SVYE.LN..L......T..........T....Q.   EgHbx1

KKKP..SF..L.IC....R..RQK..ASAE.AA..KA.KM.DA...T......T..RRQT   h TCL-3           Tlx
.KKP..SFS.S.V....RR.LRQK..ASAE.AA..KA.RM.DA...T......T..RRQT   m MUR10F

K.KS...F.NH.IF....R.LYQK..SPAD.D.I.A..G.SNAQ.IT......A.Q.RDI   d lbe             Lbx

KK.S.A.FSHA.VF...RR.AQQ...SGPE.S.M.K..R...T..........Y.T.RKQ   d bap/NK-3        NK-3

KKP.VLFSQA.V....CR.RLKK...GAE.EII.QK...SAT...........Y.S.RGD   d tin/NK-4        NK-4

QSKR.VLFNKF.ISQ...R.RKQ....AQE.Q....TIG..PT........HAY.M.RLF   EgHbx3

N.KP..PF.TQ...S...K.REKQ..SIAE.A.FSS..R...T..........A.A.RLQ   d msh/Dr          msh
N.KP..PF.TA...A..RK.RQKQ..SIAE.A.FSS..S...T..........A.A.RLQ   m Msx-1
N.KP..PFSVN...T..QK.KRKQ..SISE.A..SEL.R...T.I........A.Q.RSK   Cv cnmsh

Other Antp superclass families

K.KR.VLF.KA.TY...RR.RQQ...SAPE.EH..SLIR..PT........H.Y.T.RAQ   d vnd/NK-2        NK-2
K.KR.VLFSKA.TY...RR.RQQ...SAPE.EH..SLIR..PT........H.Y.M.RAR   m Nkx-2.2
..KR.VLFSQA.VY...RR.KQQK..SAPE.EH..SMIH..PT........H.Y.M.RQA   m TTF-1/Mlx-2.1
   R.VLFSQA.VY...RR.KQQK..SAPE.EH..SMIH..PT........H.Y.M.RQA   m Nkx-2.4
..KR.ILFSQA.IY...RR.KQQK..SAPE.EH..NLI...PT........H.Y.C.RSQ   Dt Dth-2
K.KR.VLFSKK.I....RH.RQKK..SAPE.EH..NLIG.SPT........H.Y.M.RAH   Dt Dth-1

.SWS.AVFSNL.RKG..IQ.QQQK.I.KPD.RK..AR....DA...V.........RHTR   d H2.0             Hlx
.SWS.AVFSNL.RKG...R.EIQK.V.KPD.KQ..AM.G..DA...V.........RHSK   m Hlx
GMLR.AVFSDV.RKA...T.QKQK.ISKPD.KK..SK.G.KDS............RNSK   m Dbx
GILR.AVFSED.RKA...M.QKQK.ISKTD.KK..IN.G.K.S............RNSK   c Chlx-A

PK.I...FSPS...K..HA.ES.Q.VVGAE.KA..QN...S.T...V......T.H.RMQ   d ems             ems
PK.I...FSPS...R..RA.EK.H.VVGAE.KQ..G..S.S.T...V......T.Y.RQK   m Emx1

M.KP..I.SSL..QQ.NRR.QRTQ..ALPE.A...A..G..QT..........S.Y..MM   d Dll             Dlx
I.KP..I.SSL..QA.NRR.QQTQ..ALPE.A...A..G..QT........K.S.F..LM   m Dlx
I.KP..I.SSL..QA.NHR.QQTQ..ALPE.A...A..GV.QT........K.S.Y..LI   x Xdll

.KKA..TFSGK.VF....Q.EAKK..SSSD.S...KR.DV..T..........T....IE   Ce CEH-9           NEC
KKKT..VFS.S.VFQ..ST.EVK...SSSE.AG..AN.H...T..........N...RQM   TgHbox5
KKKT..IFSKS.VFQ..ST.DVK...SSAE.AG..AA.H...T..........N.L.RQL   c Soho

EK.P...FSSE..AR.KR..NE.....E...QQ.SSE.G.N.A.I......K.A.I..ST   d en               en
DK.P...FSGT..AR.KH..NE.....EK..QQ.SGE.G.N.A.I......K.A.L..SS   d inv
DK.P...F.AE..QR.KA..QA...I.EQ..QT..QE.S.N.S.I......K.A.I..AT   m En-1
EK.P...F.GD..DR.KT..RES....EK..Q....E.G.N.S.I......K.A.L..ST   Ce CEH-16
LK.P..SF.VP..KR.SQ..EK....DEL..KK..TE.D.R.S........K.A.T..AS   Sm Smox-2
```

(b)

Fig. 5 (*Continued*)

elegans HOX cluster gene *lin-39* is about equally similar to that of Dfd family and Scr family homeodomains. Presently, it is thought that *lin-39* belongs to the Scr family.

2.1.6 Antennapedia (Antp) Family

This family comprises the fly genes *Antennapedia (Antp)*, *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*), *fushi tarazu* (*ftz*), and the vertebrate groups PG-6, PG-7,

```
....|...10....|...20....|...30....|...40....|...50....|...60
RRRKRTAYTRYQLLELEKEFHFNRYLTRRRRIELAHSLNLTERQVKIWFQNRRMKWKKEN   Consensus

Q.KA...F.DH..QT...S.ERQK..SVQE.Q....K.D.SDC...T.Y....T...RQT   d BarH1          Bar

..KA..VFSDP..SG...R.EGQ...SPPE.V...TA.G.S.T...T........H..QL   d bsh           bsh

E.KP.Q..SAR..DR..T..QTDK..SVNK..Q.SQT.....T.I.T......T....QL   Ce CEH-19       CEH19

.KGGQVRFSND.TI....K.ETQK..SPPE.KR..KM.Q.S.....T......A..RRLK   h PRH           HEX

C.P...F.SQ......HQ.KL.K..S.PK.F.V.T..M...T...............RSK   h HLXB9         Mnx

MK.I..VF.PE..EK.....LKQQ.MVGTE.VD..ST.....T...V......I..R.QS   x Xnot          Xnot

KKHS.PTFSGQ.IFA...T.EQTK..AGPE.AR..Y..GM..S...V......T..R.RH   m Gtx           Gtx

S..R...F.SE..........CKK..SLTE.SQI..A.K.S.V..........A...RIK   m Gbx-1         Gbx

Q..Q..TFSTE.T.R..V...R.E.IS.S..F...ET.R...T.I........A.D.RIE   d ro            ro

IP.R..TF.VE..YL..MY.AQSQ.VGCDE.ER..RI.S.D.Y..........IRMRR.A   Ce CEH-7
PK.P..VF.DE..EK..ES.NTSG..SGST.AK..E..G.SDN...V......T.Q..ID   Ce CEH-5
..KT..TFSNC..N...NN.NRQ....PTD.DRI.KH.G..NT..IT......A.L.R.A   Sm Smox-5
C.KP..VFSDL..MV..R..NNRK..STPQ.TN..DR.G.NQT...T.Y..........T   Cv cnox3
QK.A.VSFSSS.VHV..ER.DRQK..SSAE.A.MSRD.G.S.T..........Y.T..RA   EgHbx2
SF.N...F.D...IC..R..SHIQ..S.ID..H..QN.....K..........VR.R.R.   Sm Smox-4

Other classes

Q..C..TFSAS..D...RA.ERTQ.PDIYT.E...QRT....ARIQV..S...ARLR.QH   d prd           prd
Q..S..TF.AE..EA..RA.SRTQ.PDVYT.E...QTTA...ARIQV..S...ARLR.HS   d gsb-p
Q..S..TFSND.IDA..RI.ARTQ.PDVYT.E...Q.TG...AR.QV..S...ARLR.QL   d gsb-d
LQ.N..SF.QE.IEA.....ERTH.PDVFA.ER..AKID.P.ARIQV..S...A..RR.E   h PAX6
LQ.N..SF.ND.IDS.....ERTH.PDVFA.ER..GKIG.P.ARIQV..S...A..RR.E   d ey
LQ.N..SF.QE.IEA.....ERTH.PDVFA.ER..AKID.P.ARIQV..S...A..RR.E   m Pax-6
Q..S..TF.AE..E...RA.ERTH.PDIYT.E...QRAK...AR.QV..S...AR.R.QA   m Pax-3
Q..S..TF.AE..E....A.ERTH.PDIYT.E...QRTK...ARFQV..S...AR.R.QA   m Pax-7

Q..Y..TF.SF..E....A.SRTH.PDVFT.E...MKIG...ARIQV......A..R.QE   d al            prd-like
Q..N..TFNSS..QA..RV.ERTH.PDAFV.E...RRV..S.AR.QV......A.FRRNE   m S8
Q..I..TF.SL..K...RA.QETH.PDIYT.ED..LRID...AR.QV......A.FR.TE   Sm Smox-3
K..H..IF.Q..ID....A.QDSH.PDIYA.EV..GKTE.Q.DRIQV......A..R.TE   Ce CEH-10
...T..NFSGW..E...SA.EASH.PDVFM.EA..MR.D.L.SR.QV......A..R.RE   Ce UNC-4
QK.H..RF.PA..N...RC.SKTH.PDIFM.E.I.MRIG...SR.QV......A....RK   d W26
Q.....FF.QA..DI..QF.QT.M.PDIHH.E...RHIYIP.SRIQV......A.VRRQG   x Mix.1
K..H..IF.DE..EA..NL.QETK.PDVGT.EQ..RKVH.R.EK.EV..K...A..RRQK   m Gsc
S..E..I..PE..EAM.EV.GV...PDVSM.E...SR.GIN.SKIQV..K...A.LRNLE   EgHbx4
G..P...F..S.IEI..NV.RV.S.PGIDV.E...SK.A.D.DRIQ.......A.L.RSH   x XANF-1
Q..E..TF..A..DV..AL.GKT..PDIFM.E.V.LKI..P.SR.QV..K...A.CRQQL   d otd
Q..E..TF..S..DV..AL.AKT..PDIFM.E.V.LKI..P.SR.QV..K...A.CRQQQ   m Otx1

SKKQ.VLFSEE.KEA.RLA.ALDP.PNVGTIEF..NE.G.AT.TITN..H.H..RL.QQV   d cut           CUX
LKKP.VVLAPEEKEA.KRAYQQKP.PSPKTIED..TQ...KTST.IN..H.Y.SRIRR.L   h CDP
LKKP.VVLAPEEKEA.KRAYQQKP.PSPKTIE...TQ...KTST.IN..H.Y.SRIRR.L   dog Clox

PKKP.LVF.DV.RRT.HAI.KE.KRPSKELQ.TISQQ.G.ELST.SNF.M.A.RRSLDKW   r HNF6          ONECUT
VKVI.LTF.ET..KS.Q.S.QQ.HRP..EM.QK.SAT.E.DFST.GNF.M.S.RRLRIDQ   Ce CEH-21

KP.S.TKISLEALGI.QSFIHDVL.PDQEAIHT.SAQ.D.PKHTIIKF...Q.YHV.HHG   h SATB          SATB
```

(c)

Fig. 5 (*Continued*)

and PG-8, and the *C. elegans* gene *mab-5*. The homeodomains of this family are extremely conserved between flies and vertebrates (up to 98%). It seems likely that there existed only one common ancestral gene and that independent duplication events gave rise to the PG-6, PG-7, and PG-8 group genes in vertebrates and to the *Antp*, *Ubx*, *abd-A*, and *ftz* genes in flies. The *Drosophila ftz* gene is not involved in axial patterning, but functions in segmentation. A very cryptic Hexapeptide can be found upstream of the homeodomain, but in other insects and in arthropods, the

```
....|...10....|...20....|...30....|...40....|...50....|...60
RRRKRTAYTRYQLLELEKEFHFNRYLTRRRRIELAHSLNLTERQVKIWFQNRRMKWKKEN   Consensus

TT.V..VLNEK..HT.RTCYAA.PRPDALMKEQ.VEMTG.SP.VIRV....K.C.D..RS   r Isl-1          LIM
TK.M..SFKHH..RTMKSY.AI.HNPDAKDLKQ.SQKTG.PK.VLQV....A.A..RRMM   d ap
TK.M..SFKHH..RTMKSY.AI.HNPDAKDLKQ..QKTG..K.VLQV....A.A.FRRNL   r LH-2
..GP..TIKQN..DV.NEM.SNTPKPSKHA.AK..LETG.SM.VIQV......S.ERRLK   Ce MEC-3
..GP..TIKAK..ET.KNA.AATPKP..HI.EQ..AETG.NM.VIQV......S.ERRMK   Ce LIN-11
..GP..TIKAK..ET.KAA.AATPKP..HI.EQ..QETG.NM.VIQV......S.ERRMK   x Xlim1
NK.P..TISAKS.ET.KQAYQTSSKPA.HV.EQ..SETG.DM.V.QV......A.E.RLK   Ce CEH-14

DK.L..TI.PE..EI.YQKYLLDSNP..KMLDHI..EVG.KK.V.QV....T.ARER.GQ   h ATBF1 HD3       ZF
NK.L..TILPE..NF.YECYQSESNPS.KMLE.ISKKV..KK.V.QV....S.A.D..SR   d zfh-2 HD3
QK.F..QM.NL..KV.KSC.NDY.TP.MLECEV.GNDIG.PK.V.QV....A.A.E..SK   h ATBF1 HD4
KV.V...INEE.QQQ.KQHYSL.ARPS.DEFRMI.AR.Q.DP.V.QV....N.SRER.MQ   d zfh-1
NK.P..RI.DD..RV.RQY.DI.NSPSEEQIK.M.DKSG.PQKVI.H..R.TLF.ERQR.   h ATBF1 HD1
QK.A..RI.DD..KI.RAH.DI.NSPSEESIM.MSQKA..PMKV..H..R.TLF.ERQR.   d zfh-2 HD1
K.SS..RF.D...RV.QDF.DA.A.PKDDEFEQ.SNL...PT.VIVV....A.Q.AR.NY   h ATBF1 HD2
KAN..RF.D..IKV.QEF.EN.S.PKDSDLEY.SKL.L.SP.VIVV.....A.Q.QR.IY   d zfh-2 HD2

K.KR..TISVAAKDA..RH.GEHSKPSSQEIMRM.EE...EKEV.RV..C...QRE.RVK   m Pit-1          POU
..K...SIETNVRFA...S.LA.QKP.SEEILLI.EQ.HMEKEVIRV..C...Q.E.RI.   m Oct-2
..K...SIETNIRVA...S.ME.QKP.SEDITLI.EQ..MEKEVIRV..C...Q.E.RI.   m Oct-1
..K...SIETTIRGA...A.LA.QKP.SEEITQ..DR.SMEKEV.RV..C...Q.E.RI.   d pdm-1
K.KH..SIEDNVRHT..NY.MQCSKPSAQEIAQI.RE..MEKDV.RV..C...Q.G.RQV   x Oct-79
K.K...SIEVSVKGA..QH..KQPKPSAQEITS..D..Q.EKEV.RV..C...Q.E.RMT   d vvl/Cf1a
K.K...SIEVSVKGA..SH.LKCPKPSAQEITN..D..Q.EKEV.RV..C...Q.E.RMT   m Brn-1
K.K...SIEVNVKSR..FH.QS.QKPNAQEITQV.ME.Q.EKEV.RV..C...Q.E.RIA   Ce CEH-6
K.K...SIEANVKSI..SS.MKLSKPSAQDISS..EK.S.EKEV.RV..C...Q.E.RIT   Djpou1
.K....SIENRVRWS..TM.LKCPKPSLQQITHI.NQ.G.EKDV.RV..C...Q.G.RSS   m Oct-3/4
KK....SIAAPEKRS..AY.AVQPRPSSEKIAAI.EK.D.KKNV.RV..C.Q.Q.Q.R    r Brn-3
KK....SIAAPEKRS..AY.AVQPRPSGEKIAAI.EK.D.KKNV.RV..C.Q.Q.Q.RIV  d acj6/tI-POU
KK....SIAAPEKR...QF.KQQPRPSGE.IASI.DR.D.KKNV.RV..C.Q.Q.Q.RDF  Ce UNC-86

GEETSYCFKEKSRSV.RDWYSH.P.PSP.EKRD..EATG..TT..SN..K...QRDRAAE   d so             Six/so

Nonanimal homeobox classes

LPE.KRRL.TE.VHL...S.ETENK.EPE.KTQ..KK.G.QP...AV......AR..TKQ   HAT5             HD-ZIP I
LGE.KKRLNLE.VRA...S.ELGNK.EPE.KMQ..KA.G.QP..IA.......AR..TKQ   Athb-3

NS..KLRLSKD.SAI..ET.KDHST.NPKQKQA..KQ.G.RA...EV......ART.LKQ   HAT4             HD-ZIP II

STARKGHFGPVINQK.HEH.KTQP.PS.SVKES..EE.G..F...NK..ET..HSARVAS   Zmhox1a          PHD

KKP..QMK.PF..ET..RVYAMET.PSEAI.A..SEK.G..D..LQM..CH..L.D.NTS   Lp hox7          DDT

SKKP.PKFHSEYTPL..LY.R..A.P.YAD.RV..EKTGMLT..ITV....H.RRA.GPL   Sc A Y1          fungi
PPS.KQ.FNVHYIPV...Y.EY.A.P.AQD.AL..RKSMMSA..IEV....H.ARAR..G   Cc A 2-1
PLKTGRGHDSEAVRI..QA.KHSPNI.PAEKFR.SEVTG.KPK..T.......NRKG.K.   Um bW1
Q.P...RAKGEA.DV.KRK.EI.PTPSLVE.KKISDLIGMP.KN.R.......A.LR.KQ   yeast PHO2
A....RRTSSQE.SI.QA..EKCPAPSKEK.....E.CHM..KA.Q.....K.QAV.RQR   yeast YOX1
SPKGKSSISPQARAF..QV.RRKQS.NSKEKE.V.KKCGI.PL..RV..I.K..RS.*     yeast MATa1
```

(d)

Fig. 5 (*Continued*)

orthologous gene is better conserved, and in some species, it still seems to play a role in pattern formation.

2.1.7 Abdominal-B (Abd-B) Family

The Abd-B family of genes is located at the 5′ end of the cluster. They are more divergent than other cluster genes and they do not encode a Hexapeptide. While *Drosophila* has only one gene, *Abd-B*, there are five paralogue groups, PG-9 to PG-13, in vertebrates. The Abd-B homeodomain is 52 to 75% identical to the various human and mouse Abd-B family homeodomains; thus, these genes can be very divergent. The further removed from the center of the HOX cluster the paralogue groups are, the more divergent their homeodomain sequences become. *HoxD-9* to *HoxD-13* play important roles during pattern formation of limbs in vertebrates. In *C. elegans,* the homeodomain of PHP-3 is most similar to Abd-B (82% identical),

```
                                   Homeodomain
TALE
Consensus    VLSSLRRELSKKRKRGKLPKEATQILRAWWFQHLKNPYPSEEEKEALAEETGLTLKQINNWFINARRRHWKPMIDMQFAAMDR

d ara        SN..YGASYDLAAR.KNATR.S.AT.K..LNE.K.....TKG..IM..II.KM..T.VST..A.....LK.ENKMTWEPKNRT
d caup       AAYGYGPNYDLAAR.KNATR.S.AT.K..LSE.K.....TKG..IM..II.KM..T.VST..A.....LK.ENKMTWEPKNKT
h IRX2a      AAPLGSYPYGDPAY.KNATRD..AT.K..LNE.R.....TKG..IM..II.KM..T.VST..A.....LK.ENKMTWTPRNRS
Ce C36F7.1   AHRLYHPYGLDGI.KRNATR...AP.KD.LHS.R......KAD.VM..VG..M..T.VST..A.....LK.ENKMTWSPQNR.
h PBX1       .MILRS.F.DAR...RNFN.Q..E..NEYFYS..S.......A..E..KKC.I.VS.VS...G.K.I.YK.NIGKF.EE.NIY
m PBX1       .MILRS.F.DAR...RNFN.Q..E..NEYFYS..S.......A..E..KKC.I.VS.VS...G.K.I.YK.NIGKF.EE.NIY
h PBX3       .MILRS.F.DAR...RNFS.Q..E..NEYFYS..S.......A..E..KKCSI.VS.VS...G.K.I.YK.NIGKF.EE.NLY
h PBX2       .MILRS.F.DAR...RNFS.Q..EV.NEYFYS..S.......A..E..KKC.I.VS.VS...G.K.I.YK.NIGKF.EE.NIY
d exd        .MILRS.F.DAR...RNFS.Q.SE..NEYFYS..S.......A..E..RKC.I.VS.VS...G.K.I.YK.NIGKA.EE.NLY
CEH-20       .MILRS.F.DAR...RNFS.Q..EV.NEYFYG..S.......A..D..RQCNI.VS.VS...G.K.I.YK.N.AKA.EE.SMY
CEH-40       .IQLKK.Y.DAR...RNFS.TS.E..NEYFLANINH.....V.Q...MQCNISVA.VS...G.K.I.YK.T.AKNEDERREN
pBn HD1      RIEDV.E.IMR..RA....GDT.TV.KN..Q..C.W...T.DD.AK.V.....Q.......Q.K.N.HNNSHSLTSLKSK
pZm KNOX1    RIEDV.E.ILR..RA....GDT.S..KQ..QE.S.W...T.DD.AK.V.....Q.......Q.K.N.HNNN
pAt KNAT3    KIVDI.E.ILR..RA....GDT.SV.K...QS.S.W...T..D.AR.VQ....Q.......Q.K.N.HSNPSSSTVLKNK
pAt KNAT4    KIVDI.E.ILR..RA....GDT.SV.KS..QS.S.W...T..D.AR.VQ....Q.......Q.K.N.HSNPSSSTVSKNK
pAt KNAT5    KIVDI.E.IMR..RA....GDT.SV.KE..RT.S.W...T..D.AK.VQ....Q.......Q.K.N.NSNSSTSSTLTKN
pZm KNOX6    K.VDI.E.IMR..RA....GDTASV.K...QA.S.W...T.DD.AR.VQ....Q.......Q.K.N.HSN
pZm KNOX7    K.VDI.E.IMR..RA....GDTASV.K...QA.S.W...T.DD.AR.VQ....Q.......Q.K.N.HSN
pZm KNOX2    K.VDI.E.ILR..RA....GDTAST.K...QA.S.W...T..D.AR.VQ....Q.......Q.K.N.HNN
pGm SBH1     Y.G..KQ.FM....K.......R.Q.LE..NR.Y.W.....SQ.L....S...DQ........Q.K.....SE....VV..P
pAt STM      Y.G..KQ.FM....K.......R.Q.LD..SR.Y.W.....QQ.L....S...DQ........Q.K.....SE....VV..A
pAt KNAT1    Y....KQ....K.K.......R.K.LT..EL.Y.W.....S..V....S...DQ........Q.K.....SE....MV..G
pZm KNOX11   C..R..S.FL....K.....D.RSA.MD..NT.YRW...T..D.VR..AA...DP........Q.K.....S
pZm LG3      C..R..S.FL....K.....D.RTV.LE..NT.YRW...T..D.VR..AM...DP........Q.K.....S
pZm KNOX3    C.GN..K..C....KD......R.K.LS..EL.YRW.....M..I....S...EQ........Q.K.....S
pZm KNOX4    Y.....Q.F...K.K.......R.K.LH..EL.Y.W.....T..I....A...DQ........Q.K.....S
pOs OSH1     Y....KQ....K.K.....D.R.Q.LN..EL.Y.W.....SQ.V....S...D.........Q.K.....SDE...VM..G
pZm KN1      Y....KQ....K.K.......R.Q.LS..D..Y.W.....TQ.V....S...D.........Q.K.....SEE.HHLM..G
pAt ATK1     HI...KL.F...K.K....R..R.A.LD..NV.N.W...T.GD.I........DQ........Q.K.....SEN.P.DM..D
pSt POTH1    H....KL.F...K.K.......R.M.L...DD.FRW...T.AD.NS...S...DP........Q.K.....SEN..L.V..N
fy CUP9      FVE.KS.RKQNSGR.SN....TV...NT.LLN..N....TQQ..RE.LIK....KI.LS.....V...KIFSDYYTLVNSIPN
fy YGL096W   GY.AKEKKHKAHG..SN...ATVS..NK.LHE.VN....TVQ..RE.LAK....KL..S........KIFSGQNDANNFRRK
pAt ATH1     LQQLK.KNHQIW.PQRG..EKSVSV..N.M..NFLH...KDS..HL..IRS...RS.VS.......V.L....EEMY.E.NK
pAt Z35398   LGMIQHPSNNAW.PQRG..ER.VSV....L.E.FLH...KDSD.HM..KQ....RS.VS.......V.L....VEEMYMEEMK
pAt BEL1     YRQMTLVDAHPW.PQRG..ER.VTT....L.E.FLH....DVD.HI..RQ...SRS.VS.......V.L....EEMYCEET.
h TGIF       LDL.SSAGSG.R.R..N....SV....D.LYE.RY.A....Q..AL.SQQ.H.STL.VC........LLPD.LRKDGKDPNQ
m TGIF       LDL.SSAASGMR.R..N....SV....D.LYE.RY.A....Q..AL.SQQ.H.STL.VC........LLPD.LRKDGKDPNQ
c AKR        LDL.SSAGSG.R.R..N....SV....D.LYE.RY.A....Q..VL.SRQ.H.STL.VC........LLPD.LRKDGKDPNQ
CEH-25b      PSTPNGGRIG.S.G..IF..Q..NR..Q.L..N.TH......Q.KQ..K.....IL.V.........IVQ...QNNR.GRS
m Meis3      EDED.DL.RRRNK...IF..V..N.M...L....SH......Q.KQ..QD....IL.V.........IVQ...QSNRTGQG
h MRG2       EDED.DQ.PRRNK...IF..V..N.M...L....SH......Q.KQ..QD....IL.V.........IVQ...QSNRTGQG
m Mrg1       DDDDPDKDKKRQK...IF..V..N.M...L....TH......Q.KQ..QD....IL.V.........IVQ...QSNR.VSQ
m Meis1      DDDDPDKDKKRHK...IF..V..N.M...L....TH......Q.KQ..QD....IL.V.........IVQ...QSNR.VSQ
x XMEIS1-1   DDDDPDK.KKRNKG..IF..V..N.M...L....TH......Q.KQ..QD....IL.V.........IVQ...QSNR.VSQ
m Meis2      DDDDPDKDKKRQK...IF..V..N.M...L....TH......Q.KQ..QDA...IL.V.........IVQ...QSNR.GFL
fSp MAT1-Pi  DDHCSDIGRPMTTV..QCS.CTKPH.MR.LLL.YD.....NS.FYD.SAA....RT.LR...S.R..*
fy MATα2     .TQDMINKST.PYRGHRFT..NVR..ES.FAKNIE...LDTKGL.N.MKN.S.SRI..K..VS.R..KEKTIT.APEL.DLLS
fUm bE2      A.KKADASSPVVGC.DLSEDLPAYHM.KHFLHT.D....TQ....G.VRL.N.EVH.LTL........SGWSH.LKK..RE..
fUm bE4      A.KKANASSPVVGC.ELSEDLPAYHM.KHFLHT.E....TQ....T.VRL.N.EVH.LTL........SGWSH.LKK..RE..
fUm bE1      .MQEFNVVSPDVGC.NLSEDLPAYHM.KHFLLT.D....TQ....T.VRL.N.EVHHVTL........SGWSH.LKK..RE..
fUm bE3      .MQEANVVSPGEGC.NLSEDLPAYHM.NHFLHT.E....TQ....G.VRL.N.EVH.LTL........SGWSH.LKK..RE.P
fCc β1-1     DKT.GLDLSDD.NEPTSPTPAYVEPCAR.LKDNWY.....G.VRTQI.RQ.RTSR.D.DA...D....IGWNEVRRKHFENK.
fCc β4-1                DASPKEPAEPAYIEPSCR.LKDNWY.....PQVRSSI.KQ..ASR.D.DA...D..K.IGWNE
fSc AαZ4     AS.A.AD.KTGAELSATPLPPYIEPCYR.LVN..D....TKAI..E.LDQAR.A.GD.D....A..A.MGWGD.RRRLFGGS.
fSc AαZ5     DKLTASIAH.NASTSSLPLPPYIAPCYE.LMDN.H.....PPV.RR.L.AAP..IDD..D..VAV.A.MGWGDLRRVKFDGS.
fSc AαZ3     EALEFSENDCEWQENMPPVPPFIGACYE.LL...H.....N.GRLTPPSSPRK.FDD.DR..SA..I.IGWTH.CDTKFGG.K

d Antp                  R..G.QTYTRYQ.LE.EKEFHFN---R.LTRRRRIEI.HALC..ER..KI..Q.R.MKWK.EN
                        .........|.........|...abc......|.........|.........|.........|
                                  10        20        30        40        50        60
```

IRO
PBC
KNOX II
KNOX I
CUP
BEL
TGIF
MEIS
M-ATYP

Fig. 6 Examples of TALE homeodomain sequences. At the top, the three α-helices of the homeodomain and a consensus derived from 58 TALE sequences are shown. On the right, the different TALE classes from animals, fungi, and plants are indicated. The fly Antp homeodomain sequence is shown at the bottom for comparison. Plant genes are denoted by a "p" and fungal genes are denoted by an "f" in front of their species code, respectively. The thin line in the loop region between helix 2 and helix 3 in the fungal genes from *Ustilago maydis* (fUm) and *Schizophyllum commune* (fSc) indicates the region where extra residues were removed from the loop for this alignment.

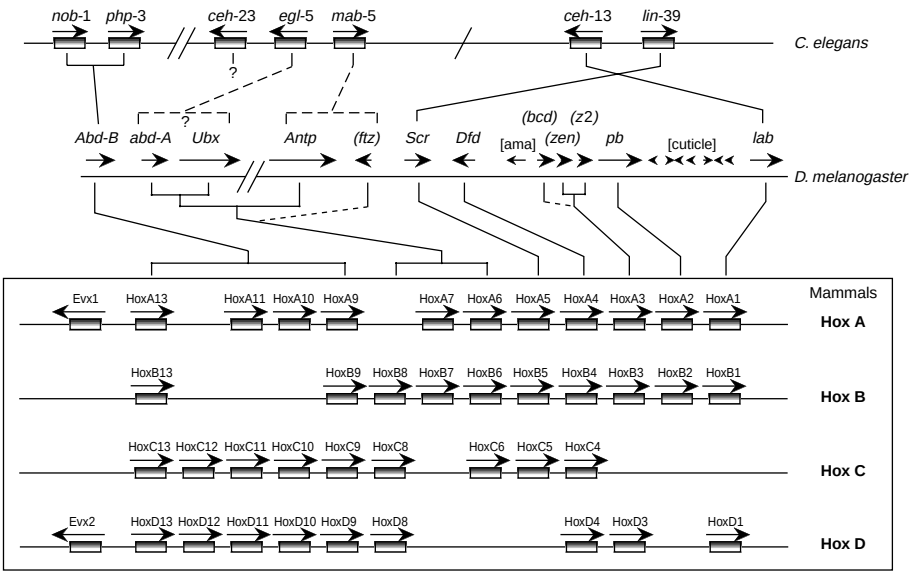

Fig. 7 Organization of the *C. elegans*, *D. melanogaster*, and mouse/human HOX clusters. At the top is a schematic representation of the *C. elegans* HOX cluster. The *Abd-B* genes (*php-3*, *nob-1*) have split far from the rest of the cluster. Of the remaining five genes in the cluster, only four genes can be assigned to fly/vertebrate homologs, in the case of *mab-5* and *egl-5*, only tentatively. The relationship of the homeobox gene *ceh-23* is unclear. In the center is a representation of the *D. melanogaster* HOX cluster, composed of the two complexes ANT-C (right) and the BX-C (left); large arrows indicate the individual transcription units of homeobox genes, while small arrows represent nonhomeobox genes (names in square brackets), and gene names of nonhomeotic homeobox genes are given in parentheses. At the bottom of the panel, the organization of the four mammalians – based on mouse and human-HOX clusters (HOXA, HOXB, HOXC, HOXD) is shown. The orientation of transcription is indicated by arrows. Black lines and brackets between the *C. elegans*, *D. melanogaster*, and mammalian clusters mark the homologous genes. Dashed lines indicate less certain sequence relationships or derived sequences.

while NOB-1 is only 52% identical, and EGL-5 53%. While *nob-1* is classified as an Abd-B gene, the assignment of *egl-5* is less certain, variously grouped either as Abd-B or as divergent Antp family gene (Fig. 7).

2.1.8 Even-skipped (eve) Family

The vertebrate eve genes (*Evx*) are located at the 5′ end of the HOX cluster (Fig. 7), suggesting that this gene family is part of the extended HOX cluster. In other organisms such as *Drosophila* and *C. elegans* though, the single eve family genes are not linked to the HOX cluster. Like the Abd-B family, the eve genes do not contain a Hexapeptide upstream of the homeodomain. Their homeodomain sequences are very different from the Antp and Abd-B family homeodomains, and in addition, the intron position is in a different location than in Hox genes. The original function of the eve family genes appears to be pattern formation in the posterior of the animal, which is consistent with its position in the HOX cluster. The eve family is very ancient since an *eve* gene has been isolated from corals, where it has

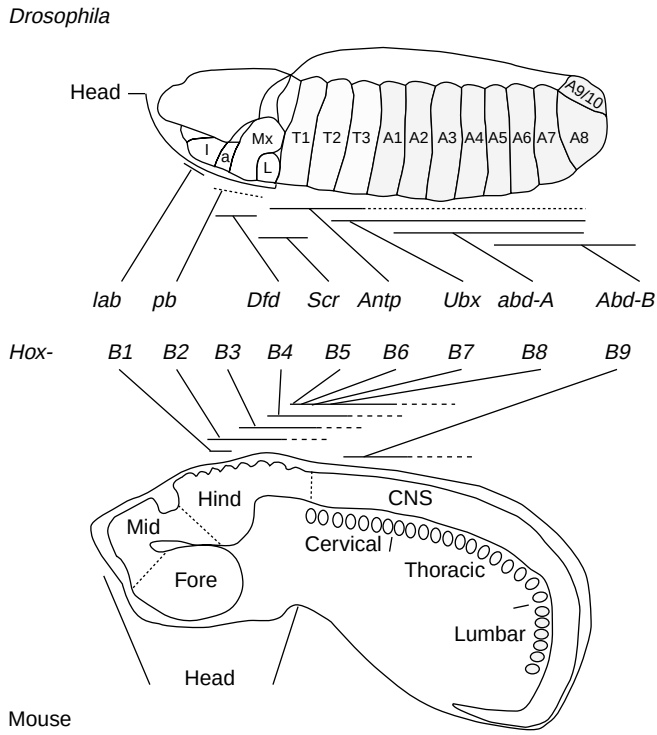

Fig. 8 Schematized domains of expression for the fruit fly HOX and the mouse HOXB cluster genes, as mapped onto a 10-h *Drosophila* embryo, or a mouse embryo at day 12, respectively. Bars indicate the approximate areas of expression for individual genes. Additional labels include I, a, Mx, L – intercalary, mandibular, maxillary, and labial segments of the presumptive head; T1–T3: thoracic segments; A1–A8: abdominal segments; Fore, Mid, Hind: forebrain, midbrain, and hindbrain; CNS: central nervous system. The labels cervical, thoracic, and lumbar indicate the somites of the respective body regions. Adapted from McGinnis, W., Krumlauf, R. (1992) Homeobox genes and axial patterning, *Cell* **68**, 283–302.

been found linked to a gene similar to HOX cluster genes.

2.1.9 Mox Family

The Mox genes were found to be linked to the HOX cluster in vertebrates, on the side of the anterior Hox genes, and thus belong to the extended HOX cluster. However, sequencewise, these genes are probably derived by duplication from the posterior genes, most likely from the eve family genes. The Mox family is conserved; a homolog, *buttonless*, has been found in *Drosophila*.

2.2 ParaHox Cluster

The ParaHox gene cluster is composed of three gene families, Gsx, Xlox, and cad (Fig. 10a). This cluster was first identified

Consensus	IYPWMK
d cad	SKPPYFDWMKKPAYPA
m Cdx-1	PRRTPYEWMRRSVAAA
Ce PAL-1	YQQSVWPFMDYQQFQG
d ems	DSYQLYPWLLSRHGRI
h EMX1	DPLHFYPWVLRNRFFG
d lab	SSIPTYKWMQLKRNVP
m HoxA1	SPAQTFDWMKVKRNPP
m HoxB1	LTPRTFDWMKVKRNPP
d pb	DSVPEYPWMKEKKTSR
h HOXB2	PPAPEFPWMKEKKSAK
m HoxB3	LTKQIFPWMKESRQTS
AmphiHox3	ANTKIYPWMKESRQNS
d Dfd	GERIIYPWMKKIHVAG
m HoxD4	QPAVVYPWMKKVHVNS
m HoxA4	KEPVVYPWMKKIHVSA
m HoxB4	KEPVVYPWMRKVHVST
m HoxC4	KQPIVYPWMKKIHVST

Consensus	IYPWMK
D Scr	NPPQIYPWMKRVHLGT
m HoxA5	AQPQIYPWMRKLHISH
m HoxB5	QTPQIFPWMRKLHISH
d Antp	MPSPLYPWMRSQFGKC
d Ubx	SNHTFYPWMAIAGECP
d abd-A	ADLPRYPWMTLTDWMG
m HoxC6	ASIQIYPWMQRMNSHS
m HoxB6	CSTPVYPWMQRMNSCN
m HoxB7	SNFRIYPWMRSSGPDR
m HoxA7	ASFRIYPWMRSSGPDR
m HoxB8	SPTQLFPWMRPQAAAG
m HoxD8	SPSQMFPWMRPQAAPG
m HoxC8	SPSLMFPWMRPHAPGR
TgHbox1	KSTSGYPWMPVAGPNV
Ce MAB-5	ISQPVFPWMKMGGAKG
Smox-1	SNVVVYPWMNPKGTDI
h TLX1	LTGLTFPWMESNRRYT
XlHbox8	RTLLPFPWMKSTKSH
h MSX1	EKPERTPWMQSPRFSP

Fig. 9 Hexapeptide sequences. The top line shows a consensus sequence for the core of the Hexapeptide. Examples of Hexapeptide sequences from various families and classes are given.

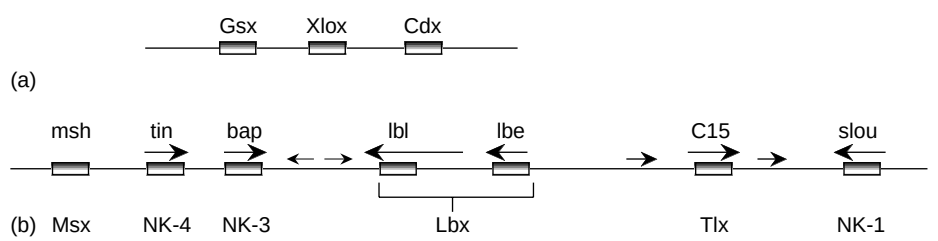

Fig. 10 (a) Schematized gene arrangement for the ParaHox gene cluster as found originally in Amphioxus. It is composed of three gene families Gsx, Xlox, and Cdx, Gsx corresponding to the most anterior and Cdx to the most posterior functioning gene. (b) Schematic representation of the NK gene cluster. This is a derived composite of the clusters found in mosquito (*Anopheles gambiae*) and *D. melanogaster*, respectively. In *Drosophila*, the *msh* gene is not linked to the cluster anymore, while in mosquito, the *NK-1* gene is not linked to the cluster anymore. Furthermore, there is only a single *Lbx* gene in mosquito.

in Amphioxus, but a cluster is also present in mammals. Nevertheless, this cluster is not as highly conserved as the HOX cluster. For instance, only one intact ParaHox cluster is found in mammals, while paralogous genes are no longer found in clusters. The *ParaHox* genes are very similar in sequence to the HOX cluster genes and they all encode a Hexapeptide upstream of the homeodomain (Fig. 9). It is thought that the ParaHox cluster probably arose through a duplication event from an ancestral HOX cluster. The ParaHox genes also function in anterior–posterior patterning, with the Gsx genes being the most anterior, and the cad genes being the most

posterior. An important functional aspect of the ParaHox cluster seems to be in the patterning of the endoderm.

2.2.1 Gsx Family

These genes were first isolated from mouse (*Gsh-1, Gsh-2*). On the basis of the homeodomain sequence, the Gsx and Xlox (see next section) proteins are very similar to the pb and PG-3 family of HOX cluster genes, confirming the notion that the ParaHox cluster arose through some duplication event of a ProtoHox cluster.

2.2.2 Xlox Family

The *Xenopus* gene *Xlhbox8* and several leech genes (*lox*) form the founding members of this family. In *Drosophila*, the *Xlox* gene seems to have been lost, while in *C. elegans* both *Xlox* and *Gsx* genes are lacking.

2.2.3 Caudal (cad) Family

The function in *Drosophila* and the expression patterns in vertebrates indicates that these genes are important in axial pattern formation in the posterior. Often, they play a role in the development of the endoderm. The *C. elegans* gene *pal-1*, while clearly a member of the cad family, does not have a readily recognizable Hexapeptide upstream of the homeodomain (Fig. 9).

2.3 The NK Cluster

The NK cluster was discovered in *Drosophila*, where a series of homeobox genes were found to be clustered in the chromosomal region 93DE (Fig. 10b), hence it is also referred to as the 93DE cluster in *Drosophila*. A corresponding cluster has also been identified in mosquito, although the gene arrangement has slightly changed since in mosquito, the NK-1 orthologue is separate from the cluster, while the Msx gene is still in the cluster (Fig. 10b). This cluster, however, has not been that well conserved in vertebrates, only fragments still exist and the genes have dispersed through the genome. A main function of the NK cluster genes is in pattern formation and development of the mesoderm.

2.3.1 NK-1 Family

The homeodomains of the NK-1 class have been highly conserved between flatworms, nematodes, flies, and vertebrates (Fig. 7). The fly gene *slouch* (*slou/NK-1/S59*) is orthologous to the vertebrate genes *Nkx-1.1* and *Nkx-1.2* (also known as *Sax*).

2.3.2 Tlx Family

The human gene *TLX1* (*TCL-3*) is one of the better-known cases of a homeobox-containing genes involved in cancer; in this case T-cell leukemia, and, thus, this gene family was named Tlx. The orthologous gene, *C*15, from *Drosophila* has not been studied well yet. The Tlx genes encode a Hexapeptide upstream of the homeodomain (Fig. 9).

2.3.3 Lbx Family

There are two Lbx family genes in *D. melanogaster, ladybird early* (*lbe*) and *ladybird late* (lbl). This seems to be a recent duplication from a single, ancestral *Lbx* gene since in *Anopheles* there is only a single *Lbx* gene. In vertebrates, the *Lbx* genes are still found linked to the *Tlx* genes; there are two paralogous miniclusters with *Lbx1* and *Tlx1*, and *Lbx2* and *Tlx2*, respectively.

2.3.4 NK-3 Family

This family is defined by the fly gene *bagpipe* (*bap*) and the vertebrate genes *Nkx-3.1* and *Nkx-3.2*.

2.3.5 NK-4 Family

This family is defined by the vertebrates genes *Nkx-2.3, Nkx-2.6,* and *Nkx-2.5* and the fly gene *tinman* (*tin*). Together with the NK-2 family, some of the NK-4 family genes share a conserved region (CP) downstream of the homeodomain (Fig. 4).

2.3.6 Msx Family

The Msx family of genes were first identified in vertebrates (as *Hox-7* genes) and in flies a corresponding gene, *msh* (*muscle-specific-homeobox*), was identified. The vertebrate genes were renamed as *Msx* genes because they were not part of the HOX cluster. The Msx family is very ancient since a member has been cloned from hydra. The sequence conservation extends outside of the homeodomain; there are at least eight amino acids immediately upstream of the homeodomain and five amino acids immediately downstream that are conserved between hydra and vertebrate Msx proteins. Also, a cryptic Hexapeptide is found in some of the genes (Fig. 9).

2.4 Noncluster Antennapedia Superclass Genes

A large group of Antp superclass gene families that are conserved from protostomes to deuterostomes are dispersed through the genome and are not, or no longer, organized in clusters. Many of them, such as NK-2, Dll, and ems, are similar to genes in the NK cluster and are loosely grouped into the NKL class.

2.4.1 NK-2 Family

The NK-2 family of homeobox genes is defined by the fly homeobox gene *NK-2* (now known as *ventral nervous system defective, vnd*). The first vertebrate gene of this family to be isolated was the transcription factor *TTF-1*. The vertebrate genes are now called *Nkx-2.1* and *Nkx-2.2*. Several of the fly and vertebrate proteins share a small conserved sequence element of about 17 amino acids downstream of the homeodomain, referred to as the *conserved peptide* (CP), or *NK2-SD* (Fig. 4). A small conserved region is also found near the amino terminus, referred to as *TN-domain*. The NK-2, NK-3, and NK-4 family of genes are closely related so that these genes were originally classified all within the NK-2 class. Thus, the vertebrate nomenclature that uses Nkx-2.x for genes that fall into separate families is confusing.

2.4.2 Distal-less (Dll) Family

Six *Dlx* genes are found in mammals, while only one fly and one nematode gene is known. The *Drosophila* gene has been shown to play an important role in limb development. In vertebrates, the *Dlx* genes are expressed in various areas of the brain, but intriguingly, some are also expressed in the limbs; thus, it is thought that *Dlx* genes play an important role in the formation of appendages.

2.4.3 Empty-spiracles (ems) Family

The genes of the ems family, though not part of any cluster, have a Hexapeptide upstream of the homeodomain. These genes function in pattern formation mainly in the head region. There are at least two genes in mammals and two in flies, which appear to have arisen from independent duplication events. In *C. elegans*, the ems family gene *ceh-2* plays a role in neuron specification in the anterior pharynx.

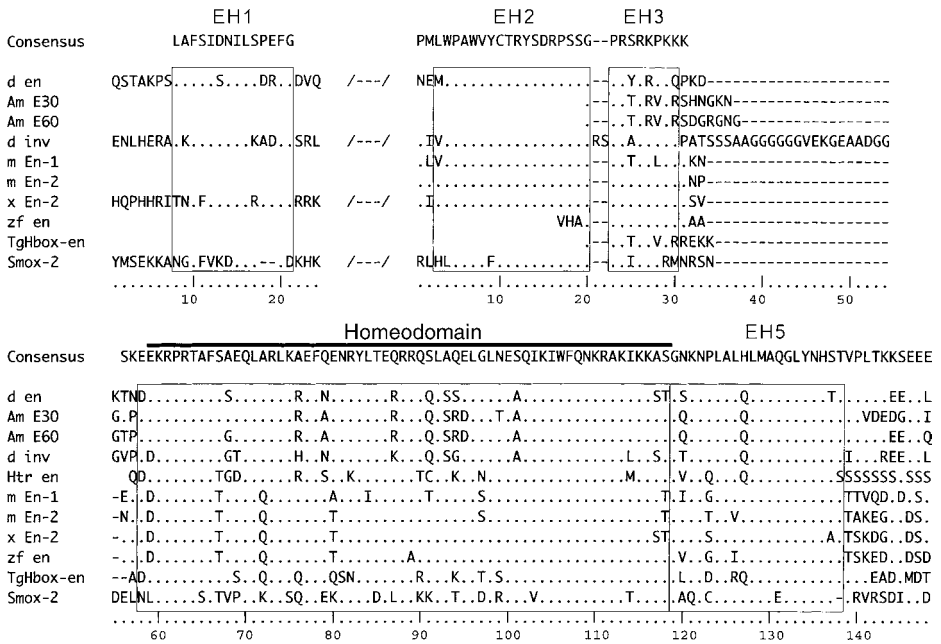

Fig. 11 Alignment of engrailed family sequences. The black bar denotes the homeodomain and the boxes indicate the domains that have been conserved between deuterostomes and protostomes. The slashes indicate a long nonconserved region of variable length. Two conserved regions can be found upstream of the homeodomain (EH1, and EH2/EH3) and a third region of conservation (EH5) is just downstream of the homeodomain, followed by a charged region.

2.4.4 **Engrailed (en) Family**

The engrailed family of genes is characterized by a series of small conserved motifs outside of the homeodomain (Fig. 11). These motifs were named EH1, EH2, EH3, EH4 (the homeodomain), and EH5. EH1 is a small conserved region (~13 amino acids) toward amino terminus of en family genes, EH2 and EH3 are adjacent conserved regions of about 20 and 8 amino acids just upstream of the homeodomain. EH2 contains residues that are somewhat reminiscent of the Hexapeptide. EH5 is a carboxy-terminal extension of the homeodomain of about 20 amino acids that is followed by a charged region. The current evidence suggests that one ancestral *engrailed* gene existed in the ancestor of vertebrates and arthropods and that, later, independent duplication events gave rise to *en* and *invected* (*inv*) in flies and *En-1* and *En-2* in vertebrates. The nematode *C. elegans* has only one en family gene, *ceh-16*.

2.4.5 **Hlx Family**

Genes similar to the *Drosophila* gene *H*2.0 have been isolated from vertebrates and ascidians. Chicken *Chlx-A* and mouse *Dbx* genes as well as the *Ascidia* gene Hal-*hlx* are quite different from *H*2.0 and the mouse *Hlx* gene; thus, this family seems to have evolved divergent members.

2.4.6 **Hex Family**

The first identified member was in mammals (Prh). This gene family is rather

distinct from Extended HOX cluster genes or NKL genes, although it is highly conserved down to cnidaria, suggesting an important role in animal development.

2.4.7 Other Families

A substantial number of additional families that have been conserved from protostomes to deuterostomes have been identified. Most of these families are similar to the NK cluster genes and can be loosely grouped into the NKL class of homeobox genes. These families include Bar, bsh, ceh19, Gbx, Gtx, Xnot, Mnx, NEC, and Vent, although the last family does not seem to be present in flies. The rough (ro) family is more similar to the ExHox class of genes. Other genes, such as the *C. elegans* genes *ceh-7* or *ceh-5*, are rather divergent from any of the other genes and no obvious homologs can be detected presently, suggesting that these genes diverged in the nematode lineage and are derived from some other homeobox gene. Similarly, other newly derived genes, as also the example of *bcd* shows, will exist only within particular animal phyla.

2.5 Antennapedia Superclass Related Genes in Lower Animals

Quite a number of homeobox genes have now been isolated from cnidaria, such as hydra or corals as well as sponges. Some are most closely related to the genes of the HOX, ParaHox, and NK cluster, as mentioned above. Others are clear orthologues of other Antp superclass genes, such as for example, en. An Antp-like gene has been isolated from coral, and this gene has been shown to be linked to an eve class gene. Thus, a primitive cluster does exist, but it may have dispersed more in lower organisms. Antp superclass homeobox genes have also been discovered in what is thought to be the most primitive animal, *Trichoplax adhaerens* (Placozoa). Once full genome projects are available, it will be possible to determine the full complement of homeobox genes and elucidate the evolutionary history of these genes.

3 Paired and prd-like Classes

3.1 Paired (prd) Class

The paired class of homeobox genes is characterized by the fact that their homeodomains contain a serine residue at position 50. In addition, a highly conserved domain of about 130 amino acids, the "paired" domain, is found toward the amino terminus of these proteins. prd class genes have been shown to be transcription factors, and the prd domain by itself has been shown to bind DNA. The structure has recently been determined and shown to consist of two globular domains with three α-helices each. Several genes have been found that encode only a prd domain but no homeodomain. In contrast, thus far, all homeodomain proteins of the prd class that contain a serine residue at position 50 also contain a prd domain. The prd domain bears resemblance to transposases, and it is thought that a prd-like homeobox gene captured a transposase early in metazoan evolution, which then evolved into the prd domain. In addition to the prd domain, the prd class genes show some extended sequence conservation just upstream of the homeodomain, and some families have an "Octapeptide" between the prd domain and the homeodomain (Fig. 4). Aligned

```
                                   ▭▭▭
```

prd domains are shown in Fig. 12. The prd class genes can be grouped into six families on the basis of this domain. In vertebrates, prd class genes are called *Pax genes*, irrespective of whether they have a homeodomain or not.

3.1.1 Family I

The genes of this family, formed by mouse *Pax-1* and *Drosophila Pox meso*, do not encode homeodomains. A mutation in mouse *Pax-1* is responsible for the *undulated* defect.

3.1.2 Family II

This family includes the *Drosophila* genes *prd, gsb-d, gsb-p* and the mouse genes *Pax-3* and *Pax-7*. Apparently, independent duplication events gave rise to the multiple genes in vertebrates and flies. This family contains a small conserved region (the Octapeptide) between the homeodomain and the prd domain. In humans, a defect in *Pax-3* causes phenotypes known as the *Waardenburg syndrome* and *craniofacial-deafness-hand syndrome*. In mouse, a mutation is known as *Splotch*.

3.1.3 Family III

Presently, this family is defined only by vertebrate genes (*Pax-2, Pax-5, Pax-8*) and may be vertebrate specific. Interestingly, these genes encode a protein piece that resembles the first third of a homeodomain. This suggests that the last two thirds of the homeodomain was only recently lost in evolution. Like in family II, these genes also encode an Octapeptide between the prd domain and the homeodomain remnant. A mutation in the human *Pax-2* gene causes kidney and retinal problems.

3.1.4 Family IV

This family is defined solely by *Pax-4*. This family may be derived from *Pax-6* genes.

3.1.5 Family V

This family is defined by *Pox neuro*, which has no homeodomain.

3.1.6 Family VI

This family is well conserved in evolution, encompassing such genes as *Pax-6* in vertebrates and *eyeless* (*ey*) in flies, and they encode both a prd domain and a homeodomain. *Pax-6* is mutated in the *Small eye* mutations in mice and rats, and in the *Aniridia* mutation in humans; thus, *Pax-6* plays a crucial role in eye development. The fly gene *ey* was shown to be homologous to *Pax-6*. Thus, these genes play important roles for eye development both in vertebrates and flies, despite the large differences in eye structure and development between these phyla. This suggests that there is a common underlying mechanism for "eye development" that has been conserved in evolution and that vertebrate eyes and insect eyes are not completely independent inventions of evolution.

3.2 Paired-like (prd-like) Class

A substantial number of homeobox genes encode a homeodomain that is similar to the prd class homeodomains but they do not encode a paired domain nor do the homeodomain sequences have a serine residue at position 50. Some of these prd-like homeodomain sequences can be as much as 75% identical to prd class homeodomain sequences. Given that the prd class genes are thought to be derived from prd-like homeobox genes, this similarity is not surprising. In recent years an increasing number of conserved prd-like families has been identified. The following families have been conserved between different phyla:

otd/Otx (fly orthodenticle), ceh10, unc4, Arix/Phox, Ptx/unc30, al (fly aristaless), otp (orthopedia), Gsc (goosecoid). Several other families seem, at present, to only occur in vertebrates, suggesting a possible vertebrate-specific expansion. While most of the prd-like genes encode glutamine residues at position 50, several families (for example, *goosecoid, otd*) encode lysine residues at that position. A feature often found in prd-like class genes is the presence of an intron between amino acids 46 and 47 of the homeodomain. Some of the families have quite extensive sequence conservation outside of the homeodomain. For example, the ceh10 family has a motif of about 70 amino acids immediately downstream of the homeodomain that is conserved from *C. elegans* to vertebrates. Many of the prd-like class genes play important roles in the developing nervous system.

4 POU Class and HNF Classes

4.1 POU Class

The POU class was originally defined on the basis of the four genes *Pit-1, Oct-1, Oct-2,* and *unc-86* (POU). The POU-specific domain is an approximately 80-amino acid long conserved domain upstream of the homeodomain with a variable linker in between (Fig. 13). Several of the POU class genes were originally isolated as transcription factors, which provided the most convincing evidence that homeobox-containing genes encode transcription factors. The POU-specific domain is required for cooperative, high affinity DNA-binding and has so far always been found in association with a POU homeodomain. The POU homeodomain is characterized by a cysteine residue in position 50. The structure of the POU domain, that is, the POU-specific domain and the POU homeodomain, bound to DNA has been determined. The POU-specific domain consists of four α-helices (indicated in Fig. 13). Helix 2 and 3 fold like a helix-turn-helix motif, although the turn between Helix 2 and 3 in the POU-specific domain is larger than that in a standard helix-turn-helix protein (such as the homeodomain). The genes have been grouped into several families and many play important roles in nervous system development.

4.1.1 Family I

Represented by the gene *Pit-1* and only found in vertebrates so far.

4.1.2 Family II

Encompasses the well-known mammalian transcription factors *Oct-1* and *Oct-2*, and at least one additional member in vertebrates, *Oct-11*. The *Drosophila* genes of this family, *nubbin* (*pdm-1*) and *pdm-2* have arisen through independent duplication events in arthropods. A divergent member of this family is the *C. elegans* gene *ceh-18*.

4.1.3 Family III

Comprises at least four vertebrate members, that is, *Brn-1, Brn-2, SCIP/Oct-6*, and *Brn-4*, one in *Drosophila* (*vvl/Cf1a*), one in *C. elegans* (*ceh-6*), as well as two genes in flatworms.

4.1.4 Family IV

This family is highly conserved between vertebrates, flies, and nematodes. The *C. elegans* gene, *unc-86*, is expressed and functions in many neurons of the sensory nervous system.

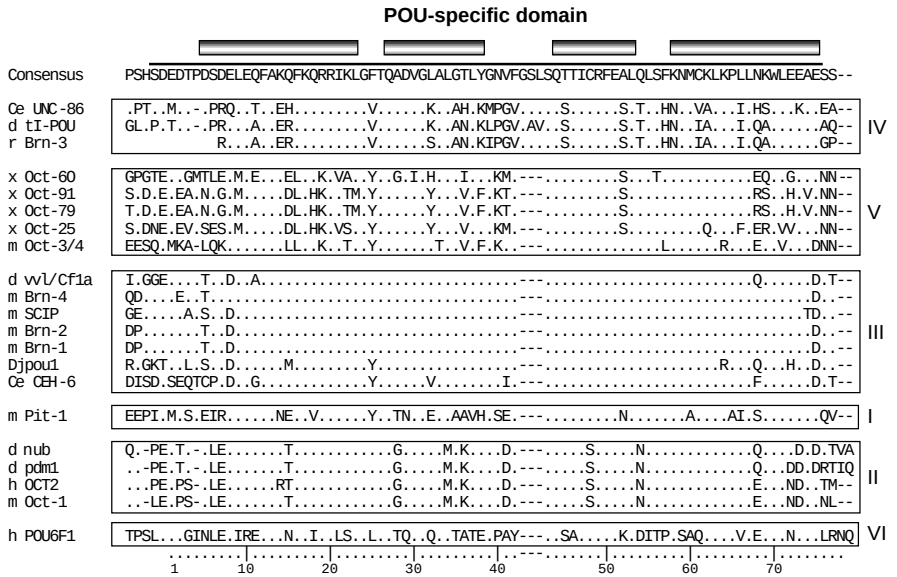

Fig. 13 Alignment of sequences for the POU-specific domain (marked by the black bar at the top). Examples of *POU* genes are given and grouped into different families (I–VI). The four shaded cylinders indicate the four α-helices of the secondary structure of the POU-specific domain. Helix 2 and 3 form a helix-turn-helix motif.

4.1.5 **Family V**

This group is a diverse assortment of genes, some of which can be quite divergent from each other. This family seems to be restricted to vertebrates.

4.1.6 **Family VI**

This family is conserved between flies and vertebrates but has been lost in *C. elegans*.

4.1.7 **Orphans**

Several additional POU genes in vertebrates exist (e.g. *Sprm-1*) that are very different from the above families and may represent novel derived genes in vertebrates.

4.2 **HNF Class**

The HNF class was originally defined by the mammalian transcription factor LFB1 or HNF1α. Figure 14 shows the homeodomain of HNF1α and the extra residues that are found between helix 2 and helix 3. The HNF class genes have been found to have a conserved domain upstream of the homeodomain (not shown). The structure of this domain has been determined and found to be similar to that of the POU domain, suggesting that the HNF class might be highly divergent and be derived from *POU* genes.

5 **Cut Superclass**

The first gene of this superclass discovered was the *Drosophila* homeobox gene *cut*, which has three copies of a conserved domain of about 80 amino acids, the cut domain (Fig. 15), upstream of the

```
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
```

```
r HNF1α/LFB1     GRRNRFKWGPASQQILFQAYERQKNPSKEERETLVEECNRAECIQRGVSPSQAQGLGSNLVTEVRVYNWFANRRKEEAFRH
d pros           TPLHSSTLTPMHLRKAKLMFFWVRYPSSAVLKMYFPDIKFNK------------------NNTAQLVKWFSNFREFYYIQM
d dve HD1        TRMRTSFDPEMELPKLQKWFADNPHPSRQQIQTYVVQLNALESRRGRKP-----------LDVNNVVYWFKNARAAQKRAE
d dve HD2        KRNRTFIDPVTEVPKLEQWFAMNTHPSHNLILKYTEDLNTMPYRQKFPR-----------LESKNVQFWFKNRRAKCKRLK

d Antp           RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALC---------------------LTERQIKIWFQNRRMKWKKEN
                 ....|....|....|....|....|....|....|....                       |....|....|....|....|
```

Fig. 14 Alignment of the atypical homeodomain sequences of rat HNF1α/LFB1, *D. prospero*, and the two homeodomains – marked by HD1 and HD2 – of *Drosophila* defective proventriculus (dve). Symbols for conserved positions and cylinders for helices are as in Fig. 1. Dashes indicate gaps. The typical homeodomain of Antp is shown at the bottom for comparison.

```
                                                Cut domain
Consensus     VGMYEELDTKEIARRVKEELKRNNIGQRIF.EKVLGLSQGSVSDLLRRPKPWSKLTQKGREPFRRMQLWLSDPE..RMLALRQA.EKRAE..R

Ce CEH-44_R1  ALNGHP.N.T...SHC.RLMIAY.....L.AKH.MNQ....L.E..SK.RH.N...D....A...IYG.I..D.AINL.CSLSPRRVWPADQN
m Cux-2_R1    GPEE.Q...A...FQ...Q.LKH.....V.GHY..........EI.A.....R...V..K...IK.KQF...EQNVLA.RTI.VRQRGSITP.
m Cux_R1      ISEG..I..A....Q...Q.IKH.......GHY..........EI.A.....N...VR.K...HK.KQF...EQNILA.RSI.GRQRENPGQS
h CDP_R1      .SEG..M..A....Q...Q.IKH.......GHY..........EI.A.....N...VR.K...HK.KQF...EQNILA.RSI.GRQRENPGQS
d CUT_R1      KYANQA...LH.S...R.L.SVH.....L.AKYI......T..E..SK....D...E...DSY.K.HA.AC.DNAVML.KSLIPKKDSGLPQY
Ce CEH-44_R2  YDA.TV...DFLVKQI..F.TM.S.S..Q.G.YI............A...T.AQ.........I....FMD.V.ASEENDEK.PKITICDEDS
m Cux-2_R2    LY..R.V..L.LT.Q...K.AK.G.C....G............M.S...............I........QLGQGQGQAPTQQPSASQASP
m Cux_R2      .Y..Q.V..I.LT.Q...K.AK.G.C....G............M.S...............I......NGELGQGV.PVQGQQQQGPVLHSV  CUX
h CDP_R2      .Y..Q.V..I.LT.Q...K.AK.G.C....G............M.S...............I......NGELGQGV.PVQGQQQQGPVLHSV
d cut_R2      FD.FNN.N.ED.V.....A.SQYS.S..L.G.S.............A.....HM.........I..KMF.E.ENAVHK.VAS.YKIAPEKLM.
Ce CEH-44_R3  SSGE.I...F..VYQ..GI.EE.G.SP.V.GDEY.HCTSSMCA..MI.T.SFENSKASEKLMYT..KTF....IAIPL.VEKEESKETVKAKI
m Cux-2_R3    .A.SP....YS.TK....V.TD..L...L.G.SI...T........S.....H..SL-.....V.........HNVEK.RDMKKL..K.YLK.
m Cux_R3      .A.SP....YG.TK....V.TD..L...L.G.TI...T........A.....H..SL-.....V......N..NNVEK.MDMKRM..K.YMK.
h CDP_R3      .A.SP....YG.TK....V.TD..L...L.G.TI...T........A.....H..SL-.....V......N..NNVEK.MDMKRM..K.YMK.
d cut_R3      AALTQD...HD.TTKI..A.LA.....K..G.A..........E..SK....HM.SI......I........ANNVER.Q.LKNERRE.SKR.
r HNF6        S.QM..IN...V.Q.ITT....YS.P.A..AQR..CR...TL.....N.......K-S...T....WK..QE..FQ..S...L.AC..K.QEH
Ce CEH-39     I.DD.....QL..H.ILD...EQC.P.ASLAV....R...TL.....K....GIMK-N..GT.Q..AN..DLDPVV.RALCFMKK.DV.RITG
Ce CEH-21     I.DD.....VD....ILS...ERC.P.TALA..I.AR...TL.....M.....VMK-N..AT.Q..SN..GLDPDV.RALCFLPK.DV.RITG
Ce F22D3.1    I.DDIYI...DLCK.IAF...NHS.P.A..A.RI.CR...TL.....N....N..K-S...T....YN.VAQ.LAT.LAI.DMKT.DVNRASG  ONE CUT
d ONECUT      SSDM..IN..DL.Q.ISA....YS.P.A..AQR..CR...TL.....N.......K-S...T....YK..QE..FQ..S...M.ACR.K.EPQ
Ce C17H12.9   FDDG...N...L.LQIAA....YS.P.A..A.R..CR...TL.....N.......K-S...T....AK..EE..FQ..S...L.AC..K.EHT
h HNF6        S.HM..IN...V.Q.ITT....YS.P.A..AQ...CR...TL.....N.......K-S...T....WK..QE..FQ..S...L.AC..K.QEH
h OC2         S.QL..IN...V.Q.ITA....YS.P.A..AQR..CR...TL.....N.......K-S...T....WK..QE..FQ..S...L.AC..K.QEP
m SATB1_R1    QVSTNTEVSS..YQW.RD....AG.S.AV.ARVAFNRT..LL.EI..KEEDPKTAS.SLLVNL.A..NF.QL..AE.DRIYQDER.RSLNAAS
h SATB1_R1    QVSTNTEVSS..YQW.RD....AG.S.AV.ARVAFNRT..LL.EI..KEEDPKTAS.SLLVNL.A..NF.QL..AE.DRIYQDER.RSLNAAS  SATB
m SATB1_R2    PENNTMNINAS.YDEIQQ.M..AKVS.AP.AKVAATK...WLCE...WKEDP.PENRTLW.NLSMIRRF..L.QPE.DAIYE.ESNAVHHHGD
h SATB1_R2    PENNTMNINAS.YDEIQQ.M..AKVS.AL.AKVAATK...WLCE...WKEDP.PENRTLW.NLSMIRRF..L.QPE.DAIYE.ESNAVHHHGD
              .........|.........|.........|.........|.........|.........|.........|.........|.........|...
                       10        20        30        40        50        60        70        80        90
```

Fig. 15 Alignment of cut domain sequences. The genes of the CUX class encode three cut domains, marked by R1, R2, and R3, the genes of the ONECUT class encode only a single cut domain, and the SATB class genes encode two highly divergent cut domains (R1, R2). The black bar indicates the cut domain.

homeodomain. The cut domain is a DNA-binding domain. Other genes with cut domains have been discovered; however, their structural organization is quite different from that of *cut* so that four distinct classes can be defined (Figs. 4, 5, 15). The homeodomain sequences of these classes are rather different from each other, possibly because the presence of additional DNA-binding domains relieves the evolutionary pressure to maintain the homeodomain.

5.1 CUX Class

This class is defined by the fly gene *cut*, *C. elegans ceh-44*, the human transcription factor CDP, and other vertebrate homologs, which have been named *Cux* genes. These genes encode three cut domains upstream of the homeodomain. Unlike other homeodomain proteins, they have a histidine residue at position 50 of the homeodomain. Biochemical analysis

of the human CDP has revealed that the cut domain is a DNA-binding domain, like the prd domain and the POU-specific domain. The CUX class genes have a most unusual structural organization because, apart from the cut domain, their amino terminus is actually shared with another gene, CASP, which is a Golgi membrane protein. The amino-terminal half of CASP can either splice to the cut-homeodomain part of the Cux genes or it can splice to the carboxy-terminal part of the CASP protein, giving rise to a fully functional CASP protein. Sometime in evolution, an ancestral Cux gene has been functionally intertwined with the *CASP* gene through alternative splicing. This organization is found in *C. elegans* and vertebrates, but in *Drosophila*, the *CASP* gene has been lost.

5.2
ONECUT Class

The class is defined by a single cut domain. Most likely, this represents the most ancestral condition. Genes in this class comprise a diverse set of genes from *C. elegans*, for example, *ceh-21* and *ceh-39*, vertebrate HNF6 and OC2, and a single gene in *Drosophila*. In this gene class, position 48 of the homeodomain exchanges the tryptophan into a phenylalanine, or into a tyrosine in one divergent gene in *C. elegans*. Further, at the DNA-binding position 50, a methionine is found.

5.3
SATB Class

The genes of the SATB class encode two highly divergent cut domains and a highly divergent homeodomain. In addition, they have a new domain, called *COMPASS*, at their amino terminus (Fig. 4). Like the ONECUT genes, they encode a phenylalanine residue at position 48 of the homeodomain. SATB class genes have presently been found only in vertebrates. In contrast to other homeobox genes that act as regular transcription factors, SATB1 has been shown to be a special global gene regulator that is involved in chromatin remodeling.

5.4
COMPASS (CMP) Class

The CMP class of homeobox genes is an unusual group of homeobox genes that encode a CMP domain upstream of two homeodomains. The two homeodomains arose through duplication from a common ancestor. These homeodomains are distinct from those of other classes because of the extra residues in the loop region between helix 2 and helix 3 (Fig. 14). Members of the CMP class have been found in *C. elegans* and in *Drosophila* (the gene *dve*), but not in vertebrates. Even though the *CMP* genes do not encode cut domains and their homeodomains are highly different from other cut homeodomains, they are classified into the cut superclass because they share the CMP domain with the SATB class of genes. Since the SATB class is only found in vertebrates and the CMP class only in insects and nematodes so far, it may be possible that these two classes are orthologous.

6
ZF and LIM Classes

These two classes share the common trait that their genes encode zinc-binding domains in addition to the homeodomain. Each class encodes distinct types of zinc-binding domains: the ZF class encodes zinc fingers of the C_2-H_2 (two cysteine

residues–two histidine residues) type, and the LIM class, two copies of the LIM domain zinc finger. Some of the homeodomains of the ZF class proteins are more similar to LIM class homeodomains than to any other homeodomain. Thus, these two homeodomain classes may have arisen from a common single ancestor, but this is presently unresolved.

6.1 ZF Class

This is without question the most bizarre class of homeobox genes. While one *Drosophila* gene *zfh-1* encodes 9 zinc-finger domains and one homeodomain, another gene *zfh-2* has 16 zinc fingers and 3 homeodomains. The human homolog of *zfh-2*, *ATBF1* encodes 17 zinc fingers and 4 homeodomains. The zinc-finger domains are of the C_2-H_2 type, the same type that is also found in the developmental control genes *Krüppel* and *hunchback* in *Drosophila*. At least two families exist, defined by Drosophila *zfh-1* and *zfh-2*, that are conserved in vertebrates. The homeodomains of these proteins are very divergent (Fig. 5); presumably, the evolutionary constraints are relaxed because of the large number of DNA-binding domains present in a single protein.

6.2 LIM Class

LIM homeobox genes encode two LIM domains upstream of the homeodomain. The LIM domain is about 60 amino acids long and contains conserved cysteine and histidine residues (Fig. 16). It was first defined in two *C. elegans* genes, *lin-11* and *mec-3*, and in the rat transcription factor *Isl-1* (*lin-11*, *Isl-1*, *mec* − 3 = LIM). The LIM domain exhibits relatively poor primary sequence conservation between different proteins, and the first and the second copy are somewhat distinct from each other. The key residues are the conserved cysteine and histidine residues. The LIM domain has been shown to be a zinc-binding domain, a so-called *zinc finger*. The cysteine residues (positions A, B, and D) and the histidine residue (position C, Fig. 16) coordinate one zinc, while the second zinc is coordinated by three cysteine residues (position E, F, and G) and one residue that can be either a cysteine, histidine, or aspartate (position H). This double zinc finger is distinct from other zinc fingers, such as for example, the ones of the ZF class. The LIM class homeodomain sequences are much better conserved than the LIM domains and can thus be used for classification. Six families have been described: apterous (ap), LHX6/7, ISLET, LMX, LIM3, and LIN-11/LIM1.

The LIM domain is also present in genes that do not encode a homeodomain. The rhombotin genes (*LMO*, i.e. LIM-only) encode two LIM domains that are similar to the two LIM domains of LIM homeobox genes. Possibly, this gene family may have lost the homeodomain. A large number of other proteins containing only LIM domains have been found that are more divergent. The following are a few examples: mammalian CRIP, a cysteine-rich intestinal protein; ESP1, an estradiol-stimulated protein in brain; hCRP, a human cysteine-rich protein; zyxin, a cytoskeletal protein; and MLP, a regulator of myogenesis. hCRP and MLP, which are very similar to each other, have two copies of LIM that are more closely related to each other than to other LIM domains. CRIP and ESP1 have only one copy, and zyxin contains three LIM domains.

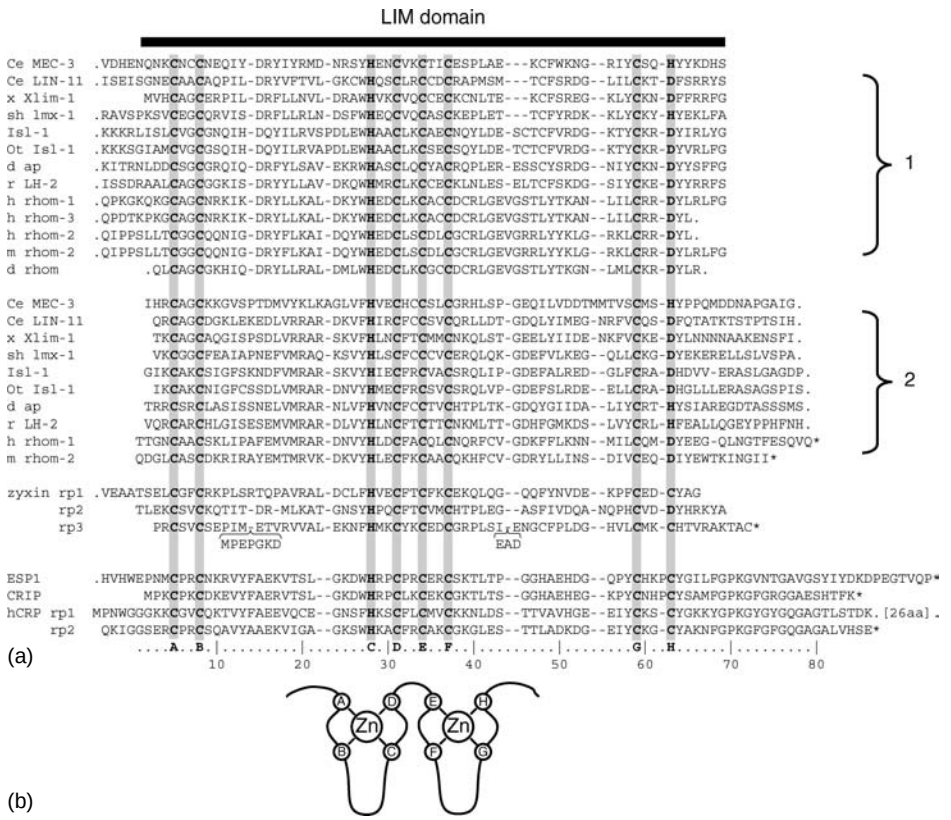

Fig. 16 Alignment of LIM domain sequences. LIM homeobox genes encode two LIM domains, the two parentheses on the right indicate the first and the second LIM domain, respectively. Several proteins, whose LIM domains are somewhat distinct and that have no homeodomain (i.e. zyxin, ESP1, CRIP, hCRP), are listed at the bottom. The conserved cysteine and histidine residues, which are the key residues for zinc binding are highlighted. At the bottom is a schematic view of the structure of the two zinc fingers; residues A to H are also indicated underneath the alignment.

The LIM domain is protein–protein interaction domain. For example, biochemical characterization of zyxin indicates that it is a zinc-binding protein that binds to α-actinin and is associated with the actin cytoskeleton. In addition, zyxin directly interacts with the chicken homolog of hCRP *in vitro*. The LIM domains of LIM homeobox genes also function in protein–protein interactions. One important interacting partner in that case is the LIM-binding protein Ldb. Thus, while the LIM domain can occur in multiple functional roles, the common denominator of these functions is protein–protein interaction.

7 Six/so Class

This class was originally defined by the *Drosophila* gene *so* and the mouse

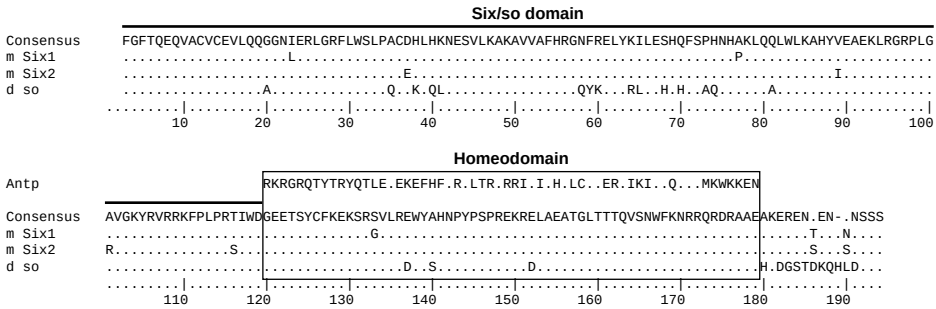

Fig. 17 Sequence alignment of Six/so class genes over the region of the Six/so domain and the homeodomain. The Six/so domain is marked by a black bar. The Six/so homeodomains are compared to the Antp homeodomain.

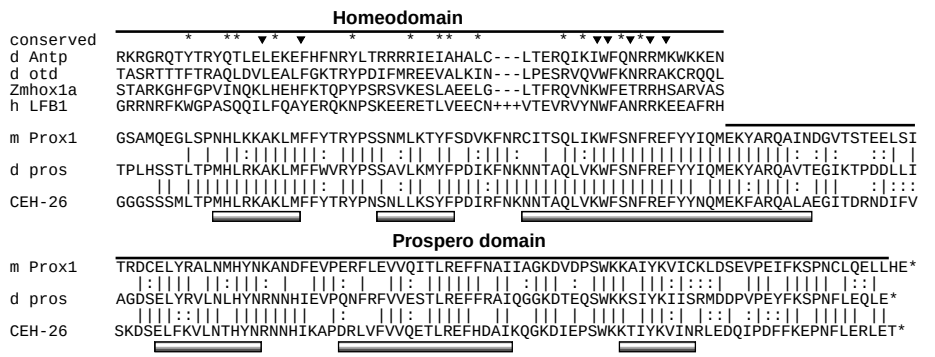

Fig. 18 Alignment of prospero class sequences. The lines between sequences indicate identities and colons mark similar amino acids. The "‡" marks matches of the pros class homeodomains to the highly conserved residues of typical homeodomains, which are shown at the top. The homeodomain as well as the pros domain are marked by black bars. The cylinders underneath the sequence show the α-helices derived from the structure of the Homeo-Pros domain.

genes *Six1* and *Six2*. These homeobox genes encode highly divergent typical homeodomains (Figs. 5, 17) that have a lysine at position 50 of the homeodomain. In addition, there is a highly conserved domain upstream of the homeodomain (Figs. 4, 17), the Six/so domain, which is about 120 amino acids long. The Six/so class of genes can be subdivided into three families, Six1/2, Six3/6, and Six4/5, which are all conserved between flies, worms, and vertebrates.

8 Prospero (Pros) Class

This class of atypical homeobox genes has been defined on the basis of the *Drosophila* gene *prospero*, the mouse gene *Prox 1*, and a *C. elegans ceh-26* (Fig. 18). It is a small class with a single member in fly and *C. elegans*. The pros class homeodomain is highly divergent and has three extra amino acids between helix 2 and helix 3. The highly conserved leucine in helix 1 is replaced with an alanine (Figs. 1,

18). Downstream of the homeodomain is a conserved sequence motif of about 100 amino acids, the pros domain, that reaches to the carboxy-terminus. The structure of the pros homeodomain and pros domain has been determined and confirmed to be a homeodomain. The pros domain forms a single structural unit with the homeodomain and is required for sequence-specific DNA binding.

9 TALE Superclass

As mentioned above, the TALE homeobox genes are characterized by having three extra residues in the loop between helix 1 and helix 2 of the homeodomain (three amino acid loop extension, Fig. 6). This group is ancient since members are found in plants, fungi, and animals. Four classes are found in animals, two in plants, and two distinct types in fungi. Many of the TALE superclass homeobox genes encode an isoleucine at position 50 of the homeobox.

9.1 PBC Class

Extensive sequence conservation outside the homeodomain has been found between the three human homeodomain proteins PBX1 (a protooncogene), PBX2, PBX3, the *C. elegans* homeodomain protein CEH-20, and the *Drosophila* protein extradenticle (exd). This new motif has been termed the *PBC domain* (Fig. 19). The PBC domain is bipartite (PBC-A and PBC-B), with a linker region between the two subregions, and about 180 amino acids long. The extended loops between helix 1 and helix 2 of the homeodomain of PBC proteins interacts with the Hexapeptide of Antp superclass homeobox genes, such as those of the lab family. Thus, PBC class genes form complexes with at least some of the HOX cluster genes. The homeodomain of PBC class genes contains a glycine residue at position 50 of the homeodomain, which is unusual as this residue can contact the DNA only via water-mediated interaction.

9.2 MEIS Class

The MEIS class of homeobox genes was first defined by the vertebrate *Meis* genes, *C. elegans unc-62 (ceh-25)*, and *Drosophila homothorax* (*hth*). As in the PBC class, a conserved, bipartite domain is found upstream of the homeodomain, called *MEIS* or *HM* (Fig. 20), which is about 130 amino acids long. Two subgroups can be

```
                                                
```

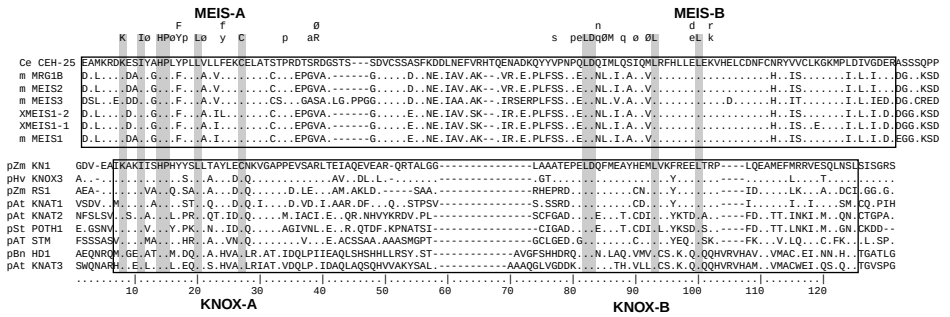

Fig. 20 Alignment of MEIS and KNOX domain sequences. The upper half shows the boxed MEIS domain, which is bipartite, consisting of MEIS-A and MEIS-B. The lower half shows the boxed KNOX domain, which is also bipartite, consisting of KNOX-A and KNOX-B. The gap in the linker region between KNOX-A and -B was increased to obtain alignment with the MEIS domain. Those residues that are absolutely conserved between MEIS and KNOX are shaded, other less-conserved positions are indicated on top of the MEIS domain; Ø represents hydrophobic residues.

defined, one based on the vertebrate *Meis* genes, and one based on the *Prep* genes. The MEIS-A region of the Prep proteins is shorter compared to Meis proteins (not shown). Fly hth and vertebrate Meis proteins interact with PBC class homeo domain proteins through their MEIS domain, which interacts with the PBC-A region. Thus, MEIS and PBC class proteins can form heterodimers. It has been shown that the interaction of hth with exd is responsible for translocating exd from the cytoplasm to the nucleus so that dimerization controls the activity of the protein complex. Further, MEIS class proteins can also interact with HOX cluster proteins that are expressed posteriorly (e.g. HoxA-10). Triple complexes of PBC, MEIS, and Hox proteins have also been reported. These combinatorial possibilities provide ample opportunities to generate many different types of DNA-binding specificity, and it is likely that the interaction between TALE homeobox genes and HOX cluster genes has been an important factor in the evolution of animal pattern formation.

9.3 IRO Class

The IRO class of TALE homeobox genes are defined by the *Drosophila* genes *araucan, mirror,* and *caupolican*. These genes are arranged in tandem in a gene cluster, the iroquois complex. Also, in mammals, there are two paralogous Irx gene clusters with three genes each. However, it is likely that the clusters evolved independently in flies and vertebrates and that only a single common ancestral *IRO* gene existed. The homeodomain of the IRO class genes contains an alanine residue at position 50 of the homeodomain (Fig. 6) and thus may have a DNA-binding activity distinct from other TALE homeobox genes. The IRO class of TALE homeobox genes share a small motif of about 12 amino acids, the IRO box, downstream of the homeodomain (not shown).

9.4 TGIF Class

The TGIF class of TALE homeodomains has been first defined by the vertebrate

TG-interacting factor (TGIF), but an orthologue has also been identified in *Drosophila*. Fly and vertebrate *TGIF* genes share an additional 20 amino acids immediately downstream of the homeodomain (not shown), *C. elegans* appears to have lost the *TGIF* gene. *TGIF* genes may have been derived from a MEIS class gene during the evolution of early animals.

10 Evolution of Homeobox Genes

10.1 Fungal Homeobox Genes

Seven homeobox genes are known in the yeast *Saccharomyces cerevisiae*. Two of these genes, *MATa1* and *MATα2* are part of the mating type locus (MAT); *MATa1* encodes a typical homeodomain and *MATα2* a TALE homeodomain. Such a dyad of a typical and TALE homeobox gene is also found in other fungi, such as *Ustilago maydis*, *Schizophyllum commune*, and *Coprinus cinereus*. That these genes are indeed evolutionarily related is further supported by the fact that the typical homeobox genes *MATa1*, Um *bW*, and Sc *AαY* have an intron in the same position in the homeobox. Also, in Microsporidia, a closely linked pair of homeobox genes has been found, one a typical and one a TALE homeobox gene. While the function of this locus is not yet known, it does show that a locus with a typical and a TALE homeobox gene is an ancestral feature in the fungal lineage. In yeast, MATa1 and MATα2 form heterodimers, and MATα2 also forms homodimers, allowing regulation of different sets of target genes. Classification of the fungal homeobox genes, apart from the distinction of typical and TALE homeobox genes, is not readily possible due to their divergence, and no additional conserved domains are encoded by these homeobox genes that would aid in classification. Phylogenetic analyses of yeast and Microsporidia homeodomain sequences suggests that within each fungal branch, duplications of homeobox genes have occurred independently. In summary, it can be estimated that there were not more than two TALE homeobox genes and two to three typical homeobox genes present in the first fungal ancestors.

10.2 Plant Homeobox Genes

In plants, 14 distinct classes of homeobox genes that have been conserved between monocots and dicots can be identified. One large group consists of homeobox genes that have a leucine-zipper, a protein interaction motif, downstream of the homeodomain. These genes can be grouped into four classes, HD-ZIP I, II, III, and IV. Samples of homeodomain sequences for two classes are shown in Fig. 5. Genes of HD-ZIP III and IV encode a START domain, a lipid-binding domain, downstream of the homeodomain. The START domain is also found in animals; however, there it is not found in homeobox genes. The homeodomain of the HD-ZIP III genes has undergone a drastic change, since in that class extra residues have been inserted between helix 2 and helix 3 of the homeodomain (not shown). Genes of the other classes also encode conserved domains outside the homeodomain. For example, the PHD class genes encode a PHD domain upstream of the homeodomain. The PHD domain is a zinc-finger motif distinct from the LIM and ZF class

zinc fingers. The DDT class genes encode a DDT domain downstream of the homeodomain. Both PHD and DDT domain are also found in animals – in fact, these two domains can occur in the same protein – but not in homeodomain proteins. Plants have two ancient classes of TALE homeobox genes, the KNOX class and the BEL class. Both encode large bipartite domains upstream of the homeodomain, termed *KNOX* and *BEL domain*, respectively. The KNOX domain (Fig. 20) can be aligned with the MEIS domain, showing that these two gene classes in plants and animals have arisen from a common ancestral TALE homeobox gene that also encoded, what we termed, the *MEINOX domain*. The BEL domain (not shown) has some weak similarities to the MEIS domain and may be derived from it. BEL and MEIS proteins have been shown to interact and this interaction is mediated through their BEL and MEIS domains, respectively.

10.3 Origin and Diversification of Homeobox Genes

The different classes of homeobox genes found in plants and animals have arisen mainly independently of each other. The only group of genes that can be clearly traced from a common ancestor are the TALE homeobox genes, in particular, the MEIS and KNOX class of homeobox genes. The PBC domain in animals has also been shown to have similarity to the MEINOX domain, and the BEL domain in plants may have similarities as well. This suggests that there was at least one MEINOX-TALE homeobox gene in the common ancestor organism (CAO) that gave rise to plants and animals. Both plants and fungi each have two rather distinct classes of TALE homeobox genes, that is, KNOX and BEL in plants and CUP and M-ATYP in yeast (Fig. 6), and animals have four, although the two main ones are PBC and MEIS. Thus, the possibility remains open that there were two TALE homeobox genes in the CAO. For typical homeobox genes, the situation is less clear. Again, there probably must have been at least one in the CAO. Since, in fungi, perhaps not more than three different types of typical homeobox genes exist, there should not have been more than three in the CAO either. Thus, the CAO must have had at least two, but not more than five homeobox genes. A single "Urhomeobox" gene must have given rise to the typical and TALE homeobox genes, but at present, we do not know whence it came from. The homeodomain is structurally related to bacterial helix-turn-helix proteins, so in some protozoa, a helix-turn-helix protein seems to have become the first homeodomain protein.

It is interesting to note that TALE homeodomain proteins form homo- or heterodimers in plants, fungi, and animals. Further, both in fungi and animals, typical homeodomain proteins, in animals, those from the HOX cluster, and TALE homeodomain proteins also interact. It suggests that these interactions are of an ancient nature and probably were already present in the CAO. In animals, the diversification of the homeobox genes has led to a large proliferation of different classes and families. The last common ancestor of all higher animals, that is, protostomes and deuterostomes, already contained the diversity of all the different homeobox classes. Counting the different conserved gene families, at least 70 different homeobox genes must have already been present in that animal.

Present day organisms such as *Drosophila* and *C. elegans* have around 100 homeobox genes, while vertebrates, because of the large-scale duplications in early vertebrate evolution, have more than 200. While there are many other types of transcription factors encoded in an animal genome, the homeobox genes are perhaps the most important group that played essential roles in shaping the evolution of animals.

See also Genetics, Molecular Basis of; Nematode (*C. elegans*), Molecular Biology of; Nematodes, Neurobiology and Development of.

Bibliography

Books and Reviews

Bürglin, T.R. (1994) A comprehensive classification of homeobox genes, in: Duboule, D. (Ed.) *Guidebook to the Homeobox Genes*, Oxford University Press, Oxford, UK, pp. 25–71.

Chi, N., Epstein, J.A. (2002) Getting your Pax straight: Pax proteins in development and disease, *Trends Genet.* **18**, 41–47.

Duboule, D. (Ed.) (1994) *Guidebook to the Homeobox Genes*, Oxford University Press, Oxford, UK.

Gehring, W.J., Ikeo, K. (1999) Pax 6: mastering eye morphogenesis and eye evolution, *Trends Genet.* **15**, 371–377.

Gehring, W.J., Affolter, M., Bürglin, T.R. (1994a) Homeodomain proteins, *Annu. Rev. Biochem.* **63**, 487–526.

Gehring, W.J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter, M., Otting, G., Wüthrich, K. (1994b) Homeodomain-DNA recognition, *Cell* **78**, 211–223.

Hobert, O., Westphal, H. (2000) Functions of LIM-homeobox genes, *Trends Genet.* **16**, 75–83.

Jagla, K., Bellard, M., Frasch, M. (2001) A cluster of Drosophila homeobox genes involved in mesoderm differentiation programs, *BioEssays* **23**, 125–133.

Kawakami, K., Sato, S., Ozaki, H., Ikeda, K. (2000) Six family genes–structure and function as transcription factors and their roles in development, *BioEssays* **22**, 616–626.

Kmita, M., Duboule, D. (2003) Organizing axes in time and space; 25 years of colinear tinkering, *Science* **301**, 331–333.

Primary Literature

Aboobaker, A.A., Blaxter, M.L. (2003) Hox gene loss during dynamic evolution of the nematode cluster, *Curr. Biol.* **13**, 37–40.

Acampora, D., D'Esposito, M., Faiella, A., Pannese, M., Migliaccio, E., Morelli, F., Stornaiuolo, A., Nigro, V., Simeone, A., Boncinelli, E. (1989) The human HOX gene family, *Nucleic Acids Res.* **17**, 10385–10402.

Aparicio, S., Hawker, K., Cottage, A., Mikawa, Y., Zuo, L., Venkatesh, B., Chen, E., Krumlauf, R., Brenner, S. (1997) Organization of the Fugu rubripes Hox clusters: evidence for continuing evolution of vertebrate Hox complexes, *Nat. Genet.* **16**, 79–83.

Arber, S., Halder, G., Caroni, P. (1994) Muscle LIM protein, a novel essential regulator of myogenesis, promotes myogenic differentiation, *Cell* **79**, 221–231.

Balling, R., Deutsch, U., Gruss, P. (1988) Undulated, a mutation affecting the development of the mouse skeleton, has a point mutation in the paired box of Pax 1, *Cell* **55**, 531–535.

Bellaoui, M., Pidkowich, M.S., Samach, A., Kushalappa, K., Kohalmi, S.E., Modrusan, Z., Crosby, W.L., Haughn, G.W. (2001) The Arabidopsis BELL1 and KNOX TALE homeodomain proteins interact through a domain conserved between plants and animals, *Plant Cell* **13**, 2455–2470.

Bender, A., Sprague Jr., G.F. (1987) MATa1 protein, a yeast transcription activator, binds synergistically with a second protein to a set of cell-type-specific genes, *Cell* **50**, 681–691.

Bertolino, E., Reimund, B., Wildt-Perinic, D., Clerc, R.G. (1995) A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif, *J. Biol. Chem.* **270**, 31178–31188.

Billeter, M., Güntert, P., Luginbühl, P., Wüthrich, K. (1996) Hydration and DNA recognition by homeodomains, *Cell* **85**, 1057–1065.

Blochlinger, K., Bodmer, R., Jack, J., Jan, L.Y., Jan, Y.N. (1988) Primary structure and expression of a product from *cut*, a locus involved in specifying sensory organ identity in *Drosophila*, *Nature* **333**, 629–635.

Bopp, D., Burri, M., Baumgartner, S., Frigerio, G., Noll, M. (1986) Conservation of a large protein domain in the segmentation gene *paired* and in functionally related genes of *Drosophila*, *Cell* **47**, 1033–1040.

Brooke, N.M., Garcia-Fernàndez, J., Holland, P.W.H. (1998) The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster, *Nature* **392**, 920–922.

Bürglin, T.R. (1994) A *Caenorhabditis elegans prospero* homologue defines a novel domain, *Trends Biochem. Sci.* **19**, 70–71.

Bürglin, T.R. (1995) The Evolution of Homeobox Genes, in: Arai, R., Kato, M. Doi, Y. (Eds.) *Biodiversity and Evolution*, The National Science Museum Foundation, Tokyo, Japan, pp. 291–336.

Bürglin, T.R. (1997) Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals, *Nucleic Acids Res.* **25**, 4173–4180.

Bürglin, T.R. (1998) The PBC domain contains a MEINOX domain: coevolution of Hox and TALE homeobox genes? *Dev. Genes Evol.* **208**, 113–116.

Bürglin, T.R. (2003) The homeobox genes of *Encephalitozoon cuniculi* (Microsporidia) reveal a putative mating-type locus, *Dev. Genes Evol.* **213**, 50–52.

Bürglin, T.R., Cassata, G. (2002) Loss and gain of domains during evolution of cut superclass homeobox genes, *Int. J. Dev. Biol.* **46**, 115–123.

Bürglin, T.R., Ruvkun, G. (1992) New motif in PBX genes, *Nat. Genet.* **1**, 319–320.

Bürglin, T.R., Finney, M., Coulson, A., Ruvkun, G. (1989) *Caenorhabditis elegans* has scores of homoeobox-containing genes, *Nature* **341**, 239–243.

Carrasco, A.E., McGinnis, W., Gehring, W.J., De Robertis, E.M. (1984) Cloning of an *X. laevis* gene expressed during early embryogenesis coding for a peptide region homologous to *Drosophila* homeotic genes, *Cell* **37**, 409–414.

Castro, L.F., Holland, P.W. (2003) Chromosomal mapping of ANTP class homeobox genes in amphioxus: piecing together ancestral genomes, *Evol. Dev.* **5**, 459–465.

Chi, Y.I., Frantz, J.D., Oh, B.C., Hansen, L., Dhe-Paganon, S., Shoelson, S.E. (2002) Diabetes mutations delineate an atypical POU domain in HNF-1alpha, *Mol. Cell.* **10**, 1129–1137.

Cho, K.W.Y., Blumberg, B., Steinbeisser, H., De Robertis, E.M. (1991) Molecular nature of Spemann's organizer: the role of the *Xenopus* homeobox gene *goosecoid*, *Cell* **67**, 1111–1120.

Clark, S.G., Chisholm, A.D., Horvitz, H.R. (1993) Control of cell fates in the central body region of *C. elegans* by the homeobox gene *lin-39*, *Cell* **74**, 43–55.

Cohen, S.M., Brönner, G., Küttner, F., Jürgens, G., Jäckle, H. (1989) Distal-less encodes a homoeodomain protein required for limb development in Drosophila, *Nature* **338**, 432–434.

Czerny, T., Schaffner, G., Busslinger, M. (1993) DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site, *Genes Dev.* **7**, 2048–2061.

Damen, W.G. (2002) fushi tarazu: a Hox gene changes its role, *BioEssays* **24**, 992–995.

Dasen, J.S., Liu, J.P., Jessell, T.M. (2003) Motor neuron columnar fate imposed by sequential phases of Hox-c activity, *Nature* **425**, 926–933.

de Rosa, R., Grenier, J.K., Andreeva, T., Cook, C.E., Adoutte, A., Akam, M., Carroll, S.B., Balavoine, G. (1999) Hox genes in brachiopods and priapulids and protostome evolution, *Nature* **399**, 772–776.

Doe, C.Q., Chu-LaGraff, Q., Wright, D.M., Scott, M.P. (1991) The *prospero* gene specifies cell fates in the *Drosophila* central nervous system, *Cell* **65**, 451–464.

Driever, W., Nüsslein-Volhard, C. (1988) The *bicoid* protein determines position in the *Drosophila* embryo in a concentration-dependent manner, *Cell* **54**, 95–104.

Duboule, D., Dollé, P. (1989) The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes, *EMBO J* **8**, 1497–1505.

Finkelstein, R., Perrimon, N. (1990) The *orthodenticle* gene is regulated by *bicoid* and *torso* and specifies *Drosophila* head development, *Nature* **346**, 485–488.

Finney, M., Ruvkun, G. (1990) The *unc-86* gene product couples cell lineage and cell identity in *C. elegans*, *Cell* **63**, 895–905.

Frain, M., Swart, G., Monaci, P., Nicosia, A., Stämpfli, S., Frank, R., Cortese, R. (1989) The liver-specific transcription factor LF-B1

contains a highly diverged homeobox DNA binding domain, *Cell* **59**, 145–157.

Freyd, G., Kim, S., Horvitz, R.H. (1990) Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gene *lin-11*, *Nature* **344**, 876–879.

Galant, R., Carroll, S.B. (2002) Evolution of a transcriptional repression domain in an insect Hox protein, *Nature* **415**, 910–913.

Galliot, B., de Vargas, C., Miller, D. (1999) Evolution of homeobox genes: Q50 paired-like genes founded the paired class, *Dev. Genes Evol.* **209**, 186–197.

Garcia-Fernàndez, J., Holland, P.W.H. (1994) Archetypal organization of the amphioxus *Hox* gene cluster, *Nature* **370**, 563–566.

Garcia-Fernàndez, J., Baguñà, J., Saló, E. (1991) Planarian homeobox genes: cloning, sequence analysis, and expression, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7338–7342.

Gauchat, D., Mazet, F., Berney, C., Schummer, M., Kreger, S., Pawlowski, J., Galliot, B. (2000) Evolution of Antp-class genes and differential expression of hydra Hox/paraHox genes in anterior patterning, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4493–4498.

Gehring, W.J. (1994) A History of the Homeobox, in: Duboule, D. (Ed.) *Guidebook to the Homeobox Genes*, Oxford University Press, Oxford, UK, pp. 1–10.

Gehring, W.J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter, M., Otting, G., Wüthrich, K. (1994) Homeodomain-DNA recognition, *Cell* **78**, 211–223.

Gómez-Skarmeta, J-L., Diez del Corral, R., de la Calle-Mustienes, E., Ferrés-Marcó, D., Modolell, J. (1996) A*raucan* and *caupolican*, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes, *Cell* **85**, 95–105.

Graham, A., Papalopulu, N., Krumlauf, R. (1989) The murine and drosophila homeobox gene complexes have common features of organization and expression, *Cell* **57**, 367–378.

Guazzi, S., Price, M., De Felice, M., Damante, G., Mattei, M-G., Di Lauro, R. (1990) Thyroid nuclear factor (TTF-1) contains a homeodomain and displays a novel DNA binding specificty, *EMBO J* **9**, 3631–3639.

Halder, G., Callaerts, P., Gehring, W.J. (1995) Induction of ectopic eyes by targeted expression of the *eyeless* gene in *Drosophila*, *Science* **267**, 1788–1792.

Hall, M.N., Johnson, A.D. (1987) Homeo domain of the yeast repressor $\alpha 2$ is a sequence-specific DNA-binding domain but is not sufficient for repression, *Science* **237**, 1007–1012.

Harada, R., Bérubé, G., Tamplin, O.J., Denis-Larose, C., Nepveu, A. (1995) DNA-binding specificity of the Cut repeats from the human Cut-like protein, *Mol. Cell. Biol.* **15**, 129–140.

Hashimoto, T., Nakano, Y., Morinaga, T., Tamaoki, T. (1992) A new family of homeobox genes encoding multiple homeodomain and zinc finger motifs, *Mech. Dev.* **39**, 125–126.

Herr, W., Sturm, R.A., Clerc, R.G., Corcoran, L.M., Baltimore, D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Finney, M., Ruvkun, G., Horvitz, H.R. (1988) The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and *Caenorhabditis elegans unc-86* gene products, *Genes Dev.* **2**, 1513–1516.

Ho, C-Y., Adamson, J.G., Hodges, R.S., Smith, M. (1994) Heterodimerization of the yeast MATa1 and MATα2 proteins is mediated by two leucine zipper-like coiled-coil motifs, *EMBO J* **13**, 1403–1413.

Ingraham, H.A., Chen, R., Mangalam, H.J., Elsholtz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L., Rosenfeld, M.G. (1988) A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype, *Cell* **55**, 519–529.

Jacobson, E.M., Li, P., Leon-del-Rio, A., Rosenfeld, M.G., Aggarwal, A.K. (1997) Structure of Pit-1 POU domain bound to DNA as a dimer: unexpected arrangement and flexibility, *Genes Dev.* **11**, 198–212.

Kappen, C., Schughart, K., Ruddle, F.H. (1993) Early evolutionary origin of major homeodomain sequence classes, *Genomics* **18**, 54–70.

Karlsson, O., Thor, S., Norberg, T., Ohlsson, H., Edlund, T. (1990) Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain, *Nature* **344**, 879–882.

Kaufman, T.C., Seeger, M.A., Olsen, G. (1990) Molecular and genetic organization of the Antennapedia gene complex of Drosophila melanogaster, *Adv. Genet.* **27**, 309–362.

Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B., Pabo, C.O. (1990) Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a

framework for understanding homeodomain-DNA interactions, *Cell* **63**, 579–590.

Klemm, J.D., Pabo, C.O. (1996) Oct-1 POU domain-DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage, *Genes Dev.* **10**, 27–36.

Klemm, J.D., Rould, M.A., Aurora, R., Herr, W., Pabo, C.O. (1994) Crystal structure of the Oct-1 POU domain bound to an Octamer site: DNA recognition with tethered DNA-binding modules, *Cell* **77**, 21–32.

Kmita-Cunisse, M., Loosli, F., Bièrne, J., Gehring, W.J. (1998) Homeobox genes in the ribbonworm *Lineus sanguineus*: evolutionary implications, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3030–3035.

Laughon, A., Scott, M.P. (1984) Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins, *Nature* **310**, 25–31.

Lemaigre, F.P., Durviaux, S.M., Truong, O., Lannoy, V.J., Hsuan, J.J., Rousseau, G.G. (1996) Hepatocyte nuclear factor 6, a transcription factor that contains a novel type of homeodomain and a single *cut* domain, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9460–9464.

Lewis, E.B. (1978) A gene complex controlling segmentation in Drosophila, *Nature* **276**, 565–570.

Luke, G.N., Castro, L.F., McLay, K., Bird, C., Coulson, A., Holland, P.W. (2003) Dispersal of NK homeobox gene clusters in amphioxus and humans, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5292–5295.

Martin, C.H., Mayeda, C.A., Davis, C.A., Ericsson, C.L., Knafels, J.D., Mathog, D.R., Celniker, S.E., Lewis, E.B., Palazzolo, M.J. (1995) Complete sequence of the bithorax complex of *Drosophila*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8398–8402.

McGinnis, W., Krumlauf, R. (1992) Homeobox genes and axial patterning, *Cell* **68**, 283–302.

McGinnis, W., Hart, C.P., Gehring, W.J., Ruddle, F.H. (1984b) Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of Drosophila, *Cell* **38**, 675–680.

McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A., Gehring, W.J. (1984a) A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans, *Cell* **37**, 403–408.

Miles, A., Miller, D.J. (1992) Genomes of diploblastic organisms contain homeoboxes: sequence of *eveC*, an *even-skipped* homologue from the cnidarian *Acropora formosa*, *Proc. R. Soc. Lond. B.* **248**, 159–161.

Miller, D.J., Miles, A. (1993) Homeobox genes and the zootype, *Nature* **365**, 215–216.

Neufeld, E.J., Skalnik, D.G., Lievens, P.M-J., Orkin, S.H. (1992) Human CCAAT displacement protein is homologous to the *Drosophila* homeoprotein *cut*, *Nat. Genet.* **1**, 50–55.

Oliver, G., Sosa-Pineda, B., Geisendorf, S., Spana, E.P., Doe, C.Q., Gruss, P. (1993) Prox 1, a *prospero*-related homeobox gene expressed during mouse development, *Mech. Dev.* **44**, 3–16.

Oliver, G., Wehr, R., Jenkins, N.A., Copeland, N.G., Cheyette, B.N.R., Hartenstein, V., Zipursky, S.L., Gruss, P. (1995) Homeobox genes and connective tissue patterning, *Development* **121**, 693–705.

Otting, G., Qian, Y.Q., Billeter, M., Müller, M., Affolter, M., Gehring, W.J., Wüthrich, K. (1990) Protein-DNA contacts in the structure of a homeodomain - DNA complex determined by nuclear magnetic resonance spectroscopy in solution, *EMBO J* **9**, 3085–3092.

Passner, J.M., Ryoo, H.D., Shen, L., Mann, R.S., Aggarwal, A.K. (1999) Structure of a DNA-bound Ultrabithorax-extradenticle homeodomain complex, *Nature* **397**, 714–719.

Phillips, K., Luisi, B. (2000) The virtuoso of versatility: POU proteins that flex to fit, *J. Mol. Biol.* **302**, 1023–1039.

Poole, S.J., Kauvar, L.M., Drees, B., Kornberg, T. (1985) The engrailed locus of Drosophila: structural analysis of an embryonic transcript, *Cell* **40**, 37–43.

Porter, S.D., Smith, M. (1986) Homoeo-domain homology in yeast MATα2 is essential for repressor activity, *Nature* **320**, 766–768.

Qian, Y.Q., Billeter, M., Otting, G., Müller, M., Gehring, W.J., Wüthrich, K. (1989) The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors, *Cell* **59**, 573–580.

Quiring, R., Walldorf, U., Kloter, U., Gehring, W.J. (1994) Homology of the *eyeless* gene of *Drosophila* to the *small eye* gene in mice and *Aniridia* in humans, *Science* **265**, 785–789.

Rauskolb, C., Peifer, M., Wieschaus, E. (1993) *Extradenticle*, a regulator of homeotic gene

activity, is a homolog of the homeobox-containing human proto-oncogene *pbx1*, *Cell* **74**, 1101–1112.

Read, A.P. (1995) Pax genes - *paired* feet in three camps, *Nat. Genet.* **9**, 333–334.

Rieckhof, G.E., Casares, F., Ryoo, H.D., Abu-Shaar, M., Mann, R.S. (1997) Nuclear translocation of extradenticle requires *homothorax*, which encodes an extradenticle-related homeodomain protein, *Cell* **91**, 171–183.

Ronshaugen, M., McGinnis, N., McGinnis, W. (2002) Hox protein mutation and macroevolution of the insect body plan, *Nature* **415**, 914–917.

Ruberti, I., Sessa, G., Lucchetti, S., Morelli, G. (1991) A novel class of plant proteins containing a homeodomain with a closely linked leucine zipper motif, *EMBO J* **10**, 1787–1791.

Ryter, J.M., Doe, C.Q., Matthews, B.W. (2002) Structure of the DNA binding region of prospero reveals a novel homeo-prospero domain, *Structure (Camb)* **10**, 1541–1549.

Sadler, I., Crawford, A.W., Michelsen, J.W., Beckerle, M.C. (1992) Zyxin and cCRP: two interactive LIM domain proteins associated with the cytoskeleton, *J. Cell. Biol.* **119**, 1573–1587.

Sánchez-Herrero, E., Vernós, I., Marco, R., Morata, G. (1985) Genetic organization of *Drosophila* bithorax complex, *Nature* **313**, 108–113.

Schmeichel, K.L., Beckerle, M.C. (1994) The LIM domain is a modular protein-binding interface, *Cell* **79**, 211–219.

Schummer, M., Scheurlen, I., Schaller, C., Galliot, B. (1992) HOM/HOX homeobox genes are present in hydra (*Chlorohydra viridissima*) and are differentially expressed during regeneration, *EMBO J* **11**, 1815–1823.

Scott, M.P., Weiner, A.J. (1984) Structural relationships among genes that control development: sequence homology between the *Antennapedia*, *Ultrabithorax*, and *fushi tarazu* loci in *Drosophila*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4115–4119.

Shepherd, J.C.W., McGinnis, W., Carrasco, A.E., De Robertis, E.M., Gehring, W.J. (1984) Fly and frog homoeo domains show homologies with yeast mating type regulatory proteins, *Nature* **310**, 70–71.

Simeone, A., Acampora, D., Gulisano, M., Stornaiuolo, A., Boncinelli, E. (1992) Nested expression domains of four homeobox genes in developing rostral brain, *Nature* **358**, 687–690.

Sun, S., Ting, C.T., Wu, C.I. (2004) The normal function of a speciation gene, Odysseus, and its hybrid sterility effect, *Science* **305**, 81–83.

Tsuchida, T., Ensini, M., Morton, S.B., Baldassare, M., Edlund, T., Jessell, T.M., Pfaff, S.L. (1994) Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes, *Cell* **79**, 957–970.

Vaessin, H., Grell, E., Wolff, E., Bier, E., Jan, L.Y., Jan, Y.N. (1991) *Prospero* is expressed in neuronal precursors and encodes a nuclear protein that is involved in the control of axonal outgrowth in *Drosophila*, *Cell* **67**, 941–953.

Vollbrecht, E., Veit, B., Sinha, N., Hake, S. (1991) The developmental gene knotted-1 is a member of a maize homeobox gene family, *Nature* **350**, 241–243.

Wada, S., Tokuoka, M., Shoguchi, E., Kobayashi, K., Di Gregorio, A., Spagnuolo, A., Branno, M., Kohara, Y., Rokhsar, D., Levine, M., Saiga, H., Satoh, N., Satoh, Y. (2003) A genomewide survey of developmentally relevant genes in Ciona intestinalis. II. Genes for homeobox transcription factors, *Dev. Genes Evol.* **213**, 222–234.

Wang, B.B., Müller-Immergluck, M.M., Austin, J., Robinson, N.T., Chisholm, A., Kenyon, C. (1993) A homeotic gene cluster patterns the anteroposterior body axis of *C. elegans*, *Cell* **74**, 29–42.

Way, J.C., Chalfie, M. (1988) *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*, *Cell* **54**, 5–16.

Wellik, D.M., Capecchi, M.R. (2003) Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton, *Science* **301**, 363–367.

Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D., Pabo, C.O. (1991) Crystal structure of MATα2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions, *Cell* **67**, 517–528.

Xu, W., Rould, M.A., Jun, S., Desplan, C., Pabo, C.O. (1995) Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations, *Cell* **80**, 639–650.

Horizontal Gene Transfer

Jack A. Heinemann and Ralph A. Bungard
University of Canterbury, Christchurch, New Zealand

1 Horizontal Gene Transfer Versus Introgression 225

2 Vectors and Pathways of HGT 227
2.1 Transformation and Pathogenic Bacteria 227
2.2 Conjugation 228
2.3 Parasites, Symbionts and Pathogens 229
2.3.1 Viruses and Transduction 229
2.3.2 Parasitic Plants 230
2.3.3 Parasitic and Pathogenic Animals 230
2.4 Hybridization 231

3 Different Ways to Observe HGT 231
3.1 Descriptive Bioinformatics 233
3.2 Experimentation 235
3.2.1 Seminal Experimentalist Approaches 235
3.2.2 Modern Process Approaches 235

4 Barriers to HGT 237
4.1 Transfer and Inheritance 238
4.2 Introgression 238

5 Final Thoughts 240

Acknowledgment 240

Bibliography 240
Books and Reviews 240
Primary Literature 241

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

Keywords

Bacteriophage (phage)
Viruses specific to bacteria.

Competition Model
Derives from an experimental demonstration of the evolutionary forces that determine when genes reproduce by HGT. On the basis of the hypothesis that different genes reproduce faster by HGT in different environments and the change in environment determines when genes become mobile. Stability of the environment then influences which genes introgress. Produces unique predictions of how genes are recruited to HGT vectors.

Complexity Hypothesis
Posits that genes least likely to introgress are of a type that are involved in highly interdependent large biochemical complexes. The interactions lead to constraints that would preclude replacement by genes that have diverged.

Conjugation
A form of DNA transfer mediated by infectious elements (called *plasmids*) normally found in bacteria. Plasmids transfer from bacteria to all known forms of life.

Hybridization (to form hybrids)
The initial product of a sexual cross between individuals that are considered to be of different varieties, subspecies, species, or genera and that might often be revealed by low fecundity or infertile offspring.

Introgression
''The permanent incorporation of genes from one set of differentiated populations (species, subspecies, races, and so on) into another'' (quote from Stewart Jr. et al.).

Phylogeny
The evolutionary connections between organisms inferred from assumed descent via a common ancestor.

Transduction
The transfer of genes by viruses to different organisms within the infectious range of the virus.

Transformation
The change in the phenotype of organisms due to the inheritance of a gene. In molecular biology jargon, it often is a term that also refers to the uptake of DNA regardless of whether it alters the phenotype.

Vectors
Viruses, gametes, organisms, plasmids, or other agents that can deliver genes by HGT.

Vertical Gene Transfer
Reproduction of genes in the context of whole genome replication and reproduction of the organism or entire cell of an organism.

Horizontal gene transfer (HGT) may be defined as any occurrence of heritable material passing between organisms, asynchronous with reproduction of the organisms. It represents replication of heritable material outside the context of parent to offspring (i.e. vertical) reproduction. Three types of evidence traditionally lead to claims of HGT. Firstly, biochemical and genetic observations of real-time (usually laboratory recreations of) gene flow between two genotypically distinguishable individuals. Secondly, DNA sequences that are common within a species but inconsistently found in that genus, and finally, direct evidence of processes that can create genes that reproduce horizontally. HGT is now confirmed to occur between all biological kingdoms and has consequences even when genes fail to introgress.

1 Horizontal Gene Transfer Versus Introgression

Awareness of HGT (horizontal gene transfer) has risen in recent times in part due to some spectacular scientific debates surrounding, particularly, the release of genetically modified organisms (GMOs) as food crops and the spread of antibiotic-resistant bacteria that cause disease. The recent controversies belie the long history of research in HGT, sometimes treating it as if it had been discovered only since genome sequences began to be completed. On the contrary, HGT is a general and pervasive biological phenomenon, but it has been packaged into separate boxes that go by other names. We attempt here to identify what is fundamental to HGT so that it may be recognized in the various other boxes it appears in, and to place it center stage in the modern scientific discourses that depend on understanding it.

There is genuine confusion about what HGT is and what it is not. Part of this confusion results from semantics, with some authors simply preferring the term lateral gene transfer (LGT) to HGT. To resolve this issue, C.I. Kado suggested that the term LGT be reserved for gene transfers between closely related partners and HGT for more distant. This proposal has some merit, but it presupposes that the important defining characteristic of HGT resides in the organisms that are donors and recipients of genes rather than the genes themselves. Besides the obvious problem of knowing when partners are close enough to call exchanges between them (LGT), it is entirely possible for two partners to swap a gene that is

phylogenetically more distant to both the donor and recipient than either is to one another.

The most important source of confusion is the tendency to substitute an outcome of HGT with a description of HGT. Two outcomes are most commonly associated with HGT. One outcome is that HGT can generate genotypic diversity in a species or offer a selective advantage to recombinants within a species. This has tempted many to consider HGT a process by which organisms, especially the asexual bacteria, recover some of the benefits of sex. There is no denying that HGT can provide genes that benefit the recipient and introduce genetic diversity into populations, but these outcomes are based on the biology of the recipient organism and are only one of the consequences of HGT.

A second outcome is introgression. Genes, and sometimes entire chromosomes, can be introduced into a separate, differentiated population. For a gene to become permanent, it almost certainly must have provided an adaptive function, or survived a fortunate bottleneck. Moreover, the recombinant genome must be found in many if not all individuals of a species. For that to occur, the recombinant genome must outcompete most or all genomes that lacked the new gene.

The concept of introgression is normally reserved for describing vertical gene transfer over the barriers erected by meiosis. However, aspects of the concept also apply to HGT. Whether an alien gene is introduced during (e.g. hybridization) or asynchronously with reproduction (i.e. HGT), there are barriers to it becoming a stable and common part of the genome of the species. Alien genes likewise must be present in a significant proportion of a population in order to be detected. Overcoming these barriers can lead to alien genes and chromosomes becoming a characteristic feature of a species.

The process by which genes introgress, however, is only part of the way genes evolve by HGT (Fig. 1). Genome sequencing reveals past gene transfers by finding alien genes in genomes. The number of alien genes found by such methods only reflects the number of genes that have introgressed, not the number or frequency of transfer. This one outcome of HGT is certainly important evidence for HGT, but it is ill suited to measuring its impact and frequency. Measurements of HGT that rest with genomic DNA sequences are essentially a measure of the range and frequencies in which genomes retain genes, not the range and frequencies of gene transfers that make the genes available in the first place (Fig. 1).

Making HGT the same as an outcome of HGT, namely, the "acquisition of 'alien genes' by a particular genome" (quote from Koonin et al. in Syvanen and Kado) substitutes historic description for understanding how HGT happens biochemically and how it influences organisms physiologically. Quite rightly, this focus leads to strong criticisms about proper identification of DNA sequences that are indeed alien.

Understanding how genes introgress, therefore, is not the same as understanding HGT. HGT is not limited to transfers that change organismal genotypes. For example, some genomes may evolve almost exclusively by HGT. Viruses and transposable elements like *mariner* are arguably the product of evolution that has little to do with organismal genomes and their vertical reproduction.

The essence of HGT as an evolutionary process, and not just an outcome, must be preserved in any definition. HGT may

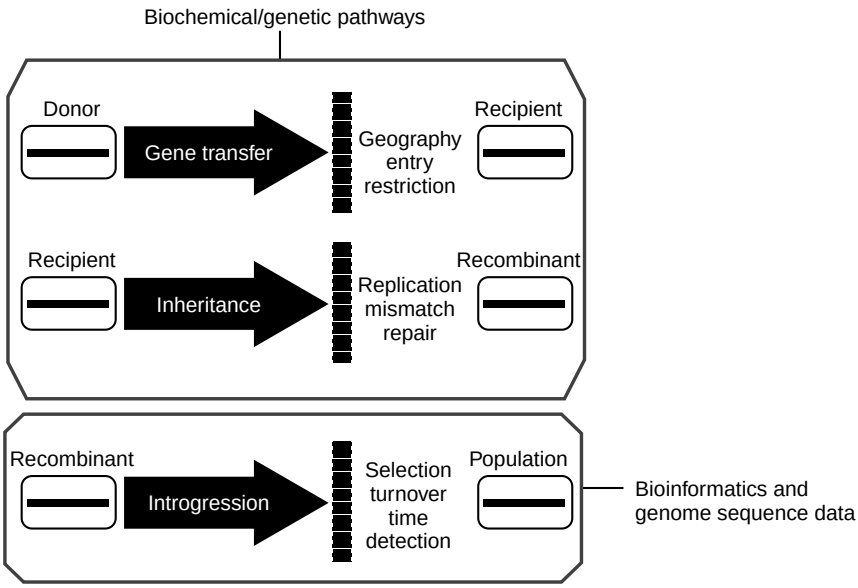

Fig. 1 Different levels of analysis used to study horizontal gene transfer. Symbols: rounded rectangles = cells or organisms, solid rectangles = mobile gene, brick wall = listed barriers. Only process experiments as discussed in Sect. 3 can cover all levels.

be defined as any occurrence of heritable material passing between organisms, asynchronous with reproduction of the organisms. It represents replication of heritable material outside the context of parent to offspring (i.e. vertical) reproduction. This definition permits different means of transfer – such as viruses or transduction, conjugation, and transformation – and different outcomes of transfer – such as serial infection or recombination – to be united in legitimate study about how genes evolve through HGT.

2 Vectors and Pathways of HGT

2.1 Transformation and Pathogenic Bacteria

Bacteria can at times become competent for the uptake of DNA from their surrounding environment. Should part or all of that DNA become stably incorporated into the genome, they are called *transformants*. Competence may be expressed constitutively in a proportion of the population or be induced. It is usually associated with the expression of a set of proteins, called *com* for competence, that include or are in addition to structures (e.g. vesicles and pili in some gram-negative bacteria) associated with the ability to bind extracellular DNA and translocate it into the cytoplasm.

The most widely studied examples of naturally competent bacteria are *Streptococcus pneumoniae* and *Bacillus subtilis* (of the gram-positives) and *Haemophilus influenzae* and *Neisseria gonorrhoeae* (of the gram-negatives). Although the molecular details of transformation varies between organisms, in each case DNA is taken into the cell in single-stranded form where it is thought to serve as a substrate for

single-strand specific DNA binding recombinases (e.g. analogs of *Escherichia coli's* RecA protein) that initiate a search for similar sequences of DNA in the recipient genome. This recombination process is aborted or reversed, depending on the activities of the endogenous mismatch-repair (MMR) pathways, if the incoming strand creates too many mismatched base pairs. Depending on the efficiency of MMR at the time, the incoming DNA may survive through a subsequent round of DNA replication and become immortalized in the recipient genome.

The uptake process has been characterized to the molecular level for a small number of bacteria and a small number of environmental conditions that induce competence. Many bacteria that are not known to be competent have homologs of *com* genes, which might indicate that they do have the potential to take up DNA from the extracellular environment. It remains unknown how many bacteria may take up DNA and all the environments in which they may be induced to take up DNA naturally. Even the icon of noncompetent bacteria, *E. coli*, was recently shown to develop a natural competence in laboratory studies.

In contrast, no special competence seems to be required for eukaryotic cells to take up DNA by transformation. DNA can pass into animal cells *in situ* simply through food and has been reported to pass through the placenta and recombine into the mammalian genome. Animal cells can also be transformed *in vivo* and *in vitro* using bacteria that cross into the cytoplasm. Viruses also can produce proteins that serve to transfer genes within multicellular organisms, bypassing the need to reinfect from the outside.

2.2 Conjugation

Bacterial conjugation in its broadest sense includes the conjugative plasmid groups discovered in gram-negative bacteria, the T-DNA of the Ti plasmid, pheromone-responsive plasmids and conjugative transposons. The focus in this overview is on the model conjugative systems defined by the IncP and IncF groups of plasmids of gram-negative bacteria.

Plasmids are similar to viruses but they lack any known extracellular form. Though loosely defined as extrachromosomal, like some viruses they can be found integrated into chromosomes at times. They have also been defined as accessory elements or parasites because hosts cured of plasmids can survive and reproduce in at least one environment. Conjugative plasmids got their name because for a time it was thought that they replaced the need for sex in prokaryotes. This idea derived from the use of genes with adaptive value, such as antibiotic resistance genes, to monitor the transfer of the plasmids. None of the above-mentioned characteristics is universal, so we have opted to define plasmids by their ability to reproduce in at least one environment in which the host cannot. That niche may not be physical but temporal, such as provided by HGT.

Conjugation mediated by the IncF and IncP plasmids requires, at a minimum, a cis-acting DNA sequence called *the origin of transfer* (*oriT*). All other functions (called *tra*) act in trans thus allowing plasmids with all trans-acting functions to also transfer plasmids with no or a few trans-acting functions. The trans-acting gene products are divided further into those involved in DNA metabolism (and usually specific to a particular *oriT*) and those involved in DNA transport

and cell–cell interactions (and thus most likely to interact with a greater range of other plasmids). The conjugative genes that are specific to DNA metabolism introduce a "nick" at *oriT* and initiate the unwinding and concomitant transfer of DNA to a recipient cell. Both strands are used as templates for the synthesis of a complementary strand, one in the donor cell and one in the recipient.

The tumor-inducing (Ti) plasmid from *Agrobacterium tumefaciens* is a plasmid with two conjugative systems. One system mobilizes the entire plasmid; the second system mobilizes a single region of the plasmid, called *the T-DNA*. It is the T-DNA that is recovered from crown gall tissue of plants infected by *A. tumefaciens*, a discovery that unambiguously demonstrated that conjugation was a means to transfer DNA across biological kingdoms.

It has become clear over the past decade that the DNA transport apparatus of conjugation is the ancestor, or at least a sibling, of other macromolecular transport systems such as the type IV protein secretion pathways. Bacteria that secrete proteins into human cells during infection may also secrete plasmids using the same apparatus. Consistent with this speculation is the ability of conjugative plasmids to transfer from bacteria to eukaryotic cells, and the ability of bacteria to support conjugation *in situ* and inside human cells.

2.3 Parasites, Symbionts and Pathogens

Parasites, symbionts, and pathogens are excellent gene vectors. Viruses, like many pathogens, can have much broader transfer ranges than symptom ranges. Some viruses transfer between plants and animals, and bacteriophage can penetrate the human central nervous system. These vectors also transmit a wide range of molecules in addition to DNA, such as RNA and proteins. These molecules are relevant here because prions and dsRNA are horizontally transferred epigenes capable of transforming the germline.

2.3.1 Viruses and Transduction

Transduction is formally viral-mediated transfer of nonvirus genes between hosts. While it is not uncommon for viruses to acquire genes from their hosts, and thus acquire new virulence or host range characteristics, transduction is distinctive because, like transformation, it delivers genes most recently of chromosomal origin that become stably integrated into the recipient genome.

Generalized transducing bacteriophage, the viruses specific to bacteria, package host DNA instead of viral DNA at some low frequency during infection (on the order of 0.001%). This phenomenon is called *generalized transduction* because any region of the chromosome may be packaged. The size of the phage determines how much DNA may be transduced by a single phage, which is normally only 1 to 2% of the host genome. The mature phage particle retains the ability to attach and initiate an infection, but delivers cellular rather than viral DNA to the recipient. Since the phage particle does not contain the viral genome, the infection does not result in the production of new viruses.

All temperate phages that integrate into a chromosome as a part of their life cycle have the potential to be *specialized* transducing phage. During the process of excising from the chromosome to begin an infectious cycle, DNA flanking the phage genome can be packaged at some low frequency. This phenomenon is known as *specialized* because only a small number of

genes are ever transduced by this phage. Phage integrating into the chromosome usually does not enter at random, but instead enters at special DNA sequences in the host chromosome. Thus, it is only the genes flanking integration sites that are transduced.

Bacterial genomes are littered with the skeletons of phage, many of which no longer have an infectious life cycle. These skeletons are revealed when they are the source of genes that promote bacterial virulence (pathogenicity islands) or symbiosis (symbiosis islands). They may also provide sequences into which still active temperate phages could integrate by homologous recombination or by a site-specific mechanism. These phage remnants also make it difficult to say what a host gene is and what a viral gene is not. This ambiguity is part of the reason why we consider viral infection itself a demonstration of HGT rather than simply when an arbitrary marker is acquired in the course of a phage infection.

2.3.2 Parasitic Plants

Around 1% of all flowering plants (angiosperms) are parasitic, attaching to host plant species from which they extract some, if not all, of the carbon, nutrients, and water they require for growth. Some of these parasites, like the Rafflesiaceae, have evolved to such an extent that the only outward signs of the parasite are when they exit the host to flower. It has long been known that parasitic plants not only exchange carbon and nutrients with their host but that they also exchange complex molecules like alkaloids and pathogens including viruses, bacteria, and phytoplasmas. This intimate relationship may also facilitate HGT between plant species. A multigene phylogenetic analysis of the parasitic Rafflesiaceae has shown that some genes from both the nuclear and the mitochondrial genomes are distinct to the order Malpighiales while other mitochondrial genes are closely associated with their obligate host *Tetrastigma*. The Rafflesiaceae are obligate host parasites, meaning that they only parasitize one host species. However, other parasitic plant species like the *Cuscuta* can attach to a broad range of hosts and can even simultaneously parasitize multiple hosts. This raises the possibility that any one host species may be continually exposed to DNA (either naked or within vectors like viruses and bacteria) from multiple, unrelated species. It will not be surprising if introgressions of many genes are discovered because of the long and continuing history of plant–plant parasitism.

2.3.3 Parasitic and Pathogenic Animals

The massive transfer of DNA to eukaryotes is widely believed to have happened when the different kinds of symbionts or parasites, such as those that were the free-living relatives of the mitochondria and chloroplast, fused with the common ancestor of eukaryotes. The transfer of DNA from organelles to the nucleus may have been a passive event, such as happens when invasive bacteria die inside animal cells, or an active process as indicated in laboratory studies following transgene transfer from yeast mitochondria to the nucleus.

Other intracellular parasites may transfer DNA to the host nucleus. Human, rabbit, and bird cells acquire the DNA from the parasite *Trypanosoma cruzi*, a protozoan that can cross the human placenta and transform the germline of animals.

The DNA donor *T. cruzi* is an example of a parasite/pathogen delivered to its human host by a biting insect vector. Other biting insects may be direct vectors of genes. The egg predatory mite *Proctolaelaps regalis* is

the most likely vector of P transposable elements to the Drosophilids. The mite does not have the element in its genome, so it may have introduced it into the germline of Drosophila after feeding on a carrier.

2.4 Hybridization

While hybridization is the outcome of sexual reproduction, it can also double as a vector for HGT. Even infertile hybrids can serve to promote HGT. For example, some hybrid plants convert to apomixis, an asexual mode of reproduction, increasing the chances of introgression. Infertile asexual hybrids or hybrids with low fertility may also produce offspring in time. When hybridization reduces fertility by creating two or more unmatched chromosomes, in time these chromosomes can duplicate, and thus pair. Once all the chromosomes again are partnered in meiosis, the hybrid can return to sexual modes of reproduction. Meanwhile, alien genes and transposable elements carried by alien chromosomes, whether or not the entire chromosomes should ever introgress, can transfer into the original genome.

The impact that mixing of previously separated populations has on the likelihood of interspecies gene transfer and introgression, should not be underestimated. Although there is an immediate increase in the chance of hybridization from breaking geographical boundaries between, say, a native species and an introduced one, the chances of more widespread gene transfer appear to be markedly increased if the introduced alien also acts as a bridging species, that is, a species that can link, through hybridization, two or more species that previously were genetically, geographically, morphologically, or physiologically separated to an extent that prevented hybridization.

These "hybrid-bridge" species have the potential to effectively act as multipliers of introgression, as they not only permit gene transfer between themselves and a resident native but also between multiple native and/or introduced species that were previously separated so much that the hybridization event was unlikely (Fig. 2). Simple examples of "hybrid-bridge" species can be envisaged: for example, a species with a flowering time that is intermediate between two resident species could allow gene flow between species previously separated by flowering morphology; likewise, a species with a broad habitat tolerance could link species from habitats that were previously geographically separated. In the face of current discoveries of gene introgressions through hybridization, it seems only a matter of time before an example of long-distance HGT driven by hybrid-bridge species is recognized.

3 Different Ways to Observe HGT

The science of horizontal gene transfer is well represented by researchers using two different but compatible approaches. The newest approach is to use bioinformatic descriptions of putative alien genes or tracts of alien DNA sequence to infer how and how often HGT happens, and what the consequences are to the recombinant genomes. This approach has the strength of looking over many different species simultaneously because of the accumulation of complete genome sequences and the rapid nature of *in silico* analysis. The bioinformatics approach of

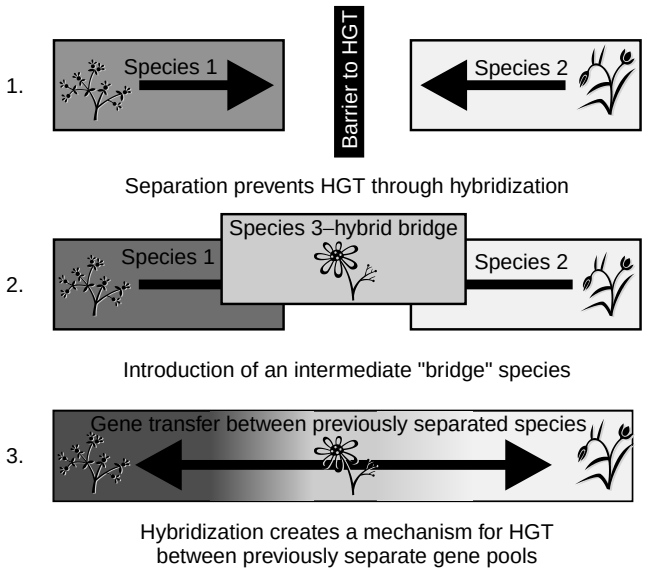

Fig. 2 Hybridization provides a mechanism for HGT between differentiated and separate multicellular eukaryotes. Symbols: rectangles represent populations of species in which HGT is confined; arrows represent the potential flow if genetic material in HGT. Removal of historical barriers to HGT, such as geographical separation, may lead to even greater potential for HGT. In this example, (1) two species may traditionally be unable to hybridize because of barriers such as habitat separation or differences in flowering morphology or physiology. (2) The introduction of a species that can hybridize with both Species 1 and 2 could bridge this separation. (3) The consequences of this ''hybrid bridge'' would mean that HGT, which previously could not occur between Species 1 and 2, would now be possible.

HGT, however, is limited by how well its techniques identify true HGT events, how accurate its predictions of gene function are, and by the very necessary fact that it is concentrating exclusively on DNA that has transferred, converted into vertically inherited material, and introgressed, that is, outcomes of HGT. Moreover, each event in the trillions of cells of a single large multicellular organism, or among the recipients of species that number 10^{20} individuals or more, is extremely easy to overlook. Below, the various bioinformatics tools for identifying HGT will be discussed.

The second approach is to observe HGT. Those observations can include biochemical descriptions of gene transfer pathways to eco-evolutionary experiments that measure how genes sort between horizontal and vertical mechanisms of inheritance. The experimental approach has the strength of using empirical biology to discover phenomena not immediately apparent from DNA sequences. It can in addition discover epigenes and other important molecules that may also transfer horizontally, and, importantly, it has the comparative power required to test

hypotheses about HGT processes. This approach is limited by the generation time of experimental organisms, the infrastructure needed to maintain different kinds of organisms, the realism of laboratory recreations of ecosystems in which gene transfers may occur, and the very limited capacity to physically sample enough biological material in order to detect HGT. Below, some novel process experiments and their relevance to understanding the ecology of mobile genes will be presented.

3.1 Descriptive Bioinformatics

Bioinformatics has produced compelling evidence from sequenced genomes of past gene transfers. Those transfer events that are detectable using bioinformatics are so because they are either

- recent;
- large fragments of DNA; or
- small but extremely different from the genome as a whole and out of place in the organismal phylogeny, in particular.

The introduction of new genes into a genome can create an anomalous phylogenetic tree, incongruent with other gene-based trees or trees built from morphological features. For example, the genes for the various aminoacyl tRNA synthetases, aaRS, create different phylogenetic trees in bacteria. This appears to result from different aaRS genes having been acquired by bacteria from various eukaryotes at different times. A particularly striking example is provided by ProRS, where a eukaryotic version is found in*Mycobacterium leprae* but not in the close relative *Mycobacterium tuberculosis*. In this case, the corresponding prokaryotic ProRS was lost from *M. leprae*; in some examples of aaRS transfer, both the prokaryotic and eukaryotic versions are maintained.

Evidence for HGT based on analyses that identify candidate genes that violate a phylogenetic order is prone to false-positive errors due, for example, to long-branch attraction. The most pervasive source of false-negative and false-positive identifications of alien genes comes from the initial construction of the tree based on putative orthologous genes. These genes diverged when the species diverged and represent the divergence of the two organisms from a common ancestor. However, they can be difficult to distinguish from paralogous genes created when a gene duplicates creating two alleles, which can then change independently over time. These reflect only the history of the genes within, instead of between, lineages.

Identifying which of multiple paralogous genes is a true orthologous gene is sometimes done by assuming that the most similar of the paralogous genes in the different species are the orthologous genes. However, accepting this *a priori* assumption can render the comparison redundant. Moreover, that test can never be independently verified because, no matter how close the sequences match, the possibility that one species lost the true orthologous gene since the speciation event can never be excluded. Indeed, the genes may be orthologous but because the gene was transferred horizontally, the comparison misrepresents the ancestry of the organism. Mistakes have been made by the inadvertent comparisons of paralogous genes.

Recent transfers may be indicated by a significant divergence in the composition of the DNA from the parameters expected on the basis of average characteristics of the genome. A horizontally transferred

gene preserved in a vertically reproducing lineage may be identified by a significant deviation from the average G + C content of the genome or from the normal codon bias of the host. Both of these indicators are quantitative. Thus, they are useful if large tracts of sequences with one or more large deviations from accepted norms are being analyzed.

The origin of sequences becomes increasingly difficult to determine the shorter they become or the closer to accepted ranges they appear. Shorter sequences may be the usual consequence of HGT that involves integration of alien genes with less sequence similarity to the recipient genome.

The outcome of many described HGT events is insertion of difficult-to-spot subgene domains. These nucleotide insertions, deletions, or replacements create "mosaic" genes that may have altered biochemical properties, for example, genes making bacteria resistant to antibiotics. Mosaics are caused by biochemical barriers, primarily MMR, that incompletely remove DNA during recombination.

MMR can saturate under stress or falter through mutation, promoting interspecies HGT ~10 000-fold and creating "mutators." Depending on the proficiency and mechanism of MMR in an organism, strands of DNA from different species may initially be paired, with the invading strand subsequently degraded. Should some mispairs escape repair, or if the MMR machinery became saturated, stretches of the recipient genome could be "massaged" into a closer match with the donor DNA over short intervals.

Recombination frequencies can vary significantly depending on the activity of MMR, resulting in mosaic genes composed of sequences from two highly divergent genes. In fact, a mutation in bacteria is 50 times more likely to have occurred by recombination than by DNA damage or replication errors. Since mutations are often revealed by very small changes in DNA sequence, it is normally impossible to distinguish between their origin as a polymerase or a mismatch-repair error. Mosaic genes can serve as molecular versions of "hybrid-bridges" (Sect. 2.4), facilitating higher frequency recombination between species that find sequence similarity in some parts of the mosaic gene.

The added difficulty of identifying genes that transfer frequently is that they may not mutate at the same rate as genes that reproduce vertically. The genomes of horizontally mobile vectors tend to be fluid. Over time, certain vectors may be replaced by related vectors carrying different genes, such as plasmids with an expanding repertoire of antibiotic resistance genes. Tracing their ancestry by sequence structure allows relationships to be determined only over very short periods of time. For example, the structure of infectious retroviruses can change at 10^4 to 10^6 times the rate of other genes and defective retroviruses that are reproducing in synchrony with the host. The phylogeny of these viruses is confined to tracing the residues of defective viruses trapped in chromosomes of organisms or monitoring divergences only on decade timescales.

Determining ancestry of organisms without relying upon nucleic acid or protein sequences is possible, but much more difficult. In the cases of ribonuclease H and the A8 subunit of mitochondria, ancestry was inferred from three-dimensional structures or other biophysical characteristics because the primary nucleic or amino acid sequences had lost such information.

3.2 Experimentation

Within the experimentalist approach are two mutually supportive schools. The first school is characterized by a quest to understand the biochemistry and pathways behind observable gene transfers. The work of that group is well represented in the literature on conjugation, transduction, and transformation. The second school elevates HGT to an evolutionary process – like sex, mutagenesis, or symbiosis – that can be characterized by how it evolved and to what selective forces it responds to.

3.2.1 Seminal Experimentalist Approaches

The origin of the experimentalist approach for studying HGT can be traced to the first description of mechanisms, including the pioneering work of the 1940s and 1950s with phage, transformation, and conjugation. Those studies were overtly focused on the biology of the organism or the vector, rather than on how HGT shaped them. For example, Luria and Delbrück's famous experiment measuring phage resistance had little to do with understanding how viruses themselves evolved, but it required a comprehensive understanding of the life cycle of the phage in order to demonstrate an evolutionary truth about whether mutations follow or precede selection. Avery et al. used natural competence in *pneumoccocus* to all but prove that DNA was the source of the genes behind the virulence traits they studied, but they used HGT to identify the transforming principle rather than using the transforming principle in a study of HGT. Hayes' use of the antibiotic streptomycin revealed the unidirectional transfer of genes from a donor to recipient by conjugation, but it would take almost 40 years before his trick was resurrected to begin to understand why horizontally mobile plasmids would come to carry antibiotic resistance genes.

In the late 1960s, research groups with a specific focus on HGT emerged. Their pioneering work provided the first hints that both proteins and DNA and RNA could physically transfer between species. The team of Maurice Stroun and Philippe Anker were arguably the first to detect hints that gene transfer from the plant pathogen *A. tumefaciens* was behind the crown gall tumors it caused. This work would take another 10 years, and a suite of new techniques, to be confirmed. Stroun and Anker also produced compelling evidence with the techniques of their time for the horizontal transfer of proteins, a theme that has gained prominence with the demonstration of the interspecies infectivity of prions and the description of the Type IV DNA and protein secretion systems that bacteria use for both conjugation and the transfer of proteins to eukaryotic hosts. What made the work of Stroun and Anker different from that of their contemporaries was that while they were inspired by the biology of the organisms they chose, their experiments were process-oriented, designed to detect gene transfer rather than to focus on the biology of particular organisms.

3.2.2 Modern Process Approaches

Research that focuses on the process of HGT is becoming more prominent, especially as the threat of gene transfer from transgenic crops and antibiotic-resistant bacteria command the attention of the scientific community. Past descriptions of gene transfers and full descriptions of biochemical mechanisms, when available, have proved of limited use in predicting

when, where, and how often gene transfer might create an unwelcome biological outcome. In this section, two contemporary models that promise predictive power are contrasted.

Why and when are some genes reproducing horizontally while other genes, or at other times these same genes, are reproducing vertically? Two different models have been constructed in answer to this question. The first model, called *the complexity hypothesis*, argues that some genes are more mobile than others simply because of the complexity of the processes in which they are involved. Those genes whose products are closely reliant on, or associated with, many other molecular partners will simply not be functional if removed from their current context. This model is supported by bioinformatic analysis of HGT events in sequenced genomes. The second model, termed *the competition model*, argues instead that the two forms of reproduction (HGT and VGT) favor different genes in different environments. This model suggests that genes transit to horizontal reproduction when that mode of reproduction outpaces alleles of the same gene reproducing vertically. This model is supported by laboratory evolution experiments.

The complexity hypothesis posits that operational genes are more mobile than informational genes. Informational genes are those with products that are associated within large complexes and may be involved in such biochemical activities as transcription, translation, and replication. Operational genes, in contrast, may work alone or in very simple assemblies and thus more easily function after transfer. From six complete prokaryotic genomes that were analyzed for 132 orthologous genes, the dual contentions that gene transfer was a continuous process and that those genes most likely to introgress were operational were supported. Again, this analysis does not say how frequently genes are transferred, but how frequently different types of transferred genes transit into stable alien genes in new species.

The competition model differs from the complexity hypothesis for predicting the attributes of genes that make them transit to horizontal reproduction. It is based on the view that the current function of a protein is not sufficient to predict future roles in different environments. These roles in different environments, rather than in their cellular context alone, determine which genes transit from vertical to horizontal and back to vertical reproductive strategies.

The competition model is not at odds with the complexity hypothesis because it would accept that operational genes may more frequently, but not exclusively, enjoy a differential reproduction as mobile genes. Consistent with this expectation, there is bioinformatics evidence of informational gene transfers and even contemporary discoveries of plasmid vectors of informational genes. Interestingly, the mobility of these genes correlates with the relatively recent introduction of modern antibiotics by humans. In the case of the ribosomal S14 protein, there is compelling evidence that transfer and introgression may be facilitated by antibiotic pressure selecting for alleles of the *rps14* gene that could confer resistance. Likewise, the eukaryotic-type IleRS has been discovered on plasmids that deliver it to bacteria susceptible to the antibiotic mupirocin. Within the expectations of the competition model, these informational genes can have a role in some environments that makes them transit from a vertical to a horizontal reproductive strategy. The competition model goes further to explain how the

genes get mobilized, not just how transferred genes introgress.

A new study comparing the complete genomes of 116 prokaryotes supports the operational–informational division that is central to the complexity hypothesis. The huge amount of sequence information that makes up this study also revealed a second class of operational genes that are as immobile as the information genes. It also showed that among the genes most commonly found to be mobile were predominantly those associated with plasmids and phage, cell surface proteins, DNA binding (particularly restriction-modification), and pathogenicity. These categories include genes that are frequently classed as PSK, or postsegregational killing genes.

The distinction of PSK genes in this new study also provides tacit support for the competition model: PSK genes are defined as any combination of genes that simultaneously produce and suppress a toxic effect on the host. PSK genes attack other infectious elements, for example, competing plasmids and viruses, but do not necessarily increase the absolute rate of reproduction of the vector or the host. PSK genes have been found to be a significant disadvantage to vectors that are forced to reproduce at the rate of cell division. Only when PSK is coupled to HGT do the vectors bearing PSK genes demonstrate competitiveness with vectors that did not have such genes.

The PSK method of attack is to kill cellular offspring that lose the genes. If a bacterium is infected by an invading virus or an incompatible plasmid, the PSK genes kill the bacterium along with the invading horizontally mobile element. In doing so, the PSK-bearing vector does not benefit with an immediate increase in reproduction, but it does eliminate a competitor.

Antibiotic resistance, novel virulence determinants, and restriction-modification genes are likely to be PSK genes and thus have evolved in the same way. The structural and functional attributes of PSK are independent of the benefits such genes might provide the organismal host in the presence of antibiotics or other lethal agents. The advantages PSK genes confer upon the HGT vector are even more important than the cost to organismal hosts for carrying the genes.

PSK genes evolve by HGT, at least some of the time. However, these phenotypes are only indirectly related to how PSK genes, like antibiotic resistance, converted from chromosomal to horizontally transferred genes. Organisms with these genes may in time have found some use for them, explaining introgression, but that is not an explanation for how the genes came to evolve on horizontally mobile elements.

The competition model predicts that genes capable of causing cell death will, under the right circumstances, transit to horizontal reproduction. It is thus noted with some irony that classic PSK systems, also called *suicide genes*, have been recommended by a prominent panel of the US National Research Council as promising bioconfinement tools for application to genetically engineered microbes.

4
Barriers to HGT

If HGT were so common and frequent, why is there not more evidence of it? The definitive answer to this question is some ways off, because, as discussed above, the preponderance of searches for it so far

have been designed to detect introgression into organismal genomes. Toward the end of the last century, however, several groups independently started asking this question. Their motivation was not to discount HGT, but to yield biologically interesting hypotheses about the biological barriers to either transfer *or* introgression (Fig. 1).

There is no doubt that there is a raft of situations, circumstances, and processes that prevent or at least reduce the rate of HGT. These are referred to here as *transfer, inheritance, and introgression barriers.*

4.1 Transfer and Inheritance

Barriers to gene transfer and inheritance include those that prevent geographical access to genes, entry of genes into the cell or organism, stability of DNA (for example, restriction enzymes) and stability of integration (recombination and mismatch repair) or an inherent compatibility between important cis-acting replication sequences and the enzymes of replication.

Geographical barriers spatially separate species to an extent that prevents exposure to their respective genes. Entry barriers include impenetrable cell walls that prevent transformation or transduction, and differences in form (like flower shape) that prevent hybridization. Virus infectious ranges may be determined by the presence or absence of a specific receptor. The genetic material, whether it is protein, RNA or DNA, may be enzymatically degraded upon entry into a cell, or specially sequestered and denatured or depolymerized. Immediately following transfer and stabilization, the gene must replicate or, in the case of nucleic acids, physically recombine with a chromosome or other cellular replicon.

4.2 Introgression

Another layer of barriers may prevent introgression. Developmental barriers specifically deny incoming genes access to the germline. Genetic barriers prevent recombination (Sect. 2.1) or they prevent chromosome pairing during meiosis (Sect. 2.4). Selection barriers block the expansion of the recombinant genotype within the population or species. This can be due to the gene having no or a negative effect on host fitness or due to expression barriers.

Expression barriers arise from incompatibility of an alien gene or polypeptide product with the host expression biochemistry, roughly that referred to as the *central dogma reactions*. The central dogma reactions are well understood as individual processes and are summarized but are treated in depth elsewhere.

Failure of cis-acting sequences of alien genes or their transcripts or products, to be properly recognized or processed by any of the reactions between transcription and the proper placement of a mature protein may prevent the gene from providing a function that could make its new host differentially competitive. Regulatory sequences on alien genes may not effectively recruit specific transcription factors to bind to promoters, terminators, and enhancers. The alien gene may produce an mRNA that is spliced differently or not at all, or which requires special editing (Fig. 3). Translated polypeptides may not fold properly in the recipient or may not sort to the correct subcellular location.

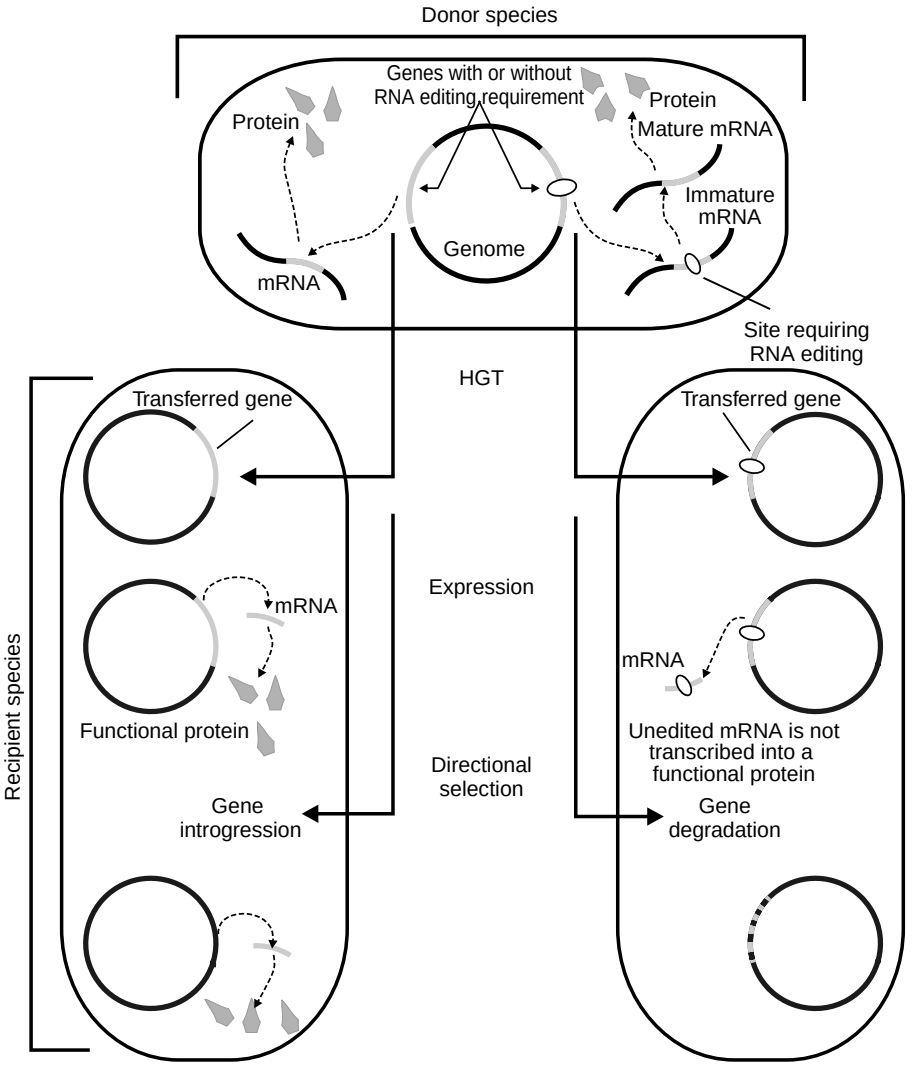

Fig. 3 RNA editing as an example of a molecular barrier to introgression. Symbols: rounded rectangles represent distinct donor or recipient species; circles represent genomes; arrows represent the linear sequence of events from HGT from donor to recipient genomes to the processes of gene expression and finally either intergression or gene degradation; protein mRNA strands and proteins are labeled as such. Some genes require donor-specific posttranscriptional RNA editing. Failure to edit transcripts could uncouple HGT from introgression if recipients do not have a similar RNA editing pathway. Genes without an RNA editing requirement are more likely to properly translate and may therefore introgress through directional selection.

5
Final Thoughts

The ultimate limitation of HGT is not transfer, but our ability to detect it. Gene transfers between closely related organisms may not be detected in genome sequences. While even small changes in existing genes may be important for function and introgression, they may be too small to distinguish from the noise of mutation through vertical descent. Small changes may arise by gene transfer and a combination of mutational amelioration and recombination, but these are almost always credited to repaired DNA damage and polymerase errors.

There are examples of genes that have overcome all the barriers to HGT described in Sect. 4. Of these barriers, only those that prevent transfer are fundamental to the evolution of all mobile genes. Barriers to inheritance and introgression only affect genes that tie their fates once again to the fate of an organism.

Acknowledgment

Jack A. Heinemann thanks the University of Canterbury for support on U6570. Ralph A. Bungard acknowledges support from the Marsden Fund of New Zealand.

See also Genetics, Molecular Basis of.

Bibliography

Books and Reviews

Brochier, C., Philippe, H., Moreira, D. (2000) The evolutionary history of ribosomal protein RpS14: horizontal gene transfer at the heart of the ribosome, *Trends Genet.* **16**, 529–533.

Bungard, R.A. (2004) Photosynthetic evolution in parasitic plants: insight from the chloroplast genome, *Bioessays* **26**, 235–247.

Christie, P.J., Vogel, J.P. (2000) Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells, *Trends Microbiol.* **8**, 354–360.

Cogoni, C., Macino, G. (2000) Post-transcriptional gene silencing across kingdoms, *Curr. Opin. Genet. Dev.* **10**, 638–643.

Doolittle, W.F. (1999) Lateral genomics, *Trends Biochem. Sci.* **24**, M5–M8.

Eberhard, W.G. (1989) Why do bacterial plasmids carry some genes and not others? *Plasmid* **21**, 167–174.

Eberhard, W.G. (1990) Evolution in bacterial plasmids and levels of selection, *Quart. Rev. Biol.* **65**, 3–18.

Ellstrand, N.C. (2003) Current knowledge of gene flow in plants: implications for transgene flow, *Phil. Tran. Roy. Soc. Lon. B* **358**, 1163–1170.

Gelvin, S.B. (2003) Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool, *Microbiol. Mol. Biol. Rev.* **67**, 16–37.

Grohmann, E., Muth, G., Espinosa, M. (2003) Conjugative plasmid transfer in gram-positive bacteria, *Microbiol. Mol. Biol. Rev.* **67**, 277–301.

Heinemann, J.A. (1991) Genetics of gene transfer between species, *Trends Genet.* **7**, 181–185.

Heinemann, J.A. (1999) How antibiotics cause antibiotic resistance, *Drug Discov. Today* **4**, 72–79.

Heinemann, J.A., Roughan, P.D. (2000) New hypotheses on the material nature of horizontally transferred genes, *Ann. N. Y. Acad. Sci.* **906**, 169–186.

Heinemann, J.A., Silby, M.W. (2003) Horizontal Gene Transfer and the Selection of Antibiotic Resistance, in: Amábile-Cuevas, C.F. (Ed.) *Multiple Drug Resistant Bacteria*, Horizon Scientific Press, Wymondham, Norfolk, pp. 161–178.

Kurland, C.G., Canback, B., Berg, O.G. (2003) Horizontal gene transfer: A critical view, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9658–9662.

Lawrence, J.G., Ochman, H. (2002) Reconciling the many faces of lateral gene transfer, *Trends Microbiol.* **10**, 1–4.

Martin, W. (2003) Gene transfer from organelles to the nucleus: frequent and in big chunks, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8612–8614.

Ragan, M.A. (2002) Reconciling the many faces of lateral gene transfer. Response from Ragan, *Trends Microbiol.* **10**, 4.

Schaecter, M. (Ed.) (2004) *The Desk Encyclopedia of Microbiology*, Elsevier Academic Press, San Diego, CA and London.

Souza, V., Eguiarte, L.E. (1997) Bacteria gone native vs. bacteria gone awry?: Plasmid transfer and bacterial evolution, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5501–5503.

Stewart Jr., C.N., Halfhill, M.D., Warwick, S.I. (2003) Transgene introgression from genetically modified crops to their wild relatives, *Nat. Rev. Genet.* **4**, 806–817.

Syvanen, M. Kado, C.I. (Eds.) (2002) *Horizontal Gene Transfer*, Academic Press, San Diego, CA.

Woolhouse, M.E.J., Taylor, L.H., Haydon, D.T. (2001) Population biology of multihost pathogens, *Science* **292**, 1109–1112.

Zhaxybayeva, O., Lapierre, P., Gogarten, J.P. (2004) Genome mosaicism and organismal lineages, *Trends Genet.* **20**, 254–260.

Primary Literature

Adler, L.S. (2000) Alkaloid uptake increases fitness in a hemiparasitic plant via reduced herbivory and increased Pollination, *Am. Nat.* **156**, 92–99.

Arber, W. (2003) Elements for a theory of molecular evolution, *Gene* **317**, 3–11.

Avery, O.T., MacLeod, C.M., McCarty, M. (1944) Studies on the chemical nature of the substance inducing transformation of Pneumococcal types, *J. Exp. Med.* **79**, 137–158.

Balsalobre, L., Ferrandiz, M.J., Linares, J., Tubau, F., de la Campa, A.G. (2003) Viridans group Streptococci are donors in horizontal transfer of topoisomerase IV genes to Streptococcus pneumoniae, *Antimicrob. Agents Chemother.* **47**, 2072–2081.

Bernstein, H., Byerly, H.C., Hopf, F.A., Michod, R.E. (1985) Genetic damage, mutation, and the evolution of sex, *Science* **229**, 1277–1281.

Carrera, M.R.A., Kaufmann, G.F., Mee, J.M., Meijler, M.M., Koob, G.F., Janda, K.D. (2004) Treating cocaine addiction with viruses, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10416–10421.

Citovsky, V., Zambryski, P. (1993) Transport of nucleic acids through membrane channels: Snaking through small holes, *Annu. Rev. Microbiol.* **47**, 167–197.

Claverys, J.-P., Martin, B. (2003) Bacterial 'competence' genes: signatures of active transformation, or only remnants? *Trends Microbiol.* **11**, 161–165.

Claverys, J.-P., Prudhomme, M., Mortier-Barriere, I., Martin, B. (2000) Adaptation to the environment: Streptococcus pneumoniae, a paradigm for recombination-mediated genetic plasticity? *Mol. Microbiol.* **35**, 251–259.

Condit, R. (1990) The evolution of transposable elements: conditions for establishment in bacterial populations, *Evolution* **44**, 347–359.

Cooper, T.F., Heinemann, J.A. (2000) Postsegregational killing does not increase plasmid stability but acts to mediate the exclusion of competing plasmids, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12543–12648.

Cupples, C.G., Cabrera, M., Cruz, C., Miller, J.H. (1990) A set of lacZ mutations in Escherichia coli that allow rapid detection of specific frameshift mutations, *Genetics* **125**, 275–280.

Davis, C.C., Wurdack, K.J. (2004) Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales, *Science* **305**, 676–678.

Denamur, E., Lecointre, G., Darlu, P., Tenaillon, O., Acquviva, C., Sayada, C., Sunjevaric, I., Rothstein, R., Elion, J., Taddei, F., Radman, M., Matic, I. (2000) Evolutionary implications of the frequent horizontal transfer of mismatch repair genes, *Cell* **103**, 711–721.

Doerfler, W., Schubbert, R., Heller, H., Kammer, C., Hilger-Eversheim, K., Knoblauch, M., Remus, R. (1997) Integration of foreign DNA and its consequences in mammalian systems, *Trends Biotechnol.* **15**, 297–301.

Doucet-Populaire, F., Trieu-Cuot, P., Andremont, A., Courvalin, P. (1992) Conjugal transfer of plasmid DNA from Enterococcus faecalis to Escherichia coli in digestive tracts of gnotobiotic mice, *Antimicrob. Agents Chemother.* **36**, 502–504.

Doucet-Populaire, F., Trieu-Cuot, P., Dosbaa, I., Andremont, A., Courvalin, P. (1991) Inducible transfer of conjugative transposon Tn*1545* from *Enterococcus faecalis* to *Listeria monocytogenes* in the digestive tracts of gnotobiotic mice, *Antimicrob. Agents Chemother.* **35**, 185–187.

Eckardt, N.A. (2001) A sense of self: the role of DNA sequence elimination in allopolyploidization, *Plant Cell* **13**, 1699–1704.

Evans, E., Alani, E. (2000) Roles for mismatch repair factors in regulating genetic recombination, *Mol. Cell. Biol.* **20**, 7839–7844.

Feil, E.J., Maynard Smith, J., Enright, M.C., Spratt, B.G. (2000) Estimating recombinational parameters in Streptococcus pneumoniae from multilocus sequence typing data, *Genetics* **154**, 1439–1450.

Ferguson, G.C., Heinemann, J.A., Kennedy, M.A. (2002) Gene Transfer between Salmonella enterica Serovar Typhimurium inside epithelial Cells, *J. Bacteriol.* **184**, 2235–2242.

Filee, J., Forterre, P., Laurent, J. (2003) The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies, *Res. Microbiol.* **154**, 237–243.

Funchain, P., Yeung, A., Stewart, J., Clendenin, W.M., Miller, J.H. (2001) Amplification of mutator cells in a population as a result of horizontal gene transfer, *J. Bacteriol.* **183**, 3737–3741.

Gerdes, K., Rasmussen, P.B., Molin, S. (1986) Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3116–3120.

Gibbs, M.J., Weiller, G.F. (1999) Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8022–8027.

Grillot-Courvalin, C., Goussard, S., Huetz, F., Ojcius, D.M., Courvalin, P. (1998) Functional gene transfer from intracellular bacteria to mammalian cells, *Nat. Biotechnol.* **16**, 862–866.

Hayes, W. (1952) Recombination in *Bact. coli K* 12: Unidirectional transfer of genetic material, *Nature* **169**, 118–119.

Haygood, R., Ives, A.R., Andow, D.A. (2003) Consequences of recurrent gene flow from crops to wild relatives, *Phil. Tran. Roy. Soc. Lon. B* **270**, 1879–1886.

Heinemann, J.A. (1999) Genetic evidence for protein transfer during bacterial conjugation, *Plasmid* **41**, 240–247.

Heinemann, J.A., Sprague Jr., G.F. (1989) Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast, *Nature* **340**, 205–209.

Heinemann, J.A., Traavik, T. (2004) Problems in monitoring horizontal gene transfer in field trials of transgenic plants, *Nat. Biotechnol.* **22**, 1105–1109.

Houck, M.A., Clark, J.B., Peterson, K.R., Kidwell, M.G. (1991) Possible horizontal transfer of Drosophila genes by the mite Proctolaelaps regalis, *Science* **253**, 1125–1128.

Hughes, A.L., Friedman, R. (2003) Genome-wide survey for genes horizontally transferred from cellular organisms to baculoviruses, *Mol. Biol. Evol.*, **Advance Access April 25, 2003**, http://mbe.oupjournals.org/cgi/reprint/msg107v101.

Humayun, M.Z. (1998) SOS and Mayday: multiple inducible mutagenic pathways in Escherichia coli, *Mol. Microbiol.* **30**, 905–910.

Jacobs, H.T. (1991) Structural similarities between a mitochondrially encoded polypeptide and a family of prokaryotic respiratory toxins involved in plasmid maintenance suggest a novel mechanism for the evolutionary maintenance of mitochondrial DNA, *J. Mol. Biol.* **32**, 333–339.

Jain, R., Rivera, M.C., Lake, J.A. (1999) Horizontal gene transfer among genomes: the complexity hypothesis, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3801–3806.

Ke, C., Tsai, J.H. (1985) Transmission of huanglungbin agent from citrus to periwinkle by dodder, *Annu. Rev. Phytopathol.* **75**, 1324–1324.

Keeling, P.J., Palmer, J.D. (2001) Lateral transfer at the gene and subgenic levels in the evolution of eukaryotic enolase, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10745–10750.

Kunik, T., Tzfira, T., Kapulnik, Y., Gafni, Y., Dingwall, C., Citovsky, V. (2001) Genetic transformation of HeLa cells by Agrobacterium, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1871–1876.

Luria, S.E., Delbrück, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance, *Genetics* **28**, 491–511.

Matic, I., Rayssiguier, C., Radman, M. (1995) Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species, *Cell* **80**, 507–515.

McClure, M.A. (2001) Evolution of the DUT gene: horizontal transfer between host and pathogen in all three domains of life, *Curr. Protein Pept. Sci.* **2**, 313–324.

Mihaylova, V.T., Bindra, R.S., Yuan, J., Campisi, D., Narayanan, L., Jensen, R., Giordano, F., Johnson, R.S., Rockwell, S., Glazer, P.M. (2003) Decreased expression of the DNA mismatch repair Gene Mlh1 under hypoxic stress in mammalian cells, *Mol. Cell. Biol.* **23**, 3265–3273.

Naito, T., Kusano, K., Kobayashi, I. (1995) Selfish behavior of restriction-modification systems, *Science* **267**, 897–899.

Nakamura, Y., Itoh, T., Matsuda, H., Gojobori, T. (2004) Biased biological functions of horizontally transferred genes in prokaryotic genomes, *Nat. Genet.* **36**, 760–766.

Nicolaisen, M. (2004) Transmission of branch-inducing phytoplasma PoiBI from poinsettia (Euphorbia pulcherrima) to crown-of-thorns (E-milii), *HortScience* **38**, 551–552.

Nitz, N., Gomes, C., de Cássia Rosa, A., D'Souza-Ault, M.R., Moreno, F., Lauria-Pires, L., Nascimento, R.J., Teixeira, A.R.L. (2004) Heritable integration of kDNA minicircle sequences from Trypanosoma cruzi into the avian genome: insights into human Chagas Disease, *Cell* **118**, 175–186.

O'Callaghan, D., Cazevieille, C., Allardet-Servent, A., Boschiroli, M.L., Bourg, G., Foulongne, V., Frutos, P., Kulakav, U., Ramuz, M. (1999) A homologue of the Agrobacterium tumefaciens VirB and Bordetella pertussis Ptl type IV secretion systems is essential for intracellular survival of Brucella suis, *Mol. Microbiol.* **33**, 1210–1220.

Rayssiguier, C., Thaler, D.S., Radman, M. (1989) The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants, *Nature* **342**, 396–401.

Resources, Bo.Aa.N., Agriculture, Bo., Sciences, Bo.L. (2004) *Biological Confinement of Genetically Engineered Organisms*, The National Academies Press, Washington, DC.

Robertson, H.M., Soto-Adames, F.N., Walden, K.K.O., Avancini, R.M.P., Lampe, D.J. (2002) The Mariner Transposons of Animals: Horizontally Mobile Jumping Genes, in: Syvanen, M., Kado, C.I. (Eds.) *Horizontal Gene Transfer*, Academic Press, San Diego, CA, pp. 173–185.

Shafer, K.S., Hanekamp, T., White, K.H., Thorsness, P.E. (1999) Mechanisms of mitochondrial DNA escape to the nucleus in the yeast Saccharomyces cerevisiae, *Curr. Genet.* **36**, 183–194.

Snow, A.A. (2002) Transgenic crops – why gene flow matters, *Nat. Biotechnol.* **20**, 542.

Stroun, M., Anker, P. (1971) Bacterial nucleic acid synthesis in plants following bacterial contact, *Mol. Gen. Genet.* **113**, 92–98.

Stroun, M., Anker, P. (1973) Transcription of spontaneously released bacterial deoxyribonucleic acid in frog auricles, *J. Bacteriol.* **114**, 114–120.

Stroun, M., Anker, P., Auderset, G. (1970) Natural release of nucleic acids from bacteria into plant cells, *Nature* **227**, 607–608.

Stroun, M., Anker, P., Maurice, P., Gahan, P. (1977) Circulating nucleic acids in higher organisms, *Int. Rev. Cytol.* **51**, 1–48.

Thomashow, M.F., Nutter, R., Postle, K., Chilton, M.-D., Blattner, F.R., Powell, A., Gordon, M.P., Nester, E.W. (1980) Recombination between higher plant DNA and the Ti plasmid of Agrobacterium tumefaciens, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6448–6452.

Tsen, S.-D., Fang, S.-S., Chen, M.-J., Chien, J.-Y., Lee, C.-C., Tsen, D.H.-L. (2002) Natural plasmid transformation in Escherichia coli, *J. Biomed. Sci.* **9**, 246–252.

Waters, V.L. (2001) Conjugation between bacterial and mammalian cells, *Nat. Genet.* **29**, 375–376.

Whitman, W.B., Coleman, D.C., Wiebe, W.J. (1998) Prokaryotes: the unseen majority, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6578–6583.

Yadav, N.S., Postle, K., Saiki, R.K., Thomashow, M.F., Chilton, M.-D. (1980) T-DNA of a crown gall teratoma is covalently joined to host plant DNA, *Nature* **287**, 458–461.

HPLC of Biological Macromolecules: *see* HPLC of Peptides and Proteins

HPLC of Peptides and Proteins

Tzong-Hsien Lee and Marie-Isabel Aguilar
Monash University, Clayton, Australia

1 Introduction 248

2 Theoretical Considerations 248

3 Reversed Phase Chromatography 249
3.1 Basic Principles 250
3.2 Factors Influencing Retention in RPC 252
3.2.1 Stationary Phases 252
3.2.2 Mobile Phases 255
3.2.3 Column Geometry 255
3.2.4 Operating Parameters 256

4 Hydrophobic Interaction Chromatography 260
4.1 Basic Principles 260
4.2 Factors Influencing Retention in HIC 261
4.2.1 Stationary Phase 261
4.2.2 Mobile Phase 262

5 Ion-exchange Chromatography 266
5.1 Basic Principles 266
5.2 Factors Influencing Retention in IEC 269
5.2.1 Stationary Phases 269
5.2.2 Mobile Phases 271
5.2.3 Operating Parameters 271
5.3 Peptide Separation in IEC 273

6 Size Exclusion Chromatography 276
6.1 Basic Principles 276
6.2 Experimental Conditions 276

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

7 Capillary Liquid Chromatography 278
7.1 General Considerations 278
7.2 Capillary RPC of Proteolytic Peptides for Proteomic Analysis 282
7.3 Capillary Ion-exchange Chromatography 284
7.4 Other LC Methods 284
7.5 Multidimensional Capillary Separations for Proteins 284
7.6 Other Approaches 286

8 LC-MS for Peptide/Protein Analysis 286
8.1 General Considerations 287
8.2 Capillary RPC-MS 288
8.3 LC/MS for Protein Analysis 289

9 Future Prospects 290

Bibliography 290
Books and Reviews 290
Primary Literature 291

Keywords

Capacity Factor
The parameter used in HPLC to measure the retention of an analyte.

Capillary Column
This term refers to a chromatographic column of "small" diameter and is used in both gas and high-performance liquid chromatography. In HPLC, the term is usually applied to columns with internal diameters of between 0.1 and 2 mm. The term *microbore column* is often used synonymously to describe these columns but is more correctly applied to columns.

Gradient Elution
The changing of HPLC mobile phase composition during the course of an analysis.

high-performance liquid chromatography (HPLC)
A technology in which compounds are partitioned between a stationary medium and a flowing liquid phase in a tubular format under high pressure, resulting in the separation of a mixture of components and the reproducible time of elution of each specific compound.

Ion Trap Mass Spectrometer
This device consists of two end-cap electrodes (entrance and exit) and a ring electrode. An ion trap mass spectrometer separates ions based on mass-to-charge ratio (m z^{-1}).

Once ions are introduced into the ion trap mass spectrometer, the radiofrequency (rf) amplitude is increased so that ions are sequentially ejected (by increasing mass) and detected.

Isocratic Elution
The use of a mobile phase of constant composition during the course of an analysis.

Mobile Phase
That part of a chromatographic system that causes the analyte to move from the point of injection to the detector – in HPLC, this is a liquid.

Peak Capacity
The number of peaks that can be separated within a defined separation or gradient time.

Peptide Mapping
The process of considering the amino acid sequence information from peptides obtained by enzyme digestion in an attempt to derive the (amino acid) sequence of the parent protein.

Resolution
A term that indicates the ability of a device/technique to separate/distinguish between closely related signals. In chromatography, it relates to the ability to separate compounds with similar retention characteristics, and in mass spectrometry it relates to the ability to separate ions of similar m/z ratios.

Stationary Phase
That part of the chromatographic system with which the analytes interact, over which the mobile phase flows.

HPLC is extremely versatile for the isolation of peptides and proteins from a wide variety of synthetic or biological sources. The complexity of the mixture to be chromatographed will depend on the nature of the source and the degree of preliminary clean up that can be performed. In the case of synthetic peptides, RPC is generally employed both for the initial analysis and the final large-scale purification. The isolation of proteins from a biological cocktail, however, often requires a combination of techniques to produce a homogenous sample. HPLC techniques are then introduced at the later stages following initial precipitation, clarification and preliminary separations using soft gel. Purification protocols therefore need to be tailored to the specific target molecule.

This chapter deals with the different HPLC techniques that are commonly employed for the analysis and purification of peptides and proteins. A brief overview of the theory of each mode of chromatography will be presented and then discussed

in terms of the parameters that control resolution and illustrated with relevant examples. The interested reader is also referred to a number of recent publications that provide a comprehensive theoretical and practical overview of this topic.

1 Introduction

The introduction of high-performance liquid chromatography (HPLC) to the analysis of peptides and proteins some 25 years ago revolutionized the biological sciences by enabling the rapid and sensitive analysis of peptide and protein structure in a way that was inconceivable 30 years ago.

Today, HPLC in its various modes has become the pivotal technique in the characterization of peptides and proteins and has therefore played a critical role in the development of peptide and protein-based pharmaceuticals. The extraordinary success of HPLC can be attributed to a number of factors. These include (1) the excellent resolution that can be achieved under a wide range of chromatographic conditions for very closely related molecules as well as structurally quite distinct molecules; (2) the experimental ease with which chromatographic selectivity can be manipulated through changes in mobile phase characteristics; (3) the generally high recoveries and hence high productivity and (4) the excellent reproducibility of repetitive separations carried out over a long period of time, which is due partly to the stability of the sorbent materials under a wide range of mobile phase conditions.

Reversed phase chromatography (RPC) is by far the most commonly used mode of separation for peptides, although ion-exchange (IEC) and size exclusion (SEC) chromatography also find application. The three-dimensional structure of proteins can be sensitive to the often harsh conditions employed in RPC, and as a consequence, RPC is employed less for the isolation of proteins where it is important to recover the protein in a biologically active form. IEC and SEC affinity chromatography are therefore the most commonly used modes for proteins, but RPC and hydrophobic interaction (HIC) chromatography are also employed.

2 Theoretical Considerations

An appreciation of the factors that control the resolution of peptides and proteins in interactive modes of chromatography can assist in the development and manipulation of separation protocols to obtain the desired separation. The capacity factor k' of a solute can be expressed in terms of the retention time t_r, through the relationship

$$k' = (t_r - t_o)/t_o \qquad (1)$$

where t_0 is the retention time of a nonretained solute. The practical significance of k' can be related to the selectivity parameter α, defined as the ratio of the capacity factors of two adjacent peaks as follows

$$\alpha = k'_i/k'_j \qquad (2)$$

which allows the definition of a chromatographic elution window in which retention times can be manipulated to maximize the separation of components within a mixture.

The optimization of high-resolution separations of peptides and proteins involves the separation of sample components through manipulation of both retention times and solute peak shape. The second factor that is involved in defining the quality of a separation is therefore the peak width σ_t. The degree of peak broadening is directly related to the efficiency of the column and can be expressed in terms of the number of theoretical plates, N, as follows

$$N = (t_r)^2/\sigma_t{}^2 \quad (3)$$

N can also be expressed in terms of the reduced plate height equivalent h, the column length L and the particle diameter of the stationary phase material d_p, as

$$N = hL/d_p \quad (4)$$

The resolution, R_s, between two components of a mixture therefore depends on both selectivity and bandwidth according to

$$R_s = 1/4\sqrt{N}(\alpha - 1)[1/(1 + k')] \quad (5)$$

This equation describes the relationship between the quality of a separation and the relative retention, selectivity and the bandwidth and also provides the formal basis upon which resolution can be manipulated to achieve a particular level of separation. Thus, when faced with an unsatisfactory separation, the aim is to improve resolution by one of three possible strategies: the first is to increase α, the second is to vary k' within a defined range normally $1 < k' < 10$, or thirdly to increase N, for example, by using very small particles in narrow bore columns.

The challenge facing the scientist who wishes to analyze and/or purify their peptide or protein sample is the selection of the initial separation conditions and subsequent optimization of the appropriate experimental parameters. This chapter provides an overview of the different techniques used for the analysis and purification of peptides and proteins and the experimental options available to achieve a high-resolution separation of a peptide or protein mixture.

3 Reversed Phase Chromatography

RPC is now an indispensable tool for the high-performance separation of complex mixtures of peptides and proteins and is used for both analytical and preparative applications. Analytical applications range from the assessment of purity of peptides following solid-phase peptide synthesis, to the analysis of tryptic maps of proteins. Preparative RPC is also used for the micropurification of protein fragments for sequencing to large-scale purification of synthetic peptides. An example of the high-resolution analysis of a tryptic digest of a protein is shown in Fig. 1. This figure, in which 150 peaks were identified, demonstrates the highly selective separation that can be achieved with enzymatic digests of proteins using RPC as part of the quality control or structure determination of a recombinant or natural protein. The chromatographic separation was obtained with a C2/C18 stationary phase packed in a column of dimensions 10 cm × 4.6 mm internal diameter (ID). Separated components can then be directly subjected to further analysis such as automated Edman N-terminal sequencing or electrospray mass spectrometry.

The purification of synthetic peptides usually involves an initial separation on an analytical scale to assess the complexity of the mixture followed by large-scale

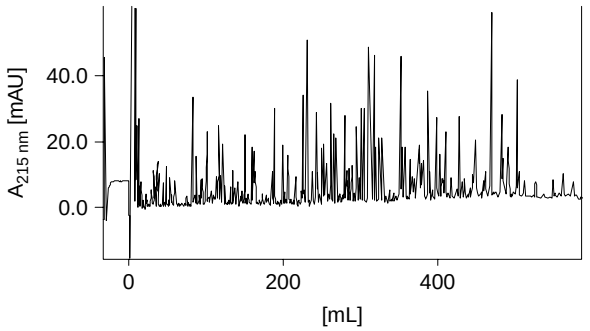

Fig. 1 High-resolution reversed phase chromatographic separation of a tryptic digest of a 165-kDa protein on a µRPC C2/C18 ST 4.6/100 column, (dimensions 10 cm × 4.6 mm ID, 3-µm particle size, 12-nm pore size). Eluent A: 0.065% trifluoroacetic acid (TFA) in water, eluent B: 0.050% TFA in 84% acetonitrile. Gradient elution was carried out with 0% B for 2 column volumes (CV), 0 to 50% for 392 CV (650 min); 50 to 100% B for 55 CV (91 min); 100% B for 10 CV (17 min), flow rate of 1 mL min^{-1} and detection was at 215 nm (Reprinted with permission from Amersham Bioscience, AKTA™ purifier Application Note No. 18-1119-53).

purification and collection of the target product. A sample of the purified material can then be subjected to RPC analysis under different elution conditions to check for purity.

The extensive use of RPC for the purification of peptides, small polypeptides with molecular weights up to 10 000 Da, and related compounds of pharmaceutical interest has not been replicated to the same extent for larger polypeptides (MW > 10 000 Da) and globular proteins. The combination of the traditionally used acidic buffering systems and the hydrophobicity of the *n*-alkylsilica supports, which can result in low mass yields or the loss of biological activity of larger polypeptides and proteins, have often discouraged practitioners from using RPC methods for large-scale protein separations. The loss of biological activity, the formation of multiple peaks for compositionally pure samples and poor yields of protein can all be attributed to the denaturation of protein solutes during the separation process using RPC. While these features detract from the use of RPC as a technique of choice in preparative purification protocols with proteins, these same characteristics can provide a unique opportunity to study protein folding and stability. Thus, the widespread practical application of RPC has been recently accompanied by a significant improvement in our understanding of the molecular basis of the retention process and its impact on conformational stability of both peptides and proteins. As a result, RPC can now also be used as a tool for the analysis of the interacting behavior of peptides and proteins at hydrophobic surfaces.

3.1 Basic Principles

The mechanism by which peptides and proteins are retained in RPC involves the hydrophobic expulsion of the solute

from the polar mobile phase followed by adsorption onto a nonpolar sorbent. Peptides and proteins are thus retained to different extents depending on their surface hydrophobicity, the eluotropicity of the mobile phase, and the nature of the hydrophobic ligands.

The physicochemical basis of RPC lies in the hydrophobic interaction between a peptide or protein and the hydrophobic matrix, and can be described in terms of the solvophobic theory where the isocratic retention factor can be expressed as

$$\ln k' = \log k_o - (N\Delta A_h + 4.836N^{1/3} \times [\kappa^e - 1]V^{2/3})\gamma/RT \qquad (6)$$

where N is Avogadro's number, ΔA_h is the hydrophobic contact area of the interacting solute and γ is the surface tension of the mobile phase. The parameter κ^e is the ratio of the energy required to create a cavity for a solvent molecule and the energy required to extend the planar surface of the liquid by the surface area of the solute molecule. Thus, simply stated, elution in RPC is achieved through a decrease in the microscopic surface tension associated with the solute–sorbent interface. Experimentally this is achieved through changes in the water content by variation in the mole fraction of organic solvent in RPC. However, peptides and proteins are generally separated by gradient elution conditions. Under these conditions, the experimentally observed retention data can be analyzed according to the linear solvent strength model (LSS) as follows

$$\log \overline{k} = \log k_o - S\overline{\varphi} \qquad (7)$$

where $\overline{k}$ is the solute median capacity factor, and $\overline{\varphi}$ is the corresponding organic mole fraction. The log k_0 value is the affinity of the solute for the sorbent in the absence of organic solvent. By analogy with the solvophobic equation above, the S-value is related to the hydrophobic contact region established between the solute and the sorbent surface. Both S and log k_0 can be derived from plots of log $\overline{k}$ versus $\overline{\varphi}$ and the LSS model provides the computational basis for the rational optimization of peptide and protein separations. An example of these plots is shown in Fig. 2 for a series of peptide analogs related to neuropeptide Y-[18–36], which differ in sequence only by the substitution of a single D-amino acid residue. These plots illustrate the ability

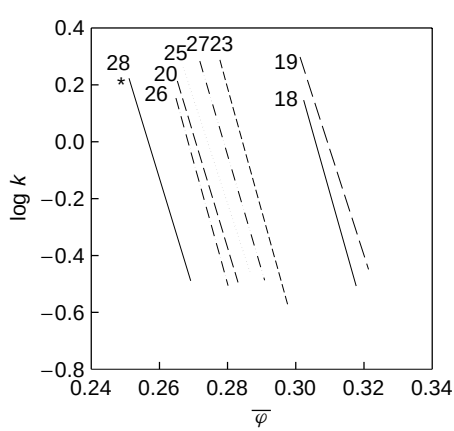

Fig. 2 Plots of log $\overline{k}$ versus $\overline{\varphi}$ for a series of D-analogs of neuropeptide Y-[18–36] separated on C18 silica with acetonitrile as the organic modifier at 4 °C. Column: Bakerbond (J T Baker, Phillipsburg, NJ) RP-C18, 25 cm × 4.6 mm ID, 5-µm particle size, 30-nm pore size. The retention plot for each analog is designated by the residue position of the D-amino acid substitution. The plots were derived from best-fit analysis to the data points (Reprinted with permission from Lazoura, E. et al. (1997) Conformational analysis of neuropeptide Y-[18–36] analogs. . ., *Biophys. J.* **72**, 238–46).

of RPC to resolve very small differences in peptide structure and also illustrate how RPC can act as a molecular probe of peptide surface topography. The data presented in these plots also provide a clear example of how to optimize separations, whereby maximal optimization will be achieved by selecting elution conditions where there is the greatest separation between the retention plots. The elution profiles of five of the NPY-[18–36] analogs are shown in Fig. 3 and demonstrate that the order of elution of the peaks corresponds to the relative position of the retention plots shown in Fig. 2 and also illustrates the corresponding degree of resolution that is obtained.

Peptides and proteins are retained in RPC through hydrophobic interactions. As depicted in Fig. 4(a), the contact region for small peptides involves the contribution from all or a large proportion of the peptide structure. As a consequence, the retention time for small peptides can be predicted on the basis of the amino acid composition through summation of the hydrophobicity coefficients for the constituent amino acid residues. In contrast, it has been well established that proteins interact with the chromatographic surface in an orientation-specific manner, in which their retention time is determined by the molecular composition of specific contact regions. For larger polypeptides and proteins that adopt a significant degree of secondary and tertiary structure, there is no correlation between retention times and the summated hydrophobicity coefficients of all constituent amino acid residues. In these cases, the chromatographic contact region comprises a small proportion of the total molecular surface as shown in Fig. 4(b, c).

3.2 Factors Influencing Retention in RPC

3.2.1 Stationary Phases

The most commonly employed experimental procedure for the RPC analysis of peptides and proteins generally involves the use of an octadecylsilica-based sorbent and a mobile phase. The chromatographic packing materials that are generally used are based on microparticulate porous silica, which allows the

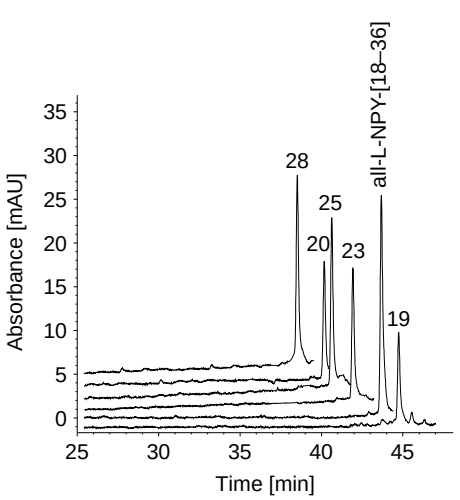

Fig. 3 Chromatograms showing the separation of the all-L-NPY-[18–36] peptide and NPY-[18–36]- analogs with D-substitutions at positions 19, 20, 23, 25, and 28 separated on a C18 silica (Bakerbond, J T Baker, Phillipsburg, NJ) RP-C18, 25 cm × 4.6 mm ID, 5-µm particle size, 30-nm pore size) with a 60-min gradient from 0 – to 50% acetonitrile at 1 mL min^{-1}. The separation was used to derive the retention plots shown in Fig.2 (Reprinted with permission from Lazoura, E. et al. (1997) The conformational analysis of NPY-[18–36] analogues of hydrophobic surfaces, *Biophys. J.* **72**, 238–46).

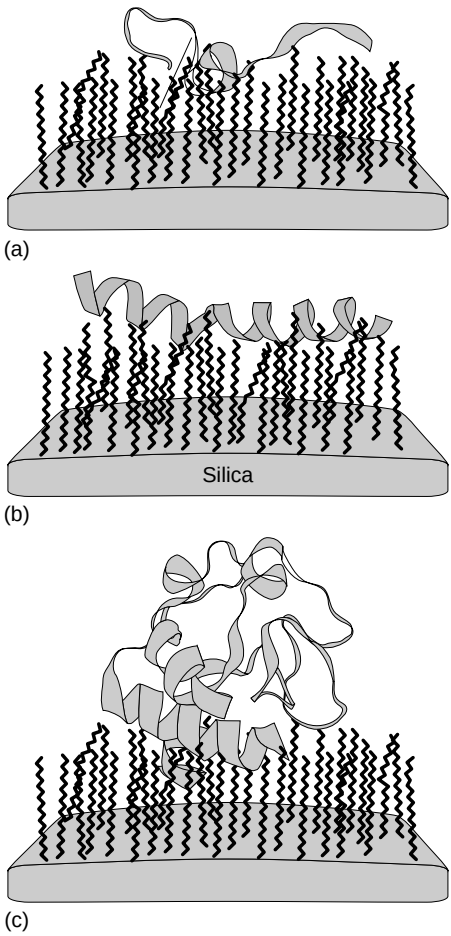

Fig. 4 Schematic representation of the binding of a polypeptide to an interactive chromatographic sorbent. The shaded areas on the polypeptide indicate the chromatographic contact region, which will be defined by the nature of the sorbent material. In reversed phase and hydrophobic interaction chromatography, the areas will be hydrophobic while in ion-exchange chromatography, the contact regions will represent the surface regions of highest charge density.

use of high linear-flow velocities resulting in favorable mass transfer properties and rapid analysis times. The silica is chemically modified by a derivatized silane bearing an *n*-alkyl hydrophobic ligand. The most commonly used ligand is *n*-octadecyl (C18), while *n*-butyl (C4) and *n*-octyl (C8) also find important application, and phenyl and cyanopropyl ligands can provide different selectivity. The process of chemical immobilization of the silica surface results in approximately half of the surface silanol group being modified. The sorbents are therefore generally subjected to further silanization with a small reactive silane to produce an end-capped packing material. The type of *n*-alkyl ligand significantly influences the retention of peptides and proteins and can therefore be used to manipulate the selectivity of peptide and protein separations. While the detailed molecular basis of the effect of ligand structure is not fully understood, a number of factors including the relative hydrophobicity and ligand chain length, flexibility and the degree of exposure of surface silanols all play a role in the retention process. An example of the effect of chain length on peptide separations can be seen in Fig. 5. It can be seen that the peaks labeled T_3 and T_{13} are fully resolved on the C4 packing but cannot be separated on the C18 material. In contrast, the peptides T_5 and T_{18} are unresolved on the C4 column but fully resolved on the C18 material. In addition to the effect on peptide selectivity, the choice of ligand type can also influence protein recovery and conformational integrity of protein samples. Generally, higher protein recoveries are obtained with the shorter and less hydrophobic *n*-butyl ligands. However, proteins have also been obtained in high yield with *n*-octadecyl silica.

The denaturation of proteins by RPC sorbents can also be controlled by using silica that has been coated with

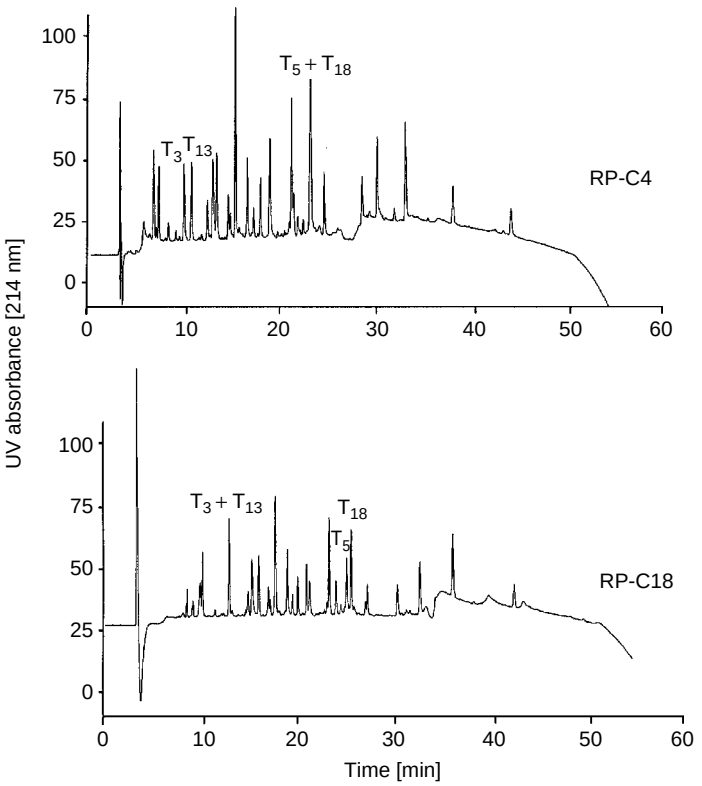

Fig. 5 The influence of *n*-alkyl chain length on the separation of tryptic peptides derived from porcine growth hormone. Top: Bakerbond (J T Baker, Phillipsburg, NJ) RP-C4, 25 cm × 4.6 mm ID, 5 µm-particle size, 30-nm pore size. Bottom: Bakerbond (J T Baker, Phillipsburg, NJ) RP-C18, 25 cm × 4.6 mm ID, 5-µm particle size, 30-nm pore size. Conditions, linear gradient from 0 to 90% acetonitrile with 0.1% TFA over 60 min, flow rate of 1 mL min^{-1}, 25 °C (Reprinted with permission from Aguilar, M-I., Hearn, M.T.W. (1996) High resolution reversed phase high performance liquid chromatography of peptides and proteins, *Methods Enzymol.* **270**, 3–26).

polymethacrylate-based polymers, which provide a series of sorbents with varying surface hydrophobicity, and in which the surface silanol groups have been masked. Silica-based packings are also susceptible to cleavage at pH values greater than 7. This limitation can severely restrict the utility of these materials for separations that require basic pH conditions to effect resolution. In these cases, alternative stationary phases have been developed including cross-linked polystyrene divinylbenzene and porous zirconia, which are all stable to hydrolysis at alkaline pHs.

The geometry of the particle in terms of the particle diameter and pore size is also an important feature of the packing material. As predicted by Eq. 4, improved resolution can be achieved by decreasing the particle diameter d_p. The most commonly used range of particle diameters for analytical scale RPC is 3 to

5 μm. There are also examples of the use of nonporous particles of smaller diameter. For preparative scale separations, 10 to 20 μm particles are utilized. The pore size of RPC sorbents is also an important factor that must be considered. For peptides, the pore size generally ranges between 100 to 300 Å depending on the size of the peptides. Porous materials of ≥300 Å pore size are necessary for the separation of proteins, as the solute molecular diameter must be at least one-tenth the size of the pore diameter to avoid restricted diffusion of the solute and to allow the total surface area of the sorbent material to be accessible. The recent development of particles with 6000 to 8000 Å pores with a network of smaller pores of 500 to 1000 Å has also allowed very rapid protein separations to be achieved.

3.2.2 Mobile Phases

One of the most powerful characteristics of RPC is the ability to manipulate solute retention and resolution through changes in the composition of the mobile phase. In RPC, peptide and protein retention is due to multisite interactions with the ligands. The practical consequence of this is that high-resolution isocratic elution of peptides and proteins can rarely be achieved as the experimental window of solvent concentration required for their elution is very narrow. Mixtures of peptides and proteins are therefore routinely eluted by the application of a gradient of increasing organic solvent concentration. RPC is generally carried out with an acidic mobile phase, with trifluoroacetic acid (TFA), the most commonly used additive due to its volatility. Phosphoric acid, perchloric acid, formic acid, hydrochloric acid, acetic acid, and heptafluorobutyric acid have also been used. The effect of ion-pairing reagents on peptide separation is illustrated in Fig. 6 for a series of peptide standards separated on a C18 column. Alternative additives such as nonionic detergents can be used for the isolation of more hydrophobic proteins such as membrane proteins.

The three most commonly employed organic solvents are acetonitrile, methanol, and 2-propanol, which all exhibit high optical transparency in the detection wavelengths used for peptide and protein analysis. Acetonitrile provides the lowest viscosity solvent mixtures and 2-propanol is the strongest eluent. An example of the influence of organic solvent where changes in selectivity can be observed for a number of peptide peaks in the tryptic map is shown in Fig. 7. In addition to the eluotropic effects, the nature of the organic solvent can also influence the conformation of both peptides and proteins and will therefore have an additional effect on selectivity through changes in the structure of the hydrophobic contact region. In the case of proteins, this may also impact on the level of recovery of biologically active material.

3.2.3 Column Geometry

The desired level of efficiency and sample loading size determines the dimension of the column to be used. For small peptides and proteins, increased resolution will be obtained with increases in column length. Thus, for applications such as tryptic mapping, column lengths between 15 to 25 cm and internal diameter (ID) of 4.6 mm are generally employed. However, for larger proteins, low mass recovery and loss of biological activity may result with these columns due to irreversible binding and/or denaturation. In these cases, shorter columns of between 2 and 20 cm in length can be used. For preparative applications in the 1 to 500 mg scale, such as the purification of synthetic peptides, so-called

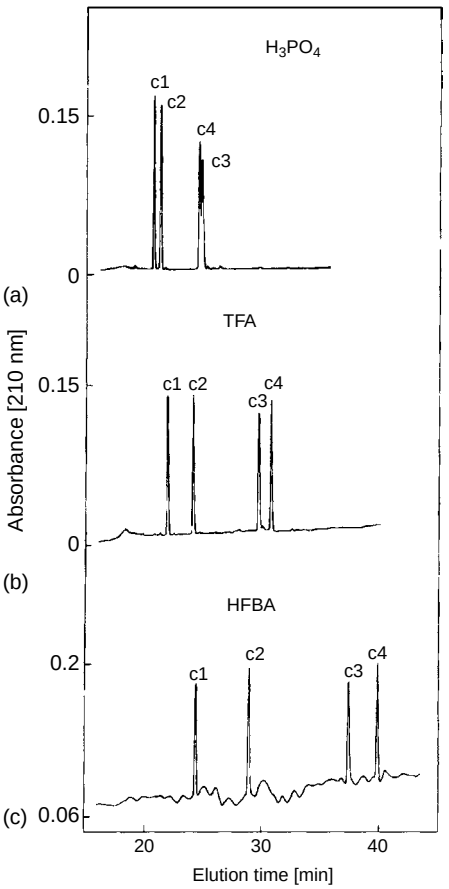

Fig. 6 The influence of ion-pairing reagent on the separation of a mixture of synthetic peptides in reversed phase chromatography using a SynChropak C18, 25 cm × 4.6 mm ID, 6.5-µm particle size, 30-nm pore size (SynChrom, Linden, IN). Conditions: linear gradient from 0 to 100% acetonitrile containing A. 0.1%H_3PO_4; B, 0.1% TFA; and C; 0.1% HFBA; flow rate of 1 mL min^{-1}, 26 °C. Peptide sequences:
C1 = Ac-GGGLGGAGGLK-amide,
C2 = Ac-KYGLGGAGGLK-amide.
C3 = Ac-GGALKALKGLK-amide,
C4 = Ac-KYALKALKGLK-amide
(Reprinted with permission from Mant, C.T., Hodges, R.S. (Eds.) (1991) *High Performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis and Conformation*, CRC Press, Boca Raton, FL, pp. 327–41).

semipreparative columns of dimensions 30 cm × 1 cm ID and preparative columns of 30 cm × 2 cm ID can be used.

The selection of the internal diameter of the column is based on the sample capacity and detection sensitivity. While most analytical applications are carried out with columns of internal diameter of 4.6-mm ID (as shown in Figs. 1, 3, and 5), for samples derived from previously unknown proteins where there is a limited supply of material, the task is to maximize the detection sensitivity. In these cases, narrow bore columns of 1- or 2-mm ID can be used, which allows the elution and recovery of samples in much smaller volumes of solvent. Capillary chromatography (see Sect. 7 below) is also finding increasing application where capillary columns of ID between 0.2 to 0.4 mm and column length of 15 cm result in the analysis of fmole of sample as shown in Fig. 8. The effect of decreasing column internal diameter on detection sensitivity is shown in Fig. 9 for the analysis of lysozyme on a C18 material packed into columns of 4.6-, 2.1-, and 0.3-mm ID.

3.2.4 Operating Parameters

There are several operating parameters that can be changed in order to manipulate

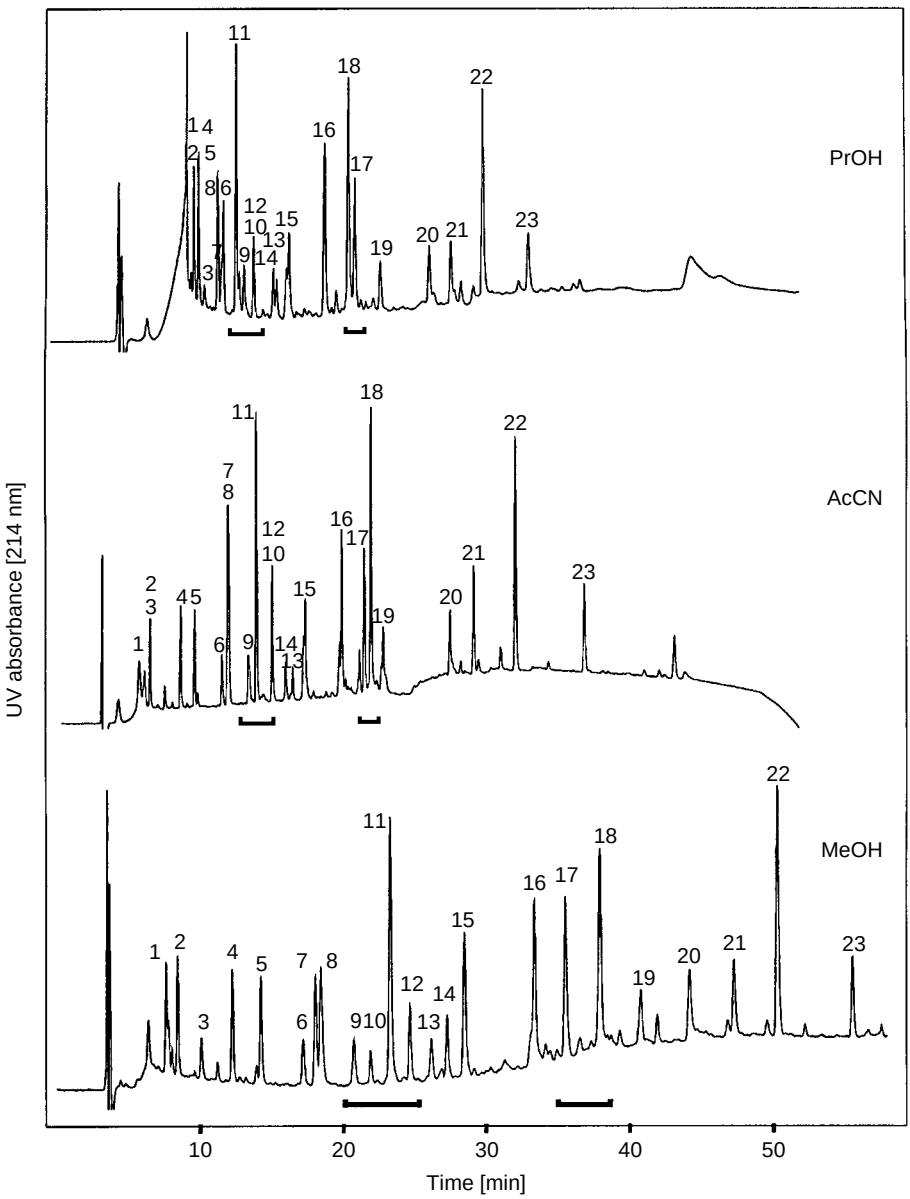

Fig. 7 The influence of organic solvent on the reversed phase chromatography of tryptic peptides derived from porcine growth hormone. Column: Bakerbond (J T Baker, Phillipsburg, NJ) RP-C4, 25 cm × 4.6 mm ID, 5-µm particle size, 30-nm pore size. Conditions, linear gradient from 0 to 90% 2-propanol (top), acetonitrile (middle) or methanol (bottom) with 0.1% TFA over 60 min, flow rate of 1 mL min^{-1}, 25 °C (Reprinted with permission from Aguilar, M-I., Hearn, M.T.W. (1996) *Methods Enzymol.* **270**, 3–26).

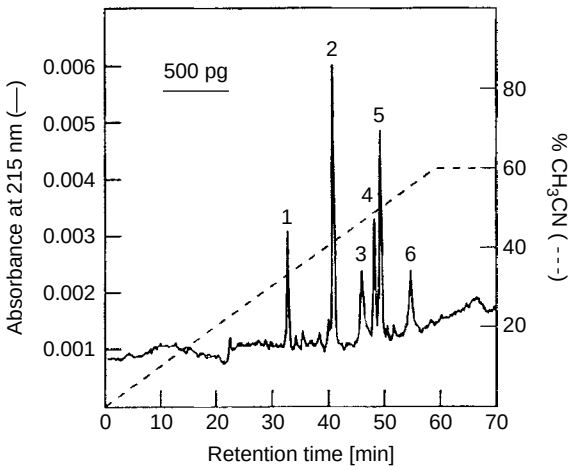

Fig. 8 The separation of proteins by reversed phase capillary chromatography. Column; Brownlee RP-300, 15 cm × 0.2 mm ID, 5-µm particle size, 30-nm pore size. Conditions; linear gradient from 0 to 60% acetonitrile over 60 min, flow rate of 1.4 µL min^{-1}. Protein; 1 = ribonuclease B, 2 = lysozyme, 3 = bovine serum albumin, 4 = carbonic anhydrase, 5 = myoglobin, 6 = ovalbumin (Reprinted with permission from Moritz, R.L., Simpson, R.J. (1993) in: Imahori, K., Sakiyama, F. (Eds.) *Methods in Protein Sequence Analysis*, Plenum Press, New York, pp. 3–10).

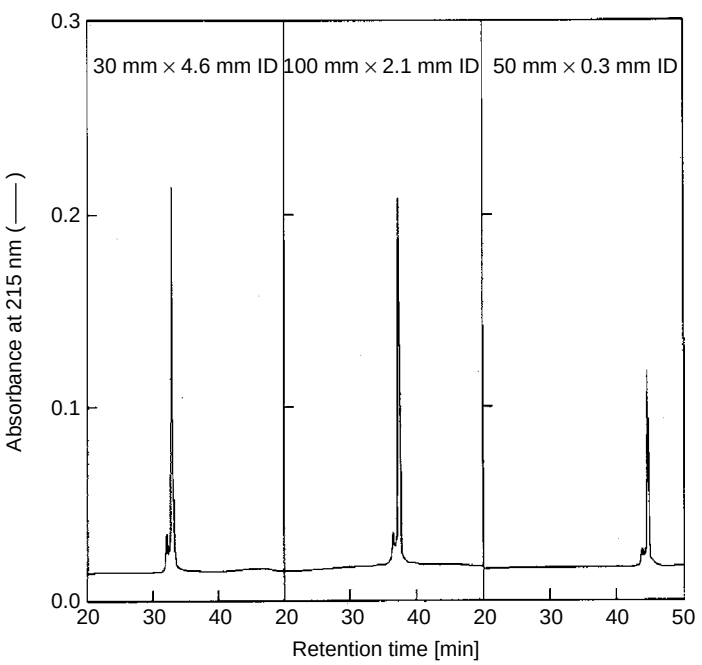

the resolution of peptide and protein mixtures in RPC. These parameters include the gradient time, the gradient shape, the mobile phase flow rate, and the operating temperature. The typical experiment with an analytical scale column would utilize a linear gradient from 5% organic solvent up to between 50 to 100% solvent over a time range of 20 to 120 min while flow rates are between 0.5 to 2.0 mL min^{-1}. With microbore columns [1–2 mm ID] flow rates of 50 to 250 μL min^{-1} are used, while for capillary columns of 0.2- to 0.4-mm ID, flow rates of 1 to 4 μL min^{-1} are applied. At the preparative end of the scale with columns of 10- to 20-mm ID, flow rates between 5 to 20 mL min^{-1} are required.

The choice of gradient conditions will depend on the nature of the molecules of interest. The influence of gradient time on the separation of a series of ribosomal proteins is shown in Fig. 10. Generally, the use of longer gradient times provides improved separation. However, these conditions also increase the residence time of the peptide or protein solute at the sorbent surface, which may then result in an increase in the degree of denaturation.

The operating temperature can also be used to manipulate resolution. While the separation of peptides and proteins is normally carried out at ambient temperature, solute retention in RPC is influenced by temperature through changes in solvent

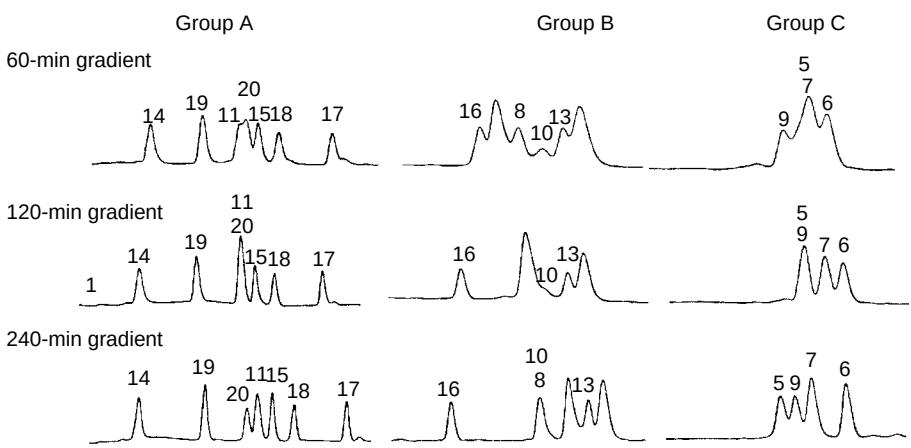

Fig. 10 The effect of gradient time on the resolution of ribosomal proteins by reversed phase chromatography. Column: Du Pont Bioseries Protein PLUS. Conditions: 25 to 46% acetonitrile with 0.1% triethylamine and 0.042% TFA, over 60 min (top), 120 min (middle) or 240 min (bottom), at a flow rate of 0.7 mL min^{-1} (Reprinted with permission from Ghrist, B.F.D. et al. (1990) in: Gooding, K.M., Regnier, F.E. (Eds.) *HPLC of Biological Macromolecules: Methods and Applications*, Dekker, New York, pp. 403–427).

Fig. 9 Effect of column internal diameter on detector sensitivity. Column: Brownlee RP-300 C8 (7 μm particle size, 30 nm pore size), 3 cm × 4.6 mm ID and 10 cm × 2.1 mm ID (Applied Biosystems) and 5 cm × 0.32 mm ID. Conditions: linear gradient from 0 to 60% acetonitrile with 0.1% TFA over 60 min, 45 °C. Flow rates, 1 mL min^{-1}, 200 μL min^{-1} and 4 μL min^{-1} for the 4.6-, 2.1-, and 0.32-mm ID columns, respectively. Sample loadings, lysozyme, 10, 4, and 0.04 μg for the 4.6-, 2.1-, and 0.32-mm ID columns respectively. (Reprinted from Moritz, R.L., Simpson, R.J. (1992) *J. Chromatogr.* **646**, 53–65, with kind permission from Elsevier Science – NL, Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands).

viscosity. In addition to this, peptide and protein conformation can also be manipulated by temperature. In the case of proteins, where biological recovery is not important, increasing temperature can be used to modulate retention via denaturation of the protein structure. For peptides, it has been shown that secondary structure can actually be enhanced through binding to the hydrophobic sorbent. Changes in temperature can therefore also be used to manipulate the structure and retention of peptide mixtures.

Chromatographic recovery of proteins from polyacrylamide gels is another important application of RPC. Inverse gradient elution chromatography has been successfully utilized for the micropreparative isolation of proteins from SDS-PAGE electroeluates. This approach is based on observations that certain RPC packings display strong interactions with proteins at high organic solvent concentrations. This allows the loading of the electroeluate under conditions where the protein is retained while the SDS and other gel-related contaminants are washed through the column. The protein is then recovered in high yield by a gradient of decreasing organic solvent.

Detection of peptides and proteins in RPC, and in all modes of chromatography, generally involves detection at between 210 to 220 nm, which is specific for the peptide bond, or at 280 nm, which corresponds to the aromatic amino acids tryptophan and tyrosine. The use of photodiode array detectors can enhance the detection capabilities by the on-line accumulation of complete solute spectra. The spectra can then be used to identify peaks specifically on the basis of spectral characteristics and for the assessment of peak purity. In addition, second derivative spectroscopy can provide information on the conformational integrity of proteins following elution.

In summary, RPC is now firmly entrenched as the central tool for the analysis of peptides and proteins and thus plays a pivotal role in the pharmaceutical industry providing the core analytical technique at all stages of the development of peptide and protein-based therapeutics.

4 Hydrophobic Interaction Chromatography

Hydrophobic interaction chromatography (HIC) is a valuable technique for the separation of proteins under nondenaturing conditions. HIC involves the use of high salt concentrations to promote hydrophobic interactions between the protein and the hydrophobic stationary phase. Solutes are then eluted in order of increasing hydrophobicity through the application of a descending salt gradient that weakens the hydrophobic interactions between the protein and the sorbent material. The ligands used for HIC materials are less hydrophobic than those in RPC. Thus, in contrast to the denaturing effects of low pH and organic solvent present in RPC systems, the mobile phases used in HIC generally stabilize the protein structure.

4.1 Basic Principles

The primary factors that control protein retention in HIC are similar to those factors that contribute to retention in RPC, that is, the surface tension of the mobile phase and the exposed hydrophobic surface area of the protein solute. The use of the solvophobic theory has been extended to describe the effects of neutral salts on protein solubility and retention

in HIC. In particular, the dependence of the logarithmic capacity factor on salt concentration m, can be expressed as

$$\log k' = \log k_0 + [(-D\mu + \upsilon + N\Delta A_h\sigma)m]/RT \quad (8)$$

where $D\mu$ and υ are terms related to the dielectric constant of the medium and the protein dipole moment. Thus, retention in HIC is dependent on the contact surface area ΔA_h, established between the solute and the sorbent and the molal surface tension increment σ of the eluting salt solution. Selectivity optimization in HIC therefore involves changes in protein solubility characteristics through either changes in the structure of the stationary phase ligand or changes in the mobile phase characteristics through the use of salts of different molal surface tension increments. A linear relationship between the capacity factor of a protein and the corresponding salt concentration is generally observed and Eq. 8 can be simplified to

$$\log k' = \lambda m + C \quad (9)$$

where C incorporates all the salt-independent terms and λ is the slope of plots of log k' versus salt concentration, m. The linear dependence of protein retention on salt concentration for a series of globular proteins is illustrated in Fig. 11. The term λ has also been shown to be linearly related to the protein surface area and the molal surface tension increment, which is consistent with the predictions of Eq. 8.

4.2 Factors Influencing Retention in HIC

4.2.1 Stationary Phase

Both polymeric and silica-based HIC supports have been produced and a range of mildly hydrophobic ligands are available to perform HIC. In particular, both alkyl and aryl ligands have proven to be successful in obtaining high levels of selectivity. Figure 12 shows the influence of a range

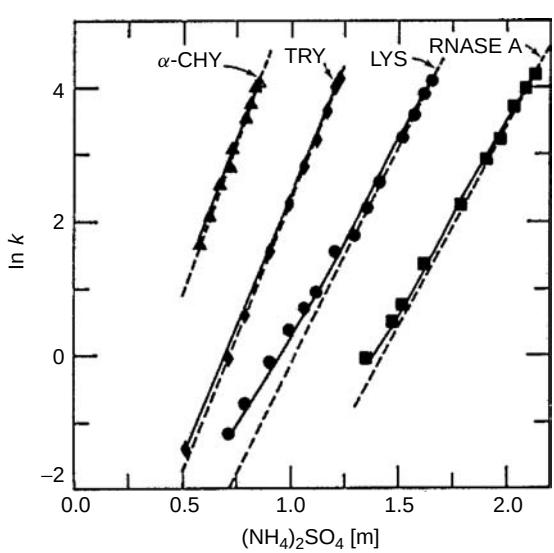

Fig. 11 The dependence of log k' on ammonium sulfate concentration for a series of proteins separated by HIC during isocratic elution. Column; TSK-phenyl (Toyo Soda). Conditions: 50 mM phosphate buffer, pH 7, flow rate 3 mL min^{-1}, 45 °C. Proteins; α-CHY = α-chymotrypsinogen, TRY = trypsinogen, LYS = lysozyme, RNASE A = ribonuclease A (Reprinted with permission from Katti, A., Maa, Y-F., Horvath, Cs. (1987) *Chromatographia* **24**, 646–650).

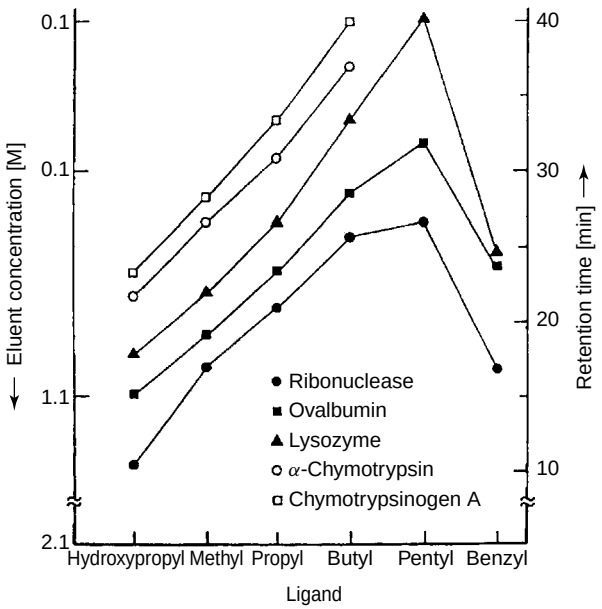

Fig. 12 The effect of ligand type on the elution of proteins in hydrophobic interaction chromatography. Column: SynChropak hydroxypropyl, propyl, methyl, butyl, pentyl and benzyl columns, 25 cm × 4.6 mm ID, 6.5-µm particle size, 30-nm pore size (SynChrom, Linden, IN). Conditions; linear gradient from 2-0M ammonium sulphate in 0.1M potassium phosphate, pH 6.8, over 30 min at a flow rate of 1 mL min^{-1}. (Reprinted from Gooding, D.L. et al. (1986) Optimization of preparative hydrophobic interaction chromatographic purification methods, *J. Chromatogr.* **359**, 331–37, with kind permission from Elsevier Science – NL, Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands).

of ligands on the retention behavior of a series of globular proteins and demonstrates that protein retention increases in the order hydroxypropyl < methyl < benzyl = propyl < isopropyl < phenyl < pentyl. This figure clearly illustrates the influence of the ligand structure on the retention of proteins in HIC. Other ligand types that can be used include silica-based ether-bonded alkyl phases, neopentylagarose, and phenylagarose.

4.2.2 Mobile Phase

Selectivity in HIC can be manipulated by changes in the nature of the eluting salt, salt concentration, pH, temperature, and the addition of mobile phase modifiers. Protein retention is strongly dependent on the type of salt employed. Salts defined as being kosmotropic or structure-making are used as they enhance hydrophobic interactions through a salting-out mechanism. This effect is formerly described in Eq. 8 in terms of the molal surface tension increment σ, the values of which are listed in Table 1 for a selection of commonly used salts in HIC. The salts with a higher σ-value will result in longer protein retention times. There have been a number of studies that have documented

Tab. 1 Molal surface tension increment for salts used in HIC[1].

Salt	σ ($\times 10^3$ *dyn-g/cm-mol*)
Na_3PO_4	2.88
Na_2SO_4	2.74
$(NH_4)_2SO_4$	2.17
Mg_2SO_4	2.06
Na_2HPO_4	2.03
NaCl	1.64
KCl	1.50

[1] *Source*: Adapted from Katti, A., Maa, Y-F., Horvath, Cs. (1987) Protein surface area and retention. . ., *Chromatographia* **24**, 646–650.

the influence of different salts on protein retention in HIC. An example of the effect of salt on the retention of lysozyme in HIC is shown in Fig. 13. $(NH_4)_2SO_4$ is the most commonly used salt in HIC, while Na_2SO_4 and NaCl also find application. $(NH_4)_2SO_4$ has a high solubility and low UV absorbance and is readily available in high purity required for HPLC. Initial salt concentrations usually range from 1 to 3 M, and the starting concentration can also influence selectivity of a separation.

The pH of mobile phases used in HIC is typically between 5 and 7 and buffered

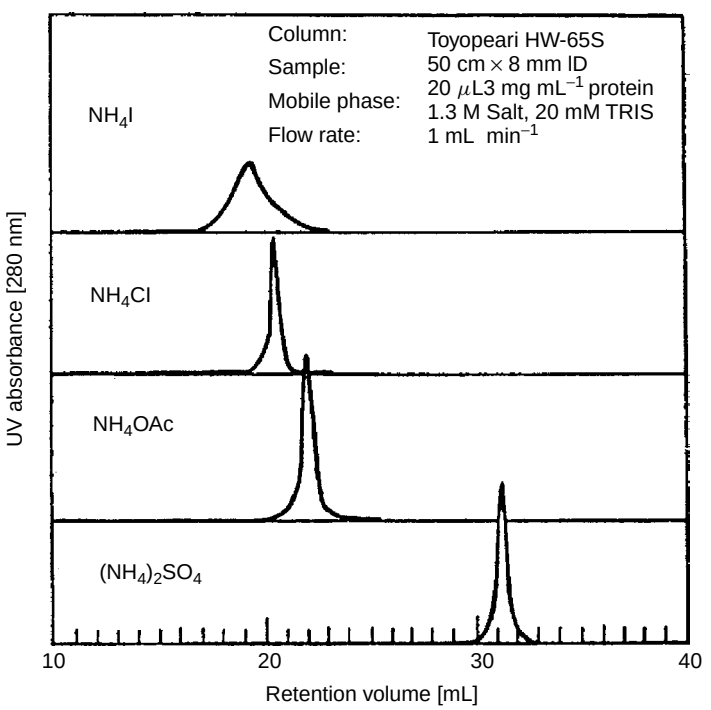

Fig. 13 The effect of salt on the elution of lysozyme in HIC. Column; Toyopearl HW-65S (Toyo Soda), 50 cm × 8 mm ID, 30-μm particle size, 100-nm pore size. Conditions; isocratic elution with 20 mM TRIS, pH 7, containing 1.3 M ammonium iodide, chloride, acetate, or sulfate, at a flow rate of 1 mL min^{-1} (Reprinted with permission from Rogetter, B.F. et al. (1990) in: Ladisch, M.R., Willison, R.C., Painton, C.C., Builder, S.E. (Eds.) *Protein Purification: From Molecular Mechanisms to Large Scale Processes*, American Chemical Society, Washington, DC, pp. 80–92).

with sodium or potassium phosphate. The influence of pH on protein retention is dependent on the particular protein as the manipulation of charges located in or near the hydrophobic binding domain will have a profound effect on the affinity of the protein for the sorbent material. Thus, changes in pH represent a useful parameter for modulating selectivity.

The gradient shape is an additional parameter that can be used to manipulate selectivity in HIC. An example of the influence of gradient time on the HIC separation of proteins is illustrated in Fig. 14. The LSS model can also be applied to the retention behavior of proteins in HIC and generally reveals a linear dependence of retention on salt concentration. The

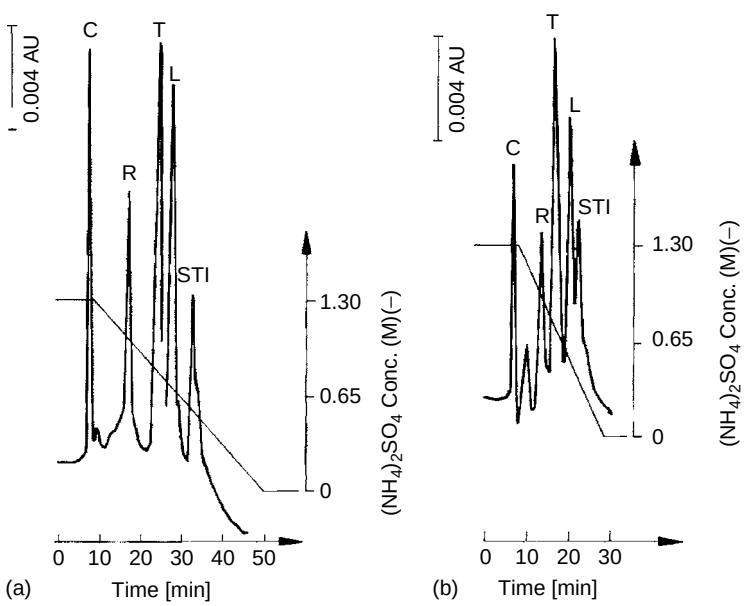

Fig. 14 The effect of gradient time on the resolution of proteins in HIC. Column: Octyl agarose, 60 × 6 mm ID, 5- to −7-µm particle size. Conditions: linear gradient from 1.33 to 0M ammonium sulfate in 0.05 M sodium phosphate, pH 7, over (a) 40 min or (b) 20 min at a flow rate of 0.2 mL min^{-1}, detection at 280 nm. Proteins: C = cytochrome *c*, R = ribonuclease A, T = transferrin, L = lysozyme and STI = soybean trypsin inhibitor (Reprinted from Hjertén, S. (et al.) (1986) Gradient and isocratic high-performance hydrophobic interaction chromatography of proteins on agarose columns, *J. Chromatogr.* **359**, 99–109, with kind permission from Elsevier Science – NL, Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands).

Fig. 15 The influence of temperature on the retention of (a) cytochrome c and (b) myoglobin separated by HIC. Column; Bio-Gel TSK-Phenyl-5-PW, 75 cm × 7.5 mm ID, 10-µm particle size, 100-nm pore size (Bio-Rad, Richmond, CA). Conditions; linear gradient from 1.7-0M ammonium sulfate in 0.1 M sodium phosphate, pH 7, over 15 min and a flow rate of 1 mL min^{-1}. (Reprinted from Ingraham, R.H. et al. (1985) Denaturation and the effects of temperature on hydrophobic interaction and reversed phase. . ., *J. Chromatogr.* **327**, 77–92, with kind permission from Elsevier Science – NL, Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands).

selection of elution conditions to maximize resolution between components therefore follows the same rationale as described for RPC in the previous section.

The addition of other solvent modifiers has also been shown to affect retention in HIC through changes in the surface tension of the mobile phase. These include detergents such as Triton X-100 and CHAPS, organic solvents such as 5 to 20% methanol, acetonitrile or even ethylene glycol and urea or guanidine hydrochloride at concentrations of 1 to 2 M. In all cases, it is possible for these additives to cause denaturation of the target proteins, so care is needed to minimize protein

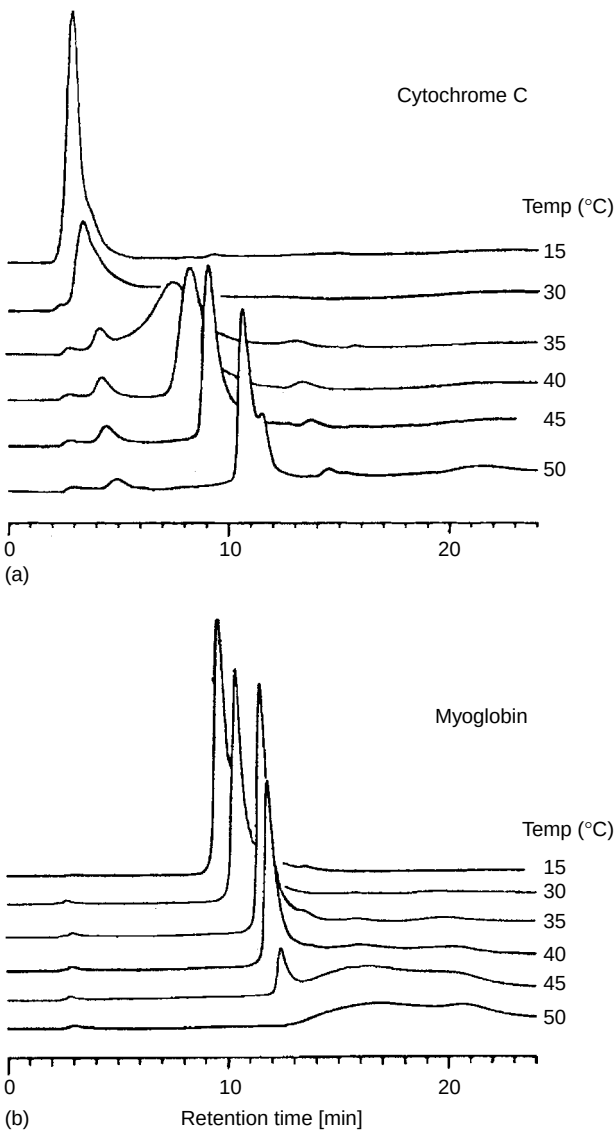

conformational changes when introducing these additives to the mobile phase.

Temperature can also be used to manipulate selectivity in HIC through changes in protein conformation. Depending on the protein solute, chromatography in the range 15 to 50 °C can be used to sharpen the shapes of individual peaks and hence improve resolution as can be seen for myoglobin in Fig. 15(b). However, significant band broadening can also be observed as a result of slow conformational interconversions, which results in a decline in resolution as is evident in Fig. 15(a) for cytochrome *C*.

Overall, HIC is a powerful tool for the purification of proteins in a biologically active form. Moreover, protein structure and conformation play a crucial role in the chromatographic behavior of proteins and subtle changes in selectivity can be achieved through changes in the relative solubility and three-dimensional structure of the protein solute.

5 Ion-exchange Chromatography

High-performance ion-exchange chromatography (IEC) is now extensively used in the analysis of proteins, and also to a lesser extent for the analysis of peptides. The early stages of protein purification generally utilize solubility-based techniques to carry out the initial fractionation. Differences in size and shape of the proteins are then exploited through application of size exclusion or preparative electrophoretic techniques. Adsorptive techniques, including IEC and RPC are then introduced to allow rapid increases in the level of resolution, recovery, and product purity.

A significant advantage of IEC over the other adsorptive modes of chromatography is the nondenaturing effects of the solutions used to elute proteins from the ion-exchange sorbents. Thus, while gross conformational changes can be observed in RPC, these are not commonly found in IEC of proteins.

5.1 Basic Principles

Protein retention in ion-exchange chromatography arises from electrostatic interactions between the peptide or protein and the charged sorbent material and solutes are eluted by increases in the concentration of a displacer salt. As a consequence, the "net charge" concept is widely used to predict the retention characteristics of proteins with both anion and cation-exchange materials. According to this model, and as illustrated in Fig. 16, a protein will be retained on a cation-exchange column if the solvent pH is lower than the pI of the protein. Conversely, a protein will be retained on an anion-exchange column if the pH is above the pI of the protein. With mobile phases operating at a pH equal to the protein's pI, the surface of the protein is considered to be overall electrostatically neutral and under these conditions, the protein should not be retained on either cation or anion exchangers.

While this model can be used to predict the retention behavior of peptides in IEC, this classical model is recognized as a simplistic approach to describing protein retention. The amphoteric nature of proteins results in the existence of localized areas of electrostatic charge at different pHs, which can allow the protein to be retained even under conditions where the protein may be at its isoelectric point.

A number of more detailed models have been developed to describe protein retention in ion-exchange chromatography.

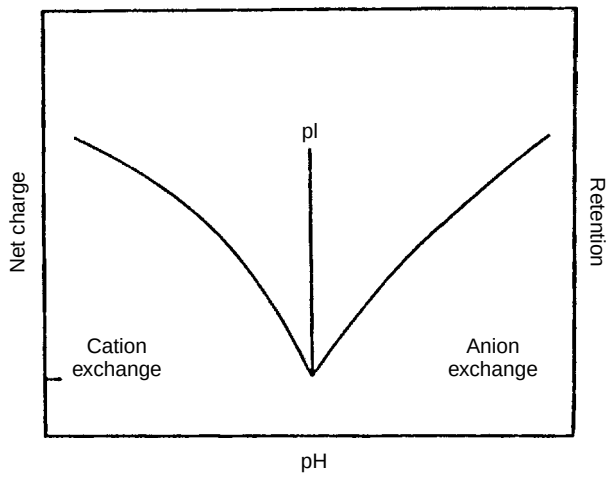

Fig. 16 The theoretical relationship between protein net charge and chromatographic retention in ion-exchange chromatography.

These include the stoichiometric displacement model and different approaches to solving the Poisson–Boltzmann equation for the interaction of charged surfaces. Each of these approaches attempts to provide a mechanistic description of the retention of proteins in IEC and provides a theoretical basis for the experimentally observed linear relationship between the solute retention and either (a) the reciprocal of the displacing salt concentration c, as follows

$$\log k' = \log K_O + Z_C \log(1/c) \qquad (10)$$

or (b) the corresponding ionic strength, I, according to

$$\log k' = \log K_O + B(1/\sqrt{I}) \qquad (11)$$

where K_0 is a constant, which incorporates a number of terms including a binding constant, the phase ratio of the packing material and various charge valency terms and can be determined by linear regression of plots of log k' versus log $(1/c)$ or plots of log k' versus $1/\sqrt{I}$. An example of the linear dependency of protein retention on log $(1/c)$ is shown in Fig. 17 for the isocratic and gradient elution of ovalbumin and carbonic anhydrase by strong anion-exchange chromatography. Z_C is the slope of these plots and can be related to the size of the electrostatic contact area established between the solute and the stationary phase. In addition, the LSS model can be used as the basis to optimize separations as described previously for RPC, through establishing the experimental conditions that maximize the separation of these plots for different components in a mixture. The variation in protein retention with changes in displacer salt concentration can be seen in Fig. 18 for the isocratic separation of myoglobin, carbonic anhydrase and ovalbumin using sodium chloride as the displacer salt.

Overall, it has now been established that the magnitude of electrostatic interactions between proteins and the charged sorbent material in IEC depends on the number and distribution of charged sites on the solute molecule that define the electrostatic contact area of the protein, the charge density of the sorbent, and the mobile phase composition. In summary,

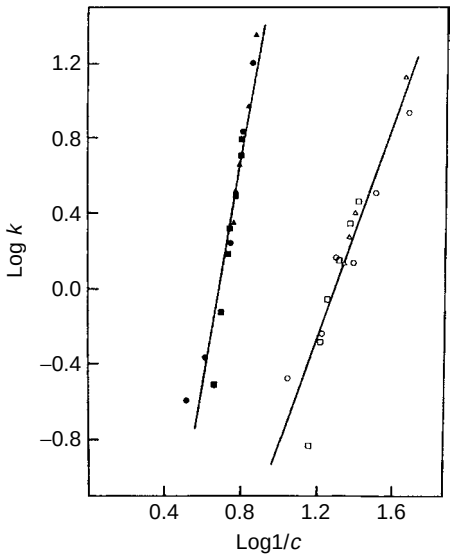

Fig. 17 Retention plots for the isocratic and gradient separation of ovalbumin (filled symbols) and carbonic anhydrase (open symbols) by strong anion-exchange chromatography. Column; Mono-Q HR 5/5, (Pharmacia, Uppsala, Sweden) Conditions; 0.02M piperazine at pH 9.6 with sodium bromide as the displacer. (Reprinted from Hodder, A.N., Aguilar, M-I., Hearn, M.T.W. (1989) The influence of different displacer salts on the retention properties of proteins separated by gradient anion-exchange chromatography, *J. Chromatogr.* **476**, 391–411, with kind permission from Elsevier Science – NL, Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands).

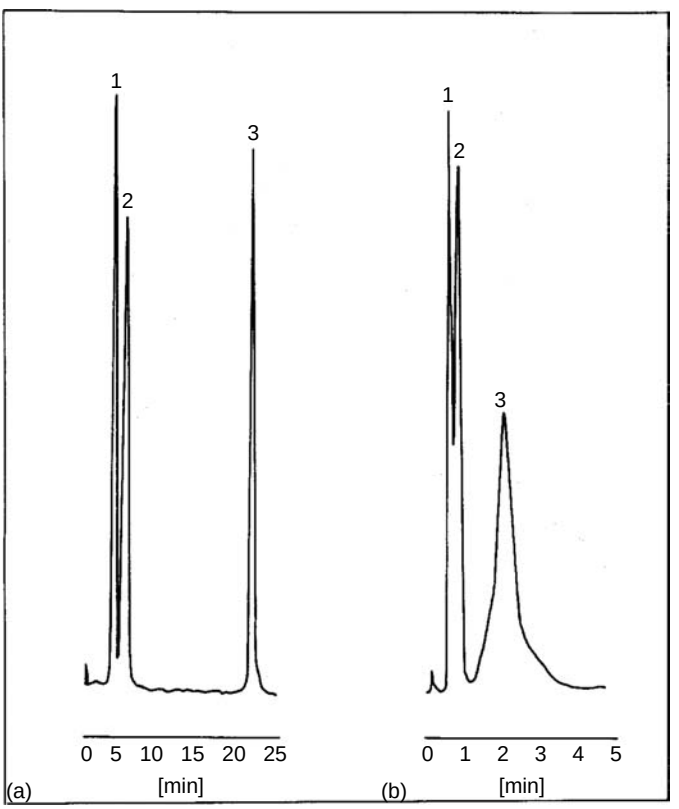

Tab. 2 Commonly encountered ligands for ion-exchange chromatography.

Ion exchanger	*Functional group*	
Strong cation exchanger (SCX)	Methylsulphonate	$-CH_2SO_3^-$
Weak cation exchanger (WCX)	Carboxymethyl	$-OCH_2COO^-$
Strong anion exchanger (SAX)	Methyl trimethyl ammonium	$-CH_2N^+ (CH_3)_3$
Weak anion exchanger (WAX)	Diethylaminoethyl	$-OCH_2H_2N^+ H(CH_2CH_3)_2$

the magnitude of the electrostatic interactions, and hence retention, in IEC is dependent on the following structural and chromatographic parameters:

1. the number and distribution of charged sites on the solute molecule that constitute the electrostatic contact area;
2. the charge density of the immobilized charged ligand;
3. the mobile phase composition.

It is these parameters that represent the factors that can be used to manipulate peptide and protein surface charge to allow optimization of selectivity in IEC.

5.2 Factors Influencing Retention in IEC

5.2.1 Stationary Phases

The support materials available for high-performance ion-exchange chromatography are generally silica-based or polymer-based materials. An ion-exchange material is selected on the basis of the desired particle and pore size, swelling characteristics at the operational pH range. However, the major factor is the capacity of the ion exchanger, which depends on the nature of the charged functional groups and the charge density as well as the pore size of the material and the charge distribution on the protein solute.

The two classes of ion exchangers are cation exchangers, which contain negatively charged functional groups, and anion exchangers, which contain positively charged functional groups. The most commonly encountered ligands are listed in Table 2. Strong cation exchangers normally contain sulphonic acid groups while strong anion exchangers contain quaternary ammonium functional groups. The charged ligands in weak cation exchangers generally contain carboxylic groups while weak anion exchangers are primary, secondary, or tertiary amines. The terms strong and weak refer to the degree of ionization with pH as strong ion exchangers are completely ionized over a much wider pH range than weak ion exchangers. The physical properties of some typical anion and cation exchangers used in protein chromatography are listed in Table 3.

Fig. 18 Isocratic elution profile of proteins separated by anion-exchange chromatography. Column; Mono-Q HR 5/5, (Pharmacia, Uppsala, Sweden) Conditions; 0.02 M piperazine at pH 9.6 with sodium chloride as the displacer at a concentration of (a) 60 mM and (b) 180 mM. Proteins; (1) = myoglobin, (2) = carbonic anhydrase, (3) = ovalbumin (Reproduced with permission from Aguilar, M-I., Hodder, A.N., Hearn, M.T.W. (1991) High-Performance Ion-exchange Chromatography of Proteins, in: Hearn, M.T.W. (Ed.) *HPLC of Proteins, Peptides and Polynucleotides – Contemporary Topics and Applications*, VCH Publications, Deerfield, FL, pp. 199–245).

Tab. 3 Properties of typical anion and cation exchangers used in protein chromatography[1].

Property	*Anion exchangers*			*Cation exchangers*		
	TSKgel DEAE-5PW (Toyo Soda)	*Mono-Q (Pharmacia)*	*SynChropak AX-300 (SynChrom)*	*TSKgel SP-5PW (Toyo Soda)*	*Mono-Q (Pharmacia)*	*SynChropak CM-300 (SynChrom)*
Support	Macroreticular polymer with hydrophilic surface	Macroreticular polymer with hydrophilic surface	Silica	Macroreticular polymer with hydrophilic surface	Macroreticular polymer with hydrophilic surface	Silica
Particle diameter (μm)	10	10	10	10	10	10
Charged ligand	$-CH_2CH_2N(C_2H_5)_2$	$-CH_2N^+(CH_3)_3$	$-NHCH_2CH_2$	$-CH_2CH_2CH_2SO_3^-$	$-CH_2SO_3^-$	$-CH_2COO^-$
Operational pH range	2–12	2–12	2–8	2–12	2–12	2–8
pKa	11.3	1.4	NA	2.5	2.6	NA
Binding Capacity (mg mL^{-1})	30	65	30	40	75	NA
Exclusion limit for polyethylene glycol (Da)	1 000 000	500 000	150 000	1 000 000	500 000	150 000

[1] *Source*: Adapted from Kato, Y. (1993) *Study: Comparison of Ion-Exchange Columns*, Tosoh Corporation, Montgomeryville, PA.
Courtesy: Modified from Choudhary, G., Horvath, Cs. (1996) Ion-exchange chromatography, *Methods Enzymol.* **270**, 47–82.

5.2.2 Mobile Phases

The selectivity of proteins in IEC can be manipulated by variation in solution pH and ionic strength, which alters the electrostatic surface potential of the protein solutes and the charged ligand thereby influencing the strength of the electrostatic interactions. Changes in the nature of the displacer ion and the buffer species represent additional methods by which protein retention can be modified.

While NaCl is the most commonly used ionic displacer, a number of other monovalent and multivalent salts can be used. The ions may influence retention through specific interactions with the ion-exchange ligand, thereby changing their ionic properties. In addition, specific salts may alter the conformation of proteins, which in turn will influence their retention behavior. At fixed ionic strengths, anions can be ranked in terms of solute retention as follows:

$$F^- < CH_3COO^- < Cl^- < HPO_4^{2-} < SO_4^{2-}$$

Similarly, cations are ranked according to the series

$$K^+ < Na^+ < NH_4^+ < Ca^{2+} < Mg^{2+}$$

KCl, NaOAc, $MgCl_2$, $Mg_3(SO_4)_2$ have all been used for the analysis and purification of a wide range of proteins. The effect of different displacer salts on protein retention in weak anion-exchange chromatography is shown in Fig. 19, which illustrates the profound influence that the nature of the salt can exert on the electrostatic interactions between proteins and ion-exchange materials.

The selection of buffer depends on the pH range required to adsorb the protein to the stationary phase. While selection of a pH can be a straightforward task based on the known pI of the protein, for proteins of unknown pI or closely related proteins such as isoforms or recombinant muteins, a map of retention versus pH can assist in the selection of mobile phase pH. An example of the changes in selectivity, which can occur over a given pH range, is shown in Fig. 20 for the separation of a series of lysozymes by cation-exchange chromatography. Once the pH range is established, additional changes in selectivity can be obtained through changes in the nature of the buffer species. The most commonly employed buffer species include phosphate and tris buffers. A range of buffers, which are commonly used in IEC of proteins, is listed in Table 4.

Tab. 4 Buffers commonly used in IEC.

pH range	*Buffer*
Cation exchange	
1.5–2.5	Maleic acid
2.6–3.6	Citric acid
3.6–4.3	Lactic acid
4.8–5.2	Acetic acid
5.0–6.0	Malonic acid
6.7–7.6	Phosphate
7.6–8.2	HEPES
Anion exchange	
4.5–5.0	*N*-methyl piperazine
5.0–6.0	Piperazine
5.8–6.4	Bis-tris
7.3–7.7	Triethanolamine
7.6–8.0	Tris
8.5–9.0	1,3-diaminopropane
9.5–9.8	Piperazine

5.2.3 Operating Parameters

While isocratic elution can be used to separate proteins in IEC, gradient elution is generally employed to obtain high-resolution separations of proteins in IEC. Linear elution over 16 to 120 min is

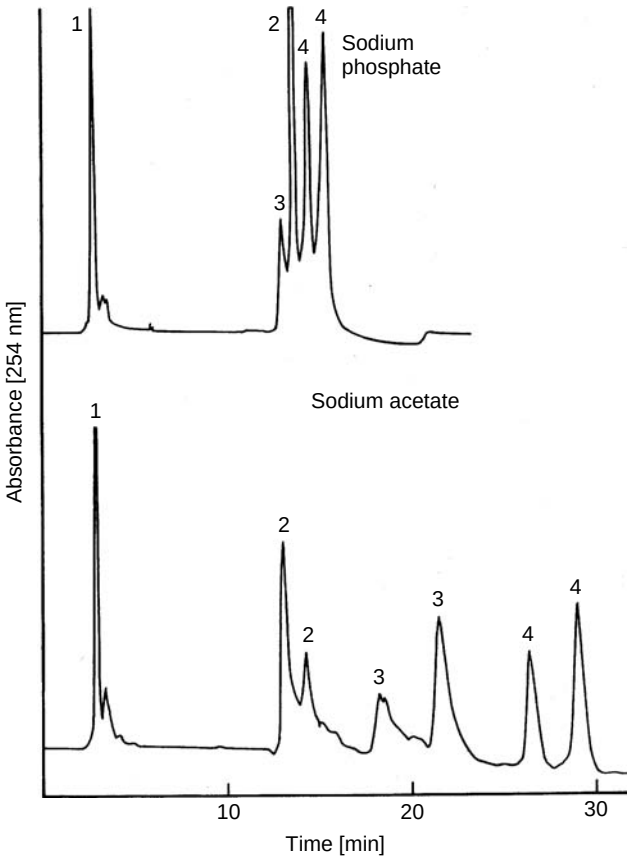

Fig. 19 The effect of different displacer salts on the retention of proteins separated by weak anion-exchange chromatography. Column, SynChropak AX-300, 25 cm × 4.1 mm ID, 6.5-μm particle size, 30-nm pore size (SynChrom, Linden, IN). Conditions; Linear gradient from 0 to 1 M salt over 30 min at a flow rate of 1 mL min^{-1}, detected at 254 nm. Top panel = sodium phosphate, lower panel = sodium acetate. Proteins; 1 = myoglobin, 2 = conalbumin, 3 = ovalbumin, 4 = β-lactoglobulins B and A. (Reprinted with permission from Nowlan, M.P., Gooding, K.M. (1991) in: Mant, C.T., Hodges, R.S. (Eds.) *High-Performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis and Conformation*, CRC Press, Boca Raton, FL, pp. 203–13).

generally applied between ionic strengths ranging from 0 to 0.5 M salt. Buffer concentrations usually range between 20 to 50 mM.

A number of additional materials can be added to the mobile phase to further enhance selectivity. For example, hydrophobic interactions may contribute to peptide and protein retention in IEC due to the nature of the stationary phase material. It has been reported that solutes cannot be eluted with some ion exchangers without

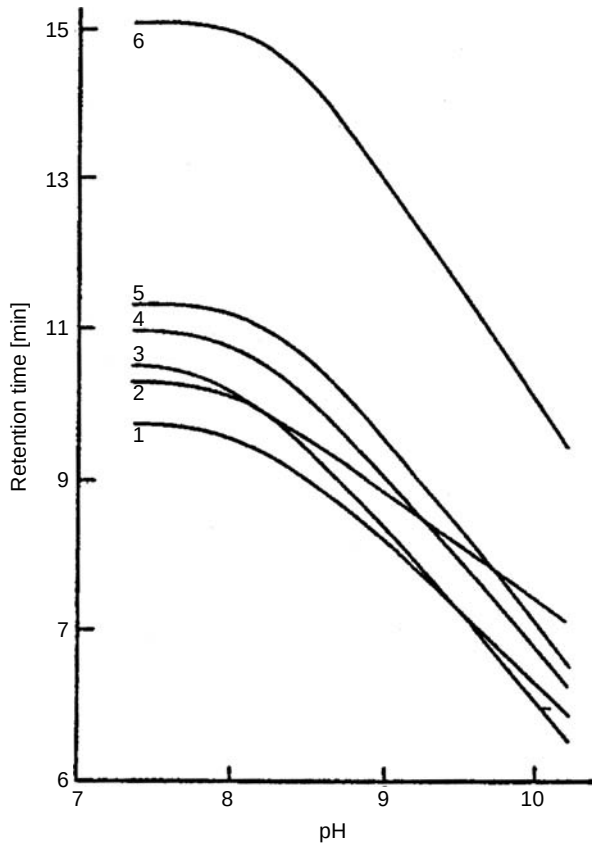

Fig. 20 The effect of pH on the retention of lysozymes in cation-exchange chromatography. Column; SynChropak CM-300, 25 cm × 4.1 mm ID, 6.5-µm particle size, 30-nm pore size (SynChrom, Linden, IN). Conditions; linear gradient from 0 to 1 M sodium chloride in 10 mM phosphate buffer (pH < 8) or 10 mM borate buffer (pH > 8), at a flow rate of 1 mL min^{-1} over 20 min. Proteins; 1 = hen egg white lysozyme, 2 = Japanese quail, 3 = ring necked pheasant. 4 = duck A, 5 = duck B, 6 = duck C. (Reprinted from Fausnaugh-Pollit, J. (et al.) (1988) Chromatographic resolution of lysozyme variants, *J. Chromatogr.* **443**, 221–28, with kind permission from Elsevier Science – NL, Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands).

the addition of acetonitrile or methanol to the mobile phase. The percentage organic modifier is usually in the range of 10 to 40%. Higher levels of solvent can cause salt precipitation and may also affect protein conformation.

5.3 Peptide Separation in IEC

For peptide applications where incomplete separation is observed in RPC, ion-exchange chromatography represents

a very useful alternative separation mode. At neutral pH, basic peptides can be separated by cation-exchange chromatography and acidic peptides can be analyzed by anion-exchange chromatography. However, ion exchange of peptides is more commonly carried out at acidic pH, in the range 2.5 to 3.0, where most peptides are

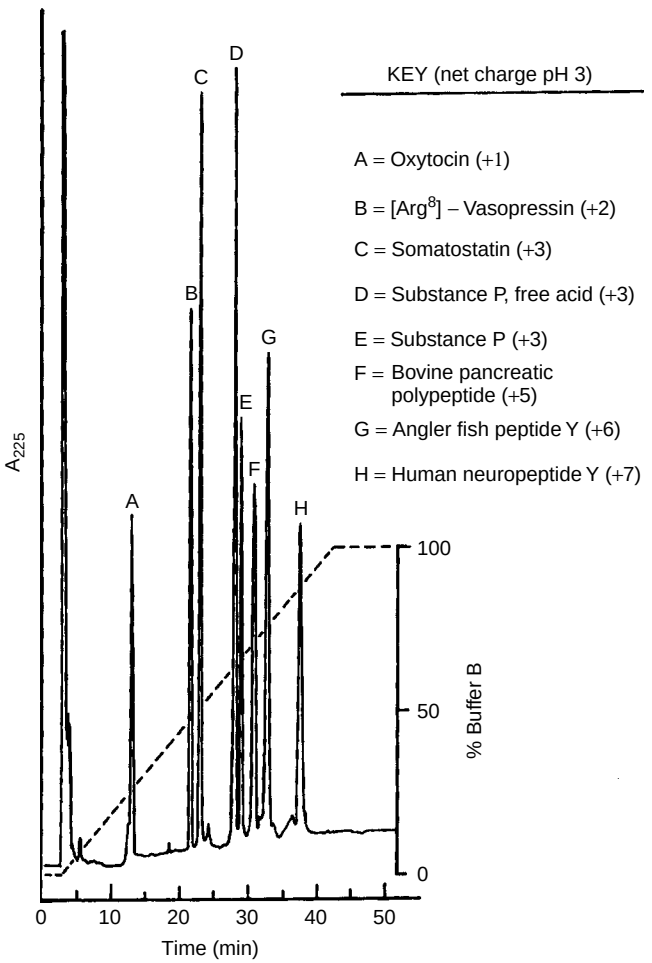

Fig. 21 Separation of a series of peptides by cation-exchange chromatography. Column; PolySULFOETHYL Aspartamide, 20 cm × 4.6 mm ID, 5 μm-particle size, 30-nm pore size (PolyLC, Columbia, MD). Conditions; linear gradient from 0 to 0.25 M potassium chloride in 5 mM potassium phosphate pH 3 and 25% acetonitrile at a flow rate of 0.7 mL min^{-1}. The peptides vary in net charge from +1 ~ +7 as indicated in the figure. (Reprinted from Alpert, A.J., Andrews, P.C. (1988) Cation-exchange chromatography of peptides on poly(2-sulfoethyl aspartamide)-silica, *J. Chromatogr.* **443**, 85–96, with kind permission from Elsevier Science – NL, Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands).

positively charged and hence cation exchange is applicable. At this pH range, the negative charges associated with aspartate and glutamate residues and the C-terminus are neutralized, while arginine, lysine, histidine residues and the N-terminus are positively charged.

Commonly used solvents for peptide IEC are usually based on phosphate buffers with NaCl or KCl as the displacer ion. For peptides up to approximately 50 residues in length with no significant secondary structure, retention is governed by the number of positive charges as illustrated in Fig. 21, which shows the separation of eight peptides ranging in overall net charge of $+1 \sim +7$. Thus, peptides differing by a single charge are generally well resolved. However, peptides with the same charge but different amino acid composition can also be separated because of differences in overall charge density. In addition, hydrophobic interactions may also contribute to the retention of peptides with ion-exchange resins. In these cases, organic solvent can be added to the mobile phases to further modulate selectivity. 10 to 40% v/v of either methanol or acetonitrile can be used and an example of the selectivity changes with the addition of acetonitrile is shown in Fig. 22 for a series

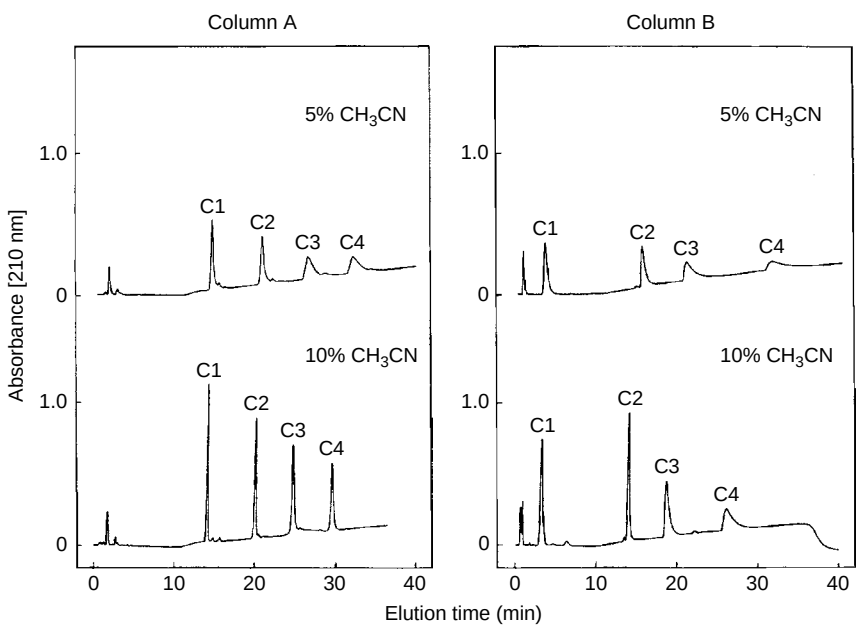

Fig. 22 The effect of organic solvent on the retention of peptides separated by cation-exchange chromatography. Column A: PolySULFOETHYL Aspartamide, 20 cm × 4.6 mm ID, 5-μm particle size, 30-nm pore size (PolyLC, Columbia, MD). Column B; Mono S HR 5/5 (Pharmacia, Uppsala, Sweden). Conditions, linear gradient from 0 to 1 M sodium chloride 5 mM potassium phosphate, pH 6.5, at a flow rate of 1 mL min^{-1}, 26 °C. Peptide sequences: C1 = Ac-GGGLGGAGGLK-amide, C2 = Ac-KYGLGGAGGLK-amide. C3 = Ac-GGALKALKGLK-amide, C4 = Ac-KYALKALKGLK-amide. (Reprinted with permission from Mant, C.T., Hodges, R.S., in: Mant, C.T., Hodges, R.S. (Eds.) (1991) *High-Performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis and Conformation*, CRC Press, Boca Raton, FL, pp. 171–185).

of peptides separated by strong cation-exchange chromatography.

In summary, high-performance ion-exchange chromatography continues to be an important technique for the analysis and purification of proteins under mild nondenaturing conditions and also provides a very useful selectivity alternative for the analysis of peptide samples.

6 Size Exclusion Chromatography

Size exclusion chromatography (SEC) is frequently used as the first step in the isolation of proteins from complex mixtures where separation is carried out according to molecular size and shape. SEC can also be used for desalting samples through buffer exchange and has also found application in the analysis of peptides. SEC has also been established as a physicochemical tool for estimating molecular size and shape of proteins and has provided insight into protein folding mechanisms by monitoring changes in protein size as a function of changes in the concentration of chemical denaturants.

6.1 Basic Principles

Separation in SEC is based on differences in molecular size in solution. Porous stationary phases with defined pore diameters and elution conditions are used that minimize interaction between the solute molecules and the stationary phase material. The larger the molecule, the smaller the amount of accessible pore volume. Molecules that are larger than the largest pore diameter cannot penetrate into the stationary phase pores and pass through the column with the fastest retention time. These molecules are eluted with V_o, while the smallest molecule is eluted in V_t, the total volume of the column. V_t is the sum of the void volume and the interstitial volume V_i, *that is*,

$$V_t = V_o + V_i \tag{12}$$

Solute elution volume in SEC is denoted by V_e which should be between V_o and V_t and can be expressed as follows

$$V_e = V_o + K_d V_i \tag{13}$$

where K_d is the distribution coefficient, which defines the fraction of internal volume which is accessible to the protein solute.

6.2 Experimental Conditions

The packing materials available for high-performance SEC are generally silica-based or polymeric. The pore diameter of SEC supports determines the exclusion limits of the material. Columns are characterized in terms of the molecular weight range, which can be adequately separated and is dependent on the pore diameter. Generally, the pore diameter ranges between 100 and 500 Å. The pore volume is also an important property of an SEC material, which must be sufficiently large to provide a high peak capacity, that is, the ability to separate seven peaks with a resolution of one. Column efficiency in high-performance SEC supports is particularly important as solutes are eluted isocratically and therefore do not exhibit band sharpening, which occurs with gradient elution. SEC supports generally have particle diameters between 5 and 10 μm.

The hydrodynamic shape and volume rather than molecular weight *per se* is the

physical property of proteins that causes separation in SEC. In order to achieve accurate estimations of molecular weight, a column must be calibrated with molecules that have the same overall shape. Under ideal conditions, K_d will be proportional to the logarithm of the molecular weight of the protein as illustrated by a typical calibration curve shown in Fig. 23 for the protein separation shown in Fig. 24. While the majority of applications of SEC have involved the analysis and purification of proteins, SEC has also recently been used in the analysis of peptides. An example of the high level of resolution that can be achieved for peptides is shown in Fig. 25, which demonstrates excellent separation in the MW range of 75 to −12 500 Da.

Mobile phase selection is important in SEC to minimize nonspecific interactions between the support and the solutes and also to avoid mobile phase induced changes in solute molecular shape. Both ionic and hydrophobic interactions can also contribute to the elution volume of proteins in SEC as most packings are weakly acidic (due to residual negative charges) and/or mildly hydrophobic. Ionic interactions can be minimized by increasing the ionic strength of the mobile phase through the addition of up to 0.5 M NaCl. However, increasing ionic strength also results in the enhancement of hydrophobic interactions. Thus, a balance is necessary to minimize both the undesired interactions. A common mobile phase composition that is employed for SEC of proteins is phosphate buffer at pH 7 with ionic strength of 0.05 to 0.1. Thus, by optimizing pH and ionic strength, secondary interactions can be almost excluded. However, ionic or hydrophobic interactions can also be used to manipulate selectivity of protein separations. The example shown in

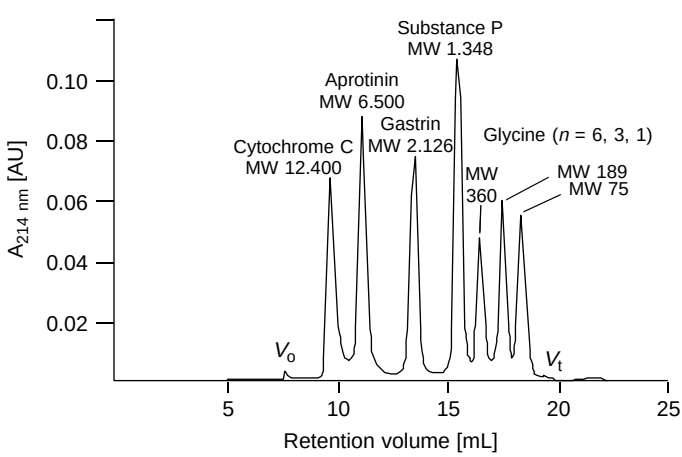

Fig. 23 A typical calibration curve for the separation of proteins by size exclusion chromatography. column; SynChropak GPC300, 30 cm × 7.8 mm ID, 5-µm particle size, 30-nm pore size (SynChrom, Linden, IN). Conditions; 0.1 M potassium phosphate, pH 7, flow rate 1 mL min. (Reprinted with permission from Gooding, K.M., Freiser, H.H. in: Mant, C.T., Hodges, R.S. (Eds.) (1991) *High-Performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis and Conformation*, CRC Press, Boca Raton, FL, pp. 135–44).

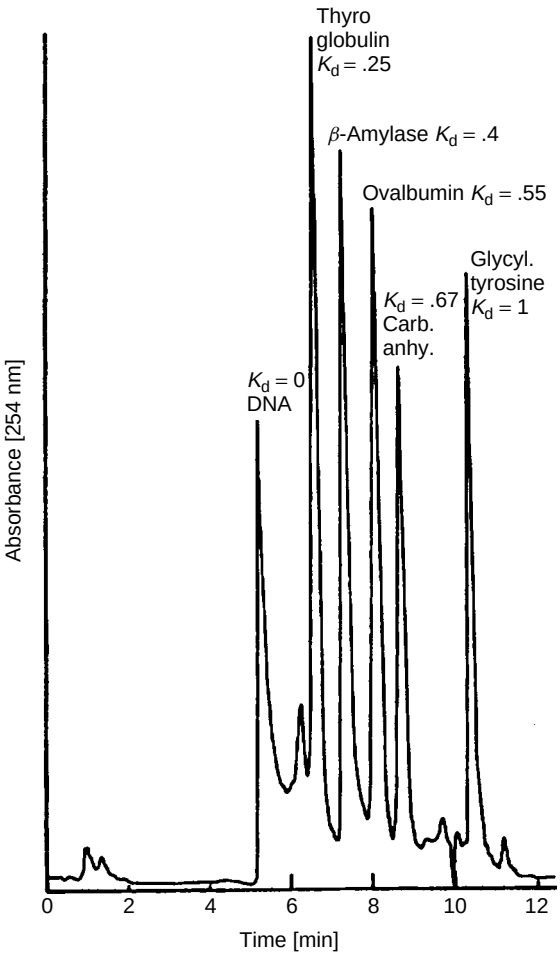

Fig. 24 The elution profile of proteins used to create the calibration curve shown in Fig. 24. (Reprinted with permission from Gooding, K.M., Freiser, H.H. in: Mant, C.T., Hodges, R.S. (Eds.) (1990) *High-Performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis and Conformation*, CRC Press, Boca Raton, FL, pp. 135–44).

Fig. 26 clearly demonstrates the effect of salt on the separation of peptides at two different pHs.

7 Capillary Liquid Chromatography

7.1 General Considerations

The transfer of analytical HPLC methods from conventional-size columns with typical 3.0- to 4.6-mm ID to capillary size dimensions of 0.1- to 0.5-mm ID has been a significant advance in recent years bringing several advantages such as reduction in the consumption and disposal of solvents, working with limited sample amounts and efficient interfacing with electrospray ionization MS. In principle, the various mechanisms utilized in conventional LC separations can be effectively implemented in packed capillary formats. However, most separations utilizing capillary LC are now performed using the reversed phase mode of separations.

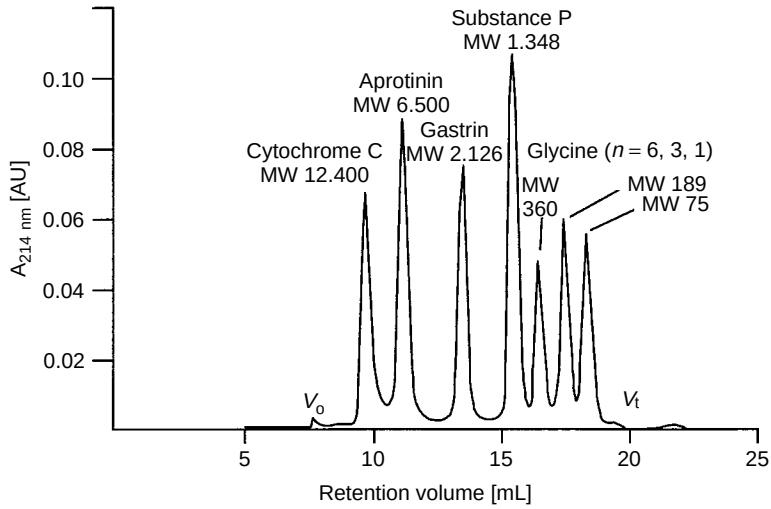

Fig. 25 The separation of peptides by size exclusion chromatography. Column; Superdex Peptide HR 10/30 (Pharmacia, Uppsala, Sweden). Conditions; 0.25 M NaCl in 0.02M phosphate buffer pH 7.2, flow rate 0.25 mL min^{-1}. (High-performance size exclusion columns Data file No 18-1106-06, 1994, courtesy of Amersham Pharmacia Biotech).

This is primarily because RPC is highly compatible with MS detection using ESI and requires only low flow rates to achieve both high separation efficiency and high detection sensitivity. Other modes of LC separation commonly utilized in peptide/protein separations include IEC, SEC, HIC, and affinity chromatography. Many of these modes of chromatography are utilized in sample clean up, preconcentrations, or fractionation before final analyses.

Compared to the various modes of LC, RPC generates the highest separation efficiencies because of the fast mass transfer of analyte between the stationary and mobile phases. Simultaneously, RPC facilitates high detection sensitivity through its compatibility with MS. Figure 27 shows the LC-MS analysis of an Arg-C digest of plasminogen activator separated on a C18 column, with the total ion current in the upper trace and the elution profile detected at 214 nm in the lower trace. The availability of on-line mass spectrometry thus significantly facilitates the identification of the peptide fragments. Other important applications involve the identification of posttranslational modifications of peptides and proteins, assignment of disulfide bonds and the identification of peptides bound to major histocompatibility complex molecules.

The separation efficiency of RPC is highly dependent on the maximum operating pressure that is determined by the column length, internal diameter, and particle size that can be used in a given separation. The primary limitation for using high-pressure capillary LC in routine peptide/protein analysis has been the need for off-line sample introduction. On-line, high-pressure-resistant switching valves with pressure limits of at least 20 000 psi, are under development. The application of a 20 000-psi capillary RPC

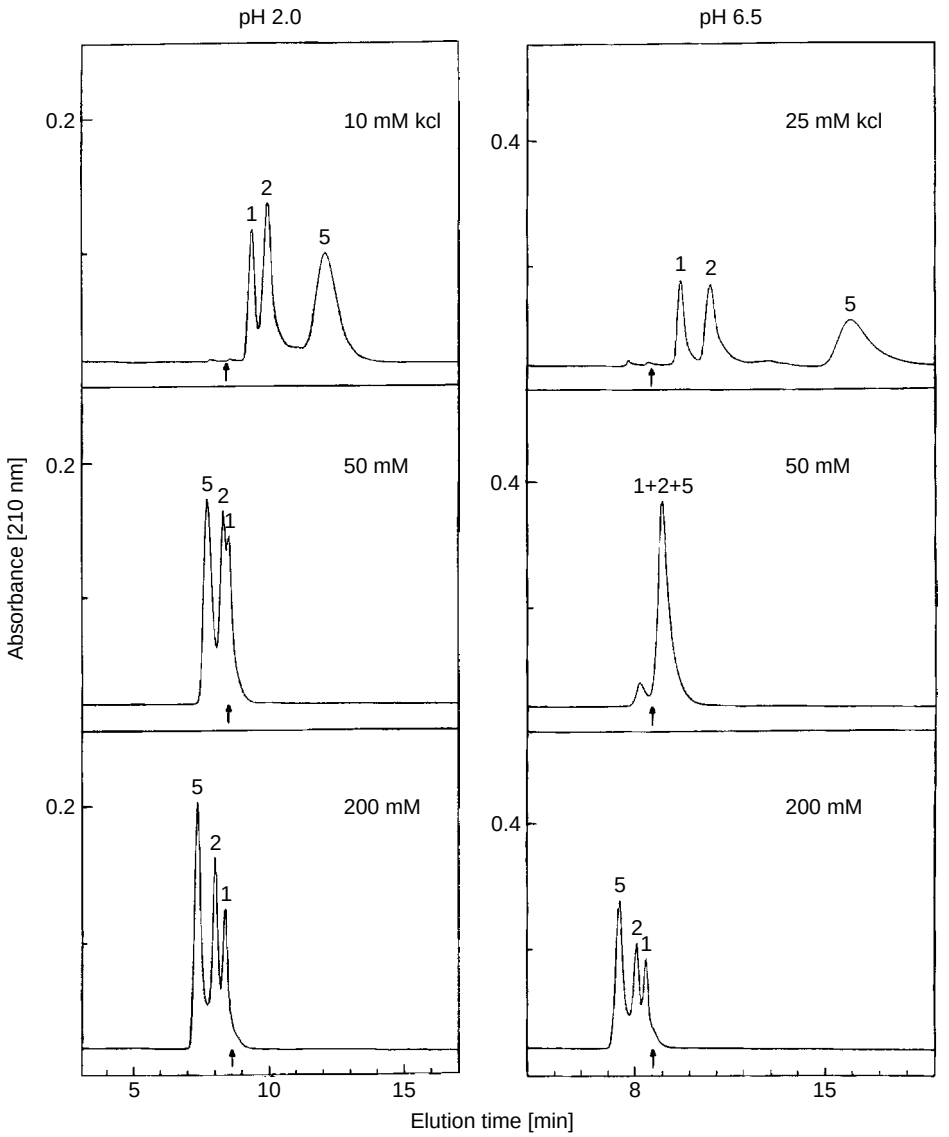

Fig. 26 Effect of salt on nonspecific interactions in size exclusion chromatography at pH 2 (left) and pH 6.5 (right). SynChropak GPC60, 30 cm × 7.8 mm ID, 5-µm particle size, 30-nm pore size (SynChrom, Linden, IN). Conditions; left −0.1% TFA containing 10, 50, or 200 mM potassium chloride; right −5 mM KH_2PO_4 containing 25, 50, or 200 mM potassium chloride, flow rate 1 mL min^{-1} Peptide sequences: 1 = AcGLGAKGAGVG-Amide, 2 = Ac$(GLGAKGAGVG)_2$-amide, 3 = Ac$(GLGAKGAGVG)_5$-amide. (Reprinted with permission from Anspach, B., Gierlich, H.U., Unger, K.K. (1988) Comparative study of Zorbax Bio Series GF 250 and GF 450 and TSK-Gel 3000 SW and SWXL columns. . ., *J. Chromatogr.* **443**, 45–54. Copyright CRC Press, Boca Raton, FL).

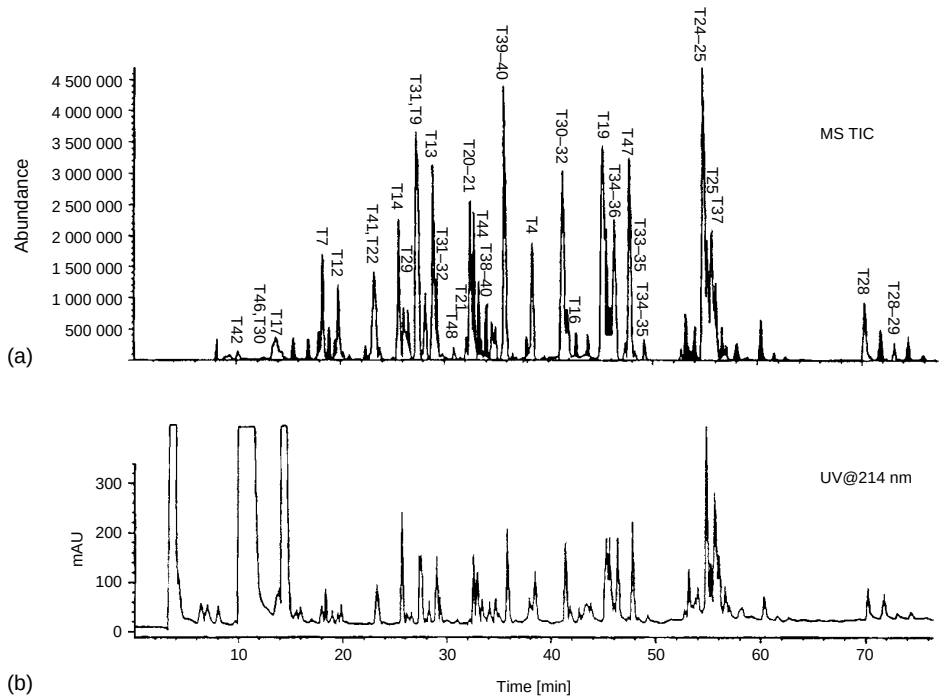

Fig. 27 LC-MS of a tryptic digest of single-chain plasminogen activator. Column:Vydac C18, 5-µm particle size, 30-nm pore size. Conditions: linear gradient from 0 to 60% acetonitrile with 0.1% TFA over 90 min, 45 °C, flow rate 0.2 mL min^{-1}. (a) electrospray mass spectrometry total ion current, (b) detection at 214 nm. (Reprinted from Apffel, A. et al. (1995) Application of capillary electrophoresis, high performance liquid chromatography, on-line electrospray mass spectrometry and matrix-assisted laser desorption ionisation. . ., *J. Chromatogr. A* **717**, 41–60. with kind permission from Elsevier Science – NL, Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands).

system for routine proteomic analyses is to be expected. High-efficiency capillary LC is now being performed at 10 000 psi using ~85-cm-long capillaries that have inner diameters from 15 µm to 150 µm packed with 3-µm porous particles. At such pressures, special switching valves with positive-feedback sealing are required for sample introduction, and separations generating peak capacities of $\sim 10^3$ for global cellular proteolytic polypeptide mixtures have been demonstrated.

For capillary separations of intact proteins, it is estimated that peak capacities of $\sim 10^2$ can be achieved, but the resolving power is not sufficient for separating complex proteomic protein mixtures (e.g. containing ~1000 proteins), and multidimensional separations are typically required. Few literature reports describe RPC applications for resolving protein complexes, probably because of the relatively strong interactions between the analyte and the stationary phase. As discussed in Sect. 3, proteins tend to denature under reversed-phase conditions (low pH and high organic solvent concentrations) making their quantitative elution rather difficult and measuring the molecular mass of a whole protein is not sufficient for

its unambiguous identification. For this approach to be successful, it is necessary to digest all of the proteins in a sample and to perform the separation at the peptide level.

The ability to effectively load significant amounts (either volume or mass) of sample onto the capillary column to improve detection capabilities for low-abundance species is particularly important to proteomic applications. In isocratic capillary LC separations, limitations for large volume sample loading result from degradation of separation efficiency. Under composition gradient conditions, however, introduction of a relatively large sample volume can be conducted without significant loss of separation efficiency. This is because the sample can be "focused" on the head of the column inlet in the weakly solvating mobile phase used for sample introduction. However, as the sample loading time is proportional to sample volume, a long-extended sample injection time is required for this case. For example, introducing a 10 μL sample with optimum LC linear velocities (e.g., ~0.1 cm min^{-1}) onto a 150-μm ID capillary column requires ~20 min. Use of high mobile phase linear velocities can shorten the sample loading time. However, the use of long capillaries packed with small particles limits the maximum mobile phase flow rate that current LC pumps can supply because of pressure limitations. In practice, relatively small volume injections having high sample concentrations can be used to reduce analysis time. For example, a *Deinococcus radiodurans* tryptic digest with concentrations (protein content) up to ~50 μg μL^{-1} has been successfully loaded onto a 150-μm ID packed capillary column. However, introducing more concentrated samples is problematic because their viscosities are too large to easily allow manipulation with a syringe.

The achievable sample mass loading capacity depends on the specific column dimension, the chromatographic packing material properties, and the analyte properties. A capacity of 20 to 200 μg g^{-1} of packing material can be estimated for RPC from these previous studies, although different sample capacities have been reported. For long packed capillaries having ~150-μm ID, the sample loading capacity can be estimated to be 0.15 to 1.5μg for an individual analyte. In global cellular proteolytic digest samples (cellular proteome), hundreds of thousands of polypeptides are potentially present, having large differences in both quantity and physicochemical properties. Using an online guard column or SPE column can be used to resolve this problem.

7.2 Capillary RPC of Proteolytic Peptides for Proteomic Analysis

Capillary RPC separations of global protein tryptic digests for proteomic studies are attracting increased attention. Such techniques are commonly carried out using acetonitrile as the organic solvent in the mobile phase(s) to achieve high-efficiency separations of complex mixtures that contain a variety of properties of proteolytic peptides. Perfluorinated carboxylic acids (e.g. trifluoroacetic or heptafluorobutyric acid) can be used as an acidic modifier in the mobile phase with ESI-MS detection conditions. Octadecyl-bonded silica particles (i.e. C18 particles) are typically used as the stationary phase. The high-quality silica matrix used for the bonding reactions is attractive for eluting strongly basic peptides. Porous particles are useful for obtaining large sample capacities. At an operating pressure of 10 000 psi (the present practical

limit for reproducible sample introductions using switching valves), ~85-cm columns packed with 3-μm particles can achieve high-efficiency separations. The use of very small particle sizes at this pressure limits the total separation efficiency. Current high-efficiency capillary LC can be performed with column inner diameters from 15 to 150 μm, corresponding to flow rates from 20 nL min^{-1} to 1.5 μL min^{-1} at their optimal linear velocities to obtain peak capacities of $\sim 10^3$. This provides great flexibility for column selection to achieve the desired MS sensitivity for a specific sample amount.

In order to improve the throughput of gradient capillary RPC, high-efficiency multiple-capillary LC separations have been implemented with a single LC pumping system consisting of several dual-capillary column devices that operate at constant pressure. These dual-capillary column devices are serially connected to two high-pressure mobile phase lines, fed using independent LC pumps (mobile phase A and mobile phase B). Each dual-capillary column device is composed of a mobile phase mixer (2.8 mL), a switching valve for mobile phase gradient, a split line (30-μm ID fused silica), a switching valve sample injector, two capillary separation columns, and a column switch valve. While constant pressure LC operation offers limited flexibility for manipulating the mobile phase gradient to optimize separation selectivity, reliably obtaining high separation efficiencies is more important for separations of highly complex samples such as global cellular protein digestions. The only limitation to the number of such capillary column devices that can be connected in a single system results from the flow rate of the high-pressure LC pumps. For example, if each dual-capillary column device operates at a mobile phase flow rate of ~20 μL min^{-1} (including splitting, typically used for our capillary LC experiments), up to ten dual-capillary column devices can be connected using pumps that supply a stable flow rate of ~0.2 mL min^{-1} at the desired pressure. The separation efficiency and reproducibility of the multiple-capillary LC separations have been experimentally evaluated. Comparing the separation results of a yeast tryptic digest for the dual-capillary column device with the results obtained using single capillary LC shows that the same separation efficiency was reproducibly obtained with both capillaries for this extremely complex sample. The average errors for elution time and peak height are 4 and 8%, respectively, for recognizable peaks with elution times of <50 min, and were only 1.8 and 5%, respectively, for recognizable peaks eluting between 50 and 170 min. Currently, the role of capillary RPC in the analysis of proteome-wide proteolytic peptides has been limited to resolving these peptides. The information from peptide elution properties on RPC columns has not yet been fully utilized for helping peptide identification. Understanding peptide elution behavior relevant to their structures will greatly improve the accuracy and confidence in MS identification of separated peptides.

Capillary RPC of proteins is limited to medium efficiencies (peak capacities of $\sim 10^2$), and its application in proteomic analysis is, at the present time, generally applied to only relatively simple protein systems. However, detailed analysis of intact proteins for yeast large ribosomal subunits has been studied by the capillary RPC. Capillary RPC of membrane proteins and construction of two-dimensional separations for complex protein mixtures are worthy of future investigation.

7.3 Capillary Ion-exchange Chromatography

IEC is primarily used for sample fractionation of either proteomic intact proteins or their proteolytic peptides, and is usually practiced using microcolumns or even conventional columns. However, because of sensitivity requirements for proteomic analysis, the use of capillary IEC is appealing. Capillary IEC can generate peak capacities of 200 to 400 for both cellular proteins and their proteolytic polypeptides. The sample capacity has been found to be significantly smaller than for capillary RPC using the same dimensions of packed capillary columns. For example, <100 μg of an yeast lysate tryptic digest can be loaded on an 85 cm × 150 μm ID capillary packed with 3-μm SCX particles (300 Å pores), while ~500 μg of the same sample can be loaded on the same dimension capillary packed with 3-μm C18 particles (300 Å pores). Desalting of samples can improve the loading capacity to some extent. These properties are important in the selection of capillary column dimensions for two-dimensional separations, particularly considering issues related to detection sensitivity (i.e. with MS). Additionally, IEC elution properties of proteins and peptides potentially provide valuable information for assisting protein identification.

7.4 Other LC Methods

Other LC separation mechanisms such as SEC and affinity chromatography have also been used for peptide/protein studies. Currently, these separations are primarily used for proteomic sample processing, such as sample purification, isolation of interested proteins, and in some situations, sample fractionation. Most of these applications are performed using larger column LC, and their implementation with capillary columns promises to improve the overall sensitivity.

7.5 Multidimensional Capillary Separations for Proteins

Because of the complexity of proteomic samples, two or even multidimensional separations appear attractive for achieving high-resolution separations. Although comprehensive two-dimensional LC separations have been developed, the fast separation in the second dimension (currently) is generally far from optimal for data-dependent MS/MS peptide identification (e.g. using ion trap MS). Additionally, the interface between the first and the second dimension separation also effectively degrades the detection of low-abundance species. The two-dimensional separations that have been used for proteomic analyses to date achieve their separation efficiencies (combined peak capacities of $\sim 10^3$) primarily from the second dimension separation, while the first dimension separation is only used for a limited number of sample fractionations (typically the fraction number is <20). The separation of total human plasma proteins digested with trypsin with high-efficiency capillary SCX/RPC is demonstrated in Fig. 28. Although used in this way, the first dimension separation efficiency is also important because it determines the number of overlapping (or repeated) components in neighboring fractions. Both CE and capillary LC can also be used as the first separation dimension for sample fractionation; however, the separation sample capacity for the first dimension and the fluid compatibility with MS for the second dimension have to be

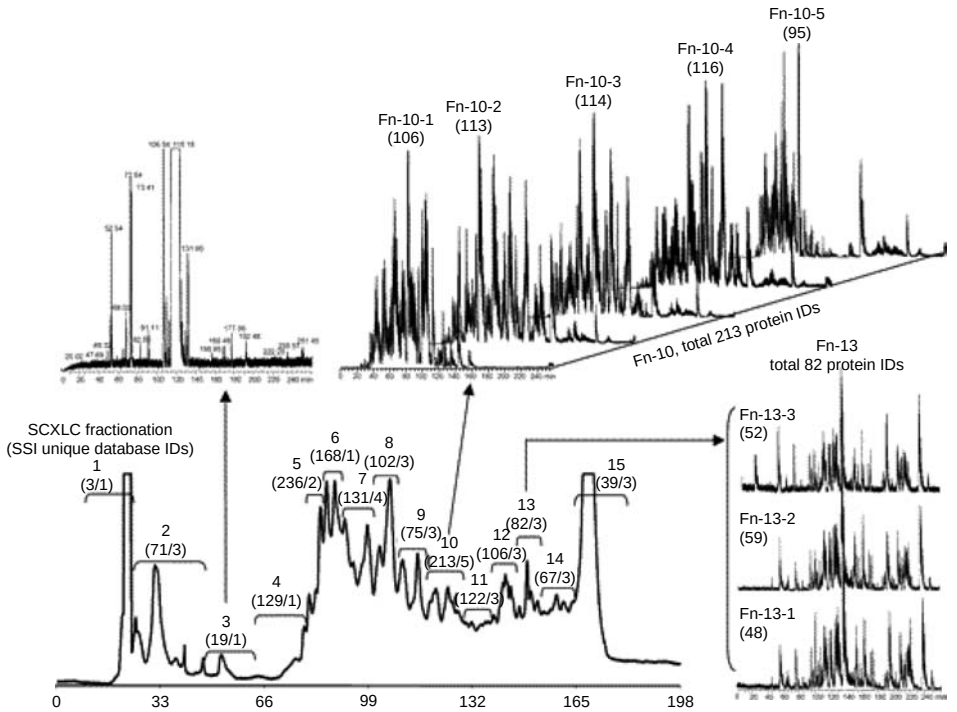

Fig. 28 High-efficiency 2D capillary SCX/RPC separation of total human plasma proteins digested with trypsin. The total plasma tryptic digests were first separated in an 80 cm × 320 µm ID fused-silica capillary packed with polysulfoethyl aspartamide-bonded silica particles (3-µm diameter, 300-Å pore size). A linear gradient of 4 mM phosphate, pH 2.5 to 0.4 M phosphate buffer in 200 min was used to separate the tryptic peptides. The peptides were separated into 15 fractions followed by RPC separation using an 85 cm × 30 µm ID capillary (packed with 3-µm C18 particles, 300-Å pore size). The number of proteins identified is given in parentheses. (Reprinted with permission from Shen, Y. et al. (2004) Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for. . ., *Anal. Chem.* **76**, 1134–44).

considered when constructing such two-dimensional separations.

A combination of size exclusion with adsorptive chromatography, the so-called restricted access chromatography, has been implemented to separate protein mixtures. In this case, only proteins and peptides below a certain size have access to the inner pore surface of the stationary phase and are thus retained while the larger proteins encounter only the hydrophilic, nonadsorptive outer surface and are found in the flow-through. Restricted access materials have found widespread use in the analysis of drug metabolites and other low-molecular weight compounds but have only recently been rediscovered for peptide/protein studies. In an example of applying restricted access chromatography to the analysis of hemofiltrate, a biofluid that is comparable to urine, it was shown that the sample preparation step could be performed in line with a two-dimensional chromatographic separation. This allowed a peak capacity of about 5000 to be reached, a number that is comparable to the number of spots obtained on a highly optimized 2DE gel for

proteins. Such high-resolution, multidimensional systems promise to provide the methodological means for biomarker discovery in combination with modern mass spectrometers and data analysis software.

Yates and coworkers have introduced a serial two-dimensional LC for proteomic analysis. In this approach, a biphasic microcolumn in which a single capillary with 100-μm ID is first packed with 5-μm C18 for 8 cm and then packed with 5-μm SCX for 4 cm, which is similar to those used in serial column chromatography. The two-dimensional separations are achieved by alternatively eluting with salt step and a 60-min linear aqueous acetonitrile gradients through the whole biphasic column. The system is used for direct analysis of large protein complexes, and 95 unique proteins have been identified for the loaded 120-μ g mixture in ~24 h of a single run (12 cycles). After optimization of this multidimensional protein identification technology (MudPIT), it was applied to proteome-wide protein identification. Optimizations included fractionation of a yeast derived proteome sample into three fractions (i.e. soluble, lightly washed, and heavily washed fractions) prior to the tryptic digestion, extension of C18 packed column length to 10 cm, varying mobile phase composition by adding 0.02% heptafluorobutyric acid (HFBA), using shallow gradient, and utilizing 15 fractionation cycles. In ~84 h required for three runs of the three fractions, a combined 5540 peptides were assigned, leading to identification of 1480 proteins. In ~28 h of a single run, ~630 soluble proteins from 1665 assigned peptides were achieved for the soluble protein fraction. By analyzing the data set against the database, 72 out of 231 possible peripheral membrane proteins were detected and identified. The addition of HFBA to the mobile phase has significantly improved the sensitivity and separation resolution in the biphasic column. The separation efficiency of the MudPIT system was evaluated, and a peak capacity of 216 was achieved for the 5-μm C18 packed 10-cm capillary. Combined with 15 cycles (i.e. first stage fractions), a total peak capacity of 3240 was estimated.

7.6
Other Approaches

While off-line capillary isoelectric focusing (CIEF)/capillary RPC separations have been investigated for complex proteomic sample analysis, the use of either on-line CIEF/capillary RPC, capillary isotachophoresis (CITP)/capillary RPC, or capillary LC/capillary zone electrophoresis (CZE) for proteomic analysis has not yet been reported. However, their use is expected because of the relatively large sample capacity and separation efficiency of the first dimension and high separation efficiency supplied by the second dimension. Utilization of these approaches can potentially provide information that makes peptide and protein identification simpler and more accurate.

8
LC-MS for Peptide/Protein Analysis

One of the most significant recent advances in bioanalytical technology is the advent of mass spectrometry for the analysis and measurement of peptide and protein molecular mass. In particular, the development of on-line electrospray mass spectrometry following RPC (LC-ES-MS) has provided a powerful detection system for the rapid analysis of peptide and proteins. The identification of proteins, peptide fragments and various modifications

is essential for understanding biological processes and the function of proteins for normal healthy and at various disease states. The resulting insights lead to the development of therapies for intervention, and ultimately, the cure of disease. The information and knowledge derived from this type of study are extremely valuable for activities involved with target identification activities during drug discovery.

As described in previous sections, chromatographic analysis of peptide and protein is conventionally detected and quantified by the ultraviolet and/or fluorescence detection. However, quantitative methods combining separation techniques with UV or fluorescence detection are of limited applicability when analyzing highly complex biomolecule mixtures. Such samples require high-resolution baseline separations or exceedingly high concentrations of analytes to overcome the impact of the matrix, because of unselective UV/fluorescence detection. Therefore, effective and sensitive peptide/protein detection and quantification from biological samples demands highly selective, robust, and accurate LC-MS methods.

8.1 General Considerations

LC-MS is a versatile combination of a commonly used chromatographic technology and mass spectrometry, a powerful identification tool not only for proteins but for all classes of organic molecules. The coupling of LC with MS has had an enormous impact on small molecule and protein profiling, and has proven to be an important alternative method to two-dimensional gel electrophoresis. In the vast majority of LC–MS methods, the chromatographic conditions employed have significant influence on the operation of the mass spectrometer. This means that the chromatographic conditions will invariably have a significant, molecule-dependent effect on the mass spectrometry conditions required to obtain informational analytical results. Since there are a substantial number of parameters that may influence the final results that are obtained, formal experimental design procedures are increasingly being utilized to determine the experimental conditions that give optimum analytical performance, adequate accuracy and precision at appropriate levels. There are many examples in the literature of methodology that allows this to be developed and achieved without the need to use complex experimental design simply by varying individual factors that are thought to affect the experimental outcome. Examples include the conditions required to give complete chromatographic resolution of the analytes present, the minimum analysis time for a complex mixture or the most intense detector signal from some or all of the analytes. As an analytical method becomes more complex, the number of factors is likely to increase and the likelihood is that the simple approach to experimental design described above will not be successful. In particular, the possibility of interaction between factors that will have an effect on the experimental outcome must be considered and factorial design allows such interactions to be probed. Factors may be classified as ''quantitative'' when they take particular values, for example, concentration or temperature, or ''qualitative'' when their presence or absence is of interest.

For an LC–MS experiment, the ''factors'' could include the composition of the mobile phase employed, its pH and flow rate, the nature and concentration of any mobile-phase additive, for example, buffer

or ion-pair reagent, the makeup of the solution in which the sample is injected, the ionization technique, spray voltage for electrospray, nebulizer temperature for APCI, nebulizing gas pressure, mass spectrometer source temperature, cone voltage in the mass spectrometer source, and the nature and pressure of gas in the collision cell if MS–MS is employed. For quantification, the assessment of results is likely to be on the basis of the selectivity and sensitivity of the analysis, that is, the chromatographic separation and the maximum production of molecular species or product ions if MS–MS is employed. Among the various ionization methods developed for coupling LC to MS, including thermospray, continuous-flow fast atom bombardment, and particle beam techniques, electrospray ionization (ESI) is the most widely used interface technique.

8.2 Capillary RPC-MS

Capillary RPC-MS currently plays a dominant role in proteomic analyses involving peptide mixtures due to its robustness for a variety of peptide properties (highly acidic/basic and hydrophilic/hydrophobic). Following enzymatic digestion of the global whole-proteome proteins, the resultant extremely complex peptide mixtures can be directly analyzed using capillary RPC-MS or MS/MS. Sample fractionation prior to RPC has also been applied in some situations. Peptide identification can be conducted using either accurately measured masses (measured with sufficient accuracy to discriminate the identified peptides from others predicted for a proteome or genome) or, more conventionally, peptide fragments produced by MS/MS. The latter approach can be carried out using relatively simple MS such as ion trap MS, while both approaches can be performed using high-quality MS (such as Fourier transform ion cyclotron resonance (FTICR) MS) with improved identification confidence, sensitivity, and dynamic range. Recent results with smaller ID capillaries have shown that comparable results are achievable with much smaller sample sizes, for example, nanoproteomics with sample size of ~10 ng, equivalent to 100 eukaryotic cells. Relatively low separation efficiencies limit the use of single-dimensional capillary RPC-MS for intact protein analysis to relatively simple proteomic systems. This type of application has been successfully demonstrated by the capillary RPC-FTICR characterization of intact proteins from the large subunit of the yeast ribosome. For the capillary RPC separations, C-5 bonded silica particles (5 μm, 300 Å pores) are used as the stationary phase in a 120 cm × 150 μm ID fused-silica capillary (currently, desirable 3-μm porous particles are not commercially available) and acidic acetonitrile aqueous solutions (0.1% TFA as the acid modifier) are used in the mobile phases. For FTICR measurements, an internal calibrant delivered from a dual ESI source was used to achieve high mass measurement accuracy (but which are not as high as achievable for lower M_r peptides). The relatively high resolution achieved by capillary RPC permitted differentiation of protein isoforms having high structural similarity and proteins from their modified forms, thus facilitating unequivocal protein assignments. By providing information on intact proteins and their modifications, the capability of identifying intact proteins holds promise as a complement to the mass spectrometric approach based on protein digestion.

8.3 LC/MS for Protein Analysis

LC-MS has found widespread use in the analysis of peptides from natural sources or generated by proteolytic digestion of larger proteins. LC-MS has also been applied to large-scale global protein characterization and identification. While analyzing biofluids directly without prior digestion is a definitive option in biomarker discovery, prior digestion gives access to the higher molecular weight proteins; however, the analysis of the digested peptide fragments from larger protein mixtures is much more complex. A dynamic range of 4 orders of magnitude could be reached in the expression level of the proteins present. However, it might be that the order of 1000 proteins in a biofluid is still an underestimate, especially if one includes possible splicing variants and posttranslational modifications. This indicates that analysis of such complex mixtures will likely remain a challenge for some time to come no matter which analytical approach is taken. Application of LC-MS to biomarker discovery is not yet very widespread partly because the method generates large and highly complex data sets that require powerful algorithms and software tools to handle and analyze them. LC-MS generates information about the components in a sample both in terms of their physicochemical properties and their molecular mass. Insights given by chromatography such as hydrophobicity or charge of the analytes may provide additional information that can help in sample comparisons. This additional information is, however, presently not systematically used. Presently, LC-MS is more suitable for initial discovery efforts than for larger clinical validation studies. However, these challenges are being tackled so that LC-MS will likely find more use in the future.

The use of HPLC on-line or off-line is an essential feature for global protein analysis to integrate the removal of buffers and salts (purification) and the separation of analytes (preconcentration) with mass spectrometry. With on-line LC-MS approaches, low flow rates ($<100\ \mu L\ min^{-1}$) have been demonstrated to provide maximum sensitivity with ESI techniques for the analysis of proteins. In this type of analysis, a precolumn split of the mobile phase ($\sim 200\ \mu L\ min^{-1}$) was used to obtain a flow rate of $0.2 \pm 0.5\ \mu L\ min^{-1}$. This microcolumn HPLC method minimizes the column elution volume and maximizes the concentration of peptide that enters the ESI interface of the mass spectrometer. Other examples of low flow approaches include peak parking techniques, where the solvent flow and gradient are stopped and the sample continues to flow into the mass spectrometer. This procedure provides longer sampling times and an improved signal-to-noise ratio because the signal can be averaged for a greater length of time. On-line LC-MS approaches for peptide mapping are also used with ESI-time of flight (TOF) mass spectrometers. A nonporous silica support has also shortened the separation of peptide mixtures in less than 35 s. The resulting chromatographic peak widths are less than 1 s wide. The unique nature of the TOF mass spectrometry allows for the complete analysis of these narrow peaks. Off-line approaches also feature sample clean-up, preconcentration, and the batch elution of peptides. This offline feature is commonly used with MALDI-TOF techniques for the simultaneous analysis of peptide mixtures. Because these methods are decoupled from the actual analysis, highly

automated procedures can be used to prepare and manipulate samples.

9
Future Prospects

The number of applications of HPLC in peptide and protein purification continue to expand at an extremely rapid rate. Solid-phase peptide synthesis and recombinant DNA techniques have allowed the production of large quantities of peptides and proteins, which need to be highly purified. The design of multidimensional purification schemes to achieve high levels of product purity highlight the power of HPLC techniques in the production of peptide and proteins-based therapeutics.

Following purification, mass spectrometry can be used to confirm the structural identity of synthetic peptides, while recombinant proteins require further structural analysis by high-resolution analytical fingerprinting to confirm the amino acid sequence. RPC will therefore continue to be the central method for the characterization and quality control analysis of synthetic peptides and recombinant proteins. Moreover, the coupling of mass spectrometry to allow on-line identification of samples will become routine and continue to expand the analytical power of HPLC. Other coupled techniques such as LC-CE (capillary electrophoresis) and LC-biosensor will also allow more rapid on-line analysis of sample purity and bioactivity.

Other areas of separation technology in which significant advances are anticipated to emerge are in the areas of miniaturization and high-speed analysis to allow the efficient purification of femtomolar to attomolar levels of material. These techniques will have important impact in the discovery of new bioactive peptides and novel proteins as potential candidates for new therapeutics. However, in the postgenomic era, the overwhelming application of HPLC will be in the continuing development of new approaches to proteomic mapping.

Finally, there has been a strong synergy between the biotechnology industry and the field of biomacromolecular HPLC as several significant recent advances in separation technology have been driven by the stringent and continually evolving regulatory requirements of the biotechnology industry. HPLC will therefore continue to be at the heart of the analytical techniques with which scientists in both academia and the pharmaceutical industry must arm themselves to be able to fully characterize the identity, purity and potency of peptides and proteins.

Bibliography

Books and Reviews

Aguilar, M-I. (Ed) (2004) HPLC of Peptides and Proteins; Methods and Protocols, *Methods in Molecular Biology*, Vol. 251, Human Press, Totowa, NJ.

Cutler, P. (Ed) (2003) Protein Purification Protocols, *Methods in Molecular Biology*, Vol. 244, 2nd edition, Human Press, Totowa, NJ.

Mant, C.T., Hodges, R.S. (Eds.) (1991) *High Performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis and Conformation*, CRC Press, Boca Raton, FL.

Gooding, K.M., Regnier, F.E. (Eds.) (1990) *HPLC of Biological Macromolecules: Methods and Applications*, Dekker, New York.

Karger, B.L., Hancock, W.S. (Eds.) (1996) High resolution separation and analysis of biological macromolecules Part A: Fundamentals, *Methods Enzymol.* **270**, entire volume.

Karger, B.L., Hancock, W.S. (Eds.) (1996) High resolution separation and analysis of biological macromolecules Part B: Applications, *Methods Enzymol.* **271**, entire volume.

Primary Literature

Adkins, J.N., Varnum, S.M., Auberry, K.J., Moore, R.J., Angell, N.H., Smith, R.D., Springer, D.L., Pounds, J.G. (2002) Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry, *Mol. Cell Proteomics* **1**, 947–955.

Afeyan, N.B., Gordon, N.F., Mazsaroff, I., Varady, L., Fulton, S.P., Yang, Y.B., Regnier, F.E. (1990) Flow-through particles for the high-performance liquid chromatographic separations of biomolecules: perfusion chromatography, *J. Chromatogr.* **519**, 1–29.

Aguilar, M-I., Hearn, M.T.W. (1991) Reversed Phase and Hydrophobic Interaction Chromatography of Proteins, in: Hearn, M.T.W. (Ed.) *HPLC of Proteins, Peptides and Polynucleotides – Contemporary Topics and Applications*, VCH, Deerfield, FL, pp. 246–290.

Aguilar, M-I., Hodder, A.N., Hearn, M.T.W. (1991) High Performance Ion-exchange Chromatography of Proteins, in: Hearn, M.T.W. (Ed.) *HPLC of Proteins, Peptides and Polynucleotides – Contemporary Topics and Applications*, VCH Publications, Deerfield, FL, pp. 199–245.

Aguilar, M-I., Mougos, S., Boublik, J., Rivier, J., Hearn, M.T.W. (1993) The effect of D-amino acid substitutions on the RP-HPLC behaviour of neuropeptide-Y [18–36] analogues, *J. Chromatogr.* **646**, 53–65.

Albert, K., Bayer, E. (1991) Characterisation of bonded phases by solid-state NMR spectroscopy, *J. Chromatogr.* **544**, 345–370.

Blondelle, S.E., Houghten, R.A. (1992) Design of model peptides having potent antimicrobial activities, *Biochemistry* **31**, 12688–12694.

Blondelle, S.E., Buttner, K., Houghten, R.A. (1992) Evaluation of peptide-peptide interactions using reversed phase high performance liquid chromatography, *J. Chromatogr.* **625**, 199–206.

Buckley, J.J., Wetlaufer, D.B. (1989) Use of the surfactant 3-(3-cholamidopropyl)-dimethylammoniopropane sulfonate in hydrophobic interaction chromatography of proteins, *J. Chromatogr.* **464**, 61–71.

Burke, T.W.L., Mant, C.T., Black, J.A., Hodges, R.S. (1989) Strong cation-exchange high performance liquid chromatography of peptides. Effect of non-specific hydrophobic interactions and linearization of peptide retention behaviour, *J. Chromatogr.* **476**, 377–389.

Carr, S.A., Hemling, M.E., Bean, M.F., Roberts, G.D. (1991) Integration of mass spectrometry in analytical biotechnology, *Anal. Chem.* **63**, 2802–2824.

Chlenov, M.A., Kandyba, E.I., Nagornaya, L.V., Orlova, I.L., Volgin, Y.V. (1993) High performance liquid chromatography of human glycoprotein hormones, *J. Chromatogr.* **631**, 261–267.

Choudhary, G., Horvath, Cs. (1996) Ion-exchange chromatography, *Methods Enzymol.* **270**, 47–82.

Choudhary, G., Wu, S.L., Shieh, P., Hancock, W.S. (2003) Multiple enzymatic digestion for enhanced sequence coverage of proteins in complex proteomic mixtures using capillary LC with ion trap MS/MS, *J. Proteome Res.* **2**, 59–67.

Ding, J., Burkhart, W., Kassel, D.B. (1994) Identification of phosphorylated peptides from complex mixtures using negative-ion orifice-potential stepping and capillary liquid chromatography/electrospray ionisation mass spectrometry, *Rapid Commun. Mass Spectrom.* **8**, 94–98.

Engelhardt, H., Schom, U.M. (1986) Optimal conditions for size exclusion chromatography of proteins, *Chromatographia* **22**, 388–394.

Fang, F., Aguilar, M-I., Hearn, M.T.W. (1996) Influence of temperature on the retention behaviour of proteins in cation-exchange chromatography, *J. Chromatogr. A* **729**, 49–66.

Fausnaugh, J.L., Regnier, F.E. (1986) Solute and mobile phase contributions to retention in hydrophobic interaction chromatography of proteins, *J. Chromatogr.* **359**, 131–146.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M. (1989) Electrospray ionization for mass spectrometry of large biomolecules, *Science* **246**, 64–71.

Frank, J., Braat, A., Duine, J.A. (1987) Assessment of protein purity by chromatography and multiwavelength detection, *Anal. Biochem.* **162**, 65–73.

Frenz, J., Cachia, J., Quan, C.P., Sliwkowski, M.B., Vasser, M. (1993) Protein sorting by high performance liquid chromatography of recombinant deoxyribonuclease I on polyionic stationary phases, *J. Chromatogr.* **634**, 229–239.

Fridman, M., Aguilar, M-I., Hearn, M.T.W. (1990) A comparative study of the equilibrium

refolding of bovine, porcine and human growth hormone using size exclusion chromatography, *J. Chromatogr.* **512**, 57–75.

Fujii, k., Nakano, T., Kawamura, T., Usui, F., Bando, Y., Wang, R., Nishimura, T. (2004) Multidimensional protein profiling technology and its application to human plasma proteome, *J. Proteome Res.* **3**, 712–718.

Gatlin, C.L., Kleemann, G.R., Hays, L.G., Link, A.J., Yates, J.R. (1998) Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography-microspray and nanospray mass spectrometry, *Anal. Biochem.* **263**, 93–101.

Guo, D., Mant, C.T., Taneja, A.K., Hodges, R.S. (1986) Prediction of peptide retention times in reversed phase high performance liquid chromatography. II. Correlation of observed and predicted peptide retention times and factors influencing the retention times of peptides, *J. Chromatogr.* **359**, 519–532.

Hanson, M., Unger, K.K., Mant, C.T., Hodges, R.S. (1992) Polymer-coated reversed-phase packings with controlled hydrophobic properties. I. Effect on the selectivity of protein separations, *J. Chromatogr.* **599**, 65–76.

Hanson, M., Unger, K.K., Mant, C.T., Hodges, R.S. (1992) Polymer-coated reversed-phase packings with controlled hydrophobic properties. II. Effect on the selectivity of peptide separations, *J. Chromatogr.* **599**, 77–86.

Harald, J., Walden, M., Schafer, S., Genz, S., Forssmann, W-G. (2004) Analytical procedures for quantification of peptides in pharmaceutical research by liquid chromatography-mass spectrometry, *Anal. Bioanal. Chem.* **378**, 883–897.

Hearn, M.T.W., Hodder, A.N., Stanton, P.G., Aguilar, M-I. (1987) Evaluation of retention and bandwidth relationships for proteins separated by isocratic anion-exchange chromatography, *Chromatographia* **24**, 769–776.

Heinitz, M.L., Kennedy, L., Kopaciewicz, W., Regnier, F.E. (1988) Chromatography of proteins on hydrophobic interaction and ion-exchange chromatographic matrices: mobile phase contributions to selectivity, *J. Chromatogr.* **443**, 173–182.

Henry, M. (1991) Design requirements of silica-based matrices for biopolymer chromatography, *J. Chromatogr.* **544**, 413–443.

Herold, M., Leistler, B. (1991) Unfolding of truncated and wild type aspartate aminotransferase studied by size exclusion chromatography, *J. Chromatogr.* **539**, 383–391.

Hodder, A.N., Aguilar, M-I., Hearn, M.T.W. (1989) The influence of different displacer salts on the retention properties of proteins separated by gradient anion exchange chromatography, *J. Chromatogr.* **476**, 391–411.

Hodder, A.N., Machin, K.J., Aguilar, M-I., Hearn, M.T.W. (1990) Identification and characterisation of coulombic interactive regions on hen lysozyme by high performance liquid anion-exchange chromatography and computer graphic analysis, *J. Chromatogr.* **517**, 317–331.

Hodges, R.S., Zhu, B-Y., Zhou, N.E., Mant, C.T. (1994) Reversed-phase liquid chromatography as a useful probe of hydrophobic interactions involved in protein folding and protein stability, *J. Chromatogr.* **676**, 3–15.

Horvath, Cs., Melander, W., Molnar, I. (1976) Solvophobic interactions in liquid chromatography with non-polar stationary phases, *J. Chromatogr.* **125**, 129–156.

Hunt, D.F., Henderson, R.A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A.L., Appella, E., Engelhard, V.H. (1992) Characterisation of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry, *Science* **255**, 1261–1263.

Jilge, G., Janzen, R., Unger, K.K., Kinkel, K.N., Hearn, M.T.W. (1987) Evaluation of advanced silica packings for the separation of biopolymers by high performance liquid chromatography. III. Retention and selectivity of proteins and peptides in gradient elution on non-porous monodisperse 1.5 μm reversed phase silicas, *J. Chromatogr.* **397**, 71–80.

Kato, Y. (1993) *Study: Comparison of Ion-Exchange Columns*, Tosoh Corporation, Montgomeryville, PA.

Kopaciewicz, W., Regnier, F.E. (1983) Mobile phase selection for the high performance ion-exchange chromatography of proteins, *Anal. Biochem.* **133**, 251–259.

Kopaciewicz, W., Rounds, M.A., Fausnaugh, J., Regnier, F.E. (1983) Retention model for high-performance ion-exchange chromatography, *J. Chromatogr.* **266**, 3–21.

Lee, T-H., Thompson, P.T., Hearn, M.T.W., Aguilar, M-I. (1997) Conformational stability of a type II' β-turn motif in human growth hormone [6–13] peptide analogues

at hydrophobic surfaces, *J. Pept. Res.* **49**, 394–403..

leMaire, M., Veil, A., Moller, J.V. (1989) Size exclusion chromatography and universal calibration of gel columns, *Anal. Biochem.* **177**, 50–56.

Lin, S., Karger, B.L. (1990) Reversed phase chromatographic behaviour of proteins in different unfolded states, *J. Chromatogr.* **499**, 89–102.

Link, A.J., Eng, J., Schietz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., Yates, J.R.I.I.I. (1999) Direct analysis of protein complexes using mass spectrometry, *Nat. Biotechnol.* **17**, 676–682.

Lu, X.M., Benedek, K., Karger, B.L. (1986) Conformational effects in the high-performance liquid chromatography of proteins. Further studies of the reversed phase chromatographic behaviour of ribonuclease A, *J. Chromatogr.* **359**, 19–29.

Lubman, D.M., Kachman, M.T., Wang, H., Gong, S., Yan, F., Hamler, R.L., O'Neil, K.A., Zhu, K., Buchanan, N.S., Barder, T.J. (2002) Two-dimensional liquid separations-mass mapping of proteins from human cancer cell lysates, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **782**, 183–196.

Mant, C.T., Parker, J.M.R., Hodges, R.S. (1987) Size-exclusion high performance chromatography of peptides. Requirement for peptide standards to monitor column performance and non-ideal behaviour, *J. Chromatogr.* **397**, 99–112.

Marshall, J., Jankowski, A., Furesz, S., Kireeva, I., Barker, L., Dombrovsky, M., Zhu, W., Jacks, K., Ingratta, L., Bruin, J., Kristensen, E., Zhang, R., Stanton, E., Takahashi, M., Jackowski, G. (2004) Human serum proteins preseparated by electrophoresis or chromatography followed by tandem mass spectrometry, *J. Proteome Res.* **3**, 364–382.

Melander, W., Horvath, Cs. (1977) Salt effects on hydrophobic interactions in precipitation and chromatography of proteins: an interpretation of the lyotropic series, *Arch. Biochem. Biophys.* **183**, 200–215.

Melander, W.R., Corradini, D., Horvath, Cs. (1984) Salt-mediated retention of proteins in hydrophobic interaction chromatography, *J. Chromatogr.* **317**, 67–85.

Melander, W.R., El Rassi, Z., Horvath, Cs. (1989) Interplay of hydrophobic and electrostatic interactions in biopolymer chromatography. Effect of salts on the retention of proteins, *J. Chromatogr.* **469**, 3–27.

Mock, K., Hail, M., Mylchreest, I., Zhou, J., Johnson, K., Jardine, I. (1993) Rapid high sensitivity peptide mapping by liquid chromatography-mass spectrometry, *J. Chromatogr.* **646**, 169–174.

Moy, F.J., Li, Y-C., Rauenbeuhler, P., Winkler, M.E., Scheraga, H.A., Montelione, G.T. (1993) Solution structure of human type-α transforming growth factor determined by heteronuclear NMR spectroscopy and refined by energy minimisation with restraints, *Biochemistry* **32**, 7334–7353.

Muller, W. (1990) New ion-exchangers for the chromatography of biopolymers, *J. Chromatogr.* **510**, 133–140.

Nice, E. (1996) Micropreparative high performance liquid chromatography of proteins and peptides: principles and applications, *Biopolymers* **40**, 319–341.

Noinville, V., Vidal-Madjar, C., Sebille, B. (1995) Modelling of protein adsorption on polymer surfaces – computation of adsorption potential, *J. Phys. Chem.* **99**, 1516–1522.

Nyberg, F., Pernow, C., Moberg, U., Eriksson, R.B. (1986) High performance liquid chromatography and diode array detection for the identification of peptides containing aromatic amino acids in studies of endorphin-degrading activity in human cerebrospinal fluid, *J. Chromatogr* **359**, 541–551.

Oroszlan, P., Wicar, S., Teshima, G., Wu, S-L., Hancock, W.S., Karger, B.L. (1992) Conformational effects in the reversed phase chromatographic behaviour of recombinant human growth hormone (rhGH) and N-methionyl recombinant human growth hormone (met-hGH), *Anal. Chem.* **64**, 1623–1631.

Poll, D.J., Harding, D.R.K. (1989) Formic acid as a milder alternative to trifluoroacetic acid and phosphoric acid in two dimensional peptide mapping, *J. Chromatogr.* **469**, 231–239.

Potschka, M. (1987) Universal calibration of gel permeation chromatography and determination of molecular shape in solution, *Anal. Biochem.* **162**, 47–64.

Purcell, A.W., Aguilar, M-I., Hearn, M.T.W. (1995) Conformational effects in the RP-HPLC of polypeptides. I: The resolution of insulin variants, *J. Chromatogr.* **711**, 61–70.

Purcell, A.W., Aguilar, M-I., Wettenhall, R.E.H., Hearn, M.T.W. (1995) The induction of

amphipathic helical structures in peptides during RP-HPLC, *J. Pept. Res.* **8**, 160–170.

Randhawa, Z., Witkowska, H.E., Cone, J., Wilkins, J.A., Hughes, P., Yamanishi, K., Yashuda, S., Masui, Y., Arthur, P., Kletke, C., Bitsch, F., Shackleton, C.H.L. (1994) Incorporation of norleucine at methionine positions in recombinant human macrophage colony stimulating factor (m-CSF,4–153), *Biochemistry* **33**, 4352–4362.

Regnier, F.E. (1987) The role of protein structure in chromatographic behaviour, *Science* **238**, 319–323.

Richards, K.L., Aguilar, M-I., Hearn, M.T.W. (1994) A comparative study of the retention behaviour and stability of cytochrome c in RP-HPLC, *J. Chromatogr.* **676**, 17–31.

Roth, C.M., Unger, K.K., Lenhoff, A.M. (1996) Mechanistic model of retention in protein ion-exchange chromatography. Reversed phase chromatographic behaviour of proteins in different unfolded states, *J. Chromatogr. A* **726**, 45–56.

Shen, Y., Jacobs, J.M., Camp, D.G., Fang, R., Moore, R.J., Smith, R.D., Xiao, W., Davis, R.W., Tompkins, R.G. II (2004) Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome, *Anal. Chem.* **76**, 1134–1144.

Shen, Y., Tolic, N., Masselon, C., Pasa-Tolic, L., Camp, D.G., Hixon, K.K., Zhao, R., Anderson, G.A., Smith, R.D. II (2004) Ultrasensitive proteomics using high efficiency on-line micro-SPE-nanoLC-nanoESI MS and MS/MS, *Anal. Chem.* **76**, 144–154.

Simpson, R.J., Moritz, R.L., Nice, E.C., Grego, B. (1987) A high performance liquid chromatography procedure for recovering subnanomole amounts of protein from SDS-gel electroeluates for gas-phase sequence analysis, *Eur. J. Biochem.* **165**, 21–29.

Snyder, L. (1980) Gradient Elution, in: Horvath, Cs. (Ed.) *HPLC-Advances and Perspectives*, Vol. 1, Academic Press, New York, pp. 207–316.

Snyder, L.R. (1996) Automated method development in high performance liquid chromatography, *Methods Enzymol.* **270**, 151–175.

Stadalius, M.A., Gold, H.S., Snyder, L.R. (1984) Optimisation model for the gradient elution separation of peptide mixtures by reversed phase high performance liquid chromatography, *J. Chromatogr.* **296**, 31–59.

Stahlberg, J., Jonsson, B. (1996) Influence of charge regulation in electrostatic interaction chromatography of proteins, *Anal. Chem.* **68**, 1536–1544.

Stout, R.W., Sivakoff, S.I., Ricker, R.D., Snyder, L.R. (1986) Separation of proteins by gradient elution from ion-exchange columns. Optimizing experimental conditions, *J. Chromatogr.* **353**, 439–463.

Sun, L., Carr, P.W. (1995) Chromatography of proteins using polybutadiene-coated zirconia, *Anal. Chem.* **67**, 3717–3721.

Tanaka, N., Kimata, K., Mikawa, Y., Hosoya, K., Araki, T., Ohtsu, Y., Shiojima, Y., Tsuboi, R., Tsuchiya, H. (1990) Performance of wide-pore silica- and polymer-based packing materials in polypeptide separation: effect of pore size and alkyl chain length, *J. Chromatogr.* **535**, 13–32.

Thevenon, G., Regnier, F.E. (1989) Reversed-phase liquid chromatography of proteins with strong acids, *J. Chromatogr.* **476**, 499–511.

Uversky, V.N. (1993) Use of fast protein size exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule, *Biochemistry* **32**, 13288–13298.

Velayudhan, A., Horvath, Cs. (1986) On the stoichiometric model of electrostatic interaction chromatography for biopolymers, *J. Chromatogr.* **367**, 160–162.

Washburn, M.P., Wolters, D., Yates, J.R.I.I.I. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology, *Nat. Biotechnol.* **19**, 242–247.

Welinder, B.S. (1991) Use of polymeric reversed-phase columns for the characterisation of polypeptides extracted from human pancreata. II. Effect of the stationary phase, *J. Chromatogr.* **542**, 83–99.

Welinder, B.S., Sorenson, H.H., Hansen, B. (1986) Reversed-phase high performance liquid chromatography of insulin. Resolution and recovery in relation to column geometry and buffer components, *J. Chromatogr.* **361**, 357–363.

Welling, G.W., Van der Zee, R., Welling-Wester, S. (1987) Column liquid chromatography of integral membrane proteins, *J. Chromatogr.* **418**, 223–243.

Wetlaufer, D.B., Koenigbauer, M.R. (1986) Surfactant-mediated protein hydrophobic

interaction chromatography, *J. Chromatogr.* **359**, 55–60.

Wilce, M.C.J., Aguilar, M-I., Hearn, M.T.W. (1995) The physicochemical basis of hydrophobicity scales: Evaluation of four new sets of amino acid coefficients derived from reversed phase high performance liquid chromatography, *Anal. Chem.* **67**, 1210–1219.

Wirth, H-J., Eriksson, K-O., Holt, P., Aguilar, M-I., Hearn, M.T.W. (1993) Ceramic based particles as chemically stable chromatographic supports, *J. Chromatogr.* **646**, 129–141.

Wu, S-L., Benedek, K., Karger, B.L. (1986) Thermal behaviour of proteins in high performance hydrophobic-interaction chromatography. On-line spectroscopic and chromatographic characterisation, *J. Chromatogr.* **359**, 3–17.

Wu, S-L., Figueroa, A., Karger, B.L. (1986) Protein conformational effects in hydrophobic interaction chromatography. Retention characterisation of and the role of mobile phase additives and stationary phase hydrophobicity, *J. Chromatogr.* **371**, 3–27.

Yarovsky, I., Aguilar, M-I., Hearn, M.T.W. (1995) Influence of the chain length and surface density on the conformation and mobility of *n*-alkyl ligands chemically immobilised to a silica surface, *Anal. Chem.* **67**, 2145–2153.

Young, P.M., Wheat, T.E. (1990) Optimisation of high performance liquid chromatographic peptide separations with alternative mobile and stationary phases, *J. Chromatogr.* **512**, 273–228.

Zhou, N.E., Mant, C.T., Hodges, R.S. (1990) Effect of preferred binding domains on retention behaviour in reversed-phase chromatography, *J. Pept. Res.* **3**, 8–20.

Zhou, N.E., Mant, C.T., Kirkland, J.J., Hodges, R.S. (1991) Comparison of silica-based cyanopropyl and octyl reversed phase packings for the separation of peptides and proteins, *J Chromatogr.* **548**, 179–193.

Human and Veterinary Classical Vaccines against Bacterial Diseases

Thomas Ebensen, Claudia Link, and Carlos A. Guzmán
GBF-German Research Centre for Biotechnology, Braunschweig, Germany

1 Bacterial Vaccines: Introductory Remarks 299

2 Inactivated Vaccines 300
2.1 Methods of Inactivation 301
2.2 Advantages and Limitations of Inactivated Vaccines 301

3 Live Vaccines 302
3.1 Attenuation 302
3.2 Advantages and Limitations of Live Bacterial Vaccines 303

4 Vaccines for Human Bacterial Diseases 304
4.1 Anthrax (*Bacillus anthracis*) 304
4.2 Cholera (*Vibrio cholerae*) 304
4.3 Enterotoxigenic *Escherichia coli* 305
4.4 Plague (*Yersinia pestis*) 306
4.5 Shigellosis (*Shigella* species) 306
4.6 Tuberculosis (*Mycobacterium tuberculosis*) 307
4.7 Typhoid Fever (*Salmonella enterica* serovar Typhi) 308
4.8 Tularemia (*Francisella tularensis*) 309
4.9 Whooping Cough (*Bordetella pertussis*) 309

5 Veterinary Bacterial Vaccines 310
5.1 Infections Caused by *Bordetella* and *Pasteurella* species 311
5.2 Brucellosis (*Brucella* spp.) 312
5.3 Porcine Pleuropneumonia (*Actinobacillus pleuropneumoniae*) 313
5.4 Diseases Caused by *Mycoplasma* spp. 314
5.5 Salmonellosis in Animals 315
5.6 Leptospirosis (*Leptospira* spp.) 315
5.7 Other Commercially Relevant Animal Diseases 316

6 Conclusions 316

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

Acknowledgments 317

Bibliography 317
Books and Reviews 317
Primary Literature 317

Keywords

DNA Vaccine
Vaccine formulation in which eukaryotic expression vectors coding for the antigens of interest are given to the vaccines.

Inactivated Vaccine
Vaccine formulation prepared using killed organisms.

Live Vaccine
Vaccine formulation prepared using a live attenuated organism, which is not able to cause disease (i.e. nonvirulent organism).

Subunit Vaccine
Vaccine formulation prepared using purified subcellular components from pathogenic organisms.

Vaccine
Preparation administered to humans or animals, which is able to stimulate a protective immune response against a pathogenic organism.

Vaccination
Administration of a vaccine preparation to humans or animals.

Whole-cell vaccine
Vaccine formulation prepared using an intact organism (e.g. bacteria, virus).

Immunization, also called *vaccination*, is the most cost-effective intervention to prevent infectious diseases. This is achieved by stimulating the production of pathogen-specific antibodies and/or cellular immune responses. Vaccines have been mainly used for the prophylaxis of infections, that is, they are administered to a large number of healthy individuals to confer protection against disease. However, the development of therapeutic vaccines against infectious and noninfectious diseases is gaining interest. Classic antibacterial vaccines, which will be the main subject of this article, have played a critical role in the efficient control of major diseases, both in humans and animals. These classic vaccines were mainly generated using inactivated or live attenuated organisms. More recently, a new generation of well-defined, safer, and effective vaccines has emerged, which is based on the use of recombinant technology and/or subcellular components.

1
Bacterial Vaccines: Introductory Remarks

The successful control of bacterial infections depends on two opposing forces: on the one hand, host-specific and unspecific defense mechanisms, which are designed to restrict bacterial survival and multiplication; on the other hand, the coordinated expression of bacterial virulence factors, which are responsible for their subversion and/or circumvention. The host immune system provides two different lines of defense against infectious agents. The first nonspecific defense line is given by the innate immune system, and the pathogen-specific adaptive immune response constitutes the second. Cellular components of the innate immune system also represent the critical bridge between these two responses (e.g. dendritic cells).

Vaccination constitutes the most cost-efficient strategy for preventing bacterial diseases through the stimulation of pathogen-specific adaptive responses, which are able to protect against subsequent challenge. Most bacterial pathogens are either restricted to the mucosae or need to transit across them to cause disease. Thus, elicitation of a local mucosal immune response after vaccination represents a clear advantage. In fact, the stimulation of a potent local response at the site at which the first line of defense is laid allows promoting protection against both disease and infection (i.e. colonization), thereby reducing the risk of pathogen transmission to other susceptible hosts.

Human vaccination against bacterial diseases has been practiced with success for many years. Despite this success, however, human society continues to suffer millions of deaths and multibillion dollar economic losses annually due to bacterial diseases. This is in part due to the lack of vaccines against many diseases, as well as to the fact that not all available vaccines are completely satisfactory in terms of immunogenicity, efficacy, and safety. In addition, not all efficacious vaccines are used properly. Antivaccine movements in developed countries and economic constraints in the developing world prevent optimal vaccine use. To reduce these losses, there is a critical need to develop new, better, safer, and cheaper vaccines, which are able to promote long-lasting immune responses.

Various strategies have been exploited to develop antibacterial vaccines (Fig. 1). Whole-cell vaccines are produced by inactivating the virulent organisms (e.g. by chemicals, heat). Alternatively, bacteria can be attenuated in such a way that the organisms are not able to cause disease in their natural hosts, but are still able to induce protective immunity. Thus, attenuated organisms are not able to produce clinical disease, but they retain the ability for transient colonization and/or self-limiting replication. These attenuated bacteria can also be used as live carriers (i.e. bacterial vectors) for the delivery of heterologous antigens from other pathogenic organisms. It is also possible to use organisms from a related but nonpathogenic species, as an alternative to attenuated organisms. However, it is essential that they share enough similarity to promote a cross-protective immune response. In other cases, mixed approaches have been established, such as the use of a microorganism from a different species that has been attenuated (e.g. attenuated *Mycobacterium bovis* BCG to protect against tuberculosis). Purified subcellular components from pathogenic

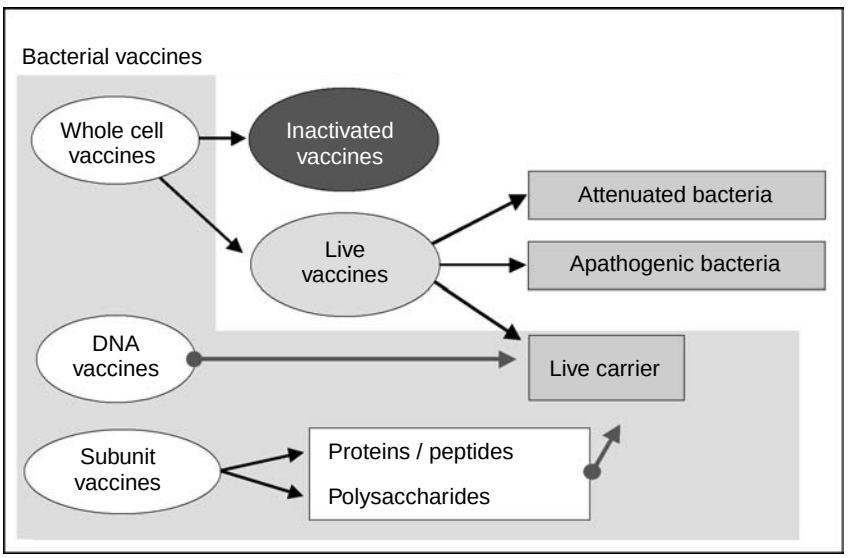

Fig. 1 Main vaccination strategies.

bacteria, which are critical targets for the elicitation of protective responses, can also be used to develop subunit vaccines. Subunit vaccine components can be inactivated toxins, capsular polysaccharides, structural proteins, or nonstructural components. The antigens used can be native or obtained by recombinant DNA technology. Subunit vaccines usually exhibit a better safety profile than whole-cell vaccines. However, they are often less immunogenic, have higher production costs, and demand expensive storage procedures.

A new approach has emerged in the vaccinology field in the last decade, the so-called *DNA vaccines*. This approach consists of the use of eukaryotic expression vectors encoding the antigen of interest, which are delivered to the vaccinees. Then, the biosynthetic machinery of the eukaryotic cells is responsible for synthesis of the vaccine-specific antigens. More recently, attenuated bacteria have been successfully exploited as a delivery system for DNA vaccines against both infectious agents and tumors. The availability of different mutants and highly sophisticated expression tools, as well as the possibility to coadminister immune modulators (e.g. cytokines) allow predictable immune responses to be triggered, according to the specific needs.

2 Inactivated Vaccines

Many vaccines against human or animal infectious diseases are composed of whole organisms, which have been inactivated to prevent *in vivo* multiplication, but retain their immunogenicity. Many of them have been empirically designed, without having a clear understanding of either the protective antigens or the immune protective mechanisms. Furthermore, these vaccines are generally associated with a high incidence of side effects.

2.1 Methods of Inactivation

Whole-cell vaccines are usually inactivated by heating and/or treating with chemicals, and the final vaccine preparation does not usually undergo further purification. Formaldehyde was the reagent first used for chemical inactivation, but glutaraldehyde and phenol have also been used. However, this procedure may affect the immunogenic epitopes present in the antigens. The use of colicin E2, a potent DNA endonuclease that enters target bacterial cells without disrupting cellular integrity, has also been proposed for bacterial inactivation.

A special inactivation approach is based on the use of phage lysis systems. The controlled expression of PhiX174 gene E in gram-negative bacteria results in the formation of a transmembrane tunnel through the cell envelope complex. The resulting so-called *bacterial ghosts* have intact envelope structures, but are devoid of cytoplasmic contents. This process was successfully exploited in a variety of gram-negative bacteria, including *Escherichia coli*, *Salmonella* spp., *Vibrio cholera*, and *Actinobacillus pleuropneumoniae*. The well-preserved surface structures of bacterial ghosts stimulate efficient immune responses when administered by various routes. This seems to be due, at least in part, to the specific targeting of antigen-presenting cells, in which they promote maturation through the potent danger signal delivered by their structural components (e.g. lipopolysaccharides). Ghosts can be used not only as vaccines against diseases caused by the inactivated microorganisms but also as a technology platform for antigen delivery. In this approach, they can be directly loaded with antigens or, alternatively, recombinant antigens can be expressed before lysis.

2.2 Advantages and Limitations of Inactivated Vaccines

Efficient bacterial inactivation leads to vaccine candidates with an extremely good safety profile, which can be administered even to immune-deficient individuals. However, nonreplicating organisms are generally less immunogenic than live bacteria, and multiple doses are required (Table 1). On the other hand, all virulence factors can be expressed by the strain used in the preparation of the vaccine, whereas key virulence factors may be missing in the attenuated vaccine strain. However, *in vitro* cultivation of bacteria does not necessarily lead to the expression of all virulence factors, whereas live vaccines are capable of eliciting immune responses also against *in vivo* expressed antigens. The reactogenicity of inactivated and live vaccines is usually greater than that of other types of vaccines (e.g. subunit vaccines). Vaccines may loose efficacy as a result of accumulated mutations in circulating strains of pathogens. This risk is lower for inactivated than for subunit vaccines, due to the presence of all potential antigens from the pathogen. However, the presence of multiple antigens may result in reduced vaccine efficacy, due to the suppression of immune responses against subdominant antigens by the immune dominants. In contrast to live vaccines, in which the immune response closely resembles that against natural infection, the immune responses against parenterally administered inactivated vaccine are limited to the systemic compartment. In addition, humoral rather than cellular responses are stimulated. Finally, inactivated vaccines are not

Tab. 1 Comparison of live and killed whole-cell vaccines.

Live/attenuated vaccines	*Killed/inactivated vaccines*
Attenuated microorganisms are used	Virulent bacteria can be used
Replicate in the host	Do not replicate
Single or few doses are required	Multiple doses are required
Smaller number of microorganisms	Larger number of microorganisms
Less stable	More stable
No adjuvants are required	Adjuvants often required
Can be delivered by natural infection route	Generally given by parenteral route
Stimulate antibody and cellular responses	Induce good humoral but poor cellular responses
Promote cellular and humoral memory	Stimulate weaker cellular memory
Sometimes spread from vaccinated to unvaccinated individuals	No spread to unvaccinated individuals
Variable safety in immunocompromised patients	Excellent safety in immunodeficient patients
Shorter shelf-life	Usually longer shelf-life
Easy transport and storage	Easy transport and storage
Higher development costs	Lower development costs

shed; thus, they are unable to infect unvaccinated individuals. On the one hand, they do not lead to herd immunity. On the other hand, they do not represent a risk for immunosuppressed contacts. Finally, the development costs of inactivated vaccines are usually lower than those of live or subunit vaccines. However, quality control (e.g. potency tests, killing tests) of inactivated vaccines may be very difficult.

3 Live Vaccines

Live vaccines lead to a self-limited asymptomatic infection, thereby stimulating similar immune responses to those observed after natural infections. Therefore, this kind of vaccine closely matches some of the criteria for an ideal vaccine, since it stimulates long-lasting protection with minimal reactogenicity after a single or a few doses.

3.1 Attenuation

Attenuation of the microorganism eliminates its disease-causing capacity and can be obtained by biological or technical manipulations (e.g. multiple passages through unusual hosts, growth at suboptimal conditions, chemical mutagenesis). In the past, attenuation was performed by empirical techniques. However, the explosive development of our knowledge in the fields of microbial pathogenesis and recombinant DNA technology allows the identification of molecular targets for attenuation and thereby enables rational and precise construction of vaccine strains containing well-defined deletions in these genes. In strains generated by recombinant DNA technology, the attenuation should be sufficient to limit pathogenicity without risk of reverting to the wild-type phenotype. Thus, it is desirable to introduce at least two independent attenuating mutations. The attenuation should not rely

on a fully functional host immune system. Furthermore, it must not be reversible by diet or by host modification of diet constituents, including the host's resident microbial flora. Finally, it is also necessary that the attenuation does not induce a persistent carrier state in vaccinees.

3.2 Advantages and Limitations of Live Bacterial Vaccines

Live vaccines have several advantages and disadvantages in comparison to subunit vaccines (Table 2). Although live attenuated vaccines replicate in the host, subunit or inactivated vaccines are unable to replicate. Therefore, attenuated vaccines usually promote long-lasting immune responses, which mimic those obtained after natural infections. However, the presence of preexisting or cross-reactive immune responses against the vaccine strain may affect its overall efficacy. Live vaccines are less stable and may be easily damaged or destroyed by heat or light. However, cold storage is required for all vaccines registered up to now. Depending on the degree of attenuation and the specific vaccine, multiple doses may be required. The risk of reversion to the wild-type phenotype, which was certainly an issue with the undefined first-generation attenuated vaccines, is now made negligible by the introduction of multiple well-defined independent deletions in second-generation vaccines. However, live vaccines can generally be considered more reactogenic. This also depends on the route of administration, since oral attenuated vaccines against typhoid fever are less reactogenic than parenteral ones. A specific issue to be taken into consideration is the potential risk associated with administration or transmission (i.e. shedding) to immunodeficient individuals. In addition, a careful and exhaustive evaluation of the potential impact of environmental release, including duration and rate of shedding, as well as the risk of horizontal gene transfer, needs to be carried out. A clear advantage of attenuated vaccines is the fact that they have lower production costs than subunit vaccines. In addition, technology transfer to developing countries is facilitated by the simplicity of the upstream and downstream processes. They are also associated with much simpler and less expensive

Tab. 2 Comparison of live and subunit vaccines.

Live vaccines	*Subunit vaccines*
Live/attenuated microorganisms	Subcellular components of microorganisms
Replicate in the host	Do not replicate in the host
Variable safety	Excellent safety
Can spread to unvaccinated individuals	Not transmissible to contacts
Potential risk of reversion	No risk of reversion
Potential risk to immunosuppressed patients	No risk to immunosuppressed patients
Usually more reactogenic	Usually less reactogenic
Stimulate humoral and cellular responses	Trigger mainly humoral responses
May require fewer doses	Multiple doses are required
Generally long-lasting protection	Periodic boosters required
Lower production costs	Higher production costs
Easy administration logistics	More complicated/expensive administration logistics

administration logistics when they are given via the mucosal route. The mucosal route also provides better safety, since parenteral vaccination is an important cause of spreading of HIV and hepatitis in the developing world.

4
Vaccines for Human Bacterial Diseases

Bacterial infections play a major role in human disease. Although the development and use of anti-infective agents has resulted in reduced morbidity and mortality, increases in antibiotic resistance complicate the clinical management of infected patients. Furthermore, the implementation of therapeutic approaches does not reduce the human suffering and costs associated with disease. Vaccines have played a major role in the fight against bacterial diseases that have cost mankind heavily in terms of human lives, such as diphtheria, tetanus, whooping cough, meningitis, and tuberculosis. Thus, it is necessary to foster the use of available vaccines and to develop new ones against those diseases for which no vaccine is yet available.

4.1
Anthrax (*Bacillus anthracis*)

Anthrax is primarily a disease of herbivorous animals acquired via the ingestion of *Bacillus anthracis* spores. Humans occasionally acquire the disease through contact with infected animals or contaminated animal products via the skin (95%) or by ingestion or inhalation. Without treatment, the cutaneous forms of anthrax have 20% mortality, whereas gastrointestinal and pulmonary anthrax reach mortality rates of 25 to 75% and more than 80% respectively. The emerging threat of bioterrorism has awakened the general interest in both the pathogenesis of *B. anthracis* and vaccine development.

Factors contributing to the pathogenesis process are the plasmid (pXO1 and pXO2) – encoded capsule components, the protective antigen, the lethal factor (LF), and the edema factor (EF). The capsule helps the bacterium to evade immune clearance, and the toxins are responsible for inducing edema and shock. Anthrax was the first bacterial disease for which vaccination was successfully demonstrated. In 1881, Pasteur observed protection in sheep after injection of a heat-attenuated *B. anthracis* strain. In 1937, the Sterne vaccine strain was developed, which is nonencapsulated (cured from pXO2) but retains the toxins. The worldwide implementation of this vaccine has contributed to the successful control of the disease in livestock and wild animals. However, it cannot be used in sensitive animals, due to its residual virulence. More recently, several live vaccine candidates lacking EF and LF were developed. However, they were considered not suitable for human use, since they still retain a certain degree of virulence. In 1970, an anthrax vaccine was licensed for human use, which consists of a sterile filtrate of microaerophilic cultures of a nonencapsulated avirulent strain.

4.2
Cholera (*Vibrio cholerae*)

Cholera is a serious water- and foodborne epidemic disease caused by *V. cholerae*, which has killed millions of people and continues to be a major health problem worldwide. Bacteria colonize the small intestine and produce an exotoxin, cholera toxin (CT), consisting of an enzymatically active subunit A and five identical

B subunits. CT disrupts the function of ion pumps, leading to changes in the net flows of sodium, chloride, and water, and resulting in massive diarrhea, electrolyte imbalance, and circulatory collapse. According to the lipopolysaccharide (LPS), two serogroups are circulating (O1 and O139). O1 organisms are classified as classical or El Tor biotypes, according to the phenotype. *V. cholerae* O1 El Tor biotype is the predominant cause of cholera worldwide; however, a minority of cholera cases are caused by *V. cholerae* O139.

Convalescents develop a protective immune response, as demonstrated by both volunteer studies and patients follow-up. Serological responses against LPS and CT are observed after a cholera episode, and can block colonization and neutralize CT. The first recorded attempt to develop a vaccine against cholera was in the late 1880s. This killed whole-cell vaccine was injected into the bloodstream, but failed to elicit protective immunity and was associated with high rates of side effects. However, a phenol-killed *V. cholerae* Ogawa and Inaba vaccine administered by parenteral route is licensed, which stimulates high titers of vibriocidal antibodies in 50 to 90% of vaccinees and shows an efficacy of about 50% for about 6 months.

To obtain a good local immune response in the gut, an oral vaccine based on heat-killed classic Inaba and Ogawa and formalin-killed El Tor Inaba and classic Ogawa, coadministered with 1 mg of the CT B subunit (CTB), has been developed and showed 60 to 85% efficacy in humans. Another oral vaccine is based on a genetically attenuated *V. cholerae* O1 classical Inaba strain (CVD 103-Hg®), contains a deletion in the CT enzymatically active subunit, and has been marked with a gene coding for mercury resistance that permits differentiation between recombinant and wild-type circulating isolates. This vaccine was licensed in Switzerland in 1994 and is now also available in several other countries. Placebo-controlled trials in several countries have shown the safety and immunogenicity of a single dose of CVD 103-Hg®, even in HIV-infected individuals. Side effects, such as mild nausea, abdominal cramping, and diarrhea, were rarely observed. A single dose conferred 95 and 65% protection against *V. cholerae* classical and El Tor, respectively, in volunteers in challenge studies. Another candidate vaccine is the *V. cholerae* O1 El Tor Inaba Peru-15 strain, which was obtained by incorporating a series of genetic deletions and modifications. A seroconversion rate of 97% for vibriocidal antibodies was observed, and none of the volunteers showed moderate or severe diarrhea after challenge with *V. cholerae* O1 El Tor Inaba strain. None of these vaccines is expected to confer protection against *V. cholerae* O139. However, candidates based on these approaches against O139-expressing *V. cholerae* are being developed.

4.3 Enterotoxigenic *Escherichia coli*

Enterotoxigenic *E. coli* (ETEC) is the most common cause of traveler's diarrhea, which is acquired by ingestion of contaminated food or water. Whole-cell vaccines, which mimic natural infections, seem to promote longer immunoprotection and are being evaluated in clinical trials. A nonreplicating whole-cell anti-ETEC vaccine was prepared by treating the enterotoxigenic *E. coli* strain H-10407 (O78:H11, CFA/I), which produces heat-stable and heat-labile (LT) toxins, with colicin E2. After two oral doses, 77.3 and 86.4% of the volunteers showed anti-CFA/I and anti-LT sIgA responses, respectively, and

90% showed antibodies to CFA/I, LT, or both. On the other hand, a formalin-killed oral ETEC vaccine expressing the most common colonization-factor antigens (i.e. CFA/I, CFA/II, and CFA/IV) and recombinant CTB was given to volunteers in Sweden. An increment in IgA-secreting cells and in the levels of intestinal sIgA against CTB and various CFA components were detected in the majority of volunteers after two doses. Oral vaccination with a killed ETEC plus CTB vaccine in Egyptian children evoked a fourfold rise in antitoxic IgA and IgG titers in 93 and 81% of the vaccinees. The vaccine was also safe and immunogenic in 2- to 12-year-old children. Besides the use of whole-cell *E. coli* vaccines, other approaches have been pursued to stimulate efficient anti-ETEC immune responses, such as the use of purified fimbriae, toxoids, or carrier organisms expressing ETEC antigens.

4.4 Plague (*Yersinia pestis*)

Yersinia pestis is the etiological agent of plague, a disease that has caused over 200 million human deaths. Plague is primarily a disease of rodents, but it can also affect humans. Three main forms of plague occur in humans: bubonic, septicemic, and pneumonic. The common form is bubonic plague, which arises following a bite from a flea that previously fed on an infected animal. The most-feared form is pneumonic plague, which is transmitted from person to person via droplets and has mortality rates of up to 100%. To prevent the disease, both killed whole-cell and live attenuated vaccines have been used. Killed *Y. pestis* bacteria have been used as veterinary vaccines since 1897, and the first killed vaccine for human use was developed in 1946. Various methods have been exploited for inactivation, including formaldehyde and heat treatment. The Commonwealth Serum Laboratories are currently producing a heat-killed *Y. pestis* vaccine (strain 195/P), which is given subcutaneously in three doses over a period of 2 months. This vaccine is mainly used in individuals who may be exposed to the pathogen, for example, veterinarians, researchers, and persons who are deployed to work in areas where the disease is endemic. However, 10% of immunized individuals show side effects such as malaise, headache, fever, and lymphadenopathy. In addition, no clinical trials have demonstrated the efficacy of killed whole-cell vaccines, and some studies suggest that killed vaccines do not provide protection against pneumonic plague.

A live attenuated strain of *Y. pestis* (EV76) was derived from a fully virulent strain by *in vitro* passages. The vaccine has been in use since 1908, especially in the former Soviet Union and the French colonies. Studies in mice indicate that this vaccine provides protection against bubonic and pneumonic plague. However, the safety of this vaccine in humans is questionable. In recent years, we have seen a renewed interest in the development of efficient and safer vaccines against this old pathogen, due to the potential risk of illegitimate use of *Y. pestis* as a biological weapon.

4.5 Shigellosis (*Shigella* species)

Shigellosisis a diarrheal disease, which is transmitted via contaminated food and water or through person–to–person contact. As few as 10 organisms can cause infection, leading to diarrhea or dysentery with frequent mucous bloody stools and severe abdominal cramps. Shigellosis is caused by four main species,

Shigella flexneri, the most common in endemic areas; *Shigella dysenteriae* type 1, most frequent in overcrowded areas and associated with large epidemic outbreaks; *Shigella sonnei*, more active in developed countries; and *Shigella boydii*. Infections caused by *S. dysenteriae* type 1 can also lead to severe complications, such as hemolytic-uremic syndrome. Between 1966 and 1997, the annual number of *Shigella* episodes worldwide was estimated to be 164.7 million, 69% of them being children under the age of 5. Approximately 500 000 deaths occur each year due to this disease.

Soon after isolation of *S. dysenteriae* in 1898, heat-killed cultures were tested to assess their potential as vaccines. However, there is still no vaccine available to prevent *Shigella* infections. In 1950, the first efforts to develop an attenuated vaccine strain were undertaken. The strains T32-ISTRATI, which was isolated after 32 passages on nutrient agar, and *S. flexneri* 2a SmD, which was dependent on streptomycin, yielded promising results, but they were associated with side effects. Subsequent efforts were focused on the use of a hybrid *E. coli* K-12 containing the invasion plasmid that codes for somatic antigens of *S. flexneri* 2a (*E. coli* EcSf2a-2 hybrid). More recently, work has addressed the development of recombinant strains containing mutations in the genes responsible for the biosynthesis of essential aromatic compounds (e.g. *aroA, aroD*) and/or virulence factors (*icsA, iuc, virG, sen, set, stxAB*). These mutants undergo only minimal intracellular replication or are deficient in intracellular and cell-to-cell spread. Promising results were obtained in clinical trials with volunteers after vaccination with some of these candidates. Finally, subunit vaccines, such as O antigen coupled to a carrier protein (conjugate vaccines), vesicles of *Neisseria* outer-membrane proteins and O-polysaccharide (proteosome vaccines), and noncovalent complexes of O-polysaccharide and ribosomal particles (ribosomal vaccines) also gave encouraging results in preclinical and clinical studies.

4.6 Tuberculosis (*Mycobacterium tuberculosis*)

Mycobacterium tuberculosis remains a major health problem worldwide, with about 2 million people dying of tuberculosis each year. *M. bovis* Bacillus Calmette–Guérin (BCG), is the only available vaccine against tuberculosis and constitutes the most widely used live bacterial vaccine in the world. The original strain was attenuated by performing >200 successive *in vitro* passages over 13 years. This strain was then distributed to several laboratories worldwide, where it was produced and maintained. Of the vaccines currently in use, more than 90% belong to 4 main strains: the French Pasteur strain 1173 P2, the Danish strain 1331, the Glaxo strain 1077, and the Tokyo strain 172. Despite the WHO's attempts to standardize production and vaccine characteristics, the concentration ranges from 50 000 to 3 million bacteria per dose. Some vaccines (Pasteur 1173 P2 and Danish 1331) induce strong immunogenicity in animal models, whereas the strains Glaxo 1077 and Tokyo 172 stimulate weak responses. Side effects differ from strain to strain and are predominantly related to infection or to errors in achieving intradermal inoculation (e.g. local reactions, lymphadenitis, osteitis). Recently, the BCG strains have been carefully characterized and shown to lack expression of certain antigens present in *M. tuberculosis*, which might be important for protective immunity. Therefore,

BCG is the most controversial vaccine in current use. It is generally accepted that BCG protects children from meningeal and miliary tuberculosis. However, protection of adults from the most common form of the disease, pulmonary tuberculosis, varies in different studies from 80% in the United Kingdom to 0% in South India. This variable efficacy can be due to many reasons, such as vaccine strain, vaccine quality, host genetics, nutrition, exposure of the vaccine to UV, infection incidence, and clinical study protocols. Fundamental questions need to be addressed concerning the rational design of novel tuberculosis vaccines, preclinical and clinical tests, and future implementation in the field.

4.7 Typhoid Fever (*Salmonella enterica* serovar Typhi)

Typhoid fever affects about 17 million people worldwide each year, with approximately 600 000 deaths. The disease is caused by *Salmonella enterica* serovar Typhi, after ingestion of contaminated food or water. Typhoid fever and infections caused by other serovars in humans and animals are a serious medical and veterinary problem (see Sect. 5.5). Thus, the development of vaccines against this pathogen has constituted an important focus of research. Whole-cell inactivated or attenuated vaccines, as well as subunit vaccines, have been used with variable results. However, most licensed typhoid vaccines confer only about 70% protection, show poor efficacy in young children, and are not used for routine vaccination.

The first parenteral whole-cell typhoid fever vaccine was introduced in 1896 in England. Using heat, alcohol, acetone, or formalin, different whole-cell vaccines were constructed and tested in field trials. These vaccines conferred 51 to 88% protection to children and young adults, lasting for up to 12 years. However, inactivated vaccines often lead to local and systemic side effects, such as fever (6–30%), headache (10%), and severe local pain (up to 35%). This type of vaccine also requires regular boosts; thus, they are rarely included in public health programs. The first human trials using attenuated strains against salmonellosis were made in the early 1970s with a streptomycin-dependent mutant of serovar Typhi. However, no significant protection was conferred by this vaccine. Then, a live attenuated vaccine strain (Ty21a) was developed by chemical mutagenesis. This strain is administered by the oral route and carries, among others, the *galE* mutation, which renders bacteria extremely susceptible to lysis in the presence of galactose. Ty21a constitutes one of the few licensed attenuated bacterial vaccines for human use and the only live vaccine against serovar Typhi. Two different formulations of Ty21a live oral typhoid vaccine have been commercialized. The liquid formulation provides better protection than enteric-coated capsules. Three doses of Ty21a in liquid formulation stimulated 77% protection over three years and 78% over five years of follow-up. Ty21a caused minimal side effects, with nausea being the only adverse event appearing more frequently in vaccinees than in the control group. A postmarketing surveillance study showed only 743 spontaneous reports of adverse effects in more than 38 million vaccinees. On the other hand, a parenteral subunit vaccine based on purified capsular polysaccharide (Vi antigen) was shown to provide 64 to 72% protection over 21 months. A conjugate of Vi with nontoxic recombinant *Pseudomonas*

aeruginosa exotoxin A (rEPA) showed enhanced immunogenicity in adults and also in children 5 to 14 years of age and stimulated a booster response in 2- to 4-year-old children. Interestingly, an efficacy of over 90% in 2- to 5-year-old children was reported after intradermal immunization with the Vi-rEPA conjugate vaccine. The improved knowledge of bacterial physiology, pathogenesis mechanisms, and host immune response has allowed developing a new generation of live attenuated vaccine strains. They have been generated by introducing defined nonreverting mutations into genes coding for metabolic or virulence functions. These vaccines have given promising results in preclinical and clinical studies.

4.8 Tularemia (*Francisella tularensis*)

Tularemia, caused by *Francisella tularensis*, is a primary disease of a wide variety of wild mammals and birds, but it can also be transmitted to humans. Type A strains of *F. tularensis* found in North America are more virulent than type B strains found in Asia, Europe, and North America. Human infections can be caused by as few as 10 bacteria, which enter the host via the skin, mucous membranes, gastrointestinal tract, or the lungs. The high disease incidence in the United States and Russia during the first half of the 20th century led to intensive research aimed at the development of efficient vaccines, because it was observed that previously infected individuals were protected against reinfection. However, vaccine candidate strains inactivated by heat, acetone, or phenol treatment did not show satisfactory efficacy in animal studies or human volunteers. During the same period, the Soviet Union developed attenuated strains, which were used to immunize millions of individuals living in tularemia-endemic areas by intradermal route until 1960. Attenuation was obtained by repeated subculturing in media supplemented with antiserum. One of those attenuated strains, *F. tularensis* LSV, was brought to the United States in 1956, and was routinely used to protect technical personnel against laboratory-acquired infections. *F. tularensis* has been considered a potential biological weapon since 1932. Because of the current danger of bioterrorism, the LSV vaccine is currently under review by the US Food and Drug Administration to assess its future potential as a vaccine.

4.9 Whooping Cough (*Bordetella pertussis*)

Whooping cough is a highly contagious respiratory infection caused by the gram-negative bacillus *Bordetella pertussis*. This pathogen still remains an important killer of children, although there has been a global decline in incidence, consistent with the overall increase in immunization. In 2001, over 12 million cases of pertussis were estimated to occur worldwide, which led to more than 285 000 deaths. After the isolation and propagation of the causative agent in 1906, major efforts were focused in the development of a vaccine. In 1914, a whole-cell "whooping cough vaccine" was listed in the United States. Over the following decades, several whole-cell vaccines were generated, which consist of chemically or heat-killed *B. pertussis* cells. In 1942, the inactivated whole-cell *B. pertussis* vaccine was combined with diphtheria and tetanus toxoids, leading to a triple vaccine (DTP) that is still in use and confers more than 80% protection. The incidence of whooping cough has been

greatly reduced by mass immunization with DTP.

Whole-cell vaccines were frequently associated with local side effects (pain, redness, swelling). Some children also showed transient fever, protracted inconsolable crying, hypotonic–hyporesponsive episodes, and seizures. Additional studies showed that the vaccine does not cause sudden infant death syndrome, spasms, or epilepsy and that the risk of developing an acute encephalopathy was less than 1 : 100 000. Thus, the benefits associated with whole-cell vaccines clearly outweigh the risks of side effects, and vaccine-mediated neurological damage or death has never been conclusively proven. However, increasing concern about adverse effects has led to a reduction in public vaccine acceptance and a consequent increase in the incidence of the disease. Thus, efforts were made to develop acellular subunit vaccines against whooping cough based on key virulence factors (pertussis toxoid, pertactin, filamentous hemagglutinin), which are now in the market. These vaccines show much lower rates of adverse effects and have replaced whole-cell vaccines in developed countries. However, due to their lower costs, whole-cell vaccines are still used in developing countries. One of the major advantages of subunit vaccines is that they can be also used in older children and adults. Thus, they are particularly useful for boosting. This is particularly important, because both natural infection and vaccination with whole-cell or acellular vaccines induce protection against reinfection only for a limited time (about 5–15 years). Adolescents and adults become gradually susceptible, as demonstrated by the increased incidence of atypical whooping cough cases in these age groups.

5 Veterinary Bacterial Vaccines

Bacterial pathogens cause a wide range of diseases in animals and have a tremendous economic impact. In addition, individuals with occupational or chronic exposure can acquire zoonotic disease (e.g. veterinarians, farm workers, butchers, pet owners). Thus, animal diseases should be combated both to improve the quality of animal products and to prevent human disease. The use of antibiotics does not constitute a valid alternative, due to both actual regulations for human food and the increased risk of emergence of antibiotic resistance. Therefore, vaccination is the best alternative for tackling this problem. The use of vaccines allows reduction of (1) the clinical symptoms and the consequent economic impact of veterinary diseases, (2) the risk of carrier development, and (3) bacterial shedding and horizontal transfer to susceptible hosts.

Since the early 1900s, massive research efforts have been invested in preventing bacterial diseases in animals. Incomplete knowledge of the causative agents and limited information about clearance mechanisms rendered this task difficult. However, by the end of the twentieth century the use of bacterial vaccines has led to a significant reduction in the incidence of infectious diseases in the veterinary field. Thus, there are many examples of veterinary diseases for which vaccines were successfully implemented. Interestingly, the number of inactivated or attenuated classical bacterial vaccines is consistently larger in the veterinary than in the human field. This is due, at least in part, to broader margins in terms of acceptable side effects and to the critical need of keeping costs low. Furthermore, combined formulations containing different bacterial and/or viral

components are the rule rather than the exception. Although the presentation of a fully comprehensive report of all veterinary vaccines goes beyond the scope of this article, several examples are given below. The focus is on those diseases that have the biggest impact, with particular emphasis on those for which classical bacterial vaccines have been introduced into the market.

5.1 Infections Caused by *Bordetella* and *Pasteurella* species

Symptomatic and asymptomatic infections caused by *Bordetella bronchiseptica* are extremely widespread in the veterinary field, with infection rates that may reach 80 to 100% in intensively reared animals. This microorganism produces a variety of pathological syndromes and predisposes infected animals to viral or bacterial superinfections (e.g. *Pasteurella multocida* Type D and *Mycoplasma hyopneumoniae* in pigs). Pigs infected with *B. bronchiseptica* and toxigenic strains of *P. multocida* can develop atrophic rhinitis. The dermonecrotic toxin released by *P. multocida* affects osteocytes and osteoblasts, promoting bone resorption and leading to systemic manifestations, and is responsible for progressive atrophic rhinitis. The signs of the disease appear by 8 to 12 weeks of age and progress throughout the growing period. Infected animals show severe atrophy of the nasal turbinate bones, accompanied by lateral deviation or shortening of the nose. A British survey on abattoir pigs showed that 92% of herds were infected, with 36% of animals exhibiting atrophic changes. This may result in a 5 to 15% reduction on live weight gain (25–40 g/day), which in turn may lead to an annual loss of output or resource wastage of 1 to 4 million £. Thus, parenteral vaccines against atrophic rhinitis, which are based on inactivated *B. bronchiseptica* in combination with *P. multocida* Type D dermonecrotic toxoid alone or together with chemically inactivated *Pasteurella*, are widely used in pigs. Vaccinated animals and their progeny develop high titers of antibodies against *B. bronchiseptica* and *P. multocida* dermonecrotic toxoid. Both sows and gilts and their piglets are vaccinated, leading to reduced bacterial colonization, toxin neutralization, better growth, increased average daily weight gain (ADG), better snout scores, and reduced mortality. In some cases, these vaccines are combined with inactivated *Erysipelothrix rhusiopathiae* to prevent erysipelas in sow herds. In addition to inactivated *B. bronchiseptica* vaccines for pigs, field studies have been performed with live attenuated prototypes, which stimulate immune responses that can reduce morbidity rates relative to unvaccinated controls.

Vaccines against *Bordetella* spp. and *Pasteurella* spp. have also been developed for other animal species. Among them, attenuated *B. bronchiseptica* vaccines, for intranasal administration to dogs to prevent infectious tracheobronchitis (kennel cough), are generally combined with other antigens (e.g. attenuated canine parainfluenza virus). Live attenuated vaccines against the poultry pathogen *Bordetella avium* were introduced into the market to protect turkeys against rhinotracheitis.

Pasteurella pneumonia is estimated to be the top killer of cattle in the United States, and *Pasteurella haemolytica* is the main contributor to shipping fever, costing millions of dollars annually in deaths and treatment costs. Thus, attenuated *P. multocida* strains are combined with *P. haemolytica* in formulations for cattle. Viral components, such as rhinotracheitis

virus, bovine respiratory syncytial virus, parainfluenza, and bovine diarrhea virus are added to the basic formulations to cover a broad range of pathogens.

Infections of domestic and wild birds with *P. multocida* lead to fowl cholera, a contagious, widely distributed disease with high morbidity and mortality. The disease is typically characterized by sudden septicemia and fibrinous pneumonia; however, chronic and asymptomatic infections also occur. *P. multocida* subspecies *multocida* is the most common cause of fowl cholera, although subspecies *septica* and *gallicida* may also cause a fowl cholera-like disease. *P. multocida* is classified into 16 serotypes based on its lipopolysaccharide antigens, and the strains most frequently causing fowl cholera are A:1, A:3, and A:4. Live attenuated and inactivated (e.g. Cholervac-PM-1®) *P. multocida* vaccines have been introduced into the market. Whole-cell inactivated vaccines (bacterins) can provide protection, but only against the homologous serotype. On the other hand, live attenuated vaccines can protect under *in vivo* conditions against heterologous serotypes. However, attenuated vaccines retain a low level of virulence, and these strains can be implicated in infection outbreaks. For the prevention of fowl cholera, the most commonly used live vaccines are the *P. multocida* M-9 and PM-1 strains. Current experimental studies revolve around identification of the specific cross-protective antigens that might be incorporated into vaccine formulations, such as the outer membrane protein Oma87, a type-4 fimbrial subunit (PtfA), and a transferrin binding protein (Tbpl). Although conclusive evidence concerning a potential role of the capsule in virulence is still missing, the genes involved in capsule biosynthesis have been characterized as potential attenuation targets. More recently, attenuated auxotrophic *P. multocida aroA* derivatives have been generated by recombinant DNA technologies and are able to protect against homologous and heterologous challenges in chickens.

5.2 Brucellosis (*Brucella* spp.)

Brucellosis is a worldwide distributed zoonotic disease caused by *Brucella* spp. Infective cells can persist in the environment for weeks, and dried preparations can retain virulence for years. *B. melitensis* commonly infects goats and sheep and is particularly virulent in humans, whereas *B. abortus, B. ovis*, and *B. suis* preferentially infect cattle, sheep, and swine, respectively. The disease can be transmitted to humans by consumption of raw animal products or by direct contact with infected animals or their carcasses, through skin lesions or by inhalation.

Because of the serious economic and medical consequences of brucellosis, consistent efforts have been invested to develop vaccines. However, no human vaccine is currently available against brucellosis, and the disease is mainly controlled by animal vaccination. Antibodies and cell-mediated immunity can influence the course of the infection. However, cell-mediated responses (e.g. CTL) are required for clearance of intracellular bacteria. These responses are best stimulated by using live vaccines or by multiple injections of protective antigens with adjuvants. Only a few effective candidates have been identified so far that reduce abortion but not necessarily infection. Live vaccines retain some virulence and, depending on the administered dose, abortions may occur during pregnancy. The strain *B. abortus*

19 has been the most widely used to prevent brucellosis in bovines. However, the efficacy differs according to the age, dose, route of administration, and disease prevalence in vaccinated herds. The strain was originally isolated from cow milk in 1923, and, after being kept at room temperature, it was found to be attenuated. The live attenuated *B. melitensis* strain Rev.1 can stimulate protection against *B. melitensis* in sheep and goats. This strain also protects rams against infection with *B. ovis*. Strains 19 and Rev.1 can be horizontally transferred from vaccinated to unvaccinated animals and change to a rough form, since they persist in some animals. Another disadvantage is that the vaccine strains are fully virulent to humans, and many accidental infections have been reported. Other attenuated strains have not been widely adopted, such as a mucoid derivative of *B. suis* (M vaccine), the *B. abortus* 104-M vaccine, *B. suis* strain 2, and *B. melitensis* strain 5, which were used as oral vaccines in China.

The smooth vaccine strains 19 and Rev.1 stimulate anti-O-chain antibodies, thereby preventing differentiation between vaccinated and infected animals. Therefore, efforts were addressed toward the development of a marker vaccine that would not interfere with eradication programs. The smooth *B. abortus* strain 45/20 was isolated in 1922 and became rough after 20 passages in Guinea pigs. This strain was protective but tended to revert to the virulent form. Thus, it was used as an inactivated preparation with adjuvants based on water and oil emulsions. However, this vaccine was discontinued because its efficacy varied greatly from batch to batch and was associated with severe local reactions. *B. abortus* strain RB51, which lacks the O-chain, was very stable after multiple *in vitro* and *in vivo* passages. Its efficacy after subcutaneous vaccination was similar to that of strain 19, but so far it has not led to deleterious side effects. Thus, strain RB51, which also induces protection in swine against *B. suis* and in goats against *B. melitensis*, is replacing strain 19 in several countries. Oral vaccination with RB51 stimulates protective immunity in mice and cattle; hence, a practical approach to wildlife immunization seems possible.

Prevention of human brucellosis is best achieved by control or eradication of the disease in animals, combined with adequate heat treatment of potentially contaminated food products. However, in some countries this is not feasible. Therefore, attempts have been made to protect humans by vaccination. This has been reinforced by the renewed interest in *Brucella* resulting from its potential as a biological weapon; however, the results obtained have been generally disappointing. A derivative of strain 19 (19-BA), which was administered by scarification in the former USSR, stimulates short-time protection but leads to severe reactions. Strain RB51 shows minimal pathogenicity in humans, but is resistant to rifampicin, which makes it unsuitable for human use.

5.3 Porcine Pleuropneumonia (*Actinobacillus pleuropneumoniae*)

Actinobacillus pleuropneumoniae is the etiologic agent of a highly contagious porcine pleuropneumonia, characterized by fibrinous, hemorrhagic, and necrotic lung lesions. The disease shows a morbidity of 15 to 30% and is associated with high mortality rates in acute outbreaks (100% of affected pigs) and a decrease in the ADG of up to 35%. So far, 15 serotypes have been identified. Different vaccines have

been introduced into the market; some are based on inactivated bacteria (bacterins), and other formulations are based on purified and detoxified Apx toxins I, II, and III (subunit vaccines). Many of these vaccines are combined with other agents, such as chemically inactivated *E. rhusiopathiae* or *Haemophilus parasuis*. Natural or experimental infection with live bacterial cells seems to offer the greatest level of protection against reinfection, suggesting that a live strain would be the ideal vaccine candidate. Several attenuated *A. pleuropneumoniae* strains have been constructed and conferred various degrees of protection against disease; these included temperature-sensitive, metabolic, nonencapsulated, and nontoxigenic mutants. More recently, an attenuated mutant deficient in the biosynthesis of aromatic compounds (HS25 *aroQ*) has been safely delivered to pigs.

5.4 Diseases Caused by *Mycoplasma* spp.

Mycoplasma spp. cause upper and lower respiratory tract infections in various animal species. These microorganisms contain the smallest bacterial genomes, lack cell walls, and are strictly dependent on the host for survival. *Mycoplasma hyopneumoniae* is a main player in the porcine respiratory disease complex, leading to significant economic losses. The prevalence of enzootic pneumonia ranges from 30 to 80% in different countries. *M. hyopneumoniae* also creates a port of entry for other pathogens. This phenomenon, known as *Mycoplasma-induced respiratory disease* (MIRD), aggravates the seriousness of the clinical signs and increases economic losses. MIRD can considerably affect the feed efficiency index (up to 14%) and the ADG (up to 17.4%). This situation results in an increase in the fattening period, which in turn leads to decreased rotation rates, increased labor and feeding costs, and reduced annual production. Infections spread from the infected sow to its litter and from older to younger susceptible pigs. Commercial vaccines have been available since 1991. They are based on inactivated and adjuvanted bacteria and are administered to baby or feeder pigs, as well as to breeding stocks. Some of the vaccine preparations have been combined with antigens from other pathogens, such as inactivated *E. rhusiopathiae* or *H. parasuis*, the etiologic agent of Glasser disease (i.e. polyserositis).

Mycoplasma gallisepticum is also an important pathogen that causes chronic respiratory disease in chicken. Vaccination with killed or living vaccines is an option for controlling infections in poultry flocks when hygienic measures fail. Live attenuated *M. gallisepticum* vaccines include the strains F, ts-11, and 6/85. The F-strain (administered in drinking water or by aerosol) reduces the decline in egg production but has residual virulence. On the other hand, strain ts-11 is less virulent and immunogenic, but still provides effective long-term protection. Temperature-sensitive mutants of *M. gallisepticum* (TS100) were used to immunize newly hatched chickens and protected them from developing air-sac lesions.

Mycoplasma agalactiae is the causal agent of contagious agalactia in small ruminants. The disease, which is more frequent in the Mediterranean region, North Africa, and the Middle East, is characterized by mastitis, arthritis, and pneumonia. Killed organisms are usually administered in combination with adjuvants. Inactivation with phenol seems to confer better protection than formalin- or heat-inactivation.

The lack of protection observed in the field may be due to infections with other *Mycoplasma* spp. Improved protection in comparison to killed vaccines was observed with live attenuated strains of *M. agalactiae*.

5.5 Salmonellosis in Animals

Salmonella causes diseases characterized by acute or chronic enteritis and septicemia in many animal species, which are the principal reservoirs for human infections. Transmission to humans occurs by contaminated drinking water or food products (e.g. milk, meat, eggs), and their incidence has increased worldwide due to the intensification of livestock production. *S. enterica* serovar Typhimurium became the second most important cause of salmonellosis in humans. Since vaccination is compulsory for poultry meat producers, and EU regulations also require monitoring of broiler-breeder flocks and slaughter of infected flocks, efforts have been made to develop efficacious multivalent vaccines for poultry. One of these vaccines consists of iron-restricted inactivated *S. typhimurium* and *S. enteritidis*. After intramuscular vaccination, chickens are protected against an oral challenge, but colonization still occurs. Other mutants, for example, *dam*, show better results with respect to protection and colonization. The first live oral vaccine against *S. typhimurium* and *S. enteritidis* was launched by Lohmann Animal Health in 2003 (TAD Salmonella). Three doses administered into the birds' drinking water give protection throughout the laying period.

The efficacy of live vaccine candidates against *Salmonella choleraesuis* (e.g. Argus SC), which causes infections in pigs, has also been evaluated. Vaccinated pigs were able to maintain normal weight after challenge, and field studies demonstrated that this type of vaccine can lower the prevalence of *Salmonella* in swine herds.

Salmonella gallinarum and *Salmonella pullorum* cause fowl typhoid and pullorum disease in birds. The clinical manifestations include anorexia, diarrhea, hepatitis, splenitis, myocarditis, pneumonia, ophthalmitis, and high mortality due to septicemia. This leads to decreased egg production and fertility. Thus, inactivated and live attenuated vaccine strains have been developed for poultry. Since 1950, attenuated derivatives of *S. gallinarum* strain 9 have been extensively assessed. One of them was attenuated by passage in a medium of low nutritional quality, leading to the so-called *9R fowl typhoid strain*, which confers strong protection to adult chickens after intramuscular injection. However, this strain retains some virulence, may persist for many months, and can be transmitted through the egg. To obtain a more avirulent strain, an *aro A* mutant was developed, but it was less effective. A single oral immunization with an *S. gallinarum nuoG* derivative reduced the mortality in 2-week-old chickens following challenge with virulent *S. gallinarum* from 75% to less than 8%. Another approach was the intramuscular application of attenuated "phage type" PT4 *S. enteritidis*.

5.6 Leptospirosis (*Leptospira* spp.)

Leptospirosis is one of the most widespread diseases in livestock and is caused by over 250 immunologically distinct serovars. Bacteria are spread by contact with skin or mucous membranes with contaminated products. Infected animals contribute to further

spreading of the disease by actively shedding bacteria. Infections may be asymptomatic or symptomatic (e.g. fever, icterus, hemoglobinuria, renal failure, infertility, abortion, and death). Of the distinct pathogenic serovars, only a small number have been isolated from domestic animals. It is remarkable that only moderate cross-immunity between serovars is observed. Cattle, swine, and horses are the most economically relevant affected species. Because leptospirosis in domestic animals is leading to significant economic losses and infected animals are a potential source of human infection, vaccination with polyvalent vaccines have been widely implemented. Certain occupational groups, such as farmers, sewer workers, and meat workers, have special risks of acquiring the disease. Outdoor leisure activities can also lead to exposure to infection. Inactivated polyvalent vaccines containing different Leptospiras (e.g. *grippotyphosa, hardjo, icterohaemorrhagiae, canicola, pomona*) are commercially available for preventing abortions and stillborns in livestock and pets. Most of them have been combined with components against other pathogens, such as *E. rhusiopathiae*, parvovirus, *Campylobacter*, bovine rhinotracheitis virus, bovine diarrhea virus, and bovine respiratory syncytial virus.

5.7 Other Commercially Relevant Animal Diseases

Vaccines that are able to prevent other diseases in livestock and pets have also been developed and implemented in the field. Among them are inactivated oil-adsorbed vaccines against *Haemophilus paragallinarum*, which is responsible for infectious coryza in chickens. This disease is characterized by swelling of the infraorbital sinuses, facial edema, oculonasal discharge, and sneezing and may cause significant losses in the poultry industry. Inactivated adjuvanted vaccines based on various strains of *Fusobacterium necrophorum* and/or *Bacteroides nodosus* are also available to prevent and treat chronic foot rot in sheep and acute foot rot in cattle. It is also possible to immunize against equine monocytic ehrlichiosis by using a whole-cell inactivated *Ehrlichia risticii* preparation. For pigs, killed vaccines against *Streptococcus suis* have also been developed. In addition, inactivated vaccines that are based on inactivated enterotoxigenic and enteropathogenic *E. coli* combined with toxoids are available for preventing post-weaning diarrhea. *Chlamydia psittaci* is responsible for respiratory infections in cats, leading to serous conjunctivitis, sneezing, and nasal discharge. Vaccinated cats have mild clinical disease and are protected from severe clinical disease after challenge exposure. Owing to the frequent emergence of adverse effects (e.g. lethargy, anorexia, lameness, and fever), the use of vaccines is frequently restricted to cats at risk of exposure. Live attenuated vaccines may cause atypical reactions in about 3% of vaccinated cats. Finally, vaccines have been also developed to prevent bacterial infections in fish. A live attenuated vaccine is used to prevent the enteric septicemia caused by *Edwardsiella ictaluri*, which costs millions of dollars to the catfish industry.

6 Conclusions

During the past 200 years, various vaccines have been developed and exploited to control major diseases. In many instances, however, the exact mechanisms of action

of successful vaccines were not fully understood. A new generation has emerged, of well-defined recombinant and/or subunit vaccines that exhibit a considerably improved safety profile. However, classical bacterial vaccines have not only played a critical role in the efficient control of infectious diseases in the past, they are still instrumental in the prevention of infections both in humans and, particularly, in animal populations.

Acknowledgments

We thank Joan Plana Duran (Fort Dodge Veterinaria S.A.), Guido Dietrich (Berna Biotech Ltd.), and Kai Schulze (GBF) for helpful discussions and critical reading of the manuscript. This work was supported in part by grants from DFG (GU 482/2–1 and GU482/2–2), EU (QLK2-CT-1999–00310), and BMBF (PathoGenoMik 601III4–1/1VIIZV-21) to CAG.

See also Antigen Presenting Cells (APCs); Dendritic Cells; Immune Defence, Cell Mediated; Immunologic Memory; Immunology.

Bibliography

Books and Reviews

Dorner, F., Barrett, P. (1999) Vaccine technology: looking to the future, *Ann. Med.* **31**, 51–60.

Friedlander, A., Welkos, S., Ivins, B. (2002) Anthrax vaccines, *Curr. Top. Microbiol. Immunol.* **271**, 33–60.

Hunter, R.L. (2002) Overview of vaccine adjuvants: present and future, *Vaccine* **20**, S7–S12.

Makela, P.H. (2000) Vaccines, coming of age after 200 years, *FEMS Microbiol. Rev.* **24**, 9–20.

Mollenkopf, H., Dietrich, G., Kaufmann, S.H. (2001) Intracellular bacteria as targets and carriers for vaccination, *Biol. Chem.* **382**, 521–532.

Paterson, Y. (1999) *Intracellular Bacterial Vaccine Vectors, Immunology, Cell Biology, and Genetics*, John Wiley & Sons, Philadelphia, USA.

Stanley, A., Plotkin, M.D., Walter, A., Orenstein, M.D. (1999) *Vaccines*, 3rd edition, W.B. Saunders Company, A Division of Harcourt Brace & Company, Philadelphia, USA.

Primary Literature

Adler, B., Bulach, D., Chung, J., Doughty, S., Hunt, M., Rajakumar, K., Serrano, M., van Zanden, A., Zhang, Y., Ruffolo, C. (1999) Candidate vaccine antigens and genes in Pasteurella multocida, *J. Biotechnol.* **73**, 83–90.

Ahren, C., Jertborn, M., Svennerholm, A-M. (1998) Intestinal immune responses to an inactivated oral enterotoxigenic Escherichia coli vaccine and associated immunoglobulin A responses in blood, *Infect. Immun.* **66**, 3311–3316.

Baraff, L.J., Shields, W.D., Beckwith, L., Strome, G., Marcy, S.M., Cherry, J.D., Manclark, C.R. (1988) Infants and children with convulsions and hypotonic-hyporesponsive episodes following diphtheria-tetanus-pertussis immunization: follow-up evaluation, *Pediatrics* **81**, 789–794.

Baron, S. (1996) *Medical Microbiology*, 4th edition, The University of Texas Medical Branch at Galveston, Galveston, USA.

Behr, M.A., Wilson, M.A., Gill, W.P., Salamon, H., Schoolnik, G.K., Rane, S., Small, P.M. (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray, *Science* **284**, 1520–1523.

Brossier, F., Weber-Levy, M., Mock, M., Sirard, J.C. (2000) Role of toxin functional domains in anthrax pathogenesis, *Infect. Immun.* **68**, 1781–1786.

Burke, D.S. (1977) Immunization against tularemia: analysis of the effectiveness of live Francisella tularensis vaccine in prevention of laboratory-acquired tularemia, *J. Infect. Dis.* **135**, 55–60.

Byrd, W., Hooke, A. (1997) Immunization with temperature-sensitive mutants of Actinobacillus pleuropneumoniae induces

protective hemolysin-neutralizing antibodies in mice, *Curr. Microbiol.* **34**, 149–154.

Campins-Marti, M., Cheng, H.K., Forsyth, K., Guiso, N., Halperin, S., Huang, L.M., Mertsola, J., Oselka, G., Ward, J., Wirsing von Konig, C.H., Zepp, F. (2001) Recommendations are needed for adolescent and adult pertussis immunisation: rationale and strategies for consideration, *Vaccine* **20**, 641–646.

Clemens, J.D., Sack, D.A., Harris, J.R., Van Loon, F., Chakraborty, J., Ahmed, F., Rao, M.R., Khan, M.R., Yunus, M., Huda, N. et al. (1990) Field trial of oral cholera vaccines in Bangladesh: results from three-year follow-up, *Lancet* **335**, 270–273.

Clemens, J.D., van Loon, F., Sack, D.A., Rao, M.R., Ahmed, F., Chakrabort, Y.J., Kay, B.A., Khan, M.R., Yunus, M.D., Harris, J.R. et al. (1991) Biotype as determinant of natural immunising effect of cholera, *Lancet* **337**, 883–884.

Clifton-Hadley, F.A., Breslin, M., Venables, L.M., Sprigings, K.A., Cooles, S.W., Houghton, S., Woodward, M.J. (2002) A laboratory study of an inactivated bivalent iron-restricted Salmonella enterica serovars Enteritidis and Typhimurium dual vaccine against Typhimurium challenge in chickens, *Vet. Microbiol.* **89**, 167–179.

Cohen, D., Ashkenazi, S., Green, M.S., Gdalevich, M., Robin, G., Slepon, R., Yavzori, M., Orr, N., Block, C., Ashkenazi, I. et al. (1997) Double-blind vaccine-controlled randomised efficacy trial of an investigational Shigella sonnei conjugate vaccine in young adults, *Lancet* **349**, 155–159.

Colditz, Gwarning.A., Brewer, T.F., Berkey, C.S., Wilson, M.E., Burdick, E., Fineberg, H.V., Mosteller, F. (1994) Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature, *JAMA* **271**, 698–702.

Collins, H.L., Kaufmann, S.H. (2001) Prospects for better tuberculosis vaccines, *Lancet Infect. Dis.* **1**, 21–28.

Confer, A. (1993) Immunogens of Pasteurella, *Vet. Microbiol.* **37**, 353–368.

Coriell, L., King, E., Smith, M. (1948) Studies on tularaemia IV. Observations on tularaemia in control and vaccinated monkeys, *J. Immunol.* **58**, 183–202.

Dennis, D.T., Inglesby, T.V., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Fine, A.D., Friedlander, A.M., Hauer, J., Layton, M. et al. (2001) Tularemia as a biological weapon: medical and public health management, *JAMA* **285**, 2763–2773.

Dietrich, G., Spreng, S., Favre, D., Viret, J.F., Guzman, C.A. (2003) Live attenuated bacteria as vectors to deliver plasmid DNA vaccines, *Curr. Opin. Mol. Ther.* **5**, 10–19.

Dueger, E.L., House, J.K., Heithoff, D.M., Mahan, M.J. (2001) Salmonella DNA adenine methylase mutants elicit protective immune responses to homologous and heterologous serovars in chickens, *Infect. Immun.* **69**, 7950–7954.

Dueger, E.L., House, J.K., Heithoff, D.M., Mahan, M.J. (2003) Salmonella DNA adenine methylase mutants prevent colonization of newly hatched chickens by homologous and heterologous serovars, *Int. J. Food Microbiol.* **80**, 153–159.

Dye, C., Scheele, S., Dolin, P., Pathania, V., Raviglione, M.C. (1999) Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project, *JAMA* **282**, 677–686.

Ebensen, T., Link, C., Guzman, C.A. (2004) Classical and Novel vaccination strategies: A comparison of Classical Bacterial Vaccines, in: Kaufmann, S.H.E. (Ed.) *Novel Vaccination strategies*, Wiley-VCH, Weinheim, Germany, 221–240.

Ebensen, T., Paukner, S., Link, C., Kudela, P., de Domenico, C., Lubitz, W., Guzman, C.A. (2004) Bacterial ghosts are an efficient delivery system for DNA vaccines, *J. Immunol.* **172**, 6858–6865.

Ehser, U., Mehlhorn, G., Mietke, H. (1993) The effect of a Bordetella live vaccine on the occurrence and manifestation of atrophic rhinitis suum and the aerogenous infection burden in field strains of the agent, *Dtsch. Tierärztl. Wochenschr.* **100**, 355–359.

Elias, B., Albert, M., Tuboly, S., Rafai, P. (1992) Interaction between Bordetella bronchiseptica and toxigenic Pasteurella multocida on the nasal mucosa of SPF piglets, *J. Vet. Med. Sci.* **54**, 1105–1110.

Ellis, R.W. (1999) New technologies for making vaccines, *Vaccine* **17**, 1596–1604.

Ellis, R.W. (2003) *New Bacterial Vaccines*, Landes Bioscience, Georgetown, USA.

Elzer, P., Enright, F., Colby, L., Hagius, S., Walker, J., Fatemi, M., Kopec, J., Beal, V., Schurig, G. (1998) Protection against infection

and abortion induced by virulent challenge exposure after oral vaccination of cattle with Brucella abortus strain RB51, *Am. J. Vet. Res.* **59**, 1575–1578.

Engels, E.A., Falagas, M.E., Lau, J., Bennish, M.L. (1998) Typhoid fever vaccines: a meta-analysis of studies on efficacy and toxicity, *Br. Med. J.* **316**, 110–116.

Evans, D., Evans, D., Opekun, A., Graham, D. (1988a) Immunoprotective oral whole cell vaccine for enterotoxigenic Escherichia coli diarrhea prepared by in situ destruction of chromosomal and plasmid DNA with colicin E2, *FEMS Microbiol. Immunol.* **1**, 9–18.

Evans, D., Evans, D., Opekun, A., Graham, D. (1988b) Non-replicating oral whole cell vaccine protective against enterotoxigenic Escherichia coli (ETEC) diarrhea: stimulation of anti-CFA (CFA/I) and anti-enterotoxin (anti-LT) intestinal IgA and protection against challenge with ETEC belonging to heterologous serotypes, *FEMS Microbiol. Immunol.* **1**, 117–125.

Foggie, A., Etheridge, J.R., Erdag, O., Arisoy, F. (1970) Contagious agalactia of sheep and goats. The serial passage in goats of an attenuated strain of Mycoplasma agalactiae AIK40, *Res. Vet. Sci.* **11**, 477–479.

Fries, L.F., Montemarano, A.D., Mallett, C.P., Taylor, D.N., Hale, T.L., Lowell, G.H. (2001) Safety and immunogenicity of a proteosome-Shigella flexneri 2a lipopolysaccharide vaccine administered intranasally to healthy adults, *Infect. Immun.* **69**, 4545–4553.

Fuller, T., Thacker, B., Mulks, M. (1996) A riboflavin auxotroph of Actinobacillus pleuropneumoniae is attenuated in swine, *Infect. Immun.* **64**, 4659–4664.

Garmory, H.S., Brown, K.A., Titball, R.W. (2002) Salmonella vaccines for use in humans: present and future perspectives, *FEMS Microbiol. Rev.* **26**, 339–353.

Gordon, J., Hood, H. (1951) Whooping cough and its epidemiological anomalies, *Am. J. Med. Sci.* **222**, 333–361.

Griffin, H.G., Barrow, P.A. (1993) Construction of an aroA mutant of Salmonella serotype Gallinarum: its effectiveness in immunization against experimental fowl typhoid, *Vaccine* **11**, 457–462.

Gurunathan, S., Klinman, D.M., Seder, R.A. (2000) DNA vaccines: immunology, application, and optimization, *Annu. Rev. Immunol.* **18**, 927–974.

Hoffman, H.J., Hunter, J.C., Damus, K., Pakter, J., Peterson, D.R., van Belle, G., Hasselmeyer, E.G. (1987) Diphtheria-tetanus-pertussis immunization and sudden infant death: results of the National Institute of Child Health and Human Development Cooperative Epidemiological Study of Sudden Infant Death Syndrome risk factors, *Pediatrics* **79**, 598–611.

Ingham, A., Zhang, Y., Prideaux, C. (2002) Attenuation of Actinobacillus pleuropneumoniae by inactivation of aroQ, *Vet. Microbiol.* **84**, 263–273.

Inzana, T., Todd, J., Veit, H. (1993) Safety, stability, and efficacy of noncapsulated mutants of Actinobacillus pleuropneumoniae for use in live vaccines, *Infect. Immun.* **61**, 1682–1686.

Janeway, C.A., Travers, J.P., Walport, M., Shlomchik, M.J. (2001) *Immunobiology: the Immune System in Health and Disease*, 5th edition, Garland Publishing, New York, USA.

Kadull, P., Reames, H., Coriell, L., Foshay, L. (1950) Studies on tularaemia V. Immunization of man, *J. Immunol.* **65**, 425–435.

Kenner, J.R., Coster, T.S., Taylor, D.N., Trofa, A.F., Barrera-Oro, M., Hyman, T., Adams, J.M., Beattie, D.T., Killeen, K.P., Spriggs, D.R. et al. (1995) Peru-15, an improved live attenuated oral vaccine candidate for Vibrio cholerae O1, *J. Infect. Dis.* **172**, 1126–1129.

Klugman, K.P., Gilbertson, I.T., Koornhof, H.J., Robbins, J.B., Schneerson, R., Schulz, D., Cadoz, M., Armand, J. (1987) Protective activity of Vi capsular polysaccharide vaccine against typhoid fever, *Lancet* **2**, 1165–1169.

Kolar, J. (1989) Brucellosis in Eastern European Countries, in: Young, E., Corbel, M. (Eds.) *Brucellosis: Clinical and Laboratory Aspects*, Chap. 12, CRC Press, Boca Raton, FL, pp. 163–172.

Kotloff, K., Winickoff, J., Ivanoff, B., Clemens, J., Swerdlow, D., Sansonetti, P., Adak, G., Levine, M. (1999) Global burden of Shigella infections: implications for vaccine development and implementation of control strategies, *Bull. World Health Organ.* **77**, 651–666.

Kramer, T.T., Roof, M.B., Matheson, R.R. (1992) Safety and efficacy of an attenuated strain of Salmonella choleraesuis for vaccination of swine, *Am. J. Vet. Res.* **53**(4), 444–448.

Lam, K., Lin, W., Yamamoto, R., Farver, T. (1983) Immunization of chickens with

temperature-sensitive mutants of Mycoplasma gallisepticum, *Avian Dis.* **27**, 803–812.

Levenson, V.J., Mallett, C.P., Hale, T.L. (1995) Protection against local Shigella sonnei infection in mice by parenteral immunization with a nucleoprotein subcellular vaccine, *Infect. Immun.* **63**, 2762–2765.

Levine, M.M., Kaper, J.B. (1993) Live oral vaccines against cholera: an update, *Vaccine* **11**, 207–212.

Levine, M.M., Kaper, J.B., Black, R.E., Clements, M.L. (1983) New knowledge on pathogenesis of bacterial enteric infections as applied to vaccine development, *Microbiol. Rev.* **47**, 510–550.

Levine, M.M., Ferreccio, C., Abrego, P., Martin, O.S., Ortiz, E., Cryz, S. (1999) Duration of efficacy of Ty21a, attenuated Salmonella typhi live oral vaccine, *Vaccine* **17**, S22–S27.

Levine, M.M., DuPont, H.L., Hornick, R.B., Snyder, M.J., Woodward, W., Gilman, R.H., Libonati, J.P. (1976) Attenuated, streptomycin-dependent Salmonella typhi oral vaccine: potential deleterious effects of lyophilization, *J. Infect. Dis.* **133**, 424–429.

Lin, F.Y.C., Ho, V.A., Khiem, H.B., Trach, D.D., Bay, P.V., Thanh, T.C., Kossaczka, Z., Bryla, D.A., Shiloach, J., Robbins, J.B. et al. (2001) The efficacy of a Salmonella typhi Vi conjugate vaccine in two-to-five-year-old children, *N. Engl. J. Med.* **344**, 1263–1269.

Lord, V., Schurig, G., Cherwonogrodzky, J., Marcano, M., Melendez, G. (1998) Field study of vaccination of cattle with Brucella abortus strains RB51 and 19 under high and low disease prevalence, *Am. J. Vet. Res.* **59**, 1016–1020.

Lubitz, W. (2001) Bacterial ghosts as carrier and targeting systems, *Expert Opin. Biol. Ther.* **1**, 765–771.

McCluskie, M., Weeratna, R. (2001) Novel adjuvant systems, *Curr Drug Targets Infect. Disord.* 2001, **1**(3), 263–271.

Meitert, T., Pencu, E., Ciudin, L., Tonciu, M. (1984) Vaccine strain Sh. flexneri T32-ISTRATI. Studies in animals and volunteers. Antidysentery immunoprophylaxis and immunotherapy by live vaccine, *Arch. Roum. Pathol. Exp. Microbiol.* **43**, 251–278.

Mel, D.M., Papo, R.G., Terzin, A.L., Vuksic, L. (1965) Studies on vaccination against bacillary dysentery. 2. Safety tests and reactogenicity studies on a live dysentery vaccine intended for use in field trials, *Bull. World Health Organ.* **32**, 637–645.

Melchior, J.C. (1977) Infantile spasms and early immunization against whooping cough. Danish survey from 1970 to 1975, *Arch. Dis. Child.* **52**(2), 134–137.

Mielcarek, N., Alonso, S., Locht, C. (2001) Nasal vaccination using live bacterial vectors, *Adv. Drug Deliv. Rev.* **51**, 55–69.

Miller, D., Wadsworth, J., Diamond, J., Ross, E. (1985) Pertussis vaccine and whooping cough as risk factors in acute neurological illness and death in young children, *Dev. Biol. Stand.* **61**, 389–394.

Mock, M., Fouet, A. (2001) Anthrax, *Annu. Rev. Microbiol.* **55**, 647–671.

Moingeon, P., de Taisne, C., Almond, J. (2002) Delivery technologies for human vaccines, *Br. Med. Bull.* **62**, 29–44.

Murphy, E., Sathiyaseelan, J., Parent, M., Zou, B., Baldwin, C. (2001) Interferon-gamma is crucial for surviving a Brucella abortus infection in both resistant C57BL/6 and susceptible BALB/c mice, *Immunology* **103**, 511–518.

Orme, I.M., McMurray, D.N., Belisle, J.T. (2001) Tuberculosis vaccine development: recent progress, *Trends Microbiol.* **9**, 115–118.

Palmer, M., Olsen, S., Cheville, N. (1997) Safety and immunogenicity of Brucella abortus strain RB51 vaccine in pregnant cattle, *Am. J. Vet. Res.* **58**, 472–477.

Perry, R., Fetherston, J. (1997) Yersinia pestis–etiologic agent of plague, *Clin. Microbiol. Rev.* **10**, 35–66.

Potter, A., Babiuk, L. (2001) New approaches for antigen discovery, production and delivery: vaccines for veterinary and human use, *Curr. Drug Targets Infect. Disord.* **1**(3), 249–262.

Prideaux, C., Pierce, L., Krywult, J., Hodgson, A. (1998) Protection of mice against challenge with homologous and heterologous serovars of Actinobacillus pleuropneumoniae after live vaccination, *Curr. Microbiol.* **37**, 324–332.

Riising, H., van Empel, P., Witvliet, M. (2002) Protection of piglets against atrophic rhinitis by vaccinating the sow with a vaccine against Pasteurella multocida and Bordetella bronchiseptica, *Vet. Rec.* **150**, 569–571.

Rosenthal, S., Clifford, J. (2002) Development of vaccines for bio-warfare agents, *Dev. Biol. (Basel)* 2002, **110**, 99–105.

Ryan, E.T., Calderwood, S.B. (2000) Cholera vaccines, *Clin. Infect. Dis.* **31**, 561–565.

Sack, D., Cadoz, M. (1999) Cholera Vaccines, in: Plotkin, S., Orenstein, W.A. (Eds.) *Vaccines*, W.B. Saunders Company, Philadelphia, USA, 639–649.

Savarino, S., Hall, E., Bassily, S., Brown, F., Youssef, F., Wierzba, T., Peruski, L., El-Masry, N., Safwat, M., Rao, M., et al., PRIDE Study Group (1999) Oral, inactivated, whole cell enterotoxigenic Escherichia coli plus cholera toxin B subunit vaccine: results of the initial evaluation in children, *J. Infect. Dis.* **179**, 107–114.

Schurig, G.G., Roop, R., Bagchi, T., Boyle, S., Buhrman, D., Sriranganathan, N. (1991) Biological properties of RB51; a stable rough strain of Brucella abortus, *Vet. Microbiol.* **28**, 171–188.

Schurig, G.G., Sriranganathan, N., Corbel, M.J. (2002) Brucellosis vaccines: past, present and future, *Vet. Microbiol.* **90**, 479–496.

Scott, P., Markham, J., Whithear, K. (1999) Safety and efficacy of two live Pasteurella multocida aro-A mutant vaccines in chickens, *Avian Dis.* **43**, 83–88.

Sebunya, T., Saunders, J. (1983) Haemophilus pleuropneumoniae infection in swine: a review, *J. Am. Vet. Med. Assoc.* **182**, 1331–1337.

Shata, M.T., Stevceva, L., Agwale, S., Lewis, G.K., Hone, D.M. (2000) Recent advances with recombinant bacterial vaccine vectors, *Mol. Med. Today* **6**, 66–71.

Stevens, M., Olsen, S., Palmer, M., Pugh, G. Jr. (1996) Immune responses and resistance to brucellosis in mice vaccinated orally with Brucella abortus RB51, *Infect. Immun.* **64**, 4534–4541.

Stewart, G.T. (1977) Vaccination against whooping-cough. Efficacy versus risks, *Lancet* **1**, 234–237.

Suazo, F.M., Escalera, A.M., Torres, R.M. (2003) A review of M. bovis BCG protection against TB in cattle and other animals species, *Prev. Vet. Med.* **58**, 1–13.

Tacket, C.O., Cohen, M.B., Wasserman, S.S., Losonsky, G., Livio, S., Kotloff, K., Edelman, R., Kaper, J.B., Cryz, S.J., Giannella, R.A. et al. (1999) Randomized, double-blind, placebo-controlled, multicentered trial of the efficacy of a single dose of live oral cholera vaccine CVD 103-HgR in preventing cholera following challenge with vibrio cholerae O1 El Tor Inaba three months after vaccination. *Infect. Immun.* **67**, 6341–6345.

Titball, R.W., Eley, S., Williamson, E., Dennis, D. (1999) Plague, in: Plotkin, S., Orenstein, W.A. (Eds.) *Vaccines*, W. B. Saunders, Philadelphia, PA, 734–742.

Titball, R.W., Williamson, E.D. (2001) Vaccination against bubonic and pneumonic plague, *Vaccine* **19**, 4175–4184.

Tola, S., Manunta, D., Rocca, S., Rocchigiani, A.M., Idini, G., Angioi, P.P., Leori, G. (1999) Experimental vaccination against Mycoplasma agalactiae using different inactivated vaccines, *Vaccine* **17**, 2764–2768.

Turnbull, P. (1991) Anthrax vaccines: past, present and future, *Vaccine* **9**, 533–539.

Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., Marth, J. (1999) *The Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Weill-Halle, B. (1957) Oral Vaccination, in: Rosenthal, S.R. (Ed.) *BCG Vaccination against Tuberculosis*, Little, Brown, Boston, MA, 175–182.

Welkos, S.L. (1991) Plasmid-associated virulence factors of non-toxigenic (pX01-) Bacillus anthracis, *Microb. Pathog.* **10**, 183–198.

Whithear, K. (1996) Control of avian mycoplasmoses by vaccination, *Rev. Sci. Tech.* **15**, 1527–1553.

WHO (1997) *Typhoid Fever*, Report 149.

WHO (2002) *World Health Report*.

Wilson, C.B., Marcuse, E.K. (2001) Vaccine safety–vaccine benefits: science and the public's perception, *Nat. Rev. Immunol.* **1**, 160–165.

Xin, X. (1986) Orally administrable brucellosis vaccine: Brucella suis strain 2 vaccine, *Vaccine* **4**, 212–216.

Zhan, Y., Yang, J., Cheers, C. (1993) Cytokine response of T-cell subsets from Brucella abortus-infected mice to soluble Brucella proteins, *Infect. Immun.* **61**, 2841–2847.

Zhang-Barber, L., Turner, A., Dougan, G., Barrow, P. (1998) Protection of chickens against experimental fowl typhoid using a nuoG mutant of Salmonella serotype Gallinarum, *Vaccine* **16**, 899–903.

Zhang-Barber, L., Turner, A.K., Barrow, P.A. (1999) Vaccination for control of Salmonella in poultry, *Vaccine* **17**, 2538–2545.

Human Genetic Variation and Disease

Lynn B. Jorde
University of Utah Health Sciences Center, Salt Lake City, Utah, USA

1 Introduction 325

2 Mutation 326
2.1 Gene Size and Duchenne Muscular Dystrophy 327
2.2 Mutation Hot Spots 327

3 Natural Selection 328
3.1 Heterozygote Advantage and Sickle Cell Disease 328
3.2 Natural Selection and HIV 329

4 Genetic Drift and Founder Effect 329
4.1 Haplotype Analysis 330
4.2 Mutation Analysis 330

5 Drift or Selection? 331
5.1 Hemochromatosis in Europe 331
5.2 G6PD Deficiency in Africa and the Mediterranean 332

6 Gene Flow 332
6.1 Cystic Fibrosis in African-Americans 332
6.2 Tay–Sachs Disease in French Canadians 332

7 Genetic Variation and Common Diseases 333

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

8 **Perspectives** 333

Bibliography 334
Books and Reviews 334
Primary Literature 334

Keywords

Fitness
An individual's relative success in transmitting his or her genes to the next generation.

Genetic Drift
Random intergenerational change in gene frequencies due to finite population size.

Haplotype
The allelic constitution of several loci on a single chromosome ("haploid genotype").

Heterozygote Advantage
A fitness differential in which the heterozygote has higher fitness than any of the homozygotes.

Natural Selection
An evolutionary process in which individuals with a specific genetic variant have higher fitness than those lacking the variant.

Polymorphism
The occurrence of two or more alleles at a locus in a population, where at least two alleles occur with frequencies greater than 1%.

Variation in the prevalence of genetic diseases results from the action of the evolutionary processes of mutation, natural selection, genetic drift, and gene flow. Genetic markers can be studied in populations to help determine the ways in which these processes interact to influence disease prevalence. This enhances our understanding of the origins and evolution of genetic diseases. This article will review the ways in which genetic markers are used to study disease variation in populations, using specific genetic diseases as examples.

1 Introduction

The prevalence of many genetic diseases varies widely among human populations (see Fig. 1 for examples). Cystic fibrosis (CF) is the most common lethal single-gene disorder among Caucasians, affecting approximately 1 in 2500 individuals. Yet, it is quite uncommon among Asians, with a prevalence of only 1 in 90 000 births. Sickle cell disease affects 1 in 400 to 600 African Americans, but it is rare among individuals of northern European descent. Tay–Sachs disease, a lethal recessive disorder seen in 1 in 3600 Ashkenazi Jewish births, is uncommon in most other populations.

Variation in the prevalence of genetic disease is but one aspect of human genetic variation. The evolutionary processes affecting "normal" human variation (e.g. traits such as height, skin color, blood groups, and DNA polymorphisms) also affect genetic disease variation. Four major evolutionary processes can be identified: mutation, natural selection, genetic drift, and gene flow. The relative contribution of each of these processes to genetic disease variation has been the subject of considerable controversy.

Much of this controversy arises from difficulties in specifying the precise nature of genetic variation. Until fairly recently, population variation in genetic disease was assessed primarily in terms of

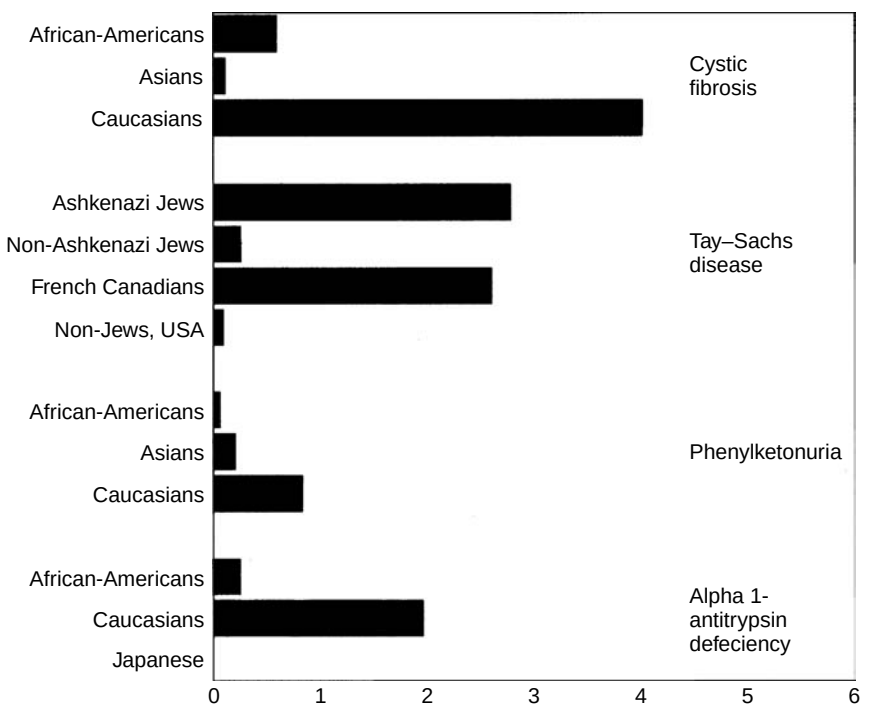

Fig. 1 Prevalence of selected genetic diseases (per 10 000 births) for a series of populations. Note that the French Canadian population refers only to those living in eastern Quebec.

protein variants. Protein electrophoresis could sometimes establish that different populations had different amino acid substitutions, leading to a disease (e.g. for sickle cell disease), but variation at the DNA level was largely unmeasurable. Furthermore, for most disease genes, the gene product, and thus the amino acid sequence, was unknown. Consequently, tracing the origins and evolution of genetic diseases in populations was often a highly speculative endeavor.

The molecular genetic techniques developed during the past two decades have improved this situation substantially. With millions of polymorphisms now identified in the human genome, it is feasible to examine polymorphisms near and within any disease locus. These polymorphisms, or markers, can take the form of restriction fragment length polymorphisms (RFLPs), short tandem repeat polymorphisms (STRPs), or single nucleotide polymorphisms (SNPs). It is becoming increasingly common to sequence entire genes or regions containing many genes in each study subject so that all variation in a region can be assessed. If several polymorphisms are measurable in the vicinity of the disease locus, haplotypes can be constructed for chromosomes carrying the disease mutation and for normal chromosomes. For example, a series of four two-allele SNPs, in which alleles are labeled 1 or 2, could form a haplotype composed of alleles 1, 2, 1, and 1 on the chromosomes bearing a particular disease mutation. Other haplotypes may be observed on normal chromosomes. Because recombination between closely linked polymorphisms is rare, these haplotypes can often be used to trace the evolutionary history of a disease mutation (i.e. under appropriate conditions, all affected individuals with the same chromosome haplotype should be descendants of the same individual). As will be shown in this article, comparisons of haplotypes in different populations can reveal whether a genetic disease had independent origins in each population.

As more and more disease genes are cloned, it is becoming possible to analyze not only marker haplotypes but also disease mutations themselves. This can be accomplished through techniques such as single-strand conformation polymorphism analysis (SSCP) or through direct DNA sequencing. Analysis of mutations sometimes reveals a tremendous diversity of different mutations at a single disease locus. Since the cloning of the CF gene (cystic fibrosis transmembrane conductance regulator, *CFTR*) in 1989, more than 1000 different mutations have been identified at this locus. As will be discussed below, these mutations vary in frequency from one population to another.

As a result of advances in molecular technology, our understanding of the role of evolutionary forces in creating and maintaining genetic variation among populations (including variation in genetic diseases) has increased considerably. This article will review the evolutionary mechanisms responsible for such variation, illustrating each mechanism with specific disease examples. It will also discuss the ways in which molecular genetic markers have contributed to our understanding of these mechanisms.

2 Mutation

Mutation, which is a heritable alteration in DNA sequence, is the ultimate source of genetic variation. While many mutations have no biological consequence for the

organism, some may produce changes in amino acid sequence that result in genetic disease. Mutations can occur as single-base substitutions (missense mutations that alter a single amino acid or nonsense mutations that produce a stop codon), deletions and duplications (some of which may produce frameshifts), alterations of splice sites (resulting in inappropriate inclusion of introns or exclusion of exons in the mature messenger RNA), promoter alterations (resulting in decreased transcription of messenger RNA), or insertions of mobile elements such as *Alu* or LINE1 repeats.

The average mutation rate in humans is approximately 10^{-9} per nucleotide per year. There is little evidence that mutation rates vary substantially among human populations. Thus, mutation itself does not account for much interpopulation variation in genetic disease. However, mutation rates do vary substantially from one disease gene to another and thus help explain why some diseases are relatively common and others are relatively rare.

2.1 Gene Size and Duchenne Muscular Dystrophy

Duchenne muscular dystrophy (DMD) is one of the more common lethal genetic diseases, affecting approximately 1 in 3500 males in most surveyed populations. Because it is an X-linked recessive disorder, it seldom affects females. Affected males usually die of cardiac or respiratory failure before age 20 and seldom reproduce. It is estimated that the DMD gene has one of the highest known mutation rates, approximately 10^{-4} (mutations per gene per generation). This high mutation rate helps to account for the fact that DMD is quite common in spite of its lethality in males.

Cloning the DMD gene has helped to explain why the mutation rate is so high. The DMD gene, which spans 2.3 million DNA bases and includes 14 000 bp (14 kilobases or kb) of coding DNA, is the largest gene known in the human. Its large size, quite simply, presents a large "target" for mutation. A similar example is given by neurofibromatosis type 1 (NF1), an autosomal dominant disorder that produces benign tumors associated with peripheral nerves. This disorder also has a high mutation rate, estimated as approximately 10^{-4}. The NF1 gene is again very large, spanning over 350 kb. The size of the coding DNA, 13 kb, is similar to that of the DMD gene. In general, mutation rate appears to correlate positively with gene size.

2.2 Mutation Hot Spots

A second factor influencing mutation rates involves DNA base-pair composition. It is known that the CG dinucleotide is highly susceptible to mutation. The cytosine base is usually methylated when it occurs next to a guanine, and methylated cytosine is likely to spontaneously lose an amino group. This produces a transition from cytosine to thymine. Surveys have shown that an inordinately large proportion of disease-causing mutations occur at CG dinucleotides. Thus, loci rich in such mutational "hot spots" are likely to have higher mutation rates. For example, achondroplasia, a reduced-stature condition, affects up to 1 in 15 000 live births, but the responsible gene (*FGFR3*, fibroblast growth factor receptor 3) spans 16.5 kb and is thus much smaller than the *NF1* or *DMD* genes. Remarkably, 99% of all mutations that cause achondroplasia occur at a single nucleotide (a methylated CG

hot spot) of *FGFR3*, where the mutation rate is several orders of magnitude higher than that of other nucleotides. Examples of other genetic diseases that have such mutational hot spots include phenylketonuria (PKU) and hemophilia A and B.

3 Natural Selection

Natural selection can be thought of as a screening process to which genetic variation is subjected. Advantageous variants are positively selected: individuals carrying the variants are more likely to survive and reproduce. This increases the frequency of the gene in the population. Disadvantageous variants will be selected against, reducing the frequency of the variant. For most genetic diseases, an equilibrium state will eventually be reached in which mutation introduces new copies of a disease-causing gene into the population while selection removes them at the same rate.

3.1 Heterozygote Advantage and Sickle Cell Disease

Certain recessive diseases present a slightly more complicated picture. In some cases, the homozygous recessive genotype may undergo negative selection, while the heterozygote may have a selective advantage. In this situation, the disease allele will be maintained at a higher frequency than if the heterozygote had no advantage. A good example of this is found in sickle cell disease. This autosomal recessive disorder, in which the erythrocytes assume a characteristic sickle shape and thus cannot move easily through the capillaries, is often fatal in homozygotes. The heterozygote has no health problems and, in a malarial environment, has a decided survival advantage. This is because sickle cell heterozygotes are resistant to infection by the *Plasmodium falciparum* malaria parasite. Consequently, heterozygotes have a survival advantage over both types of homozygotes. The frequency of the sickle cell gene is quite high in those parts of the world in which malaria has been endemic. It reaches its highest frequency in west-central Africa, where up to one in four individuals is a heterozygous carrier, and one in 45 is affected with the disorder. Sickle cell disease is also seen in parts of the Middle East, India, and the Mediterranean region. It is interesting that the frequency of the sickle cell mutation among African-Americans is substantially lower than in western Africa. This partly reflects admixture with Caucasians (see below), but it is likely also a result of a lack of natural selection in the malaria-free North American environment.

The best-studied of the sickle cell mutations is a single-base substitution (thymine instead of adenine) at the DNA triplet that specifies the sixth amino acid of the β-globin chain. This mutation results in a substitution of valine for glutamic acid known as the hemoglobin S (HbS) variant. At one time, it was thought that the presence of HbS in India and the Mideast was due to migration of Africans who carried the *HbS* gene. However, more recent investigations using RFLPs show that this is probably not the case. Marker haplotypes have been examined, using RFLPs in and around the β-globin locus. They show that the disease haplotypes occurring in Africa differ substantially from those occurring in the Mideast and India. Some haplotype variation can be expected as a result of gene conversion or recombination subsequent to the origin

of a mutation, but it is often possible to distinguish these events from multiple mutations. The RFLP haplotypes show that the HbS mutation arose at least twice, once in Africa and once in Saudi Arabia or India. Natural selection then operated in each environment independently to raise the frequency of the mutant gene. Molecular studies have shown that $\Psi\beta$, a pseudogene (i.e. a nontranscribed gene highly similar to β-globin in DNA sequence), does *not* increase in frequency among populations in malarial environments.

3.2
Natural Selection and HIV

Another example of natural selection came to light because of the human immunodeficiency virus (HIV). Some strains of this virus gain entry to macrophages and helper T cells via a cell-surface receptor that normally binds to a cytokine. The resulting viral destruction of helper T cells, a critical component of the body's immune system, leads to acquired immune deficiency syndrome (AIDS). Individuals who are homozygous for a 32-bp deletion of the gene that encodes one of the cytokine receptors (*CCR5*) lack the receptor on their cell surfaces and are thus remarkably resistant to HIV infection (because other cell-surface receptors can perform the same role as CCR5, the mutation has no ill effects). This deletion is especially common in northeastern European populations, where its frequency reaches 0.20. It is virtually absent in Asian and African populations. Analysis of haplotype variation in the chromosomal region that contains *CCR5* indicates that the deletion arose in European populations about 700 to 2000 years ago. Since HIV appeared in humans only a few decades ago, the high deletion frequency in northeastern Europe must be due to a selective force other than HIV or perhaps to genetic drift. Considering the age of the deletion, it was first suggested that it had conferred resistance against the bacterium, *Yersinia pestis*, that causes bubonic plague. However, recent research with mouse models demonstrated that *Y. pestis*' entry into cells is not affected by the absence of the CCR5 molecule. Other evidence suggests that the selective agent may instead have been the smallpox virus.

4
Genetic Drift and Founder Effect

Often, small, isolated populations exhibit high frequencies of genetic diseases that are rare in other populations. For example, Ellis van Creveld syndrome, a very rare disorder that involves reduced stature, is seen with increased frequency among the genetically isolated Old Order Amish population of Lancaster County, Pennsylvania. In fact, nearly as many cases have been observed in a single Amish kindred as have been observed in the rest of the world. In the Finnish population, which has also been quite isolated until recently, over two dozen otherwise rare autosomal recessive diseases occur with substantially elevated frequencies. Such variation in disease prevalence is usually the result of the related processes of genetic drift and founder effect.

Genetic drift refers to the random fluctuation in gene frequencies that occurs from one generation to the next as a result of sampling a limited number of gametes. In small populations, the degree of fluctuation in gene frequency increases (just as a coin-tossing experiment using only 10 coins is quite likely to result in a large deviation from the expected proportion of

50% heads and 50% tails). Genetic drift can thus lead to high frequencies of genetic diseases in small populations. Founder effect occurs when a small number of individuals form a new population. Because these individuals form a small sample of an original population, their genes are unlikely to form a representative sample of the original population from which they came. This can again result in a highly skewed distribution of disease genes among their descendants. The Old Order Amish population, for example, was founded by only about 50 couples. This effect is further exaggerated when certain founders have a disproportionately large number of descendants.

Traditionally, demographic history and genealogies have been used to demonstrate drift and founder effect. Genealogical analysis demonstrated, for example, that all 30 000 cases of porphyria variegata in the South African Afrikaner population could be traced to a single couple who emigrated from Holland in the 1680s. This approach is obviously limited by the availability of extensive genealogies.

4.1 Haplotype Analysis

More recently, molecular genetic markers have made it possible to assess founder effect at the DNA level. Haplotype analysis is particularly helpful in this regard. It is reasonable to assume that each different haplotype background associated with a disease represents a different founder. Thus, the number of distinct disease haplotypes can be counted to infer the number of founders who introduced a disease into a population. In northern Finland, it was shown that all individuals affected with X-linked choroideremia, a disease that causes blindness, had the same RFLP haplotype (using markers closely linked to the choroideremia gene). This provided evidence that each of these individuals was descended from the same common ancestor. Genealogical analysis proved this to be the case: all affected individuals shared a common ancestor that lived 12 generations ago. A similar exercise showed that the high frequency of PKU among Yemeni Jews can be attributed to the descendants of a single founding couple who lived in the seventeenth century. Haplotype analysis of markers closely linked to the *CFTR* locus demonstrated that all studied cases of CF in the Hutterite population could be traced to three founders.

The disease consequences of founder effect and genetic drift have been especially well characterized in the Ashkenazi Jewish population. For example, torsion dystonia, a rare recessive disorder, has an elevated frequency in this population. Analysis of haplotype variation has shown that the disease-causing mutation is only about 350 years old and has likely been amplified in frequency because of population bottlenecks and subsequent expansion. Similarly, three mutations in the *BRCA1* and *BRCA2* genes have reached a combined frequency of 2.5% in Ashkenazi Jews and account for a substantial proportion of inherited breast cancer cases. A mutation in the *APC* gene is found in about 1/20 members of this population and confers a twofold increased risk of developing colon cancer. This mutation alone accounts for about 10% of colon cancer cases in the Ashkenazi population.

4.2 Mutation Analysis

The cloning of disease genes often permits direct analysis of population variation

in mutations. One of the best examples is the ΔF508 (deletion of a phenylalanine residue at position 508) mutation in the *CFTR* locus, a three-base deletion that produces a loss of function and accounts for approximately 70% of the *CFTR* mutations found in Caucasians. Extensive population analysis has shown that the proportion of *CFTR* mutations accounted for by ΔF508 has a north–south cline, with a maximum of about 70 to 90% in northern Europe (where the disease itself is most common) and a minimum of about 50% in southern Europe. The proportion of ΔF508 mutations is even lower among other surveyed populations, such as African-Americans (37%) and Ashkenazi Jews (30%). The elevated frequency of CF among Europeans can be traced primarily to the ΔF508 mutation, since the frequencies of non-ΔF508 mutations are quite similar among Europeans, Asians, and Africans. Among Europeans, the great majority (>90%) of ΔF508 mutations occur on a single haplotype background, indicating that most current copies of the mutation may have descended from the same common ancestor.

Cloning of the *CFTR* gene has helped explain why CF has a high frequency in European populations. *CFTR* encodes a chloride channel found on the surfaces of epithelial cells that line the gut and the airway. The number of chloride channels in these cells is substantially reduced in transgenic mice that carry one copy of the ΔF508 mutation. The bacterium that causes typhoid fever, *Salmonella typhi*, gains access to intestinal cells via these chloride channels, and, as predicted, heterozygous mice are quite resistant to typhoid fever infection. It is reasonable to conclude that humans who were heterozygous for this mutation also benefited from resistance to typhoid fever.

5 Drift or Selection?

When a genetic disease is seen with high frequency in a population, it is often unclear whether it is caused by a selective advantage or genetic drift/founder effect. In a few cases (such as sickle cell disease or CF), a demonstrable selective agent is known. But in most cases, the evidence is far more equivocal. Here again, molecular markers have been helpful.

5.1 Hemochromatosis in Europe

Idiopathic hemochromatosis is an autosomal recessive disorder in which excess iron is absorbed in the gut and deposited in many parts of the body, including the skin, joints, liver, pancreas, and heart. Consequently, some homozygotes develop liver disease, heart failure, arthropathy, diabetes, and other disease-related conditions late in life. Approximately 1 in 300 Europeans is homozygous for mutations in the responsible gene, *HFE*. Although several different *HFE* mutations have been reported, 85% of them consist of a cysteine→tyrosine substitution that results in a loss of function of the *HFE* product. Haplotype analysis has shown that this mutation has occurred only once and is confined almost exclusively to the European population. Analysis of variation in markers near *HFE* has shown that the primary disease-causing mutation took place only about 60 to 70 generations ago. The mutation could have achieved a high frequency in this population either through drift or natural selection, but extensive analysis of marker variation near the *HFE* gene has shown that

it is extremely improbable that drift alone could have elevated the mutation frequency so quickly. In addition, population surveys show that heterozygotes have a significant reduction in the frequency of iron-deficiency anemia. Thus, a heterozygote advantage is likely to account for a rapid increase in the frequency of this mutation.

5.2
G6PD Deficiency in Africa and the Mediterranean

G6PD (glucose-6-phosphate-dehydrogenase) deficiency is in X-linked disorder that affects 400 million individuals worldwide and can result in hemopathology when mutation carriers are exposed to specific infections, drugs (e.g. primaquine), or foods (e.g. fava beans). G6PD deficiency, like sickle cell disease, confers resistance to infection by the malaria parasite, *P. falciparum*. Molecular analysis has shown that a major mutation, A-, is common in Africa, while another major mutation, Med, is common in Mideastern and Mediterranean populations. The A-mutation occurred approximately 6000 years ago, and the Med mutation occurred approximately 3000 years ago. The recent ages of these mutations reflect limited haplotype diversity and indicate that genetic drift alone could not have elevated these mutations to high frequency in just a few thousand years. Natural selection, as a result of malarial resistance, must have played an important role in raising the frequencies of these mutations. It is interesting that the form of *P. falciparum* responsible for severe malaria probably did not arise until about 10 000 years ago, after the introduction of agriculture in these areas.

6
Gene Flow

Gene flow reduces genetic differences among populations. When two populations both manifest a genetic disease, the question often arises whether the disease mutation arose independently in the two populations or was communicated from one to another via gene flow. When only a disease phenotype can be observed, it is difficult to answer this question. However, more precise molecular characterization of disease loci often illuminates the issue.

6.1
Cystic Fibrosis in African-Americans

An example is given by CF in the African-American population. This disease affects approximately 1 in 17 000 African-Americans, a prevalence much lower than that of Europeans. One hypothesis is that the disorder was introduced into the African-American population through gene flow from the European-American population. However, analysis of CF mutations shows that ΔF508 is quite rare among African-Americans and that two-thirds of the observed CF mutations are unique to this population. Thus, CF almost certainly had independent origins in the African-American population and is not principally derived from European-Americans.

6.2
Tay–Sachs Disease in French Canadians

Studies of Tay–Sachs disease in Quebec provide another example. As one would expect, the Ashkenazi Jewish population of this region has an elevated prevalence of this disorder, and the frequencies of

the two major mutations closely match those of other Ashkenazi Jewish populations. Tay–Sachs disease is also seen among the French Canadian population of Quebec. In fact, in the eastern portion of the province, the disease gene frequency is as high as among Ashkenazi Jews. Does this reflect gene flow from the Jewish population, or was there an independent origin? Direct examination of the Tay–Sachs gene shows that the French Canadians usually carry a 7.6-kb deletion that is not found among Jews. Thus, an independent origin is established. Further investigations may ultimately establish whether the high frequency of the disorder among French Canadians is the result of founder effect or natural selection.

7 Genetic Variation and Common Diseases

The examples used thus far have all involved relatively rare single-gene diseases. In each case, mutations have increased in frequency as a result of factors such as drift or selection. Increasing attention is now being devoted to mutations responsible for common diseases such as diabetes, heart disease, and common cancers. The most significant of these mutations are those that have reached high frequency (e.g. >10%). Typically, such mutations are relatively old and are thus shared among most worldwide populations. Their frequencies vary among populations, however. For example, the Є 4 allele of the apolipoprotein E (*APOE*) gene, which confers an increased risk of developing Alzheimer's disease, exhibits frequencies above 30% among some African populations (e.g. Pygmies, Khoi San, Sudanese, and Nigerians) but substantially lower (<15%) among southern Europeans and Asians (there is also evidence, however, that the association between Є 4 and Alzheimer's disease is weaker among individuals of African descent than among those of European descent). A gain-of-function mutation in the *SCN5A* gene, which encodes a sodium channel present in cardiac myocytes, confers an eightfold increased risk of cardiac arrhythmia and was found in approximately 13% of African-Americans but was absent in samples of 511 Europeans and 578 Asians. A promoter variant of the angiotensinogen (*AGT*) gene is associated with an increased risk of hypertension and reaches high frequencies in African populations. Haplotype analysis has demonstrated that natural selection has acted to decrease the frequency of this variant in non-African populations, possibly as a result of their adaptation to nontropical, high-sodium environments after they left the African continent. Differences in the frequencies of mutations such as these may help account for differences in the prevalence of common diseases in populations, although environmental factors such as diet must also have important effects. It is critical to point out that, because of the recent origin of modern humans from a small founder population and continued gene flow among populations through time, most of these variants tend to be shared among major populations and typically vary only in their frequencies.

8 Perspectives

This review has highlighted some ways in which the study of disease variation in populations can be related to evolutionary processes. Such studies are significant

in understanding evolutionary biology. They have other implications as well. An awareness of population variation in disease can aid in clinical diagnosis. For example, a physician evaluating a child who presents with recurrent respiratory infections and failure to thrive should include ethnic background in considering a differential diagnosis. If the patient is of African descent, sickle cell disease is a distinct possibility. If the patient is of northern European descent, sickle cell disease would be unlikely (but not impossible), whereas cystic fibrosis should be considered. In some cases, knowledge of the potential adaptive significance of a genetic disease (e.g. the heterozygote advantage for sickle cell disease, cystic fibrosis, or hemochromatosis) may lead to increased understanding of gene function and pathophysiology.

When considering genetic variants that contribute to common diseases, as well as those that influence response to therapeutic drugs, it should be emphasized that most such variants are shared among most major human populations, varying only in their frequencies. Consequently, population affiliation (or "race") is, at best, a very approximate indicator of critical factors such as drug response.

Advances in molecular genetics have greatly improved our understanding of the evolutionary processes leading to population variation in genetic disease. Yet many important questions remain unanswered, particularly for common, complex diseases. As more polymorphic markers are analyzed, more DNA sequenced, and more disease genes cloned, new and more profound insights are certain to be gained.

See also Genetic Analysis of Populations; Oncology, Molecular.

Bibliography

Books and Reviews

Bamshad, M., Wooding, S.P. (2003) Signatures of natural selection in the human genome, *Nat. Rev. Genet.* **4**(2), 99–111.

Crow, J.F. (2000) The origins, patterns and implications of human spontaneous mutation, *Nat. Rev. Genet.* **1**(1), 40–47.

Goldstein, D.B., Chikhi, L. (2002) Human migrations and population structure: what we know and why it matters, *Annu. Rev. Genomics Hum. Genet.* **3**, 129–152.

Jobling, M.A., Hurles, M.E., Tyler-Smith, C. (2003) *Human Evolutionary Genetics: Origins, Peoples, and Disease*, Garland Science, New York.

Jorde, L.B., Watkins, W.S., Bamshad, M.J. (2001) Population genomics: a bridge from evolutionary history to genetic medicine, *Hum. Mol. Genet.* **10**, 2199–2207.

Jorde, L.B., Carey, J.C., Bamshad, M.J., White, R.L. (2003) *Medical Genetics*, 3rd edition, Mosby, St. Louis, MO.

Kwiatkowski, D. (2000) Genetic susceptibility to malaria getting complex, *Curr. Opin. Genet. Dev.* **10**(3), 320–324.

McKusick, V.A. (Ed.) (1978) *Medical Genetic Studies of the Amish: Selected Papers*, Johns Hopkins University Press, Baltimore.

Ostrer, H. (2001) A genetic profile of contemporary Jewish populations, *Nat. Rev. Genet.* **2**(11), 891–898.

Tishkoff, S.A., Verrelli, B.C. (2003) Patterns of human genetic diversity: implications for human evolutionary history and disease, *Annu. Rev. Genomics Hum. Genet.* **4**(1), 293–340.

Primary Literature

Abadie, V., Lyonnet, S., Maurin, N., Berthelon, M., Caillaud, C., Giraud, F., Mattei, J.F., Rey, J., Rey, F., Munnich, A. (1989) CpG dinucleotides are mutation hot spots in phenylketonuria, *Genomics* **5**(4), 936–939.

Ajioka, R.S., Jorde, L.B., Gruen, J.R., Yu, P., Dimitrova, D., Barrow, J., Radisky, E., Edwards, C.Q., Griffen, L.M., Kushner, J.P. (1997) Haplotype analysis of hemochromatosis: evaluation of different linkage-disequilibrium measures and evolution of disease chromosomes, *Am. J. Hum. Genet.* **60**, 1439–1447.

Allison, A.C. (1961) Genetic factors in resistance to malaria, *Ann. N.Y. Acad. Sci.* **91**, 710–729.

Avigad, S., Cohen, B.E., Bauer, S., Schwartz, G., Frydman, M., Woo, S.L., Niny, Y., Shiloh, Y. (1990) A single origin of phenylketonuria in Yemenite Jews, *Nature* **344**(6262), 168–170.

Bamshad, M., Wooding, S., Salisbury, B.A., Stephens, J.C. (2004) Deconstructing the relationship between genetics and race, *Nat. Rev. Genet.* **5**(8), 598–609.

Bamshad, M.J., Mummidi, S., Gonzalez, E., Ahuja, S.S., Dunn, D.M., Watkins, W.S., Stone, A.C., Jorde, L.B., Weiss, R.B., Ahuja, S.K. (2002) A strong signature of balancing selection in the promoter of CCR5, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10539–10544.

Bobadilla, J.L., Macek, M. Jr, Fine, J.P., Farrell, P.M. (2002) Cystic fibrosis: a worldwide analysis of CFTR mutations-correlation with incidence data and application to screening, *Hum. Mutat.* **19**(6), 575–606.

Cawthon, R.M., Weiss, R., Xu, G.F., Viskochil, D., Culver, M., Stevens, J., Robertson, M., Dunn, D., Gesteland, R., O'Connell, P. (1990) A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations, *Cell* **62**(1), 193–201.

Chen, S.-H., Zhang, M., Lovrien, E.W., Scott, C.R., Thompson, A.R. (1991) CG dinucleotide transitions in the factor IX gene account for about half of the point mutations in hemophilia B patients: a Seattle series, *Hum. Genet.* **87**, 177–182.

Cutting, G.R., Curristin, S.M., Nash, E., Rosenstein, B.J., Lerer, I., Abeliovich, D., Hill, A., Graham, C. (1992) Analysis of four diverse population groups indicates that a subset of cystic fibrosis mutations occur in common among Caucasians, *Am. J. Hum. Genet.* **50**, 1185–1194.

Dean, M., Carrington, M., Winkler, C., Huttley, G.A., Smith, M.W., Allikmets, R., Goedert, J.J., Buchbinder, S.P., Vittinghoff, E., Gomperts, E., Donfield, S., Vlahov, D., Kaslow, R., Saah, A., Rinaldo, C., Detels, R., O'Brien, S.J. (1996) Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, A, *Science* **273**(5283), 1856–1862.

Diamond, J.M., Rotter, J.I. (1987) Observing the founder effect in human evolution, *Nature* **329**, 105–106.

Eisensmith, R.C., Okano, Y., Dasovich, M., Wang, T., Guttler, F., Lou, H., Guldberg, P., Lichter-Konecki, U., Konecki, D.S., Svensson, E. (1992) Multiple origins for phenylketonuria in Europe, *Am. J. Hum. Genet.* **51**, 1355–1365.

European Working Group on CF Genetics. (1990) Gradient of distribution in Europe of the major CF mutation and of its associated haplotype, *Hum. Genet.* **85**, 436–445.

Farrer, L.A., Cupples, L.A., Haines, J.L., Human, B., Kukull, W.A., Mayeux, R., Myers, R.H., Pericak-Vance, M.A., Risch, N., van Duijn, C.M. (1997) Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis, *JAMA* **278**, 1349–1356.

Feder, J.N., Gnirke, A., Thomas, W., Tsuchihashi, Z., Ruddy, D.A., Basava, A., Dormishian, F., Domingo, R. Jr., Ellis, M.C., Fullan, A., Hinton, L.M., Jones, N.L., Kimmel, B.E., Kronmal, G.S., Lauer, P., Lee, V.K., Loeb, D.B., Mapa, F.A., McClelland, E., Meyer, N.C., Mintier, G.A., Moeller, N., Moore, T., Morikang, E., Wolff, R.K. (1996) A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis, *Nat. Genet.* **13**(4), 399–408.

Frisch, A., Colombo, R., Michaelovsky, E., Karpati, M., Goldman, B., Peleg, L. (2004) Origin and spread of the 1278insTATC mutation causing Tay-Sachs disease in Ashkenazi Jews: genetic drift as a robust and parsimonious hypothesis, *Hum. Genet.* **114**(4), 366–376.

Galvani, A.P., Slatkin, M. (2003) Evaluating plague and smallpox as historical selective pressures for the CCR5-Delta 32 HIV-resistance allele, *Proc. Natl. Acad. Sci. U.S.A.* **100**(25), 15276–15279.

Hill, A.V.S. (1989) Molecular Markers of Ethnic Groups, in: Cruickshank, J.K., Beevers, D.G. (Eds.) *Ethnic Factors in Health and Disease*, Wright, London, pp. 25–31.

Hoyer, L.W. (1994) Hemophilia A, *N. Engl. J. Med.* **330**, 38–47.

Jeunemaitre, X., Soubrier, F., Kotelevstev, Y.V., Lifton, R.P., Williams, C.S., Charru, A., Hunt, S.C., Hopkins, P.N., Williams, R.R., Lalouel, J.-M., Corvol, P. (1992) Molecular basis of human hypertension: role of angiotensinogen, *Cell* **71**, 169–180.

Kere, J. (2001) Human population genetics: lessons from Finland, *Annu. Rev. Genomics Hum. Genet.* **2**, 103–128.

Kerem, B., Rommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A., Buchwald, M., Tsui, L. (1989) Identification of the cystic fibrosis gene: genetic analysis, *Science* **245**, 1073–1080.

Kirchhoff, T., Satagopan, J.M., Kauff, N.D., Huang, H., Kolachana, P., Palmer, C., Rapaport, H., Nafa, K., Ellis, N.A., Offit, K. (2004) Frequency of BRCA1 and BRCA2 mutations in unselected Ashkenazi Jewish patients with colorectal cancer, *J. Natl. Cancer Inst.* **96**(1), 68–70.

Klinger, K., Horn, G.T., Stanislovitis, P., Fujiwara, T.M., Morgan, K. (1990) Cystic fibrosis mutations in the Hutterite brethren, *Am. J. Hum. Genet.* **46**, 983–987.

Koenig, M., Hoffman, E.P., Bertelson, C.J., Monaco, A.P., Feener, C., Kunkel, L.M. (1987) Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals, *Cell* **50**(3), 509–517.

Kondrashov, A.S. (2003) Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases, *Hum. Mutat.* **21**(1), 12–27.

Kwiatkowski, D. (2000) Genetic susceptibility to malaria getting complex, *Curr. Opin. Genet. Dev.* **10**(3), 320–324.

Myerowitz, R., Hogikyan, N.D. (1986) Different mutations in Ashkenazi Jewish and non-Jewish French Canadians with Tay-Sachs disease, *Science* **232**, 1646–1648.

Nakajima, T., Jorde, L.B., Ishigami, T., Umemura, S.,
Emi, M., Lalouel, J.M., Inoue, I. (2002) Nucleotide diversity and haplotype structure of the human angiotensinogen gene in two populations, *Am. J. Hum. Genet.* **70**(1), 108–123.

Nakajima, T., Wooding, S., Sakagami, T., Emi, M., Tokunaga, K., Tamiya, G., Ishigami, T., Umemura, S., Munkhbat, B., Jin, F., Guan-Jun, J., Hayasaka, I., Ishida, T., Saitou, N., Pavelka, K., Lalouel, J.M., Jorde, L.B., Inoue, I. (2004) Natural selection and population history in the Human Angiotensinogen Gene (AGT): 736 complete AGT sequences in chromosomes from around the world, *Am. J. Hum. Genet.* **74**(5), 898–916.

Niell, B.L., Long, J.C., Rennert, G., Gruber, S.B. (2003) Genetic anthropology of the colorectal cancer-susceptibility allele APC I1307K: evidence of genetic drift within the Ashkenazim, *Am. J. Hum. Genet.* **73**(6), 1250–1260.

O'Brien, S.J. (1998) AIDS: a role for host genes, *Hosp. Pract. (Off. Ed.)* **33**(7), 53–6, 59–60, 66–7. passim.

Pier, G.B., Grout, M., Zaidi, T., Meluleni, G., Mueschenborn, S.S., Banting, G., Ratcliff, R., Evans, M.J., Colledge, W.H. (1998) Salmonella typhi uses CFTR to enter intestinal epithelial cells, *Nature* **393**, 79–82.

Risch, N., de Leon, D., Ozelius, L., Kramer, P., Almasy, L., Singer, B., Fahn, S., Breakfield, X., Bressman, S. (1995) Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population, *Nat. Genet.* **9**, 152–159.

Rousseau, F., Bonaventure, J., Legeal-Mallet, L., Pelet, A., Rozet, J., Maroteaux, P., Le Merrer, M., Munnich, A. (1994) Mutations in the gene encoding fibroblast growth factor receptor-3 in achondroplasia, *Nature* **371**, 252–254.

Ruwende, C., Khoo, S.C., Snow, R.W., Yates, S.N.R., Kwiatkowski, D., Gupta, S., Warn, P., Allsopp, C.E., Gilbert, S.C., Peschu, N. (1995) Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to malaria, *Nature* **376**, 246–249.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E.S. (2002) Detecting recent positive selection in the human genome from haplotype structure, *Nature* **419**(6909), 832–837.

Sankila, E.-M., de la Chapelle, A., Kärnä, J., Forsius, H., Frants, R., Eriksson, A. (1987) Choroideremia: close linkage to DXYS1 and DXYS12 demonstrated by segregation analysis

and historical-genealogical evidence, *Clin. Genet.* **31**, 315–322.

Scriver, C.R. (2001) Human genetics: lessons from Quebec populations, *Annu. Rev. Genomics Hum. Genet.* **2**, 69–101.

Splawski, I., Timothy, K.W., Tateyama, M., Clancy, C.E., Malhotra, A., Beggs, A.H., Cappuccio, F.P., Sagnella, G.A., Kass, R.S., Keating, M.T. (2002) Variant of SCN5A sodium channel implicated in risk of cardiac arrhythmia, *Science* **297**(5585), 1333–1336.

Stephens, J.C., Reich, D.E., Goldstein, D.B., Shin, H.D., Smith, M.W., Carrington, M., Winkler, C., Huttley, G.A., Allikmets, R., Schriml, L., Gerrard, B., Malasky, M., Ramos, M.D., Morlot, S., Tzetis, M., Oddoux, C., Giovine, F.S., Nasioulas, G., Chandler, D., Aseev, M., Hanson, M., Kalaydjieva, L., Glavac, D., Gasparini, P., Dean, M. (1998) Dating the origin of the CCR5-Δ32AIDS-resistance allele by the coalescence of haplotypes, *Am. J. Hum. Genet.* **62**(6), 1507–1515. di.

Struewing, J.P., Hartge, P., Wacholder, S., Baker, S.M., Berlin, M., McAdams, M., Timmerman, M.M., Brody, L.C., Tucker, M.A. (1997) The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews, *N. Engl. J. Med.* **336**, 1401–1408.

Tishkoff, S.A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J., Piro, A., Stoneking, M., Tagarelli, A., Tagarelli, G., Touma, E.H., Williams, S.M., Clark, A.G. (2001) Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance, *Science* **293**(5529), 455–462.

Toomajian, C., Ajioka, R.S., Jorde, L.B., Kushner, J.P., Kreitman, M. (2003) A method for detecting recent selection in the human genome from allele age estimates, *Genetics* **165**(1), 287–297.

Volkman, S.K., Barry, A.E., Lyons, E.J., Nielsen, K.M., Thomas, S.M., Choi, M., Thakore, S.S., Day, K.P., Wirth, D.F., Hartl, D.L. (2001) Recent origin of Plasmodium falciparum from a single progenitor, *Science* **293**(5529), 482–484.

Immune Defence, Cell Mediated

Martin F. Bachmann[1] *and Thomas M. Kundig*[2]
[1] *Cytos Biotechnology AG, Zürich-Schlieren, Switzerland*
[2] *University of Zürich, Zürich, Switzerland*

1 Introduction 341

2 Innate Immunity 343
2.1 Epithelial Cells 343
2.2 Neutrophils 344
2.3 Monocytes and Macrophages 345
2.4 Natural Killer Cells 345
2.5 Eosinophil Granulocytes 346
2.6 Basophils and Mast Cells 347
2.7 Tissue Cells 347
2.8 Platelets 348

3 Adaptive Immunity 348
3.1 Cell Mediated versus Antibody Mediated Immunity 348
3.2 Induction of Cell-mediated Immunity: T-Cell Activation 349
3.2.1 Costimulation 349
3.2.2 Duration of Antigen Presentation 350
3.2.3 DC Maturation 350
3.3 Extravasation 351
3.4 Delayed Type Hypersensitivity Reaction 351
3.5 Lymphoid Organ-like Structures in Chronic Inflammation 352

4 Mechanisms of T-Cell Mediated Immunity 353
4.1 Th Cells (CD4+) 353
4.2 Cytotoxic T Cells (CD8+) 354

5 Silencing of the Immune Response: What Goes Up Must Come Down 354

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

6 T-Cell Memory 355

Bibliography 357
Books and Reviews 357
Primary Literature 358

Keywords

B cells
Lymphocytes generated in the bone marrow that secrete antibodies after activation.

T cells
Lymphocytes generated in the bone marrow and educated in the thymus to recognize peptides presented by major histocompatibility antigens.

T helper cells (Th cells)
A subset of T cells orchestrating the immune response and helping B cells to produce antibodies.

Th1 cells
A subset of Th cells, characterized by the production of IFNγ and driving effector responses against intracellular pathogens, bacteria, and fungi.

Th2 cells
A subset of Th cells, characterized by the production of IL-4, IL-5, and IL-13, driving effector responses against multicellular parasites.

Cytotoxic T cells (CTLs)
A subset of T cells killing infected host cells and tumor cells.

Dendritic cells (DCs)
The most potent antigen-presenting cells. Essential for the induction of T-cell responses.

Many cell types make up the arsenal of the immune system and are directly involved in the destruction of invading pathogens. Broadly, two types of cells can be distinguished: firstly, some cells recognize signals that are associated with pathogens, such as tissue damage or patterns found on pathogens but not in the host. These cells belong to the innate immune system and appeared first during evolution. The second type of cells, the lymphocyte, has developed a complex machinery that allows each lymphocyte to express a single and unique receptor with random specificity,

the B- and T-cell receptors. Once these receptors recognize a pathogen, the lymphocyte encoding the receptor undergoes clonal expansion and the lymphocyte population differentiates into effector cells, ridding the pathogen. Lymphocytes therefore make up the specific or adaptive immune system. This article discusses how the various cell types protect from infection with pathogens and elucidates some of the complex interactions between the two systems.

1 Introduction

Metazoan organisms have evolved defense systems to a level of near perfection, such that severe or sustained infections are rare. The immune system is divided into an innate system and an adaptive system. The innate immune system acts rapidly and is the most important part of immune defense, as most organisms are able to survive through innate immune mechanisms alone. Only vertebrates have evolved additional mechanisms for pathogen recognition and elimination, so-called *adaptive immunity*.

The innate immune system involves a broad array of different cell types, as many cells and organs of the host have evolved to defend themselves against microbes. For decades, the innate immune system has been regarded as somewhat "unspecific" or even "primitive." However, in recent years it has become more and more apparent that the strategy of the innate immune system is to detect molecules that are common and conserved in a broad range of microorganisms and that are different from molecules of the host itself. It has been discovered that the innate immune system recognizes certain molecular patterns that are present in microbes but not in the host, so-called *pathogen-associated microbial patterns* (PAMPs). The receptors recognizing those microbial patterns are called *pattern recognition receptors* (PRRs). Many cell types of the innate immune system express PRRs, so that the innate immune system can react very rapidly.

In contrast to the innate immune system that has certain specificity for microbes, late evolution in vertebrates has developed an even more specific and more complex system of so-called *adaptive immunity*. The adaptive immune system is able to specifically recognize any protein (or even sequences thereof) that is not normally present in the host, and thus represents an invader. The key players in mounting an adaptive immune response are dendritic cells that induce T- and B-cell responses. The adaptive immune system's high specificity has the disadvantage that early in an infection, only few T and B cells that can recognize the microbe are present and these few specific T and B cells first have to proliferate to an "army of relevant size." This proliferation takes several days during which the innate immune system is the one fighting the microbes. Also, the innate immune system produces an array of different signals in the form of cytokines and chemokines as well as cognate interactions to support the adaptive immune system. Thus, the battle of the innate immune system against a microbe may be interpreted as a "wake-up call" to the adaptive immune system. Once an adaptive immune response has been generated, which includes a microbe

specific antibody- and T-cell response, the adaptive immune system also supports the innate immune system, for example, by coating microbes with antibodies so that they can be better phagocytosed (so-called *opsonization*), or by providing T-cell dependent cytokines that enhance the phagocytic activity of the innate immune cells.

Only very few of the microorganisms present in the environment pass all the lines of defense in a vertebrate host. Already in order to pass the first barrier, that is, chemical and physical barriers such as mucus, saliva, cilia movements, and epithelia with tight junctions, specialized pathogen receptors are required for pathogen adherence and penetration. Even fewer microorganisms will pass the innate immune system that has cells positioned in close association with the skin and mucosal surfaces, which are the usual sites of microbe entry. Microbes that have been eliminated either by physical or chemical barriers or by the innate immune system do not usually cause clinically manifest disease. Only those microbes that have passed all these lines of defense are dangerous for the host and will elicit a specific immune response and should actually be called *pathogens*. The adaptive immune response, in concert with an enhanced innate immune response, finally eliminates the pathogen. The immune

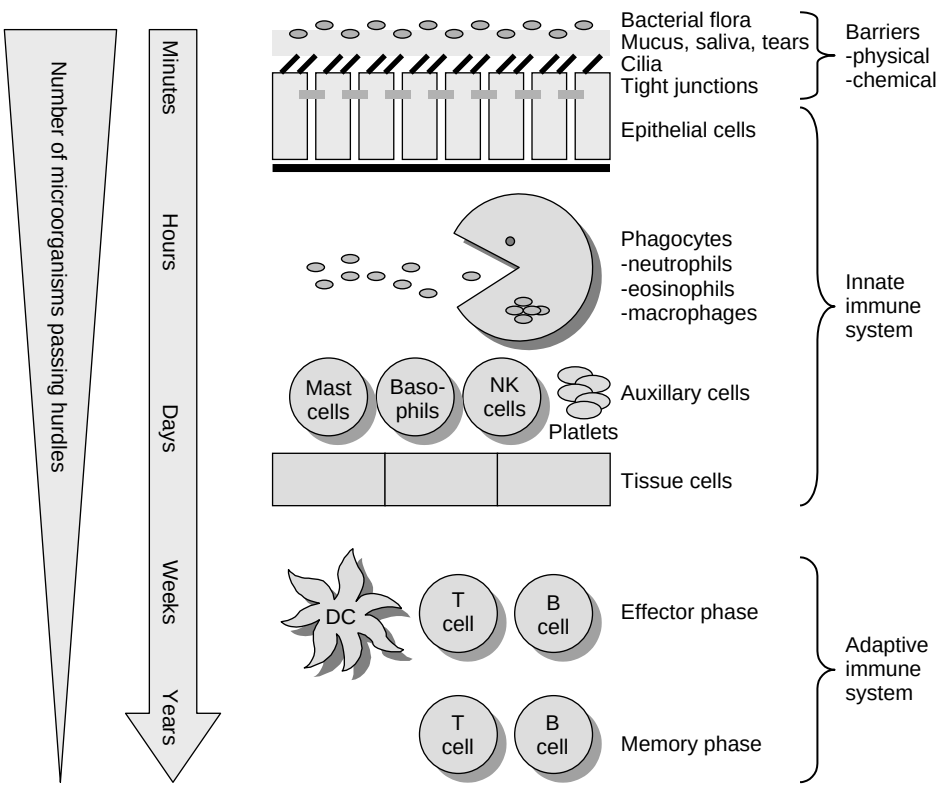

Fig. 1 Cells of the innate and adaptive immune response. Number of microorganisms that pass lines of defense and induce a response, as well as approximate time for response.

system thereafter tries to keep these dangerous microbes in mind by generating an immunological memory. Even here there are different levels of immunological memory that again depend on the virulence and persistence of the pathogen. Against persisting pathogens, for example, tuberculosis or other pathogens inducing granulomas, immunological memory is strong with pathogen-specific T cells constantly activated and alert. In addition, B cells are repeatedly restimulated to produce high and persistent levels of antibody. In contrast, if the pathogen does not persist, the immune system keeps the pathogen-specific B- and, in particular, T cells at a functionally lower level.

Figure 1 summarizes the three major lines of defense, that is, physical and chemical barriers, followed by the innate immune system consisting of phagocytes, auxiliary cells, and tissue cells. Only very few pathogens pass these first lines of defense so that an adaptive immune response is generated. An arrow indicates the approximate time frames it takes for the immune system to react.

2 Innate Immunity

2.1 Epithelial Cells

Epithelial cells have a variety of immunological functions. Most microorganisms that an individual encounters do not penetrate the body surface, but are prevented from entering by external or internal epithelial surfaces. Epithelial cells are joined by tight junctions that form an effective seal against the outside environment. Moreover, the epithelial lining of the respiratory tract is covered by sticky mucus. Possible invaders are trapped in the mucus and are then transported outward by moving cilia on the respiratory epithelium.

In addition to these mechanical defense mechanisms, the skin and mucosa also produce antimicrobial chemical substances that cover their surfaces. Tears and saliva contain the antimicrobial enzyme lysozyme. Epithelial cells also release low-molecular-weight antimicrobial peptides. The best characterized peptides belong to the family of so-called *defensins*, which have molecular masses between 3 to 6 kD and have the physicochemical properties to insert into membranes and disrupt their structure. Defensins are subdivided into two structural classes, alpha (α) and beta (β). So far, six α-defensins and three β-defensins have been described in humans. The α-defensins are produced primarily by intestinal Paneth cells, whereas airway epithelial cells produce mostly β-defensins. Defensins have a broad-spectrum antimicrobial activity ranging from gram-negative to gram-positive bacteria, as well as some viruses and fungi. The digestive tract produces antimicrobially active acid, digestive enzymes, and α-defensins. The lungs produce surfactant proteins A and D that coat the pathogen surface and enhance phagocytosis by macrophages. Last but not least, most epithelia are covered with a flora of nonpathogenic bacteria that compete with pathogenic bacteria for nutrients and can also produce additional antibacterial peptides and other substances.

Apart from the above-mentioned direct antimicrobial activity, epithelial cells are important regulators of the immune response. Epithelial cells produce numerous cytokines, cytokine antagonists, chemokines, colony-stimulating factors, and growth factors that all have the capacity to recruit multiple inflammatory cells.

Epithelial cells of mucosal surfaces also contribute to mucosal immunity through secretion of IgA. The basal epithelial cell surface expresses the secretory component, acting as a specific receptor for IgA dimers. This complex is then transported and secreted through the apical surface. Such secretory IgA contributes to antiviral and antibacterial immunity.

Epithelial cells can further internalize antigen and express major histocompatibility complex (MHC) class I and class II molecules. Epithelial cells also express costimulatory molecules, such as CD80 and CD86, which are essential for activating T cells. Thus, epithelial cells may be able to recruit inflammatory cells and sustain a local T-cell response through antigen presentation and costimulation.

2.2 Neutrophils

Neutrophils constitute the majority of the blood leukocytes. They develop in the bone marrow from the same precursors as monocytes and macrophages. After release from the bone marrow, mature neutrophils have a short transit time of only 6 to 8 h in the circulation. Then, neutrophils are stored in tissue sites through so-called *margination*. This mechanism guarantees that upon infection, a large pool of mature neutrophils is rapidly available. Because neutrophils are present in larger numbers than any other inflammatory cell in the circulation, they are the most important cells in initiating the acute tissue response to injury and infection.

In order to enter the tissue, neutrophils must adhere to the endothelium and migrate through the blood vessel, a process that is regulated by cell adhesion molecules (CAMs). These CAMs are present on neutrophils and endothelial cells where they are upregulated at sites of inflammation. Apart from CAMs, neutrophil migration also requires chemotactic mediators. Neutrophil chemotactic mediators include chemokines, most importantly Interleukin-8 (IL-8), bacterial products, complement split products, as well as leukotrienes, and other lipid mediators.

Classically, neutrophils are regarded as phagocytes. Indeed, neutrophils are highly efficient at taking up and digesting bacteria and other pathogens. Since neutrophils express a variety of receptors for the constant part of antibodies (Fc), they are most potent at eliminating antibody-decorated pathogens. Moreover, neutrophils have recently been reported to function as scavenger cells, which deliver antigens, in particular bacteria, into the spleen for induction of an adaptive immune response. However, neutrophils are far more than that and produce antimicrobial agents such as proteases, oxygen radicals, and defensins upon activation, as well as lipid mediators that lead to further cellular recruitment.

A variety of highly active substances are stored preformed in the granules of granulocytes. The serine proteinases neutrophil elastase, cathepsin G, and proteinase-3 are "ready-to-use" stored in the azurophilic granules of mature neutrophils. From there, they can either be transferred to the phagolysosomes or are released upon stimulation. Once released, neutrophil serine proteinases digest a number of extracellular matrix proteins, including elastin, collagen (types I to IV), fibronectin, laminin, and proteoglycans, which may lead to extensive tissue damage.

Release of oxygen radicals from neutrophils or other inflammatory cells apparently plays an important role in causing tissue damage at sites of inflammation.

Among the lipid mediators released by activated neutrophils are leukotrienes (LTB_4) and platelet-activating factor (PAF). Both cause further neutrophil migration and activation, thus amplifying the neutrophil inflammatory response.

Human neutrophil defensins, or human neutrophil peptides (HNP-1 to HNP-4) are small (relative molecular mass 3.54 kD) cationic peptides defined by their antimicrobial activity, but may, however, have further immunoregulatory functions.

2.3 Monocytes and Macrophages

Monocytic cells are released from the bone marrow and reach the blood as monocytes. From there, monocytes leave the circulation to reach their resident site in tissues where they differentiate to so-called *macrophages*. Macrophages are recruited to foci of inflammation.

Monocytes and macrophages play a critical role in host defense, not only in innate immunity but also in regulation of the subsequent adaptive immune response. Monocytes and macrophages produce an array of inflammatory mediators and cytokines, such as IL-1β, TNF-α, GM-CSF, IFN-γ, IL-8, iNOS, RANTES and also lipid mediators such as leukotrienes (LTC_4) and prostaglandins (PGE_2). Their function is mostly proinflammatory; under certain circumstances, monocytes and macrophages can, however, also suppress an immune response by producing IL-10.

Macrophages and monocytes are also important in adaptive immunity, as they express receptors for immunoglobulins G and E (IgG, IgE). Also, T cells and NK cells activate the phagocytic and microbicidal function of macrophages and monocytes by interferon gamma (IFN-γ). IFN-γ enhances the capacity of monocytes and macrophages to present antigen, including the upregulation of MHC class I and II molecules.

2.4 Natural Killer Cells

Natural killer (NK) cells are bone marrow-derived lymphocytes. They comprise the third major lymphocyte population and can be distinguished from other lymphocytes by the absence of B- and T-cell antigen receptors, that is, sIg and T-cell receptor (TCR) respectively. NK cells are capable of lysing certain tumor cells without prior sensitization. Because of this capacity, termed *natural killing*, NK cells were initially thought to serve primarily in defense against tumor cells. Over the years, however, NK cells have been shown to have additional functions. They produce cytokines that regulate the development of acquired, specific immunity, and are involved in the regulation of Th1 versus Th2 development.

NK cells provide early defense against pathogenic organisms during the initial response period against a variety of microorganisms, most importantly against viruses, but also against bacteria and protozoa. NK cells recognize their targets via two general types of receptors. NK cells express inhibitory receptors specific for MHC class I molecules on target cells. This inhibitory receptor prevents the NK cell from activation and killing healthy tissues of the host (''missing-self'' hypothesis). NK cells also express activation receptors that recognize target cell ligands and can trigger perforin-dependent natural killing. Many of these activating receptors, such as Ly49D and Ly49H, are structurally related to the inhibitory receptors. The two

types of receptors, inhibitory and activating, cross-regulate each other and the balanced signal of the two will determine NK cell activation. Stimulation of NK cells through the activation receptors can lead to production of proinflammatory cytokines such as IFNγ, TNF-α, and granulocyte-macrophage colony-stimulating factor (GM-CSF).

NK cells are also activated by cytokine stimulation. Infected or activated dendritic cells (DCs) and macrophages produce cytokines and chemokines such as IFNα/β, IL-12, IL-15, and IL-18 that stimulate NK cells to rapidly produce other cytokines (including IFNγ, TNFα, and GM-CSF) and chemokines (such as ATAC/lymphotactin, Mig, and MIP-1α), resulting in further activation of immune cells.

2.5 Eosinophil Granulocytes

Eosinophils derive from the bone marrow. They are considered to merely pass through the blood circulation to finally reside in the tissues. The ratio of tissue eosinophils to blood is around 100 : 1. The highest density of eosinophils is found in mucosal tissues, a localization pattern that attests to their role in the mucosal immune response.

Eosinophils possess an arsenal of numerous highly basic and cytotoxic granule proteins, as well as enzymes designed to inflict oxidative damage on biologic targets. This weaponry is released on eosinophil activation.

The eosinophil contains an array of cytotoxic granule proteins. Major Basic Protein-2 (MBP-2) is a potent cytotoxin and helminthotoxin. It is also bactericidal, and virucidal. Eosinophil Cationic Protein (ECP) is a highly potent toxin for parasites. Eosinophil-Derived Neurotoxin (EDN) is a potent neurotoxin. Eosinophil peroxidase (EPO) in the presence of H_2O_2 kills microorganisms and tumor cells. A variety of other enzymes are also found in the eosinophil, including Arylsulfatase B and lysozyme.

Apart from a direct cytotoxic and antimicrobial function, the eosinophil also acts as a regulator of the immune response. It contains lipid bodies that serve as a storage site for the production of lipid mediators, predominantly leukotriene C_4 (LTC_4). Eosinophils are also sources of growth factors as well as regulatory or proinflammatory cytokines and chemokines, such as IL-3, IL-4, IL-5, IL-8, IL-16, GM-CSF, TGF-α, TGF-β_1, RANTES, PAF, TNF-α, and MIP-1α.

Eosinophils express various membrane receptors that regulate their trafficking and activation, including adhesion receptors, receptors for the constant portion of immunoglobulins (Fc), as well as receptors for cytokines and lipid mediators.

The major role of eosinophils appears to lie in defense against large organisms, such as multicellular parasites, which are too large for phagocytosis. Eosinophils may respond to carbohydrate ligands on the parasite surface, such as the Lewis^X-related molecules, and cell adhesion molecules that appear similar to selectins. Interestingly, eosinophil action in the lung is under the control of Th cells. Specifically, while eosinophils may be recruited to the lung during an allergic reaction in a Th cell independent fashion, the presence of Th cells in the lung is required for the degranulation of eosinophils. This is yet another example of the complexity of the interaction between cells of the innate and adaptive immune system and how the systems

cross-regulate each other in order to focus the responses on potentially dangerous pathogens.

2.6 Basophils and Mast Cells

Basophils account for less than 0.2% of all leukocytes. The granules in basophiles contain an abundance of mediators, the most important ones being histamine, heparin, and leukotrienes. Upon cross-linking of the high affinity IgE receptor (FcεRI), basophils degranulate and can cause severe systemic symptoms of allergy including anaphylactic shock. However, basophils also play a role in the immune response against parasites by augmenting inflammation.

Mature human mast cells are rounded or spider-like cells of 9 to 12-μm diameter, which contain a striking amount of granules that, when stained, behave metachromatically. Mast cells are present in nearly every organ tissue, but high numbers of mast cells are found in potential entry ports for pathogens, such as the skin, and the mucosal surfaces of the eye, the gastrointestinal tract, and the respiratory tract.

Two types of mast cells, differing in their granule contents and in their patterns of mast cell–specific proteases can be distinguished. Human mast cell nomenclature is based on neutral protease composition. Whereas MC_T cells express tryptase alone, MC_{TC} cells produce tryptase, chymase, mast cell carboxypeptidase, and cathepsin G. It can be said that MC_T cells predominate in the normal lung, particularly in the alveoli, and in the small intestine mucosa. MC_{TC} cells are the predominant type found in normal skin, blood vessels, gastrointestinal submucosa, and synovium.

Mast cells are well-known effectors of allergic disease, as they express the high-affinity receptor for immunoglobulin E (IgE). If cross-linked by an allergen, the mast cell immediately degranulates an array of mediators, one of the most rapidly acting being histamine.

While it is clear that mast cells are responsible for making patients suffer from their allergies against harmless substances such as pollen or food, it is extremely difficult to find a positive role for mast cells. It is not until recently that a new and positive role of the mast cell in protective immunity against parasites and bacteria is emerging. Animal studies with parasites showed a correlation of mast cell numbers with parasite expulsion that can be readily attributed to mast cell mediators such as histamine, proteases, and cytokines. The immune response initiated by mast cell-released mediators also involves eosinophils, and Th2 cells. More recent animal studies using mast cell-deficient mice, however, suggest that it is not the mast cells themselves, but the Th2 type immune response that is the key to antiparasite immunity. Interestingly, mast cell–deficient mice also appear more susceptible to bacterial infections, suggesting a role for the mast cell in innate immunity. Mast cells produce tumor necrosis factor alpha (TNF-α), a strong chemo attractant for neutrophils, offering an explanation for the antibacterial activity of mast cells.

2.7 Tissue Cells

Although usually not regarded as part of the immune system, tissue cells possess antimicrobial defense mechanisms that are the key to host survival. Here, interferons must be mentioned. Host

cells, upon viral infection, start producing IFN-α and IFN-β. These interferons induce resistance to viral replication in neighboring uninfected cells. Interferons also upregulate class I MHC molecules, thus making virally infected cells more susceptible to CD8+ T cell-mediated killing, so that the virus is more rapidly eliminated. Finally, interferons activate NK cells, which then kill virus-infected cells.

2.8 Platelets

In addition to their role in blood clotting, platelets are involved in inflammation. Each day an adult human produces 10^{11} platelets that are split off from megakaryocytic cells in the bone marrow. Injury to endothelial cells induces platelet adherence and aggregation, leading to platelet activation and degranulation. Platelet granula contain serotonin so that capillary permeability is increased, as well as fibrinogen that activates complement and attracts leukocytes. Interestingly, platelets express CD40 ligand, a molecule critical for activation of B cells, macrophages, and dendritic cells, suggesting a possible role of platelets in the regulation of innate and adaptive immune responses.

3 Adaptive Immunity

3.1 Cell Mediated versus Antibody Mediated Immunity

The adaptive immune system is based on two pillars, T cells and B cells. T cells are responsible for cell-mediated immunity while B cells produce antibodies, the basis of humoral immunity. Although the two arms are intimately linked with T cells regulating B-cell responses (in particular, through T help) and vice versa, B cells influencing T-cell responses, the basic functions and tasks of B and T cells are quite distinct (Table 1). Antibodies are important for protection against a variety of pathogens, in particular, many viruses and bacteria. Although antibodies are essential for protection against most secondary infections, they also play a major role during many primary infections and absence of antibodies results in increased susceptibility to many viruses, such as rotaviruses and polio virus. T cells also play a major role during many primary viral infections. Cytotoxic T cells (CTLs) are the key to the elimination of many non- or poorly cytopathic viruses, such as hepatitis B and C. CTLs also play a major role in keeping persistent infections under control. Th cells, on the other hand, play a

Tab. 1 Effector mechanisms for the elimination of pathogens and tumors.

Type of disease	*Dominant effector mechanism*	*Example*
Primary infection	Antibody, T cell, innate	Rota virus (Ab), hepatitis B (T cell), Streptococcus (innate)
Secondary infection	Antibody	Rota virus, polio, hepatitis B
Chronic infection	T cells	HIV, HCV
Solid tumor	T cells	Melanoma
Single-cell tumors	Antibody (monoclonal)	B-cell lymphoma

more complex role. Th cells are critical for the B-cell response, which remains abortive in the absence of Th. This not only helps in eliminating pathogens in acute infections but may also be essential in preventing the loss of CTL-memory during chronic infection. In addition, Th cells act more directly on CTLs by facilitating the induction of proliferation competent memory CTLs (see Sect. 4).

The situation for cancer appears different since antibodies play a minor role in control of tumor growth. As antibodies do not efficiently penetrate tissues, their ability to kill tumor cells remains limited. The monoclonal antibody herceptin, which is directed against breast cancer cells, and has proven successful in patients, is probably a rather exceptional case. Another exception is single-cell tumors, such as lymphomas, as they are more accessible to antibodies than solid tumors. This is well documented by the clinical efficacy of a variety of monoclonal antibodies, such as anti-CD20 antibodies (Rituximab), which eliminate B cells in an Fc receptor dependent fashion. Yet, the importance of naturally induced lymphoma specific antibodies remains to be documented. In addition to a possible direct antitumor activity, antibodies may facilitate presentation of tumor-associated antigens by DCs and other antigen-presenting cells (APCs), leading to heightened T-cell responses. It is this T-cell response that appears most apt to reduce tumor burden, since T cells have the ability to enter solid tumor tissues. The key antitumor effector cells are CTLs, which can induce direct lysis of tumor cells. In addition, Th cells locally produce cytokines in the tumor mass, creating a proinflammatory milieu, facilitating elimination of tumor cells by recruitment and activation CTLs and nonspecific cells such as macrophages or eosinophils.

3.2 Induction of Cell-mediated Immunity: T-Cell Activation

T-cell activation is a carefully orchestrated process optimally driven by professional antigen-presenting cells, such as DCs, which express high levels of accessory and costimulatory molecules. Several additional parameters regulate T-cell activation, including the kinetics and duration of antigen presentation as well as the activation status of DCs. The following sections will discuss these parameters in more detail.

3.2.1 Costimulation

Antigen-specific stimulation of T cells through the TCR alone is not usually sufficient for full T-cell activation. On the contrary, T-cell responses remain abortive in the absence of costimulation by accessory molecules, resulting in T-cell tolerance or anergy, rather than activation. T-cell anergy characterizes nonfunctional T cells that fail to proliferate upon exposure to antigen and therefore are not able to mediate an effective immune response. The most prominent costimulatory molecule on T cells is CD28, a receptor triggered by the B-7 family members CD80 and CD86 expressed on professional antigen-presenting cells, in particular dendritic cells. Indeed, engaging CD28 together with the TCR results in a reduced number of TCRs that need to be triggered for T-cell activation and enhances TCR-mediated signal transduction. Moreover, CD28 delivers an additional, poorly understood, signal into T cells, preventing induction of T-cell anergy. Yet, many viruses and other pathogens are able to induce strong and protective T-cell responses in the absence of CD28.

This may not reflect truly costimulation-independent T-cell activation but may rather be explained by the usage of alternative costimulatory molecules, such as ICOS. Indeed, ICOS has been shown to be responsible for CD28-independent activation of T-cell responses upon infection with vesicular stomatitis virus (VSV). OX40 and 41BB, members of the TNF-receptor family are additional molecules involved in T-cell activation, presumably through costimulation-like processes. In addition to such costimulatory molecules, there are accessory molecules that facilitate T-cell activation through adhesion of T cells to DCs. LFA-1 and CD2 may be good examples, which, in addition to a postulated signaling function, enhance TCR triggering by facilitating T-cell APC interactions. Interestingly, effector function of T cells occurs rather independent of costimulation, ensuring proper activity of effector cells in the periphery, where costimulatory molecules are scarce. Thus, T cells go through an initial phase of costimulation-dependent activation followed by costimulation-independent effector function.

3.2.2 Duration of Antigen Presentation

It has been shown that a single brief exposure to antigen allows T cells to undergo a full program of activation, proliferation, and effector cell differentiation. Yet, extended antigen presentation is essential for the generation of protective T-cell responses and also for the induction of CD28-independent T-cell activation as observed after viral infection. In fact, one of the key differences between pathogens and recombinant, nonreplicating antigens is the duration of antigen presentation. While antigens derived from viruses, bacteria, and other pathogens are presented for at least as long as the pathogens replicate, recombinant antigens are presented for much shorter time spans; peptides are, for example, usually only presented for a few hours. This, in part, explains the poor immune response usually obtained with recombinant antigens and also indicates why adjuvants enhance immune responses, since adjuvants, such as aluminum hydroxide or Freund's adjuvants, form antigen depots, leading to a prolonged antigen release. In addition to forming antigen depots, modern vaccines also contain substances that activate the innate immune system, as discussed in the next Section (3.2.3).

3.2.3 DC Maturation

DCs are cells that are potent in taking up and processing antigens. In addition, DCs are the most efficient cell type for the induction of protective T-cell responses. Yet, DCs are surprisingly efficient at inducing T-cell tolerance if they are not activated by additional stimuli. Thus, only activated DCs induce protective T-cell responses. When DCs are resting or immature, they induce T-cell tolerance rather than activation. The inability of immature DCs to trigger T-cell activation may in part be explained by reduced levels of costimulatory molecules; indeed, during DC maturation, costimulatory molecules such as CD80/CD86 are upregulated, leading to increased costimulation. Nevertheless, resting DCs do express significant levels of CD86 and it seems unlikely that absence of costimulation alone is responsible for inducing tolerance rather then immunity. Additional factors, such as lack of cytokine secretion by resting DCs or their short life span that leads to a short duration of antigen presentation are probably also contributing to the inability of immature DCs to induce protective T-cell responses.

Two types of stimuli cause activation and maturation of dendritic cells: cytokines and other proteins produced by T cells, macrophages, DCs, and/or other cell types and patterns associated with pathogens (PAMPs) recognized by specific receptors on DCs. Good examples of such patterns are LPS or other cell wall components of bacteria as well as the double-stranded RNA of viruses. The best described receptors for such patterns (PRRs) are toll-like receptors 1 to 11. Interestingly, different toll-like receptors appear to stimulate the immune system in slightly different ways. DCs, for example, express, amongst others, toll-like receptors 2, 4, and 9. For the induction of CTL responses, toll-like receptor 9 appears to be much more important than the other two receptors, while for enhancing antibody responses against allergens, toll-like receptor 2 appears to be most efficient.

Simulation of DCs and other APCs by toll-like receptors is most important for the early immune response, directing the type of immune response that is induced. Specifically, most toll-like receptors direct the immune response toward a Th1 type and CTLs, namely the type of T- cell response that is critical to rid viruses and bacteria. Many multicellular parasites are eliminated by the immune system mounting a Th2 response. In line with this, most parasites are not known to engage toll-like receptors.

Cytokines produced by activated T cells are playing a major role in reinforcement, sustaining and amplifying the type of response that has been initiated by the toll-like receptors. In this way, the innate and adaptive immune system interacts intimately early in the response and the adaptive immune system only becomes independent of the innate immune system once the T-cell response has reached a certain threshold required for a sustained response.

3.3 Extravasation

Once activated, T cells need to leave lymphoid organs and emigrate into tissues. A key step in this process is the trans-endothelial migration of lymphocytes. Three steps dominate this migration (Fig. 2). In the first step, lymphocytes attach loosely to the endothelial layer by way of interaction between selectins expressed on T cells and adressins expressed on endothelial cells. In this phase, lymphocytes roll over the endothelium of blood vessels. In the second step, lymphocytes become activated by local chemokines bound to the extracellular matrix on endothelial cells. This leads to cellular activation and increased avidity of LFA-1 to its ligand I-CAM, resulting in arrest of the cells at the local site and trans-endothelial migration (diapedesis). At the site of injury or infection, CTLs may then directly lyse target cells or inhibit pathogen replication by secretion of cytokines. Similarly, Th cells, if activated by local MHC class II expressing APCs, also secrete cytokines, chemokines, and other mediators, leading to increased permeability of the endothelial cell layer and further cell-recruitment. In addition, cytokines and chemokines may trigger local activation of APCs and effector cells, such as granulocytes and CTLs.

3.4 Delayed Type Hypersensitivity Reaction

A well-known example of T-cell activation, extravasation, and local release of

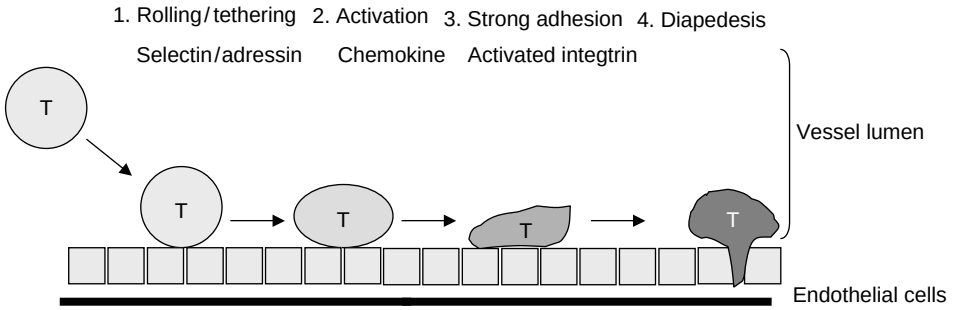

Fig. 2 Multistep model of lymphocyte activation for extravasation from blood into the tissue.

immunological mediators is the delayed type hypersensitivity reaction. Upon injection of proteins into the skin, antigen drains into T-cell areas of the lymph node, leading to MHC class II-associated antigen presentation. At the same time, Langerhans cells in the skin take up the antigen, process it, and transport it into the lymph node, further increasing antigen presentation. This leads to local activation of T cells. If the organism is exposed for the first time to the antigen, the kinetics of the T-cell response induced is slow and by the time activated T cells leave the lymph node, the antigen in the skin has been eliminated and no local reaction becomes visible. In contrast, in the presence of memory T cells, activated T cells leave the lymph node quickly while the injected antigen is still present in the skin. This results in a local inflammatory response, which is readily visible as a local rash and swelling, due to increased permeability of the vessels. Since no inflammatory response occurs in the absence of preexisting T-cell memory, this delayed type hypersensitivity (DTH) reaction is often used to measure T-cell memory and to probe for persistent infections, as is done with the tuberculin test for diagnosis of latent infection with tuberculosis.

3.5 Lymphoid Organ-like Structures in Chronic Inflammation

Induction and maintenance of T-cell responses usually requires the transport of antigen into lymphoid organs, where T-cell activation occurs. This is true both for the induction of protective T-cell responses against pathogens as well as for the generation of T-cell responses directed against self, potentially causing autoimmunity. For these latter autoimmune responses, however, the reverse may also occur, namely, that lymphocytes are attracted to self-antigens in the periphery, causing sustained T-cell responses and damage in the afflicted organs. An important molecule for this process is the chemokine BLC (CXCL13, BCA-1), which attracts B cells. Under normal conditions, BLC orchestrates the generation and maintenance of B-cell follicles in lymph nodes and spleen, by attracting CXCR5 receptor expressing B cells. B cells that have been stimulated by BLC express lymphotoxin β(LT-β), which is essential for the generation of follicular dendritic cells and formation of B-cell follicles. Around these B-cell follicles, T-cell areas are formed leading to the generation of the typical structure of lymphoid organs. Expression of BLC in the periphery

is sufficient to drive the generation of such structures within peripheral organs, causing chronic stimulation of self-specific T cells within these ectopic structures. Indeed, in a mouse model of lupus erythematosus, a severe autoimmune disease in humans, BLC expression was increased up to 10 000 fold in organs such as thymus and kidney, that are typically infiltrated by lymphocytes. It is likely that lymphoid organ-like structures observed in patients with Hashimoto's thyroiditis, type I diabetes, or rheumatoid arthritis may have a similar origin and are caused by local expression of BLC or lymphotoxin-β.

4 Mechanisms of T-Cell Mediated Immunity

4.1 Th Cells (CD4+)

Th cells have a limited ability to directly interfere with pathogen replication. Through secretion of cytokines, in particular IFNγ, Th cells can induce an antiviral state in neighboring cells, halting replication of viruses and intracellular bacteria. In addition, members of the TNF superfamily (TNF and Fas-ligand) may have the ability to cause apoptosis in infected target cells, also reducing pathogen replication. The most important function of Th cells is, however, the orchestration of the various antimicrobial effector cells (Fig. 3). Through the secretion of chemokines and cytokines, Th cells are able to recruit innate effector cells, such as macrophages, granulocytes, or eosinophils and activate DCs, enhancing T-cell priming. In addition, Th cells are able to cause isotype-switching in B cells and the generation of protective memory antibody levels. Moreover, Th cells also act directly on CD8$^+$ T cells facilitating the induction of CTL responses against weak antigens and enabling development of proliferation competent memory CTLs. Thus, although Th

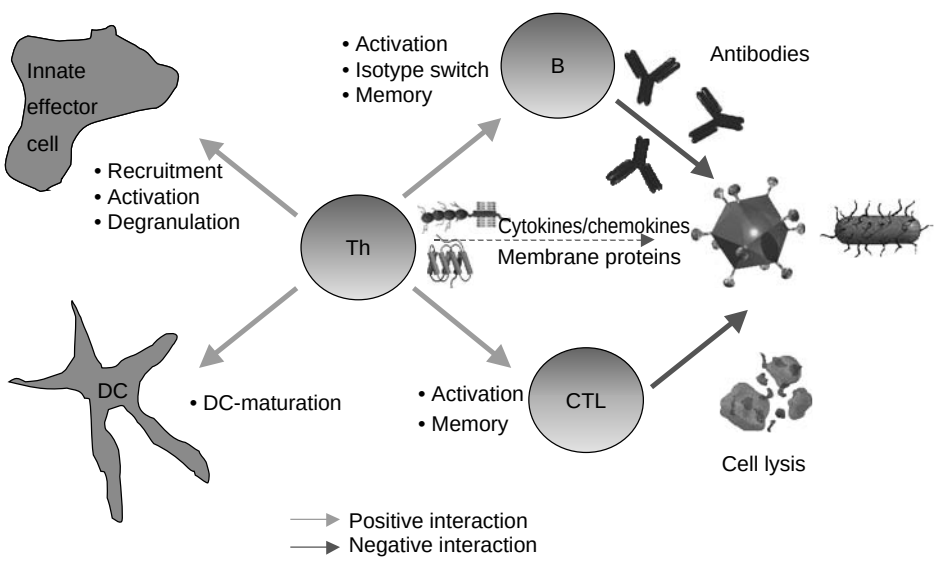

Fig. 3 Cross-talk between the innate and adaptive immune system and the critical role of Th cells in orchestrating the immune response.

cells have little immediate effector function, they are key cells for the induction and maintenance of immunity by activating an army of innate effector cells and facilitating the generation of B- and T-cell memory.

4.2 Cytotoxic T Cells (CD8+)

In contrast to Th cells, CTLs are key effector cells of the adaptive immune system. Similar to Th cells, CTLs are secreting a variety of chemokines, cytokines, and IFNγ enhancing cellular recruitment and activation as well as causing a local antiviral state. In addition, CTLs harbor lytic granules, which upon stimulation are released in a directed fashion at the immunological synapse onto the recognized target cells. The lytic granules contain a variety of proteins that can cause death of the recognized target cells. The most important proteins are perforin, which acts by "punching" holes into target cells as well as serine proteases belonging to the family of granzymes. Although it was originally thought that perforin mediates cell death by destruction of membrane integrity, it has recently become clear that the main function of perforin is to enable granzymes to enter target cells. This causes death of the target cells, because granzymes activate caspases, leading to apoptosis.

Granzymes are inactive at the low pH while they are stored in the granules and they only become activated once they are released from the granules, that is, in the target cells. The importance of the lytic perforin/granzyme pathway versus the secretion of cytokines by CTLs varies with the pathogen. For lytic viruses, which rapidly kill infected target cells, the perforin/granzymes pathway is of limited importance and cytokines and TNF family members are more important for protection. For nonlytic viruses, which do not kill the infected target cells, the perforin/granzymes pathway is more pivotal and cytokines as well as other soluble mediators play a minor role.

5 Silencing of the Immune Response: What Goes Up Must Come Down

So far, we have mainly been concerned with the buildup of protective T-cell responses. However, the silencing of ongoing immune responses to limit damage in the host caused by the immune system (immunopathology) is equally important. What mechanism(s) are available to switch off T-cell responses? For responses against acute infections, the most efficient mechanism is the elimination of antigen. Once the pathogen has been cleared from the host, the majority of T cells undergo apoptosis and, in the absence of further stimulation, the T cells become quiescent memory T cells (Fig. 4). As an additional mechanism, antigen-specific attenuation of T-cell responses by membrane-bound receptors has been postulated for CTLA-4, PD-1, a CTLA-4 homolog and further homologs of this family of receptors (e.g. BTLA). Indeed, mice deficient in the respective genes have a striking phenotype and rapidly develop symptoms of autoimmune disease. However, the mechanism of this attenuation is still a matter of debate. In fact, rather than downregulating T cells in an antigen-specific manner, receptors such as CTLA-4 may facilitate the induction of regulatory T helper cells that secrete anti-inflammatory cytokines.

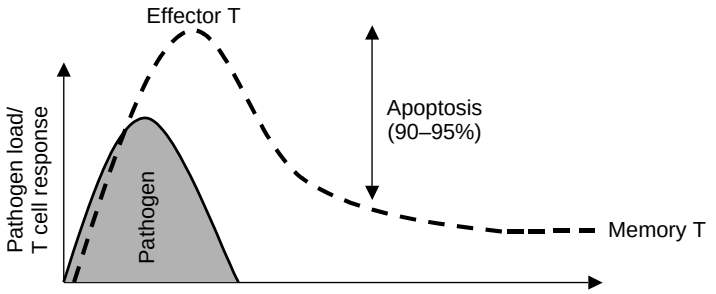

Fig. 4 Regulation of T-cell responses and establishment of memory.

TGF-β and IL-10 are the most potent immunosuppressive cytokines. Production of these anti-inflammatory cytokines has been shown to be of great importance during an infection with parasites, which may otherwise induce overshooting T cells responses resulting in the death of the host. The ambivalent function of immunosuppressive or anti-inflammatory cytokines is well illustrated by IL-10. On the one hand, acute infection of IL-10 deficient mice with Toxoplasma gondii or the malaria parasite *Plasmodium* results in a fulminate inflammatory response, causing death of the infected mice due to immunopathology usually kept in check by IL-10. On the other hand, due to increased inflammatory T-cell responses, IL-10-deficient mice are better protected from primary and secondary systemic infection with *Listeria monocytogenes* and show improved clearance of pathogens such as *Mycobacteria* or *Leishmania major* that usually persist in their hosts.

In case the pathogen cannot be cleared from the host, as is the case for the parasite *L. major*, the member of the TNF-receptor family Fas has been shown to be important by causing apoptosis of highly stimulated T cells, preventing their accumulation. Indeed, Fas-deficient mice amass antigen-specific T cells upon infection with *L. major*, resulting in grossly enlarged draining lymph nodes.

In conclusion, T-cell responses induced by acute infection that can be resolved by the host are usually downregulated by clearance of the antigen. Under condition of excessive antigen load or undue stimulation of innate immunity or when the pathogen cannot be eliminated, additional mechanisms may kick in, such as inhibitory receptors and, in particular, the production of anti-inflammatory cytokines.

6 T-Cell Memory

Upon infection or vaccination, specific T cells proliferate and expand as a population. Peak frequencies are usually reached within one or two weeks. Once the pathogen has been cleared, T-cell frequencies start to drop and more than 90% of the T cells undergo apoptosis. The remaining cells differentiate into memory cells and are maintained at relatively constant frequencies over extended time periods (Fig. 4). If the pathogen is reencountered, these T cells respond with accelerated kinetics, leading to an enhanced secondary

T-cell response. Hence the term "T-cell memory"; the immune systems remember that it has previously responded to the pathogen.

This enhanced T-cell response is not only due to increased numbers of memory T-cell precursor. Memory T cells also show an altered physiology when compared to naïve T cells. Most importantly, memory T cells can secrete cytokines within hours after stimulation while naïve T cells require days for initiation of cytokine secretion. Moreover, memory T cells can be activated in the absence of costimulation, apparently due to a rearranged signal transduction machinery. Thus, if compared to naïve T cells, memory T cells are also more potent at the single-cell level.

Various memory T-cell subsets have been described. Particularly prominent is the concept of central versus effector memory T cells. Originally, it was thought that central memory T cells reside in lymph nodes and the spleen and rapidly respond with proliferation to antigen reexposure. Effector memory T cells, on the other hand, were suggested to reside in peripheral tissues and exhibit immediate effector function. Hence, effector memory T cells were thought to be mainly responsible for immediate protection against reinfection. However, this concept has meanwhile been challenged in many ways and it was found that central memory T cells are often found in peripheral tissues and exhibit similar immediate effector function as effector memory T cells. In addition, it was even reported that central memory T cells are more potent at mediating protection against reinfection than effector memory T cells. Thus, the physiological significance of the two types of memory T cells remains to be established.

The importance of persisting antigen for the maintenance of T-cell memory and for the balance between central and effector memory T cells has been a matter of extensive debate. Nevertheless, it is now emerging that memory T cells do not need persisting antigen for their survival. Yet, presence of persisting antigen keeps the T cells in a more activated state, facilitating maintenance of immediate effector function. Thus, optimal long-term protective T-cell memory is induced by viruses that persist at a low level in the host, such as lymphocytic choriomeningitis virus or cytomegalovirus. In fact, it is the chronically stimulated T cells that keep the virus in check and prevent widespread replication. Indeed, containment of persisting viruses and other pathogens is probably one of the most important tasks of T-cell memory. The importance of chronic antigen exposure for the maintenance of T-cell effector function is also illustrated by the examples described above of lymphoid organ-like structures in peripheral organs of patients with autoimmune diseases. It is the continuous antigenic exposure in an immunostimulatory environment that keeps T cells activated and capable of mediating tissue destruction.

In conclusion, upon elimination of the bulk of the antigen, most of the effector T cells undergo apoptosis and few of the cells differentiate into memory T cells. These memory cells are maintained at increased frequencies and are able to respond to antigenic reexposure with enhanced kinetics. One key function of memory T cells is to keep persistent infections under control and prevent widespread replication of the pathogens.

See also Antigen Presenting Cells (APCs); Immunologic Memory; Immunology; Innate Immunity.

Bibliography

Books and Reviews

Ahmed, R., Gray, D. (1996) Immunological memory and protective immunity: understanding their relation, *Science* **272**(5258), 54–60.

Akira, S. (2003) Mammalian Toll-like receptors, *Curr. Opin. Immunol.* **15**(1), 5–11.

Bachmann, M.F., Kopf, M. (2002) Balancing protective immunity and immunopathology, *Curr. Opin. Immunol.* **14**(4), 413–419.

Beutler, B. (2002) Toll-like receptors: how they work and what they do, *Curr. Opin. Hematol.* **9**(1), 2–10.

Biron, C.A. (2001) Interferons alpha and beta as immune regulators–a new look, *Immunity* **14**(6), 661–664.

Campbell, D.J., Kim, C.H., et al. (2003) Chemokines in the systemic organization of immunity, *Immunol. Rev.* **195**, 58–71.

Coyle, A.J., Gutierrez-Ramos, J.C. (2001) The expanding B7 superfamily: increasing complexity in costimulatory signals regulating T cell function, *Nat. Immunol.* **2**(3), 203–209.

Delves, P.J., Roitt, I.M. (2000) The immune system. First of two parts. The immune system. Second of two parts, *N. Engl. J. Med.* **343**(1), 37–49; *N. Engl. J. Med.* **343**(2), 108–117.

Dutton, R.W. (1967) In vitro studies of immunological responses of lymphoid cells, *Adv. Immunol.* **6**, 253–336.

Freedman, J.E. (2003) CD40-CD40L and platelet function: beyond hemostasis, *Circ. Res.* **92**(9), 944–946.

Galli, S.J., Maurer, M., Lantz, C.S. (1999) Mast cells as sentinels of innate immunity, *Curr. Opin. Immunol.* **11**(1), 53–59.

Germain, R.N., Stefanova, I. (1999) The dynamics of T cell receptor signaling: complex orchestration and the key roles of tempo and cooperation, *Annu. Rev. Immunol.* **17**, 467–522.

Gleich, G.J., Adolphson, C.R. (1986) The eosinophilic leukocyte: structure and function, *Adv. Immunol.* **39**, 177–253.

Hoebe, K., Janssen, E., Beutler, B. (2004) The interface between innate and adaptive immunity, *Nat. Immunol.* **5**(10), 971–974.

Hoffmann, J.A., Kafatos, F.C., et al. (1999) Phylogenetic perspectives in innate immunity, *Science* **284**(5418), 1313–1318.

Janeway, C.A. Jr., Goodnow, C.C., et al. (1996). Danger – pathogen on the premises! Immunological tolerance, *Curr. Biol.* **6**(5), 519–522.

Kitamura, Y. (1989) Heterogeneity of mast cells and phenotypic change between subpopulations, *Annu. Rev. Immunol.* **7**, 59–76.

Klion, A.D., Nutman, T.B. (2004) The role of eosinophils in host defense against helminth parasites, *J. Allergy Clin. Immunol.* **113**(1), 30–37.

Lee, W.L., Harrison, R.E., Grinstein, S. (2003) Phagocytosis by neutrophils, *Microbes Infect.* **5**(14), 1299–1306.

Lieberman, J. (2003) The ABCs of granule-mediated cytotoxicity: new weapons in the arsenal, *Nat. Rev. Immunol.* **3**(5), 361–370.

Mellman, I., Steinman, R.M. (2001) Dendritic cells: specialized and regulated antigen processing machines, *Cell* **106**(3), 255–258.

Parker, D.C. (1993) T cell-dependent B cell activation, *Annu. Rev. Immunol.* **11**, 331–360.

Raulet, D.H. (2004) Interplay of natural killer cells and their receptors with the adaptive immune response, *Nat. Immunol.* **5**(10), 996–1002.

Sallusto, F., Lanzavecchia, A. (2000) Understanding dendritic cell and T-lymphocyte traffic through the analysis of chemokine receptor expression, *Immunol. Rev.* **177**, 134–140.

Scapini, P., Lapinet-Vera, J.A., Gasperini, S., Calzetti, F., Bazzoni, F., Cassatella, M.A. (2000) The neutrophil as a cellular source of chemokines, *Immunol. Rev.* **177**, 195–203.

Schutte, B.C., McCray, P.B. Jr. (2002) [beta]-defensins in lung host defense, *Annu. Rev. Physiol.* **64**, 709–748.

Sprent, J., Surh, C.D. (2002) T cell memory, *Annu. Rev. Immunol.* **20**, 551–579.

Takeda, K., Kaisho, T., Akira, S. (2003) Toll-like receptors, *Annu. Rev. Immunol.* **21**, 335–376.

Toes, R.E., Schoenberger, S.P., et al. (1998) CD40-CD40Ligand interactions and their role in cytotoxic T lymphocyte priming and anti tumor immunity, *Semin. Immunol.* **10**(6), 443–448.

Unanue, E.R., Askonas, B.A. (1968) The immune response of mice to antigen in macrophages, *Immunology* **15**(2), 287–296.

Zinkernagel, R.M., Bachmann, M.F., et al. (1996) On immunological memory, *Annu. Rev. Immunol.* **14**, 333–367.

Zinkernagel, R.M., Hengartner, H. (2001) Regulation of the immune response by antigen, *Science* **293**(5528), 251–253.

Primary Literature

Bachmann, M.F., Kohler, G., et al. (1999) Cutting edge: lymphoproliferative disease in the absence of CTLA-4 is not T cell autonomous, *J. Immunol.* **163**(3), 1128–1131.

Bachmann, M.F., McKall-Faienza, K., et al. (1997) Distinct roles for LFA-1 and CD28 during activation of naive T cells: adhesion versus costimulation, *Immunity* **7**(4), 549–557.

Bachmann, M.F., Wong, B.R., et al. (1999) TRANCE, a tumor necrosis factor family member critical for CD40 ligand-independent T helper cell activation, *J. Exp. Med.* **189**(7), 1025–1031.

Bachmann, M.F., Zinkernagel, R.M., et al. (1998) Immune responses in the absence of costimulation: viruses know the trick, *J. Immunol.* **161**(11), 5791–5794.

Bjorkander, J., Bake, B., Oxelius, V.A., Hanson, L.A. (1985) Impaired lung function in patients with IgA deficiency and low levels of IgG2 or IgG3, *N. Engl. J. Med.* **313**(12), 720–724.

DeBenedette, M.A., Shahinian, A., et al. (1997) Costimulation of CD28-T lymphocytes by 4-1BB ligand, *J. Immunol.* **158**(2), 551–559.

Ehrlich, P. (1878). *Beiträge zur Theorie und Praxis der Histologischen Färbung*, Thesis in Leipzig.

Fleming, A. (1922) On a remarkable bacteriolytic element found in tissues and secretions, *Proc. Roy. Soc. Ser. B* **93**, 306–317.

Greenwald, R.J., Boussiotis, V.A., et al. (2001) CTLA-4 regulates induction of anergy in vivo, *Immunity* **14**(2), 145–155.

Huang, S., Hendriks, W., Althage, A., Hemmi, S., Bluethmann, H., Kamijo, R., Vilcek, J., Zinkernagel, R.M., Aguet, M. (1993) Immune response in mice that lack the interferon-gamma receptor, *Science* **259**(5102), 1742–1745.

Ishizaka, T., De Bernardo, R., Tomioka, H., Lichtenstein, L.M., Ishizaka, K. (1972) Identification of basophil granulocytes as a site of allergic histamine release, *J. Immunol.* **108**(4), 1000–1008.

Kazura, J.W., Grove, D.I. (1978) Stage-specific antibody-dependent eosinophil-mediated destruction of Trichinella spiralis, *Nature* **274**(5671), 588–589.

Kopf, M., Coyle, A.J., et al. (2000) Inducible costimulator protein (ICOS) controls T helper cell subset polarization after virus and parasite infection, *J. Exp. Med.* **192**(1), 53–61.

Kopf, M., Ruedl, C., et al. (1999) OX40-deficient mice are defective in Th cell proliferation but are competent in generating B cell and CTL Responses after virus infection, *Immunity* **11**(6), 699–708.

Kundig, T.M., Shahinian, A., et al. (1996) Duration of TCR stimulation determines costimulatory requirement of T cells, *Immunity* **5**(1), 41–52.

Oxenius, A., Campbell, K.A., et al. (1996) CD40-CD40 ligand interactions are critical in T-B cooperation but not for other anti-viral CD4+ T cell functions, *J. Exp. Med.* **183**(5), 2209–2218.

Ruedl, C., Kopf, M., et al. (1999) CD8(+) T cells mediate CD40-independent maturation of dendritic cells in vivo, *J. Exp. Med.* **189**(12), 1875–1884.

Sallusto, F., Lenig, D., et al. (1999) Two subsets of memory T lymphocytes with distinct homing potentials and effector functions, *Nature* **401**(6754), 708–712.

Schoenberger, S.P., Toes, R.E., et al. (1998) T-cell help for cytotoxic T lymphocytes is mediated by CD40-CD40L interactions, *Nature* **393**(6684), 480–483.

Thomsen, A.R., Nansen, A., et al. (1998) CD40 ligand is pivotal to efficient control of virus replication in mice infected with lymphocytic choriomeningitis virus, *J. Immunol.* **161**(9), 4583–4590.

Unsoeld, H., Krautwald, S., et al. (2002) Cutting edge: CCR7+ and CCR7− memory T cells do not differ in immediate effector cell function, *J. Immunol.* **169**(2), 638–641.

Viola, A., Lanzavecchia, A. (1996) T cell activation determined by T cell receptor number and tunable thresholds, *Science* **273**(5271), 104–106.

Immunoassays

James P. Gosling
Department of Biochemistry, National University of Ireland, Galway, Ireland

1 Antibodies for Immunoassays 363
1.1 Immunoglobulin G Structure 363
1.2 Monoclonal, Polyclonal, and Anti-hapten Antibodies 365

2 Antibody–antigen Binding 365
2.1 Structural Complementarity 365
2.2 Molecular Mobility Influences Binding 366
2.3 Epitopes 366

3 Labeling Substances 367

4 Classification of Immunoassays 367
4.1 Nomenclature 367
4.2 The Variety of Immunoassay Formats 369
4.3 Representing Assay Complexes in Text 369
4.4 The Four Main Groups of Immunoassays 370
4.4.1 Label-free Assays 371
4.4.2 Reagent-excess Assays 371
4.4.3 Reagent-limited Assays 372
4.4.4 Ambient Analyte Assays 372

5 Established Immunoassay Systems 373
5.1 Automated Systems 373
5.2 General Laboratory Assays 373
5.3 Simple to use Test Devices 374

6 Immunoassays of the Future 375
6.1 Multiplexed Immunoassays 375
6.2 New Assay Principles and Systems 375

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

7 The Limitations of Immunoassays 376

Bibliography 378
Books and Reviews 378
Primary Literature 378

Keywords

Automated
Automated immunoassays require complex, computer-controlled instrumentation that takes care of all repetitive steps such as sample handling, pipetting, washing, end point determination and the calculation of results. New larger machines integrate the functions of immunoassay analyzers and clinical chemistry analyzers into one system.

Binding Affinity
Binding affinity is the probability of association of a ligand (antigen) with a binding molecule (antibody). High affinity depends positively on the rate of association of the two molecules and negatively on the rate of dissociation of the complex formed.

Classification
In this context, ''classification'' refers to any system that may be used to explain how the very large numbers of different immunoassays are related to each other.

Epitope
An epitope is that portion of an antigen with which the binding site of an antibody (paratope) interacts specifically.

Hapten
A hapten is a low molecular weight antigen that is not inherently immunogenic.

Hybridoma
Monoclonal hybridoma cells are ''immortal'' cells that (in the present context) are capable of synthesizing antibodies with desirable properties. They are generated by fusing lymphocytes and myeloma cells, followed by dilution, culture, and selection steps to ensure enduring viability, homogeneity, and ability to secrete desirable monoclonal antibodies.

Immunoglobulin Structure
All immunoglobulin molecules are glycoproteins with similar characteristic complex structures that are rugged and flexible. They consist of multiple, paired domains that are all derived from ''the immunoglobulin fold,'' a protein structural motif also found in many other proteins.

Interference
By interference is meant the phenomenon whereby an incorrect result is obtained in an immunoassay because of the presence, in the sample being tested, of a substance or substances (other than the analyte) capable of interacting with assay reagents so as to influence their normal roles.

Label
The term ''label'' is usually applied to the labeled antigen, labeled hapten, or labeled antibody employed as a reagent in immunoassay.

Microarray
In the context of immunoassay methodology, a microarray is usually an array or matrix of microdots each coated with a different specific antibody or antigen.

Monoclonal Antibody
A homogeneous population of antibody molecules, each with identical binding affinity and specificity, synthesized by a monoclonal cell line.

Multiplexed
Multiplexed immunoassays are capable of detecting or measuring two or more analytes following the execution of a single analytical procedure that usually has no additional steps as compared to an ordinary assay. Multiplex is a synonym of multiple.

Paratope
A paratope consists of complementarity determining regions (CDR) of the variable domains of an immunoglobulin and is that portion of an antibody that constitutes an antigen binding site.

Polyclonal Antibody
A heterogeneous population of antibody molecules raised by immunizing a suitable animal with a specific antigen.

Rapid Test Device
Rapid test devices (RTD) are small constructs usually consisting of one or more components, including moulded plastic structures, membranes, adsorbent strips, absorbent pads, and immobilized (or otherwise incorporated) assay reagents. An immunoassay formulated as a RTD enables the detection of tiny quantities of specific molecules at virtually any location by untrained persons.

Reagent-excess
In reagent-excess immunoassays, the principal reagents, such as immobilized antibody and labeled antibody or labeled antigen, are used in excess. Most reagent-excess immunoassays may also be described as ''sandwich'' immunoassays.

Reagent-limited
For reagent-limited immunoassays, the principal reagents, such as immobilized antibody and labeled antibody or labeled antigen, are used at limited concentrations, and a plot of "label bound" against analyte concentration displays an inverse relationship. Reagent-limited immunoassays are commonly referred to as "competitive" immunoassays. Separation-free immunoassays are usually "reagent-limited."

Separation-free
A separation-free (homogeneous) immunoassay has no separation step. Separation-free immunoassays are usually "reagent-limited." In a normal immunoassay, after incubation of reagents with analyte, bound label must be separated from free label to enable measurement of the bound fraction. In separation-free immunoassay, the binding of label to antibody modulates the activity of label such that analyte concentration determines *total* label activity.

Solid Phase
In immunoassay, a solid-phase component may be the internal wall of a tube or well, a bead or microparticle, a membrane or a thread. Linking a basic assay reagent such as antibody or antigen to a solid-phase component allows the easy separation of bound from free label, thereby facilitating the accurate measurement of bound (or free) label.

Standardization
By standardization is meant the development and validation of assays in such a way that all assays for the same analyte give the same (numeric) results if applied to the same samples. The same results should be obtained whether the assays were developed and manufactured by different companies, used at different locations, operated by different users, or used at different times.

Structural Complementarity
Structural complementarity is the term used to explain the high affinity binding of the paratope of a specific antibody to the epitope of an antigen. The paratope and epitope have complementary shapes, complementary distributions of electric charges, and opposing groups that form good noncovalent bonds. Structural complementarity, in all its variety and complexity, explains all strong and spontaneous molecular associations and is therefore an essential part of the ultimate explanation for all biological processes.

Specific, biologically related, intermolecular interaction is the basis of all "binding assays," recombinant DNA procedures, and, accompanied by catalysis, of enzymatic analysis. "Binding assays," also referred to by the term *saturation analysis*, depend on the "bio-recognition" of analytes by high affinity, specific binding proteins. In the last forty years "protein binding assays" have found so many applications in diverse areas of biochemical and chemical analysis that they now represent a standard analytical principle, like colorimetry or chromatography.

While there are many kinds of high affinity, specific binding proteins in nature (hormone receptors, for example), the special properties of antibodies have made them the most popular choice by far, for such methods. The term "immunoassay" applies to measurement procedures that use antibodies as specific binding reagents, as well as to assays that use antigens for the detection or quantification of specific antibodies. It is because of the extraordinary variety, specificity, and affinity of antibody–antigen binding reactions, that immunoassays are used for routine analyses and for research purposes throughout the biological and medical sciences.

Often an immunoassay procedure is chosen to measure a particular analyte because no other type of assay is technically feasible, which is true for specific antibodies, for most proteins and for many other complex biomolecules. But frequently, immunoassays are used because they match or exceed analytical and practical requirements and are more convenient and cost-effective than any alternative method.

When a broad perspective is adopted, it is clear that the history of immunotechnology as applied to the detection and quantitation of antigens did not begin with the invention of radioimmunoassay in the late 1950s. Rather, experimental immunologists have always used specific antibodies and "purified" antigens in their investigations, and major developments with respect to "immunoassay methodology" occurred when these were applied clinically and/or formulated to maximize their utility to clinical microbiologists, clinical chemists, and others. Table 1 lists some of the significant advances in this long history of the development of immunoassay methodology.

1 Antibodies for Immunoassays

1.1 Immunoglobulin G Structure

Antibodies, also defined as immunoglobulins and found in blood and other body fluids of vertebrates, are essential components of the adaptive, humoral immune system. Immunoglobulin G (IgG), the class of antibody used predominantly in immunoassays, is a 150 000 Da glycoprotein containing two identical heavy (H ~420 residues) and two identical light (L ~215 residues) polypeptide chains. The two identical antigen-binding sites (also called paratopes) of IgG are composed of the atoms of about fifty amino acid residues found in groups (complementarity determining regions, CDR) scattered along the variable segments of the H and L chains. In the 3-D structure of all immunoglobulins the six CDR occur as loops and it is CDR variability that accounts for almost all of the immense range of antibody specificity.

For analytical purposes, IgG molecules can be linked covalently to labeling substances and to solid-phase materials by any of a range of standard chemical procedures. They also readily bind noncovalently to plastics and glass to give "antibody coated" tubes, wells, or beads.

Tab. 1 Milestones in the development of immunoassay methodology.

Development	*References*
Detection of bacilli by antibody agglutination	Freund (1925)
Titration of antisera by antigen precipitation	Libby (1938)
Immunofluorescence staining	Coons (1942)
Albumin measurement by antibody precipitation	Chow (1947)
RIA for antigen	Berson and Yalow (1959); Ekins (1960)
Immunoenzymatic staining	Nakane and Pierce (1966); Avrameas and Uriel (1966)
Labeled antibody RIA	Miles and Hales (1968)
IRMA for antigen	Addison and Hales (1971), Wide (1971)
EIA for antigen	VanWeeman and Schuurs (1971)
EMIT®	Rubinstein et al. (1972)
FPIA®	Dandiker et al. (1973)
Microtitre-plate IEMA for antibody	Voller (1974)
IEMA for antigen	Maiolini and Masseyeff (1975)
Acridinium ester label	Simpson et al. (1979)
Theoretical basis for immunoassays	Jackson and Ekins (1986)
Chemiluminescence enzyme assay in EIA	Arakawa (1981); Thorpe et al. (1985)
Europium chelates in IFMA	Pettersson et al. (1983)
Fab'-enzyme conjugates in IEMA	Ishikawa et al. (1983)
ELISPOT for specific antibody-secreting cells	Czerkinsky et al. (1983)
Solid-phase separation with hapten anchor	Rattle et al. (1984)
Amplification enzyme assay in IEMA	Self (1985)
Immunoconcentration IEMA	Valkirs (1985)
CEDIA®	Henderson et al. (1986)
Microspot, multianalyte immunoassay	Ekins et al. (1990)

[a]Abbreviations: RIA, radioimmunoassay; EIA, enzyme immunoassay; IRMA, immunoradiometric assay; IEMA, immunoenzymometric assay; IFMA, immunofluorometric assay.

Immunoglobulins are flexible molecules, and this facilitates their binding simultaneously to two (or more) adjacent identical antigens on the surface of a pathogen and enables the aggregation of soluble antigens. While all interdomain regions can flex and bend, the greatest flexibility is associated with the "hinge regions" of the H chains, near the C-termini of the L chains. These regions are highly susceptible to proteolytic cleavage, and pepsin specifically hydrolyzes a site located below the inter-H chain disulphydryl links to give $F(ab')_2$ dimers that are easily reduced to give Fab' fragments ($M_r \sim 50\,000$). Fab' fragments have at least one sulphydryl group (–SH) distally located from the binding site, which can be used to establish a covalent linkage to labeling molecules or solid-phase materials. Also because of their relatively small size and reduced tendency to bind "nonspecifically," Fab' fragments are highly suitable for the preparation of soluble reagents for use in immunoassays.

Fv fragments ($M_r \sim 25\,000$), which include only the variable segments of the

H and L chains, are minimal antibodies. Recombinant DNA methods enable the preparation of stable, single-chain Fv (sFv) fragments for any monoclonal antibody, and of chimeric proteins incorporating antibody-binding sites. (Phage display techniques allow the creation of vast repertoires of genes coding for heavy and light chains, *in vitro* selection of recombinant clones producing antibody of a desired specificity, and *in vitro* "affinity maturation.")

1.2 Monoclonal, Polyclonal, and Anti-hapten Antibodies

Although during the 1980s the use of monoclonal antibodies in new immunoassays increased dramatically, even now in the twenty-first century, polyclonal antibodies are still preferred as secondary reagents in all commercial immunoassays, and is a primary reagent in a large minority of immunoassay kits. Obviously, differences with respect to cost, ease of preparation and availability are still sufficient to outweigh the advantages (homogeneity and continuity of supply) associated with monoclonal (or recombinant) antibodies.

Immunoassay antibodies come from a range of species including mice, rabbits, goats, and sheep. Avian IgY class antibodies, which may have much lower "nonspecific binding properties" are also used, and camel antibodies, which lack light chains, may become significant in the future. Usually, monoclonal antibodies are raised in mice but all the prerequisites are available for the use of rats.

The power and versatility of immunoassays is largely dependent on the availability of specific antibodies for most biologically important molecules with a distinctive chemical structure and of molecular weight >150 g mol^{-1}, whether or not they are inherently immunogenic. Immunologically, haptens are defined as molecules of limited molecular weight that are not naturally antigenic, but when conjugated to a carrier (bovine serum albumin or BSA for example, with 20–30 hapten molecules per molecule of BSA) give an immunogen that can be used to generate antibodies that specifically bind free hapten with high affinity. This approach is also highly effective for larger molecules that are weakly immunogenic. Therefore, antibodies for molecules that are newly identified or synthesized for the first time are readily prepared, in most cases.

2 Antibody–antigen Binding

2.1 Structural Complementarity

Antigen–antibody interactions are reversible and for binding to be defined as high affinity there must be a combination of easy association and reluctant dissociation. Association can be regarded as occurring in two phases, with the first requiring the overcoming of repulsive forces and the second involving the establishment of multiple hydrogen, ionic, and van der Waals attractive interactions between atoms on the antibody-binding site (paratope) and the antigenic determinant region (epitope).

When the epitope of a large protein antigen is bound in a high affinity complex with a paratope specific for it, the apparent area of close contact is about 700 to 900 Å^2, with about 20 amino acid residues on both reactants apparently participating.

An appropriate pH, which is usually near pH 7, ensures that acidic and basic groups on both species are suitably dissociated or associated. However, the contribution to binding of any close sets of opposing amino acid residues may be either negligible or crucial and this may account for the fact that a genetically determined variation in an antigen may render it virtually undetectable by an immunoassay. For example, Pettersson and coworkers discovered that, in a significant subpopulation of normal women with a genetic variant of luteinizing hormone (LH), LH was almost undetectable when measured with an immunoassay employing a particular kind of monoclonal antibody.

When the high affinity complex is between a small antigen (hapten) and its specific paratope, the hapten is found to be located in a pocket, cavity, or groove on the antibody surface. Indeed, this may be an absolute requirement for *high* affinity. When bound, the hapten is surrounded by functional groups of the peptide residues and backbone that are structurally complementary to adjacent portions of the hapten molecule. Frequently, to give the number of weak bonds to the hapten that are necessary for tight binding, the hapten is buried to the extent of 80 to 90%, or even greater. Nevertheless, as established by Arevalo and coworkers, other compounds with significant structural differences from the original specific hapten, may be capable of binding tightly to the same paratope, provided they also are sufficiently complementary with respect to size, shape, and charge distribution.

As binding of epitope to paratope occurs, there may be significant structural adjustments of the paratope (and of the epitope) and this is termed *induced fit*. However, frequently there is still sufficient interstitial space to accommodate water molecules or ions, some of which may contribute significantly to binding. Therefore, the ionic composition of buffers may be relevant to the selection and performance of an antibody, and optimal assay conditions can only be estimated experimentally.

2.2 Molecular Mobility Influences Binding

Any interaction between antibody and antigen must be preceded by a collision. The rate at which collisions occur is dependent on rates of diffusion, which, in turn, are influenced by molecular mass and shape, temperature, and, of course, on whether both reactants are mobile or not. Therefore, since smaller molecules diffuse more rapidly, haptens tend to associate more readily with their specific antibodies than with large antigens, and hapten analytes associate more readily than the large hapten-enzyme conjugates used as labels in enzymeimmunoassays. In addition, antigen–antibody interactions reach equilibrium much more slowly when the antibody or antigen is immobilized to a solid phase as opposed to when both are free to diffuse. Consequentially, many automated immunoassays, which require short incubation times, involve the use of microparticulate solid-phase media and regular mixing, or are designed such that the binding reaction occurs in liquid phase and separation is effected later.

2.3 Epitopes

Each large antigen has a theoretically unlimited number of epitopes, a number that is limited in practice by immune tolerance in the immunized animal toward surface regions that share structural features with self-proteins. It is also difficult to differentiate between overlapping

epitopes. Structurally, protein epitopes are often very similar to paratopes, being made up of a number of amino acid residues scattered, singly or in small groups, along the amino acid sequence, but close together on the surface of the intact antigen. Thus, many epitopes are "discontinuous." However, a single short peptide may represent a large fraction of an epitope and may, in isolation, be specifically bound by an antibody raised against the whole protein that it represents. Correspondingly, antibodies raised against such a peptide may bind to the region of the cognate protein containing the same sequence. Such epitopes may be referred to as "continuous." Not unusually, continuous epitopes correspond to exposed termini.

The chemical nature of epitopes is highly relevant to the development of sophisticated vaccines, and is also pertinent to immunoassays. For example, assays designed to detect specific antibodies against well-characterized viral pathogens (especially HIV), may employ single peptides (or more often sets of carefully chosen peptides), immobilized and/or linked to a labeling substance, instead of whole viral antigens. This allows precisely defined epitope specificity.

3 Labeling Substances

Although, as is explained further below, many immunoassays employ no discrete labeling "compound," immunoassays only came "of age" with the invention of radioimmunoassay by Rosaleen Yalow and Roger Ekins in the late 1950s, and the gradual development of reagent excess, sandwich assays in the 1970s. During this period also, data obtained with the aid of high-throughput immunoassays with lower detection limits of 10^{-9} to 10^{-15} M helped to revolutionize many branches of biological and medical science.

The sophistication of current immunoassay methodology allows the assay developer a wide choice of labeling substances and assay endpoints, but a small number of label types dominate. Crucial attributes of labels include ease of use and the availability of simple procedures and suitable equipment, and the ability to provide very high specific activity, that is, a very strong signal per unit mass of label. These attributes are provided to a greater or lesser extent by enzymes, chemiluminescent compounds, fluorescent compounds, latex particles, microspheres, and radioisotopes. Of particular interest to the future are the range of "rare earth" ions that can be used, each with a distinct peak wavelength of fluorescent light, facilitates dual and multiassays that simultaneously measure two or more analytes. Similarly, a set of populations of microspheres with fixed intensities of one or two fluorescent dyes, allows multiplexed (up to "100-plexed") immunoassays when measured in flow-cytometry systems with two lasers.

Ligands, such as biotin, are also used as primary labels in combination with another label type that provides the endpoint. Table 2 lists a selection of general-purpose labeling substances (including atoms, ions, small and large molecules, and chelates) and some of the methods used for detecting and measuring them.

4 Classification of Immunoassays

4.1 Nomenclature

The names of immunoassays can be confusing and make it difficult to understand

Tab. 2 Some biomolecules, compound, ions, and other substances used for labeling and methods used to measure them.

Type	*Substance*	*Endpoint detection*
Enzyme	Alkaline phosphatase	Choice of endpoint including: amplified colorimetric colorimetric electrochemical luminometric visual assessment
	β-D-galactosidase	Choice of endpoint
	Glucose-6-P dehydrogenae	Colorimetric
	Horseradish peroxidase	Choice of endpoint
	Malate dehydrogenase	Colorimetric
Fluorescent	Fluorescein and other fluorescent dyes chemically coupled to antibodies or antigens, or/and in microspheres	Fluorimetric
		Polarimetric–fluorimetric
		Fluorescence flow cytometric
	Rare earth ions/chelates Dy^{3+}, Eu^{3+}, Sm^{3+} or Tb^{3+}	Time-resolved fluorimetric
Ligand	Biotin derivative	Avidin-acridinium ester, luminometric Avidin-enzyme, choice of endpoint Avidin-Eu^{3+} chelator, TR fluorimetric
	FITC	AntiFITC-enzyme, choice of endpoint
	oligonucleotide	Polymerase chain reaction or other amplification, choice of endpoint
Luminescent	Acridinium ester	Luminometric
	Isoluminol derivatives	Luminometric
	Ruthenium(II) chelate, trispropylamine	Electroluminometric
Microparticle	Colloidal gold	Visual assessment
	Colored latex	Visual assessment
	Latex etc.	Particle counting Visual assessment
Radioisotopic	^{3}H	Liquid scintillation counting
	^{125}I,	Solid scintillation counting
	^{125}I-Bolton & Hunter reagent	Solid scintillation counting

similarities and differences, and the diversity of immunoassay designs. In general, most assay names contain "immuno," the combining form of the adjective "immune," and another combining word indicating the type of label employed, for example, radioimmunoassay (RIA) or enzyme immunoassay (EIA). If an alternative biological binding protein to an antibody is employed, such as a hormone receptor, the corresponding name is radioreceptor assay.

The use of the terms RIA and EIA is usually (but not always) restricted to reagent-limited, competitive assays (see below), and reagent-excess assays (see below) are commonly distinguished by reversing the order of the combining forms,

as in immunoradiometric assay (IRMA), immunofluorometric assay (IFMA), or immunoenzymometric assay (IEMA). The term *immunometric assay* is used to refer to reagent-excess assays in general. However, similar names have occasionally been used for competitive assays with labeled antibody.

The acronym ELISA (enzyme-linked immunosorbent assay) is generally applied to all kinds of microtiter plate assays with enzymatic labels and, when encountered, cannot be relied upon to indicate a particular assay mechanism. For clarity, EIA is to be preferred over ELISA for competitive assays equivalent to RIA and IEMA for reagent-excess assays equivalent to IRMA.

Commercially developed immunoassays with patented features are often named with "acronyms" that may or may not be defined. Three separation-free assays are EMIT® (enzyme modulated immunotest), FPIA® (fluorescence polarization immunoassay) and CEDIA® (combined enzyme donor immunoassay).

4.2 The Variety of Immunoassay Formats

The versatility of immunoassays is made possible largely by the wide variety of formats that can be implemented for assays of similar basic design. For example a basic, two-site, reagent excess, sandwich assay design for an antigen analyte may be used for high performance, high throughput, laboratory-based applications; or for simple, test devices to be used by untrained persons and sold over the counter in normal retail outlets. Much of the variety in format is dependent on the different solid-phase materials and configurations (e.g. microtiter wells, microparticles, membranes or simple test devices), on the different labels used (e.g. radioactive, fluorescent, enzymatic or colored particles), and on how the endpoint is measured (with a complex electronic instrument or visually).

4.3 Representing Assay Complexes in Text

In this article, the primary reagents used and the complexes of antibodies, antigens, and other components that form during the operation of immunoassays will be represented in the text by means of a standard notation system (not all aspects of which are exploited here).

According to this system, the analyte is shown in bold print, for example, "**Ag**," "**Ha**" or "**Ab**" for general antigen, hapten and antibody analytes, or more specifically, human chorionic gonadotropin (**hCG**) for the human pregnancy hormone. Where any one of a number of labeling substances could be used, it is represented by "L," otherwise an abbreviation such as "enz," or "^{125}I" is used.

Standard abbreviations are used when the class (IgG, IgM) or type of antibody (monoclonal, mAb, mAb1, mAb2), antibody fragment (Fab') or single-chain recombinant fragment (scFv) is to be specified. The origins of antibodies can be shown with single letter abbreviations (e.g. goat, G; human, H; mouse, M; rabbit, R; sheep, S) if this is relevant. The names of unusual components, or components that must be more exactly specified, can be abbreviated or spelled out.

The hyphen ("-") sign is used to indicate associations established before commencement of the assay such as an antibody-label conjugate or an antibody-enzyme conjugate (Ab-L, Ab-enz, respectively) or an antibody immobilized on a solid phase (sp-Ab). In contrast, the en

dash ("–") represents associations formed during the course of the assay procedure. For example, the final complex of a solid-phase ELISA for an antigen is described as: sp-Ab–**Ag**–Ab-enzyme. For a reagent-limited competitive assay for a hapten analyte two complexes are formed, sp-Ab–**Ha** and sp-Ab–Ha-L, and, as explained below, only the complex that does *not* contain the analyte is measured.

4.4 The Four Main Groups of Immunoassays

Table 3 represents a system whereby most, if not all, immunoassays can be classified into four groups. In this context, it should also be borne in mind (as mentioned above) that binding proteins other than antibodies can be used as specific binding reagents in assays that are the exact equivalents to those mentioned below. For example, intrinsic factor and β-lactoglobin are used as specific binding reagents in group 2-type assays for vitamin B_{12} and folate, respectively. Also, a combination of an antibody and a particular lectin may be used to measure specifically glycoproteins with particular carbohydrate side chains, by means of a reagent-excess, sandwich assay.

Most quantitative immunoassays developed for research purposes or for applications of limited scale are simple immunoprecipitation assays (Group 1), reagent excess, sandwich assays with either direct or indirect labels (Group 2) or reagent limited, competitive assays with a separation step (Group 3). The first three groups have long been very well represented among commercially developed immunoassays sold as

Tab. 3 Classification of immunoassays into four groups.

Group	*Type*	*Examples*
1. Label free	Agglutination	Hemagglutination Latex agglutination
	Precipitation	Ouchterlony gel diffusion Rocket immunoelectrophoresis Immunoturbidimetry/nephelometry
	Immunosensors	Resonant mirror waveguide (IAsys™) Surface plasmon resonance (BIAcore™)
2. Reagent excess	One-site	Immunostaining Western blotting Fluorescence activated flow cytometry
	Two-site	Immunoenzymometric assay, (IEMA, ELISA) Immunofluorometric assay Immunoradiometric assay
3. Reagent limited	Labeled antigen or antibody	Radioimmunoassay Enzymeimmunoassay Fluoroimmunoassay
	Separation free	EMIT, CEDIA, FPIA
4. Ambient analyte	Microarray/microchip	Microspot®

stand-alone kits or run on automated analyzers.

4.4.1 Label-free Assays

Provided that the concentrations are sufficient, the molecular complexes ([Ab–Ag]*n*) generated by antibody–antigen interaction are visible to the naked eye, but smaller amounts may also be detected and measured due to their ability to scatter a beam of light. The formation of complexes indicates that both reactants are present, and in immunoprecipitation assays a constant concentration of a reagent antibody is used to measure specific antigen ([Ab–**Ag**]*n*), and reagent antigens are used to detect specific antibody ([**Ab**–Ag]*n*). If the reagent species is previously coated onto cells (as in hemagglutination assay) or very small particles (as in latex agglutination assay), "clumping" of the coated particles is visible at much lower concentrations.

A variety of assays based on these elementary principles are in common use, including Ouchterlony immunodiffusion assay, rocket immunoelectrophoresis, and immunoturbidometric and nephelometric assays. The main limitations of such assays are restricted sensitivity (lower detection limits) in comparison to assays employing labels and, in some cases, the fact that very high concentrations of analyte can actually inhibit complex formation, necessitating safeguards that make the procedures more complex. Some of these Group 1 assays date right back to the discovery of antibodies (Table 1) and none of them have an actual "label" (e.g. Ag-enz).

Other kinds of immunoassays that are label free depend on immunosensors, and a variety of instruments that can directly detect antibody–antigen interactions are now commercially available. Most depend on generating an evanescent wave on a sensor surface with immobilized ligand, which allows continuous monitoring of binding to the ligand. Immunosensors allow the easy investigation of kinetic interactions and, with the advent of lower-cost specialized instruments, may in the future find wide application in immunoanalysis.

4.4.2 Reagent-excess Assays

Group 2 contains assays that are typified by the use of an excess concentration of labeled antibody or antigen for the detection specific antigen or antibody.

"One-site" reagent-excess assays include western blotting, whereby proteins absorbed to nitrocellulose filters are located by "probing" with labeled specific antibodies (nitrocellulose-**Ag**–Ab-L) and immunostaining, such as immunofluorescence assay whereby antigen in a tissue section is visualized with specific antibodies conjugated to fluorescein (cell matrix-**Ag**–Ab-fluorescein). Fluorescence-activated cell sorting (FACS) also depend on specific immunostaining of free-floating cells (**cell-Ag**–Ab-fluorescein).

However, most often, separate binding reactions specific for two different sites on the analyte are involved, giving rise to a tri-molecular, or larger, complex with the analyte in the middle, hence the terms "two-site assay" and "sandwich assay." For example, two monoclonal antibodies against different, distally located epitopes on the analyte may be used to detect a protein hormone such as hCG with high specificity (sp-mAb1–**hCG**–mAb2-L). Similarly, two-site assays can be used to detect specific antibodies of a particular immunoglobulin class or subclass. For example, a purified allergen and an antibody against human IgE may be used to detect specific antibodies of the IgE class that bind to the allergen in question (sp-allergen–**IgE**–Ab-L). Please note also

that to detect viral (e.g. HIV) infection, the antigen employed may be derived from a crude viral lysate, be prepared by a recombinant method; or be a synthetic peptide corresponding to an appropriate viral epitope, or a defined set of such peptides (see Sect. 2.3).

Assays to detect specific antibodies that employ immobilized-antigen to capture the analyte antibody and labeled antigen to titrate the bound antibody via its free antigen-binding site (IgG) or sites (IgM or IgA) are, strictly speaking, two-site assays (sp-Ag–**Ab**–Ag-L). However, since all sites on a single antibody are identical, such "antibody bridge" assays may be regarded as resembling one-site assays with respect to their specificity. "Antibody bridge" assays are very suitable for the early detection of viral infections such as HIV because IgM class antibodies are prominent in the early immune response, and in such assays the captured antiviral IgM antibodies may each bind multiple copies of the labeled antigen, enhancing the lower detection limit.

4.4.3 Reagent-limited Assays

Group 3 contains reagent-limited assays, typified by classical RIA. These are also commonly referred to as *competitive assays* and may have labeled antigen or labeled antibody. They are used to measure the concentrations of antigens or, more commonly, of smaller analytes such as steroids or drugs. In contrast to almost all other immunoassays, the immune complexes measured or detected (sp-Ab–Ag-L or sp-Ag–Ab-L) do not contain the analyte and, therefore, all such assays give inverse (i.e. falling as opposed to rising) standard curves when the concentration of bound label is plotted against increasing analyte concentration. Because of their "indirect" nature, competitive assays are inherently less sensitive than reagent-excess assays.

However, competitive assays can readily be designed to obviate the need for a separation step before label measurement, and such assays are referred to as *separation-free* or *homogeneous assays*. To avoid the need for a separation step, the binding reaction must influence the ability of the label to generate the endpoint signal. Ideally, such modulation should be 100% efficient and the limited lower detection limits generally achievable testify to the fact that, in general, this is not realized. Separation-free assays include the commercial systems EMIT, FPIA, and CEDIA which are widely employed in the manual and automated measurement of therapeutic, toxic, and illegal drugs. (See also Sect. 5.2.)

4.4.4 Ambient Analyte Assays

Ambient analyte immunoassays (Group 4) were proposed by Roger Ekins and are performed under conditions in which the concentration of immobilized captured antibodies (or antigen molecules) is so low that, as analyte binds to them, the concentration of analyte in the reaction medium is unaffected. As with many receptors in endocrine systems, fractional binding site occupancy is independent of both the amount of capture agent and reaction volume. For such a system, very high specific activity, fluorescent labels are necessary, and the small amount of capture antibody (or antigen) may be coated at very high density within a "microspot," facilitating the development of microarray-based multianalyte systems. A range of multianalyte microassay immunoassay systems for clinical applications have been proposed and are under development by different groups, but many do not

correspond exactly to the kinetic aspects of Ekins' model.

5 Established Immunoassay Systems

Many immunoassay systems used in human clinical laboratories were first introduced in the 1970s and 1980s and have had long lives. For example, the separation-free systems EMIT (based on glucose 6-phosphate dehydrogenase labels) and FPIA were first described in the early 1970s and are still in widespread use across the world.

5.1 Automated Systems

Early clinical immunoassay analyzer systems, which provided automation of repetitive steps such as pipetting and washing, soon led to fully automated "batch analyzers;" and more recent systems are "random access." Sample and reagent volumes have also decreased and incubation times have shortened. The panels of analytes are now more extensive, so that hormones, specific antibodies and tumor markers may be available on a single instrument. New larger machines integrate the functions of immunoassay analyzers and clinical chemistry analyzers into one system.

Another trend has been away from the colorimetric determination of endpoints to chemiluminescent-based systems, whether to measure the activity of an enzyme label (e.g. DPC Immulite®) or to directly measure a luminescent label incorporating acridinium ester (Bayer ADVIA Centaur®) or Ruthenium(II) chelate (Roche Diagnostics Elecsys®).

The precision of automated immunoassays has improved in parallel with the above developments. As reported by Wheeler in 2001, the results of the UK National External Quality Assessment Schemes indicate that levels of within and between sample assay precision of less than 5% are now commonly achievable over much of the concentration range of analytes such as luteinizing hormone.

5.2 General Laboratory Assays

Most immunoassays in use today are variations on reagent-excess two-site systems for the measurement of larger antigen analytes and specific antibodies. In addition, reagent-limited assays are employed for the measurement of smaller molecules such as steroids or common drugs, usually with labeled hapten, rather than with labeled antibody. For the measurement of low concentrations of analyte (nmol L^{-1} and below), assays with a separation step (usually involving antibody immobilized on a solid phase) are used. All of these formats, adapted for use on 96 well microtiter plates, are operated manually in many laboratories for a huge range of applications covering all areas of routine and research biology. In some areas with high throughput, such as human clinical chemistry, they have been adapted to the automated systems described briefly above; in others they have been adapted to give the rapid, easy-to-use tests described below.

For small analytes such as steroids in the low pmol L^{-1} range, an assay with an extraction step to minimize sample matrix effects is to be preferred. Radioactive labels are now rarely used outside large laboratories where the appropriate traditions, facilities, and expertise persist.

Separation-free, reagent-limited immunoassays dominate in some areas, particularly for drug screening when concentrations are often in the μmol L^{-1} range. The most widely used systems have been EMIT®, FPIA® (Abbot Laboratories) and CEDIA®.

EMIT®, is a registered name of the Syva Corporation and is an acronym for "enzyme multiplied immunotest." EMIT® is almost as old as EIA and dates from the classic paper of Rubinstein, Schneider, and Ullman in 1972. The invention of EMIT is one of the most important milestones in the history of immunoassays. In one stroke, radioisotopes were avoided, procedures simplified, "turn-around-times" greatly reduced, and automation on standard clinical chemistry analyzers made feasible.

CEDIA® is a major competitor of EMIT that depends on a detailed understanding of enzymology and protein structure. To devise CEDIA, recombinant DNA technology was exploited to produce new strains of *Escherichia coli* synthesizing inactive variants of β-D-galactosidase. These were of two complementary kinds, large fragments (enzyme acceptors, EA) and small fragments (enzyme donors, ED). Some matched pairs of EA and ED fragments could associate readily to give fully active enzyme and these were exploited (with EA as label) to develop a new, versatile separation-free immunoassay system, that is suited to automation by means of many general-purpose clinical analyzers.

There are also variations of standard immunoassays that play important specific roles in basic research. One example is the ELISPOT assay that for the last 20 years has been used in immunology research laboratories to quantify T-cell immune responses in individuals infected with different pathogens. ELISPOT tests are crucial to the evaluation of vaccines and other therapies when they are under trial, and are contributing to the fights against diseases such as AIDS.

5.3 Simple to use Test Devices

Many of the most impressive inventions involve the simplification of a previously complex task, but a long road with a series of separate inventions or developments may be necessary in some cases. Searching a database for articles on pregnancy tests reveals that in the 1950s live toads or frogs were used and that days passed before results were obtained. Obviously, and to say the least, such tests required many steps, laboratory facilities, and a skilled technician. Now the "patient" just holds a simple plastic "rapid test device" (RTD) in her urine stream for a few seconds, and minutes later she has the result.

The adoption of immunoassays, such as latex agglutination tests, to detect the pregnancy hormone hCG in urine was an important first step on the road. Later came dipsticks with multiple steps and reagents, and liquid waste to dispose of. Still later came what is now the "Beckman Coulter Hybritech ICON® hCG test," which combines simplicity of operation and a low detection limit in a reagent-excess IEMA with two monoclonal antibodies, an alkaline phosphatase label, and a visible result. Two drops serum or 5 of urine, appropriately prediluted, are added to a simple small test "pot." After two further reagent additions and washes with a simple dropper and suitable short delays, one colored dot indicates a valid negative result and two dots indicate a pregnancy. All liquid waste is retained in the "pot."

One-step operation, like that described above, arrived with the development of the

Unipath ClearView® test device. There are no liquid reagents or wash solutions to add and no need to precollect the urine sample. The "mobile" antibody is linked to intensely colored microbeads, obviating the need for a color development step. The simple RTD hides a basic reagent-excess sandwich immunoassay and has all the reagents preloaded, ready to be set in motion by the addition of water, and all the water needed comes with the urine sample.

Like for the ICON® RTD, there are now many similar devices from many companies and these have been adapted for a range of tests including fertility monitoring by detecting the preovulatory surge of LH in women. Similar RTDs are also widely used to detect specific antibodies for infectious diseases such as HIV infection, Lyme disease, etc.

6 Immunoassays of the Future

6.1 Multiplexed Immunoassays

Immunoassays capable of detecting simultaneously two or up to 10 or more analytes have been available for many years. These include the extraordinary MAST® system with up to 35 different allergens coated on cellulose threads for detecting elevated levels of specific IgE class antibodies in human serum. However, many of the multiplexed immunoassay systems of the future will be fully quantitative, capable of detecting larger numbers of varied and distinct analytes, and automated.

Developments in multianalyte microarray immunoassay systems (see also Sect. 4.4.4) have benefited from parallel advances in the methodology of genomics and proteomics. Computer controlled equipment is available for the immobilization of just hundreds of capture molecules per microspot on heterogeneous arrays of hundreds of spots. There are also automated liquid pipettors that deliver submicroliter volumes of reagents and of samples to be analyzed, where and when required. In addition, charge-coupled device (CCD) detectors can scan and quantify luminescent or fluorescent light emission from each spot in a large array. Planar waveguide evanescent field technology enables microarray assays without separation or washing steps.

Multiplexed assays based on heterogeneous populations of microbeads have also been developed in a number of laboratories. These benefit greatly from the availability of diverse sets of identifiable microspheres. Each set or population of microspheres may incorporate one or two fluorescent dyes at set densities (enabling the creation of a large number of sets) and is easily and specifically quantifiable by a suitable flow cytometer. Recent flow cytometers, dedicated to suspension-array assays, have dual lasers and are capable of up to 100-plexed operations. Immunoassays for the measurement of specific antibodies and antigens developed for such systems usually correspond to Group 2 and 3 assays as described in Sect. 4.

6.2 New Assay Principles and Systems

The detailed history of immunoassay methodology is dotted by hundreds of reports of new and original systems, labels and detection methods, each presented as offering multiple advantages such as ease of use, low detection limit, accuracy, versatility, and so on. Most disappear from

view immediately, or after a flurry of articles from a single laboratory. Of the successful ones, many have endured, and have been refined gradually, over decades. Here are some new systems that may (or may not) prosper.

"Open sandwich" (OS) immunoassay could not have been envisaged before recombinant DNA methods had been used to prepare V_H and V_L polypeptides (see Sect. 1.1), because OS assays depend on antigen stabilization of the triple complex V_H–Ag–V_L. When free V_H and V_L were first studied, their association was found to give unstable Fv fragments, and that is why, for applications requiring minimal antibodies, single-chain sFv fragments, in which the two variable polypeptides are joined by a linker peptide, are always used. In contrast, Hiroshi Ueda found, in experiments with a surface plasmon resonance biosensor, that the strength of the interaction between V_H and V_L increases greatly when specific antigen is bound also. From this realization, he went on to devise a range of potentially promising one-site "sandwich" immunoassays, and separation-free assays involving fluorescence resonance energy transfer (FRET) or bioluminescence resonance energy transfer (BRET).

Colin Self is best known as the inventor of amplified enzyme assays and their application to the development of ultrasensitive immunoassays, but he has been long aware of the basic limitations of competitive immunoassays, which are the only standard immunoassays suitable for small molecular weight (<1000 g mol^{-1}) "hapten" analytes. As summarized recently by Ashton, he has devised and patented a number of sophisticated reagent-excess immunoassay systems for small analytes. The "selective antibody system," involves a blocking molecule (BM) for free sites on the immobilized primary antibody (sp-Ab, sp-Ab–BM) and a labeled anti-idiotypic antibody (aiAb-L) that is added later and binds specifically to the occupied binding site of the primary antibody (sp-Ab–**Ha**, sp-Ab–**Ha**–aiAb-L) (an anti-idiotypic antibody is an antibody that binds to the variable region of another monoclonal antibody.) In this case, because of the role of the "blocker," the anti-idiotypic antibody does not have to be specific for the *occupied* site. Note that, unlike a competitive assay, the complex measured at the end of the assay *does* contain the analyte.

Colin Self's later "apposition system" is similar to the above but does not require the difficult task of raising a suitable anti-idiotypic antibody for *every* single primary antibody used. Rather, the challenge is to link, to an appropriate residue near the binding site of the primary antibody, a ligand to which a labeled binding protein can bind. Again, a blocking molecule is used to block all empty binding sites. In principle, the same ligand can be linked to all primary antibodies to be used in a range of assays (perhaps via a suitably located cysteine sulphydryl group inserted in the peptide sequence by "site-selective mutation" of the gene coding for the V_H or V_L), making the labeled binding protein a "universal reagent."

7 The Limitations of Immunoassays

Absolute specificity that depends on a binding reaction alone is unattainable. Whether the binding protein is an antibody, a hormone receptor or a lectin, the existence of significant concentrations of cross-reacting molecules in (unrefined) samples is a continuous possibility. After

all, the science of pharmacology depends largely on finding or devising substances that strongly cross-react with "specific" binding sites.

With hapten analytes, a structurally similar cross-reactant may be able to bind to the specific paratope in a slightly different orientation, or a molecule with a single different functional group may bind because of the existence of a sufficiently large "pocket" in the binding site. There is also the well-accepted fact that changes at, or near, the site via which the hapten was conjugated to the immunogen carrier molecule often have little effect on affinity. "Dissimilar" cross-reactants, perhaps defined as such by crude comparisons of chemical formulae, are also always possible.

Large protein, glycoprotein or complex carbohydrate antigens interact with a large area of the paratope, thereby making effective epitope uniqueness more feasible. Nevertheless, cross-reacting molecules do occur and, in addition, the achievement of adequate specificity may be exacerbated by analyte heterogeneity. Two-site assays require the recognition of two spatially distinct epitopes and can be exploited to improve specificity when analytes are large, but then there are *two* interactions that are susceptible to general interference instead of one. In contrast, absolute specificity may not always be desirable. For example, as protein analytes are subject to genetic variation, subpopulations of individuals, from whom samples to be analyzed are obtained, may exist that express a variant epitope that is not recognized by a reagent antibody (see Sect. 2.1).

When the analyte is a complex biomolecule, standardization may be especially difficult to achieve because suitable primary standards and "gold-standard" reference analytical methods are unavailable. Lately, it has become possible to prepare by recombinant DNA methods primary standards for some complex analytes such as hCG and human kallikrein 2. Reference methods for such analytes also require epitope mapping of the analyte, well-characterized antibodies, and international collaboration to ensure their acceptance and use.

When the analytes are specific antibodies, the required specificity may be quite different from one application to another. The purity and suitability of the antigen reagent is critical and, when class (or subclass) specificity is also required, a two-site design is normally used, so that the specificity of the class-specific antibody used is also vital. In all these designs, antibodies against other antigens or polyspecific antibodies may be detected, and may even give rise to false-positive diagnoses.

However, perhaps no analytical method can be guaranteed in advance to be absolutely specific 100% of the time when the samples to be analyzed are complex and of variable composition. For some types of analyte, isotope dilution mass spectrometry (ID-MS) is used as the "gold standard" and this performs as well as it does because it involves at least two stages, one to separate the analyte from the general constituents of its sample matrix and one to measure it. Therefore, when the analyte is presented eventually to the mass spectrometer, it has been extensively purified by means of a high-resolution chromatographic method. In contrast, most immunoassay methods are expected to be accurate and precise when presented with untreated samples. The only preliminary treatment that is accepted as practicable in most situations is dilution.

Therefore, any really effective answer to interference and cross-reactivity as they may affect immunoassays must

also involve preliminary purification of samples. Otherwise the accuracy of every result is completely dependent on the care taken in developing and operating the test. It depends on the careful selection and purity of antibodies (and/or antigens) and other reagents used, on the careful optimization, thorough analytical and biological validation of the assay, and on thorough quality control procedures.

See also Genetic Variation and Molecular Evolution; Immunology.

Bibliography

Books and Reviews

Crowther, J.R. (2000) *ELISA Handbook*, Humana Press, Totowa, NJ.

Diamandes, E.P., Chrisopoulos, T.K. (Eds.) (1996) *Immunoassay*, Academic Press, San Diego, CA.

Gosling, J.P. (1990) A decade of development in immunoassay methodology, *Clin. Chem.* **36**, 1408–1427.

Gosling, J.P. (Ed.) (2000) *Immunoassays: A Practical Approach*, IRL Press, Oxford, UK.

Gosling, J.P., Basso, L.V. (Eds.) (1994) *Immunoassay: Laboratory Analysis and Clinical Applications*, Butterworth Heineman, Boston, MA.

Joos, T.O., Stoll, D., Templin, M.F. (2001) Miniaturised multiplexed immunoassays, *Curr. Opin. Chem. Biol.* **6**, 76–80.

Levinson, S.S., Miller, J.J. (2002) Towards a better understanding of heterophile (and the like) antibody interference with modern immunoassays, *Clin. Chim. Acta* **325**, 1–15.

Mashishi, T., Gray, C.M. (2002) The ELISPOT assay: an easily transferable method for measuring cellular responses and identifying T cell epitopes, *Clin. Chem. Lab. Med.* **40**, 903–910.

Price, C.P., Newman, D.J. (Eds.) (1997) *Principles and Practice of Immunoassay*, 2nd edition, Macmillan, London, UK.

Wild, D. (Ed.) (2001). *The Immunoassay Handbook*, 2nd edition, Nature Publishing Group, UK.

Wheeler, M.J. (2001) Automated immunoassay analysers, *Ann. Clin. Biochem.* **38**, 217–229.

Primary Literature

Abdul-Ahad, W.G., Gosling, J.P. (1987) An enzyme-linked immunosorbent assay (ELISA) for bovine LH capable of monitoring fluctuations in baseline concentrations, *J. Reprod. Fertil.* **80**, 653–61.

Addison, G.M., Hales, C.N., (1971) The Immunometric Assay [and discussion], in: Kirkham K.E., Hunter W.M. (Eds.) *Radioimmunoassay Methods*, Churchill Livingstone, Edinburgh, Scotland, pp. 447–461 [and 467–488].

Arakawa, H., Maeda, M., Tsuji, A., Kambegawa, A. (1981) Chemiluminescence enzyme immunoassay of dehydroepiandrosterone and its sulphate using peroxidase as label, *Steroids* **38**, 453–464.

Ashton, V. (2004) Measurements in miniature, *Chemistry World* **10**, 42–45.

Avrameas, S., Uriel, J. (1966) Méthode de marquage d'antigénes at d'anticorps avec enzymes et son application en immunodiffusion, *Comptes Rendus Hebdomedaires des Seances de l'Academie des Sciences: D: Sciences naturelles (Paris)* **262**, 2543–2545.

Berson, S.A., Yalow, R.S. (1959) Assay of plasma insulin in human subjects by immunological methods, *Nature* **184**, 1648–1649.

Blackburn, G.F., Shah, H.P., Kenten, J.H., Leland, J., Kamin, R.A., Link, J., Peterman, J., Powell, M.J., Shah, A., Talley, D.B., Tyagi, S.K., Wilkins, E., Wu, T.G., Massey, R.J. (1991) Electrochemiluminescence detection for development of immunoassays and DNA probe assays for clinical diagnostics, *Clin. Chem.* **37**, 1534–1539.

Brown, C.R., Higgins, K.W., Frazer, K., Schoelz, L.K., Dyminiski, J.W., Marinkovich, V.A., Miller, S.P., Burd, J.F. (1984) Simultaneous determination of total IgE and allergen-specific IgE in serum by the MAST chemiluminescent system, *Clin. Chem.* **31**, 1500–1505.

Carson, R.T., Vignali, D.A. (1999) Simultaneous quantification of 15 cytokines in a multiplexed flow cytometric assay, *J. Immunol. Methods* **227**, 41–52.

Chow, B.F. (1947) The determination of plasma or serum albumin by means of a precipitin reaction, *J. Biol. Chem.* **167**, 757–763.

Coons, A.H., Creech, H.J., Jones, R.N., Berliner, E. (1942) Demonstration of pneumococcal antigen in tissues by use of fluorescent antibody, *J. Immunol.* **45**, 159–170.

Czerkinsky, C.C., Nilsson, L.A., Nygren, H., Ouchterlony, O., Tarkpowshy, A. (1983) A solid-phase enzyme-linked immunospot (ELISPOT) assay for enumeration of specific antibody-secreting cells, *J. Immunol. Methods* **65**, 109–121.

Dandiker, W.B., Kelly, R.J., Dandiker, J., Farcuhar, J., Levin, J. (1973) Fluorescence polarisation immunoassay. Theory and experimental methods, *Immunochemistry* **10**, 219–227.

Ekins, R.P. (1960) The estimation of thyroxine in human plasma by an electrophoretic technique, *Clin. Chim. Acta* **5**, 453–459.

Ekins, R.P., Chu, F., Biggart, E. (1990) The development of microspot, multianalyte ratiometric immunoassay using dual fluorescent labelled antibodies, *Anal. Chim. Acta* **227**, 73–96.

Emerson, J.F., Ngo, G., Emerson, S.S. (2003) Screening for interference in immunoassays, *Clin. Chem.* **49**, 1163–1169. (See also Span et al. 2003, below.)

Engel, W.D., Khanna, P.L. (1992) CEDIA in vitro diagnostics with a novel homogeneous immunoassay technique: current status and future prospects, *J. Immunol. Methods* **150**, 99–102.

Finckh, P., Berger, H., Karl, J., Eichenlaub, U., Weindel, K., Hornauer, H., Lenz, H., Sluka, P., Ehrleich Weinreich, G., Chu, F., Ekins, R. (1998) Proceedings of the UK NEQAS Endocrinology Meeting, Vol. 4, Association of Clinical Biochemists, London, 155.

Freund, J. (1925) Agglutination of tubercule bacilli, *Am. Rev. Tuberc.* **12**, 124–141.

Gonzalez, C., Guevara, P., Alarcon, I., Hernando, M., Navajo, J.A., Gonzalez-Buitrago, J.M. (2002) Antinuclear antibodies (ANA) screening by enzyme immunoassay with nuclear HEp-2 cell extract and recombinant antigens: analytical and clinical evaluation, *Clin. Biochem.* **35**, 463–469.

Gosling, J.P. (1997) Enzyme-immunoassay: With and Without Separation, in: Price, C.P., Newman D.J. (Eds.) *Principles and Practice of Immunoassay*, 2nd edition, Academic Press, San Diego, CA, pp. 350–388.

Gosling, J.P., Middle, J., Siekmann, L., Read, G. (1993) Standardization of hapten immunoprocedures: total cortisol, *Scand. J. Clin. Lab. Invest.* **53**(Suppl. 216), 3–41.

Gillis, E.H., Gosling, J.P., Sreenan, J.M., Kane, M. (2002) Development and validation of a biosensor-based immunoassay for progesterone in bovine milk, *J. Immunol. Methods* **267**, 131–138.

Haese, A., Vaisanen, V., Finlay, J.A., Pettersson, K., Rittenhouse, H.G., Partin, A.W., Bruzek, D.J., Sokoll, L.J., Lilja, H., Chan, D.W. (2003) Standardisation of two immunoassays for human glandular kallikrein 2, *Clin. Chem.* **49**, 601–610.

Henderson, D.R., Friedman, S.B., Harris, J.B., Manning, W.B., Zoccoli, M.A. (1986) 'CEDIA', a new homogeneous immunoassay system, *Clin. Chem.* **32**, 1637–41.

Herold, D.A., Fitzgerald, R.L. (2003) Immunoassays for testosterone in women: Better than a guess? *Clin. Chem.* **49**, 1250–1251.

Huang, R.P. (2001) Simultaneous detection of multiple proteins with an array-based enzyme-linked immunosorbent assay (ELISA) and enhanced chemiluminescence, *Clin. Chem. Lab. Med.* **39**, 209–214.

Ishikawa, E. (1973) Enzyme immunoassay on insulin by fluorimetry of insulin-glucoamylase complex, *J. Biochem.* **73**, 1319–21.

Ishikawa, E., Imagawa, M., Hashida, S., Yoshitake, S., Hamaguchi, Y., Ueno, T. (1983) Enzyme-labelling of antibodies and their fragments for enzyme immunoassay and immunohistochemical staining, *J. Immunoassay* **4**, 209–327.

Jackson, T.M., Ekins, R.P. (1986) Theoretical limitations on immunoassay sensitivity. Current practice and potential advantages of fluorescent Eu3+ chelates as non-radioisotopic tracers, *J. Immunol. Methods* **87**, 13–20.

Jani, I.V., Janossy, G., Brown, D.W.G., Mandy, F. (2002) Multiplexed immunoassays by flow cytometry for diagnosis and surveillance of infectious diseases in resource-poor settings, *Lancet Infect. dis.* **2**, 243–250.

Libby, R.L. (1938) A new and rapid quantitative technique for the determination of potency of types I and II antipneumococcal serum, *J. Immunol.* **34**, 269–279.

Luzzi, V.I., Saunders, A.N., Koenig, J.W., Turk, J., Lo, S.F., Carg, U.C., Dietzen, D.J. (2004) Analytic performance of

immunoassays for drugs of abuse below established cutoff values, *Clin. Chem.* **50**, 717–722.

Maiolini, R., Masseyeff, R. (1975) A sandwich method of enzymeimmunoassay. I. Application to rat and human alpha-fetoprotein, *J. Immunol. Methods* **8**, 223–234.

Miles, L.E.M., Hales, C.N. (1968) Labelled antibodies and immunological assay systems, *Nature* **219**, 186–169.

Nakane, P.K., Pierce, G.B. (1966) Enzyme-labelled antibodies: preparation and application to the localization of antigens, *J. Histochem. Cytochem.* **14**, 929–931.

Nayak, P.N. (1981) The kinetics of solid-phase immunoassays, *Ligand Q.* **4**, 34–42.

O'Connor, A., Gosling, J.P. (1997) The dependence of detection limit on antibody affinity in competitive radioimmunoassays for progesterone, *J. Immunol. Methods* **208**, 181–189.

Odell, W.D., Griffin, J. (1987) Two-monoclonal-antibody "sandwich"-type assay of human lutropin, with no cross reaction with choriogonadotrophin, *Clin. Chem.* **33**, 1603–1607.

O'Rorke, A., Kane, M.M., Gosling, J.P., Tallon, D.F., Fottrell, P.F. (1994) Development and validation of a monoclonal antibody enzymeimmunoassay for the measurement of progesterone in saliva, *Clin. Chem.* **40**, 400–410.

Pettersson, K., Ding, Y.Q., Huhtaniemi, I. (1992) An immunologically anomalous luteinizing hormone variant in a healthy woman, *J. Clin. Endocrinol. Metab.* **74**, 164–171.

Pettersson, K., Siitari, H., Hemmilä, I., Soini, E., Lövgren, T., Hänninen, V., Tanner, P., Stenman, U-H. (1983) Time-resolved fluoroimmunoassay of human choriogonadotrophin, *Clin. Chem.* **29**, 60–64.

Rattle, S.J., Purnell, D.R., Williams, P.I.M., Siddle, K., Forrest, G.C. (1984) New separation method for monoclonal immunoradiometric assays and its application to assays for thyrotropin and human choriogonadotrophin, *Clin. Chem.* **30**, 1457–1461.

Rowe-Taitt, C.A., Hazzard, J.W., Hoffman, K.E., Cras, J.J., Golden, J.P., Ligler, F.S. (2000) Simultaneous detection of six biohazardous agents using a planar waveguide array biosensor, *Biosens. Bioelectron.* **15**, 579–589.

Rubinstein, K.E., Schneider, R.S., Ullman, E.F. (1972) Homogeneous enzyme immunoassay: a new immunochemical technique, *Biochem. Biophys. Res. Commun.* **47**, 846–851.

Schneyer, A.L., Sluss, P.M., Whitcomb, R.W., Hall, J.E., Crowley, W.F. Jr., Freeman, R.G. (1991) Development of a radioligand receptor assay for measuring follitropin in serum: application to premature ovarian failure, *Clin. Chem.* **37**, 508–514.

Self, C.H. (1985) Enzyme amplification – a general method applied to provide an immunoassisted assay for placental alkaline phosphatase, *J. Immunol. Methods* **76**, 389–393.

Span, P.N., Grebenchtchikov, N., Sweep, C.G.J. (2003) Screening for interference in immunoassays, *Clin. Chem.* **49**, 1708–1709.

Stenman, U-H., Bidart, J-M., Birkin, S., Mann, K., Nisula, B., Oconnor, J. (1993) Standardization of immunoprocedures: chorionic gonadotrophin (CG), *Scand. J. Clin. Lab. Invest.* **53**(Suppl. 216), 42–78.

Taieb, J., Mathian, B., Millot, F., Patricot, M-C., Mathieu, E., Queyrel, N., Lacroix, I., Somma-Delpero, C., Boudou., P. (2003) Testosterone measured in 10 immunoassays and by isotope-dilution gas chromatography-mass spectrometry in sera from 116 men, women and children, *Clin. Chem.* **49**, 1381–1395. (See also Herold and Fitzgerald 2003, above.)

Tamate, K., Kane, M.M., Charleton, M., Egan, D., Gosling, J.P., Ishikawa, M., Fottrell, P.F. (1997) Direct colorimetric monoclonal antibody enzymeimmunoassay for estradiol-17b in saliva, *Clin. Chem.* **43**, 1159–1164.

Thorpe, G.H.G., Kricka, L.J., Mosely, S.B., Whitehead, T.P. (1985) Phenols as enhancers of the chemiluminescent horseradish peroxidase-luminol-hydrogen peroxide reaction: application in luminescence monitored enzyme immunoassays, *Clin. Chem.* **31**, 1335–1341.

Ueda, H. (2002) Open sandwich immunoassay: a novel immunoassay approach based on the interchain interaction of an antibody variable region, *J. Biosci. Bioeng.* **94**, 614–619.

Valkirs, G.E., Barton, R. (1985) ImmunoConcentration™ – a new format for solid phase immunoassays, *Clin. Chem.* **31**, 1427–1431.

Van Weeman, B.K., Schuurs, A.H.W.M. (1971) Immunoassay using antigen-enzyme conjugates, *FEBS Lett.* **15**, 232–236.

Voller, A., Bidwell, D.E., Huldt, G., Engvall, E. (1974) A microplate method of enzyme-linked

immunosorbent assay and its application to malaria, *Bull. W. H. O.* **51**, 209–211.

Wide, L. (1971) Solid Phase Antigen-antibody Systems, in: Kirkham, K.E., Hunter, W.M. (Eds.) *Radioimmunoassay Methods*, Churchill Livingstone, Edinburgh, Scotland, pp. 405–412.

Immunologic Memory

Alexander Ploss[1,2] *and Eric G. Pamer*[1]
[1] *Sloan-Kettering Institute, New York, USA*
[2] *Weill Graduate School of Medical Sciences of Cornell University, New York, USA*

1 Introduction 386

2 Characteristics of Memory Cells 388

3 CD8+ T-Cell Memory 388
3.1 Phenotyping Memory CD8+ T Cells 388
3.2 Enhanced Responsiveness of Memory CD8+ T Cells: Potential Mechanisms 389
3.3 Generation of Memory CD8+ T Cells 390
3.4 Maintaining CD8+ T-Cell Memory 392
3.5 Models of CD8+ T-Cell Memory Generation 393

4 CD4+ T-Cell Memory 396
4.1 Differentiation of Effector and Memory CD4+ T Cells 396
4.2 Phenotype of Memory CD4+ T Cells 396
4.3 Memory Generation and Maintenance 397
4.4 Trafficking of Memory CD4+ T Cells 397

5 B-Cell Memory 398
5.1 Generation of B-Cell Memory 398
5.2 Maintenance of B-Cell Memory 399

6 Conclusions 399

Bibliography 399
Books and Reviews 399
Primary Literature 400

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

Keywords

Adaptive Immune Response
In contrast to the innate or inborn immune response, the adaptive or acquired immune response is the response of antigen-specific lymphocytes to antigen and is generated by clonal expansion of lymphocytes.

Affinity Maturation
A process during the humoral immune response, which selects for survival of B cells with high affinity for their antigen, which becomes particularly apparent during secondary and subsequent immunizations.

Antigen
Any substance that stimulates the production of antibodies, hence the name is derived from their ability to *gen*erate *anti*bodies. Here, the term is also used for molecules that are recognized by T cells.

Apoptosis
In contrast to necrosis, which is induced by poisoning or anoxia, apoptosis is a programmed cell death in which the cell activates a specialized cellular process to destroy itself. This cell suicide mechanism enables metazoans to control cell numbers and is characterized by nuclear condensation, nuclear degeneration, and DNA degradation.

Chemokine
Proteins secreted by cells that stimulate the activity of other cells. Chemokines have chemoattractant properties, act as messengers between cells and have a central role in inflammatory processes.

Clusters of Differentiation (CD)
A term that was originally coined to define cell-surface molecules that are recognized by a given set of monoclonal antibodies.

Cytokine
Any of various proteins secreted by cells of the immune system that serve to affect the behavior of other cells.

Cytotoxic T cell (CTL)
CTLs are lymphocytes that can kill other cells and are particularly important for the host defense against intracellular pathogens. Most CTLs are MHC class I restricted CD8+ T cells but MHC class II restricted CD4+ T cells can also kill in some cases.

Dendritic Cells (DCs)
Their name originates from their branched or dendritic morphology. The DCs found in lymphoid tissues are the most potent stimulators of T-cell responses. Follicular DCs present antigen to B cells.

Helper CD4+ T Cell (Th Cells)
Th cells are CD4+ T lymphocytes that can help B cells to make antibodies and have been implicated in the generation of functional effector and memory CD8+ T cells. Depending on their cytokine expression profile helper T cells can be divided into Th1 cells, which produce IFN-γ and Th2 cells, which characteristically produce IL-4 and IL-5.

Homeostasis
A state of sustained equilibrium in which cells maintain stable numbers. In the case of lymphocytes, it refers to maintenance of normal numbers of lymphocytes in a noninfected individual.

Immunity
Ability to resist infections.

Immunoglobulin
General term for proteins that function as antibodies.

Interleukin (IL)
Generic term for cytokines produced by lymphocytes.

Lymphocyte
A lymphocyte is any of a group of white blood cells of crucial importance to the adaptive immune system. The two major subsets of lymphocytes are *thymus-dependent lymphocytes* or *T cells* and *B cells*, which are generated in the bone marrow.

Lymphatic System
A system of lymphoid channels and tissues that drains extracellular fluid from the periphery via the thoracic duct to the blood. It includes lymph nodes, Peyer's patches, and other organized lymphoid organs apart from the spleen, which communicates directly with the blood.

Major Histocompatibility Complex (MHC)
It is the most polymorphic gene clusters in the human genome and is encoded on human chromosome 6 (mouse chromosome 17). MHC molecules are membrane glycoproteins that present peptides to T cells. Other gene products of this region included proteins involved in antigen processing and other aspects of host defense.

Pathogen
Any disease-causing organism, such as a virus, bacterium or fungus.

Plasma Cells
Terminally differentiated B cells that are the main antibody-secreting cells of the body. They are found in the medulla of the lymph nodes, in splenic red pulp, and in the bone marrow.

T-Cell Receptor (TCR)
One of the main characteristics of T cells that distinguishes them from all other lymphocytes. It consists of a disulfide-linked heterodimer of the highly variable α- and β-chains, which form a stable complex with the invariant CD3 chains on the cell surface. T lymphocytes expressing this composition of receptor are referred to as α:β T cells, distinguishing them from a smaller subset of T cells that express an alternative receptor on the cell surface consisting of variable γ- and δ-chains in complex with CD3.

Telomere
The end of a chromosome. This specialized structure is involved in the replication and stability of linear DNA molecules.

Transgene
A gene that is integrated into the germline that is expressed *in vivo*.

Vaccination
Process of inducing an adaptive immune response against a given pathogen by stimulation with a dead or attenuated form of the pathogen.

Immunologic memory provides long-term protection against infectious pathogens and is the basis for all vaccinations. Long-term protection is mediated by specialized antigen-specific cells of the adaptive immune system, memory T- and B cell, as well as plasma cells. In response to an invading pathogen, naïve T cells specific for foreign antigen are primed by dendritic cells, expand and differentiate into effector cells that contribute to the clearance of the infection. After the initial immune response, most expanded T cells die and a small population of long-lived memory cells remains. Memory cells are qualitatively distinct from their antigen-inexperienced precursors since they proliferate and acquire their effector functions more rapidly upon repeat exposure to antigen.

1 Introduction

Immunologic memory, manifested as specific resistance to a second encounter with a pathogen, has been appreciated for more than 2000 years. In his book *The history of the Peleponnesian war*, Thucydides describes how a plague – the true pathogen is unknown – swept through Athens in 431 BCE. Thucydides made several observations on the plague including the fact

that those who survived an infection were "*never attacked twice-never at least fatally*." In other words, the few individuals surviving a first episode of plague did not develop disease when similar epidemics recurred.

Although many similar examples of immunity following survival of infection were appreciated over the millennia, intentional exploitation of the mammalian immune system to prevent disease is a relatively recent activity. Between 1718 and 1721, Lady Mary Wortley Montagu, the wife of the British Ambassador to Constantinople, described in her letters a local practice among the Turks for the prevention of smallpox. Turkish women infected healthy individuals intravenously with a small inoculum of matter collected from smallpox lesions, which caused a mild form of the disease, but rendered the person immune to severe smallpox. Almost 80 years later, in 1796, the British physician Edward Jenner reported successful protection of humans against smallpox infection by vaccination with cowpox virus. Vaccination, a term derived from vacca (Latin for cow), became widespread upon confirmation of Jenner's results.

Nearly a century later, Louis Pasteur, while attempting to demonstrate the bacterial cause of chicken cholera, discovered that bacteria grown in culture lose their virulence but maintain the ability to elicit protective immunity. His discovery that chickens immunized with attenuated bacteria developed only mild disease but became resistant to infection with fully virulent bacteria of the same species validated Jenner's concept of immunization. Thus, in the late-nineteenth century, immunization with attenuated vaccines was developed for a number of diseases and applied in Europe and America.

However, many vaccines, particularly those focusing on syphilis, tuberculosis, and salmonellosis, were unsuccessful. Neither Pasteur nor Robert Koch, who outlined the parameters required to characterize an infectious disease, could explain mechanisms of infectious disease and immune defense. Groundbreaking work by Emil von Behring and Shibasaburo Kitasato led to the discovery of antibodies, opening the field of "humoral" immunity. Serum transfer from animals infected with diphtheria into healthy animals and subsequent infection with the pathogen made them resistant to the normally lethal inoculum. This method of conferring "passive immunity" provided the first real opportunity to define immune mechanisms. The nature of the antitoxin antibody was recognized by von Behring and Kitasato to be diphtheria-specific since it could not confer protection to any other disease.

The intense focus on serum-mediated immunity relegated the demonstration by Metchnikoff of cell-mediated antimicrobial defense to a back burner, and it took many decades for this aspect of protective immunity to gain a position at center stage. Indeed, a series of seminal papers by James Murphy, including experiments that demonstrated the "role of lymphoid tissue in the resistance to experimental tuberculosis in mice," remained widely unrecognized. Not until the 1940s did Peter Medawar and Thomas Gibson redirect attention to cell-mediated inflammatory processes. Medawar and Gibson observed that skin graft rejection correlates with cellular infiltration. The pioneering experiments performed by Karl Landsteiner and Merrill Chase showed that resistance to tuberculosis could be established by passive transfer of mononuclear cells. Additional studies by

Mackaness and colleagues demonstrated the critical role of cellular immunity in certain bacterial infections. Subsequently, James Gowans demonstrated that adaptive immunity is mediated by recirculating, long-lived lymphocytes. The thymic origin of these lymphocytes, which were therefore termed T cells, was recognized, and it was shown that they play a central role in mediating resistance to intracellular pathogens. The role of CD4+ and CD8+ T-lymphocyte subsets in adoptive protection against intracellular infection was elucidated shortly thereafter. Both CD4+ and CD8+ T cells provide long-term protective immunity against infections, and much has been learned in the last decade about T cell–mediated immunity. This article will review our current understanding of immunologic memory, focusing on more recent advances.

2 Characteristics of Memory Cells

Memory T- and B-lymphocytes mediate long-term protective immunity to infectious pathogens. A characteristic feature of memory cells is their long-term survival in the absence of reexposure to the pathogen. Additional characteristics of memory T cells are their increased frequency and enhanced responsiveness to antigen. Thus, upon reexposure to antigen following secondary infection, memory cells expand more rapidly and vigorously and acquire effector functions faster than naïve T cells. Indeed, lower concentrations of antigen and less costimulation are required for memory T cell–activation. The physiologic basis for these hallmarks of immune memory has become clearer over recent years.

3 CD8+ T-Cell Memory

3.1 Phenotyping Memory CD8+ T Cells

Memory T-cell subsets are commonly distinguished from naïve cells on the basis of surface marker expression. A number of surface molecules associated with memory and naïve CD8+ T cells are involved in cell adhesion and chemotaxis, consistent with distinct trafficking of different T-cell populations. During a primary immune response, IL-7Rα (CD127) is downregulated on most CD8+ T cells and only a small subset of cells continues to express this marker during the effector phase. Only those cells that are CD127^{hi} survive and contribute to the pool of memory cells. Hence, CD127 is a marker for precursor cells that are destined to become memory cells. The expression of the adhesion molecules CD44 and CD62L has been applied to define memory cells. While naive cells are CD44^{lo} and CD62L^{hi}, they become CD44^{hi} and CD62L^{lo} upon activation. To make matters more complicated, however, CD62L is reexpressed on a fraction of memory cells, and expression of CD44 varies among different mouse strains. Recent studies have demonstrated that memory CD8+ T cells can be divided into effector memory (CD62L^{low} CCR7^{-}) and central memory (CD62L^{hi} CCR7^{+}) populations, on the basis of L-selectin (CD62L) and the CC-chemokine receptor 7 (CCR7) expression.

The expression of tyrosine phosphatase CD45 isoforms is also used to distinguish between T-cell differentiation stages. The CD45RB isoform is generally detected on naive cells while memory cells express the CD45RA and CD45RO isoforms. However, because CD45 isoform expression

on memory versus naïve T cells is complex and can change with nonspecific activation and cytokine treatment, CD45 is a relatively unreliable marker for memory T-cell definition.

The IL-2 receptor α-chain (CD25) and CD69 are rapidly expressed by T cells upon encountering cognate antigen, but their expression is not sustained on memory T cells. Similarly, O-glycosylation on mucin-type surface glycoproteins such as CD43 can distinguish effector from memory cells. Some memory T cells, however, express the IL-2 receptor β-chain (CD122), a component common to both IL-2 and IL-15 receptors, the latter cytokine being essential for memory CD8+ T-cell maintenance. Ly-6C has also been used as a marker to distinguish effector from memory T cells. Low and intermediate expression of this marker is characteristic of effector T cells, while it is highly expressed on memory T cells.

More recently, classification of human virus-specific T cells into naïve, effector, and memory cells on the basis of CD27 and CD28 expression was described. Human immunodeficiency virus (HIV), Epstein–Barr virus (EBV), cytomegalovirus (CMV) and hepatitis C virus (HCV) specific T cells, while similar during primary infection, give rise to distinct memory T-cell populations during chronic infection, which are characterized by differential expression of various surface antigens and effector molecules. Naïve CD8+ T cells are CD27+ CD28+ and express CCR7. Early and intermediate effector T cells downregulate CD28, CCR7, and CD45RA and express cytotoxic factors such as perforin, granzyme A, and GMP-17 (a marker for cytotoxic granules and lysosomes). Memory cells appear to lose the CD28, CD27, and CCR7 markers and maintain CD45 expression at low levels. They exhibit a greater cytotoxic potential than nonprimed cells and their telomeres are shortened. In this study, T-cell populations responding to persistent viral infections were more diverse than expected. EBV and HCV promoted the development of early effector memory T cells while HIV induced intermediate and CMV induced terminally differentiated memory T cells. In aggregate, recent studies demonstrate the remarkable diversity within memory T-cell populations and illustrate the importance of integrating phenotypic and functional characteristics when assessing memory T cells.

3.2 Enhanced Responsiveness of Memory CD8+ T Cells: Potential Mechanisms

Enhanced sensitivity of memory T cells for cognate antigen likely results at least in part from more rapid signal transduction. Priming of naïve T cells leads to association of the tyrosine kinase Lck with the CD8 coreceptor, thereby enhancing TCR signaling. The association between Lck and CD8 is maintained in memory cells, which contributes to the lower stringency for activation of effector and memory cells. PEST domain-enriched tyrosine phosphatase (PEP) has recently been implicated as a negative regulator for specific aspects of T-cell development and function. Mice deficient in PEP demonstrate enhanced activation of Lck, which results in increased expansion and function of the effector and memory T-cell pool. Additional qualitative differences have been reported further downstream in the signaling cascade, involving the mitogen-activated protein (MAP) kinases ERK1 and 2 (extracellular signal-regulated

kinase), which activate transcription factors essential for T-cell activation. In contrast to naïve T cells, memory cells show an increased ability to phosphorylate ERK1 and ERK2, thereby accelerating signal propagation.

In addition to enhancing the rate of T-cell activation, regulation of signaling molecules can also influence activation-induced differentiation of memory cells. For example, transgenic mice expressing a partially calcium-independent mutant of the calcium/calmodulin kinase II (CaMKII) γB show an increase in the number of T cells in secondary lymphoid organs with a memory phenotype suggesting that this kinase is important for the development of memory.

The stage of the cell cycle that most naïve and antigen experienced T cells occupy also differs. Naïve T cells are arrested in the G0/G1 stage, with high expression of the cyclin dependent kinase (CDK) inhibitor p27Kip1 and low activity of the CDK6 and CDK2. In contrast, memory CD8+ T cells are also in G0/G1 but have low expression of p27Kip1 and high CDK6 kinase activity. This results in a preactivated cell cycle state of memory T cells that favors rapid division after antigen stimulation.

Lymphocyte differentiation is associated with programmed alterations in gene expression, which is largely regulated by structural changes in chromatin. Distinct gene expression profiles have been reported for memory cells. For example, mRNA levels for several cytokines including RANTES, interferon-γ (IFN-γ), and cytotoxic molecules such as perforin and granzyme B are elevated in memory T cells. Transcripts for these genes are rapidly translated in memory T cells following TCR stimulation, circumventing the time required for transcriptional activation.

3.3 Generation of Memory CD8+ T Cells

At the end of an immune response to an invading pathogen, most of the clonally expanded T cells die, leaving a small population of memory T cells, which is maintained for long periods of time. How do memory T cell–precursors survive during the contraction phase of the immune response and which cells become long-lived memory cells? Recent experiments revealed that the homotypic form of CD8, CD8$\alpha\alpha$, is selectively expressed by CD8+ memory T cell–precursors and is required for their survival. Mice that are deficient in the CD8 enhancer ($E8_I$) express the usual heterotypic CD8$\alpha\beta$ molecule but not the homotypic form CD8$\alpha\alpha$. $E8_I$ knock-out mice mount a normal primary immune response but show diminished memory expansion. CD8$\alpha\alpha$ interacts with the non-classical MHC class I like molecule, thymic leukemia (TL) antigen. This interaction has been shown to modulate T-cell responses and leads to enhanced expression of IL-2/IL-15Rβ and IL-7Rα. CD8+ memory T-cell precursors express IL-7Rα chain, which renders them responsive to cytokine mediated survival signals. IL-7 is a potent survival factor for both naïve and memory CD8+ T cells due to the induction of antiapoptotic factors such as Bcl-2 and Bcl-X_L.

Studies of CD8+ T-cell memory have been greatly facilitated by the increasing spectrum of mice with targeted genetic deletions. For example, mice deficient in CD27, a member of the TRAF-linked tumor necrosis factor (TNF) receptor family that plays a costimulatory role during T-cell priming, have a markedly diminished virus-specific memory CD8+ T-cell response. On the other hand, mice lacking the CD28 costimulatory molecule, while

deprived of CD86-mediated helper signals, mount nearly normal memory CD8+ T-cell responses in response to lymphocytic choriomeningitis virus (LCMV) or *Listeria monocytogenes* infection, supporting the notion that the costimulatory networks contain built-in redundancies that can compensate for the absence of CD28. However, costimulatory requirements for memory T-cell generation can vary significantly, since different pathogens create distinct inflammatory milieus. For example, induction of long-term immunity to influenza virus infection in contrast to LCMV and *L. monocytogenes* infection, requires CD28 signaling. 4-1BB, a TNF receptor family member, is another costimulatory molecule that has an impact on memory T-cell formation. Mice lacking 4-1BB ligand have diminished late expansion of virus-specific effector and memory CD8+ T cells, suggesting that 4-1BB signals play a role in sustaining T-cell activation and memory generation.

TNFR-associated factor 2 (TRAF2) is an adapter protein, which associates with the cytoplasmic tail of several TNF receptor family members including CD27, CD40, OX40, and 4-1BB and thereby links signals to downstream pathways. Hence, the absence of TRAF2 is likely to abrogate signaling through multiple costimulatory pathways. However, studies using dominant negative forms of TRAF2 demonstrated that in response to influenza virus infection, only secondary CD4+ and CD8+ T-cell expansion is decreased, while the primary expansion is not affected.

Cytokines are also implicated in the generation of antigen-specific memory T cells. Infection of IL-15- or IL-15Rα-deficient mice with LCMV generates a potent primary response but the memory pool decreases gradually over time, leading to impaired secondary CD8+ T-cell expansion. This demonstrates differential IL-15 requirements for naïve and memory CD8+ T cells. In the absence of IL-15 signaling, the memory response is also diminished following vesicular stomatitis virus (VSV) infection; however, the primary CD8+ T-cell expansion also appeared to be impaired. These studies suggest that in some circumstances, IL-15 may play a role in the generation of memory CD8 T cells.

CD4+ T cells have also been implicated in the generation of CD8+ T-cell memory; however, their importance varies with the type of infection. In the setting of LCMV infection, CD4+ T cells are not necessary for primary CD8+ T-cell responses, but they are indispensable for long-term CD8+ T-cell memory. Remarkably, CD4+ T cells contribute to CD8+ T-cell memory at the time of initial priming, since activation of memory CD8+ T cells occurs independently of CD4+ T-cell help. In contrast to viral infections, generation of memory CD8+ T cells during bacterial infections appeared to be less dependent on CD4+ T cells. For example, although mice lacking the class II transactivator (CIITA) have diminished numbers of CD4+ T cells, they generate memory CD8+ T cells following infection with the *L. monocytogenes*. Consistent with this result, CD4+ T cell-depleted, CD4−/−, and MHC class II−/− mice immunized with *L. monocytogenes* 3 to 4 weeks prior to rechallenge, mount protective recall responses. However, the consequences of the lack of CD4+ T cells help in generating functional CD8+ T-cell memory become more apparent many months following the initial antigen encounter. Recent data demonstrated that memory CD8+ T cells generated without CD4 help are defective in their ability to respond to secondary encounters with antigen. Memory CD8+

T-cell generation seems to be dependent on the presence of Th cells during, but not after, priming, suggesting that T-cell help is "programmed" into CD8+ T cells during priming. However, this concept was challenged by studies in which memory CD8+ T cells were adoptively transferred into wild type or CD4+ T cell-deficient recipients. The presence of CD4+ T cells was only important after, but not during, the early CD8+ T-cell programming phase, suggesting that CD4+ T cells are required only during the maintenance phase of long-live memory CD8+ T cells.

An important mechanism by which CD4+ T cells provide help to CD8+ T cell is through CD40-CD154 interactions. Antigen-presenting cells (APCs) were originally thought to be the mediators for these signals. Recent data demonstrates, however, that activated CD8+ T cells that express CD40 may receive direct help for CD4+ T cells expressing CD154. Administration of agonistic anti-CD40 monoclonal antibodies greatly enhances the protective potential of adoptively transferred, *L. monocytogenes*-specific CD8 T cells, providing further support for the role of CD40 in memory T-cell generation and/or maintenance.

Signals transmitted from CD4+ T cells are not uniformly stimulatory. In recent years, CD4+ CD25+ T regulatory cells have been implicated in the control of autoimmunity. During infection with *L. monocytogenes*, CD4+ CD25+ T regulatory cells can suppress memory CD8+ T-cell expansion. Similar observations were reported for memory CD8+ T-cells responses to infection with herpes simplex virus (HSV). Thus, CD4+ T cells play both positive and negative roles in the control of CD8+ T-cell memory.

Primary infection with *L. monocytogenes* induces memory T cells specific for multiple distinct peptides. CD8+ T cells restricted by the nonclassical MHC class Ib molecule H2-M3 and specific for *N*-formyl methionine peptides expand more rapidly than MHC class Ia-restricted CD8+ T cells during a primary infection. However, the memory responses of H2-M3-versus MHC class Ia-restricted T cells are dramatically different. H2-M3-restricted memory T cells express activation markers upon reencountering antigen, but do not proliferate, while MHC class Ia-restricted T cells undergo explosive expansion. It was initially suggested that this disparity may reflect differences in thymic selection of these two CD8+ T-cell types. However, more recent data show that MHC class Ia-restricted memory T cells inhibit the H2-M3 memory response by rapidly terminating *in vivo* antigen presentation of H2-M3 antigen following secondary infection.

3.4 Maintaining CD8+ T-Cell Memory

Most infections are eliminated by the mammalian host, yet memory T cells persist afterwards for many years. Determining the mechanisms that enable long-term memory T-cell persistence in the absence of antigen has been a great challenge. Pioneering studies with adoptively transferred, antigen-specific memory T cells demonstrated the ability of these cells to survive in MHC-deficient mice. These observations provide strong evidence that once the differentiation program is initiated, memory CD8+ T cells can persist in the absence of antigen. However, immunological T-cell memory that is maintained without MHC contact may become functionally impaired. For example, memory CD4+ T cells specific for the male (HY) antigen show defective responses upon rechallenge and also alter

their surface phenotype expression if they are kept in MHC-deficient hosts.

If maintenance of memory T cells is MHC and antigen independent, what regulates their persistence? Recent studies have demonstrated that IL-15 is crucial for *in vivo* maintenance of memory CD8+ T cells. Both primary and memory CD8+ T-cell expansion is diminished in IL-15- or IL-15-receptor-deficient mice, and maintenance of memory T-cell populations following immunization is markedly attenuated in the absence of IL-15 signaling. The gradual decline in memory CD8+ T-cell numbers in the absence of IL-15 results from decreased proliferation by memory cells. Interestingly, IL-7 can compensate for the missing survival signals in the absence of IL-15. When IL-7 is overexpressed in IL-15−/− mice, memory CD8+ T cells are generated, while removal of IL-7 and IL-15 completely inhibits the homeostatic proliferation of CD8+ memory T cells. These findings emphasize the joint contribution of these two cytokines for promoting CD8+ memory T-cell proliferation. Surprisingly, IL-2, a factor best known for its ability to stimulate T-cell proliferation, appears to play a negative role in memory T-cell survival.

In comparison to effector CD8+ T cells, memory CD8+ T cells express high levels of bcl-2 mRNA and protein, explaining their relative resistance to apoptosis. One effect of IL-15 is to increase bcl-2 levels in memory T cells, providing a second IL-15-mediated mechanism for memory CD8+ T-cell maintenance.

Cap structures that protect telomeres from degradation and terminal fusion play an essential role in stabilizing chromosome ends. Telomeres tend to progressively shorten when cells proliferate, eventually resulting in death of the proliferating cell. Telomere length, however, can be maintained in cells with activated telomerase, which adds DNA to telomeres with each cell division. Naïve as well as memory T cells undergo massive proliferation following antigen encounter. In order to compensate for telomere shortening, telomerase is more active in memory than naïve or effector T cells, potentially contributing to their long-term survival.

3.5 Models of CD8+ T-Cell Memory Generation

While some aspects of memory T-cell maintenance have been characterized, the mechanisms that generate memory CD8+ T cells are less clear. Several models for memory generation have been proposed and are outlined in this section. The recent introduction of methods that precisely quantify antigen-specific T cells during immune responses has provided immunologists with a framework upon which to build these models.

The *linear differentiation model* of memory T-cell formation proposes progressive differentiation of T cells from naïve to effector to memory cells (see Fig. 1). Following initial stimulation, an activation program drives naïve T-cell differentiation into effector cells, followed by further differentiation of a subset of effector T cells into memory T cells. Early support for this model came from TCR repertoire analyses of effector and memory T cells. The TCR repertoire of primary CD8+ T-cell responses is similar to that of memory CD8+ T cells, suggesting that memory cells are stochastically chosen from the population of effector CD8+ T cells activated during the primary infection. While linear differentiation of memory T cells was supported by adoptive transfer studies with T-cell receptor (TCR), transgenic (tg) CD8+ T cells specific for the male H-Y antigen, the most

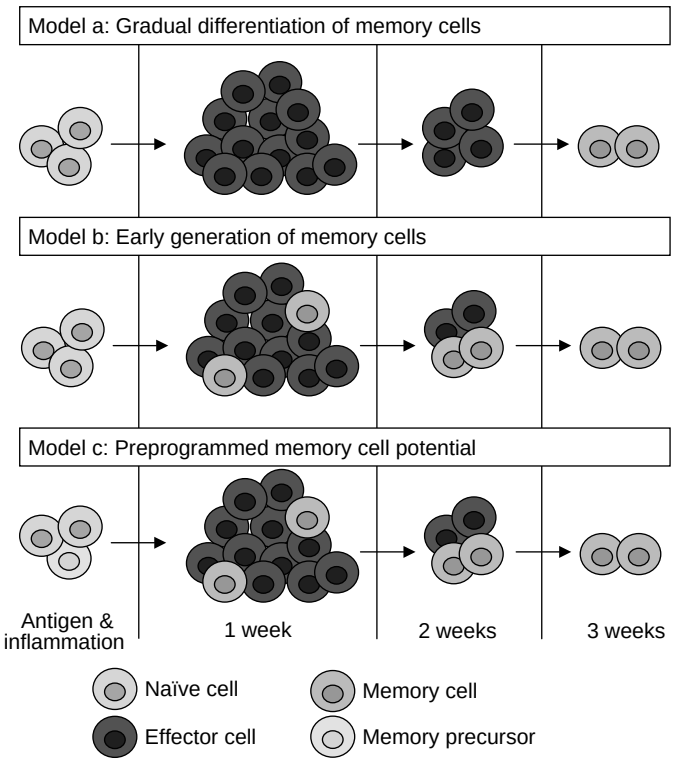

Fig. 1 Models of programmed memory T-cell generation.

direct evidence in support of this model came from studies using a clever system for genetically marking T cells that had acquired effector function. In this study, memory T cells were derived from effector T cells. It also remains controversial when memory cells form during the course of an immune response (see Fig. 1). One recent study demonstrated that T cells require several weeks from the time of antigen encounter to differentiate into memory T cells. Other studies showed that rechallenge of mice with a high inoculum of antigen 5 days after primary immunization resulted in a memory-like CD8+ T-cell response, suggesting that memory T cells may already be present during the primary immune response.

A variation of the linear differentiation model proposes that different memory T-cell subsets develop at different times following T-cell priming. According to this model, primed effector T cells can, in the absence of further stimuli, develop into CCR7+ CD62L+ central memory cells, which home to lymph nodes and lack effector functions. In contrast, effector memory cells, which are CCR7− CD62L−, derive from effector cells that have undergone more prolonged stimulation with antigen. This subset of memory cells produces high levels of effector cytokines, maintains cytolytic activity, and trafficks to peripheral tissues. Some evidence suggests that effector memory cells can convert to central memory cells upon antigen clearance.

Evidence in support of this model comes from the analysis of T-cell priming by live versus killed bacterial vaccines. Live bacteria, which induce substantial inflammation, result in large effector memory T-cell populations, while killed bacteria, which induce an attenuated inflammatory response, give rise to T cells with a central memory phenotype.

Another model for memory T-cell development is the *decreasing-potential model,* which is based on the progressive loss of proliferative capacity. The underlying idea of this model is that successful memory T-cell generation requires a short duration of antigenic exposure during T-cell priming. According to this model, larger numbers of functional memory T cells should be produced when effector T cells are briefly exposed to antigen, thereby avoiding antigen driven apoptosis (AICD). Support for this model comes from situations of overwhelming antigen dose or during chronic infections, such as HIV or hepatitis C, where antigen-specific T cells either disappear or become dysfunctional.

A variant of the decreasing-potential model suggests that the timing of T-cell priming during an immune response determines which cells enter the memory lineage. The circulation of lymphocytes between the blood and lymphatic system is a rather slow process with turnover times of 12 to 24 h. Thus, the time that it takes for individual lymphocytes to encounter antigen presented in secondary lymphoid organs can vary significantly. Since antigen presentation is transient, cells recruited later will encounter antigen for a shorter time period than cells that are initially recruited. Brief exposure to antigen would be sufficient to induce their proliferation and differentiation into effector cells; however, the signal would be inadequate to trigger death pathways. Further studies are required to provide experimental support for this model.

The third model for memory T-cell generation is the *instructive model*, which posits that effector and memory T-cell differentiation diverges during T-cell priming. According to this model, memory T cells directly differentiate from naïve T cells without passing through the effector stage. Although experimental support for this model is sparse, the notion that naïve T cells might take their cues from the inflammatory and costimulatory context and undergo differentiation into either effector or memory T cells remains plausible.

T-cell programming is a relatively new concept that should be incorporated into models of memory T-cell generation (see Fig. 2). Early studies demonstrated that CD8+ T cells specific for different epitopes undergo coordinate expansion

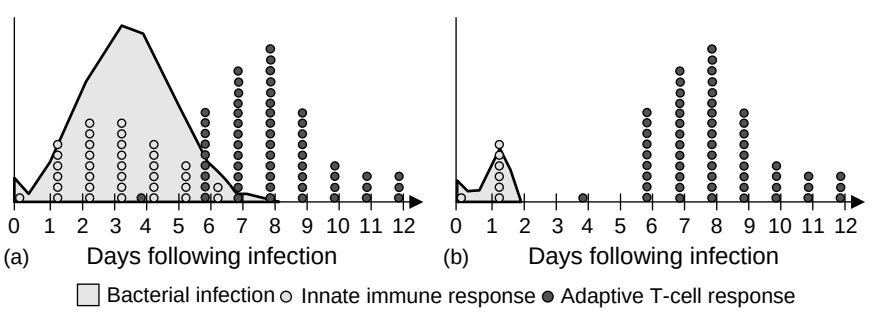

Fig. 2 Antigen-independent proliferation and memory generation.

and contraction in response to infection, despite substantial differences in antigen expression levels and stability of peptide-MHC complexes. One explanation for this finding is the relatively short duration of effective *in vivo* antigen presentation, which results from downregulation by the nascent T-cell response. While feedback inhibition limits T-cell priming to the very early phase of the infection, the brief period of priming is sufficient to initiate a program of antigen-independent T-cell expansion. Remarkably, the process of T-cell expansion, contraction, and memory formation occur in the absence of both antigen and inflammation.

4 CD4+ T-Cell Memory

While CD8+ T-cell memory has been extensively investigated, less is known about the generation and maintenance of memory CD4+ T cells. One explanation for this disparity is the difficulty to directly identify antigen-specific CD4+ T cells during the course of an immune response. One recent study, however, has compared CD4+ and CD8+ memory T-cell responses to LCMV and found that virus-specific memory CD4+ T cells decline in frequency over time while memory CD8+ T cells are maintained at high frequencies. In this section, we will concentrate on recent studies focusing on memory CD4+ T cells.

4.1 Differentiation of Effector and Memory CD4+ T Cells

Naïve CD4+ T cells, upon *in vivo* priming, differentiate into either Th1 or Th2 effector cells. While the mechanisms leading to Th1 or Th2 differentiation remain incompletely understood, the innate inflammatory response elicited by the invading pathogen plays a major role in this process. Many viruses and bacteria induce IL-12 and interferon-γ (IFN-γ) secretion, driving the differentiation of naïve CD4+ T cells into Th1 cells. In contrast, Th2 immunity, which is induced by intestinal helminth infection, is associated with innate inflammatory responses that produce IL-4, driving naïve CD4+ T-cell differentiation into Th2 cells. While the Th1/Th2 lineage decision has important consequences for the primary immune response, long-term memory T cells maintain the phenotype induced at the time of priming. Thus, while memory CD4 T cells do not produce cytokines in the absence of stimulation, upon reencountering antigen they rapidly reexpress the cytokines elicited during their primary activation. Several studies have demonstrated that epigenetic chromatin remodeling during the primary T-cell response accounts for specific cytokine expression profiles of reactivated memory T cells.

4.2 Phenotype of Memory CD4+ T Cells

Memory CD4+ T cells have a lower threshold for activation than naïve T cells. Indeed, memory CD4+ T cells respond more rapidly to antigen, with enhanced cytokine secretion and greater proliferation in response to low antigen doses. In addition to greater responsiveness to antigen, memory CD4+ T cells are also present at higher frequencies in immune animals. While cytokine production by naïve CD4+ T cells is restricted to IL-2 and IL-3, memory CD4+ T cells secrete a

greater variety of cytokines. Adoptive transfer of *in vitro* generated antigen-specific Th1 and Th2 cells leads to long-lived memory cells, which produce a pattern of cytokines closely related to that of the originally transferred cells. As mentioned in the previous section, epigenetic imprinting accounts for the fidelity of memory T-cell cytokine production.

4.3 Memory Generation and Maintenance

Mechanisms for CD4+ memory T-cell generation have not been defined. While models similar to those outlined for CD8+ T-cell memory formation may be applicable to CD4+ T-cell memory generation, existing data most strongly support linear differentiation from naïve to effector to memory CD4+ T cells. *In vitro* studies, for example, demonstrated that effector CD4+ T cells can differentiate into memory cells without the requirement for cell division. However, the factors involved in the transition from effector to memory T cells remain unclear. Remarkably, CD4+ T-cell memory is efficiently generated in the absence of IL-2, IL-4, or IFN-γ. As with CD8+ T cells, initial antigen stimulation is sufficient to initiate programmed expansion in CD4+ T cells, and extended antigenic stimulation is not required for differentiation into memory cells or for long-term persistence. Indeed, memory CD4+ T cells persist in the absence of both antigen and MHC class II molecules.

The cytokine requirements for long-term memory CD4+ T-cell survival are distinct from those described for memory CD8+ T cells. For example, while memory CD8+ T cells decrease in frequency in the absence of IL-15, memory CD4+ T cells are maintained independently of IL-15. Consistent with this finding, CD122 is expressed at low levels on memory CD4+ T cells. In addition, memory CD4+ T cells undergo homeostatic proliferation in the absence of both IL-7 and IL-15, and CD4+ T cells from mice deficient in the common cytokine receptor γ–chain, which is shared by receptors for IL-2, -4, -7, -9, and 15, can generate long-lived CD4+ memory T cells. IL-7 might not be essential for the homeostasis of CD4+ memory T cells, but contributes to the maintenance of all naïve and memory T-cell subsets, and thereby influences the control of the overall size of the T-cell pool. Differences in CD8+ and CD4+ T cell–memory maintenance have been detected following viral infection: virus-specific CD4+ memory T cells gradually decline, while CD8+ memory T cells remain stable. While many factors have been excluded as essential to CD4+ T cell–memory formation, the underlying mechanisms that drive differentiation of effector into memory T cells remains mysterious.

4.4 Trafficking of Memory CD4+ T Cells

CD8+ as well as CD4+ T cells differentiate in memory subsets with distinct effector functions and migration capacities. The differences in homing and trafficking of CD4+ T cells can be attributed to their expression of the selectin CD62L and the chemokine receptor CCR7. Central memory (CCR7+, CD62L+) CD4+ T cells are preferentially home to lymphoid tissues, whereas effector memory (CCR7−, CD62L−) recirculate in peripheral tissues long after immunization. The localization of central and effector memory CD4+ T cells in different organs has been directly visualized by immunohistology on whole body sections.

5
B-Cell Memory

5.1
Generation of B-Cell Memory

Similar to memory-T cells, memory-B cells respond rapidly to secondary exposure to antigen, proliferating and producing high affinity antibodies. B cells require accessory signals provided by helper T cells for full differentiation and are also activated by certain microbial stimuli. There are two distinct developmental pathways for antigen-specific B cells. B-1 and marginal-zone B cells proliferate and differentiate into antibody-secreting plasma cells, thereby providing the most rapid humoral response to antigen. While many plasma cells have a life span limited to a few days or weeks, some develop into long-lived cells that mediate long-term humoral immunity. Some antigen activated B cells migrate into follicles where they proliferate and form germinal centers (GCs), an important site of antibody secretion and affinity maturation. B cells expressing high affinity immunoglobulins receive positive signals from T-helper cells, leading to B-cell proliferation, survival, and maturation.

Lymphotoxin (LT) is important for development of peripheral lymphoid organs since mice lacking lymphotoxin-α (LT-α) receptor have no lymph nodes or Peyer's patches and fail to form germinal centers. These structural deficits are linked to defective isotype-switching following primary or secondary immunization with T cell–dependent antigens. The important role of germinal centers in memory B-cell formation was demonstrated in adoptive transfer experiments. B cells from LT-$\alpha-/-$ donors were capable of giving rise to a memory population, whereas a memory IgG B-cell responses could not be elicited in LT-$\alpha-/-$ hosts. These studies clearly support previous observations, which point out the important role of germinal center structures for memory B-cell development.

Some of the factors determining the lineage decision between plasma cells and memory B cells have been identified, and include cell-surface receptors, cytokines, and transcriptional regulators. Stimulation through CD40-CD154, IL-4, expression of PAX5 and BCL-6 inhibit plasma cell differentiation. On the other hand, absence of CD40, IL-3, IL-6, and IL-10 and the commensurate degradation of BCL-6 and induction of Blimp-1 (B-lymphocyte-induced maturation protein 1) promote plasma cell differentiation.

CD4+ T cells are crucial for the generation of long-term humoral immunity. Direct antigen-specific T-B-cell interactions have been visualized with adoptive transfer systems, and the molecular requirements for CD4+ T-cell help have been identified recently. One important participant in this process is the signaling lymphocyte activation molecule (SLAM), SLAM-associated protein (SAP), which was originally identified as altered in X-linked lymphoproliferative syndrome (XLP). SAP controls signaling via the family of SLAM surface receptors, including CD84, CD150 (SLAM), CD229, and CD244 and hence plays a fundamental role in T-cell function. Mice deficient in the SAP gene (SAP$-/-$) mount strong antibody responses during an acute viral infection, but they fail to generate long-lived virus-specific plasma and memory-B cells. Adoptive transfer experiments demonstrated that SAP expression by CD4+ T cells, but not B cells, is

essential for generation of long-term humoral immunity.

5.2 Maintenance of B-Cell Memory

It has been proposed that long-term humoral immunity is maintained by memory B cells, which continuously differentiate into short-lived plasma cells. Recently, however, it was shown that long-lived plasma cells survive for prolonged periods of time and maintain humoral immunity in the absence of memory B cells. Maintenance of humoral immunity appears to be antigen independent since memory B cells persist in the absence of antigen. Some experiments support the hypothesis that survival of long-lived plasma cells is determined by microenvironment rather than by intrinsic factors. *In vitro* experiments, for example, show that plasma cells isolated from tonsils undergo apoptosis unless rescued by stromal cells. Further support for this concept comes from the observation that plasma cell maintenance is restricted by the capacity of the splenic red pulp to provide "space," suggesting that plasma cell homeostasis is regulated by competition for specific survival niches.

Nerve growth factor (NGF) regulates neuronal cell development and survival. The finding that patients with autoimmune disorders often have elevated NGF plasma levels, and that NGF receptor shares structural similarities with some cytokine receptors suggested a role for this cytokine in the immune system. B cells can produce NGF under basal conditions and constitutively express NGF-receptors. It turns out that NGF acts as an autocrine survival factor for memory B cells and its neutralization completely inhibits humoral memory responses.

6 Conclusions

The journey to decipher memory T- and B-cell development continues to be a great scientific adventure, and there remain many unanswered questions that will require additional clever models and experiments. The motivation for moving forward, however, is substantial. Delineating basic mechanisms is important and applying the knowledge gained from an enhanced understanding of immunologic memory to the development of vaccines, against both infectious diseases and malignancy, will be especially satisfying.

See also Antigen Presenting Cells (APCs); Dendritic Cells; Immunology; Innate Immunity.

Bibliography

Books and Reviews

Ahmed, R., Gray, D. (1996) Immunological memory and protective immunity: understanding their relation, *Science* **272**, 54–60.

Bevan, M.J. (2004) Helping the CD8(+) T-cell response, *Nat. Rev. Immunol* **4**, 595–602.

Calame, K.L., Lin, K.I., Tunyaplin, C. (2003) Regulatory mechanisms that determine the development and function of plasma cells, *Annu. Rev. Immunol.* **21**, 205–230.

Janeway, C.A., Travers, P., Walport, M., Shlomchik, M. (2004) *Immunobiology*, 6th edition, New York and London.

Kaech, S.M., Wherry, E.J., Ahmed, R. (2002) Effector and memory T-cell differentiation: implications for vaccine development, *Nat. Rev. Immunol.* **2**, 251–262.

Ploss, A., Pamer, E.G. (2004) Memory, in: Kaufmann S.H.E. (Ed.) *Novel Vaccination Strategies*, 1st edition. WILEY-VCH, Weinheim, Germany, pp. 73–88.

Sallusto, F., Geginat, J., Lanzavecchia, A. (2004) Central memory and effector memory T

cell subsets: function, generation, and maintenance, *Annu. Rev. Immunol.* **22**, 765–787.

Schluns, K.S., Lefrancois, L. (2003) Cytokine control of memory T-cell development and survival, *Nat. Rev. Immunol.* **3**, 269–279.

Silverstein, A.M. (1989) *A History of Immunology*, 1st edition, Academic Press, Inc., San Diego, CA.

Primary Literature

Agarwal, S., Rao, A. (1998) Modulation of chromatin structure regulates cytokine gene expression during T cell differentiation, *Immunity* **9**, 765–775.

Appay, V., Dunbar, P.R., Callan, M., Klenerman, P., Gillespie, G.M., Papagno, L., Ogg, G.S., King, A., Lechner, F., Spina, C.A., et al. (2002) Memory CD8+ T cells vary in differentiation phenotype in different persistent virus infections, *Nat. Med.* **8**, 379–385.

Bachmann, M.F., Gallimore, A., Linkert, S., Cerundolo, V., Lanzavecchia, A., Kopf, M., Viola, A. (1999) Developmental regulation of Lck targeting to the CD8 coreceptor controls signaling in naive and memory T cells, *J. Exp. Med.* **189**, 1521–1530.

Badovinac, V.P., Porter, B.B., Harty, J.T. (2002) Programmed contraction of CD8(+) T cells after infection, *Nat. Immunol.* **3**, 619–626.

Becker, T.C., Wherry, E.J., Boone, D., Murali-Krishna, K., Antia, R., Ma, A., Ahmed, R. (2002) Interleukin 15 is required for proliferative renewal of virus-specific memory CD8 T cells, *J. Exp. Med.* **195**, 1541–1548.

Bennett, S.R., Carbone, F.R., Karamalis, F., Flavell, R.A., Miller, J.F., Heath, W.R. (1998) Help for cytotoxic-T-cell responses is mediated by CD40 signalling, *Nature* **393**, 478–480.

Bird, J.J., Brown, D.R., Mullen, A.C., Moskowitz, N.H., Mahowald, M.A., Sider, J.R., Gajewski, T.F., Wang, C.R., Reiner, S.L. (1998) Helper T cell differentiation is controlled by the cell cycle, *Immunity* **9**, 229–237.

Bourgeois, C., Rocha, B., Tanchot, C. (2002) A role for CD40 expression on CD8+ T cells in the generation of CD8+ T cell memory, *Science* **297**, 2060–2063.

Bradley, L.M., Duncan, D.D., Tonkonogy, S., Swain, S.L. (1991) Characterization of antigen-specific CD4+ effector T cells in vivo: immunization results in a transient population of MEL-14-, CD45RB-helper cells that secretes interleukin 2 (IL-2), IL-3, IL-4, and interferon gamma, *J. Exp. Med.* **174**, 547–559.

Bui, J.D., Calbo, S., Hayden-Martinez, K., Kane, L.P., Gardner, P., Hedrick, S.M. (2000) A role for CaMKII in T cell memory, *Cell* **100**, 457–467.

Busch, D.H., Kerksiek, K.M., Pamer, E.G. (2000) Differing roles of inflammation and antigen in T cell proliferation and memory generation, *J. Immunol.* **164**, 4063–4070.

Busch, D.H., Pilip, I., Pamer, E.G. (1998) Evolution of a complex T cell receptor repertoire during primary and recall bacterial infection, *J. Exp. Med.* **188**, 61–70.

Busch, D.H., Pilip, I.M., Vijh, S., Pamer, E.G. (1998) Coordinate regulation of complex T cell populations responding to bacterial infection, *Immunity* **8**, 353–362.

Chase, M.W. (1945) The Cellular transfer of cutaneous hypersensitivity to tuberculin, *Proc. Soc. Exp. Biol. Med.* **59**, 134–135.

Cho, B.K., Wang, C., Sugawa, S., Eisen, H.N., Chen, J. (1999) Functional differences between memory and naive CD8 T cells, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 2976–2981.

Crotty, S., Kersh, E.N., Cannons, J., Schwartzberg, P.L., Ahmed, R. (2003) SAP is required for generating long-term humoral immunity, *Nature* **421**, 282–287.

Garside, P., Ingulli, E., Merica, R.R., Johnson, J.G., Noelle, R.J., Jenkins, M.K. (1998) Visualization of specific B and T lymphocyte interactions in the lymph node, *Science* **281**, 96–99.

Gibson, T., Medawar, P.B. (1943) The fate of skin homografts in man, *J. Anat.* **77**, 299.

Goldrath, A.W., Bogatzki, L.Y., Bevan, M.J. (2000) Naive T cells transiently acquire a memory-like phenotype during homeostasis-driven proliferation, *J. Exp. Med.* **192**, 557–564.

Goldrath, A.W., Sivakumar, P.V., Glaccum, M., Kennedy, M.K., Bevan, M.J., Benoist, C., Mathis, D., Butz, E.A. (2002) Cytokine requirements for acute and basal homeostatic proliferation of naive and memory CD8+ T cells, *J. Exp. Med.* **195**, 1515–1522.

Gowans, J.L., Knight, E.J. (1964) The route of recirculation of lymphocytes in the rat, *Proc. R. Soc. Lond. Ser. B. Biol. Sci.* **159**, 257–282.

Hamilton, S.E., Porter, B.B., Messingham, K.A., Badovinac, V.P., Harty, J.T. (2004) MHC class Ia-restricted memory T cells inhibit expansion

of a nonprotective MHC class Ib (H2-M3)-restricted memory response, *Nat. Immunol.* **5**, 159–168.

Harrington, L.E., Galvan, M., Baum, L.G., Altman, J.D., Ahmed, R. (2000) Differentiating between memory and effector CD8 T cells by altered expression of cell surface O-glycans, *J. Exp. Med.* **191**, 1241–1246.

Hasegawa, K., Martin, F., Huang, G., Tumas, D., Diehl, L., Chan, A.C. (2004) PEST domain-enriched tyrosine phosphatase (PEP) regulation of effector/memory T cells, *Science* **303**, 685–689.

Hendriks, J., Gravestein, L.A., Tesselaar, K., van Lier, R.A., Schumacher, T.N., Borst, J. (2000) CD27 is required for generation and long-term maintenance of T cell immunity, *Nat. Immunol.* **1**, 433–440.

Homann, D., Teyton, L., Oldstone, M.B. (2001) Differential regulation of antiviral T-cell immunity results in stable CD8+ but declining CD4+ T-cell memory, *Nat. Med.* **7**, 913–919.

Hu, H., Huston, G., Duso, D., Lepak, N., Roman, E., Swain, S.L. (2001) CD4(+) T cell effectors can become memory cells with high efficiency and without further division, *Nat. Immunol.* **2**, 705–710.

Huster, K.M., Busch, V., Schiemann, M., Linkemann, K., Kerksiek, K.M., Wagner, H., Busch, D.H. (2004) Selective expression of IL-7 receptor on memory T cells identifies early CD40L-dependent generation of distinct CD8+ memory T cell subsets, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 5610–5615.

Jacob, J., Baltimore, D. (1999) Modelling T-cell memory by genetic marking of memory T cells in vivo, *Nature* **399**, 593–597.

Janssen, E.M., Lemmens, E.E., Wolfe, T., Christen, U., Von Herrath, M.G., Schoenberger, S.P. (2003) CD4(+) T cells are required for secondary expansion and memory in CD8(+) T lymphocytes, *Nature* **421**, 852–856.

Jenner, E. (1798) *An Inquiry into the Causes and Effects of the Variolæ Vaccinæ, a Disease Discovered in Some of the Western Counties of England, Particularly Gloucestershire, and Known by the Name of the Cow Pox*, 1st edition, Low, Sampson, London, UK.

Kaech, S.M., Ahmed, R. (2001) Memory CD8+ T cell differentiation: initial antigen encounter triggers a developmental program in naive cells, *Nat. Immunol.* **2**, 415–422.

Kaech, S.M., Hemby, S., Kersh, E., Ahmed, R. (2002) Molecular and functional profiling of memory CD8 T cell differentiation, *Cell* **111**, 837–851.

Kaech, S.M., Tan, J.T., Wherry, E.J., Konieczny, B.T., Surh, C.D., Ahmed, R. (2003) Selective expression of the interleukin 7 receptor identifies effector CD8 T cells that give rise to long-lived memory cells, *Nat. Immunol.* **4**, 1191–1198.

Kassiotis, G., Garcia, S., Simpson, E., Stockinger, B. (2002) Impairment of immunological memory in the absence of MHC despite survival of memory T cells, *Nat. Immunol.* **3**, 244–250.

Kaufmann, S.H., Simon, M.M., Hahn, H. (1979) Specific Lyt 123 cells are involved in protection against Listeria monocytogenes and in delayed-type hypersensitivity to listerial antigens, *J. Exp. Med.* **150**, 1033–1038.

Kedl, R.M., Mescher, M.F. (1998) Qualitative differences between naive and memory T cells make a major contribution to the more rapid and efficient memory CD8+ T cell response, *J. Immunol.* **161**, 674–683.

Kerksiek, K.M., Busch, D.H., Pilip, I.M., Allen, S.E., Pamer, E.G. (1999) H2-M3-restricted T cells in bacterial infection: rapid primary but diminished memory responses, *J. Exp. Med.* **190**, 195–204.

Kieper, W.C., Tan, J.T., Bondi-Boyd, B., Gapin, L., Sprent, J., Ceredig, R., Surh, C.D. (2002) Overexpression of interleukin (IL)-7 leads to IL-15-independent generation of memory phenotype CD8+ T cells. *J. Exp. Med.*, **195**, 1533–1539.

Kitasato, S. (1977) *Collected Papers of Shibasaburo Kitasato*, Kitasato Institute, Tokyo, Japan.

Koch, R. (1878) *Untersuchungen über die Aetiologie der Wundinfectionskrankheiten*, F.C.W. Vogel, Leipzig, Germany.

Ku, C.C., Murakami, M., Sakamoto, A., Kappler, J., Marrack, P. (2000) Control of homeostasis of CD8+ memory T cells by opposing cytokines, *Science* **288**, 675–678.

Kursar, M., Bonhagen, K., Fensterle, J., Kohler, A., Hurwitz, R., Kamradt, T., Kaufmann, S.H., Mittrucker, H.W. (2002) Regulatory CD4+ CD25+ T cells restrict memory CD8+ T cell responses, *J. Exp. Med.* **196**, 1585–1592.

Landsteiner, K., Chase, M.W. (1942) Experiments on transfer of cutaneous sensitivity to simple compounds, *Proc. Soc. Exp. Biol. Med.* **49**, 688–689.

Lane, F.C., Unanue, E.R. (1972) Requirement of thymus (T) lymphocytes for resistance to listeriosis, *J. Exp. Med.* **135**, 1104–1112.

Lantz, O., Grandjean, I., Matzinger, P., Di Santo, J.P. (2000) Gamma chain required for naive CD4+ T cell survival but not for antigen proliferation, *Nat. Immunol.* **1**, 54–58.

Lauvau, G., Vijh, S., Kong, P., Horng, T., Kerksiek, K., Serbina, N., Tuma, R.A., Pamer, E.G. (2001) Priming of memory but not effector CD8 T cells by a killed bacterial vaccine, *Science* **294**, 1735–1739.

Leishman, A.J., Naidenko, O.V., Attinger, A., Koning, F., Lena, C.J., Xiong, Y., Chang, H.C., Reinherz, E., Kronenberg, M., Cheroutre, H. (2001) T cell responses modulated through interaction between CD8alphaalpha and the nonclassical MHC class I molecule, TL, *Science* **294**, 1936–1939.

Mackaness, G.B., Blanden, R.V. (1967) Cellular immunity, *Prog. Allergy* **11**, 89–140.

Madakamutil, L.T., Christen, U., Lena, C.J., Wang-Zhu, Y., Attinger, A., Sundarrajan, M., Ellmeier, W., von Herrath, M.G., Jensen, P., Littman, D.R., Cheroutre, H. (2004) CD8alpha alpha-mediated survival and differentiation of CD8 memory T cell precursors, *Science* **304**, 590–593.

Marchesi, V.T., Gowans, J.L. (1964) The migration of lymphocytes through the endothelium of venules in the lymph nodes: an electron microscope study, *Proc. R. Soc. Lond. Ser. B. Biol. Sci.* **159**, 283–290.

Maruyama, M., Lam, K.P., Rajewsky, K. (2000) Memory B-cell persistence is independent of persisting immunizing antigen, *Nature* **407**, 636–642.

Masopust, D., Vezys, V., Marzo, A.L., Lefrancois, L. (2001) Preferential localization of effector memory cells in nonlymphoid tissue, *Science* **291**, 2413–2417.

Matsumoto, M., Mariathasan, S., Nahm, M.H., Baranyay, F., Peschon, J.J., Chaplin, D.D. (1996) Role of lymphotoxin and the type I TNF receptor in the formation of germinal centers, *Science* **271**, 1289–1291.

Medawar, P.B. (1944) The behaviour and fate of skin autografts and skin homografts in rabbits, *J. Anat.* **78**, 176.

Medawar, P.B. (1945) A second study of the behaviour and fate of skin homografts in rabbits, *J. Anat.* **79**, 157.

Mercado, R., Vijh, S., Allen, S.E., Kerksiek, K., Pilip, I.M., Pamer, E.G. (2000) Early programming of T cell populations responding to bacterial infection, *J. Immunol.* **165**, 6833–6839.

Merville, P., Dechanet, J., Desmouliere, A., Durand, I., de Bouteiller, O., Garrone, P., Banchereau, J., Liu, Y.J. (1996) Bcl-2+ tonsillar plasma cells are rescued from apoptosis by bone marrow fibroblasts, *J. Exp. Med.* **183**, 227–236.

Metchnikoff, E. (1905) *Immunity in Infective Diseases*, Cambridge University Press, Cambridge, UK.

Murali-Krishna, K., Lau, L.L., Sambhara, S., Lemonnier, F., Altman, J., Ahmed, R. (1999) Persistence of memory CD8 T cells in MHC class I-deficient mice, *Science* **286**, 1377–1381.

Murphy, J.B., Ellis, A.W.M. (1914) Experiments on the role of lymphoid tissue in the resistance to experimental tuberculosis in mice, *J. Exp. Med.* **20**, 397–403.

North, R.J. (1973) Importance of thymus-derived lymphocytes in cell-mediated immunity to infection, *Cell. Immunol.* **7**, 166–176.

Opferman, J.T., Ober, B.T., Ashton-Rickardt, P.G. (1999) Linear differentiation of cytotoxic effectors into memory T lymphocytes, *Science* **283**, 1745–1748.

Pasteur, L. (1880) *De l'attenuation du virus du cholera des poules, Vol. 91*, Comptes rendus de l'Academie des sciences, Paris, France.

Pihlgren, M., Dubois, P.M., Tomkowiak, M., Sjogren, T., Marvel, J. (1996) Resting memory CD8+ T cells are hyperreactive to antigenic challenge in vitro, *J. Exp. Med.* **184**, 2141–2151.

Reinhardt, R.L., Khoruts, A., Merica, R., Zell, T., Jenkins, M.K. (2001) Visualizing the generation of memory CD4 T cells in the whole body, *Nature* **410**, 101–105.

Ridge, J.P., Di Rosa, F., Matzinger, P. (1998) A conditioned dendritic cell can be a temporal bridge between a CD4+ T-helper and a T-killer cell, *Nature* **393**, 474–478.

Sallusto, F., Lenig, D., Forster, R., Lipp, M., Lanzavecchia, A. (1999) Two subsets of memory T lymphocytes with distinct homing potentials and effector functions, *Nature* **401**, 708–712.

Schoenberger, S.P., Toes, R.E., van der Voort, E.I., Offringa, R., Melief, C.J. (1998) T-cell help for cytotoxic T lymphocytes is mediated by CD40-CD40L interactions, *Nature* **393**, 480–483.

Seddon, B., Tomlinson, P., Zamoyska, R. (2003) Interleukin 7 and T cell receptor signals regulate homeostasis of CD4 memory cells, *Nat. Immunol.* **4**, 680–686.

Shedlock, D.J., Shen, H. (2003) Requirement for CD4 T cell help in generating functional CD8 T cell memory, *Science* **300**, 337–339.

Slifka, M.K., Antia, R., Whitmire, J.K., Ahmed, R. (1998) Humoral immunity due to long-lived plasma cells, *Immunity* **8**, 363–372.

Sourdive, D.J., Murali-Krishna, K., Altman, J.D., Zajac, A.J., Whitmire, J.K., Pannetier, C., Kourilsky, P., Evavold, B., Sette, A., Ahmed, R. (1998) Conserved T cell receptor repertoire in primary and memory CD8 T cell responses to an acute viral infection, *J. Exp. Med.* **188**, 71–82.

Sun, J.C., Bevan, M.J. (2003) Defective CD8 T cell memory following acute infection without CD4 T cell help, *Science* **300**, 339–342.

Sun, J.C., Williams, M.A., Bevan, M.J. (2004) CD4(+) T cells are required for the maintenance, not programming, of memory CD8(+) T cells after acute infection, *Nat. Immunol.* **5**, 927–933.

Suvas, S., Kumaraguru, U., Pack, C.D., Lee, S., Rouse, B.T. (2003) CD4+CD25+ T cells regulate virus-specific primary and memory CD8+ T cell responses, *J. Exp. Med.* **198**, 889–901.

Swain, S.L. (1994) Generation and in vivo persistence of polarized Th1 and Th2 memory cells, *Immunity* **1**, 543–552.

Swain, S.L., Hu, H., Huston, G. (1999) Class II-independent generation of CD4 memory T cells from effectors, *Science* **286**, 1381–1383.

Swanson, B.J., Murakami, M., Mitchell, T.C., Kappler, J., Marrack, P. (2002) RANTES production by memory phenotype-T cells is controlled by a posttranscriptional, TCR-dependent process, *Immunity* **17**, 605–615.

Sze, D.M., Toellner, K.M., Garcia de Vinuesa, C., Taylor, D.R., MacLennan, I.C. (2000) Intrinsic constraint on plasmablast growth and extrinsic limits of plasma cell survival. *J. Exp. Med.*, **192**, 813–821.

Tan, J.T., Ernst, B., Kieper, W.C., LeRoy, E., Sprent, J., Surh, C.D. (2002) Interleukin (IL)-15 and IL-7 jointly regulate homeostatic proliferation of memory phenotype-CD8+ cells but are not required for memory phenotype-CD4+ cells, *J. Exp. Med.* **195**, 1523–1532.

Teague, T.K., Hildeman, D., Kedl, R.M., Mitchell, T., Rees, W., Schaefer, B.C., Bender, J., Kappler, J., Marrack, P. (1999) Activation changes the spectrum but not the diversity of genes expressed by T cells, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 12691–12696.

Torcia, M., Bracci-Laudiero, L., Lucibello, M., Nencioni, L., Labardi, D., Rubartelli, A., Cozzolino, F., Aloe, L., Garaci, E. (1996) Nerve growth factor is an autocrine survival factor for memory B lymphocytes, *Cell* **85**, 345–356.

Urdahl, K.B., Sun, J.C., Bevan, M.J. (2002) Positive selection of MHC class Ib-restricted CD8(+) T cells on hematopoietic cells, *Nat. Immunol.* **3**, 772–779.

Veiga-Fernandes, H., Rocha, B. (2004) High expression of active CDK6 in the cytoplasm of CD8 memory cells favors rapid division, *Nat. Immunol.* **5**, 31–37.

Veiga-Fernandes, H., Walter, U., Bourgeois, C., McLean, A., Rocha, B. (2000) Response of naive and memory CD8+ T cells to antigen stimulation in vivo, *Nat. Immunol.* **1**, 47–53.

Wherry, E.J., Teichgraber, V., Becker, T.C., Masopust, D., Kaech, S.M., Antia, R., Von Andrian, U.H., Ahmed, R. (2003) Lineage relationship and protective immunity of memory CD8 T cell subsets, *Nat. Immunol.* **4**, 225–234.

Wong, P., Pamer, E.G. (2001) Cutting edge: antigen-independent CD8 T cell proliferation, *J. Immunol.* **166**, 5864–5868.

Wong, P., Pamer, E.G. (2003) Feedback regulation of pathogen-specific T cell priming, *Immunity* **18**, 499–511.

Wong, P., Lara-Tejero, M., Ploss, A., Leiner, I., Pamer, E.G. (2004) Rapid development of T cell memory, *J. Immunol.* **172**, 7239–7245.

Yang, Y., Chang, J.F., Parnes, J.R., Fathman, C.G. (1998) T cell receptor (TCR) engagement leads to activation-induced splicing of tumor necrosis factor (TNF) nuclear pre-mRNA, *J. Exp. Med.* **188**, 247–254.

Zhang, X., Sun, S., Hwang, I., Tough, D.F., Sprent, J. (1998) Potent and selective stimulation of memory-phenotype CD8+ T cells in vivo by IL-15, *Immunity* **8**, 591–599.

Zimmermann, C., Prevost-Blondel, A., Blaser, C., Pircher, H. (1999) Kinetics of the response of naive and memory CD8 T cells to antigen: similarities and differences. *Eur. J. Immunol.*, **29**, 284–290.

Immunology of Cancer

Peter R. Holman and Edward D. Ball
University of California, San Diego, La Jolla, CA, USA

1 Tumor Antigens 431

2 Antigen Presentation 432

3 Immune Response to Tumor-associated Antigens 435
3.1 Adaptive Immunity 435
3.1.1 Control of T-cell Proliferation 436
3.1.2 Innate Immunity 437

4 Immunotherapy 439
4.1 Passive Immunotherapy 439
4.2 Active Immunotherapy 441
4.3 *Ex vivo* Generation of CTL 443
4.4 *Ex vivo* Generation of Dendritic Cells (DC) 443
4.5 *Ex vivo* Generation of NK Cells 444
4.6 Summary and Conclusions 444

Bibliography 444
Books and Reviews 444
Primary Literature 445

Keywords

Antigen-presenting Cells (APC)
Cells such as dendritic cells or macrophages that are capable of presenting peptide antigens to T cells for immune recognition and activation.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

B-cell Receptor (BCR)
The surface immunoglobulin expressed on the surface of each B cell that uniquely recognizes an antigenic determinant.

Cluster Determinant (CD)
An internationally recognized nomenclature that identifies and classifies cell surface antigens on leukocytes that are recognized by monoclonal antibodies

Cytotoxic T Cells (CTL)
T lymphocytes that are capable of killing target cells, either antigen-specific or polyclonal.

Major Histocompatibility Complex (MHC)
In humans, referred to as human leukocyte antigen (HLA).

Monoclonal Antibodies (mAb)
Antibodies captured by cell fusion technology or recombinant DNA techniques that are produced in large quantities for diagnostic and therapeutic use.

Natural Killer (NK) Cells
Cells of the hematopoietic system that are neither lymphocytes nor classical myeloid cells that are capable of mediating cytotoxicity against a variety of tumor cells.

T-cell Receptor (TCR)
A dimeric cell surface protein on the surface of T cells that recognizes antigens expressed on antigen-presenting cells.

Tumor-associated Antigens (TAA)
Antigens uniquely expressed or overexpressed on tumor cells.

Cancer or tumor immunology refers to the field of study that seeks to define the role of the immune system in controlling and regulating the growth and survival of malignant cells. From animal studies, it has been known for decades that tumor cells can be recognized and eliminated by the immune system. The basis of immune recognition is the expression of tumor-associated or tumor-specific molecules by tumor cells in a manner that is recognizable to the various components of the immune system. Much has been learned about the antigens, antigen-presenting cells, and effector cells of the immune system over the past several years. This is still a discipline in its infancy, which is yet to fulfill the dream of effective control of human tumors by natural immunological processes. This chapter will review the current state of the science and the prospects of meaningful application of the immune system to human disease.

1 Tumor Antigens

The fundamental notion of tumor immunology is that the host immune system has the capacity to recognize antigens on tumor cell surfaces that are different from normal antigens, to which immune tolerance has already developed. The evidence for the existence of tumor antigens comes from decades of study in animal models. For example, classical experiments in the 1950s demonstrated that animals "immunized" with irradiated syngeneic tumor cells protected the host from subsequent challenge with viable tumor cells.

The molecular basis of tumor-associated antigens (TAA) has been elucidated in recent years, in concert with an unraveling of the mechanism by which antigens are perceived by the immune system. It is now known that polypeptide antigens are processed and presented as small peptides in the groove of class I and II HLA molecules and thus recognized by specific T cells through the T-cell receptor (TCR).

There are several mechanisms by which tumor antigens may arise (Table 1). It is increasingly clear that cancer is a genetic disease, resulting from various mutagenic events in the DNA of the subject. TAA may be generated through mutations in the genome of the cancer cell mediated by carcinogenic chemicals, leading to the generation of new protein molecules that are different from normal self-antigens. Point mutations, chromosomal translocations resulting in fusion genes and their encoded molecules, breaks in DNA strands, and numerous other molecular changes in DNA lead to altered protein structures in cancer cells.

Other candidates for tumor antigens include tissue-specific antigens that are overexpressed on tumor cells as a consequence of the neoplastic transformation. It would not be expected that these "normal" self-antigens would generate an immune response. However, as we will discuss below, these overexpressed antigens may be targets for targeted immunotherapy with monoclonal antibodies or cellular effectors.

Tab. 1 Tumor-associated antigens.

Origin of TAA
• Chemical mutagenesis
• Spontaneous mutation with mismatch repair defects
• Viral mutagenesis
• Chromosomal translocations (bcr-abl, AML/ETO1)
• Altered glycosylation (MUC-1)
• Idiotypic epitopes on B-cell malignancy (CLL, NHL, MM)
• Overexpression of normal tissue-specific antigens on transformed tissue
Classes of tumor-associated antigens
• Cancer testis antigens
• Differentiation antigens
• Viral antigens
• Mutated tumor suppressor antigens
• Oncogene products
• Unique antigens

Tumor antigens in human solid tumors were first described in malignant melanoma and prostate carcinoma. Through the creation of cytotoxic T-lymphocyte cell lines from patients with malignant melanoma, antigens such as MAGE-1, MART-1/Melan-A and NY-ESO-1 were found to be expressed on cells from these cancers. These peptide antigens are also expressed on normal testis tissue, leading to the classification of such antigens as ''cancer testis antigens.'' Since then, literally hundreds of peptide antigens and their HLA restricting alleles (both class I and II) have been described through similar techniques. Recently employed technologies including microarray, subtractive hybridization, and representational-difference analysis have revealed many more candidate human TAA.

Despite the theoretical and demonstrated existence of tumor antigens in human tumors, clearly the immune system in man is quite ineffectual in recognizing these antigens since cancer continues to spontaneously emerge at an alarming frequency. This sobering fact leads to the conclusion that tumor cells may be immunosuppressive to the immune system, poor ''presenters'' of their altered self-antigens, or that they can elude the immune system by other mechanisms such as continuous mutation or rapid growth kinetics. There are also examples of humoral (antibody) responses to TAA. In this instance, a cell surface or circulating antigen is recognized by the B-cell antigen receptor (BCR) leading to its internalization, induction of antibody synthesis, and processing of the antigen for expression by class II HLA molecules and subsequent presentation to $CD4^+$ T cells. Thus, both a cellular and humoral response may be generated to the same TAA.

2
Antigen Presentation

The immune system has evolved with 2 major arms, the innate arm and the adaptive arm. Innate immunity recognizes nonself-antigens in a nonspecific fashion. Each time a similar antigen is encountered, a similar response pattern is initiated. No memory component is involved. Components of the adaptive system are also activated as a result. Macrophages, interdigitating dendritic cells, and natural killer (NK) cells are important cellular components of the innate immune system. Various cytokines, complement and acute phase proteins also play key roles.

The adaptive immune system includes both humoral and cellular arms. It responds primarily to exogenous antigens that are first processed by cells, and then presented in the context of self-HLA to specialized receptors on B and T lymphocytes. Many animal and human studies support the concept that the cellular arm has a greater role in both tumor surveillance and tumor-directed immune responses. Observations include the efficacy of high doses of interleukin-2 (IL-2) in some patients with malignant melanoma and renal cell cancer and also studies demonstrating that rejection of transplantable tumors was mediated primarily by components of the cellular immune system.

For a TAA to induce an effective cellular immune response, antigen has to be presented directly or via an antigen-presenting cell (APC) to a lymphocyte bearing a T-cell receptor (TCR) that recognizes epitopes on the presented antigen. Cross linking of the TCR can then occur but additional cell–cell interactions are critical for T-cell activation. These involve essential costimulatory molecules as shown in Fig. 1. If, following

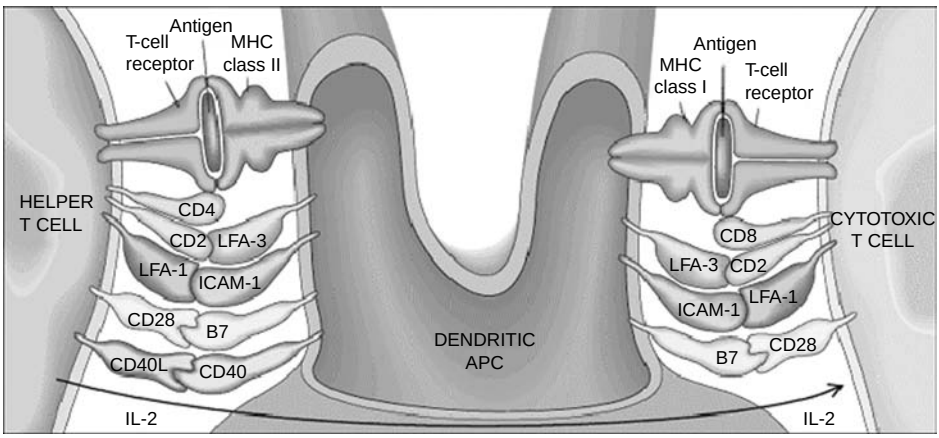

Fig. 1 Antigen presentation by APC through class I and II HLA. Helper T cells recognize MHC class II-associated antigens. Cytotoxic T cells recognize MHC class I-associated antigens. Release of interleukin-2 (IL-2) by the helper T cell provides a so-called help signal to the cytotoxic T cells. (ICAM-intercellular adhesion molecule; LFA-leukocyte function–associated antigen). Reprinted with permission from "Tumor Immunology" by Robert F. Todd, III, M.D., Ph.D., and Bruce O. Redman, D.O., Section 12, Oncology, and Section 6, Immunology/Allergy, published in ACP Medicine, edited by Dale DC and Federman DD, published by WebMD Corporation, New York, 2004. URL of product www.acpmedicine.com.

engagement of the TCR, T-cell activation does not occur, tolerance develops.

Tumor antigens have an inherent disadvantage in their ability to induce an immune response, as they are often 'self-antigens' to which tolerance frequently exists. Many tumor antigens have demonstrated the ability to induce an immune response but generally they do so only inefficiently. The goal of therapeutic immunological approaches is to augment this response and overcome immunological tolerance and ignorance.

An essential quality of the adaptive immune system is the ability to distinguish self from nonself. This is achieved via the induction of tolerance to self. This process occurs primarily in the thymus and results in a state of immunological unresponsiveness. It results primarily from the interplay of an APC presenting self-antigen to T lymphocytes in the context of self-Major Histocompatibility Complex (MHC) molecules. Tolerance can also develop at peripheral sites such as the lymph nodes and spleen.

The MHC is the key to understanding tolerance. It is composed of the Human Leukocyte Antigen (HLA) system and consists of greater than 200 genes residing on chromosome 6 (see Fig. 2). Among these, the genes involved with immunity are divided into two classes, HLA class I and class II. The class I genes important for the immune response are the HLA-A, -B and -C genes. They encode for the α chain of their corresponding HLA molecule. The β chain is identical for all class I molecules and is $\beta 2$ microglobulin, encoded from a gene on chromosome 15. Both chains of all class II molecules are encoded by genes on chromosome 6. The class II HLA molecules belong to the D class and include 5 different families (M, O, P, Q, or R). Genes on chromosome 6 also encode for HLA

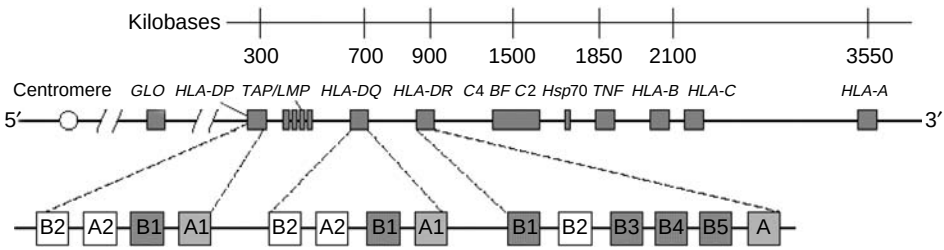

Fig. 2 Organization of *HLA* genes on human chromosome 6. The best-characterized loci of the human major histocompatibility complex (MHC), located in the *HLA* region of the short arm of chromosome 6, are depicted. Distances are shown in recombination units (centimorgans), as determined by crossover frequencies in family studies, and in kilobases, as determined by sequence analysis of fragments produced by DNases having defined cleavage sites. MHC class II molecules are encoded in the *HLA-DP*, *HLA-DQ*, and *HLA-DR* genes, and MHC class I molecules are encoded by *HLA-B*, *HLA-C*, and *HLA-A* genes. A cluster of closely linked complement genes-*C4*, *BF*, and *C2*-lies in the center of the region. There are two structural genes for *C4*, interspersed with two genes for the adrenal enzyme 21-hydroxylase. Next is the heat shock protein gene, *Hsp70*, followed by the tumor necrosis factor (TNF) genes, *A* and *B*. The orientation of the complement cluster and the *TNF* cluster has not been established, but an expanded view of this area could be depicted as-(*C4–210HA-C4B-210HB-BF-C2*)-(*HSP70*)-(*TNFA-TNFB*)-. *GLO* is a marker gene for the enzyme glyoxylase. An expansion of the class II region is in the lower portion of the figure. Each class II molecule is a heterodimer of an a and a ß chain, which are encoded in the *A* and *B* genes, respectively. Pseudogenes, which are not expressed on the cell surface, are shown in white boxes. *HLA-DP* and *HLA-DQ* have one expressed heterodimer, A1B1; *HLA-DR* has only one A chain but nine genes for B chains (four are shown in the figure). The principal expressed heterodimers for *HLA-DR* are AB1, AB3, AB4, and AB5. In the region between *HLA-DP* and *HLA-DQ* lie the closely linked *TAP1*, *TAP2*, *LMP2*, and *LMP7* genes. The *TAP* genes encode peptide transporters, whereas the *LMPf* genes encode proteasomes that fragment proteins into peptides. This cytoplasmic system is believed to be responsible for production and delivery of peptides to MHC class I molecules before their movement to the cell surface.
Reprinted with permission from "Adaptive Immunity: Histocompatibility Antigens and Immune Response Genes" by Edgar Louis Milford, M.D., and Charles B. Carpenter, M.D., Section 12, Oncology, and Section 6, Immunology/Allergy, published in ACP Medicine, edited by Dale DC and Federman DD, published by WebMD Corporation, New York, 2004. URL of product www.acpmedicine.com.

class III molecules. Both class I and II HLA molecules function as antigen-presentation molecules. They have a three-dimensional structure forming a groove on their surface in which a peptide can bind. It is this peptide–HLA complex that is recognizable to the TCR.

Peptide presentation in the context of HLA molecules allows the essential differentiation of self from nonself. HLA class I molecules generally present endogenous peptides derived from normal cellular housekeeping and turnover. They are present on most nucleated cells and are the predominant mechanism of TAA presentation. Class II molecules present peptides derived from exogenous sources. Their distribution is more limited, being present on phagocytic cells including dendritic cells and macrophages, B cells, activated T cells, and thymic epithelial cells.

Tolerance develops as a consequence of the deletion of self-reacting lymphocytes during maturation in the thymus. As TCR gene rearrangement is random and results in a vast repertoire, it is essential that those T cells whose receptors recognize self-antigen be inactivated. This process continues throughout life. Lymphocytes originate in the bone marrow. After migration to the thymus, Tlymphocytes whose receptors recognize self-MHC molecules are positively selected. If cells are not positively selected, as is the case for the majority of T cells, they undergo apoptosis. If a TCR recognizes self-HLA class II molecules presenting self-peptide, they also undergo apoptosis. This process is called negative selection. Tolerance developing via this mechanism requires antigens to be presented in the thymus. Many antigens however, including most tumor antigens, are not expressed in the thymus. Self-reactive T cells have been shown to exist but usually will not be immunologically reactive. These cells are not tolerant but are anergic, demonstrating immunologic ignorance. Anergy may be determined by the lack of a sufficient antigen dose, lack of adequate costimulation or by physical separation across tissue or anatomical barriers.

Antigen presented in the context of HLA class I is presented primarily to $CD8^+$ lymphocytes that are often directly cytotoxic. Antigen bound to HLA class II is preferentially presented to $CD4^+$ lymphocytes that generally are helper T cells. Following activation, secretion of cytokines results in the recruitment of additional immunological effector cells. The cellular distribution of HLA class II molecules is more limited, being primarily expressed on cells that function as professional APCs and also B lymphocytes.

3 Immune Response to Tumor-associated Antigens

As discussed above, immune effector cells capable of mediating cancer cell death fall into two broad categories. There are cells of the adaptive immune system, capable of responding in an antigen-specific manner, such as T and B lymphocytes. In addition, there are the cells of the innate immune system that recognize tumor cells in an antigen-nonspecific manner, such as natural killer (NK) cells, macrophages, and monocytes. Unfortunately, it appears that these responses are usually thwarted by a variety of means relating to properties of tumor cells that allow them to evade detection by the immune system. The immune response to tumors will be discussed below.

3.1 Adaptive Immunity

In order for a tumor antigen to incite an immune response, a carefully orchestrated and regulated sequence of events is required. Following antigen presentation, its recognition by a TCR is required. Subsequently, T-cell activation is necessary for subsequent effector function. If recognition between the TCR and the HLA–peptide complex alone occurs, without the involvement of additional costimulatory signals, the T cell will not be activated and a state of tolerance will develop. The TCR binding to the HLA–peptide complex is a relatively weak interaction and additional cell-to-cell interactions need to occur, strengthening the bond. Various adhesion molecules including LFA-1, CD2 and CD5 expressed on the T cells can bind to corresponding ligands expressed on the APC. Some of these adhesion molecules

may also be involved in cell signaling (see Fig. 1). A critical signal for T-cell activation is provided by the B7 family of ligands (CD80 and CD86) expressed on the surface of APC interacting with CD28 on the surface of T lymphocytes. Similarly, CD154 (CD40 ligand) interacts with CD40 on the surface of APCs. Additional molecular interactions between the lymphocyte and the APC also occur. Some of these are stimulatory and others are inhibitory. CTLA-4 is a structural homolog of CD28, also expressed on T cells, and its ligation by B7 expressed on APC's results in lymphocyte downregulation. Specialized APCs such as dendritic cells tend to express high concentrations of the costimulatory molecules, enhancing their ability to initiate an efficient immune response. As a consequence of these additional cell-to-cell interactions, the T cell becomes activated. IL-2 production is upregulated, resulting in further T-cell recruitment and cell survival necessary for an effective immune response.

The HLA class II molecules that present exogenous protein are synthesized in the endoplasmic reticulum. From there, they travel to the cytoplasm in vesicles that fuse with endosomes carrying exogenous peptides. The peptides are loaded into the peptide groove and the complex is transported to the cell surface for presentation to $CD4^+$ T cells (T helper cells). As most tumor antigens are self-antigens, they are primarily presented via the MHC class I pathway to $CD8^+$ cytotoxic lymphocytes. Endogenous proteins are processed by housekeeping structures called proteasomes. The resulting antigenic peptides are delivered to the endoplasmic reticulum by specialized molecules called *transporters associated with antigen processing* or *TAP* proteins. Within the endoplasmic reticulum, the class I HLA molecules bind the appropriate peptide before translocating to the cell surface for presentation to $CD8^+$ lymphocytes.

Upon recognition of antigen, cytotoxic $CD8^+$ T cells are stimulated to respond by releasing the contents of cytoplasmic granules (granzymes, perforins) that enter the target (cancer) through endocytosis and induce apoptosis.

$CD4^+$ cells respond to antigen-presenting cells with the release of a variety of cytokines. The pattern of cytokine release can be classified into a TH1 or TH2 pattern, which broadly can be defined as promoting cellular versus humoral responses, respectively. TH1 cytokines are interferon gamma (INFγ), IL-2 and tumor necrosis factor (TNF) beta while TH2 cytokines are IL-4, IL-5, IL-6, IL-10, and IL-13.

3.1.1 Control of T-cell Proliferation

After a T lymphocyte has been stimulated to proliferate, it is necessary to eventually limit the clonal expansion of these cells, lest they overwhelm the host. Described as cytotoxic, T-lymphocyte antigen-4, CTLA-4 is an important regulatory molecule expressed only on activated T cells. Nascent CTLA-4 is only expressed after T-cell activation, peaking at 2 to 3 days after the initiating event. CTLA-4, also known as CD152, has a higher affinity for CD80 and CD86 than the CD28 ligand on T cells responsible for "signal 2" (a costimulatory signal). The cytoplasmic domain of CD152 has an inhibitory domain, termed an *ITIM* (*inhibitory tyrosine immunoglobulin domain*), which contributes to an inhibitory signal for T-cell growth. Binding of CD80/86 to CD152 thus attenuates the immune response. The important role of CTLA-4 in moderating T-cell growth is apparent from the observation that CTLA-4 knockout mice experience lethal T-cell proliferation.

Other mechanisms of immunologic control include active suppression by certain regulatory T cells expressing CD4 and CD25, the alpha chain of the IL-2 receptor (CD25). Formerly called "suppressor" cells, this subset of $CD4^+/CD25^+$ T cells is emerging as a potent regulatory component of the immune system that is involved in immune tolerance. Since tumors that have already developed have evaded immunologic recognition, tolerance has apparently developed. Targeting $CD25^+$ cells in tumor-bearing animals has had therapeutic effects accompanied by immune responses to TAA.

3.1.2 Innate Immunity

3.1.2.1 **NK Cells** This remarkable cell is capable of killing cancer, virus-infected, and allogeneic cells through granule exocytosis, perforin, and granzymes. How NK cells recognize their targets, without recognizing normal cells, has been partially elucidated. NK cells express a family of inhibitory receptors: KIR (killer inhibitory receptors) that bind class I HLA. Thus, on autologous cells, NK cells are inhibited from mediating cytolysis of class I expressing cells. However, if HLA is downregulated, as it often is on cancer cells, the absence of HLA is the recognition element for the NK cell. The activity of NK cells is also mediated by their ability to release cytokines such as IFNγ and other TH1 cytokines. A subset of cells termed NKT cells exists, expressing both NK cell markers as well as TCR with limited Vβ usage, showing specificity for lipid or glycolipid antigens presented by CD1.

3.1.2.2 **LAK Cells** Lymphokine-activated killer (LAK) cells can be generated from T lymphocytes and NK cells. Using high concentrations of IL-2 in *ex vivo* cultures, large numbers of LAK cells can be generated with broad specificity against autologous and allogeneic target cells.

3.1.2.3 **Monocytes/Macrophages** Part of the innate immune system, macrophages and their precursors (monocytes) are able to directly engulf or phagocytose cancer cells. In addition, they secrete cytokines such as TNF and IFNγ, which also can kill cancer cells. Macrophages are also potent antigen-presenting cells, expressing class II HLA, that further promote immunity to cancer cells by this mechanism.

Monocyte/macrophages also link the cellular and humoral arms of the immune system through the cell surface expression of receptors for the Fc portion of immunoglobulin. Antibodies binding to soluble antigens, such as a TAA, thus target the antigen to mononuclear phagocytes by binding to Fc receptors, which leads to internalization of the antigen–antibody complex and subsequent antigen processing for display on class II HLA molecules. In turn, $CD4^+$ T lymphocytes then recognize the antigen(s) expressed by these professional antigen-presenting cells.

3.1.2.4 **Tumor Escape Mechanisms from Immune Control** In many cancers, such as malignant melanoma, there is an infiltration of immune effector cells frequently. The infiltrate generally contains both $CD4^+$ and $CD8^+$ cells, which can be shown to proliferate in response to TAA. In addition, following vaccination with TAA, an immune response frequently develops. Using tetramer analysis (the use of four linked MHC molecules with matching peptide to enhance binding to specific T cells and increase the ability to identify an antigen-specific T-cell population via flow

cytometry), tumor antigen–specific T cells have also been identified in the circulation of patients with cancer. However, even with tumor infiltrating lymphocytes or circulating tumor-specific T cells, tumors frequently progress and metastasize or do not respond to vaccination, suggesting that the effector cells are functionally impaired, or that the tumor antigen is too weak (i.e. there is tolerance). Any impairment of the immune systems' ability to eliminate tumor can affect both the innate and adaptive arms. Given the complexity involved in the initiation of an immune response, it is hardly surprising that there are many possible mechanisms that tumors can use to evade an immune response. Understanding these mechanisms and determining ways to overcome them will be essential to the development of effective immunologic therapeutic approaches. This is the subject of the following paragraphs.

Most malignancies express HLA class I molecules but not class II molecules. This mirrors the situation on normal cells. Tumor cells, including those from melanoma, breast, bladder, cervix, colon, head and neck, kidney, and lung cancers have been demonstrated to downregulate expression of class I molecules, resulting in deficient antigen presentation to $CD8^+$ T cells. One beneficial consequence of this may actually be enhanced NK-cell directed activity. NK cells have the ability to destroy infected or malignant cells. They recognize their target either via Fc receptor binding or via a pair of receptors called the "killer activating receptor" and the "killer inhibitory receptor" (KIR). The killer inhibitory receptor recognizes the MHC class I molecule and in its absence, the killer activating receptor is dominant, resulting in perforin and granzyme mediated cell lysis. However, as the malignancy generally persists following downregulation of the MHC class I molecule, the importance of this phenomenon remains unclear.

There are a number of mechanisms by which HLA loss occurs. There can be complete HLA loss, HLA haplotype loss, HLA locus loss, or HLA allele loss. Examples of $\beta 2$ microglobulin loss, deficiency of TAP-1 or TAP-2, or structural defects in MHC genes have all been reported, resulting in complete HLA loss and the inability to present antigen for recognition by $CD8^+$ T cells.

Proteasome subunit alterations have also been reported in a number of cancers. The ζ chain is an important TCR-associated signal transduction molecule essential for signaling following TCR ligation. Reduced expression of this molecule has been reported in a number of malignancies including malignant melanoma, colon cancer, and renal cell cancer. Loss of the ζ chain has been associated with a poorer survival in oral squamous cell cancers.

NFκB is a transcription factor involved in the expression of a number of molecules involved in mediating an effective immune response. Impaired NFκB activation has also been implicated in the defective T-cell responses in some malignancies including renal cell cancer (RCC). In RCC, this suppression appeared to be induced by gangliosides derived from malignant cells.

Expression of nonclassical HLA molecules, including HLA-G and HLA-E may also be important mechanisms of immune escape utilized by malignant cells. HLA-G has been reported to inhibit NK-mediated cytotoxicity and also the function of antigen-specific CTL. Similarly, expression of HLA-E has been demonstrated to inhibit the function of immune effector cells. Although classical HLA class I downregulation decreases antigen presentation to CTL, it enhances

recognition by NK cells. The expression of these nonclassical HLA class I molecules appears to interfere with this effect.

Although conflicting data has been reported, tumor-induced apoptosis of T cells may be an important mechanism of immune escape. This phenomenon has been termed *tumor counterattack*. It results from the interaction of CD95L on tumor cells with CD95 on immune effector cells, resulting in effector cell apoptosis. Another death ligand, TRAIL, and chemokines have also been implicated in tumor-induced apoptosis, resulting in immune escape.

An additional tumor escape mechanism may be secretion of immunosuppressive factors by malignant cells or from normal supportive cell populations. TGF-β and vascular endothelial growth factor (VEGF) are well-characterized examples. TGF-β is a microenvironment growth inhibitory regulatory cytokine. It may be involved in the prevention of neoplasia via its growth inhibitory properties, but, in many established cancers, resistance to the inhibitory effects has been reported. Many cancers produce TGF-β, which then inhibits the immune response, allowing enhanced tumor growth and metastasis. VEGF, in addition to enhancing tumor angiogenesis, has also been demonstrated to inhibit dendritic cell maturation and, thus, antigen presentation.

These abnormalities may occur initially in the microenvironment of the tumor cells but lymphocytes bearing similar abnormalities may also be found in the peripheral circulation. Generally, the abnormalities have been most marked in the tumor infiltrating lymphocytes, suggesting that the effects are initiated in the local tumor microenvironment. Their frequent inability to prevent metastasis suggests that the tumor inhibitory effect is dominant.

4 Immunotherapy

4.1 Passive Immunotherapy

Immunotherapy can be described as active or passive. Active immunotherapy refers to specific maneuvers taken to induce endogenous immune reactivity directed against TAA. Passive immunotherapy refers to the therapeutic administration of *ex vivo* generated intact antibodies, antibody fragments, or lymphocytes that are directed at TAA.

As a result of the expression of well-defined antigens on the surface of lymphoma and leukemia cells, a number have been identified as suitable targets for antibody-directed therapies. Ideally, these antigens will not be expressed on hematopoietic stem cells or on non-hematopoietic cells in order to minimize toxicity.

Recently, a number of monoclonal antibodies (mAb) have been approved for use in the United States of America. The first mAb to receive United States Food and Drug Administration (FDA) approval was Rituximab (Rituxan, Biogen IDEC Pharmaceuticals, San Diego, CA; Genentech, Inc, South San Francisco, CA). Rituximab is a chimeric antibody directed at the CD20 protein expressed on the surface of mature B cells. It is generated by cloning the light and heavy variable chain of 2B8, a murine monoclonal antibody, into a cDNA expression vector containing the human IgG1 heavy chain and kappa light-chain constant regions. As it contains a human antibody *backbone*, it has the ability to bind complement and also lyse target cells via antibody-dependant cellular cytotoxicity. It has also been shown to activate apoptotic pathways

and appears to enhance chemotherapy sensitivity. Rituximab is approved for use in indolent NHL but also has activity in aggressive lymphoma, especially when combined with chemotherapy.

Alemtuzumab (Campath-1H) (Wellcome, Beckenham, UK and Ilex Oncology, San Antonio, TX) is a mAb directed against CD52, a cell surface protein present on most normal and malignant B and T cells, monocytes, and macrophages. As with CD20, it is not expressed on hematopoietic stem cells. Its mechanism of action appears to be via antibody-dependant cellular cytotoxicity and complement fixation.

Epratuzumab (Amgen, Thousand Oaks, CA) is an anti-CD22 monoclonal antibody that has also demonstrated activity in NHL and continues to undergo clinical evaluation. The CD22 protein is expressed on 75% of B lymphocytes.

Two radioimmunoconjugates are currently licensed in the United States. The first is Ibritumomab Tiuxetan (Zevalin, Biogen IDEC Pharmaceuticals, San Diego, CA). This molecule is composed of the β emitter ^{90}Y linked by the chelator tiuxetan to the murine IgG1 kappa monoclonal antibody ibritumomab (IDEC-2B8). The second radioimmunoconjugate is Tositumomab (Bexxar, Corixa Pharmaceuticals, South San Francisco, CA). The radiation molecule in this case is ^{131}I, which is both a β- and γ-emitter. Both radioimmunoconjugates are FDA-approved for use in certain indolent lymphoma patients. They both target the CD20 protein on the surface of normal and malignant lymphocytes. The response rates appear to be higher than those seen following Rituximab use, most likely because of the ability of a single molecule to target greater numbers of malignant cells as a result of the path length of the β and γ emissions.

Another mAb approved for therapy is trastuzumab (Herceptin, Genentech, South San Francisco, CA). This mAb is directed to the HER-2 molecule, a member of the erbB epidermal growth factor receptor tyrosine kinase family. This receptor is overexpressed, mostly as a result of gene amplification, in 20 to 30% of breast cancers and is associated with a poor prognosis. Trastuzumab is a chimeric antibody composed of 95% human and 5% murine components. As with Rituximab, this results in a markedly reduced ability to form human antimouse antibodies and allows increased efficacy through the use of human effector mechanisms including antibody-dependant cellular cytotoxicity and complement fixation. Trastuzumab has activity both as a single agent and when combined with chemotherapy for HER2 overexpressing cancers. It is of note that patients who have received an anthracycline, either previously or concurrently, have an increased incidence of cardiotoxicity. The mechanism is unclear but does not appear to relate to HER2 expression in cardiac myocytes.

Gemtuzumab ozogamicin (Mylotarg, Wyeth-Ayerst Laboratories, Philadelphia, PA) is a humanized anti-CD33 monoclonal antibody conjugated to calicheamicin. CD33 is expressed in 85 to 90% of acute myelogenous leukemia cells. Calicheamicin is a potent antitumor antibiotic that induces double-stranded breaks in DNA. The conjugate has shown efficacy in patients with relapsed CD33^{+} acute myelogenous leukemia. Unlike the previously discussed antibody therapies, this conjugate is rapidly internalized, following which the calicheamicin is released via hydrolysis.

Another approach to immunotherapy has involved the generation of bispecific antibodies. These are hybrid antibodies

constructed from two parent monoclonal antibodies. One antibody is specific for the tumor cell whereas the other is specific for an immune effector cells. Thus, bispecific antibodies can direct the cytotoxic activities of T cells, NK cells, monocytes, and monocyte-derived macrophages.

4.2 Active Immunotherapy

Active immunotherapy indicates that a treatment is intended to induce a specific antitumor immune response. This term generally implies the use of vaccination to treat cancer or prevent its recurrence, differing from the normal use of this term, which refers to the primary prevention of a disease, usually of an infectious nature. Although a number of vaccine approaches have been evaluated or are in clinical trials for various malignancies, no therapy evaluated thus far has been approved for routine clinical use.

The development of cancer vaccines has evolved directly from the identification of specific TAA. In addition to the use of specific known antigens, modified whole-cell preparations can function as vaccines. Whole tumor cells may be irradiated and transduced with immunostimulatory molecules such as granulocyte/monocyte colony-stimulating factor (GM-CSF). Such an approach has the potential to induce a response to a wide array of tumor antigens, but the possibility of an autoimmune response also is increased. An advantage is that the structure or identity of specific antigens does not need to be known.

For a specific antigen to be useful as a vaccine, it should be both tumor specific and immunogenic. Unfortunately, most candidate tumor vaccines are only weakly immunogenic and require modification in order to induce an immune response. The most widely applied cancer vaccine approach has utilized the immunoglobulin idiotype (the hypervariable heavy- and light-chain regions that result from a unique set of recombination events that occur during B-cell ontogeny). Every mature B cell expresses surface immunoglobulin and the idiotype of a particular B cell and all its progeny is identical and unique. In B-cell lymphomas, where at least initially all the cells are derived from the same clone, the idiotype is an ideal vaccination target. A potential drawback is that as lymphoma progresses, a specific antigen may be lost or mutate, in which case the vaccine would no longer be effective.

Animal studies have demonstrated the ability of idiotype vaccines to provide protection from tumor challenges and also to induce responses in established tumors. In indolent lymphoma patients, this approach has also induced molecular remissions in patients with minimal residual disease following chemotherapy. At present, clinical trials, currently in phase 3, are being conducted. The most common approach has been to generate idiotype protein and conjugate it to a hapten such as keyhole limpet hemocyanin, KLH). In addition, an immunostimulatory molecule such as GM-CSF is usually administered in the proximity of the vaccine to augment immunity. The schema of one trial is illustrated in Fig. 3. In this clinical trial, a lymph node representative of the patient's lymphoma is surgically excised. From this tissue, a recombinant idiotypic protein is generated, coupled to KLH, and administered subcutaneously along with GM-CSF as a vaccine. Such an approach can be taken in patients following chemotherapy when they have achieved a remission, with the goal of inducing an immune response to the idiotypic protein, thus facilitating immunologic control of

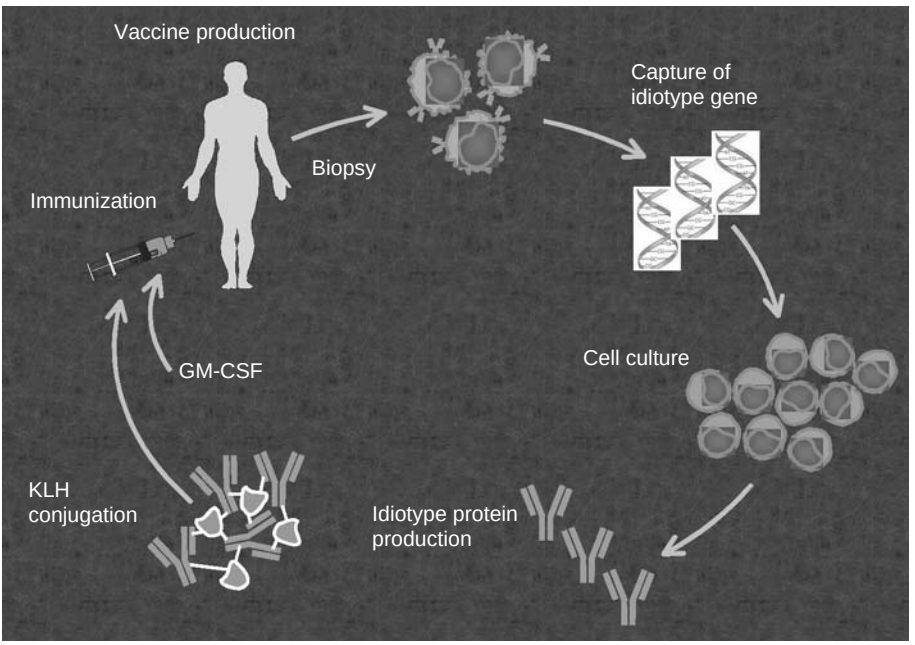

Fig. 3 Anti-idiotype vaccine preparation and administration.

any residual lymphoma cells that might lead to relapse eventually.

Another variation on this approach is the use of dendritic cells obtained from peripheral blood mononuclear cells by leukapheresis, which are cultured and then pulsed *ex vivo* with idiotype protein with or without the addition of KLH. The resulting mature dendritic cells are then returned to the patient as a cellular vaccine. Combinations of these approaches have also been utilized.

DNA vaccines are also undergoing evaluation in clinical trials and have shown efficacy in terms of the generation of specific immune responses. DNA vaccines are relatively simple to prepare. For the preparation of a DNA idiotype vaccine, variable heavy- and light-chain cDNA is isolated from tumor samples and cloned into a plasmid. In order to enhance immunogenicity, an unrelated immunostimulatory gene sequence such as that for tetanus toxoid can also be incorporated. The plasmid is bacterial derived and incorporates eukaryotic or viral promotor/enhancer transcription elements that allow efficient transcription of the target gene within the host cell. The plasmid can be delivered directly to host cells where intracellular antigen production is initiated. The main appeal of this approach, apart from simplicity, is intracellular antigen production that allows antigen presentation on the cell surface in association with HLA class I molecules, resulting in the stimulation of a cytotoxic $CD8^+$ T-cell response. In addition, secreted antigen is processed and presented by APC, resulting in both T helper response and antibody responses. In contrast, following the administration of a protein vaccine, antigen presentation is largely via professional APC, resulting predominantly in a T helper response.

Given the importance of the CTL response, if proven effective, DNA vaccination may hold certain advantages. Plasmid DNA preparation is the simplest of approaches to vaccine preparation and is in early clinical development.

The effect of tumor vaccines thus far has been difficult to evaluate. The ultimate evaluation will depend on documenting improvements in clinically relevant outcomes such as time-to-relapse or overall survival. In a disease such as indolent lymphoma, such studies would require very large number of patients followed for very long time periods, making conduct of such studies very cumbersome. Therefore, to date, surrogate immunological outcomes have been used, namely, the development of cellular or humoral immune responses. However, there remains significant controversy regarding the relative importance of each of these components of the immune system in mediating responses to tumor vaccination. Further, as discussed earlier, the development of a response may not correlate with therapeutic benefit because of local immune suppressive factors produced locally by tumors. Large-scale clinical trials with clinical outcomes will ultimately be required to demonstrate the benefits of this approach and the particular clinical situations where maximal benefit can be achieved.

4.3 *Ex vivo* Generation of CTL

Efforts by countless scientists in academia and industry are continuing to harness the cells and other components of the immune system to fight cancer. The generation of cytotoxic T-cell lines or polyclonal populations of cytotoxic T cells has been a major component of this multiplicative approach. CTLs can be grown from primary tumors in the case of melanoma and other solid tumors. In addition, they can be grown from the peripheral blood of patients with hematologic malignancy. Using cytokines such as IL-2, IL-12, IL-13 and monoclonal antibodies stimulating the costimulatory receptors CD28 and CD3 allows the generation of large numbers of CTL, which can then be cloned by limiting dilution. Efforts to achieve clinical responses in patients with solid tumors have been relatively unsuccessful to date. However, armed with increased knowledge of tumor antigens, a multitude of cytokines, and greater insight into the requirements for cellular activation, substantial progress is finally being made. Recently, it was shown that complete responses could be achieved in melanoma patients given CTLs after a brief treatment with cytotoxic chemotherapy to allow "space" for the incoming lymphocytes to proliferate and mediate cytotoxicity. This clinical trial illustrated the principle that attacking cancer requires a combination of various treatment modalities. Investigators are currently working on expanding T cells to treat leukemia, lymphoma, and a multitude of solid tumors. Other approaches under study include the blockade of CTLA-4 on T cells reacting to TAA. Such blockade allows more robust growth of T cells by neutralizing the negative regulatory activity of the CTLA-4 molecule.

4.4 *Ex vivo* Generation of Dendritic Cells (DC)

Dendritic cells are very potent antigen-presenting cells. They can be generated *ex vivo* from peripheral blood monocytes or bone marrow hematopoietic progenitor cells. Dendritic cells can also be derived

directly from myeloid leukemia cells by G- and GM-CSF and IL-4. Once generated, dendritic cells can be pulsed with tumor peptide epitopes for display on surface HLA. In addition, dendritic cells may be loaded with tumor cell lysates or apoptotic bodies, allowing display of uncharacterized tumor antigens from heterogeneous cell populations. Dendritic cells(DC) may be infused *in vivo* or may be injected intratumorally to present antigens to local T cells for antitumor effects. Promising clinical trials have been reported at the time of this writing and there is considerable hope that this form of treatment will be useful for patients with cancer.

4.5
Ex vivo Generation of NK Cells

Generation of large numbers of NK cells can be achieved through *ex vivo* expansion using IL-2. However, clinical therapeutic results have been somewhat disappointing to date. This probably is due to the inability of infused natural killer cells to track *in vivo* to sites of tumor. Further insights into the mechanisms and controls of cell trafficking will hopefully improve the results of NK-cell mediated therapy.

4.6
Summary and Conclusions

Prevention and effective treatment of human malignant disease by the immune system is the goal of modern immunological and clinical research. Efforts thus far have suffered in the clinic from a lack of basic understanding of the molecular biology and physiology of tumors and the immune system. Recent advances in the understanding of the biochemical nature of tumor-associated antigens, the mechanism of antigen presentation to the immune system, and the means by which tumor cells evade immune elimination have given rise to a number of more rational and targeted approaches to immunotherapy. In addition, the advent of powerful engineered molecules including monoclonal antibodies, cytokines, and chemokines and the ability to manipulate, purify, and expand cells of the immune system *in vitro* have provided the tools to test many new therapeutic approaches to the treatment of human malignancy.

See also Antigen Presenting Cells (APCs); Cytokines: Interleukins; Immunologic Memory; Immunology; Immunosuppressive Factors in Cancer; Innate Immunity; Molecular Mediators: Cytokines.

Bibliography

Books and Reviews

Boon, T., Cerotini, J., van den Eynde, B., van der Bruggen, P., van Pel, A. (1994) Tumor antigens recognized by T lymphocytes, *Annu. Rev. Immunol.* **12**, 337–342.

Chien, K.R. (2000) Myocyte survival pathways and cardiomyopathy: implications for trastuzumab cardiotoxicity, *Semin. Oncol.* **27**(6 Suppl. 11), 9–14; discussion 92–100.

Garrido, F., Ruiz-Cabello, F., Cabrera, T., Perez-Villar, J.J., Lopez-Botet, M., Duggan-Keen, M., Stern, P.L. (1997) Implications for immunosurveillance of altered HLA class I phenotypes in human tumours, *Immunol. Today* **18**(2), 89–95.

May, M.J., Ghosh, S. (1998) Signal transduction through NF-kappa B, *Immunol. Today* **19**(2), 80–88.

Restifo, N.P., Wunderlich, J.R. (2001) Essentials of Immunology, in: DeVita, V.T., Jr., Hellman, S and Rosenburg, S.A. (Eds.) *Cancer Principles and Practices of Oncology*, Lippincott Williams & Wilkins, Philadelphia, 43–73.

Sievers, E.L. (2003) Antibody-targeted chemotherapy of acute myeloid leukemia using gemtuzumab ozogamicin (Mylotarg), *Blood Cells Mol. Dis.* **31**(1), 7–10.

Stevenson, F.K. (1999) DNA vaccines against cancer: from genes to therapy, *Ann. Oncol.* **10**(12), 1413–1418.

Primary Literature

Balaian, L., Ball, E.D. (2001) Direct effect of bispecific anti-CD33 x anti-CD64 antibody on proliferation and signaling in myeloid cells, *Leuk. Res.* **25**(12), 1115–1125.

Bendandi, M., Gocke, C.D., Kobrin, C.B., Benko, F.A., Sternas, L.A., Pennington, R., Watson, T.M., Reynolds, C.W., Gause, B.L., Duffey, P.L., Jaffe, E.S., Creekmore, S.P., Longo, D.L., Kwak, L.W. (1999) Complete molecular remissions induced by patient-specific vaccination plus granulocyte-monocyte colony-stimulating factor against lymphoma, *Nat. Med.* **5**(10), 1171–1177.

Borrello, I., Sotomayor, E.M., Rattis, F.M., Cooke, S.K., Gu, L., Levitsky, H.I. (2000) Sustaining the graft-versus-tumor effect through posttransplant immunization with granulocyte-macrophage colony-stimulating factor (GM- CSF)-producing tumor vaccines, *Blood* **95**(10), 3011–3019.

Campbell, M.J., Esserman, L., Byars, N.E., Allison, A.C., Levy, R. (1990) Idiotype vaccination against murine B cell lymphoma. Humoral and cellular requirements for the full expression of antitumor immunity, *J. Immunol.* **145**(3), 1029–1036.

Cerny, T., Borisch, B., Introna, M., Johnson, P., Rose, A.L. (2002) Mechanism of action of rituximab, *Anticancer Drugs* **13**(Suppl. 2), S3–10.

Cheson, B.D. (2003) Radioimmunotherapy of non-Hodgkin lymphomas, *Blood* **101**(2), 391–398.

Coiffier, B., Lepage, E., Briere, J., Herbrecht, R., Tilly, H., Bouabdallah, R., Morel, P., Van Den Neste, E., Salles, G., Gaulard, P., Reyes, F., Lederlin, P., Gisselbrecht, C. (2002) CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma, *N. EngI. J. Med.* **346**(4), 235–242.

Cretney, E., Degli-Esposti, M.A., Densley, E.H., Farrell, H.E., Davis-Poynter, N.J., Smyth, M.J. (1999) m144, a murine cytomegalovirus (MCMV)-encoded major histocompatibility complex class I homologue, confers tumor resistance to natural killer cell-mediated rejection, *J. Exp. Med.* **190**(3), 435–444.

Dudley, M.E., Wunderlich, J.R., Robbins, P.F., Yang, J.C., Hwu, P., Schwartzentruber, D.J., Topalian, S.L., Sherry, R., Restifo, N.P., Hubicki, A.M., Robinson, M.R., Raffeld, M., Duray, P., Seipp, C.A., Rogers-Freezer, L., Morton, K.E., Mavroukakis, S.A., White, D.E., Rosenberg, S.A. (2002) Cancer regression and autoimmunity in patients after clonal repopulation with antitumor lymphocytes, *Science* **298**(5594), 850–854.

Egen, J.G., Kuhns, M.S., Allison, J.P. (2002) CTLA-4: new insights into its biological function and use in tumor immunotherapy, *Nat. Immunol.* **3**(7), 611–618.

Farag, S.S., Fehniger, T.A., Ruggeri, L., Velardi, A. Caligiuri, M.A. (2002) Natural killer cell receptors: new biology and insights into the graft-versus-leukemia effect, *Blood* **100**(6), 1935–1947.

Gabrilovich, D.I., Chen, H.L., Girgis, K.R., Cunningham, H.T., Meny, G.M., Nadaf, S. Kavanaugh, D., Carbone, D.P. (1996) Production of vascular endothelial growth factor by human tumors inhibits the functional maturation of dendritic cells, *Nat. Med.* **2**(10), 1096–1103.

Giovarelli, M., Musiani, P., Garotta, G., Ebner, R., Di Carlo, E., Kim, Y., Cappello, P., Rigamonti, L., Bernabei, P., Novelli, F., Modesti, A., Coletti, A., Ferrie, A.K., Lollini, P.L., Ruben, S., Salcedo, T., Forni, G. (1999) A stealth effect: adenocarcinoma cells engineered to express TRAIL elude tumor-specific and allogeneic T cell reactions, *J. Immunol.* **163**(9), 4886–4893.

Igney, F.H., Krammer, P.H. (2002a) Immune escape of tumors: apoptosis resistance and tumor counterattack, *J. Leukoc. Biol.* **71**(6), 907–920.

Igney, F.H., Krammer, P.H. (2002b) Death and anti-death: tumour resistance to apoptosis, *Nat. Rev. Cancer* **2**(4), 277–288.

Jager, E., Chen, Y.T., Drijfhout, J.W., Karbach, J., Ringhoffer, M., Jager, D., Arand, M., Wada, H., Noguchi, Y., Stockert, E., Old, L.J.,

Knuth, A. (1998) Simultaneous humoral and cellular immune response against cancer-testis antigen NY-ESO-1: definition of human histocompatibility leukocyte antigen (HLA)-A2-binding peptide epitopes, *J. Exp. Med.* **187**(2), 265–270.

Jones, E., Dahm-Vicker, M., Simon, A.K., Green, A., Powrie, F., Cerundolo, V., Gallimore, A. (2002) Depletion of CD25$^+$ regulatory cells results in suppression of melanoma growth and induction of autoreactivity in mice, *Cancer Immun.* **2**, 1.

Juweid, M.E. (2002) Radioimmunotherapy of B-cell Non-Hodgkin's Lymphoma: from clinical trials to clinical practice, *J. Nucl. Med.* **43**(11), 1507–1529.

Lee, P.P., Yee, C., Savage, P.A., Fong, L., Brockstedt, D., Weber, J.S., Johnson, D., Swetter, S., Thompson, J., Greenberg, P.D., Roederer, M., Davis, M.M. (1999) Characterization of circulating T cells specific for tumor-associated antigens in melanoma patients, *Nat. Med.* **5**(6), 677–685.

Maeurer, M.J., Gollin, S.M., Martin, D., Swaney, W., Bryant, J., Castelli, C., Robbins, P., Parmiani, G., Storkus, W.J., Lotze, M.T. (1996) Tumor escape from immune recognition: lethal recurrent melanoma in a patient associated with downregulation of the peptide transporter protein TAP-1 and loss of expression of the immunodominant MART-1/Melan- A antigen, *J. Clin. Invest.* **98**(7), 1633–1641.

Mellado, M., de Ana, A.M., Moreno, M.C., Martinez, C., Rodriguez-Frade, J.M. (2001) A potential immune escape mechanism by melanoma cells through the activation of chemokine-induced T cell death, *Curr. Biol.* **11**(9), 691–696.

Pasche, B. (2001) Role of transforming growth factor beta in cancer, *J. Cell. Physiol.* **186**(2), 153–168.

Prehn, R., Main, J. (1957) Immunity to methylcholanthrene-induced sarcoma, *J. Natl. Cancer Inst.* **18**, 769–778.

Reft, M.E., Carner, K., Chambers, K.S., Chinn, P.C., Leonard, J.E., Raab, R., Newman,R.A., Hanna, N., Anderson, D.R. (1994) Depletion of B cells in vivo by a chimeric mouse human monoclonal antibody to CD20, *Blood* **83**(2), 435–445.

Reichert, T.E., Day, R., Wagner, E.M., Whiteside, T.L. (1998) Absent or low expression of the zeta chain in T cells at the tumor site correlates with poor survival in patients with oral carcinoma, *Cancer Res.* **58**(23), 5344–5347.

Reichert, T.E., Rabinowich, H., Johnson, J.T., Whiteside, T.L. (1998) Mechanisms responsible for signaling and functional defects, *J. Immunother.* **21**(4), 295–306.

Rosenberg, S.A. (2001) Progress in human tumour immunology and immunotherapy, *Nature* **411**(6835), 380–384.

Uzzo, R.G., Clark, P.E., Rayman, P., Bloom, T., Rybicki, L., Novick, A.C., Bukowski, R.M., Finke, J.H. (1999) Alterations in NFkappaB activation in T lymphocytes of patients with renal cell carcinoma, *J. Natl. Cancer Inst.* **91**(8), 718–721.

Uzzo, R.G., Rayman, P., Kolenko, V., Clark, P.E., Bloom, T., Ward, A.M., Molto, L., Tannenbaum, C., Worford, L.J., Bukowski, R., Tubbs, R., Hsi, E.D., Bander, N.H., Novick, A.C., Finke, J.H. (1999) Mechanisms of apoptosis in T cells from patients with renal cell carcinoma, *Clin. Cancer Res.* **5**(5), 1219–1229.

Immunology

Cindy Takeuchi[1] *and Paul Wentworth*[1,2]
[1] *The Scripps Research Institute, La Jolla, California, USA*
[2] *University of Oxford, Oxford, UK*

1 Essential Components and Functions of the Immune System 407
1.1 Innate and Adaptive Immunity 407
1.2 Constitutive and Inducible Defenses 408

2 Molecular Basis of Innate Immunity 410
2.1 Evolution 410
2.2 Affinity and Specificity 411

3 Innate Immune Recognition 412
3.1 Pathogen-associated Molecular Patterns 412
3.2 Pattern-recognition Receptors 413
3.3 Toll-like Receptors 415

4 Cellular Effector Mechanisms 419
4.1 Cell-mediated Inflammation 419
4.2 Oxygen-independent Destruction of Pathogens 420
4.3 Oxygen-dependent Killing 421

5 Humoral Effector Mechanisms 422
5.1 Complement 422
5.2 Antimicrobial Peptides 424
5.3 Cytokines 427

Bibliography 427
Books and Reviews 427
Primary Literature 428

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

Keywords

Cytokines
Low molecular weight, soluble proteins that mediate signaling and affect cellular behavior by specific binding to receptors on cell surfaces. Cytokines are pleiotropic, redundant, and multifunctional.

Chemokines
A class of cytokines with chemoattractant properties, which enables migration of leukocytes to sites of inflammation.

Inflammation
Recruitment of body defenses against an invading organism or tissue damage and ensuing repair of damage.

Leukocytes
White blood cells including the lymphocytes and most myeloid cells, which perform various significant immune functions.

Phagocytosis
The process of engulfment of invading particles within an invagination of the phagocyte cell membrane.

Prior to the late-1960s, most studies on vertebrate immune systems focused on the recognition and killing mechanisms of adaptive immunity. The evolutionarily ancient mechanisms of innate immunity, generally thought to be nonspecific and obsolete, were, for the most part, overlooked. However, the relevance of innate systems has become apparent in recent decades, especially with the discovery of Toll-like receptors and with the ability of pattern-recognition receptors to stimulate effective immune responses via the induction of genes and cytokine expression. This review focuses on the complex interplay at the molecular level, between innate immune activation and adaptive immune responses and on how this relationship is essential for host defense.

Current molecular studies of innate immunity seek to clarify the interconnectivity between innate and adaptive phenomena and to understand the mechanisms of innate defense in their own right. These include examinations of the ligand specificities and structural interactions of the innate immune receptors, gene expression regulation during the maturation of immune responses, and the interplay of signal transduction pathways that culminate in induced innate responses to various pathogens. The perception of the innate immune system as a highly

evolved biological design rather than a relic from ancestral species is allowing significant advances to be made in modern immunology.

1 Essential Components and Functions of the Immune System

1.1 Innate and Adaptive Immunity

The principal responsibilities of an immune system are to recognize a wide variety of pathogens and to destroy those pathogens while tolerating the tissues of the host organism. Innate immunity is the most universal and expeditious type of immunity, having evolved over the last hundred million years or so to accomplish these tasks. It includes physical barriers, such as the skin and epithelia that protect our internal organs, and the cells that bear innate immune receptors, such as macrophages, dendritic cells (DCs), neutrophils, basophils, eosinophils, mast cells, and natural killer (NK) cells. These white blood cells derive from myeloid and lymphoid precursors in the bone marrow. An encounter with a pathogen leads to the prompt removal of the agent or recruitment, activation, and differentiation of short-lived effector cells, which have further capacity to rid the host of most infections. Innate immune mechanisms work effectively against a broad range of pathogens during the first encounter with a pathogen. The response does not require an extended period of time to develop, due to the fact that the complete protein receptors are encoded in the germ-line DNA. Expression of innate immune receptors thus does not seem to involve either gene rearrangement or clonal expansion.

While the majority of organisms survive utilizing only innate immune mechanisms, vertebrates have developed secondary means for pathogen detection and eradication. When innate immunity is circumvented or overwhelmed, an adaptive system is utilized. Adaptive immunity relies on activated lymphocytes to generate highly variable antigen receptors, which provide specific recognition of protein, carbohydrate, lipid, and nucleic acid moieties of pathogens and foreign material. These recombinant receptors, the immunoglobulins and T-cell receptors (TCR), are capable of recognizing virtually any pathogen. Each lymphocyte generates a unique antigen receptor by rearrangement of multiple gene segments. Activated cells undergo selection and clonal expansion for four to five days in order to produce large quantities of these antigen-specific receptors. Binding events by these molecules activate downstream humoral and cellular effectors, many of which are components of the innate defenses. The adaptive immune system is thus dependent on the same effector mechanisms as innate immunity but imparts activation and antigen specificity to the process. Adaptive immunity also bestows immunological memory to a host, providing enhanced protection against reinfection by the same pathogen.

Host defense by innate immune mechanisms controls infection by pathogens with common molecular surface motifs and by those that induce interferons and nonspecific defenses. But innate defenses do not have the ability to recognize and fight a broad range of pathogens. Prior to infection, viruses are rarely recognizable by macrophages as the viruses do not have many invariant structures. Other pathogenic organisms can produce a capsule that conceals the otherwise

susceptible surface molecules to avoid detection. Significantly, innate defenses are also unable to elicit long-lasting memory of specific pathogens. This ability is limited to the adaptive immune processes found exclusively in vertebrates. Nevertheless, the large number of invertebrate species that exist on earth (much greater than vertebrates) attests to the success of the innate immune system.

The beneficial features unique to the adaptive immune responses include, great diversity for antigen recognition, specificity for a particular antigen or pathogen, and the ability to confer immunological memory after an encounter. However, adaptive immune responses do not become activated until approximately three days postinfection and the production of their primary components is delayed by five days. During this delay, the innate immune system limits the proliferation and spread of pathogens until the mechanisms of adaptive immunity achieve maximum potential. In the absence of innate immunity, acquired immune responses are inadequate for host protection. In contrast, the potency and pathogenicity of antibody-mediated effector functions can be seen in autoimmune diseases, allergic responses, and allograft rejection, which all result from the misregulation of adaptive immunity. This inopportune recognition of nonpathogenic material or self is generally confined to the adaptive immune response; autoreactivity has not been observed in organisms that accommodate only the components of innate immunity.

Immunological memory, preserved in the form of long-lived cells that persist in tissues and bone marrow, can last for months, years, or a lifetime and accounts for the efficacy of vaccines. These memory cells are readily activated by a subsequent encounter with antigen and can rapidly differentiate into effector cells. Unfortunately, this memory is confined to the individual challenged by the pathogen. Apart from passive immunization, such as transfer of antibody from mother to child, there is little carryover to future generations.

Despite the limitations of the individual systems, the benefits of incorporating two systems for immune defense are obvious. The strength of the adaptive immune response lies in its great diversity and specificity, providing complexity to the innate systems in the ability to recognize exquisite details of pathogenic organisms. Meanwhile, the innate defenses allow rapid discrimination of self from nonself and delay the need for adaptive immunity during the time required to generate effector lymphocytes. Chemokine secretion recruits leukocytes to appropriate sites, and cytokine-mediated signals provide optimal interplay between the divisions. In this manner, the innate immune system can direct the adaptive immune system in combating the infection. Thus, the synergy between the innate and adaptive mechanisms creates successful host defenses in vertebrates.

1.2 Constitutive and Inducible Defenses

The components of innate immune systems, found in most multicellular organisms, can be divided conceptually by function into the afferent and efferent limbs (Table 1). These defense mechanisms may also be discussed in terms of the point at which they are active during the course of infection (Table 2). The constitutive defense mechanisms include physical, mechanical, and chemical barriers that prevent initial invasion

Tab. 1 Afferent and efferent components of the innate immune system.

	Afferent function	*Efferent function*
Humoral components	Soluble innate immune receptors such as LPS-binding protein, CD14, collectins, properdin, complement component C3b, and C-reactive protein (pentraxin)	Secreted and cytosolic cytokines, antimicrobial peptides, enzymes, late components of complement, lactoferrin, and acute-phase proteins
Cellular components	Surface-bound receptors including toll-like receptors, MSR, MMR, and NOD proteins	Cellular enzymes, cell adhesion molecules, reactive oxygen species, and other antimicrobial substances

Tab. 2 Constitutive and inducible innate defense mechanisms.

Component	*Location*	*Function*
Physical factors		
Intact skin	External	Physical barrier to the entrance of microbes.
Mucous membranes	Internal cavities	Physical barrier to the entrance of microbes
Mucus	Respiratory and digestive tracts	Trap microbes
Epiglottis	Respiratory tract	Prevent microbes and particles from entering trachea
Hair	Nose	Filter microbes and dust
Mechanical factors		
Cilia	Upper respiratory tract	Trap and remove microbes and dust
Tears	Eyes	Dilute and remove irritating substances and microbes
Saliva	Oral cavity	Wash microbes from mucosal surfaces
Urine	Urogenital tract	Wash microbes from urethra
Chemical factors		
Gastric juice	Digestive tract	Destroy bacteria and toxins in stomach
Acid pH	External	Discourage growth of microbes
Unsaturated fatty acids	Various	Antibacterial in sebum
Other defenses		
Lysozyme	Various	Antimicrobial substance in perspiration, tears, saliva, nasal secretions, and tissue fluids.
Interferon (IFN)	Circulation	Protect host cells from viral infection
Complement	Circulation	Lyse microbes, promote phagocytosis, contribute to inflammation (leukocyte recruitment)
Phagocytosis	Peripheral tissues and circulation	Ingestion and destruction of foreign particles
Inflammation	Peripheral tissues	Confine and destroy microbes, repair tissues
Fever	General	Inhibit microbial growth, aid repair

by potential pathogens. Rigid junctions between epithelial cells effectively separate the interior milieu from the external environment. Production of a glycoprotein-rich mucus layer prevents adhesion of the microorganism to the epithelium, and movement of hairlike cilia aids in foreign-particle clearance. Antimicrobial enzymes and peptides (see Sect. 5.2) are also continuously expressed at mucosal epithelia, which have frequent exposure to microbes. The destructive capability of these substances affects microorganisms but is not directed at host cells and tissues. This distinction between self and nonself is accomplished by compartmentalization, or specific activity against microbes (e.g. susceptibility of bacterial cell walls, but not host cell membranes, to lysozyme and defensin). Yet another mechanism of defense is the colonization of epithelial surfaces with commensal microbes. These normally nonpathogenic bacteria compete with harmful microorganisms for nutrients and sites of attachment. Some bacteria even produce their own antimicrobial substances, or bacteriocins, to help outcompete other species.

If a focus of infection is established, inducible host defense mechanisms including local inflammatory responses are initiated at sites of infection. Heat, pain, redness, and swelling associated with inflammation are due to increased vascular diameter, increased endothelial adhesion, and increased vascular permeability mediated by cytokines. Pathogens that have penetrated host tissue are recognized by macrophages, which can phagocytose the offending microbes. Macrophages may also release cytokines, chemokines, and inflammatory mediators, which include tumor necrosis factor-α (TNF-α), prostaglandins, leukotrienes, and platelet activating factor. Another set of inflammatory mediators is produced via the activation of the complement cascade (discussed in Sec. 5.1), which also occurs in response to pathogen binding. These molecules can stimulate an acute-phase response, activate natural killer (NK) cells, and trigger interferon production and nonspecific defenses by regulating transcriptional activation. Cytokine production is also responsible for instructive cross talk to antigen-presenting cells and adaptive immune mechanisms.

2 Molecular Basis of Innate Immunity

2.1 Evolution

The innate immune system has its origins in the bilateria approximately one billion years ago. Today, nearly all multicellular organisms rely on derivative innate strategies for protection from infection. For example, mammals utilize skin as a barrier to pathogen invasion, whereas insects have an exoskeleton that performs the same function. Following this first tier of immune defense, specialized immune cells and tissues are induced to contain infections. White blood cells and liver of mammals function analogously to the hemocytes and fat body of certain insects in this respect. Diverse organisms have also evolved similar mechanisms for the recognition of infection and for swift responses to infection. Animals and plants alike have acquired the ability to recognize invariant patterns associated with pathogenic microbial organisms. In plants, these patterns are termed *general elicitors of plant defense*, while in animals

they are designated *pathogen-associated molecular patterns*. Amazingly, plant receptors recognize patterns similar to those reported to activate innate defense mechanisms in mammals and insects, and some of the specific receptors and signaling molecules used appear to share homology. Genomic evidence also exists for common pathways for transcriptional activation in plants, vertebrates, and invertebrates. But while *Drosophila* produce many different antimicrobial peptides to combat fungal and bacterial infection, mammalian systems produce fewer of these types of peptides and rely more on cytokine-mediated inflammatory responses. It has not been conclusively determined whether similarities between diverse animal and plant defense pathways are due to conservation through a billion years of evolution or whether the systems converged independently.

The mechanisms of adaptive immunity are thought to have developed much later on the evolutionary timeline in the context of a functioning innate immune system. The genes of the TCR and antibody first appeared in the cartilaginous fish, such as shark, approximately 400 million years ago. The origin of gene rearrangement and junctional diversity in immune receptors is thought to be an invasion by a retrotransposon. The recombinase-activating genes (RAGs), required to generate diversity in the receptors, are encoded in opposite orientations in the genome, and lack introns, similar to the bacterial transposases. Additionally, recognition signal sequences (RSSs) are present at the 3′-end of the V-gene segments and 5′-end of the J-gene segments. These properties support the notion of the integration of RAGs and RSSs into a gene for an ancient antigen receptor.

2.2 Affinity and Specificity

The innate and adaptive arms of the immune system differ in their modes of recognition. While antibodies are known for their exquisite specificity for antigens, the innate immune system recognizes a few highly conserved structures found in a large variety of microorganisms. The molecules targeted are vital components of microbes, so they are not often altered by mutation. Central to our understanding of the molecular basis of innate immune recognition are investigations of the receptor complexes that recognize these molecules. Binding of these receptors mediates ingestion of pathogens, allows cells to recognize chemotactic substances and migrate toward sites of infection, and induces effector molecules that regulate further immune responses. The affinities of innate immune receptors for their targets can run in the nanomolar range, for example, soluble CD14 binds Lipopolysaccharide (LPS) from a *Salmonella* strain with dissociation constants of 4.1×10^{-8} M and 7.1×10^{-9} M, in the absence and in the presence of LPS-binding protein (LBP) respectively. In comparison, affinity constants for antibody–ligand binding usually range from 10^{-8} M to 10^{-10} M, and TCR MHC–antigen affinities run in the range of 10^{-4} M to 10^{-5} M.

Importantly, innate receptors can distinguish nonself from self. While the receptors of T- and B lymphocytes allow great variability in adaptive immune recognition, they cannot determine the nature of the antigen. Although usually rare, adaptive immune responses to nonpathogen antigens can potentiate autoimmunity, allergy, or tissue-graft rejection.

3
Innate Immune Recognition

3.1
Pathogen-associated Molecular Patterns

Pathogen-associated molecular patterns (PAMPs) are the main targets of innate immune recognition. PAMPs are usually formed by constitutively expressed structures essential to the survival of a microbe and are often shared among members of large groups of microorganisms. Significantly, PAMPs are motifs that are not found on host cells. Host immune systems have evolved to take advantage of the following three characteristics. First, constitutive expression of the PAMP allows constant recognition regardless of pathogen life cycle, and because the structure is indispensable, it is unlikely to acquire mutations. Second, the host can employ a limited number of germ-line-encoded receptors to recognize a very broad range of microorganisms and can customize its responses to them because the target molecules are class specific. And third, the presence of a PAMP allows the host to distinguish itself from the foreign microbe. Lipopolysaccharide, peptidoglycan, and lipoteichoic acids (LTA) are examples of PAMPs that are fundamental components of bacterial cell walls and are not present on eukaryotic cells. Other PAMPs include the CpG motif present in bacterial DNA, components of flagella, and viral components. The detection of these PAMPs provides some information on the identity of the pathogen.

Lipopolysaccharide is the principal outer-membrane constituent of gram-negative bacteria. It is an amphipathic molecule consisting of a hydrophilic polysaccharide core and O-antigen and a hydrophobic glycolipid anchor termed *lipid A* (Fig. 1). The core region contains Kdo (3-deoxy-D-*manno*-octulopyranosonic acid) and heptose-sugar residues, while O-antigen is composed of a sequence of repeating units of oligosaccharide, which varies between bacterial species or strain. Lipid A is composed of a phosphorylated diglucosamine carrying acyl residues and is the primary immunostimulatory component of LPS. The primary effects of LPS were first reported over a century ago, but properties of this molecule continue to be studied today. LPS, the source of bacterial endotoxin and lipid A, is, to a degree, responsible for systemic inflammatory response syndrome (SIRS) and septic shock associated with gram-negative bacteremia. LPS stimulates the secretion of cytokines from leukocytes, namely, TNF-α, interleukin (IL)-1, and IL-6, and causes changes in body temperature, heart rate, respiratory

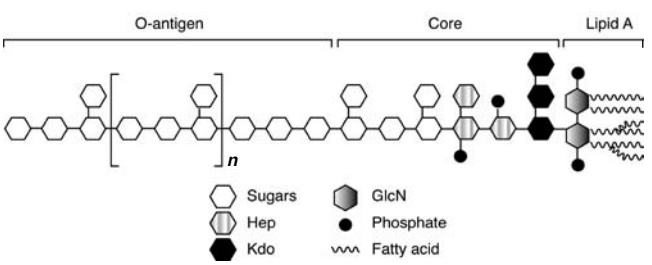

Fig. 1 General structure of LPS.
Kdo = 3-deoxy-D-manno-octulopyranosonic acid; GlcN = *N*-acetylglucosamine; Hep = heptose.

function, and leukocyte circulation. The resultant hypotension, systemic organ failure, and other significant complications lead to hundreds of thousands of deaths per year from this syndrome. However, most LPS-activated responses resolve infections quickly by initiating killing of the microbe.

Peptidoglycan and LTA may be the causative reagents in gram-positive bacterial septic shock. Peptidoglycan is a network of carbohydrate polymers covalently cross-linked by proteins, which composes the structural core of the gram-positive bacterial cell wall. Linear macromolecules are made of alternating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) residues joined by β – (1–4) linkages. Tetrapeptide side chains attached to MurNAc form covalent bonds with adjacent polymers directly through its side chains or through peptide interbridges. Teichoic acids, also found in the cell wall of gram-positive organisms, are polymers of glycerol and ribitol, linked by phosphodiester groups. Teichoic acids are connected to either the peptidoglycan itself or to plasma membrane lipids (LTA). The biological functions of LTA are known to include binding of metal cations. The LTA structure consists of two fatty acid tails and a long charged hydrophilic chain. The classic *Staphylococcal* LTA is an unbranched 1,3-linked glycerophosphate chain bound covalently to a membrane diacylglyceroglycolipid. LTAs are structurally distinct between bacterial species, varying in fatty acid composition, chain length, glycosyl substituents, glycolipid structures, spatial distributions, and charge densities. One or more of these features may explain the variable immunological properties of LTA from different sources. The fatty acid-containing lipid anchor and additional structural components on the polyglycerophosphate backbone contribute to the immunostimulatory activity of LTA. Recent studies have suggested a key role of D-alanine substituents in inducing cytokine secretion. Although the effects of LTA alone are not particularly severe, it can act synergistically with peptidoglycan, leading to systemic inflammation and multiple systems failure via induction of TNF-α and IFNγ.

3.2 Pattern-recognition Receptors

Pattern-recognition receptors (PRRs) detect conserved structures on microorganisms during innate immune defense. PRRs are structurally and functionally heterogeneous. They may contain distinct motifs such as C-type lectin domains, scavenger receptor cysteine-rich domains, and leucine-rich repeat (LRR) domains. The receptors may be secreted, or expressed on the cell surface, within intracellular compartments or in the cytoplasm. PRRs function through opsonization, complement activation, proinflammatory signaling pathways, phagocytosis, and the induction of apoptosis.

Secreted recognition molecules include the proteins synthesized in the liver during the acute-phase response, such as mannan-binding lectin (MBL), C-reactive protein (CRP), and serum amyloid P component (SAP). MBL is a collectin (contains both collagen-like and C-type lectin domains) that binds specifically to the carbohydrate arrays accessible on many microorganisms. MBL–ligand binding may be inhibited by mannose, fucose, *N*-acetylglucosamine, and DNA, but avidity is greatest for the terminal mannose residues typically present on microbial surfaces. By comparison, on vertebrate

cells, such carbohydrate patterns are usually masked by sialic acid residues, so for the most part, MBL recognizes only pathogens. MBL can effectively neutralize bacteria by binding or can opsonize bacteria and activate complement via an associated serum protease, MASP-2, initiating the lectin pathway by cleaving complement proteins C2 and C4. CRP and SAP are pentraxins that opsonize bacterial surfaces containing phosphorylcholine (PC). Both proteins display calcium-ion-dependent binding, but SAP has a slightly lower avidity for PC than does CRP. The distribution of PC in polysaccharides of foreign molecules and damaged host membranes allows CRP and SAP to recognize pathogens and altered self. CRP is found in trace amounts in human serum, but its concentration can elevate dramatically in response to inflammation and tissue necrosis. Upregulation occurs at the transcriptional level in response to cytokines IL-1β and IL-6. Multivalent binding of CRP activates complement by initiating assembly of a C3 convertase. Additionally, the protein interacts with FcγRI receptors, mediating phagocytosis and leading to the generation of proinflammatory cytokines. The topology and chemical composition of the CRP ligand-binding site has been determined by crystallography. SAP shares many features with CRP, including remarkable structural similarity, 59% sequence homology, and the ability to activate complement. Plasma LBP mediates the binding of minute amounts of LPS to glycosylphosphatidylinositol-anchored membrane CD14, triggering a proinflammatory response in macrophages, which is crucial to control infection by gram-negative pathogens. The CD14 LPS receptor complex may also involve molecules of TLR-4 and MD-2.

Several cell surface receptors function as PRRs to facilitate direct phagocytosis of microorganisms. The macrophage mannose receptor (MMR) is a C-type lectin found on macrophages but not on monocytes or neutrophils. The protein contains eight carbohydrate-recognition domains, the fourth of which is central to ligand binding, and like MBL, it can bind mannose, fucose, and *N*-acetylglucosamine. MMR is closely related to the dendritic cell PRR, DEC205, which is also a mediator of phagocytosis. Macrophage scavenger receptors (MSRs) bind a variety of charged ligands such as double-stranded RNA (dsRNA), LPS, and LTA, coordinating responses to a variety of viral and bacterial pathogens. MSRs also participate in lipid homeostasis by endocytosing acetylated and oxidized low-density lipoprotein. PRRs for a broad range of ligands have also been identified in invertebrates. These include the broad range β2integrin CR3, peptidoglycan recognition protein (PGRP), and gram-negative binding protein (GNBP), which have counterparts in the mammalian Toll-like receptors (TLRs).

Recognition systems also exist in the cytosol to prevent replication of intracellular pathogens. Activation of dsRNA-dependent protein kinase (PKR) inhibits initiation of viral and cellular protein synthesis by phosphorylating the translation initiation factor eIF2α. Interferon-inducible 2′,5′-oligodenylate synthase (OAS) also binds viral dsRNA within cells. Resulting synthesis of ppp(A2'p)nA oligoadenylates activate the RNaseL endonuclease, which cleaves RNA. This OAS/RNaseL pathway is particularly important in the antiviral action against picornaviruses. Members of the NOD (nucleotide-binding oligomerization domain) family are proteins containing a CARD, NOD, and LRR region. After PAMP

binding, protein kinase RIP2 associates with CARD domains and activates MAP kinase and NFκB signaling pathways. NOD proteins have been shown to bind LPS, but the full extent of ligand specificity for these molecules has not been elucidated. Downstream transcriptional activation by PKR (in TLR signaling) and NOD induces IFN genes.

3.3 Toll-like Receptors

Multiple innate immune receptors are able to distinguish nonself from self, as well as recognize individual classes of pathogens. Among diverse organisms, there exist genetic and structural homologies between the protein receptors and similarities in the pathways of immune-mechanism activation. The identification of the mammalian Toll-like receptor (TLR) family, homologs of the *Drosophila* Toll (Toll-1) protein, provided the primary evidence for evolutionary conservation within eukaryotic immunity. Both Toll and the TLRs are single-transmembrane spanning receptors possessing a variable number of extracellular amino-terminal LRR motifs, followed by a cysteine-rich domain. At the carboxy-end of the Toll/TLR transmembrane domain is an intracellular signaling domain, homologous to that of the cytoplasmic interleukin-1 receptor (IL-1R). It is therefore called the Toll/IL-1R (TLR) domain (Fig. 2).

The original function identified for Toll was the development of dorsoventral polarity in the fruit-fly embryo. It has since been discovered that Toll is required to combat systemic fungal infection in *Drosophila* by inducing the expression of drosomycin, an antifungal peptide. In addition to Toll, there are at least eight related homologs (Tolls-2–9), which appear to have functions in *Drosophila* development.

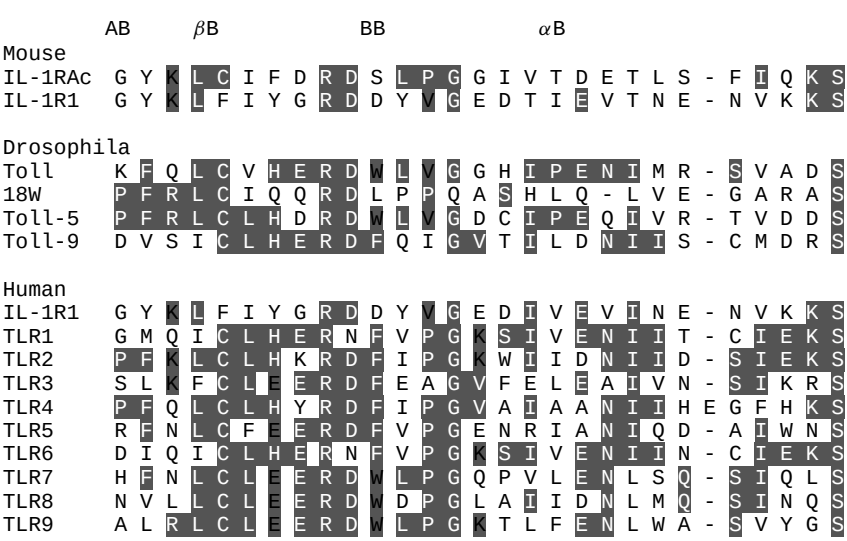

Fig. 2 Sequence alignment of the TLR domain-signaling region of mouse and human IL-1R proteins, *Drosophila* Toll proteins, and human TLRs. Residues conserved in at least four sequences are shaded. The greatest homology between *Drosophila* and human proteins occurs between the cytoplasmic domains of Toll-9 and TLRs 1, 2, 4, and 6.

Of these proteins, Toll-2 (or 18 Wheeler, 18 W) and Toll-5 have been found to have involvement in immune responses. Mutagenesis has shown that 18 W facilitates the recognition of gram-negative bacterial pathogens and the subsequent production of attacin, an anti-gram-negative bacterial peptide. Likewise, it has been suggested that Toll-5 acts as a pattern-recognition receptor mediating local antifungal responses in the absence of Toll-1/Spätzle signaling. These Toll family members respond to different pathogen classes and induce appropriate effector responses, presumably on the basis of PAMP recognition, that is, antibacterial and antifungal peptides are differentially expressed in response to the presence of bacteria or fungi. Significantly, these findings demonstrated that an organism lacking an adaptive immune system can detect infection as well as determine the type of pathogen present, entirely on the basis of innate immune mechanisms.

Mammalian TLRs may also be able to distinguish between pathogen classes. The TLRs function as PRRs in mammals, binding to microbial products to activate host defense responses. To date, 11 identified mammalian TLRs have been described, differing in sites of expression, ligand specificity, and downstream effector activation. Ten TLR genes (TLRs 1–10) are dispersed throughout the human genome, whereas mice encode nine of these, plus two additional paralogs not present in humans. Surface expression is usually low in monocytes ($\sim 10^3$ per cell) and DCs ($\sim 10^2$ per cell) but may be enhanced in response to various stimuli. TLR expression is also observed in adipocytes, myocytes, vascular endothelial cells, and intestinal epithelial cells. TLR4, the first mammalian TLR identified, when activated induces the expression of inflammatory cytokines and costimulatory molecules in macrophages. The murine TLR4 protein responds to the class of gram-negative bacteria. The molecular signature detected in this case is LPS. Although other PRRs were previously known to initiate signals in response to LPS, the existence of an additional coreceptor was hypothesized. The cellular response to LPS occurs through interaction with a circulating LBP and a GPI-linked surface receptor, CD14, followed by activation of TLR4. Point mutations in the TLR4 TIR domain led to defective LPS signal transduction, as did the absence or deletion of TLR4. These results demonstrated the necessity for a functional TLR4 in the recognition of this essential component of gram-negative bacteria. TLR4 has also been shown to mediate signals initiated by other endogenous and exogenous compounds (e.g. Taxol), respiratory syncytial virus protein F, heat shock protein 60, and domain A of fibronectin. Other conserved PAMPs have been shown to signal immune responses via other members of the TLR family, such as peptidoglycan by TLR2 and flagellin by TLR5. It seems likely that elucidation of newly discovered mammalian TLR homologs will elucidate the specificity of one or more of the homologs for mannan or glucan, and of fungal and mycobacterial pathogens for lipoarabinomannan or muramyldipeptide PAMPs, respectively.

Structural variation among TLRs is mainly due to differences in the numbers of LRRs, which affect the proteins' sequence length and molecular weight (Fig. 3). LRRs are found in a wide array of molecules, including proteoglycans, adhesion molecules, enzymes, tyrosine kinase receptors, and G-protein-coupled receptors. Toll and 18 Wheeler have 22 and 31 LRR ectodomains respectively, comparable to the 18, 19, 24, and 22 LRR-containing

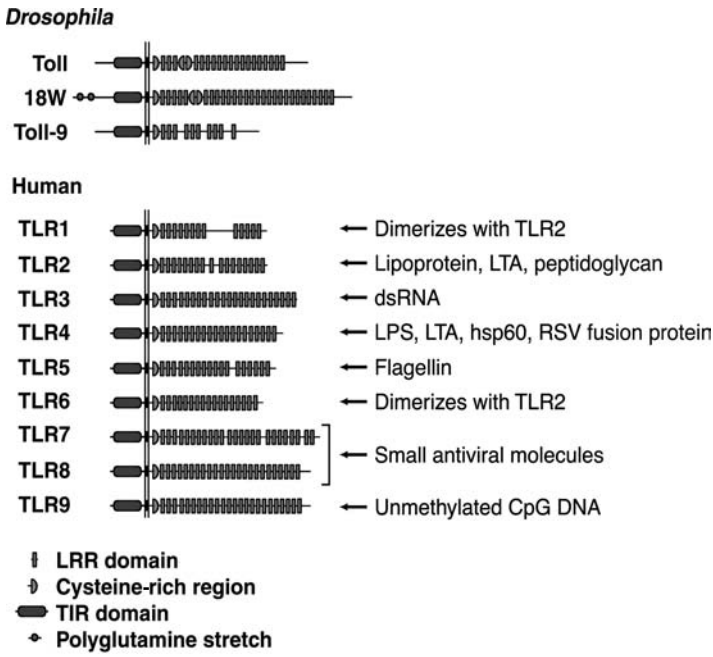

Fig. 3 Schematic representation of human TLR domains and their stimuli. Adapted from Imler, J.-L., Hoffmann, J.A. (2001) Toll receptors in innate immunity, *Trends Cell Biol.* **11**, 304–311.

TLRs 1–4. The LRRs are terminated by a domain of cysteine residues; however, the human TLR chains have lost an approximately 90-residue cysteine-rich region found variably in the *Drosophila* proteins. The extracellular domain of 18 W contains 16 putative N-glycosylation sites and a single glycosaminoglycan site. The cysteine clusters allow the receptors to form a compact module with paired termini that can be inserted between any pair of LRRs without altering the overall fold of the extracellular region. Two hydrophobic regions are apparent in the protein structure of 18 W. One corresponds to a signal peptide and the second to a transmembrane domain. The intracellular moiety of 18 W is about 380-aa long and consists of two domains, the first of which shows significant homology to the cytoplasmic domain of IL-1R. Interleukin-1 receptor, Toll, and 18 W share six conserved stretches of amino acids. The second domain consists of a glutamine-rich repeat.

Upon ligand binding, TLRs activate expression of host defense genes. Some of the intracellular signaling pathways induced are similar to those employed by IL-1R (Fig. 4). These pathways are coupled to the nuclear translocation of Rel-type transcription factors. An adaptor, such as MyD88, is first recruited to the receptor complex, mediating an interaction with the receptor via its amino-terminal TIR domain and associating with protein kinase IRAK1 through a carboxy-terminal death domain. IRAK1 undergoes autophosphorylation, which is thought to be followed by IKK-β activation via TRAF6 and by degradation of IκB. As a result, NFκB may translocate into the nucleus

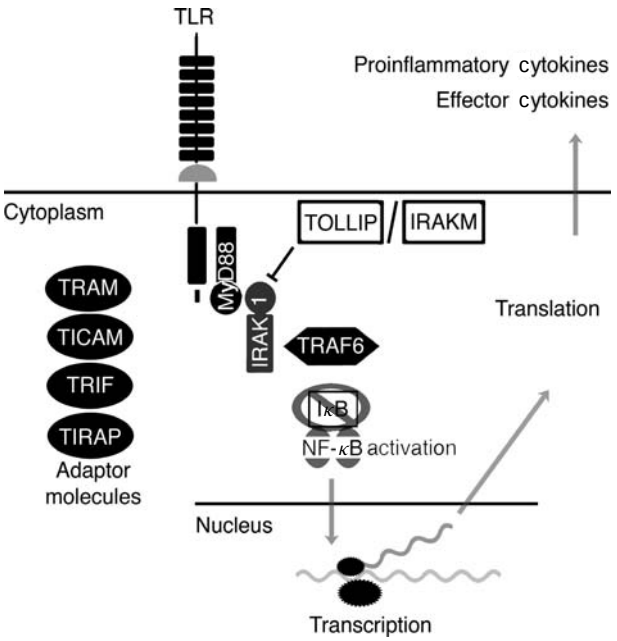

Fig. 4 MyD88-dependent and -independent TIR domain signaling. The MyD88-dependent pathway is common to all TLRs. Upon binding, TLRs induce the recruitment of MyD88 via its TIR domain, which activates IRAK1 by phosphorylation. IRAK1 then associates temporarily with TRAF6, eliciting downstream signaling, NFκB activation, and resultant production of proinflammatory cytokines such as TNF-α, IL-1, and IL-12. IRAK1 activity is negatively regulated by Tollip and IRAK-M. Some differences between signaling pathways induced by individual TLRs may be attributed to distinct adaptor protein usage. TLR4 and TLR2 MyD88-dependent signaling requires the adaptor TIRAP/Mal. IFN-β production in response to double-stranded RNA recognition by TLR3 is mediated by TRIF/TICAM-1 in a MyD88-independent manner. TRAM/TICAM2 is another adaptor pertinent to the TLR4 MyD88-independent pathway.

and allow transcription of appropriate targets. Although some TLRs are dependent on this conserved Toll signaling pathway (TLR2 and TLR9), MyD88 is not required for other TLR signaling events, such as the MyD88-independent pathway associated with the stimulation of IFN-β and the maturation of DCs. Additional regulatory mechanisms and downstream signals must also be in place to effect responses appropriate to the initiating stimulus. In particular, TLR4 activates transcription of NF-kappaB- and interferon regulatory factor (IRF)-3/IFN-inducible genes in macrophages and dendritic cells in a MyD88-independent manner, potentially by another TIR domain-containing adaptor protein, TIRAP. Data has suggested that TLRs 1, 2, and 6 can also use this alternative adaptor molecule. A third TIR-containing adaptor molecule (TICAM-1), which binds TLR3, inducing interferon-β

has also been described. Suppression of signaling may also occur by association of either of the proteins Tollip or IRAK-M with TRAF6. TLRs have been found to heterodimerize with one another, which may also contribute to the repertoire of specificities. It remains to be determined whether differences in adaptor usage and involvement of suppressor molecules can account for the activation of distinct immune responses among the TLRs. It is plausible other unidentified adaptors play a role in TLR signaling or that combinations of adaptor usage provide more distinct signals than those involving individual TIR proteins.

4 Cellular Effector Mechanisms

4.1 Cell-mediated Inflammation

Myeloid cells are critical to the defenses of the innate immune system. These include the monocyte-derived macrophages and dendritic cells and the granulocytic polymorphonuclear neutrophils (PMN), basophils, and eosinophils. These are the first cells to counter infectious processes by internalizing and degrading foreign material. The lymphoid NK cells also play a role in innate immune responses to viral infection.

Macrophages are most concentrated in lymphoreticular organs (the lymph nodes and spleen) but are present throughout the body, forming morphologically distinct populations in different tissues. Tissue histiocytes, osteoclasts, hepatic Küpffer cells, as well as microglial cells of the central nervous system are all considered macrophages. These phagocytic cells survey the blood and connective tissue for potential pathogens. Macrophages are usually the first cell type to encounter pathogen in tissues. Upon encountering a microbe, the cells can engulf and kill it directly, in the process of phagocytosis, or can recruit other myeloid cells, especially PMNs, to the sites of damage or infection. Cytokines secreted by macrophages, including IL-6, 8, and 12 and TNF-α, are all inducers of fever and inflammation associated with an acute-phase response. Secretion of TNF-α and IL-1 alter expression of vascular adhesion molecules such as E-selectin and VCAM-1. The leukocyte CD18 complex integrins allow adhesion to the vascular endothelium. These include CD11a/CD18, LFA-1 and CD11b/CD18, or Mac-1. Interactions mediated by these molecules between the leukocytes and vessel walls allow firm adhesion and diapedesis. Activation of the macrophage by IFNγ and TNF-α from NK cells allow it to kill most intracellular pathogens. Macrophages are also able to present antigens to T-lymphocytes to activate adaptive immune responses and to modulate the type of response by the production of other cytokines. These include secretion of IL-12 for Th1 and IL-10 for Th2 responses. Consequently, macrophages provide one link between innate and specific immunity.

Dendritic cells are specialized phagocytes that are particularly efficient at presenting antigen to T-lymphocytes. Dendritic cells regularly ingest bits of their immediate environment by endocytosis and macropinocytosis. When a pathogen is detected, TLR stimulation and signaling initiates maturation, enabling the DCs to produce costimulatory molecules and enhancing the capacity to present antigen. MHC molecule expression is upregulated and cells proceed to the lymph nodes for T-cell priming. In this way, DCs are poised for activation of adaptive immune responses. Like the macrophages, DCs

take on multiple identities, such as the Langerhans cells of the skin and plasmacytoid cells of the spleen. However, DCs are less numerous than the macrophage.

Polymorphonuclear neutrophils are the most abundant white blood cells in circulation that play a critical role in the destruction of circulating pathogens. These cells are not normally found in tissues but will migrate in response to the release of chemotactic factors by injured tissues or in response to the activation of other cells. At these inflammatory sites, PMNs may become activated and produce large amounts of reactive oxygen species, and release a variety of other toxic products from cytosolic granules. These cells have a short lifetime, spending approximately six hours in circulation, before undergoing apoptosis.

Eosinophils are responsive to cytokines, and themselves produce a variety of chemokines that mediate inflammation. Under normal conditions, these cell types are found in limited numbers, but counts are elevated during parasitic infection and during an allergic adaptive immune response. Most reside in the subepithelium of respiratory, digestive, and urinary tracts. Mast cells reside in nearly all tissues of the body and play major roles in the defense against parasites and in allergic inflammation. Basophils are very similar to mast cells, except that they reside in blood. These cells are distinguishable by large cytoplasmic granules and high-affinity IgE receptors. Degranulation of mast cells and basophils releases histamine, other vasoactive mediators, and certain proinflammatory cytokines, for example, TNF-α and IL-5. These compounds trigger vasodilation and increase vascular permeability, mucus secretion, bronchiolar smooth muscle constriction, leukocyte recruitment, platelet activation, tissue destruction, fever, and pain.

Natural Killer cells migrate in response to chemotactic factors and kill virally infected cells and tumor cells. Binding to target cells is followed by release of perforins into the cell membrane and lysis or stimulation of programmed cell death.

4.2 Oxygen-independent Destruction of Pathogens

Phagocytosis of pathogens occurs following ligation of certain surface receptors on the cell (Fig. 5). Macrophages and neutrophils are highly efficient at phagocytosis, utilizing the external receptors FcR, complement receptors, mannose receptors, and TLRs to mediate binding to foreign or opsonized material. It is an active process involving actin and microtubule restructuring, which surrounds the pathogen with the phagocyte membrane and internalizes it within a phagosome. Intracellular killing takes place when this vesicle is acidified and introduced to lysosomal granules in the macrophage or to azurophil granules in neutrophils, which contain toxic enzymes, proteins, and peptides. Lysosomal proteases and glycolases will disturb parasitic, bacterial, and viral membrane function. Lysozyme does so by cleaving the $\beta-(1-4)$ glycosidic bond within the peptidoglycan of gram-positive bacterial cell walls. Insertion of defensins into pathogen membranes also disrupts membrane permeability. Acid hydrolases and neutral proteinases digest multiple components of bacteria. The antibacterials transferrin and lactoferrin sequesters free iron, thereby preventing the growth of ingested microorganisms.

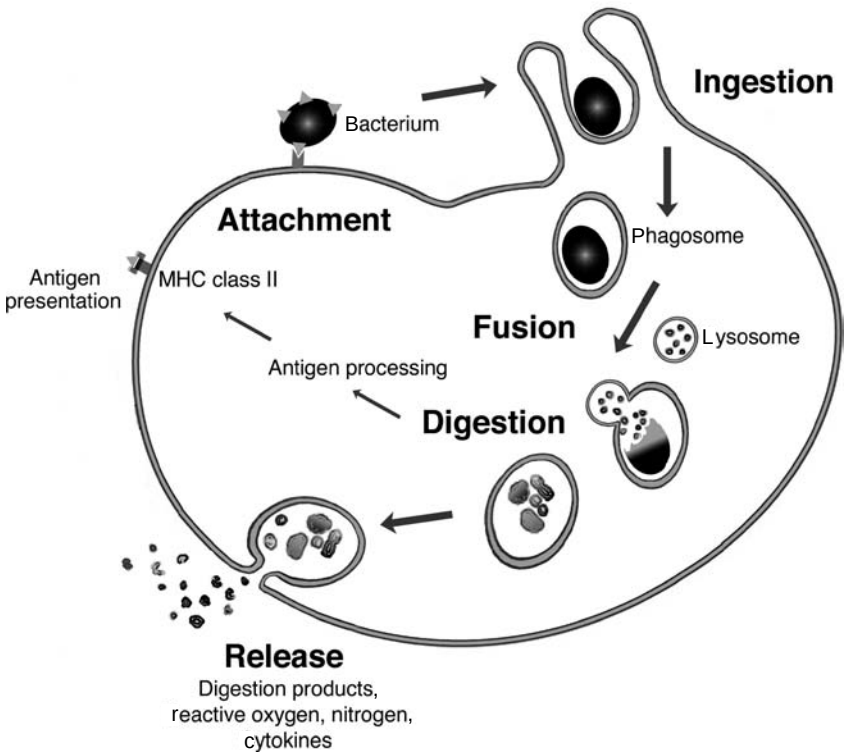

Fig. 5 Processes of phagocytosis. Attachment of the bacterium via FcRs, CRs, TLRs, and other receptors is followed by ingestion of the bacterium, forming a phagosome. Acidification of the phagosome and fusion with the lysosome allows toxic products to kill microbes and proteolytic enzymes to digest the ingested material. Release of digestion products from the cell is accompanied by antigen presentation and release of effectors.

While azurophil granules function predominantly intracellularly within the phagolysosomal vacuole, neutrophil-specific granules are prone to release their contents to the extracellular milieu and appear to have a significant role in the initiation of inflammation.

4.3 Oxygen-dependent Killing

Phagocytes may also undergo a respiratory burst to help kill engulfed microorganisms. This involves a reduction and metabolism of oxygen by cellular enzymes to superoxide ($O_2^{\bullet -}$) and hydrogen peroxide (H_2O_2), ultimately forming toxic products such as hypohalous acid (HOCl/HOBr) and hydroxyl radical (•OH). Subunits of the enzyme NADPH oxidase assemble in the phagosomal and cellular membranes in response to neutrophil activation. This activation arises from the interaction between opsonized particles or pathogens and neutrophil receptors. The phagocyte oxidase transfers electrons from NADPH to oxygen, forming superoxide anion (Eq. 1). Superoxide may spontaneously dismutate or

the enzyme superoxide dismutase may catalyze the formation of H_2O_2 and ground state O_2 (Eq. 2). Myeloperoxidase in neutrophil azurophilic granules catalyzes the formation of HOCl from H_2O_2 and chloride ions (Eq. 3). In the same manner, brominated species may also be formed, however, at much lower concentrations.

$$2O_2 + NADPH \longrightarrow 2O_2{\bullet}^- + NADP^+ + H^+ \quad (1)$$

$$2H^+ + 2O_2{\bullet}^- \longrightarrow 2H_2O_2 + O_2 \quad (2)$$

$$X^- + H_2O_2 + H^+ \longrightarrow HOX + H_2O \qquad X = Cl, Br \quad (3)$$

Superoxide ion, hydrogen peroxide, and hypochlorite are all efficient oxidizing agents that can kill bacteria and fungi by oxidizing membrane lipids and proteins. Further reactions of these oxidants catalyzed by transition metals can produce •OH (Eq. 4). Hydroxyl radical is the most biologically reactive of the aforementioned oxidants. Singlet oxygen ($^1O_2$) is an excited state of dioxygen that debatably forms downstream of HOCl or •OH, which has deleterious effects on macromolecules (Eqs. 5 and 6).

$$O_2{\bullet}^- + HOCl \longrightarrow O_2{\bullet}^- + {\bullet}OH + Cl^- \quad (4)$$

$$OCl^- + H_2O_2 \longrightarrow {}^1O_2 + Cl^- + H_2O \quad (5)$$

$$O_2{\bullet}^- + {\bullet}OH \longrightarrow {}^1O_2 + HO^- \quad (6)$$

Singlet oxygen is also thought to be formed during the dismutation of peroxyl radical and during the decomposition of 1,2-hydroperoxides. The presence of $^1O_2$ during the oxidative burst provides a means by which the proposed oxidation of water by antibodies that has been postulated leads to trioxygen species generation (Eq. 7).

$$x^1O_2 + yH_2O \longrightarrow [H_2O_3(y-1)H_2O] \longrightarrow O_3 + H_2O_2 + (x-1)O_2 \quad (7)$$

While this reaction has mainly been characterized as a function of antibodies, other immunoglobulin fold-containing proteins may also catalyze the process. Its contribution to innate defenses and immunity as a whole has not yet been fully elucidated. The reactive oxygen species produced by activated phagocytes act in concert to effect microbial killing. Reactive nitrogen species such as nitric oxide (NO) and peroxynitrite ($OONO^-$) are also important for immune defense in that they mediate intercellular signaling and are also toxic to bacteria.

5 Humoral Effector Mechanisms

5.1 Complement

The complement system is a component of innate immunity concerned with the lysis and clearance of foreign particles and host debris. Activation of this system involves binding and cleavage of multiple proteins and glycoproteins, both soluble and membrane bound. The formation of active enzyme complexes and deposition of proteolytic fragments in this process recruits phagocytes to sites of infection, prompts direct cell lysis, and opsonizes pathogens for phagocytosis. Induction of inflammatory mediators also occurs as a result of complement activation (Fig. 6).

Complement enzyme precursors are activated locally and initiate inflammatory events in response to specific stimuli such

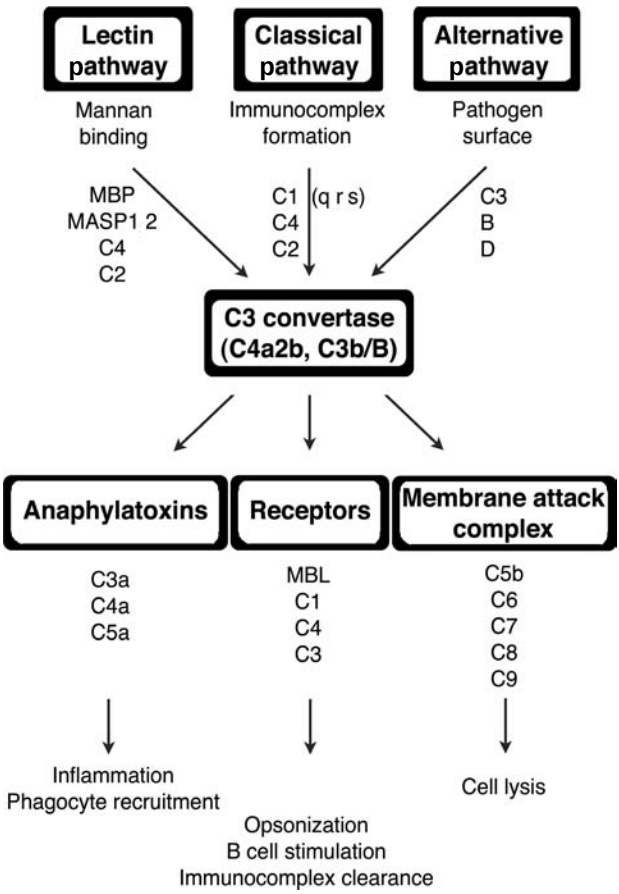

Fig. 6 Complement pathways and proteins. Complement regulatory proteins not shown include C1INH, C4bp, CR1, MCP, DAF, H, I, P, and CD59.

as pathogen binding and immunocomplex formation. This may occur by three pathways. In the classical activation pathway, the pentameric protein complex C1, made up of a C1q subunit and two subunits each of C1r and C1s, binds directly to pathogens or to antibodies that are in complex with pathogens. In the latter case, C1 links adaptive immunity to the innate complement system. Proteolytic cleavage of C1r and C1s activates C1 to sequentially cleave other complement proteins (C4 then C2) to form an active C3 convertase (C4b2a). The mannan-binding lectin pathway uses a similar route to form an identical C3 convertase: serum PRRs bind mannose residues on bacterial or viral carbohydrates, forming an MBL/MASP complex, to activate complement proteins C4 and C2 and to form an active C3 convertase. Finally, alternative activation occurs when complement component C3 becomes spontaneously activated, for example, by microbial compounds LPS or

LTA, and gets deposited onto the surface of a pathogen. Properdin is a positive regulator of the alternative pathway of complement activation, released by peripheral blood cells and the endothelium. Binding and cleavage of plasma factor B produces a fluid-phase C3 convertase.

The C3 convertases cleave C3 to produce C3a, an anaphylactic by-product that mediates inflammatory responses, and C3b, which acts as an opsonin for phagocytes with complement receptors. C3b also activates a C5 convertase, thereby producing C5a, an important inflammatory mediator, and C5b, which initiates late events of complement. These entail polymerization of the terminal complement components, C6 to C9, to form porelike membrane attack complexes (MACs), leading to loss of bacterial cell integrity and lysis. Phagocyte ingestion of complement-opsonized pathogens is mediated by specific complement receptors. There are five known types of complement receptors. The receptor for C3 (CR1) is present on neutrophils and macrophages and can promote binding and phagocytosis.

The complement system is efficient in the defense against many viral and bacterial infections as well as in the removal of cellular debris. However, if improperly regulated, complement can cause severe host tissue damage, as in rheumatoid arthritis, acute hemolytic anemia, or glomerulonephritis. Other conditions including mechanical injury and myocardial infarction can stimulate inopportune complement-mediated damage as well.

5.2 Antimicrobial Peptides

Cationic antimicrobial peptides produced by vertebrates, arthropods, plants, insects, and bacteria are small stretches of less than 50 amino acids that possess activity against bacteria, fungi, and protozoa. Of the more than 800 antimicrobial peptides submitted in the Antimicrobial Sequences Database (http://www.bbcm.univ.trieste.it/~tossi/amsdb.html), over 175 are produced by insects. These peptides are particularly important in the immune defenses of the well-studied fruit fly, *Drosophila melanogaster*. Proteins include the antibacterials Andropin, Attacin, Cecropin, Diptericin (gram negative), Defensin (gram positive), and Drosocin; and the antifungals Drosomycin and Metchnikowin, among many others. The genes encoding most of these protein effectors are induced in response to infection and the peptides secreted from *Drosophila* fat bodies. Andropin is the exception to the rule in that it is constitutively expressed in males. In contrast, most mammalian antimicrobial proteins, including the cryptidins or α-defensins present in the intestines and β-defensins produced in the skin and respiratory tract, are ubiquitously produced as part of innate constitutive defenses.

Two general mechanisms of action of these molecules have been indicated and segregated on the basis of being either membrane disruptive or nonmembrane disruptive. Initial associations of these peptides with microbes occur through electrostatic interactions, for example, between the peptide and negatively charged LPS in bacterial cell membranes. Destabilization of the outer membrane creates sites on the cytoplasmic membrane that are more susceptible to translocation or membrane disruption. It is postulated that transient pores are formed by the peptide that span the membrane, resulting in membrane depolarization and disintegration. Alternatively,

peptides may form micellar aggregates in the outer leaflet, which collapse allowing translocation into the cytoplasm. Inside the bacterium, peptides interact with nucleic acids and cellular proteins, inducing transcriptional changes and inhibiting further synthesis of macromolecules.

The peptides are classified on the basis of secondary structure into four major classes: β-sheet, α-helix, loop, or extended peptides (Fig. 7). The majority of known peptides belong to the first two groups, according to circular dichroism spectroscopy. High-resolution structures of these peptides, obtained by 2-d ^{1}H NMR, exist for about 50 of these molecules. The α-helical antimicrobials are characterized by an α-helical conformation containing a slight bend in the center of the molecule. This bend may be crucial for suppressing undesirable hemolytic activity. Structure–activity studies of magainin *2*, a typical α-helical peptide, suggest no difference in D- and L-enantiomers, ruling out chiral receptors or enzyme targets

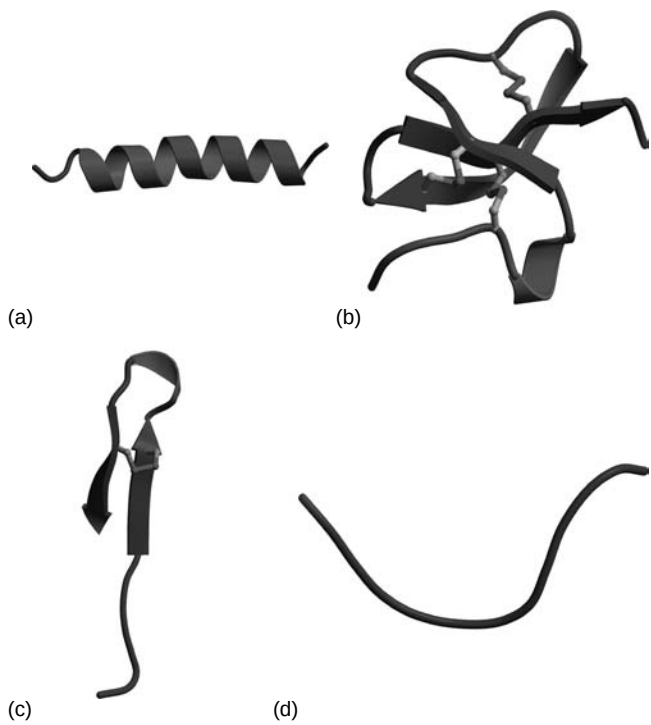

Fig. 7 Antimicrobial peptides. Representative member of each structural class: (a) magainin 2, alpha-helix; (b) β-defensin, beta-sheet; (c) thanatin, loop; (d) indolicidin, extended. Pictures generated by Matt Kelker, using bobscript and raster3D. Coordinates obtained from the RCSB Protein Data Bank under the accession codes 1BNB (defensin), 1G89 (indolicidin), 2MAG (magainin), and 8TFV (thanatin). Raster3D: Merritt & Bacon (1997) *Meth. Enzymol.* **277**, 505–524. bobscript: Esnouf, RM (1997) *J. Mol. Graph. Model.* **132–134**, 112–113.

as mechanisms of action. Furthermore, an increase in cationic charge, up to +5, leads to an increase in antimicrobial activity. This is presumably due to enhanced association with the negatively charged cell surface. Peptides containing antiparallel β-sheets, such as tachyplesin, disrupt lipid organization and undergo membrane translocation. Stabilizing disulfide bonds within the structure suggest that flexibility is pertinent for translocation to occur. The class of peptides with extended conformations are high in proline or glycine content and hence lack overall secondary structure. These extended peptides tend to form hydrogen bonds and Van der Waals interactions with membrane lipids. Evidence exists that extended peptides, such as indolicidin, translocate into the cytoplasm to inhibit DNA synthesis; however, the mechanism has not been explicitly determined. Finally, loop peptides such as the insect antimicrobial thanatin contain a single disulfide, amide, or isopeptide bond. The exact mechanism of killing by thanatin is unknown, but the molecule evidently does not perturb membrane permeability. The antimicrobial mechanism appears to be target specific, as only L-thanatin

Tab. 3 Effects of cytokines on the innate immune system.

Cytokine	*Origin*	*Function*
IL-1α and -1β	Monocytes, macrophages, epithelium, endothelium, fibroblasts	costimulation of APCs and T cells, inflammation, adhesion and fever, acute-phase response
IL-2	NK cells, activated TH_1 cells	proliferation of B cells and activated T cells, NK functions
IL-4	Mast cells, basophils TH_2 cells	B-cell proliferation, eosinophil and mast-cell growth and function, IgE and class-II MHC expression on B cells, inhibition of monokine production
IL-5	TH_2 and mast cells	B-cell and eosinophil growth and function
IL-6	Activated TH_2 cells, monocytes, macrophages, epithelium, endothelium	acute-phase response, B-cell proliferation, thrombopoiesis, synergistic with IL-1 and TNF-α
IL-8	Monocytes, macrophages, epithelium	chemoattractant for neutrophils and T cells
IL-10	Macrophages, activated TH_2 cells, CD8+ B cells	inhibits cytokine production (IL-1, -2 and TNF-α), promotes B-cell proliferation and antibody production, suppresses cellular immunity, mast-cell growth
IL-12	Monocytes, macrophages	NK and TH_1 cell activation
IFN-α and -β	DCs, epithelium, fibroblasts	Antiviral, increases MHC I expression, suppresses proliferation
IFN-γ	Macrophages, NK cells, activated TH_1 cells	Antiviral, increases MHC I &II expression, suppresses proliferation
TNF-α	Macrophages, NK cells, mast cells, epithelium	Phagocyte activation, fever, antitumor activity
G-CSF, M-CSF, GM-CSF	Various	Growth and maturation of PMNs and monocytes

is effective against gram-negative bacteria but both enantiomeric forms of the peptides possess activity against gram-positive and fungal species.

Cationic charge and amphipathic conformation appear to be the greatest requirements for antimicrobial activity. Significantly, components of the target membrane may also determine the efficacy of these peptides. Further study of peptide–membrane interactions may explain activity of these molecules with individual organisms.

5.3 Cytokines

Whereas peptides seem to predominate in the immune defenses of insects, cytokine-mediated inflammation is highly significant in mammals. Cytokines are primarily secreted by leukocytes, stimulating both humoral and cellular immune responses as well as activation of phagocytosis (Table 3). The proinflammatory cytokines TNF-α, IL-1, and IL-6, and chemokines such as IL-8 are released in response to encounters with pathogens, specifically by PAMP/PRR interactions. TNF-α is the primary mediator of acute inflammation, and the interleukins enhance this action by increasing endothelial adhesion, increasing chemokine secretion, activating cell-mediated immunity, and stimulating hepatocytes to secrete acute-phase proteins. Type-I interferons produced by virus-infected cells inhibit viral replication and kill infected cells by degrading mRNA. Interferons also activate TH_1 cells, which activate cytotoxic T cells and NK cells, which kill pathogens.

See also Antibody Molecules, Genetic Engineering of; Antigen Presenting Cells (APCs); Autoantibodies and Autoimmunity; Combinatorial Phage Antibody Libraries; Immune Defence, Cell Mediated; Immunoassays; Innate Immunity; Nucleic Acid and Protein Single Molecule Detection and Characterization.

Bibliography

Books and Reviews

Aderem, A., Ulevitch, R.J. (2000) Toll-like receptors in the induction of the innate immune response, *Nature* **406**, 782–787.

Akira, S., Takeda, K., Kaisho, T. (2001) Toll-like receptors: critical proteins linking innate and acquired immunity, *Nat. Immunol.* **2**, 675–680.

Beutler, B. (2004) Innate immunity: an overview, *Mol. Immunol.* **40**, 845–849.

Cook, D.N., Hollingsworth II, J.W., Schwartz, D.A. (2003) Toll-like receptors and the genetics of innate immunity, *Curr. Opin. Allergy Clin. Immunol.* **3**, 523–529.

Coutinho, A., Poltorack, A. (2003) Innate immunity: from lymphocyte mitogens to toll-like receptors and back, *Curr. Opin. Immunol.* **15**, 599–602.

Fearon, D.T. (1999) Innate immunity and the biological relevance of the acquired immune response, *Q. J. Med.* **92**, 235–237.

Imler, J.-L., Hoffmann, J.A. (2001) Toll receptors in innate immunity, *Trends Cell Biol.* **11**, 304–311.

Kimbrell, D.A., Beutler, B. (2001) The evolution and genetics of innate immunity, *Nat. Rev. Genet.* **2**, 255–267.

Means, T.K., Golenbock, D.T., Fenton, M.J. (2000) Structure and function of toll-like receptor proteins, *Life. Sci.* **68**, 241–258.

Medzhitov, R., Janeway Jr., C. (2000) Innate immune recognition: mechanisms and pathways, *Immunol. Rev.* **173**, 89–97.

Medzhitov, R., Janeway Jr., C. (2002) Innate immune recognition, *Annu. Rev. Immunol.* **20**, 197–216.

O'Neill, L.A. (2002) Signal transduction pathways activated by the IL-1 receptor/toll-like receptor superfamily, *Curr. Top. Microbiol. Immunol.* **270**, 47–61.

Primary Literature

Adachi, O., et al. (1998) Targeted disruption of the MyD88 gene results in loss of IL-1- and IL-18-mediated function, *Immunity* **9**, 143–150.

Dziarski, R., Tapping, R.I., Tobias, P.S. (1998) Binding of bacterial peptidoglycan to CD14, *J. Biol. Chem.* **273**, 8680–8690.

Horng, T., et al. (2002) The adaptor molecule TIRAP provides signaling specificity for toll-like receptors, *Nature* **420**, 329–333.

Jurk, M., et al. (2002) Human TLR7 or TLR8 independently confer responsiveness to the antiviral compound R-848, *Nat. Immunol.* **3**, 499.

Kobayashi, K., et al. (2002) IRAK-M is a negative regulator of toll-like receptor signaling, *Cell* **110**, 191–202.

Medzhitov, R., et al. (1997) A human homologue of the Drosophila Toll protein signals activation of adaptive immunity, *Nature* **388**, 394–397.

Miyake, K. (2003) Innate recognition of lipopolysaccharide by CD14 and toll-like receptor 4-MD-2: unique roles for MD-2, *Int. Immunopharmacol.* **3**, 119–128.

Ozinsky, A., et al. (2000) The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13766–13771.

Palaniyar, N., Nadesalingam, J., Clark, H., Shih, M.J., Dodds, A.W., Reid, K.B. (2004) Nucleic acid is a novel ligand for innate immune pattern recognition collectins surfactant proteins A and D and mannose-binding lectin, *J. Biol. Chem* Published online before print.

Powers, J.S., Hancock, R.E.W. (2003) The relationship between peptide structure and antibacterial activity, *Peptides* **24**, 1681–1691.

Radons, J., Dove, S., Neumann, D., Altmann, R., Botzik, A., Martin, M.U., Falk, W. (2003) The interleukin 1 (IL-1) receptor accessory protein Toll/IL-1 receptor domain, *J. Biol. Chem.* **278**, 49145–49153.

Tabeta, K., et al. (2004) Toll-like receptors 9 and 3 as essential components of innate immune defense against mouse cytomegalovirus infection, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3516–3521.

Underhill, D.M., Ozinsky, A. (2002) Toll-like receptors: key mediators of microbe detection, *Curr. Opin. Immunol.* **14**, 103–110.

Welch, R.A., et al. (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 17020–17024.

Wentworth Jr., P., McDunn, J.E., Wentworth, A.D., Takeuchi, C., Nieva, J., Jones, T., Bautista, C., Ruedi, J.M., Gutierrez, A., Janda, K.D., Babior, B.M., Eschenmoser, A., Lerner, R.A. (2002) Evidence for antibody-catalyzed ozone formation in bacterial killing and inflammation, *Science* **298**, 2195–2199.

Yamamoto, M., et al. (2002) Cutting edge: a novel Toll/IL-1 receptor domain-containing adaptor that preferentially activates the IFN-beta promoter in the toll-like receptor signaling, *J. Immunol.* **169**, 6668–6672.

Yamamoto, M., et al. (2003) TRAM is specifically involved in the toll-like receptor 4-mediated MyD88-independent signaling pathway, *Nat. Immunol.* **4**, 1144–1150.

Zhang, G., Ghosh, S. (2002) Negative regulation of toll-like receptor-mediated signaling by Tollip, *J. Biol. Chem.* **277**, 7059–7065.

Zhang, D., et al. (2004) A toll-like receptor that prevents infection by uropathogenic bacteria, *Science* **303**, 1522–1526.

Immuno-PCR

Takeshi Sano[1] *and Charles R. Cantor*[2,3]
[1] *Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA*
[2] *Sequenom, Inc., San Diego, CA, USA*
[3] *Boston University, Boston, MA, USA*

1 Concept 448

2 Reagents 449
2.1 Molecular Linkers between Antibody and Marker DNA 449
2.2 Antibodies 450
2.3 Marker DNA 450
2.4 Conjugation of Marker DNA to Molecular Linkers 451

3 Methods 452
3.1 Basic Microtiter Plate Format 452
3.2 Sandwich Assay Format 454
3.3 Preconjugation Format 456
3.4 Other Formats 457
3.5 Sensitivity 458
3.6 Detection of PCR Products 458
3.7 Quantitation of Antigen (Epitope) 460
3.8 Contamination Control 460
3.9 General Precautions 460

4 Applications and Perspectives 461

Bibliography 462
Books and Reviews 462
Primary Literature 462

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

Keywords

Avidin
A tetrameric glycoprotein found in chicken egg white, which specifically binds biotin with an extremely high affinity ($K_d \sim 10^{-15}$ M) (see also Streptavidin below).

Biotin
A water-soluble vitamin (vitamin H) that can be incorporated relatively easily into various biological materials.

Biotinylation
The incorporation of biotin or its derivatives into biological material.

Protein A
A cell wall protein of *Staphylococcus aureus*, which specifically binds the Fc domain of an immunoglobulin G molecule without disturbing its antigen-binding ability.

Streptavidin
A tetrameric nonglycosylated protein produced by *Streptomyces avidinii*, which specifically binds biotin with an extremely high affinity ($K_d \sim 10^{-15}$ M).

Immuno-PCR is an antigen detection system, in which the polymerase chain reaction (PCR) is used to amplify a segment of marker DNA that has been attached specifically to antigen–antibody complexes. Because of the enormous amplification capability and specificity of PCR, immuno-PCR allows considerable enhancement in detection sensitivity of a specific antigen over conventional antigen detection systems, such as enzyme-linked immunosorbent assays and radioimmunoassays. A variety of superior characteristics and the versatility of immuno-PCR, including high but controllable sensitivity, simplicity, and easy contamination control, offer great promise for its application in various aspects of biological and medical sciences, as well as clinical diagnostics and forensic medicine.

1 Concept

The concept of immuno-PCR (polymerase chain reaction), illustrated in Fig. 1, is quite simple, and is similar to those of conventional antibody-based antigen detection systems, such as enzyme-linked immunosorbent assays (ELISA) and radioimmunoassays (RIA).

In the original immuno-PCR scheme, a molecular linker, which has a bispecific binding affinity for antibody and DNA, is used to attach a marker DNA molecule specifically to an antigen–antibody complex. A segment of the attached marker DNA is amplified by PCR with appropriate primers, and the resulting PCR products are analyzed. The presence of specific PCR

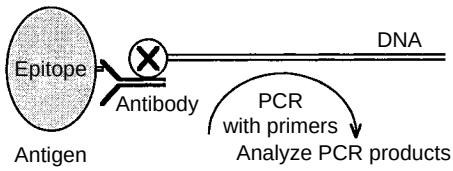

Fig. 1 Concept of immuno-PCR. See Sect. 1 for explanation. [Reprinted, with permission, from Sano, *Cell Technol.* **13**:77–80 (1994)].

products demonstrates that the marker DNA molecules are attached specifically to antigen–antibody complexes, indicating the presence of antigen.

2 Reagents

2.1 Molecular Linkers between Antibody and Marker DNA

In conventional antigen detection systems, secondary antibodies, directed against a primary antibody, are generally used to attach labels, such as reporter enzymes in ELISA or radioactive labels in RIA, to complexes between antigen and primary antibodies. In contrast, immuno-PCR uses specific molecular linkers to attach marker DNA specifically to antigen–antibody complexes. Thus, such molecular linkers are key components in immuno-PCR.

Molecular linkers used in immuno-PCR must possess bispecific binding affinity for antibody and DNA to attach marker DNA specifically to antigen–antibody complexes.

One unique and particularly versatile molecular linker for immuno-PCR is a streptavidin-protein A chimera. This streptavidin-protein A chimera was designed by genetic engineering of the streptavidin and protein A genes, which were isolated from *Staphylococcus aureus* and *Streptomyces avidinii*, respectively. A part of the protein A gene encoding two IgG-binding domains was fused to a modified streptavidin gene. The gene fusion can be expressed in *Escherichia coli* by using the T7 expression system, and the expressed streptavidin-protein A chimera can be purified to homogeneity by a simple purification procedure. This chimera possesses two independent binding abilities; one is to biotin, derived from the streptavidin moiety, and the other is to the Fc domain of an immunoglobulin G (IgG) molecule, derived from the protein A moiety. The chimera forms a subunit tetramer, which binds four biotins and four IgG molecules. This bifunctional multivalent binding specificity for biotin and antibody allows the specific conjugation of any biotinylated DNA molecule to antigen–antibody complexes.

Other simple, yet very useful molecular linkers, which can be used in immuno-PCR, are monospecific, multivalent binding molecules, such as streptavidin and avidin, each molecule of which can bind up to four biotins. Although these proteins do not bind antibody and biotinylated marker DNA independently, biotinylated marker DNA can be attached to antigen–antibody complexes by the use of biotinylated antibodies. In these methods, the biotinylated antibody is first bound to antigen (formation of complexes between antigen and biotinylated antibody). Then, streptavidin or avidin is directed to biotin conjugated to the antibody (formation of complexes consisting of antigen, biotinylated antibody, and streptavidin or avidin). Finally, biotinylated marker DNA is directed to

free biotin-binding sites of streptavidin or avidin molecules that have been attached to complexes between the antigen and the biotinylated antibody. This is followed by amplification of a segment of bound biotinylated marker DNA by PCR.

Specific chemical conjugates of antibody with marker DNA can also be used in immuno-PCR without the need for molecular linkers. Such chemical conjugates can be produced by attaching derivatized DNA molecules covalently to antibody molecules. Such chemical conjugates between antibody and marker DNA should be very useful for large-scale analysis of selected antigens, such as in clinical diagnostics, since they can be produced in large quantities and stored until used. However, because of laboriousness of many chemical conjugation procedures and the need for purification of functional conjugates, the overall versatility of performance of immuno-PCR could be reduced, limiting their use in certain settings where a number of different antigens need to be analyzed.

2.2 Antibodies

Both monoclonal and polyclonal antibodies can be used in immuno-PCR. In addition, genetically engineered antibodies, including single-chain antibodies, can also be used. When the streptavidin-protein A chimera is used as a molecular linker between antibody and marker DNA, the antibodies used must have the Fc portion, to which the protein A moiety of the streptavidin-protein A chimera can bind. The Fc portion of certain IgG subclasses does not have sufficient binding affinity for protein A. For example, mouse IgG_1, sheep immunoglobulins, and horse immunoglobulins have only very low affinity for protein A. These antibodies cannot be used directly in immuno-PCR with the streptavidin-protein A chimera as a molecular linker between primary antibody and biotinylated marker DNA. However, such antibodies can be used indirectly by letting them serve as targets for secondary antibodies, which have high binding affinity for protein A, directed against the primary antibodies.

Because of the extremely high sensitivity of immuno-PCR, very low concentrations of antibody are generally used. For example, the concentration of primary antibody can be diluted by several orders of magnitude below those used in conventional ELISA. For given antigen samples and antibody preparations, optimal concentrations of antibody, where the highest sensitivity and signal-to-noise ratios are obtained, should be determined by calibration using different concentrations of antibody.

2.3 Marker DNA

Marker DNA used in immuno-PCR is purely arbitrary. Therefore, any DNA molecules can be used unless their amplified regions are contained in sample-derived nucleic acids.

The marker DNA used most frequently in immuno-PCR is the end-biotinylated linear DNA molecule. Incorporation of multiple biotins into a marker DNA molecule should be avoided. Since molecular linkers used in immuno-PCR are generally multivalent, conjugation of marker DNA containing multiple biotins to linker molecules causes aggregation, resulting in the generation of background signals.

Conventionally, plasmid DNA is used to prepare marker DNA. The use of plasmid

DNA offers ease in handling and manipulations. Purified plasmid DNA is digested with two or more restriction endonucleases. Biotin is incorporated into one of the termini of the linear plasmid DNA by filling-in reactions with a template-dependent DNA polymerase in the presence of a biotin-labeled deoxynucleotide, such as biotin-dATP. Generally, Klenow fragment of *E. coli* DNA polymerase I or T7 DNA polymerase and its derivatives are used. Because all four deoxynucleotides containing biotin are available, restriction endonuclease digestion sites, where biotin can be incorporated, can be chosen with great flexibility. After the incorporation of biotin, the resulting end-biotinylated marker DNA should be purified by an appropriate method to remove unincorporated biotin-labeled deoxynucleotides. This step is important to avoid the binding of unincorporated biotin-labeled deoxynucleotides to the streptavidin moiety of the streptavidin-protein A chimera, streptavidin, or avidin, which serves as a molecular linker.

Another simple, efficient method to prepare an end-biotinylated marker DNA is PCR with a biotinylated primer. A segment of a template DNA is amplified by PCR with a set of primers, a biotinylated primer and a nonbiotinylated primer. In this way, it is important to remove unincorporated biotinylated primers to avoid the binding of the biotinylated primers to the streptavidin moiety of the streptavidin-protein A chimera, streptavidin, or avidin, which is used as a molecular linker. In addition, primer dimers containing biotin, if any, should also be removed.

The efficiency of biotinylation of marker DNA can be estimated by simple gel retardation assays. End-biotinylated marker DNA is mixed with streptavidin or avidin at various ratios. Each of the resulting mixtures is separated by conventional agarose gel electrophoresis. In the presence of excess amounts of streptavidin or avidin, retardation of end-biotinylated marker DNA, which results from the binding of streptavidin or avidin, can be seen. Marker DNAs as large as several kilobase pairs can be analyzed by this method. Larger marker DNAs would not be suitable for immuno-PCR. By using one of the two biotinylation methods described above, the efficiency of biotinylation should be very close to 100%.

End-biotinylated marker DNAs generated by one of the above methods are very stable, and can be stored in a frozen state for many months.

2.4 Conjugation of Marker DNA to Molecular Linkers

Conjugation of end-biotinylated marker DNA to the streptavidin-protein A chimera, used as a molecular linker, can be performed by simply mixing the two components (the chimera and end-biotinylated marker DNA) at a molar ratio of biotin to biotin-binding site of 1 or below. Because of the extremely high affinity of the streptavidin moiety of the streptavidin-protein A chimera for biotin, few unconjugated end-biotinylated marker DNA molecules should be present. The resulting conjugate will contain up to four end-biotinylated marker DNA molecules per chimera molecule.

When streptavidin or avidin is used as a molecular linker, end-biotinylated marker DNA is generally not conjugated to streptavidin or avidin prior to analysis. Instead, streptavidin or avidin, and end-biotinylated marker DNA are applied sequentially during analysis to direct the marker DNA to unoccupied biotin-binding

sites of streptavidin or avidin that has been bound to biotinylated antibody. This allows *in situ* formation of complexes consisting of biotinylated antibody, (strept)avidin, and marker DNA on antigens. Preconjugation of end-biotinylated marker DNA to streptavidin or avidin prior to analysis is possible. However, end-biotinylated marker DNA must be attached to streptavidin or avidin under saturation, since the resulting conjugates must have free biotin-binding sites for their binding to biotinylated antibodies.

3 Methods

There are several formats of immuno-PCR. One of the most versatile ways to perform immuno-PCR uses microtiter plates as solid supports, which allow simultaneous analysis of large numbers of samples. A few representative immuno-PCR formats using the streptavidin-protein A chimera as a molecular linker between antibody and biotinylated marker DNA are described, along with brief descriptions of the modification of the procedures when streptavidin or avidin is used as a molecular linker and when chemical conjugates of antibody with marker DNA is used.

3.1 Basic Microtiter Plate Format

In the microtiter format, antigen is immobilized directly on the surface of microtiter plate wells. The protocol for the detection of antigen immobilized on microtiter plate wells, illustrated schematically in Fig. 2, consists of six steps:

Step 1: immobilization of antigen on the surface of the microtiter plate wells;
Step 2: blocking of free binding sites on the well surfaces with appropriate blockers;
Step 3: binding of antibody to antigen;
Step 4: binding of the streptavidin-protein A chimera, bound to end-biotinylated marker DNA, to antigen–antibody complexes;
Step 5: amplification of a segment of the marker DNA by PCR with appropriate primers; and
Step 6: analysis of PCR products.

The key factor in this protocol is the efficient and complete removal of unbound and nonspecifically bound antibody and marker DNA, which contribute to background signals. Therefore, extensive washing must be carried out after each step. One of the advantages of the use of microtiter plates as a solid support is that washing steps can be performed quite easily even with large numbers of samples.

When streptavidin or avidin is used as a molecular linker, the above procedure is slightly modified. The primary antibody is biotinylated, as described in Sect. 2.2, prior to its application to antigen-containing samples immobilized on the surface of the microtiter plate wells. Step 4, where the streptavidin-protein A chimera containing end-biotinylated marker DNA is applied to antigen-containing samples immobilized on the surface of the microtiter plate wells, is divided into two steps. The first step is the application of streptavidin or avidin to samples. After unbound streptavidin or avidin molecules are removed, end-biotinylated marker DNA is directed to unoccupied biotin-binding sites of streptavidin or avidin molecules that have been bound to biotinylated antibody.

When chemical conjugates of antibody with marker DNA are used, the above

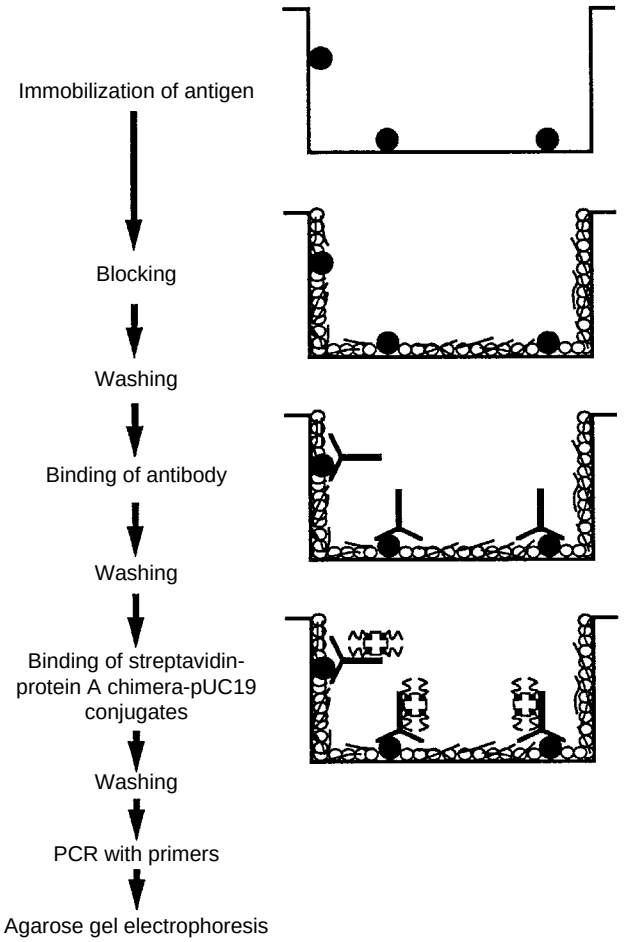

Fig. 2 Schematic illustration of basic microtiter plate format of immuno-PCR. [Reprinted, with permission, from Sano, *Cell Technol.* **13**:77–80 (1994)].

procedure is simplified. Instead of applying antibody and the streptavidin-protein A chimera containing end-biotinylated marker DNA sequentially (Steps 3 and 4), the chemical conjugates can be added to samples in one step, followed by the removal of unbound conjugates and the amplification of a segment of the marker DNA.

An example of the detection of antigen using the basic microtiter plate format with the streptavidin-protein A chimera as a molecular linker is shown in Figs. 3 and 4. In this system, bovine serum albumin (BSA) was used as antigen, and mouse monoclonal antibody (subclass, IgG_{2a}) against BSA was directed to the immobilized antigen. By using 30 amplification cycles in the PCR step, as few as 580 molecules (9.6×10^{-22} mol) of antigen were specifically and reproducibly detected. Direct comparison of

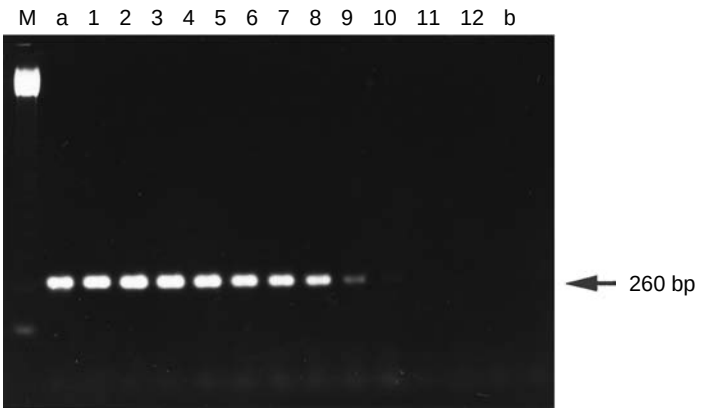

Fig. 3 Detection of BSA immobilized on a microtiter plate by immuno-PCR. Immobilized BSA was detected by immuno-PCR using monoclonal anti-BSA and a streptavidin-protein A chimera containing a 2.67-kb end-biotinylated linear plasmid as a marker. A 260-bp segment of the marker DNA was amplified by PCR, and the resulting PCR products were analyzed by agarose gel electrophoresis, stained with ethidium bromide. Lanes 1 to 9 contain PCR amplification mixtures with immobilized antigen: Lane 1, 96 fmol (5.8×10^{10} molecules); 2, 9.6 fmol (5.8×10^{9} molecules); 3, 960 amol (5.8×10^{8} molecules); 4, 96 amol (5.8×10^{7} molecules); 5, 9.6 amol (5.8×10^{6} molecules); 6, 0.96 amol (5.8×10^{5} molecules); 7, 9.6×10^{-20} mol (5.8×10^{4} molecules); 8, 9.6×10^{-21} mol (5.8×10^{3} molecules); 9, 9.6×10^{-22} mol (5.8×10^{2} molecules). Lanes 10 to 12 are derived from control wells, where no antigen was immobilized. [Reprinted, with permission, from Sano, T., Smith, C.L., Cantor, C.R. (1992) Immuno-PCR: Very sensitive antigen detection by means of specific antibody-DNA conjugates, *Science* **258**, 120–122.]

a conventional antigen detection system, ELISA, using alkaline phosphatase as a reporter enzyme indicated approximately 10^5 times enhancement in detection sensitivity was obtained by using immuno-PCR.

There are several key factors that determine the overall sensitivity and signal-to-noise ratios for given antigen samples and antibody preparations. Generally, optimization of these factors is carried out by empirical methods. The use of the microtiter plate format greatly facilitates this optimization step, because many samples can be easily analyzed simultaneously.

3.2 Sandwich Assay Format

The basic microtiter plate format, described in Sect. 3.1, is likely to generate high background signals when the antigen in physiological fluids containing endogenous immunoglobulins, such as blood samples, is to be detected and the streptavidin-protein A chimera is used as a molecular linker between the antibody and marker DNA. This occurs because of the binding of the protein A moiety of the streptavidin-protein A chimera to sample-derived immunoglobulins.

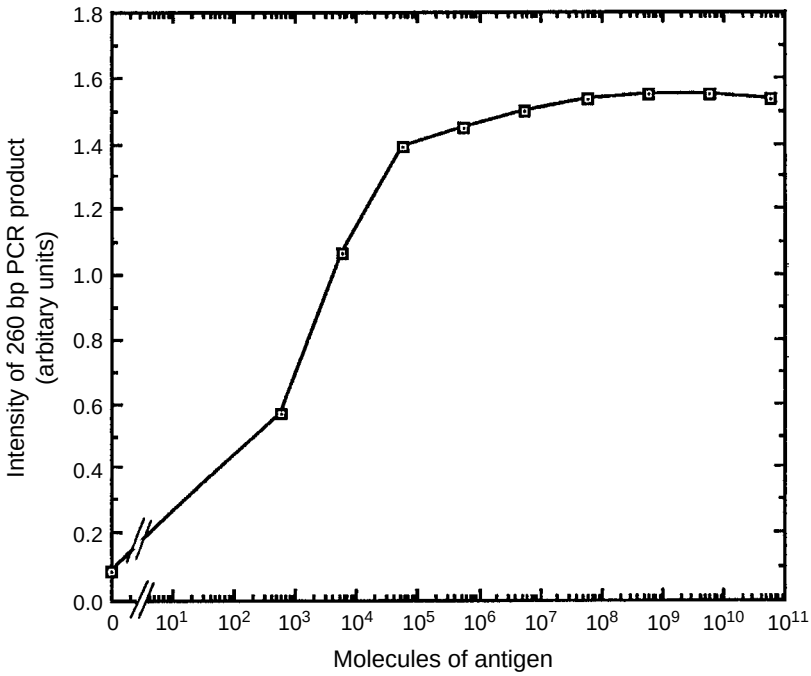

Fig. 4 Quantitation of PCR product generated by immuno-PCR. The 260-bp PCR fragment shown in Fig. 2 was quantitated, and plotted as a function of the number of antigens. Note that PCR amplification was not saturated in a range from 10^2 to 10^5 molecules of antigen, where the number of antigen molecules can be estimated. [Reprinted, with permission, from Sano, T., Smith, C.L., Cantor, C.R. (1992) Immuno-PCR: Very sensitive antigen detection by means of specific antibody-DNA conjugates, *Science* **258**, 120–122.]

One format that can be used for the detection of antigen in immunoglobulin-containing samples using the streptavidin-protein A chimera as a molecular linker is sandwich assays. In this sandwich assay format, the antigen is captured by immobilized binding molecules, which have high binding affinity for the antigen. For example, immobilized primary antibody fragments, such as Fab and $F(ab')_2$, which lack the Fc portion and, thus, do not have binding affinity for protein A, can be used to capture a specific antigen (see Fig. 5). This step allows the purification of specific antigen, thereby removing sample-derived immunoglobulins and other nonantigen molecules. Subsequently, the basic immuno-PCR procedure, described in Sect. 3.1, can be used to detect the captured antigen. To maximize the binding of the secondary antibody to the captured antigen, polyclonal antibodies are used as the secondary antibody. Monoclonal antibodies can also be used as the secondary antibody if the binding of secondary antibodies to captured antigens is inhibited or considerably disturbed by immobilized primary antibodies.

Another attractive, yet universal group of molecules to capture specific antigen is mouse monoclonal antibodies of subclass IgG_1. Antigen can be specifically captured

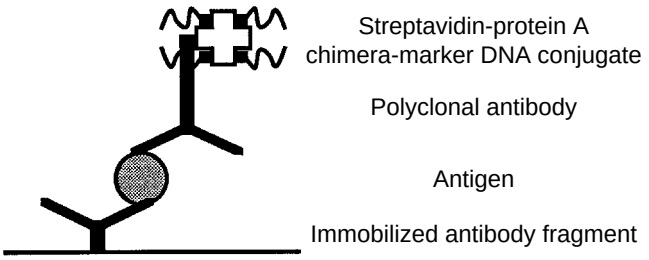

Fig. 5 Schematic illustration of sandwich assay format. In this example, antigen is captured by primary antibody fragments immobilized on a solid support. *See* Sect. 3.2 for explanation.

by immobilized mouse monoclonal antibodies, and then sample-derived immunoglobulins can be removed by washing. Mouse IgG_1 has very low affinity for protein A. Therefore, the binding of the protein A moiety of the streptavidin-protein A chimera to mouse IgG_1 monoclonal antibodies immobilized on a solid support is ignorable. Because many of the monoclonal antibodies available are mouse IgG_1, this format provides greater versatility in the detection of various biological molecules by immuno-PCR.

If specific binding molecules to antigen are available, such molecules can also be immobilized to capture antigen, thereby removing sample-derived immunoglobulins. Such binding molecules include ligands, receptors, and cofactors. For example, ligand molecules can be specifically captured by immobilized receptor molecules of this ligand, and thus sample-derived immunoglobulins can be removed by washing. Then the captured ligand can be detected by the basic immuno-PCR procedure, described in Sect. 3.1, with the use of polyclonal antibodies against the ligand.

When streptavidin or avidin is used as a molecular linker between the antibody and biotinylated marker DNA or when chemical conjugates of antibody with marker DNA is used, the presence of endogenous immunoglobulins in samples should, in principle, pose little problem since these molecules have no apparent binding affinity for immunoglobulins. However, the sandwich assay format often offers greater specificity and higher signal-to-noise ratios, as seen with ELISA, because of the removal of nontarget molecules from samples prior to the application of antibody and marker DNA.

3.3 Preconjugation Format

An attractive alternative method for the detection of antigen in immunoglobulin-containing samples is to preconjugate antibody and end-biotinylated marker DNA to the streptavidin-protein A chimera prior to the application to samples (preconjugation format) (Fig. 6). This is very similar to chemical conjugates of antibody with marker DNA. Because the streptavidin-protein A chimera binds both biotin and IgG stoichiometrically, specific conjugates, consisting of the streptavidin-protein A chimera, antibody, and end-biotinylated marker DNA, can be produced by simply mixing the three components in the appropriate ratios. By using an appropriate mixing ratio of antibody to the streptavidin-protein A chimera, the IgG-binding domain of the protein A moiety of

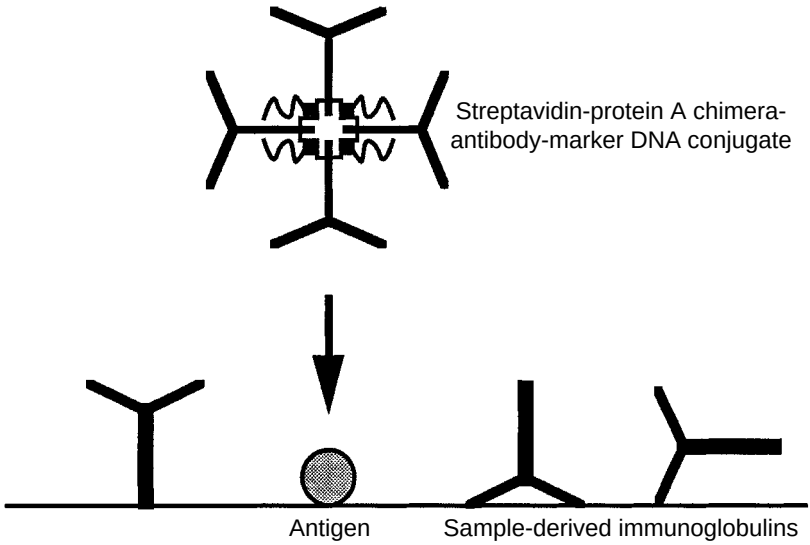

Fig. 6 Schematic illustration of preconjugation format. A specific conjugate, consisting of the streptavidin-protein A chimera, antibody, and end-biotinylated marker DNA, is applied to antigen in an immunoglobulin-containing sample. See Sect. 3.3 for explanation.

the chimera can be saturated. Such conjugates should not bind to immunoglobulins present in samples, because of saturation of the IgG-binding sites with antibodies. In addition, few exchange reactions should occur during a short period of time used in this type of assay, because of the high binding affinity of protein A for antibodies.

In the basic microtiter plate format, described in Sect. 3.1, antibodies and the streptavidin-protein A chimera containing end-biotinylated marker DNA are applied to samples sequentially. In the preconjugation format, specific conjugates consisting of the chimera, antibodies, and end-biotinylated marker DNA, are applied to samples at one time (Fig. 6). Therefore, an additional advantage of the preconjugation format is that the number of steps in a protocol can be reduced. In general, procedures with fewer steps are particularly useful to analyze large numbers of samples simultaneously, such as in clinical diagnostics. In addition, such procedures should facilitate the development of fully automated immuno-PCR assay systems.

3.4 Other Formats

With simple modifications of the protocols described above, several other immuno-PCR formats are available without loss in detection sensitivity. One of these is the combination of the sandwich assay format and the preconjugation format, for example, by using immobilized primary antibody fragments, a specific antigen can be captured and purified. Then, specific conjugates, consisting of the streptavidin-protein A chimera, antibody, and end-biotinylated marker DNA, are applied to the captured antigen. This format not only

allows the specific detection of antigen in immunoglobulin-containing samples but it can also reduce the number of steps in a protocol.

Another attractive solid support that can be used in immuno-PCR is paramagnetic microbeads (or nanoparticles). Paramagnetic microbeads are available with surfaces coated covalently with binding molecules, such as antibodies and streptavidin. Because paramagnetic microbeads can be separated and transferred by the application of magnetic fields, the use of paramagnetic microbeads in immuno-PCR should greatly facilitate the manipulation of samples during analysis.

Another simple, yet attractive format would be the detection of antigen molecules on cell surfaces by immuno-PCR (cell surface immuno-PCR). Centrifugation or filtration steps can be incorporated into the original protocol to separate unbound antibody and marker DNA from sample cells. Because of the extremely high sensitivity of immuno-PCR, this format should allow the analysis and detection of a variety of cell surface molecules with much smaller numbers of cells than conventional antigen detection methods. It should be possible to analyze and detect such cell surface molecules even at single-cell levels.

In situ immuno-PCR has been developed for sensitive detection of antigens in tissue samples. In this method, *in situ* PCR is used to amplify a segment of the marker DNA that has been attached to antigen molecules present in tissue sections. Amplified PCR products are detected by *in situ* hybridization using a probe that is complementary to PCR products. This *in situ* immuno-PCR method has shown greater sensitivity over conventional immunohistochemistry for the detection of antigen in tissue samples.

3.5 Sensitivity

Primarily because of the enormous amplification capability of PCR, immuno-PCR has a sensitivity greater than any other antigen detection method, except for a comparable antibody-based proximity ligation method followed by PCR or rolling circle amplification. In addition to this extremely high sensitivity, overall sensitivity of immuno-PCR can be controlled systematically by varying several key factors. When the streptavidin-protein A chimera is used as a molecular linker, these key factors include concentrations of antibody, concentrations of streptavidin-protein A chimera containing end-biotinylated marker DNA, numbers of PCR amplification cycles, and detection methods for PCR products.

One such example is shown in Figs. 7 and 8. The use of a reduced concentration of the streptavidin-protein A chimera containing end-biotinylated marker DNA shifted the detection range to larger numbers of antigen. Thus, overall sensitivity of immuno-PCR can be controlled by varying the concentration of the streptavidin-protein A chimera bound to end-biotinylated marker DNA.

For given antigen samples and antibody preparations, optimal conditions where maximum sensitivity and signal-to-noise ratios can be obtained should be determined by appropriate calibration. The use of microtiter plates as a solid support greatly facilitates this optimization step, because many conditions can be tested simultaneously.

3.6 Detection of PCR Products

A variety of methods, commonly used in molecular biology and genetic

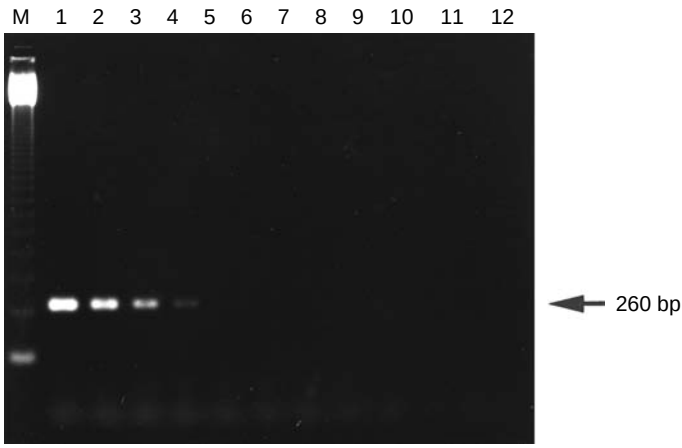

Fig. 7 Effect of a reduced concentration of the streptavidin-protein A chimera bound to end-biotinylated marker DNA on the sensitivity of immuno-PCR. The concentration of the streptavidin-protein A chimera bound to end-biotinylated marker DNA was reduced by a factor of 10, whereas all the other conditions remained the same as those in Fig. 3. Lanes 1 to 9 contain PCR amplification mixtures with immobilized antigen: Lane 1, 96 fmol (5.8×10^{10} molecules); 2, 9.6 fmol (5.8×10^{9} molecules); 3, 960 amol (5.8×10^{8} molecules); 4, 96 amol (5.8×10^{7} molecules); 5, 9.6 amol (5.8×10^{6} molecules); 6, 0.96 amol (5.8×10^{5} molecules); 7, 9.6×10^{-20} mol (5.8×10^{4} molecules); 8, 9.6×10^{-21} mol (5.8×10^{3} molecules); 9, 9.6×10^{-22} mol (5.8×10^{2} molecules). Lanes 10 to 12 are derived from control wells, where no antigen was immobilized. [Reprinted, with permission, from Sano, T., Smith, C.L., Cantor, C.R. (1992) Immuno-PCR: Very sensitive antigen detection by means of specific antibody-DNA conjugates, *Science* **258**, 120–122.]

engineering, can be used for the detection of PCR products generated by immuno-PCR. For example, conventional gel electrophoresis, in which PCR products are stained with a fluorescent dye such as ethidium bromide, can be used to analyze PCR products. This method is sufficient to perform immuno-PCR in a laboratory-scale, where relatively small numbers of samples are analyzed at a time.

A potential advantage of the use of gel electrophoresis for the detection of PCR products generated by immuno-PCR is that more than one species of PCR products can be separated and discriminated. Thus, multiple PCR products generated from single species of marker DNA with the use of more than one set of primers can be analyzed. This would provide higher accuracy in the analysis of antigen. In addition, many different antigens could be detected simultaneously, if each antigen is labeled with a different-sized marker DNA.

Analysis of a large number of samples, such as in clinical diagnostics, is facilitated by using other detection methods that are designed for this purpose and the ease of automation. For example, direct incorporation of a label, such as fluorescent dyes, haptens, and radioisotopes, into PCR products, followed by the detection

of incorporated labels, should offer simple analytical formats. Mass spectrometry of PCR products (or DNA fragments generated by primer extension of PCR products to shift to lower mass) should allow highly parallel, rapid analysis of a number of samples. It should also allow simultaneous analysis of multiple antigens, each of which is coupled to the generation of a different-sized PCR product.

3.7 Quantitation of Antigen (Epitope)

Because PCR allows the exponential amplification of a target segment, the quantitation of PCR products generated by immuno-PCR can provide an estimate of the number of antigens or epitopes in samples. The use of quantitative PCR, which has become a routine analytical method, is very useful for such effort. Mass spectrometry of PCR products (or DNA fragments generated by primer extension of PCR products) should also be quite useful. The accuracy of such estimation can be maximized by a calibration with known numbers of antigens (or epitopes). The range, where the number of antigens can be estimated by quantitation of the PCR products followed by estimation of the number of antigen, depends on the conditions used. To maximize the accuracy of estimation, several different conditions should be used. One such example can be found by comparing Figs. 3 and 4 with Figs. 7 and 8, where different concentrations of the streptavidin-protein A chimera bound to end-biotinylated marker DNA were used.

3.8 Contamination Control

One of the most important factors in avoiding background signals and false-positive signals is control of contamination, which is a problem common to all sensitive detection systems. Even though all procedures are performed very carefully, repeated use of the same marker DNA and primers may generate false-positive signals. One of the major advantages of immuno-PCR is that marker DNA is purely arbitrary. Therefore, marker DNA molecules and their primers can be changed frequently, as needed, to avoid the generation of false-positive signals caused by contamination. This characteristic offers easier control of false-positive signals than other PCR-based detection methods, in which sample-derived nucleic acids are directly amplified.

3.9 General Precautions

Immuno-PCR possesses extremely high sensitivity, derived primarily from the enormous amplification capability of PCR. Thus, any nonspecific binding of antibody and marker DNA will cause serious background problems. Extensive washing after the application of antibody and marker DNA is indispensable. Even though some fraction of specifically bound antibody and marker DNA is removed by washing, the overall sensitivity can be recovered easily by using, for example, additional amplification cycles in the PCR step.

In addition, the use of effective blocking agents is of great importance in avoiding nonspecific binding. Both protein blockers, such as nonfat dried milk and BSA, and nucleic acid blockers, such as sheared sperm DNA, are used. Because of the high specificity of PCR to target DNA defined by primers, the presence of other nucleic acid molecules does not cause the generation of background signals or false-positive signals.

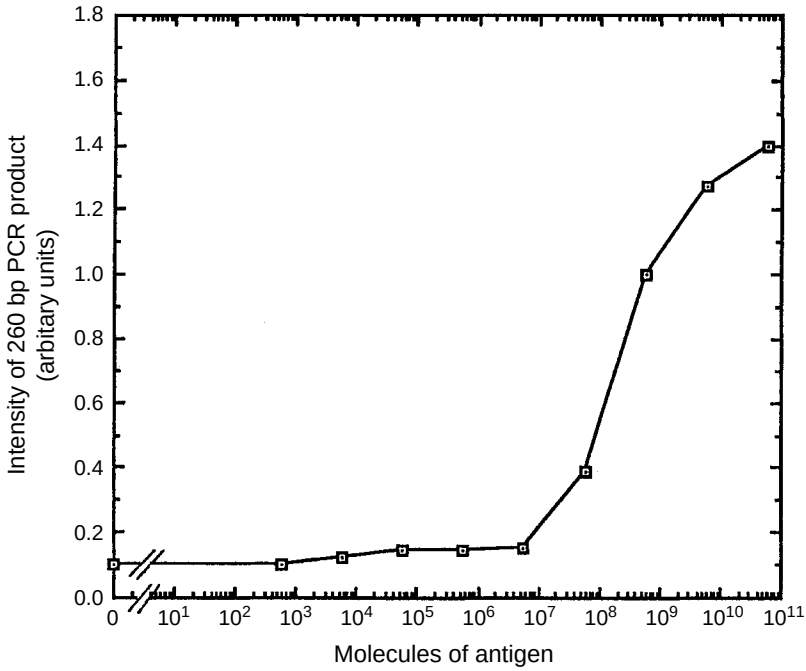

Fig. 8 Quantitation of PCR product generated by immuno-PCR shown in Fig. 7. The intensity of the 260-bp PCR fragment shown in Fig. 7 was quantitated, and plotted as a function of the number of the antigen. Note that PCR amplification was not saturated in a range from 10^7 to 10^{11} molecules of antigen, which is 10^4 to 10^5 times higher than that shown in Fig. 4. [Reprinted, with permission, from Sano, T., Smith, C.L., Cantor, C.R. (1992) Immuno-PCR: Very sensitive antigen detection by means of specific antibody-DNA conjugates, *Science* **258**, 120–122.]

4 Applications and Perspectives

Potentially, there are many areas where immuno-PCR can be applied practically and proficiently. In biological and biomedical sciences, the specific detection of biological molecules of interest is one of the most important steps in analysis. The extremely high sensitivity of immuno-PCR should allow the detection of antigens that cannot be detected by conventional antigen detection systems. Thus, the use of immuno-PCR will allow the analysis of specific antigens at microscopic scales, for example, at single-cell levels.

One of the most practical applications of immuno-PCR is in clinical diagnostics. The extremely high sensitivity of immuno-PCR will enable the specific detection of rare antigens that are present only in very small numbers. This characteristic should allow the diagnosis of diseases and infections at earlier stages of disease or infection development. Another important characteristic of immuno-PCR is its simplicity, which should allow the development of fully automated assay systems. Such automated systems are extraordinarily useful in clinical diagnostics, where large numbers of samples are analyzed repeatedly.

The distinct power and generality of immuno-PCR will allow a tremendous expansion in the applications of antibody-based detection methods, not only to biological but also to nonbiological materials. For example, the extremely high sensitivity of immuno-PCR should enable this technology to be applied to the detection of single antigen molecules, for which no method is currently available.

See also Antibody Molecules, Genetic Engineering of; Immunoassays; Immunology; Infectious Disease Testing by LCR.

Bibliography

Books and Reviews

Sano, T., Smith, C.L., Cantor, C.R. (1995a) Very Sensitive Antigen Detection By Immuno-PCR, in: Nelson, J.O., Karu, A.E., Wong, R.B. (Eds.) *Immunoanalysis of Agrochemicals: Emerging Technologies*, American Chemical Society Symposium Series Number 586, American Chemical Society, Washington, DC, pp. 175–185.

Sano, T., Smith, C.L., Cantor, C.R. (1995b) Detecting minute amounts of antigen, *CHEMTECH* **25**, 24–30.

Sano, T., Smith, C.L., Cantor, C.R. (1996a) Immuno-polymerase Chain Reaction (Immuno-PCR), in: Meyers, R.A. (Ed.) *Encyclopedia of Molecular Biology and Molecular Medicine Volume 3*, VCH Verlagsgesellschaft mbH, Weinheim, Germany, pp. 288–295.

Sano, T., Smith, C.L., Cantor, C.R. (1996b) Immuno-polymerase Chain Reaction (Immuno-PCR), in: Meyers, R.A. (Ed.) *Molecular Biology and Biotechnology: A Comprehensive Desk Reference*, VCH Publishers, New York, pp. 458–460.

Sano, T., Smith, C.L., Cantor, C.R. (1997) Deoxyribonucleic acids as unique markers in molecular detection, *Genet. Anal. – Biomol. Eng.* **14**, 37–40.

Schiavo, S., El-Shafey, A., Jordan, N., Orazine, C., Wang, Y., Zhou, X., Chiu, N.H.L., Krull, I.S. (2004) Pushing the limits of detection with immuno-PCR, *PharmaGenomics* 36–45.

Primary Literature

Adler, M., Wacker, R., Niemeyer, C.M. (2003) A real-time immuno-PCR assay for routine ultrasensitive quantification of proteins, *Biochem. Biophys. Res. Commun.* **308**, 240–250.

Adler, M., Langer, M., Witthohn, K., Eck, J., Blohm, D., Niemeyer, C.M. (2003) Detection of rViscumin in plasma samples by immuno-PCR, *Biochem. Biophys. Res. Commun.* **300**, 757–763.

Cao, Y. (2002) In situ immuno-PCR. A newly developed method for highly sensitive antigen detection in situ, *Methods Mol. Biol.* **193**, 191–196.

Cao, Y., Kopplow, K., Liu, G.Y. (2000) In situ immuno-PCR to detect antigens, *Lancet* **356**, 1002–1003.

Case, M.C., Burt, A.D., Hughes, J., Palmer, J.M., Collier, J.D., Bassendine, M.F., Yeaman, S.J., Hughes, M.A., Major, G.N. (1999) Enhanced ultrasensitive detection of structurally diverse antigens using a single immuno-PCR assay protocol, *J. Immunol. Methods* **223**, 93–106.

Chang, T.C., Huang, S.H. (1997) A modified immuno-polymerase chain reaction for the detection of β-glucuronidase from *Escherichia coli*, *J. Immunol. Methods* **208**, 35–42.

Chao, H.Y., Wang, Y.C., Tang, S.S., Liu, H.W. (2004) A highly sensitive immuno-polymerase chain reaction assay for *Clostridium botulinum* neurotoxin type A, *Toxicon* **43**, 27–34.

Chye, S.M., Lin, S.R., Chen, Y.L., Chung, L.Y., Yen, C.M. (2004) Immuno-PCR for detection of antigen to *Angiostrongylus cantonensis* circulating fifth-stage worms, *Clin. Chem.* **50**, 51–57.

Furuya, D., Yagihashi, A., Yajima, T., Kobayashi, D., Orita, K., Kirimoto, M., Watanabe, N. (2000) An immuno-polymerase chain reaction assay for human interleukin-18, *J. Immunol. Methods* **238**, 173–180.

Gullberg, M., Gustafsdottir, S.M., Schallmeiner, E., Jarvius, J., Bjarnegard, M., Betsholtz, C, Landegren, U., Fredriksson, S. (2004) Cytokine detection by antibody-based

proximity ligation, *Proc. Natl. Acad. Sci. USA* **101**, 8420–8424.

Hendrickson, E.R., Truby, T.M., Joerger, R.D., Majarian, W.R., Ebersole, R.C. (1995) High sensitivity multianalyte immunoassay using covalent DNA-labeled antibodies and polymerase chain reaction, *Nucleic Acids Res.* **23**, 522–529.

Joerger, R.D., Truby, T.M., Hendrickson, E.R., Young, R.M., Ebersole, R.C. (1995) Analyte detection with DNA-labeled antibodies and polymerase chain reaction, *Clin. Chem.* **41**, 1371–1377.

Kakizaki, E., Yoshida, T., Kawakami, H., Oseto, M., Sakai, T., Sakai, M. (1996) Detection of bacterial antigens using immuno-PCR, *Lett. Appl. Microbiol.* **23**, 101–103.

Komatsu, M., Kobayashi, D., Saito, K., Furuya, D., Yagihashi, A., Araake, H., Tsuji, N., Sakamaki, S., Niitsu, Y., Watanabe, N. (2001) Tumor necrosis factor-α in serum of patients with inflammatory bowel disease as measured by a highly sensitive immuno-PCR, *Clin. Chem.* **47**, 1297–1301.

Landegren, U., Schallmeiner, E., Nilsson, M., Fredriksson, S., Baner, J., Gullberg, M., Jarvius, J., Gustafsdottir, S., Dahl, F., Soderberg, O., Ericsson, O., Stenberg, J. (2004) Molecular tools for a molecular medicine: analyzing genes, transcripts and proteins using padlock and proximity probes, *J. Mol. Recognit.* **17**, 194–197.

Liang, H., Cordova, S.E., Kieft, T.L., Rogelj, S. (2003) A highly sensitive immuno-PCR assay for detecting Group A Streptococcus, *J. Immunol. Methods* **279**, 101–110.

Maia, M., Takahashi, H., Adler, K., Garlick, R.K., Wands, J.R. (1995) Development of a two-site immuno-PCR assay for hepatitis B surface antigen, *J. Virol. Methods* **52**, 273–286.

MaKie, A., Samuel, D., Cohen, B., Sanders, N.A. (2002a) Development of a quantitative immuno-PCR assay and its use to detect mumps-specific IgG in serum, *J. Immunol. Methods.* **261**, 167–175.

MaKie, A., Samuel, D., Cohen, B., Sanders, N.A. (2002b) A quantitative immuno-PCR assay for the detection of mumps-specific IgG, *J. Immunol. Methods.* **270**, 135–141.

Mweene, A.S., Ito, T., Okazaki, K., Ono, E., Shimizu, Y., Kida, H. (1996) Development of immuno-PCR for diagnosis of bovine herpesvirus 1 infection, *J. Clin. Microbiol.* **34**, 748–750.

Niemeyer, C.M., Adler, M., Blohm, D. (1997) Fluorometric polymerase chain reaction (PCR) enzyme-linked immunosorbent assay for quantification of immuno-PCR products in microplates, *Anal. Biochem.* **246**, 140–145.

Niemeyer, C.M., Wacker, R., Adler, M. (2003) Combination of DNA-directed immobilization and immuno-PCR: very sensitive antigen detection by means of self-assembled DNA-protein conjugates, *Nucleic Acids Res.* **31**, e90.

Niemeyer, C.M., Adler, M., Gao, S., Chi, L. (2001) Nanostructured DNA-protein aggregates consisting of covalent oligonucleotide-streptavidin conjugates, *Bioconjugate Chem.* **12**, 364–371.

Niemeyer, C.M., Adler, M., Pignataro, B., Lenhert, S., Gao, S., Chi, L., Fuchs, H., Blohm, D. (1999) Self-assembly of DNA-streptavidin nanostructures and their use as reagents in immuno-PCR, *Nucleic Acids Res.* **27**, 4553–4561.

Ozaki, H., Sugita, S., Kida, H. (2001) A rapid and highly sensitive method for diagnosis of equine influenza by antigen detection using immuno-PCR, *Jpn. J. Vet. Res.* **48**, 187–195.

Ren, J., Chen, Z., Juan, S.J., Yong, X.Y., Pan, B.R., Fan, D.M. (2000) Detection of circulating gastric carcinoma-associated antigen MG7-Ag in human sera using an established single determinant immunopolymerase chain reaction technique, *Cancer* **88**, 280–285.

Ruzicka, V., Marz, W., Russ, A., Gross, W. (1993) Immuno-PCR with a commercially available avidin system, *Science* **260**, 698–699.

Saito, K., Kobayashi, D., Sasaki, M., Araake, H., Kida, T., Yagihashi, A., Yajima, T., Kameshima, H., Watanabe, N. (1999) Detection of human serum tumor necrosis factor-α in healthy donors, using a highly sensitive immuno-PCR assay, *Clin. Chem.* **45**, 665–669.

Sanna, P.P., Weiss, F., Samson, M.E., Bloom, F.E., Pich, E.M. (1995) Rapid induction of tumor necrosis factor alpha in the cerebrospinal fluid after intracerebroventricular injection of lipopolysaccharide revealed by a sensitive capture immuno-PCR assay, *Proc. Natl. Acad. Sci. USA* **92**, 272–275.

Sano, T., Smith, C.L., Cantor, C.R. (1992) Immuno-PCR: Very sensitive antigen detection by means of specific antibody-DNA conjugates, *Science* **258**, 120–122.

Sano, T., Smith, C.L., Cantor, C.R. (1993) Immuno-PCR with a commercially available avidin system (response), *Science* **260**, 699.

Sugawara, K., Kobayashi, D., Saito, K., Furuya, D., Araake, H., Yagihashi, A., Yajima, T., Hosoda, K., Kamimura, T., Watanabe, N. (2000) A highly sensitive immuno-polymerase chain reaction assay for human angiotensinogen using the identical first and second polyclonal antibodies, *Clin. Chim. Acta* **299**, 45–54.

Suzuki, A., Itoh, F., Hinoda, Y., Imai, K. (1995) Double determinant immuno-polymerase chain reaction: a sensitive method for detecting circulating antigens in human sera, *Jpn. J. Cancer Res.* **86**, 885–889.

Wu, H.C., Huang, Y.L., Lai, S.C., Huang, Y.Y., Shaio, M.F. (2001) Detection of *Clostridium botulinum* neurotoxin type A using immuno-PCR, *Lett. Appl. Microbiol.* **32**, 321–325.

Zhang, H.T., Kacharmina, J.E., Miyashiro, K., Greene, M.I., Eberwine, J. (2001) Protein quantification from complex protein mixtures using a proteomics methodology with single-cell resolution, *Proc. Natl. Acad. Sci. USA* **98**, 5497–5502.

Zhang, Z., Irie, R.F., Chi, D.D., Hoon, D.S. (1998) Cellular immuno-PCR. Detection of a carbohydrate tumor marker, *Am. J. Pathol.* **152**, 1427–1432.

Zhou, H., Fisher, R.J., Papas, T.S. (1993) Universal immuno-PCR for ultrasensitive target protein detection, *Nucleic Acids Res.* **21**, 6038–6039.

Immuno-PCR (Immuno-Polymerase Chain Reactions): *see* Immuno-PCR

Immunosuppressive Factors in Cancer

Jeremy E. Moss[1], *Christine N. Metz*[2], *and Richard Bucala*[1]
[1] *Yale University School of Medicine, New Haven, CT, USA*
[2] *North Shore-Long Island Jewish Research Institute, Manhasset, NY, USA*

1 Introduction 468

2 Transforming Growth Factor Beta (TGFβ) 469
2.1 Sources of TGFβ 469
2.1.1 Effects of TGFβ on Monocytes/macrophages 470
2.1.2 Effects of TGFβ on T lymphocytes 470
2.1.3 Effects of TGFβ on NK and Lymphokine Activated Killer (LAK) Activity 472
2.1.4 Effects of TGFβ on Dendritic Cells (DCs) 472
2.2 Inhibition of TGFβ: Implications for Therapy 473

3 Interleukin 10 (IL-10) 474
3.1 Sources of IL-10 474
3.1.1 Effects of IL-10 on Monocytes/macrophages 475
3.1.2 Effects of IL-10 on T Lymphocytes 475
3.1.3 Effects of IL-10 on NK Cells 476
3.1.4 Effects of IL-10 on DCs 476
3.2 Inhibition of IL-10: Implications for Therapy 477

4 Macrophage Migration Inhibitory Factor (MIF) 479

5 Prostaglandin E2 (PGE2) 480
5.1 Sources of PGE2 480
5.1.1 Effects of PGE2 on Monocytes/Macrophages 480
5.1.2 Effects of PGE2 on T Lymphocytes 480
5.1.3 Effects of PGE2 on NK and LAK Cell Activity 481
5.1.4 Effects of PGE2 on DCs 481
5.2 Inhibition of PGE2: Implications for Therapy 481

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

6 Indoleamine 2,3-dioxygenase (IDO) 481
6.1 Sources of IDO 482
6.2 Effects of IDO on Immune Cells 482
6.3 Implications for Therapy 482

7 Cell Surface Molecules 483
7.1 Fas/FasL 483
7.1.1 Sources of FasL 483
7.1.2 Tumor Evasion of FasL-mediated Toxicity 483
7.1.3 Tumor-derived FasL is Functional 483
7.1.4 FasL Expression is Correlated with Improved Prognosis in Some Tumor Models 484
7.2 MHC Class I: Classical and Nonclassical Receptors 484
7.2.1 Classical MHC Class I 484
7.2.2 HLA-G 485
7.3 B7-H1 486
7.4 RCAS1 486
7.5 Other Cell Surface Molecules 487

8 Polyamines 487
8.1 Sources of Polyamines 487
8.1.1 Effects of Polyamines on Monocytes/macrophages 488
8.1.2 Effects of Polyamines on T Lymphocytes 488
8.1.3 Effects of Polyamines on NK Cells 488
8.2 Inhibition of Polyamine Biosynthesis: Implications for Therapy 488

9 Tumor Shed Immunosuppressive Molecules 488

10 Conclusion 489

Bibliography 490
Books and Reviews 490
Primary Literature 491

Keywords

Anergy
Inability to generate an immune reaction to a substance that is usually antigenic.

Indoleamine 2,3-dioxygenase
Enzyme that degrades the amino acid tryptophan. The expression of this enzyme by cells has been associated with immunosuppressive properties.

Polyamines
Naturally occurring organic amines that are critical for cellular proliferation and differentiation. Examples include putrescine, spermine, and spermadine.

Tolerance
Lack of immune response to an antigen as a result of either central (thymic) or peripheral clonal T-cell deletion or anergy.

Tumor Infiltrating Lymphocyte
Lymphocyte that infiltrates and destroys cancer cells. Maximizing the ability of these cells to kill tumor cells is the focus of much immunotherapeutic research.

Abbreviations

AICD	activation induced cell death
APC	antigen presenting cell
CEA	carcinoembryonic antigen
COX-1, COX-2	cyclooxygenase-1 and 2
CTL	cytotoxic T lymphocytes
DC	dendritic cell
DFMO	(D, L-2-(difluoromethyl) ornithine)
FAP	familial adenomatous polyposis
FasL	fas-ligand
IDO	indoleamine 2,3-dioxygenase
IFNγ	interferon gamma
LAK	lymphokine activated killer
LAP	latency associated peptide
LPS	lipopolysaccharide
MHC	major histocompatibility
MIF	macrophage migration inhibitory factor
MLR	mixed lymphocyte reaction
NK	natural killer
NO	nitric oxide
ODC	ornithine decarboxylase
PGE2	prostaglandin E2
TAP	transporter of antigen
TGFβ	transforming growth factor beta
TIL	tumor infiltrating lymphocytes
TNFα	tumor necrosis factor alpha
VEGF	vascular endothelial growth factor

Tumor cells avoid attack from the host immune system by expressing a variety of different molecules including: cytokines, prostaglandins, polyamines, enzymes, cell surface receptors, and shed antigens. Cancers that are successful in evading host antitumor defenses tend to persist, enlarge, and ultimately metastasize. Thus, an accurate identification and understanding of the various immunosuppressive factors that tumor cells employ to impair the host immune response is critical for the development of effective treatments for cancer patients. Overcoming these evasive mechanisms may be the last major obstacle to the generation of universally successful cancer therapeutics.

1 Introduction

The ability of the immune system to detect and destroy cancerous growths has been intensively debated for many years. Only recently has the consensus opinion shifted toward favoring a key role of the immune system in antitumor immunosurveillance. Several findings point to this conclusion. One is the low frequency of tumors in immunocompetent hosts and the elevated rate of malignancies in patients with immunodeficiency (even when virally induced cancers are excluded). Another argument suggesting an important role of the immune system in tumor clearance is the fact that the degree of infiltration of immune cells into tumors correlates with improved survival for patients with many different solid tumors. Mouse models have been crucial for elucidating the role of the immune system in antitumor immunity. Investigators initially concluded that murine data argued against immune system–mediated clearance of naturally developing tumors since immunocompromised mice did not develop spontaneous tumors at a higher rate than their immunocompetent counterparts. However, more recently, a strain of severely immunocompromised mice with more global immune system defects was developed. This strain displays a significantly increased number of tumors with age. Finally, as will be described in more detail below, a variety of murine models of cancer now exist which clearly demonstrate immune-mediated tumor clearance of cancers, even when tumor cells are syngeneic to the host.

It is now generally accepted, moreover, that the presence of tumor cells in an organism may not only stimulate cells of the immune system but their presence can also suppress the immune system under certain circumstances. When immunogenicity and stimulation of the host immune system predominate, destruction and clearance of cancerous cells generally occurs. Conversely, when tumor cells succeed in evading the host immune system through immunosuppressive mechanisms, they tend to survive, enlarge, and ultimately metastasize. Therefore, successful tumors acquire the ability to produce factors that subvert the host antitumor immune response. These immunosuppressive factors can repress host antitumor immune cells directly, by preventing tumor recognition and inhibiting immune cell–mediated tumor killing, but they also sometimes act indirectly by promoting host suppressor cell activity. In this article, we will discuss the effects of tumor-derived

mediators, cell surface receptors, and shed antigens on host immune cells and on tumor cells. An accurate identification and understanding of the immunosuppressive factors that hinder the host antitumor immune response will be critical for the effective treatment of cancer patients, whether by conventional means or by newly developed immunotherapeutic strategies. Overcoming the evasive mechanisms used by tumor cells to escape immune detection may be the last major obstacle to the generation of universally successful cancer therapeutics.

2 Transforming Growth Factor Beta (TGFβ)

Transforming growth factor beta (TGFβ) was originally described for its ability to induce rat kidney fibroblasts to form large colonies in soft agar in the presence of epidermal growth factor. Since that time, TGFβ has been shown to play critical roles in cell growth, differentiation, and transformation. Despite its name, however, TGFβ is generally growth inhibitory for cells. In fact, mutation of molecules in the TGFβ signaling pathway are associated with important human cancers including the inherited disease, hereditary nonpolyposis colon cancer syndrome, as well as spontaneously occurring gastric tumors, gliomas, pancreatic cancers, breast cancers, ovarian cancers, and colon cancers. Moreover, *Tgfb1* mutant mice display an increased incidence of chemically induced and spontaneous tumors while overexpression of TGFβ1 in keratinocytes makes mice resistant to chemical carcinogens.

In addition to its ability to augment cell growth, TGFβ has a variety of other effects on different cells types. TGFβ can suppress the host immune response to tumor growth and metastases. For example, in an interesting experiment, it was found that transgenic mice engineered with dominant negative, and thus nonfunctional, TGFβ receptors in either CD4$^+$ or CD8$^+$ T cells cleared melanoma and thymoma tumor cells more efficiently than wild-type mice. Moreover, TGFβ has proangiogenic and matrix inducing effects that are believed to promote tumor invasion and metastasis. Thus, in light of TGFβ's tumor-stimulating activities and despite its growth inhibitory activities, anti-TGFβ therapy is actively being pursued for the treatment of cancer.

The TGFβ superfamily consists of TGFβ1, β2, and β3, as well as an additional 40 different TGFβ-related proteins. The three mammalian isoforms (TGFβ1, 2, and 3) are synthesized as pre-pro TGFβ polypeptides that contain a signal sequence, the pro-region (latency associated peptide; LAP), and mature TGFβ (N-terminal region). The majority of TGFβ is secreted as an inactive complex associated with a second gene product known as the *latent TGFβ binding protein*. Thus, activation of the TGFβ complex is a key regulatory step for controlling the cytokine's activity. The biologic effects of TGFβ1, 2, and 3 are mediated through the same high affinity serine-threonine kinase type I and II receptors. Upon ligation of the receptors, a series of downstream signaling events occur, which include changes in proteins termed *Smads*.

2.1 Sources of TGFβ

Many cultured cells can be stimulated to secrete TGFβ, including activated (but not resting) macrophages, T cells, B cells, and natural killer (NK) cells. In addition,

numerous tumor cell types including: pancreatic, prostate, lung, colon, melanoma, and breast carcinoma cells can secrete the molecule. Elevated levels of TGFβ in serum correlate with cancer progression and reduced survival for many cancers.

While TGFβ can inhibit MHC class II expression on tumor cells, most tumor cells resist other effects of the cytokine including tumor suppressor and growth inhibitory effects. Tumor cells are rendered insensitive to the suppressive effects of TGFβ by: (1) receptor downregulation, (2) inactivating mutations within the TGFβ receptor, and/or (3) anomalies in the postreceptor TGFβ signaling pathway that includes the Smad proteins. By contrast, the cytokine has profound direct effects on immune cells expressing TGFβ receptors. Furthermore, TGFβ indirectly compromises the function of immune cells via its modification of the production of different cytokines and chemical messengers from a variety of host cells.

2.1.1 Effects of TGFβ on Monocytes/macrophages

Macrophages are important cellular coordinators of the host immune response to tumors. Macrophages serve as antigen presenting cells and also produce cytotoxic molecules, such as tumor necrosis factor alpha (TNFα) and nitric oxide (NO) that directly kill tumor cells. Macrophages promote both T- and NK-cell antitumor activities. The suppressive effects of TGFβ on monocytes/macrophages are complex and depend on the cells' origin, differentiation state, and cytokine milieu. TGFβ blocks colony stimulating factor-1-dependent proliferation of bone marrow precursor cells and hence, *affects macrophage production*. It also inhibits lipoprotein a-induced *macrophage proliferation*. Furthermore, tumor-derived TGFβ is well known for its ability to "deactivate" tissue macrophages that express high levels of TGFβR1 and TGFβR2. For example, TGFβ *suppresses the antitumor cytolytic activity* of bone marrow–derived macrophages preactivated with interferon gamma (IFNγ) and lipopolysaccharide (LPS) and *inhibits cytokine (TNFα) and NO production by macrophages*. Finally, *macrophages rendered dysfunctional by TGFβ have powerful inhibitory effects on the development of the host T-cell-mediated antitumor immune response*.

2.1.2 Effects of TGFβ on T lymphocytes

Multiple T-cell types, including CD4$^+$, CD8$^+$ cytotoxic T lymphocytes (CTLs), and tumor infiltrating lymphocytes (TILs) play prominent roles in the host immune response to tumors. TGFβ directly *suppresses T-lymphocyte activity*. It reduces pore-forming protein expression in human peripheral CD8$^+$ T cells and inhibits the cytolytic activity of both anti-CD3 stimulated CD8$^+$ T cells and CTLs directed against fresh tumor targets. Furthermore, TGFβ *inhibits the generation of CTLs in vitro*. Similar to its effect on macrophages, *TGFβ suppresses cytokine release by T lymphocytes*, including IFNγ, TNFα, IL-2, IL-6, IL-4 and IL-5.

The effects of TGFβ on T-cell immune responses are dependent on numerous factors, including T-cell type and activation status. Recently, in an examination of the effect of TGFβ on T cells (stimulated versus nonstimulated) exhibiting a defined phenotype (CD4$^+$ versus CD8$^+$ and Th1 versus Th2) and a defined maturational stage (naïve versus memory T cells) one group found that: (1) *TGFβ inhibits both the proliferation and differentiation of Th1 and Th2 cells*, even when TGFβ is present during primary stimulation; (2) *CD4$^+$ T cells primed in the presence of TGFβ exhibit*

reduced immune responses induced by receptor cross-linking and/or by specific antigen stimulation, which is independent of costimulatory molecule and IL-2 expression; (3) the presence of TGFβ during primary T-cell stimulation has a long-lasting or "imprinting" effect; (4) the immunosuppressive effects of TGFβ on T cells cannot be reversed with IL-2; and (5) *TGFβ directly suppresses memory Th1 cytokine production* by downregulating the IL-12 receptor β2 chain on T cells, while TGFβ has no effect on memory Th2 cytokine production. Therefore, tumor-derived TGFβ assist in the escape of tumors from the host cell–mediated immune response by *suppressing the host Th1 response and promoting an ineffective Th2 response.*

Similar to its many other activities, the effect of TGFβ on *T-cell proliferation* is dependent on the cell population and activation status. The addition of TGFβ to mixed lymphocyte tumor cultures results in a decrease in the number of effector CD8$^+$ CTLs generated because the cytokine inhibits their proliferation. Moreover, TGFβ inhibits the proliferation of intraepithelial CD8$^+$ lymphocytes following activation. By contrast, when naïve T cells are activated in the presence of TGFβ, they proliferate and secrete IL-2. The antiproliferative effect of TGFβ has been attributed to downregulation of the IL-2 receptor and the inhibition of IL-2 and IL-12-mediated proliferation signals. IL-2 and IL-12 are immunostimulating cytokines capable of inducing cytokine production, enhancing cytolytic activity, and promoting T-cell proliferation. There are two opposing reports describing the mechanism by which TGFβ inhibits IL-12-mediated proliferation. One shows that TGFβ blocks IL-12 induced phosphorylation and activation of the Jak-2, Tyk-2 kinases and STAT3 and 4 transcription factors in activated lymphocytes. The other states that TGFβ does not block cytokine (IL-12 and IL-2) induced JAK and STAT phosphorylation in T cells. Thus, the mechanism by which TGFβ inhibits IL-12 induced T-cell proliferation is not clear and may depend on different factors in different circumstances.

Conflicting reports of the effects of TGFβ on *T-cell apoptosis* have been described. While TGFβ has been shown in some studies to *induce T-cell apoptosis*, in other studies it has been shown to *prevent T-cell apoptosis*. These contrasting results suggest that multiple factors (microenvironment, origin of T cell, as well as the cell's differentiation and activation states) influence the T-cell apoptotic response to TGFβ. TGFβ has been reported to promote apoptosis in T cells independent of the IL-2/IL-2 receptor pathway and through the activation of caspase-like proteases in activated murine T cells. By contrast, when present during the course of T-cell activation, TGFβ *inhibits activation induced cell death* (*AICD*) *and Fas-mediated apoptosis*. AICD by T cells is one mechanism for controlling the size and the duration of the primary immune response. Further studies showed that *TGFβ prevents apoptosis in memory Th1 CD4$^+$ T cells*. The addition of TGFβ to human T cells and T-cell hybridomas significantly reduced apoptosis after activation by anti-CD3. TGFβ blockade of T-cell apoptosis was associated with decreased FasL expression. These findings suggest that TGFβ, when present at the right time, functions as a "viability factor" allowing the clonal expansion of effector T cells and the generation of long-lived memory T cells during the immune response. The importance of the timing of TGFβ administration for inducing T-cell apoptosis has been demonstrated *in*

vitro. When added at the initiation of T-cell mitogenesis, this cytokine does not induce T-cell apoptosis, whereas TGFβ added postactivation enhances apoptosis in T cells.

2.1.3 Effects of TGFβ on NK and Lymphokine Activated Killer (LAK) Activity

NK cells are a subset of bone marrow–derived immune cells that proliferate in response to viral and cellular invasion. Once activated, NK cells directly kill tumor cells, but do not require matching of major histocompatibility (MHC) molecules between effector and target cells. Overall, *TGFβ has an inhibitory effect on peripheral blood NK cells*. TGFβ *inhibits NK proliferation, cytokine production (IFNγ, TNFα, GM-CSF), activation, and cytotoxicity*.

NK cells represent the main precursors of lymphokine activated killer (LAK) activity. *TGFβ inhibits the induction of LAK cell activity in vitro*. Moreover, it *suppresses the production and the antitumor cytotoxic activity of LAKs*.

2.1.4 Effects of TGFβ on Dendritic Cells (DCs)

Dendritic cells (DCs) are the most powerful antigen presenting cells (APCs) of the immune system and their function is vital to the induction of primary immune responses, immunologic tolerance, and regulation of Th1/Th2 immune responses. DCs coordinate both innate and adaptive immunity-based mechanisms of tumor killing. Immature DCs express abundant intracellular levels of MHC class I and II molecules, which can present specific tumor antigens. Once matured, DCs display the MHC bound antigen on their surface together with multiple costimulatory molecules (CD40, CD50, CD54, CD58, CD80, CD83, and CD86) and potently activate resting or naïve $CD4^+$ and $CD8^+$ T cells to promote their differentiation into functional helper and effector (tumor-killing) T cells. Thus, the most important function of DCs is to sensitize the immune system to specific antigens. Because DCs are ten to one hundred times more potent APCs than B cells and monocytes/macrophages, much work has focused on using these cells to stimulate immunity against cancer.

TGFβ is immunosuppressive to DCs. The cytokine *blocks GM-CSF-induced maturation of DCs* (defined as CD86 and MHC class II expression and mixed lymphocyte reaction (MLR) stimulating activity) from mouse bone marrow progenitor cells. In addition, *TGFβ overexpressing murine bone marrow–derived DCs exhibit lower allogeneic T-cell stimulatory activity* (assessed by MLR and CTL assays) than nontransduced cells. Thus, TGFβ suppresses important DC activities required during host antitumor responses.

To function as APCs, DCs must localize, presumably from the bone marrow to the tumor site where specific antigens are present, and then antigen-loaded DCs must migrate to lymphoid tissue. Little is known about the directed migration of DCs *in vivo*. Recent studies show that *TGFβ reversibly regulates human DC chemotaxis in vitro*. TGFβ enhances the expression of the chemokine receptors CCR1, CCR3, CCR5, and CXCR4 on immature DCs and promotes their chemotaxis *in vitro*. However, the transcriptional expression of another chemokine receptor, CCR7, by mature DCs is suppressed by TGFβ. Therefore, *TGFβ might promote the migration of immature DCs, but inhibit the localization of mature DCs to the lymphoid organs during the host immune response to tumors*.

2.2 Inhibition of TGFβ: Implications for Therapy

Numerous methods for inhibiting the immunosuppressive effects of TGFβ during tumor growth have been described in experimental models. The biologic activity of mature TGFβ can be controlled by suppressing the production of TGFβ, inhibiting the activity of mature TGFβ, and/or by altering the molecule's interaction with receptors on immune cells. Examples of direct inhibition of TGFβ activity/production during tumor growth include the use of: soluble TGFβ receptors, nonsignaling TGFβ binding proteins such as decorin and alpha2 macroglobulin, neutralizing anti-TGFβ antibodies, and antisense oligonucleotides. Of note, in recent murine studies, a soluble anti-TβRII (TGFβ receptor antagonist) was found to decrease mouse mammary tumor metastasis to the lung. Also, treatment of nude mice with anti-TGFβ inhibited metastases of colonic adenocarcinomas. In humans, TGFβ immunoneutralization significantly enhanced the sensitivity of human breast carcinoma cells to cisplatin-mediated cytotoxicity *in vitro*, suggesting that TGFβ blockade during chemotherapy treatment may promote tumor cell killing. Humanized monoclonal antibodies to TGFβ are currently being studied not only for the treatment of cancers but also for systemic sclerosis, rhinitis, glaucoma, and prevention of coronary artery restenosis following angioplasty.

Although there are a few reports describing the secretion of biologically active TGFβ, the majority of TGFβ is secreted as a latent complex, as described above. This latent TGFβ complex is subsequently activated to release mature TGFβ, which consists of two identical disulfide-linked polypeptide homodimers. *Therefore, activation of the latent TGFβ complex appears to be the major regulator of TGFβ activity in vivo*. Activation of latent TGFβ occurs by both proteolytic mechanisms including: subtilisin-like proprotein convertases such as furin, plasmin, transglutaminase, and tissue and urokinase plasminogen activator, and by nonproteolytic mechanisms such as thrombospondin and mannose-6-phosphate/insulin growth factor II receptor interactions. In addition to being sources of TGFβ itself, tumor cells also express a number of proteases that can cleave and activate latent TGFβ.

Recent investigations highlight the importance of TGFβ activation by subtilisin-like convertases, such as furin. Like TGFβ, furin is ubiquitously expressed. The formation of active TGFβ by glioma cells *in vitro* can be blocked by agents that specifically inhibit the enzymatic activity of furin. *Thus, the development of agents that block the activation of the latent TGFβ complex may represent a new class of anti-immunosuppressive agents for the treatment of cancer*.

Another means of tumor therapy that is being investigated involves attempting to restore TGFβ sensitivity to tumors. Histone deacetylase inhibitors, captopril, and farnesyl transferase inhibitors have all shown promise in pharmacologic attempts to upregulate TGFβ signaling in tumor cells. Thus, these and other similar functioning drugs are being considered in the search for anticancer therapies.

Although much is known about the specific suppressive effects of TGFβ on immune cells *in vitro*, the precise role of TGFβ in host immunosuppression during tumor growth *in vivo* is not well understood. Obviously, numerous factors play a role in the host response to TGFβ during tumor growth and it has been

found that the predominant effect of this cytokine *in vivo* is not always immunosuppression. For example: (1) high expression of TGFβ is not always associated with a poor prognostic outcome in human cancer; (2) overexpression of TGFβ in experimental models does not always lead to tumor formation and immunosuppression; (3) mutations of TGFβ and the pathways through which it signals are associated with development of cancer in both humans and mice and; (4) the host response to tumors overexpressing TGFβ can vary depending on the site of tumor implantation. Despite these findings, it is generally agreed that further studies investigating the use of TGFβ inhibitors to reduce cancer-induced immunosuppression and inhibit metastases are justified on the basis of current understanding of the numerous functions of this interesting molecule.

3 Interleukin 10 (IL-10)

Human interleukin 10 (IL-10) functions as a homodimeric 17 to 20 kD glycoprotein and exhibits a high degree of homology with its nonglycosylated mouse IL-10 counterpart. Unlike TGFβ, IL-10 is synthesized and released as an active cytokine. It signals through the IL-10 receptor, a member of the IFN receptor family. The IL-10 receptor, which is expressed on numerous immune cells, is composed of two subunits: IL-10 receptor 1 (ligand binding subunit) and IL-10 receptor 2 (signaling subunit).

3.1 Sources of IL-10

IL-10 was originally identified as a product of Th2 cells in response to antigen stimulation in the presence of APCs. It is now thought that macrophages are the main physiologic source of IL-10, but the cytokine is also expressed by other immune cells including naïve and memory T cells, B cells, eosinophils, mast cells, DCs, monocytes, and tumor (melanoma)–infiltrating T cells and NK cells. Many solid tumors express IL-10 *in vitro* and *in vivo* including lung tumors, melanomas, and gliomas. In addition, leukemic tumors that originate from both the peripheral circulation and the bone marrow (both T- and B-cell tumors) express high levels of IL-10. Finally, Hodgkins as and non-Hodgkin's lymphomas can overexpress IL-10.

Elevated serum IL-10 is considered a negative prognostic indicator for numerous cancers including Hodgkin's lymphomas, metastatic melanoma, and pancreatic cancer. Circulating IL-10 often serves as a marker for (1) the progression from adenoma to carcinoma; (2) the presence of metastatic disease, and (3) the degree of malignancy of a cancer. On the basis of the significant relationships between serum IL-10 levels and the time-to-treatment failure and patients' overall survival, IL-10 is emerging as an important indicator for monitoring disease progression in cancer patients.

The general effect of IL-10 on tumor cells themselves is protective. Pretreatment of tumor target cells with IL-10 defends them from allospecific CTL lysis. Additionally, IL-10 treatment protects tumor cells by downregulating their expression of the cell surface target molecule that NK cells use for antitumor cytotoxicity. Tumor cells use IL-10 as an antiapoptotic factor and this function can protect them from the activity of cytotoxic drugs. Aside from the different protective effects that IL-10 has on tumor cells directly, perhaps its most

important function for promoting tumor growth involves its ability to suppress the host immune response.

3.1.1 Effects of IL-10 on Monocytes/macrophages

IL-10 was first described as "cytokine synthesis inhibitory factor". *IL-10 inhibits LPS-induced monocyte/macrophage synthesis of numerous proinflammatory cytokines including: IL-1, IL-6, IL-8, and IL-12.* In addition, *IL-10 positively regulates macrophage production of prostaglandin (another potent immunosuppressive agent), inhibits macrophage NO production, and attenuates the cytotoxic activity* of monocytes/macrophages.

IL-10 *reduces the antigen presenting capacity of certain classes of macrophages* by multiple mechanisms. *IL-10 suppresses cell surface MHC class II expression* by macrophages, in part by inhibiting the transport of peptide loaded MHC molecules to the plasma membrane. It *decreases the surface expression of CD80 and CD86,* as well as other important costimulatory molecules. Finally, IL-10 inhibits the differentiation of monocytes into dendritic cells.

By manipulating macrophage cytokine production and expression of important cell surface molecules, IL-10 reduces the APC of macrophages to specific T cells and thus, suppresses the host antitumor immune response. Interestingly, macrophages obtained from tumor-bearing hosts are more susceptible to the immunosuppressive activities of IL-10 and less sensitive to immunostimulatory factors necessary to mount an effect antitumor response.

3.1.2 Effects of IL-10 on T Lymphocytes

Beyond the indirect effects that IL-10 has on T cells through its suppressive effects on APCs (mentioned above), IL-10 also directly affects different classes of T cells by interacting with IL-10 receptors. Similar to its effect on monocytes/macrophages, IL-10 directly affects cytokine production by some T cells. For example, IL-10 *inhibits type I cytokine (IL-2, TNFα) production* by stimulated $CD4^+$ T cells. By contrast, however, IL-10 does not inhibit IFNγ secretion by $CD8^+$ T cells in response to tumor stimulation *in vivo*.

A growing literature has been developing on the role of a subset of $CD4^+$ T cells, the Th3/Tr1 or regulatory T cells, which suppress T-cell mediated immune responses and assist in the development of immunologic tolerance. IL-10 appears to be necessary for the generation of this immunomodulatory set of cells. Furthermore, IL-10 (as well as TGFβ) is a secreted cytokine that Th3 cells use to inhibit the activity of both Th1 and Th2 $CD4^+$ T cells.

The effects of IL-10 on host antitumor $CD8^+$ CTL responses are somewhat contradictory. The cytokine exhibits both immunosuppressive and immunostimulatory effects on $CD8^+$ T cells *in vitro* and *in vivo*. Under certain experimental conditions, however, the immunosuppressive effects predominate. *In vitro, IL-10 inhibits proliferative and cytotoxic T-cell responses* generated in primary MLRs. *In vivo,* IL-10 *represses CTL activity* directly by inhibiting granzyme B expression and indirectly by inhibiting the generation of functional CTLs. IL-2 or IFNγ treatment can restore the IL-10-suppressed CTL alloreactivity if administered at the time of tumor challenge, suggesting that the immunosuppressive effect of IL-10 can be blocked by Th1 type cytokines. In a mouse model examining the growth of Lewis lung tumors in mice engineered to overexpress T cell–derived IL-10, *overexpression of IL-10 significantly enhanced tumor growth.* This

enhanced growth was associated with decreased cytokine production by T cells (IL-2 and IFNγ), and suppressed cytolytic activity of CD8$^+$ T cells. Furthermore, mice produced dysfunctional splenic APCs that expressed significantly reduced cell surface levels of MHC class I, CD80, CD86, and CD11c molecules. The dysfunctional APCs could not support CTL proliferation in MLRs or promote antitumor CTL induction. DCs obtained from the IL-10 transgenic mice failed to induce antitumor CD8$^+$ T cell reactivity *in vivo* following exposure to tumor-specific peptides. Thus, this finding supports previous work which showed that IL-10 treatment of DCs reduces their allostimulatory activity for CD8$^+$ T cells and can induce alloantigen-specific anergy in CD8$^+$ T cells.

In contrast to the numerous immunosuppressive functions of IL-10 on T cells described above, under certain conditions this cytokine stimulates the host immune system and contributes to limiting the growth and metastasis of tumors. *IL-10 enhances CD8$^+$ cellular recruitment, promotes CD8$^+$ T-cell proliferation, augments the outgrowth of CTL precursor cells*, and sometimes *enhances CTL cytolytic activity*. *In vivo*, anti-IL-10 reduced the host primary CTL response in mice following immunization with B7-1$^+$ P815 tumor cells, *suggesting an important role of IL-10 in the host antitumor CTL response*. Furthermore, in transgenic mice engineered to overexpress IL-10 in APCs, a biphasic pattern of tumor growth was observed in which tumors grew rapidly during the first two weeks, but then completely regressed. Thus, *the effects of IL-10 on the host CTL response are complex*; T cell-derived IL-10 may function as an immunosuppressive factor, whereas APC-derived IL-10 might act as an immune stimulating factor.

Although IL-10 has direct effects on the function of various subsets of T cells, most often the inhibitory effects of IL-10 on T cells are mediated indirectly through its effect on APC functions. As mentioned previously, *IL-10 treatment of DCs induces alloantigen-specific anergy (unresponsiveness) in both CD4$^+$ and CD8$^+$ T cells* when present during antigen-specific activation. IL-10 induced CD4$^+$ T-cell anergy can be reversed by IL-2 administration or receptor cross-linking (anti-CD3/anti-CD28), as well as by immunization with mature dendritic cells. These finding may prove to be very important therapeutically in the future.

3.1.3 Effects of IL-10 on NK Cells

Overall, *IL-10 exerts an immunostimulatory effect on NK cells*. NK cells exhibit low level IL-10 receptor expression and treatment of NK cells with *IL-10 potentiates NK cytokine production and enhances NK cytotoxicity* toward tumors *in vitro*. Furthermore, although IL-10 has no effect by itself on NK proliferation, it *augments IL-2 induced NK proliferation*. These *in vitro* findings are supported *by in vivo* studies. In some experimental tumor models, IL-10 injection or IL-10 secreting tumor cells inhibit tumor metastases by an NK-dependent mechanism. Thus, it is important to recognize that unlike the majority of the host immune system that would probably be stimulated by IL-10 blockade during cancer therapy, NK-cell production and activity may be reduced with IL-10 inhibition.

3.1.4 Effects of IL-10 on DCs

DCs are central to the induction of the host antitumor immune response because of their potent antigen presenting capacity and T-cell costimulatory activity.

The majority of published studies support the immunosuppressive effects of IL-10 on DCs. *IL-10 treatment of DCs reduces their capacity to stimulate CD4$^+$ and CD8$^+$ T-cell responses and promotes T-cell anergy*. Tumor-expression of IL-10 *prevents the migration and accumulation of DCs* within tumors and can suppress GM-CSF-induced antitumor responses. *IL-10 also induces spontaneous DC apoptosis* and makes the cells more susceptible to lysis by NK cells.

The functional activities of DCs during cancer and cancer immunotherapy depend on their state of maturation. *In vitro*, exogenous *IL-10 inhibits the generation of mature monocyte-derived DCs* in response to GM-CSF and IL-4. Furthermore, DC-derived IL-10 can act in an autocrine fashion to prevent the cell's spontaneous maturation. IL-10 *promotes the differentiation of immature DCs* into macrophage-like cells with decreased MHC class II expression, suppressed phagocytic capacity, and a complete loss of the ability to express IL-12. A similar reduction of MHC class II and costimulatory molecule expression was observed in IL-10 transduced DCs when compared with control gene-modified DCs. While these IL-10 transduced DCs could not induce T-cell proliferation, they still retained the ability to augment CTL generation and NK-cell activity. Treatment of immature DCs with an anti-IL-10 antibody led to the opposite of the expected findings: it enhanced the expression of MHC, costimulatory molecules (CD80, CD83, CD86), and cytokines (TNFα, IL-12).

Thus, IL-10 exhibits complex and multifunctional activities on DCs. The immunosuppressive effects of IL-10 on DC generation, activity, migration, and survival may limit the effectiveness of the DC-based antitumor vaccines that are now being studied in clinical trials for the treatment of various forms of cancer. An improved understanding of the regulators of tumor-derived IL-10 expression and the development of appropriate anti-IL-10 strategies may improve DC trial outcomes.

3.2 Inhibition of IL-10: Implications for Therapy

Although not as well developed as anti-TGFβ strategies, several IL-10 inhibitory treatments have been considered in efforts to improve the host immune response during tumor growth *in vivo*. These include: neutralizing IL-10 antibodies, anti-IL-10 receptor antibodies, IL-10 antisense strategies, and the removal of IL-10 producing suppressor T cells. As mentioned above, aside from the expected positive effects of inhibitory therapy on relieving IL-10-induced immunosuppression, there is also reason to believe that inhibitory therapy may impinge directly on tumor cells' growth and immune-evasive tactics. For example, some melanoma cells express receptors for IL-10. When stimulated by the cytokine, growth augmentation, decreased HLA expression, and decreased adhesion molecule expression can result. Furthermore, pretreatment of melanoma cells with IL-10 completely inhibited autologous CTL-mediated lysis in one study. Thus, therapies aimed at blocking IL-10's binding to its receptor may not only make tumors more susceptible to an efficient attack by stimulating cells of the immune system, they may also directly suppress tumor-cell growth and make tumor cells more recognizable to the reinvigorated immune cells.

Antibody therapy Anti-IL-10 antibodies upregulate anticancer autoreactive T-cell responses by inhibiting the suppressive effects of the cytokine on macrophages *in vitro*. *In vivo*, anti-IL-10 promotes survival

in tumor-bearing mice, and enhances host CTL activity. It was notable in one study that antibody treatment resulted in an increased production of the Th1 cytokines IL-2, IL-12, and IFNγ by host spleen cells. Another study showed that neutralization of IL-10 with anti-IL-10 antibodies allowed an enhanced DTH response after BCG administration and augmented antitumor activity. Finally, anti-IL-10 receptor antibodies were able to abrogate the inhibition of PHA-stimulated T-cell proliferation induced by peritoneal monocytes from ovarian cancer patients. Together, these findings suggest that inhibition of IL-10 signaling in immune cells might promote the production of effective antitumor T cells.

Drug therapy Cyclophosphamide (low dose) has been used during active specific immunotherapy in patients with advanced melanoma and other metastatic cancers. The principle of using cyclophosphamide in this manner is that it may selectively reduce suppressor cell activity and decrease the production of inhibitory factors such as TGFβ, IL-10, and NO. Low-dose cyclophosphamide treatment of lymphoma-bearing rats exhibiting an impaired lymphocyte proliferative response led to a decrease in IL-10 production by T cells and, consequently, a shift in the immune response from suppression toward immunopotentiation. This state of immunostimulation was accompanied by reduced tumor growth and metastases in the cyclophosphamide-treated animals.

Removal of the source of IL-10 Gamma delta T cells accumulate in early tumor lesions and spleens of tumor-bearing animals and express both TGFβ and IL-10. Therefore, depletion of these immunosuppressive cells might lead to enhanced antitumor immune responses. In an experimental tumor model, lysis of tumor-associated gamma delta cells using a daunomycin-conjugated anti-gamma delta antibody improved host tumor-specific NK and CTL responses and resulted in tumor regression.

Complications Despite the many promising studies suggesting that therapies aimed at inhibiting IL-10 may attenuate the growth of cancers *in vivo*, not all studies report improved host antitumor immune responses following IL-10 inhibition or depletion. *In vivo, IL-10* gene transfected melanoma and mammary adenocarcinoma cells displayed significantly reduced tumorigenic and metastatic potential. Similarly, injection of IL-10 resulted in the inhibition of both B16-F10 melanoma metastases and spontaneous melanoma development in other experimental models. It is suggested that IL-10 therapy may inhibit tumor growth by downregulating the production of angiogenic molecules such as vascular endothelial growth factor (VEGF) or by augmenting NK-mediated tumor cell lysis. In several provocative studies looking at in-transit melanoma metastases in patients undergoing therapeutic vaccinations, IL-10 was found to be upregulated in responding more than in progressing lesions. Furthermore, the IL-10 level in the tumor cells was positively correlated with level of tumor regression. Taken together, these preliminary studies suggest an improved and more complete understanding of the role of IL-10 on the host immune system is needed to maximize the chances of developing a safe and effective immunomodulatory therapy.

4
Macrophage Migration Inhibitory Factor (MIF)

Macrophage migration inhibitory factor (MIF) was originally described in the 1960s as a product of activated T cells that inhibited the random migration of guinea pig macrophages. Later studies revealed the critical proinflammatory role of MIF in the host response to endotoxin. Over the past decade, a large body of work that shows multiple roles for MIF in the regulation of host inflammatory responses has been amassed. Inhibition of MIF using neutralizing antibodies significantly reduces the severity of inflammatory diseases including arthritis, peritonitis, and glomerulonephritis. In addition to its proinflammatory activities, however, our laboratory and others have identified immunosuppressive activities of MIF with regard to T-cell activation and antigen-specific T-cell responses *in vitro* and *in vivo*.

It is now known that numerous cell types constitutively express MIF, including immune cells. However, the expression of MIF is significantly higher in tumor cells than in normal cells. While enhanced tumor MIF expression often positively correlates with a more invasive phenotype and tumor recurrence after treatment, in some tumors the opposite effect is found. Recent studies suggest that in addition to its immunomodulatory effects, MIF also positively regulates cellular proliferation. Thus, MIF not only has tumor-promoting inhibitory effects on the adaptive immune system, it also promotes tumor-cell growth directly. Evidence suggesting that this is indeed the case is that transfection of antisense MIF into colon and bladder cancer cells results in decreased cellular proliferation. In addition to its pro-growth effects, the cytokine functionally inactivates the p53 tumor suppressor and inhibits p53-dependent apoptosis. Finally, MIF has pro-angiogenic properties that probably contribute to enhanced tumor growth and invasion.

The identification of the potential growth-promoting function of MIF led our group to examine the effect of MIF neutralization on tumor growth *in vivo*. We found that MIF blockade significantly inhibited tumor growth and tumor-associated angiogenesis. Similar observations demonstrating the growth promoting and pro-angiogenic functions of MIF were reported by other groups using experimental models of melanoma and colon adenocarcinoma. In agreement with the pro-tumor effects of MIF derived from anti-MIF antibody experiments, more recent work has revealed an enhanced tumor growth rate in transgenic mice overexpressing MIF.

Our laboratory examined the effect of MIF neutralization on CTL responses during tumor growth. Using the OVA-specific EG7 tumor cell model, we found that MIF immunoneutralization delays tumor growth and is associated with increased CTL activity, increased $CD8^+$ T-cell migration to the tumor, an accumulation of intratumoral $CD8^+$ T cells, and enhanced tumor-associated apoptosis. These observations are important because, as mentioned previously, cell-mediated tumor destruction relies on a rapid, extensive, and sustained infiltration of killer T cells. Furthermore, in other model systems, MIF was shown to inhibit NK-mediated lysis of endothelial cells and uveal melanoma cells. Therefore, tumor cells may secrete high levels of MIF to protect themselves from host NK- and CTL-mediated destruction. Thus, we hypothesize that tumor-derived MIF promotes host immune suppression, tumor growth, and angiogenesis

and thereby prevents tumor destruction. No cancer immunotherapeutic trials have studied MIF inhibition as of yet, but the preliminary evidence suggests that inhibition of the activity of this molecule may benefit cancer patients.

5 Prostaglandin E2 (PGE2)

Prostaglandin E2 (PGE2) is a powerful lipid mediator that influences inflammation, promotes tumor growth, and suppresses the host antitumor immune response. PGE2 belongs to a family of eicosanoids that are derived from n-3 and n-6 polyunsaturated fatty acids. Its synthesis is regulated by a series of steps following the cyclooxygenation of arachidonic acid through the action of two key regulatory enzymes: cyclooxygenase-1 and −2 (COX-1, -2). COX-1 is constitutively expressed by virtually all cells, whereas the inducible *COX-2* ("oncogene-responsive COX") is overexpressed by many premalignant and malignant neoplasms. Higher levels of COX-2 expression by cancer cells often correlate with poorer prognosis for patients with cancer.

5.1 Sources of PGE2

PGE2 is expressed by numerous tissues and cell types, including breast fibroblasts, colonic muscle and mucosal cells, peritoneal mesothelial cells, vascular endothelial cells, and macrophages, as well as many tumors and tumor cell lines. PGE2 expression by macrophages can be induced by tumor cell factors in tumor-bearing mice.

In tumors and tumor cells, PGE2 promotes proliferation, downregulates MHC class II expression, enhances invasiveness, inhibits apoptosis, and increases vascularization. Interestingly, although IL-10, a cytokine whose production is stimulated by PGE2, is generally able to attenuate PGE2 production (in a negative feedback loop), tumor cells sometimes lose this sensitivity due to loss of expression of the IL-10 receptor. In addition to its effects of the tumor itself, PGE2 also promotes host immunosuppression during tumor progression.

5.1.1 Effects of PGE2 on Monocytes/Macrophages

Host antitumor macrophage activity inversely correlates with PGE2 expression. While host macrophages are considered to be the major source of PGE2 during tumor growth, they express PGE2 receptors and can be rendered dysfunctional by this prostaglandin. Similar to IL-10, *PGE2 inhibits Th1 cytokine synthesis (TNFα, IL-12) by macrophages while promoting ineffective Th2 cytokine production* (IL-10). The PGE2-induced shift in the host antitumor activity toward an ineffective Th2 response encourages host immunosuppression during tumor growth.

5.1.2 Effects of PGE2 on T Lymphocytes

PGE2 is immunosuppressive to various types of T cells. It has inhibitory effects on lymphocyte proliferation, cytokine production, migration, and effector activity. Very early studies showed that immunization with PGE2 significantly reduced tumor-induced immunosuppression in mice (as measured by spleen cell proliferation assays), suggesting the role of PGE2 in *suppressing T-cell proliferation.* Further studies using anti-PGE2 antibodies in experimental tumor models demonstrated the role of this lipid in *suppressing polyclonal T-cell activation and CTL generation* following MLRs

with splenic T cells. PGE2 treatment of stimulated peripheral blood lymphocytes and $CD4^+$ T cells directly *decreases IL-2 and IFNγ production and promotes IL-10 production*. $CD8^+$ T-cell effector functions are reported to be more sensitive to the suppressive effects of PGE2 than those of $CD4^+$ T cells. PGE2 inhibits T-cell migration, promotes T-cell anergy, and suppresses $CD8^+$ CTL activity *in vitro* and *in vivo*.

5.1.3 Effects of PGE2 on NK and LAK Cell Activity

Similar to its immunosuppressive effects on CTLs, *PGE2 inhibits NK and LAK cell activity*. NK activity inversely correlates with monocyte-derived PGE2 levels in patients. Moreover, removal of monocytes from the PBMCs of patients with breast cancer restores NK and LAK levels to levels comparable to those seen in healthy controls.

5.1.4 Effects of PGE2 on DCs

PGE2 inhibits DC differentiation from bone marrow precursors and also inhibits several immune stimulatory DC functions. The molecule diminishes expression of MHC class II by DCs and decreases T-cell proliferation during MLRs (using DCs as stimulatory cells). The mechanism of action of these functions includes PGE2 stimulation of IL-10 production and inhibition of IL-12 production by DCs.

5.2 Inhibition of PGE2: Implications for Therapy

Slowed tumor growth and the inhibition of metastases has been observed by *in vivo* blockade of PGE2 activity using PGE2 immunization, anti-PGE2 antibodies, and inhibition of PGE2 production using COX inhibitors. Inhibition of COX-2 (and thus PGE2 production) in the Lewis lung carcinoma model resulted in a significant decrease in IL-10 production together with a significant increase in IL-12 secretion, and restored host antitumor reactivity. Similar findings have been observed in COX knockout mice and EP2 knockout mice, a strain of mice that lacks the receptor for PGE2. In mouse models of familial adenomatous polyposis (FAP), COX-2 inhibitor treatment resulted in decreased polyp formation in the colon. Furthermore, human patients with FAP that were treated with sulindac, a nonselective COX inhibitor, displayed a significant decline in size and number of colorectal polyps. Finally, it is known that patients with colorectal and lung cancer taking NSAIDs have a significantly lower mortality rate as compared with those not taking COX inhibitors.

The precise mechanisms by which COX inhibitors slow tumor growth are not well understood, but human epidemiologic data such as the findings mentioned above suggest antineoplastic activity of COX inhibitors. COX inhibitors enhance the host CTL responses in some tumor model systems. COX inhibitors also have been associated with the inhibition of angiogenesis, tumor growth, and telomerase elongation activity in tumor cells. In light of all of this data, pharmacologic COX-2 inhibition is actively being pursued as a treatment for cancer, especially using it in conjunction with more traditional therapies, with some promising early results.

6 Indoleamine 2,3-dioxygenase (IDO)

Indoleamine 2,3-dioxygenase (IDO), a tryptophan-degrading enzyme, has potent immune suppressive capabilities. Notable recent studies detail how expression of

this enzyme prevents immune rejection of the developing fetus despite antigenic differences from the mother in which it is developing. Administration of a drug that blocks IDO resulted in abortion as a result of immune attack on the fetus in these studies. The immunosuppressive abilities of IDO have also been observed in mouse allograft rejection studies and animal models of autoimmunity.

6.1 Sources of IDO

IDO is mainly expressed by monocytes, macrophages, and DCs but can be detected in lymphocytes and NK cells as well. A large number of tested tumor cell lines and primary human tumors express the enzyme. Since tryptophan is an essential amino acid and IDO mediated-tryptophan degradation has a growth inhibitory effect on tumor cells, it is important to consider why tumor cells would retain such an enzyme. A clue comes from a study which shows that expression of IDO by tumor cells imparts on the cells the ability to inhibit T-cell proliferation. Thus, despite the growth inhibitory effect of IDO, it is now believed that expression of the enzyme confers a selective advantage on tumor cells because it allows them to avoid immune attack.

6.2 Effects of IDO on Immune Cells

In mouse models, IDO expression by tumor cells generally affords tumors protection from immune attack. In one study, a tumor that is normally rejected by tumor-specific T cells persisted and grew when it expressed IDO. The magnitude of the T-cell response to the tumor in this model was diminished when IDO was expressed by the tumor cells. In another mouse model, a drug that blocks the effects of IDO delayed tumor formation.

The postulated mechanisms to explain how IDO, an amino acid degrading enzyme, modulates the efficacy of cells of immune system are (1) causing a depletion of local stores of tryptophan and/or (2) producing toxic metabolites through metabolism of the amino acid. In fact, although tryptophan depletion is growth inhibitory to the tumor cells themselves, T cells display an even more extreme sensitivity to low tryptophan levels and growth arrest in mid-G1 phase when exposed to low tryptophan conditions. Furthermore, some IDO tryptophan metabolites are indeed directly toxic to T cells *in vitro*.

Since expression of IDO by APCs causes them to inhibit T-cell proliferation, another means by which tumors could exploit the immunosuppressive ability of IDO would be to upregulate IDO expression in local DCs or to preferentially recruit IDO-expressing DCs to tumor sites. Preliminary work shows that tumor draining lymph nodes in patients with melanoma can contain large numbers of APCs that express IDO and that the presence of IDO positive cells in sentinel lymph node cells at the time of initial diagnosis is correlated with a bad prognosis for patients. Furthermore, tumor draining lymph nodes in a mouse melanoma model contained plasmacytoid DCs that constitutively expressed IDO and could potently suppress T-cell responses. Adoptive transfer of these cells into naïve mice resulted in T-cell anergy to antigens expressed by these DCs.

6.3 Implications for Therapy

Given the preliminary promising findings in tumor models, it is not surprising

that consideration is now being given to developing anti-IDO therapies for use in cancer, especially as adjuvant therapy in combination with more traditional chemotherapeutic approaches. While the preponderance of experimental evidence to date suggests that anti-IDO therapy may be fruitful for the treatment of cancer, a degree of caution is warranted. Two studies have shown that, in contrast to its immunosuppressive effects on T cells, IDO expression is important for proper NK cell function. Inhibition of the enzyme adversely affected NK cytotoxic activity *in vitro* and *in vivo* in these studies. Furthermore, in one of these studies, pharmacologic inhibition of IDO resulted in enhanced growth of B16 melanoma cells after subcutaneous injection.

7 Cell Surface Molecules

7.1 Fas/FasL

Fas is a membrane receptor in the tumor necrosis factor receptor family that induces apoptosis after binding Fas-ligand (FasL). In fact, this cell surface molecule was first described in 1989 when an anti-Fas antibody was shown to trigger T-cell apoptosis. FasL is a 40-kDa membrane protein related to TNF that functions not only in its membrane-bound form but also in a soluble form, soluble FasL (sFasL). Soluble FasL is derived from the proteolytic cleavage of surface-expressed FasL. Fas/FasL interactions play critical roles in the deletion of autoreactive T cells, CTL-mediated cell death, and activation-induced cell death of T cells. In addition to its proapoptotic effects, FasL binding to Fas can also stimulate the production of proinflammatory cytokines.

7.1.1 Sources of FasL

Constitutive expression of FasL is found in the lung, testis, and eye. With regard to immune cells, FasL expression is found in activated B cell, T cell, macrophage, and NK cell cultures. A wide array of human tumor cell lines and human tumors *in vitro* and *in situ* produce FasL. Moreover, within tumors FasL can be cleaved from the cell surface by matrix metalloproteinases to produce soluble FasL. Elevated FasL expression in several human tumors is significantly associated with poorer prognosis, enhanced tumor metastatic rate, and increased mortality.

7.1.2 Tumor Evasion of FasL-mediated Toxicity

Many human tumors express more Fas on their surface than FasL. It might therefore be assumed that the expression of FasL by tumor cells would be detrimental for cancers as it could initiate an apoptotic cascade in neighboring tumor cells. In fact, downregulation of Fas on the tumor cell surface does correlate with poorer prognosis in many tumors. However, this is not the only mechanism by which tumor cells acquire resistance to FasL-mediated death. Several other mechanisms enable neoplastic cells to avoid cell death triggered by Fas/FasL interactions, including: (1) expression of nonfunctional Fas on the cell surface (2) overexpression of inhibitors of the Fas/FasL apoptotic pathway such as cFLIP, bcl-2, bcl-xL, and Fas-associated phosphatase-1 and (3) secretion of decoy receptors that bind and neutralize FasL.

7.1.3 Tumor-derived FasL is Functional

FasL expressed by human tumors induces apoptosis of activated Fas^+ T lymphocytes *in vitro*. Cell death can be prevented by anti-FasL antibodies, suggesting that it is

indeed initiated by Fas/FasL interaction. Moreover, augmenting expression of FasL by tumor cells results in enhanced degree of apoptosis of lymphocytes. In pathologic sections from human tumors, higher local FasL expression in particular areas correlates with elevated levels of lymphocyte apoptosis in those areas and lower numbers of viable lymphocytes. In several murine models, FasL expression on tumor cells facilitated evasion from immune system detection.

7.1.4 FasL Expression is Correlated with Improved Prognosis in Some Tumor Models

As mentioned above, FasL binding to Fas triggers not only apoptosis but also proinflammatory signals in some cells. For example, Fas/FasL interaction in DCs results in release of IL-1β and TNFα and in smooth muscle cells results in production of IL-8, IL-1α, MCP-1, and IL-6. In contrast to the previously mentioned *in vivo* data, in some tumor models the proinflammatory effects of Fas/FasL interactions dominate over the immune suppressive capabilities of tumor-expressed FasL. Tumor cells transfected with FasL were rapidly rejected after implantation in several different model systems. The rejected tumor sites were characterized by an influx of neutrophils, which were believed to mediate the tumor destruction. In a different model system, tumor cells transfected with antisense FasL displayed no improvement in clinical course as compared with controls. The precise mechanisms that govern whether the effects of Fas/FasL interaction will be predominantly pro- or anti-inflammatory are an area of active investigation. Work to date has focused on the level of FasL expression and the role of local cytokines in shaping the immune response to Fas/FasL binding. As of now, there is a debate as to whether or not FasL expressed by human tumor cells truly acts as an important immunosuppressive mediator. With further study, a means of exploiting the FasL expression displayed by many tumor cells for cancer immunotherapy may still prove to be viable.

7.2 MHC Class I: Classical and Nonclassical Receptors

7.2.1 Classical MHC Class I

CTLs recognize their targets by peptides bound to cell surface MHC class I and are crucial for the generation of antitumor immune responses. Given the requirement of MHC class I expression for recognition by CTLs, it is interesting to note that a number of tumors display decreased expression of cell surface MHC class I proteins. Indeed, loss of HLA molecule expression is associated with a poorer prognosis of some cancers. Mechanisms tumors use to decrease cell surface HLA expression include downregulation of expression of specific HLA alleles: HLA-A, HLA-B, HLA-C and also downregulation of expression of β2 microglobulin or transporter of antigen (*TAP*) genes. Dysfunction or downregulation of the gene products from the latter two can cause a complete loss of MHC class I expression. IL-10, whether tumor or immune cell-derived, is probably integrally involved in the downregulation of MHC class I on tumor cells.

In a mouse model of lung cancer, expression of MHC class I by metastatic cells depended on the immune status of the host. In immune competent hosts, lung metastases were MHC class I negative; positivity was seen only in immunocompromised mice. This suggests that in immune competent hosts, tumors downregulate MHC

class I in an attempt to "hide" from the immune system. Furthermore, the tumors derived from the immunodeficient mice were rejected when transplanted into the immunocompetent mice. Aside from this evidence, however, there is a paucity of data to directly support the hypothesis that downregulated MHC class I expression is crucial for tumor immune evasion *in vivo*. Furthermore, downregulation of MHC class I is associated with enhanced sensitivity of cells to NK cell-mediated cytotoxicity. In fact, in an animal model, loss of $\beta 2$ microglobulin, leading to a downregulation of MHC class I on tumor cells, resulted in increased susceptibility of the cells to NK-mediated cytotoxicity. Thus, in sum, while downregulation of MHC class I remains an intuitively logical candidate mechanism for tumor immune evasion, further studies will be needed to clarify this concept.

7.2.2 HLA-G

In addition to downregulating classical HLA antigens, tumor cells can also upregulate expression of the alternative HLA molecule HLA-G to avoid immune detection. HLA-G, first described with regard to its immunosuppressive effects in the context of maternofetal tolerance, encodes proteins with similarities to proteins of the MHC class I family. HLA-G proteins can be either membrane bound (the full-length isoform HLA-G1 and the truncated isoforms -G2, -G3, and -G4) or soluble (HLA-G5, -G6, or, -G7). HLA-G is expressed by a number of cancer cell types. Enhanced HLA-G expression in melanocytes correlates with malignancy better than loss of classical MHC class I expression and elevated levels of soluble HLA-G portend a poor prognosis for melanoma patients. HLA-G proteins have become an area of active interest in tumor biology as a result of their immunosuppressive effects on NK cells, T cells, and DCs.

Expression of the full-length HLA-G1 or the truncated HLA-G2, -G3, and -G4 protects tumor cells from NK- and CTL-mediated cytotoxicity. This protective effect can be blocked with an anti-HLA-G1 specific monoclonal antibody. HLA-G positive glioma cells were found to inhibit priming of CTLs and alloproliferative responses. Both cell surface and soluble HLA-G can inhibit $CD4^+$ T-cell proliferation in MLRs. Many of HLA-G1's immunosuppressive effects are a result of its direct stimulation of inhibitory receptors on the surface of immune cells including the receptors KIR2DL4, ILT2, and ILT4. Interaction of HLA-G proteins with ILT2 and ILT4 receptors on monocytes can inhibit secretion of proinflammatory cytokines and chemokines.

The soluble protein HLA-G5, generated as a product of alternative splicing of the *HLA-G* gene transcript that lacks the transmembrane and cytoplasmic domains, blocks NK and CTL-mediated cell death. Since soluble HLA-G is a nonclassical MHC class I molecule, it is not surprising that this molecule is able to bind CD8 expressed on T cells. Subsequent to an HLA-G/CD8 interaction, $CD8^+$ T cells undergo apoptosis mediated by Fas/FasL interaction. A subpopulation of NK cells that express CD8 on their cell surfaces are also susceptible to apoptosis induced by soluble HLA-G5. Still, some authors have not found an increase in $CD8^+$ T cell apoptosis when these cells are cocultured with HLA-G5-positive cells. Therefore, T-cell cytotoxicity mediated by soluble HLA-G5 may be dependent on other mitigating or promoting factors in the immune cell's environment. In addition to its apoptosis-inducing

effects, soluble HLA-G impairs the alloproliferative response and decreases the Th1/Th2 secretion profile from $CD4^+$ T cells.

HLA-G can also produce immunomodulatory effects by upregulating functional tumor cell surface HLA-E, another nonclassical HLA molecule. HLA-E, in turn, binds to an inhibitory receptor on NK and T cells to block their cytotoxic and immune stimulatory functions.

Unfortunately, as there is no rodent homolog of HLA-G, there is a lack of well-controlled data to support the hypothesis that HLA-G proteins are important for tumor-induced immune suppression *in vivo*. Nevertheless, the manipulation of MHC class I molecules on the cell surface of tumor cells, whether classical or nonclassical, is a reasonable target for anticancer immunotherapy. The finding that nonclassical molecules act by providing inhibitory signals to killer immune cells through cell surface receptors implies that interruption of those interactions with antibodies or small molecule inhibitors may lift tumor-induced immune suppression.

7.3 B7-H1

Another surface molecule tumor cells use to actively inhibit immune responses is B7-H1, a member of the B7 family of costimulatory molecules. This molecule, also known as programmed death ligand 1 (PD-L1), can induce T-cell apoptosis upon binding the programmed death-1 (PD-1) receptor on a cell's surface. PD-1 is expressed on activated $CD4^+$ T cells, $CD8^+$ T cells, B cells, and myeloid cells. B7-H1 is normally expressed only in macrophages, but it is also found in many human tumors. The tumor-expressed molecule is able to stimulate CTL cell death through induction of apoptosis. Furthermore, anti-B7-H1 antibodies can abrogate this apoptotic effect. While the importance of this cell surface molecule in human cancers *in vivo* is currently unclear, therapy aimed at interrupting the interaction between B7-H1 and PD-1 may ultimately prove to be a useful means to alleviate tumor-induced host immunosuppression.

7.4 RCAS1

RCAS1 was initially described as a cell surface molecule, found on many tumor cells, capable of inducing immunosuppression. In early studies, it was suggested that binding of RCAS1 to a putative receptor inhibited growth and induced apoptosis in a number of cells including NK, B cells, and T cells. More recently, some of these claims have been questioned. RCAS1 was shown by another group to be a predominantly Golgi-located protein that altered cell surface expression of expression of tumor-associated, normally cryptic *O*-linked glycan structures. The structures are believed to be associated with adhesion, invasion, and metastasis of cancer cells. Thus, although possibly not a cell surface receptor as originally thought, RCAS1 may still prove to be an important factor for tumor growth and survival. Notably elevated levels of tumor RCAS expression correlate with poorer prognosis in a number of human cancers. Studying the effect of RCAS1-expressing tumor cells on neighboring immune cells warrants further consideration. Disruption of these effects, whether mediated directly by RCAS1 or by cell surface glycan structures, may still prove to be a means to interfere with tumor cell survival *in vivo*.

7.5
Other Cell Surface Molecules

A number of other cell surface molecules have been associated with tumor-induced immune evasion. These include: complement receptors, CD40, and CD70. Furthermore, it has been suggested that tumors express receptors for growth factors, such as IL-2, as a means of depleting these factors from the environment so they cannot be used by local immune cells. It is extremely unlikely that all of the different cell surface molecules tumors can use to avoid immune attack have been considered at this time. The list of molecules used by tumors for this purpose will probably continue to grow for years to come.

8
Polyamines

Polyamines (putrescine, spermine, and spermadine) are naturally occurring small organic cations. Polyamines are critical for both cell proliferation and differentiation. Much of our knowledge of polyamine bioactivity stems from the use of inhibitors of polyamine biosynthesis (see below).

8.1
Sources of Polyamines

Polyamines are found in all living cells. Ingested foods are a significant exogenous source of polyamines. Polyamine levels (in the urine, serum, and red blood cells) have been positively correlated with the stage of malignancy and tumor burden for several human cancers. In addition, because dying tumor cells release high levels of polyamines, circulating polyamine levels have been proposed to serve as indicators of the success of chemotherapeutic agents.

Polyamine biosynthesis can be selectively blocked by DFMO (D, L-2-(difluoromethyl) ornithine). DFMO is an irreversible inhibitor of ornithine decarboxylase (ODC), which mediates the first step in polyamine synthesis. DFMO depletes cells of intracellular putrescine and spermidine (but not spermine) and inhibits cell proliferation without causing cell death. Since DFMO treatment is associated with weight loss and toxicity, less toxic polyamine analogs have been developed. Polyamine analogs such as N1, N8-bis(ethyl)spermidine are reported to rapidly deplete intracellular polyamines and induce cell death. Because exogenous polyamines found in food contribute to the polyamine status of an individual, blockade of endogenous and exogenous (polyamine deficient diet) polyamines is sometimes required to identify the role of polyamines on host immunosuppression during tumor growth.

Inhibitors of polyamine biosynthesis decrease tumor growth and metastases *in vivo*. Interestingly, these inhibitors prevent growth and metastases in experimental models of B16 melanoma and Lewis Lung carcinoma when tumor cells are subcutaneously implanted, but not when tumor cells are intravenously injected. Much of the antitumor activity of ODC inhibitors is attributed to their antiproliferative activity. However, polyamines also appear to play a role in the differentiation of tumor cells, protease expression, invasiveness, protection against apoptosis, malignant transformation, and tumor-associated angiogenesis. Inhibitors of polyamine biosynthesis prevent proliferation of normal and tumor cells *in vitro,* including small cell lung carcinoma, mammary tumor, and melanoma cells. Tumor cells may be more dependent on polyamines for proliferation than normal host cells due to their

increased proliferative rate. In addition to their myriad effects on tumor cells, polyamines serve as suppressors of the host immune system.

8.1.1 Effects of Polyamines on Monocytes/macrophages

Polyamine levels are associated with numerous functional activities of macrophages/monocytes that may contribute to host immunosuppression during tumor growth. Inhibitors of polyamine biosynthesis impair TNFα-mediated macrophage activation and decrease the respiratory burst of macrophages. These inhibitors also block both NO and cytokine synthesis by macrophages. A correlation between total polyamine levels and monocyte phagocytosis is found in patients with colorectal cancer. Accordingly, treatment of tumor-bearing animals with polyamine deprivation combined with low dose cyclophosphamide resulted in a synergistic inhibition of tumor growth and metastasis associated with enhanced macrophage tumoricidal activities.

8.1.2 Effects of Polyamines on T Lymphocytes

Polyamine synthesis is required for the generation of CTLs and IL-2 responsiveness in these T cells. CTL differentiation is affected by the inhibition of polyamine biosynthesis to a greater extent than CTL proliferation. *These data suggest that polyamines are immunostimulating for T cells in vitro. In vivo,* while polyamine deprivation in animals grafted with Lewis lung carcinoma tumors reversed tumor-induced immunosuppression (as measured by splenic IL-2 production and total $CD4^+$ and $CD8^+$ T-cell populations), it was not associated with any changes in host CTL activity.

8.1.3 Effects of Polyamines on NK Cells

Inhibition of polyamine biosynthesis and polyamine deprivation *in vivo* is associated with decreased tumor growth and metastases and with significantly enhanced NK cytotoxic activity. Thus, *polyamines are inhibitory for NK cells.*

8.2 Inhibition of Polyamine Biosynthesis: Implications for Therapy

As described above, inhibitors of ODC, polyamine analogs, and polyamine deficient diets have been used in animal models to suppress polyamine levels. Both ODC inhibitors and polyamine analogs have been useful in identifying the role of polyamines on tumor growth and host immunosuppression. However, polyamine inhibition as a target for cancer treatment is complicated by the ubiquitous expression of polyamines and the contribution of food sources to an individual's polyamine status. Furthermore, polyamines play a critical role in the proliferation and survival of normal cells as well as cancer cells. Future studies are required to identify mechanism(s) to target polyamine inhibitors specifically to neoplastic cells. This will improve the therapeutic index of this class of medicine.

9 Tumor Shed Immunosuppressive Molecules

Tumors can shed antigens into the environment to hinder host immune attacks. The most well-characterized immunosuppressive antigens shed from tumors are the mucins. Mucins are high molecular weight glycoproteins present on most secretory epithelial tissues. Most human

epithelial tumors overexpress aberrantly glycosylated mucins referred to as MUC-1 and MUC-2. MUC-1 has been employed as an immunotherapeutic target for cancer patients because of its ability to induce both anti-MUC-1 antibodies and CTL responses. However, soluble MUC-1 and MUC-2 have been shown to be immunosuppressive tumor antigens. Accordingly, high serum levels of MUC-1 correlate with poor survival and enhanced metastatic disease for adenocarcinoma (breast, ovarian, pancreatic, and colorectal) following active specific immunotherapy.

The release of MUC antigens from the surface of tumor cells might promote tumorigenesis and subvert the host immune system by two mechanisms. First, *release of MUC-1 by tumor cells could inhibit the effectiveness of the host anti-MUC-specific CTL response* due to the downregulation of specific tumor antigen expression. Second, the release of tumor-associated MUC-1 molecules might have *direct immunosuppressive effects on the host immune system. In vitro,* MUC1 transfection into melanoma cells decreases their susceptibility to CTL induced cytolysis. Mucins also *inhibit T-cell proliferation, suppress IL-2* and *IFNγ production* by $CD4^+$ T cells, and *negatively regulate T-cell activation* at the resting stage.

Similar to mucins, the soluble forms of annexin II and carcinoembryonic antigen (CEA), two other tumor-shed antigens, appear to function as immunosuppressive factors. Annexin is a membrane-associated glycoprotein expressed by normal and tumor cells. Annexin II expression has been correlated with higher grade tumors and poorer prognosis in cancer patients. In its soluble form, *annexin II inhibits T-cell proliferation* and *suppresses IgG and IgM secretion from peripheral blood mononuclear cells.* CEA is another example of a "tumor antigen" currently being evaluated for use in anticancer immunotherapeutic strategies. CEA is frequently present on colorectal tumors. Elevated circulating CEA levels correlate with poorer survival rates for patients. *The shedding of CEA* from the surface of tumor cells has been proposed to *suppress LAK cytotoxic activity* and *inhibit the activities of NK and Th1 cells.*

Finally, tumor-shed gangliosides, which are glycosphingolipids released into the tumor cell microenvironment, have immunosuppressive properties. These molecules inhibit antigen process and presentation, lymphocyte proliferation, and NK cytotoxicity *in vitro. In vivo,* inhibition of syngeneic antitumor immune responses has been shown with injection of highly purified gangliosides.

Thus, multiple shed tumor-associated antigens may function as immunosuppressive agents and inhibition of these factors should be considered in attempts to optimize anticancer immunotherapy.

10 Conclusion

Unfortunately, the struggle between hosts and tumors often results in defeat for the host. Despite the fact that tumors possess antigens capable of stimulating the host immune system, they succeed in escaping the innate and acquired antitumor host immune responses through multiple and complex mechanisms. Tumors evade the immune response by expressing inhibitory cell surface molecules and/or by releasing factors including: cytokines, lipids, toxic metabolites, polyamines, and tumor antigens. These molecules can suppress the host immune response and at the same time stimulate tumor growth and survival.

Much is known regarding how single tumor-derived suppressive factors impair various immune cell subtypes in isolation. However, when the effects of numerous factors on different cell types are studied in combinations in more complex systems, their immunoregulatory activities can sometimes prove to be inconsistent with expected results. Thus, the outcome of tumor rejection or growth in a host depends on the converging influence of many different host antitumor molecules and tumor-derived survival elements. These come together to influence the ultimate outcome of tumor destruction or survival in complicated and interrelated manners. Important factors that determine the tumor and host's fate include: tumor cell type, host immune competence, tumor location, cytokine milieu, timing of appearance of different factors, origin of specific immune cells, and activation status of specific immune cells.

It is almost certain that we have not yet identified all of the tumor-derived immunosuppressive factors. Thus, more extensive studies are required to improve our understanding of how tumors impair the host immune response. This information will be invaluable for the future development of improved therapeutics designed to block tumor-associated host immunosuppression in cancer patients. Ultimately, the aim of this research is to allow physicians to improve patients' antitumor immunity, whether through conventional cancer treatments, novel immunotherapeutic approaches, or more likely combinations of the two. It is humbling to recognize that despite all of the tremendous advances made to date in the field of anticancer immunotherapy, we remain at a very early stage in the practical application of this knowledge to patients.

See also Antigen Presenting Cells (APCs); Cytokines: Interleukins; Dendritic Cells; Immunology; Immunology of Cancer; Innate Immunity; Molecular Mediators: Cytokines; Oncology, Molecular.

Bibliography

Books and Reviews

Asadullah, K., Sterry, W., Volk, H.D. (2003) Interleukin-10 therapy – Review of a new approach, *Pharmacol. Rev.* **55**, 241–269.

Baldus, S.E., Engelmann, K., Hanisch, F.G. (2004) MUC1 and the MUCs: A family of human mucins with impact in cancer biology, *Crit. Rev. Clin. Lab. Sci.* **41**, 189–231. &

Dunn, G.P., Bruce, A.T., Ikeda, H., Old, L.J., Schreiber, R.D. (2002) Cancer immunoediting: from immunosurveillance to tumor escape, *Nat. Immunol.* **3**, 991–998.

French, L.E., Tschopp, J. (2002) Defective death receptor signaling as a cause of tumor immune escape, *Semin. Cancer Biol.* **12**, 51–55.

Metz, C.N., Bucala, R. (1997) Role of macrophage migration inhibitory factor in the regulation of the immune response, *Adv. Immunol.* **66**, 197–223.

Mitchell, R.A., Bucala, R. (2000) Tumor growth-promoting properties of macrophage migration inhibitory factor (MIF), *Semin. Cancer Biol.* **10**, 359–366.

Pawelec, G. (2004) Tumour escape: antitumour effectors too much of a good thing? *Cancer Immunol. Immunother.* **53**, 262–274.

Rouas-Freiss, N., Moreau, P., Menier, C., Carosella, E.D. (2003) HLA-G in cancer: a way to turn off the immune system, *Semin. Cancer Biol.* **13**, 325–336.

Seo, N., Hayakawa, S., Tokura, Y. (2002) Mechanisms of immune privilege for tumor cells by regulatory cytokines produced by innate and acquired immune cells, *Semin. Cancer Biol.* **12**, 291–300.

Siegel, P.M., Massague, J. (2003) Cytostatic and apoptotic actions of TGF-beta in homeostasis and cancer, *Nat. Rev. Cancer* **3**, 807–820.

Weiss, E., Mamelak, A.J., La Morgia, S., Wang, B., Feliciani, C., Tulli, A., Sauder, D.N. (2004) The role of interleukin 10 in the pathogenesis and potential treatment of skin diseases, *J. Am. Acad. Dermatol.* **50**, 657–675.

Primary Literature

Aarli, A., Matre, R. (1998) Suppression of immunoglobulin secretion by soluble annexin II, *Scand. J. Immunol.* **48**, 522–526.

Abe, R., Peng, T., Sailors, J., Bucala, R., Metz, C.N. (2001) Regulation of the CTL response by macrophage migration inhibitory factor, *J. Immunol.* **166**, 747–753.

Agrawal, B., Krantz, M.J., Reddish, M.A., Longenecker, B.M. (1998) Cancer-associated MUC1 mucin inhibits human T-cell proliferation, which is reversible by IL-2, *Nat. Med.* **4**, 43–49.

Alleva, D.G., Burger, C.J., Elgert, K.D. (1994) Tumor-induced regulation of suppressor macrophage nitric-oxide and TNF-alpha production. Role of tumor-derived IL-10, TGF-beta, and prostaglandin E(2), *J. Immunol.* **153**, 1674–1686.

Ashley, D.M., Kong, F.M., Bigner, D.D., Hale, L.P. (1998) Endogenous expression of transforming growth factor beta 1 inhibits growth and tumorigenicity and enhances Fas-mediated apoptosis in a murine high-grade glioma model, *Cancer Res.* **58**, 302–309.

Attiga, F.A., Fernandez, P.M., Weeraratna, A.T., Manyak, M.J., Patierno, S.R. (2000) Inhibitors of prostaglandin synthesis inhibit human prostate tumor cell invasiveness and reduce the release of matrix metalloproteinases, *Cancer Res.* **60**, 4629–4637.

Bacher, M., Metz, C.N., Calandra, T., Mayer, K., Chesney, J., Lohoff, M., Gemsa, D., Donnelly, T., Bucala, R. (1996) An essential regulatory role for macrophage migration inhibitory factor in T-cell activation, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 7849–7854.

Barth, R.J., Coppola, M.A., Green, W.R. (1996) In vivo effects of locally secreted IL-10 on the murine antitumor immune response, *Ann. Surg. Oncol.* **3**, 381–386.

Bellone, G., Asteamezaga, M., Trinchieri, G., Rodeck, U. (1995) Regulation of NK cell functions by TGF-beta-1, *J. Immunol.* **155**, 1066–1073.

Cerwenka, A., Kovar, H., Majdic, O., Holter, W. (1996) Fas- and activation-induced apoptosis are reduced in human T cells preactivated in the presence of TGF-beta, *J. Immunol.* **156**, 459–464.

Cerwenka, A., Bevec, D., Majdic, O., Knapp, W., Holter, W. (1994) TGF-beta-1 is a potent inducer of human effector T-cells, *J. Immunol.* **153**, 4367–4377.

Chamaillard, L., Quemener, V., Havouis, R., Moulinoux, J.P. (1993) Polyamine deprivation stimulates natural-killer-cell activity in cancerous mice, *Anticancer Res.* **13**, 1027–1034.

Chamaillard, L., Catros-Quemener, V., Delcros, J.G., Bansard, J.Y., Havouis, R., Desury, D., Commeurec, A., Genetet, N., Moulinoux, J.P. (1997) Polyamine deprivation prevents the development of tumour-induced immune suppression, *Br. J. Cancer* **76**, 365–370.

Chesney, J., Metz, C., Bacher, M., Peng, T., Meinhardt, A., Bucala, R. (1999) An essential role for macrophage migration inhibitory factor (MIF) in angiogenesis and the growth of a murine lymphoma, *Mol. Med.* **5**, 181–191.

Ding, A., Nathan, C.F., Graycar, J., Derynck, R., Stuehr, D.J., Srimal, S. (1990) Macrophage deactivating factor and transforming growth factor-beta-1, factor-beta-2, and factor-beta-3 inhibit induction of macrophage nitrogen-oxide synthesis by IFN-Gamma, *J. Immunol.* **145**, 940–944.

Ding, L., Linsley, P.S., Huang, L.Y., Germain, R.N., Shevach, E.M. (1993) IL-10 inhibits macrophage costimulatory activity by selectively inhibiting the up-regulation of B7 expression, *J. Immunol.* **151**, 1224–1234.

Dong, H., Strome, S.E., Salomao, D.R., Tamura, H., Hirano, F., Flies, D.B., Roche, P.C., Lu, J., Zhu, G., Tamada, K., Lennon, V.A., Celis, E., Chen, L. (2002) Tumor-associated B7-H1 promotes T-cell apoptosis: A potential mechanism of immune evasion, *Nat. Med.* **8**, 793–800.

Espevik, T., Figari, I.S., Ranges, G.E., Palladino, M.A. (1988) Transforming growth factor-beta1 (TGF-Beta1) and recombinant human-tumor necrosis factor-alpha reciprocally regulate the generation of lymphokine-activated killer cell-activity – comparison between natural porcine platelet-derived TGF-beta1 and TGF-beta2, and recombinant human TGF-beta1, *J. Immunol.* **140**, 2312–2316.

Fargeas, C., Wu, C.Y., Nakajima, T., Cox, D., Nutman, T., Delespesse, G. (1992) Differential effect of transforming growth-factor-beta on

the synthesis of Th1-Lymphokines and Th2-like lymphokines by human lymphocytes-T, *Eur. J. Immunol.* **22**, 2173–2176.

Fingerle-Rowson, G., Petrenko, O., Metz, C.N., Forsthuber, T.G., Mitchell, R., Huss, R., Moll, U., Muller, W., Bucala, R. (2003) The p53-dependent effects of macrophage migration inhibitory factor revealed by gene targeting, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9354–9359.

Fiorentino, D.F., Bond, M.W., Mosmann, T.R. (1989) Two types of mouse T-helper cell. IV. Th2 clones secrete a factor that inhibits cytokine production by Th1 clones, *J. Exp. Med.* **170**, 2081–2095.

Fiorentino, D.F., Zlotnik, A., Mosmann, T.R., Howard, M., Ogarra, A. (1991) IL-10 inhibits cytokine production by activated macrophages, *J. Immunol.* **147**, 3815–3822.

Genestier, L., Kasibhatla, S., Brunner, K., Green, D.R. (1999) Transforming growth factor beta 1 inhibits Fas ligand expression and subsequent activation-induced cell death in T cells via downregulation of c-Myc, *J. Exp. Med.* **189**, 231–239.

Gorelik, L., Flavell, R.A. (2001) Immune-mediated eradication of tumors through the blockade of transforming growth factor-beta signaling in T cells, *Nat. Med.* **7**, 1118–1122.

Gridley, D.S., Sura, S.S., Uhm, J.R., Lin, C.H., Kettering, J.D. (1993) Effects of anti-transforming growth-factor-beta antibody and interleukin-2 in tumor-bearing mice, *Cancer Biother.* **8**, 159–170.

Groux, H., Bigler, M., Devries, J.E., Roncarolo, M.G. (1996) Interleukin-10 induces a long-term antigen-specific anergic state in human CD4(+) T cells, *J. Exp. Med.* **184**, 19–29.

Heby, O. (1981) Role of polyamines in the control of cell-proliferation and differentiation, *Differentiation* **19**, 1–20.

Huang, M., Sharma, S., Mao, J.T., Dubinett, S.M. (1996) Non-small cell lung cancer-derived soluble mediators and prostaglandin E(2) enhance peripheral blood lymphocyte IL-10 transcription and protein production, *J. Immunol.* **157**, 5512–5520.

Hudson, J.D., Shoaibi, M.A., Maestro, R., Carnero, A., Hannon, G.J., Beach, D.H. (1999) A proinflammatory cytokine inhibits p53 tumor suppressor activity, *J. Exp. Med.* **190**, 1375–1382.

Jun, C.D., Choi, B.M., Kim, S.U., Lee, S.Y., Kim, H.M., Chung, H.T. (1995) Downregulation of transforming growth-factor-beta gene-expression by antisense oligodeoxynucleotides increases recombinant interferon-gamma-induced nitric-oxide synthesis in murine peritoneal-macrophages, *Immunology* **85**, 114–119.

Kaneko, H., Hibasami, H., Mori, K., Kawarada, Y., Nakashima, K. (1998) Apoptosis induction in human breast cancer MRK-nu-1 cells by a polyamine synthesis inhibitor, methylglyoxal his (cyclopentylamidinohydrazone) (MGBCP), *Anticancer Res.* **18**, 891–896.

Kasid, A., Bell, G.I., Director, E.P. (1988) Effects of transforming growth factor-beta on human lymphokine-activated killer cell precursors – autocrine inhibition of cellular proliferation and differentiation to immune killer cells, *J. Immunol.* **141**, 690–698.

Kehrl, J.H., Wakefield, L.M., Roberts, A.B., Jakowlew, S., Alvarez-Mon, M., Derynck, R., Sporn, M.B., Fauci, A.S. (1986) Production of transforming growth-factor-beta by human lymphocytes-T and its potential role in the regulation of T-cell growth, *J. Exp. Med.* **163**, 1037–1050.

Kim, J.A., Dayton, M.A., Aldrich, W., Triozzi, P.L. (1999) Modulation of CD4 cell cytokine production by colon cancer-associated mucin, *Cancer Immunol. Immunother.* **48**, 525–532.

Koppelman, B., Neefjes, J.J., Devries, J.E., Malefyt, R.D. (1997) Interleukin 10 inhibits arrival of mature MHC class II alpha beta peptide complexes to the plasma membrane in human monocytes, *J. All. Clin. Immunol.* **99**, 1830.

Kubota, S., Kiyosawa, H., Nomura, Y., Yamada, T., Seyama, Y. (1997) Ornithine decarboxylase overexpression in mouse 10T1/2 fibroblasts: Cellular transformation and invasion, *J. Natl. Cancer Inst.* **89**, 567–571.

Kundu, N., Fulton, A.M. (1997) Interleukin-10 inhibits tumor metastasis, downregulates MHC class I, and enhances NK lysis, *Cell. Immunol.* **180**, 55–61.

Kundu, N., Beaty, T.L., Jackson, M.J., Fulton, A.M. (1996) Antimetastatic and antitumor activities of Interleukin 10 in a murine model of breast cancer, *J. Natl. Cancer Inst.* **88**, 536–541.

Lala, P.K., Parhar, R.S., Singh, P. (1986) Indomethacin therapy abrogates the prostaglandin-mediated suppression of natural-killer activity in tumor-bearing mice and prevents tumor-metastasis, *Cell. Immunol.* **99**, 108–118.

Liau, L.M., Fakhrai, H., Black, K.L. (1998) Prolonged survival of rats with intracranial C6 gliomas by treatment with TGF-beta antisense gene, *Neurol. Res.* **20**, 742–747.

Ludviksson, B.R., Seegers, D., Resnick, A.S., Strober, W. (2000) The effect of TGF-beta 1 on immune responses of naive versus memory CD4(+) Th1/Th2 T cells, *Eur. J. Immunol.* **30**, 2101–2111.

Maeda, H., Tsuru, S., Shiraishi, A. (1994) Improvement of macrophage dysfunction by administration of anti-transforming growth-factor-beta antibody in El4-bearing hosts, *Jpn. J. Cancer Res.* **85**, 1137–1143.

Maggard, M., Meng, L., Ke, B., Allen, R., Devgan, L., Imagawa, D.K. (2001) Antisense TGF-beta 2 immunotherapy for hepatocellular carcinoma: Treatment in a rat tumor model, *Ann. Surg. Oncol.* **8**, 32–37.

Malefyt, R., Haanen, J., Spits, H., Roncarolo, M.G., te Velde, A., Figdor, C., Johnson, K., Kastelein, R., Yssel, H., de Vries, J.E. (1991) Interleukin-10 (Il-10) and Viral-Il-10 strongly reduce antigen-specific human T-Cell proliferation by diminishing the antigen-presenting capacity of monocytes via down-regulation of class-II major histocompatibility complex expression, *J. Exp. Med.* **174**, 915–924.

Malefyt, R.D., Yssel, H., Devries, J.E. (1993) Direct effects of IL-10 on subsets of human CD4+ T-Cell clones and resting T-Cells – specific-inhibition of IL-2 production and proliferation, *J. Immunol.* **150**, 4754–4765.

Malefyt, R.D., Abrams, J., Bennett, B., Figdor, C.G., Devries, J.E. (1991) Interleukin-10(IL-10) inhibits cytokine synthesis by human monocytes – An autoregulatory role of IL-10 produced by monocytes, *J. Exp. Med.* **174**, 1209–1220.

Malygin, A.M., Meri, S., Timonen, T. (1993) Regulation of natural-killer-cell activity by transforming growth-factor-beta and prostaglandin-E2, *Scand. J. Immunol.* **37**, 71–76.

Marx, M., Townsend, C.M., Barranco, S.C., Glass, E.J., Thompson, J.C. (1987) Treatment of hamster pancreatic-cancer with alpha-difluoromethylornithine, an inhibitor of polyamine biosynthesis, *J. Natl. Cancer Inst.* **79**, 543–548.

Matsuda, M., Salazar, F., Petersson, M., Masucci, G., Hansson, J., Pisa, P., Zhang, Q.J., Masucci, M.G., Kiessling, R. (1994) Interleukin-10 pretreatment protects target-cells from tumor-specific and allo-specific cytotoxic T-cells and down-regulates HLA class-I expression, *J. Exp. Med.* **180**, 2371–2376.

Mule, J.J., Schwarz, S.L., Roberts, A.B., Sporn, M.B., Rosenberg, S.A. (1988) Transforming growth factor-beta inhibits the in vitro generation of lymphokine-activated killer cells and cyto-toxic T-cells, *Cancer Immunol. Immunother.* **26**, 95–100.

Munn, D.H., Sharma, M.D., Hou, D., Baban, B., Lee, J.R., Antonia, S.J., Messina, J.L., Chandler, P., Koni, P.A., Mellor, A.L. (2004) Expression of indoleamine 2,3-dioxygenase by plasmacytoid dendritic cells in tumor-draining lymph nodes, *J. Clin. Invest.* **114**, 280–290.

Nakashima, M., Sonoda, K., Watanabe, T. (1999) Inhibition of cell growth and induction of apoptotic cell death by the human tumor-associated antigen RCAS1, *Nat. Med.* **5**, 938–942.

Oswald, I.P., Gazzinelli, R.T., Sher, A., James, S.L. (1992) IL-10 synergizes with IL-4 and transforming growth-factor-beta to inhibit macrophage cytotoxic activity, *J. Immunol.* **148**, 3578–3582.

Park, J.A., Wang, E., Kurt, R.A., Schluter, S.F., Hersh, E.M., Akporiaye, E.T. (1997) Expression of an antisense transforming growth factor-beta 1 transgene reduces tumorigenicity of EMT6 mammary tumor cells, *Cancer Gene. Ther.* **4**, 42–50.

Pellegrini, P., Berghella, A.M., Del Beato, T., Maccarone, D., Cencioni, S., Adorno, D., Casciani, C.U. (1997) The sCEA molecule suppressive role in NK and TH1 cell functions in colorectal cancer, *Cancer Biother. Radiopharm.* **12**, 257–264.

Pierce, D.F. Jr., Gorska, A.E., Chytil, A., Meise, K.S., Page, D.L., Coffey, R.J. Jr., Moses, H.L. (1995) Mammary-tumor suppression by transforming growth-factor-beta-1 transgene expression, *Proc. Natl. Acad. Sci. U. S. A.* **92**, 4254–4258.

Pinson, D.M., Leclaire, R.D., Lorsbach, R.B., Parmely, M.J., Russell, S.W. (1992) regulation by transforming growth factor-beta-1 of expression and function of the receptor

for IFN-gamma on mouse macrophages, *J. Immunol.* **149**, 2028–2034.

Pitti, R.M., Marsters, S.A., Lawrence, D.A., Roy, M., Kischkel, F.C., Dowd, P., Huang, A., Donahue, C.J., Sherwood, S.W., Baldwin, D.T., Godowski, P.J., Wood, W.I., Gurney, A.L., Hillan, K.J., Cohen, R.L., Goddard, A.D., Botstein, D., Ashkenazi, A. (1998) Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer, *Nature* **396**, 699–703.

Porter, C.W., Sufrin, J.R. (1986) Interference with polyamine biosynthesis and/or function by analogs of polyamines or methionine as a potential anticancer chemotherapeutic strategy, *Anticancer Res.* **6**, 525–542.

Rook, A.H., Kehrl, J.H., Wakefield, L.M., Roberts, A.B., Sporn, M.B., Burlington, D.B., Lane, H.C., Fauci, A.S. (1986) Effects of transforming growth-factor-beta on the functions of natural-killer-cells – depressed cytolytic activity and blunting of interferon responsiveness, *J. Immunol.* **136**, 3916–3920.

Salazar-Onfray, F., Petersson, M., Franksson, L., Matsuda, M., Blankenstein, T., Karre, K., Kiessling, R. (1995) IL-10 converts mouse lymphoma-cells to a CTL-resistant, NK-sensitive phenotype with low but peptide-inducible MHC class-I expression, *J. Immunol.* **154**, 6291–6298.

Sato, K., Kawasaki, H., Nagayama, H., Enomoto, M., Morimoto, C., Tadokoro, K., Juji, T., Takahashi, T.A. (2000) TGF-beta 1 reciprocally controls chemotaxis of human peripheral blood monocyte-derived dendritic cells via chemokine receptors, *J. Immunol.* **164**, 2285–2295.

Schrey, M.P., Patel, K.V. (1995) Prostaglandin E(2) production and metabolism in human breast-cancer cells and breast fibroblasts – regulation by inflammatory mediators, *Br. J. Cancer* **72**, 1412–1419.

Seo, N., Hayakawa, S., Takigawa, M., Tokura, Y. (2001) Interleukin-10 expressed at early tumour sites induces subsequent generation of CD4(+) T-regulatory cells and systemic collapse of antitumour immunity, *Immunology* **103**, 449–457.

Seo, N., Tokura, Y., Takigawa, M., Egawa, K. (1999) Depletion of IL-10-and TGF-beta-Producing regulatory gamma delta T cells by administering a daunomycin-conjugated specific monoclonal antibody in early tumor lesions augments the activity of CTLs and NK cells, *J. Immunol.* **163**, 242–249.

Smyth, M.J., Strobl, S.L., Young, H.A., Ortaldo, J.R., Ochoa, A.C. (1991) Regulation of lymphokine-activated killer activity and pore-forming protein gene-expression in human peripheral-blood Cd8+ lymphocytes-T – inhibition by transforming growth-factor-beta, *J. Immunol.* **146**, 3289–3297.

Sombroek, C.C., Stam, A.G., Masterson, A.J., Lougheed, S.M., Schakel, M.J., Meijer, C.J., Pinedo, H.M., van den Eertwegh, A.J., Scheper, R.J., de Gruijl, T.D. (2002) Prostanoids play a major role in the primary tumor-induced inhibition of dendritic cell differentiation, *J. Immunol.* **168**, 4333–4343.

Specht, C., Bexten, S., Kolsch, E., Pauels, H.G. (2001) Prostaglandins, but not tumor-derived IL-10, shut down concomitant tumor-specific CTL responses during murine plasmacytoma progression, *Int. J. Cancer* **91**, 705–712.

Stander, M., Naumann, U., Dumitrescu, L., Heneka, M., Loschmann, P., Gulbins, E., Dichgans, J., Weller, M. (1998) Decorin gene transfer-mediated suppression of TGF-beta synthesis abrogates experimental malignant glioma growth in vivo, *Gene Ther.* **5**, 1187–1194.

Steinbrink, K., Jonuleit, H., Muller, G., Schuler, G., Knop, J., Enk, A.H. (1999) Interleukin-10-treated human dendritic cells induce a melanoma-antigen-specific anergy in CD8(+) T cells resulting in a failure to lyse tumor cells, *Blood* **93**, 1634–1642.

Stolina, M., Sharma, S., Lin, Y., Dohadwala, M., Gardner, B., Luo, J., Zhu, L., Kronenberg, M., Miller, P.W., Portanova, J., Lee, J.C., Dubinett, S.M. (2000) Specific inhibition of cyclooxygenase 2 restores antitumor reactivity by altering the balance of IL-10 and IL-12 synthesis, *J. Immunol.* **164**, 361–370.

Strassmann, G., Cole, M.D., Newman, W. (1988) Regulation of colony-stimulating factor 1-dependent macrophage precursor proliferation by type-beta transforming growth-factor, *J. Immunol.* **140**, 2645–2651.

Strassmann, G., Patilkoota, V., Finkelman, F., Fong, M., Kambayashi, T. (1994) Evidence for the involvement of interleukin-10 in the differential deactivation of murine peritoneal-macrophages by prostaglandin-E(2), *J. Exp. Med.* **180**, 2365–2370.

Sumitani, K., Kamijo, R., Toyoshima, T., Nakanishi, Y., Takizawa, K., Hatori, M., Nagumo, M.

(2001) Specific inhibition of cyclooxygenase-2 results in inhibition of proliferation of oral cancer cell lines via suppression of prostaglandin E-2 production, *J. Oral. Pathol. Med.* **30**, 41–47.

Sunkara, P.S., Rosenberger, A.L. (1987) Antimetastatic activity of DL-alpha-difluoromethylornithine, an inhibitor of polyamine biosynthesis, in mice, *Cancer Res.* **47**, 933–935.

Sunkara, P.S., Chang, C.C., Prakash, N.J., Lachmann, P.J. (1985) Effect of inhibition of polyamine biosynthesis by DL-alpha-difluoromethylornithine on the growth and melanogenesis of B-16 melanoma in vitro and in vivo, *Cancer Res.* **45**, 4067–4070.

Uyttenhove, C., Pilotte, L., Theate, I., Stroobant, V., Colau, D., Parmentier, N., Boon, T., Van den Eynde, B.J. (2003) Evidence for a tumoral immune resistance mechanism based on tryptophan degradation by indoleamine 2,3-dioxygenase, *Nat. Med.* **9**, 1269–1274.

Weller, M., Constam, D.B., Malipiero, U., Fontana, A. (1994) Transforming growth factor-beta(2) induces apoptosis of murine T-Cell clones without down-regulating bcl-2 messenger-RNA expression, *Eur. J. Immunol.* **24**, 1293–1300.

Wojtowicz-Praga, S., Verma, U.N., Wakefield, L., Esteban, J.M., Hartmann, D., Mazumder, A., Verma, U.M. (1996) Modulation of B16 melanoma growth and metastasis by anti-transforming growth factor beta antibody and interleukin-2, *J. Immunother. Emphasis Tumor. Immunol.* **19**, 169–175.

Won, J., Kim, H., Park, E.J., Hong, Y., Kim, S.J., Yun, Y. (1999) Tumorigenicity of mouse thymoma is suppressed by soluble type II transforming growth factor beta receptor therapy, *Cancer Res.* **59**, 1273–1277.

Yang, L., Yamagata, N., Yadav, R., Brandon, S., Courtney, R.L., Morrow, J.D., Shyr, Y., Boothby, M., Joyce, S., Carbone, D.P., Breyer, R.M. (2003) Cancer-associated immunodeficiency and dendritic cell abnormalities mediated by the prostaglandin EP2 receptor, *J. Clin. Invest.* **111**, 727–735.

Young, M.R., Knies, S. (1984) Prostaglandin-e production by Lewis lung-carcinoma – mechanism for tumor establishment in vivo, *J. Natl. Cancer Inst.* **72**, 919–922.

Zhang, X.H., Giangreco, L., Broome, H.E., Dargan, C.M., Swain, S.L. (1995) Control of CD4 effector fate – transforming growth-factor-beta-1 and interleukin-2 synergize to prevent apoptosis and promote effector expansion, *J. Exp. Med.* **182**, 699–709.

Zheng, L.M., Ojcius, D.M., Garaud, F., Roth, C., Maxwell, E., Li, Z., Rong, H., Chen, J., Wang, X.Y., Catino, J.J., King, I. (1996) Interleukin-10 inhibits tumor metastasis through an NK cell-dependent mechanism, *J. Exp. Med.* **184**, 579–584.

Immunotoxins and Recombinant Immunotoxins in Cancer Therapy

Yoram Reiter and Avital Lev
Technion-Israel Institute of Technology, Haifa, Israel

1 Introduction 499

2 First- and Second-generation Immunotoxins 502

3 The Development of Recombinant DNA-based Immunotoxins: Design of Recombinant Immunotoxins 503
3.1 The Toxin Moiety 503
3.1.1 Plant Toxins 503
3.1.2 Bacterial Toxins: DT and DT Derivatives 504
3.1.3 Bacterial Toxins: PE and PE Derivatives 505
3.2 The Targeting Moiety – Recombinant Antibody Fragments 506

4 Construction and Production of Recombinant Immunotoxins 507

5 Preclinical Development of Recombinant Immunotoxins 510

6 Application of Recombinant Immunotoxins 511
6.1 Recombinant Immunotoxins against Solid Tumors 511
6.2 Recombinant Immunotoxins against Leukemias and Lymphomas 512

7 Isolation of New and Improved Antibody Fragments as Targeting Moieties: Display Technologies for the Improvement of Immunotoxin Activity 514

8 Improving The Therapeutic Window of Recombinant Immunotoxins: The Balance of Toxicity, Immunogenicity, and Efficacy 517
8.1 Immune Responses and Dose-limiting Toxicity 518
8.2 Specificity Dictated by the Targeting Moiety 519

9 Conclusions and Perspectives 520

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

Bibliography 520
Books and Reviews 520
Primary Literature 520

Keywords

Cancer Therapy
Strategies to kill and eliminate transformed malignant cells using various approaches such as chemotherapy, immunotherapy, gene therapy, and others.

Immunotoxins
A form of targeted therapy approach in which a potent toxin is conjugated or genetically fused to an anitbody or other targeting moiety to specifically target the toxin to the malignant cell.

Recombinant Antibodies
Antibody molecules or their fragments generated and produced by recombinant DNA technology.

Targeted Therapy
Therapeutic strategy in which a cytotoxic drug or an effector function is delivered specifically to the malignant transformed cell by using a specific marker related to the cancerous phenotype.

Targeted cancer therapy in general and immunotherapy in particular, combines rational drug design with the progress in understanding cancer biology. This approach takes advantage of our recent knowledge of the mechanisms by which normal cells are transformed into cancer cells, thus, using the special properties of cancer cells to device novel therapeutic strategies.

Recombinant immunotoxins are excellent examples for such processes, combining the knowledge of antigen expression by cancer cells with the enormous developments in recombinant DNA technology and antibody engineering. Recombinant immunotoxins are composed of a very potent protein toxin fused to a targeting moiety such as recombinant antibody fragment or growth factor.

These molecules bind to surface antigens specific for cancer cells and kill the target cells by catalytic inhibition of protein synthesis. Recombinant immunotoxins are developed for solid tumors and hematological malignancies and have been characterized intensively for their biological activity *in vitro* on cultured tumor cell lines as well as *in vivo* in animal models of human tumor xenografts. The excellent *in vitro* and *in vivo* activities of recombinant immunotoxins have lead to their preclinical development and to the initiation of clinical trial protocols. Recent trial results have demonstrated potent clinical efficacy in patients with malignant diseases that are refractory to traditional modalities of cancer treatment; surgery, radiation therapy, and chemotherapy.

1 Introduction

The rapid progress in understanding the molecular biology of cancer cells has made a large impact on the design and development of novel therapeutic strategies. These have been developed because treatment of cancer by chemotherapy is limited by a number of factors, and usually fails in patients whose malignant cells are not sufficiently different from normal cells in their growth and metabolism. Other limiting factors are the low therapeutic index of most chemotherapeutic agents, the emergence of drug-resistant populations, tumor heterogeneity, and the presence of metastatic disease.

The concept of targeted cancer therapy is thus an important means of improving the therapeutic potential of anticancer agents and lead to the development of novel approaches such as immunotherapy. The approach of cancer immunotherapy and targeted cancer therapy combines rational drug design with the current advances in our understanding of cancer biology. This approach takes advantage of some special properties of cancer cells – many of them contain mutant or overexpressed oncogenes on their surface and these proteins are attractive antigens for targeted therapy. The first cell surface receptor to be linked to cancer was the epidermal growth factor (EGF) receptor present in lung, brain, kidney, bladder, breast, and ovarian cancer. Several other members of the EGF receptor family, that is, *erb*B2, *erb*B3, and *erb*B4 receptors, appear to be abundant on breast and ovary tumors, and *erb*B2, for example, is the target for phase I and II immunotherapy clinical trials.

Other promising candidates for targeted therapy are differentiation antigens that are expressed on the surface of mature cells, but not on the immature stem cells. The most widely studied examples of differentiation antigens currently being used for targeted therapy are expressed by hematopoietic malignancies, and include CD19, CD20, and CD22 on B-cell lymphomas and leukemias, and the interleukin (IL-)-2 receptor on T-cell leukemias. Differentiation antigens have also been found on ovarian, breast, and prostate cancer.

Another class of antigen, termed *tumor-associated antigens* (TAAs), is made up of molecules that are tightly bound to the surface of cancer cells and are associated with the transformed cancer cells. An example is the carbohydrate antigen Lewis Y (LeY) that is found in many types of solid tumors. Another class of TAAs includes cancer peptides that are presented by class I MHC molecules on the surface of tumor cells.

It should be possible to use these molecular cell surface markers as targets to eliminate the cancer cells while sparing the normal cells. For this approach to be successful, we must generate a targeting moiety that will bind very specifically to the antigen or receptor expressed on the cancer cell surface and arm this targeting moiety with an effector cytotoxic moiety. The targeting moiety can be a specific antibody directed toward the cancer antigen or a ligand for a specific overexpressed receptor. The cytotoxic arm can be a radioisotope, a cytotoxic drug, or a toxin. One strategy to achieve this is to arm antibodies that target cancer cells with powerful toxins that can originate from both plants and bacteria. The molecules generated are termed *recombinant immunotoxins*.

The goal of immunotoxin therapy is to target a very potent cytotoxic agent to cell surface molecules, which will then

internalize the cytotoxic agent, resulting in cell death. Developing this type of therapy has gained much interest in recent years. Since immunotoxins differ greatly from chemotherapy in their mode of action and toxicity profile, it is hoped that immunotoxins will have the potential to improve the systemic treatment of tumors incurable with existing modes of therapy.

As shown in Fig. 1, immunotoxins can be divided into two groups: chemical conjugates (or first-generation immunotoxins) and second-generation (or recombinant) immunotoxins. They both contain toxins that have their cell binding domains either mutated or deleted to prevent them from binding to normal cells and are either fused or chemically conjugated to a ligand or an antibody specific for cancer cells (Table 1).

This article will summarize our current understanding of the design and application of second-generation recombinant Fv–immunotoxins, which utilize recombinant antibody fragments as the targeting moiety, in the treatment of cancer, and will also discuss briefly the use of recombinant

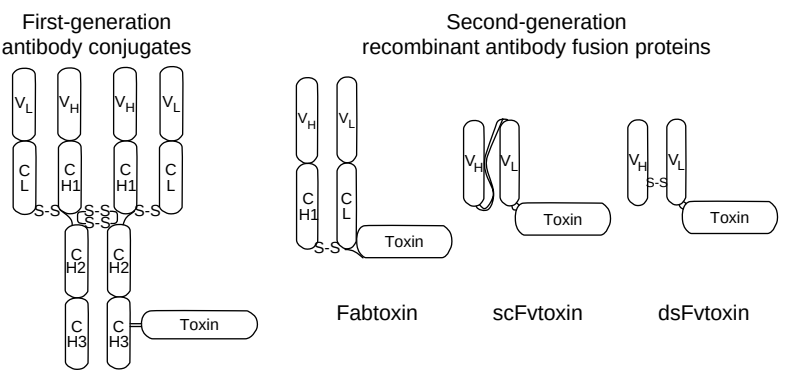

	IgGtoxin	Fabtoxin	scFvtoxin	dsFvtoxin
MW:	200 kDA	90 kDA	66 kDA	64 kDA
Composition:	Heterogeneous	----------Homogeneous ----------		
IC_{50} in culture:	3 $ngmL^{-1}$		1.5 $ngmL^{-1}$	1.5 $ngmL^{-1}$
Dose for CR:	0.75 mg kg^{-1}		0.063 $mgkg^{-1}$	0.075 $mgkg^{-1}$
T½ in circulation:	8 h		20 min	23 min
Tumor penetration:	Fair		Good	Good

Fig. 1 Immunotoxins for targeted cancer therapy. First-generation immunotoxins are whole mAbs to which the toxin is chemically conjugated. Second-generation immunotoxins are made by recombinant DNA technology by fusing recombinant antibody fragments to the toxin (usually a truncated or mutated form of the toxin). Three types of recombinant antibody fragments are used as the targeting moiety in recombinant immunotoxins. Fabs are composed of the light-chain and the heavy-chain Fd fragment (V_H and C_H1), connected to each other via the interchain disulfide bond between C_L and C_H1. scFv fragments are stabilized by a peptide linker, which connects the C-terminus of V_H or V_L with the N-terminus of the other domain. The V_H and V_L heterodimer in dsFv is stabilized by engineering a disulfide bond between the two domains. The biochemical and biological properties described in the lower part of the figure are depicted for B3–lysPE38 (LMB-1) (a first-generation antibody–PE chemical conjugate), B3(Fv)–PE38 (LMB-7) (second-generation recombinant scFv–immunotoxin for an scFv–immunotoxin) and B3(dsFv)–PE38 (LMB-9) (for a second-generation recombinant dsFv–immunotoxin).

Tab. 1 Examples of recombinant immunotoxins against cancer.

Immunotoxin	*Antigen*	*Toxin*	*Cancer*	*Clinical Trial*
Anti-CD7-dgA	CD7	Ricin	Non-Hodgkin's lymphoma	Phase I
DAB_{389}-IL2	IL-2R	DT	T-cell lymphoma Hodgkin's disease	Phase III
Anti-Tac (Fv)-PE38 (LMB-2)	CD25	PE	B and T lymphoma, leukemias	Phase I
DT-Anti-Tac(Fv)	CD25	DT	Leukemias, lymphoma	–
RFB4(dsFv)-PE38	CD22	PE	B leukemias	Phase I
Di-dgA-RFB4	CD22	Ricin	Leukemias, non-Hodgkin's lymphoma	–
B3-lysPE38 (LMB-1)	Lewis Y	PE	Carcinomas	Phase I
B3(Fv)-PE38 (LMB-7)	Lewis Y	PE	Carcinomas	Phase I
B3(dsFv)-PE38 (LMB-9)	Lewis Y	PE	Carcinoma	Phase I
BR96(sFv)-PE40	Lewis Y	PE	Carcinoma	–
e23(Fv)-PE38	*erb*B2/HER2	PE	Breast cancer	Phase I
FRP5(scFv)ETA	*erb*B2/HER2	PE	Breast cancer	–
Tf-CRM107	Transferrin-R	DT	Glioma	Phase I
HB21(Fv)-PE40	Transferrin-R	PE	Various	–
MR1(Fv)-PE38	Mutant EGF-R	PE	Liver, brain tumors	–
SS1(Fv)-PE38	Mesothelin	PE	Ovarian cancer	–

antibody fragments for other modes of cancer therapy and diagnosis.

2 First- and Second-generation Immunotoxins

First-generation immunotoxins were made in the early 1970s and were composed of cancer-specific monoclonal antibodies (mAbs) to which native bacterial or plant toxins were chemically conjugated. The understanding of toxin structure–function properties and the advancement in recombinant DNA technology and antibody engineering led to important breakthroughs in the late 1980s to construct second-generation recombinant immunotoxins that are composed of recombinant antibody fragments derived from cancer-specific antibodies or phage-display libraries and truncated forms of toxins. These molecules are produced in large amounts, needed for preclinical and clinical studies, in bacteria and feature better clinical properties.

As shown in Fig. 1, first- and second-generation immunotoxins contain toxins that have their cell binding domains either mutated or deleted to prevent them from binding to normal cells, and are either chemically conjugated or fused to a ligand or an antibody specific for cancer cells.

First-generation immunotoxins, composed of whole antibodies chemically conjugated to toxins, demonstrated the feasibility of this concept. Cancer cells cultured *in vitro* could be killed under conditions in which the immunotoxin demonstrated low toxicity toward cultured normal cells. Clinical trials with these agents had some success; however, they also revealed several problems, such as nonspecific toxicity toward some normal cells, difficulties in production and, particularly for the treatment of solid tumors, poor tumor penetration due to their large size.

Second-generation immunotoxins have overcome many of these problems. Progress in the elucidation of the toxin's structure and function combined with the techniques of protein engineering facilitated the design and construction of recombinant molecules with a higher specificity for cancer cells and reduced toxicity to normal cells. At the same time, advances in recombinant DNA technology and antibody engineering enabled the generation of small antibody fragments. Thus, it was possible to decrease the size of immunotoxins significantly and to improve their tumor-penetration potential *in vivo*. The development of advanced methods of recombinant protein production enabled the large-scale production of recombinant immunotoxins of high purity and quality for clinical use in sufficient quantities to perform clinical trials.

Another strategy to target cancer cells is to construct chimeric toxins in which the engineered truncated portion of the toxin [*Pseudomonas* exotoxin (PE) or diphtheria toxin (*DT*)] gene is fused to cDNA encoding growth factors or cytokines. These include transforming growth factor (TGF)-β, insulin-like growth factor (IGF)-1, acidic and basic fibroblast growth factor (FGF), IL-2, IL-4, and IL-6. These recombinant toxins (oncotoxins) are designed to target specific tumor cells that overexpress these receptors (Fig. 1).

In the following sections, we will summarize the rationale and current knowledge on the design and application of second-generation recombinant Fv–immunotoxins that utilize recombinant antibody fragments as the targeting moiety. Recent results of clinical trials

are summarized. We will also discuss the powerful new technologies for selecting new antibodies with unique specificities and improved properties.

3 The Development of Recombinant DNA-based Immunotoxins: Design of Recombinant Immunotoxins

3.1 The Toxin Moiety

The toxins that are most commonly used to make immunotoxins are ricin, DT, and PE. These toxins belong to a group of polypeptide enzymes that catalytically inactivate protein synthesis leading to cell death. Some of these toxins have been shown to induce apoptosis.

The genes for these toxins have been cloned and expressed in *Escherichia coli*, and the crystal structures of all three proteins have been solved. This information, in combination with mutational studies, has elucidated which toxin subunits are involved in their biological activity and, most importantly, the different steps of the cytocidal process. DT, PE, ricin, and their derivatives have all been successfully used to prepare immunotoxin conjugates, but only PE- and DT-containing fusion proteins generate active recombinant immunotoxins. This is because the toxic moiety must be separated from the binding moiety after internalization. PE and DT fusion proteins generate their free toxic moieties by proteolytic processing. Ricin does not possess such a proteolytic processing site and therefore cannot be attached to the targeting moiety with a peptide bond without losing cytotoxic activity. Recently, proteolytic processing sites were introduced into ricin by recombinant DNA techniques to try to overcome this problem.

3.1.1 Plant Toxins

Ricin, or the ricin A-chain fragment, has been a commonly used toxin for conjugation to antibodies. Ricin is synthesized as single polypeptide chains and processed posttranslationally into two subunits A and B linked through a disulfide bond. Ricin is a 65-kDa glycoprotein purified from the seeds of the castor bean (*Ricinus communis*). It is composed of an A subunit, which kills cells by catalytically inactivating ribosomes. The A subunit is linked by a disulfide bond to a B subunit, which is responsible for cell binding. The B chain is a galactose-specific lectin that binds to galactose residues present on cell surface glycoproteins and glycolipids. Once the B subunit of ricin binds to the cell membrane, the protein enters the cell through coated pits and endocytic vesicles. The A and B subunits of ricin are separated by a process involving disulfide bond reduction. The A subunit of ricin translocates across an intracellular membrane to the cell cytosol, probably with the assistance of the B subunit. In the cytosol, it arrests protein synthesis by enzymatically inactivating the 28S subunit of eukaryotic ribosomes. Because native ricin is highly toxic and lacks specificity, several modified forms of ricin have been developed to prepare immunotoxins that are better tolerated by patients.

To decrease the nonspecific binding of whole ricin, the A chain alone has been coupled to antibodies. The A chain is obtained by reducing the disulfide bond that links it to the B chain. Immunotoxins composed of the ricin A chain coupled to well-internalized antibodies can be highly cytotoxic. In the absence of the B chain (the binding subunit), however, immunotoxins

made with poorly internalized antibodies are not cytotoxic.

Immunotoxins containing the ricin A chain are rapidly cleared from the circulation by the liver by the binding of mannose and fucose residues of the A chain to receptors present on the reticuloendothelial system and hepatocytes. To circumvent this problem, these carbohydrate residues were chemically modified, resulting in a deglycosylated A chain (dgA) molecule. In preclinical studies, dgA-containing immunotoxins were found to have longer half-lives in the circulation and better antitumor efficacy *in vivo*. Also, recombinant A chain produced in *E. coli* can be used in place of dgA because it is devoid of carbohydrate and is not rapidly cleared by the liver.

Another strategy to decrease the nonspecific toxicity of native ricin is to block the galactose binding sites of the B chain by cross-linking with glycopeptide or to use short cross-linkers to connect the antibody to the toxin so that the galactose binding site is sterically blocked by the antibody. Blocked ricin retains a low affinity for galactose binding sites, which enhances internalization and cytotoxicity of an antibody that binds to a poorly internalized antigen.

Other plant toxins commonly used for clinical immunotoxin construction are saporin and pokeweed antiviral protein, which are single polypeptide chains that inactivate ribosomes in a similar fashion to ricin. Because these toxins lack the binding chain (B chain), they are relatively nontoxic to cells and are used for immunotoxin production.

3.1.2 Bacterial Toxins: DT and DT Derivatives

DT is a 58-kDa protein, secreted by pathogenic *Corynebacterium diphtheria*, which contains a lysogenic β phage. DT ADP-ribosylates eukaryotic elongation factor (EF)-2 at a "diphthamide" residue located at His415, using NAD^+ as a cofactor. This modification arrests protein synthesis and subsequently leads to cell death. Only a few, and perhaps only one, DT molecules need to reach the cytosol in order to kill a cell. When DT is isolated from the culture medium of *C. diphtheria*, it is composed of an N-terminal 21-kDa A subunit and a C-terminal 37-kDa B subunit held together by a disulfide bond. DT is the expression product of a single gene, which when secreted into the medium is processed into two fragments by extracellular proteases. When DT is produced as a recombinant single-chain protein in *E. coli*, it is not cleaved by the bacteria, but is instead cleaved by a protease in the target cells. The A domain of DT contains its enzymatic activity. The N-terminus of the B subunit of DT (or the region between A and B in single-chain DT) mediates translocation of the A subunit into the cytoplasm. The B domain, especially its C-terminus, is responsible for the binding of DT to target cells. Deletions or mutations in this part of the molecule abolish or greatly diminish the binding and toxicity of DT. DT enters cells via coated pits and is proteolytically cleaved within the endocytic compartment, if it is not already in the two-chain form, and reduced. It also undergoes a conformational change at the acidic pH present in endosomes, which probably assists translocation of the A chain into the cytosol, perhaps via a porelike structure mediated by the B chain. Derivatives of DT that are used to make immunotoxins have the C-terminus altered by mutations or partially deleted (DAB486, DAB389, and DT388), but retain the translocation and ADP-ribosylation activity of DT. Recombinant antibody fusion proteins with such

derivatives target only cells that bind the antibody moiety of the immunotoxin.

3.1.3 Bacterial Toxins: PE and PE Derivatives

Two major research studies have enabled the use and genetic manipulation of PE for the design of immunotoxins: (1) the elucidation of the crystal structure of PE, showing the toxin to be composed of three major structural domains, and (2) the finding that these domains are different functional modules of the toxin.

PE is a single-chain 66-kDa molecule secreted by *Pseudomonas aeruginosa* that, like DT, irreversibly ADP-ribosylates the diphthamide residue of EF-2, using NAD^+ as cofactor. As a consequence, protein synthesis is inhibited and cell death ensues. PE is composed of three major domains. Different functions have been assigned to each domain by mutational analysis. The N-terminal domain Ia mediates binding to the α_2-macroglobulin receptor. Domain Ib is a small domain that lies between domains II and III, and has no known function. Domain II mediates translocation of domain III, the C-terminal ADP-ribosylating domain, into the cytosol of target cells (Fig. 2). Translocation occurs after internalization of the toxin and after a variety of other steps including a pH-induced conformational change, proteolytic cleavage at a specific site in domain II, and a reductive step that separates the amino and carboxyl fragments. Ultimately, the C-terminal portion of PE is translocated from the endoplasmic reticulum into the cytosol. Despite a similar mode of action of PE and DT, which is ADP-ribosylation, and a similar initial pathway of cell entry (internalization via coated pits and endocytic vesicles) and of processing (proteolytic cleavage and a reductive step), they share almost no sequence homology.

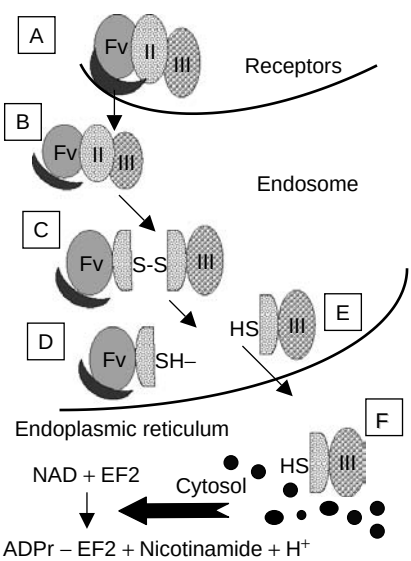

Fig. 2 The biological activity of PE A. The Fv portion of the immunotoxin targets domains II and III of PE to a cell surface receptor or other target molecule on the tumor cell (A). The immunotoxin enters the cell by internalization and is transferred into the endosome (B). Within the endosome, the molecule unfolds owing to a fall in pH. The conformational change exposes a proteolytic site, and a proteolytic cleavage occurs in the translocation domain between amino acids 279 and 280 (C). A disulfide bond is then broken, thus creating two fragments: the Fv moiety and a small part of domain II, and the rest of domain II connected to domain III (D). The C-terminal fragment containing the ADP-ribosylation domain (domain III) and most of the translocation domain (domain II) is carried into the endoplasmic reticulum (E), and translocation occurs from the endoplasmic reticulum into the cytosol (F). The enzymatically active domain ADP-ribosylates EF-2 at a diphthamide residue located at His415, using NAD^+ as a cofactor. This modification arrests protein synthesis and subsequently leads to cell death by apoptosis. In DT, the proteolytic processing occurs between residues 193 and 194. The catalytic A chain (amino acids 1–193) then translocates to the cytosol through the endosome with the help of translocation domain residues 326–347 and 358–376, which form an ion channel.

The only similarity is the spatial arrangement of key residues in their active sites that are arranged around residue Glu553 in PE and Glu145 in DT.

When the whole toxin is used to make an immunotoxin, nonspecific toxicity occurs mainly because of binding of the toxin portion to cells, mediated by the binding domain. Consequently, the goal of making improved derivatives of PE-based immunotoxins has been to inactivate or remove the binding domain. Molecules in which the binding domain has been retained but inactivated by mutations were made; however, a better alternative to inactivating the cell binding domain by mutations is to remove it from PE. The prototype molecule with this sort of deletion is PE40 (amino acids 253–613, MW 40 kDa]. Because PE40 and its derivatives described below lack the binding domain (amino acids 1–252), they have very low nonspecific toxicity, but make very active and specific immunotoxins when fused to recombinant antibodies. Currently, almost all PE-derived recombinant immunotoxins are constructed with PE38 (MW 38 kDa), a PE40 derivative that has, in addition to the deletion of domain Ia, a second deletion encompassing a portion of domain Ib (amino acids 365–379). Another useful mutation is to change the C-terminal sequence of PE from REDLK to KDEL. This improves the cytotoxicity of PE and its derivatives, presumably by increasing their delivery to the endoplasmic reticulum where translocation takes place.

3.2
The Targeting Moiety – Recombinant Antibody Fragments

The antibody moiety of the recombinant immunotoxin is responsible for specifically directing the immunotoxin to the target cell; thus, the usefulness of the immunotoxin depends on the specificity of the antibody or antibody fragment that is connected to the toxin. Consequently, for the construction of recombinant immunotoxins, the only antibodies that should be used are those that recognize antigens that are expressed on target cancer cells and are not present on normal cells, present at very low levels, or are only present on less-essential cells (Table 1). Receptors for growth factors like EGF, IL-2, IL-4, IL-6, or *erb*B2 are common targets for targeted cancer therapy because they are highly expressed on many cancer cells. Other carcinoma-related antigens include developmental antigens such as complex carbohydrates, which are often highly abundant on the surface of cancer cells.

The use of antibodies for immunotoxin production also requires that the antibody–antigen complex be internalized, because the mechanism of PE-toxin killing requires endocytosis as a first step in the entry of the toxin into the cell.

Recombinant immunotoxins contain antibody fragments as the targeting moiety. These fragments can be produced in *E. coli*, and are the result of intensive research and development in recombinant antibody technologies. Several antibody fragments have been used to construct recombinant immunotoxins (Fig. 1). One type contains Fab fragments in which the light chain and the heavy chain Fd fragment (V_H and C_H1) are connected to each other via an interchain disulfide bond between C_L and C_H1. The toxin moiety can be fused to the carboxyl end of either C_L or C_H1. Fabs can be produced in *E. coli*, either by secretion, with coexpression of light chains and Fd fragments or by expression of the chains in intracellular inclusion bodies in separate cultures; in the latter case, they are reconstituted by a

refolding reaction using a redox-shuffling buffer system. Several immunotoxins with Fab fragments have been constructed and produced in this way.

The smallest functional modules of antibodies required for antigen binding are Fv fragments. This makes them especially useful for clinical applications, not only for generating recombinant immunotoxins but also for tumor imaging, because their small size improves tumor penetration. Fv fragments are heterodimers of the variable heavy chain (V_H) and the variable light chain (V_L) domains. Unlike whole IgG or Fab, in which the heterodimers are held together and stabilized by interchain disulfide bonds, the V_H and V_L of Fvs are not covalently connected and are consequently unstable; this instability can be overcome by making recombinant Fvs that have the V_H and V_L covalently connected by a peptide linker that fuses the C-terminus of the V_L or V_H to the N-terminus of the other domain (Fig. 1). These molecules are termed *single-chain Fvs* (scFvs), and many retain the specificity and affinity of the original antibody. The cloning, construction, and composition of recombinant Fv fragments of antibodies and Fv–immunotoxins are described in Fig. 3.

Many recombinant immunotoxins have been constructed using scFvs, in which molecules, the *scFv* gene is fused to PE38 to generate a potent cytotoxic agent with targeted specificity (Figs. 1 and 3).

Until recently, the construction of scFvs was the only general method available to make stable Fvs. However, many scFvs are unstable or have reduced affinity for the antigen compared with the parent antibody or Fab fragment. This is because the linker interferes with binding or because the linker does not sufficiently stabilize the Fv structure, leading to aggregation and loss of activity. This is particularly true at physiological temperatures (37 °C). To overcome these problems, an alternative strategy has been developed that involves generating stable Fvs by connecting the V_H and V_L domains by an interchain disulfide bond engineered between structurally conserved framework residues of the Fv; these molecules are termed *disulfide-stabilized Fvs* (dsFvs). The positions at which the cysteine residues were to be placed were identified by computer-based molecular modeling; as they are located in the framework of each V_H and V_L, this location can be used as a general method to stabilize almost all Fvs without the need for any structural information. Many dsFvs have been constructed in the past three years (mainly as dsFv–immunotoxins, in which the dsFv is fused to PE38) and they show several advantages over scFvs. In addition to their increased stability (due to a decreased tendency to aggregate), they are often produced in higher yields than scFvs; in several cases, the binding affinity of the dsFv was significantly improved over that of the scFv.

4 Construction and Production of Recombinant Immunotoxins

In the recombinant immunotoxins derived from PE, the recombinant antibody fragments are fused to the N-terminus of the truncated derivative of PE (with the cell binding domain deleted, for example, PE40 or PE38). This restores the original domain arrangement of PE, which consists of an N-terminal binding domain followed by the translocation domain and the C-terminal ADP-ribosylation domain.

Only fusions of an antigen binding domain (Fv) to the N-terminus of truncated PE are active; C-terminal fusions are

not active because the bulky antigen binding domain blocks translocation of the C-terminal fragment into the cytoplasm.

DT immunotoxins are fusions of mutated DT with antigen binding regions of a recombinant antibody. However, in this case the antigen binding domain must be fused to the C-terminus of DT. This corresponds to the inverse arrangement of the functional modules of PE and DT (see Fig. 3). DT immunotoxins are active only when the enzymatically active N-terminal domain is free to translocate into the cytosol.

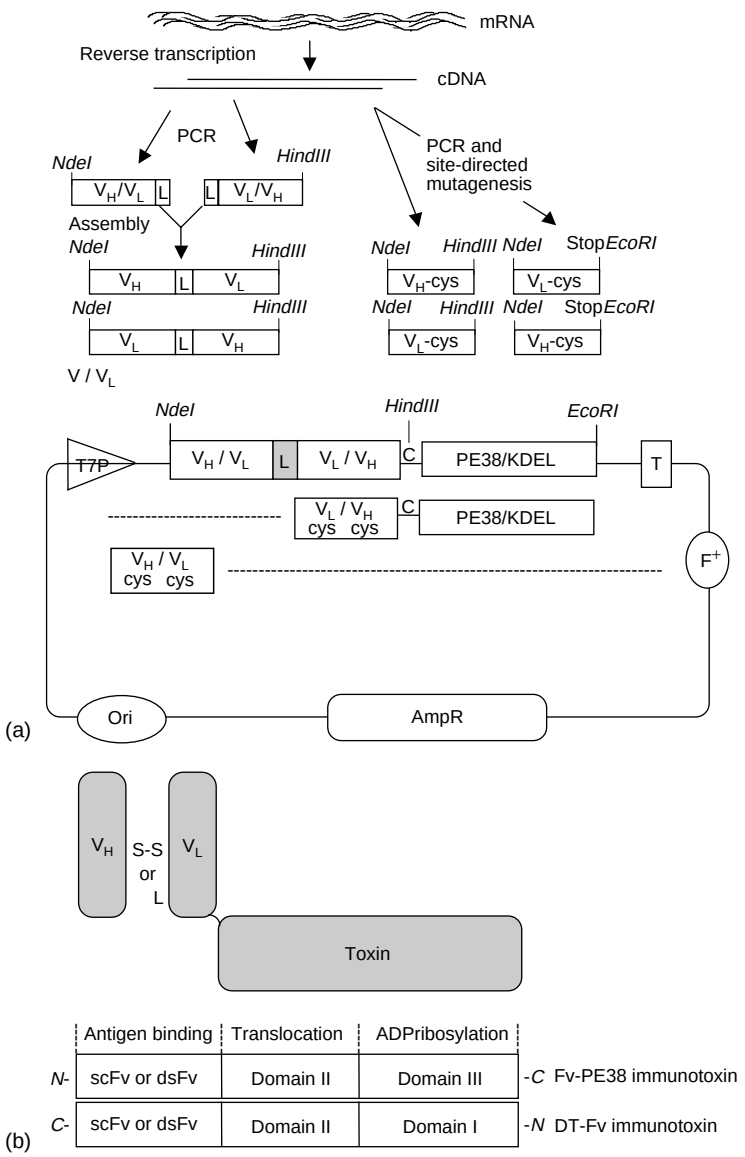

The expression vectors used for DT immunotoxins are very similar to those used with PE with the exception that the DNA fragments encoding the binding moiety are ligated to the 3′-end of the DT coding region. The cloning of the antibody variable regions is performed using cloning techniques that are now well established (Fig. 3). The plasmid vector for the expression of scFv–immunotoxins or the components of dsFv–immunotoxins is a high copy–number plasmid derived from vectors made and described by Studier and Moffatt. These contain the T7 promoter, translation initiation signals, and a transcription terminator, as well as an F+ phage replication origin to generate single-stranded DNA to be used for site-directed mutagenesis.

When these plasmids are transformed into *E. coli* BL21/DE3 (which contain the *T7 RNA* polymerase gene under the control of the *lac*UV5 promoter) they generate large amounts of recombinant protein upon IPTG induction. The recombinant scFv–immunotoxin or the components of the dsFv–immunotoxin accumulate in insoluble intracellular inclusion bodies. [dsFv–immunotoxins require two cultures, one expressing the V_H and one expressing the V_L; the toxin moiety (PE38) can be fused to either the V_H or the V_L.] The inclusion bodies are then isolated, purified, solubilized, reduced, and subsequently used in a refolding reaction that is controlled for oxidation (redox shuffling). In the case of dsFv–immunotoxins, solubilized inclusion bodies of V_H and V_L

Fig. 3 Cloning, construction, and composition of scFv–and dsFv–immunotoxins. (a) Cloning and construction of recombinant scFv–and dsFv–immunotoxins. The genes encoding the V_H and V_L variable domains are cloned usually from hybridoma mRNA by reverse transcription, cDNA synthesis, and subsequent PCR amplification using degenerate primers that are complementary to the 5′ or 3′ end of the V_H and V_L genes, or by primers that are designed according to the N-terminal amino acid sequence of the mAb to be cloned and conserved sequences at the N-terminal of the heavy and light constant regions. The variable genes can be also cloned by constant domains primers and using the RACE rapid amplification of cDNA ends (method). Restriction sites for assembling the peptide linker sequence that connects the V_H and V_L domains and for cloning into the expression vector are also introduced by PCR. Construction of dsFv involves the generation of two expression plasmids that encode the two components of the dsFv V_H-Cys and V_L-Cys. The cysteines are introduced in position 44 in FR2 of V_H and position 100 of FR4 of V_L or position 105 of FR4 in V_H and position 43 of FR2 in V_L (numbering system of Kabat et al.) by site-directed mutagenesis using as a template a uracil-containing single-stranded DNA of the scFv construct from the F+ origin present in the expression plasmid and cotransfection with M13 helper phage. In addition to the cysteines, cloning sites, ATG translation initiation codons and stop codons are introduced at the 5′ and 3′ ends and of the V_H and V_L genes as shown by site-directed mutagenesis or PCR. The antibody variable genes are subcloned into an expression vector, which contains the gene for a truncated form of PE. This expression vector is controlled by the T7 promoter and upon induction of the T7 RNA polymerase, which is under the control of the *lac*UV5 promoter, in *E. coli* BL21λDE3 by IPTG, large amounts of recombinant protein are produced. (b) Composition of recombinant immunotoxins. In PE-derived recombinant Fv–immunotoxins, the Fv region of the targeting antibody is fused to the N-terminus of a truncated form of PE, which contains the translocation domain (domain II) and enzymatically active ADP-ribosylation domain (domain III]. The cell binding domain of whole PE (domain I) is replaced by the Fv targeting moiety thus preserving the relative position of the binding domain function to the other functional domains of PE. In the dsFv–immunotoxins, there are two components. In one, the V_H or V_L domains are fused to the N-terminus of the truncated PE and, in the other, the variable domain is covalently linked by the engineered disulfide bond. DT-derived immunotoxins are fused to the C-terminus due to the inverse arrangement of the functional modules of PE and DT.

(with the toxin fused to either) are mixed in a 1:1 molar ratio into the refolding solution. The formation of the interchain disulfide bond between the V_H and V_L domains is promoted by inducing oxidation using excess oxidized glutathione or by refolding at high pH. The immunotoxins are then purified from the refolding mixtures by ion-exchange and size-exclusion chromatography. Approximately 20 mg of clinical-grade active immunotoxin can be obtained from 1 L of a fermentor culture induced with IPTG.

5 Preclinical Development of Recombinant Immunotoxins

A wide variety of recombinant immunotoxins have been made and tested against cancer target cells. If found to be active and are considered to be tested in clinical trials, they undergo several years of preclinical development to determine their efficacy and toxicity in several *in vitro* and *in vivo* experimental models (Table 2).

The initial phase is the characterization of the biological activity of the immunotoxin on cultured tumor cells. These assays include measurement of cell free enzymatic activity, that is, ADP-ribosylation activity in the case of bacterial toxins, and the binding affinity of the immunotoxin to the target antigen, which can be determined on purified antigen, on cells by binding displacement assays or by surface plasmon resonance assays. Cytotoxicity assays are performed on antigen-bearing cells and measure either inhibition of protein synthesis, proliferation, colony counts or cell viability.

Cytotoxicity assays on malignant, single-cell suspensions directly obtained from patients are a very useful test, if available, since such cells contain the physiological numbers of receptor or target density, which, in many cases is lower than

Tab. 2 Functional properties *in vitro* and *in vivo* of PE-based recombinant Fv-immunotoxins.

Immunotoxin	*Specificity*	*Activity in vitro (IC_{50}) [ng mL^{-1}]*	*Binding affinity (Kd) [nM]*	*Antitumor activity in vivo (xenograft model)*
Anti-Tac(Fv)-PE38 (LMB-2)	CD25	0.15	1.4	Complete regressions/ cures (ATAC4)
B3(Fv)-PE38 (LMB-7)	Lewis Y	1.5	1300	Complete regressions/ cures (A431)
B3(dsFv)-PE38 (LMB-9)	Lewis Y	1.5	24 000	Complete regressions/ cures (A431)
e23(Fv)-PE38 (erb-38)	*erb*B2/HER2	0.3	40	Partial regressions (A431)
RFB4(dsFv)-PE38 (BL22)	CD22	10	10	Partial and some complete regressions (CA46)
SS1(Fv)-PE38	Mesothelin	0.5	11	Complete regressions/ cures (A431-K5)
MR1(Fv)-PE38	Mutant EGF-R	3.0	11	Partial regressions (glioblastoma)
55.1(Fv)-PE38	Mucin carbohydrate	0.3	80	Complete regressions (Colo205)

established cell lines. The stability of recombinant immunotoxins *in vitro* in various physiological buffers or human serum is also an important test to predict their stability *in vivo*. *In vivo* efficacy of recombinant immunotoxins is usually demonstrated in immunodeficient mice bearing xenografts of human tumor cells. The tumor xenografts can be established as subcutaneous solid tumors, orthotopic implants, or disseminated leukemia.

Initial toxicity and pharmacokinetics studies have also been performed in mice; however, many target antigens are present at some level on some normal tissues, and thus toxicology and pharmacokinetics studies should be tested in an animal that has normal cells capable of binding the target antigen. For most immunotoxins, this requires studies in monkeys to test for targeted damage to normal tissues to predict whether such damage will occur in humans.

6 Application of Recombinant Immunotoxins

6.1 Recombinant Immunotoxins against Solid Tumors

The treatment of solid tumors with immunotoxins is challenging because of their physiological nature of tight junctions between tumor cells, high interstitial pressure within tumors and heterogeneous blood supply, and also antigen expression. The greatest need for new therapies is in the treatment of metastatic epithelial cancers, and immunotoxins can be a useful addition to the standard procedures of surgery, radiation, and chemotherapy.

As already described, the use of recombinant fragments of antibodies for making recombinant immunotoxins is especially useful for the treatment of solid tumors because their small size improves tumor penetration. Over recent years, several recombinant immunotoxins that target solid tumors have been developed (Table 2); targets include breast, lung, gastric, bladder, and central nervous system cancers. They are at different stages of clinical development and some are already employed in clinical trials.

mAb B3 is an antibody that reacts with the LeY antigen present on cancers of the colon, breast, stomach, lung, and bladder. Early trials with a first-generation immunotoxin (LMB-1) in which an antibody to LeY (mAb B3) was used to make a chemical conjugate with PE38 showed significant clinical activity, with responses in colon and breast cancer. The one complete response and one partial response observed in this trial were the first major responses to immunotoxins documented for metastatic breast and colon cancer respectively.

The B3 antibody was then used to make a single-chain immunotoxin termed *B3(Fv)–PE38 or LMB-7*. LMB-7 has shown good activity against human cancer xenografts growing in mice and it is also able to cure carcinomatous meningitis in rats when given by the intrathecal route. A phase I clinical trial with LMB-7 began in 1995 and is nearing completion. During the trial, it became evident that LMB-7 lost activity when incubated at 37 °C because of aggregation, which greatly limited its ability to penetrate solid tumors.

B3(dsFv)–PE38 (LMB-9) is the dsFv version of LMB-7 with stability improved over that of LMB-7. This improved stability also allowed it to be used in a continuous-infusion mode in mice bearing human tumor xenografts; this route of administration showed an improved therapeutic

window over a bolus injection. Clinical trials with LMB-9 started in the middle of 1998. A different recombinant single-chain immunotoxin, BR96(scFv)–PE40 was derived from the anti-LeY mAb BR96 and is also currently undergoing clinical testing.

Monoclonal antibody e23 is directed at *erb*B2 (Her2/*neu*), which is highly expressed in many breast, lung, ovarian, and stomach cancers. e23(dsFv)–PE38 is a dsFv–immunotoxin composed of the Fv portion of the e23 antibody and PE38. This dsFv–immunotoxin has a significantly improved binding affinity and stability compared with its scFv analog, e23(Fv)–PE38. FRP5scFv–ETA is also a recombinant immunotoxin targeting *erb*B2. Clinical trials with e23(Fv)–PE38 were initiated in early 1998. In a phase I study on breast cancer patients, hepatotoxicity was observed in all patients. Immunohistochemistry showed the presence of *erb*B2 on hepatocytes, explaining the liver toxicity of the immunotoxin. This study demonstrated that targeting of tumors with antibodies to *erb*B2 armed with toxic agents or radioisotopes may result in unexpected organ toxicity due to the expression of the target antigen on normal cells.

Other recombinant immunotoxins that have been constructed and have antitumor activities *in vitro* and in mouse models *in vivo* include: B1(Fv)–PE38, also directed against the LeY antigen; 55.1(Fv)–PE38 and 55.1(dsFv)–PE38, which are directed at a carbohydrate mucin antigen overexpressed in colon cancers; MR1(Fv)–PE38, constructed by antibody phage-display technology and directed to a mutant EGF receptor overexpressed in liver and brain tumors; and SS(Fv)–PE38, a new recombinant immunotoxin specific for mesothelin, a differentiation antigen present on the surface of ovarian cancers, mesotheliomas, and several other types of human cancers. SS(Fv)–PE38 was constructed from an Fv fragment that was isolated by antibody-phage display from mice that underwent DNA immunization with a plasmid expressing the cloned antigen. This approach to antibody formation eliminates the need for the production of proteins for immunization.

Immunotoxins were also used to target tumors of the central nervous system. Since the transferrin receptor is expressed on tumor and normal hepatic cells but not in normal brain, several trials have targeted antitransferrin receptor immunotoxins to brain tumors. These include a conjugate of mAb 454A12 with a recombinant form of ricin-A (plant toxin), a conjugate of human transferrin with a mutant form of DT and chimeric toxin of recombinant IL-4–PE38 fusion.

6.2 Recombinant Immunotoxins against Leukemias and Lymphomas

Conventional immunotoxins, in which IgGs or Fabs are coupled to toxins, have also been used to target leukemias and lymphomas. This approach should be quite effective because many of the tumor cells are in the blood and bone marrow, where they are readily accessible to the drug. Moreover, fresh cells from patients may be easily tested for immunotoxin binding and cytotoxic activity.

Immunotoxins have also been developed for indirect treatment of malignancies by their killing of T cells that mediate graft-versus-host disease (GvHD) in the setting of allogeneic transplantation. Clinical trials using ricin-based immunoconjugates for treatment of leukemias have shown some promising results, but dose escalation has been limited by the side effects of the toxin. In addition, it is important

to eliminate not only easily accessible tumor cells but also malignant cells that are less accessible. Therefore, even for leukemias, there is a need to develop small recombinant immunotoxins that will reach cells outside the circulation. Recombinant immunotoxins targeted at leukemia and lymphoma antigens have been made with antibody fragments specific for the subunit of the IL-2 receptor (CD25) and for CD22. In addition, growth factor fusion proteins have been made that target the IL-2, IL-4, IL-6, and granulocyte macrophage colony stimulating factor (GM-CSF) receptors.

The most potent immunotoxin produced against leukemia cells is anti-Tac(Fv)–PE38 (LMB-2); this targets CD25, which is overexpressed on many T-cell leukemias. LMB-2 is very active against leukemia cell lines *in vitro* and has very good activity in animal models. It also selectively kills cells *in vitro* obtained from patients with adult T-cell leukemia without harming hematopoietic stem cells. Phase I clinical trials with LMB-2 are showing promising results. The immunotoxin was administered to 35 patients for a total of 59 treatment cycles. One hairy cell leukemia (HCL) patient achieved a complete remission, which is ongoing at 20 months. Seven partial responses were observed in cutaneous T-cell lymphoma, HCL, chronic lymphocytic leukemia, Hodgkin's disease and adult T-cell leukemia. Responding patients had a 2 to 5 log reduction of circulating malignant cells, improvement in skin lesions, and regression of lymphomatous mass and splenomegally. All four patients with HCL responded to the treatment (one with complete responses and three had 98–99.8% reductions in malignant circulating cells). A phase II trial is planned in patients with $CD25^+$ hematologic malignancies and phase I trials are planned for the prevention of GvHD in patients undergoing high-risk allotransplantation.

The conventional immunotoxin RFT5–SMPT–dgA has also been developed to target CD25 and has resulted in several responses in Hodgkin's disease, one of which lasted over two years. It is already undergoing testing for the prevention of GvHD in patients undergoing allotransplantation and has recently been shown *ex vivo* to remove alloreactive donor T cells while preserving antileukemia and antiviral T-cell responses.

A new agent, RFB4(dsFv)–PE38 (BL22), is a dsFv–immunotoxin directed at the CD22 differentiation antigen present on most B-cell leukemias. It has high cytotoxic activity on cultured tumor cells as well as in animal models and preclinical tests have been completed. This recombinant immunotoxin recently entered clinical trials in patients with leukemias. Initial phase I trials in 16 chemotherapy-resistant HCL patients resulted in 11 complete responses and two partial responses, including two partial responses in patients ineligible for LMB-2 because of $CD25^-$ HCL cells. Responses to BL22 were associated with at least a 99.5% reduction in circulating HCL cells. BL22 also induced responses in chronic lymphocytic leukemia. These recent results demonstrate that recombinant Fv–immunotoxins containing truncated PE are particularly effective in patients with chemotherapy-refractory HCL and other hematological malignancies. Other targets for the development of B-cell leukemia–specific recombinant immunotoxins include the CD19 and CD20 differentiation antigens in B-cell tumors, and CD30 in Hodgkin's lymphoma.

The B-cell lymphoma markers CD22 and CD19 were also targeted using conventional first-generation immunotoxins with

dgA – IgG–RFB4–dgA (targeting CD22) and IgG–HD37–dgA (targeting CD19). Leukemias and lymphomas were also targeted with recombinant fusions of IL-2 with truncated DT.

7 Isolation of New and Improved Antibody Fragments as Targeting Moieties: Display Technologies for the Improvement of Immunotoxin Activity

The generation of mAbs made by immunizing animals and allowing *in vivo* processes, such as immune tolerance and somatic hypermutation, to shape the antigen-combining site is a key issue for the generation of specific antibodies. The unique features required from these molecules were already described in Sect. 1 of this article. These antibodies created *in vivo* can be used for many research and diagnostic applications. Mouse mAbs might be made less immunogenic and more effective for human therapy by reformatting the binding site into chimeric or complementary-determining region (CDR)-grafted antibodies. The advances in recombinant DNA technology and antibody engineering have also led to the ability to manipulate the size of the antigen binding domain as described in this article. The two variable domains of the binding site can be cloned and arranged into a large array of possible molecular formats and sizes, and expressed in a variety of hosts, ranging from bacteria, lower eukaryotes such as yeast and fungi, to the higher eukaryotes, including mammalian cells, transgenic animals and plants. Extraordinary progress in engineering and selecting small antibody fragments for immunotherapeutic approaches has been made over the past decade, when molecular display technologies have been developed that allow us to create very large repertoires of mouse or fully human antibodies that are displayed on filamentous phage or other molecular display systems. These technologies are now revolutionizing the way in which we can build high-affinity binding sites from scratch, from any species (including humans) and use them for clinical applications such as the targeting of a drug or toxin to cancer cells as in recombinant Fv–immunotoxins.

The concept of molecular display technology relays on the physical linkage between the genotype (the antibody variable region genes) and the phenotype (antigen-binding capability) to allow simultaneous selection of the genes that encode a protein with the desired binding function. This concept can be viewed as an *in vitro* mimicking system for the natural antibody response function of the immune system. This concept was first applied by George Smith in 1985 to small peptides. The display of functional antibody repertoires on phages required several additional discoveries. First, a procedure for accessing large collections of antibody variable domains was needed; this was first described in 1989, when partially degenerate oligonucleotides priming to the 5′ and 3′ end of variable region genes and the polymerase chain reaction (PCR) were used to amplify hybridoma or large collections of variable genes. Second, as whole antibodies cannot yet be functionally expressed in bacteria, a crucial discovery was that antibody fragments (Fab or scFv) were functionally expressed in *E. coli* when they were secreted into the periplasm of the bacteria, which simulated the naturally oxidizing environment of the endoplasmic reticulum. By providing restriction sites in

the oligonucleotides used for PCR amplification, antibody libraries could thus be cloned for expression in *E. coli*. Initially, such antibody libraries were expressed from phage-λ vectors; a plaque-screening assay with labeled antigen was then used to identify antigen-specific binding sites. Such time-consuming procedures were rapidly replaced by the third seminal development: the provision of a link between the phenotype and the genotype using phages. In 1990, McCafferty et al. showed that antibody fragments could be displayed on the surface of filamentous phage particles by fusion of the antibody variable genes to one of the phage coat proteins. Multiple rounds of affinity selection could subsequently enrich antigen-specific phage antibodies, because the phage particle carries the gene encoding the displayed antibody. This was originally reported for scFv fragments, and later for Fab fragments and other antibody derivatives such as diabodies, as well as extended to various display systems. With these advances in place, it became possible to make phage antibody libraries by PCR cloning of large collections of variable region genes expressing each of the binding sites on the surface of a different phage particle and harvesting the antigen-specific binding sites by *in vitro* selection of the phage mixture on a chosen antigen. In the early 1990s, Clackson et al. showed for the first time that phage-display technology could be used to select antigen-specific antibodies from libraries made from the spleen B cells of immunized mice, thereby bypassing the requirement to immortalize the antigen-specific B cells, as in the hybridoma technology. Similarly, libraries were made from human B cells taken from animals or individuals immunized with antigen, exposed to infectious agents, with autoimmune diseases or with cancer. Thus, phage-display technology in the early 1990s had already shown the potential to replace hybridoma technology by rescuing *V* genes from immune B cells. Further advances were reported in the mid-1990s that would bypass the use of immunization and animals altogether. First, it was shown that antibodies against many different antigens could be selected from nonimmune libraries, made from the naive light-chain and heavy-chain *IgM V*-gene pools of B cells of a non-immunized, healthy individual. Second, libraries of synthetic antibody genes, with variable genes not harvested from immune sources but consisting of germline segments artificially provided with diversity by oligonucleotide cloning, were shown to behave in a similar way to naive antibody libraries. It thus became possible to use primary antibody libraries, with huge collections of binding sites with different specificities, to select *in vitro* binding sites against most antigens, including nonimmunogenic molecules, toxic substances, and targets conserved between species.

Since these key discoveries, there have been numerous reports on applications of phage antibody libraries, ranging from basic research to drug development. In addition, many novel, related molecular display methods for antibodies have been described, including display systems on ribosomes, bacteria, and yeast cells. These technologies follow similar concepts for *in vitro* selection and improvement of binding sites. Novel selection strategies of phage-display libraries and other molecular display systems are being developed for the identification of novel antigen-binding fragments. These include selection for binding using purified or nonpurified antigen, selection for function, selection based on display capability and phage infectivity, subtractive selection procedures, and also using high-throughput selection and

screening. The use of phage-display systems will revolutionize the field of targeted drug therapy in general and the recombinant immunotoxin field in particular, because advances in this field are dependent not only on the identification of new targets on cancer cells but also on the development of new and very specific targeting moieties such as antibody fragments (scFvs). Phage-display technology now enables one to select such molecules against unique targets, especially when hybridoma technology fails to produce antibodies against an antigen or when nonimmunogenic or conserved targets between species are being used. Alternatives to phage display for making fully human antibodies are technologies developed using transgenic mice (xenomice). These transgenic mice have been engineered to lack the native murine immune repertoire and instead harbor most of the human immune system *V* genes in the germline. Injection of these "humanized" animals with a foreign antigen or hapten effectively evokes an immune response and a humanlike antibody is produced in the B cells. The antibody genes can be recovered from B cells either by PCR and library selection or by fusion into a monoclonal cell line by classic hybridoma technology. Several examples of recombinant Fv–immunotoxins that were constructed from scFvs isolated by phage display have already been reported and are being considered for use in clinical trials.

The phage-display approach has been used to isolate an scFv that binds with high affinity to a mutant form of the EGF receptor in which a deletion of a portion of the extracellular domain of the receptor generates a tumor-specific antigen. Another novel target for cancer therapy could be cancer-specific peptides presented on human leukocyte antigen (HLA) molecules on the surface of tumor cells. To accomplish this, it will be necessary to isolate antibodies that recognize tumor-specific peptides associated with class I MHC molecules on tumor cells. As a first step in this direction, a recombinant immunotoxin has been constructed using an antibody that was isolated by phage display and that binds specifically to peptide/MHC complexes found on virally infected cells. This recombinant immunotoxin was cytotoxic only to cells specifically expressing hemagglutinin peptide HA255–262 in complex with H-2K^k (mouse class I MHC) and was not cytotoxic to cells that express other peptides associated with H-2K^k or to cells not expressing H-2K^k. These studies indicate that, if antibodies that recognize tumor-specific peptides in the context of class I MHC molecules can be developed, they should be very useful agents for targeted cancer immunotherapy. Recently, the isolation of a human antibody directed against a peptide encoded by the melanoma-associated antigen MAGE-A1 presented by HLA-A1 molecules was reported. A large phage-Fab antibody repertoire was selected on a recombinant version of the complex. One of the selected phage antibodies shows binding to HLA-A1 complexed with the MAGE-A1 peptide, but does not show binding to HLA-A1 complexed with a peptide encoded by gene *MAGE-A3* and differing from the MAGE-A1 peptide by only three residues. Phages carrying this recombinant antibody bind to HLA-A1$^+$ cells only after *in vitro* loading with MAGE-A1 peptide. It remains now to see if such human anti-MHC/peptide complexes may prove useful for monitoring the cell surface expression of these complexes and, eventually, as a targeting reagent for the specific killing of tumor cells expressing tumor peptide/MHC complexes. The isolation of such rare antibodies against unique tumor

targets is a proof of the powerful abilities of antibody phage-display technology for the development of new generations of targeting molecules for cancer therapy and diagnosis.

Phage-display technology can be used not only to create new scFv antibodies but also to improve the properties of existing scFvs. Improvements in antibody stability, expression, and binding affinity can be achieved by using a combination of strategies including random and directed mutagenesis of CDR regions, DNA shuffling, and error-prone PCR. These mutagenesis strategies combined with the powerful selection methods available to *screen* antibody phage-display libraries can yield scFv molecules with significantly improved properties for clinical applications. For example, phage display was used to improve antibody affinity by mimicking somatic hypermutation *in vitro*. *In vivo* affinity maturation of antibodies involves mutation of hot spots in the DNA encoding the variable regions. This information was used to develop a strategy to improve antibody affinity *in vitro* using phage-display technology. The anti-mesothelin scFv, SS(scFv), was used to identify DNA sequences in the variable regions that are naturally prone to hypermutations. In a few selected hot spot regions encoding nonconserved amino acids, random mutations were introduced to make libraries with a size requirement between 10^3 and 10^4 independent clones. Panning of the hot spot libraries yielded several mutants with a 15- to 55-fold increase in affinity compared with a single clone with a 4-fold increased affinity from a library in which mutagenesis was done outside the hot spots (Table 2). This is an example of a powerful phage-display-based strategy that should be generally applicable for the rapid isolation of higher-affinity mutants of Fvs, Fabs, and other recombinant antibodies from antibody-phage libraries that are smaller in size.

In another example, random CDR mutagenesis to obtain mutants of MR1(Fv)–PE38, a single-chain recombinant immunotoxin that targets a mutant form of the EGF receptor, EGFRvIII, that is frequently overexpressed in malignant glioblastomas, was performed (Table 2). Initially, nine residues of heavy chain CDR3 were randomly mutagenized and several mutants with increased binding affinity were isolated. All mutations were in regions that correspond to a DNA hot spot. The mutant MR1Fvs with an increased affinity for EGFRvIII had an increased activity when converted to recombinant immunotoxins. A specific region of the variable region of the antibody light chain CDR3 that corresponded to a hot spot was mutagenized, and a mutant antibody with an additional increase in affinity and cytotoxic activity was isolated. These studies further show that targeting hot spots in the CDRs of Fvs is an effective approach for obtaining Fvs with increased affinity.

8 Improving The Therapeutic Window of Recombinant Immunotoxins: The Balance of Toxicity, Immunogenicity, and Efficacy

Although some of the problems, including design, large-scale production and stability, associated with the initial recombinant immunotoxins have been solved, other fundamental problems need to be addressed that are relevant to much of the immunotherapy field. Specificity, toxicity, and immunogenicity are major factors that will determine the usefulness and success of recombinant immunotoxins.

8.1 Immune Responses and Dose-limiting Toxicity

As with any cytotoxic agent, side effects such as nonspecific toxicity and immunogenicity can occur when multiple injections of immunotoxins are given. One class of side effects is due to inappropriate targeting of the immunotoxin to normal cells because of the poor specificity of the antibody. In addition, the toxin or the Fv portion of the antibody can bind nonspecifically to various tissues. For example, in mice, which usually do not contain target antigens, liver damage occurs when large amounts of immunotoxins are given. Molecular modeling combined with site-directed mutagenesis may help in the design of new versions of the toxin with decreased toxicity caused by nonspecific binding.

The development of neutralizing antibodies usually occurs after 10 days and limits the therapeutic application of immunotoxins to this 10-day period. Recent data from clinical trials indicate that patients with solid tumors develop antibodies much more readily then those with hematologic tumors. It is speculated that some hematologic tumors may be associated with less immunogenicity than others. For example, none of 14 patients with chronic lymphocytic leukemia treated with LMB-2 or BL22 have shown any evidence of antibodies.

Several approaches have been taken to reduce immunogenicity. One is to make small molecules, which appear to be less immunogenic; another is to use immunosuppressive agents such as deoxyspergualin or CTLA-4–Ig, an inhibitor of the costimulation pathways required for T-cell help and activation through the CD28/CTLA-4–CD80/CD86 complex. Another approach is to use the anti-CD20 mAb, Rituximab, which induces B-cell depletion in the majority of patients and is itself nonimmunogenic.

The dose-limiting toxicity of many immunotoxins is vascular leak syndrome (VLS). Recent studies indicate that recombinant toxins, including those containing mutated forms of PE, produce VLS in rats and that inflammation, which can be suppressed by steroids or nonsteroidal anti-inflammatory agents, mediates the VLS. VLS can also be mediated indirectly by the activation of endothelial cells and/or macrophages via cytokines such as tumor necrosis factor-α and interferon-γ. The activated cells produce nitric oxide (NO), which then can mediate oxidative damage to the endothelial cells and result in increased permeability.

Some studies demonstrate direct endothelial cell damage caused by binding the toxin to the cells. The direct damage to the cell is mediated by the enzymatic activity of the toxin, while others show an indirect damage that is mediated by binding of the targeting moiety. For example, experiments with human umbilical vein endothelial cells exposed to LMB-1 (antibody conjugate with truncated PE38) indicated that the mAb B3 rather than PE38 was binding to the LeY antigen on endothelial cells. Recent experiments using an *in vivo* model composed of human neonatal foreskin xenografts in severe combined immuno-deficient (SCID) mice identified a 3-amino acid motif present in protein toxins and in IL-2 that causes VLS without other toxin activity. Thus, VLS can be blocked in future trials with anti-inflammatory agents to block cytokine action, or by mutations or peptide inhibitors that will prevent the binding of the toxin or the targeting moiety to endothelial cells.

Toxicity can also be reduced by modifications in the scFv-targeting moiety. For example, reduction of the nonspecific animal toxicity of recombinant Fv–immunotoxin anti-Tac(Fv)–PE38 (which targets the IL-2 receptor) was achieved by introducing mutations in the framework regions of the Fv, which lower the isoelectric point (p*I*). The dose escalation with this recombinant Fv–immunotoxin (that has produced eight responses, including a durable clinically complete remission in a recently completed phase I trial of leukemias and lymphomas) was limited by liver toxicity. It was noted that the Fv of anti-Tac has a p*I* of 10.2, which brought about the hypothesis that the overall positive charge on the Fv portion of anti-Tac(Fv)–PE38 contributes to nonspecific binding to liver cells and results in dose-limiting liver toxicity. A mouse model was used to investigate the basis of this toxicity, and it was found that lowering the p*I* of the Fv of anti-Tac from 10.2 to 6.82 by selective mutation of surface residues causes a 3-fold decrease in animal toxicity and hepatic necrosis. This change in p*I* did not significantly alter the CD25 binding affinity, the cytotoxic activity toward target cells, or antitumor activity, resulting in a 3-fold improvement in the therapeutic index. If this decreased toxicity occurs in humans, it should greatly increase the clinical utility of this immunotoxin.

Another strategy to overcome the problems of nonspecific toxicity and antigenicity is by the chemical modification of the recombinant Fv–immunotoxins. An example of this was also demonstrated recently in which site-specific chemical modification with polyethylene glycol (PEGylation) of anti-Tac(Fv)–PE38 (LMB-2) improved its antitumor activity, and reduced animal toxicity and immunogenicity. PEGylation can increase plasma half-lives, stability, and therapeutic potency. To produce a PEGylated recombinant immunotoxin with improved therapeutic properties, a mutant form of anti-Tac(Fv)–PE38 (LMB-2) in which one cysteine residue was introduced into the peptide connector (ASGCGPE) between the Fv and the toxin was constructed. This mutant LMB-2 (Cys1-LMB-2), which retained full cytotoxic activity, was then site-specifically conjugated with 5 or 20 kDa of polyethylene glycol–maleimide. When compared with unmodified LMB-2, both PEGylated immunotoxins showed similar cytotoxic activities *in vitro*, but superior stability at 37 °C in mouse serum, a 5- to 8-fold increase in plasma half-lives in mice and a 3- to 4-fold increase in antitumor activity. This was accompanied by a substantial decrease in animal toxicity and immunogenicity. Site-specific PEGylation of recombinant immunotoxins may thus increase their therapeutic potency in humans.

8.2 Specificity Dictated by the Targeting Moiety

Specificity of the recombinant immunotoxin is determined by the distribution of the target antigens; several target antigens are relatively cancer specific, but are present on some normal cells in small amounts. For example, *erb*B2, although overexpressed on tumor cells, is also expressed on a limited number of normal cells. This reactivity with normal cells may cause side effects during immunotoxin therapy. It was discovered during a clinical trial that small amounts of the LeY antigen are expressed on the surface of endothelial cells and damage to these cells caused VLS. To overcome such problems, new specific targets and new reagents

against the cancer antigens that will recognize only the tumor-associated molecules must be identified and developed. The construction of large phage-display antibody libraries may result in the isolation and characterization of new reagents with improved specificity and affinity for cancer-targeted therapy.

9 Conclusions and Perspectives

Over the past decade, several second-generation recombinant immunotoxins with improved properties have been developed and are currently being evaluated in clinical trials. Several of these show clinical activity and promising results in phase I trials. The outcome of these clinical trials demonstrates that the promising preclinical results with these new agents can be translated into more substantial clinical responses, and that similar agents that target other cancer antigens merit further clinical development. These accumulating results suggest that Fv–immunotoxins merit further development as a new modality for targeted cancer treatment.

See also Antigen Presenting Cells (APCs); Dendritic Cells; Immunology; Innate Immunity; Oncology, Molecular.

Bibliography

Books and Reviews

Kreitman, R.J. (1999) Immunotoxins in cancer therapy, *Curr. Opin. Immunol.* **11**, 570–8.

Kreitam, R.J., Pastan, I. (1998) Immunotoxins for targeted cancer therapy, *Adv. Drug Deliv. Rev.* **31**, 53–88.

Pastan, I., Chaudhary, V., FitzGerald, D.J. (1992) Recombinant toxins as novel therapeutic agents, *Annu. Rev. Biochem.* **61**, 331–54.

Vitteta, E.S. (1994) From the basic science of B cells to biological missiles at the bedside, *J. Immunol.* **153**, 1407–20.

Primary Literature

Allured, V.S., Collier, R.J., Carroll, S.F., McKay, D.B. (1986) Structure of exotoxin A of *Pseudomonas aeruginosa* at 3.0 Angstrom resolution, *Proc. Natl. Acad. Sci. U. S. A.* **83**, 1320–4.

Amlot, P.L., Stone, M.L., Cunningham, D. et al. (1993) A phase I study of an anti-CD22–deglycosylated ricin A chain immunotoxin in the treatment of B-cell lymphomas resistant to conventional therapy, *Blood* **82**, 2624–33.

Andersen, P.S., Stryhn, A., Hansen, B.E., Fugger, L., Engberg, J., Buus, S.A. (1996) Recombinant antibody with the antigen-specific, major histocompatibility complex-restricted specificity of T cells, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1820–4.

Baluna, R., Vitetta, E.S. (1997) Vascular leak syndrome: a side effect of immunotherapy, *Immunopharmacology* **37**, 117–32.

Baluna, R., Vitetta, E.S. (1999) An *in vivo* model to study immunotoxin-induced vascular leak in human tissue, *J. Immunother.* **22**, 41–47.

Baluna, R., Rizo, J., Gordon, B.E., Ghetie, V., Vitetta, E.S. (1999) Evidence for a structural motif in toxins and interleukin-2 that may be responsible for binding to endothelial cells and initiating vascular leak syndrome, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 3957–62.

Barbas, C.F. III, Bain, J.D., Hoekstra, D.M., Lerner, R.A. (1992) Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem, *Proc. Natl. Acad. Sci. U. S. A.* **89**, 4457–61.

Batra, J.K., Kasprzyk, P.G., Bird, R.E., Pastan, I., King, C.R. (1992) Recombinant anti-erbB2 immunotoxins containing *Pseudomonas* exotoxin, *Proc. Natl. Acad. Sci. U. S. A.* **89**, 5867–71.

Batra, J.K., Jinno, Y., Chaudhary, V.K. et al. (1989) Antitumor activity in mice of an immunotoxin made with anti-transferrin receptor and a recombinant form of

Pseudomonas exotoxin, *Proc. Natl. Acad. Sci. U. S. A.* **86**, 8545–9.

Beers, R., Chowdhury, P., Bigner, D., Pastan, I. (2000) Immunotoxins with increased activity against epidermal growth factor receptor vIII-expressing cells produced by antibody phage display, *Clin. Cancer Res.* **6**, 2835–43.

Benhar, I., Pastan, I. (1995) Characterization of B1(Fv)PE38 and B1(dsFv)PE38: single-chain and disulfide-stabilized Fv–immunotoxins with increased activity that cause complete remissions of established human carcinoma xenografts in nude mice, *Clin. Cancer Res.* **1**, 1023–9.

Benhar, I., Reiter, Y., Pai, L.H., Pastan, I. (1995) Administration of disulfide-stabilized Fv–immunotoxins B1(dsFv)–PE38 and B3(dsFv)–PE38 by continuous infusion increases their efficacy in curing large tumor xenografts in nude mice, *Int. J. Cancer* **62**, 351–5.

Better, M., Chang, C.P., Robinson, R.R., Horwitz, A.H. (1988) *Escherichia coli* secretion of an active chimeric antibody fragment, *Science* **240**, 1041–3.

Bird, R.E., Hardman, K.D., Jacobson, J.W. et al. (1988) Single-chain antigen-binding proteins, *Science* **242**, 423–6.

Blakey, O.S., Watson, G.J., Knowles, P.P. et al. (1987) Effect of chemical deglycosylation of ricin A-chain on the *in vivo* fate and cytotoxic activity of an immunotoxin composed of ricin A-chain and anti-Thy 1. 1 antibody, *Cancer Res.* **47**, 947–52.

Boder, E.T., Wittrup, K.D. (1997) Yeast surface display for screening combinatorial polypeptide libraries, *Nat. Biotechnol.* **15**, 553–7.

Bolognesi, A., Tazzari, P.L., Olivieri, F. et al. (1996) Induction of apoptosis by ribosome-inactivating proteins and related immunotoxins, *Int. J. Cancer* **68**, 349–55.

Brandhuber, B.J., Allured, V.S., Falbel, T.G., McKay, D.B. (1988) Mapping the enzymatic active site of *Pseudomonas aeruginosa* exotoxin A, *Proteins* **3**, 146–54.

Brinkmann, U., Pastan, I. (1995) Recombinant immunotoxins: from basic research to cancer therapy, *Methods Enzymol.* **8**, 143–56.

Brinkmann, U., Pai, L.H., FitzGerald, D.J., Willingham, M., Pastan, I. (1991) B3(Fv)–PE38KDEL, a single-chain immunotoxin that causes complete regression of a human carcinoma in mice, *Proc. Natl. Acad. Sci. U. S. A.* **88**, 8616–20.

Brinkmann, U., Reiter, Y., Jung, S.H., Lee, B., Pastan, I. (1993) A recombinant immunotoxin containing a disulfide-stabilized Fv fragment, *Proc. Natl. Acad. Sci. U. S. A.* **90**, 7538–42.

Burton, D.R., Barbas, C.F. III, Persson, M.A., Koenig, S., Chanock, R.M., Lerner, R.A. (1991) A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals, *Proc. Natl. Acad. Sci. U. S. A.* **88**, 10134–7.

Cai, X., Garen, A. (1995) Anti-melanoma antibodies from melanoma patients immunized with genetically modified autologous tumor cells: selection of specific antibodies from single-chain Fv fusion phage libraries, *Proc. Natl. Acad. Sci. U. S. A.* **92**, 6537–41.

Carroll, S.F., Collier, R.J. (1987) Active site of *Pseudomonas aeruginosa* exotoxin A. Glutamic acid 553 is photolabeled by NAD and shows functional homology with glutamic acid 148 of diphtheria toxin, *J. Biol. Chem.* **262**, 8707–11.

Carroll, S.F., Collier, R.J. (1988) Amino acid sequence homology between the enzymic domains of diphtheria toxin and *Pseudomonas aeruginosa* exotoxin A, *Mol. Microbiol.* **2**, 293–6.

Chames, P., Hufton, S.E., Coulie, P.G., Uchanska-Ziegler, B., Hoogenboom, H.R. (2000) Direct selection of a human antibody fragment directed against the tumor T-cell epitope HLA-A1–MAGE-A1 from a nonimmunized phage-Fab library, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7969–74.

Chang, C.N., Landolfi, N.F., Queen, C. (1991) Expression of antibody Fab domains on bacteriophage surfaces. Potential use for antibody selection, *J. Immunol.* **147**, 3610–4.

Chang, K., Pastan, I. (1996) Molecular cloning of mesothelin, a differentiation antigen present on mesothelium, mesotheliomas, and ovarian cancers, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 136–40.

Chang, K., Pai, L.H., Batra, J.K., Pastan, I., Willingham, M.C. (1992) Characterization of the antigen (CAK1) recognized by monoclonal antibody K1 present on ovarian cancers and normal mesothelium, *Cancer Res.* **52**, 181–6.

Chaudhary, V.K., FitzGerald, D.J., Pastan, I. (1991) A proper amino terminus of diphtheria toxin is important for cytotoxicity, *Biochem. Biophys. Res. Commun.* **180**, 545–51.

Chaudhary, V.K., FitzGerald, D.J., Adhya, S., Pastan, I. (1987) Activity of a recombinant fusion protein between transforming growth

factor type α and *Pseudomonas* toxin, *Proc. Natl. Acad. Sci. U. S. A.* **84**, 4538–42.

Chaudhary, V.K., Gallo, M.G., FitzGerald, D.J., Pastan, I. (1990) A recombinant single-chain immunotoxin composed of anti-Tac variable regions and a truncated diphtheria toxin, *Proc. Natl. Acad. Sci. U. S. A.* **87**, 9491–4.

Chaudhary, V.K., Jinno, Y., FitzGerald, D., Pastan, I. (1990) *Pseudomonas* exotoxin contains a specific sequence at the carboxyl terminus that is required for cytotoxicity, *Proc. Natl. Acad. Sci. U. S. A.* **87**, 308–12.

Chaudhary, V.K., Queen, C., Junghans, R.P., Waldmann, T.A., FitzGerald, D.J., Pastan, I. (1989) A recombinant immunotoxin consisting of two antibody variable domains fused to *Pseudomonas* exotoxin, *Nature* **339**, 394–7.

Chiang, Y.L., Sheng-Dong, R., Brow, M.A., Larrick, J.W. (1989) Direct cDNA cloning of the rearranged immunoglobulin variable region, *Biotechniques* **7**, 360–6.

Choe, M., Webber, K.O., Pastan, I. (1994) B3(Fab)–PE38M: a recombinant immunotoxin in which a mutant form of *Pseudomonas* exotoxin is fused to the Fab fragment of monoclonal antibody B3, *Cancer Res.* **54**, 3460–7.

Choe, S., Bennett, M.J., Fujii, G. et al. (1992) The crystal structure of diphtheria toxin, *Nature* **357**, 216–22.

Chowdhury, P.S., Pastan, I. (1999) Improving antibody affinity by mimicking somatic hypermutation *in vitro*, *Nat. Biotechnol.* **17**, 568–72.

Chowdhury, P.S., Viner, J.L., Beers, R., Pastan, I. (1998) Isolation of a high-affinity stable single-chain Fv specific for mesothelin from DNA-immunized mice by phage display and construction of a recombinant immunotoxin with anti-tumor activity, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 669–74.

Clackson, T., Hoogenboom, H.R., Griffiths, A.D., Winter, G. (1991) Making antibody fragments using phage display libraries, *Nature* **352**, 624–8.

Conry, R.M., Khazaeli, M.B., Saleh, M.N. et al. (1995) Phase I trial of an anti-CD19 deglycosylated ricin A chain immunotoxin in non-Hodgkin's lymphoma: effect of an intensive schedule of administration, *J. Immunother.* **18**, 231–41.

Crameri, A., Cwirla, S., Stemmer, W.P. (1996) Construction and evolution of antibody-phage libraries by DNA shuffling, *Nat. Med.* **2**, 100–2.

Dall'Acqua, A., Carter, P. (1998) Antibody engineering, *Curr. Opin. Struct. Biol.* **8**, 443–50.

de Haard, H.J., van Neer, N., Reurs, A. et al. (1999) A large non-immunized human Fab fragment phage library that permits rapid isolation and kinetic analysis of high affinity antibodies, *J. Biol. Chem.* **274**, 18218–30.

Debinski, W., Pastan, I. (1995) Recombinant F(ab') C242–*Pseudomonas* exotoxin, but not the whole antibody-based immunotoxin, causes regression of a human colorectal tumor xenograft, *Clin. Cancer Res.* **1**, 1015–22.

Debinski, W., Puri, R.K., Kreitman, R.J., Pastan, I. (1993) A wide range of human cancers express interleukin 4 (IL-4) receptors that can be targeted with chimeric toxin composed of IL-4 and *Pseudomonas* exotoxin, *J. Biol. Chem.* **268**, 14065–70.

Eiklid, K., Olsnes, S., Pihl, A. (1980) Entry of lethal doses of abrin, ricin and modeccin into the cytosol of HeLa cells, *Exp. Cell Res.* **126**, 321–9.

Endo, Y., Tsurigi, K. (1987) RNA *N*-glycosidase activity of ricin A-chain. Mechanism of action of the toxic lectin ricin on eukaryotic ribosomes, *J. BioI. Chem.* **262**, 8128–30.

Engberg, J., Krogsgaard, M., Fugger, L. (1999) Recombinant antibodies with the antigen-specific, MHC restricted specificity of T cells: novel reagents for basic and clinical investigations and immunotherapy, *Immunotechnology* **4**, 273–8.

Engert, A., Marlin, G., Amlot, P., Wijdenes, J., Diehl, V., Thorpe, P. (1991) Immunotoxins constructed with anti-CD25 monoclonal antibodies and deglycosylated ricin A-chain have potent anti-tumor effects against human Hodgkin cells *in vitro* and solid Hodgkin tumors in mice, *Int. J. Cancer* **49**, 450–6.

Engert, A., Oiehl, V., Schnell, R. et al. (1997) A phase-1 study of an anti-CD25 ricin A-chain immunotoxin (RFT5–SMPT–dgA) in patients with refractory Hodgkin's lymphoma, *Blood* **89**, 403–410.

Flavell, O.J. (1998) Saporin immunotoxins, *Curr. Topics Microbiol. Immunol.* **234**, 57–61.

Frankel, A.E., Laver, J.H., Willingham, M.C., Burns, L.J., Kersey, J.H., Vallera, D.A. (1997) Therapy of patients with T-cell lymphomas and leukemias using an anti-CD7 monoclonal antibody–ricin A chain immunotoxin, *Leuk. Lymph.* **26**, 287–98.

Garrard, L.J., Yang, M., O'Connell, M.P., Kelley, R.F., Henner, D.J. (1991) Fab assembly and enrichment in a monovalent phage display system, *Biotechnology* **9**, 1373–7.

Georgiou, G., Stathopoulos, C., Daugherty, P.S., Nayak, A.R., Iverson, B.L., Curtiss, R. III (1997) Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines, *Nat. Biotechnol.* **15**, 29–34.

Ghetie, M-A., Richardson, J., Tucker, T., Jones, D., Uhr, J.W., Vitetta, E.S. (1991) Antitumor activity of Fab' and IgG-anti-CD22 immunotoxins in disseminated human B lymphoma grown in mice with severe combined immunodeficiency disease: effect on tumor cells in extranodal sites, *Cancer Res.* **51**, 5876–80.

Giannini, G., Rappuoli, R., Ratti, G. (1984) The amino-acid sequence of two non-toxic mutants of diphtheria toxin: CRM45 and CRM197, *Nucleic Acids Res.* **12**, 4063–9.

Glockshuber, R., Malia, M., Pfitzinger, I., Pluckthun, A. (1990) A comparison of strategies to stabilize immunoglobulin Fv-fragments, *Biochemistry* **29**, 1362–7.

Gram, H., Marconi, L.A., Barbas, C.F. III, Collet, T.A., Lerner, R.A., Kang, A.S. (1992) In vitro selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library, *Proc. Natl. Acad. Sci. U. S. A.* **89**, 3576–80.

Graus, Y.F., de Baets, M.H., Parren, P.W. et al. (1997) Human anti-nicotinic acetylcholine receptor recombinant Fab fragments isolated from thymus-derived phage display libraries from myasthenia gravis patients reflect predominant specificities in serum and block the action of pathogenic serum antibodies, *J. Immunol.* **158**, 1919–29.

Greenfield, L., Johnson, V.G., Youle, R.J. (1987) Mutations in diphtheria toxin separate binding from entry and amplify immunotoxin selectivity, *Science* **238**, 536–9.

Greenfield, L., Bjorn, M.J., Horn, G. et al. (1983) Nucleotide sequence of the structural gene for diphtheria toxin carried by corynebacteriophage β, *Proc. Natl. Acad. Sci. U. S. A.* **80**, 6853–7.

Grossbard, M.L., Lambert, J.M., Goldmacher, V.S. et al. (1993) Anti-B4-blocked ricine: a phase I trial of 7-day continuous infusion in patients with B-cell neoplasms, *J. Clin. Oncol.* **11**, 726–737.

Hanes, J., Pluckthun, A. (1997) *In vitro* selection and evolution of functional proteins by using ribosome display, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 4937–42.

Hoogenboom, H.R. (1997) Designing and optimizing library selection strategies for generating high-affinity antibodies, *Trends Biotechnol.* **15**, 62–70.

Hoogenboom, H.R., Chames, P. (2000) Natural and designer binding sites made by phage display technology, *Immunol. Today* **21**, 371–8.

Hoogenboom, H.R., Griffiths, A.D., Johnson, K.S., Chiswell, D.J., Hudson, P., Winter, G. (1991) Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains, *Nucleic Acids Res.* **19**, 4133–7.

Hudson, P.J. (1999) Recombinant antibody constructs in cancer therapy, *Curr. Opin. Immunol.* **11**, 548–57.

Hudson, P.J. (2000) Recombinant antibodies: a novel approach to cancer diagnosis and therapy, *Exp. Opin. Invest. Drugs* **9**, 1231–42.

Hung, M.C., Lau, Y.K. (1999) Basic science of HER-2/*neu*: a review, *Semin. Oncol.* **26**(4 Suppl 12), 51–9.

Huse, W.D., Sastry, L., Iverson, S.A. et al. (1989) Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda, *Science* **246**, 1275–81.

Huston, J.S., Levinson, D., Mudgett-Hunter, M. et al. (1988) Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*, *Proc. Natl. Acad. Sci. U. S. A.* **85**, 5879–83.

Hwang, J., FitzGerald, D.J., Adhya, S., Pastan, I. (1987) Functional domains of *Pseudomonas* exotoxin identified by deletion analysis of the gene expressed in *E. coli*, *Cell* **48**, 129–36.

Idziorek, T., FitzGerald, D., Pastan, I. (1990) Low pH-induced changes in *Pseudomonas* exotoxin and its domains: increased binding of Triton X-114, *Infect. Immun.* **58**, 1415–20.

Iglewski, B.H., Kabat, D. (1975) NAD-dependent inhibition of protein synthesis by *Pseudomonas aeruginosa* toxin, *Proc. Natl. Acad. Sci. U. S. A.* **72**, 2284–8.

Irvin, J.O., Uckun, F.M. (1992) Pokeweed antiviral protein: ribosome inactivation and therapeutic applications, *Pharmacol. Ther.* **55**, 279–302.

Jain, R.K. (1996) Delivery of molecular medicine to solid tumors, *Science* **271**, 1079–80.

Jakobovits, A. (1998) Production and selection of antigen-specific fully human monoclonal antibodies from mice engineered with human Ig loci, *Adv. Drug. Deliv. Rev.* **31**, 33–42.

Jiang, J.X., London, E. (1990) Involvement of denaturation-like changes in *Pseudomonas* exotoxin a hydrophobicity and membrane penetration determined by characterization of pH and thermal transitions. Roles of two distinct conformationally altered states, *J. Biol. Chem.* **265**, 8636–41.

Jinno, Y., Chaudhary, V.K., Kondo, T., Adhya, S., FitzGerald, D.J., Pastan, I. (1988) Mutational analysis of domain I of *Pseudomonas* exotoxin. Mutations in domain I of *Pseudomonas* exotoxin which reduce cell binding and animal toxicity, *J. Biol. Chem.* **263**, 13203–7.

Jinno, Y., Ogata, M., Chaudhary, V.K., Willingham, M.C., Adhya, S., FitzGerald, D., Pastan, I. (1989) Domain II mutants of *Pseudomonas* exotoxin deficient in translocation, *J. Biol. Chem.* **264**, 15953–9.

Johnson, V.G., Youle, R.J. (1989) A point mutation of proline 308 in diphtheria toxin B chain inhibits membrane translocation of toxin conjugates, *J. Biol. Chem.* **264**, 17739–44.

Kiyokawa, T., Shirono, K., Hattori, T. et al. (1989) Cytotoxicity of interleukin 2–toxin toward lymphocytes from patients with adult T-cell leukemia, *Cancer Res.* **49**, 4042–6.

Komatsu, N., Oda, T., Muramatsu, T. (1998) Involvement of both caspase-like proteases and serine proteases in apoptotic cell death induced by ricin, modeccin, diphtheria toxin, and *Pseudomonas* toxin, *J. Biol. Chem.* **124**, 1038–44.

Kondo, T., FitzGerald, D., Chaudhary, V.K., Adhya, S., Pastan, I. (1988) Activity of immunotoxins constructed with modified *Pseudomonas* exotoxin A lacking the cell recognition domain, *J. Biol. Chem.* **263**, 9470–5.

Kounnas, M.Z., Morris, R.E., Thompson, M.R., FitzGerald, D.J., Strickland, D.K., Saelinger, C.B. (1992) The α_2-macroglobulin receptor/low density lipoprotein receptor-related protein binds and internalizes *Pseudomonas* exotoxin A, *J. Biol. Chem.* **267**, 12420–3.

Kreitman, R.J., Pastan, I. (1995) Targeting *Pseudomonas* exotoxin to hematologic malignancies, *Semin. Cancer Biol.* **6**, 297–306.

Kreitman, R.J., Puri, R.K., Paslan, I. (1994) A circularly permuted recombinant interleukin 4 toxin with increased activity, *Proc. Natl. Acad. Sci. U. S. A.* **91**, 6889–93.

Kreitman, R.J., Bailon, P., Chaudhary, V.K., FitzGerald, D.J.P., Pastan, I. (1994) Recombinant immunotoxins containing anti-Tac(Fv) and derivatives of *Pseudomonas* exotoxin produce complete regression in mice of an interleukin-2 receptor-expressing human carcinoma, *Blood* **83**, 426–34.

Kreitman, R.J., Chang, C.N., Hudson, D.V., Queen, C., Bailon, P., Pastan, I. (1994) Anti-Tac(Fab)–PE40, a recombinant double-chain immunotoxin which kills interleukin-2-receptor-bearing cells and induces complete remission in an *in vivo* tumor model, *Int. J. Cancer* **57**, 856–64.

Kreitman, R.J., Chaudhary, V.K., Kozak, R.W., FitzGerald, D.J.P., Waldmann, T.A., Pastan, I. (1992) Recombinant toxins containing the variable domains of the anti-Tac monoclonal antibody to the interleukin-2 receptor kill malignant cells from patients with chronic lymphocytic leukemia, *Blood* **80**, 2344–52.

Kreitman, R.J., Chaudhary, V.K., Waldmann, T., Willingham, M.C., FitzGerald, D.J., Pastan, I. (1990) The recombinant immunotoxin anti-Tac(Fv)–*Pseudomonas* exotoxin 40 is cytotoxic toward peripheral blood malignant cells from patients with adult T-cell leukemia, *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8291–5.

Kreitman, R.J., Wilson, W.H., Robbins, D. et al. (1999) Responses in refractory hairy cell leukemia to a recombinant immunotoxin, *Blood* **94**, 3340–8.

Kreitman, R.J., Wilson, W.H., Whie, J.D. et al. (2000) Phase I trial of recombinant immunotoxin anti-tac(Fv)–PE38 (LMB-2) in patients with hematological malignancies, *J. Clin. Oncol.* **18**, 1622–36.

Kreitman, R.J., Wyndham, H., Bergeron, K. et al. (2001) Efficacy of the anti-CD22 recombinant immunotoxin BL22 in chemotherapy-resistant hairy cell leukemia, *N. Engl. J. Med.* **345**, 241–7.

Kuan, C., Pai, L.H., Pastan, I. (1995) Immunotoxins containing *Pseudomonas* exotoxin targeting Ley damage human endothelial cells in an antibody-specific mode: relevance to vascular leak syndrome, *Clin. Cancer Res.* **1**, 1589–94.

Kuan, C.T., Pastan, I. (1996) Improved antitumor activity of a recombinant anti-Lewis(Y) immunotoxin not requiring proteolytic activation, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 974–8.

Laske, D.W., Youle, R.J., Oldfield, E.H. (1997) Tumor regression with regional distribution of the targeted toxin TF-CRM1 07 in patients with malignant brain tumors, *Nat. Med.* **3**, 1362–8.

Laske, D.W., Muraszko, K.M., Oldfield, E.H. et al. (1997) Intraventricular immunotoxin therapy for leptomeningeal neoplasia, *Neurosurgery* **41**, 1039–49.

Lau, J.L., Fowler, J., Ghosh, L. (1988) Epidermal growth factor in normal and neoplastic kidney and bladder, *J. Urol.* **139**, 170–5.

LeMaistre, C.F., Saleh, M.N., Kuzel, T.M. et al. (1998) Phase I trial of a ligand fusion-protein (DAB389IL-2) in lymphomas expressing the receptor for interleukin-2, *Blood* **91**, 399–405.

Lindstrom, A.L., Erlandsen, S.L., Kersey, J.H., Pennell, C.A. (1997) An *in vitro* model for toxin-mediated vascular leak syndrome: ricin toxin A chain increases the permeability of human endothelial cell monolayers, *Blood* **90**, 2323–34.

Lorberboum-Galski, H., FitzGerald, D., Chaudhary, V., Adhya, S., Pastan, I. (1988) Cytotoxic activity of an interleukin 2–*Pseudomonas* exotoxin chimeric protein produced in *Escherichia coli*, *Proc. Natl. Acad. Sci. U. S. A.* **85**, 1922–6.

Lorimer, I.A., Keppler-Hafkemeyer, A., Beers, R.A., Pegram, C.N., Bigner, D.D., Pastan, I. (1996) Recombinant immunotoxins specific for a mutant epidermal growth factor receptor: targeting with a single chain antibody variable domain isolated by phage display, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 14815–20.

Mansfield, E., Amlot, P., Pastan, I., FitzGerald, D.J. (1997) Recombinant RFB4 immunotoxins exhibit potent cytotoxic activity for CD22-bearing cells and tumors, *Blood* **90**, 2020–6.

Marks, J.D., Hoogenboom, H.R., Bonnert, T.P., McCafferty, J., Griffiths, A.D., Winter, G. (1991) By-passing immunization. Human antibodies from V-gene libraries displayed on phage, *J. Mol. Biol.* **222**, 581–97.

Mavroudis, O.A., Jiang, Y.Z., Hensel, N. et al. (1996) Specific depletion of alloreactivity against haplotype mismatched related individuals: a new approach to graft-versus-host disease prophylaxis in haploidentical bone marrow transplantation, *Bone Marrow Transplant.* **17**, 793–9.

McCafferty, J., Griffiths, A.D., Winter, G., Chiswell, D.J. (1990) Phage antibodies: filamentous phage displaying antibody variable domains, *Nature* **348**, 552–4.

McGuinness, B.T., Walter, G., FitzGerald, K. et al. (1996) Phage diabody repertoires for selection of large numbers of bispecific antibody fragments, *Nat. Biotechnol.* **14**, 1149–54.

McLaughlin, P., Grillo-Lopez, A.J., Link, B.K. et al. (1998) Rituximab chimeric anti-CD20 monoclonal antibody therapy for relapsed indolent lymphoma: half of patients respond to a four-dose treatment program, *J. Clin. Onco1.* **16**, 2825–33.

Montagna, O., Yvon, E., Calcaterra, V. et al. (1999) Depletion of alloreactive T cells by a specific anti-interleukin-2 receptor p55 chain immunotoxin does not impair *in vitro* antileukemia and antiviral activity, *Blood* **93**, 3550–7.

Moroney, S.E., O'Alarcao, L.J., Goldmacher, V.S. et al. (1987) Modification of the binding site(s) of lectins by an affinity column carrying an activated galactose-terminated ligand, *Biochemistry* **26**, 8390–8.

Moskaug, J.O., Sletten, K., Sandvig, K., Olsnes, S. (1989) Translocation of diphtheria toxin A-fragment to the cytosol. Role of the site of interfragment cleavage, *J. Biol. Chem.* **264**, 15709–13.

Multani, P.S., O'Oay, S., Nadler, L.M., Grossbard, M.L. (1998) Phase II clinical trial of bolus infusion anti-B4 blocked ricin immunoconjugate in patients with relapsed B-cell non-Hodgkin's lymphoma, *Clin. Cancer Res.* **4**, 2599–604.

Myers, D.A., Villemez, C.L. (1988) Specific chemical cleavage of diphtheria toxin with hydroxylamine. Purification and characterization of the modified proteins, *J. Biol. Chem.* **263**, 17122–7.

Ogata, M., Pastan, I., FitzGerald, D. (1991) Analysis of *Pseudomonas* exotoxin activation and conformational changes by using monoclonal antibodies as probes, *Infect. Immun.* **59**, 407–14.

Ogata, M., Chaudhary, V.K., Pastan, I., FitzGerald, D.J. (1990) Processing of *Pseudomonas* exotoxin by a cellular protease results in the generation of a 37,000-Da toxin fragment that is translocated to the cytosol, *J. Biol. Chem.* **265**, 20678–85.

Olsnes, S., Pihl, A. (1982) Toxic Lectins and Related Proteins, in: Cohen, P. van Heyningen, S. (Eds.) *Molecular Action of Toxins*

and Viruses, Elsevier Science, New York, pp. 51.

Omura, F., Kohno, K., Uchida, T. (1989) The histidine residue of codon 715 is essential for function of elongation factor 2, *Eur. J. Biochem.* **180**, 1–8.

Onda, M., Kreitman, R.J., Vasmatzis, G., Lee, B., Pastan, I. (1999) Reduction of the nonspecific animal toxicity of anti-Tac(Fv)–PE38 by mutations in the framework regions of the Fv which lower the isoelectric point, *J. Immunol.* **163**, 6072–7.

Orlandi, R., Gussow, D.H., Jones, P.T., Winter, G. (1989) Cloning immunoglobulin variable domains for expression by the polymerase chain reaction, *Proc. Natl. Acad. Sci. U. S. A.* **86**, 3833–7.

Ouvic, M., Kuzel, T., Olsen, E. et al. (1998) Quality of life is significantly improved in CTCl patients who responded to DAB3891l-2 (ONTAK) fusion protein, *Blood* **92**(Suppl. 1), 2572.

Pai, L.H., Pastan, I. (1998) Clinical trials with *Pseudomonas* exotoxin immunotoxins, *Curr. Topics Microbiol. Immunol.* **234**, 83–96.

Pai, L.H., Wittes, R., Setser, A., Willingham, M.C., Pastan, I. (1996) Treatment of advanced solid tumors with immunotoxin LMB-1: an antibody linked to *Pseudomonas* exotoxin, *Nat. Med.* **2**, 350–3.

Pai-Scherf, L.H., Vill, J., Pearson, D. et al. (1999) Hepatotoxicity in cancer patients receiving erb-38, a recombinant immunotoxin that targets the erbB2 receptor, *Clin. Cancer Res.* **5**, 2311–5.

Pasqualucci, L., Wasik, M., Teicher, B.A. et al. (1995) Antitumor activity of anti-CD30 immunotoxin (Ber-H2/saporin) *in vitro* and in severe combined immunodeficiency disease mice xenografted with human CD30$^+$ anaplastic large-cell lymphoma, *Blood* **85**, 2139–46.

Pastan, I., FitzGerald, D. (1991) Recombinant toxins for cancer treatment, *Science* **254**, 1173–7.

Pastan, I., Lovelace, E., Rutherford, A.V., Kunwar, S., Willingham, M.C., Peehl, D.M. (1993) PR1–a monoclonal antibody that reacts with an antigen on the surface of normal and malignant prostate cells, *J. Natl. Cancer Inst.* **85**, 1149–54.

Pastan, I., Lovelace, E.T., Gallo, M.G., Rutherford, A.V., Magnani, J.L., Willingham, M.C. (1991) Characterization of monoclonal antibodies B1 and B3 that react with mucinous adenocarcinomas, *Cancer Res.* **51**, 3781–7.

Pastan, I.H., Pai, L.H., Brinkmann, U., Fitzgerald, D.J. (1995) Recombinant toxins: new therapeutic agents for cancer, *Ann. N.Y. Acad. Sci.* **758**, 345–54.

Pastan, I.H., Archer, G.E., McLendon, R.E. et al. (1995) Intrathecal administration of single-chain immunotoxin, LMB-7 [B3(Fv)–PE38], produces cures of carcinomatous meningitis in a rat model, *Proc. Natl. Acad. Sci. U. S. A.* **92**, 2765–9.

Persson, M.A., Caothien, R.H., Burton, D.R. (1991) Generation of diverse high-affinity human monoclonal antibodies by repertoire cloning, *Proc. Natl. Acad. Sci. U. S. A.* **88**, 2432–6.

Prior, T.I., Helman, L.J., FitzGerald, D.J., Pastan, I. (1991) Cytotoxic activity of a recombinant fusion protein between insulin-like growth factor I and *Pseudomonas* exotoxin, *Cancer Res.* **51**, 174–80.

Puri, R.K., Leland, P., Kreilman, R.J., Paslan, I. (1994) Human neurological cancer cells express interleukin-4 (IL-4) receptors which are targets for the toxic effects of IL-4–*Pseudomonas* exotoxin chimeric protein, *Int. J. Cancer* **58**, 574–81.

Puri, R.K., Hoon, D.S., Leland, P. et al. (1996) Preclinical development of a recombinant toxin containing circularly permuted interleukin 4 and truncated *Pseudomonas* exotoxin for therapy of malignant astrocytoma, *Cancer Res.* **56**, 5631–7.

Raag, R., Whitlow, M. (1995) Single-chain Fvs, *FASEB J.* **9**, 73–80.

Re, G.G., Waters, C., Poisson, L., Willingham, M.C., Sugamura, K., Frankel, A.E. (1996) Interleukin 2 (lL-2) receptor expression and sensitivity to diphtheria fusion toxin DAB(389) lL-2 in cultured hematopoietic cells, *Cancer Res.* **56**, 2590–5.

Reiter, Y., Pastan, I. (1996) Antibody engineering of recombinant Fv–immunotoxins for improved targeting of cancer: disulfide-stabilized Fv–immunotoxins, *Clin. Cancer Res.* **2**, 245–52.

Reiter, Y., Brinkmann, U., Lee, B., Pastan, I. (1996) Engineering antibody Fv fragments for cancer detection and therapy: disulfide-stabilized Fv fragments, *Nat. Biotechnol.* **14**, 1239–45.

Reiter, Y., Wright, A.F., Tonge, D.W., Pastan, I. (1996) Recombinant single-chain and disulfide-stabilized Fv–immunotoxins that cause complete regression of a human colon cancer xenograft in nude mice, *Int. J. Cancer* **67**, 113–23.

Reiter, Y., Di Carlo, A., Fugger, L., Engberg, J., Pastan, I. (1997) Peptide-specific killing of antigen-presenting cells by a recombinant antibody–toxin fusion protein targeted to major histocompatibility complex/peptide class I complexes with T cell receptor-like specificity, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 4631–6.

Reiter, Y., Pai, L.H., Brinkmann, U., Wang, Q.C., Pastan, I. (1994) Antitumor activity and pharmacokinetics in mice of a recombinant immunotoxin containing a disulfide-stabilized Fv fragment, *Cancer Res.* **54**, 2714–8.

Reiter, Y., Brinkmann, U., Jung, S.H., Lee, B., Kasprzyk, P.G., King, C.R., Pastan, I. (1994) Improved binding and antitumor activity of a recombinant anti-erbB2 immunotoxin by disulfide stabilization of the Fv fragment, *J. Biol. Chem.* **269**, 18327–31.

Rosenberg, S.A. (1999) A new era for cancer immunotherapy based on the genes that encode cancer antigens, *Immunity* **10**, 281–7.

Ross, J.S., Fletcher, J.A. (1999) HER-2/*neu* (c-erb-B2) gene and protein in breast cancer, *Am. J. Clin. Pathol.* **112**(1 Suppl 1), S53–67.

Rozemuller, H., Rombouts, W.J., Touw, I.P., FitzGerald, D.J., Kreitman, R.J., Pastan, I., Hagenbeek, A., Martens, A.C. (1996) Treatment of acute myelocytic leukemia with interleukin-6 *Pseudomonas* exotoxin fusion protein in a rat leukemia model, *Leukemia* **10**, 1796–803.

Sandvig, K., Olsnes, S. (1980) Diphtheria toxin entry into cells is facilitated by low pH, *J. Cell. Biol.* **87**, 828–32.

Sausville, E.A., Headlee, D., Stetler-Stevenson, M. et al. (1995) Continuous infusion of the anti-CD22 immunotoxin IgG–RFB4–SMPT–dgA in patients with B-cell lymphoma: a phase I study, *Blood* **85**, 3457–65.

Scheuermann, R.H., Racila, E. (1995) CD19 antigen in leukemia and lymphoma diagnosis and immunotherapy, *Leuk. Lymph.* **18**, 385–97.

Schnell, R., Vitetta, E., Schindler, J. et al. (1998) Clinical trials with an anti-CD25 ricin A-chain experimental and immunotoxin (RFT5–SMPT–dgA) in Hodgkin's lymphoma, *Leuk. Lymph.* **30**, 525–37.

Seetharam, S., Chaudhary, V.K., FitzGerald, D., Pastan, I. (1991) Increased cytotoxic activity of *Pseudomonas* exotoxin and two chimeric toxins ending in KDEL, *J. Biol. Chem.* **266**, 17376–81.

Senderowicz, A.M., Vitetta, E., Headlee, D. et al. (1997) Complete sustained response of a refractory post-transplantation, large B-cell lymphoma to an anti-CD22 immunotoxin, *Ann. Intern. Med.* **126**, 882–5.

Shen, G-L., Li, J-L., Ghetie, M.A., Ghetie, V. et al. (1988) Evaluation of four C022 antibodies as ricin A-chain containing immunotoxins for the *in vivo* therapy of human B-cell leukemias and lymphomas, *Int. J. Cancer* **42**, 792–7.

Siegall, C.B., Chaudhary, V.K., FitzGerald, D.J., Pastan, I. (1989) Functional analysis of domains II, Ib, and III of *Pseudomonas* exotoxin, *J. Biol. Chem.* **264**, 14256–61.

Siegall, C.B., Liggitt, D., Chace, D., Tepper, M.A., Fell, H.P. (1994) Prevention of immunotoxin-mediated vascular leak syndrome in rats with retention of antitumor activity, *Proc. Natl. Acad. Sci. U. S. A.* **91**, 9514–8.

Siegall, C.B., Schwab, G., Nordan, R.P., FitzGerald, D.J., Pastan, I. (1990) Expression of the interleukin 6 receptor and interleukin 6 in prostate carcinoma cells, *Cancer Res.* **50**, 7786–8.

Siegall, C.B., Liggitt, D., Chace, D., Mixan, B., Sugai, J., Davidson, T., Steinitz, M. (1997) Characterization of vascular leak syndrome induced by the toxin component of *Pseudomonas* exotoxin-based immunotoxins and its potential inhibition with nonsteroidal anti-inflammatory drugs, *Clin. Cancer Res.* **3**, 339–45.

Siegall, C.B., Chace, D., Mixan, B. et al. (1994) *In vitro* and *in vivo* characterization of BR96 sFv–PE40. A single-chain immunotoxin fusion protein that cures human breast carcinoma xenografts in athymic mice and rats, *J. Immunol.* **152**, 2377–84.

Siegall, C.B., Epstein, S., Speir, E. et al. (1991) Cytotoxic activity of chimeric proteins composed of acidic fibroblast growth factor and *Pseudomonas* exotoxin on a variety of cell types, *FASEB J.* **13**, 2843–9.

Skerra, A., Pluckthun, A. (1988) Assembly of a functional immunoglobulin Fv fragment in *Escherichia coli*, *Science* **240**, 1038–41.

Skrepnik, N., Zieske, A.W., Robert, E., Bravo, J.C., Mera, R., Hunt, J.D. (1998) Aggressive

administration of recombinant oncotoxin AR209 (anti-ErbB-2) in athymic nude mice implanted with orthotopic human non-small cell lung tumours, *Eur. J. Cancer* **34**, 1628–33.

Smith, G.P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface, *Science* **228**, 1315–7.

Stone, M.J., Sausville, E.A., Fay, L.W. et al. (1996) A phase I study of bolus versus continuous infusion of the anti-CD19 immunotoxin, IgG–HD37–dgA, in patients with B-cell lymphoma, *Blood* **88**, 1188–97.

Studier, F.W., Moffatt, B.A. (1986) Use of bacteriophage T7 polymerase to direct selective expression of cloned genes, *J. Mol. Bio1.* **189**, 113–30.

Thorpe, P.E., Ross, W.C.J., Brown, A.N.F. et al. (1984) Blockade of the galactose-binding sites of ricin by its linkage to antibody. Specific cytotoxic effects of the conjugates, *Eur. J. Biochem.* **140**, 63–71.

Thorpe, P.E., Wallace, P.M., Knowles, P.P. et al. (1988) Improved anti-tumor effects of immunotoxins prepared with deglycosylated ricin-A chain and hindered disulfide linkages, *Cancer Res.* **48**, 6396–403.

Tsutsumi, Y., Onda, M., Nagata, S., Lee, B., Kreitman, R.J., Pastan, I. (2000) Site-specific chemical modification with polyethylene glycol of recombinant immunotoxin anti-Tac(Fv)–PE38 (LMB-2) improves antitumor activity and reduces animal toxicity and immunogenicity, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8548–53.

Uckun, F.M., Reaman, G.H. (1995) Immunotoxins for treatment of leukemia and lymphoma, *Leuk. Lymph.* **18**, 195–201.

Veale, D., Kerr, N., Gibson, G.J., Harris, A.L. (1989) Characterization of epidermal growth factor receptor in primary human non-small cell lung cancer, *Cancer Res.* **49**, 1313–7.

Vitetta, E.S., Fulton, R.J., May, R.D., Till, M., Uhr, J.W. (1987) Redesigning nature's poisons to create anti-tumor reagents, *Science* **238**, 1098–104.

Vitteta, E.S., Sonte, M., Amlot, P. et al. (1991) Phase I immunotoxin trial in patients with B-cell lymphoma, *Cancer Res.* **51**, 4052–8.

Waldmann, T.A., Pastan, I., Gansow, O.A. et al. (1992) The multichain interleukin-2 receptor: a target for immunotherapy, *Ann. Intern. Med.* **116**, 148–60.

Wang, R.F., Rosenberg, S.A. (1999) Human tumor antigens for cancer vaccine development, *Immunol. Rev.* **170**, 85–100.

Ward, E.S., Gussow, D., Griffiths, A.D., Jones, P.T., Winter, G. (1989) Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*, *Nature* **341**, 544–6.

Waters, C.A., Schimke, P.A., Snider, C.E. et al. (1990) Interleukin 2 receptor-targeted cytotoxicity. Receptor binding requirements for entry of a diphtheria toxin-related interleukin 2 fusion protein into cells, *Eur. J. Immunol.* **20**, 785–91.

Waurzyniak, B., Schneider, E.A., Tumer, N. et al. (1997) *In vivo* toxicity, pharmacokinetics, and antileukemic activity of TXU (anti-CD7)–pokeweed antiviral protein immunotoxin, *Clin. Cancer Res.* **3**, 881–90.

Wels, W., Harwerth, I.M., Mueller, M., Groner, B., Hynes, N.E. (1992) Selective inhibition of tumor cell growth by a recombinant single-chain antibody-toxin specific for the erbB-2 receptor, *Cancer Res.* **52**, 6310–7.

Williams, D.P., Wen, Z., Watson, R.S., Boyd, J., Strom, T.B., Murphy, J.R. (1990) Cellular processing of the interleukin-2 fusion toxin DAB486–IL-2 and efficient delivery of diphtheria fragment A to the cytosol of target cells requires Arg194, *J. Biol. Chem.* **265**, 20673–7.

Williams, D.P., Parker, K., Bacha, P. et al. (1987) Diphtheria toxin receptor binding domain substitution with interleukin-2: genetic construction and properties of a diphtheria toxin-related interleukin-2 fusion protein, *Protein Eng.* **1**, 493–8.

Wilson, B.A., Collier, R.J. (1992) Diphtheria toxin and *Pseudomonas aeruginosa* exotoxin A: active-site structure and enzymic mechanism, *Curr. Topics Microbiol. Immunol.* **175**, 27–41.

Winter, G., Milstein, C. (1991) Man-made antibodies, *Nature* **349**, 293–9.

Winter, G., Griffiths, A.D., Hawkins, R.E., Hoogenboom, H.R. (1994) Making antibodies by phage display technology, *Annu. Rev. Immunol.* **12**, 433–55.

Wozniak, D.J., Hsu, L.Y., Galloway, D.R. (1988) His-426 of the *Pseudomonas aeruginosa* exotoxin A is required for ADP-ribosylation

of elongation factor II, *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8880–4.

Yang, X.D., Jia, X.C., Corvalan, J.R., Wang, P., Davis, C.G., Jakobovits, A. (1999) Eradication of established tumors by a fully human monoclonal antibody to the epidermal growth factor receptor without concomitant chemotherapy, *Cancer Res.* **59**, 1236–43.

In vitro Fertilization

Carol A. Brenner[1,3], *Hans Michael Kubisch*[2], *and Barry D. Bavister*[1,3]
[1] *University of New Orleans, New Orleans, LA, USA*
[2] *Tulane National Primate Research Center, Covington, LA, USA*
[3] *Tulane Institute for Reproductive Medicine, Center for Excellence in Women's Health, Tulane University Health Sciences Center, New Orleans, LA, USA*

1 Introduction 533

2 Biology of Spermatozoa 533
2.1 Sperm Chemotaxis 533
2.2 Sperm Capacitation and Hyperactivation 534

3 Fertilization 535
3.1 Sperm Binding to the Zona Pellucida and the Acrosome Reaction 535
3.2 Molecular Basis of Sperm–Egg Membrane Adhesion/Binding 537
3.3 Oocyte-sperm Fusion and Oocyte Activation 537
3.4 *In vitro* Fertilization 539

4 Mitochondria 540
4.1 The Role of Mitochondria in ART 540
4.2 Mitochondrial Inheritance following Ooplasmic Transplantation 544

5 Summary and Conclusions 545

Bibliography 545
Books and Reviews 545
Primary Literature 545

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

Keywords

Assisted Reproductive Technology
Refers to any intervention designed to help or improve fertility, including artificial insemination and *in vitro* fertilization.

Mitochondria
The major energy producing organelles of the oocyte; however, it is not completely clear when their metabolic role begins during early embryonic development. The sperm mitochondria are eliminated during fertilization or early embryonic development.

Oocyte
The female gamete; in most mammals, the oocyte (or egg) is arrested in metaphase of the second meiotic division before it participates in fertilization.

Spermatozoa
The male gametes; supply the paternal genome during fertilization and activate the egg to begin developing into an embryo.

Zona Pellucida
A glycoprotein envelope enclosing the oocyte; in most mammals, the zona prevents more than one spermatozoon from reaching the oocyte.

Fertilization in mammals takes place within the female reproductive tract, so information about mechanisms and molecules regulating them must come largely from studies on *in vitro* fertilization. Such studies have been most informative about fertilization mechanisms, including chemotaxis, the sperm acrosome reaction, the molecular nature of sperm–egg binding, and events occurring during and after sperm–egg plasma membrane fusion. However, many details remain to be elucidated, such as the molecular nature of sperm capacitation, the precise way in which the spermatozoon traverses the egg's zona pellucida, the identity of the sperm-derived signal that initiates egg activation, the mechanistic linkage between calcium oscillations and subsequent developmental events, and the mechanism by which paternal mitochondria are normally eliminated so that only maternal mtDNA is inherited. A major concern is whether use of novel technologies for alleviating human infertility can introduce anomalies such as paternal mtDNA inheritance, and if so, what consequences this may have for pre- or postnatal development. Until recently, it was difficult to detect small amounts of anomalous mtDNA that might be present within an embryo. However, the application of highly sensitive techniques such as RT-PCR and molecular beacon analysis can compensate for the very small numbers of oocytes that are available for analysis in mammals and so will rapidly provide further insights into the molecular mechanisms of fertilization and subsequent development.

1 Introduction

Until about 1960, most of what we knew about fertilization in animals came from studies with species such as sea urchins and frogs, for the simple reason that external fertilization allows detailed study of mechanisms and the molecules responsible for them. In contrast, very little can be deduced about mammalian fertilization as it occurs within the oviduct, except for ultrastructural observations on *in situ* fixed material. It was not until the advent of *in vitro* fertilization (IVF) technology for mammals that detailed, mechanistic studies could be undertaken. The first convincing reports of successful IVF were published in the late 1950s and the first human IVF was achieved in 1969. Even though IVF technology has been greatly improved (e.g. the ability to support sperm capacitation *in vitro*), and it is now routine in many species, there are still some inherent constraints. The relatively low number and small size of eggs that can be obtained from a single female mammal (compared with sea urchins or frogs) still impedes research progress. Moreover, mammalian IVF requires specialized and highly controlled culture conditions, such as precise gas mixtures, controlled temperatures (which vary among species), and (sometimes) specific chemical activators for spermatozoa to achieve fertilization. All these factors need to be optimized to maximize success and thus, make the most use of very limited amounts of material (eggs or embryos).

Because of the great reliance on IVF technology for advancing understanding of mammalian fertilization mechanisms, these two research areas have gone hand in hand for the past 40 years. Along the way, much has been learned about the process of fertilization, but it should be recognized that we cannot always be certain that events *in vitro* precisely represent those occurring in the natural milieu. The following sections describe some of the important findings derived from studies with mammalian gametes *in vitro* and indicate key areas where more knowledge is needed.

2 Biology of Spermatozoa

2.1 Sperm Chemotaxis

Before a spermatozoon can fertilize an egg, it has to reach the site of fertilization within the oviduct. Once it reaches there, the spermatozoon still has to make contact with the oocyte, which raises the intriguing question whether this is a random process or whether it is guided by some chemical signal released by the oocyte.

The fact that oocytes produce signals to attract spermatozoa has been well documented in marine species that reproduce by external fertilization, such as sea urchins. However, finding evidence of sperm chemotaxis in mammals as well as deciphering the nature of the signals, and the response pathways they may elicit has proven to be a much more arduous task. Early evidence suggested that follicular fluid contained substances with chemotactic properties. However, often this evidence was difficult to confirm, mainly because observations were limited to *in vitro* conditions and, more pertinently, because of the uncertainties inherent in trying to distinguish chemotaxis from random sperm accumulations that can be caused by other factors.

New support for the existence of chemotaxis in mammalian fertilization has come from recent reports of olfactory

receptors on human sperm cells. While it had been known for some time that olfactory receptor (OR) proteins can be found not only in human testicular tissues but also in the membranes of mature dog spermatozoa, the recent discovery of the odorant receptor hOR17-4 on spermatozoa has given new impetus to the idea that such receptors may be important in the ability of spermatozoa to receive chemical signals. *hOR17-4* belongs to an OR gene family that has 500 to 1000 members, many of which are expressed in the testis and several other tissues where their role is unclear. Several of the ~50 testicular ORs are made predominantly or exclusively in spermatogenic cells derived from human testicular tissue. These new ORs belong to the greater superfamily of G protein-coupled receptors, which includes the opsins and a large variety of hormone and neurotransmitter receptors. Through heterologous expression of *hOR17-4* in a human kidney cell line, it was possible to identify several agonists. One of these, bourgeonal, was not only able to induce chemotactic behavior in spermatozoa but also to produce some of the changes in intracellular Ca^{2+} usually associated with sperm modifications seen *in vitro*.

2.2 Sperm Capacitation and Hyperactivation

An important aspect of male fertility is the physiological priming of spermatozoa by a multifactorial process called *capacitation*. The culmination of this process is when spermatozoa are capable of binding to the sugar residues on the egg's extracellular envelope called the *zona pellucida* (ZP). Early attempts to use ejaculated spermatozoa for *in vitro* fertilization of oocytes met with little success. The explanation for these failures came from the work of two investigators in 1951 who discovered that spermatozoa needed to remain in the female reproductive tract for several hours before they gained the capacity to fertilize an oocyte. This process was termed *capacitation*. The modifications that occur in spermatozoa during this period are still poorly understood. Interestingly, the need for capacitation appears to be restricted to mammals where it represents the first step in a cascade of events that allows spermatozoa to rapidly undergo the necessary changes to penetrate the oocyte. As capacitation proceeds, a number of biochemical changes occur in spermatozoa (1) increased adenylyl cyclase activity and increased levels of cyclic adenosine monophosphate (cAMP), (2) protein tyrosine kinase phosphorylation of a subset of sperm components, (3) elevated intrasperm pH, (4) Ca^{2+} influx, (5) loss of sperm surface components, (6) modification/alteration of the sperm plasma membrane, and (7) changes in the surface lectin-binding patterns. Capacitation is not accompanied by visible morphological changes, but is manifested as modifications in the composition of membrane surface antigens that become accessible over the acrosome. Many intra-acrosomal contents including glycohydrolases appear on mouse and human capacitating/capacitated spermatozoa in a time-dependent manner.

In rhesus and cynomolgus monkeys, capacitation of spermatozoa has also been induced *in vitro* by exposure to the cyclic nucleotide analog dbcAMP together with caffeine. Exposure of cynomolgus spermatozoa to these reagents results in increased tyrosine phosphorylation of sperm tail proteins, which may be one of the mediators of the changes in motility associated with capacitation in this species. These motility changes, which are

characterized by a pronounced increase in the amplitude of flagellar bending, have been termed *hyperactivation* and were first reported by Yanagimachi. Other factors that have been associated with hyperactivation are an increased level of intracellular cAMP as well as a change in the level of Ca^{2+} in the flagellum and sperm head. Whether hyperactivation and capacitation are tightly coupled *in vivo* has not yet been resolved with certainty. A number of studies using various sperm treatments have shown that at least *in vitro* the two processes can be uncoupled. It is interesting to note that in humans only capacitated spermatozoa appear to be chemotactically responsive to follicular fluid, indicating that capacitation assists spermatozoa to accomplish fertilization in more than one respect. It will be interesting to determine if capacitation is a prerequisite for chemotaxis involving the olfactory receptors described by Spehr et al.

3 Fertilization

3.1 Sperm Binding to the Zona Pellucida and the Acrosome Reaction

The mammalian oocyte is surrounded by a mass of small somatic cells called *cumulus cells*, which in turn are embedded in an extracellular matrix they secrete that is rich in hyaluronic acid and contains various proteins. Spermatozoa pass through this matrix presumably with the aid of a membrane-bound hyaluronidase. The presence of one of these proteins, termed PH-20, has been confirmed in macaque and cynomolgus spermatozoa. However, Talbot showed that spermatozoa from some animals, such as frogs, which do not contain hyaluronidase, nevertheless could pass through the mammalian cumulus matrix. This makes the role of hyaluronidase in cumulus penetration somewhat inconclusive.

The zona pellucida of the oocyte contains several glycoproteins that are crucial for the ability of spermatozoa to gain access to the egg. The structure and function of these molecules have been best characterized in the mouse zona pellucida, in which there are three glycoproteins termed ZP1, ZP2, and ZP3 (alternately termed ZPA, ZPB, and ZPC). Each of these has a specific function, including binding of spermatozoa and induction of the acrosome reaction. Homologs exist in other species but there appear to be more than three zona pellucida proteins and/or their molecular weights vary over a range because of differential glycosylation. The genes encoding the zona pellucida proteins have been cloned in a variety of species including the cynomolgus monkey and baboon and these proteins have been localized histochemically on the zona pellucida in cynomolgus monkey oocytes. Binding of spermatozoa to the zona pellucida appears to be completed in two phases, comprising an initial loose attachment that is quickly followed by tighter binding. Evidence gained mostly from work in the mouse suggests that ZP3 is the protein primarily involved in the initial binding of spermatozoa and the induction of the acrosome reaction while ZP1 acts as a secondary receptor for stronger binding.

The structure of the human zona pellucida surrounding mature, preovulatory oocytes has not been routinely examined. Using antisera generated against synthetic ZP3 peptides, human metaphase II oocytes obtained after insemination *in*

vitro have been studied under a variety of conditions. Interestingly, in cases of failed fertilization with either a poor ovarian response or with morphologically defective oocytes (and in the presence of normal sperm parameters), there is a marked alteration in the zona pellucida immunoreactivity to the antisera. These findings indicated the presence of structural defects of the protein backbone of ZP3 within the zona pellucida, and suggest that this defect may also be a possible cause of fertilization failure. There is virtually no information about putative abnormalities that could operate at the level of receptor–ligand interaction during human gamete binding. Several monosaccharides located on ZP3 have been implicated as the binding epitopes including mannose, α-galactose, and β-*N*-acetylglucosamine. Much of this evidence relies on the use of synthetic oligosaccharides in competition assays *in vitro*, which allow for basic binding studies, but these may not necessarily reflect the affinities found *in vivo*.

In contrast, the ligand on the sperm membrane that mediates binding to ZP3 in primates remains to be identified. Several putative receptors have been isolated from mouse or pig spermatozoa including zonadhesin, proacrosin, and spermadhesins. The most extensively studied receptor is β1,4-galactosyltransferase. Support for its role in sperm binding has come from studies with genetically engineered mice, where overexpression has lead to increased binding to ZP3. However, it is likely that β1,4-galactosyltransferase acts in cooperation with other molecules because male mice in which β1,4-galactosyltransferase expression has been ablated remain fertile, although they exhibit altered sperm-binding characteristics and slower penetration of the zona pellucida. Further studies are necessary to elucidate the role of β1,4-galactosyltransferase in human sperm binding to the zona pellucida.

When a spermatozoon makes contact with the zona pellucida it undergoes several morphological changes that are collectively termed the *acrosome reaction*. The acrosome is a large, membrane-bound secretory vesicle, located at the anterior of spermatozoa, that contains a number of enzymes, including hyaluronidase and acrosin. Spermatozoa of many mammals (rat, bull, ram, and human) have very thin or lightly fitted acrosomes, which makes their acrosome reactions difficult to detect by light microscopy. Several methods are currently available for monitoring acrosome reactions using transmission electron microscopy, light microscopy, or by staining spermatozoa with specific fluoresceinated plant lectins. In rhesus spermatozoa, the acrosome reaction can be observed as quickly as 11 seconds after tight binding to the zona pellucida and is associated with changes in the pattern of tail movements.

During the acrosome reaction, the acrosomal membrane is modified by fusion with the overlying sperm plasma membrane resulting in the release of acrosomal enzymes. The molecular events within spermatozoa that are responsible for the acrosome reaction are not well understood. However, there is evidence that ZP3 binding activates two pathways, one leading to the opening of a channel that produces a transient influx of Ca^{2+}, the other eliciting a rise in internal pH. Both mechanisms appear necessary for the sustained release of Ca^{2+} from an intracellular Ca^{2+} pool. The acrosome reaction also causes the release and activation of acrosin, a proteolytic enzyme that is most likely one of the enzymes involved in the digestion of the zona

pellucida that facilitates the transport of spermatozoa into the perivitelline space.

3.2 Molecular Basis of Sperm–Egg Membrane Adhesion/Binding

During the last 15 years, a model for sperm–egg plasma membrane binding has been developed that suggests that this event results from the adhesion between an integrin on the egg and an integrin ligand (i.e. the disintegrin domain of a member of the ADAM family of proteins) on the spermatozoon. Initial studies showed that the sperm surface protein called *fertilin* was able to inhibit fusion with the oocyte membrane. This protein, which is composed of two subunits, is now named fertilin α (ADAM 1) and fertilin ß (ADAM 2). Sequencing of these two subunits revealed three regions of interest: (1) a metalloprotease domain that has a consensus sequence indicating enzyme activity in the fertilin α subunit but not in the fertilin ß subunit, (2) a disintegrin domain in both subunits, previously shown to be an integrin ligand in snake venom peptides, and (3) a region in the fertilin α subunit that could be modeled as an amphipathic α helix and was reminiscent of the fusion peptides of some viruses.

Both subunits of fertilin have also been identified in other mammalian species. The identification and sequencing of fertilin ß and fertilin α subunits in the mouse has led to the discovery of many other members in this new family of proteins. As new proteins have been identified, additional functions have been determined for members of this family, and it has been found that domains other than the disintegrin and metalloprotease domains may be functional. The finding that fertilin and other members of the ADAM family are present on the surface of mature spermatozoa has led to the hypothesis that spermatozoa bind via the disintegrin domain of a sperm ADAM to an egg integrin.

To further determine the involvement of specific members of the ADAM family in sperm binding (and possibly fusion), knockout mice have been generated in which the genes coding for fertilin ß, cyritestin, or both (double knockout) have been ablated. Spermatozoa from each of these mouse lines showed dramatically (90%) reduced binding to zona-free oocytes. The most plausible interpretation of these results is that each of these proteins is necessary to mediate sperm–egg plasma membrane binding.

The question of which, if any, membrane-bound integrins are required on the oocyte for sperm–egg binding (and subsequent fusion) has also been addressed using knockout mice. Recently, oocytes from mice in which expression of the $\alpha 6$ integrin subunit was ablated were shown to have no reduction in their capacity to bind to or fuse with sperm, suggesting that if $\alpha 6$ integrins do participate in sperm–egg adhesion, they are not the only molecule involved in this process.

3.3 Oocyte-sperm Fusion and Oocyte Activation

In most mammals, mature oocytes are arrested at metaphase II (MII), which is maintained by maturation promoting factor (MPF). MPF is a protein complex that consists of a kinase termed $p34^{cdc2}$ and cyclin B. Its primary function is to promote spindle assembly, chromatin condensation, and the breakdown of the nuclear envelope. The activity of MPF

appears to be regulated by the presence of cytostatic factor (CSF), which in turn requires regulation by the products of two genes, the *c-mos* proto-oncogene and *Emi1*.

After traversing the zona pellucida, the spermatozoon enters the perivitelline space and binds to the oocyte membrane, usually via the plasma membrane along the equatorial segment of the sperm head. Binding is mediated by a complex of egg membrane proteins, which probably include integrin, and bind to complementary sperm membrane molecules. Fusion with the spermatozoon results in oscillating increases in Ca^{2+} within the oocyte that in turn triggers fusion of cortical granules with the oocyte membrane and release of enzymes into the perivitelline space. These enzymes induce changes in the zona pellucida that prevent further passage of spermatozoa. Most pertinent of these changes is a modification of the structure of ZP3, which prevents further sperm binding to the zona pellucida and/or the induction of acrosome reactions in spermatozoa already bound to the zona.

The precise mechanism by which the fusion of the spermatozoon with the oocyte plasma membrane elicits the Ca^{2+} increase remains unclear, but there is strong evidence that the trigger is some factor released by the spermatozoon into the oocyte cytoplasm. This is supported by observations that rhesus monkey sperm extracts injected into oocytes can elicit Ca^{2+} oscillations similar to those observed during sperm binding. This is surprising in light of the fact that "intracytoplasmic sperm injection" (ICSI), in which spermatozoa are directly injected into the oocyte, not only bypasses fusion of the gametes but also alters the series of signaling events that occur in the spermatozoon prior to and during interaction with the oocyte's vestments. Recent reports suggest that ICSI and IVF induce similar initial Ca^{2+} responses, although ICSI zygotes exhibit shorter durations of Ca^{2+} oscillations and exhibit diminished degradation of the inositol 1-, 4-, 5-trisphosphate receptor-1.

Fusion of the sperm plasma membrane with the oocyte is followed by decondensation of the sperm head during which some of the proteins (protamines) used to tightly "package" its chromosomes are replaced by oocyte-derived histones. This is followed by formation of a pronuclear membrane and the male pronucleus. Following entry of the sperm tail into the oocyte, sperm-derived mitochondria, which were marked by ubiquitin tags during spermatogenesis, are selectively degraded by oocyte-derived factors, a phenomenon that has been observed in several species including rhesus monkeys.

How the Ca^{2+} oscillations in the oocyte are "translated" into the series of modifications that are observed during activation remains largely unknown. In this context, it is interesting to note that the number and amplitude of sperm-induced Ca^{2+} oscillations appear to differentially regulate the initiation, completion, and, possibly, the temporal order of most, if not all, of the events that occur during activation. One mediator of the Ca^{2+} signal appears to be Ca^{2+}/calmodulin-dependent kinase II (CaMKII). CaMKII activity undergoes transient increases in response to Ca^{2+} oscillations, while conversely inhibition of CaMKII prevents oocyte activation. CaMKII appears to be involved in the regulation of a variety of targets including cyclin degradation, secretion-promoting proteins, and control of oocyte calcium channels.

One important consequence of oocyte activation is the inactivation of MPF through degradation of its cyclin B subunit. Similarly, the mitogen-activated

protein kinase (MAP), which is involved in regulating the cell cycle and is responsible for maintenance of chromatin in its condensed state, is inactivated. Both steps are necessary for the oocyte to resume meiosis.

As a result of oocyte activation, meiosis resumes and the oocyte enters Anaphase II after which the second polar body is extruded, rendering the oocyte haploid. The chromosomes are incorporated into a female pronucleus, which is surrounded by a pronuclear membrane. Replication of male and female-derived DNA normally occurs synchronously, the pronuclear membranes disappear, and the maternal and paternal homologous chromosomes combine, a process called *syngamy*. Following syngamy, the resulting zygote initiates the first mitotic cell division.

Pronuclear migration and proper chromosomal segregation are mediated by the function of a centriole that serves initially as an organizing center for assembly of microtubules by polymerization of α-tubulin molecules. In most mammals, with the exception of mice and possibly other rodents, centrioles appear to be exclusively contributed by spermatozoa, while oocytes contain only a rudimentary similar structure. The centriole initially forms the sperm aster, which is necessary to move the female pronucleus toward its male counterpart. It is replicated during the pronuclear stage, and subsequently organizes the bipolar assembly of the mitotic spindle. In humans, defects in centrioles have been linked to reduced sperm motility and developmental arrest of embryos. The advantage of maternally derived centrioles for fertilization in rodents is unknown, but clearly this strategy is equally as effective as the exclusive reliance on sperm-derived centrioles in other species.

Understanding of the microtubule-mediated motility events that ensure normal sperm-oocyte interactions has been helped by the use of nonhuman primate gametes. Studies with rhesus monkey gametes have shown that the cytoskeletal events during fertilization by IVF and ICSI are very similar to those of human fertilization, and these manipulations of nonhuman primate gametes may help test the safety and improve the efficacy of current assisted reproductive technologies, as well as to develop new techniques. However, in the rhesus monkey, ICSI results in abnormal sperm nuclear remodeling, in part because of the persistence of vesicle-associated membrane protein (VAMP), the acrosome and the perinuclear theca on the sperm head, all of which are normally removed at, or close to, the oocyte cortex during natural and *in vitro* fertilization. Progression through the first cell cycle in ICSI oocytes cannot be completed until these structures have been removed from the developing male pronucleus, demonstrating fundamental differences between ICSI and IVF. While ICSI is of enormous therapeutic value for the treatment of male infertility, basic research using clinically relevant animal models is only now unraveling the cellular and molecular events that allow fertilization by sperm microinjection.

3.4 *In vitro* Fertilization

More than 1 million babies have been born with the help of assisted reproductive technologies since the birth of the first test tube baby more than 25 years ago. It is estimated that one in six couples are infertile. A quarter of a century later, the techniques are far from perfect and it is still important to study the basic mechanisms in animal models. While laboratory and domestic

animals have provided most of the current basic information on fertilization mechanisms and technologies, nonhuman primates are also invaluable because their reproductive physiology is very similar in most respects to that of humans. *In vitro* fertilization in nonhuman primates has been described in several species including squirrel monkeys, rhesus monkeys, cynomolgus monkeys, pigtail macaques, baboons, marmosets, chimpanzees, and lowland gorilla, although in most early experiments the rates of *in vitro* development of IVF embryos were low. Development to the morula stage in rhesus embryos generated by *in vitro* fertilization was first described by Bavister et al., who subsequently provided the first report of the birth of a rhesus infant following transfer of cleavage stage embryos generated by *in vitro* fertilization. Since these initial studies, numerous laboratories have reported successful IVF in nonhuman primates, mostly with rhesus macaques.

In spite of the wealth of information available from experimental studies with nonhuman primates and other animals, the practice of IVF in humans remains inefficient. In the year 2000, the average clinical pregnancy rate (CPR) and live baby rate (LBR) in the United States were 38 and 33% respectively, which is approximately the same as rates reported for natural conceptions. However, it can be estimated that an average of three embryos was transferred in each treatment cycle; thus, the actual CPR and LBR rates per embryo transferred were 13 and 11%, respectively. Although other factors are involved, these rates suggest that the average quality of embryos produced in human IVF is low, and defects in oocyte are suspected to play a major role in this. Stimulating the ovaries with exogenous gonadotropins greatly increases the number of oocytes available for IVF, from the normal number of 1 or 2 up to 20 or so, but the quality and competence of these cohorts of oocytes are quite variable. In laboratory animals and other species, ovarian stimulation with exogenous gonadotropins ("superovulation") reduces the average viability of embryos. This is because superovulation rescues some potentially atretic follicles containing defective oocytes.

The inherent variability of oocyte quality in humans is exacerbated by a wide variety of epigenetic effects on the oocytes and/or embryos resulting from IVF. These effects are due to exposure of oocytes and embryos to the artificial culture milieu that can alter gene expression, metabolism, mitochondrial localization and/or activity, and embryo development. On top of this, the quality or health of human oocytes is progressively diminished with increasing age because of inherent or genetic factors, manifested as reduced fertility and decreased success in assisted reproductive technology (ART). Other extrinsic factors such as nutrition, smoking, and environmental effects can all cause further reductions in the quality of oocytes and spermatozoa. More intensive research is needed to define the mechanisms by which epigenetic factors affect gamete quality, and to devise improved technologies for overcoming these deficiencies and for selecting the most viable gametes for use in clinical IVF.

4 Mitochondria

4.1 The Role of Mitochondria in ART

In spite of intensive research, the precise role of mitochondria in the development

of preimplantation mammalian embryos is still uncertain. Substantial changes in the organization and localization of active mitochondria occur during fertilization in some mammals. Pronounced morphological alterations take place early in embryogenesis that may reflect changes in mitochondrial metabolic functions. In addition, the numbers of mitochondria are substantially reduced during early embryo development.

Mitochondria are now recognized as a central factor in tissue biogenesis and as playing a key role in aging. Mitochondria are semiautonomous organelles each possessing one or more copies of their own restricted genome. In humans, this 16.6-kb circular strand of DNA encodes 37 genes including 13 components of the oxidative phosphorylation (OXPHOS) pathway, two ribosomal RNAs and 22 transfer RNAs. The vast majority of control genes for mitochondrial function, over 200, are found in the nuclear genome of the cell. Therefore, mitochondrial function is dependent upon the exchange of information between the cytoplasm and the nucleus. Mutations in both nuclear and mitochondrial genes have been associated with potentially lethal disorders affecting OXPHOS, prompting much interest in developing animal models to examine this disease process.

Replication and transcription of mtDNA is semiautonomous and requires interaction between the nucleus and the mitochondria. For example, mitochondrial transcription factor A (TFAM), a member of the high mobility group (HMG) of proteins, contains two HMG-box domains and is essential for the maintenance of mtDNA. Because it binds to mtDNA molecules, TFAM was first cloned and co-purified as a transcription factor. It is unlikely that TFAM or mtDNA molecules exist in an unbound form in the mitochondria. TFAM is sufficiently abundant to wrap mtDNA entirely, suggesting that human mtDNA is packaged with TFAM. The 25-kDa TFAM protein is translated from a nuclear gene whose primary function is to regulate the mitochondrial transcripts for both the heavy strand promoters (HSP) and the light strand promoters (LSP). TFAM is thought to enhance transcription via interaction with other factors such as DNA polymerase, which is the polymerase specific to the mtDNA and the nuclear transcription factor 1 (NRF-1). Both transcription factors are essential for replication of mtDNA. Consistent with this notion, low levels of mitochondrial TFAM (mtTFAM) are associated with mtDNA depletion and several severe mtDNA diseases. Targeted disruption of mouse TFAM or NRF-1 genes cause lethal mutations resulting in depletion of mtDNA during embryonic stages. Binding of NRF-1 is important for mtDNA activity *in vitro*, and its disruption leads to reduced expression of mtTFAM and mtDNA copy number. Therefore, studying mRNA profiles relevant to mitochondrial transcription factors such as TFAM and NRF-1 is necessary for understanding replication and transcription of the mitochondrial genome as well as other aspects of mitochondrial biogenesis. The need to understand the close harmony and complex mechanisms that exist between nuclear and mitochondrial genomes is important in the context of development of more aggressive forms of ART such as cloning and ooplasmic transfer. Apart from the problems that arise from genetic reprogramming of paternal and maternal imprinted genes, the requirement for strict concordance between the two genomes may lead to problems after nuclear transfer since we must consider the specific mitochondrial haplotype of the

recipient cells in relation to the inserted nuclei.

Maternally inherited mtDNA defects can result in debilitating or lethal conditions for affected individuals (so-called mitochondrial genetic diseases) if genes coding for respiratory chain enzymes are involved (OXPHOS diseases). For the oocyte and early embryo, it has been proposed that mutations in the mitochondrial genome may contribute to impaired function and developmental incompetence. The extent to which specific mtDNA defects are related to early reproductive failure is not clear, but genetic mutations may result in diminished ATP content leading to embryonic abnormalities such as cleavage arrest, slow development, blastomere loss by apoptosis, and, ultimately, implantation failure. As described below, the mitochondrial complement in human oocytes can be in excess of 150 000 organelles, so any potential adverse affect of mtDNA mutations associated with respiratory function would depend upon the magnitude of the mutant population (mutant load). It has also been suggested that embryonic failures could be a consequence of asymmetrical mitochondrial distribution in the pronuclear embryo. This could lead to disproportionate mitochondrial inheritance and perhaps produce blastomeres with diminished ATP generating capacity that could reduce developmental competence. Although it is generally assumed that mitochondria are critical determinants of competence, it is not known to what extent differences in early human embryo development observed within and between cohorts are associated with mitochondrial distribution and metabolic capabilities.

It has long been thought that the transmission of mitochondria between generations is strictly maternal. However, the development of more invasive techniques of assisted reproduction has increased concern that the mechanisms of uniparental (maternal) mitochondrial inheritance may be altered and as a result, the genetic integrity of the offspring could be compromised. For example, there is evidence of the persistence of human paternal mtDNA in "abnormal" unfertilized oocytes and embryos. Whether human sperm mtDNA is entirely eliminated or whether certain embryos are incapable of destroying sperm mitochondria is unknown. It is unclear whether the detection of paternal mtDNA sequences at extremely low levels is of any biological significance or simply coincidental and unrelated to embryo competence. With current experimental efforts designed to evaluate the potential therapeutic applications of nuclear and cytoplasmic transplantation for curing inheritable mitochondrial disease or known chromosomal abnormalities, it is necessary to understand the role(s) of mitochondria and the mechanisms that regulate their function in early mammalian embryogenesis. An understanding of the regulation and function of mitochondria in early human development becomes especially relevant in clinical IVF due to recent reports of three children born after interooplasmic transfer (insertion of a small amount of cytoplasm from an oocyte of a young woman into the oocyte of an older woman). In these children, a significant expansion of donor mtDNA has occurred, resulting in two genetically distinct populations of mtDNA coexisting within the same cytoplasm, a condition known as heteroplasmy.

The issue of concordance between the mitochondrial and nuclear genomes must be considered in light of the developmental

paradox regarding the transmission of paternal mitochondrial DNA, or lack thereof, in mammals. The fertilizing spermatozoon introduces up to 100 functional mitochondria into the oocyte cytoplasm at the time of insemination. However, these paternal mitochondria typically do not persist beyond early embryonic stages. It is presumed that there is mandatory destruction of the sperm mitochondria that appears to be evolutionarily and developmentally advantageous because the paternal mitochondria and their DNA may be severely compromised because of high levels of mitochondrial deletions and point mutations generated by the action of reactive oxygen species created by the spermatozoa during spermatogenesis and fertilization. Therefore, it is believed that mitochondria are inherited via strict maternal pathways while sperm mitochondria are selectively destroyed during fertilization and early embryogenesis and/or diluted below detection threshold levels with maternal mitochondria. Although a number of studies support the notion that sperm mitochondria are selectively destroyed by the egg, the actual molecular and cellular mechanisms are not known. Sutovsky et al. suggested that prohibitin could be one of the substrates that mark sperm mitochondria for elimination by proteolytic enzymes within the oocyte. In mice, paternal transmission of mtDNA has been observed in first-generation offspring resulting from interspecific crosses, and the persistence of paternal mtDNA has been detected in some abnormal human blastocysts and polyploid embryos. Despite the incompetence of these embryos, it may be relevant to ask what are the cytoplasmic conditions that allow the persistence of paternal mitochondria in order to further our understanding of normal regulation of mitochondrial transmission. Another suggestion to account for the absence of paternal mtDNA in human offspring is segregation of paternal mitochondria into extraembryonic tissues that develop into the placenta. However, no viable hypothesis has been advanced to explain how such a selective mechanism could operate. While transmission of paternal mitochondria may be more common in "poor quality" oocytes and embryos, it remains to be determined whether detection of paternal mtDNA could be a useful marker of embryo competence.

Although it is generally assumed that paternal mtDNA transmission does not occur after normal fertilization, the use of more invasive techniques in clinical IVF such as ICSI, ooplasm transfer, and nuclear transfer may alter the strict maternal inheritance of mtDNA. Consider the case of a 28-year-old man with severe mitochondrial myopathy due to a novel 2-bp mtDNA deletion in the *mND2* gene that encodes a subunit of the enzyme complex of the mitochondrial respiratory chain. The mtDNA harboring this mutation was paternal in origin and accounted for 90% of the patient's muscle mtDNA. The occurrence of paternal mtDNA in this patient may be an exceptional or unique situation resulting from the survival of some sperm mitochondria that ordinarily would never have been recognized had the pathogenic mutation not conferred an apparent selective proliferative advantage. While this case is assumed to be atypical, it is also apparent that underlying mechanisms responsible for elimination of sperm mtDNA in normal embryos are currently not well understood. As a consequence, further investigation of paternal mtDNA transmission and inheritance is critical if certain methods of nuclear or cytoplasmic transfer are to be used in assisted reproduction with any assurance

of safety. Furthermore, there is increasing evidence that mtDNA anomalies in spermatozoa may be associated with infertility. For example, mitochondrial point mutations and deletions have been correlated with poor sperm quality. It is unknown whether mitochondria of such spermatozoa, if introduced into an oocyte by ICSI, are eliminated in the same way as in their normal counterparts. If not, it is an open question as to whether mitochondrial diseases could be unknowingly transmitted by this method of fertilization.

4.2 Mitochondrial Inheritance following Ooplasmic Transplantation

As new technologies have been introduced and offered to patients seeking assisted reproduction, there is always considerable debate regarding the safety and efficacy surrounding these procedures. For example, the use of ooplasmic transfer has generated much attention. Ooplasmic transfer was developed to treat infertility patients with persistent poor embryonic development and recurrent implantation failure after IVF. In this experimental procedure, 5 to 15% of the ooplasm from a presumably fertile donor oocyte is microinjected into a compromised recipient oocyte. Transfer of small amounts of donor ooplasm introduces cytoplasmic proteins, mRNAs, small molecules, mitochondria, and other organelles into the recipient oocytes. Although the donor mtDNA may play no role in rescuing poor quality oocytes, ooplasmic transplantation clearly generates a mixture of mtDNA from two different parental origins (heteroplasmy) in any resulting embryos.

In two reports, relative proportions of donor and recipient mitochondria were assayed in embryos, amniocytes, fetal tissues, and blood using real-time qualitative fluorescent DNA sequencing. Donor mtDNA was detected in samples with the following frequencies: embryos 59.3%, amniocytes 22.2%, and fetal cord blood 25%. Recently, scientific attention has focused on the fact that three children produced by ooplasmic transplantation possess mitochondrial genetic material from two maternal sources, suggesting that donor mitochondrial populations not only persist but replicate in the developing children. Since the mitochondrial fingerprinting technique only qualitatively evaluates the presence of mtDNA heteroplasmy in these children, a sensitive, quantitative, allele-specific molecular beacon analysis was employed to document these changes accurately. The quantification of donor mitochondria in samples from three children using this method has shown that the donor mtDNA level is over 30% in the blood of these children and differs from the levels found in fetal cord blood and placenta in each case. Abandoning ooplasmic transfer as well as nuclear transfer in human IVF clinics has left many unanswered questions. Therefore, experimental studies on mitochondrial inheritance using nonhuman primates may be the best way to answer important questions concerning mitochondrial transmission, such as: (1) What is the normal level of heteroplasmy in primates? (2) Is there strictly uniparental, maternal mitochondrial inheritance using the IVF procedure? Nonhuman primates can answer questions concerning triparental inheritance whether the source is from paternal, maternal, or donor nuclei since we can manipulate oocytes and spermatozoa both inter- and intraspecifically within a controlled experimental environment. (3) How specific is

the elimination of the paternal genome via the ubiquitination process? (4) What changes occur in mitochondrial transmission after more aggressive assisted reproductive techniques such as nuclear transfer using blastomeres and eventually somatic cells as the donor cell source? This leads us to examine the fundamental problems associated with sophisticated ART procedures and compel us to revisit the issues of mtDNA transmission, the role of mitochondria in postfertilization development and more specifically, nuclear-mitochondrial communication.

5 Summary and Conclusions

The molecular basis of fertilization is still poorly understood. It is impractical to examine fertilization *in vivo*, so IVF is the best way to provide information, bearing in mind that events may not be identical to those occurring *in vivo*. Fortunately, the technology for achieving IVF is well developed for many species including humans and other primates, so we can address key questions about mechanisms and molecules. The advent of sensitive techniques such as RT-PCR and molecular beacon analysis compensates for the small amounts of material (oocytes and embryos) that are available. In particular, the normal events occurring at the molecular level during fertilization need further elucidation not only to further our knowledge but also so that any anomalies introduced by advanced ART can be analyzed and appropriate measures taken to avoid them. This should be the focus of future research investigations on mammalian fertilization.

See also Cellular Interactions; Developmental Cell Biology.

Bibliography

Books and Reviews

Cosson, M.P. (1990) Sperm chemotaxis, in: Gagnon, C. (Ed.) *Control of Sperm Motility: Biological and Clinical Aspects*, CRC Press, Boca Raton, FL, pp. 103–135.

Dunbar, B.S., Prasad, S.V., Timmons, T.M. (1991) Comparative Structure and Function of Mammalian Zonae Pellucidae, in: Dunbar, B.S., O'Rand, M.G. (Eds.) *A Comparative Overview of Mammalian Fertilization*, Plenum Press, New York, pp. 97–114.

Yanagimachi, R. (1994) Mammalian fertilization, in: Knobil, E., Neill, J.D. (Eds.) *The Physiology of Reproduction*, 2nd. edition, Raven Press, New York, pp. 189–317.

Primary Literature

Alam, T.I., Kanki, K., Muta, T., Ukaji, K., Abe, Y., Nakayama, H., Takio, K., Hamasaki, N., Kang, D. (2003) Human mitochondrial DNA is packaged with TFAM, *Nucleic Acid Res.* **6**, 1640–1645.

Almeida, E.A., Huovila, A.P., Sutherland, A.E., Stephens, L.E., Calarco, P.G., Shaw, L.M., Mercurio, A.M., Sonnenberg, A., Primakoff, P., Myles, D.G., White, J.M. (1995) Mouse egg integrin α_6ß$_1$ functions as a sperm receptor, *Cell* **81**, 1095–1104.

Aoki, F., Sakai, S., Kohmoto, K. (1999) Regulation of flagellar bending by cAMP and Ca^{2+} in hamster sperm, *Mol. Reprod. Dev.* **53**, 77–83.

Asch, R., Simerly, C., Ord, T., Ord, V.A., Schatten, G. (1995) The stages at which human fertilization arrests: microtubule and chromosome configurations in inseminated oocytes which failed to complete fertilization and development in humans, *Hum. Reprod.* **10**, 1897–1906.

Austin, C.R. (1951) Observations on the penetration of the sperm in the mammalian egg, *Aust. J. Sci. Res.* **B4**, 581–589.

Balmaceda, J.P., Pool, T.B., Arana, J.B., Heitman, T.S., Asch, R.H. (1984) Successful

in vitro fertilization and embryo transfer in cynomolgus monkeys, *Fertil. Steril.* **42**, 791–795.

Barnett, D.K., Bavister, B.D. (1996) What is the relationship between the metabolism of preimplantation embryos and their development in vitro? *Mol. Reprod. Dev.* **43**, 105–133.

Barritt, J.A., Brenner, C.B., Willadsen, S., Cohen, J. (2000) Spontaneous and artificial changes in human ooplasmic mitochondria, *Hum. Reprod.* **15**, 207–217.

Barritt, J.A., Brenner, C.B., Malter, H., Cohen, J. (2001) Mitochondria in human offspring derived from ooplasmic transplantation, *Hum. Reprod.* **16**, 513–516.

Bavister, B.D. (1995) Culture of preimplantation embryos: facts and artifacts, *Hum. Reprod. Update* **1**, 91–148.

Bavister, B.D. (2002a) Early history of in vitro fertilization, *Reproduction* **124**, 181–196.

Bavister, B.D. (2002b) How animal embryo research led to the first documented human IVF, *Reprod. Biomed. Online* **4**(Suppl. 1), 24–29.

Bavister, B.D., Boatman, D.E., Leibfried, L., Loose, M., Vernon, M.W. (1983) Fertilization and cleavage of rhesus monkey oocytes in vitro, *Biol. Reprod.* **28**, 983–999.

Bavister, B.D., Boatman, D.E., Collins, K., Dierschke, D.J., Eisele, S.G. (1984) Birth of rhesus monkey infant after in vitro fertilization and nonsurgical embryo transfer, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2218–2222.

Bleil, J.D., Wassarman, P.M. (1981) Mammalian sperm-egg interaction: fertilization of mouse eggs triggers modification of the major zona pellucida glycoprotein, ZP2, *Dev. Biol.* **86**, 189–197.

Bleil, J.D., Wassarman, P.M. (1983) Sperm-egg interactions in the mouse: sequence of events and induction of the acrosome reaction by a zona pellucida glycoprotein, *Dev. Biol.* **95**, 317–324.

Bleil, J.D., Wassarman, P.M. (1988) Galactose at the nonreducing terminus of O-linked oligosaccharides of mouse egg zona pellucida glycoprotein ZP3 is essential for the glycoprotein's sperm receptor activity, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6778–6782.

Bleil, J.D., Greve, J.M., Wassarman, P.M. (1988) Identification of a secondary sperm receptor in the mouse egg zona pellucida: role in maintenance of binding of acrosome-reacted sperm to eggs, *Dev. Biol.* **128**, 376–385.

Blobel, C.P., Wolfsberg, T.G., Turck, C.W., Myles, D.G., Primakoff, P., White, J.M. (1992) A potential fusion peptide and an integrin ligand domain in a protein active in sperm-egg fusion, *Nature* **356**, 248–252.

Boatman, D.E., Bavister, B.D. (1984) Stimulation of rhesus monkey sperm capacitation by cyclic nucleotide mediators, *J. Reprod. Fertil.* **71**, 357–366.

Boatman, D.E., Robbins, R.S. (1991) Bicarbonate: carbon-dioxide regulation of sperm capacitation, hyperactivated motility, and acrosome reactions, *Biol. Reprod.* **44**, 806–813.

Brenner, C.B., Barritt, J.A., Willadsen, S., Cohen, J. (2000) Mitochondrial DNA heteroplasmy after human ooplasmic transplantation, *Fertil. Steril.* **74**, 573–578.

Bronson, R.A., Fusi, F.M. (1996) Integrins and human reproduction, *Mol. Hum. Reprod.* **2**, 153–168.

Chang, M.C. (1951) Fertilizing capacity of spermatozoa deposited into the fallopian tubes, *Nature* **168**, 697.

Chen, H., Sampson, N.S. (1999) Mediation of sperm-egg fusion: evidence that mouse egg α_6ß$_1$ integrin is the receptor for sperm fertilin ß, *Chem. Biol.* **6**, 1–10.

Cherr, G.N., Meyers, S.A., Yudin, A.I., VandeVoort, C.A., Myles, D.G., Primakoff, P., Overstreet, J.W. (1996) The PH-20 protein in cynomolgus macaque spermatozoa: identification of two different forms exhibiting hyaluronidase activity, *Dev. Biol.* **175**, 142–153.

Cho, C., Bunch, D.O., Faure, J.E., Goulding, E.H., Eddy, E.M., Primakoff, P., Myles, D.G. (1988) Fertilization defects in sperm from mice lacking fertilin beta, *Science* **281**, 1857–1859.

Cohen, J., Scott, R., Schimmel, T., Willadsen, S. (1997) Birth of infant after transfer of anucleate donor oocyte cytoplasm into recipient eggs, *Lancet* **350**, 186–187.

Cohen, J., Scott, R., Alikani, M., Schimmel, T., Munne, S., Levron, J., Wu, L., Brenner, C., Warner, C., Willadsen, S. (1998) Ooplasmic transfer in mature human oocytes, *Mol. Hum. Reprod.* **4**, 269–280.

Cohen-Dayag, A., Tur-Kaspa, Y., Dor, J., Mashiach, S., Eisenbach, M. (1995) Sperm capacitation in humans is transient and correlates with chemotactic responsiveness to

follicular factors, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11039–11043.

Colledge, W.H., Carlton, M.B., Udy, G.B., Evans, M.J. (1994) Disruption of c-mos causes parthenogenetic development of unfertilized mouse eggs, *Nature* **370**, 65–68.

Cranfield, M.R., Schaffer, N., Bavister, B.D., Berger, N., Boatman, D.E., Kempske, S., Miner, N., Panos, M., Adams, J., Morgan, P.M. (1989) Assessment of oocytes retrieved from stimulated and unstimulated ovaries of pig-tailed macaques (Macaca nemestrina) as a model to enhance the genetic diversity of captive lion-tailed macaques (Macaca silenus), *Zool. Biol.* (Suppl. 1), 33–46.

Cummins, J. (2001) Mitochondria: potential roles in embryogenesis and nucleocytoplasmic transfer, *Hum. Reprod. Update* **7**, 217–228.

Cummins, J. (2002) The role of maternal mitochondria during oogenesis, fertilization and embryogenesis, *Reprod. BioMed. Online* **4**, 176–182.

Doherty, A.S., Mann, M.R., Tremblay, K.D., Bartolomei, M.S., Schultz, R.M. (2000) Differential effects of culture on imprinted H19 expression in the preimplantation mouse embryo, *Biol. Reprod.* **62**, 1526–1535.

Ducibella, T., Huneau, D., Angelichio, E., Xu, Z., Schultz, R., Kopf, G., Fissore, R., Madoux, S., Ozil, J.-P. (2002) Egg-to-embryo transition is driven by differential responses to Ca(2+) oscillation number, *Dev. Biol.* **250**, 280–291.

Eberspaecher, U., Becker, A., Bringmann, P., van der Merwe, L., Donner, P. (2001) Immunohistochemical localization of zona pellucida proteins ZPA, ZPB and ZPC in human, cynomolgus monkey and mouse ovaries, *Cell Tissue Res.* **303**, 277–287.

Edwards, R.G., Bavister, B.D., Steptoe, P.C. (1969) Early stages of fertilization in vitro of human oocytes matured in vitro, *Nature* **221**, 632–635.

Evans, J.P. (1999) Sperm disintegrins, egg integrins, and other cell adhesion molecules of mammalian gamete plasma membrane interactions, *Front. Biosci.* **4**, D114–D131.

Evans, J.P. (2001) Fertilin beta and other ADAMs as integrin ligands: insights into cell adhesion and fertilization, *BioEssays* **23**, 628–639.

Evans, J.P., Kopf, G.S., Schultz, R.M. (1997) Characterization of the binding of recombinant mouse sperm fertilin ß subunit to mouse eggs: evidence for adhesive activity via an egg $ß_1$ integrin-mediated interaction, *Dev. Biol.* **187**, 79–93.

Florman, H.M., Wassarman, P.M. (1985) O-linked oligosaccharides of mouse egg ZP3 account for its sperm receptor activity, *Cell* **41**, 313–324.

Florman, H.M., Arnoult, C., Kazam, I.G., Li, C., O'Toole, C.M.B. (1998) A perspective on the control of mammalian fertilization by egg-activated ion channels in sperm: a tale of two channels, *Biol. Reprod.* **59**, 12–16.

Gahlay, G.K., Srivastava, N., Govind, C.K., Gupta, S.K. (2002) Primate recombinant zona pellucida proteins expressed in Escherichia coli bind to spermatozoa, *J. Reprod. Immunol.* **53**, 67–77.

Giles, R.E., Blanc, H., Cann, H.M., Wallace, D.C. (1980) Maternal inheritance of human mitochondrial DNA, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6715–6719.

Grzybowski, T., Malyarchuk, B.A., Czarny, J., Miscicka-Sliwka, D., Kotzbach, R. (2003) High levels of mitochondrial DNA heteroplasmy in single hair roots: reanalysis and revision, *Electrophoresis* **24**, 1159–1165.

Gould, K.G. (1983) Ovum recovery and in vitro fertilization in the chimpanzee, *Fertil. Steril.* **40**, 378–383.

Gould, K.G., Cline, E.M., Williams, W.L. (1973) Observations on the induction of ovulation and fertilization in vitro in the squirrel monkey (Saimiri sciureus), *Fertil. Steril.* **24**, 260–268.

Hardy, C.M., Holland, M.K. (1996) Cloning and expression of recombinant rabbit fertilin, *Mol. Reprod. Dev.* **45**, 107–116.

Harris, J.D., Piersen, C.E. (2003) Cloning and expression of cynomolgus monkey and baboon zona pellucida proteins, *Mol. Reprod. Dev.* **65**, 237–244.

Hartmann, J.F., Gwatkin, R.B., Hutchison, C.F. (1972) Early contact interactions between mammalian gametes in vitro: evidence that the vitellus influences adherence between sperm and zona pellucida, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 2767–2769.

Hewitson, L., Schatten, G. (2002) The use of primates as models for assisted reproduction, *Reprod. Biomed. Online* **5**, 50–55.

Ho, H.-C., Suarez, S.S. (2001) An inositol 1,4,5-trisphosphate receptor-gated intracellular Ca(2+) store is involved in regulating sperm hyperactivated motility, *Biol. Reprod.* **65**, 1606–1615.

Ho, H.-C., Granish, K.A., Suarez, S.S. (2002) Hyperactivated motility of bull sperm is triggered at the axoneme by Ca(2+) and not cAMP, *Dev. Biol.* **250**, 208–217.

Hubner, K., Fuhrmann, G., Christenson, L.K., Kehler, J., Reinbold, R., De La Fuente, R., Wood, J., Strauss, J.F. 3rd, Boiani, M., Scholer, H.R. (2003) Derivation of oocytes from mouse embryonic stem cells, *Science* **300**, 1251–1256.

Irsigler, U.M., Van der Merwe, J.V., Botes, A.D. (1984) Fertilization and cleavage in vitro of baboon oocytes, *S. Afr. Med. J.* **66**, 447–450.

Jaenisch, R., Wilmut, I. (2001) Don't clone humans! *Science* **291**, 2552.

Jansen, S., Ekhlasi-Hundrieser, M., Topfer-Petersen, E. (2001) Sperm adhesion molecules: structure and function, *Cells Tissue Organs* **168**, 82–92.

Johnston, D.S., Wright, W.W., Shaper, J.H., Hokke, C.H., Van den Eijnden, D.H., Joziasse, D.H. (1998) Murine sperm-zona binding, a fucosyl residue is required for a high affinity sperm-binding ligand. A second site on sperm binds a nonfucosylated, beta-galactosyl-capped oligosaccharide, *J. Biol. Chem.* **273**, 1888–1895.

Jury, J.A., Frayne, J., Hall, L. (1997) The human fertilin αgene is nonfunctional: implications for its proposed role in fertilization, *Biochem. J.* **321**, 577–581.

Keyon, L., Moraes, C.T. (1977) Expanding the functional human mitochondrial DNA database by establishment of primate xenomitochondrial hybrids, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9131–9135.

Kraemer, D.C., Flow, B.L., Schriver, M.D., Kinney, G.M., Pennycook, J.W. (1979) Embryo transfer in the nonhuman primate, feline and canine, *Theriogenology* **11**, 51–62.

Kuehl, T.J., Dukelow, W.R. (1975) Fertilization in vitro of Saimiri sciureus follicular oocytes, *J. Med. Primatol.* **4**, 209–216.

Kurokawa, M., Fissore, R.A. (2003) ICSI-generated mouse zygotes exhibit altered calcium oscillations, inositol 1,4,5-trisphosphate receptor-1 down-regulation, and embryo development, *Mol. Hum. Reprod.* **9**, 523–533.

Lanzendorf, S.E., Zelinski-Wooten, M.B., Stouffer, R.L., Wolf, D.P. (1990) Maturity at collection and the developmental potential of rhesus monkey oocytes, *Biol. Reprod.* **42**, 703–711.

Larsson, N.G., Wang, J., Wilhelmsson, H., Oldfors, A., Rustin, P., Lewandoski, M., Barsh, G.S., Clayton, D.A. (1998) Mitochondrial transcription factor A is necessary for mtDNA maintenance and embryogenesis in mice, *Nat. Genet.* **18**, 231–236.

Lawrence, Y., Ozil, J.-P., Swann, K. (1998) The effects of a Ca^{2+} chelator and heavy-metal-ion chelators upon Ca^{2+} oscillations and activation at fertilization in mouse eggs suggest a role for repetitive Ca^{2+} increases, *Biochem. J.* **335**, 335–342.

Lax, Y., Rubinstein, S., Breitbart, H. (2004) Acrosin activity assay for the evaluation of Mammalian sperm acrosome reaction, *Methods Mol. Biol.* **253**, 135–140.

Lin, Y., Kimmel, L.H., Myles, D.G., Primakoff, P. (1993) Molecular cloning of the human and monkey sperm surface protein PH-20, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10071–10075.

Lin, Y., Mahan, K., Lathrop, W.F., Myles, D.G., Primakoff, P. (1994) A hyaluronidase activity of the sperm plasma membrane protein PH-20 enables sperm to penetrate the cumulus cell layer surrounding the egg, *J. Cell Biol.* **125**, 1157–1163.

Lopata, A., Summers, P.M., Hearn, J.P. (1988) Births following the transfer of cultured embryos obtained by in vitro and in vivo fertilization in the marmoset monkey (Callithrix jacchus), *Fertil. Steril.* **50**, 503–509.

Lu, Q., Shur, B.D. (1997) Sperm from beta 1,4-galactosyltransferase-null mice are refractory to ZP3-induced acrosome reactions and penetrate the zona pellucida poorly, *Development* **124**, 4121–4131.

Lum, L., Blobel, C.P. (1997) Evidence for distinct serine protease activities with a potential role in processing the sperm protein fertilin, *Dev. Biol.* **191**, 131–145.

Mahony, M.C., Gwathmey, T. (1999) Protein tyrosine phosphorylation during hyperactivated motility of cynomolgus monkey (Macaca fascicularis) spermatozoa, *Biol. Reprod.* **60**, 1239–1243.

Makler, A., Stoller, J., Reichler, A., Blumenfeld, Z., Yoffe, N. (1995) Inability of human sperm to change their orientation in response to external chemical stimuli, *Fertil. Steril.* **63**, 1077–1082.

Markoulaki, S., Matson, S., Abbott, A.L., Ducibella, T. (2003) Oscillatory CaMKII activity in mouse egg activation, *Dev. Biol.* **258**, 464–474.

May-Panloup, P., Chretien, M., Savagner, F., Vasseur, C., Jean, M., Malthiery, Y., Reynier, P.

(2003) Increased sperm mitochondrial DNA content in male infertility, *Hum. Reprod.* **18**, 550–556.

McKiernan, S.H., Bavister, B.D. (1998) Gonadotropin stimulation of donor females decreases post-implantation viability of cultured 1-cell hamster embryos, *Hum. Reprod.* **13**, 724–729.

Meng, L., Wolf, D.P. (1997) Sperm-induced oocyte activation in the rhesus monkey: nuclear and cytoplasmic changes following intracytoplasmic sperm injection, *Hum. Reprod.* **12**, 1062–1068.

Mitalipov, S.M., Yeoman, R.R., Nusser, K.D., Wolf, D. (2002) Rhesus monkeys produced by nuclear transfer from embryonic or somatic cells, *Biol. Reprod.* **66**, 1367–1373.

Miller, D.J., Macek, M.B., Shur, B.D. (1992) Complementarity between sperm surface beta-1,4-galactosyltransferase and egg-coat ZP3 mediates sperm-egg binding, *Nature* **357**, 589–593.

Miller, B.J., Georges-Labouesse, E., Primakoff, P., Myles, D.G. (2000) Normal fertilization occurs with eggs lacking the integrin $\alpha_6ß_1$ and is CD9-dependent, *J. Cell Biol.* **149**, 1289–1296.

Moos, J., Visconti, P.E., Moore, G.D., Schultz, R.M., Kopf, G.S. (1995) Potential role of mitogen-activated protein kinase in pronuclear envelope assembly and disassembly following fertilization of mouse eggs, *Biol. Reprod.* **53**, 692–699.

Myles, D.G., Kimmel, L.H., Blobel, C.P., White, J.M., Primakoff, P. (1994) Identification of a binding site in the disintegrin domain of fertilin required for sperm-egg fusion, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4195–4198.

Niemann, H., Wrenzycki, C. (2000) Alterations of expression of developmentally important genes in preimplantation bovine embryos by in vitro culture conditions: implications for subsequent development, *Theriogenology* **53**, 21–34.

Nishimura, H., Cho, C., Branciforte, D.R., Myles, D.G., Primakoff, P. (2001) Analysis of loss of adhesive function in sperm lacking cyritestin or fertilin beta, *Dev. Biol.* **233**, 204–213.

Nixon, V., Levasseur, M., McDougall, A., Jones, K. (2002) Ca(2+) oscillations promote APC/C-dependent cyclin B1 degradation during metaphase arrest and completion of meiosis in fertilizing mouse eggs, *Curr. Biol.* **12**, 746–750.

Oehninger, S. (2003) Biochemical and functional characterization of the human zona pellucida, *Reprod. Biomed. Online* **7**, 641–648.

Pal, S.K., Torry, D., Serta, R., Crowell, R.C., Seibel, M.M., Cooper, G.M., Kiessling, A.A. (1994) Expression and potential function of the c-mos proto-oncogene in human eggs, *Fertil. Steril.* **61**, 496–503.

Palermo, G., Munne, S., Cohen, J. (1994) The human zygote inherits its mitotic potential from the male gamete, *Hum. Reprod.* **9**, 1220–1225.

Parmentier, M., Libert, F., Schurmans, S., Schiffmann, S., Lefort, A., Eggerickx, D., Ledent, C., Mollereau, C., Gerard, C., Perret, J. (1992) Expression of members of the putative olfactory receptor gene family in mammalian germ cells, *Nature* **355**, 453–455.

Paterson, M., Wilson, M.R., Morris, K.D., van Duin, M., Aitken, R.J. (1998) Evaluation of the contraceptive potential of recombinant human ZP3 and human ZP3 peptides in a primate model: their safety and efficacy, *Am. J. Reprod. Immunol.* **40**, 198–209.

Perry, A.C., Gichuhi, P.M., Jones, R., Hall, L. (1995) Cloning and analysis of monkey fertilin reveals novel α subunit isoforms, *Biochem. J.* **307**, 843–850.

Pierce, K.E., Rice, J.E., Sanchez, J.A., Brenner, C., Wangh, L.J. (2000) Real-time PCR using molecular beacons for accurate detection of the Y chromosome in single human blastomeres, *Mol. Hum. Reprod.* **6**, 1155–1164.

Pope, C.E., Dresser, B.L., Chin, N.W., Lui, J.H., Loskutoff, N.M., Behnke, E.J., Brown, C., McRae, M.A., Sinoway, C.E., Campbell, M.K., Cameron, K.N., Owens, O.M., Johnson, C.A., Evans, R.R., Cedars, M.I. (1997) Birth of a western lowland gorilla (Gorilla gorilla gorilla) following in vitro fertilization and embryo transfer, *Am. J. Primatol.* **41**, 247–260.

Poueymirou, W.T., Conover, J.C., Schultz, R.M. (1989) Regulation of mouse preimplantation development: differential effects of CZB medium and Whitten's medium on rates and patterns of protein synthesis in 2-cell embryos, *Biol. Reprod.* **41**, 317–322.

Primakoff, P., Myles, D.G. (2000) The ADAM gene family: surface proteins with adhesion and protease activity, *Trends Genet.* **16**, 83–87.

Primakoff, P., Hyatt, H., Tredick-Kline, J. (1987) Identification and purification of a sperm

surface protein with a potential role in sperm-egg membrane fusion, *J. Cell Biol.* **104**, 141–149.

Ralt, D., Goldenberg, M., Fetterolf, F.P., Thompson, D., Dor, J., Mashiach, S., Garbers, D.L., Eisenbach, M. (1991) Sperm attraction to a follicular factor(s) correlates with human egg fertilizability, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2840–2844.

Reimann, J.D., Jackson, P.K. (2002) Emi1 is required for cytostatic factor arrest in vertebrate eggs, *Nature* **416**, 850–854.

Ryder, T.A., Mobberley, M.A., Hughes, L. (1990) A survey of the ultrastructural defects associated with absent or impaired human sperm motility, *Fertil. Steril.* **53**, 556–560.

Saling, P.M. (1991) How the egg regulates sperm function during gamete interaction: facts and fantasies, *Biol. Reprod.* **44**, 246–251.

Sathananthan, A.H., Kola, I., Osborne, J., Trounson, A., Ng, S.C., Bongso, A., Ratnam, S.S. (1991) Centrioles in the beginning of human development, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4806–4810.

Schatten, G. (1994) The centrosome and its mode of inheritance: the reduction of the centrosome during gametogenesis and its restoration during fertilization, *Dev. Biol.* **165**, 299–335.

Schwartz, M., Vissing, J. (2002) Paternal inheritance of mitochondrial DNA, *N. Engl. J. Med.* **347**, 576–579.

Schwartz, R., Brook, W., Zinsser, H.H. (1958) Evidence of chemotaxis as a factor in sperm motility, *Fertil. Steril.* **9**, 300–308.

Spehr, M., Gisslemann, G., Poplawski, A., Riffell, J., Wetzel, C.H., Zimmer, R.K., Hatt, H. (2003) Identification of a testicular odorant receptor mediating human sperm chemotaxis, *Science* **299**, 2054–2057.

Squirrell, J.M., Lane, M., Bavister, B.D. (2001) Altering intracellular pH disrupts development and cellular organization in preimplantation hamster embryos, *Biol. Reprod.* **64**, 1845–1854.

Squirrell, J.M., Schramm, R.D., Paprocki, A.M., Wokosin, D.L., Bavister, B.D. (2003) Imaging mitochondrial organization in living primate oocytes and embryos using multiphoton microscopy, *Microsc. Microanal.* **9**, 190–201.

Srivasta, P.N., Farooqui, A.S., Gould, K.G. (1981) Studies on hydrolytic enzymes of chimpanzee semen, *Biol. Reprod.* **25**, 363–369.

Stambaugh, R., Buckley, J. (1970) Comparative studies of the acrosomal enzymes of rabbit, rhesus monkey, and human spermatozoa, *Biol. Reprod.* **3**, 275–282.

Stambaugh, R., Seitz, H.M., Mastroianni, L. (1974) Acrosomal proteinase inhibitors in rhesus monkey (Macaca mulatta) oviduct fluid, *Fertil. Steril.* **25**, 352–357.

St. John, J.C. (2002) The transmission of mitochondrial DNA following assisted reproductive techniques, *Theriogenology* **57**, 109–123.

St. John, J.C., De Jonge, C. (2000) A hypothesis for transmission of paternal mitochondrial DNA, *Reprod. Med. Rev.* **8**, 173–185.

St. John, J.C., Cooke, I.D., Barratt, C.L. (1997) Mitochondrial mutations and male infertility, *Nat. Med.* **2**, 1065–1068.

St. John, J.C., Sakkas, D., Dimitriadi, K., Barnes, A., Maclin, V., Ramey, J., Barratt, C., De Jonge, C. (2000) Failure of elimination of paternal mitochondrial DNA in abnormal embryos, *Lancet* **15**, 199–200.

Sutovsky, P. (2003) Ubiquitin-dependent proteolysis in mammalian spermatogenesis, fertilization, and sperm quality control: killing three birds with one stone, *Microsc. Res. Tech.* **61**, 88–102.

Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., Dominko, T., Simerly, C., Schatten, G. (1999) Ubiquitin tag for sperm mitochondria, *Nature* **402**, 371–372.

Sutovsky, P., Moreno, R., Ramalho-Santos, J., Dominko, T., Simerly, C., Schatten, G. (2000) Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos, *Biol. Reprod.* **63**, 582–590.

Talbot, P., DiCarantonio, G., Zao, P., Penkala, J., Haimo, L.T. (1985) Motile cells lacking hyaluronidase can penetrate the hamster oocyte cumulus complex, *Dev. Biol.* **108**, 387–398.

Tesarik, J., Drahorad, J., Testart, J., Mendoza, C. (1990) Acrosin activation follows its surface exposure and precedes membrane fusion in human sperm acrosome reaction, *Development* **110**, 391–400.

Tollner, T.L., Yudin, A.I., Cherr, G.N., Overstreet, J.W. (2003) Real-time observations of individual macaque sperm undergoing tight binding and the acrosome reaction on the zona pellucida, *Biol. Reprod.* **68**, 664–672.

Tulsiani, T.R.P., Nagdas, S.K., Cornwall, G.A., Orgebin-Crist, M.C. (1992) Evidence for the presence of high-mannose/hybrid oligosaccharide chain(s) on the mouse ZP2 and ZP3, *Biol. Reprod.* **46**, 93–100.

Van Blerkom, J., Davis, P., Alexander, S. (2000) Differential mitochondrial distribution in human pronuclear embryos leads to disproportionate inheritance between blastomeres: relationship to microtubular organization, ATP content and competence, *Hum. Reprod.* **15**, 2621–2633.

Van Blerkom, J., Sinclair, J., Davis, P. (1998) Mitochondrial transfer between oocytes: potential applications of mitochondrial donation and the issue of heteroplasmy, *Hum. Reprod.* **13**, 2857–2868.

Vanderhaeghen, P., Schurmans, S., Vassart, G., Parmentier, M. (1993) Olfactory receptors are displayed on dog mature sperm cell, *J. Cell Biol.* **123**, 1441–1452.

VandeVoort, C.A., Tollner, T.L., Overstreet, J.W. (1992) Sperm-zona pellucida interaction in cynomolgus and rhesus macaques, *J. Androl.* **13**, 428–432.

VandeVoort, C.A., Tollner, T.L., Overstreet, J.W. (1994) Separate effects of caffeine and dbcAMP on macaque sperm motility and interaction with the zona pellucida, *Mol. Reprod. Dev.* **37**, 299–304.

Waters, S.I., White, J.M. (1997) Biochemical and molecular characterization of bovine fertilin α and ß (ADAM 1 and ADAM 2): a candidate sperm-egg binding/fusion complex, *Biol. Reprod.* **56**, 1245–1254.

Williamson, R.A., Koehler, K.K., Smith, W.D., Stenchever, M.A. (1984) Ultrastructural sperm tail defects associated with sperm immotility, *Fertil. Steril.* **41**, 103–107.

Wolf, D.P., Alexander, M., Zelinski-Wooten, M., Stouffer, R.L. (1996) Maturity and fertility of rhesus monkey oocytes collected at different intervals after an ovulatory stimulus (human chorionic gonadotropin) in in vitro fertilization cycles, *Mol. Reprod. Dev.* **43**, 76–81.

Wolf, D.P., VandeVoort, C.A., Meyer-Haas, G.R., Zelinski-Wooten, M.B., Hess, D.L., Baughman, W.L., Stouffer, R.L. (1989) In vitro fertilization and embryo transfer in the rhesus monkey, *Biol. Reprod.* **41**, 335–346.

Wolfsberg, T.G., Straight, P.D., Gerena, R.L., Huovila, A.P., Primakoff, P., Myles, D.G., White, J.M. (1995) ADAM, a widely distributed and developmentally regulated gene family encoding membrane proteins with a disintegrin and metalloprotease domain, *Dev. Biol.* **169**, 378–383.

Wright, V.C., Schieve, L.A., Reynolds, M.A., Jeng, G. (2003) Assisted reproductive technology surveillance–United States, 2000, *MMWR Surveill. Summ.* **29**, 1–16. (Erratum in *MMWR* **52**, 942).

Yanagimachi, R. (1969) In vitro capacitation of hamster spermatozoa by follicular fluid, *J. Reprod. Fertil.* **18**, 275–286.

Yeoman, R.R., Gerami-Naini, B., Mitalipov, S., Nusser, K.D., Widmann-Browning, A.A., Wolf, D.P. (2001) Cryoloop vitrification yields superior survival of Rhesus monkey blastocysts, *Hum. Reprod.* **16**, 1965–1969.

Youakim, A., Hathaway, H.J., Miller, D.J., Gong, X., Shur, B.D. (1994) Overexpressing sperm surface beta 1,4-galactosyltransferase in transgenic mice affects multiple aspects of sperm-egg interactions, *J. Cell Biol.* **126**, 573–1583.

Yuan, R., Primakoff, P., Myles, D.G. (1997) A role for the disintegrin domain of cyritestin, a sperm surface protein belonging to the ADAM family, in mouse sperm-egg plasma membrane adhesion and fusion, *J. Cell Biol.* **137**, 105–112.

Yudin, A.I., Tollner, T.L., Li, M.-W., Treece, C.A., Overstreet, J.W., Cherr, G.N. (2003) ESP13.2, a member of the beta-defensin family, is a macaque sperm surface-coating protein involved in the capacitation process, *Biol. Reprod.* **69**, 1118–1128.

In-vitro Fertilization: *see In vitro* Fertilization

Infectious Disease Testing by LCR

Max A. Chernesky[1], *Ron L. Marshall*[2], *and Astrid K. Petrich*[1]
[1] *Department of Pathology and Molecular Medicine, Hamilton, ON, Canada*
[2] *Molecular Diagnostics, Abbott Park, IL, USA.*

1 Principles of Ligase Chain Reaction (LCR) 555
1.1 Amplification Steps/Probe Synthesis 556
1.2 Gapped LCR (G LCR) 556

2 Techniques 557
2.1 Automation Methods 557
2.2 Detection of Amplified Product 557

3 Applications 558
3.1 Detection of *C. trachomatis* 558
3.2 Detection of GC 558
3.3 Detection of *M. tuberculosis* 558
3.4 Detection of Other Microorganisms 559

4 Perspectives 559
4.1 Specimen Preparation 559
4.2 Advantages 560
4.3 Disadvantages 560

Bibliography 560
Secondary References (Books/Review Papers) 560
Primary Reference Papers 561

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

Keywords

Amplification Inhibitors
Compounds in clinical specimens may inhibit amplification by reducing enzyme activity (DNA polymerase and/or DNA ligase), resulting in the occurrence of false-negative results.

DNA Polymerase
A thermostable DNA polymerase allows primer annealing and filling in of the gap at a higher temperature in G LCR. This increases specificity and eliminates the need to add additional enzyme following each cycle of denaturation.

Gap
Primer sets are chosen to anneal side by side to a single denatured template DNA target strand with a short gap of a few bases between the 3′ end of the first primer and the 5′ end of the adjacent primer. The gap consists of three or fewer of the four possible bases. This gap is filled in by the thermostable DNA polymerase. Since fewer than all four nucleotides are provided, the gap is precisely terminated at the beginning of the downstream probe. The resultant nick between the two annealed primers is sealed by DNA ligase.

Ligation
Formation of a covalent link between two immediately adjacent oligonucleotides annealed to a target strand of DNA by a thermally stable enzyme, DNA ligase.

Microparticles
Use of a carbazole label on the first probe and an adamantane label on the second probe allows for ligated probe capture with anti-carbazole microparticles and detection with anti-adamantane alkaline phosphatase. Microparticles with alternate surface chemistry have also been used by Abbott Laboratories in an automated system for purification of nucleic acids from clinical specimens.

Probe Amplification
Ligase chain reaction (LCR) is a nucleic-acid amplification technique in which sequential rounds of template-dependent ligation of four oligonucleotide probes result in the exponential synthesis of probes.

Ligase Chain Reaction (LCR) is a highly sensitive method for detecting nucleic-acid sequences. The LCR uses two pairs of oligonucleotide probes to hybridize to denatured target DNA strands. The positive-strand probe pairs are designed to bind immediately adjacent to each other on the target molecule. The negative-strand probe pairs bind similarly on the negative strand of the target DNA. A thermostable

DNA ligase is used to seal the nick between the annealed probes on each strand. Denaturation and annealing is achieved by varying the reaction temperature (thermocycling), and after strand separation, the newly ligated probe serves as a target template for subsequent reactions. Gapped LCR (G LCR) involves the use of DNA polymerase and DNA ligase to fill and seal a short gap between the annealed probes. In gapped LCR, the positive-sense probe pair binds to the target sequence, but they are separated by a few bases consisting of three or fewer of the four possible nucleotide bases. The negative-sense probe pairs bind similarly on the opposite-sense target strand. The thermostable DNA polymerase, provided with only a specific subset of the four possible nucleotides, extends the probe to a defined length until it is immediately adjacent to its downstream probe partner. The resulting nick between the extended probe and its partner is then ligated to complete the process. Similar to LCR, the newly ligated probe serves as a target template for subsequent reactions. GLCR has been commercialized for several microorganisms. The process can be qualitative and quantitative and has been shown to be more sensitive and specific than nonamplified assays for *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in traditional swab specimens and noninvasive clinical specimens such as urine. LCR assays have also been developed for other organisms such as *Mycobacterium tuberculosis*, human immunodeficiency virus DNA sequences, orthopoxviruses, human papillomavirus, herpes simplex virus, and hepatitis C virus RNA.

1 Principles of Ligase Chain Reaction (LCR)

Nucleic-acid amplification (NAA) assays have impacted the clinical microbiology laboratory by providing sensitive means of detecting specific nucleic acids. The amplification aspect of these assays has been achieved by a number of unique mechanisms including probe amplification, which is seen in ligase chain reaction (LCR). Unlike PCR where the target itself is amplified, LCR ligates two probes on each target strand. These ligated probes can then act as a template for subsequent cycles and can exponentially accumulate during the reaction. The principle of LCR is that two adjacent synthetic oligonucleotide probes anneal to a single strand of denatured target nucleic acid and are ligated to each other. LCR was first introduced in the late 1980s as the *ligase amplification reaction* and used sequential rounds of template-dependent ligation to identify specific nucleic-acid targets. Similar to PCR, LCR requires the use of a thermal cycler, detects DNA, and goes through a series of cycles for amplification to be achieved. Use of a thermostable ligase minimizes nonspecific interactions as ligation can be performed at or near the melting temperature (Tm) of the probes. Also, the thermostable ligase is able to maintain its activity even after multiple rounds of denaturation. The requirement of needing specific annealing of four oligonucleotide probes also increases specificity.

Achieving accurate results from LCR requires attention to primer design and reaction conditions. The melting temperature

of all four primers used should be within a narrow temperature range, ideally, 70 °C ± 2 °C, and having a GC basepair at the 3′ end of the first probe or at the 5′ end of the adjacent probe results in a stronger interaction, which leads to more successful ligation.

1.1 Amplification Steps/Probe Synthesis

Ligase chain reaction uses four oligonucleotide probes to recognize the target nucleic acid. In the standard LCR reaction, following a denaturation step, a pair of oligonucleotide probes anneal immediately beside each other on a strand of target nucleic acid. The second pair of probes anneals to the complementary strand in the same manner, leaving only a nick between the probes. The nick is sealed for each pair of probes with a thermostable ligase, and each newly ligated reaction product as well as the original target DNA can serve as template for subsequent cycles of denaturation, annealing, and ligation.

1.2 Gapped LCR (G LCR)

Gapped LCR (G LCR) is a modification of the standard LCR reaction that increases the specificity of the assay and reduces background due to blunt-end ligation (Fig. 1). The two pairs of oligonucleotide probes are designed to anneal to the template, with a small gap consisting of only three of fewer of the four possible nucleotides. The reaction mixture includes a thermostable DNA polymerase and the required subset of the four dNTPs. The thermostable DNA polymerase adds the specific dNTPs to fill the gap, resulting in an extended probe ending immediately adjacent to its downstream partner. The newly formed nick is then sealed by the DNA ligase. Nonspecific binding of the probes will not allow for ligation to occur. Other aspects of the amplification assay are unchanged.

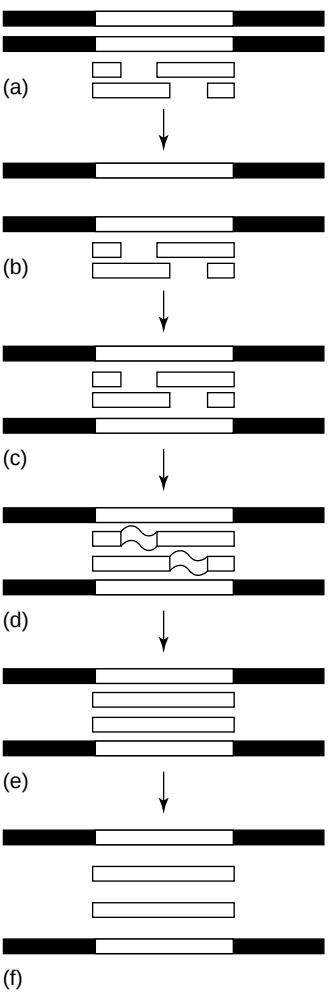

Fig. 1 Gapped ligase chain reaction amplification of DNA. (a) specimen with target; (b) heat to separate strands; (c) cool to allow probes to bind (d) PCR to fill gaps; (e) ligase to link two adjoining probes; (f) heat to separate new targets.

2 Techniques

2.1 Automation Methods

Automation has been a challenge for all NAA assays. The Abbott LCx system consists of a thermal cycler for amplification and a semiautomated enzyme immunoassay analyzer for detection of LCR or PCR products. Low-throughput manual specimen-preparation techniques have been superseded by higher-throughput systems that have employed several strategies. Abbott Laboratories has produced an M1000 automated system designed for the purification of nucleic acids from clinical specimens, using magnetic microparticles. The system uses uncoated iron-oxide particles to extract and purify viral RNA from clinical specimens, and reagent pipetting, heating, incubating, capturing, washing, and eluting do not require user intervention. HIV and hepatitis C virus (HCV) assays performed on the M1000 detected 50 to 1-million copies of RNA per ml and interfering substances were removed in the process. The process has also been demonstrated for DNA, using silica-based microparticles. Another approach has been to automate the specimen-pipetting component of sample preparation in the LCx Chlamydia assay by using a commercially available liquid-handling system called *Tecan Genesis RSP 100* (Tecan Inc., Research Triangle Park, NC USA). A mean time of 2.5 min was required to pipette a complete LCx assay carousel containing 20 samples and 4 controls, compared to 8.4 min manually, with equal precision and accuracy.

Other approaches have shown that pooling of swabs or urines in pools of four or eight have increased the number of specimens that could be processed for *C. trachomatis* and *N. gonorrhoeae*, maintaining good sensitivity and specificity and generating substantial cost savings (~50%) when the population tested has less than 10% prevalences of these two bacteria.

2.2 Detection of Amplified Product

Initial detection methods utilized a ^{32}P-label on the 3′ end of the upstream primer. The ligated product was separated from the nonligated primer by denaturing gel electrophoresis and the LCR product was detected by autoradiography. This detection procedure was followed by nonradioactive methods, one of which used fluorescent-labeled probes in a microtitre plate format. Detection of the fluorescently labeled LCR product was achieved using a fluorescent DNA sequencer and a Genescanner (Applied Biosystems). A less sophisticated nonradioactive method utilized one digoxigenin-labeled probe that was incorporated in the final LCR product and was detected by Southern blotting after separation by gel electrophoresis. An improvement for commercial applications is the Abbott LCx Analyzer, which incorporated microparticle enzyme immunoassay (MEIA). Use of a carbazole label at the 5′ end of the first probe and an adamantane label on the 3′ end of the second probe allows for ligated probe capture with antibody-coated anti-carbazole microparticles. The automation transfers the amplification product sample to a glass-fiber matrix to which the microparticle complexes bind the product irreversibly. A wash step removes unligated probe with only the detection label. The bound microparticle complexes are then incubated with alkaline phosphatase-conjugated antibody that

binds to the adamantane hapten. The complex is then detected by adding a substrate that is dephosphorylated to a fluorescent molecule by the alkaline phosphatase. The fluorescence rate is detected in the MEIA optical assembly in the LCx Analyzer. The presence or absence of specific target is determined by comparing the specimen result to the cutoff (CO) value. The CO value is the mean rate of LCx Calibrator duplicates multiplied by a factor determined in clinical studies (0.45 for *C. trachomatis*, 0.25 for *N. gonorrhoeae*). The sample rate/CO (S/CO) value is determined by calculating a ratio of the sample rate to the CO value. An S/CO ratio >1.0 indicates a positive LCx assay result.

3 Applications

3.1 Detection of *C. trachomatis*

The first use of LCR for amplifying *C. trachomatis* DNA appeared in 1993. After several years of improvement, the assay was quickly used in studies on clinical specimens and was made commercially available in 1995. The test demonstrated high sensitivity and specificity on cervical and urethral swabs compared to culture and detection of *C. trachomatis* antigens and nucleic acids by hybridization. These comparisons led to the conclusion that LCR and the other NAA assays were approximately 30% more sensitive than all other preexisting diagnostic technologies. The timely publications of the ability of LCR to diagnose *C. trachomatis* infections by noninvasive screening of female urine was rapidly confirmed by several authors. More recently, other noninvasive specimens such as vulva, introital, and vaginal swabs have been shown to be effective for *C. trachomatis* diagnosis using LCR. Other clinical samples now being pursued are pharyngeal, glans penis, rectal swabs, and synovial fluid.

3.2 Detection of GC

Birkenmeyer and Armstrong published the first evaluation of LCR for the diagnosis of *N. gonorrhoeae* (GC). Like the results achieved on clinical specimens for the detection of *C. trachomatis* nucleic acids, LCR clinical evaluations for GC on cervical and urethral swabs and urine have shown similarly high sensitivity and specificity. LCR has frequently diagnosed additional GC infections compared to culture. The LCR test has performed very well on vaginal swabs and extragenital specimens such as anorectal and pharyngeal specimens.

3.3 Detection of *M. tuberculosis*

Application of LCR for bacterial pathogens has not been limited to the STD pathogens; Iovannisci and Winn-Deen used LCR to detect the *M. tuberculosis* insertion sequence IS6110. Using fluorescently labeled probes and a fluorescent DNA sequencer, they were able to detect 100 copies of the target even in the presence of other exogenous nucleic acid. Ausina et al. described the first evaluation of a commercial LCR assay (Abbott Diagnostics) for direct detection of *M. tuberculosis* in respiratory specimens. The four primers recognize the *M. tuberculosis*-specific protein antigen b gene in a G LCR format, followed by detection using the LCx Analyzer. Similar to other commercial LCR assays, there are three main steps to the protocol: specimen preparation, amplification, and detection.

As a large proportion of respiratory specimens sent to the Mycobacteriology laboratory are sputums, specimen preparation in the *M. tuberculosis* assay has been modified. Pretreated SDS–NaOH (decontamination/mucous treatment) concentrates of respiratory specimens, which are used routinely for smear and culture detection of *M. tuberculosis*, are the starting materials for extraction. The specimen is further treated with lysis buffer and the mycobacterial DNA is released by mechanical lysis in the LCx Lysor. Extrapulmonary specimens have been evaluated in a number of studies, and sensitivities are similar to other amplification assays. The proposed use of this assay to identify *M. tuberculosis* from culture-positive bottles shows high sensitivity and specificity.

The Commercial Mtb LCx assay has been compared to other amplification methods, including PCR (COBAS AMPLICOR *M. tuberculosis* Test, Roche Diagnostics Systems, Brachburg, NJ) and Transcription-mediated Amplification (Amplified *M. tuberculosis* Direct Test, Gen-Probe, San Diego, Calif) on both respiratory and extrapulmonary specimens, and has demonstrated comparable sensitivity to other commercial assays. There are two reports of a specific isolate of *M. tuberculosis* not being detected by the LCx assay. Further characterization of the isolate in one of the studies indicated that a deletion in the target region for the probes was responsible for the failure in detection. This appears to be a fairly rare incident; however, it should be considered when using the assay as a screening tool. Generally, this is a highly reliable assay on smear-positive respiratory specimens and a multicenter evaluation of the LCx assay suggested that the sensitivity of the assay might be improved with the inclusion of a gray zone.

3.4 Detection of Other Microorganisms

Although LCR assays have found their greatest application in detection of the STD organisms *C. trachomatis* and *N. gonorrhoeae*, and more recently, in the detection of *M. tuberculosis*, since its inception, the LCR has also been used for numerous other applications. Early viral targets for LCR assay development have included human immunodeficiency virus, human papillomaviruses, and herpes simplex viruses. Detection of other viral targets included an LCR reaction that was able to differentiate Cowpoxvirus from other Orthopoxviruses by targeting two adjacent nucleotides and an asymmetric G LCR to detect Hepatitis C virus RNA. Utilizing the unique ability of the G LCR to discriminate DNA sequence differences, Karthigesu et al. and Osiowy described the use of G LCR to identify Hepatitis B virus (HBV) surface gene variants. LCR was also used to detect HBV precore mutants. Mutations associated with antimicrobial resistance can also be detected using G LCR, as described for ganciclovir resistance in cytomegalovirus (CMV) and for characterizing eliminate SHV B-lactamases. Bacterial targets for LCR development have included *Borrelia burgdorferi*. The LCR assay has also been applied to the agriculture and environmental fields, with its use in the identification and discrimination of *Listeria monocytogenes* and *Erwinia stewartii*.

4 Perspectives

4.1 Specimen Preparation

Clinical specimens may contain intracellular or extracellular nucleic acids for detection in NAA assays. The relative

concentration of DNA and inhibitors of amplification and/or ligation may impact on whether the sample is recorded as positive or negative in the LCR assay. Specimen adequacy has been used to monitor positive or negative *C. trachomatis* results in cervical swabs. It has been postulated that higher reactive levels might be achieved in the *M. tuberculosis* LCR if greater amounts of DNA could be released from the bacteria. Each NAA test has the CO for positivity and negativity established by statistical calculation performed on dilutions of nucleic acids and large numbers of positive and negative clinical specimens. Levels are arbitrarily raised or lowered to yield optimum sensitivity and specificity. Each commercial assay is required to be reproducible, and therefore, strict adherence to defined specimen preparation is important. When an assay is used on a new specimen type, the results can only be trusted after appropriate validation has been achieved.

Inhibitors of the *C. trachomatis* LCx have been reported and have been found to be associated with female urinary substances. These studies should be interpreted with caution as it is recognized that inhibitors are detected by adding certain strengths of *C. trachomatis* DNA to specimens, and too little DNA suggests more inhibition and too much implies none. Added DNA may also be lost during processing creating "false inhibition". Reproducibility problems have been reported for the LCx *C. trachomatis* assay. This phenomenon has been mainly experienced on specimens near the test CO and is present with other amplification tests such as PCR.

4.2 Advantages

The advantages of LCR have been its clinical robustness in the LCx format. The assay is virtually without contamination in the LCx instrument. Small volumes of clinical specimens such as urine are used and samples are stabilized at ambient temperature for sufficient periods to allow mailed and home-collected specimens to yield positive results.

4.3 Disadvantages

The LCR can be susceptible to inhibitors in clinical specimens, requiring a removal maneuver early in the protocol. Low throughput in the existing instrumentation is a disadvantage, which may be addressed as discussed in Sect. 2.1 above.

See also Genomic Sequencing (Core Article); Immuno-PCR; Preparedness for Bioterrorism.

Bibliography

Secondary References (Books/Review Papers)

Barany, F. (1991) Genetic disease detection and DNA amplification using cloned thermostable ligase, *Proc. Natl. Acad. Sci.* **88**, 189–193.

Birkenmeyer, L.G., Mushahwar, I.K. (1991) Mini-review: DNA probe amplification methods, *J. Virol. Methods* **35**, 117–126.

Burczak, J.D., Ching, S.F., Hu, H.Y., Lee, H.H. (1995) Ligase chain reaction for the detection of infectious agents, in Wiebrauk D., Farkas D.H (Eds.), *Molecular Methods for Virus Detection*, Academic Press, New York pp. 315–328

Carrino, J.J., Lee, H.H. (1995) Nucleic acid amplification methods, *J. Microbiol. Methods* **23**, 3–20.

Laffler, T.G., Carrino, J.J., Marshall, R.L. (1993) The ligase chain reaction in DNA-based diagnosis, *Ann. Biol. Clin.* **50**, 821–826.

Leckie, G.W., Lee, H.H. (1995) Infectious disease testing by ligase chain reaction, in R.A. Myers (Ed.), *Molecular Biology and*

Biotechnology Testing, VCH Publishers, New York pp. 463–466.

Shimer, G.H., Jr., Backman, K.C. (1995) Ligase chain reaction, *Methods Mol. Biol.* **46**, 269–278.

Wiedmann, M., Wilson, W.J., Czajka, J., Luo, J., Barany, F., Batt, C.A. (1994) Ligase chain reaction (LCR)-overview and applications, *PCR Methods Appl.* **3**, S51–S64.

Primary Reference Papers

Andersen, A.B., Hansen, E.B. (1989) Structure and mapping of antigenic domains of protein antigen b, a 38,000 molecular-weight protein of *Mycobacterium tuberculosis*., *Infect. Immun.* **57**, 2481–2488.

Ausina, V., Gamboa, F., Gazapo, E., Manterola, J.M., Lonca, J., Matas, L., Manzano, J.R., Rodrigo, C., Cardona, P.J., Padilla, E. (1997) Evaluation of the semiautomated Abbott LCx *Mycobacterium tuberculosis* assay for direct detection of *Mycobacterium tuberculosis* in respiratory specimens, *J. Clin. Microbiol.* **35**, 1996–2002.

Backman, K., December (1987) Method for detecting a target nucleic acid sequence. European Patent Office-320-308.

Barany, F., Gelfand, D.H. (1991) Cloning, overexpression, and nucleotide sequence of a thermostable DNA ligase-encoding gene, *Gene* **109**, 1–11.

Bassiri, M., Hu, H.Y., Domeika, M.A., Burczak, J., Svensson, L.O., Lee, H.H., Mardh, P.A. (1995) Detection of *Chlamydia trachomatis* in urine specimens from women by ligase chain reaction, *J. Clin. Microbiol.* **33**, 898–900.

Berg, E.S., Anestad, G., Moi, H., Storvold, G., Skaug, K. (1997) False-negative results of a ligase chain reaction to detect *Chlamydia trachomatis* due to inhibitors, *Eur. J. Clin. Microbiol. Infect. Dis.* **16**, 727–731.

Birkenmeyer, L., Armstrong, A.S. (1992) Preliminary evaluation of the ligase chain reaction for specific detection of *Neisseria gonorrhoeae*, *J. Clin. Microbiol.* **30**, 3089–3094.

Blocker, M.E., Krysiak, R.G., Behets, F., Cohen, M.S., Hobbs, M.M. (2002) Quantification of *Chlamydia trachomatis* elementary bodies in urine by ligase chain reaction, *J. Clin. Microbiol.* **40**, 3631–3634.

Bond, Carrino, S., J., Hampl, H., Hanley, K. Rinehardt, L., Laffler, T. (1990) New methods of detection of HPV. in Monsonego, J. (Ed.) *Serona Symposia*, Raven Press, Paris, France pp. 425–434.

Bourgeois, C., Sixt, N., Bour, J.B., Pothier, P. (1997) Value of a ligase chain reaction assay for detection of ganciclovir resistance-related mutations in UL97 gene of human cytomegalovirus, *J. Virol. Methods* **67**, 167–175.

Buimer, M., Van Doornum, D.J., Ching, S., Peerbooms, P.G., Plier, P.K., Ram, D., Lee, H.H. (1996) Detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* by ligase chain reaction-based assays with clinical specimens from various sites: implications for diagnostic testing and screening, *J. Clin. Microbiol.* **34**(10), 2395–2400.

Carrino, J.J., Laffler, T.G. (1991) Detection of HIV DNA sequences using the ligase chain reaction (LCR), *Clin. Chem.* **37**, 1059.

Carroll, K.C., Aldeen, W.E., Morrison, M., Anderson, R., Lee, D., Mottice, S. (1998) Evaluation of the Abbott LCx ligase chain reaction assay for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in urine and genital swab specimens from a sexually transmitted disease clinic population, *J. Clin. Microbiol.* **36**(6), 1630–1633.

Castriciano, S., Luinstra, K., Jang, D., Patel, J., Mahony, J., Kapala, J., Chernesky, M. (2002) Accuracy of results obtained by performing a second ligase chain reaction assay and PCR analysis on urine samples with positive or near cut-off results in the LCx for *Chlamydia trachomatis*, *J. Clin. Microbiol.* **40**, 2632–2634.

Cheng, H., Macaluso, M., Vermund, S., Hook, III, E.W. (2001) Relative accuracy of nucleic acid amplification tests and culture in detecting *chlamydia* in asymptomatic men, *J. Clin. Microbiol.* **39**(11), 3927–3937.

Chernesky, M., Chong, S., Jang, D., Luinstra, K., Faught, M., Mahony, J. (1998) Inhibition of amplification of *Chlamydia trachomatis* plasmid DNA by the ligase chain reaction associated with female urines, *Clin. Microbiol. Infect.* **4**, 397–404.

Chernesky, M., Chong, S., Jang, D., Luinstra, K., Sellors, J., Mahony, J.B. (1997b) Ability of ligase chain reaction and PCR assays to diagnose *Chlamydia trachomatis* infections in men by testing first-void urine, *J. Clin. Microbiol.* **35**, 982–984.

Chernesky, M.A., Lee, H., Schachter, J., Burczak, J.D., Stamm, W.E., McCormack, W.M.,

Quinn, C. (1994b) Diagnosis of *Chlamydia trachomatis* urethral infection in symptomatic and asymptomatic men by testing first-void urine in a ligase chain reaction, *J. Infect. Dis.* **170**, 1308–1311.

Chernesky, M.A., Jang, D., Lee, H., Burczak, J.D., Hu, H., Sellors, J., Tomazic-Allen, S.J., Mahony, J.B. (1994a) Diagnosis of *Chlamydia trachomatis* infections in men and women by testing first-void urines by ligase chain reaction, *J. Clin. Microbiol.* **32**, 2682–2685.

Cheng, S., Lee, H., Hook, III, E.W., Jacobs, M.R., Zenilman, J. (1995) Ligase chain reaction for detection of *Neisseria gonorrhoeae* in urogenital swabs, *J. Clin. Microbiol.* **33**(12), 3111–3114.

Chong, S., Jang, D., Song, X., Mahony, J., Petrich, A., Barriga, P., Chernesky, M. (2003) Specimen processing and concentration of *Chlamydia trachomatis* added can influence false-negative rates in the LCx assay but not in the APTIMA combo 2 assay when testing for inhibitors, *J. Clin. Microbiol.* **41**, 778–782.

Dille, B.J., Butzen, C.C., Birkenmeyer, L.G. (1993) Amplification of *Chlamydia trachomatis* DNA by ligase chain reaction, *J. Clin. Microbiol.* **31**, 729–731.

Gamboa, F., Dominguez, J., Padilla, E., Manterola, J.M., Gazapo, E., Lonca, J., Matas, L., Hernandez, A., Cardona, P.J., Ausina, V. (1998) Rapid diagnosis of *extrapulmonary tuberculosis* by ligase chain reaction amplification, *J. Clin. Microbiol.* **36**(5), 1324–1329.

Garland, S.M., Tabrizi, S., Hallo, J., Chen, S. (2000) Assessment of *Chlamydia trachomatis* prevalence by PCR and LCR in women presenting for termination of pregnancy, *Sex. Transm. Inf.* **76**, 173–176.

Gaydos, C.A., Howell, M.R., Quinn, T.C., Gaydos, J.C., McKee, Jr., K.T. (1998) Use of ligase chain reaction with urine versus cervical culture for detection of *Chlamydia trachomatis* in an asymptomatic military population of pregnant and nonpregnant females attending papanicolaou smear clinics, *J. Clin. Microbiol.* **36**(5), 1300–1304.

Gilpin, C., Dawson, D.J., O'Kane, G., Armstrong, J.G., Coulter, C. (2002) Failure of commercial ligase chain reaction to detect *Myobacterium tuberculosis* DNA in sputum samples from a patient with smear-positive pulmonary tuberculosis due to a deletion of the target region, *J. Clin. Microbiol.* **40**(6), 2305–2307.

Gronowski, A.M., Copper, S., Baorto, D., Murray, P.M. (2000) Reproducibility problems with the Abbott Laboratories LCx Assay for *Chlamydia trachomatis* and N*eisseria gonorrhoeae*, *J. Clin. Microbiol.* **38**, 2416–2418.

Hanson, K.L., Cartwright, C.P. (2001) Evaluation of an automated liquid-handling system (Tecan Genesis RSP 100) in the Abbott LCx assay for *Chlamydia trachomatis*, *J. Clin. Microbiol.* **39**(5), 1975–1977.

Hook, III, E.W., Ching, S.F., Stephens, J., Hardy, K.F., Smith, K.R., Lee, H.H. (1997) Diagnosis of *Neisseria gonorrhoeae* Infections in Women by Using ligase chain reaction on patient-obtained vaginal swabs, *J. Clin. Microbiol.* **35**(8), 2129–2132.

Hu, H., Elmore, K., Facey, I., Jenderzak, D. (1991) Detection of *Borrelia burgdorferi* by ligase chain reaction, *Abstr. 1st Gen. Meet. Am. Soc. Microbiol.* **91**, 365.

Hu, H.-Y.Y., Burczak, J.D., Leckie, G.W., Ray, K.A., Muldoon, S., Lee, H.H. (1996) Analytic performance and contamination control methods of a ligase chain reaction DNA amplification assay for detection of C*hlamydia trachomatis* in urogenital specimens, *Diagn. Microbiol. Infect. Dis.* **24**, 71–76.

Iovannisci, D.M., Winn-Deen, E.S. (1993) Ligation amplification and fluorescence detection of *Mycobacterium tuberculosis* DNA, *Mol. Cell. Probes* **7**, 35–43.

Kacena, K.A., Quinn, S.B., Hartman, S.C., Quinn, T.C., Gaydos, C.A. (1998) Pooling of urine samples for screening for *Neisseria gonorrhoeae* by ligase chain reaction: accuracy of this application, *J. Clin. Microbiol.* **36**, 3624–3628.

Kapala, J., Copes, D., Sproston, A., Patel, J., Jang, D., Petrich, A., Mahony, J., Biers, K., Chernesky, M. 2000. Pooling cervical swabs and testing by ligase chain reaction are accurate and cost-saving strategies for diagnosis of *Chlamydia trachomatis*, *J. Clin. Microbiol.* **38**, 2480–2483.

Karthigesu, V.D., Mendy, M., Fortuin, M., Whittle, H.C., Howard, C.R., Allison, L.M. (1995) The ligase chain reaction distinguishes hepatitis B virus S-gene variants, *FEMS Microbiol. Lett* **131**, 127–132.

Kim, J., Lee, H.-J. (2000) Rapid and discriminatory detection of genes coding for SHV β-lactamases by ligase chain reaction, *Antimicrob. Agents Chemother.* **44**, 1860–1864.

Kissin, D.M., Holman, S., Minkoff, H.L., DeMeo, L., McCormack, W.M., DeHovitz, J.A.

(2002) Epidemiology and natural history of ligase chain reaction detected chlamydial and gonococcal infections, *Sex. Trans. Infect.* **78**, 208–209.

Krepel, J., Patel, J., Sproston, A., Hopkins, F., Jang, D., Mahony, J., Chernesky, M. (1999) The impact on accuracy and cost of ligase chain reaction screening for genital *Chlamydia trachomatis* infection in asymptomatic women, *J. Clin. Microbiol.* **36**, 481–485.

Lauderdale, T., Landers, L., Thorneycroft, I., Chapin, K. (1999) Comparison of the PACE 2 assay, two amplification assays, and clearview EIA for detection of *Chlamydia trachomatis* in female endocervical and urine specimens, *J. Clin. Microbiol.* **37**(7), 2223–2229.

Lee, H.H., Chernesky, M.A., Schachter, J., Burczak, J.D., Andrews, W.W., Muldoon, S., Leckie, G., Stamm, W.E. (1995) Diagnosis of *Chlamydia trachomatis* genitourinary infection in women by ligase chain reaction assay of urine, *Lancet* **28**, 213–216.

Loeffelholz, M.J., Jirsa, S.J., Teske, R.K., Woods, J.N. (2001) Effect of endocervical specimen adequacy on ligase chain reaction detection of *chlamydia trachomatis, J. Clin. Microbiol.* **39**(11), 3838–3841.

Lumb, R., Davies, K., Dawson, D., Gibb, R., Gottlieb, T., Kershaw, C., Kociuba, K., Nimmo, G., Sangster, N., Worthington, M., Bastian, I. (1999) Multicenter evaluation of the Abbott LCx *Mycobacterium tuberculosis* ligase chain reaction assay.

Mahony, J., Chong, S., Jang, D., Luinstra, K., Faught, M., Dalby, D., Sellors, J., Chernesky, M. (1998) Urinary specimens from pregnant and nonpregnant women inhibitory to amplification of *Chlamydia trachomatis* nucleic acid by PCR, ligase chain reaction, and transcription-mediated amplification: identification of urinary substances associated with inhibition and removal of inhibitory activity, *J. Clin. Microbiol.* **36**, 3122–3126.

Marshall, R.L., Laffler, T.G., Cerney, M.B., Kratochvil, J.C., Morgan, R.L. (1994) Detection of HCV RNA by the asymmetric gap ligase chain reaction, *PCR Methods Appl.* **4**, 80–84.

Minamitani, S., Nishiguchi, S., Kuroki, T., Otani, S., Monna, T. (1997) Detection by ligase chain reaction of precore mutant of hepatitis B virus, *Hepatology* **25**, 216–222.

Moller, J.K., Anderson, B., Olesen, F., Lignell, T., Ostergaard, L. (1999) Impact of menstrual cycle on the diagnostic performance of LCR, TMA, and PCE for detection of *Chlamydia trachomatis* in home obtained and mailed vaginal flush and urine samples, *Sex. Transm. Inf.* **75**(4), 228–230.

Morre, S.A., van Valkengoed, I.G., de Jong, A., Boeke, A.J., van Eijk, J.T., Meijer, C.J., van den Brule, A.J. (1999) Mailed, home-obtained urine specimens: a reliable screening approach for detecting asymptomatic *Chlamydia trachomatis* infections, *J. Clin. Microbiol.* **37**, 976–980.

Nikkari, S., Puolakkainen, M., Yli-Kerttula, U., Luukkainen, R., Lehtonen, O.P., Toivanen, P. (1997) Ligase chain reaction in detection of Chlamydia DNA in synovial fluid cells, *Br. J. Rheumatol.* **36**, 763–765.

Osiowy, C. (2002) Sensitive detection of HBsAg mutants by a gap ligase chain reaction assay, *J. Clin. Microbiol.* **40**, 2566–2571.

Page-Shafer, K., Graves, A., Kent, C., Balls, J.E., Zapitz, V.M., Klausner, J.D. (2002) Increased sensitivity of DNA amplification testing for the detection of pharyngeal gonorrhea in men who have sex with men, *Clin. Infect. Dis.* **34**, 173–176.

Pfeffer, M., Meyer, H., Wiedmann, M. (1994) A ligase chain reaction targeting two adjacent nucleotides allows the differentiation of cowpox virus from other orthopoxvirus species, *J. Virol. Methods* **49**, 353–360.

Piersimoni, C., Callegaro, A., Scarparo, C., Penati, V., Nista, D., Bornigia, S., Lacchini, C., Scagnelli, M., Santini, G., DeSio, G. (1998) Comparative evaluation of the new Gen-Probe *Mycobacterium tuberculosis* amplified direct test and the semiautomated Abbott LCx *Mycobacterium tuberculosis* assay for direct detection of *Mycobacterium tuberculosis* complex in respiratory and extrapulmonary specimens, *J. Clin. Microbiol.* **36**(12), 3601–3604.

Puolakkainen, M., Hiltunen-Back, E., Reunala, T., Suhonen, S., Lahteenmaki, P., Lehtinen, M., Paavonen, J. (1998) Comparison of performances of two commercially available tests, a PCR and a ligase chain reaction test, in detection of urogenital *Chlamydia trachomatis* Infection, *J. Clin. Microbiol.* **36**(6), 1489–1493.

Rinehardt, L., Hampl, H., Laffler, T.G. 1991. Ultrasensitive non-radioactive detection of herpes simplex virus by LCR, the ligase chain reaction. *In* 20th *Annual Meeting of the Keystone Symposia on molecular and cellular biology*, p. 101.

Schachter, J., Stamm, W.E., Quinn, T.C., Andrews, W.W., Burczak, J.D., Lee, H.H. (1994) Ligase chain reaction to detect *Chlamydia trachomatis* infection of the cervix, *J. Clin. Microbial.* **32**, 2540–2543.

Schachter, J., McCormack, W.M., Chernesky, M.A., Martin, D.H., Van Der Pol, B., Rice, P.A., Hook, III, E.W., Stamm, W.E., Quinn, T.C., Chow, J.M. (2003) Vaginal swabs are appropriate specimens for diagnosis of genital tract infection with *Chlamydia trachomatis*, *J. Clin. Microbiol.* **41**(8), 3784–3789.

Smith, K.R., Ching, S., Lee, H., Ohhashi, Y., Hu, H.Y., Fisher, H.C., Hook, 3rd, E.W. (1995) Evaluation of ligase chain reaction for use with urine for identification of *Neisseria gonorrhoeae* in females attending a sexually transmitted disease clinic, *J. Clin Microbiol.* **33**(2), 455–457.

Stary, A., Najim, B., Lee, H.H. (1997a) Vulval swabs as alternative specimens for ligase chain reaction detection of genital chlamydial infections in women, *J. Clin Microbiol.* **35**, 836–838.

Stary, A., Ching, S.F., Teodorowicz, L., Lee., H. (1997b) Comparison of ligase chain reaction and culture for detection of *Neisseria gonorrhoeae* in genital and extragenital specimens, *J. Clin. Microbiol.* **35**(1), 239–242.

Thomas, B.J., Pierpoint, T., Taylor-Robinson, D., Renton, A.M. (1998) Sensitivity of the ligase chain reaction assay for detecting *Chlamydia trachomatis* in vaginal swabs from women who are infected at other sites, *Sex. Transm. Inf.* **74**, 140–141.

Tortoli, E., Lavinia, F., Simonetti, M.T. (1997) Evaluation of a commercial ligase chain reaction kit (Abbott LCx) for direct detection of *Mycobacterium tuberculosis* in pulmonary and extrapulmonary specimens, *J. Clin. Microbiol.* **35**, 2424–2426.

Tortoli, E., Lavinia, F., Simonetti, M.T. (1998) Early detection of *Mycobacterium tuberculosis* complex in BACTEC cultures by ligase chain reaction, *J. Clin. Microbiol.* **36**, 2991–2792.

Van Doornum, G.J., Schouls, L.M., Pijl, A., Cairo, I., Buimer, M., Brusiten, S. (2001) Comparison between the LCx probe system and the COBAS AMPLICORD system for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* infections in patients attending a clinic for treatment of sexually transmitted diseases in Amsterdam, The Netherlands, *J. Clin. Microbiol.* **39**(3), 829–835.

Van Doornum, G.J., Buimer, M., Prins, M., Henquet, C.J.M., Plier, R.A., Tomazic-Allen, S., Hu, H. (1995) Detection of *Chlamydia trachomatis* infection in urine samples from men and women by ligase chain reaction, *J. Clin. Microbiol.* **33**, 2042–2047.

Van Dyck, E., Ieven, M., Pattyn, S., Van Damme, L., Laga, M. (2001) Detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* by enzyme immunoassay, culture, and three nucleic acid amplification Tests, *J. Clin. Microbiol.* **39**(5), 1751–1756.

Waites, K.B., Smith, K.R., Crum, M.A., Hockett, R.D., Wells, A.H., Hook, III, E.W. (1999) Detection of *Chlamydia trachomatis* endocervical infections by ligase chain reaction versus ACCESS *Chlamydia* antigen assay, *J. Clin. Microbiol.* **37**(9), 3072–3073.

Wang, S.X., Tay, L. (1999) Evaluation of three nucleic acid amplification methods for direct direction of *Mycobacterium tuberculosis* complex in respiratory specimens, *J. Clin. Microbiol.* **37**(6), 1932–1934.

Wang, S.X., Tay, L. (2002) Early identification of *Mycobacterium tuberculosis* complex in BACTEC cultures by ligase chain reaction, *J. Med. Microbiol.* **51**, 710–712.

Wiedmann, M., Barany, F., Batt, C.A. (1993) Detection of *Listeria monocytogenes* with a nonisotopic polymerase chain reaction-coupled ligase chain reaction, *Appl. Environ. Microbiol.* **59**, 2743–2745.

Wiedmann, M., Czajka, J., Barany, F., Batt, C.A. (1992) Discrimination of *Listeria monocytogenes* from other Listeria species, *Appl. Environ. Microbiol.* **58**, 3443–3447.

Wiesenfeld, H.C., Heine, R.P., Rideout, I., Macio, I., DiBiasi, F., Sweet, R.L. (1996) The vaginal introitus: a novel site for *Chlamydia trachomatis* testing in women, *Am. J. Obstet. Gynecol.* **174**, 1542–1546.

Wilson, W.J., Wiedmann, M., Dillard, H.R., Batt, C.A. (1993) Identification of *Erwinia stewartii* by a ligase chain reaction assay, *Appl. Environ. Microbiol.* **60**, 278–284.

Winn-Deen, E.S., Batt, C.A., Wiedmann, M. (1993) Non-radioactive detection of *Mycobacterium tuberculosis* LCR products in a microtitre plate format, *Mol. Cell. Probes* **7**, 179–186.

Winter, A.J., Gilleran, G., Eastick, K., Ross, J.D.C. (2000) Comparison of a ligase chain reaction-based assay and cell culture for detection of pharyngeal carriage of *Chlamydia trachomatis*, *J. Clin. Microbiol.* **38**(9), 3502–3504.

Wu, D.Y., Wallace, R.B. (1989) The ligation amplification reaction (LAR)–Amplification of specific DNA sequences using sequential rounds of template–dependent ligation, *Genomics* **4**, 560–569.

Xu, K., Glanton, V., Johnson, S.R., Beck-Sague, C., Bhullar, V., Candal, D., Pettus, K.S., Farshy, C.E., Black, C.M. (1998) Detection of *Neisseria gonorrhoeae* infection by ligase chain reaction testing of urine among adolescent women with and without *Chlamydia trachomatis* infection, *Sex. Transm. Dis.* **25**(10), 533–538.

Informatics (Computational Biology)

Thomas Lengauer
Computational Biology and Applied Algorithmics,
Max-Planck-Institute for Informatics, Saarbrücken, Germany

1 Introduction 570

2 Sequencing Genomes 572

3 Molecular Sequence Analysis 573
3.1 Sequence Alignment 573
3.2 Phylogeny Construction 576
3.3 Finding Genes 578
3.4 Analyzing Regulatory Regions 579
3.5 Finding Repetitive Elements 580
3.6 Analyzing Genome Rearrangements 581

4 Molecular Structure Prediction 582
4.1 Protein Structure Prediction 582
4.1.1 Secondary Structure Prediction of Proteins 582
4.1.2 Similarity-based Protein Structure Prediction 583
4.1.3 *De Novo* Protein Structure Prediction 585
4.2 RNA Secondary Structure 585

5 Analysis of Molecular Interactions 587
5.1 Protein–Ligand Docking and Drug Screening 588
5.2 Protein–protein Docking 590
5.3 Other Molecular Docking Problems 591

6 Molecular Networks 591
6.1 Metabolic Networks 591
6.2 Regulatory Networks 594
6.3 The Virtual Cell 595

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 6
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30548-3

7 Analysis of Expression Data 595
7.1 The cDNA Microarray 596
7.2 Configuration of Experiments and Low-level Analysis 596
7.3 Classification of Samples 597
7.4 Classification of Probes 598
7.5 Beyond cDNA 598

8 Protein Function Prediction 599
8.1 What is Protein Function? 599
8.2 Function from Sequence 600
8.3 Protein Interaction Networks 602
8.4 Genomic Context Methods 602
8.5 Function from Structure 603
8.6 Text Mining 604
8.7 Information Integration 605

9 Analysis of Genetic Variations 605
9.1 Genetic Variations in the Human: Analyzing Predispositions 606
9.2 Genetic Variations in Pathogens: Analyzing Resistance 607

10 Basic Bioinformatics Technologies 608
10.1 Databases 608
10.2 Visualization 609
10.3 Internet 610

11 Can we Trust Bioinformatics Predictions? 611

Acknowledgment 612

References 612

Keywords

Metabolic Network
Graph structure that displays connected metabolic interactions in one or several organisms.

Molecular Docking
The computational analysis of the geometry and binding affinity of a complex to two interacting molecules.

Ontology
Structured vocabulary for talking about a specific subject, such as diseases or protein function.

Phylogeny
Mostly tree-like arrangement of several related molecules or species in order to display their evolutionary relationship.

Protein Interaction Network
Graph structure that displays pairwise interactions between proteins. The nature of the interactions can be diverse (molecular, genetic etc.).

Protein Structure Prediction
Computational generation of a three-dimensional model of a protein, given its sequence.

RNA Secondary Structure Prediction
Computational generation of a model of the base pairings in an RNA molecule.

Regulatory Network
Graph structure that displays connected protein interactions that regulate cellular processes such as gene expression.

Sequence Alignment
The stacked arrangement of several molecular sequences in rows such that matching characters in different sequences are arrayed below each other in columns. Sequence alignment is performed in order to display correspondences between characters in different sequences. Correspondences can be of an evolutionary or structural nature.

Text Mining
Automatic extraction of biological content from the computer-accessible literature.

Virtual Cell
Comprehensive computer model of the structure and inner workings of a cell.

In recent years, biological research is increasingly being based on large-scale experimentation that collects data on an organismic scale. These data are voluminous and often very noisy. Not just their interpretation but also the configuration of the involved experiments necessitates complex computer analysis. The respective computer methods are themselves an object of intensive research in a scientific discipline called *computational biology* or *bioinformatics*. Computational biology has a wide variety of facets ranging from experiment configuration and low-level data analysis to hypothesis generation by computer.

1 Introduction

The term *Bioinformatics* denotes the field that employs computer-based methods to investigate the molecular basis of living systems. The foundation for the field is provided by molecular biology data (molecular sequences, 3D structures of molecules and molecular complexes, data on mRNA and protein expression levels in cell etc.). The term *Bioinformatics* is used synonymously for the term *Computational Biology* in several countries. In the United States, Bioinformatics addresses more the infrastructural aspects of the field (accessibility of data and methods, data consistency and quality control, networking, etc.) while the scientific aspects are addressed by *Computational Biology*. Also, the term *Bioinformatics* suggests a focus on the molecular subcellular level, whereas *Computational Biology* is more open toward investigating higher levels of biological organization [1, 2].

Biology has experienced a dramatic shift in recent decades from a science that was mostly phenomenological and taxonomic in character to a field that, today, is mainly based on quantitative analysis. The molecular basis of living systems has moved more and more into the focus of biological research. In addition, in recent years, a type of experimentation has gained significance that aims at the large-scale collection of data. Rather than analyzing very local systems such as a single gene or protein and its effects, this type of experimentation, which is also called *screening*, aims at covering a substantial part or all data of a certain kind that can be obtained for an organism. The first screening projects addressed the genome sequence and started in the late eighties with the decision of the US Congress to sequence the complete human genome. It took roughly 15 years to complete this task. Along the way, the experimental technology had to be developed, and it was "tried out" on a series of more and more complex organisms (*see* Table 1 for the list of genomes whose completion marked the progress in genome sequencing technology). As this text goes to print, there are 209 published complete genomes (most of which are prokaryotic), and 956 genome sequencing projects are under way (See http://www.genomesonline.org/ for an up-to-date status of genome sequencing projects). The sequencing of small genomes does not involve major bioinformatics efforts (though the interpretation of the genome sequence does). However, as the genomes become larger and more complex, assembling the sequence data involves substantial bioinformatics. Major bioinformatics efforts were needed to assemble the genomes of *Drosophila melanogaster*, *Mus musculus*, and *Homo sapiens*. Section 2 of this chapter summarizes the bioinformatics needs of genome sequencing projects.

The availability of a complete genome sequence immediately raises a multitude of research questions that can be addressed with the help of bioinformatics methods. In Fig. 1, these questions are arrayed in order from the genotype (five o'clock direction) to the phenotype (counterclockwise up to the four o'clock direction). Some of these questions can be addressed without resorting to additional experimental data; for other questions, additional experimental data are needed. We will now quickly review these questions and go into more detail in the succeeding sections.

Tab. 1 Landmark genome sequencing projects.

Organism	*Colloquial name*	*Genome length (in kb)*	*Date sequenced*
Haemophilus influenzae	Bacterium (ear infections)	1830	1995
Escherichia coli (K12)	Bacterium in the intestine	4639	1997
Saccharomyces cerevisiae	Bakers yeast	12 069	1997
Caenorhabditis elegans	Nematode, worm	97 000	1998
Drosophila melanogaster	Fruitfly	137 000	2000
Arabidopsis thaliana	Mouse-ear cress	115 428	2000
Mus musculus	Mouse	2 497 000	2002
Homo sapiens	Human	2 866 000	2003

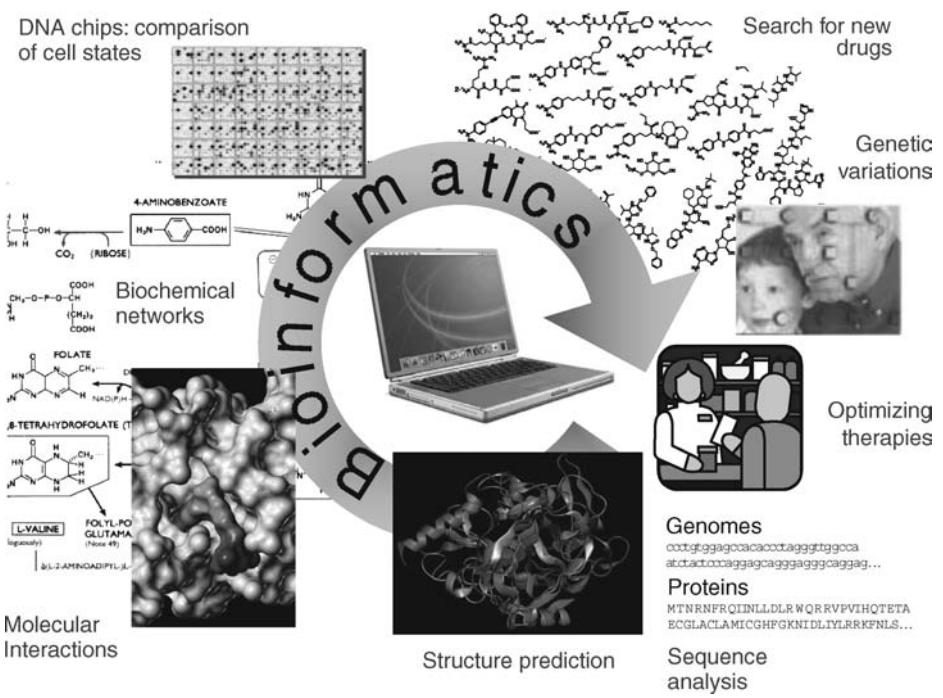

Fig. 1 Bioinformatics – from the genotype to the phenotype.

Among them are the identification of genes in the genome sequence and the analysis of noncoding regions (Sect. 3). The prediction of protein structure from protein sequence is another important problem that deserves a separate section (Sect. 4). The analysis of molecular interactions on the basis of the structure of the involved molecules is the foundation for understanding binding properties, and also serves as a starting point for drug design. We detail this field in Sect. 5. Processes of life are composed of complex biochemical networks made up of pathways and cycles of intermolecular interactions. The bioinformatics analysis of these networks is the topic of Sect. 6.

Since all cells of an organism have the same genome, we need additional data in order to differentiate between different tissues and cell states. mRNA expression data are a current-day approach to measuring this kind of information. Section 7 discusses bioinformatics methods for analyzing these data. The determination of protein function is the Holy Grail of modern biology, in general, and of bioinformatics, in particular. Protein function signals are contained in many diverse biological data. Thus, protein function prediction is an integrative task, which will be discussed in Sect. 8. Section 9 summarizes bioinformatics approaches to analyzing genetic variations in organisms, both in the human and in its microbial and viral parasites and infectious agents. Section 10 comments on the basic information infrastructure technologies relevant to bioinformatics. Finally, Sect. 11 wraps up this chapter with a general appraisal of the potential and limitations of bioinformatics.

2 Sequencing Genomes

The basis of the bioinformatics support for sequencing genomes is quickly explained. Experimental methods for genome sequencing can only handle a sequence of a few hundred base pairs length at a time. Therefore, the genomic or chromosomal sequence has to be cut into fragments of about that length before sequencing can begin. (The fragments can be longer, but then only their ends can be sequenced.) After the fragments have been sequenced, the original genome sequence has to be pieced together from the fragments. It is this assembly process that requires critical bioinformatics support.

The basic fragment assembly problem is quite simple. The sequencing process generates multiple coverage of the complete genome sequence (or of a well-defined part of it that has been identified in the previous process of constructing a physical map of the genome) with small fragments of several hundred base pairs in length. The coverage of the genome sequence must be sufficient to ensure with high probability that the segments overlap along the complete genome sequence (e.g. for the human genome, a coverage of 7X was required [3]). The fragment assembler then aligns the fragments by comparing their sequences to form the complete genome. If adjacent fragments are unique and overlap along a stretch of sufficient length, then the alignment is also unique. The only variations between different fragments covering the same part of the genome sequence can originate from reading errors or polymorphisms (in a heterozygous individual) and there are only a few of those. Therefore, the alignment problem is much simpler than the alignment of different (especially of evolutionarily only distantly related) sequences. Fragment assembly experiences its major complication by the occurrence of long repeats in the genome sequence. It is evident that in the presence of repeats the fragment assembly process just described is corrupted. For a more detailed overview of fragment assembly, *see* [4].

Fragment assembly has experienced its culmination and most dynamic progress in the process of sequencing the human genome. The human genome was sequenced by two different groups concurrently that used different experimental procedures. Both groups presented a draft of the human genome sequence in February 2001.

The method used by the public consortium is sometimes called *hierarchical*

shotgun (HS) sequencing [5]. Using this method, genomic DNA is cut into fragments by partial digestion with restriction enzymes. These fragments are then used to produce libraries of BAC (bacterial artificial chromosome) clones. These clones are between 150 and 200 kb in length. A physical map of these BACs is constructed in order to identify those clones that minimally overlap and are therefore suitable for sequencing. The selected BACs are subcloned into shotgun libraries (libraries of short randomly cut pieces of the clone) and sequenced. Finally, the genome is reconstructed using data from the overlapping sequences, guided by the map. The HS approach provides an accurate finished genome, but requires substantial additional preliminary laboratory work in creating the BAC libraries. The draft assembly of the public consortium was computed by Jim Kent at the University of California, Santa Cruz [6].

An alternative method, used by Craig Venter's group at Celera Genomics to provide a draft sequence of the human genome is whole-genome shotgun (WGS) sequencing [3]. In this approach, shotgun sequencing is applied to the whole genome, by directly sequencing clones that are 2, 10, and 50 kb in length. Additional information comes from known markers, which are used for anchoring assembled contigs (continuous stretches of sequences arising from overlapping clones) to the genome. The genome assembler for Celera was written by Gene Myers and his bioinformatics team [7, 8]. (Other algorithms for fragment assembly in shotgun sequencing have been presented in [9, 10].)

In general, the choice of the sequencing method depends on the complexity of the genome under study and the desired level of completeness of the genome sequence. The WGS approach is well established for shorter genomes without many repeats, especially for prokaryotic genomes. For long and repetitive genomes, the WGS approach is effective in obtaining a draft covering, say 80%, of the genome sequence. Finishing sequencing of a genome is difficult with this approach.

The WGS approach was since applied to drafting the dog genome [11] and probes from the Sargasso Sea containing many organisms [12]. A hybrid approach, which utilizes aspects of both methods, was used for finishing the human and the mouse genome [13]. The public consortium finished the human genome sequence in April 2003.

3 Molecular Sequence Analysis

There are a wide variety of questions that one can ask about a genome sequence. There are low-level analyses, such as analyzing the information content of a sequence or finding repetitive patterns and high-level analyses such as finding genes and analyzing the upstream regulatory regions. There are widely used packages for performing a variety of low-level and high-level analyses, such as the GCG Wisconsin package [14]. In this section, we will only address the most important and widespread sequence analysis problems.

3.1 Sequence Alignment

The most basic sequence analysis method is that of sequence alignment. Sequence

```
Hemoglobin alpha-1  1  MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-D     48
                        :.|:|.:|:.|.|.||||  :.:|.|:|||.|::::.:|.|:.:|..| |
Hemoglobin beta     1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD     48

Hemoglobin alpha-1 49 LSH-----GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR     93
                      ||.     |:::||:|||||.:|::::::||:|::.::::::||:||.:||:
Hemoglobin beta    49 LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLH     98

Hemoglobin alpha-1 94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR    142
                      |||.||:||:::.|:..||.|:..||||:|:|:.:|.:|:|:..|:.||:
Hemoglobin beta    99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH    147
```

(a)

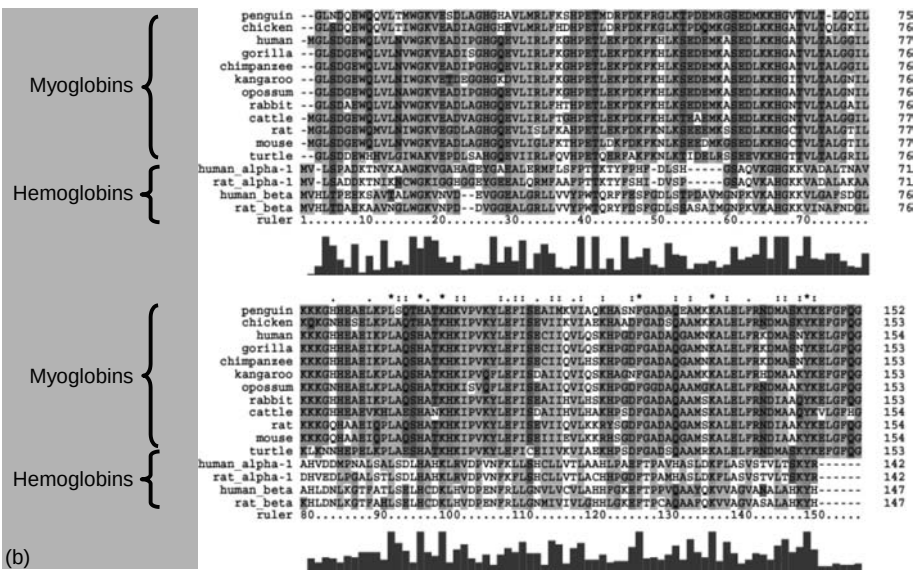

Fig. 2 (a) Global alignment of the human hemoglobin α_1 and hemoglobin β sequences. The numbers at the left and right side of the alignment represent sequence positions at the end of each row. The signs in the middle row signify similarity between the matched amino acid residues: | identical, : similar,. not similar. (b) Multiple alignment of several myoglobin and hemoglobin sequences. The alignment was generated with ClustalW. The colors represent chemical groups of amino acid residues. The bar graph at the bottom of the figure represents the degree of conservation of the respective alignment columns.

alignment arrays two (pairwise alignment) or more (multiple alignment) similar nucleic acid or protein sequences such that corresponding nucleotides and amino acid residues respectively are matched to each other. *Correspondence* can mean different things, in this context. Often, we mean *evolutionary* correspondence in the sense that the two monomers matched have the same evolutionary origin. But sometimes we will also mean *structural* correspondence in the sense that the two matched monomers play the same role inside the molecular structure. The sequence alignment method can be applied to DNA, RNA, and protein sequences alike. Figure 2(a) shows an alignment of the human hemoglobin α_1- and β-chains. Figure 2(b) depicts a multiple alignment of myoglobin and hemoglobin sequences

from different organisms. The additional gap character (-) serves to introduce insertions and deletions during the alignment such that corresponding positions in the different sequences match each other. There are exponentially many alignments in n and k between k sequences of length n. In order to find the desired alignment, we attribute a cost to each alignment. Then, we look for the alignment with maximum or minimum cost, depending on the scoring function used. Let us discuss cost maximization of a pairwise alignment of protein sequences, for concreteness. Then the cost function is a sum of terms. There is a term for each alignment column that does not contain gap characters (match/mismatch terms), and there are terms penalizing for gaps (gap terms). The cost terms are usually derived statistically. In the evolutionary interpretation of the alignment score, the match/mismatch terms are collected in a so-called substitution matrix and reflect the expectation to see the mutations witnessed by an alignment column. A basic simplifying assumption is that the mutations witnessed by different alignment columns are independent. This results in quite efficient alignment algorithms that are based on the dynamic programming paradigm [15]. The gap terms are usually set heuristically or by fitting them to a dataset of accepted alignments [16]. Pairwise alignments can align complete sequences (global alignment) and they can be used to find shorter sequence fragments in longer sequences (local alignment). The runtime of the alignment algorithm grows linearly in the product of the length of the two sequences to be aligned. If one wants to screen against a large database of sequences, even this is too much. Then one resorts to faster heuristics that are based on gapless alignment of locally similar stretches. The most widely used alignment program BLAST is based on this approach [17]. It owes its popularity not only to its efficiency but also to the fact that it returns a significance value (p-value) with the alignment that quantifies how improbable it is to observe the sequence similarity witnessed by the alignment. The more improbable an alignment, the smaller is its p-value and the more significant the finding. BLAST is probably the only major bioinformatics tool, which returns a significance value that is based on solid statistical theory [18].

Multiple alignments are harder to optimize than pairwise alignments. The straightforward extension of the pairwise gapped alignment algorithm leads to an exponential procedure and raises the problem of having to derive many cost terms based on scarce data. These problems can be circumvented, by using heuristics. PSI-BLAST [19] uses a heuristic based on BLAST in order to find sequences related distantly to a query sequence and align them multiply. PSI-BLAST takes a query sequence and computes a multiple alignment of sequences taken from a sequence database. A statistically well-founded approach to multiply aligning n given query sequences is provided by the so-called Hidden Markov models (HMMs), a stochastic procedure by which a certain Markov chain is trained to a family of related protein sequences. The result is a probabilistic weighting of the state transitions and output behavior of the Markov chain that causes it to generate new sequences of the same protein family with significantly higher probability than other sequences. This chain can then be used to multiply align protein sequences of that family and decide whether new chains belong to the family or not [20]. With HMMs, an alignment of k sequences of length n can be obtained in time $O(kn^2)$. Other very

popular multiple alignment programs are ClustalW [21, 22] and T-COFFEE [23] (*see also* Sect. 3.2). Databases of manually curated multiple alignments can be used to benchmark alignment methods [24, 25].

Sequence alignments are the basis of analyzing relationships between DNA, RNA, and protein sequences, respectively. On the DNA level, alignment algorithms are used to find genes [26] and analyze regulatory regions [27]. On the protein level, the relationships to be searched for can be evolutionary [28–30] or structural [31, 32]. The description of the kind of relationship looked for is contained in the scoring matrix. Whereas early approaches used the same amino acid–substitution matrix for all alignment columns, newer approaches use so-called position-specific scoring matrices (PSSM), that is, they adapt the scoring matrix to the specific needs of an alignment column. For example, in a structure alignment, each column stands for a local structural neighborhood in the two proteins to be aligned. This neighborhood has specific features that can be reflected by a scoring matrix tailored to this neighborhood. The same phenomenon can be taken into account, when aligning DNA sequences of upstream regulatory regions of genes, in order to find binding sites for transcription factors. Therefore, position-specific scoring matrices are used in this context, as well [33, 34]. HMMs are the basis of a number of protein family classifications [35–37], and they have also been applied to gene finding and to the analysis of promoter regions [38].

3.2 Phylogeny Construction

In phylogeny construction, one is interested in learning about the evolutionary relationship of a set of sequences. The basic idea is to generate a tree whose leaves are the sequences to be related and whose edges have appropriate lengths such that the unique path between two sequences in the tree approximates the respective evolutionary distance as accurately as possible. Figure 3(a) shows a phylogenetic tree of the myoglobin sequences used in Fig. 2(b). It is desirable to annotate the interior vertices of the tree with "ancestral" sequences such that the sequences that annotate the vertices adjacent to an edge align with a score that equals the length of the edge. Then the tree corresponds to a multiple alignment of the given sequences and the inferred ancestral sequences.

There are several ways of constructing a phylogenetic tree. One approach is *distance-based*. Here, the evolutionary distance between two sequences can be represented by the score of their pairwise alignment or by some other measure of similarity (e.g. the distance can be estimated with a maximum-likelihood method). A tree is searched for that best approximates the resulting distance matrix of the set of sequences involved. The optimization problem of finding a tree that best fits this distance measure given by the pairwise alignments of the sequences to be related, is difficult (NP-hard) and thus can only be solved approximately [15]. The problem of finding such a tree is intimately related to the problem of finding a high-quality multiple alignment of the set of sequences. Thus, there are also methods that combine these two tasks. For instance, ClustalW (see also Sect. 3.1) constructs a multiple alignment and a phylogenetic tree bottom-up by successively clustering sequences that are most closely related. The interior nodes of the tree are annotated with sequence profiles that give the preferences for amino acids (or nucleotides)

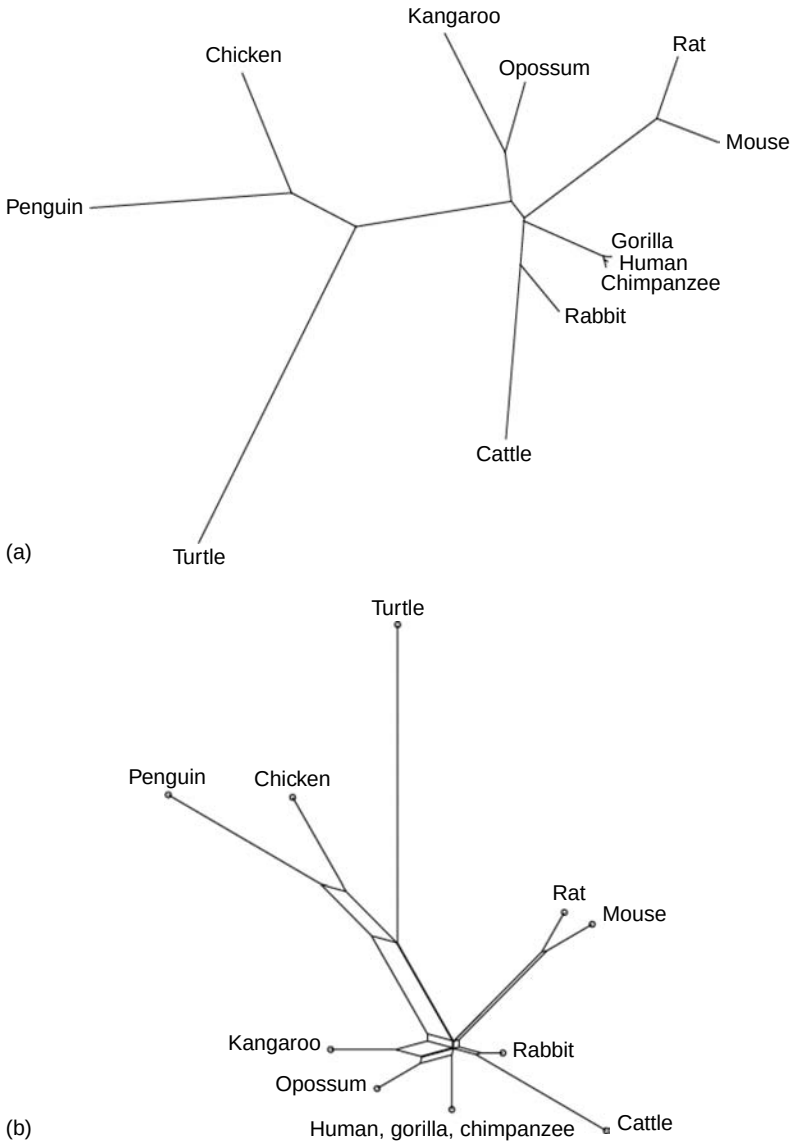

Fig. 3 (a) A phylogenetic tree of the sequences aligned in Fig. 2(b). (b) A split diagram of the more closely related sequences from Fig. 2(b).

at each position in the respective cluster. These profiles can be turned into "ancestral" sequences with consensus methods.

The *maximum parsimony* approach aims at inferring a tree that explains the evolutionary relationship between a set of sequences by introducing as few mutations as possible. This is a strong assumption that is not realized by nature in general but serves to simplify the

problem of generating phylogenies. The *maximum likelihood* approach is based on a probabilistic model of evolution and aims at generating the phylogeny that is most likely, given the observed data. It is generally much more time-consuming than the other approaches but has the advantage of being founded on a statistically more sound and generalizable theoretical basis [39].

PHYLIP (http://evolution.genetics.washington.edu/phylip.html) and PAUP* (http://paup.csit.fsu.edu/) are two widely used software libraries for constructing phylogenies (For a list of phylogeny programs, see http://evolution.genetics.washington.edu/phylip/software.html).

Phylogenies are hard to validate, because the biological process that generates phylogenies is hard to model and observe. This process has two players, namely, the stochastic process of *sequence variation* and the highly complex and unstructured process of *selection*. Because of the complexity of this situation, each accepted phylogeny is the result of a consensus among biologists rather than an imperative consequence of the given sequence data. Specifically, (a) the alignment score reflects evolutionary distance only incompletely and (b) the hypothesis that the phylogeny is tree shaped is not valid in many biological settings, for example, because recombination events can take place. The split-decomposition method [40] takes account of the fact that the distance measure derived from the sequence alignments need not be treelike. The method generates so-called split diagrams that are more treelike if the underlying distance measure allows for this, but that show deviation from the tree shape where they occur. Figure 3(b) shows a split diagram of the myoglobin sequences from Fig. 2(b).

A comprehensive up-to-date overview of methods for constructing phylogenies is given in [41]. A short recent overview can be found in [42].

3.3 Finding Genes

Finding genes is one of the grand challenges of computational biology. Figure 4 depicts the structure of a eukaryotic gene. Identifying genes entails the elucidation of all parts of the gene, not just the coding regions. Exons have to be distinguished

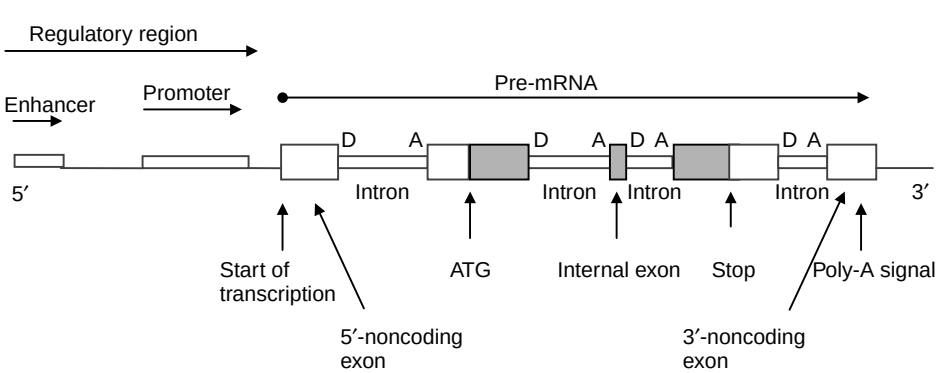

Fig. 4 The structure of a eukaryotic gene. Only the shaded parts represent coding regions.

from introns; transcription and translation, initiation and termination sites have to be identified.

There are basically two different approaches to identifying genes in genomic sequences, *content sensors* and *signal sensors* [43]. Content sensors aim at classifying DNA segments into types such as coding and noncoding. The so-called *extrinsic* (or *homology-based*) content sensors do this by comparison of the DNA sequence with biologically confirmed known gene sequences. Sequence alignment methods are used for this purpose. With the growing number of genomes, one can try to use this approach to render the gene finding procedure more accurate [44] or even to capture aspects of the evolutionary development of genes [45]. If sequences similar to the query sequence are not available, one has to distinguish the gene regions on their own right, that is, by *intrinsic* properties of short sequence fragments such as compositional bias or codon usage (*de novo* approaches). Statistical classifiers use these properties as input and are trained on biologically characterized sequences (*see* e.g. [46, 47]). HMMs are also used to model intrinsic properties of gene sequences [48].

The signal sensor approach tries to locate functional sites in genes such as splice sites, binding sites for transcription factors, and so on. A popular approach is to describe the neighborhood of a functional site with a position-specific weight matrix that is trained on a multiple alignment of the neighborhoods of confirmed functional sites [49, 50]. The alignments used for deriving such matrices are gapless. By using HMMs, one can introduce gaps in the respective alignment [50]. In fact, most of the current-day programs for finding functional sites use this technology.

The accuracy of gene prediction programs can be measured on the nucleotide and on the exon level [51]. For mammalian sequences, the accuracy of the best gene prediction programs in correctly predicting coding nucleotides is above 90%. On the exon level, the accuracy is markedly lower. Less than 80% of the exons had all their exon boundaries predicted correctly. The correct prediction of complete genes is even lower: less than 50% of predicted genes correspond exactly to the actual genes [52, 53]. The problem lies in predicting the exact limits of the coding regions. Also, gene prediction accuracy drops dramatically as the density of genes decreases and the introns become longer. Overall, exact gene prediction is still an unsolved problem.

3.4 Analyzing Regulatory Regions

Regulatory regions are the regions in DNA that lie in the (mostly upstream) neighborhood of the genes and that contain binding sites for the transcription machinery that is responsible for expressing the gene. Therefore, these regions harbor crucial information on the regulation of gene expression. Polymorphisms in these regions are responsible for many of the small and large differences between organisms of the same species and also for many predispositions and diseases. Thus, their understanding is of great importance. The most intensively studied regulatory regions are the promoters that harbor the binding sites of transcription factors (TF). Promoters exhibit a modular architecture, that is, they are composed of several TF-binding sites. Each binding site is a short segment of DNA (length 10 to 15 kb), and different binding sites are located at

varying distances to each other. The characteristics on each of the blocks can be captured with a weight matrix (see also Sect. 3.3). The modular architecture has to be captured with specifically tailored organizational models. Today, such models tend to be tailored to specific classes of promoters and are not generally applicable. Therefore, statistical methods that analyze intrinsic sequence properties are also employed [27]. The methods used are not unlike those for gene identification. Since TF-binding sites are short and highly conserved, a special approach to finding them is based on the analysis of almost identical short sequence segments in a set of DNA sequences [54, 55]. The methods used in practice today are reported to find up to about half of the existing promoters but output a significant number of false positives [56].

Enhancers, silencers, and matrix-attached regions are other types of regulatory regions that are attempted to be predicted with similar technology. Compared to protein-based bioinformatics, the analysis of regulatory regions is still in an early stage of development. It is especially difficult because the processes of molecular recognition based on proteins binding to DNA are very intricate and cannot be analyzed on a structural level yet. Thus, we have to derive all information from the rather indirect methods of sequence analysis and comparison.

3.5 Finding Repetitive Elements

As we have already seen in Sect. 2, repeats are a characteristic and complicating feature of many complex eukaryotic genomes. They present a problem with genome assembly, but they also require specific methods of sequence analysis. A few programs have been developed specifically to find different kinds of repeats in genomes.

The function of repeats in genomes is not yet well understood. Palindromic repeats (those where a sequence and its reverse complement are arranged back to back) may point at structural features such as hairpins in RNA. Repeats also have been shown to affect bacterial virulence in infectious diseases.

There are a variety of repeat finding programs. The most obvious kind is based on the observation that repeats tend to throw off sequence analysis programs; thus, they are located and masked out in the first step. Repeat masking programs such as RepeatMasker (http://ftp.genome.washington.edu/RM/RepeatMasker.html) use a dictionary of known repeats and find all exact or approximate matches to the dictionary entries in the genome sequence under investigation.

Repeat masking programs can only rediscover repeat patterns. A more challenging task is to find repeats, just given the genome sequence.

A very important subtype of repeats is tandem repeats. Here, two exactly matching short sequences immediately follow each other. There are programs that specialize in finding tandem repeats [57].

The largest challenge is to find general repeats without prior knowledge, just using the genome sequence as a basis. Such programs usually find exact repeats in a first step. These are then used as seeds for finding so-called degenerate repeats that are similar but not exactly alike. The programs REPuter [58] and MUMmer [59] store the genome sequence in a data structure called a *suffix tree* that enables the fast identification of exact repeats. The FORRepeat program uses so-called *factor oracles* for this purpose. An important

feature of repeat finding programs is an effective visualization of the output.

Repeat finding programs can be exercised for a multitude of tasks. (1) Checking genome assemblies. Assembly programs (Sect. 2) are imperfect and thus the sequence assemblies they produce may contain errors. Repeat finding programs have detected palindromic repeats in the human genome sequence that were due to wrong assembly. (2) Low copy repeats in human diseases. Several human diseases are associated with deletions or duplications of specific genomic regions. Such repeat patterns can be identified by a repeat finding program. (3) Checking uniqueness of hybridization probes. This application is technological. On a microarray, sequence probes that are unique in the genome under consideration can be deposited. This uniqueness can be ensured with a repeat finder. (4) Comparative genomics. Concatenating two different genomes and looking for repeats with varying rates of divergence allows for comparing the two genomes on different grades of similarity.

3.6 Analyzing Genome Rearrangements

With the availability of a growing number of genomes, comparison of different genomes becomes an even more valuable source of insight into the evolutionary development of species. It turns out that nature rearranges genomes in intricate ways to form new genomes. The indivisible operations that nature uses for rearranging genomes are translocation, that is, integrating some sequence fragment at another location in the same or a different chromosome, reversal, that is, translocation and integration into the oppositely directed DNA strand, as well as fusion and fission of chromosomes. We are concerned here with rearrangement events on the species level, not those on the level of the individual (which are mutation and recombination).

The comparison of two genomes from different species aims at exhibiting which parts of the genome 1 correspond to which parts in genome 2 and thereby reveal the respective sequence of rearrangement operations. Of course, we have to deal with mutations that distinguish the two sequences at the same time. In general, this is a sequence analysis problem similar to sequence alignment, in spirit, but with the rearrangement operations as basic operations instead of mutation and gap insertion. We can attribute a cost to each operation, as we did in sequence alignment, and optimize the cost of the sequence of operations transforming one genome sequence into the other. Algorithmically, the problem is more complex than that of sequence alignment ([60] Chapter 10, [61]). A basic ingredient of the algorithms is the breakpoint graph, which represents the available options for a rearrangement.

The problem becomes even more difficult if several genomes are to be compared and we are looking for a phylogeny [62]. These algorithmic developments still make several biologically unjustifiable assumptions. For instance, recent comparative investigations of the genomes of human and mouse have revealed that rearrangements of genomes happen at different scales. Intrachromosomal micro-rearrangements affect sequence segments less than 1 Mb in length; macro-rearrangements affect much longer sequence segments and can happen inside a chromosome as well as between chromosomes. Recently, the algorithmic

technology has been adapted to this finding [63].

Computational analysis of genome rearrangements is just on the verge of transcending a basic algorithmic investigation to generating useful insight into the evolution of species genomes [64, 65] and the relationship between genome rearrangement and disease [66].

4 Molecular Structure Prediction

In many cases, the 3D structure of a molecule can provide essential clues to its molecular function. This is the case especially for proteins and RNA that exercise their functional role by binding to other molecules in a unique and predictable fashion. This happens especially for enzymes and molecules that are involved in regulation and information transfer within the cell or between cells. While DNA has a very uniform double-helical structure, RNA and proteins can fold in a large variety of structures. These molecules are stable in the sense that they retain their overall fold, yet they can vary their structure in subtle ways. Examples of this phenomenon are domain movements of allosteric proteins or the phenomenon of induced fit, by which the surface of the binding site of a protein changes subtly to accommodate its binding partner.

The close relationship between molecular structure and molecular function gives the structure prediction problem its great relevance. Today, we cannot reliably predict the structures of proteins or RNA, in general. However, the protein structure prediction field is in a more advanced state than RNA structure prediction. We will discuss both fields, in turn.

4.1 Protein Structure Prediction

The protein structure prediction problem has often been called a *Grand Challenge* problem in computational biology. It has received wide attention in the scientific community and even in the media. Computer manufacturers have used it as one of the main motivators for their high-performance computer developments. The reason for this popularity is probably that the problem is quite easily communicated, while a solution still evades us to this day. Here, we discuss three versions of the problem that have received wide scientific attention. All versions take the protein sequence as primary input. Some of the methods use additional input, such as the evolutionary neighborhood of the query protein or a database of known protein structures.

There are a large number of structure prediction methods available on the Internet, today. The community also provides sophisticated means of constantly evaluating these methods. Every two years the CASP contest (critical assessment of structure prediction methods [67]) stages an international experiment for predicting protein structures that are unknown at the time of prediction but are resolved by the time of evaluation of the predictions. Furthermore, there are servers on the Internet that continually assess and compare the performance of protein structure prediction methods [68, 69].

4.1.1 Secondary Structure Prediction of Proteins

The secondary structure prediction problem aims at deciding, for each residue in the protein sequence, whether it is located in an α-helix (H), a β-strand (E), or neither (C). We call the latter regions *coil*

regions. There are variants of the problem that distinguish different helices (α-helix, 3_{10}-helix, etc.) and subdivide coil regions further, for example, by introducing different turns. Here, we will not discuss these variants.

In modern methods, the first step is to scan through a protein sequence database, retrieve homologs of the query protein, and compute a multiple alignment of all retrieved proteins. This provides critical additional input to the structure prediction procedure: For each sequence position, we now have not just a single residue but a profile of residues that related proteins provide. The introduction of this critical information was the key ingredient of boosting the structure prediction accuracy above 70% in the early nineties [70]. Given this information, a variety of classification methods can be used to predict the residues states H,E,C; among them, neural nets [70, 71] are the most popular variant. Additional improvement on the structure prediction accuracy can be obtained by running several methods and forming a consensus between them [72–74]. Such methods achieve an accuracy of just under 80%. The ends of secondary structure elements tend to be predicted more unreliably than their central regions. For the profile to be unbiased, the sequence database from which we draw the profile has to be unbiased in the sense that there are no protein families that are over- or underrepresented in the database.

The accuracy of secondary structure prediction methods is assessed by comparing their prediction to secondary structure assignments that are automatically derived from resolved 3-D protein structures. Different methods for this purpose, such as DSSP [75] and STRIDE [76] differ in their assignments. Owing to these differences, one cannot expect the prediction accuracy to increase much above 90%.

4.1.2 Similarity-based Protein Structure Prediction

The most successful approach to the prediction of 3D protein structure takes a database of resolved protein structures, the so-called template structures as secondary input. Rather than assembling the protein structure of the sequence from scratch, we repeatedly ask the question whether the protein structure attains any of the template structures. We only collect reasonably different structures in the template database – one convention is to require the sequence of any two template structures to have at most 40% sequence identity. This results in a database with roughly 5700 structures today.

How do we decide how likely it is for the sequence of the query protein to attain a given template structure? We perform a pairwise alignment of the query sequence with the sequence of the template structure. Here we use a position-specific amino acid substitution matrix (PSSM) that represents the structural preferences for a residue on the query sequence to be located at the location of the matched residue in the template structure [31, 32]. As in secondary structure prediction, we can enhance this approach by introducing information from evolutionarily related proteins in the form of a sequence profile on the side of the query sequence as well as the template sequence plus additional information, for example, stemming from secondary structure predictions [77].

This kind of aligning of the query sequence with the sequence of a template structure is called *protein threading* [78]. The score of the alignment tells us the level of preference for the query sequence to attain the template structure. Ideally, the

template that can be aligned to the query sequence with the highest alignment score should provide the structural model for the query protein. However, since alignment scores reflect structural preference only inadequately, this model selection procedure is fallible. Thus, the score has to be accompanied with a confidence value that rates how much we can believe in the prediction. Often, confidence scores are based on statistical significance, which rates how unlikely it is to obtain the alignment by chance. But other, more heuristic choices of confidence values have proven effective as well [79].

The improvement of the quality of protein structure prediction rests mainly on the improvement of the scoring function. There are two ways to improve the scoring function. The first is to find ever more suitable functional forms for scoring. In scoring, one has to make a tradeoff between the number of terms to include into the function and the computational efforts for scoring. Interaction terms between a growing number of partners within the protein afford more accurate scoring (if the interaction terms are correct) but also take more time to compute. The second is, given a fixed functional form to improve the scoring parameters. These parameters are typically fitted to existing protein structures. Thus, with the growing number of protein structures available, these parameters can be refitted regularly and should automatically lead to better scoring functions, given that the functional form for scoring is suitable.

The performance of protein threading methods is typically assessed by *fold recognition* benchmarks [68, 69]. Here the object is to retrieve the template structure that is most similar to the structure of the query protein. The performance can be quantified in terms of the number of correctly assigned folds or, in a more detailed fashion, by rating the quality of the alignment, on which the fold assignment is based. The accuracy of protein threading methods depends greatly on the similarity between the query sequence and the template protein. It is low in the case of low sequence similarity and high in the case of high sequence similarity. Over the whole protein structure database, we can achieve an accuracy of above 70% correctly assigned folds today.

Aligning the query sequence to a template protein is only the first step of producing a full-atom protein structure model for the query protein. While it is a critical step for the similarity range of below about 40% between query and template sequence, it is a trivial exercise in high similarity ranges above, say, 70%. A faulty alignment leads invariably to a wrong protein model.

The alignment only provides the structure of part of the backbone of the protein. Gaps in the alignment represent parts of the query sequence that we cannot map onto the template structure (if gaps occur in the template sequence) or tears and rips in the backbone model of the query sequence (if gaps occurs in the query sequence). The former gaps mostly stem from loops in the query protein that have no counterpart in the template structure. These loops have to be modeled in a separate loop-modeling step. Modeling loops can be achieved by taking loop templates from a database of structures of protein fragments, or by using energy optimization methods. Rips in the backbone of the query protein have to be mended and finally the side chains of the query protein have to be attached to the backbone. Here, the variants are to use a database of side-chain rotamers or to do energy optimization.

Similarity-based protein modeling tools combine these different steps in different ways and using different algorithmic procedures such as discrete optimization, nonlinear optimization, and constraint programming [80–82].

While similarity-based modeling is quite successful today, it has to be mentioned that this approach cannot discover yet unseen protein structures. Rather, it can only rediscover structures that have been seen before as attained by different protein sequences. As we can only assume to have uncovered about a third of all protein structures used by nature [83], this approach has strong limitations.

Therefore, homology-based modeling has been combined with experimental procedures for protein structure prediction in *Structural Genomics* initiatives [84]. The basic idea is that with bioinformatics methods, one first selects a set of yet structurally unresolved protein sequences that (1) are likely to fold into a yet unseen structure and (2) are likely to submit to the experimental technique for the structure elucidation that is to be used, that is, the protein should crystallize, if X-ray diffraction is to be used for structure determination [85, 86]. The selected proteins are then submitted to experimental structure elucidation. For this purpose, high-throughput versions of experimental structure determination methods have to be developed. Any newly discovered protein structure can then provide a new template for homology-based structure prediction to further expand the space of protein structure models. With the most ambitious structural genomics initiatives, the goal is no less than to cover all (or as much as possible) of protein structure space. The throughput of experimental structure prediction methods is not yet large enough to consider high-throughput structure determination a reality today [87].

4.1.3 *De Novo* Protein Structure Prediction

De novo structure prediction does not use structural templates but attempts to construct the protein structure from the sequence from scratch. This affords not only the rediscovery of known template folds attained by new protein sequences but the actual discovery of new (never seen) folds. Currently, there are basically two approaches to this problem. One uses energy minimization. This approach is of quite limited effect. The only reports have been on being able to model short stretches (a few dozen residues) in small proteins. In the late nineties, David Baker and his group invented a second approach that takes up the idea of modeling the protein structure from structural templates of small protein fragments (nonapeptides) that are taken from resolved protein structures [88, 89]. This assembly procedure is equipped with an effective scoring function [90]. This ROSETTA procedure has given the field a major push and provided significant progress in recent CASP events [91].

The ROSETTA method has also been applied to improve secondary structure prediction [92].

4.2 RNA Secondary Structure

In principle, the aspects of RNA structure prediction are similar to those encountered with proteins. However, the state of RNA structure prediction is not as mature as that of protein structure prediction. There are two aspects in which RNA structure prediction differs from protein structure prediction. The first is that the secondary structure patterns of RNA are different

from those of proteins. Whereas we have α-helices and β-strands in proteins, RNA secondary structure is made up of double-helical base pairing patterns within the RNA sequence. Thus, the secondary structure prediction problem for RNA differs from that of proteins. The second is that 3D structure prediction

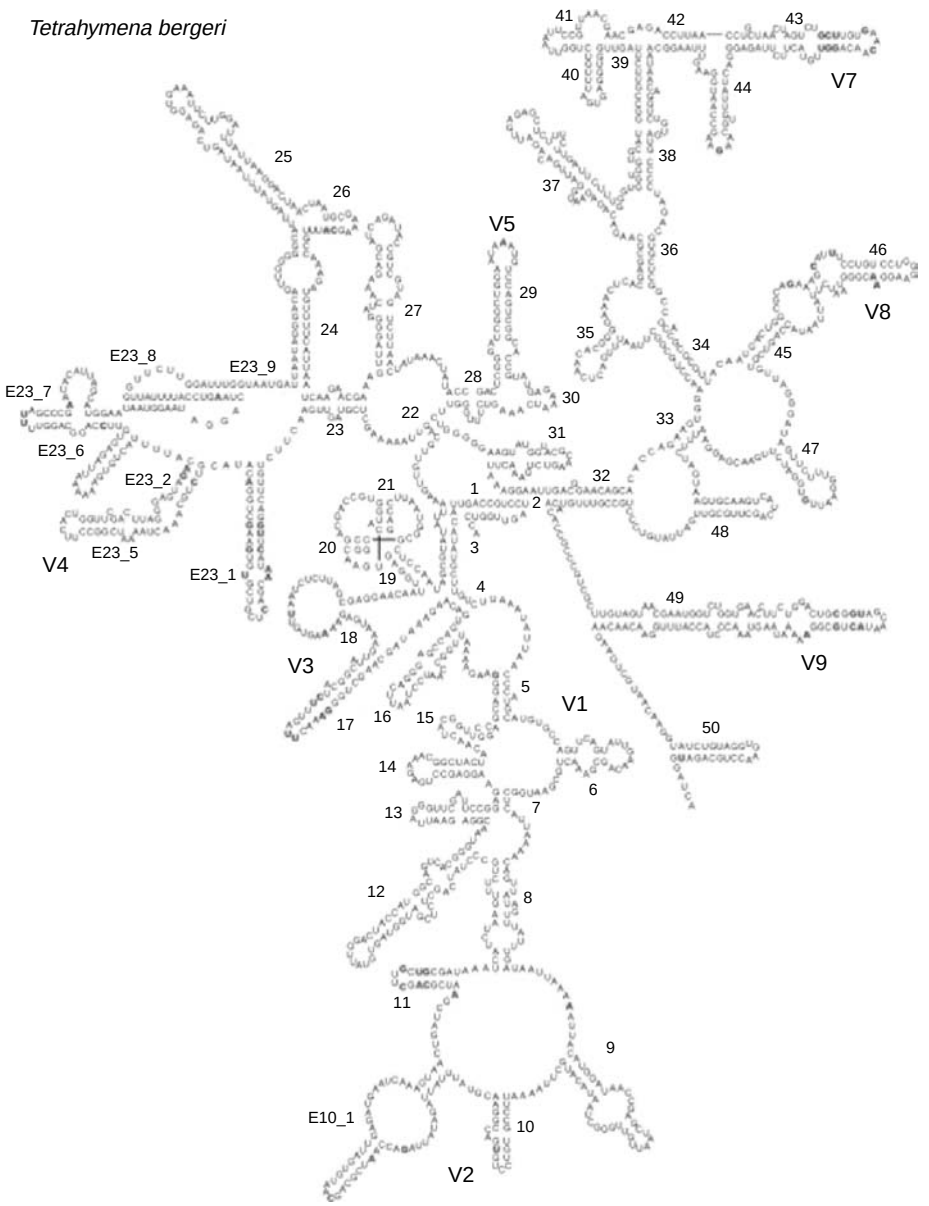

Fig. 5 Secondary structure of the RNA from the small ribosomal subunit of *Tetrahymena bergeri*. The long lines denote pseudoknots.

of RNA is hampered by a lack of data for resolved RNA structures, as compared to proteins. We have over 20 000 protein structures available today, but only a few hundred RNA structures (see http://www.rnabase.org/).

An intensively studied problem on the way to RNA 3D structure is the RNA secondary structure prediction problem. RNA exhibits secondary structure due to the base pairing occurring within the molecule. The usual Watson–Crick base pairs (A–U, C–G) occur with preference, but other pairings (U···G, A···G) and even (U···U) are also seen. Figure 5 shows a typical RNA secondary structure and indicates that RNA secondary structure basically has a tree shape, if one disregards the usually few base pairings violating this shape. (These base pairings are called *pseudoknots* [93, 94], see Fig. 5.)

There are basically two approaches to RNA secondary structure prediction. The first is called *comparative sequence analysis* and deduces the likely location of base pairings from correlated mutations in the respective positions along multiple alignments of homologous RNA sequences. Traditionally, the process was mostly performed by hand, but recently algorithmic versions of this scheme have been presented [95–97]. The second approach is energy-based. An additive energy model is assumed in which each local structural feature, such as a stacked base pair and each of the various bulges and loops in the secondary structure (*see* Fig. 5), is attributed a certain energetic contribution. These energetic values are deduced from experiments and simple model assumptions. Then, a combinatorial optimization algorithm searches for the RNA secondary structure with lowest energy. The optimization can be done efficiently using dynamic programming, if we assume that the RNA structure is treelike as described above. This method has been pioneered by Zuker and Stiegler [98], and it is quite popular, since it lends itself to many generalizations. One can select different tradeoffs between the complexity of the energy model and the runtime of the minimization algorithm [99]. One can compute not only the optimal but also a set of near-optimal foldings and display them together in a so-called *energy dot plot* in order to identify the putatively most stable regions of the structure [100]. Finally, since the energy contributions from local structure elements can be chosen to be temperature-dependent, one can assemble a "movie" of the energetically most favorable RNA structures as the temperature rises (or falls). Also, thermodynamic quantities such as the heat capacity can be computed [101] with the same efficient dynamic programming approach. All of these methods disregard pseudoknots; however, these have to be factored in by special algorithms that make the computation more expensive [102, 103].

5 Analysis of Molecular Interactions

The analysis of molecular interactions is a major component of understanding protein function as well as a basic technology for searching for new drugs. By *interaction* we mean noncovalent (only in rare cases does one have to consider the formation of covalent bonds) binding of two or more molecules. Here, we only discuss interactions between two molecules. We distinguish large molecules (macromolecules, such as proteins, DNA or RNA) and small molecules (such as metabolites or drugs). Here, we will mainly discuss the case in which the

macromolecule is a protein. Accordingly, we consider the protein–protein docking problem, which considers the formation of a complex between two proteins, and the protein–ligand docking problem, which analyzes the complex of a (small-molecule) ligand binding to a protein. Both problems are surprisingly different, in terms of their relevance and their bioinformatics solution.

5.1 Protein–Ligand Docking and Drug Screening

All docking problems are vehicles for understanding molecular interactions from a basic research point of view. In addition, the protein–ligand docking problem is of central interest mainly because of its application in drug design. This is why protein–ligand docking receives much more attention from pharmaceutical industry than the protein–protein docking problem. This application also places important requirements on protein–ligand docking software. Specifically, the demand is on high throughput: In order to screen large databases of putative drug molecules, each instance of the protein–ligand docking problem has to be solved in as short a time as possible. Today, the relevant compute times for drug screening by docking are in the second-to-minute range for docking programs that analyze full ligand and restricted protein flexibility. The ligand is generally assumed to be druglike, with molecular weights up to about 500 Da and up to a dozen rotatable bonds. Peptides, which can be larger and have more rotatable bonds, exercise the limits of general protein–ligand docking software and require special approaches.

Two main issues have to be accounted for in a docking program. One is to find the geometric fit between the two interacting molecules. This entails finding the right conformation of the – generally highly flexible – ligand and the protein (involving the analysis of induced-fit conformational changes in the protein if the software allows for it). In general, a docking program will generate several putative protein–ligand complexes that are feasible geometrically. Therefore, we need a second selector in order to distinguish the correct binding mode(s) from wrong ones. This selector is an estimate of the differential free energy of binding between the complexed and the uncomplexed molecules. According to the laws of thermodynamics, we are looking for the conformational state that minimizes this figure. In addition, the energy score can be used to distinguish good from bad lead compounds; the lower the energy score, the stronger the binding affinity of the ligand to the protein and more interesting the ligand is as a putative lead compound. Therefore, energy scoring is at the heart of docking. Unfortunately, energy scoring is also the least well-understood part of docking. In addition, in order to abide by the strict time constraints, generally reduced energy models have to be used that result in inaccurate energy estimates. Figure 6 exemplifies the result of docking a small-molecule ligand into a protein.

There are a handful of widely used docking programs (e.g. DOCK [104], GOLD [105], FlexX [106], Fred (OpenEye Software), AutoDock [107], LigandFit (Accelrys), ICM [108]) that use a few successful approaches to subproblems of protein–ligand docking in different combinations. Several recent reviews compare these programs [109–112]. For handling ligand flexibility, some programs precompute a limited set of rigid conformers that are docked into the

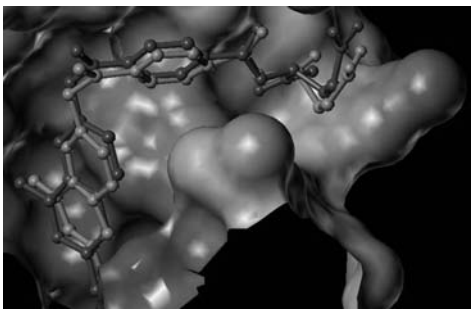

Fig. 6 Docking of the drug methotrexate into the active site of the protein dihydrofolate reductase. Green: observed crystal structure. Red: docking Result computed by the docking program FlexX.

binding site of the protein one after the other. Other programs incrementally build up the ligand inside the binding pocket of the protein out of more or less rigid fragments. Again, others use more generic conformational sampling schemes based on genetic algorithms, Monte Carlo techniques, or molecular dynamics. For energy, scoring force fields that are based on pair potentials are used as well as more rudimentary and therefore faster *empirical* scoring functions that only score additive contributions of short-range interactions between the protein and the ligand [113–115]. Among the pair potentials, the more classical force fields are derived from physical principles [116, 117], newer ones are based on an abstract functional form and fitted to the data of existing protein–ligand complexes. The latter class of scoring functions is generally called *knowledge-based* scoring functions [118, 119]. Improving scoring functions is the bottleneck for protein–ligand docking.

Docking programs that build up the ligand inside the binding pocket of the protein lend themselves to more efficient screening procedures on combinatorial ligand libraries. Here, the alternatives of the so-called R-groups in the ligand scaffold can be tried out during the buildup procedure. This can increase the screening throughput significantly [120, 121].

As an aside, it shall be noted that there is a substantial spectrum of methods for drug screening in case that the structure of the protein is not available. In this case, the query for the search through ligand databases is not the protein but a compound that is known to bind to the binding pocket of the protein. The search looks for compounds that are similar to the query compound. The notion of molecular similarity is critical here. It is usually based on one or a set of molecular descriptors, that is, abstractions of the molecule that entail information about relevant molecular properties. Different kinds of descriptors are in use. Popular descriptors consist of binary vectors that list (sometimes up to a hundred thousand) molecular properties and code, for each property, whether the molecule has that property or not. In this case, molecular similarity analysis is reduced to a comparison of binary vectors, which can be done very quickly. Thus, the resulting screening procedures have a throughput of several ten thousand molecules per second [122]. More complex descriptors involve a topological analysis of the molecule. Such descriptors incur more complex similarity analysis algorithms that only have a throughput of about 100 molecules per second. However, with such descriptors, the conformational

flexibility of ligands can be handled [123]. Overviews of approaches to drug screening by molecular similarity can be found in [124–126]

These initial drug-screening procedures only look for *lead compounds* that are (more or less) tightly binding to the protein. In order to convert a lead compound into a drug, its so-called *ADMET* properties have to be optimized. Here A stands for absorption (by the intestinal barrier, if given orally), D for distribution (throughout the body, involving deliverance to the location where the drug is supposed to act), M for metabolism (the tendency of drug to be metabolized before it reaches its place of action), E for excretion (of the drug from the body) and T for toxicity. Computational procedures for assessing these properties are summarized in [127, 128]

5.2 Protein–protein Docking

Protein–protein docking is central to understanding regulatory events in living systems. Therefore, the protein–protein docking problem is a vehicle in understanding the inner workings of cells and the molecular basis of diseases. Since proteins are not preferred drugs, these days, the protein–protein docking problem cannot help in finding new drugs, though it can be an aspect of finding target proteins – for instance, when the binding of two proteins is supposed to be inhibited by a drug. The screening aspect is not as important for the protein–protein docking problem as for the protein–ligand-docking problem. One could think of docking a protein against all structurally resolved or modeled proteins of an organism, but the number of resulting instances is still orders of magnitude smaller than for drug screens.

In protein–protein docking, the notion of molecular surface is more prominent than in protein–ligand docking, because both binding partners are basically rigid – aside from the events of induced fit occurring along the contact surface. These events need to be considered. Not considering protein flexibility incurs higher penalties in protein–protein docking than in protein–ligand docking. The energetic differences between the bound and unbound complex are usually higher for drugs (that are supposed to stick to the protein) than for proteins (that are supposed to disassociate again). All this makes protein–protein docking an exceptionally hard variant of the docking problem.

Approaches to protein–protein docking usually build on a procedure for rigid docking and incorporate induced fit on top of this algorithmic basis. Methods based on Fourier analysis or Fourier transform [129], geometric matching [130], and genetic algorithms [131] are approaches for the rigid-docking part. The induced fit can be incorporated rudimentarily by "softening" the protein surface using appropriate energy potentials or explicitly by sampling different side-chain conformations. In general, the energy-scoring problem is the same for protein–protein docking as for protein–ligand docking. If a rudimentary scoring function is used, the ranking of the different complexes that are generated by the docking program by energy is inaccurate, and rescoring of the high-ranking complexes with a more accurate scoring function may be necessary. Explicit sampling of side-chain conformations can be done by a combinatorial search procedure [132] or with genetic algorithms [133]. The dead-end elimination

algorithm is an effective method for reducing the number of conformers to be searched [134, 135].

State-of-the-art protein–protein docking procedures are ICM-DISCO [136], 3D-DOCK [137, 138], and ZDOCK/RDOCK [139, 140]. These programs and others took part in the CAPRI protein–protein docking contest (Critical Assessment of Prediction of Interactions http://capri.ebi.ac.uk/ [141]) that was started in 2001 [136, 142, 143]; the contest is going into its fifth round now. Reviews of protein–protein docking can be found in [144–146].

5.3 Other Molecular Docking Problems

Protein–DNA docking is important in understanding regulatory events in the cell. Because of the high regularity of the DNA double-helix structure, the usually large structural changes in the protein and the importance of the solvent in protein–DNA binding, this version of the docking problem is an especially difficult variant of the protein–protein docking problem. A few attempts have been made to adapt existing protein–protein docking software to protein–DNA docking [146] or to develop methods specially tailored to this problem [147, 148].

The other variants of the docking problem (protein-RNA, DNA-ligand, RNA-ligand) are covered even less in the literature.

6 Molecular Networks

While the interactions of pairs of molecules are essential basic components of cellular function, the inner workings of a cell only become apparent at the level of complex networks of such interactions. Therefore, the modeling and analysis of biochemical networks is a major aspect of understanding the cell, as a whole, a goal that has been formulated by the newly appearing field of systems biology [149].

Biochemical networks in cells are very complex. Therefore, their study is usually confined to restricted contexts. The classification that will guide the structure of this section is divided into metabolic networks and regulatory networks. Both subclasses of networks have their specially targeted databases and a growing set of analysis procedures. A more general view of protein interaction networks is afforded by cell-wide protein interaction screening experiments. This topic is taken up in Sect. 8.3.

6.1 Metabolic Networks

The metabolic network of a cell summarizes the set of reactions inside the cell that metabolize small molecules (called the *metabolites*) with the help of catalytic proteins (called *enzymes*). Coherent sequences of successive reactions are called *pathways*; if a pathway closes in on itself, it is called a *metabolic cycle*. The object of a pathway or a cycle can be to synthesize or degrade a macromolecule or a metabolite or to convert or release usable energy. Metabolic networks are very uniform. They are composed of elements as shown in Fig. 7, with two vertices called the *educt* and the *product*, an edge signifying the reaction labeled with the catalyzing enzyme, and possibly attached cofactors. Reactions can be reversible or irreversible, the latter case being denoted by a pointed arrow in the unique direction of the reaction. The composition of such elements at the metabolite nodes

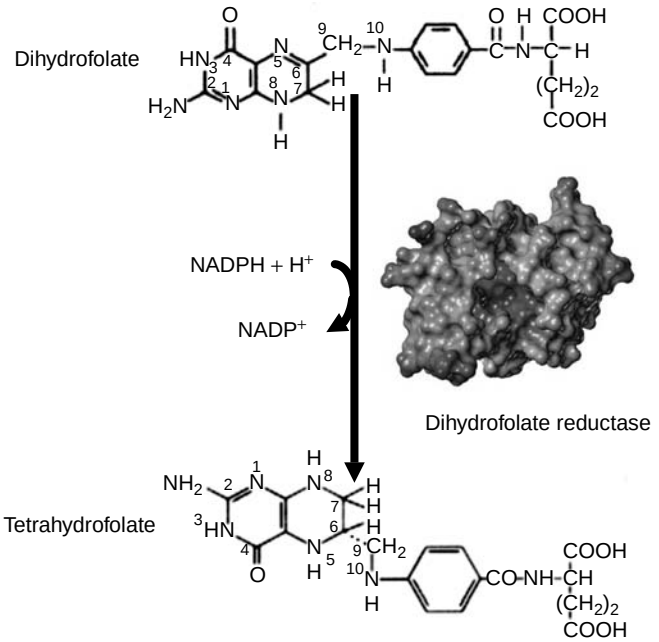

Fig. 7 A metabolic reaction. Educt: dihydrofolate. Product: tetrahydrofolate. Enzyme: dihydrofolate reductase (DHFR). This is the reaction blocked by the inhibitor bound to DHFR in Fig. 6.

yields complex networks such as shown in Fig. 8.

The regular nature of metabolic networks affords the opportunity to build large uniform databases and to devise systematic analysis procedures of these networks. In order to completely describe a single reaction such as the one in Fig. 7, we have to supply the kinetics of the reaction, that is, the concentration gradients of the involved molecular partners in terms of their starting concentrations. While the form of many reactions is known today, the reaction kinetics is a sorely missing piece of data for most reactions.

Metabolic network databases include organism-specific databases such as EcoCyc [150] (for *E. coli*), HinCyc [151] (for *H. influenzae*), PseudoCyc [152] (for *P. aeruginosa*) (Other computationally derived metabolic databases are available at the BioCyc website (http://biocyc.org)) and databases spanning several organisms ranging even over different kingdoms of life, such as KEGG [153], MPW [154] and MetaCyc [155]. For these databases, the basic unit of description is the metabolic pathway. The databases provide data records and images describing such pathways. In recent times, the images have become increasingly processable by computer. Then there are databases that concentrate on the single reaction, such as BRENDA [156] ENZYME [157], and the new integrated effort IntEnz [158]. All metabolic databases are equipped with their own set of tools for querying the databases, for navigation and for visualization of the search results.

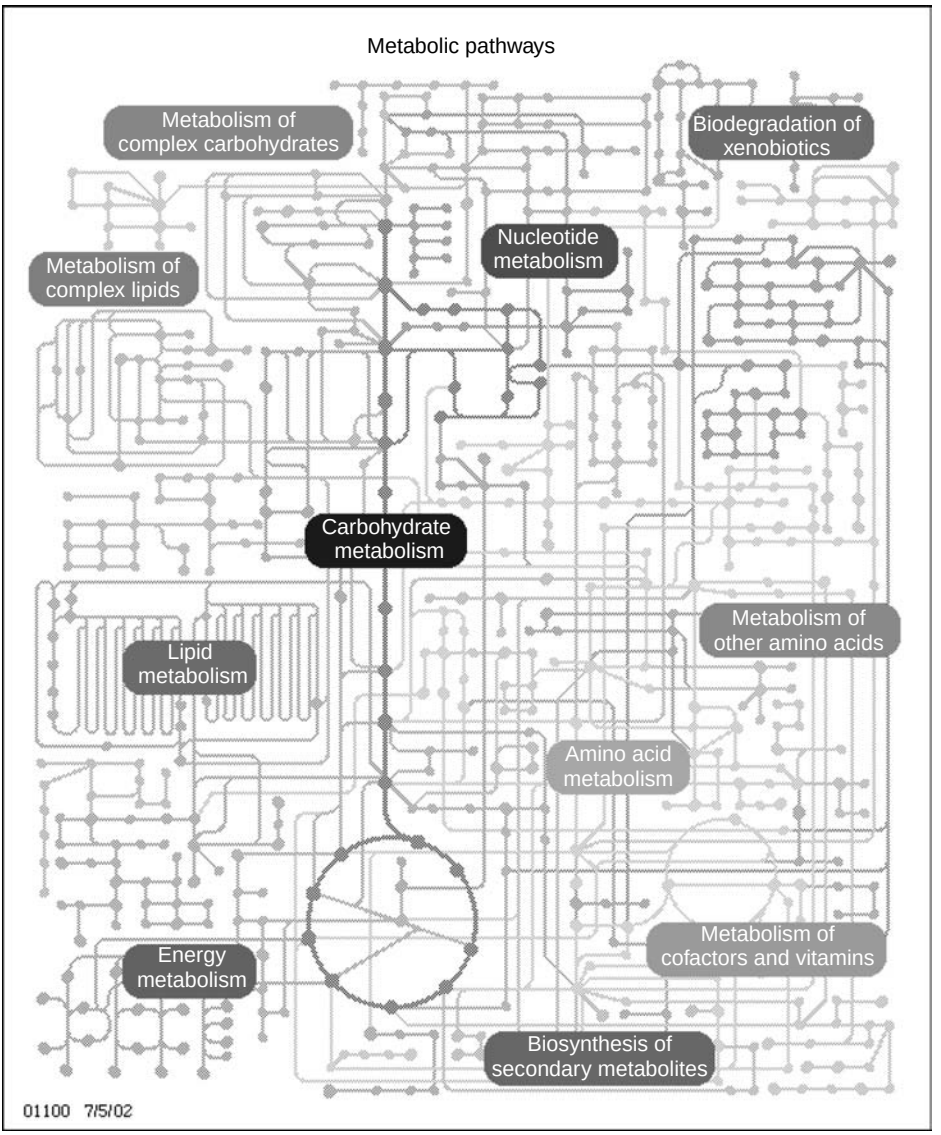

Fig. 8 Schematic diagram of part of the metabolic network of *E. coli* (from http://www.genome.ad.jp/kegg/pathway/map/map01100.html)

These databases are comprehensive enough to enable comparisons of the factual or putative topology of metabolic pathways in different species with fully sequenced genomes. Such comparisons have been undertaken for several metabolic pathways [159], partly involving genomic context methods for protein function prediction (*see* Sect. 8.4). Furthermore, mathematical methods have been developed to analyze the topological structure of pathways. The so-called flux balance analysis

method uses linear optimization to assign flux values to the edges of a metabolic network that maximize some objective function, such as growth per substrate uptake, under steady state conditions [160]. In contrast, metabolic flux analysis decomposes pathways into the so-called elementary flux modes [161, 162]. The elementary flux modes are the minimal components of a pathway that can act coherently and independently under steady state conditions. Both types of analysis aim at yielding insight into the biological structure of a metabolic network from its topology. There is software to perform this analysis on medium-sized pathways [163]. Of course, the steady state is a special case for a cell. If we drop it, we have to resort to more general analysis methods based on ordinary differential equations that model the reaction kinetics in the pathway under consideration. There has been extensive software development for this purpose [164–167]. However, the data are missing in many situations. Large metabolite screening projects aim at removing this deficit [168–170].

6.2 Regulatory Networks

The regulatory network is generally thought of as being the part of the biochemical network inside the cell that facilitates information and communication, rather than turning around matter. Signal transduction pathways and protein interaction networks that facilitate transcription of genes are two essential parts of the regulatory network. The regulatory network is much more difficult to approach than the metabolic network. One reason is that it is more inhomogeneous. In a regulatory context, there are many ways in which a molecule can affect another that is of importance, for example, agonistic, antagonistic, activating, deactivating, inhibitory and so on. Thus, currently, there is not even a generally accepted notation for representing regulatory relationships. Furthermore, the few databases of regulatory networks are more cursory than the well-developed metabolic databases. Where metabolic databases contain up to well over a hundred thousand genes, regulatory databases have grown to a few thousand genes. Most of the knowledge on regulatory pathways is still buried in the literature, that is, it is not in computer-digestible form. However, the amount of knowledge existing in the literature is so large that several attempts are being made at systematically mining the literature with automated methods to extract that knowledge. Of course, owing to unsystematic notation and difficulties in parsing text generated by humans, such efforts generated many false positives and the resulting data are therefore to be taken with the greatest caution (see also Sect. 8.6).

TRANSPATH [171] is a prominent database of signal transduction pathways. The database web interface includes tools for querying and visualization as well as an integrated tool for the analysis of microarray data in the context of the regulatory networks (*see* Sect. 7). The protein interaction databases DIP [172] and BIND [173] also describe molecules in terms of pathways and provide reaction maps.

There are basically two ways of analyzing regulatory networks. One is to simulate and analyze established networks. A straightforward approach here is to model regulatory relationships with the same framework of differential equations that is used for the nonstationary analysis of metabolic networks. This approach is general and can be accurate, given sufficient data, but it may incur performance

problems and lack the model hierarchy that biologists usually attribute to biochemical networks. The signal pathway database DOQCS [174] provides a repository of models of signal pathways on this basis. Qualitative simulation methods simplify the general differential equations model to simpler function forms [175] or even to discrete models [176]. The other, much more challenging, question is how to uncover yet unknown regulatory relationships and pathways from experimental data. One set of data from which one hopes to be able to do so is expression data (Sect. 7).

6.3 The Virtual Cell

The ultimate goal of systems biology is to provide a computational simulation of the whole cell. The *E-Cell Project* is undertaking a comprehensive effort toward this goal encompassing the computational work to provide the simulator [177, 178] and being embedded into a large metabolite screening project that is supposed to generate the necessary data for the simulation [179]. *Genomic Object Net* is another virtual cell project that is based on combinations of discrete and continuous simulation mechanisms [180].

7 Analysis of Expression Data

In the early 1990s, the first steps of developments were undertaken toward measuring the profile of genes transcribed in a cell in a certain cell state. This information goes way beyond the information contained in the genome sequence, since it can help distinguish between different tissues and different cell states within the same tissue (e.g. healthy/diseased). Thus, it was hoped to be the basis for entirely new approaches towards diagnostics and therapy of diseases. Technically, the most convenient way of measuring gene expression profiles is on the basis of the transcribed mRNA, rather than protein. The reason is the much more uniform molecular form of DNA that lends itself more conveniently to laboratory procedures than proteins. There is a tradeoff in choosing mRNA as the level on which gene expression is to be measured. On the one hand, mRNA expression levels do not always correlate well with protein expression levels, which are governed by regulation of protein translation and degradation, in addition to transcription regulation. Also, posttranslational modifications of proteins, which have an essential effect on protein function, cannot be measured on the mRNA level. Still, because of the technical accessibility, mRNA expression measurement became very popular, first on the basis of ESTs (expressed sequence tags) and currently mainly on the level of cDNA microarrays.

The advent of cDNA microarrays in the late 1990s has created a large surge of statistics and bioinformatics activities targeted both to configuring the arrays and to interpreting the resulting data. Today, not only the potential but also the limits of microarray measurements are becoming apparent. Thus, many of the hopes of making protein function accessible directly through microarray measurements have not materialized. But it is becoming clear that diagnostic procedures progress significantly through expression measurements and in concert with other biological data, expression measurements can help gain some insight into the biological role of genes and proteins.

7.1 The cDNA Microarray

Without going into the details of the technology, we give a quick summary of the essentials of the cDNA microarray here, in so far, as it is relevant to bioinformatics. The expression profile measurement begins with extracting all mRNA from a homogeneous cell lysate. The mRNA is then back-translated to cDNA and labeled radioactively or fluorescently for detection. The cDNA is then run over a microarray. On the microarray, an arrangement of complementary sequence probes for relevant genes is arrayed. The length of the probes differs in different technologies from 25 mers across about 70 mers to full-length genes. The cDNA hybridizes to the respective probes and a detection device measures the quantity of hybridized cDNA. The detected signal levels provide the raw data for analysis and interpretation.

Today, cDNA microarrays can have many thousand probes on them and collect all genes for small organisms or many putative genes for diseases in complex organisms (heart chip, cancer chip, etc.). There are basically two classes of arrays: one-channel arrays and two-channel arrays. One-channel arrays are the ones described above [181]. Two-channel arrays allow for hybridizing cDNA from two different cell lysates, each colored differently, say, red and green, in the same experiment [182]. Thus, we achieve a differential profile of two cell states with a single experiment. Several companies have proprietary microarray technologies (for an overview see http://www.bio.davidson.edu/Biology/GCAT/GSI/Microarrays.html). The Pat Brown Lab at Stanford University has provided a free protocol for making two-channel arrays that is widely used today (http://cmgm.stanford.edu/pbrown/)

7.2 Configuration of Experiments and Low-level Analysis

The configuration of microarray experiments is a nontrivial technical issue that had been taken too lightly in the early days of the technology. Today, the importance and difficulty of the consistent preparation of a microarray experiment has gained widespread attention [183]. This includes rigorous quality control of the purified RNA populations, the appropriate selection, and configuration of the probes on the chip to reliably select the desired cDNAs and avoid inadequate or nonspecific hybridization [184], a sufficient number of replications of the experiment [185], and a sufficient logging of the experiment. For the latter purpose, the MIAME standard (minimum information about a microarray experiment) has been developed [186]. The public repository of expression data ArrayExpress, which follows this standard, is in existence and growing in volume [187]. The National Center for Computational Biology (NCBI) is providing another database of expression data, the Gene Expression Omnibus (GEO) [188].

The configuration of a microarray experiment is closely related to the low-level analysis of the data generated by the experiment [189, 190]. By this we mean the procedure that turns the raw readouts of the experiment into a number signifying expression level, say, in terms of RNA copies per cell, or in terms of fold changes based on some reference profile. Low-level analysis has to incorporate and appropriately model the variances in

expression levels stemming from technological and biological sources of error. The respective statistical procedure is called *normalization* of expression data. Quite a few methods for normalizing expression data have been published, and the choice of the appropriate method is a central aspect of low-level analysis. Overviews of basic normalization methods are given in [191, 192], newer methods can be found in [193, 194].

The result of low-level analysis is an *expression data matrix* of normalized expression levels. This matrix has the dimensions number of probes × number of samples. Here the number of probes (genes) is usually very large, in the several ten thousands. The number of samples (experiments) is quite low, at most in the dozens. Thus, the expression data matrix is quite unbalanced. This has strong ramifications for the analysis of expression data.

The final step in low-level analysis is to decide what a differentially expressed gene is. To this end, we have to fix a threshold value of change in expression level, above which we regard a gene as being "conspicuous." This is a statistical question that has to be resolved in terms of a model of background variation specific to the gene and the experiment.

7.3 Classification of Samples

The considerably easier problem is to classify the (few) samples on the basis of the expression levels of the (many) probes. The related application problem is a diagnostic one: to distinguish different types or stages of a disease based on the related expression data. This problem is of high clinical relevance and has been investigated as soon as the microarray technology became available. An especially widely studied type of disease in this context is cancer. One of the first studies on classifying microarray samples targeted two types of leukemia (acute myeloid leukemia, AML and acute lymphoblastic leukemia, ALL) that are hard to distinguish histologically [195]. Since then, many studies on different types of cancer followed.

For classifying samples, a variety of statistical learning techniques can be used. Supervised learning techniques use labeled data – in this case, samples that are known to originate from patients that have contracted a certain type of the disease. Then a statistical classifier is generated that classifies future samples. Learning methods include decision trees, neural nets, support vector machines, and others [196, 197]. Depending on the types of disease to be distinguished, high classification accuracy can be achieved, sometimes over 90%. This means that the type of disease can be determined quite reliably on the basis of mRNA expression data. This has been substantiated quite impressively on a wide variety of cancers [198]. An important part of supervised learning is feature selection, that is, the identification of small sets of genes whose expression levels are indicative of the type of disease. Often, one can identify a few dozen genes whose expression levels can determine the disease type with almost the same accuracy as the classifier operating on the expression levels of all genes. Genes selected in this way can also be candidates for primary actors in the disease process. Feature selection is a more general way of determining genes that are relevant to a disease than detecting differentially expressed genes, since it takes sets of genes into account instead of just single genes.

In contrast to the supervised learning techniques, unsupervised learning techniques cluster the samples into groups that may reproduce known disease types or uncover new ones [199].

7.4 Classification of Probes

Elucidating the role of genes in the disease process using expression data is much more difficult and even hopeless, in many cases. The reason is again the shape of the expression data matrix. Now, we have to learn about the differences between a large number of genes based on very few experiments – a very difficult situation. Nevertheless, there have been various approaches in this direction.

Again, people tried clustering the genes first. The rationale was that coexpressed genes might be in functional association. It became apparent soon that such a procedure is quite inaccurate [200]. Therefore, quickly the idea came up to analyze expression data together with the upstream regulatory regions of the genes. After all, it is these regions that determine transcriptional regulation. The rationale now was that coexpressed genes that share functional sites in their upstream regulatory regions should be under the same or a similar regulatory regime and therefore share a functional relationship [201–205]. Recently, this idea has been extended to include an explicit model of the involved regulators [206].

Upstream regulatory regions are not the only kind of information that one can use to enhance the analysis of expression data. Functional context as exhibited by membership in a common metabolic or regulatory pathway can also be exploited. By analyzing the expression levels of genes in the same pathway in concert, one can obtain a statistical model of this pathway. If this model is significant in the statistical sense, this can be an indication that the pathway plays some role in the disease process. Furthermore, genes with unknown functions can be tested against the pathway model. A significant test response can indicate a functional relationship of the gene with the pathway [207, 208]. This approach can be extended by including not only data from biochemical networks but also functional relationships witnessed in the scientific literature (*see* Sect. 8.6). Protein interaction data are another supplementary source of information (*see* Sect. 8.3). Such data have been analyzed in concert with expression data to elucidate molecular pathways [209, 210]. Here, the working hypothesis is that the protein product of genes that exhibit the same expression profile tend to interact. In all cases, the analysis methods are again statistical learning schemes. Especially, if several sources of information have to be integrated, the quite general machinery of Bayesian networks proves very effective. However, Boolean networks and other graph models are also used. The relevant statistical learning procedures can come up with regulatory networks that may be quite abstract and not closely linked to the molecular regulation process [211]. Therefore, their interpretation presents certain problems.

7.5 Beyond cDNA

As mentioned above, mRNA represents an intermediate step from the genotype (DNA) to the phenotype (post-translationally modified protein), which has been chosen for expression profiling mainly for technical reasons. As the

experimental technology progresses, arrays are being developed that measure gene expression closer to the phenotype than mRNA, and even analyze aspects of protein function. Peptide and protein arrays are one such technology. Here, the goal is to affix functional proteins on the chip and expose this chip to a certain ligand for binding, or the other way around, bind a set of ligands to the chip and expose them to a protein solution. Maintaining the original functionality of the molecules while affixing them to the chip substrate is the main challenge [212–214]. Also, carbohydrate arrays are being tested for probing protein–sugar interactions in glycosylation. Finally, tissue and cell microarrays are under development. While the experimental procedures are quite different for each type of array, we expect the bioinformatics procedures to be analogous for many array types. Thus, the bioinformatics technology that is being developed for mRNA arrays now should be extendible to other types of arrays in a comparatively straightforward fashion.

8 Protein Function Prediction

Elucidating protein function is one of the central goals in molecular biology. We know quite little about the function of many proteins. Today, we have hints on the function of at most two thirds of the proteins in an organism with a completely sequenced genome. Many of these hints are based on quite spurious evidence and, in fact, quite a few probably lead in the wrong direction. Thus, while we have many genomes on hand, understanding their biology is far off. This section summarizes the widely varying efforts of approaching protein function with a mixture of experimental and bioinformatics methods. For a recent review, *see* [215].

8.1 What is Protein Function?

Protein function is a colorful term. It can mean quite different things in different contexts. Sometimes, the localization of the protein (cytosolic, membrane-standing, extracellular etc.) has already been taken as a qualification of its function. In another context, the function of the protein is understood in terms of its binding partners and the reaction that it may catalyze. Again, in other situations, the role of the protein in a complex biological process (cell cycle, apoptosis etc.) is taken to be its function. It is clear that coming up with an appropriate ontology, that is, a structured vocabulary for talking about protein function, is the first step toward systematic analysis of this issue [216].

Here, we mention three efforts for providing ontologies for protein function. The oldest is the classification of enzymes by the Enzyme Commission (http://www.chem.qmul.ac.uk/iubmb/enzyme/) [217]. This is a hierarchical classification of enzymes according to the reactions that they catalyze. This classification only applies to enzymes. More recently, a larger variety of protein function has been covered by the MIPS Functional Catalog(http://mips.gsf.de/services/funcat) in the context of the annotation of the yeast genome [218]. The catalog has been extended to other species since then. As per the classification by the Enzyme Commission, it exhibits a hierarchical treelike structure (with up to six levels, where the EC code has four). The most extensive effort of creating an ontology for protein function is by the

international Gene Ontology Consortium (http://www.geneontology.org/) [219]. Again, the ontology is hierarchical. This time, however, we have three hierarchies in order to distinguish between the three fundamentally different approaches to protein function discussed above: cellular component, molecular function, and biological process. Also, instead of the tree hierarchy, a directed acyclic graph is used such that a child node can have different parent nodes. The ontologies are expanded continuously and function annotations of genes and proteins in databases and in the literature are linked to the ontologies. This whole body of knowledge provides an increasingly voluminous gold standard by which computational methods for protein function prediction can be evaluated. Nevertheless, the structure of current ontologies provides only a structured vocabulary and not a guideline for the functional annotation of proteins. Without such a guideline, it is difficult to express functional relationships between proteins, multiple protein function, and function variability of a protein under different physiological conditions. The GO consortium does provide a semi-formal annotation guide with its ontology (http://www.geneontology.org/GO.annotation.html). More work will have to be invested into a highly expressive and computer-digestible framework for representing protein function. As the first larger regulatory networks are being understood [220], initial efforts are being undertaken to develop the expressive tools. An example of an application in a supracellular scenario is given in [221].

8.2 Function from Sequence

Nature's prime method of conserving function is by preserving (or only slightly changing) sequence. Clearly, some parts of a protein are more relevant to its function than others. Thus, there are functionally relevant motifs in protein sequences that have to be especially conserved. There are several methods of protein function prediction that exploit these facts.

In principle, one would expect that two proteins with sequences that are very similar globally also have similar or the same function. This is not always true, however. Function is handed over from one species to another via homologous proteins, those that are closely related evolutionarily. The relationship between sequence similarity and homology is intricate and subject to much debate [222]. In particular, protein function is not always carried over between homologous proteins. We have to distinguish between orthologous and paralogous proteins, in this context. Orthologous proteins occur in different species and share a recent common ancestor. Therefore, orthologous proteins are likely to have the same function. Paralogous proteins can occur in the same species and arise by a recent gene duplication event, which also allows for function divergence. Therefore, an accurate distinction between orthology and paralogy has to incorporate phylogenetic analysis. This is a complex issue, both in the way of models and algorithms [223, 224]. Therefore, various heuristics are in use. A method that only analyzes phylogenetic relationships locally with respect to two preselected groups of proteins is implemented in the program Orthostrapper [225]. Another approach that is targeted toward whole-genome analysis does away with phylogenetic analysis altogether and reduces the notion of orthology to pure sequence similarity search. The model is that two proteins in two different genomes are orthologous if each protein retrieves the

other as the top hit in a sequence database search of the respective genome [226, 227]. This model is only a crude approximation to orthology, and therefore the resulting COGs database has to be manually curated. Another protein database that provides orthology information based on this model is SMART [228].

Orthology is a notion that applies to a pair of complete proteins. However, protein function is not only conferred to a protein, as a whole. Rather, it comes in *packets*. Short stretches of sequence, so-called *motifs* carry information on protein function. For example, the well-known ATP binding site motif, the P-loop, is characterized by a sequence fragment of length 8 respecting the motif [AG]-x(4)-G-K-[ST]. Here, the characters in brackets denote choices of amino acid residues according to the one-letter code at a single sequence position. Single letters represent unique amino acid residues (no choice), and x(4) denote a stretch of four adjacent arbitrary amino acid residues. Motifs carrying information on protein function can be found by multiply aligning protein sequences that are known to have the same function and selecting highly conserved sequence regions in the alignment. Then a motif can be derived as some kind of a consensus of these regions. The PROSITE database was the first collection of this sort [229, 230]. Initially, the motifs were mainly generated by hand. Recently, methods have been put forth to derive motifs automatically [231]. Here, the issue of sensitivity versus specificity is central: a motif that covers many of the sequences with the respective function necessarily also generates *false positives* – protein sequences that contain the motif but do not share the function. Excluding such false positives therefore means reducing coverage, that is, losing *true positives*: proteins that have the function that is to be described by the motif. However, several specific motifs (with few false positives) can together cover many true positives in the sense that each true positive contains at least one of the motifs. A database of motifs based on this idea is described in [232].

Position-specific scoring matrices or Hidden Markov models (*see also* Sects. 3.1, 3.3, 3.4, and 4.1.2) are an even more general descriptive mechanism for motifs. With them, one typically describes longer sequence fragments, such as whole protein domains. They are also easy to derive from multiple alignments. Many databases have been developed around this paradigm [233]. Most of these databases have been integrated in the domain database InterPro [234].

Finally, supervised learning methods – mostly based on neural nets – have been developed for predicting quite a few aspects of protein function including cellular localization (through the analysis of signal peptides and the prediction of transmembrane helices) and posttranslational modification features (glycosylation- and phosphorylation sites). A server offering many such methods is offered by the Center for Biological Sequence Analysis at the Technical university of Denmark, Lyngby (http://www.cbs.dtu.dk/services/). One of the methods called *ProtFun* makes a comprehensive classification of proteins with respect to their function (e.g. enzyme class or participation in a biological process like amino acid biosynthesis or energy metabolism) based on general sequence features that can be calculated (like sequence length, charge, and amino acid composition etc.) and prediction through the other methods on the server. Some of the resulting classifiers achieve a high level of 90% accuracy [235].

8.3 Protein Interaction Networks

In the late 1990s, cell-wide assays for measuring protein–protein interaction were developed. These assays are an extension of the yeast-two-hybrid (Y2H) method [236] to a large scale [237]. Without going into the details of the experimentation, the procedure basically aims at providing a binary matrix of size number of proteins × number of proteins that tells for each pair of proteins, whether they interact or not. With a different procedure called *tandem-affinity purification*, one can latch onto a protein and pull out with it a whole attached complex of proteins. The result is not a binary matrix but a set of protein complexes [238]. These data facilitate the bioinformatics analysis of cell-wide protein networks. The main problem is that both procedures are hampered by significant numbers of false negatives and false positives. A reason for this is that the laboratory procedures do not faithfully recreate the conditions under which the proteins bind in the cell. As a consequence, especially, the laboratory procedures cannot distinguish between transient binding partners and those that bind to each other over longer time periods, nor can they distinguish between binding events in different physiological states. Nevertheless, there have been preliminary attempts to mine the resulting protein interaction data with bioinformatics methods [239–242]. For an overview, *see* [243]. Most of these methods are based on the unsupervised learning paradigm; they are some form of clustering. The idea is that interacting proteins share some function. This guilt by association approach can also point to interesting drug targets [244]. Global topological properties of the networks are investigated, as well [245, 246]. However, it seems evident, that protein interaction data are most meaningfully analyzed in context with other data on protein function (*see* Sect. 8.7). Public data repositories for protein interaction data are also growing in volume [247] [248].

8.4 Genomic Context Methods

Even in the absence of mRNA expression or protein interaction data, there are several ways to learn about protein function from sequence data alone. The main prerequisite is that we need to have the complete genome in order to be able to reason not only about the presence but also about the absence of a protein. Since today there are many completely sequenced genomes available, especially in the prokaryotic domain, these so-called genomic context methods are quite effective. As a rule of thumb, we need roughly 30 genomes to reason about functional associations in proteins on the basis of the genome sequence. For eukaryotes, this number is not quite reached yet. Thus, genomic context methods are not yet effective in the eukaryotic domain.

In general, the genomic context methods uncover functional associations between proteins that are more general than the proteins that are binding partners. The proteins can also be functionally associated by taking part in the same biological process (cell cycle, apoptosis, etc.) or by cooperating in producing a phenotype (genetic association). We now list several genomic context methods.

Gene neighborhood and gene order Two proteins in two different species are more likely to be functionally associated if they occur in close proximity and in the same order along the genome in many species.

There are two variants of this method, one that respects the order in which the genes occur along the genome [249] and the other that does not [250]. A generalization of this method to more general conservation patterns that present cycles of association that alternate between homology and neighborhood has been presented in [251, 252].

Domain fusion This method rests on the observation that two proteins whose genes are fused in some other organism are more likely to be binding partners [253, 254].

Phylogenetic profiles This method rests on the observation that two proteins that are either present together in a species or absent together have a higher likelihood to be functionally related [255]. The cross-species comparison is based on an orthology test that has originally been reduced to a minimum value for evolutionary distance of the two proteins that is supposed to represent orthology. More recent and effective versions use gradual levels of evolutionary distance.

In general, genomic context methods are applicable nicely to prokaryotes but not as much to eukaryotes. The reasons are that too few eukaryotic genomes are fully sequenced to draw solid conclusions from genomic context hand also that the organization of genes in prokaryotes (e.g. their collection in operons) lends itself better to genomic context methods.

8.5 Function from Structure

Protein structure harbors detailed information on molecular function, since it is because of its structure that a protein acts the way it does. Still, the automatic deduction of information on protein function from protein structure is difficult and not extensively developed yet. Protein structure by itself is not indicative of protein function, since folds are often reused for different functions through gene duplications or convergent evolution. On the other hand, the same function (e.g. enzymatic class) can be realized with several protein folds [256]. A detailed orthology analysis can help here (*see* Sect. 8.2). Beyond this, remote homologies can hint at functional relationships, if not identify in function. For instance, two remotely homologous enzymes can share the same reaction mechanism, while substrate specificity is not conserved [257].

Just as sequence alignment is the methodical basis for analyzing the similarities and evolutionary relationships between protein sequences, structure comparison serves this purpose for protein structures. Structure comparison of proteins is mostly performed by structurally superposing the two (rigid) protein structures. The structural superposition of two proteins entails two subproblems. One is to align the two protein chains structurally. This means that we match the amino acid residue pairs, one in each protein that correspond to each other spatially, that is, take each other's place in the protein structure. Algorithmically, this alignment problem is the same as in sequence alignment, but the scoring function now has to represent structural similarity rather than sequence evolution, and is therefore different. Also, it is harder to assume the independence between alignment columns (*see also* Sect. 3.1), since residues that are adjacent in the protein structure influence each other. Several automatic methods for the structural superposition of proteins that have been developed are available on the Internet,

and have given rise to databases of structural protein classifications [258–262]. A recent evaluation of these servers is presented in [263]. The most widely used automatically derived structural classification of proteins is CATH [264] that uses the structural superposition method SSAP [265]. This database is only rivaled by the database SCOP that achieves a higher classification consistency, because it is manually curated [266].

Structural motifs in proteins, notably, specific ligand binding sites, can hint at functional aspects of proteins. Methods for automatically detecting such patterns have been presented in [267–271]. The calculation of statistical significance for the occurrence of a motif is a critical issue, just as it is for sequence search. Progress in this direction has been made, as presented in [272].

The location of functional sites along the protein surface is a very important aspect of bridging the gap from structure to function in proteins. Early methods had concentrated on geometric analysis, that is, they had looked for clefts and invaginations along the protein surface [273]. Physicochemical properties had soon been included in the analysis [274–276]. More recently, the evolutionary conservation of functional residues have been explored to identify functionally relevant regions on the protein surface [277, 278]. An important ingredient of the method is an accurate molecular clock measuring the speed of evolution [279]. The ConSurf Server makes the methods available over the Internet (http://consurf.tau.ac.il/) [280].

8.6 Text Mining

The most voluminous data source for protein function information is the literature but it is also hardest to digest by a computer. Text mining subdivides into two major disciplines, namely, information retrieval and information extraction. Whereas the first task is concerned with retrieving documents according to user-defined criteria, the more ambitious task of information extraction tries to ascertain facts about prespecified types of events, entities, or relationships described in the literature. Text mining methods have been established predominantly in the newswire domain. Whereas articles in that domain aim at a general audience, biomedical articles usually target a small community of domain experts, and, thus, are more difficult to interpret. The adaptation of methods to the characteristics of the biomedical literature has only begun. Available approaches employ natural language processing techniques to varying extents ranging from direct pattern matching approaches [281] to customization of established natural language processing systems [282].

In the biomedical context, one of the ultimate goals is the correct extraction of relationships among biological entities, such as proteins, genes, or diseases. Current text mining methods fail in this general setting owing to (1) general natural language processing problems, for example, tracking references to one object throughout the text or correctly identifying negations or hypotheses, and (2) domain-specific issues, such as the highly variable gene and protein nomenclature. Special purpose approaches with promising performance have been reported, however. For example, the extraction of facts pertaining to protein subcellular localization [283] or protein–protein interactions [282] has been addressed. Recently, an approach combining information retrieval and extraction for determination of protein–protein interactions has been reported, which relies

on machine learning techniques [284]. The system is used as an aid for curating the BIND database of protein interactions. The details of the approaches are beyond this review, but recent overviews are given by [285–288].

Still, text mining is an exploratory field that needs to improve methods and develop an evaluation standard [289]. In general, information retrieval methods provide sufficient performance to be of immediate use [284]. Information extraction results, however, should be assessed in the context of additional data or taken as starting points for database entries maintained by domain experts. To foster an exchange of ideas and determine most suitable algorithms for retrieval and extraction tasks, critical assessment contests must be implemented [289]. Toward this goal, the established TREC (http://trec.nist.gov/) conference on text mining methods has started a genomics track. Also, the BioCreAtIvE (Critical Assessment for Information extraction Systems in Biology) (http://www.pdg.cnb.uam.es/BioLINK/BioCreative.eval.html) competition has been launched in 2003 for comparison of text mining methods in the biomedical domain.

8.7 Information Integration

No single method for predicting protein function can do the trick by itself. Many of the methods even have problems reaching well above the noise level with respect to the accuracy of their predictions. Therefore, a central goal is to combine the different methods, in order to increase the reliability of the predictions. By now, we have collected quite a few methods that can contribute: sequence analysis and motifs, protein interaction data, literature data (extracted by hand or automatically), mRNA expression data, and protein interaction data. There are quite a few efforts to combine these methods. Methods evaluating protein interaction data in concert with mRNA expression data have been put forth [290, 291]. Marcotte et al. have presented the first study integrating several genomic context methods with primary experimental data and mRNA expression data [292]. The Bork group has followed first with two studies, a paper with the center around genomic context methods [293] and one with the center around protein interaction data [294]. More recently, they provided a world wide web-accessible database STRING (http://string.embl.de/) [295] of precomputed functional associations based on genomic context data that they validated in an experiment of recovering functional modules in *E. coli* [296]. All of these studies indicate that the methods are much stronger in concert than a single method by itself [297–299]. Also, domain fusion provides strong signals but is applicable only to a small subset of the proteins. Phylogenetic profiles seem to be the most powerful genomic context method. Surprisingly, mRNA expression data and protein interaction data carry comparatively little signal. Protein interaction data have been put into the context of other sources of signals for protein function by [294, 300, 301].

9 Analysis of Genetic Variations

So far, we have only considered the generic genome of a species and not made differences between the individuals. Of course, there is a lot of significance in analyzing these differences. In the human, the differences explain what distinguishes us

from each other, not only in appearance but, more importantly in our predisposition to diseases and our response to certain drugs. But not just the differences between individual human genomes are relevant. The individual differences in the genomes of the infectious agents (bacteria and viruses) that are the basis of diseases hold the key to understanding phenomena of resistance to drug treatment and, in concert with the genome of the host, to understanding problems or benefits of immune response. Therefore, we have to commit a few remarks to genetic variations in both contexts.

9.1 Genetic Variations in the Human: Analyzing Predispositions

Human genomes differ from each other in about 1 in 1200 base pairs. These differences, the so-called *single-nucleotide polymorphisms* (SNPs) are the data to be mined for the analysis of genetic predispositions for diseases and individuals responses to drug therapy. Monogenic diseases, that is, diseases that are caused by alterations in a single gene, such as sickle-cell anemia (caused by an SNP in the hemoglobin gene), cystic fibrosis, or hemophilia, have been studied for many decades. Roughly 5000 such diseases are known. Information on disease-relevant mutations and the related phenotypes is collected in the OMIM database [302].

Genes for monogenetic diseases can be found by classical methods of statistical genetics. Specifically, there are two such methods. Both of them rely on molecular markers – short, polymorphic stretches of repetitive DNA or SNPs – which vary between people. These polymorphisms can be identified for different people, some of which carry the disease and some of which do not. With the help of statistical procedures, one can ascertain, which instance of the marker is related to the disease. If a significant correlation between a certain instance of the marker and the disease can be detected, this is evidence of the fact that a gene involved in the disease may lie close to the location of the marker. The probability that the gene is involved in the disease decreases with the distance of the gene from the marker. A statistical score, the so-called LOD (log-odds) score quantifies the likelihood of the gene being involved in the disease. Experimental or bioinformatics procedures follow that screen through a high-scoring genomic region, in order to uncover candidate genes and determine their function. The two methods from statistical genetics differ in the sort of data that they analyze. *Linkage analysis* analyzes familial data, that is, pedigrees of families in which the disease is running. These data are usually sparse and hard to come by, but if they are available, they provide more detailed insight into the genetic basis of the disease. *Association analysis* analyzes more voluminous data taken from larger populations, however, without being given any familial relationships [303–305]. Though no monogenetic disease is curable, today, we have a lot of knowledge on the molecular basis of many of these diseases much of which has implications on therapy. Unfortunately, most of the widespread diseases cannot be reduced to a defect in a single gene. Rather, their causes are a mixture of alterations in many genes and influences of the environment. These so-called complex diseases are the center of genomic research targeting genetic differences between individuals. The identification of genes involved in complex diseases is much harder than that of monogenetic diseases [306, 307]. The Iceland genotyping effort is providing a unique database

for studying the genetic basis of complex diseases [308]. This program aims at genotyping the total Icelandic population, in principle, and relating these data to the extensive pedigree information that this country has collected for almost one thousand years. The company deCODE Genetics, Inc. [309] has been formed, for this purpose. The project is carried out in cooperation with the pharmaceutical company Hoffman–LaRoche and has stirred much controversy [310–314] and required special legislation in Iceland [315]. A few years after its start, the project has turned up 15 disease genes and mapped genes linked to more than 25 common complex diseases (http://www.decode.com/).

The next step after identifying genes related to a disease is selecting targets for drug design and developing related drugs. The field that uses genomics for drug-related research, not necessarily regarding inter-individual differences, is often called *pharmacogenomics* [316, 317]. (However, frequently, this notion is also used with the meaning of the word *pharmacogenetics* below.) All experimental and bioinformatics technologies described above can be employed for this purpose. Notably, mRNA microarrays are under intensive investigation [318]. Furthermore, comparative genomics analysis can help transfer knowledge that is relevant for target finding and drug design from organisms that are easier to handle experimentally to the human [319, 320].

As inter-individual differences enter the investigation, the term *pharmacogenetics* [321] is used increasingly frequently, instead of *pharmacogenomics*. (However, there are also different distinctions between the two notions in the literature [322, 323].) Special challenges for bioinformatics in this context are reviewed by Altman and Klein [324]. This group also takes on a major challenge, namely the collection of relevant data in the database PharmGKB (http://www.pharmgkb.org/) [325]. This database collects genomic, phenotypical, and clinical information relevant for pharmacogenomics, and the related project is developing tools for interlinking and querying the data and visualizing the responses, as well as developing strategies for maintaining confidentiality and privacy of critical data. Beyond data handling, bioinformatics analysis of pharmacogenetic data include the investigation of structural and functional consequences of nonsynonymous SNPs, that is, polymorphisms that alter the protein sequence [326]. SNPs in the regulatory regions of genes have a more subtle effect on the binding of transcription factors and are harder to analyze with bioinformatics methods. The author is not aware of any result in this direction.

9.2 Genetic Variations in Pathogens: Analyzing Resistance

It is evident that the analysis of the genome of a pathogen and its relationship to the host can uncover the molecular basis of pathogenicity. There is a multitude of research in this area, too much to survey here. Overviews are presented in [327–331].

Here, we focus on the phenomenon of resistance based in genetic variations of the pathogen. Infectious diseases are characterized by a dynamic and dramatic battle between a population of a microbial pathogen and the immune system of the host. Drug therapy is supposed to either target the pathogen directly or support the action of the host's immune system. In

the course of this process, the pathogen responds to the selective pressure exerted by the drug treatment by preferentially evolving into resistant variants. Understanding this process necessitates an analysis on the genomic level, at best involving both the pathogen and the host. Bioinformatics can help in this process. The research area is quite new. Therefore, we will only present one example relating to HIV/AIDS.

HIV is a rapidly mutating virus that attacks the immune system of the host. Currently, there are almost twenty drugs in the marketplace that target viral proteins. Since the virus changes rapidly under drug pressure, in the *Highly Active Anti-retroviral Therapy* (HAART) scheme, combinations of drugs targeting different viral proteins are given. Still, within weeks, a resistant strain of the virus develops, and a new drug combination has to be selected. This combination should not only be effective against the currently existing viral strain but also make it as hard as possible for the virus to become resistant. The selection of drug combinations can be aided with bioinformatics methods. To this end, two kinds of data are collected: (1) genotypic resistance data relating the genotype of the viral strain prevalent inside the patient to clinical disease measures (here the number of virus particle in one μL of blood serum) and (2) phenotypic resistance data originating from laboratory experiments, in which viral strains are exposed to HIV drugs and the so-called resistance factor is measured. The resistance factor is the quotient of the drug concentration that halves the growth of the tested virus strain divided by the respective concentration for the wildtype. There are several databases collecting resistance information (http://resdb.lanl.gov/Resist_DB/default.htm) (http://hivdb.stanford.edu/) (A new development that has not generated any publications yet is *the HIV Resistance Response Database Initiative* http://www.hivrdi.org/) [332, 333]. Bioinformatics approaches to data analysis take a viral genome as input, and report estimate of effectivity of single drugs [334, 335] and drug combinations, including an estimate of the shortest mutational path to resistance [336].

10 Basic Bioinformatics Technologies

So far, this chapter was organized according to application-oriented problem domains in bioinformatics. At the end of the chapter, we want to take a more methodical viewpoint and summarize basic information technology concepts as they pertain to the field.

10.1 Databases

As has become evident throughout the chapter, data handling, curating, and interlinking is central to bioinformatics. There are a tremendous number and variety of biological databases that contribute to the field, many of them accessible via the Internet (*see* Sect. 11). The databases are developed in very different settings and use different solutions for IT concepts, data curation, and data presentation. This results in an inhomogeneous, if not to say, chaotic data landscape that integrative concepts in bioinformatics have to deal with – an almost unmanageable task.

Several approaches have been followed to integrate the various databases [337].

Link integration This approach leaves the databases as they are and, indeed, where

they are and maintains cross-links between the related data records in different databases, using the application interfaces provided by the databases. This strategy is haphazard, but it is practical since, while it is helpful to cooperate with the providers of the database, it is not necessary to do so. Link integration is the technology for most small and medium integrative bioinformatics projects. The sequence retrieval system (SRS) (http://srs.ebi.ac.uk/) [338] is a variation of this technique. It adds a central keyword indexing and search system to the linking paradigm. SRS is more sophisticated than the usual Internet search engines, because it recognizes the existence of structured fields in source databases and can link between related fields in different databases. The advantages of link integration are also its disadvantages: it is basically an uncoordinated effort of database integration, and therefore it is especially vulnerable to changes of the source databases and notational and data-structural inconsistencies.

View integration This is somewhat more complex than link integration. Again the data are left in their source databases. This time, however, an environment is built around the source databases that makes them all look like one unified database. To this end, a front-end query machine is created that parses a query, decides which databases have to be accessed, generates the subqueries to these databases, retrieves the results, and combines them to answer the query in a unified manner. The most complex system of this sort reported in the literature is the K2/Kleisli System [339]. This approach is not as popular with the users as link integration, possibly because of performance issues.

Data warehousing This is the most complex variant. Here, all data are collected from their source databases in one system under a unified database model. Tools for querying and navigation can now be integrated deeply with the data model and one has maximum control over issues of data consistency and updating. However, the effort to create a data warehouse is gigantic, and the incremental import of new information (and possibly update of old information that has been revised) is a highly nontrivial issue. Also, data models for warehouses tend to evolve, and this causes significant strain on the system, which has to stay operable in the process. An early ambitious data warehouse project was the Integrated Genome Database (IGD) [340], which was aimed at integrating sequence data with physical and genetic maps. This warehouse included more than a dozen source databases, but only held for a year. A current data warehouse project is the Genomics Unified Schema (GUS) project by IBM [341]. This project is more modest in that it only supports a number of in-house research projects and does not aspire to become a public resource.

10.2 Visualization

Visualization of bioinformatics data and analysis results are central to the acceptance of bioinformatics tools, as a whole. There are several issues to visualization.

Molecular visualization In the eighties, molecular structures found their way from the laboratory table into the computer. Beyond the revolutionary technology in hardware and software that afforded us with methods for rendering complex 3D molecular structures under controllable and realistic light conditions, central contributions

to the field included algorithms for computing molecular surfaces [342]. In principle, the graphical representation of molecular structures by computer is a mature field, whose main remaining challenges are in performance increases in order to allow for the interactive manipulation of ever more complex structures and in usability issues.

The situation is a little different if molecular motion is involved. Here, we can benefit from more help in order to visually structure complex modes of motion in large molecules or molecular complexes. Morphing algorithms are employed here that take molecular characteristics, specifically the normal modes of molecular vibrations, into account if only the starting and final state of the motion are available [343, 344]. Similar algorithms do not seem to exist for actual molecular dynamics trajectories.

Visualization of high-dimensional data Almost all analysis procedures in bioinformatics have a serious visualization problem, when it comes to presenting the results of the analysis to the user. For instance, many bioinformatics methods deal with high-dimensional data. The number of dimensions can be extraordinarily large. It can reach into the hundred thousands. Often, statistical procedures are applied to the data. The results then have to be visualized in some fashion. Usually, some kind of dimension reduction is applied, for example, using principal component analysis or independent component analysis [196]. Self-organizing maps [345] and locally linear embedding [346] are nonlinear methods for reducing the number of dimensions in a dataset while preserving neighborhood relationships between the data points. However, in general, the problem of visualizing high-dimensional data sets is far from being solved.

Visualization of graphs Other structures that have to be visualized are large unstructured graphs (e.g. protein interaction networks [347, 348]) or structured graphs with complex application specific boundary conditions (complex metabolic or regulatory networks, *see* [349, 350]). Large trees (phylogenies or hierarchical clusterings [351]) also present a challenge for visualization. Here, zooming and refinement approaches can help. Similar methods are used in genome browsers for covering several orders of magnitude in scale regarding genome sequences. In general, methods of visualization will have to be developed that utilize generic methods but are tailored to the biological application domain.

10.3 Internet

The Internet has become the central medium of bioinformatics. In addition to being the forum for providing literature, data, and methods, it is the backbone for many integrative approaches. Standards have been developed for exchanging data over the Internet, such as the distributed annotation system (DAS), which supports the exchange of genomic annotation information. The genome browser Ensembl (http://www.ensembl.org/) [352, 353] uses DAS to allow third-party providers to add their additional annotations to its canonical annotations of genes. There are two projects extending DAS, BioMOBY, and ^{my}Grid. BioMOBY [354] allows for registering services together with a description of their input and output, in order to better chain bioinformatics procedures.

myGrid [355] goes a step further by using grid computing technology [356] to describe precisely how different bioinformatics services are related, how they can be located and invoked on a large resource of processors available via the Internet. This necessitates the use of ontologies for describing bioinformatics services. Thereby, myGrid does not require a central registration system.

In some sense, genome browsers are the ultimate bioinformatics resource. They are the search engines tailored to biology. Ensembl, the University of California Santa Cruz Genome Browser (http://genome.ucsc.edu/) [357], and the Entrez Genome Site at NCBI (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome) [358] are the most popular genome browsers on the Internet. Ideally, they should assemble all biological information relevant to and derivable from the genome sequence. As it is, they currently focus on gene identification and are still lacking much of the downstream structural and functional annotation. The future will see the browsers continually fill with biological information, as it is uncovered by exercising bioinformatics and experimental methods, in concert.

11
Can we Trust Bioinformatics Predictions?

Throughout this chapter, not only the potential but also the limitations of today's bioinformatics have become apparent. Bioinformatics is plagued by inconsistent and often defective data, by heuristic algorithms that do not find the optimum that they are looking for and by models and scoring function that represent the situation they want to model only inaccurately. As a result, we cannot expect the prediction of bioinformatics procedures to be flawless. Rather, up to now, we have had to deal sometimes with substantial margins of error. There are three ways to cope with this phenomenon.

1. Be aware of it. As long as results from bioinformatics procedures are treated as suggestions and not as facts, we basically do not have a major problem. All bioinformatics predictions have to be thoroughly validated, most of them involving laboratory procedures. Still, even in the presence of high noise levels, bioinformatics procedures can limit the search space significantly, for example, when looking for drug targets or drugs. In the latter case, for instance, this is the basis for the scientific and commercial success of protein–ligand docking and drug-screening programs. Rather than finding the drug candidate unfailingly, they enrich the number of hits among the high-ranking molecules in the screen. This is sufficient in practice (at least for now) and leaves room for continual improvement. The situation is not much different in many other domains of bioinformatics.
2. Invest in data curation. Primary data are at the beginning of the bioinformatics food chain. If they are faulty, the errors will propagate. The landscape of bioinformatics databases is populated with a dominant and growing number of *derived* databases. These are databases that apply bioinformatics procedures to primary data and produce derived data, such as protein structure models, gene predictions and so on. If the primary data are faulty, these errors will propagate through the food chain and spoil the whole bioinformatics meal. Actually, this is today's situation. Attempts are being made to

remedy it by introducing quality standards for data generation (such as the MIAME standard for mRNA expression data) and by investing much effort into curating at least the central primary databases. Also, scientific journals continually adapt and step up their policies for accepting papers reporting on experimental data with respect to quality control of the reported data. Conversely, bioinformatics procedures can help curate primary data. There are quite a few cases in which bioinformatics analyses turn up sanity checks for biological data as a side product (*see*, e.g. [58, 359–361]), and it is confounding how many errors are being found. Nevertheless, the data curation problem remains difficult if not insurmountable and is one of the central bottlenecks of the whole field.

3. Assess the reliability or confidence of a bioinformatics prediction. Doing so requires a significant amount of work. In some very restricted settings, there are theoretical results that allow for deriving a significance value (sometimes called *p*-value). One prominent example is gapless alignment [18] (*see also* Sect. 3.1). In other cases, there is no theory, but heuristic statistical procedures allow for assessing the probability that the resulting prediction is just a product of chance rather than a consequence of some underlying biological phenomenon. This is the case whenever we have a reproducible distribution of scores rating the predictions. Then, we can formulate a null model against which we rate the prediction. The global sequence alignment with gaps has been done by [362], for instance. Statistical significance has been studied in a growing number of bioinformatics domains, such as protein threading, finding motifs in protein sequences and structures, and the analysis of mRNA expression data. Since exact predictions are not to be expected, accurate measures of statistical significance are the primary methodical contribution toward making bioinformatics predictions more usable.

In the future, we will see a considerable expansion of the field of bioinformatics, in terms of the variety and volume of the data, as well as in terms of the number of methods and the amount of integration. Bioinformatics is an essential part of biology and medicine already today. As it develops further and integrates more with laboratory experimentation, it will expand its role as the third major pillar of modern life science in addition to laboratory experimentation and theoretical biology.

Acknowledgment

I am grateful to the following people for providing useful comments on and suggestions to this text: Mario Albrecht, Niko Beerenwinkel, Francisco Domingues, Andreas Hahn, Daniel Hanisch, Andreas Kämper, Hugo Kubinyi, Jochen Maydt, Christian von Mering, Jörg Rahnenführer, Knut Reinert, Kirsten Roomp, Oliver Sander, Tobias Sing, Ingolf Sommer, and Priti Talwar.

References

1. Noble, D. (2002) The rise of computational biology, *Nat. Rev. Mol. Cell Biol.* **3**(6), 459–463.
2. Luscombe, N.M., Greenbaum, D., Gerstein, M. (2001) What is bioinformatics? a proposed definition and overview of the field, *Methods Inf. Med.* **40**(4), 346–358.

3. Venter, J.C., et al. (2001) The sequence of the human genome, *Science* **291**(5507), 1304–1351.
4. Huang, X. (2001) Bioinformatics Support of Genome Sequencing Projects, in: Lengauer, T. (Ed.) *Bioinformatics – from Genomes to Drugs*, Wiley-VCH, Weinheim, pp. 25–48.
5. Lander, E.S., et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**(6822), 860–921.
6. Kent, W.J., Haussler, D. (2001) Assembly of the working draft of the human genome with GigAssembler, *Genome Res.* **11**(9), 1541–1548.
7. Huson, D.H., et al. (2001) Design of a compartmentalized shotgun assembler for the human genome, *Bioinformatics* **17**(Suppl. 1), S132–S139.
8. Huson, D.H., Reinert, K. Myers, E.W. (2002) The greedy path-merging algorithm for contig scaffolding, *J. ACM* **49**(5), 603–615.
9. Pevzner, P.A., Tang, H., Waterman, M.S. (2001) An eulerian path approach to DNA fragment assembly, *Proc. Natl. Acad. Sci. U.S.A.* **98**(17), 9748–9753.
10. Batzoglou, S., et al. (2002) ARACHNE: a whole-genome shotgun assembler, *Genome Res.* **12**(1), 177–189.
11. Kirkness, E.F., et al. (2003) The dog genome: survey sequencing and comparative analysis, *Science* **301**(5641), 1898–1903.
12. Venter, J.C., et al. (2004) Environmental genome shotgun sequencing of the Sargasso sea, *Science*. **304**(5667), 66–74.
13. Waterston, R.H., et al. (2002) Initial sequencing and comparative analysis of the mouse genome, *Nature* **420**(6915), 520–562.
14. Womble, D.D. (2000) GCG: The wisconsin package of sequence analysis programs, *Methods Mol. Biol.* **132**, 3–22.
15. Gusfield, D. (1997) *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*, xviii, Cambridge University Press, Cambridge [England]; New York, p. 534
16. Zien, A., Zimmer, R., Lengauer, T. (2000) A simple iterative approach to parameter optimization, *J. Comput. Biol.* **7**(3), 483–501.
17. Altschul, S.F., et al. (1990) Basic local alignment search tool, *J. Mol. Biol.* **215**(3), 403–410.
18. Karlin, S., Altschul, S.F. (1990) Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes, *Proc. Natl. Acad. Sci. U.S.A.* **87**(6), 2264–2268.
19. Altschul, S.F., Koonin, E.V. (1998) Iterated profile searches with PSI-BLAST–a tool for discovery in protein databases, *Trends Biochem. Sci.* **23**(11), 444–447.
20. Eddy, S.R. (1998) Profile hidden markov models, *Bioinformatics* **14**(9), 755–763.
21. Chenna, R., et al. (2003) Multiple sequence alignment with the clustal series of programs, *Nucleic Acids Res.* **31**(13), 3497–3500.
22. Higgins, D.G., Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer, *Gene* **73**(1), 237–244.
23. Notredame, C., Higgins, D.G., Heringa, J. (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment, *J. Mol. Biol.* **302**(1), 205–217.
24. Thompson, J.D., Plewniak, F., Poch, O. (1999) BAliBASE: a benchmark alignment database for the evaluation of multiple alignment programs, *Bioinformatics* **15**(1), 87–88.
25. Raghava, G., et al. (2003) OXBench: a benchmark for evaluation of protein multiple sequence alignment accuracy, *BMC Bioinform.* **4**(1), 47.
26. Solovyev, V. (2001) Structure, Properties and Computer Identification of Eukaryotic Genes, in: Lengauer, T. (Ed.) *Bioinformatics – from Genomes to Drugs*, Wiley-VCH, Weinheim, Germany, pp. 59–111.
27. Werner, T. (2001) Analyzing Regulatory Regions in Genomes, in: Lengauer, T. (Ed.) *Bioinformatics – from genomes to Drugs*, Wiley-VCH, Weinheim, Germany, pp. 113–143.
28. Gonnet, G.H., Cohen, M.A., Benner, S.A. (1992) Exhaustive matching of the entire protein sequence database [see comments], *Science* **256**(5062), 1443–1445.
29. Dayhoff, M.O. (1978) *Protein Segment Dictionary 78: From the Atlas of Protein Sequence and Structure*, Vol. 5, Supplements 1, 2, 3,

vii, National Biomedical Research Foundation; Georgetown University Medical Center, Silver Spring, MD, p. 470.
30. Henikoff, S., Henikoff, J.G. (1992) Amino acid substitution matrices from protein blocks, *Proc. Natl. Acad. Sci. U.S.A.* **89**(22), 10915–10919.
31. Bowie, J.U., Luthy, R., Eisenberg, D. (1991) A method to identify protein sequences that fold into a known three – dimensional structure, *Science* **253**(5016), 164–170.
32. Kelley, L.A., MacCallum, R.M., Sternberg, M.J. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM, *J. Mol. Biol.* **299**(2), 499–520.
33. Claverie, J.M., Audic, S. (1996) The statistical significance of nucleotide position-weight matrix matches, *Comput. Appl. Biosci.* **12**(5), 431–439.
34. Kondrakhin, Y.V., et al. (1995) Eukaryotic promoter recognition by binding sites for transcription factors, *Comput. Appl. Biosci.* **11**(5), 477–488.
35. Bateman, A., et al. (2002) The Pfam protein families database, *Nucleic Acids Res.* **30**(1), 276–280.
36. Letunic, I., et al. (2002) Recent improvements to the SMART domain-based sequence annotation resource, *Nucleic Acids Res.* **30**(1), 242–244.
37. Haft, D.H., Selengut, J.D., White, O. (2003) The TIGRFAMs database of protein families, *Nucleic Acids Res.* **31**(1), 371–373.
38. Pavlidis, P., et al. (2001) Promoter region-based classification of genes, *Pacific Symposium on Biocomputing*. http://helix-web.stanford.edu/psb01, pp. 151–163.
39. Yang, Z. (1996) Phylogenetic analysis using parsimony and likelihood methods, *J. Mol. Evol.* **42**(2), 294–307.
40. Bandelt, H.J., Dress, A.W. (1992) Split decomposition: a new and useful approach to phylogenetic analysis of distance data.PG-242–52, *Mol. Phylogenet. Evol.* **1**(3), 242–252.
41. Felsenstein, J. (2003) *Inferring Phylogenies*, 1st edition, Sinauer Associates, Sunderland, MA.
42. Sawa, G., Dicks, J., Roberts, I.N. (2003) Current approaches to whole genome phylogenetic analysis, *Brief. Bioinform.* **4**(1), 63–74.
43. Mathe, C., et al. (2002) Current methods of gene prediction, their strengths and weaknesses, *Nucleic Acids Res.* **30**(19), 4103–4117.
44. Alexandersson, M., Cawley, S., Pachter, L. (2003) SLAM: cross-species gene finding and alignment with a generalized pair hidden Markov model, *Genome Res.* **13**(3), 496–502.
45. Pedersen, J.S., Hein, J. (2003) Gene finding with a hidden Markov model of genome structure and evolution, *Bioinformatics* **19**(2), 219–227.
46. Solovyev, V., Salamov, A. (1997) The Gene-Finder computer tools for analysis of human and model organisms genome sequences, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **5**, 294–302.
47. Snyder, E.E., Stormo, G.D. (1995) Identification of protein coding regions in genomic DNA, *J. Mol. Biol.* **248**(1), 1–18.
48. Burge, C., Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA, *J. Mol. Biol.* **268**(1), 78–94.
49. Brunak, S., Engelbrecht, J., Knudsen, S. (1991) Prediction of human mRNA donor and acceptor sites from the DNA sequence, *J. Mol. Biol.* **220**(1), 49–65.
50. Reese, M.G., et al. (1997) Improved splice site detection in Genie, *J. Comput. Biol.* **4**(3), 311–323.
51. Burset, M., Guigo, R. (1996) Evaluation of gene structure prediction programs, *Genomics* **34**(3), 353–367.
52. Rogic, S., Mackworth, A.K., Ouellette, F.B. (2001) Evaluation of gene-finding programs on mammalian sequences, *Genome Res.* **11**(5), 817–832.
53. Rogic, S., Ouellette, B.F., Mackworth, A.K. (2002) Improving gene recognition accuracy by combining predictions from two gene-finding programs, *Bioinformatics* **18**(8), 1034–1045.
54. Pevzner, P.A., Sze, S.H. (2000) Combinatorial approaches to finding subtle signals in DNA sequences, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **8**, 269–278.
55. Buhler, J., Tompa, M. (2002) Finding motifs using random projections, *J. Comput. Biol.* **9**(2), 225–242.

56. Qiu, P. (2003) Recent advances in computational promoter analysis in understanding the transcriptional regulatory network, *Biochem. Biophys. Res. Commun.* **309**(3), 495–501.
57. Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences, *Nucleic Acids Res.* **27**(2), 573–580.
58. Kurtz, S., et al. (2001) REPuter: the manifold applications of repeat analysis on a genomic scale, *Nucleic Acids Res.* **29**(22), 4633–4642.
59. Delcher, A.L., et al. (1999) Alignment of whole genomes, *Nucleic Acids Res.* **27**(11), 2369–2376.
60. Pevzner, P.A. (2000) *Computational Molecular Biology, An Algorithmic Approach*, MIT Press, Cambridge, MA.
61. Tesler, G. (2002) Efficient algorithms for multichromosomal genome rearrangements, *J. Comput. Syst. Sci.* **65**(3), 587–609.
62. Sankoff, D., Blanchette, M. (1999) Phylogenetic invariants for genome rearrangements, *J. Comput. Biol.* **6**(3–4), 431–445.
63. Pevzner, P., Tesler, G. (2003) Genome rearrangements in mammalian evolution: lessons from human and mouse genomes, *Genome Res.* **13**(1), 37–45.
64. Blaxter, M. (2003) Comparative genomics: two worms are better than one, *Nature* **426**(6965), 395–396.
65. Eichler, E.E., Sankoff, D. (2003) Structural dynamics of eukaryotic chromosome evolution, *Science* **301**(5634), 793–797.
66. Raphael, B.J., et al. (2003) Reconstructing tumor genome architectures, *Bioinformatics* **19**(Suppl. 2), II162–II171.
67. Venclovas, C., et al. (2003) Assessment of progress over the CASP experiments, *Proteins* **53**(Suppl. 6), 585–595.
68. Bujnicki, J.M., et al. (2001) LiveBench-2: large-scale automated evaluation of protein structure prediction servers, *Proteins, Structure, Function and Genetics*, **45**(Suppl. 5), p. 184–191.
69. Koh, I.Y., et al. (2003) EVA: evaluation of protein structure prediction servers, *Nucleic Acids Res.* **31**(13), 3311–3315.
70. Rost, B., Sander, C. (1993) Prediction of protein secondary structure at better than 70% accuracy, *J. Mol. Biol.* **232**(2), 584–599.
71. Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices, *J. Mol. Biol.* **292**(2), 195–202.
72. Cuff, J.A., et al. (1998) JPred: a consensus secondary structure prediction server, *Bioinformatics* **14**(10), 892–893.
73. Selbig, J., Mevissen, T., Lengauer, T. (1999) Decision tree-based formation of consensus protein secondary structure prediction, *Bioinformatics* **15**(12), 1039–1046.
74. Albrecht, M., et al. (2003) Simple consensus procedures are effective and sufficient in secondary structure prediction, *Protein Eng.* **16**(7), 459–462.
75. Kabsch, W., Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* **22**(12), 2577–2637.
76. Frishman, D., Argos, P. (1995) Knowledge-based protein secondary structure assignment, *Proteins* **23**(4), 566–579.
77. McGuffin, L.J., Jones, D.T. (2003) Improvement of the GenTHREADER method for genomic fold recognition, *Bioinformatics* **19**(7), 874–881.
78. Jones, D.T., Miller, R.T., Thornton, J.M. (1995) Successful protein fold recognition by optimal sequence threading validated by rigorous blind testing, *Proteins* **23**(3), 387–397.
79. Sommer, I., et al. (2002) Confidence measures for protein fold recognition, *Bioinformatics* **18**(6), 802–812.
80. Sanchez, R., Sali, A. (1997) Evaluation of comparative protein structure modeling by MODELLER-3, *Proteins, Structure, Function andGenetics*, **41**(Suppl. 1), p. 50–58.
81. Dunbrack, R.L. Jr. (1999) Comparative modeling of CASP3 targets using PSI-BLAST and SCWRL, *Proteins, Structure, Function and Genetics*, **43**(Suppl. 3), p. 81–87.
82. Schwede, T., et al. (2003) SWISS-MODEL: An automated protein homology-modeling server, *Nucleic Acids Res.* **31**(13), 3381–3385.
83. Wolf, Y.I., et al. (1999) Distribution of protein folds in the three superkingdoms of life, *Genome Res.* **9**(1), 17–26.
84. Rodrigues, A., Hubbard, R.E. (2003) Making decisions for structural genomics, *Brief. Bioinform.* **4**(2), 150–167.

85. Watson, J.D., et al. (2003) Target selection and determination of function in structural genomics, *IUBMB Life* **55**(4–5), 249–255.
86. Goh, C.S., et al. (2004) Mining the structural genomics pipeline: identification of protein properties that affect high-throughput experimental analysis, *J. Mol. Biol.* **336**(1), 115–130.
87. Bourne, P.E., et al. (2004) The status of structural genomics defined through the analysis of current targets and structures, *Pacific Symposium on Biocomputing.* http://helix-web.stanford.edu/psb04, pp. 375–386.
88. Simons, K.T., et al. (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions, *J. Mol. Biol.* **268**(1), 209–225.
89. Bonneau, R., et al. (2002) De novo prediction of three-dimensional structures for major protein families, *J. Mol. Biol.* **322**(1), 65–78.
90. Simons, K.T., et al. (1999) Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins, *Proteins* **34**(1), 82–95.
91. Bradley, P., et al. (2003) Rosetta predictions in CASP5: successes, failures, and prospects for complete automation, *Proteins* **53**(Suppl. 6), 457–468.
92. Meiler, J., Baker, D. (2003) Coupled prediction of protein secondary and tertiary structure, *Proc. Natl. Acad. Sci. U.S.A.* **100**(21), 12105–12110.
93. van Batenburg, F.H., Gultyaev, A.P., Pleij, C.W. (2001) PseudoBase: structural information on RNA pseudoknots, *Nucleic Acids Res.* **29**(1), 194–195.
94. Han, K., Byun, Y. (2003) PSEUDOVIEWER2: visualization of RNA pseudoknots of any type, *Nucleic Acids Res.* **31**(13), 3432–3440.
95. Parsch, J., Braverman, J.M., Stephan, W. (2000) Comparative sequence analysis and patterns of covariation in RNA secondary structures, *Genetics* **154**(2), 909–921.
96. Hofacker, I.L., Stadler, P.F. (1999) Automatic detection of conserved base pairing patterns in RNA virus genomes, *Comput. Chem.* **23**(3–4), 401–414.
97. Grate, L. (1995) Automatic RNA secondary structure determination with stochastic context-free grammars, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **3**, 136–144.
98. Zuker, M., Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information, *Nucleic Acids Res.* **9**(1), 133–148.
99. Lyngso, R.B., Zuker, M., Pedersen, C.N. (1999) Fast evaluation of internal loops in RNA secondary structure prediction, *Bioinformatics* **15**(6), 440–445.
100. Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule, *Science* **244**(4900), 48–52.
101. McCaskill, J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure, *Biopolymers* **29**(6–7), 1105–1119.
102. Lyngso, R.B., Pedersen, C.N. (2000) RNA pseudoknot prediction in energy-based models, *J. Comput. Biol.* **7**(3–4), 409–427.
103. Cai, L., Malmberg, R.L., Wu, Y. (2003) Stochastic modeling of RNA pseudoknotted structures: a grammatical approach, *Bioinformatics* **19**(Suppl. 1), I66–I73.
104. Ewing, T.J., et al. (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases, *J. Comput. Aided Mol. Des.* **15**(5), 411–428.
105. Jones, G., Willett, P., Glen, R.C. (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation, *J. Mol. Biol.* **245**(1), 43–53.
106. Rarey, M., et al. (1996) A fast flexible docking method using an incremental construction algorithm, *J. Mol. Biol.* **261**(3), 470–489.
107. Morris, G.M., et al. (1998) Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem.* **19**(14), 1639–1662.
108. Abagyan, R., Totrov, M., Kuznetsov, D. (1994) ICM-a method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation, *J. Comput. Chem.* **15**(5), 488–506.
109. Schulz-Gasch, T., Stahl, M. (2003) Binding site characteristics in structure-based virtual screening: evaluation of current docking tools, *J. Mol. Model. (Online)* **9**(1), 47–57.

110. Lyne, P.D. (2002) Structure-based virtual screening: an overview, *Drug Discov. Today* **7**(20), 1047–1055.
111. Taylor, R.D., Jewsbury, P.J., Essex, J.W. (2002) A review of protein-small molecule docking methods, *J. Comput. Aided Mol. Des.* **16**(3), 151–166.
112. Bissantz, C., Folkers, G., Rognan, D. (2000) Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations, *J. Med. Chem.* **43**(25), 4759–4767.
113. Böhm, H.J. (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure, *J. Comput. Aided Mol. Des.* **8**(3), 243–256.
114. Eldridge, M.D., et al. (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes, *J. Comput. Aided Mol. Des.* **11**(5), 425–445.
115. Gehlhaar, D.K., et al. (1995) Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming, *Chem. Biol.* **2**(5), 317–324.
116. Cornell, W.D., et al. (1995) A second generation force field or the simulation of proteins, nucleic acids, and organic molecules [Review], *J. Am. Chem. Soc.* **117**(19), 5179–5197.
117. Wu, G., et al. (2003) Detailed analysis of grid-based molecular docking: a case study of CDOCKER-A CHARMm-based MD docking algorithm, *J. Comput. Chem.* **24**(13), 1549–1562.
118. Gohlke, H., Hendlich, M., Klebe, G. (2000) Knowledge-based scoring function to predict protein-ligand interactions, *J. Mol. Biol.* **295**(2), 337–356.
119. Muegge, I. (2000) A knowledge-based scoring function for protein-ligand interactions: probing the reference state [Review], *Perspect. Drug Discov. Des.* **20**(1), 99–114.
120. Rarey, M., Lengauer, T. (2000) A recursive algorithm for efficient combinatorial library docking, *Perspect. Drug Discov. Des.* **20**, 63–81.
121. Lamb, M.L., et al. (2001) Design, docking, and evaluation of multiple libraries against multiple targets, *Proteins* **42**(3), 296–318.
122. Matter, H. (1997) Selecting optimally diverse compounds from structure databases: a validation study of two-dimensional and three-dimensional molecular descriptors, *J. Med. Chem.* **40**(8), 1219–1229.
123. Rarey, M., Dixon, J.S. (1998) Feature trees: a new molecular similarity measure based on tree matching, *J. Comput. Aided Mol. Des.* **12**(5), 471–490.
124. Bajorath, J. (2001) Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening, *J. Chem. Inf. Comput. Sci.* **41**(2), 233–245.
125. Chen, X., Reynolds, C.H. (2002) Performance of similarity measures in 2D fragment-based similarity searching: comparison of structural descriptors and similarity coefficients, *J. Chem. Inf. Comput. Sci.* **42**(6), 1407–1414.
126. Lengauer, T., et al. (2004) Novel technologies for virtual screening, *Drug Discov. Today* **9**(1), 27–34.
127. Ekins, S., et al. (2002) Towards a new age of virtual ADME/TOX and multidimensional drug discovery, *J. Comput. Aided Mol. Des.* **16**(5–6), 381–401.
128. Yu, H., Adedoyin, A. (2003) ADME-Tox in drug discovery: integration of experimental and computational technologies, *Drug Discov. Today* **8**(18), 852–861.
129. Katchalski-Katzir, E., et al. (1992) Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques, *Proc. Natl. Acad. Sci. U.S.A.* **89**(6), 2195–2199.
130. Norel, R., et al. (1995) Molecular surface complementarity at protein–protein interfaces: the critical role played by surface normals at well placed, sparse, points in docking, *J. Mol. Biol.* **252**(2), 263–273.
131. Gardiner, E.J., Willett, P., Artymiuk, P.J. (2001) Protein docking using a genetic algorithm, *Proteins* **44**(1), 44–56.
132. Althaus, E., et al. (2002) A combinatorial approach to protein docking with flexible side chains, *J. Comput. Biol.* **9**(4), 597–612.
133. Jones, G., et al. (1997) Development and validation of a genetic algorithm for flexible docking, *J. Mol. Biol.* **267**(3), 727–748.

134. Desmet, J., et al. (1992) The dead-end elimination theorem and its use in protein side-chain positioning, *Nature* **356**, 539–542.
135. Pierce, N.A., et al. (2000) Conformational splitting: a more powerful criterion for dead-end elimination, *J. Comput. Chem.* **21**(11), 999–1009.
136. Fernandez-Recio, J., Totrov, M., Abagyan, R. (2003) ICM-DISCO docking by global energy optimization with fully flexible side-chains, *Proteins* **52**(1), 113–117.
137. Gabb, H.A., Jackson, R.M., Sternberg, M.J. (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information, *J. Mol. Biol.* **272**(1), 106–120.
138. Moont, G., Gabb, H.A., Sternberg, M.J. (1999) Use of pair potentials across protein interfaces in screening predicted docked complexes, *Proteins* **35**(3), 364–373.
139. Chen, R., Li, L., Weng, Z. (2003) ZDOCK: an initial-stage protein-docking algorithm, *Proteins* **52**(1), 80–87.
140. Li, L., Chen, R., Weng, Z. (2003) RDOCK: refinement of rigid-body protein docking predictions, *Proteins* **53**(3), 693–707.
141. Janin, J., et al. (2003) CAPRI: a Critical Assessment of PRedicted Interactions, *Proteins* **52**(1), 2–9.
142. Smith, G.R., Sternberg, M.J. (2003) Evaluation of the 3D-Dock protein docking suite in rounds 1 and 2 of the CAPRI blind trial, *Proteins* **52**(1), 74–79.
143. Chen, R., et al. (2003) ZDOCK predictions for the CAPRI challenge, *Proteins* **52**(1), 68–73.
144. Halperin, I., et al. (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions, *Proteins* **47**(4), 409–443.
145. Smith, G.R., Sternberg, M.J. (2002) Prediction of protein–protein interactions by docking methods, *Curr. Opin. Struct. Biol.* **12**(1), 28–35.
146. Sternberg, M., Moont, G. (2002) Modelling Protein–Protein and Protein-DNA Docking, in: Lengauer, T. (Ed.) *Bioinformatics – from Genomes to Drugs*, Wiley-VCH, Weinheim, Germany, pp. 361–404.
147. Knegtel, R.M., et al. (1994) MONTY: a Monte Carlo approach to protein-DNA recognition, *J. Mol. Biol.* **235**(1), 318–324.
148. Zacharias, M., Sklenar, H. (1999) Conformational analysis of single-base bulges in A-form DNA and RNA using a hierarchical approach and energetic evaluation with a continuum solvent model, *J. Mol. Biol.* **289**(2), 261–275.
149. Ideker, T., Galitski, T., Hood, L. (2001) A new approach to decoding life: systems biology, *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372.
150. Karp, P.D., et al. (2002) The EcoCyc database, *Nucleic Acids Res.* **30**(1), 56–58.
151. Karp, P.D., Ouzounis, C., Paley, S. (1996) HinCyc: a knowledge base of the complete genome and metabolic pathways of H. influenzae, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **4**, 116–124.
152. Romero, P., Karp, P. (2003) PseudoCyc, a pathway-genome database for Pseudomonas aeruginosa, *J. Mol. Microbiol. Biotechnol.* **5**(4), 230–239.
153. Kanehisa, M., et al. (2002) The KEGG databases at GenomeNet, *Nucleic Acids Res.* **30**(1), 42–46.
154. Selkov, E.Jr., et al. (1998) MPW: the metabolic pathways database, *Nucleic Acids Res.* **26**(1), 43–45.
155. Karp, P.D., et al. (2002) The MetaCyc database, *Nucleic Acids Res.* **30**(1), 59–61.
156. Schomburg, I., et al. (2004) BRENDA, the enzyme database: updates and major new developments, *Nucleic Acids Res.* **32**(1), D431–D433.
157. Bairoch, A. (1999) The ENZYME data bank in 1999, *Nucleic Acids Res.* **27**(1), 310–311.
158. Fleischmann, A., et al. (2004) IntEnz, the integrated relational enzyme database, *Nucleic Acids Res.* **32**(1), D434–D437.
159. Osterman, A., Overbeek, R. (2003) Missing genes in metabolic pathways: a comparative genomics approach, *Curr. Opin. Chem. Biol.* **7**(2), 238–251.
160. Schilling, C.H., et al. (2000) Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems, *Biotechnol. Bioeng.* **71**(4), 286–306.
161. Schuster, S., Dandekar, T., Fell, D.A. (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering, *Trends Biotechnol.* **17**(2), 53–60.
162. Stelling, J., et al. (2002) Metabolic network structure determines key aspects of

functionality and regulation, *Nature* **420**(6912), 190–193.
163. Pfeiffer, T., et al. (1999) METATOOL: for studying metabolic networks, *Bioinformatics* **15**(3), 251–257.
164. Mendes, P. (1997) Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3, *Trends Biochem. Sci.* **22**(9), 361–363.
165. Goryanin, I., Hodgman, T.C., Selkov, E. (1999) Mathematical simulation and analysis of cellular metabolism and regulation, *Bioinformatics* **15**(9), 749–758.
166. Voit, E.O. (2000) *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*, Cambridge University Press, Cambridge.
167. Wiechert, W. (2002) Modeling and simulation: tools for metabolic engineering, *J. Biotechnol.* **94**(1), 37–63.
168. Allen, J., et al. (2003) High-throughput classification of yeast mutants for functional genomics using metabolic footprinting, *Nat. Biotechnol.*, **21**(6), p. 692–696.
169. Nielsen, K.F., Smedsgaard, J. (2003) Fungal metabolite screening: database of 474 mycotoxins and fungal metabolites for dereplication by standardised liquid chromatography-UV-mass spectrometry methodology, *J. Chromatogr., A* **1002**(1–2), 111–136.
170. Soga, T., et al. (2003) Quantitative metabolome analysis using capillary electrophoresis mass spectrometry, *J. Proteome Res.* **2**(5), 488–494.
171. Krull, M., et al. (2003) TRANSPATH: an integrated database on signal transduction and a tool for array analysis, *Nucleic Acids Res.* **31**(1), 97–100.
172. Xenarios, I., et al. (2002) DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions, *Nucleic Acids Res.* **30**(1), 303–305.
173. Bader, G.D., Betel, D., Hogue, C.W. (2003) BIND: the biomolecular interaction network database, *Nucleic Acids Res.* **31**(1), 248–250.
174. Sivakumaran, S., et al. (2003) The database of quantitative cellular signaling: management and analysis of chemical kinetic models of signaling networks, *Bioinformatics* **19**(3), 408–415.
175. de Jong, H., et al. (2003) Genetic network analyzer: qualitative simulation of genetic regulatory networks, *Bioinformatics* **19**(3), 336–344.
176. Peleg, M., Yeh, I., Altman, R.B. (2002) Modelling biological processes using workflow and Petri Net models, *Bioinformatics* **18**(6), 825–837.
177. Takahashi, K., et al. (2002) Computational challenges in cell simulation: a software engineering approach, *IEEE Intell. Sys.* **17**(5), 64–71.
178. Takahashi, K., et al., (2003) E-Cell 2: multi-platform E-Cell simulation system, *Bioinformatics* **19**(13), 1727–1729.
179. Triendl, R. (2002) Computerized role models, *Nature* **417**(6892), 7.
180. Matsuno, H., et al., (2003) Biopathways representation and simulation on hybrid functional Petri net, *In Silico Biol.* **3**(3), 389–404.
181. Lipshutz, R.J., et al., (1999) High density synthetic oligonucleotide arrays, *Nat. Genet.* **21**(Suppl. 1), 20–24.
182. Brown, P.O., Botstein, D. (1999) Exploring the new world of the genome with DNA microarrays, *Nat. Genet.* **21**(Suppl. 1), 33–37.
183. Forster, T., Roy, D., Ghazal, P. (2003) Experiments using microarray technology: limitations and standard operating procedures, *J. Endocrinol.* **178**(2), 195–204.
184. Tobler, J.B., et al. (2002) Evaluating machine learning approaches for aiding probe selection for gene-expression arrays, *Bioinformatics* **18**(Suppl. 1), S164–S171.
185. Zien, A., et al. (2003) Microarrays: how many do you need? *J. Comput. Biol.* **10**(3–4), 653–667.
186. Brazma, A., et al. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data, *Nat. Genet.* **29**(4), 365–371.
187. Brazma, A., et al. (2003) ArrayExpress–a public repository for microarray gene expression data at the EBI, *Nucleic Acids Res.* **31**(1), 68–71.
188. Edgar, R., Domrachev, M., Lash, A.E. (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository, *Nucleic Acids Res.* **30**(1), 207–210.

189. Krajewski, P., Bocianowski, J. (2002) Statistical methods for microarray assays, *J. Appl. Genet.* **43**(3), 269–278.
190. Yang, Y.H., Speed, T. (2002) Design issues for cDNA microarray experiments, *Nat. Rev. Genet.* **3**(8), 579–588.
191. Quackenbush, J. (2002) Microarray data normalization and transformation, *Nat. Genet.* **32**(Suppl.), 496–501.
192. Park, T., et al. (2003) Evaluation of normalization methods for microarray data, *BMC Bioinform.* **4**(1), 33.
193. Zien, A., et al. (2001) Centralization: a new method for the normalization of gene expression data, *Bioinformatics* **17**(Suppl. 1), S323–S331.
194. Huber, W., et al. (2003) Parameter estimation for the calibration and variance stabilization of micorarray data, *Stat. Appl. Genet. Mol. Biolo.* **2**(1), Article 3.
195. Golub, T.R., et al. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science* **286**(5439), 531–537.
196. Hastie, T., Tibshirani, R., Friedman, J. (2001) *The Elements of Statistical Learning*, Springer-Verlag, New York.
197. Friedman, N., Kaminski, N. (2002) Statistical methods for analyzing gene expression data for cancer research, *Ernst Schering Res. Found. Workshop* (38), 109–131.
198. Russo, G., Zegar, C., Giordano, A. (2003) Advantages and limitations of microarray technology in human cancer, *Oncogene* **22**(42), 6497–6507.
199. Sharan, R., Elkon, R., Shamir, R. (2002) Cluster analysis and its applications to gene expression data, *Ernst Schering Res, Found. Workshop* (38), 83–108.
200. Iyer, V.R., et al. (1999) The transcriptional program in the response of human fibroblasts to serum, *Science* **283**(5398), 83–87.
201. Lyons, T.J., et al. (2000) Genome-wide characterization of the Zap1p zinc-responsive regulon in yeast, *Proc. Natl. Acad. Sci. U.S.A.* **97**(14), 7957–7962.
202. Tavazoie, S., et al. (1999) Systematic determination of genetic network architecture, *Nat. Genet.* **22**(3), 281–285.
203. Werner, T. (2001) Target gene identification from expression array data by promoter analysis, *Biomol. Eng.* **17**(3), 87–94.
204. Palin, K., et al. (2002) Correlating gene promoters and expression in gene disruption experiments, *Bioinformatics* **18**(Suppl. 2), S172–S180.
205. Yu, H., et al. (2003) Genomic analysis of gene expression relationships in transcriptional regulatory networks, *Trends Genet.* **19**(8), 422–427.
206. Segal, E., Yelensky, R., Koller, D. (2003) Genome-wide discovery of transcriptional modules from DNA sequence and gene expression, *Bioinformatics* **19**(Suppl. 1), I273–I282.
207. Zien, A., et al. (2000) Analysis of gene expression data with pathway scores, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **8**, 407–417.
208. Hanisch, D., et al. (2002) Co-clustering of biological networks and gene expression data, *Bioinformatics* **18**(90001), 145S–1154.
209. Ideker, T., et al. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network, *Science* **292**(5518), 929–934.
210. Segal, E., Wang, H., Koller, D. (2003) Discovering molecular pathways from protein interaction and gene expression data, *Bioinformatics* **19**(Suppl. 1), I264–I272.
211. de Jong, H. (2002) Modeling an simulation of genetic regulatory systems, *J. Comput. Biol.* **9**(1), 67–103.
212. Howbrook, D.N., et al. (2003) Developments in microarray technologies, *Drug Discov. Today* **8**(14), 642–651.
213. Wilson, D.S., Nock, S. (2003) Recent developments in protein microarray technology, *Angew. Chem. Int., Ed. Engl.* **42**(5), 494–500.
214. Templin, M.F., et al. (2002) Protein microarray technology, *Trends Biotechnol.* **20**(4), 160–166.
215. Domingues, F.S., Lengauer, T. (2003) Protein function from sequence and structure data, *Appl. Bioinform.* **2**(1), 3–12.
216. Lan, N., Montelione, G.T., Gerstein, M. (2003) Ontologies for proteomics: towards a systematic definition of structure and function that scales to the genome level, *Curr. Opin. Chem. Biol.* **7**(1), 44–54.
217. Webb, E.C. (1992) *Enzyme Nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology*, Acemdic Press, New York.

218. Mewes, H.W., et al. (2000) MIPS: a database for genomes and protein sequences, *Nucleic Acids Res.* **28**(1), 37–40.
219. Ashburner, M., et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nat Genet.* **25**(1), 25–29.
220. Davidson, E.H., et al. (2002) A genomic regulatory network for development, *Science* **295**(5560), 1669–1678.
221. Kam, N., et al. (2003) Modeling of C. elegans development: a scenario-based approach, *International Workshop on Computational Methods in Systems Biology (CMSB 2003)*, Rovereto, Italy.
222. Fitch, W.M. (2000) Homology a personal view on some of the problems, *Trends Genet.* **16**(5), 227–231.
223. Eulenstein, O., Mirkin, B., Vingron, M. (1998) Duplication-based measures of difference between gene and species trees, *J. Comput. Biol.* **5**(1), 135–148.
224. Yuan, Y.P., et al. (1998) Towards detection of orthologues in sequence databases, *Bioinformatics* **14**(3), 285–289.
225. Storm, C.E., Sonnhammer, E.L. (2002) Automated ortholog inference from phylogenetic trees and calculation of orthology reliability, *Bioinformatics* **18**(1), 92–99.
226. Tatusov, R.L., Koonin, E.V., Lipman, D.J. (1997) A genomic perspective on protein families, *Science* **278**(5338), 631–637.
227. Tatusov, R.L., et al. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes, *Nucleic Acids Res.* **29**(1), 22–28.
228. Letunic, I., et al. (2004) SMART 4.0: towards genomic data integration, *Nucleic Acids Res.* **32**(1), D142–D144.
229. Bucher, P., Bairoch, A. (1994) A generalized profile syntax for biomolecular sequence motifs and its function in automatic sequence interpretation, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 53–61.
230. Falquet, L., et al. (2002) The PROSITE database, its status in 2002, *Nucleic Acids Res.* **30**(1), 235–238.
231. Nevill-Manning, C.G., Wu, T.D., Brutlag, D.L. (1998) Highly specific protein sequence motifs for genome analysis, *Proc. Natl. Acad. Sci. U.S.A.* **95**(11), 5865–5871.
232. Huang, J.Y., Brutlag, D.L. (2001) The EMOTIF database, *Nucleic Acids Res.* **29**(1), 202–204.
233. Attwood, T.K. (2000) The role of pattern databases in sequence analysis, *Brief. Bioinform.* **1**(1), 45–59.
234. Mulder, N.J., et al. (2003) The InterPro database, 2003 brings increased coverage and new features, *Nucleic Acids Res.* **31**(1), 315–318.
235. Jensen, L.J., et al. (2002) Prediction of human protein function from post-translational modifications and localization features, *J. Mol. Biol.* **319**(5), 1257–1265.
236. Fields, S., Song, O. (1989) A novel genetic system to detect protein–protein interactions, *Nature* **340**(6230), 245–246.
237. Uetz, P., et al. (2000) A comprehensive analysis of protein–protein interactions in Saccharomyces cerevisiae, *Nature* **403**(6770), 623–627.
238. Gavin, A.C., et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes, *Nature* **415**(6868), 141–147.
239. Schwikowski, B., Uetz, P., Fields, S. (2000) A network of protein–protein interactions in yeast, *Nat. Biotechnol.* **18**(12), 1257–1261.
240. Saito, R., Suzuki, H., Hayashizaki, Y. (2003) Construction of reliable protein–protein interaction networks with a new interaction generality measure, *Bioinformatics* **19**(6), 756–763.
241. Bader, G.D., Hogue, C.W. (2003) An automated method for finding molecular complexes in large protein interaction networks, *BMC Bioinform.* **4**(1), 2.
242. Krause, R., von Mering, C., Bork, P. (2003) A comprehensive set of protein complexes in yeast: mining large scale protein–protein interaction screens, *Bioinformatics* **19**(15), 1901–1908.
243. Schachter, V. (2002) Protein-interaction networks: from experiments to analysis, *Drug Discov. Today* **7**(Suppl. 11), S48–S54.
244. Legrain, P., Wojcik, J., Gauthier, J.M. (2001) Protein–protein interaction maps: a lead towards cellular functions, *Trends Genet.* **17**(6), 346–352.
245. Wuchty, S. (2002) Interaction and domain networks of yeast, *Proteomics* **2**(12), 1715–1723.
246. Barabasi, A.L., Oltvai, Z.N. (2004) Network biology: understanding the cell's functional

organization, *Nat. Rev. Genet.* **5**(2), 101–113.
247. Hermjakob, H., et al. (2004) IntAct: an open source molecular interaction database, *Nucleic Acids Res.* **32**(1), D452–D455.
248. Brun, C., et al. (2003) Functional classification of proteins for the prediction of cellular function from a protein–protein interaction network, *Genome Biol.* **5**(1), R6.
249. Dandekar, T., et al. (1998) Conservation of gene order: a fingerprint of proteins that physically interact, *Trends Biochem. Sci.* **23**(9), 324–328.
250. Overbeek, R., et al. (1999) The use of gene clusters to infer functional coupling, *Proc. Natl. Acad. Sci. U.S.A.* **96**(6), 2896–2901.
251. Kolesov, G., Mewes, H.W., Frishman, D. (2001) SNAPping up functionally related genes based on context information: a colinearity-free approach, *J. Mol. Biol.* **311**(4), 639–656.
252. Kolesov, G., Mewes, H.W., Frishman, D. (2002) SNAPper: gene order predicts gene function, *Bioinformatics* **18**(7), 1017–1019.
253. Enright, A.J., et al. (1999) Protein interaction maps for complete genomes based on gene fusion events [see comments], *Nature* **402**(6757), 86–90.
254. Marcotte, E.M., et al. (1999) Detecting protein function and protein–protein interactions from genome sequences, *Science* **285**(5428), 751–753.
255. Pellegrini, M., et al. (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles, *Proc. Natl. Acad. Sci. U.S.A.* **96**(8), 4285–4288.
256. Thornton, J.M., et al. (2000) From structure to function: approaches and limitations, *Nat. Struct. Biol.* **7**(Suppl.), 991–994.
257. Todd, A.E., Orengo, C.A., Thornton, J.M. (2001) Evolution of function in protein superfamilies, from a structural perspective, *J. Mol. Biol.* **307**(4), 1113–1143.
258. Holm, L., Sander, C. (1993) Protein structure comparison by alignment of distance matrices, *J. Mol. Biol.* **233**(1), 123–138.
259. Holm, L., Sander, C. (1997) Dali/FSSP classification of three-dimensional protein folds, *Nucleic Acids Res.* **25**(1), 231–234.
260. Alexandrov, N.N., Fischer, D. (1996) Analysis of topological and nontopological structural similarities in the PDB: new examples with old structures, *Proteins* **25**(3), 354–365.
261. Alexandrov, N.N. (1996) SARFing the PDB, *Protein Eng.* **9**(9), 727–732.
262. Shindyalov, I.N., Bourne, P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path, *Protein Eng.* **11**(9), 739–747.
263. Novotny, M., Madsen, D., Kleywegt, G.J. (2004) Evaluation of protein fold comparison servers, *Proteins* **54**(2), 260–270.
264. Pearl, F.M., et al. (2003) The CATH database: an extended protein family resource for structural and functional genomics, *Nucleic Acids Res.* **31**(1), 452–455.
265. Taylor, W.R., Orengo, C.A. (1989) Protein structure alignment, *J. Mol. Biol.* **208**(1), 1–22.
266. Lo Conte, L., et al. (2002) SCOP database in 2002: refinements accommodate structural genomics, *Nucleic Acids Res.* **30**(1), 264–267.
267. Russell, R.B. (1998) Detection of protein three-dimensional side-chain patterns: new examples of convergent evolution, *J. Mol. Biol.* **279**(5), 1211–1227.
268. Kleywegt, G.J. (1999) Recognition of spatial motifs in protein structures, *J. Mol. Biol.* **285**(4), 1887–1897.
269. Bradley, P., Kim, P.S., Berger, B. (2002) TRILOGY: discovery of sequence-structure patterns across diverse proteins, *Proc. Natl. Acad. Sci. U.S.A.* **99**(13), 8500–8505.
270. Barker, J.A., Thornton, J.M. (2003) An algorithm for constraint-based structural template matching: application to 3-D templates with statistical analysis, *Bioinformatics* **19**(13), 1644–1649.
271. Spriggs, R.V., Artymiuk, P.J., Willett, P. (2003) Searching for patterns of amino acids in 3D protein structures, *J. Chem. Inf. Comput. Sci.* **43**(2), 412–421.
272. Stark, A., Sunyaev, S., Russell, R.B. (2003) A model for statistical significance of local similarities in structure, *J. Mol. Biol.* **326**(5), 1307–1316.
273. Liang, J., Edelsbrunner, H., Woodward, C. (1998) Anatomy of protein pockets and cavities: measurement of binding site

geometry and implications for ligand design, *Protein Sci.* **7**(9), 1884–1897.
274. Wei, L., Altman, R.B. (1998) Recognizing protein binding sites using statistical descriptions of their 3D environments, *Pac. Symp. Biocomput.* 497–508.
275. Stahl, M., Taroni, C., Schneider, G. (2000) Mapping of protein surface cavities and prediction of enzyme class by a self-organizing neural network, *Protein Eng.* **13**(2), 83–88.
276. Liang, M.P., et al. (2003) WebFEATURE: an interactive web tool for identifying and visualizing functional sites on macromolecular structures, *Nucleic Acids Res.* **31**(13), 3324–3327.
277. Lichtarge, O., Bourne, H.R., Cohen, F.E. (1996) An evolutionary trace method defines binding surfaces common to protein families, *J. Mol. Biol.* **257**(2), 342–358.
278. Armon, A., Graur, D., Ben-Tal, N. (2001) ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information, *J. Mol. Biol.* **307**(1), 447–463.
279. Pupko, T., et al. (2002) Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues, *Bioinformatics* **18**(90001), 71S–777.
280. Glaser, F., et al. (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information, *Bioinformatics* **19**(1), 163–164.
281. Blaschke, C., et al. (1999) Automatic extraction of biological information from scientific text: protein–protein interactions, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 60–67.
282. Friedman, C., et al. (2001) GENIES: a natural-language processing system for the extraction of molecular pathways from journal articles, *Bioinformatics* **17**(Suppl. 1), S74–S82.
283. Craven, M., Kumlien, J. (1999) Constructing biological knowledge bases by extracting information from text sources, *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, AAAI Press, Heidelberg, 77–86.
284. Donaldson, I., et al. (2003) PreBIND and Textomy–mining the biomedical literature for protein–protein interactions using a support vector machine, *BMC Bioinform.* **4**(1), 11.
285. Andrade, M.A., Bork, P. (2000) Automated extraction of information in molecular biology, *FEBS Lett.* **476**(1–2), 12–17.
286. de Bruijn, B., Martin, J. (2002) Getting to the (c)ore of knowledge: mining biomedical literature, *Int. J. Med. Inf.* **67**(1–3), 7–18.
287. Hirschman, L., et al. (2002) Accomplishments and challenges in literature data mining for biology, *Bioinformatics* **18**(12), 1553–1561.
288. Shatkay, H., Feldman, R. (2003) Mining the biomedical literature in the genomic era: an overview, *J. Comput. Biol.* **10**(6), 821–855.
289. Yeh, A.S., Hirschman, L., Morgan, A.A. (2003) Evaluation of text data mining for database curation: lessons learned from the KDD Challenge Cup, *Bioinformatics* **19**(Suppl. 1), I331–I339.
290. Kemmeren, P., et al. (2002) Protein interaction verification and functional annotation by integrated analysis of genome-scale data, *Mol. Cell* **9**(5), 1133–1143.
291. Kemmeren, P., Holstege, F.C. (2003) Integrating functional genomics data, *Biochem. Soc. Trans.* **31**(Pt. 6), 1484–1487.
292. Marcotte, E.M., et al. (1999) A combined algorithm for genome-wide prediction of protein function, *Nature* **402**(6757), 83–86.
293. Huynen, M., et al. (2000) Predicting protein function by genomic context: quantitative evaluation and qualitative inferences, *Genome Res.* **10**(8), 1204–1210.
294. von Mering, C., et al. (2002) Comparative assessment of large-scale data sets of protein–protein interactions, *Nature* **417**(6887), 399–403.
295. von Mering, C., et al. (2003) STRING: a database of predicted functional associations between proteins, *Nucleic Acids Res.* **31**(1), 258–261.
296. Von Mering, C., et al. (2003) Genome evolution reveals biochemical networks and functional modules, *Proc. Natl. Acad. Sci. U.S.A.* **100**(26), 15428–15433.
297. Pellegrini, M. (2001) Computational methods for protein function analysis, *Curr. Opin. Chem. Biol.* **5**(1), 46–50.
298. Valencia, A., Pazos, F. (2002) Computational methods for the prediction of protein

interactions, *Curr. Opin. Struct. Biol.* **12**(3), 368–373.
299. Huynen, M.A., et al. (2003) Function prediction and protein networks, *Curr. Opin. Cell. Biol.* **15**(2), 191–198.
300. Tucker, C.L., Gera, J.F., Uetz, P. (2001) Towards an understanding of complex protein networks, *Trends Cell Biol.* **11**(3), 102–106.
301. Ideker, T., et al. (2002) Discovering regulatory and signalling circuits in molecular interaction networks, *Bioinformatics* **18**(90001), 233S–2240.
302. Hamosh, A., et al. (2002) Online Mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders, *Nucleic Acids Res.* **30**(1), 52–55.
303. Elston, R.C. (1998) Linkage and association, *Genet. Epidemiol.* **15**(6), 565–576.
304. Liu, B.-H. (1998) *Statistical Genomics: Linkage, Mapping, and QTL Analysis*, xxix, CRC Press, Boca Raton, p. 611.
305. March, R.E. (1999) Gene mapping by linkage and association analysis, *Mol. Biotechnol.* **13**(2), 113–122.
306. Page, G.P., et al. (2003) ''Are we there yet?'': deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits, *Am. J. Hum. Genet.* **73**(4), 711–719.
307. Brookes, A.J. (2001) Rethinking genetic strategies to study complex diseases, *Trends Mol. Med.* **7**(11), 512–516.
308. Gulcher, J., Stefansson, K. (1998) Population genomics: laying the groundwork for genetic disease modeling and targeting, *Clin. Chem. Lab. Med.* **36**(8), 523–527.
309. Hakonarson, H., Gulcher, J.R., Stefansson, K. (2003) deCODE genetics, Inc, *Pharmacogenomics* **4**(2), 209–215.
310. Abbott, A. (1999) Sweden sets ethical standards for use of genetic 'biobanks', *Nature* **400**(6739), 3.
311. Gulcher, J.R., Stefansson, K. (1999) Ethics of population genomics research, *Nature* **400**(6742), 307–308.
312. Palsson, B., Thorgeirsson, S. (1999) Genetic databases. Decoding developments in Iceland, *Nat. Biotechnol.* **17**(5), 407.
313. Gulcher, J., Stefansson, K. (1999) An Icelandic saga on a centralized healthcare database and democratic decision making, *Nat. Biotechnol.* **17**(7), 620.
314. Abbott, A. (1999) 'Strengthened' Icelandic bioethics committee comes under fire, *Nature* **400**(6745), 602.
315. Hodgson, J. (1999) deCODE looks forward as database law passes, *Nat. Biotechnol.* **17**(2), 127.
316. Evans, W.E., Relling, M.V. (1999) Pharmacogenomics: translating functional genomics into rational therapeutics, *Science* **286**(5439), 487–491.
317. Lindpaintner, K. (2003) Pharmacogenetics and the future of medical practice, *J. Mol. Med.* **81**(3), 141–153.
318. Gerhold, D.L., Jensen, R.V., Gullans, S.R. (2002) Better therapeutics through microarrays, *Nat. Genet.* **32**(Suppl.), 547–551.
319. Stoll, M., et al. (2000) New target regions for human hypertension via comparative genomics, *Genome. Res.* **10**(4), 473–482.
320. Hoopengardner, B., et al. (2003) Nervous system targets of RNA editing identified by comparative genomics, *Science* **301**(5634), 832–836.
321. Schmith, V.D., et al. (2003) Pharmacogenetics and disease genetics of complex diseases, *Cell. Mol. Life Sci.* **60**(8), 1636–1646.
322. Tsai, Y.J., Hoyme, H.E. (2002) Pharmacogenomics: the future of drug therapy, *Clin. Genet.* **62**(4), 257–264.
323. Goldstein, D.B., Tate, S.K., Sisodiya, S.M. (2003) Pharmacogenetics goes genomic, *Nat. Rev. Genet.* **4**(12), 937–947.
324. Altman, R.B., Klein, T.E. (2002) Challenges for biomedical informatics and pharmacogenomics, *Annu. Rev. Pharmacol. Toxicol.* **42**, 113–133.
325. Hewett, M., et al. (2002) PharmGKB: the pharmacogenetics knowledge base, *Nucleic Acids Res.* **30**(1), 163–165.
326. Sreekumar, K.R., Aravind, L., Koonin, E.V. (2001) Computational analysis of human disease-associated genes and their protein products, *Curr. Opin. Genet. Dev.* **11**(3), 247–257.
327. Francois, I. (2001) Conquering antibacterial resistance, *Drug News Perspect.* **14**(1), 46–49.
328. McGrath, K.M., et al. (2001) Using HIV-1 sequence variability to explore virus biology, *Virus Res.* **76**(2), 137–160.

329. Paine, K., Flower, D.R. (2002) Bacterial bioinformatics: pathogenesis and the genome, *J. Mol. Microbiol. Biotechnol.* **4**(4), 357–365.
330. Sassetti, C., Rubin, E.J. (2002) Genomic analyses of microbial virulence, *Curr. Opin. Microbiol.* **5**(1), 27–32.
331. Wilson, J.W., et al. (2002) Mechanisms of bacterial pathogenicity, *Postgrad. Med. J.* **78**(918), 216–224.
332. Shafer, R.W., et al. (2000) Human immunodeficiency virus reverse transcriptase and protease sequence database, *Nucleic Acids Res.* **28**(1), 346–348.
333. Beerenwinkel, N., et al. (2002) Diversity and complexity of HIV-1 drug resistance: a bioinformatics approach to predicting phenotype from genotype, *Proc. Natl. Acad. Sci. U.S.A.* **99**(12), 8271–8276.
334. Draghici, S., Potter, R.B. (2003) Predicting HIV drug resistance with neural networks, *Bioinformatics* **19**(1), 98–107.
335. Beerenwinkel, N., et al. (2003) Geno2pheno: estimating phenotypic drug resistance from HIV-1 genotypes, *Nucleic Acids Res.* **31**(13), 3850–3855.
336. Beerenwinkel, N., et al. (2003) Methods for optimizing antiviral combination therapies, *Bioinformatics* **19**(Suppl. 1), I16–I25.
337. Stein, L.D. (2003) Integrating biological databases, *Nat. Rev. Genet.* **4**(5), 337–345.
338. Zdobnov, E.M., et al. (2002) The EBI SRS server-new features, *Bioinformatics* **18**(8), 1149–1150.
339. Davidson, S.B., et al. (2001) K2/Kleisli and GUS: experiments in integrated access to genomic data sources, *IBM Syst. J.* **40**(2), 512–531.
340. Ritter, O., et al. (1994) Prototype implementation of the integrated genomic database, *Comput. Biomed. Res.* **27**(2), 97–115.
341. Bahl, A., et al. (2003) PlasmoDB: the plasmodium genome resource. A database integrating experimental and computational data, *Nucleic Acids Res.* **31**(1), 212–215.
342. Connolly, M.L. (1983) Solvent-accessible surfaces of proteins and nucleic acids, *Science* **221**(4612), 709–713.
343. Krebs, W.G., et al. (2002) Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic, *Proteins* **48**(4), 682–695.
344. Echols, N., Milburn, D., Gerstein, M. (2003) MolMovDB: analysis and visualization of conformational change and structural flexibility, *Nucleic Acids Res.* **31**(1), 478–482.
345. Kohonen, T. (1990) The self-organizing map, *Proc. IEEE* **78**(9), 1464–1480.
346. Roweis, S.T., Saul, L.K. (2000) Nonlinear dimensionality reduction by locally linear embedding, *Science* **290**(5500), 2323–2326.
347. Han, K., Ju, B.H. (2003) A fast layout algorithm for protein interaction networks, *Bioinformatics* **19**(15), 1882–1888.
348. Ju, B.H., Han, K. (2003) Complexity management in visualizing protein interaction networks, *Bioinformatics* **19**(Suppl. 1), I177–I179.
349. Becker, M.Y., Rojas, I. (2001) A graph layout algorithm for drawing metabolic pathways, *Bioinformatics* **17**(5), 461–467.
350. Rojdestvenski, I. (2003) Metabolic pathways in three dimensions, *Bioinformatics* **19**(18), 2436–2441.
351. Bingham, J., Sudarsanam, S. (2000) Visualizing large hierarchical clusters in hyperbolic space, *Bioinformatics* **16**(7), 660–661.
352. Hubbard, T., et al. (2002) The Ensembl genome database project, *Nucleic Acids Res.* **30**(1), 38–41.
353. Birney, E., et al. (2004) Ensembl 2004, *Nucleic Acids Res.* **32**(1), D468–D470.
354. Wilkinson, M.D., Links, M. (2002) BioMOBY: an open source biological web services proposal, *Brief. Bioinform.* **3**(4), 331–341.
355. Stevens, R.D., Robinson, A.J., Goble, C.A. (2003) myGrid: personalised bioinformatics on the information grid, *Bioinformatics* **19**(Suppl. 1), I302–I304.
356. Foster, I., Kesselman, C. (1999) *The Grid: Blueprint for a New Computing Infrastructure*, xxiv, Morgan Kaufmann Publishers, San Francisco, p. 677, [8] of plates.
357. Karolchik, D., et al. (2003) The UCSC genome browser database, *Nucleic Acids Res.* **31**(1), 51–54.
358. Wheeler, D.L., et al. (2003) Database resources of the national center for

biotechnology, *Nucleic Acids Res.* **31**(1), 28–33.
359. Brunak, S., Engelbrecht, J., Knudsen, S. (1990) Cleaning up gene databases, *Nature* **343**(6254), 123.
360. Hooft, R.W., et al. (1996) Errors in protein structures, *Nature* **381**(6580), 272.
361. Laskowski, R.A., et al. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, *J. Biomol. NMR* **8**(4), 477–486.
362. Waterman, M.S., Vingron, M. (1994) Rapid and accurate estimates of statistical significance for sequence data base searches, *Proc. Natl. Acad. Sci. U.S.A.* **91**(11), 4625–4628.

Innate Immunity

Osamu Takeuchi[1,2] *and Shizuo Akira*[1,2]
[1] *Osaka University, Osaka, Japan*
[2] *ERATO, Japan Science and Technology Agency, Osaka, Japan*

1 Innate Immunity 3

2 *Drosophila* Immune Response 4

3 The Lectin-complement Pathway 5

4 Toll-like Receptors (TLRs) 5
4.1 TLR2 7
4.2 TLR1 and TLR6 7
4.3 TLR3 7
4.4 TLR4 8
4.5 TLR5 8
4.6 TLR7 and TLR8 9
4.7 TLR9 9
4.8 TLR11 9

5 The TLR Signaling Pathway 10
5.1 TIR Domain-containing Adaptor Molecules 11
5.1.1 MyD88 11
5.1.2 TRIF/TICAM-1 11
5.1.3 TIRAP/Mal 12
5.1.4 TRAM/TICAM-2 12
5.1.5 Sterile α and HEAT-Armadillo Motifs (SARM) 12
5.2 The IRAK Family 12
5.3 TNF Receptor-associated Factor 6 (TRAF6) 13
5.4 TAK1 and TAB1, TAB2 and TAB3 13

6 Intracellular Recognition of PAMPs 14

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

7 Pathways Activating IFN Response 15

8 Perspectives 16

Bibliography 17
Books and Reviews 17
Primary Literature 17

Keywords

Innate Immunity
An evolutionally conserved immune system utilizing germline-encoded pattern-recognition receptors.

PAMPs
Pathogen-specific molecular patterns recognized by host pattern-recognition receptors.

TLRs
Pattern-recognition receptors that detect pathogen-derived components and are composed of extracellular leucine-rich repeats and a cytoplasmic TIR domain.

TIR Domain
A domain shared by the cytoplasmic portions of TLRs, IL-1R family members, and adaptor proteins that mediate signal transduction.

NOD-LRR Family
Pattern-recognition receptors composed of C-terminal leucine-rich repeats, a central nucleotide-binding oligomerization domain, and N-terminal protein–protein interaction motifs, such as caspase recruitment domains, pyrin domains, or a TIR domain.

The innate immune system utilizes a limited number of germline-encoded pattern-recognition receptors to sense invading pathogens, and is evolutionally conserved from *Drosophila* to vertebrates. The mammalian innate immune system is subdivided into three different mechanisms. The first mechanism is the lectin-complement pathway, which recognizes invading microorganisms using plasma proteins such as lectin. The second mechanism is the Toll-like receptor (TLR) pathway, which detects pathogen-specific molecular patterns (PAMPs) shared by broad classes of microorganisms via TLRs on the membranes of innate immune cells. Each TLR activates a distinct signaling pathway by recruiting different TIR-domain containing adaptor molecules. The pathways lead to activation of the transcription factors NF-κB, AP-1, and interferon regulatory factor 3 (IRF3), followed by rapid induction

of proinflammatory cytokines and type I interferons (IFNs). The third mechanism is mediated by cytoplasmic pattern-recognition receptors that detect microorganisms in the cytosol, and also leads to the production of cytokines and IFNs. These innate responses also instruct acquired immunity by expressing costimulatory molecules and presenting antigens for recognition by lymphocytes.

1
Innate Immunity

In vertebrates, the immune system consists of both innate and acquired immunity, which are defined on the basis of their mechanisms for detecting the presence of infectious organisms. The innate immune system is characterized by the use of a limited number of germline-encoded receptors that recognize diverse pathogens invading the host. The innate system, therefore, targets a set of molecular structures that are absent from host cells, but are unique to microorganisms and shared by various pathogens. By recognizing these "pathogen-specific" patterns, the innate system is able to prevent autoimmune responses. For instance, lipopolysaccharide (LPS), bacterial lipoprotein, peptidoglycan (PGN), lipoteichoic acid (LTA) and double-stranded RNA (dsRNA) are synthesized by bacteria or viruses, but not by host cells. Furthermore, carbohydrates such as mannose and *N*-acetylglucosamine on the surface of pathogens also represent targets for recognition.

The innate immune recognition is mediated by three different mechanisms. The first mechanism is the lectin-complement pathway, which is mediated by lectins and complement proteins secreted into the extracellular space. Binding of lectin to a microorganism triggers a chain of reactions on the surface of that microorganism leading to its destruction or opsonization. The second mechanism is the Toll-like receptor (TLR) pathway, which is mediated by transmembrane receptors on innate immune cells such as macrophages and dendritic cells (DCs). Detection of microorganisms by TLRs activates the cells to produce proinflammatory cytokines and chemokines to evoke inflammation and to recruit immune cells to the site of the infection. In addition, TLR pathways can induce maturation of DCs to instruct and adequately activate acquired immunity. The third mechanism is mediated by cytoplasmic pattern-recognition receptors that detect viruses and bacteria that have successfully invaded cells.

In contrast, the acquired immune system generates a highly diverse repertoire of antigen receptors, T- and B-cell receptors, via DNA rearrangement then followed by clonal selection that eliminates self-reacting cells. After encountering a pathogen, the lymphocytes bearing appropriately high affinity antigen receptors for that pathogen expand and eliminate the pathogens in the late stage of infection and contribute to the establishment of immunological memory. However, recent studies have revealed that adequate instruction by innate immune cells is a prerequisite for the activation of lymphocytes.

The innate immune system is evolutionally conserved from *Caenorhabditis elegans* and *Drosophila melanogaster* to vertebrates. In fact, it should be noted that *Toll* was originally identified during

research into *Drosophila* developmental biology, and its role in *Drosophila* immune responses has been extensively studied. Moreover, the knowledge gained from that research has had a tremendous influence on advancing research into mammalian innate immunity.

2
Drosophila Immune Response

Drosophila does not possess the adaptive immune system, and mounts its host defense by discriminating between classes of pathogens and inducing immune responses. An important part of *Drosophila* immune response is the synthesis of antimicrobial peptides against microbial infection. *Drosophila* has seven antimicrobial peptides, and among those, Drosomycins and Metchnikowin have activity against fungi, Defensin acts against gram-positive bacteria and Attacins, Cecropins, Drosocin, and Diptericins have spectra against gram-negative bacteria. The induction of appropriate antimicrobial peptides is regulated by two distinct signaling pathways, *Toll* and *Imd*. The *Toll* pathway is activated in response to infections by fungi and gram-positive bacteria. Toll is a receptor that contains extracellular leucine-rich repeats and a cytoplasmic Toll/IL-1R homology domain. It was originally identified as an essential component of the pathway that determines the dorsal-ventral axis in early *Drosophila* embryogenesis, but *Toll*-mutant flies were also found to be susceptible to fungal infection. Invading gram-positive bacteria are recognized by PGN-recognition proteins (PGRP) followed by activation of a Toll ligand, Spaetzle. Binding of Spaetzle to Toll activates the *Drosophila*homolog of MyD88 and a serine/threonine kinase, Pelle, followed by the phosphorylation and degradation of Cactus, an IκB like inhibitory protein of Rel-type transcription factors. The degradation of Cactus releases the transcription factors Dorsal and Dorsal-related immunity factor (DIF) and initiates their nuclear translocation to induce the expression of Drosomycin genes. The pathway closely resembles the mammalian TLR- and interleukin-1 receptor (IL-1R)-mediated signaling pathway. In *Drosophila*, the Toll family consists of nine members, but the other Toll family members may not be involved in antimicrobial immune response.

The *Imd* pathway is involved in host defense against gram-negative bacteria by inducing Attacins, Cecropins, Drosocin, and Diptericins. The IMD protein contains a death domain (DD) that has high homology to that of mammalian TNF-receptor interacting protein (RIP). *Drosophila* homologs of FADD, TGF-β-activated kinase 1 (TAK1), IκB kinase (IKK) β are involved in the *Imd* pathway. Activated *Drosophila* IKKβ phosphorylates another NF-κB like protein, Relish, which results in its cleavage and subsequent nuclear translocation. In turn, Relish induces the expression of genes encoding antimicrobial peptides against gram-negative bacteria. Gram-negative-binding proteins (GNBPs) and a member of the PGRP family containing a transmembrane domain have been implicated in the *Imd* pathway, but the precise mechanisms for *Imd* pathway activation by these molecules remain to be clarified.

Taken together, *Drosophila* distinguishes among different types of invading microorganisms by activation of distinct signaling pathways, resulting in the production of appropriate antimicrobial peptides for the particular type of invading pathogen.

3
The Lectin-complement Pathway

In mammalian innate immunity, the lectin-complement pathway is an important mechanism for recognizing a wide range of pathogens outside of host cells, and is mediated by a number of proteins that are secreted into the serum. The lectin pathway involves carbohydrate recognition by mannose-binding lectin (MBL) and ficolins (Fig. 1). MBL and ficolins specifically recognize *N*-acetyl-D-glucosamine (GlcNAc), mannose, and fucose, and do not interact with D-galactose or sialic acid, which are abundant on mammalian cells. Binding of MBL and ficolins to microbial carbohydrates activates members of the MBL-associated serine protease (MASP) family on the surface of bacterial cells. MASPs, which form complexes with MBL, are enzymes that cleave complement components to activate the complement pathway (Fig. 1). The cleavage leads to direct destruction of the invading bacteria or to their opsonization, which facilitates phagocytosis by macrophages, and in human, MBL-deficiency was shown to increase susceptibility to a variety of infectious diseases. Among MASPs, MASP-2 is responsible for activation of C4 and C2, which leads to C4bC2a enzyme complex formation. C4bC2a then acts as a C3 convertase and cleaves C3, generating products that can bind and destroy bacteria. In contrast, MASP-1 is able to cleave C3 directly (Fig. 1). Together, the lectin-complement pathway plays an important role in detecting microorganisms in the bloodstream and in extracellular space.

4
Toll-like Receptors (TLRs)

In 1997, Janeway and Medzhitov identified a mammalian homolog of *Drosophila* Toll and showed that it was involved in the innate immune response. Eleven mammalian TLRs have been reported to

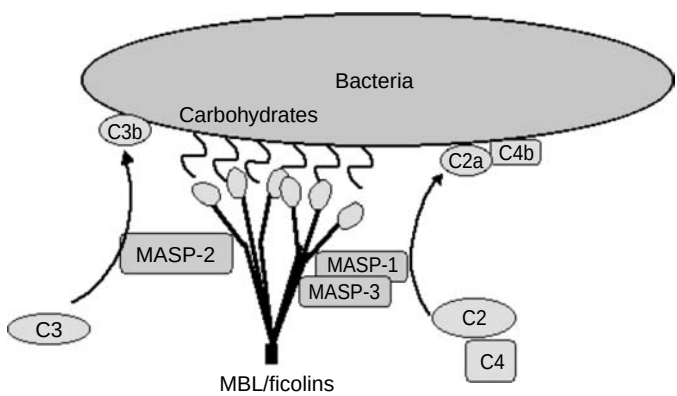

Fig. 1 The lectin-complement pathway. MBL or ficolin binds bacterial carbohydrates on the surface of a bacterial cell. This results in the activation of MASP family members that associate with MBL or ficolin. MASP-2 is responsible for catalyzing cleavage of C4 and C2 to generate C4b and C2a. A C4bC2a complex catalyzes the cleavage of C3, resulting in the generation of C3b that binds to and destroys the bacteria. In contrast, MASP-1 and -3 cleaves C3 directly.

date, and 10 of them have been shown to recognize specific PAMPs. TLRs (Toll-like receptors) are expressed not only on macrophages, DCs, and B cells, but also on fibroblasts, vascular endothelial cells, and intestinal epithelial cells. The TLR family proteins consist of extracellular leucine-rich repeat (LRR) motifs, a transmembrane region, and a cytoplasmic tail containing a Toll/IL-1 receptor homology (TIR) domain. The LRR motifs are responsible for ligand recognition and the TIR domain is essential for triggering intracellular signaling pathways. The activation of TLR signaling pathways leads to the expression of proinflammatory cytokine and chemokine genes. This induces migration of immune cells from peripheral blood to the site of infection. DCs play a pivotal role in T cell activation and differentiation into T helper 1 cells. DCs phagocytose pathogens, process them, and present the antigenic peptides on MHC molecules. TLR signaling is critical for the maturation of DC by regulating phagosome maturation, costimulatory molecule expression and cytokine production.

Homeostasis of acquired immunity is controlled by regulatory T cells, which can suppress self-reacting T cells. However, regulatory T cells do not interfere with T cell activation in the course of pathogenic infection. Recent studies revealed that TLR signaling is involved in suppression of the inhibitory function of regulatory T cells. This is controlled by IL-6 secreted by TLR stimulation.

In addition, ligands for TLR3, TLR4, TLR7, and TLR9, but not for TLR2 and TLR5, are known to upregulate type I interferons (IFNs) and IFN-inducible genes. These TLRs recognize viral as well as bacterial components, indicating that TLR signaling is also involved in antiviral immune responses. Roles of each TLR in the recognition of bacterial components are described in the following section (Fig. 2).

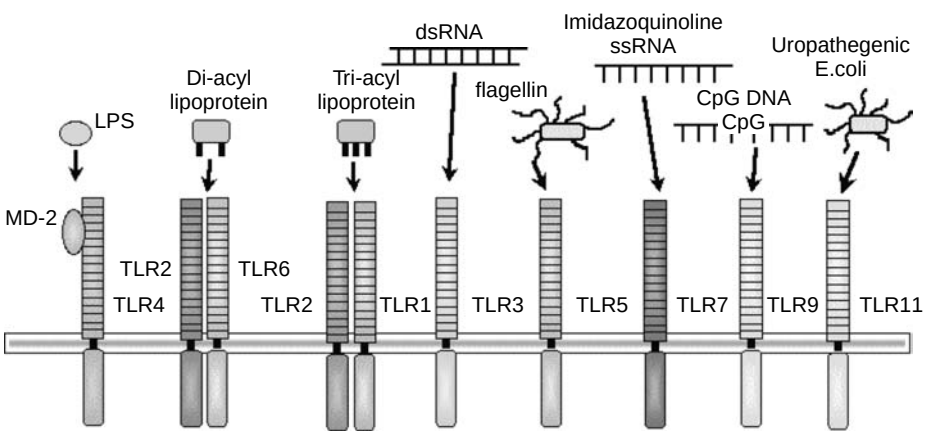

Fig. 2 TLRs recognize various PAMPs. TLR2 recognizes bacterial lipoproteins in concert with TLR1 or TLR6. TLR1 and TLR6 discriminate subtle difference in the lipid moiety of tri-acyl and di-acyl lipoproteins, respectively. TLR4 forms a complex with MD-2 and detects LPS. TLR3 is involved in extracellular recognition of viral dsRNA and TLR5 recognize bacterial flagellin. TLR7 and TLR9 are responsible for viral ssRNA and unmethylated CpG-DNA, respectively. TLR11 detects uropathogenic bacteria in mice.

4.1
TLR2

TLR2 recognizes various PAMPs from bacteria, fungi, and parasites including bacterial lipoprotein, PGN, *Saccharomyces cerevisiae* zymosan, and glycosylphosphatidylinositol (GPI)-anchors from *Trypanosoma cruzi*. In addition, TLR2 is involved in the recognition of LPS from *Porphyromonas gingivalis* and *Leptospira interrogans*, which have different structures to LPS from enterobacteria. Macrophages derived from TLR2$^{-/-}$ mice did not produce proinflammatory cytokines in response to TLR2 ligands, such as PGN and bacterial lipoproteins, and TLR2 signaling did not induce type I IFNs. Corresponding to the failure to detect these components through TLR2, TLR2$^{-/-}$ mice were susceptible to infections by various gram-positive bacteria and fungi, including *Staphylococcus aureus*, Group B *Streptococcus*, *Borrelia burgdorferi*, *Chlamydia trachomatis*, *Candida albicans*, *Streptococcus pneumoniae*, and so on. On the other hand, recent reports have shown that TLR2 signaling suppresses immunity against *Candida* infection. The population of regulatory T cells and IL-10 secretion were partially reduced in TLR2$^{-/-}$ mice, which may explain the improved resistance to *Candida* infection. Although the precise mechanism of this suppression is to be clarified, this is an interesting phenomenon that TLR signaling can work as both stimulatory and inhibitory manner.

4.2
TLR1 and TLR6

TLR1 and TLR6 are highly homologous with each other and are also homologous to TLR2. Heterodimerization of TLR2 with TLR1 or TLR6 determines the ligand specificity. Although expression of either TLR1 or TLR6 alone failed to confer any response, coexpression of TLR1 or TLR6 with TLR2 resulted in cell activation in response to triacyl- or diacyl-lipopeptide, respectively. The cellular response to mycoplasmal diacyl-lipopeptide was abolished in TLR6$^{-/-}$ mice. In contrast, TLR1$^{-/-}$ mice showed an impaired response to triacyl-lipopeptide, while the response to diacyl-lipopeptide was normal. Overexpression studies revealed that TLR1 and TLR6 each interacted with TLR2 in cells. These observations suggest that TLR1 and TLR6 recognize the difference in the lipid portion of lipoproteins through heterodimerization with TLR2.

4.3
TLR3

Infection with RNA viruses leads to dsRNA generation in the cytoplasm of infected cells. Such dsRNA is a virus-specific molecular pattern that can be recognized by the host to stimulate immune responses. Cells expressing TLR3 respond to dsRNA or its synthetic mimic, polyinosinic-polycytidylic acid (poly I:C), and activate an intracellular signaling pathway leading to upregulation of type I IFNs and proinflammatory cytokines. Myeloid DCs and lung fibroblasts derived from TLR3$^{-/-}$ mice were hyporesponsive to poly I:C stimulation, and TLR3$^{-/-}$ mice were reported to be susceptible to cytomegalovirus (CMV) infection. However, inoculation of poly I:C into TLR3$^{-/-}$ mice resulted in normal production of serum IFNα, indicating that other molecule(s) are also involved in the dsRNA-induced immune responses.

4.4 TLR4

Two strains of mice, C3H/HeJ and C57BL/10ScCr, were known to be hyporesponsive to LPS, a component of outer cell membrane of gram-negative bacteria. Positional cloning of the gene responsible for LPS recognition in these mice identified TLR4. In C3H/HeJ mice, a missense point mutation resulted in replacement of a proline in the cytoplasmic domain of the TLR4 protein with histidine, and created a dominant-negative form of the protein. In C57BL/10ScCr mice, the *Tlr4* locus was entirely deleted. TLR4$^{-/-}$ mice generated by homologous recombination of embryonic stem (ES) cells were highly resistant to LPS-induced shock, and macrophages and B cells derived from these mice were unresponsive to LPS in terms of proinflammatory cytokine production and the proliferative response. The activation of NF-κB and MAP kinase in response to LPS was also abrogated. Moreover, C3H/HeJ mice were highly susceptible to infection by *Salmonella* and *Escherichia coli*, indicating that recognition of bacteria by TLR4 is critical to host defense

The entire LPS recognition machinery is more complicated. When LPS is present in the bloodstream, it is immediately captured by LPS-binding protein (LBP), and the LPS–LBP complex is transferred to CD14, a GPI-anchored protein, located on the surface of macrophages. An extracellular protein, MD-2, directly binds TLR4 and enhances LPS-induced NF-κB activation in the TLR4-expressing cells. LBP$^{-/-}$ and CD14$^{-/-}$ mice are hyporesponsive to LPS stimulation. The response of MD-2$^{-/-}$ mice to LPS was also severely impaired. Interestingly, surface expression of TLR4 was abolished in MD-2$^{-/-}$ macrophages. LPS stimulation initiates the dimerization of TLR4 to ignite the downstream signaling pathway. An LRR containing transmembrane protein, RP105, is also involved in LPS recognition in B cells, since targeted disruption of RP105 resulted in an impaired proliferative response of B cells to LPS.

In addition to LPS, TLR4 recognizes several other PAMPs such as respiratory syncytial virus F protein and mouse mammary tumor virus envelope protein. In addition, it has also been reported to recognize endogenous products. Under inflammatory conditions, components of the extracellular matrix are degraded or processed. The products, such as fibronectin fragments and hyaluronic acid degradation products can stimulate cells through TLR4. Furthermore, several studies have demonstrated that both human and chlamydial HSP60 are putative ligands for TLR2 and TLR4. However, it is possible that these ligand preparations were contaminated with LPS or other PAMPs, since most studies utilize recombinant proteins synthesized in *E. coli*.

4.5 TLR5

TLR5 recognizes bacterial flagellin, a unique PAMP composed of protein. Flagellin is a protein component of the flagellum, a structure used by some bacteria to move through liquid media. Chinese hamster ovary (CHO) cells expressing human TLR5 can activate NF-κB in response to flagellin. Moreover, *Salmonella typhimurium* lacking flagellin failed to activate TLR5, indicating that TLR5 is critical for flagellin recognition. A stop codon within the open-reading frame (ORF) of human TLR5 in many individuals renders them incapable of responding adequately to flagellated bacteria. The

MOLF/Ei mouse strain is shown to be susceptible to *S. typhimurium* infection, and one of the MOLF/Ei susceptibility loci was mapped to chromosome 1, which contains the mouse *Tlr5* gene. Furthermore, TLR5 expression was reduced in MOLF/Ei mice.

4.6 TLR7 and TLR8

TLR7 is closely related to TLR8 and shows significant homology with TLR9. TLR7 and TLR8 are both located on the X chromosome. Recent studies have revealed that the ligands of these three TLRs are nucleotides and their derivatives. Initially, imidazoquinoline derivatives, including imiquimod and resiquimod (R-848), were identified as TLR7 ligands in mice. These derivatives are small synthetic compounds that are approved for use as antiviral drugs due to their potency for inducing type I IFNs. TLR7$^{-/-}$ mice failed to respond to these compounds by producing proinflammatory cytokines or type I IFNs. In humans, R-848 strongly activates plasmacytoid DCs (pDCs) to produce IFNα. TLR7 is also involved in the recognition of other synthetic anticancer drugs, such as loxoribine (7-allyl-8-oxoguanosine), bropirimine (2-amin-5-bromo-6-phenyl-4(3)-pyrimidinone) and certain guanosine analogs. Recently, TLR7 was found to be essential for DCs to recognize single-stranded RNAs (ssRNAs) (Fig. 2).

Expression of human TLR8 as well as human TLR7 in HEK293 cells conferred the responsiveness to ssRNAs to activate an NF-κB reporter gene. In contrast, the expression of mouse TLR7, but not TLR8, could restore ssRNA response in culture cells. Moreover, antigen-presenting cells from TLR8$^{-/-}$ mice produced proinflammatory cytokines normally in response to ssRNAs, suggesting that human and mouse TLR8 recognize different PAMPs.

4.7 TLR9

Bacterial DNA is characterized by an abundance of CpG motifs and the lack of methylation. This PAMP is recognized by the host immune cells such as DCs, macrophages, and B cells, and strongly induce T helper 1 development. Furthermore, synthetic oligodeoxynucleotides containing an unmethylated CpG motif (CpG-DNA) show strong immunostimulatory activity.

TLR9 was shown to be essential for CpG-DNA detection, since the immune response to CpG-DNA stimulation was abolished in TLR9$^{-/-}$ mice. TLR9 is localized in the endoplasmic reticulum (ER) where CpG-DNA is transported after its incorporation into a tubular lysosomal compartment. DNA viruses, such as Herpes simplex virus (HSV) types I and II, also contain genomic DNA with unmethylated CpG motifs, and pDCs from TLR9$^{-/-}$ mice failed to produce IFNα upon HSV I or II infection. Moreover, mice with a mutation in the *Tlr9* locus were susceptible to CMV infection, indicating that TLR9 is critically involved in the immune response against DNA viruses.

4.8 TLR11

Mouse TLR11 is expressed in macrophages and liver, kidney, and bladder epithelial cells. Cells expressing TLR11 specifically respond to uropathogenic bacteria, but not to known other TLR ligands. TLR11$^{-/-}$ mice were highly susceptible to kidney infections by uropathogenic bacteria, and macrophages derived from TLR11$^{-/-}$

mice failed to produce proinflammatory cytokines in response to these bacteria. However, human TLR11 is likely to be nonfunctional, since the stop codons were observed in its ORF.

5 The TLR Signaling Pathway

By detecting PAMPs, TLRs trigger downstream Signaling pathways that lead to nuclear translocation of NF-κB and the activation of MAP kinases (Fig. 3). TIR domain-containing adaptors are recruited to the receptors first, and then interact with IL-1R-associated kinases (IRAKs). Activated IRAK1 recruits TNF receptor-associated factor 6 (TRAF6), which activates two ubiquitination proteins, Ubc13 and Uev1A. Ubc13 and Uev1A form a complex, and this UBC complex catalyzes the formation of a lysine 63-linked polyubiquitin chain. Ubiqutinated TRAF6 activates a complex of TGFβ activating kinase 1 (TAK1) and associated proteins, TAB1, TAB2, and TAB3, to directly phosphorylate the IκB kinase (IKK) complex. The activated IKK complex, composed of IKKα,

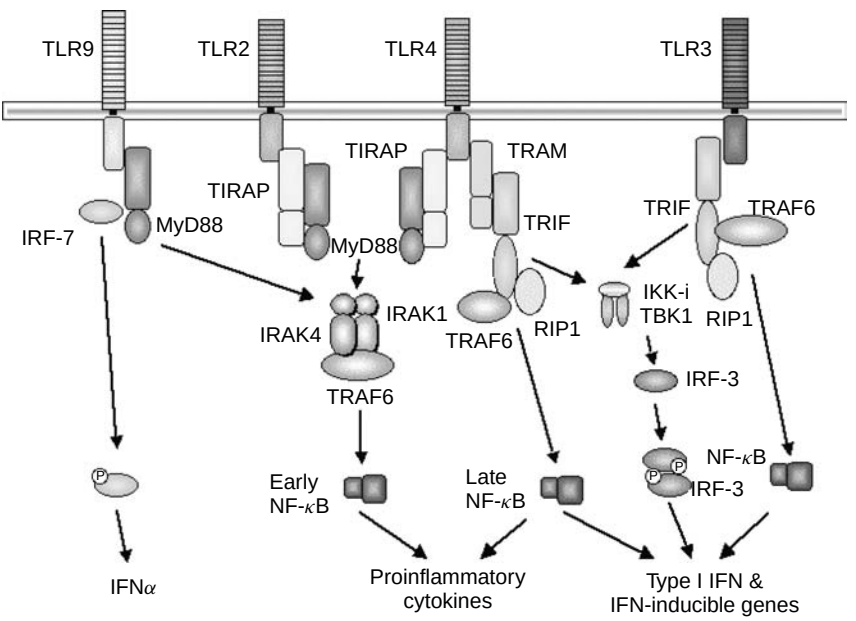

Fig. 3 The signaling pathways emanated from TLRs. Ligand binding with TLRs except TLR3 recruits MyD88 to the cytoplasmic TIR domain. MyD88, in turn, associates with IRAK4 and IRAK1. IRAK1 then activates TRAF6 leading to the nuclear translocation of NF-κB. Activated NF-κB upregulates transcription of proinflammatory cytokine genes. Another TIR-domain adaptor molecule, TIRF, is recruited to TLR3 and TLR4. TRIF activates late NF-κB via RIP1 and TBK1/IKK-i to induce phosphorylation of IRF-3. Subsequently, IRF3 homodimerizes, translocates to the nucleus and upregulates transcription of type I IFNs and IFN-inducible genes. TIRAP is specifically involved in a TLR2- and TLR4-mediated signaling pathway via MyD88, but is dispensable for the activation of TRIF-dependent pathway. TRAM bridges TLR4 and TRIF to activate TRIF-dependent pathway. TLR9 signaling is dependent on MyD88. MyD88 directly associates with IRF7 in pDC to induce phosphorylation. IRF7 subsequently activate IFNα gene expression.

IKKβ, and NF-κB essential modulator (NEMO), phosphorylates IκB, leading to its proteasome-mediated degradation, and the released NF-κB translocates from the cytosol to the nucleus to mediate the expression of proinflammatory cytokines. Next, we focus on the signaling molecules that act adjacent to the receptors, and act on the TIR domain-containing adaptors in particular.

5.1 TIR Domain-containing Adaptor Molecules

All IL-1R and TLR family members possess a TIR domain, a motif of ~160 amino acids, in their cytoplasmic region, which is critical for emanating downstream signaling pathways. Following ligand binding, TLRs dimerize and recruit cytoplasmic adaptors that also contain a TIR domain. Five TIR domain-containing adaptor molecules have been identified in mammals to date, namely, myeloid differentiation factor 88 (MyD88), TIR domain-containing adaptor inducing IFNβ (TRIF)/TIR domain-containing adaptor molecule-1 (TICAM-1), TIR domain-containing adaptor protein (TIRAP)/MyD88 adaptor like (MAL), TRIF-related adaptor molecule (TRAM)/TICAM-2 and sterile α and HEAT-Armadillo motifs (SARM). Individual TLRs can activate distinct signaling pathways through recruiting distinct adaptor molecules.

5.1.1 MyD88

MyD88 is an adaptor molecule consisting of an N-terminal DD and a C-terminal TIR domain. All IL-1R/TLR family members, except for TLR3, share at least one common signaling pathway via MyD88. After ligand stimulation, TLR/IL-1R interacts with MyD88 via the TIR domains, and MyD88 then recruits and associates with IRAK proteins by homophilic interaction via the DD. Analysis of MyD88$^{-/-}$ mice revealed that MyD88 is a gateway for signaling through all the TLR/IL-1R family members, except for TLR3. Cells derived from MyD88$^{-/-}$ mice failed to produce proinflammatory cytokines in response to various TLR ligands, such as bacterial lipoprotein, PGN, LPS, flagellin, ssRNA and CpG-DNA, in addition to IL-1β and IL-18. MyD88$^{-/-}$ cells showed defective activation of NF-κB and MAP kinases in response to TLR ligands, except for the TLR4 ligand (LPS) and TLR3 ligand (poly I:C). When MyD88$^{-/-}$ cells were stimulated with LPS, delayed activation of NF-κB and MAP kinases was observed. The defect in TLR signaling renders MyD88$^{-/-}$ mice highly susceptible to infections with various pathogens, including bacteria, fungi, viruses, and parasites, such as *S. aureus*, *Mycobacterium tuberculosis*, *Mycobacterium avium*, *Listeria monocytogenes*, Group B *Streptococcus*, *B. burgdorferi*, *C. albicans* and *T. cruzi*. Taken together, these observations indicate that MyD88 is essential for host defense against various pathogens by activating innate immune responses.

5.1.2 TRIF/TICAM-1

TRIF contains a TIR domain, a TRAF6 binding motif, and a C-terminal receptor interacting protein (RIP) homotypic interaction domain. Overexpression of TRIF activated an NF-κB-dependent reporter gene, and also activated the IFNβ reporter much more strongly than other TIR domain-containing adaptors. TRIF can associate with TRAF6 through its TRAF6 binding motif, and also interact with RIP1 via the RIP homotypic interaction motif. TRIF can also associate with TBK1, a kinase known to phosphorylate IRF3 to

induce IFNβ gene induction. Poly(I:C)-mediated NF-κB activation was abolished in RIP$^{-/-}$ cells, indicating that RIP is essential for TLR3-mediated NF-κB activation. TRIF$^{-/-}$ mice were defective in TLR3- and TLR4-mediated proinflammatory cytokine production and the IFN response. Although LPS-mediated initial activation of NF-κB occurred in TRIF$^{-/-}$ cells, it was attenuated more rapidly than that in wild-type cells. LPS-induced signaling pathways were not activated in the absence of both MyD88 and TRIF, indicating that LPS signaling is entirely dependent on these two adaptors. Overexpression of TRIF in cells induced apoptosis via caspase-8 activation, suggesting a role for TRIF in bacteria-induced cell death. Interestingly, neither TLR7- and TLR9-mediated cytokine production nor the IFN response was impaired in TRIF$^{-/-}$ cells, indicating that the function of TRIF is limited to TLR3 and TLR4 signaling.

5.1.3 TIRAP/Mal

TIRAP/Mal was cloned as another TIR domain-containing adaptor molecule from a database search. Although TIRAP/Mal was initially implicated in an MyD88-independent signaling pathway based on *in vitro* studies, generation of TIRAP/Mal$^{-/-}$ mice revealed that TIRAP/Mal was required for an MyD88-dependent pathway emanating from TLR2 and TLR4 but not for MyD88-independent pathway.

5.1.4 TRAM/TICAM-2

TRAM/TICAM-2 is a 235-amino acids protein with a C-terminal TIR domain. Like TRIF, expression of TRAM/TICAM-2 activates IRF-3 and NF-κB-dependent gene induction. In TRAM/TICAM-2$^{-/-}$ cells, LPS-induced cytokine and IFN-inducible gene expression was severely impaired. Furthermore, TRAM/TICAM-2$^{-/-}$ cells failed to sustain LPS-induced NF-κB activation compared to wild-type cells. In contrast, there were no defects in the poly I:C and CpG-DNA responses. TRAM/TICAM-2 can physically interact with both TLR4 and TRIF/TICAM-1, whereas no direct association between TLR4 and TRIF/TICAM-1 was detected. When taken together, TRAM/TICAM-2 functions to bridge TLR4 and TRIF/TICAM-1 to activate the downstream signaling pathway.

5.1.5 Sterile α and HEAT-Armadillo Motifs (SARM)

SARM is a 690-amino acid protein containing a TIR domain, two sterile α-motif (SAM) domains and an Armadillo repeat motif (ARM). SARM is evolutionally conserved from *C. elegans* and *D. melanogaster*. *C. elegans* only contains two genes that encode TIR-domain proteins. One is *tol-1*, a Toll homolog, and the other is *tir-1*, a SARM homolog. *Caenorhabditis elegans* SARM is required for resistance to microbial pathogens. However, overexpression of human SARM in mammalian cells was not sufficient to activate NF-κB and IRF3. The establishment of SARM$^{-/-}$ mice has not yet been reported.

5.2 The IRAK Family

Four IRAK family members including IRAK1, IRAK2, IRAK4, and IRAK-M have been identified in mammals, and are composed of an N-terminal DD and a subsequent serine/threonine kinase domain. While IRAK1 and IRAK4 have bona fide kinase activity, IRAK2 and IRAK-M are both inactive due to mutation of a critical residue in their kinase domains. After ligand stimulation with IL-1R/TLR, IRAK4 is

recruited to TLR/IL-1R via the DD to interact with MyD88 or TIRAP/Mal. Subsequently, IRAK4 can associate and directly phosphorylate IRAK1, which triggers the transient recruitment of TRAF6. Studies using IRAK4$^{-/-}$ mice revealed that IRAK4 is essential for activation of the IL-1R/TLR-induced signaling pathway. In contrast, although embryonic fibroblasts (MEFs) derived from IRAK1$^{-/-}$ mice exhibited an attenuated response to stimulation with IL-1β, LPS or other PAMPs, significant cytokine production was observed. This may be due to compensation by IRAK2. It is interesting to explore whether the kinase activity of IRAKs is critical for signaling, since kinase-deficient IRAK1 was still able to restore IL-1β-mediated NF-κB activation in IRAK1-deficient HEK293 cells. The role of kinase activity of IRAK4 is still controversial. Some reports showed that expression of kinase-deficient IRAK4 strongly impaired IL-1-mediated NF-κB activation, while a report showed that a kinase-inactive mutant of IRAK4 partially restored IL-1β-mediated IL-6 production as well as NF-κB activation in IRAK4$^{-/-}$ cells. Future studies will clarify the roles of the kinase activities of IRAK family members. IRAK2 also interacts with MyD88 and TIRAP/Mal, and may mediate the TIRAP/Mal signaling pathway emanating from TLR2 and TLR4. A future IRAK2 knockout study will be necessary to reveal its exact physiological functions. IRAK-M expression is restricted to monocytes/macrophages and is upregulated upon stimulation with TLR ligands. IRAK-M$^{-/-}$ mice showed enhanced production of proinflammatory cytokines in response to TLR ligands. IRAK-M perturbs the association of the IRAK4-IRAK1 complex resulting in the prevention of IRAK1-IRAK4 complex, suggesting that IRAK-M is a negative regulator of TLR signaling pathways.

These observations suggest that IRAKs play distinct roles in IL-1R/TLR signaling. IRAK4 acts upstream for IRAK1, IRAK2 can signal with TIRAP/Mal, and IRAK-M is a negative regulator of the signaling pathway.

5.3
TNF Receptor-associated Factor 6 (TRAF6)

TRAF6 belongs to a TRAF family, which is characterized by an N-terminal RING finger domain and a conserved C-terminal TRAF-domain. The RING finger domain is also found in a large family of E3 ubiquitin ligases, and TRAF6 was recently shown to function as an ubiquitin ligase, along with an E2 ligase complex consisting of Ubc13 and Uev1A/Mms2, to catalyze the formation of a polyubiquitin chain through lysine-63 (K63) of ubiquitin. TRAF6$^{-/-}$ mice exhibited severe osteopetrosis due to a defect in osteoclast differentiation. B cells and fibroblasts from TRAF6$^{-/-}$ mice showed defective responses to IL-1β, TLR ligands, and CD40 ligation. However, LPS-induced IFN-inducible gene expression was not impaired in TRAF6$^{-/-}$ cells, suggesting that TRAF6 is dispensable for the MyD88-independent pathway.

5.4
TAK1 and TAB1, TAB2 and TAB3

TAK1 was originally identified as a TGFβ-activated kinase and subsequently reported to be involved in the IL-1β signaling pathway. Ubiquitin-dependent activation of TAK1 results in phosphorylation of IKKβ and MKK6 leading to NF-κB and MAP kinase activation. Knockdown of TAK1 abolished the IKK activation induced by TNFα and IL-1β stimulation.

TAB1 and TAB2 were identified as adaptor proteins that associate with TAK1. TAB3, a TAB2 homolog, is also involved in IL-1β-mediated NF-κB and MAP kinase activation. TAB2 and TAB3 bind preferentially to lysine-63-linked polyubiquitin chains through a zinc finger domain. However, IL-1β-mediated activation of NF-κB and MAP kinases was not impaired in MEFs from TAB2$^{-/-}$ mice. TAB2 and TAB3 may therefore compensate for each other's functions *in vivo*.

6
Intracellular Recognition of PAMPs

TLRs have a transmembrane domain and are localized on the plasma membrane, ER and lysosome membrane, suggesting that TLRs are not suitable for recognizing pathogens invaded into the cytosol of cells. Various pathogens such as intracellular bacteria and viruses are supposed to be detected by PRRs in the cytosol. In mammals, the NOD-LRR family is thought to play a role in detecting cytosolic pathogens. The members of this family share a tripartite domain structure consisting of a C-terminal LRR motifs, a central nucleotide-binding oligomerization domain (NOD), and N-terminal protein–protein interaction motifs, such as caspase recruitment domains (CARDs), pyrin domains, or a TIR domain. NOD1 and NOD2, containing N-terminal CARD domains have been extensively studied for their ligands. NOD1 detects γ-D-glutamyl-meso diaminopimelic acid (iE-DAP) found in gram-negative bacterial peptidoglycan. NOD1$^{-/-}$ macrophages fail to produce cytokines in response to iE-DAP. In contrast, NOD2 is a receptor for muramyl-dipeptide (MDP) derived from bacterial PGN (Fig. 4). A missense point mutation in the human *Nod2* gene is correlated with susceptibility to Crohn's disease, an inflammatory bowel disease. A recent study revealed that NOD2 functioned to suppress TLR2-mediated activation of NF-κB in response to PGN. NOD2$^{-/-}$ mice showed enhanced IL-12 production in response to PGN stimulation, which may explain the mechanism by which NOD2 controls the susceptibility to the disease. In addition, a mutation in the NOD domain of NOD2 is the cause of Blau disease, an autosomal dominant disorder characterized by granulomatous arthritis, uveitis, and skin rash. Ligand binding of NOD1 and NOD2 causes their oligomerization and results in the activation of NF-κB through recruitment of a serine/threonine kinase, RIP2/RICK. NODs associate with RIP2/RICK through their CARD domains by a homophilic interaction. Although the number of the NOD-LRR family members is growing, the functions of members other than NOD1 and NOD2 have yet to be clarified.

Viral dsRNA is synthesized in infected cells for the viral replication. Viral dsRNA is a virus-specific molecular structure to be recognized by hosts. Most cells infected with virus produce type I IFNs in a TLR3 independent manner, implying that the mechanism recognizing dsRNA in cytoplasm plays an important role in IFN response. Cytoplasmic proteins, protein kinase R (PKR), and retinoic acid inducible protein-I (RIG-I) are implicated in viral dsRNA recognition. PKR belongs to a protein family containing dsRNA-binding domains, and is induced in response to stimulation with IFNs. Activation of PKR results in growth arrest of the cells via phosphorylating eIF2α. Several reports have shown that virus-induced type I IFN production is modestly impaired in PKR$^{-/-}$ mice. In addition, PKR$^{-/-}$ mice show defective induction of apoptosis

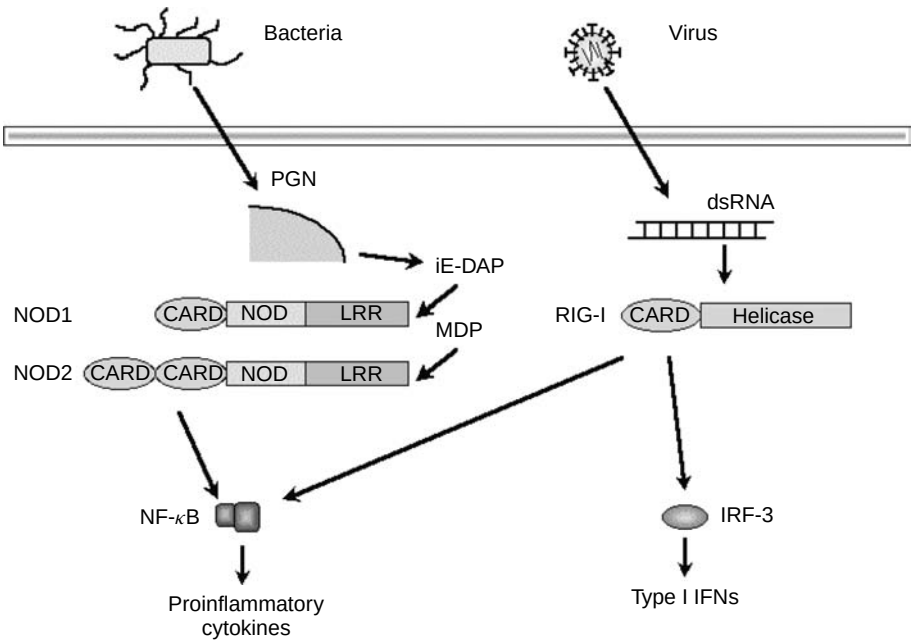

Fig. 4 Cytoplasmic recognition of microorganism. Intracellular bacteria are recognized by NOD-LRR family members in the cytoplasm. Components of PGN, iE-DAP, and MDP, are recognized by NOD1 and NOD2, respectively. NOD1 and NOD2 are composed of C-terminal LRR, central NOD domain, and N-terminal CARD domain. They detect bacterial components via LRR motifs and signal through CARD domain to activate NF-κB. When RNA viruses invade into a cell, dsRNA is generated for viral replication. RIG-I detects viral dsRNA via its helicase domain. CARD domain of RIG-I is required for activating IRF3 and NF-κB.

in response to poly I:C and bacterial burden. The other candidate of dsRNA recognition molecules, RIG-I, is also an IFN-inducible protein. RIG-I contains an N-terminal CARD domain and a DExD/H box RNA helicase domain. RIG-I interacts with poly I:C, and overexpression of RIG-I in cells conferred Newcastle Disease virus (NDV)- and dsRNA-mediated IFN response (Fig. 4). Knockdown of RIG-I by RNA interference abolished the IFN response upon NDV infection, suggesting that RIG-I is responsible for detecting RNA viruses in the cytosol of cells. It will be interesting to explore the relationships between cytoplasmic and extracellular virus recognition in host defense.

7 Pathways Activating IFN Response

Type I IFNs mainly consist of IFNα and IFNβ. The IFNα family is encoded by a cluster of multiple genes, whereas there is only a single *IFNβ* gene. Type I IFNs are strongly induced upon viral infection and stimulation with TLR ligands. Among the TLR ligands, stimulation with the TLR3, TLR4, TLR7, and TLR9 ligands, but not the TLR2 ligand, is known to upregulate type I IFN production.

Recent studies have revealed the involvement of two IKK-related kinases, inducible IκB kinase (IKK-I; also known as IKKε) and TRAF-family member-associated

NF-κB activator (TANK) TANK-binding kinase 1 (TBK1 also known as T2K), in IFN production upon TLR stimulation and viral infection. IKK-i and TBK1 can directly phosphorylate IRF3 and IRF7 *in vitro* and activate the IFNβ promoter. TBK1$^{-/-}$, but not IKK-i$^{-/-}$, MEFs showed severely impaired induction of IFNβ and IFN-inducible genes in response to LPS, intracellular introduction of poly I:C and RNA virus infection. Activation of IRF3, but not NF-κB, in response to LPS and poly(I:C) was also diminished in TBK1$^{-/-}$ cells. IKK-i/TBK1 double-deficient MEFs failed to express any detectable levels of IFNβ and IFN-inducible genes in response to poly I:C, indicating that both IKK-i and TBK1 contribute to the IFN pathway. Furthermore, these observations suggest that the signaling pathways triggered by TLR stimulation as well as cytoplasmic viral recognition converge at the level of TBK1/IKK-i (Fig. 3).

The transcription factors, IRF3 and IRF7 are essential for the expression of type I IFNs in response to viral infection and TLR stimulation. Phosphorylated IRF3 forms homodimers that translocate into the nucleus and bind to the IFN-stimulated regulatory element (ISRE) present in the promoters of a set of IFN-inducible genes to induce their expression. Initially, secreted type I IFNs activate their receptors in both autocrine and paracrine manners. This activates IFN-stimulated gene factor 3 (ISGF3), which consists of signal transducer and activator of transcription (STAT) 1, STAT2, and p40/IRF9 to amplify the response. IRF7 is also implicated in the production of IFNα. Cells deficient in IRF3 and IRF7 expression are defective in type I IFN production in response to viral infection. Moreover, IRF3$^{-/-}$ mice showed defective induction of IFNβ in response to LPS, indicating a critical role for IRF3 in the TLR4-induced IFN response.

Plasmacytoid DCs (pDCs) are known to be recruited to inflamed lymph nodes and produce large amounts of type I IFNs. pDCs express TLR7 and TLR9, and stimulation by their ligands induces IFNα production. Since this IFNα induction is independent of TBK1, the existence of a novel mechanism is hypothesized. It was revealed that MyD88 and TRAF6 can directly interact with IRF7, and induce its nuclear translocation and IFNα upregulation. (Fig. 3). Since IRF7 constitutively expresses only in pDCs, the interaction can explain the mechanism on how TLR7 and TLR9 signaling produce a large amount of IFNα in pDCs.

In summary, TLR3- and TLR4-induced IFN production is mediated through a TRIF-TBK1/IKK-i-IRF3 signaling pathway. In contrast, TLR7- and TLR9-induced IFNα production in pDC is mediated by direct association of MyD88 and IRF7, suggesting that type I IFN production is regulated by stimulus- and tissue-specific mechanisms.

8 Perspectives

As described above, the mammalian innate immune system comprised of three different mechanisms in the recognition of microorganisms. The lectin-complement pathway can detect microorganisms at the extracellular space to destroy them directly and/or cause their opsonization to facilitate phagocyte recognition. TLRs, located on the cell membrane, are responsible for activating innate immune cells by detecting microorganisms on the cell surface or in the lysosome. Finally, there are cytoplasmic PRRs recognizing viruses infected into cells and some intracellular

bacteria. Secreted type I IFNs have a strong antiviral activity by expressing IFN-inducible genes.

Recent extensive studies have revealed ligands of most TLRs and their signaling pathways. Although the cytoplasmic domains of TLRs resemble each other, the adaptor molecule(s) required for the signaling pathways differ depending on the TLR. However, most studies are implemented using PAMPs purified from pathogens to simplify the experimental system. Therefore, for comprehensive understanding of host–pathogen interactions, the next logical step would be to study the host immune responses against whole pathogens in addition to purified PAMPs and to analyze how the responses are modulated by the TLR system.

In contrast, studies on the cytoplasmic recognition of pathogens are just beginning. Although a gene family encoding putative PRR, NOD-LRR proteins, has been identified, the functions of the members are yet to be identified. The response to intracellular bacteria is more complex than previously expected. For example, in *L. monocytogenes* infection, the type I IFN response exacerbates the course of infection. IFNα/β $R^{-/-}$ mice as well as IRF3$^{-/-}$ mice are more resistant to *Listeria* infection compared to wild-type mice. *Listeria*-induced IFN response is activated independent of the TLR pathway, implying that the bacteria may take advantage of host cytoplasmic recognition system for improving their environment in hosts. Virus recognition by the host occurs both inside and outside of cells. Since viruses are complex and diverse, the host must develop an elaborate mechanism(s) to sense the invasion of various viruses.

Together, the innate immune system is not evoking a nonspecific response and random inflammation, but controlling a coordinated and adaptive response depending on the pathogen. Future studies will clarify the intricate pathogen–host relationship, allowing for more targeted therapeutic treatment of infection via modulating the innate immune system.

See also Antigen Presenting Cells (APCs); Autoantibodies and Autoimmunity; Cellular Interactions; Dendritic Cells; Immune Defence, Cell Mediated; Immunologic Memory; Immunology; Pathogens: Innate Immune Reponses.

Bibliography

Books and Reviews

Akira, S., Takeda, K. (2004) Toll-like receptor signalling, *Nat. Rev. Immunol.* **4**, 499–511.

Barton, G.M., Medzhitov, R. (2002) Toll-like receptors and their ligands, *Curr. Top. Microbiol. Immunol.* **270**, 81–92.

Beutler, B., Rietschel, E.T. (2003) Innate immune sensing and its roots: the story of endotoxin, *Nat. Rev. Immunol.* **3**, 169–176.

Hoffmann, J.A. (2003) The immune response of Drosophila, *Nature* **426**, 33–38.

Inohara, N., Nunez, G. (2003) NODs: intracellular proteins involved in inflammation and apoptosis, *Nat. Rev. Immunol.* **3**, 371–382.

Janeway, C.A., Travers, P., Walport, M., Shlomchik, M. (2001) *Immunobiology: The Immune System in Health & Disease*, Garland Publishing, New York.

Turner, M.W. (2003) The role of mannose-binding lectin in health and disease, *Mol. Immunol.* **40**, 423–429.

Primary Literature

Alexopoulou, L., Holt, A.C., Medzhitov, R., Flavell, R.A. (2001) Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3, *Nature* **413**, 732–738.

Alexopoulou, L., Thomas, V., Schnare, M., Lobet, Y., Anguita, J., Schoen, R.T., Medzhitov, R., Fikrig, E., Flavell, R.A. (2002) Hyporesponsiveness to vaccination with Borrelia burgdorferi OspA in humans and in TLR1- and TLR2-deficient mice, *Nat. Med.* **8**, 878–884.

Auerbuch, V., Brockstedt, D.G., Meyer-Morse, N., O'Riordan, M., Portnoy, D.A. (2004) Mice lacking the type I interferon receptor are resistant to Listeria monocytogenes, *J. Exp. Med.* **200**, 527–533.

Blander, J.M., Medzhitov, R. (2004) Regulation of phagosome maturation by signals from toll-like receptors, *Science* **304**, 1014–1018.

Carrero, J.A., Calderon, B., Unanue, E.R. (2004) Type I interferon sensitizes lymphocytes to apoptosis and reduces resistance to listeria infection, *J. Exp. Med.* **200**, 535–540.

Chamaillard, M., Hashimoto, M., Horie, Y., Masumoto, J., Qiu, S., Saab, L., Ogura, Y., Kawasaki, A., Fukase, K., Kusumoto, S., Valvano, M.A., Foster, S.J., Mak, T.W., Nunez, G., Inohara, N. (2003) An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid, *Nat. Immunol.* **4**, 702–707.

Cheung, P.C., Nebreda, A.R., Cohen, P. (2004) TAB3, a new binding partner of the protein kinase TAK1, *Biochem. J.* **378**, 27–34.

Couillault, C., Pujol, N., Reboul, J., Sabatier, L., Guichou, J.F., Kohara, Y., Ewbank, J.J. (2004) TLR-independent control of innate immunity in Caenorhabditis elegans by the TIR domain adaptor protein TIR-1, an ortholog of human SARM, *Nat. Immunol.* **5**, 488–494.

Diebold, S.S., Kaisho, T., Hemmi, H., Akira, S., Reis e Sousa, C. (2004) Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA, *Science* **303**, 1529–1531.

Diebold, S.S., Montoya, M., Unger, H., Alexopoulou, L., Roy, P., Haswell, L.E., Al-Shamkhani, A., Flavell, R., Borrow, P., Reis e Sousa, C. (2003) Viral infection switches non-plasmacytoid dendritic cells into high interferon producers, *Nature* **424**, 324–328.

Fitzgerald, K.A., McWhirter, S.M., Faia, K.L., Rowe, D.C., Latz, E., Golenbock, D.T., Coyle, A.J., Liao, S.M., Maniatis, T. (2003) IKKepsilon and TBK1 are essential components of the IRF3 signaling pathway, *Nat. Immunol.* **4**, 491–496.

Fitzgerald, K.A., Palsson-McDermott, E.M., Bowie, A.G., Jefferies, C.A., Mansell, A.S., Brady, G., Brint, E., Dunne, A., Gray, P., Harte, M.T., McMurray, D., Smith, D.E., Sims, J.E., Bird, T.A., O'Neill, L.A. (2001) Mal (MyD88-adapter-like) is required for Toll-like receptor-4 signal transduction, *Nature* **413**, 78–83.

Georgel, P., Naitza, S., Kappler, C., Ferrandon, D., Zachary, D., Swimmer, C., Kopczynski, C., Duyk, G., Reichhart, J.M., Hoffmann, J.A. (2001) Drosophila immune deficiency (IMD) is a death domain protein that activates antibacterial defense and can promote apoptosis, *Dev. Cell.* **1**, 503–514.

Girardin, S.E., Boneca, I.G., Viala, J., Chamaillard, M., Labigne, A., Thomas, G., Philpott, D.J., Sansonetti, P.J. (2003) Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection, *J. Biol. Chem.* **278**, 8869–8872.

Girardin, S.E., Boneca, I.G., Carneiro, L.A., Antignac, A., Jehanno, M., Viala, J., Tedin, K., Taha, M.K., Labigne, A., Zahringer, U., Coyle, A.J., DiStefano, P.S., Bertin, J., Sansonetti, P.J., Philpott, D.J. (2003) Nod1 detects a unique muropeptide from gram-negative bacterial peptidoglycan, *Science* **300**, 1584–1587.

Hawn, T.R., Verbon, A., Lettinga, K.D., Zhao, L.P., Li, S.S., Laws, R.J., Skerrett, S.J., Beutler, B., Schroeder, L., Nachman, A., Ozinsky, A., Smith, K.D., Aderem, A. (2003) A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to legionnaires' disease, *J. Exp. Med.* **198**, 1563–1572.

Hayashi, F., Smith, K.D., Ozinsky, A., Hawn, T.R., Yi, E.C., Goodlett, D.R., Eng, J.K., Akira, S., Underhill, D.M., Aderem, A. (2001) The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5, *Nature* **410**, 1099–1103.

Heil, F., Hemmi, H., Hochrein, H., Ampenberger, F., Kirschning, C., Akira, S., Lipford, G., Wagner, H., Bauer, S. (2004) Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8, *Science* **303**, 1526–1529.

Hemmi, H., Kaisho, T., Takeuchi, O., Sato, S., Sanjo, H., Hoshino, K., Horiuchi, T., Tomizawa, H., Takeda, K., Akira, S. (2002) Small anti-viral compounds activate immune cells via the TLR7 MyD88-dependent signaling pathway, *Nat. Immunol.* **3**, 196–200.

Hemmi, H., Takeuchi, O., Sato, S., Yamamoto, M., Kaisho, T., Sanjo, H., Kawai, T., Hoshino, K., Takeda, K., Akira, S. (2004) The roles of two IkappaB kinase-related kinases in lipopolysaccharide and double stranded RNA signaling and viral infection, *J. Exp. Med.* **199**, 1641–1650.

Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., Matsumoto, M., Hoshino, K., Wagner, H., Takeda, K., Akira, S. (2000) A Toll-like receptor recognizes bacterial DNA, *Nature* **408**, 740–745.

Hochrein, H., Schlatter, B., O'Keeffe, M., Wagner, C., Schmitz, F., Schiemann, M., Bauer, S., Suter, M., Wagner, H. (2004) Herpes simplex virus type-1 induces IFN-alpha production via Toll-like receptor 9-dependent and -independent pathways, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11416–11421.

Hoebe, K., Du, X., Georgel, P., Janssen, E., Tabeta, K., Kim, S.O., Goode, J., Lin, P., Mann, N., Mudd, S., Crozat, K., Sovath, S., Han, J., Beutler, B. (2003) Identification of Lps2 as a key transducer of MyD88-independent TIR signalling, *Nature* **424**, 743–748.

Horng, T., Barton, G.M., Medzhitov, R. (2001) TIRAP: an adapter molecule in the Toll signaling pathway, *Nat. Immunol.* **2**, 835–841.

Horng, T., Barton, G.M., Flavell, R.A., Medzhitov, R. (2002) The adaptor molecule TIRAP provides signalling specificity for Toll-like receptors, *Nature* **420**, 329–333.

Hoshino, K., Takeuchi, O., Kawai, T., Sanjo, H., Ogawa, T., Takeda, Y., Takeda, K., Akira, S. (1999) Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the Lps gene product, *J. Immunol.* **162**, 3749–3752.

Hsu, L.C., Park, J.M., Zhang, K., Luo, J.L., Maeda, S., Kaufman, R.J., Eckmann, L., Guiney, D.G., Karin, M. (2004) The protein kinase PKR is required for macrophage apoptosis after activation of Toll-like receptor 4, *Nature* **428**, 341–345.

Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J.P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C.A., Gassull, M., Binder, V., Finkel, Y., Cortot, A., Modigliani, R., Laurent-Puig, P., Gower-Rousseau, C., Macry, J., Colombel, J.F., Sahbatou, M., Thomas, G. (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease, *Nature* **411**, 599–603.

Ishitani, T., Takaesu, G., Ninomiya-Tsuji, J., Shibuya, H., Gaynor, R.B., Matsumoto, K. (2003) Role of the TAB2-related protein TAB3 in IL-1 and TNF signaling, *EMBO J.* **22**, 6277–6288.

Kanayama, A., Seth, R.B., Sun, L., Ea, C.K., Hong, M., Shaito, A., Chiu, Y.H., Deng, L., Chen, Z.J. (2004) TAB2 and TAB3 activate the NF-kappaB pathway through binding to polyubiquitin chains, *Mol. Cell.* **15**, 535–548.

Kawai, T., Adachi, O., Ogawa, T., Takeda, K., Akira, S. (1999) Unresponsiveness of MyD88-deficient mice to endotoxin, *Immunity* **11**, 115–122.

Kawai, T., Takeuchi, O., Fujita, T., Inoue, J., Muhlradt, P.F., Sato, S., Hoshino, K., Akira, S. (2001) Lipopolysaccharide stimulates the MyD88-independent pathway and results in activation of IFN-regulatory factor 3 and the expression of a subset of lipopolysaccharide-inducible genes, *J. Immunol.* **167**, 5887–5894.

Kobayashi, K., Hernandez, L.D., Galan, J.E., Janeway, C.A. Jr., Medzhitov, R., Flavell, R.A. (2002) IRAK-M is a negative regulator of Toll-like receptor signaling, *Cell* **110**, 191–202.

Krug, A., Luker, G.D., Barchet, W., Leib, D.A., Akira, S., Colonna, M. (2004) Herpes simplex virus type 1 activates murine natural interferon-producing cells through toll-like receptor 9, *Blood* **103**, 1433–1437.

Kurt-Jones, E.A., Popova, L., Kwinn, L., Haynes, L.M., Jones, L.P., Tripp, R.A., Walsh, E.E., Freeman, M.W., Golenbock, D.T., Anderson, L.J., Finberg, R.W. (2000) Pattern recognition receptors TLR4 and CD14 mediate response to respiratory syncytial virus, *Nat. Immunol.* **1**, 398–401.

Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.M., Hoffmann, J.A. (1996) The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in Drosophila adults, *Cell* **86**, 973–983.

Liberati, N.T., Fitzgerald, K.A., Kim, D.H., Feinbaum, R., Golenbock, D.T., Ausubel, F.M. (2004) Requirement for a conserved Toll/interleukin-1 resistance domain protein in the Caenorhabditis elegans immune response, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6593–6598.

Lomaga, M.A., Yeh, W.C., Sarosi, I., Duncan, G.S., Furlonger, C., Ho, A., Morony, S., Capparelli, C., Van, G., Kaufman, S., van der Heiden, A., Itie, A., Wakeham, A., Khoo, W., Sasaki, T., Cao, Z., Penninger, J.M., Paige, C.J., Lacey, D.L., Dunstan, C.R., Boyle, W.J., Goeddel, D.V., Mak, T.W. (1999) TRAF6 deficiency results in osteopetrosis and defective interleukin-1, CD40, and LPS signaling, *Genes Dev.* **13**, 1015–1024.

Lund, J., Sato, A., Akira, S., Medzhitov, R., Iwasaki, A. (2003) Toll-like receptor 9-mediated recognition of Herpes simplex virus-2 by plasmacytoid dendritic cells, *J. Exp. Med.* **198**, 513–520.

Lye, E., Mirtsos, C., Suzuki, N., Suzuki, S., Yeh, W.C. (2004) The role of IRAK-4 kinase activity in IRAK-4-mediated signaling, *J. Biol. Chem.* **279**, 40653–40658.

Nagai, Y., Akashi, S., Nagafuku, M., Ogata, M., Iwakura, Y., Akira, S., Kitamura, T., Kosugi, A., Kimoto, M., Miyake, K. (2002) Essential role of MD-2 in LPS responsiveness and TLR4 distribution, *Nat. Immunol.* **3**, 667–672.

Naito, A., Azuma, S., Tanaka, S., Miyazaki, T., Takaki, S., Takatsu, K., Nakao, K., Nakamura, K., Katsuki, M., Yamamoto, T., Inoue, J. (1999) Severe osteopetrosis, defective interleukin-1 signalling and lymph node organogenesis in TRAF6-deficient mice, *Genes Cells* **4**, 353–362.

Netea, M.G., Sutmuller, R., Hermann, C., Van der Graaf, C.A., Van der Meer, J.W., van Krieken, J.H., Hartung, T., Adema, G., Kullberg, B.J. (2004) Toll-like receptor 2 suppresses immunity against Candida albicans through induction of IL-10 and regulatory T cells, *J. Immunol.* **172**, 3712–3718.

Ninomiya-Tsuji, J., Kishimoto, K., Hiyama, A., Inoue, J., Cao, Z., Matsumoto, K. (1999) The kinase TAK1 can activate the NIK-I kappaB as well as the MAP kinase cascade in the IL-1 signalling pathway, *Nature* **398**, 252–256.

O'Connell, R.M., Saha, S.K., Vaidya, S.A., Bruhn, K.W., Miranda, G.A., Zarnegar, B., Perry, A.K., Nguyen, B.O., Lane, T.F., Taniguchi, T., Miller, J.F., Cheng, G. (2004) Type I Interferon Production Enhances Susceptibility to Listeria monocytogenes Infection, *J. Exp. Med.* **200**, 437–445.

Ogura, Y., Bonen, D.K., Inohara, N., Nicolae, D.L., Chen, F.F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R.H., Achkar, J.P., Brant, S.R., Bayless, T.M., Kirschner, B.S., Hanauer, S.B., Nunez, G., Cho, J.H. (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease, *Nature* **411**, 603–606.

Ozinsky, A., Underhill, D.M., Fontenot, J.D., Hajjar, A.M., Smith, K.D., Wilson, C.B., Schroeder, L., Aderem, A. (2000) The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 13766–13771.

Perry, A.K., Chow, E.K., Goodnough, J.B., Yeh, W.C., Cheng, G. (2004) Differential Requirement for TANK-binding Kinase-1 in Type I Interferon Responses to Toll-like Receptor Activation and Viral Infection, *J. Exp. Med.* **199**, 1651–1658.

Poltorak, A., He, X., Smirnova, I., Liu, M.Y., Van Huffel, C., Du, X., Birdwell, D., Alejos, E., Silva, M., Galanos, C., Freudenberg, M., Ricciardi-Castagnoli, P., Layton, B., Beutler, B. (1998) Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene, *Science* **282**, 2085–2088.

Qin, J., Jiang, Z., Qian, Y., Casanova, J.L., Li, X. (2004) IRAK4 kinase activity is redundant for interleukin-1 (IL-1) receptor-associated kinase phosphorylation and IL-1 responsiveness, *J. Biol. Chem.* **279**, 26748–26753.

Sakaguchi, S., Negishi, H., Asagiri, M., Nakajima, C., Mizutani, T., Takaoka, A., Honda, K., Taniguchi, T. (2003) Essential role of IRF-3 in lipopolysaccharide-induced interferon-beta gene expression and endotoxin shock, *Biochem. Biophys. Res. Commun.* **306**, 860–866.

Sato, M., Suemori, H., Hata, N., Asagiri, M., Ogasawara, K., Nakao, K., Nakaya, T., Katsuki, M., Noguchi, S., Tanaka, N., Taniguchi, T. (2000) Distinct and essential roles of transcription factors IRF-3 and IRF-7 in response to viruses for IFN-alpha/beta gene induction, *Immunity* **13**, 539–548.

Sharma, S., tenOever, B.R., Grandvaux, N., Zhou, G.P., Lin, R., Hiscott, J. (2003) Triggering the interferon antiviral response through an IKK-related pathway, *Science* **300**, 1148–1151.

Shimazu, R., Akashi, S., Ogata, H., Nagai, Y., Fukudome, K., Miyake, K., Kimoto, M. (1999) MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4, *J. Exp. Med.* **189**, 1777–1782.

Suzuki, N., Suzuki, S., Duncan, G.S., Millar, D.G., Wada, T., Mirtsos, C., Takada, H., Wakeham, A., Itie, A., Li, S., Penninger, J.M., Wesche, H., Ohashi, P.S., Mak, T.W., Yeh, W.C. (2002) Severe impairment of interleukin-1 and Toll-like receptor signalling in mice lacking IRAK-4, *Nature* **416**, 750–756.

Takaesu, G., Kishida, S., Hiyama, A., Yamaguchi, K., Shibuya, H., Irie, K., Ninomiya-Tsuji, J., Matsumoto, K. (2000) TAB2, a novel adaptor protein, mediates activation of TAK1 MAPKKK by linking TAK1 to TRAF6 in the IL-1 signal transduction pathway, *Mol. Cell.* **5**, 649–658.

Takeuchi, O., Hoshino, K., Akira, S. (2000) Cutting edge: TLR2-deficient and MyD88-deficient mice are highly susceptible to Staphylococcus aureus infection, *J. Immunol.* **165**, 5392–5396.

Takeuchi, O., Hoshino, K., Kawai, T., Sanjo, H., Takada, H., Ogawa, T., Takeda, K., Akira, S. (1999) Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components, *Immunity* **11**, 443–451.

Takeuchi, O., Kawai, T., Muhlradt, P.F., Morr, M., Radolf, J.D., Zychlinsky, A., Takeda, K., Akira, S. (2001) Discrimination of bacterial lipoproteins by Toll-like receptor 6, *Int. Immunol.* **13**, 933–940.

Takeuchi, O., Sato, S., Horiuchi, T., Hoshino, K., Takeda, K., Dong, Z., Modlin, R.L., Akira, S. (2002) Cutting edge: role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins, *J. Immunol.* **169**, 10–14.

Tanabe, T., Chamaillard, M., Ogura, Y., Zhu, L., Qiu, S., Masumoto, J., Ghosh, P., Moran, A., Predergast, M.M., Tromp, G., Williams, C.J., Inohara, N., Nunez, G. (2004) Regulatory regions and critical residues of NOD2 involved in muramyl dipeptide recognition, *EMBO J.* **23**, 1587–1597.

Underhill, D.M., Ozinsky, A., Hajjar, A.M., Stevens, A., Wilson, C.B., Bassetti, M., Aderem, A. (1999) The Toll-like receptor 2 is recruited to macrophage phagosomes and discriminates between pathogens, *Nature* **401**, 811–815.

Watanabe, T., Kitani, A., Murray, P.J., Strober, W. (2004) NOD2 is a negative regulator of Toll-like receptor 2-mediated T helper type 1 responses, *Nat. Immunol.* **5**, 800–808.

Yamamoto, M., Sato, S., Hemmi, H., Uematsu, S., Hoshino, K., Kaisho, T., Takeuchi, O., Takeda, K., Akira, S. (2003) TRAM is specifically involved in the Toll-like receptor 4-mediated MyD88-independent signaling pathway, *Nat. Immunol.* **4**, 1144–1150.

Yamamoto, M., Sato, S., Hemmi, H., Hoshino, K., Kaisho, T., Sanjo, H., Takeuchi, O., Sugiyama, M., Okabe, M., Takeda, K., Akira, S. (2003) Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway, *Science* **301**, 640–643.

Yamamoto, M., Sato, S., Hemmi, H., Sanjo, H., Uematsu, S., Kaisho, T., Hoshino, K., Takeuchi, O., Kobayashi, M., Fujita, T., Takeda, K., Akira, S. (2002) Essential role for TIRAP in activation of the signalling cascade shared by TLR2 and TLR4, *Nature* **420**, 324–329.

Yang, Y.L., Reis, L.F., Pavlovic, J., Aguzzi, A., Schafer, R., Kumar, A., Williams, B.R., Aguet, M., Weissmann, C. (1995) Deficient signaling in mice devoid of double-stranded RNA-dependent protein kinase, *EMBO J.* **14**, 6095–6106.

Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., Taira, K., Akira, S., Fujita, T. (2004) The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses, *Nat. Immunol.* **5**, 730–737.

Zhang, D., Zhang, G., Hayden, M.S., Greenblatt, M.B., Bussey, C., Flavell, R.A., Ghosh, S. (2004) A toll-like receptor that prevents infection by uropathogenic bacteria, *Science* **303**, 1522–1526.

Insulin Resistance, Diabetes and its Complications

Dominic S. Ng
University of Toronto, Toronto, Ontario, Canada

1 Introduction 26
1.1 Gene–Environment Interactions 27

2 Peripheral Insulin Resistance 28
2.1 Insulin Signaling 28
2.2 Exercise and Glucose Uptake by Skeletal Muscle 29
2.2.1 Skeletal Muscle Insulin Resistance in Diabetics 30
2.2.2 Exercise as Therapeutic Intervention in Insulin Resistance 31
2.2.3 Acute Exercise and Insulin Resistance 32
2.2.4 Chronic Exercise and Insulin Resistance 32
2.3 Adipocytes, Obesity, and Insulin Resistance 32
2.3.1 Insulin and Adipocytes 32
2.3.2 Adipose Tissue in Glucose Homeostasis and Energy Storage 33
2.3.3 Lipotoxicity 34
2.3.4 Visceral Obesityff 35
2.3.5 Effect of Adipocyte Size on Insulin Resistance 35
2.3.6 Adipose Tissue as Endocrine Organ 37
2.3.7 Leptin 37
2.3.8 Adiponectin (Acrp30/AdipQ) 38

3 Pancreatic β-cell Dysfunction 39
3.1 Effect of FFA on β-cell Apoptosis 39

4 Insulin Resistance and the Liver 40
4.1 Hepatic Insulin Resistance and Dyslipidemia 41
4.2 The Role of SREBP in Hepatic Insulin Resistance 42

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

5 Insights from Gene-targeted Mouse Models 43
5.1 Insulin Receptor (IR) Knockout Mice 43
5.2 β-cell-specific IR Knockout Mice (βIRKO) 43
5.3 Muscle-specific IR Knockout Mice (MIRKO) 44
5.4 Liver-specific IR Knockout Mice (LIRKO) 44
5.5 Fat-specific IR Knockout Mice (FIRKO) 45

6 Hyperglycemic Complications 45
6.1 Clinical Impact – Microvascular Complications 45
6.1.1 Intracellular Oxidative Stress 46
6.1.2 The Polyol Pathway Flux 47
6.1.3 Advanced Glycation End Products 47
6.1.4 Activation of Protein Kinase C 47
6.1.5 The Hexosamine Pathway 48
6.1.6 The Effect of Hyperglycemia on Insulin Signaling 48
6.1.7 The Effect of Hyperglycemia on β-cell Function 49
6.1.8 The Effect of Hyperglycemia on Oxidative Stress 49
6.2 Clinical Impact – Macrovascular Complications 50
6.2.1 Relevance of NADPH Oxidase in Atherogenesis 50
6.2.2 The Effect of Hyperglycemia on Thrombosis 51
6.2.3 The Effect of Hyperglycemia on Endothelial Dysfunction 52

7 Genetic Approaches to New Genes – Candidate Genes and Susceptibility Genes 52
7.1 Monogenic Syndromes of Insulin Resistance and Diabetes Mellitus 53
7.1.1 Insulin Receptor Mutations 53
7.1.2 Insulin Gene Mutations 53
7.1.3 Familial Partial Lipodystrophy (FPLD) 53
7.1.4 Insulin Receptor Substrates 54
7.1.5 Maturity Onset Diabetes of the Young (MODY) 54
7.2 Common Genetic Variants of Other Candidate Genes 55
7.2.1 ADRB3 55
7.2.2 PPARγ 55
7.3 Genome Scanning and Positional Cloning Strategies 55

8 Summary and Conclusions 56

Acknowledgment 57

Bibliography 57
Books and Reviews 57
Primary Literature 58

Keywords

Apoptosis
A cell-death sequence that is largely genetically programmed.

Atherogenesis
The complex array of pathologic processes that result in the development of the atherosclerotic plaques, lesions responsible for the most common type of occlusion arterial diseases.

β-cells
The endocrine cell type in the pancreas that specializes in the secretion of insulin in response to ambient glucose concentrations and other signals.

Dyslipidemia
Altered blood levels of various lipoprotein fractions when compared with normal subjects.

Insulin Resistance
A reduced cellular response of any tissue to the binding of insulin to its cell surface insulin receptors. At the whole organism level, insulin resistance impairs insulin-induced glucose uptake by the tissues, a major predisposing factor to elevated blood glucose.

Insulin Signaling
The binding of insulin to its receptors results in a sequence of biochemical alterations, primarily phosphorylation, of a series of downstream intracellular signaling molecules, which eventually leads to the target cellular events, for example, translocation of the glucose transporter, GLUT4, induction of target gene transcription, etc.

Knockout Mice
Genetically engineered mice with a gene or a selected segment of DNA being disrupted structurally; a powerful mouse model to study the functional importance of the segment of DNA being disrupted.

Lipodystrophy
A clinical condition of either complete or partial loss of fat depot. Many specific syndromes are genetically determined, but it can also be drug-induced, most notably those caused by protease inhibitors commonly used for the treatment of AIDS.

Lipotoxicity
Dysfunction of nonadipose tissues as a result of fat accumulation.

Diabetes mellitus, a chronic progressive metabolic disorder of glucose homeostasis, is a major cause of morbidity and mortality. It is broadly categorized into type 1 and type 2, each with distinctly different pathogenesis. Diabetes mellitus is extremely prevalent, with type 2 making up nearly 90% of the cases and the incidence continues to increase at an alarming rate, in parallel with aging and the increasing prevalence of obesity worldwide. Although the pathogenesis of these two major categories is different, they share similar long-term complications as a result of inadequate glycemic control. It is well established that poor glycemic control is the major causative factor in the development of microvascular complications, affecting primarily the kidney, eyes, and peripheral nerves. Recent data suggest that hyperglycemia also plays a critical role in the development of macrovascular complications. To date, a complete normalization of glucose metabolic disturbances in diabetes remains an elusive goal. Even excellent glycemic control evades a large number of affected subjects. However, advances in our understanding of the pathogenesis of the disorder at cellular and molecular levels over the past several decades had led to significant improvements in treatment outcomes.

Insulin resistance is highly prevalent, and particularly so in Western societies. It can be found in a number of metabolic disorders, including the metabolic syndrome, impaired glucose tolerance, and in its extreme form, type 2 diabetes. In the current paradigm, insulin resistance plays a central role in the predisposition of an individual to the development of type 2 diabetes. Initially, compensatory overproduction of insulin by pancreatic β-cells enables an individual to sustain normal blood glucose levels. Over time, progressive β-cell dysfunction and loss of cell mass leads to relative insulin deficiency and diabetes ensues. Obesity not only directly causes insulin resistance, but also contributes to the progressive failure of β-cells through a number of mechanisms, including free fatty acid–induced β-cell apoptosis. Such maladaptive responses are also associated with an increased ectopic deposition of fat to nonadipose tissues, in turn, aggravating the preexisting insulin resistance.

The development of clinical diabetes is a result of combined genetic susceptibility and environmental stressors. With only a handful of exceptions, diabetes is a polygenic disorder. Identification of candidate genes may provide a more refined framework for either better understanding of the pathophysiological processes and/or for identifying novel targets for pharmacological interventions.

1 Introduction

Diabetes mellitus is a chronic progressive disease caused by inherited and/or acquired deficiency in the production of insulin, or by the ineffectiveness of the insulin produced, resulting in elevated blood glucose levels. Diabetes is broadly categorized as type 1, in which the pancreas fails to produce insulin, and type 2, in which the body fails to respond properly to the action of insulin. Diabetes is extremely prevalent. Recent data suggest

that approximately 150 million people worldwide have diabetes mellitus and amongst them, approximately 90% have type 2. This number will increase to at least 300 million by 2025 (WHO 2003). Much of this increase will occur in developing countries due to aging, unhealthy lifestyle, and the increasing epidemic of obesity, all predisposing factors for type 2 diabetes. Although the two types of diabetes differ fundamentally in their pathogenesis, they share common features in the clinical complications including those directly attributable to chronic hyperglycemia. With regard to pathogenesis, this review will focus on insulin resistance, the cardinal pathogenic factor for the development of type 2 diabetes mellitus.

Insulin resistance of the peripheral tissue predates the development of frank diabetes. This has been documented in a variety of conditions, including obesity, as well as in individuals with impaired fasting glucose. To maintain glucose homeostasis, the pancreatic β-cells overproduce insulin, leading to hyperinsulinemia. As the disease progresses, β-cell decompensation develops and frank diabetes ensues.

1.1 Gene–Environment Interactions

Type 2 diabetes is increasingly being recognized as a polygenic disorder. Nevertheless, environmental factors are considered responsible for the alarming rise in diabetes incidence in all age groups including the young since no major change in the gene pool has occurred. Furthermore, the success of intervention trials in preventing the progression to diabetes in vulnerable populations by reducing the glycemic load through acarbose (Study to Prevent Non-Insulin-Dependent Diabetes Mellitus, or STOP-NIDDM trial) or modest lifestyle change of low saturated fat weight-loss diets with regular exercise (Finnish Diabetes Prevention Study and the US Diabetic Prevention Program) has emphasized the role of environmental factors, especially when combined. Current thinking on the genesis of diabetes links genes and environment through insulin resistance. A progression is seen where insulin resistance increases over time, promoted by adiposity, especially intra-abdominal fat. Insulin levels rise and then decline over time as blood glucose levels rise due to pancreatic insufficiency. Genetic factors influencing insulin resistance, body fat distribution and β-cell failure are obviously key factors in this scenario. Interestingly, more than 80% of subjects with diabetes are obese. Observational studies support the notion that diet high in glycemic load increases the risk of development of type 2 diabetes, whereas a high fiber diet reduces the risk. Current data also support the notion that the quality of dietary fat significantly influences the development of insulin resistance and diabetes. Physical activities also confer beneficial effects on insulin resistance and improvement of glycemic control. A single bout of aerobic exercise can significantly reduce plasma glucose levels. Likewise, endurance exercise training also improves insulin resistance and glycemic control.

From the genetic standpoint, the age-adjusted increase in diabetes and obesity in the last 100 years in western nations cannot be attributed to a change in the gene pool. Rather, it is likely, in large measure, to be related to changes in diet and lifestyle. From the evolutionary perspective, our genetic makeup is ill-adapted to life in the twenty-first century. Such a lifestyle is typified by lack of need to exercise and a continuous abundance of food.

Furthermore, the nature of the food has changed. For example, traditional starchy foods that are digested more slowly, such as beans, peas, lentils, pumpernickel breads, and parboiled rice, have been replaced in the modern diet by white bread, potatoes, and glutinous white rice that tend to taste sweeter due to salivary amylolytic digestion in the mouth. As a result, lower glycemic index foods have tended to be exchanged for higher glycemic index foods. This dietary change has been even more marked in traditional cultures where contemporary carbohydrate foods have replaced very low glycemic index traditional foods in the diets, Pima Indians and Australian Aborigines of the groups that have seen some of the greatest increases in diabetes incidence. The thrifty genome has therefore been challenged by twenty-first century lifestyles. Much of the difficulty in this situation of identifying the single gene or group of genes responsible is due to the fact that the whole genome evolved to deal with a radically different lifestyle. The genetic cause of our present epidemic is therefore polygenic.

2 Peripheral Insulin Resistance

Clearance of blood glucose by target tissues is mediated by a family of facilitative transporters (GLUTs). While genes for 11 GLUT isoforms have been identified, only 7 have been shown to transport sugars (GLUTs 1–5, 8, 9). GLUT1 is expressed in erythrocytes and in vascular endothelial cells in the brain, whereas GLUT3 is expressed in neurons. GLUT1 and GLUT3 act in concert to mediate glucose uptake past the blood–brain barrier into the neurons. GLUT2 is expressed widely in a number of tissues including liver, intestine, kidney, and pancreatic β-cells. These transporters participate as part of the glucose sensing system in the β-cells and as the transport system in the intestinal cells for the absorption of dietary glucose. GLUT5 is a fructose transporter, GLUT8 appears to be important in blastocyst development, and GLUT9 is expressed in the brain and leukocytes. GLUT4, found predominantly in skeletal muscle and adipose tissues, is quantitatively the most important insulin-mediated transporter in the clearance of blood glucose. Insulin-mediated rapid cell surface translocation of the preformed and intracellularly sequestered GLUT4 remains the most investigated hypothesis for the mechanism of how insulin mobilizes glucose uptake.

2.1 Insulin Signaling

The binding of insulin to its heterotetrameric receptor (Fig. 1) results in intracellular autophosphorylation of the β-subunit and increases tyrosine kinase activity. These activated insulin receptors modulate the activities of their target effector proteins through phosphorylating a number of proximal intermediate proteins including members of the insulin receptor substrate family (IRS1-4), Shc, Cbl, and so on. An important example of this sequential activation is the tyrosine phosphorylation of IRS proteins resulting in the availability of docking sites of p85, the regulatory subunit of the type 1A phosphatidylinositol 3-kinase (PI3K). PI3K has long been recognized to play a critical role in mediating insulin-stimulated GLUT4 translocation. Activation of PI3K leads to the formation of 3′ phosphoinositide, which in turn, sequentially phosphorylates and activates phophoinositidyl-dependent protein kinase (PDK1) and PKB (also

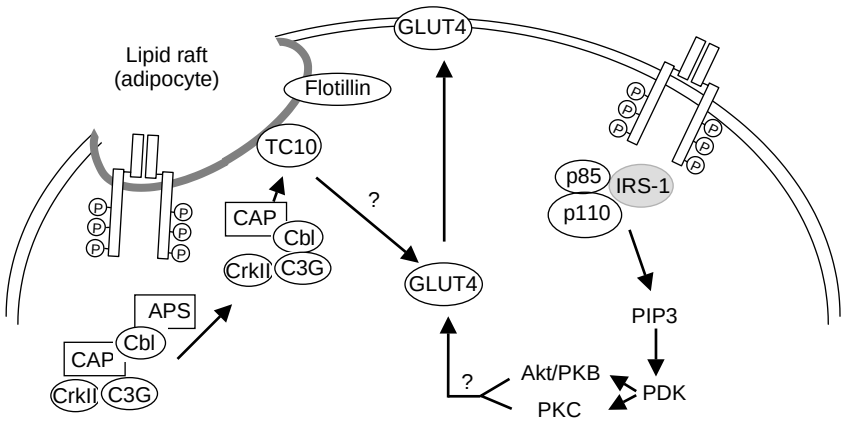

Fig. 1 Signaling cascades in insulin receptor mediated GLUT4 translocation – cross talk between IRS-mediated pathways and lipid raft-associated pathways.

known as Akt). In addition, protein kinase Cζ (PKCζ) has also been shown to participate in insulin-stimulated GLUT4 translocation. However, data to date have not identified a dominant role for either pathway.

Several lines of evidence strongly suggest that activation of PI3K in the generation of phosphatidylinositol 3,4,5-triphosphate is necessary but not sufficient for mediating the insulin-stimulated GLUT4 translocation. Novel insights recently developed were built on the observations of the comparative studies on the proximal insulin receptor tyrosine–phosphorylated substrates. It was found that in adipocytes, only insulin can induce tyrosine phosphorylation of Cbl. Two accessory proteins, APS and Cbl-associated adaptor protein (CAP), appear to act in concert. The former probably enables the insulin receptor to tyrosine phosophorylate Cbl and the latter, through its ability to interact with the caveolar protein flotillin, recruits the complex to the lipid raft in adipocytes. During this process, a small adaptor protein CrkII is also recruited to the lipid raft through its association with the phosphorylated Cbl. CrkII, in turn, interacts with the guanylnucleotide exchange factor, C3G, through one of the two tandem SH3 domains. Putting it together, insulin induces tyrosine phosphorylation of Cbl and recruits the APS/Cbl/CAP/CrkII/C3G complex to the lipid raft at the plasma membrane. Subsequent screens for small GTP binding proteins lead to the discovery of TC10, a member of the Rho family, which is instrumental in the mediation of insulin-induced, PI3K-independent stimulation of GLUT4 translocation. It therefore appears that the TC10 pathway functions in parallel with PI3K to stimulate fully GLUT4 translocation in response to insulin and these two major components are spatially separated and distinctly compartmentalized. The discovery of this novel lipid raft–associated pathway has provided a new framework for unraveling the complexity of insulin-mediated GLUT4 translocation.

2.2 Exercise and Glucose Uptake by Skeletal Muscle

Glucose can also be taken up by skeletal muscles by a number of factors

independent of insulin, including muscle contractions, hypoxia, nitric oxide, and bradykinin. Increased GLUT4 translocation has been demonstrated in mediating the increased glucose uptake for each of the above factors. In spite of its great therapeutic potential, the cellular and molecular signaling mechanism of the contraction-induced glucose uptake by skeletal muscles remains to be fully elucidated. Many candidates, including glycogen, adenosine, nitric oxide (NO), PKC, and 5′ AMP-activated protein kinase (AMPK) have been studied.

It has been shown that glycogen content of skeletal muscle modulates glucose uptake during contractions. Physiologically, glycogen-store depletion after exercise requires repletion to restore contractile functions. If carbohydrate supplementation is inadequate after exercise, glycogen is stored only incompletely. On the other hand, with the availability of carbohydrates, the restoration mechanism may indeed result in supercompensated state, rebuilding glycogen levels to beyond that of the resting levels. Studies showed that surface membrane GLUT4 protein content after contraction was inversely related to the initial glycogen levels in fast-twitch fibres. This modulator effect of glycogen has also been reported in humans. These authors reported a near doubling of plasma glucose clearance rate during one hour of exercise at 70% of maximum oxygen consumption when muscle glycogen content was 185 μmol gm^{-1} versus 800 μmol gm^{-1}. On the contrary, the stimulatory effect of insulin on glucose transport/GLUT4 recruitment was decreased when glycogen was supercompensated. The mechanism behind acute enhanced insulin action after a bout of exercise remains poorly understood. Three to four hours post exercise (when most of the acute effects of exercise ought to have subsided), insulin's effects on glucose uptake and glycogen synthase activity are increased but there was no evidence of increase in insulin-signaling intermediaries. It was suggested that this acute effect of insulin post exercise might be mediated by pathways downstream of Akt and glycogen synthase kinase 3 (GSK3). In isolated muscles, inhibition of protein synthesis with cycloheximide did not prevent muscle contraction–induced increases in insulin sensitivity, suggestive of a protein synthesis independent pathway. The molecular mechanisms underlying the modulatory effects of glycogen on insulin action are not known. Current *in vivo* data point to the possibilities of (1) increased insulin-induced activation of Akt and (2) increased AMPK activities in association with a low glycogen content. However, it is intriguing to note that glycogen depletion is not a necessary condition to observe AMPK activation with 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) and further research is required to elucidate the precise role of glycogen in postexercise insulin action.

2.2.1 Skeletal Muscle Insulin Resistance in Diabetics

It has been well established that skeletal muscle is the major site of insulin-stimulated glucose uptake in the whole body glucose homeostasis. *In vivo* studies consistently showed significant impairment in postprandial insulin mediated glucose uptake in both type 2 diabetic patients and in glucose tolerant first-degree relatives. Tracer studies further localized the defect to the nonoxidative pathways for glucose metabolism. *Ex vivo* assays quantifying insulin-stimulated glucose uptake on abdominal muscle surgical biopsy specimens were developed to show a

marked impairment in glucose transport in morbidly obese patients with or without diabetes. In concordance with this finding, muscle biopsy specimens similarly obtained from lean to moderately obese type 2 diabetic subjects revealed a significant reduction in insulin-stimulated transport of 3-O-methylglucose, a non-metabolizable glucose analog, at both physiological and pharmacological concentrations of insulin. Similar defects in insulin-mediated glucose transport have also been reported in rodent models of insulin resistance. The defects of glucose transport observed in diabetics on the basis of *in vitro* and *ex vivo* studies correlate exceedingly well with those of *in vivo* analyses.

In recent years, the technique of open biopsy of the vastus lateralis muscle under local anesthetics has made *ex vivo* specimens more assessable, and therefore, has permitted more systematic analyses. Using an exofacial *bis*-mannose photolabelling technique, Lund et al. were able to show that insulin stimulation of the muscle strip leads to an increase in the cell surface GLUT4 protein level. Since this surface level of GLUT4 is strongly correlated with the 3-O-methylglucose uptake and because of the fact that the GLUT4 turnover rate is not affected by insulin, it is concluded that the major mechanism of glucose transport in human skeletal muscle is GLUT4 translocation to the plasma membrane.

In type 2 diabetic individuals, insulin resistance is not generally associated with a reduction in skeletal muscle GLUT4 levels, but the insulin-induced translocation of GLUT4 is markedly impaired. In insulin-resistant subjects, insulin stimulation of muscle IRS-1 tyrosine phosphorylation and of IRS-1 immunoprecipitable PI3K activity is significantly less compared to insulin-sensitive controls. A reduction in insulin-stimulated Akt/PKB kinase activity in skeletal muscle from insulin-resistant type 2 diabetic individuals has also been reported, albeit not consistently. The role of Akt signaling in the impairment of glucose transport seen in type 2 diabetic individuals remains unclear. There have been a number of reports showing a reduction in insulin-stimulated glycogen synthase activity in diabetic subjects. Insulin-stimulated mitogen-activated protein kinase (MAPK) phosphorylation is not significantly altered in muscles from type 2 diabetic individuals. It appears that skeletal muscle insulin resistance in type 2 diabetic individuals is largely a result of defects in the proximal insulin-signaling steps. Interestingly, a survey of the literature on rodent models of insulin resistance, especially those data based on Zucker *fa/fa* rats, yielded remarkably similar conclusions regarding the molecular defects in mediating the insulin-stimulated glucose transport.

2.2.2 Exercise as Therapeutic Intervention in Insulin Resistance

Although our current state of knowledge in the impact of exercise on glucose homeostasis at the molecular level remains to be fully explored, many lines of evidence point to exercise as a highly desirable form of intervention in the treatment and/or prevention of type 2 diabetes among susceptible individuals. The most compelling evidence comes from the published results on the Diabetes Prevention Program trial and the Finnish Diabetes Prevention Study. Rigorous lifestyle changes, including a minimum of 150 min of physical activity per week to achieve at least a 7% reduction in body weight remains the most efficacious way of preventing the development of type 2 diabetes in susceptible subjects.

2.2.3 Acute Exercise and Insulin Resistance

It is well established that in type 2 diabetes patients, significant glucose lowering can be achieved during an acute, single bout of physical activity. A concomitant reduction in plasma insulin level has also been observed. Mechanistic studies in both type 2 diabetic subjects and in rodent models yielded very similar conclusions. Immediately after a bout of exercise (45–60 min of cycling at 60–70% of V_{max} in the human study), Kennedy et al. reported a 74% increase in plasma membrane GLUT4 protein in vastus lateralis muscle, an improvement comparable to the postexercise increase seen in nondiabetic subjects. This improvement, on the other hand, was not related to a change in GLUT4 protein expression. Similar findings were reported by King et al. in the obese Zucker rat model. These findings are consistent with the observations that contractile-induced skeletal muscle glucose transport remains intact both in animal models of insulin resistance and in type 2 diabetic subjects.

2.2.4 Chronic Exercise and Insulin Resistance

The potential beneficial effect of endurance exercise training in glucose metabolism and insulin resistance has been extensively evaluated. To date, both rodent models and human studies yielded positive results in the effect of exercise training on glucose metabolism. In rodent models, prolonged exercise consistently results in improvement in whole body glucose tolerance, whole body glucose disposal, insulin-stimulated glucose transport, GLUT4 translocation, and GLUT4 protein expression. In insulin-resistant and type 2 diabetics, endurance exercise training leads to improvements in glucose tolerance and whole body insulin-mediated glucose disposal.

The mechanisms underlying the beneficial effects of endurance training on glucose metabolism in insulin-resistant rodent models, insulin-resistant, and type 2 diabetic subjects remain poorly understood. In obese Zucker rats, exercise training results in long-term adaptive changes including upregulation of GLUT4 protein expression, increase GLUT4 translocation, and cell surface GLUT4 after insulin stimulation. However, the impact of exercise on the insulin receptor signaling pathway is less certain. Data pertaining to the effect of exercise on skeletal muscle insulin signaling in diabetic subjects are equally sparse. A better understanding of these molecular determinants will greatly enhance our ability to develop more effective therapeutic exercise strategies to combat insulin resistance and prevent onset of type 2 diabetes.

2.3 Adipocytes, Obesity, and Insulin Resistance

It has long been well recognized that obesity is a strong risk factor for the development of insulin resistance and diabetes. Likewise, up to 80% of type 2 diabetic subjects are either overweight or obese. This strong correlation is also applicable to many genetic isolates with extraordinary high prevalence of type 2 diabetic subjects including the Pima Indians and the Sandy Lake Oji-Cree. While insulin action is key to virtually all facets of adipocyte differentiation, maturation, and physiologic functions, excess fat mass promotes insulin resistance.

2.3.1 Insulin and Adipocytes

Insulin is a key regulator of nearly all aspects of fat-cell biology including

the promotion of the differentiation of preadipocytes to adipocytes. In mature adipocytes, insulin action simultaneously promotes lipogenesis and inhibits lipolysis. Meanwhile, insulin stimulates the uptake of glucose in insulin-sensitive tissues via activation of the glucose transporter and the uptake of fatty acids from circulating triglycerides (TG) in very low density lipoproteins and chylomicrons through increasing the lipoprotein lipase activity. In addition, insulin has been known to be mitogenic to the adipose tissues as well as promoting coordinated gene expressions through transcription factors of either the sterol response element-binding protein 1 (SREBP1) variants or the forkhead protein family.

In adipocytes, insulin also acts through binding to and activating the insulin receptor with a subsequent activation of PI3K, the latter being a necessary but not a sufficient condition for the induction of the Glut transporters for glucose uptake. The non-PI3K factor in adipocyte insulin signaling remains to be fully elucidated. Previous studies implicated both the Akt/PKB pathway and the PKCλ/ζ pathway, but the exact role of each in mediating glucose transport remains controversial. As discussed earlier, the lipid raft–associated TC10 pathway is also implicated in the non-PI3K signaling. In the case of adipocytes, it is further suggested that this TC10 pathway interacts with contical actin, the adipocyte-specific unique actin cytoskeleton organization, in the modulation of GLUT4 translocation.

It has been suggested that pathways that mediate the various metabolic effects of insulin may diverge downstream of the PI3K step, providing a partial explanation for the differential sensitivity of different pathways to the action of the same ambient concentration of insulin. For example, a significantly lower concentration of insulin is adequate to inhibit adipocyte lipolysis but not sufficient to reverse the hyperglycemia through increased glucose uptake. In insulin-resistant and diabetic states, adipose tissue shares similar functional defects in insulin signaling including impaired binding of insulin to the receptors, phosphorylation of the receptors, and the downstream IRS. Meanwhile, adipose tissues also manifest tissue-specific defects that are distinct from those of skeletal muscles, including a selective reduction in IRS-1 expression resulting in the recruitment of IRS-2 as the main docking protein for PI3K. In skeletal muscles, both IRS-1 and IRS-2 protein levels are normal and the signaling defect was caused by a reduction in PI3K activity in association with both IRS-1 and -2.

2.3.2 Adipose Tissue in Glucose Homeostasis and Energy Storage

The quantitative importance of adipose tissue in the regulation glucose disposal has been continuously revised. On the basis of earlier metabolic studies, it has been held that insulin-mediated glucose uptake by adipose tissue, at least in lean individuals, is relatively insignificant when compared with skeletal muscle. More recent studies in humans as well as in transgenic animal models seem to suggest the contrary. Selective overexpression of the *GLUT4* gene in adipose tissue resulted in a significant increase in whole body insulin sensitivity. A selective disruption of the *GLUT4* gene in adipose tissue led to a degree of insulin resistance comparable to that seen in the skeletal muscle-specific *glut4* knockout mice.

In humans, treatment with the β3 selective agonist CL316,243 increases glucose uptake selectively in fat, which is also associated with an increase in whole body

glucose uptake and suppression of hepatic glucose production. While these observations collectively attest to the possibility that direct insulin-mediated uptake of glucose by adipose tissue may parallel that of skeletal muscle quantitatively, one needs to allow for contributions of other nondirect mediators.

Adipose tissue has long been viewed as a major energy depot. Excess energy intake is stored in adipocytes in the form of TGs. Insulin is a key activator of not only the action of lipoprotein lipase in the uptake of lipoprotein-derived fatty acids but also inhibits the activity of hormone-sensitive lipase to prevent lipolysis. In addition, insulin also acts directly on the insulin-responsive sterol response element–binding protein 1 (*SREBP1*) gene whose gene products, SREBP1a and SREBP1c serve as transcription factors in regulating the expressions of a series of genes including those involved in promoting endogenous lipogenesis. Insulin-deficient and insulin-resistant states including obesity are associated with enhanced lipolysis of the adipose tissue, resulting in an elevated level of circulating free fatty acids (FFA). Increased adipose lipolysis may result in a variety of dysmetabolic consequences including peripheral insulin resistance, hepatic insulin resistance, increased hepatic TG overproduction, and aggravation of pancreatic β-cell insulin secretory function. More recently, elevated FFA concentration has also been shown to result in myocardial dysfunction. Collectively, this FFA-induced multiorgan dysfunction has been dubbed lipotoxicity. The unifying concept behind such otherwise unrelated organ dysfunction is that of excessive intracellular accumulation of TGs as a consequence of nutrient excess and altered energy disposal. In this paradigm, leptin action seems to be the key switch in the proper channeling of the metabolic fate of the excess FA flux commonly seen in insulin-resistant states and obesity.

2.3.3 Lipotoxicity

Caloric intake in excess of expenditure results in an increase in adiposity and associated insulin resistance and hyperinsulinemia. Insulin resistance at the adipose tissue level results in increased lipolysis but the compensatory hyperinsulinemia may impact directly on lipid metabolism in nonadipose tissues. *SREBP1*, a gene encoding a lipogenic transcription factor, remains insulin-responsive in the insulin-resistant milieu and has been shown to be upregulated in a variety of nonadipose tissues and may play an important role in endogenous lipogenesis observed in these tissues. Excess circulating FFAs are capable of entering nonadipose tissues in the body in a fatty acid binding protein independent manner. Under normal circumstances, such excess free FA is efficiently channeled to the β-oxidation pathway and the excess energy dissipated as heat. However, in the absence of leptin action (either due to leptin deficiency or leptin resistance), nonoxidative metabolism of FFAs prevail leading to tissue steatosis, an accumulation of TGs in the nonadipose tissues. There is also evidence suggesting that high levels of FFAs and their subsequent metabolism increase the consumption of antioxidants and cause oxidative damage. In the mouse model of leptin deficiency (*ob/ob*), leptin receptor deficiency (*db/db*), and a rat model of leptin unresponsiveness (*fa/fa*), marked obesity is associated with a generalized form of steatosis, namely, significant accumulation of TG in skeletal muscle, cardiac muscle, pancreatic beta cells, kidneys, and liver. In a murine model of congenital lipodystrophy, a paucity of

adipose tissue also results in severe insulin resistance and steatosis in multiple nonadipose tissue depots. More intriguingly, the metabolic defect can be reversed by either treatment with leptin or with white fat transplantation. On the other hand, a similar transplantation experiment with white adipose without leptin fails to correct the metabolic abnormalities in another model for lipoatrophy, attesting to the critical importance of leptin in the reversal of insulin resistance and diabetes. In this particular model, a paucity of adipose tissue results in an obligate shunting of excess energy to nonadipose tissues. A concomitant lack of leptin thus promotes nonoxidative disposal of FA in nonadipose tissues, resulting in TG accumulation.

Studies on the relevance of this lipotoxicity model in humans have entered the literature. Treatment of human subjects with lipodystrophy-associated diabetes and insulin resistance with recombinant leptin has shown very encouraging beneficial effects. In the first study, 8 of the 9 lipodystrophic female subjects had diabetes, low plasma leptin levels, and fatty livers. A twice-daily leptin injection for 4 months resulted in a mean 1.9% reduction in HgA1C, a 6% reduction in fasting plasma TGs, and a 28% reduction in liver volume. In the latter study, 3 patients with severe lipodystrophy, diabetes, hepatic steatosis, and marked leptin deficiency underwent a chronic leptin-replacement regime, which resulted in a marked reversal in insulin resistance, hepatic steatosis, and hypertriglyceridemia.

2.3.4 Visceral Obesityff

Numerous lines of evidence suggest that visceral fat depots are more important in the promotion of dysmetabolic status and coronary risk than subcutaneous fat. In the Quebec Cardiovascular Study, visceral obesity was characterized by a clustering of a metabolic triad, namely, hyperinsulinemia, hyperapolipoprotein B, and small dense LDL, and was associated with a 20-fold increase in the risk of coronary heart disease over 5 years. Similar findings have also been reported in other cohorts.

Since insulin resistance and dyslipidemia are more closely associated with visceral fat mass than subcutaneous fat, a variety of biochemical and metabolic factors have been explored in explaining such differences. Visceral fat is characterized by enhanced lipolysis, resulting in an increase in FFA flux, especially into the portal circulation. Kissebah et al. showed that, *in vitro*, palmitate exposure caused a dose-dependent reduction in cell surface receptor binding of insulin in isolated hepatocytes, suggestive of a reduction in hepatic extraction of insulin. Differences in gene regulation in the different fat depots have also been reported. Collectively, these studies suggest that the deep abdominal peritoneal fat depot is less effective in insulin-mediated glucose uptake, expresses relatively less leptin, and more PPARγ. Plasminogen activator inhibitor-1 (PAI-1) is produced by adipocytes and its plasma level is strongly correlated with the degree of visceral adiposity but not subcutaneous fat in humans. Circulating PAI-1 has been implicated to play a role in thrombus formation and continues to be investigated as an emerging coronary risk factor.

2.3.5 Effect of Adipocyte Size on Insulin Resistance

Several lines of experimental evidence suggest that adipocyte size influences the impact of adipose tissues on the development of diabetes. Enlarged adipocyte size appears to be a better clinical predictor of

diabetes independent of insulin resistance. Paolisso et al. showed that the predictive value of plasma FFA concentrations on diabetes can be accounted for entirely by the adipocyte size. A large fat cell size is predictive of the development of type 2 diabetes in Pima Indians independent of age, sex, percent body fat, and fat distribution.

Though intriguing, the biological basis of such association remains speculative. Early stage obesity, secondary to overnutrition, is characterized by overrepresentation by large adipocytes (hypertrophic obesity). Such large adipocytes might be more resistant to differentiation upon further nutrient overload. As a result, continued caloric intake will progress into an accumulation of fat in nonadipose tissues including muscle, liver, and pancreatic β-cells, resulting in insulin resistance and in case of β-cells, increased apoptosis, and loss of β-cell mass ensue. Additional putative contributing factors include a disproportionally high rate of secretion of TNFα and release of FFA by the large adipocytes, aggravating insulin resistance.

Adipose differentiation is a complex, highly regulated process. The ability for a given adipose depot to recruit preadipocyte differentiation and proliferation influences the cell number and size of the fat cells. Of the various genes and hormones that regulate and modulate the fat cell differentiation machinery, three transcription factors play particularly important roles in our understanding of adipogenesis, namely, peroxisome proliferator activator receptor γ (PPARγ), CAAT enhancer binding proteins (C/EBP)s, and adducin 1 (ADD1)/SREBP1.

PPARγ, a member of the PPAR subfamily of the nuclear hormone receptor superfamily, is expressed at high levels in adipose tissue. Two splice variants, PPARγ-1 and PPARγ-2, have been identified and the former has been shown to be expressed at low levels in a variety of other tissues. This receptor is induced very early in adipocyte differentiation and has been identified as a regulator of many fat-specific genes and a *master* regulator that can trigger the entire adipogenic program. PPARγ forms heterodimers with the retinoic acid receptor α (RARα), which is recognized by the DR-1 sequence in a variety of fat-specific genes including aP2. The key regulatory role of PPARγ in adipogenesis is best exemplified by the efficient way in which *PPARγ* gene transfected fibroblasts can be induced to differentiate into adipocytes.

C/EBPα is induced relatively late in the adipocyte differentiation program. Similar to PPARγ, C/EBPα also binds to many adipose-specific genes. At a physiologic level of expression, C/EBPα is a relatively weak inducer of adipogenesis. However, when coexpressed with PPARγ, C/EBPα can become extremely potent in the induction of adipocyte differentiation.

ADD1/SREBP1 is a basic helix–loop–helix protein independently cloned as an early gene induced in adipogenesis and as a sterol response element–binding protein involved in lipogenesis. SREBP1 and its close homolog SREBP2 are products of two different genes, but the former is expressed as two splice variants SREBP1a and SREBP1c. The primary protein products of each gene are large transmembrane proteins inserted in the endoplasmic reticulum membrane in a hairpin fashion. Two sequential proteolytic steps initiated by SREBP cleavage-activating protein (SCAP) result in the release of the terminal portion of the protein into the nucleus to impart its transcriptional activity. It has been shown that ectopic expression of

SREBP1 in fibroblasts under conditions favorable for adipogenesis augment PPARγ-induced differentiation into adipocytes.

Recent studies have added new insights into the once elusive mechanism by which thiazolidenediones (TZD), a class of synthetic PPARγ ligands already approved for clinical use, ameliorate insulin resistance. Treatment of obese female *fa/fa* Zucker rats with pioglitazone for up to 28 days decreased the insulin resistant and hyperlipidemic states and caused a shift in body adiposity. In the TZD-treated rats, there was a depot-specific increase in the number of smaller fat cells due to the appearance of new fat cells and the shrinkage and/or disappearance of the existing large mature fat cells. This is consistent with the notion that treatment of rats with a PPARγ agonist might have induced the formation of new small adipocytes that may have better lipid storage capacity as well as being more insulin-sensitive, leading to a lower lipolytic activity. These changes may in turn result in a reduction in ectopic, nonadipose TG accumulation. In the case of the skeletal muscles, this may result in a significant degree of improvement in skeletal muscle insulin sensitivity. In humans, adipocyte restructuring after treatment with a TZD (troglitazone or pioglitazone) was shown to increase total body adiposity, but there was a shift from visceral fat to the subcutaneous fat in conjunction with a reduction in insulin resistance. In this case, it appears that an increase in adipose tissue mass after treatment with a TZD due to a redistribution from the large visceral fat to the smaller subcutaneous fat may be critical in the amelioration of insulin resistance.

2.3.6 Adipose Tissue as Endocrine Organ

Recent studies in adipocyte biology have expanded our conventional view of adipose tissue being primarily a fat depot for storage of energy in the form of TGs. In this traditional paradigm, insulin action on the adipocytes plays the key role of stimulating glucose uptake and lipogenesis and simultaneously inhibits lipolysis, promoting fat storage. Defects in fuel partitioning including increased adipose mass and/or increased lipolysis with increased circulating free fatty acids would result in insulin resistance, dyslipidemia, and β-cell dysfunction. We have now accrued increasing evidence that adipose tissues are programmed to secrete a large variety of substances that may act as signaling molecules in a hormonal, paracrine, and/or autocrine fashion. The secretory factors identified to date span a large variety including enzymes (lipoprotein lipase and adipsin), cytokines (tumor necrosis factor α (TNFα), interleukin 6 (IL6)), hormones (adiponectin, leptin, resistin, acylation stimulation protein, and corticosteroids), growth factors (vascular endothelial growth factor), and thrombogenesis modulators (plasminogen activator inhibitor-1 (PAI-1)). These substances are differentially regulated and are functionally diverse. Amongst them, leptin, adiponectin, TNFα, and resistin are most studied for their roles in energy homeostasis, insulin signaling, and glucose/lipid metabolism. The rapid pace of new research findings in this area does not permit an updated discussion within the scope of this paper. Readers are referred to more updated research papers and reviews.

2.3.7 Leptin

The leptin gene was first cloned when the *circulating factor* in the obese *ob/ob* mice was sought. The gene product, leptin, is a 16-kDa protein synthesized and secreted mainly by the white adipose tissue. The *ob/ob* mice, now known to

be leptin deficient, develop severe hyperphagia, reduced metabolic rate, massive obesity, and a variety of endocrinal abnormalities including infertility. Leptin acts through binding with the leptin receptor in the central nervous system with its main function in regulating food intake. Defective leptin receptors have been identified as the cause of obesity, insulin resistance, and diabetes in the *db*/*db* mice. In humans, monogenic causes of leptin deficiency–related obesity have been described but are extremely rare. The large majority of obese humans have an elevated level of circulating leptin, suggesting that *leptin resistance* is a key metabolic defect in obesity. Not surprisingly, administration of leptin to obese humans led to minimal benefit in the reversal of excess body weight and insulin resistance. On the contrary, in the relatively uncommon cases of subjects with partial lipodystrophy, affected subjects have been found to be hypoleptinemic, and leptin treatment resulted in a reversal of the metabolic abnormalities.

The leptin receptor was first isolated from mouse choroid plexus by expression cloning. This receptor is normally expressed at high levels in hypothalamic neurons and in other cell types, including T cells and vascular endothelial cells. There are several isoforms of the leptin receptor with the long form being predominantly expressed in the hypothalamus. Binding and activation of the leptin receptor results in downstream alterations of gene expression through the JAK/STAT pathway. One functional implication of this signaling is its possible linkage with the metabolic action of protein tyrosine phosphatase 1B (PTP1B). In addition to playing a role in the inactivation of insulin signaling through dephosphorylation of the insulin receptor, PTP1B also can dephosphorylate the leptin receptor–associated tyrosine kinase JAK. Leptin-deficient mice made PTP1B deficient were found to have much reduced weight gain and an increased metabolic rate, supportive of the notion that the signaling pathway cross talk is biologically significant at the whole organism level.

Leptin expression is regulated by a number of factors. Positive modulators include intracellular glucosamine (a product of glucose flux), insulin, TNFα, glucocorticoids, interleukin-1, and so on. It is intriguing to note that in cultured adipocytes, leptin mRNA expression is correlated with the lipid content and adipocyte size, but the underlying mechanism for this adaptive change remains unknown. Negative modulators include PPARγ activation, catecholamines, cAMP agonists, β-adrenergic receptor agonists, and thyroid hormones. In addition to the extensively characterized action of leptin in the regulation of energy balance, glucose homeostasis and insulin signaling, leptin has also been implicated in other biological functions, including immune activities by stimulating proliferation of T lymphocytes and in maturation of reproductive organs and in gestation. Future studies are needed to fully understand the biological function of this adipocyte-derived hormone and its role as a therapeutic in the treatment of obesity and diabetes.

2.3.8 Adiponectin (Acrp30/AdipQ)

Adiponectin was originally identified as a highly expressed gene during adipocyte differentiation. The gene was expressed exclusively by adipocytes, and the secreted protein contains an amino-terminal collagenous domain and a carboxy-terminal globular domain, the latter bearing significant homology with collagens X and VIII and complement C1q. Plasma level of adiponectin is inversely associated with

BMI in humans and is reduced in a number of murine models of insulin resistance. Furthermore, adiponectin level is lower in CHD patients compared to subjects without CHD, even after adjustment for BMI. Several lines of evidence suggest that adiponectin may have direct biological action as an insulin sensitizer by reducing hepatic glucose output as well as reducing skeletal TG content. Treatment of mice with adiponectin induced a large and sustained weight loss in a high fat/high sucrose–fed model mouse. Recent data also suggest that adiponectin confers antiatherogenic actions. Data available to date suggest the adiponectin may be a promising therapeutic candidate for the prevention of diabetes and CHD.

3 Pancreatic β-cell Dysfunction

In pancreatic β-cells, ambient fatty acids modulate the glucose-induced insulin secretion. The potency of fatty acids to promote glucose-induced insulin secretion varies with chain length and the degree of unsaturation. Acute lowering of plasma fatty acid levels is associated with a decreased insulin response to glucose. However, longer exposure of islets to FFA results in a reduction in insulin secretion in the hyperglycemic milieu. This effect of differential impact of acute versus chronic exposure of plasma FFA on insulin secretion was also observed in humans. These observations are consistent with the original Randle hypothesis based on studies in cardiac and skeletal muscle. An elevation of FFA would promote oxidation of fatty acids, which in turn inhibits pyruvate dehydrogenase activity, resulting in an accumulation of cytosolic glucose-6-phosphate with associated inhibition of hexokinase for glucose uptake. Since glycolysis is the pathway for glucose metabolism in β-cells, this FA-induced alteration in glucose uptake may play an important role in the development of β-cell secretory dysfunction.

3.1 Effect of FFA on β-cell Apoptosis

Pancreatic islets in diabetics are characterized by a reduced β-cell mass. This morphologic change is postulated to be largely a result of an accelerated β-cell apoptosis during the progression from prediabetic state to diabetes. The notion of a progressive reduction in β-cell mass caused by accelerated β-cell apoptosis as part of the natural history of the development of type 2 diabetes is supported by autopsy studies on obese and diabetic subjects when compared with lean, nondiabetic controls. Excess accumulation of TGs in β-cells, similar to many other nonadipose tissues, is a cardinal feature of the insulin-resistant state. The progressive TG accumulation has been attributed to both an elevated plasma level of FFA as substrate and a shift toward nonoxidative metabolic pathway in conjunction with an increase in endogenous lipogenesis. *In vitro* data strongly suggest that the modest increase in TG content in islets is causative of the compensatory islet cell hyperplasia and hyperinsulinemia. However, further increase in islet TG eventually leads to decompensation, triggering an accelerated apoptosis and precipitous loss of β-cell mass. In conjunction with the progression of the disease, cellular content of ceramide accumulation causes β-cell decompensation by inducing cellular apoptosis. As such, it has been suggested that FFA-induced β-cell apoptosis may, at least in part, be due to

an increased supply of substrate for the synthesis of ceramide. Inhibition of ceramide synthetase with fumonisin B1 can inhibit FFA-induced apoptosis. In parallel, increased plasma FFA is also associated with an induction of iNOS, promoting nitric oxide (NO)-induced apoptosis.

The mechanism of FFA-induced β-cell apoptosis remains incompletely understood. Recent studies suggest that the PKB pathway may play a key role. Selective transgene expression of a constitutively active form of PKB in β-cells resulted in markedly increased β-cell mass largely through preserving cell survival. This may be explained on the basis that PKB phosphorylation prevents the nuclear translocation of a number of its downstream substrates including forkhead box (Fox) transcription factors, glycogen synthase kinase-3α/β (GSK3$\alpha\beta$), bcl-2 associated death agonist (BAD), caspase 9, and murine double minute 2 (MDM2). Nuclear translocation of such factors would promote proapoptotic transactivation. *In vitro* experiments using a pancreatic β-cell line INS-1 showed that not only can FFA induce apoptosis by inhibition of PKB activation but also can inhibit the IGF-1-induced PKB phosphorylation.

Recent studies suggest that leptin can effectively protect the islet cells by promoting β-oxidative metabolism of surplus FFAs and reducing lipogenesis, especially during periods of overnutrition. Intriguingly, all fatty acids are not equally capable of inducing apoptosis. Saturated fatty acids are highly effective in the induction of apoptosis while unsaturated fatty acids are not. Human β-cell culture studies revealed that when mixing the two types of fatty acids, the unsaturated FAs are capable of preventing saturated FA-induced apoptosis.

Functional characterizations of glucagon-like peptides (GLPs) over the recent years have led to the discovery that GLP-1 may protect pancreatic β-cells from apoptosis. GLP-1 confers its biological actions through binding to the GLP1-receptor, a G-protein coupled receptor expressed in pancreatic β-cells. Not only does GLP-1 stimulate insulin secretion, it also exerts a number of other antidiabetic actions including inhibition of glucagon secretion, reduction of food intake and regulation of cell proliferation, differentiation, and apoptosis. Several lines of experimental evidence support the notion that the GLP-1 receptor is the key mediator of the antiapoptotic properties of GLP-1. In Zucker diabetic rats and in *db/db* mice, short-term infusion of GLP-1 was associated with a marked increase in β-cell mass and the development of new islet clusters. *In vivo* treatment of mice with a long-acting GLP-1 analog NN2211 and *in vivo* treatment of exendin-4, a GLP-1 receptor agonist, both led to a reduction in β-cell apoptosis. A more detailed mechanistic understanding of how GLP-1 regulates cell proliferation and apoptosis is potentially useful in the development of GLP-1 based therapeutics for the treatment or prevention of diabetes.

4
Insulin Resistance and the Liver

Hepatic glucose overproduction is a cardinal feature of type 2 diabetes mellitus. In addition, hypertriglyceridemia in association with low levels of high density lipoprotein cholesterol (HDL-C) levels, elevated levels of apolipoprotein (apo) B100, and less buoyant (small dense) low-density lipoproteins (LDL) are frequently observed

in subjects with diabetes and insulin resistance. Recent studies continue to unravel the complex underlying mechanisms both in the cause of hepatic insulin resistance and its metabolic sequelae.

An elevated blood level of FFA has been shown to promote skeletal muscle insulin resistance and β-cell dysfunction. The mechanisms by which elevated levels of FFA contribute to hepatic glucose overproduction are less well known. A recent study by Massillon et al. showed that the glucose-6-phosphatase gene can be induced by an increased level of FFA by Intralipid infusion on a background of pancreatic clamping of the FFA levels. In a somewhat more physiological model, Oakes et al. examined the effect of high-fat feeding in rats on the development of insulin resistance. Compared to chow fed rats, high fat feeding significantly impaired insulin sensitivity, resulting in hyperinsulinemia. While the compensatory increase in insulin level was able to suppress FFA release, it did not normalize the intracellular availability of FA in skeletal muscle. In the liver, glycogen synthesis was reduced because of the down regulation of glycogen synthase. Glucogenesis was enhanced in conjunction with increased FFA oxidation. However, an acute blockade of FFA oxidation with etomoxir reduced gluconeogenesis and yet, increased glucose output was unaffected. This is likely because of an alternate pathway of a sustained increase in the glucose-6-phosphatase/glucose kinase activity ratio, favoring increased glucose output. In humans, a progressive increase in FFA load through Intralipid infusion resulted in an increasing percent contribution of gluconeogenesis to the observed endogenous glucose overproduction. The relative importance of glycogenolysis and gluconeogeneis in FFA-induced hepatic glucose overproduction has been studied in healthy human subjects. Physiologically, insulin suppresses hepatic glucose output more by inhibiting glycogenolysis than by gluconeogenesis. Elevated plasma FFA levels attenuated the suppression of glucose production by interfering with insulin suppression of glycogenolysis. More recently, Lam et al. demonstrated that cell membrane translocation of PKC-δ and increased FFA oxidation may also be major contributors to the FFA-induced hepatic glucose overproduction.

FFAs have also been proposed as the link between intra-abdominal fat and peripheral insulin resistance. Obesity is associated with higher fasting plasma FFA concentrations. Intra-abdominal fat drains via the portal vein directly to the liver. At high concentrations, it has been shown to interfere with hepatic insulin binding and clearance, which even at low FFA concentrations, is only 50% efficient. It is suggested that increased insulin levels in the peripheral circulation secondary to reduced hepatic clearance downregulate peripheral glucose receptors in muscle and in the long term lead to progressive insulin resistance.

4.1 Hepatic Insulin Resistance and Dyslipidemia

Insulin resistance and diabetes are commonly associated with plasma hypertriglyceridemia, elevated apoB100, and a reduced level of HDL-C. While reduced lipoprotein lipase in the light of insulin resistance may have contributed to the accumulation of very low-density lipoproteins (VLDL) in this dyslipidemic state, a hepatic overproduction of VLDL seems to be of primary importance. Many metabolic factors influence the rate of secretion of VLDL (production rate). Since the

backbone apoB100 is constitutively synthesized in great excess and each VLDL particle contains one apoB100 molecule, the rate of apoB100 (hence VLDL) entering the circulation is governed by a number of factors in the VLDL synthetic machinery, including the translocation efficiency, lipidation of apoB with TGs, and the rate of degradation of the protein. Grossly, it appears that it is the lipids available in the ER lumen that dictate the amount of apoB eventually secreted.

Lipidation of apoB with TGs is largely mediated by the microsomal transfer protein (MTP). In addition to participating in the assembly and maturation of the VLDL particles, MTP also plays a role in stimulating apoB production through inhibition of ubiquitination, a key step in apoB degradation. Hepatic expression of MTP has been shown to be modulated by insulin through interaction of an insulin response element in its promoter. In addition, response elements to other transcription factors including adaptor protein complex (AP-1), hepatic nuclear factor 1 (HNF-1), and hepatic nuclear factor 4 (HNF-4) have also been identified. MTP mRNA expression is also regulated by hepatic sterol content and is thought to be mediated by SREBPs.

The cytosolic content of TGs, the major neutral lipid in the mature VLDL, correlates with the rate of VLDL secretion. Increased availability of cytosolic TG, either through de novo synthesis, or through esterification of FFA derived from the circulation, stimulates the VLDL production and secretion. Recently, a series of studies performed on a hamster model of insulin resistance through fructose feeding have advanced our understanding of the molecular mechanisms of insulin resistance–associated hepatic VLDL overproduction. In this model of oral fructose-induced insulin resistance, the authors observed an increase in intracellular apoB stability, and elevated levels of MTP in association with enhanced VLDL particle assembly. Impaired hepatic insulin signaling was evidenced by a reduced tyrosine phosphorylation of the insulin receptor and IRS-1. In addition, elevated protein mass and activity of PTP-1B, and a reduced level of PI3K activity were observed. The level of ER-60, a cysteine protease, was significantly reduced, consistent with the observed increase in apoB stability.

4.2 The Role of SREBP in Hepatic Insulin Resistance

SREBP1a and 1c are two variants of the same *SREBP1* gene, each using a different transcription start site. SREBP1a is a potent transcription activator of all SREBP responsive genes, including those involved in both endogenous cholesterol and fatty acid and TG synthesis. SREBP1c, on the other hand, primarily transactivates the enzymes involved in lipogenesis. In adult liver, SREBP1c is the predominant SREBP-1 isoform. The selective coordinated transduction by SREBP1c of the fatty acid program of synthesis has been demonstrated *in vivo* in a transgenic mouse model, which overexpresses the truncated version of SREBP1c that terminates prior to the membrane attachment domain. In mice, fasting markedly reduces the amount of SREBP1c in the nuclei. There is a corresponding reduction in the mRNA abundance of the SREBP1c target genes, that is, those involved in *de novo* lipogenesis. On refeeding a high carbohydrate/low fat diet, Horton et al. reported a markedly exaggerated overabundance of SREBP1c (increased 4- to 5-fold over nonfasting baseline) in association with a parallel

increase in hepatic mRNA levels of the lipogenic genes.

Recent data on a rodent model showed that the insulin-induced hepatic lipogenesis is mediated by modulating SREBP1c at the transcriptional level. By comparing two distinct murine models of severe insulin resistance, namely, the *ob/ob* (leptin-deficient) mice and the adipose-specific SREBP transgenic (lipoatrophic) mice, marked hyperinsulinemia differentially caused a significant reduction in IRS-2 mRNA but continued to stimulate SREBP1c, resulting in hepatic lipogenesis and VLDL overproduction. A more recent study further defined the role SREBP1c in the obesity–insulin resistance connection. Targeted disruption of SREBP1 in the *ob/ob* mice by cross breeding the two knockout strains selectively ameliorated the enhanced lipogenesis seen in the *ob/ob* mice without any significant alterations in fat mass and insulin resistance.

Diets rich in polyunsaturated fatty acids (PUFA) including 18 : 2 (ω-6), 20 : 5, and 22 : 6 (ω-3) have been shown to significantly lower plasma lipid levels. In a rodent model, dietary PUFA suppresses hepatic lipogenesis by reducing the nuclear content of mature SREBP1. PUFA reduces the nuclear abundance of SREBP1 by first acutely inhibiting the proteolysis of SREBP1 followed by a reduction in hepatic SREBP1 mRNA and precursor protein content. More recently, Yoshikawa et al. also showed that the suppressive effects of PUFA on SREBP1-induced lipogenesis is mediated through inhibition of the binding of the liver X receptor (LXR) to LXR response element (LRE) on the SREBP1 gene. Our understanding of the molecular mechanism of the interaction between fatty acids and transcription factors in lipid and glucose metabolism may continue to define more specific targets for treatment of hepatic insulin resistance and its associated deleterious phenotypes.

5 Insights from Gene-targeted Mouse Models

Transgenic and knockout mouse technology has provided very powerful tools for dissecting the effects of individual genes involved in insulin signaling and effector pathways. The outcomes of these studies have been extensively reviewed and will be summarized here.

5.1 Insulin Receptor (IR) Knockout Mice

The insulin receptor gene has been knocked out in mice both globally and in a tissue-specific manner. The homozygous IR knockout mice (IR−/−) appear normal at birth but rapidly develop severe metabolic phenotypes of insulin resistance at the time of suckling. These include severe hyperglycemia, ketoacidosis, hypertriglyceridemia, and steatohepatitis. Hepatic content of glycogen was reduced. Muscle was partially atrophic and the amount of adipose tissue was reduced. The mice generally died within one week of birth. On the contrary, the phenotype of the heterozygous knockout mice (IR+/−) was relatively mild. One report showed that the mice developed hyperinsulinemia at 4 to 6 months of age but only about 10% eventually developed diabetes. The insulin receptor gene has been disrupted selectively in a number of tissues, which collectively provided great insight into glucose homeostasis.

5.2 β-cell-specific IR Knockout Mice (βIRKO)

By using the cre-loxP mediated conditional gene inactivation technique, the murine IR

gene was disrupted selectively in pancreatic β-cells. The βIRKO mice were normoglycemic and began to develop significant hyperinsulinemia by 6 months of age. However, insulin secretion in response to a glucose load injected peritoneally was impaired, representing a specific impaired first-phase secretory response to glucose and not to arginine at early age. A chronic progressive impairment of glucose tolerance to an intraperitoneal glucose load was detected as early as 2 months of age. The same glucose tolerance test unmasked a reduced level of insulin levels at 4 months of age. This selective loss of first-phase secretory response is remarkably similar to the early stage of development of diabetes in humans. Furthermore, this sercretory defect was observed in spite of a quantitative preservation of β-cell GLUT2 transporter. Although it is unlikely that acute first-phase secretory defects observed in humans is caused by specific insulin receptor gene defects, these data are suggestive of β-cell insulin signaling playing an important role.

5.3 Muscle-specific IR Knockout Mice (MIRKO)

A cre-loxP system was also used to generate muscle-specific IR knockout mice. Skeletal muscle IR expression was reduced by more than 95% in conjunction with a comparable loss of insulin-stimulated IR and insulin receptor substrate-1 (IRS-1) phosphorylation. At a whole body level, these mice showed a selective increase in fat mass, plasma TG, and FFA levels, but their blood glucose levels, insulin levels, and glucose tolerance were normal.

Resting uptake of 2-deoxyglucose (2DG) by skeletal muscle is normal in the MIRKO mice. Likewise, exercise-induced 2DG uptake with and without insulin is also normal. Glycogen synthase activity is also normal in both the resting state and with exercise but not after insulin stimulation.

To further evaluate the role of muscle IR in the development of type 2 diabetes, the authors cross-bred the MIRKO into the heterozygous IR knockout mice IR+/−, and the resultant mice of interest are designated MIRKO-50%. These mice begin to develop hyperinsulinemia by 3 months. At 6 months, hyperinsulinemia becomes more severe and many animals showed hyperglycemia in a fed state.

The relatively mild glycemia phenotype in the MIRKO mice was addressed by Kim et al. In the absence of IR in skeletal muscles, there was a redistribution of the substrate to liver and adipose tissue, resulting in an expanded fat mass, hypertriglyceridemia, and elevated FFA levels, features of the prediabetes syndrome. Such intriguing findings challenge the conventional view that skeletal muscle insulin resistance is the main driving force in the development of type 2 diabetes.

5.4 Liver-specific IR Knockout Mice (LIRKO)

The liver-specific IR knockout mice develop severe insulin resistance and impaired glucose tolerance. There is a loss of inhibition of gluconeogenesis and hepatic glucose overproduction despite hyperinsulinemia, the latter was the result of a combined increase in insulin secretion and reduced clearance. The hepatic glucose overproduction seen in the LIRKO mice is also not responsive to exogenously administered insulin, bringing up the concept that hepatic glucose overproduction can continue even in the presence

of hyperglycemia in the LIRKO mice and that glucose autoregulation of hepatic glucose production does not occur in the absence of insulin signaling in liver. In this model, the severe hyperinsulinemia had led to a simultaneous differential regulation of a number of insulin-responsive genes. While phosphoenolpyruvate carboxykinase (*PEPCK*) gene expression was found to be upregulated in the LIRKO mouse, the expression of glucokinase and liver pyruvate kinase were coordinately downregulated. Likewise, the transcription of the *SREBP1c* gene continues to be positively responsive to insulin. The loss of IR in the LIRKO mouse liver results in a failure of the insulin receptor through divergent pathways that regulate gene expression, which in turn, contribute to the severe metabolic phenotypes. More interestingly, it is the analyses of these coordinated changes in gene expression that led to the identification of cis-acting DNA regulatory elements (insulin response element), which mediate insulin's effect.

5.5
Fat-specific IR Knockout Mice (FIRKO)

Fat-specific IR knockout mice were recently reported. These animals showed a sustained reduction in total fat mass and are protected from age-related glucose intolerance despite consuming higher amount of calories. There was also a reduction in total body TG content and a polarized distribution of fat cell size. In addition, these mice enjoyed a significantly prolonged life span as compared to wild-type mice. More intriguing is the observation that the phenotype of these mice is quite similar to those of calorie-restricted mice, leading the authors to suggest that it is the leanness that is the key determinant to longevity, not caloric restriction *per se*.

6
Hyperglycemic Complications

6.1
Clinical Impact – Microvascular Complications

The clinical spectrum of diabetic microvascular complications has been extensively reviewed. Two landmark clinical trials over the past decade, the Diabetes Control and Complications Trial (DCCT) for type 1 diabetic subjects and United Kingdom Prospective Diabetes Study (UKPDS) for type 2 diabetic subjects, have provided compelling evidence that hyperglycemia is the primary causative factor in the development and progression of microvascular complications, namely, retinopathy, neuropathy, and nephropathy. In these trials, excellent glycemic control has been demonstrated unequivocally to retard the progression of the microvascular complications (DCCT 1993, UKPDS 1998). Diabetic microvascular disease in retina, glomerulus, and vasa nervorum share similar pathogenic changes. At the onset of hyperglycemia, elevated intracellular glucose concentrations directly impact on vascular tone regulators, including reduced bioavailablity of nitric oxide (NO), increased production of vasoconstrictors including endothelin-1 and angiotensin II, and elaboration of permeability factors such as vascular endothelial growth factor (VEGF). Microvascular cell loss and the compensatory endothelial cell proliferation and qualitative and quantitative changes in extracellular matrix predispose to increased vascular permeability. Excessive matrix accumulation promoted

by transforming growth factor beta 1 (TGFβ1) and deposition of periodic acid Schiff (PAS) positive plasma proteins also contribute to microvascular occlusions, leading to local ischemic changes.

Our understanding of how hyperglycemia acts as the common denominator for the development of such an array of microvascular complications has recently been conceptualized by establishing a linkage between four major metabolic pathways: increased polyol pathway, increased advance glycation end-product (AGE) formation, activation of PKC, and increased hexosamine flux. A unifying hypothesis in linking these four pathways has recently been put forth. Evidence available to date suggests that the intracellular excess glucose–induced mitochondrial superoxide (O_2^-) overproduction may be the key in the simultaneous activation of all of these pathways (Fig. 2). In this hypothesis, excess mitochondrial O_{2-} production partially inhibits the glycolytic enzyme GAPDH, which in turns diverts its upstream glycolytic metabolites from glycolysis into pathways of glucose overutilization. Accumulation of cellular glucose would result in the formation of sorbitol and fructose through the polyol pathway. Excess fructose-6-phosphate is metabolized into glucosamine. Glyceraldehyde-3-phosphate in excess would provide a substrate for the increased formation of diacylglycerol for PKC activation as well as increased formation of advanced glycation end (AGE) products.

6.1.1 Intracellular Oxidative Stress

An alternate pathway by which hyperglycemia may confer oxidative stress is the overproduction of superoxide by the mitochondrial electron-transport chain. In the mitochondria, the generation of ATP is coupled to the electron-transport chain, through which molecular oxygen (O_2) may be reduced to water by electron donors NADH and $FADH_2$ from the Krebs cycle. Protons are pumped through the inner membrane of the mitochondria, sustaining an electrochemical gradient and electrons flow through intermediates

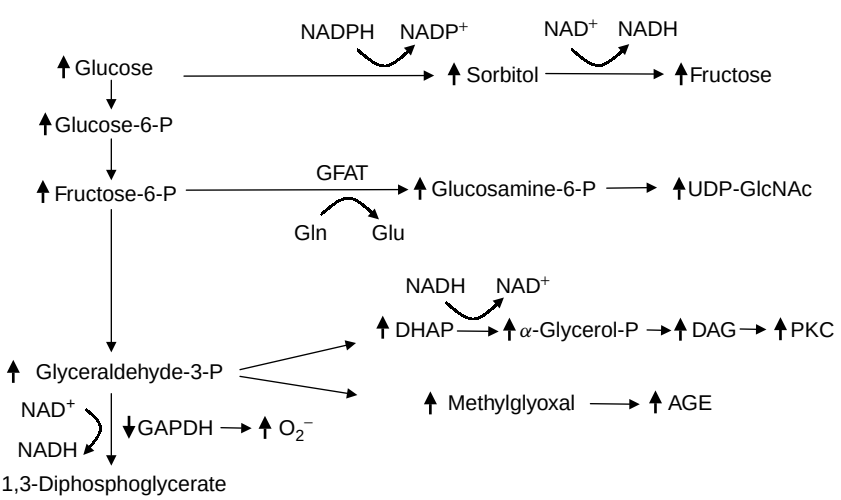

Fig. 2 Impact of elevated cellular glucose levels on multiple metabolic pathways in the promotion of diabetic complications.

such as ubisemiquinone. It has been shown that high glucose availability results in increased production of electron donors, which in turns raises the proton gradient. This, in turn, prolongs the lifetime of the electron-transport intermediates. When this number goes beyond a certain threshold, the intermediate will redirect its reactivity, greatly increasing the conversion of molecular oxygen to superoxide anion (O_2^-). This pathway of hyperglycemia-induced mitochondrial O_2^- overproduction can be attenuated by overexpression of mitochondrial superoxide dismutase (MnSOD) and uncoupling protein-1 (UCP-1).

6.1.2 The Polyol Pathway Flux

Aldose reductase is the rate-limiting enzyme mediating the reduction of glucose and other carbonyl compounds. The reduction is NADPH-dependent and has a low K_m for glucose and with elevated glucose, aldose reductase converts glucose to sorbitol, which in turn can be converted to fructose. These reactions cause simultaneous consumption of NADPH and NAD^+ respectively, leading to a reduction of cellular concentration of NADPH and increased ratio of NAD/NAD^+. A cellular depletion of NADPH would result in a reduction in the regeneration of the reduced form of glutathione (GSH), which in turn aggravates the intracellular oxidative stress. On the other hand, an increased $NADH/NAD^+$ ratio would inhibit the enzyme glyceraldehyde-3 phosphate dehydrogenase (GAPDH), leading to an accumulation of triose phosphate. The downstream effect of this includes an increased formation of methylglyoxal, a precursor of AGEs and diacylglycerol (DAG), the latter leading to the activation of protein kinase C (PKC).

6.1.3 Advanced Glycation End Products

Intracellular glucose at elevated concentration may also be metabolized through nonenzymatic pathways to several reactive dicarbonyls, which in turn react with amino groups of intracellular and extracellular proteins to form AGEs. Intracellular auto-oxidation of glucose leads to the formation of glyoxal. Decomposition of the Amadori product has been shown to generate 3-deoxyglucosone and fragmentation of glyceraldehyde-3-phosphate and dihydroxyacetone phosphate will form methylglyoxal. The two reactive dicarbonyls glyoxal and methylglyoxal may be detoxified by the glyoxalase system. The biological action of intracellular AGEs precursors may damage cells through a number of mechanisms. AGEs' modification of intracellular proteins results in alteration of both the structure and function of cellular proteins. AGEs' formation on extracellular matrix interferes with both matrix–matrix and cell–matrix interactions. AGEs-modified plasma proteins are capable of binding to AGEs receptors (RAGEs) on a variety of cell types including endothelial cells, mesangial cells, and macrophages. In the latter case, binding of AGEs to RAGEs in macrophages may induce production of reactive oxygen species.

6.1.4 Activation of Protein Kinase C

More than 10 isoforms of protein kinase C (PKC) have been described and nine of them are known to be activated by diacylglycerol (DAG). Elevated intracellular levels of glucose promote the *de novo* synthesis of DAG through the glycolytic intermediate dihydroxyacetone. Increased DAG primarily activates the two isoforms PKC-β and -δ, which explains the activation of a large array of downstream cellular and molecular processes responsible for the

pathogenesis of a variety of hyperglycemia-induced complications, most notably, those related to the microvascular diabetic complications.

PKC-β activation has been shown to mediate abnormal blood flow in both retinal and renal vasculatures, possibly through reduced NO production and/or an increased activity of endothelin-1, one of the most potent vasoconstrictors. In mesangial cells, PKC activation induces endothelin-1 stimulated MAP-kinase activity, which may contribute to diabetic nephropathy. PKC activation has also been invoked in hyperglycemia-induced expression of VEGF in smooth muscle cells, which may account for the observed increase in vascular permeability and angiogenesis in chronic hyperglycemia. Increased microvascular matrix accumulation in long standing hyperglycemia may also be attributable to PKC activation. The increase in PAI-1 expression in the hyperglycemic state is a predisposing factor for increased thrombogenicity and the increased NFκB activity may account for the increased expression of a wide array of proinflammatory genes. Lastly, hyperglycemia has been shown to be associated with an increased activity of the vascular NADPH oxidase system and PKC activation as a mediator has also been implicated.

6.1.5 The Hexosamine Pathway

Intracellular accumulation of fructose-6-phosphate of the glycolytic pathway may be shunted to the hexosamine pathway to produce glucosamine. The rate-limiting enzyme for this pathway, glutamine:fructose-6-phosphate amidotransferase (GFAT), converts fructose-6-phosphate to glucosamine phosphate, which will subsequently be converted to UDP-*N*-acetylglucosamine. The pathological significance of this pathway can be related to the hyperglycemia-induced upregulation of the transcription of TGF-α, TGF-β1, and PAI-1. Several lines of evidence suggest that the hexosamine-mediated transcription upregulation of *PAI-1* gene in a number of tissues including vascular cells and mesangial cells is through Sp1. Glycosylation modification of the transcription factor Sp1 may also contribute to the enhancement in its transcription activation activity. In addition to PAI-1, hyperglycemia-induced O-acetylglucosaminylation is implicated in the modulation of the activity and function of a number of other cellular proteins including eNOS modification at the Akt site and a number of PKC isoforms, promoting diabetes-related complications.

6.1.6 The Effect of Hyperglycemia on Insulin Signaling

The direct effect of hyperglycemia on insulin signaling has recently been reviewed. A 4-h incubation of rat entensor digitorum longus (EDL) muscle *ex vivo* with 25 mM glucose resulted in a 50% reduction in insulin-stimulated glucose incorporation into glycogen and a significant reduction in the insulin-stimulated phosphorylation and activation of Akt/PKB. The insulin receptor tyrosine phosphorylation, PI3K activation, and MAP-kinase activation were not affected. Furthermore, the same laboratory also demonstrated previously, both *in vivo* and *in vitro*, that acute hyperglycemia causes cellular accumulation of malonyl coA, an inhibitor of palmitoyl transferase. The resultant accumulation of palmitoyl-coA could be a source of cellular ceramide. Interestingly, incubating the EDL muscle with C2-ceramide also resulted in similar signaling defects, raising

the speculation that ceramide may play a role in mediating this signaling defect.

In cultured human vascular endothelial cells, incubation of high glucose has resulted in a reduced activation of Akt/PKB in conjunction with an increase in apoptosis, a known downstream effect of Akt/PKB. Concomitantly, cellular concentration of diacyl glyceral (DAG) and the PKC activity were both elevated.

6.1.7 The Effect of Hyperglycemia on β-cell Function

It has been well accepted that it is the development of β-cell dysfunction on the background of progressive insulin resistance that precipitates the development of frank type 2 diabetes. The early clinical defect of β-cell dysfunction, a loss of the first-phase insulin secretion in response to glucose stimulation, precedes fasting hyperglycemia by up to 5 years. Initially, hypersecretion of insulin is able to compensate for the insulin resistance, maintaining euglycemia. It is the progressive loss of β-cell mass, in part through accelerated apoptosis, and other progressive dysfunction that precipitates the decompensation, and fasting hyperglycemia ensues. Recent studies examined the potential mechanism underlying glucose-induced β-cell apoptosis. In light of the fact that β-cells have the highest level of expression of *O*-linked monosaccharide *N*-acetylglucosamine (*O*-GlcNAc) transferase (OGT), recent studies suggest that hyperglycemia may cause β-cell apoptosis through first being metabolized through the hexosamine pathway, resulting in an excessive rate of *O*-GlcNAc modification of a variety of proteins and eventually accelerated apoptosis.

Several lines of evidence suggest that hyperglycemia *per se* in turn further aggravates β-cell dysfunction as well as insulin resistance. Persistent hyperglycemia has been shown to desensitize the β-cells to glucose stimulation. The mechanisms include a reduction in the intracellular concentration of signaling molecules, for example, malonyl coA, which increases fatty acid oxidation, and desensitization of the ATP-dependent potassium channels. Meanwhile, hyperglycemia also contributes to the aggravation of peripheral insulin resistance through downregulation of the glucose transporter. This sets up a vicious cycle and the worsening of hyperglycemia will in turn aggravate β-cell function.

6.1.8 The Effect of Hyperglycemia on Oxidative Stress

Oxidant stress continues to be considered to play a critical role in atherogenesis despite the disappointing results of antioxidant trials (HOPE, BHF/MRC). Reactive oxygen species (ROS) generally consist of a number of free radicals, including superoxide anion (O_2^-) and hydroxyl radicals (OH) and their downstream reactive intermediates including peroxynitrite radicals ($ONOO^-$) and hydrogen peroxide (H_2O_2). The free radicals are highly reactive chemically and are able to initiate chain reactions. Superoxide anion radicals are generated by a number of enzyme systems in the vessel walls, including NAD(P)H oxidase, xanthine oxidase, lipoxygenase, cycloxygenase, mitochondrial oxidative phosphorylation, and so on. In defense, a family of superoxide dismutases (SOD) has evolved to transform O_2^- into hydrogen peroxide (H_2O_2). However, the H_2O_2 can, through either a Haber–Weiss reaction or the Fenton reaction, in turn lead to the formation of highly reactive hydroxide (OH^-) and hydroxyl radicals (OH). Alternatively, two endogenous defense systems, catalase and glutathione peroxidase, are able to reduce

H_2O_2. The former reduces to water and molecular oxygen and the latter, using glutathione as cosubstrate, is able to reduce H_2O_2 and lipid peroxides to water and lipid alcohols respectively. In the vessel wall, O_2^- can rapidly inactivate nitric oxide (NO) by reacting with it to produce a highly reactive species, peroxinitrite ($ONOO^-$). Therefore, O_2^- produced in excess not only provides a source for continued production of reactive oxidative species to induce tissue and lipid peroxidation, it also attenuates the potentially beneficial vasodilatory effect of NO in the vessel wall.

NAD(P)H oxidase was initially described in phagocytes as an enzyme complex capable of producing bursts of O_2^- upon stimulation for bacterial killing. The neutrophil NADH oxidase is an enzyme complex consisting of a membrane-bound cytochrome complex b558, made up of two main subunits gp91phox and p22phox and three cytosolic components, p47phox, p67phox, and rac-1. In its quiescent state, the enzyme system is inactive, but upon activation, the cytosolic components are recruited to associate with the membranous cytochrome complex. NADPH, FAD, and heme are bound to their binding sites in the gp91phox subunit and the electron transfer from NADPH to O_2 ensues, resulting in a rapid burst of O_2^- production. In recent years, nonphagocytic NAD(P)H oxidase homologs have been identified in a variety of cell types including endothelial cells, smooth muscle cells, and adventitial fibroblasts in the vessel wall. The relative abundance of each of the components in vascular cells varies with cell type. Specifically, vascular smooth muscle cells contain only trace amounts of gp91phox and p67phox. On the other hand, nox-1, a member of a family of gp91phox-like homologs, is abundant in vascular smooth muscle cells. In endothelial cells and fibroblasts, all five components are present. In both endothelial cells and smooth muscle cells, p22phox and distinct gp91phox homologs appear to be functionally important in the generation of O_2^-.

6.2 Clinical Impact – Macrovascular Complications

6.2.1 Relevance of NADPH Oxidase in Atherogenesis

Several lines of evidence suggest that NADPH oxidases play important roles in the development of atherosclerosis. First, lesion macrophage NADPH oxidase, together with 12/15 lipoxygenase and myeloperoxidase may contribute to ongoing oxidative modification of LDL, perpetuating the progression of the atherosclerotic lesion. Intracellular release of reactive oxygen species (ROS) from the NAD(P)H oxidase system may mediate the action of growth factors and cytokines, contributing to vascular hypertrophy and inflammation. Data from studies of animal models also showed consistent results. Rabbits fed on a long-term high-cholesterol diet show increased vascular O_2^- production and endothelial dysfunction. The potential role of NADPH oxidase in early atherogenesis is further supported by the observation that oscillatory shear stress, but not laminar shear stress, causes a sustained increase in NADPH oxidase–derived O_2^-. These findings may, in part, explain the propensity for atherosclerotic lesions to form at bifurcations. In addition to the flow perturbation, a number of growth factors and cytokines known to contribute to atherogenesis, including IL-6, angiotensin II, and TNF-α, have been shown to activate the NADPH oxdiases. In turn, the NADPH oxidase–derived O_2^- has been shown to

activate the downstream intracellular signaling pathways including p38MAPK (mitogenesis) and Akt/PKB (cell survival kinase) proinflammatory markers including vascular adhesion molecules and MCP-1. Vascular smooth muscle cell proliferation and migration, two important features of atherogenesis, have also been shown to be upregulated by NADPH oxidase–derived O_2^- production.

The direct effect of glucose on NADPH oxidase activity has recently been investigated. By using cultured bovine aortic smooth muscle cells and endothelial cells, Inoguchi et al. demonstrated that incubation of the cells with a high concentration of glucose for 72 h resulted in a PKC-dependent increase in superoxide production. Park et al. recently reported that high glucose concentration (25 mM) directly upregulates endothelin-1 mRNA in bovine retinal endothelial cells in a PKC-β and -δ dependent manner. In human umbilical vascular endothelial cells (HUVEC), endothelin-1 increases the production of O_2^- through upregulation of gp91phox mRNA in a dose-dependent manner and the activation is mediated mainly by the type B endothelin receptor. We recently demonstrated that a 1-h exposure of the freshly isolated C57Bl/6 mouse aorta to 30 mM of glucose increased the aortic O_2^- production by 4 fold, as determined by the lucigenin-derived chemiluminescence method.

6.2.2 The Effect of Hyperglycemia on Thrombosis

It has long been recognized that diabetes mellitus is associated with a hypercoagulable state. Patients with diabetes are significantly more prone to die of a thrombotic event. In concordance with the observed increased thrombotic event rates in diabetic patients, clinical indicators of abnormal coagulation, clotting, fibrinolytic factors, and platelet functions have been studied extensively and have been comprehensively reviewed recently. Virtually all the parameters considered have been found to be abnormal in diabetic patients. On the other hand, the relative contribution of elevated blood glucose on these abnormalities is less well studied. Limited studies are available in exploring the effect of acute hyperglycemia on healthy volunteers. It has been reported that factor VII activity, factor VIII activity, and tissue factor pathway inhibitor (TFPI) can be induced acutely through glucose infusion. Shechter et al. reported a positive correlation between glucose level and platelet-dependent thrombosis in diabetic patients with coronary heart disease.

Elevated plasma levels of plasminogen activator inhibitor type-1 (PAI-1) is one of the established markers of the metabolic syndrome. In addition, increased plasma levels of PAI-1 have also been reported in a number of acute prothrombotic states, including myocardial infarction, deep vein thrombosis, and disseminated intravascular coagulation. PAI-1 is expressed in a variety of tissues and may mediate tissue-specific thrombosis. Obese human subjects also have an elevated level of PAI-1 and the main source of circulating PAI-1 is visceral adipose tissue. In addition, *PAI-1* is also expressed in vascular endothelial cells and smooth muscle cells, and is regulated by a number of external stimuli including insulin. In insulin-resistant states, both *in vivo* and *in vitro* experiments suggest that PAI-1 gene expression continues to be responsive to insulin. Thus, glucose transport and *PAI-1* gene expression are mediated by different pathways.

Recent studies continue to support a positive relationship between elevated

glucose and PAI-1 expression. Gabriely et al. showed that hyperglycemia induces the gene expression of PAI-1 in adipose tissue and is mediated by the activation of the hexosamine pathway. Hyperglycemia has also been shown to induce *PAI-1* expression in vascular endothelial cells, which is at least mediated by the activation of the MAP kinase and PKC pathways.

6.2.3 The Effect of Hyperglycemia on Endothelial Dysfunction

The potential causative role of hyperglycemia in atherogenesis continues to draw a great deal of attention. In the UKPDS trial, Stratton et al. reported a positive association between the updated HgA1C and the incidence of the cardiovascular events. For a 1% reduction in the HgA1C, there was an associated 14% reduction in the incidence of myocardial infarction (8–21%, $p < 0.0001$). The underlying mechanism for this rather modest association remains incompletely understood. In addition to the indirect effect through increased PAI-1 levels, the potential impact of hyperglycemia on endothelial function has been extensively explored but will only be briefly reviewed here.

Endothelial dysfunction is a hallmark of atherosclerosis and may predate the appearance of atherosclerotic plaques. Several lines of evidence suggest that elevated levels of glucose directly impact on endothelial function. In a recent dog model study, Gross et al. showed that acute hyperglycemia-induced by intravenous infusion of D-glucose resulted in the development of endothelial dysfunction as assessed by coronary blood flow responses to acetylcholine, and this effect was mediated by an increased production of ROS. In a cultured endothelial cell model, exposure of bovine aortic endothelial cells to a high concentration of glucose led to a reduction in endothelial NO synthase mRNA expression and glycogen synthase activity. Elevated glucose resulted in reduction in the eNOS gene expression, glycogen synthetase activity, ERK 1,2 signaling, p38, Akt expression, and Cu/Zn superoxide-dismutase activity. Uncoupling of eNOS has also been described in diabetic subjects. Tetrahydrobiopterin (BH_4), a cofactor for endothelial nitric oxide synthase, becomes oxidized in the hyperglycemic milieu. As a result, instead of producing NO to react with O_2^- and yield $ONOO^-$, the uncoupled eNOS is capable of producing more O_2^-. Lowering the availability of BH4 in diabetics has been shown to be causative of inducing eNOS uncoupling. This reaction further results in increased production of O_2^- but a reduced availability of nitric oxide.

7 Genetic Approaches to New Genes – Candidate Genes and Susceptibility Genes

It has been well accepted that type 2 diabetes mellitus is a complex metabolic disease with strong genetic influences. Although monogenic causes of insulin resistance and diabetes such as MODY (maturity-onset diabetes of the young) have been identified, the overwhelming majority of the cases are polygenic in nature. Identification of such genes and their functional roles in the development of insulin resistance and diabetes may provide important insights for the identification of novel targets for therapeutic interventions. The candidate gene approach has been informative in identifying a number of monogenic causes of insulin resistance, albeit only accounting for a very small proportion of all type 2 diabetes. In addition,

this approach also provided insights into the role of common genetic variants of a few additional genes.

7.1 Monogenic Syndromes of Insulin Resistance and Diabetes Mellitus

7.1.1 Insulin Receptor Mutations

Over 50 different mutations in the insulin receptor gene have been described. In addition to severe insulin resistance, these affected subjects are often found to have other physical abnormalities including abnormal facies, intrauterine growth retardation, acanthosis nigricans, and so on. The phenotype of these mutations is typically transmitted in an autosomal recessive fashion, but the obligate heterozygous parents and some of the other heterozygous relatives often develop milder forms of insulin resistance comparable to those seen in type 2 diabetic subjects. Functional characterizations of these naturally occurring mutations have provided important insights into the structure–function relationships in various domains of the receptor.

7.1.2 Insulin Gene Mutations

Rare mutations in the insulin gene (*INS*) can cause hyperinsulinemia and insulin resistance. Mechanisms of impaired signaling include defective binding to the insulin receptor and changes in proinsulin leading to inefficient cleavage of proinsulin. On the other hand, a number of *INS* mutations and polymorphisms have been detected with no associated susceptibility to diabetes.

7.1.3 Familial Partial Lipodystrophy (FPLD)

Familial partial lipodystrophy (FPLD) of the Dunnigan type (also called *lipoatrophy*), is a rare form of insulin resistance associated with a characteristic form of regional loss of subcutaneous fat tissue after puberty. Other clinical phenotypes include insulin resistance, hypertriglyceridemia, glucose intolerance/diabetes, and severe hypoleptinemia. The causative defective gene was recently identified as the *LMNA*, lamin A/C gene, and the syndrome is transmitted in an autosomal dominant fashion. Several mutations of the *LMNA* gene have been identified to cause FPLD. The *LMNA* gene codes for two splice variants lamin A and C that heterodimerize, and the gene products are nuclear envelope proteins. The mechanism by which *LMNA* gene mutations cause FPLD remains unknown. In light of the remarkable similarities in the metabolic abnormalities when compared to the adipose-specific *SREBP1* transgenic mice, a murine model of generalized lipoatrophy, it has been speculated that a physical interaction between SREBP and the nuclear envelope might be essential for insulin signaling. It is of interest to note that mutations in other regions of the same *LMNA* gene can cause Emery–Dreifuss muscular dystrophy and a cardiomyopathy. In addition to mutations in *LMNA* gene, two mutations in the *PPARγ* gene have also been reported to be associated with the FPLD syndrome. Glucose transporters GLUT1, GLUT2, and GLUT4 are encoded by the *SLC2A1, SLC2A2*, and *SLC2A4* genes respectively. The *SCL2A1* gene is expressed ubiquitously in a constitutive manner but is particularly abundant in the brain. Rare mutations of the *SLC2A1* gene cause seizure disorders due to ineffective transport of glucose across the blood–brain barrier. The *SCL2A2* gene is expressed primarily in the liver and pancreas. Rare mutations of this gene have been described with the phenotype to include impaired

glucose uptake and glycogen accumulation in the liver and kidneys. *SCL2A4* encodes for GLUT4 in myocytes and adipocytes and plays a crucial role in glucose uptake by these tissues for utilization. Surprisingly, there are no single gene mutations or common polymorphisms that have been consistently demonstrated to influence diabetes susceptibility.

7.1.4 Insulin Receptor Substrates

Being located at the proximal arm of the insulin-signaling cascade, insulin receptor substrates genes (*IRS*) are natural candidates for insulin resistance and diabetes. A rare mutation in the *IRS1* gene has been detected in a subject with type 2 diabetes. On the other hand, common polymorphisms have been reported to be associated with type 2 diabetes in a number of population studies.

7.1.5 Maturity Onset Diabetes of the Young (MODY)

This is a relatively rare group of monogenic causes of early-onset type 2 diabetes (~2–5% of all type 2 diabetes cases) transmitted in an autosomal dominant manner. The affected are typically not obese and they develop diabetes as early as in childhood, but most by the third decade. All but one type of MODY, namely, MODY2, are caused by mutation of specific transcription factors and the typical primary metabolic defects are characterized by glucose-stimulated insulin secretion rather than insulin resistance. MODY2 results from mutations in the glucokinase (*GCK*) gene, encoding for glucose kinase in the glycolytic pathway. In pancreatic β-cells, a defective glycolytic pathway results in an impairment of signaling for glucose-induced insulin secretion. Genes identified to be responsible for the other MODY syndromes through either positional cloning or candidate gene approaches include a variety of transcriptional factors. Once again, defective β-cell insulin secretion is the common primary defect. Genes mutated to cause the other MODY types include hepatic nuclear factor 4α (*HNF4*α) gene in MODY1, hepatic nuclear factor 1α (*HNF*1α) gene for MODY3, insulin promoting factor 1 (*IPF1*) for MODY4, and transcription factor 2 gene (*TCF2*) for MODY5. The heterogeneity in the genes responsible for different MODY syndromes is reflected by the heterogeneity in clinical phenotypes.

Mutation in the *HNF*1α gene can cause MODY3 and represents the most commonly described MODY subtype. Hyperglycemia in MODY3 is typically severe and is frequently associated with microvascular complications. It is of interest to note that most mutations leading to MODY3 are dominant negative because mice heterozygous for the null mutation in the murine *HNF*1α gene have a normal phenotype. Data available to date suggest that only mutations located in the transactivation domain of HNF1α have a dominant negative effect.

MODY1 is caused by mutations in the transcription factor *HNF*4α, but are much less prevalent than MODY3. The mutations result in defective binding to the target DNA sequences. HNF4α is a member of the steroid/thyroid hormone receptor superfamily and an upstream regulator of HNF1α expression. Its activity is in turn modulated by a number of metabolic factors including long-chain fatty acids. The binding may result in either the activation or inhibition of HNF4α transcriptional activity as a function of chain length degree of saturation, and therefore, makes this transcription factor a

potential therapeutic target for dietary fatty acid manipulation.

A deletion mutation in the transcription factor insulin promoter factor-1 (*IPF1*) gene is found to cosegregate with a large MODY kindred. Subjects heterozygous for the mutation develop diabetes of variable severity and one homozygote for the mutated gene developed pancreatic agenesis at birth and suffered both endocrine and exocrine pancreatic deficiency. The importance of *IPF-1* in pancreatic development was confirmed by the pancreatic agenesis phenotype seen in the embryo of the *ipf-1* knockout mice.

Transcription factor 2 (TCF2)/HNF1β mutations have been described in subjects with early-onset diabetes consistent with MODY. In addition, these subjects also developed severe nondiabetic kidney disease and Mullerian aplasia.

7.2 Common Genetic Variants of Other Candidate Genes

Common variants in a number of functional candidate genes including β2- and β3-adrenergic receptor genes, PPARγ, uncoupling protein 1 gene, and so on, have been studied extensively. Two such genes, β3-adrenergic receptor (*ADRB3*) and *PPARγ*, will be briefly reviewed here.

7.2.1 ADRB3

ADRB3 is a G-protein coupled receptor expressed largely in visceral adipose tissue. Activation by β-agonists results in cellular accumulation of cAMP that will enhance lipolysis and thermogenesis. A Trp64Arg in the *ADRB3* gene polymorphism was first discovered in Pima Indians, and homozygotes have an increased tendency to develop type 2 diabetes. A follow-up study in a Finn cohort found a positive association between this polymorphism and abdominal obesity and insulin resistance, generating further interest in exploring the potential clinical utility of this genetic variant. However, an increasing number of studies reporting negative associations dampened the pursuit of this candidate gene.

7.2.2 PPARγ

PPARγ gene encodes for two splice variants, PPARγ1 and PPARγ2. The former is expressed in a variety of tissues including muscle and adipose, but PPARγ2 is expressed virtually exclusively in adipose tissue. To date, two rare forms of dominant negative PPARγ mutations have been implicated as monogenic causes of insulin resistance and diabetes but they appear to cause lipodystrophy. Of the numerous common variants described for PARγ2, the Pro12Ala PPARγ polymorphism has received a great deal of attention. While the association between this polymorphism with diabetes and obesity has been variable, the association between this polymorphism and insulin resistance has been more consistent. Several populations of different ethnicity showed enhanced insulin sensitivity amongst those carrying the Ala12 allele.

7.3 Genome Scanning and Positional Cloning Strategies

The genome-wide linkage approach permits a systematic search throughout the genome for susceptibility genes in complex diseases such as diabetes without any prior knowledge of the functions of the genes. This technique complements the candidate gene approach because it is increasingly clear that type 2 diabetes susceptibility genes are likely to encode for

proteins of no obvious functional links to current established pathways.

Many genome scans have been reported and the results summarized in a number of recent reviews.

The *NIDDM1* locus on chromosome 2q was identified in Mexican Americans. Mahtani et al. reported a linkage at a locus near *MODY3* on chromosome 12 in Finnish families with the major phenotype being an insulin secretion defect. The identification of the calpain 10 (*CAPN10*) gene, which encodes the intracellular nonlysosomal calcium-activated cysteine protease calpain-10 as the gene involved on the *NIDDM1* locus, is a successful example of the positional cloning approach. The identification of calpain-10 holds promise in discovering novel biochemical pathways in glucose homeostasis and insulin resistance.

By using both positional cloning and a candidate gene approach, Hegele et al. have identified, in the Oj-Cree of Sandy Lake in Northern Ontario, a novel mutation in the *HNF1A* gene (*HNF1A Gly319Ser*), which is strongly linked to early-onset type 2 diabetes. The homozygotes and heterozygotes for this mutation are expected to have odd ratios of 4 and 2 in the development of type 2 diabetes, respectively. It is intriguing to observe that despite the same gene as in MODY3 being mutated, the clinical features of subjects with this *HNF1A Gly319Ser* mutation bear little resemblance to the MODY3 phenotypes.

8
Summary and Conclusions

Type 2 diabetes mellitus is an emerging epidemic worldwide, with the incidence rising in parallel with a continuing rise in the prevalence of obesity. With only a handful of exceptions, type 2 diabetes is a polygenic disorder with the clinical phenotype of insulin resistance and frank diabetes brought on by environmental factors. Treatment modalities have originally been focusing on the normalization of hyperglycemia by lifestyle modification and antidiabetic pharmacological agents, singly or in combination. More recently, rigorous lifestyle measures as well as a number of glucose lowering agents have been found to be effective in the prevention of type 2 diabetes. However, much remains to be done in our understanding of the known players in the pathogenesis of insulin resistance, β-cell dysfunction, and diabetes at the cellular and molecular levels. To significantly advance our understanding of this disorder, a multidisciplinary approach is essential.

In the area of insulin signaling, new players continue to be discovered and our understanding of the critical steps in signaling and the interactions between them and external environmental factors will prove to be informative for new therapeutic targets. The field of cross talk between specific signaling pathways and their interactions with cell membrane microdomains represent an exciting area of future research. A systematic dissection of the signaling pathways by the generation of knockout and transgenic mice will also offer novel insights into the functional roles of these individual players at the whole mammalian organism level. The genome projects will soon make available an enormous amount of DNA sequence data from various species, including humans, and will facilitate discovery of novel genes involved in insulin resistance and diabetes. Our evolving view of the adipose tissue as not just a passive fat depot

but rather as being a dynamic and active secretory organ, together with a better understanding of biological functions and how they interact with external modulating factors, will be critical for us in developing effective strategies to combat increasing adiposity and their metabolic and functional sequelae. The biological actions of adiponectin and its interactions with other adipocyte-derived hormones and cytokines appear to be an area critical to our understanding of how obesity may contribute to insulin resistance. In addition, partitioning of energy as mediated by distribution and redistribution of body fat may hold the key for our future understanding of links between obesity, insulin resistance, and diabetes and for better design of therapeutics. The discovery of glucagon-like peptides (GLPs) and their respective receptors holds great promise to effectively control hyperglycemia and to reestablish the β-cell proliferation/apoptosis balance with the ultimate goal of preserving β-cell mass.

Uncontrolled hyperglycemia continues to be the main threat for the development of microvascular disease and our currently available therapeutics frequently fail to completely normalize glycemia or to maintain control. Tracking the pathophysiology of high glucose–induced long-term diabetic microvascular complications to the function at subcellular organelle function will facilitate better preventive measures. Likewise, the role of hyperglycemia in the development of macrovascular complications continues to stimulate great interest. Current understanding of the role of elevated blood glucose, especially postprandial hyperglycemia on arterial wall function, including endothelial function, oxidative stress, and the interaction with lipoproteins, may lead to the identification of novel targets for atherogenesis in diabetes.

Acknowledgment

The author wishes to thank Dr. David Jenkins for his contribution and his critical review of the manuscript. D. Ng is a recipient of a Canadian Institute for Health Research New Investigator Career Scholarship.

See also Adipocytes; Diabetes Insipidus, Molecular Biology of; Immunology; Innate Immunity.

Bibliography

Books and Reviews

Adeli, K., Taghibiglou, C., Van Iderstine, S.C., Lewis, G.F. (2002) Mechanisms of hepatic very low-density lipoprotein overproduction in insulin resistance, *Trends Cardiovasc. Med.* **11**, 170–176.

Brownlee, M. (2002) Biochemistry and molecular cell biology of diabetic complications, *Nature* **414**, 813–820.

Busch, C.P., Hegele, R.A. (2001) Genetic determinants of type 2 diabetes mellitus, *Clin. Genet.* **60**, 243–254.

Chisolm, G.M., Steinberg, D. (2000) The oxidative modification hypothesis of atherogenesis: an overview, *Free Radical Biol. Med.* **28**, 1815–1826.

Friedman, J., Halaas, J.L. (1998) Leptin and the regulation of body weight in mammals, *Nature* **395**, 763–770.

Horton, J.D., Goldstein, J.L., Brown, M.S. (2002) SERBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver, *J. Clin. Invest.* **109**, 1125–1131.

Kahn, B.B., Flier, J.S. (2000) Obesity and insulin resistance, *J. Clin. Invest.* **106**, 473–481.

Kahn, A.H., Pessin, J.E. (2002) Insulin regulation of glucose uptake: a complex interplay of intracellular signalling pathways, *Diabetologia* **45**, 1475–1483.

Mauvais-Jarvis, F., Kulkarni, R.N., Kahn, C.R. (2002) Knockout models are useful tools to

dissect the pathophysiology and genetics of insulin resistance, *Clin. Endocrinol.* **57**, 1–9.

Sheetz, M.J., King, G.L. (2002) Molecular understanding of hyperglycemia's adverse effects for diabetic complications, *JAMA* **288**, 2579–2588.

Sorescu, D., Szocs, K., Griendling, K.K. (2001) NAD(P)H oxidases and their relevance ot atherosclerosis, *Trends Cardiovasc. Med.* **11**, 124–131.

Unger, R.H., Orci, L. (2001) Diseases of liporegulation: new perspective on obesity and related disorders, *FASEB J.* **15**, 312–321.

Primary Literature

Accili, D., Drago, J., Lee, E.J., Johnson, M.D., Cool, M.H., Salvatore, P., Asico, L.D., Jose, P.A., Taylor, S.I., Westphal, H. (1996) Early neonatal death in mice homozygous for a null allele of the insulin receptor gene, *Nat. Genet.* **12**, 106–109.

Agarwal, A.K., Garg, A. (2002) A novel heterozygous mutation in peroxisome proliferator-activated receptor-gamma gene in a patient with familial partial lipodystrophy, *J. Clin. Endocrinol. Metab.* **87**, 408–411.

Aguilar-Bryan, L., Bryan, J. (1999) Molecular biology of adenosine triphosphate-sensitive potassium channels, *Endocr. Rev.* **20**, 101–135.

Ahima, R.S., Dushay, J., Flier, S.N., Prabakaran, D., Flier, J.S. (1997) Leptin accelerates the onset of puberty in normal female mice, *J. Clin. Invest.* **99**, 391–395.

Amer, P. (1995) Differences in lipolysis between human subcutaneous and omental adipose tissues, *Ann. Med.* **27**, 435–438.

Andreasson, K., Galuska, D., Thorne, A., Sonnenfeld, T., Wallberg-Henriksson, H. (1991) Decreased insulin-stimulated 3-0-methylglucose transport in in vitro incubated muscle strips from type II diabetic subjects, *Acta Physiol. Scand.* **142**, 255–260.

Azevedo, J.L. Jr., Carey, J.O., Pories, W.J., Morris, P.G., Dohm, G.L. (1995) Hypoxia stimulates glucose transport in insulin-resistant human skeletal muscle, *Diabetes* **44**, 695–698.

Babior, B.M. (1999) NADPH oxidase: an update, *Blood* **93**, 1464–1498.

Balon, T.W., Nadler, J.L. (1996) Nitric oxide mediates skeletal glucose transport, *Am. J. Physiol.* **270**(6 Pt 1), E1058–E1059.

Balon, T.W., Nadler, J.L. (1997) Evidence that nitric oxide increases glucose transport in skeletal muscle, *J. Appl. Physiol.* **82**, 359–363.

Barroso, I., Gurnell, M., Crowley, V.E., Agostini, M., Schwabe, J.W., Soos, M.A., Maslen, G.L., Williams, T.D., Lewis, H., Schafer, A.J., Chatterjee, V.K., O'Rahilly, S. (1999) Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension, *Nature* **402**(6764), 880–883.

Berg, A.H., Combs, T.P., Du, X., Brownlee, M., Scherer, P.E. (2001) The adipocyte-secreted protein Acrp30 enhances hepatic insulin action, *Nat. Med.* **7**, 947–953.

Bjorntorp, P. (1990) "Portal" adipose tissue as a generator of risk factors for cardiovascular disease and diabetes, *Arteriosclerosis* **10**, 493–496.

Bluher, M., Kahn, B.B., Kahn, C.R. (2003) Extended longevity in mice lacking the insulin receptor in adipose tissue, *Science* **299**(5606), 572–574.

Bluher, M., Michael, M.D., Peroni, O.D., Ueki, K., Carter, N., Kahn, B.B., Kahn, C.R. (2002) Adipose tissue selective insulin receptor knockout protects against obesity and obesity-related glucose intolerance, *Dev. Cell* **3**, 25–38.

Boden, G., Cheung, P., Stein, T.P., Kresge, K., Mozzoli, M. (2002) FFA cause hepatic insulin resistance by inhibiting insulin suppression of glycogenolysis, *Am. J. Physiol. Endocrinol. Metab.* **283**, E12–E19.

Bonadonna, R.C., Del Prato, S., Bonora, E., Saccomani, M.P., Gulli, G., Natali, A., Frascerra, S., Pecori, N., Ferrannini, E., Bier, D., Cobelli, C., DeFronzo, R.A. (1996) Roles of glucose transport and glucose phosphorylation in muscle insulin resistance of NIDDM, *Diabetes* **45**, 915–925.

Brozinick, J.T. Jr., Etgen, G.J. Jr., Yaspelkis, B.B. III, Ivy, J.L. (1992) Contraction-activated glucose uptake is normal in insulin-resistant muscle of the obese Zucker rat, *J. Appl. Physiol.* **73**, 382–387.

Brozinick, J.T. Jr., Etgen, G.J. Jr., Yaspelkis, B.B. III, Kang, H.Y., Ivy, J.L. (1993) Effects of exercise training on muscle GLUT-4 protein content and translocation in obese Zucker rats, *Am. J. Physiol.* **265**(3 Pt 1), E419–E427.

Brun, R.P., Tontonoz, P., Forman, B.M., Ellis, R., Chen, J., Evans, R.M., Spiegelman, B.M. (1996) Differential activation of adipogenesis by multiple PPAR isoforms, *Genes Dev.* **10**, 974–984.

Bruning, J.C., Michael, M.D., Winnay, J.N., Hayashi, T., Horsch, D., Accili, D., Goodyear, L.J., Kahn, C.R. (1998) A muscle-specific insulin receptor knockout exhibits features of the metabolic syndrome of NIDDM without altering glucose tolerance, *Mol. Cell* **2**, 559–569.

Bruning, J.C., Winnay, J., Bonner-Weir, S., Taylor, S.I., Accili, D., Kahn, C.R. (1997) Development of a novel polygenic model of NIDDM in mice heterozygous for IR and IRS-1 null alleles, *Cell* **88**, 561–572.

Butler, A.E., Janson, J., Bonner-Weir, S., Ritzel, R., Rizza, R.A., Butler, P.C. (2003) Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes, *Diabetes* **52**, 102–110.

Calles-Escandon, J., Garcia-Rubi, E., Mirza, S., Mortensen, A. (1999) Type 2 diabetes: one disease, multiple cardiovascular risk factors, *Coron. Artery Dis.* **10**, 23–30.

Carayannopoulos, M.O., Chi, M.M., Cui, Y., Pingsterhaus, J.M., McKnight, R.A., Mueckler, M., Devaskar, S.U., Moley, K.H. (2000) GLUT8 is a glucose transporter responsible for insulin-stimulated glucose uptake in the blastocyst, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7313–7318.

Carr, M.E. (2001) Diabetes mellitus: A hypercoagulable state, *J. Diabetes Complications* **15**, 44–54.

Carpentier, A., Mittelman, S.D., Lamarche, B., Bergman, R.N., Giacca, A., Lewis, G.F. (1999) Acute enhancement of insulin secretion by FFA in humans is lost with prolonged FFA elevation, *Am. J. Physiol. Endocrinol. Metab.* **276**, E1055–E1066.

Casanueva, F.F., Dieguez, C. (1999) Neuroendocrine regulation and actions of leptin, *Front. Neuroendocrinol.* **20**, 317–363.

Cerielo, A., Motz, E., Cavarape, A., Lizzio, S., Russo, A., Quatraro, A., Giugliano, D. (1997) Hyperglycemia counterbalances the antihypertensive effect of glutathione in diabetic patients: evidence linking hypertension and glycemia through the oxidative stress in diabetes mellitus, *J. Diabetes Complications* **11**, 250–255.

Chehab, F.F., Qiu, J., Mounzih, K., Ewart-Toland, A., Ogus, S. (2002) Leptin and reproduction, *Nutr. Rev.* **60**(10 Pt 2), S39–S46.

Cheng, A., Uetani, N., Simoncic, P.D., Chaubey, V.P., Lee-Loy, A., McGlade, C.J., Kennedy, B.P., Tremblay, M.L. (2002) Attenuation of leptin action and regulation of obesity by protein tyrosine phosphatase 1B, *Dev. Cell* **2**, 497–503.

Chiang, S.H., Baumann, C.A., Kanzaki, M., Thurmond, D.C., Watson, R.T., Neudauer, C.L., Macara, I.G., Pessin, J.E., Saltiel, A.R. (2001) Insulin-stimulated GLUT4 translocation requires the CAP-dependent activation of TC10, *Nature* **410**(6831), 944–948.

Chiasson, J.L., Josse, R.G., Gomis, R., Hanefeld, M., Karasik, A., Laakso, M. (2002) Acarbose for prevention of type 2 diabetes mellitus: the STOP-NIDDM randomised trial, *Lancet* **359**, 2072–2077.

Cline, G.W., Petersen, K.F., Krssak, M., Shen, J., Hundal, R.S., Trajanoski, Z., Inzucchi, S., Dresner, A., Rothman, D.L., Shulman, G.I. (1999) Impaired glucose transport as a cause of decreased insulin-stimulated muscle glycogen synthesis in type 2 diabetes, *N. Engl. J. Med.* **341**, 240–246.

Cnop, C.M., Hannaert, J.C., Hoorens, A., Eizirik, D.L., Pipeleers, D.G. (2001) Inverse relationship between cytotoxicity of free fatty acids in pancreatic islet cells and cellular triglyceride accumulation, *Diabetes* **50**, 1771–1777.

Colombo, C., Cutson, J.J., Yamauchi, T., Vinson, C., Kadowaki, T., Gavrilova, O., Reitman, M.L. (2002) Transplantation of adipose tissue lacking leptin is unable to reverse the metabolic abnormalities associated with lipoatrophy, *Diabetes* **51**, 2727–2733.

Craven, P.A., Studer, R.K., DeRubertis, F.R. (1994) Impaired nitric oxide-dependent cyclic guanosine monophosphate generation in glomeruli from diabetic rats. Evidence for protein kinase C-mediated suppression of the cholinergic response, *J. Clin. Invest.* **93**, 311–320.

Czech, M.P., Corvera, S. (1999) Signaling mechanisms that regulate glucose transport, *J. Biol. Chem.* **274**, 1865–1868.

De Keulenaer, G.W., Chappell, D.C., Ishizaka, N., Nerem, R.M., Alexander, R.W., Griendling, K.K. (1998) Oscillatory and steady laminar shear stress differentially affect human endothelial redox state: role of a superoxide-producing NADH oxidase, *Circ. Res.* **82**, 1094–1101.

de Souza, C.J., Eckhardt, M., Gagen, K., Dong, M., Chen, W., Laurent, D., Burkey, B.F. (2001) Effects of pioglitazone on adipose tissue

remodeling within the setting of obesity and insulin resistance, *Diabetes* **50**, 1863–1871.

de Souza, C.J., Hirshman, M.F., Horton, E.S. (1997) CL-316,243, a beta3-specific adrenoceptor agonist, enhances insulin-stimulated glucose disposal in nonobese rats, *Diabetes* **46**, 1257–1263.

Deeb, S.S., Fajas, L., Nemoto, M., Pihlajamaki, J., Mykkanen, L., Kuusisto, J., Laakso, M., Fujimoto, W., Auwerx, J. (1998) A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity, *Nat. Genet.* **20**, 284–287.

DeFronzo, R.A., Gunnarsson, R., Bjorkman, O., Olsson, M., Wahren, J. (1985) Effects of insulin on peripheral and splanchnic glucose metabolism in noninsulin-dependent (type II) diabetes mellitus, *J. Clin. Invest.* **76**, 149–155.

DeFronzo, R.A., Jacot, E., Jequier, E., Maeder, E., Wahren, J., Felber, J.P. (1981) The effect of insulin on the disposal of intravenous glucose. Results from indirect calorimetry and hepatic and femoral venous catheterization, *Diabetes* **30**, 1000–1007.

Dela, F., Larsen, J.J., Mikines, K.J., Ploug, T., Petersen, L.N., Galbo, H. (1995) Insulin-stimulated muscle glucose clearance in patients with NIDDM. Effects of one-legged physical training, *Diabetes* **44**, 1010–1020.

Derave, W., Lund, S., Holman, G.D., Wojtaszewski, J., Pedersen, O., Richter, E.A. (1999) Contraction-stimulated muscle glucose transport and GLUT-4 surface content are dependent on glycogen content, *Am. J. Physiol.* **277**(6 Pt 1), E1103–E1110.

Despres, J.P. (2001) Health consequences of visceral obesity, *Ann. Med.* **33**, 534–541.

Diabetes Control and Complications Trial (DCCT), The Diabetes Control and Complications Trial Research Group. (1993) The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus., *N. Engl. J. Med.* **329**, 977–986.

Dobbins, R.L., Chester, M.W., Daniels, M.B., McGarry, J.D., Stein, D.T. (1998) Circulating fatty acids are essential for efficient glucose-stimulated insulin secretion after prolonged fasting in humans, *Diabetes* **47**(10), 1613–1618.

Doege, H., Bocianski, A., Scheepers, A., Axer, H., Eckel, J., Joost, H.G., Schurmann, A. (2001) Characterization of human glucose transporter (GLUT) 11 (encoded by SLC2A11), a novel sugar-transport facilitator specifically expressed in heart and skeletal muscle, *Biochem. J.* **359**(Pt 2), 443–449.

Dolan, P.L., Tapscott, E.B., Dorton, P.J., Dohm, G.L. (1993) Contractile activity restores insulin responsiveness in skeletal muscle of obese Zucker rats, *Biochem. J.* **289**(Pt 2), 423–426.

Dohm, G.L., Tapscott, E.B., Pories, W.J., Dabbs, D.J., Flickinger, E.G., Meelheim, D., Fushiki, T., Atkinson, S.M., Elton, C.W., Caro, J.F. (1988) An in vitro human muscle preparation suitable for metabolic studies. Decreased insulin stimulation of glucose transport in muscle from morbidly obese and diabetic subjects, *J. Clin. Invest.* **82**, 486–494.

Drucker, D.J. (2003) Glucagon-like peptides: regulators of cell proliferation, differentiation, and apoptosis, *Mol. Endocrinol.* **17**, 161–171.

Du, X.L., Edelstein, D., Dimmeler, S., Ju, Q., Sui, C., Brownlee, M. (2001) Hyperglycemia inhibits endothelial nitric oxide synthase activity by posttranslational modification at the Akt site, *J. Clin. Invest.* **108**, 1341–1348.

Du, X.L., Edelstein, D., Rossetti, L., Fantus, I.G., Goldberg, H., Ziyadeh, F., Wu, J., Brownlee, M. (2000) Hyperglycemia-induced mitochondrial superoxide overproduction activates the hexosamine pathway and induces plasminogen activator inhibitor-1 expression by increasing Sp1 glycosylation, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12222–12226.

Duerrschmidt, N., Wippich, N., Goettsch, W., Broemme, H.J., Morawietz, H. (2000) Endothelin-1 induces NAD(P)H oxidase in human endothelial cells, *Biochem. Biophys. Res. Commun.* **269**, 713–717.

Eitel, K., Staiger, H., Brendel, M.D., Brandhorst, D., Bretzel, R.G., Haring, H.U., Kellerer, M. (2002) Different role of saturated and unsaturated fatty acids in beta-cell apoptosis, *Biochem. Biophys. Res. Commun.* **299**, 853–856.

Eriksson, J., Koranyi, L., Bourey, R., Schalin-Jantti, C., Widen, E., Mueckler, M., Permutt, A.M., Groop, L.C. (1992) Insulin resistance in type 2 (non-insulin-dependent) diabetic patients and their relatives is not associated with a defect in the expression of the insulin-responsive glucose transporter (GLUT-4) gene in human skeletal muscle, *Diabetologia* **35**, 143–147.

Etgen, G.J. Jr., Jensen, J., Wilson, C.M., Hunt, D.G., Cushman, S.W., Ivy, J.L. (1997) Exercise training reverses insulin resistance in muscle by enhanced recruitment of GLUT-4 to the cell surface, *Am. J. Physiol.* **272**(5 Pt 1), E864–E869.

Farilla, L., Hui, H., Bertolotto, C., Kang, E., Bulotta, A., Di, Mario, U., Perfetti, R. (2002) Glucagon-like peptide-1 promotes islet cell growth and inhibits apoptosis in Zucker diabetic rats, *Endocrinology* **43**, 4397–4408.

Fisher, J.S., Gao, J., Han, D.H., Holloszy, J.O., Nolte, L.A. (2002) Activation of AMP kinase enhances sensitivity of muscle glucose transport to insulin, *Am. J. Physiol. Endocrinol. Metab.* **282**, E18–E23.

Foretz, M., Guichard, C., Ferre, P., Foufelle, F. (1999) Sterol regulatory element binding protein-1c is a major mediator of insulin action on the hepatic expression of glucokinase and lipogenesis-related genes, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12737–12742.

Froguel, P., Velho, G. (2001) Genetic determinants of type 2 diabetes, *Recent Prog. Horm. Res.* **56**, 91–105.

Fruebis, J., Tsao, T.S., Javorschi, S., Ebbets-Reed, D., Erickson, M.R., Yen, F.T., Bihain, B.E., Lodish, H.F. (2001) Proteolytic cleavage product of 30-kDa adipocyte complement-related protein increases fatty acid oxidation in muscle and causes weight loss in mice, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2005–2010.

Furuta, H., Furuta, M., Sanke, T., Ekawa, K., Hanabusa, T., Nishi, M., Sasaki, H., Nanjo, K. (2002) Nonsense and missense mutations in the human hepatocyte nuclear factor-1 beta gene (TCF2) and their relation to type 2 diabetes in Japanese, *J. Clin. Endocrinol. Metab.* **87**, 3859–3863.

Gabriely, I., Yang, X.M., Cases, J.A., Ma, X.H., Rossetti, L., Barzilai, N. (2002) Hyperglycemia induces PAI-1 gene expression in adipose tissue by activation of the hexosamine biosynthetic pathway, *Atherosclerosis* **160**, 115–122.

Goldberg, H.J., Scholey, J., Fantus, I.G. (2000) Glucosamine activates the plasminogen activator inhibitor 1 gene promoter through Sp1 DNA binding sites in glomerular mesangial cells, *Diabetes* **49**, 863–871.

Gagnon, J., Mauriege, P., Roy, S., Sjostrom, D., Chagnon, Y.C., Dionne, F.T., Oppert, J.M., Perusse, L., Sjostrom, L., Bouchard, C. (1996) The Trp64Arg mutation of the beta3 adrenergic receptor gene has no effect on obesity phenotypes in the Quebec Family Study and Swedish Obese Subjects cohorts, *J. Clin. Invest.* **98**, 2086–2093.

Gautier, J.F., Mourier, A., de Kerviler, E., Tarentola, A., Bigard, A.X., Villette, J.M., Guezennec, C.Y., Cathelineau, G. (1998) Evaluation of abdominal fat distribution in noninsulin-dependent diabetes mellitus: relationship to insulin resistance, *J. Clin. Endocrinol. Metab.* **83**, 1306–1311.

Gavrilova, O., Marcus-Samuels, B., Graham, D., Kim, J.K., Shulman, G.I., Castle, A.L., Vinson, C., Eckhaus, M., Reitman, M.L. (2000) Surgical implantation of adipose tissue reverses diabetes in lipoatrophic mice, *J. Clin. Invest.* **105**, 271–278.

Ghosh, S., Langefeld, C.D., Ally, D., Watanabe, R.M., Hauser, E.R., Magnuson, V.L., Nylund, S.J., Valle, T., Eriksson, J., Bergman, R.N., Tuomilehto, J., Collins, F.S., Boehnke, M. (1999) The W64R variant of the beta3-adrenergic receptor is not associated with type II diabetes or obesity in a large Finnish sample, *Diabetologia* **42**, 238–244.

Ghilardi, N., Ziegler, S., Wiestner, A., Stoffel, R., Heim, M.H., Skoda, R.C. (1996) Defective STAT signaling by the leptin receptor in diabetic mice, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6231–6235.

Glogowski, E.A., Tsiani, E., Zhou, X., Fantus, I.G., Whiteside, C. (1999) High glucose alters the response of mesangial cell protein kinase C isoforms to endothelin-1, *Kidney Int.* **55**, 486–499.

Gnudi, L., Shepherd, P.R., Kahn, B.B. (1996) Over-expression of GLUT4 selectively in adipose tissue in transgenic mice: implications for nutrient partitioning, *Proc. Nutr. Soc.* **55**(1B), 191–199.

Goodyear, L.J., Giorgino, F., Sherman, L.A., Carey, J., Smith, R.J., Dohm, G.L. (1995) Insulin receptor phosphorylation, insulin receptor substrate-1 phosphorylation, and phosphatidylinositol 3-kinase activity are decreased in intact skeletal muscle strips from obese subjects, *J. Clin. Invest.* **95**, 2195–2204.

Gorlach, A., Brandes, R.P., Nguyen, K., Amidi, M., Dehghani, F., Busse, R. (2000) A gp91phox containing NADPH oxidase selectively expressed in endothelial cells is a major source of oxygen radical generation in the arterial wall, *Circ. Res.* **87**, 26–32.

Gould, G.W., Holman, G.D. (1993) The glucose transporter family: structure, function and

tissue-specific expression, *Biochem. J.* **295**, 329–341.

Griendling, K.K., Sorescu, D., Ushio-Fukai, M. (2000a) NAD(P)H oxidase: role in cardiovascular biology and disease, *Circ. Res.* **86**, 494–501.

Griendling, K.K., Sorescu, D., Lassegue, B., Ushio-Fukai, M. (2000b) Modulation of protein kinase activity and gene expression by reactive oxygen species and their role in vascular physiology and pathophysiology, *Arterioscler. Thromb. Vasc. Biol.* **20**, 2175–2183.

Gross, E.R., LaDisa, J.F., Weihrauch, D., Olson, L.E., Kress, T.T., Hettrick, D.A., Pagel, P.S., Warltier, D.C., Kersten, J.R. (2003) Reactive oxygen species modulate coronary wall shear stress and endothelial function during hyperglycemia, *Am. J. Physiol. Heart Circ. Physiol.* **284**, 1552–1559 Jan 23 [epub ahead of print].

Guilherme, A., Czech, M.P. (1998) Stimulation of IRS-1-associated phosphatidylinositol 3-kinase and Akt/protein kinase B but not glucose transport by beta1-integrin signaling in rat adipocytes, *J. Biol. Chem.* **273**, 33119–33122.

Hanis, C.L., Boerwinkle, E., Chakraborty, R., Ellsworth, D.L., Concannon, P., Stirling, B., Morrison, V.A., Wapelhorst, B., Spielman, R.S., Gogolin-Ewens, K.J., Shepard, J.M., et al. (1996) A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2, *Nat. Genet.* **13**, 161–166.

Hara, K., Okada, T., Tobe, K., Yasuda, K., Mori, Y., Kadowaki, H., Hagura, R., Akanuma, Y., Kimura, S., Ito, C., Kadowaki, T. (2000) The Pro12Ala polymorphism in PPAR gamma2 may confer resistance to type 2 diabetes, *Biochem. Biophys. Res. Commun.* **271**, 212–216.

Heart Protection Study Collaborative Group. (2002) MRC/BHF heart protection study of antioxidant vitamin supplementation in 20,536 high-risk individuals: a randomised placebo-controlled trial, *Lancet* **360**, 23–33.

Hegele, R.A. (1995) Dyslipidaemia and coronary heart disease: nature vs nurture, *Br. J. Hosp. Med.* **54**, 143–146.

Hegele, R.A., Cao, H., Anderson, C.M., Hramiak, I.M. (2000) Heterogeneity of nuclear lamin A mutations in Dunnigan-type familial partial lipodystrophy, *J. Clin. Endocrinol. Metab.* **85**, 3431–3435.

Hegele, R.A., Cao, H., Frankowski, C., Mathews, S.T., Leff, T. (2002) PPARG F388L, a transactivation-deficient mutant, in familial partial lipodystrophy, *Diabetes* **51**, 3586–3590.

Hegele, R.A., Cao, H., Harris, S.B., Hanley, A.J., Zinman, B. (1999) The hepatic nuclear factor-1alpha G319S variant is associated with early-onset type 2 diabetes in Canadian Oji-Cree, *J. Clin. Endocrinol. Metab.* **84**, 1077–1082.

Henriksen, E.J. (2002) Effects of acute exercise and exercise training on insulin resistance, *J. Appl. Physiol.* **93**, 788–796.

Herman, W.H., Fajans, S.S., Ortiz, F.J., Smith, M.J., Sturis, J., Bell, G.I., Polonsky, K.S., Halter, J.B. (1994) Abnormal insulin secretion, not insulin resistance, is the genetic or primary defect of MODY in the RW pedigree, *Diabetes* **43**, 40–46.

Hertz, R., Magenheim, J., Berman, I., Bar-Tana, J. (1998) Fatty acyl-CoA thioesters are ligands of hepatic nuclear factor-4alpha, *Nature* **392**(6675), 512–516.

Holloszy, J.O., Kohrt, W.M., Hansen, P.A. (1998) The regulation of carbohydrate and fat metabolism during and after exercise, *Front. Biosci.* **3**, D1011–D1027.

Horikawa, Y., Iwasaki, N., Hara, M., Furuta, H., Hinokio, Y., Cockburn, B.N., Lindner, T., Yamagata, K., Ogata, M., Tomonaga, O., Kuroki, H., Kasahara, T., Iwamoto, Y., Bell, G.I. (1997) Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY, *Nat. Genet.* **17**, 384–385.

Horikawa, Y., Oda, N., Cox, N.J., Li, X., Orho-Melander, M., Hara, M., Hinokio, Y., Lindner, T.H., Mashima, H., Schwarz, P.E., del Bosque-Plata, L., Horikawa, Y., Oda, Y., Yoshiuchi, I., Colilla, S., Polonsky, K.S., Wei, S., Concannon, P., Iwasaki, N., Schulze, J., Baier, L.J., Bogardus, C., Groop, L., Boerwinkle, E., Hanis, C.L., Bell, G.I. (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus, *Nat. Genet.* **26**, 163–175.

Horton, J.D., Bashmakov, Y., Shimomura, I., Shimano, H. (1998) Regulation of sterol regulatory element binding proteins in livers of fasted and refed mice, *Proc. Natl. Acad.Sci. U.S.A.* **95**, 5987–5992.

Hu, E., Tontonoz, P., Spiegelman, B.M. (1995) Transdifferentiation of myoblasts by the adipogenic transcription factors PPARγ and C/EBPα, *Proc. Natl. Acad. Sci.* **92**, 9856–9860.

Hughes, V.A., Fiatarone, M.A., Fielding, R.A., Kahn, B.B., Ferrara, C.M., Shepherd, P., Fisher, E.C., Wolfe, R.R., Elahi, D., Evans, W.J. (1993) Exercise increases muscle GLUT-4 levels and insulin action in subjects with impaired glucose tolerance, *Am. J. Physiol.* **264**(6 Pt 1), E855–E862.

Ido, Y., Carling, D., Ruderman, N. (2002) Hyperglycemia-induced apoptosis in human umbilical vein endothelial cells: inhibition by the AMP-activated protein kinase activation, *Diabetes* **51**, 159–167.

Inoguchi, T., Li, P., Umeda, F., Yu, H.Y., Kakimoto, M., Imamura, M., Aoki, T., Etoh, T., Hashimoto, T., Naruse, M., Sano, H., Utsumi, H., Nawata, H. (2000) High glucose level and free fatty acid stimulate reactive oxygen species production through protein kinase C–dependent activation of NAD(P)H oxidase in cultured vascular cells, *Diabetes* **49**, 1939–1945.

Inoue, N., Ramasamy, S., Fukai, T., Nerem, R.M., Harrison, D.G. (1996) Shear stress modulates expression of Cu/Zn superoxide dismutase in human aortic endothelial cells, *Circ. Res.* **79**, 32–37.

Isakoff, S.J., Taha, C., Rose, E., Marcusohn, J., Klip, A. (1995) The inability of phosphatidylinositol 3-kinase activation to stimulate GLUT4 translocation indicates additional signaling pathways are required for insulin-stimulated glucose uptake, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10247–10251.

Ishii, H., Jirousek, M.R., Koya, D., Takagi, C., Xia, P., Clermont, A., Bursell, S.E., Kern, T.S., Ballas, L.M., Heath, W.F., Stramm, L.E., Feener, E.P., King, G.L. (1996) Amelioration of vascular dysfunctions in diabetic rats by an oral PKC beta inhibitor, *Science* **272**(5262), 728–731.

Isomaa, B., Henricsson, M., Lehto, M., Forsblom, C., Karanko, S., Sarelin, L., Haggblom, M., Groop, L. (1998) Chronic diabetic complications in patients with MODY3 diabetes, *Diabetologia* **41**, 467–473.

Koch, M., Rett, K., Maerker, E., Volk, A., Haist, K., Deninger, M., Renn, W., Haring, H.U. (1999) The PPARgamma2 amino acid polymorphism Pro 12 Ala is prevalent in offspring of Type II diabetic patients and is associated to increased insulin sensitivity in a subgroup of obese subjects, *Diabetologia* **42**, 758–762.

Jacob, S., Stumvoll, M., Becker, R., Koch, M., Nielsen, M., Loblein, K., Maerker, E., Volk, A., Renn, W., Balletshofer, B., Machicao, F., Rett, K., Haring, H.U. (2000) The PPARgamma2 polymorphism pro12Ala is associated with better insulin sensitivity in the offspring of type 2 diabetic patients, *Horm. Metab. Res.* **32**, 413–416.

Jenkins, D.J., Thorne, M.J., Wolever, T.M., Jenkins, A.L., Rao, A.V., Thompson, L.U. (1987) The effect of starch-protein interaction in wheat on the glycemic response and rate of in vitro digestion, *Am. J. Clin. Nutr.* **45**, 946–951.

Jenkins, D.J., Wolever, T.M., Taylor, R.H., Ghafari, H., Jenkins, A.L., Barker, H., Jenkins, M.J. (1980) Rate of digestion of foods and postprandial glycaemia in normal and diabetic subjects, *Br. Med J.* **281**, 14–17.

Jiang, T., Sweeney, G., Rudolf, M.T., Klip, A., Traynor-Kaplan, A., Tsien, R.Y. (1998) Membrane-permeant esters of phosphatidylinositol 3,4,5-trisphosphate, *J. Biol. Chem.* **273**, 11017–11024.

Jonsson, J., Carlsson, L., Edlund, T., Edlund, H. (1994) Insulin-promoter-factor 1 is required for pancreas development in mice, *Nature* **371**(6498), 606–609.

Joshi, R.L., Lamothe, B., Cordonnier, N., Mesbah, K., Monthioux, E., Jami, J., Bucchini, D. (1996) Targeted disruption of the insulin receptor gene in the mouse results in neonatal lethality, *EMBO J.* **15**, 1542–1547.

Kadonaga, J.T., Courey, A.J., Ladika, J., Tjian, R. (1988) Distinct regions of Sp1 modulate DNA binding and transcriptional activation, *Science* **242**, 1566–1570.

Kadowaki, T., Hara, K., Kubota, N., Tobe, K., Terauchi, Y., Yamauchi, T., Eto, K., Kadowaki, H., Noda, M., Hagura, R., Akanuma, Y. (2002) The role of PPARgamma in high-fat diet-induced obesity and insulin resistance, *J. Diabetes Complications* **16**, 41–45.

Kaur, J., Singh, P., Sowers, J.R. (2002) Effects of glycemic control and other determinants on vascular disease in type 2 diabetes, *Am. J. Med.* **113**Suppl. 6A, 12S–22S.

Kawai, T., Takei, I., Oguma, Y., Ohashi, N., Tokui, M., Oguchi, S., Katsukawa, F., Hirose, H., Shimada, A., Watanabe, K., Saruta, T. (1999) Effects of troglitazone on fat distribution in the treatment of male type 2 diabetes, *Metabolism* **48**, 1102–1107.

Kawanaka, K., Han, D.H., Nolte, L.A., Hansen, P.A., Nakatani, A., Holloszy, J.O.

(1999) Decreased insulin-stimulated GLUT-4 translocation in glycogen-supercompensated muscles of exercised rats, *Am. J. Physiol.* **276**(5 Pt 1), E907–E912.

Kawanaka, K., Nolte, L.A., Han, D.H., Hansen, P.A., Holloszy, J.O. (2000) Mechanisms underlying impaired GLUT-4 translocation in glycogen-supercompensated muscles of exercised rats, *Am. J. Physiol. Endocrinol. Metab.* **279**, E1311–E1318.

Kennedy, J.W., Hirshman, M.F., Gervino, E.V., Ocel, J.V., Forse, R.A., Hoenig, S.J., Aronson, D., Goodyear, L.J., Horton, E.S. (1999) Acute exercise induces GLUT4 translocation in skeletal muscle of normal human subjects and subjects with type 2 diabetes, *Diabetes* **48**, 1192–1197.

Kim, J.K., Michael, M.D., Previs, S.F., Peroni, O.D., Mauvais-Jarvis, F., Neschen, S., Kahn, B.B., Kahn, C.R., Shulman, G.I. (2000) Redistribution of substrates to adipose tissue promotes obesity in mice with selective insulin resistance in muscle, *J. Clin. Invest.* **105**, 1791–1797.

Kim, Y.B., Nikoulina, S.E., Ciaraldi, T.P., Henry, R.R., Kahn, B.B. (1999) Normal insulin-dependent activation of Akt/protein kinase B, with diminished activation of phosphoinositide 3-kinase, in muscle in type 2 diabetes, *J. Clin. Invest.* **104**, 733–741.

Kim, J.B., Spiegelman, B.M. (1996) ADD1/SREBP1 promotes adipocyte differentiation and gene expression linked to fatty acid metabolism, *Genes Dev.* **10**, 1096–1107.

King, P.A., Betts, J.J., Horton, E.D., Horton, E.S. (1993) Exercise, unlike insulin, promotes glucose transporter translocation in obese Zucker rat muscle, *Am. J. Physiol.* **265**(2 Pt 2), R447–R452.

Kissebah, A.H., Peiris, A.N. (1989) Biology of regional body fat distribution: relationship to non-insulin-dependent diabetes mellitus, *Diabetes Metab. Rev* **5**, 83–109.

Kitamura, T., Ogawa, W., Sakaue, H., Hino, Y., Kuroda, S., Takata, M., Matsumoto, M., Maeda, T., Konishi, H., Kikkawa, U., Kasuga, M. (1998) Requirement for activation of the serine-threonine kinase Akt (protein kinase B) in insulin stimulation of protein synthesis but not of glucose transport, *Mol. Cell. Biol.* **18**, 3708–3717.

Knowler, W.C., Barrett-Connor, E., Fowler, S.E., Hamman, R.F., Lachin, J.M., Walker, E.A., Nathan, D.M. (2002) Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin, *N. Engl. J. Med.* **346**, 393–403.

Kobayashi, H., Nakamura, T., Miyaoka, K., Nishida, M., Funahashi, T., Yamashita, S., Matsuzawa, Y. (2001) Visceral fat accumulation contributes to insulin resistance, small-sized low-density lipoprotein, and progression of coronary artery disease in middle-aged non-obese Japanese men, *Jpn. Circ. J.* **65**, 193–199.

Korshunov, S.S., Skulachev, V.P., Starkov, A.A. (1997) High protonic potential actuates a mechanism of production of reactive oxygen species in mitochondria, *FEBS Lett.* **416**, 15–18.

Kotani, K., Ogawa, W., Matsumoto, M., Kitamura, T., Sakaue, H., Hino, Y., Miyake, K., Sano, W., Akimoto, K., Ohno, S., Kasuga, M. (1998) Requirement of atypical protein kinase clambda for insulin stimulation of glucose uptake but not for Akt activation in 3T3-L1 adipocytes, *Mol. Cell. Biol.* **18**, 6971–6982.

Kraegen, E.W., James, D.E., Storlien, L.H., Burleigh, K.M., Chisholm, D.J. (1986) In vivo insulin resistance in individual peripheral tissues of the high fat fed rat: assessment by euglycaemic clamp plus deoxyglucose administration, *Diabetologia* **29**, 192–198.

Krook, A., Roth, R.A., Jiang, X.J., Zierath, J.R., Wallberg-Henriksson, H. (1998) Insulin-stimulated Akt kinase activity is reduced in skeletal muscle from NIDDM subjects, *Diabetes* **47**, 1281–1286.

Krook, A., Whitehead, J.P., Dobson, S.P., Griffiths, M.R., Ouwens, M., Baker, C., Hayward, A.C., Sen, S.K., Maassen, J.A., Siddle, K., Tavare, J.M., O'Rahilly, S. (1997) Two naturally occurring insulin receptor tyrosine kinase domain mutants provide evidence that phosphoinositide 3-kinase activation alone is not sufficient for the mediation of insulin's metabolic and mitogenic effects, *J. Biol. Chem.* **272**, 30208–30214.

Krook, A., Bjornholm, M., Galuska, D., Jiang, X.J., Fahlman, R., Myers, M.G. Jr., Wallberg-Henriksson, H., Zierath, J.R. (2000) Characterization of signal transduction and glucose transport in skeletal muscle from type 2 diabetic patients, *Diabetes* **49**, 284–292.

Ku, D.N., Giddens, D.P., Zarins, C.K., Glagov, S. (1985) Pulsatile flow and atherosclerosis in the

human carotid bifurcation. Positive correlation between plaque location and low oscillating shear stress, *Arteriosclerosis* **5**, 293–302.

Kulkarni, R.N., Bruning, J.C., Winnay, J.N., Postic, C., Magnuson, M.A., Kahn, C.R. (1999) Tissue-specific knockout of the insulin receptor in pancreatic beta cells creates an insulin secretory defect similar to that in type 2 diabetes, *Cell* **96**, 329–339.

Lam, T.K., Yoshii, H., Haber, C.A., Bogdanovic, E., Lam, L., Fantus, I.G., Giacca, A. (2002) Free fatty acid induced hepatic insulin resistance: a potential role for protein kinase C-delta, *Am. J. Physiol. Endocrinol. Metab.* **283**, E682–E691.

Lambeth, J.D., Cheng, G., Arnold, R.S., Edens, W.A. (2000) Novel homologs of gp91phox, *Trends Biochem. Sci.* **25**, 459–461.

Larsen, J.J., Dela, F., Kjaer, M., Galbo, H. (1997) The effect of moderate exercise on postprandial glucose homeostasis in NIDDM patients, *Diabetologia* **40**, 447–453.

Lebovitz, H.E. (2000) Pathogenesis of type 2 diabetes, *Drug Benefit Trends* **12**Suppl. A, 8–16.

Lee, Y., Hirose, H., Zhou, Y.T., Esser, V., McGarry, J.D., Unger, R.H. (1997) Increased lipogenic capacity of the islets of obese rats: a role in the pathogenesis of NIDDM, *Diabetes* **46**, 408–413.

Lefebvre, A.M., Laville, M., Vega, N., Riou, J.P., van Gaal, L., Auwerx, J., Vidal, H. (1998) Depot-specific differences in adipose tissue gene expression in lean and obese subjects, *Diabetes* **47**, 98–103.

Lewis, G.F. (1997) Fatty acid regulation of very low density lipoprotein production, *Curr. Opin. Lipidol.* **8**, 146–153.

Li, Y., Hansotia, T., Yusta, B., Ris, F., Halban, P.A., Drucker, D.J. (2003) Glucagon-like peptide-1 receptor signaling modulates beta cell apoptosis, *J. Biol. Chem.* **278**, 471–478.

Lisinski, I., Schurmann, A., Joost, H.G., Cushman, S.W., Al-Hasani, H. (2001) Targeting of GLUT6 (formerly GLUT9) and GLUT8 in rat adipose cells, *Biochem. J.* **358**, 517–522.

Listenberger, L.L., Ory, D.S., Schaffer, J.E. (2001) Palmitate-induced apoptosis can occur through a ceramide-independent pathway, *J. Biol. Chem.* **276**, 14890–14895.

Liu, K., Paterson, A.J., Chin, E., Kudlow, J.E. (2000) Glucose stimulates protein modification by O-linked GlcNAc in pancreatic beta cells: linkage of O-linked GlcNAc to beta cell death, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2820–2825.

Lord, G.M., Matarese, G., Howard, J.K., Baker, R.J., Bloom, S.R., Lechler, R.I. (1998) Leptin modulates the T-cell immune response and reverses starvation-induced immunosuppression, *Nature* **394**, 897–901.

Lund, S., Holman, G.D., Zierath, J.R., Rincon, J., Nolte, L.A., Clark, A.E., Schmitz, O., Pedersen, O., Wallberg-Henriksson, H. (1997) Effect of insulin on GLUT4 cell surface content and turnover rate in human skeletal muscle as measured by the exofacial bis-mannose photolabeling technique, *Diabetes* **46**, 1965–1969.

MacDonald, P.E., El-Kholy, W., Riedel, M.J., Salapatek, A.M., Light, P.E., Wheeler, M.B. (2002) The multiple actions of GLP-1 on the process of glucose-stimulated insulin secretion, *Diabetes* **51**Suppl. 3, S434–S442.

Maedler, K., Spinas, G.A., Dyntar, D., Moritz, W., Kaiser, N., Donath, M.Y. (2001) Distinct effects of saturated and monounsaturated fatty acids on beta-cell turnover and function, *Diabetes* **50**, 69–76.

Maffei, M., Fei, H., Lee, G.H., Dani, C., Leroy, P., Zhang, Y., Proenca, R., Negrel, R., Ailhaud, G., Friedman, J.M. (1995) Increased expression in adipocytes of ob RNA in mice with lesions of the hypothalamus and with mutations at the db locus, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6957–6960.

Mahtani, M.M., Widen, E., Lehto, M., Thomas, J., McCarthy, M., Brayer, J., Bryant, B., Chan, G., Daly, M., Forsblom, C., Kanninen, T., Kirby, A., Kruglyak, L., Munnelly, K. et al. (1996) Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families, *Nat. Genet.* **14**, 90–94.

Malik, N.M., Carter, N.D., Murray, J.F., Scaramuzzi, R.J., Wilson, C.A., Stock, M.J. (2001) Leptin requirement for conception, implantation, and gestation in the mouse, *Endocrinology* **142**, 5198–5202.

Mammarella, S., Romano, F., Di Valerio, A., Creati, B., Esposito, D.L., Palmirotta, R., Capani, F., Vitullo, P., Volpe, G., Battista, P., Della Loggia, F., Mariani-Costantini, R., Cama, A. (2000) Interaction between the G1057D variant of IRS-2 and overweight in the pathogenesis of type 2 diabetes, *Hum. Mol. Genet.* **9**, 2517–2521.

Marette, A., Mauriege, P., Marcotte, B., Atgie, C., Bouchard, C., Theriault, G., Bukowiecki, L.J.,

Marceau, P., Biron, S., Nadeau, A., Despres, J.P. (1997) Regional variation in adipose tissue insulin action and GLUT4 glucose transporter expression in severely obese premenopausal women, *Diabetologia* **40**, 590–598.

Massillon, D., Barzilai, N., Hawkins, M., Prus-Wertheimer, D., Rossetti, L. (1997) Induction of hepatic glucose-6-phosphatase gene expression by lipid infusion, *Diabetes* **46**, 153–157.

Mathers, J. (2002) Choosing your carbohydrates to prevent diabetes, *Br. J. Nutr.* **88**, 107–108.

Mathias, S., Pena, L.A., Kolesnick, R.N. (1998) Signal transduction of stress via ceramide, *Biochem. J.* **335**(Pt 3), 465–480.

Matsuzawa, Y., Funahashi, T., Nakamura, T. (1999) Molecular mechanism of metabolic syndrome X: contribution of adipocytokines adipocyte-derived bioactive substances, *Ann. N.Y. Acad. Sci.* **892**, 146–154.

Michael, M.D., Kulkarni, R.N., Postic, C., Previs, S.F., Shulman, G.I., Magnuson, M.A., Kahn, C.R. (2000) Loss of insulin signaling in hepatocytes leads to severe insulin resistance and progressive hepatic dysfunction, *Mol. Cell* **6**, 87–97.

Miller, F.J. Jr., Gutterman, D.D., Rios, C.D., Heistad, D.D., Davidson, B.L. (1998) Superoxide production in vascular smooth muscle contributes to oxidative stress and impaired relaxation in atherosclerosis, *Circ. Res.* **82**, 1298–1305.

Mohamed-Ali, V., Pinkney, J.H., Coppack, S.W. (1998) Adipose tissue as an endocrine and paracrine organ, *Int. J. Obes. Relat. Metab. Disord.* **22**, 1145–1158.

Moldovan, L., Moldovan, N.I., Sohn, R.H., Parikh, S.A., Goldschmidt-Clermont, P.J. (2000) Redox changes of cultured endothelial cells and actin dynamics, *Circ. Res.* **86**, 549–557.

Montague, C.T., O'Rahilly, S. (2000) The perils of portliness: causes and consequences of visceral adiposity, *Diabetes* **49**, 883–888.

Montague, C.T., Prins, J.B., Sanders, L., Zhang, J., Sewter, C.P., Digby, J., Byrne, C.D., O'Rahilly, S. (1998) Depot-related gene expression in human subcutaneous and omental adipocytes, *Diabetes* **47**, 1384–1391.

Mori, Y., Murakawa, Y., Okada, K., Horikoshi, H., Yokoyama, J., Tajima, N., Ikeda, Y. (1999) Effect of troglitazone on body fat distribution in type 2 diabetic patients, *Diabetes Care* **22**, 908–912.

Musi, N., Fujii, N., Hirshman, M.F., Ekberg, I., Froberg, S., Ljungqvist, O., Thorell, A., Goodyear, L.J. (2001) AMP-activated protein kinase (AMPK) is activated in muscle of subjects with type 2 diabetes during exercise, *Diabetes* **50**, 921–927.

Nesher, R., Karl, I.E., Kipnis, D.M. (1985) Dissociation of effects of insulin and contraction on glucose transport in rat epitrochlearis muscle, *Am. J. Physiol.* **249**(3 Pt 1), C226–C232.

Ng, D., Xuan W.L. (2002) Unpublished data.

Nishigori, H., Yamada, S., Kohama, T., Tomura, H., Sho, K., Horikawa, Y., Bell, G.I., Takeuchi, T., Takeda, J. (1998) Frameshift mutation, A263fsinsGG, in the hepatocyte nuclear factor-1beta gene associated with diabetes and renal dysfunction, *Diabetes* **47**, 1354–1355.

Noyman, I., Marikovsky, M., Sasson, S., Stark, A.H., Bernath, K., Seger, R., Madar, Z. (2002) Hyperglycemia reduces nitric oxide synthase and glycogen synthase activity in endothelial cells, *Nitric Oxide* **7**, 187–193.

Oakes, N.D., Cooney, G.J., Camilleri, S., Chisholm, D.J., Kraegen, E.W. (1997) Mechanisms of liver and muscle insulin resistance induced by chronic high-fat feeding, *Diabetes* **46**, 1768–1774.

Okamoto, Y., Kihara, S., Ouchi, N., Nishida, M., Arita, Y., Kumada, M., Ohashi, K., Sakai, N., Shimomura, I., Kobayashi, H., Terasaka, N., Inaba, T., Funahashi, T., Matsuzawa, Y. (2002) Adiponectin reduces atherosclerosis in apolipoprotein E-deficient mice, *Circulation* **106**, 2767–2770.

O'Brien, R.M., Granner, D.K. (1996) Regulation of gene expression by insulin, *Physiol. Rev.* **76**, 1109–1161.

Opara, E.C., Garfinkel, M., Hubbard, V.S., Burch, W.M., Akwari, O.E. (1994) Effect of fatty acids on insulin release: role of chain length and degree of unsaturation, *Am. J. Physiol.* **266**(4 Pt 1), E635–E639.

Oral, E.A., Simha, V., Ruiz, E., Andewelt, A., Premkumar, A., Snell, P., Wagner, A.J., DePaoli, A.M., Reitman, M.L., Taylor, S.I., Gorden, P., Garg, A. (2002) Leptin-replacement therapy for lipodystrophy, *N. Engl. J. Med.* **346**, 570–578.

Ouchi, N., Kihara, S., Arita, Y., Nishida, M., Matsuyama, A., Okamoto, Y., Ishigami, M., Kuriyama, H., Kishida, K., Nishizawa, H., Hotta, K. et al. (2001) Adipocyte-derived plasma protein, adiponectin, suppresses lipid accumulation and class A scavenger receptor

expression in human monocyte-derived macrophages, *Circulation* **103**, 1057–1063.

Paolisso, G., Tataranni, P.A., Foley, J.E., Bogardus, C., Howard, B.V., Ravussin, E. (1995) A high concentration of fasting plasma non-esterified fatty acids is a risk factor for the development of NIDDM, *Diabetologia* **38**, 1213–1217.

Park, J.Y., Takahara, N., Gabriele, A., Chou, E., Naruse,K., Suzuma, K., Yamauchi, T., Ha, S.W., Meier, M., Rhodes, C.J., King, G.L. (2000) Induction of endothelin-1 expression by glucose: an effect of protein kinase C activation, *Diabetes* **49**, 1239–1248.

Perez-Martin, A., Raynaud, E., Mercier, J. (2001) Insulin resistance and associated metabolic abnormalities in muscle: effects of exercise, *Obes. Rev.* **2**, 47–59.

Petersen, K.F., Oral, E.A., Dufour, S., Befroy, D., Ariyan, C., Yu, C., Cline, G.W., DePaoli, A.M., Taylor, S.I., Gorden, P., Shulman, G.I. (2002) Leptin reverses insulin resistance and hepatic steatosis in patients with severe lipodystrophy, *J. Clin. Invest.* **109**, 1345–1350.

Prentki, M., Corkey, B.E. (1996) Are the β-cell signaling molecules rmalonyl-CoA and cytosolic long-chain acyl-CoA implicated in multiple tissue defects of obesity and NIDDM?, *Diabetes* **45**, 273–283.

Rahier, J., Boebbles, R.M., Henquin, J.C. (1983) Cellular composition of the human diabetic pancreas, *Diabetologia* **24**, 366–371.

Randle, P.J., Garland, P.B., Hales, C.N. (1963) The glucose fatty acid cycle: its role in insulin sensitivity and the metabolic disturbances of diabetes mellitus, *Lancet* **1**, 785–789.

Rao, A.K., Chouhan, V., Chen, X., Sun, L., Boden, G. (1999) Activation of the tissue factor pathway of blood coagulation during prolonged hyperglycemia in young healthy men, *Diabetes* **48**, 1156–1161.

Ravussin, E., Smith, S.R. (2002) Increased fat intake, impaired fat oxidation, and failure of fat cell proliferation result in ectopic fat storage, insulin resistance, and type 2 diabetes mellitus, *Ann. N.Y. Acad. Sci* **967**, 363–378.

Richelsen, B., Pedersen, S.B., Moller-Pedersen, T., Bak, J.F. (1991) Regional differences in triglyceride breakdown in human adipose tissue: effects of catecholamines, insulin, and prostaglandin E2, *Metabolism* **40**, 990–996.

Richter, E.A., Wojtaszewski, J.F., Kristiansen, S., Daugaard, J.R., Nielsen, J.N., Derave, W., Kiens, B. (2001) Regulation of muscle glucose transport during exercise, *Int. J. Sport Nutr. Exerc. Metab.* **11**(Suppl.) S71–S77.

Rivellese, A.A., de Natale, C., Lilli, S. (2002) Type of dietary fat and insulin resistance, *Ann. N.Y. Acad. Sci.* **967**, 329–335.

Roden, M., Stingl, H., Chandramouli, V., Schumann, W.C., Hofer, A., Landau, B.R., Nowotny, P., Waldhausl, W., Shulman, G.I. (2000) Effects of free fatty acid elevation on postabsorptive endogenous glucose production and gluconeogenesis in humans, *Diabetes* **49**, 701–707.

Rondinone, C.M., Wang, L.M., Lonnroth, P., Wesslau, C., Pierce, J.H., Smith, U. (1997) Insulin receptor substrate (IRS) 1 is reduced and IRS-2 is the main docking protein for phosphatidylinositol 3-kinase in adipocytes from subjects with non-insulin-dependent diabetes mellitus, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4171–4175.

Ruderman, N.B., Saha, A.K., Vavvas, D., Witters, L.A. (1999) Malonyl-CoA, fuel sensing, and insulin resistance, *Am. J. Physiol.* **276**(1 Pt 1), E1–E18.

Ryder, J.W., Yang, J., Galuska, D., Rincon, J., Bjornholm, M., Krook, A., Lund, S., Pedersen, O., Wallberg-Henriksson, H., Zierath, J.R., Holman, G.D. (2000) Use of a novel impermeable biotinylated photolabeling reagent to assess insulin- and hypoxia-stimulated cell surface GLUT4 content in skeletal muscle from type 2 diabetic patients, *Diabetes* **49**, 647–654.

Saengsirisuwan, V., Kinnick, T.R., Schmit, M.B., Henriksen, E.J. (2001) Interactions of exercise training and lipoic acid on skeletal muscle glucose transport in obese Zucker rats, *J. Appl. Physiol.* **91**, 145–153.

Saha, A.K., Laybutt, D.R., Dean, D., Vavvas, D., Sebokova, E., Ellis, B., Klimes, I., Kraegen, E.W., Shafrir, E., Ruderman, N.B. (1999) Cytosolic citrate and malonyl-CoA regulation in rat muscle in vivo, *Am. J. Physiol.* **276**(6 Pt 1), E1030–E1037.

Saha, A.K., Vavvas, D., Kurowski, T.G., Apazidis, A., Witters, L.A., Shafrir, E., Ruderman, N.B. (1997) Malonyl-CoA regulation in skeletal muscle: its link to cell citrate and the glucose-fatty acid cycle, *Am. J. Physiol.* **272**(4 Pt 1), E641–E648.

Salmeron, J., Ascherio, A., Rimm, E.B., Colditz, G.A. et al. (1997) Dietary fiber, glycemic load, and risk of NIDDM in men, *Diabetes Care* **20**, 545–550.

Samad, F., Pandey, M., Bell, P.A., Loskutoff, D.J. (2000) Insulin continues to induce plasminogen activator inhibitor 1 gene expression in insulin-resistant mice and adipocytes, *Mol. Med.* **6**, 680–692.

Samad, F., Yamamoto, K., Loskutoff, D.J. (1996) Distribution and regulation of plasminogen activator inhibitor-1 in murine adipose tissue in vivo. Induction by tumor necrosis factor-alpha and lipopolysaccharide, *J. Clin. Invest.* **97**, 37–46.

Sato, R., Miyamoto, W., Inoue, J., Terada, T., Imanaka, T., Maeda, M. (1999) Sterol regulatory element-binding protein negatively regulates microsomal triglyceride transfer protein gene transcription, *J. Biol. Chem.* **274**, 24714–24720.

Schneider, B.S., Faust, I.M., Hemmes, R., Hirsch, J. (1981) Effects of altered adipose tissue morphology on plasma insulin levels in the rat, *Am. J. Physiol.* **240**, E358–E362.

Shechter, M., Merz, C.N., Paul-Labrador, M.J., Kaul, S. (2000) Blood glucose and platelet-dependent thrombosis in patients with coronary artery disease, *J. Am. Coll. Cardiol.* **35**, 300–307.

Shepherd, P.R., Gnudi, L., Tozzo, E., Yang, H., Leach, F., Kahn, B.B. (1993) Adipose cell hyperplasia and enhanced glucose disposal in transgenic mice overexpressing GLUT4 selectively in adipose tissue, *J. Biol. Chem.* **268**, 22243–22246.

Shepherd, P.R., Withers, D.J., Siddle, K. (1998) Phosphoinositide 3-kinase: the key switch mechanism in insulin signalling, *Biochem. J.* **333**(Pt 3), 471–490.

Sherman, W.M., Friedman, J.E., Gao, J.P., Reed, M.J., Elton, C.W., Dohm, G.L. (1993) Glycemia and exercise training alter glucose transport and GLUT4 in the Zucker rat, *Med. Sci. Sports Exerc.* **25**, 341–348.

Shimabukuro, M., Zhou, Y.-T., Levi, M., Unger, R.H. (1998) Fatty acid-induced β cell apoptosis: A link between obesity and diabetes, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2498–2502.

Shimano, H., Horton, J.D., Shimomura, I., Hammer, R.E., Brown, M.S., Goldstein, J.L. (1997) Isoform 1c of sterol regulatory element binding protein is less active than isoform 1a in livers of transgenic mice and in cultured cells, *J. Clin. Invest.* **99**, 846–854.

Shimomura, I., Bashmakov, Y., Ikemoto, S., Horton, J.D., Brown, M.S., Goldstein, J.L. (1999) Insulin selectively increases SREBP-1c mRNA in the livers of rats with streptozotocin-induced diabetes, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13656–13661.

Shimomura, I., Hammer, R.E., Ikemoto, S., Brown, M.S., Goldstein, J.L. (1999) Leptin reverses insulin resistance and diabetes mellitus in mice with congenital lipodystrophy, *Nature* **401**(6748), 73–76.

Shimomura, I., Hammer, R.E., Richardson, J.A., Ikemoto, S., Bashmakov, Y., Goldstein, J.L., Brown, M.S. (1998) Insulin resistance and diabetes mellitus in transgenic mice expressing nuclear SREBP-1c in adipose tissue: model for congenital generalized lipodystrophy, *Genes Dev.* **12**, 3182–3194.

Shimomura, I., Matsuda, M., Hammer, R.E., Bashmakov, Y., Brown, M.S., Goldstein, J.L. (2000) Decreased IRS-2 and increased SREBP-1c lead to mixed insulin resistance and sensitivity in livers of lipodystrophic and ob/ob mice, *Mol. Cell* **6**, 77–86.

Shulman, G.I., Rothman, D.L., Jue, T., Stein, P., DeFronzo, R.A., Shulman, R.G. (1990) Quantitation of muscle glycogen synthesis in normal subjects and subjects with non-insulin-dependent diabetes by 13C nuclear magnetic resonance spectroscopy, *N. Engl. J. Med.* **322**, 223–228.

Skyler, J.S. (1998) Insulin therapy in type 2 diabetes mellitus, in: DeFronzo R.A (Ed.) *Current Therapy of Diabetes Mellitus*, Mosby-Year Book, St. Louis, AMO, pp. 108–116.

Smith, U. (1985) Regional differences in adipocyte metabolism and possible consequences in vivo, *Int. J. Obes.* **9**Suppl. 1, 145–148.

Sobel, B.E. (2002) Effects of glycemic control and other determinants on vascular disease in type 2 diabetes, *Am. J. Med.* **113**Suppl. 6A, 12S–22S.

Song, X.M., Kawano, Y., Krook, A., Ryder, J.W., Efendic, S., Roth, R.A., Wallberg-Henriksson, H., Zierath, J.R. (1999) Muscle fiber type-specific defects in insulin signal transduction to glucose transport in diabetic GK rats, *Diabetes* **48**, 664–670.

Spiegel, S., Merrill, A.H. Jr. (1996) Sphingolipid metabolism and cell growth regulation., *FASEB J.* **10**, 1388–1397.

Stratton, I.M., Adler, A.I., Neil, H.A., Matthews, D.R., Manley, S.E., Cull, C.A., Hadden, D., Turner, R.C., Holman, R.R. (2000) Association of glycaemia with macrovascular and

microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study, *BMJ* **321**(7258), 405–412.

Staubs, P.A., Nelson, J.G., Reichart, D.R., Olefsky, J.M. (1998) Platelet-derived growth factor inhibits insulin stimulation of insulin receptor substrate-1-associated phosphatidylinositol 3-kinase in 3T3-L1 adipocytes without affecting glucose transport, *J. Biol. Chem.* **273**, 25139–25147.

Steppan, C., Lazar, M. (2002) Resistin and obesity-associated insulin resistance, *Trends Endcrinol. Metab.* **13**, 18–23.

Stoffel, M., Duncan, S.A. (1997) The maturity-onset diabetes of the young (MODY1) transcription factor HNF4alpha regulates expression of genes required for glucose transport and metabolism, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13209–13214.

Sundaresan, M., Yu, Z.X., Ferrans, V.J., Irani, K., Finkel, T. (1995) Requirement for generation of H2O2 for platelet-derived growth factor signal transduction, *Science* **270**, 296–299.

Suzuki, M., Akimoto, K., Hattori, Y. (2002) Glucose upregulates plasminogen activator inhibitor-1 gene expression in vascular smooth muscle cells, *Life Sci.* **72**, 59–66.

Thorburn, A.W., Brand, J.C., O'Dea, K., Spargo, R.M., Truswell, A.S. (1987) Plasma glucose and insulin responses to starchy foods in Australian aborigines: a population now at high risk of diabetes, *Am. J. Clin. Nutr.* **46**, 282–285.

Thorburn, A.W., Gumbiner, B., Bulacan, F., Wallace, P., Henry, R.R. (1990) Intracellular glucose oxidation and glycogen synthase activity are reduced in non-insulin-dependent (type II) diabetes independent of impaired glucose uptake, *J. Clin. Invest.* **85**, 522–529.

Thornalley, P.J. (1990) The glyoxalase system: new developments towards functional characterization of a metabolic pathway fundamental to biological life, *Biochem. J.* **269**, 1–11.

Tomas, E., Lin, Y.S., Dagher, Z., Saha, A., Luo, Z., Ido, Y., Ruderman, N.B. (2002) Hyperglycemia and insulin resistance: possible mechanisms, *Ann. N.Y. Acad. Sci.* **967**, 43–51.

Tontonoz, P., Hu, E., Spiegelman, B.M. (1994) Stimulation of adipogenesis in fibroblasts by PPARγ2, a lipid-activated transcription factor, *Cell* **79**, 1147–1156.

Topper, J.N., Cai, J., Falb, D., Gimbrone, M.A. Jr. (1996) Identification of vascular endothelial genes differentially responsive to fluid mechanical stimuli: cyclooxygenase-2, manganese superoxide dismutase, and endothelial cell nitric oxide synthase are selectively upregulated by steady laminar shear stress, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10417–10422.

Tozzo, E., Gnudi, L., Kahn, B.B. (1997) Amelioration of insulin resistance in streptozotocin diabetic mice by transgenic overexpression of GLUT4 driven by an adipose-specific promoter, *Endocrinology* **138**(4), 1604–1611.

Trent, D.F., Fletcher, D.J., May, J.M., Bonner-Weir, S., Weir, G.C. (1984) Abnormal islet and adipocyte function in young B-cell-deficient rats with near-normoglycemia, *Diabetes* **33**, 170–175.

Tuomilehto, J., Lindstrom, J., Eriksson, J.G., Valle, T.T., Hamalainen, H., Ilanne-Parikka, P., Keinanen-Kiukaanniemi, S., Laakso, M., Louheranta, A., Rastas, M., Salminen, V., Uusitupa, M. (2001) Finnish Diabetes Prevention Study Group. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance, *N. Engl. J. Med.* **344**, 1343–1350.

Tuttle, R.L., Gill, N.S., Pugh, W., Lee, J.P., Koeberlein, B., Furth, E.E., Polonsky, K.S., Naji, A., Birnbaum, M.J. (2001) Regulation of pancreatic beta-cell growth and survival by the serine/threonine protein kinase Akt1/PKBalpha, *Nat. Med.* **7**, 1133–1137.

Unger, R.H. (1995) Lipotoxicity in the pathogenesis of obesity-dependent NIDDM. Genetic and clinical implications, *Diabetes* **44**, 863–870.

Unger, R.H. (2001) Lipotoxicity of beta-cells in obesity and in other causes of fatty acid spillover, *Diabetes* **50**Suppl. 1, S118–S121.

United Kingdom Prospective Diabetes Study (UKPDS), UK Prospective Diabetes Study (UKPDS) Group. (1998) Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34), *Lancet* **352**, 854–865.

Ushio-Fukai, M., Zafari, A.M., Fukui, T., Ishizaka, N., Griendling, K.K. (1996) p22phox is a critical component of the superoxide-generating NADH/NADPH oxidase system and regulates angiotensin II-induced hypertrophy in vascular smooth muscle cells, *J. Biol. Chem.* **271**, 23317–23321.

Vaxillaire, M., Abderrahmani, A., Boutin, P., Bailleul, B., Froguel, P., Yaniv, M., Pontoglio, M. (1999) Anatomy of a homeoprotein revealed by the analysis of human MODY3 mutations, *J. Biol. Chem.* **274**, 35639–35646.

Velho, G., Froguel, P. (1998) Genetic, metabolic and clinical characteristics of maturity onset diabetes of the young, *Eur. J. Endocrinol.* **138**, 233–239.

Vinik, A., Bell, G. (1988) Mutant insulin syndromes, *Horm. Metab. Res.* **20**, 1–10.

Wallberg-Henriksson, H. (1992) Exercise and diabetes mellitus, *Exerc. Sport Sci. Rev.* **20**, 339–368.

Wang, Q., Somwar, R., Bilan, P.J., Liu, Z., Jin, J., Woodgett, J.R., Klip, A. (1999) Protein kinase B/Akt participates in GLUT4 translocation by insulin in L6 myoblasts, *Mol. Cell. Biol.* **19**, 4008–4018.

Watson, R.T., Pessin, J.E. (2001a) Intracellular organization of insulin signaling and GLUT4 translocation, *Recent Prog. Horm. Res.* **56**, 175–193.

Watson, R.T., Pessin, J.E. (2001b) Subcellular compartmentalization and trafficking of the insulin-responsive glucose transporter, GLUT4, *Exp. Cell Res.* **271**, 75–83.

Watson, R.T., Shigematsu, S., Chiang, S.H., Mora, S., Kanzaki, M., Macara, I.G., Saltiel, A.R., Pessin, J.E. (2001c) Lipid raft microdomain compartmentalization of TC10 is required for insulin signaling and GLUT4 translocation, *J. Cell Biol.* **154**, 829–840.

Wells-Knecht, K.J., Zyzak, D.V., Litchfield, J.E., Thorpe, S.R., Baynes, J.W. (1995) Mechanism of autoxidative glycosylation: identification of glyoxal and arabinose as intermediates in the autoxidative modification of proteins by glucose, *Biochemistry* **34**, 3702–3709.

Weyer, C., Foley, J.E., Bogardus, C., Tataranni, P.A., Pratley, R.E. (2000) Enlarged subcutaneous abdominal adipocyte size, but not obesity itself, predicts type II diabetes independent of insulin resistance, *Diabetologia* **43**, 1498–1506.

Widen, E., Lehto, M., Kanninen, T., Walston, J., Shuldiner, A.R., Groop, L.C. (1995) Association of a polymorphism in the beta 3-adrenergic-receptor gene with features of the insulin resistance syndrome in Finns, *N. Engl. J. Med.* **333**, 348–351.

Williams, B., Gallacher, B., Patel, H., Orme, C. (1997) Glucose-induced protein kinase C activation regulates vascular permeability factor mRNA expression and peptide production by human vascular smooth muscle cells in vitro, *Diabetes* **46**, 1497–1503.

Wojtaszewski, J.F., Hansen, B.F., Gade, J., Kiens, B., Markuns, J.F., Goodyear, L.J., Richter, E.A. (2000) Insulin signaling and insulin sensitivity after exercise in human skeletal muscle, *Diabetes* **49**, 325–331.

Wojtaszewski, J.F., Higaki, Y., Hirshman, M.F., Michael, M.D., Dufresne, S.D., Kahn, C.R., Goodyear, L.J. (1999) Exercise modulates postreceptor insulin signaling and glucose transport in muscle-specific insulin receptor knockout mice, *J. Clin. Invest.* **104**, 1257–1264.

Wojtaszewski, J.F., Jorgensen, S.B., Hellsten, Y., Hardie, D.G., Richter, E.A. (2002a) Glycogen-dependent effects of 5-aminoimidazole-4-carboxamide (AICA)-riboside on AMP-activated protein kinase and glycogen synthase activities in rat skeletal muscle, *Diabetes* **51**, 284–292.

Wojtaszewski, J.F., Nielsen, J.N., Richter, E.A. (2002b) Invited Review: Effect of acute exercise on insulin signaling and action in humans, *J. Appl. Physiol.* **93**, 384–392.

WHO Fact Sheets. www.who.int/ncd/dia/.

Wrede, C.E., Dickson, L.M., Lingohr, M.K., Briaud, I., Rhodes, C.J. (2002) Protein kinase B/Akt prevents fatty acid-induced apoptosis in pancreatic beta -cells (INS-1), *J. Biol. Chem.* **277**, 49676–49684.

Xu, J., Nakamura, M.T., Cho, H.P., Clarke, S.D. (1999) Sterol regulatory element binding protein-1 expression is suppressed by dietary polyunsaturated fatty acids. A mechanism for the coordinate suppression of lipogenic genes by polyunsaturated fats, *J. Biol. Chem.* **274**, 23577–23583.

Xu, J., Teran-Garcia, M., Park, J.H., Nakamura, M.T., Clarke, S.D. (2001) Polyunsaturated fatty acids suppress hepatic sterol regulatory element-binding protein-1 expression by accelerating transcript decay, *J. Biol. Chem.* **276**, 9800–9807.

Yahagi, N., Shimano, H., Hasty, A.H., Matsuzaka, T., Ide, T., Yoshikawa, T. et al. (2002) Absence of sterol regulatory element-binding protein-1 (SREBP-1) ameliorates fatty livers but not obesity or insulin resistance in Lep(ob)/Lep(ob) mice, *J. Biol. Chem.* **277**, 19353–19357.

Yamauchi, T., Kamon, J., Waki, H., Terauchi, Y., Kubota, N., Hara, K., Mori, Y., Ide, T., Murakami, K., Tsuboyama-Kasaoka, N., Ezaki, O.

et al. (2001) The fat-derived hormone adiponectin reverses insulin resistance associated with both lipoatrophy and obesity, *Nat. Med.* **7**, 941–946.

Yen, C.J., Beamer, B.A., Negri, C., Silver, K., Brown, K.A., Yarnall, D.P., Burns, D.K., Roth, J., Shuldiner, A.R. (1997) Molecular scanning of the human peroxisome proliferator activated receptor gamma (hPPAR gamma) gene in diabetic Caucasians: identification of a Pro12Ala PPAR gamma 2 missense mutation, *Biochem. Biophys. Res. Commun.* **241**, 270–274.

Yerneni, K.K., Bai, W., Khan, B.V., Medford, R.M., Natarajan, R. (1999) Hyperglycemia-induced activation of nuclear transcription factor kappaB in vascular smooth muscle cells, *Diabetes* **48**, 855–864.

Yokoyama, C., Wang, X., Briggs, M.R., Admon, A., Wu, J., Hua, X., Goldstein, J.L., Brown, M.S. (1993) SREBP-1, a basic-helix-loop-helix-leucine zipper protein that controls transcription of the low density lipoprotein receptor gene, *Cell* **75**, 187–197.

Yoshikawa, T., Shimano, H., Yahagi, N., Ide, T., Amemiya-Kudo, M., Matsuzaka, T. et al. (2002) Polyunsaturated fatty acids suppress sterol regulatory element-binding protein 1c promoter activity by inhibition of liver X receptor (LXR) binding to LXR response elements, *J. Biol. Chem.* **277**, 1705–1711.

Yusuf, S., Sleight, P., Pogue, J., Bosch, J., Davies, R., Dagenais, G. (2000) Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators, *N. Engl. J. Med.* **342**, 145–153.

Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L., Friedman, J.M. (1994) Positional cloning of the mouse obese gene and its human homologue, *Nature* **372**(6505), 425–432.

Zhou, Y.P., Grill, V.E. (1995) Palmitate-induced beta cell insensitivity to glucose is coupled to decreased pyruvate dehydrogenase activity and enhanced kinase activity in rat pancreatic islets, *Diabetes* **44**, 194–199.

Zierath, J.R., He, L., Guma, A., Odegoard Wahlstrom, E., Klip, A., Wallberg-Henriksson, H. (1996) Insulin action on glucose transport and plasma membrane GLUT4 content in skeletal muscle from patients with NIDDM, *Diabetologia* **39**, 1180–1189.

Zierath, J.R., Krook, A., Wallberg-Henriksson, H. (2000) Insulin action and insulin resistance in human skeletal muscle, *Diabetologia* **43**, 821–835.

Zobolotny, J.M., Bence-Hanulee, K.K., Stricker-Krongrad, A. et al. (2002) PTP1B regulates leptin signal transduction in vivo, *Dev. Cell* **2**, 385–387.

Zou, M.H., Shi, C., Cohen, R.A. (2002) Oxidation of the zinc-thiolate complex and uncoupling of endothelial nitric oxide synthase by peroxynitrite. *J. Clin. Invest.* **109**, 817–826.

Intracellular Fatty Acid Binding Proteins and Fatty Acid Transport

Judith Storch[1] *and Lindsay McDermott*[2]
[1] *Rutgers University, New Brunswick, NJ*
[2] *University of Glasgow, Glasgow, UK*

1 **Equilibrium Binding of Fatty Acids to FABPs 74**

2 ***In vitro* Fatty Acid–Transfer Properties of FABPs 76**

3 **Transfection Studies of FABP Function 80**

4 **Cellular Fatty Acid Transport via FABP–Protein Interactions 81**

5 **Insights into FABP Function from Null Mice 83**

6 **Perspectives 86**

Bibliography 87
Books and Reviews 87
Primary Literature 87

Keywords

Cell Transfections
This is a method whereby a biologically active protein is introduced (chemically or electrically) to cultured mammalian cells. Live cells are then assayed to determine the effects of the introduced protein. Results from such experiments may reveal how the protein behaves within the cell.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

Fatty Acids
This is a long carbon chain carboxylic acid that condenses with glycerol to form fat. The carbon chain can vary in length and number and positioning of doulble bonds.

Knockout Mouse
This is a scientific model used to explore the role that a particular gene may play in humans. Within the knockout mouse, the gene of interest is removed using artifical methods; all cells within the resulting mutant mouse contain the same mutation. The appearance, biochemical characteristics, and behavior of the knockout mouse can provide some indication of the gene's normal role in the mouse.

Lipid-binding Proteins
These are small, soluble intracellular proteins capable of non–covalently binding fatty acids and other small hydrophobic ligands. It is believed to participate in lipid transport.

Lipid Transport
This is the movement of lipids from the site of synthesis to their site of utilization.

Long-chain fatty acids (FA) are required by cells as membrane phospholipid constituents, metabolic substrates, precursors for signaling molecules, and mediators of gene expression. They are in constant flux and need to enter and leave cells rapidly and, presumably, in a regulated manner. The relatively low aqueous solubility of fatty acids would strongly suggest that specific and efficient mechanisms must exist for their intracellular transport. High levels of fatty acid–binding proteins (FABPs) are found within cells, and, although it has been shown that these proteins noncovalently bind fatty acids with high affinity, their true *in vivo* functions have remained elusive. This chapter focuses on recent findings assessing the transport function of FABPs, and on data supporting putative mechanisms by which FABPs may be involved in cellular FA uptake, efflux, and intracellular transport.

1
Equilibrium Binding of Fatty Acids to FABPs

There are 12 members of the mammalian FABP family, each with specific tissue expression and, with the exception of the retinol and retinoic acid–binding proteins, each is named after the first tissue of isolation (Table 1). In the tissues where they are found, FABPs are typically expressed abundantly, at reported levels of up to 13% cytosolic protein and, as their name suggests, are capable of noncovalently binding long-chain fatty acids with high affinity and a 1 : 1 molar stoichiometry. Ileal lipid-binding protein (I-LBP) and liver fatty acid–binding protein (L-FABP) prove exceptions to this rule: both can bind more bulky, hydrophobic ligands such as lysophospholipids, bile acids, eicosanoids,

Tab. 1 Members of the family of mammalian intracellular fatty acid–binding proteins.

Name	*Occurrence*
E-FABP (K-FABP)	Epidermis, adipose, mammary tissue, tongue epithelia, testis
H-FABP	Heart muscle, cardiac and skeletal muscle, brain, mammary gland, kidney, adrenals, ovaries, testis
B-FABP	Brain, central nervous system
M-FABP	Peripheral nervous system
A-FABP	Adipose, macrophages
I-FABP	Small intestine
I-LBP	Small intestine (distal)
L-FABP	Liver, small intestine
CRABP-I	Brain, skin, testis
CRABP-II	Epidermis, adrenal
CRBP-I	Liver, kidney, testis, lung
CRBP-II	Small intestine

and some drugs, I-LBP does not bind fatty acids, and L-FABP is capable of binding two fatty acids simultaneously. A recent NMR spectroscopic study showed two distinct binding environments for these FA, and suggested that binding of the first fatty acid precedes and may facilitate binding of the second.

Equilibrium binding studies have been used in an attempt to elucidate the functional characteristics of each FABP type. Most recently, the ligand-binding specificity of eight human FABPs (heart, liver, intestine, adipocyte, myelin, epidermal, brain fatty acid–binding protein, and ileal lipid-binding protein) were directly compared by Zimmerman et al. using the Lipidex assay. By determining the equilibrium distribution of FAs between the resin and FABP, binding affinities were measured and, with the exception of I-LBP, ranged from 0.2 to 4.01 μM. The results obtained showed that the proteins have a lower affinity for palmitic acid than for oleic and arachidonic acids. Contrary to these findings and using the ADIFAB (acrylodated intestinal fatty acid–binding protein) method, Richieri and colleagues showed that human FABPs from brain, heart, intestine, liver, and myelin pertained little or no selectivity for a particular FA, and obtained K_d values ranging from 2 to 400 nM. ADIFAB consists of intestinal fatty acid–binding protein covalently modified with the fluorescent acrylodan group, which exhibits a marked red-shift in fluorescence emission maximum upon fatty acid binding, enabling unbound concentrations of FAs to be accurately measured. A further study using isothermal titration calorimetry (ITC) to determine the binding of FAs to human L-FABP yielded a stearate K_d value approximately 100 times larger (weaker binding) than that determined by Richieri et al., and concluded, again in contrast to Richieri et al., that L-FABP preferentially binds unsaturated relative to saturated FAs. It is possible, however, that the poor solubility of stearate, as well as the high fatty acid concentrations necessary for the ITC injection method, resulted in a lower stearate concentration in the reaction vessel, thus producing lower values for heat change, and therefore uncertain K_d values.

In spite of these discrepancies over absolute K_d values, it is clear from these and earlier *in vitro* binding experiments that FABPs bind long-chain saturated and unsaturated FA. Indeed, NMR and X-ray crystallographic structures of holo FABPs reveal the position of the bound FA within the individual protein structures. In some structures, including intestinal FABP (I-FABP) and adipocyte FABP (A-FABP), the fatty acid adopts a bent conformation, while in others, such as heart FABP (H-FABP), it adopts a U-shaped conformation.

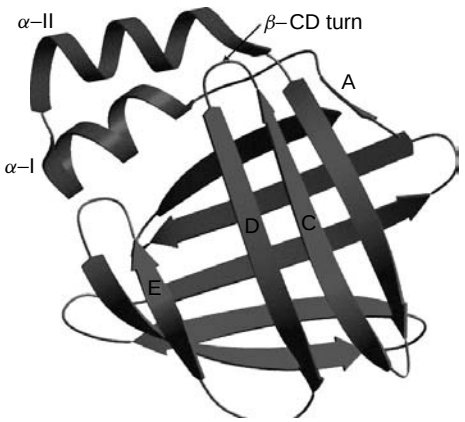

Fig. 1 The three-dimensional structure of a representative cytoplasmic fatty acid–binding protein. The protein comprises a flattened β-barrel capped by two short α-helices. The helices and closely positioned α-turns are believed to behave as a portal for ligand entry and exit.

Notwithstanding a wide variance in primary sequence, all members of the FABP family consist of a β-barrel structure capped by two α-helices; the latter is believed to behave as a portal for ligand entry and release (Fig. 1). This hypothesis was supported recently by fluorescence-based experiments comparing A-FABP and a triple mutant (V32G, F57G, K58G), designed to enlarge the putative portal opening by reducing the size of portal amino acids. By comparing analinonaphthalene sulfonic acid (ANS) and oleate-binding affinities and ANS-binding rates, it was found that enlargement of the A-FABP portal region increased ligand accessibility into the cavity with only modest effects on ligand-binding affinity, suggesting that dynamic fluctuations in this region regulate cavity access. Indeed, the solution structure of apo I-FABP was shown to exhibit a higher degree of mobility in this portal area in comparison with that of the ligand-bound I-FABP complex relative to other domains in these proteins. In an attempt to better understand the mechanism by which FAs bind to and dissociate from the binding cavities of FABPs, Richieri et al. constructed 31 single amino acid mutants within the portal region and in the region of the gap between the βD- and βE-strands of I-FABP, and determined binding affinities and rate constants for FA binding. Together with experiments examining these parameters as a function of ionic strength, it was suggested that the FA initially binds through an electrostatic interaction to Arg56 on the surface of the protein, before inserting into the binding cavity, with a reversal of these steps for the dissociation reaction.

The structural and biochemical studies of FA binding to FABPs do not directly demonstrate a ligand transport function for the proteins, however, they nevertheless, are consistent with their participation in such processes. It has been suggested that upon binding to FABP, the FAs traverse the aqueous cytoplasm in a more energetically favorable manner owing to increased aqueous solubility of FAs, in that lateral diffusion of FABP-bound FAs proceeds more rapidly than that of membrane-bound FAs.

2 *In vitro* Fatty Acid–Transfer Properties of FABPs

The mechanisms by which members of the FABP family transfer FAs to

membranes has been examined in a series of *in vitro* experiments. These studies examined the rate of anthroyloxy-labeled fatty acid (AOFA) movement from FABPs to model acceptor vesicles containing a nonexchangeable fluorescence quencher, using a fluorescence resonance energy transfer assay.

Transfer of fluorescent FA analogues from adipocyte, heart, intestinal, epidermal, brain, and myelin FABPs (A-, H-, I-, E-, B-, and M-FABPs), and of retinol from cellular retinal–binding protein I (CRBP-I), to membrane targets appears to involve collision of the holo protein with both zwitterionic and anionic membranes, with more effective collisional interactions occurring with the latter. FA transfer rates were directly proportional to acceptor membrane phospholipid concentration, and were modulated by changes in the acceptor vesicle charge, implying that FABP and membrane come into direct physical contact, with ligand transfer occurring during the collision. Further studies suggested that positively charged (lysine) residues on the protein surface are involved in this ligand transfer process and are likely to participate in the formation of effective FABP:membrane complexes that involve electrostatic interactions. Site-directed mutagenesis of A-FABP and H-FABP demonstrated that lysines in the helical cap domain are important for establishing these ionic interactions. In particular, lysine residues on αI, αII, and the βCD-turn of both proteins, and possibly the βA-strand of A-FABP, but not H-FABP, were shown to be directly involved in these charge:charge interactions. Indeed, by removing I-FABPs two α-helices, fatty acid transfer no longer occurred via a collisional mechanism. Furthermore, creation of chimeric proteins of A- and H-FABPs revealed that the αII-helix is important in determining the absolute fatty acid transfer rates, while the αI-helix appears to be particularly important in regulating protein sensitivity to the negative charge of membranes. The FABP α-helical domain, part of the same region believed to enable FA entry and exit, is therefore extremely important not only for the direct interaction with membrane phospholipids during ligand transfer to membranes but also in regulating FA transfer rates.

The direct interaction of the FABPs with membranes was further assessed using fluorescence-based assays and direct physical measurements and, as suggested by the transfer experiments, the α-helical domain and surface lysine residues therein proved particularly important for membrane association. Preincubation of anionic vesicles with I-FABP prevented the subsequent binding of the peripheral membrane protein cytochrome c, suggesting that the I-FABP was membrane-bound. In contrast, helix-less I-FABP demonstrated 80% less efficiency in preventing cytochrome c binding than intact I-FABP. Further direct support for I-FABP:membrane interactions was provided by surface pressure measurements, Brewster angle microscopy, and infrared reflection-absorption spectroscopy (IRRAS), which revealed that I-FABP interacted with 1,2-dimyristoyl phosphatidic acid monolayers to a stronger extent than its helix-less variant. IRRAS studies also showed I-FABP to induce a stronger conformational ordering of the lipid acyl chains than helix-less I-FABP. The interaction of A-FABP with vesicles was also directly measured using FTIR spectroscopy, and it was found that A-FABP interacts much more strongly with acidic than zwitterionic membranes, and that neutralization of A-FABP positive surface charges by acetylation considerably weakens its interactions with negatively

charged vesicles. As for I-FABP, these data supported observations gained from the cytochrome c–binding assay, whereby A-FABP, but not acetylated A-FABP, was able to prevent subsequent cytochrome c binding to model anionic membranes. Notably, in all these studies, the degree of membrane interaction correlated directly with the rate of fatty acid transfer, indicating that FABP: membrane interactions are functionally related to their fatty acid transport properties. While the primary mode of FABP:membrane interaction appears to involve the establishment of ion pairs between positive charges on the protein surface and negative membrane charges, recent mutagenesis studies have suggested that hydrophobic interactions between I-FABP helix II residues and membrane phospholipids may also play a role in establishing the collisional interactions that promote FA transfer. Indeed, it has been noted that those FABPs displaying a collisional mechanism for fatty acid transfer, possess a conserved, solvent-exposed, bulky hydrophobic side chain located on αII, namely, a phenylalanine, leucine, isoleucine, or methionine. L-FABP on the other hand, displaying a diffusional mechanism for fatty acid transfer, possesses a Glu in this position. The aforementioned results for I-FABP thus suggest that exposed hydrophobic residues could be involved in membrane association, thereby rendering a subsequent interaction with a receptor protein more efficient and permitting the exchange of ligand without its entry into an aqueous phase.

L-FABP and CRBP-II are the only members of the FABP family that were found not to transfer FA, or retinol in the case of CRBP-II, via a collisional mechanism. L-FABP transfers FAs to membranes almost 50-fold slower than members of the family, exhibiting a collisional FA transfer mechanism. Despite overall structural similarity, this particular FABP appears to transfer its ligand via aqueous diffusion, a mechanism that does not involve direct protein–membrane contact. In support of this mechanism, the rate of FA transfer from L-FABP to membranes was modulated by neither the concentration of acceptor membranes nor their composition; however, changes in the ionic strength of the buffer directly affected the transfer rates, indicating that the transfer rate is regulated by aqueous solubility of the fatty acid. Recently, it was demonstrated that the alpha-helical region of L-FABP is responsible for its diffusional mechanism of fatty acid transfer to membranes. This was deduced by the creation of a pair of chimeric proteins, one containing the ligand-binding domain of I-FABP and the alpha-helical region of L-FABP (alphaLbetaIFABP) and the other containing the ligand-binding pocket of L-FABP and the helical domain of I-FABP (alphaIbetaLFABP). Transfer rates from the chimeric proteins compared to those of the wild type indicated that the slower rate of FA transfer observed for L-FABP relative to that of I-FABP is, in part, determined by the helical domain of the proteins. In addition, absolute FA transfer from alphaLbetaIFABP to membranes occurred by aqueous diffusion whereas FA transfer from alphaIbetaLFABP occurred via protein–membrane collisional interactions. Like L-FABP, CRBP-II also displayed a diffusional mechanism, and indeed proved ineffective in preventing cytochrome c from binding to phospholipid vesicles. Davies et al., however, were able to observe L-FABP binding to anionic phospholipid vesicles using the fluorescent probe DAUDA, an undecanoate (11 : 0) derivative. Upon binding to FABP,

the fluorescence maximum of DAUDA becomes blue- shifted and exhibits a substantial increase in intensity. By mixing a preformed DAUDA:L-FABP complex with anionic vesicles, an immediate decrease in DAUDA fluorescence occurred, indicating a release of DAUDA from the protein. As this did not occur with zwitterionic vesicles, the results suggested that interaction of L-FABP with the anionic membrane interface induces a rapid conformational change, resulting in a reduced affinity of DAUDA for the protein. The nature of this interaction was suggested to involve both electrostatic and nonpolar forces. It is important to note, however, that these L-FABP–membrane interactions were observed at very low ionic strength assay conditions, but were not found at physiologic ionic strength.

From these experiments, it was hypothesized that those FABPs exhibiting a collisional mechanism for FA transfer are most likely to be involved in the targeted transfer of FAs, whereby the proteins interact either with specific membrane lipids and/or membrane protein domains to transfer their fatty acid ligand. Recently, it was also discovered that transfer of anthroyloxylated fatty acids (AOFAs) from phospholipid membranes to I-FABP (in the opposite direction from that initially examined) also occurs via a collisional mechanism. It is possible then that FABPs may utilize membrane–protein interactions not only for the acquisition of ligand but also for their delivery. By displaying a diffusional mechanism for fatty acid transfer whereby no direct protein–membrane interaction occurs, it is likely that L-FABP and CRBP-II may function in the capacity of cytosolic fatty acid or retinol reservoirs. Nonetheless, just as the apparent membrane interactions may be surrogates for FABP–protein interactions, the absence of apparent L-FABP or CRBP-II–membrane interactions does not preclude protein–protein interactions; however, the nature of any such interactions is likely to be different from that of the collisional FABPs.

It is important to note that much of the work examining the mechanisms of FA transfer to model membranes has involved the use of rodent FABPs. With numerous links between FABPs and several chronic diseases in humans becoming ever clearer, a direct comparison of fatty acid transfer properties of rodent and human proteins demonstrated that the rates and mechanisms of FA transfer from L-, I-, A- and H-FABPs observed for rodent proteins were the same for their human counterparts. An additional approach used to examine the transport function of the FABPs is fluorescence recovery after photobleaching (FRAP) for solution conditions. In such studies, the effective diffusion (Deff) of the fluorescent probe *N*-(7-nitro-2,1,3-benzoxadiazol-4-yl)-stearate (NBDS) is evaluated in individual cells or in so-called model cytosol. Using this approach, it has been found that the rate of NBDS movement correlates directly with the intracellular level of L-FABP in HepG2 cells and hepatocytes, the level of I-FABP in embryonic stem cells, the total FABP level (L-FABP + I-FABP) in rat enterocytes isolated from different intestinal segments, and the L-FABP concentration in solutions prepared to resemble cytosol with intracellular membranes. The mechanism for these effects likely involves FABPs acting to limit fatty acid partitioning into immobile membranes, thereby increasing the rate of movement of the fatty acid. Additionally, specific effects of FABPs in cells were also suggested by results in which permeabilized HepG2

cells were used to generate various cytosolic compositions by incubation with different protein-containing solutions; in these studies, albumin was found to be only 4-fold more effective than equal concentrations of L-FABP in increasing the Deff of NBDS, despite the fact that its fatty acid–binding capacity is of far greater magnitude greater than that of L-FABP, as is its FA-binding affinity.

3 Transfection Studies of FABP Function

To assess FABP function within a more physiological milieu, FABP genes have been transfected into model cell cultures and subsequent changes in FA uptake and metabolism examined.

A series of experiments examining L-cell fibroblasts or embryonic stem cells transfected with L-FABP and/or I-FABP appeared to suggest a role for L-FABP in cellular FA uptake; when NBDS or the fluorescent fatty acid analog *cis*-parinaric acid (cPnA) were added to cells expressing L-FABP and compared with control cells, or cells expressing I-FABP, a 2-fold increase in fluorescence intensity was observed. However, care should be taken in the interpretation of these results given that the quantum yield for cPnA binding to L-FABP differs by a similar degree to that of cPnA binding to I-FABP. Indeed, in contrast to these results, it was reported more recently that expression of L-FABP in transfected L-cell fibroblasts did not reduce the uptake of ^{3}H palmitic acid, its oxidation, or induction of lipid accumulation, although uptake of ^{3}H phytanic acid was found to be reduced. Thus, apparently similar studies provide divergent results. More recently, nevertheless, Wolfrum et al. used peroxisome proliferators to increase L-FABP levels in HepG2 cells, and antisense L-FABP to decrease L-FABP mRNA expression, and the net oleate uptake was shown to correlate directly with the L-FABP content of the cells. When I-FABP was expressed at 2-fold higher concentrations in L-cells, it was found that cPnA uptake was lower relative to cells with lesser I-FABP levels. In contrast, the decreased level of I-FABP expression in differentiated relative to undifferentiated embryonic stem cells was also correlated with a decrease in fatty acid uptake. Results of experiments involving overexpression of Ala^{54} and Thr^{54}-I-FABP (two I-FABP forms created by a single base pair alteration in the human I-FABP gene) in Caco-2 intestinal cells must also be viewed carefully. A 2-fold increase in net fatty acid uptake was obtained from Thr^{54}-I-FABP-transfected cells and approximately 5-fold more triacylglycerol was secreted into the basolateral medium relative to Ala^{54}-I-FABP-transfected cells. In differentiated cells of both lines, the endogenous levels of L-FABP were decreased relative to control cells, although L-FABP levels, nevertheless, remained 2- to 3-fold higher than levels of I-FABP. Moreover, it has subsequently been demonstrated that, in contrast to earlier indications, Caco-2 cells do in fact express I-FABP. A recent publication, nevertheless, reported that parent Caco-2 cells as well as mock-transfected cells fail to express detectable levels of I-FABP mRNA or protein at any stage of differentiation; upon transfecting cells with I-FABP, radiolabeled oleic acid was used to monitor fatty acid metabolism, and it was deduced that I-FABP expression in intestinal cells leads to reduced triacylglycerol secretion. Clearly, the precise function of I-FABP in the enterocyte remains uncertain. Given the additional variable of high levels of expression of L-FABP in this

cell type as well, an understanding of the functional properties of enterocyte FABPs, in particular, remains a challenge.

CHO cells transfected with A-FABP showed a 1.5- to 2-fold increase in the net uptake of oleate and, interestingly, expression of equivalent levels of a mutated form of A-FABP with reduced fatty acid binding failed to produce a change in uptake. In contrast, expression of A-FABP in L6 myoblasts did not alter the net fatty acid uptake, although the degree of differentiation in these cells differed from that of control cells. Recently, electroporation of A-FABP into 3T3-L1 preadipocytes was used as an alternative technique for modifying the level of cellular FABP directly. Incorporation of A-FABP was found to result in an increase in the initial rate of palmitate uptake relative to that of control cells, supporting an A-FABP-mediated transport function.

An early report suggested an increase in net fatty acid uptake in cells overexpressing H-FABP, whereas overexpression of H-FABP in L6 myoblasts was reported not to alter uptake. Expression of CRBP-I in Caco-2 cells increased net retinol uptake by about 2-fold; in this clone, a large decrease in endogenous Caco-2 expression of CRBP-II was also found. Results from experiments involving CRABPI knockout ABI cell lines suggested that CRABPI functions to regulate the intracellular concentrations of retinoic acid and to maintain high levels of oxidized retinoic acid metabolites such as 4-oxoretinoic acid within cells.

Thus, some studies have yielded conflicting results, although some have provided strong evidence for an intracellular FABP-mediated transport function. A general concern with stable transfections in cultured cell lines is that parallel alterations may be occurring due to clonal variability, and secondary changes due to the altered expression of a specific gene, in this case an FABP, may also occur. The latter issue is not as serious as the former, as it at least points to the potential involvement of the FABP in a cellular process, if not to its precise role at the molecular level. It would seem of great importance for transfection studies, especially, to demonstrate a dose-dependent functional response to FABP expression. Thus far, only the studies of Wolfrum et al. have been so rigorously performed. The use of direct protein transfer techniques, including streptolysis, electroporation, and lipid- and peptide-based protein transfer reagents, avoid entirely the issue of clonal variability, and in large part the concern about secondary changes in cellular processes. These approaches have not yet been widely applied to the FABPs.

4
Cellular Fatty Acid Transport via FABP–Protein Interactions

As noted above, *in vitro* studies suggested a potential for FABPs to act as targeting proteins, conveying their ligands to particular domains on organellar membranes, and/or to specific protein receptors. Recently, a number of protein–protein interactions involving FABPs have indeed been reported, suggesting that fatty acids may be transported around the cell in a regulated manner.

Using yeast two-hybrid assays, an interaction between A-FABP and hormone-sensitive lipase (HSL) was discovered and further confirmed with experiments including GST-pulldowns and coimmunoprecipitation of the HSL:A-FABP complex. The A-FABP interaction domain of HSL was found to reside in the N-terminal portion of the protein, whereas the catalytic domain is known to be localized to the

C-terminus. Recent studies showed that HSL residues His194 and Glu199 appear to be critical for interactions with A-FABP. Furthermore, it has been shown that A- and E-FABPs, but not I-or L-FABPs, bind to HSL in a 1 : 1 molar stoichiometry only in the presence of oleate. However, A-, E-, L-, H-, and I-FABPs all stimulated HSL activity to an equivalent and relatively modest extent, approximately two-fold. A point mutant of A-FABP which does not display FA binding did not cause this modest stimulation. These results suggest that the effect of the FABPs on HSL is likely via binding of fatty acid but that a protein–protein interaction is not required for the FABP effect. It appears then that binding and activation of HSL by FABPs are separate and distinct functions. Taken as a whole, results from these experiments suggest that A-FABP may function to traffic fatty acids away from the triglyceride droplet after hydrolysis by HSL, thereby promoting further lipolysis by diminishing end-product inhibition. It is still unclear exactly where A-FABP is taking the fatty acid. However, given the evidence of an association between H-FABP and the cytoplasmic domain of the putative transmembrane fatty acid transporter CD36/FAT in milk-fat globule membranes, as determined by gel filtration and coimmunoprecipitation, it is possible that A-FABP transports its bound fatty acid to the plasma membrane and, via an interaction with CD36/FAT, promotes the efflux of the fatty acid out of the cell (Fig. 2). Alternatively, in the adipocyte or in other cell types, an FABP could transport its bound fatty acid to internal sites for re-esterification.

Again using the yeast two-hybrid system, an interaction between L-FABP and the lipid-activated transcription factor

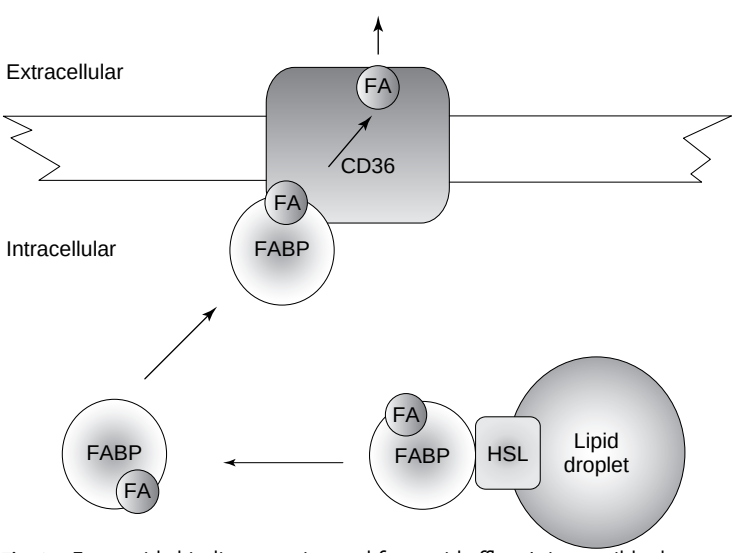

Fig. 2 Fatty acid–binding proteins and fatty acid efflux. It is possible that upon hydrolysis of lipid by hormone sensitive lipase (HSL) or other lipases, the resulting fatty acid (FA) is bound by fatty acid–binding protein (FABP) and transported through the cytoplasm to CD36/FAT, whereupon the fatty acid is off-loaded for efflux out of the cell. Adipocyte FABP-HSL and heart FABP-CD36 complexes have been identified.

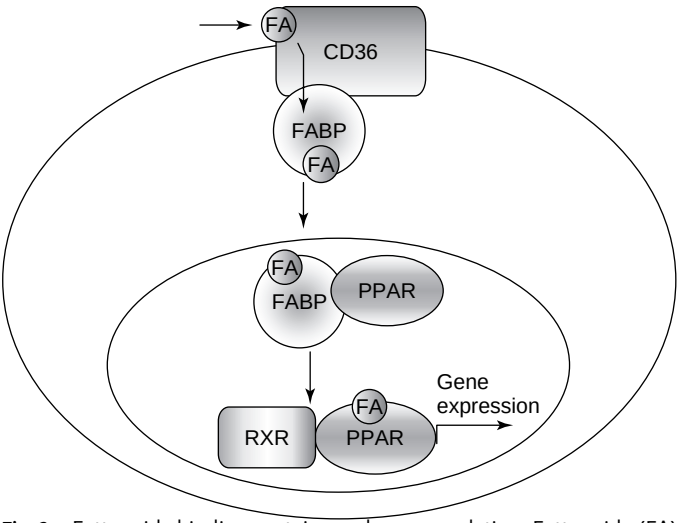

Fig. 3 Fatty acid–binding proteins and gene regulation. Fatty acids (FA) taken up by CD36/FAT may then be bound in the intracellular space by fatty acid–binding proteins (FABP) for transport through the cytoplasm to the nucleus. Upon entering the nucleus, the FABP:FA complex binds to PPAR, the FA becomes PPAR-bound, and gene expression is initiated. CD36 has been shown to play a role in fatty acid uptake, FABPCD36 complexes have been identified, and an association between liver FABP and PPARa has been shown. There is also evidence to suggest that adipocyte and epidermal FABPs may be partially localized to the nucleus.

peroxisome proliferator activated receptor α (PPARα) was found. This was further assessed by pull-down assays and immunoprecipitation, and was shown to be independent of ligand binding. PPARα is believed to be a nuclear target for fatty acids and initiates gene expression of enzymes involved in lipid metabolism. Such an association, therefore, suggests that L-FABP serves to directly traffic its fatty acid ligand, gained possibly from an interaction with CD36/FAT and/or other transmembrane transporters, to the nucleus and, thereby, directly functions in the regulation of gene expression (Fig. 3). Indeed, recent results indicate that A- and E-FABPs localize in the nuclei of 3T3-LI adipocytes as well as the cytoplasm, suggesting that these two FABPs may also exert their action at the level of the nucleus.

An interaction between the extracellular domain of CD36 and the S100A8/S100A9: arachidonic acid complex has also been identified in endothelial cells, implying a role for this protein in transcellular eicosanoid metabolism. Given that the coordinate regulation of gene expression and similar abundance of cytoplasmic FABPs and membrane FA transporters have been repeatedly demonstrated, it is highly possible that a concerted, vectorial mechanism for FA transport exists.

5 Insights into FABP Function from Null Mice

Several mouse models null for different FABPs have been created by targeted gene disruption. These are providing important support for the transport functions

of these proteins, as well as new insights into additional potential functions. Of the four FABP knockouts described thus far, the H-FABP-null mouse provides perhaps the clearest example of functional consequences. Mice lacking expression of H-FABP displayed a substantial decrease in long-chain fatty acid uptake into the heart. Further studies in cardiac myocytes isolated from wild-type and H-FABP$^{-/-}$ animals demonstrated that defective fatty acid uptake and oxidation appears to be the underlying cause of the phenotype observed in the whole animal. Interestingly, the CD36/FAT-null mouse displays a very similar FA uptake phenotype, again suggesting a concerted action by the two types of FA transport proteins in cellular FA trafficking. In the H-FABP-null animals, as well as in the CD36 nulls, physiological compensation for the decreased FA uptake appears to occur by an increase in glucose uptake and oxidation in the heart, and not by compensatory increases in other members of the FABP family. H-FABP is also expressed in the mammary gland; however, the H-FABP$^{-/-}$ mice were reported to have no overt phenotype in this tissue.

Mice null for I-FABP were also reported to show no apparent compensation with other FABPs. The I-FABP$^{-/-}$ mice developed hyperinsulinemia that was independent of body weight gain, an unusual dissociation. The I-FABP knockout mice gained more weight and had higher levels of serum triglycerides. This could indicate an involvement of the protein in lipid absorption, metabolism, or secretion, which requires further investigation. An intriguing finding in this model was that the effects on weight and serum triglycerides were observed only in male mice. This gender dependency suggests a previously unexplored interaction between I-FABP and sex hormones.

In contrast to the H-FABP and I-FABP nulls, mice null for the ap2 gene, which encodes A-FABP, showed a dramatic increase in expression of another FABP, keratinocyte FABP (K-FABP), in adipose tissue, perhaps accounting for the absence of a dramatic phenotype in the animals. Initial investigations of low fat–fed mice showed few differences between wild-type and A-FABP$^{-/-}$ animals; however, feeding a high-fat diet resulted in lower levels of plasma insulin and reduced adipocyte mRNA levels of tumor necrosis factor α (TNFα) relative to wild-type mice; the absence of hyperinsulinemia appeared to occur despite the presence of high-fat diet-induced obesity. Conversely, however, it was found that younger aP2$^{-/-}$ mice, despite maintaining lower glucose levels, did in fact develop hyperinsulinemia on a high-fat diet; the plasma insulin levels were directly correlated with the degree of adiposity in both wild-type and A-FABP$^{-/-}$ mice. Further, adipocyte TNFα secretion was not reduced relative to wild-type mice. These results indicate that a dissociation between the development of obesity and the development of hyperinsulinemia, a hallmark of obesity, is not apparent in the A-FABP$^{-/-}$ mice under all circumstances.

The aP2$^{-/-}$ mice have also been reported to have modest decreases in lipolysis in some, but not all, investigations, as well as a small increase in basal levels of *de novo* fatty acid synthesis. Interestingly, the pancreatic insulin secretory response to β-adrenergic stimulation was suppressed in aP2$^{-/-}$ mice, and when aP2 deficiency was introduced into the genetically obese ob/ob mouse, animals lacking A-FABP showed decreased pancreatic insulin secretion as well as improved glucose tolerance. Thus, despite the discrepancies and despite the fact that the mechanisms of the

null phenotypic changes are not clear, the collective results indicate that further explorations of the role of A-FABP in fatty acid flux and systemic lipid and glucose metabolism are warranted. A recent report of dramatic increases in serum levels and adipose expression of a bone morphogenic protein in the A-FABP-null mice may point to a heretofore unrecognized interaction between fatty acid metabolism and bone development. At this point, however, the nature of any such association remains to be explored.

A potentially critical involvement for A-FABP in the development of diet-induced atherosclerosis has been recently revealed by studies of A-FABP$^{-/-}$ mice crossed with mice deficient in apoE, the latter being a well-established model of dietary atherosclerosis. ApoE$^{-/-}$ animals develop severe coronary arterial occlusion on a high-fat diet; in dramatic contrast, the A-FABP$^{-/-}$ /apoE$^{-/-}$ mice developed only trivial lesions, strongly indicating a role for A-FABP in the accumulation of lipid-rich foam cells in the arterial intima. As A-FABP is highly expressed not only in adipose tissue but also in macrophages, the results suggested that it plays a critical role in the development of hypercholesterolemia-induced atherosclerosis, likely at the level of the macrophage. Indeed, recent bone marrow transplantation studies showed that macrophage-expressed A-FABP, rather than adipocyte A-FABP, was likely to be primarily involved in the development of dietary atherosclerosis. K-FABP was also found to be expressed in macrophages; however, unlike the compensatory upregulation of expression observed in adipose tissue of the ap2$^{-/-}$ mice, macrophage K-FABP expression remained unchanged. The mechanisms for the apparently proatherosclerotic effects of macrophage A-FABP do not appear to be primarily related to cholesterol ester accumulation, as this was altered to a modest extent, if at all, in macrophages from the A-FABP$^{-/-}$ /apoE$^{-/-}$ mice. Interestingly, levels of interleukin 6, interleukin 1β, macrophage inflammatory proteins 1α and 1β, and macrophage chemoattractant protein 1 were decreased in the apoE/A-FABP-deficient macrophages relative to apoE$^{-/-}$, indicating a role for A-FABP in inflammatory cytokine and chemokine expression.

Recently, the generation of a skin-type FABP (K- or E-FABP)-null mouse (E-FABP$^{-/-}$) was reported. Here, too, it appears that compensatory changes in another FABP are found: mice lacking E-FABP showed increased expression of H-FABP in skin. A dramatic phenotype was not observed in the E-FABP$^{-/-}$ mice; however, changes in the rate, although not the extent of transepidermal water loss, were observed. The mechanisms by which the FABPs may participate in the water barrier function of skin are not yet known. Recently, the construction of transgenic mice overexpressing the *FABP5* gene encoding E-FABP in adipocytes was reported. In adipocytes from FABP5 transgenic mice, the total FABPprotein level was found to increase 150% as compared to the wild type due to a 10-fold increase in the level of E-FABP and an unanticipated 2-fold downregulation of the A-FABP. These results suggested that there is a positive relationship between lipolysis and the total level of FABP, but not between lipolysis and a specific FABP type.

L-FABP knockout mice displayed a dramatically reduced cytosolic fatty acid–binding capacity in the mouse liver. Moreover, a dramatic shift in liver lipid distribution in favor of cholesterol, cholesterol esters, and phospholipids was observed. L-FABP, sterol carrier protein 2

(SCP-2) and SCP-x may contribute to this change in liver lipid distribution; levels of SCP-2 (a known cholesterol-binding protein) increased by approximately 75% while levels of SCP-x decreased in L-FABP null mice. As yet, however, the precise mechanism and functional consequences of the shift in hepatocellular lipid distribution in L-FABP null mice are unknown. It is also worth noting that in an independently generated L-FABP$^{-/-}$ mouse, no changes in SCP-2 levels were detected. In other studies the role of L-FABP in peroxisomal oxidation and metabolism of branched-chain fatty acids was investigated using cultured primary hepatocytes isolated from livers of L-FABP null and wild-type mice. L-FABP gene ablation reduced maximal, but not initial, uptake of phytanic acid (the most common branched-chain fatty acid), and inhibited phytanic acid peroxisomal oxidation and microsomal esterification. L-FABP gene-ablated hepatocytes were also found to contain decreased levels of free fatty acids and triglyceride. L-FABP null mice have been further used to test the hypothesis that L-FABP limits the availability of long-chain fatty acids for oxidation and for PPARα, a fatty acid–binding transcription factor that determines the capacity of hepatic fatty acid oxidation. Indeed, it was found that the mechanisms whereby L-FABP affects fatty acid oxidation may vary with physiological condition; under fasting conditions, hepatic L-FABP appears to contribute to hepatic LCFA oxidation and ketogenesis by a nontranscriptional mechanism, whereas L-FABP appears to activate ketogenic gene expression in fed mice. Independently of these experiments Newberry et al. generated mice with a targeted deletion of the endogenous L-FABP gene and have characterized their response to alterations in hepatic fatty acid flux following prolonged fasting. The resulting data point to an inducible defect in fatty acid utilization in fasted L-FABP null mice that involves targeting of substrate for use in triglyceride metabolism.

6 Perspectives

It has often been suggested that the expression of more than one type of FABP in a single tissue or even a single cell type, is a strong predictor that FABPs perform functions other than or in addition to bulk binding and transport. Recent studies of Widstrom et al., using displacement of the fluorescent probe ANS, showed that H-FABP can bind arachidonic and linoleic acid metabolites, which serve as components of cell signaling cascades. If the FABPs do in fact act to target as well as bind FAs, as the data reviewed here suggest, then there must be specific signals for protein trafficking. An interesting feature of A-FABP is that Phe57 on the β-CD turn rotates by more than 90° into the binding cavity of the apoprotein relative to the holoprotein. As discussed earlier, differences in apo-FABP versus holo-FABP tertiary structures are especially notable in the helix-turn-helix portal domain. Differences in the regulation of fluorescent FA transfer from I-FABP to membranes relative to transfer in the opposite direction further suggests differential trafficking properties of apo- and holo-FABPs. Thus, structural changes in apo- versus holo-FABPs might act to alter interactions with other cellular structures, prevent competition between apo- and holo-proteins for interaction with the same receptor, act as a signal for a ligand-bound protein, or change FABP affinity for a membrane. Evidently, there is much to be learned about the precise

mechanisms by which FABPs participate in the uptake, efflux, and intracellular transport and metabolism of their small hydrophobic ligands.

See also Annexins; Membrane Traffic: Vesicle Budding and Fusion; Membrane Transport.

Bibliography

Books and Reviews

Glatz, J.F.C., Luiken, J.J.F.P., van Bilsen, M., van der Vusse, G.J. (2002) Cellular lipid binding proteins as facilitators and regulators of lipid metabolism, *Mol. Cell. Biochem.* **239**, 3–7.

Hamilton, J.A. (2002) Fatty acids bind to proteins: the inside story from protein structures, *Prostaglandins Leukot. Essent. Fatty Acids* **67**, 65–72.

Hamilton, J.A. (2004) Fatty acid interactions with proteins: what X-ray crystal and NMR solution structures tell us, *Prog. Lipid Res.* **43**, 177–199.

Hanhoff, T., Lucke, C., Spener, F. (2002) Insights into binding of fatty acids by fatty acid binding proteins, *Mol. Cell. Biochem.* **239**, 45–54.

Van der Vusse, G.J., van Bilsen, M., Glatz, J.F.C., Hasselbaink, D.M., Luiken, J.J.F.P. (2002) Critical steps in cellular fatty acid uptake and utilization, *Mol. Cell. Biochem.* **239**, 35–43.

Zimmerman, A.W., Veerkamp, J.H. (2002) New insights into the structure and function of fatty acid binding proteins, *Cell. Mol. Life Sci.* **59**, 1096–1116.

Primary Literature

Abumrad, N., Coburn, C., Ibrahimi, A. (1999) Membrane proteins implicated in long-chain fatty acid uptake by mammalian cells: CD36, FATP and FABPm, *Biochim. Biophys. Acta* **1441**, 4–13.

Alpers, D.H., Strauss, A.W., Ockner, R.K., Bass, N.M., Gordon, J.I. (1984) Cloning of a cDNA encoding rat intestinal fatty acid binding protein, *Proc. Natl. Acad. Sci. U. S. A.* **81**, 313–317.

Atshaves, B.P., Foxworth, W.B., Frolov, A., Roths, J.B., Kier, A.B., Oetama, B.K., et al. (1998) Cellular differentiation and I-FABP protein expression modulate fatty acid uptake and diffusion, *Am. J. Physiol.* **274**, C633–C644.

Atshaves, B.P., McIntosh, A.M., Lyuksyutova, O.I., Zipfel, W., Webb, W.W., Schroeder, F. (2004) Liver fatty-acid binding protein gene ablation inhibits branched-chain fatty acid metabolism in cultured primary hepatocytes, *J. Biol. Chem.* **279**, 30954–30965.

Atshaves, B.P., Storey, S.M., Petrescu, A., Greenberg, C.C., Lyuksyutova, O.I., Smith, R., Schroeder, F. (2002) Expression of fatty acid binding proteins inhibits lipid accumulation and alters toxicity in L-cell fibroblasts, *Am. J. Physiol. Cell Physiol.* **283**, 688–703.

Baier, L.J., Bogardus, C., Sacchettini, J.C. (1996) A polymorphism in the human intestinal fatty acid binding protein alters fatty acid transport across Caco-2 cells, *J. Biol. Chem.* **271**, 10892–10896.

Balendiran, G.K., Schnutgen, F., Scapin, G., Borchers, T., Xhong, N., Lim, K., et al. (2000) Crystal structure and thermodynamic analysis of human brain fatty acid- binding protein, *J. Biol. Chem.* **275**, 27045–27054.

Bernlohr, D.A., Angus, C.W., Lane, M.D., Bolanowski, M.A., Kelly, T.J. Jr. (1984) Expression of specific mRNAs during adipose differentiation: identification of an mRNA encoding a homologue of myelin P2 protein, *Proc. Natl. Acad. Sci. U. S. A.* **81**, 5468–5472.

Binas, B., Danneberg, H., McWhir, J., Mullins, L., Clark, A.J. (1999) Requirement for the heart-type fatty acid binding protein in cardiac fatty acid utilization, *FASEB J.* **13**, 805–812.

Chen, A.C., Yu, K., Lane, M.A., Gudas, L.J. (2003) Homozygous deletion of the CRABPI gene in AB1 embryonic stem cells results in increased CRABPII gene expression and decreased intracellular retinoic acid concentration, *Arch. Biochem. Biophys.* **411**, 159–173.

Claffey, K.P., Crisman, T.S., Ruiz-Opazo, N., Brecher, P. (1988) Expression of rat heart fatty acid binding protein cDNA in rat L6 myoblasts, *FASEB J.*, **2**, A1783.

Clark, A.J., Neil, C., Gusterson, B., McWhir, J., Binas, B. (2000) Deletion of the gene encoding H-FABP/MDGI has no overt effects in the mammary gland, *Transgenic Res.* **9**, 439–444.

Coburn, C.T., Knapp, F.F., Febbraio, M., Beets, A.L., Silverstein, R.L., Abumrad, N.A. Jr.

(2000) Defective uptake and utilization of long chain fatty acids in muscle and adipose tissues of CD36 knockout mice, *J. Biol. Chem.* **275**, 32523–32529.

Coe, N.R., Bernlohr, D.A. (1998) Physiological properties and functions of intracellular fatty acid-binding proteins, *Biochim. Biophys. Acta* **1391**, 287–306.

Coe, N.R., Simpson, M.A., Bernlohr, D.A. (1999) Targeted disruption of the adipocyte lipid-binding protein (aP2 protein) gene impairs fat cell lipolysis and increases cellular fatty acid levels, *J. Lipid Res.* **40**, 967–972.

Corsico, B., Liou, H.L., Storch, J. (2004) The alpha-helical domain of liver fatty acid bindingprotein is responsible for the diffusion-mediated transfer of fatty acids to phospholipid membranes, *Biochemistry* **43**, 3600–3607.

Corsico, B., Cistola, D.P., Frieden, C., Storch, J. (1998) The helical domain of intestinal fatty acid binding protein is critical for collisional transfer of fatty acids to phospholipid membranes, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 12174–12178.

Darimont, C., Gradoux, N., Cumin, F., Baum, H.P., De Pover, A. (1998) Differential regulation of intestinal and liver fatty acid-binding proteins in human intestinal cell line (Caco-2): role of collagen, *Exp. Cell Res.* **244**, 441–447.

Davies, J.K., Thumser, A.E., Wilton, D.C. (1999) Binding of recombinant rat liver fatty acid-binding protein to small anionic phospholipid vesicles results in ligand release: a model for interfacial binding and fatty acid targeting, *Biochemistry* **38**, 16932–16940.

Eads, J., Sacchettini, J.C., Kromminga, A., Gordon, J.I. (1993) Escherichia coli-derived rat intestinal fatty acid binding protein with bound myristate at 1.5 A resolution and I-FABPArg106→Gln with bound oleate at 1.74 A resolution, *J. Biol. Chem.* **268**, 26375–26385.

Ellinghaus, P., Wolfrum, C., Assmann, G., Spener, F., Seedorf, U. (1999) Phytanic acid activates the peroxisome proliferator-activated receptor alpha (PPARalpha) in sterol carrier protein 2-/ sterol carrier protein x-deficient mice, *J. Biol. Chem.* **274**, 2766–2772.

Erol, E., Kumar, L.S., Cline, G.W., Shulman, G.I., Kelly, D.P., Binas, B. (2004) Liver fatty acid binding protein is required for high rates of hepatic fatty acid oxidation but not for the action of PPARalpha in fasting mice, *FASEB J.* **18**, 347–349.

Feng, L., Hatten, M.E., Heintz, N. (1994) Brain lipid-binding protein (BLBP): a novel signaling system in the developing mammalian CNS, *Neuron* **12**, 895–908.

Fraser, H., Coles, S.M., Woodford, J.K., Frolov, A.A., Murphy, E.J., Schroeder, F., Bernlohr, D.A., Grund, V. (1997) Fatty acid uptake in diabetic rat adipocytes, *Mol. Cell. Biochem.* **167**, 51–60.

Gedde-Dahl, A., Kulseth, M.A., Ranheim, T., Drevon, C.A., Rustan, A.C. (2002) Reduced secretion of triacylglycerol in Caco-2 cells transfected with intestinal fatty acid binding protein, *Lipids* **37**, 61–68.

Gericke, A., Smith, E.R., Moore, D.J., Mendelsohn, R., Storch, J. (1997) Adipocyte fatty acid-binding protein: interaction with phospholipid membranes and thermal stability studied by FTIR spectroscopy, *Biochemistry* **36**, 8311–8317.

Giguere, V., Lyn, S., Yip, P., Siu, C.H., Amin, S. (1990) Molecular cloning of cDNA encoding a second cellular retinoic acid- binding protein, *Proc. Natl. Acad. Sci. U. S. A.* **87**, 6233–6237.

Glatz, J.F., van der Vusse, G.J. (1996) Cellular fatty acid-binding proteins: their function and physiological significance, *Prog. Lipid Res.* **35**, 243–282.

Gordon, J.I., Alpers, D.H., Ockner, R.K., Strauss, A.W. (1983) The nucleotide sequence of rat liver fatty acid binding protein mRNA, *J. Biol. Chem.* **258**, 3356–3363.

Helledie, T., Antonius, M., Sorensen, R.V., Hertzel, A.V., Bernlohr, D.A., Kolvraa, S., et al. (2000) Lipid-binding proteins modulate ligand-dependent trans-activation by peroxisome proliferator-activated receptors and localize to the nucleus as well as the cytoplasm, *J. Lipid Res.* **41**, 1740–1751.

Herr, F.M., Aronson, J., Storch, J. (1996) Role of portal region lysine residues in electrostatic interactions between heart fatty acid binding protein and phospholipid membranes, *Biochemistry* **35**, 1296–1303.

Herr, F.M., Matarese, V., Bernlohr, D.A., Storch, J. (1995) Surface lysine residues modulate the collisional transfer of fatty acid from adipocyte fatty acid binding protein to membranes, *Biochemistry* **34**, 11840–11845.

Herr, F.M., Li, E., Weinberg, R.B., Cook, V.R., Storch, J. (1999) Differential mechanisms of retinoid transfer from cellular retinol-binding proteins types I and II to phospholipid membranes, *J. Biol. Chem.* **274**, 9556–9563.

Hertzel, A.V., Bennaars_eiden, A., Bernlohr, D.A. (2002) Increased lipolysis on transgenic animals overexpression the epithelial fatty acid binding protein in adipose cells, *J. Lipid Res.* **43**, 2105–2111.

Hodsdon, M.E., Cistola, D.P. (1997) Discrete backbone disorder in the nuclear magnetic resonance structure of apo intestinal fatty acid-binding protein: implications for the mechanism of ligand entry, *Biochemistry* **36**, 1450–1460.

Hotamisligil, G.S., Johnson, R.S., Distel, R.J., Ellis, R., Papaioannou, V.E., Spiegelman, B.M. (1996) Uncoupling of obesity from insulin resistance through a targeted mutation in aP2, the adipocyte fatty acid binding protein, *Science* **274**, 1377–1379.

Hsu, K.T., Storch, J. (1996) Fatty acid transfer from liver and intestinal fatty acid-binding proteins to membranes occurs by different mechanisms, *J. Biol. Chem.* **271**, 13317–13323.

Issemann, I., Green, S. (1990) Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators, *Nature* **347**, 645–650.

Jefferson, J.R., Powell, D.M., Rymaszewski, Z., Kukowska-Latallo, J., Lowe, J.B., Schroeder, F. (1990) Altered membrane structure in transfected mouse L-cell fibroblasts expressing rat liver fatty acid-binding protein, *J. Biol. Chem.* **265**, 11062–11068.

Jefferson, J.R., Slotte, J.P., Nemecz, G., Pastuszyn, A., Scallen, T.J., Schroeder, F. (1991) Intracellular sterol distribution in transfected mouse L-cell fibroblasts expressing rat liver fatty acid-binding protein, *J. Biol. Chem.* **266**, 5486–5496.

Jenkins, A.E., Hockenberry, J.A., Nguyen, T., Bernlohr, D.A. (2002) Testing of the portal hypothesis: analysis of a V32G, F57G, K58G mutant of the fatty acid binding protein of the murine adipocyte, *Biochemistry* **41**, 2022–2027.

Jenkins-Krutchen, A.E., Bennaars-Eiden, A., Ross, J.R., Shen, W.J., Kraemer, F.B., Bernlohr, D.A. (2003) Fatty acid-binding protein-hormone-sensitive lipase interaction. Fatty acid dependence on binding, *J. Biol. Chem.* **278**, 47636–47643.

Kennedy, M.W., Beauchamp, J. (2000) Sticky-finger interaction sites on cytosolic lipid-binding proteins? *Cell. Mol. Life Sci.* **57**, 1379–1387.

Kerkhoff, C., Sorg, C., Tandon, N.N., Nacken, W. (2001) Interaction of S100A8/S100A9-arachidonic acid complexes with the scavenger receptor CD36 may facilitate fatty acid uptake by endothelial cells, *Biochemistry* **40**, 241–248.

Kim, H.K., Storch, J. (1992a) Free fatty acid transfer from rat liver fatty acid-binding protein to phospholipid vesicles. Effect of ligand and solution properties, *J. Biol. Chem.* **267**, 77–82.

Kim, H.K., Storch, J. (1992b) Mechanism of free fatty acid transfer from rat heart fatty acid-binding protein to phospholipid membranes. Evidence for a collisional process, *J. Biol. Chem.* **267**, 20051–20056.

Kodukula, S., Corsico, B., Storch, J. (2002) Role of Polar Residues in the Helical Domain of Enterocyte Fatty Acid Binding Proteins (FABP) in the mechanism of Fatty Acid (FA) Transfer to Phospholipid (PL) Membranes, *FASEB J.*

Krey, G., Braissant, O., L'Horset, F., Kalkhoven, E., Perroud, M., Parker, M.G., Wahli, W. (1997) Fatty acids, eicosanoids, and hypolipidemic agents identified as ligands of peroxisome proliferator-activated receptors by coactivator- dependent receptor ligand assay, *Mol. Endocrinol.* **11**, 779–791.

Krieg, P., Feil, S., Furstenberger, G., Bowden, G.T. (1993) Tumor-specific overexpression of a novel keratinocyte lipid-binding protein. Identification and characterization of a cloned sequence activated during multistage carcinogenesis in mouse skin, *J. Biol. Chem.* **268**, 17362–17369.

Layne, M.D., Patel, A., Chen, Y.H., Rebel, V.I., Carvajal, I.M., Pellacani, A., et al. (2001) Role of macrophage-expressed adipocyte fatty acid binding protein in the development of accelerated atherosclerosis in hypercholesterolemic mice, *FASEB J.* **15**, 2733–2735.

Le Beyec, J., Delers, F., Jourdant, F., Schreider, C., Chambaz, J., Cardot, P., Pincon-Raymond, M. (1997) A complete epithelial organization of Caco-2 cells induces I-FABP and potentializes apolipoprotein gene expression, *Exp. Cell Res.* **236**, 311–320.

Levin, M.S. (1993) Cellular retinol-binding proteins are determinants of retinol uptake and metabolism in stably transfected Caco-2 cells, *J. Biol. Chem.* **268**, 8267–8276.

Li, E., Demmer, L.A., Sweetser, D.A., Ong, D.E., Gordon, J.I. (1986) Rat cellular retinol-binding protein II: use of a cloned cDNA to define its

primary structure, tissue-specific expression, and developmental regulation, *Proc. Natl. Acad. Sci. U. S. A.* **83**, 5779–5783.

Liou, H.L., Storch, J. (2001a) Role of surface lysine residues of adipocyte fatty acid-binding protein in fatty acid transfer to phospholipid vesicles, *Biochemistry* **40**, 6475–6485.

Liou, H.L., Storch, J. (2001b) The role of adipocyte fatty acid binding protein (AFABP) in fatty acud uptake in 3T3-LI cells: altering cellular AFABP levels by electroporation [abstract], *FASEB J.* **15**, A1090.

Liou, H.L., Kahn, P.C., Storch, J. (2002) Role of the helical domain in fatty acid transfer from adipocyte and heart fatty acid-binding proteins to membranes: analysis of chimeric proteins, *J. Biol. Chem.* **277**, 1806–1815.

Luxon, B.A. (1996) Inhibition of binding to fatty acid binding protein reduces the intracellular transport of fatty acids, *Am. J. Physiol.* **271**, G113–G120.

Luxon, B.A., Weisiger, R.A. (1993) Sex differences in intracellular fatty acid transport: role of cytoplasmic binding proteins, *Am. J. Physiol.* **265**, G831–G841.

Luxon, B.A., Milliano, M.T. (1997) Cytoplasmic codiffusion of fatty acids is not specific for fatty acid binding protein, *Am. J. Physiol.* **273**, C859–C867.

Luxon, B.A., Milliano, M.T. (1999) Cytoplasmic transport of fatty acids in rat enterocytes: role of binding to fatty acid-binding protein, *Am. J. Physiol.* **277**, G361–G366.

Makowski, L., Boord, J.B., Maeda, K., Babaev, V.R., Uysal, K.T., Morgan, M.A., et al. (2001) Lack of macrophage fatty-acid-binding protein aP2 protects mice deficient in apolipoprotein E against atherosclerosis, *Nat. Med.* **7**, 699–705.

Martin, G.G., Danneberg, H., Kumar, L.S., Atshaves, B.P., Erol, E., Bader, M., Schroeder, F., Binas, B. (2003) Decreased liver fatty acid binding capacity and altered liver lipid distribution in mice lacking the liver fatty acid-binding protein gene, *J. Biol. Chem.* **278**, 21429–21438.

McCullagh, K.J., Juel, C., O'Brien, M., Bonen, A. (1996) Chronic muscle stimulation increases lactate transport in rat skeletal muscle, *Mol. Cell. Biochem.* **156**, 51–57.

Murphy, E.J. (1998) L-FABP and I-FABP expression increase NBD-stearate uptake and cytoplasmic diffusion in L cells, *Am. J. Physiol.* **275**, G244–G249.

Murphy, E.J., Prows, D.R., Jefferson, J.R., Schroeder, F. (1996) Liver fatty acid-binding protein expression in transfected fibroblasts stimulates fatty acid uptake and metabolism, *Biochim. Biophys. Acta* **1301**, 191–198.

Nemecz, G., Hubbell, T., Jefferson, J.R., Lowe, J.B., Schroeder, F. (1991) Interaction of fatty acids with recombinant rat intestinal and liver fatty acid-binding proteins, *Arch. Biochem. Biophys.* **286**, 300–309.

Newberry, E.P., Xie, Y., Kennedy, S., Han, X., Buhman, K.K., Luo, J., Gross, R.W., Davidson, N.O. (2003) Decreased hepatic triglyceride accumulation and altered fatty acid uptake in mice with deletion of the liver fatty-acid binding protein gene, *J. Biol. Chem.* **278**, 51664–51672.

Owada, Y., Takano, H., Yamanaka, H., Kobayashi, H., Sugitani, Y., Tomioka, Y., et al. (2002) Altered water barrier function in epidermal-type fatty acid binding protein-deficient mice, *J. Invest. Dermatol.* **118**, 430–435.

Pelsers, M.M., Lutgerink, J.T., Nieuwenhoven, F.A., Tandon, N.N., van der Vusse, G.J., Arends, J.W., et al. (1999) A sensitive immunoassay for rat fatty acid translocase (CD36) using phage antibodies selected on cell transfectants: abundant presence of fatty acid translocase/CD36 in cardiac and red skeletal muscle and up- regulation in diabetes, *Biochem. J.* **337**(3), 407–414.

Pelton, P.D., Zhou, L., Demarest, K.T., Burris, T.P. (1999) PPARgamma activation induces the expression of the adipocyte fatty acid binding protein gene in human monocytes, *Biochem. Biophys. Res. Commun.* **261**, 456–458.

Prinsen, C.F., Veerkamp, J.H. (1998) Transfection of L6 myoblasts with adipocyte fatty acid-binding protein cDNA does not affect fatty acid uptake but disturbs lipid metabolism and fusion, *Biochem. J.* **329**(2), 265–273.

Prows, D.R., Schroeder, F. (1997) Metallothionein-IIA promoter induction alters rat intestinal fatty acid binding protein expression, fatty acid uptake, and lipid metabolism in transfected L-cells, *Arch. Biochem. Biophys.* **340**, 135–143.

Prows, D.R., Murphy, E.J., Schroeder, F. (1995) Intestinal and liver fatty acid binding proteins differentially affect fatty acid uptake and esterification in L-cells, *Lipids* **30**, 907–910.

Repa, J.J., Mangelsdorf, D.J. (2000) The role of orphan nuclear receptors in the regulation of

cholesterol homeostasis, *Annu. Rev. Cell Dev. Biol.* **16**, 459–481.

Richieri, G.V., Ogata, R.T., Kleinfeld, A.M. (1994) Equilibrium constants for the binding of fatty acids with fatty acid- binding proteins from adipocyte, intestine, heart, and liver measured with the fluorescent probe ADIFAB, *J. Biol. Chem.* **269**, 23918–23930.

Richieri, G.V., Low, P.J., Ogata, R.T., Kleinfeld, A.M. (1999) Binding kinetics of engineered mutants provide insight about the pathway for entering and exiting the intestinal fatty acid binding protein, *Biochemistry* **38**, 5888–5895.

Richieri, G.V., Ogata, R.T., Zimmerman, A.W., Veerkamp, J.H., Kleinfeld, A.M. (2000) Fatty acid binding proteins from different tissues show distinct patterns of fatty acid interactions, *Biochemistry* **39**, 7197–7204.

Rolf, B., Oudenampsen-Kruger, E., Borchers, T., Faergeman, N.J., Knudsen, J., Lezius, A., Spener, F. (1995) Analysis of the ligand binding properties of recombinant bovine liver- type fatty acid binding protein, *Biochim. Biophys. Acta* **1259**, 245–253.

Sacchettini, J.C., Said, B., Schulz, H., Gordon, J.I. (1986) Rat heart fatty acid-binding protein is highly homologous to the murine adipocyte 422 protein and the P2 protein of peripheral nerve myelin, *J. Biol. Chem.* **261**, 8218–8223.

Schaap, F.G., Binas, B., Danneberg, H., van der Vusse, G.J., Glatz, J.F. (1999) Impaired long-chain fatty acid utilization by cardiac myocytes isolated from mice lacking the heart-type fatty acid binding protein gene, *Circ. Res.* **85**, 329–337.

Scheja, L., Makowski, L., Uysal, K.T., Wiesbrock, S.M., Shimshek, D.R., Meyers, D.S., et al. (1999) Altered insulin secretion associated with reduced lipolytic efficiency in aP2−/− mice, *Diabetes* **48**, 1987–1994.

Schroeder, F., Jefferson, J.R., Powell, D., Incerpi, S., Woodford, J.K., Colles, S.M., et al. (1993) Expression of rat L-FABP in mouse fibroblasts: role in fat absorption, *Mol. Cell. Biochem.* **123**, 73–83.

Sha, R.S., Kane, C.D., Xu, Z., Banaszak, L.J., Bernlohr, D.A. (1993) Modulation of ligand binding affinity of the adipocyte lipid-binding protein by selective mutation. Analysis in vitro and in situ, *J. Biol. Chem* **268**, 7885–7892.

Shaughnessy, S., Smith, E.R., Kodukula, S., Storch, J., Fried, S.K. (2000) Adipocyte metabolism in adipocyte fatty acid binding protein knockout mice (aP2−/−) after short-term high-fat feeding: functional compensation by the keratinocyte [correction of keratinocyte] fatty acid binding protein, *Diabetes* **49**, 904–911.

Shen, W.J., Sridhar, K., Bernlohr, D.A., Kraemer, F.B. (1999) Interaction of rat hormone-sensitive lipase with adipocyte lipid- binding protein, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 5528–5532.

Shen, W.J., Liang, Y., Hong, R., Patel, S., Natu, V., Sridhar, K., et al. (2001) Characterization of the functional interaction of adipocyte lipid- binding protein with hormone-sensitive lipase, *J. Biol. Chem.* **276**, 49443–49448.

Smith, E.R., Storch, J. (1999) The adipocyte fatty acid-binding protein binds to membranes by electrostatic interactions, *J. Biol. Chem.* **274**, 35325–35330.

Spitsberg, V.L., Matitashvili, E., Gorewit, R.C. (1995) Association and coexpression of fatty-acid-binding protein and glycoprotein CD36 in the bovine mammary gland, *Eur. J. Biochem.* **230**, 872–878.

Storch, J., Veerkamp, J.H., Hsu, K.T. (2002) Similar mechanisms of fatty acid transfer from human and rodent fatty acid-binding proteins to membranes:Liver, intestine, heart, muscle and adipose tissue FABPs, *Mol. Cell. Biochem.* **239**, 25–33.

Sundelin, J., Laurent, B.C., Anundi, H., Tragardh, L., Larhammar, D., Bjorck, L., et al. (1985) Amino acid sequence homologies between rabbit, rat, and human serum retinol-binding proteins, *J. Biol. Chem.* **260**, 6472–6480.

Thompson, J., Winter, N., Terwey, D., Bratt, J., Banaszak, L. (1997) The crystal structure of the liver fatty acid-binding protein. A complex with two bound oleates, *J. Biol. Chem.* **272**, 7140–7150.

Thumser, A.E., Storch, J. (2000) Liver and intestinal fatty acid-binding proteins obtain fatty acids from phospholipid membranes by different mechanisms, *J. Lipid Res.* **41**, 647–656.

Thumser, A.E., Wilton, D.C. (1996) The binding of cholesterol and bile salts to recombinant rat liver fatty acid-binding protein, *Biochem. J.* **320**(3), 729–733.

Thumser, A.E., Tsai, J., Storch, J. (2001) Collision-mediated transfer of long-chain fatty

acids by neural tissue fatty acid-binding proteins (FABP): studies with fluorescent analogs, *J. Mol. Neurosci.* **16**, 143–150.

Uysal, K.T., Scheja, L., Wiesbrock, S.M., Bonner-Weir, S., Hotamisligil, G.S. (2000) Improved glucose and lipid metabolism in genetically obese mice lacking aP2, *Endocrinology* **141**, 3388–3396.

van der Lee, K.A., Vork, M.M., De Vries, J.E., Willemsen, P.H., Glatz, J.F., Reneman, R.S., et al. (2000) Long-chain fatty acid-induced changes in gene expression in neonatal cardiac myocytes, *J. Lipid Res.*. **41**, 41–47.

Van Nieuwenhoven, F.A., Willemsen, P.H., van der Vusse, G.J., Glatz, J.F. (1999) Co-expression in rat heart and skeletal muscle of four genes coding for proteins implicated in long-chain fatty acid uptake, *Int. J. Biochem. Cell Biol.* **31**, 489–498.

Walz, D.A., Wider, M.D., Snow, J.W., Dass, C., Desiderio, D.M. (1988) The complete amino acid sequence of porcine gastrotropin, an ileal protein which stimulates gastric acid and pepsinogen secretion, *J. Biol. Chem.* **263**, 14189–14195.

Wang, H., He, Y., Kroenke, C.D., Kodukula, S., Storch, J., Palmer, A.G., Stark, R.E. (2002) Titration and exchange studies of liver fatty acid-binding protein with (13)C-labeled long-chain fatty acids, *Biochemistry* **41**, 5453–5461.

Weise, M.J., Hsieh, D., Hoffman, P.M., Powers, J.M., Brostoff, S.W. (1980) Bovine peripheral nervous system myelin P2 protein: chemical and immunological characterization of the cyanogen bromide peptides, *J. Neurochem.* **35**, 393–399.

Widstrom, R.L., Norris, A.W., Spector, A.A. (2001) Binding of cytochrome P450 monooxygenase and lipoxygenase pathway products by heart fatty acid-binding protein, *Biochemistry* **40**, 1070–1076.

Wilkinson, T.C., Wilton, D.C. (1986) Studies on fatty acid-binding proteins. The detection and quantification of the protein from rat liver by using a fluorescent fatty acid analogue, *Biochem. J.* **238**, 419–424.

Witthuhn, B.A., Bernlohr, D.A. (2001) Upregulation of bone morphogenetic protein GDF-3/Vgr-2 expression in adipose tissue of FABP4/aP2 null mice, *Cytokine* **14**, 129–135.

Wolfrum, C., Borchers, T., Sacchettini, J.C., Spener, F. (2000) Binding of fatty acids and peroxisome proliferators to orthologous fatty acid binding proteins from human, murine, and bovine liver, *Biochemistry* **39**, 1469–1474.

Wolfrum, C., Borrmann, C.M., Borchers, T., Spener, F. (2001) Fatty acids and hypolipidemic drugs regulate peroxisome proliferator-activated receptors alpha- and gamma-mediated gene expression via liver fatty acid binding protein: a signaling path to the nucleus, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 2323–2328.

Wolfrum, C., Buhlmann, C., Rolf, B., Borchers, T., Spener, F. (1999a) Variation of liver-type fatty acid binding protein content in the human hepatoma cell line HepG2 by peroxisome proliferators and antisense RNA affects the rate of fatty acid uptake, *Biochim. Biophys. Acta* **1437**, 194–201.

Wolfrum, C., Ellinghaus, P., Fobker, M., Seedorf, U., Assmann, G., Borchers, T., Spener, F. (1999b) Phytanic acid is ligand and transcriptional activator of murine liver fatty acid binding protein, *J. Lipid Res.* **40**, 708–714.

Wootan, M.G., Storch, J. (1994) Regulation of fluorescent fatty acid transfer from adipocyte and heart fatty acid binding proteins by acceptor membrane lipid composition and structure, *J. Biol. Chem.* **269**, 10517–10523.

Wu, F., Corsico, B., Flach, C.R., Cistola, D.P., Storch, J., Mendelsohn, R. (2001) Deletion of the helical motif in the intestinal fatty acid-binding protein reduces its interactions with membrane monolayers: Brewster angle microscopy, IR reflection-absorption spectroscopy, and surface pressure studies, *Biochemistry* **40**, 1976–1983.

Xu, Z., Bernlohr, D.A., Banaszak, L.J. (1993) The adipocyte lipid-binding protein at 1.6-A resolution. Crystal structures of the apoprotein and with bound saturated and unsaturated fatty acids, *J. Biol. Chem.* **268**, 7874–7884.

Young, A.C., Scapin, G., Kromminga, A., Patel, S.B., Veerkamp, J.H., Sacchettini, J.C. (1994) Structural studies on human muscle fatty acid binding protein at 1.4 A resolution: binding interactions with three C18 fatty acids, *Structure* **2**, 523–534.

Zimmerman, A.W., van Moerkerk, H.T., and Veerkamp, J.H. (2001) Ligand specificity and conformational stability of human fatty acid-binding proteins, *Int. J. Biochem. Cell Biol.*. **33**, 865–876.

Intracellular Fatty Acid Binding Proteins in Metabolic Regulation

John M. Stewart
Mount Allison University, Sackville, New Brunswick, Canada

1 Introduction 97
1.1 Fatty Acid Binding Proteins 97
1.2 General Introduction to Fatty Acid Binding Protein Function 98
1.3 Established Interactions between Carbohydrate and Fatty Acid-based Energy Production 98
1.4 The Involvement of FABP in Metabolism: Working Hypothesis 99
1.5 Criteria for Physiological Relevance of Metabolite Modulation of Fatty Acid Binding to FABP 100
1.6 Liver FABP 100
1.7 Heart/Muscle FABP 102

2 Potential of Formation of Schiff Bases: Nonenzymatic Glycation of FABPs 103

3 Theoretical Effects and Implications of Reciprocal Cross Talk 104
3.1 How Much Fatty Acid Would Be Available to Interact with Hexokinase? 104

4 Difference in Modulation of Fatty Acids and between Different Types of FABP 106

5 Where Else Should We Look? 106
5.1 Other Enzymes Influenced by Fatty Acids 106
5.2 The PPAR System 106

6 Summary 107

Acknowledgments 108

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

Bibliography 109
Books and Reviews 109
Primary Literature 109

Keywords

Allosteric
Literally "other space." The action of a modifying molecule on a protein at a site other than the usual site at which the protein performs its function.

Anabolic
Those reactions in biological systems that synthesize more complex molecules from simpler molecules, usually requiring energy.

Binding Site
A cavity formed by the folding of the protein backbone with the ability to bind a specific molecule in a discrete three-dimensional relationship to the bulk of the protein structure.

Catabolism
Those reactions in biological systems that break down complex molecules to simpler molecules, usually releasing energy that is trapped and converted to ATP.

Dissociation Constant
The dissociation constant (K_d) is a measure of the equilibrium between protein-bound and unbound ligand. K_d is also the concentration of ligand that occupies half of the protein binding sites in the system. A measure of the strength of ligand binding; a smaller K_d value represents stronger binding than a larger K_d.

Enzyme
A protein synthesized by the cell that behaves as a catalyst to a chemical reaction. Almost all chemical reactions in a biological system are catalyzed by a separate and specific enzyme. Enzyme names usually end in "-ase."

Fatty Acids
A class of biological molecules that consist of long, hydrophobic carbon chains with a single carboxylic acid functional group; poorly soluble in water and, hence, hydrophobic.

Gluconeogenesis
The synthesis of "new" glucose from smaller molecules such as pyruvate. In some senses, a reverse of the results of glycolysis.

Glycation
The chemical reaction between a carbohydrate and an amine group of a protein to form a covalent C=N bond (a Schiff base). The surface structure and operation of the protein is usually altered by this reaction.

Glycolysis
The series of 10 biochemical, enzyme-catalyzed reactions that converts glucose into pyruvate and is located in the cytoplasm of the cell.

Hyperlipidemia
A concentration of lipids larger than consistent with normal levels.

Inhibition Constant
The concentration of an enzyme inhibitor that reduces the enzyme rate by one half (K_i).

Isoform
In proteins, a variant of a protein coded for by a different gene and contains one, or more, amino acid substitutions.

Isosteric
Literally, ''same space.'' The action of a modifying molecule on a protein at a site, which is the same as that where the protein function occurs.

Ketone Bodies
When carbohydrate is curtailed, as in fasting, starvation, or diabetes, fatty acids are converted to a class of small, four-carbon molecules called *ketone bodies*, usually in the liver. Ketone bodies are more soluble and are transported in the circulation to muscles where they are converted to energy. One of the ketone bodies decomposes spontaneously to produce acetone that is exhaled.

Krebs Cycle
A central, cyclic series of eight biochemical, enzyme-catalyzed reactions taking place in the mitochondria that further degrade carbon resources from carbohydrates, fatty acid oxidation, and protein break down to carbon dioxide, extracting the chemical energy inherent in these structures. This is also known as the *citric acid cycle*.

Ligand
A molecule that binds to a specific protein binding site. Here, fatty acids are ligands to fatty acid binding proteins.

Metabolism
A combination of all catabolic and anabolic reactions.

Metabolic Flux
The rate at which a specific metabolic pathway, or a segment of that pathway, processes metabolites through the system. For example, the flux of glucose through glycolysis

would be the rate at which the combined reactions of glycolysis convert glucose to pyruvate.

Monomeric
A protein with only one subunit; opposite of multimeric, where a number of individual protein molecules form a complex.

Null mice
Laboratory mice in which a gene has been deleted.

Nuclear Hormone Receptors
A large family of proteins in the nucleus that bind various ligands, associate with other transcription factor proteins to form dimers, and then target expression sequences on genes, activating their transcription.

β-Oxidation
A series of enzyme-catalyzed reactions that occur mainly in the mitochondria that converts fatty acids into two carbon units for entrance into the Krebs Cycle.

Phosphorylation
The addition of a phosphate group, usually from ATP, to a simple molecule or to a protein through an ester bond formation. Protein phosphorylation is usually a switch to turn the protein activity on or off.

Peroxisome
A subcellular organelle or compartment.

Polyunsaturated Fatty Acids
Long-chain fatty acids that contain two or more carbon–carbon double bonds.

Protein
A polymer of amino acid linked by a peptide bond. The sequence of the 20 types of amino acids used in protein synthesis is dictated by the gene sequence.

Long-chain fatty acids are paradoxical: they are fuel, structural components, signaling molecules, and, at elevated levels, are toxic. Because of this nature of fatty acids, control of their concentration and movement within cells is important. Central to the biological and biochemical role of fatty acids is a family of proteins referred to as the *fatty acid binding proteins* (FABPs). After early discoveries that the FABPs were the intracellular transport system for fatty acids, their other potential roles, spurred on by the discovery that there were many types of FABPs, have continued to mount. This ancient family of proteins emerged before the divergence of the vertebrate and

invertebrate lines (more than one billion years ago) in response to the necessity of manipulating energy-rich fatty acids. A number of ancillary functions seem to have accrued to this diverse protein family that place them in a central position of energy metabolism, fuel sensing, and monitoring the concentrations of fatty acids derived from the diet.

1 Introduction

1.1 Fatty Acid Binding Proteins

Fatty acid binding proteins (FABPs) are a family of small intracellular proteins with a molecular weight between 14 000 and 15 000 (about 133 amino acids). With intracellular concentrations ranging from 200 to 400 μM, they can represent between 3 and 15% of soluble protein inside cells, depending upon the animal species and the tissue. For example, in humans, heart cells' FABP represents between 3 and 5% of soluble protein, while in the flight muscle cells of migrating Western Sandpipers, it can be 20% of soluble cellular protein. While the protein family has been observed in all animals (vertebrates and invertebrates), there is great three-dimensional structural similarity. The molecule is compact (about 250 × 350 × 400 nm) and, although generally referred to as a *β-clam motif* with one "side" pinched inward, has a barrel shape. Liver- and muscle-type FABP are shown in Fig. 1. All strands are hydrogen-bonded to the preceding and succeeding strands with only one exception, known as the *gap*. This rift in the continuous surface of the barrel (facing the reader in Fig. 1) allows flexibility of the barrel structure without significant disruption of the H-bonded strands around the rest of its circumference. At the "top" of the gap region (the structure is usually oriented with the helix-turn-helix motif at the top and back) is the portal region where the base of the helix-turn-helix and right turns of two β-strands are proximal. It is through the portal region that the ligand probably enters and exits the internal binding site. The internal volume to accommodate ligands has been determined and can vary from about 250 to 440 A^3, with the largest observed for liver FABP that accommodates two, rather than the usual single, fatty acid molecule.

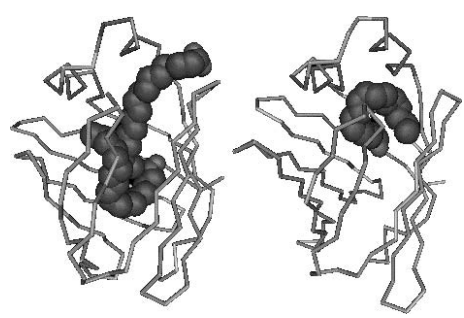

Fig. 1 The peptide backbone structures of rat liver FABP (left panel) with both binding sites filled with oleate and human muscle FABP (right panel) with its single binding site occupied. The images were derived from Cn3D v 4.0 from NCBI with accessions codes 1HMS (muscle protein) and 1LFO (liver protein).

1.2 General Introduction to Fatty Acid Binding Protein Function

FABPs have functions arising from their ability to bind long-chain fatty acids (FAs) and to interact with either the membranous or aqueous fraction of the cell (predominantly the cytoplasm). The potential of interactions between FABPs and other proteins has been supported by evidence that FABPs can alter enzyme kinetics, interact directly with both hormone-sensitive lipase and a cell membrane fatty acid transport protein (CD36), and with peroxisome proliferator–activated receptors (PPAR) in the nucleus that are involved in turning on the expression of various genes. Studies with null mice in which the gene for heart FABP was ablated reflected not only the role of the protein in upregulating genes of FA metabolism but also the compensatory activation of the glycolytic system and apparent reduction of the usual controls of glycolysis.

FABPs can also bind and transport signaling molecules such as bilirubin, prostaglandin E1, and lipoxygenase metabolites of the fatty acid arachidonic acid, although some signaling lipids such as sleep-inducing oleamide in the central nervous system do not bind to brain FABPs. Further, FABPs may protect polyunsaturated fatty acids from free radical damage (peroxidation) through preferential binding of them while greater partitioning of oxidation products of polyunsaturated fatty acids into FABPs, when compared with unilamellar vesicles, may indicate protection of membranes from these monohydroxy fatty acids. Thus, FABPs are involved in intracellular signaling, cell development, and membrane protection, and it appears that they are more than intracellular ferries of poorly soluble hydrophobic molecules.

The idea is developed herein of an interaction between carbohydrate and fatty acid energy metabolism (also known as *fuel sensing* or *fuel preference*) that involves FABPs. Evidence is presented to support the contention that interactions between nonlipid metabolites and FABPs can modulate the partitioning of long-chain FAs into FABP binding sites. Additionally, experimental results showing the potential for reciprocal interaction between FAs and enzymes of glycolysis will be presented. Finally, different FABP types (muscle and liver) respond differently to nonlipid metabolites. The suggested cross talk between the two fuel streams in energy metabolism is centered on substrate cycles involving the enzyme pairs hexokinase-glucose 6-phosphatase and phosphofructokinase-fructose 1,6-phosphatase in muscle and glucokinase-glucose 6-phosphatase in liver. Differential involvement of different FABP types with separate metabolic loci may provide some insight into why there are multiple intracellular fatty acid binding proteins but only a single major extracellular fatty acid binding protein.

1.3 Established Interactions between Carbohydrate and Fatty Acid-based Energy Production

The exchange of rudimentary status information (predominantly concentration) between carbohydrate and FA catabolisms by substrate level control and inhibition/activation of participating enzymes is well established. The reciprocal interaction between FA levels and glucose use, referred to as the *Randle cycle*, suggests that modulation of glycolytic flux depends upon

the mitochondrial production of citrate which, when transported to the cytoplasm, inhibits the major flux generator of this pathway, phosphofructokinase-I (PFK-I). In effect, an elevated citrate concentration inhibits PFK-I and slows the flow of carbohydrate through this point in the pathway. The other major effect of FA is to activate a pyruvate dehydrogenase kinase that phosphorylates mitochondrial pyruvate dehydrogenase and decreases its activity. This curtails the movement of pyruvate into mitochondrial pathways (Krebs Cycle), and an overall decrease is observed in glucose oxidation. This accounts well for the commonly observed decrease of glucose oxidation when lipid levels are increased.

The control of mitochondrial oxidation of fatty acids (β-oxidation) is vested in gating entry of FAs into the mitochondrial matrix through coordinated actions of carnitine palmitoyltransferase I (CPT I), palmitoylcarnitine translocase, and matrical carnitine palmitoyltransferase II. A powerful inhibitor of CPT I and part of the fuel sensing system is malonyl-CoA, the first committed metabolite in FA synthesis and itself a product of cytoplasmic citrate. CPT I is also under phosphorylation control in liver, where phosphorylation by cAMP-protein kinase increases the CPT I activity by 30 to 80%. Thus, an increase in mitochondrially produced citrate (from carbohydrate or other sources) increases its partition to the cytoplasm. An increased cytosolic concentration of citrate inhibits glucose processing at PFK-I and shifts the carbon resources it represents into the pathway that synthesizes fatty acids. The first metabolite along this anabolic pathway is malonyl-CoA that "shuts down" entry of fatty acid carbon into the mitochondria where it is oxidized.

1.4 The Involvement of FABP in Metabolism: Working Hypothesis

An additional level of lipidic control of glucose phosphorylation in muscle has recently been reported: CoA esters of long-chain FAs (in rat and human skeletal muscle) and long-chain FAs (in bovine heart) inhibit hexokinase I. This enzyme, by adding a phosphate group to glucose, traps glucose in the cell. There are many older reports that long-chain FAs inhibit a number of other enzymes, but uncertainty arises from those reports since unrealistically large FA concentrations were employed, and detergental effects cannot be ruled out. One careful kinetic study from that pre-FABP era was the inhibitory effect of palmitate on muscle PFK-I with an inhibition constant (K_i) of 25 μM, although the physical binding of FAs to PFK-I has not been studied. Thus, if FAs can influence enzymes of glycolysis (and elsewhere, see below), an involvement of FABP in delivering the FA to the sites of action, presumably through protein:protein interaction, is strongly implicated since such fatty acids have vanishingly small solubility in the aqueous cytosol. Other areas where long-chain FAs have influence, such as the ADP/ATP translocase-porin-hexokinase complex of mitochondria, might also be examined productively.

If FABPs modulate FA metabolism above the level of the genetic machinery, to what signals do they respond? Clearly, any productive modulation of FABP function must act in response to the instantaneous status of the cell/tissue. Some years ago, we suggested that the metabolic status of the cell might be part of the signaling mechanism influencing FA binding

characteristics of various FABPs. Starting with this idea, we began to search for metabolites or conditions within the cell that might influence the FABP operation through modulation of formation of the protein:FA complex. This would probably be an allosteric modulation (in the strictest sense of the term, i.e. not at the fatty acid binding site) since competitive interaction, such as observed with ferri- and ferroheme, would be a binary system (on/off) and not optimally adaptive to a dynamic intracellular milieu. It must be pointed out that, in a strictly thermodynamic sense (i.e. mass action), the effect would result from repartitioning effects. Although it might seem unusual that a small, monomeric protein with a relatively small surface area that changed little upon binding FAs would display allosteric properties usually observed with large, multimeric enzyme systems, such regulation has been demonstrated for other small proteins: cytochrome *c* has a saturable binding site for ATP that regulates the electron transport system and hence oxygen consumption and ATP production. If FABPs are involved in the productive modulation of lipid metabolism and in targeting of FABPs, for example, to the mitochondria, there must be an exchange of information between the binding of FAs and the metabolite status of the cell.

1.5 Criteria for Physiological Relevance of Metabolite Modulation of Fatty Acid Binding to FABP

In the search for possible modulators of FABPs, we compared the degree of binding of reporter ligands (radioactively labeled or fluorescent fatty acids) in the presence and absence of physiological concentrations of metabolites. The assay conditions used in such studies are particularly important since FA binding is effected by ionic strength, pH, and by temperature, and it is vital to mimic the intracellular conditions as closely as possible. Statistically significant changes in binding of reporter ligands to FABP must occur at modulator concentrations encountered under physiological conditions. We examined potential modulators at concentrations found in unstressed tissue since metabolic flux is normally curtailed compared to the maximal flux attainable.

A second level criterion we employ is that the equilibrium dissociation constant (the apparent K_d) of any modulator must also fall within physiological concentration ranges to be part of a normal regulatory network. Finally, interactions that influence FA binding should be consistent with what is known about particular metabolic pathways and physiological states.

1.6 Liver FABP

As an initial test of the working hypothesis, we examined the effect of glycolytic metabolites on the binding of both ^{14}C-labeled oleate and the fluorescent probe *cis*-parinarate (cPnA) to rat L-FABP [27] at 37 °C. To do this, we set the concentration of reporter molecules at the value of the K_d of the ligand, where binding is most sensitive to change. After establishing the FABP:FA complex, we challenged it with test metabolites. We reported that, of the early glycolytic intermediates, the most effective modulators of the ability of L-FABP to bind FAs were glucose and glucose-6-phosphate (G6P). Both of these metabolites increase the ability of FABP to bind long-chain FAs 30 and 40% for 6 mM glucose and 0.25 mM G6P respectively. An alternate binding

assay using the fluorescent probe cPnA provided the same result. In examining the effect of increasing concentrations of these two metabolites on oleate and cPnA binding, the concentrations required to produce a 50% increase in the observed FA binding change were 6 mM and 0.2 mM respectively, within the normal physiological range of glucose and G6P in rat liver. As part of this study, we also examined glucose-1-phosphate (G1P), phosphate ion, fructose-6-phosphate (F6P), fructose-1,6-bisphosphate (F-1,6-P_2), and fructose-2,6-bisphosphate (F-2,6-P_2), phosphoenol pyruvate (PEP), ATP, ADP, AMP and cAMP, NAD^+ and NADH, $NADP^+$ and NADPH, and acetyl-CoA, none of which had an effect on FA binding to rat L-FABP.

A recent study included an examination of the effect of glucose and G6P on [1–^{14}C]oleate binding to human H-FABP and L-FABP. While there was no effect of glucose or G6P on oleate binding as compared to that on the control, the human muscle-type FABP was affected and was similar (although smaller in magnitude) to the effects on bovine H-FABP, as described below.

We extended this study to L-FABP of another rodent species, Richardson's ground squirrel (*Spermophilus richardsonii*), arguing that such a mechanism may be more generally observed. In this system, glucose and G6P also increased the binding of cPnA to the FABP at 37 °C compared to that of the control by $20.6 \pm 0.5\%$ and $41.4 \pm 2.2\%$ respectively ($n = 4$; $p < 0.05$), when examined under the same conditions, as reported before, for rat and using *cis*-parinarate as a fluorescent probe.

In summary, Glc and G6P increased the ability of L-FABP to bind FAs in three mammalian systems and to different degrees. Since the number of ligand molecules bound to the protein was not likely to increase (2 : 1 binding stoichiometry for L-FABP), the effect was probably centered on changes in the binding strength for fatty acid. It should be noted that the effect of glucose on liver-type FABP to glucose is exhibited in frogs (*Rana pipens* and *Rana sylvatica*) and turtle (*Chrysemus picta belli*). The similarity of response in mammals, amphibians, and reptiles suggests a more general phenomenon. We are presently determining amino acid sequences of these proteins and carrying out topological mapping of three-dimensional structures in an attempt to identify surface features of the protein that would best provide information for a phase of site-directed mutational studies that could create the glucose effect or eliminate it.

Because of different experimental protocols and few reports of work with pure proteins, it is difficult to find a general metabolic scenario into which the L-FABP work can be fit. However, recent examination of the effect of oleate on the glycolytic/gluconeogenic balance and glycogen systems of fasted rat hepatocytes is consistent with a role for FABP. In the presence of S4048, an inhibitor of endoplasmic reticulum G6P translocase, oleate increased intracellular G6P 3-fold and glycogen production 1.5-fold. Further, oleate did not have an inhibitory effect on glucokinase (the liver form of hexokinase), but stimulated G6P phosphatase flux (part of the putative hexokinase/G6P phosphatase substrate cycle) and decreased overall lactate production. The increase in G6P was attributed to increased gluconeogenesis. Thus, it appears that long-chain fatty acids, while interacting with early glycolytic enzymes, foster gluconeogenesis in liver. Presumably, an increase in G6P or glucose could increase the partitioning of

FAs to the FABP and behave as a communicative link to lipid status.

Other studies in our laboratory with rat L-FABP have examined the role of the ketone body, D-3-hydroxybutyrate, where, at normal physiological levels (0.14 mM), there was only a small increase in the binding of FA (10%). At levels mirroring diabetic or fasting/starvation hepatic levels (about 7 mM), there was a 45.5% increase in FA binding. The effect was saturable with an inhibition constant, K_i (3-hydroxybutyrate), of 4.9 mM, which is not physiological under euglycemic or nonstarvation conditions. Increased FA binding with elevated 3-hydroxybutyrate (when carbohydrate supply is curtailed as in diabetes or fasting/starvation) could reverse a tendency to gluconeogenesis and modify delivery of available FAs.

1.7 Heart/Muscle FABP

For heart tissue, which has different hexokinase (HK) isoforms and a different FABP type from liver (H-FABP or M-FABP), the situation appears equally complex (note: muscle-type and heart-type FABPs are the same protein and will be referred to as H-FABP here). We showed that bovine heart hexokinase (type I) is inhibited by long-chain fatty acids in a manner that correlates positively with chain length and degree of saturation, but is activated by FAs containing less than 12 carbons. Hexokinase also has a saturable binding site for cPnA and oleate with dissociation constant (K_d) values of 3 μM and 1.3 μM respectively (by fluorescence competition assay), while the short and medium-chain FAs that activated HK did not bind to this site. In other words, long-chain fatty acids bind strongly to hexokinase in a discrete binding site to inhibit the trapping of blood borne glucose within the cell. We can now report work on the FABP purified from bovine heart and for heart-type FABP from the spadefoot toad (*Scaphiopus couchii*).

Figure 2 shows the effects of a number of early glycolytic intermediates on FA binding to the FABP Bovine H-FABP was insensitive to glucose and G6P, as reported for human H-FABP and with toad heart FABP. There was no significant effect on the binding of FAs by physiological concentrations of F6P, but there was significant inhibition of FA binding by $F16P_2$. G1P and P_i did not have any effect on the binding of the fluorescent probe. In contrast to mammalian L-FABP, FA binding by bovine H-FABP was unchanged by 3-hydroxybutyrate.

Unlike hepatic tissue, muscle and heart tissue do not actively produce glucose or ketone bodies. Heart and red skeletal muscles import carbon resources (carbohydrates, fatty acids, and ketone bodies) in response to energy needs and are predominantly aerobic. This essentially unidirectional resource flow is reflected in the poise of muscle metabolic pathways, where the major control points are PFK-I, pyruvate dehydrogenase, and mitochondrial import of FAs. As discussed above, both long-chain FAs and their CoA derivatives inhibit glucose phosphorylation in heart and skeletal muscle. The focus of the interactions in muscle tissue appears to be PFK-I. Long-chain FAs inhibit muscle PFK-I ($K_i = 25$ μM) and F-1,6-P_2 (at 0.03 mM) inhibits FA binding, establishing the possibility of a reciprocal interaction. Elevated FA could decrease glycolytic flux through inhibiting actions on both HK and PFK-I, while decreased F-1,6-P_2 could lift resting state inhibition of FA binding to this FABP type.

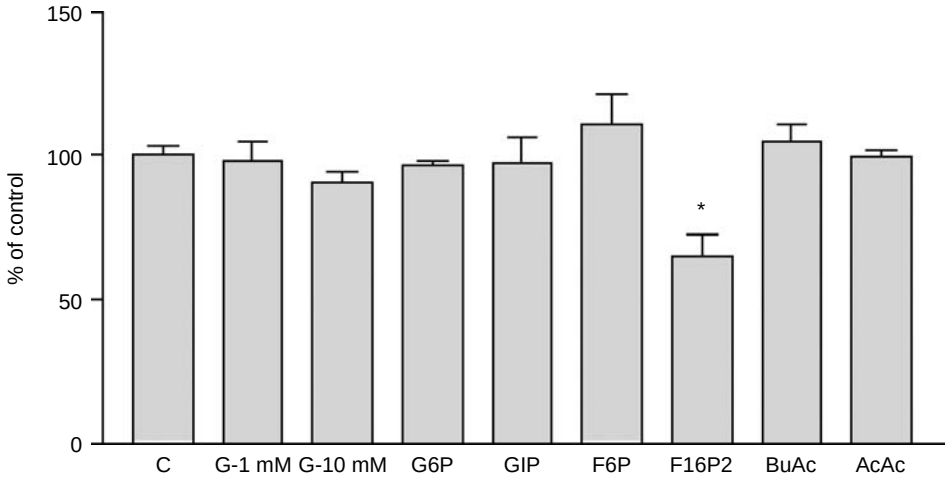

Fig. 2 The effect of various metabolites on the binding of *cis*-parinaric acid (1 μM) to bovine heart FABP compared with the control (no nonlipid metabolite). The assays were carried out at 37 °C, in 50 mM sodium phosphate buffer, pH 7. The assays contained 20 μg of pure proteins per milliliter. The concentrations of metabolites not indicated were G6P, 0.5 mM; F6P, 0.1 mM; F1,6P2, 0.03 mM; 3-hydroxybutyrate (ButAc), 0.1 mM; acetoacetate (AcAc), 0.02 mM. The asterisk indicates statistically significant differences from the control at the 95% confidence limit. The experiments were replicated three to five times with the data representing the means and the error bars the standard deviation.

2
Potential of Formation of Schiff Bases: Nonenzymatic Glycation of FABPs

Carbohydrates that undergo ring opening and formation of a carbonyl group can form Schiff bases with primary amines such as lysine residues that are prevalent in FABP and which appear involved in the collisional mechanism proposed for loading and unloading muscle-type FABPs. Since many of the glycolytic intermediates that can ring open to produce the carbonyl function group influence binding of FAs to FABPs (G6P, glucose, F-1,6-P_2) while G1P and F-2,6-P_2 (which are locked in the closed hemiacetal and hemiketal rings) do not affect FA binding, it is possible that formation of Schiff bases with protein amino groups may cause the observed altered sequestration of FAs. To examine this possibility, we incubated a number of purified FABPs and bovine serum albumin with fluorescently tagged glucose (2-(*N*-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)amino)-2-deoxyglucose; 2-NBDG) at room temperature and in 50 mM sodium phosphate, pH 7.0. We monitored the proteins for indications of covalent fluorescent tagging that would result from Schiff base formation. Periodically, aliquots of the incubation mixtures were removed, and 2-NBDG was separated from protein with a small size-exclusion column. The protein fraction was examined spectrofluorometrically to detect fluorescent tagging (excitation at 473 nm, emission at 532 nm). Figure 3 shows the effect of incubating a number of FABPs and BSA with the fluorescent tag for three days. Only after 48 h of incubation were proteins tagged with the fluorescent

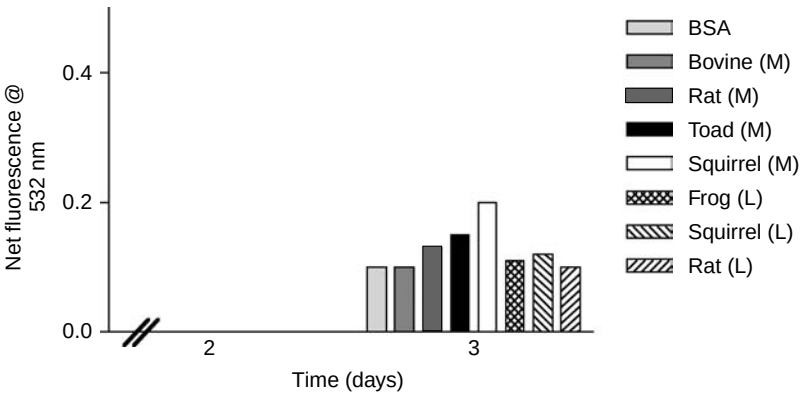

Fig. 3 The time course of glycation of bovine serum albumin and assorted FABPs (30 μg/ml) with D-glucose analog NBDG (80 μM) in 10 mM sodium phosphate buffer, pH 7.0, and 1% (w/v) sodium azide at 25 °C. Over the time course, as proteins (after separation from small molecules with size exclusion) became glycated, they became fluorescently tagged (473-nm excitation, 532-nm emission). A fluorescence value of 0.5 represents 1-nmol fluorescent tag.

carbohydrate derivative at detectable levels. The slowness of the nonenzymatic glycations we observed was consistent with nonenzymic glycation rates of human serum albumin (days) using other fluorometric methods. Thus, it is unlikely that Schiff base formation was significant during the time required for making the measurements of the effects of metabolites on FA binding (minutes).

3 Theoretical Effects and Implications of Reciprocal Cross Talk

3.1 How Much Fatty Acid Would Be Available to Interact with Hexokinase?

The similarity of dissociation constants of HK and FABP for oleate and cPnA leads to an important question surrounding such mutual interactions. Given the large cellular concentration of FABP, is there enough unbound FA to interact with HK? To model the amount of unbound FA in the presence of two different binding proteins requires the concentration value of both proteins, the total FA concentration, and the equilibrium dissociation constants (K_d) of both binding proteins to be known. The intracellular concentration of FABP has been reported to be 85 μM, while some estimates provide ranges of 200 to 400 μM. An intermediary value of 250 μM muscle-type FABP was used in the following model.

Determining the molar concentration of HK in heart tissue is more difficult since it is not directly reported in the literature. We can estimate from the maximal tissue activity (V_{max}) and the measured catalytic rate constant (k_{cat}) of the enzyme a reasonable enzyme concentration since $V_{max} = k_{cat}[E]_T$, where $[E]_T$ is the total enzyme. The k_{cat} of type I HK is 64 s^{-1}, while the maximal activity in mammalian heart has been reported between 5 and 10 μmol min^{-1} g^{-1} tissue for mammals. A value of 8 μmol min^{-1} g^{-1} tissue (porcine heart) was used here.

Finally, the estimate must convert the tissue weight–based activities to concentrations: the assumption of 75% water content will suffice. Thus, with $V_{max} = 8\ \mu mol\ min^{-1}\ g^{-1}$ and $k_{cat} = 64\ s^{-1}$, we can estimate a minimum $[E]_T = 3\ \mu M$. Using a K_d of 1 μM for FABP and 1.3 μM for HK, we can model equilibrium concentrations of unbound FA and degree of saturation of both protein systems at different initial FA concentrations. Figure 4(a and b) shows the result of this very simple equilibrium model.

The amount of free FA in the simple model in which FABP is present is enough to saturate HK, as indicated by Fig. 4(a) that shows the binding isotherms of FABP at various free FA concentrations, and

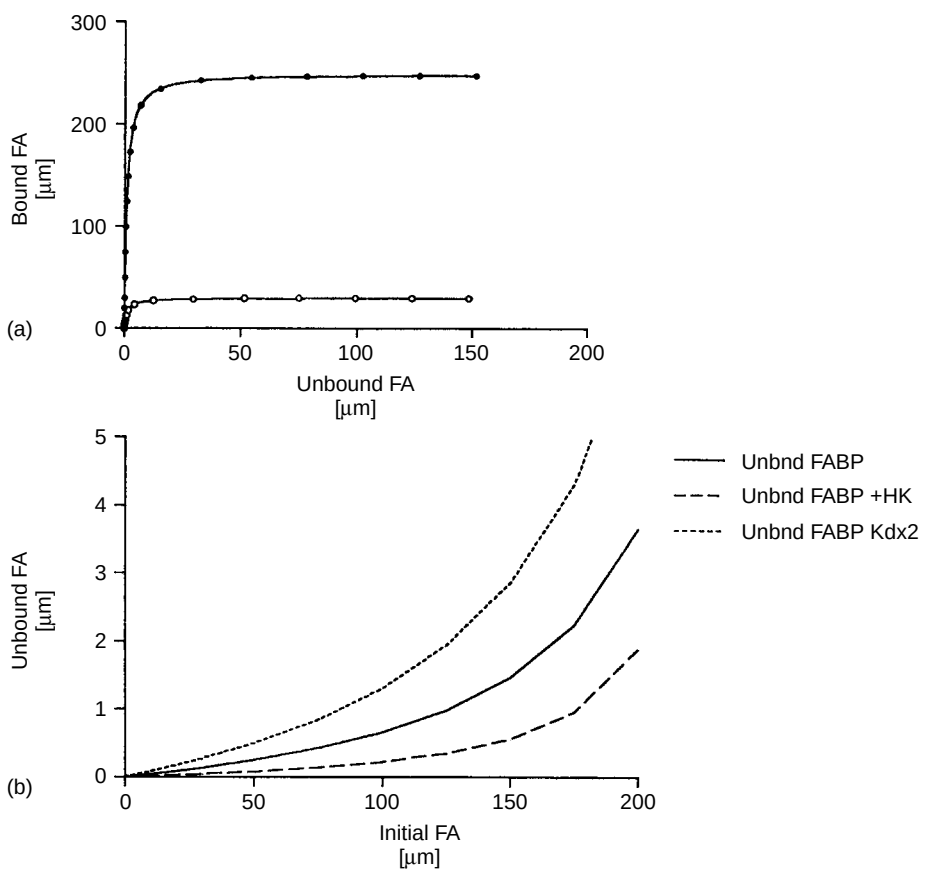

Fig. 4 (a) The degree of saturation of a muscle-type FABP and hexokinase versus the amount of unbound fatty acid (oleate) in the model system. The saturation curve of hexokinase (lower curve) was generated using the concentration remaining after FABP equilibrated its binding site with ligand. The hexokinase binding curve was magnified by 10 to be observable at this scale. (b) The residual concentration of unbound fatty acid at various initial fatty acid concentrations in the presence of FABP and of FABP plus hexokinase. The upper line indicates the effect on the amount of unbound fatty acids if binding strength of FABP was decreased by a factor of two (i.e. K_d increased twofold).

the binding curve of HK using the FA remaining unbound in the presence of FABP. Figure 4(b) shows the estimated concentration of unbound FA in the presence of FABP, of FABP plus HK, and the effect on the amount of unbound FA when the K_d value of the FABP for oleate is doubled (i.e. binding is weakened), as is seen in toad muscle FABP in the presence of F-1,6-P_2. Thus, even with FABP present, providing very low concentrations of unbound FAs, unbound FA could be sequestered on HK.

4 Difference in Modulation of Fatty Acids and between Different Types of FABP

The data suggest that liver FABPs and muscle types of FABP respond different nonlipid metabolites. While the liver types increase the degree of FA binding in the presence of glucose and G6P, with an activation of binding also elicited by ketone bodies, muscle type does not respond to glucose or G6P, is inhibited by F-1,6-P_2, and does not respond to ketone body. If FABPs are involved in fuel selection, the different metabolic poise of these two tissues would demand that they operate differently. The differential response may offer some part of an explanation of the presence of different FABPs in different tissues and perhaps multiple forms in some tissues.

5 Where Else Should We Look?

5.1 Other Enzymes Influenced by Fatty Acids

The influence of long-chain FAs on enzyme kinetics, specifically hexokinase, and PFK-I have been noted above. FAs also have an effect on the myoglobin system and bind strongly to cytochrome *c* altering their functions in the cell. As part of an examination of potential sites of interference during hyperlipidemia, we examined a battery of enzymes to see if elevated FA concentrations altered their reaction kinetics. We examined the enzymes of glycolysis (HK, PFK, PK, lactate dehydrogenase, aldolase, triosephosphate isomerase, pentose phosphate cycle (G6P dehydrogenase), the citric acid cycle (citrate synthase, malate dehydrogenase, and α-ketoglutarate dehydrogenase), and β-oxidation (CPT-I)). Additionally, we examined the effects of fatty acids on the enzymes that degrade reactive oxygen species (catalase, superoxide dismutase). Table 1 shows the results of incubation of enzymes of major metabolic systems with oleate. In these studies, all enzymes were obtained from commercial sources, and assays were established that provided maximal specific activities. We compared the maximal enzyme activities over a range of FA concentrations with the highest being 0.2 mM. The enzymes listed above, and not in Table 1, showed no effect upon incubation with long-chain fatty acids.

The clearest conclusion from this survey is that most enzymes are not sensitive to concentrations of FAs. Those enzymes that are sensitive to FAs are those that have controlling functions on various metabolic pathways and indicate that there is a closer connection between lipid and carbohydrate metabolism than has generally been appreciated.

5.2 The PPAR System

The peroxisomal proliferator-activated receptors (PPAR) are lipid-binding proteins,

Tab. 1 The effect of oleate on the specific activity of selected enzymes and proteins. The values are mean specific activity ± standard deviation and the number of trials are given in parentheses.

Enzyme	***No oleate***	***0.2 mM oleate***
	[μmol min^{-1} mg^{-1} protein]	***[μmol min^{-1} mg^{-1} protein]***
Glycolysis		
Hexokinase (bovine heart)	2.74 ± 0.10 (3)	0.55 ± 0.03 (3)[b]
Phosphofructokinase (rabbit muscle)	18.6 ± 0.9 (3)	0.32 ± 0.07 (3)[b]
Pyruvate kinase (rabbit muscle)	9.8 ± 1.6 (6)	10.2 ± 1.6 (3)
Lactate dehydrogenase (rabbit muscle)	1.8 ± 0.1 (5)	1.7 ± 0.01 (5)
Krebs cycle		
Citrate synthase (porcine heart)	54.8 ± 2.5 (3)	51.3 ± 3.1 (3)
Pentose phosphate cycle		
G-6-P dehydrogenase (yeast)	26.6 ± 0.7 (3)	0.20 ± 0.04 (3)[b]
G-6-P dehydrogenase (bovine adrenals)	0.67 ± 0.04 (6)	0.55 ± 0.20 (6)
Electron transport		
Ferrocytochrome *c* oxidation[a]	0.050 ± 0.005 (4)	0.100 ± 0.015 (4)[b]
Reactive oxygen		
Catalase (bovine liver)	11.2 ± 0.4 (6)	11.4 ± 0.3 (5)
Superoxide dismutase (SOD) (bovine erythrocyte)		
L-DOPA oxidation-no SOD	0.250 ± 0.002 (3)	0.260 ± 0.005 (3)
L-DOPA oxidation-10 mg SOD	0.13 ± 0.1 (3)	0.14 ± 0.01 (3)

[a] Cytochrome *c* autoxidation was measured with Hansatech DW2/2 oxygen electrode at pH 7, 37 °C, in 50 mM sodium phosphate buffer with an initial [cytochrome *c*] = 0.25 mM. Units are μmol O_2 min^{-1} mg^{-1} protein.
[b] Indicates values that were significantly different ($p < 0.05$).

members of the nuclear hormone receptors family, and transcription factors that initiate expression of genes after forming a complex with the retinoid X receptor proteins (RXR). The PPAR:RXR complex binds to specific expression elements of the gene. PPARs (types α and γ interact with FA-loaded FABPs in the nucleus, and a fatty acid is transferred to the PPAR. It is the PPAR:FA complex than can then bind to the RXR. Fatty acid binding proteins along with bound fatty acids are translocated into the nucleus where they interact with the PPARs. FABPs without bound FAs are excluded from the nuclear matrix. Conceivably, alteration of the binding strength of FABPs for their fatty acids would influence the amount of import of the loaded FABP and perhaps the ease at which the FA is transferred to a PPAR. If, in their interaction with nonlipid metabolites, FABPs mediate a potential interaction between glycolytic status and the expression of FA metabolic enzymes and proteins, they are more than transport barges for poorly soluble fatty acids.

6 Summary

The binding of FA to FABP can be modulated by nonlipid metabolites.

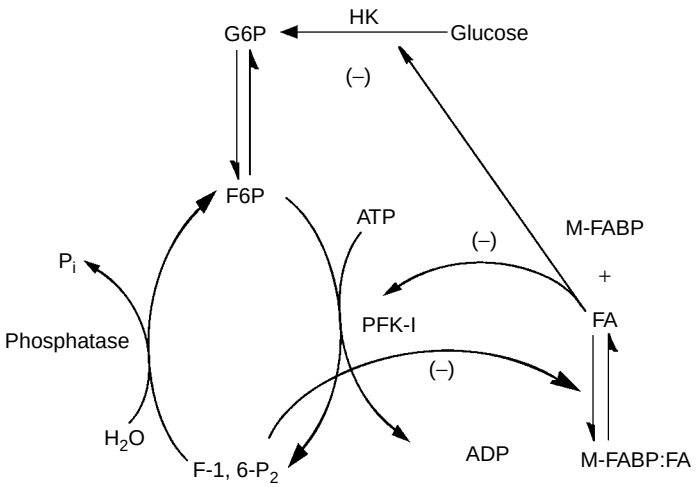

Fig. 5 A model of the interaction scheme between early glycolysis and FABP in mammalian muscle or heart. The bracketed negative signs indicate inhibition of either enzyme activity or binding to FABP. The values of the inhibition constants are given.

Figure 5 summarizes one model (muscle FABP/PFK-I) that forms the basis of present endeavors. There is a differential response of L- and M-FABP to metabolites. These differential responses may offer insight into why there are different FABP types expressed in the same tissue. While the ability of FAs to bind to cytoplasmic enzymes may alter the concentrations of unbound FAs (usually sequestered in membranes), the reverse, where FABPs can influence enzyme substrates, must be evaluated. A consequence of FABP interactions with glycolytic intermediates may influence the concentration of available glucose in the case of hexokinase and F-1,6-P_2 in the case of PFK-I. Significant fractions of cytoplasmic glycolytic substrates could be associated with FABPs. As in conservation of critical, thermodynamic characteristics of fatty acids binding to FABPs across species, the effects of metabolites on FA binding to FABPs being retained across diverse species support this possible new role for FABPs.

Acknowledgments

Many thanks to John A. Blakely, my long-time research associate, for discussions, reading of the manuscript, and technical expertise, and to my colleagues Janet M. Storey and Kenneth B. Storey, Carleton University, for stimulating discussions over the years. The work of a number of students has been essential in this work, notably (in no particular order), Steven Wiseman, Jaime F. Snow (nee Claude), Stephen J. Crozier, Daniel Jardine, Dr. Monica Henry, Philip Karpowicz, Dr Kevin Larade, Shauna Flavelle, Megan Miller, Ashley Woods (glycation work), Luke E. Gauthier, Nathalie Boudreau, Dr Justin MacDonald, Ted Wright, Vanessa Dewling, Christopher Martin, Jonathan D. Schinkel, R. James Allan, and Dr Frank O'Dea. The National Sciences and Engineering Research Council of Canada, the Heart and Stroke Foundation of New Brunswick and Mount Allison University must be acknowledged

for the funding that allowed the research to proceed.

See also Biological Regulation by Protein Phosphorylation; Lipid and Lipoprotein Metabolism; Metabolic Basis of Cellular Energy.

Bibliography

Books and Reviews

Alkadi, H.A., Fox, K.A. (2004) Do we need additional markers of myocyte necrosis: the potential value of heart fatty-acid-binding protein? *QJM* **97**, 187–198.

Besnard, P., Niot, I., Poirier, H., Clement, L., Bernard, A. (2002) New insights into the fatty acids-binding protein (FABP) family in the small intestine, *Mol. Cell. Biochem.* **239**, 139–147.

Duplus, E., Forest, C. (2002) Is there a single mechanism for fatty acid regulation of gene transcription, *Biochem. Pharmacol.* **64**, 893–901.

Duttaroy, A.K., Spener F. (Eds.) (2003) *Cellular Proteins and their Fatty Acids in Health and Disease*, Wiley-VCH GmbH & Co. KgaA, Weinheim.

Glatz, J.F., Bonen, A., Luiken, J.J. (2002) Exercise and insulin increase muscle fatty acid uptake by recruiting putative fatty acid transporters to the sarcolemma, *Curr. Opin. Clin. Nutr. Metab. Care* **5**, 365–370.

Hanhoff, T., Lucke, C., Spener, F. (2002) Insights into binding of fatty acids by fatty acid binding proteins, *Mol. Cell. Biochem.* **239**, 45–54.

Laakso, M. (2004) Gene variants, insulin resistance, and dyslipidaemia, *Curr. Opin. Lipidol.* **15**, 115–120.

Londraville, R.L. (1996) Intracellular fatty acid-binding proteins: putting lower vertebrates in perspective, *Braz. J. Med. Biol. Res.* **29**, 707–720.

Randle, P.J. (1998) Regulatory interactions between lipids and carbohydrates: the glucose fatty acid cycle after 35 years, *Diabetes Metab. Rev.* **14**, 263–283.

Scopes, R.K. (1987) *Protein Purification: Principles and Practice*, 2nd edition, Springer-Verlag, New York.

Stewart, J.M. (2000) The cytoplasmic fatty-acid-binding proteins: thirty years and counting, *Cell. Mol. Life Sci.* **57**, 1345–1358.

Stremmel, W., Pohl, L., Ring, A., Herrmann, T. (2001) A new concept of cellular uptake and intracellular trafficking of long-chain fatty acids, *Lipids* **36**, 981–989.

Van der Vusse, G.J., van Bilsen, M., Glatz, J.F., Hasselbaink, D.M., Luiken, J.J. (2002) Critical steps in cellular fatty acid uptake and utilization, *Mol. Cell. Biochem.* **239**, 9–15.

Weiss, E.P., Brown, M.D., Shuldiner, A.R., Hagberg, J.M. (2002) Physiol, *Genomics* **10**, 145–157.

Primary Literature

Binas, B., Danneberg, H., McWhir, J., Mullins, L., Clark, A.J. (1999) Requirement for the heart-type fatty acid binding protein in cardiac fatty acid utilization, *FASEB* **13**, 805–812.

Chu, R., Lim, H., Brumfield, L., Liu, H., Herring, C., Ulintz, P., Reddy, J.K., Davison, M. (2004) Protein profiling of mouse livers with peroxisome proliferators-activated receptor alpha activation, *Mol. Cell. Biol.* **24**, 6288–6297.

Craig, D.B., Wallace, C.J. (1993) ATP binding to cytochrome c diminishes electron flow in the mitochondrial respiratory pathway, *Protein Sci.* **2**, 966–976.

Dutta-Roy, A.K., Gopalswamy, N., Trulzsch, D.V. (1987) Prostaglandin E1 binds to Z protein of rat liver, *Eur. J. Biochem.* **162**, 615–619.

Ek, B.A., Cistola, D.P., Hamilton, J.A., Kaduce, T.L., Spector, A.A. (1997) Fatty acid binding proteins reduce 15-lipoxygenase-induced oxygenation of linoleic acid and arachidonic acid, *Biochim. Biophys. Acta* **17**, 75–85.

Ek-von Mentzer, B.A., Zhang, G., Hamilton, J.A. (2001) Binding of 13-HODE and 15-HETE to phospholipid bilayers, albumin and intracellular fatty acid binding proteins. Implications for transmembrane and intracellular transport and for protection from lipid peroxidation, *J. Biol. Chem.* **276**, 15575–15580.

Fang, T.-Y., Alechina, O., Aleshin, A.E., Fromm, H.J., Honzatko, R.B. (1998) Identification of

a phosphate regulatory site and a low affinity binding site for glucose 6-phosphate in the N-terminal half of human brain hexokinase, *J. Biol. Chem.* **273**, 19548–19553.

Font, B., Vial, C., Gautheron, D.C. (1975) Intracellular and submitochondrial localization of pig heart hexokinase, *FEBS Letts.* **56**, 24–29.

Glatz, J.F., Vork, M.M., Cistola, D.P., van der Vusse, G.J. (1993) Cytoplasmic fatty acid binding protein: significance for intracellular transport of fatty acids and putative role on signal transduction pathways, *Prostaglandins Leukot. Essent. Fatty Acids* **48**, 33–41.

Gustafson, L.A., Neeft, M., Reijngoud, D.-J., Kuipers, F., Sauerwein, H.P., Romijn, J.A., Herling, A.W., Burger, H.-J., Meijer, A.J. (2001) Fatty acid and amino acid modulation of glucose cycling in isolated rat hepatocytes, *Biochem. J.* **358**, 665–671.

Harano, Y., Kashiwagi, A., Kojima, H., Suzuki, M., Hashimoto, T., Shigeta, Y. (1985) Phosphorylation of carnitine palmitoyltransferase and activation by glucagon in isolated rat hepatocytes, *FEBS Letts.* **188**, 267–272.

Jia, Y., Qi, C., Kashireddi, P., Surapureddi, S., Zhu, Y.J., Rao, M.S., Le Roith, D., Chambon, P., Gonzalez, F.J., Reddy, J.K. (2004) Transcription coactivator PBP, the peroxisome proliferators-activated receptor (PPAR)-binding protein, is required for PPARalpha-regulated gene expression in liver, *J. Biol. Chem.* **279**, 24427–24434.

Jolly, C.A., Hubbell, T., Behnke, W.D., Schroeder, F. (1997) Fatty acid binding protein: stimulation of microsomal phosphatidic acid formation, *Arch. Biochem. Biophys.* **341**, 112–121.

Kaikaus, R.M., Bass, N.M., Ockner, R.K. (1990) Functions of fatty acid binding proteins, *Experientia* **46**, 617–630.

Kerner, J., Hoppel, C. (2000) Radiochemical malonyl-CoA decarboxylase assay: activity and subcellular distribution in heart and skeletal muscle, *Biochim. Biophys. Acta* **1486**, 1–17.

Kerner, J., Distler, A.M., Minkler, P.E., Parland, W., Peterman, S.M., Hoppel, C.L. (2004) Phosphorylation of rat liver mitochondrial carnitine palmitoyltransferase-I (CPT-I): effect on the enzyme's kinetic properties, *J. Biol. Chem.* **279**, 41104–41113.

Kim, H.K., Storch, J. (1992) Mechanism of free fatty acid transfer of rat heart fatty acid-binding protein to phospholipids membranes. Evidence for a collisional process, *J. Biol. Chem.* **267**, 20051–20056.

Liou, H.-L., Storch, J. (2001) Role of surface lysine residues of adipocyte fatty acid-binding protein in fatty acid transfer to phospholipids vesicles, *Biochemistry* **40**, 6475–6485.

Nemecz, G., Hubbell, T., Jefferson, J.R., Lowe, J.B., Schroeder, F. (1991) Interaction of fatty acids with recombinant rat intestinal and liver fatty acid-binding proteins, *Arch. Biochem. Biophys.* **286**, 300–309.

Qi, C., Zhu, Y., Reddy, J.K. (2000) Peroxisome proliferators-activated receptors, coactivators, and downstream targets, *Cell Biochem. Biophys.* **32**, 187–204.

Ramadoss, C.S., Uyeda, K., Johnston, J.M. (1976) Studies on the fatty acid inactivation of phosphofructokinase, *J. Biol. Chem.* **251**, 98–107.

Randle, P.J., Priestman, D.A., Mistry, S.C., Halsall, A. (1994) Glucose fatty acid interactions and the regulation of glucose disposal, *J. Cell. Biochem.* **55**(Suppl.), 1–11.

Richieri, G.V., Ogata, R.T., Zimmerman, A.W., Veerkamp, J.H., Kleinfeld, A.M. (2000) Fatty acid binding proteins from different tissues show distinct patterns of fatty acid interactions, *Biochemistry* **39**, 7197–7204.

Ruderman, N.B., Saha, K.A., Vavvas, D., Witters, L.A. (1999) Malonyl-CoA, fuel sensing, and insulin resistance, *Am. J. Physiol. Endocrinol. Metab.* **276**, E1–E18.

Schaap, F.G., Binas, B., Danneberg, H., van der Vusse, G.J., Glatz, J.F. (1999) Impaired long-chain fatty acid utilization by cardiac myocytes isolated from mice lacking the heart-type fatty acid binding protein gene, *Circ. Res.* **85**, 329–337.

Shen, W.-J., Sridhar, K., Bernlohr, D.A., Kraemer, F.B. (1999) Interaction of rat hormone-sensitive lipase with adipocyte lipid-binding protein, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5528–5532.

Shen, W.J., Liang, Y., Hong., R., Patel, S., Natu, V., Sridhar, K., Jenkins, A., Bernlohr, D.A., Kraemer, F.B. (2001) Characterization of the functional interaction of adipocyte lipid-binding protein with hormone-sensitive lipase, *J. Biol. Chem.* **276**, 49443–49448.

Stewart, J.M. (1990) Free fatty acids enhance the oxidation of oxymyoglobin and inhibit the peroxidase activity of metmyoglobin, *Biochem. Cell Biol.* **68**, 1096–1102.

Stewart, J.M., Blakely, J.A. (2000) Long chain fatty acids inhibit and medium chain fatty acids activate mammalian cardiac hexokinase, *Biochim. Biophys. Acta Mol. Cell. Biol. Lipids* **1484**, 278–286.

Stewart, J.M., Blakely, J.A., Johnson, M.D. (2000) The interaction of ferrocytochrome *c* with long-chain fatty acids and their CoA and carnitine esters, *Biochem. Cell Biol.* **78**, 675–681.

Stewart, J.M., Dewling, V.F., Wright, T.G. (1998) Fatty acid binding to rat liver fatty acid-binding protein is modulated by early glycolytic intermediates, *Biochim. Biophys. Acta* **1391**, 1–6.

Stewart, J.M., English, T.E., Storey, K.B. (1998) Comparisons of the effects of temperature on the liver fatty acid binding proteins from hibernator and nonhibernator mammals, *Biochem. Cell Biol.* **76**, 593–599.

Stewart, J.M., Blakely, J.A., Boudreau, N.M., Storey, K.B. (2001) A comparison of oleamide in the brains of hibernating and non-hibernating Richardson's ground squirrel (*Spermophilus richardsonii*) and its inability to bind to brain fatty acid binding protein, *J. Therm. Biol.* **27**, 309–315.

Stewart, J.M., Claude, J.F., MacDonald, J.A., Storey, K.B. (2000) The muscle fatty acid binding protein of spadefoot toad (*Scaphiopus couchii*), *Comp. Biochem. Physiol. B* **125**, 347–357.

Stewart, J.M., Slysz, G.W., Pritting, M., Muller-Eberhard, U. (1996) Ferriheme and ferroheme are isosteric inhibitors of fatty acid binding to rat liver fatty acid binding protein, *Biochem. Cell Biol.* **74**, 249–255.

Spitsberg, V.L., Matitashvili, E., Gorewit, R.C. (1995) Association and coexpression of fatty-acid-binding protein and glycoprotein CD36 in the bovine mammary gland, *Eur. J. Biochem.* **230**, 872–878.

Storch, J., Thumser, A.E.A. (2000) The fatty acid transport function of fatty acid-binding proteins, *Biochem Biophys. Acta* **1486**, 28–44.

Thompson, A.L., Cooney, G.J. (2000) Acyl-CoA inhibition of hexokinase in rat and human skeletal muscle is a potential mechanism of lipid-induced insulin resistance, *Diabetes* **49**, 1761–1765.

Thumser, A.E., Tsai, J., Storch, J. (2001) Collision-mediated transfer of long-chain fatty acids by neural tissue fatty binding proteins (FABP): studies with fluorescent analogs, *J. Mol. Neurosci.* **16**, 143–150.

Tuominen, E.K., Zhu, K., Wallace, C.J., Clark-Lewis, I., Craig, D.B., Rytomaa, M. (2001) ATP induces a conformation change in lipid-bound cytochrome c, *J. Biol. Chem.* **276**, 19356–19362.

Veerkamp, J.H., Zimmerman, A.W.J. (2001) Fatty acid-binding proteins of nervous tissue, *J. Mol. Neurosci.* **16**, 133–142.

Veerkamp, J.H., van Moerkerk, T.B., Zimmerman, A.W. (2000) Effect of fatty acid-binding proteins on intermembrane fatty acid transport studies on different types and mutant proteins, *Eur. J. Biochem.* **267**, 5959–5966.

Vork, M.M., Glatz, J.F.C., van der Vusse, G.J. (1993) Significance of cytoplasmic fatty acid-binding protein for the ischemic heart, *J. Theor. Biol.* **160**, 207–222.

Wieckowski, M.R., Brdiczka, D., Wojtczak, L. (2000) Long-chain fatty acids promote opening of the reconstituted mitochondrial permeability transition pore, *FEBS Letts.* **484**, 61–64.

Williamson, D.H., Brosnan, J.T. (1994) Concentrations of metabolites in animal tissues, *Methods Enzyme Anal.* **4**, 2266–2302.

Wolfrum, C., Borrmann, C.M., Borchers, T., Spener, F. (2001) Fatty acids and hypolipidemic drugs regulate peroxisome proliferator-activated receptors alpha- and gamma-mediated gene expression via liver fatty acid binding protein: a signaling path to the nucleus, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2323–2328.

Zoellner, H., Hou, J.Y., Hochgrebe, T., Poljak, A., Duncan, M.W., Golding, J., Henderson, T., Lynch, G. (2001) Fluorometric and mass spectrometric analysis of nonenzymatic glycosylated albumin, *Biophys. Res. Commun.* **284**, 83–89.

Intracellular Signaling in Cancer

Chittam U. Thakore[1], *Brian D. Lehmann*[1], *James A. McCubrey*[2,3], *and David M. Terrian*[1,3]
[1] *The Brody School of Medicine at East Carolina University, Anatomy & Cell Biology, Greenville, NC, USA*
[2] *The Brody School of Medicine at East Carolina University, Microbiology & Immunology, Greenville, NC, USA*
[3] *Leo Jenkins Cancer Center, Greenville, NC, USA*

1 Overview of the Prototypical Growth-signaling Circuit 116
1.1 Signaling Pathways in Normal Human Cells 116
1.2 Genetic Rewiring of Signaling Circuits in Cancer Cells 118
1.3 Identification of Activating Mutations in Cancer Genes Encoding Signaling Molecules 118

2 Ras/Raf/Mek/Erk Pathway 119
2.1 Ras 119
2.2 Raf 120
2.3 Mek 122
2.4 Erk 123
2.5 Roles of the Ras/Raf/Mek/Erk Pathway in Neoplasia 124

3 Protein Kinases Encoded by Cancer Genes 126
3.1 Protein-tyrosine Kinase Family 126
3.1.1 Receptor Protein-tyrosine Kinase Subfamily 126
3.1.2 Cytoplasmic Protein-tyrosine Kinase Subfamily 135
3.2 Protein-serine/Threonine Kinase Family 138
3.2.1 Tyrosine Kinase-like Subfamily 138
3.2.2 Homologs of Yeast Sterile 7, Sterile 11, Sterile 20 Kinase Subfamily 138
3.2.3 Calcium/Calmodulin-dependent Protein Kinase Subfamily 138
3.2.4 Containing Pka, Pkg, Pkc Subfamilies 141

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

3.2.5 Containing Cdk, Mapk, Gsk3, Clk Subfamilies 142
3.2.6 Phosphatidylinositol Kinase (Pik)-Related Subfamily 142

Acknowledgments 143

Bibliography 143
Books and Reviews 143
Primary Literature 144

Keywords

Cancer Gene
A mutated gene that has been causally implicated in oncogenesis (i.e. conferring a clonal growth advantage) by at least two independent reports confirming somatic/germline mutations in primary patient specimens.

Chromosomal Translocation
Genomic rearrangements that create chimeric genes; the most common mutation in cancer genes, often resulting in fusion proteins that include an oncogenic kinase catalytic domain and regulatory elements of an unrelated protein.

Human Kinome
A catalog of the 518 protein kinase genes encoded by the human genome and viewable at www.kinase.com.

Kinase Domain
A conserved sequence of amino acids composed of 12 subdomains and responsible for the catalytic transfer of the γ-phosphate of ATP to protein substrates.

Neoplasm
Any new growth of tissue that is both uncontrolled and progressive.

Abbreviations

ALCL anaplastic large-cell lymphoma
ALK anaplastic lymphoma kinase
ALL acute lymphoblastic leukemia
ARG Abl-related gene
ATM ataxia telangiectasia mutated kinase
CDK cyclin-dependent kinase

CML	chronic myelogenous leukemia
EGFR	epidermal growth factor receptor
ERK	extracellular signal regulated kinase
FGFR	fibroblast growth factor receptor
FLT3	FMS-like tyrosine kinase 3
GDNF	glial cell line–derived neurotrophic factor
GOF	gain-of-function
GRB	growth factor receptor bound
JAK	Janus kinase
JNK	Jun-amino-terminal kinase
JNKK	JNK kinase
LCK	T-lymphocyte protein–tyrosine kinase
MAP	mitogen-activated protein
MEK	MAP kinase–kinase
NSCLC	non–small-cell lung cancer
NTRK	neurotrophin tyrosine receptor kinase
PDGFR	platelet-derived growth factor receptor
PDK1	3-phosphoinositide-dependent protein kinase-1
PI3K	phosphoinositide 3-kinase
PIM	proviral integration of the moloney murine leukemia virus
PLCγ	phospholipase C γ
PTEN	phosphatase and tensin homolog
PTK	protein–tyrosine kinase
RET	rearranged in transformation
RTK	receptor protein–tyrosine kinase
SCF	stem cell growth factor
SGK	serum glucocorticoid regulated kinase
SH	Src homology domain
STAT	signal transducers and activators of transcription
STRAD	STE20-related adapter
STK11	serine–threonine kinase 11
TCR	T-cell receptor
TPR	translocated promoter region

Cancer genes that encode protein kinases are the most commonly mutated class of genes to be causally linked to the progressive outgrowth of malignant neoplasms. Protein kinases operate as important molecular switches that impinge, directly or indirectly, upon an integrated signaling circuitry that often becomes genetically reprogrammed, through successive mutations, to transmit many of the growth-stimulating and survival signals that govern the transformation of normal human cells into a cancerous cell mass. Of the 291 genes recently included in a comprehensive catalog of known human cancer genes, 9.3% (27 distinct genes) encode kinase domain sequences. This number far exceeds a random prediction of 6.3

genes (2%) and reinforces the hypothesis that a vast majority of human cancer cells carry genetic defects in the growth-signaling circuitry that enable these cells to grow autonomously and evade cellular senescence and death. First, we provide a general overview of a prototypical growth-signaling circuit and a detailed analysis of the canonical Ras/Raf/MEK/ERK pathway, because it has a central role in cancer cell signaling. This analysis is followed by a discussion of each of the kinases known to be mutated in cancer; we will describe the domain organization and normal function(s) of the protein, its oncogenic alterations in cancer and how this alters its activity, and identify the tumor/cancer types in which these changes are most common.

1 Overview of the Prototypical Growth-signaling Circuit

1.1 Signaling Pathways in Normal Human Cells

Protein phosphorylation is the most common posttranslational protein modification in mammalian cells and is a fundamental mechanism for the direct or indirect control of physiological processes of crucial importance to cell growth and survival. Protein phosphorylation is mediated by a superfamily of protein kinases comprising ~2% of the human genome, and a diverse collection of unrelated substrate proteins, comprising more than 10% of the genome. Collectively, these proteins control the responsiveness of a cell to its environment by coordinating complex functions such as normal embryonic development, cell division, differentiation, adhesion, motility, and metabolic activity.

All eukaryotic protein kinases have conserved a ~250 to 300 amino acid "kinase domain," composed of 12 subdomains and 12 relatively (>95%) invariant residues. The kinase domain uses the γ-phosphate of ATP or GTP to generate phosphate monoesters using two types of acceptors: (a) protein phenolic groups of tyrosine amino acids (i.e. the protein-tyrosine kinases, PTK) and (b) protein alcohol groups on serine/threonine amino acids (i.e. the protein Ser/Thr kinases). A third class, called dual specificity protein kinases, are able to phosphorylate either substrate (both Tyr and Ser/Thr). The kinase domain folds into a two-lobed three-dimensional structure separated by a deep hydrophobic cleft. When in a catalytically active conformation, the kinase domain is able to perform three interdependent functions: (a) binding and orienting the ATP/GTP phosphate donor as a complex with a divalent cation (Mg^{2+} or $Mn2^{2+}$), (b) binding and orienting the protein substrate, and (c) transferring the γ-phosphate from ATP/GTP to the acceptor hydroxyl residue of the protein substrate. The latter reaction occurs within a subdomain referred to as the *catalytic loop*. In a normal cell, the catalytic activity of most protein kinases is constrained by an intrasteric inhibition at both the ATP/GTP and substrate peptide binding sites within the kinase domain. This suppression of catalytic activity is often achieved when a "pseudosubstrate" sequence, located outside of the kinase domain and containing the consensus recognition sequence of a true

peptide substrate lacking the phosphoryl acceptor, interacts with the kinase domain and thereby blocks or distorts the active site. This form of autoinhibition is common among both the PTK and Ser/Thr kinase families and serves as just one example of the multiple layers of regulatory constraint normally exerted on a protein kinase in order to maintain a balance between the rate of cell-cycle progression and cell death. Other important examples of mechanisms capable of relieving the autoinhibition of a protein kinase include the role of protein–protein interactions (both homo- and hetero-dimerization), and activating ligands and/or second messengers. Limiting the intensity/duration of changes in protein phosphorylation are protein phosphatases and a diverse array of mechanisms controlling the biogenesis and half-life of the protein kinase.

Our understanding of how multiple protein kinases normally function within a signaling pathway has been derived largely from studies focused on delineating immediate upstream and downstream interactions through gain- and loss-of-function manipulations, arranging these interactions into orderly and conditional cascades of molecular events, and ultimately linking cell surface receptors to an effector that is capable of eliciting a biological response. Analyses of such linear pathways have provided valuable insights into the interactions between components, both within and between certain pathways.

Many of the genes that cause cancer encode receptor PTKs (RTK), a subclass of transmembrane-spanning receptor that responds to certain extracellular cues (ligands) with a disinhibition of their intrinsic PTK activity. The cytoplasmic Ser/Thr protein kinases Akt and Raf function intracellularly as important points for the convergence of many of the signals emanating from these upstream oncogenic RTKs. A

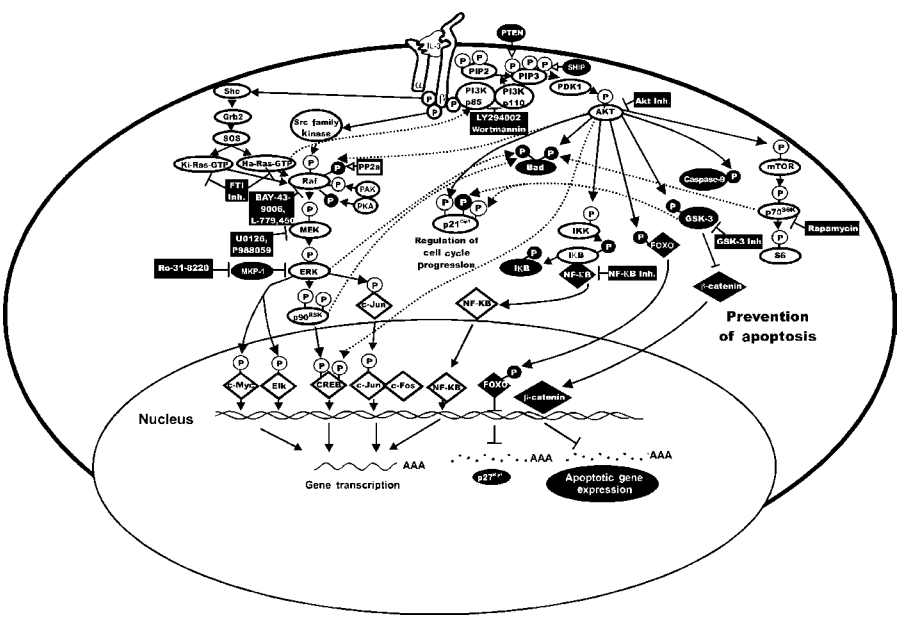

Fig. 1 Interactions between Raf/MEK/ERK and PI3K/Akt pathways and sites of action of small molecular weight signal transduction pathway inhibitors.

typical graphic representation of such a signaling network is shown in Figure 1, encompassing Akt, Raf, and several interactive signaling pathways regulating the phosphorylation of its downstream effectors, such as c-Myc, Elk, CREB, NF-kappa B, and Forkhead. These phosphorylated proteins, in turn, mediate the effects of Akt and Raf on gene transcription, and on the proliferation, growth, differentiation, and survival of the cell.

1.2 Genetic Rewiring of Signaling Circuits in Cancer Cells

Most PTKs that have thus far been implicated in cancer undergo mutations that result in a loss of their normal autoinhibition of catalytic activity. In certain signaling "cascades," such as the mitogen-activated protein (MAP) kinase pathway (Fig. 1), this loss of repression would be sufficient to trigger the constitutive activation of all downstream components, since the activity of MAP kinases is dependent on activation of the upstream kinase. However, the regulation of kinase activity is generally more complex, making it more difficult to predict how the oncogenic deregulation of a given kinase may contribute to cellular transformation. For example, the activation of 3-phosphoinositide-dependent protein kinase-1 (PDK1) is not sufficient to induce the phosphorylation of certain of its downstream substrates (e.g. p70 ribosomal S6 kinase). Rather, these substrate proteins must first undergo conformational rearrangements that render their PDK1 phosphorylation sites accessible to PDK1. Such modes of substrate-directed regulation are not frequently integrated into a typical signaling diagram. A traditional view of signaling networks provides a valuable and empirical foundation for generating meaningful hypotheses pertaining to the potential downstream effects that a constitutively active (mutant) oncogenic kinase might have within a network such as that shown in Figure 1. However, these predictions are often intuitive and can be somewhat simplistic when applied to the downstream effectors of a protein kinase such as PDK1.

Systems biology and computational modeling have now entered the field of cell biology and are changing the way scientists think about signaling networks in complex biological systems. A goal of systems biology is to understand, from a global perspective, the constellation of biological behaviors that can emerge from single/multiple mutations within the components of a cell's extensive network of signaling modules. This approach to interrogating cells, together with high-throughput genomic and proteomic screening approaches, promises to lead to the identification of unexpected "cluster groups" of genes/proteins of similar or related functions that are mutually influenced by experimentally defined mutational events. Though this approach presents a formidable challenge, it constitutes a major new frontier in the study of signaling in cancer, and has gained a significant degree of traction with the recent identification of a limited and well-characterized set of cancer genes encoding signaling molecules.

1.3 Identification of Activating Mutations in Cancer Genes Encoding Signaling Molecules

Of the 291 genes recently included in a comprehensive catalog of known human cancer genes, 27 distinct genes encode

kinase domain sequences (see Cancer Genome Project: http://www.sanger.ac.uk/CGP/). With the identification of each of these mutated kinases, it will now be possible to apply the principles of systems biology to a global analysis of how the introduction of each mutant kinase alters cellular behavior, and to exploit genomic and proteomic strategies to extract responsive cluster groups. With such an expansive inventory of new data, computational modeling can then begin to evaluate whether a specific combination of mutations ("hits") would be capable of transforming a normal cell into a cancer cell. The oncogenic themes likely to arise from such an effort can be expected to vary among cancers and according to the sequence in which genetic mutations are introduced within a given cell.

The success of the epidermal growth factor receptor (EGFR) kinase inhibitor gefitinib (Iressa®) in the treatment of non–small-cell lung cancer (NSCLC) and the ABL kinase inhibitor imatinib (Gleevec®) in treating chronic myeloid leukemia (CML) demonstrate the potential for targeting specific activating mutations in protein kinases that promote proliferative signals in cancer cells. However, it is also now apparent that the effectiveness of such treatments will be limited, and will be critically dependent on identifying the dominant genetic lesion responsible for activating a given kinase in each malignancy, and thereafter developing drugs to target this Achilles' heel. For example, only about 10 to 19% of patients with NSCLC respond to treatment with Iressa®, although a majority of these patients express the EGFR. However, a subset of patients, representing ~10% of lung cancers with tumors positive for mutations clustered around the ATP-binding pocket of the tyrosine kinase domain, respond dramatically to Iressa®. Therefore, the patient's response to therapy will be dependent on correctly matching a given kinase inhibitor to the causal oncogenic event in each case. As a prelude to this emerging effort, this article will describe each of the 27 protein kinases known to be capable of functioning as dominant human cancer genes in turn. We will describe the domain organization and normal function(s) of the protein, its oncogenic alterations in cancer and how this alters its activity, and identify the tumor/cancer types in which these changes are most common. First, however, we begin with an analysis of the canonical Ras/Raf/MEK/ERK pathway that is of such importance in cancer cell signaling.

2 Ras/Raf/Mek/Erk Pathway

The Ras/Raf/MEK/ERK pathway is a central signal transduction pathway, which transmits signals from multiple cell surface receptors to transcription factors in the nucleus. This pathway is frequently referred to as the MAP kinase pathway because it can be stimulated by mitogens, cytokines, and growth factors. The pathway can be activated by Ras stimulating the membrane translocation of Raf. This pathway also interacts with many different signal transduction pathways including phosphatidylinositol 3-kinase (PI3K)/Akt and Jak/STAT (See below).

2.1 Ras

Ras is a small GTP-binding protein, which is the common upstream molecule of several signaling pathways including Raf/MEK/ERK, PI3K/Akt, and RalEGF/Ral. The switch regions of the Ras proteins are in part responsible for the switch

between the inactive and active states of the protein. Switching between these states has been associated with conformational changes in the switch regions, which allows the binding of GTPase activating proteins and guanine nucleotide exchange factors. When Ras is active, GTP is bound, whereas when Ras is inactive GDP is bound. GTPases inactivate the Ras proteins while guanine nucleotide exchange factors activate the Ras proteins by stimulating the removal of phosphate from GTP or addition of GTP, respectively.

Different mutation frequencies have been observed between *Ras* genes in human cancer. There are three different Ras family members, Ha-Ras, Ki-Ras and N-Ras. The Ras proteins show varying abilities to activate the Raf/MEK/ERK and PI3K/Akt cascades; Ki-Ras has been associated with Raf/MEK/ERK while Ha-Ras is associated with PI3K/Akt activation. Amplification of *Ras* proto-oncogenes and activating mutations that lead to the expression of constitutively active Ras proteins are observed in approximately 30% of all human cancers. The effects of Ras on proliferation and tumorigenesis have been documented in immortal cell lines. However, antiproliferative responses of oncogenic Ras have also been observed in non-transformed fibroblasts, primary rat Schwann cells, and primary fibroblast cells of human and murine origins. This premature G_1 arrest and subsequent senescence is dependent on the Raf/MEK/ERK pathway, and has been shown to be mediated by many cell-cycle inhibitory molecules including $p15^{Ink4b}/p16^{Ink4a}$ and $p21^{Cip1}$.

2.2 Raf

The Raf protein family consists of A-Raf, B-Raf, and Raf-1, which are involved in the regulation of proliferation, differentiation, and apoptosis, induced after cytokine stimulation. The Raf proteins have three distinct functional domains: CR1, CR2, and CR3. The CR1 domain is necessary for Ras binding and subsequent activation. The CR2 domain is a regulatory domain. CR3 is the kinase domain. Deletion of the CR1 and CR2 domains produces a constitutively active Raf protein due, in part, to the removal of phosphorylation sites, which serve to negatively regulate the kinase in the CR2 domain.

The *A-Raf* gene is located on the X chromosome in humans and produces a 68-kDa protein. The highest level of A-Raf expression in the adult is in the urogenital tract. A-Raf deletion results in postnatal lethality, attributed to neurological and gastrointestinal defects. The role of A-Raf has remained elusive. A-Raf is the weakest Raf kinase in terms of activation of ERK activity. ERK activation occurs in A-Raf knock-out mice indicating that other Raf isoforms can compensate for this deficiency. Some studies have indicated that A-Raf has an important role in stimulating the growth of hematopoietic cells. Moreover, we have observed that overexpression of activated A-Raf will abrogate the cytokine-dependence of hematopoietic cells.

The *B-Raf* gene is located on chromosome 7q34. B-Raf encodes a 94-kDa protein. The highest expression of B-Raf is in the testes and neuronal tissue. The *B-Raf* gene produces splice variants, the physiological roles of which have not yet been elucidated. The loss of B-Raf expression in mice results in intrauterine death between days 10.5 and 12.5. The B-Raf knock-out mouse embryo displays enlarged blood vessels and increased apoptosis of differentiated endothelial cells. This is an indication that Raf kinases can regulate

apoptosis. B-Raf and Raf-1 activity are negatively regulated by Akt phosphorylation. Akt also has other effects on the presentation of apoptosis, and will be discussed in more detail later. In contrast, other studies have shown that overexpression of B-Raf in Rat-1 cells results in decreased apoptosis because of inhibition of caspase activity. Historically, B-Raf is the strongest Raf in terms of induction of MEK activity, as determined by *in vitro* kinase assays. B-Raf activation is different from either Raf-1 or A-Raf activation; it requires the phosphorylation of only one regulatory residue, whereas activation of Raf-1 and A-Raf require two phosphorylation events. B-Raf is also regulated by Akt and the serum glucocorticoid regulated kinase (SGK). Recently it has been shown that B-Raf may be required in the activation of Raf-1, and that B-Raf and Raf-1 may be induced at different times following stimulation. Moreover, the three different Raf proteins may have different subcellular pools, and Raf-1 and A-Raf may be localized, in some cases, to the mitochondrion. The roles of B-Raf in human neoplasia will be discussed later.

The *Raf-1* gene is located on chromosome 3p25 and produces a 74-kDa protein. The *Raf-1* proto-oncogene was the first *Raf* gene to be cloned. It is the cellular homolog of *v-Raf* contained in the naturally occurring acute retrovirus 3611-MSV. *Raf-1* is ubiquitously expressed in adult tissues with highest expression in muscle, cerebellum, and fetal brain. It is the most studied Raf isoform. A dominant-negative version of Raf-1 inhibits Ha-Ras-induced transformation and tumor formation. Raf-1 has important roles in apoptosis as it phosphorylates and inactivates Bad. Raf-1 phosphorylates and co-immunoprecipitates with Bcl-2 and also regulates Bag and Bad expression, in BCR/ABL expressing cells. Recently, Raf-1 has been proposed to have roles independent of MEK/ERK that are often involved in regulating cell-cycle progression and apoptosis. These new roles for Raf-1 will be discussed below.

The activation of the three Raf isoforms is complex and not totally understood. Ras is known to activate B-Raf independently of Src activity while Raf-1 and A-Raf require an Src-like kinase for complete activation. An Src-like kinase phosphorylates Tyr340 and Tyr341 on Raf-1 and Tyr301 and Tyr302 on A-Raf, which are not present in B-Raf. B-Raf contains aspartic acid at the corresponding residues. Aspartic acid residues are negatively charged and are believed to confer a constitutively active configuration at this site. B-Raf is constitutively phosphorylated on Ser445 (a site equivalent to Ser338 in Raf-1), which is phosphorylated as a result of Ras activation. The Ras binding domains of B-Raf and Raf-1 have a greater affinity for Ras than A-Raf. This suggests that A-Raf may have a primary activator other than Ras. Phosphorylation of the activation loop of Raf-1 and B-Raf is necessary, but not sufficient, for activation. The deletion or mutation of inhibitory phosphorylation sites present at certain residues by site-directed mutagenesis activates Raf proteins. Akt also phosphorylates Raf-1 at Ser259, which has been associated with inhibition of Raf-1 activity.

Several protein kinases are known to regulate Raf-1 activity. cAMP-dependent protein kinase inhibits Raf-1. cAMP activates PKA, resulting in phosphorylation of Raf-1 on Ser43 and Ser621, and inhibiting Raf-1 activity. This contrasts with B-Raf activation in response to cAMP-dependent protein kinase. Raf-1 is activated through $p21^{Cip1}$-associated protein. Activation of conventional protein kinase C isoforms (α, β, and γ) stimulates Raf-1 activity in

a baculovirus system. Protein kinase Cε is associated with c-N-Ras and Raf-1 and is necessary for activation of Raf-1 by phorbol 12myristate 13acetate in fibroblasts. RKIP, a Raf-1-interacting protein, inhibits Raf-1 activation of MEK1. 14-3-3 proteins are known to bind Raf. The interaction of 14-3-3 proteins with phosphorylated Ser259 and Ser621 inhibits Raf-1. Protein phosphatase 2A can remove the phosphate on Ser259, allowing Raf to become disassociated from 14-3-3. This allows Raf-1 to assume a conformation in which it can be phosophorylated and activated by other kinases including $p21^{Cip1}$-associated protein and Src family kinases.

Raf-1 was originally characterized as phosphorylating MEK1, but B-Raf is the primary MEK1 activator in bovine brain extracts, NIH 3T3 cells, PC12 cells and other cells. Raf-dependent activation of p90 ribosomal S6 kinase ($p90^{Rsk}$) and NF-κB are activated by an Raf-1 mutant, which no longer binds or phosphorylates MEK1. This suggests that Raf-1 has physiological substrates other than MEK1. Recently, the possible roles of Raf-1 in the Raf/MEK/ERK signal transduction pathway have become a matter of controversy due to the discovery that B-Raf is the more potent activator of MEK and that many of the "functions" of Raf-1 still persist in Raf-1 knock-out mice, but not in B-Raf knock-out mice. Raf-1 has been postulated to have important roles in cell-cycle progression, activation of the p53 and NF-κB transcription factors, and the prevention of apoptosis. Raf-1 has been postulated to have nonenzymatic functions and serve as a docking protein. Raf-1 has been proposed to have important functions at the mitochondrial membrane. Interestingly, it has been shown that Bcr-Abl (see Sect. 3.1.2) may interact with Raf-1 to alter the distribution of Bag at the mitochondrial membrane and hence regulate (prevent) apoptosis in hematopoietic cells.

Recently, the mechanism of B-Raf regulation has been more intensively investigated. B-Raf activation may occur through the GTP-binding protein Rap-1, which is in turn activated by RA-guanine nucleotide exchange factor-1 and Src. cAMP-dependent protein kinase may activate Src in some cells, resulting in Rap-1 activation. There are three regulatory residues on B-Raf – Ser444, Thr598, and Ser601. The Ser/Thr kinases that phosphorylate these residues are not known. Activated B-Raf can interact with Ras, which in turn results in the activation of Raf-1. B-Raf can activate MEK, which results in the activation of ERK and downstream kinases and transcription factors. Rheb and 14-3-3 proteins can bind and inhibit B-Raf activity. B-Raf can also be negatively affected by phosphorylation by either SGK or Akt. Akt has been shown to phosphorylate B-Raf on three residues – Ser364, Ser428, and Thr439 (see below). However, Akt may have greater affinity for other substrates besides B-Raf. SGK has been proposed to be a more relevant kinase for regulating B-Raf through phosphorylation of Ser364. B-Raf may also activate Raf-1 *via* Ras-GTP. While Rap-1 activates B-Raf, it may also serve to inactivate Raf-1. Thus the pathways for activation and inactivation of B-Raf and Raf-1 are complex, and may at times appear contradictory.

2.3 Mek

MEK proteins are the primary downstream targets of Raf. The *MEK* family of genes consists of five genes: *MEK1, MEK2, MEK3, MEK4, and MEK5*. This family of dual specificity kinases has both Ser/Thr and Tyr kinase activity. The structure

of MEK consists of an amino-terminal negative regulatory domain and a carboxy-terminal MAP kinase-binding domain that is necessary for binding and activation of ERKs. Deletion of the regulatory MEK1 domain results in constitutive MEK1 and ERK activation.

MEK1 is a 393 amino acid protein with a molecular weight of 44 kDa. MEK1 is modestly expressed in embryonic development and is elevated in adult tissue, with the highest levels detected in brain tissue. Knock out of functional MEK1 results in lethality due to placental vascularization problems. Mice with dominant-negative MEK1 have also been generated; these mice were viable although defects in T-cell development occurred. MEK1 requires phosphorylation of Ser218 and Ser222 for activation. The substitution of these residues with aspartate or glutamic acid led to an increase in activity and foci formation in NIH3T3 cells. Mutated MEK1 constructs demonstrate that the activation of ERK does not require a functional MEK1 kinase domain. Replacement of the amino-terminus of MEK1 with the hormone-binding domain of the estrogen receptor produces a construct with kinase activity responsive to the presence of estrogen analogs. This construct is an invaluable tool in research into MEK1 signaling. Studies with this construct demonstrated that activated MEK1 could abrogate the cytokine dependency of certain hematopoietic cells. Constitutive activity of MEK1 in primary cell culture promotes senescence and induces p53 and $p16^{INK4a}$, whereas the opposite was observed in immortalized cells or cells lacking either p53 or $p16^{INK4a}$. Constitutive activity of MEK1 inhibits NF-κB transcription, by negatively regulating $p38^{MAPK}$ activity. Pharmaceutical companies have developed inhibitors of MEK. The two most widely used are U0126 and PD98059, because they are commercially available. These two inhibitors have IC_{50}s of 70 nM and 2 μM, respectively. PD98059 inhibits activation, while U0126 inhibits activity. Thus, these two inhibitors have a noncompetitive mechanism of inhibiting ERK activity.

2.4 Erk

The main physiological substrates of MEK are the members of the ERK (extracellular signal regulated kinase), or MAP kinase, family of genes. The ERK family consists of four distinct groups of kinases: ERK, Jun amino-terminal kinases (JNK1/2/3), $p38^{MAPK}$ (p38 $\alpha/\beta/\gamma/\delta$), and ERK5. In addition, there are ERK3, ERK4, ERK6, ERK7, and ERK8 kinases, which are related to ERK1 and ERK2, but have different modes of activation, with biochemical roles that are not as well characterized. The ERK 1 and 2 proteins are the most studied with regards to Raf signaling in hematopoietic cells.

ERK1 and *ERK2* encode 42 and 44 kDa proteins, respectively. These proteins were originally isolated by their ability to phosphorylate microtubule-associated protein 2. ERKs are activated through dual phosphorylation of threonine and tyrosine residues by MEK1 kinases. This dual phosphorylation activates ERK; however, ERK activity is downregulated by serine/threonine phosphatases, tyrosine phosphatases, or dual specificity phosphatases. ERK dimerization occurs subsequent to phosphorylation. This dimerization maintains the activation of ERK and promotes nuclear localization of the protein. Activated ERK preferentially phosphorylates serine/threonine residues preceding a proline.

Mice lacking ERK1 have defective T-cell development, similar to transgenic mice containing a dominant-negative MEK1. Recently, different roles have been suggested for ERK1 and ERK2. While these two proteins are often detected at similar levels, their expression patterns may change. ERK1 has been postulated to inhibit ERK2 expression. Increased ERK2 activity is associated with cell proliferation. Interestingly, ERK1 has been associated with cognitive brain functions including learning.

2.5 Roles of the Ras/Raf/Mek/Erk Pathway in Neoplasia

Mutations at three different Ras codons (12, 13, and 61) convert Ras into a constitutively active protein. These point mutations can be induced by environmental mutagens. Given the high level of mutations that have been detected at Ras, this pathway has always been considered a key target for therapeutic intervention. Approximately 30% of human cancers have Ras mutations. Ras mutations are frequently observed in certain hematopoietic malignancies including myelodysplastic syndromes, juvenile myelomonocytic leukemia, and acute myeloid leukemia. Ras mutations are often one step in tumor progression and mutations at other genes (chromosomal translocations, gene amplification, and tumor suppressor gene inactivation) have to occur for a complete malignant phenotype to be manifested. Pharmaceutical companies have developed many farnesyl transferase inhibitors, which suppress the farnesylation of Ras, preventing it from localizing to the cell membrane.

As stated previously, there are three different *Ras* genes – *Ki-Ras*, *Ha-Ras*, and *N-Ras*. The biochemical differences between these Ras proteins have remained elusive. *Ki-Ras* mutations have been more frequently detected in human neoplasia than *Ha-Ras* mutations. Ras has been shown to activate both the Raf/MEK/ERK and the PI3K/Akt pathways. Thus, mutations at Ras should theoretically activate both pathways simultaneously. Ras mutations have a key role in malignant transformation as both of these pathways can prevent apoptosis as well as regulate cell-cycle progression. Recently it was shown that there is specificity in terms of the ability of Ki-Ras and Ha-Ras to induce the Raf/MEK/ERK and PI3K/Akt pathways. Ki-Ras preferentially activates the Raf/MEK/ERK pathway, while Ha-Ras preferentially activates the PI3K/Akt pathway. Therefore, if Ras inhibitors could be developed which would specifically inhibit one particular Ras protein, it might be possible to inhibit one of the downstream pathways. This might, under certain circumstances, be advantageous. Furthermore, decreases in ERK expression may affect differentiation responses. Thus in certain tumors, one might desire to inhibit the effects that Ras has on the PI3K/Akt pathway as opposed to the effects it has on Raf. Targeting of Ha-Ras as opposed to Ki-Ras might inhibit apoptosis suppression by Ha-Ras, but not the effects Ki-Ras has on cell-cycle progression and differentiation.

Overexpression of the Raf/MEK/ERK cascade has been frequently observed in human neoplasia. A prime consequence of this activation may be the increased expression of growth factors, which can potentially further activate this cascade by an autocrine loop. Many cytokine and growth factor genes contain transcription factor binding sites, which are bound by transcription factors (Ets, Elk, Jun, Fos, CREB) that are often activated by the

Raf/MEK/ERK cascade. Identification of the mechanisms responsible for activation of this cascade has remained elusive. Genetic mutations at Raf, MEK, or ERK were thought to be relatively rare in human neoplasia. For many years, numerous scientists were of the opinion that the activation of the Raf/MEK/ERK cascade was mainly due to mutations at Ras, and hence studies aimed at elucidating the mechanisms of Ras activation were launched.

While mutations at the *Raf* gene in human neoplasia have been detected, they have not until recently gained the clinical importance that Ras mutations readily achieved. Because of more innovative, high-throughput DNA sequencing, scientists have recently discovered that the *B-Raf* gene is frequently mutated in certain cancers including hematopoietic tumors. Approximately 60% of the melanomas surveyed in one study were observed to have mutations at *B-Raf* (see Sect. 3.2.1). This result provides relevance to the investigation of signal transduction pathways because, by understanding how *B-Raf* is activated, one Ras-dependent and one Ras-independent event, scientists could predict why a single missense mutation in *B-Raf* permitted ligand-independent activation, whereas in similar mutation events it could not be predicted whether the mutation events would result in either Raf-1 or A-Raf activation, as they requires multiple activation events. Interestingly, in 22 tumors (melanomas, colorectal, and NSCLC tumors) that were examined, there were 10 with mutations at *B-Raf* and 10 with mutations at Ras. Out of these two tumors had mutations at *B-Raf* and Ras; two did not have mutations at either gene. Thus *B-Raf* transformation does not appear to require Ras, and many tumors had mutations at one or the other, but not both, genes. Recently, it has been suggested that B-Raf mutations occur during the process of progression of the tumor, and not during the establishment of the tumor. This was suggested by analysis of *B-Raf* mutations in different developmental stages of melanomas.

Raf inhibitors have been developed and some are being used in clinical trials. Raf-1 has at least 13 regulatory phosphorylation residues. This makes inhibition of Raf activity difficult, as certain phosphorylation events stimulate Raf activity while others inhibit and promote Raf association with 14-3-3 proteins, rendering it inactive though present in the cytoplasm. Certain Raf inhibitors were developed which inhibit the Raf kinase activity as determined by assays with purified Raf proteins and substrates (e.g. MEK). Some Raf inhibitors may affect a single Raf isoform (e.g. Raf-1), others may affect Raf proteins, which are more similar (Raf-1 and A-Raf), while other Raf inhibitors may affect all three Raf proteins (Raf-1, A-Raf and B-Raf). We have observed that the L-779,450 inhibitor suppresses the effects of A-Raf and Raf-1 more than those of *B-Raf*. Knowledge of the particular *Raf* gene that is mutated or overexpressed in certain tumors may provide critical information for planning treatment for the patient. Inhibition of certain *Raf* genes might prove to be beneficial, while inhibition of other *Raf* genes under certain circumstances might prove detrimental. Thus, the development of unique Raf inhibitors may prove useful in human cancer therapy.

Activation of Raf is complex and requires numerous phosphorylation and dephosphorylation events. Prevention of Raf activation by targeting kinases (e.g. Src or Akt) and phosphatases involved in Raf activation may be a mechanism to inhibit/regulate Raf activity. It is worth noting that some of these

kinases normally inhibit Raf activation (e.g. Akt). A major limitation of this approach would be the specificities of these kinases and phosphatases. Inhibiting these kinases/phosphatases could result in activation or inactivation of other proteins and would have other effects on cell physiology.

Dimerization of Raf proteins is critical for their activity. We often think of a single Raf protein carrying out its biochemical activity. However, Raf proteins dimerize with themselves and other Raf isoforms to become active. Drugs such as coumermycin, which inhibit Raf dimerization, and others such as geldanmycin, which prevent interaction of Raf with 14-3-3 proteins, suppress Raf activity. Various Raf isoforms may dimerize and result in chimeric molecules, which may have different biochemical activities. Little is known about the heterodimerization of Raf proteins.

Downstream of Raf lies MEK. Currently, it is believed that MEK1 is not frequently mutated in human cancer. However, aberrant expression of MEK1 has been detected in many different types of cancer, and mutated forms of MEK1 will transform fibroblast, hematopoietic, and other cell types. Useful inhibitors to MEK have been developed, which display high degrees of specificity. The successful development of MEK inhibitors may be due to the relatively few phosphorylation sites on MEK involved in activation/inactivation. MEK inhibitors are in clinical trials.

Downstream of MEK lies ERK. To our knowledge no small-molecular-weight ERK inhibitors have been developed yet. However, inhibitors to ERK could prove very useful as ERK can phosphorylate many targets (Rsk, c-Myc, Elk, etc.), which have growth-promoting effects. There are at least 2 ERK molecules regulated by the Raf/MEK/ERK cascade, ERK1 and ERK2. Little is known about the different *in vivo* targets of ERK1 and ERK2. However, ERK2 has been postulated to have pro-proliferative effects while ERK1 has antiproliferative effects. Development of specific inhibitors to ERK1 and ERK2 might eventually prove useful in the treatment of certain diseases.

The MAP kinase phosphatase-1 (*DUSP1*) removes the phosphates from ERK. MAP kinase phosphatase-1 is mutated in certain tumors and could be considered to be a tumor suppressor gene. An inhibitor to this phosphatase has been developed (Ro-31-8220).

3 Protein Kinases Encoded by Cancer Genes

3.1 Protein-tyrosine Kinase Family

3.1.1 Receptor Protein-tyrosine Kinase Subfamily

More than half (55%) of the protein kinases included in the current census of human cancer genes are RTKs. Somatic chromosomal translocations that result in the fusion of a given tyrosine kinase domain of one gene with the regulatory region of another unrelated gene are most common. In many instances, a number of different genes will form chimeric transcripts with the same *RTK* gene. Generally, the fusion proteins encoded by these translocated cancer genes phosphorylate their protein substrates in a ligand-independent manner. Such mutations are often found in lymphomas, leukemia, epithelial neoplasms, and solid tumors.

ALK The anaplastic lymphoma kinase (*ALK*) gene, located on human chromosome 2p23, is expressed only in the

small intestine, testis, and brain, and encodes a Type Ia membrane protein (carboxy-terminal kinase domain oriented intracellularly) with a glycosylated amino-terminal extracellular region that binds the ligands pleiotrophin and midkine. Ligand-binding specificity is dependent on two cysteine-rich clusters that are adjacent to the major ligand-binding domain. The extracellular domain of ALK also contains a low-density lipoprotein-A module, suggesting that ALK may also interact with low-density lipoproteins. Other motifs present in the extracellular domain include a glycine-rich patch and MAM domain (Meprin/A-5/protein-tyrosine phosphatase Mu), which may play a role in mediating certain cell–cell interactions. Like other RTKs, discussed below, tyrosine residues in the intracellular domain of ALK regulate its intrinsic kinase activity, and provide phosphorylation-dependent recruitment sites for the Src homology 2 (SH-2) domains of various adapter proteins and enzymes that propagate intracellular signaling cascades. For ALK, these include the insulin receptor substrate-1, Shc protein, and phospholipase C-γ. The adapter protein Shc forms complexes with GRB2 (growth factor receptor-bound protein 2) and the Son of Sevenless protein to activate the Ras/Raf/MEK/ERK signaling cascade (Fig. 1). Although ALK-deficient mice do not display any marked developmental defects, the *drosophila* homolog (i.e. *DAlk*) is required for normal development of the gut musculature. A common feature among all ALK chimeric proteins is the presence of an oligomerization domain in the partner protein and oligomerization-mediated kinase activation. Although the intracellular localization of these fusion proteins are predominantly determined by the amino-terminal fusion partner, all fusion proteins result in aberrant activation of multiple downstream signaling cascades that are responsible for cellular transformation. Approximately half the non-Hodgkin's lymphomas classified as anaplastic large-cell lymphomas (ALCL) are associated with a common somatic chromosomal translocation, t(2;5)(p23;q35). This chimeric gene encodes an oncogenic fusion protein comprised of the complete cytosolic and kinase domains of ALK joined to the amino-terminal 117 amino acid residues of the nucleolar phosphoprotein nucleophosmin. The transforming activity of the fusion protein (nucleophosmin-ALK) requires the activation of its kinase function, a result of the oligomerization mediated by the nucleophosmin segment. Although less common, other ALK fusion partners include TPM3, TFGxl, TFGs, ATIC, CLTC, MSN, TPM4, MyH9, RANBP2, ALO17, and CARS. Constitutive ALK signaling activates pathways that either enhance proliferation or target the PI3K/Akt survival pathway. Proliferation is enhanced by activation of the MAP kinase pathway through Ras or phospholipase C-γ (PLC-γ). PLC-γ-mediated hydrolysis of membrane phospholipids produces localized increases in inositol triphosphate and diacyglycerol, which, in turn, stimulate the release of Ca^{2+} from intracellular stores and the activity of the Ser/Thr protein kinase C. Although the precise nature of the interaction between PI3K and ALK is unknown, a potential complex includes nucleophosmin-ALK and the regulatory p85 subunit of PI3K with GAB2, Shc, and/or CrkL, an interaction recently identified in the human Karpass 299 cell line.

EGFRs The EGFR subfamily of RTKs encompasses four closely related receptors, EGFR (also known as ErbB-1 or HER1),

ErbB-2 (HER2), ErbB-3 (HER3), and ErbB-4 (HER4), which drive proliferation, differentiation, survival, and cell motility. Of these, *EGFR* and *ErbB-2*, located on chromosomes 7p12 and 17q21, respectively, are best characterized for their role in human cancers. EGFRs are comprised of an extracellular domain (containing two receptor ligand-binding domains and a furin-like domain), a transmembrane region, and the cytoplasmic domain (containing the catalytic kinase domain). EGFR signaling is normally triggered by the binding of its natural ligand, such as EGF, betacellulin, heparin-binding EGF, and transforming growth factor-α. These EGFR ligands are all membrane-bound proteins containing one or more EGF-like motifs that interact with the receptor. Members of the EGFR subfamily are capable of forming as many as 10 different homodimers and heterodimers when occupied by an appropriate ligand and the affinity of each EGFR combination differs for each ligand. For example, an EGFR homodimer binds EGF, betacellulin, and heparin-binding EGF with about the same affinity, whereas an ErbB2:3 heterodimer preferentially binds the heregulin β isoform versus the heregulin α isoform. Receptor dimerization initiates the transphosphorylation of the PTK activation loop, which is required for optimal kinase function. Multiple tyrosine residues in the cytoplasmic tail are also phosphorylated, permitting the recruitment of adapter signaling proteins. Adapter proteins containing SH-2 domains recognize and bind phosphorylated Tyr residues and feed into signaling pathways including the Ras/Raf/MEK/ERK, PI3K/Akt, and PLCγ/PKC pathways. The result is the activation of proliferative and cell-survival signals. Potential biological responses are complex and diverse, depending upon receptor composition, ligand identity, and duration of signaling. The *EGFR (ErbB-1)* and *ErbB-2* genes are commonly amplified, rearranged, or mutated in many human cancers, including head and neck cancer, NSCLC, breast cancer, and gliomas. For example, EGFR is overexpressed in 40to 80% of NSCLCs. As discussed above (see Sect. 1.3), however, only a subset of these patients responds to the EGFR kinase inhibitor Iressa®. Therefore, overexpression by itself is not sufficient to predict responses, and the evidence suggests that activating mutations are also required. In NSCLC, these activating mutations are somatic and clustered around the ATP-binding pocket of the EGFR catalytic domain. All mutations are heterozygous and consist either of small, in-frame deletions, or amino acid substitutions within exons 17 to 19. Importantly, these mutations lead to increased responsiveness to EGF and also confer susceptibility to the inhibitor Iressa®. In contrast, rearrangements within the extracellular domain of EGFR are common in gliomas and colorectal cancer, and are associated with kinase activation. A number of monoclonal antibodies targeting the extracellular ligand-binding region (e.g. cetuximab) have been developed to specifically block such mutant EGFRs. The *erbB-2* gene is amplified in 20 to 30% of breast and ovarian cancers. ErbB-2 does not directly bind to ligands and yet its overexpression promotes oncogenesis through decreases in attenuation pathways and signal potentiation. ErbB-2 amplifies signaling indirectly by increasing the ligand affinity of its heterodimeric binding partners (i.e. EGFR, ErbB-3, and ErbB-4) and prolonging the surface expression of these RTKs by slowing their rate of internalization. In addition, ErbB-2 enhances EGFR recycling and reduces lysosomal targeting.

FGFRs The fibroblast growth factor receptor (FGFR) family is comprised of four high-affinity RTKs (FGFR1-4) that share several structural features in common, including two or three immunoglobulin-like domains and an acidic box in their amino-terminal extracellular domains, a single transmembrane domain, and an intracellular PTK domain split by a short (14 amino acid) insertion. While three of these RTKs have now been classified as cancer genes (*fgfr1–3*), the fourth (*fgfr4*) is also known to be overexpressed in mammary and ovarian carcinomas. The chromosomal locations of the *fgfr1-fgfr3* genes are 8p11.2, 10q26, and 4p16.3, respectively. Distinct FGFR mutations have been shown to mediate a variety of autosomal disorders of bone growth and embryonic development and a discussion of each would be beyond the scope of the present article. However, it is important to note that the majority of mutant FGFRs identified to date possess elevated PTK activity relative to their wild-type counterparts, resulting in ligand-independent activation. While mutations in the extracellular domain of FGFRs have been associated with dwarfism, Pfeiffer syndrome, and other developmental abnormalities, the most common mutations associated with human cancer involve a subset of acquired (somatic) translocations that target chromosome band 8p11 (*fgfr1*), and will serve as an instructive example of how the constitutive activation of FGFRs might contribute to oncogenesis. Crystal structures of an inactive FGFR1 reveal that the activation loop normally occludes substrate tyrosine binding to the catalytic site. In solution, equilibrium exists between this inactive conformation (called *cis* inhibition) and the active loop conformation. The family of fibroblastic growth factors (FGFs) capable of activating the FGFRs includes at least 24 members sharing 30 to 50% amino acid sequence homology, and are designated FGF-1 through FGF-24. Certain FGF family members are considered to be oncogenic (e.g. FGF-3, FGF-4, FGF-5, and FGF-6) and all have been implicated in the control of a variety of biological processes. In response to ligand occupancy, FGFRs undergo a ligand-induced conformational shift that promotes dimerization and *trans*phosphorylation of tyrosine residues in the activation loops of the binding partner. This leads to a movement of the activation loop away from the catalytic site that permits substrate access, while *trans*phosphorylation provides binding sites for recruitment of signal transducers and activators of transcription (STATs) *via* their SH-2 domains. In acute myelogenous leukemia, chromosomal rearrangements lead to a fusion between the ZNF198 mRNA at exon 17 and FGFR1 at exon 9, containing the kinase domain. *ZNF198* is a widely expressed gene that encodes a 155-kDa protein with five Zn finger–related motifs (known as MYM domains) that are retained in the ZNF198-FGFR1 chimeric protein. ZNF198-FGFR1 has constitutive PTK activity and is thought to rely on STAT5 activation in the transformation of hemopoietic cells. FGFR1 phosphorylation of STAT5 promotes dimerization, and dimeric STATs are exported to the nucleus, where these latent transcription factors associate with the promoters of early response genes such as *c-fos*, *c-jun*, and *c-myc* and initiate transcription of cell-cycle regulatory genes. FGFRs are mitogenic for endothelial cells and play a significant role in angiogenesis by stimulating the production and release of agents that break down basement membranes.

FLT3 FMS-like tyrosine kinase 3 (FLT3) is a member of the platelet-derived growth

factor receptor (PDGFR) family and is located on human chromosome 13q12. Four of the five members of the PDGFR gene family are regarded as authentic cancer genes. These include KIT and the PDGRs type A (PDGFRA) and PDGFR type B, in addition to FLT3. FLT3 contains five extracellular ligand-binding immunoglobulin-like (Ig) domains and a cytoplasmic split tyrosine kinase motif. FLT3 is expressed in hematopoietic cells, placenta, gonads, and brain. The ligand for FLT3 is a hematopoietic growth factor termed FL (Flt3 ligand). The *fl* gene encodes a transmembrane protein, but alternative splicing can result in the generation of a soluble form of the FLT3 ligand. Flt3 binds both the membrane-bound and soluble forms of FL. FL binding activates tyrosine kinase activity and stimulates proliferation. Although FL does not efficiently induce proliferation of normal cells by itself, it does reinforce signaling *via* other growth factor receptors. FLT3 deficient mice develop normally, although they possess specific deficiencies in primitive B-lymphoid progenitors, and $FLT^{-/-}$ stem cells are impaired in their ability to reconstitute T cells and myeloid cells when transplanted into bone marrow. Thus, FLT3 signaling is critical for normal development of stem cells and B cells. FLT3 signaling is mediated by phosphorylation-dependent interactions between its carboxy-terminal cytoplasmic tail and enzymes and adapter proteins, such as the regulatory p85 subunit of PI3K, RAS/GTPase activating protein, PLC-γ, Shc, and GRB2. The activity of FLT3 is normally autoinhibited by intramolecular interactions between the cytosolic juxtamembrane domain and kinase domain, preventing ATP and substrate from binding to the activation loop. Although FLT3 is not normally expressed in mature hematopoietic cells, somatic mutants of the *flt3* gene are frequently expressed in hematological malignancies. Examples include FLT3 expression in 70% of acute myelogenous leukemias, about 30% of acute lymphoblastic leukemias (ALL), and some CML cases in lymphoid blast crisis. Although a number of different activating mutations of the *flt3* gene have been documented, two types are by far the most common. Small tandem duplications of amino acids within the juxtamembrane domain and point mutations within the activation loop occur in 24 and 7% of acute myelogenous leukemia patients, respectively. Although duplications within the juxtamembrane domain are highly variable, each is thought to move this cytosolic domain away from the activation loop, triggering kinase activity. Point mutations of the nucleotides encoding Asp835, within the FLT3 activation loop, similarly disrupt the autoinhibition of kinase activity. Since FLT3 mutations tend to be associated with a poor prognosis in leukemia, it is thought that mutant FLT3 may cooperate with other leukemia oncogenes to confer a more aggressive phenotype of this disease.

KIT The proto-oncogene *c-kit*, located on human chromosome 4q12, encodes the stem cell growth factor (SCF) receptor KIT/SCFR, a human homolog of the *v-kit* Hardy-Zuckerman 4 feline sarcoma viral oncogene. The domain organization of KIT has been conserved by all members of the PDGFR family of glycosylated, single-pass transmembrane proteins containing five extracellular immunoglobulin-like domains and a split intracellular PTK domain. The cytosolic juxtamembrane region appears to provide a crucial level of KIT autoinhibition, as mutations within this region are often associated with constitutive, ligand-independent, kinase activity.

The importance of this receptor in normal hematopoiesis and melanogenesis was established by the finding that naturally occurring germline loss-of-function mutations result in developmental defects in both processes. In a normal cell, KIT activation stimulates the proliferation and differentiation of cells such as hematopoietic stem cells, mast cells, and germ cells through the canonical RTK pathway; that is, SCF binding initiates KIT dimerization and the transphosphorylation of the dimeric KIT subunits. These phosphotyrosine residues enable the docking of SH-2 domain–containing signaling molecules including GRB2, SHP-1 (a protein-tyrosine phosphatase expressed in hematopoietic cells), and members of the Src family of cytosolic tyrosine kinases. The dephosphorylation of KIT by SHP-1 functionally limits the activity of this RTK and the half-life of the receptor is limited by the adapter protein APS, which promotes ubiquitin-mediated degradation of KIT through the recruitment of E3 ubiquitin ligase c-Cbl. Thus, both SHP-1 and APS may function as tumor suppressors for KIT-induced malignancies. Downstream of KIT, PI3K-dependent activation of Akt and the phosphorylation/inactivation of the proapoptotic protein Bad appear to be crucial mediators of KIT-induced cellular transformation. More than 30 gain-of-function (GOF) mutations in *KIT* result in a constitutively active kinase and have now been identified in a variety of different human malignancies. These mutations can generally be found in one of two regions. Deletions of a few codons within the autoinhibitory juxtamembrane region, encoded by exon 11, are observed in gastrointestinal stromal tumors and single amino acid changes in exon 17, in the carboxy-terminal half of the kinase domain, are associated with mast cell leukemia and to a lesser extent with T-cell lymphoma and acute myelogenous leukemia. Other GOF point mutations have been identified in exon 13, encoding the amino-terminal half of the KIT kinase domain. In addition to these somatic mutations, GOF mutations of the *KIT* gene have also been associated with germline tumors such as testicular seminomas and ovarian dysgerminomas. Experimentally, the co-expression of KIT and its ligand SCF has been shown to enhance the autocrine growth of cells and tumorigenesis. Approximately 80% of the patients with activating *KIT* mutations in exon 11 respond to the EGFR inhibitor Gleevec®.

MET The human *Met* (HGF receptor) gene is located on chromosome 7q21-q31 and is expressed in a variety of organs and cell types. Following its glycosylation, the MET precursor protein (190 kDa) is cleaved into a 50 kDa α-chain and 140-kDa β-chain that are linked *via* disulfide bonds. After membrane insertion of the β-subunit, these heterodimers form a mature receptor consisting of an extracellular α-subunit joined to a single-pass membrane protein (i.e. β-subunit) with a large ectodomain and intracellular PTK domain. Stimulation of MET by its natural ligand, hepatocyte growth factor/scatter factor, can lead to a broad range of responses in normal cells, such as increased proliferation, scattering, enhanced motility, and angiogenesis. The downstream effects of MET activation include the stimulation of both the PI3K and MAP kinase pathways (see Fig. 1). A series of autophosphorylation events, including phosphorylation within the β-chain activation loop at Tyr1234 and Tyr1235, and a substrate docking site at Tyr1249 and Tyr1356, stimulate MET catalytic activity and facilitate the docking of SH-2 domain- and

MET-binding domain (MBD)-containing proteins including Src, GRB2, PI3K, and PLC-γ. Proteosomal degradation of MET is mediated through interactions between the multisubstrate docking site of MET and c-Cbl. A score of MET-activating mutations, several of which are located in the PTK domain of the β-subunit, have been identified in various solid tumors and metastatic carcinomas. Of particular interest are various germline as well as somatic missense mutations within exons 16 to 19 of the *Met* gene. These mutations commonly generate amino acid substitutions within the kinase domain. For example, the N1118Y substitution impacts the highly conserved ATP-binding site while Y1253D affects the regulatory site on MET in squamous cell cancer. MET overexpression has been observed in various tumor specimens, including NSCLC, mesothelioma, pancreatic cancer, osteosarcoma, ovarian carcinoma, as well as metastatic lesions. Studies of small cell lung carcinoma have revealed that interactions between PI3K and MET sustain survival and motility. Direct involvement of MET in the process of oncogenesis is implicated by the presence of germline as well as somatic missense mutations, which characterize hereditary papillary renal carcinoma.

NTRK1 and NTRK3 Neurotrophins are a family of four secreted and growth-promoting polypeptides known to be widely distributed in both neural and non-neural tissues – that is, nerve growth factor, brain-derived neurotrophic factor, and neurotrophins 3 and 4 (NT3 and NT4). Three homologous neurotrophin tyrosine receptor kinases (NTRK1–3) are responsive to one or more of the neurotrophins and both NTRK1 and NTRK3 are known to form oncogenic fusion proteins in soft tissue tumors. The human *Nrtk1* and *Nrtk3* genes are located on chromosomes 1q21-q22 and 15q25, respectively. While nerve growth factor is the preferred ligand of NTRK1, NTRK3 binds neurotrophin 3 with high affinity. NTRK receptors share a common structural organization of their amino-terminal extracellular domains, which includes three leucine-rich 24-residue motifs flanked by two cysteine clusters and a pair of juxtamembrane C2-type immunoglobulin-like domains. The major ligand-binding domains are localized to the membrane-proximal immunoglobulin-like domains, which are also required to inhibit spontaneous dimerization and activation of the receptors in the absence of ligand. The intracellular carboxy-terminus of these receptors contains a tyrosine kinase domain plus several Tyr-containing motifs. These motifs are involved in regulating kinase activity and also in the phosphorylation-dependent recruitment of several small G proteins (e.g. Ras, Rap-1, and members of the Cdc-42/Rac/Rho family) and intermediates in the MAP kinase, PI3K, and PLCγ signaling cascades, ultimately resulting in activation of gene expression in a cell-dependent manner. Chromosomal translocations that form oncogenic fusion proteins containing the kinase domains of either NTRK1 or NTRK3, have been identified in various carcinomas. In the case of NTRK1, oncogenic activity may be conferred by a chromosomal translocation involving the 5′-dimerization domain of a non-muscle tropomyosin gene and the 3′-kinase domain of *Ntrk1*, giving rise to a constitutively active PTK. In other cases, rearrangements have been identified involving either the 5′-*TPR* (translocated promoter region) from the *Met* oncogene, or the 5′-region of *TFG* (*trk* fused gene). The *TPR-Ntrk1* translocation

forms a fusion protein composed of the amino-terminal portion of TPR and the carboxy-terminal tyrosine kinase domain of NTRK1. Transgenic mice that express human TPR-NTRK1 in the thyroid develop follicular hyperplasia and papillary carcinoma. Chromosomal translocations involving *Ntrk3* often involve the breakpoint regions t(12;15)(p13;q25) and result in a chimera consisting of the NTRK3 kinase domain and the dimerization domain of the ETV6/Tel transcription factor. The oncogenic fusion proteins produced by any of the rearrangements discussed above exhibit three common characteristics that are absent from NTRK1 or NTRK3; ubiquitous cytosolic expression, ligand-independent dimerization *via* coiled-coil domain interactions, and constitutive kinase activity. These fusions have been implicated in the genesis of human colon carcinoma and papillary thyroid carcinoma in the case of NTRK1, and in secretory breast carcinoma and congenital fibrosarcoma in the case of NTRK3.

PDGFRA and PDGFRB PDGFRA and PDGFRB are encoded by genes located in human chromosomes 4q11-q13 and 5q31-q32, respectively. PDGFRA exhibits a pronounced structural and sequence similarity with members of the PDGFR family, including PDGFRB. The main distinction between PDGFRA and PDGFRB is exemplified by their ligand preferences, with PDGFRB selectively interacting with PDGF homodimers containing B type polypeptide chains (PDGF BB) while PDGFRA interacts more promiscuously with the PDGF AA, BB, and AB dimers. In addition, these receptors exhibit temporally distinct and tissue-specific activities in the mammalian embryo and adult. For example, PDGFRA expression in the nervous system coincides with the development of glial cells, which includes processes of glial cell migration and proliferation. The functional PDGFRA consists of an extracellular domain containing five immunoglobulin-like domains punctuated by eight N-linked glycosylation sites, a transmembrane domain, and two cytoplasmic kinase domains (domains 1 and 2). The extracellular immunoglobulin-like domains are essential for the proper folding of the PDGFRA protein, thereby enabling receptor–ligand interactions. Ligand engagement induces receptor dimerization and autophosphorylation, which generates phosphorylation-dependent recruitment sites for adapter proteins and enzymes capable of activating the PI3K and MAP kinase pathways (e.g. Src, PI3K, PLC-γ, the Crk group of adapter proteins, and SHP-2). PDGFRA is also capable of stimulating c-Jun N-terminal kinase (JNK1) to promote apoptosis, and inhibit cell-cycle progression *via* the cyclin-dependent kinase (CDK) inhibitor $p21^{Cip1}$. Signaling through the PDGFRA also stimulates membrane ruffling, cytoskeletal rearrangements, and migration in a cell-type-specific manner. PDGFRA activity is limited by ligand-induced Src activation and receptor trafficking/internalization. In cancer, a potentially activating deletion of an 81 amino acid stretch within the fifth immunoglobulin-like domain of the extracellular stalk of PDGFRA has been described. This deletion mutant is found amplified in glioblastoma, where it is likely to give rise to an alternatively folded receptor capable of sustaining PDGFRA catalytic activity in the absence of ligand. Recently, a transforming *PDGFRA* deletion mutant lacking immunoglobulin-like domain-encoding exons 8 and 9 has also been identified in gliomas. This mutant protein is capable of autophosphorylation and, therefore, activation in the absence

of PDGF. In addition, mutations and deletions within the PDKFRA activation loop (exon 18) and juxtamembrane region (exon 12), that have been detected in gastrointestinal pacemaker cell tumors (i.e. "gastrointestinal stromal tumors"), are also capable of producing ligand-independent activation. PDGFRA and its ligands have also been reported to facilitate tumorigenesis in the absence of genetic mutations, by establishing autocrine signaling loops, as in the case of medulloblastomas where autocrine stimulation promotes metastasis.

RET The glial cell line–derived neurotrophic factor (GDNF) family, consisting of GDNF, neurturin, artemin, and persephin, all transmit signals promoting neuronal survival, proliferation, and differentiation *via* the transmembrane receptor tyrosine kinase RET (rearranged in transformation) and its co-receptors GFRα-1 to -4. RET is normally expressed throughout the nervous system, within neural crest derived cells, and the urogenital system, and is required for neuronal and kidney development. The GFRα1 to 4 co-receptors are glycosyl-phosphatidylinositol-anchored proteins that selectively recognize and bind certain members of the GDNF family and associate with RET kinase to trigger intracellular signaling. For instance, GDNF preferentially binds to the GFRα1-RET heterodimer. The *RET* gene has been mapped to human chromosome 10q11.2 and is known to give rise to a transcript which is alternatively spliced to yield the short, middle, and long isoforms of RET, each containing an extracellular domain with a cadherin-related motif and a cysteine-rich region. When bound to GDNF, GFRα1 dimerizes with RET, and stimulates transphosphorylation of specific Tyr residues within the cytoplasmic domain of RET. The three RET isoforms have a number of Tyr residues in common, including those residing within the kinase domain and carboxy-terminus, though the long RET isoform contains an additional carboxy-terminal Tyr1096 residue. Once phosphorylated, these residues serve as scaffolding sites for signaling complexes. The conserved Tyr residues bind GRB10, PLCγ, and Shc, while Tyr1096 binds GRB2 specifically and links RET activation to the cell proliferation and survival pathways controlled by the Ras/Raf/MEK/ERK and PI3K/AKT pathways respectively. Underscoring the importance of Tyr1062, are studies that demonstrate that a point mutation within this codon significantly decreases the transforming activity of the RET protein. RET-mediated transformation results from various activating germline mutations in early-onset multiple endocrine neoplasia types 2A and 2B and late-onset familial medullary thyroid carcinoma. In addition, genetic rearrangements have been identified in papillary thyroid carcinoma. Chromosomal translocations involving the intracellular domain of *RET* have given rise to 10 known fusion proteins (RET/PTC1 -9 and RET/ELKS) prevalent in papillary thyroid carcinoma in a geographically distinct manner. These include translocations between *RET* and *H4*, *RIα*, *ELE1*, *RFG5*, *hTIF*, *RFG7*, *kinectin*, *RFG9*, and *ELKS*. The most common rearrangements, RET/PTC1 and RET/PTC3, involve inversions of the *H4* inv(10)(q11.2q21) and *ELE1* genes and account for 90% of all rearrangements in papillary thyroid carcinoma. Studies suggest that RET/PTC becomes ectopically localized to the cytoplasm where it activates STAT3 by phosphorylation of Tyr705. The constitutively active STAT3 promotes cellular transformation *via* stimulation of

cyclin-D1, *D2*, *D3*, and *bcl-XL* transcription. RET/PTC has also been implicated in the transcriptional upregulation of the osteopontin gene, resulting in an increase in the proliferative and invasive activity of cells that respond to autocrine stimulation by osteopontin. While RET/PTC rearrangements predominate in papillary thyroid carcinoma, germline mutations within *RET* are responsible for the onset of multiple endocrine neoplasia types 2A and 2B, and familial medullary thyroid carcinoma. Point mutations or base pair duplications of the *RET* gene have been reported in patients with multiple endocrine neoplasia 2A and familial medullary thyroid carcinoma. Missense mutations within *RET* have also been described in patients with multiple endocrine neoplasia 2A or 2B. Point mutations observed in multiple endocrine neoplasia 2A and familial medullary thyroid carcinoma, target the extracellular cysteine residues that normally contribute to intramolecular disulfide bonds, and promote RET dimerization. Formation of RET homodimers leads to its ligand-independent activation. On the other hand, RET mutations in multiple endocrine neoplasia 2B do not induce dimerization; rather, mutations target the catalytic core and may alter substrate specificity by inducing conformational changes in the region containing Tyr905, a conserved residue important for catalysis.

3.1.2 Cytoplasmic Protein-tyrosine Kinase Subfamily

ABL1/ABL2 The *ABL1* gene (chromosomal band 9q34.1), along with the closely related *ABL2* or *ARG* (Abl-related gene; chromosomal band 1q24-q25), belongs to the Abelson family of genes that encode prototypic non-receptor tyrosine kinases involved in various cellular processes. The *ABL1* gene product, c-Abl is a ubiquitously expressed protein that colocalizes with actin-based cytoskeletal structures in the cytoplasm, and binds chromatin when imported to the nucleus. These intermolecular interactions are mediated by the carboxy-terminal actin- and DNA-binding domains of c-Abl. The amino-terminus of c-Abl contains an SH-3 domain and an SH-2 domain. Autoinhibition of c-Abl is achieved by a novel variation on a familiar theme, and the details of this mechanism only recently became apparent when the crystal structure of the protein became available for examination. As in the case of the closely related Src kinases, the amino-terminus of the protein functions as an autoinhibitory "Cap" region preventing substrate docking with the kinase domain. However, this is not achieved by the binding of its SH-2 domain to a phosphotyrosine residue, as in other Src-like kinases. Rather, an amino-terminal myristate is responsible for binding to a hydrophobic pocket in the kinase domain that stabilizes an autoinhibited conformation. It is this distinction in the autoinhibitory mechanisms employed by c-Abl as compared with those of the other Src kinases that explains the ability of the drug imatinib (Gleevec®) to selectively inhibit the catalytic activity of c-Abl, but not that of c-Src. Targeted disruption of c-Abl in mice is characterized by lymphopenia, failure to thrive, and neonatal mortality. In the cytoplasm, c-Abl participates in the propagation of mitogenic signals via the Ras/Raf/MEK/ERK and PI3K/Akt pathways. This signaling is negatively regulated by the interactions of c-Abl with PSTPIP1, a PEST-type protein Tyr phosphatase that dephosphorylates c-Abl. c-Abl signaling in the nucleus results in a suppression of cell growth *via* the c-Abl-dependent transactivation of various proapoptotic genes. During the early G1

phase of the cell cycle, this nuclear activity of c-Abl is repressed due to its sequestration by the pocket protein Rb. Following an oncogenic and reciprocal chromosomal translocation, control of c-Abl activity is lost in cells, and this generates what has been termed the *Philadelphia chromosome*. This mutation is present in most cases of CML and many cases of ALL. Although similar translocations occur in CML and ALL, their c-Abl products are distinct, producing 210 kDa and 180 to 185 kDa proteins respectively. Both involve a t(9;22) (q34, q11) translocation of the *ABL* gene to the center of the *BCR* gene, producing a fused transcript that translates into a chimeric Bcr-Abl protein that is constitutively active. C-Abl activity is sustained by the loss of its autoinhibitory (myristoylated) amino-terminal tail and the dimerization triggered by the presence of coiled-coil domains within the amino-terminus of Bcr. Bcr-Abl dimerization, in turn, allows for the transphosphorylation of the fusion protein and activation of oncogenic signaling. The formation of Bcr-Abl homo-oligomers and hetero-oligomers with Bcr further promotes oncogenesis by localizing the protein in the cytosol and preventing DNA binding. The use of the small molecular tyrosine kinase inhibitor Gleevec® in the treatment of CML patients who are Bcr-Abl-positive has met with success, but drug resistance occurs frequently. Mutations of residues in or near the ATP-binding pocket of c-Abl prevent Imatinib binding and account for a majority of drug-resistant cases. The homologous *ARG* gene is involved in a similar translocation, but with a different fusion partner, Ets variant gene 6 (*ETV6/Tel*). ETV6/Tel is a member of the Ets family of transcription factors and is frequently found to be rearranged with a variety of translocation partners in human leukemias. This fusion protein contains the complete SH-3, SH-2, and PTK domains of ARG along with the helix–loop–helix oligomerization domain of ETV6/Tel. The translocation t(1;12)(q25;p13) causes malignancy in adult patients with acute myelogenous leukemia.

JAK-2 Janus kinase 2 (JAK-2) is a receptor-associated PTK essential for propagating certain cytokine induced signals. Located on human chromosome 9p24, the *JAK-2* gene encodes five JH domains, which are regions common to all members of the JAK kinase family and are sites of interaction between the protein kinase and its associated cytokine receptor. An amino-terminal JH-1 domain contains the kinase domain responsible for the phosphorylation of its preferred substrate; the Tyr, neutral, basic amino acid motif. Carboxy-terminal to the catalytic domain is a regulatory pseudokinase domain, referred to as the JH-2 domain. Normally, JAK-2 activity is cytokine-dependent. Interleukin-3 (IL-3), IL-5, and/or IL-6 interactions with cytokine receptors initiate receptor dimerization and the activating transphosphorylation of the constitutively associated JAK-2 proteins. Once active, JAK-2 phosphorylates Tyr residues on the carboxy-terminal region of cytokine receptors, providing binding sites for recruitment of STATs *via* their SH-2 domains. This in turn permits JAK-2 to phosphorylate STAT proteins, thereby promoting the disassociation of STATs from cytokine receptors, and their dimerization. When exported to the nucleus, dimeric STATs associate with the promoters of early response genes such as *c-fos*, *c-jun*, and *c-myc* and initiate transcription of cell-cycle regulatory genes such as *cyclin D1* and pro-survival genes such as

bcl-XL. Consequently, the constitutive activation of JAK-2 significantly enhances cell proliferation and survival. Deregulation of JAK-2 activity arises as a result of a somatic chromosomal translocation between the dimerization domain of the ETS-like transcription factor Tel and JAK-2. The resulting fusion protein contains the amino-terminal protein dimerization domain of Tel and carboxy-terminal JAK-2 catalytic domain (JH-1), and is able to form homodimers capable of transphosphorylating its binding partner, thus leading to constitutive activation. The Tel-JAK-2 fusion protein is ectopically localized within cells and may undergo unusual interactions with members of the early gene induction pathway in a cytokine-independent manner, thereby supporting cellular transformation. The presence of the *Tel-JAK-2* translocation and resulting transformation of hematopoietic cells underscore the importance of JAK-2 signaling in hematopoiesis. Examples of the Tel-JAK-2 driven transformation include cases of T-cell childhood acute lymphocytic leukemia, acute myeloid leukemia, acute lymphocytic leukemia, and atypical chronic myeloid leukemia. The finding that JAK-2 inhibition by a specific kinase blocker (i.e. AG-490) induces apoptosis of leukemic cells further supports the direct linkage between constitutive JAK-2 activation and malignant transformation. It is of note that while neither JAK-1 nor JAK-3 meets the criteria for inclusion in the cancer genome, both these cytoplasmic PTKs also signal to the nucleus *via* STATs and are overexpressed in various leukemias.

LCK The T-lymphocyte protein-tyrosine kinase (*LCK*) is a member of the Src-tyrosine kinase family and is located on the short arm of human chromosome 1 (1p34). LCK expression predominates in T cells, where its activity is crucial for normal T-cell development and activation. LCK is closely associated with the T-cell receptors (TCR) CD4 and CD8. LCK is a multidomain protein, typical of the Src kinases, including SH-3, SH-2, tyrosine kinase domains, and a unique domain responsible, among other things, for the membrane anchoring of LCK. During the process of acquiring adaptive immunity, T-cell activation is initiated by the binding of CD4 and CD8 to antigen-presenting cells *via* type I and type II glycoproteins of the major histocompatibility complex. The proximity of LCK to these activated TCRs allows for the transduction of signals responsible for the T-cell response, which includes the production of IL-2 and inositide (1,4,5) triphosphate. LCK activation requires autophosphorylation of Tyr394 and dephosphorylation of Tyr505, in the carboxy-terminal tail. The protein Tyr phosphatase CD45 regulates LCK activity in a bimodal fashion, not only by mediating Tyr505 dephosphorylation (enhanced activity) but also by dephosphorylation of Tyr394 (suppressed activity). LCK promotes T-cell activation by phosphorylating immunoreceptor Tyr-based activation motifs present within TCR complexes. When phosphorylated, these motifs enable recruitment of SH-2 domain–containing proteins (e.g. ZAP70) to the TCR. In addition, LCK phosphorylates the membrane–cytoskeleton linker protein ezrin to promote T-cell activation. In T cells, the LKC-dependent phosphorylation of the linker protein Shc also initiates mitogenic signaling *via* the Ras/Raf/MEK/ERK pathway. Because of the critical position occupied by LCK in the TCR signaling pathway, its necessarily transient activity is highly regulated through ubiquitination under the control of the Cbl ubiquitin ligase. LCK mutations resulting in constitutive

autophosphorylation and/or deregulated catalytic function have been described in T-cell leukemia, where they contribute to the oncogenic potential of this kinase. Insertion of the tripeptide QKP between the cytosolic SH-2 and kinase domains may also enhance catalytic function by disrupting the normal repressive interaction of its SH-2 domain with its catalytic site. Three other recorded mutations, resulting in a substitution within the region critical for interaction between LCK and CD4 and two alternative substitutions within the kinase domain, may result in a GOF for LCK. In addition to these mutations, a chromosomal translocation (t(1;7)(p34;q34)) has been observed in T-cell acute lymphoblastic leukemia, resulting in an oncogenic fusion protein (i.e. βTCR:LCK) that contains the βTCR enhancer region and the 5′ promoter region of LCK. In the case of the βTCR:LCK fusion, a GOF would result from an alleviation of transcriptional repression by the βTCR enhancer region.

3.2 Protein-serine/Threonine Kinase Family

3.2.1 Tyrosine Kinase-like Subfamily

BRAF BRAF is discussed above in Sect. 2.2.

3.2.2 Homologs of Yeast Sterile 7, Sterile 11, Sterile 20 Kinase Subfamily

JNKK The c-Jun terminal kinase–kinase (*JNKK*) gene is located on human chromosome 17p11. This oncogene encodes a dual specificity kinase catalyzing Thr180 and Tyr182 phosphorylation of p38 MAP kinase and JNK1, 2, and 3 in response to environmental stress, including changes in cellular osmolarity. The functional consequences of JNKK signaling are cell-type-, environment-, and stimulus-specific. Regulation of JNKK occurs upstream of the kinase as a result of the ubiquitin-mediated degradation of MAPK kinase–kinase (MEKK1). Ubiquitination prevents phosphorylation of JNKK by MEKK1 and abrogates downstream activation of JNK. JNK activation requires stable complex formation between JNKK and JNK. The amino-terminal region of JNKK facilitates this interaction as it codes for a conserved MAP kinase-docking site (D-domain). JNKK is able to modulate JNK signaling to AP-1 transcription factors by competitive sequestration of JNK *via* D-domain interactions. In a classic response to UV irradiation, JNKK mediated activation of JNK results in apoptosis through a p53-dependent mechanism. Downregulation of JNKK or its downstream signaling components provides cells with a growth advantage. Moreover, a loss of JNKK expression has been observed in various cancers, including prostate cancer. An overall *JNKK* mutation rate of ~5% has been observed in various types of tumors. The mutations include two nonsense mutations, five missense mutations, and one splice site mutation. Not surprisingly, these mutations generally serve to inactivate JNKK signaling. In keeping with its tumor suppressor classification, the *JNKK* gene is a known target for homozygous deletions in pancreatic adenocarcinomas, biliary adenocarcinomas, and breast carcinomas.

3.2.3 Calcium/Calmodulin-dependent Protein Kinase Subfamily

PIM-1 A common site for proviral integration of the moloney murine leukemia virus (PIM) in mice encodes the *PIM-1* gene. The Ser/Thr kinase PIM-1 is the product of this gene, which is located

on the p12 segment of human chromosome 6. In addition to PIM-1, PIM-2 and PIM-3 make up the PIM family of kinases and may serve redundant functions as observed in PIM-1-deficient mice, where compensation for the loss of *PIM-1* was evident. PIM-1 functions to promote hematopoiesis in response to growth factors such as IL-2 and IL-3 and its expression is confined, throughout fetal development, to hematopoietic sites within the liver and spleen, as well as within the thymus, lymph-nodes, and testis. High expression of PIM-1 has also been observed in myeloid cell and B-cell lines. PIM-1 has been characterized as a 33-kDa cytoplasmic protein, though recent studies have revealed that its nuclear localization is crucial for its antiapoptotic function in Burkitt's lymphoma. Upregulation of PIM-1 has been implicated in BCR/ABL mediated leukemogenesis where it acts downstream of STAT5 to confer protection from apoptosis and promote cell-cycle progression. PIM-1 overexpression has also been shown to positively influence a variety of human leukemias, though overexpression alone is not sufficient for cellular transformation, and cooperation between PIM-1 and c-myc is often observed during this process. The identification of a breakpoint region on chromosome 6 has raised the possibility of a translocation between *PIM-1* and the proto-oncogene *c-ABL* (t(6;9)(p21; q33)) in some cases of myeloid leukemia. In response to cytokines and various mitogens, *PIM-1* expression is induced *via* the JAK/STAT5 and MAP kinase pathways and its activity may be modulated through a series of autophosphorylation events. With the exception of Etk, which promotes the activation of PIM-1 *via* phosphorylation of Tyr218, upstream kinase regulators of PIM-1 have not been identified. Recent studies show that phosphorylation of the nuclear mitotic apparatus protein by PIM-1 facilitates interactions between this protein, the heterochromatic protein-1β, and the spindle proteins dynein and dynactin, which are crucial for mitosis. Furthermore, PIM-1 alleviates transcriptional repression by directly phosphorylating HPI, a member of the transcriptional repression machinery, thereby regulating chromatin structure. Other putative substrates of PIM-1 include p100, a coactivator of the transcription factor c-myb, the cell-cycle phosphatase Cdc25A, and the cyclin-dependent kinase inhibitor $p21^{cip1}$. It is evident that PIM-1 regulates the activity of many proteins that have the potential to promote tumor growth and, as expected, its expression is tightly regulated *via* transcriptional and translational mechanisms. The presence of dyad symmetry elements enables the *PIM-1* gene to form stem loop structures, and to block transcriptional elongation in the absence of appropriate stimuli. Another regulatory element is present in the form of an A/U rich region within the 3′-unstranslated region of PIM-1 mRNA. This region promotes RNA instability, which can be overcome in the presence of certain cytokines and mitogens. Translation of the PIM-1 message is regulated by the presence of a 5′-untranslated region containing an internal ribosomal entry site. PIM-1 activity may also be regulated *via* autophosphorylation, while its turnover rate is dependent upon ubiquitination. In the event of deregulated PIM-1 expression, oncogenesis is most likely facilitated by the PIM-1-dependent stimulation of cell-cycle progression and mitosis.

STK11 Serine–threonine kinase 11 (STK11), also known as LKB1, is a crucial regulator of cellular metabolism and polarity. This cytoplasmic kinase is ubiquitously expressed in human tissues and exhibits

catalytic activity only in the presence of its binding partner STRAD (STE20-related adapter). As STRAD lacks intrinsic kinase activity, the association of the pseudokinase domain of STRAD with the kinase domain of STK11 is postulated to result in intermolecular interactions capable of stimulating the STK11 catalytic core. In the cytoplasm, the association of STK11 with STRAD gives rise to a series of STK11 autophosphorylation events and results in the phosphorylation of STRAD. Induction of cell polarity by STK11 may result from its association with the Ser/Thr kinase PAR1. This leads to destabilization and polarized localization of microtubules. STK11 also plays a role in axis induction during development by stimulating the Wnt signaling pathway. The tumor suppressor activity of STK11 is mediated by the ataxia-telangiectasia mutated kinase (ATM, see below). In response to the cellular stress caused by ATP depletion, STK11 directly phosphorylates Thr172 of the AMP-activated kinase. This results in the downregulation of ATP-consuming processes, such as the translational activity mediated by the mammalian target of rapamycin, and upregulation of ATP-producing processes such as glucose uptake. Therefore, the loss of STK11 expression, observed in various carcinomas, offers cells a distinct growth advantage by removing the inhibition normally placed on protein synthesis and mitogenesis. The tumor suppression activity of STK11 makes it a target for inactivating mutations such as those observed in patients with the hereditary Peutz-Jeghers syndrome. Indeed, the *STK11* gene was mapped to region p13.3 of human chromosome 19 – a site susceptible to germline mutations in patients with Peutz-Jeghers syndrome. In addition to the formation of benign intestinal hamartomatus polyps, individuals with Peutz-Jeghers syndrome exhibit a predisposition toward gastrointestinal, pancreatic, ovarian, testicular, uterine, and breast carcinomas. Studies have identified various deviations of the wild-type STK11 transcript. These are caused by translational frameshift insertions or deletions, nonsense mutations, missense mutations, and disruption of STK11 mRNA splicing through exonal splice site mutations. The functional consequence of these mutations is the generation of an inactive STK11 protein. Germinal *STK11* mutations are often accompanied by somatic mutations. Somatic mutations leading to homozygous deletion of *STK11* or *STK11* truncations have also been reported in patients with non-Peutz-Jeghers syndrome related pancreatic cancer, mucinous minimal deviation adenocarcinoma of the uterine cervix, lung adenocarcinoma, and malignant melanoma. In general, *STK11* mutations affect the biological activity of STK11 by one of two common means. Large truncations or targeted mutations of invariant residues within the STK11 kinase domain destroy its ability to bind ATP and catalyze phosphotransfer. The SL26 mutant present in some cases of Peutz-Jeghers syndrome does not affect kinase activity; rather it promotes the loss of cytoplasmic retention, which is accompanied by nuclear localization of STK11. It is hypothesized that nuclear localization results from the inability of the SL26 mutant to interact with its cytoplasmic substrates and thereby remain in the cytoplasm. Nuclear localization also leads to silencing of STK11 signaling because of the lack of substrates and activating factors. Paradoxically, owing to a marked increase in the AMP/ATP ratio, cells with loss of STK11 expression are rendered more susceptible to cell death caused by agents such as the

AMP analog AICAR. This response may represent a unique therapeutic opportunity for those patients whose carcinomas exhibit a loss of STK11 activity.

3.2.4 Containing Pka, Pkg, Pkc Subfamilies

AKT2 The *AKT2* gene, located at 9q13.1-q13.2, produces a phosphoinositide-dependent Ser/Thr kinase, crucial for regulating cellular events including apoptosis, differentiation, proliferation, and metabolism. The gene product of AKT2, also referred to as protein kinase B beta (PKB-β), is a broadly expressed protein containing a single 3-phosphoinositide-binding pleckstrin homology (PH) domain at its amino terminal. Adjacent to this pleckstrin homology domain is the kinase domain, followed by a carboxy-terminal regulatory domain. The AKT isoforms (AKT1 and AKT2) function downstream of various RTKs, and the biogenesis of an active AKT requires the sequential recruitment of the inactive protein to the plasma membrane followed by the transphosphorylation of a conserved Thr residue within the activation loop (Thr308 in AKT1 and Thr306 in AKT2) and Ser residue within a carboxy-terminal hydrophobic motif (Ser473 in AKT1 and Ser474 in AKT2). AKT is recruited to the plasma membrane through the binding of its amino-terminal PH domain to phosphatidylinositol-3, 4, 5-triphosphate (PIP_3), a lipid product of PI3K. This protein–-lipid interaction is tightly controlled by the lipid phosphatase *PTEN* (phosphatase and tensin homolog), which is also a cancer (tumor) suppressor gene that undergoes both somatic and germline inactivating mutations. By anchoring AKT to the membrane lipid bilayer, PIP_3 alters the conformation of the AKT kinase domain and exposes the activation loop phosphorylation site to PDK1. Following this "priming" step, AKT is phosphorylated on its hydrophobic motif by a yet-to-be-identified protein Ser/Thr kinase. Integrin-linked kinase was shown to be essential for this final step in the maturation of AKT, but whether this kinase is directly responsible for AKT phosphorylation remains unresolved. The authentic AKT hydrophobic motif kinase may be the DNA-dependent protein kinase. Good evidence for this has recently been presented, but has not yet been confirmed. Once activated, AKT is capable of enhancing cellular proliferation and the generation of survival signals through some of the pathways and effector molecules included in Figure. 1. *AKT2* was included in the census of cancer genes because of its amplification in ovarian and pancreatic cancers. The region that is amplified in ovarian carcinomas (19q13.1-q13.2) spans the entire *AKT2* gene. However, no activating mutations have been documented for this kinase in cancer and, therefore, tumors in which *AKT2* has been amplified may not be responsive to small-molecule inhibitors of kinase activity. Depending on the genetic profile presented by a given patient, other oncogenes might contribute to the transforming activity of AKT2. For example, the coexistence of an *AKT2* amplification and loss-of-function mutation in *PTEN* would allow an enhancement of AKT2 signaling in the presence of endogenous growth factors. The human prostate epithelial cancer cell line LNCaP harbors such a *PTEN* mutation and possesses constitutive AKT kinase activity. Another interesting candidate in ovarian cancer is *SEI-1*. *SEI-1* is located in a genetic region of chromosome 19q13.1-q13.2, neighboring *AKT2*. This genetic region is coamplified along with *AKT2* in a subset of ovarian cancer

cell lines. The *SEI-1* gene encodes a CDK4-binding protein, which renders the activity of cyclin D/CDK4 complexes resistant to the inhibitory effect of $p16^{INK4a}$. Given the proliferation-related functions of AKT2 and SEI-1, it is possible that both gene amplifications at chromosome 19 have the potential to advance the growth of certain ovarian tumors.

3.2.5 Containing Cdk, Mapk, Gsk3, Clk Subfamilies

CDK4 and the INK4 inhibitors The human cyclin-dependent kinase 4 (*CDK4*) gene is located at chromosome 12q14 and encodes a small (303 amino acid) cell-cycle regulatory protein Ser/Thr kinase essential for coordinating the cell's progression through the early growth (G1) phase of the cell cycle. CDK4 is activated by the binding of D-type cyclins (which impart basal activity to the kinase) and, by phosphorylation (which fully activates the kinase). When active, CDK4 interferes with the tumor suppressor activity of the retinoblastoma protein Rb and its homologs. CDK4 is inactivated by the binding of the INK4 family of CDK inhibitors. G1 progression depends on extracellular mitogenic signals that upregulate cyclin D expression, committing the cell to another round of cell division. INK4 proteins bind monomeric CDK4, preventing its stable interaction with cyclin D. They also bind the cyclin D – CDK4 complex, forming an inactive ternary complex. There are four INK4 proteins ($P16^{INK4a}$, $P15^{INK4b}$, $P18^{INK4c}$, and $P19^{INK4d}$), all of which possess ankyrin repeats involved in binding CDK4. The binding of these proteins substantially alters the conformation of the CDK4 ATP-binding site and renders the kinase inactive. Mutations leading to a loss of control over the CDK4/INK4/Rb pathway are among the most common in human cancer. CDK4 is often amplified and overexpressed in glioblastoma (50%), uterine cervical carcinoma (26%), breast carcinoma (16%), and osteosarcoma (16%). In addition, a germline arginine-to-cysteine substitution at codon 24 has been reported in some cases of melanoma. This mutation contributes to malignant transformation by preventing the interaction between $p16^{INK4a}$ and CDK4, but is rarely encountered in other human neoplasms. More common are mutations in the binding region of the *INK4* genes, lowering the affinity of the INK4 proteins for CDK4. In the case of $p16^{INK4a}$, such mutations occur frequently in leukemia (58%), bladder carcinoma (50%), glioma (35%), nasopharyngeal carcinoma (35%), and pancreatic cancer (21%). Promoter methylation is also a common mechanism of inactivation of the *INK4* genes in cancer.

3.2.6 Phosphatidylinositol Kinase (Pik)-Related Subfamily

ATM The human tumor suppressor gene ataxia-telangiectasia mutated (*ATM*) is located on chromosome 11q22-q23 and is frequently mutated in a rare autosomal recessive disorder causing neuronal degeneration, immunologic deficiency, radiosensitivity, and cancer predisposition. Over one-third of ataxia-telangiectasia patients develop lymphoid cancers including non-Hodgkin's lymphoma, Hodgkin's lymphoma, and other leukemias. The *ATM* gene encodes a large (3056 amino acid) nuclear protein that is expressed in testis, spleen, and thymus of adult mice and plays a critical role in maintaining genetic stability in response to DNA damage. ATM is a member of the recently identified PIK-related subfamily that phosphorylates Ser/Thr residues, rather

than lipids. All members contain a conserved carboxy-terminal catalytic domain, referred to as the PI-kinase domain, and a novel amino-terminal FAT (FRAP, ATM, and TRRAP) domain that possibly regulates the conformation of the PI-kinase domain. Structurally, ATM is composed of two regions referred to as the *head* and an *arm*. Three-dimensional reconstructions of ATM bound to DNA suggest that the kinase uses its arm to clamp around the double helix. Whether this protein–DNA interaction is required for the ATM-dependent repair of double-strand breaks is not known. ATM activation in response to DNA double-strand breaks is rapidly triggered following ionizing radiation functions to arrest the cell-cycle progression of damaged cells in G1, S, or G2. The G1/S checkpoint is activated through the induction of p53. ATM positively regulates the p53 pathway by increasing the half-life of the p53 protein, either by directly phosphorylating Ser15, or indirectly through activation of other kinases such as Chk2 and c-Abl. ATM, Chk2, and c-Abl are all capable of phosphorylating p53, thus stabilizing the later protein by interfering with its binding to the p53 regulatory protein, murine double minute 2. In addition to regulating the G1/S checkpoint, ATM engages the S and G2/M checkpoints in response to ionizing radiation. It triggers the G2/M checkpoint by Chk2-mediated phosphorylation of Ser216 on Cdc25C, thereby promoting the binding of 14-3-3 proteins. The Cdc25C/14-3-3 complex is then exported from the nucleus, preventing Cdc25C from activating the nuclear Cdc2/cyclinB complex required for mitosis. In a similar fashion, ATM controls the S-phase checkpoint through the Chk2-dependent phosphorylation of Cdc25A Ser123. This leads to Cdc25A degradation. Most of the ATM mutations associated with cancer are derived from large deletions, causing protein truncations that inactivate the kinase. Other inactivating mutations are represented by missense mutations (e.g. L1420F) or small in-frame deletions/insertions. This loss of ATM function is thought to predispose to cancer by increasing genetic instability.

Acknowledgments

National Institutes of Health Grant (R01 CA98195 to JAM) and the DOD Prostate Cancer Research Program (DAMD 17-02-1-0053 to DMT) supported this work.

See also Cancer Stem Cells; Cell Signaling During Primitive Hematopoiesis; Cellular Interactions; Epigenetic Mechanisms in Tumorigenesis; Homeodomain Proteins; Mutagenesis, Malignancy and Genome Instability; Oncology, Molecular.

Bibliography

Books and Reviews

Blume-Jensen, P., Hunter, T. (2001) Oncogenic kinase signaling, *Nature* **411**, 355–365.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., Stratton, M.R. (2004) A census of human cancer genes, *Nature Rev.* **4**, 177–183.

Hanks, S.K., Hunter, T. (1995) Protein kinases 6. the eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification, *FASEB J.* **9**, 576–596.

Hubbard, S.R., Till, J.H. (2000) Protein tyrosine kinase structure and function, *Annu. Rev. Biochem.* **69**, 373–398.

Steelman, L.S., Pohnert, S.C., Shelton, J.G., Franklin, R.A., Bertrand, F.E., McCubrey, J.A. (2004) JAK/STAT, Raf/MEK/ERK, PI3K/Akt and BCR-ABL in cell cycle progression and leukemogenesis, *Leukemia* **18**, 189–218.

Terrian, D.M., (Ed.), (2003) *Cancer Cell Signaling*, Humana Press, Totowa.

Primary Literature

Agnes, F., Shamoon, B., Dina, C., Rosnet, O., Birnbaum, D., Galibert, F. (1994) Genomic structure of the downstream part of the human FLT3 gene: exon/intron structure conservation among genes encoding receptor tyrosine kinases (RTK) of subclass III, *Gene* **145**, 283–288.

Alessi, D.R., Andjelkovic, M., Caudwell, B., Cron, P., Morrice, N., Cohen, P., Hemmings, B.A. (1996) Mechanism of activation of protein kinase B by insulin and IGF-1, *EMBO J.* **15**, 6541–6551.

Bai, R.Y., Ouyang, T., Miething, C., Morris, S.W., Peschel, C., Duyster, J. (2000) Nucleophosmin-anaplastic lymphoma kinase associated with anaplastic large-cell lymphoma activates the phosphatidylinositol 3-kinase/Akt antiapoptotic signaling pathway, *Blood* **96**, 4319–4327.

Bischof, D., Pulford, K., Mason, D.Y., Morris, S.W. (1997) Role of the nucleophosmin (NPM) portion of the non-Hodgkin's lymphoma-associated NPM-anaplastic lymphoma kinase fusion protein in oncogenesis, *Mol. Cell. Biol.* **17**, 2312–2325.

Blume-Jensen, P., Janknecht, R., Hunter, T. (1998) The kit receptor promotes cell survival via activation of PI 3-kinase and subsequent Akt-mediated phosphorylation of Bad on Ser136, *Curr. Biol.* **8**, 779–782.

Bosch, E., Cherwinski, H., Peterson, D., McMahon, M. (1997) Mutations of critical amino acids affect the biological and biochemical properties of oncogenic A-Raf and Raf-1, *Oncogene* **15**, 1021–1033.

Bossotti, R., Isacchi, A., Sonnhammer, E.L.L. (2000) FAT: a novel domain in PIK-related kinases, *TIBS* **25**, 225–227.

Brown, V.L., Harwood, C.A., Crook, T., Cronin, J.G., Kelsell, D.P., Proby, C.M. (2004) p16INK4a and p14ARF tumor suppressor genes are commonly inactivated in cutaneous squamous cell carcinoma, *J. Invest. Dermatol.* **122**, 1284–1292.

Carow, C.E., Levenstein, M., Kaufmann, S.H., Chen, J., Amin, S., Rockwell, P., Witte, L., Borowitz, M.J., Civin, C.I., Small, D. (1996) Expression of the hematopoietic growth factor receptor FLT3 (STK-1/Flk2) in human leukemias, *Blood* **87**, 1089–1096.

Cazzaniga, G., Tosi, S., Aloisi, A., Giudici, G., Daniotti, M., Pioltelli, P., Kearne, L., Biondi, A. (1999) The tyrosine kinase abl-related gene ARG is fused to ETV6 in an AML-M4Eo patient with a t(1;12)(q25;p13): molecular cloning of both reciprocal transcripts, *Blood* **94**, 4370–4373.

Chang, F., Lee, J.T., Navolanic, P.M., Steelman, J.G., Blalock, W.L., Franklin, R.A., McCubrey, J.A. (2003) Involvement of PI3K/Akt pathway in cell cycle progression, apoptosis, and neoplastic transformation: a target for cancer chemotherapy, *Leukemia* **17**, 590–603.

Chang, F., Steelman, L.S., Lee, J.T., Shelton, J.G., Navolanic, P.M., Blalock, W.L., Franklin, R.A., McCubrey, J.A. (2003) Signal transduction mediated by the Ras/Raf/MEK/ERK pathway from cytokine receptors to transcription factors: potential targeting for therapeutic intervention, *Leukemia* **17**, 1263–1293.

Clark, S.S., McLaughlin, J., Crist, W.M., Champlin, R., Witte, O.N. (1987) Unique forms of the abl tyrosine kinase distinguish Ph1-positive CML from Ph1-positive ALL, *Science* **235**, 85–88.

Cong, F., Spencer, S., Cote, J.F., Wu, Y., Tremblay, M.L., Lasky, L.A., Gogg, S.P. (2000) Cytoskeletal protein PSTPIP1 directs the PEST-type protein tyrosine phosphatase to the c-Abl kinase to mediate Abl dephosphorylation, *Mol. Cell.* **6**, 1413–1423.

Crews, C.M., Alessandrini, A., Erikson, R.L. (1992) The primary structure of MEK, a protein kinase that phosphorylates the ERK gene product, *Science* **258**, 478–480.

Cunningham, D., Humblet, Y., Siena, S., Khayat, D., Bleiberg, H., Santoro, A., Bets, D., Mueser, M., Harstrick, A., Verslype, C., Chau, I., Van Cutsem, E. (2004) Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer, *N. Engl. J. Med.* **351**, 337–345.

de Klein, A., van Kessel, A.G., Grosveld, G., Bartram, C.R., Hagemeijer, A., Bootsma, D., Spurr, N.K., Heisterkamp, N., Groffen, J., Stephenson, J.R. (1982) A cellular oncogene is translocated to the Philadelphia chromosome

in chronic myelocytic leukemia, *Nature* **300**, 765–767.

Dong, J., Phelps, R.G., Qiao, R., Yao, S., Benard, O., Ronai, Z., Aaronson, S.A. (2003) B Raf oncogenic mutations correlate with progression rather than initiation of human melanoma, *Cancer Res.* **63**, 3883–3886.

Druker, B.J. (2004) Molecularly targeted therapy: have the floodgates opened? *The Oncologist* **9**, 357–360.

Druker, B.J., Talpaz, M., Testa, D.J., Peng, B., Buchdunger, E., Ford, J.M., Lydon, N.B., Kantarjian, H., Capdeville, R., Ohno-Jones, S., Sawyers, C.L. (2001) Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia, *N. Engl. J. Med.* **344**, 1031–1037.

Eley, G., Frederick, L., Wang, X.Y., Smith, D.I., James, C.D. (1998) 3′ end structure and rearrangements of EGFR in glioblastomas, *Genes Chromosomes Cancer* **23**, 248–254.

Falck, J., Mailand, N., Syljuasen, R.G., Bartek, J., Lukas, J. (2001) The ATM-Chk2-Cdc25A checkpoint pathway guards against radioresistant DNA synthesis, *Nature* **410**, 766–767.

Feng, J., Park, J., Cron, P., Hess, D., Hemmings, B.A. (2004) Identification of a PKB/Akt hydrophobic motif Ser-473 kinase as DNA-dependent protein kinase, *J. Biol. Chem.*: **279**, 41189–41196.

Gatti, R.A., Berkel, I., Boder, E., Braedt, G., Charmley, P., Concannon, P., Ersoy, F., Foroud, T., Jaspers, N.G.J., Lange, K. et al. (1988) Localization of an ataxia-telangiectasia gene to chromosome 11q22–23, *Nature* **336**, 577–580.

Gilliland, D.G., Griffin, J.D. (2002) The roles of FLT3 in hematopoiesis and leukemia, *Blood* **100**, 1532–1542.

Goldman, J.M., Melo, J.V. (2003) Chronic myeloid leukemia – advances in biology and new approaches to treatment, *N. Engl. J. Med.* **349**, 1451–1464.

Golub, T.R., Goga, A., Barker, G.F., Afar, D.E., McLaughlin, J., Bohlander, S.K., Rowley, J.D., Witte, O.N., Gilliland, D.G. (1996) Oligomerization of the ABL tyrosine kinase by the Ets protein TEL in human leukemia, *Mol. Cell Biol.* **16**, 4107–4116.

Haeder, M., Rotsch, M., Bepler, G., Hennig, C., Havemann, K., Heimann, B., Moelling, K. (1988) Epidermal growth factor receptor expression in human lung cancer cell lines, *Cancer Res.* **48**, 1132–1136.

Hall, M., Peters, G. (1996) Genetic alterations of cyclins, cyclin-dependent kinases, and Cdk inhibitors in human cancer, *Adv. Cancer Res.* **68**, 67–108.

Hantschel, O., Superti-Furga, G. (2004) Regulation of the c-Abl and Bcr-Abl tyrosine kinases, *Nat. Rev. Mol. Cell. Biol.* **5**, 33–44.

Heath, C., Cross, N.C. (2004) Critical role of STAT5 activation in transformation mediated by ZNF198-FGFR1, *J. Biol. Chem.* **279**, 6666–6673.

Huang, W., Kessler, D.S., Erikson, R.L. (1995) Biochemical and biological analysis of Mek1 phosphorylation site mutants, *Mol. Biol. Cell* **6**, 237–245.

Iijima, Y., Okuda, K., Tojo, A., Tri, N.K., Setoyama, M., Sakaki, Y., Asano, S., Tokunaga, K., Kruh, G.D., Sato, Y. (2002) Transformation of Ba/F3 cells and Rat-1 cells by ETV6/ARG, *Oncogene* **21**, 4374–4383.

Iwahara, T., Fujimoto, J., Wen, D., Cupples, R., Bucay, N., Arakawa, T., Mori, S., Ratzkin, B., Yamamoto, T. (1997) Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system, *Oncogene* **14**, 439–449.

Jeffrey, P.D., Tong, L., Pavletich, N.P. (2000) Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors, *Genes Dev.* **14**, 3115–3125.

Johnson, D.E., Williams, L.T. (1993) Structural and functional diversity in the FGF receptor multigene family, *Adv. Cancer Res.* **60**, 1–41.

Keung, Y.K., Beaty, M., Steward, W., Jackle, B., Pettnati, M. (2002) Chronic myelocytic leukemia with eosinophilia, t(9;12)(q34;p13), and ETV6-ABL gene rearrangement: case report and review of the literature, *Cancer Genet. Cytogenet.* **138**, 139–142.

Khosravi, R., Maya, R., Gottlieb, T., Oren, M., Shiloh, Y., Shkedy, D. (1999) Rapid ATM-dependent phosphorylation of MDM2 precedes p53 accumulation in response to DNA damage, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14973–14977.

King, C.R., Kraus, M.H., Aaronson, S.A. (1985) Amplification of a novel v-erbB-related gene in a human mammary carcinoma, *Science* **229**, 974–976.

Kulkarni, S., Reiter, A., Smedley, D., Goldman, J.M., Cross, N.C.P. (1999) The genomic structure of ZNF198 and location of breakpoints in the t(8;13) myeloproliferative syndrome, *Genomics* **55**, 118–121.

Lacronique, V., Boureux, A., Monni, R., Dumon, S., Mauchauffe, M., Mayeux, P., Gouilleux, F., Berger, R., Gisselbrecht, S., Ghysdael, J., Bernard, O.A. (2000) Transforming properties of chimeric TEL-JAK proteins in Ba/F3 cells, *Blood* **95**, 2076–2083.

Leonard, J.H., Kearsley, J.H., Chenevix-Trench, G., Hayward, N.K. (1991) Analysis of gene amplification in head-and-neck squamous-cell carcinoma, *Int. J. Cancer* **48**, 511–515.

Llorca, O., Rivera-Calzada, A., Grantham, J., Willison, K.R. (2003) Electron microscopy and 3D reconstruction reveal that human ATM kinase uses an arm-like domain to clamp around double-stranded DNA, *Oncogene* **22**, 3867–3874.

Lyman, S.D., Stocking, K., Davison, B., Fletcher, F., Johnson, L., Escobar, S. (1995) Structural analysis of human and murine flt3 ligand genomic loci, *Oncogene* **11**, 1165–1172.

Lynch, T.J., Bell, D.W., Sordella, R. et al. (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to Gefitinib, *N. Engl. J. Med.* **350**, 2129–2038.

Ma, P.C., Maulik, G., Christensen, J., Salgia, R. (2003) c-Met: structure, functions and potential for therapeutic inhibition, *Cancer Metastasis Rev.* **22**, 309–325.

Marais, R., Light, Y., Paterson, H.F., Mason, C.S., Marshall, C.J. (1997) Differential regulation of Raf-1, A-Raf, and B-Raf by oncogenic ras and tyrosine kinases, *J. Biol. Chem.* **272**, 4378–4383.

Maroe, N., Rottapel, R., Rosnet, O., Marchetto, S., Lavezzi, C., Mannoni, P., Birnbaum, D., Dubreuil, P. (1993) Biochemical characterization and analysis of the transforming potential of the FLT3/FLK2 receptor tyrosine kinase, *Oncogene* **8**, 909–918.

Matsuoka, S., Huang, M., Elledge, S.J. (1998) Linkage of ATM to cell cycle regulation by the Chk2 protein kinase, *Science* **282**, 1893–1897.

McWhirter, J.R., Galasso, D.L., Wang, J.Y. (1993) A coiled-coil oligomerization domain of Bcr is essential for the transforming function of Bcr-Abl oncoproteins, *Mol. Cell. Biol.* **13**, 7587–7595.

Meydan, N., Grunberger, T., Dadi, H., Shahar, M., Arpaia, E., Lapidot, Z., Leeder, J.S., Freedman, M., Cohen, A., Gazit, A., Levitzki, A., Roifman, C.M. (1996) Inhibition of acute lymphoblastic leukemia by a Jak-2 inhibitor, *Nature* **379**, 645–648.

Mora, A., Komander, D., van Aalten, D.M.F., Alessi, D.R. (2004) PDK1, the master regulator of AGC kinase signal transduction, *Semin. Cell Dev. Biol.* **15**, 161–170.

Morris, S.W., Kirstein, M.N., Valentine, M.B., Dittmer, K.G., Shapiro, D.N., Saltman, D.L., Look, A.T. (1994) Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma, *Science* **263**, 1281–1284.

Muenke, M., Schell, U., Hehr, A., Robin, N.H., Losken, H.W., Schinzel, A., Pulleyn, L.J., Rutland, P., Reardon, W., Malcolm, S., Winter, R.M. (1994) A common mutation in the fibroblast growth factor receptor 1 gene in Pfeiffer syndrome, *Nat. Genet.* **8**, 269–274.

Neilson, K.M., Friesel, R. (1996) Ligand-independent activation of fibroblast growth factor receptors by point mutations in the extracellular, transmembrane, and kinase domains, *J. Biol. Chem.* **271**, 25049–25057.

Passos-Bueno, M.R., Wilcox, W.R., Jabs, E.W., Sertie, A.L., Alonso, L.G., Kitoh, H. (1999) Clinical spectrum of fibroblast growth factor receptor mutations, *Hum. Mutat.* **14**, 115–125.

Pluk, H., Dorey, K., Superti-Furga, G. (2002) Autoinhibition of c-Abl, *Cell* **108**, 247–259.

Roche-Lestienne, C., Soenen-Cornu, V., Grardel=Duflos, N., Lai, J.L., Phillip, N., Facon, T., Fenaux, P., Preudhomme, C. (2002) Several types of the Abl gene can be found in chronic myeloid leukemia patients resistant to STI571, and they can pre-exist to the onset of treatment, *Blood* **100**, 1014–1018.

Russell, J.P., Powell, D.J., Cunnane, M., Greco, A., Portella, G., Santoro, M., Fusco, A., Rothstein, J.L. (2000) The TRK-T1 fusion protein induces neoplastic transformation of thyroid epithelium, *Oncogene* **19**, 5729–5735.

Saltzman, A., Stone, M., Franks, C., Searfoss, G., Munro, R., Jaye, M., Ivashchenko, Y. (1998) Cloning and characterization of human Jak-2 kinase: high mRNA expression in immune cells and muscle tissue, *Biochem. Biophys. Res. Commun.* **246**, 627–633.

Shtivelman, E., Lifshitz, B., Gale, R.P., Canaani, E. (1985) Fused transcript of abl and bcr genes in chronic myelogenous leukemia, *Nature* **315**, 550–554.

Sommer, S.S., Jiang, Z., Feng, J., Buzin, C.H., Zheng, J., Longmate, J., Jung, M., Moulds, J., Dritschilo, A. (2003) ATM missense mutations

are frequent in patients with breast cancer, *Cancer Genet. Cytogenet.* **145**, 115–120.

Sozeri, O., Vollmer, K., Liyanage, M., Frith, D., Kour, G., Mark, G.E., Stabel, S. 3rd (1992) Activation of the c-Raf protein kinase by protein kinase C phosphorylation, *Oncogene* **7**, 2259–2262.

Tang, T.C.-M., Sham, J.S.T., Xie, D., Fang, Y., Huo, K.-K., Wu, Q.-L., Guan, X.-Y. (2002) Identification of a candidate oncogene *SEI-1* within a minimal amplified region at 19q13.1 in ovarian cancer cell lines, *Cancer Res.* **62**, 7157–7161.

Thompson, F.H., Nelson, M.A., Trent, J.M., Guan, X.Y., Liu, Y., Yang, J.M., Emerson, J., Adair, L., Wymer, J., Balfour, C., Massey, K., Weinstein, R., Alberts, D.S., Taetle, R. (1996) Amplification of 19q13.1-q13.2 sequences in ovarian cancer. G-band, FISH, and molecular studies, *Cancer Genet. Cytogenet.* **87**, 55–62.

Thorstenson, Y.R., Roxas, A., Kroiss, R., Jenkins, M.A., Yu, K.M., Bachrich, T., Muhr, D., Wayne, T.L., Chu, G., Davis, R.W., Wagner, T.M., Oefner, P.J. (2003) Contributions of ATM mutations to familial breast and ovarian cancer, *Cancer Res.* **63**, 3325–3333.

Tybulewicz, V.L., Crawford, C.E., Jackson, P.K., Bronson, R.T., Mulligan, R.C. (1991) Neonatal lethality and lymphopenia in mice with a homozygous disruption of the c-abl proto-oncogene, *Cell* **65**, 1153–1163.

Van Etten, R.A., Jackson, P., Baltimore, D. (1989) The mouse type IV c-abl gene product is a nuclear protein, and activation of transforming ability is associated with cytoplasmic localization, *Cell* **58**, 669–678.

Weber, C.K., Slupsky, J.R., Herrmann, C., Schuler, M., Rapp, U.R., Block, C. (2000) Mitogenic signaling of Ras is regulated by differential interaction with Raf isozymes, *Oncogene* **19**, 169–176.

Wolfel, T., Hauer, M., Schneider, J., Serrano, M., Wolfel., C., Klehmann-Hieb, E., De Plaen, E., Hankeln, T., Meyer, K.H., Beach, D. (1995) A p16INK4a-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma, *Science* **269**, 1281–1284.

Wong, A.J., Ruppert, J.M., Bigner, S.H., Grzeschik, C.H., Humphrey, P.A., Bigner, D.S., Vogelstein, B. (1992) Structural alterations of the epidermal growth factor receptor gene in human gliomas, *Proc. Natl. Acad. Sci. USA* **89**, 2965–2969.

Zhang, B.H., Guan, K.L. (2000) Activation of B-Raf kinase requires phosphorylation of the conserved residues Thr598 and Ser601, *EMBO J.* **19**, 5429–5439.

Zimmermann, S., Moelling, K. (1999) Phosphorylation and regulation of Raf by Akt (protein kinase B), *Science* **286**, 1741–1744.

Zuo, L., Weger, J., Yang, Q., Goldstein, A.M., Tucker, M.A., Walker, G.J., Hayward, N., Dracopoli, N.C. (1996) Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma, *Nat. Genet.* **12**, 97–99.

Ionizing Radiation Damage to DNA

Clemens von Sonntag
Leibniz-Institut für Oberflächenmodifizierung (IOM), Leipzig, Germany

1 **Energy Absorption 150**

2 **Radiation-induced DNA Lesions 151**

3 **Reaction of DNA Radicals 152**

4 **Model Studies 153**

5 **DNA Strand Breakage and Cross-linking 154**

Bibliography 156
Books and Reviews 156
Primary Literature 157

Keywords

Cross-Link
Covalent bond between two macromolecular moieties, usually formed when two macromolecular free radicals recombine.

DNA Strand Breakage
Event induced by certain kinds of chemical damage, through free-radical processes involving the DNA molecule but also through the action of enzymes that try to repair the DNA damage.

Free Radical
Chemical species that is highly reactive because it possesses an unpaired electron.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

Heteroatom
Atom in a molecule of an organic compound that is neither a carbon nor a hydrogen atom.

Ionizing Radiation
Energetic photons (X rays, γ-rays) and energetic particles (β-rays, fast electrons from electron accelerators, positrons, α-rays and other fast-moving ions, neutrons) that cause multiple ionization events along their trajectory through matter.

Peroxyl Radical
Generated by addition of dioxygen to a free radical.

Pulse Radiolysis
Technique for studying the kinetics of the chemical reactions of short-lived intermediates: for example, the aqueous solution of a substrate is exposed to a short (nanoseconds-to-microseconds range) burst of high-energy electrons from an electron accelerator.

Radical Ion
Free radical with a positive or negative charge.

Ionizing radiation represents an important risk factor to the living organism. It has always been part of the natural environment, hut it is also now a technological phenomenon, lending added urgency to the study of its effects on living matter. It has been established that the most sensitive target in the living cell is the DNA, which undergoes radiation damage through free-radical processes. This effect has two practical aspects beyond its purely scientific interest: to assess the radiation risk with a view to minimize it, perhaps by chemical means, and to apply ionizing radiation selectively for therapeutic purposes (e.g. in cancer treatment). These goals will not be optimally achieved without adequate knowledge of the chemical processes that ionizing radiation sets in train in DNA.

1
Energy Absorption

Absorption of ionizing radiation by matter leads to the formation of radical cations and electrons (reaction 1) and electronically excited molecules (reaction 2).

$$\mathrm{M} + \text{ionizing radiation} \longrightarrow \mathrm{M}^{\bullet +} + \mathrm{e}^{-} \quad (1)$$

$$\mathrm{M} + \text{ionizing radiation} \longrightarrow \mathrm{M}* \quad (2)$$

Energy deposition is inhomogeneous. In the case of sparsely ionizing radiation

(e.g. γ-rays or high-energy electrons), the average distance between two regions of energy deposition is about 200 Å in matter of about unit density (i.e. in aqueous media). With densely ionizing radiation (e.g. α-particles) these regions, which are called *spurs* and are centered on the location of a primary ionization event, strongly overlap to form a continuous ionization track. The rate of energy deposition is proportional to the electron density. In the living cell, about 70% of the energy is absorbed by the water and about 30% by the organic matter and other solutes.

The radiolysis of water leads to the formation of $^{\bullet}OH$ radicals, solvated electrons, and $H^{\bullet}$ atoms. The water radical cation readily transfers a proton to the neighboring water molecules, thereby yielding an $^{\bullet}OH$ radical (reaction 3), and the electron becomes solvated (reaction 4). Electronically excited water may decompose into $^{\bullet}OH$ and $H^{\bullet}$ (reaction 5).

$$H_2O^{\bullet +} \longrightarrow {}^{\bullet}OH + H^+ \qquad (3)$$

$$e^- + nH_2O \longrightarrow e^-_{aq} \qquad (4)$$

$$H_2O* \longrightarrow H^{\bullet} + {}^{\bullet}OH \qquad (5)$$

For sparsely ionizing radiation, the radiation-chemical yields, or G values, (The G value is expressed in molecules per 100 eV, which corresponds to 1.037×10^{-7} mol J^{-1}.) are $G(^{\bullet}OH) = 2.8 \times 10^{-7}$ mol J^{-1}, $G(e^-_{aq}) = 2.7 \times 10^{-7}$ mol J^{-1}, $G(H^{\bullet}) = 0.6 \times 10^{-7}$ mol J^{-1}. In the spurs, some radical combination occurs yielding the "molecular products" H_2 and H_2O_2 ($G(H_2) = 0.45 \times 10^{-7}$ mol J^{-1}, $G(H_2O_2) = 0.8 \times 10^{-7}$ mol J^{-1}). The water-derived radicals are all highly reactive. Hence, their formation and reactions cause an important part, possibly the major part, of the radiation damage to the living cell.

2 Radiation-induced DNA Lesions

Ionizing radiation is absorbed by the various cell components with practically equal probability. This is in contrast with UV radiation, which is predominantly absorbed by the cellular nucleic acids. While the mechanism of inactivation in the two cases (UV vs ionizing radiation) is quite different, DNA again is by far the most sensitive target to damage induced by ionizing radiation. Subionization UV radiation mainly causes base dimerizations (e.g. formation of thymine dimers at 260 nm) with barely any strand breakage. Ionizing radiation produces very little of this type of damage but causes, characteristically, DNA strand breakage. Ionizing radiation base damage is largely due to free-radical reactions, including those caused by the $^{\bullet}OH$ radicals generated by the radiation absorbed in the aqueous medium surrounding the DNA. In this way, a considerable number of different kinds of damage arise whose products one may group into the following categories:

Altered bases

Altered sugar moiety
- DNA strand break
- Release of unaltered bases

Cross-links
- DNA-DNA
- DNA-protein

Apart from phosphate-ester bond cleavage, which constitutes a strand break, the phosphate moiety as such is not modified. It is important to realize that the cluster-type energy deposition of the ionizing radiation will cause some of the damaging events also to appear in clusters; that is, there is a considerable likelihood

that, especially in double-stranded DNA, one damaged site will be close to another, or even to several others. These "clustered lesions" have also been termed *locally multiply damaged sites* (LMDS). They may involve damaged bases and one or two single-strand breaks. Two opposite single-strand breaks will result in a DNA double-strand break.

In the living cell, the DNA is surrounded not only by proteins (e.g. histones in eukaryotic cells) but also by a high concentration of low molecular weight organic material, among others the thiol glutathione. These react very efficiently (i.e. at practically diffusion-controlled rates) with the water-derived radicals ($^{\bullet}OH$, e^{-}_{aq}, $H^{\bullet}$) formed in the aqueous surrounding of DNA. Hence, these are effectively scavenged, and so the DNA is largely protected against, say, $^{\bullet}OH$-attack. Only $^{\bullet}OH$ radicals generated in the close vicinity of DNA stand some chance of reacting with and hence doing damage to the genetic material.

3 Reaction of DNA Radicals

Damage by ionizing radiation, induced either by the direct effect or by water radicals, will result in DNA radicals; these are the precursors of the final (nonradical) damage at the product stage. The thiol glutathione (GSH), which is present in living cells at comparatively high concentrations (approaching 10^{-2} mol dm^{-3}) can react with the DNA radicals ($R^{\bullet}$) by H-transfer (reaction 6).

$$R^{\bullet} + GSH \longrightarrow RH + GS^{\bullet} \qquad (6)$$

This reaction is thermodynamically slightly favored because the S−H bond is relatively weak. Recently, it has been realized that the peptide C−H bond in proteins is even somewhat lower, but it has not yet been established whether proteins attached to DNA (e.g. histones) reduce $^{\bullet}OH$-induced damage not only by scavenging $^{\bullet}OH$ radicals but also by H-donation to a DNA radical.

In competition with this H-donation reaction, dioxygen can add to the DNA radicals, thereby forming the corresponding peroxyl radicals (reaction 7).

$$R^{\bullet} + O_2 \longrightarrow ROO^{\bullet} \qquad (7)$$

These peroxyl radicals give rise to the final products. Some of the peroxyl radicals are capable of eliminating $HO_2^{\bullet}/O_2^{\bullet-}$, but the major part will decay bimolecularly (reaction 8).

$$2ROO^{\bullet} \longrightarrow \text{products} \qquad (8)$$

Low-molecular–weight peroxyl radicals often terminate at dose to diffusion-controlled rates. In high-molecular–weight material such as DNA, the diffusion of radical-bearing segments is considerably restricted. Hence, they may persist for relatively long times and may undergo reactions other than reaction (8) or the elimination of $HO_2^{\bullet}/O_2^{\bullet-}$.

The reaction of the primary DNA radicals with glutathione is usually termed *chemical repair*. However, a repair in the true sense is achieved only if the radical site to be repaired has been created by H-abstraction (e.g. reaction 9).

$$RH + {}^{\bullet}OH \longrightarrow R^{\bullet} + H_2O \qquad (9)$$

In DNA, such sites are present only at the sugar moiety or at the methyl group of thymine. For the most part, $^{\bullet}OH$ reacts by addition to a C=C or C=N bond (e.g. reaction 10). Subsequent "repair" by H-donation leads to the formation of

a hydrate (reaction 11) rather than the original molecule.

$$R_2C{=}CR_2 + {}^{\bullet}OH \longrightarrow HOCR_2{-}CR_2^{\bullet} \quad (10)$$

$$HOCR_2{-}CR_2^{\bullet} + GSH \longrightarrow HOCR_2{-}CHR_2 + GS^{\bullet} \quad (11)$$

Apparently, the cellular enzymatic repair system can cope with this kind of damage much better than the kind resulting from peroxyl radical reactions (cf the "oxygen effect" in radiobiology).

4 Model Studies

Although the chemical nature of some kinds of base damage has been studied *in vivo*, most of our knowledge of radiation-induced DNA damage is derived from model studies. Electron spin resonance studies for the identification of radicals have mainly been carried out in the solid state at low temperatures. In contrast, practically all product and kinetic (pulse radiolysis) studies have been done in dilute aqueous solutions. Thus, they reflect mostly the indirect effect.

In the direct effect, radical cations (and electrons) are the species primarily produced (reaction 1). To mimic this reaction and to study the fate of radical cations in aqueous solutions, photoionization with a laser (monophotonic at $\lambda = 193$ nm or biphotonic at $\lambda = 248$ nm) or electron transfer to photoexcited quinones (Q) were used with some advantage (e.g. reaction 12).

$$Q* + \text{nucleobase} \longrightarrow Q^{\bullet -} + \text{nucleobase}^{\bullet +} \quad (12)$$

The nucleobase radical cations are strong acids. Quick deprotonation ensues at a heteroatom, but in thymine, deprotonation can eventually materialize at the exocyclic methyl group carbon, in competition with a nucleophilic addition of water to the carbon 6 position. Although after deprotonation the heteroatom-centered radical predominates, the radical cation in near-neutral media is always present at low "equilibrium" concentrations. In contrast to the situation at the heteroatom, deprotonation at the methyl position is practically irreversible under these conditions.

Nucleobase cations are strong oxidants (for a sugar-derived radical cation see below), and in DNA those derived from thymine, cytosine, and adenine may oxidize a neighboring guanine (G), the nucleobase with the lowest reduction potential. Hole transfer through DNA is now a well-documented process. The ultimate sinks of the hole are GG doublets and, even better, GGG triplets whose reduction potentials are substantially lower than that of a single G. Mechanistically, this hole transfer can be described by an incoherent hopping process involving quantum-mechanical channels.

The electrons formed in the ionization process are readily scavenged by the nucleobases (a diffusion-controlled reaction). The radical anions thus formed are strong bases. Hence, it is not surprising that those derived from adenine, guanine, and cytosine are protonated by water on the submicrosecond timescale. The thymine radical anion has a pK_b value of about 7.0, and therefore is much longer lived than the other radical anions.

Most of these radical anions (and to a lesser extent their heteroatom-protonated forms) have pronounced reducing properties; that is, they are capable of retransferring the electron to an oxidant.

However, they are metastable species with respect to ultimate conversion into carbon-protonated intermediates (e.g. in thymine, the final protonation site is C-6). These radicals no longer have reducing properties.

This sequence of events also has a bearing on the reactions of the radical anions with dioxygen. For example, the thymine radical anion and its heteroatom-protonated form react with dioxygen by forming $O_2^{\bullet -}/HO_2^{\bullet}$ thereby regenerating the nucleobase, while in the reaction of dioxygen with the C-6-protonated radical anion, the thymine molecule is destroyed.

Besides the formation of nucleobase radical cations (on account of the direct effect) and radical anions (by e^- attachment), one must consider the reactions of the $^{\bullet}OH$ radical as a major contributor to radiation-induced DNA damage. It mostly adds to the double bonds of the nucleobases, but it also abstracts H atoms from the sugar moiety and the methyl group in thymine.

For the investigation of $^{\bullet}OH$-induced reactions in isolation, in radiation-chemical experiments it is standard practice to convert e_{aq}^- (from reaction 4) into $^{\bullet}OH$ by saturating the solution to be irradiated with nitrous oxide (reaction 13).

$$e_{aq}^- + N_2O \longrightarrow {}^{\bullet}OH + N_2 + OH^- \tag{13}$$

As a result, the radical species now consist of 90% $^{\bullet}OH$ and 10% $H^{\bullet}$; that is, the observed reactions and their products are dominated by the effects of the $^{\bullet}OH$ radical. In the pyrimidines, $^{\bullet}OH$-addition to the C-5 position yields a reducing radical, while an addition to the C-6 position yields a radical with oxidizing properties. These properties can be defined using suitable redox probes. The pulse radiolysis technique has allowed the characterization of these radicals and the determination of their yields. In the case of the purines, the redox titration technique is not as straightforward, and the assignment of the sites of $^{\bullet}OH$-addition is complicated by rapid ring-opening and water elimination reactions.

Our present knowledge of pyrimidine free-radical chemistry (in particular, uracil and thymine, their methyl derivatives as well as their nucleosides) is much more extensive than that of the purines. It is obvious from the published data that only a fraction of the primary purine $^{\bullet}OH$-adduct radicals have shown up in the form of products. Considerable effort will be required to bring the purine (and cytosine) free-radical chemistry to a satisfactory level of understanding. This situation is also reflected in the determination of the base product yields from irradiated DNA, where a considerable deficit in the product yields (related to the primary $^{\bullet}OH$ radical yield) is observed.

5
DNA Strand Breakage and Cross-linking

Solvated electrons do not cause DNA strand breakage, but $^{\bullet}OH$ radicals do. In competition with addition to the nucleobases, they also abstract H atoms from the sugar moiety. This has two possible consequences: in the subsequent reactions a strand break is induced and an unaltered base is released, or base release occurs without giving rise to strand breakage. Hence, base release always predominates somewhat over strand breakage. A number of altered sugars that are related to these processes have been identified, both in the absence and in the presence of dioxygen. On the basis of detailed model studies, the reactions and their kinetics

leading to strand breakage in the absence of dioxygen are fairly well understood. The primary step is the abstraction of the H atom at C-4′:

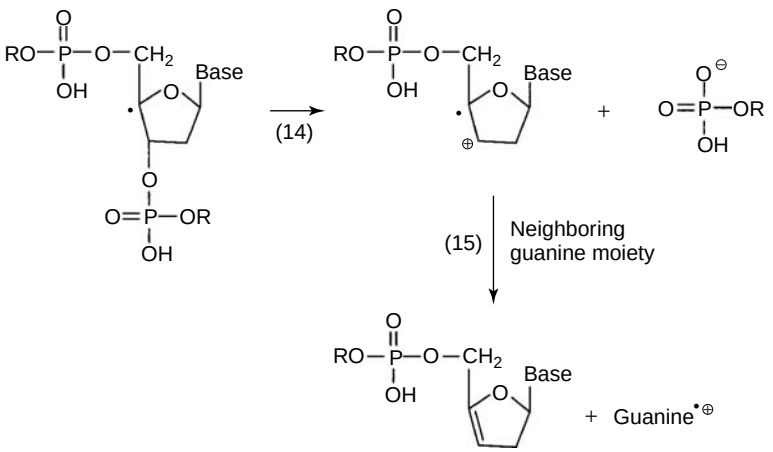

This radical then eliminates a neighboring phosphate (linked to a fragment of the DNA strand) (reaction 14), leaving behind a radical cation. This radical cation has oxidizing properties and can oxidize a neighboring guanine moiety (reaction 15) with the consequence of hole migration through DNA (see above). In competition, the sugar radical cation reacts with water, either at the position that has eliminated the phosphate, or at C-4′. In the former case, the other phosphate function may be eliminated by the same mechanism; in double-stranded DNA, this sequence of events produces a clean gap in the affected strand (the end groups of the two fragments are phosphate groups) with the loss of some information because of the disappearance of the damaged nucleoside. In the latter case, the base is also lost, but additionally at the end of one of the fragment strands, a damaged sugar remains linked to the phosphate group. Enzymatically speaking, this is a "dirty" end group. Details of the mechanism of DNA strand breakage under conditions of oxygenation are less well understood, but some of the relevant sugar lesions have been detected. Model systems (ribose 5-phosphate) indicate that under these conditions C-5′ should be an additional site of attack and, in analogy, one would expect (no experimental evidence yet) the C-3′ peroxyl radical also to be a potential precursor for strand breakage.

So far, sugar damage has been discussed only in terms of •OH radicals attacking this moiety. A contribution of the direct effect (ionization of the sugar moiety and the phosphate groups) must also be considered, but experimental evidence is not yet available. However, there is another interesting aspect. In polynucleotides such as poly(U) and poly(C), there is convincing evidence that base radicals are the major precursors of the sugar radicals that lead to strand breakage and the release of an

unmodified base at the site of the damaged sugar. It is less clear whether such a radical transfer from the base to the sugar moiety can also occur in DNA.

In mammalian cells, DNA double-strand breaks are observed alongside single-strand breaks approximately in the ratio of 1 : 25. This poses the question of how these double-strand breaks are formed. It has been argued here that they result from clustered lesions. In the literature, an additional one-hit route has been suggested that involves a radical transfer from the already broken strand to the sugar moiety of the opposite strand, followed by breakage of this strand.

Carbon-centered radicals are known to add to the C=C bonds of nucleobases. Such reactions, as well as radical–radical combination reactions involving macroradicals, in principle allow the formation of DNA–protein and DNA–DNA cross-links. In special cases, such products have been observed with biological material, albeit in yields considerably lower than DNA double-strand breaks.

See also Free Radicals in Biochemistry and Medicine; Mutagenesis, Malignancy and Genome Instability; Nucleic Acids (DNA) Damage and Repair.

Bibliography

Books and Reviews

Dizdaroglu, M. (1991) Chemical determination of free radical-induced damage to DNA, *Free Radical Biol. Med.* **10**, 225–242.

Hüttermann, J. (1991) Radical ions and their reactions in DNA and its constituents, in: Lund, A., Shiotani, M. (Eds.) *Radical Ionic Systems*, Kluwer, Dordrecht, pp. 435–462.

Michael, B.D., Held, K.D., Harrop, H.A. (1983) Biological Aspects of DNA Radioprotection, in: Nygaard, O., Simic, M.G. (Eds.) *Radioprotectors and Anticarcinogens*, Academic Press, New York, pp. 325–338.

Teebor, G.W., Boorstein, R.J., Cadet, J. (1988) The repairability of oxidative free radical mediated damage to DNA: A review, *Int. J. Radiat. Biol.* **54**, 131–150.

von Sonntag, C. (1987) *The Chemical Basis of Radiation Biology*, Taylor & Francis, London.

von Sonntag, C., Hagen, U., Schön-Bopp, A., Schulte-Frohlinde, D. (1981) Radiation-induced strand breaks in DNA: chemical and enzymatic analysis of end groups and mechanistic aspects, *Adv. Radiat. Biol.* **9**, 109–142.

Ward, J.F. (1988) DNA damage produced by ionizing radiation in mammalian cells: identities, mechanisms of formation, and repairability, *Prog. Nucleic Acid Res. Mol. Biol.* **35**, 95–125.

Primary Literature

Giese, B., Amaudrut, J., Köhler, A., Spormann, M., Wessely, S. (2001) Direct observation of hole transfer through DNA by hopping between adenine bases and by tunnelling, *Nature* **412**, 318–320.

Schulte-Frohlinde, D., Simic, M.G., Görner, H. (1990) Laser-induced strand break formation in DNA and polynucleotides, *Photochem. Photobiol.* **52**, 1137–1151.

Labeling, Biophysical

Gertz I. Likhtenshtein
Ben-Gurion University of the Negev, Beer-Sheva, Israel

1 Principles 159
1.1 Spin Labeling 159
1.1.1 General 159
1.1.2 Rotational Diffusion of Nitroxide Label 160
1.1.3 Nitroxides as Dielectric, pH, Redox, and Imaging Probes 160
1.1.4 Double-labeling Technique. Spin Label–Spin Probe Method: Spin-oxymetry 161
1.1.5 New ESR Techniques 162
1.2 Luminescent Labeling 162
1.2.1 General 162
1.2.2 Rotational Diffusion of Luminescent Labels 162
1.2.3 Molecular Dynamics and Micropolarity of the Media 163
1.2.4 Resonance Energy Transfer Between Chromophores; Quenching of Fluorescence; Fluorescent Recovery after Photobleaching (FRAP) 164
1.2.5 Triplet, Photochrome, Triplet–photochrome, and Spin-triplet-photochrome Labeling 164
1.2.6 Dual Fluorophore-spin Labeling: High-sensitivity Redox Probes, Spin and Nitric Oxide Traps 165
1.2.7 New Fluorescence Technique 166
1.3 Miscellaneous Labeling Methods 168
1.3.1 Mössbauer Labels 168
1.3.2 Nuclear Magnetic Resonance Probes 168
1.3.3 Electron and X-ray Scattering Labels 168

2 Applications 168
2.1 Enzymes and Proteins 168
2.1.1 Active Sites of Enzymes 168
2.1.2 Conformational Changes and Molecular Dynamics 170

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

2.2 Membranes 172
2.3 Nucleic Acids 173
2.4 Cells and Organs 174
2.5 Miscellaneous 175
2.5.1 Biologically Active Compounds, Biological Analysis, Pharmokinetics 175

3 New Trends in Biophysical Labeling 176

Bibliography 176
Books and Reviews 176
Primary Literature 177

Keywords

Dipole–dipole Interaction
Arises from Coulomb or magnetic fields mutually induced by electrostatic or magnetic dipoles respectively.

Exchange Interaction
Originates from inter- or intramolecular interaction between electrons due to their quantum mechanical delocalization (e.g. overlap of orbitals populated with the electrons).

Label
A compound that binds covalently to an object under study and whose properties enable monitoring by appropriate physical methods.

Probe
A compound that either adds on non-covalently or diffuses freely in the medium.

Spin
The intrinsic angular momentum of an unpaired electron that induces magnetic momentum.

In solving problems of molecular biology, molecular biophysics and biochemistry of proteins, nuclear acids, and biomembranes, it is necessary to investigate molecular dynamics properties of biological objects and spatial disposition of their individual parts. One must also know the depth of immersion of definite centers in a biological matrix, that is, the availability of enzyme sites to substrates, distance of electron tunneling between a donor and an acceptor group, position of a biophysical label in a membrane and in a protein globule, distribution of the electrostatic field around molecules, and so on. The biophysical labeling approach consists, in principle, of

the modification of chosen sites of an object of interest by special compounds (labels and probes) whose properties make it possible to trace the state of the biological matrix by appropriate physical methods. Four types of compound are most often used as labels and probes: (1) centers with unpaired electrons, spins (stable nitroxide radicals and paramagnetic metal complexes); (2) fluorescent and phosphorescent chromophores; (3) electron-scattering groups of heavy atoms (e.g. polygold or polymercury compounds); and (4) Mössbauer atoms (e.g. ^{57}Fe), which give the nuclear γ-resonance spectra. Three types of problems can be solved by means of labeling technique on a molecular level: (1) recording of conformational changes in proteins, enzymes, membranes, and other structures; (2) the investigation of the microrelief, micropolarity, and intra- and intermolecular mobility of biological objects; and (3) the determination of the distance between chosen parts of systems of interest. The labeling approach makes it possible to work in whole native cellular and subcellular structures in aqueous, diluted solutions. The measurements are fast and widely available.

1 Principles

1.1 Spin Labeling

1.1.1 General

In the overwhelming majority of stable nitroxide radicals (NRs) used as spin labels and probes, the nitroxide group is stabilized by a substituent on the α-carbon atoms; methyl groups, for example (Fig. 1). Spin labels generally include fragments of already known, widely accepted specific reagents for functional groups of proteins, lipids, nucleic acids, and other biological compounds. The nitroxide group with unpaired electron produces simple, well resolved, and comparatively easily treated electron spin resonance (ESR) spectra that are sensitive to the molecular motion of the nitroxide band and its interactions with other paramagnetic materials. In a constant magnetic field of strength H_0, electron spins orient themselves in two directions, that is, along the field and against the field (the Zeeman Effect). Superposition of a perpendicular electromagnetic field of frequency ν in the microwave region results in reorientation of the spins and resonance absorption of the microwave energy occurs. The condition of the resonance is the equality

$$h\nu = g\beta H_0, \qquad (1)$$

where β is the Bohr magneton and g is the g-factor that is characteristic of magnetic momentum of the nitroxide.

The unpaired electron of nitroxide is delocalized over the oxygen and nitrogen. In the spherical symmetrical S state, the delocalized electron undergoes a contact interaction with the spin of the nitrogen nucleus that causes (for ^{14}N) isotropic hyperfine splitting (A_{iso}) into three components corresponding to orientation of the nuclear spin with magnetic quantum number $m = 1$ along ($m = -1$), against ($m = +1$), and perpendicular to ($m = 0$) the constant magnetic field, correspondingly. The A_{iso} is proportional to the electron spin density on the nuclear spin.

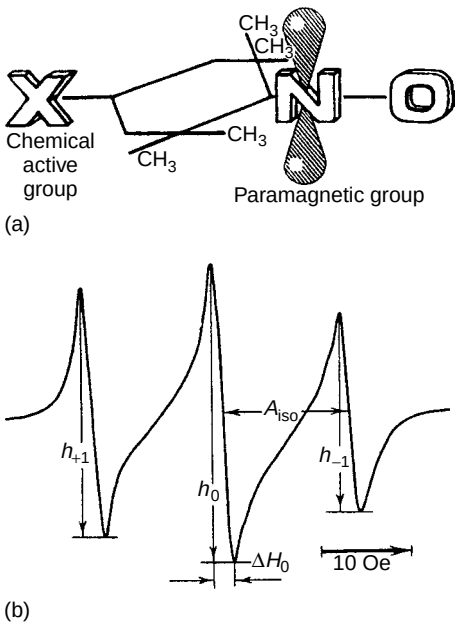

Fig. 1 (a) Schematic presentation for a nitroxide spin label, and (b) schematic spectrum of a nitroxide in solution.

Another type of electron–nuclear spin interaction is associated with localization of the electron on the p-orbital of nitrogen, and therefore, depends on the angle between the direction of the axes of the p-orbital and the constant magnetic field. Such an interaction gives rise to anisotropic hyperfine splitting (A_{aniso}).

1.1.2 Rotational Diffusion of Nitroxide Label

The nitroxides are asymmetric particles (Fig. 1a) characterized by three values of the g-factor (g_x, g_y, g_z) and A_{aniso} (A_x, A_y, A_z). In the immobilized state (frozen or supercooled solutions), the ESR spectra are superpositions of the *microspectra* of an enormous number of radicals with different orientations. The motion of the radicals results in averaging anisotropic effects. According the theory, a value of the time of the nitroxide isotropic rotational diffusion by one radian (correlation time, τ_c) in the region $\tau_c = 1 \times 10^{-10} \div 2 \times 10^{-9}$ s can be calculated from ESR spectra by using the formula,

$$\tau_c = \frac{2.7 \times 10^{-10}}{\left\{\left[\left(\frac{I_0}{I_-}\right)^{1/2} - 1\right]\Delta H_0\right\}} \quad (2)$$

where ΔH_0 is the width of the central component of the spectrum in milliteslas (mT) and I_0 and I_{-1} are the intensities of components with $m = 0$ and $m = -1$ respectively (Fig. 1b). Special methods were developed for the measurement $\tau_c = 10^{-4} \div 10^{-8}$ s.

1.1.3 Nitroxides as Dielectric, pH, Redox, and Imaging Probes

The magnetic parameters of NRs, the g-factor, and the hyperfine splittings A_{iso} and A_{aniso} are sensitive to the properties of the medium (e.g. micropolarity and ability to form hydrogen bonds with the $N - \dot{O}$ group). The physical reason for such dependence is that the contribution of the resonance structure on the right

$$N - \dot{O} \longleftrightarrow \dot{N}^+ - O^-$$

causes an increase in the spin density of unpaired electrons on the nitrogen atom and therefore increases the A_{iso} and A_{aniso} values. For example, $A_{iso} = 1.4 + 0.064\ (\varepsilon_0 - 1)/(\varepsilon_0 + 1)$, where ε_0 is the dielectric constant and A_{iso} is given in millitesla. Protonation of the $N - \dot{O}$ group leads to the value of $\Delta A_{iso} = 0.57$ mT. At protonation of other groups of the nitroxide rings, the value ΔA_{iso} can reach 0.35 mT. The g-factor is also reported to be sensitive to the medium.

NRs are relatively stable toward oxidation, but they can be readily reduced to the

corresponding hydroxylamines. The redox potential of a nitroxide (0.2–0.3 eV) is high enough to oxidize such biological compounds as ascorbic acids, semiquinones, superoxide radical, and Fe^{+2}. Therefore, NRs can serve as redox probes.

1.1.4 Double-labeling Technique. Spin Label–Spin Probe Method: Spin-oxymetry

Under certain circumstances, the ESR spectra of paramagnetic centers will respond suitably to the approach of other centers. Two types of spin–spin interactions are distinguished: (1) dipole–dipole interaction associated with induction by the magnetic dipole of one paramagnetic group of a local magnetic field at the site of another paramagnetic center; and (2) exchange interaction caused by an overlap of the orbitals of unpaired electrons.

For a pair of dipoles, the value of magnetic interaction is

$$\Delta H_{dd} = \frac{\mu_1\mu_2}{r^3} - \frac{(\mu_1 r)(\mu_2 r)}{r^5} \qquad (3)$$

where μ_1 and μ_2 are magnetic moments and r is the distance between the spins.

The degree of overlap of orbitals with unpaired electrons is quantitatively characterized by the value of the exchange integral I. The value of I is about 10^{13} s^{-1} at the van der Waals distance and decreases exponentially with increasing distance between the spins. The exchange interaction usually prevails over the dipole–dipole interaction at $r < 0.8$–1.0 nm and therefore is suitable for the study of closely disposed paramagnetic materials.

Dipole–dipole interaction is applicable to systems with remote centers. For example, the distance between the pairs of nitroxide spin labels up to 3.0 to 3.5 nm can be calculated by formula:

$$r = 1.6(\Delta H_{dd})^{-1/3}, \qquad (4)$$

where ΔH_{dd} (mT) is the broadening of the ESR line.

Recently, many modifications of pulse ESR have been designed that allowed the improvement of the distance-measurement accuracy and to expand range of distance available for ESR spectroscopy. The principal advantage of such pulse methods is the direct determination of spin-relaxation parameters that, in turn, are directly related to spin–spin interactions, depending on distances.

The investigation of the paramagnetic contribution of ions to the value of the spin–lattice relaxation rate spin label has resulted in the extension of the distance that can be measured up to 6 nm. When a nitroxide spin label encounters a spin probe, that is, a chemically inert paramagnetic species (e.g. ferricyanide or diphenylchromium) that diffuses freely in solution, the magnitude of broadening of the label ESR line ΔH (mT) is related to the rate constant of the exchange interaction k_{ex} ($M^{-1}s^{-1}$) by the following equation:

$$\Delta H = 6.52 \times 10^{-9} k_{ex} C \qquad (5)$$

where C is the molar spin probe concentration. Measurement of ESR spectra parameters allows one to determine the value of k_{ex} in the range of 10^6 to 10^{10} M^{-1} s^{-1}. For a particular pair of paramagnetic species, the value of k_{ex} is proportional to the collision frequency, and therefore, depends on chemical structure, microviscosity, steric hindrances, concentration of O_2, and distribution of electrostatic charges in the region of encounters with the use nitroxide probes of different electrostatic charges. Hence, after preliminary empirical calibration, the spin label–spin probe method based on examining k_{ex} can be used for experimental study of these factors in biological objects.

1.1.5 New ESR Techniques

In high-frequency ESR spectroscopy (HF ESR) that generates strong magnetic fields (up to 9 T), the increased magnetic field leads to increasing spectral sensitivity to motion dynamics sensitivity in a slow-motion regime. Recently, significant progress in HF ESR has been achieved with the use of millimeter-wave quasi-optic technique, permitting the construction of a 9-T, 250-GHz (1.2 mm) spectrometer.

The 2D ESR technique relies on selective irradiation of a particular resonance line with a microwave frequency field and observation of the resulting effects in the rest of the spectrum. This method provides a 2D display of the homogeneous linescape across an inhomogeneous ESR spectrum. This approach allows direct study of dynamic processes (rotational and translational diffusion, electron transfer) and static dipole and exchange spin–spin interactions. The possibility of the two-dimensional approach has been significantly extended with the use Fourier transform electron spin resonance (2D FT ESR) and electron spin–echo techniques. Multiple quantum ESR recently developed for measuring distances between spins (r) longer than 1.2 nm is based upon double-quantum coherence pulsed ESR methods. Introducing an extensive cycling of a four-pulse sequence allows the determination of dipole–dipole splitting in the homogeneous ESR spectrum. The latter is directly connected to the r value. The electron nuclear double resonance (ENDOR) method provides direct structural information about a paramagnetic molecule orientation and conformation. Another method that is important for structure assignment is the electron–nuclear–nuclear triple resonance spectroscopy, which is an extension of the ENDOR method.

Among applications of new ESR techniques based on continuous wave illumination (CW ESR), the following ones can be pointed out: (1) site-directed spin labeling; (2) rigid incorporation of a spin label in proteins backbone alpha chain; (3) using fast Fourier transform deconvolution; (4) ENDOR of labeled enzyme-active centers; (5) pairwise interaction spin–spin interaction on a solely tumbling macromolecule; (6) and measurements of depth of immersion and location of paramagnetic centers.

1.2 Luminescent Labeling

1.2.1 General

The main parameters of both fluorescence and phosphorescence are their intensity I, the line shape and frequency of the maxima of spectra ν_{max}, the quantum yield ϕ, and the lifetime of the excited state τ^*. A significant difference between fluorescence and phosphorescence is that the lifetimes of the excited state during fluorescence ($\tau^*_f = 10^{-8} - 10^{-10}$ s) are much shorter than phosphorescence lifetimes ($\tau^*_{ph} = 10^{-3} - 10^{2}$ s). Another distinguishing feature of a phosphorescent molecule is the formation of the triplet, a paramagnetic state after excitation, and such a chromophore may serve as a triplet label.

1.2.2 Rotational Diffusion of Luminescent Labels

Chromophores that are being used as luminescent labels usually have asymmetric adsorption and emission transition moments and are sensitive to the polarization of incident light. If the rotation time τ_R of the chromophore is comparable to or shorter than its excited lifetime τ^*, the emitted light will be depolarized (Fig. 2). A characteristic parameter of depolarization

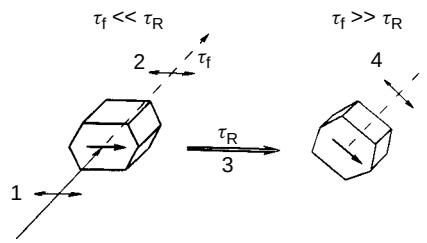

Fig. 2 Schematic representation of the dependence of fluorescent depolarization of a chromophore on the rotational correlation time τ_R and the excited state lifetime τ_f: 1, irradiation by polarized incident light; 2, polarized fluorescence from *immobile* chromophore; 3, rotation of the chromophore; 4, depolarized fluorescence from mobile chromophore.

is the anisotropy,

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}}, \tag{6}$$

where $I_{\parallel}$ and $I_{\perp}$ are the intensities of emitted light passing through the polarizer with the electric axis directed parallel or perpendicular, respectively, to the polarization of the incident light. For a spherical molecule,

$$\frac{r_0}{r} = 1 + \frac{\tau^*}{\tau_R} = 1 + \frac{3k_B T\tau^*}{4\pi b^3 \eta} \tag{7}$$

where r_0 is the anisotropy for a chromophore without rotation, β is the radius of rotation, and η is the viscosity of the medium. Measurement of the dependence of r/r_0 on T/η allows determination of the value of τ_R if τ^* is known. The value of τ_R can be measured directly from the change in the anisotropy after pulse radiation of the sample under study. The aforementioned technique is concerned with rotational diffusion of protein molecules in solution or lipids in membranes with $\tau_R = 10^{-8} - 10^{-9}$ s (fluorescence) and macromolecules in membranes $\tau_R = 10^{-5} - 10^{-6}$ s (phosphorescence).

1.2.3 Molecular Dynamics and Micropolarity of the Media

The excitation of a chromophore is accompanied by a change in the electric dipole moment of the molecule. This involves a change in the energy of interaction with the surrounding dipole molecules. If the characteristic dipole rotational correlation time τ_R in the medium is much longer than the lifetime of the fluorescence or phosphorescence τ^*, the dipoles have no chance to follow the change in the light-induced electric field, and the transition from the *unsolvated* excited state to the ground state takes place. In another limiting case, where $\tau_R \ll t^*$, the interaction between the dipoles and excited molecule lowers the energy of the system, and emission is effected from the *solvated* level. The relaxation can be followed by time-resolved spectroscopy, measuring the kinetic value of the shift of the maximum of the fluorescent or phosphorescent spectra $\nu_{max}(t)$ (relaxation shift):

$$\nu_{max}(t) - \nu_{max}(\infty) = [\nu_{max}(0) - \nu_{max}(\infty)] \times \exp\left(-\frac{t}{\tau_r}\right), \tag{8}$$

where the indexes t, ∞, and 0 are related to the ν_{max} of the time-resolved emission spectrum at a given moment t, $t \to \infty$ and $t = 0$. The value of τ_R can be also estimated by measuring the temperature dependence of the ν_{max} of steady state emission spectra. The value of the relaxation shift depends on the electrostatic dipole moment of

the excited chromophore and on the parameters of polarity (in the simplest case, dielectric constant) of the medium. Hence, the latter can be estimated for biological objects (e.g. active centers of enzymes, membranes) after preliminary calibration, using model systems with known properties.

1.2.4 Resonance Energy Transfer Between Chromophores; Quenching of Fluorescence; Fluorescent Recovery after Photobleaching (FRAP)

The electrostatic dipole–dipole interaction between chromophores permits the nonradiative transfer of energy from an excited singlet state of a molecule (a donor, D) to an unexcited singlet state of another molecule (an acceptor, A) via an inductive resonance mechanism. According the theory developed by Förster, the efficiency of the transfer:

$$E_T = 1 - \frac{\phi}{\phi_0} = \frac{R_0^6}{R_0^6 + R^6} \qquad (9)$$

where ϕ and ϕ_0 are the quantum yields of the donor fluorescence with and without the acceptor; R is the distance between D and A, and R_0 is the critical distance at which the probability of energy transfer is equal to that of donor excited state deactivation:

$$R_0 = \frac{162K^2}{\pi^5 n^4} \int_0^\infty \frac{f_D(\nu)\varepsilon_A(\nu)\,\mathrm{d}\nu}{\nu^4}, \qquad (10)$$

where the n is refractive index of the medium, $\int$ is the spectrum overlap of the donor fluorescence $f_{D(u)}$ and the acceptor adsorption $\varepsilon_{A(\upsilon)}$, and K^2 is the orientation factor of the dipole pair. For random reorientation of the donor and acceptor transition moments, $K^2 = 2/3$. Measurement of E makes it possible to estimate distances between the centers by 1.0 to 8.0 nm. Lanthanide (chelates of terbium or europium) based fluorescence resonance energy transfer (FRET) is a recent modification of the technique with a number of technical advantages, yet relies on the same fundamental mechanism. The advantages are (1) the value of the critical radius is 5.0 to 7.0 nm; (2) since the donor's emission is unpolarized, the efficiency of energy transfer depends primarily on the distance between donor and acceptor, and not on their relative orientation.

The phenomenon of quenching excited singlet states has been used in the investigation of collisions between fluorescent labels and quenching molecules. The process can be described by the Stern–Volmer equation:

$$\phi_0\phi = 1 + K_\mathrm{q}[Q] = 1 + \tau^* k_q[Q], \qquad (11)$$

where K_q is the Stern–Volmer constant and k_q is the quenching rate constant. The range of the k_q values available for the fluorescence quenching technique is 10^8 to 10^9 $\mathrm{M^{-1}s^{-1}}$. Molecules with polar groups and heavy atoms, many aromatic molecules, and electron acceptors are efficient quenchers, with $k_\mathrm{q} = (3\text{–}6) \times 10^9$ $\mathrm{M^{-1}s^{-1}}$.

The relatively slow diffusion of fluorescent probes and labeled macromolecules is studied by Fluorescent Recovery after Photobleaching (FRAP) techniques, which are based on photobleaching of small portions of the objects under study by a laser pulse, followed by monitoring of fluorescence recovery as a result of diffusion from an adjacent portion of the system.

1.2.5 Triplet, Photochrome, Triplet–photochrome, and Spin-triplet-photochrome Labeling

The excited triplet state has two unpaired electrons and can be considered a specific biradical, triplet label. Therefore, the processes of deactivation of the triplet

state can be accelerated owing to exchange spin–spin interaction with other paramagnetic molecule by means of triplet–triplet (T–T) energy transfer or intercrossing. In particular, the experimental values of k_{TT}, the rate constant of the triplet state deactivation, can be written as follows:

$$k_{TT} = 10^{15} \exp(-26R), \qquad (12)$$

where R (nm) is the distance between the triplet label and a quenching molecule separated by *nonconducting* media (e.g. molecules or groups with saturated chemical bonds) and the result is given in reciprocal seconds. Dynamic exchange interactions upon collisions between exchanging centers in solution are very effective. The long lifetime of a triplet label in the excited state (τ^*_{ph} is in the range 10^{-1}–10^{-6} s) offers possibilities for measuring the rates of such slow dynamic processes as collisions of proteins in solution and diffusion of small molecules in rigid membranes. Nitroxide spin probes, complexes of paramagnetic metals, biological molecules (say, quinones, vitamins, etc.), and O_2 have proved to be effective quenchers of triplet labels.

The phenomenon of photochromism involves light-induced reversible transitions of a chromophore (A) to another form (B); for example, the photoisomerization a fluorescent *trans*-stilbene derivative to the non-fluorescent *cis*-stilbene. The latter includes the rotation of phenyl groups as a necessary step in the transformation. In a viscous medium and in systems with strong steric hindrance, the rotation can be the rate-determining stage; therefore, the experimental rate constant of the photoisomerization k_{is} of the photochromic label can be a characteristic of the microviscosity of an object of interest.

Photochrome labeling allows one to investigate relaxation processes in biological systems using sensitive and widely available fluorescence techniques. An approach was developed to determine the rate constant of very rare collisions, including those between macromolecules in rigid media such as biological membranes. The photochrome–triplet labeling approach is based on monitoring the rate at which the photochromic stilbene label *cis–trans* isomerization is sensitized by the triplet–triplet energy transfer between this label and a triplet label (say, Erythrosine). The relatively long lifetime of the excited triplet state of the triplet label and nonreversible character of the photoisomerization make it possible to *integrate* information on the energy transfer by accumulating the photoreaction product, a *trans*-stilbene derivative. The characteristic time of the photochrome–triplet labeling method is about 1 s. These unique characteristics of the approach allow one to determine the rate constant of very rare collisions, including those between macromolecules in such rigid media as proteins in biological membranes. An additional step in the cascade reaction scheme is the quenching of the sensitizer triplet state with relatively low-concentration radicals (Fig. 3). This technique combines the three types of biophysical probes: stilbene photochrome probe, triplet probe, and stable nitroxide-radical spin probe, which depresses the sensitizer exited triplet state.

1.2.6 Dual Fluorophore-spin Labeling: High-sensitivity Redox Probes, Spin and Nitric Oxide Traps

The quantitative characterization of quenching and redox processes is based upon the use of two molecular subfunctionalities (a fluorescent chromophore and a stable nitroxide radical) tethered together

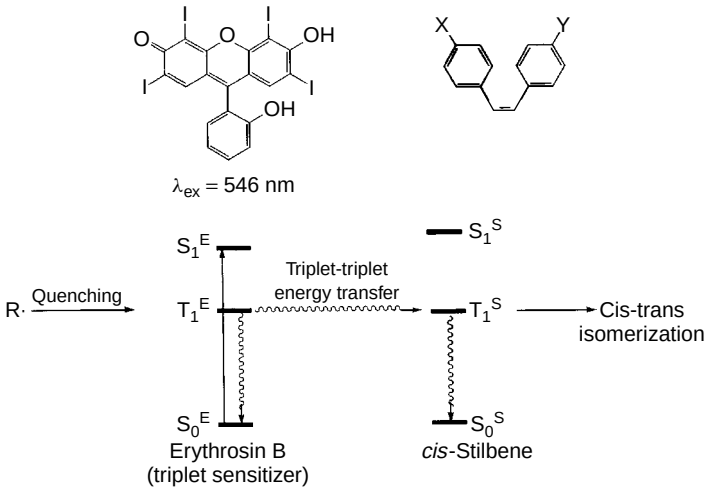

Fig. 3 Representation of the energy levels of the cascade reactants and a competition between the $T_1^E \rightarrow T_1^S$ and $T_1^E \rightarrow S_0^S$ processes.

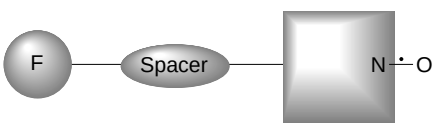

Fig. 4 Schematic representation of dual fluorophore–nitroxide molecule.

by a spacer. A typical double probe consists of fluorophore and nitroxide fragments connected by functional groups (a spacer) (Fig. 4). In the dual molecule, the nitroxide is a strong intramolecular quencher of the fluorescence from the chromophore fragment. Chemical or photoreduction of this fragment to a hydroxylamine derivative, oxidation of the nitroxide fragment, or addition of an active radical yields the fluorescence increase and the parallel decay of the fragment ESR signal. A reducing substrate (e.g. antioxidant analyte, semiquinone and superoxide radicals, nitric oxide in the nanomolar concentration scale) reduces the nitroxide function resulting in a decay of the nitroxide ESR signal plus concomitant enhancement of the chromophore fluorescence. Organic synthetic chemistry allows playing with the chemical structure of dual molecules of different fluorescence, ESR, and redox properties with a variety of bridges (spacers) tethering the fluorophore and nitroxide fragments. It was demonstrated that variation of the chemical structure of nitroxide moieties allows measuring of the concentration of nitric oxide, superoxide, and antioxidants (ascorbic and uric acids, flavons) in biological liquids (human blood plasma, fruits juices), that is to say, establishing the reducing status of these objects.

1.2.7 New Fluorescence Technique

In confocal microscopy, a tiny open volume element of about 0.2 fL is created by a focused laser beam. When a fluorescent molecule enters the confocal volume, it becomes excited. In this technique, the space distribution of a fluorescent fluorophore and the fluctuation of fluorescence intensity

are analyzed. Recent developments in fluorescence spectroscopy and microscopy have made it possible to detect and image single molecules. These techniques allow the conducting of spectroscopic measurements for studying chemical and biological species and their interaction with the environment. Single molecular measurement offers time resolution to monitor dynamic processes such as translation, orientation, and enzymatic reactions on a timescale from milliseconds to ten seconds. Fluorescent quantum efficiencies greater than 0.1 can be studied with this technique. The techniques that have evolved to the level of single molecule sensitivity at room temperature are flow cytometry, confocal fluorescence correlation spectroscopy, and microdroplet technique.

Two-photon molecular excitation is performed by very high local intensity provided by tight focusing in a laser-scanning microscopy. This technique is combined with the temporal concentration of femtosecond pulsed lasers that produce a stream of pulses with a pulse duration of about 100 fs at a repetition rate of about 80 MHz. Advantages of the two-photon laser spectroscopy are as follows: high resolution, tolerance of infrared light by biological objects, different selection rule, and vibronic coupling. In home dual-color cross-correlation fluorescence spectroscopy experiments, a sample containing two fluorophores, with different emissions in each molecule, is irradiated with two lasers (or with one laser) to perform simultaneous excitation of the fluorophores. This technique in combination with confocal laser microscopy allows the separation of a microscopic volume with two different fluorophores from a volume with only one fluorophore, and therefore, the monitoring of dissociation of the dual-labeled molecules or association of two single-labeled molecules.

With the development of the near-infrared (NIR) laser diodes, the synthesis of fluorescent dyes with excitation and emission in the NIR wavelength regime has accelerated in the past decade for microscopy application. Fluorescence lifetime imaging is one of the techniques that attract much attention. Fluorescence lifetimes possess the benefit of being independent of local intensity, concentration and photobleaching of the fluorophore, and scattering in complex structure. This

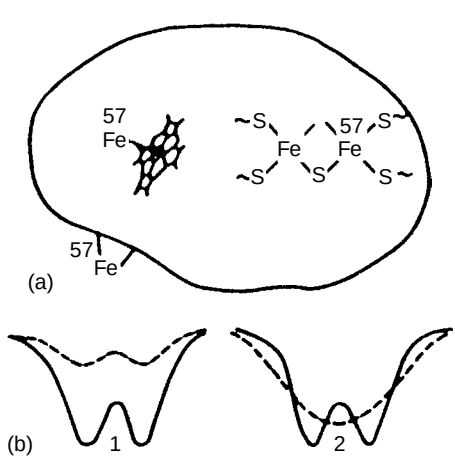

Fig. 5 The Mössbauer labeling methods. (a) Possible methods of introducing labels (^{57}Fe) into proteins. (b) Expected changes in Mössbauer spectral parameters for labels mobilities with $\tau_c = 10^{-7}$ s and amplitude of 0.01–0.02 nm: solid lines, absence of such a mobility; dashed lines, show presence. Curve 1, vibrational motion (f′decreases; line width does not change); curve 2, diffusional motion (line width increases, f′does not change).

technique in combination with FRET allows the investigation of interactions of proteins and nucleic acids within a cell.

1.3 Miscellaneous Labeling Methods

1.3.1 Mössbauer Labels

The Mössbauer method is based on the phenomenon of nuclear gamma resonance (NGR), for example, the transition of a nucleus (e.g. ^{57}Fe) from the ground state to the excited state upon absorption of a γ-quantum emitted by ^{57}Co. The dependence of the intensity the emitting and absorbing nuclei is indicated in the so-called *Mössbauer spectrum* (Fig. 5). Parameters of the NGR spectra are extremely sensitive to the amplitude x of the Mössbauer atom oscillation and the diffusion coefficient D, and therefore, to the dynamics of an object of interest. For example, for ^{57}Fe:

$$f' = 10^{-2300\langle x\rangle^2} \qquad (13)$$

and when the amplitude increases by 0.05 nm, the f' value decreases by a factor of 10^5. Mössbauer atoms can be introduced into various portions of biological structures, say active centers of Fe-containing enzymes, via chemical reactions, absorption, or biosyntheses. In a complimentary technique of the Rayleigh scattering of Mössbauer radiation, the radiation is incident upon the sample. The sample does not need to be labeled with ^{57}Fe; nevertheless, the analysis of the radiation gives the mean square displacement of the scattering atoms averaging over all the atoms of the sample.

1.3.2 Nuclear Magnetic Resonance Probes

Probes modified specifically with isotopes (^{2}H, ^{13}C, ^{15}N, ^{19}F, ^{17}O, ^{28}Na, etc.) are characterized by distinct and easily interpreted NMR spectra. Such modifications can be exemplified by the introduction of ^{13}C into the C=O group of lysine, ^{15}N into the amide group of glycine, and ^{2}H into specific positions of lipid probes. The most widely used NMR probes that are commonly combined with spin labeling are deuterated probes. Isotopes ^{13}C and ^{15}N are the most suitable for studying the molecular dynamics of biological molecules.

1.3.3 Electron and X-ray Scattering Labels

The relative intensity of electron scattering and X-ray scattering (I/I_0) increases, other conditions being equal, with atomic number (Z) of the component atom: $I/I_0 \sim Z^2$. A number of polynuclear complexes and clusters of heavy metals are used nowadays as electron-scattering labels: colloid particles and clusters of gold, polymercury mercarbides, and compounds of palladium and iodine. Analysis of electron micrographs and X-ray scattering of the labeled biological objects allows one to investigate the location of specifically labeled portions.

2 Applications

2.1 Enzymes and Proteins

2.1.1 Active Sites of Enzymes

Monitoring the mobility of spin and luminescence labels and measuring the magnetic and electrostatic interactions between them can be used to study many phenomena, including the chemical nature and mutual disposition of modified catalytic and binding chemical groups, the size and form of a cleft of enzyme-active

centers, the distribution of electrostatic charges around the centers, the location of specific zones on an enzyme complex, and the rotation of protein globules.

Serine proteases are used as a *proving ground* for various versions of the spin-labeling method. The principal catalytic groups of the α-chymotrypsin active center, namely, Ser195, His57, and Met192, were modified by nitroxide spin labels. The ESR spectra of spin-labeled preparations were highly sensitive to fine changes in the structure of the active centers at the interactions with the substrate analogs. When nitroxide derivatives of sulfamides having bridges of varying lengths between the aromatic groups and nitroxide fragment were included in the enzyme-active center, the mobility of such a fragment could be drastically enhanced by using a probe longer than 1.45 nm. Physical labeling methods were applied to determine the mutual arrangement of Fe_nS_n clusters, the ATP center of nitrogenase and the location of these centers on the enzyme macromolecules, and for estimation of distances between the nucleotide-binding centers of H^+ ATP synthase (4.0–5.0 nm).

The labeling method was used to probe the molecular nature of the binding pocket of a G protein–coupled receptor and the events immediately following the binding and activation. The substance P peptide, a potent agonist for the neurokinin-1 receptor, has been modified with a nitroxide spin probe specifically attached at Lys-3. The EPR spectrum of the bound peptide indicates that the Lys-3 portion of the agonist is highly flexible. A slight increase in the mobility of the bound peptide in the presence of a non-hydrolyzable analog of GTP was detected, indicative of the alternate conformational states described for this class of receptor. *S*-adenosylmethionine (AdoMet) synthetase (I) of *Escherichia coli* is one of numerous enzymes that have a flexible polypeptide loop that moves from gate access to the active site in a motion that is closely coupled to catalysis. Site-directed spin labeling was used to introduce nitroxide groups at 2 positions in the loop to illuminate how the motion of the loop is affected by substrate binding. The results suggest that the motion of the loop may be an intrinsic property of the protein and not be strictly ligand-modulated.

The confocal fluorescence coincidence analysis has been employed for a rapid homogeneous assay for restriction endonuclease *Eco*RI. This methodology has been improved by the application of two-photon excited, dual-color, cross-correlation spectroscopy on the level of single diffusing molecules. A double-strand of DNA was labeled with Rhodamin green and Texas red. The kinetics of the enzymatic cleavage of the labeled DNA by restriction endonuclease was monitored by this new technique. The single-molecule fluorescence technique was used for the study of differences in the chemical reactivity of individual molecules of enzymes. Enzyme molecules are presented at very low concentration (7.6×10^{-17} to 1.5×10^{-16} M) in a narrow capillary and each discrete molecule produces a discrete zone. The kinetics of synthesis of a fluorescent product, 2′-(2-benzothiazol)-6′-hydrobenzthiazol, produced by single alkaline phosphatase molecule has been investigated.

Rhodanase was labeled at its catalytic site with the phosphorescence probe eosin isothiocyanide. The accessibility of molecules to the probe was determined by phosphorescence lifetime-quenching studies. The use of a novel pulsed ESR technique for distance measurement based on the detection of double-quantum

coherence (DQC), which yields high quality dipolar spectra, significantly extended the range of measurable distances in proteins using nitroxide spin-labels. Eight T4 lysozyme (T4L) mutants, doubly labeled with methanethiosulfonate spin-label (MTSSL), have been studied using DQC-ESR at 9 and 17 GHz. The distances span the range from 2.0 to 4.7 nm.

The physical basis of an approach rests upon the effect of the local electrostatic potential upon dynamic interactions at encounters with charged quenching molecules resulting in fluorescence (phosphorescence), or between a stable radical, for example, nitroxide, and another charged paramagnetic species. In such cases, the relaxation parameters, that is, the lifetime of the fluorescence (phosphorescence) chromophore or spin–spin and spin–lattice relaxation rates of paramagnetic species are dependent upon the frequency of encounters, and therefore, on local electrostatic fields. Electron-carrier horse heat cytochrome c and dioxygen-carrier sperm-whale myoglobin served as models for determining local electrostatic charges in the vicinity of paramagnetic active sites of metalloenzymes and metalloproteins. Calculations of local charge Z_x in the vicinity of a paramagnetic particle (such as the active site of a metalloprotein or a spin label) colliding with a nitroxide or metallocomplex with known charge Z_p can be carried out with the use of the Debye equation.

Nitroxide spin molecules are convenient probes for the ENDOR application for solving some problems of enzyme catalysis. From ENDOR studies of molecular structures and conformations, several spin-labeled amino acid derivatives incorporated into enzyme-active sites have been reported. For instance, the spin-labeled transition-state analog in the α-chymotripsin reaction, *N*-(2,2,5,5-tetramethyl-1-oxypyrrolinyl)-L-phenylalaninal, has been synthesized. The conformation of the acyl- moiety of the substrate analog in the active site of the reaction intermediate differs significantly from that of the free substrate in solution.

2.1.2 Conformational Changes and Molecular Dynamics

Labeling methods were applied for detecting conformational changes, including large-scale denaturation transitions, predenaturing phenomena, allosteric effects (transmitted from one enzyme subunit to another), conformational changes in enzyme structures other than quaternary (transglobular allosteric transition), and changes in the state of various segments of the active centers of enzymes due to the action of specific reagents on the adjacent segments.

Allosteric effects were first reported by McConnell and associates in studies of horse hemoglobin with the β-SH group modified by a spin label 1.3 to 1.4 nm from the heme active center. Transglobular transitions were studied in a number of proteins, including lysozyme, myoglobin, aspartate aminotransferase, and myosin.

Biophysical labeling methods provide a unique possibility for monitoring local intermolecular dynamics properties in a wide range of correlation times (τ_c is in the range 10^2 to 10^{-10} s). They enable the investigation of the dynamics of different parts of protein globules in the vicinity of the spin, fluorescence, phosphorescence, and Mössbauer labels under various conditions (temperatures 30–330 K, water content, viscosity, substrates, inhibitor additions, etc.). The experiments revealed the following tendency: appropriate physical methods detect the mobility of labels starting from the temperature of liquid

nitrogen; the lower the value of the characteristic frequency of the method, the lower the temperature at which the label mobility can be recorded. Thus, the mobility recorded results from a gradual softening of the protein–water matrix rather than from an individual phase transition.

At physiological ambient conditions, all labeling methods indicate the mobility of labels and surrounding media in the nanosecond temporal regions. A correlation between the temperature dependencies of nonharmonic parameters of Mössbauer atoms in heme, the heme spectra Soret band Gaussian broadening, non-elastic neutron scattering, and the molecular dynamics simulation in myoglobin was demonstrated. According to these data, the anharmonic intramolecular mobility of proteins increases dramatically at $T > 200$ K. This conclusion was confirmed by low-temperature experiments with the use of the fluorescence dynamic Stokes shift, neutron scattering, and molecular dynamics simulation in other proteins, superoxide dismutase, lysozyme, elastase, bacteriorhodopsin, and RNAse. It is necessary to stress that the anharmonic nanosecond dynamics of the proteins recorded with labeling methods only appear if the water content of the sample being studied exceeds a critical value by about 10 to 25%. According to experimental data on Mössbauer spectroscopy, at ambient temperatures, the myoglobin heme group exhibits the unharmonic nanosecond motion with $\langle \Delta x \rangle_{nh} > 0.02$ nm. The flexibility of the cavity of the myoglobin active site is evidenced by the mobility of a spin probe, a derivative of isocyanate attached to the heme group in the single crystal.

The mobility of a single deuterated tryptophane located in a loop of the active site of triosephosphate isomerase has been investigated by solid-state deuterium NMR and solution state 19F NMR. The rate of the loop's opening and closing was detected using samples of the enzyme in the presence and in the absence of a substrate analog DL-glycerol 3-phosphate at temperatures ranging from -15 to $+45\,^{\circ}$C. It was shown that the rate of the loop's opening and closing is of order $10^4\ s^{-1}$. Hinge-bending and substrate-induced conformational transitions in T4 lysozyme in solution were confirmed in a study by site-directed labeling. Both single and pairs of nitroxide spin labels were introduced into different domains of the protein followed by monitoring distances between the labels by ESR technique. In the absence of a substrate, the results are consistent with a hinge-bending motion, which opens the active site cleft. When substrate binding takes place, the relative domain movement occurs.

Simulation studies of typical intramolecular energy transfer experiments reveal that both static and dynamic conformational distribution information can thus be obtained at a single temperature and viscosity. This method was used for the investigation of the refolding transition of *E. coli* adenylate kinase (AK) by monitoring the refolding kinetics of a selected 20-residue helical segment in the CORE domain of the protein.

The local and global dynamics of the *Sulfolobus solfataricus* β-glycosidase were studied by ESR and time-resolved fluorescence techniques. For EPR investigations, the protein was covalently modified by the maleimido nitroxide spin label, which is specific for cysteine -SH groups, at position 344 and at position 101, where Ser101 was changed into a cysteine by site-directed mutagenesis. The labeled proteins underwent temperature perturbation in the range 290 to 335 K. The general dynamic

information was deduced from the analysis of the fluorescence emission decay of the tryptophanyl residues that are present in each region of the protein structure.

2.2 Membranes

Specific modification of individual lipid and protein parts of biological and model membranes by whole sets of labels and probes allows the biophysical labeling method to be extremely effective in investigating membrane dynamics and microstructure. Monitoring of the rotational diffusion of nitroxide, fluorescent, and phosphorescent derivatives of lipids and proteins, and their collisions, enables one to determine thickness, the microviscosity profile of membranes, membrane heterogeneity, and phase transitions, as well as to discover phenomena related to the lateral diffusion of lipid and protein associations in membranes. Spin and fluorescence labeling approaches have indicated significant similarities between the structural and dynamic properties of model bilayer liposomes and the lipid phase of such biological membranes such as erythrocytes, the sarcoplasic reticulum, and microsomes. A number of specific problems associated with the intrinsic inhomogeneity of cells, biochemical and biophysical processes (including redox ones), and cell organization can be readily solved by means of labeling approaches. In differentiated cells (ovarian cells, spermatozoa), the FRAP method indicated the presence of protein domains and of different microcompartments in various parts of the cells that differed with respect to diffusion coefficients and mobile fractions. The values of intercellular pH monitored by a fluorescent probe (fluorescein after enzymatic deacetylation) were found to differ in various compartments of the cells.

The distribution of LW peptide between coexisting ordered and disordered lipid domains was probed by measuring the LW Trp fluorescence quenched by a nitroxide-labeled phospholipid that is concentrated in disordered lipid domains. Strong quenching of the Trp fluorescence (relative to quenching in model membranes lacking domains) showed that LW peptide was concentrated in quencher-rich disordered domains and was largely excluded from ordered domains. Nitroxide spin label studies with high-field/high-frequency ESR and two-dimensional Fourier transform ESR enable one to accurately detect distances in biomolecules, unravel the details of the complex dynamics in proteins, characterize the dynamic structure of membrane domains, and discriminate between bulk lipids and boundary lipids that coat transmembrane peptides or proteins.

The EPR spectroscopic characterization of a recently developed magnetically oriented spin-labeled model membrane system was reported. The oriented membrane system is composed of a mixture of a bilayer-forming phospholipid and a short-chain phospholipid that breaks up the extended bilayers into bilayered micelles or bicelles that are highly hydrated (approx. 75% aq.). Paramagnetic lanthanide ions (Tm^{3+}) were added to align the bicelles such that the bilayer normal is collinear with the direction of the static magnetic field. The nitroxide spin probe 3β-doxyl-5α-cholestane (cholestane) was used to demonstrate the effects of macroscopic bilayer alignment through the measurement of orientational dependent hyperfine splittings that were close to A_{yy}. The EPR signals of cholestane inserted into oriented and randomly dispersed DMPC-rich

bilayers have been investigated over the temperature range of 298 to 348 K.

2D-ELDOR spectroscopy has been employed to study the dynamic structure of the liquid-ordered (L_o) phase versus that of the liquid-crystal (L_c) phase in multibilayer phospholipid vesicles without (L_c) and with (L_o) cholesterol, using end-chain and headgroup labels and spin-labeled cholestane. Fluorescence correlation spectroscopy was used for investigation of the size and the size distribution of liposomes loaded with a fluorescence probe.

2.3 Nucleic Acids

The following methods have been developed for modification of nucleic acids: (1) covalent attachment of spin and fluorescent labels to functional groups of the nucleotide bases; (2) covalent attachment of labels to hydroxyl groups on the ribose glucosyl rings; (3) alkylation or acylation of the SH group of the 4-thiouracil or the NH_2 group on the aminoacyl residue of tRNA; (4) enzymatic synthesis of nucleic acids with the use of a labeled base; and (5) intercalation of spin and fluorescent probes between bases in the double-helix portion of nucleic acids, and complexation with paramagnetic (Mn^{+2}) or Mössbauer (^{57}Fe) atoms. The extent of the immobilization of the labels and probes and the spin–spin interactions between them are determined by the structure of the nucleic acid, the length and flexibility of the label, and ambient conditions (pH, temperature, ionic strength, destructive forces, and degrees of binding of various compounds). The thermal- and pH-induced transitions of nucleic acids at their melting temperatures are accompanied by a drastic increase in the mobility of the attached spin labels. The ESR spectra of spin-labeled preparations of nucleic acids are sensitive to fine conformational changes as well.

The dynamics of iron nuclei in the condensates obtained by interaction of Fe^{III} with DNA, (DNA monomer), has been investigated by ^{57}Fe Mössbauer spectroscopy. Functions of the parameters isomer shift and nuclear quadrupole splitting in temperature ranges 20 to 260 K were employed to investigate the dynamics of Fe nuclei and showed linear trends in the temperature ranges 20 to 150 and 150 to 260 K, respectively, the latter with larger slopes. Site-directed spin labeling utilizes site-specific attachment of a stable nitroxide radical to probe the structure and dynamics of nucleic acids. 4-thiouridine base was introduced at each of six different positions in a 23-nucleotide RNA molecule. The 4-thiouridine derivatives were subsequently modified with one of three methanethiosulfonate nitroxide reagents to introduce a spin label at specific sites. The motion is similar to that found for a structurally related probe on helical sites in proteins, suggesting a similar mode of motion. The nitroxide spin-labels are covalently linked to a deoxyuridine residue using either a monoacetylene or diacetylene tether. A detailed analysis of the EPR spectrum of duplex DNA in solution, spin-labeled using the diacetylene tether, demonstrates that the motion of the nitroxide can be modeled in terms of this independent uniaxial rotation together with motion of the DNA, which is consistent with the global tumbling of the duplex.

New genome sequence information is rapidly increasing the number of nucleic acid (NA) targets of use for characterizing and treating diseases. Detection of these targets by fluorescence-based assays is often limited by fluorescence background from unincorporated or unbound probes

that are present in large excess over the target. To solve this problem, energy transfer–based probes have been developed and used to reduce the fluorescence from unbound probes. The quenching approach of a two-color NA assay with a correlated, two-color, single-molecule fluorescence detection was employed. A method for generating fluorescent probes from RNA or DNA samples was described. In this method, an amine-modified base was added to the ends of the primers so that a dye could be introduced into each DNA and onto the end of each product. This allows labeling of a very small amount of total RNA in such a way that reproducible experiments were performed. It was shown that FRET between fluorescein labeled to poly(C) and an intercalator agent takes place when single-stranded poly(C) hybridizes with poly(I).

The following application of the labeling approach can be also pointed out: (1) protein-induced DNA blending studied by labeling a 30-base pair strand of DNA, one with europium donor chelate and the other with the acceptor dye Cy5; (2) detection of the association between actin and the protein dystropin inside a muscle cell; (3) time-resolved identification of individual label mononucleotides in water; (4) DNA sequences on the single molecule level; (5) detection of individual tumor marker molecules in neat human.

2.4 Cells and Organs

Fluorescence lifetime imaging microscopy using multiphoton excitation techniques is now finding an important place in the imaging of protein–protein interactions and intracellular physiology. An example of a typical application of the system was provided in which the fluorescence resonance energy transfer between a donor–acceptor pair of fluorescent proteins within a cellular specimen is measured. The recent development of automated fluorescence imaging systems has enabled fluorescence microscopy to be used for the purposes of compounds screening. This technique has found various applications, including screening for the effect of kinase inhibitors on the cytoskeleton, agonist-stimulated receptor internalization, and protein phosphorylation and acetylation.

L-band (low-frequency) EPR imaging instrumentation has been developed in order to elucidate spatially defined differences in tissue metabolism and the redox status. For example, 3D images of nitroxyl compounds in a rat tail and isolated rat heart have been reported, which reflect the local metabolic and oxygenation status. Very recently, a new magnetic resonance imaging method (MRI) was presented that can "visualize" free radical generation in animals by means of MRI. The advantages of MRI, the combination of MRI and EPR, and PEDRI (proton electron double resonance imaging), and recent progress of these methods were demonstrated.

The kinetics of reductions of the radical R*, 5-dimethylaminonaphthalene-1-sulfonyl-4-amino-2,2,6,6-tetramethyl-1-piperidine-oxyl, by blood and its components were studied using the EPR technique. The results demonstrate that: (1) the erythrocytes catalyze the redaction of R* by ascorbate; (2) the rate of redaction of the radical is high though it does not penetrate the cells; and (3) in human erythrocytes, there is an efficient electron transfer route through the cell membrane. The study points out that R* is a suitable spin label for

measuring the reduction kinetics and antioxidant capacity in blood as expressed by reduction by ascorbate. The nitroxides are reduced to corresponding hydroxylamine in cellular incubations as well as *in vivo* incubations. *Tissue hypoxia* is known to significantly enhance the rate of the nitroxide reduction.

Fluorescence *in situ* hybridization (FISH)-based techniques were used in the clinical detection of genetic alterations in tumor cells. Fluorescent DNA probes also enabled screening for very subtle chromosomal changes. A growing number of FISH-based cytogenetic tests are employed in clinical laboratories to support a physician's diagnoses of the causes and the course of a disease. FISH-based analyses have been applied very successfully to the analysis of single cells and have demonstrated the existence of cell clones of different chromosomal make-up within human tumors. An immunofluorescence assay was developed to identify proteins specifically binding to oligonucleoside phosphorodithioate (ODN) aptamers from a bead-bound ODN library. It was shown that high-resolution localization of genes on chromosomes is possible with specially correlated introduction of fluorescent DNA probes (fluorescence *in situ* hybridization). This method allows the mapping of cloned genes to a locus defined by a chromosome band.

2.5 Miscellaneous

2.5.1 Biologically Active Compounds, Biological Analysis, Pharmokinetics

Syntheses and applications have been reported for many types of spin- and fluorescent-labeled analogs of amino acids, steroids, phospho-organic compounds, alkaloids and terpenes, nucleotides and their fragments, coenzymes (vitamins B_{12} and B_6, protohemin, NADH, etc.), and haptens. Advances in the synthetic chemistry of nitroxide analogs of medical preparations such as morphine, local anesthetics, anticancer drugs (analogs of thiophosphoamin, rubromycin, adriamycin, nitroarylaziridnes, nitrosourea, diethyleneimineurethanic acids, and anthracycline drugs) have provided the basis for further development in basic and applied biochemical and pharmacokinetic investigations.

Some of the most promising applications of fluorescence and spin labeling appear to entail the analysis of biologically active compounds, metal ions, biopolymers, and enzymes. Changes in fluorescence, rotational diffusion parameters, dipole–dipole electrostatic and spin–spin interactions under contact between physical probes, and the compound being tested allow one to work out an appropriate analytical procedure in each particular case. A number of procedures featuring fluorescent and spin-labeled chelating reagents have been developed for quantitative assays of Ca^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Co^{2+}, and Na^{+} ions, and for determination of enzymatic activity. The ability of chromophores (propidium, acridine orange, etc.) to change the intensity of fluorescence upon binding with nucleic acids has been widely used in analytical biochemistry. Lanthanide fluorescence chelates have been successfully developed as the fluorescence labels for highly sensitive bioassays. The applications of lanthanide fluorescence labels for time-resolved fluoroimmunoassay (TR-FIA), DNA hybridization assay, cell activity assay, and fluorescence imaging microscopy have been widely investigated.

Nitroxide probes are used as *in vivo* imaging reagent and oximetry probes. The

pharmacokinetics and spatial distribution of the nitroxide in tumor tissue were followed and compared with those in normal tissue. The tumor tissues showed significant heterogeneity in the nitroxide distribution and higher reducing rate compared to the normal tissue.

3
New Trends in Biophysical Labeling

The following main trends in development and application biophysical labeling may be pointed out:

- Synthesis of fluorescent and phosphorescent dyes with excitation and emission in the NIR wavelength regime for microscopy and analysis application.
- Syntheses and application of dual fluorophore–nitroxide probes of different redox potential and fluorophore fragment nature.
- Site-directed incorporation of spin and luminescent label in protein, nuclear acids followed by the investigation with the use modern physical methods.
- Growing employment of recently developed ESR techniques (high-frequency ESR spectroscopy, Fourier transform 2D ESR, ELDOR electron spin-echo techniques, ELDOR and ENDOR, and multiple-quantum ESR) and fluorescence methods such as confocal microscopy, single molecular measurement, multiphoton molecular excitation, fluorescence lifetimes imaging, cascade systems, etc.
- Expanding objects under investigation including cells, organs, and organisms in prospect.
- Wide application in biomedical research and medical diagnostics.

Bibliography

Books and Reviews

Berliner, L. (Ed.) (1998) *Spin Labeling. The Next Millennium*, Academic Press, New York, p. 14.

Berliner, L., Eaton, S., Eaton, G. (Eds.) (2000) *Distance Measurements in Biological Systems by ESR*, Kluwer Academic Publishers, Dordrecht.

Blake, R.A. (2001) Cellular screening assays using fluorescence microscopy, *Curr. Opin. Pharmacol.* **1**, 533–539.

Blank, A., Dunnam, C.R., Borbat, P.P., Freed, J.H. (2003) High resolution electron spin resonance microscopy, *J. Magn. Reson.* **165**, 116–127.

Borbat, P.P., Costa-Filho, A.J., Earle, K.A., Moscicki, J.K., Freed, J.H. (2001) Electron spin resonance in studies of membranes and proteins, *Science* **291**, 266–269.

Kocherginsky, N., Swartz, H.M. *Nitroxide Spin Labels*, CRC Press, Boca Raton, FL, 1995, pp. 114–172.

Krinichnyi, V.I. (1994) *2-mm Wave Band ESR Spectroscopy of Condensed Systems*, CRC Press, Boca Raton, FL.

Krishnan, R.V., Masuda, A., Centonze, V.E., Herman, B. (2003) Quantitative imaging of protein-protein interactions by multiphoton fluorescence lifetime imaging microscopy using a streak camera, *J. Biomed. Opt.* **8**, 362–367.

Lakowicz, J.R. (1983) *Principles of Fluorescence Spectroscopy*, 2nd edition, Plenum Press, New York.

Lakowicz, J.R., Gryczynski, I., Shen, Y., Malicka, J., D'Auria, S., Gryczynski, Z. (2002) Fluorescence spectral engineering – biophysical and biomedical applications, *Springer Ser. Fluoresc.* **2**, 43–68.

Likhtenshtein, G.I. (1976) *Spin Labeling Method in Molecular Biology*, Wiley-Interscience, New York.

Likhtenshtein, G.I. (1993) *Biophysical Labeling Methods in Molecular Biology*, Cambridge University Press, Cambridge, New York.

Likhtenshtein, G.I. (1996) Spin and fluorescence immunoassay in solution, in: Lefkovits, I., Nezlin, R. (Eds.) *Immunology Methods Manual*, Pergamon Press, London, pp. 540–550.

Likhtenshtein, G.I. (2003) *New Trends in Enzyme Catalysis and Mimicking Chemical Reactions*,

Kluwer Academic Publishers, Dordrecht, New York.

Likhtenshtein, G.I., Febrario, F., Nucci, R. (2000) Intramolecular dynamics and conformational transitions in proteins studied by biophysical labeling methods. Common and specific features of proteins from thermophylic microorganisms, *Spectrochem. Acta Part A: Biomol. Spectrosc.* **56**, 2011–2031.

Papper, V., Likhtenshtein, G.I. (2001) Substituted stilbenes: A new view on well-known systems: new application in chemistry and biophysics, *J. Photochem. Photobiol., A: Chem.* **140**, 39–52.

Parak, F., Reinish, L. (1986) Mössbauer effect in the study of structure dynamics, *Methods Enzymol.* **131**, 568–607.

Ratner, V., Kahana, E., Eichler, M., Haas, E. (2002) A general strategy for site-specific double labeling of globular proteins for kinetic FRET studies, *Bioconjugate Chem.* **13**, 1163–1170.

Rettig, W., Stremel, B., Schrader, S., Seifer, H. (Eds.) (1999) *Applied Fluorescence in Chemistry, Biology and Medicine*, Springer-Verlag, Berlin.

Selvin, P.R. (1999) Luminescent lanthanide chelates for improved resonance energy transfer and application to biology, in: Rettig, W., Stremel, B., Schrader, S., Seifer, H. (Eds.) *Applied Fluorescence in Chemistry, Biology and Medicine*, Springer-Verlag, Berlin, pp. 457–487.

Swartz, H.M. (2002) Measuring real levels of oxygen in vivo: opportunities and challenges, *Biochem. Soc. Trans.* **30**, 248–252.

Trommer, W.E., Vogel, P.D. (1992) Photoaffinity spin labeling, in: Zhdanov, R.I. (Ed.) *Bioactive Spin Labels*, Springer, Berlin, pp. 405–427.

Vo-Dinh, T. (2003) life time-based imaging, in: Herman, P., Lin, H.-J., Lakowicz, J.R. (Eds.) *Biomedical Photonics Handbook 9/1–9/30*, CRC Press LLC, Boca Raton, FL.

Weier, H.-U.G., Greulich-Bode, K.M., Ito, Y., Lersch, R.A., Fung, J. (2002) FISH in cancer diagnosis and prognostication: From cause to course of disease, *Expert Rev. Mol. Diagn.* **2**, 109–119.

Xiang, C.C., Brownstein, M.J. (2003) Preparing fluorescent probes for microarray studies, *Methods Mol. Biol.* **224**, 55–60.

Yuan, J., Wang, G., Matsumoto, K. (2001) Lanthanide fluorescence chelates and their applications in bioassays, *Trends Inorg. Chem.* **7**, 109–117.

Primary Literature

Boehmer, M., Enderlein, J. (2003) Fluorescence spectroscopy of single molecules under ambient conditions: methodology and technology, *Chem. Phys. Chem.* **4**, 792–808.

Borbat, P.P., Mchaourab, H.S., Freed, J.H. (2002) Protein structure determination using long-distance constraints from double-quantum coherence ESR: study of T4 lysozyme, *J. Am. Chem. Soc.* **124**, 5304–5314.

Costa-Filho, A.J., Shimoyama, Y., Freed, J.H. (2003) A 2D ELDOR study of the liquid ordered phase in multilamellar vesicle membranes, *Biophys. J.* **84**, 2619–2633.

Fujii, H., Berliner, L.J. (2003) In vivo detection and visualization of reactive oxygen free radicals by EPR and MRI, *Biomed. Res. Trace Elem.* **14**, 6–10.

Hustedt, E.J., Kirchner, J.J., Spaltenstein, A., Hopkins, P.B., Robinson, B.H. (1995) Monitoring DNA dynamics using spin-labels with different Independent mobilities, *Biochemistry* **34**, 4369–4375.

Krupyanskii, Y.u.F., Esin, S.V., Eshenko, G.V., Mikhailyuk, M.G. (2002) Equilibrium fluctuations in lysozyme and myoglobin, *Hyperfine Interact.* **141/142**, 273–277.

Mangels, M.L., Cardon, T.B., Harper, A.C., Howard, K.P., Lorigan, G.A. (2000) Spectroscopic characterization of spin-labeled magnetically oriented phospholipid bilayers by EPR spectroscopy, *J. Am. Chem. Soc.* **122**, 7052–7058.

McNulty, J.C., Thompson, D.A., Carrasco, M.R., Millhauser, G.L. (2002) Dap-SL: a new site-directed nitroxide spin labeling approach for determining structure and motions in synthesized peptides and proteins, *FEBS Lett.* **529**, 243–248.

Nakazumi, H., Colyer, C.L., Kaihara, K., Yagi, S., Hyodo, H. (2003) Red luminescent squarylium dyes for noncovalent HSA labeling, *Chem. Lett.* **32**, 804,805.

Nolan, R.L., Cai, H., Nolan, J.P., Goodwin, P.M. (2003) A simple quenching method for fluorescence background reduction and its application to the direct, quantitative detection of specific mRNA, *Anal. Chem.* **75**, 6236–6243.

Qin, P.Z., Hideg, K., Feigon, J., Hubbell, W.L. (2000) Monitoring RNA base structure and

dynamics using site-directed spin labeling, *Biochemistry* **42**, 6772–6783.

Saphier, O., Silberstein, T., Shames, A.I., Likhtenshtein, G.I., Maimon, E., Mankuta, D., Mazor, M., Katz, M., Meyerstein, D., Meyerstein, N. (2003) The reduction of a nitroxide spin label as a probe of human blood antioxidant properties, *Free Radical Res.* **37**, 301–308.

Schunemann, V., Winkler, H. (2000) Structure and dynamics of biomolecules studied by Mossbauer spectroscopy, *Rep. Prog. Phys.* **63**, 263–353.

Shafer, A.M., Bennett, V.J., Kim, P., Voss, J.C. (2003) Probing the binding pocket and endocytosis of a G protein-coupled receptor in live cells reported by a spin-labeled substance P agonist, *J. Biol. Chem.* **278**, 34203–33421.

Talavera, E.M., Bermejo, R., Crovetto, L., Orte, A., Alvarez-Pez, J.M. (2003) Fluorescence energy transfer between fluorescein label and DNA intercalators to detect nucleic acids hybridization in homogeneous media, *Appl. Spectrosc.* **57**, 208–215.

Taylor, J.C., Markham, G.D. (2003) Conformational dynamics of the active site loop of S-adenosylmethionine synthetase illuminated by site-directed spin labeling, *Arch. Biochem. Biophys.* **415**, 164–171.

Trotta, A., Barbieri Paulsen, A., Silvestri, A., Ruisi, G., Assunta Girasolo, M., Barbieri, R. (2002) The dynamics of ^{57}Fe nuclei in FeIII-DNA condensate, *J. Inorg. Biochem.* **88**, 14–18.

Ligase-mediated Gene Detection

Johan Stenberg, Mats Nilsson, and Ulf Landegren
Department of Genetics and Pathology, Uppsala University, Sweden

1 Principles 180
1.1 Basis of Ligase-mediated Analyses 180
1.2 Biology of Ligation 181
1.3 Properties of Ligases 182
1.4 Chemistry of Enzymatic Ligation 182
1.5 Substrate Requirements in Enzymatic Ligation Reactions 184
1.6 Ribozyme Ligases 184
1.7 Chemical Ligation 184

2 Techniques 186
2.1 Oligonucleotide Joining Reactions for Detection of Nucleic Acids 186
2.2 Padlock Probes 188
2.3 Proximity Ligation 190
2.4 Other Uses of Ligation Reactions 190

3 Outlook 190

Bibliography 191
Books and Reviews 191
Primary Literature 191

Keywords

Ligase
An enzyme capable of joining the 5′ and 3′ ends, respectively, of two adjoining nucleic acid molecules, creating a phosphodiester bond.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

5′ and 3′
The two ends of a nucleic acid strand are termed the 5′ and the 3′ ends. Enzymatic synthesis proceeds in the 5′ to 3′ direction.

Ligation template
A nucleic acid molecule with a sequence that allows two probe molecules to hybridize adjacent to each other, allowing ligation of the 5′ end of one to the 3′ end of the other to take place.

DNA ligation is a popular mechanism in analyses of gene sequences. The ligation of pairs of DNA probes upon hybridization to a target nucleic acid sequence creates a contiguous DNA sequence that can be used to infer the presence and nature of target molecules. The ligatable DNA strands can also be coupled to other molecules interacting with targets of interest to increase the range of molecules detectable in ligase-based assays.

In ligase-mediated detection of nucleic acid sequences, the requirements for substrate recognition by ligases may be used to distinguish target sequences that differ in even single nucleotide positions. The covalent link that forms between the probe pairs can be used to connect retrievable functions on one probe to detectable groups on the other. Even more usefully, the resulting contiguous DNA strand can be amplified, along with sequence elements that encode the target specificity of the probes. These principles are being applied in a rapidly increasing range of molecular analyses, including high-throughput genotyping, *in situ* detection of nucleic acid sequences, and highly sensitive protein assays.

1 Principles

1.1 Basis of Ligase-mediated Analyses

Specific nucleic acid sequences can be analyzed using hybridization probes that base pair to the target sequence. For enhanced performance, hybridization probes are often used in combination with nucleic acid–specific enzymes in molecular genetic analyses. One example is the polymerase chain reaction (PCR), where two oligonucleotide probes and a DNA polymerase are used to detect a nucleic acid sequence. In ligase-mediated detection reactions, pairs of DNA probes are joined by the action of a ligase. The joining event is used as a measure of the presence and nature of the nucleic acid sequences. This is possible because ligation critically depends on the ability of the probes to base-pair to a template nucleic acid molecule (Fig. 1). This assay mechanism has several valuable properties:

1. The ligation reaction is inherently very specific since it requires the coincident hybridization of two independent probe molecules immediately next to each other on a template. It is therefore

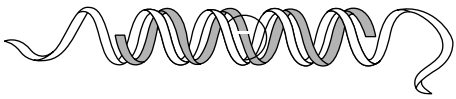

Fig. 1 Oligonucleotide ligation assay. Configuration of two oligonucleotide probes (shaded), hybridizing immediately next to each other on a target nucleic acid strand (white). Where the two probes meet (circled region), they can be connected by a ligase.

unlikely that the probes will join in the absence of the correct template nucleic acid molecule. As a consequence, the assay permits detection of unique nucleic acid sequences in complex populations of DNA molecules, for example, the complete human genome.

2. Variants of sequences can be conveniently distinguished because ligation is strongly inhibited by a mismatch even in a single nucleotide position at the junction of the probes.
3. The act of ligation of probes gives rise to a new DNA sequence not previously present in the reaction. This circumstance can be exploited for detection of the reaction products. For example, detection can be enhanced by exploiting the increased hybridization stability of the ligation product as compared to that of each individual probe. Alternatively, pairs of probes may be used in which one member carries a detectable moiety and the other is bound to a support before or after the ligation assay. Thereby, ligation products can be efficiently isolated and excess probes removed by washes before detection. If the two ligatable probe sequences are located at either end of the same DNA molecule, then the ligation will result in a circularized probe, wound around the target molecule. Finally, ligation products can also be recruited as templates in subsequent nucleic acid-amplification reactions, resulting in highly sensitive detection of target sequences.

Ligation-mediated analysis has undergone a strong development from the characterization of ligase enzymes and the first analytical ligase experiments carried out in the late 1960s and early 1970s, respectively. We will first discuss some fundamental properties of ligation reactions before describing some of the methods that have been developed on the basis of the probe ligation principle.

1.2 Biology of Ligation

Ligases are proteins that serve to join the ends of polynucleotide strands. They use cofactors such as NAD or ATP to supply the energy required to form the phosphodiester bonds. By contrast, the so-called breakage-reunion enzymes, exemplified by the topoisomerases, first break and then reseal DNA strands in a concerted fashion, and do not require an external supply of energy.

In the cell, ligases complete excision-repair reactions in damaged DNA. In addition, they seal interrupted DNA strands during discontinuous DNA replication, and during meiosis they participate in altering the continuity of paternally and maternally inherited DNA sequences through genetic recombination.

The first biochemical characterization of enzymes that join DNA strands appeared in 1967. The isolated ligases soon proved to be of great value as biotechnological tools for the analysis of DNA structure and construction of novel DNA molecules. Using

these enzymes, it famously proved possible to assemble recombinant molecules from separate DNA segments *in vitro*, and then clone, propagate, and express these genes in new organisms.

1.3 Properties of Ligases

Several enzymes are available that can join nucleic acid strands in a template-dependent manner. The DNA ligase most commonly used in laboratory work is derived from bacteriophage T4. T4 DNA ligase is a monomeric protein with a molecular weight of 68 kDa. It uses ATP as a cofactor and can join both DNA and RNA strands, provided that these are immediately juxtaposed by being base-paired to complementary strands, which can be either DNA or RNA. Reactions involving RNA strands proceed much less efficiently, however. Ends of double-stranded DNA molecules can be joined by the T4 DNA ligase, even when these have no single-stranded extensions that can hybridize and thus stabilize the interaction between the two ends.

The *E. coli* ligase has a molecular weight of 74 kDa. Unlike the ligases derived from bacteriophages, archaea, and eukaryotes, all of which use ATP as a cofactor, the *E. coli* ligase requires NAD as a cofactor, as is true for other bacterial ligases.

Ligases have also been isolated from thermophilic organisms such as *Thermus thermophilus*, *Thermus aquaticus*, and *Desulfurolobus ambivalens*. These enzymes operate at high reaction temperatures and survive conditions that denature DNA. At the optimal reaction temperature of these enzymes, oligonucleotides are less likely to interact with mismatched target sequences. Furthermore, the thermostable ligases exhibit no or little tendency to join DNA molecules with base-paired (blunt) ends or with short complementary (sticky) ends. Because of these properties, they offer increased selectivity of detection and convenience in many analytical reactions.

T4 DNA ligase can also be employed to construct nucleic acid strands that branch into two distinct 3′ end sequences by attaching the 5′ phosphate of an oligonucleotide to the 3′ hydroxyl of a nucleotide residue internal to another oligonucleotide. This requires that this second oligonucleotide is constructed such that the internal nucleotide is connected to the downstream sequence by a 2′ to 5′ phosphodiester bond, rather than the regular 3′ to 5′ bond, leaving a free 3′ hydroxyl.

In contrast to the enzymes just mentioned that all require double-stranded substrates, the T4 RNA ligase joins DNA or RNA strands that are not base-paired to a complementary polynucleotide strand. The probability of ligation is increased, however, if the two ends to be ligated are positioned near each other. This enzyme requires ATP as a cofactor.

1.4 Chemistry of Enzymatic Ligation

DNA ligases join DNA strands by using the energy of a pyrophosphate linkage in a nucleotide cofactor to create a new phosphodiester bond, connecting the two strands. This bond is formed between a 5′-phosphate group and a 3′-hydroxyl on two adjacent nucleic acid molecules. The enzyme-catalyzed reaction can be described in four steps (Fig. 2): first, a high-energy enzyme intermediate is formed when an adenylyl group from a cofactor (ATP or NAD) is coupled to the ligase while releasing pyrophosphate

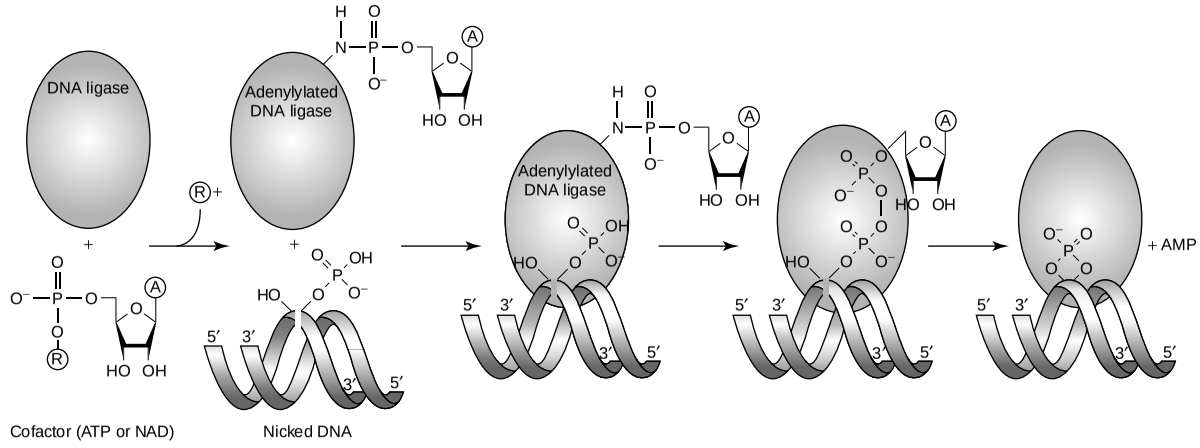

Fig. 2 The chemistry of enzymatic ligation of nucleic acids. The reaction takes place in four reversible reaction steps. In the first step, an adenylyl group from a cofactor, NAD or ATP, is transferred to the ε-amino group of a lysine residue in the DNA ligase enzyme. Thereby the energy required to join two nucleic acid strands is transferred from the cofactor and stored temporarily in the form of a high-energy bond between an adenylyl group and the enzyme. The activated ligase then binds to a nicked site on a DNA duplex. Next, the adenylyl group is transferred to the 5′-phosphate to be connected to the 3′-hydroxyl by ligation. This step recreates a pyrophosphate linkage and preserves the high-energy bond. Finally, the activated 5′ end is attacked by the nearby 3′-hydroxyl group, resulting in the formation of a phosphodiester bond between the 5′ and 3′ ends and releasing AMP. The circled A indicates an adenine group and the circled R represents either a pyrophosphate (in the case of the cofactor ATP) or nicotinamide mononucleotide monophosphate (in reactions in which NAD is a cofactor).

or NMN. The adenylylation occurs at a lysine residue in a conserved motif in the ligase. In the second step, the activated enzyme binds to a nicked site in the substrate. The adenylyl group is then transferred to the 5′-phosphate at the nick, and finally this activated phosphate is attacked by a nearby 3′-hydroxyl group, forming a phosphodiester bond between the two adjoining strands and AMP is released. The enzyme remains bound to the ligation center throughout the ligation process and recognizes the adenylylated phosphate moiety together with the nearby 3′-hydroxyl. Probably, the reactivity of the activated 5′-pyrophosphate is further increased by bridge complexation to the enzyme via a magnesium ion required in the reaction.

1.5 Substrate Requirements in Enzymatic Ligation Reactions

In contrast to the ligation of restriction enzyme–digested DNA fragments, probes in ligation assays generally hybridize stably under assay conditions. The reaction is therefore more similar to the sealing of an interrupted DNA strand in a duplex molecule. However, since the ligation reaction can trap hybridized molecules through the formation of a covalent bond, it is also possible to use short oligonucleotide probes involving just a few hybridizing bases in ligation assays. By contrast, such short probes would exhibit very limited stability in conventional hybridization-based assays.

As mentioned earlier, either RNA or DNA molecules can be used as probes in reactions catalyzed by T4 DNA ligase, although most commonly, DNA oligonucleotides are employed. Enzymatic ligation requires the presence of a phosphate group at the 5′ end of one probe molecule to be joined to a hydroxyl group at a nearby 3′ end. Ligases are very sensitive to mismatches at the nucleotides on either side of the nick to be sealed (Fig. 4). A mismatch at the 3′ end generally is more inhibitory than one at the 5′ end. Mismatches in other positions of the probe-target hybrid can also significantly reduce the efficiency of ligation. The *Tth* DNA ligase has been shown to ligate substrates with a mismatch nine nucleotides away from the nick in the 5′-direction with a lower efficiency than the corresponding matched substrate. Ligases also require a minimal length of base-paired sequences, which differs between enzymes. As an example, T7 DNA ligase is able to join a 5′ hexamer to a nonamer, while *Tth* DNA ligase is not.

1.6 Ribozyme Ligases

RNA molecules have been found that have the capacity to catalyze ligation of two RNA molecules hybridizing in juxtaposition on a complementary nucleic acid molecule. Such ribozyme ligases have been isolated from large pools of combinatorially synthesized RNA sequences by an iterative process of *in vitro* selection for the ability to ligate themselves to another RNA molecule, and amplification of ligation products. Molecules with a ligation capacity greatly increased over background levels were isolated and sequenced. Nonetheless, RNA ligases remain substantially less efficient than the protein enzymes.

1.7 Chemical Ligation

Also, chemical means can be employed to join nucleic acid strands through a

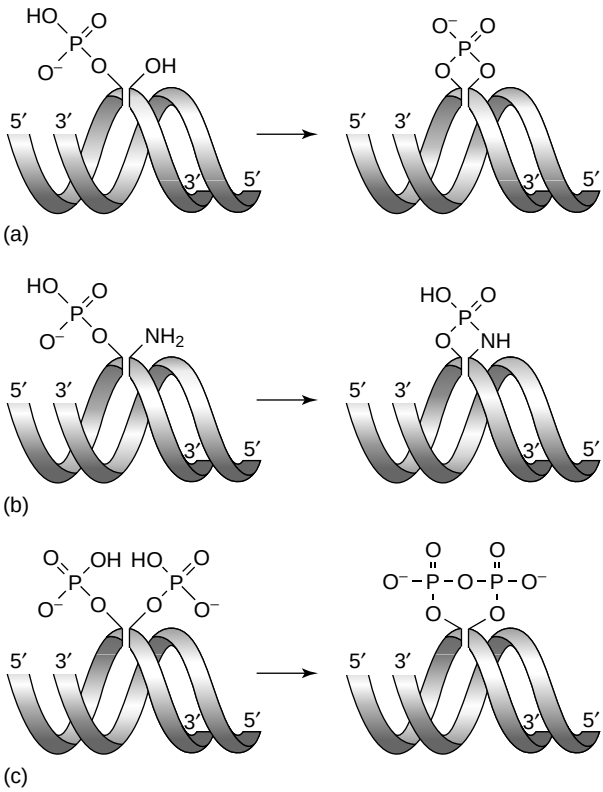

Fig. 3 Chemically induced nucleic acid ligation reactions. Using reagents such as cyanogen bromide or carbodiimide, nucleic acid molecules that are not substrates for the protein enzymes may be joined by chemical ligation. Examples of such reactions are (a) a strand with a 5′-hydroxyl group can be joined to a 3′ phosphate on a nearby strand to form a regular phosphodiester linkage; (b) a 3′ amino function may be joined to a 5′ phosphate, creating a phosphoamide linkage; and (c) two adjoining ends, both bearing phosphate groups, may be joined to form a pyrophosphate linkage.

phosphodiester linkage in a template-dependent manner. Substrate requirements in these reactions may differ from those of enzyme-catalyzed reactions (Fig. 3). For example, cyanogen bromide preferentially participates in the ligation of strands with a 5′-hydroxyl and a 3′-phosphate group, respectively, while protein ligases strictly require a 5′-phosphate group and a 3′-hydroxyl group. It is also possible to modify the ends of nucleic acid molecules so that they can be joined covalently upon hybridization to a template strand, but with no requirement for enzymes or catalysts. For example, upon hybridization of a 3′-phosphorothioate next to a 5′-iodothymidine residue, the two strands will be ligated, forming a phosphorothioate bond.

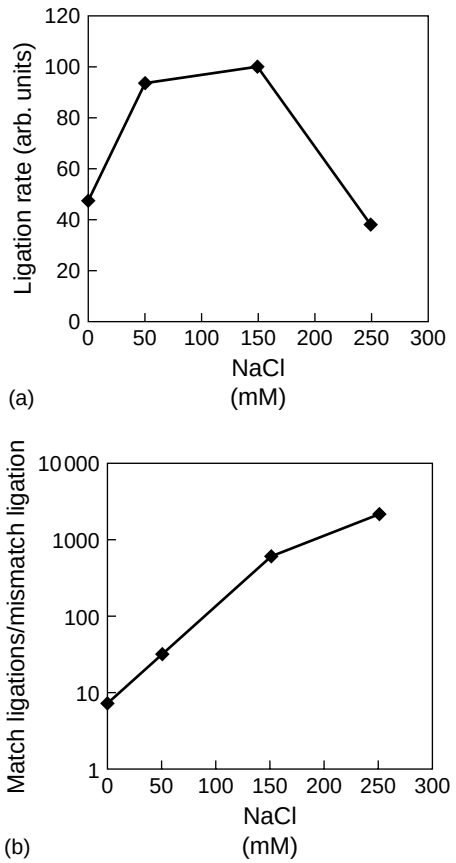

Fig. 4 Ligation at different concentrations of NaCl of two oligonucleotide probes using T4 DNA ligase under limiting conditions. (a) The rate of ligation in the presence of a template oligonucleotide fully matched to the probes. (b) The ratio of the number of ligations on a matched template to the number of ligations on a template mismatched at the position opposite the 3′ ultimate nucleotide at the site of ligation.

Also, pairs of oligonucleotides that form an interrupted third strand in a triple helix by binding to the major groove of a DNA duplex may be joined chemically. Since there is no requirement for substrate recognition by an enzyme, chemical ligation is not limited to the joining of natural nucleotide residues; residues containing other sugars or bases may become covalently joined to nearby strands. Moreover, besides phosphodiester bonds, phosphoroamide, phosphorothioate, and pyrophosphate bonds may be created. It is also possible to join fragments that are shorter than what is required by ligase enzymes. Chemical ligation thus extends the repertoire of substrate molecules that may be joined by ligation.

Because of the properties of nucleic acid-ligation reactions discussed above, ligation reactions have proven useful in a number of assays for analysis of target molecules. Next, we will describe some of these assays.

2 Techniques

2.1 Oligonucleotide Joining Reactions for Detection of Nucleic Acids

Ligation of pairs of oligonucleotide probes was used by Khorana and coworkers in the early 1970s to identify target molecules. Since the end of the 1980s, this has become a popular mechanism for genetic analyses, with the demonstration that DNA sequence variants can be efficiently detected and distinguished by comparing the ligation of pairs of probes that are matched at the junction to one but not the other of two target sequence variants. Over a wide range of ligase concentrations, this oligonucleotide ligation assay (OLA) reaction is inhibited by a single mismatched base pair where the two

strands to be joined abut. This mechanism is now in routine use in clinical diagnostics.

The amount of signal achieved from an OLA reaction performed on genomic DNA samples is generally not sufficient for unambiguous detection. This can be solved by generating more targets, more ligation events per target, or by amplifying the signal after ligation. More targets can be generated by performing a PCR amplification before ligation. If a thermostable ligase is used, thermal cycling may also be employed during the ligation to allow ligated pairs of probes to dissociate from their targets, giving room for new probes to hybridize and be ligated. In the ligase chain reaction (LCR), these two methods are combined. A second pair of probes, complementary to the first pair, is introduced. The ligation of the first pair of probes generates a template for the second pair and *vice versa*. In this manner, thermal cycling of the ligation reaction results in an exponential increase of ligated probes. The method is associated with a tendency to template-independent probe ligation through blunt-end ligation of the probe pairs. This problem can be avoided by using a gap-fill approach, where a gap of one or a few nucleotides first has to be filled in by the action of a polymerase before ligation can occur. The extension can be performed in an allele-specific manner by designing the 3′ ends of the probes to be extended so that they will be correctly matched to one but not the other of the target sequence variant.

A gap-fill polymerization step may also be included in the OLA test to decrease the number of probes required for an assay. In the regular OLA outlined above, the ability of one locus-specific probe to be joined to two different allele-specific probes is investigated. If the probes instead leave a gap corresponding to the variable position, then combined polymerase extension and ligation reactions can be performed in the presence of either allele-specific nucleotide triphosphate. With this method, allele distinction occurs at two levels; first the right nucleotide has to be introduced, and if a mistake is made, the ligation step will still be inhibited. Also, the number of probes that have to be constructed is decreased, at the cost of performing multiple reactions.

As mentioned, signal amplification can also be performed after ligation. This is most commonly done by introducing primer sequences in the probes so that ligation products can be amplified by PCR. Another amplification method is ligation-initiated Qβ amplification. In this reaction, the bacteriophage Qβ replicase enzyme, which has the capacity to replicate certain RNA molecules and then copy both the original strand and the replication product, is used. With this enzyme, a 10^9-fold amplification can be achieved in 30 min. To use this enzyme in a ligation assay, two RNA probes are constructed in such a way that they form a substrate for the Qβ replicase only after ligation to each other.

Several methods are available for detection of the ligated probes. For example, the members of the probe pair can be modified so that the reaction results in a detectable combination of functions in the ligation product. One of the probes may be biotinylated while the other probe carries a fluorophore, allowing ligation products to be captured on a support and detected. In another method, the remote ends of the ligation probes have complementary sequences carrying two different fluorophores. Upon ligation, the probes

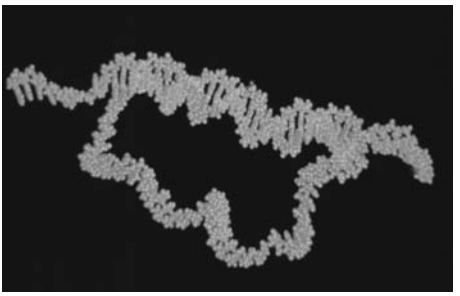

Fig. 5 A padlock probe, designed to be ligated into a circle upon interaction with the proper target sequence. The two ends of the linear probe (yellow) hybridize in juxtaposition on the target sequence (blue). The 5′-phosphate group on one end of the probe about to be joined by ligation to the 3′-hydroxyl at the opposite end is shown in red. Circularized molecules are wound around the target strand and can detach only by strand breakage or by sliding off a nearby end on the target molecule (see color plate p. xxii).

assume a hairpin-loop structure, and exhibit an altered emission spectrum due to the closeness of the two fluorophores. The method has been shown to permit analysis of point mutations in unamplified genomic DNA.

Since any nucleotide mismatches can be distinguished from perfect matches under standard reaction conditions, the oligonucleotide ligation procedure is suitable for automation and can be performed on large series of samples. The throughput of such analyses can be further increased by multiplexing – using many pairs of probes in the same reaction. The detection method must then be adjusted to separate the signals from different pairs of probes. For example, one of the probes of each pair may be size-coded to distinguish different loci by electrophoresis, while the other may be labeled with different fluorophores to identify separate alleles. The method has also been used to measure the copy number of genes. Alternatively, the locus may be represented by a hybridization sequence tag on one probe. Ligation products may be identified after hybridization to DNA microarrays with complements of the tag sequences. With this strategy, it is possible to identify large numbers of genotypes from a single reaction.

2.2 Padlock Probes

In a variant of the OLA technique, the two probes to be joined are in fact the two ends of the same linear probe molecule. Upon hybridization to the target sequence and subsequent ligation, the probe becomes circularized, wound around the target molecule in a padlock-like fashion (Fig. 5). Padlock probes may also be used in a gap-fill reaction to decrease the number of probes required for genotyping by half. If the target has no free ends, then ligated padlock probes will remain linked to the target even after denaturing washes. Thus, in solid-phase assays, unligated probes and any intermolecular ligation products may be efficiently washed away, reducing background signals. The topological link allows padlock probes to be used for localized detection of target DNA or RNA sequences via detectable moieties attached to the probe (Fig. 6).

Besides localized detection, padlock probes also offer particular advantages in multiplex assays. As more probes are combined in one reaction, the risk increases for unwanted ligation between different probes, fortuitously hybridizing next to each other somewhere in the added DNA sample. These ligation products

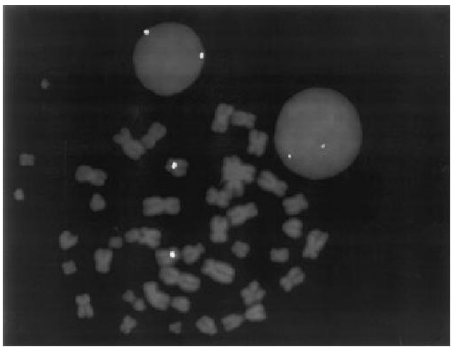

Fig. 6 *In situ* analysis of a repeated DNA sequence, located in the centromeric region of human chromosome 12, using a biotin-modified padlock probe, detected via fluorescein-conjugated streptavidin.

can be ignored, however, since properly ligated padlock probes can be conveniently distinguished from products of inter-probe ligations: Correct ligation of padlock probes results in circular molecules, while ligation of noncognate probe ends results in linear molecules that may be degraded by exonucleolytic treatment, and that fail to template rolling circle replification, as described below.

After ligation, circularized padlock probes may be amplified by PCR across the ligation junction, allowing thousands of different amplification products to be identified by hybridizing to tag microarrays. Circularized padlock probes are also suitable templates for rolling circle replication (RCR) reactions. This reaction is a linear amplification procedure that requires a single primer, and it results in long, single-stranded products that consist of many complements of the DNA circles repeated in tandem. The RCR reaction allows precise quantification and may be monitored in real time by the use of molecular beacons.

The hyper-branched rolling circle amplification (HRCA) reaction is a variant of the RCR procedure that allows a rapid isothermal amplification of circularized probes. Two primers are used; one primer complementary to the circular DNA strand and another one that hybridizes to each complement of the circle in the RCR product. Extension products from the primers along a strand catch up with each other and result in strand displacement of the downstream sequence. This process, in turn, generates new single-stranded sequences where new primers can anneal. The extension and strand displacement process yields a mixture of single- and double-stranded fragments of lengths representing multiples of the starting DNA circle. The amplification reaction is capable of billionfold amplification in a 90-min isothermal incubation. By using distinct primers for allele-specific padlock probes, the HRCA reaction has been used for genotyping.

Padlock probes have also been used to recognize specific target sequences in double-stranded DNA molecules. This can be done by first using PNA (peptide nucleic acid) probes that pry open the DNA duplexes, allowing padlock probes to be attached as the so-called ear ring probes. A similar effect has been achieved with linear probes having midsections that form triplex structures with duplex target DNA molecules. The probes become topologically linked to the target molecules when an oligonucleotide is added that allows the free ends of the probes to be joined. The mechanism has been used to couple

targeting peptides or biotin molecules to plasmid molecules.

2.3 Proximity Ligation

Also, proteins and other macromolecules can be detected via DNA ligation reactions using the so-called proximity ligation method. Pairs of binding reagents, such as antibodies or DNA aptamers binding separate determinants on a specific target protein, bring into proximity the 5′ and 3′ ends of oligonucleotides attached to the binding reagents (Fig. 7). By introducing a ligation template, referred to as a connector oligonucleotide, the DNA sequence extensions of proximity probes that have bound the same target molecule can be joined, forming a DNA strand with sequence elements required for amplification, identification, and visualization. By contrast, 5′ and 3′ ends on probes that have failed to bind in pairs to a target molecule rapidly hybridize to one connector oligonucleotide each, preventing any subsequent hybridization to a common connector and ligation. The method can be performed without any washes and offers a sensitivity that is orders of magnitude greater than conventional sandwich ELISA protein detection assays. Both individual and interacting proteins can be demonstrated. The method can also be performed in solid-phase mode for further increased sensitivity of detection, and more recently a variant of the method has been shown to permit sensitive detection of the location of target proteins *in situ*.

2.4 Other Uses of Ligation Reactions

Ligation reactions have also been used to detect bending and twisting of short double-stranded DNA molecules. By studying the efficiency with which a ligase joins the ends of DNA double helices to form circles, nucleotide sequence-dependent or protein-induced bending of the stiff double-stranded DNA molecule may be estimated. In a similar fashion, binding by transcriptional factor complexes can juxtapose separate regulatory gene sequences, allowing otherwise remote DNA fragment ends to be brought together and be ligated.

3 Outlook

Large-scale SNP genotyping studies currently appear to shift from hybridization- and polymerization-based assays to assays that also include a ligation step. Gene expression profiling may well follow the same trend. Furthermore, the proximity

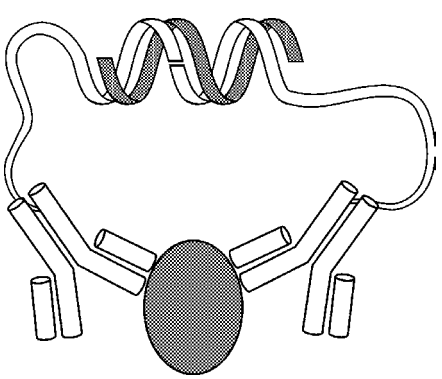

Fig. 7 Proximity ligation. Two proximity probes (white), consisting of antibodies conjugated to oligonucleotides, simultaneously bind to the same target protein molecule (dark shaded). Thereby an oligonucleotide (light shaded) and a ligase, added to template ligation of the two probes, can direct the formation of a DNA strand that can then be amplified and detected to reflect the presence of the target protein.

ligation mechanism can similarly transform large-scale protein expression and interaction studies to a form where DNA strands are created that represent the recognized proteins, and that can be amplified and detected, for example, after hybridization to tag microarrays. The unique event of a ligation reaction connecting two previously separate DNA sequences, or the two ends of individual probe molecules, can be designed to reflect the presence of target DNA, RNA, or protein molecules with excellent specificity, efficiently distinguishing closely similar target molecules. Once formed, the ligation products serve as a class of universally amplifiable, information-carrying molecules. They can therefore be detected with great sensitivity and allow very large numbers of target molecules to be investigated in parallel. In a similar manner, the probe ligation mechanism will also be important to demonstrate the location or the colocation of target molecules in biological specimens. In conclusion, ligation reactions are promising as a general means to investigate large, diverse sets of target molecules, and they may thus serve as the *molecular transistors* that will be required to standardize, miniaturize, and scale-up biological analyses for future highly resolving biological investigations in research, and in a wide range of application areas.

See also Anthology of Human Repetitive DNA; Infectious Disease Testing by LCR; Oligonucleotides.

Bibliography

Books and Reviews

Banér, J., Nilsson, M., Isaksson, A., Mendel-Hartvig, M., Antson, D.O., Landegren, U. (2001) More keys to padlock probes: mechanisms for high-throughput nucleic acid analysis, *Curr. Opin. Biotechnol.* **12**, 11–15.

Demidov, V.V. (2003) Earrings and padlocks for the double helix: topological labeling of duplex DNA, *Trends Biotechnol.* **21**, 148–151.

Doherty, A.J., Dafforn, T.R. (2000) Nick recognition by DNA ligases, *J. Mol. Biol.* **296**, 43–56.

Engler, M.J., Richardson, C.C. (1982) DNA ligases, in: Boyer, P.D. (Ed.) *The Enzymes*, Vol. 15, Academic Press, New York, pp. 3–29.

Gullberg, M., Fredriksson, S., Taussig, M., Jarvius, J., Gustafsdottir, S., Landegren, U. (2003) A sense of closeness: protein detection by proximity ligation, *Curr. Opin. Biotechnol.* **14**, 82–86.

Higgins, N.P., Cozzarelli, N.R. (1979) DNA-joining enzymes: a review, *Methods Enzymol.* **68**, 50–71.

Jarvius, J., Nilsson, M., Landegren, U. (2003) Oligonucleotide ligation assay, *Methods Mol. Biol.* **212**, 215–228.

Nilsson, M., Baner, J., Mendel-Hartvig, M., Dahl, F., Antson, D.O., Gullberg, M., Landegren, U. (2002) Making ends meet in genetic analysis using padlock probes, *Hum. Mutat.* **19**, 410–415.

Primary Literature

Abravaya, K., Carrino, J.J., Muldoon, S., Lee, H.H. (1995) Detection of point mutations with a modified ligase chain reaction (Gap-LCR), *Nucleic Acids Res.* **23**, 675–682.

Alsmadi, O.A., Bornarth, C.J., Song, W., Wisniewski, M., Du, J., Brockman, J.P., Faruqi, A.F., Hosono, S., Sun, Z., Du, Y., Wu, X., Egholm, M., Abarzua, P., Lasken, R.S., Driscoll, M.D. (2003) High accuracy genotyping directly from genomic DNA using a rolling circle amplification based assay, *BMC Genomics* **4**, 21.

Antson, D.O., Isaksson, A., Landegren, U., Nilsson, M. (2000) PCR-generated padlock probes detect single nucleotide variation in genomic DNA, *Nucleic Acids Res.* **28**, E58.

Banér, J., Isaksson, A., Waldenstrom, E., Jarvius, J., Landegren, U., Nilsson, M. (2003) Parallel gene analysis with allele-specific padlock probes and tag microarrays, *Nucleic Acids Res.* **31**, e103.

Banér, J., Nilsson, M., Mendel-Hartvig, M., Landegren, U. (1998) Signal amplification of

padlock probes by rolling circle replication, *Nucleic Acids Res.* **26**, 5073–5078.

Barany, F. (1991) Genetic disease detection and DNA amplification using cloned thermostable ligase, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 189–193.

Besmer, P., Miller, R.C. Jr., Caruthers, M.H., Kumar, A., Minamoto, K., Van de Sande, J.H., Sidarova, N., Khorana, H.G. (1972) Studies on polynucleotides. CXVII. Hybridization of polydeoxynucleotides with tyrosine transfer RNA sequences to the r-strand of phi80psu + 3 DNA, *J. Mol. Biol.* **72**, 503–522.

Cherepanov, A.V., de Vries, S. (2002) Dynamic mechanism of nick recognition by DNA ligase, *Eur. J. Biochem.* **269**, 5993–5999.

Crothers, D.M., Drak, J., Kahn, J.D., Levene, S.D. (1992) DNA bending, flexibility, and helical repeat by cyclization kinetics, *Methods Enzymol.* **212**, 3–29.

Dolinnaya, N.G., Tsytovich, A.V., Sergeev, V.N., Oretskaya, T.S., Shabarova, Z.A. (1991) Structural and kinetic aspects of chemical reactions in DNA duplexes. Information on DNA local structure obtained from chemical ligation data, *Nucleic Acids Res.* **19**, 3073–3080.

Ekland, E.H., Szostak, J.W., Bartel, D.P. (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences, *Science* **269**, 364–370.

Fan, J.B., Chen, X., Halushka, M.K., Berno, A., Huang, X., Ryder, T., Lipshutz, R.J., Lockhart, D.J., Chakravarti, A. (2000) Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays, *Genome Res.* **10**, 853–860.

Faruqi, A.F., Hosono, S., Driscoll, M.D., Dean, F.B., Alsmadi, O., Bandaru, R., Kumar, G., Grimwade, B., Zong, Q., Sun, Z., Du, Y., Kingsmore, S., Knott, T., Lasken, R.S. (2001) High-throughput genotyping of single nucleotide polymorphisms with rolling circle amplification, *BMC Genomics* **2**, 4.

Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gustafsdottir, S.M., Ostman, A., Landegren, U. (2002) Protein detection using proximity-dependent DNA ligation assays, *Nat. Biotechnol.* **20**, 473–477.

Grossman, P.D., Bloch, W., Brinson, E., Chang, C.C., Eggerding, F.A., Fung, S., Iovannisci, D.M., Woo, S., Winn-Deen, E.S., Iovannisci, D.A. (1994) High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation, *Nucleic Acids Res.* **22**, 4527–4534.

Hardenbol, P., Baner, J., Jain, M., Nilsson, M., Namsaraev, E.A., Karlin-Neumann, G.A., Fakhrai-Rad, H., Ronaghi, M., Willis, T.D., Landegren, U., Davis, R.W. (2003) Multiplexed genotyping with sequence-tagged molecular inversion probes, *Nat. Biotechnol.* **21**, 673–678.

Hsuih, T.C., Park, Y.N., Zaretsky, C., Wu, F., Tyagi, S., Kramer, F.R., Sperling, R., Zhang, D.Y. (1996) Novel, ligation-dependent PCR assay for detection of hepatitis C in serum, *J. Clin. Microbiol.* **34**, 501–507.

Landegren, U., Kaiser, R., Sanders, J., Hood, L. (1988) A ligase-mediated gene detection technique, *Science* **241**, 1077–1080.

Lizardi, P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C., Ward, D.C. (1998) Mutation detection and single-molecule counting using isothermal rolling-circle amplification, *Nat. Genet.* **19**, 225–232.

Mendel-Hartvig, M., Kumar, A., Landegren, U. (2004) Ligase-mediated construction of branched DNA strands: a novel DNA joining activity catalyzed by T4 DNA ligase, *Nucleic Acids Res.* **32**,, e2.

Nilsson, M., Antson, D.O., Barbany, G., Landegren, U. (2001) RNA-templated DNA ligation for transcript analysis, *Nucleic Acids Res.* **29**, 578–581.

Nilsson, M., Barbany, G., Antson, D.O., Gertow, K., Landegren, U. (2000) Enhanced detection and distinction of RNA by enzymatic probe ligation, *Nat. Biotechnol.* **18**, 791–793.

Nilsson, M., Gullberg, M., Dahl, F., Szuhai, K., Raap, A.K. (2002) Real-time monitoring of rolling-circle amplification using a modified molecular beacon design, *Nucleic Acids Res.* **30**, e66.

Nilsson, M., Krejci, K., Koch, J., Kwiatkowski, M., Gustavsson, P., Landegren, U. (1997) Padlock probes reveal single-nucleotide differences, parent of origin and in situ distribution of centromeric sequences in human chromosomes 13 and 21, *Nat. Genet.* **16**, 252–255.

Nilsson, M., Malmgren, H., Samiotaki, M., Kwiatkowski, M., Chowdhary, B.P., Landegren, U. (1994) Padlock probes: circularizing oligonucleotides for localized DNA detection, *Science* **265**, 2085–2088.

Odell, M., Shuman, S. (1999) Footprinting of Chlorella virus DNA ligase bound at a

nick in duplex DNA, *J. Biol. Chem.* **274**, 14032–14039.

Park, Y.N., Abe, K., Li, H., Hsuih, T., Thung, S.N., Zhang, D.Y. (1996) Detection of hepatitis C virus RNA using ligation-dependent polymerase chain reaction in formalin-fixed, paraffin-embedded liver tissues, *Am. J. Pathol.* **149**, 1485–1491.

Pickering, J., Bamford, A., Godbole, V., Briggs, J., Scozzafava, G., Roe, P., Wheeler, C., Ghouze, F., Cuss, S. (2002) Integration of DNA ligation and rolling circle amplification for the homogeneous, end-point detection of single nucleotide polymorphisms, *Nucleic Acids Res.* **30**, e60.

Pritchard, C.E., Southern, E.M. (1997) Effects of base mismatches on joining of short oligodeoxynucleotides by DNA ligases, *Nucleic Acids Res.* **25**, 3403–3407.

Qi, X., Bakht, S., Devos, K.M., Gale, M.D., Osbourn, A. (2001) L-RCA (ligation-rolling circle amplification): a general method for genotyping of single nucleotide polymorphisms (SNPs), *Nucleic Acids Res.* **29**, E116.

Roulon, T., Coulaud, D., Delain, E., Le Cam, E., Helene, C., Escude, C. (2002) Padlock oligonucleotides as a tool for labeling superhelical DNA, *Nucleic Acids Res.* **30**, E12.

Roulon, T., Helene, C., Escude, C. (2002) Coupling of a targeting peptide to plasmid DNA using a new type of padlock oligonucleotide, *Bioconjugate Chem.* **13**, 1134–1139.

Samiotaki, M., Kwiatkowski, M., Parik, J., Landegren, U. (1994) Dual-color detection of DNA sequence variants by ligase-mediated analysis, *Genomics* **20**, 238–242.

Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F., Pals, G. (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification, *Nucleic Acids Res.* **30**, e57.

Tyagi, S., Landegren, U., Tazi, M., Lizardi, P.M., Kramer, F.R. (1996) Extremely sensitive, background-free gene detection using binary probes and beta replicase, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5395–5400.

Wabuyele, M.B., Farquar, H., Stryjewski, W., Hammer, R.P., Soper, S.A., Cheng, Y.W., Barany, F. (2003) Approaching real-time molecular diagnostics: single-pair fluorescence resonance energy transfer (spFRET) detection for the analysis of low abundant point mutations in K-ras oncogenes, *J. Am. Chem. Soc.* **125**, 6937–6945.

Weiss, B., Thompson, A., Richardson, C.C. (1968) Enzymatic breakage and joining of deoxyribonucleic acid. VII. Properties of the enzyme-adenylate intermediate in the polynucleotide ligase reaction, *J. Biol. Chem.* **243**, 4556–4563.

Wu, D.Y., Wallace, R.B. (1989) Specificity of the nick-closing activity of bacteriophage T4 DNA ligase, *Gene* **76**, 245–254.

Xu, Y., Karalkar, N.B., Kool, E.T. (2001) Nonenzymatic autoligation in direct three-color detection of RNA and DNA point mutations, *Nat. Biotechnol.* **19**, 148–152.

Yeakley, J.M., Fan, J.B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M.S., Fu, X.D. (2002) Profiling alternative splicing on fiber-optic arrays, *Nat. Biotechnol.* **20**, 353–358.

Zhang, D.Y., Brandwein, M., Hsuih, T.C., Li, H. (1998) Amplification of target-specific, ligation-dependent circular probe, *Gene* **211**, 277–285.

Lipid and Lipoprotein Metabolism

Clive R. Pullinger and John P. Kane
University of California, San Francisco, CA

1 Lipid Transport and Disorders 199
1.1 Exogenous Transport 199
1.1.2 Chylomicron Assembly 199
1.1.3 Chylomicron Catabolism 200
1.1.4 Lipoprotein Lipase 200
1.1.5 Chylomicron Remnant Uptake 202
1.2 Endogenous Transport 204
1.2.1 VLDL Assembly 204
1.2.2 VLDL Catabolism 205
1.2.3 LDL Catabolism 207
1.3 Disorders of Lipid Transport 207
1.3.1 Abetalipoproteinemia 207
1.3.2 Hepatic Lipase Deficiency 208
1.3.3 Familial Hypercholesterolemia 208
1.3.4 Familial Ligand-defective Apolipoprotein B-100 209
1.3.5 Familial Hypobetalipoproteinemia 210
1.3.6 Familial Combined Hyperlipidemia 210
1.3.7 Autosomal Recessive Hypercholesterolemia 211
1.3.8 Autosomal Dominant Hypercholesterolemia 211
1.3.9 Chylomicron Retention Disease and Anderson Disease 212
1.3.10 Sitosterolemia 212

2 Reverse Cholesterol Transport 213
2.1 ATP-Binding Cassette Transporter A1 214
2.1.1 Tangier Disease 214
2.2 Caveolin 214
2.3 Lecithin Cholesterol Acyl Transferase 215
2.4 Phospholipid Transfer Protein 215

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

2.5 Cholesteryl Ester Transfer Protein 216
2.6 Endothelial Lipase 216

3 Cholesterol Metabolism 217
3.1 Synthesis 217
3.1.1 HMGCoA Synthase and Reductase 217
3.1.2 Smith–Lemli–Opitz Syndrome 218
3.1.3 Acyl CoA: Cholesterol Acyltransferases 218
3.2 Catabolism: Bile Acid Metabolism and Disorders 219
3.2.1 Cholesterol 7α-hydroxylase Deficiency 221
3.2.2 Oxysterol 7α-Hydroxylase Deficiency 221
3.2.3 Cerebrotendinous Xanthomatosis 221
3.3 Disorders of Intracellular Lipoprotein Degradation 221
3.3.1 Wolman Disease and Cholesteryl-ester Storage Disease 221
3.3.2 Niemann–Pick Disease Type C 222

4 Triglyceride and Phospholipid Metabolism 222

5 Role of Nuclear Hormone Receptors 223
5.1 Sterol Regulatory Element Binding Proteins 223
5.2 Peroxisome Proliferator-activated Receptors 224
5.3 Liver X Receptors 224
5.4 Farnesoid X Receptor and Pregnane X Receptor 225

6 The Apolipoproteins 225
6.1 The Apolipoprotein Multigene Family 225
6.2 Apolipoprotein A-I 227
6.3 Apolipoprotein A-II 227
6.4 Apolipoprotein A-IV 228
6.5 Apolipoprotein A-V 228
6.6 Apolipoprotein B 229
6.7 Apolipoprotein C-I 230
6.8 Apolipoprotein C-II 231
6.8.1 Apolipoprotein C-II Deficiency 231
6.9 Apolipoprotein C-III 231
6.10 Apolipoprotein C-IV 232
6.11 Apolipoprotein E 233
6.11.1 Dysbetalipoproteinemia and ApoE Deficiency 234
6.11.2 Alzheimer Disease and ApoE 234
6.12 Apolipoprotein (a) 235
6.13 Other Apolipoproteins 236
6.13.1 Apolipoprotein D 236
6.13.2 Apolipoprotein H 236
6.13.3 Apolipoprotein J 236
6.13.4 Apolipoprotein L Gene Family 237

7 Scavenger Receptors 237
7.1 SR-BI/CLA1 237
7.2 Modified LDL Receptors: MSR1, CD36 and LOX1 237

8 Lipids and Atherosclerosis 239

Bibliography 239
Books and Reviews 239
Primary Literature 239

Keywords

Apolipoproteins
Proteins associated with the plasma lipoproteins.

Chylomicrons
A class of triglyceride-rich lipoproteins of intestinal origin, isolated by ultracentrifugation at a density of 0.93 g mL^{-1}.

HDL
High-density lipoproteins: density range 1.063 to 1.21 g mL^{-1}.

IDL
Intermediate-density lipoproteins: density range 1.006 to 1.019 g mL^{-1}.

Kringles
Named from their resemblance to Danish pretzels, these triloop peptide domains of 80 to 90 amino acid residues are held rigidly together by three internal disulfide bonds in a 1–6, 2–4, 3–5 pattern. They are found in a number of proteins including apolipoprotein (a), prothrombin, factor XII, plasminogen, tissue plasminogen activator, and urokinase.

LDL
Low-density lipoproteins: density range 1.019 to 1.063 g mL^{-1}.

VLDL
Very low-density lipoproteins, a class of triglyceride-rich lipoproteins: density range 0.93 to 1.006 g mL^{-1}.

Abbreviations

CE Cholesteryl ester
CETP cholesteryl-ester transfer protein

LCAT	lecithin:cholesterol acyltransferase
LPL	lipoprotein lipase
LRP	LDL-receptor-like protein
MTP	microsomal triglyceride transfer protein
FFA	free fatty acids
PLTP	phospholipid transfer protein
PPAR	peroxisome proliferator-activated receptor
SREBP	sterol regulatory element binding protein
TM	transmembrane

Lipoprotein metabolism can be subdivided into transport of exogenous lipids, transport of endogenous lipids from the liver to peripheral tissues, and reverse cholesterol transport with eventual excretion of cholesterol and its catabolic end products, bile salts, into the bile. Lipids are transported through the bloodstream as macromolecular complexes called *lipoproteins*, of which there are a number of classes that can be separated by sequential ultracentrifugation. Lipoprotein particles consist of a hydrophobic lipid core that is stabilized by a monolayer of phospholipids, free cholesterol, and apolipoproteins. Nine apolipoproteins are members of a multigene family. Eight of these are clustered at two loci: apoA-I, apoA-IV, apoA-V, and apoC-III on chromosome 11; and apoC-I, apoC-II apoC-IV, and apoE on chromosome 19. The *apoB* gene codes for two major structural proteins, one of which is produced by a novel mRNA editing mechanism.

There are an increasing number of key genes recognized that play important roles in lipid metabolism and transport. Those that code for intracellular proteins involved in cholesterol synthesis, storage, or catabolism include HMGCoA synthase (HMGCS1), HMGCoA reductase (HMGCR), acyl-CoA:cholesterol acyltransferase (ACAT1/SOAT1 and ACAT2/SOAT2), and cholesterol 7α-hydroxylase (CYP7A1). The diacylglycerol O-acyltransferases (DGAT1 and DGAT2) are important in triglyceride synthesis. The microsomal triglyceride transfer protein (MTP) is a major participant in the assembly of very low density lipoproteins (VLDL). There is a family of lipoprotein receptor genes, most notable of which is the LDL receptor. Similarly, there is a family of lipases including lipoprotein lipase (LPL), hepatic triglyceride lipase (LIPC), and endothelial lipase (LIPG). In addition to the apolipoproteins, but also associated with lipoproteins, are a number of gene products that are important for lipid transport. These include lecithin:cholesterol acyltransferase (LCAT), phospholipid transfer protein (PLTP), and cholesterol ester transfer protein (CETP). Key genes involved in reverse cholesterol transport are ATP-binding cassette A1 (ABCA1), scavenger receptor SR-BI/CLA1 (SCARB1), and caveolin (CAV1). Other members of the ABC transporter family are being increasingly seen as major players in lipid transport. For example, ABCA7 and ABCB4 are phospholipid transporters, and ABCB11 is the bile salt export protein.

Over the last decade, a number of nuclear receptors that regulate the genes mentioned above have been identified. These include the sterol regulatory element

binding proteins (SREBPs), the liver X receptors (LXRα, NR1H3, and LXRβ, NR1H2), the farnesoid X receptor (FXR, NR1H4), and the peroxisome proliferator-activated receptors (PPARα, PPARγ, and PPARδ).

Recent advances in molecular cell biology and molecular medicine, involving studies of these genes and their products, have significantly increased our understanding of lipoprotein metabolism and atherosclerotic heart disease.

1 Lipid Transport and Disorders

1.1 Exogenous Transport

Exogenous transport, illustrated in Fig. 1, describes the mechanism by which large quantities of dietary triglycerides and cholesterol are moved through the blood for storage. Prior to absorption, triglycerides are hydrolyzed by lipases to fatty acids and β-monoacylglycerol. The lipids are absorbed in the small intestine with the aid of emulsifying bile acids. Triglycerides are resynthesized and much of the cholesterol is esterified in the enterocytes.

1.1.2 Chylomicron Assembly

The hydrophobic triglycerides and cholesterol esters are packaged in the smooth endoplasmic reticulum (ER) of enterocytes as large stable microemulsion particles called *chylomicrons*. The nascent chylomicrons, many over 100 nm in diameter, are exocytosed *via* the Golgi apparatus into the lymphatics, whence they migrate through the thoracic duct into the blood. In addition to a single molecule of the structural protein apoB-48 they contain, primarily, apoA-I, apoA-II, and apoA-IV. As a result of exchange with HDL they lose the A apolipoproteins and gain C apolipoproteins.

The large polypeptide subunit of microsomal triglyceride transfer protein, which has homology to vitellogenin, is an 88-kDa protein the gene for which, *MTP* (chromosome 4q24), is expressed in intestine and liver and is required for lipoprotein assembly. It can be shown, *in vitro*, to enhance the transfer of triglycerides, cholesteryl esters, and phospholipids between phospholipid membranes and to mediate the lipidation of apoB. The activity, present in the microsomal fraction, resides in a heterodimer of MTP with a 58-kDa multifunctionalprotein disulfide isomerase (PDI), which acts as a chaperone that ensures correct protein folding. The MTP promoter is regulated positively by cholesterol and negatively by insulin, the latter *via* the MAPK(erg) cascade. It contains the recognition sequences for AP-1 and for the liver-specific factors HNF-1 and HNF-4. It also contains a modified sterol response element (SRE) at −124 to −116 that binds SREBP. Negative response elements have been found for ethanol and IL-1 at −612 to −142, and −121 to −88, respectively. In a large population-based study, a functional SNP, −493 G/T, was not associated with lipoprotein levels or CAD risk. Other reports have been contradictory, showing either decreased or increased LDL cholesterol in subjects homozygous for the less common −493 T variant. The degree of impact of this SNP is thought to depend on the level of visceral

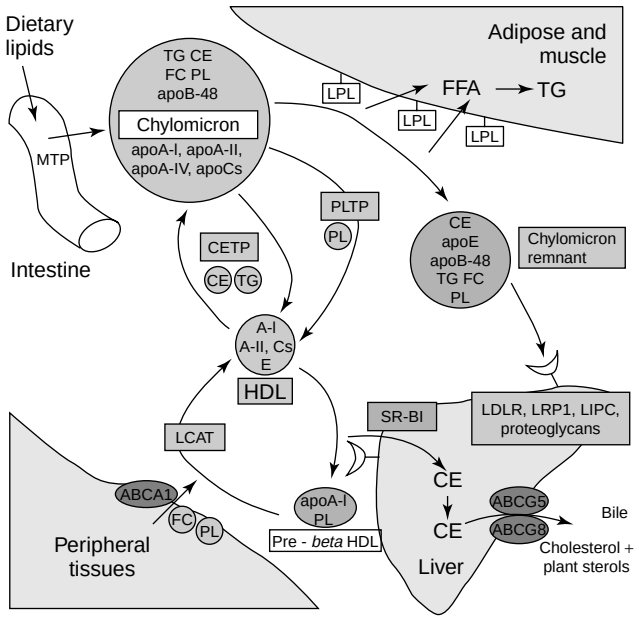

Fig. 1 Exogenous lipid transport (CE, cholesteryl ester; CETP, cholesteryl-ester transfer protein; FC, free cholesterol; LCAT, lecithin:cholesterol acyltransferase; LDLR, LDL receptor; LIPC, hepatic lipase; LPL, lipoprotein lipase; LRP1, LDL receptor-related protein 1; MTP, microsomal triglyceride transfer protein; FFA, free fatty acid; PL; phospholipid; PLTP, phospholipid transfer protein; SR-BI, scavenger receptor class B, 1; TG, Triglyceride.).

adipose tissue, and it may contribute to dyslipidemia in diabetes. The level of MTP mRNA was increased in both intestine and liver when hamsters were fed a high fat diet. The response was more rapid in the intestine.

1.1.3 Chylomicron Catabolism

1.1.4 Lipoprotein Lipase

Most of the triglyceride core of chylomicrons is hydrolyzed by the enzyme lipoprotein lipase (LPL), a triglyceride hydrolase that is synthesized at a number of sites, notably adipose tissue, skeletal muscle, heart, adrenal glands, placenta, and mammary gland. LPL, originally referred to as *clearing factor lipase*, is secreted from parenchymal cells and is transported to the vascular endothelial luminal surface where it attaches, as a noncovalent homodimer, to a high-affinity binding site. The precise nature of this interaction is unknown but is thought to involve cell membrane-anchored heparan sulfate proteoglycans such as perlecan, the glypicans, and the syndecans, and a specific heparin-sensitive binding protein. The hydrolysis of triglyceride releases free fatty acids (FFA) and β-monoacylglycerol. FFA are taken up by adipose tissue where they are re-esterified and stored as triglyceride. In other tissues, such as the heart, FFA can be oxidized as a source of energy. LPL can also facilitate the cellular binding and uptake of lipoproteins by lipoprotein receptors.

The gene for LPL (chromosomal location, 8p21.3) spans 30 kb, is comprised of 10 exons, and codes for a 475-amino acid-residue protein with a 27-residue signal peptide. Two sizes of mRNA, 3.35 and 3.75 kb, are observed. Under different physiological conditions, LPL is regulated hormonally primarily by post-transcriptional control. This control, in the case of insulin, is *via* stabilization of mRNA, and, with thyroid hormone and adrenaline, this is at the translational level. Regulatory cis-acting sequences and transcription factors responsible for basal promoter activity have been identified, as has a promoter region responsible for estrogen suppression of transcription. In adults, tissue-specific suppression of *LPL* gene expression in the liver involves the binding of transcription factor RF-2-LPL to a NF-1-like site in the region −591 to −288 of the transcription initiation site. The role of a peroxisome proliferator response element (PPRE), which binds peroxisome proliferator-activated receptors α or γ (PPARα and PPARγ) is enigmatic. Induction of PPARs with their ligands has been reported to either induce or suppress LPL transcription. A further complication is that changes in LPL mRNA do not always correlate with enzyme activity. An SRE at −90 to −81 that binds SREBP is responsible for the upregulation seen in adipose tissue as a result of cholesterol depletion. In tissues other than adipose or muscle, a positive response to sterols is mediated by an LXRα response element, a DR4 sequence in the first intron at +635 to +650.

The mature enzyme is a 421-residue, 55-kDa, N-linked glycoprotein. Ser132, Asp156 and His241 make up the active site, which is homologous to that in pancreatic, hepatic, and endothelial lipase. All these enzymes have a catalytic site pocket covered by a lid (LPL residues 216 to 239) that opens upon binding of the substrate. These structures play a vital role in determining the substrate specificity of each lipase. Unlike the other lipases in the family, LPL requires apoC-II as a cofactor. Asn43, an N-linked glycosylation site, is essential for catalysis and secretion of LPL. Sites in LPL have been identified that bind apoC-II and lipid. There are two sites rich in arginine and lysine residues, at 279 to 282 and 292 to 304, that bind to heparin.

The rare autosomal recessive disorder, familial lipoprotein lipase deficiency is characterized by diminished LPL activity and chylomicronemia. Numerous mutations in the *LPL* gene have been identified. Many are single-base substitutions; one is a 2-kb insertion and one a 6-kb deletion. Exon 5 appears to be a mutational hot spot, with numerous missense mutations that produce inactive LPL. Homozygosity for G188E or A221del, two of the most common mutations, causes chylomicronemia. Another relatively common mutant, N291S, decreases LPL activity, but is not severe enough on its own to cause chylomicronemia.

Four out of five cases of pregnancy-induced chylomicronemia were shown to be due to partial LPL deficiency caused by mutations in the *LPL* gene. In these subjects, hypertriglyceridemia became much more pronounced during pregnancy, when there is normally a much higher rate of production of VLDL.

Two SNPs in the *LPL* gene, −93 T/G and D9N, are in strong linkage disequilibrium in Caucasians and are associated with differences in the levels of plasma triglyceride. The Asn9 allele is present at a population frequency of 1.5%, with carriers having triglyceride values 25 mg dL^{-1} higher than noncarriers and a higher risk of heart disease. More common

is the S447X variant. The 447X allele, with a population frequency as high as 25%, is associated with decreased triglycerides and lower risk of vascular disease.

In transgenic mice expressing a high activity of human LPL in heart, skeletal muscle, and adipose tissue, the plasma levels of VLDL triglyceride were much decreased and HDL2 levels were raised. The clearance of VLDL and the conversion of VLDL to LDL were enhanced. There was no increase in VLDL levels in these animals after sucrose feeding as seen with normal mice. When placed on a high-cholesterol diet, the animals were protected against hypercholesterolemia.

1.1.5 Chylomicron Remnant Uptake

The fatty acids released by LPL are taken up by the surrounding tissue, a process facilitated by fatty acid transport proteins. They are stored as triglyceride, or used as fuel. Along with free cholesterol and phospholipid, apoC-II is shed, as the chylomicron particles get progressively smaller. During this process they acquire apoE and cholesterol esters, and lose most of the other apoproteins, except apoB-48, by transfer with HDL. Ultimately, chylomicron remnant particles rich in cholesterol esters are formed. These are rapidly removed at the microvillus surface of hepatocytes by a high-affinity process. Earlier it was thought that a distinct chylomicron remnant receptor existed. Despite much evidence for its existence it eluded purification and cloning. It is now accepted that the LDL receptor can initially bind to chylomicron remnants with apoE acting as the ligand. Remnants bound to the LDL receptor move rapidly to clathrin-coated pits at the base of the microvilli and are endocytosed. Hepatic lipase, which is bound to cell-surface heparan sulfate proteoglycans, is also involved in initial binding to remnants. The space of Disse is rich in apoE, which is also bound to the proteoglycans. The remnant particles accumulate apoE, increasing their affinity for the LDL-receptor-related protein or LRP1 (also referred to as the *α_2-macroglobulin receptor*).

LRP1 (Table 1) is a large multifunctional cell-surface receptor comprised of a large 515-kDa extracellular domain and an 85-kDa membrane-spanning subunit. It is expressed in a number of tissues including liver, adrenal cortex, ovary, fibroblasts, macrophages, and some neurons. In addition to binding to α_2-macroglobulin-protease complexes LRP1 has been shown to bind apoE-enriched βVLDL, Pseudomonas exotoxin A and plasminogen activator-inhibitor complexes. The binding and uptake of these ligands is blocked by a 39-kDa receptor-associated protein (RAP) that copurifies with the receptor. LRP1 binds with high affinity to the C-terminal domain of lipoprotein lipase (residues 378 to 423 of human LPL). Knockout of LRP1 is lethal to the embryo.

Another member of the lipoprotein receptor family, LDL receptor-related protein 2 (LRP2), also referred to as *glycoprotein 330 (gp330)* (Table 1), is expressed by the epithelial cells of a number of tissues. It is found mainly as a cell-surface protein concentrated in clathrin-coated pits and is structurally related both to the LRP1 receptor and to the LDL receptor. Like LRP1, LRP2 binds to RAP, LPL, and apoE-enriched βVLDL. It also binds with high affinity to apoJ. Because LRP2 binds LPL with high affinity, like LRP1, it may be involved in the hepatic clearance of LPL-associated lipoproteins. It is mainly expressed in specialized epithelial cells, particularly in proximal renal tubules. Whatever the precise role of LRP1

Tab. 1 Six human lipoprotein receptor genes.

Gene	*Alternative name*	*HUGO symbol*	*Gene size [kb]*	*Exons*	*Chromosomal location*	*Initial translated protein amino acid residues*	*Mature protein molecular weight [kDa]*
ApoB48 receptor		APOB48R	4.3	4	16p12.1	1088	200
LDL receptor		LDLR	43	18	19p13.2	860	120[a]
LDL receptor-related protein 1	alpha2-macroglobulin receptor	LRP1	92	89	12q13.3	4544	515 and 85
LDL receptor-related protein 2	Glycoprotein 330 or megalin	LRP2	234	79	2q31.1	4655	519
LDL receptor-related protein 8	apoE receptor-2	LRP8	81	19	1p32.3	963	154
VLDL receptor		VLDLR	33	19	9p24.2	873	130

[a]Apparent size is 160 kDa on SDS gels.

and LRP2 in chylomicron remnant removal may be, evidence in mice suggests a greater role for the LDL receptor itself.

A recently recognized receptor expressed on monocytes, macrophages, and endothelial cells is the apolipoprotein B48 receptor (APOB48R), which is capable of binding apoB-48-containing triglyceride-rich lipoproteins. It is possible that this protein plays a role in the formation of foam cells in the artery wall.

1.2 Endogenous Transport

The transport of endogenous lipids, illustrated in Fig. 2, describes the movement through the circulation of lipids synthesized in the liver.

1.2.1 VLDL Assembly

The liver secretes triglyceride-rich lipoproteins analogous to chylomicrons. These nascent VLDL particles are smaller than chylomicrons but have a similar lipid composition. In humans, they contain one copy of apoB-100 as their structural protein as well as apoE and the A and C families of apoproteins. The triglyceride used for their assembly in the liver is derived from chylomicron remnants and from fatty acids released by LPL and taken up by the liver, or from fatty acids synthesized *de novo*. As with chylomicrons, MTP plays a critical role in the assembly of nascent VLDL. Unlike chylomicron secretion, which is maintained only throughout the postprandial period, VLDL production continues, albeit at a reduced rate, even during periods of starvation, when it is

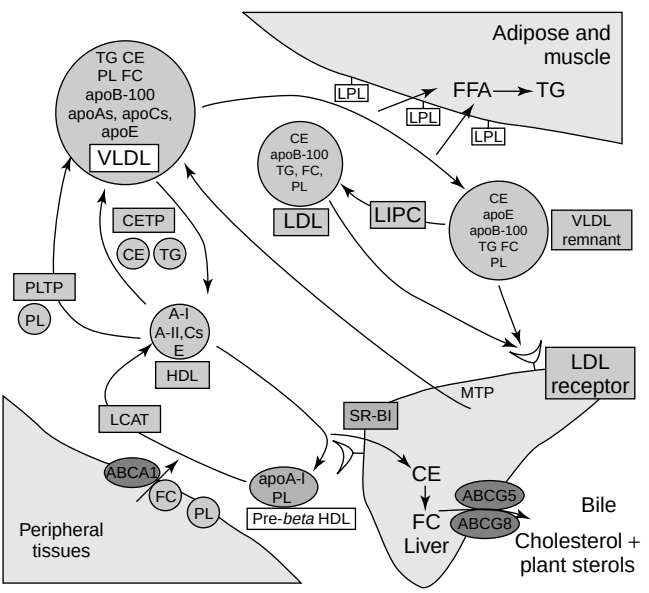

Fig. 2 Endogenous lipid transport (CE, cholesteryl ester; CETP, cholesteryl-ester transfer protein; FC, free cholesterol; LCAT, lecithin:cholesterol acyltransferase; LIPC, hepatic lipase; LPL, lipoprotein lipase; MTP, microsomal triglyceride transfer protein; FFA, free fatty acid; PL, phospholipid; PLTP, phospholipid transfer protein; SR-BI, scavenger receptor class B, 1; TG, Triglyceride.).

an important means of providing muscle fuel.

In the relatively common disorder familial combined hyperlipidemia (see Sect. 1.3.6), overproduction and secretion of VLDL particles by the liver is the cause of the increased level of circulating apoB and triglyceride. The important role of stearoyl-CoA desaturase (SCD), which catalyzes the formation of the monounsaturated fatty acids oleate and palmitoleate, in regulating the supply of triglyceride, phospholipid, and cholesteryl esters for VLDL production, has been recognized.

1.2.2 VLDL Catabolism

Nascent VLDL loses apoA-I and apoE and gains more C apoproteins by a process of exchange with HDL. The triglyceride core of the mature VLDL is hydrolyzed by LPL in a similar manner to chylomicrons, though at a slower rate.

As the hydrolysis of VLDL triglyceride proceeds, transfer with HDL restores apoE and removes most of the C proteins. The resulting VLDL remnants contain a single molecule of apoB-100 and several copies of apoE. About half the remnants are removed by the hepatic LDL receptor, which has high-affinity binding sites for apoE. The remaining remnants lose more triglyceride by transfer to HDL and by the action of hepatic lipase (LIPC), and are converted to LDL particles (Fig. 3) that are rich in cholesterol esters. During this process, the particle decreases in size to about 22 nm in diameter and the apoE and C proteins are lost, leaving apoB-100 as the sole protein.

Hepatic lipase has similar substrate specificity for triglycerides as LPL, but hydrolyses phospholipid two to three times more effectively. It is not activated by apoC-II and does not require Ca^{2+}. Like LPL, its optimum activity is at pH 8.0 to 9.0. Evidence points to a role for hepatic lipase, in addition to LPL, in the production of remnant particles that can be efficiently taken up by the liver. It contributes, *via* an interaction with extracellular proteoglycans, to the process of binding and uptake of remnants by the liver.

The gene for hepatic lipase (chromosome 15q21.3) is 35 kb in length and is composed of 9 exons. The promoter region has tissue-specific glucocorticoid and cAMP response elements. Thyroid hormone has been shown to increase the activity of hepatic lipase by a posttranslational mechanism. Mature hepatic lipase is a glycoprotein (60 kDa) of 476 residues. As a result of alternative splicing, the adrenal gland produces an isoform that is different from the liver enzyme. Four polymorphisms in the LIPC promoter, at −250, −514, −710 and −763, are in almost complete linkage disequilibrium. The rare haplotype, which has a frequency of 0.20, is associated with increased HDL levels, decreased hepatic lipase activity, and more buoyant LDL particles. Hence, LIPC variants play an important role in determining the amount of small dense LDL, which is believed to be associated with increased risk of CAD.

The VLDL receptor (Table 1), a cell-surface receptor that binds with high affinity to VLDL and cholesteryl ester-rich βVLDL, but not to LDL, was first cloned in rabbits. It is expressed in the heart, skeletal muscle, and macrophages, with less expression in lung, kidney, placenta, pancreas, and brain. In contrast to the LDL receptor, to which it shows a very high degree of homology, it is not expressed in the liver. The human VLDL receptor protein is 96% homologous to the rabbit receptor. The human gene gives rise to two forms of the receptor, one with five

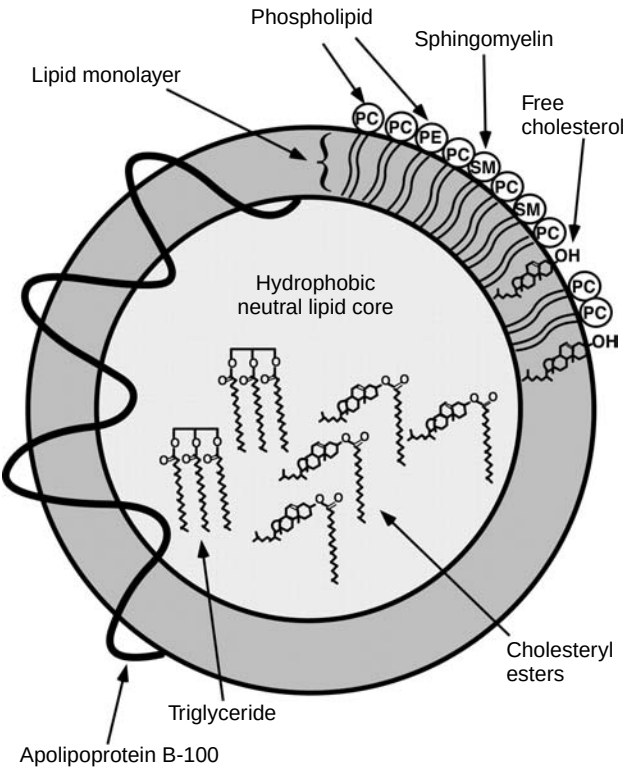

Fig. 3 Structure of low-density lipoprotein (PC, phosphatidylcholine; SM, sphingomyelin; PE, phosphatidylethanolamine.)

domains and one lacking the O-linked sugar domain. It is thought that the VLDL receptor functions in nonhepatic tissues in which fatty acids are actively metabolized. When compared to the LDL receptor gene, the VLDL receptor has an extra exon coding for a ligand binding domain. The VLDL receptor binds to triglyceride-rich lipoproteins *via* apoE, and not apoB, in concert with lipoprotein lipase (LPL), to which it also binds. Like the LDL receptor, it also binds to the receptor-associated protein (RAP). Its role in lipid metabolism has been hard to establish. Studies with VLDL receptor-deficient mice have been difficult to interpret, but it has now been shown that the VLDL receptor does indeed play a role in peripheral triglyceride hydrolysis. Another apoE receptor, homologous to both the LDL receptor and the VLDL receptor, is the low-density lipoprotein receptor-related protein 8 (LRP8) also known as the *apoE receptor-2* (APOER2). It binds with high affinity to apoE-rich βVLDL, and poorly to VLDL and LDL. LRP8 is expressed most highly in the brain and placenta. Both LRP8 and the VLDL receptor have neuronal signaling functions and play important roles during the development of the brain.

The LDL receptor is responsible for the uptake by the liver of VLDL remnants, the binding affinity of which is considerably

higher than that of LDL. The ligand on VLDL is apolipoprotein E (apoE).

1.2.3 LDL Catabolism

The liver removes approximately 70% of the LDL particles circulating in plasma, mainly *via* the LDL receptor: the remainder are removed by a nonspecific low-affinity process. On LDL particles the ligand for the receptor is apolipoprotein B (apoB).

The LDL receptor is present on the surfaces of most nucleated cells. The highest concentration of mRNA is in the adrenal gland followed by the corpus luteum of the ovary and then the liver. The half-life of LDLR mRNA in liver is 30 min.

The LDL receptor transcript, a 5.3-kb mRNA, has a long 2.5-kb 3′ UTR, which contains a dinucleotide AT repeat polymorphism. The gene has three 300-bp Alu repeats that may be a recombination hot spot.

The N-terminal 292 residues of LDLR, encoded by exons 1 to 6, make up the ligand binding domain with seven cysteine-rich tandem repeats. Each repeat of approximately 40 residues has three disulfide bonds with a coordinated Ca^{2+} ion. Exons 7 to 14 encode a domain that has a strong homology with the EGF precursor. This 400-residue region, which contains a six-bladed β-propeller motif, is central for endosomal release of the lipoprotein and for receptor recycling to the cell surface. A 50-residue O-linked sugar domain is encoded by exon 15. This is followed by a membrane-spanning domain of 22 residues encoded by exon 16 and the beginning of exon 17. The end of exon 17, and part of exon 18, code for a cytoplasmic C-terminus of 50 residues. This has a signal, NPVY, which is necessary for the incorporation of the receptor into clathrin-coated pits, and subsequent endocytosis. There is another motif between residues 790 and 839, which directs the receptor to the basolateral surface of hepatocytes. The remainder of exon 18 encodes the 3′UTR. At the low pH found in endosomes, the β-propeller region displaces lipoprotein particles bound to the receptor by means of a strong interaction with ligand-binding repeats 4 and 5.

LDL receptor activity is under sterol-dependent regulation, being tightly coupled to the extracellular level of LDL, with most of the regulation at the level of gene transcription. In addition to the major means of regulation by repression of transcription by sterols, the hepatic expression of the LDL receptor is stimulated by thyroid hormone. The proximal promoter of LDL receptor contains a sterol regulatory element (SRE) flanked on either side by Sp1 sites. For a high level of expression, the active form of SREBP acts in synergy with Sp1.

1.3 Disorders of Lipid Transport

1.3.1 Abetalipoproteinemia

The MTP protein and microsomal triglyceride transfer activity are absent in intestinal biopsies from patients with abetalipoproteinemia. In several subjects, this has been shown to be due to homozygous mutations in the gene for the large subunit of MTP. Abetalipoproteinemia is a rare autosomal recessive disease, which is characterized by the absence of apoB-containing lipoproteins in the plasma because of an inability to secrete triglyceride-rich lipoproteins from intestine or liver. The failure to absorb lipids results in a deficiency of the fat-soluble vitamins A, D, E, K and β-carotene. In the absence of dietary vitamin supplementation, this results in neurological and ophthalmological complications during childhood.

1.3.2 Hepatic Lipase Deficiency

Familial hepatic lipase deficiency, a rare condition, is marked by elevated levels of VLDL remnants and by the presence of triglyceride- and phospholipid-rich LDL and HDL. Compound heterozygosity for S267F and T383M is associated with high levels of cholesteryl-ester rich βVLDL in plasma, and premature atherosclerosis.

1.3.3 Familial Hypercholesterolemia

The discovery of the LDL receptor and its subsequent cloning has led to an understanding of the molecular basis of familial hypercholesterolemia (FH). FH is an autosomal codominant disorder characterized by elevated LDL levels, the presence of cholesterol deposits in skin and tendons (xanthomas) and in arteries (atheromas), and premature coronary heart disease (CAD). Clinical disease occurs earlier in men than in women. More extensive xanthomatosis and atherosclerosis are seen in homozygotes, who generally die of myocardial infarction (MI) by age 30. FH is caused by one of several hundred of the known mutations in the LDL receptor gene. The heterozygous frequency is 1 in 500 and the homozygous frequency 1 in a million. This monogenic disorder accounts for 4% of those individuals having type IIa hyperlipidemia as designated by the World Health Organization (those with LDL levels above the 95th percentile and with no other lipoprotein abnormalities).

A number of biochemical studies first led to the recognition that the LDLR receptor defects cause FH. Radiolabeled autologous LDL was found to have a lower fractional catabolic rate (FCR) when injected into FH subjects, when compared with normal individuals. Normally the half-life of LDL is 2.9 days whereas it is 4.7 days in heterozygous patients and 6.6 days in homozygotes. The slightly altered lipid composition of LDL prepared from FH patients did not affect metabolism of the LDL particles when injected into control subjects. Deficient high-affinity specific binding or internalization of LDL was seen in fibroblasts from FH homozygous patients. Kinetic studies indicated that, in addition to lower FCR, there is an increase in the rate of LDL production in FH. This is explained by a reduced rate of clearance of VLDL remnants by the LDL receptor, and a consequential increase in the proportion converted to LDL. The receptor normally clears VLDL remnants much more rapidly than it does LDL because of its higher affinity for apoE than for apoB. In a subset of FH patients the underlying mutation affects only LDL binding, and not that of VLDL remnants, and there is a more modest rise in plasma LDL levels.

Normally the liver takes up 70% of LDL particles through the LDLR, and the remainder through a less well-understood low-affinity nonspecific mechanism. This so-called *scavenger pathway* plays a more dominant role in the removal of LDL in FH. The scavenger pathway operates in macrophages in the artery wall where the cells take up modified LDL particles and can develop into the lipid-laden foam cells that are found in atheromas. Therefore, in FH, the longer residence time of LDL, which leads to more oxidation and other modifications of the LDL particles, is a factor underlying the observed premature CAD. This is especially true of homozygous patients.

A diagnosis of FH has come to be synonymous with LDL receptor gene defects, and many individuals so diagnosed can be

shown to have LDL receptor mutations. A few patients, though, do not, and defects at other loci can cause a phenocopy disorder. Where an LDL receptor mutation is known to be present, it has been occasionally observed that other loci can modulate its severity.

Five classes of LDL receptor defect have been established at the molecular level. Class 1 is null alleles caused by different types of mutations that result in the absence of protein synthesis and account for 17% of FH cases. Class 2, comprising the majority of cases (54%), is due to mutations that cause defective transport of the receptor from the ER to the Golgi apparatus, with most mutations occurring in repeats 4 and 5 of the ligand binding domain. Receptor molecules that do not bind normally even though they are found on the cell surface constitute class 3. Most of these are due to in-frame mutations that affect the EGF precursor homology or ligand binding domains. Class 4 mutations are found in the cytoplasmic domain and result in receptors that fail to endocytose due to an inability to cluster in clathrin-coated pits. Class 5 comprises another common set of mutations (22%). Here, binding and endocytosis are normal, but there is a failure to release the LDL particle in the endosomal compartment. Many of these are due to mutations in the β-propeller domain.

The *LDL* receptor gene has been targeted in mice using homologous recombination to produce an FH model. These $Ldlr^{-/-}$ mice have a twofold increase in cholesterol levels due to substantial increases in IDL and LDL, but do not develop substantial atherosclerosis. The addition of cholesterol to the diet accentuates the phenotype. The phenotype in these mice can be temporarily reversed using a recombinant LDL receptor adenovirus.

1.3.4 Familial Ligand-defective Apolipoprotein B-100

As stated above, FH accounts for only 4% of cases with elevated LDL and much effort has been made to identify other genetic loci that are involved, besides the LDL receptor. Autologous LDL, prepared from non-FH hypercholesterolemic patients, was often found to have a low FCR. It was surmised that this could be due to the presence of ligand-defective LDL, and patients were identified whose LDL showed poor binding to LDL receptors on cultured fibroblasts. Because the lipid composition of the LDL was unaffected, it was concluded that an abnormality in apoB due to a genetic defect was responsible. This autosomal codominant disorder is called *familial ligand-defective apoB* (FLDB).

After separation of the two LDL allotypes from one patient with FLDB the particles were shown to have more than 90% reduced affinity for the receptor. In three families, an apoB R3500Q mutation was shown to associate with FLDB. The R3500Q mutation, with a frequency in the general population of 1 in 600, is much more common among patients attending lipid clinics. The hypercholesterolemia seen with R3500Q heterozygotes is milder compared with that seen in FH heterozygotes, and the presence of xanthomas and premature CAD is less frequent. The few homozygotes that have been reported are more severely affected than heterozygotes. Two other FLDB mutations have been reported. The R3531C variant, which is less common than R3500Q, results in milder hypercholesterolemia, consistent with the finding that 3531C LDL particles retain 27% of normal binding affinity for the receptor. FH/FLDB patients with compound heterozygosity for apoB R3531C and LDL receptor P663L are more severely

affected than those with the FH mutation alone. The other known mutation, R3500W, is similar in its effects to the R3500Q disorder.

An interesting mechanism has been proposed to explain why the LDL particles are defective in FLDB. The topological arrangement of the apoB polypeptide on the LDL surface, mapped using monoclonal antibodies, displays a "ribbon and bow" structure. Mutations in the region around residues 3500 to 3531 cause conformational changes that disrupt the topology and mask the receptor binding site. A direct interaction between W4369 and R3500 probably helps to preserve the bow arrangement.

1.3.5 Familial Hypobetalipoproteinemia

Individuals with familial hypobetalipoproteinemia (FHB) have apoB mutations that result in the production of C-terminally truncated variants. The length of the truncated protein is monotonically related to the diameter and buoyant density of the lipoprotein formed. Numerous such mutations have been reported, many of which are small deletions. Four mutations that were predicted, B-2, B-9, B-25 and B-29 proteins, were not detected in plasma. B-31 was found only in the HDL density interval, whereas, B-37 was found in HDL, LDL, and VLDL intervals. Those longer than B-46 were found mainly in LDL and VLDL. Three mutants, B-75, B-87 and B-89, had increased binding to the LDL receptor. In one especially interesting case of FHB, a frameshift mutation created a sequence of eight consecutive adenines, caused by a single base pair deletion in exon 26 that predicted the formation of a truncated apoB species. However, this allele was shown to produce full-length apoB-100 in addition to the truncated protein. This was due to an additional adenine being inserted in a portion of the mRNA molecules during transcription, leading to restoration of the correct reading frame.

The phenotype associated with FHB is variable. Simple heterozygotes have low levels of LDL cholesterol but are usually clinically asymptomatic. At its most severe, with the presence of 2 null alleles, FHB is similar to the recessive disorder abetalipoproteinemia (see Sect. 1.3.1 above). Mouse models for FHB have been produced by targeting the *apoB* gene using homologous recombination.

1.3.6 Familial Combined Hyperlipidemia

An independent phenotype termed *familial combined hyperlipidemia* (FCH) was originally described, more than 30 years ago, in families of patients who had survived heart attacks. FCH was originally thought to be a dominant disorder, but more recently it has been shown to obey a multigenic mode of inheritance. It displays high penetrance and is present at a frequency of 1 to 2% in the general population. It is thought to play a role in 15% of cases of premature CAD. Any of three phenotypes can occur, and in any one individual the phenotype can vary over time. Because the genetic basis of FCH is not well understood, it is still identified by phenotype. Elevation of either VLDL or LDL level is seen in some subjects. However, levels of both these lipoproteins are elevated in most individuals with FCH. LDL cholesterol levels are almost always above 100 mg dL^{-1} and IDL is usually raised. High levels of apoB are a hallmark of this disease. Patients also tend to have HDL and LDL particles that are smaller, denser, and more atherogenic. There is much evidence that overproduction of VLDL, by about two-fold, is responsible for the observed dyslipidemia. Such overproduction

is thought to be due, in part, to an increase in flux of FFA to the liver. Obesity and insulin resistance, with elevated FFA, are often seen in cases of FCH. The molecular causes of FCH have proved difficult to decipher. Individual linkage studies of FCH families have identified genetic loci with high lod scores, but when data were combined from several studies the associations were considerably diminished, suggesting that FCH is caused by defects at alternate loci and may be multigenic. Thus, FCH is a genetically heterogeneous disorder, and there is evidence to support the involvement of several candidate genes. It has been suggested that the lipoprotein lipase gene plays a modifier role. A number of individuals with FCH have low LPL activity. A major candidate for FCH has been the ApoA-I/ApoC-III/ApoA-IV/ApoA-V locus at chromosome 11q23.3. Some, but not all, linkage studies have identified this locus. Other gene candidates are LCAT (see Sect. 2.3), CETP (see Sect. 2.5), manganese superoxide dismutase (an enzyme involved in lipid oxidation), intestinal fatty acid–binding protein 2 (FABP2), and tumor necrosis factor receptor superfamily, member 1B (TNFRSF1B).

1.3.7 Autosomal Recessive Hypercholesterolemia

The molecular basis of the rare disorder, autosomal recessive hypercholesterolemia (ARH), has recently been established. Defects in the LDL receptor and *apoB* genes were ruled out by cosegregation studies. ARH was mapped to chromosome 1p35–p36. This disease was originally referred to as *pseudo-homozygous FH*. The mean plasma cholesterol level of 29 patients was 561 mg dL^{-1}. The probands in two Sardinian families with this rare disorder had clinical symptoms that were analogous to those seen in the homozygous form of FH. One striking difference between ARH and FH was that fibroblasts from ARH patients bound and endocytosed LDL normally even though kinetic studies had shown decreased FCR for LDL. However, the degradation of LDL by monocyte-derived macrophages and EBV-lymphocytes isolated from ARH patients was defective. Normal processing of LDL was seen when the lymphocytes were transfected with the normal *ARH* gene, which codes for a receptor adapter protein. The ARH protein contains a phosphotyrosine binding domain (PTB). These domains bind to NPXY sequences found in the cytoplasmic tails of members of the LDL receptor family. The NPXY sequence targets receptors to clathrin-coated pits. One FH mutation, Y807C (in the *LDLR* gene) which results in defective internalization of the LDL receptor, disrupts the NPXY consensus sequence. Besides binding to the LDL receptor, the ARH protein binds to the receptor adapter protein AP-2, and the clathrin heavy chain. In some signaling proteins, PTB domains are essential for signal transduction. The ARH PTB domain is similar to that found in the two endocytic proteins, numb and Disabled-2 (Dab-2). It is likely that these two proteins, which are expressed in fibroblasts, can substitute for the ARH protein in those cells.

1.3.8 Autosomal Dominant Hypercholesterolemia

Besides FH and FLDB a third type of autosomal dominant hypercholesterolemia has been reported. In a large kindred with hypercholesterolemia, in which the LDL receptor and *ApoB* genes had been excluded, linkage was found to chromosome 1q32. The gene responsible for the disorder is *PCSK9*, which codes for the proprotein convertase subtilisin/kexin type 9.

This protein is mainly expressed in liver, kidney, and intestine. The missense mutations detected were thought to cause a gain of function. Its overexpression in mice results in a doubling of total serum cholesterol with no change in HDL. The related protease, PCSK8, cleaves and activates SREBP transcription factors.

1.3.9 Chylomicron Retention Disease and Anderson Disease

An inherited disorder of severe fat malabsorption was reported in 1961 and termed *Anderson disease*. A similar disorder, chylomicron retention disease (CMRD) was reported later, and differed from Anderson disease on the basis of partitioning between membrane-bound and cytosolic compartments of enterocytes, which accumulate large amounts of lipid in both cases. A variant of CMRD is associated with Marinesco–Sjogren syndrome, a neuromuscular disorder. All these diseases are characterized by failure to thrive in infancy, fat-soluble vitamin deficiency, low plasma cholesterol, and the absence of circulating chylomicrons. Hepatic secretion of VLDL appears to be normal. Recently, a coat protein complex, COPII, which facilitates transport from ER to Golgi, has been shown to be required for VLDL and chylomicron secretion. Mutations in the gene *SARA2*, which codes for the protein Sar1B, a GTPase component of COPII, have been shown to underlie Anderson disease and the two forms of CMRD.

1.3.10 Sitosterolemia

Sitosterolemia, also referred to as *phytosterolemia*, is a monogenic disorder that causes elevated levels of LDL. This rare autosomal recessive disease was first described over thirty years ago, and is characterized by highly elevated plasma levels of plant sterols, notably β-sitosterol and campesterol. By 1998, there were only 45 known cases of sitosterolemia. Plant sterols also accumulate in tissues, and are found in atherosclerotic plaque and in xanthomas, where they account for 20% of the sterol content. A characteristic of sitosterolemia is that during the first decade of life patients develop both tendon and tuberous xanthomas. Early CAD and MI are often associated with this disease.

In patients with sitosterolemia, although the absorption of cholesterol is only slightly higher than normal, that of plant sterols is dramatically elevated, from a normal 5% to between 15 and 63%. The underlying molecular basis for sitosterolemia remained a conundrum for decades. Both decreased hepatic *de novo* synthesis of cholesterol and HMGCoA reductase activity were seen. Conversion of cholesterol to bile acids was normal even though cholesterol 7α-hydroxylase activity was lower. There have been contradictory reports as to whether the cholesterol content of liver is affected. Considering that the increased absorption of sterols is insufficient to explain the levels of phytosterols in plasma, an important observation was that overall sterol excretion, particularly of plant sterols, into bile was low.

It has now been determined that mutations in one or the other of the ABC transporter genes *ABCG5* and *ABCG8* cause sitosterolemia. Mutations in the *ABCG5* gene were detected in only some of the patients. An inspection of the *ABCG5* locus at chromosome 2p21 revealed an additional ABC transporter, *ABCG8*. Mutations in the *ABCG8* gene account for the other cases of sitosterolemia. Both *ABCG5* and *ABCG8* are half transporters, each gene has 13 exons and codes for proteins (sterolin-1 and sterolin-2) each with one nucleotide binding fold (NBF) and a single transmembrane domain (TM) comprised

of 6 membrane-spanning α-helices. The genes are orientated in a head-to-head arrangement with the transcription start sites 140-bp apart. *ABCG5* and *ABCG8* interact to form a heterodimer in the ER, constituting a functional complex. In all cases, homozygous or compound heterozygous mutations were seen in either *ABCG5* or *ABCG8*. The presence of a single normal copy of each gene rules out the disease.

It seems that *ABCG5* and *ABCG8* have evolved as sterol transporters, which are expressed to the highest degree in liver and small intestine. When coexpressed they are found with apical plasma membrane markers. It is thought that an ER retention motif localizes these proteins in the ER until heterodimerization occurs, the motif is masked, and transportation to the Golgi proceeds. When these genes were overexpressed in mice, the animals had a decreased rate of net absorption of dietary cholesterol and an increased rate of excretion of sterol into bile. Contrary to what is seen in sitosterolemia, hepatic cholesterol synthesis was increased in these mice. With *ABCG5/ABCG8* knockout mice there was a higher fractional absorption of plant sterol from the diet and a high concentration of sitosterol in plasma.

2 Reverse Cholesterol Transport

The hepato-biliary system is the major means by which cholesterol is excreted. Whereas most of the bile salts and free cholesterol in bile are reabsorbed by the ileum and returned to the liver, a fraction is lost in the feces. Some of this is derived from cholesterol on remnant particles or from hepatic synthesis of cholesterol. The rest is from cholesterol returned to the liver from peripheral tissues by an active, specific mechanism. This latter process is termed *reverse cholesterol transport*. This is initiated in the extravascular compartment in cells of extrahepatic tissues, where free cholesterol effluxes with the aid of ATP-binding cassette transporter A1 (ABCA1) and the class B scavenger receptor SR-BI (see Sect. 7.1). In the case of ABCA1, especially in macrophages, an extracellular primary acceptor, pre-*beta*-1 HDL, comprised of phospholipid and two copies of apoA-I, takes up cholesterol. Pre-*beta*-1 HDL has an apparent molecular weight of 67 kDa. Some evidence also indicates that a lipoprotein particle with gamma electrophoretic mobility, containing only apoE, may also play a role. SR-BI facilitates the diffusion of free cholesterol between the plasma membrane and *alpha* HDL particles. Of the total efflux of cholesterol from peripheral tissues, possibly a greater amount is transferred by the SR-BI facilitated process than by the ABCA1 mediated one.

The cholesterol transferred to pre-*beta* HDL and *alpha* HDL is now a substrate for LCAT. The resulting cholesterol esters (CE) accumulate to form a hydrophobic core of larger pre-*beta* HDL particles and ultimately appear in large mature spherical HDL particles of alpha mobility. This CE can be taken up directly by liver *via* the scavenger receptor SR-BI, or transferred to LDL, VLDL, VLDL remnants, or chylomicron remnants by a process that is mediated by CETP.

Several subspecies of HDL are revealed by electrophoresis in nondenaturing gels and by isoelectric focusing. Some of these contain apoA-I but not apoA-II, and are referred to as *LpAI*. Others contain both of these proteins and are referred to

as *LpAI/AII*. LpAI and LpAI/AII play metabolically different roles, especially with regard to reverse cholesterol transport. Both are able to promote the efflux of cholesterol from the plasma membrane, but LpAI appears to mediate the translocation of intracellular cholesterol to the cell surface. LpAI/AII is the preferred substrate for hepatic lipase. The level of LpAI in plasma is a better measure of the protective antiatherogenic potential of HDL than LpAI/AII is.

2.1 ATP-Binding Cassette Transporter A1

Phospholipids can move readily ("flip-flop") across the bilayer of certain membranes such as those of the endoplasmic reticulum. However, in the plasma membrane (and also late Golgi, endosomal and bile canalicular membranes) such passive equilibration is restricted and requires an ATP-dependent process involving members of a superfamily of ATP-binding cassette transporter proteins. Unlike the half transporters, ABCG5 and ABCG8, referred to above (Sect. 1.3.10), ABCA1 is a full transporter with 2 TM and 2 NBF domains. ABCA1 is highly expressed in liver where it is involved in the secretion of nascent HDL. In macrophages, particularly in the artery wall, ABCA1 is thought to be important for the removal of a toxic pool of cholesterol, thereby preventing apoptosis.

2.1.1 Tangier Disease

Tangier disease is a rare autosomal disorder, thought originally to be recessive, but now accepted to be codominant, and is characterized by the almost complete absence of HDL. What little HDL is present is pre-*beta* HDL. CE accumulates in many tissues, notably the intestine and reticuloendothelial system, with the accumulation of orange-pigmented lipids in the tonsils and rectal mucosa. Peripheral neuropathies and a syringomyelia-like disorder accompany the disease, together with premature CAD. Fibroblasts cultured from affected patients showed very low efflux of cholesterol to acceptor apoA-I compared to normal cells. Efflux of cholesterol to HDL was little affected while that of phospholipid was greatly decreased. After numerous candidate genes were ruled out, linkage and subsequent studies confirmed that homozygous or compound heterozygous mutations in *ABCA1* underlie all cases of Tangier disease. Heterozygous *ABCA1* mutations account for some cases of primary hypoalphalipoproteinemia, though these are probably less than 10% of cases. The ABCA1 defects that cause Tangier disease appear to result primarily in impaired phospholipid efflux, which leads to secondary impairment of cholesterol efflux.

2.2 Caveolin

All cells actively acquire and efflux cholesterol to the retrieval pathway, to maintain homeostasis in plasma membranes. The process involves movement of cholesterol to the plasma membrane and is probably facilitated by the caveolin proteins coded by three caveolin genes, *CAV1, CAV2*, and *CAV3*. There are two isoforms of caveolin-1, *alpha* and *beta*, because there are two transcription start sites. The human *CAV1* and *CAV2* genes each have 3 exons and are located 18-kb apart at chromosome 7q31.2. The caveolin family of scaffold proteins, notably the 22-kDa caveolin-1, is found in caveolae, which are clathrin-free invaginations of the plasma membrane, rich in free cholesterol and sphingolipids.

Multiple caveolin molecules form highly ordered assemblies within caveolae. Concentrated in the caveolae are components of lipid signaling pathways, transmembrane receptor kinases, protein kinase C, protein kinase A, adenyl cyclase, and MAP signaling intermediates. In addition, the scavenger receptors SR-BI (SCARB1) and CD36 (SCARB3) are also located in caveolae. The caveolin polypeptides, which bind free cholesterol, are embedded in the plasma membrane with both the N- and C-termini in the cytoplasm. Short, approximately 20-residue, motifs in the caveolins bind to and recruit free cholesterol and proteins.

2.3 Lecithin Cholesterol Acyl Transferase

The enzyme LCAT rapidly esterifies the free cholesterol that is transferred to pre-*beta*-1 HDL from plasma membranes. The cholesteryl esters are transferred to HDL with *alpha* electrophoretic mobility. HDL are also acceptors of free cholesterol from other lipoproteins, notably LDL, and this transfer may be LCAT dependent.

The *LCAT* gene (chromosomal location 16q22.1) has six exons. Mature LCAT, secreted by the liver, is a 60-kDa, 416-residue glycoprotein with a hydrophilic C-terminus. There are regions with homology to apoE, LPL, and hepatic lipase (LIPC). Cofactor activation of LCAT has been attributed to apoA-I and apoA-IV.

A number of mutations causing LCAT deficiency have been reported, which result in decreased levels of HDL- and LDL-cholesterol esters. There is also abnormal architecture of virtually all lipoprotein species. These mutations are dispersed in various regions of LCAT, indicating that there are several important domains. Two mutations have been shown to cause cases of Fish-eye disease, which is associated with markedly reduced levels of HDL cholesterol. In one of these cases (T123I), the mutant protein has been shown to be capable of esterifying LDL cholesterol but not HDL cholesterol.

Studies using *in vitro* mutagenesis have shown that, of the four potential N-glycosylation sites, Asn84 is important for full enzymatic activity, though not for intracellular processing, and Asn272 is essential for secretion. The LCAT variant R158C appears to be a natural polymorphism.

2.4 Phospholipid Transfer Protein

Although the liver secretes nascent discoidal HDL, HDL particles are also formed in plasma, with the apolipoproteins and phospholipids being derived from triglyceride-rich lipoproteins (chylomicrons and VLDL). The supply of phospholipid for the formation of HDL particles is mediated by PLTP. Studies on transgenic animals have provided evidence that pre-*beta*-1HDL particles are generated by the activities of PLTP. PLTP-knockout mice have reduced levels of HDL, owing to decreased production and increased catabolism. As well as being important for pre-*beta*-1-HDL formation, PLTP plays an important role in remodeling larger HDL species, by converting smaller HDL_3 particles to larger HDL_2 particles. In addition to PLTP, hepatic lipase too can metabolize HDL_2 to particles of smaller diameter, and plays a role in pre-*beta*-HDL production.

While PLTP is expressed in many tissues, the source for most of the circulating protein is liver and adipose tissue. There is evidence that the *PLTP* gene is upregulated at the transcriptional level by PPARα, LXR, and possibly by FXR. PLTP

is a highly glycosylated protein and has an apparent molecular mass of 81 kDa, with 476 amino acid residues. The human gene (13.3 kb) at chromosome 20q13.12 is composed of 16 exons and is organized in a manner similar to that of the gene for CETP. These two genes appear to have evolved from a common ancestor. PLTP has homology with (in addition to CETP) lipopolysaccharide-binding protein and neutrophil-bactericidal-permeability-increasing protein. PLTP is nonspecific in its transfer capabilities, being able to transfer most phospholipids in addition to diacylglycerol. It has also been shown to promote the exchange transfer of *alpha*-tocopherol (vitamin E) among lipoproteins.

2.5 Cholesteryl Ester Transfer Protein

CETP is responsible for the transfer of the cholesteryl esters formed by LCAT from HDL to the apoB-containing lipoproteins and facilitates the reverse transfer of triglycerides. The HDL triglycerides are subsequently hydrolyzed by hepatic lipase. CETP is an important factor in determining the ultimate size profile of LDL particles. It is a 70-kDa hydrophobic glycoprotein. Most of the circulating CETP is associated with the HDL fraction of plasma. The human gene, on chromosome 16q13, comprises 16 exons spanning 25 kb and codes for a 493-residue prepeptide. It is expressed in the liver, small intestine, adrenal glands, spleen, and adipose tissue. There is evidence for the presence of an alternatively spliced CETP mRNA species, lacking exon 9, in certain tissues and for only limited secretion of the corresponding protein.

Individuals with homozygous familial CETP deficiency have high levels of large, apoE-rich HDL particles and low levels of LDL cholesterol, but do not suffer from premature atherosclerosis. These subjects have markedly decreased catabolism of apoA-I and apoA-II. Two particular mutant alleles, D442G and a splicing defect in intron 14, account for 10% of the variance of HDL in the Japanese population. These alleles are relatively common, with heterozygosities of 7 and 2% respectively. Analysis of the CETP gene promoter revealed eight major haplotypes among Japanese subjects, with one being associated with lower CETP levels and higher HDL levels in plasma.

High-level expression of CETP in transgenic mice (a species in which plasma CETP is normally absent) is associated with an increased rate of catabolism of apoA-I. This leads to a decrease, both in the level of HDL cholesterol and in HDL particle size. These animals develop atherosclerosis more easily than control mice.

In addition to the transfer of HDL cholesterol esters by CETP to other lipoproteins and subsequent uptake by the liver, there are two other minor routes for cholesterol ester clearance. Firstly, apoE-rich HDL particles are endocytosed by hepatic receptors that recognize apoE, and secondly there is selective hepatic uptake of HDL cholesterol esters, but not apoA-I, by SR-BI, a process involving hepatic lipase.

2.6 Endothelial Lipase

Endothelial lipase (LIPG) is a recently recognized member of the lipase gene family, which includes lipoprotein lipase (LPL), hepatic lipase (LIPC), pancreatic lipase (PNLIP), phosphatidylserine-specific phospholipase A1 (PLA1A), and pancreatic lipase-related proteins 1 and 2 (PNLIPRP1 and PNLIPRP2). It has 44 and

41% homology with LPL and hepatic lipase, respectively. In particular, the active site catalytic triad, Ser169, Asp193 and His274, and heparin binding sites are conserved. A notable exception is the lid region responsible for substrate recognition. Unlike these two lipases it does not hydrolyze triglycerides; rather, its preferred substrate is phosphatidylcholine releasing the sn-1 acyl chain. It is expressed in a number of tissues, notably liver, thyroid, smooth muscle, bronchial epithelial cells, macrophages, and placenta. The human gene is at chromosome 18q21.1, and is 31 kb-long with 10 exons. The precise physiological function of endothelial lipase is not clear at present. However, evidence points to a role in HDL metabolism. Endothelial lipase knockout mice (*Lipg*−/−), and mice treated with an antiendothelial lipase antibody, had increased levels of HDL. The overexpression of this lipase in mice led to decreased HDL. Endothelial lipase hydrolyzes the main lipid constituent of HDL, phosphatidylcholine. It is not clear whether this occurs before or after binding of HDL to its receptor, SR-BI, that is, before or after much of the cholesterol content of the HDL particle has been transferred.

3 Cholesterol Metabolism

3.1 Synthesis

Cholesterol is required primarily to maintain the structure and characteristics of cellular membranes. It is also required as a precursor for steroid hormones and bile acids. Much of the cholesterol in cells is obtained from circulating LDL by uptake through the LDL receptor, and by a nonspecific pathway. However, most tissues, in particular, the liver, can synthesize cholesterol.

The isoprenoid biosynthetic pathway, of which cholesterol is the end product, can also supply cells with other compounds such as dolichol and ubiquinone. An isoprenoid intermediate in the cholesterol synthetic pathway, 15-carbon farnesyl pyrophosphate, and the related isoprenoid 20-carbon geranylgeranyl pyrophosphate, are used for the posttranslational prenylation of certain proteins (notably G proteins) at cysteine residues through thioether bonds. Prenylation, essential for membrane localization of a number of proteins, is catalyzed by several prenyltransferases. The gene products that are prenylated include many that are involved in signal transduction pathways. Examples are the ras, Rho, and Rab GTPases, Lamin A and B, and phosphorylase kinases. Cholesterol itself is covalently attached to the N-terminus of Shh, the product of the sonic hedgehog homolog gene (*SHH*) *via* a catalytic activity residing in the C-terminus. The Shh-cholesterol adduct is thereby restricted to zones within the developing cells in which Shh acts.

3.1.1 HMGCoA Synthase and Reductase

The initial precursor for cholesterogenesis is acetyl CoA. After conversion to acetoacetyl CoA, two key regulatory enzymes, hydroxymethylglutarylcoenzymeA (HMG-CoA) synthase and HMG-CoA reductase are responsible, sequentially, for the formation of mevalonic acid. This is the substrate for the formation, *via* geranyl pyrophosphate and farnesyl pyrophosphate, of squalene. Cyclization of squalene produces the first sterol, lanosterol. This is modified further by a series of enzymes forming cholesterol. The enzymatic reduction of

HMG-CoA to mevalonic acid, which utilizes 2 molecules of NADPH, represents the first and rate-determining step in isoprenoid and cholesterol biosynthesis.

The genes for HMG-CoA synthase, HMG-CoA reductase, farnesyl pyrophosphate synthase, squalene synthase (all regulatory enzymes), and the LDL receptor are coordinately regulated by the amount of cholesterol available to the cell. As well as having unique enhancers, they share common cis-acting sterol regulatory elements (SREs) in their promoters.

The gene for HMG-CoA synthase (HMGCS1) on chromosome 5p12 spans 23 kb and has 9 exons, and the gene for HMG-CoA reductase (HMGCR) on chromosome 5q13.3 spans 24 kb and has 19 exons. Seven membrane-spanning segments, in the 339-residue N-terminus, that anchor the 97-kDa HMG-CoA reductase protein to the smooth ER, act as a sterol-sensing domain, which controls the regulation of enzyme degradation. The catalytic site is located in the 548-residue C-terminus, which projects into the cytosol.

Regulation of both enzymes is *via* negative feedback control by sterol and nonsterol products of mevalonic acid. In addition to transcriptional and translational regulation, HMG-CoA reductase is regulated posttranslationally by phosphorylation of the catalytic domain with AMP-activated protein kinase. This latter mode of cross regulation, which is independent of the feedback control, coordinates isoprenoid synthesis with cellular energy balance.

Several inhibitors of HMG-CoA reductase, namely, the statin class of drugs, cause hepatic induction of LDL receptor expression. This leads to increased clearance of VLDL remnants and LDL, which results in a highly effective lowering of plasma levels of LDL cholesterol. What seemed to be evidence for the existence of an additional gene, coding for a second HMG-CoA reductase isoform expressed in peroxisomes, has been discounted. It is now thought to be merely the result of alternative targeting of the ER enzyme.

3.1.2 Smith–Lemli–Opitz Syndrome

First described in 1964, Smith–Lemli–Opitz Syndrome (SLOS) is a rare developmental disorder characterized by high levels of 7-dehydrocholesterol and low levels of cholesterol in plasma. It is caused by a defect in the gene for the cholesterol biosynthetic pathway enzyme, 7-dehydrocholesterol reductase (DHCR7). This enzyme has a TM sterol-sensing domain similar to that found in HMG-CoA reductase, SREBP cleavage activation protein (SCAP), the Niemann–Pick type C1 protein (NPC-1) and Patched (the Shh receptor). Patients have multiple congenital malformations that vary in clinical severity. The lack of sufficient cholesterol interferes with Shh signaling which is crucial for normal embryogenesis. Congenital disorders less common than SLOS, affecting other sterol modification enzymes, have also been reported.

3.1.3 Acyl CoA: Cholesterol Acyltransferases

Acyl-CoA: cholesterol acyltransferase (ACAT) catalyzes the long-chain fatty acyl esterification of cholesterol and is crucial for the regulation of the level of free cholesterol within cells. Cholesterol esters are used for chylomicron and VLDL production. In steroidogenic tissues, such as the adrenal gland, stored CE is used as the precursor for steroid hormone. Free cholesterol released after uptake of LDL is esterified by ACAT and stored in cytoplasmic lipid droplets. This process is important in the macrophages and smooth

muscle cells of the artery wall where it can lead to the formation of CE-loaded foam cells, features of early atherosclerotic lesions. ACAT activity in liver is regulated by the availability of unesterified cholesterol. Hepatic ACAT activity helps to maintain the free cholesterol concentration within the liver and plays a role in regulating the secretion of cholesterol, as a constituent of VLDL, into plasma. Hence, ACAT is involved in the regulation of apoB secretion from the liver. It also plays a role in regulating the efflux of free cholesterol into bile. ACAT activity in the liver is decreased in patients with cholesterol gallstones, and this decrease contributes to the increased availability of free cholesterol for bile secretion.

ACAT is a membrane-bound protein of the endoplasmic reticulum, which initially proved difficult to purify to homogeneity. Two human genes with ACAT activity have been identified. Because the gene for acetoacetyl-CoA thiolase has been given the symbol ACAT, the acyl-CoA: cholesterol acyltransferases have been renamed as *sterol O-acyltransferases*, SOAT1 and SOAT2. The *SOAT1* gene is on chromosome 1q25.2, spans 61 kb, has 16 exons, and codes for a 550-residue protein. SOAT2 is on chromosome 12q13.13, spans 21 kb, has 15 exons, and codes for a 522-residue protein. SOAT2 (ACAT2) was discovered on the basis of high homology with SOAT1 from codon 102 to the C-terminus. The rare phenomenon of trans-splicing appears to be the origin of a minor 4.3 kb ACAT1 mRNA, which has an additional exon from chromosome 7.

There has been confusion regarding the sites of expression of these two genes. *ACAT1* is more widely expressed than *ACAT2*, which is mainly expressed in small intestine at the apical region of the villi and in liver. In humans, ACAT1 activity dominates that of ACAT2 in human liver. However, in ACAT1 knockout mice hepatic CE levels and synthetic rates were normal. The only tissues affected were the adrenals and macrophages. ACAT2 disruption in mice resulted in very low ACAT enzyme activity in liver and intestine. These animals showed resistance to diet-induced hypercholesterolemia and to the formation of cholesterol gallstones because of the decreased intestinal cholesterol absorption. Whatever the relative expression levels of ACAT1 and ACAT2 are in human liver, they are probably different from those in mouse liver.

Both enzymes are membrane bound, each with five TM domains, being found in microsomal cell fractions. It has been proposed that CE that is formed within the ER bilayer by ACAT is processed in one of two possible ways. It can either be directed as lipid droplets into the cytosol or, in cells producing lipoproteins, transferred to apoB in the ER lumen by MTP.

3.2 Catabolism: Bile Acid Metabolism and Disorders

The main bile acids, taurine or glycine conjugates of cholic, deoxycholic, and chenodeoxycholic acids, are synthesized from cholesterol in the liver. This is the primary means by which cholesterol is catabolized, the hepato-biliary system being the main pathway for cholesterol excretion. The first, rate-controlling step of the major pathway of bile acid synthesis, is the formation of 7α-hydroxycholesterol. This is catalyzed by cholesterol 7α-hydroxylase (CYP7A1), a liver-specific, cytochrome P-450-dependent, mixed function oxidase. In humans, the pathway leads to two

primary bile acids, cholic acid and chenodeoxycholic acid.

CYP7A1, located in the smooth ER, is subject to feedback regulation at the transcriptional level, with mRNA levels decreased by high levels of bile acids and positively correlated with the availability of cholesterol. *CYP7A1* gene transcription is subject to feedback regulation by bile acids through a bile acid response element (BARE), and is positively correlated with the availability of free cholesterol. CYP7A1 mRNA levels are regulated by thyroid hormone and insulin. There have been numerous studies on the regulation of CYP7A1, and the promoter region has been extensively analyzed. However, there remain a number of unresolved questions, especially with regard to the human gene, and a number of contradictions exist among the studies. It has been emphasized that there are important differences in the types of bile acids among species and the response of CYP7A1 to high-cholesterol diets, especially between humans and widely studied rodents. Nevertheless, there is a current, commonly accepted, model that attempts to explain the regulation of CYP7A1 and bile acid homeostasis, albeit based on rodent data.

The concentration of oxysterols, which are activating ligands for the nuclear receptor LXR-α (NR1H3), is thought to reflect the level of cell cholesterol. High levels of cholesterol, and hence of oxysterols, increase the transcription of CYP7A1, basal transcription being the result of the binding of CPF (or LRH-1, the mouse homolog) encoded by the *NR5A2* gene, and of HNF4-α, to the promoter. There is evidence that HNF-1 binds to and activates the CYP7A1 promoter in humans but not in the rat. A number of bile acids, notably chenodeoxycholic acid (CDCA), are activating ligands for the nuclear receptor FXR (NR1H4). FXR (see Sect. 5.4) is an orphan nuclear hormone receptor expressed within hepatocytes and enterocytes in the terminal ileum. It is thought that FXR regulates CYP7A1 indirectly because no consensus binding site has been identified in the promoter. FXR upregulates the gene for SHP (small heterodimer partner), and the SHP protein binds to and inhibits the activation of CPF (LRH-1) and HNF4-alpha. This model holds up better in rodents than in humans. Unlike its action on the rat cyp7a1 promoter, LXR has only a small effect on the human promoter to which it binds only weakly. It has been reported that FXR only modestly increases SHP transcription in human HepG2 cells. SREBP can activate the human SHP promoter, but not that of the mouse gene. The transcriptional coactivator PGC-1α activates *CYP7A1* gene transcription. *CYP7A1* is repressed by the c-Jun N-terminal kinase pathway, a process that involves fibroblast growth factor 19 (FGF-19), which is regulated by FXR.

The human gene for cholesterol 7α-hydroxylase, which is 10-kb long, lies at chromosome 8q12.1, contains 6 exons, and codes for a 504-residue peptide, which contains sterol and heme binding sites. Overexpression of the gene in hamsters caused a large decline in LDL, and transgenic mice had resistance to diet-induced gallstone formation and to atherosclerosis. There have been conflicting reports as regards the effect on lipid levels in plasma, and the phenotype in general, in knockout mice. In one colony, most homozygous animals died within 18 days of birth unless mothers were supplemented with vitamins and cholic acid was added to the diet. There were no changes in serum lipid levels. The phenotype consisted of fat malabsorption, abnormal lipid excretion, and skin and behavioral abnormalities. A second colony

survived on an unsupplemented diet, had elevated serum cholesterol, and had a proatherogenic phenotype. CYP7A1 polymorphisms, notably an SNP at −204 in the promoter, have been associated with differences in LDL-cholesterol levels. Screening of hypercholesterolemic patients, using a differential DNA melting technique, revealed several rare CYP7A1 variants.

Cholesterol 7α-hydroxylase initiates the classical or neutral bile acid synthesis pathway. An additional pathway, the alternative or acidic pathway, does not involve CYP7A1 and is initiated by mitochondrial sterol 27-hydroxylase (CYP27A1). Subsequently, a microsomal enzyme, oxysterol 7α-hydroxylase (CY7B1), introduces a 7α-hydroxyl group. The rest of the pathway is similar to the neutral pathway. However, chenodeoxycholic acid is the main product here, and not cholic acid.

3.2.1 Cholesterol 7α-hydroxylase Deficiency

A number of CYP7A1 coding region variants have been found, one of which is a frameshift mutation (L413fsX414) that results in loss of catalytic activity due to deletion of the heme binding domain of the enzyme, and leads to the disorder *cholesterol 7α-hydroxylase deficiency*. Two homozygous males had highly reduced bile acid excretion, significant hypercholesterolemia and hypertriglyceridemia, and were profoundly resistant to treatment by HMG-CoA reductase inhibitors (statins). They both had early gallstone disease. Liver biopsy and stool bile acid analysis indicated limited upregulation of the alternative bile acid synthetic pathway. The observed doubling of liver cholesterol content, due to low conversion to bile acids, may have resulted in downregulation of hepatic LDL receptors that could account for the raised plasma cholesterol. In terms of the hypercholesterolemia, there appears to be a gene dosage effect, with heterozygotes also affected. Thus, cholesterol 7α-hydroxylase deficiency is an autosomal codominant disorder.

3.2.2 Oxysterol 7α-Hydroxylase Deficiency

A single case of this disorder, associated with lack of synthesis of primary bile acids, was reported in a human infant due to a nonsense mutation in exon 5 of the *CYP7B1* gene. The child had cholestasis and early liver failure, due to accumulation of toxic monhydroxycholenoic acids. CYP7A1 activity was low in this patient, as in normal infants, and therefore was not able to compensate for the absence of CYP7B1. In contrast, *Cyp7b1* knockout mice had normal bile acid metabolism, with plasma cholesterol and triglycerides unchanged.

3.2.3 Cerebrotendinous Xanthomatosis

First reported in 1937, cerebrotendinous xanthomatosis (CTX) is caused by a deficiency of sterol 27-hydroxylase (CYP27A1). It is characterized by a marked reduction in bile acid synthesis, an accumulation of cholesterol and cholestanol in many tissues, and by ataxia, dementia, and premature atherosclerosis. Cyp27a1 knockout mice had even lower bile acid production, were hypertriglyceridemia, but had no CTX-like symptoms.

3.3 Disorders of Intracellular Lipoprotein Degradation

3.3.1 Wolman Disease and Cholesteryl-ester Storage Disease

Degradation of lipoproteins following their uptake by tissues involves lysosomal acid lipase, which catalyzes the hydrolysis of

cholesteryl esters and triglycerides. A deficiency in the activity of this enzyme can leads to two disorders, Wolman disease, and cholesteryl ester storage disease (CESD). Both are autosomal recessive disorders caused by mutations in the gene for lysosomal acid lipase (LIPA) on chromosome 10q23.31. The gene has 12 exons and spans 201 kb. Transcription leads, by alternative splicing, to two mRNAs that code for 399- and 401-residue proteins, respectively. Wolman disease is a severe disorder invariably fatal in the first year of life. It is associated with a failure to thrive, multiple gastrointestinal symptoms, including steatorrhea and hepatosplenomegaly, and adrenal calcification. CESD is associated with hepatomegaly, but its symptoms are generally less severe. It sometimes does not manifest in childhood. Owing to hyperlipidemia, severe premature atherosclerosis is common.

3.3.2 Niemann–Pick Disease Type C

Caused by defects in at least two genes, Niemann–Pick Disease Type C, first described in 1958, is a rare autosomal recessive disorder characterized by aberrant intracellular trafficking of endocytosed LDL cholesterol. The cholesterol accumulates in lysosomes. It is often a fatal disease with progressive visceral and neurological symptoms. The human *NPC1* gene spans 55 kb on chromosome 18q11.2 with 25 exons, and encodes a 1278-residue integral membrane protein with a sterol-sensing domain similar to HMG-CoA reductase, SCAP and Patched. The second gene associated with the disorder, NPC2, is on chromosome 14q24.3, spans 13 kb, and codes for a 151-residue protein. This has a lipid-binding motif and functions, together with the NPC1 protein, to facilitate cholesterol transport in the late lysosomal/endosomal pathway.

A gene related to *NPC1* is *NPC1L1* (Niemann–Pick disease, type C1, gene-like 1). The NPC1L1 protein product, highly expressed in enterocyte brush border membranes, plays a poorly understood, but critical, role in the intestinal absorption of cholesterol. It functions in the pathway of cholesterol absorption in which the cholesterol-lowering drug ezetimibe acts. Ezetimibe failed to inhibit the uptake of cholesterol in NPC1L1 knockout mice.

4 Triglyceride and Phospholipid Metabolism

Many triglyceride and phospholipid pathway enzymes are intimately bound to membranes and have proved difficult to purify, sequence, and clone. The steps leading to the formation of diacylglycerol from acyl CoA, *via* phosphatidate, are common to both the triglyceride and phospholipid pathways. Diacylglycerol is synthesized *via* the action of phosphatidic acid phosphatase (PAP). There is a family of PAP enzymes encoded by three genes, *PPAP2A, PPAP2B*, and *PPAP2C*. In each case, isoforms resulting from alternative splicing exist. They have an N-glycosylation site and 6 TM domains, being integral membrane glycoproteins. The PAPs also function in phospholipase D signal transduction.

The last step in the synthesis of phosphatidyl choline, the major phospholipid, is formed by the action of CDP-choline: 1,2-diacylglycerol choline phosphotransferase (or choline/ethanolaminephosphotransferase) an intrinsic, ER and Golgi enzyme. The human gene *CEPT1* located on chromosome 1p13.2 has 9 exons, spans 45 kb, and encodes a 396-residue protein. CEPT also catalyses the formation of phosphatidylethanolamine. First cloned from

yeast, the protein has a molecular mass of 46 kDa and seven membrane-spanning helices. CDP-choline itself is formed by the action of CTP:phosphocholine cytidyltransferase (CT). Rat CT has been cloned and has a molecular mass of 42 kDa. It lacks a hydrophobic domain and is attached to membranes by a 58-residue amphipathic α-helix. There are two human CT genes: *PCYT1A* and *PCYT1B*.

In humans, there are two enzymes that catalyze the final step in triglyceride synthesis. The diacylglycerol acyltransferases DGAT1 and DGAT2 utilize fatty acyl-CoA and diacylglycerol as substrates to form triglyceride. The *DGAT1* gene, first recognized because of its homology to ACAT, is at chromosome 8q24.3, has 17 exons, and spans 11 kb. *DGAT2* on chromosome 11q13.3 has 8 exons spanning 33 kb. It is expressed widely, but at high levels in liver and white adipose tissue.

5 Role of Nuclear Hormone Receptors

5.1 Sterol Regulatory Element Binding Proteins

The 5′ flanking region of the LDL receptor (LDLR) gene was found to contain a sterol regulatory element (SRE). In the presence of sterols, this element loses its ability to synergistically enhance transcription *via* two direct repeats that bind to the Sp1 transcription factor. A nuclear protein was identified that binds to the SRE and is termed *sterol regulatory element binding protein* (SREBP). SREs have since been identified in the promoters of other genes involved in sterol and fatty acid synthesis. Three similar transcription factors, SREBP-1a, SREBP-1c, and SREBP-2, all of which are 120-kDa proteins, have been identified, and they all bind to SREs. SREBP-1a, and SREBP-1c differ as a result of separate transcription starts sites in the *SREBF1* gene involving alternative initial exons (human gene: chromosome 17p11.2). The SREBP-2 protein is encoded by the *SREBF2* gene (human gene: chromosome 22q13.2). In mammals, SREBP-1a and SREBP1-c primarily activate the genes of fatty acid synthesis, and SREBP-2 activates those of sterol synthesis. The SREBPs are basic helix–loop–helix leucine zipper (bHLH-zip) transcription factors found in a wide variety of animals. The precursor peptides are attached to the ER membrane with the N-terminal bHLH-zip domain and the C-terminal regulatory domain located in the cytoplasm. The presence of sterols blocks a two-step proteolytic cleavage of SREBP that occurs in sterol-depleted cells when a released N-terminal fragment of the protein enters the nucleus and activates transcription of sterol-regulated genes *via* the SREs. In the presence of sterols, SREBP is bound by its regulatory domain to the SREBP-cleavage-activating protein (SCAP). SCAP has eight membrane-spanning helices in its N-terminus and a C-terminal propeller region composed of multiple WD repeats that binds to SREBP. The SREBP–SCAP complex is retained in the ER by proteins that interact with the sterol-sensing domains of SCAP, and which are encoded by insulin-induced genes 1 and 2 (*INSIG-1* and *INSIG-2*). Other proteins, not yet identified, are thought to retain the Insig proteins in the ER since they do not possess ER retention signals. In the absence of sterols the Insig proteins release SREBP–SCAP, which migrates in COPII (coat protein complex) vesicles to the Golgi apparatus. Here, the initial cleavage of SREBP by *site 1 protease (S1P)* (also called

PCSK8), encoded by the *MBTPS1* gene, occurs in a protein loop in the lumen of the ER. A second cleavage by the metalloproteinase S2P, encoded by the *MBTPS2* gene at site 2 in a transmembrane domain, releases the N-terminal active peptide.

5.2 Peroxisome Proliferator-activated Receptors

The peroxisome proliferator-activated receptor (PPAR) family of transcription factors was discovered in rodents. In rodents, PPARs are activated by chemicals, resulting in larger and more numerous peroxisomes. This is accompanied by increased β-oxidation through activation of the acyl CoA oxidase gene, which contains, in its promoter, a peroxisome proliferator response element (PPRE) consisting of a direct repeat of the motif PuGGTCA, to which PPAR binds as a heterodimer with the 9-cis retinoic receptor (RXR, NR2B1). PPARs play roles in differentiation, cell division, apoptosis, and inflammation. They are also key regulators of lipid and lipoprotein metabolism and of the pathways of gluconeogenesis and glycolysis. There are three human *PPAR* genes, *PPARA* (chromosome 22q13.31), *PPARG* (chromosome 3p25.2), and *PPARD* (chromosome 6p21.31) that encode, respectively, PPARα, PPARγ, and PPARδ. The first, PPARα is expressed at high levels in liver, kidney, muscle, and heart where it activates fatty acid β-oxidation. The second, PPARγ, in contrast, is highly expressed in intestine, adipose, and mammary tissue. It promotes the storage of lipids. The third receptor, PPARδ is more widely expressed than the other two members of the family. There are a number of fatty acid and fatty acid–derived compounds that are ligands for PPARs. In particular, a hydroxyeicosatetraenoic acid (HETE) derivative 8S-HETE and leukotriene B4 are potent activators of PPARα. Natural ligands for PPARγ include prostaglandin J2, 15-HETE and the hydroxyoctadecadienoic acids (HODE) 9-HODE and 13-HODE. The hypolipidemic fibrate drugs and antidiabetic glitazone drugs are synthetic PPAR activators.

5.3 Liver X Receptors

The orphan nuclear receptors LXRα and LXRβ are activated by oxysterols, though the nature of the precise physiological ligand is still unclear. Examples of natural compounds known to activate LXR are 24(S),25-epoxycholesterol and 20(S)-, 22(R)-, 24(S)-, and 27-hydroxycholesterol. The LXRs are key regulators of numerous genes involved in lipid and carbohydrate metabolism, the inflammatory response, and energy homeostasis. They bind as heterodimers with RXR to DR-4 (or LXRE) response elements in the promoters of target genes. These are two hexanucleotide repeats with the consensus (T/A/G)G(G/T)(G/T)(T/C)A, separated by four nucleotides. Lipid transfer and transporter genes that are targets for LXRs include *ABCA1*, *ABCG1*, *ABCG4*, *ABCG5/G8*, *CETP*, and *PLTP*. Gene targets involved in fatty acid synthesis are, *ACC* (acetyl CoA carboxylase), *FAS* (fatty acid synthase), and *SCD* (stearoyl CoA desaturase). Two apolipoprotein genes, *APOC2* and *APOE*, are also upregulated by LXRs as is *SREBP-1c* and *LPL*. The *Cyp7a1* gene in rodents, but not in humans is a target for LXRs. The *LXR* genes themselves are subject to positive feedback autoregulation by LXRs. The human gene for LXRα (*NR1H3*, chromosome 11p11.2) is expressed at high levels in

the liver, kidney, adipose tissue, and intestine, whereas that for LXRβ (*NR1H2*, chromosome 19q13.33) is more ubiquitously expressed. In liver, LXRα is induced by insulin, leading to upregulation of lipogenesis and downregulation of gluconeogenesis *via* repression of the rate-limiting enzyme phosphoenolpyruvate carboxykinase gene (*PCK1*).

5.4 Farnesoid X Receptor and Pregnane X Receptor

At high concentrations, farnesol was found to activate an orphan nuclear receptor. Thus, this receptor was named FXR (farnesoid X receptor). Since then FXR (human gene, *NR1H4*, chromosome 12q23.1) has been shown to be a bile acid sensor with the highest activation by chenodeoxycholic acid. Some 25- and 26-hydroxylated bile acid intermediates also have high affinity for FXR. With RXR, it binds as a heterodimer to the bile acid response elements (BARE) of a number of genes involved in bile acid and lipid metabolism. Target genes upregulated by FXR include *ABCC2* (multidrug resistance protein: MRP2), *ABCB11* (canalicular bile salt export pump: BSEP), *FABP6* (fatty acid–binding protein 6, the ileal bile acid-binding protein), *SHP* (see Sect. 3.2 above), *ApoC-II*, and *PLTP*.

A related nuclear receptor, the pregnane X receptor, PXR (human gene, *NR1I2*, chromosome 3q13.33), also binds as a heterodimer with RXR. Natural ligands include lithocholic acid. It activates several cytochrome P450 enzyme genes, notably the cluster of four *CYP3A* family genes, *CYP2B*, *ABCB1 (MDR1)* and *ABCC2 (MRP2)*. These enzymes are involved in drug metabolism, and some of them in lipid synthesis and transport also.

6 The Apolipoproteins

6.1 The Apolipoprotein Multigene Family

Nine apolipoproteins, all of which are exchangeable among lipoproteins, are members of a multigene family. These are *apoA-I*, *apoA-II*, *apoA-IV*, *apoA-V*, *apoC-I*, *apoC-II*, *apoC-III*, *apoC-IV* and *apoE*. The genes have a similar structure and have evolved from a common ancestor that was similar to apoC-I. Each has 4 exons, except apoA-IV, and apoA-V in which exons 1 and 2 have fused. Within exon 3 there is, in each case, a conserved region of 33 codons. This amphipathic α-helical lipid binding domain is a trimer of 11 amino acids. The fourth exon of each codes for variable numbers of consensus, 22-residue and, in some cases, 11-residue, lipid-binding tandem repeats.

The *apoA-I*, *apoC-III*, *apoA-IV* and *apoA-V* genes (Table 2) are clustered in a 48-kb region on chromosome 11 (11q23.3). This cluster is regulated by an enhancer region, necessary for intestinal expression of apoA-I and apoA-III, upstream of the *apoA-III* gene. Transcription factors involved include SP1 and HNF4. In a patient with premature atherosclerosis, a 6-kb inversion affecting the *apoA-I* and *apoC-III* genes resulted in apoA-I and apoC-III deficiency. In another case the entire gene cluster was deleted, resulting in apoA-I, apoC-III and apoA-IV deficiency and very low HDL levels. Population studies using polymorphic markers suggest an association between the A-I/C-III/A-IV/A-V locus and the frequency of hyperlipidemia and atherosclerosis. A case-control study of SNPs at this locus showed significant association between the FCH phenotype and the *ApoA-I* and *ApoA-V* genes.

Tab. 2 The human apolipoprotein genes.

Protein	*Alternative name*	*HUGO symbol*	*Chromosomal location*	*Gene size [kb]*	*Exons*	*Initial translated protein amino acid residues*	*Mature protein molecular weight [kDa]*	*Normal fasting concentration in plasma [mg dL^{-1}]*	*Lipoprotein found on*
ApoA-I		APOA1	11q23.3	1.9	4	267	28	130	HDL, chylomicrons
ApoA-II		APOA2	1q21-q23	1.3	4	100	8.7	40	HDL, VLDL, chylomicrons
ApoA-IV		APOA4	11q23.3	2.6	3	396	46	16	HDL, chylomicrons
ApoA-V		APOA5	11q23.3	2.5	3	363	39	<1	HDL
ApoB		APOB	2p24-p23	45	29	4563 (apoB-100) 2179 (apoB-48)	549 264	80 <1	VLDL, IDL, LDL chylomicrons and chylomicron remnants
ApoC-I		APOC1	19q13.2	4.6	4	83	6.6	6	HDL, IDL, VLDL
ApoC-II		APOC2	19q13.2	3.3	4	101	8.9	3	HDL, IDL, VLDL
ApoC-III		APOC3	11q23.3	3.1	4	99	8.9	12	HDL, IDL, VLDL
ApoC-IV		APOC4	19q13.2	3.3	3	127	11.5	<1	VLDL, HDL
ApoD		APOD	3q26.2-qter	15.2	5	189	19–32	10	HDL
ApoE		APOE	19q13.2	3.6	4	317	34	5	HDL, IDL, VLDL
ApoH	beta-2-glycoprotein I	APOH	17q23-qter	17	8	345	50	20	HDL, VLDL, chylomicrons
ApoJ	Clusterin, SP-40/40	CLU	8p21-p12	17	9	427	70	1.4	
ApoLI		APOL1	22q13.1		7			<1	HDL, VLDL
Apo(a)[a]	lipoprotein Lp(a)	LPA	6q26–27	212	39	4548	550	variable[a]	Lp(a)

[a]Data for apo(a) is given for the gene in the public domain genome sequence. The protein size can vary from 300 to 800 kDa. The concentration in human plasma can vary from undetectable to over several hundred mg dL^{-1}.

The genes for apoC-I, apoC-II, apoC-IV, apoE, (Table 2) and an apoC-I pseudogene are clustered on chromosome 19 (19q13.32) spanning 44 kb. The regulation of this cluster is discussed below (Sect. 6.11).

6.2 Apolipoprotein A-I

The major protein component of HDL, apoA-I is an activator of LCAT. The secreted 249-residue propeptide is processed in the plasma to the mature 243-residue (28 kDa) form. A number of rare charge variants have been detected by isoelectric focusing. Two of these, P165R and R173C (apoA-I Milano), in addition to a number of truncated variants, are associated with low levels of HDL. In the case of apoA-I Milano, the decrease in the level of HDL is due to an elevated rate of catabolism of the abnormal apoA-I species. Recently, clinical trials of intravenously infused recombinant apoA-I Milano have shown regression of coronary atherosclerosis. The practicality of this approach *versus* more conventional drug therapy in treating large numbers of patients is questionable. Several mutations, including G26R and L60R, cause amyloidosis. Two others, K107del and P165R, are mildly defective in activating LCAT.

An XmnI RFLP in the 5′flanking region is associated with hypertriglyceridemia in some populations. A possible functional single-base G/A substitution polymorphism in the promoter at nucleotide −75, in one of two inverted repeats, was associated with differences in HDL levels in some populations.

A number of cis-acting elements upstream of the gene have been identified as required for full liver-specific expression of apoA-I. However, while only 256-bp 5′ of the *apoA-I* gene are required for hepatic expression, a construct containing 5.5-kb 5′ and 4-kb 3′ of the human gene was not sufficient for intestinal expression in transgenic mice. This was achieved only when a construct was used that contained both the *apoA-I* and *apoC-III* genes together with 300-bp 5′ to the *apoA-I* gene and 1.4-kb 5′ to the *apoC-III* gene.

ApoA-I production and the level of the mRNA in rat liver are stimulated by thyroid hormone, an effect due both to increased transcription and to stabilization of apoA-I nuclear RNA. HNF-4, a member of the thyroid/steroid hormone receptor superfamily, has been shown to interact with a sequence in the 5′ flanking region and to increase promoter activity. Inhibitory effects have been shown for another DNA-binding protein, ARP-1, at the same site.

The overexpression of human apoA-I in transgenic mice increased the HDL cholesterol levels and prevented early diet-induced atherosclerosis. When these animals were crossed with apoE-deficient mice there was a marked decrease in the development of atherosclerosis, which is a characteristic of the apoE-deficient animals.

Homozygous apoA-I-deficient mice produced by gene targeting had only 20% of the normal level of HDL cholesterol but were generally healthy.

6.3 Apolipoprotein A-II

A secondary protein component of HDL, apoA-II is synthesized and secreted mainly by the liver and intestine. Unlike apoA-I it does not activate LCAT. Like apoA-I, it is regulated both transcriptionally and translationally.

Following cleavage of the 18-residue signal sequence some of the 82-residue

propeptide is cleaved prior to secretion, forming the mature 77-residue peptide. The rest is processed in the plasma. ApoA-II contains one free cysteine residue, enabling it to form heterodimers with apoE and apoD, though most of it exists in the form of a homodimer. The physiological function of apoA-II is largely unknown. It may be involved in the modulation of LCAT and HTGL activity.

Numerous studies have explored the regulation of the human gene expression. A diverse set of promoter elements have been identified. An MspI polymorphism at the 3′ end of the gene, in an Alu repeat sequence, has been reported to be associated with differences in the levels of apoA-I and apoA-II in serum.

An increased tendency to develop atherosclerotic lesions was seen in transgenic mice that overexpress human apoA-II. These animals had elevated levels of HDL and the particles were larger with a higher ratio of apoA-II to apoA-I when compared with normal mice.

6.4
Apolipoprotein A-IV

The 46-kDa glycoprotein, ApoA-IV is synthesized mainly in the small intestine and exists in a number of isoforms. The precise role of apoA-IV is unknown, but, like apoA-I, it activates LCAT and may be involved in reverse cholesterol transport. It has been reported to be a satiety factor. ApoA-IV is rapidly catabolized, having a tenfold higher fractional catabolic rate than apoA-I. The rate of synthesis determines the level of apoA-IV in plasma.

Three common polymorphisms exist: N127S, T347S and Q360H. With the latter polymorphism, the fractional catabolic rate of the *His* allele product has been reported to be lower than that of the more common *Gln* allele (frequency 92%). Genotypic variation in the *APOA-1V* gene affects the response of LDL cholesterol to diet.

6.5
Apolipoprotein A-V

Unlike the other apolipoproteins, apoA-V (and apoC-IV) was not discovered first as a protein isolated from plasma. Rather, the presence of apoA-V was revealed in the rat as the most highly upregulated of several novel genes following partial hepatectomy, and as a result of a comparison of the mouse and human genomes. It shows noticeable homology with apoA-I and apoA-IV, and the gene lies close to the cluster on chromosome 11q23 together with the genes for these two apolipoproteins as well as apoC-III. The apoA-V protein was then identified in plasma in the HDL fraction. Overexpression, either as a transgene, or by adenovirus transfection, resulted in decreased levels of plasma triglycerides. In contrast, knockout mice showed a marked increase in triglycerides. Together with the confirmation that functional FXR and PPARα response elements exist in the promoter for apoA-V, this strongly suggests a role in the regulation of triglyceride metabolism in the liver. The upregulation of apoA-V after hepatectomy suggests it has a role in the provision and trafficking of lipid for formation of new cell membranes. The apoA-V sequence contains amphipathic α-helices with an overall 60% α-helical content. Like ApoA-I, it can organize dimyristoylphosphatidylcholine vesicles into discoidal structures with diameters of 15 to 20 nm. However, unlike ApoA-I, it only poorly activates LCAT and does not increase cholesterol efflux from cAMP-stimulated macrophages.

6.6 Apolipoprotein B

In plasma from fasting humans, 90% of circulating apoB is in LDL, 8% in IDL and 2% in VLDL and chylomicrons. A centile nomenclature is used to describe apoB species based on their molecular weights. There are two circulating forms, apoB-48 and apoB-100 (Table 2). C-terminally truncated apoB-48, found on chylomicrons and their remnants, is produced in humans in the intestine. ApoB mRNA (14 kb in length) is edited by an editing complex in the nucleus of intestinal enterocytes in an enzyme-mediated process, such that codon 2153 becomes UAA (a stop codon) in place of CAA (Glu) in the gene sequence. A 26-nucleotide domain, 6662 to 6687, is highly conserved in mammals and necessary for editing to occur. The editing enzyme, Apobec-1, a 27-kDa protein with cytidine deaminase activity, was isolated and cloned from rat intestine. Additional factors are required for the editing to occur. The first to be identified is the Apobec-1 complementation factor (ACF), a 65-kDa protein.

In humans, the liver makes only the apoB-100 found in plasma on VLDL, IDL, and LDL. ApoB mRNA editing is undetectable in liver from humans, pigs, cows, sheep, cats, and monkeys. In liver from rabbits and guinea pigs, editing activity is very low. However, in several species, including dogs, horses, rats, and mice, an appreciable degree of editing has been found. The extent of this hepatic editing may be an important determinant of the level of apoB-containing lipoproteins in the circulation.

There is little acute metabolic regulation of hepatic levels of apoB mRNA. The availability of lipid and the extent of intracellular degradation of apoB are important in regulating secretion rates. The proteasome–ubiquitin pathway is the major means for the degradation of the apoB that fails to be lipidated.

Much larger than the exchangeable apolipoproteins, apoB has a globular N-terminus followed by four alternating amphipathic α-helical and β-sheet domains. This "pentapartite" structure is given as NH_2-$\beta\alpha1$-$\beta1$-$\alpha2$-$\beta2$-$\alpha3$-CO_2H. Through interaction with MTP, the $\beta\alpha1$ domain forms a lipid pocket that promotes lipid accumulation in the early stages of VLDL formation. The $\beta1$ and $\beta2$ hydrophobic surfaces interact with, and organize, the lipoprotein lipid core. The overall structure of native LDL is generally thought to be spherical. ApoB is highly glycosylated *via* 16 N-linked sites, with carbohydrate contributing 35 kDa of the molecular weight of apoB-100. Two regions (residues 1280 to 1320 and 3180 to 3282) are particularly sensitive to cleavage by endoproteases, an indication that they are exposed on the surface of LDL. Thrombin cuts at 1297 and 3249, forming three fragments, T4 (N-terminal), T3 (central) and T2 (C-terminal). The region from residues 3000 to 3600 that flanks the T2/T3 junction contains sequences responsible for binding to the LDL receptor. Receptor-blocking monoclonal antibodies have been mapped to this region. Two positively charged sequences in this region (residues 3147 to 3157 and 3359 to 3367) bind strongly to heparin. The second of these has close homology to the LDL receptor binding domain of apoE.

The tissue-specific regulation of expression of the *apoB* gene is complex with many cis- and trans-acting factors. Regulation of intestinal expression remains largely enigmatic. Studies with transfected cells, *in vitro*, and with transgenic mice have provided an understanding of the

liver-specific regulation. Binding sites for a number of nuclear regulatory proteins including ARP-1, EAR-2, EAR-3, HNF-4 and C/EBP have been identified within the promoter region. HNF-4 and C/EBP activate transcription synergistically. ARP-1, EAR-2, and EAR-3 repress transcription. There is a core enhancer in intron 2 that binds HNF-1, C/EBP, and an uncharacterized protein. A weaker enhancer lies in intron 3. In addition to these elements, nuclear matrix association regions (MARs), identified at the 5′ and 3′ ends of the gene, have been shown to play an important role in apoB expression. Tissue-specific transcription in liver and intestine correlates with undermethylation of the 5′ flanking region in these tissues.

There is a hypervariable region, composed of 15-bp AT-rich tandem repeats, located 200-bp downstream of the gene, with at least 14 common alleles ranging from 25 to 52 repeats. Five single-amino-acid substitution polymorphisms, the antigenic group (Ag) series, were first detected in antisera from multiply transfused subjects. Four are in association with unique common apoB polymorphisms. In the case of the fifth, the Ag(x/y) polymorphism, two SNPs in complete linkage disequilibrium are involved. These are BfaI and Eco57I RFLPs, associated with P2712L and N4311S substitutions, respectively. Pro2712/Asn4311 represents the more common Ag(y) epitope. A silent XbaI RFLP in the third base of the codon for Thr2488 is in strong linkage disequilibrium with the Ag(x/y) polymorphism. An association of the XbaI RFLP with differences in lipid levels has been reported in a number of studies, with this polymorphism accounting for possibly 3 to 5% of the variance seen in the amount of LDL cholesterol in the general population. A common insertion/deletion polymorphism results in either a 27- or 24-residue signal sequence. A rare third allele has 29 residues. A number of studies have examined the association between particular polymorphisms or haplotypes and either lipid levels or incidence of atherosclerotic disease.

6.7 Apolipoprotein C-I

A minor component of HDL, IDL, and VLDL, apoC-I is found mainly in HDL (97%). More than 90% of the circulating apoC-I is of hepatic origin, with moderate expression in skin and trace levels of expression in brain. Little is known concerning its physiological function. It was shown to be a highly effective inhibitor of the binding of apoE-enriched βVLDL to the α_2-macroglobulin receptor/LRP. The expression of the human *apoC-I* gene in transgenic mice resulted in high levels of triglyceride-rich lipoproteins, indicating that apoC-I is involved in the regulation of their metabolism. Evidence in baboons indicates that an N-terminal fragment of apoC-I interacts with apoA-I to inhibit CETP activity. It may also be a cofactor for LCAT. ApoC-I, which has a domain that binds to lipopolysaccharide seems to play a role in infection and inflammation.

A functional apoC-I promoter polymorphism at −317 (a HpaI RFLP), between the *apoC-I* and *apoE* genes, is associated with type III hyperlipidemia *via* the apoE genotype. There is a synonymous SNP in codon 6, and BglI and DraI RFLPs have been reported. Common liver-specific regulation of expression of the *apoC-I* and *apoE* genes is discussed below (Sect. 6.11). ApoC-I deficient mice produced by gene targeting have a moderately elevated level

of serum triglycerides. The rate of clearance of remnant lipoprotein particles in these animals is decreased.

6.8 Apolipoprotein C-II

The apolipoprotein apoC-II gene is expressed mainly in the liver, with intestinal cells containing only small amounts of mRNA. It is glycosylated within the cell but is subsequently deglycosylated in the plasma, where it exists mainly in the 8.9 kDa, 79-residue form, although some of it is further processed by removal of the N-terminal 6-amino acids to produce a minor isoform. The only known role of apoC-II is as the requisite cofactor for LPL. The activator and primary LPL binding domains are located in the C-terminal third of apoC-II. Like the other C apolipoproteins, apoC-II is transferred during hydrolysis from chylomicrons and VLDL to HDL. Most of the circulating apoC-II is present in HDL; 10% is in IDL, and 30% is in VLDL and chylomicrons.

The first intron of the *apoC-II* gene contains 4 Alu type sequences and a $(TG)_n(AG)_m$ microsatellite with at least 15 different alleles. The third intron contains a minisatellite composed of 37-bp tandem repeats. There are two common alleles of six and seven repeats with frequencies of 82 and 18%, respectively. AvaII, BanI, BglI, NcoI, PstI, and TaqI, RFLPs have been reported. In human hepatoma cells, apoC-II is upregulated by the nuclear receptor FXR.

Three charge variants of apoC-II, K19T, E38K, and K55Q, all detected by isoelectric focusing, have been reported. The first two are rare, but the K55Q variant is common among individuals of African decent. Whether they are associated with hyperlipidemia is unclear.

6.8.1 Apolipoprotein C-II Deficiency

Autosomal recessive familial chylomicronemia, due to homozygous apoC-II deficiency, has been described in ten instances. It is characterized by fasting chylomicronemia and severe hypertriglyceridemia, and is often accompanied by the presence of xanthomas and pancreatitis. Six cases were due to apoC-II gene mutations that introduced premature stop codons, resulting in either the absence of apoC-II or very low levels of truncated species. One other was due to an intron 2 donor splice defect, and one to an initiation defect M-22V. In one case, a single C insertion between codons 69 and 70 resulted in low levels of a 96-residue frameshifted protein with severely impaired cofactor activity. A major disruption of the *apoC-II* gene, and neighboring *apoC-IV* gene, by a 7.5-kb deletion, has also been reported.

6.9 Apolipoprotein C-III

The glycoprotein, apoC-III is the most abundant protein in VLDL after apoB100, comprising 25 to 30% of the protein mass. Most is found in HDL (60%), 10% in LDL, 10% in IDL and 20% in VLDL. There are three isoforms: apoC-III_{-0}, apoC-III_{-1} and apoC-III_{-2}, with 0, 1, and 2 residues of sialic acid per molecule, respectively. These have different isoelectric points and can be separated by isoelectric focusing.

Little is known about the precise function of apoC-III in lipoprotein metabolism. It inhibits LPL *in vitro*. Although a number of hydrophilic sequences within the peptide interact with LPL, the major inhibitory effect resides in an NH_2 terminal domain. ApoC-III does not compete with apoC-II at the apoC-II activation site on LPL. Glycosylation of apoC-III does not seem to be important for LPL inhibition.

Overproduction of VLDL in rats fed a high-sucrose diet does not alter the rate of apoC-III gene transcription, or the level of mRNA, in liver. Inflammation is linked to a reduction in the level of hepatic apoC-III mRNA and with a decrease in circulating HDL apoC-III content. Thyroid hormone has been reported to improve the efficiency of maturation of the apoC-III mRNA.

Studies using transgenic mice suggest that apoC-III is involved in the apoE-mediated clearance of triglyceride-rich lipoproteins. Overexpression of the human gene in transgenic mice causes elevated levels of VLDL triglyceride with more than double the normal amount of apoC-III per particle, and is accompanied by an increase in circulating nonesterified fatty acids. The VLDL particles have a larger mean diameter and are taken up poorly by lipoprotein receptors. Tissue levels of LPL are normal and it is likely that the hypertriglyceridemia is a result of the low fractional catabolic rate of VLDL and a low rate of clearance of apoB48 remnants in these animals. This low rate is probably due to the combined effect of the increased amount of apoC-III and the decreased amount of apoE, on VLDL. These mice were crossed with those that overexpress apoE to produce animals overexpressing both apoC-III and apoE. These animals had normal levels of triglycerides and chylomicron remnants.

ApoC-III deficient mice, produced by gene targeting, have decreased fasting levels of triglycerides and show an absence of postprandial lipemia with an increased rate of clearance of chylomicrons. An interpretation of these findings is complicated by the fact that the intestinal expression of the neighboring *apoA-I* and *apoA-IV* genes are decreased in these animals.

In a family of Mexican origin an apoC-III Q38K mutant was associated with moderate hypertriglyceridemia. Rare variants, D45N and A23T do not seem to be associated with dyslipidemia. Another mutant, T74A, is not glycosylated. Hyperalphalipoproteinemia with abnormal HDL containing large apoE-rich particles was seen in subjects with an apoC-III K58E mutation. There is an SstI SNP in the 3′ untranslated region of exon 4 and a tetranucleotide repeat polymorphism, with 7 common alleles, in intron 3, at the end of an Alu sequence. The SstI SNP and a polymorphism in a promoter insulin response element are associated with differences in triglyceride levels.

6.10 Apolipoprotein C-IV

That an additional apolipoprotein gene might exist at the 19q13.2 locus along with *apoC-I*, *apoC-II*, and *apoE* was first suggested in a study of the corresponding rat locus. Later, the human *apoC-IV* gene was characterized and shown to be located only 555-bp upstream of *apoC-II*. It lacks a TATA box and is transcribed in liver at a level well below that of the *apoC-II* gene. A higher expression level of apoC-IV in the rabbit has been attributed to the presence of a unique purine-rich tandem repeat sequence in the promoter and 5′ UTR that functions as a GAGA box. The peptide has two amphipathic α-helical regions and has but a limited homology with apoC-I, apoC-II, and apoE. Very low levels of apoC-IV are found in human plasma, mainly in VLDL with some in HDL. Much higher levels of the protein are found in rabbit plasma, where it is primarily associated with VLDL particles. Rabbit apoC-IV was the first to be characterized. The human protein can be O- and N-glycosylated, and exists as a number of sialylation isoforms. When the human protein was expressed

in mice there was an associated increase in plasma levels of triglyceride.

6.11 Apolipoprotein E

ApoE mRNA is found in a number of tissues, notably liver, brain, kidney, adrenal gland, macrophages, spleen, and ovary. The sequence between residues 140 and 160 of the mature peptide constitutes the ligand that binds to the LDL receptor. The C-terminal residues 245 to 299 are important for the association of apoE with lipoprotein particles and for tetramer formation. Most of the apoE in the circulation is synthesized in the liver. Much of the newly synthesized apoE, like apoB, is broken down by a degradative pathway. The synthesis and secretion of apoE is reduced by insulin. Half of the apoE present in serum is in HDL, 10% in LDL, 20% in IDL, and 20% in VLDL. It is a glycoprotein with a single O-linked chain at Thr194. The carbohydrate contains one or more sialic acid residues.

There are three common alleles of the apoE gene, $\varepsilon 2$, $\varepsilon 3$, and $\varepsilon 4$, present at frequencies of 9, 77, and 14%, which give rise to E2, E3, and E4 isoforms, respectively. The proteins can be separated by isoelectric focusing, with E4 the most basic, and E2 the most acidic. They differ at residues 112 and 158. E2 is Cys112/Cys158, E3 is Cys112/Arg158 and E4 is Arg112/Arg158. Some of the E3 in plasma exists as a homodimer and some as a heterodimer with apoA-II. The binding affinity of E2 for the LDL receptor is less than 5% that of E3 or E4. This is thought to be due to reorganized salt bridges as a result of the loss of the Arg158 residue. This changes the charge disposition of the surface that is presented to the receptor.

Expression of apoE in liver in transgenic mice was shown to require the presence of a hepatic control region (HCR) located 18-kb and 9-kb downstream of the *apoE* and *apoC-I* genes, respectively. This HCR lies between the *apoC-I* gene and the *apoC-I* pseudogene. Constructs containing a 5-kb apoE promoter but lacking the HCR resulted in expression only in kidney and skin. A second HCR was later located 27-kb downstream of the *apoE* gene. The four genes in the 19q13.2 cluster, *apoC-I*, *apoC-II*, *apoC-IV*, and *apoE*, are coordinately regulated. In liver, this regulation is *via* the two distal 350-bp HCRs (HCR.1 and HCR.2), which show close homology. In macrophages, adipose tissue, brain, and skin regulation is through two homologous 620-bp multienhancer regions (ME.1 and ME.2) situated 3.3-kb and 15.9-kb downstream of the *apoE* gene, respectively. ME.1 and ME.2 contain LXR response elements, and in macrophages the mRNAs for all four genes were upregulated in response to LXR ligands. It has been hypothesized that the induction of these four genes may be important for promoting the efflux of lipid from macrophage foam cells in atherosclerotic lesions.

Transgenic mice that overexpress apoE have a striking decrease in plasma levels of VLDL and LDL due to increased clearance rates of triglyceride-rich lipoproteins.

Homologous recombination was used to target and inactivate the *apoE* gene in mice. These *Apoe*$^{-/-}$ animals have hypercholesterolemia, which is especially severe when they are fed a high fat diet. Plasma levels of VLDL and VLDL remnants, containing mainly apoB-48 rather than apoB-100, are raised. The mice exhibit rapid development of atherosclerotic lesions that are thought to be a result simply of the very high cholesterol levels (400

to 500 mg dL^{-1}) rather than to the atherogenicity of the VLDL and VLDL remnant particles. Mice resulting from crossing *Apoe*$^{-/-}$ mice with those that exclusively produce apoB-100 (*Apob*$^{100/100}$ mice) had lower levels of total cholesterol, similar to the levels in *Ldlr*$^{-/-}$ mice crossed with *Apob*$^{100/100}$ mice. The latter animals had much more severe atherosclerosis due to higher numbers of small LDL.

The abnormal phenotype of *Apoe*$^{-/-}$ mice was to a large extent corrected following bone marrow transplantation from a donor mouse with the normal *apoE* gene, indicating the extent to which apoE synthesized by macrophages can contribute to the pool of lipoprotein apoE.

6.11.1 Dysbetalipoproteinemia and ApoE Deficiency

The 1% of individuals who are homozygous for the $\varepsilon 2$ allele have a tendency to accumulate βVLDL in plasma. Additional unknown genetic or other factors cause 5% of these subjects to develop clinically important dysbetalipoproteinemia with marked accumulation of chylomicron and VLDL remnants. This disorder is referred to as *type III hyperlipidemia*.

Rare isoforms of apoE have been discovered, many of which are defective in binding to the LDL receptor. In four of these cases the mutation results in dominant transmission, and in one case recessive transmission, of type III hyperlipidemia. A double mutant, apoE-4$_{\text{Philadelphia}}$ (E13K and R145C) is associated in the homozygous state with a severe form of type III hyperlipidemia and in the heterozygous state with a moderate form. Transgenic mice have been produced that express the human mutant apoE$_{\text{Leiden}}$, which contains a seven-amino acid insertion that is a tandem repeat of residues 121 to 127. In these animals there is dominant transmission of hyperlipoproteinemia, characterized by elevated levels of VLDL and LDL.

ApoE deficiency is a rare genetic disorder characterized by less than 1% of normal levels of apoE in plasma. Patients have xanthomas, type III hyperlipidemia and premature atherosclerosis. There is delayed clearance of chylomicron and VLDL remnant particles, leading to elevated levels of cholesterol-rich VLDL and IDL.

6.11.2 Alzheimer Disease and ApoE

The frequency in the population of the $\varepsilon 4$ allele declines with age and that of the $\varepsilon 2$ allele increases. There is much evidence for a highly significant genetic association between the $\varepsilon 4$ allele and late onset Alzheimer's disease (LOAD), both in the familial and the sporadic forms. The effect is proportional to gene dosage, with the $\varepsilon 4/\varepsilon 4$ genotype being a greater risk factor than the $\varepsilon 4/\varepsilon 3$ or $\varepsilon 4/\varepsilon 2$ genotypes. These, in turn, are greater risk factors than the other genotypes ($\varepsilon 3/\varepsilon 3$, $\varepsilon 3/\varepsilon 2$ or $\varepsilon 2/\varepsilon 2$). In families with late-onset Alzheimer's disease, the presence of the $\varepsilon 4/\varepsilon 4$ genotype alone is associated with the disease in almost all of those over 80 years of age. Additional evidence, supporting the link between apoE and Alzheimer's disease, is that the gene locus is in a region of chromosome 19, previously associated with LOAD. The apoE protein has been found in senile plaques and neurofibrillary tangles along with the β-amyloid protein. *In vitro*, both apoE4 and apoE3 bind with high specificity, and irreversibly, to β-amyloid, with apoE residues 244 to 272 being crucial for the interaction. The formation of the complex of β-amyloid with apoE4 is much more rapid than with apoE3.

6.12 Apolipoprotein (a)

Apolipoprotein (a) [apo(a)], synthesized in the liver, is found on a lipoprotein species Lp(a), where it is covalently attached to apoB *via* a disulfide bridge. The two proteins also interact noncovalently. Lp(a) floats in the 1.050 to 1.080 g mL^{-1} density fraction in the ultracentrifuge and can be separated from HDL by gel filtration. The level of Lp(a) within the population varies from less than 0.10 mg dL^{-1} to as much as 500 mg dL^{-1}. Of this variation, 90% is genetically determined, primarily at the *apo(a)* gene locus with a small amount possibly attributed to the *apoE* genotype. One factor that affects the plasma concentration is the size of a particular isoform, with an inverse correlation between the size and the rate of secretion, this rate being determined by the efficiency of posttranslational processing.

High levels of Lp(a) have been shown, in most epidemiological studies, to be an independent risk factor for premature atherosclerosis, but the physiological function of Lp(a) is still essentially unknown, although the high degree of homology of apo(a) with plasminogen appears to be responsible for its inhibitory effect on thrombolysis. Plasminogen and apo(a) compete in binding to fibrin. Recombinant apo(a) binds with high affinity *via* intrachain lysine residues of intact fibrin and to carboxy-terminal lysines on the degraded peptides. To what degree Lp(a) inhibits fibrinolysis physiologically has been questioned. It has been postulated that Lp(a), *via* a reduction in the formation of plasmin, enhances vascular smooth muscle cell proliferation as a result of lower activation of TGFβ.

Thyroid hormone status, besides modulating the plasma concentration of total apoB also affects Lp(a) levels. In the hypothyroid state, there is impaired catabolism. Secretion is suppressed and catabolism elevated in the hyperthyroid state.

It is not clear to what extent Lp(a) is secreted directly from the liver or formed in plasma as a result of association between free apo(a) and LDL. It does not appear to originate from a triglyceride-rich lipoprotein precursor. Contact with LDL is made through only a few kringle repeats with most of the apo(a) extended away from the particle.

The human gene for apo(a), *LPA*, is closely linked to that of plasminogen (*PLG*), the two genes being separated by 38 kb. Two *apo(a)*/plasminogen pseudogenes are present at the same locus. The size of the *apo(a)* gene is highly polymorphic and depends on the number of exons that code for tri-loop peptide structures held together by three internal disulfide bonds and termed *kringles*. Plasminogen has five kringles numbered I to V. Apo(a) has a variable number of kringles, with high homology to plasminogen kringle IV, and a single copy homologous to kringle V. It also has a region that is homologous to the plasminogen protease domain, but it does not have a tissue plasminogen activator cleavage site. Cys4057, in kringle 36 of the published cDNA sequence, one of the kringle IV repeats, forms a disulfide bond with apoB Cys4326. In the noncovalent interaction, apoB lysine 680 plays a critical role. The size polymorphism, which is reflected in the molecular weight range ($<$300 to $>$800 kDa) of more than 34 different apo(a) isoforms, can be detected by pulsed-field gel electrophoresis after digestion with KpnI. In addition to these size

polymorphisms, 14 SNPs have been reported. There is a common Met to Thr polymorphism in kringle IV type 10. A Trp-to-Arg mutation, also in kringle IV type 10, results in Lp(a) that is defective in binding to the lysine residues of fibrin. A C/T polymorphism in the 5′ untranslated region of the gene is associated with decreased *in vitro* translation.

When fed a high-fat diet, transgenic mice (a species in which Lp(a) is absent) that express human apo(a) had larger areas of lipid-staining atherosclerotic lesions than control animals did. The apo(a) in these animals is not in the form of Lp(a); that is, it is not covalently attached to apoB. However, when these mice were crossed with mice expressing human apoB100, high levels of human Lp(a) were found in the plasma.

6.13 Other Apolipoproteins

6.13.1 Apolipoprotein D

The sialoglycoprotein apoD is related to retinol-binding protein and both are members of the lipocalin superfamily. ApoD shows no homology with the other apolipoproteins. In humans, the gene is expressed in a number of tissues, notably the adrenal glands, cerebellum, spleen, kidney, and pancreas. There is much less expression in the liver and intestine. Cleavage of the 20-residue signal peptide gives rise to the mature form in which the N-terminal Gln is blocked by cyclization. Much of it exists as a 38-kDa heterodimer with apoA-II. Isoelectric focusing detects a number of isoforms with variable numbers of sialic acid residues. It is mainly found in the HDL fraction where it constitutes about 5% of the total protein. It binds to a number of ligands, including progesterone, cholesterol, bilirubin, and arachidonic acid. ApoD possibly plays a role, alongside LCAT and CETP in reverse cholesterol transport. There is increasing evidence that apoD is important for nerve regeneration.

Putative regulatory domains have been reported in the apoD promoter region. Response elements include those for estrogen, progesterone, glucocorticoid, and thyroid hormone. The promoter contains a 26-bp element that could form a Z-DNA structure. MspI and TaqI RFLPs have been studied.

6.13.2 Apolipoprotein H

The physiologic function of apoH has not been determined with any certainty. It is also referred to as *beta-2-glycoprotein I*. In addition to being involved in lipoprotein metabolism, it has also been reported to be involved in coagulation, and is a cofactor for the binding of antiphospholipid autoantibodies to anionic phospholipid. These antibodies are found, for example, in patients with lupus erythematosus. The gene for apoH is alternatively spliced and probably produces a number of distinct polypeptide products.

6.13.3 Apolipoprotein J

An abundant and widely dispersed glycoprotein, ApoJ, also referred to as *clusterin* and *SP-40/40*, is expressed at a high level in the epithelial cells of many tissues. A constituent of the terminal complement complex, apoJ is a multifunctional glycoprotein that inhibits complement-mediated cytolysis. It is found on an unstable subclass of HDL particles that have $\alpha 2$ electrophoretic mobility, only a small proportion of which contain apoA-I. ApoJ and apoA-I associate *in vitro*. ApoJ is an inhibitor of the complement membrane-attack complex. ApoJ is

stored in platelet granules and released after platelet activation, and it may be involved in wound repair. The other functions of apoJ are thought to include roles in sperm maturation, lipid transport, programmed cell death, and membrane recycling.

6.13.4 Apolipoprotein L Gene Family

ApoL-I, reported to be the trypanosome lytic factor, was discovered on a subpopulation of HDL particles in HDL isolated by selected-affinity immunosorption. The gene for apoL-I is a member of a family of six genes; however, only apoL-I seems to be truly an apolipoprotein. Microarray analysis of prefrontal cortex tissue from schizophrenia subjects showed upregulation of the *ApoL-I, ApoL-II,* and *ApoL-IV* genes.

7 Scavenger Receptors

7.1 SR-BI/CLA1

The scavenger receptor class B type I, SR-BI, which binds directly to cholesterol, is able to alter the distribution of membrane cholesterol pools, and is localized to cholesterol-rich rafts. It facilitates the selective uptake by the liver of cholesteryl esters from HDL, thereby playing a crucial role in reverse cholesterol transport. SR-BI binds through a C-terminal domain to the multifunctional PDZK1 protein, which may play a scaffolding role. PDZK1 knockout mice have increased serum levels of cholesterol.

Previously, it was thought that initial binding of HDL to the receptor was followed by diffusion of lipid into the plasma membrane down a concentration gradient. It is now believed that, following binding to SR-BI, the complete HDL particle undergoes endocytosis. The cholesteryl esters are hydrolyzed in early endosomes and the HDL particle is then recycled to the plasma membrane and secreted. SR-BI expression correlates with the secretion of cholesterol into bile, SR-BI being expressed in canalicular as well as sinusoidal membranes. Hence, it is thought that SR-BI helps to mediate the secretion of cholesterol into bile, possibly in conjunction with phospholipid transporters such as ABCB4 and sterol transporters ABCG5/G8. In peripheral tissues, SR-BI mediates the diffusion of free cholesterol from the plasma membrane to *alpha* HDL particles. In macrophages, it plays an important antiatherogenic role.

The human gene for SR-BI (also known as *CLAI* and *CD36L1*) is *SCARB1* and is located at chromosome 12q24.31. It spans 86 kbp and has 13 exons. It codes for a 509-residue, 57-kDa, protein, which is N-glycosylated at a number of sites, giving a total molecular weight of 82 to 86 kDa. Both the N- and C-termini are located in the cytoplasm with two transmembrane domains separated by a 408-residue extracellular domain. As a result of alternative splicing, there is a second isoform called *SR-BII,* with a different C-terminus.

7.2 Modified LDL Receptors: MSR1, CD36 and LOX1

Oxidized or chemically modified LDL are taken up by macrophages *via* receptor-mediated endocytosis, independent of the LDL receptor. This may be the mechanism, *in vivo*, for foam cell formation, leading to atherosclerotic lesions. Lipoproteins containing apoB have been

isolated from normal and atheromatous arteries.

Initially, from bovine liver, a glycoprotein was isolated that binds to acetylated LDL, oxidized LDL, and endotoxin (bacterial lipopolysaccharide), and was referred to as the *acetylated LDL receptor* or *scavenger receptor*. The ligands for this receptor are numerous and include polyanionic compounds, such as polyribonucleotides. Cloning of the receptor, SR-A, now referred to as the *macrophage scavenger receptor 1 (MSR1)*, revealed the presence of two homologous proteins, the type I and type II class A scavenger receptors. These are produced from a single MSR1 gene by alternative splicing. The human gene at chromosome 8p22 has 10 exons. Isoform type I has 451 residues (50 kDa) and type II 388 residues (38 kDa). They have identical structures except for the C-terminus. The common structures include an NH_2-terminal cytoplasmic tail, a transmembrane region and a collagen-like domain. The active receptors exist as trimers. A third isoform exists, that does not internalize acetyl-LDL and by inhibiting the first two isoforms plays a dominant negative function. Despite increased foam cell formation in cells overexpressing MSR1 *in vitro*, decreased atherosclerosis was observed in MSR1 transgenic mice.

The macrophage receptor, CD36, was found to bind oxidized LDL with high affinity. It belongs to the family of class B scavenger receptors, which includes SR-BI and LIMP-2. CD36 is a multifunctional receptor that is also referred to as the *collagen type I receptor*, and the *thrombospondin receptor*. In phagocytic cells, it is a scavenger receptor that recognizes and removes apoptotic cells. Among the numerous roles attributed to it, CD36 is also thought to function as a fatty acid translocase (FAT) in muscle and adipose tissue. CD36 knockout mice develop hyperlipidemia, and transgenic animals develop hypolipidemia. Supporting the notion that CD36, as a scavenger receptor for oxidized atherogenic lipoproteins, is of major importance in the development of atherosclerosis are data with CD36 knockout mice crossed with apoE-deficient animals. These animals, when put on a high fat diet, developed greatly decreased aortic lesions.

The *CD36* gene has 16 exons, spans 38 kbp, and is located on chromosome 7q21.11. It encodes a 471-amino acid-residue transmembrane protein that is highly glycosylated. Feed-forward upregulation of the *CD36* gene by oxidized LDL is *via* a PPARγ response element in its promoter that is activated by PPARγ ligands such as prostaglandin J2. CD36 is also induced by interleukin-4. During inflammation, it is downregulated *via* TGF-β1 and lipopolysaccharide.

A member of the C-type lectin gene family, the oxidized LDL (lectin-like) receptor 1 (OLR1), also referred to as *LOX1*, has been shown to bind and internalize oxidized LDL. The *OLR1* gene is located on chromosome 12p13.2 with a cluster of natural killer cell receptor genes. Its six exons span 14 kb, and encode a 273-residue protein. It is expressed in a number of tissues, including brain, testis, kidney, and aorta, with the highest levels of mRNA in lung, spinal cord, and placenta. OLR1 is upregulated by a number of cytokines. Case-control studies indicate that polymorphisms at the OLR1 locus are associated with risk of CAD and acute myocardial infarction. There have been conflicting reports on associations with Alzheimer disease.

8
Lipids and Atherosclerosis

Disorders in the transport and metabolism of lipids contribute materially to the development of arteriosclerotic heart disease, stroke, and peripheral vascular disease. The histological hallmark of atherogenesis is the lipid-filled macrophage, which is transformed into a proinflammatory element. Smooth muscle cells can also be transformed into lipid-filled cells that participate in atherogenesis. The accumulation of oxidized lipids leads to the formation of neoantigens and increases oxidative stress, evoking an array of inflammatory mechanisms, ultimately leading to the attenuation of coronary blood flow or its cessation in some regions of the coronary vasculature, resulting in myocardial infarction.

The cell biology of these processes is highly complex, involving many factors that reflect genomic dependency. Rapid progress in the study of the genomic basis of disorders of lipid metabolism and of coronary disease is leading to the investigation of novel genes, thereby opening new avenues for understanding of the mechanisms underlying vascular diseases.

See also Adipocytes; Gene Therapy and Cardiovascular Diseases; Intracellular Fatty Acid Binding Proteins in Metabolic Regulation.

Bibliography

Books and Reviews

Brunzell, J.D., Deeb, S.S. (2001) Familial Lipoprotein Lipase Deficiency, Apo C-II Deficiency, and Hepatic Lipase Deficiency, in: Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D. (Eds.) *The Metabolic and Molecular Bases of Inherited Disease*, 8th edition, McGraw-Hill, New York, pp. 2789–2816.

Goldstein, J.L., Hobbs, H.H., Brown, M.S. (2001) Familial Hypercholesterolemia, in: Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D. (Eds.) *The Metabolic and Molecular Bases of Inherited Disease*, 8th edition, McGraw-Hill, New York, pp. 2863–2913.

Havel, R.J., Kane, J.P. (2001) Introduction: Structure and Metabolism of Plasma Lipoproteins, in: Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D. (Eds.) *The Metabolic and Molecular Bases of Inherited Disease*, 8th edition, McGraw-Hill, New York, pp. 2705–2716.

Kane, J.P., Havel, R.J. (2001) Disorders of the Biogenesis and Secretion of Lipoproteins Containing the B-apolipoproteins, in: Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D. (Eds.) *The Metabolic and Molecular Bases of Inherited Disease*, 8th edition, McGraw-Hill, New York, pp. 2717–2752.

Pullinger, C.R., Kane, J.P., Malloy, M.J. (2003) Primary hypercholesterolemia: genetic causes and treatment of five monogenic disorders, *Expert Rev. Cardiovasc. Ther.* **1**, 107–119.

Tall, A.R., Breslow, J.L., Rubin, E.M. (2001) Genetic Disorders Affecting High-Density Lipoprotein Metabolism, in: Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D. (Eds.) *The Metabolic and Molecular Bases of Inherited Disease*, 8th edition, McGraw Hill, New York, pp. 2915–2936.

Utermann, G. (2001) Lipoprotein(a), in: Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D. (Eds.) *The Metabolic and Molecular Bases of Inherited Disease*, 8th edition, McGraw-Hill, New York, pp. 2753–2787.

Primary Literature

Aalto-Setala, K., Weinstock, P.H., Bisgaier, C.L., Wu, L., Smith, J.D., Breslow, J.L. (1996) Further characterization of the metabolic properties of triglyceride-rich lipoproteins from human and mouse apoC-III transgenic mice, *J. Lipid Res.* **37**, 1802–1811.

Abifadel, M., Varret, M., Rabes, J.P., Allard, D., Ouguerram, K., Devillers, M., Cruaud, C., Benjannet, S., Wickham, L., Erlich, D., Derre, A., Villeger, L., Farnier, M., Beucler, I., Bruckert, E., Chambaz, J.,

Chanu, B., Lecerf, J.M., Luc, G., Moulin, P., Weissenbach, J., Prat, A., Krempf, M., Junien, C., Seidah, N.G., Boileau, C. (2003) Mutations in PCSK9 cause autosomal dominant hypercholesterolemia, *Nat. Genet.* **34**, 154–156.

Adimoolam, S., Jin, L., Grabbe, E., Shieh, J.J., Jonas, A. (1998) Structural and functional properties of two mutants of lecithin-cholesterol acyltransferase (T123I and N228K), *J. Biol. Chem.* **273**, 32561–32567.

Allan, C.M., Walker, D., Segrest, J.P., Taylor, J.M. (1995) Identification and characterization of a new human gene (APOC4) in the apolipoprotein E, C-I, and C-II gene locus, *Genomics* **28**, 291–300.

Altmann, S.W., Davis, H.R., Jr., Zhu, L.J., Yao, X., Hoos, L.M., Tetzloff, G., Iyer, S.P., Maguire, M., Golovko, A., Zeng, M., Wang, L., Murgolo, N., Graziano, M.P. (2004) Niemann-Pick C1 Like 1 protein is critical for intestinal cholesterol absorption, *Science* **303**, 1201–1204.

Anderson, C.M., Townley, R.R., Freemanm, Johansen, P. (1961) Unusual causes of steatorrhea in infancy and childhood, *Med. J. Aust.* (2), 617–622.

Aouizerat, B.E., Allayee, H., Bodnar, J., Krass, K.L., Peltonen, L., de Bruin, T.W., Rotter, J.I., Lusis, A.J. (1999) Novel genes for familial combined hyperlipidemia, *Curr. Opin. Lipidol.* **10**, 113–122.

Beffert, U., Stolt, P.C., Herz, J. (2004) Functions of lipoprotein receptors in neurons, *J. Lipid. Res.* **45**, 403–409.

Berge, K.E., Tian, H., Graf, G.A., Yu, L., Grishin, N.V., Schultz, J., Kwiterovich, P., Shan, B., Barnes, R., Hobbs, H.H. (2000) Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters, [Comment In: *Science* 2000 Dec 1;**290**(5497), 1709–11 UI: 20559920] *Science* **290**, 1771–1775.

Bhattacharyya, A.K., Connor, W.E. (1974) Beta-sitosterolemia and xanthomatosis. A newly described lipid storage disease in two sisters, *J. Clin. Invest.* **53**, 1033–1043.

Bodzioch, M., Orsó, E., Klucken, J., Langmann, T., Böttcher, A., Diederich, W., Drobnik, W., Barlage, S., Büchler, C., Porsch-Ozcürümez, M., Kaminski, W.E., Hahmann, H.W., Oette, K., Rothe, G., Aslanidis, C., Lackner, K.J., Schmitz, G. (1999) The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease [see comments], *Nat. Genet.* **22**, 347–351.

Boren, J., Ekstrom, U., Agren, B., Nilsson-Ehle, P., Innerarity, T.L. (2001) The molecular mechanism for the genetic disorder familial defective apolipoprotein B100, *J. Biol. Chem.* **276**, 9214–9218.

Breitling, R., Krisans, S.K. (2002) A second gene for peroxisomal HMG-CoA reductase? A genomic reassessment, *J. Lipid Res.* **43**, 2031–2036.

Briggs, M.R., Yokoyama, C., Wang, X., Brown, M.S., Goldstein, J.L. (1993) Nuclear protein that binds sterol regulatory element of low density lipoprotein receptor promoter. I. Identification of the protein and delineation of its target nucleotide sequence, *J. Biol. Chem.* **268**, 14490–14496.

Brooks-Wilson, A., Marcil, M., Clee, S.M., Zhang, L.H., Roomp, K., van Dam, M., Yu, L., Brewer, C., Collins, J.A., Molhuizen, H.O., Loubser, O., Ouelette, B.F., Fichter, K., Ashbourne-Excoffon, K.J., Sensen, C.W., Scherer, S., Mott, S., Denis, M., Martindale, D., Frohlich, J., Morgan, K., Koop, B., Pimstone, S., Kastelein, J.J., Hayden, M.R. et al (1999) Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency [see comments], *Nat. Genet.* **22**, 336–345.

Brunzell, J.D., Deeb, S.S. (2001) Familial Lipoprotein Lipase Deficiency, Apo C-II Deficiency, and Hepatic Lipase Deficiency, in: Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D. (Eds.) *The Metabolic and Molecular Bases of Inherited Disease*, 8th edition, McGraw-Hill, New York, pp. 2789–2816.

Buhman, K.K., Accad, M., Novak, S., Choi, R.S., Wong, J.S., Hamilton, R.L., Turley, S., Farese, R.V. Jr. (2000) Resistance to diet-induced hypercholesterolemia and gallstone formation in ACAT2-deficient mice, *Nat. Med.* **6**, 1341–1347.

Carstea, E.D., Morris, J.A., Coleman, K.G., Loftus, S.K., Zhang, D., Cummings, C., Gu, J., Rosenfeld, M.A., Pavan, W.J., Krizman, D.B., Nagle, J., Polymeropoulos, M.H., Sturley, S.L., Ioannou, Y.A., Higgins, M.E., Comly, M., Cooney, A., Brown, A., Kaneski, C.R., Blanchette-Mackie, E.J., Dwyer, N.K., Neufeld, E.B., Chang, T.Y., Liscum, L., Tagle, D.A. et al (1997) Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis, *Science* **277**, 228–231.

Chang, T.Y., Chang, C.C., Lu, X., Lin, S. (2001) Catalysis of ACAT may be completed within the plane of the membrane: a working hypothesis, *J. Lipid Res.* **42**, 1933–1938.

Chatterton, J.E., Phillips, M.L., Curtiss, L.K., Milne, R., Fruchart, J.C., Schumaker, V.N. (1995) Immunoelectron microscopy of low density lipoproteins yields a ribbon and bow model for the conformation of apolipoprotein B on the lipoprotein surface, *J. Lipid Res.* **36**, 2027–2037.

Chen, W.J., Goldstein, J.L., Brown, M.S. (1990) NPXY, a sequence often found in cytoplasmic tails, is required for coated pit-mediated internalization of the low density lipoprotein receptor, *J. Biol. Chem.* **265**, 3116–3123.

Chester, A., Scott, J., Anant, S., Navaratnam, N. (2000) RNA editing: cytidine to uridine conversion in apolipoprotein B mRNA, *Biochim. Biophys. Acta* **1494**, 1–13.

Chiang, J.Y. (2002) Bile acid regulation of gene expression: roles of nuclear hormone receptors, *Endocr. Rev.* **23**, 443–463.

Chiang, J.Y., Kimmel, R., Stroup, D. (2001) Regulation of cholesterol 7alpha-hydroxylase gene (CYP7A1) transcription by the liver orphan receptor (LXRalpha), *Gene* **262**, 257–265.

Chinetti, G., Fruchart, J.C., Staels, B. (2000) Peroxisome proliferator-activated receptors (PPARs): nuclear receptors at the crossroads between lipid metabolism and inflammation, *Inflamm. Res.* **49**, 497–505.

Cohen, J.C. (1999) Contribution of cholesterol 7alpha-hydroxylase to the regulation of lipoprotein metabolism, *Curr. Opin. Lipidol.* **10**, 303–307.

Cohen, P., Miyazaki, M., Socci, N.D., Hagge-Greenberg, A., Liedtke, W., Soukas, A.A., Sharma, R., Hudgins, L.C., Ntambi, J.M., Friedman, J.M. (2002) Role for stearoyl-CoA desaturase-1 in leptin-mediated weight loss, *Science* **297**, 240–243.

Cohn, J.S., Tremblay, M., Boulet, L., Jacques, H., Davignon, J., Roy, M., Bernier, L. (2003) Plasma concentration and lipoprotein distribution of ApoC-I is dependent on ApoE genotype rather than the Hpa I ApoC-I promoter polymorphism, *Atherosclerosis* **169**, 63–70.

Connelly, P.W., Hegele, R.A. (1998) Hepatic lipase deficiency, *Crit. Rev. Clin. Lab. Sci.* **35**, 547–572.

Couture, P., Otvos, J.D., Cupples, L.A., Lahoz, C., Wilson, P.W., Schaefer, E.J., Ordovas, J.M. (2000) Association of the C-514T polymorphism in the hepatic lipase gene with variations in lipoprotein subclass profiles: The Framingham Offspring Study, *Arterioscler. Thromb. Vasc. Biol.* **20**, 815–822.

Duchateau, P.N., Pullinger, C.R., Orellana, R.E., Kunitake, S.T., Naya-Vigne, J., O'Connor, P.M., Malloy, M.J., Kane, J.P. (1997) Apolipoprotein L, a new human high-density lipoprotein apolipoprotein expressed by the pancreas. Identification, cloning, characterization, and plasma distribution of apolipoprotein L, *J. Biol. Chem.* **272**, 25576–25582.

Dugi, K.A., Dichek, H.L., Santamarina-Fojo, S. (1995) Human hepatic and lipoprotein lipase: the loop covering the catalytic site mediates lipase substrate specificity, *J. Biol. Chem.* **270**, 25396–25401.

Edwards, P.A., Kast, H.R., Anisfeld, A.M. (2002) BAREing it all: the adoption of LXR and FXR and their roles in lipid homeostasis, *J. Lipid Res.* **43**, 2–12.

Endemann, G., Stanton, L.W., Madden, K.S., Bryant, C.M., White, R.T., Protter, A.A. (1993) CD36 is a receptor for oxidized low density lipoprotein, *J. Biol. Chem.* **268**, 11811–11816.

Erickson, S.K., Lear, S.R., Deane, S., Dubrac, S., Huling, S.L., Nguyen, L., Bollineni, J.S., Shefer, S., Hyogo, H., Cohen, D.E., Shneider, B., Sehayek, E., Ananthanarayanan, M., Balasubramaniyan, N., Suchy, F.J., Batta, A.K., Salen, G. (2003) Hypercholesterolemia and changes in lipid and bile acid metabolism in male and female cyp7A1-deficient mice, *J. Lipid Res.* **44**, 1001–1009.

Febbraio, M., Podrez, E.A., Smith, J.D., Hajjar, D.P., Hazen, S.L., Hoff, H.F., Sharma, K., Silverstein, R.L. (2000) Targeted disruption of the class B scavenger receptor CD36 protects against atherosclerotic lesion development in mice, *J. Clin. Invest.* **105**, 1049–1056.

Forman, B.M., Goode, E., Chen, J., Oro, A.E., Bradley, D.J., Perlmann, T., Noonan, D.J., Burka, L.T., McMorris, T., Lamph, W.W. (1995) Identification of a nuclear receptor that is activated by farnesol metabolites, *Cell* **81**, 687–693.

Garcia, C.K., Wilund, K., Arca, M., Zuliani, G., Fellin, R., Maioli, M., Calandra, S., Bertolini, S., Cossu, F., Grishin, N., Barnes, R., Cohen, J.C., Hobbs, H.H. (2001) Autosomal

recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein, *Science* **292**, 1394–1398.

Gillespie, J.G., Hardie, D.G. (1992) Phosphorylation and inactivation of HMG-CoA reductase at the AMP-activated protein kinase site in response to fructose treatment of isolated rat hepatocytes, *FEBS Lett.* **306**, 59–62.

Goldstein, J.L., Schrott, H.G., Hazzard, W.R., Bierman, E.L., Motulsky, A.G. (1973) Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia, *J. Clin. Invest.* **52**, 1544–1568.

Graf, G.A., Li, W.P., Gerard, R.D., Gelissen, I., White, A., Cohen, J.C., Hobbs, H.H. (2002) Coexpression of ATP-binding cassette proteins ABCG5 and ABCG8 permits their transport to the apical surface, *J. Clin. Invest.* **110**, 659–669.

Groenendijk, M., Cantor, R.M., de Bruin, T.W., Dallinga-Thie, G.M. (2001) The apoAI-CIII-AIV gene cluster, *Atherosclerosis* **157**, 1–11.

Grundy, S.M., Vega, G.L. (1985) Influence of mevinolin on metabolism of low density lipoproteins in primary moderate hypercholesterolemia, *J. Lipid. Res.* **26**, 1464–1475.

Hall, S., Chu, G., Miller, G., Cruickshank, K., Cooper, J.A., Humphries, S.E., Talmud, P.J. (1997) A common mutation in the lipoprotein lipase gene promoter, −93T/G, is associated with lower plasma triglyceride levels and increased promoter activity in vitro, *Arterioscler. Thromb. Vasc. Biol.* **17**, 1969–1976.

Hegele, R.A., Little, J.A., Vezina, C., Maguire, G.F., Tu, L., Wolever, T.S., Jenkins, D.J., Connelly, P.W. (1993) Hepatic lipase deficiency. Clinical, biochemical, and molecular genetic characteristics, *Arterioscler. Thromb.* **13**, 720–728.

Hirata, K., Dichek, H.L., Cioffi, J.A., Choi, S.Y., Leeper, N.J., Quintana, L., Kronmal, G.S., Cooper, A.D., Quertermous, T. (1999) Cloning of a unique lipase from endothelial cells extends the lipase gene family, *J. Biol. Chem.* **274**, 14170–14175.

Holt, J.A., Luo, G., Billin, A.N., Bisi, J., McNeill, Y.Y., Kozarsky, K.F., Donahee, M., Wang da, Y., Mansfield, T.A., Kliewer, S.A., Goodwin, B., Jones, S.A. (2003) Definition of a novel growth factor-dependent signal cascade for the suppression of bile acid biosynthesis, *Genes Dev.* **17**, 1581–1591.

Innerarity, T.L., Balestra, M.E., Arnold, K.S., Mahley, R.W., Vega, G.L., Grundy, S.M., Young, S.G. (1988) Isolation of defective receptor-binding low density lipoproteins from subjects with familial defective apolipoprotein B-100, *Arteriosclerosis* **8**, 551a.

Issemann, I., Green, S. (1990) Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators, *Nature* **347**, 645–650.

Janowski, B.A., Willy, P.J., Devi, T.R., Falck, J.R., Mangelsdorf, D.J. (1996) An oxysterol signalling pathway mediated by the nuclear receptor LXR alpha, *Nature* **383**, 728–731.

Jaye, M., Lynch, K.J., Krawiec, J., Marchadier, D., Maugeais, C., Doan, K., South, V., Amin, D., Perrone, M., Rader, D.J. (1999) A novel endothelial-derived lipase that modulates HDL metabolism, *Nat. Genet.* **21**, 424–428.

Jiang, X.C., Bruce, C., Mar, J., Lin, M., Ji, Y., Francone, O.L., Tall, A.R. (1999) Targeted mutation of plasma phospholipid transfer protein gene markedly reduces high-density lipoprotein levels, *J. Clin. Invest.* **103**, 907–914.

Jira, P.E., Waterham, H.R., Wanders, R.J., Smeitink, J.A., Sengers, R.C., Wevers, R.A. (2003) Smith-Lemli-Opitz syndrome and the DHCR7 gene, *Ann. Hum. Genet.* **67**, 269–280.

Jones, B., Jones, E.L., Bonney, S.A., Patel, H.N., Mensenkamp, A.R., Eichenbaum-Voline, S., Rudling, M., Myrdal, U., Annesi, G., Naik, S., Meadows, N., Quattrone, A., Islam, S.A., Naoumova, R.P., Angelin, B., Infante, R., Levy, E., Roy, C.C., Freemont, P.S., Scott, J., Shoulders, C.C. (2003) Mutations in a Sar1 GTPase of COPII vesicles are associated with lipid absorption disorders, *Nat. Genet.* **34**, 29–31.

Kane, J.P., Aouizerat, B.E., Luke, M.M., Shiffman, D., Iakoubova, O., Liu, D., Rowland, C.M., Catanese, J.J., Leong, D.U., Lau, K.F., Louie, J.Z., Tong, C.H., McAllister, L.B., Dabby, L.F., Ports, T.A., Michaels, A.D., Zellner, C., Pullinger, C.R., Malloy, M.J., Devlin, J.J. (2004) Novel Genetic Markers for Structural Coronary Artery Disease, Myocardial Infarction, and Familial Combined Hyperlipidemia: Candidate and Genome Scans of Functional SNPS, in: Matsuzawa, T.K.Y., Nagai, R., Teramoto, T. (Eds.) *Atherosclerosis XIII: Proceedings of the 13th International Symposium*, Elsevier, Kyoto, Japan, pp. 309–312.

Karathanasis, S., Ferris, E., Haddad, I. (1987) DNA inversion with the apolipoproteins

AI/CIII/AIV encoding gene clusters of certain patients with premature atherosclerosis, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7198–7202.

Kast, H.R., Nguyen, C.M., Sinal, C.J., Jones, S.A., Laffitte, B.A., Reue, K., Gonzalez, F.J., Willson, T.M., Edwards, P.A. (2001) Farnesoid X-activated receptor induces apolipoprotein C-II transcription: a molecular mechanism linking plasma triglyceride levels to bile acids, *Mol. Endocrinol.* **15**, 1720–1728.

Kelley, R.L., Roessler, E., Hennekam, R.C., Feldman, G.L., Kosaki, K., Jones, M.C., Palumbos, J.C., Muenke, M. (1996) Holoprosencephaly in RSH/Smith-Lemli-Opitz syndrome: does abnormal cholesterol metabolism affect the function of Sonic hedgehog? *Am. J. Med. Genet.* **66**, 478–484.

Koschinsky, M.L., Marcovina, S.M. (2004) Structure-function relationships in apolipoprotein(a): insights into lipoprotein(a) assembly and pathogenicity, *Curr. Opin. Lipidol.* **15**, 167–174.

Kuivenhoven, J.A., Pritchard, H., Hill, J., Frohlich, J., Assmann, G., Kastelein, J. (1997) The molecular pathology of lecithin:cholesterol acyltransferase (LCAT) deficiency syndromes, *J. Lipid Res.* **38**, 191–205.

Langer, T., Strober, W., Levy, R.I. (1972) The metabolism of low density lipoprotein in familial type II hyperlipoproteinemia, *J. Clin. Invest.* **51**, 1528–1536.

Lawn, R.M., Wade, D.P., Garvin, M.R., Wang, X., Schwartz, K., Porter, J.G., Seilhamer, J.J., Vaughan, A.M., Oram, J.F. (1999) The Tangier disease gene product ABC1 controls the cellular apolipoprotein-mediated lipid removal pathway [see comments], *J. Clin. Invest.* **104**, R25–R31.

Lee, M.H., Lu, K., Hazard, S., Yu, H., Shulenin, S., Hidaka, H., Kojima, H., Allikmets, R., Sakuma, N., Pegoraro, R., Srivastava, A.K., Salen, G., Dean, M., Patel, S.B. (2001) Identification of a gene, ABCG5, important in the regulation of dietary cholesterol absorption, *Nat. Genet.* **27**, 79–83.

Li, B.L., Li, X.L., Duan, Z.J., Lee, O., Lin, S., Ma, Z.M., Chang, C.C., Yang, X.Y., Park, J.P., Mohandas, T.K., Noll, W., Chan, L., Chang, T.Y. (1999) Human acyl-CoA:cholesterol acyltransferase-1 (ACAT-1) gene organization and evidence that the 4.3-kilobase ACAT-1 mRNA is produced from two different chromosomes, *J. Biol. Chem.* **274**, 11060–11071.

Linton, M.F., Pierotti, V., Young, S.G. (1992) Reading frame restoration with an apolipoprotein B gene frameshift mutation, *Proc. Natl. Acad. Sci. U.S.A.*; **89**, 11431–11435.

Lu, H., Inazu, A., Moriyama, Y., Higashikata, T., Kawashiri, M.A., Yu, W., Huang, Z., Okamura, T., Mabuchi, H. (2003) Haplotype analyses of cholesteryl ester transfer protein gene promoter: a clue to an unsolved mystery of TaqIB polymorphism, *J. Mol. Med.* **81**, 246–255.

Lu, T.T., Makishima, M., Repa, J.J., Schoonjans, K., Kerr, T.A., Auwerx, J., Mangelsdorf, D.J. (2000) Molecular basis for feedback regulation of bile acid synthesis by nuclear receptors, *Mol. Cell.* **6**, 507–515.

Mak, P.A., Laffitte, B.A., Desrumaux, C., Joseph, S.B., Curtiss, L.K., Mangelsdorf, D.J., Tontonoz, P., Edwards, P.A. (2002) Regulated expression of the apolipoprotein E/C-I/C-IV/C-II gene cluster in murine and human macrophages. A critical role for nuclear liver X receptors alpha and beta, *J. Biol. Chem.* **277**, 31900–31908.

Mar, R., Pajukanta, P., Allayee, H., Groenendijk, M., Dallinga-Thie, G., Krauss, R.M., Sinsheimer, J.S., Cantor, R.M., de Bruin, T.W., Lusis, A.J. (2004) Association of the APOLIPOPROTEIN A1/C3/A4/A5 gene cluster with triglyceride levels and LDL particle size in familial combined hyperlipidemia, *Circ. Res.* **94**, 993–999.

Mehta, A., Kinter, M.T., Sherman, N.E., Driscoll, D.M. (2000) Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA, *Mol. Cell. Biol.* **20**, 1846–1854.

Mimmack, M.L., Ryan, M., Baba, H., Navarro-Ruiz, J., Iritani, S., Faull, R.L., McKenna, P.J., Jones, P.B., Arai, H., Starkey, M., Emson, P.C., Bahn, S. (2002) Gene expression analysis in schizophrenia: reproducible up-regulation of several members of the apolipoprotein L family located in a high-susceptibility locus for schizophrenia on chromosome 22, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 4680–4685.

Mishra, S.K., Watkins, S.C., Traub, L.M. (2002) The autosomal recessive hypercholesterolemia (ARH) protein interfaces directly with the clathrin-coat machinery, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16099–16104.

Naureckiene, S., Sleat, D.E., Lackland, H., Fensom, A., Vanier, M.T., Wattiaux, R., Jadot, M., Lobel, P. (2000) Identification of HE1 as the second gene of Niemann-Pick C disease, *Science* **290**, 2298–2301.

Nishimaki-Mogami, T., Une, M., Fujino, T., Sato, Y., Tamehiro, N., Kawahara, Y., Shudo, K., Inoue, K. (2004) Identification of intermediates in the bile acid synthetic pathway as ligands for the farnesoid X receptor, *J. Lipid. Res.* **45**, 1538–1545.

Paszty, C., Maeda, N., Verstuyft, J., Rubin, E.M. (1994) Apolipoprotein AI transgene corrects apolipoprotein E deficiency-induced atherosclerosis in mice, *J. Clin. Invest.* **94**, 899–903.

Pennacchio, L.A., Olivier, M., Hubacek, J.A., Cohen, J.C., Cox, D.R., Fruchart, J.C., Krauss, R.M., Rubin, E.M. (2001) An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing, [Comment In: *Science*. 2001 Oct 5; **294**(5540), 33 UI: 21472613 Comment In: *Science*. 2001 Oct 5; **294**(5540), 33 UI: 21472614] *Science* **294**, 169–173.

Porter, J.A., Young, K.E., Beachy, P.A. (1996) Cholesterol modification of hedgehog signaling proteins in animal development, *Science* **274**, 255–259.

Pullinger, C.R., Malloy, M.J., Shahidi, A.K., Ghassemzadeh, M., Duchateau, P., Villagomez, J., Allaart, J., Kane, J.P. (1997) A novel apolipoprotein C-III variant, apoC-III(Gln38→Lys), associated with moderate hypertriglyceridemia in a large kindred of Mexican origin, *J. Lipid Res.* **38**, 1833–1840.

Pullinger, C.R., Hennessy, L.K., Chatterton, J.E., Liu, W.Q., Love, J.A., Mendel, C.M., Frost, P.H., Malloy, M.J., Schumaker, V.N., Kane, J.P. (1995) Familial ligand-defective apolipoprotein B – Identification of a new mutation that decreases LDL receptor binding affinity, *J. Clin. Invest.* **95**, 1225–1234.

Pullinger, C.R., Eng, C., Salen, G., Shefer, S., Batta, A.K., Erickson, S.K., Verhagen, A., Rivera, C.R., Mulvihill, S.J., Mary, J., Malloy, M.J., Kane, J.P. (2002) Human Cholesterol 7α-Hydroxylase (CYP7A1) Deficiency has a Hypercholesterolemic Phenotype, *J. Clin. Invest.* **110**, 109–117.

Qin, S., Kawano, K., Bruce, C., Lin, M., Bisgaier, C., Tall, A.R., Jiang, X. (2000) Phospholipid transfer protein gene knock-out mice have low-high density lipoprotein levels, due to hypercatabolism, and accumulate apoA-IV-rich lamellar lipoproteins, *J. Lipid Res.* **41**, 269–276.

Rassart, E., Bedirian, A., Do Carmo, S., Guinard, O., Sirois, J., Terrisse, L., Milne, R. (2000) Apolipoprotein D, *Biochim. Biophys. Acta* **1482**, 185–198.

Rawson, R.B. (2003) The SREBP pathway–insights from Insigs and insects, *Nat. Rev. Mol. Cell Biol.* **4**, 631–640.

Remaley, A.T., Schumacher, U.K., Stonik, J.A., Farsi, B.D., Nazih, H., Brewer, H.B. Jr. (1997) Decreased reverse cholesterol transport from Tangier disease fibroblasts. Acceptor specificity and effect of brefeldin on lipid efflux, *Arterioscler. Thromb. Vasc. Biol.* **17**, 1813–1821.

Rose, H.G., Kranz, P., Weinstock, M., Juliano, J., Haft, J.I. (1973) Inheritance of combined hyperlipoproteinemia:evidence for a new lipoprotein phenotype, *Am. J. Med.* **54**, 148–160.

Rudenko, G., Henry, L., Henderson, K., Ichtchenko, K., Brown, M.S., Goldstein, J.L., Deisenhofer, J. (2002) Structure of the LDL receptor extracellular domain at endosomal pH, *Science* **298**, 2353–2358.

Rust, S., Rosier, M., Funke, H., Real, J., Amoura, Z., Piette, J.C., Deleuze, J.F., Brewer, H.B., Duverger, N., Denefle, P., Assmann, G. (1999) Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1, *Nat. Genet.* **22**, 352–355.

Schonfeld, G. (2003) Familial hypobetalipoproteinemia: a review, *J. Lipid Res.* **44**, 878–883.

Schwarz, M., Lund, E.G., Setchell, K.D.R., Kayden, H.J., Zerwekh, J.E., Björkhem, I., Herz, J., Russell, D.W. (1996) Disruption of cholesterol 7alpha-hydroxylase gene in mice. II. Bile acid deficiency is overcome by induction of oxysterol 7alpha-hydroxylase, *J Biol. Chem.* **271**, 18024–18031.

Segrest, J.P., Jones, M.K., De Loof, H., Dashti, N. (2001) Structure of apolipoprotein B-100 in low density lipoproteins, *J. Lipid Res.* **42**, 1346–1367.

Setchell, K.D., Schwarz, M., O'Connell, N.C., Lund, E.G., Davis, D.L., Lathe, R., Thompson, H.R., Weslie Tyson, R., Sokol, R.J., Russell, D.W. (1998) Identification of a new inborn

error in bile acid synthesis: mutation of the oxysterol 7alpha-hydroxylase gene causes severe neonatal liver disease, *J. Clin. Invest.* **102**, 1690–1703.

Shen, P., Howlett, G.J. (1992) Two coding regions closely linked to the rat apolipoprotein E gene: nucleotide sequences of rat apolipoprotein C-I and ECL cDNA, *Arch. Biochem. Biophys.* **297**, 345–353.

Silver, D.L., Tall, A.R. (2001) The cellular biology of scavenger receptor class B type I, *Curr. Opin. Lipidol.* **12**, 497–504.

Simonet, W.S., Bucay, N., Lauer, S.J., Taylor, J.M. (1993) A far-downstream hepatocyte-specific control region directs expression of the linked human apolipoprotein E and C-I genes in transgenic mice, *J. Biol. Chem.* **268**, 8221–8229.

Smith, D.W., Lemli, L., Opitz, J.M. (1964) A newly recognized syndrome of multiple congenital anomalies, *J. Pediatr.* **64**, 210–217.

Smith, J.L., Hardie, I.R., Pillay, S.P., de Jersey, J. (1990) Hepatic acyl-coenzyme A:cholesterol acyltransferase activity is decreased in patients with cholesterol gallstones, *J. Lipid Res.* **31**, 1993–2000.

Smith, J.L., Rangaraj, K., Simpson, R., Maclean, D.J., Nathanson, L.K., Stuart, K.A., Scott, S.P., Ramm, G.A., de Jersey, J. (2004) Quantitative analysis of the expression of ACAT genes in human tissues by real-time PCR, *J. Lipid Res.* **45**, 686–696.

Soria, L.F., Ludwig, E.H., Clarke, H.R.G., Vega, G.L., Grundy, S.M., McCarthy, B.J. (1989) Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 587–591.

Steffensen, K.R., Gustafsson, J.A. (2004) Putative metabolic effects of the liver X receptor (LXR), *Diabetes* **53**(1), S36–S42.

Tacken, P.J., Hofker, M.H., Havekes, L.M., van Dijk, K.W. (2001) Living up to a name: the role of the VLDL receptor in lipid metabolism, *Curr. Opin. Lipidol.* **12**, 275–279.

van der Vliet, H.N., Sammels, M.G., Leegwater, A.C., Levels, J.H., Reitsma, P.H., Boers, W., Chamuleau, R.A. (2001) Apolipoprotein A-V: a novel apolipoprotein associated with an early phase of liver regeneration, *J. Biol. Chem.* **276**, 44512–44520.

Vanhamme, L., Paturiaux-Hanocq, F., Poelvoorde, P., Nolan, D.P., Lins, L., Van Den Abbeele, J., Pays, A., Tebabi, P., Van Xong, H., Jacquet, A., Moguilevsky, N., Dieu, M., Kane, J.P., De Baetselier, P., Brasseur, R., Pays, E. (2003) Apolipoprotein L-I is the trypanosome lytic factor of human serum, *Nature* **422**, 83–87.

Veniant, M.M., Withycombe, S., Young, S.G. (2001) Lipoprotein size and atherosclerosis susceptibility in Apoe(−/−) and Ldlr(−/−) mice, *Arterioscler. Thromb. Vasc. Biol.* **21**, 1567–1570.

Venkatesan, S., Cullen, P., Pacy, P., Halliday, D., Scott, J. (1993) Stable isotopes show a direct relation between VLDL apoB overproduction and serum triglyceride levels and indicate a metabolically and biochemically coherent basis for familial combined hyperlipidemia, *Arterioscler. Thromb.* **13**, 1110–1118.

Wetterau, J.R., Combs, K.A., McLean, L.R., Spinner, S.N., Aggerbeck, L.P. (1991) Protein disulfide isomerase appears necessary to maintain the catalytically active structure of the microsomal triglyceride transfer protein, *Biochemistry* **30**, 9728–9735.

Wetterau, J.R., Aggerbeck, L.P., Bouma, M.E., Eisenberg, C., Munck, A., Hermier, M., Schmitz, J., Gay, G., Rader, D.J., Gregg, R.E. (1992) Absence of microsomal triglyceride transfer protein in individuals with abetalipoproteinemia, *Science* **258**, 999–1001.

Yamanaka, S., Zhang, X.Y., Miura, K., Kim, S., Iwao, H. (1998) The human gene encoding the lectin-type oxidized LDL receptor (OLR1) is a novel member of the natural killer gene complex with a unique expression profile, *Genomics* **54**, 191–199.

Yang, T., Espenshade, P.J., Wright, M.E., Yabe, D., Gong, Y., Aebersold, R., Goldstein, J.L., Brown, M.S. (2002) Crucial step in cholesterol homeostasis: sterols promote binding of SCAP to INSIG-1, a membrane protein that facilitates retention of SREBPs in ER, *Cell* **110**, 489–500.

Yokode, M., Pathak, R.K., Hammer, R.E., Brown, M.S., Goldstein, J.L., Anderson, R.G. (1992) Cytoplasmic sequence required for basolateral targeting of LDL receptor in livers of transgenic mice, *J. Cell Biol.* **117**, 39–46.

Zannis, V.I., Kan, H.Y., Kritis, A., Zanni, E., Kardassis, D. (2001) Transcriptional regulation of the human apolipoprotein genes, *Front Biosci.* **6**, D456–D504.

Zuliani, G., Arca, M., Signore, A., Bader, G., Fazio, S., Chianelli, M., Bellosta, S., Campagna, F., Montali, A., Maioli, M., Pacifico, A., Ricci, G., Fellin, R. (1999) Characterization of a new form of inherited hypercholesterolemia: familial recessive hypercholesterolemia, *Arterioscler. Thromb. Vasc. Biol.* **19**, 802–809.

Lipids, Microbial

Colin Ratledge
Department of Biological Sciences, University of Hull, Hull, UK

1 Introduction 250

2 Archaea (or Archaebacteria) 251

3 Bacteria 255
3.1 Unsaturated Fatty Acids 256
3.2 Polyunsaturated Fatty Acids 258
3.3 Branched Chain Fatty Acids 258
3.4 Cyclopropane Fatty Acids 259
3.5 Hydroxy Fatty Acids 259
3.6 Mycolic Acids 259
3.7 Polyesters 260
3.8 Biosurfactant Lipids 262

4 Yeasts 263
4.1 Fatty Acids and Fatty Acyl Lipids 263
4.2 Single Cell Oils (SCO) 264
4.3 Biochemistry of Oleaginicity 265
4.4 Other Lipids 267

5 Molds 268
5.1 Fatty Acids 268
5.2 Polyunsaturated Fatty Acids 268
5.3 Single Cell Oils 269
5.4 Other Lipids 272

6 Unicellular Algae 273
6.1 Cyanobacterial Lipids 273

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

6.2 Eukaryotic Algae Lipids 274
6.2.1 Carotenoids 274
6.2.2 Polyunsaturated Fatty Acids 274

Bibliography 275
Books and Reviews 275
Primary Literature 276

Keywords

Ceramide
A branched chain lipid based on a long-chain (C_{18}) polyhydroxyamine (= sphingolipid) to which a fatty acid group is attached via the 2-amino group.

Fatty Acids
Long-chain (usually 16 or 18C atoms) aliphatic carboxylic acids normally found esterified to glycerol; see **Triacylglycerol**.

Glycolipid
A sugar (usually a mono- or disaccharide) linked to one or more fatty acyl groups; acts as an emulsifying agent by having both water- and lipid-soluble components.

Lipoglycan
A macromolecular complex where the glucan or glycan (polysaccharide) backbone molecule has a number of fatty acyl substituents. Overall, the molecule retains its carbohydrate properties.

Neutral Lipid
A lipid having no charged group; encompasses mono-, di-, and triacylglycerols plus hydrocarbons, including carotenoids, sterols, and so on.

Phospholipid
Strictly, any lipid containing a phospho group, but usually refers to lipids based on 1-, 2-diacyglycerol-3-phosphate (also known as *phosphatidic acid*); a polar substituent (often choline) may also occur on the phospho group, thereby creating a range of phospholipid types. Phospholipids form the major component of most membranes of cells.

Polar Lipids
A lipid having a polar (charged) group; often taken to be synonymous with phospholipids but will include sphingolipids and other related lipids.

Polyunsaturated Fatty Acids (PUFA)
Fatty acids containing three or more double bonds that are usually methylene interrupted, namely, $-CH{=}CH-CH_2-CH{=}CH-$. Polyunsaturated fatty acids can

therefore be designated according to the position of the double bond nearest to the CH_3-terminus: thus, $CH_3-CH_2-CH=CH-$would be known as an ω-3 (or n-3) fatty acid.

Sulfolipid
Any lipid containing a sulfo group, usually $-SO_3-$.

Terpenoid Lipids
Lipids that are based on multiples of the isopentenyl group ($CH_2{:}C(CH_3)\cdot CH_2\cdot CH_2-$) giving rise to sterols, carotenoids, polyprenols, and the side chains of chlorophylls, quinones, and so on.

Triacylglycerol
Member of a class of lipids in which the three alcohol groups of glycerol ($CH_2OH\cdot CHOH\cdot CH_2OH$) are each esterified with a fatty acid. These, along with phospholipids, are the predominant lipid types in most eukaryotic cells.

Unsaturated Fatty Acids
Fatty acids containing one or more double ($-CH=CH-$) bonds. The position of the double bond is indicated by the first carbon numbering from the carboxylic acid group: $^{1}COOH\cdot{}^{2}CH_2\cdot{}^{3}CH_2\cdot{}^{4}CH_2\cdot{}^{5}CH_2\cdot{}^{6}CH_2\cdot{}^{7}CH_2\cdot{}^{8}CH_2\cdot{}^{9}CH=^{10}CH\cdot{}^{11}CH_2\cdot{}^{12}CH=$ $^{13}CH\cdot{}^{14}CH_2\cdot{}^{15}CH_2\cdot{}^{16}CH_2\cdot{}^{17}CH_2\cdot{}^{18}CH_3$ is 9, 12-octadecadienoic acid (or linoleic acid). The orientation of the double bond is usually in the *Z* (or cis) configuration though *E* (or trans) bonds are known. The shorthand nomenclature for the acid above is therefore 18 : 2 (*Z*9, *Z*12), abbreviated to 18 : 2 (9, 12) if the stereospecificity is not required to be stated. Sometimes it is convenient just to give the position of the double bond nearest the terminal methyl group because, once this is defined, the position of all the other double bonds, being methylene interrupted, are also defined: thus linoleic acid could be given as 18 : 2 (ω-6) or alternatively as 18 : 2 (n-6). This system is particularly useful for polyunsaturated fatty acids and is used in this chapter. Almost all double bonds in naturally occurring fatty acids are of the *Z*- or cis configuration.

Microorganisms, like all other living cells, contain lipids. These lipids are probably the most diverse of all the major kingdoms, and range from simple fatty acids as occurring in plants and animals, to complex multibranched ether lipids found only in the most primitive of bacteria, the *Archaea*. Lipids are used for both structural and functional activities. The former include the principal component of all membranes – whether it is of the cytoplasm or of the major organelles of the eukaryotic microorganisms: yeasts, fungi, and algae. The functional aspects of lipids include their role as storage reserve materials as well as fulfilling key roles in photosynthetic microorganisms. Both bacteria and eukaryotic microorganisms have the propensity to accumulate during periods of carbon nutrient excess large

amounts of lipids, which they can utilize during periods of starvation as sources of carbon, energy, and even water. Present biotechnological applications of microbial lipids are aimed at realizing their potential in the nutritional and health care industries. The molecular biology of lipids, though, is a relatively neglected area but is likely to become a major focus of activity as the exploitability of microbial genes for the improvement of plant oils and fats is recognized.

1 Introduction

Lipids are divided into two major classes: the components based on fatty acids (or fatty acyls) and those based on the terpene, or isoprenoid structure. The former includes all the triacylglycerols (**I**) and acylglycerophospholipids (**II**) that occur throughout all eukaryotic cells, including plants, animals as well as microorganisms except the *Archaea*. The latter class of lipids includes the sterols (**III**) and carotenoid (**IV**) lipids, which are synthesized from a common pathway known as the mevalonate or isoprenoid (**V**) route. The archaeal lipids (see Sect. 2) are also derived from mevalonate. Fig. 1 shows the biosynthetic origins of these two groups.

A further distinction that can be made with lipids is to categorize them as polar or neutral. Polar lipids, which are characteristically the lipids of the cell bilayer membranes, owe their polarity to the presence of a charged group such as ethanolamine or choline, attached to the glycerol group (**II**). These lipids are therefore often referred to as phospholipids, an ambiguous and somewhat imprecise term that merely denotes any lipid containing a phospho group. Other polar lipids that do not contain a phospho group include a number of glycolipids, involving condensation of a fatty acid with a sugar or sugar groups, sulfolipids where a SO_3^- group can be found, peptidolipids involving one or more amino acids, and a number of sphingolipids that are more usually thought of as animal-type lipids, but are found in small but significant amounts in many yeasts and fungi. The entire subject of microbial lipids, including lipids of viruses, has been covered in an extensive two-volume treatise edited by the present author and his colleague, Professor S. G. Wilkinson.

Microorganisms are now divided into three domains: *Archaea, Bacteria*, and *Eukarya*. The first two comprise all the prokaryotic organisms that were originally referred to as the *Archaebacteria* and the *Eubacteria*, respectively, and have been combined under the single order of the *Bacteria*. In view of the considerable disparity between the 16S rRNA sequences and the structures of RNA polymerases of these two orders, they are now considered to be distinct evolutionary groups and, therefore, have been given equal ranking. The *Eukarya* domain comprises a group of microorganisms loosely known as the fungi, which are then divided into the yeasts and the molds as useful, but somewhat imprecise terminologies. *Eukarya* also includes the photosynthetic algae. Blue-green algae are now referred to as cyanobacteria and are included in *Bacteria*.

The genetics and molecular biology of microbial lipids is very much in its

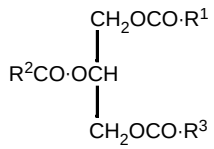

where R^1CO-, R^2CO- and R^3CO- are fatty acyl groups.

I

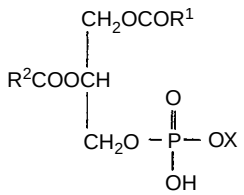

General structure

R^1CO- and R^2CO- are fatty acyl groups.
X = polar group (choline, serine, ethanolamine etc.)

II

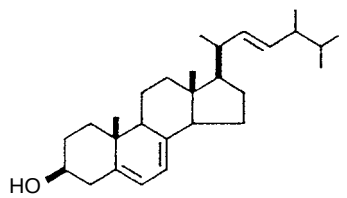

III

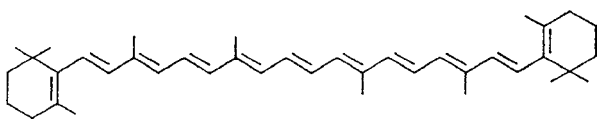

IV

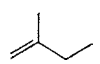

V

infancy and only a few areas have received much attention. The application of genetic techniques have mainly been confined to *Escherichia coli* and *Saccharomyces cerevisiae*, neither of which produces unusual or interesting lipids. Nevertheless, the basis of a genetic understanding of microbial lipids has now been laid. This article is aimed at providing factual information concerning the diversity of lipid molecules found in the different major groups of microorganisms.

2 Archaea (or Archaebacteria)

The distinction between the *Archaea*, the most primitive of the prokaryotic domains, and the *Bacteria* was made principally on the differences in the 16S rRNA sequences. This distinction is also clearly seen in the nature of the lipid components of the two domains. The *Archaea* uniquely contain lipids that are made up of repeating isopentanyl (**VI**) units. The isopentanyl

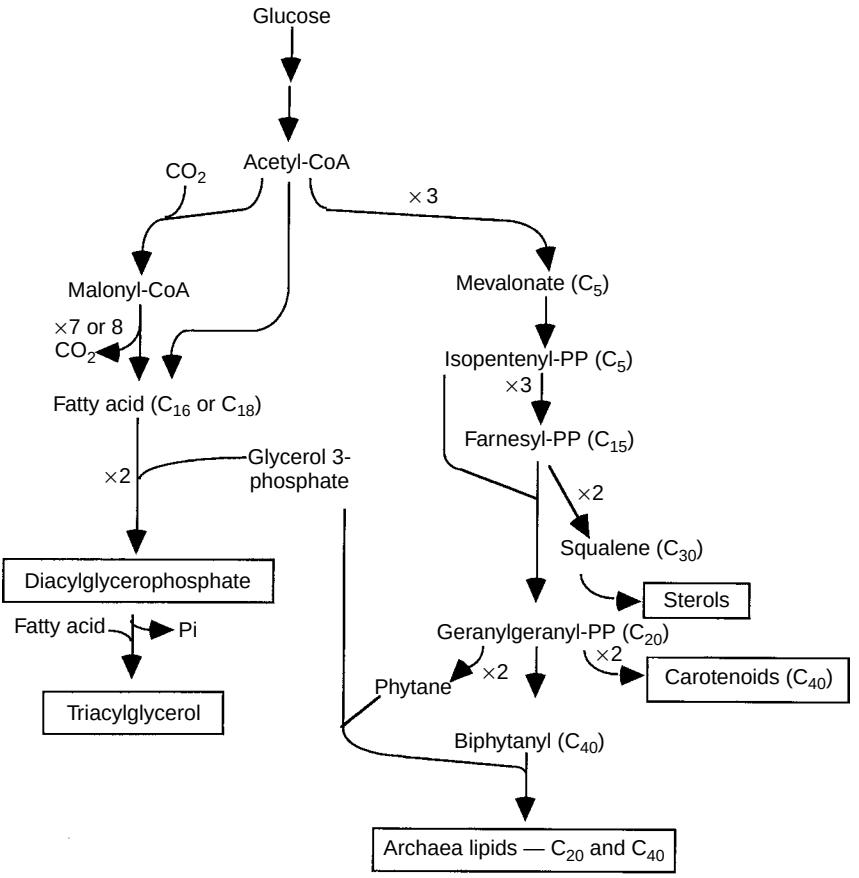

Fig. 1 Biosynthetic origins of major groups of microbial lipids.

unit is derived from the mevalonate pathway of lipid biosynthesis (see Fig. 1), but although this pathway is found in other microorganisms as a source of the terpenoid lipids, in the *Archaea* the entire lipid family are derived from this source.

VI

A further distinction that arises between the archaeal lipids and those of all other organisms is that they are linked to glycerol via ether linkages ($-CH_2-O-CH_2-$) rather than the more common ester link ($-CH_2-O-CO-$). The ether bond is much less susceptible to hydrolysis than the ester linkage and, not surprisingly, then confers considerable stability to the membrane structures of the *Archaea* that are made up with these lipids. The *Archaea* are, indeed, very resistant to extremes of the environment: they comprise the extreme halophiles (*Halobacterium, Halococcus, Natronococcus, Natronobacterium,* and so on), extreme thermophiles (*Thermus, Caldariella, Sulfobolus,* and so on), and organisms that can grow on methane – the methanotrophs

or methanogens (*Methanospirillum*, *Methanosarcina*, *Methanococcus*, and so on). The somewhat loose term "extremophile" has been given to this group of organisms as a convenient epithet, but it lacks the formal precision of *Archaea* and may often be used to include representatives of the *Bacteria* that show some tolerance to the extremes of the environment.

A further point to note in connection with these glycerol diether lipids (Fig. 2) is that the two isopranoid chains (also called *phytanyl chains*) are attached to the *sn*2- and 3- positions of glycerol rather than the conventional 1- and 2- positions of the diacylglycerols and diacylphosphoglycerols found in all other living cells, including bacteria.

Typical isopranoid ether lipids of *Archaea* are shown in Fig. 2. The lipid given in Fig. 2(a), which is the core lipid of the *Archaea*, is also referred to as *archaeol*, which indicates its association with this domain of microorganisms as well as its status as an alcohol. The lipid given in Fig. 2(c) is known as *caldarchaeol*. As Fig. 2 indicates, not only are there a variety of configurations giving rise to both diethers and tetraether lipids but there are also phosphoglycolipids and sulfolipids, as well as lipoglycans that form part of the cell wall of these microorganisms and are all composed of these phytanyl lipids.

Thus, not only are the lipid membranes of the extremophilic microorganisms composed of highly stable isopranoid ether lipids but so too are many of the lipid components associated with the cell wall.

Figure 3 gives the biosynthesis of the archaeol phospholipids and glycolipids in the extreme halophiles. The C_{20} product of the mevalonate pathway, probably, geranylgeranyl pyrophosphate is considered to condense with dihydroxyacetone and gives a "pre-diether." This becomes the precursor of both the phospholipids and the glycolipids, with the reduced, unsaturated bonds of the geranylgeranyl group being saturated after condensation with the dihydroxyacetone. The biphytanyl (C_{40}) chains are thought to arise by direct head-to-head condensation of two molecules of archaeol (Fig. 2a). The substituted glycerol and phosphoglycero lipids (Figs. 2e and f) are probably synthesized before this condensation reaction; this would include the sugar-containing lipids (glycolipids).

The function of these branched chain ether lipids is to impart stability to the membranes of these bacteria under extremes of temperature, pH, salinity, and so on. Under these extreme conditions, conventional ester-linked lipids would quickly hydrolyze; moreover, unsaturated fatty acyl groups would be quickly oxidized, and thus, "conventional" microorganisms would not, and cannot, survive under these conditions. Since the alkyl chains are completely saturated and furthermore have only short side chains (methyl groups), which allows for close molecular packing, the branched chain ether lipids become liquid crystalline at ambient temperatures. Consequently, their presence in the extreme thermophilic organisms means that these organisms are unable to function until the temperature is at least 70 or even 80 °C. Thus, the high stability of the membranes and cell walls of these organisms means that the organisms cannot grow below temperatures at which conventional microorganisms would quickly die. The presence of pentacyclic rings in the biphytanyl chains (Fig. 2d) ensures that the fluidity of the membranes containing these lipids remains fairly constant as the temperature is increased above 80 °C. Many of these organisms have been isolated from deep-sea thermal vents,

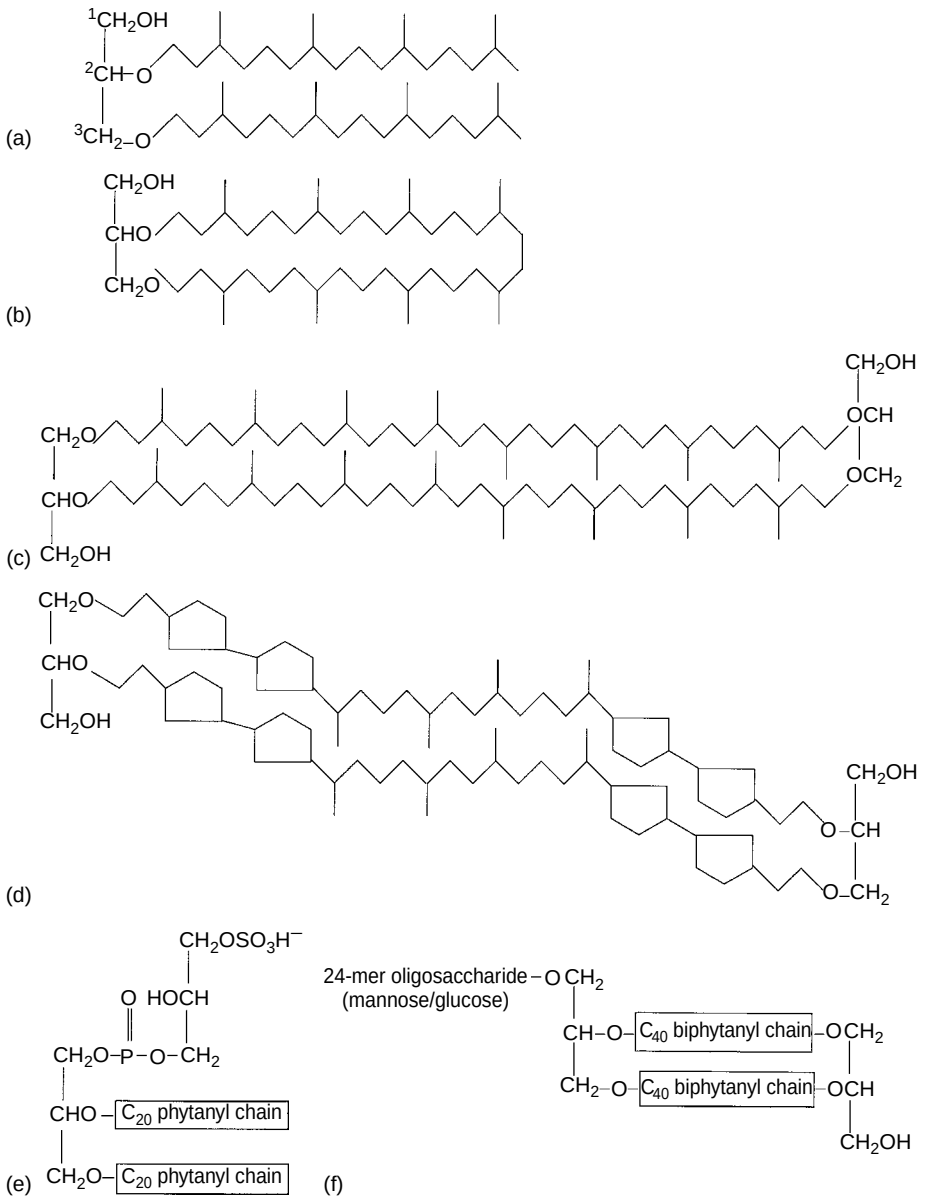

Fig. 2 Isoprenoid lipids of *Archaea:* (a) 2,3-di-*O*-phytanyl*sn*glycerol (archaeol), (b) 2,3-di-*O*-biphytanyl-*sn*-glycerol, (c) glycerol-dialkylglycerol tetraether (caldarchaeol), (d) tetracyclized glycerol dialkylnonitol tetraether (cyclized nonitolcaldarchaeol), (e) phosphoglycerolsulfate dialkylglycerol diether, (f) lipoglycan from *Thermoplasma* mannosyl-glucosyl oligosaccharide of (c).

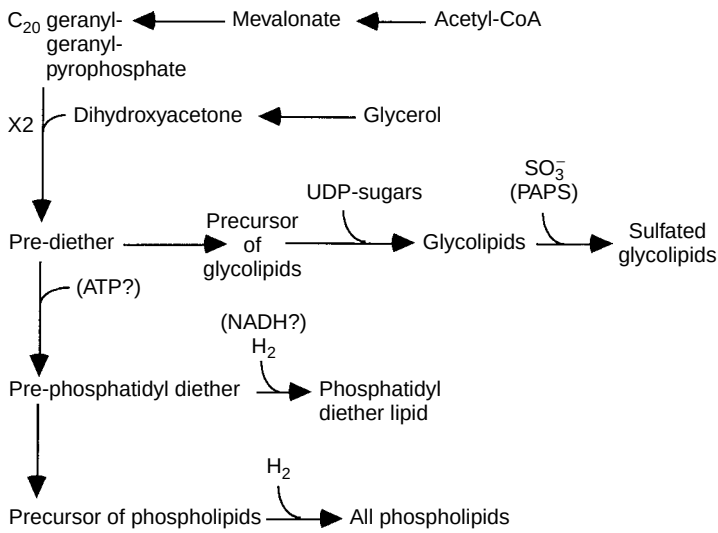

Fig. 3 Proposed pathways for the biosynthesis of the archaeol-containing phospho- and glycolipids in the extreme halophilic *Archaea*. (Adapted from the work of Morris Kates, University of Ottawa, Canada.)

demonstrating that their tolerance to temperatures in excess of 110 °C must be excellent.

Since the taxonomy of *Archaea* is still very much in its infancy, there is a considerable amount of work to be carried out on the structures of many of these organisms. Undoubtedly, some very complex lipids and lipid-containing molecules will be found in the next few years, and these will add to the bewildering array of the di- and tetraether lipids that have already been described.

3 Bacteria

"*Bacteria*" is the term now applied to all bacteria except those now within the *Archaea*. It, therefore, comprises the bulk of "conventional" bacteria that were originally classified as the *Eubacteria*. These organisms synthesize ester-based lipids, usually based on glycerol, from fatty acids that are synthesized via the acetyl-malonyl pathway (see Fig. 1). In most cases, the resultant lipids are used as structural and functional entities in bacteria, being the component of the main phospholipid membranes of the cell. Also, however, they are used in conjunction with polysaccharides and peptides in the construction of the cell wall and other parts of the cell envelope. Bacteria are divided into two major groups, the gram-positive and the gram-negative organisms, reflecting differences in their cell wall structures. The former have a peptide-glycan outer shell acting as the wall, which is then anchored to the inner phospholipid bilayer membrane via lipoteichoic acids (**VII**). In gram-negative bacteria, there is an outer membrane, beyond the peptidoglycan layer, that is a matrix of phospholipids, complex lipopolysaccharides in which the porin uptake proteins are anchored. Lipoproteins are then used to link the outer membrane

to the peptidoglycan layer that abuts the inner membrane. Most of the phospholipids are conventional diacylphosphoglycero derivatives (**II**) that are found in all living cells. There are, though, some variations in that ethanolamine, monomethyl-, and dimethylethanolamine may be used instead of choline (= trimethylethanolamine) as the polar basic group attached to the phospho group. Bacteria, apart from the mycobacteria, do not synthesize significant amounts of triacylglycerols (**I**).

The genetics of bacterial lipid metabolism are under active investigation by several major groups worldwide; the group led by John E. Cronan, Jr. (Urbana, IL) is particularly active in this respect.

3.1 Unsaturated Fatty Acids

In many bacteria, unsaturated fatty acids are synthesized by a so-called anaerobic pathway that is not found in other microorganisms. This pathway (Fig. 4) has been mainly studied in *E. coli*. It involves a modification of the existing fatty acid synthetase system (see Fig. 1) where, at the C_{10} level, there occurs an alternative dehydratase enzyme that forms 3*(Z)*-decenoate instead of the 2*(E)*-decenoate. (These terms replace the older descriptors of cis-β, γ-and *trans*-$\alpha\beta$-decenoates, respectively.) The latter intermediate is reduced and leads, via the continuation of fatty acid synthesis, to palmitoyl- and stearoyl-ACP (where ACP is a separable acyl carrier protein in *E. coli* having relative molecular mass of 8847). The 10 : 1 (*Z*3) intermediate cannot be reduced and is thus successively elongated so that the double bond remains in the molecule and, as it is elongated, the position of the double bond advances down the chain: 10 : 1 (*Z*3) → 12 : 1 (*Z*5) → 14 : 1 (*Z*7) → 16 : 1 (*Z*9) → 18 : 1 (*Z*11), which is *cis* (or *Z*)-vaccenic acid. This latter product is thus distinct from oleic acid, 18 : 1 (*Z*9), found in some bacteria and in all yeasts and molds. The presence of *cis*-vaccenic acid in extracted microbial lipids can be clearly distinguished from oleic acid by good gas chromatographic procedures, especially those using capillary gas chromatography.

The final elongation reaction of the "anaerobic" route is temperature sensitive. The enzyme that condenses 16 : 1 (*Z*9)-ACP (palmitoleoyl-ACP) with malonyl-ACP has a higher rate of reaction at lower temperatures than at

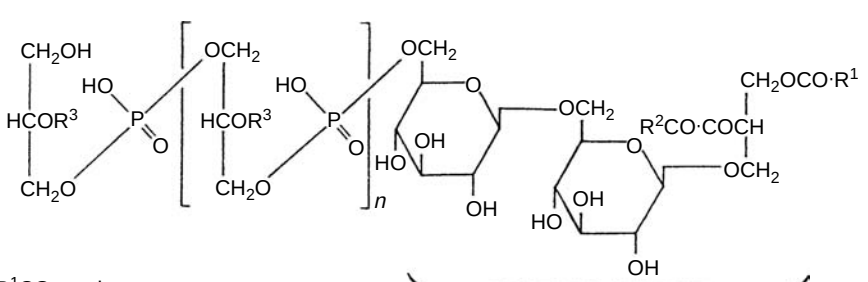

$R^1CO-$and
$R^2CO-$are conventional fatty acyl groups
R^3 is mainly H, D-alanyl or *N*-acetylglucosaminyl
$n = 30$ approx.

VII

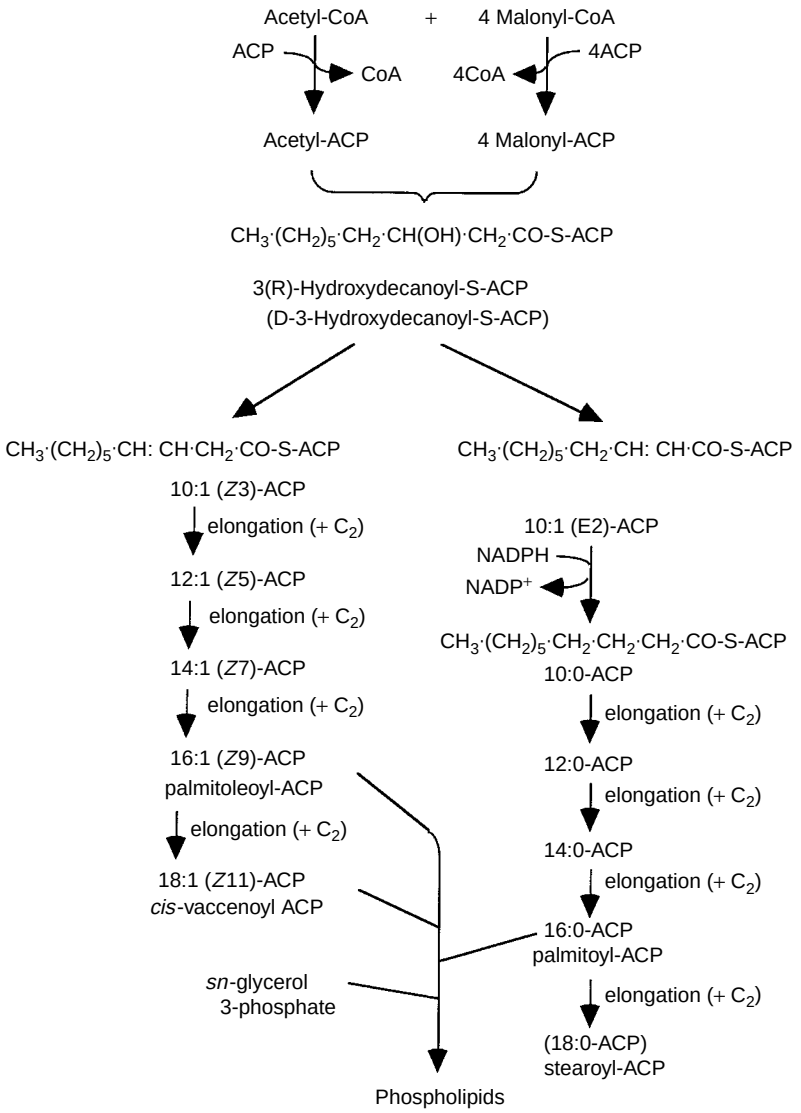

Fig. 4 Biosynthesis of saturated and unsaturated fatty acids by the fatty acid synthetase of *E. coli*. The conventional cyclic pathway of fatty acid biosynthesis condenses acetyl-ACP and malonyl-ACP in a succession of reactions leading to the formation of 3(*R*)-hydroxydecanoyl-*S*-ACP. This can either continue through the cyclic sequence to give the series of saturated fatty acids (right-hand side) or be dehydrated to give a fatty acid with the double bond now in a different position but also of a different isomeric configuration. Since this intermediate, the 3(*Z*)-decenoyl-ACP, now cannot participate in the reductive reaction, the double bond remains as the acyl chain continues to undergo further condensation reactions in the fatty acid synthetase cycle. The final conversion of 16 : 1 (*Z*9)-ACP to 18 : 1 (*Z*11)-ACP is a temperature-controlled reaction in that at low temperatures the condensing enzyme has a higher relative rate of reaction than at higher temperatures (see text).

higher ones. Consequently, at low growth temperatures, more 18:1-ACP is formed than at higher temperatures, and it competes with palmitoyl-ACP for incorporation into phospholipids (see Fig. 4) and results in the formation of a higher proportion of diunsaturated acylphospholipids. This, in turn, ensures a lower temperature for the liquid phase transition of the membrane lipids. This temperature sensitivity of unsaturated fatty acid biosynthesis allows *E. coli* and other bacteria to adapt and to survive at low temperatures by modifying the fatty acyl constituents of the membrane phospholipids, ensuring that they remain as fluid as possible at lower temperatures.

3.2
Polyunsaturated Fatty Acids

Although the vast majority of bacteria do not synthesize polyunsaturated fatty acids because of the lack of specific desaturase enzymes, mycobacteria do synthesize fatty acids with multiple double bonds: for example, phleic acid is a 36 : 5 (4, 8, 12, 16, 20) fatty acid found in *Mycobacterium phlei* and *Mycobacterium smegmatis*. However, the double bonds are not of the methylene-interrupted type as found elsewhere. Such fatty acids, though, have been recognized in a number of bacteria recovered from fish intestines and include eicosapentaenoic acid, 20 : 5 (n-3), and docosahexaenoic acid, 22 : 6 (n-3). Although some interest has been indicated in these bacteria as potential sources of these very long-chain polyunsaturated fatty acids, better sources are various fungi and marine algae (see Sects. 5.3 and 6.2). The bacteria, however, could be useful providers of the genes needed to clone into plants if genetically modified plants are to be created that can synthesize these nutritionally desirable fatty acids.

3.3
Branched Chain Fatty Acids

Branched chain fatty acids with a methyl group at the penultimate or antepenultimate carbon atom (see **VIIIA** and **B**) are synthesized by both gram-positive and gram-negative bacteria. The principal genera among the former group that produce branched chain fatty acids are *Bacillus, Clostridium, Listeria, Micrococcus, Sarcina*, and *Staphylococcus*; of the gram-negatives, *Flavobacterium, Bacteroides, Legionella, Vibrio*, and *Desulphovibrio* are the principal genera, though some species of *Pseudomonas* and *Xanthomonas* also contain these acids. The branched chain fatty acids are synthesized from the derivatives arising from the branched chain amino acids: leucine (or valine) for the iso-fatty acids and isoleucine for the anteiso-fatty acids (see **VIIIA** and **B**).

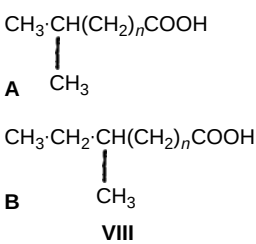

Fatty acids with a methyl branch in the middle of the chain are also known. The best-known example of this is tuberculostearic acid, 10-methyloctadecanoate (**IX**), which was originally isolated from the tubercle bacillus *Mycobacterium tuberculosis* but has now been found in more than 20 difference species and strains of *Mycobacterium* besides also occurring in a number of other actinomycetes.

A number of other methyl-branched fatty acids also occur in mycobacteria, including small amounts of the iso- and anteiso-fatty acids, as well as multi-methyl-branched fatty acids of which the C_{32} mycocerosic acid 2,4,6,8-tetramethyloctadecanoate is but one of several complex branched chain fatty acids found in this genera (see also Sect. 3.6, Mycolic Acids).

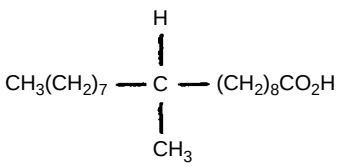

IX

3.4 Cyclopropane Fatty Acids

Fatty acids that contain a cyclopropane ring are fairly common in bacterial lipids. The most widely distributed example is lactobacillic acid, Z-11, 12-methyleneoctadecanoic acid (**X**), abbreviated as 19:0 cy. This, as the name would suggest, was originally isolated from *Lactobacillus* spp. but is recognized now in many organisms including *E. coli*, most *Pseudomonas* spp., and some *Rhizobium* and *Clostridium* spp. The corresponding 17 : 0 cy fatty acid also occurs. The presence of these acids can be regarded as indicative of a unique bacterial lipid because they do not occur in eukaryotic microorganisms: a related fatty acid, sterculic acid (an unsaturated 19 : 1 cy fatty acid), though, does occur in some plant lipids.

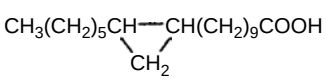

X

3.5 Hydroxy Fatty Acids

Hydroxy fatty acids are constituents of lipopolysaccharides of the cell envelope of the gram-negative bacteria and have been found in most species that have been studied. 3-Hydroxyalkanoic acids, ranging from C_{11} to C_{21}, have been identified in *E. coli* and species of *Salmonella, Klebsiella, Yersinia, Vibrio, Pseudomonas, Xanthomonas,* and *Rhizobium.* 2-Hydroxy fatty acids also occur to a lesser extent; however, the steric configuration of this hydroxy fatty acid is the *S* (or *L*) form and is thus opposite to the *R* (or *D*) form of the 3-hydroxy fatty acid. This probably indicates separate biosynthetic origins for the two fatty acids. 3-Hydroxy iso-fatty acids may be detected in some species as minor constituents.

The arrangement of the 3-hydroxy fatty acids (of which 3-hydroxy tetradecanoate is the most abundant type) is shown in Fig. 5 as part of the lipid A structure of some enterobacteria. Lipid A is the anchor region of the complex lipopolysaccharide structure in the outer membrane of these organisms. The fatty acids attached to the diglucosamine will then adjoin the periplasmic space.

3.6 Mycolic Acids

Very long chain, hydroxylated, branched chain fatty acids occur in bacteria belonging to the *Mycobacterium, Nocardia, Corynebacterium* group, *Rhodococcus,* and related groups of bacteria. The basic structure is that of a 2-alkyl, 3-hydroxy fatty acid (**XI**). The alkyl side chain may be up to 24 carbon atoms long (as occurs in *M. tuberculosis*); it is shorter in other mycobacteria. In *Corynebacterium* spp. it is only six carbons long. The long-chain group attached

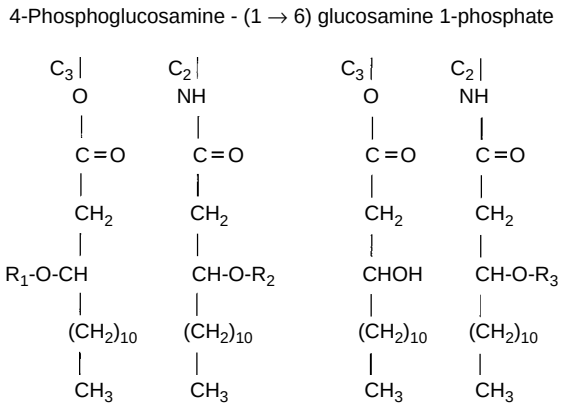

Lipid A	R_1	R_2	R_3
Salmonella minnesota	14:0 and 2-HO-14:0	12:0	16:0
Escherichia coli	14:0	12:0	H
Proteus mirabilis	14:0	14:0	H and 16:0

Fig. 5 Structure of lipid A in three enterobacteria.

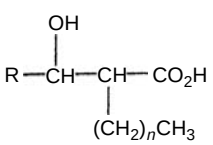

n = 5–23, odd numbers only

XI

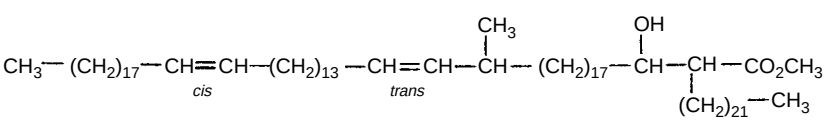

XII

at C-3 of the mycolic acid is up to 61 carbon atoms, making some of the mycolic acids the largest unsubstituted lipids that occur in any organism. A typical mycolic acid, as found in *M. smegmatis* as its methyl ester ($C_{80}H_{156}O_3$), is shown in structure **XII**.

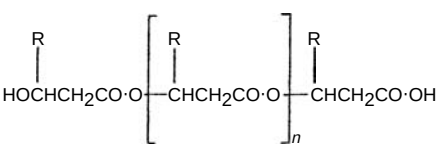

(*n* = up to 30 000)

XIII

3.7 Polyesters

Bacteria, apart from some exceptions in the *Mycobacterium* genus, do not synthesize triacylglycerols, which are the usual storage lipid of eukaryotic microorganisms (see Sects. 4.1 and 5.3). Lipid storage, nevertheless, still occurs

in bacteria by the biosynthesis of high molecular weight polyesters of β-hydroxybutyrate (PHB) or polyesters of β-hydroxyalkanoate (PHA). Poly-β-hydroxybutyrate ($R = -CH_3$ in **XIII**), which is the *D* (or *R*) isomer, is synthesized by a large number of bacteria; species of *Alcaligenes, Azotobacter, Bacillus, Nocardia, Pseudomonas, Rhizobium, Rhodococcus,* and *Zoogloea*, plus *E. coli*, have been investigated in numerous laboratories. In some bacteria, the more ubiquitous PHB is replaced by a poly-β-hydroxyalkanoate (PHA) in which the side chain (R in **XIII**) may be up to 12 carbon atoms long. PHB has also been found in some fungi.

PHA, which can reach up to 80% of the biomass of *Alcaligenes eutrophus*, and also in a recombinant strain of *E. coli*, can attain molecular sizes of over 1 million Daltons and, moreover, be produced by the bacteria with a combination of side chains (see **XIII**) which then endows the resultant biopolymer with properties that can be designed with a particular function in mind. Nevertheless, in spite of all these advantages, including biodegradability, neither PHB nor PHA has achieved commercial production mainly because the costs involved are about ten times those of producing polyethylene and related chemical polymers.

Some hope for commercial production was considered to be by way of introducing the requisite three key genes for PHB synthesis (see Fig. 6) into a plant such as rapeseed, sunflower, or soybean. Unfortunately, yields of the biopolymer were not as good as in the bacteria and costs of production were again found to be higher than costs of producing conventional plastics. A limited amount of commercial interest only remains in these unique bacterial polymers once considered as the "green" alternative to petrochemically produced plastics.

Since the pathway for PHB biosynthesis requires only three new enzymes (see

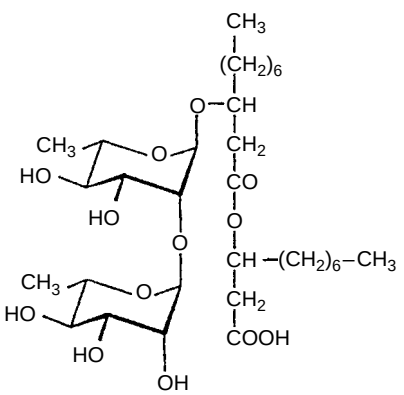

XIV

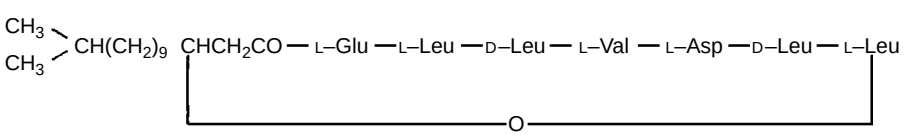

XV

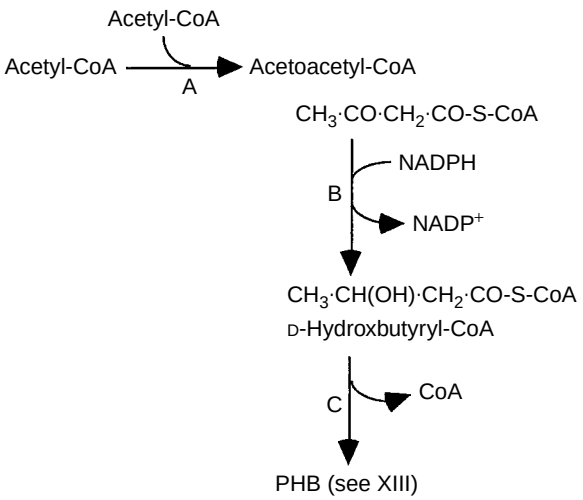

Fig. 6 Biosynthesis of poly-β-hydroxybutyrate in bacteria. Enzymes are 3-ketothiolase (A), acetoacetyl-CoA reductase (B), and PHB synthetase (C).

Fig. 6), prospects of cloning the corresponding three genes into plants has been under consideration. All three genes have been isolated and sequenced. Plants that normally synthesize considerable amounts of oil, thereby already having good supplies of acetyl-CoA, are the obvious recipients of the gene transfer: rapeseed, sunflower, and soybean are all currently receiving attention. However, the cost of producing these materials, even using genetically modified plants, appears to be more than the market is willing to pay for biodegradable plastics.

3.8 Biosurfactant Lipids

Bacteria produce a number of biosurfactants of diverse chemical structures that have the ability to form stable emulsions of oil and water mixtures. Most of the biosurfactants owe their properties to being glycolipids (see **XIV**); that is, they possess both water-soluble and oil-soluble components. Some, however, are lipopeptides (see **XV**). Highest concentrations of these compounds are produced during bacterial growth on oils, either petroleum or plant in origin. Several of these biosurfactants are produced commercially or are undergoing industrial investigation. They include emulsan, which is produced by *Acinecobacter calcoaceticus* RAG-1 and is a polyanionoic, amphipathic, heterolipopolysaccharide of high molecular weight (1 MDa), rhamnolipids (**XIV**) produced by *Pseudomonas* spp., and surfactin (**XV**) produced by *Bacillus subilis*. Other biosurfactants of commercial interest are produced by yeasts and molds (see Sects. 4.4 and 5.4).

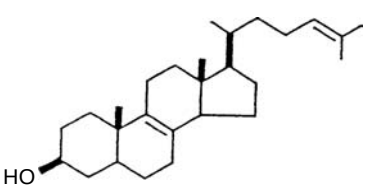

XVI

4 Yeasts

4.1 Fatty Acids and Fatty Acyl Lipids

Yeasts are eukaryotic, unicellular fungi that do not form a mycelium and, thus, are differentiated from the filamentous fungi. There are over 600 identified species of yeasts. Yeasts, like all eukaryotic organisms, synthesize their fatty acids via the acetate–malonate route (Fig. 1) and possess specific desaturases for the conversion of saturated fatty acids to unsaturated ones. Most fatty acids are 16 to 18 carbon atoms in length, though small amounts of fatty acids as long as 28 or 30 carbons can be discerned by careful analysis of some species.

Yeasts tend to have a much more limited range of fatty acids than molds: oleic (18 : 1), palmitic (16 : 0), linoleic (18 : 2), and palmitoleic (16 : 1) acids usually account for more than 90% of the total fatty acids. Typical fatty acid profiles of yeasts are shown in Table 1.

Tab. 1 Typical fatty acid profiles of some selected yeasts.

Organism	*Relative amount [% w/w] of major acyl groups*						
	14 : 0	*16 : 0*	*16 : 1*	*18 : 0*	*18 : 1*	*18 : 2*	*18 : 3*
Ascosporogenous yeasts							
Debaryomyces hansenii	1	20	12	2	28	30	8
Hansenula polymorpha	1	16	2	3	25	30	23
Kluyveromyces marxianus	<1	11	24	5	45	12	3
Lipomyces lipofer[a]	<1	37	4	7	48	3	0
Lipomyces starkeyi[a]	<1	34	6	5	51	3	0
Pichia stipitis	<1	12	5	2	33	40	7
Saccharomyces cerevisiae	1	9	46	2	42	0	0
Schizosaccharomyces pombe	1	4	23	<1	70	0	0
Schwanniomyces occidentalis	<1	16	11	1	42	21	8
Yarrowia lipolytica[a]	3	21	7	7	17	35	0
Zygosaccharomyces rouxii	<1	6	11	3	33	46	0
Basidiosporogenous yeasts							
Leucosporidium scotti	<1	5	2	3	17	34	28
Rhodosporidium toruloides[a]	1	25	<1	13	46	12	2
Sporidiobolus salmonicolor	<1	24	<1	6	36	36	3
Asporogenous yeasts							
Brettanomyces lambicus	2	16	48	0	21	11	0
Candida albicans	<1	20	7	1	22	27	20
Candida boidinii	<1	15	20	3	27	31	0
Candida curvata D[a,b]	<1	29	<1	12	51	6	2
Candida tropicalis	<1	22	5	9	29	26	4
Cryptococcus albidus[a]	<1	12	1	3	73	12	<1
Rhodotorula graminis[a]	1	30	1	12	36	15	4
Trichosporon cutaneum[a]	3	13	0	22	50	13	0
Trichosporon pullulans[a]	<1	15	0	2	57	24	1

[a]Considered to be oleaginous species with oil contents >25% of cell dry weight.
[b]Now reclassified as *Cryptococcus curvatus*.

Fatty acid profiles of all the major genera and of the key species within each genus now have been published. The most systematic and extensive survey was conducted by Lodewyk Kock and his associates in Bloemfontein, South Africa. Although most yeasts (see Table 1) synthesize di- and polyunsaturated fatty acids (18 : 2 and 18 : 3), *Saccharomyces cerevisiae* (baker's and brewer's yeast) does not synthesize unsaturated fatty acids beyond oleic acid (18 : 1). Reports of the occurrence of 18 : 2 and 18 : 3 in *S. cerevisiae* continue to be made, though their presence is attributable to the yeast being grown on a culture medium that contains animal- or plant-derived material bearing traces of these acids. Other yeasts, though, synthesize both 18 : 2 and 18 : 3, which is the α-isomer [i.e. 18 : 3 (9, 12, 15)].

Fatty acids are incorporated into a range of acylglycerophospholipids (**II**) of which phosphatidylcholine, -serine, -ethanolamine, and -inositol are the major ones; diphosphatidylglycerol (cardiolipin) is also common. Many yeasts also synthesize small amounts of sphingolipids (whose function is largely unknown). Unlike bacteria, all yeasts synthesize some triacylglycerols (**I**) along with the functional phospholipids. Triacylglycerols function as a reserve storage product; in some yeasts (about 25 or so species), the amount of triacylglycerol that is stored can be up to 80% of the cell volume. Such yeasts are known as oleaginous species and have been examined as potential sources of oils because the extracted triacylglycerol is similar in composition to several of the commercially produced plant oils.

4.2 Single Cell Oils (SCO)

Microbial lipids that have been considered as sources of edible oils are referred to as *single cell oils* (SCO), as an equivalent term to *single cell proteins*, which describes microorganisms as potential sources of protein or food. In view of the high costs of biotechnological production of such oils, commercial interests have centered on the highest valued oils. For yeasts, this focus has been in the production of a substitute for cocoa butter. (With molds, commercial interest has been directed mainly on the polyunsaturated fatty acids: see Sect. 5.) The main yeast to have been examined was *Candida curvata* (also known as *Apiotrichum curvatum*), now renamed *Cryptococcus curvatus*. To replicate the fatty acid composition of cocoa butter, it was necessary to delete the Δ9-desaturase gene converting stearic acid to oleic acid, thus causing the accumulation of stearic acid up to 40% in the yeast oil (see Table 2). The high content (35%) of stearic acid in cocoa butter gives the product its characteristic hardness at ambient temperatures, with a sharp transition to the molten state at 30 to 32 °C.

The genetic modification of *Candida curvata* was carried out by Henk Smit and colleagues at the Free University of Amsterdam. The yeast has also been explored commercially in New Zealand by Julian Davies (Industrial Research Ltd., formerly the Department of Scientific and Industrial Research of the New Zealand government) using as a low cost substrate the waste lactose from whey arising during cheese manufacture. Even with such a cheap feedstock, however, the biotechnological route has not proved to be commercially competitive against cocoa butter or equivalent fats that are produced from palm oil by fractional recrystallization. Interest in this particular SCO has therefore ceased.

Tab. 2 Cocoa butter fatty acids and the fatty acyl composition of yeast triacylglycerols[a].

Yeast	*Relative amounts [% w/w] of major fatty acyl groups*					
	16:0	*18:0*	*18:1*	*18:2*	*18:3*	*24:0*
WT	17	12	55	8	2	1
WT-NZ[b]	23	23	42	3	1	4
Ufa 33[c]	20	50	6	11	4	4
R22.72	16	43	27	7	1	2
F33.10	24	31	30	6		4
Cocoa butter	23–30	32–37	30–37	2–4		Trace

[a]From *Cryptococcus curvata* [formerly *Candida curvata (Apiotrichum curvatum)*] wild-type (WT) strain, an unsaturated fatty acid auxotrophic mutant (Ufa 33), a revertant mutant (R22.72), and a hybrid derived from Ufa 33 (F33.10) compared to the best results obtained with the wild-type strain grown with limited O_2 supply to diminish desaturase activity.
[b]Wild type grown (in New Zealand) on whey lactose.
[c]Grown with 0.2 g oleic acid L^{-1}.

4.3 Biochemistry of Oleaginicity

The biochemical reason for the relatively sparse distribution of oleaginicity among yeasts (of 600 plus species, only some 25 or so yeasts are known that accumulate >25% lipid) has been identified as the key presence of ATP: citrate lyase (ACL) in the oleaginous species (see Fig. 7). In this scheme, citric acid, which is produced within the mitochondrion, is not further metabolized through the reactions of the tricarboxylic acid cycle because, under the conditions that lead to the accumulation of lipid (exhaustion of N from the culture medium), isocitrate cannot be converted to α-ketoglutarate. This conversion is prevented because adenosine monophosphate (AMP) reaches such a low intracellular concentration under these conditions that isocitrate dehydrogenase, which specifically requires AMP for activity, cannot become fully operational. Thus, when cells have exhausted their supply of N in the medium, the concentration of AMP falls, preventing citrate metabolism; then, provided a supply of carbon (e.g. glucose) is still available, the cells begin to accumulate citrate, which exits the mitochondrion and is cleaved by ACL to yield acetyl–CoA units. The other product from the ACL reaction is oxaloacetate, which is recycled to pyruvate via malic acid; which is decarboxylated to pyruvate by malic enzyme and simultaneously produces the NADPH that is needed for fatty acid biosynthesis. Although all oleaginous yeasts possess ACL, not all possess malic enzyme; at least some species, therefore, must have available an alternative means of generating NADPH and recycling the oxaloacetate. The underlying biochemistry of oleaginicity in yeasts and molds has been largely worked out in the author's laboratory.

The genetics of fatty acid and phospholipid biosynthesis has been examined principally in *S. cerevisiae*, which is not

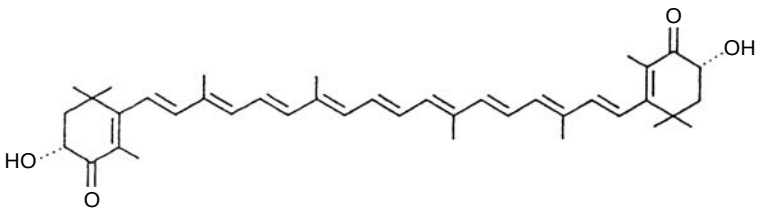

XVII

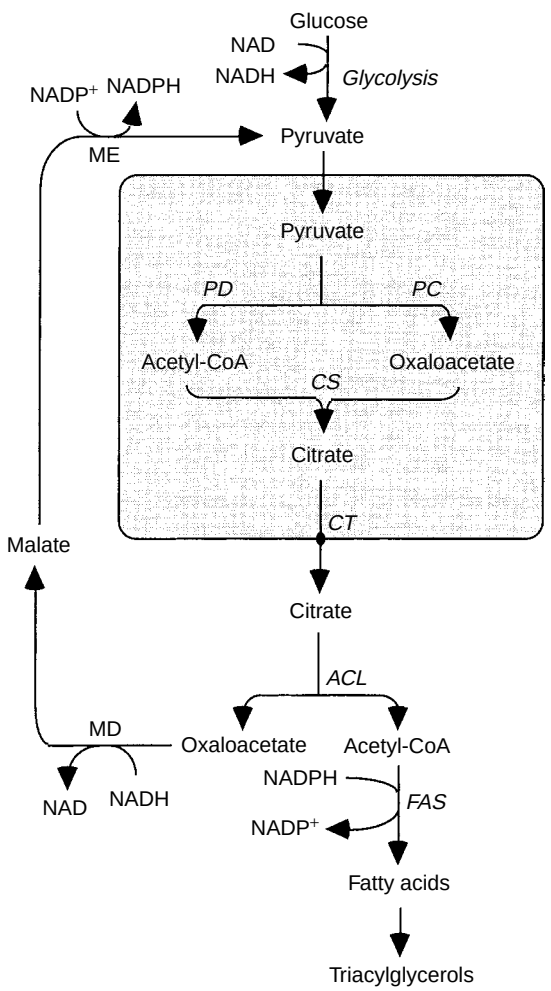

Fig. 7 Mechanism of acetyl–CoA production in oleaginous yeasts and molds leading to the accumulation of single cell oils (triacylglycerols). The metabolism of glucose to pyruvate occurs by *glycolysis* in the cytoplasm; pyruvate is transported into the mitochondrion (shaded box) and is converted to acetyl–CoA and oxaloacetate (by pyruvate dehydrogenase, *PD*, and pyruvate carboxylase, *PC*, respectively), which are synthesized into citrate via citrate synthetase (*CS*). Citrate is translocated via citrate translocase (*CT*) out of the mitochondrion, to be cleaved by ATP: citrate lyase (*ACL*). The acetyl–CoA then is used for fatty acid synthesis (*FAS* = fatty acid synthetase) and oxaloacetate is converted by malate dehydrogenase (*MD*) and malic enzyme (*ME*) back to pyruvate. Malate dehydrogenase uses NADH, which will be provided by the key reactions of glycolysis: *ME* generates the NADPH needed for fatty acid biosynthesis.

an oleaginous species and does not contain ACL activity. Leading investigators of these processes are Susan A. Henry at Cornell University and George M. Carman at Rutgers University, in New Jersey.

4.4 Other Lipids

Yeasts synthesize varying amounts of terpenoid lipids. Sterols, mainly ergosterol (**III**) and zymosterol (**XVI**), occur in all yeasts but usually at no more than 1% of the cell dry weight. However, in some species of *S. cerevisiae*, ergosterol has been reported up to 10%, and, in one case, up to 20% of the cell dry weight. However, this sterol is not of commercial importance.

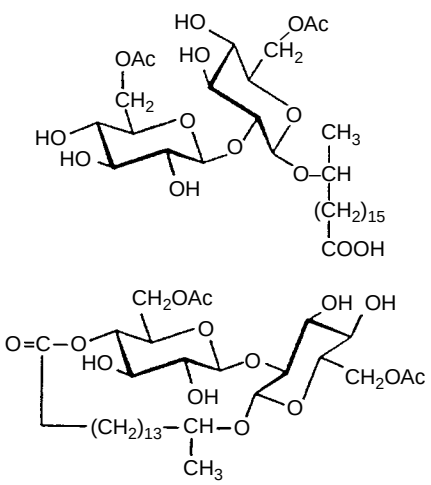

XVIII

Many yeasts do not synthesize carotenoids lipids. These include, obviously, the *Candida* (from *candidus*, Latin for white) yeasts. Various carotenoids, though, are produced in some abundance by the pink and red yeasts: principally, *Sporobolomyces*, *Rhodotorula/Rhodosporidium*, and *Phaffia*. The carotenoid produced by *Phaffia rhodozyma* is astaxanthin (**XVII**), which is now produced commercially as a supplement to fish feed, being a source of red pigmentation for the coloration of salmonid fish.

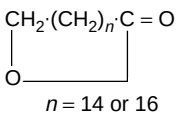

$n = 14$ or 16

XIX

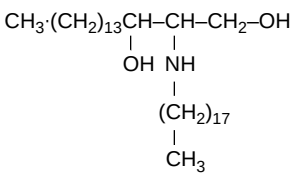

XX

Yeasts, like bacteria, also produce a number of biosurfactants when grown in hydrocarbons or other water-insoluble substrates. The most widely known of these are the sophorolipids (**XVIII**) produced by *Candida bombicola* and related species. Yields may exceed 100 g L^{-1}, but commercial uptake has been very limited because the sophorolipids, though efficacious as emulsificants, have little advantage over the much cheaper, chemically produced surfactants. The key component fatty acyl group of the sophorolipid from *C. bombicola* is a ω- or ω-1-hydroxy fatty acid, which may be lactonized into a macrocyclic lactone (**XIX**) that has some potential use as perfumery intermediate.

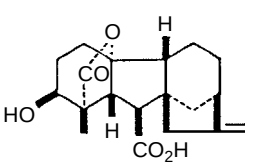

XXI

Although ceramide lipids are usually only minor components of microbial lipids, *N*-stearolyphytosphingosine (**XX**) is now being produced commercially using *Pichia* (formerly *Hansenula*) *ciferri*. The ceramide, after purification, is incorporated into various preparations being used in the cosmetics industry, and is also of some potential in the health care industry for treatment of various skin disorders.

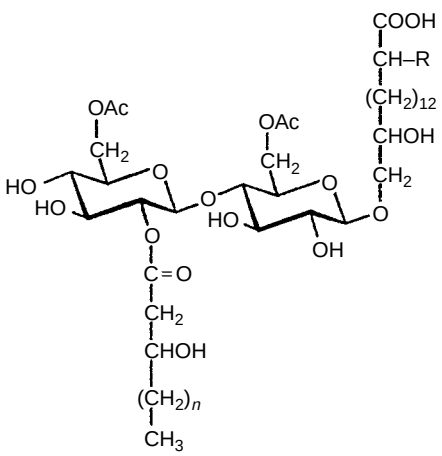

XXII

5 Molds

5.1 Fatty Acids

Some 60 000 individual species of filamentous fungi are known. These microorganisms, which are also referred to as *molds*, synthesize a range of fatty acids, from 12 to 24 carbon atoms in chain length and include polyunsaturated fatty acids up to 22 : 6 (docosahexaenoic acid). Although no systematic survey of mold lipids has been carried out, analysis of most of the easily cultivatable genera has been accomplished. Table 3 gives some of the different ranges of fatty acids that are found. Very little work has been carried out in the genetics of molecular biology of fungal lipids. In a number of species, there may be extensive accumulation of triacylglycerols as storage reserves; oil contents exceeding 70% have been recorded in several instances. The biochemistry of lipid accumulation and of biosynthesis appears to be similar to those processes elucidated with yeasts.

Besides the conventional straight chain fatty acids, some species of the order of phycomycetes produce iso- and anteiso-fatty acids (**VIIIA, B**). Ricinoleic acid, 15-hydroxyoleic acid, which is the principal fatty acid found in castor oil, occurs in *Claviceps* species.

5.2 Polyunsaturated Fatty Acids

One of the few differences in fatty acid composition between representatives of the three main orders of molds – the phycomycetes (or lower fungi), the ascomycetes (also known as the fungi imperfecti), and the higher fungi or basidiomycetes – is that the phycomycetes synthesize unsaturated fatty acids of the n-6 series. The ascomycetes and basidiomycetes fungi, in contrast, synthesize n-3 fatty acids. The biosynthetic route to these two series of fatty acids is shown in Fig. 8. These two pathways are similar to those that occur elsewhere: the n-6 route is an "animal"-only route, and the n-3 route is found in plants. Uniquely, some fungi are able to convert arachidonic acid as a member of the n-6 series to eicosapentaenoic acid, which belongs to the n-3 series, by the action of a Δ17-desaturase that is not found in other organisms. The ability of fungi to synthesize fatty acids longer than C_{18} is very restricted, however,

Tab. 3 Typical fatty acid profiles of some selected filamentous fungi.

Organism	***Relative amount [% w/w] of major acyl groups***							
	12 : 0 and shorter	***14 : 0***	***16 : 0***	***18 : 0***	***18 : 1***	***18 : 2***	***18 : 3***	***Others***
Phycomycetes (lower fungi)								
Conidiobolus nanodes	<1	<1	15	7	26	3	3[a]	20 : 1, 16% 20 : 4, 22%
Entomophthora coronata	40	31	9	2	14	2	1[a]	
Mortierella alpina	0	0	14	7	10	6	5[a]	20 : 3, 6% 20 : 4, 52%
Mucor hiemalis	0	2	15	10	33	19	19[a]	
Mucor miehei	0	1	23	6	44	15	6[a]	
Pythium ultimum	0	10	16	3	18	17	2[a]	20 : 1, 3% 20 : 4, 14% 20 : 5, 10%
Rhizopus arrhizus	0	19	18	6	22	10	12[a]	
Thraustochytrium aureum	0	2	27	1	35	3	1[a]	20 : 3, 2% 20 : 5, 6% 22 : 5, 6% 22 : 6, 11%
Ascomycetes (fungi imperfecti)								
Aspergillus niger	2	1	14	8	28	36	9	
Aspergillus terreus	0	2	23	<1	14	40	21	
Claviceps purpurea	0	<1	23	2	19	8	0	12-OH-18 : 1, 42%
Fusarium moniliforme	0	1	15	11	30	42	1	
Fusarium oxysporum	0	<1	31	12	35	18	1	
Penicillium spinulosum	0	<1	15	7	42	31	1	
Basidiomycetes (higher fungi)								
Agaricus campestris	2	1	15	3	5	66	0	
Tolyposporium ehrenbergii	<1	1	7	5	81	2	0	
Ustilago zeae	0	<1	9	3	54	17	0	20 : 1, 2% 20 : 2, 8% 22 : 0, 1%

[a] γ-linolenic acid, 18 : 3 (6, 9, 12).

and tends to occur to any appreciable extent only in some representatives of the phycomycetes fungi.

5.3 Single Cell Oils

Commercial interest in fungal oils has centered on the species that produce an abundance of certain polyunsaturated fatty acids. Oils (triacylglycerols) containing such fatty acids are used as dietary supplements, and numerous claims are made for their efficiency in the treatment of disorders as varied as childhood eczema and premenstrual tension, and in the prevention of heart disease. The inclusion of particular polyunsaturated fatty acids into infant milk formulations has also been advocated because these acids, which occur

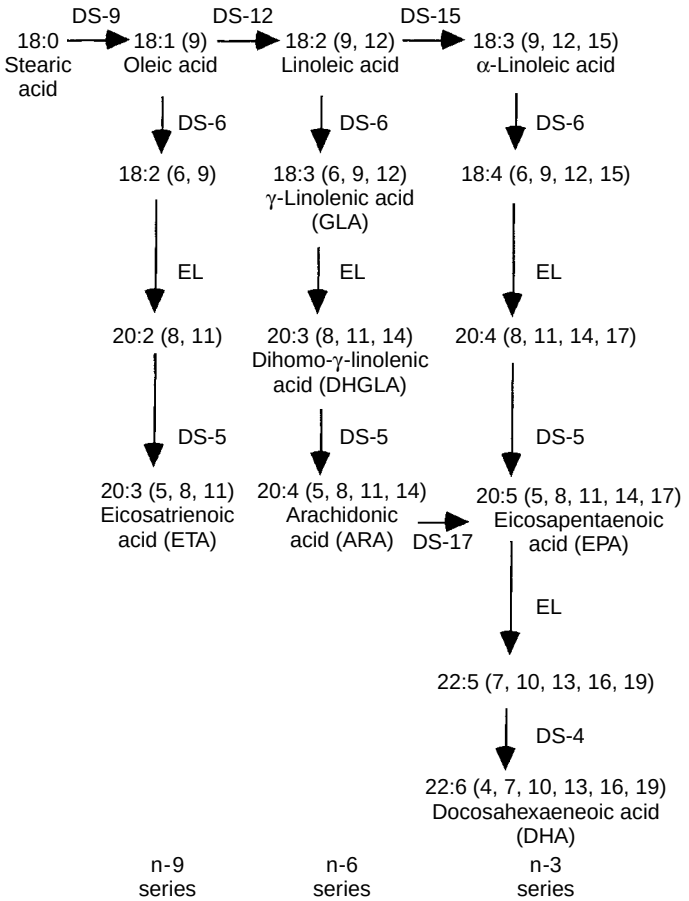

Fig. 8 Pathways of polysaturated fatty acid biosynthesis in fungi: DS-N, desaturase operating at *N*th carbon atom in chain; EL, elongase (acetyl–CoA dependent). The notations n-9, n-6, and n-3 indicate the position of the last double bond that is furthest away from the respective carboxylic acid group.

in human milk, are completely lacking in cow's milk.

Of particular interest are fungi that produce γ-linolenic acid (GLA), 18 : 3 (6, 9, 12), which is currently available only in the seed oil of the evening primrose (*Oenothera bienesis*), borage (*Borago officinalis*), and, as a mixture with the α-isomer, 18 : 3 (9, 12, 15), in *Ribes* spp. (blackcurrant, redcurrant, and so on). The high cost of these plant oils (between \$10 and \$15 per kg, though prices are very variable and have been both lower and higher in recent years) has made the biotechnological route using species of *Mucor, Mortierella*, and *Rhizobium* particularly attractive. Commercial processes using species of the former two genera were developed in the 1980s and ran into the 1990s, though, neither operates today (2004). In the United Kingdom, a process

was operated by J. & E. Sturge Ltd, Yorkshire, using *Mucor circinelloides* grown in fermenters of up to 220 m^3 with the oil being sold under the trade name of Oil of Javanicus. This became the world's first commercially produced single cell oil.

The oil had the highest possible purity and satisfied all regulatory guidelines for its quality. In Japan, a similar process was developed later by Idemitsu Kosan Co. Ltd, Tokyo, using *Mortierella isabellina*. Production of both oils ceased primarily because of the cheaper plant oils that became available, especially borage oil (also known as starflower oil) which had a superior content of GLA over the mold oils (see Table 4 for comparisons between the fungal oils and plant oils containing GLA).

The market for oils rich in GLA now appears to be met by borage and evening primrose oils and both have developed their own niche positions making it difficult for alternative sources of GLA to displace them. This includes the possible production of GLA using a genetically modified plant, such as sunflower, in which the gene coding for the D6-desaturase, that converts the existing linoleic acid (18 : 2) into GLA (see Fig. 8), has been inserted into the plant. The reluctance of the commercial owners of this technology to develop the GM crop for GLA oil production is perhaps understandable because of the current political pressures (especially in the United Kingdom and other European countries) against such technology and the advent of the cheaper borage oil.

One single cell oil process using a mold continues to be of increasing commercial importance and this is for the production of an oil rich in arachidonic acid (ARA), 20 : 4 (n-6) – see Fig. 8. This polyunsaturated fatty acid is now recommended for inclusion in infant formulae along with docosahexaenoic acid (DHA) 22 : 6 (n-3) whose production as another single cell oil is described in Sect. 6.2. The functional roles of ARA and DHA are considered to be in the development of retinal and neural membrane lipids and are thought to be of diminished availability if babies are not being fed upon mother's milk.

Tab. 4 Fatty acid profiles of two commercial fungal oil products compared with evening primrose oil, borage oil, and black currant oil containing γ-linolenic acid.

Oil content, [% w/w]	*Mucor circinelloides*[a] *20*	*Mortierella isabellina*[b] *ND*	*Evening primrose 18–20*	*Borage*[c] *30*	*Blackcurrant 30*
Fatty acid	*Relative amount [% w/w]*				
16 : 0	22–25	27	6–10	9–13	6
18 : 0	5–8	6	1.5–3.5	3–5	1
18 : 1	38–41	44	6–12	15–17	10
18 : 2	10–12	12	65–75	37–41	48
γ-18 : 3	15–18	8	8–12	19–25	17
α-18 : 3	0.2		0.2	0.5	13

[a] Production organism used by J. & E. Sturge Ltd., UK
[b] Production organism used by Idemitzu Ltd., Japan; ND, not disclosed but probably about 45–50%.
[c] Also contains 20 : 1 (4.5%), 22 : 1 (2.5%), and 24 : 0 (1.5%); 22 : 1 is erucic acid.

Supplementation of infant formulae by both these fatty acids has now been recommended in over 60 countries, including the United States and most European countries. Infants receiving these supplements are considered to have improved development of memory and of eyesight.

The organism of choice for ARA production is *Mortierella alpina*. The mold is grown in large-scale fermenters and produces an oil content of about 40% of its biomass with ARA at about 45% of the total fatty acids. Various descriptions of this process have been given. The final oil is diluted with some sunflower oil in order to bring the ARA content down to a standard level of 40%. The oil itself has passed every stringent test and is regarded as entirely safe for ingestion by infants as well as adults, including pregnant females. A profile of the fatty acids of this oil is shown in Table 5.

5.4 Other Lipids

Molds, like yeasts, contain a wide variety of lipids. Some of these lipids are associated with various aspects of fungal differentiation, and certain lipids are found only in association with fungal spores or with particular stages of fungal development. As an example of this, ricinoleic acid (15HO-18 : 1), is found in *Claviceps purpurea*, but only in association with the sclerotial stage of development of this fungus. No 15HO-18 : 1 is found when *C. purpurea* is grown in its mycelial form in submerged culture.

Sterols occur in all molds. Ergosterol (**III**) is the commonest one, but many others (>40) are known.

Carotenoids occur in about 60% of all fungi. β-Carotene (IV) is produced commercially using *Blakeslea trispora* and, though most processes have now ceased owing to the availability of cheaper, chemically produced material, at least one process still continues in eastern Europe.

The group of compounds known as the gibberellins are diterpenes and are biosynthetically derived from a common intermediate of the carotenoid pathway (geranyl geranylpyrophosphate). These are commercially produced as plant growth hormones using *Gibberella fujikuroi*, which

Tab. 5 Fatty acid profiles of single cell oils rich in arachidonic acid and docosahexaenoic acid that are produced commercially by cultivation in large (100 m^3) fermenters.

Oil	*Fatty acid composition (Rel.% w/w)*													
	12 : 0	*14 : 0*	*16 : 0*	*16 : 1*	*18 : 0*	*18 : 1*	*18 : 2 (n-6)*	*18 : 3 (n-6)*	*18 : 3 (n-3)*	*20 : 3 (n-6)*	*20 : 4 (n-6)*	*22 : 5 (n-6)*	*22 : 6 (n-3)*	*24 : 0*
ARASCO[a]	–	0.4	8	0	11	14	7	4	–	4	49	–	0	1
DHASCO[b]	4	20	18	2	0.4	15	0.6	–	–	–	–	–	39	–
DHAGold[c]	–	13	29	12	1	1	2	–	3	1	–	12	25	–

[a] Production organism: *Mortierella alpina*.
[b] Production organism: *Crypthecodinium cohnii*.
[c] Production organism: *Schizochytrium* sp. and now (2004) renamed as DHASCO-S.
Note: ARASCO, DHASCO, and DHAGold are registered trade names of Martek Bioscience Corp., MD, United States of America. ARASCO and DHASCO are combined together at a 2 : 1 ratio and used for incorporation into infant formulae under the trade name of Formulaid.

is the telomorphic state of *Fusarium moniliforme*. The principal product is gibberellic acid (**XXI**), though some 80 different variations of this compound are known.

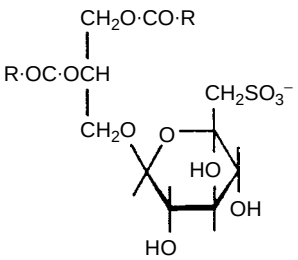

XXIII

A number of surfactants are also produced by fungi. Like those produced by other microorganisms, they are effective in creating stable oil–water mixtures. They may, therefore, be functionally useful to microorganisms that inhabit the phyllosphere and, as such, may improve attachment of the fungi (mycelial cells or spores) to plant leaves. A surfactant based on cellobiose from the maize rust fungus *Ustilago maydis* is known as ustilagic acid (**XXII**).

6 Unicellular Algae

The term "algae" encompasses 14 major groups or classes ranging from prokaryotic bacteria through the yellow, green, golden, and brown macroalgae, which includes both the seaweeds and diatoms. Most algae are photosynthetic organisms, though some may grow heterotrophically – that is, using fixed carbon compounds rather than CO_2, from which both carbon and energy may be obtained. When CO_2 is used as the carbon source, sunlight is then usually used, via photosynthesis, as the source of energy.

6.1 Cyanobacterial Lipids

The prokaryotic algae, or cyanobacteria, were originally referred to as the *blue-green algae*. The principal genera are *Anabena, Anacystis, Nostoc, Oscillatoria, Spirulina*, and *Synechococcus*. The lipids of these genera are diverse and, of course, include chlorophyll, which is an essential part of the photosynthetic apparatus. Whereas bacteriochlorophylls are found in the green and purple photosynthetic bacteria (*Rhodospirillaceae, Chromatiaceae*, and *Chlorobiaceae*), plant chlorophyll is found in the cyanobacteria. This then suggests that cyanobacteria are the ancestors of the chloroplasts of plants. The neutral lipids of cyanobacteria include many carotenoid pigments, hydrocarbons, hopanoids, quinones, and, in some species, poly-β-hydroxybutyrate (**XIII**). The principal polar lipids of the cyanobacteria are phosphatidylglycerol and sulfoquinovosyldiacylglycerol (**XXIII**), which are regarded as the sulfolipid characteristic of plants. The fatty acyl substituents of these lipids are "plant-like" rather than "bacteria-like" and include oleic acid (18 : 1), linoleic acid (18 : 2), and either γ- or α-linolenic acid (18 : 3), according to the genus. A number of conjugated lipids also occur that are associated with the cell envelope; these include lipopolysaccharides forming lipid A (see Fig. 5) as part of the cell envelope. The cyanobacterial cells are usually low in total lipid contents (<15%) and, because of the wide diversity of types, are not usually considered commercially exploitable, given the impracticality of purifying individual lipids. However, an exception may be possible

with *Spirulina*, which has been researched with some enthusiasm in Israel by Zvi Cohen and his colleagues, at the Ben Gurion University, in the Negev Desert.

Spirulina platensis and *Spirulina maxima* are multicellular, filamentous cyanobacteria that profusely colonize lakes in Africa and Mexico. They have been used as indigenous supplies of food and feed and are of some commercial interest because both have been known for some time to contain γ-linolenic acid. Although the algae grow well in outdoor lagoons in hot, sunny climates and can, indeed, produce in the extracted lipid a GLA content exceeding 20%, the distribution of GLA between neutral lipid (e.g. triacylglycerols), glycolipid, and phospholipid fractions is about 1 : 9 : 4. This, coupled with the low (<6%) content of lipid in the cells, probably makes it a commercially unattractive source of the fatty acid. With appropriate marketing, however, it could be used to boost the nutritional claims for whole-cell *spirulina* as a dietary supplement.

6.2 Eukaryotic Algae Lipids

6.2.1 Carotenoids

Of the nonfatty acyl lipids, carotenoids have been of some commercial interest. The production of β-carotene (**IV**) by *Dunaliella salina* has been investigated by a number of groups worldwide, with those in Israel, Western Australia, and California being particularly active. Commercialization of this process has now been developed in Western Australia by Aqua-Carotene Ltd., which uses outdoor cultivation ponds to produce the biomass.

Also of increasing commercial interest is another carotenoid, astaxanthin, which is used principally as a colorant for feeding farmed fish. This is produced by *Haematococcus pluvialis*, which can contain over 2% of its biomass as astaxanthin. It is grown in large-scale ponds and is sold as an alga meal by a number of companies. Chemically produced carotenoids, however, remain cheaper than the algal sources but do not have the benefit of being able to claim that they are "natural" ingredients. This is possibly not a deterrent when considering large-scale use where cost is the priority, though, legislation in individual countries may dictate otherwise.

6.2.2 Polyunsaturated Fatty Acids

Several commercial processes now exist for the production of the very long-chain fatty acid, docosahexaenoic acid, 22 : 6 n-3, (DHA). This fatty acid is now recommended for inclusion in infant formulae along with arachidonic acid, see Sect. 5.3. It is also recommended for dietary supplementation for adults, where it may offer protection against coronary heart diseases. It is preferred, in some cases, to fish oil which, though containing DHA, also contains eicosapentaenoic acid, 20 : 5 n-3, (EPA) and which may be contraindicated as it can interfere with the uptake and functional role of DHA. Fish oil is not allowed as a supplement for infant formulae in some countries, including the United States, because of its possible contamination with environmental toxins and heavy metals.

The main production organism for DHA is *Crypthecodium cohnii*, which is a unicellular, nonphotosynthetic dinoflagellate. It is grown in large-scale fermenters of up to 100 m^3. The principal company behind the development of the process and the marketing of the oil is Martek Biosciences Corp., Maryland, United States. The oil is known by the trade name of DHASCO and, when mixed with arachidonic acid for

incorporation into infant foods, is known as Formulaid. The fatty acid profile of the oil is shown in Table 5, where it can be seen that DHA is the sole polyunsaturated fatty acid thereby making the oil highly desirable as a source of this single PUFA.

Another related process for the production of DHA uses another marine organism known as *Schizochytrium*. The inclusion of this organism as an alga is uncertain as the taxonomic position of the group of organisms, that includes *Schizochytrium* and the related group of thraustochytrid organisms, continues to be researched. *Schizochytrium*, like *Crypthecodinium cohnii*, is nonphotosynthetic and is also grown in large-scale bioreactors using glucose as substrate. The oil produced is slightly different from the oil of *C. cohnii* in that it contains a small but appreciable percentage of docosapentaenoic acid, 22 : 5 n-6. The full profile of the fatty acids are shown in Table 5.

The Schizochytrium process was originally developed by OmegaTech Ltd, Boulder, CO, which is now owned by Martek Biosciences Corp. The oil was originally given the trade name of DHAGold and was used as an animal feed supplement to boost the content of DHA in the final product. It was particularly useful with poultry where it led to eggs being enriched in DHA and thus, could attract a premium price. The oil itself has now received approval as a direct supplement for humans, including infants, and will, in the future, be used for this purpose. Its current trade name is DHASCO-S.

Some interest has continued in the use of photosynthetically grown algae for polyunsaturated fatty acid production. Although the organisms can be grown outdoors using sunlight as energy source and CO_2 as a carbon source, the costs of production remain very high because of the slow growth rate of the cells and the low cell densities that are produced. It may take four to six weeks to generate algal cells at no more than 5 g L^{-1} in an outdoor system, whereas *Schizochytrium*, as discussed earlier, when grown heterotrophically in a stirred fermenter it can reach biomass densities of up to 200 g L^{-1} in about four days. However, there are still several groups who continue to use microalgae for the production of various fatty acids not easily obtainable from other sources. Of particular interest is the possible production of EPA for which no source, as a single PUFA, currently exists apart from very expensive purification from fish oils that contain both EPA and DHA. The demand for an oil rich in EPA as a single PUFA is likely to increase as it has been recommended for use in the treatment of various disorders of the brain, including bipolar disorder and schizophrenia. There have also been suggestions that a high dosage of EPA may help in the treatment of certain cancers though this is highly contentious. For these reasons, there is likely to be considerable interest in obtaining a good supply of this PUFA from whatever source and, currently, photosynthetic algae appear to offer the best prospects.

See also Bacterial Cell Culture Methods; FTIR of Biomolecules; Microbial Development.

Bibliography

Books and Reviews

Cohen, Z. (Ed.) (1999) *Chemicals from Microalgae*, Taylor and Francis, London.

Cohen, Z., Ratledge, C. (Eds.) (2005) *Single Cell Oils – 2*, American Oil Chemists' Society, Champaign, IL.

Cronan, J.E. (2003) Bacterial membrane lipids: where do we stand? *Annu. Rev. Microbiol.* **57**, 203–224.

Horikoshi, K., Grant, W.D. (Eds.) (1998) *Extremophiles: Microbial Life in Extreme Environments*, Wiley Liss, NJ.

Kroes, R., Schaefer, E.J., Squire, R.A., Williams, G.M. (2003) A review of the safety of DHA oil, *Food Chem. Toxicol.* **41**, 1433–1446.

Ratledge, C., Wilkinson, S.G. (Eds.) (1988,1989) *Microbial Lipids*, Vols. 1 and 2. Academic Press, London and New York, 963–726.

Primary Literature

Murphy, D.J. (2001) The biogenesis and functions of lipid bodies in animals, plants and microorganisms, *Prog. Lipid Res.* **40**, 325–438.

Nikaido, H. (2003) Molecular basis of bacterial outer membrane permeability revisited, *Microbiol. Mol. Biol. Rev.* **67**, 593–656.

Ratledge, C., Wynn, J.P. (2001) The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms, *Adv. Appl. Microbiol.* **51**, 1–51.

Sorger, D., Daum, G. (2003) Triacylglycerol biosynthesis in yeast, *Appl. Microbiol. Biotechnol.* **61**, 289–299.

Wen, Z.-Y., Chen, F. (2003) Heterotrophic production of eicosapentaenoic acid by microalgae, *Biotechnol. Adv.* **21**, 273–294.

Lipoprotein Analysis

Alan T. Remaley[1] *and G.Russell Warnick*[2]
[1] *National Institutes of Health, MD, USA*
[2] *Pacific Biometrics Research Foundation Issaquah, WA, USA*

1 Overview of Lipoproteins 278
1.1 Lipid Biochemistry 278
1.2 Lipoprotein Structure and Function 280

2 Analysis of Lipoprotein Components 283
2.1 Total Cholesterol Assay 283
2.2 Triglyceride Assay 284
2.3 Phospholipid Assay 285
2.4 Apolipoprotein Assays 285

3 Fractionation of Lipoproteins 285
3.1 HDL-C Assays 286
3.2 LDL-C Assays 287
3.3 Electrophoresis of Lipoproteins 288
3.4 Subfractionation of Lipoproteins 289

4 Diagnostic Application of Lipoprotein Analysis 289
4.1 Standardization of Lipoprotein Analysis 289
4.2 Interpretation of Lipoprotein Assay Results 291

5 Perspective 292

Bibliography 293
Books and Reviews 293
Primary Literature 293

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

Keywords

Apolipoprotein
A protein found on lipoprotein particles that functions to maintain the structural integrity of the lipoprotein particle and to modulate lipoprotein metabolism.

Atherosclerosis
Deposition of lipid in the intimal wall of large- to medium-sized arteries that impinges on blood flow and/or predisposes the vessel to thrombosis.

Cholesterol
A 27-carbon sterol that is a common lipid constituent of cell plasma membranes and lipoproteins.

Coronary Artery Disease
The presence of atherosclerosis in the coronary artery vessels of the heart, which can interfere with blood flow.

Lipoprotein
A small complex of lipid and protein that serves as the main carrier of extracellular lipid.

Lipoproteins are submicron- to micron-sized particles that are composed of both lipids and proteins. Lipoproteins are the major extracellular carrier of lipids, such as cholesterol, and thus, they play an important role in the pathogenesis of atherosclerosis. The analysis of lipoproteins is performed in most clinical laboratories to identify patients at risk of coronary heart disease (CHD). Lipoprotein analysis is also essential in monitoring the effectiveness of cholesterol-lowering therapy.

The focus of this chapter will be on methods for lipoprotein analysis that are used for routine diagnostic purposes, but in addition, new promising types of lipoprotein analysis that are still in the realm of basic research will also be highlighted. A brief description of lipid biochemistry and lipoprotein metabolism will be followed by a review of commonly used methods for analyzing lipoproteins. Finally, the chapter will conclude with a summary of how the results of lipoprotein analysis are used in the diagnosis and management of patients with dyslipidemias.

1 Overview of Lipoproteins

1.1 Lipid Biochemistry

Lipids, commonly referred to as *fats*, are composed of mostly carbon and hydrogen (Fig. 1), and are therefore hydrophobic. Two major classes of lipids, based on their relative hydrophobicity, are found on lipoproteins. Phospholipids and cholesterol are classified as amphipathic lipids because, in addition to C-H bonds, they also contain polar groups. In contrast,

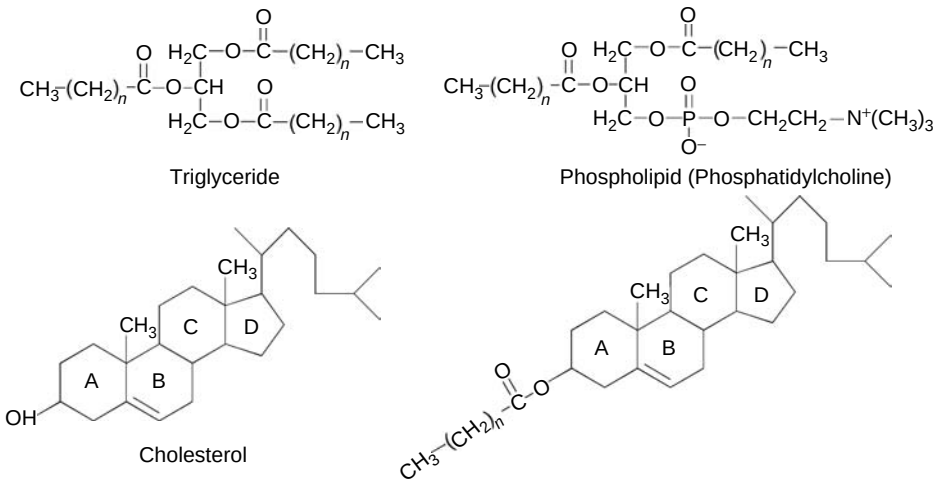

Fig. 1 Chemical structures of lipids found on lipoproteins.

cholesteryl esters and triglycerides do not contain any polar constituents and are therefore more hydrophobic and are classified as neutral lipids.

A prototypical structure of a lipoprotein particle is shown in Fig. 2. Phospholipids and cholesterol form a monolayer on the surface of lipoprotein particles. Phospholipids are oriented so that their charged head groups are facing outward, away from the neutral hydrophobic core. Cholesterol is also oriented so that its polar hydroxyl group on the A ring (Fig. 1) is pointing away from the lipoprotein surface. Cholesteryl esters and triglycerides are found in the core of lipoprotein particles in a disorganized liquid crystalline state.

Except for cholesterol, all of the other major lipids found on lipoprotein have numerous distinct chemical forms (Fig. 1). The different forms of triglyceride can be identified on the basis of the type of fatty acids that they contain. Fatty acids can be classified on the basis of the number of carbon atoms as short-chain (4–6 carbon atoms), medium-chain (8–12 carbon atoms), or long-chain (>12 carbon atoms) fatty acids. Most dietary fatty acids are long chain and contain an even number of carbon atoms. Fatty acids also differ on the basis of the number of unsaturated or double carbon bonds that they contain. Fatty acids with no double bonds are classified as saturated fatty acids, whereas fatty acids with one double bond are monounsaturated and fatty acids with two or more double bonds are polyunsaturated. The different fatty acids are commonly designated by their carbon-chain length and by the number of double bonds. For example, arachidonate, a polyunsaturated fatty acid that has 20 carbon atoms and 4 double bonds is designated as 20 : 4. The common fatty acids found in human serum triglycerides are palmitate (16 : 0), oleate (16 : 1), and stearate (18 : 0). Triglycerides containing shorter chain fatty acids or unsaturated fatty acids have a lower melting point and are usually liquid at room temperature. Saturated fatty acids are commonly found in food from animal sources, whereas unsaturated fatty acids are abundant in food derived from plants.

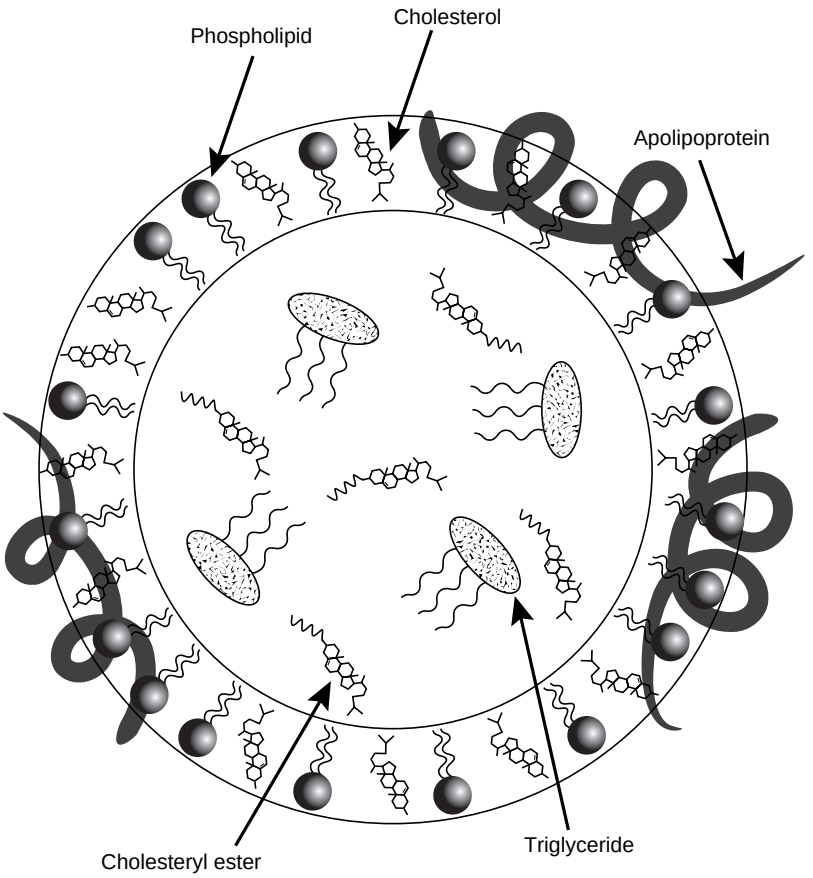

Fig. 2 Model of lipoprotein particle.

Similar to triglycerides, there are different forms of cholesteryl esters, depending on the type of fatty acid that is esterified to the A ring of cholesterol. Cholesteryl linoleate, which is primarily produced by the serum enzyme lecithin: cholesterol acyltransferase (LCAT), and cholesteryl oleate, which is primarily produced by the intracellular enzyme acylcoenzyme A: cholesterol acyltransferase (ACAT), are the two most common forms of cholesteryl esters in humans.

Phospholipids can differ not only by the type of fatty acids that they contain but also by their different head groups (choline, inositol, serine, glycerol, and ethanolamine), which serves as the basis for their nomenclature. Phosphatidylcholine (Fig. 1), which contains a choline head group, is the most common type of phospholipid found on lipoproteins.

1.2 Lipoprotein Structure and Function

Lipoproteins are typically spherical in shape and range in size from 5 to 1200 nm (Table 1). In addition to lipids, lipoproteins also contain proteins, called *apolipoproteins*, which reside on the surface of

Tab. 1 Characteristics of the major human lipoproteins.

Features	*Chylos*	*VLDL*	*LDL*	*HDL*
Density (g mL^{-1})	<0.93	0.93–1.006	1.019–1.063	1.063–1.21
Molecular weight (kDa)	$(0.4–30) \times 10^9$	$(10–80) \times 10^6$	2.75×10^6	$(1.75–3.6) \times 10^5$
Diameter (nm)	80–1200	30–80	18–30	5–12
Total lipid (% weight)	98	89–96	77	50
Triglyceride (% weight)	84	44–60	11	3
Total cholesterol (% weight)	7	16–22	62	19
Site of synthesis	Intestine	Liver	Catabolism of VLDL	Liver, intestine
Electrophoretic mobility	Origin	Pre-beta	Beta	Alpha

Tab. 2 Characteristics of the major human apolipoproteins.

Apolipoprotein	*Molecular weight [kDa]*	*Plasma concentration [mg dL^{-1}]*	*Major lipoprotein location*	*Function*
ApoA-I	28 000	100–200	HDL	Structural, LCAT activator, ABCA1 lipid acceptor
ApoA-II	17 400	20–50	HDL	Structural
ApoA-IV	44 000	10–20	Chylos, VLDL, HDL	Structural
ApoB-100	5.4×10^5	70–125	LDL, VLDL	Structural, LDL receptor ligand
ApoB-48	2.6×10^5	<5	Chylos	Structural, remnant receptor ligand
ApoC-I	6630	5–8	Chylos, VLDL, HDL	Structural
ApoC-II	8900	3–7	Chylos, VLDL, HDL	Structural, LPL cofactor
ApoC-III	9400	10–12	Chylos, VLDL, HDL	Structural, LPL inhibitor
ApoE	34 400	3–15	VLDL, HDL	Structural, LDLreceptor, and remnant receptor ligand
Apo(a)	$(3–7) \times 10^5$	<30	Lp(a)	Unknown

lipoprotein particles (Fig. 2). Apolipoproteins can serve both a structural role in maintaining the integrity of the lipoprotein particle and also a functional role in lipoprotein metabolism. Apolipoproteins contain a structural motif called an *amphipathic helix,* which enables them to bind to lipids. Amphipathic helices are α-helices with approximately half of the helix composed of hydrophilic amino acid residues and the other half composed of hydrophobic amino acid residues. They are oriented on the surface of lipoprotein particles to allow the interaction of the hydrophobic face of the helix with the surface lipids. Apolipoproteins also play a major role in modulating lipoprotein metabolism by acting as ligands for various lipoprotein receptors and by acting as activators and inhibitors of several lipid-modifying enzymes (Table 2).

Several different classification systems based on their varying physical properties can be used for categorizing lipoprotein particles. A density classification system (Chylomicrons (Chylos), very low density

lipoproteins (VLDL), intermediate density lipoproteins (IDL), low density lipoproteins (LDL), and high density lipoproteins (HDL)), which was developed on the basis of the original ultracentrifugation system for isolating lipoproteins, is most common (Table 1). The size of the lipoprotein particles is inversely related to their density. Larger lipoprotein particles have relatively more core lipid and less protein and are, therefore, lighter in density.

Three major pathways in lipoprotein metabolism and how the different types of lipoprotein particles are involved in each pathway are shown in Fig. 3. Chylomicrons are the largest and the least-dense lipoprotein particles and contain a single molecule of apoB-48 as well as several other types of apolipoproteins (Table 2). Chylomicrons are produced by the intestine and they participate in the exogenous pathway, which delivers dietary lipids to the liver and to other peripheral tissues. The turbidity of postprandial serum is due to the presence of chylomicrons, which reflect light because of their large size. Once chylomicrons enter the circulation, they are quickly catabolized within a few hours by lipoprotein lipase (LPL) and hepatic lipase (HL), which are on the surface of endothelial cells and hydrolyze triglycerides and phospholipids (Fig. 3). The free fatty acids that are generated are taken up by tissues and are either reesterified and stored in lipid drops or are used as a source of energy. Chylomicrons are transformed by this process into the more dense chylomicron remnants, which are primarily removed by the liver via a chylomicron remnant receptor that binds apoE.

VLDL, IDL, and LDL participate in the endogenous pathway of lipid metabolism, which transports hepatic-derived lipids to peripheral tissues (Fig. 3). VLDL has a single copy of apoB-100; it is produced in the liver and is rich in triglycerides synthesized by the liver. Like chylomicrons, the core lipids of VLDL are acted upon by LPL and HL, and the fatty acids released are taken up by peripheral tissues. As a consequence of the lipolysis, a significant fraction of VLDL is transformed into IDL and ultimately into LDL. The majority of LDL is returned to the liver via the LDL receptor; however, LDL can also be taken up in peripheral tissues by the LDL receptor.

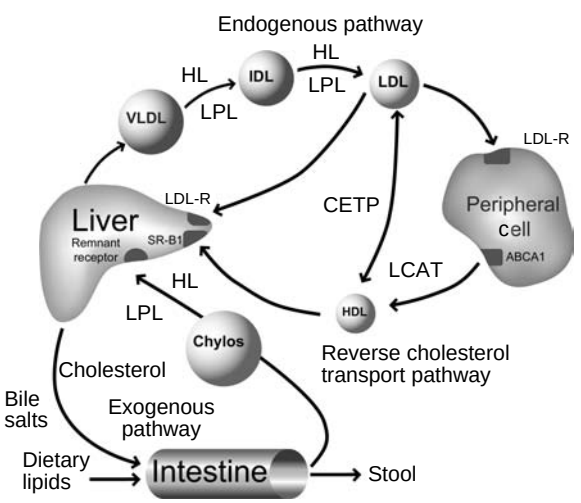

Fig. 3 Diagram of lipoprotein metabolism.

Because of its small size, LDL can also readily infiltrate into the vessel wall and become trapped there by proteoglycans. LDL in the vessel wall is prone to oxidation, which then makes it a better ligand for scavenger receptors on macrophages. Macrophages that take up too much lipid become filled up with intracellular lipid drops and are transformed into foam cells. A collection of foam cells in the vessel walls forms a fatty streak, which is an early precursor of atherosclerotic plaques. Oxidized LDL can also initiate the inflammation of the vessel wall by acting as a chemotactic agent and by inducing the production of various proinflammatory cytokines by macrophages. Another proatherogenic lipoprotein similar to LDL in structure is Lipoprotein (a) (Lp(a)). It contains a single copy of apoB-100, which is covalently linked by a disulfide bond to apo (a), a plasminogen-like protein of unknown function.

High-density lipoprotein is the principal lipoprotein that mediates the reverse cholesterol-transport pathway, delivering the excess cholesterol in peripheral cells back to the liver for excretion (Fig. 3). Because most peripheral cells cannot catabolize cholesterol, the reverse cholesterol-transport pathway is one of the main mechanisms for maintaining cellular cholesterol homeostasis. The principal apolipoprotein on HDL is apoA-I, which is sometimes used as a marker for HDL. In contrast to LDL, HDL has anti-inflammatory properties and is considered to be antiatherogenic because it also mediates the removal of excess cholesterol from peripheral cells. The process of reverse cholesterol transport is initiated when lipid-poor apoA-I interacts with the ABCA1 transporter on the cell surface and extracts excess cholesterol from cells. Cholesterol that is removed from cells is then trapped on HDL after it is transformed into cholesteryl ester by LCAT (Fig. 3). HDL cholesterol is then returned either directly to the liver by the SR-B1 receptor or indirectly after being transferred to LDL by the cholesteryl ester transfer protein (CETP) (Fig. 3).

2 Analysis of Lipoprotein Components

The initial analysis of lipoproteins usually begins with a direct measurement of total cholesterol and triglyceride from serum or plasma, without any prior extraction or fractionation step. Apolipoproteins can also be measured directly on serum or plasma. There are several chemical methods for measuring the various lipids on lipoproteins, which are still used as reference methods by the Centers for Disease Control and Prevention (CDC) and the associated Cholesterol Reference Method Laboratory Network (CRMLN). These chemical methods, however, are complex, laborious, and expensive to perform and have largely been replaced in routine clinical laboratories by automated enzymatic methods.

2.1 Total Cholesterol Assay

A coupled enzyme reaction (Fig. 4), which depends on the specificity of cholesterol oxidase for cholesterol, is the most common enzyme method for measuring total cholesterol. Before measuring cholesterol, cholesteryl ester, which makes up about two-thirds of total cholesterol, is first converted to free cholesterol by cholesteryl esterase. The free cholesterol generated then reacts with cholesterol oxidase, which results in the production of cholestenone

- Cholesteryl ester + H_2O $\xrightarrow{\text{Cholesteryl esterase}}$ Cholesterol + Fatty acid
- Cholesterol + O_2 $\xrightarrow{\text{Cholesterol oxidase}}$ Cholestenone + H_2O_2
- H_2O_2 + Dye $\xrightarrow{\text{Cholesterol oxidase}}$ Color

Fig. 4 Enzymatic assay for cholesterol.

and hydrogen peroxide. The amount of cholesterol in the sample is quantified by measuring the amount of hydrogen peroxide generated by using a horseradish peroxidase–dye system (Fig. 4). The assay is monitored spectrophotometrically at approximately 500 nm and can be performed on serum or plasma without any prior extraction with an organic solvent. Modern reagents include other constituents that eliminate most sources of interference, although specimens with marked elevations of vitamin C or bilirubin and specimens with gross hemolysis can experience interference with the peroxidase catalyzed color reaction. Most commercial assays are developed for specific analyzers and should yield results within 3% of the result obtained by the reference method, with a coefficient of variation of less than 3%. Although it is not frequently performed for diagnostic purposes, free (nonesterified) cholesterol can be simply measured by omitting the cholesterol ester–hydrolase step.

2.2 Triglyceride Assay

There are several different colorimetric enzyme assays for triglycerides, which primarily differ in how they are coupled to an indicator dye. As shown in Fig. 5, all of these methods begin with the enzymatic hydrolysis of fatty acid from the glycerol backbone, with a lipase. In one of the more common methods (Fig. 5), the amount of triglyceride is measured by quantifying the glycerol, using glycerokinase and glycerophosphate oxidase and the same dye-peroxidase system used in the cholesterol assay. One problem with this approach is that the assay will also measure any endogenous free glycerol in a sample. For most samples, this will result in a clinically insignificant positive bias of less than 10 to 20 mg dL^{-1}, but in some disorders, such as diabetes, or in patients receiving total parenteral nutrition, the bias can become clinically significant. It is for this reason

- Triglyceride + H_2O $\xrightarrow{\text{Bacterial lipase}}$ Fatty acid + Glycerol
- Glycerol + ATP $\xrightarrow{\text{Glycerokinase}}$ Glycerophosphate + ADP
- Glycerophosphate + O_2 $\xrightarrow{\text{Glycerophosphate oxidase}}$ Dihydroxyacetone + H_2O_2
- H_2O_2 + Dye $\xrightarrow{\text{Peroxidase}}$ Color

Fig. 5 Enzymatic assay for triglyceride.

that some laboratories compensate for this problem by measuring and correcting for the free glycerol in a blanking procedure or by performing an assay that consumes any free glycerol in the first step, thus eliminating the measurement of any endogenous glycerol later on in the reaction. Alternatively, the manufacturers of some triglyceride assays compensate for this bias by adjusting the value of their calibrators.

2.3 Phospholipid Assay

The measurement of phospholipids has not been shown to have any diagnostic value and is commonly done only in research laboratories. The choline-containing phospholipids, lecithin, lysolecithin, and sphingomyelin, account for at least 95% of total phospholipids in serum and can be measured enzymatically with commercial kits, using phospholipase D, choline-oxidase, and horseradish peroxidase. To measure total phospholipids, independent of the head group, phosphorus is measured chemically after it is released by acid hydrolysis from an extracted sample. The individual types of phospholipids can be measured semi-quantitatively after separation by thin-layer chromatography.

2.4 Apolipoprotein Assays

Because of their relatively unique distribution on the different lipoprotein particles (Table 2), apoA-I and apoB-100 in serum are used as an index of the amount of HDL and LDL respectively. Because these two proteins are relatively abundant, they can be measured with specific antibodies by either turbidometric or nephelometric procedures. These assays are now routinely available on standard clinical analyzers and are free from most interferences except for problems related to light scattering from turbid samples with elevated chylomicrons and VLDL. Tests for apoA-I and apoB-100 have been shown to be more precise and accurate than tests for HDL-C and LDL-C. The apolipoproteins have been proposed as being possibly superior for predicting CHD (coronary heart disease) risk, but they have not yet been incorporated into the National Cholesterol Education Program (NCEP) practice guidelines for lipoprotein analysis.

The proatherogenic Lp(a) lipoprotein can also be measured by quantifying the presence of the apo(a) apolipoprotein, using various ELISA or nephelometric methods. There are, however, different size polymorphisms of apo(a), which complicate the immunologic measurement and differ in their correlation with CHD risk. A qualitative assessment of isoform distribution by electrophoresis or by genotyping is also used for characterizing Lp(a) and for estimating CHD risk.

3 Fractionation of Lipoproteins

Numerous methods for fractionation of the various lipoproteins have been developed on the basis of their different physical properties, such as size, density, charge, solubility, and apolipoprotein content. Historically, ultracentrifugation was the primary means of separating the different lipoprotein fractions on the basis of density and is still the foundation of the current lipoprotein classification system (Table 1). The primary reason for fractionating lipoproteins is to provide a measure

of their relative abundances. Fractionation is particularly useful for measuring the cholesterol content of LDL (LDL-C) and HDL (HDL-C), which are used in calculating CHD risk. Electrophoresis is the other commonly used method in clinical laboratories for fractionating lipoproteins.

3.1 HDL-C Assays

Until recently, the most common method for measuring HDL-C has been a multistep precipitation procedure, which involved the removal of apoB-containing lipoproteins (chylos, VLDL, IDL, LDL, and Lp(a)) by precipitation and centrifugation, followed by the enzymatic measurement of cholesterol in the supernatant. A wide variety of polyanions, such as heparin, phosphotungstate, and dextran sulfate will bind to positively charged groups on apoB-containing lipoproteins and cause their aggregation and precipitation, in the presence of divalent cations, such as manganese and magnesium. Magnesium has become the divalent cation of choice because of interference with enzymatic lipid assays by manganese. Numerous commercial assays have been developed with the different precipitating reagents. Initially, results from these different assays often yielded discordant results, but this problem has largely been resolved by the ongoing efforts to improve the standardization of lipoprotein assays.

A major limitation of the various HDL-precipitation methods is that specimens with elevated triglycerides often have a positive bias for HDL-C. Because of their low density, specimens with high levels of chylomicrons and triglyceride-rich VLDL will often form aggregates that do not fully sediment upon centrifugation. This leads to a false elevation for HDL-C, because, when cholesterol is measured in the supernatant, not only cholesterol from HDL is measured but cholesterol from any residual aggregates is also measured. This problem can at least be partially solved by predilution of the sample, which facilitates the sedimentation of the aggregates. Although it is laborious, ultracentrifugation or ultrafiltration of the sample to remove any residual aggregates will also eliminate this source of bias.

Unlike the precipitation-based HDL-C tests, most other commonly measured clinical laboratory tests do not require any significant specimen preprocessing. This limitation of the HDL-C-precipitation tests has prompted the recent development of homogenous assays for HDL-C that can be performed directly on serum or plasma, without the physical removal of non-HDL lipoproteins. These assays work by using reagents that chemically mask the non-HDL lipoproteins during the enzymatic measurement of cholesterol associated with HDL. This can be accomplished by using different strategies that involve the use of polymers, detergents, antibodies, or even modified enzymes. These various reagents bind to non-HDL lipoproteins and block their accessibility to cholesterol esterase and cholesterol oxidase. Other homogenous assays selectively deplete non-HDL-C in the first step, thus preventing its measurement later on in the reaction. Several commercial homogenous HDL-C assays are now available on automated instruments, and these assays show relatively good accuracy and typically better precision than precipitation-based assays. Precipitation-based HDL-C assays have now largely been replaced by homogenous HDL-C assays in most routine clinical laboratories, but they still have a role in research laboratories and in specialty lipid laboratories because the

homogenous HDL-C assays do not always yield good results in patients with unusual dyslipidemias and/or in patients with severe liver or kidney disease.

3.2 LDL-C Assays

Traditionally, methods for measuring LDL-C have been more difficult to perform than HDL-C assays because there is no convenient precipitation procedure to cleanly separate LDL from the other apoB-containing lipoproteins. As a consequence, the original LDL-C methods involved a procedure commonly referred to as *β-quantification*, which involves both a precipitation and an ultracentrifugation step. In the first step, chylomicrons and VLDL are removed from a sample by ultracentrifugation (105 000 G for 18 h at 10 °C) at the native density of plasma (1.006 g L^{-1}), which results in their flotation at the top of the tube. Chylomicrons and VLDL are then removed with a tube slicer and cholesterol in the infranate, which consists of HDL-C and LDL-C, is measured using an enzymatic assay. HDL-C is then measured either from the infranate or from the original sample after precipitation and removal of LDL. LDL-C is calculated by subtracting HDL-C from the cholesterol value obtained from the infranate. The β-quantification method serves as the basis for the LDL-C reference method, but because of its complexity and the requirement of an ultracentrifuge, the procedure is primarily used by specialty lipid laboratories.

Instead of directly measuring LDL-C, most clinical laboratories calculate it from the results of a fasting sample, using the following Friedewald equation: (LDL-C = Total cholesterol – (HDL-C + triglyceride/5); all units in mg dL^{-1}). In a fasting sample without chylomicrons, cholesterol can be mostly found in LDL, HDL, and VLDL and triglycerides are mostly found in VLDL. The term triglyceride/5 in the equation has been shown to closely approximate VLDL-C because there is about five times more triglyceride than cholesterol in a typical VLDL particle when both values are expressed in units of mg dL^{-1}. LDL-C can, therefore, be simply calculated by subtracting the measured HDL-C and the calculated VLDL-C from the measured total cholesterol. However, the Friedewald equation has been shown to yield inaccurate results in samples that have VLDL particles with abnormal lipid composition. This rarely occurs but can be observed in patients with fasting triglycerides greater than 400 mg dL^{-1}, who have triglyceride-rich VLDL, and in patients with type-III hyperlipidemia, who have cholesterol-rich VLDL. Thus, by the Friedewald equation, LDL-C can be reasonably estimated in most samples by simply measuring total cholesterol, triglyceride, and HDL-C. These three tests and the calculated LDL-C comprise what has now become commonly known as a "lipid panel."

One potential limitation of both the β-quantification procedure and the Friedewald calculation is that they both measure, in addition to LDL-C, the relatively small amounts of cholesterol that are associated with Lp(a) and IDL. In addition, the β-quantification procedure is complex and there have been concerns about the precision and accuracy of the Friedewald equation because of the cumulative effect of errors in the three underlying tests that are used in the calculation. These limitations have led to the development of direct homogenous LDL-C assays. Like the homogenous HDL-C assays, these assays use reagents that selectively block or solubilize and consume the different lipoprotein

classes, allowing the direct measurement of LDL-C from an unfractionated sample. Several automated commercial homogenous LDL-C assays are currently available. In general, these assays show good precision and appear to be accurate on normal specimens. Homogenous LDL-C assays, however, have not yet replaced the use of calculated LDL-C in most clinical laboratories for several reasons. First, the homogenous LDL-C assays are relatively new and, in many cases, have not been fully evaluated in patients with the full spectrum of dyslipidemias. Early results, in fact, suggest that the homogenous LDL-C assays are not highly specific in their separation of the various lipoproteins and do not consistently agree with β-quantification and are typically no more reliable than Friedewald calculation. In addition, because the three tests that make up the lipid panel are usually necessary for a complete lipoprotein evaluation, it has not yet been clearly established that performing a separate direct test for LDL-C offers any real advantages in the diagnosis and management of patients with dyslipidemias.

3.3 Electrophoresis of Lipoproteins

Electrophoresis is a relatively convenient method for separating the major classes of lipoproteins. Electrophoresis of lipoproteins can be performed on a variety of matrices, but separation on agarose gels is relatively easy to perform and it readily resolves the major lipoprotein classes on the basis of both size and charge. Some typical electrophoretic patterns of lipoproteins are shown in Fig. 6. HDL shows the fastest migration on agarose gels and migrates to what is called the *α-position*. Chylomicrons barely migrate past the origin of the gel, and LDL typically migrates to the intermediate β-position. VLDL migrates to the pre-β-position between HDL and LDL, but in type-III hyperlipidemia, it usually migrates as a broad band in the β-region. The lipoproteins in agarose gels are usually stained with a dye that stains neutral lipids, such as Oil Red O.

Early on, the electrophoretic separation of lipoproteins was an important diagnostic tool because the original phenotypic classification of dyslipidemias was based on this method. With the discovery of the genetic basis of many of the common dyslipidemias and with the advent of more specific tests, a genetic classification of lipoproteins is now more commonly used. As a consequence, electrophoresis is no longer as valuable diagnostically, but it is still useful for identifying and monitoring patients with rare variants of dyslipidemias. Standard electrophoretic methods are also only semiqualitative, which limits their use. Recently, however, commercial

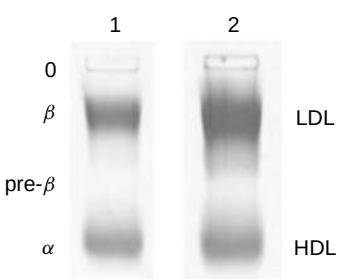

Fig. 6 Agarose gel electrophoresis of a normal subject (Lane 1) and a patient with familial hypercholesterolemia, with an elevated LDL (Lane 2).

automated electrophoretic systems have been developed, which have been shown to be relatively precise and accurate in quantifying the major lipoprotein fractions. There has also been a recent resurgence in the interest of electrophoretic techniques for separating lipoproteins because of their ability to resolve subfractions within each of the major lipoprotein classes.

3.4 Subfractionation of Lipoproteins

Besides the major lipoprotein classes shown in Table 2, it is possible to detect distinct subspecies or subclasses within each class. Most of the methods used in fractionation of the major subclasses of lipoproteins can be modified to improve their resolution for performing subfraction analysis. In addition, nuclear magnetic resonance spectroscopy and density gradient ultracentrifugation in a vertical rotor are particularly useful for performing a comprehensive subfraction analysis of lipoproteins. Nondenaturing gradient gel electrophoresis in polyacrylamide is also a good method for separating the subfractions of lipoproteins by size. When combined with agarose gel electrophoresis, in a two-dimensional gel-electrophoresis system, as many as 10 different species of HDL can be observed.

Subfraction analysis of lipoproteins has largely been performed for basic research applications, but increasingly, subfractions within both LDL and HDL have been shown to potentially have diagnostic importance. A subfraction of LDL called *small dense LDL* has been proposed to be particularly proatherogenic. This form of LDL also appears to be at least partly inheritable. Approximately 25% of the Caucasian population has what is called the *subclass B phenotype* of LDL, which is associated with an increase in small dense LDL and an increased risk for CHD.

In the case of HDL, some forms of HDL have been shown to be particularly efficient in promoting reverse cholesterol transport and may even be a stronger negative predictor of CHD than total HDL. One such form is a small lipid-poor form of HDL, which has a pre-β-type migration on agarose gel electrophoresis. Pre-β-HDL can also be measured by an ELISA assay. Unlike other forms of HDL, pre-β-HDL is discoidal in shape because it lacks a central lipid core of neutral lipids. It consists of a bilayer of phospholipids around which apolipoproteins surround the acyl chains of the phospholipids in a beltlike configuration. Pre-β-HDL has been proposed to be particularly antiatherogenic because it is the form of HDL that initially interacts with cells for removing excess cholesterol by the ABCA1 transporter (Fig. 3).

4 Diagnostic Application of Lipoprotein Analysis

4.1 Standardization of Lipoprotein Analysis

There are three important sources of variability or error that can affect the results of lipoprotein analysis, namely, biological, preanalytical, and analytical variability. There has been great improvement in the last several years in analytical systems for measuring lipids and lipoproteins, and systematic biases among the various lipid and lipoprotein assays have significantly decreased over the last several years. This has occurred because of the partnership between the clinical laboratory community, diagnostic companies, and

various governmental regulatory bodies in improving the accuracy of lipoprotein analysis. In order to minimize the misdiagnosis of patients, the National Cholesterol Education Program or NCEP, a National Institute of Health sponsored organization that produces practice guidelines for the diagnosis and treatment of patients for coronary artery disease, has developed accuracy and precision goals for lipoprotein analysis (Table 3). The CDC created the Cholesterol Reference Method Laboratory Network (CRMLN; http://www.cdc.gov/nceh/dls/crmln/crmln.htm), which provides a mechanism for diagnostic companies and routine clinical laboratories to compare the performance of their assays to the reference methods, which are traceable to the definitive assays established by the National Institute of Standards and Technology (NIST). Clinical laboratories are also required, as part of their biannual recertification process, to compare themselves, multiple times each year, to their peers that use the same assays by analyzing unknown control material, through one of the national proficiency testing programs, such as the one offered by the College of American Pathologists. It is also mandatory for clinical laboratories to test their assays daily by measuring control materials. In addition, it is important to use a reagent system with an analyzer for which the assay has been developed, in order to ensure the proper performance of the assay. Because of the underlying errors in total cholesterol and HDL-C assays, it is often particularly difficult, however, for routine clinical laboratories to achieve the NCEP accuracy goals for LDL-C by using the Friedewald equation. Further improvements are considered necessary in the current total cholesterol and HDL-C assays or in direct LDL-C assays to improve the accuracy and precision of LDL-C measurements.

It is important also to consider preanalytical sources as a source of error. It is recommended that a blood sample be collected from subjects in a sitting position because posture can affect the results of lipoprotein analysis, through changes in hemoconcentration. In addition, different types of collection tubes can lead to differences in results. Most clinical laboratories use serum for their analysis, but EDTA plasma is often preferred when lipoproteins are isolated by ultracentrifugation because EDTA helps reduce proteolysis and oxidation of the lipoproteins. EDTA in the collection tube will, however, cause slight dilution of the sample because it increases the osmotic strength of plasma, which causes water from red blood cells to shift to the plasma compartment. Lipoprotein results from EDTA plasma should, therefore, be multiplied by 1.03 to match serum results.

Tab. 3 NCEP analytical performance goals.

	Precision	*Bias*	*Total error*
Total cholesterol:	3% CV	±3%	±8.9%
HDL cholesterol:			
$\geq$42 mg dL^{-1}	4% CV	±5%	±12.8%
<42 mg dL^{-1}	SD < 1.7 mg dL^{-1}		
LDL cholesterol:	4% CV	±4%	±11.8%
Triglycerides:	5% CV	±5%	±14.8%

Other types of anticoagulants also have varying osmotic effects. Fasting (9–12 h) specimens are preferred for lipoprotein analysis because of problems related to measuring LDL-C in the presence of high triglycerides and because of the changes that occur in the lipoprotein levels in the postprandial state.

There is also a large biological variability for lipids and lipoproteins. For total cholesterol, the coefficient of variation for biological variability is approximately 6% so that results from the same individual that are collected only a few days apart can vary by as much as 12%. For triglycerides, the coefficient of biological variability is as large as 20%. Some of the potential sources of biological variability include effects due to acute or subacute illnesses, changes in diet, exercise frequency, body weight, or smoking. It is for this reason that the NCEP recommends that any treatment decisions be based on the average of two to three test results that should be taken at least 2 weeks apart to reduce the effect of any biologic and analytical variability.

4.2 Interpretation of Lipoprotein Assay Results

Adult reference range of lipids and lipoproteins are shown in Table 4. Unlike most other laboratory-test results, these reference ranges do not represent the central 95% interval. For example, approximately 50% of the adult US population has a total cholesterol value greater than the upper limit of 200 mg dL^{-1}. The reference values for lipids and lipoproteins instead represent the ideal test range chosen by expert consensus for reducing the risk of developing cardiovascular disease.

A flowchart of how lipoprotein analysis is used in implementation of the NCEP-III guidelines (http://www.nhlbi.nih.gov/about/ncep) for the diagnosis and treatment of cardiovascular disease is shown in Fig. 7. These guidelines represent the consensus of an expert panel on the scientific literature supporting the treatment of dyslipidemia for the prevention of CHD. It is recommended that a complete lipid

Tab. 4 Adult reference ranges for lipids.

Total cholesterol	140–200 mg dL^{-1}
HDL-C	40–75 mg dL^{-1}
LDL-C	50–130 mg dL^{-1}
Triglyceride	60–150 mg dL^{-1}

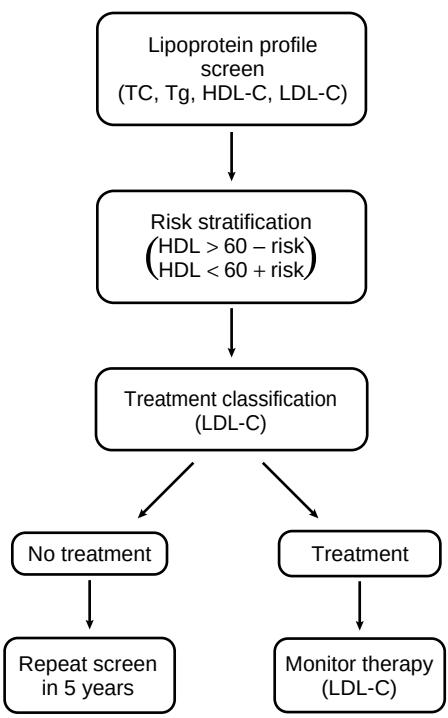

Fig. 7 Flow chart on the use of lipoprotein analysis in NCEP-III guidelines for the diagnosis and treatment of dyslipidemia.

panel test (total cholesterol, triglycerides, HDL-C, and a calculated LDL-C) be performed, as part of the initial screen, on a 9 to 12 h fasting sample. If the total triglycerides are more than 400 mg dL^{-1}, a β quantitation of LDL-C may be necessary or, possibly, a direct LDL-C test. For nonfasting samples, just a total cholesterol and HDL-C test should be performed. If the total cholesterol is less than 200 mg dL^{-1} and the HDL-C is greater than 60 mg dL^{-1} on a nonfasting sample and if the patient is at low risk for CHD, no further analysis is recommended. Each subject is then stratified into four levels of risk for CHD. This stratification is based on previous history of cardiovascular disease and/or its risk equivalent, such as diabetes, as well as other known risk factors for CHD, such as hypertension, smoking, and family history. HDL-C is also used in this risk stratification process. HDL-C greater than 60 mg dL^{-1} is used as a negative risk factor, whereas HDL-C less than 40 mg dL^{-1} is used as a positive risk factor. Each patient is then categorized into a treatment category on the basis of his or her LDL-C. Each level of risk has an ideal upper limit for LDL-C, and the LDL-C limit is set lower for those patients at higher risk. Initially, all patients that have an LDL-C that exceeds the recommended level are counseled to initiate lifestyle changes, such as reducing weight, increasing physical activity, and switching to a low-fat diet. If the lifestyle changes are ineffective in reaching the desirable range of LDL-C and/or if patients have an LDL-C that is beyond a recommended drug-treatment threshold for LDL-C, concurrent treatment with drugs to lower cholesterol is also recommended. The effect of any therapy is monitored, thereafter, by measuring LDL-C. HDL-C and triglycerides may also be important to be monitored, depending on their pretreatment levels and the type of drug therapy used. Other newly developed lipoproteins tests, such as Lp(a), remnant lipoproteins, small dense LDL, and tests for other markers of coronary artery disease, such as CRP and homocysteine, are currently not used in the initial risk stratification of a patient but these can be considered when making treatment decisions. For patients at the lowest risk for CHD, with an LDL-C below 160 mg dL^{-1}, no treatment is recommended and a repeat screen should be performed within 5 years.

5 Perspective

Nearly 50 years have passed since the first methods were developed for separating and analyzing lipoproteins by ultracentrifugation. During this time, great advances have been made in unraveling the various pathways of lipoprotein metabolism and the role of the individual lipoproteins. This has been coupled with the identification and elucidation of many of the genetic causes of dyslipidemias. Of paramount importance have been the epidemiological and animal studies that have revealed the causal association between lipoproteins and coronary heart disease. In many cases, advances in analytical techniques for lipoprotein analysis were pivotal to these discoveries. On the basis of the rapid advances to date in lipoprotein analysis, it is envisioned that there will be many future improvements in these tests, which will lead to even more detailed insights into lipoprotein metabolism and diagnostic tests with even greater clinical utility for preventing coronary heart disease.

See also FTIR of Biomolecules; Gene Therapy and Cardiovascular Diseases; Macromolecules, X-Ray Diffraction of Biological.

Bibliography

Books and Reviews

Albers, J.J., Segrest, J.P. (Eds.) (1986) *Plasma Lipoproteins, Part A: Plasma Lipoproteins: Preparation, Structure, and Molecular Biology. Methods in Enzymology*, Vol. 129. Academic Press, London and New York.

Converse, C.A., Skinner, E.R. (Eds.) (1992) *Lipoprotein Analysis: A practical Approach*. IRL Press, Oxford, New York, and Tokyo.

Li, W.H., Tanimura, M., Luo, C.C., Datta, S., Chan, L. (1988) The apolipoprotein multigene family: biosynthesis, structure, structure-function relationships, and evolution, *J. Lipid Res.* **29**, 245.

Libby, P., Aikawa, M., Schonbeck, U. (2000) Cholesterol and atherosclerosis, *Biochim. Biophys. Acta* **1529**, 299.

Nwokoro, N.A., Wassif, C.A., Porter, F.D. (2001) Genetic disorders of cholesterol biosynthesis in mice and humans, *Mol. Genet. Metab.* **74**, 105.

Pauciullo, P. (2002) Lipoprotein transport and metabolism: a brief update, *Nutr. Metab. Cardiovasc. Dis.* **12**, 90.

Ridker, P.M., Antman, E.M. (1999) Pathogenesis and pathology of coronary heart disease syndromes, *J. Thromb. Thrombolysis* **8**, 167–189.

Rifai, N., Warnick, G.R., Dominiczak, M.H., (Eds.) (2000) *Handbook of Lipoprotein Testing*. AACC Press, Washington, DC.

Segrest, J.P., Albers, J.J. (Eds.) (1986) *Plasma Lipoproteins, Part: B Plasma Lipoproteins: Characterization, Cell Biology, and Metabolism*. Methods in Enzymology, Vol. 128. Academic Press, London and New York.

Primary Literature

Abell, L.L., Levy, B.B., Brody, B.B., Kendall, F.C. (1952) A simplified method for the estimation of total cholesterol in serum and demonstration of its specificity, *J. Biol. Chem.* **195**, 357–66.

Allain, C.C., Poon, L.S., Chan, C.S., et al. (1974) Enzymatic determination of total serum cholesterol, *Clin. Chem.* **20**, 470–5.

Arrese, M.A., Crawford, J.M. (1997) Of plaques and stones: the SR-B1 (scavenger receptor class B, type 1), *Hepatology* **26**, 1072–4.

Bachorik, P.S., Ross, J.W., for the National Cholesterol Education Program Working Group on Lipoprotein Measurement. (1995) National Cholesterol Education Program recommendations for measurement of low-density lipoprotein cholesterol: executive summary, *Clin. Chem.* **41**, 1414–20.

Barrans, A., Jaspard, B., Barbaras, R., et al. (1996) Pre-beta HDL: structure and metabolism, *Biochim. Biophys. Acta* **1300**, 73–85.

Bartlett, G.R. (1959) Phosphorus assay in column chromatography, *J. Biol. Chem.* **234**, 466–8.

Castelli, W.P., Garrison, R.J., Wilson, P.W.F., et al. (1986) Incidence of coronary heart disease and lipoprotein cholesterol levels: the Framingham Study, *JAMA* **256**, 2835–8.

Chapman, M.J., Goldstein, S., Lagrange, D., et al. (1981) A density gradient ultracentrifugal procedure for the isolation of the major lipoprotein classes from human serum, *J. Lipid. Res.* **22**, 339–58.

Cohen, A., Hertz, H.S., Mandel, J., et al. (1980) Total serum cholesterol by isotope dilution/mass spectrometry: a candidate definitive method, *Clin. Chem.* **26**, 854–60.

Cole, T.G. (1990) Glycerol blanking in triglyceride assays: is it necessary? *Clin. Chem.* **36**, 1267–8.

Cooper. G.R., Smith, S.J., Myers, G.L., et al. (1994) Estimating and minimizing effects of biologic sources of variation by relative range when measuring the mean of serum lipids and lipoproteins, *Clin. Chem.* **40**, 227–32.

Dobiasova, M., Frohlich, J.J. (1999) Advances in understanding of the role of lecithin cholesterol acyltransferase (LCAT) in cholesterol transport, *Clin. Chim. Acta* **286**, 257–71.

Eckfeldt, J.H., Copeland, K.R. (1993) Accuracy verification and identification of matrix effects. The College of American Pathologist's protocol, *Arch. Pathol. Lab. Med.* **117**, 381–6.

Erbagci, A.B., Tarakcioglu, M., Aksoy, M., et al. (2002) Diagnostic value of CRP and Lp(a) in

coronary heart disease, *Acta Cardiol.* **57**, 197–204.

Fredrickson, D.S., Levy, R.I., Lees, R.S. (1967) Fat transport in lipoproteins: an integrated approach to mechanisms and disorders, *N. Engl. J. Med.* **276**, 148–56.

Friedewald, W.T., Levy, R.I., Fredrickson, D.S. (1972) Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge, *Clin. Chem.* **18**, 499–502.

Frohlich, J., Lear, S.A. (2002) Old and new risk factors for atherosclerosis and development of treatment recommendations, *Clin. Exp. Pharmacol. Physiol.* **29**, 838–42.

Genest, J.J., Corbett, H, McNamara J.R, et al. (1988) Effect of hospitalization on high-density lipoprotein cholesterol in patients undergoing elective coronary angiography, *Am. J. Cardiol.* **61**, 998–1000.

Genest, J.J., McNamara, J.R., Salem, D.N., et al. (1991) Prevalence of risk factors in men with premature coronary artery disease, *Am. J. Cardiol.* **67**, 1185–9.

Harris, N., Galpchian, V., Rifai, N. (1996) Three routine methods for measuring high-density lipoprotein cholesterol compared with the reference method, *Clin. Chem.* **42**, 738–43.

Havel, R.J., Eder, H.A., Bragdon, J.H. (1955) The distribution and chemical composition of ultracentrifugally separated lipoproteins in human serum, *J. Clin. Invest.* **34**, 1345–53.

Hevonoja, T., Pentikainen, M.O., Hyvonen, M.T., et al. (2000) Structure of low density lipoprotein (LDL) particles: basis for understanding molecular changes in modified LDL, *Biochim. Biophys. Acta* **1488**, 189–210.

Klotzsch, S.G., McNamara, J.R. (1990) Triglyceride measurements: a review of methods and interferences, *Clin. Chem.* **36**, 1605–13.

Kromhout, D., Menotti, A., Bloemberg, B., et al. (1995) Dietary saturated and trans fatty acids and cholesterol and 25-year mortality from coronary heart disease: the Seven Countries Study, *Prev. Med.* **24**, 308–15.

Kulkarni, K.R., Garber, D.W., Schmidt, C.F., et al. (1992) Analysis of cholesterol in all lipoprotein classes by single vertical ultracentrifugation of fingerstick blood and controlled-dispersion flow analysis, *Clin. Chem.* **38**, 1898–905.

Lamarche, B., Lemieux, I., Despres, J.P. (1999) The small, dense LDL phenotype and the risk of coronary heart disease: epidemiology, patho-physiology and therapeutic aspects, *Diabetes Metab.* **25**, 199–211.

Li, Z., McNamara, J.R., Ordovas, J.M., et al. (1994) Analysis of high-density lipoproteins by a modified gradient gel electrophoresis method, *J. Lipid. Res.* **35**, 1698–711.

Lofland, Jr., H.B. (1964) A semiautomated procedure for the determination of triglycerides in serum, *Anal. Biochem.* **9**, 393–400.

Miremadi, S., Sniderman, A., Frohlich, J. (2002) Can measurement of serum apolipoprotein B replace the lipid profile monitoring of patients with lipoprotein disorders? *Clin. Chem.* **48**, 484–8.

National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). (2002) Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III) final report, *Circulation* **106**, 3143–421.

Nauck, M., Warnick, G.R., Rifai, N. (2002) Methods for measurement of LDL-cholesterol: a critical assessment of direct measurement by homogeneous assays versus calculation, *Clin. Chem.* **48**, 236–54.

Noble, R.P. (1968) Electrophoretic separation of plasma lipoproteins in agarose gel, *J. Lipid Res.* **9**, 693–700.

Ornish, D., Brown, S.E., Scherwitz, L.W., et al. (1990) Can lifestyle changes reverse coronary heart disease? The lifestyle heart trial, *Lancet* **336**, 129–33.

Rifai, N., Iannotti, E., DeAngelis, K., et al. (1998) Analytical and clinical performance of a homogeneous enzymatic LDL-cholesterol assay compared with the ultracentrifugation-dextran sulfate-Mg^{2+} method, *Clin. Chem.* **44**, 1242–50.

Schaefer, E.J., Lichtenstein, A.H., Lamon-Fava, S., et al. (1995) Lipoproteins, nutrition, aging, and atherosclerosis, *Am. J. Clin. Nutr.* **61**(Suppl. 3), 726S–740S.

Schaefer, E.J., McNamara, J.R., Genest, Jr., J., et al. (1988) Clinical significance of hypertriglyceridemia, *Semin. Thromb. Hemost.* **14**, 142–8.

Schmitz, G., Becker, A., Aslanidis, C. (1996) ACAT/CEH and ACEH/LAL: two key enzymes in hepatic cellular cholesterol homeostasis

and their involvement in genetic disorders, *Z. Gastroenterol.* **34**(Suppl. 3), 68–72.

Shepherd, J., Cobbe, S.M., Ford, I., et al. (1995) Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia: the West of Scotland Coronary Prevention Study, *N. Engl. J. Med.* **333**, 1301–7.

Silver, D.L., Jiang, X.C., Arai, T., et al. (2002) Receptors and lipid transfer proteins in HDL metabolism, *Ann. N. Y. Acad. Sci.* **902**, 103–11.

Steele, B.W., Koehler, D.F., Azar, M.M., et al. (1976) Enzymatic determinations of cholesterol in high-density-lipoprotein fractions prepared by a precipitation technique, *Clin. Chem.* **22**, 98–101.

Stein, E.A., Myers, G.L., for the National Cholesterol Education Program Working Group on Lipoprotein Measurement. (1995) National Cholesterol Education Program recommendations for triglyceride measurement: executive summary, *Clin. Chem.* **41**, 1421–6.

Sugiuchi, H., Irie, T., Uji, Y., et al. (1998) Homogeneous assay for measuring low-density lipoprotein cholesterol in serum with triblock copolymer and a-cyclodextrin sulfate, *Clin. Chem.* **44**, 522–31.

Sugiuchi, H., Uji, Y., Okabe, H., et al. (1995) Direct measurement of high-density lipoprotein cholesterol in serum with polyethylene glycol-modified enzymes and sulfated a-cyclodextrin, *Clin. Chem.* **41**, 717–23.

Sviridov, D., Nestel, P. (2002) Dynamics of reverse cholesterol transport: protection against atherosclerosis, *Atherosclerosis* **161**, 245–54.

Takayama, M., Itoh, S., Nagasaki, T., et al. (1977) A new enzymatic method for determination of serum choline-containing phospholipids, *Clin. Chim. Acta* **79**, 93–8.

Tall, A.R., Costet, P., Wang, N. (2002) Regulation and mechanisms of macrophage cholesterol efflux, *J. Clin. Invest.* **110**, 899–904.

Warnick, G.R., Wood, P.D.J.W., for the National Cholesterol Education Program Working Group on Lipoprotein Measurement. (1995) National Cholesterol Education Program recommendations for measurement of high-density lipoprotein cholesterol: executive summary, *Clin. Chem.* **41**, 1427–33.

Warnick, G.R., Knopp, R.H., Fitzpatrick, V., et al. (1990) Estimating low-density lipoprotein cholesterol by the Friedewald equation is adequate for classifying patients on the basis of nationally recommended cutpoints, *Clin. Chem.* **36**, 15–9.

Warnick, G.R., Nauck, M., Rifai, N. (2001) Evolution of methods for measurement of HDL-cholesterol: from ultracentrifugation to homogeneous assays, *Clin. Chem.* **47**, 1579–96.

Warnick, G.R., Nguyen, T., Bergelin, R.O., et al. (1982) Lipoprotein quantification: an electrophoretic method compared with the Lipid Research Clinics method, *Clin. Chem.* **28**, 2116–20.

Wilson, P.W., D'Agostino, R.B., Levy, D., et al. (1998) Prediction of coronary heart disease using risk factor categories, *Circulation* **97**, 1837–47.

Wu, L.L., Warnick, G.R., Wu, J.T., et al. (1989) A rapid micro-scale procedure for determination of the total lipid profile, *Clin. Chem.* **35**, 1486–91.

Yamagishi, M., Terashima, M., Awano, K., et al. (2000) Morphology of vulnerable coronary plaque: insights from follow-up of patients examined by intravascular ultrasound before an acute coronary syndrome, *J. Am. Coll. Cardiol.* **35**, 106–11.

Zilversmit, D.B. (1995) Atherogenic nature of triglycerides, postprandial lipemia, and triglyceride-rich remnant lipoproteins, *Clin. Chem.* **41**, 153–8.

Liposome Gene Transfection

Nancy Smyth Templeton
Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX, USA

1 **Introduction 299**

2 **Optimization of Cationic Liposome Formulations for Use *in vivo* 301**

3 **Liposome Morphology and Effects on Gene Delivery and Expression 302**

4 **Optimal Lipids and Liposome Morphology: Effects on Gene Delivery and Expression 305**

5 **Liposome Encapsulation, Flexibility, and Optimal Colloidal Suspensions 305**

6 **Overall Charge of Complexes and Entry into the Cell 307**

7 **Ligands Used for Targeted Delivery 308**

8 **Attachment of Ligands 309**

9 **Serum Stability of Optimized Nucleic Acid–Liposome Complexes for Use *in vivo* 310**

10 **Optimized Half-life in the Circulation 311**

11 **Broad Biodistribution of Optimized Liposome Formulations 312**

12 **Optimization of Targeted Delivery 312**

13 **Efficient Dissemination Throughout Target Tissues and Migration Across Tight Barriers 313**

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

14 Optimization of Plasmids for *in vivo* Gene Expression 314

15 Optimization of Plasmid DNA Preparations 316

16 Detection of Gene Expression 317

17 Optimization of Dose and Frequency of Administration 318

18 Summary 318

Acknowledgment 318

Bibliography 319
Books and Reviews 319
Primary Literature 319

Keywords

Colloidal Suspensions
The density of complexes in a specific volume.

Formulation
Procedures used to make liposomes and complexes.

Gene Therapy
Treatments that are mediated by either the addition of needed genes or the destruction of transcripts encoding unwanted gene products.

Lamellar Structures
Alternation of a lipid bilayer and a water layer.

Liposomes
Bilayers made from lipids.

Nonviral Delivery
Delivery vehicles that are not viruses or any other live organism.

Optimization
Creation of optimal conditions.

Serum Stability
Complexes that do not precipitate in specific percentage of serum.

Systemic Delivery
Delivery in the bloodstream.

Varied results have been obtained using cationic liposomes for *in vivo* delivery. Furthermore, optimization of cationic liposomal complexes for *in vivo* applications is complex involving many diverse components. These components include nucleic acid purification, plasmid design, formulation of the delivery vehicle, administration route and schedule, dosing, detection of gene expression, and others. This chapter will also focus on optimization of the delivery vehicle formulation. These formulation issues include morphology of the complexes, lipids used, flexibility versus rigidity, colloidal suspension, overall charge, serum stability, half-life in circulation, biodistribution, delivery to and dissemination throughout target tissues. Broad assumptions have been frequently made on the basis of the data obtained from focused studies using cationic liposomes. However, these assumptions do not necessarily apply to all delivery vehicles and, most likely, do not apply to many liposomal systems when considering these other key components that influence the results obtained *in vivo*. Optimizing all components of the delivery system is pivotal and will allow broad use of liposomal complexes to treat or cure human diseases or disorders. This chapter will highlight the features of liposomes that contribute to successful delivery, gene expression, and efficacy.

1 Introduction

Many investigators are focused on the production of effective nonviral gene therapeutics and on creating improved delivery systems that mix viral and nonviral vectors. Use of improved liposome formulations for delivery *in vivo* is valuable for gene therapy and would avoid several problems associated with viral delivery. Delivery of nucleic acids using liposomes is promising as a safe and nonimmunogenic approach to gene therapy. Furthermore, gene therapeutics composed of artificial reagents can be standardized and regulated as drugs rather than as biologics. Cationic lipids have been used for efficient delivery of nucleic acids to cells in tissue culture for several years. Much effort has also been directed toward developing cationic liposomes for efficient delivery of nucleic acids in animals and in humans. Most frequently, the formulations that are best to use for transfection of a broad range of cell types in culture are not optimal for achieving efficacy in small and large animal disease models.

Much effort has been devoted to the development of nonviral delivery vehicles due to the numerous disadvantages of viral vectors that have been used for gene therapy. Following viral delivery *in vivo*, immune responses are generated to expressed viral proteins that subsequently kill the target cells required to produce the therapeutic gene product. An innate humoral immune response can be

produced to certain viral vectors due to previous exposure to the naturally occurring virus. Random integration of some viral vectors into the host chromosome could occur and cause activation of proto-oncogenes, resulting in tumor formation. Clearance of viral vectors delivered systemically by complement activation can also occur. Viral vectors can be inactivated upon readministration by the humoral immune response. Potential for recombination of the viral vector with DNA sequences in the host chromosome that generates a replication-competent infectious virus also exists. Specific delivery of viral vectors to target cells can be difficult because two distinct steps in engineering viral envelopes or capsids must be achieved. First, the virus envelope or capsid must be changed to inactivate the natural tropism of the virus to enter specific cell types. Then, sequences must be introduced that allow the new viral vector to bind and internalize through a different cell surface receptor. Other disadvantages of viral vectors include the inability to administer certain viral vectors more than once, the high costs for producing large amounts of high-titer viral stocks for use in the clinic, and the limited size of the nucleic acid that can be packaged and used for viral gene therapy. Attempts are being made to overcome the immune responses produced by viral vectors after administration in immune competent animals and in humans, such as the use of gutted adenoviral vectors or encapsulation of viral vectors in liposomes. However, complete elimination of all immune responses to viral vectors may be impossible.

Use of liposomes for gene therapy provides several advantages. A major advantage is the lack of immunogenicity after *in vivo* administration, including systemic injections. Therefore, the nucleic acid-liposome complexes can be readministered without harm to the patient and without compromising the efficacy of the nonviral gene therapeutic. Improved formulations of nucleic acid–liposome complexes can also evade complement inactivation after *in vivo* administration. Nucleic acids of unlimited size can be delivered ranging from single nucleotides to large mammalian artificial chromosomes. Furthermore, different types of nucleic acids can be delivered including plasmid DNA, RNA, oligonucleotides, DNA–RNA chimeras, synthetic ribozymes, antisense molecules, RNAi, viral nucleic acids, and others. Certain cationic formulations can also encapsulate and deliver viruses, proteins or partial proteins with a low isoelectric point (pI), and mixtures of nucleic acids and proteins of any pI. Creation of nonviral vectors for targeted delivery to specific cell types, organs, or tissues is relatively simple. Targeted delivery involves elimination of nonspecific charge interactions with nontarget cells and addition of ligands for binding and internalization through target cell surface receptors. Other advantages of nonviral vectors include the low cost and relative ease in producing nucleic acid-liposome complexes in large scale for use in the clinic. In addition, greater safety for patients is provided using nonviral delivery vehicles due to few or no viral sequences present in the nucleic acids used for delivery, thereby precluding generation of an infectious virus. The disadvantage of nonviral delivery systems had been the low levels of delivery and gene expression produced by "first-generation" complexes. However, recent advances have dramatically improved transfection efficiencies and efficacy of liposomal vectors. Reviews

of other *in vivo* delivery systems and improvements using cationic liposomes have been published recently.

Cationic liposome–nucleic acid complexes can be administered via numerous delivery routes *in vivo*. Routes of delivery include direct injection (e.g. intratumoral), intravenous, intraperitoneal, intra-arterial, intrasplenic, mucosal (nasal, vaginal, rectal), intramuscular, subcutaneous, transdermal, intradermal, subretinal, intratracheal, intracranial, and others. Much interest has focused on noninvasive intravenous administration because many investigators believe that this route of delivery is the ''holy grail'' for the treatment or cure of cancer, cardiovascular, and other inherited or acquired diseases. Particularly for the treatment of metastatic cancer, therapeutics must reach not only the primary tumor but also the distant metastases.

Optimization of cationic liposomal complexes for *in vivo* applications and therapeutics is complex, involving many distinct components. These components include nucleic acid purification, plasmid design, formulation of the delivery vehicle, administration route and schedule, dosing, detection of gene expression, and others. Often I make the analogy of liposome optimization to a functional car. Of course, the engine of the car, analogous to the liposome delivery vehicle, is extremely important. However, if the car does not have wheels, adequate tyres, and so on, the motorist will not be able to drive the vehicle to its destination. This chapter will focus on optimization of these distinct components for use in a variety of *in vivo* applications. Optimizing all components of the delivery system will allow broad use of liposomal complexes to treat or cure human diseases or disorders.

2 Optimization of Cationic Liposome Formulations for Use *in vivo*

Much research has been directed toward the synthesis of new cationic lipids. Some new formulations led to the discovery of more efficient transfection agents for cells in culture. However, their efficiency measured *in vitro* did not correlate with their ability to deliver DNA after administration in animals. Functional properties defined *in vitro* do not assess the stability of the complexes in plasma or their pharmacokinetics and biodistribution, all of which are essential for optimal activity *in vivo*. Colloidal properties of the complexes, in addition to the physicochemical properties of their component lipids, also determine these parameters. In particular, in addition to efficient transfection of target cells, nucleic acid–liposome complexes must be able to traverse tight barriers *in vivo* and penetrate throughout the target tissue to produce efficacy for the treatment of disease. These are not issues for achieving efficient transfection of cells in culture with the exception of polarized tissue culture cells. Therefore, we are not surprised that optimized liposomal delivery vehicles for use *in vivo* may be different than those used for efficient delivery to cells in culture.

In summary, *in vivo* nucleic acid–liposome complexes that produce efficacy in animal models of disease have extended half-life in the circulation, are stable in serum, have broad biodistribution, efficiently encapsulate various sizes of nucleic acids, are targetable to specific organs and cell types, penetrate across tight barriers in several organs, penetrate evenly throughout the target tissue, are optimized for nucleic acid:lipid ratio and colloidal suspension *in vivo*, can be size fractionated

to produce a totally homogenous population of complexes prior to injection, and can be repeatedly administered. Recently, we demonstrated efficacy of a robust liposomal delivery system in small and large animal models for lung, breast, head and neck, and pancreatic cancers, and for Hepatitis B and C (Clawson and Templeton, unpublished data). On the basis of efficacy in these animal studies, this liposomal delivery system will be used in upcoming clinical trials to treat these cancers. Our studies have demonstrated broad efficacy in the use of liposomes to treat disease and have dispelled several myths that exist concerning the use of liposomal systems.

3 Liposome Morphology and Effects on Gene Delivery and Expression

Efficient *in vivo* nucleic acid–liposome complexes have unique features including their morphology, mechanisms for crossing the cell membrane and entry into the nucleus, ability to be targeted for delivery to specific cell surface receptors, and ability to penetrate across tight barriers and throughout target tissues. Liposomes have different morphologies based upon their composition and the formulation method. Furthermore, the morphology of complexes can contribute to their ability to deliver nucleic acids *in vivo*. Formulations frequently used for the delivery of nucleic acids are lamellar structures including small unilamellar vesicles (SUVs), multilamellar vesicles (MLVs), or bilamellar invaginated vesicles (BIVs) recently developed in our laboratory (Fig. 1). Several investigators have developed liposomal delivery systems using hexagonal structures; however, they have demonstrated efficiency primarily for the transfection of some cell types in culture and not for *in vivo* delivery. SUVs condense nucleic acids on the surface and form "spaghetti and meatballs" structures. DNA–liposome complexes made using SUVs produce little or no gene expression upon systemic delivery, although these complexes transfect

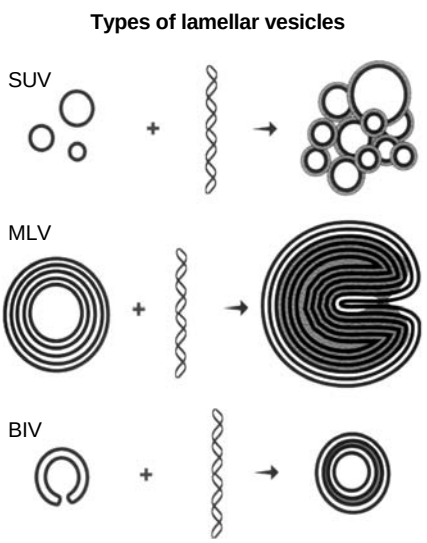

Fig. 1 Diagrams drawn from cryoelectron micrographs of cross sections through vitrified films of various types of liposomes and DNA–liposome complexes. SUVs are small unilamellar vesicles that condense nucleic acids on the surface and produce "spaghetti and meatballs" structures. MLVs are multilamellar vesciles that appear as "Swiss Rolls" after mixing with DNA. BIVs are bilamellar invaginated vesicles produced using a formulation developed in our laboratory. Nucleic acids are efficiently encapsulated between two bilamellar invaginated structures (BIVs).

numerous cell types efficiently *in vitro*. Furthermore, SUV liposome–DNA complexes cannot be targeted efficiently. SUV liposome–DNA complexes also have a short half-life within the circulation, generally about 5 to 10 min. Polyethylene glycol (PEG) has been added to liposome formulations to extend their half-life; however, PEGylation created other problems that have not been resolved. PEG seems to hinder delivery of cationic liposomes into cells due to its sterically hindering ionic interactions, and it interferes with optimal condensation of nucleic acids onto the cationic delivery vehicle. Furthermore, extremely long half-life in the circulation, for example, several days, has caused problems for patients because the bulk of the PEGylated liposomal formulation doxil that encapsulates the cytotoxic agent, doxorubicin, accumulates in the skin, hands, and feet. For example, patients contract mucositis and "Hand and Foot Syndrome" that cause extreme discomfort to the patient. Attempts to add ligands to doxil for delivery to specific cell surface receptors has not resulted in much cell-specific delivery, and the majority of the injected targeted formulation still accumulates in the skin, hands, and feet. Addition of PEG into formulations developed in our laboratory also caused steric hindrance in the bilamellar invaginated structures that did not encapsulate DNA efficiently, and gene expression was substantially diminished.

Some investigators have loaded nucleic acids within SUVs using a variety of methods; however, the bulk of the DNA does not load or stay within the liposomes. Furthermore, most of the processes used for loading nucleic acids within liposomes are extremely time-consuming and not cost effective. Therefore, SUVs are not the ideal liposomes for creating nonviral vehicles for targeted delivery.

Complexes made using MLVs appear as "Swiss rolls" when viewing cross-sections by cryoelectron microscopy. These complexes can become too large for systemic administration or deliver nucleic acids inefficiently into cells due to the inability to "unravel" at the cell surface. Addition of ligands onto MLV liposome–DNA complexes further aggravates these problems. Therefore, MLVs are not useful for the development of targeted delivery of nucleic acids.

Using a formulation developed in our laboratory, nucleic acids are efficiently encapsulated between two bilamellar invaginated vesicles, BIVs. We created these unique structures using 1,2-bis(oleoyloxy)-3-(trimethylammino)propane (DOTAP) and cholesterol (Chol) and a novel formulation procedure. This procedure is different because it includes a brief, low frequency sonication followed by manual extrusion through filters of decreasing pore size. The 0.1 and 0.2 μm filters used are made of aluminum oxide and not polycarbonate that is typically used by other protocols. Aluminum oxide membranes contain more pores per surface area, evenly spaced and sized pores, and pores with straight channels. During the manual extrusion process, the liposomes are passed through each of the four different sized filters only once. This process produces 88% invaginated liposomes. Use of high-frequency sonication and/or mechanical extrusion produces only SUVs.

The BIVs produced condense, unusually large amounts of nucleic acids of any size Fig. 2 or viruses Fig. 3. Furthermore, addition of other DNA condensing agents including polymers is not necessary. For example, condensation of plasmid DNA onto polymers first before encapsulation in the BIVs did not increase condensation or subsequent gene

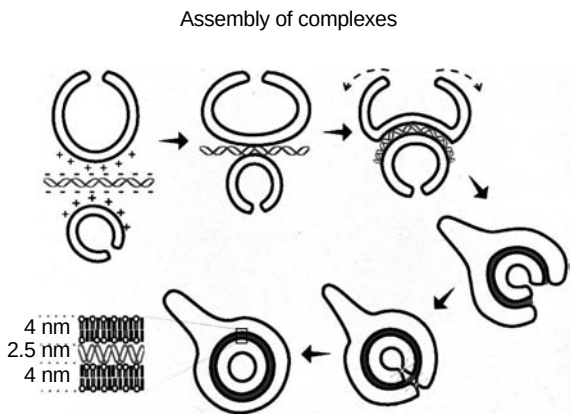

Fig. 2 Proposed model showing cross sections of extruded DOTAP:Chol liposomes (BIVs) interacting with nucleic acids. Nucleic acids adsorb onto a BIV via electrostatic interactions. Attraction of a second BIV to this complex results in further charge neutralization. Expanding electrostatic interactions with nucleic acids cause inversion of the larger BIV and total encapsulation of the nucleic acids. Inversion can occur in these liposomes because of their excess surface area, which allows them to accommodate the stress created by the nucleic acid–lipid interactions. Nucleic acid binding reduces the surface area of the outer leaflet of the bilayer and induces the negative curvature due to lipid ordering and reduction of charge repulsion between cationic lipid headgroups. Condensation of the internalized nucleic acid–lipid sandwich expands the space between the bilayers and may induce membrane fusion to generate the apparently closed structures. The enlarged area shows the arrangement of nucleic acids condensed between two 4 nm bilayers of extruded DOTAP:Chol.

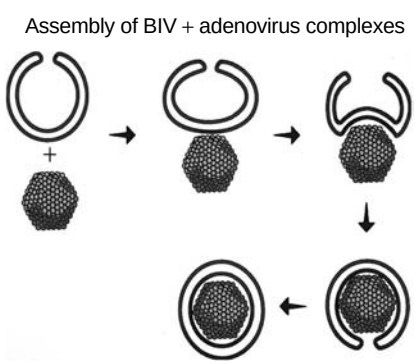

Fig. 3 Proposed model showing cross sections of an extruded DOTAP:Chol liposome (BIV) interacting with adenovirus. Adenovirus interacts with a BIV causing negative curvature and wrapping around the virus particle.

expression after transfection *in vitro* or *in vivo*. Encapsulation of nucleic acids by these BIVs alone is spontaneous and immediate, and, therefore, cost effective, requiring only one step of simple mixing. The extruded DOTAP:Chol–nucleic acid complexes are also large enough so that they are not cleared rapidly by Kupffer cells in the liver and yet extravasate across tight barriers, including the endothelial cell barrier of the lungs in a normal mouse, and diffuse through target organs efficiently. Our recent work demonstrating efficacy for treatment of nonsmall cell lung cancer showed that only BIV DOTAP:Chol-p53 DNA–liposome complexes produced efficacy, and SUV DOTAP:Chol-p53 DNA–liposome complexes produced no efficacy. Therefore, the choice of lipids alone is not sufficient for optimal DNA delivery, and the morphology of the complexes is essential.

4
Optimal Lipids and Liposome Morphology: Effects on Gene Delivery and Expression

Choosing the best cationic lipids and neutral lipids is also essential for producing the optimal *in vivo* formulation. For example, using our novel manual extrusion procedure does not produce BIVs using the cationic lipid dimethyldioctadecylammonium bromide (DDAB). Furthermore, DOTAP is biodegradable, whereas DDAB is not biodegradable. Use of biodegradable lipids is preferred for use in humans. Furthermore, only DOTAP- and not DDAB-containing liposomes produced highly efficient gene expression *in vivo*. DDAB did not produce BIVs and was unable to encapsulate nucleic acids. Apparently, DDAB- and DOTAP-containing SUVs produce similar efficiency of gene delivery *in vivo*; however, these SUVs are not as efficient as BIV DOTAP:Chol. In addition, use of L-α dioleoyl phosphatidylethanolamine (DOPE) as a neutral lipid creates liposomes that cannot wrap or encapsulate nucleic acids. Several investigators have reported efficient transfection of cells in culture using DOPE in liposomal formulations. However, our data showed that formulations consisting of DOPE were not efficient for producing gene expression *in vivo*.

Investigators must also consider the source and lot of certain lipids purchased from companies. For example, different lots of cholesterol from the same vendor can vary dramatically and will affect the formulation of liposomes. Currently, we are using synthetic cholesterol (Sigma, St. Louis, MO). The Food and Drug Administration prefers synthetic cholesterol instead of natural cholesterol that is purified from the wool of sheep for use in producing therapeutics for injection into humans.

Our BIV formulations are also stable for a few years as liquid suspensions. Freeze-dried formulations can also be made that are stable indefinitely even at room temperature. Stability of liposomes and liposomal complexes is also essential, particularly for the commercial development of human therapeutics.

5
Liposome Encapsulation, Flexibility, and Optimal Colloidal Suspensions

A common belief is that artificial vehicles must be 100 nm or smaller to be effective for systemic delivery. However, this belief is most likely true only for large, inflexible delivery vehicles. Blood cells are several microns (up to 7000 nm) in size, and yet have no difficulty circulating in the blood, including through the smallest capillaries. However, sickle cell blood cells, that are rigid, do have problems in the circulation. Therefore, we believe that flexibility is a more important issue than small size. In fact, BIV DNA–liposome complexes in the size range of 200 to 450 nm produced the highest levels of gene expression in all tissues after intravenous injection. Kupffer cells in the liver quickly clear delivery vehicles (including nonviral vectors and viruses) that are not PEGylated and are smaller than 200 nm. Therefore, increased size of liposomal complexes could extend their circulation time, particularly when combined with injection of high colloidal suspensions. BIVs were able to encapsulate nucleic acids and viruses apparently due to the presence of cholesterol in the bilayer (Fig. 4). Whereas, formulations including DOPE instead of cholesterol could not assemble nucleic acids by a "wrapping

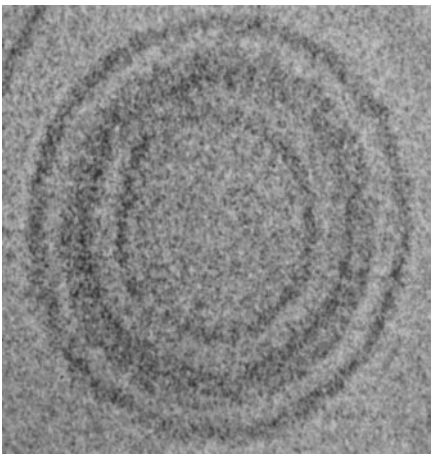

Fig. 4 Cryoelectron micrograph of BIV DOTAP:Chol–DNA–liposome complexes. The plasmid DNA is encapsulated between two BIVs.

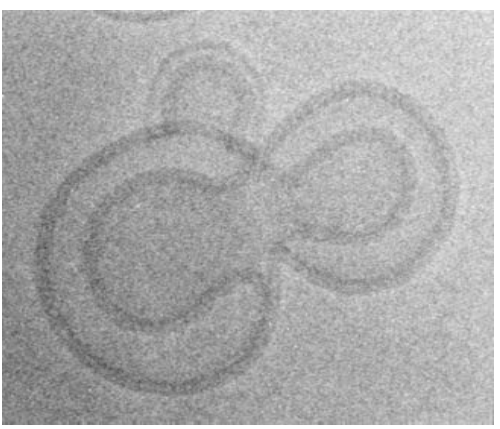

Fig. 5 Cryoelectron micrograph of extruded DOTAP:DOPE liposomes complexed to plasmid DNA. Although these liposomes were prepared by the same protocol that produces BIV DOTAP:Chol, these vesicles cannot wrap and encapsulate nucleic acids. The DNA condenses on the surfaces of the liposomes shown.

type'' of mechanism (Fig. 5), and produced little gene expression in the lungs and no expression in other tissues after intravenous injections. Because the extruded DOTAP:Chol BIV complexes are flexible and not rigid, are stable in high concentrations of serum, and have extended half-life, they do not have difficulty circulating efficiently in the bloodstream.

We believe that colloidal properties of nucleic acid–liposome complexes also determine the levels of gene expression produced after *in vivo* delivery. These properties include the DNA:lipid ratio that determines the overall charge density of the complexes and the colloidal suspension that is monitored by its turbidity. Complex size and shape, lipid composition and formulation, and encapsulation efficiency of nucleic acids by the liposomes also contribute to the colloidal properties of the complexes. The colloidal properties affect serum stability, protection from nuclease degradation, blood circulation time, and biodistribution of the complexes.

Our *in vivo* transfection data showed that an adequate amount of colloids in suspension was required to produce efficient gene

expression in all tissues examined. The colloidal suspension is assessed by measurement of adsorbance at 400 nm using a spectrophotometer optimized to measure turbidity. Our data showed that transfection efficiency in all tissues corresponded to OD400 of the complexes measured prior to intravenous injection.

6 Overall Charge of Complexes and Entry into the Cell

In addition, our delivery system is efficient because the complexes deliver DNA into cells by fusion with the cell membrane and avoid the endocytic pathway (Fig. 6). Cells are negatively charged on the surface, and specific cell types vary in their density of negative charge. These differences in charge density can influence the ability of cells to be transfected. Cationic complexes have nonspecific ionic charge interactions with cell surfaces. Efficient transfection of cells by cationic complexes is, in part, contributed by adequate charge interactions. In addition, recent publications report that certain viruses have a partial positive charge around key subunits of viral proteins on the virus surface responsible for binding to and internalization through target cell surface receptors. Therefore, this partial positive charge is required for virus entry into the cell. Thus, maintenance of adequate positive charge on the surface of targeted liposome complexes is essential for optimal delivery into the cell. Different formulations of liposomes interact with cell surfaces via a variety of mechanisms. Two major pathways for interaction are by endocytosis or by direct fusion with the cell membrane. Preliminary data suggest that nucleic acids delivered *in vitro* and *in vivo* using complexes developed in our lab enter the cell by direct fusion (Fig. 6). Apparently, the bulk of the nucleic acids do not enter endosomes, and, therefore, far more nucleic acid enters the nucleus. Cell transfection by direct fusion produces orders of magnitude increased levels of gene expression and numbers of cells transfected versus cells transfected through the endocytic pathway.

We believe that maintenance of adequate positive charge on the surface of complexes is essential to drive cell entry by direct fusion. Therefore, we create targeted delivery of our complexes *in vivo* without

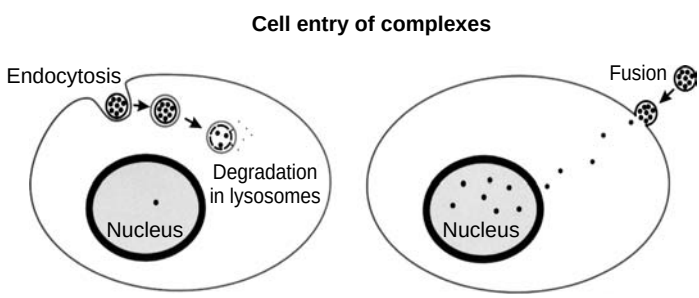

Fig. 6 Mechanisms for cell entry of nucleic acid–liposome complexes. Two major pathways for interaction are by endocytosis or by direct fusion with the cell membrane. Complexes that enter the cell by direct fusion allow delivery of more nucleic acids to the nucleus because the bulk of the nucleic acids do not enter endosomes.

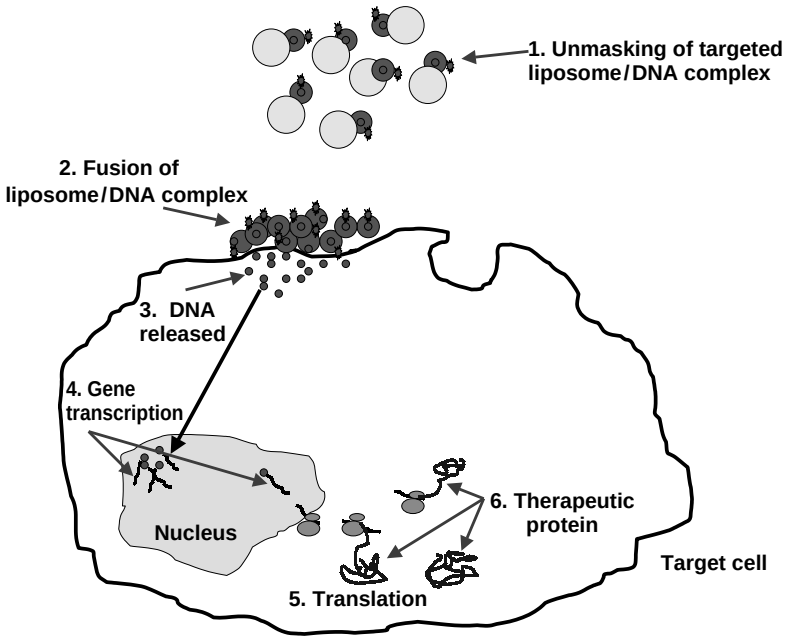

Fig. 7 Optimized strategy for delivery and gene expression in the target cell. Optimization of many steps is required to achieve targeted delivery, deshielding, fusion with the cell membrane, entry of nucleic acids into the cell and to the nucleus, and production of gene expression of a cDNA cloned in a plasmid.

the use of PEG. These ligand-coated complexes also reexpose the overall positive charge of the complexes as they approach the target cells. Through ionic interactions or covalent attachments, we have added monoclonal antibodies, Fab fragments, proteins, partial proteins, peptides, peptide mimetics, small molecules, and drugs to the surface of our complexes after mixing. These ligands efficiently bind to the target cell surface receptor, and maintain entry into the cell by direct fusion. Using novel methods for addition of ligands to the complexes for targeted delivery results in further increased gene expression in the target cells after transfection. Therefore, we design targeted liposomal delivery systems that retain predominant entry into cells by direct fusion versus the endocytic pathway. Figure 7 shows our optimized strategy to achieve targeted delivery, deshielding, fusion with the cell membrane, entry of nucleic acids into the cell and to the nucleus, and production of gene expression of a cDNA cloned in a plasmid.

7 Ligands Used for Targeted Delivery

Using liposomes that encapsulate nucleic acids, ligands can be coated onto the surface of the complexes formed (Fig. 8). We have added monoclonal antibodies, Fab fragments, proteins, partial proteins, peptides, peptide mimetics, small molecules, and drugs to the surface of the complexes after mixing. Ligands are chosen by their ability to efficiently bind to a

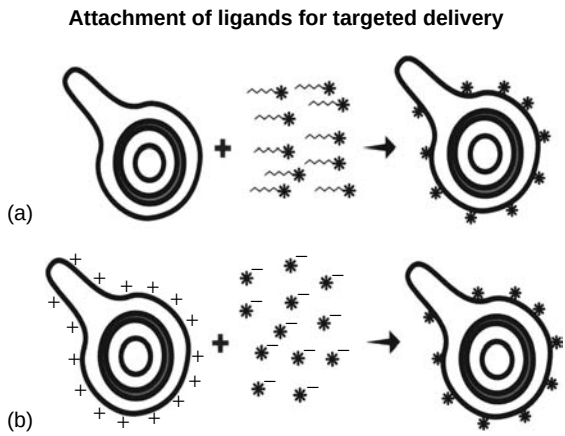

Fig. 8 Cross sections of extruded DOTAP:Chol-nucleic acid–liposome complexes (left) that are coated with ligands (right). Ligands are attached to the surface of preformed nucleic acid–liposome complexes by covalent attachments through "linker lipids" (a) or by ionic interactions (b).

target cell surface receptor while maintaining entry into the cell by direct fusion. Entry into the cell will be discussed further below. The ligands most useful for gene therapeutics in humans will be those that are smallest and possess high affinity for the target receptor. Nonviral systems are desirable because they can be repeatedly administered. Therefore, immune responses may be generated in animals or people upon repeated administration of complexes containing too much ligand or too large a ligand on the surface. These immune responses could cause the targeted therapeutics to be unsafe and/or ineffective for treatments in the clinic.

However, often the "best ligand" is not always available immediately. An investigator could also wait for years for the most appropriate ligand to be generated or produced in the amounts required for large experiments. Our experience shows that much useful information can be generated concerning targeting of a particular cell surface receptor of interest using the "less-than-ideal ligand" *in vitro* and *in vivo* in pilot experiments while creating the best ligand concurrently.

8 Attachment of Ligands

Generally, investigators attach ligands to PEG for incorporation into liposomes and other conjugates or for coating onto the surface of complexes after mixing. After extensive work with PEG in our lab, we have chosen or created alternative methods to use for the attachment of ligands, and have avoided the use of PEG due to its numerous disadvantages discussed above and below in this chapter.

Alternative strategies to the use of PEG include attachment of ligands through ionic interactions or by covalent attachment to "linker lipids". The ligands listed in the section above are generally negatively charged. Therefore, the ligands can simply be adsorbed onto the surface of complexes with encapsulated nucleic acids after mixing (Fig. 8b). Additional moieties can be added to the ligand to increase the amount of negative charge, and yet do not interfere with the ability of the modified ligand to efficiently bind to the appropriate cell surface receptor. For example, we used succinylated asialofetuin to target delivery of DNA–liposome complexes to the asialoglycoprotein receptor on

hepatocytes in the liver. The succinic acid amides provided greater negative charge to asialofetuin, and, therefore, bound to the surface of complexes more efficiently than asialofetuin alone. The amount used for adsorption onto the surface of complexes is ligand dependent. Titration studies must be performed to determine the optimal amount of ligand to coat onto the surface of complexes. Ultimately, *in vivo* transfection experiments must be performed to verify the optimal amount of ligand to use to provide delivery to the target cells and the highest levels of gene expression in these cells with no generation of an immune response.

Ligands or modified ligands containing reactive groups can be covalently attached to linker lipids (Fig. 8a). These ligand–lipid conjugates must be checked for optimal activity of the ligand to bind to its receptor. Furthermore, the covalent linkage must not be immunogenic in animals or people after repeated administration. Ligand–lipid conjugates can be spontaneously inserted into the outside membrane of complexes in which the nucleic acids are encapsulated within liposomes (Figure 8b). The amount of ligand–lipid used for insertion into the surface of complexes is also ligand dependent. Titration studies must be performed to determine the optimal amount of ligand–lipid to insert into the surface of complexes. Again, *in vivo* transfection experiments must be performed to verify the optimal amount of ligand–lipid to use to provide delivery to the target cells and the highest levels of gene expression in these cells with no generation of an immune response.

Using the alternative approaches described above, we have produced complexes that provide delivery of nucleic acids to target cells. Furthermore, the gene expression in the target cells using the targeted complexes is higher than that using the "generic" extruded DOTAP:Chol complexes.

9 Serum Stability of Optimized Nucleic Acid–Liposome Complexes for Use *in vivo*

Serum stability of cationic complexes is complicated and cannot be assessed by simply performing studies at a random concentration of serum. Figure 9 shows results from serum stability studies of DNA–liposome complexes that have been optimized in our laboratory for systemic delivery. Serum stability of these complexes was studied at 37 °C out to 24 h at concentrations of serum ranging from 0 to 100%. Two different serum stability assays were performed. The first assay measured the OD400 of BIV DOTAP:Chol–DNA–liposome complexes added into tubes containing a different concentration of serum in each tube, ranging from 0 to 100%. The tubes were incubated at 37 °C, and small aliquots from each tube were removed at various time points out to 24 h. The OD400 of each aliquot was measured on a spectrophotometer calibrated to accurately measure turbidity. Previous work in our laboratory demonstrated that the OD400 predicted both the stability of the complexes and the transfection efficiency results obtained for multiple organs after intravenous injections. Percent stability for this assay is defined as the transfection efficiency that is obtained at a particular OD400 of the complexes used for intravenous injections. Therefore, this assay is rigorous because slight declines in OD400 of these complexes result in obtaining

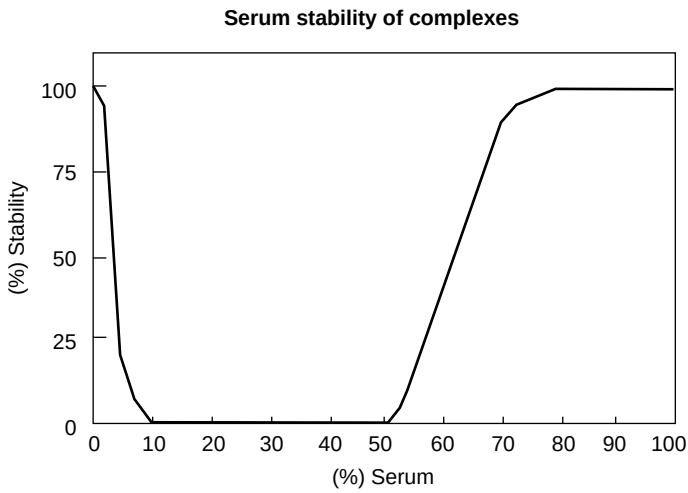

Fig. 9 Serum stability profile for DNA–liposome complexes optimized for systemic delivery. Serum stability of these complexes was studied at 37 °C out to 24 h at concentrations of serum ranging from 0 to 100%. Serum stability at the highest concentrations of serum, about 70 to 100%, that are physiological concentrations of serum found in the bloodstream is required.

no transfection *in vivo*. Declines in the OD400 also measure precipitation of the complexes.

A second assay was performed to support the results obtained from the OD400 measurements described above. A different concentration of serum, ranging from 0 to 100%, was placed into each well of a 96-well micro titer dish. BIV DOTAP:Chol–DNA–liposome complexes were added to the serum in the wells, and the plate was incubated at 37 °C. The plate was removed at various time points out to 24 h and complexes in the wells were observed under the microscope. Precipitation of complexes in the wells was assessed. 100% stability was set at no precipitation observed. Results from this assay were compared with those obtained in the first assay. 100% stability of complexes was set at no decline of OD400 in assay #1 and no observed precipitation in assay #2 at each % serum concentration, and the results were plotted (Fig. 9).

The results showed serum stability at the highest concentrations of serum, about 70 to 100%, that are physiological concentrations of serum found in the bloodstream. In addition, these complexes were also stable in no or low concentrations of serum. The complexes were unstable at 10 to 50% serum, perhaps due to salt bridging. Therefore, *in vitro* optimization of serum stability for formulations of cationic complexes must be performed over a broad range of serum concentration to be useful for applications *in vivo*.

10 Optimized Half-life in the Circulation

As stated above, the extruded BIV DOTAP:Chol–nucleic acid complexes are large enough so that they are not cleared

rapidly by Kupffer cells in the liver and yet extravasate across tight barriers and diffuse through the target organ efficiently. Further addition of ligands to the surface of extruded BIV DOTAP:Chol–nucleic acid complexes does not significantly increase the mean particle size. Extravasation and penetration through the target organ and gene expression produced after transfection are not diminished. These modified formulations are positively charged and deliver nucleic acids efficiently into cells *in vitro* and *in vivo*. Because extruded BIV DOTAP:Chol–nucleic acid complexes with or without ligands have a 5-h half-life in the circulation, these complexes do not accumulate in the skin, hands, or feet. Extended half-life in the circulation is provided primarily by the formulation, preparation method, injection of optimal colloidal suspensions, and optimal nucleic acid:lipid ratio used for mixing complexes, serum stability, and size (200 to 450 nm). Therefore, these BIVs are ideal for use in the development of effective, targeted nonviral delivery systems that clearly require encapsulation of nucleic acids.

11
Broad Biodistribution of Optimized Liposome Formulations

Our "generic" BIV nucleic acid–liposome formulation transfects many organs and tissues efficiently after intravenous injection, and has demonstrated efficacy in animal models for lung cancer, pancreatic cancer, breast cancer, Hepatitis B and C (Clawson and Templeton, unpublished data), and cardiovascular diseases. Therefore, optimization of the morphology of the complexes, the lipids used, flexibility of the liposomes and complexes, colloidal suspension, overall charge, serum stability, and half-life in circulation allows for efficient delivery and gene expression in many organs and tissues other than the lung. Apparently, these extruded DOTAP:Chol BIV nucleic acid–liposome complexes can overcome the tendency to be adsorbed only by the endothelial cells lining the circulation surrounding the lungs described by other investigators. However, as discussed above and below, we can further direct delivery to specific target tissues or cells by our targeted delivery strategies in combination with reversible masking used to bypass nonspecific transfection.

12
Optimization of Targeted Delivery

Much effort has been made to specifically deliver nucleic acid–liposome complexes to target organs, tissues, and/or cells. Ligands that bind to cell surface receptors are usually attached to PEG and then attached to the cationic or anionic delivery vehicle. Owing to the shielding of the positive charge of cationic complexes by PEG, delivery to the specific cell surface receptor can be accomplished by only a small fraction of complexes injected systemically. Furthermore, delivery of PEGylated complexes into the cell occurs predominantly through the endocytic pathway, and subsequent degradation of the bulk of the nucleic acid occurs in the lysosomes. Thus, gene expression is generally lower in the target cell than when using the nonspecific delivery of highly efficient cationic complexes.

As discussed above, the vast majority of the injected PEGylated complexes bypasses the target cell. Apparently, the PEGylated complexes cannot utilize critical charge interactions for optimal transfection into cells by direct fusion. Inability to expose positive charge on the surface of

optimized complexes results in the transfection of fewer cells. PEGylation was first used to increase the half-life of complexes in the circulation and to avoid uptake in the lung. However, this technology also destroys the ability to efficiently transfect cells. We were able to increase the half-life in circulation of BIVs to 5 h without the use of PEG. Because the extended half-life of BIVs is not too long, this delivery system does not result in the accumulation of complexes in nontarget tissues that circulate for one to three days. Some investigators have now reported targeted delivery that produces increased gene expression in the target cell over their nontargeted complexes. However, these nontargeted and targeted delivery systems are inefficient compared to efficient delivery systems such as the BIVs.

In using the extruded BIV DOTAP: Chol–nucleic acid:liposome complexes, we produced an optimal half-life in the circulation without the use of PEG. Extended half-life was produced primarily by the formulation, preparation method, injection of optimal colloidal suspensions, serum stability, and optimal nucleic acid:lipid ratio used for mixing complexes, and size (200 to 450 nm). Furthermore, we avoided uptake in the lungs using the negative charge of the ligands and "shielding/deshielding compounds" that can be added to the complexes used for targeting just prior to injection or administration *in vivo*. Our strategy to bypass nonspecific transfection is called *reversible masking*. Addition of ligands using the novel approaches that we developed, adequate overall positive charge on the surface of complexes is preserved. In summary, we achieve optimal circulation time of the complexes, reach and deliver to the target organ, avoid uptake in nontarget tissues, and efficiently interact with the cell surface to produce optimal transfection.

13
Efficient Dissemination Throughout Target Tissues and Migration Across Tight Barriers

A primary goal for efficient *in vivo* delivery is to achieve extravasation into and penetration throughout the target organ/tissue ideally by noninvasive systemic administration. Without these events, therapeutic efficacy is highly compromised for any treatment, including gene and drug therapies. Achieving this goal is difficult because of the many tight barriers that exist in animals and people. Furthermore, many of these barriers become tighter in the transition from neonates to becoming adults. Penetration throughout an entire tumor is further hindered because of the increased interstitial pressure within most tumors. We believe that nonviral systems can play a pivotal role in achieving target organ extravasation and penetration needed to treat or cure certain diseases. Our preliminary studies have shown that extruded BIV DOTAP:Chol–nucleic acid:liposome complexes can extravasate across tight barriers and penetrate evenly throughout entire target organs, and viral vectors cannot cross identical barriers. These barriers include the endothelial cell barrier in a normal mouse, the posterior blood retinal barrier in adult mouse eyes, complete and even diffusion throughout large tumors, and penetration through several tight layers of smooth muscle cells in the arteries of pigs. Diffusion throughout large tumors was measured by expression of ß-galactosidase or the proapoptotic gene *p*53 in about half of the p53-null tumor cells after a single injection of BIV DOTAP:Chol – DNA–liposome

complexes into the center of a tumor. Transfected cells were evenly spread throughout the tumors. Tumors injected with complexes encapsulating plasmid DNA encoding p53 showed apoptosis in almost all of the tumor cells by TUNEL staining. Tumor cells expressing p53 mediate a bystander effect on neighboring cells perhaps due to upregulation by Fas ligand that causes nontransfected tumor cells to undergo apoptosis. Currently, we are investigating the mechanisms used by extruded DOTAP:Chol–nucleic acid:liposome complexes to cross barriers and penetrate throughout target organs. By knowing more about these mechanisms, we hope to develop more robust nonviral gene therapeutics.

14
Optimization of Plasmids for *in vivo* Gene Expression

Delivery of DNA and subsequent gene expression may be poorly correlated. Investigators may focus solely on the delivery formulation as the source of poor gene expression. In many cases, however, the delivery of DNA into the nucleus of a particular cell type may be efficient, although little or no gene expression is achieved. The causes of poor gene expression can be numerous. The following issues should be considered independent of the delivery formulation, including suboptimal promoter-enhancers in the plasmid, poor preparation of plasmid DNA, and insensitive detection of gene expression.

Plasmid expression cassettes typically have not been optimized for animal studies. For example, many plasmids lack a full-length CMV promoter-enhancer. Over one hundred variations of the CMV promoter-enhancer exist, and some variations produce greatly reduced or no gene expression in certain cell types. Even commercially available plasmids contain suboptimal CMV promoters-enhancers, although these plasmids are advertised for use in animals. Furthermore, upon checking the company data for these plasmids, one would discover that these plasmids have never been tested in animals and have been tested in only one or two cultured cell lines. Conversely, plasmids that have been optimized for overall efficiency in animals may not be best for transfection of certain cell types *in vitro* or *in vivo*. For example, many investigators have shown that optimal CMV promoters-enhancers produce gene expression at levels several orders of magnitude less in certain cell types. In addition, one cannot assume that a CMV promoter that expresses well within the context of a viral vector, such as adenovirus, will function as well in a plasmid-based transfection system for the same cell context. Virus proteins produced by the viral vector are required for producing high levels of mRNA by the CMV promoter in specific cell nuclei.

Ideally, investigators design custom promoter-enhancer chimeras that produce the highest levels of gene expression in their target cells of interest. Recently, we designed a systematic approach for customizing plasmids used for breast cancer gene therapy using expression profiling. Gene therapy clinical trials for cancer frequently produce inconsistent results. We believe that some of this variability could result from differences in transcriptional regulation that limit expression of therapeutic genes in specific cancers. Our systemic liposomal delivery of a nonviral plasmid DNA showed efficacy in animal models for several cancers. However, we

observed large differences in the levels of gene expression from a CMV promoter-enhancer between lung and breast cancers. To optimize gene expression in breast cancer cells *in vitro* and *in vivo*, we created a new promoter-enhancer chimera to regulate gene expression. Serial analyses of gene expression (SAGE) data from a panel of breast carcinomas and normal breast cells predicted promoters that are highly active in breast cancers, for example, the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) promoter. Furthermore, GAPDH is upregulated by hypoxia, which is common in tumors. We added the GAPDH promoter, including the hypoxia enhancer sequences, to our *in vivo* gene expression plasmid. The novel CMV-GAPDH promoter-enhancer showed up to 70-fold increased gene expression in breast tumors compared to the optimized CMV promoter-enhancer alone. No significant increase in gene expression was observed in other tissues. These data demonstrate tissue-specific effects on gene expression after nonviral delivery, and suggest that gene delivery systems may require plasmid modifications for the treatment of different tumor types. Furthermore, expression profiling can facilitate the design of optimal expression plasmids for use in specific cancers.

Several reviews have stated that nonviral systems are intrinsically inefficient compared to viral systems. However, as discussed above, one must separate issues of the delivery vehicle versus the plasmid that is delivered. Case in point, we have shown that our extruded liposomes optimized for systemic delivery could outcompete delivery using a lentivirus. For example, we have compared SIVmac239, a highly noninfectious virus, with nonviral delivery of SIVmac239 DNA complexed to BIVs in adult rhesus macaques after injection into the saphenous vein of the leg. Our data showed that the monkeys injected with SIV DNA encapsulated in DOTAP:Chol BIVs were infected four days postinjection, and high levels of infection were produced in these monkeys at 14 days postinjection. Furthermore, higher levels of SIV RNA in the blood were produced using our BIV liposomes for delivery versus that using the SIV virus. CD4 counts were measured before and after injections. CD4 levels dropped in all monkeys to the lowest levels ever detected in the macaques in any experiment by 28 days postinjection, the first time point at which these counts were measured postinjection. All monkeys had clinical SIV infections and lost significant weight by day 28. These results were surprising because SIVmac239 is not highly infectious, and monkeys become sick with SIV infection only after several months or years postinjection with SIVmac239 virus. Therefore, we were able to induce SIV infection faster using our nonviral delivery of SIV plasmid DNA. In this case, we delivered a replication-competent plasmid so that gene expression increased over time posttransfection. Our delivery system was highly efficient and exceeded that of the lentivirus. The critical feature in this nonviral experiment was the plasmid DNA that was delivered.

Plasmids can be engineered to provide for specific or long-term gene expression, replication, or integration. Persistence elements, such as the inverted terminal repeats from adenovirus or adeno-associated virus, have been added to plasmids to prolong gene expression *in vitro* and *in vivo*. Apparently, these elements bind to the nuclear matrix thereby retaining the plasmid in cell nuclei. For regulated gene expression, many different inducible promoters are used that promote expression only in the presence of a positive regulator or in

the absence of a negative regulator. Tissue specific promoters have been used for the production of gene expression exclusively in the target cells. As discussed in the previous paragraph, replication-competent plasmids or plasmids containing sequences for autonomous replication can be included that provide prolonged gene expression. Other plasmid-based strategies produce site-specific integration or homologous recombination within the host cell genome. Integration of a cDNA into a specific "silent site" in the genome could provide long-term gene expression without disruption of normal cellular functions. Homologous recombination could correct genetic mutations upon integration of wild-type sequences that replace mutations in the genome. Plasmids that contain fewer bacterial sequences and that produce high yield upon growth in *Escherichia coli* are also desirable.

15
Optimization of Plasmid DNA Preparations

The transfection quality of plasmid DNA is dependent on the preparation protocol and training of the person preparing the DNA. For example, we performed a blinded study asking three people to make DNA preparations of the same plasmid from the same box of a Qiagen Endo-Free Plasmid Preparation kit. One person then mixed all of the DNA–liposome complexes on the same morning using a single vial of liposomes. One person performed all tail vein injections, harvesting of tissues, preparation of extracts from tissues, and reporter gene assays on the tissue extracts. *In vivo* gene expression differed 30-fold among these three plasmid DNA preparations.

One source for this variability is that optimized methods to detect and remove contaminants from plasmid DNA preparations have not been available. We have identified large amounts of contaminants that exist in laboratory and clinical grade preparations of plasmid DNA. These contaminants copurify with DNA by anion exchange chromatography and by cesium chloride density gradient centrifugation. Endotoxin removal does not remove these contaminants. HPLC cannot detect these contaminants. Therefore, we developed three proprietary methods for the detection of these contaminants in plasmid DNA preparations. We can now make clinical grade good manufacturing practices (GMP) DNA that does not contain these contaminants. To provide the greatest efficacy and levels of safety, these contaminants must be assessed and removed from plasmid DNA preparations. These contaminants belong to a class of molecules known to inhibit both DNA and RNA polymerase activities. Therefore, gene expression posttransfection can be increased by orders of magnitude if these contaminants are removed from DNA preparations. The presence of these contaminants in DNA also precludes high dose delivery of DNA–liposome complexes intravenously. Our group and other investigators have shown that intravenous injections of high doses of improved liposomes alone cause no adverse effects in small and large animals.

Some investigators have removed the majority of CpG sequences from their plasmids and report reduced toxicity after intravenous injections of cationic liposomes complexed to these plasmids. However, only low doses containing up to 16.5 μg of DNA per injection into each mouse were shown to reduce toxicity. Therefore, no significant dose response to CpG

motifs in plasmid DNA was demonstrated. To achieve efficacy for cancer metastases, particularly in mice bearing aggressive tumors, most investigators are interested in injecting higher doses in the range of 50 to 150 μg of DNA per mouse. Therefore, removal of CpG sequences from plasmid-based gene therapy vectors will not be useful for these applications because no difference in toxicity was shown after intravenous injections of these higher doses of plasmids, with or without reduced CpG sequences, complexed to liposomes. Therefore, we believe that removal of the other contaminants in current DNA preparations, discussed above, is the major block to the safe intravenous injection of high doses of DNA–liposome complexes.

16
Detection of Gene Expression

Thought should also be given to choosing the most sensitive detection method for every application of nonviral delivery rather than using the method that seems most simple. For example, detection of ß-galactosidase expression is far more sensitive than that for the green fluorescent protein (GFP). Specifically, 500 molecules of ß-galactosidase (ß-gal) per cell are required for detection using X-gal staining. Whereas, about one million molecules of GFP per cell are required for direct detection. Furthermore, detection of GFP may be impossible if the fluorescence background of the target cell or tissue is too high. Detection of chloramphenicol acetyltransferase (CAT) is extremely sensitive with little or no background detected in untransfected cells. Often, assays for CAT expression can provide more useful information than using ß-gal or GFP as reporter genes.

Few molecules of luciferase in a cell can be detected by luminescence assays of cell or tissue extracts posttransfection. The sensitivity of these assays is highly dependent on the type of instrument used to measure luminescence. However, luciferase results may not predict the therapeutic potential of a nonviral delivery system. For example, if several hundred or thousand molecules per cell of a therapeutic gene are required to produce efficacy for a certain disease, then production of only few molecules will not be adequate. If only few molecules of luciferase are produced in the target cell using a specific nonviral delivery system, then the investigator may be misled in using this system for therapeutic applications.

Furthermore, noninvasive detection of luciferase expression *in vivo* is not as sensitive as luminescence assays of cell or tissue extracts posttransfection. Recently, some colleagues of mine tried cooled charge coupled device (CCD) camera imaging on live mice after intravenous injection of other cationic liposomes complexed to plasmid DNA encoding luciferase, and they were not able to detect any transfection. However, these liposomal delivery systems had been used to detect luciferase by luminescence assays of organ extracts. My colleagues detected luciferase expression by CCD imaging after intravenous injections of BIV DOTAP:Chol-luciferase DNA–liposome complexes. Using the same CCD imaging system following intravenous injections of PEI-DNA complexes, extremely poor transfection efficiency was observed in all tissues. Because the luciferase protein is short-lived, maximal expression was detected at 5 h posttransfection. Whereas, detection of HSV-TK gene expression using microPET imaging in the same mice was highest at 24 h posttransfection. In contrast to

luciferase, the CAT protein accumulates over time, and, therefore, the investigator is not restricted to a narrow time frame for assaying gene expression. Furthermore, detection of CAT seems to be more sensitive than CCD imaging of luciferase following intravenous injections of DNA–liposome complexes. However, the animals must be sacrificed in order to perform CAT assays on tissue or organ extracts. In summary, further work is still needed to develop *in vivo* detection systems that have high sensitivity and low background.

17 Optimization of Dose and Frequency of Administration

To establish the maximal efficacy for the treatment of certain diseases or for the creation of robust vaccines, injections, or administrations of the nonviral gene therapeutic, and so on via different routes may be required. For particular treatments, one should not assume that one delivery route is superior to others without performing the appropriate animal experiments. In addition, people with the appropriate expertise should perform the injections and administrations. In our experience, only a minority of people who claim expertise in performing tail vein injections can actually perform optimal injections.

The optimal dose should be determined for each therapeutic gene or other nucleic acid that is administered. The investigator should not assume that the highest tolerable dose is optimal for producing maximal efficacy. The optimal administration schedule should also be determined for each therapeutic gene or other nucleic acid. To progress faster, some investigators have simply used the same administration schedule that they used for chemotherapeutics, for example. The investigator should perform *in vivo* experiments to determine when gene expression and/or efficacy drops significantly. Most likely, readministration of the nonviral gene therapeutic is not necessary until this drop occurs. Loss of the therapeutic gene product will vary with the half-life of the protein produced. Therefore, if a therapeutic protein has a longer half-life, then the gene therapy could be administered less frequently.

18 Summary

Overcoming some hurdles remains in the broad application of nonviral delivery; however, we are confident that we will successfully accomplish the remaining challenges soon. Furthermore, we predict that eventually the majority of gene therapies will utilize artificial reagents that can be standardized and regulated as drugs rather than biologics. We will continue to incorporate the molecular mechanisms of viral delivery that produce efficient delivery to cells into artificial systems. Therefore, the artificial systems, including liposomal delivery vehicles, will be further engineered to mimic the most beneficial parts of the viral delivery systems while circumventing their limitations. We will also maintain the numerous benefits of the liposomal delivery systems discussed in this chapter.

Acknowledgment

I thank Dr. David D. Roberts at the National Cancer Institute, National Institutes

of Health, Bethesda, MD for preparation of the figures.

See also Phospholipids; Plasmids.

Bibliography

Books and Reviews

Templeton, N.S. *Gene Targeting Protocols*, (2000) Vol. 133, Humana Press, Totowa, NJ, pp. 1–244.

Jain, R.K. (1994) Barriers to drug delivery in solid tumors, *Sci. Am.* **271**, 58–65.

Jain, R.K. (1999) Transport of molecules, particles, and cells in solid tumors, *Annu. Rev. Biomed. Eng.* **1**, 241–263.

Li, S., Ma, Z., Tan, Y., Liu, F., Dileo, J., Huang, L. (2002) Targeted Delivery Via Lipidic Vectors, in: Curiel, D.T., Douglas, J.T. (Eds.) *Vector Targeting for Therapeutic Gene Delivery*, Wiley-Liss Inc., Hoboken, NJ, pp. 17–32.

Marshall, E. (2003) Second child in French trial is found to have leukemia, *Science* **299**, 320.

Pirollo, K.F., Xu, L., Chang, E.H. (2002) Immunoliposomes: A Targeted Delivery Tool for Cancer Treatment, in: Curiel, D.T., Douglas, J.T. (Eds.) *Vector Targeting for Therapeutic Gene Delivery*, Wiley-Liss Inc., Hoboken, NJ, pp. 33–62.

Templeton, N.S., Lasic, D.D. (1999) New directions in liposome gene delivery, *Mol. Biotechnol.* **11**, 175–180.

Primary Literature

Aksentijevich, I., Pastan, I., Lunardi-Iskandar, Y., Gallo, R.C., Gottesman, M.M., Thierry, A.R. (1996) In vitro and in vivo liposome-mediated gene transfer leads to human MDR1 expression in mouse bone marrow progenitor cells, *Hum. Gene Ther.* **7**, 1111–1122.

Behr, J.-P., Demeneix, B., Loeffler, J.P., Perez-Mutul, J. (1989) Efficient gene transfer into mammalian primary endocrine cells with lipopolyamine-coated DNA, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6982–6986.

Felgner, P.L., Ringold, G.M. (1989) Cationic liposome-mediated transfection, *Nature* **337**, 387–388.

Felgner, J.H., Kumar, R., Sridhar, C.N., Wheeler, C.J., Tsai, Y.J., Border, R., Ramsey, P., Martin, M., Felgner, P.L. (1994) Enhanced gene delivery and mechanism studies with a novel series of cationic lipid formulations, *J. Biol. Chem.* **269**, 2550–2561.

Felgner, P.L., Gadek, T.R., Holm, M., Roman, R., Chan, H.W., Wenz, M., Northrop, J.P., Ringold, G.M., Danielson, H. (1987) Lipofection: a highly efficient lipid-mediated DNA transfection procedure, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7413–7417.

Gabizon, A., Catane, R., Uziely, B., Kaufman, B., Safra, T., Cohen, R., Martin, F., Huang, A., Barenholz, Y. (1994) Prolonged circulation time and enhanced accumulation in malignant exudates of doxorubicin encapsulated in polyethylene-glycol coated liposomes, *Cancer Res.* **54**, 987–992.

Gordon, K.B., Tajuddin, A., Guitart, J., Kuzel, T.M., Eramo, L.R., VonRoenn, J. (1995) Hand-foot syndrome associated with liposome-encapsulated doxorubicin therapy, *Cancer* **75**, 2169–2173.

Gustafsson, J., Arvidson, G., Karlsson, G., Almgren, M. (1995) Complexes between cationic liposomes and DNA visualized by cryo-TEM, *Biochim. Biophys. Acta* **1235**, 305–312.

Handumrongkul, C., Zhong, W., Debs, R.J. (2002) Distinct sets of cellular genes control the expression of transfected, nuclear-localized genes, *Mol. Ther.* **5**, 186–194.

Hildebrandt, I.J., Iyer, M., Wagner, E., Gambhir, S.S. (2003) Optical imaging of transferrin targeted PEI/DNA complexes in living subjects, *Gene Ther.* **10**, 758–764.

Hood, J.D., Bednarski, M., Frausto, R., Guccione, S., Reisfeld, R.A., Xiang, R., Cheresh, D.A. (2002) Tumor regression by targeted gene delivery to the neovasculature, *Science* **296**, 2404–2407.

Iyer, M., Berenji, M., Templeton, N.S., Gambhir, S.S. (2002) Noninvasive imaging of cationic lipid-mediated delivery of optical and PET reporter genes in living mice, *Mol. Ther.* **6**, 555–562.

Jain, R.K. (1991) Haemodynamic and transport barriers to the treatment of solid tumours, *Int. J. Radiat. Biol.* **60**, 85–100.

Leventis, R., Silvius, J.R. (1990) Interactions of mammalian cells with lipid dispersions containing novel metabolizable cationic amphiphiles, *Biochim. Biophys. Acta* **1023**, 124–132.

Liu, F., Qi, H., Huang, L., Liu, D. (1997) Factors controlling the efficiency of cationic lipid-mediated transfection in vivo via intravenous administration, *Gene Ther.* **4**, 517–523.

Liu, Y., Liggitt, D., Zhong, W., Tu, G., Gaensler, K., Debs, R. (1995) Cationic liposome-mediated intravenous gene delivery, *J. Biol. Chem.* **270**, 24864–24870.

Liu, Y., Mounkes, L.C., Liggitt, H.D., Brown, C.S., Solodin, I., Heath, T.D., Debs, R.J. (1997) Factors influencing the efficiency of cationic-liposome mediated intravenous gene delivery, *Nat. Biotechnol.* **15**, 167–173.

Loeffler, J.P., Behr, J.-P. (1993) Gene transfer into primary and established mammalian cell lines with lipopolyamine-coated DNA, *Methods Enzymol.* **217**, 599–618.

Lu, H., Zhang, Y., Roberts, D.D., Osborne, C.K., Templeton, N.S. (2002) Enhanced gene expression in breast cancer cells in vitro and tumors in vivo, *Mol. Ther.* **6**, 783–792.

Mislick, K.A., Baldeschwieler, J.D. (1996) Evidence for the role of proteoglycans in cation-mediated gene transfer, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12349–12354.

Papahadjopoulos, D., Allen, T.M., Gabizon, A., Mayhew, E., Matthay, K., Huang, S.K., Lee, K., Woodle, M.C., Lasic, D.D., Redemann, C., Martin, F.J. (1991) Sterically stabilized liposomes: improvements in pharmacokinetics and antitumor therapeutic efficacy, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 11460–11464.

Philip, R., Liggitt, D., Philip, M., Dazin, P., Debs, R. (1993) In vivo gene delivery: efficient transfection of T lymphocytes in adult mice, *J. Biol. Chem.* **268**, 16087–16090.

Pinnaduwage, P., Huang, L. (1989) The role of protein-linked oligosaccharide in the bilayer stabilization activity of glycophorin A for dioleoylphosphatidylethanolamine liposomes, *Biochim. Biophys. Acta* **986**, 106–114.

Ramesh, R., Saeki, T., Templeton, N.S., Ji, L., Stephens, L.C., Ito, I., Wilson, D.R., Wu, Z., Branch, C.D., Minna, J.D., Roth, J.A. (2001) Successful treatment of primary and disseminated human lung cancers by systemic delivery of tumor suppressor genes using an improved liposome vector, *Mol. Ther.* **3**, 337–350.

Rose, J.K., Buonocore, L., Whitt, M.A. (1991) A new cationic liposome reagent mediating nearly quantitative transfection of animal cells, *Biotechniques* **10**, 520–525.

Senior, J., Delgado, C., Fisher, D., Tilcock, C., Gregoriadis, G. (1991) Influence of surface hydrophilicity of liposomes on their interaction with plasma protein and clearance from the circulation: studies with poly(ethylene glycol)-coated vesicles, *Biochim. Biophys. Acta* **1062**, 77–82.

Shi, H.Y., Liang, R., Templeton, N.S., Zhang, M. (2002) Inhibition of breast tumor progression by systemic delivery of the maspin gene in a syngeneic tumor model, *Mol. Ther.* **5**, 755–761.

Solodin, I., Brown, C.S., Bruno, M.S., Ching-Yi, C., Eun-Hyun, J., Debs, R., Heath, T.D. (1995) A novel series of amphiphilic imidazolinium compounds for in vitro and in vivo gene delivery, *Biochemistry* **34**, 13537–13544.

Sternberg, B. (1996) Morphology of cationic liposome/DNA complexes in relation to their chemical composition, *J. Liposome Res.* **6**, 515–533.

Templeton, N.S., Lasic, D.D., Frederik, P.M., Strey, H.H., Roberts, D.D., Pavlakis, G.N. (1997) Improved DNA: liposome complexes for increased systemic delivery and gene expression, *Nat. Biotechnol.* **15**, 647–652.

Templeton, N.S., Alspaugh, E., Antelman, D., Barber, J., Csaky, K.G., Fang, B., Frederik, P., Honda, H., Johnson, D., Litvak, F., Machemer, T., Ramesh, R., Robbins, J., Roth, J.A., Sebastian, M., Tritz, R., Wen, S.F., Wu, Z. (1999) Non-viral Vectors for the Treatment of Disease, *Keystone Symposia on Molecular and Cellular Biology of Gene Therapy*, Salt Lake City, UT.

Thierry, A.R., Lunardi-Iskandar, Y., Bryant, J.L., Rabinovich, P., Gallo, R.C., Mahan, L.C. (1995) Systemic gene therapy: biodistribution and long-term expression of a transgene in mice, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9742–9746.

Tirone, T.A., Fagan, S.P., Templeton, N.S., Wang, X.P., Brunicardi, F.C. (2001) Insulinoma induced hypoglycemic death in mice is prevented with beta cell specific gene therapy, *Ann. Surg.* **233**, 603–611.

Tsukamoto, M., Ochiya, T., Yoshida, S., Sugimura, T., Terada, M. (1995) Gene transfer and expression in progeny after intravenous DNA

injection into pregnant mice, *Nat. Genet.* **9**, 243–248.

Uziely, B., Jeffers, S., Isacson, R., Kutsch, K., Wei-Tsao, D., Yehoshua, Z., Libson, E., Muggia, F.M., Gabizon, A. (1995) Liposomal doxorubicin: antitumor activity and unique toxicities during two complementary phase I studies, *J. Clin. Oncol.* **13**, 1777–1785.

Xu, Y., Szoka, F.C. (1996) Mechanism of DNA release from cationic liposome/DNA complexes used in cell transfection, *Biochemistry* **35**, 5616–5623.

Yew, N.S., Zhao, H., Przybylska, M., Wu, I.-H., Tousignant, J.D., Scheule, R.K., Cheng, S.H. (2002) CpG depleted plasmid DNA vectors with enhanced safety and long-term gene expression in vivo, *Mol. Ther.* **5**, 731–738.

Yotnda, P., Chen, D.-H., Chiu, W., Piedra, P.A., Davis, A., Templeton, N.S., Brenner, M.K. (2002) Bilamellar cationic liposomes protect adenovectors from preexisting humoral immune responses, *Mol. Ther.* **5**, 233–241.

Zhu, N., Liggitt, D., Liu, Y., Debs, R. (1993) Systemic gene expression after intravenous DNA delivery in adult mice, *Science* **261**, 209–211.

Liver Cancer, Molecular Biology of

Mehmet Ozturk and Rengul Cetin-Atalay
Bilkent University, Bilkent, Ankara, Turkey

1 Epidemiology of Liver Cancers and Hepatocellular Carcinoma 324

2 Molecular Biology of Hepatocellular Carcinoma 325
2.1 Hepatogenesis and Homeostasis in the Adult Liver 325
2.2 Hepatocyte Regeneration During Chronic Liver Disease 326
2.3 Genetic and Epigenetic Changes in Hepatocellular Carcinoma 327
2.3.1 Major Oncogenes 328
2.3.2 Major Tumor Suppressor Genes 329
2.4 Molecular Mechanisms of Hepatocellular Carcinogenesis 330

Acknowledgments 331

Bibliography 331
Books and Reviews 331
Primary Literature 332

Keywords

Aflatoxins
A family of natural toxins produced by *Aspergillus* species, inducing mutations in animals and involved in liver cancer development.

Axin1
The gene encoding a protein that facilitates the degradation of oncogenic β-catenin protein that is inactivated by mutation in liver cancers.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

β-catenin
A proto-oncogene when mutated contributes to the development of liver cancer.

Cyclin D1
A gene-encoding cyclin partner of cyclin-dependent kinase 4 that is amplified in liver cancers.

Hepatitis B Virus
A small DNA virus that infects hepatocytes in the liver, causing acute or chronic hepatitis.

Hepatitis C Virus
A small RNA virus infecting hepatocytes in the liver, causing acute or chronic hepatitis.

$P16^{INK4a}$
The gene encoding a protein that inhibits the activity of cyclin-dependent kinase 4 that is inactivated by *de novo* methylation in different cancers, including liver cancers.

p53
A tumor suppressor gene that is inactivated by mutation in different cancers including liver cancers.

Primary liver cancer is the fifth most common cancer worldwide. Hepatocellular carcinoma (HCC), which constitutes 80% of liver cancers, is linked etiologically to hepatitis B virus (HBV), hepatitis C virus (HCV), and aflatoxins. Hepatocellular carcinoma results from malignant transformation of hepatocytes and/or its progenitor cells. Viral factors appear to contribute to hepatocellular carcinogenesis nonspecifically by inducing chronic liver regeneration leading to cirrhosis, followed by HCC. Major cellular genes that are mutated in HCC are *β-catenin* and *Cyclin D1* oncogene, as well as *p53* and *axin1* tumor suppressor genes. *$P16^{INK4a}$*, another tumor suppressor gene is inactivated by promoter hypermethylation. These mutations affect p53, retinoblastoma, and wnt pathways in HCC, leading to a loss of proliferation control, escape from apoptosis and senescence, and probably, aberrant differentiation process.

1 Epidemiology of Liver Cancers and Hepatocellular Carcinoma

Primary liver cancer is a major public health problem, worldwide. It is the fifth most common cancer in the world, and third most common cause of cancer-related death. Hepatocellular carcinoma, cholangiocarcinoma, and hepatoblastoma are the main forms of cancer of the liver. However, the great majority (more than

80%) of primary liver tumors are HCCs. More than 500 000 new cases of HCC are diagnosed each year. Long considered as a disease of people living in Africa and Asia, HCC is now becoming a serious health threat in Europe and North America also because of its rising incidence.

Chronic infection with HBV, HCV, as well as naturally occurring aflatoxins were classified as carcinogenic to humans by International Agency for Cancer Research (IARC) working groups. These three etiologic factors constitute the majority of liver cancer risk factors in the world. Other risk factors for liver carcinomas are alcohol, nonviral cirrhosis, autoimmune chronic active hepatitis, and metabolic diseases, such as α-1-antitrypsin deficiency and hemochromatosis. Hepatitis B virus is the major etiology in Asia and Africa, whereas HCV appears to be more involved in HCCs occurring in Japan, Europe, and North America.

All major etiological factors of HCC have in common the ability to cause chronic liver disease leading to viral cirrhosis, which represents the single most common cause of these tumors with a contribution to over 80% of these cancers. The chronic process leading to cirrhosis with further development into HCC may take many years. Therefore, when analyzing the molecular etiology of HCC, it is often to dissociate viral from host-related contributions. Although involved as major causes, HBV and HCV differ from *bona fide* oncogenic viruses such as human papillomavirus 16 in their contributions to cancer. Their contribution to HCC appears to be rather nonspecific, mainly through the chronic liver regeneration process that they are known to be more directly involved in.

2 Molecular Biology of Hepatocellular Carcinoma

The liver is the largest internal organ that exhibits endocrine, exocrine, metabolic, and detoxification activities. Hepatocytes carry out the main functions of the liver and account for approximately 80% of all liver cells. Hepatocellular carcinoma is believed to originate from these mature cells, and/or from their progenitors. Therefore, a short overview of hepatogenesis and homeostasis of normal and diseased liver may be helpful for a better understanding of HCC biology.

2.1 Hepatogenesis and Homeostasis in the Adult Liver

Hepatogenesis is a process that consists of distinct developmental stages: competence, commitment, differentiation, and morphogenesis of hepatocytes. The liver develops from the ventral gut endoderm juxtaposing the developing heart. The entire embryonic endoderm is competent to adopt a hepatic fate and expresses the transcription factor HNF-3. Signals, including FGFs, released by cardiac mesoderm act on the underlying ventral endoderm of the foregut, inducing it to commit to a hepatic cell fate. In response to this, the prehepatic endoderm begins to express genes associated with a differentiated hepatocyte phenotype. At this stage of early development, these cells are called *hepatoblasts* or embryonic liver stem cells as they are at least bipotential, giving rise to both hepatocyte and bile epithelial cell lineages. The mechanisms controlling liver stem cell phenotype and the bipotential cell lineage commitment are mostly unknown. As far as the hepatocyte lineage is concerned,

expression of the full array of genes necessary for hepatocyte function requires the transcription factor HNF-4α. This differentiation step is followed by a period of rapid cell growth and morphogenesis that results in a functioning liver. Growth of the liver requires several transcription factors expressed in hepatocytes including NF-κB, c-Jun, and XBP-1.

The adult liver is estimated to be replaced by normal tissue renewal approximately once a year. This very slow process is maintained mainly by proliferation of differentiated hepatocytes. These hepatocytes are also able to restore the replacement of large volumes of liver tissue losses, following partial hepatectomy, for example, and during chronic liver diseases. However, under experimental conditions during which the proliferation of mature hepatocytes is suppressed, the liver mass can also be restored by adult liver cells, also called *oval* cells believed to reside in the liver. Furthermore, it has been shown that liver hepatocytes can also be derived from bone marrow cell populations in humans. Thus, the adult liver appears to be equipped with a powerful cell renewal system using mature hepatocytes, liver stem cells, or even bone marrow–derived stem cells.

2.2 Hepatocyte Regeneration During Chronic Liver Disease

The major etiological factors of HCC, namely, HBV and HCV may induce a chronic disease characterized by repeated inflammation, degeneration, and regeneration cycles in the liver. This chronic process is maintained by repetitive cycles of "exposure, inflammation, necrosis/apoptosis and regenerative response". After tens of years, the chronic hepatitis will lead to a cirrhosis stage, which is manifested by metabolic and replicative failure accompanied with the formation of fibrosis in the diseased liver. The HCC may arise at this stage from dysplastic nodules of hepatocytes without an overt stage of adenoma. The major limiting factors against the occurrence of malignant transformation are cell death and senescence. Therefore, HCC cells must acquire new abilities to escape from these limiting factors, namely, resistance to both apoptosis and senescence.

The process of inflammation in chronic liver disease creates a sustained state of oxidative stress. Reactive oxygen species can be generated during the inflammatory response to viral hepatitis or internally by the metabolic stress of hepatocytes. This oxidative stress is usually taken care of, at least initially, by antioxidant processes. Selenium is considered to play an indirect antioxidant role as an essential constituent of a selenoprotein, namely, glutathione peroxidase (Gpx), in elimination of cellular reactive oxygen species. Cytosolic or classical Gpx (Gpx-1) is considered as the most critical enzyme for the protective effects of selenium. It has recently been shown that selenium deficiency leads to apoptosis in hepatocyte-derived cells, due to oxidative stress. On the other hand, most HCC cells are tolerant to selenium deficiency and oxidative stress-inducing conditions. Thus, selenium deficiency in liver may confer growth advantage to malignant tumor cells, since they appear to adapt themselves to a selenium-deficient cellular environment.

Resistance to oxidative stress is probably one of the many ways in which malignant hepatocytes become resistant to apoptosis. Additional mechanisms involving the apoptotic machinery itself exist. Among

others, resistance to *Fas*-mediated apoptosis, changes in the expression of *bcl-2* family of genes, as well as *p53* gene mutations appear to confer apoptosis resistance to HCC cells.

2.3 Genetic and Epigenetic Changes in Hepatocellular Carcinoma

Hepatocellular cancers display many chromosomal changes (aneuploidies) such as polyploidy, loss of heterozygosity, allelic imbalance, amplifications as well as translocations. Aneuploidy is a general feature of solid tumors, so its occurrence in HCC is not surprising. In addition, it is well established that HBV DNA integrates into the host genome causing many chromosomal rearrangements. It is expected that the chromosomal regions that undergo tumor-specific changes harbor critical genes involved in carcinogenesis.

Historically, the oncogenes were first discovered as transforming viral genes. It was later shown that most viral oncogenes are activated forms or homologs of cellular proto-oncogenes involved in positive regulation of cell proliferation. Cellular proto-oncogenes can also be activated by somatic or even germ-line mutations, in the absence of viral participation. There are several cellular oncogenes (β-catenin, cyclins, etc.), which are considered as oncogenic in human hepatocellular carcinogenesis. There are also suspected viral oncogenes of HBV (mainly *X* gene) and hepatitis C virus (mainly core gene), which may contribute to hepatocellular cancer development. However, despite endless efforts, the true oncogenic role of these viral genes in human hepatocellular carcinogenesis is still a matter of debate (Table 1).

Tab. 1 The status of oncogenes and tumor suppressor genes in hepatocellular cancers.

Gene	*Mutation (%)*	*Other alterations*
p53	28	HBx interaction
β-catenin	13–89[c]	–
APC[a]	0–62[c]	LOH
Axin	5	LOH
E-cadherin	–	LOH, methylation
Cyclin D[b]	11–13	–
Cyclin A[b]	19	HBV integration
RB1	15	LOH, HD
p15^{INK4B}	0	HD
p16^{INK4A}	0–55[a]	HD, methylation
p14^{ARF}	0	–
p21$^{KIP1/CIP1}$	5	–
M6P/IGF2R	0–33	HD
TGFRB2	0–44	–
Smad2	0–2	–
Smad4	0–6	–
IGF-2	–	LOI
c-met	30[d]	–
PTEN	0–5	LOH, HBx interaction
K-ras	0–17	–
N-ras	0–16	Amplification
H-ras	0–10	Methylation
c-myc[b]	0–50	–
N-myc[b]	0	–
BRCA2[a]	5	LOH
hMLH1	–	LOH
hMSH2	–	LOH
DLC-1	–	HD, hyper methylation

[a]Somatic & germ-line mutations.
[b]Amplification.
[c]In HCC and hepatoblastoma.
[d]In childhood hepatocellular carcinoma only.
Notes: LOH: Loss of heterozygosity; HD: homozygous deletion; LOI: Loss of imprinting.
Source: This table was constructed by the modification of a table presented in Ozturk M., Cetin-Atalay R. (2003) Biology of Hepatocellular Cancer, in: Rustgi A. (Ed.) *Gastrointestinal Cancers: Biology, Diagnosis and Therapy*, 2nd edition, Lippincott-Raven Press, New York. Ozturk & Cetin-Atalay, 2003, to include recent data for additional alterations.

Tumor suppressor genes (or anti-oncogenes) are another group of cellular genes whose inactivation contributes to carcinogenesis. These genes, discovered much later than oncogenes, appear to be the most frequently mutated genes in different cancers, including hepatocellular cancers. To date, many tumor suppressor genes including *p53*, *mannose-6-phosphate/insulin-like factor 2 receptor* (*M6P/IGF2R*), *retinoblastoma* (*RB1*), *p16*^{INK4A}, *adenomatosis polyposis coli* (*APC*), *axin* (*axin1*), and *BRCA2* are known to be mutated in hepatocellular cancers (Table 1).

The list of hepatocellular cancer-associated oncogenes and tumor suppressor genes will probably grow over the coming years to include many more genes. There are at least two reasons to explain the high number of altered genes in HCC. Firstly, solid tumors in the adult may need the accumulation of many genetic alterations before they become clinically detectable. Indeed, the well-known "latent period" between the first exposure to an etiological agent (i.e. infection with HBV) and the development of HCC is in favor of such a hypothesis. Secondly, the multiplicity of genetic alterations in HCC may indicate that different etiological factors affect different sets of target genes in hepatocytes. This etiologically defined genetic heterogeneity of HCC results in a phenotypic heterogeneity of these tumors.

Here, we will review major oncogenes and tumor suppressor genes involved in HCC development.

2.3.1 Major Oncogenes

***β-catenin* (*CTNNB1*) gene** This gene encodes a pivotal component of the *Wnt* signaling pathway, which controls cell fate decisions during animal development. Inappropriate deregulation of this pathway leads to cancer in a number of tissues. The stability of β-catenin protein in the cytoplasm is low in unstimulated cells because of proteasome-mediated degradation by a protein complex, which consists of Axin, Adenomatosis polyposis coli (APC), and glycogen synthase kinase 3β (GSK3β). Wnt signaling inhibits the kinase activity of the quaternary complex. As a consequence, β-catenin accumulates in the cytoplasm, translocates to the nucleus, and becomes a transcriptional coactivator of T cell factor (TCF), the ultimate nuclear target of Wnt signaling. The *β-catenin* gene has recently been identified as one of the most frequently mutated genes in both hepatocellular carcinomas and hepatoblastomas. Both missense mutations and in-frame deletions mostly affect only exon 3 of *β-catenin* gene. These mutations affect, as a rule, the N-terminal region of β-catenin protein, which carry serine/threonine phosphorylation motifs involved in "marking" of the protein for ubiquitin-mediated degradation. Reported frequencies of β-catenin mutations in human HCC vary between 13 and 41%. Nuclear accumulation of β-catenin was reported to occur in 35% of HCCs. This could be due to mutations on other genes involved in β-catenin degradation. *Axin* gene mutations were found in 5% of HCCs. We have recently reported that, p53 mutations cause *in vitro* accumulation of β-catenin. It appears that β-catenin mutations in HCCs are as frequent as p53 mutation rates. Unlike p53, β-catenin mutations are also frequent (about 30%) in chemically induced rodent tumors. Childhood hepatoblastomas carry an even higher rate of β-catenin mutations. The reported frequencies vary between 48 and 89%, the overall rate being 58%.

Cyclin genes Cyclin genes comprise a gene family involved in the stepwise activation of cyclin-dependent kinases (CDK) during the cell cycle. Historically, the first cyclin gene known to be mutated in HCC is *CYCLIN A2* (*CCNA2*), which was identified in the integration site of HBV in a single tumor. Although this type of cyclin gene alteration has not been found in other tumors, this observation provided new clues about the implications of cyclin genes in HCC. Another cyclin known to be altered in HCC is the Cyclin D1 (*CCND1*; also called *PRAD1*) gene whose product forms active complexes with CDK4 enzyme during early G1 phase. The amplification of *Cyclin D* gene was observed in 10 to 13% of HCCs. *Cyclin D* gene amplification in HCC appears to be a late event, associated with more aggressive tumors.

2.3.2 Major Tumor Suppressor Genes

***P53* (*TP53*) gene** Many reports now indicate that the *p53* gene, which is located at chromosome 17p, is mutated in nearly one third of HCCs, worldwide. All of the reported p53 mutations in HCC are somatic. Therefore, germ-line mutations of p53 appear not to predispose to HCC. Both the frequency and the type of p53 mutations are different depending on geographical location and suspected etiology of these tumors. An HCC-specific codon 249 mutation (AGG->AGT leading to R249S), suspected to be induced by aflatoxins, was found in most HCCs from geographical areas with high incidence of HCC and with a high risk of exposure to aflatoxins. The codon 249 mutation is present in 36% of tumors from Africa and 32% of tumors from China. These two regions of the world are known for high incidence of HCC, where both HBV and aflatoxins are known to be the main etiological factors. In contrast, the codon 249 mutation is seen in less than 4% of HCCs from Japan, Europe, and North America, where HBV and HCV, but not aflatoxins are the main etiological factors. The overall frequency of codon 249 mutations in the world is 11%. Mutations affecting other codons of the *p53* gene are detected in HCC and their worldwide frequency is 18%. The frequency of all p53 mutations in HCC varies between 15% in Europe and 42% in China with a worldwide frequency of 27%. Thus, *p53* gene is mutated in about a third of HCCs, but only a third of these can be etiologically linked to a high risk of aflatoxin exposure. Therefore, p53 mutations can occur in HCC independent of aflatoxin risk, as well as HBV or HCV infection.

Cyclin-dependent kinase inhibitor genes Two closely located genes at chromosome 9p, namely, Cyclin-dependent Kinase (CDK) inhibitor 2a (*CDKN2A*, also called CDK4 inhibitor $p16^{INK4a}$ or MTS1) and cyclin-dependent kinase inhibitor 2b (*CDKN2B*, also called CDK4 inhibitor $p15^{INK4b}$ or *MTS2*) display genomic alterations in different human cancers. A number of studies demonstrated that the *CDKN2A* gene is strongly implicated in HCC. This gene codes for two alternatively spliced transcripts. One of the transcripts is for $p16^{CIP1KIP1}$ protein, an inhibitor of CDK4 and CDK6. The other transcript encodes for $p14^{ARF}$ protein, mouse homolog of which was shown to regulate wild-type p53 activity. More than 50% of HCCs display *de novo* methylation of *CDKN2A* gene, whereas its mutations are rare. It is known that *de novo* methylation is a mechanism involved in gene silencing. Therefore, HCC cells with methylated *CDKN2A*

gene are unable to express the gene, leading to the loss of a CDK inhibitor protein.

***Axin* (*AXIN1*) gene** *Axin* gene product is involved in the negative regulation of the *wnt* signaling. Like APC protein, the axin protein is required for active proteolytic degradation of β-catenin protein. *Axin* gene is mutated in about 5% of HCCs. The loss of the remaining *axin* gene allele was also demonstrated, suggesting that this gene acts as a tumor suppressor gene for HCC.

2.4 Molecular Mechanisms of Hepatocellular Carcinogenesis

The major genetic and epigenetic changes affecting mostly *p53*, *Retinoblastoma*, *p16*^{INK4a}, *β-catenin*, and *axin* genes allow us to link three major signaling pathways to malignant transformation of hepatocytes, namely, p53, retinoblastoma, and Wnt-β-catenin pathways. As shown in Fig. 1, these three pathways are involved in different processes that play key roles in malignant phenotype of not only HCC but also many other tumor types.

The p53 pathway, when it is active, controls three major cellular processes; namely, induction of apoptotic cell death, cell cycle arrest, and induction of senescence. Therefore, it is expected that p53 mutations that are observed at high frequency in HCC cells will lead to the inactivation of p53 pathway, together with a loss of p53-mediated apoptotic, antiproliferative, and antisenescence responses. Similarly, retinoblastoma pathway is involved in the control of cell cycle and permanent cell cycle arrest that is equivalent to senescence. The loss of p16^{INKa} expression in HCC may inactivate retinoblastoma pathway in these cells allowing them to escape from cell cycle control as well as from senescence.

The aberrant activation of Wnt-β-catenin pathway by mutations affecting β-catenin and Axin1 may confer new abilities to these cells such as autonomous proliferation ability, as well as cellular plasticity. Although hepatocellular effects of Wnt-β-catenin pathway activation are poorly

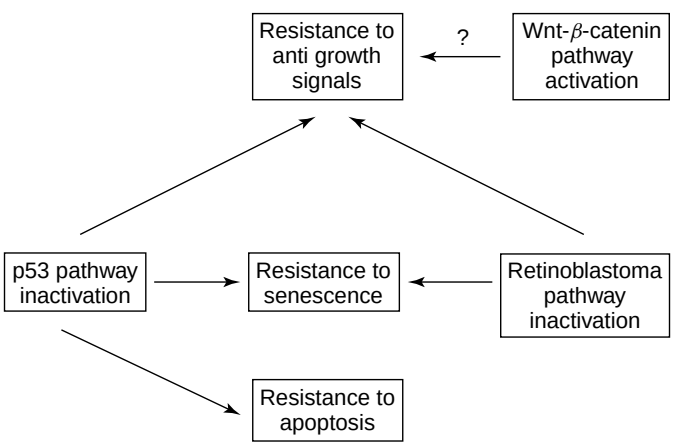

Fig. 1 Molecular mechanisms of hepatocellular carcinogenesis.

known, experimental data from colorectal cancers suggest that this activation may upregulate *Cyclin D1* and *c-myc* gene expression in these cells leading to an ability to proliferate even in the absence of mitogenic stimuli. In addition, activated Wnt-β-catenin pathway may confer HCC cells to gain a stem cell–like phenotype with the ability of unlimited self-renewal and generate progeny of differentiated cells that may count for an increased cell plasticity. With this regard, the origin of HCC is still under debate. According to the conventional theory, HCC results from the ''dedifferentiation'' of mature hepatocytes into a less differentiated state. A ''stem cell origin'' has been proposed as an alternative mechanism. According to this theory, HCC originates from ''oval'' cells displaying stem cell–like properties that are detected in the liver prior to the chemical induction of HCC in rats. Although ''oval'' cell-like structures have been described in association with different liver diseases, there is no convincing evidence for a direct contribution of liver stem cells to HCC in humans.

The molecular events leading to the development of HCC, whether it derives from a mature hepatocyte or a progenitor cell, are poorly understood. Future studies will probably focus on the role of Wnt-β-catenin pathway activation in HCC cells that will provide experimental evidence about whether such predictions are warranted.

Acknowledgments

The authors' work is supported by grants from TUBITAK, TUBA, and Bilkent University (Turkey).

See also Cancer Chemotherapy, Theoretical Foundations of; Epigenetic Mechanisms in Tumorigenesis; Intracellular Signaling in Cancer; Oncogenes; Oncology, Molecular.

Bibliography

Books and Reviews

Andrisani, O.M., Barnabas, S. (1999) The transcriptional function of the hepatitis B virus X protein and its role in hepatocarcinogenesis, *Int. J. Oncol.* **15**, 373–379.

Buendia, M.A., Hepatitis, B. (1998) viruses and cancerogenesis, *Biomed. Pharmacother.* **52**(1), 34–43.

Cadenas, E., Davies, K.J. (2000) Mitochondrial free radical generation, oxidative stress, and aging, *Free Radical Biol. Med.* **29**, 222–230.

Colombo, M. (1999) Natural history and pathogenesis of hepatitis C virus related hepatocellular carcinoma, *J. Hepatol.* **31**, 25–30.

Darlington, G.J. (1999) Molecular mechanisms of liver development and differentiation, *Curr. Opin. Cell Biol.* **11**, 678–682.

Fausto, N. (2000) Liver regeneration, *J. Hepatol.* **32**, 19–31.

Galle, P.R., Krammer, P.H. (1998) CD95-induced apoptosis in human liver disease, *Semin. Liver Dis.* **18**, 141–151.

Green, D.R., Reed, J.C. (1998) Mitochondria and apoptosis, *Science* **281**, 1309–1312.

Hanahan, D., Weinberg, R.A. (2000) The hallmarks of cancer, *Cell* **100**, 57–70.

Hurlstone, A., Clevers, H. (2002) T-cell factors: turn-ons and turn-offs, *EMBO J.* **21**, 2303–2313.

IARC. (1993) Some naturally occurring substances: food items and constituents, heterocyclic aromatic amines and mycotoxins, *IARC Monogr. Eval. Carcinog. Risks Hum.* **56**, 245–395.

IARC. (1994) Hepatitis viruses, *IARC Monogr. Eval. Carcinog. Risks Hum.* **59**, 39–221.

Llovet, J.M., Burroughs, A., Bruix, J. (2003) Hepatocellular carcinoma, *Lancet* **6**, **362**, 1907–1917.

Matsubara, K. (1991) Chromosomal Changes Associated with Hepatitis B virus DNA Integration and Hepatocarcinogenesis, in: McLacchlan, A. (Ed.) *Molecular Biology of Hepatitis B Viruses*, CRC Press, Boca Raton, FL, pp. 245–261.

Michalopoulos, G.K., DeFrances, M.C. (1997) Liver regeneration, *Science* **276**, 60–66.

Ozturk, M. (1999) Genetic aspects of hepatocellular carcinogenesis, *Semin. Liver Dis.* **19**, 235–242.

Ozturk, M., Cetin-Atalay, R. (2003) Biology of Hepatocellular Cancer, in: Rustgi, A. (Ed.) *Gastrointestinal Cancers: Biology, Diagnosis and Therapy*, 2nd edition, Lippincott-Raven Press, New York, pp. 575–596.

Sell, S., Dunsford, H.A. (1989) Evidence for the stem cell origin of hepatocellular carcinoma and cholangiocarcinoma, *Am. J. Pathol.* **134**, 1347–1363.

Shafritz, D.A., Dabeva, M.D. (2002) Liver stem cells and model systems for liver repopulation, *J. Hepatol.* **36**, 552–564.

Sherr, C.J., McCormick, F. (2002) The RB and p53 pathways in cancer, *Cancer Cell* **2**, 103–112.

Primary Literature

Bressac, B., Galvin, K.M., Liang, T.J., Isselbacher, K.J., Wands, J.R., Ozturk, M. (1990) Abnormal structure and expression of p53 gene in human hepatocellular carcinoma, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1973–1997.

Bressac, B., Kew, M., Wands, J., et al. (1991) Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa, *Nature* **350**, 429–431.

Cagatay, T., Ozturk, M. (2002) p53 mutation as a source of aberrant β-catenin accumulation in cancer cells, *Oncogene* **21**, 7971–7980.

Chaubert, P., Gayer, R., Zimmermann, A., Fontolliet, C., Stamm, B., Bosman, F., Shaw, P. (1997) Germ-line mutations of the p16INK4(MTS1) gene occur in a subset of patients with hepatocellular carcinoma, *Hepatology* **25**, 1376–1381.

de La Coste, A., Romagnolo, B., Billuart, P., Renard, C.A., Buendia, M.A., Subrane, O., Fabre, M., Chelly, J., Belford, C., Kahn, A., Perret, C. (1998) Somatic mutations of the beta-catenin gene are frequent in mouse and humanhepatocellular carcinomas, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8847–8851.

Hsu, I.C., Metcalf, R.A., Sun, T., Welsh, J.A., Wang, N.J., Harris, C.C. (1991) Mutational hotspot in the p53 gene in human hepatocellular carcinomas, *Nature* **350**, 427–428.

Hui, A.M., Sakamoto, M., Kanai, Y., Ino, Y., Gotoh, M., Yokota, J., Hirohashi, S. (1996) Inactivation of p16INK4 in hepatocellular carcinoma, *Hepatology* **24**, 575–579.

Irmak, M.B., Ince, G., Ozturk, M., Cetin-Atalay, R. (2003) Acquired tolerance of hepatocellular carcinoma cells to selenium deficiency: a selective survival mechanism? *Cancer Res.* **63**, 6707–6715.

Jung, J., Zheng, M., Goldfarb, M., Zaret, K.S. (1999) Initiation of mammalian liver development from endoderm by fibroblast growth factors, *Science* **284**, 1998–2003.

Miyoshi, Y., Iwao, K., Nagasawa, Y., Aihara, T., Sasaki, Y., Imaoka, S., Murata, M., Shimano, T., Nakamura, Y. (1998) Activation of the beta-catenin gene in primary hepatocellular carcinomas by somatic alterations involving exon 3, *Cancer Res.* **58**, 2524–2547.

Nishida, N., Fukuda, Y., Komeda, T., Kita, R., Sando, T., Furukawa, M., Amenomori, M., Shibagaki, I., Nakao, K., Ikenaga, M. (1994) Amplification and overexpression of the cyclin D1 gene in aggressive human hepatocellular carcinoma, *Cancer Res.* **54**, 3107–3110.

Ozturk, M. (1991) p53 mutation in hepatocellular carcinoma after aflatoxin exposure, *Lancet* **338**, 1356–1359.

Parkin, D.M., Bray, F., Ferlay, J., Pisani, P. (2001) Estimating the world cancer burden: GLOBOCAN 2000, *Int. J. Cancer* **94**, 153–56.

Satoh, S., Daigo, Y., Furukawa, Y., Kato, T., Miwa, N., Nishiwaki, T., Kawasoe, T., Ishiguro, H., Fujita, M., Tokino, T., Sasaki, Y., Imuoka, S., Murata, M., Shimano, T., Yamaoka, Y., Nakamura, Y. (2000) AXIN1 mutations in hepatocellular carcinomas, and growth suppression in cancer cells by virus-mediated transfer of AXIN1, *Nat. Genet.* **24**, 245–250.

Shimada, M., Hasegawa, H., Gion, T., Utsunomiya, T., Shitabe, K., Takenaka, K., Otsuka, T., Maehara, Y., Sugimachi, K. (2000)

The role of telomerase activity in hepatocellular carcinoma, *Am. J. Gastroenterol.* **95**, 748–752.

Ueda, H., Ullrich, S.J., Gangemi, J.D., Kappel, C.A., Ngo, L., Feitelson, M.A., Jay, G. (1995) Functional inactivation but not structural mutation of p53 causes liver cancer, *Nat. Genet.* **9**, 41–47.

Unsal, H., Yakicier, C., Marcais, C., Kew, M., Volkmann, M., Zentgraf, H., Isselbacher, K.J., Ozturk, M. (1994) Genetic heterogeneity of hepatocellular carcinoma, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 822–826.

Wang, J., Chenivesse, X., Henglein, B., Brechot, C. (1990) Hepatitis B virus integration in a cyclin A gene in a hepatocellular carcinoma, *Nature* **343**, 555–557.

Yu, S.Y., Zhu, Y.J., Li, W.G. (1997) Protective role of selenium against hepatitis B virus and primary liver cancer in Qidong, *Biol. Trace Elem. Res.* **56**, 117–124.

Livestock Genomes (Bovine Genome)

John Lewis Williams
Roslin Institute, Roslin, Midlothian, Scotland, UK

1 Introduction 337

2 Domestication of Cattle 338
2.1 Origins of Domesticated Cattle 338
2.2 Cattle Breeds 339
2.3 Conservation of Diversity 344

3 Developing Maps of the Bovine Genome 344
3.1 Somatic Cell Hybrids 346
3.2 *In situ* Hybridization 346
3.3 Linkage Mapping 348
3.4 High-resolution Maps 350
3.5 RH Mapping 350
3.6 Sequence Ready BAC Contigs 352
3.7 Sequencing the Bovine Genome 354

4 Identification of Genes Controlling Traits 354
4.1 Mapping Quantitative Trait Loci (QTL) 355
4.2 Dairy Traits 356
4.3 Beef Traits 358
4.4 Health Traits 359
4.5 Identifying the Trait Genes 361

5 Functional Genomics 363

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

6 In Conclusion 363

Bibliography 364
Books and Reviews 364
Primary Literature 364

Keywords

Genome
The complete nuclear DNA complement of an organism.

Genotype
The genetic make up of an individual, either at a specific locus, or considered across the whole genome.

Marker-assisted Selection (MAS)
Using DNA markers to aid the selection of animals of given genotype to achieve particular phenotypes.

Microsatellite Loci
A position in the genome where a short sequence (typically 2 or 3 bases long) is repeated several time in tandem.

Phenotype
The physical appearance or a measurable trait.

Quantitative trait loci (QTL)
A genetic locus that accounts for a proportion of the variation in a continuously distributed phenotype.

Selection
Choosing individuals with particular characteristics in order to improve the population.

Trait
A particular variable characteristic of individuals in a population.

Cattle have adapted to survive in a wide variety of environments, which has given rise to extensive diversity at the phenotypic and genetic level. Today, this diversity provides the opportunity for selective breeding and improvement of cattle for commercially desirable traits. Up to now, selection has been based on phenotype, focusing on the traits that are most easily measured. This approach has been successful, with

spectacular improvement in some traits, although there have also been associated penalties for other traits. However, intensive selection based on a limited number of phenotypes could reduce the diversity present in the population, which may have important implications for the success of future breeding objectives. In addition, selective improvement has been focused on a limited number of breeds, which are now used internationally, with the consequence that many of the genetically different breeds of cattle are being lost worldwide. To make the selection process more efficient, it is important to gain knowledge of the genes controlling particular traits and to understand the way that variation within these genes affects the traits. The first step toward identifying the trait genes has been to develop genetic and physical maps of the bovine genome. This has been achieved at an international level using several genome-mapping methods. The ultimate genome map, the bovine genome sequence, is now being determined. This genomic information is being used in specific cattle populations to identify, first the genetic location of genes controlling particular traits, then, starting from the genomic locations, to identify the genes themselves. Knowledge of the genes controlling complex traits, such as feed-conversion efficiency, health, fertility, and product quality, would allow these traits to be included in breeding objectives, with the potential for improved commercial viability, while safeguarding welfare and genetic diversity.

1 Introduction

There are more than 750 genetically different breeds of cattle worldwide, in addition to innumerable crossbred and undefined populations. This wealth of genetic diversity has enabled cattle to be raised in most environments inhabited by man; there are cattle populations adapted to survive in extremes of temperature, in arid areas and in the face of disease and parasite challenge. This environmental adaptation has, for the most part, been achieved by natural selection, insofar as those individuals that were best suited to their environment thrived and were therefore used to breed subsequent generations. The uses to which cattle have been put are equally as diverse as the environments in which they are kept. Cattle provide meat and milk for feed, are used for traction, and as a source of building materials, clothing, tallow, and glue, with numerous other uses found for the by-products of cattle production. The wealth of genetic diversity in cattle populations has provided the opportunity for selective breeding and improvement of cattle for particular uses. Today, selection has moved beyond the need for adaptation to environmental conditions to selecting on commercially desirable traits, at least in the developed world. Traditionally, breeding was carried out at the local level and those cattle that were selected for breeding were the ones that best fulfilled the local uses. More recently, cattle and semen have been moved around the world and selection has been carried out to achieve peak production under ideal, rather than local, conditions of management.

In modern breeding, the selection criteria adopted differ depending on the goals; for example, whether the end market is

beef or dairy production. However, in general, the traits used in the selection programs are those that are most readily accessible to measurement in commercial systems. Although, for example, feed efficiency is a very important factor in profitability for cattle production, this trait is often not taken into account because the trait is not readily recorded. Instead, selection is based on easily measured traits, in the case of dairy production on milk yield, protein, and fat content, and in the case of beef animals on growth rates and conformation. Over the past 40 years methods for breeding selection have improved, so that today selection of breeding sires is achieved by testing their progeny and applying sophisticated statistical methods, usually based on a ''best linear unbiased prediction'' (BLUP) model to calculate their relative genetic value for the desired traits. This approach is of greatest value in traits that are restricted by sex and age, for example, the prediction of the genetic merit of dairy-breed sires for milk traits. Application of this approach has given impressive results, for example, milk production from Holstein cows has doubled over the past 40 or so years. The progeny-testing approach has also resulted in significant improvements in beef-associated traits in the specialized beef breeds. Nevertheless, the gains achieved in a limited number of production traits come at the price of losses in other traits. The spectacular increases in milk yields achieved in the Holstein breed have been accompanied by decreased fertility and increased lameness. These losses are in part the result of an increased level of inbreeding, which is the inevitable penalty of a progeny-testing approach, as only a limited number of individuals can be tested in each generation. Along with losses in traits that are not under selection, the major problem associated with this approach, and associated with the consequent high level of inbreeding, is the potential concentration of deleterious recessive defects. As the level of inbreeding, defined by the inheritance of the same ancestral genes from both parents increases, the likelihood that an individual will inherit the same deleterious mutation from both parents also increases. This has been manifest in the elite Holstein dairy population by the appearance of genetic diseases such as DUMPS (deficiency of uridine mono-phosphate synthase), BLAD (bovine leukocyte adhesion deficiency), and most recently a bulldog calf syndrome.

In order to improve the process of genetic selection, it is important to develop a better knowledge of the genetic control of the various traits. The long-term goal is knowledge of the genes involved in particular traits, and an understanding of the variations within these genes, and how their products interact. This will ultimately allow the phenotype to be predicted from the genotype. In the shorter term, genetic maps and physical maps of the bovine genome are allowing markers for genes involved in particular traits and diseases to be localized, and in some cases, the genes themselves to be identified. The progress in unraveling the organization of the bovine genome is discussed below.

2 Domestication of Cattle

2.1 Origins of Domesticated Cattle

Analysis of both mitochondrial and nuclear DNA allows the relationships between cattle populations and their origins to be investigated. Domesticated cattle can

be classified into two groups on the basis of phenotype: *Bos indicus*, or Zebu cattle, are characterized by a hump at the shoulder, and *Bos taurus*, or Taurine cattle, which are humpless cattle. These two groups are sometimes referred to as subspecies, but they freely interbreed. *Bos indicus* cattle are better adapted to arid regions and crossbreeding between the two subspecies is often used to produce cattle with improved productivity in unfavorable dry regions. Comparison of mitochondrial "D-loop" DNA sequence confirms that the phenotypic distinction is reflected at the molecular level. The D-loop sequences fall into two groups, or clades, that coincide broadly with their geographic distribution. The exceptions are the African Zebu cattle, which have humps and phenotypically appear to be *B. indicus*, but have mitochondrial DNA that is typically taurine.

The distinction between *B. indicus* and *B. taurus* is also found by examining sequence diversity of nuclear DNA. Comparison of nuclear DNA is most commonly achieved by examining allelic variation at microsatellite loci (see Sect. 3.3). Construction of phylogenetic relationships between cattle using microsatellite data also produces two major clusters, which correspond with the indicus and taurine cattle. In the case of nuclear DNA the African Zebu cattle cluster with the taurine breeds, suggesting that they are hybrids. Examination of sequence variation on the Y chromosome and comparison with the maternally derived mitochondrial DNA allows the migration and dynamics of breeding to be examined, and confirms the introgression of male indicus genes into the original African taurine breeds.

The history of cattle domestication can be traced using both the mitochondrial and microsatellite data, and suggests that the indicus and taurine lineages diverged at least 22 000 years ago. The data also suggest two sites for domestication of cattle, both in the Middle East, from where domesticated cattle radiated out: the eastern domestication spreading to India, founding the indicus breeds, which were subsequently imported into east Africa, and the western site of domestication which spread via north Africa into Europe, founding the taurine breeds as shown in Fig. 1.

2.2 Cattle Breeds

Since domestication, cattle have been raised in a large variety of environmental conditions. It is therefore not surprising that extensive diversity began to appear both at the phenotypic and at the genetic level. Traditionally, breeding was carried out at a local level, often using a limited number of shared bulls. Originally cattle were selected primarily for survival, and fulfilled a range of production requirements, meat, milk, traction and so on, with equal emphasis. For the most part this is still the case, with the 70% of cattle raised in the developing world. The selection of individuals with particular characteristics suited to local environments, needs, and preferences, led to the emergence of distinctive groups of cattle with characteristic phenotypes. This phenotypic distinctiveness of groups of cattle was pursued by further selection, and later these groups were formalized into breeds, which were characterized by fixing particular features, typically colors or markings. With the establishment of breeds and the formation of societies to promote them, particularly in Europe and North America, there came the tendency for selection aimed at breeding cattle to suit particular production niches, such as breeds specializing in dairy or meat production. Some breeds that started

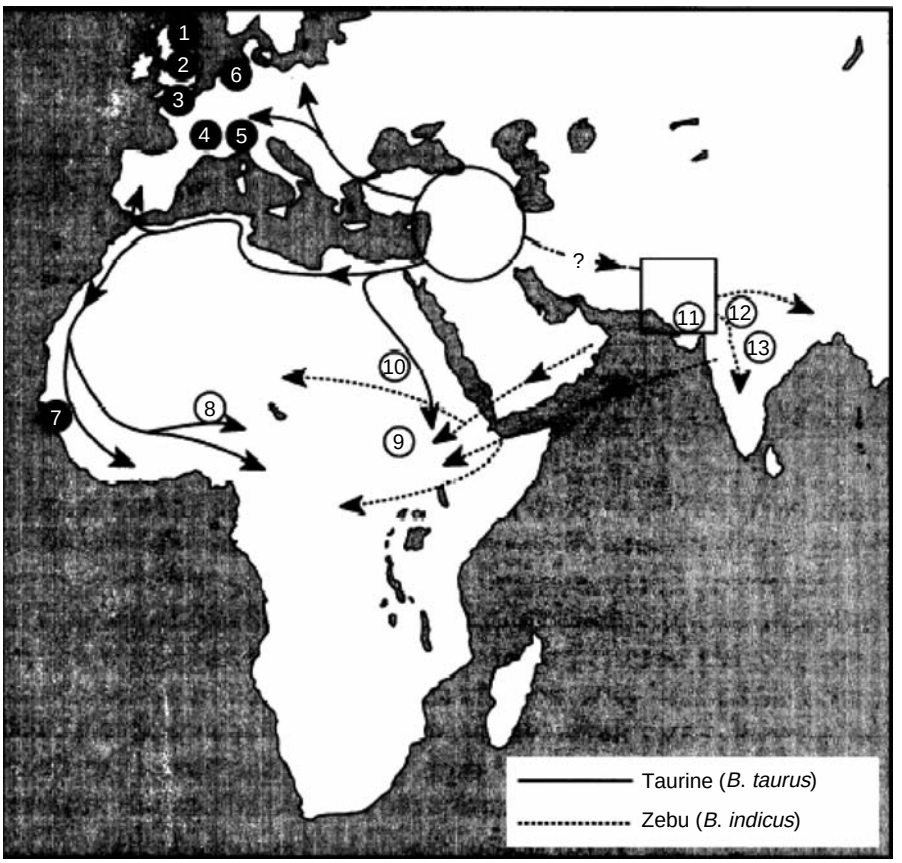

Fig. 1 Postulated migratory routes of cattle across western Asia, Africa, and Europe. Geographical origins of breed determined from mtDNA analysis of the following breeds (1) Aberdeen Angus, (2) Hereford, (3) Jersey, (4) Charolais, (5) Simmental (6) Friesian (7) N'Dama, (8) White Fulani, (9) Kenana, (10) Butana, (11) Tharparkar, (12) Sahiwal and 13 Hariana. *Bos taurus* breeds are illustrated by black symbols and *Bos indicus* breeds by open symbols (from Loftus, R.T., MacHugh, D.E., Bradley, D.G., Sharp, P.M., Cunningham, E.P. (1994) Evidence for two independent domestications of cattle, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2757–2761). Figure 1 (from Bradley et al. showing domestication origins).

with a dual function have more recently undergone strong selection for particular traits. The Friesian, which was used initially for both dairy and meat production, has become the leading dairy breed in the guise of the Holstein, in terms of quantity of milk produced. Modern dairy breeds produce far more milk than would be necessary to feed a calf, and often have trouble maintaining body weight during lactation. Conversely, selection of the meat-producing breeds has resulted in animals with little milk, in some cases insufficient milk for their calves, but with impressive muscular development, and in extreme cases are so muscle-bound that they have problems with locomotion. Thus, artificial selection has produced individuals that would not survive under natural selection.

The phenotypic distinctiveness of cattle breeds is also reflected at the genetic level. From the early 1960s through to the present day, biochemical markers, such as enzyme polymorphisms and later blood groups on the surface of red cells, detected using allo-antisera, has been used extensively by breed societies to verify the pedigrees of individuals and maintain the purity of their breeds. This biochemical information can also be used to track the genetic changes of breeds over this 40-year period, which coincides with the time during which the intensity of selection for increased productivity has also increased dramatically. Analysis of the blood-group data has shown that, in general, between-breed diversity is much larger than within-breed diversity. So individuals from the same breed cluster together. The phenotypic appearance of cattle breeds has changed in response to the intense selection that has been applied over the past 60 years. In addition, the livestock market has become international with semen, and to a lesser extent embryos, from many breeds being shipped all over the world. This has provided the opportunity for genetic divergence within the breeds as different selection criteria are applied in different regions; nevertheless, the genetic composition of breeds has remained relatively constant. Comparison of blood typing data over a 40-year period shows that although there has been some genetic drift, individuals remain more like their ancestors from the same breed, than individuals from other closely related or phenotypically similar breeds (see Fig. 2). Examination

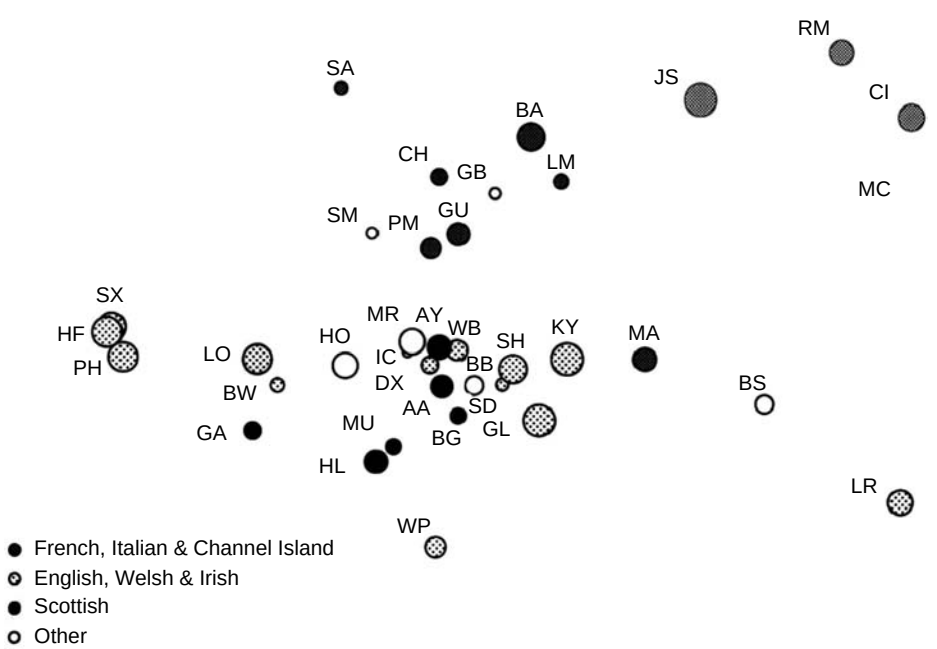

Fig. 2 Relationships between breeds based on a principal components analysis of differences between frequencies of red cell antigens (from Blott, S.C., Williams, J.L., Haley, C.S. (1998) Genetic relationships among European cattle breeds, *Anim. Genet.* **29**, 273–282). Figure 1 (from Bradley et al. showing domestication origins).

of individuals from different geographical locations reveals variations between the populations in different areas, most likely arising from a founder effect; however, the variation among individuals from a breed still remains smaller than the variation between breeds.

Other studies have used microsatellite loci to examine genetic diversity of cattle at the DNA level. A panel of 30 microstatellite loci has been selected, through a European Union network and have been adopted by the Food and Agriculture Organization of the United Nations for studies of diversity in cattle. This panel of markers is recommended for use in all studies of genetic diversity in cattle (see Table 1). The use of a common panel of markers should allow data to be integrated and compared between studies; this will allow changes in the genetic composition of breeds to be followed over time and genetic erosion to be monitored. The results from the microsatellite-based studies reflect the

Tab. 1 List of markers recommended for the study of genetic diversity in cattle.

Marker number	*Marker*	*Chromosome*	*Primer sequences (5′-3′)*
	INRA063		ATTTGCACAAGCTAAATCTAACC
1	(D18S5)	18	AAACCACAGAAATGCTTGGAAG
	INRA005[2]		CAATCTGCATGAAGTATAAATAT
2	(D12S4)	12	CTTCAGGCATACCCTACACC
	ETH225		GATCACCTTGCCACTATTTCCT
3	(D9S1)	9	ACATGACAGCCAGCTGCTACT
	ILSTS005		GGAAGCAATGAAATCTATAGCC
4	(D10S25)	10	TGTTCTGTGAGTTTGTAAGC
	HEL5[1]		GCAGGATCACTTGTTAGGGA
5	(D21S15)	21	AGACGTTAGTGTACATTAAC
	HEL1		CAACAGCTATTTAACAAGGA
6	(D15S10)	15	AGGCTACAGTCCATGGGATT
	INRA035		ATCCTTTGCAGCCTCCACATTG
7	(D16S11)	16	TTGTGCTTTATGACACTATCCG
	ETH152		TACTCGTAGGGCAGGCTGCCTG
8	(D5S1)	5	GAGACCTCAGGGTTGGTGATCAG
	INRA023		GAGTAGAGCTACAAGATAAACTTC
9	(D3S10)	3	TAACTACAGGGTGTTAGATGAACTC
	ETH10[4]		GTTCAGGACTGGCCCTGCTAACA
10	(D5S3)	5	CCTCCAGCCCACTTTCTCTTCTC
	HEL9		CCCATTCAGTCTTCAGAGGT
11	(D8S4)	8	CACATCCATGTTCTCACCAC
	CSSM66[3]		ACACAAATCCTTTCTGCCAGCTGA
12	(D14S31)	14	AATTTAATGCACTGAGGAGCTTGG
	INRA032[5]		AAACTGTATTCTCTAATAGCTAC
13	(D11S9)	11	GCAAGACATATCTCCATTCCTTT
	ETH3		GAACCTGCCTCTCCTGCATTGG
14	(D19S2)	19	ACTCTGCCTGTGGCCAAGTAGG
	BM2113		GCTGCCTTCTACCAAATACCC
15	(D2S26)	2	CTTCCTGAGAGAAGCAACACC

Tab. 1 (*continued*)

Marker number	*Marker*	*Chromosome*	*Primer sequences (5′-3′)*
	BM1824		GAGCAAGGTGTTTTTCCAATC
16	(D1S34)	1	CATTCTCCAACTGCTTCCTTG
	HEL13[5]		TAAGGACTTGAGATAAGGAG
17	(D11S15)	11	CCATCTACCTCCATCTTAAC
	INRA037		GATCCTGCTTATATTTAACCAC
18	(D10S12)	11	AAAATTCCATGGAGAGAGAAAC
	BM1818		AGCTGGGAATATAACCAAAGG
19	(D23S21)	23	AGTGCTTTCAAGGTCCATGC
	ILSTS006		TGTCTGTATTTCTGCTGTGG
20	(D7S8)	7	ACACGGAAGCGATCTAAACG
	MM12		CAAGACAGGTGTTTCAATCT
21	(D9S20)	9	ATCGACTCTGGGGATGATGT
	CSRM60		AAGATGTGATCCAAGAGAGAGGCA
22	(D10S5)	10	AGGACCAGATCGTGAAAGGCATAG
	ETH185		TGCATGGACAGAGCAGCCTGGC
23	(D17S1)	17	GCACCCCAACGAAAGCTCCCAG
	HAUT24		CTCTCTGCCTTTGTCCCTGT
24	(D22S26)	22	AATACACTTTAGGAGAAAAATA
	HAUT27		TTTTATGTTCATTTTTTGACTGG
25	(D26S21)	26	AACTGCTGAAATCTCCATCTTA
	TGLA227		CGAATTCCAAATCTGTTAATTTGCT
26	(D18S1)	18	ACAGACAGAAACTCAATGAAAGCA
	TGLA126		CTAATTTAGAATGAGAGAGGCTTCT
27	(D20S1)	20	TTGGTCTCTATTCTCTGAATATTCC
	TGLA122		CCCTCCTCCAGGTAAATCAGC
28	(D21S6)	21	AATCACATGGCAAATAAGTACATAC
	TGLA53		GCTTTCAGAAATAGTTTGCATTCA
29	(D16S3)	16	ATCTTCACATGATATTACAGCAGA
	SPS115		AAAGTGACACAACAGCTTCTCCAG
30	(D15)	15	AACGAGTGTCCTAGTTTGGCTGTG

Note: These markers were initially selected from those used across several projects and subsequently were adopted as the standard panel for work in cattle. More information can be obtained from http://www.projects.roslin.ac.uk/cdiv/markers.html.

same results as studies with biochemical markers, that is, for anonymous markers that are not subjected to selection: (1) breeds of cattle can be distinguished at the genetic level, and (2) within-breed variation is lower than between-breed variation. DNA markers are now replacing biochemical polymorphisms for pedigree verification, so a body of data is building up that could be used in the future to explore variations in the genetic diversity of cattle breeds.

The use of DNA markers has also been mooted for identifying the breed of an animal, or more appropriately in commercial terms the breed of origin of a meat sample. Unfortunately, although the analysis of DNA polymorphisms allows animals to be clustered by breed, there is overlap between the clusters. Therefore,

although it is statistically possible to assign individuals to breeds, there is likely to be considerable error in the result, at least with the number of markers currently used.

2.3
Conservation of Diversity

The introduction of artificial insemination (AI) in the more developed countries during the 1950s, coupled with improvement in management, has resulted in rapid progress in the improvement of cattle for simple production traits. This increasing use of AI meant that particular bulls with desirable characteristics became widely used in preference to local bulls. Consequently, where the economic environment supports high input agriculture, there has been a dramatic increase in milk yield and meat produced from the improved stock. The unfortunate consequence has been the reduction of genetic diversity, both within the selected breeds, as superior individuals have been used preferentially as breeding stock, and also through the replacement of traditional, less productive, breeds. In less developed and environmentally less favored areas, the use of the improved breeds presents a great cause for concern. Local breeds are usually adapted to survive in their local environments, for example, with increased tolerance of extremes in temperature or with the ability to survive and remain productive in the face of particular disease or parasite challenge. The inappropriate and/or unmanaged attempts to introduce improved breeds into some areas have met with disastrous consequences. In 1993, 112 of the 783 cattle breeds worldwide were at risk of extinction. One of the greatest risks is the replacement of local stock that are adapted for survival in the face of disease challenge with disease sensitive stock, in areas where standards and resources to provide extensive veterinary care are not available. It is important that genetic diversity is maintained, as this is the source of variation that will allow further selection. The first essential task is to record and monitor diversity across cattle populations. Screening genetic diversity using a limited number of DNA markers, such as the FAO panel discussed above, provides a crude measure of genetic diversity; however, this approach does not identify genes that are involved in specific adaptation or that control unique phenotypes. Thus, to preserve genetic diversity, much more extensive studies that consider the whole genome in more detail and that take into account variations in phenotypes are necessary. To this end, genome-mapping programs have the goal of identifying the genetic control of phenotypic variation.

3
Developing Maps of the Bovine Genome

The bovine genome is contained within 29 autosomes and the X and Y sex chromosomes, and consists of about 3 000 Mb of DNA. The autosomes are all acrocentric, that is, the centromere is at the end, while the sex chromosomes are sub-metacentric. Like most other mammalian genomes, the bovine genome is composed of less than 10% sequence that codes for proteins and about 50% of repetitive sequences. One major repetitive component is the bovine A-2 repeat, which has a core 115-bp element and comprises around 1.6% of the bovine genome. The second most common repeat element is the bovine SINE or bovine – B repeat, which is characterized by a 560-bp band seen by electrophoresis following digestion with the restriction enzyme *Pst*1,

the whole of this element is 3.1 kbp long and represents about 0.5% of the genome (see Fig. 3).

There has been considerable effort internationally to produce maps of the bovine genome driven for the most part by the objective of finding those genes that are involved in commercially important traits. Genomic maps can be compared between species to study the evolution of genomes and to reveal the organization and architecture of the DNA. This comparison will contribute to an understanding of the way the genome functions and the identification of those features that are important and that should be subjected to further study. There are several types of genomic map, which are either physical, or genetic. Physical maps can either be based on the localization of

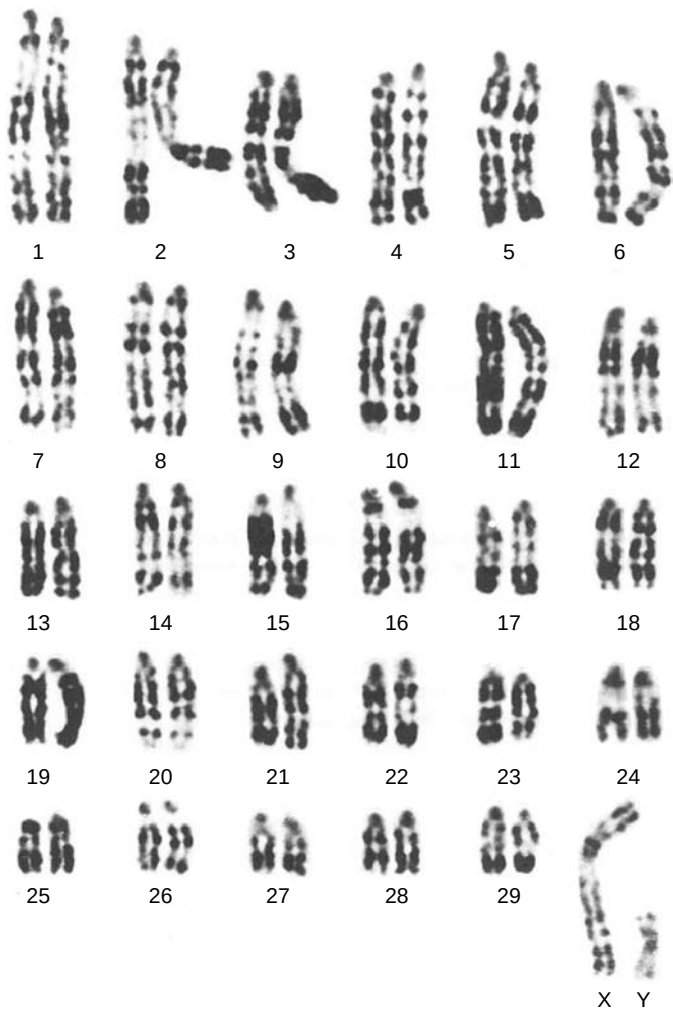

Fig. 3 R banded Bovine Chromosomes (from Fries, R., Ruvinsky, A. (1999) *The Genetics of Cattle*, CABI Publishing, Oxford, UK).

markers on the chromosomes by *in situ* hybridization, or they can be clone-based maps where the genome is fragmented and cloned and the clones physically reassembled so that the location and orientation of the pieces in relation to each other is known. A third type of physical map uses somatic cell hybrids, where fragments of the target genome are transferred into cells of another species; statistical methods are then used to calculate the proximity of markers from their coretention in the hybrid cells. In contrast, linkage maps are assembled on the basis of relationship between markers deduced from patterns of recombination in families. The ordering of loci should correspond between physical and linkage maps, but the distances between loci will vary due to the differences in the way the relationship between them is worked out. The ultimate genomic map of a species will be its DNA sequence.

3.1
Somatic Cell Hybrids

The first maps of the bovine genome were created using somatic cell hybrids made by fusing bovine cells with immortalized rodent cells (usually mouse or hamster). The immortalized cells are deficient in a selectable marker, usually thymidine kinase; thus, to survive in selection medium they need to retain the bovine Tk gene. In practice, the rodent cells retain large fragments of bovine chromosomes. Using a panel of somatic cell hybrids it is possible, using statistical methods, to identify which markers occur on the same chromosome, and to a certain extent the order of markers. In 1993, the first bovine genome map was published with 30 synteny groups, which were identified using a somatic cell hybrid panel. With the exception of the X chromosome, the synteny groups were not assigned to chromosomes and designated U1 to U29.

3.2
In situ Hybridization

In order to assign markers to bovine chromosomes, the method of *in situ* hybridization was initially used. Originally, radioactively labeled gene probes were hybridized to metaphase chromosomes spread out on microscope slides. The slides were then coated with photographic emulsion and following development the distribution of silver grains was analyzed. The distribution of silver grains was then compared with the banding pattern of the stained chromosomes using statistical analysis to assign the physical location of gene to cytogenetic bands identified on the standard karyotype. All the unassigned synteny groups identified by somatic cell hybrids were assigned to chromosomes using radioactive *in situ* hybridization. However, this approach is not very precise, as the silver grains are widely dispersed and a large number of hybridizations were required to assign a probe to a particular band with confidence, and even then the localization was at low resolution. A refinement of this methodology was to use fluorescently labeled probes, in a method called *fluorescent in situ hybridization* (FISH). With FISH, the hybridized probe is visualized directly; this increases the precision of the method and as a result reduces the number of chromosome hybridizations necessary to be confident of the localization. Currently there are around 500 genes localized on chromosomes by *in situ* mapping. In addition to mapping genes, it is possible to localize the anonymous markers used in linkage mapping to chromosomes using

FISH by labeling the cosmid clones from which the anonymous marker was derived, and using it as the probe. This approach allowed the physical bovine map and the linkage maps to be aligned.

The FISH approach allows the chromosomal band harboring the target sequence of the probe to be identified, and has a resolution at the DNA level of 2 Mbp or more. A higher resolution can be obtained if less condensed chromosomes, or indeed cloned DNA strands are used in the Fiber-FISH technique, which allows a localization of probes at a resolution of about a kilobase. The physical location of a probe to a chromosomal band obtained using the fiber-FISH approach also provides the information required for cloning of targeted regions by chromosomal microdissection.

Information obtained from somatic cell panels and *in situ* hybridization of genes demonstrated that in many cases the genes that came from the same chromosomal region in man or mouse also mapped to a single chromosome in cattle; indeed, genes from one region in one species was generally found close together on a chromosome of another, as shown in Fig. 4. This suggested the organization of the genomes of divergent species has been conserved during evolution. The extent to which this organization is conserved has been elegantly demonstrated using chromosomal probes derived from one species to "paint" the chromosomes of another using a ZOO-FISH approach. ZOO-FISH suggested that there are about 50 conserved segments of genome between the human and bovine genomes.

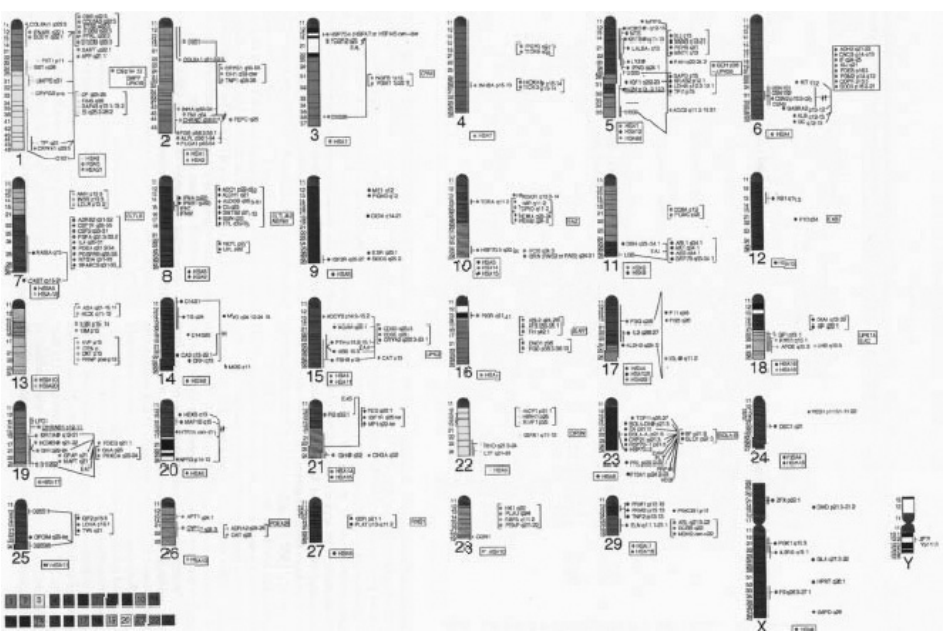

Fig. 4 Painting bovine chromosomes with human chromosome paints shows that large regions of chromosomes are conserved across species, at least at the gross level. (From Solinas-Toldo, S., Lengauer, C., Fries, R. (1995) Comparative Genome map of human and cattle, *Genomics* **27**, 489–596.)

3.3 Linkage Mapping

Linkage mapping relies on recombination occurring less frequently between markers that are close together than those that are more distant on chromosomes. Therefore, by analyzing the inheritance of markers in families it is possible to build a genetic map based on the frequency of recombination, or linkage, between the markers. To build a genetic map, there are two requirements: families of animals in which the inheritance of chromosomes can be tracked, and markers that are highly variable (polymorphic) so that in the majority of cases the parental origin of the marker can be determined. By combining the information on the inheritance of a large number of markers within families, it is possible to construct a genetic map by analyzing the recombination frequency between pairs of markers.

Creating genetic maps for cattle has two immediate problems: firstly, cows generally have singleton calves, and secondly, the generation for period for cattle is relatively long; therefore, cattle families are usually small. Producing a reasonable number of full siblings by natural breeding is both a long-term and expensive exercise. If conventional cattle families were used, in general, two parents and their calf would have to be typed for each marker. The situation in commercial herds is generally better as artificial insemination is used for breeding cattle and hence large half-sib sire families are readily available. Half-sib commercial populations have been extensively used for mapping the loci involved in the control of commercial traits, but in practice the contribution of the dam to the mapping information is limited, as she will only contribute one or two calves. For the construction of the genetic maps, large full-sib families are much more efficient. Fortunately, multiple ovulation embryo transfer (MOET) technology has been developed for cattle, which means a large number of progeny can be produced from two parents. MOET has allowed relatively large full-sib populations to be assembled as "reference populations" for genetic mapping. The advantage of full-sib families is that the effort required to type the markers is significantly reduced, while the resolution of the maps produced is increased. To obtain the highest density of markers on the genetic map, it is of benefit for all workers in the field to focus on a limited number of mapping populations. The international efforts to create bovine linkage maps focused on two reference populations: one, an international collaboration led by Australia with full-sib families provided by Australia, Kenya, Europe, and the United States to create the International Bovine Reference Pedigree (IBRP), and the second, focused on a reference population produced by the USDA Meat Animal Research Center in Nebraska. The distinction has been drawn between *reference* families, which have been specifically created and used to make genetic maps, and *resource* families, which have been assembled to localize the genes involved in particular traits.

The second requirement for constructing a genetic map is the markers. Whereas *in situ* hybridization and somatic cell hybrid mapping simply requires probes for the markers that can be hybridized to the DNA, the resolution of the linkage map is governed by the number of informative meiosis, that is, the more frequently the parental origin of markers can be determined the higher the resolution. Thus, it is important that the markers used for linkage mapping are highly polymorphic in the reference families. In general, genes

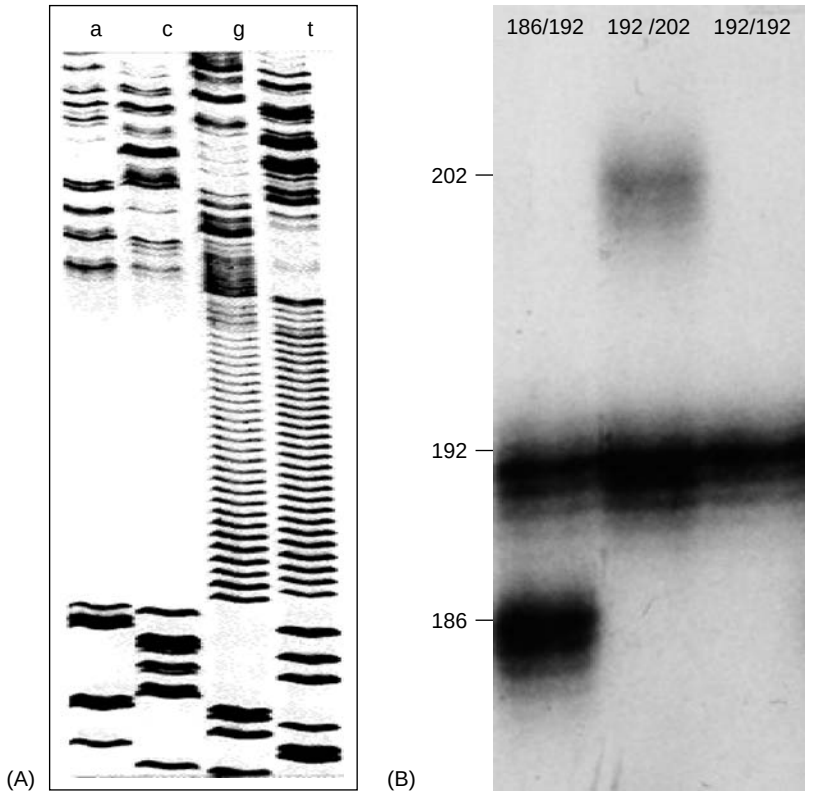

Fig. 5 Panel A. A sequencing gel showing a microsatellite locus with 31 repeats of the GT dinucleotide motif. Panel B. A microsatellite marker labeled with a radioactive isotope and visualized by exposure to photographic film. The marker has three alleles with sizes shown in base pairs along the side. The genotypes of each individual is shown at the top.

have less variation than the intervening DNA sequences; therefore, markers selected from anonymous stretches of DNA between genes have been mostly used to construct the linkage maps. The preferred type of "anonymous" marker is based on microsatellite loci. These loci are tandem repeats of simple sequences 1 to 4 bases in length, the most common being dinucleotide repeats of alternating C and T. Microsatellite loci seem to have a high mutation rate, which is thought to arise from DNA strand–slippage during replication, giving rise to variable numbers of the repeat (see Fig. 5). Therefore, each locus has a relatively large number of alleles, typically between 3 and 10. The number of alleles at a microsatellite locus is approximately proportional to the number of copies of the repeat unit. These microsatellite loci occur at about 40-kbp intervals and cover the whole genome, and therefore provide excellent markers for linkage mapping. By constructing primers flanking the microsatellite locus, the polymerase chain reaction (PRC) allows rapid genotyping of the marker.

The international mapping effort produced sequence information for a large number of these microsatellite loci during the early 1990s, which were then genotyped in the IBRP and MARC reference families to produce genetic maps. Each of these maps had around 1000 markers, mostly the microsatellite loci. Some of the markers were specific to one or the other map, while a large number of markers were common to both. In addition to the genetic maps produced using the reference families, resource families have been typed for a large number of microsatellite loci in order to localize the genes involved in various traits; some mapping studies have published genetic maps based on the information generated from the resource herd they used. As a significant number of the markers used to localize the trait loci are common between the studies, and are also in common with the genetic maps, the information can be combined to improve the resolution and confidence of ordering of the loci included in the maps. A series of international chromosome workshops, held under the auspices of the International Society for Animal Genetics, have assembled the various data into consensus maps. There are now well over 2400 microsatellite markers mapped on the bovine genome (see http://locus.jouy.inra.fr/cgi-bin/lgbc/mapping/common/intro2.pl?BASE=cattle or http://sol.marc.usda.gov/genone/cattle/cattl.html).

3.4 High-resolution Maps

Both *in situ* hybridization and linkage-mapping approaches have limitations with respect to creating detailed genome maps: the resolution of both is limited: the *in situ* approach, by the physical limits of observation, and linkage mapping, by the size of resource families used, and hence the number of informative meiosis available to order markers. A further drawback of the linkage-mapping approach is that the markers must be polymorphic, so that alleles on different parental chromosomes can be readily identified to track the inheritance of the chromosomal regions. The gross conservation of synteny between genomes discussed above may be disrupted at the level of gene order, with blocks of genes being rearranged within a broadly conserved framework. It is, therefore, important that mapping studies are able to include genes within the maps, which can then be used to compare genomic organization across species. Genes show much more limited levels of polymorphism than the microsatellite loci, and hence are more difficult to place on linkage maps accurately.

3.5 RH Mapping

Like linkage mapping, the radiation hybrid (RH) mapping approach relies on markers that are close together in the genome being separated less frequently when chromosomes are broken than more distant markers. In the case of RH mapping, the separation of markers is achieved by random fragmentation of the genome by radiation, rather than by recombination. An RH mapping-panel is constructed by irradiating cells from the target species, which are then fused with immortalized recipient hamster cells as shown in Fig. 6. The hamster cells used are deficient for a selectable marker (TK or HPRT); thus, growth in selection medium requires the retention of the region of the donor genome that carries the selectable marker; however, in addition many other fragments, typically between

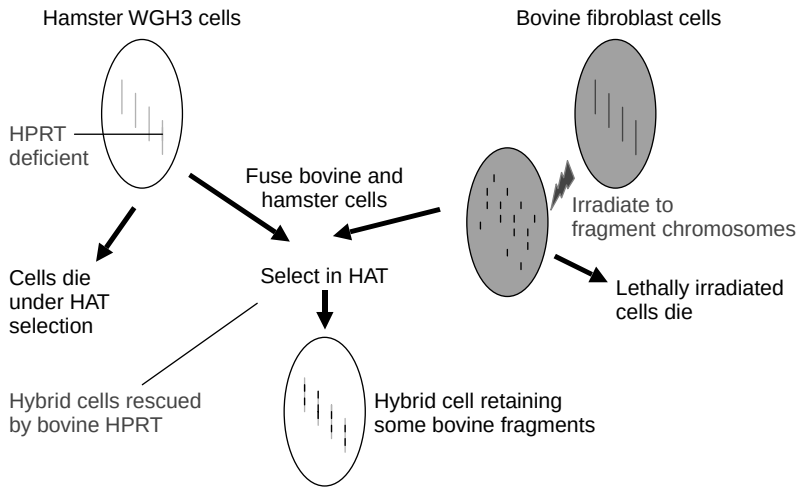

(A)

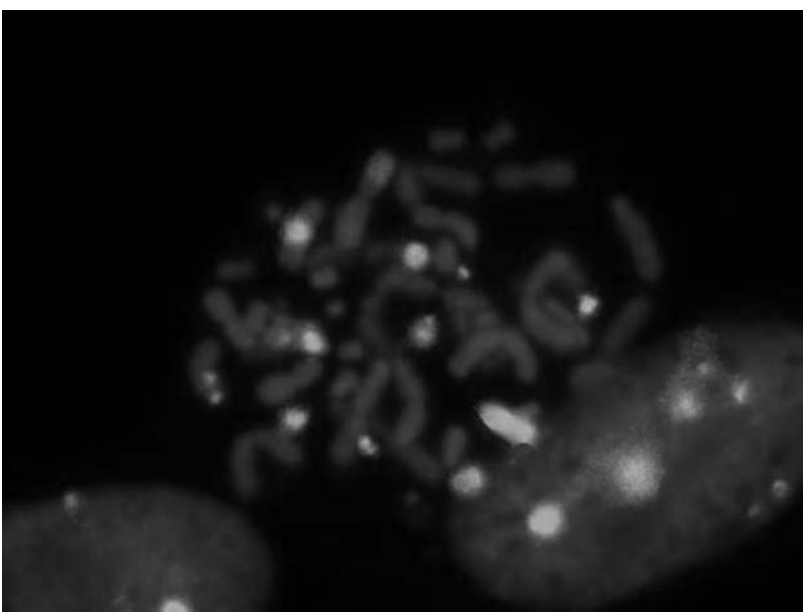

(B)

Fig. 6 Panel A: Construction of Radiation Hybrid cells. Fragments of the donor (bovine) chromosomes are retained in the immortalized rodent (hamster) cells that are rescued in selection medium by retaining the selectable marker from the donor genome. Panel B illustrates the presence of bovine genome fragments in the RH cells of the Bovine 3000 rad RH panel Williams, J.L., Eggen, A., Ferretti, L., Farr, C., Gautier, G., Amati, G., Ball, G., Caramori, T., Critcher, R., Costa, S., Hextall, P., Hills, D., Jeulin, A., Kiguwa, S.L., Ross, O., Smith, A.L., Saunier, K.L., Urquhart, B.G.D., Waddington, D. (2002) A Bovine Whole Genome Radiation Hybrid Panel and Outline Map, *Mamm. Genome* **13**, 469–474. (Photograph kindly provided by Dr C Farr Department of Genetics University of Cambridge).

15 and 40% of the donor genome is retained in each hybrid. A panel of RH cell lines is assembled in which each line contains different combinations of fragments from the donor genome. The presence or absence of markers in the cells of the hybrid panel is then used to calculate the relationship between the markers and build them into the RH map.

The RH approach has two major advantages over physical mapping by *in situ* hybridization and genetic-linkage mapping. Firstly, changing the radiation dose used in the construction of the panel allows panels with different resolutions to be created and secondly, the markers do not need to be polymorphic to be included in the map. Three RH panel have been described for cattle, two of which were constructed using a relatively low radiation dose (3–5000 rads) and one higher resolution panel (12 000 rads). There are now over 2000 markers mapped on the low-resolution panels and framework maps published. The inclusion of a proportion of microsatellite in the RH maps allows them to be aligned with the genetic maps (see Fig. 7). However, the main advantage of the RH panels is to allow the inclusion of genes or ESTs in the bovine map and so detailed comparative maps between the bovine and other genomes can be made.

3.6
Sequence Ready BAC Contigs

Identifying the genes involved in particular traits in cattle is currently carried out by linkage mapping followed by a positional candidate gene approach. This approach identifies genes that may be expected to be involved in the trait because of the function of their products and which map within the appropriate region. The candidate genes are then tests for involvement with the trait. This approach relies heavily on comparative mapping information to make use of data available for the human genome to provide the positional candidate genes. In the absence of positional candidate genes, or if the candidate genes identified do not prove to be the genes involved in the trait, the next approach is to construct a set of overlapping contiguous clones (called "Contigs") spanning the region of the chromosome that harbors the target gene. Contigs are usually constructed using clones containing large fragments of inserted genomic DNA, either cosmids (~40 kbp inserts), YACs (up to 1 Mbp inserts) or BACs (100+ kbp inserts). Construction of contigs is usually achieved by screening the large fragment libraries with markers that are available for the region of interest, then joining the clones together by "walking," which is achieved by using a probe from the end of one clone to find the next, and so on until clones covering the region of interest are found.

As the chromosomal locations for the genes controlling more traits are discovered, it becomes sensible to assemble a whole genome contig in which the clones from a large fragment library are placed in order along the each of the chromosomes. The whole genome contig then provides easy access to clones for particular regions for local sequencing and gene hunting projects. Selection of a "minimum tiling path" which contains the minimum number of overlapping clones to span all the chromosomes also provides the ideal underpinning resource from which to sequence the genome.

At the time of writing, the assembly of a whole genome BAC contig for the bovine genome is under way. The approach adopted is to use a BAC fingerprinting technique for the initial ordering

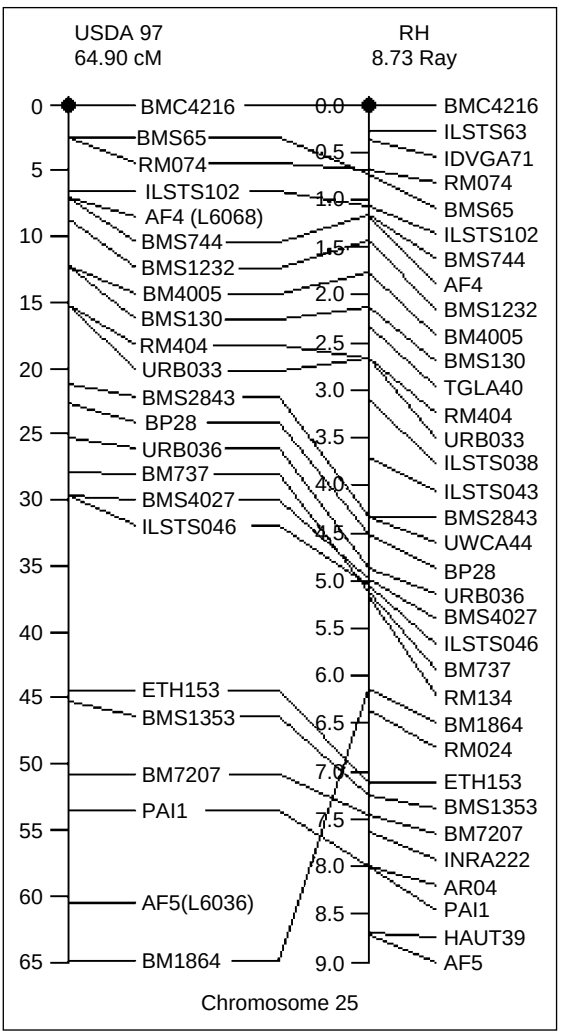

Fig. 7 Examples of genetic and RH maps for chromosome 25. The order of markers is broadly the same between methods; however, the relative distance between markers varies (from Williams et al. 2003).

of the clones. With this approach, BAC clones are digested with two restriction enzymes and sophisticated software is used to match fragment sizes and identify clones that contain regions of DNA that are common by identifying shared restriction fragments. The contig assembled from the fingerprinting is then confirmed by end sequencing the BACs and checking for overlapping sequence. With the availability of the human sequence, the BAC sequences can be used to align the bovine contigs with the human genome. However, interruptions in the conservation of synteny between human and bovine genomes may introduce errors into the ordering of the contigs; thus, it is unlikely that this approach will generate a continuous contig of all chromosomes. The order and orientation of contigs in the bovine genome will therefore be confirmed by aligning the BAC contig with the bovine

RH maps by typing markers in common between the BACs and the RH maps.

Two BAC libraries are currently being fingerprinted and the results merged. At the British Columbia Cancer Agency, 280 000 cattle BAC clones have been fingerprinted from libraries created at the BACPAC resources (Children's Hospital, Oakland Research Institute, Oakland, California), while at INRA in France, 90 000 clones have been fingerprinted. These data are currently being merged to produce the around 15-fold BAC clone coverage of the bovine genome. End sequencing of the BAC clones has just started and is being undertaken by the international bovine research community. The database of results will be maintained at The Genome Sciences Center, Vancouver, British Columbia, Canada. In parallel, the Texas and Roslin RH panels are being used to assist with the ordering of the clones.

3.7
Sequencing the Bovine Genome

The Human Genome Project announced the completion of the first draft of the entire human sequence in 2001 with the essential completion of the human genome sequence in 2003. This provides a vast resource for understanding the organization of the human genome and provides the template against which other genomes can be compared. The availability of the human sequence will also allow scientists to identify new genes and explore the evolutionary history of genomes. However, on its own the information contained within the human sequence is limited compared with the potential information that can be obtained by comparing and contrasting genomes between closely and distantly related species. Comparing the differences as well as the similarities between genomes is the key to understanding the functionally relevant features. Therefore, in addition to decoding the genetic makeup of humans, the sequence of a number of model organisms are also being analyzed, including baker's yeast, the roundworm, the laboratory mouse, and a rough draft of the sequence of the rat genome was produced in November 2002. The construction of the complete physical map of the bovine genome, by linking contiguous BAC clones to span all chromosomes, provides important resources that are the foundation for a bovine genome-sequencing project. The shotgun sequencing approach has inherent problems associated with assembling sequence data, whereas the identification of ordered BAC contigs and the clones constituting the minimum tiling path, provide the framework and the means to assemble the genome sequence efficiently. At the time of writing, a bovine genome-sequencing project is underway at Baylor College of Medicine Sequencing Center (see http://hgsc.bcm.tmc.edu/projects/bovine/)

4
Identification of Genes Controlling Traits

Genetic improvement of cattle has focused on traits that impact on profitability of agricultural enterprises; traditionally this meant increased output. Simple production traits are easiest to measure and so have been used in conventional selection programs with great success. However, using modern methods selection on a broader spectrum of traits can be considered and other traits that impact on profitability and welfare, such as

feed-conversion efficiency, health, fertility, and product quality can be included in the selection objectives. If selection is phenotype-based, improvement in one trait is compromised by simultaneously including other traits in the breeding goals. However, using information on the genes controlling the traits potentially allows for selection on genotype and hence on several traits simultaneously. In theory, once sufficient knowledge is available, individuals carrying beneficial genes for several traits could be identified using DNA markers and mated to produce progeny with the desired characteristics. While we are a long way from this level of knowledge, work is under way to identify the genes involved in a wide range of traits. The ultimate goal of molecular genetics is to understand how allelic variations in the genes controlling various traits interact and result in the observed variations in the phenotype. Progress toward an understanding of the genetic control of traits can be achieved if the location in the genome of the genes involved is known, and markers linked to the genes are found. Once information has been accumulated to predict the alleles present in an individual, these linked markers can be used to aid breeding by marker-assisted selection (MAS).

Identifying the genes involved in the control of traits can be approached in a number of ways. Information on the physiology of the trait can be used to identify the biochemical pathways involved and hence postulate the controlling genes. This information can be coupled with patterns of expression among tissues to facilitate cloning of the gene(s) likely to affect the trait. These "candidate" genes are then studied in the context of the trait to identify if they play a role in controlling the observed variation. This approach clearly requires a good *a priori* knowledge of the trait and its physiology, but even so it could result in important genes being missed if they are not obviously involved in the known physiology. The second approach is to start with no prior assumptions regarding the physiology or genes controlling the trait, and to use a genome-mapping approach. The mapping approach is often necessary, particularly with complex traits where the various gene interactions cannot be predicted. In the same way as building a genetic map, mapping the genes that control particular traits requires families in which to track the inheritance of chromosomal regions. These families have to be segregating for the trait of interest and information quantifying the trait has to be collected. The mapping process involves tracking inheritance of chromosomal regions using DNA markers and then correlating inheritance of the markers with inheritance of characters associated with the trait. In practice, identification of the "trait genes" is achieved using a combination of the genetic mapping and candidate gene approaches.

4.1 Mapping Quantitative Trait Loci (QTL)

Production associated traits in cattle usually display continuous variation, such as growth or milk yield, and are called "Quantitative Traits." Variation in these traits is controlled by several loci, called *quantitative trait loci* (QTL). Two approaches have been adopted for mapping QTL, one is to make use of commercial cattle herds and the second is to use specifically bred cattle populations to examine particular traits. The advantage of using commercial populations is that they already exist and that simple measurements are recorded at a national level. In contrast, the resource population has to be specially bred;

however, this provides the opportunity of starting from diverse founders that allows the genetic and phenotypic variation in the trait(s) of interest to be maximized. Access to experimental resource herds also extends the possibility of recording the "difficult to measure" traits such as feed-conversion efficiency or immune response to pathogens, which would be impossible in a commercial context. The choice of using commercial or resource populations is to a large extent dependent on the trait that is being considered.

4.2 Dairy Traits

Both of the approaches described above, using candidate gene or genome-mapping approaches have been used to look for genes controlling traits that are important for milk production. The candidate gene approach has focused on the genes coding for the constituent proteins in milk. Milk is mainly water, typically around 88%, containing about 5% lactose, 3.5% fat, and 3.5% protein together with a number of minor components. Of the protein fraction, 80% are the caseins, which precipitate from milk at low pH and form the curd. The remaining whey proteins include β-lactoglobulin, α-lactalbumin and lactoferrin. There are 4 casein genes in cow, α_{s1}, α_{s2}, β and κ, which are clustered within a 25-kb locus on chromosome 6 (6q31), shown in Fig. 8.

Given the high proportion of casein forming the milk protein fraction, and that they are the chief constituents of the curd, which is used in cheese making, the casein genes have been extensively studied. The role of κ is in stabilization of casein micelles in liquid milk and cleavage of κ-casein (CSN3) leads to precipitation

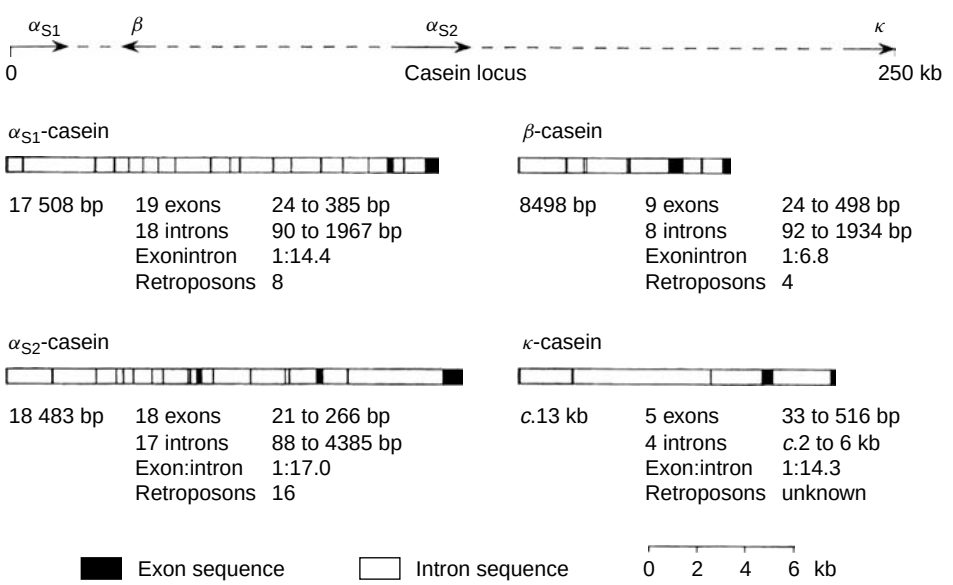

Fig. 8 Schematic representation of the bovine casein gene depicting the exon structure as well as their organization within the casein locus (from Fries, R., Ruvinsky, A. (1999) *The Genetics of Cattle*, CABI Publishing, Oxford, UK).

of the caseins in cheese production. Two alleles of the CSN*3* gene have been described, of which the B allele has been associated with improved manufacturing properties of milk. The B allele has also been associated with increased milk production, but only in Holstein cattle and not in other breeds. Determining the various alleles at each of the 4 casein loci allows casein haplotypes to be constructed, some of which have been associated with increased milk production in other breeds. This may suggest that genes linked to the casein locus, rather than the casein genes themselves are responsible for the variation. Additional interest has been focused on polymorphisms in the β-casein gene (CSN*2*), where there has been the suggestion that the A2 allele may be associated with protection from diabetes.

Selection for improved milk production has been achieved through progeny testing, which has resulted in milk yields in the Holstein breed nearly doubling in the past 40 years to 10 000 liters per 305-day lactation today. In developed countries, the Holstein represents the most numerous cattle breed and accounts for the majority of milk produced. The consequence of intensive selection has been the narrowing of the genetic base of the population and the wide-scale use of a relatively small number of "elite" sires, many of which are closely related. As a result, the dairy industry has very large half-sib families in commercial herds sired by a limited number of sires. While selection for milk yield has produced, in the Holstein breed, cattle with dramatically increased milk yield, there have been an associated decrease in protein content of the milk, decreased fertility, and increased incidence of lameness.

While resource herds have been used to address dairy associated traits, so far few of these results have been published. Most of the data available on QTL involved in dairy associated traits has come from studies on the commercial dairy population, the structure of which, as discussed above, is well suited to genetic mapping studies. However, the range of traits that is routinely recorded in commercial herds is limited, and the focus is on simple traits such as milk yield plus protein and fat composition, and somatic cell scores, the latter being related to the incidence of mastitis. In some cases, "linier type" (conformation) traits are also recorded. The major issue to be considered when addressing improvement in dairy traits is that they are often sex specific, and that the production animals are female, whereas selection is applied to the males. Thus, the industry requires females that produce large quantities of milk, have healthy udders and high fertility, while selection is on the bulls that display none of these traits. Genetic mapping studies could be carried out by tracking the inheritance of chromosomal regions from sires into their daughters, on which the traits are recorded, in a so-called daughter design study. However, this approach is inefficient for several reasons: firstly, a considerable number of daughters would need to be genotyped with DNA markers to track inheritance of the chromosomal regions from a single sire, and secondly, the sire has to be segregating for polymorphisms both in relevant genes and also for phenotypic variations in the trait, for many sires this will not be the case. An alternative approach, proposed by Weller et al., is the "Granddaughter design study" where a small number of males (bulls) are used to produce a large number of sons. DNA markers are used to track the chromosomal inheritance between the bull and his sons. The sons in their turn are used to sire a large number of daughters, on which the production traits are

measured. The phenotypic measurements on the daughters are then used to calculate the genetic merit of their sire (the son of the elite sire). This is the basis of the progeny test used to select breeding sires. By correlating DNA marker information with the predicted breeding values, trait genes can be mapped.

The use of data collected on commercial herds to map QTL was pioneered by Georges et al., who used the US Holstein population to map QTL involved in milk yield and quality. The study used 1518 progeny tested bulls with production data from over 150 000 cows to derive their performance information. The sires were genotyped with 159 microsatellite loci, and analysis of the phenotypic and DNA marker information allowed 5 QTL to be identified, on chromosomes 1, 6, 9, 10, and 20. Each of the QTL was involved in different combinations of the traits examined, namely: milk yield, fat, and protein percentage, and fat and protein yield. As each of the QTL was identified in a subset of the sire families, it is unclear if these results indicate that the different QTL affect different traits, or whether the differences are because each of the sires and grand sires was segregating for different combinations of the QTL. Subsequent studies in other populations have independently confirmed one or more of the QTLs from the Georges study. An additional QTL for milk yield and composition has been reported on chromosome 14, which has been fine mapped and the likely trait gene identified (see DGAT1 below).

4.3 Beef Traits

The structure of the beef industry is different from the dairy industry in a number of ways. Beef is produced from a large number of breeds, which have not been under as intensive selection as seen with the dairy breeds. The conditions under which beef cattle are raised also vary considerably, maintaining the need for diversity of production animals. Breeding in beef units is often by natural service, and where AI, is used the number of progeny produced per bull is considerably fewer than seen with dairy breeds. There is some systematic recording of growth, fat, and conformation traits in a number of countries providing the opportunity for QTL mapping for beef production traits. However, the majority of published information on QTL for beef-associated traits comes from specifically bred resource herds. Several resource herds have been created worldwide, including herds at Texas A&M University (USA) and CSIRO (Australia) that have been produced by crossbreeding between *B. indicus* versus *B. taurus* to produce an F2 population to address meat quality traits. *Bos indicus* cattle, while more tolerant of drought, produce meat that is typically tougher than that of *B. taurus* breeds. Work by the Meat Animal Research Center in Nebraska using crosses between several breeds, has mapped QTL for several beef-associated traits, including growth rate, muscle mass, fat, and so on. A dairy x beef (Jersey vs Limousin) F2 resource herd has been bred in Adelaide (Australia) to specifically address traits associated with beef production. Progeny from this herd have also been raised in New Zealand, which will enable QTL affecting production in *intensive* systems (Australia) to be compared with those that are relevant in *extensive*, grass-based systems (New Zealand). Results of this study are due to be published. A conceptually similar herd, produced as F2 and reciprocal cross between dairy and beef breeds

(Holstein and Charolais) at the Roslin Institute, is also addressing the identification of QTL involved in beef production; however, in addition, an extensive portfolio of traits including health and fertility traits is being recorded on this population. This approach will allow pleiotropic interaction between traits and QTL to be addressed.

4.4 Health Traits

Whereas one of the primary drivers in natural selection will be health and resistance to disease challenge, the intervention of man in the selection of livestock species has, up to now, ignored health traits as selection criteria in favor of productivity traits. In some cases, productivity is dependent on health; however, with the increasing dependency on veterinary interventions, the natural defense mechanisms of the individual have been largely ignored. With the increasing demand from consumers for naturally produced products, together with the break down in effective antibiotic treatment caused by drug resistance in pathogens, there is an increased need for selection of animals with improved natural resistance to disease.

Much attention has been focused on the bovine major histocompatibility complex (MHC) as a potential locus that regulates immune response. The MHC region in cattle, in common with other mammals, contains genes coding for the class I and class II MHC molecules. T cells, that are responsible for eliciting an immune response do not recognize foreign antigens in their native form, but require specialist cells (B-cells, macrophages etc.) to process the antigen and present it bound to MHC molecules on their surface before a specific immune response is triggered. Variations in binding affinity of the antigens to the MHC molecules results in variations in the active immune response directed toward that antigen: failure of the antigen to bind to the MHC molecule means that it is not presented to T cells in the correct context and thus no immune response ensues. The bovine MHC (BoLA – bovine lymphocyte antigen locus) is located on chromosome 23 and in addition to the MHC antigens the region codes for 30 to 40 other genes including several immunologically active proteins, such as tumor necrosis factor and proteins of the complement system. There are two distinct classes of MHC antigen. The class I antigens are composed of two chains, a class I heavy chain that is encoded within the MHC region and β_2-microglobulin whose coding gene is located elsewhere in the genome. The class I molecules present antigen to cytotoxic T cells whose function is to kill infected cells, for example, with virus. There are around 20 MHC class I genes spanning about 1.5 Kb within the BoLA region. The class II molecules are composed of α- and β-chains, both encoded by genes within the MHC region. Several class II genes exist in cattle that resemble their counterparts in other mammalian species, namely, *DQ* α and β, and *DR* α and β genes, also DYA, and the class II-like chaperone molecule DMA. In the cattle, the *DR*β gene has been duplicated in some haplotypes, while the equivalent of the human, the DP locus appears to be absent in the BoLA region. The class II genes are involved in presentation of antigen to T-helper cells that promote both cellular and humoral immune responses. Given the central role of the MHC molecules in promoting immune response, it is not surprising that associations between MHC alleles and immune response to various infectious diseases and parasites have been sought.

The most common disease in dairy cattle is mastitis, which is an infection of the udder, and can be caused by several pathogens. Mastitis is both a welfare and an economic problem, resulting in decreased milk yield, increased veterinary treatments and is the most common cause of death in dairy cattle. The incidence of mastitis is not generally recorded, except in Scandinavian countries where a national health-recording scheme keeps individual records of disease for every cow. Studies using the Scandinavian data have identified putative QTL on chromosomes 3, 4, 14 and 27 associated with mastitis incidence. More commonly, somatic cell scores (SCS) are used as an indicator of mastitis incidence, although SCS are a measure of inflammatory response, and do not necessarily indicate clinical or subclinical infection. Consequently, the correlation between clinical mastitis and SCS range from 0.37 to 0.97. However, this is a measure that is routinely recorded in commercial dairy herds. Studies to examine associations between genetic loci and SCS suggest that some MRC DRβ alleles may be associated with increased SCS and hence increased risk of mastitis.

Breeds of cattle are known that show resistance to particular parasites, for example, the N'Dama cattle of West Africa are resistant to *Trypanosoma congolense*, which causes sleeping sickness, whereas other breeds of cattle are highly susceptible to this disease. The introduction of more productive breeds into endemic areas is currently not a feasible option, as the animals would either die or be too expensive to maintain through requirement for extensive veterinary treatment. However, the disease resistant breed is small and not very productive. One approach to allow the introduction of more productive breeds in to infected areas is to identify the genes controlling the disease resistance and introduce these resistance genes into more productive animals. In order to identify the genetic control of *T. congolense* infection, an F2 cattle resource population was bred at ILRAD (now ILRI) in Kenya by crossing N'Dama with Boran cattle, the latter being susceptible to trypanosomiasis. This study has identified QTL to be involved in resistance; and fine mapping of these loci is being pursued in a mouse model.

A new cattle disease was identified in the United Kingdom in 1986, called *bovine spongiform encephalopathy* (BSE). The BSE epidemic developed rapidly to peak in 1992. Today, more than 180,000 cases have been confirmed in the United Kingdom. In sheep incidence of the related transmissible spongiform encephalopathy, scrapie, is associated with polymorphisms in the *PrP* gene, at codons 136, 154, and 171. Sheep that are homozygous for glutamine and position 171 are most susceptible to scrapie, while individuals with arginine are most resistant. In cattle, two polymorphisms have been reported in the coding region of the *PrP* gene, a silent change that affects a *Hind*II restriction site, and a difference in the number of G-C rich octa-repeat elements, with alleles that have either 5, 6, or 7 copies of the repeat. Two case-control studies found no association between variations in the octa-peptide-repeat and incidence of BSE. Complete sequence information for a 25-kb region containing the bovine *PrP* gene revealed 9 polymorphisms in the PrP coding region, including the two mentioned above, and more than 40 SNPs and insertions in noncoding regions of the gene. However, there is no evidence to date for polymorphisms within the bovine *Prnp* coding region that affect susceptibility to BSE. A whole genome scan with microsatellite

markers that examined transmission disequilibrium of alleles between BSE affected and control progeny bred from a small number of bulls revealed three loci, on chromosomes 5, 10, and 20 that were associated with incidence of BSE. The interval on bovine chromosome 10 coincides with the corresponding interval on mouse chromosome 9 that has been associated with resistance of mice to experimental scrapie and is now being investigated further.

4.5 Identifying the Trait Genes

The ultimate objective of the genome-mapping work is to identify the genes and variations within the genes that control the traits of interest. Up to now a reasonable number of genes controlling a diverse range of traits have been localized to fairly broad regions on many bovine chromosomes. Typically, linkage-mapping studies in cattle have identified, at best, a 20-cM interval (the whole bovine genome is 3000 cM) within which the gene affecting the trait of interest is located. This size of the interval represents 20 Mbp of DNA and, assuming a conservative estimate of around 30 000 genes in the bovine genome, is likely to contain at least 200 and possibly as many as 1000 genes. There are several ways of identifying the gene involved in controlling a trait from its map location; these include fine mapping using meta-analysis of data from different populations to refine the map location, examination of haplotypes to identify the limits of linkage disequilibrium between markers and the trait, positional cloning using large fragment clones to build contigs across the QTL region, and a positional candidate gene approach, where genes likely to be involved in the trait are identified within the QTL region. In conjunction with these approaches, comparative information can be used, taking information about the position and role of candidate genes from other species and using knowledge of conservation of synteny to identify candidate genes likely to be located within the QTL region in cattle.

Two studies have been published that are paradigms for identification of genes involved in meat and dairy traits in cattle. Both these studies used a mixture of linkage mapping, positional cloning, and comparative genomics to identify the gene controlling double muscling, and a gene for milk production in cattle.

Double muscling is an extreme form of muscle development that has been described in several breeds of cattle; the most extreme form of double muscling being seen in the Belgian Blue breed. Segregation patterns for the trait suggested that the phenotype was controlled by a single gene, which was subsequently mapped to chromosome 2 using a resource herd bred from double-muscled Belgian Blue cattle crossed to nondouble muscled cattle. The first approach for identifying the gene controlling the phenotype, starting from the map position, was to test candidate genes located close to the region identified from the mapping study. However, this approach failed to identify the trait gene. At the same time, unconnected work on mice identified a new member of the growth and differentiation factor β superfamily of genes (GDF-8) that was expressed in muscle and that was required for normal muscle development. Knocking-out the expression of GDF-8 resulted in hyper-muscularity in the mice and therefore was named myostatin. Using RH mapping, myostatin was shown to map to bovine chromosome 2 within the interval where the double-muscling gene had been mapped. Sequencing the

myostatin gene from double-muscled individuals revealed an 11 bp deletion in Belgian Blue cattle that truncates the protein in the bioactive c-terminal domain, and a single nonconservative base change in the double-muscled Piedmontase breed. These mutations are strongly associated with the double-muscled phenotype and make myostatin the likely causative gene.

It has been suggested that double muscling originated in the British Beef Shorthorn breed. However, the myostatin gene in cattle is highly polymorphic, with 19 polymorphisms currently described within the bovine gene. Analysis of the distribution of these polymorphisms has allowed the possible origins of the alleles found in different breeds to be investigated. There is currently no molecular evidence to support the Shorthorn as the origin of the myostatin alleles responsible for double muscling in European cattle. The 19 polymorphisms can be assembled into 10 haplotypes, which are related in a broad phylogenetic tree with a shallow timescale, suggesting that most of the mutations have accumulated independently, and fairly recently. The 11-bp deletion seen in Belgian Blue cattle is the most common mutation responsible for double muscling and it is likely that this allele was introduced into several other breeds from a single source at about the same time. Independent, rather than a stepwise accumulation of mutations then followed.

The effects of the disruptive mutations in the myostatin gene on the phenotype vary between the different breeds. The same 11-bp deletion that is found in the Belgian Blue and which is associated with an extreme form of double muscling, is associated with a milder phenotype in other breed, such as the Spanish Asturiana and British South Devon cattle. This has lead to speculation that modifiers must interact with myostatin in the regulation of muscle development and are partly responsible for the double-muscled phenotype.

A QTL with effect on milk production, particularly fat content was localized to chromosome 14 in cattle using a linkage-mapping approach in several commercial populations. The fact that the same QTL location was identified independently gives confidence both in the existence of a locus with significant effects on milk synthesis, and also on the position of the QTL. By examining the extent of linkage disequilibrium on individual chromosomes, the QTL region was narrowed. Individual recombination events between marker and the phenotype allow most of the mapped QTL region to be discounted as the location of the gene, so that the trait gene could be localized to within 3 cM, representing 3 Mb of DNA. A BAC contig spanning this region was assembled in preparation for local sequencing to identify genes within the region.

Work carried out in parallel in mice identified diacyl-glycerol *O*-acyltransferase (DGAT1) as an enzyme involved in the formation of triglyceride, which is a major component of the fat found in milk. Studies in mice also showed that females lacking both copies of DGAT1 had impaired lactation. DGAT1 was mapped within the QTL region on chromosome 14 in cattle by RH mapping and also mapped with the refined 3 cM interval on the BAC contig, making it a very strong positional candidate gene. Sequencing the DGAT1 gene revealed a polymorphism at position 232, which substitutes alanine for lysine (K232A). Individuals with lysine have a higher milk production than those with alanine.

5
Functional Genomics

Having a detailed genome map, even the ultimate map in the form of the sequence, of the genome, is only the starting point for identifying the genes controlling traits in cattle and understanding their function. The next step will be to characterize the expression of the genes, and the functions of their protein products, in order to understand their role in controlling development and survival of the individual and the variations seen among individuals.

Advances in technology mean that we can now examine the expression of large numbers of genes simultaneously, using a macro- or microarray approaches. These arrays are created using either cDNA probes, or oligonucleotide probes designed from knowledge of the gene sequence. The probes are anchored to a solid matrix, a nylon membrane in the case of the macroarray, or a glass slide for the microarray. Complex probes are then created by reverse transcribing and labeling the RNA from the cells or tissues whose gene expression is being studied. The labeled probe is then hybridized to the arrays and expression patterns deduced, either by comparison with a reference, or in the case of microarrays by simultaneous hybridization of the samples being compared, each of which is labeled with a different fluorescent dye. The prerequisite for these arrays are the gene probes that are used. Although specialist oligonucleotide probes have been made for particular genes, for example, those involved in immune responses, currently most array probes are cDNAs that have been identified by sequencing clones from cDNA libraries produced from the target tissue or tissues. The data generated from the sequencing cDNA clones produces expressed sequence tags (ESTs). At the time of writing, there are over 122 000 cattle ESTs sequences recorded on the EMBL sequence database, mainly generated by the USDA Meat Animal Research Center in Nebraska. These arrays will enable gene expression to be explored in a range of tissues in different physiological states and from animals with variable phenotypes. This information will provide an extension to the genome-mapping studies to help identify the genes involved in the control of traits, and to understand their function

6
In Conclusion

The construction of maps for the bovine genome has allowed the chromosomal regions harboring genes controlling specific traits to be identified, and in some cases the genes themselves have been found using a variety of techniques, starting from their map position. Comparative mapping information is allowing the alignment of bovine chromosomes with high-density maps in other species, and with the human genome sequence. This will provide considerable information on the range of genes present in the genome and their location in cattle. In the foreseeable future, the sequence of the bovine genome will be available. Analysis and annotation of the bovine sequence along with the sequence from other species will, in many cases, confirm the predictions made from the comparative mapping. Differences observed between genomes will also lead to new findings with the identification of new genes and a greater understanding of the regulation and expression of the genes and their role in controlling phenotypic variation.

See also Genetics, Molecular Basis of; Immunoassays.

Bibliography

Books and Reviews

Fries, R., Ruvinsky, A. (1999) *The Genetics of Cattle*, CABI Publishing, Oxford, UK.

McCarthy, L.C. (1996) Whole genome radiation hybrid mapping, *Trends Genet.* **12**, 491–493.

Ng-Kwai-Hang, K.F. Grosclaude, F. (1992) Genetic polymorphism of milk proteins, in: P.F. Fox. (Ed.) *Advanced Dairy Chemistry* Vol. 1 Elsevier, Amsterdam pp. 405–455.

Primary Literature

Band, M.R., Larson, J.H., Redeiz, M., Green, C.A., Heyen, D.W., Donovan, J., Windish, R., Steining, C., Mahyuddin, P., Womack, J.E., Lewin, H.A. (2000) An ordered map of the cattle and human genomes, *Genome Res.* **10**, 1359–1368.

Barendse, W., Vaiman, D., Kemp, S., Sugimoto, Y., Armitage, S., Williams, J.L., Sun, H., Eggen, A., Agaba, M., Aleyasin, A., Band, M., Bishop, M., Buitkamp, J., Byrne, K., Collins, F., Cooper, L., Coppettiers, W., Denys, B., Drinkwater, R., Easterday, K., Elduque, C., Ennis, S., Erhardt, G., Ferretti, L., Flavin, N., Gao, Q., Georges, M., Gurung, R., Harlizius, B., Hawkins, G., Hetzel, J., Hirano, T., Hulme, D., Joergensen, C., Kessler, M., Kirkpatrick, B., Konfortov, B., Kostia, S., Kuhn, C., Lenstra, J., Leveziel, H., Lewin, H., Leyhe, B., Li, L., Martin Burriel, I., McGraw, R., Miller, R., Moody, D., Moore, S., Nakane, S., Nijman, I., Olsaker, I., Pomp, D., Rando, A., Ron, M., Shalom, A., Soller, M., Teale, A., Thieven, I., Urquhart, B., Vage, D-I., Van de Weghe, A., Varvio, S., Velmalla, R., Vilkki, J., Weikard, R., Woodside, C., Womack, J., Zanotti, M., Zaragoza, P. (1997) A medium density genetic linkage map of the bovine genome, *Mamm. Genome* **8**, 21–28.

Blott, S.C., Williams, J.L., Haley, C.S. (1998) Genetic relationships among European cattle breeds, *Anim. Genet.* **29**, 273–282.

Bradley, D.G., Loftus, R.T., Cunningham, P., McHugh, D.E. (1998) Genetics and domestic cattle origins, *Evol. Anthropol.* **6**, 79–86.

Dunner, S., Miranda, M.E., Amigues, Y., Cañon, J., Georges, M., Hanset, R., Williams, J.L., Ménissier, F. (2003) Haplotype diversity of the myostatin gene among beef cattle breeds, *Genetic Selection and Evolution* **35**, 103–118.

Ferretti, L., Urquhart, B.G.D., Eggen, A., Olsaker, I., Harlizius, B., Castiglioni, B., Mezzelani, A., Solinas-Toldo, S., Thieven, U., Zhang, Y., Morgan, A.L.G., Teres, V.M., Schwerin, M., Martin-Burriel, I., Chowdary, B., Erhardt, G., Nijman, I.J., Cribiu, E.P., Barendse, W., Leveziel, H., Fries, R., Williams, J.L. (1997) Cosmid-derived markers anchoring the bovine genetic map to the physical map, *Mamm. Genome* **8**, 29–36.

Fries, R., Eggen, A., Womack, J.E. (1993) The bovine genome map, *Mamm. Genome* **4**, 405–428.

Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A.T., Sargeant, L.S., Sorensen, A., Steele, M.R., Zhao, X., Womack, J.E., Hoeschele, I. (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing, *Genetics* **139**, 907–920.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Ried, S., Simon, P., Spelman, R., Georges, M., Snell, R. (2002) Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition, *Genome Res.* **12**, 222–231.

Hernandez-Sanchez, J., Waddington, D., Weiner, P., Haley, C.S., Williams, J.L. (2002) Genome-wide search for markers associated with bovine spongiform encephalopathy, *Mamm. Genome* **13**, 164–168.

Kappes, S.M., Keele, J.W., Stone, R.T., Sonstegard, T.S., Smith, T.P.L., McGraw, R.A., Lopez-Corrales, N.L., Beattie, C.W. (1997) A second-generation linkage map of the bovine genome, *Genome Res.* **7**, 235–249.

Klungland, H., Sabry, A., Heringstad, B., Olsen, H.G., Gomez-Raya, L., Vage, D.I., Olsaker, I., Odegard, J., Klemetsdal, G., Schulman, N., Vilkki, J., Ruane, J., Aasland, M., Ronningen, K., Lien, S. (2001) *Mamm. Genome* **12**, 837–842.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C. et al (2001) Initial sequencing and analysis of the human genome, *Nature* **409**(6822), 860–921.

Lien, S., Gomez-Raya, L., Steine, T., Fimland, E., Rogne, S. (1995) Associations between casein haplotypes and milk yield traits, *J. Dairy Sci.* **78**, 2047–2056.

Loftus, R.T., MacHugh, D.E., Bradley, D.G., Sharp, P.M., Cunningham, E.P. (1994) Evidence for two independent domestications of cattle, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2757–2761.

McPheron, A.C., Lee, S-J. (1997) Double muscling in cattle due to mutations in the myostatin gene, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12457–12461.

Popescu, C.P., Long., S., Riggs, P., Womack, J., Schmutz, S., Fries, R., Gallagher, D.S. (1996) Standardization of cattle karyotype nomenclature: Report of the committee for the standardization of the cattle karyotype, *Cytogenet. Cell Genet.* **74**, 259–261.

Sharif, S., Mallard, B.A., Wilkie, B.N., Sargeant, J.M., Scott, H.M., Dekkers, J.C.M., Lesley, K.E. (1998) Associations of the major histocompatibility complex DRB3 (BoLA-DRB3) alleles with occurrence of disease and milk somatic cell score in Canadian dairy cattle, *Anim. Genet.* **29**, 185–193.

Solinas-Toldo, S., Lengauer, C., Fries, R. (1995) Comparative Genome map of human and cattle, *Genomics* **27**, 489–596.

Stone, R.T., Keele, J.W., Shackelford, S.D., Kappes, S.M., Koohmaraie, M. (1999) A primary screen of the bovine genome for quantitative trait loci affecting carcass and growth traits, *J. Anim. Sci.* **77**, 1379–1384.

Weller, J.I., Kashi, Y., Soller, M. (1990) Power of daughter and grand-daughter designs for determining linkage between markers and quantitative trait loci in dairy cattle, *J. Dairy Sci.* **73**, 2525–2537.

Williams, J.L., Eggen, A., Ferretti, L., Farr, C., Gautier, G., Amati, G., Ball, G., Caramori, T., Critcher, R., Costa, S., Hextall, P., Hills, D., Jeulin, A., Kiguwa, S.L., Ross, O., Smith, A.L., Saunier, K.L., Urquhart, B.G.D., Waddington, D. (2002) A Bovine Whole Genome Radiation Hybrid Panel and Outline Map, *Mamm. Genome* **13**, 469–474.

Winter, A., Kramer, W., Werner, F.A.O., Lokkers, S., Kata, S., Durstewitz, G., Buitkamp, J., Womack, J.E., Thaller, G., Fries, R. (2002) Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9300–9305.

Living Organism (Animal) Patents

William Lesser
Department of Applied Economics and Management, Cornell University, Ithaca, NY, USA

1 Introduction 369

2 Concept and Operation of Patent and Related Systems 369
2.1 Origins of Intellectual Property Rights 369
2.2 Purposes of Intellectual Property Rights 370
2.2.1 Economic Justification 370
2.2.2 Moral Justification 371
2.3 Forms of Intellectual Property Rights Applicable to Animals 371
2.3.1 Patents 371
2.3.2 Plant Breeders' Rights 372
2.3.3 Trade Secrets 374

3 Issues with Protecting Living Organisms 374
3.1 Developments in the United States 374
3.2 Developments in Europe 375
3.3 Legislation in Other Leading Countries 376
3.4 International Agreements 377

4 Issued Patents 378
4.1 Animal Patents in the United States 379
4.2 Animal Patents at the EPO 380

5 Issues with Patented Animals 380
5.1 Status of Technology 380
5.2 Benefits 381
5.2.1 Animal Models 382
5.2.2 Production of Compounds 382
5.2.3 Other Applications 382

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

5.3 Areas of Concern 382
5.3.1 Access to Patented Animals 383
5.3.2 Environmental Safety 383
5.3.3 Health 384
5.3.4 Food Safety 384
5.3.5 Ethics 385

6 Public Attitudes 386
6.1 Attitudes in the United States 386
6.2 Attitudes in Europe 388

Bibliography 388
Books and Reviews 388
Primary Literature 389

Keywords

Animal Patent
The application of patent law to the protection of higher animals.

Intellectual Property Rights (IPR)
Legal systems for providing property protection for abstract ideas and knowledge, the systems generally including patents, trademark, and copyright, among others.

Patent
Instrument covering the grant of limited exclusive property rights to inventors; also utility patent.

Transgenic Animal
An animal to whose DNA (hereditary material) has been added DNA material from another source other than parental germplasm. Patented animals are typically transgenics.

Animal Patents refers to the granting of utility patents to higher animals. The application of patents to animals is a recent development in patent law with the first formal decision (in the United States) dating to only 1987. This extension of patent law raises for some legal, practical, and moral issues, yet others see patent protection as an essential incentive for the development of specialized animals for uses in medical research, medicine production, agricultural/aquaculture uses, and even pets. By 2004, some 600 animal patents had been granted worldwide, 80% in the United States, and most relating to ''animal models'' for medical research. Indeed, the first animal patent in 1987, for the ''Harvard'' mouse, was for cancer

research. The World Trade Organization agreement, while specifying minimal Intellectual Property Rights protection allowable in member states, does specifically allow countries to exclude patents for animals while offering no alternative protection systems. Hence, besides the United States and European Union (EPO members), only three countries presently allow patents for animals. Animal patents are likely to remain geographically limited for some time, but scientific advances and commercial realities will eventually lead to pressures on governments to grant protection. Public opinion surveys show the best acceptance for patented animals providing clear benefits to humans.

1 Introduction

Animal Patents refers to the application of patents to higher animals. Patents are a means of granting property rights and, by providing limited monopoly control, an incentive for investing in inventive research and development (R&D). Patents, a form of Intellectual Property Rights (IPR), have been found to provide an especially significant incentive for easily copied inventions, which includes self-reproducible animals. As a background, the general functioning of patent systems (known more formally as utility patents) is described in Sect. 2. Also included is the description of a specialized patentlike system for plants only, which, while not directly applicable to animals, does allow some insights into the special considerations and approaches for applying IPR to self-reproducing life forms.

While, at the most basic level, there is nothing conceptually different about applying patents to animals, their use does raise two fundamental issues:

- Are animals discoveries or products of nature and hence excludable from patents?
- Should ethical/moral considerations for patenting animals exclude patentability?

The focus here is on issues associated with patented higher animals, as contrasted with other economic and social matters applied to animals in general. As a practical matter, this means the animals that have been genetically transformed. Methods are important, but only to the extent they affect the production of transgenic animals. Cloning, for example, is a method that can be applied to transgenic animals but is not associated with the animals themselves. Also, while there is no specific bar to the patenting of traditionally bred animals and, indeed, plants produced through conventional breeding approaches are routinely granted patents, as yet grants for animals have been limited to transgenics.

2 Concept and Operation of Patent and Related Systems

2.1 Origins of Intellectual Property Rights

The earliest form of Intellectual Property Rights has been traced by many scholars to Venice in the Middle Ages. Master craftsmen, it seems, were beset with competition from former apprentices. To control that

competition, a law was passed that prohibited former apprentices from going into competition with licensed craftsmen for a period equivalent to the period of the apprenticeship. That period, of about 20 years, is said to be the origin of the current 20-year patent duration from first application. Certainly, craftsmen benefited from reduced competition. Apprentices benefited less so, but without the protection from competition, many may not have found masters willing to take on apprentices. Finally, the consuming public lost in the short run through higher prices, but with the long-term assurance of trained master craftsmen available to produce the products, eventually they would benefit as well.

2.2 Purposes of Intellectual Property Rights

2.2.1 Economic Justification

Over subsequent centuries, patent and related IPR laws have undergone multiple changes, but the essence of the public/private trade-off seen in long-ago Venice is still evident. Patents are temporary (typically 20 years). The inventor benefits from that period of reduced competition and thereby has an incentive to invest in inventive activities. The public benefits from the existence of products and processes that otherwise may not have existed, or whose appearance would have been delayed. The justification recognizes the copier, who does not have to recover the R&D expenses of the inventive process, can always undersell the inventor. Hence, no one outside the public sector has an incentive to invest in easily copied new products, and technological advances stagnate. Patent laws restore that incentive through the legal prohibition on direct copying.

Contemporary patent law adds a further dimension, the requirement that the invention be described. The resultant file of descriptions creates a public storehouse of technical information while helping assure competitive products can be produced soon after the patent expires. For these benefits, the public pays a higher price than would prevail in a competitive market. A higher price can come about only if supply is reduced so that the public has less of the product available than under competitive conditions. Stated differently, product use diffuses more slowly.

In addition to the mere existence of the novel product or process, the public benefits from the so-called spillover effect. Spillover recognizes that a private owner cannot capture all the benefits from a product. A large piece of property kept undeveloped by an owner not only benefits the owner but also the public, which enjoys the views and ecological benefits provided. Similarly, an invention (gasoline engine) can lead to derivative inventions (automobiles), which further benefit the public. Or the inventor, to meet unit sales and profit goals, may need to charge a lower price than many users would be willing and able to pay. Those users gain a direct benefit at the expense of the inventor.

The investment incentive is the principal justification for patents and related IPR. There is, however, another justification for patent-type laws: the provision of an incentive for firms to distribute products to other countries. Key to appreciating this aspect is a realization that intellectual property law is strictly national: protection must be held in each country to be effective. For example, a product protected in the United States has no protection in France or Japan unless specifically sought and received there. Inventors

would be concerned about unauthorized copying in those markets as well, and are typically hesitant to supply them without the corresponding intellectual property protection. Hence, intellectual property protection can assist in the transfer of products and technologies.

Patents belong to a group of related rights known as *industrial property rights*, which also include Plant Breeders' Rights, trademarks, and trade secrets. Trademarks function somewhat differently, but with a general similarity of intent. Trademarks allow owners to develop goodwill in a product with the knowledge that it will not be dissipated by unauthorized imitators. For the consumer, a trademark serves as a signal of product quality – a Coca Cola or McDonald's hamburger has a consistency and standard taste consumers expect. True Rolex watches are generally of far higher quality than the copies. Thus, with trademarks too, both the owner of the trademark and the public benefit at least to some degree.

Copyrights fall into a separate class of protection mechanisms. Copyrights are applied to creative works like books and music, although important current uses include software. Copyrights are more of a protection by right and involve fewer formal procedures.

Trade secrets impose a penalty if secrets are acquired in an inappropriate way. The secret itself may be anything of value, such as customer lists or performance records. No formal process is required to achieve trade secret status, beyond a serious effort to keep the secret that way. In that way, a trade secret can conceivably be perpetual. The formula for Coca Cola is a well-known trade secret. In many cases, firms will achieve stronger (broader) protection by combining several forms of protection within a single product. For example, a patent might protect the product, but the least cost process for producing it is protected by a trade secret. Or a product is patented while the name is trademarked.

2.2.2 Moral Justification

A second justification for IPR is a moral, or perhaps more correctly, personal rights justification. This concept, often couched in the individualism philosophy of the Enlightenment, stresses that creation (whether artistic or technical) is from the self and hence should be granted the same rights as any personal property. IPR and personal rights are aspects of the self and hence should be granted as a right. Here, the economic justification will be stressed, as it is generally. The moral right concept, though, does continue in the idea of benefit sharing as well as requirements in most patent laws that the inventor(s) be identified by name. Perhaps, the inventor does not own the invention (employment contracts, for example, typically make any invention in the course of work the property of the employer). However, he/she/they must be at least identified on the patent as the inventor.

2.3 Forms of Intellectual Property Rights Applicable to Animals

The preceding subsection laid out the general justification for IPR and identified in broad terms the major forms. This section contains more detailed descriptions of the operation of the several systems potentially applicable to animals. Greatest attention is of course given to patents.

2.3.1 Patents

Patents are intended as an incentive and reward for developing something new and

useful, but only if the development is a nontrivial extension. The system has developed specific terms and interpretations for these concepts.

Nonobviousness or inventive step: The invention should not be obvious to someone "skilled in the art." This is the nontrivial requirement.

Novelty: The invention must be new in the sense of not having been previously known through publication or public use. The United States uses a system with a one-year "grace period," which means simply that the invention could have been revealed up to one year prior to applying for a patent. Most countries do not allow any prior revealing of the invention – in what is known as absolute novelty. Absolute novelty, in particular, is difficult for public sector researchers whose work often requires publishing of results.

Utility: Utility requires that the invention must serve some identified purpose. This does not say the invention must be practical – patent offices do not judge practicality. Rather, the requirement is that a use be identified. This is an important component of the system for it led to the rejection of applications for human genes early in the Human Genome Project. Those applications read to a gene only, with no specific knowledge of a use for it. Hence, the rejection.

Issued patents are all of the same general type, although this masks some large differences. One is the *scope* of an individual patent – a wider scope means a related product must be more different so as not to be infringing. The scope of the first animal patent, the "Harvard" mouse, is quite broad, reading in the first claim, "all non-human mammals" that exhibit the identified trait. Often, patents in new technical fields are broad, narrowing as the techniques become more standardized. Typically, a person knowledgeable in the field is needed to interpret the scope of a patent.

A second distinction is made on whether the patent may be referred to as *per se*, process, product by process, or dependent. Those are not terms that will appear on the patent, but can be inferred from a close reading. For example, a claim which reads "a means of" is usually a patent for a process. A product by process patent covers the product only if it is produced in the described way. A different production technique lies outside the patent. The concept of *per se* and dependent patents can be explained by a factitious reference to aspirin. Imagine the initial aspirin patent was for headache relief (in fact, it was a natural product with a patent for a chemical synthesis). Because (in this example) it was the first, the patent owner had rights to any subsequently developed uses. One was identified – the use to help prevent and subsequently recover from a heart attack due to the blood thinning property – and patented. The later patent though is dependent on the first one, so it cannot be used without the permission of the initial patent owner. The owner of the initial patent, for his or her part, also cannot use the improvement without permission. Often, what could be an impasse is resolved with a cross-license between the two owners.

2.3.2 Plant Breeders' Rights

Plant variety protection (PVP) is a special purpose system only for plants. Technically, it is a form of *sui generis* or special purpose system owing to the exclusive focus on plants. Plant breeder's rights (PBR) are quite recent, with origins in the 1930s and 1940s. UPOV (the French acronym

for the International Convention for the Protection of New Varieties of Plants), the international convention, was established only in 1961.

PBR replaces utility, nonobviousness, and novelty for patent systems (see above) with uniformity, stability, distinctness, and novelty (abbreviated as DUS). Novelty is the same in concept to patents, although the specific terms and use conditions differ. Uniformity and stability are technical requirements, which assure that the protected variety is identifiable after repeated multiplications. Clearly, if the protected product (variety) cannot be definitively identified, the system breaks down. For the future, genetic markers or a related technology may be used, but that is not the case at present.

Distinctness is akin to nonobviousness for patents. Distinctness may be claimed in one of a number of traits, such as disease resistance or flower color. Most countries actually grow out varieties and measure the difference from a base or reference variety. In the United States, the claims of the applicant are generally accepted, akin to a registration system. Moreover, the US Plant Variety Protection Office allows distinctness for traits of no practical merit. Hence, PBR in the United States are generally considered to be weaker than in other UPOV-member countries.

PVP have other components, which mean that they provide weaker protection than patents. These are the so-called Breeders' Rights and Farmers Privilege. Breeders' Rights specifically allow breeders to use protected varieties in a breeding program without the permission of the variety owner. For patented products, this is known as the *research exemption*. Because the research exemption is not statutory (written in the law), it is a matter of court interpretation, the focus of which is still under debate by scholars and practitioners. The Farmers Privilege expressly allows farmers to retain the crop for use as a seed source for a subsequent crop. The farmer, however, may not sell the seed or share with other farmers. The most recent UPOV Act (1991) makes the Farmers Privilege a national option, but most countries (including the United States) are choosing to allow it. In Europe, with the exception of "small farmers," a payment for seed savings set at 50% of the "normal" license fee must be paid.

PVP are intended to protect the entire plant as well as its propagating parts (seeds, etc.). This is in line with the specific focus of preventing direct competition for that variety. A weakness became apparent when overseas growers (where PVP was unavailable) shipped plant parts such as flower blossoms into a protected market. That was not illegal until the most recent version of PVP legislation was adopted, extending protection to plant parts, and potentially to products of those parts (say oil from a protected soybean variety).

For biotechnological applications though, the current scope of protection is insufficient. Breeders can simply use crossbreeding or other means to transfer a genetically engineered gene construct from one variety (protected) to another (unprotected). The breeders' rights exemption makes such a transfer perfectly legal. A patent plant would provide protection for it would read to the genetically engineered trait, so protection would not be voided through the breeding process. Moreover, the breeding process itself would be a likely infringement. Conversely, PVP covering the variety and a patent for the gene construct provides protection similar to a patented plant.

PVP-like systems have two advantages for animals. They are simpler and less

costly to secure, and hence appropriate for sequential improvements such as livestock breeding. Second, many countries have specifically prohibited patents for plants or animals but are allowing PBR systems. However, a workable system requires identifying the target population – in short, a workable definition for a "variety." This exists for plants, but the relatively small number of progeny and inherent heterogeneity of an animal line (with the possible exception of inbred experimental lines) make a direct extension of PVP to animals inappropriate. One proposal for an "Animal Breeders' Rights" system is described in Sect. 3.4.

2.3.3 Trade Secrets

Since the 1930s, corn breeders have used trade secrets to protect first generation (F-1) hybrids. Since hybrids do not reproduce true-to-type, the sale of seed does not allow the duplication of the seed. Breeders need only protect the pure lines and crossings used to gain *de facto* protection of the resultant hybrid. Hence, breeders typically fence their pure line fields and of course do not reveal the crossing sequence.

For agricultural livestock, a similar approach is used by poultry and hog breeders. There, multiple crosses of pure lines produce synthetics with characteristics similar to plant hybrids. Poultry breeders keep tight control of their pure lines and release grandparent and even parent stock only under careful supervision. Of course, this strategy is not applicable for traditionally crossbred animals.

3 Issues with Protecting Living Organisms

The preceding text describes the general considerations for protecting inventions. Living organisms though bring to bear several additional considerations. This can be inferred from the general recentness of extensions of protection to living matter. An early example is a patent granted to Pasteur in the 1880s for a yeast strain, but the Patent Office seemingly considered the product to be inanimate rather than animate. The initial specialized law was the (US) Plant Patent Act of 1930 (subsequently incorporated into the Patent Act). The Plant Patent Act applies only to asexually propagated plants. Asexually propagated materials are stable, thus ending congressional uncertainty about the identification of protected materials over repeated regeneration.

Subsequently, European interests took the lead in protecting plant varieties, beginning in the 1940s. The first international convention, UPOV, was adopted in 1961 with the most recent version dating to 1991. The United States did not adopt a version, the Plant Variety Protection Act, until 1970, which was amended in 1980 and 1994.

3.1 Developments in the United States

Of major significance was the 1980 Supreme Court decision in *Chakrabarty*. While the particular decision related to a microorganism (for more rapid decomposition of spilled crude oil), the decision was made in sweeping terms. The Court declared, "everything under the sun that is made by man" is patentable subject matter. Shortly thereafter, in 1985, the patent and trademark office (PTO) in an internal decision, *Ex parte Hibberd*, declared (higher) plants to be patentable subject matter. "Patentable subject matter" simply means that the application cannot

be rejected solely because of the area of application.

Interestingly, in 2001, there was a challenge heard by the Supreme Court to the provision of patents for plants (*J.E.M. AG Supply v. Pioneer Hi-Bred*). The challenge began (in typical fashion) as a defense for an infringement charge, declaring there was no infringement for the patent was invalid. The reason? The Congress passed the PVPA to protect plants and hence did not intend to allow for the extension of patents. The Supreme Court strongly endorsed *Hibberd,* declaring that patents and PBR could "mutually coexist."

Two years after *Hibberd*, in another internal decision in *Ex Parte Allen* (1987), the PTO declared higher animals to be patentable subject matter. Interestingly in that case, the underlying application for a "polypoidal" oyster was rejected for insufficient nonobviousness. The first animal patent was not issued until April 1988 for the "Harvard mouse," a laboratory research specimen.

3.2 Developments in Europe

Patent protection for animals (as well as other forms of living organisms) in Europe has been shaped first by the wording in the European Patent Convention (1973) (EPC), which, in Article 53(b), prohibited patents for "plant or animal varieties or essentially biological processes for the production of plants or animals." With an origin dating to an earlier and simpler era, it is now, with the advent of the new biotechnology, no longer easily determinable what is an "essentially biological process." Within Europe, there are two separate routes to patents: through national patent offices or through the European Patent Office (which grants "bundled" national patents). While there are small technical and practical differences between the patents received from the routes, the legislation harmonizing the requirement of the Convention means that the two systems operate in parallel for purposes here.

To clarify the interpretation of the phrase "essentially biological," the European Directive on the Legal protection of Biotechnological Inventions (Directive 98/44/EC, Rule 23(b)(5)) contains the definition:

"A process for the production of plants or animals is essentially biological if it consists entirely of natural phenomena such as crossing or selection."

As a consequence of that Directive, animals (and plants) that have been transformed genetically are patentable subject matter. Relevant Directive articles are as follows:

- Biological materials isolated from its natural environment or produced by means of a technical process may be the subject of an invention even if it previously occurred in nature (Article 6.3).
- The protection conferred by a patent on a biological material shall extend to any biological material derived from that biological material through propagation or multiplication in an identical or divergent form and possessing those same characteristics (Article 8).
- The protection conferred by a patent on a product containing or consisting of genetic information shall extend to all material (except the human body) in which the product is incorporated and in which the genetic information is contained and performs its function (Article 9).
- The protection shall not extend to biological material [] where the multiplication

or propagation necessarily results from the application for which the biological material was marketed, provided that material is not subsequently used for other propagation or multiplication (Article 10).

Note that, as a consequence of the genetic transformation, the products are no longer potentially construed as "discoveries," and hence are specifically unpatentable (EPC Article 52(2)(a)). Stated differently, an animal (or plant) is patentable in Europe (provided other patenting requirements are met) if the invention applies to more than the single variety.

The "Harvard mouse" invention, first to receive a US patent, also received the first animal patent in Europe, but the path was less than direct. The initial application (1985) was dismissed (1989) on the interpretation that it was a prohibited animal variety and then granted on appeal (1990, issued 1992). A subsequent opposition (1992) did not lead to a cancellation, but the revised patent was limited to rodents only (2001). Further appeals are likely.

The EPC (Article 53(a)) further allows exclusions from patentability for inventions that would be contrary to "ordre public" or "morality." Some attempts, generally unsuccessful, have been made to identify animal patenting as immoral. The Patent Office applies a test contrasting the public benefit from a successful invention with the potential suffering of the animals involved. Specifically, the Examining Division declared:

"In the case at hand three different interests are involved and require balancing: there is a basic interest of mankind to remedy widespread and dangerous diseases, on the other hand the environment has to be protected against the uncontrolled dissemination of unwanted genes and, moreover, cruelty to animals has to be avoided. The latter two aspects may well justify regarding an invention as immoral and therefore unacceptable unless the advantages, i.e. the benefit to mankind, outweigh the negative aspects."

3.3 Legislation in Other Leading Countries

Many patent statutes are silent on the patentability of higher life forms, making the patentability of animals a matter of legal interpretation. This point aside, 17 countries have granted the "Harvard mouse" patent (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, The Netherlands, Portugal, Spain, Sweden, the United Kingdom, and the United States; plus Japan has a related patent and New Zealand has issued a patent for another form of transgenic mouse). The issues in notable countries outside Europe and the United States are described below:

Australia: The Patent Office does not treat higher life forms (including animals) distinctly from lower life forms. And since lower life forms that are novel and inventive are patentable, it may be (and has been) higher animals can be patented too.

Canada: The "Harvard mouse" patent was initially rejected by the examiner on the grounds that higher life forms were not patentable subject matter in Canada. Following were a series of appeals:

- 1995 appeal to Commissioner of Patents, which upheld examiners rejection on the grounds that the creation of the mouse proper was controlled by the laws of nature.
- 1998 appeal to Trial Division of the Federal Court, which upheld the Commissioner's interpretation that higher

life forms were not patentable subject matter in Canada.

- 2000 appeal to the Federal Court of Appeals, which overturned the Trial Division decision, drawing on the reasoning of the US *Chakrabarty* decision.
- In 2002 (May), Commissioner of Patents appealed to the Supreme Court of Canada.
- In 2002 (December), the Supreme Court of Canada decided 5 to 4 that higher life forms are not patentable in Canada. The details are given below.

The sole issue before the Supreme Court was that the patentability of life forms constitute a ''manufacture'' or ''composition of matter'' within the context of Article 2 of the Canadian Patent Act. Interpreting the terms ''manufacture'' and ''composition of matter,'' the former was understood to be a ''nonliving mechanistic product or process,'' while the latter was construed, in the context of the proceeding terms of allowable subject matter (invention), ''any new and useful art, process, machine, manufacture or composition of matter'' (Patent Act, Section 2). The majority concluded that, while ''composition of matter'' can be interpreted broadly, the context of Section 2 means it is ''best read as not including higher life forms.''

''Higher life forms cannot be conceptualized as mere 'compositions of matter' within the context of the *Patent Act*. [] It is possible the Parliament did not intend to include higher life forms in the definition of 'invention.' It is also possible the Parliament did not regard cross-bred plants and animals as patentable because they are better regarded as 'discoveries.'''

The decision proceeds to note that patents for human life cannot be excluded judicially by the courts and that the existence of Canadian PVP legislation can be interpreted as possibly indicating the need for separate legislation to apply to higher life forms generally. Overall, ''clear and unequivocal legislation [by the Parliament] is required for higher life forms to be patentable.''

New Zealand: New Zealand uses the same term ''manner of manufacture'' to describe patentable subject matter as does Australia, so that applications in the two countries have progressed in parallel. Specifically, New Zealand has granted patents for higher animals.

Japan: Part VII of the Examination Guidelines for Patent and Utility Model in Japan (2000), Chap. 2, Part 4 describes when patents may be awarded for animals (Available at (verified 4/5/04) http://www.deux.jpo.go.jp/cgi/search.cgi?query=examination+guidelines&lang=en&root=short). Patents may be granted *per se,* for parts of animals or methods of creating animals. Humans are excluded. Also excluded are discoveries and those inventions not capable of industrial application (utility).

3.4 International Agreements

This European Patent Convention wording is also significant because it has carried through in modified form into the TRIPs (Trade-Related Aspects of Intellectual Property Rights) of the 1994 World Trade Organization. Signatories (all 147 of them currently) have agreed to adopt certain minimum forms of intellectual property protection. Some are straightforward – the provision of trade secret legislation is an example. The area of plants and animals though remains ambiguous, for in Article 27.3(b), countries have the option of excluding from patentability ''plants and

animals other than microorganisms, and essentially biological processes for the production of plants or animals []. However, Members shall provide for the protection of plant varieties either by patents or by an effective *sui generis* system or by any combination thereof.'' While the situation in some developed countries is ambiguous regarding the patenting of animals, as is noted above, virtually all developing countries are choosing to prohibit patents for plants and animals. In this case, the bar to the patenting of animals is more absolute than under the European Patent Convention for the wording reads as ''animals,'' not ''animal varieties.'' In consequence, it will likely be some time at best before animals are patentable in developing countries.

It is worth noting that while there is extant a form of *sui generis* (which means ''special purpose'') protection for plants in the form of PBR (see Sect. 2.3.2), nothing of the type exists for animals. Its absence means that there is no protection for animals not qualifying for a patent, which is to say nontransgenic animals. Yet the development of superior breeds is a costly and risky endeavor, just like other inventive activities, so animal breeding can be presumed to be under-funded as a consequence of the absence of a *sui geners* protection system. Lesser has attempted to conceptualize such a system, noting that the principal complexity is any workable definition of an ''animal variety.'' Some animal lines such as those typically used in experimentation are sufficiently inbred to be stable and describable. Typically though, livestock, insects, fish, and pets are sufficiently heterogeneous as to defeat the description requirement of an IPR system. Further compounding the issue is the simple operation of heredity, which shows that only half the offspring of a simple trait will inherit the trait if just a single parent is a carrier, while more complex traits are inherited at far lower levels of probability. As a result, without testing individuals there is no assurance as to which will and which will not inherit a trait.

Lesser's approach in bypassing these characteristics of many higher animals functions more like a trademark than patent or PBR law. That is, protection is applied to the name of a parent and royalties are owing if the name is evoked, but not otherwise. Hence, one of *Man O'War's* foals would be charged royalties if he were specifically identified in a blood line, but not otherwise. Many beef cattle and purebred dogs use similar systems of identifying parent(s), while chickens and some fish and hogs (known as *synthetics*) are biologically more like hybrids than traditionally bred animals. They can be effectively protected through trade secrecy (see Sect. 2.3.3) by keeping the patent pure line stock and crossing sequence as confidential.

4 Issued Patents

On-line search engines of patent files at the US Patent and Trademark Office and the European Patent Office (EPO) make it possible to determine the numbers and types of patented animals. There is some uncertainty in the counts for it is not always entirely clear if the patent reads to an animal or a method for producing one. Also, when considering the species of the animals, some are definitive while other patents read to several species or all ''nonhuman mammals.'' Nonetheless, from a review of patent files, it is possible to gain a clear understanding of the kinds of patents and uses thereof.

4.1 Animal Patents in the United States

US animal patents are classified in US class 800/13 – 20 (available at www.uspto.gov under patents, search). As of mid-2004, animal patents comprised:

Mammal (800/14)	88
Bovine (800/15)	22
Sheep (800/16)	21
Swine (800/17)	24
Mouse (800/18)	350
Bird (800/19)	1
Fish (800/20)	9

It is clear that mice (and rats) are the most commonly patented. As for the uses, animal disease models (as with the initial "Harvard mouse") dominate. An example is patent # 6,156,279, *HIV transgenic animals and uses therefore*, Dec. 5, 2000, which first claim reads (claims identify the invention):

1. A transgenic rat, whose genome contains at least one copy of a human immunodeficiency virus type 1 (HIV-1) proviral DNA [] and wherein said rat develops at least one symptom of acquired immune deficiency syndrome (AIDS).

Second in number are patents for the use of transgenic animals to produce useful compounds, most of which are excreted into milk. An example is patent # 6,727,405, *Transgenic animals secreting desired proteins into milk* (April 27, 2004):

1. A non-human mammal whose genome comprises a DNA construct [] wherein said mammal is selected from the group consisting of mouse, sheep, pig, goat and cow, and wherein said heterologous protein is expressed in the milk of the mammal.

Further in the medical area is a patent (# 6,331,658, *Genetically engineered mammals for use as organ donors*, Dec. 18, 2001) for producing a transgenic pig as a source of human transplant organs with a reduced likelihood of rejection:

10. A non-human transgenic mammal, wherein the genome of the mammal stably includes a nucleotide sequence [] such that the organ exhibits a decrease in antibody-mediated rejection. . .
11. The mammal of claim 10 wherein the mammal is a pig.

Examples of nonmedically related inventions are less common. One is (H1,065, Apr. 30, 2002, *Transgenic avian line resistant to avian leukosis virus*):

1. A chicken line, derived from a strain of chickens which is susceptible to avian leukosis virus infection, which chicken line has integrated in its genome a proviral sequence [] and is thus genetically resistant to avian leukosis virus, subgroup A.

This invention appears to anticipate commercial use, viruses being a cause of significant loss in poultry production.

Also included are – interestingly – transgenic pets. Led by patent # 6,380,458 (Apr. 30, 2002) is *Cell-lineage specific expression in transgenic zebrafish*. The claims leading to the commercial product, "GloFish," are:

1. A transgenic zebrafish that expresses a heterologous expression product, comprising a zebrafish cell lineage-specific expression sequence. . .

4. The transgenic zebrafish of claim 3 wherein the reporter protein is green fluorescent protein.

4.2
Animal Patents at the EPO

Animal patents (see Sect. 3.2) are indeed being issued in the EU. The UK Patent Office issued patent number GB2382579, "Assays and Disease Models Using Transgenic Zebrafish," in October 2003. The animal model screens suppressors that lessen the activity of a disease gene.

Current applications number approximately 70 (although not all can be expected to mature to a grant). Animal model patents represent 55 of applications, and expressions of proteins the remainder. The pattern is similar to the United States, and indeed many applications are based on inventions for which the first application was made in the United States. Examples from European inventors include:

"Transgenic Animal Model for Obesity Expressing FOXC2." Publication # EP1255774 11/02.

"C1 Inhibitor Produced in the Milk of Transgenic Mammals." Publication # 1252184, 9/01.

5
Issues with Patented Animals

This section addresses the numerous issues that arise or are associated with patented animals. The list of items includes other technical developments that are potentially patentable but have yet to be the subject of applications to what is known of effects, environmental, economic, and social, to attitudes.

5.1
Status of Technology

The current granted patents and applications (see Sect. 4) clearly indicate technologies and applications at a high level of development. It is possible that developments that are generally either (1) not sufficiently refined or novel to be the subject of an application, or (2) are of insufficient economic potential to justify the costs associated with a patent are absent. Yet, those technologies may in future years mature to become the subject of applications and so are worthy of notice. A technology of particular note is the replacement of microinjection (first applied in 1981) and retroviral vector systems (1985) as the transformation methods of choice to the use of sperm-mediated transgenics, or techniques involving nuclear transfer. Microinjection, while functional, was marked by a low level of success (less than 1%). That level means successful transformations are costly, and potentially contributes to unintended animal suffering.

Transgenic animals are produced for five underlying reasons:

- Improve animal health
- Increase productivity and improve product quality
- Mitigate environmental effects of animal production
- Produce therapeutics
- Provide models for medical research.

As is readily seen in Sect. 4, the final two reasons dominate current patents, while the first is also in evidence. The first three listed are nonetheless underrepresented in current patent documentation, and so warrant further attention here.

Sample developments are reported in Table 1.

Tab. 1 Examples of developments with transgenic animals.

Species	*Modification*	*Purpose*	*Main beneficiaries*
Tilapia	Growth hormone	Faster growth	Fish farmers
Salmon	Growth hormone	Faster growth	Fish farmers
Salmon	"Antifreeze" gene	Faster growth	Fish farmers
Sheep	Antibody in milk	Enriched milk	Consumers
Cow, ewe	Human protein in milk	Produce molecules	Humans
Cow	Higher casein	Enriched milk, heat resistance	Food industry
Pig	Phytase gene	Reduce phosphorus in excrement	Environment
Insects	Male sterility	Reduce pest populations	Farmers, public
Mosquitoes	*Plasmodium* (malaria) resistance	Health	Humans, esp. in tropics

Source: Findinier 2003, pp. 20–21.

A few of these require further explanation. Food quality/composition is a highly desirable development, but technically it has been difficult to achieve as traits are frequently controlled by multiple genes. Plant biotechnology has experienced similar complexities. Fish biotechnology is largely directed to enhanced growth rates, using several distinct approaches. Transgenic salmon have exhibited growth rates three to five times the nontransformed controls. Those with the antifreeze attribute show growth throughout the year, not only during the warmer months.

The transgenic Enviropig© uses the phytase enzyme transferred from a bacteria to help with the digesting of phosphorous in cereal grains. There is an 80 to 100% percent improvement over the figures for nontransformed pigs in the sense that there is no need for phosphorous supplements; Further, there are reductions of 57 to 64% of phosphorous in manure. Owing to the effect of phosphorous on stimulating aquatic algae growth, its reduction in manure will have distinct environmental benefits. However, the transgenic must compete in the marketplace with feed-based phosphorous-reducing approaches, so the practicality remains to be proven.

Genetic male sterility to control pest populations is seen as a more cost-effective alternative to radiation sterilization. It has the conceptual benefit of not reducing the competitiveness of the sterile males, but practicality varies across species because of the ease of sexing and degree of monogamous mating. Mosquitoes resistant to *Plasmodium* provide an alternative approach to controlling one of the greatest causes of loss of life worldwide. The need is greatest as populations are developing resistance to available prophylactic medications. However, studies have shown that resistance has to be near 100% to be effective, and resistance-delaying strategies may mean the engineering of multiple gene resistance, a complex and time consuming task.

5.2 Benefits

The preceding sections describe – or, at minimum, hint at – the benefits generated

by patented animals. In this section, we have included further documentation of the benefits that are being generated and to whom they accrue.

5.2.1 Animal Models

Numerous animal models are being developed through the mechanism of biotechnology. Firms are presently designing new models on request, as with Charles River Laboratories, one of the large supplier firms:

> Charles River and Xenogen Biosciences formed a marketing collaboration []. Xenogen Biosciences provides custom programs for the development of transgenic animals []. (www.criver.com; verified 5/04)

Products are in considerable demand. By 1989, orders for the path breaking "Harvard mouse" amounted to 10 000 at $50 each, versus $3.50 for standard lines.

5.2.2 Production of Compounds

The actual production costs of compounds using transgenics versus traditional approaches are not publicized, making direct comparisons impossible. However, complex compounds are very costly to produce using bioreactors, such as blood clotting Factor IX, which costs $100 000 to 200 000 annually per patient for hemophiliacs. Pigs can secrete Factor IX into milk at a concentration 250 to 1000 times the level of bioreactors, making purification simpler. Indeed, production is so elevated that only several hundred transgenic pigs are needed to supply all hemophiliacs worldwide. Animal numbers, however, will be so small compared to standing herd sizes as to have no effect on the farm and agribusiness sectors.

5.2.3 Other Applications

Other transgenic technologies, and particularly those used for milk production, are insufficiently commercialized at this stage to have any reliable indication of effects on producers or consumers. One can however note that animal diseases are a major cost component for prevention, treatment, and associated losses so that enhanced resistance would have marked benefits for both producers and consumers of animal products. For example, US vet and medicine costs for hogs per hundredweight of gain in 2002 was $1.10, or 2% of total costs, but 14% of residual returns minus operating costs. Worldwide, small farmers are the major owners of livestock – in India, they own 70% of the livestock but just 30% of the land area – and could be a major beneficiary.

For fish, a genetic mutation (nonbiotech) leading to genetically male Nike tilapia increased production by 30%. For Philippine farmers utilizing them, the net income doubled.

5.3 Areas of Concern

Animal patents are, to understate the current situation, very controversial, at least with certain groups. Concerns are focused on

- ethics,
- food safety,
- health,
- environmental safety, and
- access to patented animals.

Here the underlying positions will be summarized in reverse order. No attempt is made to represent a consensus position as that may not exist, or is not discernible. Rather, in the following section, attitude surveys from several countries are reported

as an indication of the acceptability of patented animals in particular, and the products of those creations.

5.3.1 Access to Patented Animals

Effective access involves a complex of cost, availability, and use provisions. Here, attention is on agriculture/aquaculture and other dispersed uses rather than medical applications, where commercial channels among limited number of users are of long-standing nature.

Cost cannot be reliably projected at this stage, other than to note that most food production have low margins, meaning high process would render the innovations uncompetitive in many cases. For example, with rBST, an injected biotech product used for stimulating milk production, analysis suggests that the supplier, Monsanto, prices the product near the breakeven point for users. That keeps adoption rates low – indeed, over time, use has been declining. Distribution of livestock-based technologies depends in part on the production rates by species. Poultry are high producers, hogs medium, while cattle naturally produce but one calf annually. For single gene traits, only half will inherit; for multiple gene ones, less. Embryo splitting and implantation speed up the rate, but at considerable cost. For cattle, then, dissemination through herds will be a slow process.

A second consideration relates to the collection of royalties, assuming that producer cash flow considerations will dictate royalties be paid when returns are realized at sale time. Lesser's analysis proposes that packing plants (livestock) or breed associations (pets) will be the royalty collectors of necessity. Until rapid, cheap tests are developed to determine the presence of traits, fees must be based on herd averages.

5.3.2 Environmental Safety

Environmental safety applies principally to concerns over the escape and competition by transgenic animals in the natural environment. Competition can affect the same species through crossbreeding, or other species by colonizing an environmental niche. The degree of the threat then depends in large measure on a species' ability to become feral, escape from captivity, and its mobility. Insects then are predicted to have substantial potential for environmental damage while domesticated cattle are ranked low (Table 2).

Some experience exists with nontransformed salmon raised in cages in the sea. Escapes are estimated at 15%, a number confirmed by the percent of farm-raised salmon caught by Norwegian fishermen. However, farm-raised salmon, like most nonnative species, have been found to be ineffective breeders compared to the wild stock. Nonetheless, if the alien species is a large portion of a local population, then a notable impact can be anticipated.

Procedures for assessing and, when needed, mitigating the environmental

Tab. 2 Environmental risk factors levels for transgenic animals.

Animal	*Ability to become feral*	*Likelihood of escape*	*Mobility*
Insects[a]	High	High	High
Fish[b]	High	High	High
Rodents	High	High	High
Cats	High	High	Moderate
Pigs	High	Moderate	Low
Rabbits	High	Moderate	Moderate
Chickens	Low	Low	Low
Cattle	Low	Low	Low

[a]Gypsy moth and African bee only.
[b]Excludes shellfish and crustaceans.
Source: FDA, reported in Findinier 2003, p. 17.

affects of transgenics are known as *biosafety*. (The term sometimes encompasses food safety, but as used here refers to environmental aspects only). National legislation specifies the review procedures involved, although they are better developed and more practiced with plants than animals. Internationally, there is an umbrella agreement, the Cartagena Protocol under the Biodiversity Convention (available at http://www.biodiv.org/biosafety/default.aspx; verified 5/04). The Protocol, established in 2000, has to date (5/04) been ratified by 98 countries. It establishes general procedures and requirements but does not contain specific provisions.

5.3.3 Health

Human health issues are associated with the development of new viruses, particularly if animal organs are used for human transplants. A retrovirus used as a transformation vector could combine with a latent virus in the new host, spreading through the food chain and into humans. This threat can be ended by discontinuing viral vectors.

Transplanted organs present a special risk for the same genetic modification that reduces the rejection of the foreign organ could also weaken the patient's immunity to the new virus. Thus situated, the new virus could adapt itself to the novel human host environment and become infectious. This possibility is not trivial for pigs – already the subject of a xenotransplantation patent (see Sect. 4.1) – have at least six known endogenous retroviruses. Overall, concerns about the generation of new diseases "require a chain of events of very low cumulative probability but for which there are some experimental precedents."

5.3.4 Food Safety

Internationally, food safety standards and procedures are established under FAO/WHO by the Codex Alimentarius Commission. The main purposes of this program are protecting health of the consumers and ensuring fair trade practices in the food trade, and promoting coordination of all food standards work undertaken by international governmental and nongovernmental organizations.

In 2003, the Commission adopted the "Principals for the Risk Assessment of Foods Derived from Modern Biotechnology" as an overarching agreement for understanding risk assessment and nutritional status of foods derived from biotechnology, including transgenic animals (available at http://www.fao.org/es/ESN/food/risk_biotech_taskforce_en.stm; verified 5/04). The principals are based on the following elements:

- Premarket assessment on a case-by-case basis
- Assessment based on a comparison with traditional foods (known as substantial equivalents)
- Risk management should be proportional to identified risks.

Food concerns of transgenic animals are focused largely on the existence of novel proteins. Those proteins can be added intentionally (as with a growth hormone), result from unanticipated compositional changes, or may be a consequence of the unpredictable trans gene expression in animals. For example, there may be a transfer from mammary glands to the blood and hence into the food supply. When testing for food safety, the traditional focus is on

- toxicity,
- allergenicity, and
- whole food value.

Toxicity is relatively straightforward to examine, although there is always a policy decision over which populations (elderly, children) are to be protected and at what presumed consumption levels. Allergenicity is problematic to screen for because very small portions of a population may be very allergic to a wide range of proteins. Whole food value – an assessment if the total nutritional balance has been altered – presents test complications because the subject product often constitutes a small portion of the diet so that feeding large-enough amounts to detect nutritional effects would be unfeasible or unhealthy. In any case, similar issues arise with transgenic plant-derived foods, and to date there is no documented example of any deleterious human health effects.

5.3.5 Ethics

Individuals may object to patenting animals either directly on the basis of a rejection of ownership and "playing God" by interfering with a natural creation or indirectly because patents can foster actions or behavior considered deleterious to animals. The first point is a difficult one for multicultural societies to resolve because opinions can be intense, but may vary widely. What is morally repugnant to some may be fully acceptable to others. The best indication of the prevalence of these objections is public opinion surveys (see Sect. 6 following).

The indirect effects of patents through transgenics can be discussed in the terms of animal rights and animal welfare. Animal rights supporters may be concerned with, for example, not changing what is essential about a pig, or allowing animals to live in a way consistent with their nature. Pigs, for example, are both social and programmed to rooting and, by one perspective, should not be prevented from those expressions. What is innate to a species is often difficult to conceptualize after millennia of being subject to human control. And rights have broad consequences that are likely not fully comprehended by many.

Animal welfare, however, has wide and growing support, largely directed to the avoidance of unnecessary suffering. Austria, for example, recently adopted legislation specifying that chickens must be reared on a floor (not in cages) and Dobermans may not have tails and ears trimmed for appearance reasons. Animal suffering is a possible issue with animal patents if, for example, territorial instincts can be reduced, allowing for more dense confinement, or if experimentation leads to suffering. How is that matter addressed?

A partial offset to overall suffering is the possibility of experimentation that will lessen future suffering. Developing disease treatment or resistance is an example. There are examples of disease-resistant animals becoming available (see Sect. 4). Yet, many experiments will benefit humans far more than the test species. In those cases, a common approach is to consider the balance of (animal) suffering compared to potential (human) benefit. In Europe, the EPO has devised a process for determining if the advantages outweigh the negative aspects (see Sect. 3.2). Outside patent offices, calls have been made for weighing "costs and benefits." For the foreseeable future, the utilitarian view of placing human needs in the primacy while minimizing the costs to animals seems likely to prevail, but views of animal rights are expanding with the general public and

could affect fundamental research with transgenics in the longer term.

6
Public Attitudes

As a background, it is important to recognize that the general public is not particularly well informed about biotechnology. As examples:

- 1999 (Gallup) 50% heard little/nothing
- 1987 (Gallup) 63% heard little/nothing
- 2001 "heard a great deal about genetically modified foods" – 44%
- 2003 "heard a great deal about genetically modified foods" – 34%

If anything, awareness in the United States is declining in an era where there are no notable headlines regarding transgenics.

Indeed, the public indicates a low level of knowledge on the subject – in 2001, respondents self-selected a median rating of only 3 (1–10 scale of "knowledgeable") – and demonstrates a limited knowledge with such replies as:

United States	33% believe tomatoes do not contain genes, 58% say they have not eaten transgenic foods
Europe	65% believe tomatoes do not contain genes

Of course, knowledge does not necessarily influence opinions, which are based on emotions. If anything, a lack of knowledge makes would-be consumers less accepting as risk is associated with the new. For that reason, more educated consumers, who presumably are better able to educate themselves, typically are more accepting. The lack of knowledge does make respondents especially vulnerable to differences in terminology. Surveys have indicated that the label biotechnology is better accepted than using "genetically modified." Hence, traditional supporters of biotechnology like food companies use the former in surveys, while opponents such as some environmental groups favor genetically modified. Overall, as always, surveys should be interpreted with some care.

6.1
Attitudes in the United States

Whatever the comfort level with eating transgenics, consumers are accepting the transforming of life forms for other reasons. While transgenic animals for food are ranked at 2.8 out of 10 (compared with 6.8 for plants), other uses of transgenic animals is less than half that level, or 2.3 (Fig. 1). Seemingly, the closer the life form is to direct human experiences, the lower the acceptance.

Support for transgenic animals is thrown into sharper contrast when evaluated against transgenic plants. For plants, there is an 81% positive position, and a 67% point spread between favorable and unfavorable views (Fig. 2). For the same question applied to animals, the figures are 49% approval and only an eight-point spread. When the issue is less cost and more specific health and safety benefits, the divergence is smaller, but still notable. For example, 57% of respondents favor the use of animals for transplant organs, while 58% support the idea of producing spider-web compounds in goat milk (which can make strong and light bulletproof vests). Another survey reports the approval for plant transgenics of 49% in 2003 versus 27% for animals.

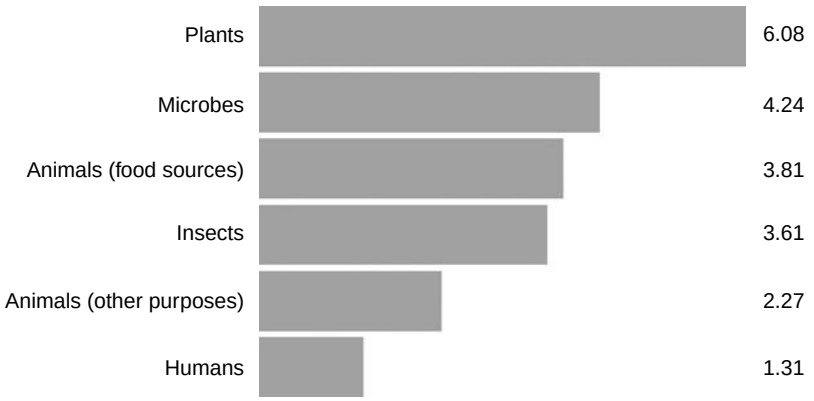

Fig. 1 Comfort with the genetic modification of animals (rank by mean (out of 10)). *Source*: Pew Initiative on Biotechnology (2003b) *Americans Are Far More Comfortable With Genetic Modifications Of Plants Than Animals*. Available at http://pewagbiotech.org/research/2003update/4.php; verified May 2004.

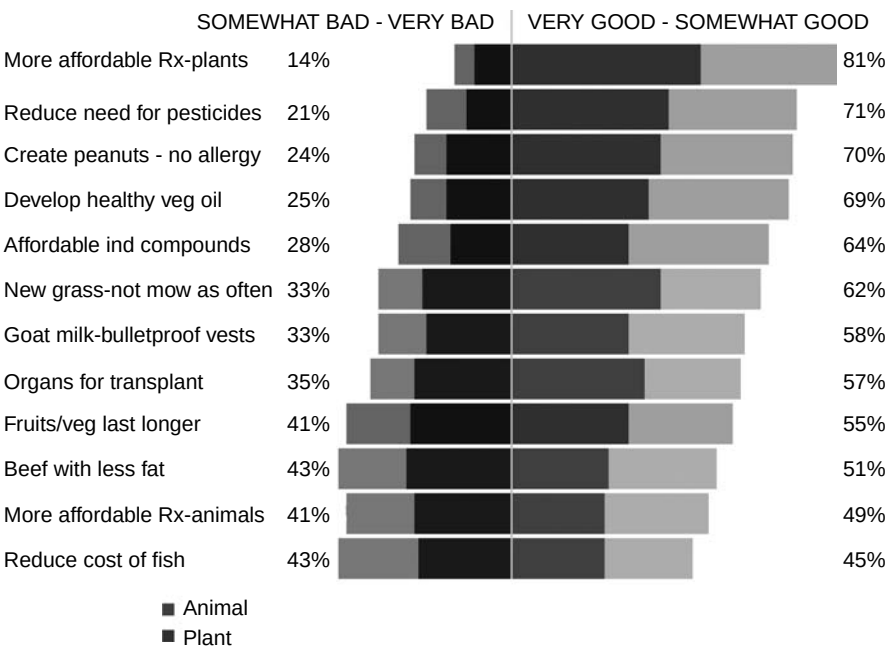

Fig. 2 Likelihood of supporting the genetic modification of plants and animals when used for (darker shading : stronger intensity). *Source*: Pew Initiative on Biotechnology (2003b) *Americans Are Far More Comfortable With Genetic Modifications Of Plants Than Animals*. Available at http://pewagbiotech.org/research/2003update/4.php; verified May 2004.

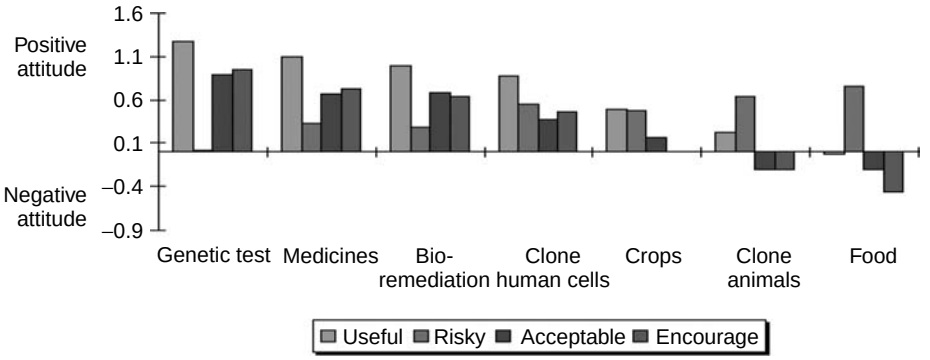

Fig. 3 European attitudes to uses of biotechnology, 2002. *Source*: Nature America, 2000, from Eurobarometer.

These opinions should be viewed in the context of public knowledge of animal biotechnology. In fact, 32% of Americans state that they do not know enough to form an opinion on the subject and another 11% are neutral (28% have a favorable impression, 29% unfavorable). Consumers distinguish among three areas of animal biotechnology. They are favorably disposed (55%) toward "genomics," the use of genetic information to improve animal care and food quality from animals. However, fewer consumers are favorably disposed toward genetic engineering in animals (36%), and even fewer toward cloning (18%).

6.2 Attitudes in Europe

While Europeans are more opposed to biotechnology than are North Americans, attitudes are nuanced by the use of the technology, as in North America. Support is far higher for medicines than food in general (Fig. 3).

By 2002, the periodic Eurobarometer series on attitudes to biotechnology had added a question about xenotransplantation, the only question with direct relevance to patented animals. The absolute rating was similar to crops, with useful and risky at nearly identical levels (Fig. 3). Weak support exists in 13 of the 15 countries polled, with Greece, Finland, and Austria in opposition (Table 2).

The question if there is a general change in attitude over time cannot be answered directly from the opinion surveys because of changes in the wording of the question. However, partitioning the respondents into supporters and risk acceptors, a general pattern of increases in support for "GM foods" is evident in all countries but France, Italy, and the Netherlands. Indeed, support in Italy actually declined – the only country where this happened.

See also Nuclear Transfer for Cloning Animals.

Bibliography

Books and Reviews

Beardmore, J.A., Porte, J.S. (2003) *Genetically Modified Organisms and Aquaculture*, FAO, Rome, Fisherise Circular No. 989 (FIRI/C989C).

Blattman, A., McCann, J., Bodkin, C., Naumoska, J. (2002) *Global Intellectual Property: International Developments in Animal Patents*, Ch. 5 63–84 in M.F.

European Patent Office (1992) *Official J.* **588**, 591–592.

Findinier, I. (2003) *Animal Biotechnologies: State of the Art, Risks and Perspectives*, FAO, Rome Available at www.fao.org/es/ESN/food/GenBiotech_en.

Gaskell, G., Allum, N., Stares, S. (2003) *Europeans and Biotechnology in 2002*. Eurobarometer 58.0. Available at http://europa.eu.int; verified 5/04.

Grubb, P.W. (1999) *Patents for Chemicals, Pharmaceuticals and Biotechnology*, Clarendon Press, Oxford.

Hallman, W.K., Hebdan, W.C., Aquino, H.L., Cuite, C.L., Lang, J.T. (2003) *Public Perceptions of Genetically Modified Foods: A National Study of American Knowledge and Opinion*, Food Policy Institute, Cook College, Rutgers, New Brunswick, NJ Available at www.foodpolicyinstitute.org; verified May 2004.

Houdebine, L.M. (2002) Transgenics to improve animal production, *Livest. Prod. Sci.* **74**, 255–268.

International Food Information Council (2004) *Support for Food Biotechnology Stable Despite News on Unrelated Food Safety Issues*. Available at http://ific.org/research/biotechres03.cfm; verified May 2004.

Ithaca J. (2004) Genetically modified pets making retailers uneasy, 8A.

Johnston, J.A., *Rochester Times-Union* (1989) B&L will make mice for cancer research, 1A, 8D.

Kesavan, P.C., Swaminathan, M.S. (2003) Ethical, Social, Environmental and Economic Issues in Animal Agriculture, *Book of Extended Synopses, FAO/IAEA International Symposium on Applications of Gene-Based Technologies for Improving Animal Production and Health in Developing Countries*, Vienna, 6–10 Oct. 2003, IAEA-CN-110/KN85.

Lavitrano, M., Bacci, M.L., Forni, M., Lazzereschi, D., Stefano, C.D., Foretti, D., Giancotti, P., Marfe., G., Pucci, L., Renzi, L., Wang, H., Stoppacciaro, A., Casa, G.D., Seren, E., Rossi, G. (2002) Efficient production by sperm-mediated gene transfer of human decay accelerating factor (hDAF) transgenic pigs for xenotransplantation, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14230–14235.

Lesser, W. (1987) The impacts of seed patents, *North Cent. J. Agr. Econ.* **9**, 37–48.

Lesser, W. (1993) Animal variety protection: a proposal for a US model law, *J. Pat. Trademark Off. Soc.* **76**, 697–715.

Lesser, W., Bernard, J., Billah, K. (1999) Methodologies for *ex ante* projections of adoption rates for agbiotech products: Lessons learned from rBST, *Agribusiness* **15**(2), 149–162.

Maskus, K.E. (2000) *Intellectual Property Rights in the Global Economy*, Institute for International Economic, Washington, DC.

National Research Council (2002) *Animal Biotechnology: Science-Based Concerns*, National Academies Press, Washington, DC.

Nature America (2000) *Biotechnology and the European Public*. Available at http://biotech.nature.com; verified May 2004.

Pew Initiative on Biotechnology (2002) *Biotechnology in the Barnyard: Implications of Genetically Engineered Animals*. Available at www.pew.org; verified 5/04.

Pew Initiative on Biotechnology (2003a) *Awareness of Genetically Modified Foods Remains Low*. Available at http://pewagbiotech.org/research/2003update/4.php; verified May 2004.

Pew Initiative on Biotechnology (2003b) *Americans Are Far More Comfortable With Genetic Modifications Of Plants Than Animals*. Available at http://pewagbiotech.org/research/2003update/4.php; verified May 2004.

Rothschild, M.F., Newman, S. (Eds.) (2002) *Intellectual Property Rights in Animal Breeding and Genetics*, CABI Publishing, Oxford.

Thompson, P.B. (1997) *Food Biotechnology in Ethical Perspective*, Blackie, London.

U.S. Department Agriculture (2004) *Hog Production Costs and Returns*. Available at http://usda.ers.gov/Data/CostsandReturns/data/current/C-hogs.xlts (verified 5/04).

Walterscheid, E.C. (1994) The early evolution of the united states patent law: antecedents (*Part I*), *J. Patent and Trademark Office Soc.* **75**, 398–426.

Primary Literature

Committee on the Judiciary, House of Representatives (1988) *Patents and the Constitution: Transgenic Animals*, U.S. Government Printing Office, Washington, DC.

Grubb, P.W. (1999) *Patents for Chemicals, Pharmaceuticals and Biotechnology*, Clarendon Press, Oxford.

Lesser, W. (Ed.) (1989) *Animal Patents: The Legal, Economic and Social Issues*, Stockton Press, New York.
National Research Council (2002) *Animal Biotechnology: Science-Based Concerns*, National Academies Press, Washington, DC.
Office Technology Assessment, Congress of the US (1989) *New Developments in Biotechnology: Patenting Life – Special Report*, OTA-BA-370. U.S. Goverment Printing Office, Washington, DC.
Rothschild, M., Newman, S. (Eds.) (2002) *Intellectual Property Rights in Animal Breeding and Genetics*, CABI Publishing, Oxford.

Lou Gehrig's Disease (Amyotrophic Lateral Sclerosis): *see* Motor Neuron Diseases: Cellular and Animal Models

Lymphokines: *see* Cytokines: Interleukins

Machines, Biological: *see* Motor Proteins

Macromolecules, X-Ray Diffraction of Biological

Albrecht Messerschmidt and Robert Huber
Max-Planck-Institut für Biochemie, Martinsried, Germany

1 Introduction 395
1.1 Crystals and Symmetry 395
1.2 Protein Solubility 397
1.2.1 Ionic Strength 397
1.2.2 pH and Counterions 399
1.2.3 Temperature 399
1.2.4 Organic Solvents 399
1.3 Experimental Techniques 400
1.4 Crystallization Screenings 402

2 Experimental Techniques 403
2.1 X-ray Sources 403
2.1.1 Conventional X-ray Generators 403
2.1.2 Synchrotron Radiation 404
2.1.3 Monochromators 405
2.2 Detectors 406
2.2.1 General Components of an X-ray Diffraction Experiment 406
2.2.2 Image Plates 407
2.2.3 Gas Proportional Detectors 407
2.2.4 Charge-coupled Device-based Detectors 409
2.3 Crystal Mounting and Cooling 409
2.3.1 Conventional Crystal Mounting 409
2.3.2 Cryocrystallography 410
2.3.3 Crystal Quality Improvement by Humidity Control 412
2.4 Data-collection Techniques 413
2.4.1 Rotation Method 413
2.4.2 Precession Method 415

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

3 Principles of X-ray Diffraction by a Crystal 416
3.1 Scattering of X-rays by an Atom 416
3.2 Scattering of X-rays by a Unit Cell 418
3.3 Scattering of X-rays by a Crystal 418
3.3.1 One-dimensional Crystal 418
3.3.2 Three-dimensional Crystal 419
3.4 The Reciprocal Lattice and Ewald Construction 420
3.5 The Temperature Factor 422
3.6 Symmetry in Diffraction Patterns 422
3.7 Electron Density Equation and Phase Problem 422
3.8 The Patterson Function 424
3.9 Integrated Intensity Diffracted by a Crystal 424
3.10 Intensities on an Absolute Scale 425
3.11 Resolution of the Structure Determination 425
3.12 Diffraction Data Evaluation 426
3.13 Solvent Content of Protein Crystals 427

4 Methods for Solving the Phase Problem 427
4.1 Isomorphous Replacement 427
4.1.1 Preparation of Heavy Metal Derivatives 427
4.1.2 Single Isomorphous Replacement 428
4.1.3 Multiple Isomorphous Replacement 430
4.2 Anomalous Scattering 431
4.2.1 Theoretical Background 431
4.2.2 Experimental Determination 432
4.2.3 Breakdown of Friedel's Law 433
4.2.4 Anomalous Difference Patterson Map 434
4.2.5 Phasing Including Anomalous Scattering Information 434
4.2.6 Multiwavelength Anomalous Diffraction Technique 435
4.2.7 Determination of the Absolute Configuration 438
4.3 Patterson Search Methods (Molecular Replacement) 438
4.3.1 Rotation Function 439
4.3.2 Translation Function 439
4.3.3 Computer Programs for Molecular Replacement 440
4.4 Phase Calculation 440
4.4.1 Refinement of Heavy Atom Parameters 440
4.4.2 Protein Phases 441
4.5 Phase Improvement 443
4.5.1 Solvent Flattening 443
4.5.2 Histogram Matching 444
4.5.3 Molecular Averaging 444
4.5.4 Phase Combination 445
4.6 Difference Fourier Technique 445

5 Model Building and Refinement 447
5.1 Model Building 447
5.2 Crystallographic Refinement 447
5.3 Accuracy and Verification of Structure Determination 449

6 Applications 450
6.1 Enzyme Structure and Enzyme–Inhibitor Complex 450
6.1.1 X-ray Structure of Cystathionine ß-Lyase 450
6.1.2 Enzyme–Inhibitor Complex Structures of Cystathione ß-Lyase 452
6.1.3 Crystal Structure of the Thrombin–Rhodniin Complex 455
6.2 Metalloproteins 456
6.2.1 Crystal Structure of the Multicopper Enzyme Ascorbate Oxidase 457
6.2.2 Crystal Structure of Cytochrome-*c* Nitrite Reductase Determined by Multiwavelength Anomalous Diffraction Phasing 458
6.3 Large Molecular Assembly 462
6.3.1 Crystal Structure of 20S Proteasome from Yeast 462

Bibliography 464
Books and Reviews 464
Primary Literature 464

Keywords

Anomalous scattering
If the wavelength of the incident X-ray beam is around the absorption edge of a given atom, its atomic scattering factor becomes complex with a normal, a dispersive, and an absorption component.

Atomic scattering factor
Mathematical expression for the X-ray scattering of an individual atom.

Crystallographic R-factor
Reliability factor for the X-ray structure determination, sum over the absolute values of the differences between the observed and calculated structure factors divided by the sum of the absolute values of the observed structure factors.

Electron density function
The spatial distribution of electrons in the crystal, which can be obtained by the Fourier transformation of the structure factors, represents the atomic arrangement in the crystal.

Ewald construction
Construction to visualize the diffraction of X-rays by a crystal; a sphere with a radius of the reciprocal of the incident X-ray wavelength is drawn with the origin of the reciprocal lattice of the crystal at the intersection of the incident beam with the surface of the sphere; diffraction occurs if a reciprocal lattice point intersects the surface of the so-called Ewald sphere; the direction of the diffracted beam is the connection between the center of the Ewald sphere and the intersection point of the relevant reciprocal lattice point.

MAD
Multiple anomalous diffraction, a method to exploit the anomalous scattering effect of a set of natural or artificially introduced anomalous scatterers for the solution of the phase problem.

MIR
Multiple isomorphous replacement, a method to solve the phase problem by preparing isomorphous heavy metal derivatives and measuring X-ray data sets of the native and the derivative crystals.

Phase problem
The phase of the diffracted waves is lost during the measurement as the X-ray intensities deliver the amplitude of the diffracted waves only.

Reciprocal lattice
A lattice in reciprocal space with lattice points, which are perpendicular to the corresponding set of lattice planes in the lattice in direct space with a length of the reciprocal of the lattice plane distance or its multiples.

Resolution
Minimal distance between two objects that can be resolved.

Structure factor
Mathematical expression for the X-ray wave diffracted by a crystal, which consists of amplitude and phase factor.

X-ray diffraction of biomolecules covers diffraction experiments on single crystals of small biological molecules such as oligopeptides, cofactors, steroid hormones, and so on, or biological macromolecules. But it also includes small angle scattering experiments on biological macromolecules in solution. This chapter is exclusively dedicated to the X-ray crystallography of biological macromolecules. It comprises the determination of the three-dimensional structures of proteins, nucleic acids, and other biological macromolecules at atomic resolution by diffraction of X-rays

on crystals of such macromolecules. The atomic structure is obtained from the electron density distribution of the macromolecular crystal, which is the Fourier transform of the waves diffracted by the crystal. The amplitudes of the diffracted waves are determined directly from the diffraction experiment. The diffraction phases are revealed (1) from additional diffraction data of isomorphous heavy atom derivatives (multiple isomorphous replacement (MIR) technique), (2) from multiple anomalous diffraction (MAD), if suitable anomalous scatterers are in the crystal and using tunable synchrotron radiation, and (3) by Patterson search techniques (molecular replacement), if structural information on the biological macromolecule under investigation is available. X-ray crystallography of biological macromolecules is the unique method for the elucidation of the spatial structures at atomic resolution of complex biomolecules with molecular masses greater than 30 kDa. The structures represent a time and spatial average of molecules packed in a crystal. The preparation of suitable crystals may be a limiting factor.

1 Introduction

1.1 Crystals and Symmetry

In a crystal, atoms or molecules are arranged in a three-dimensionally periodic manner by translational symmetry. The crystal is formed by a three-dimensional stack of unit cells, which is called the *crystal lattice* (Fig. 1a and b). The unit cell is built up by three noncollinear vectors **a**, **b**, and **c**. In the general case, these vectors have unequal magnitudes and their mutual angles deviate from 90°. The arrangement of the molecule(s) in the unit cell may be asymmetrical, but very often it is symmetrical. This is illustrated in Fig. 2(a–e) in two-dimensional lattices for rotational symmetries.

In crystals, only 1-, 2-, 3-, 4-, and 6-fold rotations are allowed. This follows from the combination of the lattice properties with rotational operations. Other possible symmetry elements are mirror plane *m*, inversion center, and combination of rotation axis with inversion center (inversion axis). These are the point group symmetries. They can only occur among each other in a few certain combinations of angles. Other angle orientations would violate the lattice properties. The number

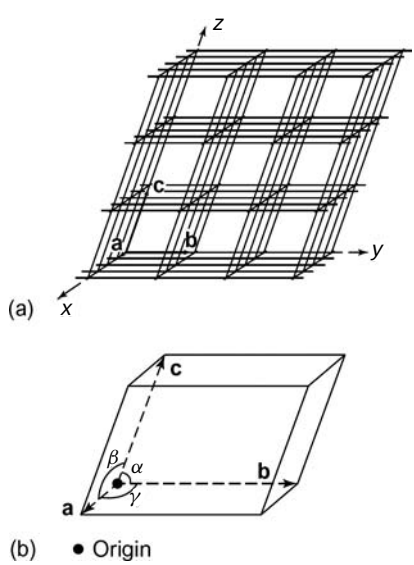

Fig. 1 (a) Crystal lattice and (b) unit cell.

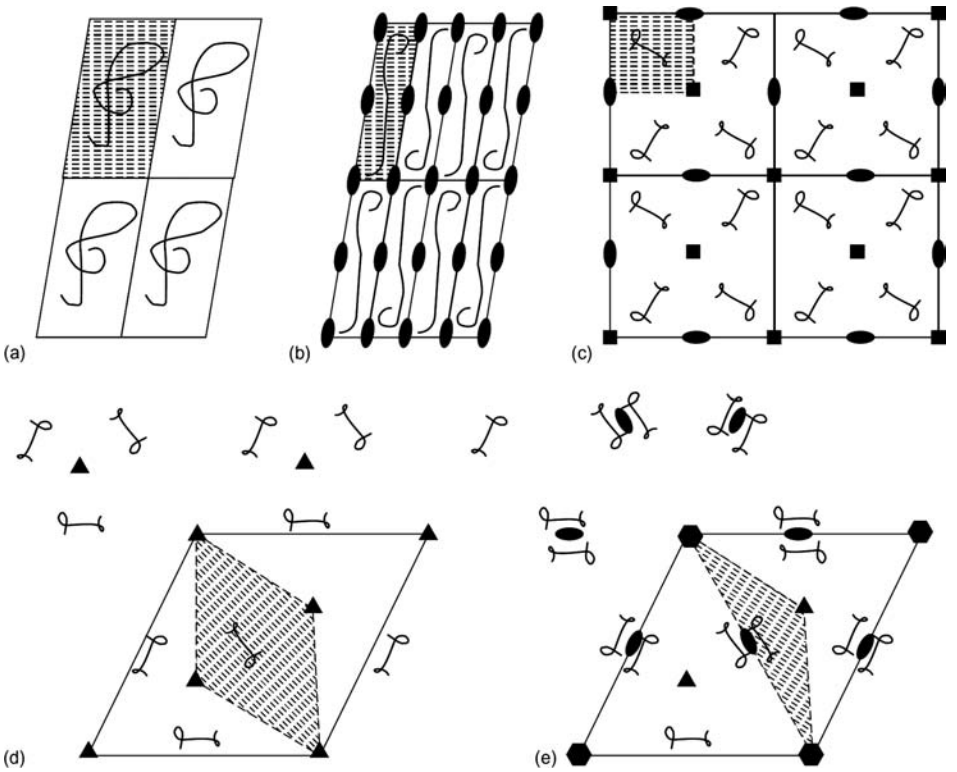

Fig. 2 Rotational symmetry elements in two-dimensional lattices: (a) 1, (b) 2, (c) 4, (d) 3, and (e) 62. The asymmetric unit is hatched.

of all possible combinations reveals the 32 point groups. The crystal morphology obeys the point group symmetries.

Adding an inversion center to the point group symmetry leads to the 11 Laue groups. These are of importance for the symmetry of X-ray diffraction patterns. Their symbols are 1, $2/m$, $2/mmm$, 3, 3 m, $4/m$, $4/mmm$, $6/m$, $6/mmm$, $m3$, and $m3m$. Proteins and nucleic acids are chiral molecules. Therefore, they can crystallize only in the 11 enantiomorphic point groups: 1, 2, 3, 4, 6, 23, 222, 32, 422, 622, and 432.

The combination of point group symmetries with lattices leads to seven crystal systems, triclinic, monoclinic, orthorhombic, trigonal, tetragonal, hexagonal, and cubic, with 14 different Bravais-lattice types, which can be primitive, face-centered, all-face-centered, and body-centered. Furthermore, additional symmetry elements are generated, having translational components such as screw axes or glide mirror planes. There exist 230 space groups of which 65 are enantiomorphic (for chiral molecules such as proteins). Figure 3 shows the graphical representation for the space group $P2_12_12_1$, as listed in the *International Tables for Crystallography*. The asymmetric unit is one-fourth of the unit cell and can contain one or several molecules. Multimeric molecules may have their own

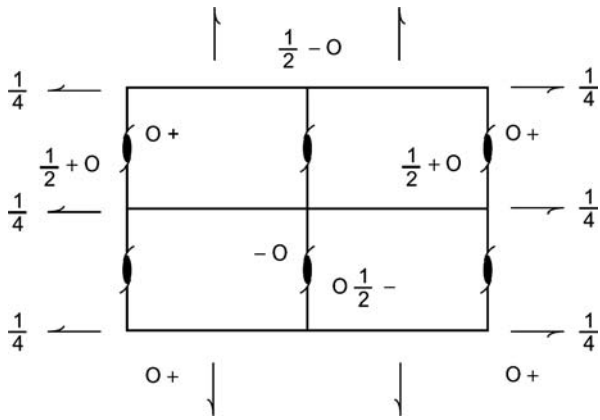

Fig. 3 Graphical representation of space group $P2_12_12_1$.

symmetries, which are called *noncrystallographic symmetries*. Here, axes that are 5-fold, 7-fold, and so on, are also allowed.

1.2 Protein Solubility

Figure 4 shows a typical phase diagram illustrating the solubility properties of a macromolecule. In the labile phase, crystal nucleation and growth compete, whereas in the metastable region, only crystal growth appears. In the unsaturated region, crystals dissolve. The solubility of proteins is influenced by several factors, as follows.

1.2.1 Ionic Strength

A protein can be considered as a polyvalent ion and, therefore, its solubility can be discussed on the basis of the Debye–Hückel theory. In aqueous solution, each ion is surrounded by an "atmosphere" of counter ions. This ionic atmosphere influences the interactions of the ion with water molecules and hence the solubility.

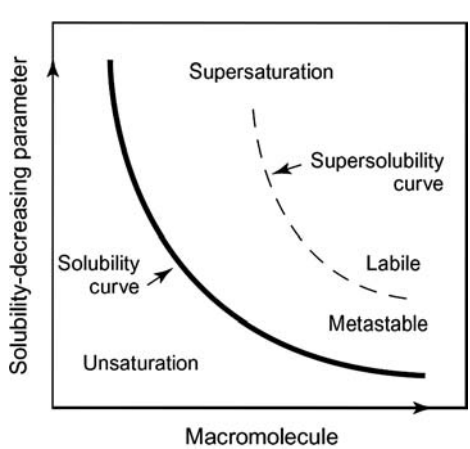

Fig. 4 Phase diagram illustrating the solubility properties of macromolecules. (Reproduced by permission of Academic Press, Inc., from Weber, P.C. (1997) Overview of protein crystallization methods, *Methods Enzymol*. **276**, 13–23.)

1.2.1.1 **"Salting-in"** At low ionic concentration, the "ionic atmosphere" increases the solubility as it increases the possibilities for favorable interactions with water molecules. We obtain (Eq. 1 and 2)

$$\log S - \log S_0 = \frac{AZ_+Z_-\sqrt{\mu}}{1 + aB\sqrt{\mu}} \quad (1)$$

$$\mu = \frac{1}{2}\sum c_j Z_j^2 \quad (2)$$

where μ = ionic strength, S = solubility of the salt at a given ionic strength μ, S_0 = solubility of the salt in absence of the electrolyte, Z_+, Z_- ionic charge of salt ions, A, B = constants depending on the temperature and dielectric constant, a = average diameter of ions, and c_j = concentration of the jth chemical component. Ions with higher charge are more effective for changes in solubility. Most salts and proteins are more soluble in low ionic strength than in pure water. This is termed as *salting-in* (Fig. 5).

1.2.1.2 **"Salting-out"** At higher ionic strength, the ions compete for the surrounding water. Therefore, water molecules are taken away from the dissolved agent and the solubility decreases according to Eq. (3):

$$\log S - \log S_0 = \frac{AZ_+Z_-\sqrt{\mu}}{1 + aB\sqrt{\mu}} - K_s\mu \quad (3)$$

The term $K_s\mu$ predominates at high ionic strengths, which means that "salting-out" is then proportional to the ionic strength (Fig. 5). In a medium with low ionic strength, the solubility of a protein can be decreased by increasing or decreasing the salt concentration. Salts with small, highly charged ions are more effective than those with large, lowly charged ions.

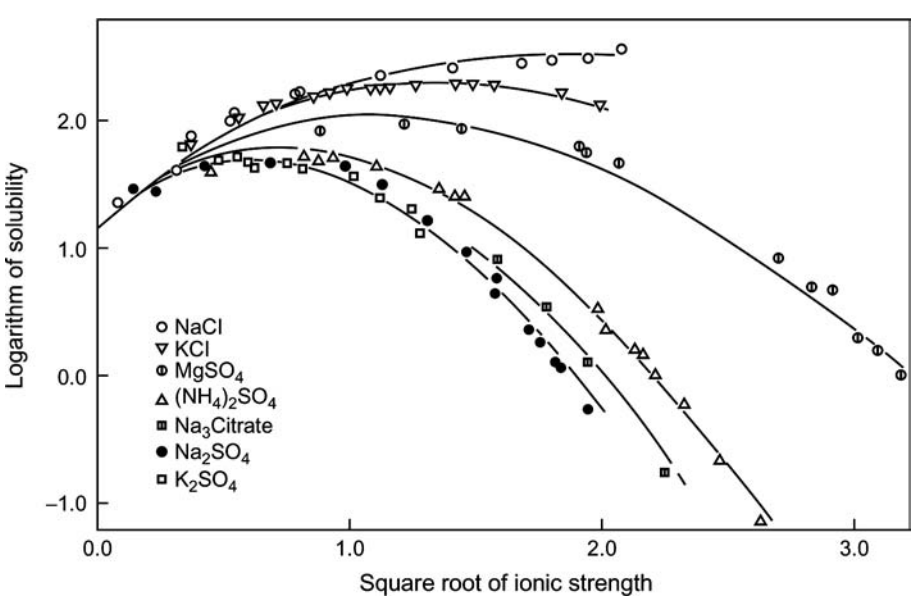

Fig. 5 Solubility of carboxyhemoglobin in various electrolytes at 25 °C. (Reproduced by permission of the American Society for Biochemistry and Molecular Biology, from Green, A.A. (1932) *J. Biol. Chem.* **95**, 47–66.)

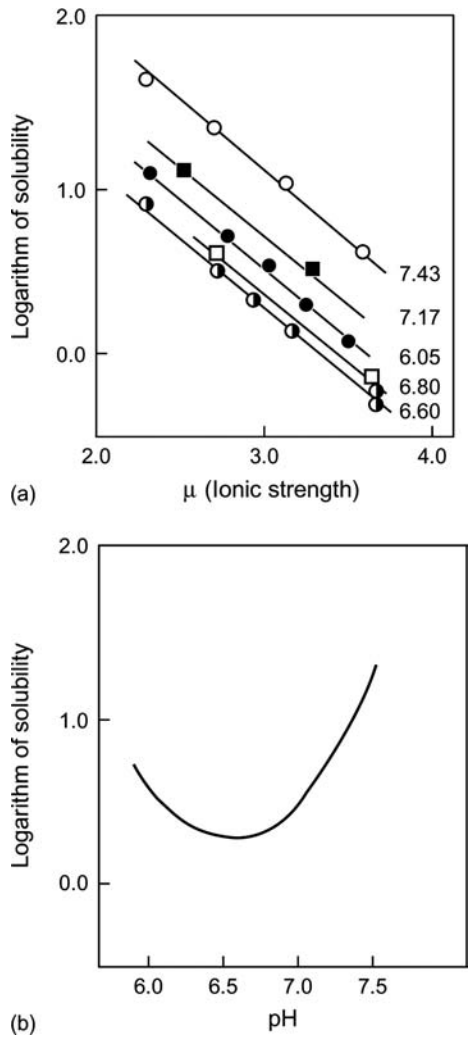

Fig. 6 (a) Solubility of hemoglobin at different pH values in concentrated phosphate buffers; (b) extracted from (a). (Reproduced by permission of the American Society for Biochemistry and Molecular Biology, from Green, A.A. (1931) *J. Biol. Chem.* **93**, 495–516.)

Ammonium sulfate is often used because of its high solubility.

1.2.2 pH and Counterions

A protein is more soluble when its net charge is larger. The minimum solubility is found at the isoelectric point. The net charge is zero and hence the packing in the solid state (in the crystal) is possible owing to electrostatic interactions without the accumulation of a net charge of high energy. All "salting-out" curves are parallel, K_s remains constant, and S_0 varies with pH (Fig. 6a and b). In some cases, the isoelectric point is different at low and high ionic strength owing to the interactions of the protein with counterions, which can cause a net charge at the pH of the isoelectric point.

1.2.3 Temperature

Many factors governing protein solubility are temperature dependent. The dielectric constant decreases with increasing temperature. In the solution energy, $\Delta G = \Delta H - T\Delta S$, the entropy term has an increasing influence with increasing temperature. The temperature coefficient of the solubility depends on other conditions (ionic strength, presence of organic solvents, etc.).

At high ionic strength, most proteins are less soluble at 25 °C than at 4 °C, for example the temperature coefficient is negative. The opposite is valid for low ionic strength.

1.2.4 Organic Solvents

The presence of organic solvents leads to a decrease in the dielectric constant. This causes an augmentation of the electric attraction between opposite charges on the surface of the protein molecule and hence to a reduction in solubility. In general, the solubility of a protein is reduced in the presence of an organic solvent if the temperature decreases. Often, organic

solvents denature proteins. Therefore, one should work at low temperatures.

1.3 Experimental Techniques

The whole field of macromolecular crystallography has been excellently reviewed in Volumes 114 and 115 and Volumes 276 and 277 of *Methods in Enzymology*. A collection of review articles concerning the theory and practice of crystallization of biomacromolecules is given in Part A of Carter and Sweet.

A protein preparation to be used in crystallization should be "pure" or "homogeneous" at a level that established chromatographic methods are providing (protein content $\geq$95%). Furthermore, it should meet the requirements of "structural homogeneity." These requirements can be enumerated as follows. It is first necessary to prepare the protein in an isotypically pure state free from other cellular proteins. It may then be necessary to maintain the homogeneity of the protein preparation against covalent modification during crystallization by adding inhibitors of sulfhydryl group oxidation, by proteolysis, and by the action of reactive metals. It may be necessary to suppress the slow denaturation/aggregation of the protein and to restrict its conformational flexibility to reduce the entropic barrier to crystallization presented by extensive conformational flexibility.

For the crystallization of biomacromolecules, a broad spectrum of crystallization techniques exists. The most common techniques are described here. The oldest and simplest method is batch crystallization (Fig. 7a). In batch experiments, vials containing supersaturated protein solutions are sealed and left undisturbed. In microbatch methods, a small (2–10 μL) droplet containing both protein and precipitant is immersed in an inert oil, which prevents droplet evaporation. In the case that ideal conditions for nucleation and growth are different, it is useful to undertake the separate optimization of these processes. This can be done by seeding, a technique where crystals are transferred from nucleation conditions to those that will support only growth (Fig. 7b). For macroseeding, a single crystal is transferred to an etching solution, then to a solution of optimal growth. In microseeding experiments, a solution containing many small seed crystals, occasionally obtained by grinding a larger crystal, is transferred to a crystal growth solution.

The method of crystallization by vapor diffusion is depicted in Fig. 8(a). In this method, unsaturated precipitant-containing protein solutions are suspended over a reservoir. Vapor equilibration of the droplet and reservoir causes the protein solution to reach a supersaturation level where nucleation and initial crystal growth occur. Changes in soluble protein concentration in the droplet are likely to decrease supersaturation over the time course of the experiment. The vapor diffusion technique can be carried out as hanging drop or sitting drop method.

In crystallization by dialysis, the macromolecular concentration remains constant as in batch methods (Fig. 8b) because the molecules are forced to stay in a fixed volume. The solution composition is changed by diffusion of low-molecular-weight components through a semipermeable membrane. The advantage of dialysis is that the precipitating solution can be easily changed. Dialysis is also uniquely suited to crystallizations at low ionic strength and in the presence of volatile reagents such as alcohols.

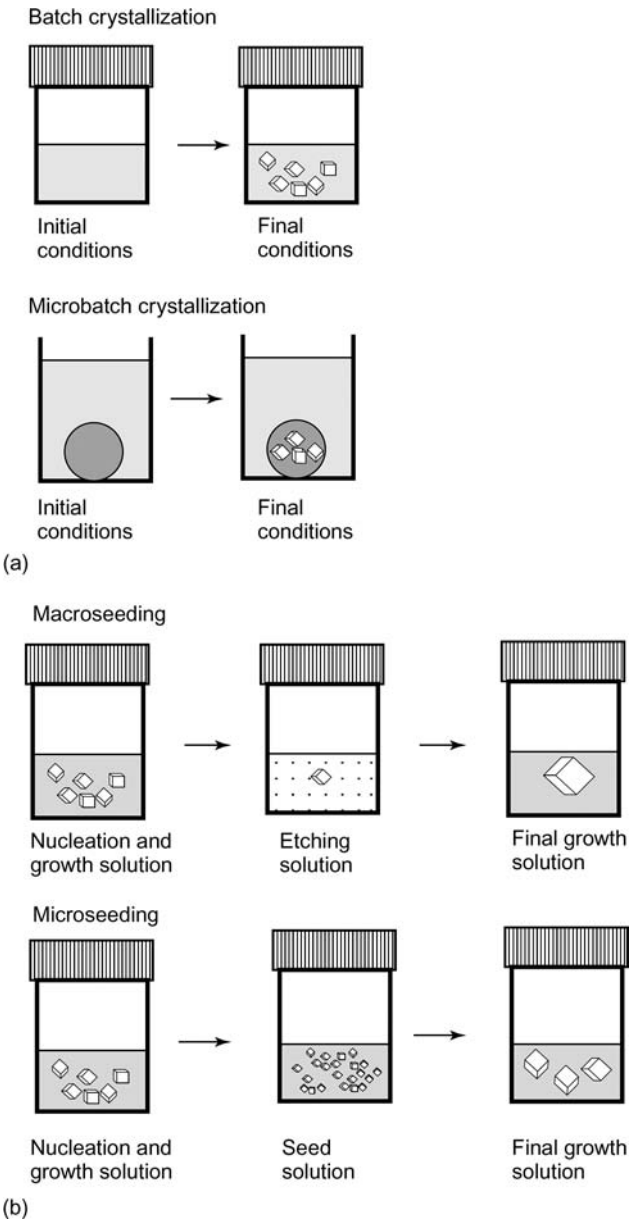

Fig. 7 Schematic presentation of (a) batch crystallization and (b) seeding techniques. (Reproduced by permission of Academic Press, Inc., from Weber, P.C. (1997) *Methods Enzymol*. **276**, 13–23.)

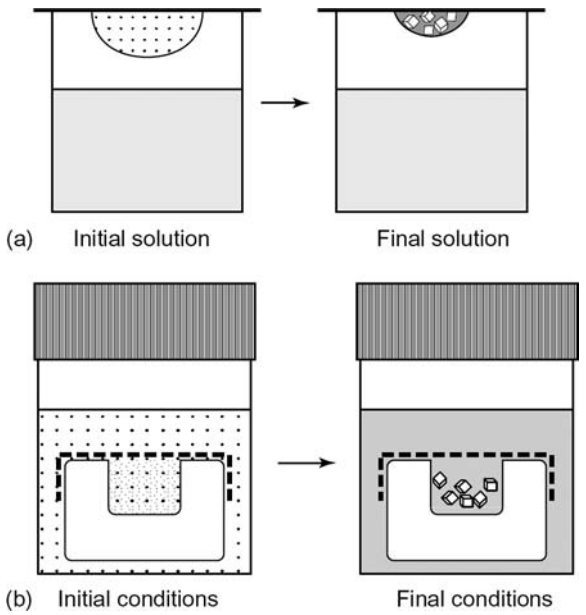

Fig. 8 Schematic representation of (a) vapor diffusion and (b) dialysis. (Reproduced by permission of Academic Press, Inc., from Weber, P.C. (1997) *Methods Enzymol.* **276**, 13–23.)

1.4 Crystallization Screenings

Screening schemes have been developed that change the most common parameters of this multiparameter problem such as protein concentration, the nature and concentration of the precipitant, pH, and temperature. Each screening can be extended by adding specific additives in low concentrations that affect the crystallization. Sparse matrix crystallization screens are widely applied. The sparse matrix formulation allows one to screen efficiently a broad range of the most popular and effective salts (e.g. ammonium sulfate, sodium and potassium phosphate, sodium citrate, sodium acetate, lithium sulfate), polymers (e.g. poly(ethyleneglycol) (PEG) of different molecular masses (from 400 to 8000)), and organic solvents (e.g. 2,4-methylpentanediol (MPD), 2-propanol, ethanol) versus a wide range of pH. Another approach is the systematic screening of the statistically most successful precipitants. A single precipitant is screened at four unique concentrations versus seven precise levels of pH between 4 and 10. Such grid screens can be done with ammonium sulfate, PEG 6000, MPD, and PEG 6000 in the presence of 1.0 M lithium chloride or sodium chloride. For the crystallization of membrane proteins, for each detergent that is necessary to make the membrane protein soluble, a whole grid screen or sparse matrix screen must be constructed. In principle, all three techniques can be applied for the different screening schemes, but mostly the vapor diffusion technique is applied because it is easy to use and the protein consumption is low. For a typical broad screening, about 2 mg of protein is sufficient. Chryschem plates (sitting drop) or Linbro plates (hanging drops) may be

used for the vapor diffusion crystallization screening experiments. Once crystals have been obtained, their size and quality can be optimized by additional fine screens around the observed crystallization conditions. General rules do not exist that indicate which method for crystallization one has to use for which type of protein. Suggestions for crystallization conditions to be tested can be obtained from the Biological Macromolecule Crystallization Database.

2 Experimental Techniques

2.1 X-ray Sources

2.1.1 Conventional X-ray Generators

X-rays are produced when a beam of high-energy electrons, which have been accelerated through a voltage V in a vacuum, hit a target. An X-ray tube run at voltage V will emit a continuous X-ray spectrum with a minimum wavelength given by Eq. (4):

$$\lambda_{\min} = \frac{hc}{eV} = \frac{12398}{V} \qquad (4)$$

with λ in angströms ($1\ \text{Å} = 10^{-10}$ m) and V in volts. The critical voltage, V_0, which is required to excite the characteristic line of a particular element, can be calculated from the corresponding wavelength for the appropriate absorption edge. For the copper absorption edge, $\lambda_{ae} = 1.380$ Å. Hence, we have (Eq. 5)

$$V_0 = \frac{12398}{\lambda_{ae}} = 8.98\ \text{kV} \qquad (5)$$

Provided that $V > V_0$, the characteristic line spectra will be produced (Fig. 9). The oldest and cheapest X-ray sources are sealed X-ray tubes. The cathode and anode are situated under vacuum in a sealed glass tube, and the heat generated at the anode is removed by a water-cooling system. For the generation of higher intensities, as needed in protein crystallography, one has to use a rotating anode (Fig. 10). Here the anode is rotated, which allows a higher power loading at the focal spot. In protein crystallography, copper targets are usually taken. The used take-off angle is near 4°, which results in apparent focal spot sizes of about 0.3×0.3 mm.

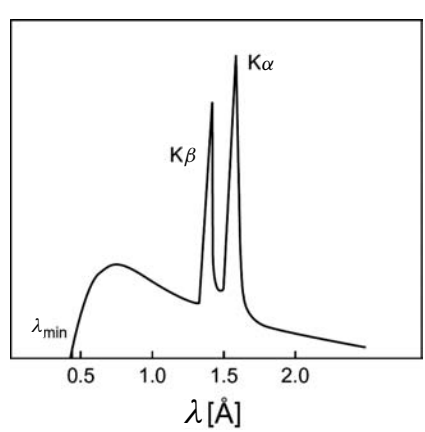

Fig. 9 X-ray spectrum emitted from a copper anode. It shows the continuous "Bremsspektrum" starting at λ_{min} and the two characteristic copper lines λ Kα = 1.5418Å (superposition of λ Kα_1 = 1.5405 Å and λ Kα_2 = 1.5443 Å) and λ Kβ = 1.3922 Å.

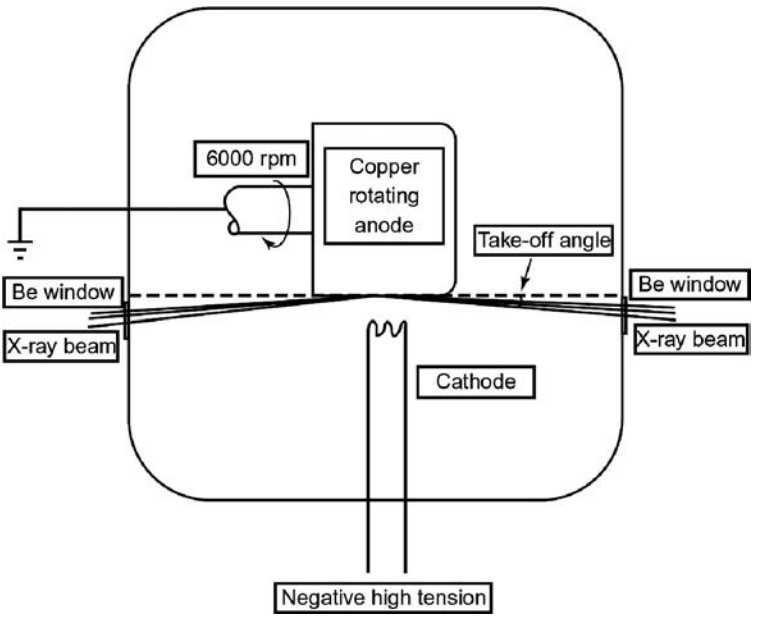

Fig. 10 Schematic drawing of a rotating anode tube. Take-off angle is near 4°. For copper, the tube is normally operated at 50-kV high tension and 100-mA cathode current.

2.1.2 Synchrotron Radiation

As electrically charged particles such as electrons or positrons of high energy are kept under the influence of magnetic fields and travel in a pseudocircular trajectory, synchrotron radiation is emitted and can be used in many different types of experiments. The particles are injected into the storage ring directly from a linear accelerator or through a booster ring. They circulate in a high vacuum for several hours at a relative constant energy. To keep the bunched particles traveling in a nearly circular path, a lattice of bending magnets is set up around the storage ring. As the particle beam traverses each magnet, the path of the beam is altered, and synchrotron radiation is emitted. The loss of energy of the particle beam is compensated by a radiofrequency input at each cycle. The synchrotron radiation can be channeled through different beamlines for use in research.

Other types of magnets – insertion devices called *wigglers* and *undulators* – can be assembled in the storage ring. Unlike the bending magnets, the primary purpose of which is to maintain the circular trajectory, wigglers and undulators are used to increase the intensity of the emitted radiation. Bending magnets and wigglers cause a continuous spectrum of radiation.

In contrast, the radiation produced by an undulator has a discontinuous spectrum, and can be tuned to various wavelengths. The importance of synchrotron radiation for macromolecular crystallography lies in the high brilliance (photons s^{-1} $mrad^{-2}$ mm^{-2} per $\Delta\lambda/\lambda$; that is, how small is the source and how well collimated are the X-rays?) of the beam, the

high intensity, and the tunability of the wavelength in the relevant range from 0.5 to 3.0 Å. The time structure of the beam is of interest for time-resolved crystallography. The particles circulate in bunches with widths of 50 to 150 ps and repeat every few microseconds.

About 15 synchrotron radiation facilities equipped with beamlines for macromolecular crystallography are available throughout the world operated at energies from about 1.5 to 6 to 8 GeV for third-generation machines. An aerial view of the European Synchrotron Radiation Facility (ESRF) in Grenoble, a third-generation machine, is shown in Fig. 11. The ESRF storage ring is operated at 6 GeV and has a circumference of 844.39 m. Its critical wavelength, λ_c, is 0.6 Å.

2.1.3 Monochromators

In the majority of applied diffraction techniques, monochromatic X-rays are used. Therefore, the emitted white radiation of X-rays must be further monochromatized. With copper Kα radiation generated by a sealed or rotating anode tube, the Kβ radiation can be removed with a nickel filter. Much better results can be achieved with a monochromator. The simplest monochromator is a piece of a graphite crystal that reflects the copper Kα radiation at a Bragg angle of 13.1° and a glancing angle of 26.2°. Improved beam focusing is obtained by a double mirror system. The mirror assembly is composed of two perpendicular bent nickel-coated glass optical flats, each with translation, rotation, and slit components housed in a helium gas-flashed chamber, which is commercially available (Molecular Structure Corporation, The Woodlands, TX, USA). The prototype and basic theory in the use of this system were discussed in detail by Phillips and Rayment.

For synchrotron radiation with its much higher intensity, germanium or silicon single crystals can be applied as monochromators, which filter out a bandwidth of $\delta\lambda/\lambda$ from 10^{-4} to 10^{-5}, two orders of magnitude smaller than that with graphite. Single or double monochromators can be used, which are either flat or bent. The bent monochromators have the advantage that

Fig. 11 Aerial view of ESRF in Grenoble. Courtesy of ESRF.

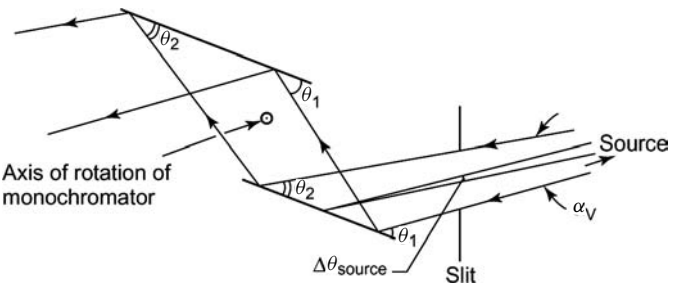

Fig. 12 Schematic drawing of a double monochromator system. (Reproduced by permission of Cambridge University Press, from Helliwell, J.R. (1992) *Macromolecular Crystallography with Synchrotron Radiation*, Cambridge University Press, Cambridge.)

they simultaneously focus the beam. The double monochromator (Fig. 12) has the advantage that the emergent monochromatic beam is parallel to and only slightly displaced from the incident synchrotron radiation beam. This makes necessary only small adjustments of the X-ray optics and detector arrangement when it is tuned to another wavelength compared to a single monochromator where the whole X-ray diffraction assembly must be moved.

2.2 Detectors

2.2.1 General Components of an X-ray Diffraction Experiment

A principal arrangement for a macromolecular X-ray diffraction experiment is depicted in Fig. 13. The primary beam leaves the X-ray source and passes the X-ray optics, which may be a simple collimator or the various types of monochromators or mirror systems described above,

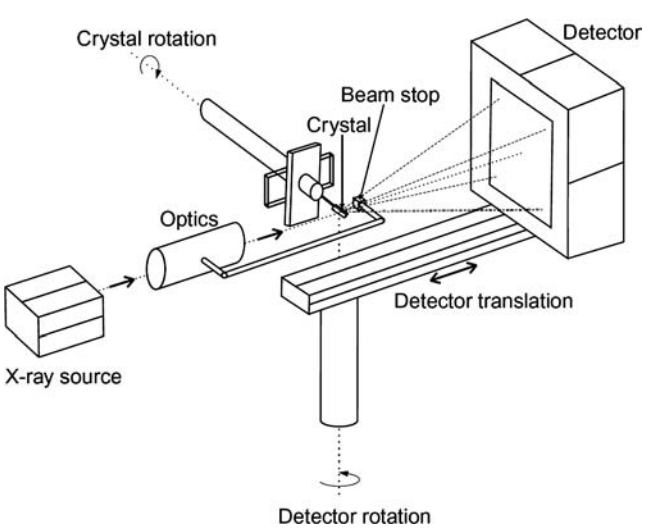

Fig. 13 Principal arrangement for a macromolecular X-ray diffraction experiment.

and terminated with a collimator. The crystal is mounted on a goniometer head either in a quartz capillary or in a cryo-loop shock-frozen at low temperature. The goniometer head is attached to a device that can perform spatial movements of the crystal around the center of the crystal. The simplest kind of such a movement is the rotation of the crystal about a spindle axis as indicated in Fig. 13. This device can be a multiple axis goniostat (2–4 axes), which allows the crystal to be brought into any spatial orientation around its center. The X-ray detector, which registers the diffracted intensities, is mounted on a device that permits the translation and rotation of the detector. If the active area of the detector is large enough to collect all generated diffracted beams at a given wavelength, detector rotation is not necessary and the detector is arranged normal to the primary beam. A small piece of lead is placed in the path of the primary beam just behind the crystal to prevent damage to the detector and superfluous gas scattering.

The classical detectors in macromolecular crystallography have been photographic films and single-photon counters. The photographic films were used on specially designed X-ray cameras and the single-photon counters on four-circle diffractometers. The main disadvantage of these detectors was their low sensitivity and with films the limited dynamic range (1 : 200). Over the last 15 years, powerful detectors have been developed, which will be discussed briefly. These new detectors have almost completely replaced photographic films and single-photon counters.

2.2.2 Image Plates

An image plate (IP) consists of a support (either a flexible plastic plate or a metal base) coated with a photostimulable phosphor (150 μm) and a protective layer (10 μm). The photostimulable phosphor is a mixture of very thin crystals of $BaF(Br,I):Eu^{2+}$ and an organic binder. This phosphor can store a fraction of the absorbed X-ray energy by electrons trapped in color centers. It emits photostimulated luminescence, whose intensity is proportional to the absorbed X-ray intensity, when later stimulated by visible light. The wavelength of the photostimulated luminescence ($\lambda \approx 390$ nm) is reasonably separated from that of the stimulating light ($\lambda \approx 633$ nm, in practice a red laser), allowing it to be collected by a conventional high quantum efficiency photomultiplier tube. The output of the photomultiplier is amplified and converted to a digital image, which can be processed by a computer. The residual image on the IP can be erased completely by irradiation with visible light, to allow repeated use.

IPs have several excellent performance characteristics as integrating X-ray area detectors that make them well suited in X-ray diffraction. The sensitivity is at least 10 times higher than for X-ray films and the dynamic range is much broader ($1:10^4$–10^5). Important for synchrotron radiation is their high sensitivity at shorter wavelengths (e.g. 0.65 Å). A disadvantage is the relatively long readout times for each exposure (from 45 s to several min). IP diffractometer systems are commercially available from several companies. All systems work reliably and deliver good-quality data. A photograph of the newest IP system produced by Mar Research (Norderstedt, Germany) is shown in Fig. 14.

2.2.3 Gas Proportional Detectors

As X-ray counters, gas proportional detectors provide unrivaled dynamic range and sensitivity for photons in the range important for macromolecular crystallography.

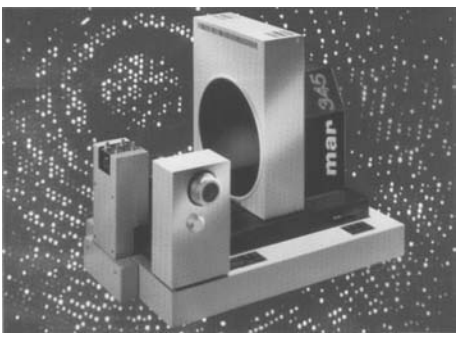

Fig. 14 Mar 345 IP diffractometer system from Mar Research, Norderstedt, Germany. The circular plate rotates for scanning and the laser is moved along a radial line. Courtesy of Mar Research.

The classical gas proportional detector is a multiwire proportional chamber (MWPC), widely used as an in-house detector with conventional X-ray sources. Two MWPC diffractometer systems are commercially available. Gas proportional detectors use as a first step the absorption of an X-ray photon in a gas mixture high in xenon or argon. This photoabsorption produces one electron–ion pair whose total energy is just the energy of the initial X-ray photon. The ion returns to its neutral state either by emission of Auger electrons or by fluorescence. Since the kinetic energy of these first electrons is far greater than the energy of the first ionization level of the xenon or argon atoms, fast collisions with atoms (or molecules) in the gas very quickly produce a cascade of new electron–ion pairs in a small region extending over a few hundred micrometers around the conversion point. The total number of primary electrons that are produced during this process is proportional to the energy of the absorbed X-ray photon and is thus a few hundred for ~10-keV photons. These primary electrons then drift to the nearest anode wire where an ionization avalanche of as many as 10 000 to 1 000 000 ion pairs results. The motion of the charged particles in this avalanche (chiefly the motion of the heavy positive ions away from the anode wire) causes a negative-going pulse on the anode wire and positive-going pulses on a few

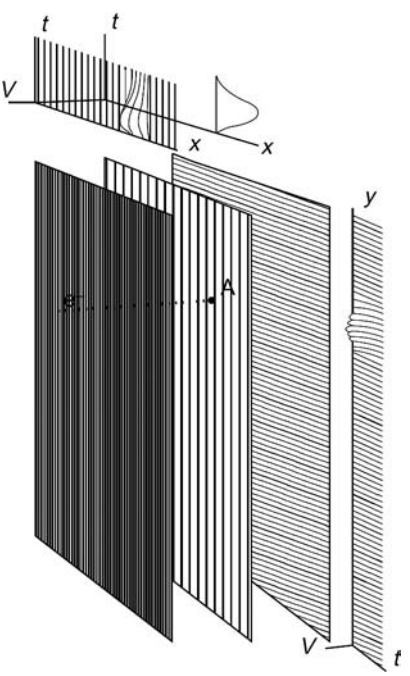

Fig. 15 Expanded view of an MWPC, showing the anode plane sandwiched between the two cathode planes. A is the position of the avalanche. The centers of the induced charge distributions are used to determine the coordinates, *x* and *y*, of the avalanche. (Reproduced by permission of Academic Press, Inc., from Kahn, R., Fourme, R. (1997) *Methods Enzymol.* **276**, 244–268.)

of the nearest wires in the back (cathode) wire plane (see Fig. 15).

Disadvantages of the MWPC detector are the limited counting rate due to the buildup of charges in the chamber and limitations in the readout electronics and the lower sensitivity at shorter wavelengths. This makes the application of MWPCs with synchrotron radiation poorly effective.

2.2.4 Charge-coupled Device-based Detectors

A remarkable development for the use with synchrotron radiation is the design and construction of charge-coupled device (CCD) detectors. CCDs were developed originally as memory devices, but the observation of localized light-induced charge accumulation in CCDs quickly led to their development as imaging sensors. These CCD detectors are integrating detectors like the conventional X-ray sensitive film, IPs, and analog electronic detectors using either silicon intensified target (SIT) or CCD sensors. Integrating detectors have virtually no upper rate limits because they measure the total energy deposited during the integration period (although individual pixels may become saturated if the signal exceeds its storage capacity).

The first commercially available analog electronic detector was the fast area television detector (FAST) detector produced by Enraf-Nonius (Delft, The Netherlands). This detector contained a SIT vidicon camera as an electronically readable sensor. The SIT vidicon exhibits higher noise than CCDs, which have therefore replaced SIT sensors during the past few years. Because of their high intrinsic noise, detectors with SIT vidicon sensors need an analog image-amplification stage and this limits the overall performance of such detectors. Several CCD detector systems have also been developed that incorporate image intensification. The most important development in detector design for macromolecular crystallography has been the incorporation of scientific-grade CCD sensors into instruments with no image intensifier. These detector designs are based on direct contact between the CCD and a fiber-optic taper. There are several commercial systems available based on this construction (Mar Research, Norderstedt, Germany; Hamlin Detector).

A schematic representation of such a detector is shown in Fig. 16. An X-ray phosphor (commonly $Gd_2O_2S:Tb$) is attached to a fiber-optic faceplate, which is tightly connected to a fiber-optic taper. The X-ray sensitive phosphor surfaces at the front convert the incident X-rays into a burst of visible-light photons. Although it is possible to permit the X-rays to strike the CCD directly, this method has several drawbacks, such as radiation damage to the CCD, signal saturation, and poor efficiency. The use of a larger phosphor as active detector area and the demagnifying fiber-optic taper is also necessary because the size of the scientific-grade CCD sensors is not as large as needed for the demands of the X-ray diffraction experiment. The fiber-optic taper is then bonded to the CCD, which is connected to the electronic readout system. The CCD must be cooled to temperatures ranging from -40 to $-90\,^{\circ}C$, depending on the various systems. The great advantage of CCD detectors is their short readout time, which lies in the range from 1 to a few seconds.

2.3 Crystal Mounting and Cooling

2.3.1 Conventional Crystal Mounting

The purpose of crystal mounting is to isolate a single crystal from its growth

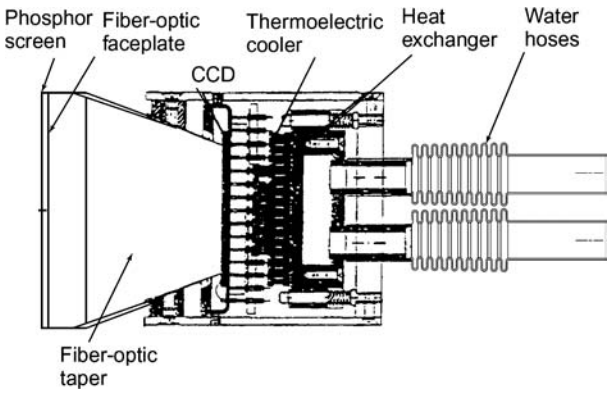

Fig. 16 Schematic representation of a CCD/taper detector. (Reproduced by permission of Academic Press, Inc., from Westbrook, E.M., Naday, I. (1997) *Methods Enzymol.* **276**, 268–288.)

medium so that it can be used in the X-ray diffraction experiment to study its diffraction properties. It is important that the manipulation of the crystal introduces as little damage as possible to its three-dimensional structure. The most important aspect of crystal mounting is to preserve the crystal in its state of hydration. This is accomplished by sealing the crystal in a thin-walled (0.001-mm thick) glass or quartz capillary tube. The important steps in conventional crystal mounting are illustrated in Fig. 17(a–c). The crystal must be dislodged from the surface on which it grew, and then it may be drawn into the capillary using suction from a small-volume (0.25 mL) syringe, micropipet, or mouth aspirator, which are connected to the funnel of the capillary by a flexible plastic hose of appropriate diameter. Next, the capillary should be inverted to allow the crystal to fall to the inner meniscus. Then, the surrounding solution may be removed using thin strips of filter paper or by a small glass pipet. The extent to which the crystal should be dried must be determined by experience. The final step is to place a small volume of mother liquor in the capillary and seal both ends. The capillary is then glued to a metal base, which can be attached to a goniometer head.

2.3.2 Cryocrystallography

Many macromolecular crystals suffer from radiation damage when exposed to X-rays with energies and intensities as used in macromolecular X-ray diffraction experiments with both conventional sources and synchrotron radiation. A possibility for reducing radiation damage of the crystal during the measurement is to cool the crystal to low temperatures, usually to 100 K. For this purpose, the crystal is flash-frozen to prevent ice formation or damage to the crystal. One method of crystal treatment is the removal of external solution by transferring the crystal in a small drop to a hydrocarbon oil and either teasing the liquid away or drawing it off with filter paper or a small pipet. The oil-coated crystal is then mounted on a glass fiber or small glass "spatula." Oil protects the crystal from drying and acts as an adhesive that hardens on cooling to hold the sample rigidly. Much more frequently used

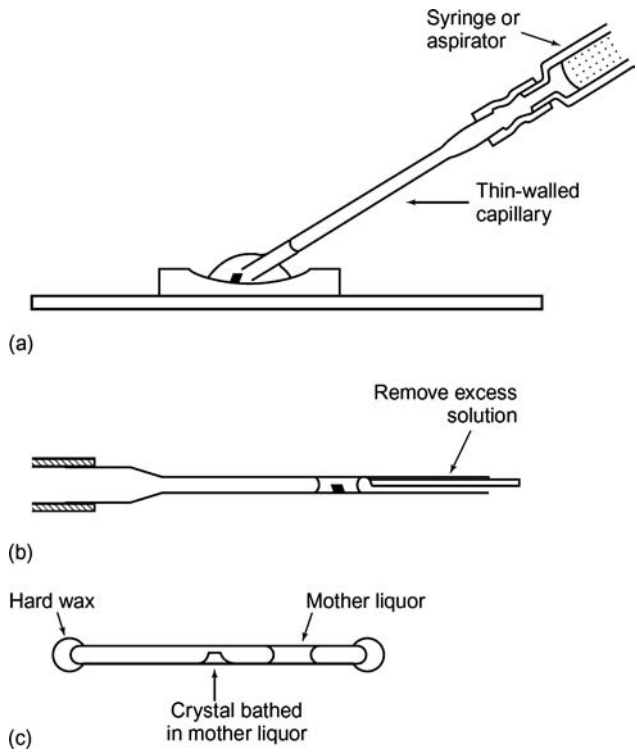

Fig. 17 Mounting of a crystal in a glass capillary. (Reproduced by permission of Academic Press, Inc., from Rayment, I. (1985) *Methods Enzymol.* **114**, 136–140.)

is a technique in which the crystal is suspended in a film of mother liquor in a small loop. This method avoids problems with damage by the oil or mechanical damage when removing the external liquid, and it has proven successful for most samples. It does, however, require the use of a cryoprotectant to prevent ice formation. The most commonly used cryoprotectants are glycerol, PEG of different molecular weights, glucose, and MPD.

The loop is produced from fine fibers that permit unobstructed data collection in nearly all sample orientations. The crystal is held within the loop, suspended in a thin film of cryoprotectant-containing harvest buffer. The loop is supported by a fine wire or pin, which itself is attached to a steel base used for placing the assembly on a goniometer head and in storing mounted crystals. Once in the loop, the crystal is cooled to a temperature at which the increasing viscosity of the liquid prevents molecular rearrangement. The rate of cooling must be rapid enough to reach this point before ice-crystal nucleation occurs. Two methods are used: cooling directly in the gas stream of a cryostat or plunging the crystal into a cryogenic liquid. The first method is explained in Fig. 18(a) and (b). The loop assembly (with crystal) is attached to the goniometer head, with the cold stream deflected. Then, the cold stream is unblocked to flash-freeze

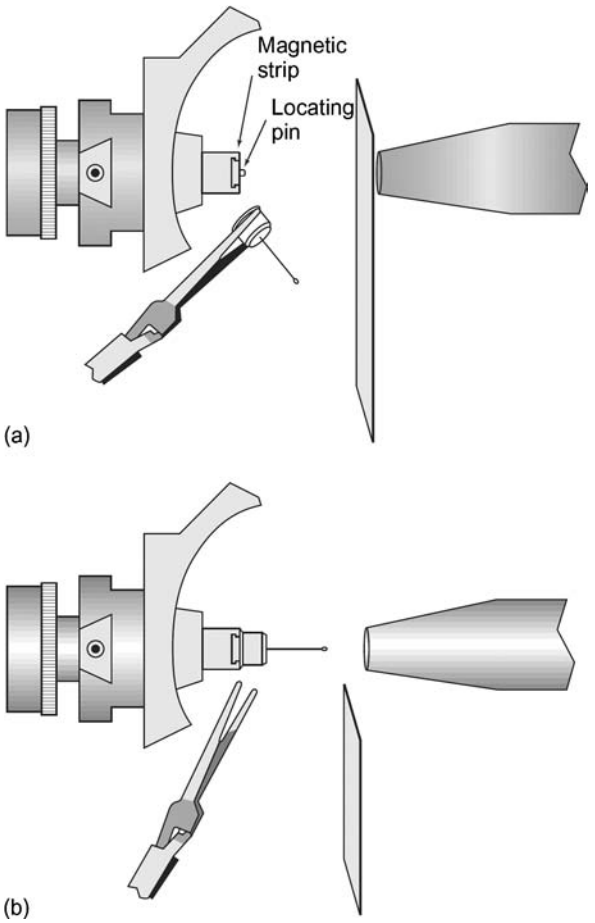

Fig. 18 Flash cooling in the direct cold gas stream of a cryostat. (Reproduced by permission of Academic Press, Inc., from Rodgers, D.W. (1997) *Methods Enzymol.* **276**, 244–268.)

the crystal. In this gas-stream position, the goniometer head must be heated to prevent ice formation on the goniometer head. Cryostats and cryocrystallographic tools are commercially available.

Cryocrystallography has had a great impact on macromolecular crystallography by dramatically increasing the lifetime of a crystal during the X-ray experiment allowing, for example, the collection of several data sets from one crystal at different wavelengths using synchrotron radiation.

2.3.3 Crystal Quality Improvement by Humidity Control

The crystal quality is of decisive significance for a successful X-ray crystal structure determination. Two principal cases must be distinguished: (1) the crystals diffract to a resolution only (>4.5 Å), which does not allow a structure determination at all and (2) the crystal quality is good enough to elucidate its 3-D structure (resolution <3.5 Å but not better than 2.5 Å) but a structure determination at

higher resolution and accuracy therewith would be necessary, for example, for the characterization of a metal center in a metalloprotein. The control of the crystal packing via the solvent content of a crystal is a useful approach for the improvement of the crystal quality of biological macromolecules. For this purpose, a so-called Free Mounting System (FMS) has been developed and successfully applied. The principal construction of the FMS is shown in Fig. 19 (a). A main part of the FMS is the humidifier unit, which consists of the humidifier, the control, and power electronics. A stream of humid air with defined moisture is produced. A flexible teflon tube transports the humid air to the crystal holder, which is depicted in Fig. 19 (b). A very compact construction allows the sample to be mounted in a controlled environment with minimal restriction for the X-ray measurement. The head part is freely rotatable relative to the insert, without axial movement. A heating element and a temperature sensor are integrated into the head part. The humid air stream through the head part is adjusted to the temperature of the head part, independent of the ambient temperature. The crystal may either be mounted in a patch-clamp pipette or conventional cryo-loop (Fig 19 (c)).

In a typical experiment to increase the crystal quality, X-ray crystal diffraction is monitored at various defined crystal humidities. Usually, a positive effect is observed at lower humidity, which causes shrinkage of the unit cell volume and allows different and possibly more favorable crystal contacts. After having found the optimal condition, the crystal, mounted in a loop, can be shock-frozen for a subsequent data collection. A remarkable improvement of crystal quality by using the FMS could be observed with about 30% of different projects under investigation.

2.4 Data-collection Techniques

2.4.1 Rotation Method

Most macromolecular X-ray diffraction systems use the rotation method for data collection. For each crystal, a reciprocal lattice can be constructed, which is very useful in the interpretation of crystallographic crystal diffraction experiments. Diffraction theory (discussed later) tells us that an X-ray reflection is generated when a point of this reciprocal lattice lies on a sphere of radius $1/\lambda$ whose origin is $1/\lambda$ away from the origin of the reciprocal lattice in the direction of the primary beam (Fig. 20). The direction of such a diffracted beam is along the connection of the center of the so-called Ewald sphere (radius $1/\lambda$) and the intersection of the reciprocal lattice point on the Ewald sphere. Owing to certain factors, which will be discussed later, the apparent reciprocal lattice extends to a given radius only, which defines the resolution sphere. To bring all the reciprocal lattice points within the resolution sphere into the reflection position, the crystal must be rotated around its center. Nearly all macromolecular X-ray diffraction systems apply the rotation technique in the normal beam case where the rotation axis is normal to the incident X-ray beam. Rotating the crystal around 360° brings all reciprocal lattice points within the resolution sphere in the reflection position except for the region between the rotation axis and the Ewald sphere, which is therefore called the *blind region*. This region can be collected when the crystal has been brought into another orientation. The diffracted beams are usually registered with a flat detector at distance D from the crystal, which is also normal to the primary beam. To avoid overlapping of reflection

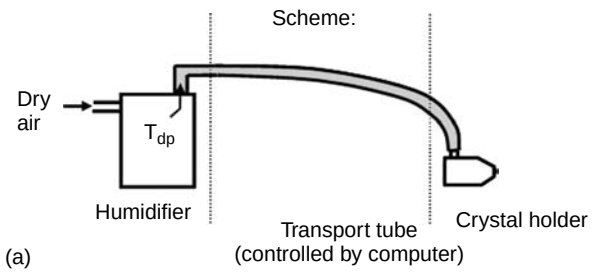

(a)

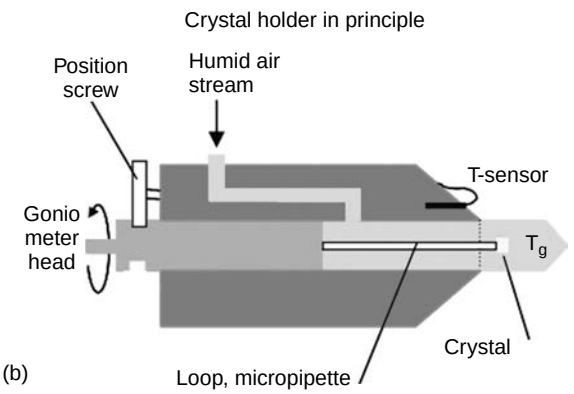

(b)

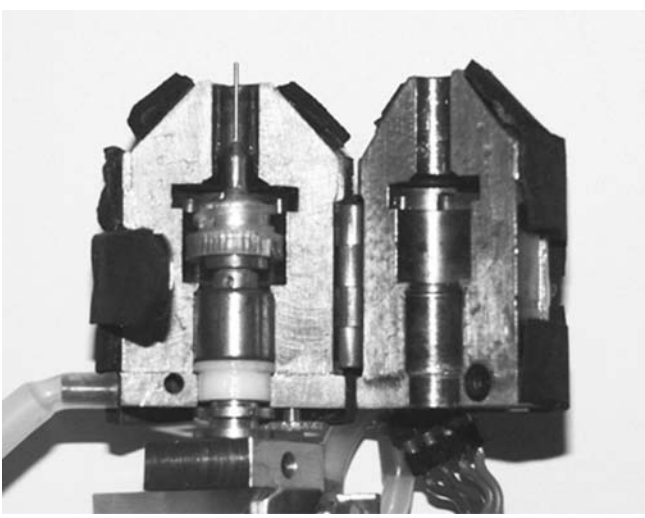

(c)

Fig. 19 The Free Mounting System: (a) principal construction; (b) schematic view of the crystal holder; (c) opened crystal holder with magnetic base and mounting loop. T_{dp} – temperature of dew point, T_g – temperature of goniometer head.

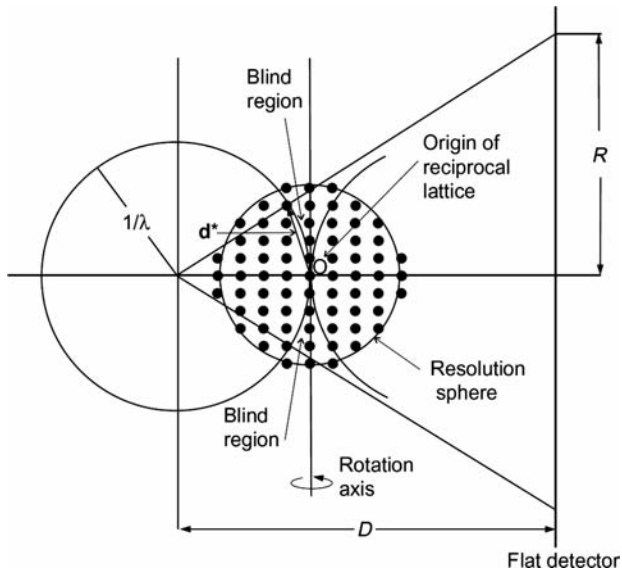

Fig. 20 Diffraction geometry in the rotation method usually applied in macromolecular X-ray diffraction systems.

spots on the detector, the crystal is rotated by rotation-angle increments, which can vary from tenths of degrees to 1 or to a few degrees, depending on the size of the crystal unit cell, crystal mosaicity, beam collimation, and other factors. Each individual exposure is processed and electronically stored in a computer. These raw data images are evaluated subsequently with relevant computer programs (discussed in some detail later) to give the intensities and geometric reference values (indices) for each collected intensity.

2.4.2 Precession Method

The rotation method delivers a distorted image of the reciprocal lattice for each geometry of the detector (flat or curved) and orientation with respect to the rotation axis. An undistorted image of the reciprocal lattice can be obtained by using the precession method. The principle of this technique is shown in Fig. 21. The detector is a flat film. During the motion of a given reciprocal lattice plane (in Fig. 21, a so-called zeroth plane going through the origin O of the reciprocal lattice), the flat detector must always be parallel to this reciprocal lattice plane to obtain an undistorted image of this plane. The normal of the reciprocal lattice plane and, consequently, also the detector are inclined with respect to the primary X-ray beam by an angle μ. When the normals of the reciprocal lattice plane and the detector carry out a concerted precession motion of angle μ around the primary X-ray beam, a circular region of the reciprocal lattice plane is registered on the detector (these regions are shown as dashed circles in Fig. 21). In a precession camera construction, the crystal and the film cassette are both held in a universal joint, which are linked so that the film and crystal move together in phase with precession angle μ. In Fig. 21, the joints are symbolized as forks and their linkage by a line. Parallel to

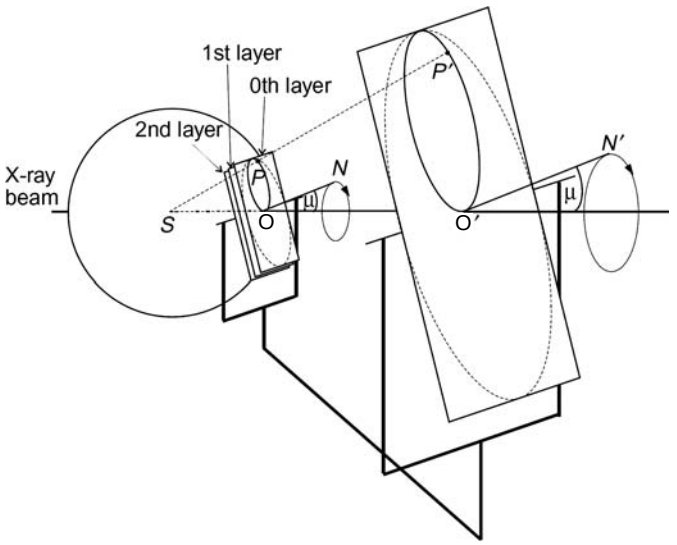

Fig. 21 Principle of the precession method (modified from Buerger, M.J. (1964) *The Precession Method in X-ray Crystallography*, John Wiley & Sons, New York).

the zeroth reciprocal lattice plane is a set of lattice planes that also carry out this precession movement. The parts of them that are swung through the Ewald sphere also give rise to an image on the flat detector (first and second reciprocal lattice layers are also indicated in Fig. 21). The images of these layers would superimpose on the detector, and the use of the technique in this way is denoted by the screenless precession method. The insertion of a screen with a suitable annular aperture between the crystal and the detector at an appropriate distance can be used to screen out the desired reciprocal layer. This screen is also inclined by the precession angle μ and is coupled to the concerted precession movement of crystal and film cassette. The strength of the precession technique is that in addition to the undistorted imaging of the reciprocal lattice planes, the indexing of the diffraction spots is straightforward and the symmetry of the diffraction pattern is readily obtained by inspection.

For this reason, the precession method has been broadly applied in macromolecular crystallography for a long time. The replacement of film by the new generation of detectors has almost completely stopped the use of the precession method. Nevertheless, the use of a precession camera has great teaching benefits in becoming familiar with the reciprocal lattice concept.

3 Principles of X-ray Diffraction by a Crystal

3.1 Scattering of X-rays by an Atom

A component of the electrical field of the incident wave has the following form (Eq. 6) in free space referred to an origin at $\mathbf{x} = 0$:

$$E(\mathbf{x}) = E_0 \exp[2\pi i(\nu t - \mathbf{x}\mathbf{s}_0)] \qquad (6)$$

where $\mathbf{s}_0$ = wave vector of incident wave, $\nu = c/\lambda$ = frequency, λ = wavelength, and $s_0 = 1/\lambda$ = absolute value of incident wave vector. This wave interacts with a scattering center at position $\mathbf{r}$ (Fig. 22). The electric field component of the incident wave causes the electron at position $\mathbf{r}$ to oscillate. Together with the positively charged nucleus of the atom, which does not oscillate, this can be considered as a classical dipole oscillator. This dipole oscillator emits a spherical wave, which is denoted as a scattered wave and can be given in the following form (Eq. 7 and 8):

$$E_{SC} = CE(\mathbf{r})\frac{\exp(-2\pi i s_0 r_1)}{r_1} \quad (7)$$

$$E_{SC} = CE_0 \exp(-2\pi i \mathbf{s}_0 \mathbf{r}) \times \frac{\exp(2\pi i(\nu t - s_0 r_1)}{r_1} \quad (8)$$

The phase-angle-dependent factor has been omitted. The amplitude of the scattered wave is proportional to the amplitude $E(\mathbf{x})$ of the wave incident at $\mathbf{r}$. This gives the factor $E(\mathbf{r})$. C is a proportionality factor, taking into account the peculiarities of the scattering center. The factor $1/r_1$ considers the conservation of the scattered energy flux. We add all wavelets scattered from different volume elements of an atom to get the total amplitude of the scattered wave at point $\mathbf{R}$ relative to the origin O in the atom.

Equations 9 to 12 follow from Fig. 22:

$$\mathbf{r} + \mathbf{r}_1 = \mathbf{R} \quad (9)$$

$$r_1^2 = (\mathbf{R} - \mathbf{r})^2 = R^2 + r^2 - 2rR\cos(\mathbf{r}, \mathbf{R}) \quad (10)$$

$$r_1 = R\left[1 - \frac{2r}{R}\cos(\mathbf{r}, \mathbf{R}) + \left(\frac{r}{R}\right)^2\right]^{1/2} \quad (11)$$

$$r_1 \approx R\left(1 - \frac{r}{R}\cos(\mathbf{r}, \mathbf{R}) + \ldots\right) \quad (12)$$

where $(r/R)^2$ and higher terms were neglected in the expansion of the square root. It follows that (Eq. 13)

$$r_1 \approx R - r\cos(\mathbf{r}, \mathbf{R}) \quad (13)$$

Now we can combine the spatial phase factors in the equation for E_{SC}. With the approximation for r_1, we obtain Eq. (14):

$$\exp(-2\pi i(\mathbf{s}_0\mathbf{r} + s_0 r_1) = \exp(-2\pi i s_0 \mathrm{R}) \times \exp(-2\pi i(\mathbf{s}_0\mathbf{r} - s_0 r\cos(\mathbf{r}, \mathbf{R})) \quad (14)$$

As $s_0 = s$ and $\mathbf{s}$ is parallel to $\mathbf{R}$, we obtain Eq. (15).

$$s_0 r\cos(\mathbf{r}, \mathbf{R}) = sr\cos(\mathbf{r}, \mathbf{s}) = \mathbf{sr} \quad (15)$$

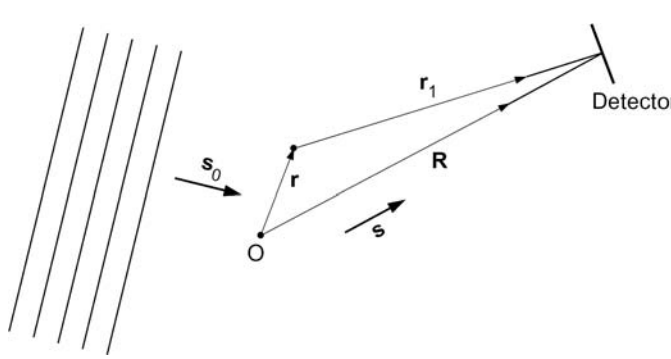

Fig. 22 Scattering of a planar X-ray wave by an electron at position **r** with respect to origin O.

It follows for the phase factor (Eq. 16):

$$\begin{aligned}\exp(-2\pi i(\mathbf{s}_0\mathbf{r} + s_0 r_1) &= \exp(-2\pi i s_0 \mathrm{R}) \\ &\times \exp(-2\pi i(\mathbf{s}_0\mathbf{r} - \mathbf{s}\mathbf{r})) \\ &= \exp(-2\pi i \mathbf{S}_0 \mathrm{R}) \\ &\times \exp(2\pi i(\vec{\mathbf{r}}(\vec{\mathbf{S}} - \vec{\mathbf{S}}_0)) \qquad (16)\end{aligned}$$

Now we can write for the wave scattered by a center at $\mathbf{r}$ (Eq. 17):

$$E_{\mathrm{SC}} = \left(\frac{CE_0 \exp(-2\pi i s_0 R)\exp(2\pi i \nu t)}{R}\right) \times \exp(2\pi i \mathbf{r}(\mathbf{s} - \mathbf{s}_0)) \qquad (17)$$

In a useful approximation, r_1 has been replaced by R in the denominator. In general, the scattering from an atom comes from the distribution of electrons in the atom. If the scattering of a volume element $\mathrm{d}v$ of the atom is proportional to the local electron density ρ ($\mathbf{r}$), then the scattering amplitude will be proportional to the integral (Eq. 18):

$$f(\mathbf{S}) = \int_{\text{vol. of atom}} \rho(\mathbf{r}) \exp(2\pi i \mathbf{r}\mathbf{S})\mathrm{d}v \qquad (18)$$

3.2 Scattering of X-rays by a Unit Cell

A unit cell may contain N atoms at positions of their internal origins at $\mathbf{r}_j$ ($j = 1, 2, 3, \ldots, N$) with respect to the origin of the unit cell (Fig. 23). For atom 1, we obtain Eq. (19):

$$\begin{aligned}\mathbf{f}_1 &= \int_{\text{vol. of atom}} \rho(\mathbf{r}) \exp[2\pi i(\mathbf{r}_1 + \mathbf{r})\mathbf{S}]\mathrm{d}v \\ &= f_1 \exp(2\pi i \mathbf{r}\mathbf{S}) \qquad (19)\end{aligned}$$

with (Eq. 20)

$$f_1 = \int_{\text{vol. of atom}} \rho(\mathbf{r}) \exp 2\pi i \mathbf{r}_1 \mathbf{S}\mathrm{d}v \qquad (20)$$

where f_1 is the atomic form factor for atom 1. It reflects the characteristics of the scattering of the individual atoms and is real if the wavelength of the incident X-ray is not close to an absorption edge of the atom. The atomic form factor f is equal to Z, the ordinary number of the scattering atom, at a diffraction angle of 0° and decreases with increasing diffraction angle. For N atoms, this adds up to the total scattered wave of a unit cell $\mathbf{G}(\mathbf{S})$ (Fig. 24) according to Eq. (21):

$$\mathbf{G}(\mathbf{S}) = \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{r}_j \mathbf{S}) \qquad (21)$$

3.3 Scattering of X-rays by a Crystal

3.3.1 One-dimensional Crystal

In a one-dimensional crystal, the unit cells are separated by the unit cell vector $\mathbf{a}$. The contribution of the scattered wave

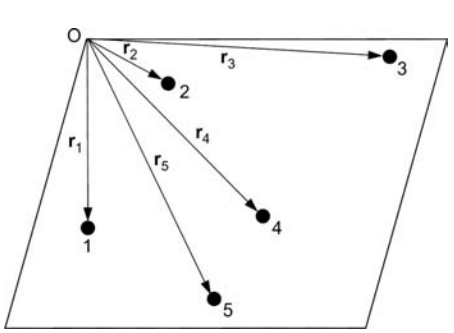

Fig. 23 Atomic positions in a unit cell.

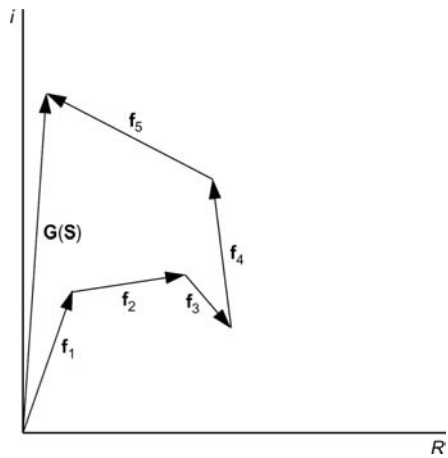

Fig. 24 Vector diagram for the total scattered wave in direction **S** added up for five atoms.

from the unit cell at the origin of the crystal is $\mathbf{G}(\mathbf{S})$. All scatterers in the second unit cell are displaced by the vector **a** relative to the origin, which introduces a corresponding phase factor and reveals for the second unit cell relative to the origin, $\mathbf{G}(\mathbf{S}) \exp 2\pi i \mathbf{aS}$. For the nth unit cell relative to the origin, we obtain $\mathbf{G}(\mathbf{S}) \exp 2\pi i(n-1)\mathbf{aS}$. This sums up for the total wave to Eq. (22):

$$\mathbf{F}(\mathbf{S}) = \sum_{n=1}^{T} \mathbf{G}(\mathbf{S}) \exp 2\pi i(n-1)\mathbf{aS} \quad (22)$$

Generally, $\mathbf{F}(\mathbf{S})$ is of the same order of magnitude as $\mathbf{G}(\mathbf{S})$, and no strong scattering effect is observed (Fig. 25a). However, when $2\pi\ \mathbf{aS} = 2\pi h$ or an integral multiple of 2π or $\mathbf{aS} = h$ (h is an integer), the waves add up constructively to a scattered wave proportional to $T|\ \mathbf{G}(\mathbf{S})|$ (Fig. 25b). $T = 10^5$ for a 1-mm long crystal with a 100-Å lattice constant. The intensity distribution of the scattered waves is concentrated around the values where **aS** is equal to an integer and depends on the number of contributing unit cells. The more unit cells are contributing, the sharper is the concentration of the intensity around these values.

3.3.2 Three-dimensional Crystal

In this case, the unit cell is spanned by the unit cell vectors **a**, **b**, and **c**, and is repeated periodically by the corresponding vector shifts in the respective spatial directions. This means that we will obtain scattered waves of measurable intensities when the three subsequent conditions are fulfilled (Eq. 23):

$$\mathbf{aS} = h;\ \mathbf{bS} = k;\ \mathbf{cS} = l \quad (23)$$

These conditions are known as *Laue equations*.

If we neglect the proportionality constant T, we obtain Eq. (24) for the total scattered wave for a three-dimensional crystal with a unit cell containing N atoms:

$$\mathbf{F}(\mathbf{S}) = \sum_{j=1}^{N} f_j \exp 2\pi i \mathbf{r}_j \mathbf{S} \quad (24)$$

with (Eq. 25)

$$\mathbf{r}_j = \mathbf{a}x_j + \mathbf{b}y_j + \mathbf{c}z_j \quad (25)$$

Hence, we have (Eq. 26 and 27)

$$\mathbf{r}_j\mathbf{S} = x_j\mathbf{aS} + y_j\mathbf{bS} + z_j\mathbf{cS} = hx_j + hy_j + lz_j \quad (26)$$

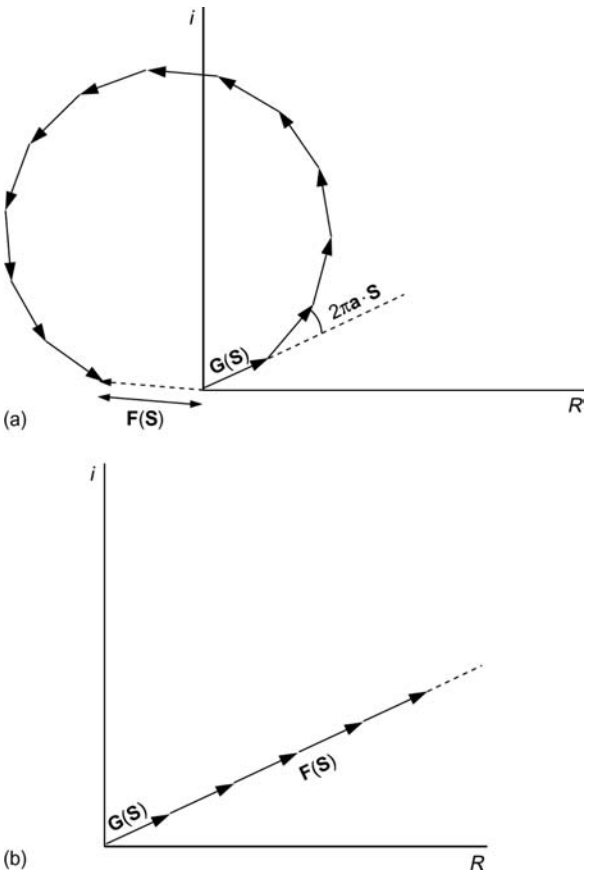

Fig. 25 Vector diagrams displaying the total wave scattered by a molecule in a crystal. (a) The phase differences between waves scattered by adjacent unit cells is $2\pi\mathbf{aS}$ and (b) the phase difference is an integral multiple of 2π (adapted from Blundell, T.L., Johnson, L.N. (1976) *Protein Crystallography*, Academic Press, New York).

(from Laue's equation) and

$$\mathbf{F}(hkl) = \sum_{j=1}^{N} f_j \exp 2\pi i(hx_j + hy_j + lz_j)$$
$$= |\mathbf{F}(hkl)| \exp i\alpha(hkl) \qquad (27)$$

with $|\mathbf{F}(hkl)|$ – amplitude and α – phase angle. We obtain the intensity of the scattered wave as the structure factor $\mathbf{F}(hkl)$ multiplied by its complex conjugate value according to (Eq. 28):

$$I(hkl) = \mathbf{F}(hkl)\mathbf{F}^*(hkl)$$
$$= |\mathbf{F}(hkl)|^2 \qquad (28)$$

3.4 The Reciprocal Lattice and Ewald Construction

In section 2.4, we mentioned the usefulness of the concept of the reciprocal lattice

in understanding the diffraction of X-rays from a crystal. Now we have the necessary relations for the derivation of the reciprocal lattice. One can write the scattering vector **S** as Eq.(29):

$$\mathbf{S} = h_x\mathbf{a}^* + k_y\mathbf{b}^* + l_z\mathbf{c}^* \tag{29}$$

where **S** is a vector in reciprocal space with the metric $\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$. The relation to the direct space with metric **a**, **b**, **c** is still unknown. The vector **S** must obey the Laue equations (Eq. 30):

$$\begin{aligned} \mathrm{a}\mathbf{S} &= \mathbf{a}(h_x\mathbf{a}^* + k_y\mathbf{b}^* + l_z\mathbf{c}^*) = h \\ &= h_x\mathbf{a}\mathbf{a}^* + k_y\mathbf{a}\mathbf{b}^* + l_z\mathbf{a}\mathbf{c}^* = h \end{aligned} \tag{30}$$

This is fulfilled only when $\mathbf{a}\mathbf{a}^* = 1$, $h_x = h$, and $\mathbf{a}\mathbf{b}^*$ and $\mathbf{a}\mathbf{c}^* = 0$. Similar equations can be derived for the other two Laue conditions. Thus, vector **S** is a vector of a lattice in reciprocal space. The relation between direct and reciprocal lattice is given by the following set of nine equations (Eq. 31–39):

$$\mathbf{a}\mathbf{a}^* = 1 \tag{31}$$

$$\mathbf{b}\mathbf{a}^* = 0 \tag{32}$$

$$\mathbf{c}\mathbf{a}^* = 0 \tag{33}$$

$$\mathbf{a}\mathbf{b}^* = 0 \tag{34}$$

$$\mathbf{b}\mathbf{b}^* = 1 \tag{35}$$

$$\mathbf{c}\mathbf{b}^* = 0 \tag{36}$$

$$\mathbf{a}\mathbf{c}^* = 0 \tag{37}$$

$$\mathbf{b}\mathbf{c}^* = 0 \tag{38}$$

$$\mathbf{c}\mathbf{c}^* = 1 \tag{39}$$

It follows from these that $\mathbf{a}^* \perp \mathbf{b}$; **c**; $\mathbf{b}^* \perp \mathbf{a}$; **c**; $\mathbf{c}^* \perp \mathbf{a}$; **b**; and vice versa. The metric relations can also be derived from these relations. It means that the inverse lattice vectors are perpendicular to the plane that is spanned by the two other noninverse lattice vectors. Bragg's law can be derived now by inspection of Fig. 26. The wave vectors for the incident wave $\mathbf{s}_0$ and the scattered wave **S** have the same absolute value of $1/\lambda$. Vector **S** must be a vector of the reciprocal lattice and its absolute value is equal to d^*. From Fig. 26, we obtain Eq. 40 to 42:

$$\sin\theta = \frac{d^*}{2}\lambda \tag{40}$$

$$\lambda = \frac{2\sin\theta}{d^*} \tag{41}$$

$$\lambda = 2d\sin\theta \quad \text{for } n = 1 \tag{42}$$

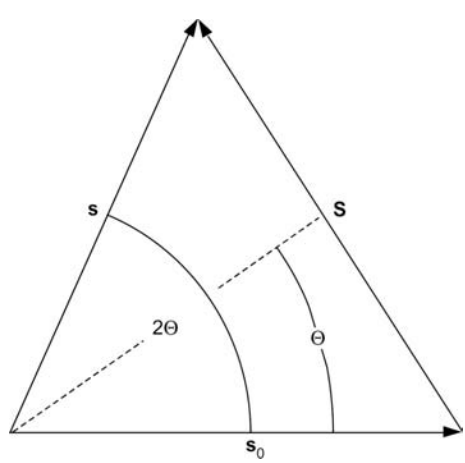

Fig. 26 Geometric representation of diffraction geometry, 2Θ, glance angle; Θ, Bragg angle.

The general equation for Bragg's law is Eq. (43):

$$2d \sin\theta = n\lambda \tag{43}$$

where n is the order of reflection and d is the interplanar distance in the direct lattice.

The Ewald construction is contained in Fig. 20. A sphere of radius $1/\lambda$ is drawn. The origin of the reciprocal lattice is located where the wave vector $\mathbf{s}_0$ ends on the Ewald sphere. A diffracted beam is generated if a reciprocal lattice vector $\mathbf{d}^*_{hkl}$ with an absolute value of $1/d_{hkl}$ cuts the Ewald sphere. The beam is diffracted in the direction of the connection of the origin of the Ewald sphere and the intersection point of the reciprocal lattice point on the Ewald sphere. The diffraction pattern of a lattice is itself a lattice with reciprocal lattice dimensions.

3.5 The Temperature Factor

The thermal motion of the atoms causes a decrease of the scattering power by a factor of exp[-B($\sin^2\theta/\lambda^2$)] with (Eq. 44)

$$B = 8\pi^2\overline{u}^2 \tag{44}$$

where $\overline{u}$ is the mean displacement of the atoms due to the thermal motion. The atomic scattering factor f must be multiplied with this factor. In this model, the thermal motion has been assumed to be isotropic. Therefore, B is denoted as the isotropic temperature factor. In molecules, this is usually not the case and the thermal motion is described by a tensor ellipsoid. Here we obtain a set of six independent anisotropic temperature factors. In protein crystallography, isotropic B values for each atom of the molecules are used normally. The thermal motion of the atoms is one main reason for the falloff of the diffraction intensity especially at higher diffraction angles. This limits the possible recordable number of diffraction spots and, as will be seen later, the resolution of the diffraction experiment.

3.6 Symmetry in Diffraction Patterns

An X-ray diffraction data set from a crystal represents its reciprocal lattice with the corresponding diffraction intensities at the reciprocal lattice points (hkl). As the reciprocal lattice is closely related to its direct partner, it reveals symmetries, lattice properties, and other peculiarities (e.g. systematic extinctions) that are connected to the direct crystal symmetry such as unit cell dimensions and space group. A detailed discussion of this problem is given in Buerger.

In the case of real atomic scattering factors f, the diffraction intensities are centrosymmetric according to Friedel's law (Eq. 45):

$$I(hkl) = I(\overline{hkl}) \tag{45}$$

This is illustrated in Fig. 27(a,b). The square of a complex number is the product of this number by its complex conjugate. This is shown for $\mathbf{F}$ (hkl) in Fig. 27(a) and for $\mathbf{F}(\overline{hkl})$ in Fig. 27(b). The resulting intensities are equal in both cases.

3.7 Electron Density Equation and Phase Problem

Inspection of the equation for the structure factor (Eq. 46)

$$\mathbf{F}(\mathbf{S}) = \sum_{j=1}^{N} f_j \exp 2\pi i \mathbf{r}_j \mathbf{S} = \int_{\substack{\text{vol. of}\\ \text{unit cell}}} \rho(\mathbf{r}) \exp 2\pi i \mathbf{r}\mathbf{S} \mathrm{d}v \tag{46}$$

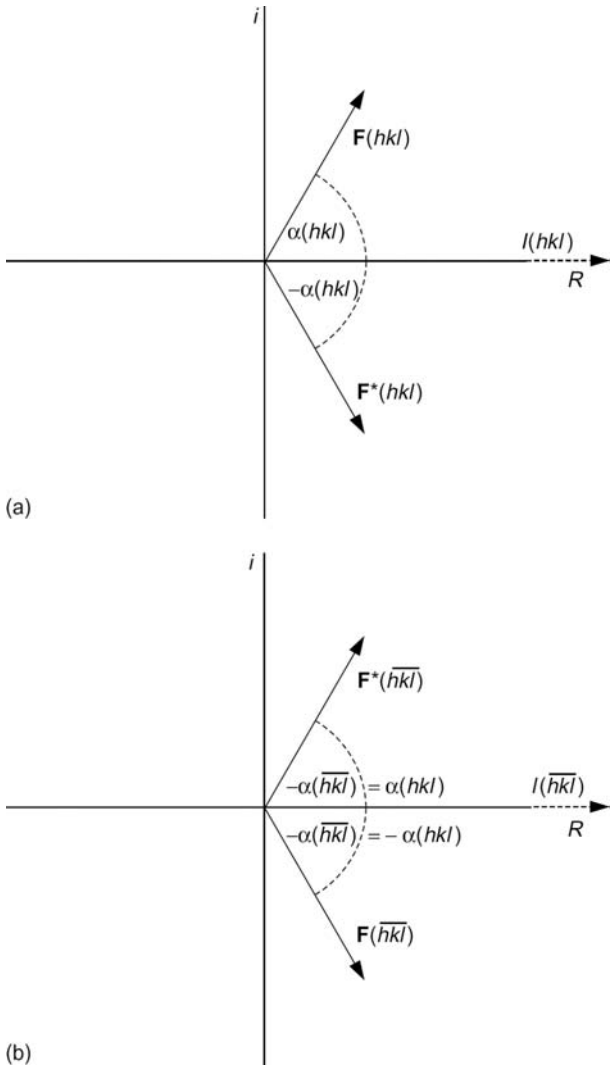

Fig. 27 Diagram illustrating the basis of Friedel's law.

shows that it is the Fourier transform of the electron density ρ (**r**).

The electron density ρ (**r**) is then the inverse Fourier transform of the structure factor **F**(**S**) according to Eq. (47):

$$\rho(\mathbf{r}) = \int_{\substack{\text{vol. of}\\ \text{diffraction space}}} \mathbf{F}(\mathbf{S}) \exp -2\pi i \mathbf{r}\mathbf{S} \mathrm{d}v_s \tag{47}$$

The integration is replaced by summation since **F**(**S**) is not continuous and is nonzero only at the reciprocal lattice points. Hence, we have (Eq. 48)

$$\rho(xyz) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \mathbf{F}(hkl) \times \exp[-2\pi i(hx + ky + lz) \tag{48}$$

Knowing the structure factors (Eq. 49)

$$\mathbf{F}(hkl) = |\mathbf{F}(hkl)| \exp i\alpha(hkl) \qquad (49)$$

one can calculate the electron density distribution in the unit cell and thus determine the atomic positions of the scattering molecule(s). Unfortunately, the measured quantities are only the absolute values $|\mathbf{F}(hkl)|$ of the structure factor. Information on the phase angles $\alpha(hkl)$ is lost during the diffraction experiment. The determination of these phases is the basic problem in any crystal structure determination, and we discuss the methods for solving the phase problem later.

3.8 The Patterson Function

The measured X-ray intensities are proportional to the square of the absolute value of the structure factor according to Eq. (28). Would it be possible to use the intensities directly to calculate from these a function that contains structural information? The answer is "yes." If one calculates a convolution of the electron density with itself, Patterson showed that this is just the Fourier transform of the intensities (Eq. 50):

$$P(uvw) = \int_{\substack{\text{vol. of}\\ \text{unit cell}}} \rho(xyz) \times \rho(x+u, y+v, z+w)\mathrm{d}v \qquad (50)$$

From this, it follows that (Eq. 51)

$$P(\mathbf{u}) = \frac{1}{V}\sum_{\mathbf{h}} F_{\mathbf{h}}^2 \exp -2\pi i\mathbf{hu} \qquad (51)$$

The function $P(\mathbf{u})$ will have maxima if the positions $\mathbf{x}$ and $\mathbf{x}+\mathbf{u}$ correspond to atoms. Thus, we obtain a function that contains the interatomic vectors as maxima. We expect N^2 peaks for N atoms. The maxima are proportional to Z_iZ_j. The Patterson function is a very useful tool to locate atoms when the number of atoms in the asymmetric unit of the unit cell is not too high (e.g. <20) or it contains a subset of heavy atoms among not too many (e.g. <100) light atoms such as C, N, O, or S. Here, the heavy atom–heavy atom vectors are clearly prominent. If a protein with 1000 or a multiple of that light atoms holds one or several heavy atoms per molecule, the signal resulting from the heavy atoms can no longer be resolved. However, when using the method of isomorphous replacement (discussed later), a Patterson function of the heavy atom structure can be calculated, from which it is possible to locate the heavy atoms.

3.9 Integrated Intensity Diffracted by a Crystal

Real crystals are not perfect. They can be regarded as consisting of small blocks of perfect crystals (sizes in the range of 0.1 μm), which have an average tilt angle among each other of 0.1 to 0.5° for protein crystals and which diffract independently of each other. Such a real crystal is denoted a mosaic crystal. The total energy, E(**h**), in a diffracted beam for a mosaic crystal rotating with uniform angular velocity w through the reflecting position in a beam of X-rays of incident intensity I_0 is given by Darwin's equation (Eq. 52):

$$\mathrm{E}(\mathbf{h}) = \frac{I_0}{\omega}\lambda^3\frac{e^4}{m^2c^4}p\frac{LAV_x}{V^2}|\mathbf{F}(\mathbf{h})|^2 \qquad (52)$$

where λ = wavelength of X-rays, e = electronic charge, m = mass of electron, c = velocity of light, p = polarization factor, L = Lorentz factor (geometrical factor taking into account the relative time each reflection spends in the reflection position), A = absorption factor, V = volume

of the unit cell, and V_x = irradiated crystal volume.

Owing to the mosaicity (0.1–0.5°), each reflection has a corresponding reflection width. The integrated intensity equation is valid under the assumption that apart from ordinary absorption, the incident intensity, I_0, is constant within the crystal (kinematic theory of X-ray diffraction) and the mosaic blocks are so small that no multiple scattering occurs within an individual mosaic block. The integrated intensity depends on λ to the third power. Increasing the wavelength causes appreciably stronger diffraction intensities but is accompanied by larger absorption. Copper Kα radiation with a wavelength of 1.5418 Å is an optimal choice for protein crystallography when using X-ray generator sources. Also important is the dependence of the integrated intensity on the unit cell volume V by its negative second power. Doubling of the unit cell volume with twice as many molecules, taking into account the increase in $|\mathbf{F(h)}|^2$ by having now $2n$ molecules per unit cell, reduces the average intensity for the reflected beams by a factor of two. In Eq. (52), $(\lambda^3/\omega V^2) \times (e^4/m^2c^4) \times V_x \times I_0$ is a constant for a given experiment. The corrected intensity on a relative scale $I(\mathbf{h})$ is obtained from Eq. (53):

$$I(\mathbf{h}) = \frac{E(\mathbf{h})}{p \times L \times A} \tag{53}$$

3.10 Intensities on an Absolute Scale

The corrected intensity on a relative scale $I(\mathbf{h})$ can be converted to an intensity given by Eq. (54)

$$I(abs, \mathbf{h}) = \mathbf{F(h)F(h)}^* = |\mathbf{F(h)}|^2 \tag{54}$$

on an absolute scale by applying a so-called Wilson plot. The basis for this plot is an equation that connects the average intensity on an absolute scale with the average intensity on a relative scale by a scale factor C and considers the isotropic thermal motion of the scattering atoms by the temperature factor given in Eq. (44). This is written in the form of Eq. (55):

$$\ln \frac{\overline{I(\mathbf{h})}}{\sum_j (f_j)^2} = \ln C - 2B\frac{\sin^2\theta}{\lambda^2} \tag{55}$$

This is the equation of a straight line. B, the overall temperature factor, and C, the scale factor, can be obtained by plotting $\ln \overline{I(\mathbf{h})}/\sum_j (f_j)^2$ against $(\sin^2\theta)/\lambda^2$.

3.11 Resolution of the Structure Determination

The concept of resolution in X-ray diffraction has the same meaning as the concept in image formation in the optical microscope. After the Abbe theory, we obtain Eq. (56):

$$d_m = \frac{\lambda}{2NA} \tag{56}$$

where NA is the numerical aperture of the objective lens. In protein crystallography, the nominal resolution of an electron density map is expressed in d_m, the minimum interplanar spacing for which Fs are included in the Fourier series. The maximum attainable resolution at a given wavelength is $\lambda/2$. For Cu Kα radiation, it is 0.7709 Å and would suffice to determine protein structures at atomic resolution (the distance of a carbon–carbon single bond is about 1.5 Å). However, usually the thermal vibrations of the atoms in

a protein crystal are so high that the diffraction data cannot be observed to the full theoretical resolution limit. The polypeptide chain fold can be determined at a resolution of better than 3.5 Å. A medium-resolution structure is in the resolution range of 3.0 to 2.2 Å, and makes the amino acid side chains clearly visible. A high-resolution structure has a nominal resolution better than 2.2 Å and can be as good as 1.2 Å. In such structures, the main-chain carbonyl oxygens become visible as prominent bumps and at a resolution better than 2.0 Å, aromatic side chains acquire a hole in the middle of their ring systems. For some very well diffracting crystals from small proteins, diffraction data extending to resolutions below 1.2 Å could be collected with synchrotron radiation. Such structures reveal real atomic resolution where each atom is visible as an isolated maximum in the electron density map.

3.12 Diffraction Data Evaluation

The analysis and reduction of diffraction data from a single crystal consists of seven main steps: (1) visualization and preliminary analysis of the raw, unprocessed data; (2) indexing of the diffraction patterns; (3) refinement of the crystal and detector parameters; (4) integration of the diffraction spots; (5) finding the relative scale factors between measurements; (6) precise refinement of crystal parameters using the whole data set; and (7) merging and statistical analysis of the measurements related by space-group symmetry. When using electronic area detectors with short readout times such as CCD or MWPC detectors, it is possible to collect diffraction images with small rotational increments (0.05–0.2°). In this case, the reflection profile over the crystal rotation angle can be registered, giving a three-dimensional picture of the spot. The evaluation of such diffraction data can be done with computer programs MADNES, XDS, the San Diego programs and related programs XENGEN, and X-GEN. IP systems with their longer readout times are operated in a film-like mode with rotational increments of 0.5 to 2.0°. Here, mainly the program systems MOSFLM and Denzo are applied. The most important developments in the data evaluation of macromolecular diffraction measurements are autoindexing, profile fitting, transformation of data to a reciprocal-space coordinate system, and demonstration that a single rotation image contains all of the information necessary to derive the diffraction intensities from that image.

Scaling, merging, and statistical analysis of the intensity data are either done with corresponding programs of the CCP4 program suite or with Scalepack. The principles of these operations are given in the manuals for these programs. With modern data-collection methods, the completeness should approach 100% (including the low-resolution data, which are very important for molecular replacement), the ratio $I/\sigma(I)$ should be significant even for the highest resolution shell, and undue emphasis should not be given to the reliability factor for merging the data (R-merge) unless factors such as multiplicity are taken into account. Nowadays, it is customary using synchrotron radiation techniques and fast CCD detectors to collect as much data as possible (before radiation damage becomes significant) in order to produce good statistics.

3.13 Solvent Content of Protein Crystals

The Matthews parameter V_M, which is defined according to Eq. (57):

$$V_M = \frac{V_{\text{unit cell}}}{M_{\text{Prot}}} \tag{57}$$

where $V_{\text{unit cell}}$ is the volume of the unit cell and M_{Prot} is the molecular mass of the protein in the unit cell, has values that are in the range 1.6 to 3.5 $Å^3\ Da^{-1}$ for proteins. This allows a rough estimation of the number of molecules in the unit cell. Furthermore, V_M can be used for the assessment of the solvent content of a protein crystal. Calling V_{Prot} the crystal volume occupied by the protein, V'_P its fraction with respect to the total crystal volume V, and M_{Prot} the mass of protein in the cell, we obtain Eq. (58):

$$V'_P = \frac{V_{\text{Prot}}}{V} = \frac{V_{\text{Prot}}/M_{\text{Prot}}}{M_{\text{Prot}}/V} \tag{58}$$

The first term is the specific volume of the protein, the second the reciprocal of V_M and, remembering that the molecular weight is expressed in daltons, we have Eq. (59):

$$V'_P = \frac{1.6604}{d_{\text{Prot}} V_M} \tag{59}$$

Taking 1.35 g cm^{-3} as the protein density, we obtain as a first approximation Eq. (60) and (61):

$$V'_P \approx \frac{1.23}{V_M} \tag{60}$$

$$V'_{\text{Solv}} \approx 1 - V'_P \tag{61}$$

The solvent content in a protein crystal may vary from 75 to 40%.

4 Methods for Solving the Phase Problem

4.1 Isomorphous Replacement

4.1.1 Preparation of Heavy Metal Derivatives

If one can attach one or several heavy metal atoms at defined binding site(s) to the protein molecules without disturbing the crystalline order, one can use such isomorphous heavy atom derivatives for the phase determination. The lack of isomorphism can be monitored by a change in the unit cell parameters compared with the native crystal and a deterioration of the quality of the diffraction pattern. The preparation of heavy atom derivatives is undertaken by soaking the crystals in mother liquor containing the dissolved heavy metal compound. Soaking times may be in the range from several minutes to months. Concentrations of the heavy metal compound may vary from tenths of millimolar to 50 mM. Favorite heavy atoms are Hg, Pt, U, Pb, Au, rare earth metals, and so on. Potential ligands can be classified as hard and soft ligands according to Pearson. Hard ligands are electronegative and undergo electrostatic interactions. In proteins, such ligands are glutamate, aspartate, terminal carboxylates, hydroxyls of serines and threonines, and in the buffer acetate, citrate, and phosphate. Soft ligands are polarizable and form covalent bonds such as cysteine, cystine, methionine, and histidine in proteins, and Cl^-, Br^-, I^-, S-ligands, CN^-, and imidazole in the buffer solution.

Metals are classified according to their preference for hard or soft ligands. Class (a) metals bind preferentially to hard ligands. They comprise the cations of A-metals such as alkali and alkaline earth metals, the lanthanides, some actinides,

and groups IIIA, IVA, and VA of the transition metals. Class (b) metals are rather soft and polarizable and can form covalent bonds to soft ligands. They include heavy metals at the end of the transition metal groups such as Hg, Pt, and Au. Thus, in the protein, the class (b) metals Hg, Pt, and Au and complex compounds of them bind to soft ligands such as cysteine, histidine, or methionine and the class (a) metals U and Pb to hard ligands such as the carboxylate groups of glutamate or aspartate.

4.1.2 Single Isomorphous Replacement

The structure factor $\mathbf{F}_{PH}$ for the heavy atom derivative structure (Fig. 28) becomes (Eq. 62)

$$\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H \tag{62}$$

where $\mathbf{F}_P$ = structure factor of the native protein and $\mathbf{F}_H$ = contribution of the heavy atoms to the structure factor of the derivative. The isomorphous differences, $F_{PH} - F_P$, which can be calculated from experimental intensity data sets of the native and derivative protein, correspond to the distance CB in Fig. 28, and are given by Eq. (63):

$$F_{PH} - F_P = F_H \cos(\alpha_{PH} - \alpha_H) - 2F_P \sin^2\left(\frac{\alpha_P - \alpha_{PH}}{2}\right) \tag{63}$$

If F_H is small compared with F_P and F_{PH}, the sine term will be very small and we have (Eq. 64)

$$F_{PH} - F_P \approx F_H \cos(\alpha_{PH} - \alpha_H) \tag{64}$$

When vectors $\mathbf{F}_P$ and $\mathbf{F}_H$ are collinear, then (Eq. 65)

$$|\mathbf{F}_{PH} - \mathbf{F}_P| = F_H \tag{65}$$

The square of the isomorphous differences, $F_{PH} - F_P$, can be used as coefficients in a Patterson synthesis. We get

$$\begin{aligned}(F_{PH} - F_P)^2 &= 4F_P^2 \sin^4\left(\frac{\alpha_P - \alpha_{PH}}{2}\right) && \text{(i)}\\ &\quad + F_H^2 \cos^2(\alpha_{PH} - \alpha_H) && \text{(ii)}\\ &\quad - 4F_P F_H \sin^2\left(\frac{\alpha_P - \alpha_{PH}}{2}\right) \times \cos(\alpha_{PH} - \alpha_H) && \text{(iii)}\end{aligned} \tag{66}$$

It is a theorem of Fourier theory that the Fourier transform of the sum

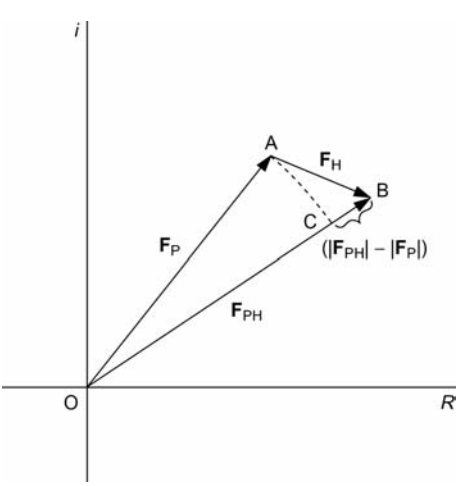

Fig. 28 Vector diagram for the vector addition of the structure factor of the native protein $\mathbf{F}_P$ and the heavy atom contribution $\mathbf{F}_H$ to the heavy atom derivative structure factor $\mathbf{F}_{PH}$.

of Fourier coefficients is equal to the sum of the Fourier transforms of the individual Fourier coefficients. Here, there are three different terms. $(\alpha_P - \alpha_{PH})$ is small if F_H is small, and term (i), which gives the protein–protein interaction will be of low weight. The transform of term (iii) is zero if sufficient terms are included. However, if $F_H \ll F_P$, $(\alpha_P - \alpha_{PH})$ is effectively random, and term (ii) will give heavy atom vectors with half the expected peak heights (Eq. 67)

$$F_H^2 \cos^2(\alpha_{PH} - \alpha_H) = \frac{1}{2}F_H^2 + \frac{1}{2}F_H^2 \cos 2(\alpha_{PH} - \alpha_H) \quad (67)$$

with the second term on the right contributing only noise to the Patterson map because the angles α_{PH} and α_H are not correlated. Such an isomorphous heavy atom difference Patterson map allows the determination of the positions of the heavy metals on the condition of isomorphism and a not-too-large heavy atom partial structure. The interpretation of these difference Patterson maps is undertaken by vector verification routines, which are part of the CCP4 program suite. In these routines, the asymmetric unit of the unit cell is systematically scanned by calculating on each scan point the corresponding heavy atom–heavy atom vectors, determining their peak height in the Patterson map and evaluating a meaningful correlation value (e.g. the sum of the correlated maxima). Prominent heavy atom sites should show up with high correlation values.

It is important to know what intensity changes are generated by the attachment of heavy atoms to the macromolecule. According to Crick and Magdoff, the relative root-mean-square intensity change is given by Eq. (68) for centric reflections:

$$\frac{\sqrt{\overline{(\Delta I)^2}}}{\overline{I_P}} = 2 \times \sqrt{\frac{\overline{I_H}}{\overline{I_P}}} \quad (68)$$

and by Eq. (69) for acentric reflections:

$$\frac{\sqrt{\overline{(\Delta I)^2}}}{\overline{I_P}} = \sqrt{2} \times \sqrt{\frac{\overline{I_H}}{\overline{I_P}}} \quad (69)$$

where $\overline{I_H}$ is the average intensity of the reflections if the unit cell were to contain the heavy atoms only and $\overline{I_P}$ is the average intensity of the reflections of the native protein. Attaching one mercury atom ($Z = 80$) to a macromolecule with varying molecular mass and assuming 100% occupancy gives the following average relative changes in intensity: 0.51 for 14 000 Da, 0.25 for 56 000 Da, 0.18 for 112 000 Da, 0.13 for 224 000 Da, and 0.09 for 448 000 Da. From this estimation it is evident that with increasing molecular mass more heavy atoms or for large molecular masses heavy metal clusters such as $Ta_6Br_{12}^{2+}$ must be introduced to generate intensity changes that can be statistically measured (precision for intensity measurements between 5 and 10%) and which are sufficient for the phasing.

The phase calculation for single isomorphous replacement can be seen from the so-called Harker construction for this case (Fig. 29). $\mathbf{F}_H$, which can be calculated from the known heavy atom positions, is drawn in its negative direction from the origin O ending at point A. Circles are drawn with radii F_P and F_{PH} from points O and A respectively. The connections of the intersection points of both circles B and C with origin O determine two possible phases for $\mathbf{F}_P$. This means that the single isomorphous replacement leaves an ambiguity in the phase determination for the acentric reflections.

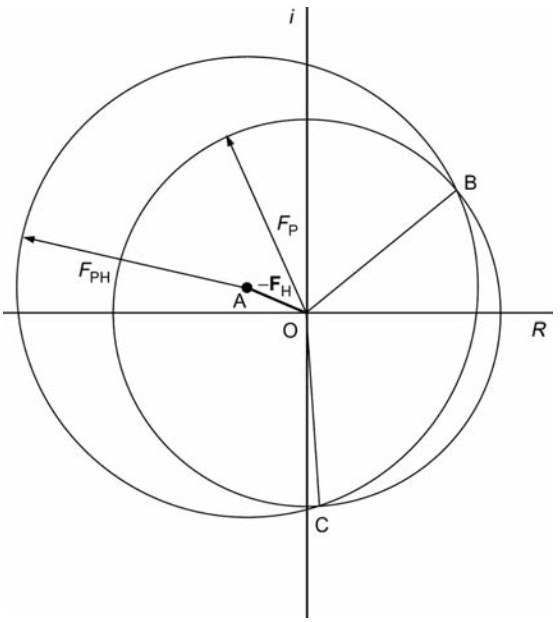

Fig. 29 Harker construction for the phase calculation by the method of single isomorphous replacement.

4.1.3 **Multiple Isomorphous Replacement**

The phase ambiguity can be overcome if two or more isomorphous heavy atom derivatives are used, which exhibit different heavy atom partial structures. In Fig. 30, the Harker construction for two different heavy atom derivatives is shown. In addition to Fig. 29, $-\mathbf{F}_{H2}$ is drawn from the origin O and a third circle with radius F_{PH2} is inserted around its endpoint B. The

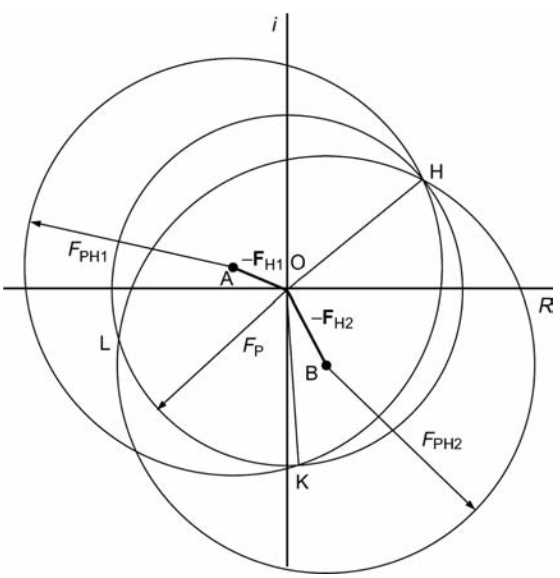

Fig. 30 Harker construction for the phase calculation by the method of MIR for two different heavy atom derivatives PH1 and PH2.

intersection point, H, of all three circles determines the protein phase, α_P. In the case of n isomorphous derivatives, there are $n+1$ circles, which have one common intersection point whose connection to origin O determines the protein phase, α_P.

4.2 Anomalous Scattering

4.2.1 Theoretical Background

So far, in the normal Thomson scattering of X-rays, the electrons in the atom have been treated as free electrons that vibrate as a dipole-oscillator in response to the incident electromagnetic radiation and generate elastic scattering of the X-rays. However, the electrons are bound to atomic orbitals in atoms, and this treatment is valid only if the frequency ω of the incident radiation is large compared to any natural absorption frequency ω_{kn} of the scattering atom. For the light atoms in biological macromolecules (H, C, N, O, S, P) with frequency ω of the used radiation (in the range of 0.4 to 3.5 Å), this condition is fulfilled and these atoms scatter normally. For heavier elements, the assumption $\omega \gg \omega_{kn}$ is no longer valid, and the frequency ω may be higher for some and lower for other absorption frequencies. If ω is equal to an absorption frequency ω_{kn}, absorption of radiation will occur, which is manifested by the ejection of a photoelectron with an energy corresponding to the ionization energy for this electron. This transition goes to a state in the continuous region because the discrete energy states are all occupied in the atom. The absorption frequencies for the K, L, or M shells are connected with the corresponding absorption edges, which are characterized by a sharp drop in the absorption curve (absorption vs λ) at the edge position. It is evident that the scattering from the electrons with their resonance frequencies close or equal to the frequency of the incident radiation will deliver a special contribution, which is called *anomalous scattering*.

The classical treatment is briefly outlined. It is assumed that the atoms scatter as if they contain electric dipole-oscillators having certain definite natural frequencies. The classical differential equation of the motion of a particle of mass m and charge e in an alternating electric field $\mathbf{E} = \mathbf{E}_0 e^{i\omega t}$ is Eq. (70):

$$\ddot{\mathbf{x}} + k\dot{\mathbf{x}} + \omega_s^2 \mathbf{x} = \frac{e\mathbf{E}_0}{m} e^{i\omega t} \tag{70}$$

where the damping factor, k, is proportional to the velocity of the displayed charge and ω_s is the natural circular frequency of the dipole if the charge is displaced. The steady state solution for this equation for the moment of the dipole that executes forced oscillations of frequency under the action of the incident wave is Eq. (71):

$$\mathbf{M} = e\mathbf{x} = \frac{e^2}{m} \frac{\mathbf{E}_0 e^{i\omega t}}{\omega_s^2 - \omega^2 + ik\omega} \tag{71}$$

The amplitude A of the scattered wave at unit distance in the equatorial plane is given by Eq. (72):

$$A = \frac{e^2}{mc^2} \frac{\omega^2 E_0}{\omega_s^2 - \omega^2 + ik\omega} \tag{72}$$

The scattering factor of the dipole, f, is now defined as the ratio of the amplitude scattered by the oscillator to that scattered by a free classical electron under the same conditions. This amplitude at unit distance and in the equatorial plane is given by Eq. (73):

$$A' = -\frac{e^2}{mc^2} E_0 \tag{73}$$

Hence, we obtain Eq. (74) for f:

$$f = \frac{\omega^2}{\omega^2 - \omega_s^2 - ik\omega} \tag{74}$$

If f is positive, the scattered wave has a phase difference of π with respect to the primary beam (introduced by the negative sign in the equation for A'). If $\omega \gg \omega_s$, f is unity. In the case of $\omega \ll \omega_s$, f is negative, and the dipole then scatters a wave in phase with the primary beam.

Equation (74) can be split into real and imaginary parts so that we obtain Eq. (75):

$$f = f' + if'' \tag{75}$$

with (Eqs. 76 and 77)

$$f' = \frac{\omega^2(\omega^2 - \omega_s^2)}{(\omega^2 - \omega_s^2)^2 + k^2\omega^2} \tag{76}$$

$$f'' = \frac{k\omega^3}{(\omega^2 - \omega_s^2)^2 + k^2\omega^2} \tag{77}$$

We now extend this for an atom consisting of s electrons, each acting as a dipole-oscillator with oscillator strength $g(s)$ and resonance frequency ω_s. We have to multiply the contribution for each electron by $g(s)$ and form the sum over all electrons. For the total real part of the atomic scattering factor, we obtain Eq. (78):

$$f' = \sum_s \frac{g(s)\omega^2}{\omega^2 - \omega_s^2} \tag{78}$$

which assumes that ω is not very nearly equal to ω_s, and a small damping. f' can be written as Eq. (79):

$$f' = f_0 + \Delta f' = \sum_s g(s) + \sum_s \frac{g(s)\omega_s^2}{\omega^2 - \omega_s^2} \tag{79}$$

For free electrons, we have $\omega_s = 0$ and $f' = f_0 = \sum_s g(s)$. The real part of the increment of the scattering factor is due to the binding of electrons. $\Delta f'$ is the dispersion component of the anomalous scattering.

If ω is comparable to ω_s but slightly greater, $ik\omega$ must not be neglected. f becomes complex (Eq. 80):

$$f = f' + if'' = f_0 + \Delta f' + i\Delta f'' \tag{80}$$

The imaginary part lags $\pi/2$ behind the primary wave, that is it is always $\pi/2$ in front of the scattered wave. $\Delta f''$ is known as the absorption component of the anomalous scattering. In the quantum mechanical treatment of the problem, the oscillator strengths are calculated from the atomic wave functions. Hönl, in theoretical work, used hydrogen-like atomic wave functions. In the frame of this approach, to each natural dipole frequency ω_s in the classical expression, there corresponds in the quantum expression a frequency ω_{kn}, which is the Bohr frequency associated with the transition of the atom from the energy state k to the state n in which it is supposed to remain during the scattering. Modern quantum mechanical calculations of anomalous scattering factors on isolated atoms, based on relativistic Dirac–Slater wave functions, have been carried out by Cromer and Liberman. It follows from the theory of the anomalous scattering of X-rays that f_0 is real, independent of the wavelength of the incident X-rays but dependent on the scattering angle. $\Delta f'$ and $\Delta f''$ depend on the wavelength, λ, of the incident radiation but are virtually independent of the scattering angle.

4.2.2 Experimental Determination

$\Delta f''$ is related to the atomic absorption coefficient μ_0 by Eq. (81):

$$\Delta f''(\omega) = \frac{mc\omega}{4\pi}\mu(\omega_0) \tag{81}$$

$\Delta f'$ can now be calculated by the Kramers–Kronig transformation (Eq. 82):

$$\Delta f''(\omega) = \frac{2}{\pi}\int_0^\infty \frac{\omega'\Delta f''(\omega')}{\omega^2 - \omega'^2}\mathrm{d}\omega' \quad (82)$$

As fluorescence is closely related to absorption, fluorescence measurements varying the X-ray radiation frequency are used to determine the frequency dependence of the dispersive components of the different chemical elements. Instead of the radiation frequency ω, the radiation is often characterized by its wavelength, λ, or photon energy, E. The dispersion correction terms $\Delta f'$ and $\Delta f''$ are often simply denoted f' and f''. Figure 31 shows the anomalous scattering factors near the absorption K edge of selenium from a crystal of *E. coli* selenomethionyl thioredoxin. The spectrum was measured with tunable synchrotron radiation. Apart from the "white line" feature at the absorption edge, f'' drops by about 4 electrons, approaching the edge from the short wavelength side; $\Delta f'$ exhibits a symmetrical drop of −8 electrons around the edge. Similar values can be observed at the K edges for Fe, Cu, Zn, and Br, whose wavelengths all lie in the range 0.9 to 1.8 Å, which is well suited for biological macromolecular X-ray diffraction experiments. For other interesting heavy atoms such as Sm, Ho, Yb, W, Os, Pt, and Hg, the LII (Sm) or LIII edges are in this range. Here, the effects are even greater. Considerably larger changes are found for several lanthanides, such as Yb, where the minimum f' is −33 electrons and the maximum f'' is 35 electrons.

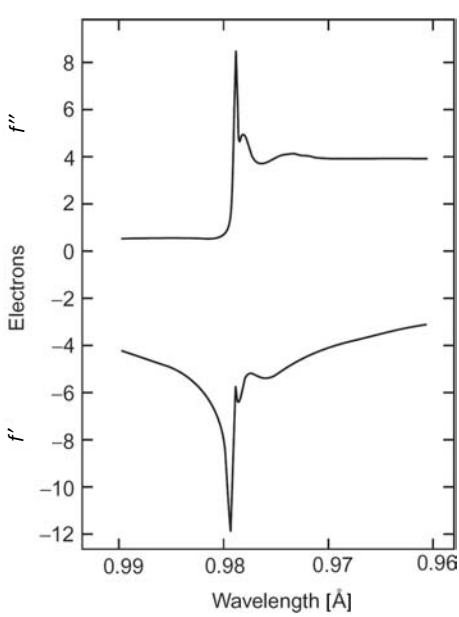

Fig. 31 Anomalous scattering factors near the absorption K edge of selenium from a crystal of *E. coli* selenomethionyl thioredoxin. (Reproduced by permission of Academic Press, Inc., from Hendrickson, W.A., Ogata, C.M. (1997) *Methods Enzymol.* **276**, 494–523.)

4.2.3 Breakdown of Friedel's Law

Under the assumption that the crystal contains a group of anomalous scatterers, one can separate the contributions from the distinctive components of the scattering factor according to Hendrickson and Ogata to obtain Eq. (83):

$$^{\lambda}\mathbf{F}(\mathbf{h}) = {}^{\circ}\mathbf{F}_{\mathrm{N}}(\mathbf{h}) + {}^{\circ}\mathbf{F}_{\mathrm{A}}(\mathbf{h}) + {}^{\lambda}\mathbf{F}'_{\mathrm{A}}(\mathbf{h}) + i^{\lambda}\mathbf{F}''_{\mathrm{A}}(\mathbf{h}) \quad (83)$$

where ${}^{\circ}\mathbf{F}_{\mathrm{N}}$ is the contribution of the normal scatterers and ${}^{\circ}\mathbf{F}_{\mathrm{A}}$, ${}^{\lambda}\mathbf{F}'_{\mathrm{A}}$, and ${}^{\lambda}\mathbf{F}''_{\mathrm{A}}$ are the contributions for the corresponding components of the complex atomic form factor. For the centrosymmetric reflection, we obtain Eq. (84):

$$\mathbf{F}(-\mathbf{h}) = {}^{\circ}\mathbf{F}_{\mathrm{N}}(-\mathbf{h}) + {}^{\circ}\mathbf{F}_{\mathrm{A}}(-\mathbf{h}) + {}^{\lambda}\mathbf{F}'_{\mathrm{A}}(-\mathbf{h}) + i^{\lambda}\mathbf{F}''_{\mathrm{A}}(-\mathbf{h}) \quad (84)$$

The geometric presentation for both structure factors is given in Fig. 32. The

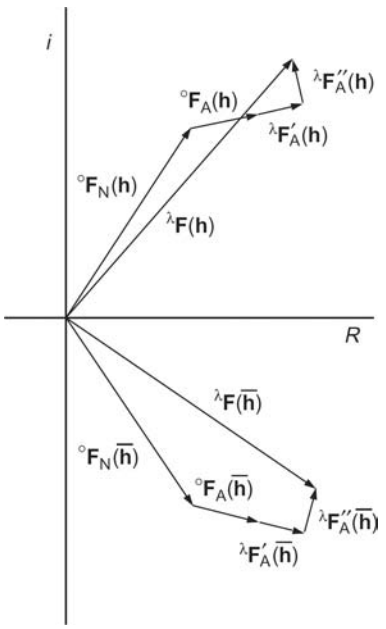

Fig. 32 Vector diagram explaining the breakdown of Friedel's law.

inversion of the sign of **h** causes a negative phase angle for all contributions where the components of the scattering factor are real. For the f''-dependent part, this is also valid, but owing to the imaginary factor i, this vector has to be constructed with a phase angle $+\pi/2$ with respect to ${}^{\circ}\mathbf{F}_{\mathrm{A}}(-\mathbf{h})$ and ${}^{\lambda}\mathbf{F}'_{\mathrm{A}}(-\mathbf{h})$. The resultant absolute values for ${}^{\lambda}\mathbf{F}(\mathbf{h})$ and ${}^{\lambda}\mathbf{F}(-\mathbf{h})$ are no longer equal, which means that their intensities (square of the amplitude) are different (breakdown of Friedel's law).

4.2.4 Anomalous Difference Patterson Map

One can show that (Eq. 85)

$$ {}^{\lambda}F(\mathbf{h}) - {}^{\lambda}F(-\mathbf{h}) \approx \frac{2}{k}[{}^{\circ}\mathrm{F}_{\mathrm{A}}(\mathbf{h}) + {}^{\lambda}F'_{\mathrm{A}}(\mathbf{h})]\sin(\alpha_{\mathbf{h}} - \alpha_{\mathrm{A}}) \qquad (85) $$

where $\alpha_{\mathbf{h}}$ is the phase angle of ${}^{\lambda}\mathbf{F}(\mathbf{h})$, α_{A} the phase angle of the anomalous scatterers, and (Eq. 86)

$$ k = \frac{{}^{\circ}\mathrm{F}_{\mathrm{A}}(\mathbf{h}) + {}^{\lambda}F'_{\mathrm{A}}(\mathbf{h})}{{}^{\lambda}F''_{\mathrm{A}}(\mathbf{h})} \qquad (86) $$

As coefficients for an anomalous difference Patterson map, we obtain Eq. (87):

$$ \Delta F^2_{\mathrm{ano}} = [{}^{\lambda}F(\mathbf{h}) - {}^{\lambda}F(-\mathbf{h})]^2 \sim \frac{4}{k^2}[{}^{\circ}\mathrm{F}_{\mathrm{A}}(\mathbf{h}) + {}^{\lambda}F'_{\mathrm{A}}(\mathbf{h})]^2 \sin^2(\alpha_{\mathbf{h}} - \alpha_{\mathrm{A}}) \qquad (87) $$

The ΔF_{ano}s will be maximal if the phase angle α_{A} is perpendicular to the phase angle $\alpha_{\mathbf{h}}$ and zero if both vectors are collinear, which is opposite to the MIR case. The anomalous Patterson map contains peaks of the anomalous scatterers with heights proportional to half of $(4/k^2)[{}^{\circ}\mathrm{F}_{\mathrm{A}}(\mathbf{h}) + {}^{\lambda}F'_{\mathrm{A}}(\mathbf{h})]^2$ owing to the $\sin^2$ term and is therefore suited to determine the structure of the anomalous scatterers.

4.2.5 Phasing Including Anomalous Scattering Information

The combination of anomalous scattering information with isomorphous replacement permits the unequivocal determination of the protein phases, as shown in Fig. 33. Using the anomalous scattering information alone gives two possible solutions for the protein phase characterized by the intersection points H and L in Fig. 33. Combining it with the corresponding intensities from the native protein without the anomalous scatterers leaves only one solution for the protein phase (vector O-H in Fig. 33). The case in Fig. 33 is called *single isomorphous replacement anomalous scattering* (*SIRAS*). Having n isomorphous heavy atom derivatives, each with anomalous scattering contributions, the Harker

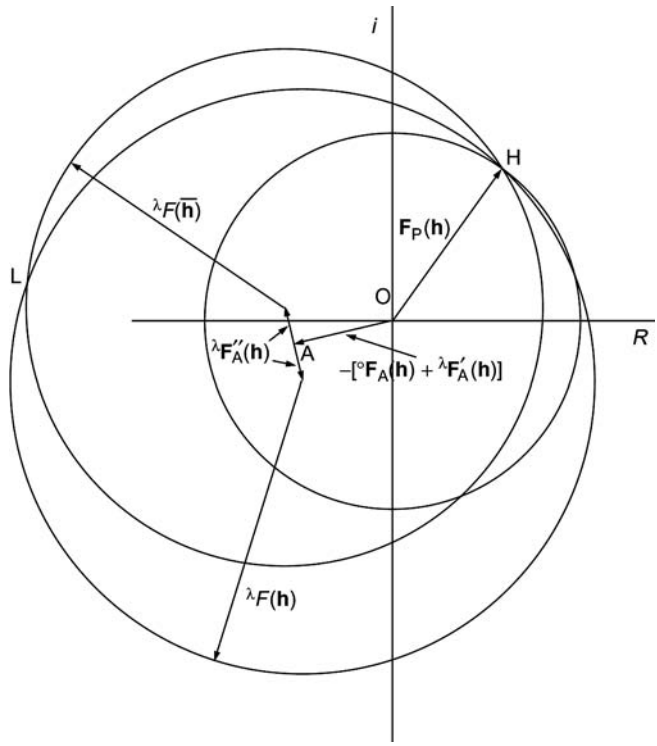

Fig. 33 Harker construction illustrating the phase determination combining information from anomalous scattering and isomorphous replacement.

construction can be extended for this situation, and the phasing method is then designated as multiple isomorphous replacement anomalous scattering (MIRAS).

4.2.6 Multiwavelength Anomalous Diffraction Technique

During the last few years, the MAD technique has matured to be a routine method and has led to a revolution in biological macromolecular crystallography. If there are one or a few anomalous scatterers in the biological macromolecule, it is possible to determine the whole spatial structure from one crystal (exact isomorphism) by the MAD technique. The anomalous scatterers may be intrinsic as in metalloproteins (e.g. Fe, Zn, Cu, Mo, Mn) or exogenous (e.g. Hg in a heavy atom derivative or Se in selenomethionyl proteins). A prerequisite for the MAD technique is well-diffracting crystals (resolution better than 2.8 Å) because the anomalous components of the atomic form factor are virtually independent of the diffraction angle and acquire increasing weight with increasing scattering angle. This advantageous property together with exact isomorphism serves for the determination of good phases down to the full resolution, and leads to the production of excellent experimental MAD-phased electron density maps. A typical MAD experiment is carried out at three different

wavelengths (tunable synchrotron radiation), at minimum f' and maximum f'' at the absorption edge of the anomalous scatterer(s), and at a remote wavelength where anomalous scattering effects are small.

The basic equations for the MAD technique as formulated by Hendrickson and Ogata are as follows. Equation (83) can be written as Eq. (88):

$$^{\lambda}\mathbf{F}(\mathbf{h}) = {}^{\circ}\mathbf{F}_{\mathrm{T}}(\mathbf{h}) + {}^{\lambda}\mathbf{F}_{\mathrm{A}}(\mathbf{h}) + i^{\lambda}\mathbf{F}''_{\mathrm{A}}(\mathbf{h}) \tag{88}$$

where (Eq. 89)

$$^{\circ}\mathbf{F}_{\mathrm{T}} = {}^{\circ}\mathbf{F}_{\mathrm{N}} + {}^{\circ}\mathbf{F}_{\mathrm{A}} \tag{89}$$

with subscript T for the totality of atoms in the structure.

Furthermore, we have Eqs. (90) to (93):

$$^{\circ}\mathbf{F}_{\mathrm{T}}(f^{\mathrm{o}}) = {}^{\circ}\mathrm{F}_{\mathrm{T}} \exp(i^{\circ}\phi_{\mathrm{T}}) \tag{90}$$

$$^{\circ}\mathbf{F}_{\mathrm{A}}(f^{\mathrm{o}}) = {}^{\circ}\mathrm{F}_{\mathrm{A}} \exp(i^{\circ}\phi_{\mathrm{A}}) \tag{91}$$

$$^{\lambda}\mathbf{F}'_{\mathrm{A}} = f(f') \tag{92}$$

$$^{\lambda}\mathbf{F}''_{\mathrm{A}} = f(f'') \tag{93}$$

In the common case of a single kind of anomalous scatterer, we obtain Eqs. (94) and (95):

$$^{\lambda}\mathbf{F}'_{\mathrm{A}} = \frac{f'(\lambda)}{f^{\circ}}\,{}^{\circ}\mathbf{F}_{\mathrm{A}} \tag{94}$$

$$^{\lambda}\mathbf{F}''_{\mathrm{A}} = \frac{f''(\lambda)}{f^{\circ}}\,{}^{\circ}\mathbf{F}_{\mathrm{A}} \tag{95}$$

Separating the experimentally observable squared amplitude into wavelength-dependent and wavelength-independent terms gives Eq. (96):

$$\begin{aligned} {}^{\lambda}F(\pm\mathbf{h})^2 = {} & {}^{\circ}\mathrm{F}_{\mathrm{T}}^2 + a(\lambda)\,{}^{\circ}\mathrm{F}_{\mathrm{A}}^2 \\ & + b(\lambda)\,{}^{\circ}\mathrm{F}_{\mathrm{T}}\,{}^{\circ}\mathrm{F}_{\mathrm{A}} \cos({}^{\circ}\phi_{\mathrm{T}} - {}^{\circ}\phi_{\mathrm{A}}) \\ & \pm c(\lambda)\,{}^{\circ}\mathrm{F}_{\mathrm{T}}\,{}^{\circ}\mathrm{F}_{\mathrm{A}} \sin({}^{\circ}\phi_{\mathrm{T}} - {}^{\circ}\phi_{\mathrm{A}}) \end{aligned} \tag{96}$$

with (Eqs. 97–99)

$$a(\lambda) = \frac{f'^2 + f''^2}{f^{\circ 2}} \tag{97}$$

$$b(\lambda) = 2\frac{f'}{f^{\circ}} \tag{98}$$

$$c(\lambda) = 2\frac{f''}{f^{\circ}} \tag{99}$$

The derivation of the formula for ${}^{\lambda}F(\mathbf{h})^2$ is illustrated in detail. ${}^{\lambda}F(\mathbf{h})^2$ is obtained from the triangle formed by the vectors ${}^{\lambda}\mathbf{F}(\mathbf{h})$, ${}^{\circ}\mathbf{F}_{\mathrm{T}}(\mathbf{h})$, and **a** (Fig. 34) by use of the cosine rule. The absolute values of the vectors are represented in italics, and the relevant angle is $(180^{\circ} - {}^{\circ}\phi_{\mathrm{A}} - \delta)$. We get

$$\begin{aligned} {}^{\lambda}F(\mathbf{h})^2 = {} & {}^{\circ}\mathrm{F}_{\mathrm{T}}^2 + \left(\frac{f'^2 + f''^2}{f^{\circ 2}}\right) {}^{\circ}\mathrm{F}_{\mathrm{A}}^2 - 2 \\ & \times {}^{\circ}\mathrm{F}_{\mathrm{T}} \times \left(\frac{f'^2 + f''^2}{f^{\circ 2}}\right)^{1/2} \times {}^{\circ}\mathrm{F}_{\mathrm{A}} \\ & \times \cos(\pi - \delta - {}^{\circ}\phi_{\mathrm{A}} + {}^{\circ}\phi_{\mathrm{T}}) \end{aligned} \tag{100}$$

The cosine term in Eq. (100) can be obtained using some basic trigonometry:

$$\begin{aligned} & \cos(\pi + ({}^{\circ}\phi_{\mathrm{T}} - {}^{\circ}\phi_{\mathrm{A}} - \delta)) \\ & \quad = -\cos({}^{\circ}\phi_{\mathrm{T}} - {}^{\circ}\phi_{\mathrm{A}} - \delta) \end{aligned} \tag{101}$$

$$\begin{aligned} & \cos(({}^{\circ}\phi_{\mathrm{T}} - {}^{\circ}\phi_{\mathrm{A}}) - \delta) = \cos({}^{\circ}\phi_{\mathrm{T}} - {}^{\circ}\phi_{\mathrm{A}}) \\ & \quad \times \cos\delta + \sin({}^{\circ}\phi_{\mathrm{T}} - {}^{\circ}\phi_{\mathrm{A}}) \times \sin\delta \end{aligned} \tag{102}$$

with

$$\cos\delta = \frac{\left(\frac{f'}{f^{\circ}}\right) \times {}^{\circ}\mathrm{F}_{\mathrm{A}}}{\left(\frac{f'^2 + f''^2}{f^{\circ 2}}\right)^{1/2}} \times {}^{\circ}\mathrm{F}_{\mathrm{A}} = \frac{\left(\frac{f'}{f^{\circ}}\right)}{\left(\frac{f'^2 + f''^2}{f^{\circ 2}}\right)^{1/2}} \tag{103}$$

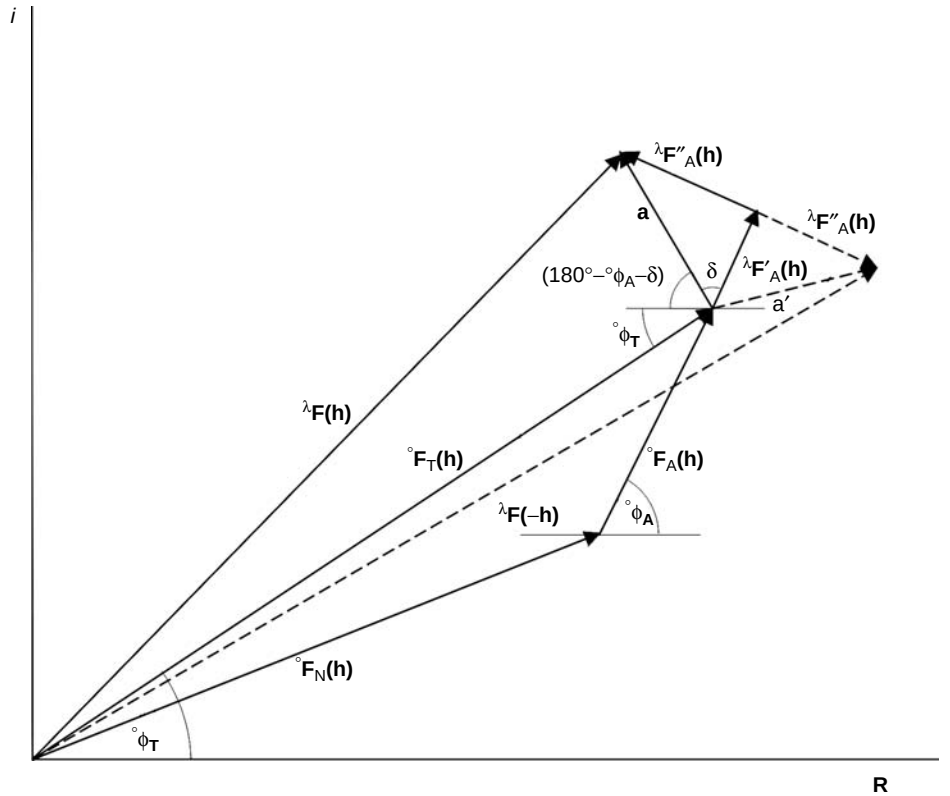

Fig. 34 Schematic drawing of structure factors of a biological macromolecule that contains one kind of anomalous scatterer.

$$\sin\delta = \frac{\left(\frac{f''}{f^\circ}\right) \times {}^\circ\mathrm{F_A}}{\left(\frac{f'^2 + f''^2}{f^\circ 2}\right)^{1/2}}$$

$$\times\ {}^\circ\mathrm{F_A} = \frac{\left(\frac{f''}{f^\circ}\right)}{\left(\frac{f'^2 + f''^2}{f^{\circ 2}}\right)^{1/2}} \qquad (104)$$

Substituting this in Eq. 99, we obtain the expression for ${}^\lambda F(\mathbf{h})^2$. ${}^\lambda F(-\mathbf{h})^2$ is determined from the triangle formed by the vectors ${}^\lambda\mathbf{F}(-\mathbf{h})$, ${}^\circ\mathbf{F}_\mathrm{T}(\mathbf{h})$, and $\mathbf{a}'$ (Fig. 34) using a similar approach. Maximum anomalous scattering effects can be expected in intensity differences of reflections that would be equal for exclusively normal scattering. This is the case for Friedel pairs, $\mathbf{h}$ and $-\mathbf{h}$, or their rotational symmetry partners, and the relation for such differences is given in Eq. (85). Of further interest are dispersive differences between structure amplitudes at different wavelengths, Eq. (105):

$$\Delta F_{\Delta\lambda} \equiv {}^{\lambda i}F(\mathbf{h}) - {}^{\lambda j}F(\mathbf{h}) \qquad (105)$$

The anomalous or dispersive intensity differences can be used to determine the structure of the anomalous scatterers. The methods are the same as for isomorphous replacement. They include vector

verification procedures of difference Patterson maps or direct methods programs such as Shake and Bake and SHELXD.

4.2.7 Determination of the Absolute Configuration

As anomalous scattering destroys the centrosymmetry of the diffraction data, this effect can be used to determine the absolute configuration of chiral biological macromolecules. The most common method is to calculate protein phases on the basis of both hands of the heavy atom or anomalous scatterer structures and check the quality of the relevant electron density map that should be better for the correct hand. Furthermore, secondary structural elements in proteins (consisting of L-amino acids) such as α-helices should be right handed.

4.3 Patterson Search Methods (Molecular Replacement)

If the structures of molecules are similar (virtually identical) or contain a major similar part, this can be used to determine the crystal structure of the related molecule if the structure of the other molecule is known. This is done by systematically exploring the Patterson function of the crystal structure to be determined with the Patterson function of the search model. Let us first consider some important features of the Patterson function. The relation between two identical molecules in the search crystal structure (Fig. 35a) can generally be formulated as Eq. (106):

$$\mathbf{X}_2 = [C]\mathbf{X}_1 + \mathbf{d} \qquad (106)$$

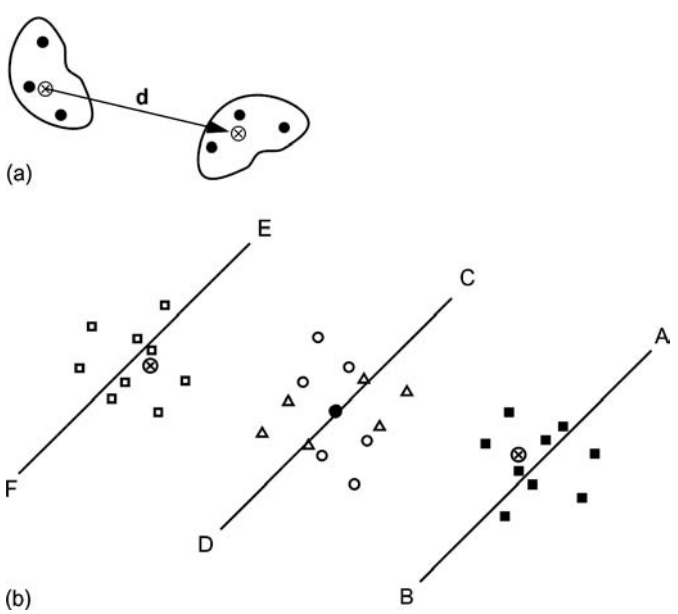

Fig. 35 Patterson function of two identical molecules separated by the spatial movement given in Eq. (106). (a) Positions of the two molecules; (b) interatomic vectors of structure in (a). △, ○, intramolecular vectors of the left and right molecule; □, ■, intermolecular vectors.

Equivalent positions $\mathbf{X}_1$ in molecule 1 are at positions $\mathbf{X}_2$ of molecule 2. $[C]$ is the rotation matrix and $\mathbf{d}$ is the translation vector of the movement of the molecule. Fig. 35(b) shows the Patterson function belonging to the molecular arrangement in Fig. 35(a). It is evident that around the origin, vectors are assembled that are intramolecular, whereas the vectors around lines AB and EF are intermolecular. The intramolecular vectors depend on the molecule orientation only and, therefore, can be used for its determination. Once the orientation of the molecule(s) has been elucidated, this can be used to reveal the translation of the molecule(s) by analyzing the intermolecular vector part of the Patterson function. The distinction between intra- and intermolecular vector sets and exploiting them for orientation and translation determination was given by Hoppe. The extension to protein crystallography and the first mathematical formulation of the rotation and translation functions were given by Rossmann and Blow.

4.3.1 Rotation Function

The intramolecular vectors are arranged in a volume around the origin of the Patterson function with a radius equal to the dimension of the molecule. The rotational search is then carried out in this volume u. The search Patterson (deduced either from the search model or from the crystal Patterson itself) is rotated to any possible rotational orientation $\mathbf{X}_2$, characterized by three rotational angles α, β, γ, which may be defined in different ways (polar angles, Eulerian angles, etc.). At each angular position, the actual functional values are correlated with those of the crystal Patterson all through the volume u and integrated over this volume. The correlation function may be the sum or the product of each corresponding pair of values. Rossmann and Blow proposed a product function, and the rotation function for this case is given by Eq. (107):

$$R(\alpha, \beta, \gamma) = \int_u P_2(\mathbf{X}_2 P_1(\mathbf{X}_1)\mathrm{d}\mathbf{X}_1 \quad (107)$$

The function has maxima if the intramolecular vector sets are coincident. The calculation can be carried out in both direct and reciprocal space.

The self-rotation function is a special form of the rotation function. If an asymmetric unit contains more than one copy of a molecule, the rotation matrix between the molecules can be determined by a self-rotation function. Here, the crystal Patterson is rotated against itself, and the integration is taken over the volume u around the origin in the same manner. The identical molecules may have an arbitrary orientation to each other or they may be related by local or the so-called noncrystallographic symmetries. Searching for local rotation axes is done best in a polar angle system. The search Patterson is brought into each polar orientation and then rotated around the angle value for the local axis being sought, for example 120° for a threefold local axis.

4.3.2 Translation Function

Once the orientation of the molecule(s) has been determined, the translation of the molecule(s) can be obtained from a translation function. The model Patterson $P_2(\mathbf{u})$ revealed from the model in the correct orientation is calculated for different translations $\mathbf{t}$ and correlated with the crystal Patterson $P_1(\mathbf{u})$. The translation function proposed by Crowther and Blow has the form shown in Eq. (108):

$$T(\mathbf{t}) = \int P_1(\mathbf{u}) P_2(\mathbf{u}, \mathbf{t})\mathrm{d}\mathbf{u} \quad (108)$$

$T(\mathbf{t})$ reveals a maximum peak at the correct translation $\mathbf{t}$ if the center of gravity of the search model was at the origin for $\mathbf{t} = 0$.

4.3.3 Computer Programs for Molecular Replacement

An early program for molecular replacement, working in direct space, was written by Huber. Nowadays, several program packages are available, either being exclusively dedicated to the molecular replacement technique or having integrated relevant modules. Pure molecular replacement programs are, for example, AMORE and GLRF. The rotational and translational search starting from the search model is fully automated in AMORE and includes a final rigid body refinement of each proposed solution. GLRF offers different types of rotation and translation functions, all operating in reciprocal space, and a Patterson correlation refinement. A peculiarity of the GLRF program is the locked rotation function. This function takes into account the possible noncrystallographic symmetries and is an average of n independent rotation functions with an improved peak-to-noise ratio. Frequently used program packages including molecular replacement modules are the CCP4 program suite, CNS, and PROTEIN.

4.4 Phase Calculation

4.4.1 Refinement of Heavy Atom Parameters

Before the protein phases can be calculated, it is necessary to refine the heavy atom parameters. These are the coordinates x, y, z, the temperature factor (either isotropic or anisotropic), and the occupancy. The refinement modifies the parameters in such a way that $|\mathbf{F}_{\mathrm{PH}}(\mathrm{obs})|$ becomes as close as possible to $|\mathbf{F}_{\mathrm{PH}}(\mathrm{calc})|$. Using the method of least squares, the refinement according to Rossmann minimizes Eq. (109):

$$\varepsilon = \sum_{\mathbf{h}} w(\mathbf{h})[(F_{\mathrm{PH}} - F_{\mathrm{P}})^2 - kF^2_{\mathrm{H\ calc}}]^2 \tag{109}$$

where k is a scaling factor to correct F^2_{Hcalc} to a theoretically more acceptable value because according to Eq. (64), $\mathbf{F}_{\mathrm{PH}} - \mathbf{F}_{\mathrm{P}}$ and $\mathbf{F}_{\mathrm{H}}$ have approximately the same length when $\mathbf{F}_{\mathrm{PH}}$, $\mathbf{F}_{\mathrm{P}}$, and $\mathbf{F}_{\mathrm{H}}$ point in the same direction. The probability for this case will be high if the difference between F_{PH} and F_{P} is large. An improvement can be obtained if the contribution from the anomalous scattering is included.

For the parameter refinement of anomalous scattering sites, the differences between the observed and calculated structure factor amplitudes for $^{\circ}\mathrm{F}_{\mathrm{A}}$ are subjected to minimization. Another approach treats the anomalous or dispersive contributions as in MIR phasing.

From the refined heavy atom parameters, preliminary protein phase angles α_{P} can be obtained as shown in the corresponding Harker construction. A further refinement of the heavy atom parameters can be achieved by the "lack of closure" method, incorporating this knowledge. The definition of this "lack of closure" ε is illustrated in Fig. 36(a) and (b). In the case of perfect isomorphism, the vector triangle $\mathbf{F}_{\mathrm{P}} + \mathbf{F}_{\mathrm{H}} = \mathbf{F}_{\mathrm{PH}}$ closes exactly (Fig. 36a). In practice, this condition will not be fulfilled, and a difference ε between the observed F_{PH} and the calculated F_{PH} will remain (Fig. 36b). $F_{\mathrm{PH}}(\mathrm{calc})$ can be obtained from the triangle OAB (Fig. 36b)

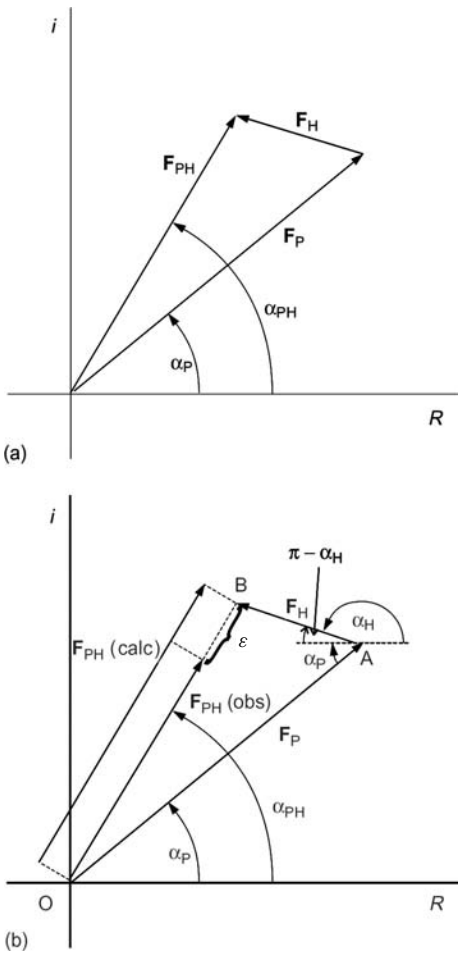

Fig. 36 Definition of "lack of closure." (a) Perfect isomorphism; (b) usually, the observed and calculated values for $\mathbf{F}_{PH}$ differ by the "lack of closure" ε.

with the cosine rule (Eq. 110):

$$F_{PH} = [F_P^2 + F_H^2 + 2F_P \times F_H \cos(\alpha_H - \alpha_P)]^{1/2} \quad (110)$$

The function that is minimized by the least-squares method is Eq. (111):

$$E_j = \sum_{\mathbf{h}} m_{\mathbf{h}} \varepsilon_j(\mathbf{h})^2 \quad (111)$$

where (Eq. (112))

$$\varepsilon_j = k_j F_{PHj}(\text{obs}) - F_{PHj}(\text{calc}) \quad (112)$$

is the "lack of closure" for the heavy atom derivative j, k_j is a scaling factor, and $m_{\mathbf{h}}$ is a weighting factor.

4.4.2 Protein Phases

As the structure factor amplitudes F_P, F_{PH}, F_H, and α_H are known, the protein phase angle α_P can be calculated. For the single isomorphous replacement situation (Fig. 28), ε is zero only for the two protein phase angles α_P where the two circles for F_P and F_{PH} intersect. In practice, all these observed quantities exhibit errors. For the treatment of these errors, it is assumed that all errors are in $\mathbf{F}_{PH}$ and that both $\mathbf{F}_H$ and $\mathbf{F}_P$ are error free. For each protein phase angle α, $\varepsilon(\alpha)$ is calculated. The smaller $\varepsilon(\alpha)$ is, the higher is the probability of a correct phase angle α. For each reflection of a derivative j, a Gaussian probability distribution is assumed for ε according to Eq. (113):

$$P(\alpha) = P(\varepsilon) = N \exp\left[-\frac{\varepsilon^2(\alpha)}{2E^2}\right] \quad (113)$$

where N is a normalization factor and E^2 is the mean square value of ε. Small values of E are related to probability curves with sharp peaks and well-determined phase angles, and the opposite is true for large E values. Such phase-angle probability curves can be calculated for each individual reflection and derivative. For single isomorphous replacement, this curve is symmetric with two high peaks corresponding to the two possible solutions for α_P. We obtain the total probability for each reflection with contributions from n heavy atom derivatives by multiplying the

individual probabilities (Eq. 114):

$$P(\alpha) = \prod_{j=1}^{n} P_j(\alpha) = N' \exp\left[-\sum_j \frac{\varepsilon_j^2(\alpha)}{2E_j^2}\right] \tag{114}$$

These curves will be nonsymmetric with one or several maxima (see Fig. 37a and b).

The question arises of which phases should be taken in the electron density equation to calculate the best electron density function. An immediate guess would be to use the phases where $P(\alpha)$ has the highest value. This approach would be appropriate for unimodal distributions but not for bimodal distributions. Blow and Crick derived the phase value that must be applied under the assumption that the mean square error in electron density over the unit cell is minimal. For one reflection, this is given by Eq. (115):

$$(\Delta\rho^2) = \frac{1}{V^2}(\mathbf{F}_s - \mathbf{F}_t)^2 \tag{115}$$

where $\mathbf{F}_t$ is the true factor and $\mathbf{F}_s$ is the structure factor applied in the Fourier synthesis. The mean square error is then obtained as Eq. (116):

$$(\Delta\rho^2) = \frac{1}{V^2}\frac{\int_{\alpha=0}^{2\pi} (\mathbf{F}_s - F\exp i\alpha)^2 P(\alpha)\mathrm{d}\alpha}{\int_{\alpha=0}^{2\pi} P(\alpha)\mathrm{d}\alpha} \tag{116}$$

$\mathbf{F}_t$ has a phase probability of $P(\alpha)$ and has been given as $\mathbf{F}_t = F\exp i\alpha$. It can be shown that the numerator integral in Eq. (116) is minimal if (Eq. 117)

$$\mathbf{F}_{s(best)} = F\frac{\int_{\alpha=0}^{2\pi} \exp(i\alpha)P(\alpha)\mathrm{d}(\alpha)}{\int_{\alpha=0}^{2\pi} P(\alpha)\mathrm{d}\alpha} = mF\exp(i\alpha_{best}) \tag{117}$$

Equation (117) corresponds to the center of gravity of the probability distribution with polar coordinates (mF, α_{best}), where m is defined as magnitude of $\mathbf{m}$ given by Eq. (118):

$$\mathbf{m} = \frac{\int_{\alpha=0}^{2\pi} P(\alpha)\exp(i\alpha)\mathrm{d}\alpha}{\int_{\alpha=0}^{2\pi} P(\alpha)\mathrm{d}\alpha} \tag{118}$$

This magnitude of $\mathbf{m}$ is equivalent to a weighting function and is designated the "figure of merit." The electron density map calculated with mF and α_{best} is known

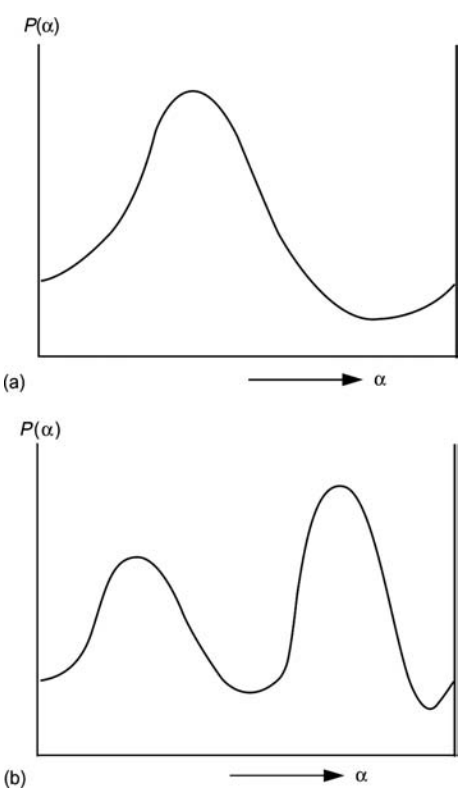

Fig. 37 Total probability curves $P(\alpha)$ for two different reflections: (a) for one derivative; (b) for more than one derivative.

as *best Fourier* and should represent a Fourier map with minimum least-squares error from the true Fourier map.

For the total error of the "best Fourier," Eq. (119) has been derived:

$$(\Delta\rho^2) = \frac{1}{V^2}\sum_{\mathbf{h}} F^2(\mathbf{h})(1 - m^2) \quad (119)$$

The order of magnitude of this error may be illustrated by the example of the structure determination of lysozyme. The root-mean-square error in the Fourier synthesis was 0.35 e $Å^{-3}$, with values of 2.0-Å resolution for the diffraction data and a mean "figure of merit" of 0.6.

The program systems CCP4 and PROTEIN contain all routines necessary to calculate protein phases according to the MIRAS technique and a number of different kinds of Fourier maps. Alternative probabilistic approaches for the phase calculation are used in programs MLPHARE and SHARP. Both programs can also carry out MAD phasing. The MADSYS program is based on the algebraic approach outlined in the MAD section of this contribution and executes all tasks of a MAD analysis from scaling to phase-angle calculation.

4.5 Phase Improvement

With the methods so far described, an experimental electron density map can be calculated, and if its quality is high enough, the atomic model can be constructed. However, there are methods for further phase improvement available, which may be applied in general or depending on the given prerequisites. Such phase improvement routines have been used routinely over the last 10 years and have had a large impact on the advancement of biological macromolecular crystallography.

4.5.1 Solvent Flattening

Protein crystals have a solvent content of 75 to 40%. In a highly refined protein crystal structure, the solvent space between the molecules is rather flat owing to the dynamic nature of this region. Usually, the initial experimental starting phases are of lower quality than the final ones and as a result, the solvent region (if the molecular boundaries can be identified) contains noise peaks. It is now obvious to set the noisy solvent space to a low constant value and calculate new improved phases by Fourier back-transforming this corrected electron density map. However, it is evident that the definition of the molecular boundaries will be tedious and depend on the quality of the electron density map. Wang has proposed an automatic procedure that smooths the electron density to define the protein region. This smoothed electron density map is traced against a threshold value that separates this map into molecule and solvent space according to their ratio of volumes in the unit cell. The space inside the molecular envelope is polished to avoid internal voids. Now, a new electron density map is calculated using the observed structure factor amplitudes and the phases revealed from the solvent corrected map. The solvent corrected map is obtained by setting all electron density values inside the molecular envelope to those of the initial map and all values outside the envelope to a low constant value. These phases from the solvent flattening procedure can be combined with the MIR or MAD phases. This procedure can be repeated in several iterative cycles because after each cycle of solvent flattening, the quality of the electron density map is improved. There are no prerequisites for the application of the method of solvent flattening. It is evident that solvent flattening is most

effective for crystals with a high solvent content.

4.5.2 Histogram Matching

Histogram matching is a technique emanating from image processing. In the application to electron density maps, it is assumed that a high-quality protein crystal structure has a characteristic frequency distribution of electron density, which serves as a standard reference distribution for other electron density maps. Such maps of lower quality exhibit a frequency distribution of electron density, which deviates from the standard distribution. The electron density map of low quality is then scaled in such a way that its frequency distribution of electron density now corresponds to the standard distribution. Histogram matching is normally used together with solvent flattening and is incorporated into the density modification programs SQUASH and DM from the CCP4 program package.

4.5.3 Molecular Averaging

If there is more than one identical subunit in the asymmetric unit of the crystal, molecular averaging can be used to improve the protein phases. The spatial relations between the single identical subunits in the asymmetric unit may be determined by Patterson search methods or from the arrangement of the heavy atoms or anomalous scatterers. The spatial relation between the identical subunits can be improper (the relevant spatial movement consists of a rotation about an unsymmetrical angle value and a translation component) or proper (the spatial movements form a symmetry group that is composed of rotational symmetry elements only). Such additional symmetries are called *noncrystallographic* or *local*, and there are no limitations concerning the Zähligkeit of the symmetry axes (e.g. five-, seven-, and higherfold axes are allowed). It is evident that averaging about the different related subunits, whose electron density should be equal in each subunit, must result in an improved electron density map and therefore in improved protein phases. Molecular averaging is best done in direct space and several programs (e.g. RAVE or MAIN) are available.

The procedure of molecular averaging is composed of several steps. First, the molecular envelope must be determined from the initial electron density map or from a molecular model that, for example, has been obtained from molecular replacement. Next, the particular electron density averaging between the related subunits is performed. This is followed by the reconstitution of the complete crystal unit cell with the averaged electron density. The space outside the molecular envelope is flattened. This map is then Fourier back-transformed. The obtained phase angles can either be taken directly or combined with known phase information to calculate a new and improved electron density map. This cycle can be repeated several times until convergence of the electron density map improvement has been reached. It is very useful to refine the local symmetry operations after every macrocycle of molecular averaging. Furthermore, molecular averaging can be applied if proteins crystallize in more than one crystal form.

Molecular averaging is especially efficient if a high noncrystallographic symmetry is present as in virus structures, but the averaging over two related subunits alone (the lowest case of local symmetry) can give a considerable improvement.

In special cases where high noncrystallographic symmetry exists and the phase information extends to low resolution only, cyclic molecular averaging can be used to extend the phase angles to the full resolution of the native protein. This was first shown in the structure analysis of hemocyanin from *Panulirus interruptus*. It is extensively used in the analysis of icosahedral structures and for large molecular assemblies.

4.5.4 Phase Combination

In the course of a crystal structure analysis of a biological macromolecule, phase information from different sources may be available, such as information from isomorphous replacement, anomalous scattering, partial structures, solvent flattening, and molecular averaging. An overall phase improvement can be expected when these factors are combined, and a useful method to do this was proposed by Hendrickson and Lattman. The probability curve for each reflection is written in an exponential form as Eq. (120):

$$P_s(\alpha) = N_s \exp(K_s + A_s \cos\alpha + B_s \sin\alpha + C_s \cos 2\alpha + D_s \sin 2\alpha) \quad (120)$$

Subscript s stands for the source from which the phase information has been derived. K_s and the coefficients A_s, B_s, C_s, and D_s depend on the structure factor amplitudes and other magnitudes, for example, the estimated standard deviation of the errors in the derivative intensity, but are independent of the protein phase angles α. The overall probability function $P(\alpha)$ is obtained by a multiplication of the individual phase probabilities, and this turns out to be a simple addition of all K_s and of the related coefficients in the exponential term. We obtain Eq. (121):

$$\begin{aligned} P(\alpha) &= \prod_s \mathrm{P}_s(\alpha) \\ &= N' \exp\left[\sum_s K_s + \left(\sum_s A_s\right)\cos\alpha \right. \\ &\quad + \left(\sum_s B_s\right)\sin\alpha + \left(\sum_s C_s\right)\cos 2\alpha \\ &\quad \left. + \left(\sum_s D_s\right)\sin 2\alpha\right] \end{aligned} \quad (121)$$

K_s and the coefficients A_s to D_s have special expressions for each source of phase information.

4.6 Difference Fourier Technique

Supposing that one has solved the crystal structure of a biological macromolecule and has isomorphous crystals of this macromolecule, which contain small structural changes caused by a substrate-analog or inhibitor binding, a metal removal or replacement or a local mutation of one or several amino acids. Then, these structural changes can be determined by the difference Fourier technique. The difference Fourier map is calculated with the differences between the observed structure factor amplitudes of the slightly altered molecule $F_{\mathrm{DERI}}(\mathrm{obs})$ and the native molecule $F_{\mathrm{NATI}}(\mathrm{obs})$ as Fourier coefficients and the phase angles of the native molecule α_{NATI} as phases according to Eq. (122):

$$\begin{aligned} \rho_{\mathrm{DERI}} - \rho_{\mathrm{NATI}} \cong \frac{1}{V}\sum_{\mathbf{h}} m[&F_{\mathrm{DERI}}(\mathrm{obs}) \\ &- F_{\mathrm{NATII}}(\mathrm{obs}) \times \exp(i\alpha_{\mathrm{NATI}}) \\ &\times \exp(-2\pi i\mathbf{hx})] \end{aligned} \quad (122)$$

where m may be the figure of merit or another weighting scheme. The difference Fourier map can alternatively be calculated with coefficients $F_{\mathrm{DERI}}(\mathrm{obs}) - F_{\mathrm{DERI}}(\mathrm{calc})$ and phases $\alpha_{\mathrm{DERI}}(\mathrm{calc})$. $F_{\mathrm{DERI}}(\mathrm{calc})$ and $\alpha_{\mathrm{DERI}}(\mathrm{calc})$ do not include the unknown contribution of the structural change.

Figure 38(a) and (b) illustrate the relation for the structure factors involved in the difference Fourier technique. We assume that the structural change is small. If F_{NATI} is large, the structure factor amplitude of the structural change F_{SC} will be small compared to F_{NATI}, and α_{DERI} will be close to α_{NATI}. This is no longer valid if F_{NATI} is small. Now, F_{SC} is comparable to F_{NATI}, and α_{DERI} may deviate considerably from α_{NATI}. This implies the necessity to introduce a weighting scheme that scales down the contributions where the probability is high that α_{NATI} differs appreciably from the correct phase angle. Various weighting schemes have been elaborated such as those of Sim and Read. The weighting scheme of Sim has the following form (Eq. 123):

$$w = \frac{I_1(X)}{I_0(X)} \tag{123}$$

for acentric reflections and (Eq. 124)

$$w = \tanh \frac{X}{2} \tag{124}$$

for centric reflection with (Eq. 125)

$$X = \frac{2F_{\mathrm{DERI}} \times F_{\mathrm{NATI}}}{\sum_{1}^{n} f_j^2} \tag{125}$$

$I_0(X)$ and $I_1(X)$ are modified Bessel functions of zeroth and first order respectively. These equations and weighting schemes can also be used for the calculation of OMIT maps (where parts of the model have been omitted from the structure factor evaluation) or when a complete structure must be developed from a known partial model. F_{DERI} must be replaced by the observed structure factor F, F_{NATI} by the structure factor of the known or included part of the model F_{K}, and α_{NATI} by the phase angle α_{K} of the known or included part of the model.

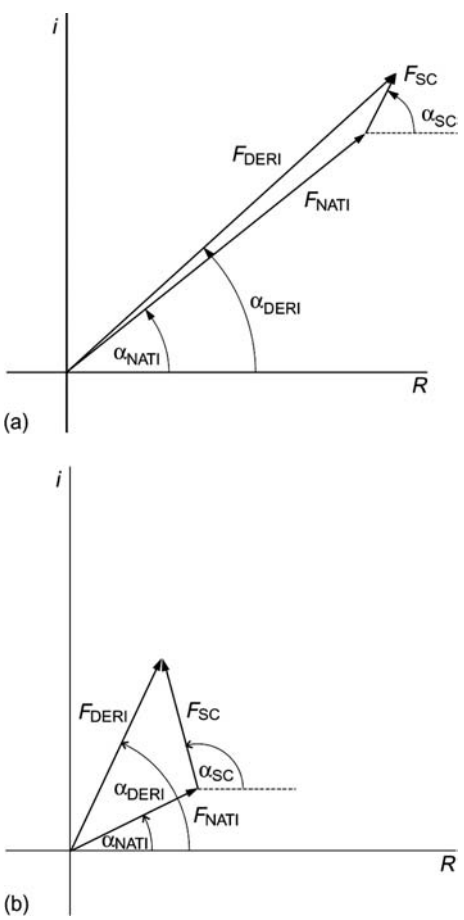

Fig. 38 Vector diagrams illustrating different situations (a) and (b) in the difference Fourier technique for the involved structure factors.

5
Model Building and Refinement

5.1
Model Building

Once the quality of the MIRAS or MAD maps is good enough, model building can be started. This is done on a computer graphics system, and the main modeling programs are O and TURBO-FRODO. An interesting alternative is the program MAIN, which additionally contains routines for molecular averaging, molecular docking, and other features. The visualization of the relevant electron density map on the computer graphics system is done as cagelike structures. For this purpose, the standard deviation from the mean value of the map is calculated, and the cagelike structure is built up for a given contour level (normally 1.0 σ). The first task in a *de novo* protein crystal structure analysis is to localize the trace of the polypeptide chain. This can be assisted by routines for automatic chain tracing such as the bones routine, an auxiliary program of O. Such automatic chain-tracing programs generate a skeleton of the electron density map. This representation was introduced by Greer. When the trace of the polypeptide chain has been identified, the atomic model can be built into the electron density. The atomic model is represented as sticks, connecting the atomic centers of bonded atoms. The individual building blocks (amino acids) of the protein molecule can be generated, interactively manipulated (e.g. linked with each other, moved, rotated, etc.), and fitted into the corresponding part of the electron density map. The geometry of the atomic model is regularized according to protein standard geometries.

The success of model building depends on the quality and resolution of the experimental electron density map. Usually, the quality of the electron density map is not so good that the complete model can be constructed in one cycle. In this case, the partial atomic model is refined crystallographically against the observed structure factor amplitudes. This phase information can be used directly to calculate a new electron density map commonly with $2F_{\mathrm{obs}} - F_{\mathrm{calc}}$ Fourier coefficient amplitudes. This kind of map is the sum of a normal F_{obs} Fourier and a difference Fourier synthesis. It displays the atomic model with normal weight and indicates errors in the model by its contribution of the difference Fourier map. The parallel determination and inspection of a difference Fourier map are also very helpful. As already mentioned, the model phases can be combined with phases present from other sources or incorporated in procedures of phase improvement. A further model-building cycle can be started with such new and improved electron density maps. After several cycles of model building and crystallographic refinement, the atomic model will be so well defined that the solvent structure of internally bound solvent molecules can be developed. The atomic model is complete now and the biochemical interpretation can be started.

5.2
Crystallographic Refinement

The structural model has to be subjected to a refinement procedure. In practical macromolecular crystallography, one does not always have atomic resolution. Therefore, the single atoms cannot be treated as moving independently. They must be refined using energy or stereochemistry restraints, taking care to

maintain a reasonable stereochemistry of the macromolecule. There exist different approaches to structure refinement of macromolecules. The minimization of a potential energy function E together with a diffraction term D according to Eq. (126):

$$S = E + D \tag{126}$$

where (Eqs. 127 and 128)

$$\begin{aligned} E = &\sum k_{\mathrm{b}}[b_{j(\mathrm{calc})} - b_j^0]^2 \\ &+ \sum k_\tau[\tau_{j(\mathrm{calc})} - \tau_j^0]^2 \\ &+ \sum k_\Theta[1 + \cos(m\Theta_k + \delta] \\ &+ \sum (Ar^{-12} + Br^{-6}) \end{aligned} \tag{127}$$

$$D = \sum_i w_i[F_{i(\mathrm{obs})} - kF_{i(\mathrm{calc})}]^2 \tag{128}$$

is applied in the programs EREF and CNS, which are now used frequently. The four terms of the right-hand side of E describe bond, valence angle, dihedral torsion angle, and nonbonded interactions, k_{b} is the bond stretching constant, k_τ is the bond angle bending force constant, k_Θ is the torsional barrier, m and δ are the periodicity and the phase of the barrier, A and B are the repulsive and long-range nonbonded parameters, D is the crystallographic contribution with w_i a weighting factor, F_{obs} is the observed structure factor, F_{calc} is the calculated structure factor, and k is a scaling factor. Programs TNT and PROLSQ use stereochemically restrained least-squares refinement. For both refinement schemes, parameters are employed that were derived from small molecule crystal structures of amino acids, small peptides, nucleic acids, sugars, fatty acids, cofactors, and so on. If noncrystallographic symmetry is present, a corresponding term may be introduced in the energy or stereochemistry part of the expression to be minimized. It is possible to divide the structural model into several individual parts and refine these parts as rigid bodies. This is especially useful with solutions from molecular replacement.

A measure of the quality of the crystallographic model is calculated from the crystallographic R-factor (Eq. 129):

$$R = \frac{\sum_i |F_{\mathrm{obs}}| - |k|F_{\mathrm{calc}}|}{\sum_i |F_{\mathrm{obs}}|} \tag{129}$$

Typical R-factors are below 0.2 for a well-refined macromolecular structure.

Beside the atomic coordinates x, y, z, the atomic temperature factor B may be refined at a resolution better than 3.5 Å. This is done in most programs in a separate step where, for example, in program CNS, the target function (Eq. 130)

$$T = E_{\mathrm{XRAY}} + E_R \tag{130}$$

is minimized, where (Eq. 131)

$$\begin{aligned} E_R = &W_B \sum_{(i,j)\text{-bonds}} \frac{(B_i - B_j)^2}{\sigma^2_{\mathrm{bonds}}} \\ &+ W_B \sum_{(i,j,k)\text{-angles}} \frac{(B_i - B_k)^2}{\sigma^2_{\mathrm{angles}} + W_B} \\ &\times \sum_{k\text{-group}} \sum_{j\text{-equivalences}} \sum_{\substack{i\text{-unique}\\ \text{atoms}}} \frac{(B_{ijk} - \overline{B_{ijk}})^2}{\sigma^2_{\mathrm{ncs}}} \end{aligned} \tag{131}$$

The last term is used only if noncrystallographic symmetry restraints should be imposed on the molecules. Normally, isotropic B-factors are applied and refined in macromolecular crystallography only. Even for a high-resolution structure (1.7 Å), the ratio of observations (observed structure factors) to parameters to be

refined (x, y, z, B for each atom) is only about 3. Therefore, as many additional "observations" (energy or stereochemistry restraints) as possible are incorporated. In some cases, it is useful to refine the individual occupancy of certain atoms such as bound metal ions or solvent atoms. This must be performed in a separate step.

All of the mentioned refinement procedures are based on the least-squares method. The radius of convergence for this method is not very high because it follows a downhill path to its minimum. If the model is too far away from the correct solution, the minimization may end in a local minimum corresponding to an incorrect structure. Brünger introduced the method of molecular dynamics (MD), which is able to overcome barriers in the S-function and find the correct global minimum. MD calculations simulate the dynamic behavior of a system of particles. The basic idea of the MD refinement technique is to increase the temperature sufficiently high for the atoms to overcome energy barriers and then to cool slowly to approach the energy minimum. This MD protocol is designated as simulated annealing (SA). The crystallographic application of the MD or SA technique includes a crystallographic term D, as given in Eq. (126), treated as a pseudoenergy term. A crystallographic MD or SA refinement is capable of overcoming a high-energy barrier occurring in the flipping of a peptide plane. It can be useful in removing model bias from the system.

A new approach is the refinement of macromolecular structures by the maximum likelihood method. Programs working on the basis of this method are REFMAC and CNS. The results derived using the maximum likelihood residual are consistently better than those from least-squares refinement.

If the resolution of a biological macromolecular crystal structure is equal to or better than 1.2 Å, it is in the range of real atomic resolution, and the ratio of observations to parameters is high enough to carry out, in principle, an unrestrained crystallographic refinement. Advances in cryogenic techniques, area detectors, and the use of synchrotron radiation enabled macromolecular data to be collected to atomic resolution for an increasing number of proteins. SHELXL is a program with all tools for the crystallographic refinement of biological macromolecules at real atomic resolution.

Since the advent of structural genomics, automation of all parts of structure analysis has become of paramount importance. A first step in this direction was the compilation of the Arp/Warp program system. This system allows the building and refinement of a protein model automatically and without user intervention, starting from diffraction data extending to a resolution higher than 2.3 Å and reasonable estimates of crystallographic phases. The method is based on an iterative procedure that describes the electron density map as a set of unconnected atoms and then searches for protein-like patterns. Automatic pattern recognition (model building) combined with refinement permits a structural model to be obtained reliably within few CPU hours.

5.3 Accuracy and Verification of Structure Determination

A measure of the quality of a structure determination is the crystallographic R-factor given in Eq. (129). For a high-resolution structure, for example 1.6 Å, it should not be much larger than 0.16. As this R-factor is an overall number, it does not indicate major local errors. This can be

obtained by the evaluation of a real space R-factor, which is calculated on a grid for nonzero elements according to Eq. (132):

$$R_{\text{real space}} = \frac{\sum |\rho_{\text{obs}} - \rho_{\text{calc}}|}{\sum |\rho_{\text{obs}} + \rho_{\text{calc}}|} \qquad (132)$$

where ρ_{obs} is the observed and ρ_{calc} is the calculated electron density.

It has been shown that the conventional R-factor may reach rather low values in a crystallographic refinement with structural models that turned out to be wrong later. To overcome this unsatisfactory situation, Brünger proposed the additional calculation of a so-called free R-factor. For this purpose, the reflections are divided into a working set (e.g. 90%) and a test set (e.g. 10%). The reflections in the working set are used in the crystallographic refinement. The free R-factor is calculated with reflections from the test set, which were not used for the crystallographic refinement and is thus unbiased by the refinement process. There exists a high correlation between the free R-factor and the accuracy of the atomic model phases.

The accuracy of the final model expressed by the mean coordinate error can be determined alternatively by a Luzzati or σ_A plot. The mean coordinate error for a macromolecular structure determined with a resolution of 2.0 Å and a crystallographic R-factor of 0.2 is in the region of ±0.2 Å.

The stereochemistry of the final model must also be checked. The root-mean-square deviation of bond lengths and bond angles from ideal geometry should not be greater than 0.015 Å and 3.0° respectively. The conformation of the main-chain folding is verified by a Ramachandran plot. The dihedral angles Φ and Ψ are plotted against each other for each residue. The data points should lie in the allowed regions of the plot, which correspond to energetically favorable secondary structures such as α-helices, ß-sheets, and defined turn structures. Exceptions are glycine residues, which may occur at any position in the Ramachandran plot. Further stereochemical parameters to be checked are bond lengths and angles, dihedral angles (e.g. determinating side-chain conformations), noncovalent interactions, geometry of H-bonds, and interactions in the solvent structure. This can be done with the programs PROCHECK or WHAT CHECK.

Nearly all spatial structures of biological macromolecules determined either by X-ray crystallography or nuclear magnetic resonance (NMR) techniques have been and will be deposited with the RCSB Protein Data Bank at Rutgers University. The information of the structural model is in a file that contains for each individual atom of the model a record with atom number, atom name, residue type, residue name, coordinates x, y, z, B-value(s), and occupancy. The header records hold useful information such as crystal parameters, amino acid sequence, secondary structure assignments, and references.

6 Applications

6.1 Enzyme Structure and Enzyme–Inhibitor Complex

6.1.1 X-ray Structure of Cystathionine ß-Lyase

Cystathionine ß-lyase (CBL) is a member of the γ-family of pyridoxal-5′-phosphate (PLP)-dependent enzymes that cleaves Cß-S bonds of a broad variety of substrates. The crystal structure of CBL from *Escherichia coli* has been solved using

MIR phases in combination with density modification. The enzyme has been crystallized by the hanging drop vapor diffusion method using either ammonium sulfate or PEG 400 as precipitating agent. The crystals belong to the orthorhombic space group $C222_1$ with unit cell parameters $a = 60.9$ Å, $b = 154.7$ Å, and $c = 152.7$ Å. There is one dimer per asymmetric unit. A native data set has been collected using synchrotron radiation (wavelength 1.1 Å) at the wiggler beamline BW6 at the storage ring DORIS at the Deutsches Elektronensynchrotron (DESY) in Hamburg, Germany. Data sets for three heavy atom derivatives (thiomersalate, 2-mercuri-4-diazobenzoic acid and platinum(II)-2,2′-6,6″-terpyridinium chloride) were registered on an imaging plate scanner (MAR Research, Norderstedt, Germany) using graphite monochromatized Cu Kα radiation from an RU200 rotating anode generator (Rigaku, Tokyo, Japan) operating at 5.4 kW. The reflection data were processed with the MOSFLM package and scaled with programs from the CCP4 program suite. All data sets reveal satisfactory symmetry consistency factors ($R_{\text{merge}} \leq 0.08$) and completenesses (>90%). The heavy atom positions were determined from isomorphous difference Patterson maps using the vector verification routines of program PROTEIN. All derivatives have one common heavy atom–binding site at Cys72, and one of the mercury derivatives shows a second binding site at Cys229. The dimer in the asymmetric unit is related by a local twofold axis that lies parallel to the x-axis. The translation of this local axis was determined from the distribution of the heavy atom sites in the unit cell. An initial MIR map was calculated, followed by solvent flattening, twofold averaging about the local symmetry, and density modification and phase extension to 2.5 Å resolution. Phase calculations, solvent flattening, density modification, and phase extension were done with programs of the CCP4 package. Program AVE was used for the averaging. The quality of the resulting electron density map was sufficiently high to build an almost complete atomic model. Model building was performed on an ESV-30 Graphic system workstation (Evans and Sutherland, Salt Lake City, UT, USA) using program O. The atomic model was refined by energy-restrained crystallographic refinement with XPLOR.

The final model of CBL is made up of two monomers with 391 amino acids each, one cofactor and one hydrogencarbonate molecule per monomer, and 581 solvent water molecules. The final crystallographic R-factor is 0.152 for data from 8.0- to 1.83-Å resolution, and the free R-factor is 0.221. The mean positional error of the atoms as estimated from a Luzzati plot is ±0.19 Å. A homotetramer with 222 symmetry is built up by crystallographic and noncrystallographic symmetry (Fig. 39). Each monomer of CBL (Fig. 40) can be described in terms of three spatially and functionally different domains. The N-terminal domain (residues 1–60) consists of three α-helices and one ß-strand. It contributes to tetramer formation and is part of the active site of the adjacent subunit. The second domain (residues 61–256) harbors PLP and has an α/ß structure with a seven-stranded ß-sheet as the central part. The remaining C-terminal domain (residues 257–395), connected by a long α-helix to the PLP-binding domain, consists of four helices packed on the solvent-accessible side of an antiparallel four-stranded ß-sheet. The fold of the C-terminal and the PLP-binding domain and the location of the active site are similar to aminotransferases. Most

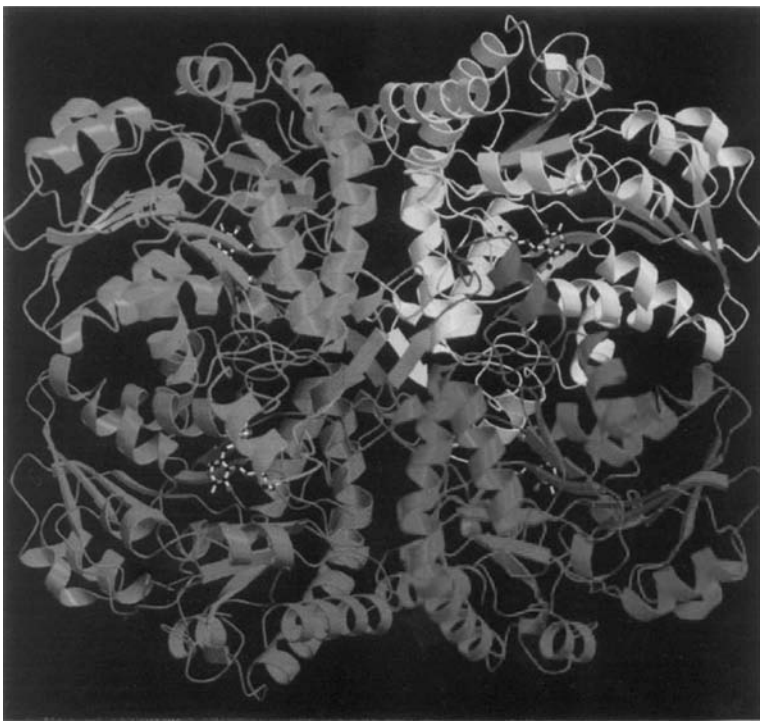

Fig. 39 Ribbon plot of the CBL tetramer viewed along the *x*-axis. The monomers are colored differently. The blue- and green-colored monomers, which are related by a crystallographic axis (horizontal, in the plane of the paper), build up one catalytic active dimer, and the yellow and red ones the other. The location of the PLP-binding site is shown in a ball-and-stick presentation; MOLSCRIPT and RASTER3D. (Reproduced by permission of Academic Press, Ltd., from Clausen, T. et al. (1996) *J. Mol. Biol.* **23**, 202–224.) (See color plate p. xxi).

of the residues in the active site are strongly conserved among the enzymes of the transulfuration pathway. Figure 41 shows the final $2F_{obs} - F_{calc}$ map superimposed with the refined atomic model around the active site of CBL. The cage-like structures for the representation of the electron density correspond to a contour level of 1.2σ. The protein part (main- and side-chain atoms), the PLP cofactor, and the hydrogencarbonate molecule are well defined in the electron density map.

The knowledge of the spatial structure of a given enzyme structure forms the basis for understanding its functional properties. It is now possible to design rational site-directed mutants or determine the structures of enzyme–substrate, enzyme–substrate analog, or enzyme–inhibitor complexes, which will deliver invaluable information in understanding the enzyme's functional properties.

6.1.2 Enzyme–Inhibitor Complex Structures of Cystathione ß-Lyase

The enzyme–inhibitor X-ray structures of ß,ß,ß-tri-fluoroalanine (TFA) and L-aminoethoxyvinylglycine (AVG) with CBL could be determined. In both cases,

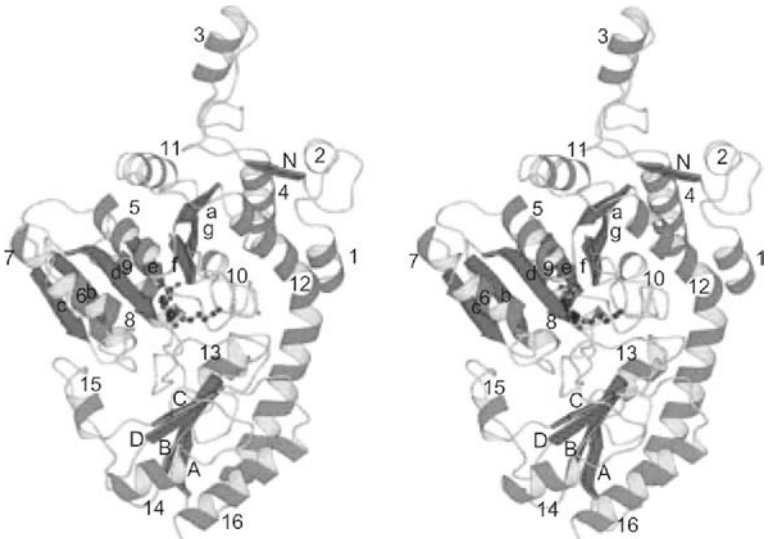

Fig. 40 Stereo ribbon presentation of the CBL monomer, emphasizing secondary structure elements. α-Helices are drawn as green spirals, ß-strands as magenta arrows. PLP and PLP-binding Lys210 are shown in a ball-and-stick representation; MOLSCRIPT and RASTER3D. (Reproduced by permission of Academic Press, Ltd., from Clausen, T. et al. (1996) *J. Mol. Biol.* **23**, 202–224.) (See color plate p. xxiii).

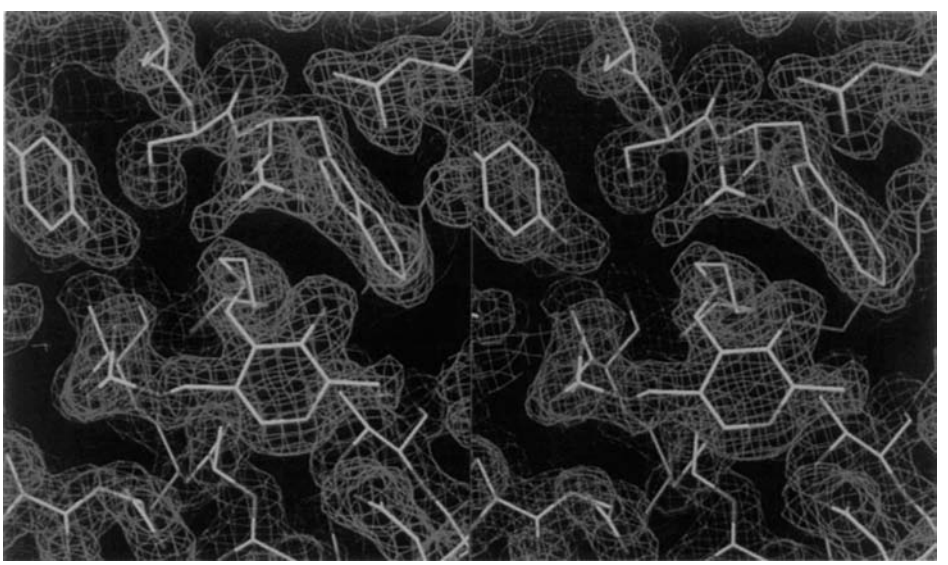

Fig. 41 Stereo plot of the electron density in the active site of CBL, superimposed with the refined model of the region around the cofactor. The $2F_{obs} - F_{calc}$ map is contoured at 1.2σ and calculated at 1.83-Å resolution. (Reproduced by permission of Academic Press, Ltd., from Clausen, T. et al. (1996) *J. Mol. Biol.* **23**, 202–224.)

crystals of the complexes were obtained by incubating the enzyme solution with inhibitor in the millimolar range and subsequent cocrystallization. The resultant crystals were isomorphous with the native enzyme, making it possible to apply the difference Fourier technique in the structure solution. The technical details for the structure analyses are given in the relevant references. The CBL/TFA complex structure was determined to substantiate that the ε-amino group of the active-site Lys210 can react with the nucleophile at the active site via Michael addition, which leads to covalent labeling and inactivation of the enzyme. The final $F_{obs} - F_{calc}$ and $2F_{obs} - F_{calc}$ electron density maps for the CBL/TFA complex around the active site are displayed in Figure 42. Clear, continuous electron density between the cofactor and Lys210 can be seen in this map, indicating a covalent lysine–inactivator–PLP product. In Fig. 42, the blue $F_{obs} - F_{calc}$ map reveals the well-defined electron density for the bound inhibitor. This binding mode of TFA to CBL corresponds to an intermediate in the reaction of TFA with CBL. The structure of the inactivation product proves that Lys210 is the active-site nucleophile reacting via Michael addition with the inactivator. It must also be the residue that transfers a proton from Cα to Sγ in the reaction with the substrates.

The CBL/AVG structure has been determined at 2.2-Å resolution and a crystallographic *R*-factor of 0.164. The X-ray structure shows that AVG binds to the PLP cofactor forming the external aldimine. Lys210 is no longer bound to the PLP cofactor. Figure 43 is an overlay of the atomic models of native CBL(magenta), CBL/TFA (yellow), and CBL/AVG (green). The main difference in inhibitor binding is the location of Cß and its substituents; in the

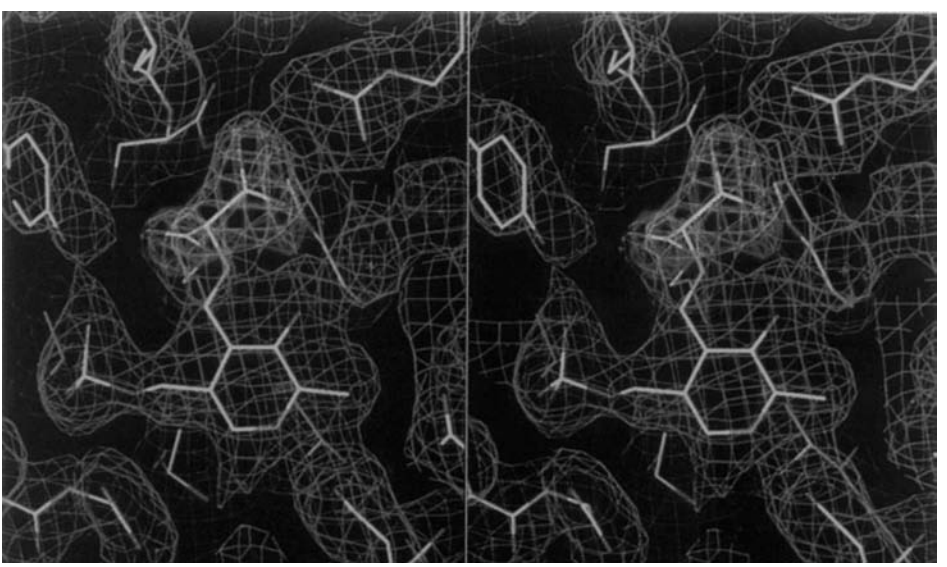

Fig. 42 $F_{obs} - F_{calc}$ (blue) and $2F_{obs} - F_{calc}$ (green) electron density map of the CBL/TFA complex around the active site contoured at 3.5σ and 1.0σ respectively, at 2.3-Å resolution. (Reproduced by permission of Academic Press, Ltd., from Clausen, T. et al. (1996) *J. Mol. Biol.* **23**, 202–224.) (See color plate p. xxiii).

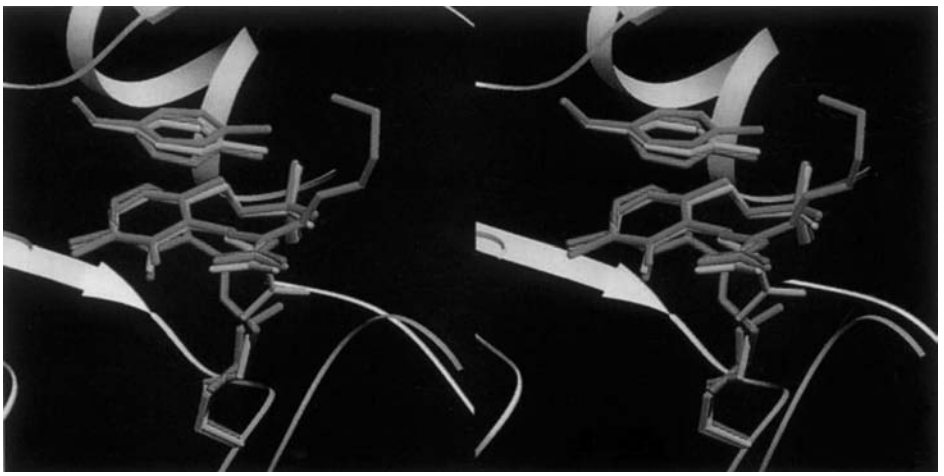

Fig. 43 Stereo view of a superposition of Tyr111 and the PLP derivative of the unliganded enzyme (magenta), the CBL/AVG adduct (green), and the TFA-inactivated enzyme (yellow); SETOR. (Reproduced by permission of the American Chemical Society, Clausen, T. et al. (1997) *Biochemistry* **36**, 12633–12643.) (See color plate p. xxiv).

TFA complex, the inactivator is directed toward the protein interior (the A face of the cofactor), whereas in CBL/AVG, Cß is located at the B side of the cofactor. The α-carboxylate group in CBL/AVG is located in the same position as the hydrogencarbonate in the native and the α-carboxylate group of TFA in the CBL/TFA complex. The terminal amino group of AVG is held in place mainly by interactions with the hydroxyl group of Tyr111.

The experimental determination of the external aldimine structure in the CBL/AVG complex is of high relevance because it can serve as a rational basis for modeling of substrate and inhibitor binding, leading to more effective herbicides.

6.1.3 Crystal Structure of the Thrombin–Rhodniin Complex

This complex structure is an example related to pharmaceutical research. The goal of this special application is the development of more efficient blood anticoagulants. The target enzyme is α-thrombin. This enzyme is a serine proteinase of trypsin-like specificity. α-Thrombin, the key enzyme in hemostasis and thrombosis, exhibits both enzymatic and hormonelike properties, and can be both pro- and anticoagulatory. Rhodniin is a highly specific inhibitor of thrombin isolated from the assassin bug *Rhodnius prolixus*. Such blood-sucking animals have developed various anticlotting mechanisms to prevent local clotting of the victim's blood. These natural thrombin inhibitors are polypeptides of 60 to 120 amino acid residues.

The crystal structure of the noncovalent complex between recombinant rhodniin and bovine a-thrombin has been determined at 2.6-Å resolution. Crystals were obtained by cocrystallization of thrombin with rhodniin in an approximately 1 : 1 molar ratio. The structure could be solved by molecular replacement because the spatial structure of the major constituent, bovine α-thrombin, was known. Only a diffraction

data set of the complex crystal had to be collected. Rotational and translational searches for the orientation and position of the thrombin molecules in the unit cell were performed with the program AMORE. The rotational search showed two solutions with correlation values of 0.22 and 0.20 over 0.09 for the next highest peak. Translational search and rigid body fitting for these two solutions resulted in a correlation value of 0.54, with the two independent complex molecules in the asymmetric unit. The quality of the electron density map calculated from the thrombin phases was good enough in principle to build the model of the rhodniin molecule (noncrystallographic averaging was also applied). The structure was refined with XPLOR to an *R*-factor of 0.189 and a free *R*-factor of 0.262.

Figure 44 shows the structure of the complex between thrombin and rhodniin as a ribbon plot, with α-helices represented as ribbon spirals and ß-strands as arrows. The N-terminal domain binds in a substrate-like manner to the narrow active-site cleft of thrombin. The C-terminal domain, whose distorted reactive-site loop cannot adopt the canonical conformation, docks to the fibrinogen recognition exosite via extensive electrostatic interactions. The peculiarity of this complex structure is that the two KAZAL-type domains of rhodniin bind to two different sites of thrombin.

6.2 Metalloproteins

Metals bound as cofactors in proteins have a great variety of functions. They may be involved in the activation of small inorganic or organic molecules, in oxygen storage and transport, in electron transport, in regulation of biological processes, or in stabilizing a transition state during enzymatic catalysis. Their role may also be solely structural. The chemistry of metals in proteins has attracted the interest of inorganic chemists, and has led to the formation of

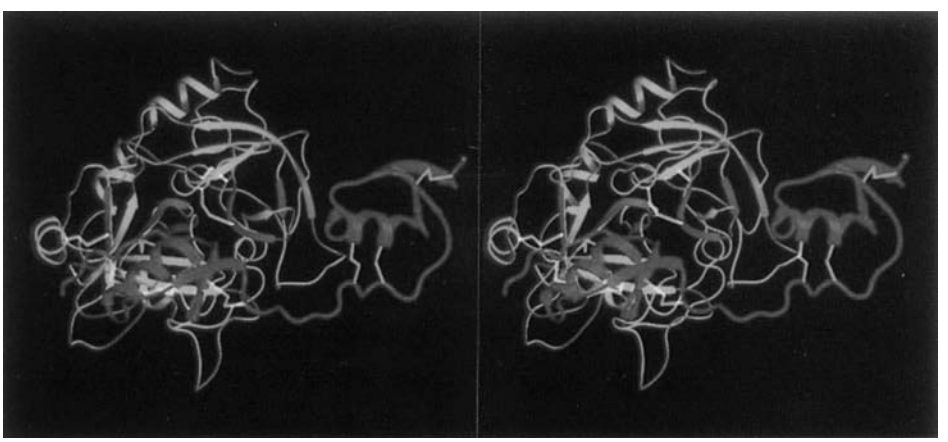

Fig. 44 Stereo view of the complex formed between thrombin (blue) and rhodniin (red) in the thrombin standard orientation, that is, with the active-site cleft facing the viewer and a bound inhibitor chain from left to right. Yellow connections indicate disulfide bridges. Rhodniin interacts through its N-terminal domain in a canonical manner with the active site and through its C-terminal domain with the fibrinogen recognition exosite of thrombin; SETOR. (Reproduced by permission of Oxford University Press, from van de Locht, A. et al. (1995) *EMBO J.* **14**, 5149–5157). (See color plate p. xxiv).

the field of biological inorganic chemistry. As an example of a complex metalloprotein, the multicopper oxidase ascorbate oxidase (AO) is presented.

6.2.1 Crystal Structure of the Multicopper Enzyme Ascorbate Oxidase

The blue protein AO belongs to the group of "blue" oxidases with laccase and ceruloplasmin. These are multicopper enzymes catalyzing the four-electron reduction of molecular oxygen with concomitant one-electron oxidation of the substrate. The crystal structure of AO has been solved by the MIR technique and refined to 1.9-Å resolution. The peculiarity of this structure determination is briefly described. It consists in the utilization of the information of two different crystal forms. In both crystal forms, the molecules arrange themselves as homotetramers with 222 symmetry, but in crystal form 1, one of these twofold axes is realized by a crystallographic twofold axis resulting in two subunits per asymmetric unit. In crystal form 2, one homotetramer is found per asymmetric unit. In crystal form 1, six isomorphous heavy atom derivatives could be found and interpreted. An initial MIR map was calculated, solvent flattened, and averaged about the local twofold axis. For crystal form 2, no phase information was available. From the averaged uninterpreted MIR map, a whole tetramer was selected and used for rotational and translational searches in crystal form 2. This was successful and provided the necessary phase information for crystal form 2 and, additionally, the local symmetry. Now, averaging could be performed both separately in the two crystal forms and subsequently between both crystal forms. The averaged electron density was transported into both unit cells and a new macrocycle of averaging could be started. These macrocycles were used to extend the phases from 3.5 Å to the full attainable resolution. This structure analysis was the first example where a molecular replacement was carried out with an uninterpreted MIR electron density–based model.

AO is a homodimeric enzyme with a molecular mass of 70 kDa and 552 amino acid residues per subunit (zucchini). The three-domain structure and the location of the mononuclear centers and trinuclear copper centers in the AO monomer as derived from the crystal structure are shown in Fig. 45. The folding of all three domains is of a similar ß-barrel type. The mononuclear copper site is located in domain 3 and the trinuclear copper species is bound between domains 1 and 3. The coordination of the mononuclear copper site is depicted in Fig. 46. It has the four canonical type-1 copper ligands (His, Cys, His, Met), also found in plastocyanin and azurin. The copper is coordinated to the ND1 atoms of His445 and His512, the SG atom of the Cys507, and the SD atom of Met517 in a distorted trigonal geometry. This unusual coordination geometry confers this copper site with its blue color. The trinuclear copper site (see Fig. 47) has eight histidine ligands symmetrically supplied by domains 1 and 3 and two oxygen ligands. The trinuclear copper site may be divided into a pair of copper (CU2, CU3), with six histidine ligands in a trigonal prismatic arrangement. The pair is bridged by an OH^-, which leads to a strong antiferromagnetic coupling and makes this copper pair electron paramagnetic resonance silent. The remaining copper has two histidine ligands and an OH^- or H_2O ligand. A binding pocket for the reducing substrate, which is complementary to an ascorbate molecule, is located near the mononuclear copper site and is accessible

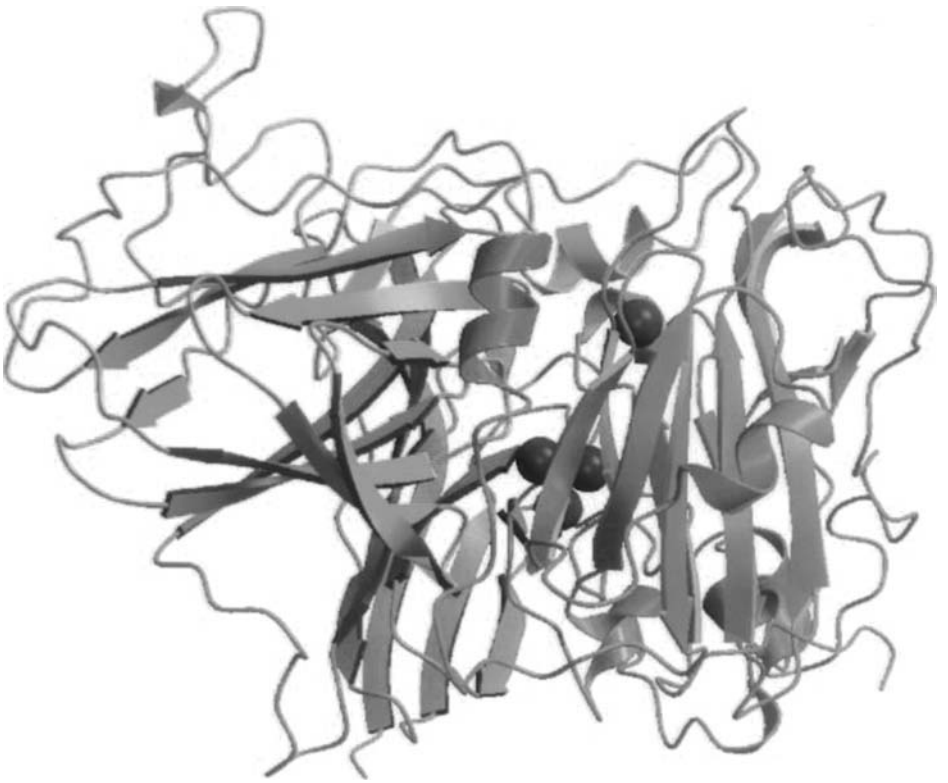

Fig. 45 Schematic representation of the monomer structure of AO. (Reproduced by permission of World Scientific Publishing Co., from Messerschmidt, A. (1997) Spatial Structures of Ascorbate Oxidase, Laccase and Related Proteins, in: Messerschmidt, A. (Ed.) *Multi Copper Oxidases*, World Scientific Publishing, Singapore.)

from solvent. A broad channel providing access from the solvent to the trinuclear copper species, which is the binding and reaction site for the dioxygen, is present in AO. During catalysis, an intramolecular electron transfer between the mononuclear copper site and the trinuclear copper cluster occurs. The distances between the mononuclear copper and the three coppers of the trinuclear center are 12.20, 12.69, and 14.87 Å, respectively. Furthermore, the crystal structures of functional derivatives of AO such as the reduced, azide, and peroxide form have been determined. They show considerable changes at the trinuclear copper site in the reduced form and the peroxide or azide binding to the trinuclear copper species in the relevant structures. The X-ray studies on AO delivered essential information in understanding the catalytic mechanism.

6.2.2 Crystal Structure of Cytochrome-*c* Nitrite Reductase Determined by Multiwavelength Anomalous Diffraction Phasing

The spatial structure of cytochrome-*c* nitrite reductase from *Sulfurospirillum deleyianum* has been solved by the MAD technique using synchrotron radiation and

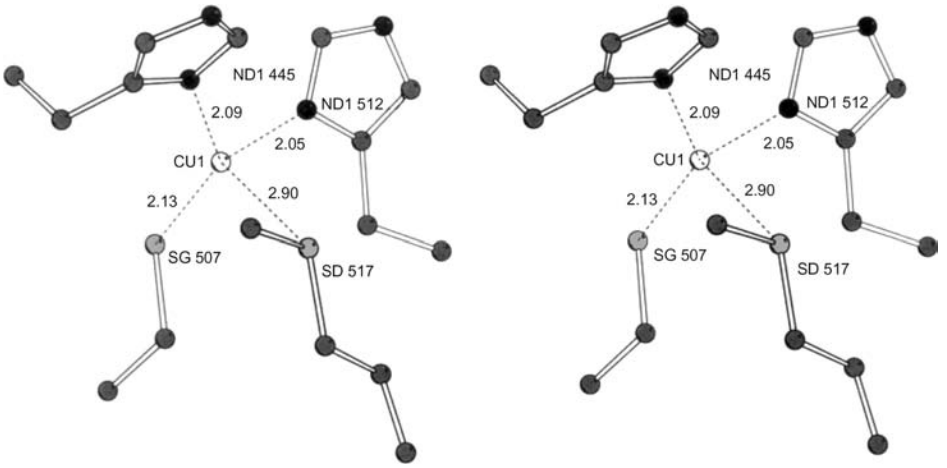

Fig. 46 Stereo drawing of the mononuclear copper site in domain 3 of AO. The displayed bond distances are for subunit A; MOLSCRIPT. (Reproduced by permission of World Scientific Publishing Co., from Messerschmidt, A. (1997) Spatial Structures of Ascorbate Oxidase, Laccase and Related Proteins, in: Messerschmidt, A. (Ed.) *Multi Copper Oxidases*, World Scientific Publishing, Singapore.)

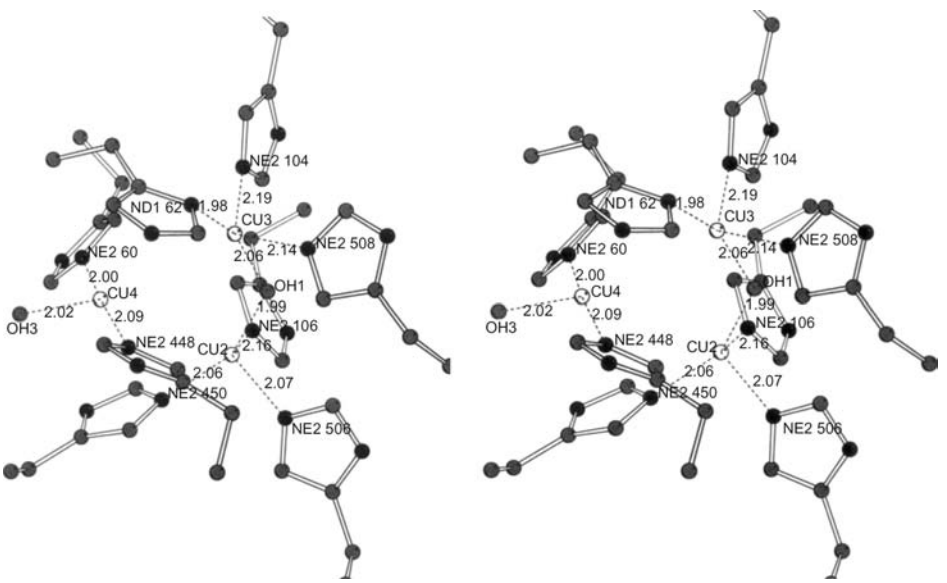

Fig. 47 Stereo drawing of the trinuclear copper site of AO. The displayed bond distances are for subunit A; MOLSCRIPT. (Reproduced by permission of World Scientific Publishing Co., from Messerschmidt, A. (1997) Spatial Structures of Ascorbate Oxidase, Laccase and Related Proteins, in: Messerschmidt, A. (Ed.) *Multi Copper Oxidases*, World Scientific Publishing, Singapore.)

refined at a resolution of 1.9 Å to a *R*-factor of 0.18. The data collection for the MAD phasing was performed at the wiggler beamline BW6 at the storage ring DORIS at DESY in Hamburg, Germany. At the beginning, an X-ray fluorescence spectrum around the Fe K absorption edge was registered using an NaI(Tl) scintillation counter. Evaluation of the spectrum with program DISCO gave the anomalous dispersion contributions, f' and f'', as a function of photon energy (Fig. 48). Subsequently, diffraction data were collected at three different photon energies. Two of them, 7141 and 7129 eV, correspond to maximum f'' and minimum f' respectively. A third high-resolution data set (1.90 Å resolution) was collected at a photon energy of 11 808 eV. All data sets were obtained from a single crystal that had been shock-frozen at 100 K. Anomalous difference Patterson maps were calculated from the data collected at f'' maximum of the Fe K absorption edge. Correlated peaks that appeared in all three Harker sections were chosen and used for phasing with MLPHARE. The highest peaks from the resulting Fourier map were analyzed for consistency with the anomalous difference Patterson map and, if correct, used for a new cycle of phase calculations. Thus, it was possible successively to find 15 iron positions, which divided into three groups of five. Phasing with program SHARP produced an interpretable electron density map, which was solvent flattened and threefold averaged with program AVE. The resulting electron density map at a resolution of 1.9 Å was of high quality and allowed the construction of the complete atomic model into that map.

Cytochrome-*c* nitrite reductase (58 kDa) catalyzes the six-electron reduction of nitrite to ammonia as one of the key steps in the biological nitrogen cycle, where it participates in the anaerobic energy metabolism of dissimilatory nitrate ammonification. The crystal structure shows that the enzyme is a homodimer

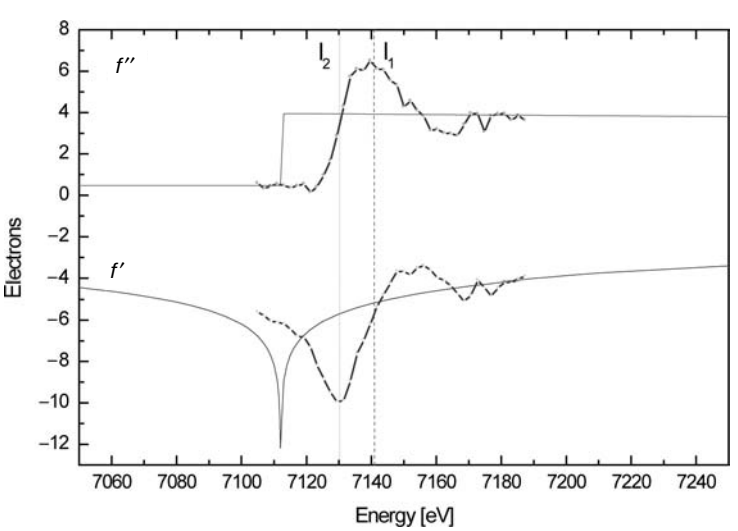

Fig. 48 Dispersion terms as a function of the photon energy as obtained from an X-ray fluorescence scan from cytochrome-*c* nitrite reductase sample crystal and evaluation with DISCO.

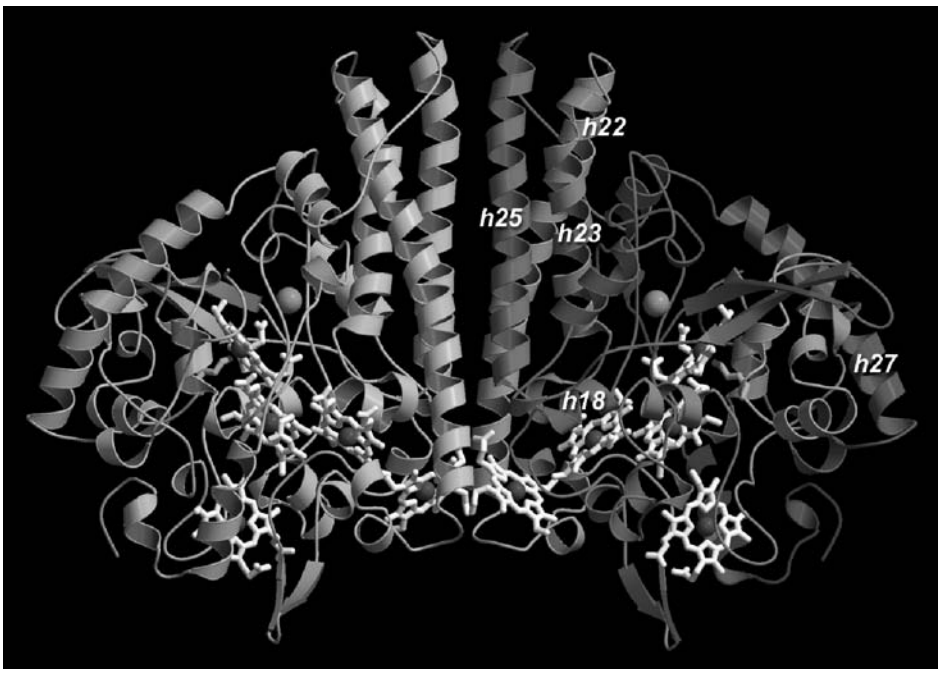

Fig. 49 Overall structure of the nitrite reductase dimer. A front view with the dimer axis oriented vertically, five hemes in each monomer (white), the Ca^{2+} ions (grey), and Lys133 that coordinates the active-site iron (yellow). The dimer interface is dominated by three long α-helices per monomer. All hemes in the dimer are covalently attached to the protein. (Reproduced by permission of Macmillan Magazines, Ltd., from Einsle, O. et al. (1999) *Nature* (*London*) **400**, 476–480.) (See color plate p. xxii).

of ~100 Å × 80 Å × 50 Å, with 10 hemes in a remarkably close packing (Fig. 49). The protein folds into one compact domain with α-helices as the predominant secondary structural motif, ranging from short helical turns to four long helices at the C-terminal end of the peptide chain (Fig. 49). There are eight 3_{10} helices three- or four-residues long, seven of which occur within the first 200 residues, ß-sheet structures are found only in two short, antiparallel strands, where one is part of a funnel-like cavity leading to the active site. Dimer formation is mediated by helices *h*22 and *h*25, with helix *h*25 as the key element, as it interacts with its counterpart in the other monomer over its full length of 28 residues, corresponding to 42 Å. The five hemes in the monomer of cytochrome-*c* nitrite reductase are in close contact, with Fe-Fe distances of between 9 and 12.8 Å. They are arranged as a group of three, almost coplanar, hemes, with heme 1 forming the active site. Hemes 2 and 5 are farther apart and are not coplanar with hemes 1, 3, and 4. All hemes except heme 1 are bis-histidynyl-coordinated, and are linked to the peptide backbone by thioether bonds to the cysteine residues of a classical heme-binding motif for periplasmic proteins, Cys-X_1-X_2-Cys-His. The propionate side chains of heme 1 form part of the active-site cavity. The site of nitrite reduction is clearly heme

1, with NZ atom of Lys133, replacing a histidine in the classical binding motif, and an oxygen atom of a sulfate ion, present in the crystallization buffer. An additional electron density maximum detected close to the active site was assigned to Ca^{2+}.

S. deleyianum cytochrome-*c* nitrite reductase reduces not only nitrite to ammonia but also the potential intermediates NO and hydroxylamine. No intermediates are released during nitrite turnover. Obviously, the active site accomodates anions and uncharged molecules and releases the ammonium cation only after full six-electron reduction. The preference for anions is reflected by a positive electrostatic potential around and inside the active-site cavity. The cationic product might make use of a second channel leading to the protein surface opposite to the entry channel. It branches before reaching the protein surface and ends with both arms in areas possessing a significant electrostatic surface potential. The existence of separate pathways for substrate and product with matched electrostatic potential could also contribute to the enzyme's high specific activity compared to the siroheme-containing reductases that catalyze the same reaction.

6.3 Large Molecular Assembly

6.3.1 Crystal Structure of 20S Proteasome from Yeast

The controlled degradation of proteins in the interior of cells is of central significance for many processes that range from cell cycle control and differentiation to cellular immune response. The target protein is labeled for destruction by the covalent linkage of a small protein called *ubiquitin* and is degraded after adenosine triphosphate (ATP)-driven unfolding in the closed internal chamber of a large protein complex, which is known as *26S proteasome* (molecular mass 2 000 000 Da). The core and the proteolytic chamber of the 26S proteasome are formed by the catalytic 20S particle (molecular mass 700 000 Da). This particle is flanked at each end by a so-called 19S cap, which seems to be responsible for the recognition and unfolding of the ubiquitin-labeled target proteins. The eukaryotic catalytic 20S machinery consists of 14 different but related protein subunits that assemble to the overall structure of 28 protein chains, a monster of the molecular world. The structure of the 20S proteasome from *Saccharomyces cerevisiae* has been elucidated at atomic resolution. This is one of the most complex structures that has been determined up to now, excluding symmetrical systems such as viral capsids. It shows that complexity itself is no limit to understanding a structure on the atomic level.

The structure analysis is briefly outlined. Crystals could be obtained by the hanging drop vapor diffusion method with a final precipitant concentration of 12% MPD in the drop (pH 6.5). The crystals are monoclinic, space group $P2_1$, cell parameters $a = 135.7$ Å, $b = 301.8$ Å, $c = 144.7$ Å, and ß $= 112.6°$, and have one 20S particle per asymmetric unit. Data sets with two different inhibitors were collected at the wiggler beamline BW6 at DESY with radiation of $\lambda = 1.1$ Å and at 90 K temperature. The mean attainable resolution was 2.4 Å. A self-rotation function calculated at 5-Å resolution revealed the orientation of the local twofold axis. The atomic model of the 20S proteasome from *Thermoplasma acidophilum* could be used for molecular replacement calculations (program AMORE) at 3.5-Å resolution, and a prominent solution could be found. The resultant electron density map was cyclically averaged about the local

twofold axis. The individual subunits were identified according to their characteristic amino acid sequences and were built into the map. The final model with the lactacystin inhibitor was refined at a resolution of 2.4 Å to an *R*-factor of 0.26, consisting of 48 888 proteins, 30 inhibitors, 18 magnesium atoms, and 1800 solvent molecules.

The structure of the eukaryotic proteasome is important because it is much more complex than its archaebacterial relative, which has two types of subunits, α and ß, only. The α-subunits do not seem to be catalytic but they may self-assemble to form sevenfold rings. In contrast, the ß-subunits show catalytic activity but are not able to assemble themselves. The two components assemble themselves to a stack of four rings, two outer α-rings, which enclose an inner pair of ß-rings. The eukaryotic proteasome retains the important property, however, that the single subunits of the sevenfold rings are different, probably reflecting the increased biological functions. The exact sevenfold symmetry of the particle is lost and a twofold symmetry remains solely to give seven different α- and seven different ß-subunits (Fig. 50a and b). In the proteasome, the unfolded protein is cut

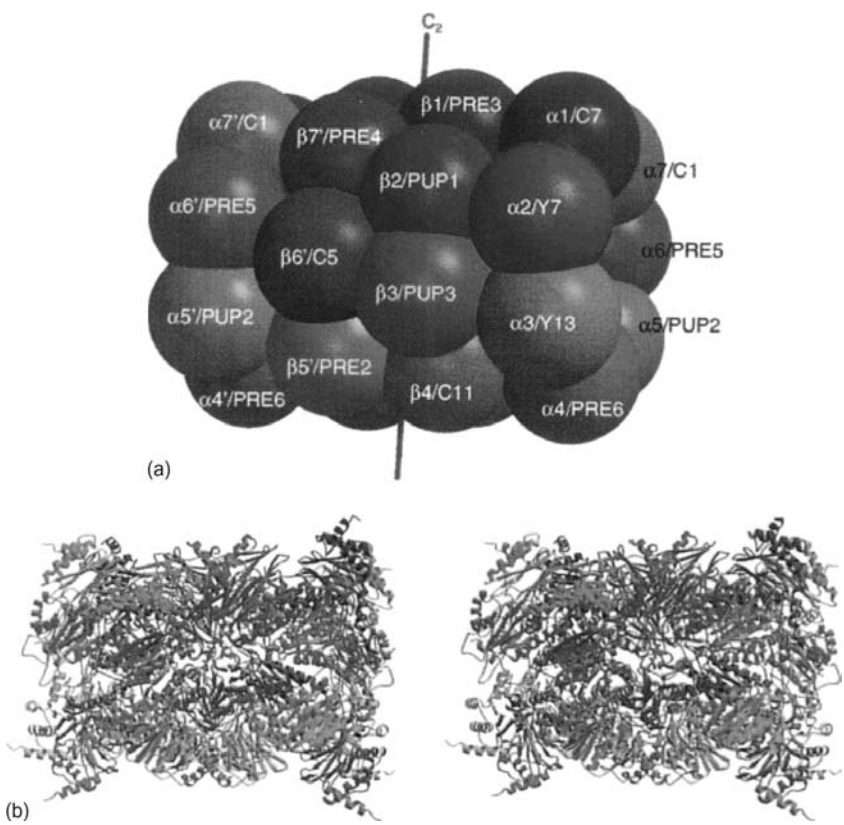

Fig. 50 Topology of the 28 subunits of the yeast 20S proteasome drawn as (a) spheres and (b) ribbon representation; MOLSCRIPT, RASTER3D. (Reproduced by permission of Macmillan Magazines Ltd., from Groll, M. et al. (1997) *Nature (London)*, **386**, 463–471.)

into peptide products in the inner chamber of the 20S particle, and these have a length distribution with a center around octa- or nonapeptides. These sizes are appropriate to bind to MHC-class-I molecules.

See also Chaperones, Molecular.

Bibliography

Books and Reviews

Arndt, U.W., Wonnacott, A.J. (Eds.) (1977) *The Rotation Method in Crystallography*, North Holland, Amsterdam, The Netherlands.

Blundell, T.L., Johnson, L.N. (1976) *Protein Crystallography*, Academic Press, New York.

Brünger, A.T. (1992) *XPLOR, Version 3.1, a System for Crystallography and NMR*, Yale University Press, New Haven, CT.

Buerger, M.J. (1961) *Crystal-structure Analysis*, John Wiley & Sons, New York.

Buerger, M.J. (1964) *The Precession Method in X-ray Crystallography*, John Wiley & Sons, New York, 1964.

Carter, C.W. Jr, Sweet, R.M. (Eds.) (1997) *Macromolecular crystallography*, Part A, *Methods Enzymol.* **276**, 1–700.

Carter, C.W. Jr, Sweet, R.M. (Eds.) (1997) *Macromolecular crystallography*, Part B, *Methods Enzymol.* **277**, 1–664.

Drenth, J. (1999) *Principles of Protein X-ray Crystallography*, Springer, New York.

Glusker, J.P., Lewis, M., Rossi, M. (1994) *Crystal Structure Analysis for Chemists and Biologists*, VCH Publishers, New York.

Hahn, T. (Ed.) (1997), *International Tables for Crystallography*, Vol. A. D. Reidel, Dordrecht, The Netherlands.

Helliwell, J.R. (1992) *Macromolecular Crystallography with Synchrotron Radiation*, Cambridge University Press, Cambridge, MA.

McRee, D. (1999) *Practical Protein Crystallography*, Academic Press, San Diego, CA.

Michel, H. (1991) *Crystallization of Membrane Proteins*, CRC Press, Boca Raton, FL.

Wyckoff, H.W., Hirs, C.H.W., Timasheff, S.N. (Eds.) (1985) *Diffraction Methods for Biological Macromolecules*, Part A, *Methods Enzymol.* **114**, 1–588.

Wyckoff, H.W., Hirs, C.H.W., Timasheff, S.N. (Eds.) (1985) *Diffraction Methods for Biological Macromolecules*, Part B, *Methods Enzymol.* **115**, 1–485.

Primary Literature

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. (2000) The protein data bank, *Nucleic Acids Res.* **28**, 235–242.

Blow, D.M., Crick, F.H.C. (1959) The treatment of errors in the isomorphous replacement method, *Acta Crystallogr.* **12**, 794–802.

Brünger, A.T. (1990) Extension of molecular replacement: new search strategy based on Patterson correlation refinement, *Acta Crystallogr.* **A46**, 46–57.

Brünger, A.T. (1992) Free R value: A novel statistical quantity for assessing the accuracy of crystal structures, *Nature (London)*, **355**, 472–475.

Brünger, A.T. (1993) Assessment of phase accuracy by cross validation: the free R value. Methods and applications, *Acta Crystallogr.* **D49**, 24–36.

Brünger, A.T., Nilges, M. (1993) Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR spectroscopy, *Q. Rev. Biophys.* **26**, 49–125.

Brünger, A.T., Kuriyan, J., Karplus, M. (1987) Crystallographic *R*-factor refinement by molecular dynamics, *Science* **235**, 458–460.

Brünger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L. (1998) Crystallography and NMR system – a new software suite for macromolecular structure determination, *Acta Crystallogr.* **D54**, 905–921.

CCP4, (1994) Collaborative computational project No. 4, *Acta Crystallogr.* **D50**, 760–763.

Clausen, T., Huber, R., Laber, B., Pohlenz, H.-D., Messerschmidt, A. (1996) Crystal structure of the pyridoxal-5′-phosphate dependent Cystathionine ß-Lyase from *Escherichia coli* at 1.83 Å, *J. Mol. Biol.* **262**, 202–224.

Clausen, T., Huber, R., Messerschmidt, A., Pohlenz, H.-D., Laber, B. (1997) Slow-binding inhibition of *Escherichia coli* Cystathionine ß-Lyase by L-aminoethoxyvinylglycine: a

kinetic and X-ray study, *Biochemistry* **36**, 12633–12643.

Crick, F.H.C., Magdoff, B.S. (1956) The theory of the method of isomorphous replacement, *Acta Crystallogr.* **9**, 901–908.

Cromer, D.T., Liberman, D. (1970) Relativistic calculation of anomalous scattering factors for X-rays, *J. Chem. Phys.* **53**, 1891–1898.

Crowther, R.A., Blow, D.M. (1967) A method of positioning a known molecule in an unknown crystal structure, *Acta Crystallogr.* **23**, 544–548.

Dauter, Z., Lamzin, V.S., Wilson, K.S. (1997) The benefits of atomic resolution, *Curr. Opin. Struct. Biol.* **7**, 681–688.

Diamond, R. (1969) Profile analysis in single crystal diffractometry, *Acta Crystallogr.* **A25**, 43–55.

Dickerson, R.E., Weinzierl, J.E., Palmer, R.A. (1968) A least-squares refinement method for isomorphous replacement, *Acta Crystallogr.* **B24**, 997–1003.

Dodson, E., Vijayan, M. (1971) The determination and refinement of heavy-atom parameters in protein heavy-atom derivatives: some model calculations using acentric reflections, *Acta Crystallogr.* **B27**, 2402–2411.

Eichhorn, K.D. (1985) DISCO: Calculation of the Anomalous Dispersion Corrections, f' and f'', to the Atomic X-ray Form Factor from Both EXAFS and Theory, Deutsches Elektronensynchrotron DESY, Hamburg.

Einsle, O., Messerschmidt, A., Stach, P., Bourenkov, G.P., Bartunik, H.D., Huber, R., Kroneck, P.M.H. (1999) Structure of cytochrome *c* nitrite reductase, *Nature (London)* **400**, 476–480.

Engh, R.A., Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement, *Acta Crystallogr.* **A47**, 392–400.

Evans, S.V. (1993) SETOR: Hardware lighted three-dimensional solid model representation of macromolecules, *J. Mol. Graphics* **11**, 134–138.

Ford, G. (1974) Intensity determination by profile fitting applied to precession photographs, *J. Appl. Crystallogr.* **7**, 555–564.

Gaykema, W.P.J., Hol, W.G.J., Vereijken, J.M., Soeter, N.M., Bak, H.J., Beintema, J.J. (1984) 3.2 Å structure of the copper-containing, oxygen-carrying protein *Panulirus interruptus* haemocyanin, *Nature (London)* **309**, 23–29.

Gilliland, G.L., Tung, M., Blakeslee, D.M., Ladner, J.E. (1994) Biological macromolecule crystallization database, version 3.0: new features, data and the NASA archive for protein crystal growth data, *Acta Crystallogr.* **D50**, 408–413.

Green, A.A. (1931) Studies in the physical chemistry of the proteins. VIII. The solubility of hemoglobin in concentrated salt solutions. A study of the salting out of proteins, *J. Biol. Chem.* **93**, 495–516.

Green, A.A. (1932) Studies in the physical chemistry of the proteins. X. The solubility of hemoglobin in solutions of chlorides and sulfates of varying concentration, *J. Biol. Chem.* **95**, 47–66.

Greer, J. (1974) Three-dimensional pattern recognition: an approach to automated interpretation of electron density maps of proteins, *J. Mol. Biol.* **82**, 279–301.

Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H.D., Huber, R. (1997) Structure of 20S proteasome from yeast at 2.4 Å resolution, *Nature (London)* **386**, 463–471.

Hendrickson, W.A. (1985) Stereochemically restrained refinement of macromolecular structures, *Methods Enzymol.* **115**, 252–270.

Hendrickson, W.A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation, *Science* **254**, 51–58.

Hendrickson, W.A., Lattman, E.E. (1970) Representation of phase probability distributions for simplified combination of independent phase information, *Acta Crystallogr.* **B26**, 136–143.

Hendrickson, W.A., Ogata, C.M. (1997) Phase determination from Multiwavelength Anomalous Diffraction measurements, *Methods Enzymol.* **276**, 494–523.

Hendrickson, W.A., Horton, J.R., LeMaster, D.M. (1990) Selenomethionyl proteins produced for analysis by Multiwavelength Anomalous Diffraction (MAD): a vehicle for direct determination of three-dimensional structure, *EMBO J.* **9**, 1665–1672.

Hönl, H. (1933) Zur Dispersionstheorie der Röntgenstrahlen, *Z. Phys.* **84**, 1–16.

Hooft, R., Vriend, G. (1996) WHAT IF Program Manual, EMBL, Heidelberg, Chapter 23.

Hoppe, W. (1957) Die Faltmolekülmethode: Eine neue Methode zur Bestimmung der Kristallstruktur bei ganz oder teilweise bekannten Molekülstrukturen, *Z. Elektrochem.* **61**, 1076–1083.

Howard, A.J., Nielsen, C., Xuong, Ng.H. (1985) Software for a diffractometer with multiwire area detector, *Methods Enzymol.* **114**, 452–472.

Howard, A.J., Gilliland, G.L., Finzel, B.C., Poulos, T.L., Ohlendorf, D.H., Salemne, F.R. (1987) The use of an imaging proportional counter in macromolecular crystallography, *J. Appl. Crystallogr.* **20**, 383–387.

Huber, R. (1965) Die Automatisierte Faltmolekülmethode, *Acta Crystallogr.* **19**, 353–356.

Jack, A., Levitt, M. (1978) Refinement of large structures by simultaneous minimization of energy and *R*-factor, *Acta Crystallogr.* **A34**, 931–935.

James, R.W. (1960) *The Optical Principles of the Diffraction of X-rays*, Vol. II The Crystalline State, Bell, London, 135–167.

Jancarik, J., Kim, S.H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins, *J. Appl. Crystallogr.* **24**, 409–411.

Jones, T.A. (1978) A graphics model building and refinement system for macromolecules, *J. Appl. Crystallogr.* **15**, 24–31.

Jones, T.A., Zou, J.Y., Cowan, S.W., Kjelgaard, M. (1991) Improved methods for building protein models in electron density maps and location of errors in these models, *Acta Crystallogr.* **A47**, 110–119.

Kabsch, W. (1988) Automatic indexing of rotation diffraction patterns, *J. Appl. Crystallogr.* **21**, 67–81.

Kabsch, W. (1988) Evaluation of single-crystal X-ray diffraction data from position-sensitive detector, *J. Appl. Crystallogr.* **21**, 916–924.

Kabsch, W. (1993) *Data Collection and Evaluation with Program XDS*, in Data Collection and Processing, Proceedings of the CCP4 Study Weekend, 29–30 January 1993, Sawyer, L. Isaac, N. Bailey S., (Eds.) SERC Daresbury Laboratory, Warrington, UK, 63–70.

Kahn, R., Fourme, R. (1997) Gas proportional detectors, *Methods Enzymol.* **276**, 268–288.

Kiefersauer, R., Than, M.E., Dobbek, H., Gremer, L., Melero, M., Strobl, S., Dias, J.M., Soulimane, T., Huber, R. (2000) A novel free-mounting system for protein crystals: transformation and improvement of diffraction power by accurately controlled humidity changes, *J. Appl. Cryst.* **33**, 1223–1230.

Kleywegt, G.J., Jones, T.A. (1994) 'Halloween . . . Masks and Bones', in *From First Map to Final Model*, Bailey, S., Hubbard, R., Waller, D., (Eds.) Proceedings of the Study Weekend, SERC Daresbury Laboratory, Warrington, UK, 59–66.

Knäblein, J., Neuefeind, T., Schneider, F., Bergner, A., Messerschmidt, A., Löwe, J., Steipe, B., Huber, R. (1997) $Ta_6Br_{12}^{2+}$, a tool for phase determination of large biological assemblies by X-ray crystallography, *J. Mol. Biol.* **270**, 1–7.

Kraulis, P.J. (1991) MOLSCRIPT:A program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr.* **24**, 946–950.

Ladenstein, R., Schneider, M., Huber, R., Bartunik, H., Schott, K., Bacher, A. (1988) Heavy riboflavin synthase from bacillus subtilis: crystal structure analysis of the icosahedral Beta-60 capsid at 3.3 Å resolution, *J. Mol. Biol.* **203**, 1045–1070.

de La Fortelle, E., Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods, *Methods Enzymol.* **276**, 472–494.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M. (1993a) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* **26**, 283–291.

Laskowski, R.A., Moss, D.S., Thornton, J.M. (1993b) Main-chain bond lengths and bond angles in protein structures, *J. Mol. Biol.* **231**, 1049–1067.

Leslie, A. (1994) *Mosflm User Guide*, Mosflm Version 5.41', in Data Collection and Processing, Proceedings of the CCP4 Study Weekend, 29–30 January 1993, Sawyer, L., Isaac, N., Bailey, S., (Eds.) HRC Laboratory of Molecular Biology, Cambridge, UK, 44–51.

Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., Huber, R. (1995) Crystal structure of the 20S proteasome from the archaeon *Thermoplasma acidophilum* at 3.4 Å resolution, *Science* **268**, 533–539.

van de Locht, A., Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kröger, B., Höffken, W., Bode, W. (1995) Two heads are better than one: crystal structure of the insect derived double domain Kazal Inhibitor rhodniin in complex with thrombin, *EMBO J.* **14**, 5149–5157.

Luzzati, V. (1952) Traitement Statistique des Erreurs dans la Determination des Structures Crystallines, *Acta Crystallogr.* **A5**, 802–810.

MacArthur, M.W., Laskowski, R.A., Thornton, J.M. (1994) Knowledge-based validation of protein-structure coordinates derived by X-ray crystallography and NMR spectroscopy, *Curr. Opin. Struct. Biol.* **4**, 731–737.

Merrit, E.A., Murphy, N.E.P. (1994) RASTER3D Version 2.0. A program for photorealistic molecular graphics, *Acta Crystallogr.* **D50**, 869–873.

Messerschmidt, A., Pflugrath, J.W. (1987) Crystal orientation and X-ray pattern prediction routines for area-detector diffractometer systems in macromolecular crystallography, *J. Appl. Crystallogr.* **20**, 306–315.

Messerschmidt, A., Luecke, H., Huber, R. (1993) X-ray structures and mechanistic implications of three functional derivatives of ascorbate oxidase from zucchini, *J. Mol. Biol.* **230**, 997–1014.

Messerschmidt, A., Ladenstein, R., Huber, R., Bolognesi, M., Avigliano, L., Petruzzelli, R., Rossi, A., Finazzi-Agro, A. Refined crystal structure of ascorbate oxidase at 1.9 Å resolution, *J. Mol. Biol.* **224**, 179–205.

Messerschmidt, A., Rossi, A., Ladenstein, R., Huber, R., Bolognesi, M., Gatti, G., Marchesini, A., Petruzzelli, R., Finazzi-Agro, A. (1989) X-ray crystal structure of the blue oxidase ascorbate oxidase from zucchini: analysis of the polypeptide fold and a model of the copper sites and ligands, *J. Mol. Biol.* **206**, 513–529.

Moffat, K. (1998) Ultrafast time-resolved crystallography, *Nature Struct. Biol., Synchrotron Suppl.* 641–643.

Murshudov, G.N., Vagin, A.A., Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum likelihood method, *Acta Crystallogr.* **D53**, 240–255.

Navaza, J. (1994) AMoRe: An automated package for Molecular Replacement, *Acta Crystallogr.* **A50**, 157–163.

Otwinowski, Z. (1991) Maximum Likelihood Refinement of Heavy Atom Parameters, in *Isomorphous Replacement and Anomalous Scattering*, Wolf, W., Evans, P.R., Leslie, A.G.W., (Eds.) SERC Daresbury Laboratory, Warrington, UK, 80–86.

Otwinowski, Z. (1993) *The Denzo Program Package*, in Data Collection and Processing, Proceedings of the CCP4 Study Weekend, 29–30 January 1993, Isaac, N., Bailey, S., (Eds.) SERC Daresbury Laboratory, Warrington, UK, 56–62.

Otwinowski, Z., Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol.* **276**, 307–326.

Patterson, A.L. (1934) A Fourier series method for the determination of the components of interatomic distances in crystals, *Phys. Rev.* **46**, 372–376.

Pearson, R.G. (1969) Hard and soft acids and bases, *Survey* **5**, 1–52.

Perrakis, A., Morris, R., Lamzin, V.S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol.* **6**, 458–463.

Phillips, W.C., Rayment, I. (1985) A systematic method for aligning double-focusing mirrors, *Methods Enzymol.* **114**, 316–329.

Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V.J. (1963) Stereochemistry of polypeptide chain configurations, *J. Mol. Biol.* **7**, 95–99.

Read, R.J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors, *Acta Crystallogr.* **A42**, 140–149.

Read, R.J. (1990) Structure-factor probabilities for selected structures, *Acta Crystallogr.* **A46**, 900–912.

Rodgers, D.W. (1997) Practical cryocrystallography, *Methods Enzymol.* **276**, 183–203.

Rossmann, M.G. (1960) The accurate determination of the position and shape of heavy-atom replacement groups in proteins, *Acta Crystallogr.* **13**, 221–226.

Rossmann, M.G. (1979) Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting, *J. Appl. Crystallogr.* **12**, 225–238.

Rossmann, M.G., Blow, D.M. (1962) The detection of subunits within the crystallographic asymmetric unit, *Acta Crystallogr.* **15**, 24–31.

Rossmann, M.G., Arnold, E., Erikson, J.W., Frankenberger, E.A., Griffith, J.P., Hecht, H.-J., Johnson, J.E., Kamer, G., Luo, M., Mosser, A.G., Rueckert, R.R., Sherry, B., Vriend, G. (1985) Structure of a human common cold virus and functional relationship to other picornaviruses, *Nature (London)* **317**, 145–153.

Roussel, A., Cambilleau, C. (1989) Turbo-Frodo in Silicon Graphics Geometry, Partners Directory, Silicon Graphics, Mountain View, CA.

Schneider, R.R., Sheldrick, G.M. (2002) Substructure solution with *SHELXD*, *Acta Crystallogr.* **D58**, 1772–1779.

Sheldrick, G.M. (1997) SHELX: High-resolution refinement, *Methods Enzymol.* **277**, 319–343.

Sim, G.A. (1959) The distribution of phase angles for structures containing heavy atoms. II. A modification of the normal heavy-atom method for noncentrosymmetrical structures, *Acta Crystallogr.* **12**, 813–815.

Steigemann, W. (1974) Die Entwicklung und Anwendung von Rechenverfahren und Rechenprogrammen zur Strukturanalyse von Proteinen am Beispiel des Trypsin-Trysininhibitorkomplexes, des freien Inhibitors und der L-Asparaginase, PhD Thesis, Technische Universität München.

Stubbs, M.T., Bode, W. (1993) A player of many parts: the spotlight falls on thrombin's structure, *Thromb. Res.* **69**, 1–58.

Teng, T.-Y. (1990) Mounting of crystals for macromolecular crystallography in a free-standing thin film, *J. Appl. Crystallogr.* **23**, 387–391.

Tong, L., Rossmann, M.G. (1990) The locked rotation function, *Acta Crystallogr.* **A46**, 783–792.

Tronrud, D.E., Ten Eyk, L.F., Matthews, B.W. (1987) An efficient general-purpose least-squares refinement program for macromolecular structures, *Acta Crystallogr.* **A43**, 489–501.

Turk, D. (1992) Weiterentwicklung eines Programmes für Molekülgrafik und Elektronendichte-Manipulation und seine Anwendung auf verschiedene Protein-Struktur-Aufklärungen, PhD Thesis, Technische Universität München.

Wang, B.-C. (1985) Resolution of phase ambiguity in macromolecular crystallography, *Methods Enzymol.* **115**, 90–112.

Westbrook, E.M., Naday, I. (1997) Charge-coupled device based area detectors, *Methods Enzymol.* **276**, 244–268.

Wilson, K.S. (1998) Illuminating crystallography, *Nat. Struct. Biol., Synchrotron Suppl.* 627–630.

Xu, H.L., Hauptmann, H.A., Weeks, C.M. (2002) Sine-enhanced *Shake-and-Bake*: the theoretical basis and applications to Se-atom substructures, *Acta Crystallogr.* **D58**, 90–96.

Yang, W., Hendrickson, W.A., Kalman, E.T., Crouch, R.J. (1990) Expression, purification, and crystallization of natural and selenomethionyl recombinant ribonuclease H from *Escherichia coli*, *J. Biol. Chem.* **265**, 13553–13559.

Zhang, K.Y.J., Main, P. (1990) The use of Sayre's equation with solvent flattening and histogram matching for phase extension and refinement of protein structures, *Acta Crystallogr.* **A46**, 377–381.

Malaria Mosquito Genome

Robert A. Holt[1] *and Frank H. Collins*[2]
[1] *Canada's Michael Smith Genome Science Centre, Vancouver, BC, Canada*
[2] *University of Notre Dame, Notre Dame, IN, USA*

1 The Mosquito and Infectious Disease 471
1.1 Scope and Impact of Mosquito-borne Disease 471
1.2 Strategies for Malaria Control 472

2 *A. gambiae*, the Principal Vector of Malaria 473
2.1 The Mosquito and the Malaria Parasite 473
2.2 *A. gambiae* is a Polymorphic Species 474

3 The *A. gambiae* Genome 475
3.1 The PEST Strain 475
3.2 The Draft Genome Sequence 476
3.3 Genome Assembly, Genome size, and Physical Mapping 477
3.4 Genetic Variation 478
3.5 Gene Content 480
3.6 Comparison of the *Anopheles* and *Drosophila* Proteomes 481
3.7 Genes Regulated by Blood Feeding 483

4 Utility of the *A. gambiae* Genome 484
4.1 Odorant Receptors and Novel Repellents/Attractants 484
4.2 Insecticide Resistance Genes 485
4.3 Insecticide Targets 487
4.4 Immune-response Genes and Genetic Vector Control Strategies 488

Bibliography 489
Books and Reviews 489
Primary Literature 491

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

Keywords

BAC
Bacterial artificial chromosome. A single copy cloning vector for large DNA fragments (on the order of 100 to 200 kbp).

Contig
A segment of contiguous sequence, uninterrupted by gaps, assembled from a number of overlapping sequence reads.

Cytogenetic Form
Subspecies of *Anopheles gambiae*, as classified by paracentric chromosomal inversions.

DDT
The pesticide dichlorodiphenyltrichloroethane.

Mate Pair
The two sequence reads derived from either end of a DNA fragment.

Orthologs
Genes in different species arising from a single ancestral gene (e.g. human α-globin and mouse α-globin).

Paralogs
Genes in the same species, arising from local gene duplication (e.g. human α-globin and human β-globin).

PEST
The pink eye standard laboratory strain of *Anopheles gambiae*.

Read
A DNA sequence, approximately 500 bp in length, that is derived from a template DNA molecule. The terms "read," "trace," "lane," and "electropherogram" are often used interchangeably.

SNP
Single nucleotide polymorphism.

Scaffold
A set of contigs separated by sequence gaps but of defined order and orientation.

Vector
An organism that carries disease-causing pathogens from one host to another.

WHO
World Health Organization.

Shotgun
A method of whole-genome sequencing where high molecular weight DNA from an organism is randomly fragmented and sequence reads are obtained from fragment ends. Sequencing is done at some level of redundancy (coverage) such that each nucleotide in the genome is represented in multiple reads. The consensus genome sequence is reconstructed computationally on the basis of overlaps among sequence reads.

Sporogeny
The stages of the *Plasmodium* development cycle that take place within the mosquito vector.

Mosquito-borne illnesses exact an enormous toll on human health; malaria parasites alone currently infect approximately 500 million people, more than 1 million of whom will die this year. *A. gambiae* is the principal malaria vector in sub-Saharan Africa, where 90% of the world's deaths due to malaria occur. Historically, the greatest successes in prevention of malaria, yellow fever, and other mosquito-borne disease have come from controlling the mosquito vector. Current work on vector control holds much promise and can be addressed more efficiently by genomic approaches. The reference genome sequence of *A. gambiae sensu stricto* (a representative member of the cryptic *A. gambiae* species complex) has yielded novel insecticide targets, has facilitated the identification of genes involved in insecticide resistance, and has led to the identification of genes that underlie the strong preference of this vector for human blood. Germ-line transformation of the mosquito has been achieved and the reference genome sequence will facilitate efforts to develop a tractable genetic control strategy for mosquitoes. The publication of the *A. gambiae* genome is a landmark in medical entomology, both because this was the first vector genome to be fully sequenced and because of the impact of this mosquito on public health.

1 The Mosquito and Infectious Disease

1.1 Scope and Impact of Mosquito-borne Disease

Mosquitoes, which are members of the insect family Culicidae of the order Diptera, the flies, are medically the most important group of arthropod vectors, or transmitters, of human pathogens. Mosquitoes transmit a large number of human viral pathogens, including the viruses that cause dengue, yellow fever, and West Nile encephalitis, as well as several nematodes species that cause

lymphatic filariasis. But, by far, the most important mosquito-borne disease is malaria, which is caused by protozoan parasites of the genus *Plasmodium*. Malaria shares with HIV/AIDS and tuberculosis the distinction of being one of the three most important disease-specific causes of human mortality in the world today. Estimates of the public health impact of malaria are imprecise, but it is widely assumed that approximately half a billion people are currently infected with malaria parasites and that more than 1 million die each year, principally infants and very young children. Although four different *Plasmodium* species infect humans, nearly all of the mortality is caused by the parasite *Plasmodium falciparum*.

1.2 Strategies for Malaria Control

The etiologic agent of human malaria and the pathogen's mode of transmission by mosquitoes were both determined in the latter years of the nineteenth century by Charles Laveran (who discovered the parasite) and Ronald Ross and Battista Grassi (they independently identified the vector), but malaria has been recognized as an important human disease for centuries. By the fifth century, Hippocrates had clearly described the clinical symptoms associated with three different species of malaria parasites, and it was widely recognized in those times that the disease was caused by association with marshy environments with bad air (thus the name malaria).

Antimalarial drugs and interventions targeted at the vector remain the primary public health measures for dealing with malaria. Two very important discoveries emerged during World War II: the antimalarial drug chloroquine, and the insecticide DDT (dichlorodiphenyltrichloroethane). Both were extremely effective, very inexpensive to produce, unencumbered by patents, and had remarkably low levels of toxicity. (The adverse environmental impacts of DDT, which is a highly stable molecule that can accumulate to toxic levels in species at or near the top of the food chain, were entirely due to its widespread use for control of insects of agricultural importance.) In the decades of the 1950s and 1960s, chloroquine and DDT (and related types of antimalarials and insecticides) formed the basis for very effective, countrywide malaria-control programs that were implemented in the framework of a World Health Organization-directed worldwide Malaria Eradication Campaign. Malaria was permanently eradicated from most developed countries in the temperate-climate regions of the world and prevalence was significantly reduced in many other malaria-endemic countries. Unfortunately, worldwide eradication was an impossible goal, and by the 1970s, the responsibility for managing and funding malaria-control programs was returned to national health programs and budgets.

Several important lessons emerged from the WHO (World Health Organization) Malaria Eradication Campaign. Vector control using insecticides like DDT sprayed on the walls of houses was an extremely effective malaria-control strategy. However, it was costly and required well-managed teams of technically skilled people. Moreover, it was effective only in countries with sufficient infrastructures like roads and bridges for vector-control teams to reach the population that was at risk. Control programs based exclusively on antimalarial drugs like chloroquine

were even more costly because of the requirement for active case detection. But as active control programs lapsed, effective and inexpensive antimalarial drugs like chloroquine found their way into malaria-endemic communities through either the national primary health systems or through the marketplace.

For most of sub-Saharan Africa, with the notable exceptions of some of the countries in southern Africa and around some of the major cities, the organized malaria-control programs developed during the WHO Malaria Eradication Campaign were not even attempted, particularly in rural areas. Nonetheless, inexpensive antimalarial drugs like chloroquine became widely available, even in relatively remote rural communities in most of Africa. Unfortunately, chloroquine-resistant *P. falciparum* appeared in coastal Kenya in 1978, and in the subsequent two decades, resistant parasites became the prevalent form almost throughout sub-Saharan Africa. Although the health impact of the publicly available and affordable antimalarials like chloroquine was never assessed, the emergence of parasite resistance to this drug has been coincident with an equally poorly documented perception that malaria-specific mortality in Africa has been increasing. Today, chloroquine is an effective antimalarial only in parts of west Africa, and unfortunately, replacement drugs like Fansidar (sulfadoxine/pyramethamine), mefloquine, and – more recently – artemisinin-based drugs (based on the ingredients in *Artemisia annua*) are either very costly or are themselves provoking the emergence of resistant parasite strains. As resistance to existing antimalarial drugs becomes more entrenched, there is increased need for novel and effective vector-control strategies.

2
A. gambiae, the Principal Vector of Malaria

2.1
The Mosquito and the Malaria Parasite

Malaria parasites are intracellular human parasites that undergo cycles of invasion, growth, and mitotic replication in erythrocytes. The parasites are transmitted among human hosts by *Anopheles* mosquitoes, which ingest parasites when they feed on the blood of an infected human host. The malaria parasites undergo a complex developmental cycle, referred to as the *sporogenic developmental cycle*, in the mosquito vector. A small fraction of the *Plasmodium*-infected erythrocytes are differentiated forms called *gametocytes* that are triggered by hypoxanthine in the mosquito's midgut to develop into male and female gametes. These fuse to form diploid zygotes (the only diploid stage in the *Plasmodium* life cycle), which traverse the mosquito's midgut cells and then undergo a meiotic division and differentiate further into growth stages called the *oocysts*. After approximately one week of growth and mitotic divisions, a single oocyst will produce several thousand specialized, haploid forms called *sporozoites*. On maturation of the oocysts, the sporozoites are released into the mosquito's open circulatory system through which they are transported passively to the salivary glands. These parasite stages invade the salivary gland cells and accumulate in the salivary ducts and acini of these cells, from where they are injected into a host capillary when the mosquito next takes a blood meal. In the human host, the sporozoites are transferred passively through the circulatory system to the liver, where they invade hepatocytes. In these cells, the parasite undergoes approximately a

dozen rounds of mitotic replication that result in the production of several thousand *merozoites*. On completion of this cycle of replication, the infected hepatocytes rupture and release the merozoites into the circulatory system, where they invade erythrocytes and initiate the blood cell cycle of infection that causes disease.

Plasmodium falciparum malaria is widely distributed throughout the widely distributed throughout the world, especially in tropical regions bounded by approximately 30° north and south latitude. Over this broad distribution, parasite prevalence in human populations with endemic *P. falciparum* ranges from low levels of a few percent to peak endemicity levels, approaching 100% in rainy seasons when transmission is maximum. Of the approximately 500 or more *Anopheles* species found worldwide, several dozen are involved in the transmission of malaria parasites. The *Anopheles* vectors are the major determinant of local levels of disease endemicity. The majority of *Anopheles* species are probably physiologically competent to support sporogonic development of the parasite. In order for malaria parasites to remain endemic in a particular human community, the vector population must effectively propagate each infected human case to – on average – more than one new human host. With an effective case reproductive rate of less than 1, the parasite cannot be stably maintained in a human population.

Four important features of mosquito biology determine whether a particular *Anopheles* species will be effective as a malaria vector: (1) It must be a physiologically suitable host. (2) It must also be a moderately abundant species in order to maintain a density sufficient to achieve transmission. (3) The mosquito must also take blood meals from people. Because the vector must both acquire the infection in one blood meal and then take a second meal on another human host after the sporogonic cycle is complete, blood-meal host choice is an important determinant of vectorial effectiveness. (4) Finally, and most important, the age structure of the vector population must be such that a sufficient portion survives long enough to allow sporogonic development to be completed.

2.2 *A. gambiae* is a Polymorphic Species

Approximately 90% of the world's malaria-specific mortality occurs in sub-Saharan Africa, where the primary vector is *A. gambiae*. A second important African vector is the mosquito *Anopheles arabiensis*, which is a close relative of *A. gambiae* in a cluster of seven related African species known as the *A. gambiae sibling-species complex*. This extraordinary concentration of malaria in Africa is largely a consequence of the close ecological association of *A. gambiae* and *A. arabiensis* with people. *A. gambiae*, in particular, takes blood meals almost exclusively from people; its larval stages develop in temporary pools of water, like tire ruts, animal hoof-prints, and irrigated agricultural fields that are produced by human activity; and when not laying eggs or mating, the adult females spend most of their time in human dwellings either blood feeding or resting while eggs develop.

Perhaps, because of the close association this mosquito has with people, it is also a highly polymorphic species. Like a number of other insects, especially in the order Diptera, *A. gambiae* and its sibling species have giant polytene chromosomes in many of their tissues, and these have facilitated population analysis. During the

past three decades, careful studies of the polytene chromosomes of field-collected *A. gambiae* from countries in west Africa have revealed patterns of distribution of a number of paracentric chromosome inversions that suggest the presence of at least five different reproductively isolated chromosome forms designated as Bissau, Bamako, Mopti, Forest, and Savanna. (All *Anopheles* genomes are based on three pairs of chromosomes, two autosomes, and an X and Y sex-determining pair, and in most *Anopheles* species, cells of either the larval salivary glands or nurse cells of developing ovaries contain polytene chromosomes that, when properly prepared, will show a reproducible, species-specific banding pattern. In addition to their value in species identification and population studies, polytene chromosomes are excellent for physical mapping of cloned DNA.) Most of the inversions that distinguish chromosomal forms are restricted to the right arm of chromosome 2. In geographic regions where two or more of the chromosome forms are present, inversion heterozygotes expected from interbreeding of forms are either totally absent or are present at levels significantly below Hardy–Weinberg expectations. Three of these forms, Bamako, Mopti, and Savanna, have been carefully studied in Mali, where they exhibit clear differences in ecology. In spite of considerable effort by several groups to identify molecular markers diagnostic of these chromosome forms, the only molecule found thus far that distinguishes different *A. gambiae* populations is the rDNA, where sequences in both the IGS and ITS regions can be used to identify what are referred to as *S and M ribosomal DNA* (rDNA) *types*. In Mali, the M rDNA type is associated exclusively with Mopti cytogenetic specimens and the S rDNA type is found only in Savanna and Bamako specimens. In other parts of west Africa, however, the association between karyotype and rDNA type is less predictable. Both S and M rDNA types are found in Forest, Bissau, and Savanna cytogenetic forms.

The frequencies of polymorphic paracentric inversions are clearly important characters in *A. gambiae* as well as in its sibling species, *A. arabiensis*. In both species, inversion frequencies have been shown to vary along clines associated with climatic and ecological parameters, and the data from Mali, supporting the presence of chromosomally distinct *A. gambiae* forms that are both ecologically and genetically differentiated is quite robust. Unfortunately, karyotyping of wild specimens is both difficult and possibly biased, and the lack of congruence between the chromosomal inversion and rDNA molecular markers means that the population structure of this important African vector remains unclear.

3 The *A. gambiae* Genome

3.1 The PEST Strain

The *A. gambiae* PEST (Pink Eye STandard) strain was chosen for genome sequencing because it had a fixed chromosomal arrangement, a sex-linked pink-eye mutation that can readily be used as an indicator of cross-colony contamination, and because clones from two different PEST-strain BAC (Bacterial Artificial Chromosome) libraries had already been end sequenced and physically mapped. The pink-eye mutation originated in a colony called *A. gambiae LPE*, established in 1951

at the London School of Hygiene and Tropical Medicine from mosquitoes collected in Lagos, Nigeria. In 1986, this mutation was introduced into a colony of *A. gambiae* from western Kenya by crossing males of the LPE strain with female offspring of wild caught Kenyan *A. gambiae* (the Savanna form), selecting males from the F2 of this cross and then crossing them again with additional female offspring of wild caught Kenyan *A. gambiae*. From the F2 offspring of this second outcross to Kenyan mosquitoes, a strain was selected that was fixed for pink eye. This outcrossing scheme was repeated once more in 1987, producing a pink-eye strain with a genetic composition largely constituted of the western Kenya Savanna cytogenetic form. In each of these crosses, several hundred female offspring of at least 20 wild caught mosquitoes were used in the cross. This strain, designated *A. gambiae PE*, was polymorphic for the inversions 2La (32%) and 2Rbc (19%). The 2Rbc inversion is characteristic of Mopti, indicating that the original LPE strain from Nigeria was the Mopti form. This inversion was apparently balanced by the uninverted form because no 2Rbc/bc individuals were detected in the colony. From this PE strain, Mukabayire and Besansky selected a set of nine families whose female parent and at least 20 female offspring were fixed for the standard chromosome karyotype. The progeny of these nine families were pooled to form the *A. gambiae* PEST strain. This strain may clearly have some Mopti-derived DNA as the standard karyotype is shared by Mopti and Savanna and the original PE strain did have the 2Rbc inversion rather than the 2Rb that is typical of Savanna. When tested, this colony was fully susceptible to *P. falciparum* from western Kenya.

3.2 The Draft Genome Sequence

Obtaining the draft genome sequence of *A. gambiae* was a collaborative effort, with sequence data provided by Celera, The Institute for Genomics Research, and the French National Sequencing Centre, Genoscope. The whole-genome shotgun method was adopted for this project, given its speed and efficiency and in consideration of its previously successful use for the *Drosophila* genome project. Paired end reads were derived from 2 254 546 plasmid clones with 2-, 10-, or 50-kb inserts, (50 kbp plasmid libraries had the bulk of their insert sequence excised during the initial steps of library construction, and only the insert ends were propagated in *Escherichia coli*.) and approximately 40 000 large-insert BAC clones, to give a total of 10.2-fold sequence coverage of the genome, estimated from CoT analysis to be about 260 Mbp. All plasmid and BAC libraries were constructed using DNA extracted from several hundred adult PEST-strain mosquitoes. Libraries from male and female mosquitoes were constructed separately and sequenced in equal proportion, giving approximately equal coverage of the autosomes and accordingly less coverage of the X and Y chromosomes. All electropherograms are downloadable from the NCBI trace archive (http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?) or Ensembl trace server (http://trace.ensembl.org).

The WGS data set was assembled using Celera's assembly software, where algorithms are executed in a stepwise manner. An initial Blastn comparison of each sequence read to every other sequence read is used to detect overlaps (the overlap criteria being 40 or more base pairs with 6% or less mismatch). Reads within each

cluster of overlapping sequences are then assembled into a contiguous consensus sequence (contig). Contigs that appear from read depth to be single copy in the genome are ordered and oriented into longer scaffolds, using mate pairs (two sequence reads derived from either end of a clone insert). Mate pairs are also used to guide computational intrascaffold gap filling, whereby one mate is located in the sequence flanking either side of a gap and its partner is placed within the gap. The resulting assembly is thus a set of typically very long scaffolds that are ordered and oriented sets of contigs whose intervening gaps are of known mean length and whose standard deviations are based on estimated library insert sizes.

The published draft of the *A. gambiae* whole-genome assembly was comprised of 8987 scaffolds, spanning 278 million base pairs (Mbp) of sequence. It is important to note that a small number of very large scaffolds cover most of the genome (91% of the sequence is represented by 303 scaffolds greater than 30 kb), and that many small scaffolds are expected to fit within intrascaffold gaps. In the entire assembly, there are 8986 gaps between scaffolds and 9964 gaps within scaffolds. Gaps within scaffolds are largely due to repetitive sequences that could not be uniquely placed on the basis of sequence overlap or mate-pair information and may also occasionally be due to a simple lack of coverage. Genbank accession numbers for the 8987 genome scaffolds are AAAB01000001 through AAAB01008987 and the scaffold sequences are downloadable as a set from ftp://ftp.ncbi.nih.gov/genbank/genomes/*Anopheles*_gambiae/Assembly_scaffolds.

While use of the whole-genome shotgun approach for this project was well justified by time and cost savings, it is important to understand that the present form of the assembly is that of a draft genome sequence and, as with all draft genomes, there are numerous gaps that will require focused finishing efforts. Further, because of the fact that the plasmid libraries were derived from whole adult mosquitoes, some contaminating sequence from commensal gut bacteria was expected and were in fact observed. Scaffolds with hits to bacteria were flagged in their Genbank record as containing putative contaminating sequence of possible bacterial origin. In total, 213 short scaffolds were flagged, which together comprised 36.5 kb or 0.013% of the genome. These scaffolds were not culled from the assembly because of the remote possibility that some may represent real horizontal transfer events.

3.3 Genome Assembly, Genome size, and Physical Mapping

The sum of all the assembled *Anopheles* genome scaffolds is 278 Mb, which is significantly larger than the genome size of 260 Mb predicted by CoT analysis. There are several reasons why the assembly is close to, but slightly larger than the genome size predicted by CoT analysis. First, many small scaffolds may fit into gaps within other larger scaffolds. Second, because the X chromosome accounts for close to 20% of the genome but is underrepresented in males relative to the autosomes, CoT analysis using a mixed-sex pool of mosquitoes is expected to underestimate genome size. The third likely cause of discrepancy is the unanticipated assembly of highly variant alleles into separate contigs, as discussed further below (Sect. 3.4 Genetic variation).

The final step in defining the reference genome for *A. gambiae* was to order and

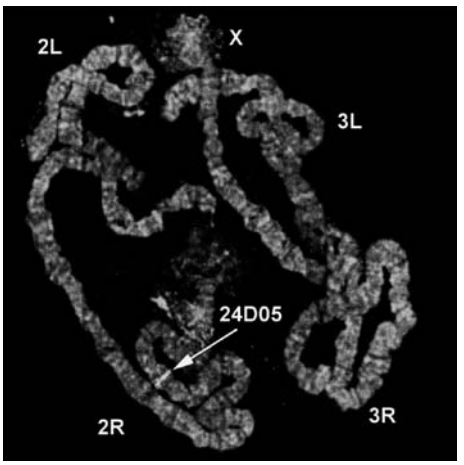

Fig. 1 Illustration of fluorescent *in situ* hybridization using a BAC clone probe. The *A. gambiae* polytene chromosome complement was isolated from the ovarian nurse cells of half-gravid mosquitoes. The probe, 24D08, is a BAC clone from the PEST library and is labeled with Cy3-dUTP red; the polytene chromosome is stained with Yoyo-1 iodide green (Molecular Probes).

orient the scaffolds on the three chromosome pairs, using physical mapping data. The chromosomal locations of assembled scaffolds were determined with sequence-tagged genomic DNA BAC clones that were mapped by *in situ* hybridization to ovarian nurse-cell polytene chromosomes (Fig. 1). For the initial assembly, approximately 2000 BAC clones were mapped and these data enabled the chromosomal location and orientation of scaffolds, constituting about 227 Mbp. Photo archive data showing the results of physical mapping are viewable at http://www.anobase.org. Since the initial assembly, the order and orientation of some of the smaller scaffolds have been confirmed or corrected, and ~7 Mbp more scaffolded sequence has been placed and oriented on the physical genome by *in situ* mapping of selected BAC and cDNA clones. Currently, 172 of the largest scaffolds have been assigned to chromosomal locations (a total of ~233 Mbp), and gaps between almost one-third of these scaffolds have been crossed by BAC clones or short BAC tiling paths (these BACs remain to be sequenced). Several dozen scaffolds (<10 Mbp) can be assigned only to nonspecific centromeric heterochromatic regions and about 1 Mbp of scaffolded sequence has been identified as possible Y-chromosome sequence; thus, about 243 Mbp of the 278 Mbp assembly has been assigned to locations with some level of confidence. Most of the remaining ~8700 unassigned scaffolds (~35 Mbp) are less than 20 kb. The majority of these appear to be repetitive heterochromatic sequences that, in *A. gambiae,* are clustered around the centromeres, the telomeres, and in a few islands in the euchromatic chromosome arms. Most of these scaffolds do not appear to contain genes and many are unlikely to be assigned to specific chromosomal locations by currently available approaches.

3.4 Genetic Variation

Analysis of high-quality mismatches within contig multiple sequence alignments revealed approximately half-a-million single nucleotide polymorphisms (SNPs) in the *A. gambiae* PEST-strain genome. The distribution of SNPs along the chromosomes is bimodal, with some regions showing roughly one SNP every 10 kb and others showing approximately one SNP every 200 bp. Given that the genome of the PEST strain appears to have resulted from a complex introgression of divergent Mopti and Savanna chromosomal forms, we expect that some genomic

regions may be derived only from one or the other form, yielding a low density of SNPs, while other genomic regions may continue to segregate both divergent forms, yielding a high SNP density. Interestingly, the X chromosome has a markedly lower average level of polymorphism and does not show this bimodal pattern, perhaps due to male hemizygosity and hence, a lower rate of introgression is seen on this chromosome. In addition, heterozygosity of the X is expected to be depressed because of the selection for homozygosity of the X-linked pink-eye mutation.

As expected, the overwhelming majority of SNPs lie in intergenic regions, but there is still an abundance of SNPs within functional genes. Introns and intergenic regions have virtually identical heterozygosities, but the silent coding positions appear to have more than twofold enrichment of variability. Generally, silent coding sites are considered as having more stringent constraints than introns or intergenic regions because of biased codon usage, and this is reflected in a lower diversity of silent sites in most organisms. The reason for elevated silent variation in *A. gambiae* is at present unknown. Nucleotides with strong functional constraints, such as splice donors, splice acceptors, and stop codons have the lowest heterozygosity, and nonsynonymous (missense) positions are also evidently low in heterozygosity.

The bimodal SNP distribution in the PEST strain, and hence, the existence of two approximately equally frequent haplotypes in a substantial fraction of the genome, presented a unique challenge for assembly of the sequence data set, and it is important to note that some assembly artifacts have resulted. Where the sequence variation between the two haplotypes is small, they have been assembled together into a single contig, with the most frequent nucleotide at each position contributing to the consensus sequence. This is the manner in which all current whole-genome assembly algorithms handle genetic variation, even though the resulting consensus sequence does not precisely represent any native haploid genome. In the case of *A. gambiae*, where the haplotypes are more divergent, they have occasionally been split into two separate contigs. This is problematic because the assembler only allows a single consensus sequence to represent any given position in the genome. As a result, in the draft assembly, there are situations in which the second redundant contig (1) has been placed in the nearest gap, causing an artificial tandem duplication within the scaffold; (2) has been left out of the scaffold as singleton; or (3) where the problem is at a scaffold end, the duplicate has been placed at the beginning of the neighboring scaffold. The phenomenon described here was initially revealed as a larger separation of mate pairs in parts of the assembly than what the library insert sizes had predicted. Comprehensive screening of the 1 644 078 total mate pairs in the assembly revealed a total of 726 regions, with this diagnostic mate-pair signature covering a total of approximately 7.7% of the assembly. The amount of sequence that is doubly represented in the genome assembly because of this artifact is estimated at about 23.5 Mbp.

While the bimodal pattern of SNP variation is a unique feature of the PEST strain and is not representative of any wild mosquito population, individual variants represent real polymorphisms in *A. gambiae* and will provide valuable marker sets for genetic studies. The approximately 500 000 PEST-strain SNPs are downloadable from

ftp://ftp.ncbi.nih.gov/genbank/genomes/Anopheles_gambiae or www.ensembl.org/anopheles_gambiae. A further 75 000 SNPs have been identified by comparing light sequence coverage (1.2×) from a lab strain of the Mopti cytogenetic form to the assembled PEST-strain sequence, and these SNP data are also available at www.ensembl.org/anopheles_gambiae.

Understanding the genetic diversity of anopheline mosquitoes and the molecular basis of cytogenetic form–associated traits will be accelerated by genomics and will likely point the way to more efficient strategies of vector-targeted malaria control. Efforts are presently underway to analyze large regions of the *A. gambiae* Y chromosome as population genetic markers and as indicators of the evolutionary relationships of the different *A. gambiae* forms and other species closely related to *A. gambiae*. Efforts to use other parts of the genome to define and identify such species and forms have been complicated by presumed interspecific and interform recombination over many regions of the different genomes. As a presumably nonrecombining part of the genome, the Y-chromosome sequence will be an invaluable source of material for teasing apart population structure and for developing rapid diagnostics for field studies. Studies are also underway to obtain wild isolates of *A. gambiae* from several different regions of Africa and sequence segments of DNA at regular intervals across the genomes of multiple individuals from each isolate. This survey of genetic diversity among different isolates of *A. gambiae* will facilitate the teasing apart of anopheline mosquito population structure and will play a crucial role in understanding mosquito biology, particularly those biological traits that influence vectorial capacity.

3.5 Gene Content

The initial automated annotation of the draft *A. gambiae* genome sequence was accomplished using established annotation pipelines at Celera and the Ensembl group at the European Bioinformatics Institute/Sanger Institute. Both pipelines use a combination of homology-based evidence and *ab initio* gene-finding algorithms to model gene structure. At the inception of the project, there were only approximately 6000 EST sequences available in public databases. However, given the evolutionary distance between *Drosophila* and *Anopheles* (~250 million years), obtaining wide representation from *Anopheles* EST libraries was essential for identifying the open reading frames of *A. gambiae* genes, and an additional approximately 80 000 EST sequences were derived from whole mosquitoes, as an integral part of the *Anopheles* genome project.

For proteome analysis, a nonredundant set of predicted protein sequences from the two pipelines (Celera and Ensembl) was constructed by selecting, for each locus, the annotation containing the largest number of exons. A total of 15 189 automated predictions were derived in this manner and, after removal of putative transposable elements and other contaminants, served as the set for initial analysis of the *A. gambiae* proteome. As is typical with automated genome annotations, it is likely that some false-positive predictions (pseudogenes, bacterial contaminants, and transposons) and false-negative predictions (*Anopheles* genes that were not computationally predicted) remain, and there are undoubtedly numerous errors in defining the precise boundaries of these putative gene structures. Automated annotation, while being a cost-effective method for an initial

high-level view of the genome, does not substitute for the careful manual curation that will be undertaken by the community over a period of time. Since the initial publication of the *Anopheles* genome, the annotation has been under continuous development by the Ensembl group in coordination with the International *A. gambiae* Genome Consortium. The present version of the annotated genome is available at www.ensembl.org/*Anopheles*_gambiae. A tutorial on the Ensembl Anopheles site is available at www.ensembl.org/*Anopheles*-gambiae/documents/mosquito_doc.pdf.

3.6 Comparison of the *Anopheles* and *Drosophila* Proteomes

Analysis of the annotated gene set has revealed many interesting features of the proteome of *A. gambiae*, particularly in comparison with *Drosophila melanogaster*, the closest relative of this vector (*A. gambiae* and *D. melanogaster* shared a common Dipteran ancestor approximately 250 million years ago.) to have a completely sequenced genome. In total, 47.2% of the *Anopheles* protein set is composed of 1 : 1 orthologs (defined as *reciprocal best Blast matches*), and 13.8% is composed of "many-to-many" orthologs, where paralogous gene families have expanded/contracted unevenly. A further 17.9% of the *Anopheles* protein set is represented by proteins that have a homologous relationship to *Drosophila* proteins, but the match cannot be easily defined as orthologous or paralogous. For example, proteins in this group might share one or more protein domains or might be divergent members of large and diverse protein families. 10.0% of *Anopheles* proteins had their best match to a noninsect species and 11.1% of *Anopheles* proteins do not show homology to proteins from any other species. Of these 1437 *Anopheles*-specific genes, 575 are supported by EST evidence and 522 show homology to other *A. gambiae* proteins. While some of these proteins may represent *de novo Anopheles* genes or rapidly evolving ancestral genes that are now mutated beyond recognition, it is likely that homologous sequences for these proteins will be discovered as the genomes of organisms with closer evolutionary relationships to *A. gambiae* are sequenced. Interestingly, in comparing the *A. gambiae* and *D. melanogaster* proteomes, sequence similarity is highest among structural molecules, transporters, and enzymes and is lowest among genes involved in defense and immunity. This is consistent with the results from other recent whole-genome projects that suggest that the most plastic gene families are those that mediate the direct interaction of the organism with its environment, such as genes involved in olfaction, detoxification, and immune response. In comparison, developmental genes are highly conserved between *D. melanogaster* and *A. gambiae*, with 85% having single 1 : 1 orthologs (best reciprocal matches) between the two species.

Given the evolutionary distance between *A. gambiae* and *D. melanogaster* conserved gene order (synteny) is recognizable only over short distances. Approximately 34% of *A. gambiae* genes colocalize in small microsyntenic clusters, defined as chromosomal regions where at least two orthologs or orthologous groups are separated by no more than five unrelated genes in the intervening DNA. In total, there are 948 such microsyntenic blocks in the *A. gambiae* genome, when compared to *D. melanogaster*, the largest block containing 31 orthologous genes.

Most clusters contain substantially fewer genes and show evidence of local duplications, inversions, and translocations. In contrast, homology between *A. gambiae* and *D. melanogaster* on the level of whole-chromosome arms is easily recognizable upon mapping of 1 : 1 orthologs. The most conserved pair of chromosomal arms are *D. melanogaster*2L (Dm2L) and *A. gambiae*3R (Ag3R), with 67% of Ag3R orthologs and 83% of microsyntenic clusters being significantly similar to Dm2L. *A. gambiae* chromosomal arm 2L appears to be largely homologous to Dm3L and Dm2R, with approximately 42% of *A. gambiae* orthologs found on each of these two *D. melanogaster* arms. Similarly, Ag2R shares homology with Dm3R. Ortholog content of Ag3L indicates that this chromosomal arm has modest homology to both Dm2R (30%) and Dm3L (22%). Homology between the X chromosomes of the two species is limited, with significant shuffling of chromosomal segments to and from autosomes.

The predicted *Anopheles* protein set has been classified on the basis of protein domains, and their functional categories, using InterPro and the Gene Ontology (GO). The relative abundance of the majority of proteins containing InterPro domains is similar between the mosquito and fly. As expected, insect-specific cuticle and chitin-binding domains and the insect-specific olfactory receptors are overrepresented relative to noninsects. However, comparing *A. gambiae* and *D. melanogaster,* only 6 of the 200 most abundant protein domains differ significantly in abundance. However, some of these differences are dramatic, such as the striking expansion of fibrinogen domain–containing proteins observed in *A. gambiae* relative to *D. melanogaster*. There are 58 *A. gambiae* and 13 *D. melanogaster* fibrinogen domain–containing proteins and only a single pair is a 1 : 1 ortholog. It is tempting to speculate that fibrinogen domains, which were originally found in proteins that mediate blood coagulation in mammals, could be conferring some function to mosquitoes that enables blood feeding, such as inhibition of blood-meal coagulation. Further, the serine proteases (central effectors of proteolytic processes) are well represented in both insect genomes, but *Anopheles* has nearly 100 additional members, perhaps again reflecting the hematophagous adaptation of this Dipteran.

Major differences in gene number are seen in the peroxidase system, which mediates the oxidation of diverse substrates. A total of 18 peroxidases were observed in *A. gambiae* and only 10 were observed in *D. melanogaster*. The *Anopheles* peroxidases have high homology to the salivary peroxidases of the mosquito *A. albimanus,* suggesting the possibility that this gene family has expanded to facilitate the blood-feeding process.

In a separate comparative analysis using the LEK algorithm to identify protein families with different representation in *A. gambiae* and *D. melanogaster*, a prominent expansion of a 19-member family of odorant receptors (ORs) was detected in *A. gambiae*. The members of this family do not show substantial sequence similarity to ORs from *D. melanogaster* or any other organism. While between *A. gambiae* and *D. melanogaster*, there is significant conservation of the G protein-coupled receptor superfamily as a whole, this *A. gambiae* OR subfamily appears to be unique and may play an important role in mosquito-specific behavior that includes host seeking, as discussed further in Sect. 4.1. In addition to the conspicuous protein family expansions discussed above, many additional

expansions were seen in diverse families. These include critical components of the visual system, structural components of the cell adhesion and contractile machinery, energy-generating glycolytic enzymes, anabolic and catabolic enzymes involved in protein and lipid metabolism, transporters, and detoxification enzymes.

Identification of gene family contractions have not been a focus of analysis of the *A. gambiae* genome, even though gene loss is considered to be an important mediator of evolution. It is ambiguous when doing a binary comparison whether there is expansion in one species or contraction in the other, relative to the state of the common ancestor. This can only be revealed by analysis of orthology across several species. Further, it is difficult to distinguish real gene losses from simple false-negative annotation results. Given that the overall gene count in *A. gambiae* and *D. melanogaster* is not significantly different, with each species having roughly 14 000 annotated gene loci, the overall number of expansions must be balanced by an equal number of reductions in each species, as they evolve to fill their ecological niches.

3.7
Genes Regulated by Blood Feeding

Newly emerged female mosquitoes are sustained by sugar meals as they seek the blood meal that is necessary to support egg development. Active host seeking is facilitated by olfactory, temperature, and visual cues. After locating a host and extracting several times their weight in blood, the female mosquito appears to undergo profound metabolic reprogramming, commensurate with the task of processing the massive blood bolus.

The *A. gambiae* genome project included an EST-sequencing program utilizing two source libraries. One library was constructed using DNA from whole adult female mosquitoes before blood feeding, when in a relatively quiescent metabolic state, and the second library was constructed 24 h after a blood meal, when the mosquitoes were under the metabolic burden of converting the blood meal into eggs. This approach facilitated identification of genes activated or inactivated by blood feeding in a hematophagous insect. Approximately 40 000 ESTs were sequenced from each library and assembled into approximately 7000 consensus sequences, each representing a distinct gene locus. 435 genes showed significantly altered transcription following blood-meal ingestion, as determined by Chi-squared analysis. Analysis of gene expression pre- and post blood feeding has revealed both expected and unexpected changes (Table 1). As expected, there is dramatic upregulation (up to 144 fold) of transcripts encoding serine proteases and other proteases, transcripts involved in transcriptional control and nuclear regulation, and transcripts representing the protein synthesis machinery that are involved in converting the digested blood proteins into new functional mosquito protein products. Further, transcripts involved in gluconeogenesis, lipid metabolism, and purine metabolism are upregulated, as are a small number of transcripts, such as ubiquitin-conjugating enzymes and cytosolic aminopeptidases, whose protein products mediate intracellular degradation of protein. Finally, there is upregulation of oogenesis-associated proteins and lipids 24 h post blood meal, including several vitellogenins (yolk constituents) and members of the phenoloxidase pathway that mediates eggshell melanization. In terms

of decreased expression, genes involved in sugar metabolism (for example, salivary and midgut maltases, glycolysis, and oxidative phosphorylation) are downregulated, which is not unexpected, given the presumed decreased reliance on sugar metabolism with the high intake of a protein-rich blood meal. More intriguing is the marked downregulation post blood meal of striated muscle proteins, including actin, myosin, tafazzin, and troponin, as well as ATP-dependent cation transporters and Ca^{2+} binding proteins involved in motor signaling. Taken together with the observed decrease in transcription of opsins, photoreceptor associated proteins, cuticular proteins associated with the eye lens, and pheromone-binding proteins, the post blood-meal expression pattern suggests that after a blood meal, mosquitoes become detached from their environment and sacrifice mobility in order to devote metabolic attention to the monumental task of blood-meal processing and egg development. The set of genes upregulated by blood-meal digestion presents a rich source of novel insecticide targets and will allow identification of promoters that can activate transgenes in a blood-meal responsive manner.

4 Utility of the *A. gambiae* Genome

4.1 Odorant Receptors and Novel Repellents/Attractants

The distinct preference of *A. gambiae* for human blood meals contributes substantially to the menace of this vector. Very

Tab. 1 Representative messages significantly more transcribed or less transcribed in female *A. gambiae* mosquitoes 24 h after human blood-meal ingestion.

Category	↕	*Gene description*	*Gene symbol*	*EST count (no blood)*	*EST count (blood)*
Extracellular protein digestion	↑	Trypsin-2 precursor	AgCP10888	1	144
Intracellular protein degradation	↑	Ubiquitin	AgCP12328	0	47
Nuclear regulation	↑	Histone	AgCP10826	2	40
Protein synthesis	↑	Elongation factor 1a	AgCP3905	163	278
Oogenesis	↑	Vitellogenin	AgCP2518	0	878
Gluconeogenesis	↑	Aspartate amino transferase	AgCP8491	1	10
Glycolysis	↓	Succinate dehydrogenase	AgCP3730	24	7
Oxidative phosphorylation	↓	Glycerol-3-phosphate dehydrogenase	AgCP3306	51	11
Extracellular sugar digestion	↓	Salivary maltase	AgCP12790	10	0
Striated muscle	↓	Flightin (striatal muscle protein 27)	AgCP3724	71	2
Vision	↓	Opsin	AgCP12420	276	184
Chemosensation	↓	Pheromone-binding protein	AgCP11481	7	0

little is known about the molecular basis of this selective behavior, although it almost certainly involves specific human odors and mosquito pathways for detecting and responding to these odors. Continuing to explore the diversity of ORs and odorant binding proteins encoded in the *A. gambiae* genome will lead to a better understanding of the mechanisms of host preference and will uncover potential targets for a new generation of specific attractants/repellents. Preliminary analysis of the draft genome sequence identified a total of 276 G protein-coupled receptors, including 79 candidate ORs. The majority of these ORs (64 of 79) show expression only in olfactory tissues, as determined by RT-PCR experiments. Comparison to *D. melanogaster* revealed a similar total number of ORs, but only a single unequivocal orthologous pair of ORs between the two species. The extensive lineage-specific expansion of ORs likely reflects the ecological and physiological relevance of these receptors and the importance of each species being able to detect relevant chemicals – for example, rotting fruit odors for *D. melanogaster* and human host odors for *Anopheles*. AgOr1, a putative *A. gambiae* OR, has recently been identified as a possible mediator of human host finding. Initial clues that AgOr1 may be important in host finding include its expression exclusively in the olfactory tissue of female mosquitoes, and the downregulation of its expression in these tissues after a blood meal. Hallem EA et al. provided further evidence by engineering *D. melanogaster* neurons lacking endogenous ORs and by transfecting these with AgOr1. Electrophysiological measurements of the transfected cells showed that a component of human sweat, 4-methylphenol, elicited a strong response.

4.2 Insecticide Resistance Genes

During the 1950s and early 1960s, the WHO malaria eradication campaign succeeded in eradicating malaria from Europe and dramatically reduced its prevalence in many other parts of the world, primarily through programs that combined mosquito control agents like DDT with antimalarial drugs like chloroquine. However, malaria and anopheline mosquitoes remain entrenched in sub-Saharan Africa and the appearance of chloroquine-resistant parasites and insecticide-resistant mosquitoes is contributing to the current rise of malaria in Africa. Even control programs based on insecticide-impregnated bed nets, which are now advocated by WHO and are being widely implemented in Africa are threatened by the development of resistance to pyrethroids, the insecticide class of choice for this application. Insecticide resistance is often metabolic, where the insecticide will induce the expression of an enzyme that degrades it. Metabolic resistance is mediated largely by one of the three classes of enzymes, depending on the particular insecticidal agent in question. These are carboxylesterases (CEs), glutathione-*S*-transferases (GSTs) and cytochrome P450s. Ranson et al. cataloged members of these three major insecticide resistance–related families in the *A. gambiae* genome. Comparison to *D. melanogaster* revealed that a considerable expansion of these families has occurred in the mosquito. For example, there are 51 versus 31 CEs and 111 versus 90 cytochrome P450 enzymes in *A. gambiae* and *D. melanogaster* respectively. Secure orthologs between *D. melanogaster* and *A. gambiae*, identified by rigorous analysis of phylogenetic trees comprise less than

15% of these supergene families, indicating that gene families have radiated independently from common ancestral genes, presumably driven by the different metabolic requirements of the two species for natural compounds. Over the past decade, several resistance loci for different agents have been mapped in the mosquito genome, using classical genetic methods, and inspection of gene content within the boundaries of mapped resistance loci has been revealing. For example, it has been observed that several large clusters of cytochrome P450 genes colocalize with a known pyrethroid resistance locus involved in oxidative metabolism of the insecticide. Further, a major DDT resistance locus colocalizes with a cluster of eight epsilon (insect-specific) GST genes on chromosome 3R. Biochemical characterization of the members of this GST cluster has shown that only one of these eight enzymes, GSTE2-2, is able to metabolize DDT. Western blots using antibodies raised against this GST indicated that its expression is elevated in a DDT-resistant strain (ZAN/U) of *A. gambiae*, relative to an insecticide-sensitive strain (Kisumu). Subsequent and more comprehensive quantitative PCR studies in these two mosquito strains have indicated that not just GSTE2-2, but in total, five of the eight GSTs genes in this cluster are significantly overexpressed in the DDT-resistant strain. Gene dosage analysis showed no evidence for gene amplification, suggesting perhaps that coexpression of genes in the epsilon cluster is controlled by a common regulatory element.

In addition to metabolic resistance, insensitivity to an insecticide can arise through mutations that disrupt binding of the insecticide molecule to its target protein. A well-characterized example is the insensitivity to two major insecticide families, organophosphates and carbamates, that arises through mutation in their target protein Acetylcholinesterase (AChE). Only one AChE gene is present in *D. melanogaster* (called ace-2) and it is well established that resistance in *D. melanogaster* is caused by mutations at this locus. Two AChE genes, ace-1 and ace-2, have been identified in *A. gambiae*, through sequence searches of the reference genome, using human and *D. melanogaster* AChEs as queries. The *A. gambiae* ace-1 gene shows 52% amino acid identity and the *A. gambiae* ace-2 gene shows 83% amino acid identity to the single *D. melanogaster* AChE, and both *Anopheles* genes contain the conserved FGESAG active site motif. Phylogenetic analysis suggests that the presence of two AChE genes is the ancestral state, and one of these genes has been lost in *D. melanogaster*. It is well established that mutation of the target site of the single *D. melanogaster* AChE gene causes resistance to organophosphates and carbamates in this species, Interestingly, however, in organophosphate and carbamate resistant *Culex pipiens* and *A. gambiae* strains, insensitivity results from a substitution (G119S) near the catalytic site of the ace-1 gene, not a mutation of the ace-2 gene, which is presumably the direct *D. melanogaster* ortholog.

Understanding the common mechanisms of insecticide resistance in *A. gambiae* is of practical importance. The possibility now exists of screening new compounds for tolerance liability by determining whether they cause mutation in known hotspots in genes previously shown to confer resistance or induction of genes with a detoxifying function. Further, specific mutations in resistance-conferring genes can be used as markers to monitor the spread of resistance to insecticides

in use in Africa, and can provide important information to help guide ongoing chemical-based vector-control programs.

4.3 Insecticide Targets

While most chemical agents used in the battle against mosquito-borne disease are broad-spectrum agents developed for the purpose of agricultural pest control, the possibility now exists of finding and exploiting physiological and biochemical systems that are unique to the mosquito, in order to produce new and highly selective agents for mosquito control. The reference mosquito genome sequence has helped reveal a variety of systems that may be good candidates for the development of more selective agents for mosquito control. For example, regulatory peptides act as neurochemicals and hormones that guide mosquito reproduction and development. A total of 35 genes encoding putative regulatory peptides have been annotated in the *A. gambiae* genome and agents that mimic or block the action of these peptides may be effective insecticides. Ecdysteroid peptide hormones, which govern gene expression during female reproduction and larval development are of particular interest. Ovary ecdysteroids released after a blood meal stimulate the fat body to begin secreting yolk proteins. The presence of an ortholog of *A. gambiae* ovary ecdysteroid hormone in the mosquito *Aedes aegypti* but the absence of an ortholog in *D. melanogaster* suggests that there is a unique role for this peptide in hematophagous insects and, thus, an opportunity for development of specific antimosquito agents. Similarly, diuresis is a process that has a critical function in the mosquito that may be exploited. After blood feeding, female mosquitoes engage in rapid diuresis to decrease the volume of the blood meal and allow flight. Genes for all four diuretic hormones known to regulate fluid secretion in insects have been identified in the *A. gambiae* genome and present additional targets for incapacitating the mosquito. Also of interest, the receptor for FRMF amide, a cardioexcitatory tetrapeptide first identified in clam, has been identified in the mosquito genome. In functional studies, the tetrapeptide significantly increased the frequency of spontaneous contractions of the heart in mosquito larvae, suggesting the possibility of developing agonists for this receptor as a larvicide.

Mining of reference genome sequences has revealed that enzyme glutathione reductase is absent from Diptera and is functionally substituted by the thioredoxin system. Thioredoxins are small ubiquitous thiol proteins that efficiently cleave disulfide bonds in a number of other proteins and take part in redox control of numerous cellular processes such as protein folding, signaling, and transcription. The key enzyme in this system is thioredoxin reductase, a single copy gene located on the *A. gambiae* X chromosome and shares 52% sequence identity with its human ortholog. The sequence of the catalytic redox center is different between the *A. gambiae* (Thr-Cys-Cys-SerOH) and human (Gly-Cys-selenocysteine-GlyOH) proteins and this difference provides an attractive avenue for the development of a selective inhibitor.

Differences between mammalian and insect metabotropic glutamate receptors have also been revealed by whole-genome comparison. The *D. melanogaster* metabotropic glutamate receptor (mGluR) shares a very similar pharmacological profile with its mammalian orthologs, being

activated or inhibited by the same natural or synthetic ligands. However, a new receptor in mGluR subclass has been found in the mosquito, the honeybee *Apis mellifera*, and *D. melanogaster*, called AmXR, HBmXR, and *Dm*XR respectively. No direct orthologs of these novel insect receptors are detectable in *C. elegans* or in human genome. The native ligand for these receptors is unknown, although it has been shown that extract from *D. melanogaster* and *A. gambiae*, but not *C. elegans* or mouse brain, caused dose-dependent receptor activation. Sensitivity to formaldehyde (which masks amino groups) and insensitivity to HCl (which disrupts peptide bonds) suggests that the native ligand is an amino acid–like molecule, but the native ligand is clearly not glutamate, given that residues contacting γ-carboxyl group of glutamate are not conserved with the mGluR, and heterologous expression showed no stimulation by glutamate or mGluR agonists AMPA, kainate, and NMDA quisqualate. Because the XR receptors appear to be insect specific, compounds that disrupt their function may, if lethal, form a promising new class of specific insecticide.

4.4 Immune-response Genes and Genetic Vector Control Strategies

To successfully complete its complex life cycle, the malaria parasite must evade both the human and mosquito immune systems. Parasites ingested with the blood meal develop within the mosquito gut into ookinetes, which subsequently burrow into the gut epithelium and form oocysts. When the oocysts burst, sporozoites are released and these travel to the salivary glands to await transfer to a new host. However, large losses in parasite number occur during this invasion because of mosquito defenses such as melanotic parasite encapsulation. These defenses do not comprise an adaptive immune system, such as that found in mammals, but rather belong to the ancient innate immune system that most metazoans rely on for dealing with invading microorganisms. Facilitating these innate mosquito defenses offers a new and promising strategy for malaria control.

In the past decade, genetic selection in the laboratory has yielded several interesting *A. gambiae* strains, including some that are completely refractory to *Plasmodium*. These strains have been the subject of intensive study in a number of different laboratories, and the *A. gambiae* genome is facilitating the identification of candidate genes for *Plasmodium* resistance. More recently, Christophides et al. screened the *A. gambiae* genome for genes implicated in the innate immune response of the mosquito and identified 242 players from 18 different gene families including, for example, pattern-recognition receptors that bind pathogen-specific molecules, second messengers that can amplify or dampen immune signals, and effector molecules such as the prophenoloxidases that mediate melanization. There is marked diversification of these genes relative to their *Drosophila* homologs, likely reflecting the adaptation of *Anopheles and Drosophila* to different immune challenges represented by different ecological and physiological conditions. Recent functional screening (gene silencing using dsRNA) of a subset of these candidate immune-response genes in *A. gambiae* has yielded compelling findings for several pattern-recognition receptors. Functional knockout of the *A. gambiae* leucine-rich repeat protein (LRIM1) led to a substantial (3.6-fold)

increase in oocyst numbers, indicating that this gene has a strong protective effect against the invading parasite. Similarly, functional knockout of the pattern-recognition receptor thioester-containing protein 1 (TEP1) in *A. gambiae* led to a fivefold increase in the number of oocysts developing on the midgut epithelium and a complete abolishment of parasite melanization. Conversely, functional knockout of two different *A. gambiae* C-type lectines (CTL4 and CTLMA2) resulted in a paradoxical enhancement of immune response, with massive melanization of invading ookinetes. It is not clear how these mosquito CTL genes act to protect the developing malaria parasite within the mosquito and no obvious homologs of these genes have been detected in any other organism. LRIM1, TEP1, CTL4, and CTLMA2 are the first mosquito genes to be identified as having an important role in the immune response of the mosquito to the invading malaria parasite. While it is unlikely that gene silencing using dsRNA will be practical as a malaria-control measure, these exciting observations are nonetheless the first steps along the road to the development of a genetic-based transmission-blocking strategy for malaria. Germ-line transformation of the mosquitoes *A. stephensi* and *A. gambiae* has recently been achieved, bringing the possibility of malaria control through the introduction of transgenic mosquitoes another step closer to reality. While genetic control strategies for malaria vectors are complicated by several factors, including the diversity of naturally occurring cytogenetic forms that are reproductively isolated, the selective pressure applied to the *Plasmodium* population, plus all of the usual risks associated with the release of transgenic insects, there is much promise that transgenics will play a key role in malaria control in the coming decades.

See also Genetics, Molecular Basis of; Immunoassays.

Bibliography

Books and Reviews

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J.,

Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M., Venter, J.C. (2000) The genome sequence of *D. melanogaster*, *Science* **287**, 2185–2195.

Besansky, N.J., Powell, J.R. (1992) Reassociation kinetics of *Anopheles gambiae* (Diptera: Culicidae) DNA, *J. Med. Entomol.* **29**, 125–128.

Catteruccia, F., Nolan, T., Loukeris, T.G., Blass, C., Savakis, C., Kafatos, F.C., Crisanti, A. (2000) Stable germline transformation of the malaria mosquito *Anopheles* stephensi, *Nature* **405**, 959–962.

Christophides, G.K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P.T., Collins, F.H., Danielli, A., Dimopoulos, G., Hetru, C., Hoa, N.T., Hoffmann, J.A., Kanzok, S.M., Letunic, I., Levashina, E.A., Loukeris, T.G., Lycett, G., Meister, S., Michel, K., Moita, L.F., Muller, H.M., Osta, M.A., Paskewitz, S.M., Reichhart, J.M., Rzhetsky, A., Troxler, L., Vernick, K.D., Vlachou, D., Volz, J., von Mering, C., Xu, J., Zheng, L., Bork, P., Kafatos, F.C. (2002) Immunity-related genes and gene families in Anopheles gambiae, *Science* **298**, 159–165.

Coluzzi, M., Sabatini, A., Petrarca, V., Di Deco, M.A. (1979) Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex, *Trans. R. Soc. Trop. Med. Hyg.* **73**, 483–497.

Hill, C.A., Fox, A.N., Pitts, R.J., Kent, L.B., Tan, P.L., Chrystal, M.A., Cravchik, A., Collins, F.H., Robertson, H.M., Zwiebel, L.J. (2002) G protein-coupled receptors in *Anopheles gambiae*, *Science* **298**, 176–178.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, J.M., Wides, R., Salzberg, S.L., Loftus, B., Yandell, M., Majoros, W.H., Rusch, D.B., Lai, Z., Kraft, C.L., Abril, J.F., Anthouard, V., Arensburger, P., Atkinson, P.W., Baden, H., de Berardinis, V., Baldwin, D., Benes, V., Biedler, J., Blass, C., Bolanos, R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chaturverdi, K., Christophides, G.K., Chrystal, M.A., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C.A., Flanigan, M., Grundschober-Freimoser, A., Friedli, L., Gu, Z., Guan, P., Guigo, R., Hillenmeyer, M.E., Hladun, S.L., Hogan, J.R., Hong, Y.S., Hoover, J., Jaillon, O., Ke, Z., Kodira, C., Kokoza, E., Koutsos, A., Letunic, I., Levitsky, A., Liang, Y., Lin, J.J., Lobo, N.F., Lopez, J.R., Malek, J.A., McIntosh, T.C., Meister, S., Miller, J., Mobarry, C., Mongin, E., Murphy, S.D., O'Brochta, D.A., Pfannkoch, C., Qi, R., Regier, M.A., Remington, K., Shao, H., Sharakhova, M.V., Sitter, C.D., Shetty, J., Smith, T.J., Strong, R., Sun, J., Thomasova, D., Ton, L.Q., Topalis, P., Tu, Z., Unger, M.F., Walenz, B., Wang, A., Wang, J., Wang, M., Wang, X., Woodford, K.J., Wortman, J.R., Wu, M., Yao, A., Zdobnov, E.M., Zhang, H., Zhao, Q., Zhao, S., Zhu, S.C., Zhimulev, I., Coluzzi, M., della Torre, A., Roth, C.W., Louis, C., Kalush, F., Mural, R.J., Myers, E.W., Adams, M.D., Smith, H.O., Broder, S., Gardner, M.J., Fraser, C.M., Birney, E., Bork, P., Brey, P.T., Venter, J.C., Weissenbach, J., Kafatos, F.C., Collins, F.H., Hoffman, S.L. (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*, *Science* **298**, 129–149.

Mongin, E., Louis, C., Holt, R.A., Birney, E., Collins, F.H. (2004) The *Anopheles gambiae* genome: an update, *Trends Parasitol.* **20**, 49–52.

Ranson, H., Claudianos, C., Ortelli, F., Abgrall, C., Hemingway, J., Sharakhova, M.V., Unger, M.F., Collins, F.H., Feyereisen, R. (2002) Evolution of supergene families associated with insecticide resistance, *Science* **298**, 179–181.

Zdobnov, E.M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R.R., Christophides, G.K., Thomasova, D., Holt, R.A.,

Subramanian, G.M., Mueller, H.M., Dimopoulos, G., Law, J.H., Wells, M.A., Birney, E., Charlab, R., Halpern, A.L., Kokoza, E., Kraft, C.L., Lai, Z., Lewis, S., Louis, C., Barillas-Mury, C., Nusskern, D., Rubin, G.M., Salzberg, S.L., Sutton, G.G., Topalis, P., Wides, R., Wincker, P., Yandell, M., Collins, F.H., Ribeiro, J., Gelbart, W.M., Kafatos, F.C., Bork, P. (2002) Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*, *Science* **298**, 149–159.

Primary Literature

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool, *J. Mol. Biol.* **215**, 403–410.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The gene ontology consortium, *Nat. Genet.* **25**, 25–29.

Bauer, H., Gromer, S., Urbani, A., Schnolzer, M., Schirmer, R.H., Muller, H.M. (2003) Thioredoxin reductase from the malaria mosquito *Anopheles gambiae*, *Eur. J. Biochem.* **270**, 4272–4281.

Blandin, S., Shiao, S-H., Moita, L.F., Janse, C.J., Waters, A.P., Kafatos, F.C., Levashina, E.A. (2004) Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*, *Cell* **116**, 661–670.

Bryan, J.H., Di Deco, M.A., Petrarca, V., Coluzzi, M. (1982) Inversion polymorphism and incipient speciation in *Anopheles* gambiae s.str. in The Gambia, West Africa, *Genetica* **59**, 167–176.

Collins, F.H., Sakai, R.K., Vernick, K.D., Paskewitz, S., Seeley, D.C., Miller, L.H., Collins, W.E., Campbell, C.C., Gwadz, R.W. (1986) Genetic selection of a Plasmodium-refractory strain of the malaria vector *Anopheles gambiae*, *Science* **234**, 607–610.

Coluzzi, M., Petrarca, V., Di Deco, M.A. (1985) Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*, *Boll. Zool.* **52**, 45–63.

Coluzzi, M., Sabatini, A., della Torre, A., Di Deco, M.A., Petrarca, V. (2002) A polytene chromosome analysis of the *Anopheles gambiae* species complex, *Science* **298**, 1415–1418.

Ding, Y., Ortelli, F., Rossiter, L.C., Hemingway, J., Ranson, H. (2003) The *Anopheles gambiae* glutathione transferase supergene family: annotation, phylogeny and expression profiles, *BMC Genomics* **4**, 35.

Duttlinger, A., Mispelon, M., Nichols, R. (2003) The structure of the FMRFamide receptor and activity of the cardioexcitatory neuropeptide are conserved in mosquito, *Neuropeptides* **37**, 120–126.

Favia, G., Lanfrancotti, A., Spanos, L., Siden-Kiamos, I., Louis, C. (2001) Molecular characterization of ribosomal DNA polymorphisms discriminating among chromosomal forms of Anopheles gambiae s.s, *Insect Mol. Biol.* **10**, 19–23.

Favia, G., Dimopoulos, G., della Torre, A., Toure, Y.T., Coluzzi, M., Louis, C. (1994) Polymorphisms detected by random PCR distinguish between different chromosomal forms of *Anopheles gambiae*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10315–10319.

Favia, G., della Torre, A., Bagayoko, M., Lanfrancotti, A., Sagnon, N.F., Toure, Y.T., Coluzzi, M. (1997) Molecular identification of sympatric chromosomal forms of Anopheles gambiae and further evidence of their reproductive isolation, *Insect Mol. Biol.* **6**, 377–383.

Fournier, D., Mutero, A., Pralavorio, M., Bride, J.M. (1993) *D. melanogaster* acetylcholinesterase: mechanisms of resistance to organophosphates, *Chem. Biol. Interact.* **87**, 233–238.

Gentile, G., Slotman, M., Ketmaier, V., Powell, J.R., Caccone, A. (2001) Attempts to molecularly distinguish cryptic taxa in *Anopheles gambiae* s.s, *Insect Mol. Biol.* **10**, 25–32.

Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E., Okwuonu, G., Hines, S., Lewis, L., DeRamo, C., Delgado, O., Dugan-Rocha, S., Miner, G., Morgan, M., Hawes, A., Gill, R., Holt, R.A., Adams, M.D., Amanatides, P.G., Baden-Tillson, H., Barnstead, M., Chin, S., Evans, C.A., Ferriera, S., Fosler, C., Glodek, A., Gu, Z., Jennings, D., Kraft, C.L., Nguyen, T., Pfannkoch, C.M., Sitter, C., Sutton, G.G., Venter, J.C., Woodage, T., Smith, D., Lee, H.M., Gustafson, E.,

Cahill, P., Kana, A., Doucette-Stamm, L., Weinstock, K., Fechtel, K., Weiss, R.B., Dunn, D.M., Green, E.D., Blakesley, R.W., Bouffard, G.G., De Jong, P.J., Osoegawa, K., Zhu, B., Marra, M., Schein, J., Bosdet, I., Fjell, C., Jones, S., Krzywinski, M., Mathewson, C., Siddiqui, A., Wye, N., McPherson, J., Zhao, S., Fraser, C.M., Shetty, J., Shatsman, S., Geer, K., Chen, Y., Abramzon, S., Nierman, W.C., Havlak, P.H., Chen, R., Durbin, K.J., Egan, A., Ren, Y., Song, X.Z., Li, B., Liu, Y., Qin, X., Cawley, S., Cooney, A.J., D'Souza, L.M., Martin, K., Wu, J.Q., Gonzalez-Garay, M.L., Jackson, A.R., Kalafus, K.J., McLeod, M.P., Milosavljevic, A., Virk, D., Volkov, A., Wheeler, D.A., Zhang, Z., Bailey, J.A., Eichler, E.E., Tuzun, E., Birney, E., Mongin, E., Ureta-Vidal, A., Woodwark, C., Zdobnov, E., Bork, P., Suyama, M., Torrents, D., Alexandersson, M., Trask, B.J., Young, J.M., Huang, H., Wang, H., Xing, H., Daniels, S., Gietzen, D., Schmidt, J., Stevens, K., Vitt, U., Wingrove, J., Camara, F., Mar Alba, M., Abril, J.F., Guigo, R., Smit, A., Dubchak, I., Rubin, E.M., Couronne, O., Poliakov, A., Hubner, N., Ganten, D., Goesele, C., Hummel, O., Kreitler, T., Lee, Y.A., Monti, J., Schulz, H., Zimdahl, H., Himmelbauer, H., Lehrach, H., Jacob, H.J., Bromberg, S., Gullings-Handley, J., Jensen-Seaman, M.I., Kwitek, A.E., Lazar, J., Pasko, D., Tonellato, P.J., Twigger, S., Ponting, C.P., Duarte, J.M., Rice, S., Goodstadt, L., Beatson, S.A., Emes, R.D., Winter, E.E., Webber, C., Brandt, P., Nyakatura, G., Adetobi, M., Chiaromonte, F., Elnitski, L., Eswara, P., Hardison, R.C., Hou, M., Kolbe, D., Makova, K., Miller, W., Nekrutenko, A., Riemer, C., Schwartz, S., Taylor, J., Yang, S., Zhang, Y., Lindpaintner, K., Andrews, T.D., Caccamo, M., Clamp, M., Clarke, L., Curwen, V., Durbin, R., Eyras, E., Searle, S.M., Cooper, G.M., Batzoglou, S., Brudno, M., Sidow, A., Stone, E.A., Payseur, B.A., Bourque, G., Lopez-Otin, C., Puente, X.S., Chakrabarti, K., Chatterji, S., Dewey, C., Pachter, L., Bray, N., Yap, V.B., Caspi, A., Tesler, G., Pevzner, P.A., Haussler, D., Roskin, K.M., Baertsch, R., Clawson, H., Furey, T.S., Hinrichs, A.S., Karolchik, D., Kent, W.J., Rosenbloom, K.R., Trumbower, H., Weirauch, M., Cooper, D.N., Stenson, P.D., Ma, B., Brent, M., Arumugam, M., Shteynberg, D., Copley, R.R., Taylor, M.S., Riethman, H., Mudunuri, U., Peterson, J., Guyer, M., Felsenfeld, A., Old, S., Mockrin, S. and Collins, F. (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution, *Nature* **428**, 493–521.

Githeko, A.K., Brandling-Bennett, A.D., Beier, M., Atieli, F., Owaga, M., Collins, F.H. (1992) The reservoir of Plasmodium falciparum malaria in a holoendemic area of western Kenya, *Trans. R. Soc. Trop. Med. Hyg.* **86**, 355–358.

Grossman, G.L., Rafferty, C.S., Clayton, J.R., Stevens, T.K., Mukabayire, O., Benedict, M.Q. (2001) Germline transformation of the malaria vector, Anopheles gambiae, with the piggyBac transposable element, *Insect Mol. Biol.* **10**, 597–604.

Hallem, E.A., Nicole Fox, A., Zwiebel, L.J., Carlson, J.R. (2004) Olfaction: mosquito receptor for human-sweat odorant, *Nature* **427**, 212–213.

Kanzok, S.M., Fechner, A., Bauer, H., Ulschmid, J.K., Muller, H.M., Botella-Munoz, J., Schneuwly, S., Schirmer, R., Becker, K. (2001) Substitution of the thioredoxin system for glutathione reductase in *D. melanogaster*, *Science* **291**, 643–646.

Knols, B.G., de Jong, R., Takken, W. (1995) Differential attractiveness of isolated humans to mosquitoes in Tanzania, *Trans. R. Soc. Trop. Med. Hyg.* **89**, 604–606.

Mason, G.F. (1967) Genetic studies on mutations in species A and B of the *Anopheles gambiae* complex, *Genet. Res.* **10**, 205–217.

Mitri, C., Parmentier, M.L., Pin, J.P., Bockaert, J., Grau, Y. (2004) Divergent evolution in metabotropic glutamate receptors. A new receptor activated by an endogenous ligand different from glutamate in insects, *J. Biol. Chem.* **279**, 9313–9320.

Mukabayire, O., Besansky, N.J. (1996) Distribution of T1, Q, Pegasus and mariner transposable elements on the polytene chromosomes of PEST, a standard strain of *Anopheles gambiae*, *Chromosoma* **104**, 585–595.

Mukabayire, O., Caridi, J., Wang, X., Toure, Y.T., Coluzzi, M., Besansky, N.J. (2001) Patterns of DNA sequence variation in chromosomally recognized taxa of *Anopheles gambiae*: evidence from rDNA and single-copy loci, *Insect Mol. Biol.* **10**, 33–46.

Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H., Remington, K.A., Anson, E.L., Bolanos, R.A.,

Chou, H.H., Jordan, C.M., Halpern, A.L., Lonardi, S., Beasley, E.M., Brandon, R.C., Chen, L., Dunn, P.J., Lai, Z., Liang, Y., Nusskern, D.R., Zhan, M., Zhang, Q., Zheng, X., Rubin, G.M., Adams, M.D., Venter, J.C. (2000) A whole-genome assembly of *D. melanogaster*, *Science* **287**, 2196–2204.

Olson, M.V. (1999) When less is more: gene loss as an engine of evolutionary change, *Am. J. Hum. Genet.* **64**, 18–23.

Osta, M.A., Christophides, G.K., Kafatos, F.C. (2004) Effects of mosquito genes on plasmodium development, *Science* **303**, 2030–2032.

Ortelli, F., Rossiter, L.C., Vontas, J., Ranson, H., Hemingway, J. (2003) Heterologous expression of four glutathione transferase genes genetically linked to a major insecticide-resistance locus from the malaria vector *Anopheles gambiae*, *Biochem. J.* **373**, 957–963.

Price, D.A., Greenberg, M.J. (1977) Structure of a molluscan cardioexcitatory neuropeptide, *Science* **197**, 670–671.

Ribeiro, J.M. (2003) A catalogue of *Anopheles gambiae* transcripts significantly more or less expressed following a blood meal, *Insect Biochem. Mol. Biol.* **33**, 865–882.

Riehle, M.A., Garczynski, S.F., Crim, J.W., Hill, C.A., Brown, M.R. (2002) Neuropeptides and peptide hormones in *Anopheles gambiae*, *Science* **298**, 172–175.

Rubin, G.M., Yandell, M.D., Wortman, J.R., Gabor Miklos, G.L., Nelson, C.R., Hariharan, I.K., Fortini, M.E., Li, P.W., Apweiler, R., Fleischmann, W., Cherry, J.M., Henikoff, S., Skupski, M.P., Misra, S., Ashburner, M., Birney, E., Boguski, M.S., Brody, T., Brokstein, P., Celniker, S.E., Chervitz, S.A., Coates, D., Cravchik, A., Gabrielian, A., Galle, R.F., Gelbart, W.M., George, R.A., Goldstein, L.S., Gong, F., Guan, P., Harris, N.L., Hay, B.A., Hoskins, R.A., Li, J., Li, Z., Hynes, R.O., Jones, S.J., Kuehl, P.M., Lemaitre, B., Littleton, J.T., Morrison, D.K., Mungall, C., O'Farrell, P.H., Pickeral, O.K., Shue, C., Vosshall, L.B., Zhang, J., Zhao, Q., Zheng, X.H., Lewis, S. (2000) Comparative genomics of the eukaryotes, *Science* **287**, 2204–2215.

Shahabuddin, M., Pimenta, P.F. (1998) Plasmodium gallinaceum preferentially invades vesicular ATPase-expressing cells in Aedes aegypti midgut, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3385–3389.

Takken, W., Knols, B.G. (1999) Odor-mediated behavior of Afrotropical malaria mosquitoes, *Annu. Rev. Entomol.* **44**, 131–157.

Toure, Y.T., Petrarca, V., Traore, S.F., Coulibaly, A., Maiga, H.M., Sankare, O., Sow, M., Di Deco, M.A., Coluzzi, M. (1994) Ecological genetic studies in the chromosomal form Mopti of *Anopheles gambiae* s.str. in Mali, West Africa, *Genetica* **94**, 213–223.

Toure, Y.T., Petrarca, V., Traore, S.F., Coulibaly, A., Maiga, H.M., Sankare, O., Sow, M., Di Deco, M.A., Coluzzi, M. (1998) The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa, *Parassitologia* **40**, 477–511.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K.,

Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Vernick, K.D., Fujioka, H., Seeley, D.C., Tandler, B., Aikawa, M., Miller, L.H. (1995) Plasmodium gallinaceum: a refractory mechanism of ookinete killing in the mosquito, *Anopheles gambiae*, *Exp. Parasitol.* **80**, 583–595.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S.E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M.R., Brown, D.G., Brown, S.D., Bult, C., Burton, J., Butler, J., Campbell, R.D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A.T., Church, D.M., Clamp, M., Clee, C., Collins, F.S., Cook, L.L., Copley, R.R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K.D., Deri, J., Dermitzakis, E.T., Dewey, C., Dickens, N.J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D.M., Eddy, S.R., Elnitski, L., Emes, R.D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G.A., Flicek, P., Foley, K., Frankel, W.N., Fulton, L.A., Fulton, R.S., Furey, T.S., Gage, D., Gibbs, R.A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T.A., Green, E.D., Gregory, S., Guigo, R., Guyer, M., Hardison, R.C., Haussler, D., Hayashizaki, Y., Hillier, L.W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D.B., Johnson, L.S., Jones, M., Jones, T.A., Joy, A., Kamal, M., Karlsson, E.K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W.J., Kirby, A., Kolbe, D.L., Korf, I., Kucherlapati, R.S., Kulbokas, E.J., Kulp, D., Landers, T., Leger, J.P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D.R., Mardis, E.R., Matthews, L., Mauceli, E., Mayer, J.H., McCarthy, M., McCombie, W.R., McLaren, S., McLay, K., McPherson, J.D., Meldrim, J., Meredith, B., Mesirov, J.P., Miller, W., Miner, T.L., Mongin, E., Montgomery, K.T., Morgan, M., Mott, R., Mullikin, J.C., Muzny, D.M., Nash, W.E., Nelson, J.O., Nhan, M.N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M.J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G., Pepin, K.H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C.S., Poliakov, A., Ponce, T.C., Ponting, C.P., Potter, S., Quail, M., Reymond, A., Roe, B.A., Roskin, K.M., Rubin, E.M., Rust, A.G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M.S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J.B., Slater, G., Smit, A., Smith, D.R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J.P., Von Niederhausern, A.C., Wade, C.M., Wall, M., Weber, R.J., Weiss, R.B., Wendl, M.C., West,

A.P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R.K., Winter, E., Worley, K.C., Wyman, D., Yang, S., Yang, S.P., Zdobnov, E.M., Zody, M.C., Lander, E.S. (2002) Initial sequencing and comparative analysis of the mouse genome, *Nature* **420**, 520–562.

Weill, M., Lutfalla, G., Mogensen, K., Chandre, F., Berthomieu, A., Berticat, C., Pasteur, N., Philips, A., Fort, P., Raymond, M. (2003) Comparative genomics: insecticide resistance in mosquito vectors, *Nature* **423**, 136–137.

Yeates, D.K., Wiegmann, B.M. (1999) Congruence and controversy: toward a higher-level phylogeny of Diptera, *Annu. Rev. Entomol.* **44**, 397–428.

Zdobnov, E.M., Apweiler, R. (2001) InterProScan: an integration platform for the signature-recognition methods in InterPro, *Bioinformatics* **17**, 847–848.

Zheng, L., Cornel, A.J., Wang, R., Erfle, H., Voss, H., Ansorge, W., Kafatos, F.C., Collins, F.H. (1997) Quantitative trait loci for refractoriness of *Anopheles gambiae* to Plasmodium cynomolgi B, *Science* **276**, 425–428.

Zieler, H., Nawrocki, J.P., Shahabuddin, M. (1999) Plasmodium gallinaceum ookinetes adhere specifically to the midgut epithelium of Aedes aegypti by interaction with a carbohydrate ligand, *J. Exp. Biol.* **202**(Pt 5), 485–495.

Male Reproductive System: Testis Development and Spermatogenesis

Kate A.L. Loveland and David M. de Kretser
Monash Institute of Reproduction and Development, Monash University, Clayton, Victoria, Australia and The Australian Research Council Centre of Excellence in Biotechnology and Development

1 Testicular Architecture, Cellular Composition, and Principles 499
1.1 The Adult Testis 499
1.1.1 Two Functional Compartments 499
1.1.2 Spermatogenesis is Cyclical 500
1.2 The Developing Testis 502
1.2.1 The Fetal Testis 502
1.2.2 Neonatal and Juvenile Somatic and Germ Cell Differentiation 503

2 Regulation of Gene Expression During Spermatogenesis 505
2.1 DNA 505
2.2 RNA 506
2.2.1 Overview of Transcription 506
2.2.2 Transcription Factors 508
2.2.3 Sex Chromosome–encoded Genes 510
2.2.4 Translational Regulation 510

3 Control Mechanisms 512
3.1 Extragonadal and Hormonal Control of Spermatogenesis 512
3.1.1 Overview 512
3.1.2 The Action of FSH and Thyroid Hormone 513
3.1.3 The Action of Testosterone 515
3.1.4 Estrogen in Spermatogenesis 517
3.2 Molecules Involved in Cell Cycle Control 517
3.2.1 Cyclins 517
3.2.2 Cyclin-dependent Kinases and Regulatory Proteins 518
3.2.3 Other Regulators of Cell Cycle Progression 519
3.2.4 The Machinery of Meiosis 520

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

3.2.5 Regulation of Apoptosis 521
3.3 Growth Factors in Testicular Function 523
3.3.1 Overview 523
3.3.2 c-Kit and Stem Cell Factor 523
3.3.3 Transforming Growth Factorβ Superfamily 525

4 Conclusion 529

Bibliography 530
Books and Reviews 530
Primary References and Specialist Reviews 530

Keywords

Gonocyte
A diploid male germ cell in fetal testis that undergoes mitosis and then becomes quiescent until the onset of spermatogenesis at puberty after birth.

Leydig Cell
A somatic cell producing testosterone, found in the interstitium between seminiferous tubules.

Peritubular Myoid Cell
A somatic cell type with contractile capacity; forms a layer around each seminiferous tubule and contributes to its basement membrane.

Primordial Germ Cell
Indifferent germ cells that exist before the onset of gonadal differentiation in a fetal testis or ovary.

Sertoli Cell
Somatic cells that surround and nourish the developing male germ cells that form the scaffolding of the seminiferous epithelium.

Spermatid
A haploid male germ cell that is transformed from a round shape to an elongated shape during spermiogenesis, as nuclear remodeling, acrosome formation, and tail elongation occur.

Spermatocyte
The male germ cell undergoing meiosis I and II; may be tetraploid (primary spermatocyte) or diploid (secondary spermatocyte).

Spermatogonium
A diploid male germ cell that can undergo mitosis.

Spermatozoon
A motile, elongated, haploid male gamete released from the seminiferous epithelium.

The determination of gonadal sex establishes the basic structure of the testis. This results in the formation of seminiferous cords containing immature Sertoli cells and gonocytes, with the testosterone-secreting Leydig cells lying within the intertubular region. Gonadotrophic stimulation of the testis and the resultant testosterone secretion stimulates spermatogenesis, which involves mitosis, meiosis, and the production of spermatozoa. Germ cell function is crucially dependent on the Sertoli cells that control the intratubular environment, partly by forming the blood–testis barrier. The biochemical nature of the signals that regulate somatic cell function and germ cell maturation is now partially understood, though the complexities of the regulatory pathways and their complete significance are yet to be clarified. This brief overview of the current state of knowledge regarding the controlling signals presents some aspects of their impact on the target cells, both somatic and germinal.

1 Testicular Architecture, Cellular Composition, and Principles

1.1 The Adult Testis

The synthesis of spermatozoa is initiated during puberty and is continuous throughout adult life. Male fertility is dependent on the maintenance of a self-renewing stem cell population and a complex set of signals derived from hormonal and local factors that drive spermatogenesis in a highly regulated manner.

1.1.1 Two Functional Compartments

The adult mammalian testis consists of two structurally distinct but functionally interrelated compartments: the *seminiferous* epithelium, wherein germ cells divide and differentiate surrounded by the Sertoli "nurse" cells, and the *interstitial compartment*. Spermatogenesis occurring within the seminiferous tubules can be divided into three processes: (1) replication of stem cells by mitosis (primordial germ cells in the fetus, gonocytes in the juvenile, and spermatogonia in the sexually mature individual); (2) reduction of chromosomal number by the process of meiosis (primary and secondary spermatocytes); and (3) a complex metamorphosis, called *spermiogenesis*, involving the transformation of round cells (spermatids) into elongated spermatozoa.

The developing germ cells, other than spermatogonia, are enveloped by Sertoli cells, which in turn have access to the interstitial space that contains the Leydig cells,

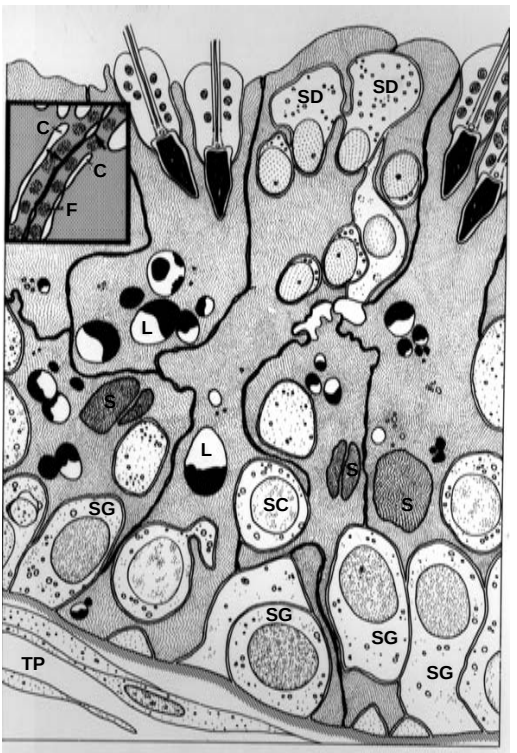

Fig. 1 Diagrammatic representation of the epithelium of the seminiferous tubule. Illustrated are spermatogonia (SG), primary spermatocytes (SC), spermatids (SD), mature spermatozoa, and Sertoli cells (S) with lipid inclusions (L). The peritubular myoid cells lie outside the tubule in the tunica propria (TP). The Leydig cells and vasculature lie external to the tunica propria. *Inset*: organization of the tight junctions between Sertoli cells that form the blood–testis barrier.

peritubular myoid cells, macrophages, and the vasculature fluids (Fig. 1).

The adjacent Sertoli cells are connected by a unique cluster of junctional specializations comprised of tight and gap junctions at their base. This junction divides the seminiferous epithelium into two functional compartments, apical and basal. Mitotic germ cells are in the basal compartment, and they cross the tight junction as they enter their meiotic phase of differentiation. Within the apical compartment of the seminiferous epithelium, the spermatocytes and spermatids are exclusively reliant on Sertoli cell products and are also invisible to the surveillance of the immune system, which might otherwise view them as ''foreign'' cells. Differentiating germ cells are connected to their clonal siblings by cytoplasmic bridges until the time of their release as spermatozoa into the tubule lumen as the result of incomplete cytokinesis. Thus they remain functionally diploid even when their nuclear compartments contain only one parental copy of each chromosome.

It is important to recognize that the seminiferous tubule is an avascular compartment. Intercellular transport into the epithelium cannot occur because there are tight junctions between adjacent Sertoli cells luminal to the spermatogonial layer. Consequently, germ cells other than spermatogonia are dependent on the Sertoli cells for transport and for the regulation of the internal milieu of the tubule.

1.1.2 Spermatogenesis is Cyclical

The process of spermatogenesis is highly organized and ordered, with precise

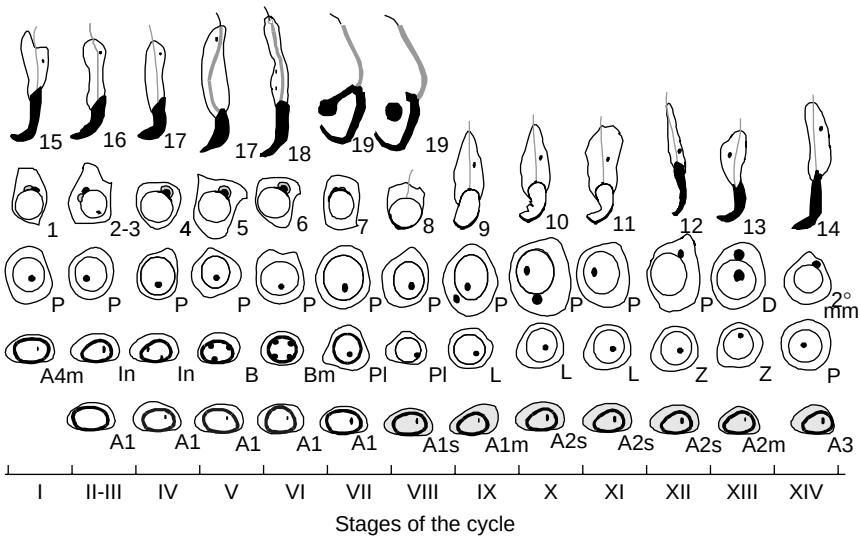

Fig. 2 Stages of the cycle of the seminiferous epithelium observed in the adult rat testis. Each column represents the invariant set of germ cell maturation types present simultaneously within a cross-section of a seminiferous tubule. The least mature germ cell types, present at the base of the seminiferous epithelium, (stem cells $\{A_s\}$, A_{pr}, A_{al}, A_8 and A_{16} spermatogonia) are infrequently observed and are not illustrated here. Type A_1, A_2, A_3 and A_4 spermatogonia (A_1-A_4) are morphologically indistinguishable, while Intermediate (In) and Type B spermatogonia have distinctive heterochromatin. Stages of spermatocytes : preleptotene (PL), leptotene (L), zygotene (Z), pachytene (P), diplotene (D) and metaphase (2°mm). Spermatid steps are indicated by numbers 1 through 19, with 1 to 8 being round spermatids and 9 to 19 elongating spermatids. Adapted from Russell et al. 1990.

associations of germ cells at specific stages of maturation (Fig. 2), creating invariant cycles within the epithelium and waves of stages along the tubule. Although disruption of germ cell development can be induced experimentally, leading to an arrest at discrete stages of spermatogenesis, the normal arrangement of cellular associations is rapidly reestablished upon recovery. The mechanisms that maintain these highly ordered patterns are poorly understood, though they are believed to be essential for attainment of the full spermatogenic potential of the testis. The underlying basis for the spermatogenic wave appears to be established in fetal life, as illustrated by the cyclic expression of the galectin gene in fetal Sertoli cells, prior to the onset of germ cell differentiation. That germ cells can in turn influence the cyclic function of the adult seminiferous epithelium has been demonstrated by interspecific transfer of rat sperm into the mouse testis; the rat sperm specifies the timing of the epithelial cycle when embedded in the mouse Sertoli cell.

The preceding description indicates that the testis contains germ cells and somatic cells involved in discrete, characteristic patterns of macromolecular synthesis. This arises from the steps in spermatogenesis that require replication of stem cells by mitosis, the reduction of chromosomal number and DNA content to the haploid state, and the complex processes constituting spermiogenesis. The latter

requires specific protein synthesis by the haploid genome and the cessation of transcription as the chromatin condenses to form the sperm head. These events and their implications are emphasized in the following sections.

1.2 The Developing Testis

1.2.1 The Fetal Testis

The indifferent gonad The gonads arise from urogenital ridges that form on either side of the midline of the developing embryo. Formed from a thickening of the coelomic epithelium on the surface of the mesonephros, the gonad is initially devoid of germ cells and has the potential to develop as either a testis or an ovary, and hence it is termed *indifferent*. The migration of cells into this area contributes to both somatic (fetal Leydig cell and peritubular myoid cell lineages from the mesonephros; some Sertoli cells from the coelomic epithelium) and germ cell (primordial germ cells from the endoderm of the yolk sac near the allantoic invagination) lineages.

The indifferent gonad is present during approximately weeks 4 to 7 postcoitum in the human and 9 to 12 days postcoitum (dpc) in the mouse. Establishment of this structure requires expression of steroidogenic factor 1 (SF-1; a transcription factor), the Wilm's tumor 1 (WT1) protein isoforms, and fibroblast growth factor-9 (Fgf-9). For each of these, the absence of the gene results in initial formation of the indifferent gonad in the mouse embryo, but the tissue degenerates and does not support germ cell survival.

Genes controlling testis formation The primary molecular trigger for gonadal sex determination is SRY (sex-determining region on the Y chromosome) gene expression, observed in the mouse gonad at 10.5 to 12.5 dpc within the somatic cells that are precursors to Sertoli cells. Introduction of the mouse SRY gene onto a mouse somatic chromosome through transgenesis produced sex-reversed male mice, that is, XX males, having male external genitalia, male sexual behavior, and testes that lack germ cells. Thus, SRY expression directs testis formation but is not sufficient to provide the environment required for spermatogenesis. A highly conserved portion of this gene encodes an HMG (high mobility group) box, characteristic of a family of DNA-binding proteins, and mutations in this region of SRY can cause sex reversal in mouse and humans. The definitive identification of SRY target genes is yet to be achieved.

The first visible sign of testicular differentiation occurs at day 12.5 postcoitum in mouse, with the organization of Sertoli cells into cords surrounding the male primordial germ cells, now termed *gonocytes*. In the female, the initiation of germ cell meiosis at day 13.5 postcoitum marks ovarian development.

One likely downstream target of SRY is SOX9, and, similar to SRY, expression of SOX9 in pre-Sertoli cells is sufficient to drive development of an XX gonad to form a testis. Also containing an HMG box, SOX9 is present in both male and female gonads when indifferent, but it is upregulated in males and downregulated in females following SRY expression onset. As testis differentiation begins, SOX9 moves to the nucleus to effect changes in transcription of downstream genes. SOX9 may switch on the genes required to form testicular cords, and it is known to induce expression of Müllerian Inhibiting Substance

(MIS). The growth factor MIS is produced by fetal Sertoli cells during cord formation, and it mediates regression of the female duct components (see section on Mullerian inhibiting substance). The MIS-gene promoter includes a binding site for SF-1. Expression of fgf9 in the mouse male gonad begins immediately after SRY, and it is required for normal Sertoli cell differentiation, germ cell proliferation, and migration of mesonephric cells into the developing gonad. The expression of neurotropin-3 (NT3) in Sertoli cells appears to regulate mesonephric cell migration and can positively regulate SOX9 expression.

The germ cells are directed to form either gonocytes or oocytes by signals from their somatic cell environment. This commitment occurs in the testis between 11.5 and 12.5 dpc, and, in the ovary, with the commencement of meiosis at 13.5 dpc. Evidence of a positive feedback loop is derived from the production of prostaglandin D2 by germ cells, which will stimulate XX somatic cells to synthesize MIS *in vitro* and form cords, indicative of their differentiation into the Sertoli cell lineage.

This period of development is classically held to be gonadotrophin-independent. The subsequent emergence of a fetal-type Leydig cell population depends on expression of the type A platelet-derived growth factor (PDGF) receptor, and retinoic acid has been implicated in the formation of cords by affecting laminin deposition. These data highlight the concept that, during fetal life, a diversity of locally derived signals are required in order to orchestrate development of the full complement of somatic cell lineages required for the support of spermatogenic cells.

The testis continues to grow through fetal life while the gonocyte remains quiescent. The rate of Sertoli cell proliferation in the rodent peaks just before birth, and the interstitial cell populations continue to mature and multiply.

1.2.2 Neonatal and Juvenile Somatic and Germ Cell Differentiation

The onset of spermatogenesis occurs in a testis that contains quiescent germ cells engulfed by proliferating Sertoli cells in the testis cords (Fig. 3). The cords are surrounded by layers of immature myoid cells, which, along with the Sertoli cells, contribute to the expanding and changing basement membrane at the cord perimeter. Nests of fetal Leydig cells are evident. Within a few days after birth in rodents, and in the prepubertal period, the germ cells are stimulated to reenter the cell cycle and migrate to contact the cord basement membrane. The gonocytes that do not migrate will die by apoptosis; those that do migrate will form the stem cell population that is the progenitor of all adult spermatogenesis. The stimulus for this differentiation, termed *the first wave of spermatogenesis*, appears to be a combination of hormonal (e.g. follicle-stimulating hormone; FSH) and growth-factor (e.g. transforming growth factor β superfamily members) signaling to both germ cells and somatic cells. In the rat testis, germ cell migration is mediated by cell adhesion (N-CAM) and growth factor signaling, with Sertoli cell–derived stem cell factor (SCF) interacting with the c-kit tyrosine kinase receptor on the maturing germ cells (Sect. 3.3.2). The situation in the mouse may be slightly different, and perhaps independent of SCF signaling at this step, with the expansion of this stem cell population being profoundly affected by the presence of glial-derived neurotropic factor (GDNF). Other factors implicated in events at this time include NT3, nerve growth factor

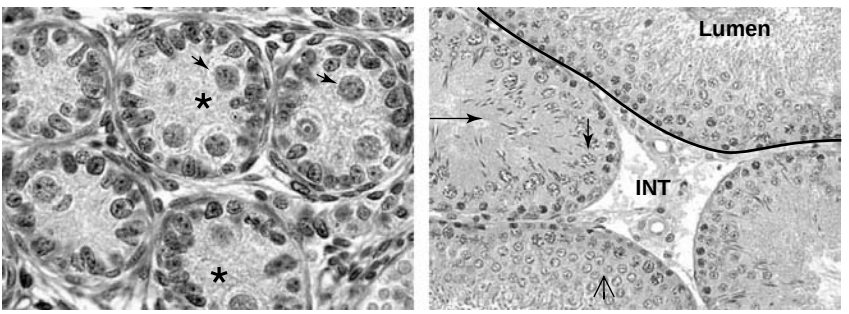

Day 2 postpartum rat testis

Adult rat testis (≥60 dpp)

(a)

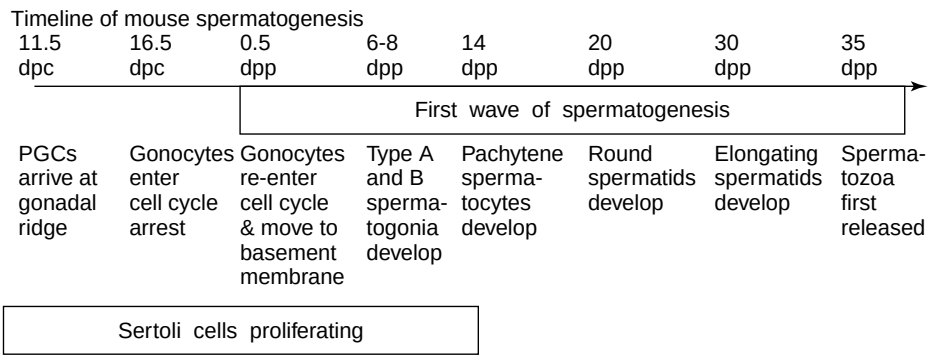

(b)

Fig. 3 Progressive changes in testis during fetal and postnatal development. Micrographs (a) and timeline (b) illustrating changes in the developing rodent testis. (a) newborn rat testis is comprised of cords (shown in cross-section) surrounded by maturing peritubular myoid cells. Gonocytes (arrows) are the only germ cell type present; these are surrounded by Sertoli cell cytoplasm (asterisk). Adult testis tubules (shown in cross-section) are surrounded by flattened layer of peritubular myoid cells; line indicates tubule boundary. Interstitium (INT) contains Leydig cells, macrophages, lymphatics, and blood vessels. Pachytene spermatocytes (arrowheads), round spermatids (open arrows) and elongating spermatids (dashed arrow) are progressively closer to the tubule lumen where mature spermatozoa are released. (b) timeline of murine male germ cell developmental events through to initial production of spermatozoa. Adapted from McCarrey (1993).

(NGF), epidermal growth factor (EGF), PDGF and follistatin (an activin and BMP antagonist).

Examining the rodent testis at discrete time points after birth allows one to observe the appearance of progressively maturing germ cell types. (Fig. 3). In contrast, these time points are highly variable in primates.

Somatic cell differentiation is an important feature of the first wave of spermatogenesis. The cessation of Sertoli cell division during this interval establishes the full complement of these cells that are present throughout the individual's life. This is of crucial importance for establishing the maximum capacity for adult sperm production

(discussed further in Sect. 3.1.2). Sertoli cell proliferation is controlled by interactions between hormones and growth factors, with FSH, thyroid hormone, GDNF, and activin representing some key *in vivo* modulators. A new Leydig cell population arises and differentiates in the interstitium, and the cells surrounding the cords gradually flatten to form the thin layer of peritubular myoid cells observed in the adult testis. Once again, the differentiation and function of these cells appear to respond to a network of hormonal and local factor cues.

To date, a limited number of genes have been identified that show differential expression within a particular cell type between the immature (i.e. first wave) testis and the adult testis. This presumably reflects the changing nature of the seminiferous epithelium with the onset of germ cell, Sertoli cell and peritubular myoid cell differentiation. For example, mRNAs encoding the prosurvival members of the bcl-2 family, bcl-2 and bcl-$_{XL}$, are present in spermatogonia and Sertoli cells during the first two postnatal weeks of mouse testis development but are absent thereafter. The bcl-2 mRNA is switched off entirely, while bcl-$_{XL}$ shifts to the newly emerging spermatocyte population. Such differences are not surprising within the Sertoli cells that transform from mitotic to terminally differentiated cells. Less predictable was the observation that spermatogonia during the first wave of spermatogenesis are functionally different from those present in adulthood. These findings emphasize the notion that germ cell mRNA and protein synthesis are governed by changing environmental cues, including those from the adjacent clones of the progressively more and less mature germ cell clones within the seminiferous epithelium.

2 Regulation of Gene Expression During Spermatogenesis

2.1 DNA

During the preleptotene stage of Meiosis I, there is a peak of DNA synthesis coincident with the peak in DNA polymerase activity that transforms the diploid spermatogonium into a tetraploid spermatocyte. A low level of DNA synthesis documented in pachytene spermatocytes is associated with the DNA repair that follows recombination between homologous chromosomes. DNA methylation patterns vary, with some single-copy genes being methylated at normal levels, some regions of repetitive DNA sequences being under-methylated, and others, such as the genes encoding the chromatin-associated protamine and transition proteins, and the germ cell–specific phosphoglycerate kinase (PGK-2) undergoing changes in methylation during spermatogenesis.

Chromatin remodeling occurs in the haploid germ cells to enable DNA compaction in spermatozoa. The chromosomes become highly organized in hairpin structures of DNA arranged into looped domains along a track of nuclear matrix, with the centromeres found in the center and the telomeres at the perimeter of the sperm nucleus. During spermiogenesis, the histones that package genomic DNA in somatic cells, spermatogonia, and spermatocytes are replaced by transition proteins that are in turn replaced by protamines. The protamines are small, basic proteins represented by 1 or 2 family members. In mouse, both protamine-1 and protamine-2 are required for normal sperm formation. A relatively small proportion of sperm chromatin remains associated with

a nucleohistone; in humans, this accounts for about 15% of sperm DNA. Association of DNA with histones has been proposed to reflect the involvement of genes in these regions with the early postfertilization events.

Inactivation of the X chromosome occurs during meiosis, but it is incomplete. In particular, the Xp terminus remains active, and includes genes encoding the GM-CSF receptor, mic2, and steroid sulfatase in humans. Coincident with X chromosome inactivation, there is activation of autosomal homologs of several genes, including those encoding PGK-2, the EI subunit of pyruvate dehydrogenase, lactate dehydrogenase (LDH-X), and Zfa. These appear to be the result of retroposon activities, as they are characterized by a lack of introns, unlike their counterparts on the X chromosome.

2.2 RNA

RNA synthesis occurs in all spermatogenic cells up through the round spermatid stage, with a constant ratio of DNA to RNA content per cell through meiosis. One study provided the estimate that one half of all mRNAs made in the adult testis increase in abundance after meiosis, and one quarter of adult testicular mRNAs are synthesized only in postmeiotic cells, indicating the requirement for specialized proteins to build a functional spermatozoon.

An important feature of postmeiotic gene expression is that mRNAs produced by the haploid male germ cell are shared between clonal siblings, rendering them functionally diploid while still in the seminiferous epithelium. This was illustrated by the observation of human growth hormone (hGH) transcripts in all spermatids of heterozygote mice bearing a transgene that encodes the hGH gene linked to the protamine promoter. The protamine-transcript sharing between spermatids has also been described, and proteins are also exchanged between clonal siblings at this stage.

As is apparent in every developmental system, unique interactions between many sets of molecules dictate the specific physiological response of the differentiating cell. Because spermatogenesis involves morphological and physiological events that do not happen in any other cell type, it is easy to anticipate that a unique set of genes and regulatory mechanisms will function in the testis. There are also many genes that are expressed in somatic cells that are important for sperm development. Control of the production of proteins required for spermatogenesis is exhibited during both transcription and translation. Many meiotic and postmeiotic gene transcripts are characterized by being structurally distinct from their premeiotic and somatic cell counterparts. The examples in Table 1 illustrate the variety of causes and consequences of this feature. The potential activity of unique RNA-splicing machinery has been postulated on the basis of this information.

2.2.1 Overview of Transcription

Regulation of RNA polymerase II activity in eukaryotic cells is effected by interaction with transcription factors that bind to specific sequences of DNA in each gene. In each cell, the expression level of a particular gene is determined by the nature and abundance of transcription factors at a specific time, and it is common that several factors are influential. Transcription factors are classified on the basis of the predicted structures of their DNA-binding

Tab. 1 Alternative splicing of mRNAs is a feature common to many meiotic and haploid transcripts. This table highlights the diversity of examples.

Protein name	*Known activity*	*Meiotic and/or haploid mRNA splicing features*	*Size [kb]*	
			Somatic or Premeiotic transcript	*Postmeiotic transcript*
c-abl	Cytoplasmic and nuclear tyrosine kinase	3′UTR uses alternative polyadenylation site; no change in protein detected	5.3, 6.5	4.7
c-kit	Receptor tyrosine kinase	Use of alternative splice site; truncated protein encoded lacking all but part of intracellular domain; protein expression not documented (see Sect. 3.4.2)	5.5	3.2, 2.3
CREM	Transcription factor	Protein encoded by spermatocyte mRNA (CREM-τ) has stimulatory rather than inhibitory effect on cAMP-response element (see Sect. 2.2.2)	2 (unique product size detected by PCR)	2
Proenkephalin	Precursor to opiods	Longer 5′UTR; no effect on protein product	1.5	1.9

domains. The common transcription factor motifs are all detected in spermatogenic cells, indicating that the mechanisms employed to effect transcription are conserved between somatic and germ cells. On the other hand, the presence of unique mRNA sizes in spermatids suggests that novel RNA-processing mechanisms exist in the haploid male germ cell.

2.2.2 Transcription Factors

The ubiquitous transcription factor SP1 contains the zinc-finger motif for DNA binding, and it binds to a GC-rich promoter region (GC box) to mediate transcription in genes lacking the TATA box motif. High levels of this protein are present in spermatids. Zinc finger–containing proteins expressed in the testis are encoded on the Y chromosome (*ZFY* in humans and *Zfy-1* and *Zfy*-2 in mouse), on autosomes (*Zfa* in mouse) and on the X chromosome (*ZFX* in human and *Zfx* in mouse). *Zfy-1* expression is first detected in mouse primordial germ cells (PGCs), while *Zfy-2* becomes abundant first between 7 and 14 days of age. Messenger RNA levels in both these increase at the beginning of meiosis. A strain of mouse (*Sxr'*) that lacks the *Zfy-1* and *Zfy-2* genes shows no progression of spermatogenesis into meiosis.

Members of the hormone receptor superfamily also bind to DNA through a zinc-finger domain. Retinoic acid receptor mRNAs are found in Sertoli cells (*Rara*) and (*Rarb)* and in germ cells (*Rara*). Vitamin A deprivation results in blocked proliferation of type A spermatogonia and meiotic prophase entry, while subsequent administration of retinol has been used to produce testes in which all germ cells progress in synchrony through the stages of the seminiferous epithelium.

Homeobox domain and POU domain family members have been detected in the adult testis. Messenger RNAs encoding several OCT proteins have been found in male germ cells at all developmental stages. *Oct-4* gene expression is an important marker of undifferentiated germ cells, as it is present specifically in cells of the embryo epiblast and then in the germ cells that arise from them. In the adult, the *Hox 1.4* gene is expressed in meiotic and haploid germ cells, where its expression may be regulated by retinoic acid.

The leucine zipper family of transcription factors includes *fos* and *jun*, which dimerize to form the AP1 transcription factor that binds to the TPA-response element (TRE). The activity of AP1 varies in cells in response to the activation of protein kinase C, through stimulation by growth factors or phorbol esters. *Jun-D* mRNA is present in all spermatogenic cells, and is predominant in spermatocytes and round spermatids, while the relatively low levels of *c-jun* and *junB* mRNA are highest in type B spermatogonia. *Fra-1* and *c-fos,* and mRNAs are also found in spermatogenic cells. Reversible induction of *c-fos, c-jun* and *junB* transcripts occurs in response to the cell dissociation procedures used to produce purified germ cells for analysis, an important observation for those wishing to study gene expression in purified germ cell populations.

Cellular responses to cAMP levels are mediated through changes in protein kinase A (PKA) activity in order to regulate transcription factors that bind to the cAMP-response element (CRE) in the promoter regions of many genes. Active PKA typically phosphorylates one of the CREB family of bZIP transcription factors, CREB, CREM, or ATF-1 isoforms, enabling its association with a CREB binding protein to affect transcription.

A distinct mechanism of CRE activation occurs in meiotic and postmeiotic germ cells, wherein PKA activation promotes association of CREM with activator of CREM in Testis (ACT), a member of the LIM-only family of transcription factors; this does not involve the activation step *via* phosphorylation required in other cell types. CREs are present in many genes that are specifically or preferentially expressed in the testis or in meiotic germ cells, including those encoding pyruvate dehydrogenase a-2, phosphoglycerate kinase-2, testis angiotensin-converting enzyme, polyubiquitin, transition proteins 1 and 2, protamines 1 and 2, calspermin, and RT-7. In addition, the expression of PKA regulatory and catalytic subunit mRNAs is elevated in meiotic and postmeiotic germ cells, and PKA regulatory subunit isoforms have discrete localizations in the mature human spermatozoan.

The CREB protein exists in a wide range of tissues at low levels, which suggests that its intracellular level is not the primary regulator of its activity, and a modulator protein (CREM) has been characterized (see below). Multiple isoforms of CREB are found in the testis, including those that lack the leucine zipper domain and the nuclear translocation signal; these would presumably regulate the activity of the other isoforms of CREB. CREM also is present in the testis in multiple isoforms. In somatic and premeiotic cells, the presence of CREM α, β, and γ isoforms antagonizes the action of cAMP by downregulating expression of genes bearing the CRE binding motif. Alternative splicing of the *CREM* mRNA in primary spermatocytes produces germ cell–specific forms, designated CREMτ, $\tau 1$, and $\tau 2$, which act as agonists in round spermatids, where these proteins are found. It is thus apparent from these specific changes in patterns of gene expression in primary spermatocytes that a cascade of changes in intracellular signaling via cAMP occurs. Mice lacking the CREM gene are sterile, with defects evident in round spermatids and a complete absence of elongating spermatids. Expression of CREM and CREB is dependent upon both FSH and androgen, while the switch of CREM from antagonist to activator is effected by FSH. Transcription of CREB appears to undergo positive autoregulation, as its promoter contains three consensus CREs.

Other DNA-binding proteins that are unique to spermatogenesis have been identified using a combination of DNA footprinting, gel-retardation assays, *in vitro* transcription assays, and transgenic mouse production. Promoter regions active in regulating transcription specifically in postmeiotic germ cells were initially identified from regions of homology in the 5'UTR of mouse protamine genes, *Prm-1* and *Prm-2*, which are transcribed only in haploid male germ cells. One of these (region D) confers negative regulation of these genes in somatic cells, while regions B and E appear to confer positive regulatory affects. D element motifs have been identified in the *RT7* promoter, in close proximity to a CREB binding site, where at least one of them appears to positively regulate transcription in nuclear extracts of seminiferous tubules. *RT7* encodes a component of the sperm tail outer dense fiber, also named Odf1, which is required for motility. A 39-kDa protein that is present in nuclear extracts of seminiferous tubules, TTF-D, appears to bind the D_1 element of both *RT7* and *c-mos* promoters. Enhancer and repressor elements have been identified in the germ cell–specific *PGK-2* gene, which appear to coordinately determine the appropriate levels of expression

for somatic and germ cells. Other DNA-binding proteins detected in association with developing germ cells include members of the HMG box (e.g. SRY; see section on Genes controlling testis formation), *ets* proto-oncogene, and *c-abl* families.

2.2.3 Sex Chromosome–encoded Genes

The absence of spermatogonia in the *SRY* transgenic *males* highlights the involvement of other gene products in germ cell development. In as many as 40% of cases of human infertility attributed to the male partner, no specific cause, such as endocrinological failure or presence of antisperm antibodies, can be identified. The concept that other genes required for the successful completion of spermatogenesis are located on the Y chromosome emerged from observations of patients with azoospermia who had small Y chromosomes. Several genes, termed azoospermia factors (AZFs), have been identified in humans by intensive mapping of the Yq11 region of such patients. Deletions in the Y chromosome are highly correlated with a low-to-absent sperm count. ($\leq$1 million sperm per ml).

One group of genes identified at multiple subintervals of Yq11encodes RNA-binding proteins, including DAZ and RBMY. Some of these (e.g. RBMY) appear required for storing and then facilitating translation of mRNAs required for spermiogenesis in elongating spermatids when the genomic DNA has already been packaged with protamine and compacted for long term storage in spermatozoa. New information derived from detailed analysis of the human Y chromosome sequence has indicated that a unique combination of mechanisms are likely to have mediated evolution of genes encoded on the male chromosome. In addition to derivation of genes from common ancestors of the X and Y chromosomes, such as RBMY, the transposition and amplification of autosomal genes (e.g. DAZL, a DAZ paralog) and retrotransposition (CDY) have been implicated.

The X chromosome also harbors genes that are specifically used in male germ cells. Following subtractive hybridization to identify mRNAs expressed in mouse spermatogonia, a disproportionate number of those uniquely expressed in male germ cells (i.e. not in ovary or in other somatic tissues) were subsequently identified as products of X-encoded genes.

2.2.4 Translational Regulation

A predominant feature of gene expression in the testis is the storage of mRNAs synthesized in the meiotic and postmeiotic germ cells for translation at later stages. Relatively high levels of certain proteins are essential at discrete stages of spermiogenesis, such as during tail formation and acrosome development. Accessibility of genes encoding the proteins required for transcription becomes limited during spermiogenesis as nuclear DNA condensation occurs. Transcription has been detected in round spermatids up to stage 8, though translation continues during acrosome and tail formation, finally ceasing with the cytoplasmic rearrangements that precede sloughing of residual cytoplasm prior to spermiation.

There are many mRNAs that are synthesized in meiotic cells but are not translated until later on, in spermiogenesis. For example, an autosomal form of the enzyme phosphoglycerate kinase, PGK-2, is synthesized in spermatids following inactivation of the X chromosome, when the X-linked *PGK-1* gene is no longer transcriptionally active. The *PGK-2* mRNA is synthesized in premeiotic and meiotic

cells but is not associated with ribosomes, and hence not translated until after meiosis. The mRNAs encoding the proteins required for chromatin packaging (transition proteins and protamines), acrosome formation, and sperm-tail structures are also under translational regulation.

Sequences that regulate transcription are present in the 5′ and 3′ UTRs of mRNAs. At the 5′ end, interactions between initiation factors (eIF-4F) and the 5′ cap must occur for translation to begin, and the level of phosphorylation of initiation and elongation factors (which determines their activity) is regulated by extracellular signals. Sequences in the 3′ UTR also influence transcription rates, and the poly(A) tail length of some mRNAs fluctuates in response to signals that mediate changes in their translation rates. Poly(A)-binding proteins (PABPs) increase mRNA stability and hence increase the amount of protein produced, while 3′ AU-rich sequences contribute to the destabilization of mRNAs.

Increased poly(A) tail length correlates with higher levels of association with polyribosomes (and therefore increased translation rates) of some meiotic cell–specific mRNAs, including cytochrome c_T and LDH C, while others show no difference. In contrast, poly(A) tail shortening occurs in association with translation of mRNAs encoding the transition proteins and protamines. These mRNAs are synthesized in round spermatids and stored until spermatid elongation, when they sequentially replace histones during nuclear rearrangement. Premature synthesis of protamine 1 protein in a transgenic mouse caused arrest of spermatid development. Through analysis of *Tpap*-/- mice, a testis-specific cytoplasmic poly(A) polymerase, TPAP, has been identified as essential for progression of round to elongated spermatids. The absence of TPAP is associated with decreased expression of genes required for spermiogenesis, including transition protein 1, protamines 1 and 2, sperm fibrous sheath component 1 (Fsc1), sperm outer dense fiber protein RT7 (Odf1), and heat shock protein Hsc70t. The *Tpap*-/- male mouse phenotype is comparable to that in mice lacking the TATA-binding protein-related factor-2 (TRF2). Analysis of the mRNA-encoding TRF2 and other transcription factors involved in RNA polymerase II-mediated transcription revealed a reduction in the length of the poly(A) tails in the *Tpap*-/- mice, and impairment of transport into the nucleus of the TAF10 protein, all contributing to the reduction in TFIID complex activity required for synthesis of the gene products that are essential for spermiogenesis.

Postmeiotic mRNA translation is mediated by RNA-binding proteins, and several of these have been shown, by gene knockout studies in mice, to be essential for postmitotic germ cell differentiation including Prpb, TLS/FUS, TB-RBP, and SPNR. Testis brain RNA-binding protein (TB-RBP/ translin), present in round spermatids but absent from elongated spermatids, binds specifically to conserved sequences (Y and H elements) in many mRNAs including transition protein 1, protamines 1 and 2, AKAP4, glyceraldehyde 3-phosphate dehydrogenase-S, myelin basic protein, calmodulin kinase II, and tau protein. TB-RBP appears to have a role in storage and transport of these mRNAs, binding them to microtubules and potentially moving these mRNAs between clonal siblings through their intracytoplasmic bridges. It has also been characterized as a DNA-binding protein that may link consensus sequences during recombination and repair during

meiosis in male germ cells and in lymphoid tumors. Spermatid perinuclear RNA-binding protein (SPNR) is also a microtubule-associated RNA-binding protein that is essential for spermatogenesis. The *SPNR* mRNA itself appears to be translationally regulated, as it is synthesized in spermatocytes, and the protein is present only after meiosis. *SPNR*-/- mice produce abnormally shaped sperm in reduced numbers, indicating that SPNR is required for normal completion of spermiogenesis. Phosphorylation of RNA-binding proteins, TLS/FUS, TB-RBP, and MSY2, mediates their dissociation from target mRNAs to enable their translation. The phosphatase that effects this translational activation step in male germ cells has yet to be identified. A potential role for a phosphatase antagonist, the Styx protein, which contains a point mutation in the catalytic domain to render it inactive, is suggested by the phenotype of abnormal and reduced spermatid differentiation and infertility.

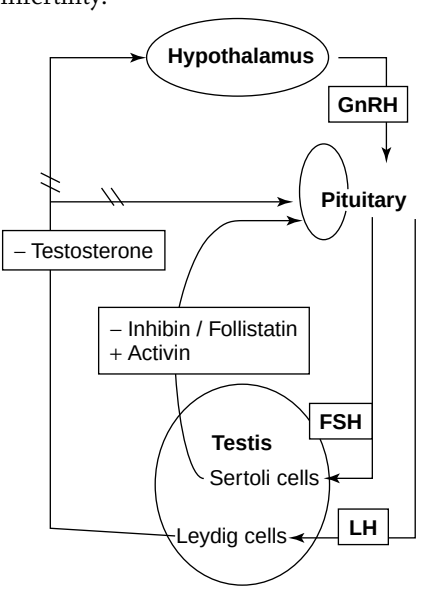

Fig. 4 Illustration of primary components of the feedback loops comprising the hypothalamal–pituitary axis.

3 Control Mechanisms

3.1 Extragonadal and Hormonal Control of Spermatogenesis

3.1.1 Overview

The production of normal sperm numbers in the adult is dependent on the actions of the pituitary gonadotrophins, follicle-stimulating hormone (FSH) and luteinizing hormone (LH). The secretion of these is regulated by gonadotrophin-releasing hormone (GnRH) from the hypothalamus and feedback from the testis *via* testosterone and inhibin (Fig. 4). In the absence of FSH or LH, spermatogenesis is compromised to an extent that varies between species and with the stage of sexual maturity at removal. The receptors for FSH and LH within the testis are present only in the somatic Sertoli cells and Leydig cells, respectively, and hence these hormones affect germ cell development indirectly. The hormone testosterone is also essential for normal spermatogenesis, and its secretion by the Leydig cell is under the influence of LH from the pituitary. Receptors for testosterone (T) are abundant on Sertoli cells, peritubular myoid cells, and Leydig cells. As is true for LH and FSH, the influence of testosterone on spermatogenesis is clearly through the local production of intermediary compounds. In addition, thyroxine has a profound influence on Sertoli cell maturation in immature animals. Recent evidence for the impact of estrogen on testis development has arisen from the analysis of infertile male mice that lack the components of the estrogen synthesis and

signaling pathway; data that reinforce the findings from prior studies that described the impact of estrogen and xenoestrogens.

3.1.2 The Action of FSH and Thyroid Hormone

The action of FSH in the testis is through its impact on Sertoli cell physiology, since receptors for FSH are exclusively located on these cells, while thyroid hormone receptors have been identified in both somatic and germ cells. The impact of these hormones on adult testicular function has been studied using approaches that manipulate their levels and function at defined periods of development and in adulthood. Through recent investigations on humans and mice that are deficient in FSH production and signaling, we now understand that spermatogenesis can proceed in the complete absence of this hormone. These individuals achieve complete spermatogenesis in the lifelong absence of FSH, although the testis size is reduced and fertility is compromised due to suboptimal sperm production and quality.

FSH in the immature testis In immature rats, FSH stimulates Sertoli cell proliferation and the onset of maturation; suppression of FSH levels prior to the normal cessation of Sertoli cell division by day 16 in the rat reduces Sertoli cell numbers in the adult testis. The timing of Sertoli cell proliferation in the juvenile has been established using observation of mitotic figures in histochemical sections and quantitative autoradiography following administration of ^{3}H-thymidine to monitor DNA synthesis. Cell division is evident in the rat fetus at day 16, and the mitotic index of Sertoli cells is maximal between 18 and 21 dpc, corresponding to the period when fetal Leydig cells increase their output of testosterone. The rate of Sertoli cell proliferation decreases after birth and ceases by day 16 in the rat. Thus, the full complement of Sertoli cells in the adult testis is present by this time. It is not known whether Sertoli cell proliferation becomes slower for all cells after birth or whether there exists a discrete population of Sertoli cells that continues to divide until a specific signal to arrest division is received. The response of Sertoli cells to FSH changes during the postnatal period in rats. The initial postnatal period is characterized by being singularly FSH-dependent for proliferation, while between days 5 and 13 postpartum, Sertoli cell proliferation can be transiently stimulated by activin. This effect of activin is augmented by the presence of FSH. Other local factors have been observed to affect Sertoli cell proliferation, including GDNF.

With the termination of mitosis, the Sertoli cell enters a phase of differentiation, wherein continued germ cell differentiation (i.e. onset of meiosis) within the seminiferous epithelium becomes evident, and the lumen of the mature epithelium is formed as Sertoli cells develop tight junctions at the base of the epithelium (forming the blood–testis barrier) and establish a pattern of vectorial secretion. The changed expressions of several mRNAs and proteins are regarded as hallmarks of the now differentiated Sertoli cell, for example, androgen-binding protein.

FSH in the adult testis The impact of FSH in adult spermatogenesis appears to be greatest in regulating spermatogonial development, though its influence on spermatid adherence to the Sertoli cell ectoplasmic specialization has also been documented. However, as stated previously, FSH is not essential for completion of spermatogenesis, but it is required to

achieve quantitatively normal sperm production. In the adult rat, expression of FSH receptor mRNA is greatest at stages IX–XII in the cycle of the seminiferous epithelium, while the protein is apparently most abundant between stages XIII–I.

The impact of FSH on testicular function in the adult varies dramatically between species, with neutralization of FSH having a limited effect on sperm production in rats and a much greater effect in nonhuman primates. Suppression of FSH action has been achieved using passive immunization, injection of high doses of testosterone, and by active immunization against GnRH. In rats, this results in a decrease in testis weight, germ cell numbers, and sperm output of about 10%; in monkeys, a 50% reduction in testis size results from active and passive immunization against FSH. Following hypophysectomy in rats, which removes LH and FSH, germ cell degeneration is most evident at stage VII and can be partially halted by administration of FSH. In this model, in the absence of exogenous testosterone, FSH supports spermatid development to step 8, while the addition of testosterone permits completion of spermiogenesis. This correlates with the finding that androgen receptors are most abundant on stage VII Sertoli cells.

Several *in vitro* studies have documented the effects of the administration or deprivation of FSH on adult rat Sertoli cell protein secretion (e.g. transferrin and inhibin), mRNA synthesis, and cytoskeletal architecture. An increase in spermatogonial and spermatocyte apoptosis, and a reduction in the number of type B spermatogonia entering meiosis is observed in response to FSH withdrawal in adult rats, while administration of FSH increases spermatogonial proliferation and DNA synthesis in spermatogonia and spermatocytes. *In vitro* studies of adult Sertoli cell function indicate that FSH stimulates inhibin secretion both in intact seminiferous tubules and in purified cell preparations. Inhibin will act as a hormone to reduce FSH secretion from the pituitary, but it clearly may have local effects through its impact on TGFβ and activin signaling (see Sect. 3.3.3), thereby affecting germ cell proliferation and differentiation. Some FSH-regulated products are known to act directly on spermatogonia to promote survival, most notably stem cell factor (See section on Interactions between SCF and c-kit affecting spermatogenesis).

Effects of FSH on adult sperm output and quality: *in vivo* and *in vitro* models In the adult rat testis, each Sertoli cell has the capacity to support approximately six to eight round spermatids. Therefore, the potential sperm output of the adult testis (i.e. number of sperm produced per day) is fixed during the postnatal period of Sertoli cell division. Intratesticular administration of an antimitotic drug to reduce neonatal Sertoli cell division resulted in a testis containing a smaller number of germ cells that were present in proportion to number of Sertoli cells at a ratio identical to that in untreated animals.

In addition to the apparent variation in impact of FSH between species, an assessment of the role of FSH on spermatogenesis is complicated by issues such as the purity and source of the FSH administered in older studies, the physiological status of the animal under treatment, and the complexity of intercellular events that occur within the normal testicular environment. With the availability of recombinant FSH for experimental work, and the use of genetically modified mouse models,

a more convincing picture of the interactions between hormones in testicular physiology is developing.

Abrogation of FSH signaling in mice was achieved through the production of mice lacking the genes encoding the FSH β subunit or the FSH receptor. In both these models, spermatogenesis proceeded to completion, but the number of spermatozoa was reduced (to 45–65% of the number in the controls), sperm morphology was abnormal, and fertility was impaired. Similar observations are recorded for patients with homozygous inactivating mutations in the FSH receptor.

Effects of the thyroid hormone on adult sperm output: *in vivo* and *in vitro* models The thyroid hormone receptor is most abundant in rat Sertoli cells during the period when Sertoli cell division is decreasing, between the end of fetal development and the first two weeks after birth. The results of *in vitro* and *in vivo* studies indicate that thyroxine normally decreases Sertoli cell proliferation and promotes Sertoli cell maturation. Thyroid hormone levels affect the cessation of Sertoli cell proliferation. *In vivo* and *in vitro* administration of T3 in rat models promotes premature reduction in Sertoli cell mitosis and the synthesis of products associated with their terminal differentiation. Information from related studies on the influence of thyroid hormone on testicular physiology has reinforced this observation that the Sertoli cell population determines the number of sperm produced in the adult. The induction of neonatal hypothyroidism by administration of prophylthiouracil to the drinking water of pregnant rats and mice results in extension of the period of Sertoli cell proliferation to day 35, with a corresponding delay in their maturation. The extended period of Sertoli cell proliferation results in larger testes in adulthood, with a marked increase in Sertoli and germ cell numbers. Evidence from *in vitro* studies supports the concept that there is synergism between FSH and tri-iodothyronine (T3) in effecting some aspects of this change; thus, the balanced influence of thyroid hormone and FSH actions on Sertoli cell proliferation will have a profound effect on adult male fertility.

3.1.3 The Action of Testosterone

Leydig cell populations and their development The pattern of Leydig cell development varies between species, though the common observation is that Leydig cells are first observed in the interstitium after cord formation. Fetal Leydig cells have been shown to derive from mesenchymal cells originating in the mesonephros under the influence of local factors, including PDGF. Their rapid proliferation is stimulated by chorionic gonadotrophin and LH. This population appears to be static during perinatal life. At the time of puberty, a different population of interstitial cells is stimulated by LH to differentiate into adult-type Leydig cells. It appears that this effect is partly mediated through the actions of FSH. In immature rats, treatment with FSH, but not LH, promoted transformation of interstitial cells into adult Leydig cells, and FSH has also been shown to stimulate both LH receptor activity and T-production in response to LH stimulation. Fetal and adult Leydig cells share the characteristic of containing large amounts of smooth endoplasmic reticulum, and they can also be identified by the presence of steroidogenic enzymes and their capacity to produce androgens. Fetal Leydig cells are easily distinguished by the presence of lipid inclusions, which, in the rat, are mostly absent in the adult Leydig cell.

The abundance of intracellular lipid stores varies in adult Leydig cells with the degree of LH stimulation and availability of dietary cholesterol.

Receptors for LH are found only on Leydig cells, and the addition of LH stimulates production and secretion of testosterone (T). The amount of T produced varies *in vitro* with the addition of various growth factors present *in vivo*, including bFGF, PDGF, and TGFα. Thus, production of T by the Leydig cell is modulated by both local and hormonal signals.

Actions of testosterone The understanding from recent information is that testosterone is required for providing the basal support for spermatogenesis, while FSH affects the quantity and quality of sperm output through its actions in the immature as well as the adult animal.

The presence of testosterone is essential for complete spermatogenesis, though the likely importance of dihydroxytestosterone (DHT) as an alternative ligand for the androgen receptor has been strongly suggested. The binding affinity of the androgen receptor for DHT is 2 to 10 times greater than for T, but T is 5 times more abundant than DHT within the testis and is hence more likely to be available for binding to the receptor. The intratesticular level of T is 50 times greater than what is measured in serum, and the physiological basis for this requirement by the testis of such high levels of T is not known. Androgen-binding protein and albumin in interstitial fluid can both bind T, and may contribute to the two to threefold higher concentrations of T measured in interstitial fluid relative to the testicular venous blood in normal adult rats.

The most profound effects of T on spermatogenesis are exerted at stages VII–VIII of the seminiferous epithelium. Conversion of step 7 spermatids to step 8 spermatids is specifically dependent on T *in vivo*. The absence of T *in vivo* leads to a change in the interaction between Sertoli cells and elongating spermatids. The mature germ cells are not released into the tubule lumen (spermiation), and are instead phagocytosed by the Sertoli cell. Genes encoding CAMs and other components of the ectoplasmic specialization, which mediates attachment of elongating spermatids to Sertoli cells, are regulated by local concentrations of T and FSH.

Restoration of T levels in long-term suppression models leads to rapid restoration of round to elongated spermatid conversion ratios, though the numbers of spermatogonia and spermatocytes are not fully restored. In contrast, the presence of high testosterone levels can inhibit spermatogonial proliferation in models of damaged or recovering testis, such as the *jsd* mouse or the irradiated rat testis, thereby indicating a role for T in facilitating stem cell proliferation or recovery.

While Sertoli cells possess androgen receptors (see above), they may also respond to factors produced by interstitial cells in response to T, to produce the androgen dependency of spermatogenesis. Proteins produced by peritubular myoid cells, such as P-Mod-S, have been implicated in such a scenario, as they stimulate secretion of transferrin, inhibin, and retinol-binding protein by Sertoli cells.

The *hpg* mouse, which bears a naturally occurring deletion of the GnRH gene, has been exploited to elucidate the distinct roles of T and FSH. Administration of T can restore spermatogenic capacity to these animals that are chronically FSH-deficient, in accord with the findings from analyses of FSHβ and FSH receptor knockout animals, though the number of

sperm produced is subnormal. A human FSHβ-subunit transgene was introduced to the *hpg* mouse, resulting in restoration of spermatogonia numbers. Restoration of pachytene spermatocytes to normal and spermatid numbers to near normal numbers was achieved with T administration beginning on day 21 postpartum for 6 weeks. These findings further reinforce the importance of T in the postmitotic differentiation of germ cells.

3.1.4 Estrogen in Spermatogenesis

There is long-standing recognition of the detrimental effects of exogenous estrogen exposure on various aspects of male reproductive tract development and function from the studies of animal models and humans. Fetal and neonatal exposure to estrogenic compounds has been implicated in an increased rate of testicular cancer, cryptorchidism, epididymal abnormalities, and decreased fertility.

In recent years, the influence of estrogens on spermatogenesis has gained a deeper appreciation. Animals lacking the genes encoding the estrogen receptor-α and -β genes (ERαKO and ERβKO mice) and the aromatase cytochrome P450 gene (*CYP19;* ARKO mice) have been generated and examined, and reagents have been developed that can discern the cellular localization of each receptor subunit. The ERαKO mice have impaired fertility due to impaired efferent duct structure and function, while the ERβKO mice are fertile. The ARKO mice suffer from a progressive loss in spermatogenic capacity; the animals are fertile and testicular histology is normal at 14 weeks postpartum, while abnormalities in both the interstitial and seminiferous tubule compartments are apparent after 18 weeks. The double ERKO mice (ER$\alpha\beta$KO) lacking both receptor subunits revealed a phenotype similar to that of the ERαKO, but with distinctively affected mounting behavior.

Because each receptor subunit bears distinct ligand binding and responsive properties, the collection of information about receptor distribution and ligand production is required in order to fully understand the etiology of these phenotypes. Much of this information remains to be sorted out, and there are conflicting results in the literature, some of which are likely to highlight species-specific differences. The ERα appears to be restricted to Leydig and accessory duct cells. ERβ is more widely distributed, found on Leydig and Sertoli cells, in accessory ducts, and in spermatogonia, pachytene spermatocytes, and round spermatids. The aromatase enzyme distribution is similarly broad, having been detected in Leydig and Sertoli cells, pachytene spermatocytes, and in round and elongating spermatids.

3.2 Molecules Involved in Cell Cycle Control

3.2.1 Cyclins

The identification of control points in spermatogenic progression has been achieved by determining cyclin family member expression patterns. Cyclins, proteins originally identified in invertebrate eggs, undergo tremendous changes in levels of expression immediately following fertilization and in subsequent rounds of cell division, with passage through the cell cycle. Individual cyclins have been associated with specific stages of the cell cycle and, to date, at least 8 familes have been identified. Family members are classified as cyclins A through H based on their amino acid homologies and/or their time of appearance during the cell cycle. Cyclin B

associates with a serine/threonine kinase (a cyclin-dependent kinase (Cdk), Cdk1; named $p34^{cdc2}$ kinase in yeast) to form the maturation promoting factor (MPF) which allows the cell to progress through the M-phase of the cell cycle in mitosis and meiosis. The kinase activity of each Cdk is regulated by a complex series of phosphorylation and dephosphorylation events, dependent upon its association with cyclin. Following M-phase, cyclin B is degraded by proteolysis following cross-linking to ubiquitin. This scenario represents the paradigm of how cell cycle passage is regulated, through a sequence of cyclin activation and deactivation events.

Analysis of cyclin gene expression during murine spermatogenesis shows that members of the cyclin A, B, and D families are expressed at discrete sites within the seminiferous epithelium in association with distinct stages of spermatogenic differentiation. There are some differences between reports of their cellular localizations that most likely reflect the nature of the reagent used (i.e. the specificity of the antibody) and the detection method applied. Functional data are predominantly derived from analyses of mice lacking a specific cyclin gene.

The cyclin A family members are usually associated with the G1/S transition of mitotic cells. Cyclin A2 mRNA is present in spermatogonia and preleptotene spermatocytes, and it appears to regulate entry into the meiotic prophase. Cyclin A1 mRNA is synthesized in late pachytene spermatocytes and is essential for the pachytene-to-diplotene transition in mouse spermatocytes. Cyclin B family members participate in the G2/M transition. Cyclin B3 is present in preleptotene and zygotene spermatocytes, and cyclin B2 protein is most abundant in pachytene spermatocytes. Cyclin B1 mRNA is most abundant in round spermatids, although it too is detected in pachytene and more advanced spermatocytes. Cyclin D family members are most abundant in G1 stage cells and are involved with the G1/S transition in many cellular contexts. In the murine testis, cyclin D3 has been observed in gonocytes, round spermatids, Sertoli cells, and Leydig cells. Cyclin D2 expression has been reported as low to negligible, with protein being detected in limited numbers of adult mouse spermatogonia and in Sertoli cells. Its expression in spermatogonia appears to be associated with those cells that are entering the differentiation pathway. The cyclin D2 knockout male mouse testes show a severe loss of germ cells. Cyclin D1, the expression of which is restricted to Sertoli cells, has also been identified as an estrogen–receptor ligand, the significance of which remains to be established.

3.2.2 Cyclin-dependent Kinases and Regulatory Proteins

Cyclin activity is clearly linked to association with a specific Cdk. Cdk4, the binding partner to cyclins D2 and D3, is most highly expressed in the mouse testis during the first spermatogenic wave, predominantly in spermatogonia and Sertoli cells. The absence of this gene in mouse leads to a loss of spermatogonia and spermatocytes. The predominant expression of *Cdc2* and *Cdk2* mRNAs in spermatocytes and their presence in postmeiotic germ cells suggests that these molecules are involved in germ cell differentiation. Other *Cdk* family genes are expressed in adult Sertoli cells, which are nonproliferating, indicating that roles other than cell cycle regulation are played by these kinases.

The cdks that function during G1 of the cell cycle are regulated by specific cyclin-dependent inhibitors (CKIs) belonging to

two families, and many of these functions facilitate rather than inhibit cyclin action. The cip/kip CKIs preferentially affect the action of cyclins A and E through Cdk2. The $p27^{Kip1}$ protein is found in terminally differentiated Sertoli cells, while $p21^{Cip1/Waf1}$ is in pachytene spermatocytes and round spermatids. The INK4 CKIs bind to and inhibit Cdk4 and Cdk6, thereby blocking the activation of the D-type cyclins, the susbsequent phosphorylation of Rb, and the entry into the S-phase of the cell cycle. The INK4 proteins $p18^{Ink4c}$ and $p19^{Ink4d}$ are synthesized in spermatocytes. The absence of either of these genes in mice results in fertile adults with smaller-than-normal testes, increased germ cell apoptosis, and reduced sperm numbers. The combined absence of these two CKIs results in a more severely infertile phenotype in male mice that exhibit delayed entry into meiosis and elevated numbers of spermatocytes undergoing apoptosis. These mice are also characterized by Leydig cell hyperplasia (as seen in the $p18^{Ink4c}$ KO mouse) and elevated FSH. The impact of deregulated cyclin D function, therefore, affects both somatic and germ cells in these animals. This finding may be related to how cyclin D/CDK4 can prevent the inhibitory action of $p27^{Kip1}$ on cyclin E/CDK2. With the loss of CDK4, $p27^{Kip1}$ activity is elevated, and cyclin E/CDK2 function is diminished.

Dephosphorylation of Cdk1 at specific tyrosine residues by the phosphatase *cdc 25* increases its serine/threonine kinase activity to allow progression through mitosis. The expression of *cdc25C* in the adult mouse is highest in the testis, with the germ cells being the predominant site of expression of its mRNA. It is suggested that interactions between Cdc25, Cdk1, and CycB2 in meiotic cells mediate progression through the G2-M transition of meiosis.

3.2.3 Other Regulators of Cell Cycle Progression

The association of a cyclin with a Cdk results in the formation of an active enzyme complex that phosphorylates the retinoblastoma protein (Rb). Rb generally acts to prevent progression through the cell cycle by binding to E2F and blocking its ability to drive transcription. Phosphorylation of Rb leads to its dissociation from E2F, and gene products required for DNA synthesis are then produced. Rb has been detected in Sertoli cells, spermatogonia, and late-stage elongated spermatids in the rat testis, and it undergoes phosphorylation in a cyclic manner through the cycle of the seminiferous epithelium, indicating that its function is regulated during spermatogenesis.

Progression through meiosis depends on the presence of Hsp70.2, a molecular chaperone required for cyclin B/Cdk1 function. Mice lacking Cks2, a mammalian homolog of a yeast Cdk-1 binding protein, enter prophase I of meiosis but fail to complete the first meiotic division (MI), indicating that a specific cyclin–cdk complex is crucial for spermatocyte progression into anaphase. Deletion of the gene encoding the DNA repair protein, Ercc1, affects murine germ cells at all stages of development, leading to the removal of damaged germ cells by apoptosis.

The c-mos protein was originally identified as a component of cytostatic factor with influence during meiosis in testes and ovaries. C-mos regulates progression through the cell cycle at least partly by limiting cyclin degradation. The testis-specific isoform is a 43-kDa protein detected predominantly in spermatocytes. Overexpression of c-mos protein in the pachytene

spermatocytes of transgenic mice resulted in no overt changes in phenotype, including fertility. Overexpression in spermatids resulted in enlargement of testicular size, apparently due to increased numbers of germ cells.

Expression of the pin1 protein has been implicated in enhanced expression of cyclin D1 in several tissues. Its deletion in a mouse knockout model produced a phenotype that was strikingly similar to that of the cyclin D1 knockout mouse in mammary and retinal tissue. However, the testicular phenotype of hypoplasia was similar to that observed in the cyclin D2 knockout mouse, suggesting that pin1 may interact with D2 in the testis.

An understanding of the roles of these and other proteins that interact with the machinery that regulates mitosis and meiosis will reveal much about how hormones and growth factors combine to regulate spermatogenesis. One example is derived from an analysis of mice lacking both *p*27^{Kip1} and *inhibin* α (a member of the transforming growth factor β superfamily; see section on Activin, Inhibin, and Follistatin). The absence of inhibin α leads to development of Sertoli cell and granulosa cell tumors beginning 4 weeks after birth, while those lacking p27^{kip1} have enlarged organs, including testes. Inhibin acts as a *tumor suppressor* in normal tissue by promoting association of p27^{Kip1} with cyclinD2/E1 and Ck4/2, holding them in an inactive complex. In a normal testis, the balance of signals between FSH and inhibin drive Sertoli cell terminal differentiation and exit from the cell cycle, leading to the onset of p27^{Kip1} expression. In the absence of either one of these factors in *p*27^{Kip1}-/-/*inhibin* α-/- mice, this signaling pathway is lost, resulting in the early onset of tumorigenesis.

3.2.4 The Machinery of Meiosis

A plethora of gene products expressed in somatic cells also affect testis development and/or germ cell maturation. The genes that govern the events of meiosis, however, have a unique relevance to the biology of gametogenesis and hence to spermatogenesis. The meiotic prophase begins with DNA replication by the preleptotene spermatocytes. The chromosomes condense and axial elements of the synaptonemal complex are formed during the leptotene stage, followed by pairing of the homologous chromosomes during zygotene. The longest stage of meiosis in the male, pachytene, is characterized by fully aligned, synapsed chromosomes that have undergone the recombination that is required for progression through diplotene (when the chiasmata resolve) and into the successive meiotic divisions, M1 and M2.

Checkpoints to ensure the fidelity of replication and DNA rejoining are key features of meiosis. While some of the crucial genes are known from analysis of infertile mice, others represent mammalian homologs of highly conserved genes that mediate cell division in yeast. DNA mismatch repair proteins, DMC1, COR1, MSH4, MSH5, and MLH1 facilitate formation of the synaptonemal complex that links homologous chromosomes during Meiosis I. The murine Brca1 protein is required for crossing over and repair of DNA damage during spermatogenesis – a function known to be required for suppressing breast tumor formation in humans. The sister chromatid proteins, SCP1, SMC1, and SMC3, facilitate DNA recombination, and the heat shock protein, HSP70-2 mediates disassembly of the synaptonemal complex.

The unique passage of the male germ cell through the G2/MI transition of

Meiosis I has been examined by treating mouse spermatocytes with the protein phosphatase inhibitor, okadaic acid. The pachytene (but not leptotene or zygotene) spermatocyte is competent to proceed through MI all through the several days in which the homologous chromosomes remain fully paired *in vivo*, and the specific machinery required to proceed through the meiotic metaphase is currently being clarified.

3.2.5 Regulation of Apoptosis

It has been estimated that up to 75% of germ cells entering the differentiation pathway will die by apoptosis without completing spermatogenesis. As in the developing ovary, there are waves of germ cell apoptosis in the rodent testis that coincide with the commitment to spermatogenesis by gonocytes and entry of spermatocytes into meiosis during the first wave. This cell death appears related to the need to maintain the optimal numerical ratio of Sertoli to germ cells, to ensure maximal production of fully fertile sperm, while also serving as a quality control mechanism to delete damaged germ cells.

Two pathways for triggering apoptosis are well understood, and both have been identified and characterized within the mammalian testis. The death receptor–mediated pathway is triggered by a signal from one of several extracellular ligands in the tumor necrosis superfamily (e.g. FasL, TNFα) upon binding to a specific cell surface receptor (e.g. Fas, TNFR1). Ligand binding causes receptor trimerization and recruitment of an adaptor protein (e.g. FADD), which leads to association with, and activation of, the central executioner complex of caspase enzymes, initially through activation of caspase 8. The complementary expression of Fas in Leydig and germ cells (quiescent primordial germ cells in the fetus and spermatocytes in the adult) and FasL in Sertoli cells, and of TNFα in spermatocytes and TNFR1 in Sertoli cells, can govern survival of each of these cell types. Treatments causing upregulation of only Fas (radiation, hyperthermia, ischemia/reperfusion) induce germ cell apoptosis, while those increasing Sertoli cell apoptosis (treatment with specific toxicants) induce upregulation of both Fas and FasL.

A separate pathway is triggered in response to a variety of extracellular insults (e.g. irradiation, cytotoxic drugs, serum deprivation) and intracellular events (DNA damage, inappropriate expression of cell cycle regulatory molecules). In this pathway, the balance of prosurvival and pro-apoptotic Bcl-2 family proteins dictates the fate of each cell. In response to injury (e.g. local heating), the upregulation of pro-apoptotic family members leads to cell death. The pro-apoptotic proteins can drive release of cytochrome c from the mitochondrion that, then, binds to Apaf-1, enabling it to oligomerize and recruit procaspase-9. This *apoptosome* now functions to recruit and activate downstream caspases. In both pathways, the end result of caspase activation is degradation of proteins required for maintenance of cell integrity, including nuclear lamins and cytoskeletal elements.

Conflicting reports regarding the cellular localization of Bcl-2 family proteins stem from the limited availability of reliable antibodies; however, it is clear that the expression of these proteins is cell-specific and alters during testis development. For example, while prosurvival *Bcl-2* mRNA is readily detected in spermatogonia and Sertoli cells of the newborn and developing

mouse testis, it is reportedly undetectable in adult testis. Prosurvival Bcl_{XL} and pro-apoptotic Bad are initially synthesized in spermatogonia and Sertoli cells but shift to being synthesized exclusively within spermatocytes as the testis progresses through the first wave of spermatogenesis. Prosurvival Bcl-w and its partner, pro-apoptotic Bax, are present in spermatogonia and Sertoli cells throughout development and adulthood, while the pro-apoptotic Bim is widely distributed.

The functional importance of the balanced expression of the Bcl-2 family proteins has been repeatedly demonstrated though production of transgenic animals that misexpress a prosurvival family member or knockout mice that lack one. In the case of Bcl-2 overexpression in spermatogonia or deletion of Bax, the enhanced survival of germ cells leads to disruption of the seminiferous epithelium about 3 weeks into development, apparently due to the inability of the seminiferous epithelium to function with the excess number of immature germ cells. In the case of Bcl-w deletion, there is a delay in the onset of germ cell death to 3 to 4 weeks postpartum, presumably due to the compensatory function of other prosurvival family members, Bcl-2 and Bcl_{XL}, during the first wave of spermatogenesis.

Expression levels of Bcl-2 family members is probably influenced by the normal milieu of signals that affect testis cells, though to date there are only limited data concerning this point. The levels of bcl-w and bcl-$_{XL}$ in stage VII tubule segments of adult rats are stimulated by *in vitro* exposure to stem cell factor, a known germ cell survival factor normally synthesized by the Sertoli cell under partial regulation by FSH (see section on Stem cell factor). Addition of FSH to these cultures supports germ cell survival through the SCF/c-kit signaling pathway. The rodent testis responds to a variety of insults with changes in apoptotic regulators. Exposure to mild hyperthermia (testicular heating to 43 °C for 15 min) causes redistribution of bax protein from the cytoplasmic to the perinuclear area of the cell within 30 min.

There are conflicting reports describing the effects of testosterone withdrawal by administration of a specific Leydig cell toxicant, ethane dimethylsulfonate, and changes in bcl-2 family members as well as in Fas, have been reported.

The p53 mRNA and protein are found in spermatocytes. A protein with a complex set of functions, inactive p53 is bound in the cell to the heat shock chaperone complex, and when activated, it can regulate expression of bcl-2 and bax to drive apoptosis. Its absence in mice has been reported to impede correct completion of meiotic division l and slightly affect sperm morphology and male fertility. Conflicting reports regarding its functional impact on spermatogenic progression are thought to arise from the differences in the mouse strains examined. Apoptosis of spermatocytes but not spermatogonia appears to be dependent on the presence of p53 in the mouse under normal conditions, whereas the converse has been reported in the case of irradiation-induced DNA damage.

Heat shock factor 1(HSF1) is also bound to the heat shock chaperone complex when inactive. Its release is triggered by various cellular stressors, and the translocation of HSF1 to the nucleus promotes synthesis of heat shock proteins. Similar to p53, HSF1 also appears to have different effects on spermatogonia and spermatocytes, protecting the former and driving apoptosis of the latter in response to heating. Unlike in other cell types, these responses appear to be independent of Hsp70 protein production.

3.3 Growth Factors in Testicular Function

3.3.1 Overview

The term growth factor is loosely applied to molecules that mediate cell proliferation, and yet it is clear that they may instead abrogate proliferation in some circumstances and facilitate differentiation in others. Given the diversity of cell types that are present simultaneously within the testis, it is difficult to distinguish the specific actions of a certain molecule. Highly purified cell preparations and the development of *in vitro* model systems continue to be important for studying the effects of growth factors on testicular cell types. Many examples of potential paracrine and autocrine interactions have been described, and the list of these continues to grow. Many have been detected in more than one cell type, a fact that should inspire caution in investigators as they assess the significance of experimental responses to modulating a growth factor signaling event. Localization and physiological data should also be addressed and interpreted with caution, because of the known artifacts induced by testis cell dissociation and culture. The availability of natural mutants or *genetically engineered* mice with altered expression of growth factors and receptors has provided some information about the roles of growth factors in the testis, but these avenues are limited by the normally diverse biological roles of these same substances outside the testis. Mice with tissue- and cell type-specific ablation of individual genes and the manipulation of immortalized somatic and germ cells *in vitro* are providing new information about local regulation of spermatogenesis.

It is not possible to review the roles of many of the growth factors that are progressively being identified in the testis. The sections that follow will provide an overview of the two growth factor systems in the testis that have been studied with relative intensity.

3.3.2 c-Kit and Stem Cell Factor

The c-kit receptor tyrosine kinase Mutations at the *White-spotting (W)* and *Steel (Sl)* loci in mice can result in animals that are anemic, sterile, and white in coat color. Naturally occurring mutations at these loci have provided a superb source of information concerning two genes that are essential for normal germ cell development. The *W* locus encodes c-kit, a receptor tyrosine kinase glycoprotein in the PDGF receptor family. In humans, the *C-kit* gene covers 20 kb and is comprised of 21 exons. Mutations involving this locus include large deletions, rearrangements and point mutations, and these have been documented in humans, mice and rats. These mutations can affect the amount of c-kit protein expressed and the level of kinase activity; all naturally occurring point mutations described to date are within the kinase domain of the protein. Phenotypic changes accompanying heterozygous mutations affect the melanocyte lineage in humans, but to date no association with human infertility has been documented, nor have homozygous mutations been documented. Dimerization of both ligand and receptor subunits is required for signal transduction, causing autophosphorylation of the c-kit receptor and subsequent recruitment of signaling molecules and activation of Akt by phosphatidylinositol-3 kinase (PI-3K), and of extracellular-signal-related kinases 1/2 (ERK1/2) through SH2 binding proteins including Grb2. This activation drives translocation of cyclin D3 to the nucleus where it effects cellular progression through G1 and into S phase.

Stem cell factor The ligand for c-kit is an integral membrane glycoprotein encoded at the *Sl* locus, and it has been named stem cell factor (SCF; also known as mast cell growth factor (MGF), kit ligand (KL) and steel factor). SCF is encoded by at least 8 exons in mouse, rat and humans. Exon 6 encodes an amino acid sequence that is readily cleaved by proteases on the surface of cells *in vitro*, and this cleavage produces a soluble form of SCF. Two SCF mRNA forms are common to every tissue examined, one including and one excluding exon 6. Differential responses by the c-kit-positive target cell to the soluble and membrane-anchored SCF isoforms have been recorded. The relative proportions of the two SCF mRNA forms changes during testicular maturation, with the membrane-anchored form predominant in early development and the soluble form prevalent in the adult. This alternative splicing is regulated *in vitro* by changing the pH of the culture medium, and it is thought to reflect *in vivo* changes in the local environment as the Sertoli cells mature. Synthesis of SCF is promoted by FSH, and this is a key mechanism by which FSH facilitates spermatogonial survival in the postnatal testis.

Interactions between SCF and c-kit affecting spermatogenesis The involvement of c-kit and SCF in spermatogenesis was originally observed during analyses of *W* and *Sl* mice. Phenotype characteristics shared by mice with mutations in either of these genes reflected a loss of stem cell function in hematopoietic, melanocyte, and germ cell lineages. These mice had varying degrees of anemia, white fur patches, and infertility.

In both *Sl*/*Sl* and *W*/*W* homozygotes, the PGCs form normally and are present in normal numbers until 8.5 dpc, when the number of PGCs declines to about 2% of normal by 12.5 dpc. By 14 dpc, the *Sl*/*Sl* mice are devoid of germ cells, and the PGCs have not migrated from the hindgut to the gonadal ridge, while some migration is seen in the *W*/*W* individuals. *SCF* mRNA is detected in mouse mesodermal cells along the migratory pathway of the PGCs from about 9 dpc and in the genital ridge at 12.5 dpc. Isolated primordial germ cells of mice have been grown in culture with SCF and other growth factors. In these systems, the addition of SCF enhances PGC survival; more limited effects on PGC proliferation and adhesion have been described. In addition, recombinant human leukemia inhibitory factor (LIF) stimulates PGC proliferation *in vitro*, and it is probable that other locally produced factors including BMPs (see sections 'The indifferent gonad' and 'Bone morphogenetic proteins'.) regulate migration, proliferation, and differentiation of these cells.

SCF/c-kit interactions are also essential in postnatal spermatogenesis. Messenger RNA encoding c-kit is found in Leydig cells, differentiating spermatogonia, primary spermatocytes, and round spermatids. C-kit receptor protein has been detected on Leydig cells and spermatogonia, and the postmitotic cell types contain mRNA encoding a truncated protein, tr-kit, consisting of only part of the intracellular domain. Tr-kit accumulates in the post-acrosomal region of the sperm head and in the mid-piece of the sperm tail. It is believed to have a role after fertilization in triggering embryo development, as its microinjection into mouse eggs leads to their activation. In contrast to the differentiating type A spermatogonia that have c-kit mRNA and a functional c-kit signaling receptor, stem cell spermatogonia are

characterized by the absence c-kit mRNA and protein. The mechanisms that govern upregulation of c-kit, as these cells become committed to differentiate, are clearly important to identify, and Bmp4 has recently been identified as playing a role (see section on Bone morphogenetic proteins).

Membrane-anchored SCF produced by Sertoli cells can interact with c-kit on spermatogonia, while the cleaved form could reach c-kit on additional distal targets such as interstitial Leydig cells. Binding of murine spermatogonia to Sertoli cells, specifically *via* membrane-anchored SCF interaction with c-kit, has been demonstrated *in vitro*. An antiserum that blocks binding of SCF to c-kit disrupts proliferation of type A_{1-4} spermatogonia *in vivo* and *in vitro*, causing increased apoptosis of spermatogonia and spermatocytes. Genetically modified mice bearing a specific point mutation required for PI-3K-mediated c-kit signaling are infertile due to the specific loss of type A spermatogonia. Abrogation of the PI-3K signaling pathway does not cause the pleiotropic phenotypes of white patches and anemia observed in other *W* locus mutants. These data indicate that Sertoli cell signaling through c-kit is required for survival and may impact on mitotic progression of spermatogonia.

3.3.3 Transforming Growth Factorβ Superfamily

The transforming growth factor, βs (TGFβs), belongs to a family of at least 50 proteins sharing a common dimeric structure that is built upon monomers containing seven conserved cysteines. Two common receptor subtypes with serine/threonine kinase activity are involved in most TGFβ superfamily signaling, type I and type II. The ligands in this family have both distinct and overlapping pathways for signal transduction (Fig. 5). Signaling by each of these many ligands is affected by the local availability of a host of soluble and inhibitory ligands, competition for receptor subunits, and competition from inhibitory and stimulatory intracellular signaling components. Owing to the wide variety of effects of these proteins on a broad spectrum of biological systems, it is clear that their expression and action in the testis will be important in a multiplicity of ways that pose a significant challenge to experimenters. In this section, we will review some aspects of this family of growth factors and receptors that relate to our understanding of their roles in testicular function.

TGFβ Three TGFβ genes have been identified in mammals, encoding TGFβ1–3. These are synthesized as large, inactive glycoproteins that must be cleaved and/or released from binding proteins in order to become activated. TGFβ has an additional type III receptor subunit, betaglycan, a transmembrane proteoglycan with the capacity to bind all three TGFβs that is required to enable TGFβ binding to the type II receptor. Betaglycan is also a coreceptor for inhibin A (see section Activin, Inhibin, and Follistatin). Endoglin is another type III receptor for TGFβs 1 and 3 that can also bind activin A and BMPs 2 and 7.

TGFβ receptors on mouse primordial germ cells may mediate suppression of their proliferation, as has been shown *in vitro* on dissociated cells. Later in fetal development, TGFβ application to cultured rat testis fragments induced apoptosis of gonocytes at the ages when the germ cells were proliferating (13.5 dpc and day 3 postpartum), but not at an age when these cells are quiescent (17.5 dpc). Rat gonocytes possess both type I and type II receptors and synthesize TGFβ3. In the adult rat, TGFβ receptor types are

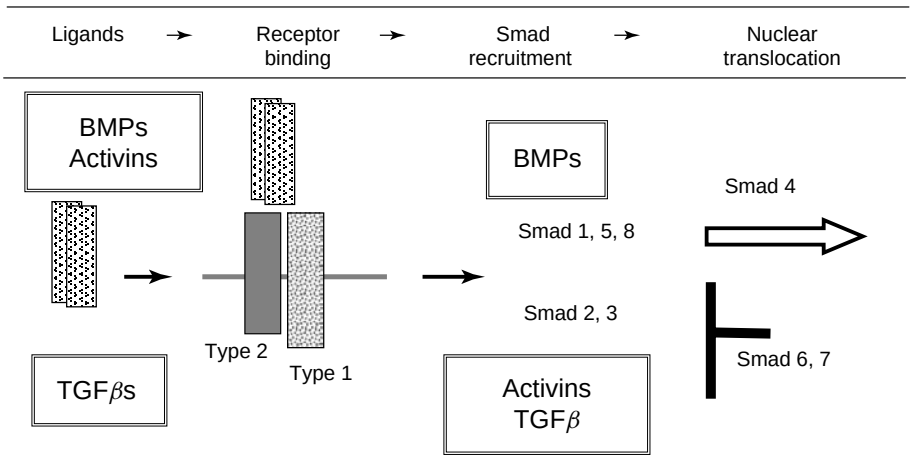

Fig. 5 The TGFβ superfamily members have overlapping and distinct signaling machinery. The local concentration of ligands, inhibitors, receptors, and signaling molecules will create a unique situation in each cell type, with the potential to generate a multiplicity of signals depending on the physiological situation. This diagram illustrates the shared use of receptor subunits by activin and BMPs, while TGFβs use their own set of type I and II receptors and rely on betaglycan for signaling. Inside the cell, activation by activin and TGFβ recruits Smads 2 and 3, while Bmp-signaling recruits Smads 1, 5, and 8. These pathways converge with their requirement for binding the shared Smad 4 to achieve nuclear translocation required for affecting gene transcription, and they are all inhibited by Smads 6 and 7.

coexpressed in pachytene spermatocytes and round spermatids.

Immature Sertoli cells do not appear to have TGFβ receptors, and only the TGFβII receptor has been detected on adult Sertoli cells. However, these cells do make TGFβ1 and 2. Sertoli cell tight junction formation and synthesis of junction-associated proteins *in vitro* is affected by TGFβ3, with *TGFβ2* and *TGFβ3* mRNA levels decreasing as these junctions are being established and increasing thereafter. Secretion of TGFβ2 by Sertoli cells and peritubular cells may relate to how they interact to form the testicular cords *in vivo*, since the addition of exogenous TGFβ2 to cocultures, but not to monocultures, promotes formation of cellular aggregates. TGFβ1 can reduce growth of the embryonic and newborn rat testis, and TGFβ3 has been observed in the cells surrounding the seminferous cords of the early postnatal testis.

TGFβ receptors and ligands are also synthesized in Leydig cells. TGFβ3 immunoreactivity is reported in these cells as early as 16.5 dpc, with functional inhibition of LH-stimulated testosterone synthesis by TGFβ described at day 20.5 postpartum. The *in vitro* regulation of Leydig cell testosterone production by TGFβ1 is at least partially achieved by modulating the numbers of cell surface–LH receptors.

Mullerian inhibiting substance (MIS) At the time of transformation of the indifferent gonad into a testis (see section on Genes controlling testis formation), SRY triggers MIS production in Sertoli cells. Acting through the Bmpr1α receptor located on mesenchymal cells of

the Müllerian ducts, MIS is essential for regression of this duct that contains the precursors of the female uterus, fallopian tube, and upper vagina. At this time, the onset of testosterone secretion by Leydig cells induces differentiation of the Wolffian duct and male external genitalia.

MIS expression is confined to Sertoli cells, with much higher levels measured in fetal and neonatal testes than those observed after Sertoli cell division has ceased. An inactivating mutation in the murine *MIS* gene results in males with normal testis and male reproductive tract development. However, these animals are infertile owing to the presence of Müllerian duct derivatives that present a physical barrier to normal sperm efflux during coitus. This phenotype can be rescued by mating homozygous *MIS* null mice with those that overexpress the human *MIS* gene, thereby demonstrating the conservative structure of the encoded proteins. MIS-deficient females exhibit normal fertility. Foci of Leydig cell hyperplasia were observed in the testes of some (27%) adult MIS-deficient mice, but no other abnormalities were evident in testis weight and growth of seminal vesicles. From results such as these, a role for MIS in regulation of Leydig cell proliferation has been proposed.

Bone morphogenetic proteins Although there is currently limited information concerning the expression and cellular localization of the 15 BMP ligands and BMP receptor subunits, the impact of BMP signaling has been demonstrated at multiple stages in spermatogenesis. Deletion of murine *BMP4* results in the absence of primordial germ cells, as both Bmp4 and Bmp8 signals from the surrounding somatic environment are required for cells to be allocated to this lineage from the embryo epiblast. A potential role for Bmps2 and 4 in governing the fate of murine stem cells has been revealed through the use of *in vitro* treatment followed by transplantation into a recipient host. In contrast to glial-derived neurotropic factor, a distantly related TGFβ superfamily member, the addition of Bmp4 (or activin; see section on Activin, Inhibin, and Follistatin) reduces the stem cell potential of spermatogonia from cryptorchid animals when cultured on a feeder layer prior to transplantation. This suggests that Bmp signaling may provide a signal to differentiate at the onset of the spermatogenic pathway in the postnatal testis. Sertoli cell–derived Bmp4 has been identified as a candidate mediator of the developing mouse testis by upregulating c-kit at this time. Deletion of *Bmp8b* leads to male mouse infertility due to reduced germ cell proliferation at the onset of spermatogenesis and increased spermatocyte loss in the adult testis.

Activin, Inhibin, and Follistatin Inhibin and activin were first identified and purified on the basis of their role in feedback regulation of the testicular–hypothalamic–pituitary axis. The secretion of inhibin by Sertoli cells acts to reduce FSH secretion by the pituitary, while activin has the capacity to stimulate FSH production (Fig. 4). The signaling activity of activin is modified by the binding of other proteins, such as $\alpha_2$macroglobulin, a component of serum, and follistatin. Follistatin is produced within the testis by Sertoli cells under the influence of FSH, and the protein is present in spermatogonia, spermatocytes, and Leydig cells.

Documenting the pattern of activin and inhibin synthesis, the action within the testis is complicated by the fact that they are each dimers, which share a common subunit, with activin being a homo- or

heterodimer of two β subunits (βA, βB, βC, βD, and βE subunits have all been identified), while inhibin is composed of one α and one β subunit (either βA or βB). Inhibin α subunit expression has been detected in Sertoli and Leydig cells, while β subunits have been observed in Sertoli cells, Leydig cells, peritubular myoid cells, gonocytes (βA only), spermatocytes (βB), and round spermatids (βB). The adult Sertoli cell produces predominantly activin B (dimers of βB), while peritubular myoid cells synthesize activin A (βA dimers). *In vivo*, the ratio of α and βB subunits changes during the cycle of the seminiferous epithelium, leading to the suggestion that the relative proportions of inhibin and activin also vary along the length of the tubule. During Sertoli cell maturation, the level of α subunit mRNA synthesis is relatively constant during the proliferative and nonproliferative (maturational) phases, while the α subunit mRNA level decreases as the cells cease to proliferate and begin to mature.

Preferential binding of ^{125}I-labelled inhibin to rat Leydig cells has been reported, and at least two genes have been identified that encode specific inhibin receptor subunits, betaglycan and InhBP/p120. Inhibin, binding to its coreceptor betaglycan (TGFβ type III receptor; see section TGFβ.) blocks activin access to its type II receptor subunit. Betaglycan and InhBP/p120 are both present in Leydig cells. The addition of inhibin to stage dissected–tubule segments in culture results in a decrease in DNA synthesis, specifically in intermediate spermatogonia and in preleptotene spermatocytes, while the addition of activin increases DNA synthesis in the same cell types in this system. This data suggests that either there are other as-yet-unidentified inhibin receptors on germ cells, or that inhibin blocks activin signaling by direct competition for activin receptor binding sites. The effect of activin βA addition on immature rat Sertoli cells is to stimulate their proliferation during a discrete window of postnatal life. Activin addition to Sertoli cell cultures also regulates androgen receptor expression and FSH-induced aromatase activity.

Messenger RNAs encoding activin receptor subunits have been detected in Sertoli and germ cells at all stages of maturation, with stage-dependent peaks during the cycle of the seminiferous epithelium in the adult. Activin, like TGFβ1, inhibits proliferation of mouse primordial germ cells in culture. In contrast, activin increases gonocyte numbers in cultures of testis fragments taken from the newborn rat testis, while blocking their differentiation into spermatogonia at the onset of spermatogenesis. This differentiation was stimulated by the combined addition of FSH and follistatin. Other studies with mouse tissues have shown that activin, like BMP, reduces the stem cell potential of germ cells in culture. The addition of activin to cocultures of Sertoli and germ cells leads to increased spermatogonial proliferation and cellular aggregation. Hence activin may play a role in regulating the proliferation or differentiation of germ cells at many stages of development, further demonstrating the intricacy of regulation, which must contribute to achievement and maintenance of normal spermatogenesis. Messenger RNA encoding a membrane-bound inhibitor of TGFβ superfamily signaling, Bambi, is upregulated as germ cells transform into spermatogonia at the onset of spermatogenesis. In the adult, Bambi mRNA appears highly expressed in pachytene spermatocytes and round spermatids, making it another potential player in the complex set

of interactions that govern regulation of TGFβ signaling in the testis.

Another role of activin in the testis may be to maintain the immunosuppressed nature of the interstitium. Activin has been shown to suppress proliferation of peripheral T lymphocytes, an activity that would limit the influence of lymphocytes when presented with novel antigens such as haploid germ cell products, as might occur during testicular damage.

Our understanding of how inhibin and activin act are regulated within the testis is incomplete. In a genetically engineered mouse lacking the inhibin α subunit gene, no effect on testicular development or adult fertility was observed, suggesting that the physiological role of inhibin is either limited or at least partially met by another protein. However, the subsequent development of stromal cell tumors indicates that inhibin can function as a tumor suppressor (see Sect. 3.2.3). Disruption of the *βB subunit* gene has no affect on male fertility, but it has a profound influence on female fecundity manifested as perinatal lethality in offspring of homozygous females. The *βA*-/- mouse dies at birth, so the importance of the βA protein on postnatal development cannot be studied in these animals. Insertion of the *βB* gene into the *βA* locus has been performed to create mice with reduced activin bioactivity (activin A is reportedly 10-fold more bioactive than activin B in terms of signaling), and the animals survive to adulthood. The males are fertile but exhibit a delay in completion of the first wave of spermatogenesis. This further highlights the importance of activin A in the first wave of spermatogenesis, correlating with downregulation of this mRNA during the first week of postnatal life. *In vitro* studies and detailed mRNA analyses conducted on newborn rats has shown that the quiescent gonocytes synthesize activinA in the fetal testis, and the protein is stored until the cells transform into spermatogonia. At this time, the germ cells begin to synthesize the activin antagonists, follistatin and Bambi, indicating that the commencement of their differentiation requires a downregulation of activin signaling.

4 Conclusion

It is clear that a complicated set of hormones and locally produced factors determines the pathway of testis development and spermatogenesis. It is a challenge for the future to further develop model systems that will allow germ and somatic cell differentiation to proceed *in vitro* and to use the techniques that allow manipulation of gene expression in discrete cell types *in vivo*. Already the potential exists, with the production of immortalized somatic and germ cell lines which can support some degree of progression through meiosis. Advances in germ cell and organ transplantation have enabled sperm development in the testis of a host. We are just learning how to identify, store, culture, and stimulate the development of spermatogonial stem cells. This new information is being applied to better understand the source of the complex regulatory signals that guide spermatogenesis, to identify contraceptive targets, to understand the genetic basis of male infertility, and to restore fertility in genetically deficient individuals.

See also Endocrinology, Molecular; Female Reproduction System, Molecular Biology of.

Bibliography

Books and Reviews

de Rooij, D.G., Russell, L.D. (2000) All you wanted to know about spermatogonia but were afraid to ask, *J. Androl.* **21**, 776–798.

Eddy, E.M. (2002) Male germ cell gene expression, *Recent Prog. Horm. Res.* **57**, 103–128.

Handel, M.A. (Ed.) (1998) *Meiosis and Gametogenesis*, Academic Press, CA.

Russell, L.D., Ettlin, R.A., Sinha Hikim, A.P., Clegg, E.D. (1990) *Histological and Histopathological Evaluation of the Testis*. Cache River Press, Clearwater, FL.

Russell, L.D., Griswold, M.D. (Eds.) (1993) *The Sertoli Cell*, Cache River Press, Clearwater, FL.

Soder, O. (Ed.) (2003) *The Developing Testis. Physiology and Pathophysiology*, Karger, Basel.

Primary References and Specialist Reviews

Adams, I.R., McLaren, A. (2002) Sexually dimorphic development of mouse primordial germ cells: switching from oogenesis to spermatogenesis, *Development* **129**, 1155–1164.

Almiron, I., Chemes, H. (1988) Spermatogenic onset. FSH modulates mitotic activity of germ and Sertoli cells in immature rats, *J. Androl.* **11**, 235–246.

Beumer, T.L., Kiyokawa, H., Roepers-Gajadien, H.L., van den Bos, L.A., Lock, T.M., Gademan, I.S., Rutgers, D.H., Koff, A., de Rooij, D.G. (1999) Regulatory role of p27kip1 in the mouse and human testis, *Endocrinology* **140**, 1834–1840.

Beumer, T.L., Roepers-Gajadien, H.L., Gademan, I.S., Kal, H.B., de Rooij, D.G. (2000) Involvement of the D-type cyclins in germ cell proliferation and differentiation in the mouse, *Biol. Reprod.* **63**, 1893–1898.

Blanco-Rodriguez, J. (2001) Mitotic/meiotic Checkpoints and Germ Cell Apoptosis, *Proceedings of the VII International Congress of Andrology*, MEDIMOND, pp. 173–184.

Brennan, J., Tilmann, C., Capel, B. (2003) Pdgfr-alpha mediates testis cord organization and fetal Leydig cell development in the XY gonad, *Genes Dev.* **17**, 800–810.

Brinster, R.L., Avarbock, M.L. (1994) Germline transmission of donor haplotype following spermatogonial transplantation, *Proc. Natl. Acad. Sci.* **91**, 11303–11307.

Burns, K.H., Agno, J.E., Sicinski, P., Matzuk, M.M. (2003) Cyclin D2 and p27 are tissue-specific regulators of tumorigenesis in inhibin {alpha} knockout mice, *Mol. Endocrinol.* **17**, 2053–2069.

Clarke, T.R., Hoshiya, Y., Yi, S.E., Liu, X., Lyons, K.M., Donahoe, P.K. (2001) Mullerian inhibiting substance signaling uses a bone morphogenetic protein (BMP)-like pathway mediated by ALK2 and induces SMAD6 expression, *Mol. Endocrinol.* **15**, 946–959.

Clermont, Y., Perey, B. (1957) Quantitative study of the cell population of the seminiferous tubules in immature rats, *Am. J. Anat.* **100**, 241–267.

Colvin, J.S., Green, R.P., Schmahl, J., Capel, B., Ornitz, D.M. (2001) Male-to-female sex reversal in mice lacking fibroblast growth factor 9, *Cell* **104**, 875–889.

Cooke, H.J., Saunders, P.T. (2003) Mouse models of male infertility, *Nat. Rev. Genet.* **3**, 790–801.

Cooke, P.E., Zhao, Y.-D., Bunick, D. (1994) Triiodothyronine inhibits proliferation and stimulates differentiation of cultured neonatal Sertoli cells: possible mechanism for increased adult testis weight and sperm production induced by neonatal goitrogen treatment, *Biol. Reprod.* **51**, 1000–1005.

Crackower, M.A., Kolas, N.K., Noguchi, J., Sarao, R., Kikuchi, K., Kaneko, H., Kobayashi, E., Kawai, Y., Kozieradzki, I., Landers, R., Mo, R., Hui, C.C., Nieves, E., Cohen, P.E., Osborne, L.R., Wada, T., Kunieda, T., Moens, P.B., Penninger, J.M. (2003) Essential role of Fkbp6 in male fertility and homologous chromosome pairing in meiosis, *Science* **300**, 1291–1295.

Cupp, A.S., Kim, G.H., Skinner, M.K. (2000) Expression and action of neurotropin-3 and nerve growth factor in embryonic and early postnatal rat testis development, *Biol. Reprod.* **63**, 1617–1628.

Fragale, A., Puglisi, R., Morena, A.R., Stefanini, M., Boitani, C. (2001) Age-dependent activin receptor expression pinpoints activin A as a physiological regulator of rat Sertoli cell proliferation, *Mol. Hum. Reprod.* **7**, 1107–1114.

Gnessi, L., Fabbri, A., Spera, G. (1997) Gonadal peptides as mediators of development and functional control of the testis: an integrated system with hormones and local environment, *Endocr. Rev.* **18**, 541–609.

Habert, R., Lejeune, H., Saez, J.M. (2001) Origin, differentiation and regulation of fetal and adult Leydig cells, *Mol. Cell Endocrinol.* **179**, 47–74.

Handelsman, D.J., Spaliviero, J.A., Simpson, J.M., Allan, C.M., Singh, J. (1999) Spermatogenesis without gonadotropins: maintenance has a lower testosterone threshold than initiation, *Endocrinology* **140**, 3938–3946.

Haywood, M., Spaliviero, J., Jimemez, M., King, N.J., Handelsman, D.J., Allan, C.M. (2003) Sertoli and germ cell development in hypogonadal (hpg) mice expressing transgenic follicle-stimulating hormone alone or in combination with testosterone, *Endocrinology* **144**, 509–517.

Heckert, L.L., Griswold, M.D. (2002) The expression of the follicle-stimulating hormone receptor in spermatogenesis, *Recent Prog. Horm. Res.* **57**, 129–148.

Hsia, K.T., Millar, M.R., King, S., Selfridge, J., Redhead, N.J., Melton, D.W., Saunders, P.T. (2003) DNA repair gene Ercc1 is essential for normal spermatogenesis and oogenesis and for functional integrity of germ cell DNA in the mouse, *Development* **130**, 369–378.

Kashiwabara, S., Noguchi, J., Zhuang, T., Ohmura, K., Honda, A., Sugiura, S.A., Miyamoto, K., Takahasi, S., Inoue, K., Ogura, A., Baba, T. (2002) Regulation of spermatogenesis by testis-specific, cytoplasmic poly(A) polymerase TPAP, *Science* **298**, 1999–2002.

Kent, J., Wheatley, S.C., Andrews, J.E., Sinclair, A.H., Koopman, P. (1996) A male-specific role for SOX9 in vertebrate sex determination, *Development* **122**, 2813–2822.

Kumar, T.R., Varani, S., Wreford, N.G., Telfer, N.M., de Kretser, D.M., Matzuk, M.M. (2001) Male reproductive phenotypes in double mutant mice lacking both FSHbeta and activin receptor IIA, *Endocrinology* **142**, 3512–3518.

Kurihara, Y., Tokuriki, M., Myojin, R., Hori, T., Kuroiwa, A., Matsuda, Y., Sakurai, T., Kimura, M., Hecht, N.B., Uesugi, S. (2003) CPEB2, a novel putative translational regulator in mouse haploid germ cells, *Biol. Reprod.* **69**, 261–268.

Lee, J., Richburg, J., Shipp, E.B., Meistrich, M.L., Boekelheide, K. (1999) The Fas system, a regulator of testicular germ cell apoptosis, is differentially up-regulated in Sertoli cell versus germ cell injury of the testis, *Endocrinology* **140**, 852–858.

Lewis, K.A., Gray, P.C., Blount, A.L., MacConell, L.A., Wiater, E., Bilezikjian, L.M., Vale, W. (2000) Betaglycan binds inhibin and can mediate functional antagonism of activin signaling, *Nature* **404**, 411–414.

Liu, D., Liao, C., Wolgemuth, D.J. (2000) A role for cyclin A1 in the activation of MPF and G2-M transition during meiosis of male germ cells in mice, *Dev. Biol.* **224**, 388–400.

Livera, G., Rouiller-Fabre, V., Durand, P., Habert, R. (2000) Multiple effects of retinoids on the development of Sertoli, germ, and Leydig cells of fetal and neonatal rat testis in culture, *Biol. Reprod.* **62**, 1303–1314.

Martineau, J., Nordqvist, K., Tilmann, C., Lovell-Badge, R., Capel, B. (1997) Male-specific cell migration into the developing gonad, *Curr. Biol.* **7**, 958–968.

McCarrey, J.R. (1993) Development of the Germ Cell, in: Desjardins, C., Ewing, L.L. (Eds.), *Cell and Molecular Biology of the Testis*, New York Oxford University Press, pp 58–89.

McCoshen, J.A., McCallion, D.J. (1975) A study of the primordial germ cells during migratory phase in steel mutant mice, *Experientia* **31**, 589–590.

McGuinness, M.P., Orth, J.M. (1992) Reinitiation of gonocyte mitosis and movement of gonocytes to the basement membrane in testes of newborn rats in vivo and in vitro, *Anat. Rec.* **233**, 527–537.

McLachlan, R.I., O'Donnell, L., Meachem, S.J., Stanton, P.G., de Kretser, D.M., Pratis, K., Robertson, D.M. (2002) Identification of specific sites of hormonal regulation in spermatogenesis in rats, monkeys, and man, *Recent Prog. Horm. Res.* **57**, 149–179.

Meehan, T., Loveland, K.L., de Kretser, D., Cory, S., Print, C.G. (2001) Developmental regulation of the bcl-2 family during spermatogenesis: insights into the sterility of bcl-w-/- male mice, *Cell Death Differ.* **8**, 225–233.

Meehan, T., Schlatt, S., O'Bryan, M.K., de Kretser, D.M., Loveland, K.L. (2000) Regulation of germ cell and Sertoli cell development by activin, follistatin, and FSH, *Dev. Biol.* **220**, 225–237.

Meng, X., Lindahl, M., Hyvonon, M.E. Parvinen, M., de Rooij, D.G., Hess, M.W., Raatikainen-Ahokas, A. Sainio, K., Rauvala, H., Lakso, M.,

Pichel, J.G., Wstphal, H., Saarma, K., Sariola, H. (2000) Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. *Science* **287**, 1489–1493.

Meachem, S.J., Mclachlan, R.I., Stanton, P.G., Robertson, D.M., Wreford, N.G. (1999) FSH immunoneutralization acutely impairs spermatogonial development in normal adult rats, *J. Androl.* **20**, 756–762; discussion 755.

Morrish, B.C., Sinclair, A.H. (2002) Vertebrate sex determination: many means to an end, *Reproduction* **124**, 447–457.

O'Donnell, L., McLachlan, R.I., Wreford, N.G., de Kretser, D.M., Robertson, D.M. (1996) Testosterone withdrawal promotes stage-specific detachment of round spermatids from the rat seminiferous epithelium, *Biol. Reprod.* **55**, 895–901.

Orth, J.M., Gunsalus, G.L., Lamperti, A.A. (1988) Evidence from Sertoli cell-depleted rats indicates that spermatid number in adults depends on numbers of Sertoli cells produced during perinatal development, *Endocrinology* **122**, 787–794.

Orth, J.M., Jester, W.F., Li, L.H., Laslett, A.L. (2000) Gonocyte-Sertoli cell interactions during development of the neonatal rodent testis, *Curr. Top. Dev. Biol.* **50**, 103–124.

Orth, J.M., Qiu, J., Jester, W.F. Jr., Pilder, S. (1997) Expression of the c-kit gene is critical for migration of neonatal rat gonocytes in vitro, *Biol. Reprod.* **57**, 676–683.

Osborne, L.R., Wada, T., Kunieda, T., Moens, P.B., Penninger, J.M. (2003) Essential role of Fkbp6 in male fertility and homologous chromosome pairing in meiosis, *Science* **300**, 1291–1295.

Pellegrini, M., Grimaldi, P., Rossi, P., Geremia, R., Dolci, S. (2003) Developmental expression of BMP4/ALK3/SMAD5 signaling pathway in the mouse testis: a potential role of BMP4 in spermatogonia differentiation, *J. Cell Sci.* **116**, 3363–3372.

Print, C.G., Loveland, K.L. (2000) Germ cell suicide: new insights into apoptosis during spermatogenesis, *BioEssays* **22**, 423–430.

Ravnik, S.E., Wolgemuth, D.J. (1999) Regulation of meiosis during mammalian spermatogenesis: the A-type cyclins and their associated cyclin-dependent kinases are differentially expressed in the germ-cell lineage, *Dev. Biol.* **207**, 408–418.

Raymond, C.S., Murphy, M.W., O'Sullivan, M.G., Bardwell, V.J., Zarkower, D. (2000) Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation, *Genes Dev.* **14**, 2587–2595.

Rossi, P., Dolci, S., Albanesi, C., Grimaldi, P., Ricca, R., Geremia, R. (1993) Follicle-stimulating hormone induction of steel factor (SLF) mRNA in mouse Sertoli cells and stimulation of DNA synthesis in spermatogonia by soluble SLF, *Dev. Biol.* **155**, 68–74.

Rossi, P., Dolci, S., Sette, C., Geremia, R. (2003) Molecular mechanisms utilized by alternative c-kit gene products in the control of spermatogonial proliferation and sperm-mediated egg activation, *Andrologia* **35**, 71–78.

Russell, L.D., Peterson, R.N. (1984) Determination of the elongate spermatid-Sertoli cell ratio in various mammals, *J. Reprod. Fertil.* **70**, 635–641.

Rozen, S., Skaletsky, H., Marszalek, J.D., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., Page, D.C. (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes, *Nature* **423**, 873–876.

Schrans-Stassen, B.H.G.J., van de Kant, H.J.G., de Rooij, D.G., van Pelt, M.M. (1999) Differential expression of *c-kit* in mouse undifferentiated and differentiating Type A spermatogonia, *Endocrinology* **140**, 5894–5900.

Schrans-Stassen, B.H., Saunders, P.T., Cooke, H.J., de Rooij, D.G. (2001) Nature of the spermatogenic arrest in Dazl-/- mice, *Biol. Reprod.* **65**, 771–776.

Shuttlesworth, G.A., de Rooij, D.G., Huhtaniemi, I., Reissmann, T., Russell, L.D., Shetty, G., Wilson, G., Meistrich, M.L. (2000) Enhancement of A spermatogonial proliferation and differentiation in irradiated rats by gonadotropin-releasing hormone antagonist administration, *Endocrinology* **141**, 37–49.

Sicinski, P., Donaher, J.L., Geng, Y., Parker, S.B., Gardner, H., Park, M.Y., Robker, R.L., Richards, J.S., McGinnis, L.K., Biggers, J.D., Eppig, J.J., Bronson, R.T., Elledge, S.J., Weinberg, R.A. (1996) Cyclin D2 is an FSH-responsive gene involved in gonadal cell proliferation and oncogenesis, *Nature* **384**, 470–474.

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovell-Badge, R., Goodfellow, P.N. (1990) A gene from the human sex-determining region encodes a protein with

homology to a conserved DNA-binding motif, *Nature* **346**, 240–244.

Spruck, C.H., de Miguel, M.P., Smith, A.P., Ryan, A., Stein, P., Schultz, R.M., Lincoln, A.J., Donovan, P.J., Reed, S.I. (2003) Requirement of Cks2 for the first metaphase/anaphase transition of mammalian meiosis, *Science* **300**, 647–650.

Tadokoro, Y., Yomogida, K., Hiroshi, O., Tohda, A., Nishimune, Y. (2002) Homeostatic regulation of germinal stem cell proliferation by the GDNF/FSH pathway, *Mech. Dev.* **113**, 29–39.

Tsutsui, T., Hesabi, B., Moons, D.S., Pandolfi, P.P., Hansel, K.S., Koff, A., Kiyokawa, H. (1999) Targeted disruption of CDK4 delays cell cycle entry with enhanced p27(Kip1) activity, *Mol. Cell Biol.* **19**, 7011–7019.

Venables, J.P., Elliott, D.J., Makarova, O.V., Makarov, E.M., Cooke, H.J., Eperon, I.C. (2000) RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2beta and affect splicing, *Hum. Mol. Genet.* **9**, 685–694.

Vergouwen, R.P., Jacobs, S.G., Huiskamp, R., Davids, J.A., de Rooij, D.G. (1991) Proliferative activity of gonocytes, Sertoli cells and interstitial cells during testicular development in mice, *J. Reprod. Fertil.* **93**, 233–243.

Vidal, V.P., Chaboissier, M.C., de Rooij, D.G., Schedl, A. (2001) Sox9 induces testis development in XX transgenic mice, *Nat. Genet.* **28**, 216–217.

Wang, P.J., McCarrey, J.R., Yang, F., Page, D.C. (2001) An abundance of X-linked genes expressed in spermatogonia, *Nat. Genet.* **27**, 422–426.

Wu, X.Q., Hecht, N.B. (2000) Mouse testis brain ribonucleic acid-binding protein/translin colocalizes with microtubules and is immunoprecipitated with messenger ribonucleic acids encoding myelin basic protein, alpha calmodulin kinase II, and protamines 1 and 2, *Biol. Reprod.* **62**, 720–725.

Wykes, S.M., Krawetz, S.A. (2003) The structural organization of sperm chromatin, *J. Biol. Chem.* **278**, 29471–29477.

Yan, W., Kero, J., Suominen, J., Toppari, J. (2001) Differential expression and regulation of the retinoblastoma family of proteins during testicular development and spermatogenesis: roles in the control of germ cell proliferation, differentiation and apoptosis, *Oncogene* **20**, 1343–1356.

Yan, W., Suominen, J., Samson, M., Jegou, B., Toppari, J. (2000) Involvement of Bcl-2 family proteins in germ cell apoptosis during testicular development in the rat and pro-survival effect of stem cell factor on germ cells in vitro, *Mol. Cell Endocrinol.* **165**, 115–129.

Yang, J., Chennathukuzhi, V., Miki, K., O'Brien, D.A., Hecht, N.B. (2003) Mouse testis brain RNA-binding protein/Translin selectively binds to the messenger RNA of the fibrous sheath protein glyceraldehyde 3-phosphate dehydrogenase-S and suppresses its translation in vitro, *Biol. Reprod.* **68**, 853–859.

Zhang, F.P., Poutanen, M. (2001) Normal prenatal but arrested postnatal sexual development of luteinizing hormone receptor knockout (LuRKO) mice, *Mol. Endocrinol.* **15**, 172–183.

Zhao, G.Q., Liaw, L., Hogan, B.L.M. (1996) Bone morphogenetic protein 8A plays a role in the maintenance of spermatogenesis and the integrity of the epididymis, *Development* **125**, 1103–1112.

Zhao, M., Shirley, C.R., Yu, Y.E., Mohapatra, B., Zhang, Y., Unni, E., Deng, J.M., Arango, N.A., Terry, N.H., Weil, M.M., Russell, L.D., Behringer, R.R., Meistrich, M.L. (2001) Targeted disruption of the transition protein 2 gene affects sperm chromatin structure and reduces fertility in mice, *Mol. Cell Biol.* **21**, 7243–7255.

Zhou, Q., Nie, R., Prins, G.S., Saunders, P.T., Katzenellenbogen, B.S., Hess, R.A. (2003) Localization of androgen and estrogen receptors in adult male mouse reproductive tract, *J. Androl.* **23**, 870–881.

Zsebo, K.M., Williams, D.A., Geissler, E.N., Broudy, V.C., Martin, F.H., Atkins, H.L., Hsu, R.Y., Birkett, N.C., Okino, K.H., Murdock, D.C. Jacobsen, F.W., Langley, K.E., Smith, K.A., Takeishi, T., Cattanach, B.M., Galli, S.J., Suggs, S.V. (1990) Stem cell factor is encoded at the *Sl* locus of the mouse and is the ligand for the *c-kit* tyrosine kinase receptor, *Cell* **63**, 213–224.

Mammalian Cell Culture Methods

Dieter F. Hülser
University of Stuttgart, Stuttgart, Germany

1 Principles 537
1.1 Development of Cultured Cells 537
1.2 Selection of Cells 538
1.3 Cell Cycle and Growth Curves 539
1.4 Cell Profiling 540
1.5 Ethical Rules 541

2 Techniques 541
2.1 Preparation of Primary and Continuous Cultures 541
2.1.1 Monolayer Cultures 543
2.1.2 Suspension Cultures 543
2.1.3 Three-dimensionally Growing Cultures 544
2.2 Cell Culture Media 544
2.3 Equipment 546
2.3.1 Vessels 546
2.3.2 Appliances 548
2.4 Safety and Biohazards 548
2.5 Storage 549

3 Applications 549
3.1 Cellular Regulations 550
3.2 Cancer Research 550
3.3 Production 551

4 Perspectives 552

Bibliography 552
Books and Reviews 552
Primary Literature 553

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

Keywords

Cell Cloning
Generation of a colony from a single cell. Subculturing results in a cell strain.

Cell Cycle
Ordered sequences (G_1, S, G_2, and M) of cellular syntheses between two cell divisions.

Cell Line
Subcultured primary cultures. A cell line can be finite or continuous.

Cell Strain
Cell line that has been purified by physical separation, selection, or cell cloning.

Continuous Cell Line
Indefinite proliferation. This immortalization of a cell line may be induced by a viral gene transfer or was already acquired by some cancer cells before cultivation.

Primary Culture
Freshly isolated cells in culture until the first passage into a subculture.

Serum
Blood fluid without cells and clotting factors.

Suspension Culture
Cells proliferate isolated from each other when suspended in growth medium.

Tissue Culture
Accustomed term for the cultivation of animal cells. Originally: Fragments of tissues maintained *in vitro*.

Mammalian cell cultures originate from tissue explants or cell suspensions as primary cell cultures that can be subcultured with a limited life span. By transformation, these cells might lose some of their original properties and establish permanent growth. Many of these continuous cell lines are aneuploid and genetically unstable, nevertheless, our knowledge of molecular, physiological, biochemical, and biophysical properties of cells is notably based on investigations with such cell lines. Since the synthesis of various bioproducts such as vaccines, monoclonal antibodies, enzymes, and hormones is accomplished with cell cultures, many efforts were made to develop and improve cell culture technology. Besides this impetus, cancer research also stimulated the progress of cell culture methods as can be seen with the development of three-dimensionally growing cultures. These cultures provided a better understanding of tumor invasion and revealed the importance of the extracellular matrix for physiological regulations of cell–cell interactions that

cannot be observed with monolayer or suspension cultures. This knowledge helped to improve the cultivation of cells that are used in clinical treatments as is the case for wound healing with implantation of epidermis or for defect organs such as liver, which can be supported with bioartificial organs during a temporary extracorporal bypass.

1 Principles

1.1 Development of Cultured Cells

Animals are three-dimensionally organized complex multicellular creations. Their tissues and organs maintain specific internal milieus and are separated from each other and from their environment by specialized endothelial and epithelial cells. The adult human body consists of hundreds of cell types and altogether of approximately 50 to 100 $\times$ 10^{12} cells from which about 10^9 cells vanish within one hour. This sounds more alarming than it really is because most of these cells are replaced within the same time. The highest turnover is found for blood and epithelial cells; after wounding, organs and tissues might start proliferation albeit they are normally almost quiescent. Even when most cells have the potential for proliferation, it is not an easy task to cultivate them *in vitro*, that is, in Petri dishes, flasks, or bioreactors.

A stimulus to study isolated cells was given in 1858 when Rudolf Virchow postulated that pathological characteristics may be detected on a cellular level. First experiments to cultivate cells were performed with non-mammalian cells such as amphibians with their high capacity to regenerate lost limbs and with chicken embryos that are easily accessed and proliferate with a high activity. Successful attempts to maintain cells *in vitro* date back to 1885 when Wilhelm Roux kept isolated nerve fibers of chicken embryos alive in a warm saltbroth. Much effort was put in the establishment of an appropriate medium in which the cells could survive and even proliferate. In 1895, Paul Ehrlich succeeded to grow and propagate mouse tumor cells by intraperitoneal injection in mice where they grew as single-cell suspension in ascites. This *in vivo* cultivation revealed that the liquid of the peritoneum obviously contains all nutrients necessary for cell proliferation. Consequently, human ascites was used to keep pieces of human skin alive *in vitro*, as was demonstrated in 1898 by Ljunggren who grafted them back to the donors, including himself. A significant step forward was achieved in 1907 by Ross Harrison, who cultivated spinal cords of frogs in a hanging drop of coagulated frog's lymph and studied the outgrowth of nerve fibers during several weeks. In 1912, Carrel reported on the permanent life of tissues outside the organism when he cultivated embryonic chicken cells in an extract of chicken embryos. He introduced aseptic methods and cultivated the cells as monolayers in glass flasks, but that he has kept these cells for more than 30 years is certainly not true since *in vitro* proliferation of normal cells is limited to about 50 divisions, which became known

later by the investigations of Hayflick and Moorhead. Carrel and coworkers had regularly fed their cultures with insufficiently filtered extracts of chicken embryos, which contained fresh cells.

Different media were adopted for specific cells, which grew as monolayers on different substrata such as glass, polypropylene, ceramics, and so on. Earle and coworkers treated mouse fibroblasts with methylcholanthrene and cultured these transformed cells in medium with a bicarbonate/CO_2 buffered salt solution. These L-cells have been growing as a continuous cell line since 1943, and at present their use might only be outnumbered by the first continuous human cell line HeLa, which was derived from a human cervical carcinoma by Gey and coworkers in 1952 (Fig. 1). Modified medium formulations were used by Dulbecco to grow animal viruses in cultured cells and Eagle described in 1955 the minimum requirements of nutrients for appropriate synthetic media, which, however, still required the addition of serum from young animals. Another important step was made by Levi-Montalcini and colleagues who isolated of a protein, which stimulates the growth of certain nerve cells. Similarly, to this nerve growth factor, other growth factors were later isolated from serum (see Sect. 2.2). With better-defined media, more complex cells could be cultivated as was the case with the successful development of a hybridoma cell line for the production of monoclonal antibodies by Köhler and Milstein.

Even from the very beginning of tissue culture, three-dimensional explants were always used, and most experiments are still performed with monolayer and suspension cultures. This might change in the near future since many regulatory processes are better studied in three-dimensionally growing cells that perfectly match the *in vivo* conditions.

1.2 Selection of Cells

Most animal cells are cultivated as monolayers since the majority of cells is anchorage dependent and not very selective with regard to a substratum. This type of cell cultivation offers advantages for microscopic inspections during growth or during experiments and it allows micromanipulation of the cells with electrodes. Normal cells in a primary culture usually stop proliferation by a so-called contact inhibition when they come in close contact on a flat surface. By serum deprivation, the proliferation of these cells can also be inhibited and they remain quiescent in the G_1- or G_0-phase

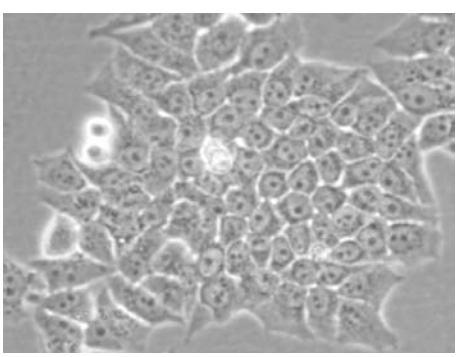

Fig. 1 Monolayer of HeLa cells during logarithmic growth. Diameter of a cell is about 15 µm. Light-microscope phase contrast picture.

of their cell cycle. After addition of serum to these cultures, the cells reenter the cell cycle and progress in a synchronized manner. Under these conditions, signals from individual cells are multiplied since effects occur synchronously in all cells as has been demonstrated for growth factor–induced channel openings in embryonic rat fibroblasts. In spite of the high variance of samples, cells of primary cultures are recommended when specialized functions should be investigated under controlled culture conditions.

Cultures with selected properties are available with transformed cells. They are not contact inhibited and continue to grow as long as the medium is not depleted of nutrients. Numerous investigations in cell biology as well as in cancer research were performed with these continuous cell lines and cell strains, and much of our knowledge on cellular regulation, synthesis, and proliferation is based on these cells. For the production of cellular material, cells were selected after transformation, hybridization, or transfection. In catalogs of the American Type Culture Collection (ATCC), of the European Collection of Cell Cultures (ECACC), or of other national culture collections, appropriate cells for many purposes can be found.

Since production is usually optimized for a high yield, cells were further selected for anchorage-independent growth, which qualifies them to proliferate and to synthesize the requested product in single-cell suspension in large volumes. However, as is evident from intact organisms, producer cells must not necessarily be in a proliferative status. This understanding finally led to cultivation methods that allow the cells to grow as three-dimensional cell aggregates. This might end in organ-like cultures where the cells stop to proliferate but still synthesize substances or function as bioartificial organs as is the case for extracorporal cultivated liver cells.

1.3 Cell Cycle and Growth Curves

Under optimal conditions in culture, cells may divide every 10 to 20 hours except when cells may be stressed by transferring from one culture vessel into the next or by exhausted medium. Between two cell divisions, a cell must synthesize all its materials such as DNA, RNA, proteins, lipids, and so on. These processes could occur continuously or in discrete phases. As has been demonstrated by Quastler and Sherman, DNA is replicated in a separate phase that starts a few hours after division and ends a few hours before the next division. This DNA synthesizing period is called *S-phase*, it is preceded by the G_1- and followed by the G_2-phase, whereas the M-phase covers the mitosis and the cell segregation between the two G-phases. Similarly, as in an adult organism, cells may stop proliferation and rest in the G_0-phase.

When cells are cultivated as suspension culture, they exist as isolated single cells and can easily be harvested without enzymatic treatment and transferred directly into a new culture vessel. Under these conditions, cells may immediately continue to proliferate at a high rate. During mitosis, one cell divides into two cells and, therefore, the number of cells increases to $N + 1$. Since one "old" cell disappears after its doubling time and two "young" cells are instead added to the culture, we must distinguish between growth rate $G = (\mathrm{d}N/\mathrm{d}t)$, which indicates how many cells are added to a culture during a certain time period, and the birth rate

$B = 2G$, which gives the number of cells that are "born" at a certain time point. A culture of continuously proliferating cells has, therefore, twice as much young cells as old cells and its age distribution declines exponentially.

Under these conditions, an exponential growth is observed, which is described by

$$N = N_0 \cdot e^{\mu \cdot t} \quad (1)$$

with N: number of cells; N_0: number of cells at the start of the culture; growth parameter $\mu = ln2/t_d$; t: time; t_d: doubling time of cells. Every cell has an individual cell age τ that starts after mitosis at $\tau = 0$ and ends at $\tau = t_d$ with the next division. Under optimal conditions, the individual doubling time of a cell could be identical with the population doubling time. However, during growth, the medium will gradually be exhausted and some cells may stop proliferating or are prolonged in their cell cycle because one or more restriction points cannot be passed. Furthermore, when monolayer cells are transferred from one vessel into the next by trypsinization and mechanical treatment, they might additionally be stressed by temperature shifts and their proliferation in the new vessel starts delayed. These delays result in characteristic growth curves as is shown in Fig. 2 for BICR/M1R$_k$ cells. For fitting these points with a curve, the growth parameter μ must be modified to $\mu = \mu_1 - \mu_2 N$ and leads to the Verhulst–Pearl equation:

$$N = \frac{\mu_1 \cdot N_0 \cdot e^{\mu_1 \cdot t}}{\mu_1 + \mu_2 \cdot N_0 (e^{\mu_1 \cdot t} - 1)} \quad (2)$$

This equation represents realistic growth conditions in which the proliferation also depends on the number of cells and the available amount of nutrients in a medium. A typical growth curve starts with a lag-phase in which the number of cells is rather constant before the cells start with exponential growth and enter the log-phase. For practical reasons, the ordinate of a growth curve is logarithmically divided and, therefore, the log-phase is characterized by a straight line. When the medium is deprived of nutrients, cells stop proliferating and for a while the number of cells is constant – the culture is in its stationary phase before cells start to die away.

1.4 Cell Profiling

Altogether, several ten thousands of cell lines may be kept worldwide; therefore, a reliable identification of the cells becomes mandatory. Besides microbial contamination and phenotypic drift, cross-contamination between cells occurs more

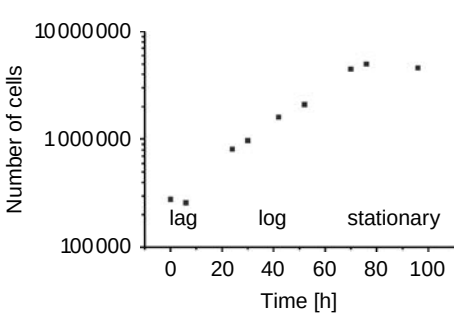

Fig. 2 Growth curve of BICR/M1R$_k$ cells with a typical lag-, log-, and stationary-phase. This continuous cell line originates from a mammary carcinoma of the Marshall rat.

often than it is detected and leads to severe misidentifications of cell cultures. Within 10 years of the start in culture of the human cancer cells HeLa, other human and animal cell lines were contaminated with these fast-growing HeLa cells as was detected by genetic markers. These alarming reports did not stop this carelessness and in 1976, Nelson-Rees and Flandermeyer reported more than 90 cell lines that were contaminated with HeLa cells. It is still estimated that up to one-third of all cell lines is of different origin or species to that claimed. These cross-contaminations result from careless handling as might be the case when several cell lines are transferred simultaneously and medium is added to all cultures with the same pipette. Aerosols and small droplets with cells might adhere to the outside of the pipette and are thus transferred into the stock medium from where they can contaminate other cultures. Cross-contamination could be detected by experienced molecular biology laboratories with several time-consuming and expensive methods, but none was suitable for use as an international reference standard. With the development of new PCR-based forensic techniques, an inexpensive method, the short tandem repeat (STR) profiling, has been applied to DNA from human cell lines. STR generates standard numerical codes for lengths of polymorphic DNA loci, which qualify as universal reference standard for human cell lines. Masters and colleagues recommend DNA profiling of cell lines as a normal practice and suggest this analysis as a prerequisite for publication. This DNA fingerprint will not only help to identify cell lines and greatly enhance the confidence in these studies but should also allow comparisons of published results on a more solid basis as momentarily possible.

1.5 Ethical Rules

Working with continuous cell cultures does not seem to cause severe ethical problems; however, when primary cells are acquired from humans and animals, several critical implications arise. For clinical treatments as well as for research, the use of human stem cells from embryos, fetuses, or adults is strictly regulated by national and international legislation. However, research with other human cellular material must also be justified and can only be approved by local research ethical committees when the demands of national guidelines are enforced. Similarly, the use of animal tissue requires ethical and legal approval, often by local animal welfare committees. Animal welfare aspects include the minimizing of pain, suffering, distress, and lasting harm as well as the restriction to the absolutely necessary number of animals. Experiments with cultures of permanently growing cells do not exclude animal welfare considerations since most cell lines are still maintained with purchased fetal bovine or newborn calf serum. Even when little or no influence can be taken on the collecting process, there are also technical arguments for considering a replacement of serum (see Sect. 2.2).

2 Techniques

2.1 Preparation of Primary and Continuous Cultures

A primary culture starts with biopsies from solid tissues or organs, preferentially from embryos, but tumor cells are

also often selected because of their high proliferation capacity. These firmly attached cells have to be disaggregated to single cells, a procedure that combines mechanical dissociation and enzymatic detachment. After chopping the tissue with fine surgical scalpels, proteolytic enzymes such as trypsin, collagenase, or pronase are applied to open proteinaceous bonds between cells, a process that might be accelerated by the appropriate temperature and thorough pipetting. This cell suspension will be diluted with medium plus serum; centrifuged and resuspended in medium for several times before the cells are transferred into Petri dishes or flasks. Most cells are anchorage dependent and will adhere to an appropriate substratum after a few hours. At that time point, the cultures must be washed free of all cellular debris and damaged cells, since only a small portion of cells may have survived the isolation procedure. Microscopic inspection will then reveal the yield of attached single cells, but also aggregates of cells will be found from which an outgrowth of cells can be observed after a couple of days. Usually, a mixture of cell types will be present in these primary cultures, the majority being fibroblasts that will proliferate better than other cell types. Since they are also motile, they can grow out from cell aggregates and spread over the substratum. Epithelial cells, on the other hand, are more or less immobile and tend to form patches on the substratum. Figure 3 shows a primary culture of embryonic mouse brain cells in which nerve cells have spread on top of other cells that are out of focus.

Adherent cells can be detached by trypsin, and enough cells might be harvested from these primary cultures to proceed with their propagation in multiple dishes or flasks. These subcultures can be used to multiply the selected cells and generate a new cell line that contains all the different cells of the primary culture. If these cells are selected from normal tissues, these cultures have a limited life span and must, therefore, be multiplied by several cultivation passages, collected, and stored in liquid nitrogen so that planned experiments can always rely on identical cell populations (see Sect. 2.5). Critical cells may not proliferate, especially when the density of transferred cells is too low. This problem can sometimes be solved either with conditioned medium or with feeder layers. The so-called conditioned medium is taken from a culture of logarithmically growing cells and added to freshly propagated cells. This medium contains cell specific factors that mimic the presence of numerous cells and stimulate the growth of critical cells. A similar effect can be obtained with irradiated feeder layer cells, which form a monolayer and serve as substratum for critical cells.

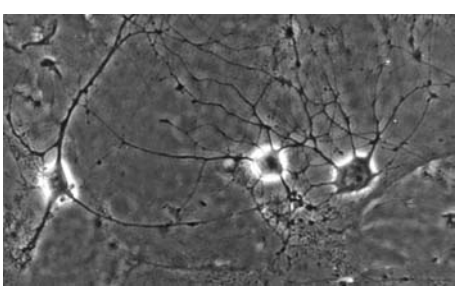

Fig. 3 Primary culture of embryonic mouse brain cells. Nerve cells have spread on top of other cells, which are out of focus. Light-microscope phase contrast picture.

By transforming and cloning cell lines, permanently growing cell strains with specific properties can be derived, and many of them are available from cell banks. A majority of experiments is, therefore, performed with these continuous cells.

2.1.1 Monolayer Cultures

Plasma membranes of animal cells carry fixed negative charges that allow cell–cell adhesion by divalent cations (Ca^{++}, Mg^{++}), but cells may also adhere to artificial surfaces that are either negatively or positively charged. Cells form a monolayer when they adhere not only to each other but also to the surface of culture vessels. Different cells form differently arranged monolayers: epithelial cells (Fig. 1) show a typical cobble stone appearance, whereas transformed fibroblasts (Fig. 4) may grow criss-cross and pile up when a high density is reached. The cells' morphology might indicate that the medium is exhausted and its replacement or a subculture is necessary. For propagating the cells, the medium is withdrawn and trypsin is added with a Ca^{++}/Mg^{++} free balanced salt solution. The time and temperature required for this treatment depend on the sensitivity of the cells, and other enzymes might also be required for a gentle disaggregation. In any case, the enzymatic treatment should be as short as possible to avoid unnecessary stress. Therefore, the cells must be washed with medium to block the activity of trypsin by its serum. The cells can then be detached from the surface by repetitive pipetting of medium over the monolayer, a procedure that also disperses floating cell aggregates. Portions of this resulting single-cell suspension are used for subcultures, and an aliquot might be used for determining the concentration of cells. The number of cells that are necessary for starting a new culture depends on the cell type, on the kind of planned experiments, and on the type of culture vessels. A few hours after starting the incubation of new cultures, they can be microscopically inspected to exclude any damage that might have occurred during the passage.

2.1.2 Suspension Cultures

Several monolayer cell lines have lost their anchorage dependence by transformation and are now also kept as suspension cultures. HeLa cells are amongst these cultures, which grow as monolayers and in suspension. Primary suspension cultures can be obtained from normal lymphocytes, which are anchorage independent. They have a limited life span, but lymphoblastoid cells with unlimited growth in culture are also available as are other ascites tumor cells that have lost their anchorage dependence already *in vivo*.

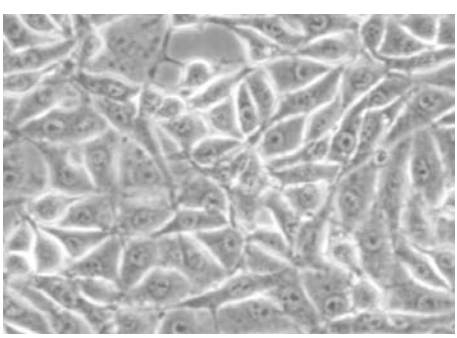

Fig. 4 Monolayer of BICR/M1R_k cells during logarithmic growth. Light-microscope phase contrast picture.

2.1.3 Three-dimensionally Growing Cultures

Not only may primary explants grow three-dimensionally but also transformed cells can be cultured as so-called multicellular spheroids in suspension. For a few anchorage-dependent cell lines, it is sufficient to start a culture of these cells in a stirrer flask for suspension cultures. When cells meet each other in this suspension, they will stay together and thus form larger spheroidal aggregates within a couple of weeks. Most other cell lines must first be kept in dishes with a nonadhesive surface to induce this three-dimensional growth. These Petri dishes are commercially available and are mainly used for cultivating bacteria on agar. After a few days of cultivation in these dishes, the cells have formed irregular aggregates, which are transferred into stirrer flasks in which they form multicellular spheroids and may be kept and treated for several weeks (Fig. 5). Spheroids can be harvested directly with pipettes, and medium is easily replaced since the spheroids precipitate quickly when they are no longer stirred. When such a spheroid is placed on a culture dish, the cells quickly adhere to this charged surface and within a day many cells will have migrated from the spheroid and proliferate again as monolayer cells.

When biopsies of organs are cultured in stirrer flasks, they will maintain morphological and physiological properties, and many cells remain fully differentiated and proliferate sparsely. Even when fresh explants are required for new experiments, these cultures offer some advantages for complex experiments that might, for instance, include immunological reactions as has been shown with lymphoid tissue and HIV infection.

Epithelium, which separates an organism from its environment or an organ from its surrounding fluid, is formed by layers of different cells. When cultured *in vitro*, epithelial cells often form flat monolayers of different cells but epidermal cells may be kept as organotypic flat multilayers in a liquid–gas interphase. This *in vitro* skin is another example for the many types of three-dimensionally growing cultures that are not only used to study tissue regeneration but are also applied in biomedical treatments.

2.2 Cell Culture Media

The basic components of cell culture media are inorganic salts, glucose, and organic substances, which are also present in blood plasma of animals. Their concentration varies with different compositions of

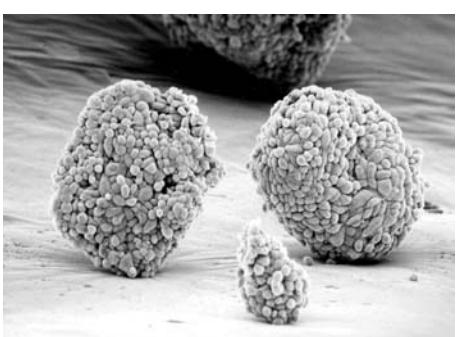

Fig. 5 Multicell spheroids of BICR/M1R_k cells. The diameter of the larger spheroids is about 350 µm. Scanning electron microscope picture.

media that were designed for specific cells with special properties. Most of these media are commercially available as solutions and also as powder that has to be dissolved in water. The selection of water, however, is not trivial since it may contain organic and inorganic material, which will not be removed by simple distillation so that other purification processes must be included. Differences in cell activity may be due to different water qualities; therefore, the use of ultrapure water is recommended. In any case, powdered media require sterilization by filtration through a pore size of 0.22 μm, preferably during bottling and before storage in a refrigerator.

All basic salt broths of cell culture media contain NaCl, KCl, $CaCl_2$, and $MgCl_2$. The concentration of these inorganic salts is set to adjust the osmotic balance of the cells and maintain their membrane potential. When this solution is buffered to a pH value of 7.2 to 7.4 with the pH buffering phosphates NaH_2PO_4 and Na_2HPO_4, it can be used to keep the cells alive for a short while. A complete medium requires much more additives for the functioning of inorganic salts as mediators for cell attachment or as enzyme cofactors. Carbohydrates (glucose) are added as energy source, and amino acids serve as nitrogen sources, although glutamine can also provide a carbon source via transamination. Essential and nonessential amino acids and water-soluble vitamins are also necessary as nutrients and additional energy sources; their concentrations vary with different medium formulations.

For proliferating mammalian cells, the buffering capacity of the medium is increased by bicarbonate ($NaHCO_3$). In this case, the medium's CO_2/HCO_3 content must be balanced by gaseous CO_2, which requires gassed incubators in which an atmosphere of 5–10% CO_2 in air can be maintained. Another buffering system is given by the zwitterion HEPES, which is often used together with the "natural" bicarbonate buffering system. With this buffer, cells can be manipulated under atmospheric conditions without drastic pH changes. For a rapid and easy control of the medium's pH, phenol red is added as indicator.

Many laboratories supplement their medium with antibiotics such as streptomycin sulfate and penicillin G. This reduces the risk of bacterial contamination but increases the development of resistant organisms and encourages the use of inadequate aseptic techniques. Transfected cells, however, may have been selected as resistant to puromycin or geneticin, and the use of these antibiotics is recommended for maintaining a selection pressure.

A very important component of a cell culture medium is serum, and in most cases newborn calf or fetal bovine serum is added. Serum is a complex mixture that contains not only essential growth factors but also plasma proteins, hormones, metabolites, and growth inhibitors. Since serum components are affected by age, health, and nutrition of the donor animals, the quality of serum will change from batch to batch and, therefore, every new batch has to be tested for the required quality standards. When bovine spongiform encephalopathy (BSE) spread in cattle, guidelines became effective, which should minimize the risk of BSE transmission via medicinal products and the necessity for media without any material from animal origin has increased.

At the beginning, the so-called serum-free media contained factors that were isolated from serum and stimulated the

growth of certain cell types, which can be seen from their name: fibroblast growth factor (FGF), nerve growth factor (NGF), epithelial growth factor (EGF), platelet-derived growth factor (PDGF), and insulin-like growth factor (IGF). Other serum-isolated additives include interferons, hormones, and attachment factors. With these components, a serum-free medium was still an undefined medium since it contained components that were subject to variation. Effects to develop defined media where the chemical structure and the concentration of every component are known were successful. These media are consistent from batch to batch and are optimized for growth of specific cell types and product synthesis. Almost every cell line requires its own expensive serum-free medium, its use is therefore still limited to productions with expensive purification processes and to some biomedical treatments.

2.3 Equipment

2.3.1 Vessels

Monolayer cells need a substratum to adhere, and this material must be of good optical quality since microscopic inspection of the cells is mandatory for most experiments, but other material might be advantageous when it comes to production. Cells preferentially adhere to negatively charged surfaces and, therefore, the plastic vessels for cell culturing are charged during the manufacturing process. Many laboratories use transparent disposable multiwell plates, Petri dishes, and T-flasks made of polystyrene. Their volume varies between 0.1 and 250 mL medium and with the corresponding surface areas. However, other mechanisms are also effective in cell adhesion. When cells are cultivated on glass surfaces, they might not adhere well when the glass flasks are first in use. After several passages of cells, the adhesion is enhanced, indicating a conditioning of these glass surfaces. This is due to specific receptors for cellular adhesion, which are part of the extracellular matrix of plasma membranes. They partly remain adherent to the glass surfaces when the cells move or are harvested. Even after cleaning and sterilizing, material is firmly connected with the glass and thus facilitates the attachment of cells of sequential cultures. This finally led to the use of special mixtures of extracellular matrix glycoproteins and proteoglycans by which any substratum could be made available for the cultivation of critical cells. Continuous cell lines such as HeLa adhere to different substrata, for example, carbon-coated glass (Fig. 6) and polished stainless steel (Fig. 7).

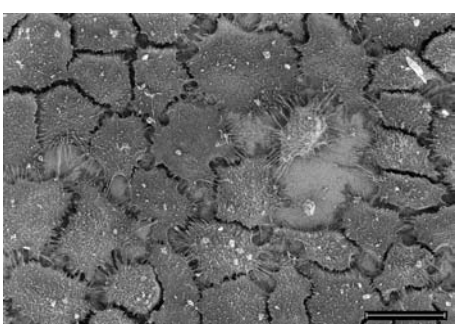

Fig. 6 Monolayer of HeLa cells during logarithmic growth on carbon- coated glass. Attached particles originate from serum and/or dead cells' debris. Scanning electron microscope picture, bar = 30 µm.

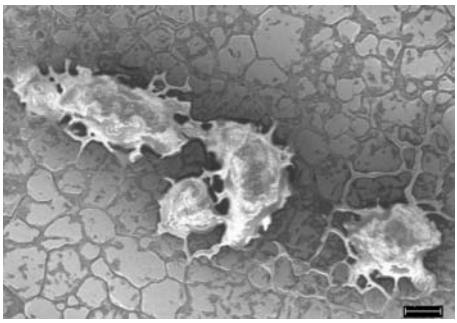

Fig. 7 HeLa cells grown on stainless steel. Note the amoeboid protrusions onto the nonpolished gaps. Scanning electron microscope picture, bar = 10 μm.

Many sophisticated systems have been developed to improve the yield of monolayer cells. Since the area of the substratum on which the cells can grow must be increased and the volume of the vessels must be expanded, risks of contamination and problems of regulation will also simultaneously rise. These tasks are solved with different strategies. A classical technique is the use of cylindrical bottles that are slowly rotated so that the total inner glass or plastic surface is available for cultivating monolayers with a small amount of medium. Many of these roller culture bottles can be rotated simultaneously on racks and thus increase the total amount of products. Other mass culture techniques make use of separated compartments in which the cells grow on permeable membranes and medium is provided in a compartment separate from the products, a strategy that reduces the cost of product purification. An example is the hollow fiber, which is not only used for adherent but also for suspended cells. The surface of substratum can also be increased when the cells are grown on microcarriers. These small beads with diameters of about 200 μm are fabricated from different materials, such as polystyrene, silica, glass, or dextran and may be coated with collagen or mixtures of cell adhesive substances. Cells can grow as monolayers on these beads, which are kept as suspension and thus mediate mass cultivation of monolayer cells.

Cells can also grow isolated in suspension cultures in which they must be kept floating with a minimum of shear stress and optimal conditions for growth and production. For small volumes, this is achieved with stirrer flasks or gyratory shakers, but cells may also be cultivated in air-lift bioreactors that may hold several ten thousand liters. In small stirrer flasks, the medium is kept in motion with magnetic stirrers and rotating paddles or pendulums, necessary controls are maintained by the incubator's devices. This is no longer possible when the volume of a stirrer flask exceeds one liter. Scaling-up in volume requires better controls of medium conditions; therefore, bioreactors are equipped with sensors for the regulation of gas, temperature, pH, and metabolites.

To prevent outgrowth of cells, multicellular spheroids and three-dimensionally growing biopsies are constantly kept in motion in stirrer flasks or gyratory shakers. For this cultivation in small volumes, an interesting alternative is given by rotating wall vessels that were designed for cell growth in low or zero gravity. Their rotation speed is so adjusted that the cells remain stationary, which not

only facilitates the formation of three-dimensional aggregates but also explants from organs and tissues can be maintained under these conditions.

2.3.2 Appliances

Essential equipments that are used in tissue culture laboratories are shortly mentioned here. A laminar flow hood protects not only cells from microbial infections but also the operator from hazardous materials (see Sect. 2.4). Incubators must allow regulation of temperature, CO_2, and humidity and provide good internal convection. An autoclave is necessary for sterilizing solutions, but medium must be filtered since many ingredients will not tolerate the high temperature in an autoclave. Wasted dishes and other disposable material that came in contact with pathogens or with cultures of primate or transfected cells must also be autoclaved before they are trashed. Glassware and glass pipettes can be sterilized in a dry heat at temperatures higher than 160 °C for at least one h. Besides refrigerators and freezers for storing substances and solutions, liquid nitrogen canisters should also be available for deep-freezing and storing cell lines. Ultrapure water might not always be necessary, but a cell culture laboratory should have access to a supply; in any case, deionized and distilled water must be available. Well-adjusted phase contrast microscopes, both upright and inverted, are important tools in a cell culture laboratory since microscopical inspection of the cells during growth and before an experiment can save time and costs when inappropriate cells are to be used. A balance should not be missing as well as centrifuges and many other helpful instruments such as pH meter, cell counter, dispensers, pipettors, and a glassware washing machine.

2.4 Safety and Biohazards

For any workplace, national regulations are effective that prevent operators from safety hazards of physical or chemical origin. Working in a tissue culture laboratory adds further sources of potential safety hazards, which originate from cell cultures contaminated with latent viruses or occult pathological organisms. A known risk comes from transfected cells for which not only the origin of the cell line but also the transfection vector outlines the risk. A low individual and community risk is given with microorganisms that are unlikely to cause disease in healthy operators or animals and with cell lines that are not of human or primate origin. These risk group I agents can be handled in a containment level I laboratory that matches a functionally designed standard microbiology laboratory. Aseptic techniques must be used and all liquid waste should be treated with bleach or similar detergents. Since during pipetting aerosols are formed that might carry cells and unidentified pathogens, it is good laboratory practice to protect operators with a laminar flow hood with an air barrier at the front opening and an exhaust filter.

Agents that are of moderate individual and limited community risk are classified as risk group II. This includes cultures of primate and human origin, recombinants, transfectants, and animal tumor cells. A separate containment level II laboratory with separate equipments is obligatory, and cells must be handled in a class II laminar-flow hood. Human cell lines that are virus-producing or are infected with pathogens have a high individual and a low community risk and are classified as risk group III. These cells can also be handled in a class II laminar-flow hood,

but any waste and material that came in contact with cells must be sterilized before leaving the separate containment level III laboratory. Only designated laboratory staff is allowed to work in this room.

The risk group IV is given when a high individual and a high community risk must be expected from the experiments. This includes human pathogens that produce very serious human or animal diseases that might be untreatable and readily transmitted. Biopsies and cell cultures carrying known human pathogens must be handled in a containment level IV laboratory. In addition to the requirements for a level III laboratory, this laboratory must be physically isolated and functionally independent of other areas and must have an air lock for entry and exit. Class III biological safety cabinets and/or positive pressure ventilated suits protect designed operators for whom a shower and change of clothes are obligatory when entering or leaving the laboratory. Authorization for handling transfected cells and human pathogens as well as the routine inspection of containment laboratories are regulated by national laws.

2.5 Storage

Primary cells have a limited life span and continuous cell lines may change properties by transformation, dedifferentiation, or because of genetic instability and infection. When production or research depend on certain properties, it is therefore essential to have a stock of cells with the original properties. Cell lines as well as small multicellular organisms and embryos can be preserved by freezing since much of the cellular volume is water that is essential for cellular metabolism. When water becomes to ice, cellular metabolism is stopped and cells can be kept for years under cryopreserved conditions when the temperature is deep enough.

Before freezing, cells should be maintained under routine conditions and must be inspected for contamination. Monolayer cells are harvested during exponential growth by trypsinization; suspension cultures are concentrated by centrifugation. Cells must be thoroughly washed and resuspended to a concentration of about 1 to 10 million cells per milliliter freezing medium. This freezing medium is composed of the appropriate growth medium plus a high concentration of serum (50% or more) plus dimethyl sulfoxide or glycerol. This cell suspension is transferred in plastic ampoules wherein cells will be frozen at a slow rate till they reach −70 °C and will be kept overnight before they are transferred into liquid nitrogen and stored at −196 °C.

For thawing, an ampoule is placed in a 37 °C water bath and slowly agitated. Its content is transferred into a flask and diluted with medium. After 6 to 8 h, when viable cells are firmly attached, medium with dead cells and cellular debris should be aspirated and fresh medium is added.

3 Applications

Considerable knowledge of cellular structures and functions is the result of investigations with cells in culture. This includes intracellular synthesis processes for numerous macromolecules, such as nucleic acids and proteins, as well as for most other relevant cellular molecules. Intracellular energy and signal transfer, cytoskeletal architecture, transport through membranes, product formation, but also cell–cell interactions, malignant transformations, and

microbial infections are representatively mentioned for the numerous issues for which answers were found in cell cultures. A few examples for applications in research and production are given that concentrate on the importance of three-dimensionally grown cell aggregates, a method that might become the standard technique in the future.

3.1 Cellular Regulations

Complex cell interactions and epithelial differentiation were studied with organotypic cocultures of flat epidermal multilayers and revealed the normal regulation and balance of these cells. For this *in vitro* epidermalization, a collagen matrix at the basal side is required, whereas the keratinocytes are exposed to air. The collagen layer also supports dermal fibroblasts that synthesize the right extracellular matrix proteins. Together with the metabolites of the medium, which is separated from the collagen layer by a filter, all necessary components are available for proliferation and differentiation of the epidermal cells.

Intracellular regulations are not only dependent on the combination of cells but also on the culture conditions. This is trivial as long as necessary metabolites or messenger molecules are missing or are too low in concentration; however, with the same concentrations available in medium, cells may behave in a completely different manner when grown three-dimensionally as cell aggregates instead of as two-dimensional monolayers. An example is the so-called contact effect, which was observed for multicellular spheroids in which the tumor cells are more sensitive to irradiation than in two-dimensional growth. Later, it was detected that cells that were coupled to their neighbors via gap junction channels showed this effect more pronounced. Gap junction channels regulate an intercellular exchange of ions and molecules of up to 900 dalton and are normally found to be open in monolayer cells. However, in multicellular spheroids of several tumor cell lines, the permeability of these channels will be regulated, they were found open in two-day old spheroids and were closed two days later. It is still unknown what controls this channel closing, but channels may open again when the cells were allowed to grow as monolayer. Another growth-dependent regulation was demonstrated with β-galactosidase synthesizing L-cells. For this production, the cells had been transfected with a lacZ gene, which was under control of a β-actin promoter. This constitutive transfection led to a constant production when the cells were cultured as monolayer but β-galactosidase activity was downregulated in cells that were cultivated as spheroids. Since monolayer L-cells are motile, a permanent synthesis of cytoskeletal elements such as β-actin is required, but in spheroids the cells are rather immobile and the synthesis of β-actin is downregulated, which also effects the β-galactosidase activity. In contrast to the layered growth of epidermal cells, multicellular spheroids contain only one type of continuously growing cells, and these are exposed to gradients of metabolites, protons, and oxygen. Furthermore, spheroids establish an extracellular matrix, which is not found when the same cells are cultivated as monolayers.

3.2 Cancer Research

Malignant growth of cells can only be understood correctly when the modes for normal growth are known. Both cases can be

investigated under the same experimental conditions with cell cultures. A successful model for studying carcinogenesis *in vitro* is again the multilayered epidermis, which allowed examination of malignant transformation and tumor progression. A problem of metastatic behavior of tumor cells is invasion, it can be investigated *in vitro* with cocultures of three-dimensionally grown aggregates of normal and malignant cells. First, attempts were made with organ cultures from chick embryo or fragments of human endometrium, which were confronted with suspended tumor cells. When precultured embryonic chicken heart fragments were confronted with tumor cell spheroids, invasive behavior could be analyzed with immuno-histological techniques, and comparative studies revealed that this *in vitro* invasiveness matched the *in vivo* situation. Interestingly, it was shown that invasive tumor cells were coupled to the host via gap junction channels and that noncoupled HeLa cells became invasive after transfection with a connexin gene, which led to an expression of host-compatible gap junction channels. Several other prerequisites for invasion have been detected with similar cell culture models: Metalloproteases were found to be active during invasive processes of tumors from brain and the cell adhesion molecule e-cadherin plays a role when breast tumor cells invade the host. Weakening the cadherin-mediated cell–cell contacts facilitates detachment and migration of single tumor cells in a mesenchymal type of movement. Recent studies with new fluorescent probes allowed a time-resolved investigation of invasive processes in three-dimensional collagen matrices and revealed a supramolecular plasticity mechanism with a transition from a proteolytic mesenchymal toward a nonproteolytic amoeboid movement.

3.3 Production

Two modes of cellular production are possible: either the product remains within the cells or it is secreted into the medium. In any case, purification processes are necessary and, therefore, the culture conditions should be adjusted and optimized to increase the yield and avoid unnecessary costs. For biomedical treatments, the cells are often the requested product, which requires safe culture conditions. These should not only perfectly match the host's physiological conditions but must also be free of nonhuman additives that may be due to limited cleaning processes.

Monoclonal antibodies are produced by hybridoma cells in suspension culture. They are secreted into the medium and can be harvested with the culture supernatant. For mass production, hybridomas are often grown in hollow-fiber cultures that facilitates harvesting; the highest yield, however, might be achieved when hybridomas are grown as ascites culture in mice.

Amongst the first products of monolayer cultures were therapeutically important substances such as viral vaccines and tissue plasminogen activator, which were harvested from roller bottle cultures or from microcarriers in suspension cultures. With the availability of genetically manipulated cells, the production of therapeutically relevant proteins with animal cell cultures increased considerably. The list of these substances encloses insulin, interferon, interleukin, plasminogen activators, blood clotting factors, hemopoietic growth factors, hormones, vaccine, and so on. Many of these proteins have a complex tertiary structure and require posttranslational modifications that can only be synthesized in animal cells but not in bacteria. For mass production, producing cells

are often additionally modified so that they tolerate high shear stress when cultivated as suspension in large bioreactors.

An often-applied method in biomedical treatment is transplantation of skin. For this purpose, small pieces of skin will be taken from the patient, cut, and expanded to a mesh. They will then be cultivated, and keratinocytes will fill the interstice. These pieces can then be retransplanted onto the wounds of the patient. Another approach, which is still under development, is the temporary support of defect organs such as liver by bioartificial organs via an extracorporal bypass.

4 Perspectives

Biomedical treatment with primary or continuous cell cultures requires complex growth conditions, and some of the current approaches might lead to new therapeutic applications. Three-dimensionally growing cells do not completely represent the cells' behavior in human or animal, and monolayer cells are not necessarily representative for more complex culture systems. This does not argue against investigations with cell cultures as long as possible limitations are taken into account.

A great potential of research with monolayer cultures is given by certain cellular functions such as channel activities, intracellular transport, or protein synthesis. Examples are cells that were transfected with a fusion construct that leads to the synthesis of proteins that are labeled with a fluorescent protein. This labeling enables investigations of supramolecular dynamics by microscopic techniques such as internal reflection fluorescence or fluorescence correlation spectroscopy. Results obtained with these techniques offer insights into general cellular transport and diffusion processes.

Cellular production has been optimized with selected cell types that grow under adequate suspension- or monolayer-conditions and might still be expanded with other transfected cells and new recipes for synthetic media. A great potential for production can be expected from cell cultures of coelenterates and arthropods because pathogens are unlikely to transfer to human cells. Successful cultivation of three-dimensional aggregates of sponges has already been reported, and continuously growing insect cells are available for a longer time.

See also Antibody Molecules, Genetic Engineering of; Bacterial Cell Culture Methods; Immunology; Prokaryotic and Eukaryotic Cells in Biotech Production.

Bibliography

Books and Reviews

Abbot, A. (2003) Biology's new dimension (news feature), *Nature* **424**, 870–872.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2002) *Molecular Biology of the Cell*, 4th edition, Garland Press, New York.

Bjerkvig, R. (Ed.) (1992) *Spheroid Culture in Cancer Research*, CRC, Boca Raton, FL.

Davis, J.M. (Ed.) (2002) *Basic Cell Culture: A practical Approach*, 2nd edition, Oxford University Press, London.

Freshney, R.I. (2000) *Culture of Animal Cells: A Manual of Basic Techniques*, 4th edition, Wiley-Liss, New York.

Freshney, R.I., Freshney, M.G. (Eds.) (2002) *Culture of Epithelial Cells*, 2nd edition, Wiley-Liss, New York.

Lodish, H., Baltimore, D., Berk, A., Darnell, J., Matsudaira, P., Zipursky, S.L. (1999) *Molecular Cell Biology*, 4th edition, W. H. Freeman, New York.

Masters, J.R.W. (Ed.) (2000) *Animal Cell Culture: A Practical approach*, 3rd edition, Oxford University Press, London.

O'Brien, S.J. (2001) Cell culture forensics, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7656–7658.

Pollack, R. (Ed.) (1981) *Readings in Mammalian Cell Culture*, 2nd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor.

Primary Literature

Abercrombie, M. (1970) Contact inhibition in tissue culture, *In Vitro* **6**, 128–142.

Abercrombie, M., Heaysman, J.E.M. (1954) Observations on the social behaviour of cells in tissue culture. II. ''Monolayering'' of fibroblasts, *Exp. Cell Res.* **6**, 293–306.

Behrens, J., Mareel, M.M., Van Roy, F.M., Birchmeier, W. (1989) Dissecting tumor cell invasion: epithelial cells acquire invasive properties after the loss of uvomorulin-mediated cell-cell adhesion, *J. Cell Biol.* **108**, 2435–2447.

Bernstein, L.R., Liotta, L.A. (1994) Molecular mediators of interactions with extracellular matrix components in metastasis and angiogenesis, *Curr. Opin. Oncol.* **6**, 106–113.

Birchmeier, W. (1991) E-Cadherin-mediated cell-cell adhesion prevents invasiveness on human carcinoma cells, *J. Cell Biol.* **113**, 173–185.

Bjerkvig, R., Lærum, O.D., Mella, O. (1986) Glioma cell interactions with fetal rat brain aggregates *in vitro* and with brain tissue *in vivo*, *Cancer Res.* **46**, 4071–4079.

Bräuner, T., Hülser, D.F. (1990) Tumor cell invasion and gap junctional communication. II. Normal and malignant cells confronted in multicell spheroids, *Invasion Metastasis* **10**, 31–48.

Carrel, A.(1912) On the permanent life of tissues outside the organism, *J. Exp. Med.* **15**, 516–528.

Custodio, M.R., Prokic, I., Steffen, R., Koziol, C., Borojevic, R., Brümmer, F., Nickel, M., Müller, W.E. (1998) Primmorphs generated from dissociated cells of the sponge Suberites domuncula: a model system for studies of cell proliferation and cell death, *Mech. Ageing Dev.* **105**, 45–59.

de Ridder, L.J., Lærum, O.D. (1981) Invasion of rat neurogenic cell lines in embryonic chick heart fragments *in vitro*, *J. Natl. Cancer Inst.* **66**, 723–728.

Dertinger, H., Hülser, D. (1981) Increased radioresistance of cells in cultured multicell spheroids. I. Dependence on cellular interaction, *Rad. Envir. Biophys.* **19**, 101–107.

Dulbecco, R. (1952) Production of plaques in monolayer tissue culture by single particles of an animal virus, *Proc. Natl. Acad. Sci. U.S.A.* **38**, 747–752.

Durand, R.E., Sutherland, R.M. (1972) Effects of intercellular contact on repair of radiation damage, *Exp. Cell Res.* **71**, 75–80.

Eagle, H. (1955) Nutrition needs of mammalian cells in tissue culture, *Science* **122**, 501–504.

Earle, W.R., Schilling, E.L., Stark, T.H., Straus, N.P., Brown, M.F., Shelton, E. (1943) Production of malignancy *in vitro*; IV: The mouse fibroblast cultures and changes seen in the living cells, *J. Natl. Cancer Inst.* **4**, 165–212.

Eckert, R., Dunina-Barkovskaya, A., Hülser, D.F. (1993) Biophysical characterization of gap-junction channels in HeLa cells, *Pflügers Arch – Eur. J. Physiol.* **424**, 335–342.

Elfgang, C., Eckert, R., Lichtenberg-Fraté, H., Butterweck, A., Traub, O., Klein, R.A., Hülser, D.F., Willecke, K. (1995) Specific permeability and selective formation of gap junction channels in connexin-transfected HeLa cells, *J. Cell Biol.* **129**, 805–817.

Engebraaten, O., Bjerkvig, R., Lund-Johansen, M., Wester, K., Pedersen, P.-H., Mørk, S., Backlund, E.-O., Lærum, O.D. (1990) Interaction between human brain tumour biopsies and fetal rat brain tissue *in vitro*, *Acta Neuropathol.* **81**, 130–140.

Fusenig, N., Boukamp, P. (1998) Multiple stages and genetic alterations in immortalization, malignant transformation, and tumor progression of human skin keratinocytes, *Mol. Carcinog.* **23**, 144–158.

Gerlach, J.C., Mutig, K., Sauer, I.M., Schrade, P., Efimova, E., Mieder, T., Naumann, G., Grunwald, A., Pless, G., Mas, A., Bachmann, S., Neuhaus, P., Zeilinger, K. (2003) Use of primary human liver cells originating from discarded grafts in a bioreactor for liver support therapy and the prospects of culturing adult liver stem cells in bioreactors: a morphologic study, *Transplantation* **76**, 781–786.

Gey, G.O., Coffman, W.D., Kubicek, M.T. (1952) Tissue culture studies of the proliferative

capacity of cervical carcinoma and normal epithelium, *Cancer Res.* **12**, 264–265.

Good, N.E., Winget, G.D., Winter, W., Conolly, T.N., Izawa, S., Singh, R.M.M. (1966) Hydrogen ion buffers and biological research, *Biochemistry* **5**, 467–477.

Graeber, S.H., Hülser, D.F. (1998) Connexin transfection induces invasive properties in HeLa cells, *Exp. Cell Res.* **243**, 142–149.

Grümmer, R., Hohn, H.P., Mareel, M.M., Denker, H.W. (1994) Adhesion and invasion of three human choriocarcinoma cell lines into human endometrium in a three-dimensional organ culture system, *Placenta* **15**, 411–429.

Hansson, A., Bloor, B.K., Sarang, Z., Haig, Y., Morgan, P.R., Stark, H.J., Fusenig, N.E., Ekstrand, J., Grafström, R.C. (2003) Analysis of proliferation, apoptosis and keratin expression in cultured normal and immortalized human buccal keratinocytes, *Eur. J. Oral. Sci.* **111**, 34–41.

Harrison, R.G. (1907) Observations on the living developing nerve fiber, *Proc. Soc. Exp. Biol. Med.* **4**, 140–143.

Hayflick, L., Moorhead, P.S. (1961) The serial cultivation of human diploid cell strains, *Exp. Cell Res.* **25**, 585–621.

Hülser, D.F., Brümmer, F. (1982) Closing and opening of gap junction pores between two and three-dimensionally cultured tumor cells, *Biophys. Struct. Mech.* **9**, 83–88.

Hülser, D.F., Frank, W. (1971) Stimulation of embryonic rat cells in culture by a protein fraction isolated from fetal calf serum. I. Electrophysiological measurements on cell surface membranes [in German], *Z. Naturforsch. B* **26**, 1045–1048.

Klünder, I., Hülser, D.F. (1993) Beta-galactosidase activity in transfected Ltk-cells is differentially regulated in monolayer and in spheroid cultures, *Exp. Cell Res.* **207**, 155–162.

Knedlitschek, G., Schneider, F., Gottwald, E., Schaller, T., Eschbach, E., Weibezahn, K.F. (1999) A tissue-like culture system using microstructures: influence of extracellular matrix material on cell adhesion and aggregation, *J. Biomech. Eng.* **121**, 35–39.

Köhler, G., Milstein, C. (1975) Continuous cultures of fused cells secreting antibodies of predefined specificity, *Nature* **256**, 495–497.

Laerum, O.D., Nygaar, S.J., Steine, S., Mork, S.J., Engebraaten, O., Peraud, A., Kleihues, P., Ohgaki, H. (2001) Invasiveness *in vitro* and biological markers in human primary glioblastomas, *J. Neuro-Oncol.* **54**, 1–8.

Levi-Montalcini, R., Angeletti, P.U. (1968) Nerve growth factor, *Physiol. Rev.* **48**, 534–569.

Lin, C.Q., Bissel, M.J. (1993) Multi-faceted regulation of cell differentiation by extracellular matrix, *FASEB J.* **7**, 737–743.

Maas-Szabowski, N., Szabowski, A., Stark, H.J., Andrecht, S., Kolbus, A., Schorpp-Kistner, M., Angel, P., Fusenig, N.E. (2001) Organotypic cocultures with genetically modified mouse fibroblasts as a tool to dissect molecular mechanisms regulating keratinocyte growth and differentiation, *J. Invest. Dermatol.* **116**, 816–820.

Mareel, M., Kint, J., Meyvisch, C. (1979) Methods of study of the invasion of malignant C3H-mouse fibroblasts into embryonic chick heart *in vitro*, *Virchows Arch., B, Cell Pathol.* **30**, 95–111.

Mareel, M.M., Bracke, M.E., Van Roy, F.M., de Baetselier, P. (1997) Molecular mechanisms of cancer invasion, *Encyclopedia of Cancer* **II**, 1072–1083.

Margolis, L.B., Fitzgerald, W., Glushakova, S., Hatfill, S., Amichay, N., Baibakov, B., Zimmerberg, J. (1997) Lymphocyte trafficking and HIV infection of human lymphoid tissue in a rotating wall vessel bioreactor, *AIDS Res. Hum. Retroviruses* **13**, 1411–1420.

Margolis, L., Hatfill, S., Chuaqui, R., Vocke, C., Emmert-Buck, M., Linehan, W.M., Duray, P.H. (1999) Long term organ culture of human prostate tissue in a NASA-designed rotating wall bioreactor, *J. Urol.* **161**, 290–297.

Masters, J.R., Thomson, J.A., Daly-Burns, B., Reid, Y.A., Dirks, W.G., Packer, P., Toji, L.H., Ohno, T., Tanabe, H., Arlet, C.F., Kelland, L.R., Harrison, M., Virmani, A., Ward, T.H., Ayres, K.L., Debenham, P.G. (2001) Short tandem repeat profiling provides an international reference standard for human cell lines, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8012–8017.

Minuth, W.W., Sittinger, M., Kloth, S. (1998) Tissue engineering – generation of differentiated artificial tissues for biomedical applications, *Cell Tissue Res.* **291**, 1–11.

Moolenaar, W.H., Jalink, K. (1992) Membrane potential changes in the action of growth factors, *Cell. Physiol. Biochem.* **2**, 189–195.

Moscona, A.A. (1968) Cell aggregation: properties of specific cell-ligands and their

role in the formation of multicellular systems, *Dev. Biol.* **18**, 250–277.

Nelson-Rees, W.A., Flandermeyer, R.R. (1976) HeLa cultures defined, *Science* **191**, 96–98.

Quastler, H., Sherman, F. (1959) Cell population kinetics in the intestinal epithelium of the mouse, *Exp. Cell Res.* **17**, 420–438.

Schleich, A., Tchao, R., Frick, M., Mayer, A. (1981) Interaction of human carcinoma cells with an epithelial layer and the underlying basement membrane. A new model, *Arch. Geschwulstforsch.* **51**, 40–44.

Smith, J.A., Martin, L. (1973) Do Cells Cycle? *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1263–1267.

Smola, H., Stark, H.J., Thiekotter, G., Mirancea, N., Krieg, T., Fusenig, N.E. (1998) Dynamics of basement membrane formation by keratinocyte-fibroblast interactions in organotypic skin culture, *Exp. Cell Res.* **239**, 399–410.

Sutherland, R.M., McCredie, J.A., Inch, W.R. (1971) Growth of multicell spheroids in tissue culture as a model of nodular carcinomas, *J. Natl. Cancer Inst.* **46**, 113–120.

Todaro, G.J., Green, H. (1963) Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines, *J. Cell Biol.* **17**, 299–313.

Wolf, E. (1970) Organ Chimeras and Organ Culture of Malignant Tumors, in: Thomas, J.A. (Ed.) *Organ Culture*, Academic press, New York, pp. 459–505.

Wolf, K., Mazo, I., Leung, H., Engelke, K., von Adrian, U.H., Deryugina, E.I., Strongin, A.Y., Bröcker, E.-B., Friedl, P. (2003) Compensation mechanism in tumor cell migration: mesenchymal-amoeboid transition after blocking of pericellular proteolysis, *J. Cell Biol.* **160**, 267–277.

Mass Spectrometry of Proteins (Proteomics)

Hiroyuki Matsumoto[1], *Sadamu Kurono*[1], *Masaomi Matsumoto*[2], *and Naoka Komori*[1]
[1] *The University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA*
[2] *Oklahoma Department of Agriculture, Oklahoma City, OK, USA*

1 Proteomics: the Primary Application of Mass Spectrometry to Biomolecules 559
1.1 What is Proteomics? 560

2 Technical Basis of Proteomics 561
2.1 Protein Display for Proteomic Investigation 561
2.2 Mass Spectrometry: an Essential Tool for Proteomics 562
2.3 Genome Information: a Prerequisite for Peptide Mass Fingerprinting 564
2.4 Peptide Mass Fingerprinting 565
2.5 Confirmation of the Candidate Protein in Peptide Mass Fingerprinting 570
2.5.1 Confirmation by "Orthogonal" Type of Data 570
2.5.2 Confirmation by Peptide Fragmentation by MS/MS 571

3 Logic in Proteomics 571
3.1 Proteomics and Abductive Inference 571
3.2 Advantage of Proteomics Approach over Conventional Hypothesis-based Investigation 575

4 Proteomics under a Particular Theme: Ocular Proteomics 575

5 Quantitative Proteomics 578

6 Proteomics and Molecular Medicine 580

Acknowledgment 581

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

Bibliography 581
Books and Reviews 581
Primary Literature 581

Keywords

Abductive Inference or Abduction
Inference to the best explanation as proposed by Charles S. Peirce.

Collision Induced Dissociation or MS/MS
In CID, a selected precursor ion is activated by collisions with neutral gas molecules and fragmented into daughter ions.

Electrospray Ionization
A method for the ionization of nonvolatile molecules such as proteins, peptides, sugars, and nucleic acids by spraying the sample solution onto a small orifice in the mass spectrometer.

Endopeptidase
An enzyme that cleaves a polypeptide chain of proteins at a specific site.

Genome
The complete DNA sequence of an organism.

Mass Spectrometry/Mass Spectrometer
In mass spectrometry, a mass spectrometer measures the mass of almost any molecule that can be ionized in the gas phase.

Matrix-Assisted Laser Desorption/Ionization
A method for the ionization of nonvolatile molecules such as proteins, peptides, sugars, and nucleic acids by irradiating the mixed crystals between sample molecules and laser dye (matrix) molecules in the mass spectrometer.

Peptide Mass Fingerprinting
Identification of a protein through the measurement of the masses of endoprotease-cleaved fragments in reference to genome information.

Post Source Decay
The fragmentation of ions after the acceleration in the ion source prior to entering the reflectron that separates the daughter ions.

Posttranslational Modification
Covalent modification of amino acids in proteins taking place during and after the translational process.

Proteome
The full complement of proteins produced by a particular genome.

Proteomics
The study of the full complement of proteins encoded by a genome.

Two-dimensional Gel Electrophoresis
A gel electrophoresis system to separate a complex mixture of proteins by an isoelectric focusing in the first dimension followed by the sodium dodecylsulfate polyacrylamide gel electrophoresis in the second dimension.

U
The symbol for a mass unit representing one-twelfth the mass of carbon; equivalent to dalton (Da) used in biochemistry.

The completion of genome projects represents a pinnacle of human enterprise in the biosciences. The orchestration of technologies in mass spectrometry, genome science, and computer sciences has created a new branch of molecular bioscience called proteomics. A proteomics approach to the study of protein expression and posttranslational modification does not require foreknowledge of the identity of target proteins. A proteomic investigation begins with the discovery of unidentified proteins of interest under well-defined physiological conditions. The approach involves (1) protein display by 2-D gel electrophoresis or other separation technique; (2) determination of protein entities of interest; (3) peptide mass fingerprinting (PMF); and (4) genome/proteome database search. The methodology of the proteomics approach is characterized neither by *deduction* nor by *induction* in the traditional sense, but is a clear example of what C. S. Peirce described as *abductive inference* a century ago. *The investigation of molecular and cellular events is often intractable to deductive and inductive methods due to its extreme complexity and nonlinearity*. Proteomics is a powerful tool to study complex biological systems because of its characteristics: (1) no *a priori* knowledge of the protein's identity is required to initiate a project; and (2) a holistic approach is possible for investigation. It is expected that proteomics will substantially contribute to the future development of molecular medicine.

1 Proteomics: the Primary Application of Mass Spectrometry to Biomolecules

In the last decade, mass spectrometry, which has longer than a century of history in physical sciences, entered a new era: two new ionization methods that enable mass measurement of nonvolatile biomolecules such as proteins, nucleic acids, and sugars have been invented and perfected. The new types of mass spectrometry, together

with the development of genome science, opened up a totally new type of inquiry into the molecular aspects of life. "Proteomics" represents such emerging genre of molecular biosciences that utilizes biomolecular mass spectrometry and genome information.

1.1 What is Proteomics?

Proteomics is a new discipline in biosciences, which aims to study the proteome. The new term "proteome" was first used in late 1994 at the Siena 2-D Electrophoresis Meeting and was first printed in a scientific journal in 1995 by Wasinger et al. There is an inherent relationship between genome and proteome and the creation of the former precedes the latter. Therefore it would be appropriate to briefly explain genome before defining proteome. The term genome was first used around 1965 to represent the whole set of information encoded by chromosomal DNA. Some of the early literature includes the following: Mauer (1965), Breeze and Cohen (1965), Winocour (1965), Diamandopoulos and Enders (1965), Easton and Hiatt (1965), and Heisenberg and Blessing (1965). Since each species of living organism possesses a particular set of chromosomal DNA, it can be said that each living species possesses one genome. Moreover, since the chromosomal DNA remains the same in every cell under any physiological conditions within a certain species, the genome also remains the same. The usage of the term genome implies that each species possesses only one genome, that is, human genome, rat genome, *Drosophila* genome, yeast genome, *Escherichia coli* genome, rice genome, and so on. Those terms are well defined scientifically although there could exist some variations in genome even among the same species. In contrast to the well-defined nature of "genome" by itself, "proteome" needs to be defined with caution. The first usage of the term proteome is as follows: "proteome" refers to the total complement of a genome or the complement able to be encoded by a given genome. Thus, the term proteome was coined about three decades after the term genome was introduced. The MEDLINE (OVID) entries of relevant manuscripts reflect this historical fact: a search with "genome" yields more than 7800 entries, whereas the search for "proteome" yields about 1700 entries as of late 2002. The early literature that used the word "proteome" include Wasinger et al. (1995), Kahn (1995), Wilkins et al. (1996a), Yan et al. (1996), Qi et al. (1996), O'Brien (1996), Wilkins et al. (1996b), Wilkins et al. (1996c), Shevchenko et al. (1996), Celis et al. (1996), Wimmer et al. (1996), and Wilkins et al. (1996d). In their 1995 paper, Wasinger et al. carefully added that it is unlikely that the totality of this potential for protein expression will be realized at any one given instant. According to the "central dogma" of gene expression, DNA encodes mRNAs, and the mRNAs encode proteins. In the past half a century, since the establishment of the double helix model of DNA and its replication, many modes of molecular mechanisms that affect and regulate the expression of proteins have been discovered. The entire process of protein expression consists of three distinguished steps: transcription, translation, and posttranslation. Every process that affects any of these steps modulates the ability of the genome to express itself, thereby modulating the proteome. Under a certain condition a certain type of cell expresses a group of proteins that is a subset of the total complement of a given genome. Because changes in physiological

parameters even in the same type of cells most likely affect protein expression transcriptionally, translationally, and/or post-translationally, the kinds and amounts of proteins expressed under one set of physiological parameters will be different from those under another set of physiological parameters. Thus, in contrast to genome, which is a rather static entity under all physiological conditions, the proteome is a dynamic entity: there is only one genome for a given species, but there are many realizations of proteomes even for the same single species. This characteristic of the proteome makes it essential to define precisely the samples for proteome analysis because without defining the biological parameters such as cell types and physiological conditions the interpretation of the proteome analysis will be ambiguous. In a certain sense, all the traditional experiments performed on protein targets can be called "proteomics studies." However, there appear to be intrinsic differences between the traditional experimental inquiries in biochemistry and molecular biology and those of proteomics studies in two ways: (1) proteomics renders high throughput of inquiries because of the advance in technology of protein separation and analysis by mass spectrometry and genome database, and (2) the advent of technology to investigate proteins on a massive scale has enabled us to utilize a type of investigational logic, which is distinct from the traditional biochemical inquiries. A new logical inference appropriate for proteomics will be discussed in the following Sect. 3.1 Proteomics and Abductive Inference.

In contrast to the static aspect of genome, the dynamic aspect of genome will be studied by a new area of discipline called "functional genomics." In functional genomics, the expressed mRNAs will be studied under a well-defined condition compared to a control. The pool of mRNAs will be reverse-transcribed into cDNAs and the resulting cDNAs will be analyzed mainly by a DNA array technique. In the interpretation of DNA array experiments, the amount of a protein and that of the corresponding mRNA are assumed to be correlated, which is not always the case, at least in the case of monocellular eukaryote yeast. Further studies need to be performed to evaluate the studies of mRNA expression in reference to the protein expression, especially in higher eukaryotes. In order to achieve this goal, a sensitive and reliable methodology for quantitative proteomics of expressed proteins will be essential. (See Sect. 5 Quantitative Proteomics.)

2 Technical Basis of Proteomics

2.1 Protein Display for Proteomic Investigation

The early proteomic inquiries can be found in works in which a mixture of proteins were separated by 2-D gel electrophoresis. Although there had been several precedents for 2-D gel systems before O'Farrell (1975) invented the modern and powerful method of separating a mixture of proteins, the O'Farrell-type 2-D gel system has been proven to be a powerful method for the separation of a complex mixture of proteins. O'Farrell's protocol combines isoelectric focusing (IEF) gel under a nonionic detergent NP-40 and urea in the first dimension and the traditional Laemmli-type sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) in the second dimension. Almost at the same time, Ames and Nikaido reported a similar 2-D gel system that

makes the use of sodium dodecyl sulfate (SDS) possible for solubilizing membrane proteins. Since O'Farrell's protocol does not solubilize many of the intrinsic membrane proteins in the sample, Ames and Nikaido's method complements that of O'Farrell's. Early work using 2-D gel electrophoresis clearly demonstrated that protein expression and posttranslational modification *in vivo* are dynamic processes (for example, see Lee et al., 1979; Giometti and Anderson, 1981; Willard and Anderson, 1981; Kosik et al., 1982; Neukirchen et al., 1982; Matsumoto et al., 1982; Matsumoto and Pak, 1984). Although these examples and others indicated the power of 2-D gel electrophoresis for studying the protein expression, its weakness is that the researchers at that time did not have the power to identify the proteins of interest. This identification of proteins of interest would have to wait another decade for the development of new technologies. In the 1980s, many new and powerful technologies in biochemistry and molecular biology were developed. Modern biotechniques relevant to early proteomic inquiries include monoclonal antibody production, membrane blotting techniques both of nucleic acids and proteins, gas phase Edman degradation performed on the membrane blot, and cDNA cloning techniques including *λgt11*-based protein expression vectors that enable library screening using antibodies. It should be noted that in the 1980 to early 1990 genome databases did not exist. Therefore, cloning of each gene encoding the protein of interest was the only way to identify the gene. The premise of proteomics is that as long as a protein of interest is distinctively displayed by a 2-D gel or other methods, one can identify the protein and its gene without gene cloning. This capability of proteomics approach gives us an opportunity to catalog all the genes expressed in a particular tissue such as, for example, retina; see Fig. 1.

2.2 Mass Spectrometry: an Essential Tool for Proteomics

In late 1980s, two important ionization methods were invented and revolutionized the modern mass spectrometry: Koichi Tanaka invented matrix-assisted laser desorption/ionization (MALDI) and John Fenn invented electrospray ionization (ESI), for which both scientists received Nobel Prize in Chemistry in 2002 (http://www.nobel.se/chemistry/laureates/2002/). Before the invention of ESI and MALDI, the analyte molecules had to be volatile for any mass spectroscopic analysis. Biopolymers such as proteins, peptides, nucleic acids such as DNA and RNA, and polysaccharides cannot be analyzed by traditional mass spectrometers unless they are chemically modified to render them volatile. This is not always possible except for some shorter polymers.

A mass spectrometer consists of three important components: (1) the ion source that ionizes the analyte molecules (for example, MALDI and ESI are two types of ion sources that can ionize nonvolatile molecules such as proteins, peptides, and nucleic acids); (2) the ion analyzer that separates each ion from the mixture according to the mass-to-charge ratio (m/z); and (3) the detector that counts the number of ions characterized by a specific m/z value. Several types of ion analyzers are currently available: for example, magnetic sector–type ion analyzer, quadrupole (Q) ion analyzer, time-of-flight (TOF) ion analyzer, quadrupole ion-trap (QIT) ion analyzer, and Fourier transform ion cyclotron resonance (FTICR) ion analyzer. Theoretically, any combination of ion source

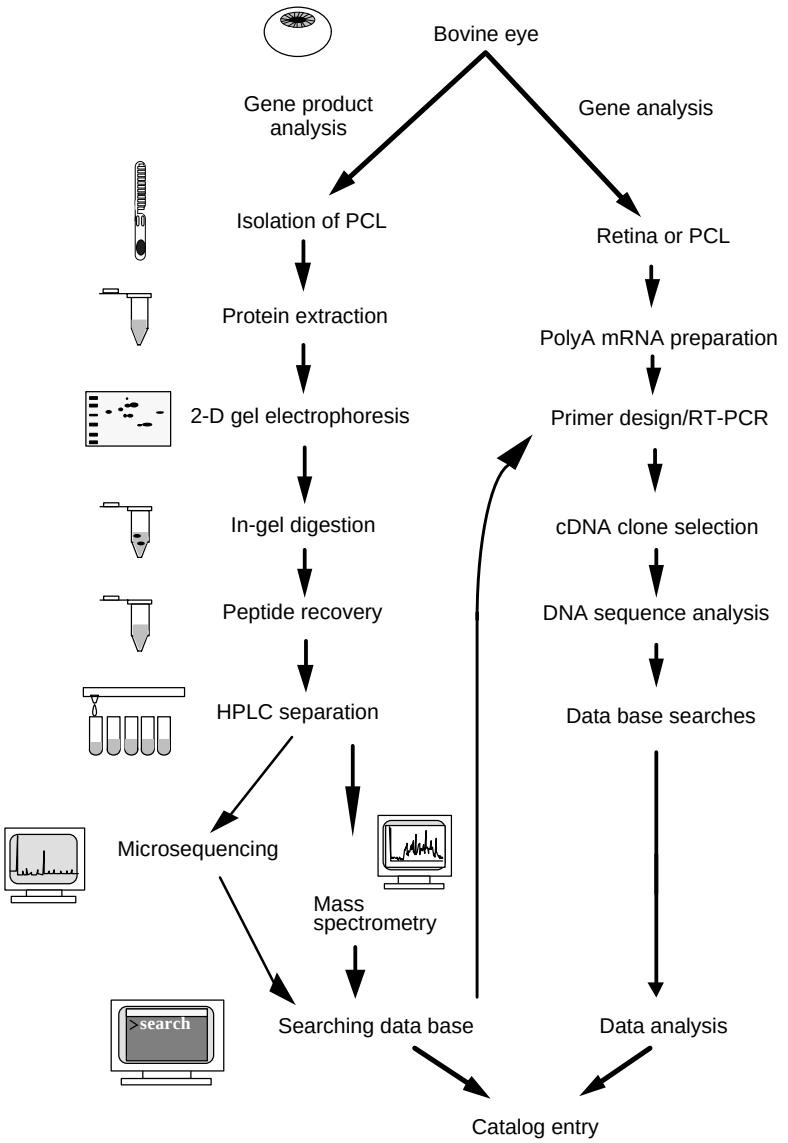

Fig. 1 A flow chart illustrating the proteomics procedures for cataloging photoreceptor proteins from the photoreceptor monolayer of bovine retina (see Fig. 9).

and ion analyzer is possible. However, the most common combinations include ESI-tandem-Q, ESI-Q-TOF, ESI-QIT, MALDI-TOF, and MALDI-QIT-TOF. FTICR is a rather expensive mass spectrometer and both ESI and MALDI ionization methods are common to FTICR. In any case, what a mass spectrometer does is to measure the mass of ionized molecules in the form of m/z.

2.3 Genome Information: a Prerequisite for Peptide Mass Fingerprinting

The ability of mass spectrometry to measure the m/z value of peptides with high sensitivity and accuracy even in a mixture constitutes the experimental basis for the identification of proteins in proteomics. The method is called "*peptide mass fingerprinting*," which is often abbreviated as PMF. In the early 1990s, several groups reported the possibility of identifying proteins based on their peptide mass fingerprints. All proteins have their unique amino acid sequences because they are all encoded by unique genes. Highly ambitious endeavors to sequence whole genome of a certain organism were planned in late 1980s. In the beginning, the genome projects progressed rather slowly. Through the early 1990s, however, the advent of automated DNA sequencing technologies together with the advancement in microcomputers and their software gradually accelerated the progress of genome projects. It is interesting to point out that since the late 1990s, the competition between the government-funded genome project led by National Institute of Health and the genome project operated by a private-sector Celera Genomics significantly accelerated the development of the human genome project. Now complete genome information is available for many bacteria, monocellular eukaryote yeast, *Caenorhabditis elegans* (worm), *Drosophila melanogaster* (fruit fly), and other model eukaryotes. The human genome has been sequenced to its rough draft in 2001, followed by its near completion in 2003. The genome projects of rodents such as rat and mice, both of which serve as excellent model systems for biomedical research, will also be completed in the near future. In the case of *D. melanogaster*, the whole DNA sequence of its genome has been known and available through the Internet since 2000. Annotation of the DNA sequence combined with all the information assembled from studies of cloning individual genes predicts 14 431 open reading frames (ORFs). This indicates that *Drosophila* is likely to have about 14 000 different kinds of proteins expressed during its entire life cycle. Suppose one discovers a protein of interest and uncovering its identity appears to be important for the further development of the ongoing project. For the practice of PMF, it is necessary that the protein of interest is isolated from other proteins and displayed in a certain reproducible way. In the early phase of proteomics around 1995, 2-D gel electrophoresis has been the major technical tool for the display and discovery of a protein of interest. However, in recent years, the capabilities of 2-D gel electrophoresis, when applied to a total protein extract of biological samples, has been challenged. The first weakness is the limitation in the capacity of 2-D gel electrophoresis to display a large number of proteins at once. The human genome is estimated to encode more than 30 000 proteins. Currently, there is no 2-D gel electrophoresis system to display proteins in the order of tens of thousands. However, as discussed in the previous chapter, it is unlikely for one type of cell to express all the proteins. Therefore, in practice, the total number of proteins that one has to display on one 2-D gel is a fraction of 30 000. We have no reliable data regarding how many proteins are actually expressed in one type of cell under a given physiological condition. It is speculated that probably the number is 5000 $\sim$ 10 000. The maximum number of proteins that can be displayed depends on the gel system used. However,

it would be safe to say that 3000 ~ 4000 is the maximum displaying capacity of a 2-D gel at best. The second weakness, which is somehow related to the first problem, is the difficulty of displaying minor proteins on a 2-D gel when major proteins are present. It is speculated that the dynamic range of protein expression can be in the order of $10^5 \sim 10^6$. Therefore, if the protein sample contains several major proteins, then the minor protein components tend to be masked on a 2-D gel. This type of situation will be observed in the analysis of body fluids such as blood. The third weakness is the difficulty of displaying membrane-bound proteins on a 2-D gel, especially of the O'Farrell's type. As mentioned earlier, the 2-D gel system developed by Ames and Nikaido would ameliorate this difficulty. However, each membrane protein has its own characteristics; there is no guarantee for the complete solubilization and the subsequent display of membrane-bound proteins on a 2-D gel. Because of these disadvantages of 2-D gel electrophoresis, other techniques for protein display and analysis have been actively sought. One major advance in this direction, sometimes claimed as a technique "beyond 2-D gel," is the development of multidimensional high performance liquid chromatography (HPLC). In this technique, the whole protein mixture will be extracted and subjected to HPLC at least twice (two-dimensional or 2-D) or more times (multidimensional). The resins packed in the HPLC columns have "orthogonal" analytical properties so that the separation in each step will not be redundant, but rather complementary, making the separation of each protein thorough. It should be noted, however, that even a carefully designed and performed multidimensional HPLC protocol could suffer from technical imperfections such as incomplete solubilization of proteins and peptides or failure in effective elution of those molecules from the columns. The effectiveness of multidimensional HPLC for displaying the proteome must be tested empirically for each proteomics application.

2.4 Peptide Mass Fingerprinting

Suppose that, after an intensive investigation, a protein is isolated and speculated to be responsible for a certain physiological and/or pathological activity. For example, the protein of interest can be displayed as a spot on a 2-D gel or can be fractionated as a distinctive elution peak on an HPLC. It should be noted that a 2-D gel that has been run and archived many years ago can still serve as a sample source for PMF (Matsumoto and Komori, the procedure for PMF is summarized in Table 1). There are three steps in PMF:

1. *Endopeptidase digestion of the protein of interest.* In most of the PMF, trypsin, which digests the carboxyl termini of lysine (K) and arginine (R), is the endopeptidase of choice. Other endopeptidases can also be used. In Fig. 2, an example is shown when the protein of interest was isolated on a 2-D gel. In-gel digestion is preferentially used for the digestion of proteins isolated on either one-dimensional or 2-D gel electrophoresis; although other methods such as digestion after electroelution of the protein or digestion after blotting on a membrane filter can be used, the in-gel digestion is less labor-intensive and yet effective. In-gel digestion of a 2-D gel protein spot followed by mass spectrometry was reported in the mid-1990s by several groups.

Tab. 1 Peptide mass fingerprinting (PMF) consists of three steps of experimental operation; (1) in-gel digestion of protein by trypsin (1 ~ 10), 2) MALDI-TOF MS (11 ~ 12) and 3) PMF search through the Internet database (13 ~ 15).

In-gel digestion

1. Excise the protein spot out. Destain the gel piece in 50% acetonitrile (MeCN)/100 mM NH_4HCO_3, pH 8.0
2. Dry the gel piece briefly.
3. Apply 2 μL (1 μL each for two times) of TPCK-treated trypsin (Promega, Madison, WI, USA) solution (1 μg/μL in 50 mM acetic acid) to the gel piece.
4. Cover the gel piece with 50 μl of 2.5 mM NH_4HCO_3.
5. Digest at 30 °C for ca. 12 h.
6. Suck up the solution and keep it.
7. Extract the digested peptides with 60% MeCN/0.1% trifluoroacetic acid (TFA) with shaking for 1 h.
8. Mix above 6 and 7 solutions and freeze-dry the peptide mixture.
9. Dissolve the residue in 5-μL water.
10. Submit 1 μL of the obtained sample to MALDI-TOF MS.

Mass spectrometry by MALDI-TOF MS

11. Mix 1 μL of the tryptic peptides with 1 μL of matrix solution consisting of 10 mg/ml α-cyano-4-hydroxycinnamic acid (α-CHCA) in 50% MeCN/0.1% TFA in a plastic tube and apply 1 mL on a sample plate.
12. Dry the sample and measure the MALDI-TOF spectrum.

Peptide mass fingerprinting (PMF)

13. Read the mass numbers of the major peaks.
14. Search PMF through MS-Fit, and so on. If the search hits a promising candidate for protein, retrieve the full sequence.
15. Reevaluate the PMF in reference to the full sequence of the candidate.

2. *Mass spectrometry.* Mass spectrometric measurement of the tryptic digest of a target protein can be made by either of the two types of mass spectrometers: electrospray-ionization (ESI) mass spectrometer or matrix-assisted laser desorption/ionization (MALDI) mass spectrometer. The ESI ion source is often interfaced with a quadrupole, quadrupole ion trap, or time-of-flight mass analyzer. The MALDI ion source is commonly interfaced with a TOF mass analyzer. The ESI ion source has a tendency to produce multiply charged peptide ions. In contrast, the MALDI ion source produces mainly singly charged peptide ions and thus the interpretation of the mass spectrum is straightforward. For the initial stage of PMF, a MALDI-TOF mass spectrometer (MALDI-TOF MS) is often used. In Fig. 3, a PMF procedure by MALDI-TOF MS is illustrated.
3. *PMF by database search.* After measuring the peptide masses created by an endoprotease such as trypsin, the mass numbers will be searched through a database. A conceptual scheme for fingerprint matching is illustrated in Fig. 4. The search is usually performed through the Internet, although the use of stand-alone databases for the search is also possible. In this particular case shown

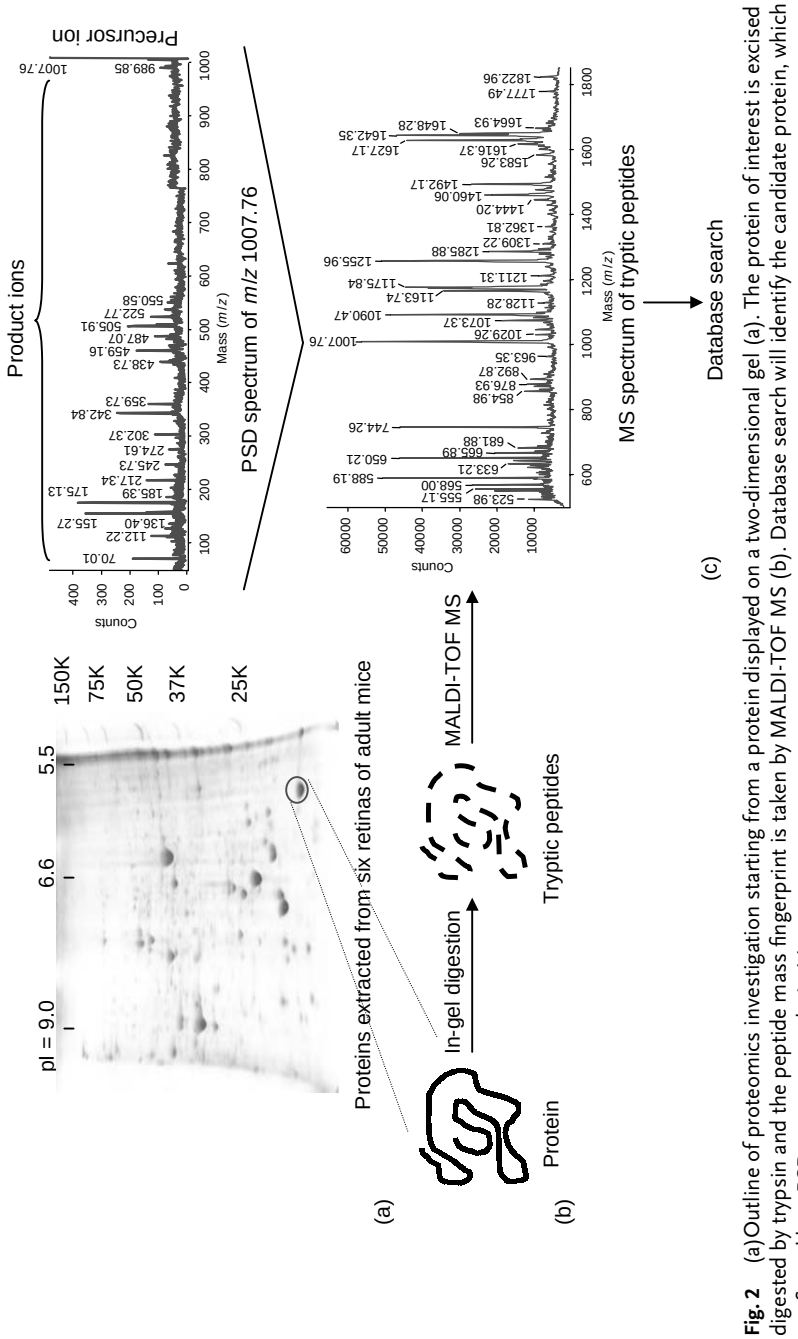

Fig. 2 (a) Outline of proteomics investigation starting from a protein displayed on a two-dimensional gel (a). The protein of interest is excised and digested by trypsin and the peptide mass fingerprint is taken by MALDI-TOF MS (b). Database search will identify the candidate protein, which will be confirmed by a PSD spectrum analysis (c).

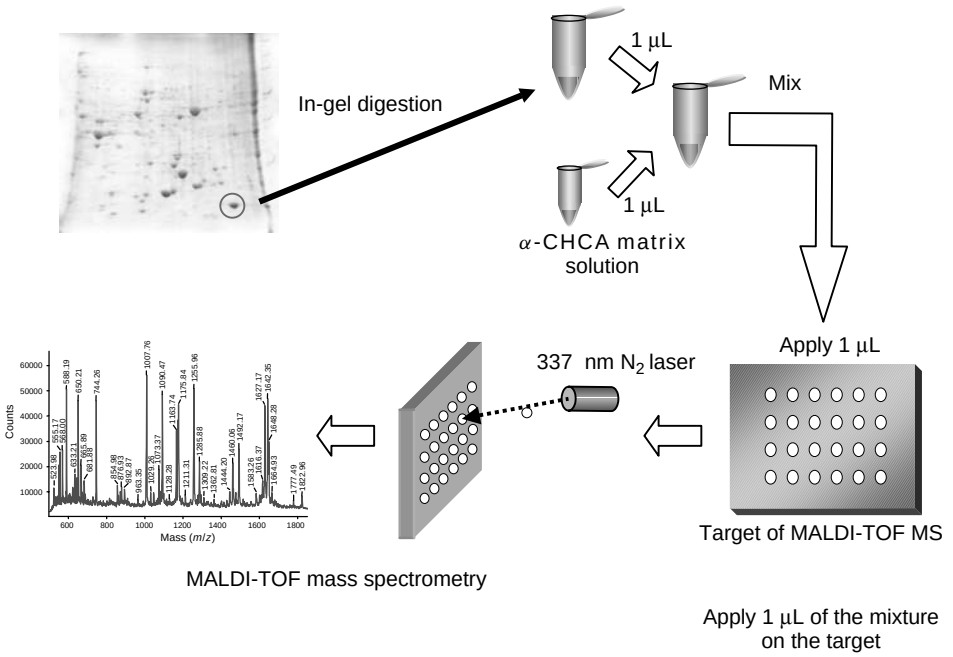

Fig. 3 Procedural outline of peptide mass fingerprinting by MALDI-TOF MS. After the in-gel digestion of the protein of interest, the mixture of tryptic digests is mixed with matrix solution containing α-cyano-4-hydroxycinnamic acid (α-CHCA) and spotted on a MALDI-TOF MS sample plate. The irradiation of the mixed crystals between the peptides and α-CHCA by 337 nm N_2 laser produces mostly singly charged ions. The m/z value of each peptide ion is measured at high accuracy and sensitivity.

Amino acid sequence of a target protein

MDVTIQHPWFKRALGPFYPSRLFDQFFGEGLFEYDLLPFLSSTISPYYRQSLFRTVLDSGISE
VRSDRDKFVIFLDVKHFSPEDLTVKVLEDFVEIHGKHNERQDDHGYISREFHRRYRLPSNVD
QSALSCSLSADGMLTFSGPKVQSGLDAGHSERAIPVSREEKPSSAPSS

(a)

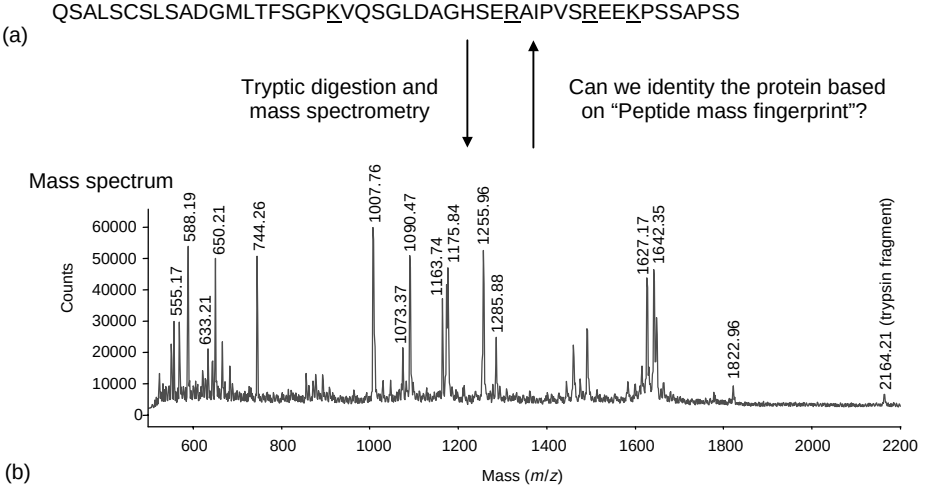

(b)

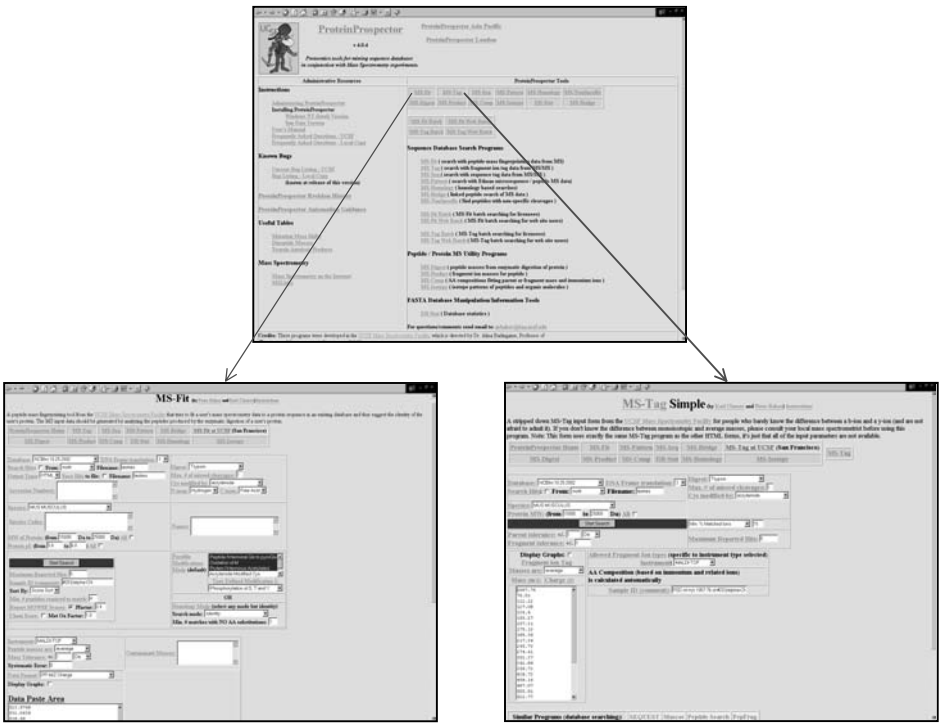

Fig. 5 An example of peptide mass fingerprinting database search: MS-Fit and MS-Tag. In MS-Fit (Lower left), the *m*/*z* values of the peptide ions are used for inquiries. In MS-Tag (Lower right), the *m*/*z* values of fragment ion tags are used for inquiries. In both cases of database search, other parameters such as ''species,'' ''the protein molecular weight range,'' and ''the tolerance of mass measurement,'' and so on can be specified. All the print outs shown in Figs. 5, 6, and 8 are the results of peptide mass fingerprinting using Protein Prospector (http://prospector.ucsf.edu/).

in Fig. 5, MS-Fit in Protein Prospector at http://www.prospector.ucsf.edu/ was used. Other database sites usable for PMF include MASCOT at http://www.matrixscience.com/, ProFound at http://prowl.rockefeller.edu/, and PeptIdent at http://www.expasy.org/. The output of MS-Fit search

Fig. 4 The concept of peptide mass fingerprinting: The amino acid sequence of the target protein is initially unknown (a). The digestion of the protein by trypsin creates a mixture of peptides. The following mass spectrometry gives the *m*/*z* values of tryptic peptide ions (b). If the amino acid sequence of the protein is known, all the peptides generated by the tryptic digest can be predicted by a theoretical (*in silico*) digestion. In a real experiment, however, it is unlikely that the mass spectrometric measurement will reveal all the digested fragments. In this particular case, the matched peptides covered only 72% (126 out of 173 amino acids) of the protein. ''Peptide mass fingerprinting'' is a practice to predict the identity of the target protein based on the observed masses of the peptide fragments.

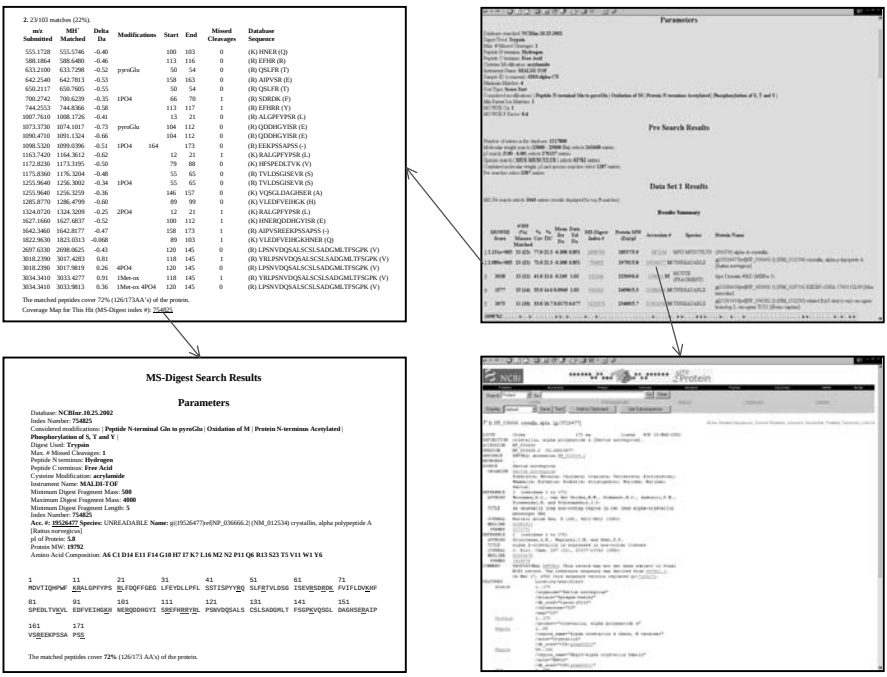

Fig. 6 The results of MS-Fit database search indicated in Fig. 5 (shown in the upper right corner). Clicking each hyperlink will output: (1) NCBI Protein Database Entry (by clicking Accession #; shown in the lower right), (2) the calculated and the observed fragments including posttranslational modifications (by clicking the candidate listing; shown in the upper left), and (3) the MS-Digest Search Result (by clicking MS-Digest Index #; shown in the lower left).

shown in Fig. 5 is shown in Fig. 6.

2.5 Confirmation of the Candidate Protein in Peptide Mass Fingerprinting

In an ideal case of PMF, all the predicted (*in silico*) fragments of the protein of interest are supposed to be confirmed in the mass spectrum. Practically, though, a total coverage of peptide mass fingerprints is unlikely to be achieved from many reasons, including incomplete digestion, low ionization efficiencies of some of the proteolytic peptides, loss of material during sample processing, and posttranslational modifications of amino acid side chains. Observation of peptides with unexpected masses, that is, "orphan masses," is rather common and is attributed to (1) posttranslational or artifactual modification, imperfect proteolysis, and/or contaminated proteins in the case of successful protein identification and (2) false positive identification of a protein.

2.5.1 Confirmation by "Orthogonal" Type of Data

In order to increase the confidence of the PMF, further experiments need to be performed. One way is to incorporate "orthogonal data types" in addition to the original PMF. The orthogonal data types include (1) chemical modification such as methyl esterification (increase of

14 mass units) on aspartic acid, Glutamic acid, and C-terminus carboxyl groups, (2) iodination on tyrosine and histidine residues (increase of 126 mass units), (3) hydrogen/deuterium exchange on all amino acids (increase of one mass unit per exchanged atom), and others. One may extend the orthogonal data types into other data types; given the amino acid sequence of the candidate protein, one may predict many properties and results that can be derived by further experiments. For example, if an antibody specific to the candidate protein is available, a western immuno blot will give further evidence. Alternatively, partial Edman sequencing of the N-terminus or the N-termini of internal peptides after appropriate endopeptidase digestion and peptide isolation will also give orthogonal type data for further confirmation. Most importantly, after this initial stage of discovery, the candidate protein will emerge and a hypothesis will be made. In all the scientific investigations, further experiments must be carried out to confirm the hypothesis. If the assignment of the protein identity is incorrect, the following experiments will produce contradictory results, resulting in the correction of the old hypothesis to a new hypothesis. This cyclic process was declared by Peirce as part of abductive inference.

2.5.2 Confirmation by Peptide Fragmentation by MS/MS

Mass spectrometry can create further evidence for the confirmation of the candidate protein if the machine is equipped with collision-induced dissociation (CID) capability. For this purpose, the mass spectrometer needs to have at least two ion mass analyzers. This can be achieved in two ways: (1) by connecting two ion mass analyzers in series, for example, such as a tandem quadrupole (tandem Q) mass spectrometer or a tandem time-of-flight mass spectrometer (TOF-TOF), and (2) by connecting two different types of ion mass analyzers, for example, such as quadrupole time-of-flight (Q-TOF) mass spectrometer or quadrupole ion trap time-of-flight (QIT) mass spectrometer. These mass spectrometers enable a real CID fragmentation to confirm the candidate proteins from the daughter ions and they are at the higher end in cost.

A MALDI-TOF mass spectrometer equipped with a reflectron can also obtain a fragmentation spectrum through "post-source decay (PSD)" mechanism. In the PSD process, the ion to be analyzed will be selected and separated through an electric mirror called "reflectron." An example is shown to confirm a peptide "ALGPFYPSR" in Fig. 7. All the expected ions in the fragmentation process can be predicted by computer algorithm, as illustrated in Fig. 8.

3 Logic in Proteomics

3.1 Proteomics and Abductive Inference

The advancement of medical biosciences in the past decades has accumulated massive amounts of information due to the rapid technological development in every relevant area. The increase in knowledge and technologies in molecular biology has played an especially significant role in directing the researchers toward molecular medicine. Molecular biology in the modern context can be said to have started when the double helical structure of DNA implying the mode of its replication was revealed by Watson and Crick in 1953. Half a century later, the draft of human

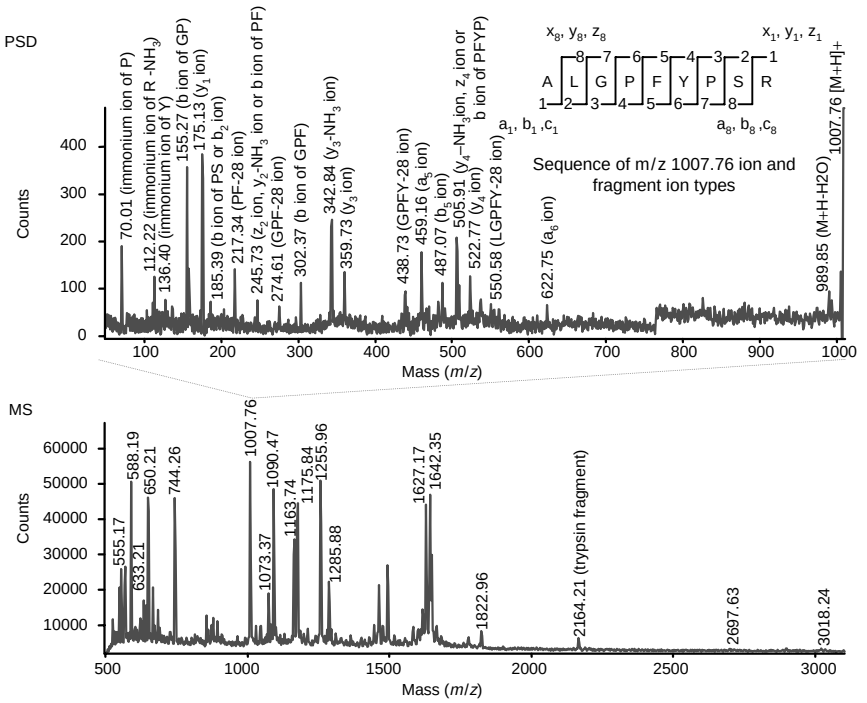

Fig. 7 An example of peptide fragmentation by PSD by MALDI-TOF MS. The ion at $m/z = 1007.76$ (shown in MS) was selected and analyzed by PSD (shown in PSD). The fragment ions shown support the assignment of the $m/z = 1007.76$ peak to be the peptide α-A-crystallin amino acid 13–21 with the sequence "ALGPFYPSR."

genome was completed in 2001. The draft of the human genome marks an epoch and summarizes all the efforts in the participating areas in biomedical sciences. During the past half century, the logical approaches that molecular biologists used are based mainly on "reductionism"; a scientific strategy and belief that a deeper understanding would be possible by investigating entities at more and more miniscule dimensions (see, for example, "Methodological reductionism" pp. 750, The Oxford Companion to Philosophy). The first and most triumphant reductionist approach in modern biochemistry took place when Eduard Büchner (1860–1917) showed that the cell-free systems of yeast are still capable of fermenting glucose in a test tube. The reductionist approach flourished since then toward the middle of the twentieth century, establishing the major metabolic pathways including glycolysis, gluconeogenesis, glycogen metabolism, citric acid cycle, pentose phosphate cycle, biosynthetic pathways of amino acids and nucleotides, catabolism of fatty acids, and others. Since the emergence of the Watson-Crick's double helix model of DNA, the trend of biochemical investigations shifted toward topics other than metabolism and a new area called molecular biology was gradually established in 1970s. The completion of the draft of human genome announced in 2001 is an epoch-making event, eloquently summarizing the past half a century of molecular

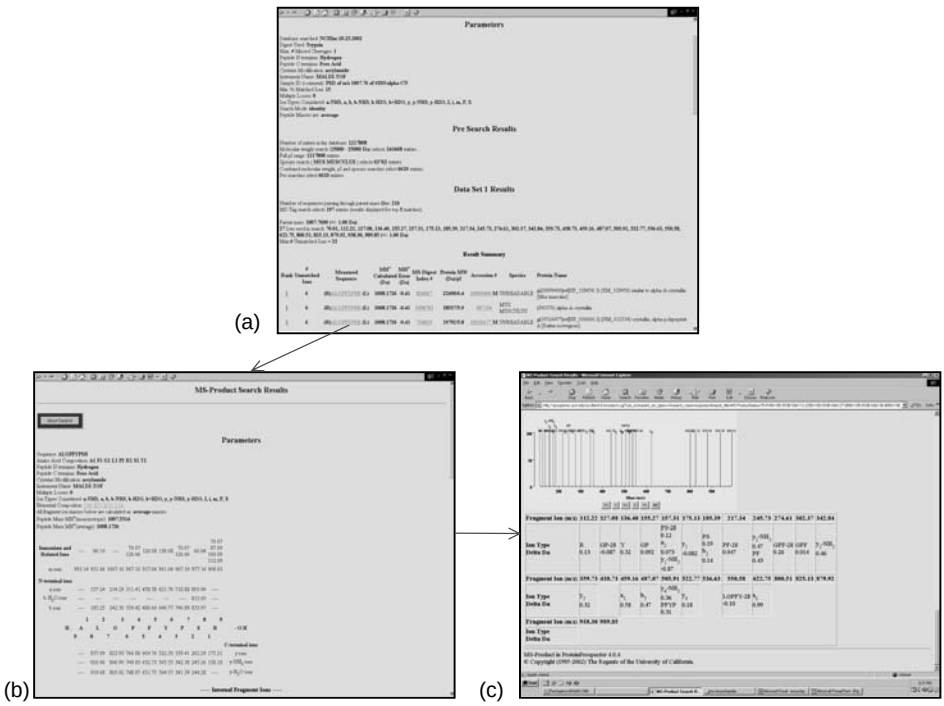

Fig. 8 The peptide fragments induced in the MS/MS process can be predicted by computer algorithm. Clicking the measured sequence (in the upper figure) outputs the MS-Product Search Results shown in (b) and (c).

biology. The initiation and substantial acceleration of progress in genome projects of bacteria, simple eukaryote models such as yeast, nematode, and fruit fly, and vertebrate models such as rodents took place in the late 1980s through the early 1990s. It is in this period that the proteomics was prepared and initiated.

The logic and strategies of proteomics investigation are novel and unique compared to those in conventional biochemistry and molecular biology. The reductionist approach in these areas has its origin in the use of cell-free systems and all subsequent efforts to fractionate the components. This approach constitutes the backbone of most of the disciplines in modern natural sciences including physics and chemistry. The arguments that have been employed in modern science are said to be classified in two categories; deduction and induction. However, it appears that the logical backbone of proteomics approach is "abductive inference or abduction," which is completely different from either deduction or induction.

Umberto Eco explains the difference between deduction, induction, and abduction as follows: "Suppose I enter a room and there find a number of bags, containing different kinds of beans. On the table there is a handful of white beans; and after some searching, I find one of the bags contains white beans only. I at once infer as a probability, or as a fair guess, that this handful was taken out of

that bag. This sort of inference is called making a hypothesis''. Eco continues as follows:

In the case of logical *deduction* there is rule from which, given a case, I deduce a result: *All the beans from this bag are white – These beans are from this bag – These beans are white*.

In the case of *induction*, given a case and a result, I infer a rule: *These beans are from this bag – These beans are white – All the beans from this bag are white* (probably).

In the case of hypothesis or *abduction*, there is the inference of a case from a rule and a result: *All the beans from this bag are white – These beans are white – These beans are from this bag* (probably).

– Umberto Eco, ''A Theory of Semiotics''

If we follow Eco's format, the logical structure of PMF can be represented by the following three statements:

1. Gene X is translated into protein Y. Protein Y, if digested, will produce a peptide mass fingerprint represented by a set of mass (y1, y2, y3, . . .).
   (This is what Eco designates a ''rule.'')
2. The digestive products of protein A gave a peptide mass fingerprint represented by a set of mass (y1, y2, y3, . . .).
   (This is what Eco designates a ''result.'')
3. Protein A is the same entity as protein Y and is encoded by gene X.
   (This is what Eco designates a ''case.'')

In Table 2, the relationship between deduction, induction, and abduction is summarized. Thus, the argument involved in PMF is an example of abduction as explained in Table 2. Note that neither rule nor result requires precise knowledge of the identity of protein encoded by gene X.

Tab. 2 Deduction, induction, and abduction.

In the case of deduction, the argument follows as indicated below:
Rule: Gene X is translated into protein Y. Protein Y, if digested, will produce a peptide mass fingerprint represented by a set of mass (y1, y2, y3, . . .).
Case: Protein A is the same entity as protein Y and is encoded by gene X.
Result: The digestive products of protein A gave a peptide mass fingerprint represented by a set of mass (y1, y2, y3, . . .).
The conclusion derived by deduction is always true.
In the case of induction, the argument follows as indicated below: Case: Protein A is the same entity as protein Y and is encoded by gene X.
Result: The digestive products of protein A gave a peptide mass fingerprint represented by a set of mass (y1, y2, y3, . . .).
Rule: Gene X is translated into protein Y. Protein Y, if digested, will produce a peptide mass fingerprint represented by a set of numbers (y1, y2, y3, . . .).
The conclusion derived by induction could be either true or false.
In the case of abduction, the argument follows as indicated below:
Rule: Gene X is translated into protein Y. Protein Y, if digested, will produce a peptide mass fingerprint represented by a set of numbers (y1, y2, y3, . . .).
Result: The digestive products of protein A gave a peptide mass fingerprint represented by a set of numbers (y1, y2, y3, . . .).
Case: Protein A is the same entity as protein Y and is encoded by gene **X**.
The conclusion derived from abduction is actually a hypothesis that needs to be proven by further experiments.

3.2 Advantage of Proteomics Approach over Conventional Hypothesis-based Investigation

The logical structure of proteomics investigation as explained in 3.1 renders a unique advantage that does not exist in conventional hypothesis-based investigation. The distinguishing characteristic of abductive inference is the fact that it requires no initial hypothesis. This is because *abduction is a process of inferring to the best explanation. Abduction, instead of being based on a hypothesis, will create a hypothesis.* This characteristic in the logical structure of proteomics approach makes it suitable for the investigation of complex systems such as cellular signaling and gene expression. The wider the system to be investigated in biomedical sciences, the more complex and nonlinear the underlying molecular mechanisms will be. The most extreme cases include diagnosis of diseases. Abduction has been evaluated to be the right and practical logic of diagnosis. In a similar context, abductive inference in proteomics investigation will be a powerful tool to study the diagnostic and pharmacological aspects of human diseases. There is no surprise in observing the fact that many pharmaceutical/biotechnology companies have initiated and invested substantial resources along this line, resulting in new areas such as pharmaceutical proteomics or proteomics of diseases. A unifying theme in all these efforts is the holistic character of approach. Expected prospective areas using proteomics approach include the following: cancer diagnosis, disease diagnosis from body fluids, pharmacological screening, toxicology, hormone research, and diagnosis of infection. In fact, the applicability of proteomics investigation to medicine is unlimited.

4 Proteomics under a Particular Theme: Ocular Proteomics

Although proteomics techniques are high throughput in nature, it is more sensible to reduce the variables in the original samples as much as possible in order to make the interpretation of the results simpler. Practically speaking, each proteomics investigator is likely to have a specific target. "Ocular proteomics" is illustrated here as an example. All the visual information is initiated when photons reflected from the visible objects get absorbed by the retinal (i.e. vitamin A aldehyde) chromophore in the visual pigment molecules rhodopsin and cone visual pigments. These visual pigments constitute the photosensitive membrane structure of photoreceptor cells in the retina. The retina contains, in addition to photoreceptor cells, other types of neural and glial cells in order to maintain its function. The neural retinal cells process visual signals prior to the visual cortex of the CNS. There are numerous diseases reported in which malfunction and deterioration of retinal cells occur. Some retinal degenerative diseases are apparently caused by mutations by retinal specific genes, implying that protein malfunction causes the retinal degeneration. Therefore, cataloguing all the proteins expressed in the retina will be a useful resource for the understanding of retinal function and its malfunction at the protein level.

An effort has been made to initiate cataloguing proteins expressed in the photoreceptor cell layer of bovine retina compared to the rest of the retina from which the photoreceptor layer has been depleted. In order to dissect the retina, a tissue printing method was used, as illustrated in Fig. 9. The variation of this

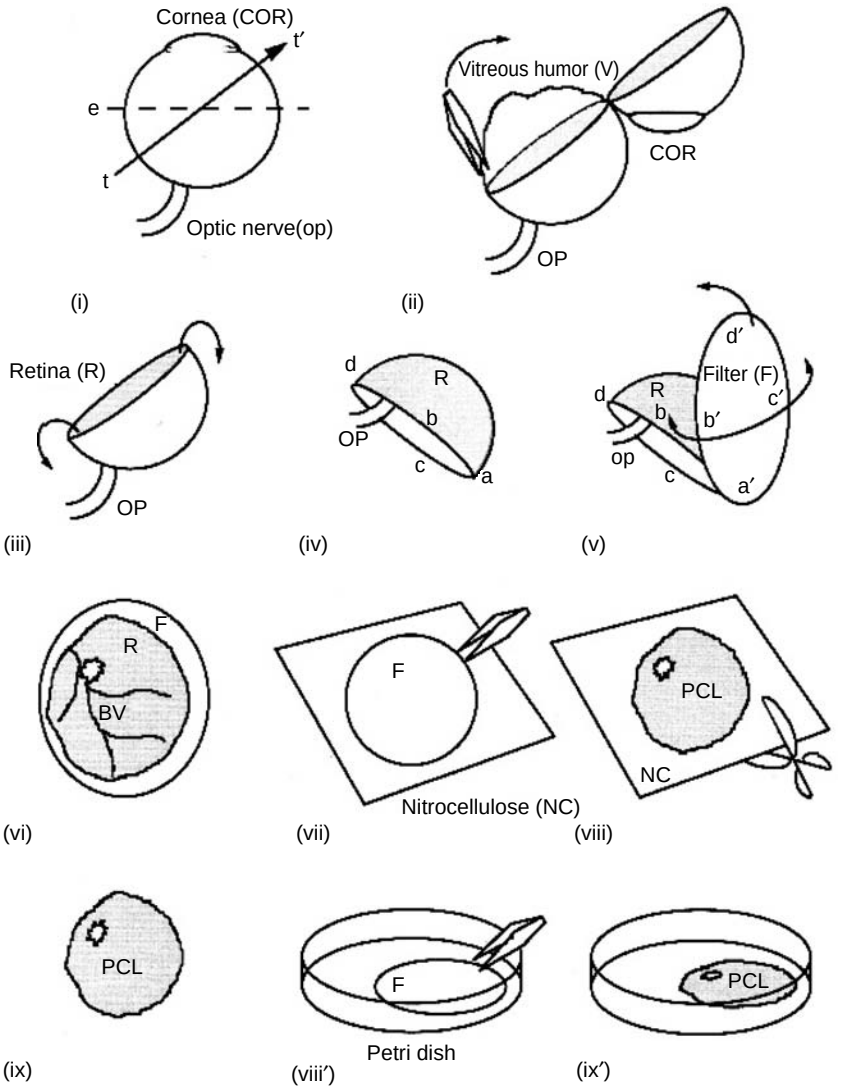

Fig. 9 A schematic diagram for preparation of the photoreceptor cell monolayer. (i) The eyeball is cut obliquely along the line t to t'. Cornea (COR), Optic nerve (OP). (ii) Vitreous humor (V) is removed from the eyecup. (iii) & (iv) The dissected eyecup is inverted. Retina (R). (v) & (vi) The retina is transferred to a Whatman filter paper (F) and lifted. (vii) The filter paper is overlaid face-side down onto a nitrocellulose membrane (NC). (viii) As the filter paper is lifted, the photoreceptor cell monolayer (PCL) remains on the nitrocellulose membrane. The excess margin of NC is removed with scissors. (ix) The PCL preparation is obtained. (viii′) As a modification, a plastic Petri dish is used instead of the nitrocellulose membrane. (ix′) The PCL transfers onto the Petri dish.

method can be used for other tissues if cells of significant similarity form a layer detachable from other types of cells. After tissue printing of the photoreceptor cells, the separated layers were examined by morphology (Fig. 10, left). Each layer of preparation was processed for 2-D gels (Fig. 10, right). The major 2-D gel spots on the photoreceptor layer (PCL) were excised and PMF was performed. The results of PMF are summarized in Table 3. By comparing the 2-D gels in Fig. 10 and Table 3, we conclude that four protein spots are expressed more abundantly or even exclusively in the PCL layer; Spot 1, Spot 8, Spot 9, and Spot 12. The protein represented by Spot 1 is IRBP, which is an extracellular protein assumed to be responsible for carrying retinoids between the pigment epithelial cells and the photoreceptor cells. The proteins represented by Spot 8 and Spot 9 are aspartate aminotransferase (AAT) and creatine kinase (CK) respectively, both of the mitochondrial type. Both of these enzymes are crucial components of the energy metabolism in the mitochondria. The protein represented by Spot 12 is the β subunit of transducin, which is involved in the G protein–coupled receptor signaling in the photoreceptor excitation and, therefore, is highly specific to the photoreceptor cells. The protein spots other than 1, 8, 9, and 12 also exist abundantly in the rest of the retina (Fig. 10). Those proteins expressed in high abundance in the entire retina are components of the glycolysis, as illustrated

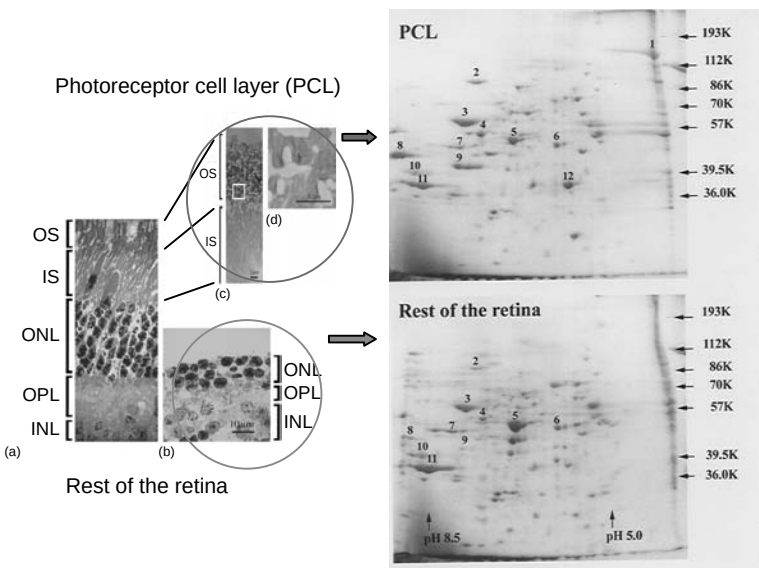

Fig. 10 Electron microscope images (Left) of the bovine retina (a), residual retina after removal of the photoreceptor cell monolayer (b), PCL adhered to nitrocellulose (c). Figure (d) illustrates the enlarged portion of OS as indicated by a square in Fig. (c). OS, outer segments; IS, inner segments; ONL, outer nuclear layer; OPL, outer plexiform layer; INL, inner nuclear layer. Each bar in the figures indicates dimension. The PCL sample and the rest of the retina were then subjected to 2-D gel electrophoresis (Upper and Lower Right).

Tab. 3 Major photoreceptor cell proteins assigned. MW, molecular weight in kilodaltons; pI, isoelectric point.

Spot #	*Assignment*	*MW/pI*	*Comments*
1	IRBP	138 kDa/5.0	Retinoid metabolism/extracellular
2	cis-Aconitase	82.7 kDa/7.6	TCA cycle/mitochondrial
3	Pyruvate kinase	55 kDa/7.7	Glycolysis/cytosolic
4	F1-ATPase	55.3 kDa/8.6	Respiratory/mitochondrial
5	Enolase	50 kDa/5.9	Glycolysis/cytosolic
6	Unknown		
7	Phosphoglycerate kinase	47 kDa/8.0	Glycolysis/cytosolic
8	(AAT)	45 kDa/9.2	Mitochondrial
9	Creatine kinase	43 kDa/7.2	Mitochondrial
10	Pyruvate dehydrogenase	41 kDa/8.6	Glycolysis/cytosolic
11	Glyceraldehyde 3-phosphate dehydrogenase	36 kDa/8.5	Glycolysis/cytosolic
12	Transducin-β	37 kDa/5.9	Phototransduction/membrane

in Fig. 11. From these results, one may conclude at least two things: (1) in bovine retina, the most abundant classes of proteins include the enzymes that are responsible for basic metabolism such as glycolysis, and (2) in the photoreceptor layer, two enzymes, that is, AAT and CK in the mitochondria responsible for the energy metabolism, are abundant compared to the rest of the retinal layer excluding the photoreceptor cells. Interesting conjectures emerge from these results. These conjectures can be evaluated by questions such as follows: (1) Do the mitochondria in the photoreceptor cells express larger quantities of AAT and CK compared to those in the rest of the retina? (2) Is the role of AAT in the photoreceptor cells to bypass the tricarboxylic acid cycle? These questions are being asked on the basis of a new set of hypotheses that did not exist before the experiments. Therefore, the whole process of proteomics approaches follows the method of *abductive inference* (or *making hypotheses*) that Peirce formulated. It should be noted that, based on these new hypotheses, another set of experiments will be performed and that the experiments should be conducted on the basis of hypotheses through conventional investigational logic, hence not by the "abductive inference" *per se* used in the very beginning of inquiries.

5 Quantitative Proteomics

The first stage of proteomics is to compare two samples, that is, a standard and the unknown sample. In order to obtain quantitative results in the comparison, there are at least two types of techniques available.

First is a rather straightforward method of obtaining quantitative data by scanning the 2-D gels by a densitometric scanner and quantifying the density of each individual spot by integration of pixels. In order to perform this type of quantitation, a scanning densitometer with an accurate calibration and software to perform the gel image analysis and pixel integration of each spot are required. Currently, there are several companies selling a scanning

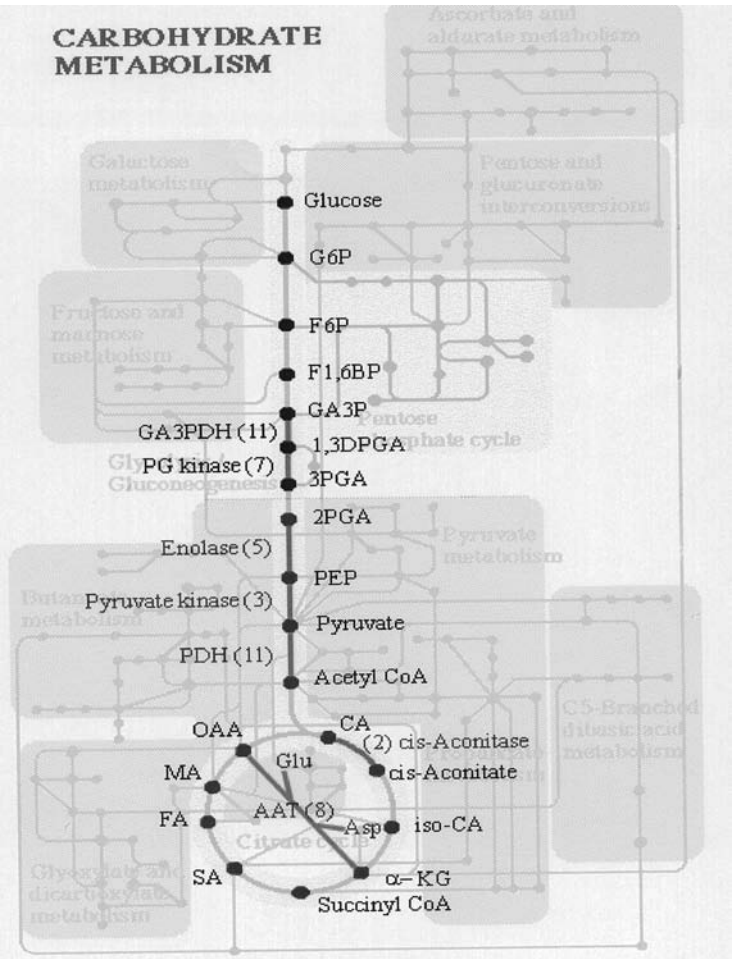

Fig. 11 The major proteins expressed in the bovine retina are enzymes in carbohydrate metabolism. The mitochondrial AAT is expressed in a larger quantity in the PCL compared to the rest of the retina. This result produces a hypothesis that the tricarboxylic cycle in the PCL operates differently from that in the rest of the retinal cells, that is, the interconversion of glutamate (Glu) and aspartate (Asp) via the bypass reaction between oxaloacetic acid (OAA) and α-ketoglutaric acid (α-KG) is more active in the PCL than the rest of the retina. The metabolic scheme is adapted from a website at Kyoto Encyclopedia of Genes and Genomes (KEGG).

densitometer and software that are suitable for this purpose.

Second group of quantitation methods utilizes modification of amino acid side chains with a couple of agents that have different physical properties such as different emission properties in a pair of fluorescent dyes or different masses caused by the incorporation of stable isotopes such as deuterium (D), which has 2 u compared to H (1 u). Carbon-13 (13 u) instead of carbon-12 (12 u) and nitrogen-15

(15 u) instead of nitrogen-14 (14 u) can also be used. In this type of quantitation, one sample (control) is labeled with a reagent with a certain property and the second sample (experiment) is labeled with a reagent with a property distinguishable from the control. Endopeptidase digestion needs to be done after or before the modification of amino acid side chains, depending on which system will be used.

For example, in a protocol called "2-D Fluorescence Differential Gel Electrophoresis (DIGE)", two protein samples will be modified by two different fluorescent dyes that emit at distinguishable wavelengths, and mixed. The mixture of fluorescent dye-labeled proteins will be separated on a 2-D gel. Since the pair of fluorescent dyes are designed so that the changes in the isoelectric point (pI) and the molecular weight are the same between the control and the experiment, each protein of the same kind, though originating from different samples, migrates and focuses at the same spot on a 2-D gel. With the use of a two-color scanning fluorescence densitometer, the amount of the modified proteins can be quantified.

Another example of quantitation in the second group is a mass spectrometric method in which a pair of light or heavy peptide modifier is used. "Isotope Coded Affinity Tag (ICAT)" is an especially successful technique that belongs to this category. In the ICAT technique, two different samples will be labeled by a derivative of a cysteine-modifier iodoacetoamide that carries an affinity tagged biotin and 8 atoms of deuterium (d_8-ICAT). A control ICAT reagent was synthesized without the 8 atoms of deuterium, making the reagent 8 u lighter than the d_8-ICAT. The two samples modified with d_8-ICAT and ICAT will be digested. The advantage of ICAT is that the affinity tag biotin can be used to purify the modified peptide fragment. After affinity purification by an avidin column, the eluate will be analyzed by an HPLC-interfaced ESI mass spectrometer. A deuterium-labeled *N*-alkylmaleimides also can be used in this type of analysis without the sophistication of affinity tag. Another method to quantify protein involves the digestion of one protein sample in the presence of ^{18}O-labeled water. The other sample to be compared will be digested in regular water (^{16}O). The former digests will make each peptide 4 u heavier than the latter. After mixing the digests and a high-resolution mass spectrometric analysis, the ratio between the two peaks separated by 4 u indicates the ratio of the protein existing in the two samples.

6 Proteomics and Molecular Medicine

The unique characteristic of proteomics approach that does not require presuppositions or hypotheses makes it an excellent tool for molecular medicine. Instead of starting from one particular protein that is supposed to be important for the understanding of molecular mechanisms underlying a disease, a proteomics approach can start from evaluating the protein expression and modification and their abnormal variance without knowing what they are in the beginning. Development of technologies in many areas in the past decades enabled us to apply a proteomics approach to virtually any area of medical biosciences. Each area of proteomics inquiry will be defined well by the set of parameters in its particular project. One can choose any input parameters that affect the system to be investigated, and, depending on the variable parameters, proteomics can be classified as, for example, "pharmaceutical

proteomics,'' ''physiological proteomics,'' ''pathological proteomics,'' and so on. For example, with a proteomics approach, it is now feasible to initiate a project to evaluate the effect of drugs on the protein expression and posttranslational modification in a particular type of cells under a particular set of physiological parameters (pharmaceutical proteomics). Another example is that a comparison of proteins from cancerous tissues and those from normal tissues may give us a hint to initiate a new project based on a new hypothesis (cancer proteomics). Thus, the applicability and versatility of proteomics approach to the study of diseases appear to be unlimited.

Acknowledgment

This work is supported by grants from National Institutes of Health EY06595, EY13877, EY12190, and RR17703.

See also Gel Electrophoresis, 2D-difference; Mass Spectrometry, High Speed DNA Fragment Sizing; Nucleic Acids (DNA) Sequencing, Transcriptional.

Bibliography

Books and Reviews

Buchler, J. (Ed.) (1955) *Philosophical Writings of Peirce*, Dover Publications, New York.

Copi, I.M. (1972) *Introduction to Logic*, 4th edition, Collier-Macmillan Canada, Ltd., Toronto, Canada.

Dowling, J.E. (1987) *Retina: An Approachable Part of the Brain*, The Belknap Press of Harvard University Press, Cambridge, MA.

Eco, U. (1979) *A Theory of Semiotics*, Indiana University Press, Bloomington, IN.

Grayson, M.A. (Ed.) (2002) *Measuring Mass: From Positive Rays to Proteins*, Chemical Heritage Press, Philadelphia, PA.

Josephson, J.R., Josephson, S.G. (Eds.) (1996) *Abductive Inference. Computation, philosophy, technology*, Cambridge University Press, Cambridge, MA.

Liebler, D.C. (2002) *Introduction to Proteomics*, Humana Press, Totowa, NJ.

Matsumoto, H., Komori, N. (2000) Ocular proteomics: cataloging photoreceptorproteins by 2-D gel electrophoresis and mass spectrometry, *Methods Enzymol.* **316**, 492–511.

Matsumoto, H., Kahn, E.S., Komori, N. (1999) Emerging Role of Mass Spectrometry in Molecular Biosciences: Studies of Protein Phosphorylation in Fly Eyes as an Example, in: Yoshizawa., T. (Ed.) *Rhodopsins and Phototransduction*, Novartis Foundation Symposium No. 224., pp. 225–248.

Morange, M. (1998) *A History of Molecular Biology*, Translated by Cobb, M., Harvard University Press, Cambridge, MA.

Rabilloud, Th. (Ed.) (2000) *Proteome Research: Two-dimensional Gel Electrophoresis and Identification Methods*, Springer, Berlin, Germany.

Siuzdak, G. (1996) *Mass Spectrometry for Biotechnology*, Academic Press, San Diego, CA.

Sparkman, O.D. (2000) *Mass Spectrometry Desk Reference*, Global View Publishing, Pittsburgh, PA.

Wilkins, M.R., Williams, K.L., Appel, R.D., Hochstrasser, D.F. (Eds.) (1997) *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin, Germany.

Primary Literature

Ames, G.F., Nikaido, K. (1976) Two-dimensional gel electrophoresis of membrane proteins, *Biochemistry* **15**, 616–623.

Breeze, D.C., Cohen, P.S. (1965) Evidence for the exponential reproduction of the genome of encephalomyocarditis virus, *Virology* **27**, 227–229.

Corthals. G.L., Gygi, S.P., Aebersold, R., Patterson, S.D. (2000) Identification of Proteins by Mass Spectrometry, in: Rabilloud, Th., (Ed.) *Proteome Research: Two-dimensional Gel Electrophoresis and Identification Methods*, Springer, Berlin, Germany pp. 197–231.

Celis, J.E., Gromov, P., Ostergaard, M., Madsen, P., Honore, B., Dejgaard, K., Olsen, E., Vorum, H., Kristensen, D.B., Gromova, I., Haunso, A., Van Damme, J., Puype, M., Vandekerckhove, J., Rasmussen, H.H. (1996) Human 2-D PAGE databases for proteome analysis in health and disease, *FEBS Lett.* **398**, 129–134.

Diamandopoulos, G.T., Enders, J.F. (1965) Studies on transformation of Syrian hamster cells by simian virus 40 (SV40): acquisition of oncogenicity by virus-exposed cells apparently unassociated with the viral genome, *Proc. Natl. Acad. Sci. U. S. A.* **54**, 1092–1099.

Easton, J.M., Hiatt, C.W. (1965) Possible incorporation of SV40 genome within capsid proteins of adenovirus 4, *Proc. Natl. Acad. Sci. U. S. A.* **54**, 1100–1104.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M. (1989) Electrospray ionization for mass spectrometry of large biomolecules, *Science* **246**, 64–71.

Fountoulakis, M., Langen, H. (1997) Identification of proteins by matrix-assisted laser desorption ionization-mass spectrometry following in-gel digestion in low-salt, nonvolatile buffer and simplified peptide recovery, *Anal. Biochem.* **250**, 153–156.

Gharbi, S., Gaffney, P., Yang, A., Zvelebil, M.J., Cramer, R., Waterfield, M.D., Timms, J.F. (2002) Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system, *Mol. Cell. Proteomics* **1**, 91–98.

Giometti, C.S., Anderson, N.G. (1981) Muscle protein analysis. III. Analysis of solubilized frozen-tissue sections by two-dimensional electrophoresis, *Clin. Chem.* **27**, 1918–1921.

Gygi, S.P., Rochon, Y., Franza, B.R., Aebersold, R. (1999b) Correlation between protein and mRNA abundance in yeast, *Mol. Cell. Biol.* **19**, 1720–1730.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R. (1999a) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nat. Biotechnol.* **17**, 994–999.

Han, D.K., Eng, J., Zhou, H., Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry, *Nat. Biotechnol.* **19**, 946–951.

Heisenberg, M., Blessing, J. (1965) Unvollstandige Viruspartikel als Folge von Mutationen im Genom des Bacteriophagen fr., *Zeitschrift fur Naturforschung – Teil B – Anorganische Chemie, Organische Chemie, Biochemie, Biophysik, Biologie* **20**, 859–864.

Henzel, W.J., Billeci, T.M., Stults, J.T., Wong, S.C., Grimley, C., Watanabe, C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases, *Proc. Natl. Acad. Sci. USA* **90**, 5011–5015.

Hochstrasser, D.F. (1997) Clinical and Biomedical Applications of Proteomics, in: Wilkins, M.R., Williams, K.L., Appel, R.D., Hochstrasser, D.F. (Eds.) *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin, Germany, pp. 187–219.

Hookway, C.J. (1995) Abduction, in: Honderich, T. (Ed.) *The Oxford Companion to Philosophy*, Oxford University Press, Oxford, UK, pp. 1.

Hurley, J.B., Yarfitz, S. (1994) Transduction mechanisms of vertebrate and invertebrate photoreceptors, *J. Biol. Chem.* **269**, 14329–14332.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of human genome, *Nature* **409**, 860–921.

James, P., Quadroni, M., Carafoli, E., Gonnet, G. (1993) Protein identification in DNA databases by peptide mass fingerprinting, *Biochem. Biophys. Res. Commun.* **195**, 58–64.

Ji, H., Whitehead, R.H., Reid, G.E., Moritz, R.L., Ward, L.D., Simpson, R.J. (1994) Two-dimensional electrophoretic analysis of proteins expressed by normal and cancerous human crypts: application of mass spectrometry to peptide-mass fingerprinting, *Electrophoresis* **15**, 391–405.

Kahn, P. (1995) From genome to proteome: looking at a cell's proteins, *Science* **270**, 369–370.

Kinumi, T., Jackson, K.W., Ohashi, M., Tobin, S.L., Matsumoto, H. (1997) The phosphorylation site and desmethionyl N-terminus of *Drosophila* phosrestin I *in vivo* determined by mass spectrometric analysis of proteins separated on two-dimensional gel electrophoresis, *Eur. Mass Spectrom.* **3**, 367–378.

Kosik, K.S., Gilbert, J.M., Selkoe, D.J., Strocchi, P. (1982) Characterization of

postmortem human brain proteins by two-dimensional gel electrophoresis, *J. Neurochem.* **39**, 1529–1538.

Larsson, T., Norbeck, J., Karlsson, H., Karlsson, K.A., Blomberg, A. (1997) Identification of two-dimensional gel electrophoresis resolved yeast proteins by matrix-assisted laser desorption ionization mass spectrometry, *Electrophoresis* **18**, 418–423.

Lee, C.Y., Charles, D., Bronson, D., Griffin, M., Bennett, L. (1979) Analyses of mouse and Drosophila proteins by two-dimensional gel electrophoresis, *Mol. Gen. Genet.* **176**, 303–311.

Li, G., Waltham, M., Anderson, N.L., Unsworth, E., Treston, A., Weinstein, J.N. (1997) Rapid mass spectrometric identification of proteins from two-dimensional polyacrylamide gels after in gel proteolytic digestion, *Electrophoresis* **18**, 391–402.

Lindemann, T., Hintelmann, H. (2002) Identification of selenium-containing glutathione S-conjugates in a yeast extract by two-dimensional liquid chromatography with inductively coupled plasma MS and nanoelectrospray MS/MS detection, *Anal. Chem.* **74**, 4602–4610.

Lockhart, D.J., Winzeler, E.A. (2000) Genomics, gene expression and DNA arrays, *Nature* **405**, 827–836.

Macdonald, N., Chevalier, S., Tonge, R., Davison, M., Rowlinson, R., Young, J., Rayner, S., Roberts, R. (2001) Quantitative proteomic analysis of mouse liver response to the peroxisome proliferator diethylhexylphthalate (DEHP), *Arch. Toxicol.* **75**, 415–424.

Matsui, N.M., Smith, D.M., Clauser, K.R., Fichmann, J., Andrews, L.E., Sullivan, C.M., Burlingame, A.L., Epstein, L.B. (1997) Identification of two-dimensional gel electrophoresis resolved yeast proteins by matrix-assisted laser desorption ionization mass spectrometry, *Electrophoresis* **18**, 409–417.

Matsumoto, H., Komori, N. (1999) Protein identification on two-dimensional gels archived nearly two decades ago by in-gel digestion and matrix-assisted laser desorption ionization time-of-flight mass spectrometry, *Anal. Biochem.* **270**, 176–179.

Matsumoto, H., Pak, W.L. (1984) Light-induced phosphorylation of retina-specific polypeptides of *Drosophila in vivo*, *Science* **223**, 184–186.

Matsumoto, H., Kurono, S., Komori, N. (2002) Proteomics as a Tool for Studying Complex Systems and the Abductive Inference of C.S. Peirce, *J. Mass Spectrom. Soc. Jpn.* **50**, 116–125.

Matsumoto, H., O'Tousa, J., Pak, W.L. (1982) Light-induced modification of *Drosophila* retinal polypeptides *in vivo*, *Science* **217**, 839–841.

Matsumoto, H., Kurien, B., Takagi, Y., Kahn, E.S., Kinumi, T., Komori, N., Yamada, T., Hayashi, F., Isono, K., Pak, W.L., Jackson, K.W., Tobin, S.L. (1994) Phosrestin I undergoes the earliest light-induced phosphorylation by a calcium/calmodulin-dependent protein kinase in *Drosophila* photoreceptors, *Neuron* **12**, 997–1010.

Mann, M., Hojrup, P., Roepstorff, P. (1993) Use of mass spectrometric molecular weight information to identify proteins in sequence databases, *Biol. Mass Spectrom.* **22**, 338–345.

Mauer, I. (1965) Chromosome and genome as conceptual models for certain types of mental retardation, *Am. J. Mental Deficiency* **70**, 191–203.

Neukirchen, R.O., Schlosshauer, B., Baars, S., Jackle, H., Schwarz, U. (1982) Two-dimensional protein analysis at high resolution on a microscale, *J. Biol. Chem.* **257**, 15229–15234.

Nishizawa, Y., Komori, N., Usukura, J., Jackson, K.W., Tobin, S.L., Matsumoto, H. (1999) Initiation of ocular proteomics for cataloging bovine retinal proteins: microanalytical techniques permit the identification of proteins derived from a novel photoreceptor preparation, *Exp. Eye Res.* **69**, 195–212.

Niwayama, S., Kurono, S., Matsumoto, H. (2001) Synthesis of d-labelled N-alkylamides and application to quantitative peptide analysis by isotope differential mass spectrometry, *Bioorg. Med. Chem. Lett.* **11**, 2257–2261.

O'Brien, C. (1996) Protein fingerprints, proteome projects and implications for drug discovery, *Mol. Med. Today* **2**, 316.

O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins, *J. Biol. Chem.* **250**, 4007–4021.

Pappin, D.J.C., Hojrup, P., Bleasby, A.J. (1993) Rapid identification of proteins by peptide-mass fingerprinting, *Curr. Biol.* **3**, 327–332.

Patterson, S.D., Aebersold, R. (1995) Mass spectrometric approaches for the identification of gel-separated proteins, *Electrophoresis* **16**, 1791–1814.

Patton, W.F. (2002) Detection technologies in proteome analysis, *J. Chromatogr., B* **771**, 3–31.

Qi, S.Y., Moir, A., O'Connor, C.D. (1996) Proteome of Salmonella typhimurium SL1344: identification of novel abundant cell envelope proteins and assignment to a two-dimensional reference map, *J. Bacteriol.* **178**, 5032–5038.

Rosenfeld, J., Capdevielle, J., Guillemot, J.C., Ferrara, P. (1992) In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis, *Anal. Biochem.* **203**, 173–179.

Sechi, S. (2002) A method to identify and simultaneously determine the relative quantities of proteins isolated by gel electrophoresis, *Rapid Commun. Mass Spectrom.* **16**, 1416–1424.

Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Boucherie, H., Mann, M. (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels, *Proc. Natl. Acad. Sci. USA* **93**, 14440–14445.

Smolka, M.B., Zhou, H., Purkayastha, S., Aebersold, R. (2001) Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis, *Anal. Biochem.* **297**, 25–31.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T. (1988) Protein and polymer analyses up to m/z 100000 by laser ionization time-of-flight mass spectrometry, *Rapid Commun. Mass Spectrom.* **2**, 151–153.

Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., Davison, M. (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology, *Proteomics* **1**, 377–396.

Turecek, F. (2002) Mass spectrometry in coupling with affinity capture-release and isotope-coded affinity tags for quantitative protein analysis, *J. Mass Spectrom.* **37**, 1–14.

Unlu, M., Morgan, M.E., Minden, J.S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts, *Electrophoresis* **18**, 2071–2077.

Venter, J.C., Adams, M.D., Myers, E.W., et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Von Eggeling, F., Gawriljuk, A., Fiedler, W., Ernst, G., Claussen, U., Klose, J., Romer, I. (2001) Fluorescent dual colour 2D-protein gel electrophoresis for rapid detection of differences in protein pattern with standard image analysis software, *Int. J. Mol. Med.* **8**, 373–377.

Wagner, K., Miliotis, T., Marko-Varga, G., Bischoff, R., Unger, K.K. (2002) An automated on-line multidimensional HPLC system for protein and peptide mapping with integrated sample preparation, *Anal. Chem.* **74**, 809–820.

Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, R., Williams, K.L., Humphery-Smith, I. (1995) Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium, *Electrophoresis* **16**, 1090–1094.

Wilkins, M.R., Sanchez, J.C., Williams, K.L., Hochstrasser, D.F. (1996b) Current challenges and future applications for protein maps and posttranslational vector maps in proteome projects, *Electrophoresis* **17**, 830–838.

Wilkins, M.R., Sanchez, J.C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F., Williams, K.L. (1996c) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it, *Biotechnol. Genet. Eng. Rev.* **13**, 19–50.

Wilkins, M.R., Ou, K., Appel, R.D., Sanchez, J.C., Yan, J.X., Golaz, O., Farnsworth, V., Cartier, P., Hochstrasser, D.F., Williams, K.L., Gooley, A.A. (1996a) Rapid protein identification using N-terminal "sequence tag" and amino acid analysis, *Biochem. Biophys. Res. Commun.* **221**, 609–613.

Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J.C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., Williams, K.L., Hochstrasser, D.F. (1996d) From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis, *Bio-Technology* **14**, 61–65.

Willard, K.E., Anderson, N.G. (1981) Two-dimensional analysis of human lymphocyte proteins: I. An assay for lymphocyte effectors, *Clin. Chem.* **27**, 1327–1334.

Wimmer, K., Kuick, R., Thoraval, D., Hanash, S.M. (1996) Two-dimensional separations of

the genome and proteome of neuroblastoma cells, *Electrophoresis* **17**, 1741–1751.

Winocour, E. (1965) Attempts to detect an integrated polyoma genome by nucleic acid hybridization II. Complementarity between polyoma virus DNA and normal mouse synthetic RNA, *Virology* **27**, 520–527.

Yan, J.X., Wilkins, M.R., Ou, K., Gooley, A.A., Williams, K.L., Sanchez, J.C., Golaz, O., Pasquali, C., Hochstrasser, D.F. (1996) Large-scale amino-acid analysis for proteome studies, *J. Chromatogr., A* **736**, 291–302.

Yao, X., Freas, A., Ramirez, J., Demirev, P.A., Fenselau, C. (2001) Proteolytic ^{18}O labeling for comparative proteomics: model studies with two serotypes of adenovirus, *Anal. Chem.* **73**, 2836–2842.

Yates, III, J.R., Speicher, S., Griffin, P.R., and Hunkapiller, T. (1993) Peptide mass maps: a highly informative approach to protein identification, *Anal. Biochem.* **214**, 397–408.

Zhou, G., Li, H., DeCamp, D., Chen, S., Shu, H., Gong, Y., Flaig, M., Gillespie, J.W., Hu, N., Taylor, P.R., Emmert-Buck, M.R., Liotta, L.A., Petricoin, E.F., III, Zhao, Y. (2002) 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers, *Mol. Cell. Proteomics* **1**, 117–124.

Mass Spectrometry, High Speed DNA Fragment Sizing

Chung-Hsuan Chen
Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

1 Introduction 589
1.1 Time-of-flight Mass Spectrometry 590
1.2 Matrix-assisted Laser Desorption/Ionization (MALDI) for DNA Analysis 592

2 DNA Sizing and Sequencing by Mass Spectrometry 596
2.1 DNA Sizing by MALDI 596
2.2 DNA Sequencing 597
2.2.1 DNA Sequencing by Mass Spectrometry with DNA Ladders 600
2.2.2 Direct Sequencing 601

3 MALDI for Disease Diagnosis 603
3.1 Base Deletion 603
3.2 Point Mutation 606
3.3 Dynamic Mutation 607

4 DNA Typing for Forensic Applications 609
4.1 Forensic DNA Sample Preparation 610
4.2 Mass Spectrometry DNA Detection for Forensic Applications 611
4.2.1 RFLP and VNTR Analysis 611
4.2.2 Short Tandem Repeats (STR) 612
4.2.3 Single Nucleotide Polymorphism (SNP) 613
4.2.4 Gender Determination 615

5 Conclusion 616

Acknowledgment 616

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 7
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30549-1

Bibliography 616
Books and Reviews 616
Primary Literature 616

Keywords

Matrix-assisted Laser Desorption and Ionization (MALDI)
Biomolecules that are cocrystallized with a much larger amount of organic compounds on a substrate are laser-ablated to accomplish desorption and ionization for mass spectrometry detection.

Time-of-Flight Mass Spectrometer (TOF)
A mass spectrometer that is used to determine the mass-to-charge ratio of gas-phase ions by measuring the flight time in a drift tube for ions.

Gel Electrophoresis
Migration of charged molecules in an electric field with gel as a medium, which is obtained for separation and/or purification of biomolecules.

DNA Hybridization
Two single-stranded DNAs with complementary sequences combine to produce a double-stranded, with hydrogen bonding between two single-stranded DNAs. Hybridization is often used for DNA sequence identification. Microarray hybridization is often used for high throughput–DNA analysis.

DNA Sequencing
A process that determines the sequence of bases in a DNA segment.

Abbreviations

ASPCR:	Allele-specific polymerase chain reaction
CF:	Cystic fibrosis
CFTR:	Cystic fibrosis transmembrane conductance regulator
ds-DNA:	Double-stranded deoxyribonucleic acid
DNA:	Deoxyribonucleic acid
DRPLA:	Dentatorubral-pallidoluysian atrophy
EDTA:	Ethylene diamine tetraacetic acid
IR-MALDI:	MALDI with an infrared laser beam for desorption
MALDI:	Matrix-assisted laser desorption/ionization

MS:	Mass spectrometry
PCR:	Polymerase chain reaction
RFLP:	Restriction fragmented length polymorphism
ss-DNA:	Single-stranded deoxyribonucleic acid
SNP:	Single nucleotide polymorphism
STR:	Short tandem repeat
TOF:	Time-of-flight
TOFMS:	Time-of-flight mass spectrometry
UV-MALDI:	MALDI with an ultraviolet laser beam for desorption
VNTR:	Variable number of tandem repeats

Recently, high-speed DNA fragment sizing has been in critical need due to its important application on genomic function, DNA sequencing, disease diagnosis, DNA typing for forensic use, and microbial analysis for bioagent identification. Mass spectrometry has played a key role in rapid DNA sizing. This article discusses the fundamental principles of mass spectrometry for DNA analysis and the advantages and disadvantages of mass spectrometry compared to other approaches. It also gives a brief historical review of the development of mass spectrometry for DNA analysis. Applications in DNA sequencing, genetic disease diagnosis, and DNA finger printing are also presented.

1 Introduction

Since the discovery of the double-helix structure of DNA by Watson and Crick, research on DNA has been a key component of biological and medical research. The major functions of DNA comprise the transmission of information through a genetic code and self-replication. Many important advances have been made in understanding the complex organization of genetic material in cells and applying genetic engineering to manipulate the replication capability. However, because of the extreme complexity of the human genome, a worldwide Human Genome Project on mapping and sequencing the entire human genome was initiated in 1988 and has been more or less completed recently. The interest in DNA sequencing has expanded to include microbial genome, mouse genome, rice genome, and others. It is also expected that the need to sequence and resequence known genes will be increased by many orders of magnitude in the coming decades. Nevertheless, nearly all the data relating to sequencing of DNA have been obtained by gel electrophoresis, which is still relatively slow and somewhat expensive. It is clear that faster and cheaper methods are in critical need. For most resequencing work, the sequencing speed and cost can be more critical than the size of DNA that can be sequenced. Mass spectrometry for DNA sizing and analysis can serve this purpose because of its short analysis time.

1.1 Time-of-flight Mass Spectrometry

The conventional method used for DNA measurement is gel electrophoresis. In this process, an electric field is applied to the gel medium to cause the drift of DNA fragments, which typically exist as charged particles. The drift velocity tends to be lower as DNA fragments become larger. However, this process is slow and needs hours to separate different sizes of DNA fragments. The DNA-sizing speed by capillary gel electrophoresis is significantly faster than slab-gel electrophoresis. Nevertheless, mass spectrometry DNA sizing is much faster than either of the methods of electrophoresis. Mass spectrometry has been used for several decades to measure the molecular weight of gas samples by mass-to-charge ratio. In order to use a mass spectrometer for molecular weight determination, the relevant molecules need to be produced in gas phase. The pressure inside the mass spectrometer needs to be less than 10^{-4} Torr in order to prevent collisions between the ions and other gaseous molecules. Owing to the practically zero-vapor pressure of DNA at room temperature, a method of producing DNA gas-phase molecules is required. Desorption of DNAs by a laser beam oran ion beam is often needed. Another approach is to introduce solvated DNA samples through electrospray or ionspray processes. Since mass spectrometry is based on the detection of gas-phase ions, a means to produce such ions is obviously needed. Ions can be produced by several different processes. These processes include charged particle ionization, photoionization, chemi-ionization, and electron-impact ionization. All these processes involve the removal of an electron from a molecule to form a positive ion. However, positive or negative ions can also be produced by attaching or detaching a charged particle such as a proton. This process is quite often achieved by a chemical process. Once the ions have been produced, various sizes of ions may be separated by electric or magnetic fields. The mass-resolved ions are then accelerated to impinge on a solid surface of low work function to generate secondary electrons, which are subsequently detected by an electron multiplier, or a channeltron or microchannel plates (MCP). Since these electron detection devices can have an amplification factor of 10^8, even a single secondary electron can be detected.

Owing to the similarity of principles between a time-of-flight mass spectrometer (TOFMS) and a gel-electrophoresis device, a time-of-flight mass spectrometer is most often used for DNA detection. A TOFMS is a device in which ions of different masses are given the same energy and are allowed to travel in a field-free space. Because of their different velocities, the ions of differing masses arrive at the end of the field-free space at different times, where an ion detector produces a time-varying electrical signal proportional to the number of ions impacting it. The amplified current signal is typically displayed on a fast digital oscilloscope, where the molecular weights can be determined by measuring the time interval between the creation of ions and the detection of ions. The ions need to be produced in a very short time interval and in a very small volume in order to achieve high resolution mass spectra. Thus, ionization by a short-pulsed laser beam is an ideal choice for a TOFMS. A schematic of a typical TOFMS is shown in Fig. 1. Ions are produced by a short-pulsed laser beam, which is typically focused to a small area in the ionization region. Ions

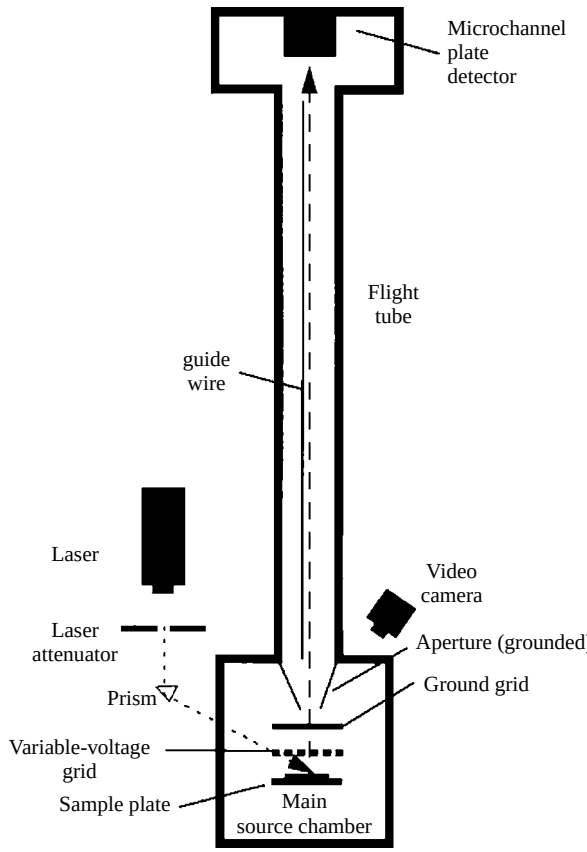

Fig. 1 Schematic of a typical MALDI Time-of-flight mass spectrometer. A laser beam is focused onto a sample for desorption and ionization through protonation/deprotonation process. UV, IR, and visible lasers have all been used for MALDI. However, most works have been done with UV-laser desorption. A video camera is typically used to observe the sample location so that the laser beam can be focused on the desired location. Pulsed-ion extraction is often used to enhance mass resolution. A guide wire is used to achieve the oscillation of desorbed ions to reduce the desorbed biomolecular ions hitting the wall of vacuum chamber. A charged particle detector such as microchannel plate is used for ion detection. Most commercial instruments have sample holder with array configuration so that multiple samples can be placed in the mass spectrometer for analysis.

are then accelerated to a uniform energy by an electric field before they go into the free drift zone. An electron multiplier or a microchannel plate is placed at the end of the drift tube to detect the ions. Since ions have different initial velocities when they are produced, the higher the acceleration energy, the lesser the effect the initial

velocity spread will have on the flight time, and the higher the mass resolution will be. On the other hand, a higher voltage gradient in the area of the ionizing laser will result in ions produced in different parts of the laser beam being accelerated through different potentials, reducing the resolution of the flight times. Better resolution is achieved with a smaller ratio of ionization radius to acceleration length. There are several ways to improve the resolution of a TOFMS. A popular approach is the reflectron TOFMS. An electrostatic lens to bend ion trajectory is typically placed at the end of the drift tube to *reflect* the ions toward the detector. This design can be used to compensate for either the initial velocity spread or the spatial spread of the ions. However, it is difficult to compensate for both velocity spread and spatial spread. Resolution ($M/\Delta M$) of 10^5 for a reflectron TOFMS has been achieved by a few groups of researchers. One major advantage of a TOFMS over other types of mass spectrometers is its speed in measuring all the masses in a sample in a very short period of time, and another is its capability of measuring ions with very high molecular weights. In principle, there are no limitations on the size of molecules that can be measured by a TOFMS if the ions can be accelerated to a high enough velocity that secondary electrons can be ejected due to the ion's impact on the ion detector. This property is critically important for the detection of DNA fragments since most DNA fragments have very high molecular weights.

A mass spectrometer can be used for molecular weight determination only under the condition that the molecular ions are produced in space. Thus, a mass spectrometer, usually, can only be used for materials with vapor pressure higher than 10^{-14} Torr at room temperature. Although the minimum vapor pressure required is very low, the vapor pressures of many materials at room temperature are even lower (lower than 10^{-14} Torr). DNA fragments are in this category. In order to use mass spectroscopy for DNA measurement, it is necessary to develop a method that places DNA fragments in space without breaking them up. Recent developments in laser desorption have been significantly successful in achieving this. A frozen double-stranded DNA fragment on a copper plate was successfully desorbed by a dye-laser beam. Gel electrophoresis was used to prove the success of laser desorption of a large amount of DNA. However, the use of mass spectrometry for DNA analysis requires both desorption and ionization.

1.2 Matrix-assisted Laser Desorption/Ionization (MALDI) for DNA Analysis

Since a mass spectrometer can only be used to detect ions, some process of ionization is necessary if only neutral DNA segments are produced in the desorption process. Ionization can be achieved either by electron detachment to form a neat positive ion or by electron attachment to form a neat negative ion. Ions can also be produced by protonation to form $(M + H)^+$ or deprotonation to produce $(M - H)^-$. Since the size of a DNA segment is usually large and some of the DNA bonding (such as the glycosidic bond) is fragile, successful soft ionization to produce parent ions without breakup is a challenge. In addition to the *soft* desorption and *soft* ionization, the third major problem is the difficulty in detecting very heavy ions. In order to use an electron multiplier or an MCP in the detection of an

ion efficiently, the ion velocity needs to be greater than 1.5×10^6 cm/s. For example, a 1000-bp DNA segment has a molecular weight of ~600 000 daltons. These ions need to be accelerated to 700 KeV in order to reach a velocity of 1.5×10^6 cm/s. Most mass spectrometer designs are not suitable for accelerating very heavy ions to the velocity required for detection. Recently, the secondary electron emission efficiency at various velocities for a few bio-organic molecules was measured and the conclusion reached was that the threshold velocity could be somewhat lower than 1.5×10^6 cm/s for larger molecules. The successful detection was attributed, at least partially, to the ejection of smaller ions when large ions hit the metal plates. Nevertheless, the low detection efficiency for very large molecular ions is a concern when using mass spectrometry for the detection of very large DNA segments.

It is clear that the use of mass spectrometers for DNA detection will require the solution of several challenging problems. However, there is a major advantage in using mass spectrometry because of the potential to achieve ultrafast speed. Using gel electrophoresis for DNA separation typically takes a few hours. If radioactive tagging is used, it can be done overnight. However, a mass spectrum can be obtained within one second. For example, a TOFMS usually needs less than one millisecond to separate different sizes of DNA fragments; thus speed is improved by many orders of magnitude. In addition, hundreds of samples can be placed in a mass spectrometer for fast analysis if an array sample holder is designed to hold multiple targets. For gel electrophoresis, tagging of radioactive materials or organic chromophore to DNA fragments is required for detection. Thus, radioactive or chemical wastes are produced. Mass spectrometry analysis eliminates the need for tagging; therefore, no waste will be produced. Although a mass spectrometer can be more expensive than a gel electrophoresis setup, the much faster speed of analysis by a mass spectrometer can drive the cost much lower on a per sample basis. For example, 1000 samples of DNA fragments can possibly be analyzed by a TOFMS within an hour. This can be particularly important for medical applications, since it may be necessary for large hospitals to analyze thousands of samples each day.

For the analysis of materials with no vapor pressure at room temperature, several desorption methods have been used to carry the analyte into space for mass analysis. Major approaches include desorption by fast-atom bombardment (FAB), secondary-ion bombardment, and laser desorption. However, none of these have been very successful for analysis of large biomolecules. Since most biomolecules are fragile, when subjected to heat or bombardment, a fragmentation process almost always occurs. However, in 1987, Hillenkamp and his coworkers discovered that large protein molecular ions can be produced without much fragmentation by laser desorption if these biomolecules are mixed with smaller organic compounds that serve as a matrix for strong absorption of a laser beam. This process is now called *matrix-assisted laser desorption* (MALD). The typical preparation technique for MALD is to dissolve biomolecular samples in solution, then prepare another solution that contains small organic compounds such as 3-hydroxypicolinic acid (3-HPA). These two solutions are subsequently mixed, and a small amount of the resulting solution is placed on a metal plate to dry. After the crystallization of the sample, the sample plate is placed in

the mass spectrometer for analysis. The molar ratio of matrix-to-analyte is typically more than 100 to 1. During a MALD process, matrix materials strongly absorb the laser energy and become vaporized, and large biomolecules are carried outward during the fast-vaporization process. Large biomolecules can be delivered into space without breakup, probably due to minimal direct absorption of laser energy; thus, *soft* desorption can be achieved. However, it has been found that biomolecule parent ions are also produced during the MALD process in addition to the expected neutral molecules. Thus, these desorbed ions can be directly detected by a mass spectrometer. This process involving matrix-assisted laser desorption and ionization at the same time is referred to in an abbreviated form as MALDI (matrix-assisted laser desorption and ionization).The mechanism of ion production has been speculated by many researchers to involve proton transfer.

Up to now, one major disadvantage of MALDI for DNA analysis has been its relatively poor mass resolution especially for large DNA fragments. In order to study the cause of poor mass resolution, velocity distributions of desorbed oligomers were studied, and it was found that the distribution more or less follows the distribution of the matrix. In general, the mass resolution of MALDI spectrum of DNA is relatively poor compared to that of small organic compounds. This relatively poor resolution is due to (1) high initial velocity distribution, (2) the inhomogeneity of DNA samples on metal plates, (3) the adducts of matrix materials or metal ions to DNAs, and (4) fragmentation of DNA molecules due to the laser ablation process. Since the velocity distribution of the DNAs are passively carried out by matrix materials, the velocity spreads are expected to be similar to the velocity distribution of the matrix material. If the molecular weight of a DNA fragment is 200 000 daltons and the molecular weight of the matrix is 100, the desorbed energy of DNA molecules can reach 50 eV, even assuming that the average kinetic energy of desorbed matrix material is near thermal. Thus, the broad distribution of ion energies of desorbed biomolecules is one important factor that limits the resolution. When a biomolecule sample mixed with matrix materials begins to dry on the plate, the process of crystal growth makes the sample distribution inhomogeneous. The thickness of the sample can also cause lower resolution. However, a reflectron TOFMS can compensate for the ionization volume to achieve a resolution higher than 1000. Since most spectra by MALDI have resolutions of only a few hundred, the limited resolution is primarily a result of adduct and fragment formation. If a biomolecule can be attached to a matrix molecule, or part of a matrix molecule, the resolution then depends on the molecular weight of the attached particle. It is also not unusual to observe alkali metal ions attached to DNA molecules. The adduct formation is strongly dependent on the matrix used. Detection sensitivity for small DNA segments (<10mer) by MALDI can reach 1 femtomole. The bigger the size of the DNA segment, the lower is the detection sensitivity.

The chemical characteristic of matrix material used in MALDI for DNA segment analysis is probably the single most critical parameter. Only a few compounds have been found to be useful matrices for oligomers. A good matrix material needs to have good solubility in the solvents used for the analyte, strong absorption at the laser wavelength, and no chemical reactivity with the analytes. In general, the matrix needs to dissolve in the solvent

in concentrations higher than 0.01 M solution. Otherwise, a high molar matrix-to-analyte ratio is difficult to achieve. The absorption coefficient at the laser wavelength needs to be higher than 3000 L $mole^{-1}cm^{-1}$; otherwise, the laser energy deposition cannot be concentrated in the top few hundred molecular layers to produce efficient desorption. It is clear that the matrix materials used cannot chemically react with the analytes. For example, strong acids tend to remove purine bases. Thus, they would not be good matrices. The glycosidic bonding of DNA is fragile and easily subject to chemical interaction. This can be a factor that makes MALDI for DNA analysis much more difficult than for protein analysis. Although a few criteria can be set up to rule out certain chemicals as good candidates for MALDI of DNA segments, the search for good matrices is still more or less trial and error.

In addition to the chemical properties of a matrix material, the amount of energy deposited by the laser is also critical for achieving good spectra. It has been shown that the relationship between biomolecule ion production and the intensity of the laser pulse is highly nonlinear. Below a threshold, no biomolecule ion signal is produced. Above a threshold, the ion production increases rapidly to a plateau. The threshold for ion production of DNA segments was found to be a few megawatts per square centimeter. However, ion signals tend to disappear when the laser intensity is so high that intense plasma is produced.

Detailed mechanisms for ion production of biomolecules by MALDI are still not well understood. However, a simple model of photoionization and photochemistry processes has been proposed to explain the production of biomolecular ions. The model suggests that a multiphoton absorption process leads to the ionization of matrix molecules, which have subsequent chemical reactions with biomolecule analytes to form ions.

The existence of metastable states of negative oligomer ions during the MALDI process has been experimentally proved. The lifetimes of small negative oligomer ions were measured to be a few hundred microseconds. The lifetimes become shorter for larger negative oligomer ions compared to that of smaller ones. These results indicate that the detection of large negative oligomer ions will be very difficult for Fourier transform, ion-cyclotron, reflectron TOF and ion-trap mass spectrometers with MALDI because of the long ion lifetimes inherently required by these instruments.

Although MALDI has been applied to detect large biomolecules with a good determination of molecular weights, it has been difficult to achieve quantitative measurements of biomolecules. The primary reasons for poor reproducibility are related to the inhomogeneous distribution of samples on substrates and the high sensitivity to changes in laser fluence. A small change in laser fluence can cause a very large fluctuation in analyte signals. When samples are prepared in a MALDI experiment, the crystallization process can make the distribution of matrix materials inhomogeneous. For example, nonhomogeneous needle-shaped crystals were observed when using 2.5-dihydroxybenzoic acid as a matrix. The signal level of analytes can be a strong function of the exact spot at which the laser desorbs. It is also known that an impurity in a MALDI process can reduce the signal levels of analytes significantly. In addition, the surface condition of the substrate can make a significant change in the threshold for the

production of plasma. In brief, it is difficult to obtain reliable measurements of absolute quantities of analytes by MALDI. If MALDI is to become a routine analytical tool, it is obvious that its capability to make quantitative measurements of analytes must be improved. When the quantity of the matrix is fixed, signals from oligomers are not linearly proportional to the quantity of the analyte. This very nonlinear relation between ion signal and analyte is mostly due to the number of layers of analyte absorbing laser energy and the detailed interaction between DNA and matrix molecules during the crystallization process. Too much of the analyte on the samples may not produce bigger signals. However, if a known biomolecule with similar chemical properties is used for an internal calibration, the signal level can become more or less independent of the laser fluence and the inhomogeneous distribution of matrix materials.

2 DNA Sizing and Sequencing by Mass Spectrometry

2.1 DNA Sizing by MALDI

Since the discovery of MALDI, many research groups have succeeded in measuring various proteins and large organic compounds by MALDI. MALDI has also been applied to DNA segments. Initially, success was limited to the detection of small DNA fragments. Then, 3-hydroxypicolinic acid (3-HPA) was found to be a good matrix for oligonucleotide detection. Both positive and negative ions were observed using 3-HPA as matrix (Fig. 2). Restriction fragment length polymorphism (RFLP) and polymerase chain reactions (PCR) are two of the most important DNA analysis methodologies discovered in the past few decades. PCR is broadly used in

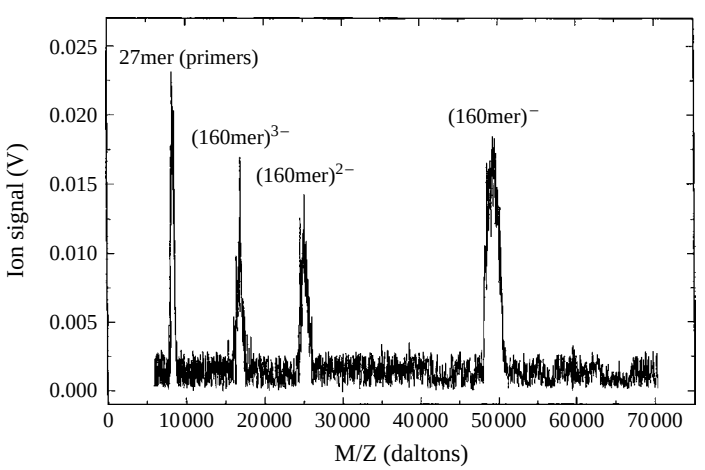

Fig. 2 Negative-ion mass spectrum of a 160-bp DNA amplified from pLB 129. The desorption laser was a fourth harmonic of a Nd-Yag laser with wavelength at 266 nm. In addition to single charged ions, doubly and triply – charged ion peaks were also observed. Although double-stranded DNA was used, only the peaks corresponding to single-stranded DNA ions were observed. It indicates that ds-DNA was probably dissociated into ss-DNA due to the interaction with acidic matrix molecule.

nearly every DNA laboratory to replicate a required segment of DNA from genomic DNAs. MALDI mass spectrometry has also been applied for RFLP and PCR products analysis. The success in using mass spectrometry for measurements of RFLP and PCR products paved the way for several important applications including disease diagnosis and DNA typing for forensic applications. With the development of an instrument to give high ion energy and the use of mixtures of matrices, MALDI was successfully used to detect double-stranded DNA (ds-DNA) of 500 bp, which was produced by a PCR amplification process (Fig. 3). However, only a single-stranded DNA was observed. This marked the first success of using MALDI for a large DNA fragment. Recently, the detection of DNA with the size of 3199 bp was reported (Fig. 4). It can be concluded that large DNA fragments can be measured by MALDI. However, the mass resolution for large DNA fragments (>300 nucleotides) is still very poor ($M/\Delta M < 50$). RFLP measurement by MALDI for S16 rDNA for a microbe has also been successfully demonstrated. (Fig. 5). It indicates that mass spectrometry is emerging as a valuable tool for DNA sizing for various applications.

2.2 DNA Sequencing

DNA sequencing is a method that not only determines the size of a DNA fragment, but also its structure. DNA contains four primary bases – adenine (A), guanine (G), cytosine (C), and thymine (T). Their chemical structures and molecular weights are shown in Fig. 6. Since there is a mechanism for translating the nucleotide sequence in DNA into the amino acid sequence of protein, DNA sequencing is critically important for biological and medical research. Current methods of DNA sequencing such as those of Sanger and Maxam-Gilbert have proven to be very useful for sequencing small DNA segments. These methods of sequencing usually

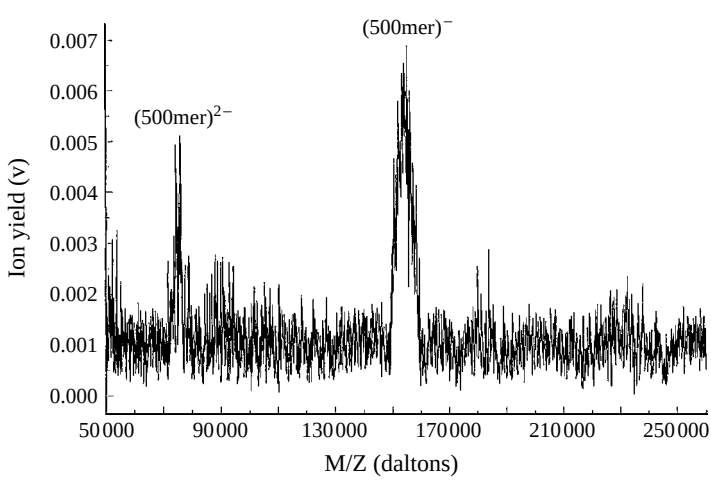

Fig. 3 Mass spectrum of a 500-bp DNA amplified by PCR from bacteriophage lamda genome. The laser fluence used was 200 mJ cm^{-2}. No peaks corresponding to ds-DNA were observed.

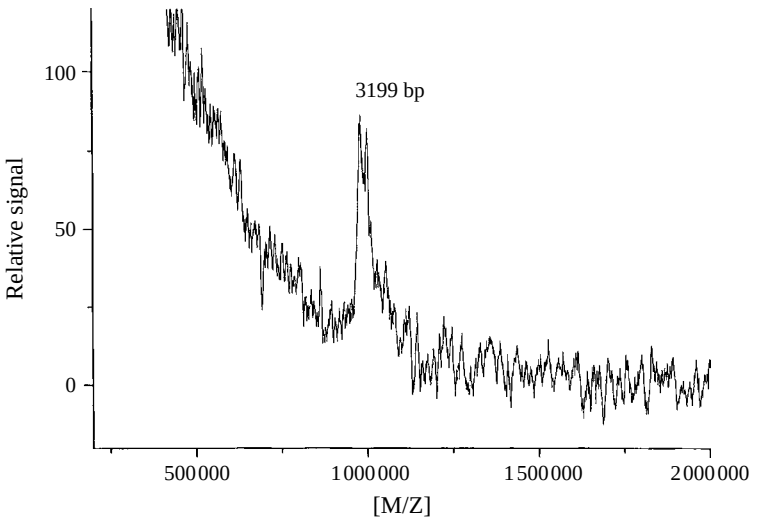

Fig. 4 Mass spectrum of a large DNA fragment of 3199 bp. It clearly indicates that large DNA segments can be detected by MALDI. However, the detection efficiency is very low due to the low yield of secondary electrons when the ion velocity is very low ($< 1 \times 10^5$ cm s^{-1}).

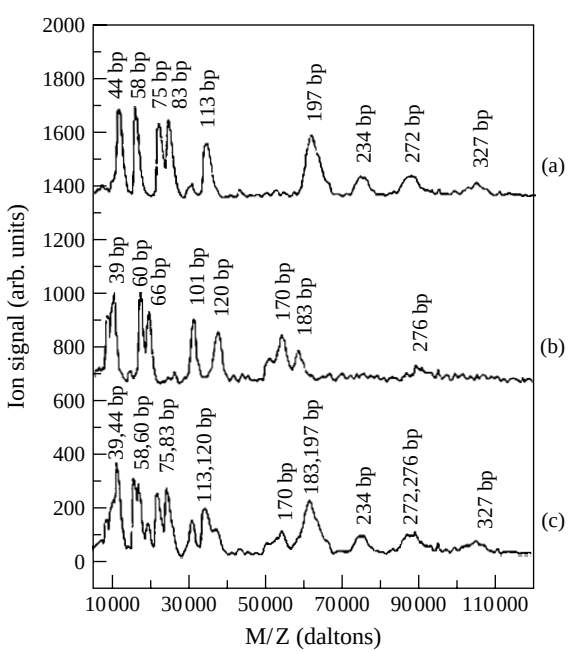

Fig. 5 MALDI spectra from enzyme digestion of PCR products of 16S rDNA from various microbial samples. (a) and (b) show the mass spectra of DNA fragments from RFLP with enzyme digestion by *Hha*I and *Has*III and *Msp*I and *Rsa*I, respectively. The pattern of fragmentation can be used to help in bacteria identification. (c) shows the results of DNA fragments from the mixtures of RFLP from samples contributing to (a) and (b).

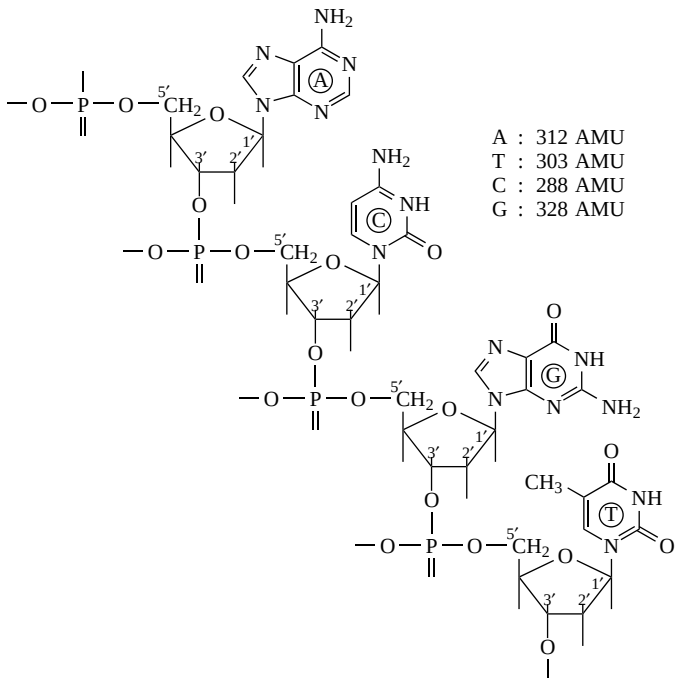

Fig. 6 Chemical structures and molecular weights of adenine (A), thymine (T), guanine (C) and cytosine (C).

involve labeling fragments of DNA for identification following gel electrophoresis. Either radioactive labels such as ^{32}P or ^{35}S, or chemical labels such as fluorescent dyes are currently used. At the present time, routine DNA sequencing in most laboratories relies on the use of radioactive isotopes. The fragments are generated either chemically or enzymatically to represent all possible positions of each of the four nucleotides (A, G, C, and T). Following electrophoresis, the radioactive labels are located by autoradiography. This method is very time-consuming. With radioactive labeling, four lanes of gel are usually required to separate the fragments from a given DNA segment that terminates in A, G, C, or T. Substitution of fluorescent labels for radioactive isotopes allows the DNA fragments to be detected continuously during electrophoresis. This method usually employs four different fluorescent labels. Using these four labels, all four types of DNA fragments (A, G, C, and T) can be run in one electrophoresis lane. The uncertainty in comparing the label positions in four adjacent lanes of the gel is eliminated. Automatic sequencing instruments based on fluorescence measurements are currently commercially available. Recently, capillary gel electrophoresis coupled with the fluorescence method has also been successfully used for improving sequencing speed. However, both radioactive methods and fluorescent dye labeling methods for DNA sequencing require the use of the time-consuming gel electrophoresis. Thus, it is natural to consider using TOFMS for DNA sequencing to reduce

the sequencing time. Instead of taking several hours by gel electrophoresis, the TOFMS takes less than one millisecond to finish the separation of different sizes of DNA segments. However, the parent DNA ions must be produced with high efficiency and without serious fragmentation. Since the MALDI process has been successfully used in producing parent ions of small DNA segments without serious fragmentation, it can have an important role in the future of fast DNA sequencing especially for re-sequencing.

2.2.1 DNA Sequencing by Mass Spectrometry with DNA Ladders

DNA sequencing has been broadly used for biomedical research and clinical applications during the past two decades. Rapid and reliable DNA sequencing can also be very valuable in forensic applications. In the conventional sequencing approach, different sizes of DNA ladders, which are produced by either Sanger's enzymatic method or Maxam–Gilbert's chemical cleavage method, are separated by gel electrophoresis to achieve sequencing. With the use of MALDI for sequencing, the speed can be significantly greater than with gel electrophoresis. In addition, MALDI sequencing does not require labeling for identification, which saves both time and cost.

During the past few years, mass spectrometry for sequencing short DNA fragments has been pursued. In 1995, the first sequencing of a 45mer ss-DNA using MALDI-TOF with enzymatic preparation of DNA ladders was demonstrated. Recently, sequencing of ss-DNA higher than 100 nucleotides by using cycle sequencing to produce DNA ladders was achieved. Figure 7 shows the negative-ion mass spectra of DNA ladders from A, C, G, and T reactions using double-stranded DNA 130 bp as the template with reverse primer. Forward primers were also used for sequencing with results similar to the data shown in Fig. 7. Since sequencing by using both forward and backward primers

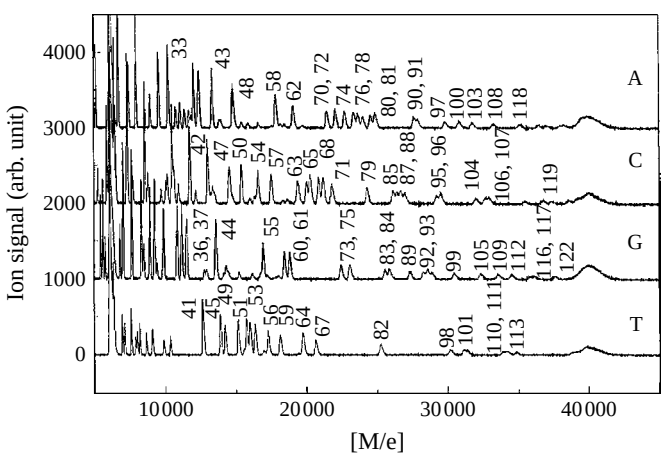

Fig. 7 Mass spectra of Sanger sequencing. DNA ladders were produced by dideoxynucleotide chain termination of the PCR product of a double-stranded template. Owing to the limitation of mass resolution ($M/\Delta M$), the largest ss-DNA can be sequenced is limited to $\sim$ 100 nt.

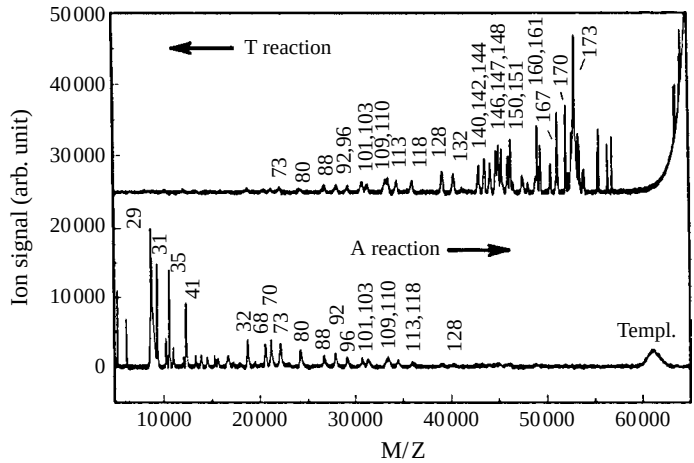

Fig. 8 Negative-ion mass spectra of Sanger sequencing. Ladders were produced by a dideoxynucleotide-chain termination of the PCR product of a double-stranded 200-bp template. The results were from the A- reaction with the reverse primer and T-reaction with forward primer.

can reach 120mer, complete sequencing of 200-bp DNA can be achieved. Figure 8 illustrates the idea of combining the spectra from forward and backward primers to sequence ds-DNA with a 200-bp template. Since mass is measured to identify the size of DNAs, there is no concern about a missing band, which can occur in the gel electrophoresis method. Thus, redundant sequencing can possibly be eliminated using MALDI DNA sequencing.

At present, automatic gel sequencers can read up to 1000 bp, but most routine sequencing ranges from 300 to 500 bases. Since the size of DNA to be sequenced by mass spectrometry is smaller than in the gel method, MALDI may not be an efficient method for *de novo* DNA sequencing for genome projects. However, with the sequencing of many different genomes completed, mass spectrometry is emerging as a useful tool for DNA resequencing due to its speed and the absence of a requirement for DNA labeling.

2.2.2 Direct Sequencing

When using mass spectrometry for sequencing DNA with ladders, the time needed for a TOFMS to separate and detect various ladders is only a few seconds. However, the time needed to prepare DNA ladders is many minutes to hours. Thus, the ideal sequencing method would be to use mass spectrometry to carry out direct DNA sequencing without the need to produce DNA ladders by chemical or enzymatic methods. This also implies that DNA ladders need to be produced during the laser ablation process.

Direct DNA sequencing can be easily understood from the illustrations in Figs. 9 and 10. Assuming the selective cleavage of P–O bond at 3′-linkage for an oligonucleotide of 5′-CTGTGA-3′, the primary fragmentation of such breaking of this bond will result in two series of fragments. These are labeled as 5′- and 3′-termini (Fig. 9). Each series has six members ranging from 1- to 6-mers. The vertical dotted line in Fig. 9 represents the position of the

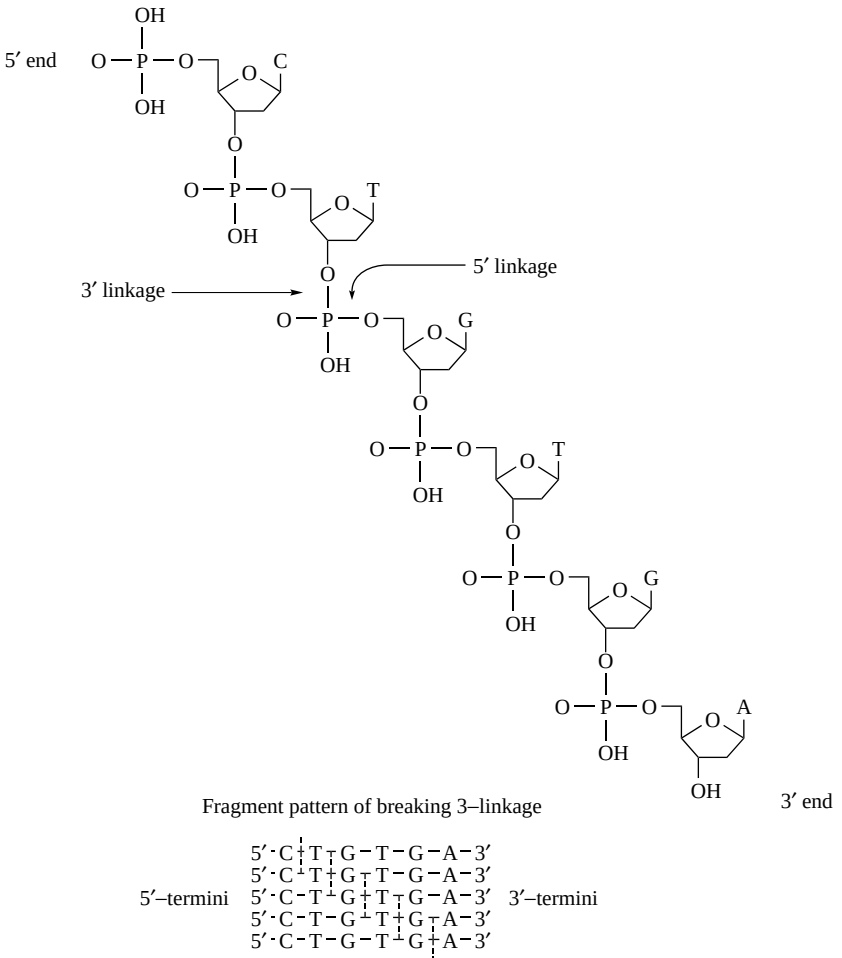

Fig. 9 Illustration of the selective cleavage P–O bond and the fragmentation pattern for a 5′-CTGTGA–3′ oligonucleotide.

cleavage. The fragments in both series can be ionized and resolved in MALDI spectra. Figure 10 shows the simulated mass spectra for these two series. Trace 'a' in Fig. 10 is the spectrum of the series with 3′-termini. The mass difference between two adjacent peaks in the series provides the information for each extra base. For example, the first member of this series is A, and the second member is GA. Owing to the selective cleavage of 3′-linkage, the mass difference between these two peaks will be exactly 329.2 daltons, which represents the mass of dGMP. The sequence information can be obtained by analyzing all values of mass difference in this series together with the total mass of 5′-CTGTGA-3′. The same information can also be obtained from the series with 5′-termini (Fig. 10, trace 'b'). If a complete series (either 5′- or 3′- termini) of such fragmentation for a DNA sample can be

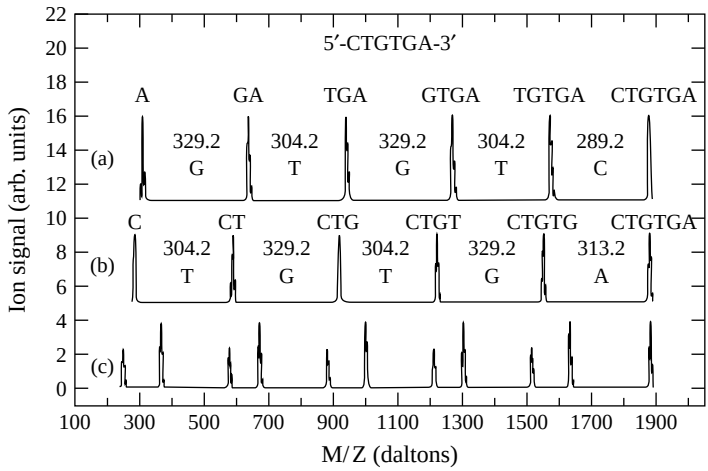

Fig. 10 Simulated spectra of the selective cleavage of P–O 3′-linkage for 5′-CTGTGA-3′ oligonucleotide: (a) simulated spectrum of fragment series with 3′-termini by the 3′-cleavage; (b) series with 5′-termini by 3′ cleavage; and (c) is the combination of (a) and (b).

experimentally resolved, the full sequence information will be obtained. Trace 'c' of Fig. 10 shows a mass spectrum of two series from the cleavage of 5'-linkage P–O bond. This sequencing method is very reliable since the sequence result from one series can be reconfirmed by the results from other series. Recently, UV-MALDI was attempted in order to achieve direct DNA sequencing by carefully adjusting the experimental conditions, which include the use of newly discovered two-component matrices and adequate laser fluence. Figure 11 shows the direct sequencing of DNA probes up to 35 nucleotides. Similar results by IR-MALDI and by electrospray high-resolution Fourier transform mass have also been reported. At present, direct sequencing is still limited to short oligonucleotides, owing to the limited mass resolution, since the mass difference between different bases needs to be identified. If a 60mer oligonucleotide is to be sequenced, the accuracy of the ion peaks needs to be better than 1 in 2000 since the mass difference between A (adenine) and T (thymine) is only 9 daltons. When a DNA segment is longer than 100mer, the effect of the abundance of ^{13}C isotope makes the distinction between A and T extremely difficult.

3 MALDI for Disease Diagnosis

DNA mutations that causes diseases can be simply classified into three major categories. These are (1) deletion or insertion, (2) point mutation, and (3) dynamic mutation. MALDI has been applied to all these different categories.

3.1 Base Deletion

In 1994, mutations in the cystic fibrosis (CF) gene in clinical samples were first demonstrated. Several measurements of CF were also carried out by mass spectrometry. In the North American

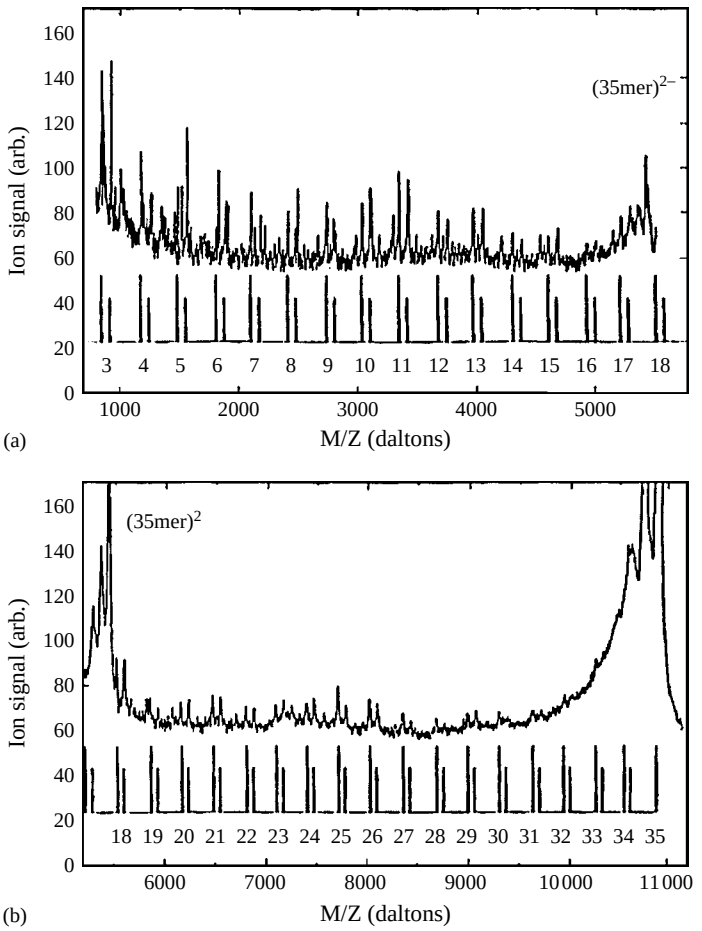

Fig. 11 MALDI mass spectrum with selective fragmentation for direct sequencing of a 35-nt oligonucleotide, 5′-GCGTGATGGAATCGATGACGTGCGATGTCGTGTTT-3′ The higher intensity series in the simulated spectra is 5′ termini and 3′ cleavage series and the lower intensity is 3′ termini and 3′ cleavage series. A mixture of 3-HPA and 2,4,6- trihydroxyacetophenone was used as a matrix. (a) and (b) are for low and high mass regions respectively.

population, about 70% of CF carriers have a 3-bp deletion in exon 10, resulting in the loss of phenylalanine residue at codon 508 (ΔF508). Two oligonucleotide primers, CF1 and CF2, were designed to amplify a DNA segment spanning the deletion, thus generating a 59-bp and 56-bp fragment of the normal CF gene and ΔF508 mutation, respectively (Fig. 12). Analysis of the PCR amplified products by MALDI resolved the 3-bp difference between the normal and the ΔF508 alleles in the spectra shown in Figure 12. To validate the efficiency of MALDI-TOFMS, a total of 30 genomic DNAs from patient samples were used as templates to amplify DNA segments

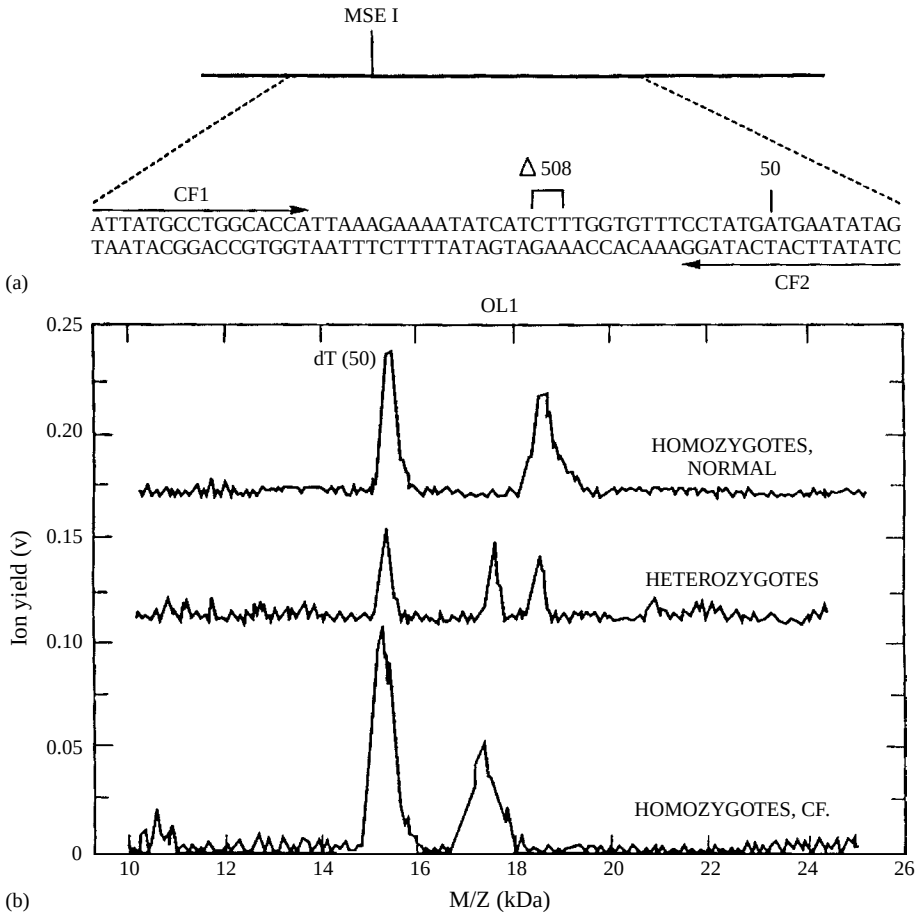

Fig. 12 Differential diagnosis of individuals with the normal and Δ F508 alleles by MALDI-TOF mass spectrometry. (a) Nucleotide sequence of 59 bp in exon 10 of the CFTR gene with the primers CF1 and CF2 indicated by arrows. A deletion of 3 bp in Δ F508 mutation is marked by the bracket. The DNA fragment amplified by PCR with forward and backward primers is schematically represented by the horizontal line on the top of the nucleotide sequence with a unique MseI site indicated. (b) MALDI-TOFMS of DNA amplified from a healthy volunteer, a Δ F508 carrier, and a CF patient. A synthetic dT_{50} was included as an internal standard for mass peak calibration.

from exon 10 or other exon/intron regions (as negative controls) of the CFTR gene. The samples were coded numerically and amplified by PCR. All the samples with ΔF508 mutations were verified earlier by size fractionation in a 10% acrylamide gel or by sequencing analysis. In a double-blind study (both PCR and mass analyses were performed without knowing whether the CFTR was normal or mutated), results from the mass spectrometry and conventional assays were in total agreement. These results indicate that MALDI mass spectrometry has the potential to become a valuable method of clinical CF diagnosis.

3.2 Point Mutation

Most DNA mutations due to contaminants are point mutations. Thus, the capability to detect point mutations by mass spectrometry is essential. MALDI has also been applied for point mutation detection. The approach is to amplify the desired region of a DNA template by PCR using two primers that have their 3'-ends extended to the site of expected mutation. To test this approach, a synthetic oligo with 60 nucleotides, whose sequence is derived from human p53 mRNA from codon 144 to 163, was used as a template with two normal allele-specific primers in a PCR (ASPCR). The lengths of the primers used were 16 and 23 nucleotides. By using a normal template and normal allele-specific primer pair, a normal PCR DNA product is expected to be produced. The size of the expected PCR product should be equal to the sum of both primers minus one. On the other hand, little PCR product will be made if the normal template is replaced with the mutant template, because the 3′-end base of normal primers results in a mismatch with the mutant template. Since Taq DNA polymerase lacks a 3′ exonucleolytic proofreading activity, the mismatch will block the amplification during PCR. The sequence of the template and the primers used were

Primer 1 (23mer)

5'-CAGCTGTGGGTTGATTCCACACC-3'

Template (60 bp)

5'-CAGCTGTGGGTTGATTCCACACCC
CCGCCCGGCACCCGCGTCCGCG
CCATGGCCATCTAC-3'

3'-GGGGCGGCCGTGGGC-5'

Primer 2 (16mer)

As shown in Fig. 13, the PCR products with the expected size of 38 nucleotides could be clearly observed. On the other hand, no PCR products were detected when the normal templates were replaced with mutant templates (Fig. 14). Since mutation of the p53 gene has been

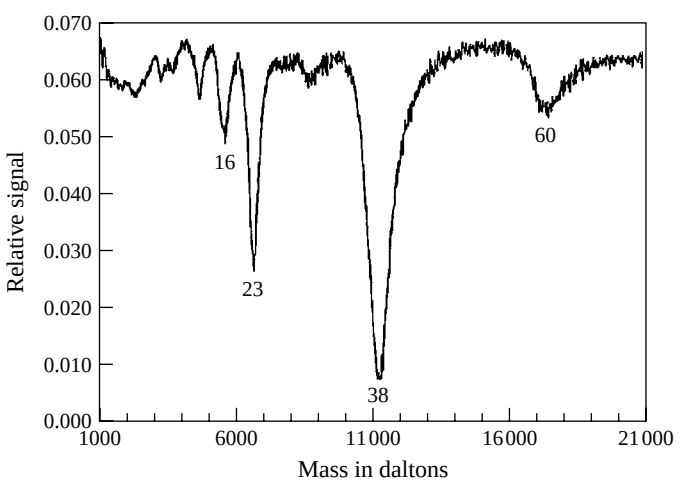

Fig. 13 MALDI mass spectrum of 38-bp PCR product with two primers overlapping at one base. The lengths of primers are 16 and 23 nucleotides.

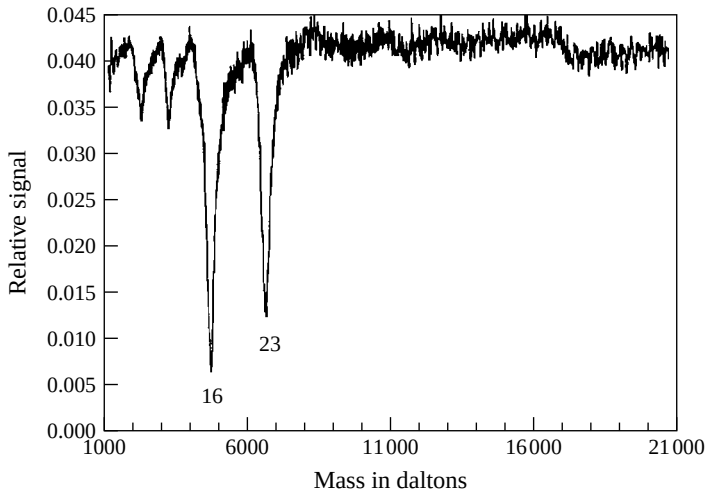

Fig. 14 MALDI mass spectrum with primers at the mutation site not complementary to the template. No PCR products, namely 38 bp were observed.

found to be responsible for more than 60% of cancers, the capability to detect p53 mutation is valuable for giving early warning of possible cancer development.

Measurements of G551D point mutation in cystic fibrosis transmembrane conductance regulator (CFTR) have also been attempted. The G551D mutation in the CFTR gene involves a G to A mutation, which results in a glycine to a spartic amino acid substitution. The approach is similar to the one adopted for point mutation in the synthetic template simulating a portion of the p53 gene. If the two primers based on the normal sequence match the target DNA sequence, a normal PCR product will be produced. However, if the alternate primers that match the mutant hybridize, a PCR product specific for mutant sequencing will be produced. Thus, the mass spectrometer can be used to identify people who are homozygous-normal, heterozygous, or homozygous-abnormal at a mutation site. A typical result is shown in Fig. 15. A heterozygous template was used. The peaks of 37 and 46 bp represent the PCR products from both normal and mutant plates. Four primers with different lengths were used in this work.

3.3 Dynamic Mutation

There are numerous genetic diseases such as Huntington's disease (HD), dentatorubral-pallidoluysian atrophy (DRPLA), Kennedy's disease, and a number of spinal–cerebellar ataxias, which are due to the abnormal trinucleotide expansion. Rapid measurements of the number of these trinucleotide repeats can give accurate identification of normal, heterozygous carrier, and homozygous patients. Examples of MALDI mass spectrometry for rapid diagnosis of Huntington's disease and DRPLA are given in Figs. 16 and 17.

From the above, it is clear that MALDI mass spectrometry can be used in the diagnosis of various diseases.

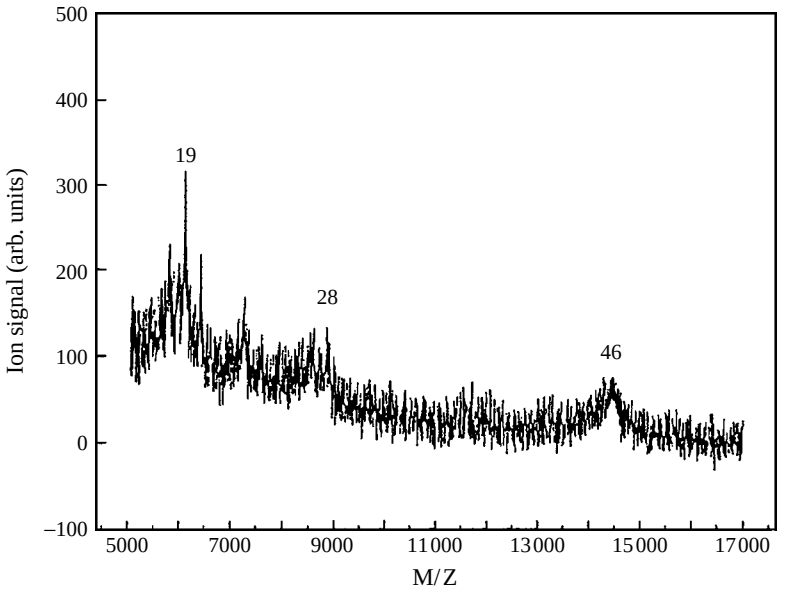

Fig. 15 MALDI mass spectrum for the detection of G551D point mutation in cystic fibrosis transmembrance conductance (CFTR). The G551D mutation in the CFTR gene involves a G to A mutation that results in a glycine to aspartic amino acid substitution. Mass spectrum of PCR products in 46 bp indicates the sample has a heterozygous template.

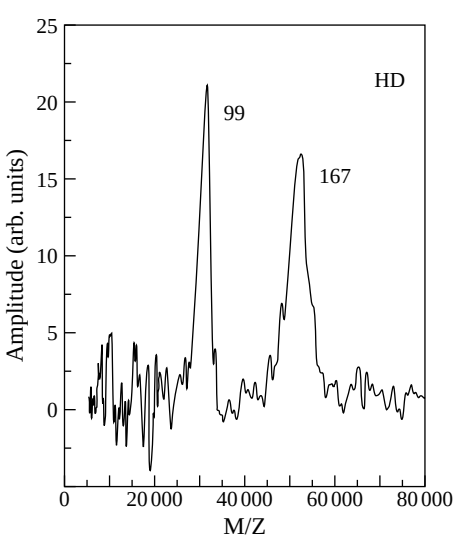

Fig. 16 Mass spectrum of a Huntington disease (HD) patient sample. Experimental results show the sample is heterozygous with CAG repeats of 17 and 40.

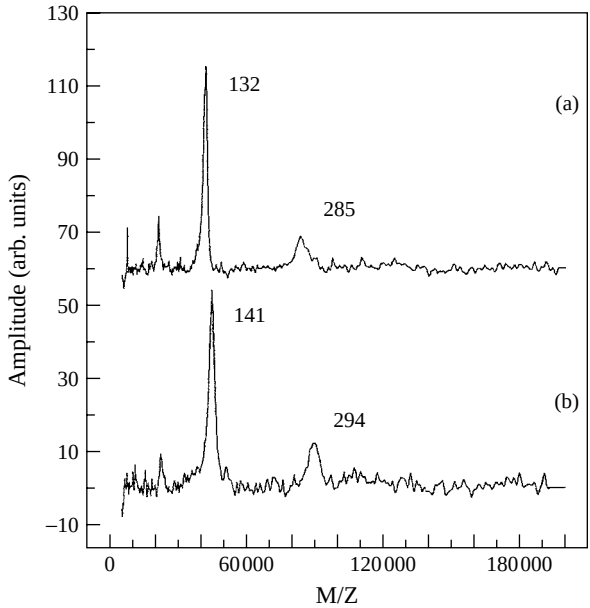

Fig. 17 Mass spectra for dentatorubral–pallidoluysian atrophy (DRPLA) patient samples. Both samples in (a) and (b) are heterozygous with the CAG repeats as 12/63 and 15/66 respectively.

4 DNA Typing for Forensic Applications

DNA fingerprinting is a technology that identifies individuals based on patterns of DNA markers detected in the genomic DNA. It can be used in forensic analysis to identify suspects or victims. It is particularly valuable at scenes of violent crimes where a body may not be available, or in instances where decomposition or dismemberment excludes the use of standard forensic techniques. There are several different approaches available to obtain DNA typing for forensic applications. These include restriction fragment length polymorphism (RFLP), short tandem repeats (STR), and single nucleotide polymorphism (SNP).

The foundation of DNA typing for person identification was established by the observation of restriction fragment length polymorphism (RFLP) in which polymorphic DNA loci are determined from restriction enzyme digestion. In 1985, Jefferys and his collaborators discovered that 33- and 16-bp repeated structures that vary in number between different individuals, referred to as *minisatellites*, could be used for fingerprinting. In 1987, the isolation and characterization of variable numbers of tandem repeats (VNTR) for single-locus DNA profiling was reported. More recently, the emphasis has shifted to the typing of STRs because of the high reliability arising from the large number of STRs in the human genome. During the past few years, it was discovered that SNPs) occur once every 300 to 1000 bp in DNA fragments. Thus, SNP typing is expected to become a very reliable tool for person identification in the future.

4.1 Forensic DNA Sample Preparation

In the process of using DNA typing for forensic applications, it is critically important to have very reliable and stringent protocols established. The major steps for DNA typing often include (1) Sample collection: blood, buccal cells, and hair follicles are the general sources for fresh DNA. For dried stains, samples should be collected and maintained in a dry, cool state. For autopsy tissue, it can be more reliable to have samples from more than one organ. All samples must be labeled as to source, tissue type, and the time and date of sample collection. (2) Storage and Transport: DNA samples can be stored and shipped as nonextracted tissue, as a lysate with tissue in an appropriate EDTA (ethylene diamine tetraacetic acid) solution, or as fully extracted pure DNA. For nonextracted tissues, sample refrigeration during shipping is often recommended. (3) DNA extraction: standard phenol–chloroform extraction and alcohol precipitation processes are often used to extract DNA from biological samples. An automated DNA extractor can achieve high quality extraction. It is desirable to complete DNA extraction as soon as possible. (4) DNA amplification: DNA amplification is often necessary to increase the quantity of sample available for analysis. PCR is the standard technique for the amplification of a selected DNA segment. Since an amplification of 10^6 or more can be achieved by PCR, any contamination by other DNAs can lead to a false conclusion. Thus, great care must be taken to prevent cross-contamination of sample material at the collection site and at any time during the analysis. (5) DNA size or sequence determination. This is the ultimate step leading to the conclusion of person identification. Until now, gel electrophoresis has been used for both DNA size determination and sequencing. However, analysis by mass spectrometry can be faster and more reliable.

The primary goal of DNA profiling for forensic application is to get the assurance of a match (inclusion) or mismatch (exclusion) of the evidence DNA sample with the suspect or victim's sample. In general, DNA profiling can be used to easily *rule out* the suspect when DNA typing between suspect and evidence shows clear differences. Thus, DNA typing is a very powerful tool for proving that an individual *does not* match the DNA collected at a crime scene. On the other hand, if the DNA profile from the evidence sample matches with the suspect, there is still a possibility that the evidence DNA could have come from another person. It might be a coincidental match between two persons having similar DNA profiles. Thus, population genetics plays an important role in suspect identification. The simplest approach is to assume that mating is random. The random mating of persons produces the same genotypes as random combinations of sperm and egg. Although the mating in local populations is not usually strictly random, the random match model provides a good approximation when no special genetic linkage is involved. For subpopulation groups such as people in an isolated village, the random mating approach is not generally a reliable assumption. However, it can probably be used as a guide for the minimum number of loci to be analyzed for confirming the matching of the suspect with the evidence.

Since the confirmation of matching the suspect with the evidence DNA often requires the analysis of several loci, the time

needed for DNA profiling can be long and the cost, high. At present, all forensic DNA samples for court evidence are analyzed by gel electrophoresis. Visualization of DNA in this procedure requires the use of either radioactive material or dye tagging. VNTR electrophoretic analysis, especially with radioactive probes, can take several weeks to complete the analysis of different loci for forensic confirmation. The human factor introduced by lengthy handling increases the risk of error or contamination and reduces reliability.

4.2 Mass Spectrometry DNA Detection for Forensic Applications

Each DNA typing involves (1) measurements of DNA fragments associated with genetic markers; (2) match determination for various genetic markers for different samples; and (3) statistical analysis for the type match to determine the possibility of an accidental match. All sequencing or size determination of DNA fragments for forensic applications to date have been carried out using gel electrophoresis. The various approaches among the forensic community include the analysis of VNTR, STR, SNP, and DNA sequencing. Laboratory protocol and standard procedures for data analysis were set up for forensic sample analysis. With the recent progress in development of mass spectrometry for DNA analysis, there is a good potential for the use of mass spectrometry for all the above applications. Since the use of mass spectrometry for DNA analysis is still in the early stages of development, standard protocols for its use in forensic analysis have not yet been established. The potential applications of using mass spectrometry in DNA fingerprinting for forensic applications is presented and discussed below.

4.2.1 RFLP and VNTR Analysis

DNA typing relies on the use of DNA markers. Among the most widely used DNA markers are restriction fragment length polymorphisms. RFLP analysis is based on differences in the positions of restriction endonuclease recognition sites. Restriction endonucleases are enzymes, which recognize and cut double-stranded DNA at sequence-specific recognition sites. Homologous DNA fragments from different individuals will contain different cut sites based on random mutations within the stretch of nucleotides. Consequently, fragments differing in length and number will be generated upon restriction digestion. These constitute RFLPs, and because they differ in size, the RFLP fragments can be readily separated by routine agarose electrophoresis.

When an individual's entire genomic DNA is digested by a restriction endonuclease, hundreds to thousands of restriction fragments are produced. When digested genomic DNA is subjected to gel electrophoresis, individual bands are not discernible. In such a case, analysis is generally facilitated by the Southern blotting technique. In Southern blotting, digested DNA is separated by electrophoresis. The digested fragments are transferred to a nitrocelluose or nylon membrane through capillary action. The DNA binds to the membrane by electrostatic attraction. Labeled probes specific for the RLFP of interest are then hybridized to the DNA-containing membrane. A banding pattern can be visualized using autoradiography or chemical detection techniques. Each individual will exhibit a characteristic pattern. Several RFLPs must be used

to statistically ensure that an individual's pattern is unique. The fact that RFLPs can occur anywhere in the genome and that there are potentially millions of RFLPs detectable with different probes makes RFLP analysis a tremendously powerful tool for forensic use.

A special class of RFLPs is based on DNA sequences that occur in tandem repeats. Tandem repeats arise when slippage mutation occurs during DNA replication. It has been observed that this type of mutation occurs frequently enough to generate significant variations over many generations. Polymorphism for these markers is so high in humans that only identical twins will have the same patterns. This factor makes the use of tandem repeats analysis extremely valuable to forensic scientists.

Variable number tandem repeat (VNTR) analysis has been used extensively in forensic applications. VNTRs consist of approximately 10 to 100 bp repeated several times in tandem. For example, a tandem repeat of the trinucleotide CTG would appear in the genome as CTGCTGCTGCTGCTGCTGCTG. In this particular example, this sequence has seven repeats. VNTRs are generally bordered by unique sequence DNA, a characteristic that can be exploited by current technology. In VNTR analysis, genomic DNA is digested with a restriction endonuclease that does not contain a cut site within the tandem repeat sequence. Since the restriction enzyme cuts outside of the tandem repeat region, the tandem repeat is left intact. Tandem repeat fragments are separated by gel electrophoresis and analyzed with Southern blotting. The length of the tandem repeat will vary between individuals. Individuals can be homozygous or heterozygous for a particular VNTR. The large number of VNTR-containing loci creates millions of possible pattern combinations. If there are n loci, there should be n homozygotes and $n(n-1)/2$ heterozygotes. If m loci are analyzed, the number of possible genotypes is $[n(n+1)/2]^m$. If $n = 20$ and $m = 4$, there are more than 1 billion genotypes. As a result, VNTR provides the forensic analyst a powerful tool for inclusion and/or exclusion.

The need for Southern blotting can be eliminated if the VNTR region is selectively amplified with the polymerase chain reaction (PCR). PCR primers complementary to sequences flanking the tandem repeat region can be designed and used to amplify the tandem repeat region. The length of an individual's tandem repeat can then be obtained by gel electrophoresis. It is PCR amplification of VNTRs that shows a lot of promise for MALDI detection. By using MALDI for RFLP detection, gel electrophoresis, Southern blotting and radioactive material or dye tagging can be eliminated. Figure 18 shows the mass spectra for restrictionenzyme–digested DNA samples. A difference of 12 bases with three 4-base tandem repeats can clearly be observed. Both primary and doubly charged ion peaks are observed. The spectrum indicates that one chromosome in this region is 12 bp longer than the same region in the other chromosomes. Thus, the result indicates the sample is from a person with heterozygous vWFII.

4.2.2 Short Tandem Repeats (STR)

For VNTR analysis by measuring RFLP without the use of PCR, one major disadvantage is that the quantity of the DNA sample is so small that a long analysis time is required. One promising alternative is to amplify and measure loci containing short tandem repeats (STR).

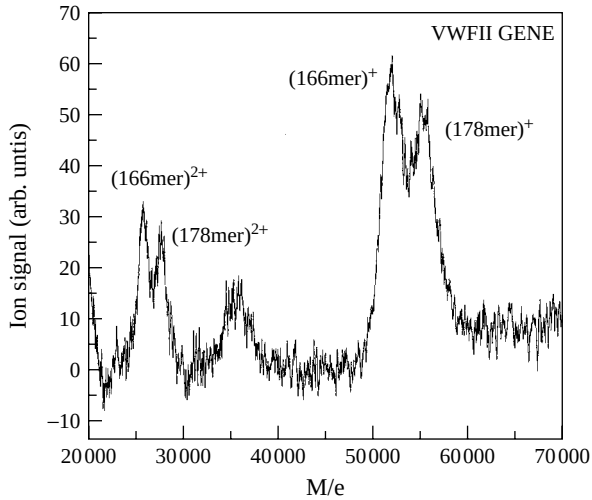

Fig. 18 Mass spectrum from the digestion of a part of *vWFII* gene. Peaks are observable for a difference of 12 bases. Both primary and doubly charged ion peaks are observed. The spectrum indicates that one chromosome in this region is 12 bp longer than the same region in the other chromosome.

There is a large number of STR distributed throughout the human genome. STR analysis can be used for as many loci as are needed for reliable identification. PCR is nearly always used in STR analysis. Since PCR can amplify the selected DNA segments by more than six orders of magnitude, only a very small quantity of DNA sample is required. However, the disadvantage is that any procedure using PCR amplification is highly susceptible to contamination. Even a trace of foreign DNA contamination can be amplified and detected to potentially lead to a false conclusion. However, careful handling of DNA samples or applying quantitative PCR techniques can more or less resolve the concern regarding DNA contamination. Since the analysis of STR is by measurement of the sizes of DNA products from the PCR process, the sizes of the DNA fragments are often less than 300 bp and the quantity is often in the range of a few picomoles, which can be readily detected by MALDI.

Typical mass spectra of a DNA sample amplified by PCR from parts of different genes are shown in Fig. 19. Since different people have different numbers of repeats in each of these genes, measurements of the number of repeats for several loci can be used for person identification. There are many tetranucleotide repeats, which can be used for forensic identification. HUMTH01 is one that is often used in forensic applications.

4.2.3 Single Nucleotide Polymorphism (SNP)

There are several types of DNA sequence variation, including deletion, insertion, difference in number of repeats such as STR and single base-pair differences. The single base difference is the most common type. This single base pair difference is often referred to as *single nucleotide*

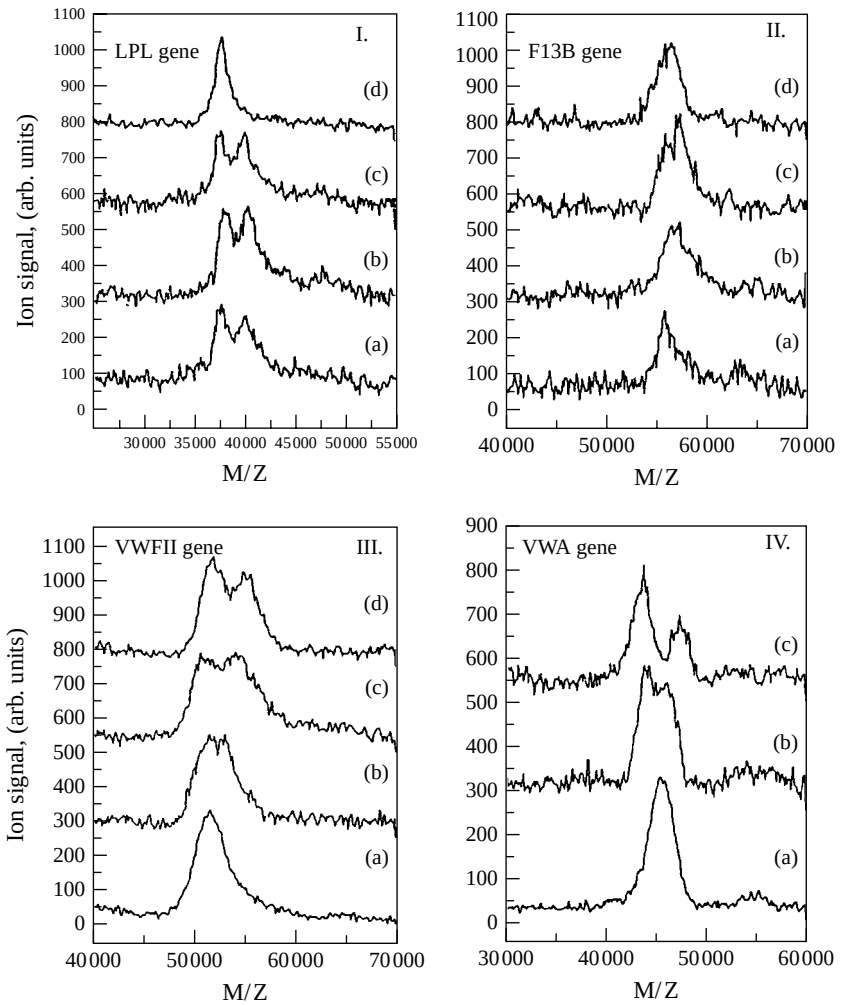

Fig. 19 MALDI mass spectra of PCR products of *LPL*, *F13B*, *vWFII*, and *vWA* genes from different persons. Homozygous and heterozygous as well as the number of the short tandem repeats can be determined. This indicates MALDI can be used for short tandem repeat determination for forensic applications.

polymorphism (SNP). There are several million SNPs in the human genome. The predominance of SNPs and their much lower mutation rate make SNP analysis extremely valuable for forensic applications. ASPCR, described above for point mutation detection, can also be applied to SNP detection.

Another approach to using MALDI for SNP detection is to hybridize a primer adjacent to the polymorphic locus and extend it by a single base by dideoxynucleotide addition. MS can be used to accurately measure the mass of the extension product to identify the base. For homozygotes, one peak will be observed.

However, two peaks should be observed for a heterozygote.

4.2.4 Gender Determination

Sex determination is often important for identifying suspects and victims. A reliable gender test is PCR amplification of a segment of the X–Y homologous gene amelogenin. A single pair of primers spanning part of the first intron generates 106-bp and 112-bp PCR products from the X and Y chromosomes respectively. MALDI can be subsequently used for the measurements of PCR products. Recently, it was found that the deletion of the Y-encoded gene occurs in a small percentage of Y chromosomes. Thus, it is more reliable to test both amelogenin and the male sex-determining gene SRY for gender determination. Male DNA generates three products of lengths 93 bp (SRY), 106 bp (Amelogenin in X chromosome) and 112 bp (amelogenin in Y chromosome). Coamplification of female DNA generates only the 106-bp product. As shown in Fig. 20, the distinction of genotyping between male and female is quite clear. Figure 20 shows examples of mass spectra of PCR products for (a) female gender, (b) male gender, and (c) mixture of male and female gender. Of special interest in forensic analysis, especially for sexual assault, is the detection of a trace amount of male-specific DNA in a female sample. Such detection can be performed by demonstrating the presence of male-derived products of 93 bp and 112 bp. When MALDI spectra of a composite sample containing mostly female species (10 ng) and only a trace amount of male sample (0.01 ng) was checked, the trace of the male sample could be detected. It

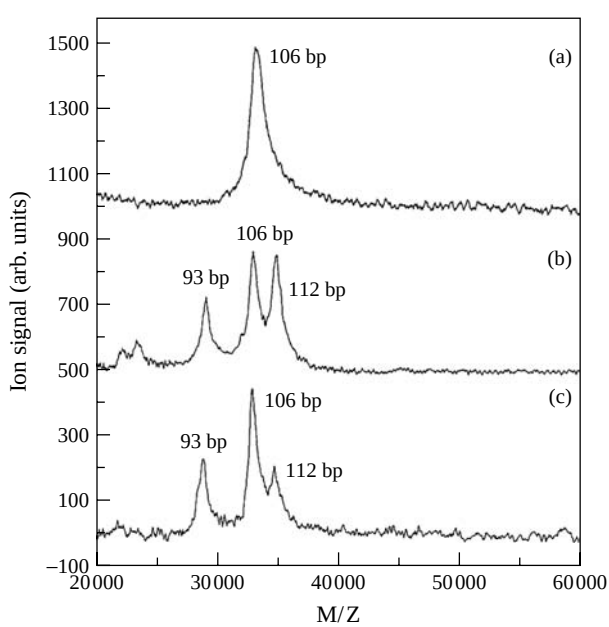

Fig. 20 Mass spectra of samples for sex-specific coamplification (Amel X/Y plus SRY) of loci for (a) female (b) male, and (c) female plus male (with sample ratio of 1 : 1) respectively.

indicates that as little as 0.1 percent of male DNA in a female DNA sample can be detected. It also indicates that the detection sensitivity can reach 0.01 ng or less which is about one order of magnitude better than by gel electrophoresis.

5 Conclusion

Mass spectrometry provides a unique method of nucleic acid detection and sequencing. Applications relating to disease diagnosis, and forensic applications in person identification have been clearly demonstrated. However, improvement in mass resolution and detection sensitivity of large DNA fragments is still required in order to enable MS to be routinely used for broad applications. A better understanding of the mechanisms of desorption and ionization will provide the basis for further improvement of MALDI.

Acknowledgment

This work is primarily supported by the National Institute of Justice, grant number 97-LB-VX-A047, and supported in part by the Office of Biological and Environmental Research, U.S. Department of Energy under contract DE-AC05-96OR22464 with UT Bartell LLC.

See also Genomic Sequencing (Core Article); Human Genetic Variation and Disease; Mass Spectrometry of Proteins (Proteomics); Nucleic Acid and Protein Sequence Analysis and Bioinformatics.

Bibliography

Books and Reviews

Chen, C.H. (1996) in: Meyers, R.A. (Ed.) *Mass Spectrometry High Speed DNA Fragment Sizing Encyclopedia of Molecular Biology and Molecular Medicine*, VCH Publication, VCH, Weinheim pp. 1–11.

Lay, J.O. Jr. (2001) MALDI-TOF mass spectrometry of bacteria, *Mass Spectrom. Rev.* **20**, 172–194.

Lechner, D., Lathrop, G.M., Gut, I.G. (2001) Large-scale genotyping by mass spectrometry: experience, advances and obstacles, *Proteomic Genomic* **6**, 31–38.

Limbach, P.A. (1996) Indirect mass spectrometric methods for characterizing and sequencing oligonucleotides, *Mass Spectrom. Rev.* **15**, 297–336.

Murray, K.K. (1996) DNA sequencing by mass spectrometry, *J. Mass Spectrom.* **31**, 1203–1215

Primary Literature

Amexis, G., Oeth, P., Abel, K., Ivshina, A., Pelloquin, F., Cantor, C.R., Braun, A., Chumakov, K. (2001) Quantitative mutant analysis of viral quasispecies by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12097–12102.

Berkenkamp, S., Kirpeka, F., Hillemkamp, F. (1998) Infrared MALDI mass spectrometry of large nucleic acid, *Science* **281**, 266–268.

Buetow, K.H., Edmonson, M., Macdonald, R., Clifford, R., Yip, P., Kelley, J., Little, D.P., Strausberg, R., Koster, H., Cantor, C.R., Braun, A. (2001) High throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry based strategies and studies among those approaches, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 581–584.

Chang, L.Y., Tang, K., Schell, M., Ringelberg, C., Matteson, K.J., Allman, S.L., Chen, C.H. (1995) Detection of ΔF508 mutation of the cystic fibrosis gene by laser mass spectrometry, *Rapid Commun. Mass Spectrom.* **9**, 772–774.

Chen, C.H., Chang, L.Y., Tang, K. (1994) Laser mass spectrometry for molecular medicine applications, *Opto News Lett.* **47**, 8–11.

Chen, C.H., Golovlev, V.V., Taranenko, N.I., Allman, S.L., Isola, N.R., Potter, N.T., Matteson, K.J., Chang, L.Y. (1999) MALDI for DNA sequencing, disease diagnosis, and forensic applications, *Am. Biotech. Lab.* **17**, 14–18.

Chen, C.H., Taranenko, N.I., Zhu, Y.F., Allman, S.L. (1995) MALDI for fast DNA analysis and sequencing, *Lab. Robotics Automat.* **8**, 87–99. Chiu, N.H.L., Tang, K., Yip, P., Braun, A., Koster, H., Cantor, C.R. (2000) Mass spectrometry of single-stranded restriction fragments captured by an undigested complementary sequence, *Nucleic Acids Res.* **28**, e31.

Ding, C., Cantor, C.R. (2003) A high throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight MS, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3059–3064.

Doktycz, M.J., Hurst, G.B., Habibi-Goudarzi, S., McLuckey, S.A., Tang, K., Chen, C.H., Uziel, M., Jacobson, K.B., Woychik, R.P., Buchanan, M.V. (1995) Analysis of PCR-amplified DNA products by mass spectrometry using matrix-assisted laser desorption and electrospray: current status, *J. Anal. Biochem.* **230**, 205–214.

Elso, C., Toohey, B., Reid, G.E., Potter, K., Simpson, R.J., Foote, S.J. (2002) Mutation detection using mass spectrometric separation of tiny oligonucleotide fragments, *Genomic Res.* **12**, 1428–1433.

Faulstich, K., Woemer, K., Brill, H., Engels, J.W. (1997) A sequencing method for RNA oligonucleotides based on mass spectrometry, *Anal. Chem.* **69**, 4349–4353.

Golovlev, V.V., Allman, S.L., Garrett, W.R., Taranenko, N.I., Chen, C.H. (1997) Laser induced acoustic desorption, *Int. J. Mass Spectrom. Ion Process.* **169/170**, 69–78.

Golovlev, V.V., Lee, S.-S.H., Allman, S.L., Taranenko, N.I., Chen, C.H. (2001) Non-resonance MALDI of oligonucleotides: mechanism of ion desorption, *Anal. Chem.* **73**, 809–812.

Haff, L.A., Smirnov, I.P. (1989) Single-nuvcleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI/TOF mass spectrometry, *Genome Res.* **7**, 378–388.

Hartmer, R., Storm, N., Boecker, S., Rodi, C.P., Hillenkamp, F., Jurinke, C., Boom, D.V. (2003) Rnase T1 mediated base-specific cleavage and MALDI-TOF MS for high-throughput comparative sequence analysis, *Nucleic Acids Res.* **31**, e47.

Isbell, D.T., Gusev, A., Taranenko, N.I., Chen, C.H., Hercules, D.M. (1999) Analysis of nucleotides directly from TLC plates using MALDI-MS detection, *Fresenius' J. Anal. Chem.* **365**, 625–630

Isbell, D.T., Gusev, A., Taranenko, N.I., Chen, C.H., Hercules, D.M. (1999) Separation and detection of carcinogen-adducted oligonucleotides *J. Mass Spectrom.* **34**, 774–776.

Isola, N.R., Allman, S.L., Golovlev, V.V., Chen, C.H. (1999) Chemical cleavage sequencing of DNA using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Anal. Chem.* **71**, 2266–2269.

Isola, N.R., Taranenko, N.I., Allman, S.L., Golovlev, V.V., Chen, C.H. (2001) MALDI-TOF mass spectrometric method for detection of hybridized DNA oligomer, *Anal. Chem.* **73**, 2126–2131.

Jones, R.B., Allman, S.L., Tang, K., Garrett, W.R., Chen, C.H. (1993) Neutralization of negatively charged oligonucleotides, *Chem. Phys. Lett.* **212**, 451–456.

Jurinke, C., van den Boom, D., Cantor, C.R., Koster, H. (2001) The use of Mass ARRAY technology for high throughput genotyping, *Adv. Biochem. Eng. Biotechnol.* **77**, 57–74.

Kirpekar, F., Nordhoff, E., Kristiansen, K., Roepstorff, P., Lezius, A., Hahner, S., Karas, M., Hillenkamp, F. (1993) Matrix-assisted laser desorption/ionization mass spectrometry of enzymatically synthesized RNA up to 150 kDa, *Nucleic Acids Res.* **22**, 3866–3870.

Koster, H., Tang, K., Fu, D., Braun, A., van den Boom, D., Smith, C.L., Cotter, R.J., Cantor, C.R. (1996) A strategy for rapid and efficient DNA sequencing by mass spectrometry, *Nature/Biotechnology* **14**, 1123–1128.

Kwon, Y., Tang, K., Cantor, C.R., Koster, H., Kang, C. (2001) DNA sequencing and genotyping by transcriptional synthesis of chain-terminated RNA, *Nucleic Acids Res.* **29**, e11.

Lefmann, M., Honisch, C., Bocker, S., Storm, N., Wintzingerode, F., Schloteburg, C., Moter, A., Boom, D.V., Gobel, U.B. (2004) Novel mass spectrometry-based tool for genotypic identification of mycobacteria, *J. Clin. Microbiol.* **42**, 339–346.

Little, D.P., Braun, A., O'Donnell, M.J., Koster, H. (1997) Mass spectrometry from

miniaturized arrays for full comparative DNA analysis, *Nat. Med.* **12**, 1413–1416.

Little, D.P., Chorush, R.A., Spier, J.P., Senko, M.W., Kelleher, N.L., McLafferty, F.W. (1994) Rapid sequencing of oligonucleotides by high-resolution mass spectrometry, *J. Am. Chem. Soc.* **116**, 4893–4897.

Maya, L., Chen, C.H., Stevenson, K.A., Kenik, E.A., Allman, S.L., Thundat, T.G. (2002) Mass spectrometric analysis of water-soluble gold nanoclusters, *J. Nanoparticle Res.* **4**, 417–422.

Muddiman, D.C., Anderson, G.A., Hofstadler, S.A., Smith, R.D. (1997) Length and base composition of PCR-amplified nucleic acids using mass measurements from electronspray ionization mass spectrometry, *Anal. Chem.* **69**, 1543–1549.

Nordhoff, E., Karas, M., Cramer, R., Hahner, S., Hillenkamp, F., Kirpekar, F., Lezins, A., Muh, J., Meier, C., Engela, J.W. (1995) Direct mass spectrometric sequencing of low picomole amount of oligonucleotides up to 21 bases by matrix-assisted laser desorption/ionoization mass spectrometry, *J. Mass Spectrom.* **30**, 99–112.

Null, A.P., Hannis, J.C., Muddiman, D.C. (2001) Genotyping of simple and compound short tandem repeat loci using electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, *Anal. Chem.* **73**, 4514–4521.

Tang, K., Allman, S.L., Chen, C.H. (1993) Matrix-assisted laser desorption ionization (maldi) of oligonucleotides with various matrices, *Rapid Commun. Mass Spectrom.* **7**, 943–948.

Tang, K., Allman, S.L., Chen, C.H., Chang, L.Y., Schell, M. (1994) Matrix-assisted laser desorption/ionization of restriction enzyme-digested DNA, *Rapid Commun. Mass Spectrom.* **8**, 183–186.

Tang, K., Allman, S.L., Jones, R.B., Chen, C.H. (1992) Comparison of rhodamine dyes as matrices for matrix-assisted laser desorption/ionization mass spectrometry, *Org. Mass Spectrom. Lett.* **27**, 1389–1392.

Tang, K., Allman, S.L., Jones, R.B., Chen, C.H. (1993) Laser mass spectrometry of polydeoxyribothymidylic acid mixtures, *Rapid Commun. Mass Spectrom.* **7**, 63–66.

Tang, K., Allman, S.L., Jones, R.B., Chen, C.H. (1993) Quantitative analysis of biopolymers by matrix-assisted laser desorption, *Anal. Chem.* **64**, 2164–2166.

Tang, K., Allman, S.L., Jones, R.B., Chen, C.H., Araghi, S. (1993) Laser mass spectrometry of oligonucleotides with isomer matrices, *Rapid Commun. Mass Spectrom.* **7**, 435–439.

Tang, K., Oeth, P., Kammerer, S., Denissenko, M.F., Ekblom, J., Jurinke, C., Boom, D.V., Cantor, C.R. Mining disease susceptibility genes through SNP analyses and expression profiling using MALDI-TOF mass spectrometry, *J. Proteomic Res.* (in Press).

Tang, K., Opalsky, D., Abel, K., van den Boom, D., Yip, P., DelMistro, G., Braun, A., Cantor, C.R. (2003) Single neucleotide polymorphism analyses by MALDI-TOF MS, *Int. J. Mass Spectrom.* **226**, 37–54.

Tang, K., Taranenko, N.I., Allman, S.L., Chang, L.Y., Chen, C.H. (1994) Detection of 500 – nucleotide DNA by laser desorption mass spectrometry, *Rapid Commun. Mass Spectrom.* **8**, 727–730.

Tang, K., Taranenko, N.I., Allman, S.L., Chen, C.H., Chang, L.Y., Jacobson, K.B. (1994) Picolinic acid as a matrix for laser mass spectrometry of nucleic acid and proteins, *Rapid Commun. Mass Spectrom.* **8**, 673–677.

Tang, W., Zhu, L., Smith, L.M. (1997) Controlling DNA Fragmentation in MALDI-MS by chemical modification, *Anal. Chem.* **69**, 302–312.

Taranenko, N.I., Allman, S.L., Golovlev, V.V., Taranenko, N.V., Isola, N.R., Chen, C.H. (1998) Sequencing DNA using mass spectrometry for ladder detection, *Nucleic Acid Res.* **26**, 2488–2490.

Taranenko, N.I., Chung, C.N., Zhu, Y.F., Allman, S.L., Golovlev, V.V., Isola, N.R., Martin, S.A., Haff, L.A., Chen, C.H. (1997) Matrix-assisted laser desorption/ionization for sequencing single-stranded and double-stranded DNA, *Rapid Commun. Mass Spectrom.* **11**, 386–392.

Taranenko, N.I., Golovlev, V.V., Allman, S.L., Taranenko, N.V., Chen, C.H., Hong, J., Chang, L.Y. (1998) Matrix-assisted laser desorption/ionization for short tandem repeat loci, *Rapid Commun. Mass Spectrom.* **12**, 413–418.

Taranenko, N.I., Hurt, R., Zhou, J.Z., Isola, N.R., Huang, H., Lee, S.H., Chen, C.H. (2002) Laser desorption mass spectrometry for microbial DNA analysis, *J. Microbiol. Methods* **48**, 101–106.

Taranenko, N.I., Matteson, K.J., Chung, C.N., Zhu, Y.F., Chang, L.Y., Allman, S.L., Haff, L.,

Martin, S.A., Chen, C.H. (1996) Laser desorption mass spectrometry for point mutation detection, *Genet. Anal.: Biomol. Eng.* **13**, 87–94.

Taranenko, N.I., Potter, N.T., Allman, S.L., Chen, C.H. (1999) DNA Analysis for Gender Identification by MALDI, *Anal. Chem.* **71**, 3974–3976.

Taranenko, N.I., Potter, N.T., Allman, S.L., Golovlev, V.V., Chen, C.H. (1999) Detection of trinucleotide expansion in neurodegenerative disease by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Genet. Anal.: Biomol. Eng.* **15**, 25–31.

Taranenko, N.I., Tang, K., Allman, S.L., Chang, L.Y., Chen, C.H. (1994) 3-Aminopicolinic acid as a matrix for MALDI of DNA, *Rapid Commun. Mass Spectrom.* **8**, 727–730.

Vallone, P.M., Devaney, J.M., Marino, M.A., Butler, J.M. (2002) A strategy for examining complex mixture of deoxyoligonucleotide using IP-RP HPLC, MALDI-TPF MS and Iinformatics, *Anal. Biochem.* **304**, 257–265.

Wada, Y., Yamamoto, M. (1997) Detection of single-nucleotide mutation including substitution and deletion by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Rapid Commun. Mass Spectrom.* **11**, 1657–1660.

Wintzingerode, F., Bocker, S., Schlotelburg, C., Chiu, N.H.L., Storm, N., Jurinke, C., Cantor, C.R., Gobel, U.B., Boom, D.V. (2002) Base specific fragmentation of amplified 16S rRNA genes analyzed by mass spectrometry: a tool for rapid bacterial identification, *Proc. Natl. Acad. Sci.* **99**, 7039–7044.

Zhu, Y.F., Lee, K.L., Tang, K., Allman, S.L., Taranenko, N.I., Chen, C.H. (1995) Revisit of MALDI for small proteins, *Rapid Commun. Mass Spectrom.* **9**, 1315–1320.

Zhu, Y.F., Taranenko, N.I., Allman, S.L., Martin, S.A., Haff, L., Chen, C.H. (1996) The study of 2,3,4-trihydroxyacetophenone and 2,4,5-trihydroxyacetophenone as matrices for DNA detection in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Rapid Commun. Mass Spectrom.* **10**, 383–388.

Zhu, Y.F., Taranenko, N.I., Allman, S.L., Martin, S.A., Haff, L., Chen, C.H. (1996) The effect of ammonium salt and matrix in the detection of DNA by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Rapid Commun. Mass Spectrom.* **10**, 1591–1596.

Mass Spectrometry-based Methods of Proteome Analysis

Boris L. Zybailov and Michael P. Washburn
Stowers Institute for Medical Research, Kansas City, MO, USA

1 Principles and Instrumentation 3
1.1 Proteome and Proteomics 3
1.2 Need for Large-scale Analyses of Gene Products at the Protein Level 3
1.3 General Problems in Proteome Analyses 4
1.4 MS Instrumentation 5
1.4.1 Ionization Methods 6
1.4.2 Mass Analyzers 8
1.4.3 MS/MS 9
1.5 Methods of Sample Fractionation 10
1.5.1 2D-electrophoretic Separation of Complex Protein Mixtures 11
1.5.2 Liquid-phase Separation Methods 12
1.5.3 Affinity Methods (Epitope Tagging) 13
1.6 Protein Identification by MS 14

2 Quantitative Methods of Proteome Analysis Using MS 18
2.1 Metabolic Labeling 18
2.2 Isotope Coded Affinity Tags 20
2.3 ^{18}O Labeling 21
2.4 Postdigestion Labeling 22
2.5 Global mRNA and Protein Expression Analyses 23

3 Specific Examples of Applications 23
3.1 Global Proteome Sampling 23
3.1.1 Global Proteome Sampling Based on 2D Page 23
3.1.2 Global Proteome Sampling Based on Multidimensional LC 24
3.1.3 MS-assisted Disease Diagnosis from Serum Samples 25
3.2 Analysis of Protein Modifications by Mass Spectrometry 25
3.2.1 Phosphorylated Proteins 27
3.2.2 Glycosylated Proteins 28
3.2.3 Ubiquitinated Proteins 29

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

3.3 Analysis of Protein–Protein Interactions by Mass Spectrometry 30
3.3.1 Computational Methods of Protein–Protein Interaction Prediction 30
3.3.2 Yeast 2-hybrid Arrays 30
3.3.3 Direct Analysis of Large Protein Complexes; Composition of the Yeast Ribosome 31
3.3.4 Analysis of Multisubunit Protein Complexes Involved in Ubiquitin-dependent Proteolysis by Mass Spectrometry 31
3.3.5 Proteomics of the Nuclear Pore Complex 31
3.3.6 High-throughput Analyses of Protein–Protein Interactions 33
3.3.7 Quantitative Proteomics Methods in the Studies of Protein Complexes 35

Bibliography 36
Books and Reviews 36
Primary Literature 36

Keywords

2D Page
Two-dimensional polyacrylamide gel electrophoresis; a technology that separates intact proteins according to their pI in the first dimension and molecular weight in the second.

MS/MS
A mass spectrometry technique in which precursor ions are selectively fragmented, thus allowing detailed structural information to be obtained.

MudPIT
A multidimensional protein identification technology; a proteomic method based on liquid chromatography of peptide mixtures with subsequent identification by tandem mass spectrometry.

Quantitative Proteomics
A group of methods that allow large-scale quantitative assessment of protein expression and identification.

Proteomics aims to identify, characterize, and map gene functions at the protein level for whole cells or organisms. A typical experimental scheme for a large-scale proteomics inquiry involves fractionation of a complex protein mixture by electrophoretic or chromatographic means followed by subsequent identification of the components in the individual fractions by mass spectrometry. Owing to

continuous and rapid improvement in instrument sensitivity, throughput capacity, software versatility, and techniques of statistical validation of the data, mass spectrometry–based approaches are becoming mainstream methods in a proteome analysis.

1 Principles and Instrumentation

1.1 Proteome and Proteomics

The term *proteome* was first introduced in mid-1990s to name the functional complement of a genome. By analogy with *genomics*, the term *proteomics* refers to studies of a gene's function at the protein level. Both these terms have a large-scale flavor to them. Indeed, studies that fall under the category "proteomics" frequently deal with large-scale analyses of proteins on the level of the whole organism, tissue, cell, or subcellular compartments.

Examples of typical biological questions addressed by modern proteomics experiments include but are not limited to establishing "news of difference" between healthy and pathological states, monitoring global gene expression during growth and development, establishing cellular localizations of a particular subset of an expressed genome, and identifying networks of protein–protein interactions. In fact, a recent review classifies proteomics studies according to the type of the addressed biological questions into "profiling proteomics," "functional proteomics," and "structural proteomics." Profiling proteomics amounts to large-scale identification of the proteins in a cell or tissue present at a certain physiological conditions. Functional proteomics refers to the studies of functional characteristics of proteins – posttranslational modifications, protein–protein interactions, and cellular localizations. Structural proteomics includes studies of protein tertiary structure, typically made by a combination of X-ray, NMR, and computational techniques. Adopting this classification, this chapter primarily focuses on the profiling and functional proteomics.

1.2 Need for Large-scale Analyses of Gene Products at the Protein Level

Rapid success of the various genome-sequencing programs is one of the major factors that led to the development of large-scale proteomics methods. In fact, a search of the genome-derived sequence databases is usually an intrinsic part of mass spectrometry–based proteome analysis. While of tremendous value, genomic information is, in principle, insufficient for understanding the complex processes of cellular function. Indeed, in addition to the projection of the information from genes into proteins, cells of the living organisms must metabolically extract useful information from the environment. Therefore, the total informational content necessarily increases in a proteome compared to the corresponding genome.

An immediate consequence of this increase in the informational content is that all of the following – protein isoforms, protein posttranslational modifications, protein conformational states, protein abundance levels, as well as dynamical

changes of these properties – have their own important functional implications in the living cells. In certain cases, genomic sequence still can be used to assess some of the functional properties of the corresponding proteins. For example, the codon adaptation index (CAI) and codon bias can be used to predict the expression level of genes. Another important case is a global sequence-based prediction of protein–protein interactions, an example of which is discussed in Sect. 3 of this chapter.

Also, to some degree, the functional properties of proteins can be assessed by analysis of the mRNA transcripts. Specifically, mRNA expression profiles are frequently used to estimate the corresponding protein levels. Most common methods that are used to measure global mRNA expression are cDNA microarray and serial analysis of gene expression (SAGE). However, in many cases, the correlation between mRNA and protein is insufficient to quantitatively predict protein expression. Potential reasons for these discrepancies in mRNA and protein levels include the differences in half-lives of proteins and mRNAs and posttranslational mechanisms that control the rates of translation. Thus, neither DNA nor mRNA sequence information is sufficient for understanding of the cellular functions. This fact provides motivation to improve the old and develop new technologies for the large-scale analyses of gene products at the protein level.

1.3 General Problems in Proteome Analyses

In a given organism, there are several times more different proteins than there are genes. Estimates for humans give such numbers as about 30 000 different genes and more than 100 000 different proteins. Factors such as posttranslational modifications, isoforms, and expression levels increase the complexity and interconnectedness of a proteome compared to a genome. Problems associated with modern proteomics methods include (1) difficulties in the detection of low-abundant proteins; (2) difficulties in obtaining quantitative information; (3) biases in a proteome coverage; (4) difficulties in characterization of posttranslational modifications; (5) difficulties in data analysis and interpretation in the large-scale experiments; (6) poor reproducibility.

1. Protein levels as low as several dozens copies per cell can be of functional significance. This is true for certain receptors, signaling proteins, and regulatory proteins. However, detection of the low-abundance proteins is often an issue in proteomics experiments. The ability of a proteomics method to identify low-abundant proteins in the presence of high-abundant ones is characterized by the dynamic range of a method. The dynamic range of a proteomics method is defined as the ratio of concentration of the most abundant to the concentration of the least abundant proteins identified by a method. Proteomics methods based on two-dimensional polyacrylamide gel-electrophoresis (2D PAGE) separations typically have a lower dynamic range than the methods based on affinity separation or multidimensional liquid chromatography.
2. Quantitative information on the protein abundance in different physiological contexts is the goal of comparative proteomics studies. 2D PAGE methods are good at solving these types of problems, when combined with scintillation

counting or fluorescence imaging spectroscopy. In the MS-based methods, quantitation is achieved by differential labeling (discussed in detail in Sect. 2).

3. Biases in proteome coverage can be inherent to a particular method, or be related to the design of an experimental scheme. For example, 2D PAGE methods are biased against highly hydrophobic proteins. At the same time, there is no inherent bias against hydrophobic proteins in MS identification. However, sample enrichment or fractionation steps necessarily preceding MS analysis are often biased against one or the other class of proteins.
4. Information on types and degrees of posttranslational modifications and protein isoforms is a necessary component of comprehensive functional description of a proteome. Unfortunately, it is difficult to design a proteomics method that would simultaneously give both good proteome coverage and identify all posttranslational modifications. Methods, such as 2D PAGE, which separate intact proteins, are good at the detection of different posttranslational modifications and isoforms present at the same time. However, owing to large diversity in protein masses and shapes, biases in proteome coverage are inherent in the intact protein methods.
5. During a large-scale proteome analysis, thousands of data points are generated. Hence, there are inherent problems related to the data reduction and extraction of the useful information. Additionally, the obtained information needs to be presented in an accessible format to the scientific community. These issues often require significant computational support and software development.
6. Most of the high-throughput, large-scale methods suffer from poor reproducibility. This is especially true in the case of 2D PAGE and MS-based schemes with several fractionation and separation steps involved. Possible ways to improve reproducibility include reduction in the number of separation and fractionation steps, adhering to standard protocols, and use of automation whenever possible.

Most of these challenges, however, are merely technical limitations. Modern technologies continuously progress toward higher sensitivity, higher range of proteome coverage, and higher reproducibility. Also, even if it may be impossible to have 100% proteome coverage with a single method, proper design of a proteomics program that takes advantage of several different methods can achieve impressive results.

1.4 MS Instrumentation

Mass spectrometry (MS) is a platform technology for all proteomics. It is the mass spectrometer that is used to identify the protein present in samples. Furthermore, there are a variety of mass spectrometry approaches that can be used to achieve this end. Figure 1 illustrates the general concept of MS analysis in proteomics. Generally, MS analysis involves creation of gaseous ions from an analyte followed by separation of the produced ions according to their mass-to-charge ratio (m/z). Instruments that perform analysis of this type are called *mass spectrometers*. Principle components of a mass spectrometer are an ion source, where ionization takes place; a mass analyzer, where the m/z is measured; and a detector, where amount

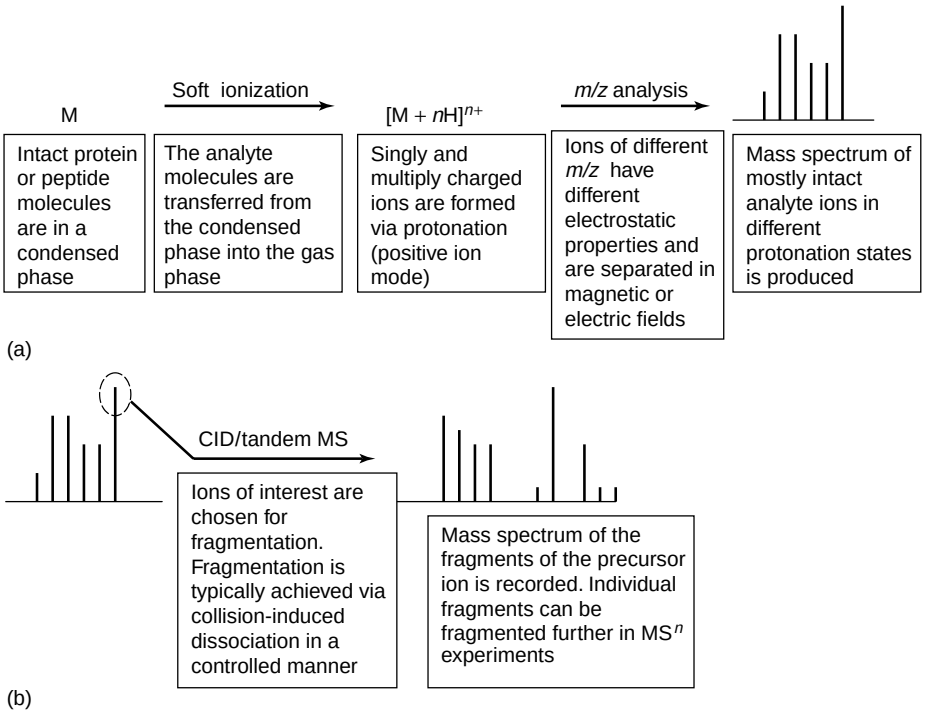

Fig. 1 Principle scheme of single MS (a) and MS/MS (b) analyses. (a) Intact proteins or peptides are transferred from the condensed phase into the gas phase and are ionized through capture (positive mode) or loss (negative mode) of protons. In high-throughput protein identification experiments, positive mode of ionization is used. Negative mode of ionization can be used for analysis of carbohydrates and nucleic acids. Ions of different m/z are discriminated by magnetic or electric fields. (b) Ions whose structure needs to be determined are chosen for fragmentation via CID reactions in the MS/MS experiment. The obtained fragmentation patterns can be searched against protein databases for identification.

of the ions corresponding to a particular m/z is recorded. Large macromolecules, such as proteins, are nonvolatile, and owing to a lack of proper ionization techniques for a long time, MS was limited to the analysis of smaller molecules. Finally, in the 1980s, protein ionization methods were developed, thus making MS-proteomics feasible.

1.4.1 Ionization Methods

The two most commonly used methods of protein ionization are electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI). In both of these techniques, ionization occurs through uptake (positive mode) or loss (negative mode) of one or several protons. In large-scale proteomics studies, the positive mode is typically used. The negative mode of ionization finds its uses in analyses of carbohydrates and nucleic acids. In any of the available ionization methods, there is no definite relationship between the amount of ions formed and the amount of the analyte. This fact makes mass spectrometers inherently nonquantitative devices.

ESI produces mostly multiply charged ions by creating a fine spray of charged droplets in a strong electric field. With the application of a dry gas, the droplets evaporate, and the electrostatic repulsion causes transfer of the analyte ions into the gas phase (Fig. 2a). The multiple charging allows analysis of very large molecules with analyzers that have relatively small m/z range. Another advantage of the multiple charging is that more accurate molecular weight can be obtained from analysis of the distribution of multiply charged peaks. ESI is used in a wide range of proteomics applications but is limited by susceptibility to high salt concentrations and to contaminants in the sample.

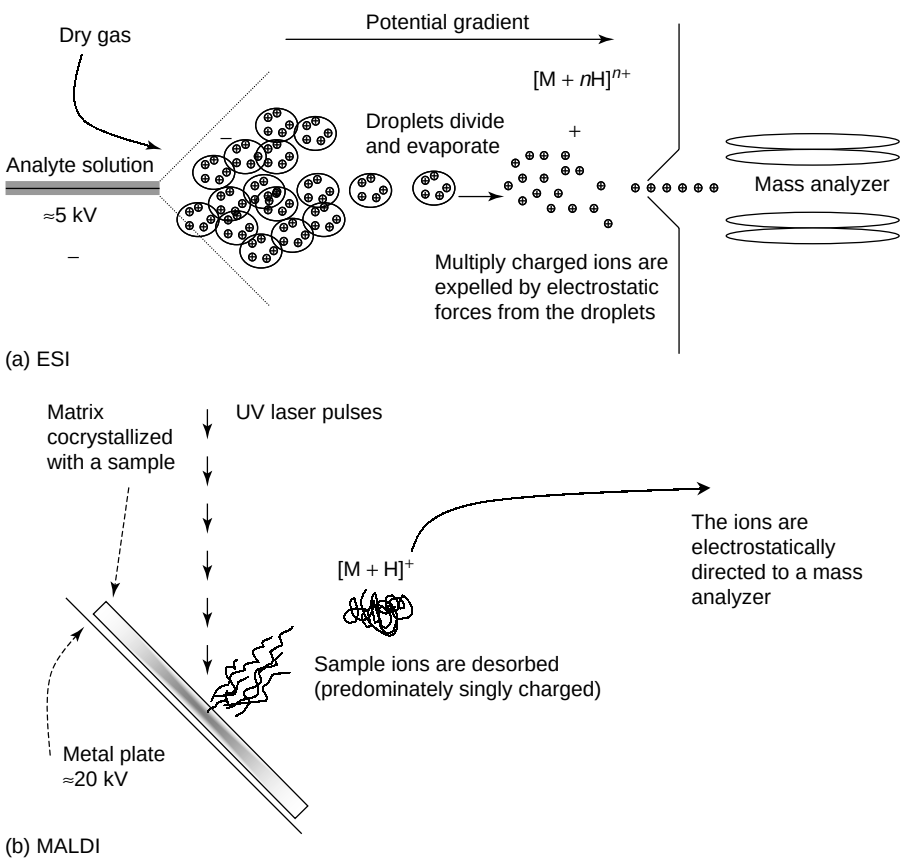

Fig. 2 Conceptual schemes of (a) Electrospray ionization (ESI) and (b) Matrix-assisted laser desorption ionization (MALDI). (a) During ESI, a fine spray of charged droplets is created. The droplets evaporate and the multiply-charged ions are expelled by electrostatic forces. ESI is frequently used for analysis of complex peptide mixtures via coupling to multidimensional liquid chromatography. (b) During MALDI, the analyte molecules are ejected by laser pulses from the sample cocrystallized with a matrix. During MALDI, mostly singly charged ions are produced. Because the analyte molecules are ejected in bundles, MALDI method is ideally suited for coupling to TOF mass analyzers. MALDI-TOF instruments are frequently used to analyze individual spots on 2D PAGE.

MALDI produces predominately singly charged ions by aiming laser pulses at a sample cocrystallized with a molecular matrix on a metal plate under high voltage (~20 kV). Typical matrices used in MALDI-assisted protein analysis include α-Cyano-4-hydroxycinnamic acid and 2-(4-Hydroxyphenylazo)-benzoic acid. The laser is tuned to the absorption maximum of the matrix. The sample ions are preformed in the condensed phase in sufficient quantities. The matrix absorbs some of the laser pulse energy, thus minimizing sample damage. The sample ions and matrix molecules gain enough kinetic energy and are ejected into a gas phase (Fig. 2b). Table 1 outlines and compares principle characteristics of these ionization methods.

Importantly, in both the MALDI and ESI sources, ion fragmentation occurs rarely and the analyte molecules remain largely intact. This nondestructiveness is what makes these methods so attractive for characterization of biomolecules. However, in certain applications (e.g. peptide sequencing), it may be necessary to fragment molecular ions (preferably in a predictable manner) in order to extract additional information. This is typically achieved via collision-induced dissociation (CID). Such types of MS analysis are called *tandem MS* (MS/MS) and often are denoted by MS^n, where n is the number of generation of fragment ions analyzed (Fig.1).

1.4.2 Mass Analyzers

Mass analyzers separate ions according to their m/z in electric or magnetic fields. Most common types of mass analyzers used in proteomics research are quadrupole (QD), ion trap (IT), time-of-flight (TOF), and Fourier transform ion cyclotron resonance (FTICR). Often in modern instruments, several analyzers of the same or different types are combined together to achieve maximum performance.

Tab. 1 Comparison of ESI and MALDI sources.

	ESI	*MALDI*
Principle of action	Uses electric field to produce sprays of fine droplets; as the droplets evaporate, ions are formed	Uses laser pulses to desorb and ionize analyte molecules cocrystallized with a matrix on a metal surface
Ions formed	Multiply charged (The larger the analyte molecule, the more likely it acquires multiple charges)	Singly charged
Mass range	>100 kDa	>100 kDa
Resolution	~2500 (with IT/QD mass analyzers)	~10 000 (with TOF mass analyzers and ion reflectors)
Typical application	10^{-15} mole LC/MS of peptide mixtures; tandem MS; protein identification by comparing experimental MS/MS spectra with theoretical MS/MS spectra produced from protein databases	10^{-15} mole Analysis of spots on 2D PAGE; determination of molecular weight; protein identification by "peptide mass fingerprinting"

QD mass analyzers are frequently used in conjunction with ESI source. They offer moderate resolution (up to 2500 mass units) and moderate sensitivity. QD consists of four parallel rods with a hyperbolic cross section. Diagonally opposite rods are connected to radio frequency and direct-current voltage sources thus establishing the quadrupole field. Ions that are produced at an ion source are electrostatically accelerated into this quadrupole field. Mass-selection is achieved by proportional changes of amplitudes of radio frequency and direct current in a way that for any given pair of these amplitudes there, only the ions of specific m/z reach the detector. QDs are known for their tolerance of high pressures (up to 10^{-4} torr), which makes them attractive to use with ESI for liquid chromatography (LC)-coupled applications.

IT mass analyzers work by confining ions to a small volume via radio-frequently oscillating electric fields. ITs offer moderate sensitivity at a relatively low monetary cost and are good for MS/MS applications. In the latter case, ITs are often used cojointly with QDs. Mass accuracy of regular ITs is rather low, since only limited amount of ions can be accumulated in a small volume. Mass accuracy, as well as resolution can be significantly improved by using two-dimensional ITs (sometimes also called *linear*), which increase the ion storage volume.

TOF mass analyzers measure the time traveled by an ion from a source to a detector. The longer the flight path, the better the resolving power. However, longer flight paths also increase the scan time. In commercial TOF analyzers, compromise is achieved with the length of the flight path on the order of several meters. The resolving power of a simple TOF analyzer is poor – several to ten times less than that of QD. Significant improvement in the resolving power of TOF is offered by additions of one or several ion reflectors. Upon reflection, the velocity distribution of ions at particular m/z narrows, thus increasing resolution and sensitivity. As a result, modern TOF and double-TOF spectrometers can achieve resolving power of 10 000 and more. Measurement of the ion's TOF can be achieved only if the analyte ions are presented to TOF analyzer in discrete bundles. Because of this need for discrete ion bundles, TOF analyzers are particularly suited to the pulsed nature of MALDI. In fact, MALDI-TOF is one of the most common types of mass spectrometers used in proteomics research.

FTICR mass analyzers are based on the resonance absorption of energy by ions that precess in a magnetic field. The recorded array of the precession time-curves is Fourier-transformed to obtain the component frequencies of the different ions. Next, the component frequencies are related to the ion's m/z. Because frequency is a parameter that is easy to measure with high precision and accuracy, FTICR has the highest resolving power amongst MS analyzers – up to 10^6 and more. With FTICR, it is also particularly easy to do MS^n experiments. Unfortunately, current FTICR instruments are cumbersome, expensive, and not readily available.

1.4.3 MS/MS

In some applications, it is necessary to fragment molecular ions produced by ESI or MALDI further to obtain additional structural information. Figure 3 illustrates a possible way to do this with a triple–QD ESI mass spectrometer. To obtain a tandem spectrum, the first quadrupole scans across a set m/z range and selects ions of interest. In the second quadrupole, CID

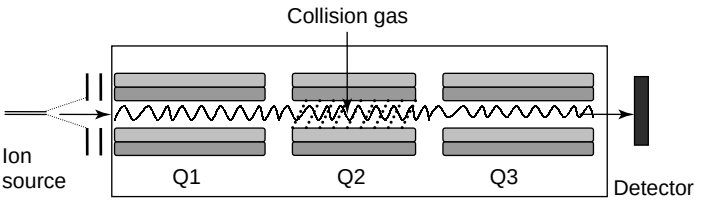

Fig. 3 A triple quadrupole mass spectrometer with MS/MS capability. Q1 selects ion to be fragmented and allows the selected ion to pass into Q2. The ions in Q2 are fragmented via collisions with inert gas (typically argon). Q3 analyzes the fragments generated in Q2. MS^n ($n \leq 4$) experiments can be performed with this setup as well. To achieve this, fragmentation must be increased at the level of ionization. In modern instruments, quadrupole-ion trap mass analyzers are used for MS^n ($n > 8$) experiments.

reaction takes place – ions that were selected by the first quadrupole undergo collisions with argon gas and fragment. Finally, the third quadrupole analyzes the resulted fragments. Importantly, the fragmentation via CID occurs in a predictable manner – protein and peptide ions break mostly at peptide bonds – which makes the large-scale sequencing easier.

1.5 Methods of Sample Fractionation

The large complexity of protein mixtures from biological samples poses additional challenges for proteomics experiments. Depending on the scope of a particular proteomics task, it is desirable to introduce sample enrichment and fractionation steps prior to MS identification. The scope of a particular task can range either from an analysis of a subset of a proteome – for example, analysis of proteins specific to a particular organelle and proteins modified in a certain way – to a simultaneous analysis of all the proteins in the proteome. In both of these cases, a proper choice of separation strategy determines the overall throughput and sensitivity of the proteomics experiment. While it may be desirable to fractionate protein mixture down to individual proteins (aim of 2D PAGE separation), significant gain in the throughput can be achieved when mixtures of proteins are introduced to the mass spectrometer.

MS analysis of intact proteins is impractical in most of the high-throughput tasks with the current instrumentation because of the large range in protein masses. To simplify the measurements, in most of the MS analyses, proteins are enzymatically digested (either individually purified or in mixtures) to produce peptide fragments. The resulted peptide mixtures are often fractionated further. Fractionation of the peptide mixtures is usually achieved via in-liquid chromatography methods.

Isolation of a particular subcellular compartment or an organelle is usually achieved by a combination of centrifugation and solubilization steps. Proteins modified posttranslationally in a specific way (e.g. phosphoproteins) can be isolated by chemical- or immunoaffinity methods (see examples in Sect. 3). Large-scale isolation of protein complexes can be done by the epitope tagging and subsequent immunoaffinity purification (see examples in Sect. 3). Another promising affinity

method is that of protein chips – a method that uses large arrays of antibodies, or other binding factors to isolate proteins of interests.

1.5.1 2D-electrophoretic Separation of Complex Protein Mixtures

2D PAGE separation methods were introduced in mid-1970s and were extensively used for analysis of complex protein mixtures. The principle of 2D PAGE separation is illustrated in Fig. 4. 2D PAGE separates proteins in the first dimension by their pI, and in the second, by their molecular weight. Prior to the development of MS-identification techniques suitable for macromolecules, identification of individual proteins in the gel spots was difficult. Typically, the identification of individual proteins in the gel spots was done by immunostaining methods or by N-terminal degradation sequencing. Nowadays, in most laboratories, analysis of 2D PAGE spots is performed with MS. Typically, the individual spots are excised, digested, and analyzed by MALDI-MS. Protein identification is achieved by peptide mapping – comparison of the observed peptide peak patterns with the predicted digest fragments of proteins in a database. However, if some of the spots on the 2D PAGE are overlapping, the method of peptide mapping can give incorrect results.

As a separation technique, 2D PAGE offers high resolution, and is able to distinguish between different protein isoforms and also between different posttranslational modification

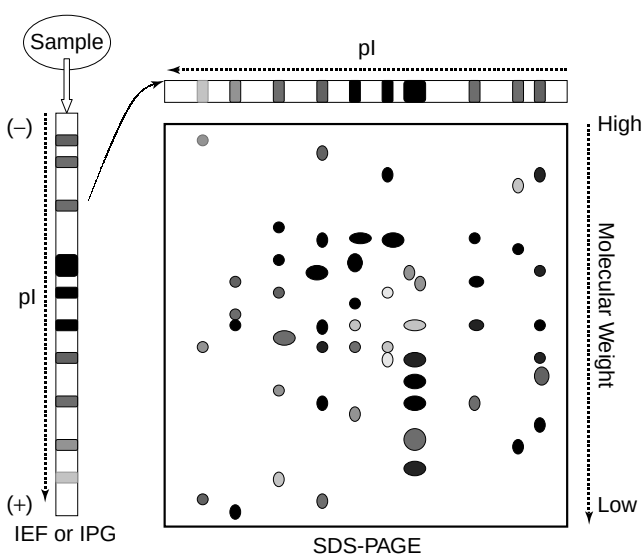

Fig. 4 Principle of the two-dimensional electrophoretic separation of intact proteins. The sample is loaded onto IEF to IPG strip, where proteins are separated according to their isoelectric points (pI). Next, the strip is loaded on SDS-PAGE, where the proteins are separated according to their molecular weights (MW). Visualization of protein spots is achieved via chemical staining methods. Individual spots can be further excised and analyzed by mass spectrometry.

states. However, 2D PAGE is a denaturing technique; hence, it is not suitable for direct analysis of protein complexes and protein–protein interactions. Also, 2D PAGE is known to discriminate against low-abundance and membrane proteins. Other limitations of 2D PAGE include biases against proteins with pIs outside the 2–10 range and biases against heavy proteins (>100 kDa). The dynamic range of 2D PAGE is also limited by resolution of the spot visualization methods. Recently, a number of ways to overcome some of these limitations of 2D PAGE have been developed.

Significant improvement to 2D PAGE was the development of immobilized pH gradients (IPGs), which increased both reproducibility and the load capacity of 2D PAGE. Sample prefractionation into discrete isoelectrical fractions and subsequent analysis by several narrow-range IPGs is yet a further improvement to this technology, which both increases the method's dynamic range and somewhat reduces the biases in proteome coverage. Despite these technological improvements, 2D PAGE remains a time-consuming and labor-intensive technique. Alternative methods of protein separations employing LC exist and can potentially overcome some of the 2D PAGE's limitations in the large-scale proteome analysis.

1.5.2 Liquid-phase Separation Methods

An ESI source allows MS characterization of proteins and peptides that are present in solutions and therefore is ideally suited for online coupling with liquid-phase separations. Liquid-phase separation can be achieved by using high performance liquid chromatography (HPLC), capillary isoelectric focusing (CIEF), capillary electrophoresis (CE), or by other methods. When two or more distinct liquid phases are used in a separation, with each of the phases relying on a unique independent physical property of the analyte, this liquid-phase separation is called *multidimensional*. Independent physical properties correspond to "dimensions" and these can be size, charge, hydrophobicity, or affinity to a particular substrate. The use of several independent dimensions significantly increases resolving power of a separation.

In-liquid separation methods can be used both for the intact proteins and for the peptide mixtures. One promising technique for characterization of global extracts of intact proteins is CIEF-FTICR mass spectrometry. In this method, the CIEF separates intact proteins by pI and subsequent analysis by FTCIR produces a two-dimensional display similar to the one obtained in 2D PAGE. Jensen et al. reported resolution of 400 to 1000 proteins in the mass range of 2–100 kDa from global protein extracts of *Escherichia coli* and *Deinococcus radiodurans*. While of good resolving power and sensitivity, this technique needs further improvement, perhaps by addition of protein fragmentation so that the resolved proteins could be identified.

Wall et al. developed a method that can potentially obtain both intact protein molecular weights as well as sequence information for a large portion of a proteome. In this method, proteins were separated by pI using isoelectric focusing (IEF) in the first dimension and by hydrophobicity using nonporous reversed-phase HPLC in the second dimension (IEF-NP RP HPLC). Next, the fractions that were eluted from HPLC were analyzed in two parallel experiments by (1) MALDI-TOF-MS and (2) ESI-TOF-MS. Analysis of human erythroleukemia cell lysate by this method resulted in resolution and identification of

several hundreds of unique proteins in a mass range from 5 to 85 kDa. Having both intact protein molecular weights as well as sequence information alleviates characterization of protein posttranslational modifications and isoforms. Thus, Wall et al. reported detection of posttranslational modifications of cytosolic actin, heat shock 90 beta, HINT and α-enolase.

The intact protein approaches are usually limited to proteins of smaller sizes, mostly because the resolving power of mass analyzers drops with increase in the ions masses. Also, biases toward one or the other types of intact proteins are unavoidable with any chromatographic separation method. A conceptually different approach, in which purification of intact proteins is avoided, involves proteolytic digestion of protein mixtures. Proteolytic digestion transforms protein mixtures into more uniform (in terms of mass and chromatographic properties) mixtures of peptides. Because of the increased uniformity, analysis of peptide mixtures essentially eliminates biases in proteome coverage. Unfortunately, the absence of information on intact proteins makes characterization of posttranslational modifications more difficult.

Chemical or enzymatic cleavage of proteins into peptides followed by multidimensional liquid-phase separation is typically used in a global proteome survey type of experiments. Generally, proteins from whole cells or tissue homogenates are first enzymatically digested and then the resulting peptide mixtures are separated by electrophoresis or liquid chromatography in a microcapillary format. A microcapillary column is attached directly to the ionization source so that the eluting fractions are ionized and introduced to a mass analyzer. While fragmentation of the peptide ions is optional with MALDI-MS/peptide mapping analysis of 2D PAGE spots, it is required with the liquid-phase/MS analysis of complex peptide mixtures. The sequence information obtained by MS/MS for individual peptide ions is used to match peptides to the corresponding proteins in a database.

Multidimensional protein identification technology (MudPIT) is a good example of a liquid-phase/MS method. In MudPIT, a biphasic SCX/RP microcapillary column is used to fractionate a complex peptide mixture (Fig. 5). The peptide fractions are eluted from the column by series of HPLC gradients into the tandem mass spectrometer. The obtained MS/MS spectra are used to search a sequence database via SEQUEST algorithm. MudPIT was used to detect and identify ~1500 proteins from *Saccaromyces cerevisiae* proteome, ~2500 proteins from the *Oryza sativa* proteome and ~2500 proteins from the *Plasmodium falciparum* proteome.

1.5.3 Affinity Methods (Epitope Tagging)

Affinity-purification methods are frequently employed in those proteomics programs that analyze specific subsets of a proteome, such as phosphorylated or glycosylated proteins. Also, affinity methods are often used in large-scale studies of protein–protein interactions and protein complexes. Affinity methods are based on targeted interactions between proteins and antibodies or between proteins and other protein-specific molecules, immobilized in a column or in the form of an array. For example, to isolate phosphorylated proteins from the rest of the proteome, one can use an immunoaffinity column prepared with antibodies specific to the phosphorylated amino acids.

Epitope tagging is used in large-scale analyses of protein–protein interactions and protein complexes (Fig. 6). In this

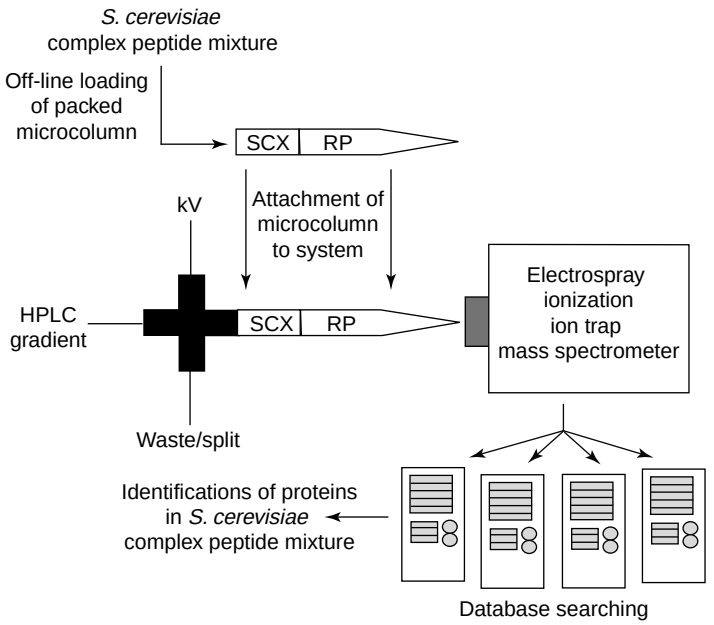

Fig. 5 Multidimensional protein identification technology (MudPIT). Complex peptide mixtures are loaded onto a biphasic microcapillary column packed with strong cation exchange (SCX) and reverse-phase (RP) materials. Peptides directly elute into the tandem mass spectrometer because a voltage (kV) supply is directly interfaced with the microcapillary column. Peptides are first displaced from the SCX to the RP by a salt gradient and are eluted off the RP into the MS/MS. The tandem mass spectra generated are correlated to theoretical mass spectra generated from protein or DNA databases by the SEQUEST algorithm. Figure is reproduced from Washburn, M.P., Wolters, D., Yates, J.R. III (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology, *Nat. Biotechnol.* **19**, 242–247 by permission of Nature Publishing Group.

strategy, proteins to be isolated are fused with a motif recognizable by a specific antibody. This fusion is typically done by incorporating the epitope sequence into C-terminal of genes that encode proteins of interest. Then the fused proteins are expressed and purified along with their interaction partners by the immunoaffinity chromatography using antibodies specific to this particular epitope. If a particular protein is of low abundance, it can be overexpressed to increase the overall recovery.

1.6 Protein Identification by MS

Accurate identification of proteins is the essential requirement of proteomics studies. High complexity of protein mixtures derived from tissues, whole cells, or subcellular compartments adds an additional requirement of high throughput in the identification of proteins. In this section, we review common strategies that match MS data with proteins present in a biological sample.

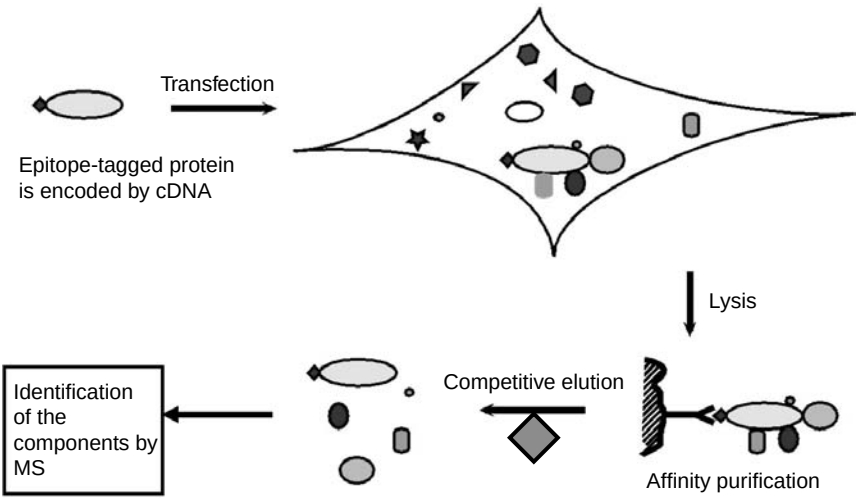

Fig. 6 Principle of affinity purification with epitope tagging. Bait proteins are expressed as fusion with a motif recognizable by a certain antibody (epitope). The cells are lysed, and proteins are purified by immunoaffinity chromatography. Next, the bound fraction is competitively eluted from the column, and the pulled proteins are analyzed by LC/MS or 1D PAGE/MS.

Typically, protein identification amounts to the deduction of the sequence-specific information from MS data followed by searches of the sequence databases. Databases that MS data can be matched against are protein, expressed sequence tag (EST), and genomic databases. Because currently there are no suitable ways to fragment intact proteins inside a mass spectrometer, high-throughput sequencing by MS is possible only if intact proteins are fragmented into peptides first. Depending on a particular type of proteomics scheme and MS instrumentation used, protein sequence information can be obtained by (1) peptide mass fingerprinting; (2) accurate mass tags (AMTs); (3) peptide fragmentation in MS/MS; (4) sequence tags; or by combinations of (1) through (4).

Peptide mass fingerprinting method is usually adopted in those cases in which individual proteins are separated during purification steps, such as in 2D PAGE. In the case of 2D PAGE, individual spots are picked, digested, and masses of the peptides are recorded. Next, the experimentally obtained peptide masses are matched against theoretical peptide libraries generated from protein sequence databases. MS/MS data can also be obtained along with mass fingerprinting, and nowadays this is a common practice.

With MALDI-TOF instrumentation – which is typically used with the peptide mass fingerprinting type of analysis – several peptide masses are needed to unambiguously identify a protein. Using more accurate instrumentation, such as FTICR-MS, it is possible to identify proteins based on the mass of a single peptide, without MS/MS data. In this case, a peptide that uniquely corresponds to a protein is called *accurate mass tag* (AMT). Conrads et al. evaluated utility of AMTs for identification of proteins from *S. cerevisiae* and *Caenorhabditis elegans*. The

authors demonstrated that up to 85% of the predicted tryptic peptides from these two organisms could be used as AMTs at mass accuracies typical of FTICR-MS instruments (~1 ppm). The authors also discussed utility of AMTs with highly accurate mass measurements in detection of phosphorylated proteins. They argued that because mass defect of P is larger than that of H, C, and O, the average mass of phosphopeptides is slightly lower than the mass of unmodified peptides of the same nominal weight; thus enabling the identification of phosphorylated peptides if the mass measurement accuracy is sufficient.

Whenever it is possible to match a single peptide to a protein, it is no longer necessary to purify samples down to individual proteins. Instead, fractions containing mixtures of proteins can be digested with trypsin followed by analysis of the resulted peptide fragments. For example, the above-discussed AMT method can be used for analysis of complex protein mixtures, because it matches single peptide to a unique protein. However, the requirement of high mass accuracy makes uses of AMT limited. Additionally, FTICR instrumentation is cumbersome and is not a widely distributed technology today. Far more superior in terms of proteome coverage per monetary cost are methods that employ peptide sequence analysis by MS/MS. The key fact that enables high-throughput peptide sequencing by MS/MS is predictability of the peptide precursor ion fragmentation in the CID reactions.

Figure 7 shows the adopted nomenclature for the fragment ions series. Ions that form through dissociation of the peptide bonds are the most abundant ones if moderate collision energies are used (30–50 V). The ions that retain their charge on the N-terminal part after fragmentation of the precursor are called *y* ions, and the

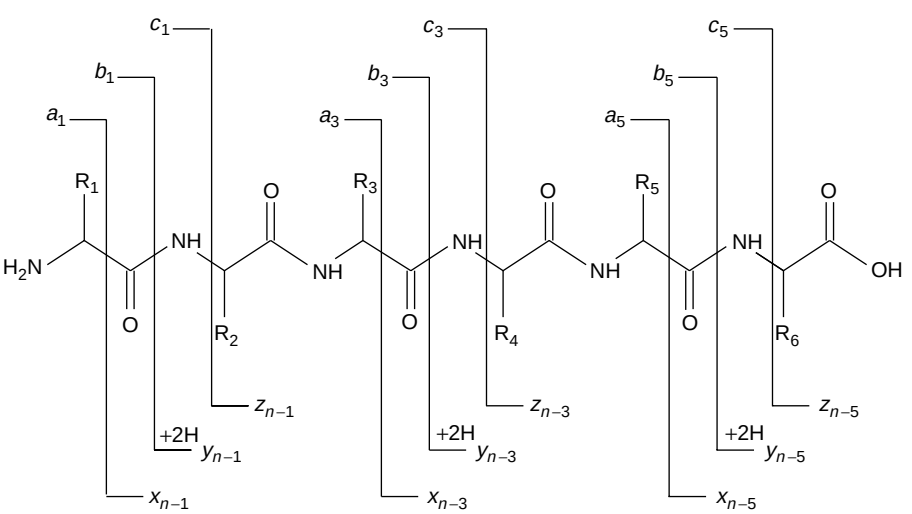

Fig. 7 Nomenclature of the ions formed during peptide fragmentation. In a typical MS/MS experiment via CID reactions, *y* and *b* ions are predominant. These two types are used for computer-assisted peptide identification. Through the loss of CO fragment, *a* ions can be produced from *b* ions. While not important for identification, *a* ion series can be used for independent result validation. Other types of ions are not produced in CID reactions.

ions that retain their charge on the C-terminal part after fragmentation of the precursor are called *b* ions. The subscript to the right of *b* and *y* symbols equals the number of amino acid residues in the corresponding fragment. The difference in m/z values between consecutive ions within a given series corresponds to the difference in the sequences of the two fragments. Because the consecutive ions within a series represent peptide fragments that differ in exactly one amino acid, and each amino acid residue has a unique nominal weight (except I and L), the pattern of m/z values of *y* and *b* ions corresponds to the amino acid sequence of the precursor peptide. Unfortunately, some expected peaks could be missing from the MS/MS. Additionally, experimental spectra can be complicated by unwanted fragmentation. Also, it is not always possible to unambiguously determine from which series a particular ion fragment comes. For these reasons, manual interpretation of MS/MS spectra can be tedious and ineffective. As a consequence, *de novo* sequencing is rarely done in high-throughput proteomics experiments. Instead of direct interpretation of the MS/MS spectra, computer programs that find the best matches to the spectra from a database are used.

Typical example of an algorithm that finds best matches to MS spectra is SEQUEST software package developed by Eng et al. Analysis of MS data by SEQUEST starts with reduction of tandem spectra complexity. Only a certain number of most abundant ions are considered and the rest are discarded. Also, the unfragmented precursor ion is removed from the spectra in order to prevent its misidentification as a fragment. Next, SEQUEST selects sequences from a database. First, SEQUEST creates list of peptides that have masses at or near the mass of the precursor ion. Second, SEQUEST generates virtual MS/MS spectrum for each of the candidate and compares them to the observed spectrum. As the result of this comparison, a cross correlation score is produced (Xcorr). Another score, DeltaCN, reflects the difference in correlation of the second ranked match from the first one. The higher this score is, the more likely it is that the first match is the correct one.

SEQUEST also has the ability to detect posttranslational modifications. In this case, SEQUEST looks for increased masses of amino acids. For example, cysteines are typically carboxymethylated by iodoacetomide prior to enzymatic digestion. In this case, masses of all cysteine residues are considered to increase by 57 Da. To look for amino acids that can be either modified or not modified, the database size is increased by allowing different weights to represent the same amino acid. Unfortunately, only known posttranslational modification can be detected by SEQUEST – the types of modifications that are being looked at need to be explicitly specified in input parameters.

Another useful technique for peptide identification is sequence tag. This technique is implemented in algorithms such as GutenTag. In GutenTag, a partial sequence is inferred directly from the tandem spectrum, and then the database is searched for matches that include this partial sequence and that match the masses on C- and N-sides of the fragment. Importantly, GutenTag method is error-tolerant, and allows detection of unknown posttranslational modifications. A recent variation of this method, MultiTag also allows protein identification from organisms with unsequenced genomes.

2 Quantitative Methods of Proteome Analysis Using MS

Quantitative proteome analysis aims at large-scale identification of differences in protein expression. Two-dimensional difference gel electrophoresis (2D-DIGE) is one of the techniques that can achieve that aim. Recent example of application of this technique is the report by Friedman et al. The authors used 2D-DIGE in combination with MALDI-TOF to analyze the proteome of human colon cancer. However, while acknowledging the importance of the gel-based methods in modern quantitative proteomics, we believe that methods based on multidimensional chromatography have greater potential, and therefore, we limit our further discussion to quantitation within the LC/LC/MS/MS paradigm.

Even though absolute quantification (i.e. clear relationship between the peak intensity and amount of the analyte) is a challenge for modern mass spectrometers, methods of finding relative abundances exist. These methods are based on labeling through mass modifications of the whole proteome or some of its subsets. Labeling through mass modification is particularly suited for quantitative analysis in differential expression profiling, where two states under study are differentially labeled via "light" or "heavy" mass modifications. Figure 8 schematically depicts a quantitative proteomics approach. Mass modification can be introduced at different steps in the sample preparation – during growth, after growth, during digestion, or after digestion. Labeling at the early stages of sample preparation minimizes losses of the analytes. Instrumentation for quantitative studies is essentially the same as instrumentation for qualitative studies. Nevertheless, quantitative analysis is more laborious and typically achieves lower proteome coverage. Also, quantitative methods in proteomics are still at the stage of development and are not as broadly applied as qualitative methods.

2.1 Metabolic Labeling

Widely used in structural biology to prepare samples for analysis by nuclear magnetic resonance (NMR) analysis, metabolic labeling methods have become useful quantitative proteomics tools. Metabolic labeling is a mass modification method that is done very early in the experiment – at the stage of cell growth. Figure 8 illustrates the principle behind this strategy. The goal is to compare protein abundances in cells grown at different conditions. To do this, the growth media from one of the conditions is enriched in stable low-abundant isotopes. The enrichment can be done either by labeling all amino acids by ^{15}N, or by supplementation with a single labeled amino acid. The processed and digested cell extracts from the two different conditions are combined in a one-to-one ratio and analyzed by liquid-phase/MS/MS methods. In the resultant MS spectra, peaks corresponding to the same peptide from different conditions are offset according to the degree of labeling. The ratio between the two peaks corresponds to the difference in abundances.

The first demonstrations of metabolic labeling in quantitative proteomics used gel electrophoresis and spot excision as the protein isolation method. For example, Oda et al. identified proteins that were altered in expression between two strains of *S. cerevisiae* grown in ^{14}N or ^{15}N media, and determined the phosphorylation levels of a specific protein in the same

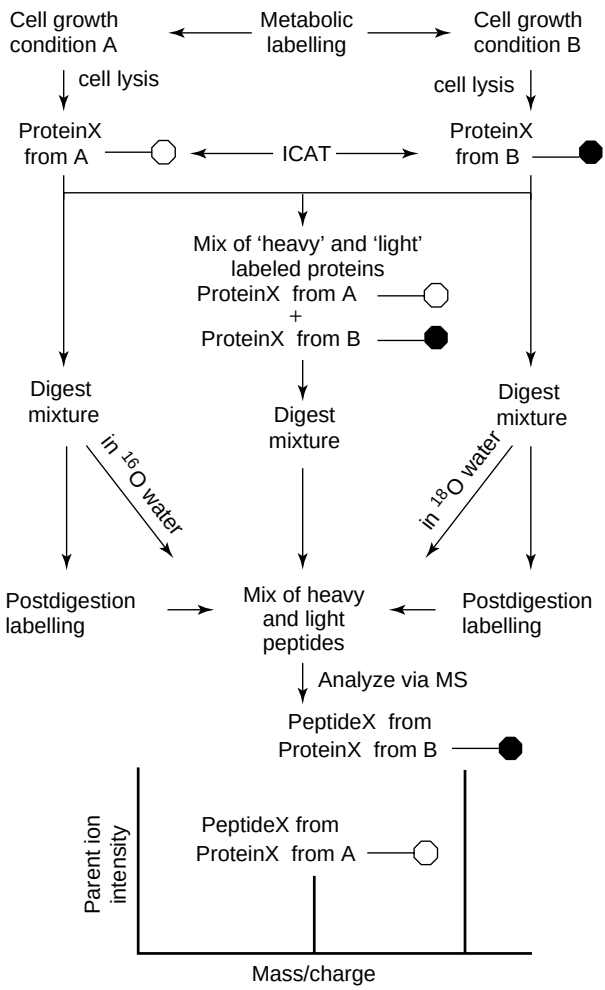

Fig. 8 Quantitative proteomics approach. When carrying out a quantitative proteomic analysis, the key is for the same peptide from two unique growth conditions to have unique masses when being analyzed by a mass spectrometer. "Heavy" and "light" peptides may be generated at many points in a sample preparation pathway. Metabolic labeling introduces a label during the growth of the organism and is therefore the earliest point of introduction of "heavy" and "light" labels. Metabolic labeling is followed by ICAT, digestion in ^{16}O and ^{18}O water, and lastly postdigestion labeling. Only after a label has been introduced can the samples be mixed and further processed. Figure and figure legend are reproduced from Washburn, M.P., Ulaszek, R., Deciu, C., Schieltz, D.M., Yates, J.R. III (2002) Analysis of quantitative proteomic data generated via multidimensional protein identification technology, *Anal. Chem.* **74**, 1650–1657 by permission of Analytical Chemistry.

sample. Complete metabolic labeling has been demonstrated using chromatography based proteomics methods on *D. radiodurans*, mouse B16 cells, and on *S. cerevisiae*.

In addition, isotopically enriched single amino acids may be used for the selective metabolic labeling of a cell type for a quantitative proteomic analysis. In *S. cerevisiae*, Jiang et al. have described the single amino acid isotopic enrichment of *S. cerevisiae* with D_{10}-Leu, and Berger et al. described the comparative analysis of *S. cerevisiae* cultured in media containing either ^{13}C-Lys or unlabeled lysine. Stable isotope labeling by amino acids in cell culture (SILAC) was explored as an alternative to ^{15}N labeling. Ong et al. studied mammalian cells grown either with deuterated or nondeuterated leucine. Importantly, parameters such as cell morphology, doubling time, and ability to differentiate did not change in deuterated sample compared to nondeuterated one. The authors used this technique to study changes in protein expression induced by muscle cell differentiation. The authors reported that glyceraldehyde-3-phosphate dehydrogenase, fibronectin, and pyruvate kinase M2 were upregulated. Another approach to metabolic labeling is the rare isotope depletion of the growth media from one of the conditions under study. If the rare isotopes are removed, then one expects m/z distributions to shift to the lighter values. With conventional instrumentation, this could work for large proteins. However, high-resolution instrumentation like FTCIR must be used to analyze peptide mixtures.

2.2 Isotope Coded Affinity Tags

Metabolic labeling with stable isotopes while analytically advantageous to other methods is available only in cases when studied cells or organisms are cultivable. For this reason, metabolic labeling is not suitable for diagnostic and clinical applications. When cultivation or controlling growth is difficult, different strategies need to be used for quantitative proteome analysis. One of the possible approaches is the method that uses cysteine-specific reagents – isotope coded affinity tags (ICATs) – to differentially label proteomics samples. ICATs have three functional elements: cysteine-specific reactive group, isotopically labeled linker, and affinity group (Fig. 9). For the comparative analysis, the cysteine residues in samples are separately labeled with either labeled or unlabeled ICAT. The derivatized samples are then combined in a one-to-one ratio and digested with a protease resulting in both labeled and unlabeled peptide fragments. The labeled peptide fragments are then purified by affinity chromatography, fractionated by reversed-phase chromatography and analyzed by tandem mass spectrometry, which provides both qualitative analysis and the relative abundance of the peptide isoforms in the samples. The MS analysis is analogous to metabolic labeling case – ratio of "heavy" to "light" peptide ions correlates with their relative abundance. Also, the ICAT approach has the obvious conceptual limitation – only peptides that contain cysteine can be detected. It has been shown that in yeast only ~10% of all tryptic peptides contain cysteine. Therefore, full sequence coverage may not be possible even for the most abundant proteins.

Significant improvement to the ICAT method is the cleavable ICAT (cICAT), which is presently supplied by Applied Biosystems in a kit format. The cICAT reagent has four essential structural elements. The first is a protein reactive group

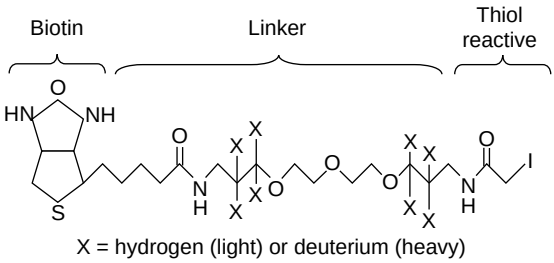

Fig. 9 Structure of isotope coded affinity tags. The ICAT reagent consists of a biotin group linked to a cysteine reactive group. The linker may be deuterated eight times or protonated at each site allowing for the generation of D_0- or D_8-ICAT. The differential masses of the linker group allow for the use of ICAT in a quantitative proteomic scheme. The figure and figure legend have been reproduced from Hunter, T.C., Washburn, M.P. (2003) The integration of chromatography and peptide mass modification for quantitative proteomics, *J. Liq Chromat. Rel. Technol.* **26**, 2285–2301 by permission of Marcel Dekker, Inc.

(iodoacetamide) that covalently links the isotope-coded affinity tag to the protein through alkylation of cysteines. The second structural element is the biotin affinity tag that allows enrichment of the tagged peptides. The third is an isotopically labeled linker ($C_{10}H_{17}N_3O_3$). Nine carbon atoms of the linker can be either ^{12}C or ^{13}C giving light and heavy version of the tag respectively. The light and heavy molecules have the same chromatographic properties, but differ in mass (9 Da). Once the sample is subjected to mass spectrometric analysis, the ratio of intensities between heavy and light peptides provides a relative quantitation of the proteins in the original sample. The fourth structural element of cICAT is an acid cleavage site that allows removal of part of the tag prior to MS analysis. After avidin-affinity purification of the cICAT-labeled peptides, biotin portion of the label and part of the linker can be removed by adding trifluoroacetic acid. This reduces the overall mass of the tag on the peptides and improves the overall peptide fragmentation efficiency.

The ICAT methodology has been successfully applied to a variety of biological questions including studies of several cell types, organelles, and different classes of proteins. Quantitative proteomic analysis via ICAT has been coupled with cDNA array analysis to investigate the galactose utilization pathway in *S. cerevisiae* and to investigate the mRNA and protein expression changes brought about by culturing *S. cerevisiae* in either galactose or ethanol. In addition, ICAT detected changes in protein expression of peripheral and integral membrane proteins by analyzing the effect on 12-phorbol 13-myristate acetate on the microsomes of HL-60 cells. Of all non-gel-based quantitative proteomic strategies, ICAT is the most mature as demonstrated by the successful use of ICAT in biologically driven analyses.

2.3 ^{18}O Labeling

The global modification of all proteolytic peptides in a mixture may be carried out via the labeling of carboxyl groups that occurs through incorporation of ^{18}O from $H^{18}O$ during proteolytic hydrolysis (Fig. 10). In order to introduce a 4-Da mass shift into the C-terminus of a peptide, proteins may be digested in the presence of $H_2{}^{18}O$. Proteases like trypsin will carry out this

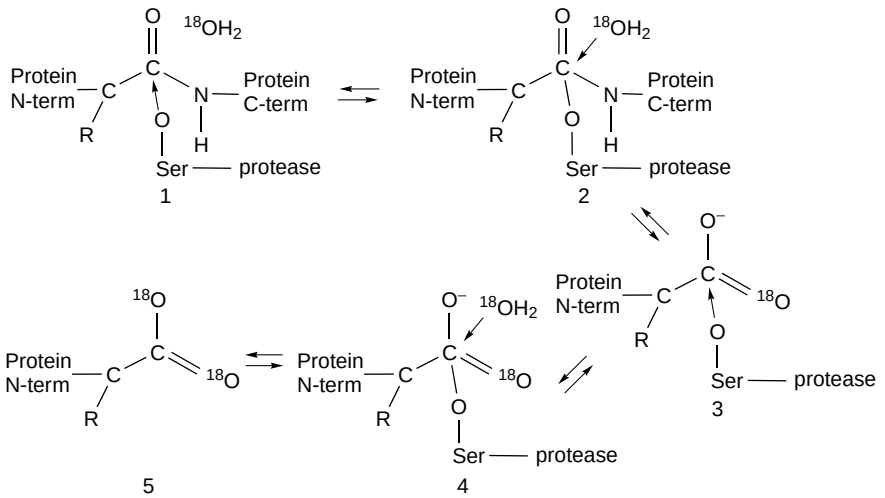

Fig. 10 C-terminal digestion modification with ^{18}O. ^{18}O may be incorporated into the C-terminus of a peptide during digestion with enzymes such as trypsin, endoproteinase Lys-C, and endoproteinase Glu-C. A simplified version of this reaction scheme is shown. (1) To begin, the peptide needs to be digested in $^{18}OH_2$ in order to then incorporate ^{18}O. The serine in the active site of the proteases listed attacks the carbonyl carbon in a peptide bond. (2) Next, $^{18}OH_2$ attacks the protein–protease intermediate also at the carbonyl carbon displacing the NH group on the peptide bond. (3) As a result, a peptide with a single ^{18}O has been generated. (4) A repeat of steps (1) and (2) is needed to drive the reaction to completion as shown in (5) where two atoms of ^{18}O have now been incorporated into the peptide C-terminus. Labeling of one sample with ^{18}O by digesting in $^{18}OH_2$ and mixing this with the other sample digested in ^{18}O depleted water allows for the determination of the relative abundance of peptides from a mixture. Figure and figure legend is reproduced from Hunter and Washburn (2003) by permission of Marcel Dekker, Inc.

reaction during the process of enzymatic cleavage. By mixing a sample with proteins digested in the presence of ^{18}O and the absence of ^{18}O, a pairwise comparison may be made to determine the relative abundance of peptides in a sample.

2.4 Postdigestion Labeling

There are several alternatives to the residue-specific modification of cysteine, which include methods for differential modification of lysine and O-phosphorylated serine residues. The phosphoprotein isotope coded affinity tag (PhIAT) method has been shown to be capable of enriching and identifying mixtures of low-abundance phosphopeptides. The PhIAT method uses a chemical modification of phosphorylated serine and threonine residues to cysteine before introduction of a standard ICAT reagent. The mass-coded abundance tag (MCAT) approach uses a residue-specific modification lysine residues by *O*-methylisourea to introduce a differential tag. In addition, 2-methoxy-4,5-dihydro-1H-imidizole has also been used to modify lysine residues for the purpose of introducing a differential mass tag.

The N/C termini of peptides after digestion may also be labeled through a variety of means. A C-terminal modification

method is methyl esterification of carboxyl groups using either methanolic HCl or the deuterated analog. In this case, multiple sites in a peptide may be modified with labels introduced at aspartic acids, glutamic acids, and C-termini. The N-terminal labeling of tryptic peptides with *N*-hydroxysuccinimide or 1-Nicotinoyloxysuccinimide esters and their stable isotope analogs is another approach that can be potentially used for quantitative proteomic analyses. In fact, coupling ^{18}O labeling and N-terminal labeling methods for protein expression profiling produced more comprehensive results than when either method was used alone.

2.5 Global mRNA and Protein Expression Analyses

An emerging application of quantitative proteomics approaches includes the large-scale analysis of protein expression correlated to large-scale mRNA expression analyses. In three independent comparisons of mRNA and protein levels in *Saccharomyces cerevisiae*, overall partial positive Spearman rank correlation coefficients ranging between 0.21, 0.45, and 0.57 were obtained. These studies employed ICAT, chromatography, and mass spectrometry, ^{15}N labeling and MudPIT or chemiluminescence of SDS-PAGE approaches to determine protein expression levels in cells grown under different conditions in each study. In all likelihood, this pattern of partial positive correlation between mRNA and protein expression levels could be expected to persist under a variety of conditions in *S. cerevisiae*. When these approaches begin to be globally applied to other organisms, it will be interesting to see if this trend persists.

3 Specific Examples of Applications

3.1 Global Proteome Sampling

The goal of global proteome sampling is the simultaneous identification of proteins in a cell or tissue at a given condition. Data obtained in such experiments can be further used to answer more specific biological questions, such as difference between healthy and pathological states. Typically, the proteins are identified by mass spectrometry and are grouped into functional categories.

3.1.1 Global Proteome Sampling Based on 2D Page

Global proteome analysis by 2D PAGE method is difficult because each spot needs to be picked and identified individually, thus increasing time and cost of the analysis. Also, as we discussed earlier, 2D PAGE separation suffers from biases – certain classes of proteins, such as hydrophobic or those of high molecular weight are difficult to detect. Apart from these limitations, 2D PAGE approach has an important advantage over other methods – it easily resolves protein isoforms.

In their analysis of *Haemophilus influenzae* proteome, Langen et al. used several techniques to maximize the 2D PAGE performance. To increase the proteome coverage, they used immobilized pH gradient strips covering several pH regions. Also, to visualize low-copy-number proteins, the authors performed a series of protein extractions, such as heparin chromatography, chromatofocusing, and hydrophobic interaction chromatography. In order to detect cell-envelope-bound proteins, the authors used immobilized pH gradient strips in

combination with a two-detergent system with a cationic detergent in the first and an anionic detergent in the second dimensions. The isolated proteins were identified by MALDI/MS and peptide fingerprinting. As a result, 502 unique proteins were identified (about 30% of all ORFs)

Analysis of the mouse brain proteome performed by Klose et al. provides a good illustration of what 2D PAGE can do for the global proteome sampling type of experiments. By using 2D PAGE, the authors performed comparative analysis of the two distantly related mouse strains, *Mus musculus* C57Bl/6 (B6) and *Mus spretus* (SPR). About 8700 proteins from the cytosolic fraction of brain proteome were compared between the two species. By analyzing 2D PAGE of B6 and SPR strains, as well as of F_1 (B6 × SPR) and B_1 (F_1 × SPR) hybrids, the authors detected 1324 species-specific polymorphisms. Among these, 466 proteins were identified by MALDI-TOF/MS using peptide mass fingerprinting. To detect the polymorphisms, the authors considered variations in electrophoretic mobility, spot intensity, and the number of different isoforms corresponding to one protein. Additionally, through the analysis of F_1 and B_1 generations, the authors established which polymorphisms were genetically dominant. The key feature that enabled this comprehensive study was the high quality of the 2D PAGE. To analyze the mouse brain proteome, the authors used the large-gel 2D PAGE, which employs IEF gel incubation, and large (46 × 30 cm) format. Implemented in this way, the 2D PAGE gives both high resolution and high sensitivity – more than 10 000 protein spots from mouse tissues can be visualized simultaneously.

3.1.2 Global Proteome Sampling Based on Multidimensional LC

When it is necessary to catalog proteins present in a cell or an organism in a given environmental context, the multidimensional LC separation of peptide mixtures followed by MS/MS is the most convenient method to use. While it is not as good at determining protein isoforms and posttranslational modifications as 2D PAGE, the biases in proteome coverage are greatly reduced.

Florens et al. performed proteomics studies of the life cycle of the human pathogen *Plasmodium falciparum* (malaria) life cycle. The authors identified 2415 proteins and assigned them to functional groups at four stages of the cycles (sporozoites, merozoites, trophozoites, and gametocytes). The sporozoite is the form in which *P. falciparum* is injected by a mosquito. The merozoite is the form that invades erythrocytes. The trophozoite is the form that multiplies in the erythrocytes. The gametocyte is the sexual stage of malaria parasite life cycle. The analysis was performed by MudPIT. The authors found that about 50% of sporozoite proteins were unique to that stage. In sporozoites, about 25% were shared with any other stage. Trophozoites, merozoites, and gametocytes had 20 to 30% unique proteins and they had 40 to 60% of their proteins shared. Only 6% of all identified proteins were shared between all four stages, which were mainly histones, ribosomal proteins, and transcription factors. Out of the 2415 identified proteins, 51% were previously annotated as hypothetical.

Koller et al. used both 2D PAGE and MudPIT to analyze *Oryza sativa* (rice) proteome. The analyses were performed on the protein extracts from leaf, root, and seed tissue. The goal of this study was to determine tissue-specific expression of

proteins. 2D PAGE separation followed by MS/MS yielded 556 unique protein identifications, comprising 348 proteins from leaf, 199 from root, and 152 from seed. MudPIT analysis resulted in significantly larger coverage: 2363 total proteins, with 867 from leaf, 1292 from root, and 822 from seed. A total of 165 proteins were uniquely detected by 2D PAGE, whereas 1972 proteins were uniquely detected by MudPIT. Next, the authors searched the nonredundant protein database by BLAST and grouped the identified proteins into functional categories. The largest category (32.8%) included proteins that had no homology to the predicted proteins. Proteins classified as involved in metabolism comprised 20.8% of all identified proteins. Out of the 2528 detected proteins, 189 were shared among all three tissues. These included housekeeping proteins that are involved in transcription, mRNA biosynthesis, translation, and protein degradation. However, most of the proteins had tissue-specific expression: 622 specific to leaf, 862 specific to root, and 512 specific to seeds.

To characterize the proteome of *S. cerevisiae* mitochondria, Sickmann et al. combined four separation methods: IEF-incubated 2D PAGE; digestion with four different proteases, followed by multidimensional LC/MS/MS; SDS/PAGE combined with multidimensional LC/MS/MS; treatment of mitochondria with trypsin, followed by SDS/PAGE or HPLC and MS/MS. The authors identified a total of 750 mitochondrial proteins. When classified into functional categories, 24.9% of all the identified proteins were of unknown function, 24.9% were involved in genome maintenance and gene expression, and 14.1% were involved in energy metabolism. The rest of the identified proteins were involved in metabolism, transport, and cell rescue.

3.1.3 MS-assisted Disease Diagnosis from Serum Samples

Proteomics technologies recently emerged as a useful tool in clinical disease diagnosis. For example, Petricoin et al. used surface-enhanced laser desorption ionization time-of-flight (SELDI-TOF) and artificial-intelligence-based informatics algorithms to discriminate between control group and ovarian cancer patients. First, the authors generated a preliminary training set of mass spectra derived from 50 unaffected women and 50 women with ovarian cancer. Next, they used an iterative searching algorithm to find the best discriminatory pattern amongst these MS data. As a result, the algorithm correctly identified all cancer cases in the masked set. Additionally, out of 66 cases of malignant disease, only 3 were recognized as cancer (false-positives). Thus, the study by Petricoin et al. demonstrated good sensitivity and predictive power of MS-based proteomics when applied to clinical disease diagnosis. For further information on this subject, we refer the interested reader to the comprehensive review by Rosenblatt et al.

3.2 Analysis of Protein Modifications by Mass Spectrometry

In living organisms, protein activity is regulated mainly by covalent modifications, which occur either co- or post-translationally. Identification of types of modifications and their locations is often a necessary requirement for an understanding of the regulation and function of a given protein. There are hundreds of known protein modifications. Among

these, phosphorylation is, perhaps, the most important and widespread – about one-third of all proteins from mammalian genomes are thought to be phosphorylated. Another functionally important modification is glycosylation – glycosylated proteins are ubiquitous components of cellular surfaces where their oligosaccharide groups participate in a wide range of cell–cell recognition events. Comprehensive analysis of glycosylated proteins is more challenging than analysis of other modifications, mainly because the structure of oligosaccharide varies.

Other commonly occurring modifications that are involved in protein regulation and function are disulfide bonds, acetylations, and ubiquitinations. Some of these and other modifications are listed in Table 2. Changes in a protein length, either as a result of alternative splicing or protein truncations, also may be considered as protein modifications. Generally, it is difficult to identify protein truncations by methods that deal with protein/peptide mixtures, and often purification down to individual proteins is required in such cases (e.g. by 2D PAGE).

Currently available methods of large-scale analysis of modified proteins can be grouped into two major classes: those that use sample enrichment or chemical treatment prior to MS, and those that rely on MS data alone. Enrichment methods include affinity purification, chemical tagging followed by affinity purification, and immunoprecipitation. MS methods of detection and identification of modified peptides include neutral loss scan, precursor ion scan, postsource decay, and others.

Sometimes it is of special interest to obtain information on several types of modifications at once. If that is the case, computer programs such as GutenTag (see Sect. 2) can be used to analyze MS/MS data. Additionally, if mass changes introduced by modifications are known, the search for modified proteins can be done by SEQUEST with input parameters modified in accordance with the mass changes. However, generally, it is difficult to analyze modified peptides in the background of nonmodified peptides. Therefore, when it is clear what type of modification needs to be analyzed, the fractionation steps that enrich that particular modification need to be introduced into the experimental scheme.

Tab. 2 Common protein modification.

Modification	*Monoisotopic/average mass change*
Phosphorylation[a]	+79.9663/79.9799
Acetylation	+42.0106/42.0373
N-acetylglucosamine (GlcNAc)[a]	+203.0794/203.1950
Disulfide bond	−2.01565/2.0159
Methylation	+14.0157/14.0269
Hydroxylation	+15.9949/15.9994
Oxidation of methionine	+15.9949/15.9994
Ubiquitination of lysines	+114.0429/114.1040[b]

[a]Occurs on tyrosine, serine, threonine. Widespread throughout the proteome. Functions include protein regulation, signal transduction.
[b]Mass change is due to Gly–Gly residue, which is left on ubiquitinated lysines after trypsin digestion.

Below we discuss several illustrative examples of analysis of phosphorylated, glycosylated, and ubiquitinated proteins from recent literature. In addition, tools and techniques used in analysis of phosphorylation are generally applicable to analysis of many other modifications such as methylation and acetylation, while analysis of glycosylation poses additional analytical and instrumental challenges due to variability in structures of oligosaccharide groups. Ubiquitination, while not as frequent as phospho- and glyco modifications, is important for protein degradation in proteasomes.

3.2.1 Phosphorylated Proteins

Main tools in large-scale identification of phosphoproteins are enrichment by immobilized metal affinity chromatography (IMAC), chemical tagging, and immunoprecipitation by phosphor-specific antibodies. IMAC technology is based on methods developed by Andersson et al. and relies on interaction of phosphate group with immobilized Fe^{3+} ions. Ficarro et al. used IMAC combined with LC/MS/MS to characterize the phosphorylated portion of the yeast proteome. The authors showed that carboxylic acid interfered with IMAC purification, and needed to be protected. The protection was achieved by esterification with methanol in the presence of HCl. Phosphorylated tryptic peptides were identified via SEQUEST. From the whole cell lysate, Ficarro et al. detected more than 1000 phosphopeptides. From these, 383 sites of phosphorylation were determined. A potential improvement to this analysis would be the use of other proteinases in parallel with trypsin to increase the sequence coverage.

An enrichment technique that could also complement IMAC is immunoprecipitation with antibodies that bind to any protein that contain phosphorylated residues. While antibodies exist for phosphorylated serine, threonine, and tyrosine, only the anti-phosphotyrosine antibody binds strongly enough to allow enrichment. Pandey et al. used phosphotyrosine immunoprecipitation to study phosphorylation in HeLa cells in response to epidermal growth factor (EGF). The phosphopeptides were immunoprecipitated from untreated and EGF-treated cell lysates and resolved by electrophoresis. Individual gel bands were excised and studied by MALDI-MS and ESI-MS/MS. As a result, the authors identified Vav-2 as a substrate of EGF-receptor.

A report by Salomon et al. also gives a nice demonstration of analysis of phosphorylation in human cells. The authors used phosphotyrosine immunoprecipitation along with methyl esterification and IMAC combined with multidimensional LC/MS to assess tyrosine phosphorylation that occurs over time in myelogenous leukemia cells in response to treatment. The authors reported identification of 64 unique tyrosine phosphorylation sites in 32 proteins.

In another report by Ficarro et al., the authors used anti-phosphotyrosine immunoblots to study capacitation of human sperm. Capacitation is a cAMP-dependent process that is necessary for fertilization. The authors performed a comparative analysis of capacitated versus noncapacitated sperm. First, they separated sperm proteins by 2D PAGE followed by western blotting with anti-phosphotyrosine antibodies. In the next step, they excised and digested spots that exhibited phosphorylation, followed by IMAC to enrich for phosphopeptides with subsequent MS/MS analysis. As a result, the authors pinpointed several proteins that undergo phosphorylation upon capacitation

of the sperm. Additionally, the authors used differential isotopic labeling to quantify the phosphorylation. The labeling was achieved at the stage of protecting the carboxy groups prior to IMAC, by treatment of the peptide mixtures from capacitated and noncapacitated digests with CD_3OD/DCl and CH_3OH/HCl respectively. These two peptide mixtures were further combined in one-to-one ratio and analyzed by IMAC/LC/MS/MS. As a result of this quantitation, the authors found 20 unique peptides that exhibited different phosphorylation levels between capacitated and noncapacitated sperm.

Metabolic labeling strategy was first described to quantitate changes in phosphorylation. In this approach, cells from two batches that have potentially different levels of phosphorylated proteins are metabolically labeled with N^{14} and N^{15}. Next, the cells are lysed; the target proteins are purified, digested, and analyzed by MS. Changes in peak intensities that correspond to modified and unmodified peptides from the two conditions provide quantitation of phosphorylation.

Another way to quantitate phosphorylation levels is to use modified ICAT strategy. In this approach, the phosphate groups in phosphopeptides derived from two different conditions are chemically replaced with either labeled or unlabeled tags. The tagging involves the following steps: (1) beta-elimination of the phosphate groups; (2) addition of 1,2-ethanedithiol containing either four hydrogens (EDT-D_0) or four deuteriums (EDT-D_4); (3) biotinylation of the EDT group using (+)-biotinyl-iodoacetamidyl-3,6-dioxaoctanediamine. The tagged peptides are further affinity purified by avidin column and analyzed by LC/MS.

3.2.2 Glycosylated Proteins

The importance of protein glycosylation, especially during cell–cell communication in multicellular organisms, is often acknowledged by using the terms *glycobiology* and *glycomics*. Owing to a wide range of possible polysaccharide structures, analysis of glycosylation is not as straightforward as analysis of other posttranslational modification. Currently, there are no satisfactory methods for global, proteome-wide analysis of all glycoprotein forms. It is possible, however, to characterize glycoproteins with glycogroups of constant structure. It is also possible to globally map glycosylation sites.

As an example, consider a broad research question such as to identify and characterize glycosylated proteins in a given biological system. In such a case, hypothetical analysis could include the following steps: (1) proteolytic digestion; (2) enrichment for glycopeptides; (3) identification of glycopeptides by MS and MS/MS; (4) structure determination of polysaccharide groups by MS^n. Alternatively, one could also separate proteins by 2D PAGE, and use glycospecific staining methods to identify spots of interest. Also, during the MS part of analyses, it could be useful to separate constant glyco structures from the variable ones, as well as N-linked from O-linked ones. One of the problems in analysis of glycopeptides is that glycogroups are very labile. Because of this lability, peptide fragmentation in CID reactions is reduced, thus making sequencing by MS/MS more difficult. An alternative way is to chemically (e.g. with beta-elimination of O-linked oligosaccharides) or enzymatically (e.g. with *N*-glycosidase F) remove glycogroups prior to analysis and to use chemical tags.

In eukaryotes, the most widespread constant type of glycosylation is O-linked *N*-acetylglucosamine (*O*-GlcNAc), which is found on many nuclear and cytoplasmic proteins. Glycosylation of serine and threonine residues by *O*-GlcNAc is believed to compete with and complement phosphorylation in mediating protein–protein interactions. The proteins that are glycosylated by *O*-GlcNAc include RNA polymerase II, transcription factors, chromatin-associated proteins, nuclear pore proteins, protooncogenes, tumor suppressors, and proteins involved in translation.

Because *O*-GlcNAc and phospho groups modify essentially the same amino acids, it is of special interest to establish methods that can characterize these modifications simultaneously. In this pursuit, Wells et al. developed a method based on beta-elimination followed by Michael addition (BEMAD) of dithiothreitol (DTT) or biotine pentylamine (BAP). This method relies on the fact that *O*-GlcNAc groups are more prone to elimination than phosphate groups. With the right conditions of elimination, *O*-GlcNAc can be tagged selectively. The DTT and BAP tags also allow enrichment by affinity chromatography and are stable in MS/MS fragmentation, thus allowing identification of the modified sites. First, the authors tested their method on synthetic peptides, and then they performed analysis of several biological samples: Synapsin I from rat brain and nuclear pore complex (NPC). In Synapsin I, three novel *O*-GlcNAc sites, as well as three previously known sites were mapped, thus validating the method. In the nuclear pore complex, BEMAD also mapped novel *O*-GlcNAc sites in Lamin B receptor and Nup155. Using BEMAD along with modification-specific antibodies and enzymes, the authors were able to distinguish between *O*-GlcNAc- and phosphopeptides.

During CID reactions, oligosaccharide moieties fragment mainly at glycosidic bonds. This fact potentially allows discerning of a primary structure of an oligosaccharide in MS^n experiments. While technology for a large-scale structural analysis of glycoforms is not developed yet, oligosaccharide structure determination is certainly possible for individually isolated glycopeptides. Notable examples include characterization of lipooligosaccharides from *Haemophilus influenzae* and *Neisseria gonorrhoeae*.

3.2.3 Ubiquitinated Proteins

Degradation of proteins in living organisms is a complex, highly regulated process, which plays important roles in many cellular pathways. The first step in protein degradation is the attachment of ubiquitin moieties to lysines of the substrate. The second step is proteolysis of the tagged protein by proteasome. Given the importance of this posttranslational modification, it is surprising that the report by Peng et al. is perhaps the only paper in the literature that addresses large-scale analysis of protein ubiquitination.

Peng et al. described a systematic approach based on LC/LC/MS/MS to analyze protein ubiquitination in yeast. In this pursuit, Peng et al. expressed His-tagged ubiquitin in *S. cerevisiae cells* followed by purification over Ni-chelating resin. Denaturing conditions were used at the enrichment step in order to minimize copurification of proteins that are not ubiquitinated, but form complexes with ubiquitin. The enriched ubiquitinated proteins were digested with trypsin and analyzed by SCX/RP/MS/MS followed by identification by SEQUEST. The tryptic

digestion of ubiquitinated peptides results in glycine–glycine fragments at the sites of modification. The corresponding increase in mass by 114 Da of the modified lysine residues allows localization of the sites of ubiquitination. As a result, the authors identified 110 ubiquitination sites present in 72 ubiquitinated proteins. However, some of the known ubiquitinations were not detected. This is probably due to the fact that the method is biased toward more abundant species. Indeed, precise localization of a posttranslational modification via SEQUEST requires a high sequence coverage. As a consequence, in their analysis, Peng et al. identified 1075 proteins totally after Ni-resin purification, but were able to confirm ubiquitination in only 10% of them.

3.3 Analysis of Protein–Protein Interactions by Mass Spectrometry

In a cell, proteins exert their functions through interactions with other proteins. Proteomics methods developed in the past few years can be applied directly to the analysis of the protein–protein interactions and protein complexes. Protein–protein interactions can be studied either on the level of individual protein complexes, or on the level of the whole proteome. Examples of different types of analyses include the possibility of predicting interactions from amino acid sequence, focused mass spectrometric analyses of individual multiprotein complexes: yeast ribosome, SAGA-like complex (SLIK), Pol II preinitiation complex (PIC), nuclear pore complex (NPC), and proteome-wide analyses of protein–protein interactions. In addition, quantitative proteomic methods may be used to analyze the dynamics of protein complexes.

3.3.1 Computational Methods of Protein–Protein Interaction Prediction

To some degree, protein–protein interactions can be deduced indirectly from the amino acid sequences. In this pursuit, Bock et al. created a learning algorithm that was trained to recognize and predict protein–protein interactions. The training was achieved on the experimentally known interactions from a variety of organisms. As an outcome of the training, the decision function was constructed, which was statistically evaluated using unseen test data. As a result, on average, about 80% of the interactions were predicted accurately from the unseen datasets. Obviously, computational methods alone do not give exhaustive description and the obtained results need to be validated experimentally. Nevertheless, the interaction datasets obtained *in silico* provide useful reference points for experimental types of analyses.

3.3.2 Yeast 2-hybrid Arrays

One of the most common approaches to analyze protein–protein interactions is the yeast-2-hybrid (Y2H) screen, which is a selection method that detects protein–protein interactions in the yeast nucleus. This is a well-developed technique that can be easily optimized for a high-throughput analysis. The Y2H screen was successfully applied to mapping of a large-scale protein interaction network in the yeast *S. cerevisiae*. The Y2H has a good resolution and can detect weak and transient interactions. However, the Y2H method detects only binary interactions between proteins and may not be used to study transcriptional activators. Another drawback of the Y2H is that interactions occur in the nucleus, so that interactions for many proteins take place out of their native environment.

3.3.3 Direct Analysis of Large Protein Complexes; Composition of the Yeast Ribosome

Link et al. employed multidimensional liquid chromatography coupled with MS/MS to study composition of the yeast ribosome. The authors named their method *direct analysis of large protein complexes*, and demonstrated its wide applicability. In this study, the ribosomes were purified by sucrose gradients and then denatured. Next, the ribosomal RNA was removed, the ribosomes were enzymatically digested, and the digests were loaded on a 2D separation column, consisting of the SCX and the RP dimensions. After the separation, peptides were eluted directly into the mass spectrometer for the MS/MS analysis. Finally, SEQUEST algorithm was used to search nucleotide databases and to match the peptide fragmentation patterns. The authors demonstrated the high-throughput capacity of this approach – more than 100 proteins could be identified in a single experiment. As a result of this study, new protein components of yeast and human 40 S subunit were discovered.

3.3.4 Analysis of Multisubunit Protein Complexes Involved in Ubiquitin-dependent Proteolysis by Mass Spectrometry

In a number of reports, Deshaies et al. (and references therein) used MS-based strategies to characterize the composition of various protein complexes involved in proteolysis. The authors used sequential epitope tagging, affinity purification, and mass spectrometry (SEAM) to study regulation and function of SCF ubiquitin ligases. In the application of this method, SCF subunits Skp1 and Cul1 were C-terminally tagged with Myc epitope. Next, the cells expressing the tagged proteins were lysed, and the soluble fractions were affinity purified, digested, and analyzed by mass spectrometry. As a result, a total of 16 Skp1- and Cul1-interacting proteins were detected. Several of these proteins were not previously known, including Hrt1, Rav1, and Rav2. These new proteins were further subjected to SEAM and this led to identification of the new complex Rav1/Rav2/Skp1. Subsequently, it was found that Rav1/Rav2/Skp1 complex interacts with V_1 component of V-ATPase, the vacuolar membrane ATPase. In different studies by biochemical methods, the authors further determined that Rav1/Rav2/Skp1 regulates the assembly of V-ATPase from V_1 and V_0 domains.

The same group used multidimensional protein identification technology (MudPIT) to identify proteins that interact with the 26 S proteasome. As it was introduced in the previous sections of this chapter, MudPIT allows analysis of immunoprecipitated fractions without a gel separation step. Instead, the immunoprecipitated fractions are digested, and the peptide mixtures are separated in two dimensions (SCX and RP) followed by MS analysis. Using MudPIT, the authors identified every known subunit within affinity-purified 26 S proteasome, as well as one subunit that was not previously known. Additionally, a set of proteins potentially interacting with the proteasome (PIPs) was found. By immunoblotting methods, six of these PIPs were further confirmed to associate with the proteasome.

3.3.5 Proteomics of the Nuclear Pore Complex

MS-based protein identification is a useful tool that can efficiently determine the composition of a given multiprotein complex. However, this method by itself does not necessarily provide information on the complex's spatial architecture. Nor does it directly answer questions about the

multiprotein complex function. Therefore, when such questions arise, the mass spectrometric tools need to be complemented with other techniques, for example, immunoblotting, immunofluorescence, electron microscopy and so on. Several studies of the nuclear pore complexes (NPCs) that are discussed here give a good illustration of these types of integrative strategies.

In the early 1990s, studies by three-dimensional cryoelectron microscopy revealed the basic shape and architecture of NPCs in *Xenopus* nucleus. It was determined that NPCs are proteinaceous structures situated in the double membrane of the nuclear envelope. Estimated sizes of NPCs vary from ~125 MDa (*Xenopus*) to 66 MDa (*Saccharomyces*). NPCs have an eightfold rotational symmetry with the rotational axis normal to the nuclear envelope membrane and a twofold mirror symmetry with the symmetry plane parallel to the nuclear envelope membrane. Current research efforts are aimed at understanding the mechanism of the biological function of the NPCs – nucleotransport. To understand the NPC's function, it is useful to catalog all the protein components of NPCs in different organisms. Ideally, this should lead to testable hypotheses on how these components contribute to the overall structure and function of NPCs. Biological problems of this kind can be addressed by proteomics methods as was elegantly demonstrated for yeast and mammalian NPCs.

In the yeast study, Rout et al. prepared highly enriched fraction of NPC proteins, followed by separation on ceramic hydroxyapatite HPLC, which gave efficient recovery of the loaded proteins. Reverse-phase TFA-HPLC was used in parallel for resolution of the low molecular mass proteins from the NPC fraction. The next step in the separation involved SDS-PAGE with visualization of protein bands by copper staining. Subsequently, peptide mixtures were prepared from bands of interest via in-gel trypsin digestion followed by analysis by MALDI-MS. The peptide mass matching method was used to search nonredundant protein sequence database. Previously known nucleoporins were genomically tagged with a protein A epitope (ProtA), which allowed further immunolocalization by fluorescence microscopy. As a result of this study, the authors identified 29 nucleoporins and 11 transport factors and NPC-associated proteins. The authors also determined stoichiometry and position of each of the nucleoporins found within the NPC by quantitative immunoblotting and by immunoelectron microscopy. In the immunoelectron microscopy analysis, the ProtA-tagged proteins were labeled using gold-conjugated antibody, which aided visualization. On the basis of the deciphered architecture of the NPC, Rout et al. proposed a model of nucleotransport called a *Brownian affinity gating model*. The core idea of the proposed model is that translocation through the NPC occurs via diffusion: diffusive movements of the filamentous nucleoporins on the cytoplasmic face of the NPC exclude macromolecules that do not bind to them, but when the binding does occur (and that happens when a cargo molecule is associated with its transport factor), the residence time of the cargo at the NPC gate increases, which in turn facilitates the diffusion of the cargo into the nucleus.

In the mammalian study, Cronshaw et al. enriched NPCs fractions from rat liver nuclei by sequential solubilization. At each step of the enrichment, the authors used electron microscopy, SDS-PAGE, and immunoblotting to confirm that NPCs remained intact. After the enrichment, NPCs were treated with detergent, which

produced a solution of monomeric nucleoporins. The individual proteins in this nucleoporin mixture were separated by C4 reverse-phase chromatography followed by SDS-PAGE. Protein identification was performed using single-MS and MS/MS. In addition to previously known 23 nucleoporins, Cronshaw et al. identified six novel nucleoporins, and also four proteins containing WD repeats. One of these four WD-containing proteins was ALADIN, the gene mutated in Allgrove syndrome.

Spatial organization of a protein complex can be assessed by the cross-linking method. In this method, a protein complex is affinity-purified and then treated with a cross-linking agent, a chemical that introduces new covalent bonds between neighbor proteins. New bands, that appear on SDS-PAGE because of this treatment can be excised and identified by mass spectrometry. As a result, if a pair of the proteins gets cross-linked, this usually means that the two proteins are located close to each other within the complex. A good example of this strategy is the study by Rappsilber et al. in which the authors used cross-linking/MS method to deduce the spatial composition of the six-member subcomplex Nup84p of the yeast NPC. The authors emphasized generic applicability of this approach. One of the significant challenges in application of this method, however, is the choice of the proper cross-linking reaction condition – usually a number of different cross-linkers has to be screened, before the right degree of cross-linking is obtained. Additionally, interaction between subunits that are hidden deep inside a complex may be inaccessible to cross-linkers. Because of these and other difficulties, the cross-linking method is limited to complexes of small sizes and is difficult to apply on the broader scale. Nevertheless, the cross-linking method may prove to be useful in the studies of conformational or compositional changes of individual protein complexes, when the specific structural states can be "frozen" through interaction with the cross-linking agents.

3.3.6 High-throughput Analyses of Protein–Protein Interactions

In most cases, MS-based analytical schemes similar to those employed in the studies of individual complexes can be redesigned for use at the proteome-wide scale. A report by Gavin et al. is one of the first examples of the MS-based proteome-wide analyses of the protein complexes. For the large-scale isolation of the protein complexes from yeast Gavin et al. used tandem affinity purification (TAP), as first introduced by Rigaut et al. In the TAP method, a gene-specific fusion cassette – which contains calmodulin-binding domain, a specific protease cleavage site, and ProtA domain – is introduced at the C- or N-terminal of yeast's ORFs of interest. Then, assuming that expression of fusion proteins is maintained close to the natural level, the first affinity purification is performed. In this step, fusion proteins along with their interaction partners (so-called protein assemblies) are isolated from cell extract by affinity selection on IgG matrix. Next, the bound proteins are released by addition of the protease. Finally, the second affinity purification is done, which involves incubation with calmodulin beads in the presence of calcium. The advantage of tandem affinity purification when compared to standard epitope tagging approaches is that it removes most of the nonspecific interactions. Out of 1548 yeast strains generated by Gavin et al., 1167 expressed the fusion proteins at detectable levels. After the purification of the protein complexes by TAP, the authors

subjected the complexes to electrophoretic separation followed by trypsin digestion, and subsequent analysis by MALDI-MS. Overall, by MS analysis, Gavin et al. identified 1440 gene products (~25% of the genome) from various organelles. However, identification of membrane proteins in this study has proven to be difficult – only 40 membrane proteins were purified successfully out of total 293 membrane proteins detected. The authors then proceeded with grouping of the identified proteins into complexes. This was done by the analysis of overlaps in composition of the pulled-down assemblies from 589 different bait proteins. The authors reported a total of 245 purifications that corresponded to 98 known complexes from the yeast protein database (YPD). Another 242 purifications out of the 589 were assembled into 134 new complexes. The authors were able to identify proteins as low-abundant as 15 copies per cell, thus showing high sensitivity of the TAP method. However, reproducibility was rather poor – the authors estimated that probability of finding the same protein from the same bait in two purifications is about 70%. Another weakness of the TAP method comes from possibility of interference of the TAP tag with complexes assembly and protein function. In fact, Gavin et al. found that when the essential genes were TAP-tagged; in about 20% of these cases, nonviable strains were obtained. Also, the authors reported significant bias against proteins with molecular weight below 15 KDa.

Another notable report of a high-throughput protein complex identification is the study by Ho et al. In this case, the bait proteins contained Flag epitope tag and were overexpressed from *GAL1* or *tet* promoters. Next, the protein assemblies were isolated in one-step immunoaffinity purification followed by resolution on SDS-PAGE, digestion and MS, and MS/MS analyses. The authors called this method "high-throughput mass spectrometric protein complex identification" (HMS-PCI). The immunoaffinity purification of complexes assembled around overexpressed baits should, in theory, generate more false-positives than TAP method would generate, because of the nonphysiological concentrations of the baits. On the other hand, weak and transient interactions that would not be detected in the TAP method could be captured by HMS-PCI. In fact, Ho et al. were able to assess certain regulatory and signaling pathways, by studying complexes pulled-down with phosphatases and kinases used as baits.

As of today none of the methods of mapping of protein–protein interactions within a proteome is comprehensive enough to provide full coverage. Hence, it is useful to compare datasets obtained with different approaches. In their article, Christian von Mering et al. evaluated all available interaction datasets obtained in yeast. These included the data from MS-based studies discussed above, as well as data from Y2H, correlated mRNA expression, synthetic lethality, and *in silico* predictions. The evaluation was done through comparisons with the reference dataset (MIPS and YPD). As a result, percentage coverage (fraction of the reference covered) and accuracy (fraction of data confirmed by the reference) were estimated for every method. According to the authors' analysis, TAP method provides both higher coverage and higher accuracy than either Y2H or HMS-PCI. Also, the analysis shows that HMS-PCI is the least accurate method amongst the three.

In a series of reports, Bader and Hogue developed algorithmic approaches for finding molecular complexes from datasets

obtained in different interaction studies. By analyzing combined data from TAP and HMS-PCI studies, they found a novel nucleolar complex of 148 proteins that included 39 proteins with unknown function. Further, they described a graph-theoretic clustering algorithm *molecular complex detection* (MCODE) that allows detection of dense regions (potential complexes) within the interaction networks. Importantly, the authors showed that MCODE algorithm is not affected by a high rate of false-positives in datasets from the high-throughput experiments.

To summarize, none of the current high-throughput experimental schemes provide sufficient coverage and accuracy. Therefore, integrative approaches that take advantage of different methods are necessary. Additionally, all of the discussed cases dealt with cells in a certain growth conditions. It is of special interest, however, to study dynamics within protein interaction networks in response to environmental stimuli, in progression through the cell cycle, or in pathological states. Some of these questions can be addressed by quantitative proteomics techniques.

3.3.7 Quantitative Proteomics Methods in the Studies of Protein Complexes

Methods of quantitative proteomics can be used to study dynamical changes in abundance, composition, and activities of multiprotein complexes. A good example of such a study is the work by Ranish et al. in which composition of a large RNA polymerase preinitiation complex (PIC) was assessed by the isotopically coded affinity tags (ICATs) method. This and other differential labeling methods are discussed in detail in the previous section of this chapter. The ICAT approach as employed by Ranish et al., consists of the four major steps summarized below:

1. The samples from two different conditions are labeled with either "light" or "heavy" tags.
2. The labeled samples are combined together and enzymatically digested.
3. The mixture is fractionated by SCX chromatography, then the labeled peptides are isolated by avidin-affinity chromatography followed by separation by the reversed-phase microcapillary chromatography.
4. The labeled peptides that are eluting from the reversed phase column are analyzed by ESI-MS/MS.

In MS analysis, the peak intensity ratios of the differentially labeled peptides on the ion chromatogram are related to the relative abundances of the corresponding proteins in the two different environments. The peptide identities are established in the MS/MS spectra and the corresponding proteins are identified by SEQUEST. With this quantitative ICAT method, the authors were able to distinguish between components of the PIC and the copurifying background proteins, some of which had higher abundances. Thus, the authors demonstrated the high analytical power of this approach. In their analysis, Ranish et al. identified a total of 326 proteins, 42% of which participate in the Pol II-mediated transcription. Also, the authors used the ICAT method to monitor changes in the PIC composition in the presence or absence of TBP. TBP is a transcription factor that binds to the TATA element and is required for the functional PIC assembly. According to Ranish et al., most of the Pol II components are increased in abundance by a factor of at least 1.9 upon addition of TBP, and several Pol II components showed no increase in abundance. In addition, potentially new component of the PIC was discovered. A

limitation of this approach, as was noted by the authors, is that using only the cysteine specific tags leaves out tryptic peptides that do not contain cysteines. In this respect, strategies that use metabolic labeling or N-terminal labeling may be more promising.

As an example, in their recent paper, Blagoev et al. used stable isotopic amino acids in cell culture (SILAC) to study EGF signaling. The control and EGF-stimulated HeLa cell populations were labeled with ^{12}C- arginine and ^{13}C-arginine respectively via metabolic incorporation. Combined cell lysates from these two conditions were affinity-purified with SH2 domain of GST-SH2 fusion protein used as bait. SH2 domain specifically binds phosphorylated EGF receptor. Protein complexes obtained in this purification were digested with trypsin and the peptide mixtures were analyzed by MS. As a result, the authors identified 228 proteins, 28 of which were enriched upon EGF simulation.

Bibliography

Books and Reviews

Aebersold, R., Mann, M. (2003) Mass spectrometry-based proteomics, *Nature* **422**, 198–207.

Bakhtiar, R., Nelson, R.W. (2000) Electrospray ionization and matrix-assisted laser desorption ionization mass spectrometry. Emerging technologies in biomedical sciences, *Biochem. Pharmacol.* **59**, 891–905.

Chalmers, M.J., Gaskell, S.J. (2000) Advances in mass spectrometry for proteome analysis, *Curr. Opin. Biotechnol.* **11**, 384–390.

Choudhary, J.S., Blackstock, W.P., Creasy, D.M., Cottrell, J.S. (2001) Matching peptide mass spectra to EST and genomic DNA databases, *Trends Biotechnol.* **19**, S17–S22.

Figeys, D. (2003) Proteomics in 2002: a year of technical development and wide-ranging applications, *Anal. Chem.* **75**, 2891–2905.

Koonin, E.V. (2001) Computational genomics, *Curr. Biol.* **11**, R155–R158.

McLachlin, D.T., Chait, B.T. (2001) Analysis of phosphorylated proteins and peptides by mass spectrometry, *Curr. Opin. Chem. Biol.* **5**, 591–602.

Noordewier, M.O., Warren, P.V. (2001) Gene expression microarrays and the integration of biological knowledge, *Trends Biotechnol.* **19**, 412–415.

Siuzdak, G. (1994) The emergence of mass spectrometry in biochemical research, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11290–11297.

Watson, J.T. (1997) *Introduction to Mass Spectrometry*, Lippincott-Raven, Philadelphia, NY.

Wilkins, M.R., Appel, K.D., Hochstrasser, D.F. (Eds.) (1997) *Proteome Research: New Frontiers in Functional Genomics (Principles and Practice)*, Springer-Verlag, New York.

Primary Literature

Adam, S.A. (2001) The nuclear pore complex, *Genome Biol.* **2**, REVIEWS0007.1–0007.6.

Aebersold, R. (2003) Quantitative proteome analysis: methods and applications, *J. Infect. Dis.* **187**(()Suppl. 2), S315–S320.

Aebersold, R., Mann, M. (2003) Mass spectrometry-based proteomics, *Nature* **422**, 198–207.

Akey, C., Radermacher, M. (1993) Architecture of the Xenopus nuclear pore complex revealed by three-dimensional cryo-electron microscopy, *J. Cell. Biol.* **122**, 1–19.

Andersson, L., Porath, J. (1986) Isolation of phosphoproteins by immobilized metal (Fe^{3+}) affinity chromatography, *Anal. Biochem.* **154**, 250–254.

Ardekani, A.M., Liotta, L.A., Petricoin, E.F. III (2002) Clinical potential of proteomics in the diagnosis of ovarian cancer, *Expert Rev. Mol. Diagn.* **2**, 312–320.

Bacher, G., Korner, R., Atrih, A., Foster, S.J., Roepstorff, P., Allmaier, G. (2001) Negative and positive ion matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and positive ion nano-electrospray ionization quadrupole ion trap mass spectrometry of peptidoglycan fragments isolated from various Bacillus species, *J. Mass Spectrom.* **36**, 124–139.

Bader, G.D., Hogue, C.W. (2002) Analyzing yeast protein-protein interaction data obtained from different sources, *Nat. Biotechnol.* **20**, 991–997.

Bader, G.D., Hogue, C.W. (2003) An automated method for finding molecular complexes

in large protein interaction networks, *BMC Bioinform.* **4**, 2.

Bakhtiar, R., Nelson, R.W. (2000) Electrospray ionization and matrix-assisted laser desorption ionization mass spectrometry. Emerging technologies in biomedical sciences, *Biochem. Pharmacol.* **59**, 891–905.

Berger, S.J., Lee, S.W., Anderson, G.A., Pasa-Tolic, L., Tolic, N., Shen, Y., Zhao, R., Smith, R.D. (2002) High-throughput global peptide proteomic analysis by combining stable isotope amino acid labeling and data-dependent multiplexed-MS/MS, *Anal. Chem.* **74**, 4994–5000.

Bock, J.R., Gough, D.A. (2001) Predicting protein–protein interactions from primary structure, *Bioinformatics* **17**, 455–460.

Bock, J.R., Gough, D.A. (2003) Whole-proteome interaction mining, *Bioinformatics* **19**, 125–134.

Cagney, G., Emili, A. (2002) De novo peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging, *Nat. Biotechnol.* **20**, 163–170.

Cagney, G., Uetz, P., Fields, S. (2000) High-throughput screening for protein-protein interactions using two-hybrid assay, *Methods Enzymol.* **328**, 3–14.

Carr, S.A., Huddleston, M.J., Bean, M.F. (1993) Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry, *Protein Sci.* **2**, 183–196.

Chalmers, M.J., Gaskell, S.J. (2000) Advances in mass spectrometry for proteome analysis, *Curr. Opin. Biotechnol.* **11**, 384–390.

Choudhary, J.S., Blackstock, W.P., Creasy, D.M., Cottrell, J.S. (2001) Matching peptide mass spectra to EST and genomic DNA databases, *Trends Biotechnol.* **19**, S17–S22.

Comer, F.I., Hart, G.W. (1999) O-GlcNAc and the control of gene expression, *Biochim. Biophys. Acta* **1473**, 161–171.

Comisarow, M.B., Marshall, A.G. (1996) The early development of Fourier transform ion cyclotron resonance (FT-ICR) spectroscopy, *J. Mass Spectrom.* **31**, 581–585.

Conrads, T.P., Alving, K., Veenstra, T.D., Belov, M.E., Anderson, G.A., Anderson, D.J., Lipton, M.S., Pasa-Tolic, L., Udseth, H.R., Chrisler, W.B., Thrall, B.D., Smith, R.D. (2001) Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and 15N-metabolic labeling, *Anal. Chem.* **73**, 2132–2139.

Conrads, T.P., Anderson, G.A., Veenstra, T.D., Pasa-Tolic, L., Smith, R.D. (2000) Utility of accurate mass tags for proteome-wide protein identification, *Anal. Chem.* **72**, 3349–3354.

Conrads, T.P., Issaq, H.J., Veenstra, T.D. (2002) New tools for quantitative phosphoproteome analysis, *Biochem. Biophys. Res. Commun.* **290**, 885–890.

Corthals, G.L., Wasinger, V.C., Hochstrasser, D.F., Sanchez, J.C. (2000) The dynamic range of protein expression: a challenge for proteomic research, *Electrophoresis* **21**, 1104–1115.

Cronshaw, J.M., Krutchinsky, A.N., Zhang, W., Chait, B.T., Matunis, M.J. (2002) Proteomic analysis of the mammalian nuclear pore complex, *J. Cell. Biol.* **158**, 915–927.

Deshaies, R.J., Seol, J.H., McDonald, W.H., Cope, G., Lyapina, S., Shevchenko, A., Verma, R., Yates, J.R. III (2002) Charting the protein complexome in yeast by mass spectrometry, *Mol. Cell. Proteomics* **1**, 3–10.

Dohmen, R.J., Madura, K., Bartel, B., Varshavsky, A. (1991) The N-end rule is mediated by the UBC2(RAD6) ubiquitin-conjugating enzyme, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7351–7355.

Dotsch, V., Wagner, G. (1998) New approaches to structure determination by NMR spectroscopy, *Curr. Opin. Struct. Biol.* **8**, 619–623.

Dreger, M. (2003) Subcellular proteomics, *Mass Spectrom. Rev.* **22**, 27–56.

Eng, J., McCormack, A., Yates, J.R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequence in a protein database, *J. Am. Soc. Mass Spectrom.* **5**, 976–989.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M. (1989) Electrospray ionization for mass spectrometry of large biomolecules, *Science* **246**, 64–71.

Fey, S.J., Larsen, P.M. (2001) 2D or not 2D. Two-dimensional gel electrophoresis, *Curr. Opin. Chem. Biol.* **5**, 26–33.

Ficarro, S., Chertihin, O., Westbrook, V.A., White, F., Jayes, F., Kalab, P., Marto, J.A., Shabanowitz, J., Herr, J.C., Hunt, D.F., Visconti, P.E. (2003) Phosphoproteome analysis of capacitated human sperm. Evidence of tyrosine phosphorylation of a kinase-anchoring protein 3 and valosin-containing

protein/p97 during capacitation, *J. Biol. Chem.* **278**, 11579–11589.

Fields, S., Song, O. (1989) A novel genetic system to detect protein-protein interactions, *Nature* **340**, 245–246.

Figeys, D. (2003) Proteomics in 2002: a year of technical development and wide-ranging applications, *Anal. Chem.* **75**, 2891–2905.

Florens, L., Washburn, M.P., Raine, J.D., Anthony, R.M., Grainger, M., Haynes, J.D., Moch, J.K., Muster, N., Sacci, J.B., Tabb, D.L., Witney, A.A., Wolters, D., Wu, Y., Gardner, M.J., Holder, A.A., Sinden, R.E., Yates, J.R., Carucci, D.J. (2002) A proteomic view of the plasmodium falciparum life cycle, *Nature* **419**, 520–526.

Fountoulakis, M., Takacs, M.F., Berndt, P., Langen, H., Takacs, B. (1999) Enrichment of low abundance proteins of escherichia coli by hydroxyapatite chromatography, *Electrophoresis* **20**, 2181–2195.

Fountoulakis, M., Takacs, M.F., Takacs, B. (1999) Enrichment of low-copy-number gene products by hydrophobic interaction chromatography, *J. Chromatogr., A* **833**, 157–168.

Friedman, D.B., Hill, S., Keller, J.W., Merchant, N.B., Levy, S.E., Coffey, R.J., Caprioli, R.M. (2004) Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry, *Proteomics* **4**, 793–811

Futcher, B., Latter, G.I., Monardo, P., McLaughlin, C.S., Garrels, J.I. (1999) A sampling of the yeast proteome, *Mol. Cell. Biol.* **19**, 7357–7368.

Gaucher, S.P., Cancilla, M.T., Phillips, N.J., Gibson, B.W., Leary, J.A. (2000) Mass spectral characterization of lipooligosaccharides from haemophilus influenzae 2019, *Biochemistry* **39**, 12406–12414.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., Weissman, J.S. (2003) Global analysis of protein expression in yeast, *Nature* **425**, 737–741.

Glickman, M.H., Ciechanover, A. (2002) The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction, *Physiol. Rev.* **82**, 373–428.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.L., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., Briggs, S. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica), *Science* **296**, 92–100.

Goodlett, D.R., Keller, A., Watts, J.D., Newitt, R., Yi, E.C., Purvine, S., Eng, J.K., von Haller, P., Aebersold, R., Kolker, E. (2001) Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation, *Rapid. Commun. Mass Spectrom.* **15**, 1214–1221.

Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., Weiss, W. (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients, *Electrophoresis* **21**, 1037–1053.

Goshe, M.B., Conrads, T.P., Panisko, E.A., Angell, N.H., Veenstra, T.D., Smith, R.D. (2001) Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses, *Anal. Chem.* **73**, 2578–2586.

Goshe, M.B., Veenstra, T.D., Panisko, E.A., Conrads, T.P., Angell, N.H., Smith, R.D. (2002) Phosphoprotein isotope-coded affinity tags: application to the enrichment and identification of low-abundance phosphoproteins, *Anal. Chem.* **74**, 607–616.

Goto, N.K., Kay, L.E. (2000) New developments in isotope labeling strategies for protein solution NMR spectroscopy, *Curr. Opin. Struct. Biol.* **10**, 585–592.

Greenbaum, D., Colangelo, C., Williams, K., Gerstein, M. (2003) Comparing protein abundance and mRNA expression levels on a genomic scale, *Genome Biol.* **4**, 117.

Greenbaum, D., Jansen, R., Gerstein, M. (2002) Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts, *Bioinformatics* **18**, 585–596.

Greis, K.D., Hayes, B.K., Comer, F.I., Kirk, M., Barnes, S., Lowary, T.L., Hart, G.W. (1996) Selective detection and site-analysis of

O-GlcNAc-modified glycopeptides by beta-elimination and tandem electrospray mass spectrometry, *Anal. Biochem.* **234**, 38–49.

Griffin, T.J., Gygi, S.P., Ideker, T., Rist, B., Eng, J., Hood, L., Aebersold, R. (2002) Complementary profiling of gene expression at the transcriptome and proteome levels in saccharomyces cerevisiae, *Mol. Cell. Proteomics* **1**, 323–333.

Gronborg, M., Kristiansen, T.Z., Stensballe, A., Andersen, J.S., Ohara, O., Mann, M., Jensen, O.N., Pandey, A. (2002) A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, frigg, as a protein kinase a substrate, *Mol. Cell. Proteomics* **1**, 517–527.

Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y., Aebersold, R. (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9390–9395.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nat. Biotechnol.* **17**, 994–999.

Gygi, S.P., Rochon, Y., Franza, B.R., Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast, *Mol. Cell. Biol.* **19**, 1720–1730.

Ha, K.S. (2001) Principle and biological applications of protein chip system, *Exp. Mol. Med.* **33**, 127–132.

Haab, B.B., Dunham, M.J., Brown, P.O. (2001) Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions, *Genome Biol.* **2**, RESEARCH0004.1–0004.13.

Han, D.K., Eng, J., Zhou, H., Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry, *Nat. Biotechnol.* **19**, 946–951.

Hansen, A.P., Petros, A.M., Mazar, A.P., Pederson, T.M., Rueter, A., Fesik, S.W. (1992) A practical method for uniform isotopic labeling of recombinant proteins in mammalian cells, *Biochemistry* **31**, 12713–12718.

Hansen, K.C., Schmitt-Ulms, G., Chalkley, R.J., Hirsch, J., Baldwin, M.A., Burlingame, A.L. (2003) Mass spectrometric analysis of protein mixtures at low levels using cleavable 13C-isotope-coded affinity tag and multidimensional chromatography, *Mol. Cell. Proteomics* **2**, 299–314.

Hart, G.W., Haltiwanger, R.S., Holt, G.D., Kelly, W.G. (1989) Nucleoplasmic and cytoplasmic glycoproteins, *Ciba. Found. Symp.* **145**, 102–112, discussion 112–108.

Herbert, B. (1999) Advances in protein solubilization for two-dimensional electrophoresis, *Electrophoresis* **20**, 660–663.

Herbert, B., Righetti, P.G. (2000) A turning point in proteome analysis: sample prefractionation via multicompartment electrolyzers with isoelectric membranes, *Electrophoresis* **21**, 3639–3648.

Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R., Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network, *Science* **292**, 929–934.

Imam-Sghiouar, N., Laude-Lemaire, I., Labas, V., Pflieger, D., Le Caer, J.P., Caron, M., Nabias, D.K., Joubert-Caron, R. (2002) Subproteomics analysis of phosphorylated proteins: application to the study of B-lymphoblasts from a patient with Scott syndrome, *Proteomics* **2**, 828–838.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4569–4574.

Jansen, R., Bussemaker, H.J., Gerstein, M. (2003) Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models, *Nucleic Acids Res.* **31**, 2242–2251.

Jensen, P.K., Pasa-Tolic, L., Anderson, G.A., Horner, J.A., Lipton, M.S., Bruce, J.E., Smith, R.D. (1999) Probing proteomes using capillary isoelectric focusing-electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, *Anal. Chem.* **71**, 2076–2084.

Jensen, P.K., Pasa-Tolic, L., Peden, K.K., Martinovic, S., Lipton, M.S., Anderson, G.A., Tolic, N., Wong, K.K., Smith, R.D. (2000) Mass spectrometric detection for capillary isoelectric focusing separations of complex protein mixtures, *Electrophoresis* **21**, 1372–1380.

Ji, J., Chakraborty, A., Geng, M., Zhang, X., Amini, A., Bina, M., Regnier, F. (2000)

Strategy for qualitative and quantitative analysis in proteomics based on signature peptides, *J. Chromatogr., B. Biomed. Sci. Appl.* **745**, 197–210.

Jiang, H., English, A.M. (2002) Quantitative analysis of the yeast proteome by incorporation of isitopically labeled leucine, *J. Proteome Res.* **1**, 345–350.

Karas, M., Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons, *Anal. Chem.* **60**, 2299–2301.

Klose, J. (1999) Large-gel 2-D electrophoresis, *Methods Mol. Biol.* **112**, 147–172.

Klose, J., Kobalz, U. (1995) Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome, *Electrophoresis* **16**, 1034–1059.

Koller, A., Washburn, M.P., Lange, B.M., Andon, N.L., Deciu, C., Haynes, P.A., Hays, L., Schieltz, D., Ulaszek, R., Wei, J., Wolters, D., Yates, J.R., III (2002) Proteomic survey of metabolic pathways in rice, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11969–11974.

Koonin, E.V. (2001) Computational genomics, *Curr. Biol.* **11**, R155–R158.

Krishna, R., Wold, F. (1993) Post-Translational Modification of Proteins, *Adv. Enzymol. Relat. Areas Mol. Biol.* **67**, 265–298.

Lahm, H.W., Langen, H. (2000) Mass spectrometry: a tool for the identification of proteins separated by gels, *Electrophoresis* **21**, 2105–2114.

Leavell, M.D., Leary, J.A., Yamasaki, R. (2002) Mass spectrometric strategy for the characterization of lipooligosaccharides from neisseria gonorrhoeae 302 using FTICR, *J. Am. Soc. Mass Spectrom.* **13**, 571–576.

Lee, Y., Mrksich, M. (2002) Protein chips: from concept to practice, *A TRENDS Guide to Proteomics II.* **20**, s14–s18.

Li, J., Steen, H., Gygi, S.P. (2003) Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents: the yeast salinity stress response, *Mol. Cell. Proteomics* **2**, 1198–1204.

Link, A. (2002) Multidimensional peptide separations in proteomics, *A TRENDS Guide to Proteomics II.* **20**, s8–s13.

Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., Yates, J.R. III (1999) Direct analysis of protein complexes using mass spectrometry, *Nat. Biotechnol.* **17**, 676–682.

Lipton, M.S., Pasa-Tolic, L., Anderson, G.A., Anderson, D.J., Auberry, D.L., Battista, J.R., Daly, M.J., Fredrickson, J., Hixson, K.K., Kostandarithes, H., Masselon, C., Markillie, L.M., Moore, R.J., Romine, M.F., Shen, Y., Stritmatter, E., Tolic, N., Udseth, H.R., Venkateswaran, A., Wong, K.K., Zhao, R., Smith, R.D. (2002) Global analysis of the deinococcus radiodurans proteome by using accurate mass tags, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11049–11054.

Liska, A.J., Shevchenko, A. (2003) Expanding the organismal scope of proteomics: cross-species protein identification by mass spectrometry and its implications, *Proteomics* **3**, 19–28.

Liu, P., Regnier, F. (2002) An isotope coding strategy for proteomics involving both amine and carboxyl labeling, *J. Proteome Res.* **1**, 443–450.

Mamyrin, B.A., Karataev V.I., Shmikk D.V., Zagulin V.A. (1973) Reflectron for TOF-MS, *Sov. Phys. JETP* **37**, 45–48.

Mamyrin, B.A., Shmikk D.V. (1979) Linear reflection, *Sov. Phys. JETP* **49**, 762–764.

Mann, M., Wilm, M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags, *Anal. Chem.* **66**, 4390–4399.

McLachlin, D.T., Chait, B.T. (2001) Analysis of phosphorylated proteins and peptides by mass spectrometry, *Curr. Opin. Chem. Biol.* **5**, 591–602.

von Mering, C., Kruse, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., Bork, P. (2002) Comparative assessment of large scale data sets of rotein-protein interactions, *Nature* **417**, 399–403.

Mirgorodskaya, O.A., Kozmin, Y.P., Titov, M.I., Korner, R., Sonksen, C.P., Roepstorff, P. (2000) Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using 18O-labeled internal standards, *Rapid. Commun. Mass Spectrom.* **14**, 1226–1232.

Molloy, M.P. (2000) Two-dimensional electrophoresis of membrane proteins using immobilized pH gradients, *Anal. Biochem.* **280**, 1–10.

Moskovets, E. (2000) Optimization of the mass reflector parameters for direct ion extraction, *Rapid. Commun. Mass Spectrom.* **14**, 150–155.

Moskovets, E., Karger, B.L. (2003) Mass calibration of a matrix-assisted laser desorption/ionization time-of-flight mass

spectrometer including the rise time of the delayed extraction pulse, *Rapid. Commun. Mass Spectrom.* **17**, 229–237.

Mrksich, M., Dike, L.E., Tien, J., Ingber, D.E., Whitesides, G.M. (1997) Using microcontact printing to pattern the attachment of mammalian cells to self-assembled monolayers of alkanethiolates on transparent films of gold and silver, *Exp. Cell. Res.* **235**, 305–313.

Munchbach, M., Quadroni, M., Miotto, G., James, P. (2000) Quantitation and facilitated de novo sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation-directing moiety, *Anal. Chem.* **72**, 4047–4057.

Muszynska, G., Andersson, L., Porath, J. (1986) Selective adsorption of phosphoproteins on gel-immobilized ferric chelate, *Biochemistry* **25**, 6850–6853.

Noordewier, M.O., Warren, P.V. (2001) Gene expression microarrays and the integration of biological knowledge, *Trends Biotechnol.* **19**, 412–415.

Oda, Y., Huang, K., Cross, F.R., Cowburn, D., Chait, B.T. (1999) Accurate quantitation of protein expression and site-specific phosphorylation, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6591–6596.

O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins, *J. Biol. Chem.* **250**, 4007–4021.

Oh-Ishi, M., Satoh, M., Maeda, T. (2000) Preparative two-dimensional gel electrophoresis with agarose gels in the first dimension for high molecular mass proteins, *Electrophoresis* **21**, 1653–1669.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics, *Mol. Cell. Proteomics* **1**, 376–386.

Pappin, D.J. (2003) Peptide mass fingerprinting using MALDI-TOF mass spectrometry, *Methods Mol. Biol.* **211**, 211–219.

Pasa-Tolic, L., Harkewicz, R., Anderson, G.A., Tolic, N., Shen, Y., Zhao, R., Thrall, B., Masselon, C., Smith, R.D. (2002) Increased proteome coverage for quantitative peptide abundance measurements based upon high performance separations and DREAMS FTICR mass spectrometry, *J. Am. Soc. Mass Spectrom.* **13**, 954–963.

Peters, E.C., Horn, D.M., Tully, D.C., Brock, A. (2001) A novel multifunctional labeling reagent for enhanced protein characterization with mass spectrometry, *Rapid. Commun. Mass Spectrom.* **15**, 2387–2392.

Petricoin, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone, C., Fishman, D.A., Kohn, E.C., Liotta, L.A. (2002) Use of proteomic patterns in serum to identify ovarian cancer, *Lancet* **359**, 572–577.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., Seraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification, *Methods* **24**, 218–229.

Rademacher, T.W., Parekh, R.B., Dwek, R.A. (1988) Glycobiology, *Annu. Rev. Biochem.* **57**, 785–838.

Reynolds, K.J., Yao, X., Fenselau, C. (2002) Proteolytic 18O labeling for comparative proteomics: evaluation of endoproteinase Glu-C as the catalytic agent, *J. Proteome Res.* **1**, 27–33.

Rosenblatt, K.P., Bryant-Greenwood, P., Killian, J.K., Mehta, A., Geho, D., Espina, V., Petricoin, E.F., Liotta, L.A. (2004) Serum proteomics in cancer diagnosis and management, *Annu. Rev. Med.* **55**, 97–112.

Rout, M.P., Aitchison, J.D., Suprapto, A., Hjertaas, K., Zhao, Y., Chait, B.T. (2000) The yeast nuclear pore complex: composition, architecture, and transport mechanism, *J. Cell. Biol.* **148**, 635–652.

Rudiger, H., Gabius, H.J. (2001) Plant lectins: occurrence, biochemistry, functions and applications, *Glycoconj. J.* **18**, 589–613.

Sanchez, J.C., Hochstrasser, D.F. (1999) High-resolution, IPG-based, mini two-dimensional gel electrophoresis, *Methods Mol. Biol.* **112**, 227–233.

Santoni, V., Molloy, M., Rabilloud, T. (2000) Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21**, 1054–1070.

Schwartz, J.C., Jardine, I. (1996) Quadrupole ion trap mass spectrometry, *Methods Enzymol.* **270**, 552–586.

Schwartz, J.C., Senko, M.W., Syka, J.E. (2002) A two-dimensional quadrupole ion trap mass spectrometer, *J. Am. Soc. Mass Spectrom.* **13**, 659–669.

Sheeley, D.M., Reinhold, V.N. (1998) Structural characterization of carbohydrate sequence, linkage, and branching in a quadrupole Ion trap mass spectrometer: neutral oligosaccharides and N-linked glycans, *Anal. Chem.* **70**, 3053–3059.

Siuzdak, G. (1994) The emergence of mass spectrometry in biochemical research, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11290–11297.

Smith, R.D., Anderson, G.A., Lipton, M.S., Pasa-Tolic, L., Shen, Y., Conrads, T.P., Veenstra, T.D., Udseth, H.R. (2002) An accurate mass tag strategy for quantitative and high-throughput proteome measurements, *Proteomics* **2**, 513–523.

Smith, R.D., Loo, J.A., Edmonds, C.G., Barinaga, C.J., Udseth, H.R. (1990) New developments in biochemical mass spectrometry: electrospray ionization, *Anal. Chem.* **62**, 882–899.

Snow, D.M., Hart, G.W. (1998) Nuclear and cytoplasmic glycosylation, *Int. Rev. Cytol.* **181**, 43–74.

Sunyaev, S., Liska, A.J., Golod, A., Shevchenko, A. (2003) MultiTag: multiple error-tolerant sequence tag search for the sequence-similarity identification of proteins by mass spectrometry, *Anal. Chem.* **75**, 1307–1315.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., Rothberg, J.M. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae, *Nature* **403**, 623–627.

Unlu, M., Morgan, M.E., Minden, J.S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts, *Electrophoresis* **18**, 2071–2077.

Varshavsky, A. (1996) The N-end rule: functions, mysteries, uses, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12142–12149.

Verma, R., Chen, S., Feldman, R., Schieltz, D., Yates, J., Dohmen, J., Deshaies, R.J. (2000) Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes, *Mol. Biol. Cell.* **11**, 3425–3439.

Vosseller, K., Wells, L., Hart, G.W. (2001) Nucleocytoplasmic O-glycosylation: O-GlcNAc and functional proteomics, *Biochimie* **83**, 575–581.

Wall, D.B., Kachman, M.T., Gong, S., Hinderer, R., Parus, S., Misek, D.E., Hanash, S.M., Lubman, D.M. (2000) Isoelectric focusing nonporous RP HPLC: a two-dimensional liquid-phase separation method for mapping of cellular proteins with identification using MALDI-TOF mass spectrometry, *Anal. Chem.* **72**, 1099–1111.

Wall, D.B., Kachman, M.T., Gong, S.S., Parus, S.J., Long, M.W., Lubman, D.M. (2001) Isoelectric focusing nonporous silica reversed-phase high-performance liquid chromatography/electrospray ionization time-of-flight mass spectrometry: a three-dimensional liquid-phase protein separation method as applied to the human erythroleukemia cell-line, *Rapid. Commun. Mass Spectrom.* **15**, 1649–1661.

Washburn, M.P., Koller, A., Oshiro, G., Ulaszek, R.R., Plouffe, D., Deciu, C., Winzeler, E., Yates, J.R. III (2003) Protein pathway and complex clustering of correlated mRNA and protein expression analyses in Saccharomyces cerevisiae, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3107–3112.

Washburn, M.P., Ulaszek, R., Deciu, C., Schieltz, D.M., Yates, J.R. III (2002) Analysis of quantitative proteomic data generated via multidimensional protein identification technology, *Anal. Chem.* **74**, 1650–1657.

Washburn, M.P., Wolters, D., Yates, J.R. III (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology, *Nat. Biotechnol.* **19**, 242–247.

Washburn, P.W., Yates, J.R. III (2000) New methods of proteome analysis: multidimensional chromatography and mass spectrometry, *Trends Biotechnol.* **18**, S27–S30.

Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, R., Williams, K.L., Humphery-Smith, I. (1995) Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium, *Electrophoresis* **16**, 1090–1094.

Watson, J.T. (1997) *Introduction to Mass Spectrometry*, Lippincott-Raven, Philadelphia, NY.

Wells, L., Whalen, S.A., Hart, G.W. (2003) O-GlcNAc: a regulatory post-translational modification, *Biochem. Biophys. Res. Commun.* **302**, 435–441.

Weng, S., Gu, K., Hammond, P.W., Lohse, P., Rise, C., Wagner, R.W., Wright, M.C., Kuimelis, R.G. (2002) Generating addressable protein microarrays with PROfusion covalent mRNA-protein fusion technology, *Proteomics* **2**, 48–57.

Wilkins, M.R., Sanchez, J.C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F., Williams, K.L. (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it, *Biotechnol. Genet. Eng. Rev.* **13**, 19–50.

Wilkins, M.R., Williams, K.L., Appel, R.D., Hochstrasser, D.F. (Eds.) (1997) *Proteome Research: New Frontiers in Functional Genomics (Principles and Practice)*, Springer-Verlag.

Wilson, N.L., Schulz, B.L., Karlsson, N.G., Packer, N.H. (2002) Sequential analysis of N- and O-linked glycosylation of 2D-PAGE separated glycoproteins, *J. Proteome Res.* **1**, 521–529.

Winslow, R.L., Boguski, M.S. (2003) Genome informatics: current status and future prospects, *Circ. Res.* **92**, 953–961.

Wolters, D.A., Washburn, M.P., Yates, J.R. III (2001) An automated multidimensional protein identification technology for shotgun proteomics, *Anal. Chem.* **73**, 5683–5690.

Yao, J., Burton, J.L., Saama, P., Sipkovsky, S., Coussens, P.M. (2001) Generation of EST and cDNA microarray resources for the study of bovine immunobiology, *Acta Vet. Scand.* **42**, 391–405.

Yates, J.R., Eng, J.K., McCormack, A.L., Schieltz, D. III (1995) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database, *Anal. Chem.* **67**, 1426–1436.

Yost, R.A., Boyd, R.K. (1990) Tandem mass spectrometry: quadrupole and hybrid instruments, *Methods Enzymol.* **193**, 154–200.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Li, J., Liu, Z., Qi, Q., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Zhao, W., Li, P., Chen, W., Zhang, Y., Hu, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Tao, M., Zhu, L., Yuan, L., Yang, H. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica), *Science* **296**, 79–92.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R.A., Gerstein, M., Snyder, M. (2001) Global analysis of protein activities using proteome chips, *Science* **293**, 2101–2105.

Matrix, Extracellular and Interstitial

Abebe Akalu and Peter C. Brooks
New York University School of Medicine, NY, USA

1 Introduction 47

2 Structural Organization and Molecular Composition of the Extracellular Matrix 47
2.1 Molecular Composition and Structural Organization of the Basement Membrane 50
2.2 Collagen Type IV 50
2.3 Laminin 51
2.4 Nidogen 52
2.5 Perlecan 53

3 Molecular Composition and Structural Organization of the Interstitial Matrix 54
3.1 Collagen 55
3.2 Fibronectin 56
3.3 Vitronectin 57
3.4 Thrombospondin 58

4 Cellular Receptors for ECM Components 59

5 Mechanisms of ECM Modification and Remodeling 62

6 Modulation of Cellular Behavior by Intact and Structurally Altered ECM Components 65

7 Control of Physiological and Pathological Processes by Interactions with the ECM 67

8 Conclusions 70

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

Acknowledgments 70

Bibliography 70
Books and Reviews 70
Primary Literature 71

Keywords

Basement Membrane (BM)
A highly ordered thin sheetlike sub-compartment of the ECM composed of networks of glycoproteins and proteoglycans that separate epithelia and endothelia from the underlying stroma.

Cryptic ECM Epitope
A specific amino acid sequence of an ECM protein that is normally not accessible for cellular interactions under normal physiological conditions.

Extracellular Matrix (ECM)
A complex acellular network of adhesive collagenous and noncollagenous glycoproteins embedded in a hydrated gel composed of a mixture of proteoglycans.

Interstitial Matrix
A loosely organized sub-compartment of the ECM composed of glycoproteins, proteoglycans, and matricellular proteins that surround individual cells and tissues.

Matricellular Proteins
ECM-associated proteins that modulate cellular functions but which do not contribute significantly to structural architecture of the matrix.

The requirement of bidirectional communication between cells and their local acellular microenvironment for proper development and homeostasis in multicellular organisms has been appreciated for decades. Interestingly, intriguing new studies have suggested that some of this molecular information may be cryptic or hidden within the three-dimensional architecture of proteins and requires structural modification of the ECM for efficient transmission and utilization by cells and tissues. Thus, the dynamic interplay between cells and both intact and structurally altered ECM proteins help establish specificity, coordinate, and refine the complex interconnected biological responses critical for normal physiological processes. In this regard, unique insight will likely be gained into pathological events by a more complete understanding of the integrated network of molecular signals regulated by cellular interactions with the ECM.

1 Introduction

It has been known for decades that multicellular organisms depend on information within the local acellular microenvironment for proper structural organization, development, and homeostasis. The complex interconnected networks of molecules that compose the extracellular environment impacts nearly every aspect of normal physiology and disease, ranging from molecular control of gene expression and cell cycle progression to mechanical and architectural organization of tissues and organs. The major interests of many early investigators within the fields of molecular and developmental biology were primarily focused on the cellular compartment, including membrane, cytoplasmic, and nuclear components. However, as more sophisticated technology and tools were developed, it quickly became apparent that cells were not only physically linked to the insoluble scaffolding proteins outside the cell, but that cell shape, function, and survival depended on unique interactions and the bidirectional flow of information from outside the cell to the cells interior. Thus, the stage was set to begin the fascinating journey into an in-depth understanding of the biological responses controlled by the extracellular matrix (ECM).

2 Structural Organization and Molecular Composition of the Extracellular Matrix

For the purpose of our discussion, the ECM can be thought of as being composed of two general compartments, including the basement membrane (BM) or basal lamina and the interstitial matrix (Fig. 1). While these two compartments can be studied as separate features, they do not exist in isolation, but rather are interconnected by a variety of anchoring and

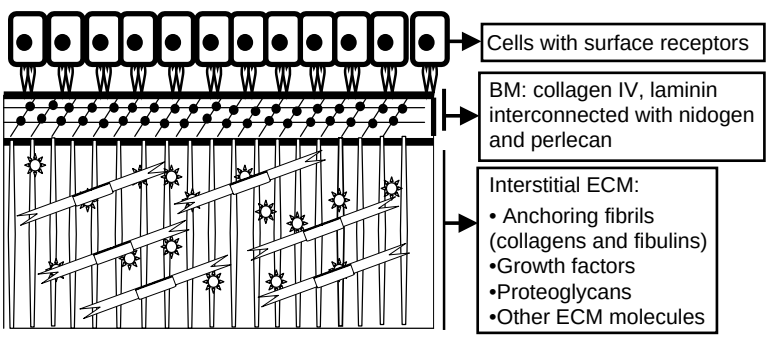

Fig. 1 Schematic representation of the structural organization and composition of ECM. The extracellular matrix can be organized into two compartments, including the basement membrane (BM) and the interstitial matrix. These two compartments are connected by a series of anchoring fibers, which consist of distinct forms of collagen, fibulins, as well as other ECM macromolecules. The major components of the BM include collagen type IV, laminin, nidogen, and perlecan. The interstitial matrix is composed of a gel-like composition of ECM proteins, including fibronectin, distinct forms of collagen, vitronectin, and various matricellular proteins such as thrombospondin. Growth factors, proteases, and a number of proteoglycans can also be found in the interstitial matrix.

interconnecting fibrils. Thus, a highly integrated continuum exists between the BM and interstitial matrix. In general terms, the BM is organized into a highly ordered meshlike network composed of structural glycoproteins, proteoglycans, and other macromolecules. BMs can be found underlying mostly all epithelial sheets and blood vessels, thus contributing to physical and functional compartmentalization. In contrast, the interstitial matrix can be characterized as a loosely organized gel-like composition of collagenous and noncollagenous glycoproteins, proteoglycans, and matricellular proteins that surround cells, tissues, and organs. The molecular composition of the BM and interstitial matrix can vary widely, depending on the repertoire of cells present within a given tissue. Moreover, the composition and structural integrity can also vary, depending on the particular physiological and/or pathological processes that may be occurring at any given time.

While heterogeneous in composition, examples of the major components of the BM include collagen type IV, laminin, nidogen, and perlecan. In addition to cross connections mediated by nidogen, collagen, and laminin, these molecules can interact with each other via a number of unique functional domains. While these macromolecules make up the majority of the BM, other minor, but no less important, components contribute to its structure, including genetically distinct forms of collagen (i.e. collagen type VIII, XV, and XVIII) and numerous proteoglycans. Examples of the major components of the interstitial matrix include an array of fiber forming collagens, including collagen type I, II, and III. Other major components that have been characterized include fibronectin, vitronectin, fibrinogen, thrombospondin, and elastin (Table 1).

Tab. 1 Partial list of the major components of the basement membrane and interstitial matrix.

1. Composition of basement membrane
Collagen IV
Laminin
Nidogen
Perlecan
2. Composition of interstitial matrix
Collagen type I, II, III
Fibronectin
Thrombospondin
Vitronectin
SPARC
Syndecan
Proteoglycans
Fibrinogen/Fibrin
Elastin

The complex sets of matrix macromolecules are not randomly organized, but are specifically ordered and arranged into supramolecular assembles. The physical interactions that facilitate the unique architectural organization are mediated, in large part, by distinct protein–protein binding domains, glucosaminoglycan (GAG) side chains, and a variety of carbohydrate moieties. For example, distinct functional domains found within nidogen bind collagen, and laminin, while unique domains within collagen can self-associate with other collagen molecules as well as other ECM proteins. These structural glycoproteins and proteoglycans also provide functional binding sites for a variety of growth factors, cytokines, proteolytic enzymes, and cell adhesion receptors. Thus, it is clear that besides participating in the structural assembly of mechanical scaffolds, ECM molecules facilitate localization and storage of a number of crucial regulatory factors. ECM localization of these regulatory molecules may serve to

protect them from degradation and/or induce conformational changes necessary for optimal presentation to cell surface receptors. As can be appreciated from the vast complexity of protein–protein interactions, it would be well beyond the scope of our discussion to analyze in-depth the vast number of molecular association that contribute to the assembly of the ECM.

While it was once thought that the ECM was merely an insoluble scaffold that primarily functioned in providing mechanical support for cells and tissues, it is now known that the ECM can regulate a variety of processes. Some of the more well-characterized events (Fig. 2) regulated by the ECM include adhesion, migration, invasion, cell cycle progression, proliferation, cell survival, differentiation, gene expression, tissue compartmentalization, and filtration. The ECM is not simply a static structural scaffold, but is rather dynamic, with changes occurring in molecular composition and structure. This dynamic feature is critical in helping to integrate the complex sets of regulatory signals transferred to cells from the ECM. In fact, fine-tuning these matrix-derived signals can be accomplished by structurally altering matrix components, which may modify existing binding sites for growth factors, cytokines, proteases, and cell surface receptors. Conformational changes in matrix molecules may expose new cryptic binding sites for secreted molecules along with creating novel binding sites for cell surface receptors such as integrins. Notably, the compositional and structural changes occurring within the ECM can be facilitated by changes in the secretion, structural organization, glycation, and cross-linking of proteins. The wealth of biochemical and molecular information stored within the ECM can be

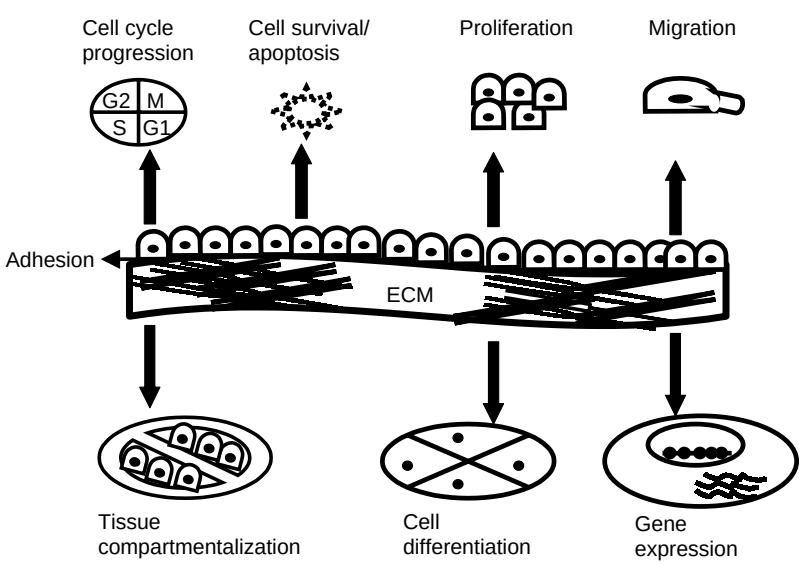

Fig. 2 Cellular processes regulated by the ECM. Cellular interactions with the extracellular matrix can regulate a wide array of important cellular processes. A partial list of the well-characterized events include cell cycle progression, cell survival, and apoptosis, proliferation, migration, adhesion, tissue compartmentalization, differentiation, and gene expression.

transmitted to the cells by binding to a number of cell surface receptors, including integrins, cell surface proteoglycans, and an emerging class of receptor tyrosine kinases. While examples of all these classes of receptors have been shown to mediate cellular interactions with the ECM, it could be argued that integrins are the most well-studied group of molecules that mediate cell–ECM interactions.

2.1 Molecular Composition and Structural Organization of the Basement Membrane

As mentioned above, the BM or basal lamina is a highly ordered meshlike network of glycoproteins and proteoglycans that separate epithelium and endothelium from the underlying stroma. Given the diversity of cell types that secrete ECM molecules, it is not surprising that the molecular composition of BMs can vary from tissue to tissue. While molecular diversity is evident, a number of major components are common to most BMs and include collagen type IV, laminin, nidogen, and perlecan. These molecular components are not meant to represent a complete list of the BM molecules, but rather, examples of well-characterized components to help illustrate its compositional and structural organization.

2.2 Collagen Type IV

The major collagenous component of the BM is collagen type IV. A variety of other minor collagen types are also present, including collagen types VIII, XV, and XVIII. Collagen type IV molecules are composed of a combination of 3 α chains organized in a triple helical manner. Six different collagen type IV chains ($\alpha1$–$\alpha6$) have been identified, each of which represents a distinct gene product. One of the most widely distributed isoforms of collagen type IV present in nearly all BMs is the triple helical molecule composed of 2 $\alpha1$(IV) chains and 1 $\alpha2$(IV) chain. Interestingly, other isotypes of type IV collagen have been detected in the BMs from distinct tissues, suggesting tissue-specific distribution patterns. Collagen type IV is organized into functional domains, including a large central triple helical region composed of repeating units with glycine at every third position. While this repeated pattern and triple helical structure exists throughout the central domain, interruptions on this unique organization also exist. The triple helical region has been shown to be critical for mediating cell adhesion, as a number of integrin-binding sites have been localized to this area. Moreover, several binding sites for other ECM proteins are also located within the triple helical region, suggesting an important role for this region in the collagen network assembly. The amino terminus of collagen type IV contains a globular region termed the *7S domain*. The 7S domain facilitates association of collagen molecules into tetramers, which is important in the formation of higher-order collagen networks. A third important structural and functional feature is the C-terminal globular noncollagenous (NC1) domain. These NC1 domains are thought to facilitate selection of individual α chains during network assembly. Interestingly, recent studies have indicated that specific integrin-binding sites are not only confined to the triple helical domain but are also located in NC1 domains. For example, studies have indicated that the NC1 domain from $\alpha1$ chain can interact with integrin $\alpha1\beta1$, while the NC1 domain from the $\alpha2$ chain supports interactions with integrins $\alpha3\beta1$, $\alpha v\beta3$, and $\alpha v\beta5$.

Moreover, NC1 domains from the $\alpha 3$ and $\alpha 6$ chains of collagen type IV specifically bind to integrin $\alpha v\beta 3$, while NC1 domains form $\alpha 4$ and $\alpha 5$ chains fail to interact with integrin receptors. Given the facts that ligation of distinct integrin receptors can initiate unique signaling cascades, integrin interactions with noncollagenous (NC) domains may regulate a diversity of cellular processes. To this end, it is becoming increasingly clear that NC1 domains from distinct forms of collagens, such as types IV, VIII, XV, and XVIII, may possess unique regulatory properties since potent antiangiogenic and antitumorgenic modules have been identified from these regions.

Several families of molecules have been shown to mediate cellular interactions with collagen (18–20, 29, 30). However, integrin receptors likely represent the predominate class of cell surface molecules, facilitating interactions with collagen type IV. A number of distinct integrin heterodimers interact with triple helical collagen IV, including integrins $\alpha 1\beta 1$, $\alpha 2\beta 1$, and $\alpha 3\beta 1$. During invasive cellular events, such as tumor invasion, metastasis, and angiogenesis, the triple helical structure of collagen can be altered, leading to disruption of $\beta 1$ integrin binding, resulting in exposure of cryptic $\alpha v\beta 3$ bindings sites. This shift in integrin-mediated recognition could result in the initiation of unique signaling cascades. Thus, minor changes in the three-dimensional structure of collagen could profoundly impact signaling cascades, thereby modifying the biological response to the ECM.

2.3 Laminin

A second large glycoprotein that forms an interconnected sheetlike network with collagen type IV is laminin. Laminin represents a family of adhesive glycoproteins comprised of high molecular weight (about 800 kDa) disulfide-linked heterotrimers with chains designated α, β, and γ. Since the original discovery of laminin many years ago, five different α-chains ($\alpha 1$ to $\alpha 5$), three different β chains ($\beta 1$ to $\beta 3$), and three different γ-chains ($\gamma 1$ to $\gamma 3$) have been identified. Although the theoretical combinations of the three individual chains could allow up to 45 different trimeric isoforms, to date 12 isoforms have been described in detail. Of the 12 isoforms known, it could be argued that laminin-1 ($\alpha 1\beta 1\gamma 1$) and laminin-5 ($\alpha 3\beta 3\gamma 2$) are perhaps the most well characterized. The individual chains of laminin share a significant degree of structural similarity, and can be organized into a number of functional domains, including globular domains (IV and VI), epidermal growth factor (EGF)-like laminin-repeats (LE-module) (III and V), and an alpha-helical coiled-coil domain.

In general, laminin is organized into a crosslike structure containing three short arms and one long arm. The short arms are formed by the N-terminal region of one of the three chains, whereas the long arm is a rod-like coiled-coil structure comprised of the carboxyl termini of all three subunits. The long arm ends in a globular structure called the *G-domain*. The G-domain can be further subdivided into five globular subdomains (G1–G5). The G-domains contain major binding-sites for cellular receptors, such as integrins, syndecans, and α-dystroglycan. Moreover, laminin also possesses binding sites for minor collagen species such as collagen type XVIII. In addition to a number of proteoglycans, the BM protein nidogen directly interacts with laminin. Many of these unique protein–protein interactions

are thought to be critical for structural assembly of the BM.

While distinct forms of laminin are found mainly in the BM, studies have also identified isoforms present within the matrix of embryonic mesenchyme, loose connective tissue, and developing central nervous system. Interestingly, throughout embryonic development and in the adult organisms, studies have identified different patterns of cell and tissue-specific expression. In fact, distinct isoforms of laminin have their own temporally and spatially regulated expression profiles. For example, while laminin-10 exhibits wide tissue distribution, the expression of other isoforms is tightly regulated. In fact, laminin-2 is expressed in the BM of skeletal muscle and plays a role in neurite out growth, while laminin-5 is a constituent of the BM of many epithelial tissues where it regulates stable adhesion and promotes migration. Thus, it is clear that interactions with laminin can regulate a variety of cellular and physiological processes.

Laminin interacts with cells through a variety of mechanisms, including proteoglycans and integrin receptors. The major integrins known to interact with laminin include $\alpha1\beta1$, $\alpha2\beta1$, $\alpha3\beta1$, $\alpha6\beta1$, $\alpha6\beta4$, $\alpha7\beta1$, whereas examples of nonintegrin receptors include the 67-kDa laminin receptors and dystroglycan. Recent studies have suggested that laminin the $\alpha4$ chain is a specific high affinity ligand for the $\alpha v\beta3$ and $\alpha3\beta1$. Notably, distinct integrin binding sites have been identified within laminin, including the classical RGD amino acid recognition sequence as well as a number of non RGD integrin binding sites. As was seen with collagen type IV, the molecular and structural integrity of laminin can impact integrin recognition and modify cellular responses. For example, studies have identified a cryptic site exposed following proteolytic cleavage of laminin-5 that promotes breast carcinoma cell migration.

2.4 Nidogen

As discussed above, the two major structural glycoproteins collagen type IV and laminin are organized into two sheetlike networks. These complex meshlike assemblies are interconnected by the protein Nidogen. Nidogen, also known as *entactin*, is a ubiquitous component of the BM and is conserved in all metazoans. Interestingly, invertebrates, including *Caenorhabditis elegans*, possess only a single gene, whereas in mammals, two nidogen isoforms have been identified termed *nidogen-1* (150 kDa) and *nidogen-2* (200 kDa). Nidogen-2 is predominately located in endothelial BM, whereas nidogen-1 can be ubiquitously distributed within the basal lamina from a variety of tissues. These tissue distribution patterns suggest that different forms of nidogen may play distinct roles in tissue organization, BM assembly, and in the regulation of physiological processes such as blood vessel development. Studies have indicated that nidogen-2 binds to endostatin, an antiangiogenic module of the NC1 domain of collagen type XVIII (41). The two nidogen isoforms share substantial domain and sequence identity with an overlapping binding repertoire for laminins, collagen IV, perlecan, and fibulin. Nidogen consists of three globular domains (G1 to G3). Domain G2 of nidogen-1 contains the binding site for collagen IV, perlecan, and fibulins. In contrast, domain G3 contains high affinity binding sites for the laminin γ chain, collagen IV, and weaker binding sites for fibulins.

Studies have suggested that nidogen may play an important role in BM organization by facilitating connections between laminin and collagen IV networks. Interestingly, *C. elegans*, harboring mutant forms of nidogen-1, exhibited relatively normal assemble of type IV collagen, suggesting that nidogen-1 may not be essential for type IV collagen assembly under these conditions. These findings are somewhat surprising, given the previous studies indicating the importance of nidogen in the structural organization of collagen IV and laminin networks. The nidogen-binding site within laminin has been localized to the laminin epidermal growth factor–like (LE) module γ1III4. Antibodies against γ1III3-5 modules were shown to disrupt early kidney, lung, and salivary gland development. In addition to nidogen's function in stabilizing the BM, it has also been implicated in other protein integration functions such as facilitating interactions with ECM components such as perlecan and fibulin. As with many ECM proteins, nidogen plays roles in signal transduction through integrins $\alpha v\beta 3$ and $\alpha 3\beta 1$, thereby influencing the biological response of the cells to the ECM.

2.5 Perlecan

Besides the major glycoprotein components of the basal lamina, a second family of molecules that contribute to the formation of the BM includes the proteoglycans. Proteoglycans are not only important components of the ECM but are also found on the cell surface. Proteoglycans can be characterized as consisting of a core protein with negatively charged polysaccharide glucosaminoglycan (GAG) side chain. On the basis of their GAG side chains, proteoglycans can be classified into four main groups, including chondrotin sulfate, dermatan sulfate, keratin sulfate, and heparan sulfate. The cellular and tissue distribution profiles along with their biological functions have been suggested to depend on both the GAG chain and the core protein content. Interestingly, the GAG side chains attached to the core proteins have been shown to vary depending on the tissue. For example, phosphacan, a chondroitin sulfate proteoglycans expressed abundantly in neural tissue contains keratin sulfate in rat embryonal cerebellum, but lack these GAGs in rat retinal tissues. Thus, variations in the structural features of these molecules may modify their functions, and thus alter cellular behavior in a tissue-specific manner.

In addition to the structural diversity associated with the GAG chains, the core proteins also consist of structurally distinct domains. In fact, proteoglycans can interact with various biologically active molecules through specific domains of the core protein, thus expanding their diverse capacity to impact biological responses. Recent studies have identified more than 30 structurally distinct proteoglycans from vertebrate tissues. Many proteoglycans bind several different ECM components and growth factors, and play critical roles in processes such as cell adhesion, migration, and differentiation. Notably, heparan sulfate proteoglycans can be subclassified as those present on the cell surface, such as syndecan, those anchored to the cell surface by glycosylphosphatidylinositol such as glypican, and those present in the BM, including perlecan and agrin.

Perlecan is a multidomain heparan sulfate proteoglycan with a wide distribution in the ECM. Perlecan has been shown to be an important element for assembly

and function of BMs. Mammalian perlecan consists of a 480-kDa core protein containing three to four heparin sulfate or chondroitin sulfate side chains. The molecular weight of the fully glycosylated protein is between 700 and 800 kDa. Structural studies have provided evidence that perlecan consists of five distinct domains. Domain I, has a molecular weight of 20 kDa and consists of 172 amino acids containing three consecutive Ser-Gly-Asp triplets that represent the attachment sites for GAGs. Domain II of mammalian perlecan is homologous to the low-density lipoprotein (LDL) receptor ligand-binding domain and contains the pentapeptide DGSDE. While domain III is homologous to the internal segments of laminin A, domain IV is homologous to neural cell adhesion molecule (N-CAM). Domain V is located at the C-terminus and consists of tandem arrays of three laminin G-type modules and four epidermal growth factor–like modules. This domain can bind to the laminin–nidogen complex, fibulin-2, sulfatides, $\beta 1$ integrins, and α-dystroglycan, thereby contributing to the BM organization and structure. Domain V also contains two SGxG tetrapeptides, which may serve as attachment sites for GAG chains other than heparan sulfate. In addition, this domain contains two Leu-Arg-Glu tripeptides that may facilitate binding of neurons to S-laminin in the synaptic BM. Finally, perlecan interacts with collagen type IV and fibronectin, further implicating it in the structural organization of the ECM.

Perlecan is produced during early and late stages of embryonic development and is localized in BMs, vessel walls, cartilage matrix, and other extracellular spaces. The importance of perlecan to mammalian development has been demonstrated in mouse knockout experiments. Although the BM can develop in the absence of perlecan, nearly half of all perlecan null mice died at embryonic day 10.5 with widespread cephalic, skeletal abnormalities, and aberrant BM formation in the heart. Studies have indicated that perlecan can regulate a variety of biological functions. Some of these well-characterized functions include its role in maintenance of the integrity of BM, and the storage of growth factors and cytokines within the ECM. Interestingly, proteolytic cleavage of perlecan may result in the release of proangiogenic factors such as fibroblast growth factors and/or antiangiogenic factors such as endorepellin, a newly identified antiangiogenic fragment of perlecan.

3 Molecular Composition and Structural Organization of the Interstitial Matrix

The ECM can be viewed as being composed of two compartments, including the BM and the interstitial matrix. While the general architecture of the BM is identifiable as a thin highly organized sheetlike structure. The interstitial matrix can be described as a loose organization of collagenous and noncollagenous glycoproteins, matricellular proteins, and proteoglycans that surrounds individual cells, tissues, and organs. It is important to note that while these compartments can be viewed separately, they are indeed intimately connected by anchoring fibers of collagen and fibulins. The components of interstitial matrix from a given microenvironment can vary, depending on the cellular repertoire of the tissue. Given the diversity of cell types that secrete ECM molecules, it would be well beyond the scope of our discussion

to attempt to describe in-depth all the components that contribute to the interstitial matrix. Thus, we have chosen to focus on a few examples of the major components, including collagen type I, the glycoproteins fibronectin, vitronectin, and the matricellular protein thrombospondin. These molecular components are not meant to represent a complete list of interstitial matrix molecules, but rather are examples of well-characterized components to help illustrate the composition, structural organization, and function of the interstitial matrix.

3.1 Collagen

The great importance of collagen in physiology, development, and tissue homeostasis has been known for decades. Its singular importance in all metazoan organisms is evident, given its high degree of conservation throughout species as diverse as sponges and man. Over the past several decades, a wealth of biochemical and genetic studies have demonstrated critical roles for collagen in structural organization and function of the ECM. In fact, a wide array of human diseases have been linked to aberrant synthesis, mutations, and biochemical modification of collagen, including Osteogenesis Imperfecta, Ehlers-Danlos Syndrome type V11, and Alport Syndrome, to name just a few. Collagen represents the most abundant ECM protein in the body. It has been estimated that at least 30 distinct genes code for specific α chains of collagen. In fact, at least 21 genetically distinct forms of collagen have been identified. The numerous forms of collagen can be grouped according to their general molecular structure, organization, and function, which include fibrillar collagens, nonfibrillar network forming collagen, FACIT (fibril-associated collagens with interrupted triple helices) collagen, and multiplexes. Examples of the multiplexin subgroup include collagen type XV and XVIII. Interestingly, these types of collagen have received much attention recently due to the discovery of potent antiangiogenic activity within isolated fragments (Restin and Endostatin) of their respective NC domains. As described above, collagen type IV is an important example of the nonfibrillar networking forming collagen, while collagen type IX and XII are examples of FACIT collagens. Some of the earliest forms of collagen to be described belong to the fibrillar collagen subgroup such as collagen type I, II, and III.

It has been estimated that collagen type I may represent up to 25% of the total protein in the body and the majority of the protein within the interstitial matrix. Therefore, we have chosen to focus on collagen type I as an example of a major collagenous protein within the interstitial matrix. Collagen type I is composed of three α chains, including two $\alpha 1$ chains and one $\alpha 2$ chain organized into a triple helical structure. These collagen type I chains exhibit the characteristic amino acid triplet repeat G-X-Y with Glycine (G) at every third position. The amino acids proline and hydroxyproline are also often represented at the X and/or Y positions within the triplet. The relatively small size of the glycine residue at the third position has been suggested to be critical for proper folding of the triple helical structure. Secretion of procollagen is followed by cleavage of both the N- and C-terminal propeptides. N- and C-terminal propeptide processing is crucial in facilitating collagen fibril formation and higher-order supra molecular assemblies. Interestingly, soluble propeptide fragments have been

detected in serum, and these correlate with enhanced collagen synthesis and bone metastasis. Moreover, propeptide collagen fragments have also been shown to exhibit a number of biological activities, including promoting chemotactic activity as well as inhibiting collagen synthesis under specific conditions. Thus, the maturation and synthesis of collagen type I appears to regulate a number of cellular processes.

Similar to what was discussed with other ECM molecules, collagen type I can interact with a variety of proteins, including other ECM molecules such as fibronectin, thrombospondin, secreted protein acidic rich in cysteine (SPARC) and a number of proteoglycans. These distinct protein–protein interactions can be altered by structural and biochemical changes, including altered hydroxylation, cross-linking, glycation, and the degree of triple helical formation. For example, structural denaturation of the triple helical region can expose unique cryptic binding sites for additional molecules. Moreover, collagen interactions with distinct proteoglycans such as decorin have been shown to specifically alter collagen structure, fibril, and matrix formation. In addition, collagen type I can interact with many other molecules, including proteolytic enzymes and cell surface receptors such as integrins, which facilitate communication between cells and the matrix. The major integrin receptors mediating cellular interactions with collagen include members of the $\beta 1$ integrin subfamily such as $\alpha 1\beta 1$ and $\alpha 2\beta 1$. Shifts in integrin interactions with collagen may result in the initiation of unique signaling pathways, leading to alterations in biological responses, including cell adhesion, migration, proliferation, survival, and gene expression.

3.2 Fibronectin

The large glycoprotein fibronectin is a multidomain molecule found in the ECM, on the surface of cells, and in plasma and other body fluids. Fibronectin can exist in multiple soluble and insoluble forms. Soluble forms of fibronectin represents a major component in the plasma and other bodily fluids, while insoluble forms are either associated with the cell surface or are incorporated into the interstitial matrix. The fibrillar form of fibronectin is of considerable interest to the present discussion because of its capacity to mediate an array of interactions with other ECM proteins. In particular, this multidomain glycoprotein exists as a dimer composed of a two 250-kDa monomers linked covalently near the C-termini by a pair of disulfide bonds. Each monomer is comprised of multiple regions termed *fibronectin type repeats*. Fibronectin is organized into multiple domains containing 12 type I, 11 type II and 15 to 17 type III repeats.

The complex structural organization of fibronectin helps facilitate interactions with a variety of molecules. For example, fibronectin contain binding sites for fibrin, native and denatured collagens, heparin and a number of growth factors. Interestingly, recently studies have indicated that fibronectin type II modules can interact with collagen types II, III, IV, V, and X. Moreover, fibronectin also contains heparin-binding domains that interact with heparin and chondroitin sulfate proteoglycans. Fibronectin type III repeats can bind to the matricellular protein tenacin. Taken together, these observations imply that the biological activities of fibronectin involve a number of critical interactions with a broad range of molecules.

Besides providing critical binding sites for many components of the interstitial matrix, fibronectin is well known for its capacity to facilitate cell adhesion, migration, proliferation, and cell survival. These important cellular processes are regulated in large part by integrins. Examples of the integrins known to bind to fibronectin include $\alpha5\beta1$, $\alpha4\beta1$, $\alpha9\beta1$, $\alpha4\beta7$, $\alpha v\beta3$, and $\alpha v\beta6$. One of the most common integrin receptors for fibronectin is $\alpha5\beta1$, which binds to fibronectin repeats I_{1-9}, $II_{1,2}$, III_9, and III_{10}. Importantly, the ability of fibronectin to facilitate cellular interactions mediated by multiple integrin binding events provides an opportunity to initiate distinct signaling cascades.

3.3 Vitronectin

Another important glycoprotein associated with the interstitial matrix is vitronectin. Vitronectin is a 75-kDa adhesive glycoprotein that is present in both the plasma and the ECM. It circulates in blood in a monomeric (native) form, but is converted into disulfide-linked multimeric (denatured) form when incorporated into the ECM. The native vitronectin molecules contain 459 amino acid residues organized into functionally distinct domains. The N-terminal domain (residues 1–44) of vitronectin is rich in cystine and is termed the *somatomedin B* (SMB) domain. Importantly, the well-known integrin recognition sequence Arg-Gly-Asp (R-G-D) is located in the C-terminus of the SMB domain residues. The most C-terminal portion (residues 348–370) of vitronectin contains sequences homologous to hemopexin and has a positively charged heparin-binding segment.

Although vitronectin is predominately produced in the liver and found in the circulation, it is also found in a variety of tissues, particularly during wound healing, inflammation, angiogenesis, and during tumor growth. Interestingly, some of the unique protein-binding properties of vitronectin depend on its structural conformation and degree of multimerization. For example, multimeric forms of vitronectin exhibit enhanced capacity to bind to a number of molecules, including other ECM proteins and specific proteoglycans. Moreover, vitronectin has been shown to interact with growth factors and growth factor–binding proteins present within the ECM such as transforming growth factor-β (TGFβ) and insulin-like growth factor–binding protein-5 (IGFBP-5). Vitronectin binding to IGFBP-5 may act to modulate its bioavailability, thus impacting the function of insulin-like growth factor-1 (IGF-1), which in turn could alter cellular migration and proliferation. Adding to the diversity of molecules known to interact with vitronectin include factors such as the C5b-8 complement complex, the thrombin-antithrombin III complex, plasminogen activator inhibitor 1 (PAI-1), urokinase plasminogen activator receptor (uPAR), heparin, plasminogen, and β-endorphin. Vitronectin has also been shown to regulate cell adhesion, migration, proliferation, and cell survival. The vast majority of these cellular functions are regulated by integrins. The major integrin receptors known to interact with vitronectin include the classical vitronectin receptor $\alpha v\beta3$ as well as a variety of other αv-containing integrins such as $\alpha v\beta1$, $\alpha v\beta5$, $\alpha v\beta6$, and $\alpha v\beta8$. Again, the variety of signaling pathways initiated by distinct integrin ligation events can serve to coordinate, transmit, and regulate cellular responses associated with vitronectin.

3.4 Thrombospondin

A number of matricellular proteins are also important components of the interstitial matrix. Matricellular proteins have been defined as those ECM-associated proteins that modulate cell–ECM interactions and/or cellular functions, but which do not contribute significantly to the structural architecture of the matrix. Some examples of well-characterized matricellular proteins include tenacin, osteopontin, thrombospondin, and SPARC. We have chosen to focus primarily on thrombospondin as one example of the important matricellular protein family.

Members of the thrombospondin family can be subdivided into at least two groups on the basis of their general structure and domain organization. For example, thrombospondin-1 (TSP1) and thrombospondin-2 (TSP2) are trimeric molecules, which contain a procollagen-like domain and type 1, and type III repeat (TRS) domains. In contrast, TSP3, TSP4, and TSP5 are pentameric molecules, which lack the procollagen-like domain. In general, members of the thrombospondin family can be characterized as multimeric, calcium-binding glycoproteins that regulate a variety of cellular functions. As an example, thrombospondin-1 is a 420-kDa homotrimeric glycoprotein, which contains N- and C-terminal globular domains. The N- and C-terminal globular domains interact with heparin, and heparin sulfate proteoglycans along with facilitating cell attachment, spreading, migration, and platelet aggregation. The procollagen-like domain plays a role in subunit assembly along with exhibiting the capacity to inhibit angiogenesis, while the type III repeat facilitates both calcium and cell binding. While this multifunctional protein can interact with other ECM molecules such as heparin sulfate proteoglycans, it is not considered to play a major role in structural organization of the interstitial matrix. Thrombospondin is secreted by a variety of cells, including macrophages, fibroblasts, and endothelial cells, and is distributed within a variety of tissues. Interestingly, in the adult organism, thrombospondin has been observed predominately at sites of injury and tissue remodeling.

As has been observed with most ECM proteins, thrombospondin can interact with a variety of distinct proteins, thereby playing roles in the regulation of many cellular processes. The functions of TSP1 are thought to be mediated by direct binding to cell surface receptors, including integrins. In fact, TSP1 binds to a number of integrins such as, $\alpha v\beta 3$, $\alpha IIb\beta 3$, $\alpha 3\beta 1$, $\alpha 4\beta 1$, and $\alpha 5\beta 1$. Moreover, a number of nonintegrin-binding proteins or receptors have also identified, including the low-density lipoprotein receptor-related protein (LRP), proteoglycans, CD36, and CD47. Beyond the ability of thrombospondin to interact with cell surface receptors, studies have revealed a complex array of other binding proteins ranging from other ECM molecules such as fibrinogen, fibronectin, and heparin sulfate proteoglycans to a variety of proteases, including thrombin, elastase, and certain MMPs. Given the capacity of thrombospondin to interact with such a diversity of distinct protein, it is not surprising that it has been shown to regulate many important cellular events, including attachment, spreading, motility, cytoskeletal organization, proliferation, cell–cell aggregation, apoptosis, transcriptional regulation, differentiation, and angiogenesis. Thus, as can be clearly appreciated from the relatively short discussion of thrombospondin,

this ECM protein plays major roles in regulating a diversity of biological responses to the interstitial ECM.

4
Cellular Receptors for ECM Components

The ability of cells to respond to their extracellular environment depends on bidirectional communication with stromal elements. In fact, this bidirectional flow of information is crucial to the development and homeostasis of multicellular organisms. Thus, to detect, integrate, and transmit information from outside the cell to the inside and, in turn, to respond to specific microenvironmental changes, cells need a network of unique sensor and responder-type receptor molecules with the capacity to assimilate, organize, and use the information within the ECM. To this end, at least five families of cell adhesion receptors have been identified with the ability to mediate either cell–cell or cell–ECM interactions (Fig. 3). The major groups of cell adhesion molecules include selectin, cadherins, members of the immunoglobulin super gene family, cell surface proteoglycans, and integrins. Interestingly, recent studies have also identified an emerging group of receptor tyrosine kinases that may facilitate cellular interactions with the ECM. These groups of cell adhesion receptors have the capacity to transmit biochemical and molecular information from the stromal environment. With some exceptions, the members of the selectin, cadherin, immunoglobulin, and proteoglycans families mediate cell–cell interactions, while integrins are

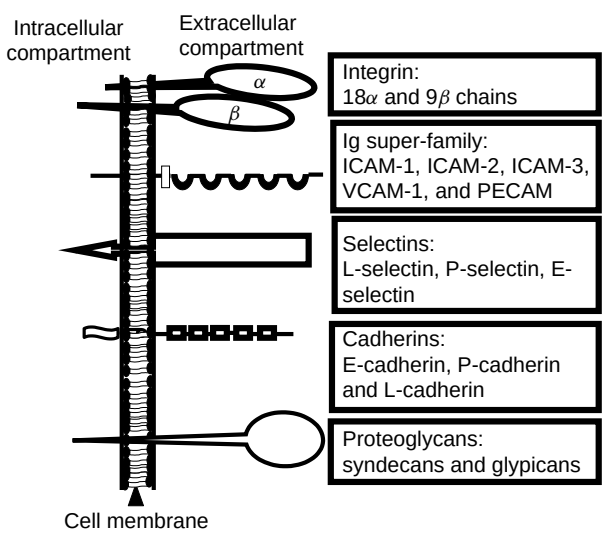

Fig. 3 Major families of cell adhesion molecules mediating cell–cell and/or cell–ECM interactions. In order for cells to sense and, in turn, respond to the acellular microenvironment, they require cell surface receptors with the capacity to facilitate the bidirectional flow of information between cells and their local ECM. Examples of the five major classes of cell adhesion receptors include integrins, Ig superfamily, selectins, cadherins, and cell surface proteoglycans.

the major receptors facilitating cell–ECM interactions. Some interesting examples of nonintegrin receptors mediating cellular communication with the ECM include the cell surface proteoglycans (Glypicans and Syndecans) as well as specific receptor tyrosine kinases.

Syndecans are transmembrane heparin sulfate proteoglycan. At least four distinct syndecans have been identified and studies have suggested that syndecans 1, 2, and 4 can bind to a number of ECM molecules. The syndecans are multidomain transmembrane molecules with an extracellular domain containing heparin sulfate GAG chains. It is thought that syndecan interactions with ECM components such as laminin and fibronectin are mediated primarily by interactions with the GAG chains rather than with core protein sequences. In fact, ECM binding specificity can be modulated by the degree of GAG sulfation as well as the length and composition of the disaccharide moieties. Syndecans have a conserved transmembrane domain and a short cytoplasmic domain containing three subdomains termed C1, V, and C2. The cytoplasmic domain has been shown to interact with cytoplasmic proteins such as src, cortactin, PIP2, and PKCα. The capacity of the syndecans to bind to these cytoplasmic proteins is thought to facilitate signal transduction. In this regard, syndecans contribute to the regulation of cellular proliferation and survival. Moreover, syndecans interact with a number of growth factors, cytokines, and protease inhibitors. These unique interactions can facilitate the regulation of protease and cytokine activity in close proximity to the cell surface.

As mentioned above, integrins are the major family of cell adhesion receptors that mediate cellular interactions with the ECM (Table 2). However, several $\beta1$ and $\beta2$ subunit containing integrins mediate cell–cell interactions by interacting with counter receptors and other binding ligands expressed on adjacent cells. Integrins represent a large family of heterodimeric transmembrane proteins composed of noncovalently associated α and β chains. The α and β chains represent distinct gene products with alternatively spliced forms identified. To date, at least 18 α and 9 β subunits have been described. Distinct combinations of the α and β chains lead to the formation of at least 24 distinct integrins with different ligand and tissues specificities. The functional integrin heterodimer is composed of a ligand-binding globular extracellular domain containing repeated regions with cation binding capacity. In addition, integrins have a transmembrane domain and a short cytoplasmic tail that has the capacity to bind to a wide array of cytosolic proteins known to be crucial in facilitating signal transduction. Studies have suggested that ligand-binding specificity

Tab. 2 Examples of selected integrins and their ECM ligands.

ECM ligand	*Integrin*
Collagens	$\alpha1\beta1$, $\alpha2\beta1$, $\alpha3\beta1$
Laminins	$\alpha1\beta1$, $\alpha2\beta1$, $\alpha6\beta1$, $\alpha7\beta1$, $\alpha6\beta4$, $\alpha v\beta3$, $\alpha3\beta1$
Nidogen	$\alpha3\beta1$, $\alpha v\beta3$
Fibronectin	$\alpha5\beta1$, $\alpha4\beta1$, $\alpha4\beta7$, $\alpha3\beta1$, $\alpha v\beta3$
Vitronectin	$\alpha v\beta1$, $\alpha v\beta3$, $\alpha v\beta5$, $\alpha IIb\beta3$, $\alpha v\beta6$, $\alpha v\beta8$
Thrombospondin	$\alpha3\beta1$, $\alpha v\beta3$, $\alpha IIb\beta3$, $\alpha4\beta1$, $\alpha v\beta1$

and relative affinity is dictated in part by structural conformation, cation binding, and specific contact residues within both the α and β chains. Thus, the affinity and ligand-binding specificity can be regulated at multiple levels, allowing fine-tuning of interactions.

Historically, integrins were thought to function as molecular glue, facilitating physical interactions between cells and the ECM. However, over the last decade, experimental evidence has expanded the repertoire of the molecular ligands and integrin functions (Fig. 4). For example, some of the more well-characterized processes regulated by integrins include cell adhesion, migration, proliferation, cell survival, differentiation, signal transduction, gene expression, and extracellular matrix assembly. Importantly, the multiplicity of functions regulated by integrins is attributable in large part to the vast number of distinct integrin-binding partners. Most of the classical integrin ligands initially identified were ECM proteins, including collagen, laminin, fibronectin, vitronectin, and thrombospondin. Early work on integrin ligands lead to the identification of the tripeptide RGD as an important sequence mediating cell adhesion. However, it has now been confirmed that a wide array of distinct non-RGD-containing proteins can interact with integrin receptors. In some respects, integrins can be thought of as molecular organizational centers on the cell surface since integrins can bind and regulate the activity of a number of different families of molecules (Fig. 4). For

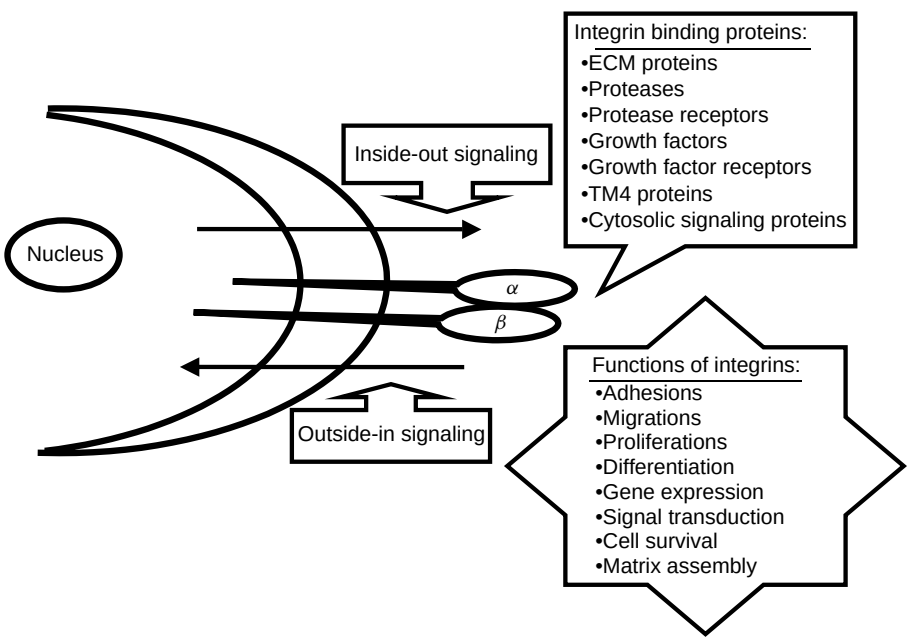

Fig. 4 Integrins: the predominate class of receptors mediating cell–ECM interactions. Integrins represent a family of multifunctional heterodimeric transmembrane receptors, which can bind a wide array of regulatory proteins. Integrin receptors can mediate ''inside-out'' signaling and ''outside-in'' signaling. These distinct signaling cascades can facility bidirectional communication between cells and the local microenvironment.

example, besides ECM proteins, recent studies have indicated that integrins can bind to and associate with such diverse sets of regulatory molecules as proteolytic enzymes, protease receptors, growth factors, growth factor receptors, and TM4 proteins. Thus, it is not surprising that integrins play major roles in regulating the structural assembly, integrity, and remodeling of the ECM. In turn, integrins serve as sensors and molecular responder proteins that facilitate integration and transmission of complex arrays of biochemical and mechanical information from the ECM to the cells. Thus, the capacity of integrin to act as bidirectional signaling receptors, facilitators, sensors, and responders makes them uniquely suited for the job of regulating biological responses to the ECM.

5 Mechanisms of ECM Modification and Remodeling

It is well known that the structural integrity of the BM and interstitial matrix can have a significant impact on cellular behavior. Thus, compositional, biochemical, and structural modification of the ECM can provide mechanisms to alter diverse sets of biological responses. In order to have a significant appreciation of how the ECM influences biological responses, one must first have some insight into the mechanisms by which the ECM is modified and how the unique modifications are recognized and/or interpreted by cells. There is a great diversity of potential mechanisms by which the various components of the ECM can be altered, and thus would be well beyond the scope of our discussion to analyze all of them in detail. Therefore, we have chosen to focus on a few interesting mechanisms as examples of how the ECM can be modulated (Fig. 5).

The molecular composition of distinct ECM compartments can vary substantially, depending on the tissue environment and physiological or pathological events that may be occurring at a particular time. The complex interplay between growth factors, integrins, proteases, and matrix components function cooperatively

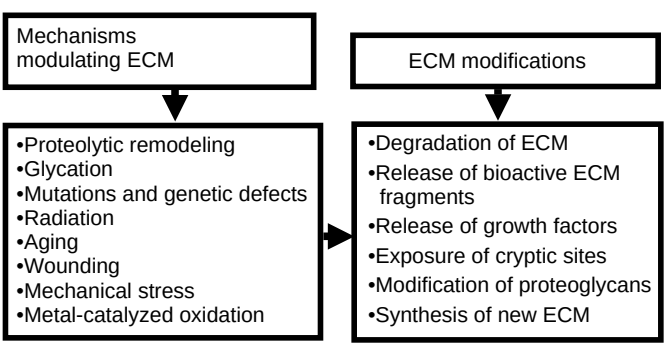

Fig. 5 Mechanisms by which ECM modification can altered cellular behavior. The extracellular microenvironment is not static, but is rather dynamic with numerous changes in structure, integrity, and composition, depending on the cellular repertoire and physiological processes that may be occurring. These important structural and biochemical modifications may lead to altered cellular interactions with the matrix, and thus modify biological responses to the ECM.

to regulate the structure and molecular composition of the ECM. During physiological processes such as embryo development, organogenesis, and wound healing responses, a variety of growth factors and cytokines, including members of the fibroblast growth factor (FGF) and TGFβ families can be released. These growth factors as well as many others can bind cognate receptors on the cell surface, initiating signaling cascades that regulate synthesis of specific ECM components, including collagen, laminin, and fibronectin. Recent studies using mutant embryoid bodies harboring mutations in FGF receptor-2 resulted in altered synthesis of collagen type IV and laminin. Interestingly, this FGF-regulated synthesis of collagen type IV and laminin may be controlled in part by the PI3 kinase and Akt/PKB signaling pathway. Altered ECM synthesis can lead to disruption of BM formation and epithelial differentiation. FGF signaling is also known to play important roles in regulating wound healing and angiogenesis. During these normal physiological responses, elevated levels of FGF have been observed, which, in turn, could lead to significant modification of the preexisting ECM.

A second important example of growth factor regulation of ECM protein synthesis involves TGFβ. It is well known that members of the TFGβ family play important roles in regulation of the expression of collagen and laminin. In fact, differentiation, morphogenesis, and dorsoventral patterning during embryo development are regulated by members of the TGFβ family. Ligation of TGFβ receptors initiate signaling through the SMAD protein family, which can ultimately translocate to the nucleus where they participate in transcriptional regulation by direct DNA-binding events or facilitating formation of transcriptionally active complexes. Thus, developmentally regulated expression patterns and concentration gradient of these growth factors can have a profound impact on the molecular composition of the ECM during physiological development. In turn, the compositional changes occurring during development could provide different binding sites for integrin receptors, other growth factors, matricellular proteins, proteoglycans, and matrix-degrading enzymes. These changes could function to alter signaling pathways, leading to differential regulation of a number of biological responses. It is important to point out that these interconnected events leading to ECM compositional changes can occur not only during normal physiological processes but also play significant roles during pathological events. For example, significant changes in ECM composition occur during inflammatory reaction, pathological angiogenesis, and tumor development. In fact, a reduction in fibronectin expression during tumor development has been observed, while an increase in fibrinogen/fibrin has been documented in association with some tumors. Moreover, studies have shown an increase in vitronectin synthesis in association with glioblastoma, in which the microenvironment is typically characterized with relatively low levels of structural ECM proteins. Finally, another well-documented example of ECM compositional changes involves the matricellular proteins thrombospondin and SPARC. While these proteins are not generally considered structural components, they are upregulated during wounding responses. These matricellular proteins bind components of the ECM and modulate integrin responses such as adhesion, migration, and proliferation.

As has been appreciated for decades, growth factor signaling can influence a

wide array of molecules that modify the ECM, including integrin receptors and proteolytic enzymes. Two major classes of proteolytic enzymes known to modify the structural integrity of the ECM include the serine and matrix metalloproteinase (MMP) families. These classes of enzymes can be regulated at multiple levels, including regulation of transcription, translation, and post translation events, including conformational activation, secretion, glycation, dimerization, and inactivation by inhibitors. Members of the serine and metalloproteinase families are secreted in a latent form and are subsequently activated by a number of mechanisms. Members of the serine protease family such as uPA, tPA, and plasmin can cleave fibronectin, fibrinogen, fibrin, and a number of proteoglycans. The metalloproteinase family, of which at least 24 members have been defined so far, can modify nearly all the components of the ECM, including triple helical collagen, denatured collagen (gelatins), laminin, fibronectin, and proteoglycans. As with most complex biological systems, there exist interconnected regulatory activities between the fibrinolytic cascade and the metalloproteinase system. For example, members of the serine protease family such as plasmin may function to activate latent procollagenase, converting it into an active enzyme. Moreover, certain MMPs such as MMP-9 may degrade serpin protease inhibitors. Interestingly, active proteases from both the fibrinolytic and metalloproteinase systems not only cleave structural components of the ECM but also activate certain growth factors and growth factor receptors, release matrix sequestered molecules, modify proteoglycans, and bind integrins.

Notably, proteolytic enzymes impact the structure and integrity of the ECM in several ways. First, and perhaps the most obvious way, is the degradation of the ECM. Both serine proteases and MMPs degrade components of both the BM and the interstitial matrix. Degradation of major ECM proteins can lead to shifts in molecular composition, alterations in integrin, proteoglycan, and other interaction sites, leading to architectural disorganization. In addition, alterations in ECM proteins can lead to modified integrin signaling, thereby altering gene expression. Second, proteolytic enzymes may not only remove restrictive matrix barriers but, in addition, may also selectively facilitate small conformational changes exposing cryptic regions of ECM proteins that modulate cellular functions. In fact, recent work has lead to the development of Mabs that specifically react with denatured or proteolyzed collagen. Importantly, small alterations in the three-dimensional structure of proteins may impact their susceptibility to biochemical modifications such as glycation, cross-linking or sensitivity to proteolysis. Proteolytic modification of the ECM may also lead to the release of matrix-bound growth factors, protease inhibitors, or other regulatory molecules that are sequestered physically within the matrix. Finally, proteolytic activity could release biologically active fragments of larger ECM proteins that, in turn, might impact cellular responses by binding to the cell surface.

Besides protease-mediated modification of ECM proteins, posttranslational changes of matrix components can also occur. For example, conditions associated with hyperglycemia can lead to the formation of advanced glycation end products or AGEs. The formation of AGEs is associated with nonenzymatic glycation of proteins. Specifically, high glucose levels lead to glucose interacting with free amino groups, forming an unstable shift base. These

functional groups undergo molecular rearrangements, resulting in the formation of Amedori products that ultimately result in the formation of AGEs. The glycation processes have been shown to modify a wide variety of ECM proteins, including distinct forms of collagen, laminin, vitronectin, and fibronectin. Studies have suggested the existence of specific cell surface receptors for these AGEs, which may regulate distinct processes such as chemotaxis, proliferation apoptosis, and angiogenesis. Thus, it is likely that glycation of ECM proteins modulate their functions. In fact, glycation alters their ability to polymerize and associate with other ECM proteins as well as disrupt their capacity to facilitate cell adhesion. In this regard, studies have demonstrated that glycation of the integrin recognition sequence RGD in collagen and fibronectin altered its capacity to promote cell adhesion and migration. In still further studies, glycation of ECM proteins was shown to disrupt processes such matrix assembly, MMP activation, integrin recognition, and signal transduction. Thus, it is clear that glycation of ECM proteins could modify both the structural and functional properties of matrix components.

Interestingly, recent experimental evidence has been provided that both ionizing and nonionizing radiation can cause structural alterations of ECM proteins. For example, it is known that UV irradiation can cause structural cross-linking of proteins such as collagen. Aberrant cross-linking could alter proteolytic and thermal stability, induce changes in protein–protein interactions, and modify integrin recognition. All these structural changes could lead to significant modification of biological responses. In addition, studies have demonstrated that ionizing radiation can cause changes in ECM protein production, thus altering the composition and integrity of the ECM. It was recently shown that a dose of ionizing radiation as small as 20cGy could result in significant loss of the exposure of the HUIV26 cryptic epitope within collagen type IV, both *in vitro* and *in vivo*. While functional expression of the HUIV26 cryptic epitope was reduced by approximately 50%, the exposure of a second cryptic epitope defined by the monoclonal antibody HU177 exhibited little change. These interesting results suggest that small doses of ionizing radiation may have a differential and significant impact on the structure and function of specific ECM proteins. These studies provide a few interesting examples of the potential mechanism by which the ECM can be altered.

6
Modulation of Cellular Behavior by Intact and Structurally Altered ECM Components

The ECM is known to regulate an array of molecular, cellular, and biochemical processes. Crucial to this regulation is the ability of cells to sense and, in turn, respond to their surrounding ECM. Integrin receptors are the major group of molecules that mediate bidirectional communication with the ECM. At least 24 different integrin heterodimers have been characterized with distinct and overlapping binding specificities. By virtue of their ability to interact with an array of cytosolic proteins, integrins are known to transmit signals from outside the cell to the inside. Biochemical changes inside the cell can also be transmitted to the outside by integrins, facilitating bidirectional communication.

Given the diversity of potential integrin signaling events, it would be well beyond the scope of our discussion to analyze them

all in detail. Thus, we will limit our description to a few general concepts crucial for a basic appreciation of integrin signaling and its importance in regulating biological responses to the ECM. Integrin recognition of ECM ligands depends on a variety of molecular parameters, including α and β subunit composition, cation binding, substrate composition, and amino acid sequences. In general, integrin interactions can lead to subunit conformational changes and multimerization. In turn, these conformational changes can lead to the association of a number of cytosolic proteins such as α-actinin and talin with integrin cytoplasmic tails. These initial protein–protein interactions lead to the recruitment of sets of adaptor molecules and kinases, leading to formation of signaling complexes. In addition to direct integrin signaling cascades, integrins can cooperate and cross talk with growth factors and growth factor receptors, thus enhancing the level of complexity and potential mechanisms controlled by integrins.

Some important examples of signaling cascade regulated by integrin–ECM ligation include the Ras-Raf-Mek pathways, the PI3K/AKT pathway, and the JNK pathway. Interestingly, integrin ligation of fibronectin has been shown to regulate apoptosis by initiating specific signaling cascades that modified the expression and activation states of effector molecules such as Bcl-2, Bax, and a number of caspases. Moreover, studies have also shown that integrin–ECM interactions regulate expression of molecules involved in cell cycle control such as cyclins and cyclin-dependent kinases. In addition, studies have documented that integrin signaling controls actin cytoskeleton reorganization and formation of focal adhesions that are involved in the control of cellular migration and invasion. Finally, integrins regulate the expression and activation of matrix altering proteases such as MMPs. Thus, it is clear from the variety of interconnected signaling cascades impacted by integrin binding that the ECM can control a diversity of biological responses.

The diversity of responses to the ECM from a given cell can be altered, depending on the structural integrity of the matrix. For example, it is known that proteolytic degradation of epithelial basement membranes can lead to induction of apoptosis in epithelial cells. MMP-mediated cleavage of interstitial collagen type I promotes melanoma cell survival and suppresses apoptosis by regulating Bcl-2 and Bax expression. Structural changes in ECM proteins can also alter other cellular processes such as migration. For example, studies have shown the MMP-mediated cleavage of laminin-5 resulted in the exposure of a cryptic epitope that promoted breast tumor cell migration. Moreover, MMP-9-dependent cleavage of collagen type IV can expose a cryptic epitope that regulates the migration of retinal endothelial cells. Finally, studies have suggested that MMP-2-mediated cleavage of fibronectin can lead to the exposure of a unique cryptic site that possesses antiadhesive properties.

A wide array of proteolytic fragments of collagen has been shown to modulate distinct biological responses. For example, proteolytic release of the endostatin molecule from the NC domain of collagen type XVIII was shown to inhibit endothelial cell proliferation and migration *in vivo*. Fragments of other NC1 domains from collagen type IV (α1NC1, α2NC1, α3NC1, and α6NC1) have all been shown to modulate cellular responses, including adhesion, migration, invasion, and proliferation. The ability of these collagen fragments to alter cellular events may be due, in part, to their capacity to bind integrin receptors.

Besides specific modules of collagen, proteolytic fragments of other ECM proteins have also exhibited modulatory effects on cells. Structurally altered fibronectin has been shown to modulate matrix assembly by exposing cryptic assembly sites. Fragments of laminin, elastin, SPARC, thrombospondin, and fibrinogen have all been shown to regulate processes such as adhesion, proliferation, motility, and protease production. Thus, cellular interactions with both native as well as structurally altered ECM components play major roles in regulating cellular behavior.

7 Control of Physiological and Pathological Processes by Interactions with the ECM

The rapid growth of gene-targeting techniques has lead to a greater understanding of the functional importance of the ECM in regulating complex sets of biological responses. In fact, greater than 40 mutant and/or ECM knockout mice and 26 integrin null mice have been developed. Studies utilizing these mice have provided unique insight into the interconnected roles of integrins and ECM proteins in the bidirectional flow of information between the ECM and cells. Owing to the number and diversity of studies using these genetic approaches, we have chosen to discuss only a few examples to illustrate the importance of the ECM in regulating biological responses.

A variety of transgenic and mutant mice have been developed that harbor alterations within collagen genes. Interestingly, studies involving mutations or gene knock outs of the $\alpha 1$ chains of collagen types I, II, and III exhibited abnormal blood vessel development, vessel rupture, skin defects, and chondrodysplasia. Mutations have also been introduced into genes encoding collagen types IV, V, VI, VII, IX, X, and XI. The mutant phenotypes of these mice vary from renal failure, skin, skeletal, and cartilage defects to hearing loss. Interestingly, mutations in the $\alpha 1$, $\alpha 2$, and $\alpha 3$ chains of the collagen-binding integrins have exhibited some of the same defects, including skin blistering, renal defects, as well as reduced tumor associated angiogenesis. Thus, alterations in expression of collagen or specific integrins clearly impact crucial biological events.

In addition to collagen, the functional roles of other ECM glycoproteins have been studied using gene-targeting techniques. Mice harboring mutations in the gene coding for fibronectin exhibited abnormal development and defects such as improper somite formation. Similar defects were also observed in mice defective in the fibronectin receptor $\alpha 5\beta 1$, including abnormal vascular and somite development. Mice carrying mutations in specific chains of laminin exhibited a variety of defects, including abnormal BM assembly as well as renal, neuromuscular, central nervous system, and vascular abnormalities. Interestingly, mice defective in laminin-binding integrins ($\alpha 3$, $\alpha 6$, and $\beta 4$ chains) showed similar sets of abnormalities, including renal defects and skin blistering and epithelial malformations. In further studies, mice harboring a mutation in perlecan exhibited defects in heart and brain development as well as chondrodysplasia. Finally, studies with thrombospondin-2 mutant mice suggested an important role for TSP-2 in control of vascular development and collagen matrix assembly.

During early embryonic development, many of the major ECM components are being produced. In fact, studies have shown that by the blastocyst stage, ECM components such as laminin, distinct

isoforms of collagen, fibronectin, and perlecan are already present. Mice lacking the laminin 1γ chain failed to develop proper BM structures and subsequently died at E5.5. This lack of a functional BM was thought to contribute to the failure of endodermal cells to differentiate, ultimately leading to extensive apoptosis and embryo death. During gastrulation, a variety of regulatory molecules and morphogens help establish specific gradients that are critical for pattern formation. Mice deficient in fibronectin exhibit defects in neural tube closure and cardiovascular abnormalities. A critical role for cellular communication with fibronectin is evident since mice lacking $\alpha 5\beta 1$ show similar cardiovascular defects.

Members of the TGF-β family such as BMP-4 are thought to play important roles in early pattern formation through signal transduction cascades mediated by activation of the SMAD family members. Inhibitors of BMP signaling, such as Chordin, contain CR-1 domains. CR-1 domains are thought to facilitate binding of these inhibitors to BMP, thereby blocking receptor binding. Interestingly, these CR-1 domains have been found to be present within a number of proforms of collagen, including procollagen type I II, III, and IV. During the maturation of collagen, the prodomains are enzymatically released. Thus, the CR-1-containing domains may bind to BMP, thereby serving as a regulator of embryonic patterning.

Recent studies have also provided convincing data that invasive and pathological events, such as angiogenesis, tumor growth, and metastasis, are tightly controlled by cellular interaction with the ECM. As with most complex physiological events, angiogenesis or the development of blood vessels from preexisting vessels require the precisely coordinated activities of numerous families of regulatory molecules. Some well-characterized examples of these regulatory proteins include growth factors and their receptors, cell adhesion receptors, proteolytic enzymes, and ECM components. It is interesting to point out that interactions with the ECM can regulate the production, activation, and biodistribution of all these molecular regulator of angiogenesis. In addition, studies have demonstrated that endothelial cells secrete and, in turn, interact with specific components of the ECM such as collagen laminin and fibronectin. In fact, endothelial cell interactions with collagen and laminin were required to initiate morphological reorganization of the endothelial cells into tubelike structures. More recent work has demonstrated that mice harboring mutations in genes coding for specific isoforms of collagen exhibit abnormal vasculature, resulting in embryo death. Moreover, several studies from transgenic mice harboring defects in collagen binding integrins also exhibited defective vascular development. Interestingly, studies have shown that administration of function-blocking antibodies directed to collagen binding integrins such as $\alpha 2\beta 1$ and $\alpha 1\beta 1$ potently blocked angiogenesis *in vivo*. One of the most well-characterized adhesion receptors known to regulate angiogenesis is the $\alpha v\beta 3$ integrin. Importantly, $\alpha v\beta 3$ is highly upregulated on the surface of activated angiogenic blood vessels, while little is expressed on normal quiescent vessels. Antagonists of $\alpha v\beta 3$, including functional blocking antibodies and peptides, have been shown to inhibit angiogenesis in multiple models. In fact, a humanized monoclonal antibody termed *vitaxin* is now being evaluated in human clinical trials for cancer therapy.

The molecular mechanisms by which blocking integrins inhibit angiogenesis are numerous. For example, blocking integrin interactions with ECM components can disrupt cellular adhesion and migration that are known to be required for specific steps during angiogenesis. Besides blocking physical interactions with ECM, disruption of integrin signaling events are also likely to contribute to the inhibitory activity. Integrin ligation of distinct domains within fibronectin differentially regulated the expression of matrix-degrading proteases such as MMPs. In fact, mice lacking either MMP-2 or MMP-9 have been shown to exhibit reduced angiogenesis *in vivo*. Further studies have shown that specific integrins such as $\alpha v\beta 3$ can contribute to the localization of proteolytically active forms of MMP-2 to the surface of invasive cells, thus modulating ECM remodeling. Modulation of protease-mediated ECM remodeling can impact angiogenesis in several ways. In particular, proteolytic release of NC domains of collagen type XVIII (endostatin), collagen type XV (restin), collagen type VIII (vastatin), and various NC domains from collagen type IV (arrestin, tumstatin, and canstatin) have all been shown to inhibit angiogenesis. A wide array of other bioactive ECM fragments have also been shown to either inhibit or induce angiogenesis, including fragments from laminin, fibronectin, fibrinogen, and thrombospondin, to name just a few. Finally, recent studies suggest that proteolytic remodeling of the ECM can expose specific matrix-immobilized cryptic sites. In fact, cellular interactions with the HUIV26 cryptic epitope are mediated by the integrin $\alpha v\beta 3$. A function-blocking monoclonal antibody directed to the HUIV26 cryptic epitope was shown to potently inhibit angiogenesis in several models. Finally, integrin ligations of ECM components are also known to regulate the production of angiogenesis, inducing growth factors such as VEGF and bFGF. Thus, cellular interactions with the ECM can have profound impact on new blood vessel development.

Recent studies have provided evidence that many of the same cellular and molecular processes that regulate angiogenesis also play important roles in tumor growth and metastasis. The concept that modulating ECM interactions with tumor cells can alter their malignant behavior has been extensively evaluated. In fact, blocking tumor cell interactions with the ECM can inhibit tumor cell proliferation, invasion, and metastasis. Thus, while contact independent cell proliferation can be a characteristic of malignant tumor cells, ECM contact is still of great importance.

Interestingly, studies suggest that loss on fibronectin expression correlates with enhanced tumor growth. These studies along with many others provided some of the early evidence that modulation of ECM contacts could influence malignant transformation and the spread of tumor cells to distant sites. For example, a number of $\beta 1$-containing integrins have been shown to play important roles in tumor growth as well as metastasis in multiple murine tumor models. In this regard, transfection of $\alpha 2$ integrin subunit allowing functional expression of $\alpha 2\beta 1$ resulted in enhanced rhabdosarcoma tumor metastasis *in vivo*. Besides $\beta 1$ integrins, a wealth of evidence suggests that many other integrins contribute to tumor growth and progression, including $\alpha 6\beta 4$ and $\alpha v\beta 5$. In studies with human melanoma biopsies, it was shown that nonmalignant skin lesions and radial growth phase melanoma had little, if any, $\alpha v\beta 3$ expression, while vertical growth phase melanoma and metastatic lesions

expressed elevated levels of $\alpha v\beta 3$. Importantly, studies have utilized a number of $\alpha v\beta 3$ antagonists to block tumor growth, including function-blocking Mabs, small peptides, nonpeptidic mimetics, and antisense approaches. The mechanisms by which blocking $\alpha v\beta 3$ function inhibits tumor growth and progression are likely to be many, and perhaps similar to those previously discussed in the context of angiogenesis. These mechanisms might include disruption of physical contact with the ECM, thereby altering traction necessary for tumor cell movement and disruption of signaling cascades. Other potential mechanisms might include altering protease production and growth factor secretion. In fact, it is likely that a combination of many of these mechanisms contribute to the inhibitory activity observed with antagonists of $\alpha v\beta 3$ integrin. Thus, it is clear from the wealth of experimental evidence that tumor cell interactions with the ECM play an important role in regulating many aspects of tumor growth and progression.

8 Conclusions

As can be appreciated from our discussion above, the development of new molecular tools, imaging techniques, as well as gene-targeting procedures have lead to unique insight into the complex structural features of the ECM and how changes within this microenvironment can dramatically alter a plethora of biological events. In fact, over the last decade, investigators have become keenly aware of the great importance of the molecular information contained within the ECM and the critical requirement of transferring this information to cells in order to regulate both physiological as well as pathological processes. Our expanding knowledge of integrin biology has provided important mechanistic understanding of the bidirectional flow of information between cells and their local microenvironment. This insight, in conjunction with the emerging concept that a substantial amount of information is functionally hidden within ECM and only exposed following structural remodeling, has opened up an interesting "Cryptic" microenvironment in which to study. This cryptic ECM environment may provide unique molecular insight and provide potential therapeutic targets for the control of genetic defects, angiogenesis, tumor growth, and metastasis. Thus, future studies into the dynamic and fascinating nature of the acellular compartment of multicellular organisms will likely fuel our quest for a more complete understanding of the impact that the ECM has on biological responses.

Acknowledgments

This work was supported in part from NIH grant ROI-CA91645 and a grant from BioStratum Inc., to PCB.

Bibliography

Books and Reviews

Adams, J.C. (2001) Thrombospondin: multifunctional regulators of cell interactions, *Annu. Rev. Cell Dev. Biol.* **17**, 25–51.

Collen, D. (2001) Role of the plasminogen system in fibrin-homeostasis and tissue remodeling, *Hematology (Am. Soc. Hematol. Educ. Program)*. **1**, 1–9.

Iozzo, R.V. (1998) Matrix proteoglycans: from molecular design to cellular function, *Annu. Rev. Biochem.* **67**, 609–652.

Roberts, A.B., Caestecker, M.P., Lechleider, R.J. (2000) Signaling from TGF-β-receptors, in:

The Molecular Basis of Cancer and Other Diseases, Humana Press, Totowa, NJ, 39–51.
Simpson-Haidaris, P.J., Rybarczyik, B. (2001) Tumors and fibrinogen. The role of fibrinogen as an extracellular matrix protein, *Ann. N. Y. Acad. Sci.* **93**, 406–425.
Xu, J., Brooks, P.C. (2002) Extracellular matrix in the regulation of angiogenesis, in: *Assembly of the Vasculature and its Regulation*, Birkhauser, Boston.

Primary Literature

Adachi, E., Hopkinson, I., Hayashi, T. (1997) Basement-membrane stromal relationships: interactions between collagen fibrils and the lamina densa, *Int. Rev. Cytol.* **173**, 73–155.
Adams, J.C., Tucker, R.P. (2000) The thrombospondin type-1 repeat (TSR) superfamily: diverse proteins with related roles in neuronal development, *Dev. Dyn.* **218**, 280–299.
Adams, J.C., Watt, F.M. (1993) Regulation of development and differentiation by the extracellular matrix, *Development* **117**, 1183–1198.
Agah, A., Kyriakides, T.R., Lawler, J., Bornstein, P. (2002) The lack of thrombospondin-1 (TSP1) dictates the course of wound healing in double-TSP1/TSP2-null mice, *Am. J. Pathol.* **161**, 831–839.
Akiyama, S.K., Olden, K., Yamada, K.M. (1995) Fibronectin and integrins in invasion and metastasis, *Cancer Met. Rev.* **14**, 173–189.
Arikawa-Hirasawa, E., Watanabe, H., Takami, H., Hassell, J.R., Yamada, Y. (1999) Perlecan is essential for cartilage and cephalic development, *Nat. Genet.* **23**, 354–358.
Badylak, S.F. (2002) The extracellular matrix as a scaffold for tissue reconstruction, *Semin. Cell Dev. Biol.* **13**, 377–383.
Barnes, D.W., Reing, J.E., Amos, B. (1985) Heparin-binding properties of human serum spreading factor, *J. Biol. Chem.* **260**, 9117–91122.
Belkin, A.M., Stepp, M.A. (2000) Integrins as receptors for laminins, *Micr. Res. Tech.* **51**, 280–301.
Berditchevski, F., Zuttter, M.M., Hemler, M.E. (1996) Characterization of novel complexes on the cell surface between integrins and proteins with 4 transmembrane domains TM4 proteins), *Mol. Biol. Cell.* **7**, 193–207.
Bernfield, M., Gotte, M., Park, P.W., Reizes, O., Fitzgerald, M.L., Lincecum, J., Zako, M. (1999) Functions of cell surface heparan sulfate proteoglycans, *Annu. Rev. Biochem.* **68**, 729–777.
Bevilacqua, M.P., Nelson, R.M. (1993) Selectins, *J. Clin. Invest.* **91**, 379–387.
Birkedal-Hansen, H. (1995) Proteolytic remodeling of extracellular matrix, *Curr. Opin. Cell Biol.* **7**, 728–735.
Bokel, C., Brown, N.H. (2002) Integrins in development: moving on, responding to, and sticking to the extracellular matrix, *Dev. Cell.* **3**, 311–321.
Bootle-Wilbraham, C.A., Tazzyman, S., Marshall, J.M., Lewis, C.E. (2000) Fibrinogen E-fragment inhibits the migration and tubule formation of human dermal microvascular endothelial cells in vitro, *Cancer Res.* **60**, 4719–4724.
Bornstein, P., Sage, E.H. (1994) Thrombospondins, *Methods Enzymol.* **245**, 62–85.
Bornstein, P., Sage, H.E. (2002) Matricellular proteins: extracellular modulators of cell function, *Curr. Opin. Cell Biol.* **14**, 608–616.
Brassart, B., Fuchs, P., Huet, E., Alix, A.J., Wallach, J., Tamburro, A.M., Delacoux, F., Haye, B., Emonard, H., Hornebeck, W., Debelle, L. (2001) Conformational dependence of collagenase (matrix metalloproteinase-1) up-regulation by elastin peptides in cultured fibroblasts, *J. Biol. Chem.* **276**, 5222–5227.
Brightman, A.O., Rajwa, B.P., Sturgis, J.E., McCallister, M.E., Robinson, J.P., Voytik-Harbin, S.L. (2000) Time-lapse confocal reflection microscopy of collagen fibrillogenesis and extracellular matrix assembly in vitro, *Biopolymers* **54**, 222–234.
Brooks, P.C. (1996) Role of integrins in angiogenesis, *Eur. J. Cancer* **32A**, 2423–2429.
Brooks, P.C. (1997) Integrin avb3: A therapeutic target, *DN&P* **10**, 456–461.
Brooks, P.C., Clark, R.A., Cheresh, D.A. (1994) Requirement of vascular integrin alpha $\alpha v\beta 3$ for angiogenesis, *Science* **264**, 569–571.
Brooks, P.C., Klemke, R.L., Schon, S., Lewis, J.M., Schwartz, M.A., Cheresh, D.A. (1997) Insulin-like growth factor receptor cooperates with integrin $\alpha v\beta 5$ to promote tumor cell dissemination in vivo, *J. Clin. Invest.* **99**, 1390–1398.
Brooks, P.C., Roth, J.M., Lymberis, S.C., DeWyngaert, K., Broek, D., Formenti, S.C. (2002) Ionizing radiation modulates the exposure of the HUIV26 cryptic epitope within collagen

type-IV during angiogenesis, *Int. J. Rad. Oncol. Biol. Phys.* **54**, 1194–1201.

Brooks, P.C., Stromblad, S., Klemke, R., Visscher, D., Sarkar, F.H., Cheresh, D.A. (1995) Anti-integrin $\alpha v\beta 3$ blocks human breast cancer growth and angiogenesis in human skin, *J. Clin. Invest.* **96**, 1815–1822.

Brooks, P.C., Montgomery, A.M., Rosenfeld, M., Reisfeld, R.A., Hu, T., Klier, G., Cheresh, D.A. (1994) Integrin $\alpha v\beta 3$ antagonists promote tumor regression by inducing apoptosis of angiogenic blood vessels, *Cell* **79**, 1157–1164.

Brooks, P.C., Stromblad, S., Sanders, L.C., von Schalscha, T.L., Aimes, R.T., Stetler-Stevenson, W.G., Quigley, J.P., Cheresh, D.A. (1996) Localization of matrix metalloproteinase MMP-2 to the surface of invasive cells by interaction with integrin $\alpha v\beta 3$, *Cell* **85**, 683–693.

Brown, J.C., Sasaki, T., Gohring, W., Yamada, Y., Timpl, R. (1997) The C-terminal domain V of perlecan promotes beta1 integrin-mediated cell adhesion, binds heparin, nidogen and fibulin-2 and can be modified by glycosaminoglycans, *Eur. J. Biochem.* **250**, 39–46.

Casaroli Marano, R.P., Preissner, K.T., Vilaro, S. (1995) Fibronectin, laminin, vitronectin and their receptors at newly-formed capillaries in proliferative diabetic retinopathy, *Exp. Eye Res.* **60**, 5–17.

Chan, B.M., Matsuura, N., Takada, Y., Zetter, B.R., Hemler, M.E. (1991) In vitro and in vivo consequences of VLA-2 expression on rhabdomysarcoma cells, *Science* **251**, 1600–1602.

Chang, C., Werb, Z. (2001) The many faces of metalloproteinases: cell growth, invasion, angiogenesis and metastasis, *Trends Cell Biol.* **11**, S37–S43.

Charonis, A.S., Tsilbary, E.C. (1992) Structural and functional changes of laminin and type-IV collagen after nonenzymatic glycation, *Diabetes* **2**, 49–51.

Chen, H., Herndon, M.E., Lawler, J. (2000) The cell biology of thrombospondin-1, *Matrix Biol.* **19**, 597–614.

Chung, A.E., Durkin, M.E. (1990) Entactin: structure and function, *Am. J. Respir. Cell Mol. Biol.* **3**, 275–282.

Colognata, H., Yurcheco, P.D. (2000) Form and function: the laminin family of heterotrimers, *Dev. Dyn.* **218**, 213–234.

Colorado, P.C., Torre, A., Kamphaus, G., Maeshima, Y., Hopfer, H., Takahashi, K., Volk, R., Zamborsky, E.D., Herman, S., Sarkar, P.K., Ericksen, M.B., Dhanabal, M., Simons, M., Post, M., Kufe, D.W., Weichselbaum, R.R., Sukhatme, V.P., Kalluri, R. (2000) Anti-angiogenic cues from vascular basement membrane collagen, *Cancer Res.* **60**, 2520–2526.

Cooper, A.R., MacQueen, H.A. (1983) Subunits of laminin are differentially synthesized in mouse eggs and early embryos, *Dev. Biol.* **96**, 467–447.

Cornelissen, M., Thierens, H., Ridder, L. (1996) Effects of ionizing radiation on cartilage: emphasis on effects on the extracellular matrix, *Scanning Microsc.* **10**, 833–840.

Costell, M., Gustafsson, E., Aszodi, A., Morgelin, M., Bloch, W., Hunziker, E., Addicks, K., Timpl, R., Fassler, R. (1999) Perlecan maintains the integrity of cartilage and some basement membranes, *J. Cell Biol.* **147**, 1109–1122.

Dallas, S.L., Rosser, J.L., Mundy, G.R., Bonewald, L.F. (2002) Proteolysis of latent transforming growth factor-beta (TGF-beta) binding protein-1 by osteoclasts. A cellular mechanisms for release of TGF-beta from bone matrix, *J. Biol. Chem.* **14**, 21352–21360.

Damsky, C.H., Ilic, D. (2002) Integrin signaling: its where the action is, *Curr. Opin. Cell Biol.* **114**, 594–602.

Damsky, C.H., Werb, Z. (1992) Signal transduction by integrin receptors for extracellular matrix: cooperative processing of extracellular information, *Curr. Opin. Cell Biol.* **4**, 772–781.

Davis, G.E. (1992) Affinity of integrins for damaged extracellular matrix: alpha v beta 3 binds to denatured collagen type-I through RGD sites, *Biochem. Biophys. Res. Commun.* **182**, 1025–1031.

Davis, G.E., Bayless, K.J., Davis, M.J., Meininger, G.A. (2000) Regulation of tissue injury responses by the exposure of matricryptic sites within extracellular matrix molecules, *Am. J. Pathol.* **156**, 1489–1498.

De Arcangelis, A., Mark, M., Kreidberg, J., Sorokin, L., Georges-Labouesse, E. (1999) Synergistic activities of alpha 3 and alpha 6 integrins are required during apical ectodermal ridge formation and organogenesis in the mouse, *Development* **126**, 3957–3968.

Di Lullo, G.A., Sweeney, S.M., Korkko, J., Ala-Kokko, L., San Antonio, J.D. (2002) Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type-I collagen, *J. Biol. Chem.* **277**, 4223–4231.

Dogic, D., Eckes, B., Aumailley, M. (1999) Extracellular matrix, integrins and focal adhesions, *Curr. Top. Pathol.* **93**, 75–85.

Dumin, J., Dickeson, K.S., Stricker, T.P., Bhattacharyya-Pakrasi, M., Roy. Santoro, S.A., Parks, W.C. (2001) Pro-collagenase-1 (Matrix Metalloproteinase-1) binds the $\alpha 2\beta 1$ integrin upon release from keratinocytes migrating on type-I collagen, *J. Biol. Chem.* **276**, 29368–29374.

Duraisamy, Y., Slevin, M., Smith, N., Bailey, J., Zweit, J., Smith, C., Ahmed, N., Gaffney, J. (2001) Effects of glycation on basic fibroblast growth factor induced angiogenesis and activation of associated signal transduction pathways in vascular endothelial cells: possible relevance to wound healing in diabetes, *Angiogenesis* **4**, 277–288.

Durkin, M.E., Wewer, U.M., Chung, A.E. (1995) Exon organization of the mouse entactin gene corresponds to the structural domains of the polypeptide and has regional homology to the low-density lipoprotein receptor gene, *Genomics* **26**, 219–228.

Eliceiri, B.P. (2001) Integrin and growth factor receptor crosstalk, *Circ. Res.* **89**, 1104–1110.

Emsley, J., Knight, C.G., Frandale, R.W., Barnes, M.J., Liddington, R.C. (2000) Structural basis of collagen recognition by integrin $\alpha 2\beta 1$, *Cell* **100**, 47–56.

Erynck, R., Zhang, Y., Feng, X.-H. (1998) Smads: transcriptional activators of TGF-β responses, *Cell* **95**, 737–740.

Exposito, J.-Y., Cluzel, C., Garrone, R., Lethias, C. (2002) Evolution of collagens, *Anat. Rec.* **268**, 302–316.

Fox, J.W., Mayer, U., Nischt, R., Aumailley, M., Reinhardt, D., Wiedemann, H., Mann, K., Timpl, R., Krieg, T., Engel, J. (1991) Recombinant nidogen consists of three globular domains and mediates binding of laminin to collagen type IV, *EMBO J.* **10**, 3137–3146.

French, M.M., Smith, S.E., Akanbi, K., Sanford, T., Hecht, J., Farach-Carson, M.C., Carson, D.D. (1999) Expression of the heparan sulfate proteoglycan, perlecan, during mouse embryogenesis and perlecan chondrogenic activity in vitro, *J. Cell Biol.* **145**, 1103–1115.

Friedlander, M., Brooks, P.C., Shaffer, R.W., Kincaid, C.M., Varner, J.A., Cheresh, D.A. (1995) Definition of two angiogenic pathways by distinct alpha v integrins, *Science* **270**, 1500–1502.

Furuyama, A., Iwata, M., Hayashi, T., Mochitate, K. (1999) Transforming growth factor-beta-1 regulates basement membrane formation by alveolar epithelial cells in vitro, *Eur. J. Cell Biol.* **78**, 867–875.

Garcia Abreu, J., Coffinier, C., Larrain, J., Oelgeschlager, M., De Robertis, E.M. (2002) Chordin-like CR domains and the regulation of evolutionarily conserved extracellular signaling systems, *Gene* **287**, 39–47.

George, E.L., Georges-Labouesse, E.N., Patel-King, R.S., Rayburn, H., Hynes, R.O. (1993) Defects in mesoderm, neural tube and vascular development in mouse embryos lacking fibronectin, *Development* **119**, 1079–1091.

Ghosh, S., Stack, S.M. (2000) Proteolytic modification of laminins: functional consequences, *Micro. Res. Tech.* **51**, 238–246.

Giannelli, G., Falk-Marzillier, J., Schiraldi, O., Stetler-Stevenson, W.G., Quaranta, V. (1997) Induction of cell migration by matrix metalloproteinase-2 cleavage of laminin-5, *Science* **277**, 225–228.

Gladson, C.L., Cheresh, D.A. (1991) Glioblastoma expression of vitronectin and the alpha V beta 3 integrin. Adhesion mechanism for transformed glial cells, *J. Clin. Invest.* **88**, 1924–1932.

Goldsmith, E.C., Borg, T.K. (2002) The dynamic interaction of the extracellular matrix in cardiac remodeling, *J. Card. Fail.* **8**, S314–S318.

Gorden, M.K., Olsen, B.R. (1990) The contribution of collagenous proteins to tissue-specific matrix assemblies, *Curr. Opin. Cell Biol.* **2**, 833–838.

Grumet, M., Milev, P., Sakurai, T., Karthikeyan, L., Bourdon, M., Margolis, R.K., Margolis, R.U. (1994) Interactions with tenascin and differential effects on cell adhesion of neurocan and phosphacan, two major chondroitin sulfate proteoglycans of nervous tissue, *J. Biol. Chem.* **269**, 12142–12146.

Gustafsson, E., Fassler, R. (2000) Insights into extracellular matrix functions from mutant mouse models, *Exp. Cell Res.* **261**, 52–68.

Gutheil, J.C., Campbell, T.N., Pierce, P.R., Watkins, J.D., Huse, W.D., Bodkin, D.J., Cheresh, D.A. (2000) Targeted antiangiogenic

therapy for cancer using Vitaxin: a humanized monoclonal antibody to the integrin $\alpha v\beta 3$, *Clin. Cancer Res.* **6**, 3056–3061.

Hangai, M., Kitaya, N., Xu, J., Chan, C.K., Kim, J.J., Werb, Z., Ryan, S.J., Brooks, P.C. (2002) Matrix metalloproteinase-9-dependent exposure of a cryptic migratory control site in collagen is required before retinal angiogenesis, *Am. J. Pathol.* **161**, 1429–1437.

Hayashi, M., Akama, T., Kono, I., Kashiwagi, H. (1985) Activation of vitronectin (serum spreading factor) binding of heparin by denaturing agents, *J. Biochem.* **98**, 1135–1138.

Hedstrom, L. (2002) Serine protease mechanism and specificity, *Chem. Rev.* **102**, 4501–4524.

Heissig, B., Hattori, K., Friedrich, M., Rafii, S., Werb, Z. (2003) Angiogenesis: vascular remodeling of the extracellular matrix involves metalloproteinases, *Curr. Opin. Hematol.* **10**, 136–141.

Hieken, T.J., Ronan, S.G., Farolan, M., Shilkaitis, A.L., Das Gupta, T.K. (1996) Beta 1 integrin expression: a marker of lymphatic metastases in cutaneous malignant melanoma, *Anticancer Res.* **16**, 2321–2324.

Hohenester, E., Engel, J. (2002) Domain structure and organization in extracellular matrix proteins, *Matrix Biol.* **212**, 115–128.

Holtkotter, O., Nieswandt, B., Smyth, N., Muller, W., Hafneer, M., Schulte, V., Krieg, T., Eckes, B. (2002) Integrin alpha 2-deficient mice develop normally, are fertile, but display partial defective platelet interaction with collagen, *J. Biol. Chem.* **277**, 10789–10794.

Hood, J.D., Bednarski, M., Frausto, R., Guccione, S., Reisfeld, R.A., Cheresh, D.A. (2002) Tumor regression by targeted gene delivery to the neovasculature, *Science* **296**, 2404–2407.

Hunter, D.D., Cashman, N., Morris-Valero, R., Bulock, J.W., Adams, S.P., Sanes, J.R. (1991) An LRE (leucine-arginine-glutamate)-dependent mechanism for adhesion of neurons to S-laminin, *J. Neurosci.* **11**, 3960–3971.

Hutter, H., Vogel, B.E., Plenefisch, J.D., Norris, C.R., Proenca, R.B., Spieth, J., Guo, C., Mastwal, S., Zhu, X., Scheel, J., Hedgecock, E.M. (2000) Conservation and novelty in the evolution of cell adhesion and extracellular matrix genes, *Science* **287**, 989–994.

Hynes, R.O. (2002) Integrins: bidirectional, allosteric signaling machines, *Cell* **110**, 673–687.

Iivanainen, E., Kahari, V.M., Heino, J., Elenius, K. (2003) Endothelial cell-matrix interactions, *Microsc. Res. Tech.* **60**, 13–22.

Inatani, M., Tanihara, H. (2002) Proteoglycans in retina, *Prog. Retin. Eye Res.* **21**, 429–447.

Inatani, M., Tanihara, H., Oohira, A., Honjo, M., Kido, N., Honda, Y. (2000) Spatiotemporal expression patterns of $\alpha 6\beta 4$ proteoglycan/phosphacan in the developing rat retina, *Invest. Ophthalmol. Vis. Sci.* **41**, 1990–1997.

Ingber, D.E. (2002) Mechanical signaling and the cellular response to extracellular matrix in angiogenesis and cardiovascular physiology, *Circ. Res.* **91**, 877–887.

Iozzo, R.V. (2001) Heparan sulfate proteoglycans: intricate molecules with intriguing functions, *J Clin. Invest.* **108**, 165–167.

Iozzo, R.V., Cohen, I.R., Grassel, S., Murdoch, A.D. (1994) The biology of perlecan: the multifaceted heparan sulphate proteoglycan of basement membranes and pericellular matrices, *Biochem. J.* **302**, 625–639.

Itoh, T., Tanioka, M., Yoshida, H., Yoshioka, T., Nishimoto, H., Itohara, S. (1998) Reduced angiogenesis and tumor progression in gelatinase A-deficient mice, *Cancer Res.* **58**, 1048–1051.

Jones, J.C.R., Dehart, G.W., Gonzales, M., Goldfinger, L.E. (2000) Laminins: an overview, *Micro. Res. Tech.* **51**, 211–213.

Jung, Y.D., Ahmad, S.A., Liu, W., Reinmuth, N., Parikh, A., Stoeltzing, O., Fan, F., Ellis, L.M. (2002) The role of the microenvironment and intercellular cross-talk in tumor angiogenesis, *Semin. Cancer Biol.* **12**, 105–112.

Kadler, K.E., Holmes, D.F., Trotter, J.A., Chapman, J.A. (1996) Collagen Fibril formation, *Biochem. J.* **316**, 1–11.

Kadoya, Y., Salmivirta, K., Talts, J.F., Kadoya, K., Mayer, U., Timpl, R., Ekblom, P. (1997) Importance of nidogen binding to laminin gamma-1 for branching epithelial morphogenesis of the submandibular gland, *Development* **124**, 683–691.

Kamphaus, G.D., Colorado, P.C., Panka, D.J., Hopfer, H., Ramchandran, R., Torre, A., Maeshima, Y., Mier, J.W., Sukhatme, V.P., Kalluri, R. (2000) Canstatin, a novel matrix-derived inhibitor of angiogenesis and tumor growth, *J. Biol. Chem.* **275**, 1209–1215.

Karumanchi, S.A., Jha, V., Ramchandran, R., Karihaloo, A., Tsiokas, L., Chan, B., Dhanabal, M., Hanai, J.-I., Venkataraman, G., Shriver, Z., Keiser, N., Kalluri, R., Zeng, H.,

Mukhopadhyay, D., Chen, R.L., Lander, A.D., Hagihara, K., Yamaguchi, Y., Sasisekharan, R., Cantley, L., Sukhatme, V.P. (2001) Cell surface glypicans are low affinity endostatin receptors, *Mol. Cell.* **7**, 811–822.

Kauppila, S., Stenback, F., Kacinski, B.M., Carcangiu, M.-L., Risteli, J., Risteli, L. (1999) Characterization of type-I collagen synthesis and maturation in uterine carcinosarcomas, *Cancer* **86**, 1299–1306.

Kern, A., Marcantonio, E.E. (1998) Role of the I-domain in collagen binding specificity and activation of the integrins $\alpha 1\beta 1$ and $\alpha 2\beta 1$, *J. Cell Physiol.* **176**, 634–641.

Kohfeldt, E., Sasaki, T., Gohring, W., Timpl, R. (1998) Nidogen-2: a new basement membrane protein with diverse binding properties, *J. Mol. Biol.* **282**, 99–109.

Kresse, H., Schonherr, E. (2001) Proteoglycans of the extracellular matrix and growth control, *J. Cell Physiol.* **189**, 2666–2674.

Kumar, C.C., Malkowski, M., Yin, Z., Tanghetti, E., Yaremko, B., Nechuta, T., Varner, J., Liu, M., Smith, E.M., Neustadt, B., Presta, M., Armstrong, L. (2001) Inhibition of angiogenesis and tumor growth by SCH221153, a dual alpha (v) beta (3) and alpha (v) beta (5) integrin receptor antagonist, *Cancer Res.* **61**, 2232–2238.

Kuzuya, M., Asai, T., Kanda, S., Maeda, K., Cheng, X.W., Iguchi, A. (2001) Glycation cross-links inhibits matrix metalloproteinase-2 activation in vascular smooth muscle cells cultured on collagen lattice, *Diabetologia* **44**, 433–436.

Kyriakides, T.R., Zhu, Y.H., Yang, Z., Huynh, G., Bornstein, P. (2001) Altered extracellular matrix remodeling and angiogenesis in sponge granulomas of thrombospondin 2-null mice, *Am. J. Pathol.* **159**, 1255–1262.

Labat-Robert, J. (2002) Fibronectin in malignancy effects of aging, *Semin. Cancer Biol.* **12**, 187–195.

Lafrenie, R.M., Yamada, K.M. (1996) Integrin-dependent signal transduction, *J. Cell. Biochem.* **61**, 545–553.

Lane, T.F., Iruela-Arispe, M.L., Johnson, R.S., Sage, E.H. (1994) SPARC is a source of copper-binding peptides that stimulate angiogenesis, *J. Cell Biol.* **125**, 929–943.

Lee, J.E., Park, J.C., Hwang, Y.S., Kim, J.K., Kim, J.G., Sub, H. (2001) Characterization of UV-irradiated dense/porous collagen membranes: morphology, enzymatic degradation, and mechanical properties, *Yonsei Med. J.* **42**, 172–179.

Lester, B.R., McCarthy, J.B. (1992) Tumor cell adhesion to the extracellular matrix and signal transduction mechanisms implicated in tumor cell motility, invasion and metastasis, *Cancer Metastasis Rev.* **11**, 31–44.

Li, G., Fridman, R., Kim, H.R. (1999) Tissue inhibitor of metalloproteinases-1 inhibits apoptosis of human breast epithelial cells, *Cancer Res.* **59**, 6267–6275.

Li, J., Zhang, Y.P., Kirsner, R.S. (2003) Angiogenesis in wound repair: angiogenic growth factors and the extracellular matrix, *Microsc. Res. Tech.* **60**, 107–114.

Li, X., Chen, Y., Scheele, S., Arman, E., Haffner-Krausz, R., Lonai, P. (2001) Fibroblast growth factor signaling and basement membrane assembly are connected during epithelial morphogenesis of the embryoid body, *J. Cell Biol.* **153**, 811–822.

Li, X., Talts, U., Talts, J.F., Arman, E., Ekblom, P., Lonai, P. (2001) Akt/PKB regulates laminin and collagen IV isotypes of the basement membrane, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14416–14421.

Lichtner, R.B., Howlett, A.R., Lerch, M., Xuan, J.A., Brink, J., Langton-Webster, B., Schneider, M.R. (1998) Negative cooperativity between alpha (3) beta (1) and alpha (2) beta (1) integrins in human mammary carcinoma MDA MB 231 cells, *Exp. Cell. Res.* **240**, 368–376.

Lin, C.Q., Bissell, M.J. (1993) Multi-faceted regulation of cell differentiation by extracellular matrix, *FASEB J.* **7**, 737–743.

Liu, X., Wu, H., Byrne, M., Jeffrey, J., Krane, S., Jaenisch, R. (1995) A targeted mutation at the known collagenase cleavage site in mouse type I collagen impairs tissue remodeling, *J. Cell Biol.* **130**, 227–237.

Liu, Z., Zhou, S.D., Sharipo, J.M., Shipley, S.S., Twining, L.A., Diaz, R.M., Senior, R.M., Werb, Z. (2000) The serpin a1-proteases inhibitor is a critical substrate for gelatinase B/MMP-9 in vivo, *Cell* **102**, 647–655.

Lode, H.N., Moehler, T., Xiang, R., Jonczyk, A., Gillies, S.D., Cheresh, D.A., Reisfeld, R.A. (1999) Synergy between an antiangiogenic integrin alpha v antagonist and an antibody-cytokine fusion protein eradicates spontaneous tumor metastases, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 1591–1596.

Lohler, J., Timpl, R., Jaenisch, R. (1984) Embryonic lethal mutation in mouse collagen I gene causes rupture of blood vessels and is associated with erythropoietic and mesenchymal cell death, *Cell* **38**, 597–607.

Luo, Z.J., King, R.H.M., Lewin, J., Thomas, P.K. (2002) Effects of nonenzymatic glycosylation of extracellular matrix components on cell survival and sensory neurite extension in cell culture, *J. Neurol.* **249**, 424–431.

Maeshima, Y., Manfredi, M., Reimer, C., Holthaus, K.A., Hopfer, H., Chandamuri, B.R., Kharbanda, S., Kalluri, R. (2001) Identification of the anti-angiogenic site within vascular basement membrane-derived tumstatin, *J. Biol. Chem.* **276**, 15240–15248.

Marneros, A.G., Olsen, B.R. (2001) The role of collagen-derived proteolytic fragments in angiogenesis, *Matrix Biol.* **20**, 337–345.

Mayer, U., Zimmermann, K., Mann, K., Reinhardt, D., Timpl, R., Nischt, R. (1995) Binding properties and protease stability of recombinant human nidogen, *Eur. J. Biochem.* **227**, 681–686.

Mayer, U., Nischt, R., Poschl, E., Mann, K., Fukuda, K., Gerl, M., Yamada, K., Timpl, R. (1993) A single EGF-like motif of laminin is responsible for high affinity nidogen binding, *EMBO J.* **12**, 1879–1885.

McCarthy, A.D., Etcheverry, S.B., Bruzzone, L., Lettier, G., Barrio, D.A., Cortizo, A.M. (2001) Non-enzymatic glycosylation of type-I collagen matrix: effects on osteoblastic development and oxidative stress, *BMC Cell Biol.* **2**, 16–26.

Mercurio, A.M., Rabinovitz, I. (2001) Towards a mechanistic understanding of tumor invasion: lesions from the $\alpha 6\beta 4$ integrin, *Semin. Cancer Biol.* **11**, 129–141.

Mitjans, F., Sander, D., Adan, J., Sutter, A., Martinez, J.M., Jaggle, C.S., Moyano, J.M., Kreysch, H.G., Piulats, J., Goodman, S.L. (1995) An anti-alpha v-integrin antibody that blocks integrin function inhibits the development of a human melanoma in nude mice, *J. Cell Sci.* **108**, 2825–2838.

Mongiat, M., Sweeney, S.M., San Antonio, J.D., Fu, J., Iozzo, R.V. (2003) Endorepellin, a novel inhibitor of angiogenesis derived from the C-terminus of perlecan, *J. Biol. Chem.* **278**, 4238–4249.

Mosher, D.F., Sottile, J., Wu, C., McDonald, J.A. (1992) Assembly of the extracellular matrix, *Curr. Opin. Cell Biol.* **4**, 810–818.

Nakamura, T., Sato, K., Hamada, H. (2002) Effective gene transfer to human melanomas via integrin-targeted adenoviral vectors, *Hum. Gene Ther.* **13**, 613–626.

Nakashimqa, J., Sumitomo, M., Miyajima, A., Jitsukawa, S., Saito, S., Tachibana, M., Murai, M. (1997) The value of serum carboxyterminal propeptide of type-I procollagen in predicting bone metastasis in prostate cancer, *J. Urol.* **157**, 1736–1739.

Nam, T., Moralez, A., Clemmons, D. (2002) Vitronectin binding to IGF binding protein-5 (IGFBP-5) alters IGFBP-5 modulation of IGF-I actions, *Endocrinology* **143**, 30–36.

Noonan, D.M., Fulle, A., Valente, P., Cai, S., Horigan, E., Sasaki, M., Yamada, Y., Hassell, J.R. (1991) The complete sequence of perlecan, a basement membrane heparan sulfate proteoglycan, reveals extensive similarity with laminin A chain, low density lipoprotein receptor, and the neural cell adhesion molecule, *J. Biol. Chem.* **266**, 22939–22947.

O'Reilly, M., Boehm, T., Shing, Y., Fukai, N., Vasies, G., Lane, W.S., Flynn, E., Birkhead, J.R., Olsen, B.R., Folkman, J. (1997) Endostatin: an endogenous inhibitor of angiogenesis and tumor growth, *Cell* **88**, 277–285.

Olsen, B.J. (1995) New insight into the function of collagens from genetic analysis, *Curr. Opin. Cell Biol.* **7**, 720–727.

Ornitz, D.M. (2000) FgFs, heparan sulfate and FGFRs: complex interactions essential for development, *Bioessays* **22**, 108–112.

Ortega, N., Werb, Z. (2002) New functional roles for non-collagenous domains of basement membrane collagen, *J. Cell Sci.* **115**, 4201–4214.

Ottani, V., Raspanti, M., Ruggeri, A. (2001) Collagen structure and functional implications, *Micron* **32**, 251–260.

Palmieri, D., Camardellqa, L., Ulivi, V., Guasco, G., Manduca, P. (2000) Trimer carboxyl propeptide of collagen type-I produced by mature osteoblasts is chemotactic for endothelial cells, *J. Biol. Chem.* **275**, 32658–32663.

Pankov, R., Yamada, K.M. (2002) Fibronectin at a glance, *J. Cell Sci.* **115**, 3861–3863.

Parise, L.V., Lee, J.W., Juliano, R.L. (2000) New aspects of integrin signaling in cancer, *Semin. Cancer Biol.* **10**, 407–414.

Pepper, M.S. (2001) Role of the matrix metalloproteinase and plasminogen activator-plasmin system in angiogenesis, *Arterioscler. Thromb. Vasc. Biol.* **21**, 1104–1117.

Perez-Moreno, M., Jamora, C., Fuchs, E. (2003) Stick business: orchestrating cellular signals at adherence junctions, *Cell* **112**, 535–548.

Petitclerc, E., Stromblad, S., von Schalscha, T.L., Mitjans, F., Piulats, J., Mongomery, A.M.P., Cheresh, D.A., Brooks, P.C. (1999) Integrin $\alpha v\beta 3$ promotes M21 melanoma growth in human skin by regulating tumor cell survival, *Cancer Res.* **59**, 2724–2730.

Petitclerc, E., Boutaud, A., Prestakyo, A., Xu, J., Sado, Y., Ninomiya, Y., Sarras, M.P., Hudson, B.G., Brooks, P.C. (2000) New Functions for non-collagenous domains of human collagen type-IV: novel integrin ligands inhibiting angiogenesis and tumor growth in vivo, *J. Biol. Chem.* **275**, 8051–8061.

Ponce, M.L., Nomizu, M., Kleinman, H.K. (2001) An angiogenic laminin site and its antagonist bind through the alpha(v)beta(3) and alpha(5) beta(1) integrins, *FASEB J.* **15**, 1389–1397.

Prockop, D.J., Kivirikko, K.I. (1995) Collagen: molecular biology, disease, and potentials for therapy, *Annu. Rev. Biochem.* **64**, 403–434.

Pupa, S.M., Menard, S., Forti, S., Tagliabue, E. (2002) New insights into the role of extracellular matrix during tumor onset and progression, *J. Cell Physiol.* **192**, 259–267.

Qian, F., Vaux, D.L., Weissman, I.L. (1994) Expression of the integrin alpha (4) beta (1) on melanoma cells can inhibit the invasive stage of metastasis formation, *Cell* **77**, 335–347.

Ramchandra, R., Dhanabal, M., Volk, R., Waterman, M.J.F., Segal, M., Lu, H., Knebelmann, B., Sukhatme, V.P. (1999) Antiangiogenic activity of restin, NC10 domain of human collagen XV: comparison to endostatin, *Biochem. Biophys. Res. Commun.* **255**, 735–739.

Ries, A., Gohring, W., Fox, J.W., Timpl, R., Sasaki, T. (2001) Recombinant domains of mouse nidogen-1 and their binding to basement membrane proteins and monoclonal antibodies, *Eur. J. Biochem.* **268**, 5119–5128.

Ross, J.J., Shimmi, O., Vilmos, P., Petryk, A., Kim, H., Gaudenz, K., Hermanson, S., Ekker, S.C., O'Conner, M.B., Marsh, J.L. (2001) Twisted gastrulation is a conserved extracellular BMP antagonist, *Nature* **410**, 479–483.

Salmivirta, K., Talts, J.F., Olson, M., Sasaki, T., Timpl, R., Ekblom, P. (2002) Binding of mouse nidogen-2 to basement membrane components and cells and its expression in embryonic and adult tissues suggest complementary functions of the two nidogens, *Exp. Cell Res.* **279**, 188–201.

Sasaki, M., Kleinman, H.K., Huber, H., Deutzmann, R., Yamada, Y. (1988) Laminin, a multidomain protein. The A chain has a unique globular domain and homology with the basement membrane proteoglycan and the laminin B chains, *J. Biol. Chem.* **263**, 16536–16544.

Sasaki, T., Gohring, W., Miosge, N., Abrams, W.R., Rosenbloom, J., Timpl, R. (1999) Tropoelastin binding to fibulins, nidogen-2 and other extracellular matrix proteins, *FEBS Lett.* **460**, 280–284.

Schittny, J.C., Yurchenco, P.D. (1989) Basement membranes: molecular organization and function in development and disease, *Curr. Opin. Cell Biol.* **1**, 983–988.

Schlessinger, J. (1997) Direct binding and activation of receptor tyrosine kinase by collagen, *Cell* **91**, 809–720.

Schneller, M., Vuori, K., Ruoslahti, E. (1997) $\alpha v\beta 3$ integrin associates with activated insulin and PDGF beta receptors and potentiates the biological activity of PDGF, *EMBO J.* **16**, 5600–5607.

Schreiner, C.L., Fisher, M., Bauer, J., Juliano, R.L. (1993) Defective vasculature in fibronectin-receptor-deficient CHO cell tumors in nude mice, *Int. J. Cancer* **55**, 436–441.

Schwartz, M.A., Ginsberg, M.H. (2002) Networks and crosstalk: integrin signaling in cell spreading, *Nat. Cell Biol.* **4**, E65–E68.

Sechler, J.L., Rao, H., Cumiskey, A.M., Vega-Colon, I., Smith, M.S., Murata, T., Schwarzbauer, J.E. (2001) A novel fibronectin binding site required for fibronectin fibril growth during matrix assembly, *J. Cell Biol.* **154**, 1081–1088.

Senger, D.R., Claffey, K.P., Benes, J.E., Perruzzi, C.A., Sergiou, A.P., Detmar, M. (1994) Angiogenesis promoted by vascular endothelial growth factor: regulation through alpha(1)beta(1) and alpha(2)beta(1)integrins, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 13612–13617.

Senger, D.R., Perruzzi, C.A., Streit, M., Koteliansky, V.E., de Fougerolles, A.R., Detmar, M. (2002) The alpha(1)beta(1) and alpha(2)beta(1)

integrins provide critical support for vascular endothelial growth factor signaling, endothelial cell migration, and tumor angiogenesis, *Am. J. Pathol.* **160**, 195–204.

Siebold, B., Qian, R.-Q., Glanville, R.W., Hofmann, H., Deutzmann, R., Kuhn, K. (1987) Construction of a model for the aggregation and cross-linking region (7S domain) of type-IV collagen based upon an evaluation of the primary structure of the $\alpha 1$ and $\alpha 2$ chains in this region, *Eur. J. Biochem.* **168**, 569–575.

Simons, M., Horowitz, A. (2001) Syndecan-4-mediated signaling, *Cell Signal.* **13**, 855–862.

Sipes, J.M., Guo, N., Negre, E., Vogel, T., Krutzsch, H.C., Roberts, D.D. (1993) Inhibition of fibronectin binding and fibronectin-mediated cell adhesion to collagen by a peptide from the second type I repeat of thrombospondin, *J. Cell Biol.* **121**, 469–477.

Smyth, N., Vatansever, H.S., Murray, P., Meyer, M., Frie, C., Paulsson, M., Edgar, D. (1999) Absence of basement membranes after targeting the LAMC1 gene results in embryonic lethality due to failure of endoderm differentiation, *J. Cell Biol.* **144**, 151–160.

Somg, Q.H., Klepeis, V.E., Nugent, M.A., Trinkaus-Randall, V. (2002) TGF-beta-1 regulates TGF-beta-1 and FGF mRNA expression during fibroblast wound healing, *Mol. Pathol.* **55**, 164–176.

Sottile, J., Hocking, D.C. (2002) Fibronectin polymerization regulates the composition and stability of extracellular matrix fibrils and cell-matrix adhesions, *Mol. Cell Biol.* **13**, 3546–3559.

Stad, R.K., Buurman, W.A. (1994) Current views on structure and function of endothelial adhesion molecules, *Cell Adhes. Commun.* **2**, 261–268.

Stenback, F., Wasenius, V.M. (1985) Basement membranes in ultraviolet light-induced skin lesions and tumors, *Photodermatol.* **2**, 347–358.

Stromblad, S., Becker, J.C., Yebra, M., Brooks, P.C., Cheresh, D.A. (1996) Suppression of p53 activity and p21WF1/CIP1 expression by vascular cell integrin $\alpha v\beta 3$ during angiogenesis, *J. Clin. Invest.* **98**, 426–433.

Stupack, D.G., and Cheresh, D.A. (2002) ECM remodeling regulates angiogenesis: endothelial integrins look for new ligands, *Sci.STKE* **119**, PE7.

Sundaramoorthy, M., Meiyappan, M., Todd, P., Hudson, B.G. (2002) Crystal structure of NC1 domains: structural basis for type-IV collagen assembly in basement membranes, *J. Biol. Chem.* **277**, 31142–31253.

Talts, J.F., Andac, Z., Gohring, W., Brancaccio, A., Timpl, R. (1999) Binding of the G domains of laminin alpha1 and alpha2 chains and perlecan to heparin, sulfatides, alpha-dystroglycan and several extracellular matrix proteins, *EMBO J.* **18**, 863–870.

Tapanadechopone, P., Hassell, J.R., Rigatti, B., Couchman, J.R. (1999) Localization of glucosaminoglycan substitution sites on domain V of mouse perlecan, *Biochem. Biophys. Res. Commun.* **265**, 680–690.

Timpl, R. (1989) Structure and biological activity of basement membrane proteins, *Eur. J. Biochem.* **180**, 487–502.

Timpl, R. (1993) Proteoglycans of basement membranes, *Experientia* **49**, 417–4282.

Tomasini, B.R., Mosher, D.F. (1991) Vitronectin, *Prog. Hemost. Thromb.* **10**, 269–305.

Tomono, Y., Naito, I., Ando, K., Yonezawa, T., Sado, Y., Irakawa, S., Arata, J., Okigaki, T., Ninomiya, Y. (2002) Epitope-defined monoclonal antibodies against multiplexin collagen demonstrates that type-XV and XVIII collagens are expressed in specialized basement membranes, *Cell Struct. Funct.* **27**, 9–20.

Tremble, P., Chiquet-Ehrismann, R., Werb, Z. (1994) The extracellular matrix ligands fibronectin and tenascin collaborate in regulating collagenase gene expression in fibroblasts, *Mol. Biol. Cell.* **5**, 439–453.

Trikha, M., De Clerck, Y.A., Markland, F.S. (1994) Contortrostatin, a snake venom disintegrin, inhibits beta 1 integrin-mediated human metastatic cell adhesion and blocks experimental metastasis, *Cancer Res.* **54**, 4993–4998.

Tuckwell, D.S., Ayad, S., Grant, M.E., Takigawa, M., Humphries, M.J. (1994) Conformation dependence of integrin-type II collagen binding: inability of collagen peptides to support $\alpha 2\beta 1$ binding, and mediation of adhesion to denatured collagen by a novel $\alpha 5\beta 1$-fibronectin bridge, *J. Cell Sci.* **107**, 993–1005.

Tunggal, P., Smyth, N., Paulsson, M., Ott, M.-C. (2000) Laminins: structure and genetic regulation, *Microsc. Res. Tech.* **51**, 214–227.

Van Belle, P.A., Elenitsas, R., Satyamoorthy, K., Wolfe, J.T., Guerry, D., Schuchter, L., Van Belle, T.J., Albelda, S., Tahin, P., Herlyn, M.,

Elder, D.E. (1999) Progression-related expression of beta 3 integrin in melanomas as nevi, *Hum. Pathol.* **30**, 562–567.

Vernon, R.B., Angello, J.C., Iruela-Arispe, M.L., Lane, T.F., Sage, E.H. (1992) Reorganization of basement membrane matrices by cellular traction promotes the formation of cellular networks in vitro, *Lab. Invest.* **66**, 536–547.

Vu, T.H., Werb, Z. (2000) Matrix metalloproteinases: effects on development and normal physiology, *Gene. Dev.* **14**, 2123–2133.

Vu, T.H., Shipley, J.M., Bergers, G., Berger, J.E., Helms, J.A., Hanahan, D., Shapiro, S.D., Senior, R.M., Werb, Z. (1998) MMP-9/gelatinase B is a key regulator of growth plate angiogenesis and apoptosis of hypertrophic chondrocytes, *Cell* **93**, 411–422.

Wagner, D., Ivatt, R., Destree, A., Hynes, R. (1981) Similarities and differences between fibronectins of normal and transformed hamster cells, *J. Biol. Chem.* **25**, 11708–11715.

Watanabe, K., Takahashi, H., Habu, Y., Kamiya-Kubushiro, N., Kamiya, S., Nakamura, H., Yajima, H., Ishii, T., Katayama, T., Miyazaki, K., Fukai, F. (2000) Interaction with heparin and matrix metalloproteinase 2 cleavage expose a cryptic anti-adhesive site of fibronectin, *Biochemistry* **39**, 7138–7144.

Watt, F.M. (2002) Role of integrins in regulating epidermal adhesion, growth and differentiation, *EMBO J.* **21**, 3919–3926.

Werb, Z., Tremble, P.M., Behrendtsen, O., Crowley, E., Damsky, C.H. (1998) Signal transduction through the fibronectin receptor induces collagenase and stromelysin gene expression, *J. Cell Biol.* **109**, 877–889.

Wever, U.M., Shaw, L.M., Albrechtsen, R., Mercurio, A.M. (1997) The integrin alpha (5) beta (1) promotes survival of metastatic human breast carcinoma cells in mice, *Am. J. Pathol.* **151**, 1191–1198.

Woods, A. (2001) Syndecans: transmembrane modulators of adhesion and matrix assembly, *J. Clin. Invest.* **107**, 935–941.

Xu, J., Rodriguez, D., Kim, J.J., Brooks, P.C. (2000) Generation of monoclonal antibodies to cryptic collagen sites by using subtractive immunization, *Hybridoma* **19**, 375–385.

Xu, J., Rodriguez, D., Petitclerc, E., Kim, J.J., Mangai, M., Moon, S.Y., Davis, G.E., Brooks, P.C. (2001) Proteolytic exposure of a cryptic site within collagen type-IV is required for angiogenesis and tumor growth in vivo, *J. Cell Biol.* **154**, 1069–1079.

Yurchenco, P.D., Ruben, G.C. (1987) Basement membrane structure in situ: evidence for lateral associations in the type-IV collagen network, *J. Cell Biol.* **105**, 2559–2568.

Ziober, B.L., Chen, Y.Q., Ramos, D.M., Waleh, N., Kramer, R.H. (1999) Expression of the alpha (7) beta (1) laminin receptor suppresses melanoma growth and metastatic potential, *Cell Growth Differ.* **10**, 479–490.

Medicinal Chemistry

David J. Triggle
State University of New York, Buffalo, NY, USA

1 Background 84

2 The Receptor Concept 86

3 Chemistry 89
3.1 Druglike Molecules 90
3.2 Combinatorial Chemistry and Compound Libraries 91
3.3 Structure-based Approaches to Drug Design 97
3.4 Peptidomimetics 101
3.5 Protein–Protein Interactions 105
3.6 Quantitative Structure–Activity Relationships 106
3.7 Stereochemistry 108

4 The Influence of Molecular Biology 111
4.1 The Expanded Genetic Code 112
4.2 Genes as Drugs 112
4.3 Antisense and Small Interfering RNA 113
4.4 Gene-based Drug Targeting and Delivery 114

5 Summary 115

Bibliography 115
Books and Reviews 115
Primary Literature 115

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

Keywords

Combinatorial Chemistry
The discipline of high output synthesis of chemical structures in solution, in the solid phase, or by the use of biological expression technologies.

High-throughput Screening
Technologies complementary to combinatorial chemistry that permit the rapid and efficient screening in assays of the biological activities of compounds in an automated manner.

Ligand
A molecule that binds to a receptor to initiate (agonist) or block (antagonist) response. Ligands may be small natural or synthetic molecules (neurotransmitters and their analogs) or large peptides, proteins, or nucleic acids.

Peptidomimetic
A synthetic molecule that mimics the biological properties of peptides and proteins but does not include typical peptide structures.

Receptor
A component of a cell that interacts specifically with (receives) other molecules and, in appropriate combination, initiates (or blocks) biological response. Receptors may be protein, nucleic acid, lipid, or carbohydrate in composition.

Stereoselectivity
The ability of molecules to interact with receptors and other macromolecules in an asymmetric manner such that one structure exhibits preferential interaction or activity. Stereoselectivity is extremely common in biological systems because the building blocks of nature – amino acids, sugars etc. – are chiral.

Structure–Activity Relationships
The definition of the relationship between chemical structure and biological activity. This may be achieved in qualititative terms or, and increasingly, in quantitative terms that involve the determination of the 3D structure of the ligand–receptor complex.

Medicinal chemistry defines one component of a sequence of events in the process of drug discovery and development. The critical steps of lead structure identification and refinement have been, and continue to be, the major contributions of medicinal chemistry to the drug discovery process: medicinal chemistry may be thought of as comprising three stages – a discovery step, an optimization step, and a production step (Fig. 1). Historically, drug development has been associated with biologically active products from natural sources, principally plants, and the development of natural product chemistry. Indeed, some 50% of currently available drugs have their origins, directly or indirectly, in natural products. With the ascendancy of synthetic

organic chemistry, emphasis was increasingly placed on the screening of synthesized molecules in a variety of biological test systems that moved progressively from *in vivo* whole-animal systems to single expressed protein systems and to targets derived from reading of the genome. This approach has acquired increasingly mechanistic underpinnings with the development and quantitation of the receptor concept and the availability of receptor-based assays. The processes of lead structure identification and refinement have become progressively more sophisticated in partnership with developments in molecular and structural biology that have led to the characterization of new potentially drugable targets, the availability of human proteins as research tools with which to study drug–receptor interactions and as biopharmaceuticals, and the availability of three-dimensional structures of human receptors and enzymes with which to perform *in silico* medicinal chemistry. Finally, the new technologies embodied in high-throughput screening and combinatorial chemistry have led to an extraordinary level of biological activity determination: the translation of this productivity into therapeutically available medicines remains, however, a significant problem.

The genomics era, initiated in 1953 with the publication of the Watson and Crick paper on the structure of DNA, has had a major impact not only on biology but also on other scientific disciplines, including chemistry. The basic themes that govern biology – diversity, replication, evolution, and self-organization – are increasingly major components of medicinal chemistry, in the development of combinatorial chemistry, self-replicating molecules, *in vitro* evolution of molecules to biological fitness, and the synthesis of molecules around a template.

The elucidation of the human genome, as well as the genomes of other species including many bacteria, was the second major event of the genomics era. From the perspective of drug discovery, the human genome was anticipated to yield an enormous increase in the number of potential drug targets and thus greatly expedite the availability of new drugs. Together with the technologies of combinatorial chemistry and high-throughput screening, the production line for new drugs was anticipated to be faster and more capacious by several orders of magnitude. However, the number of human genes, approximately 30 000 (but still an uncertain number), is relatively small by comparison with the number of genes in other species, suggesting that our complexity arises not from a simple increase in the number of proteins, but rather from an expansion in the complexity of the signaling and regulatory networks. Thus, the number of potential drug targets is likely to be significantly smaller than was originally anticipated: nonetheless, the anticipated number of 1000 to 2000 potential targets is still significantly larger than the number of current targets, estimated at approximately 120 in all. The issue remains, however, as to how many of these new targets will actually be drugable.

Knowledge of the human genome will also lead to the introduction of *personalized* medicine whereby the drug will be matched more precisely to the patient, thus generating better response, facilitating clinical trial development, and reducing drug withdrawal and related misadventures that are typically due to the reactions with an extremely small percentage of patients, but nonetheless prominent because of the large patient base with many drugs.

1
Background

Active principles from natural sources were the original drugs available to man and still remain a very significant source. References to such products date back thousands of years to the Chinese, Greek and Roman civilizations, and "natural medicines" have been used therapeutically since prerecorded times, as part of religious ceremonies, and doubtless also to relieve a life that was "nasty, short and brutal". Thus, the acetycholinesterase inhibitor physotigmine, isolated from the Calabar bean was used in the West African tribal ritual of *trial by ordeal*, and various psychoactive substances found in plants have been used for thousands of years. The nature of these principles, which were invariably crude mixtures, was not known and their activity doubtless varied considerably. However, as their chemical constitutions came to be elucidated, they became available as pure materials and also served as leads for synthetic analogs with improved properties. Among the natural products that have served as leads in this manner are the antimalarial quinine, the analgesic morphine, the antihypertensive drug reserpine and the stimulant cocaine (Fig. 2). Introductions in the twentieth century from plant natural sources include many antibiotics such as the penicillins, cephalosporins, and tetracyclines and, more recently, the cholesterol-lowering agent compactin (a lead to the statin family), the immunomodulator cyclosporin, the antimalarial agent artemisinin, and anticancer agents such as paclitaxel, camptothecin, and the bryostatins (Fig. 3). Finally, the discovery of the multiple regulatory and signaling properties of nitric oxide (NO) in neuronal, cardiovascular, immune, and inflammatory responses emphasizes the route to drug discovery through the elucidation of endogenous pathways and structures (Fig. 4).

Despite the early success of plant-based medicinal chemistry, the advent of synthetic organic chemistry resulted in a significant displacement of botanicals as drugs. There is now, however, a significant resurgence of interest in natural products as pharmaceuticals, and due to recent successes, this has led to an increasing

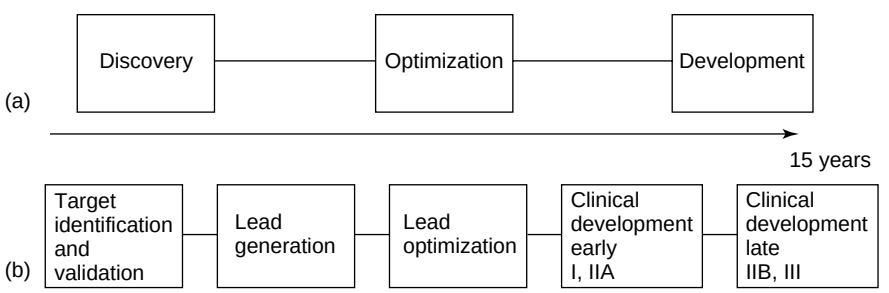

Fig. 1 (a) The drug discovery process may be viewed as composed of three principal stages – *discovery, optimization, development*. (b) The discovery phase may be viewed as composed of five stages: following target identification and validation that may arise from phenotypic or genotypic observations of the two stages of medicinal chemistry – *lead generation and lead optimization* – are initiated. During these two stages, compounds having appropriate pharmacodynamic and pharmacokinetic properties are identified and are subsequently optimized for these properties.

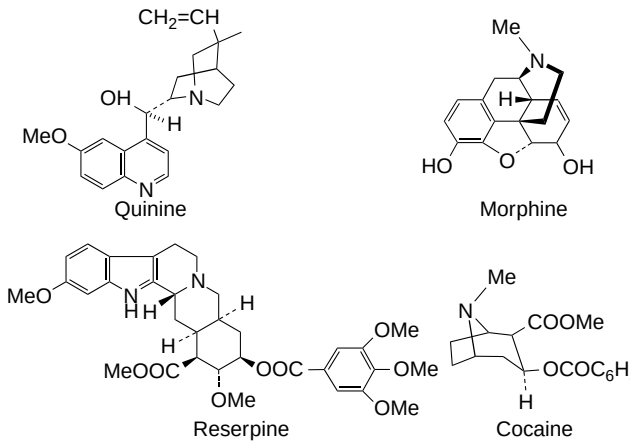

Fig. 2 Examples of potent naturally occurring pharmacologic entities.

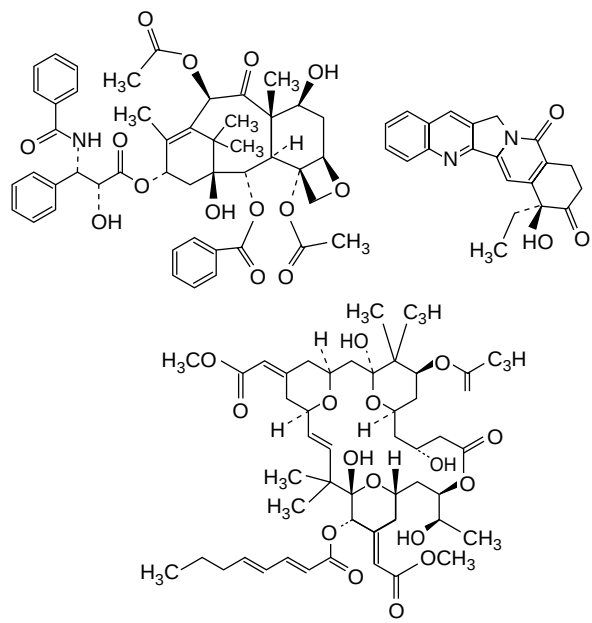

Fig. 3 Molecules derived from plant natural sources in the twentieth century including paclitaxel, camptothecin, and bryostatin 1.

understanding of indigenous peoples' healing practices, and to the increasing realization that habitat destruction is causing the loss of potentially valuable drugs and drug leads. For example, it is likely that fertility control has long been practiced by humans through the use of plant-based extracts, including

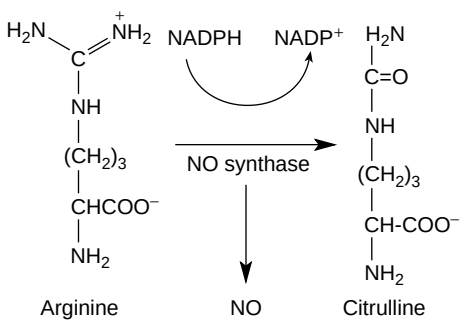

Fig. 4 The nitric oxide (NO synthase) pathway.

one from *Silphion* during the Greek and Roman civilizations:this plant species was apparently harvested to extinction.

2 The Receptor Concept

The concept that drugs act at specific sites, though long recognized, is critical to current processes of drug discovery and to the elucidation of the mechanisms of drug action. Over the past century, the contributions of Paul Ehrlich to immunology and the chemotherapy of protozoan infections, and those of John Newton Langley to drug action in the nervous system are of particular note. Ehrlich viewed receptors as the protoplasmic side chains of cells, and Langley defined a receptor as "a specially excitable constituent [which] receives the stimulus and, by transmitting it, causes contraction" [response]. These considerations, together with Emil Fischer's concept of a *lock-and-key* process of small molecule–macromolecule interactions, categorized drug action as specific rather than nonspecific, with defined structure–activity relationships including stereoselectivity, and serving to stimulate or block receptor activity, agonism and antagonism, respectively. The mutual complementarity of the drug molecule and the receptor-binding site provides the underlying basis of structure–activity relationships, the definition and analysis of dose–response relationships and the elucidation, through structural biology techniques, of new ligand structures.

Receptors are classified into several major families, the individual members of which show substantial structural and functional homology. The major classes include the following:

1. The G-protein-coupled receptors: for neurotransmitters (dopamine etc), many polypeptide hormones (substance P, etc.), odorants and light.
2. Ligand-gated ion channels: ion channels with integral receptors for neurotransmitters (acetylcholine, glutamic acid, etc.).
3. Voltage-gated ion channels: ion channels that are activated by changes in cellular electrical potential (Na^+, K^+ and Ca^{2+}).
4. Enzyme-associated receptors: tyrosine kinase, and guanylyl cyclase.
5. Steroid hormone receptors: nuclear transcriptional regulators.

The G-protein-coupled receptor (GPCR) family is one of the largest gene families constituting more than 1% of the human genome, and more than half of all modern drugs have been estimated

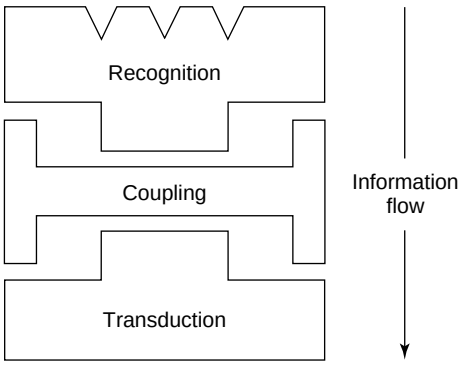

Fig. 5 The receptor, depicted as three discrete entities, as an information generating and processing system.

to be directed at these receptors. This family encompasses receptors for biogenic amines, polypeptide hormones, taste and olfactory stimuli, adhesion ligands and other signaling systems.

Receptors may be viewed as informational machines mediating a vectorial flow of information to the cellular machinery from the ligand, the receptor, or both (Fig. 5). Consistent with this scheme, the majority of receptors are multimeric protein assemblies and their activity may in principle be controlled by drug action at a variety of sites in the complex. Thus, in the G-protein-coupled receptor system depicted in Fig. 6, activity may be regulated by ligand interaction at the receptor-recognition site, at allosteric sites associated with the receptor, through drugs regulating the G-proteins themselves, or through drugs acting on the effectors. The several components of the total receptor complex, and the interfaces between the components, offer multiple opportunities for drug action in addition to the classically defined receptor itself.

Several properties including drug selectivity, affinity, biological response, coupling components, gene sequence, and protein structure determine the overall classification of pharmacological receptors. Thus, in the major family of G-protein-coupled receptors, characterized by a homologous structure with seven transmembrane sequences, there are five principal subclasses of muscarinic receptors (m1–m5) characterized by selective drugs and coupled

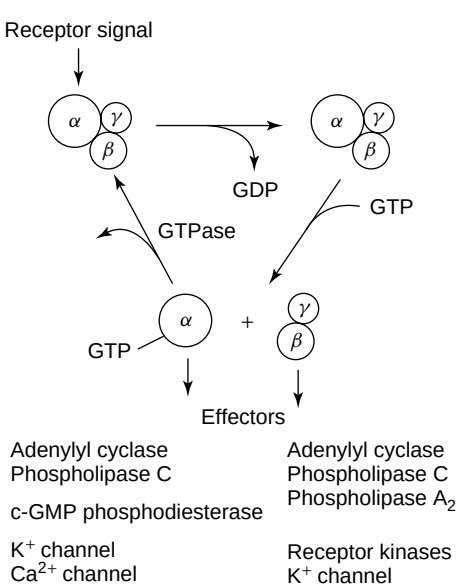

Fig. 6 The organization of heteromeric guanine nucleotide (G) binding proteins as a GTP/GDP exchanger and their coupling to a variety of biological effectors.

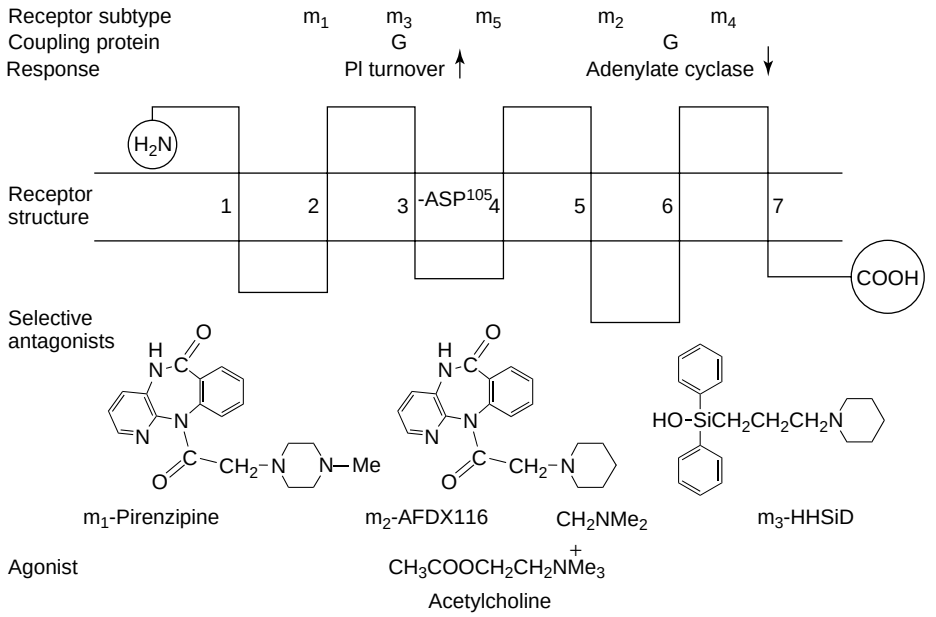

Fig. 7 The muscarinic acetylcholine receptor as a member of the G-protein-coupled family existing as five subtypes – m_1 to m_5 – recognized by selective antagonists and differentially coupled to stimulation of phosphatidylinositol (PI) hydrolysis or inhibition of adenylate cyclase.

differentially to pharmacological effectors (Fig. 7).

The traditional route to the classification, characterization, and eventual isolation of pharmacological receptors has stemmed from the recognition that a specific hormone, neurotransmitter, or drug, produce a defined biological response, that the action of the agent can be mimicked by structurally related compounds with

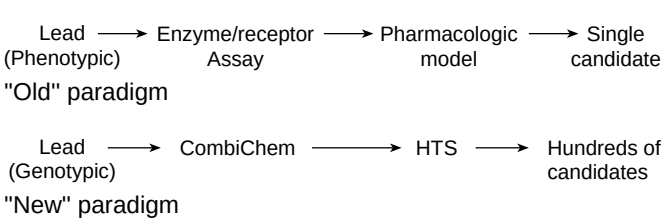

Fig. 8 The ''old'' and ''new'' paradigms of drug discovery. In the old paradigm, a lead target observation, typically a biological (phenotypic) observation such as excess gastric acid secretion or vasoconstriction was used as the basis for an assay into which were fed potential drug candidates typically synthesized ''one molecule at a time'' to generate a single clinical candidate. In the new paradigm, a lead target is derived from a genome database, and high-throughput chemistry and high-throughput screening are applied to generate, in principle, hundreds of clinical candidates.

a definable structure–activity relationship including stereoselectivity and that both agonists and antagonists exist. The actions of a natural product such as morphine led early to the conclusion that endogenous opiods exist for which morphine is a mere surrogate and that there must exist one or more opiod receptors through which the actions of endogenous, naturally occurring, and synthetic opiods occur. From the combined knowledge of receptor structure and the human and other genomes, it is now possible to identify receptors through homology searches in the absence of any knowledge of the endogenous ligand or of the biological responses controlled. The identification of such *orphan* receptors represents a reversal of the classical receptor characterization scheme and provides a further source of potential drug targets (Fig. 8).

3 Chemistry

Traditionally, the medicinal chemistry component of drug discovery has been initiated from a lead structure, either a natural product of plant or animal origin, an endogenous hormone or neurotransmitter, or a synthetic molecule exhibiting some degree of the appropriate biological activity. The lead structure is then optimized to produce the desired potency and selectivity of action and the required pharmacokinetic characteristics. Acquisition of the latter is of particular importance since many agents fail in development because of inadequate solubility, absorption or metabolic properties. This process, essentially a "one- molecule-at-a-time" scheme, has achieved many major successes including the development of the β-adrenergic blockers from isoproterenol, the development of angiotensin-converting enzyme (ACE) inhibitors from a peptide teprotide and the development of the angiotensin II receptor antagonists, including losartan, from a weakly active imidazole derivative (Fig. 9). In this process, the biological target, be it blood pressure, acid secretion, tumor growth, anxiety, or depression, has been derived from a biological or phenotypic response.

In the new postgenomic era, it is, however, possible to define targets through analysis of gene structure from the approximately 30 000 genes currently believed to comprise the human genome as well as from the genomes of bacteria, viruses or parasitic species. Although the genomic approach offers, in principle, many new targets, the issue of target validation looms far larger than in the classical process. It is vitally necessary to associate a particular gene with a particular phenotypic expression *and* to demonstrate the relevance of this to the disease process. The techniques of target validation are several and include the analysis of gene and protein expression in normal and diseased tissues. Knockout and conditional knockout animals, the creation of mutant (ethyl-nitrosourea-induced) mice, the use of model organisms including *Drosophila melanogaster* and *Caenorhabditis elegans,* the use of small interfering RNA, and small molecule–based chemical–biological techniques.

With increasing knowledge of the target structure, there is greater emphasis on structure-based drug design whereby X ray, NMR, and computational approaches are employed to provide a physical interpretation of the drug–receptor complex and, in principle, to provide a more rapid and efficient route to drug design. There are still significant limitations

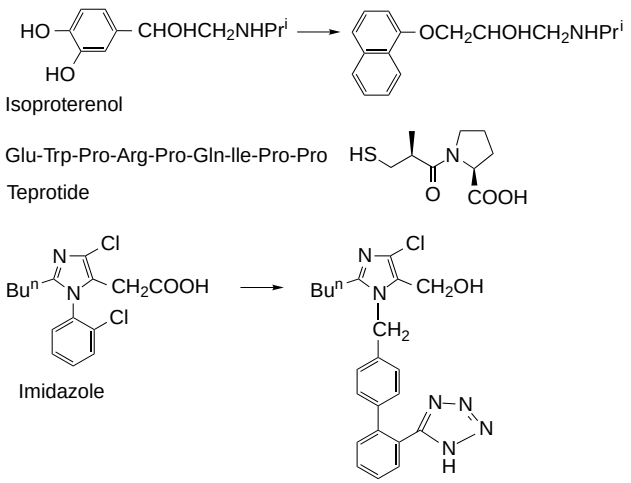

Fig. 9 Examples of lead structures (left) that gave rise to the synthetic antagonists propranolol, captopril, and losartan, active at β-adrenergic receptors, angiotensin-converting enzyme, and angiotensin II receptors, respectively.

to this approach, although it will doubtless continue to assume greater importance. First, the structures of many targets are still unknown and the structures of ligands in solution or in the solid state, although frequently easily accessible, may have no relationship to the structure of the receptor-bound ligand.

At the other extreme of drug discovery has been the development of combinatorial chemistry and high-throughput screening, whereby the ability to synthesize vast chemical libraries and to use automated techniques to screen them against a large number of (typically) genomically derived targets was intended to greatly automate and render more productive the drug discovery process (Fig. 8). This promise remains to be completely fulfilled. However, the synergies obtained by the combined application of combinatorial chemistry and structure-based designs are fueling the new approaches to drug discovery, designed to combine the intelligence of rationality with the efficiency of technology.

3.1 Druglike Molecules

Despite the ability to make chemical libraries of virtually unlimited size, the issue remains that the number of "drug-like" molecules is relatively small and that quite limiting chemical and physical properties are necessary to ensure that molecules are both biologically active *and* orally available. A number of approaches have been employed in efforts to quantitate these properties. The best known of these approaches is probably the empirical "rule-of-five" derived from a consideration of over 2000 therapeutically active drugs that indicates a likely problem when a molecule obeys any two of the following conditions:

- Molecular weight >500
- Number of H-bond acceptors >10
- Number of H-bond donors >5
- Calculated logP >5.0.

A number of refinements of this system have been implemented. For example, a

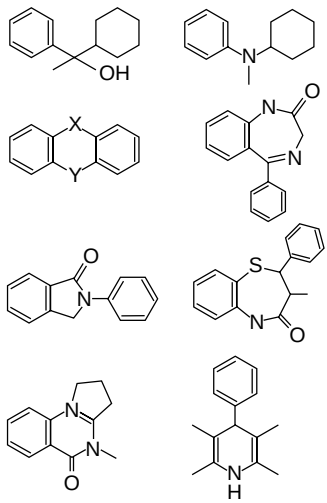

Fig. 10 Examples of *pharmacophoric* structures or building blocks that are commonly found (with appropriate molecular decoration) in pharmacologically active agents.

survey of over 1100 drug candidates at GlaxoSmithKline for which oral bioavailability measurements are available suggests that molecules with reduced flexibility (10 or fewer rotatable bonds) and a polar surface area of less than 140Å^2 (or 12 or fewer H-bond donors or acceptors) are important predictors of good oral bioavailability. These and related approaches have been reasonably successful in predicting "druglike" molecules present in large chemical databases such as the Available Chemicals Directory and the Comprehensive Medicinal Chemistry Database.

A second and related issue is that of the chemical structures of active molecules: consistent with the above considerations, these also tend to be relatively limited in terms of potential chemical space and component structures – *pharmacophores* – tend to be conservatively employed. A number of basic pharmacophoric structures or *scaffolds* exist (Fig. 10) onto which are grafted the functional groups that define specific drug families. It is proposed that these structures are those that resist "hydrophobic collapse" (the tendency to self-associate to inactive species) through their conformational nonflexibility. Prominent examples of such scaffolds include the 1,4-dihydropyridine and benzodiazepine nuclei, both of which give rise to potent and biologically diverse classes of drugs (Figs. 11 and 12). An interesting example of combinatorial chemistry practiced by Nature, the supreme combinatorial chemist, centers around the venomous snails of the *Conus* genus. These snails produce disulfide-bridged peptide toxins that exhibit both high affinity and selectivity for a variety of neurotransmitter receptors and ion channels. The peptides are synthesized as precursors with a hypervariable C-terminus region, the latter permitting amino acid changes in discrete regions to tailor pharmacological specificity.

3.2 Combinatorial Chemistry and Compound Libraries

Chemical libraries have long been an important proprietary component of the drug discovery process in the search for "lead" structures. Such libraries

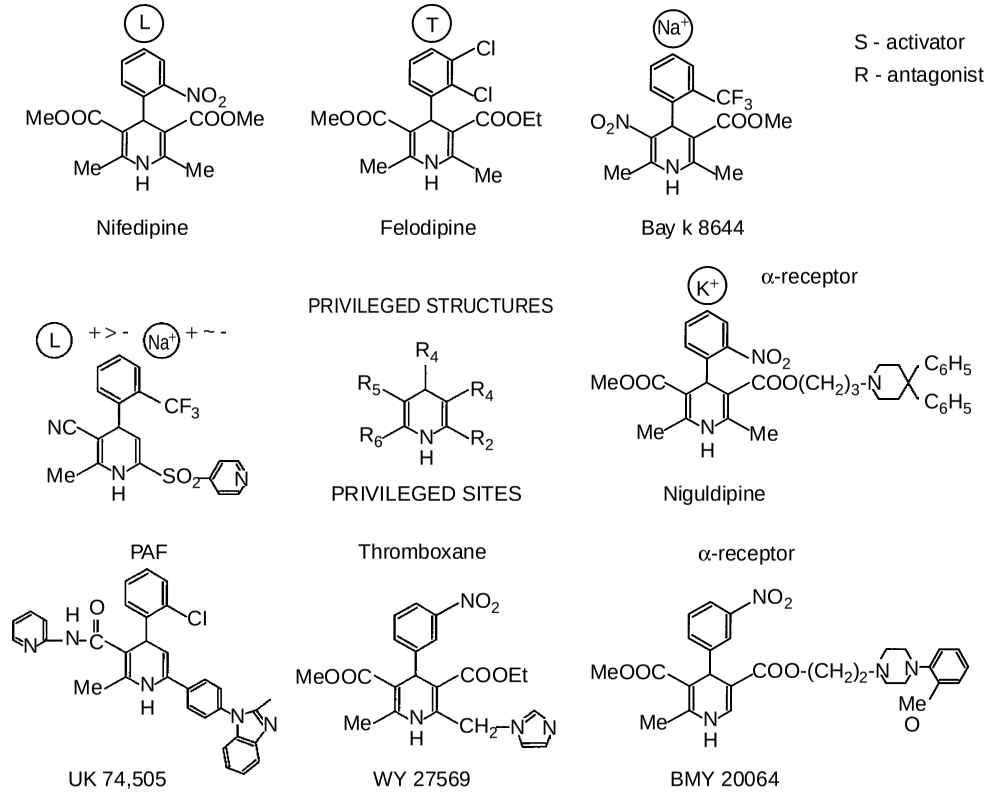

Fig. 11 The 1,4-dihydropyridine nucleus as a *privileged* structure that can be directed at a diverse set of ion channels and pharmacological receptors.

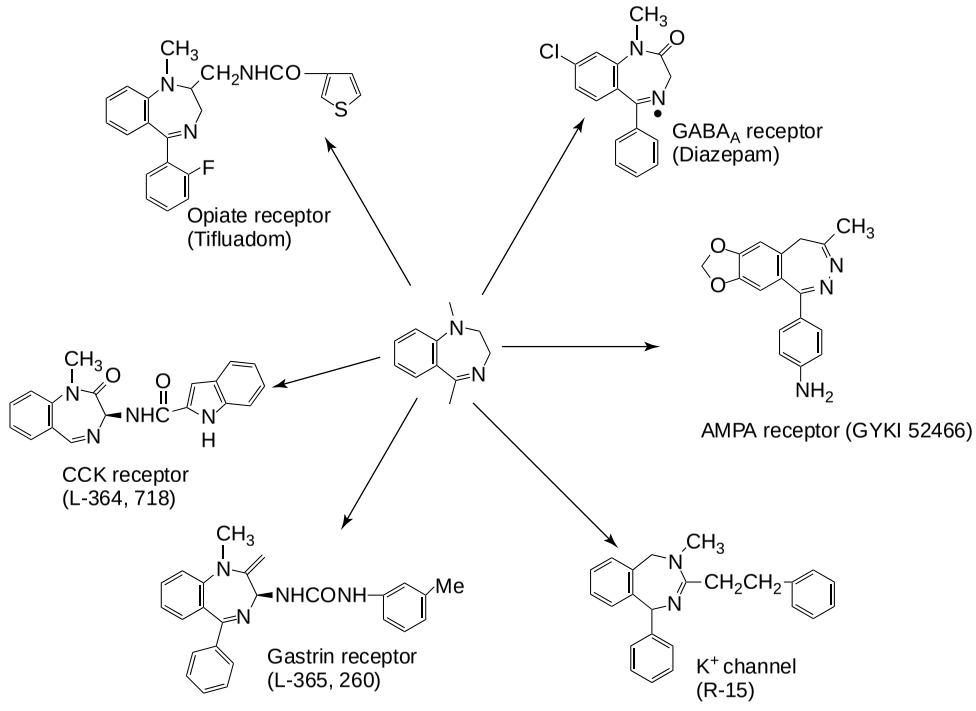

Fig. 12 The 1,4-benodiazepine nucleus as a *privileged* structure capable of interacting at a wide variety of ion channels and pharmacologic receptors.

were typically derived from the traditional "one-molecule-at-a -time" approach of drug discovery, and were typically between 1000 and 100 000 molecules in size: these libraries were based to a very significant extent on the previous chemical experience of the company and could vary widely in both molecular diversity and "drug-likeness". The technologies of parallel and combinatorial synthesis have dramatically increased library scale. A typical organic chemist can perhaps synthesize 50 to 100 compounds per year. The technique of parallel synthesis in which, with the aid of automation, multiple products can be synthesized in discrete reaction vessels can increase this chemical productivity by a 100-fold factor or more. The combinatorial approach, whether in solution or, more typically, in the solid phase, expands this scale further by performing multiple reactions in a single reaction vessel, creating all possible products from a precursor set of building blocks. A simple calculation illustrates the power of the combinatorial approach. From a basic set of 1000 building blocks, it is possible to construct 1000^5 pentapeptides, or one quadrillion molecules! In general, if the number of available building blocks is a, the number of synthetic steps is x and the same number of building blocks is used per step, then N, the number of individual compounds is given by

$$N = a^{x} \quad (1)$$

And if the number of building blocks varies in each step (e, f, g in a three-step reaction), then

$$N = \text{e.f.g} \quad (2)$$

In principle, one can make more compounds by combinatorial chemistry than there is mass in the universe! Although the original thrust of combinatorial chemistry (coupled with the simultaneous development of high-throughput screening) was designed to produce as many molecules as possible and was a retreat from rational discovery processes, the paucity of success in such large random libraries led to a refocusing of combinatorial thinking to factor in both diversity and druglike properties in the construction of smaller libraries.

The selection and choice of building blocks *and* the linking chemistry are the critical steps in the combinatorial process. For example, the development of peptide and oligonucleotide libraries is quite straightforward since the building blocks and the coupling chemistry are both well known and easily implemented. However, because of the limitations to the use of these structures as drugs, increasing attention is being directed to the use of building blocks that create small-molecule libraries. A second consideration is whether the library is to be used for random screening or for lead development. Increasingly, combinatorial chemistry is used in the generation of "focused" libraries that are designed either to cover wide areas of chemical space in as small a collection as possible or to exploit chemical space around a particular structure. Furthermore, the need for such libraries to contain "druglike" molecules is increasingly recognized. The choice of natural products as libraries has much to commend it since these molecules, complex as they appear, have evolved and have been "forged in the crucible of evolution" to have specific biological functions and to bind to proteins or nucleic acids. A number of such natural product libraries are shown in Fig. 13. In any event, library design and development increasingly includes consideration of those molecular features that provide appropriate pharmacokinetic characteristics.

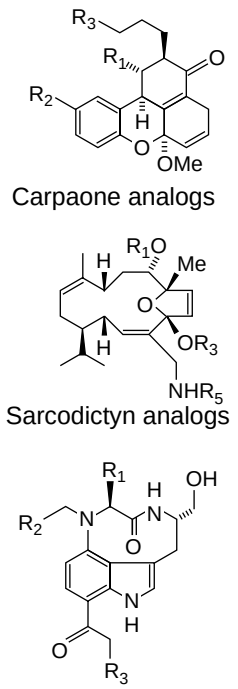

Fig. 13 Combinatorial libraries based on natural product lead structures.

The solid-phase approach of combinatorial chemistry is generally performed on resin particles or on two-dimensional arrays. An illustration of one such process, the split synthesis technique, is provided in Fig. 14. At each synthetic state, the resin support is pooled and then split: this technique overcomes the different reactivities of individual building blocks and ensures the equimolar distribution of the product, which is critical to the quantitation of the subsequent screening process. At the completion of the synthetic sequence, the compounds are cleaved from the resin support. A variety of techniques can be used to deconvolute the library and to identify the active structure(s). One approach is the encoded library in which a code or identifier tag is added at each stage of the synthetic process (Fig. 15). The library can then be "read" from the code sequence that is unique to each member of the library. Varying codes may be employed including sequencable, (nucleotides and peptides), nonsequencable, (fluorophores) or nonmolecular (radiofrequency) codes.

The process of *dynamic combinatorial chemistry* relies on reversible reactions between the building blocks of the system: thus, the composition of the library is under thermodynamic rather than kinetic control (Table 1). This is a critical differentiation from conventional

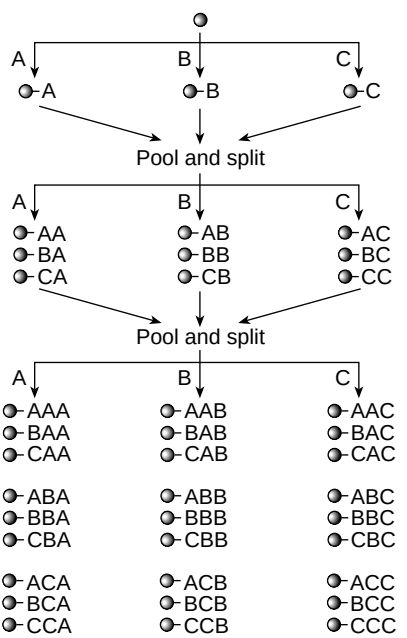

Fig. 14 The split synthesis method for the generation of combinatorial libraries in which each *monomer* is reacted to completion with the solid support and then pooled and split repeatedly to generate the resin-bound library. (From S. Hobbs DeWitt, Pharmaceutical News 1 : 11, 1994.)

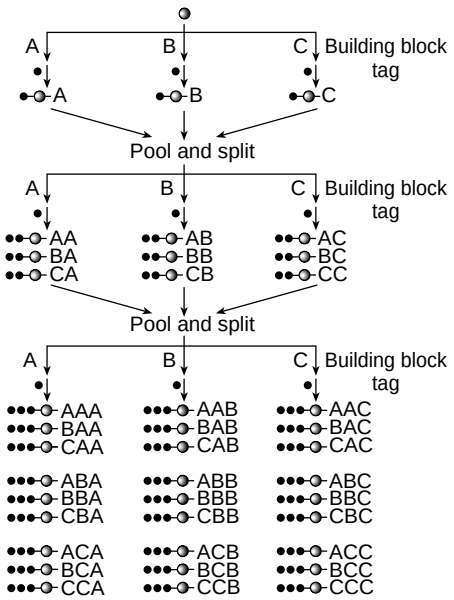

Fig. 15 The encoded combinatorial library in which the addition of each *monomer* is accompanied by a *tag*. The tag sequence that may be chemical or electronic specifically encodes the oligomeric sequence of interest. (From S. Hobbs DeWitt, Pharmaceutical News 1 : 11, 1994.)

Tab. 1 Combinatorial chemistry and dynamic combinatorial chemistry.

Combinatorial libraries	*Dynamic libraries*
Unaffected by recognition	Amplified by recognition
Selected compounds to be made independently	Selected compounds isolable from library
Complex topologies difficult to access	Complex topologies easier to access
Greater chemical access	More limited chemical access
Solubility not important	All members of library must be soluble
Under kinetic control	Under thermodynamic control

combinatorial chemistry because it indicates that the presence of a recognition process specific for one member of the equilibrium will stabilize that member by an equilibrium shift between the reactants and the products (Fig. 16). Accordingly, an *amplification* of yield is obtained. In principle, dynamic combinatorial chemistry offers the heady vision of potent and selective molecules self-assembling in the presence of the biological target and thus automating the entire process of drug discovery. This approach has been applied to a number of systems of biological interest, including the development of inhibitors against neuraminidase using a scaffold reacting with a series of aldehydes to yield a dynamic library of Schiff bases and heminanimals (Fig. 17). In the presence of the target neuraminidase, significant amplification of only the selected components occurred. In a related example, the enzyme acetylcholinesterase was used as a target: a series of acylhydrazides containing a charged amine function (to interact with the enzyme anionic site) and aldehydes yielded a dynamic library from which a compound was isolated with nanomolar inhibitory potency (Fig. 18). In practice, although proof of principle has been obtained in a number of systems, there are limitations, first in the number of suitable reversible reactions available and in both the stability of the target macromolecule under the reaction conditions, and in the presence of undesirable cross reactions between the protein and the chemical reactants.

The process of *target-accelerated synthesis* of active molecules is related to dynamic combinatorial chemistry whereby the coupling of building blocks is accelerated

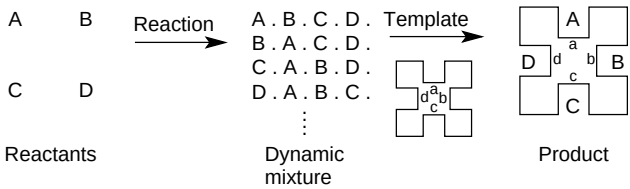

Fig. 16 In the process of dynamic combinatorial chemistry the reactant fragments, A–D, establish an equilibrium of products, one of which is selected by the presence of the template thus driving the overall reaction to select for that specific product.

Fig. 17 The use of dynamic combinatorial chemistry to develop a series of inhibitors against neuraminidase. The scaffold entity is reacted with a variety of ketones in the presence of the enzyme neuraminidase to produce by selection the active products.

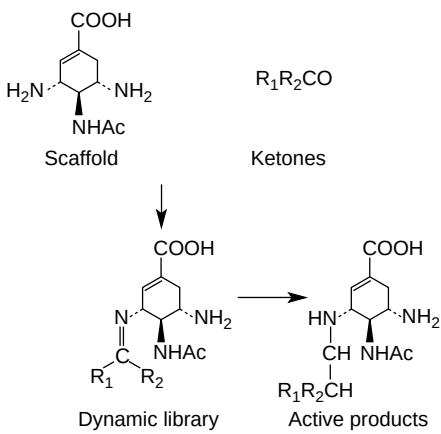

by assembling them on a target template: since the coupling is covalent, the system is under kinetic rather than under thermodynamic control. An example of target-accelerated synthesis is provided in Fig. 19, which depicts the assembly of an extremely potent (femtomolar KI) acetylcholinesterase inhibitor by 1,3-dipolar cycloaddition chemistry from the depicted starting materials.

Coming full circle back to Nature, combinatorial chemistry can also be achieved in laboratory-based *in vivo* systems through genetic manipulations of bacterial biosynthetic pathways to create new antibiotics and through phage display to generate peptide libraries. Both the polyketide and non-ribosomal classes of antibiotics are synthesized from their carboxylic acid or amino acid precursors, respectively, by large multifunctional polyketide synthases and peptide synthases. The modular construction of these enzymes makes possible their selective modification and thus the biosynthesis of novel products, including active antibiotics. Phage display technologies permit the display of very large peptide libraries that can be used *via* affinity-selection processes to define and target specific recognition sites and can be used to define new potential drug targets in cells and organs via phage targeting.

3.3 Structure-based Approaches to Drug Design

Increasingly, the intuition-based (and highly successful) approach to drug

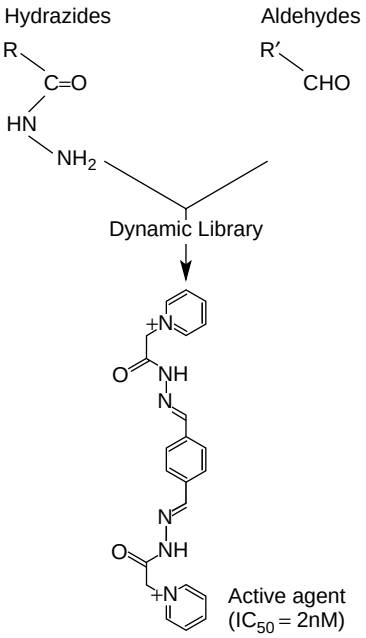

Fig. 18 The use of dynamic combinatorial chemistry to generate a dynamic library that in the presence of the target enzyme, acetylcholinesterase, yielded an active inhibitor with nanomolar affinity.

discovery is assuming a more quantitative character with the addition of structure-based components (Fig. 20). In particular, the understanding of the three-dimensional structure of the target proteins and of their ligand complexes with drugs is proving to be critical for the generation of docking processes to fit new molecules to active sites, *in silico* screening of potential drug structures, the elucidation of mechanisms of drug action, and the development of structure–activity relationships. This approach is frequently referred to as *structure-based drug design*, and with the increasing availability of macromolecular structures via X ray and NMR techniques or homology modeling, it is a rare contemporary drug discovery project that does not have inputs from these design principles. Additionally, the limitations caused by the lack of availability of protein three-dimensional structures are rapidly being resolved by detailed resolution of such membrane proteins as ion channels. With high-throughput protein crystallization processes now in place, it is likely that the next decade will see the majority of protein structures solved, either directly or indirectly, through the use of increasingly sophisticated computational methods.

The antagonists of HIV-1 protease, an enzyme critical to the process of viral assembly and maturation, provides one example of how structure-based design has led to the development of clinically effective and orally active drugs. Homology-based reasoning originally suggested that HIV-1 protease was related mechanistically to the aspartate protease family. This family is characterized by a two-domain active site, with each domain contributing an aspartate residue to the catalytic process. Direct X ray analysis confirmed this, showing the enzyme to be a homodimer with C2 symmetry (Fig. 21). Some symmetry-based antagonists designed from this structure are depicted in Fig. 22.

Most recently, the crystal structure of the angiotensin-converting enzyme complexed with the ACE inhibitor lisinopril has been obtained. ACE inhibitors are important anti-hypertensive agents that play an important role in the control of congestive heart failure. These inhibitors were developed without the knowledge of the structure of the human ACE enzyme but rather were designed on the assumption that the enzyme was mechanistically similar to carboxypeptidase A. There is, however, little three-dimensional similarity, thus indicating a comparatively common

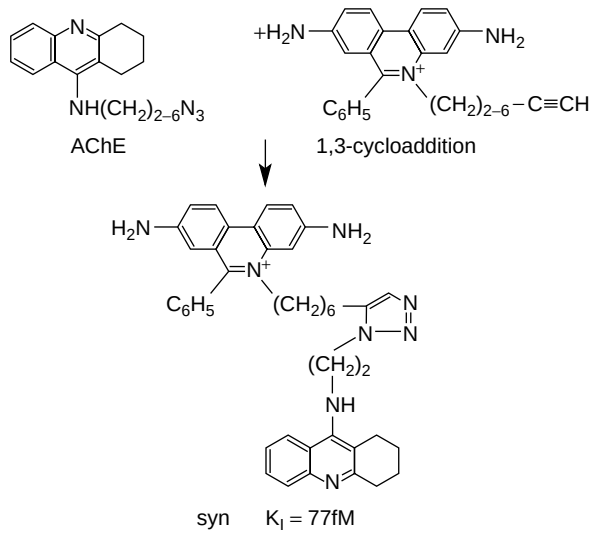

Fig. 19 An example of target-accelerated synthesis of an active molecule. In this example, acetylcholinesterase is the target macromolecule and small-molecule fragments that interact at discrete subsites of the enzyme become linked through 1,3-cycloaddition to yield the active inhibitor with femtomolar potency.

facet of scientific discovery, namely, that significant discoveries can be made on the basis of incorrect assumptions! However, the availability of the ACE structure should facilitate the development of new ACE inhibitors.

The structures of membrane-bound receptors and ion channels have, until recently, not been available at a resolution helpful to the generation of structure-based drug design. Such knowledge is, however, becoming available. In its absence, sequence data, homology modeling, and site-directed mutagenesis have permitted important conclusions to be drawn concerning the residues and topographies that are critical to drug–receptor interactions. The β-adrenergic receptor is a member of the G-protein coupled protein family characterized by the presence of 7 transmembrane helices. Site-directed mutagenesis has revealed the presence of residues critical to both agonist binding and receptor activation. A model for the interaction of the agonist isoproterenol is provided in Fig. 23. The interactions between transmembrane domains V and VI are consistent with observations that the cytoplasmic loop between these domains is critical to the coupling of the receptor and G-protein. Thus, a simple structural model can provide a mechanistic basis for receptor activation and also indicate that molecules that do not make these specific interactions are likely not to function as activators, but rather (as in the case of propranolol) as antagonists. Similar studies with receptors for the polypeptide hormone substance P and its nonpeptide antagonist CP 96 345 reveal that they occupy binding sites that are partially distinct (Fig. 24). Residues in transmembrane

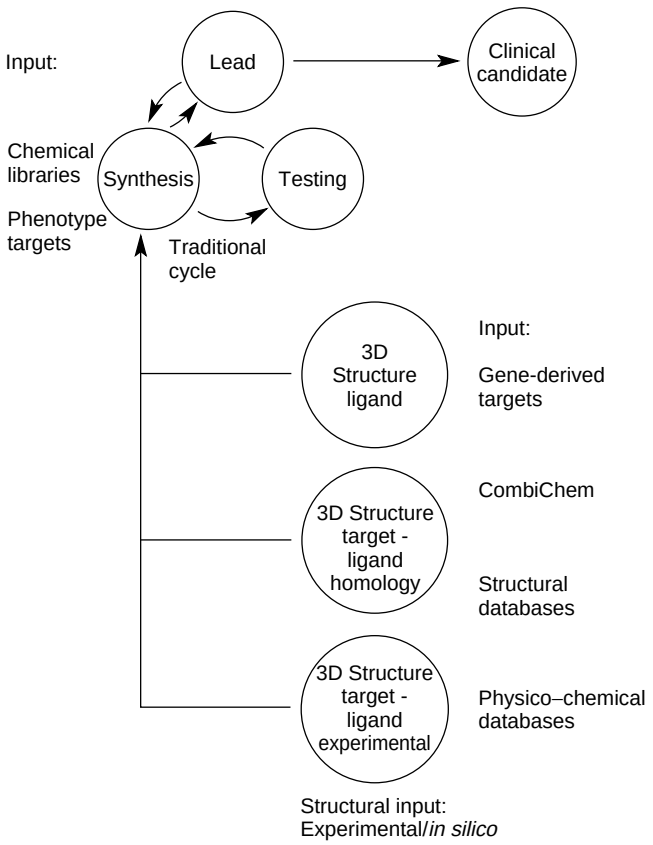

Fig. 20 The incorporation of structure-based drug design into the process of drug discovery. The traditional cycle of discovery (top left) is aided by the components of structure-based design that include inputs from both experimental and *in silico* structural design and from a structural and physicochemical databases.

domains 5 and 6 and the extracellular epitope between these domains are clearly critical for the nonpeptide antagonist, but not for peptide–agonist binding.

Ion channels are an important class of pharmacological receptors and are the loci of action of many drugs and toxins including, tetrodotoxin, pyrethroids, calcium channel antagonists, snake venoms and type II antidiabetic drugs (Fig. 25). The understanding of both channel function and drug action is being greatly facilitated by the availability of three-dimensional structures of several potassium channels. These studies, generally confirmatory of some fifty years of electrophysiological work, have elucidated the nature of the channel pore through which K^+ ions selectively flow, the subunit organization of the channel, the receptor site for quaternary ammonium inhibitors (NEt_4^+), and the nature of the voltage-sensing mechanism (Fig. 26).

These structure-based approaches are increasingly complemented by "expert" computer-based algorithms that both

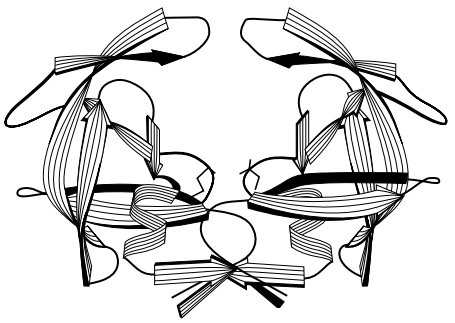

Fig. 21 The ribbon backbone of HIV-1 protease, depicting the vertical twofold axis of symmetry and the active site aspartate residues in the middle. (Reproduced with permission from the American Chemical Society from Greer, J., Erickson, J.W., Baldwin, J.J. and Varney, M.D. (1994) *J. Med. Chem.* **37**, 1035.)

generate *and* screen, on a virtual basis, structures that fit protein-receptor sites. Complemented with algorithms that predict pharmacokinetic characteristics (ADME) true *in silico* drug design is likely to be an important component of actual drug discovery in the coming decade.

3.4 Peptidomimetics

The physiologically important ligands for many receptors and enzymes are peptides,

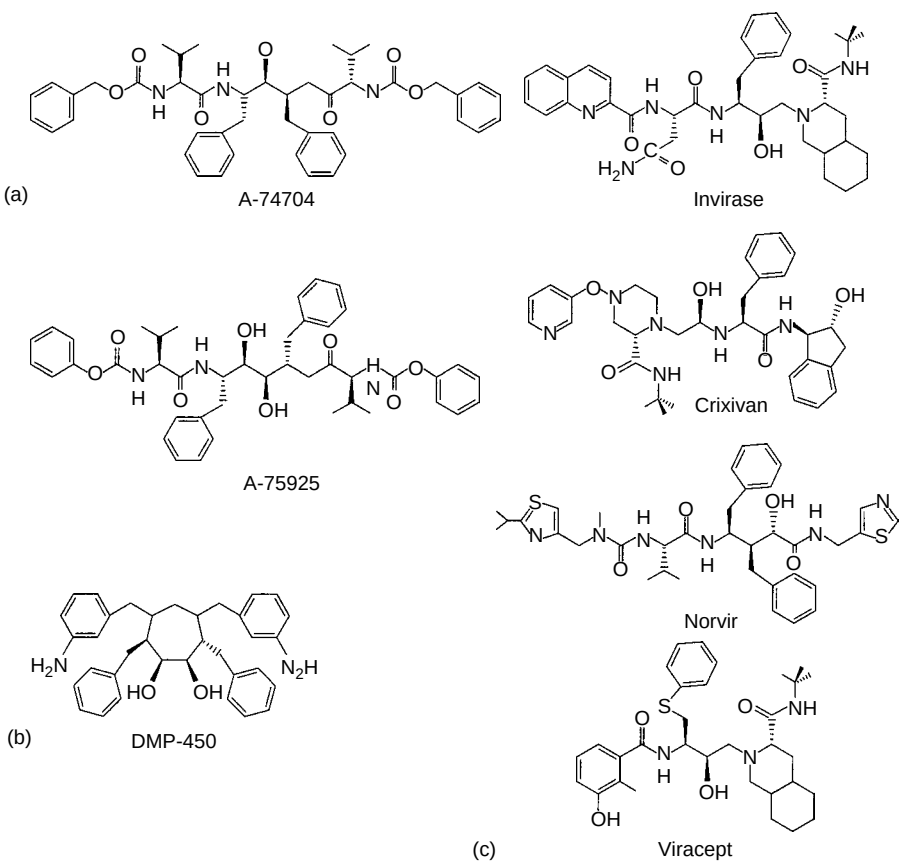

Fig. 22 HIV-1 protease inhibitors.

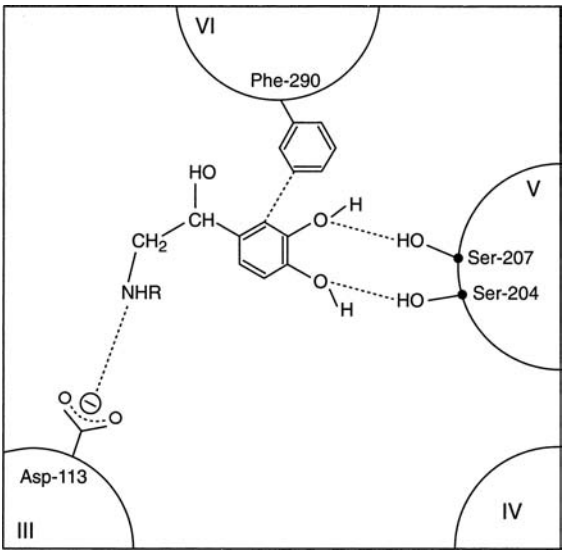

Fig. 23 Schematic representation of the interaction of a catecholamine with the β-adrenergic receptor and depicting specific functional interactions with residues on transmembrane helices III, V, and VII.

	Wild type NK-1	CR NK1 (NK3-TM7)	CR NK1 (NK3-TM6-7)	CR NK1 (NK3-TM5-7)	Wild type NK-3
Substance P	0.14 ± 0.03	0.49 ± 0.04	1.1 ± 0.2	0.9 ± 0.1	300 ± 80
Eledoisin	16 ± 5	4.6 ± 1.3	12 ± 3.5	4.5 ± 0.6	4.7 ± 1.1
(CP 96345)	14 ± 6	5.4 ± 0.9	330 ± 60	>>10 000	>>10 000

Fig. 24 Substance P and the nonpeptide antagonist CP 96 345 interactions at chimeric NK-1/NK-3 receptors. (Data from Getler, Nature 362 : 345, 1994.)

and the impact of biotechnology on drug discovery has given further emphasis to peptides, both as drugs and as lead structures. However, the disadvantages of short biological half-life, poor bioavailability, and inability to penetrate the central nervous system generally outweigh the potential advantages of potency and selectivity. The search for peptidomimetics – small synthetic molecules that mimic peptide activity – is an increasingly important component of medicinal chemistry. The opiate alkaloids, including morphine, are examples of such peptidomimetics: they

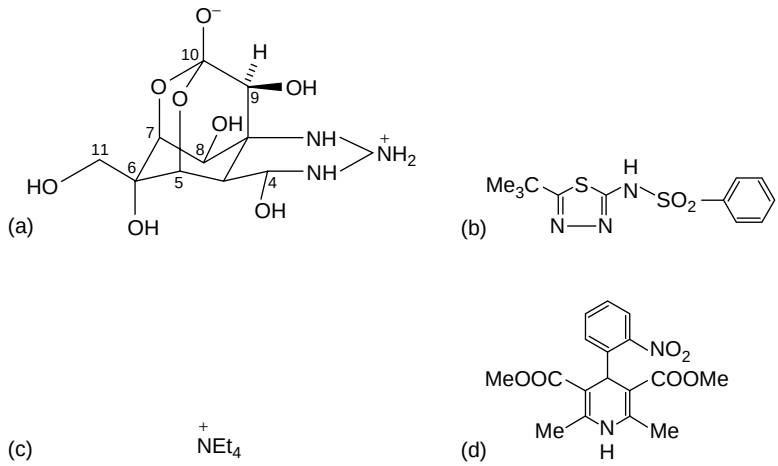

Fig. 25 Drugs that interact with ion channels (a) tetrodotoxin at Na^+ channels, (b) glybuzole at K^+_{ATP} channels, (c) tetraethylammonium at K^+ channels and (d) nifedipine at Ca^{2+} channels.

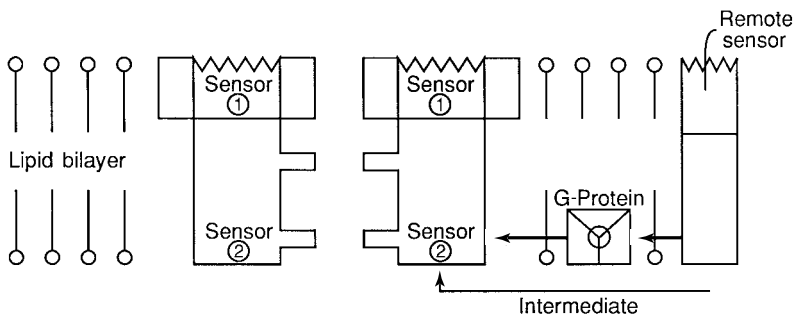

Fig. 26 Schematic representation of an ion channel depicting the pore with "gates" that open and close in response to chemical or physical stimuli, intra- and extracellular sensors associated with the subunits of the ion channel (which is invariably an oligomeric construct) and a remote sensor that is coupled to the channel via cytosolic or membrane signals.

illustrate the principles of peptidomimetic design and demonstrate that peptides and peptidomimetics are frequently structurally quite distinct (Fig. 27). Indeed, the activities of morphine and related compounds were well described long before the endogenous opiod peptides had been discovered!

Peptidomimetics may be considered in three principal categories. *Reaction pathway mimetics* are chemical species that mimic an intermediate, typically a transition state, in an enzyme-catalyzed pathway. *Peptide topographic intermediates* are chemical entities that mimic the topography of non-adjacent peptide regions, and *peptide backbone mimetics* mimic local peptide topography (amide bonds, β-turns etc).

The naturally occurring pepsin inhibitor, pepstatin, contains the amino acid statine.

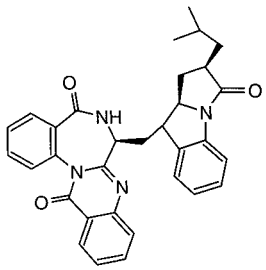

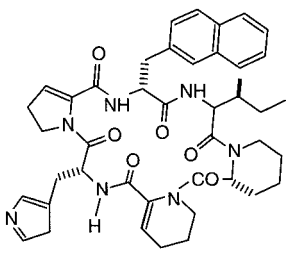

Fig. 27 Examples of nonpeptides that interact at peptide receptors: asperlicin at the cholecystokinin receptor; an oxytocin antagonist interacting at the oxytocin receptor; morphine acting at the opiate receptor; losartan interacting at the angiotensin II receptor.

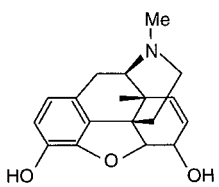

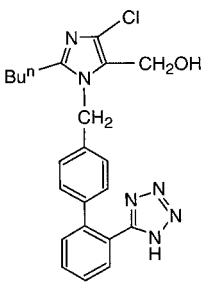

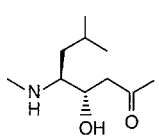

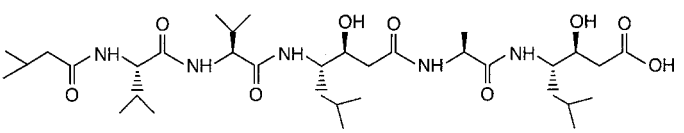

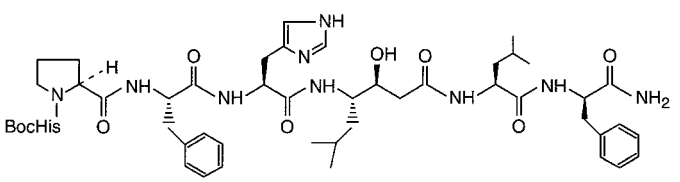

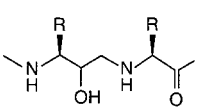

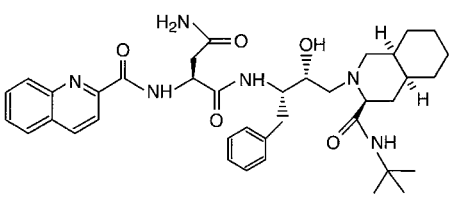

Fig. 28 The statine and hydroxyethylamine structures as peptide mimetics incorporated into the enzyme inhibitors pepstatin, L-363,564, and Ro-31-8959.

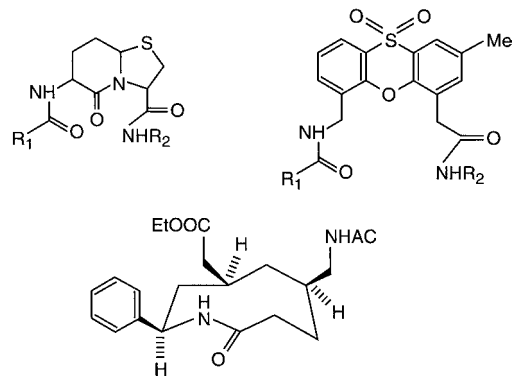

Fig. 29 Examples of peptidomimetic scaffolds.

Statine mimics a dipeptidyl tetrahedral hydrolytic transition state: this mimicking of a transition state likely underlies the ability of pepstatin to inhibit a variety of aspartic proteases, including renin and human immunodeficiency virus (HIV-1). The incorporation of the statin residue into the renin sequence was an important first step in the generation of potential renin inhibitors (Fig. 28). Other replacements serving a similar role in peptidomimetic design include the hydroxyethylene and dihydroxyethylene functionalities, which have generated potent renin inhibitors, and the hydroxyethylamine dipeptide mimetics that are used for generating potent HIV-1 protease inhibitors.

Peptide topographic intermediates are represented by an increasingly large group of compounds, frequently of natural product origin that compete with or mimic peptide actions. Prominent examples include asperlicin, which interacts at cholecystokinin receptors, imidazoles that interact at angiotensin II receptors, and, of course, morphine and its analogs that interact at opiate receptors (Fig. 27).

As part of the process of drug discovery, considerable effort has been devoted to the discovery of compounds that may mimic local protein conformations including β-turns and α-helices. The availability of such structures facilitates the construction of peptidomimetic scaffolds that can be decorated with the appropriate functional groups to generate receptor and enzyme selectivity. Some examples are shown in Fig. 29. The presence of the benzodiazepine nucleus in a variety of receptor-selective ligands is a likely example of such a peptide backbone–mimetic process (Fig. 12).

3.5 Protein–Protein Interactions

The traditional targets of drug design and discovery have been receptors and enzymes, typically viewed as acting independently and communicating through soluble biochemical intermediates, and in which small molecules are the effectors of the function. Increasingly, it is being realized that protein–protein interactions serve as critical mediators of cellular responses. Many receptors, enzymes, and ion channels function as modules composed of communicating subunits, and cells possess anchoring structures that serve to localize and spatially coordinate these protein–protein signaling complexes. Additionally, many receptors such

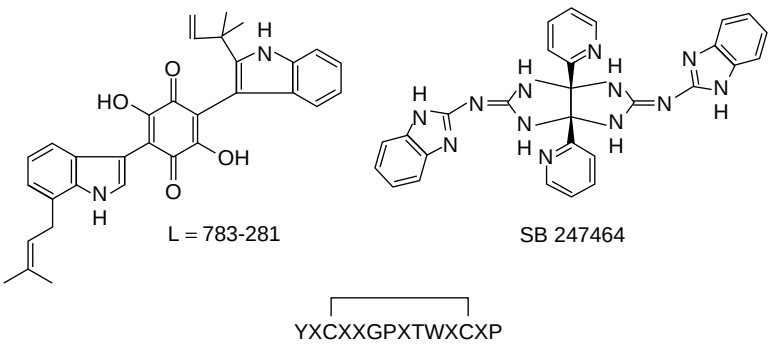

Fig. 30 Small-molecule mimetics of: insulin (L-783-381); erythropoietin (SB 24 7464) and granulocyte-stimulating hormone (cyclic peptide).

as tyrosine kinases that exist as monomers form oligomers under the influence of the activating hormone. The analysis of protein–protein interactions has been greatly facilitated by such technologies as the yeast two-hybrid system. And an increasing number of protein–protein interactions will be identified through the genome project. Hence, protein–protein interactions have become important drug targets. However, the relatively large size of interacting domains in such interactions has raised questions about whether small molecules can effectively interfere with such broad interactions, and whether such molecules would also possess sufficient specificity of action. However, an increasing number of such small molecules are known. For example, Aggrestat and Integrilin mimic the adhesive RGD epitope of fibrinogen interaction with GPIIb/IIIa, and are employed to reduce restenosis in coronary angioplasty. It can be anticipated that advances in our understanding of protein folding and construction will facilitate the design of small-molecule modulators of such interactions.

Of particular interest are studies that have demonstrated that protein hormones can be effectively reduced to small-molecule surrogates that effectively *mimic* the activating properties of the parent hormones. These include small-molecule mimics of insulin, erythropoietin, and granulocyte-stimulating hormone (Fig. 30).

An increasing number of diseases are now known to be associated with protein misfolding, including such degenerative disorders as Alzheimer's and the spongiform encephalopathies. Similarly, in cancer, the p53 gene is frequently mutated giving rise to mutant p53 protein. Small molecules that can restore the wild-type protein conformation may therefore serve as useful therapeutic agents.

3.6 Quantitative Structure–Activity Relationships

The application of quantitative structure–activity relationships rests on the accepted thesis that a mutual structural complementarity exists between the ligand and its binding site and that the biological properties of the ligand are a function of its discrete physicochemical properties, including pKa, electronic configuration, lipophilicity, polar surface area, molar volume, and stereochemistry. An early application of this principle is the

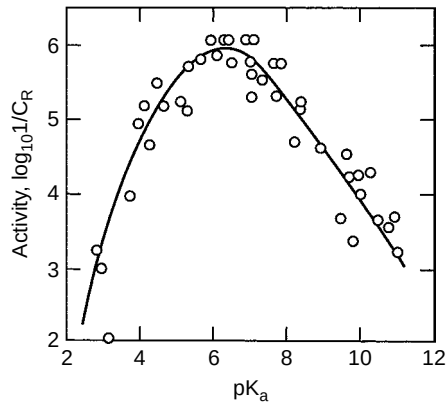

Fig. 31 The relationship between antibacterial activity of sulfonamides (log1/CR) and the pKa of the sulfonamide NH group. (Reproduced with permission from the American Chemical Society from Bell, P.H. and Roblin, R.O., (1942) *J. Amer. Chem. Soc.* **64**, 2905.)

inverse linear relationship observed by Overton and Meyer between water solubility and narcosis. A similarly simple relationship between structure and function was observed in the antibacterial sulfonamides, in which the bell-shaped curve of Fig. 31 indicates that the sulfonamide is active as the anionic species, but permeates the bacterial cell as the neutral molecule.

The definition of QSARs owes much to the pioneering work of Corwin Hansch who, building on the linear free energy relationship expressed in the Hammett equation of physical–organic chemistry, was able to correlate biological activity with hydrophobic, steric, and electronic substituent constants, through multiple linear regression analyses. Thus, for the inhibition of the enzyme acetylcholinesterase by a series of diethyl (substituted phenyl)phosphates:

$$\log 1/C_{50} = 5.77\sigma + 2.71 \qquad (3)$$

indicating that the activity in this homologous series of compounds is determined solely by the electronic properties of the substituent in the phenyl ring. More commonly, multiparameter equations are used, giving, for example, the following relationship for the local anesthetic activity of a series of 2-dialkylaminoethylbenzoates:

$$\log 1/C_{50} = 0.58\pi - 1.26\sigma + 0.96 \qquad (4)$$

A more complex relationship is shown in Fig. 32 for a series of substituted 1,4-dihydropyridine calcium channel blockers, selected members of which are widely employed as cardiovascular agents. Activity is dependent upon the hydrophobic, electronic, and steric properties of the aryl substituents.

The QSAR approach in several variations has become a routine component of the practice of medicinal chemistry. There are, however, weaknesses associated with the method: substituent constants must be available and a large and varied experimental data set is necessary to have a statistically powerful analysis. Furthermore, the method defines only the properties of the ligand and usually within a restricted structural set. Thus, the method tends to be retrospective and interpolative rather than predictive and extrapolatable. Nonetheless, its usefulness should not be underestimated, particularly in conjunction with other components of rational drug design.

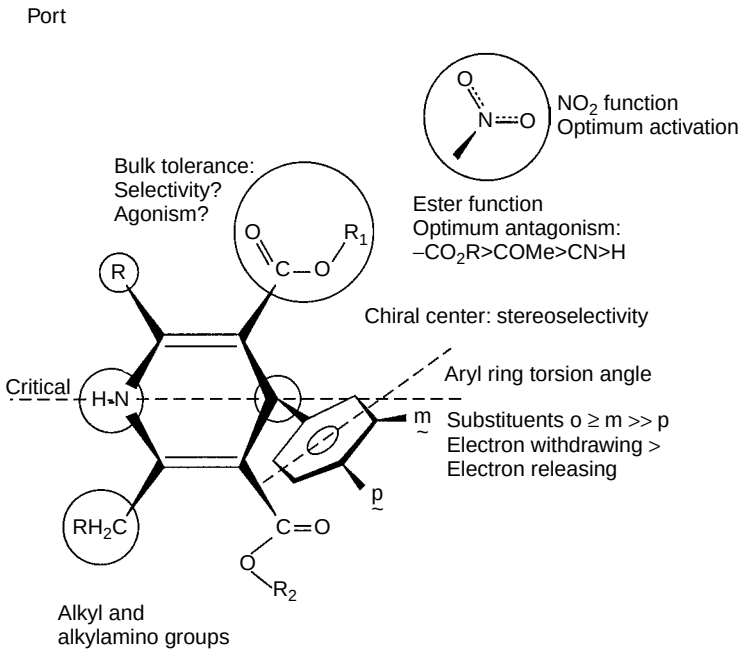

$$\log 1/C = 0.62\pi + 1.96\sigma_{meta} - 0.44L_{meta} - 3.26B_{1para} - 1.51L_{meta} + 14$$
$$n = 46,\ r = 0.90$$

Fig. 32 The structure–activity relationship for aryl-1,4-dihydropyridines active at the L-type voltage-gated Ca^{2+} channel. The structural formula depicts the general structural features that confer antagonist or activator properties. The equation describes the antagonist activity of a series of 1,4-dihydropyridines with the general formula -2,6-disubstituted-3,5-dicarboalkoxy-4-[substituted phenyl]-1,4-dihydropyridine.

3.7 Stereochemistry

That drug–receptor interactions exhibit stereoselectivity has been well recognized since the time of Pasteur. Indeed, so common is such stereoselectivity that it is generally considered to be an important index of the specificity of such interactions. This stereoselectivity of interaction represents a challenge to our definition of drug actions long recognized chemically, but now increasingly encountered at the clinical and regulatory levels. Clinical significance derives from considerations of the efficacy of a single enantiomer versus a racemate, from considerations of stereoselective metabolism and disposition, and from the impact of the route of administration and interpatient variability of response. Regulatory issues derive from considerations that racemic drugs can represent separate agents in fixed combinations, and development issues derive from considerations of costs, including those for chemical synthesis, and

of pursuing a single enantiomer versus a racemate. (We consider here only optically active isomerism: the same principles apply, however, to geometrical and other isomers. Also, for the sake of brevity, we consider only the case of molecules with a single chiral (stereogenic) center).

Enantiomers can differ both quantitatively and qualitatively in their biological activities. At one extreme, one enantiomer may be totally devoid of activity and at the other extreme both enantiomers may have potent, but distinct, activities. Thus, the following categories of activity are possible:

1. Both the isomers are equally active and there is no stereoselectivity of action. This situation is not common.
2. The isomers differ quantitatively in their activity. Most drugs that occur as stereoisomers fall into this category (in the limiting situation one enantiomer possesses all of the activity).
3. The isomers differ qualitatively in their activity and have discrete biological activity at the same or distinct receptors. Such behavior may translate to agonist and antagonist behavior at the same receptor, subtype selectivity, or activity at distinct receptor classes.
4. The behavior of the isomers in the racemate is different from that of the enantiomers alone.

Examples of racemates in which the enantiomers differ quantitatively in their activity are quite common. They include the positive inotrope dobutamine, the anti-inflammatory ibuprofen, the anticoagulant warfarin, and the local anesthetic bupivacaine. The differences in activity may be large enough to regard one enantiomer as effectively inactive: examples include the β-blocker propranolol in which the *S*-enantiomer is active and the cholinergic agonist bethanchol in which the *S*-enantiomer is also the active entity. Enantiomers that differ qualitatively in their activities include: nicotine where the *R*- and the *S*-enantiomers are substrates and inhibitors respectively of the enzyme *N*-methyltransferase; the 1-methyl-5-phenyl-5-propylbarbituric acid enantiomers in which the S- and R-forms are convulsant and anticonvulsant respectively; propoxyphene in which the enantiomers are analgesic and antitussive; sotalol in which the enantiomers are a β-adrenergic blocker and a K^+ channel blocker; and the enantiomers of Bay K 8644 which are activator and antagonist. Some examples are shown in Fig. 33.

In principle, stereoselectivity of drug action is derived from *both* pharmacodynamic and pharmacokinetic processes and can occur at any or all of the stages depicted in Fig. 34: transport to and from the site of action, storage in depots, interaction with binding proteins, interaction with receptors, metabolic enzymes, and transport in excretory pathways. The Ca^{2+} channel antagonist verapamil is available only as the racemate and is employed as an antianginal, anti-hypertensive and as a selective Class IV antiarrhythmic. The more active *S*-enantiomer is cleared more rapidly through first-pass metabolism. Accordingly, verapamil offers a different clinical profile according to whether it is administered orally or intravenously, being more active in the latter case. Even more subtle distinctions can be made in a comparison of the enantiomers of verapamil on cardiac and vascular smooth muscle: stereoselectivity is very low in the latter and significantly higher in the former (Table 2).

Important differences can exist between the toxicologic or side-effect profiles of enantiomers. For L-DOPA, the levo (available) form is less toxic, vomiting

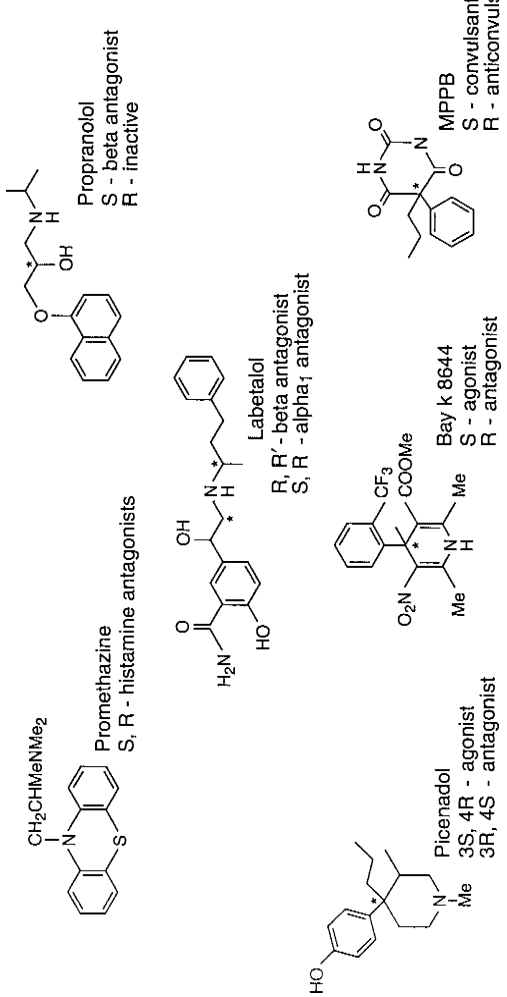

Fig. 33 The stereoselectivity of drug action depicting the quantitative and qualitative differences in biological activity that can exist between enantiomers.

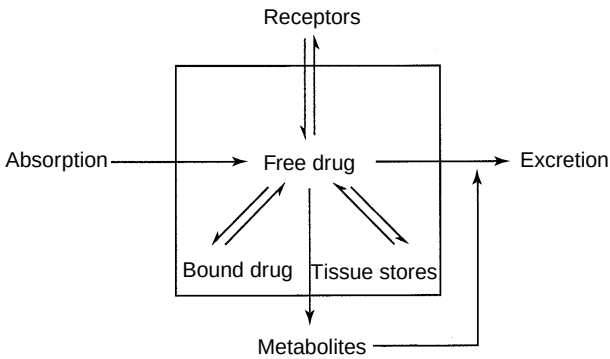

Fig. 34 The observed pharmacological activity of a drug depends upon both pharmacodynamic and pharmacokinetic properties as depicted. Each of the processes depicted has its own structural and stereochemical requirements. These requirements are not necessarily identical at each site and, in fact, may well conflict.

Tab. 2 Stereoselectivity of Verapamil in the cardiovascular system.

	Vasodilation	*Cardiodepression*
	−log EC50 (M)	−log EC50 (M)
Enantiomer		
R(+)−Verapamil	7.31	5.7
S(−)−Verapamil	7.70	7.2

is associated with the D-isomer of levamisole, and the teratogenic effects of thalidomide may be associated with the *S*-enantiomer (however, this has no clinical significance since the enantiomers epimerize very rapidly in solution). Stereoselectivity of general anesthesia has been demonstrated for the volatile agents isoflurane and halothane; although modest, this difference may be of clinical significance and expand the safety margin of these agents.

These considerations are part of an increasing scientific, clinical, and regulatory pressure to evaluate the chirality of drug action at all stages of the drug development process. The use of single enantiomer drugs continues to increase as do conversions of racemates to single enantiomers. Even those in which a racemate may be introduced clinically, this will be done only after the properties of the enantiomers have been thoroughly investigated, and in which it can be demonstrated clinically that there is no disadvantage in the use of a racemate rather than a single enantiomer.

4 The Influence of Molecular Biology

The principles and tools of molecular biology exert an increasing influence on the theory, practice, and application of medicinal chemistry. Molecular biology provides not only the structures of the receptors and targets for drug action, but validates these targets, defines disease states, provides the logic of chemical design, generates the rapid screening protocols, and directly provides drugs and drug candidates. These principles have been elaborated upon previously, but some brief additional comments follow. More

comprehensive discussion will be found elsewhere in this Encyclopedia.

4.1 The Expanded Genetic Code

Our existing genetic code contains 4 nucleotides that specify 64 unique nucleotide triplets that in turn encode the 20 amino acids that are the building blocks for all the proteins in at least this corner of the universe, apart from the rare organism that incorporates selenocysteine and pyrrolysine. The ability to incorporate novel amino acids into proteins has the potential to answer important questions concerning the evolution of life, produce new proteins with novel biological functions and activities, *and* to produce organisms that are biologically unique. Amino acid analogs have been incorporated into proteins by a variety of techniques including posttranslational modification and the use of chemically aminoacylated t-RNA that can be recognized by existing stop codons. However, the generation of autonomous bacteria with an expanded genetic code that can incorporate nonnatural amino acids with efficiency and fidelity represents a major step in harnessing the genetic code for the *in vivo* design and production of novel potential protein therapeutics.

4.2 Genes as Drugs

In principle, gene insertion to replace a defective gene or genes or to enhance expression of a particular gene product provides a permanent curative treatment. The critical issues here for effective gene delivery are the nature of the disease itself, the delivery system and its specificity, the extent and duration of the resultant gene expression, and the safety of the process. Additionally, there are significant ethical issues surrounding gene therapy that are absent with conventional drugs. A number of the potential targets for gene therapy are given in Table 3: clearly, the most plausible targets are monogenic diseases where the disease is due to the presence or misexpression of a single gene.

Gene therapy may be carried out *in vivo* or *in vitro* and the insertion methods are of several kinds. They include adenoviruses and retroviruses that are used as insertional carriers, liposomes to transfer plasmid DNA, and other physical methods including conjugation of the DNA with a protein that targets a specific receptor through which it is internalized.

Although there are in excess of some six hundred protocols, the majority of them being in the field of cancer, clinical success has been limited thus far. In particular,

Tab. 3 Targets and potential targets for gene therapy.

Monogenic diseases
X-Linked severe combined immunodeficiency
ADA deficiency
Cystic fibrosis
Mucopoylaccharidosis
Factor IX deficiency (Hemophilia B)
Familial hypercholesterolemia
Chronic granulomatous disease
Gene types transferred in clinical trials
Cytokines 26%
Antigen 14%
Tumor suppressor 12%
Suicide 8%
Deficiency 8%
Drug resistance 6%
Receptor 4%
Replication inhibitor 3%

Source: Data from Journal of Gene Medicine Database, 2004: www.wiley.co.uk/genmed/clinical/

the issue of the specificity of the gene insertion process has assumed particular importance, with observations that individuals being treated for X-linked severe combined immunodeficiency syndrome ("bubble boy" disease) have developed leukemia as a consequence of random insertion of the therapeutic gene near the LM02 gene, a gene linked to leukemia.

4.3 Antisense and Small Interfering RNA

Oligonucleotides bind with great specificity to a variety of nucleic acid and protein targets. Additionally, nucleic acids are particularly well suited for combinatorial chemistry *and* selection processes: oligonucleotides selected from such a process and referred to as *aptamers* can serve both as targeting species (research reagents) to identify small-molecule ligands, and as actual drug species. For example, by using an iterative cycle of RNA library generation, target incubation, removal of non-binding RNA, recovery of bound RNA, and amplification of binding molecules by RT-PCR (Fig. 35), these techniques have led to the identification of small-molecule inhibitors of HIV-1 replication and platelet-derived growth factor binding.

In principle, oligonucleotides with specificity conferred by Watson–Crick pairing will constitute highly specific drugs for defined nucleic acid RNA targets. Thus, an oligonucleotide of some 15–19 bases will have a unique sequence relative to the entire human genome and could, theoretically, offer absolute specificity of interaction. Such oligonucleotides will base-pair to a specific segment of RNA to block translation, block splicing, or facilitate degradation through RNAseH. This is not necessarily realized in practice since the oligonucleotide does not have equal access to all sites on the target RNA. Additionally, oligonucleotides also interact with proteins in a non-antisense manner. Many derivatives of phosphodiester oligonucleotides have been examined, particularly the backbone-modified methylphosphonate, phosphorothioate, and phosphorodithioate analogs (Fig. 36). These modifications serve to enhance chemical stability, reduce non-specific interactions, and enhance cell permeation. However, only one antisense therapeutic, Vitavene(™), is currently clinically available.

RNA interference is a gene-silencing mechanism originally discovered in lower organisms, but now shown to be widely applicable and to have considerable potential

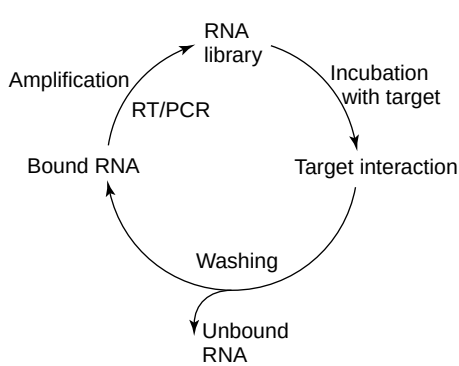

Fig. 35 An iterative cycle whereby RNA molecules that bind to a selected target are identified, amplified and subject to a further round of binding and selection to amplify the desired property.

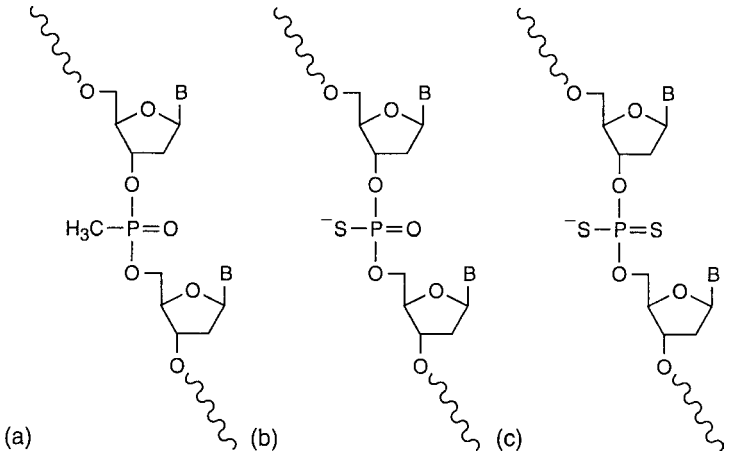

Fig. 36 Backbone-modified oligonucleotides (a) methyphosphonates, (b) phosphorothioates, and (c) phosphorodithioates.

for both target validation and for the generation of actual drugs. In lower organisms, the introduction of double-stranded RNA is followed by enzymatic processing to yield short, 21–22 double-stranded (ds) sRNAs, that guide sequence-specific degradation of the homologous mRNAs. The process is designed to recognize the introduction of RNA from foreign genes during, for example, virus infection, or when transposons and transgenes are randomly inserted. RNA interference (RNAi) can be produced in vertebrates by the introduction of short dsRNA, and serves as a mechanism for specifically silencing gene function. RNAi offers major promise for target validation, being faster than the production of animal knockouts, and the ultimate promise for actual drug delivery. It offers, in principle, significant advantages over antisense technology, since much lower concentrations of the dsRNA are needed, the effects are longer lasting and, of particular significance, siRNA can be produced intracellularly from RNA polymerase III promoters.

Despite the attractiveness of these nucleic acid–based technologies, it cannot be assumed that they are automatically superior to the existing small-molecule techniques. In particular, in protein targets that are part of an oligomeric complex, elimination of one protein could lead to the destruction of that complex and a cascade of effects that would not be produced by the action of a single small-molecule alone acting on the complex.

4.4 Gene-based Drug Targeting and Delivery

Gene-based mechanisms can also be used to confer or to enhance drug sensitivity in an unresponsive cell type. The transfer to cells, for example, of the herpes simplex thymidine kinase(HSV-tk) confers sensitivity to such nucleoside analogs as the antiherpes drug, ganciclovir. Such transfer can be mediated by a retrovirus that incorporates only into actively dividing cells and thus offers particular selectivity for malignant cells. HSV-tk converts the nontoxic ganciclovir into the phosphorylated derivative, which is a chain terminator of DNA synthesis.

5 Summary

Medicinal chemistry, a discipline firmly rooted in synthetic organic chemistry, has increasingly close relationships to structural and computational chemistry, particularly to molecular biology at the discovery interface, to toxicology and pharmacology at the development interface, and to medicine at the clinical interface. The linkage to molecular biology has assumed major significance in the past decade, and chemistry has increasingly been adopting characteristics that are dictated by the basic themes that drive biology – *diversity, replication, evolution,* and *self-organization*. It is entirely conceivable that, as this biological paradigm increasingly permeates science and, in particular, medicinal chemistry, we shall ultimately have Paul Ehrlich's "magic bullet" – self-synthesizing, self-replicating, self-evolving, and self-targeting nanomachines that circulate in our body, constantly repairing damage and destroying biological intruders.

See also Antitumor Agents: Taxol and Taxanes – Production by Yew Cell Culture; Gene Targeting; Oncology, Molecular; Pharmacogenomics and Drug Design.

Bibliography

Books and Reviews

Anderson, A. (2003) The process of structure-based drug design, *Chem. Biol.* **10**, 787–797.

Drug Discovery Today. (2002) *20th Anniversary of Combinatorial Chemistry*. Parts (1, 2). 7: Issues 1 and 2.

Gohlke, H., Klebe, G. (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors, *Angew. Chem., Ed. Int.* **41**, 2644–2676.

Gordon, E.M., Kerwin, J.F. (Eds.) (1998) *Combinatorial Chemistry and Molecular Diversity in Drug Discovery*, Wiley-Liss, New York.

Houk, K.N., Leach, A.G., Kim, S.P., Zhang, Y. (2003) Binding affinities of host-guest, protein-ligand, and protein-transition-state complexes, *Angew. Chem., Int. Ed.* **42**, 4872–4897.

Lough, W.J., Wainer, I.W. (Eds.) (2003) *Chirality in Natural and Applied Science*, Blackwell Publishing, Oxford, UK.

Rothstein, M.A. (Ed.) (2003) *Pharmacogenomics. Social, Ethical and Clinical Dimensions*, Wiley-Liss, New York.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G.,. Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Wess, G., Urmann, M., Sickenburger, B. (2001) Medicinal chemistry: challenges and opportunities, *Angew. Chem., Int. Ed.* **40**, 3341–3350.

Wiesner, J., Ortmann, R., Jomaa, H., Schlitzer, M. (2003) New antimalarial drugs, *Angew. Chem., Int. Ed.* **42**, 5274–5293.

Primary Literature

Breinbauer, R., Vetter, I.R., Waldmann, H. (2002) From protein domains to drug candidates – natural products as guiding principles in the design and synthesis of compound libraries, *Angew. Chem., Int. Ed.* **41**, 2878–2890.

Carroll, F.I. (2003) Monoamine transporters and opiod receptors. Targets for addiction therapy, *J. Med. Chem.* **46**, 1775–1794.

Chanda, S.K., Caldwell, J.S. (2003) Fulfilling the promise: drug discovery in the post-genomic era, *Drug Discovery Today* **8**, 168–174.

Crooke, S.T. (2000) Potential roles of antisense technology in cancer chemotherapy, *Oncogene* **19**, 6651–6659.

Croston, G.E. (2002) Functional cell-based uHTS in chemical genomic drug discovery, *Trends Biotechnol.* **20**, 110–115.

Daly, J.W. (2003) Ernest Guenther award in chemistry of natural products. Amphibian skin: a remarkable source of biologically

active arthropod alkaloids, *J. Med. Chem.* **46**, 45–452.

Famulok, M., Mayer, G., Blind, M.. (2000) Nucleic acid aptamers – from selection *in vitro* to applications *in vivo*, *Acc. Chem. Res.* **33**, 591–599.

Flower, D.R. (1999) Modeling G protein-coupled receptors and signal transduction for drug design, *Biochim. Biophys. Acta* **1422**, 207–234.

Fredrickson, R., Lagerstrom, M.C., Lundin, L.-G., Schioth, H.B. (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints, *Mol. Pharmacol.* **63**, 1256–1272.

Freidinger R.M. (2003) Design and synthesis of novel bioactive peptides and peptidomimetics, *J. Med. Chem.* **46**, 5553–5566.

Gadek, T.R., Nicholas, J.B. (2003) Small molecule antagonists of proteins, *Biochem. Pharmacol.* **65**, 1–8.

Giannis, A., Rubsma, F. (1997) Peptidomimetics in drug design, *Adv. Drug Res.* **29**, 1–78.

Gohlke, H., Klebe, G. (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed.* **41**, 2644–2676.

Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B.T., MacKinnon, R. (2003) X-ray structure of a voltage-dependent K^+ channel, *Nature* **423**, 33–41.

John, V., Beck, J.P., Bienkowski, M.J., Sinha, S., Heinrikson, R.L. (2003) Human β-secretase (BACE) and BACE inhibitors, *J. Med. Chem.* **46**, 4625–4630.

Kay, B.K., Kurakin, A.V., Hyde-DeRuyscher, R. (1998) From peptides to drugs via phage display, *Drug Discovery Today* **3**, 370–378.

Kolb, H.C., Finn, M.G., Sharpless, K.B. (2001) Click chemistry: diverse chemical function from a few good reactions, *Angew. Chem., Int. Ed.* **40**, 2004–2021.

Lane, D.P., Hupp, T.R. (2003) Drug discovery and p53, *Drug Discovery Today* **8**, 347–355.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development, *Adv. Drug Delivery Rev.* **46**, 3–26.

Mehl, R., Anderson, C., Santoro, S.W., Wang, L., Martin, A.B., King, D.S., Horn, D.M., Schultz, P.G. (2003) Generation of a bacterium with a 21 amino acid genetic code, *J. Am. Chem. Soc.* **125**, 935–939.

Natesh, R., Schwager, L.U., Sturrock, E.D., Acharya, K.R. (2003) Crystal structure of the human angiotensin-converting enzyme-lisinopril complex, *Nature* **421**, 551–554.

Olivera, B.M., Hillyard, D.R., marsh, M., Yoshikami, D. (1995) Combinatorial peptide libraries in drug design: lessons from venomous cone snails, *Trends Biotechnol.* **13**, 422–426.

Patchett A.A. (2002) Natural products and design: interrelated approaches in drug discovery, *J. Med. Chem.* **45**, 5609–5616.

Patchett, A.A., Nargund, R.P. (2000) Privileged structures – an update, *Annu. Rep. Med. Chem.* **35**, 269–298.

Pearlstein, R., Vaz, R., Rampe, D. (2003) Understanding the structure-activity relationship of the human ether-a-go-go-related gene cardiac K^+ channel. A model for bad behavior, *J. Med. Chem.* **46**, 2017–2023.

Ruoslahti, E. (2000) Special delivery of drugs by targeting to tissue-specific receptors in the vasculature, *Pharm. News* **7**(4) 35–40.

Tan, D.S., Foley, M.A., Stockwell, B.R., Shair, M.D., Schreiber, S.L. (1999) Synthesis and preliminary evaluation of a library of polycyclic small molecules for use in chemical genetic assays, *J. Am. Chem. Soc.* **121**, 9073–9087.

Thompson, J.D. (2002) Applications of antisense and siRNAs during preclinical drug development, *Drug Discovery Today* **7**, 912–917.

Triggle, D.J. (2003) Medicines in the 21st century. Pills, potions, politics and profits. Where is public policy? *Drug Dev. Res.* **59**, 269–291.

Triggle, D.J. (2002) Ion channels: structure, function and pharmacology, in: Krogsgaard-Larsen, P., Liljefors, T., Madsen, U. (Eds.) *Textbook of Drug Design and Discovery*, Francis & Taylor, London, pp 173–204.

Tuschl, T. (2001) RNA interference and small interfering RNAs, *ChemBioChem* **2**, 239–245.

Walsh, C.T. (2002) Combinatorial biosynthesis of antibiotics: challenges and opportunities, *ChemBioChem* **3**, 124–134.

Weng, Z., DeLisi, C. (2002) Protein therapeutics: promises and challenges for the 21st century, *Drug Discovery Today* **20**, 29–35.

Wiley, R.A., Rich, D.H. (1993) Peptidomimetics derived from natural products, *Med. Res. Rev.* **13**, 327–384.

Zaman, G.J.R., Michiels, P.J.A., van Boeckel, C.A.A. (2003) Targeting RNA: new opportunities to address drugless targets, *Drug Discovery Today* **8**, 297–306.

Membrane Traffic: Vesicle Budding and Fusion

Michael J. Clague and Sylvie Urbé
University of Liverpool, Liverpool, UK

1 Overview of Cellular Trafficking Pathways 121
1.1 Maintaining Cellular Compartmentalization 121
1.2 Secretory Pathway 121
1.3 Endocytic Pathway 123
1.4 Principle Steps of Vesicular Transport 123

2 Methodologies 124
2.1 Subcellular Fractionation 124
2.2 Microscopy 124
2.3 Yeast Genetics 124
2.4 Cell-free Assays of Membrane Transport 125

3 Evolutionary Conservation of Fundamental Mechanisms 126

4 Signal-mediated versus Bulk Flow Transport 126

5 Coated Vesicle Formation 127
5.1 Coats 127
5.2 Adaptors 128
5.3 Principles of Cargo Selection 128
5.4 Induction of Curvature 129
5.5 Vesicle Scission 130

6 Vesicle Targeting 130
6.1 Rab Proteins 130
6.2 Tethering Molecules 131

7 Vesicle Fusion 131
7.1 Understanding Energy Barriers to Fusion 131

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

7.2 Membrane Destabilization 132
7.3 SNARE Proteins 132
7.4 Regulation of SNARE Complex Assembly 134

8 Vesicle Trafficking and Disease 134

9 Conclusions 135

Bibliography 136
Books and Reviews 136
Primary Literature 136

Keywords

Clathrin
A large protein complex with a triskelion structure built from three heavy and three light chains, capable of self-associating to form lattices, which coat various vesicles.

Coated Vesicle
60- to 100-nm diameter membrane vesicle that buds from a subcellular compartment, encased in a regular protein coat, which selects vesicle cargo and drives budding. Examples include clathrin-coated and COP-coated vesicles; COP refers to coat protein complex.

Endocytosis
Process by which material is taken up into the cell from the external environment, through incorporation into vesicles, which bud from the plasma membrane.

Membrane Fusion
Merger of lipid bilayer membranes, which is accompanied by the establishment of continuity between the aqueous interiors that are bounded by the originally distinct membranes.

Rab Protein
A member of a large family of small GTPases, which regulate vesicle trafficking, notably through the recruitment of tethering molecules.

Secretory Pathway
A series of compartments that are traversed by material, destined to be, secreted from the cell and which are connected by vesicular transport. Principally comprising the endoplasmic reticulum (ER), ER-Golgi intermediate compartment (ERGIC) Golgi apparatus, and trans-Golgi network (TGN).

SNARE Protein
Member of a conserved family of proteins believed to impart specificity to and catalyze intracellular vesicle fusion.

Tight control of the flux of membrane components through vesicle budding and fusion is essential to cellular organization, giving rise to subcellular compartments with distinct functions. Cargo is generally selected for inclusion into vesicles by specific sorting signals, which are recognized by adaptor proteins, that link to protein components of vesicle coats. The membrane deformation required for vesicle formation is promoted by the coat complex in conjunction with accessory proteins, which may partially insert into the bilayer membrane. Specific targeting of vesicles is initially mediated through "tethering" molecules and then cemented through interactions between SNARE proteins, which also catalyze fusion. The essential mechanisms are conserved throughout the cell and from yeast to man.

1 Overview of Cellular Trafficking Pathways

1.1 Maintaining Cellular Compartmentalization

The advent of electron microscopy revealed the complexity of subcellular organization, allowing a first glimpse of a bewildering array of membrane-bounded compartments (Fig. 1). This raised deep questions that are still pertinent to modern day cell biology. How are distinct organelles generated and maintained in the face of a constant flux of material? To what degree is this dependent on spatial organization versus biochemical specification? How should compartments be defined? by morphology, function, or biochemical composition? Much is accomplished through vesicle budding and fusion. In order for this to be so, there must be very specific mechanisms of cargo selection and targeting. Delivery of lysosomal degradative enzymes to the cell nucleus, for example, would court disaster.

1.2 Secretory Pathway

As its name implies, this pathway is taken by molecules destined for cellular secretion. Following translocation of nascent proteins across the ER membrane, proteins must traverse a series of compartments (Fig. 1), including the Golgi stack, where successive posttranslational processing events, such as glycosylation, may take place. The striking stacked cisternae of the Golgi represent a polarized structure, such that the cis-most and trans-most cisternae are highly enriched in distinct enzymes. The number of Golgi stacks may vary dramatically between cell types, but its function is nevertheless conserved.

The trans-Golgi network (TGN) represents the principal sorting station of the secretory pathway. From this compartment, material may be constitutively secreted (e.g. extracellular matrix proteins) or packaged into regulated secretory vesicles (e.g. insulin), which normally undergo

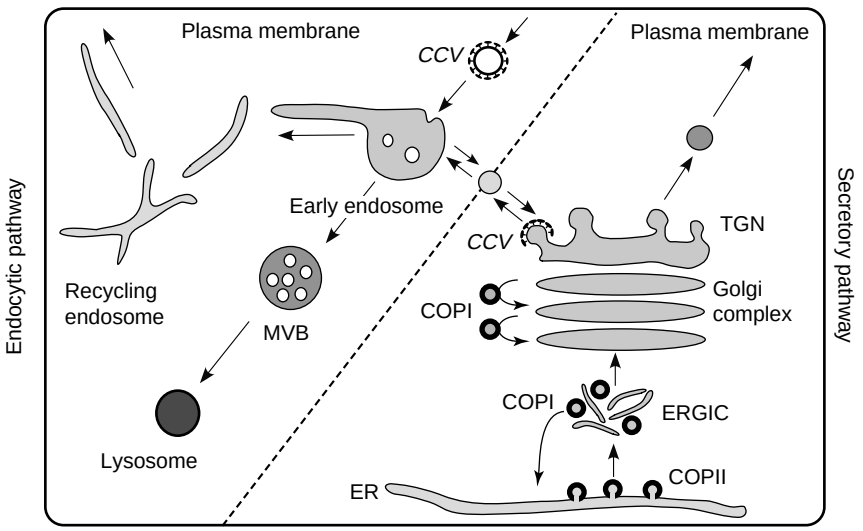

Fig. 1 Schematic diagram of cellular vesicular trafficking pathways in the eukaryotic cell. Newly synthesized proteins translocate into the endoplasmic reticulum (ER) and then progress along the secretory pathway, first by inclusion into COPII coated vesicles, which deliver material to the ER-Golgi intermediate compartment (ERGIC). Tubular elements of the ERGIC may coalesce to form a cis-Golgi cisterna. Anterograde transport toward the plasma membrane, is thought to proceed by maturation of Golgi cisternae, followed by budding of secretory vesicles from the trans-Golgi network (TGN). ER and Golgi resident enzymes are continually retrieved by inclusion into COPI vesicles, which move in a retrograde direction. Material internalized from the plasma membrane may be included in clathrin-coated vesicles (CCVs), which use the AP2 adaptor. The CCVs fuse with the early endosome from where internalized cargo can either be recycled back to the plasma membrane through the tubular recycling endosome or progress toward lysosomes by inclusion in multivesicular bodies (MVBs). Crossover between these two principal highways is mediated by vesicular transport between the TGN and endosomes. Note that other subcellular organelles, such as the nucleus and mitochondria, which do not form part of this endomembrane system are not depicted.

fusion with the plasma membrane (exocytosis) in response to elevated Ca^{2+} levels. Some proteins are diverted at this stage from the secretory pathway toward the endocytic pathway (e.g. lysosomal enzymes).

Progression along the secretory pathway is mediated by vesicular transport. This involves the selection and packaging of cargo into vesicles, which bud off from the limiting membrane and then fuse with specific target membranes. This mechanism can be used to designate forward (anterograde) transport of cargo or retrieval (retrograde transport) of specific components that function earlier in the pathway. Consider transport through the polarized Golgi stacks: according to one view, this is comprised of cisternae, each representing individual stable entities maintained by a delicate balance of retrograde and anterograde vesicular transport. However, it is currently considered more likely that individual cisternae progress through the Golgi complex, gradually undergoing maturation through the selective retrieval of ''earlier'' Golgi markers by incorporation into COP-coated vesicles. This ''cisternal maturation'' model of Golgi transport has

the advantage that it can account for the secretion of large molecules, such as algal scales or collagen, which are too big to be packaged into individual COP-coated vesicles.

1.3 Endocytic Pathway

The best-understood pathway for internalization of material at the plasma membrane is the clathrin-coated vesicle (CCV) pathway, first described by Roth and Porter in mosquito oocytes. However, it has now become clear that several other non-clathrin-mediated pathways operate in parallel. Once internalized, most material is transported to a tubulovesicular compartment, referred to as the *early or sorting endosome*. From this point, endocytosed proteins can enter a pathway for recycling (e.g. transferrin receptor) or be sorted toward the lysosome (e.g. EGF receptor), the major degradative compartment of the cell (Fig. 1). Lysosomal sorting involves the inward budding of cargo containing vesicles from the limiting membrane of sorting endosomes to create multivesicular bodies (MVBs), which then fuse either with late endosomes or directly with lysosomes. The recycling pathway from the early endosome probably involves tubular structures, which break off and direct material through a perinuclear recycling compartment back to the plasma membrane.

1.4 Principle Steps of Vesicular Transport

This article is limited to consideration of membrane trafficking mediated by small vesicles, for which common principles can be extracted. Other trafficking events, for example, generation of tubules for endosomal recycling or inward invagination of vesicles into the endosome lumen will not be considered here. The general mechanism of vesicle transport normally includes the following steps, which are depicted in Fig. 2. A coat assembles on an intracellular membrane through interactions with cargo proteins. Membrane deformation results in vesicular profiles, which must undergo scission to create coated vesicles. Upon uncoating, initial contact with target membranes is made through "tethering" molecules. Finally, the vesicle is committed to fusion through further molecular associations, which may provide the energy to overcome

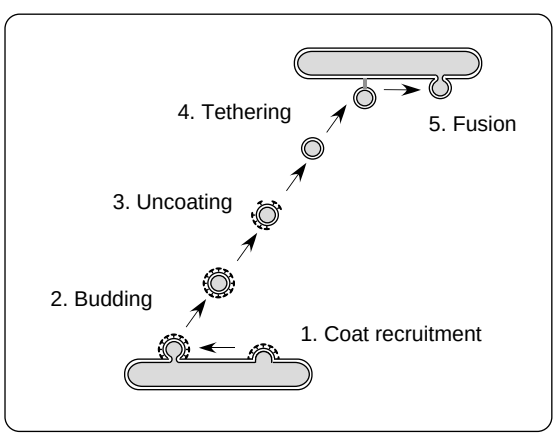

Fig. 2 Basic steps of the vesicle transport cycle (see text for details).

the barriers to membrane fusion, resulting in continuity of both lipid bilayers and lumenal contents.

2 Methodologies

2.1 Subcellular Fractionation

De Duve and colleagues pioneered the use of analytical centrifugation to separate subcellular compartments derived from homogenized tissue. By examining the concomitant separation of specific enzymatic activities, they were able to propose that biochemical pathways were more compartmentalized than previously imagined. Centrifugation techniques have been vital in the purification of various vesicle fractions, including clathrin-coated vesicles, COP-coated vesicles, and synaptic vesicles, thus allowing biochemical analysis of their composition. The introduction of mass spectroscopic techniques for proteomic analysis now allows for in depth profiling of vesicle-associated proteins. A recent study has identified 209 distinct proteins associated with highly purified brain-derived clathrin-coated vesicles.

2.2 Microscopy

The biochemical work of De Duve was brilliantly complemented by the morphological studies of Palade, Claude, Porter, and coworkers, who harnessed the power of electron microscopy to visualize the interior of the cell with a particular focus on the secretory pathway of the pancreatic acinar cell. Jamieson and Palade introduced an *in vitro* "pulse-chase" protocol to follow the secretory pathway of nascent proteins. A short incubation of cells with [^{3}H]-leucine selectively labels newly synthesized proteins, which can be visualized by autoradiography and subsequent electron microscopic examination. By varying the time of incubation in the absence of label (chase period) following the labelling "pulse," the progression of proteins through successive secretory compartments could be visualized. Palade frequently observed vesicular profiles in his micrographs and proposed that these may be intermediate vessels for transport between compartments. The introduction of immunocytochemistry and, in particular, the introduction of cryo-electron microscopy by Tokuyasu and Singer paved the way for a quantitative analysis of specific cargo proteins and structural components of vesicles. More recently, the ability to tag proteins with multiple colored fluorescent proteins has made it possible to directly observe protein dynamics through fluorescence microscopy of live cells. Thus, the flux of material in response to a given perturbation can now be followed in real time at the subcellular level.

2.3 Yeast Genetics

A genetic screen in yeast for mutants defective in secretion of specific enzymes identified 23 complementation groups that provided the first set of candidate genes (*sec* genes) for regulation of membrane traffic. The first *sec* gene to be cloned and characterized, *Sec4*, is a ras-like small GTPase belonging to the rab family that acts to target mature secretory vesicles to the plasma membrane. Further screens for mutants defective in vacuolar protein sorting (*vps* genes) or endocytosis (*end* genes) have produced equally informative inventories of genes responsible for specific regulation of membrane traffic.

2.4 Cell-free Assays of Membrane Transport

James Rothman pioneered the development of *in vitro* assays for the reconstitution of membrane trafficking steps. Rothman and coworkers first established a biochemical assay that reconstitutes a specific intracellular fusion event, namely, transport between cis and medial cisternae of the Golgi stack (Fig. 3). They combined wild-type Golgi fractions isolated from CHO cells with Golgi fractions isolated from mutant cells lacking a sugar-modifying (GlcNAc transferase) activity, but containing a model cargo protein, vesicular stomatitis virus glycoprotein (VSVG protein) by virtue of prior infection. This protein is a substrate for GlcNAc, but can only acquire [^{3}H]-radiolabeled GlcNAc by transfer between heterologous Golgi stacks. Rothman and coworkers were able to show that this transfer requires both cytosol and ATP, and thus embarked on a dissection of the cytosolic factors required to support the assay. A major insight derived from this work was that specificity of fusion is principally determined biochemically, rather than by relying on the architecture of the cell interior.

One drawback of this original assay was that it required generation of a vesicular intermediate, and essential factors could therefore operate at the level of vesicle budding, vesicle docking, or vesicle fusion. Subsequent refinements led to the

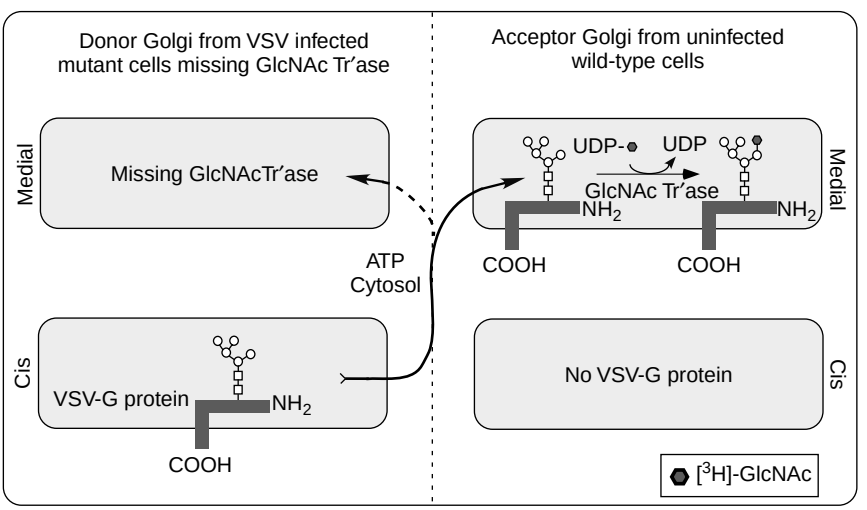

Fig. 3 Schematic diagram of the *in vitro* reconstitution of Golgi transport. The first *in vitro* assay of membrane transport opened the way for a biochemical dissection. Golgi fractions are prepared from two sets of homogenized tissue culture cells, one wild type and the other a mutant cell line lacking the medial-Golgi enzyme GlcNAc transferase, which adds a GlcNAc modification to the carbohydrate chain on secretory proteins. The mutant cells are infected with vesicular stomatitis virus (VSV), for which the spike glycoprotein VSVG provides a model secretory protein. VSVG is accumulated in the cis-Golgi cisternae prior to homogenization. The only way VSVG can acquire [^{3}H] labelled GlcNAc is through contents mixing between the medial-Golgi of wild-type cells and cis-Golgi of mutant cells. Under most conditions, this mixing requires COPI vesicle budding and fusion. This assay was used to fractionate cytosol for factors necessary to obtain efficient radiolabeling of VSVG and led to the identification of NSF, α-SNAP, and characterization of COPI coat proteins.

dissection of individual steps. Assays that reconstitute other transport steps have since been established and developed to the point where both vesicle budding and fusion can be reconstituted with simple model membranes and a minimal set of components. One powerful strategy has been to combine a biochemical approach with yeast genetics, notably in an assay of yeast vacuole fusion. This has allowed the manipulation of membrane proteins in the assay by deletion of specific genes in the parent yeast strains, which then enables the testing of the roles of individual proteins and the topological constraints for their participation in fusion.

3
Evolutionary Conservation of Fundamental Mechanisms

One of the earliest proofs of evolutionary conservation was the isolation of the NEM-sensitive factor (NSF), identified as a critical protein component in cytosol necessary to support mammalian Golgi transport *in vitro*. Rothman and colleagues showed that this protein is functionally equivalent to the *Sec18* gene product of the yeast *Saccharomyces cerevisiae*, which was known to be essential for vesicle-mediated transport from the endoplasmic reticulum to the Golgi apparatus. At one stroke, this not only suggested that NSF might operate at the heart of a fusion mechanism conserved from yeast to mammals but also that it may be required in multiple membrane transport events. Confirmation of this general role of NSF was obtained using *in vitro* assays of endosome fusion and of ER to Golgi transport. This led to the attractive idea that NSF may act as a component of a universal fusion machinery required for every intracellular fusion event. This notion, which is still held to be broadly true (only less direct than initially envisaged), raises the question of how specificity of fusion events is attained.

NSF is recruited to membranes by α-SNAP (Sec17), which in turn binds to a family of proteins called *SNAREs*, that are also conserved from yeast to mammals. SNAREs are believed to be key players in determining specificity of intracellular fusion events and are also the best candidates for catalyzing the membrane fusion event itself. Their intracellular distribution is such that each membrane-bounded compartment is associated with a specific complement of SNAREs that form the basis of a combinatorial code for fusion (see Sect. 7). Recent genome-wide analysis suggests that there are 24 yeast SNAREs with similar numbers in flies and worms. Thirty-five SNAREs have been identified in the human genome, wherein the additional SNAREs predominantly represent tissue-specific isoforms. The current view of NSF function is that it is an ATPase, which configures SNARE proteins into a fusogenic conformation, but is not actively engaged in the fusion step.

This evolutionary conservation has meant that biochemical studies carried out with mammalian cell fractions have dovetailed with yeast genetic studies. The large-scale yeast screens can identify proteins that unerringly have mammalian counterparts, which are identifiable through homology.

4
Signal-mediated versus Bulk Flow Transport

Selective inclusion into transport vesicles is thought to be a major mechanism for establishing discrete compartments. An alternative "bulk flow" model posits that

resident proteins are selectively retained, whilst vesicle contents are simply samples of the residual components of the parent organelle. To begin to understand this process, one must be able to compare the contents of the vesicle membrane with the parent membrane. Two approaches have been applied to this problem, quantitative immunoelectron microscopy and biochemical analysis of vesicle fractions. Thus, many studies have established that cargo is actively concentrated in clathrin-coated vesicles budding from the plasma membrane. This may be constitutive, as in the case of the transferrin receptor or regulated, as is the case for epidermal growth factor receptor (EGFR), which requires activation through ligand binding.

This issue has been more contentious when considering anterograde transport through the secretory pathway. Measurements using a glycosylated acyltripeptide as a "bulk flow" marker showed a high rate of secretion, suggesting that secretory proteins were nonselectively transported between the ER and the cell surface. However, subsequent analysis indicated that this may reflect a direct efflux pathway, following transport of the glycosylated peptide across the ER membrane to the cytosol.

Immunoelectron microscopical analysis of a SNARE protein (rBet1) has shown that it is enriched approximately sixfold in coat protein complex II (COPII) vesicles that bud from the ER. In contrast, in the same study, two soluble secretory proteins of pancreatic acinar cells, amylase, and chymotrypsinogen were not enriched in these vesicles and were instead concentrated further along the pathway in the tubular structures of the ER-Golgi intermediate compartment (ERGIC), where they are probably enriched by exclusion from COPI vesicles that mediate retrograde transport to the ER. In a separate study utilizing biochemical analysis, Malkus et al. have compared the content of COPII vesicles generated *in vitro* from yeast microsomal fractions with that of the starting material. They clearly show that a soluble protein, glycosylated pro-α-factor, is 20-fold enriched in these vesicles relative to bulk flow markers. A plasma membrane protein, the general amino acid permease (Gap1p), is also enriched in these vesicles by virtue of a di-acidic sorting signal in its COOH-terminal cytosolic domain. Hence, some, but not all, soluble secretory proteins may be selected for entry into COPII vesicles. Others may leave the ER by bulk flow, only to be concentrated in later compartments by selective exclusion from the retrograde COPI vesicle pathway.

Vesicles also contain lipids, which are more than just a permeability barrier to hydrophilic molecules. Brügger et al. have used nanoelectrospray ionization tandem mass spectroscopy to conduct a quantitative analysis of COPI-coated vesicle lipids compared to the parent Golgi composition. Vesicles were found to be de-enriched in both cholesterol and sphingomyelin, indicating that lipids too are subject to selection for inclusion into vesicles.

5 Coated Vesicle Formation

5.1 Coats

Morphological examination of vesicles frequently reveals protein dense coatings on the cytosolic surface. Coat assembly may drive vesicle formation. The first coat to be identified is comprised of hexagonal arrays of clathrin triskelia on the plasma membrane, which must rearrange to a mixture of hexagons and pentagons as the

spherical vesicle forms and buds. Isolated clathrin triskelia can self-assemble into these "basket" structures *in vitro*. The principal sites of clathrin-coated vesicle formation are the plasma membrane and the TGN, from which both are delivered to the early endosome.

The other major class of coated vesicle so far characterized are the COP (coat protein complex) coated vesicles, which mediate retrograde trafficking between the Golgi and the ER, and between Golgi stacks (COPI), and anterograde trafficking from the ER to Golgi (COPII). The COPI coat is formed by Arf-dependent recruitment of "coatamer" (a preassembled heteroheptameric complex) from the cytosol. Arf is a small GTPase that recruits coatamer when it is in a GTP-bound active configuration. Similarly, the Arf family GTPase, Sar1p, choreographs the sequential recruitment of Sec23/24 and Sec13/31 complexes that comprise the COPII coat. Arf family proteins associate with membranes in their GTP-bound state by insertion of a hydrophobic anchor. This is withdrawn upon GTP hydrolysis and Arf dissociation can promote coat disassembly. The COPII component Sec23, is a GTPase-activating protein (GAP) for Sar1, and, so, upon recruitment of the Sec23/24 complex, the vesicles have only a brief opportunity to recruit cargo before coat disassembly begins.

5.2 Adaptors

Adaptor molecules link cargo-selected membrane proteins to the vesicle coat proteins. The first adaptors to be identified were the AP -1 and AP-2 adaptors (adaptor protein), which are highly enriched in TGN- and plasma membrane-derived clathrin-coated vesicles respectively. AP-1 and AP-2 are both heterotetramers of related proteins (AP-1: $\gamma\beta\mu$ and σ, AP-2: $\alpha\beta\mu$ and σ which promote clathrin self-assembly *in vitro*. Phosphoinositide lipids play a role in recruitment of these adaptor complexes; thus AP-1 binds to the TGN-enriched lipid PtdIns4P, whereas AP-2 binds to the plasma membrane enriched lipid PtdIns(4,5)P_2. The crystal structure of the AP-2 complex has been solved, revealing potential binding sites for phosphoinositides. Two further adaptor complexes, AP-3 and AP-4 were found by searching sequence databases for homologs of the AP-1 and AP-2 subunits. In distinction to AP-1 and AP-2, these are not enriched in clathrin-coated vesicles and may act independently or in conjunction with an unidentified scaffold to mediate membrane trafficking between the TGN and endosomes. Another class of adapter protein has recently been identified as a family of monomeric proteins called *GGAs* (Golgi-localized γear-containing, ARF-binding proteins), which contain a domain homologous to γ-adaptin. It is clear that they can bind to cargo receptors and clathrin, but, mysteriously, they are not enriched in clathrin-coated vesicles.

5.3 Principles of Cargo Selection

Specific signals mediate sorting of material into clathrin-coated vesicles. These signals may be embedded in the amino acid sequence or otherwise result from either constitutive or regulated posttranslational modification. The first sorting signal to be defined was the NPXY motif in LDL receptor, which resulted from Brown and Goldstein's pioneering studies of familial hypercholesterolemia. Mutation of the

tyrosine residue within this motif leads to failure of LDL receptor endocytosis. Other well-defined signals have since been determined including YXXø where ø is a bulky hydrophobic residue that determines transferrin receptor internalization amongst others. Constitutive posttranslational modification serving as a lumenal sorting signal is exemplified by lysosomal enzymes that are phosphorylated on sugar residues in the cis-Golgi, leading to capture in the TGN by the cation-independent mannose 6-phosphate receptor (ci-M6PR). ci-M6PR receptors are then packaged into AP1 clathrin-coated vesicles destined for the endocytic pathway, through cooperative interactions between GGA proteins and AP-1. A good example of regulated sorting is the ligand-dependent phosphorylation of tyrosine kinase receptors, such as EGFR, which leads to recruitment of accessory factors such as the Cin85/Cbl complex that downregulates receptor from the plasma membrane through incorporation into clathrin-coated vesicles. The number of molecules capable of fulfilling an adaptor function has steadily extended beyond the AP complexes, leading to the notion that there is both redundancy between and specialization of adaptor proteins. Examples include β-arrestin, which couples to seven-transmembrane G-protein-coupled receptors and ARH, which is mutated in patients with autosomal recessive hypercholesterolemia (ARH). ARH patients have a clinical phenotype almost indistinguishable from familial hypercholesterolemia but have normal *LDLR* alleles.

Binding of cargo to Sec24 appears to be the principal means of selection into COPII vesicles budding from the ER. However, the diverse array of cargo molecules utilize more than one binding site, two of which have now been mapped through elegant structural and biochemical studies. Vesicles must also incorporate specific targeting molecules; an interesting study of SNARE incorporation into COPII vesicles has suggested that only fusogenic forms of the SNAREs are selected for transport.

5.4 Induction of Curvature

Producing vesicles requires bending the membrane, which in turn requires energy. A standard bilayer bending modulus is estimated to be around 20 $k_B T$, where k_B is Boltzmann's constant and T the absolute temperature in Kelvin. One view is that the polymerization of the coat on a nascent vesicle induces membrane curvature. Several accessory proteins such as endophilin, amphiphysin, and epsin, implicated in vesicle formation, will tubulate or vesiculate liposomes when incubated *in vitro*. Recently, it has been proposed that BAR (Bin/Amphiphysin/Rvs) domain containing proteins can induce curvature through tight electrostatic association with acidic membranes in the absence of substantial hydrophobic insertion. The structure of the dimeric BAR domain from *Drosophila amphiphysin* is rigid, elongated, and will present a concave face to the membrane, which could accommodate a curved membrane of 11 nm outer radius. This, of course, does not fit well with the radius of coated vesicles (50 – 100 nm). It may be that rather than inducing curvature under physiological conditions, these proteins sense curvature of membranes, as their affinity for preexisting curved membranes will be higher. Bigay et al. have proposed that the BAR domain protein ArfGAP may act as a curvature sensor in COPI vesicle coat disassembly.

An alternative scenario, by which curvature may be induced, is through insertion of hydrophobic groups into one leaflet of the bilayer membrane, which increase the lateral area of this monolayer relative to the other and, consequently, promote curvature. An interesting case is the endocytic regulator epsin, which upon binding the lipid PtdIns(4,5)P_2 undergoes a conformational change that induces an amphipathic α-helix, which can insert into one leaflet of the bilayer. Clathrin-coated vesicle formation at the plasma membrane is promoted by generation of this phosphoinositide, which may also play a role in recruitment of AP-1.

5.5 Vesicle Scission

Budding vesicles can frequently be seen using electron microscopy as omega-shaped profiles protruding from parent membranes. The final step of vesicle scission will require membrane fusion to produce a continuous bilayer. The *shibire* mutation in Drosophila was identified as a temperature-sensitive mutation that caused paralysis through inhibition of synaptic vesicle recycling. It was shown that this defect leads to an accumulation of vesicles at the synaptic membrane that are unable to complete the final scission step. Later, the *shibire* gene was shown to encode a GTPase protein called *dynamin*. Dynamin is able to undergo large-scale oligomerization, wrapping itself around lipid tubules *in vitro*, or around the necks of budding vesicles in *vivo*, to form a helical collar. The pitch of this helix increases as dynamin oligomers undergo highly cooperative GTP hydrolysis. This has led to the "poppase" model of vesicle formation, in which this cooperative extension of a dynamin collar is used to provide the force to pop off a budding vesicle.

6 Vesicle Targeting

Once formed, a vesicle must identify its target membrane. The ultimate arbiters of the commitment to vesicle fusion are the SNARE proteins (Sect. 7.3), but prior to their involvement, looser associations are made that allow exploration of the propriety of any union.

6.1 Rab Proteins

One identifier of membrane compartments are the small GTPases of the rab family, of which there are more than 40 members in the human genome. Attention was first focused on this family when the budding yeast protein Sec4 was shown to be a member. Shortly thereafter, work on early endosome fusion in mammalian cells implicated rab5 as a key regulator. It has now been realized that the function of rab proteins is extremely complex. The GTP-bound form of rab5, for example, interacts with upwards of 22 proteins. However, for our purposes, their preeminent role is in their conformation-dependent recruitment of a family of extended coiled-coil proteins or multisubunit complexes, which are believed to act as tethers between interacting membranes (see next section). By recruiting membrane tethers through the GTP-bound form of rab proteins, which possess an intrinsic decay rate due to GTP hydrolysis, it is ensured that the tethering interaction is finite. Thus, if the appropriate downstream interactions, such as SNARE complex formation, are not available, the relationship can be severed before any damage can be done. This provides a form of kinetic proof reading as a potential means of ensuring fidelity.

6.2 Tethering Molecules

Tethering factors are heterogeneous in sequence and structure, but they do fall into distinct classes. A number of extended coiled-coil proteins, which exist as flexible rod-shaped structures of parallel homodimers, have been linked with tethering at the Golgi complex (Golgins) and on endosomes (EEA1, rabaptin 5) and may confine the search for trans-SNARE partners to the tethered membranes. Only a couple of examples for which there is reasonable experimental evidence for this role will be discussed here. EEA1 is proposed to play a role in tethering of endosomes prior to fusion. It is recruited to endosomes through a dual interaction with PtdIns3*P* and GTP-rab5 at its C-terminus, and also has an additional rab5-binding site at its N-terminus. Depletion of EEA1 from the cytosol markedly reduces the efficiency of endosome fusion *in vitro*. The C-terminus of EEA1 also interacts with the SNARE proteins syntaxin 6 and syntaxin 13, both of which have been implicated in endosome fusion.

Attachment of ER-derived vesicles to yeast Golgi *in vitro* was found to require the cytosolic factor Uso1 (a homolog of p115 in mammalian cells) and the small GTPase Ypt1, a member of the Rab family. Removal of Ypt1 from membranes led to a corresponding loss of Uso1, although a direct interaction between the two proteins has so far not been demonstrated. Ypt1 may either bind directly to Uso1 or regulate availability of an alternative receptor-binding site. Interestingly, both factors are also implicated in sorting of cargo at the ER, suggesting a coupling between cargo selection and vesicle tethering.

Another class of tethering molecules exists as multisubunit complexes. Seven large conserved complexes have been proposed to have roles in vesicle tethering at distinct transport steps. These can be divided into two families. One, termed the *quatrefoil complexes*, shares a common N-terminal domain in the majority of complex subunits, which all assemble in multiples of four. They include the Sec6/8 complex (or exocyst), the conserved oligomeric Golgi complex (COG) and the Golgi associated retrograde protein (GARP) complex. The exocyst binds to plasma membrane in budding yeast through recruitment by the Sec4 GTPase, although it is also an effector of other small GTPases. The COG complex localizes to the Golgi complex, although its function is not precisely clear. The GARP complex localizes to the TGN, binds to the SNARE protein Tlg1p, and is implicated in retrograde traffic from endosomes to the Golgi. Other multisubunit complexes, which have rather hazily been ascribed tethering functions, include the TRAPP (transport protein particle) complexes I and II, the class C Vps complex and Dsl1p. Several of these particles contain guanine nucleotide exchange factors (GEFs) that will activate rab GTPases. Thus, subsequent to tethering COPII vesicles to the Golgi, TRAPPI is proposed to activate Ypt1, which, in turn, recruits a second, downstream tether, the coiled-coil protein Uso1p, which has been discussed above.

7 Vesicle Fusion

7.1 Understanding Energy Barriers to Fusion

A condition of membrane fusion is that the membranes must become very closely apposed. Measurement of changes in bilayer spacing in multilamellar vesicles,

as a function of externally applied osmotic or hydrostatic pressure, revealed a surprisingly strong exponentially growing (characteristic decay length of 0.1–0.2 nm) repulsive force between phospholipid bilayers. Under normal circumstances then, bilayers do not spontaneously fuse. Insensitivity of this force to salt concentration and hence membrane electrostatics led to the initial conclusion that the repulsive force is due to hydration of the polar lipid head-groups by water. However, it is now generally believed that other forces also contribute to this repulsion, including steric interactions along with elastic pressures. In addition, an attractive force between closely apposed model membranes has been observed by measurements of adhesion of bilayers supported on mica surfaces in aqueous solutions. This force is believed to arise from exposure of hydrophobic moieties of phospholipids as a result of the stress exerted on bilayer surfaces that approach within about 1 nm.

7.2 Membrane Destabilization

A theoretical analysis of interactions between apposed membranes proposed that out-of-plane thermal fluctuations of the bilayers lead to the formation of close (less than 0.5 nm) contact between the membranes within a small area (approximately 10 nm^2). According to this theory, increasing hydration repulsion between apposed polar heads of lipid molecules in this area causes the rupture of interacting monolayers. This rupture results in monolayer fusion of the membranes, that is, in the formation of a bridge connecting the monolayers, which is usually named the "stalk" or hemifusion intermediate that can further evolve to form a hemifusion diaphragm, which then ruptures leading to full fusion.

7.3 SNARE Proteins

The "SNARE hypothesis" was proposed in 1993 by Rothman and co-workers to explain the Specificity of intracellular fusion events. It proposes that the core of the fusion machinery is comprised of SNARE proteins that are localized to specific subcellular compartments. Only cognate SNAREs on partner membranes can form a complex that promotes fusion. Following vesicle tethering, partner membranes have the opportunity to form trans-SNARE complexes. The first SNARE complex to be characterized was the one that specifies fusion of synaptic vesicles with the plasma membrane. This interaction, which is believed to be characteristic of all SNARE complexes, involves the formation of a parallel four-helix bundle, termed a "*SNAREpin*," in which one helix is contributed by the vesicle-associated SNARE (v-SNARE) synaptobrevin, whilst the other three are contributed by SNAREs on the target membrane (t-SNAREs). In this case, two of the t-SNARE helices are provided by SNAP-25 and one by syntaxin 1. Sequence analysis has shown that all SNAREs probably derive from a single ancestral gene that has been duplicated in the case of SNAP-25.

SNARE complexes do not simply lock vesicles into a closely associated state; they also drive membrane fusion. Substitution of the transmembrane domains of syntaxin and synaptobrevin with lipid anchors allows the docking interaction of vesicles to proceed, but fusion is no longer observed. The resolution of the SNARE complex structure immediately suggested an attractive mechanism by which fusion is promoted that has been termed

the zipper model. In this model, the assembly of the parallel four-helix bundle occurs initially at the N-termini of the helices and zips up toward the membrane anchors, consequently pulling the partner membranes into close apposition. The SNAREs themselves must tilt toward the membrane during this process, and it is attractive to speculate that this is coupled to deformation of the membrane through the transmembrane anchor (Fig. 4). Insertion of a flexible linker between the transmembrane domain and the coiled-coil domain reduces SNARE-dependent fusion efficiency systematically with increasing length of the linker. Note that after fusion, a SNARE complex will persist, but now in a cis configuration until it is disassembled by the action of NSF, which utilizes ATP hydrolysis.

Evidence that SNAREs participate in a late stage of physiological fusion reactions has been provided by an assay of regulated secretion of dense core vesicles in PC12 cells. In this system, the assembly of the SNARE complex occurred after the rise in Ca^{2+}, and could not be experimentally dissociated from the fusion process. In adrenal chromaffin cells, an antibody that inhibits SNARE assembly also reduces the initial fast component of exocytosis, indicating that even the vesicles poised for the quickest release require the SNARE assembly step.

A finding that the three synaptic SNAREs, when reconstituted into model liposomes and distributed according to cellular physiology, could promote liposome fusion led to a highly influential proposal that the SNAREpin complex represented the minimal fusion machinery. McNew et al. further developed this form of assay by taking advantage of the functional identification of yeast SNARE complexes associated with three specific transport steps, namely fusion of ER-derived transport vesicles with the Golgi, homotypic vacuole fusion, and fusion

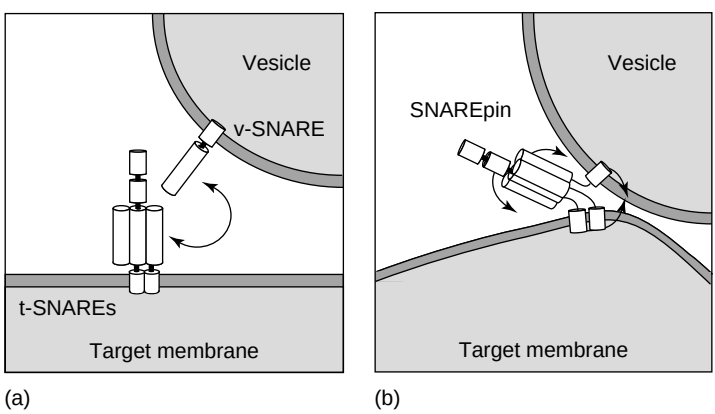

Fig. 4 Model for membrane deformation and fusion mediated by SNARE proteins. (a) The four helical sections of a SNARE complex in a prefusion state are arranged in a 3 : 1 distribution between partner membranes. (b) Progressive zipping of the four helical sections of SNARE proteins leads to the assembly of a tight parallel four-helix bundle (SNAREpin). This process forces membranes into close membrane apposition and will generate stresses at the membrane that are coupled to imposed bending of the participating proteins (indicated by arrows).

of secretory vesicles with the plasma membrane. They tested the fusion activity of liposomes bearing the set of t-SNAREs associated with each of these three steps with liposomes bearing any one of 11 actual or potential v-SNAREs in the yeast genome. The *in vitro* system recapitulated the cellular specificity to a remarkable degree, testifying to the high level of discrimination encoded within the SNARE complex.

These experiments built a three-component t-SNARE on one vesicle population to interact with a unitary v-SNARE on another. It is known that the four-helix bundle can also be formed in solution from soluble cytoplasmic domains of SNARE proteins. However, there are topological constraints on SNARE interactions when confined to membranes, which are not imposed in solution. On varying the distribution of each of four cognate SNAREs between donor and acceptor vesicles, Parlati et al. found that the original combination of Bet1 (v-SNARE) liposomes with Sed5–Bos1–Sec22 (t-SNARE complex) is uniquely fusogenic. Thus, the distribution of cognate SNAREs between partner membranes determines fusion specificity. This is important because the cellular distribution of SNARE proteins is not static.

7.4 Regulation of SNARE Complex Assembly

What is the effect of including a fifth SNARE in the liposome fusion system? In most cases, this has no effect, but instances have been found in which the additional SNARE can compete with and substitute for a fusogenic subunit, thereby inhibiting fusion. Components of the cis-Golgi fusion machinery are inhibited by components of the trans-Golgi machinery and vice versa. This has led to the concept of the inhibitory SNARE (i-SNARE). If SNAREs can regulate fusion in both a positive and negative manner, then this will provide tighter control of specificity at biological membranes containing overlapping SNARE complements. The ratio of fusogenic SNAREs to i-SNARES will determine propensity for specific fusion.

The N-terminal domain of the syntaxin-like SNARE proteins, such as neuronal syntaxin1 and yeast Sso1p, fold into three-helix bundles and inhibit SNARE pairing when folded back against their SNARE domains. In this configuration, syntaxin can form a tight complex with the neuronal SM protein (Sec1/Munc18 homolog) Munc18-1, but it must change to the "open" conformation to engage in SNARE complex assembly. This interaction with Munc18-1 has thus been proposed to inhibit syntaxin SNARE function, but this remains controversial. The proteins Munc13 and RIM have been proposed to promote synaptic vesicle fusion through displacement of Munc18-1 from syntaxin. Likewise, some tethering factors such as EEA1, known to interact with syntaxin family members, may also act to displace SM proteins and promote an open syntaxin configuration.

8 Vesicle Trafficking and Disease

The origins of many human pathologies lie in defective vesicle trafficking. In many cases, this is due to a specific failure of cargo selection. For example, hypercholesterolemia leading to atherosclerosis can result from mutations in the cytoplasmic domain of the LDL receptor required for endocytosis or in the LDL receptor–specific adaptor protein, ARH. The hypertensive Liddle's syndrome is due to

failure to endocytose epithelial Na+ channels. It has also been argued that various proteins are oncogenic because of crucial roles that they may play in downregulating activated growth factor receptors through endocytic trafficking to lysosomal degradative compartments. For example, the oncogene c-Cbl is an E3-ubiquitin ligase, which mediates monoubiquitination of activated tyrosine kinase receptors, such as EGFR. Ubiquitinated receptors are specifically selected for downregulation by inclusion into lumenal vesicles of endosomal multivesicular bodies that routes them to lysosomes away from the recycling pathway. Failure to direct receptors along this route will lead to retention of activated receptors and increased mitogenic signaling.

One would expect defects in the core components of vesicle budding and fusion to be lethal, but there are examples where heterozygosity or point mutations lead to specific pathologies. Knockout mice heterozygous for syntaxin 4 develop muscle insulin resistance, consistent with a role for this SNARE molecule in GLUT4 (glucose transporter) exocytosis. A point mutation in α-SNAP leads to a developmental defect in mice, known as *hydrocephaly with hop gait*, which is characterized by a reduced size in the cerebral cortex, owing to premature production of cerebral cortical neurons and depletion of the progenitor pool. This is accompanied by abnormal localization of many apical proteins implicated in regulation of cell fate in neuroepithelial cells.

A number of bacterial pathogens subvert the cellular trafficking machinery to fashion intracellular compartments conducive to their requirements. For example, *Legionella pneumophila* is a bacterial pathogen that infects eukaryotic host cells and replicates inside a specialized organelle that is morphologically similar to the endoplasmic reticulum (ER). The Legionella-containing vacuole acquires rab1 and the v-SNARE protein Sec.22b, which facilitate fusion of the vacuole with ER derived COPII vesicles, allowing formation of a specialized compartment that can support bacterial replication. The clostridial neurotoxins, botulinum B and tetanus toxin are zinc endopetidases, which cause paralysis through specific cleavage of the neuronal v-SNARE, synaptobrevin.

9 Conclusions

The basic components of the intracellular budding and fusion machineries have been identified through a combination of biochemistry and yeast genetics. Current frontiers are concerned with understanding the regulation of these components. For example, what special mechanisms allow a synaptic vesicle to be released within milliseconds of an elevated Ca^{2+} concentration and what then limits the release of vesicles at the synapse? There is an increasing awareness of the role of membrane traffic in disease due to inherited or acquired mutations and due to utilization of established routes or subversion of trafficking machineries by various pathogens such as viruses and bacteria.

See also Cell Junctions, Structure, Function, and Regulation; Cell Nucleus Biogenesis, Structure and Function; Intracellular Fatty Acid Binding Proteins and Fatty Acid Transport; Membrane Transport; Phagocytosis.

Bibliography

Books and Reviews

Antonny, B., Schekman, R. (2001) ER export: public transportation by the COPII coach, *Curr. Opin. Cell. Biol.* **13**, 438–443.

Blumenthal, R., Clague, M.J., Durell, S.R., Epand, R.M. (2003) Membrane fusion, *Chem. Rev.* **103**, 53–69.

Boal, D. (2002) *Mechanics of the Cell*, Cambridge University Press, ISBN: 0521796814, http://titles.cambridge.org/catalogue.asp?isbn=0521796814, http://www.sfu.ca/~boal/moc.html.

Bonifacino, J.S., Traub, L.M. (2003) Signals for sorting of transmembrane proteins to endosomes and lysosomes, *Annu. Rev. Biochem.* **72**, 395–447.

Chernomordik, L.V., Zimmerberg, J. (1995) Bending membranes to the task: structural intermediates in bilayer fusion, *Curr. Opin. Struct. Biol.* **5**, 541–547.

Meresse, S., Steele-Mortimer, O., Moreno, E., Desjardins, M., Finlay, B., Gorvel, J.P. (1999) Controlling the maturation of pathogen-containing vacuoles: a matter of life and death, *Nat. Cell Biol.* **1**, E183–E188.

Pfeffer, S.R. (2001) Rab GTPases: specifying and deciphering organelle identity and function, *Trends Cell Biol.* **11**, 487–491.

Rizo, J., Sudhof, T.C. (2002) Snares and Munc18 in synaptic vesicle fusion, *Nat. Rev. Neurosci.* **3**, 641–653.

Robinson, M.S. (2004) Adaptable adaptors for coated vesicles, *Trends Cell Biol.* **14**, 167–174.

Whyte, J.R.C., Munro, S. (2002) Vesicle tethering complexes in membrane traffic, *J. Cell Sci.* **115**, 2627–2637.

Primary Literature

Antonny, B., Madden, D., Hamamoto, S., Orci, L., Schekman, R. (2001) Dynamics of the COPII coat with GTP and stable analogues, *Nat. Cell Biol.* **3**, 531–537.

Balch, W.E., Dunphy, W.G., Braell, W.A., Rothman, J.E. (1984) Reconstitution of the transport of protein between successive compartments of the Golgi measured by the coupled incorporation of N-acetylglucosamine, *Cell* **39**, 405–416.

Bankaitis, V.A., Johnson, L.M., Emr, S.D. (1986) Isolation of yeast mutants defective in protein targeting to the vacuole, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9075–9079.

Beckers, C.J., Block, M.R., Glick, B.S., Rothman, J.E., Balch, W.E. (1989) Vesicular transport between the endoplasmic reticulum and the Golgi stack requires the NEM-sensitive fusion protein, *Nature* **339**, 397–398.

Bickford, L.C., Mossessova, E., Goldberg, J. (2004) A structural view of the COPII vesicle coat, *Curr. Opin. Struct. Biol.* **14**, 147–153.

Bigay, J., Gounon, P., Robineau, S., Antonny, B. (2003) Lipid packing sensed by ArfGAP1 couples COPI coat disassembly to membrane bilayer curvature, *Nature* **426**, 563–566.

Blondeau, F., Ritter, B., Allaire, P.D., Wasiak, S., Girard, M., Hussain, N.K., Angers, A., Legendre-Guillemin, V., Roy, L., Boismenu, D., Kearney, R.E., Bell, A.W., Bergeron, J.J., McPherson, P.S. (2004) Tandem MS analysis of brain clathrin-coated vesicles reveals their critical involvement in synaptic vesicle recycling, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 3833–3838.

Bock, J.B., Matern, H.T., Peden, A.A., Scheller, R.H. (2001) A genomic perspective on membrane compartment organization, *Nature* **409**, 839–841.

Bonifacino, J.S. (2004) The GGA proteins: adaptors on the move, *Nat. Rev. Mol. Cell. Biol.* **5**, 23–32.

Brown, M.S., Goldstein, J.L. (1986) A receptor-mediated pathway for cholesterol homeostasis, *Science* **232**, 34–47.

Brügger, B., Erben, G., Sandhoff, R., Wieland, F.T., Lehmann, W.D. (1997) Quantitative analysis of biological membrane lipids at the low picomole level by nanoelectrospray ionization tandem mass spectrometry, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2339–2344.

Burri, L., Lithgow, T. (2004) A complete set of SNAREs in yeast, *Traffic* **5**, 45–52.

Cao, X., Ballew, N., Barlowe, C. (1998) Initial docking of ER-derived vesicles requires Uso1p and Ypt1p but is independent of SNARE proteins, *EMBO J.* **17**, 2156–2165.

Chae, T.H., Kim, S., Marz, K.E., Hanson, P.I., Walsh, C.A. (2004) The hyh mutation uncovers roles for alpha Snap in apical protein localization and control of neural cell fate, *Nat. Genet.* **36**, 264–270.

Chen, W.J., Goldstein, J.L., Brown, M.S. (1990) NPXY, a sequence often found in cytoplasmic

tails, is required for coated pit-mediated internalization of the low density lipoprotein receptor, *J. Biol. Chem.* **265**, 3116–3123.

Chen, Y.A., Scales, S.J., Patel, S.M., Doung, Y.-C., Scheller, R.H. (1999) SNARE complex formation is triggered by Ca2+ and drives membrane fusion, *Cell* **97**, 165–174.

Christoforidis, S., McBride, H.M., Burgoyne, R.D., Zerial, M. (1999) The rab5 effector EEA1 is a core component of endosome docking, *Nature* **397**, 621–625.

Chvatchko, Y., Howald, I., Riezman, H. (1986) Two yeast mutants defective in endocytosis are defective in pheromone response, *Cell* **46**, 355–364.

Clague, M.J. (1998) Molecular aspects of the endocytic pathway, *Biochem. J.* **336**, 271–282.

Clague, M.J. (1999) Membrane transport: take your fusion partners, *Curr. Biol.* **9**, R258–R260.

Collins, B.M., McCoy, A.J., Kent, H.M., Evans, P.R., Owen, D.J. (2002) Molecular architecture and functional model of the endocytic AP2 complex, *Cell* **109**, 523–535.

Conibear, E., Cleck, J.N., Stevens, T.H. (2003) Vps51p mediates the association of the GARP (Vps52/53/54) complex with the late Golgi t-SNARE Tlg1p, *Mol. Biol. Cell.* **14**, 1610–1623.

Conner, S.D., Schmid, S.L. (2003) Regulated portals of entry into the cell, *Nature* **422**, 37–44.

De Duve, C. (1965) The separation and characterization of subcellular particles, *Harvey Lect.* **59**, 49–87.

Di Fiore, P.P., Gill, G.N. (1999) Endocytosis and mitogenic signalling, *Curr. Opin. Cell Biol.* **11**, 483–488.

Diaz, R., Mayorga, L., Weidman, P.J., Rothman, J.E., Stahl, P.D. (1989) Vesicle fusion following receptor-mediated endocytosis requires a protein active in Golgi transport, *Nature* **339**, 398–400.

Doray, B., Ghosh, P., Griffith, J., Geuze, H.J., Kornfeld, S. (2002) Cooperation of GGAs and AP-1 in packaging MPRs at the trans-Golgi network, *Science* **297**, 1700–1703.

Dumas, J.J., Merithew, E., Sudharshan, E., Rajamani, D., Hayes, S., Lawe, D., Corvera, S., Lambright, D.G. (2001) Multivalent endosome targeting by homodimeric EEA1, *Mol. Cell* **8**, 947–958.

Farsad, K., De Camilli, P. (2003) Mechanisms of membrane deformation, *Curr. Opin. Cell Biol.* **15**, 372–381.

Fasshauer, D. (2003) Structural insights into the SNARE mechanism, *Biochim. Biophys. Acta* **1641**, 87–97.

Ford, M.G., Mills, I.G., Peter, B.J., Vallis, Y., Praefcke, G.J., Evans, P.R., McMahon, H.T. (2002) Curvature of clathrin-coated pits driven by epsin, *Nature* **419**, 361–366.

Fries, E., Rothman, J.E. (1980) Transport of vesicular stomatitis virus glycoprotein in a cell-free extract, *Proc. Natl. Acad. Sci. U. S. A.* **77**, 3870–3874.

Gillingham, A.K., Munro, S. (2003) Long coiled-coil proteins and membrane traffic, *Biochim. Biophys. Acta* **1641**, 71–85.

Gorvel, J.P., Chavrier, P., Zerial, M., Gruenberg, J. (1991) Rab 5 controls early endosome fusion in vitro, *Cell* **64**, 915–925.

Guo, W., Roth, D., Walch-Solimena, C., Novick, P. (1999) The exocyst is an effector for Sec4p, targeting secretory vesicles to sites of exocytosis, *EMBO J.* **18**, 1071–1080.

Haas, A., Wickner, W. (1996) Homotypic vacuole fusion requires sec17p (yeast a-SNAP) and sec18p (yeast NSF), *EMBO J.* **15**, 3296–3305.

Hanson, P.I., Roth, R., Morisaki, H., Jahn, R., Heuser, J. (1997) Structure and conformational changes in NSF and its membrane receptor complexes visualized by quick-freeze/deep-etch electron microscopy, *Cell* **90**, 523–535.

He, G., Gupta, S., Yi, M., Michaely, P., Hobbs, H.H., Cohen, J.C. (2002) ARH is a modular adaptor protein that interacts with the LDL receptor, clathrin, and AP-2, *J. Biol. Chem.* **277**, 44044–44049.

Helm, C.A., Israelachvili, J.N., McGuiggan, P.M. (1989) Molecular mechanisms and forces involved in the adhesion and fusion of amphiphilic bilayers, *Science* **246**, 919–922.

Hirst, J., Lui, W.W., Bright, N.A., Totty, N., Seaman, M.N., Robinson, M.S. (2000) A family of proteins with gamma-adaptin and VHS domains that facilitate trafficking between the trans-Golgi network and the vacuole/lysosome, *J. Cell Biol.* **149**, 67–80.

Hong, H.K., Chakravarti, A., Takahashi, J.S. (2004) The gene for soluble N-ethylmaleimide sensitive factor attachment protein alpha is mutated in hydrocephaly with hop gait (hyh) mice, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1748–1753.

Huttner, W.B., Schiebler, W., Greengard, P., De Camilli, P. (1983) Synapsin 1 (protein 1), a nerve terminal-specific phosphoprotein. (III)

Its association with synaptic vesicles studied in a highly purified synaptic vesicle preparation, *J. Cell Biol.* **96**, 1374–1388.

Jamieson, J.D., Palade, G.E. (1967) Intracellular transport of secretory proteins in the pancreatic exocrine cell. I. Role of the peripheral elements of the Golgi complex, *J. Cell Biol.* **34**, 577–596.

Kagan, J.C., Stein, M.P., Pypaert, M., Roy, C.R. (2004) Legionella subvert the functions of rab1 and sec22b to create a replicative organelle, *J. Exp. Med.* **199**, 1201–1211.

Katzmann, D.J., Odorizzi, G., Emr, S.D. (2002) Receptor downregulation and multivesicular-body sorting, *Nat. Rev. Mol. Cell. Biol.* **3**, 893–905.

Leikin, S.L., Kozlov, M.M., Chernomordik, L.V., Markin, V.S., Chizmadzhev, Y.A. (1987) Membrane fusion: overcoming of the hydration barrier and local restructuring, *J. Theor. Biol.* **129**, 411–425.

Lippincott-Schwartz, J., Patterson, G.H. (2003) Development and use of fluorescent protein markers in living cells, *Science* **300**, 87–91.

Lipschutz, J.H., Mostov, K.E. (2002) Exocytosis: the many masters of the exocyst, *Curr. Biol.* **12**, R212–R214.

Malkus, P., Jiang, F., Schekman, R. (2002) Concentrative sorting of secretory cargo proteins into COPII-coated vesicles, *J. Cell Biol.* **159**, 915–921.

Martinez-Menarguez, J.A., Geuze, H.J., Slot, J.W., Klumperman, J. (1999) Vesicular tubular clusters between the ER and Golgi mediate concentration of soluble secretory proteins by exclusion from COPI-coated vesicles, *Cell* **98**, 81–90.

McNew, J.A., Weber, T., Engelman, D.M., Sollner, T.H., Rothman, J.E. (1999) The length of the flexible SNAREpin juxtamembrane region is a critical determinant of SNARE-dependent fusion, *Mol. Cell* **4**, 415–421.

McNew, J.A., Weber, T., Parlati, F., Johnston, R., Melia, T.J., Sollner, T.H., Rothman, J.E. (2000b) Close is not enough: SNARE-dependent membrane fusion requires an active mechanism that transduces force to membrane anchors, *J. Cell Biol.* **150**, 105–117.

McNew, J.A., Parlati, F., Fukuda, R., Lohnston, R.J., Paz, K., Paumet, F., Sollner, T.H., Rothman, J.E. (2000a) Compartmental specificity of cellular membrane fusion encoded in SNARE proteins, *Nature* **407**, 153–159.

Mills, I.G., Jones, A.T., Clague, M.J. (1998) Involvement of the endosomal autoantigen EEA1 in homotypic fusion of early endosomes, *Curr. Biol.* **8**, 881–884.

Mills, I.G., Urbe, S., Clague, M.J. (2001) Relationships between EEA1 binding partners and their role in endosome fusion, *J. Cell Sci.* **114**, 1959–1965.

Mironov, A.A., Beznoussenko, G.V., Nicoziani, P., Martella, O., Trucco, A., Kweon, H.S., Di Giandomenico, D., Polishchuk, R.S., Fusella, A., Lupetti, P., Berger, E.G., Geerts, W.J., Koster, A.J., Burger, K.N., Luini, A. (2001) Small cargo proteins and large aggregates can traverse the Golgi by a common mechanism without leaving the lumen of cisternae, *J. Cell Biol.* **155**, 1225–1238.

Mossessova, E., Bickford, L.C., Goldberg, J. (2003) SNARE selectivity of the COPII coat, *Cell* **114**, 483–495.

Nicholls, B.J., Ungermann, C., Pelham, H.R.B., Wickner, W., Haas, A. (1997) Homotypic vacuolar fusion mediated by v- and t-SNAREs, *Nature* **387**, 199–202.

Novick, P., Field, C., Schekman, R. (1980) Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway, *Cell* **21**, 205–215.

Orci, L., Palmer, D.J., Amherdt, M., Rothman, J.E. (1993) Coated vesicle assembly in the Golgi requires only coatomer and ARF proteins from the cytosol, *Nature* **364**, 732–734.

Padron, D., Wang, Y.J., Yamamoto, M., Yin, H., Roth, M.G. (2003) Phosphatidylinositol phosphate 5-kinase Ibeta recruits AP-2 to the plasma membrane and regulates rates of constitutive endocytosis, *J. Cell Biol.* **162**, 693–701.

Painter, R.G., Tokuyasu, K.T., Singer, S.J. (1973) Immunoferritin localization of intracellular antigens: the use of ultracryotomy to obtain ultrathin sections suitable for direct immunoferritin staining, *Proc. Natl. Acad. Sci. U. S. A.* **70**, 1649–1653.

Palade, G. (1975) Intracellular aspects of the process of protein secretion, *Science* **189**, 347–358.

Parlati, F., McNew, J.A., Fukuda, R., Miller, R., Sollner, T.H., Rothman, J.E. (2000) Topological restriction of SNARE-dependent membrane fusion, *Nature* **407**, 194–198.

Parsegian, V.A., Fuller, N., Rand, R.P. (1979) Measured work of deformation and repulsion of lecithin bilayers, *Proc. Natl. Acad. Sci. U. S. A.* **76**, 2750–2754.

Pearse, B., Robinson, M.S. (1990) Clathrin, adaptors and sorting, *Ann. Rev. Cell Biol.* **6**, 151–172.

Pearse, B.M.F. (1975) Coated vesicles from pig brain: purification and biochemical characterization, *J. Mol. Biol.* **97**, 93–98.

Pelham, H.R.B. (1999) SNAREs and the secretory pathway-lessons from yeast, *Exp. Cell Res.* **247**, 1–8.

Piper, R.C., Luzio, J.P. (2001) Late endosomes: sorting and partitioning in multivesicular bodies, *Traffic* **2**, 612–621.

Poodry, C.A., Edgar, L. (1979) Reversible alteration in the neuromuscular junctions of Drosophila melanogaster bearing a temperature-sensitive mutation, shibire, *J. Cell Biol.* **81**, 520–527.

Robinson, M.S., Bonifacino, J.S. (2001) Adaptor-related proteins, *Curr. Opin. Cell. Biol.* **13**, 444–453.

Romisch, K., Ali, B.R. (1997) Similar processes mediate glycopeptide export from the endoplasmic reticulum in mammalian cells and Saccharomyces cerevisiae, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 6730–6734.

Roth, M.G. (2004) Phosphoinositides in constitutive membrane traffic, *Physiol. Rev.* **84**, 699–730.

Roth, T.F., Porter, K.R. (1964) Yolk protein uptake in the oocyte of the mosquito Aedes Aegypti L, *J. Cell Biol.* **20**, 313–332.

Rothman, J.E. (1994) Mechanisms of intracellular protein transport, *Nature* **372**, 55–63.

Sacher, M., Barrowman, J., Wang, W., Horecka, J., Zhang, Y., Pypaert, M., Ferro-Novick, S. (2001) TRAPP I implicated in the specificity of tethering in ER-to-Golgi transport, *Mol. Cell.* **7**, 433–442.

Salminen, A., Novick, P.J. (1987) A ras-like protein is required for a post-Golgi event in yeast secretion, *Cell* **49**, 527–538.

Schiavo, G., Benfenati, F., Poulain, B., Rossetto, O., Polverino de Laureto, P., DasGupta, B.R., Montecucco, C. (1992) Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin, *Nature* **359**, 832–835.

Serafini, T., Stenbeck, G., Brecht, A., Lottspeich, F., Orci, L., Rothman, J.E., Wieland, F. (1991) A coat subunit of Golgi-derived non-clathrin-coated vesicles with homology to the clathrin-coated vesicle coat protein B-adaptin, *Nature* **349**, 215–220.

Sheetz, M.P., Singer, S.J. (1974) Biological membranes as bilayer couples. A molecular mechanism of drug-erythrocyte interactions, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4457–4461.

Snyder, P.M. (2002) The epithelial Na+ channel: cell surface insertion and retrieval in Na+ homeostasis and hypertension, *Endocr. Rev.* **23**, 258–275.

Sollner, T., Whiteheart, S.W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P., Rothman, J.E. (1993) SNAP receptors implicated in vesicle targeting and fusion, *Nature* **362**, 318–324.

Soubeyran, P., Kowanetz, K., Szymkiewicz, I., Langdon, W.Y., Dikic, I. (2002) Cbl-CIN85-endophilin complex mediates ligand-induced downregulation of EGF receptors, *Nature* **416**, 183–187.

Stowell, M.H., Marks, B., Wigge, P., McMahon, H.T. (1999) Nucleotide-dependent conformational changes in dynamin: evidence for a mechanochemical molecular spring, *Nat. Cell Biol.* **1**, 27–32.

Sutton, R.B., Fasshauer, D., Jahn, R., Brunger, A.T. (1998) Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Åresolution, *Nature* **395**, 347–353.

Thien, C.B.F., Langdon, W.Y. (2001) Cbl:many adaptions to regulate protein tyrosine kinases, *Nature Rev.* **2**, 294–305.

Toonen, R.F., Verhage, M. (2003) Vesicle trafficking: pleasure and pain from SM genes, *Trends Cell Biol.* **13**, 177–186.

Traub, L.M. (2003) Sorting it out: AP-2 and alternate clathrin adaptors in endocytic cargo selection, *J. Cell Biol.* **163**, 203–208.

van der Bliek, A.M., Meyerowitz, E.M. (1991) Dynamin-like protein encoded by the drosophilia shibire gene associated with vesicular traffic, *Nature* **351**, 411–414.

Varlamov, O., Volchuk, A., Rahimian, V., Doege, C.A., Paumet, F., Eng, W.S., Arango, N., Parlati, F., Ravazzola, M., Orci, L., Sollner, T.H., Rothman, J.E. (2004) i-SNAREs: inhibitory SNAREs that fine-tune the specificity of membrane fusion, *J. Cell Biol.* **164**, 79–88.

Weber, T., Zemelman, B.V., McNew, J.A., Westermann, B., Gmachl, M., Parlati, F., Sollner, T.H., Rothman, J.E. (1998) SNAREpins: minimal machinery for membrane fusion, *Cell* **92**, 759–772.

Weimbs, T., Mostov, K., Low, S.H., Hofmann, K. (1998) A model for structural similarity between different SNARE complexes

based on sequence relationships, *Trends Cell Biol.* **8**, 260–262.

Whiteheart, S.W., Schraw, T., Matveeva, E.A. (2001) N-ethylmaleimide sensitive factor (NSF) structure and function, *Int. Rev. Cytol.* **207**, 71–112.

Wieland, F.T., Gleason, M.L., Serafini, T.A., Rothman, J.E. (1987) The rate of bulk flow from the endoplasmic reticulum to the cell surface, *Cell* **50**, 289–300.

Wilson, D.W., Wilcox, C.A., Flynn, G.C., Chen, E., Kuang, W.-J., Hazel, W.J., Block, M.R., Ullrich, A., Rothman, J.E. (1989) A fusion protein required for vesicle mediated transport in both mammalian cells and yeast, *Nature* **339**, 355–359.

Xu, T., Rammner, B., Margittai, M., Artalejo, A.R., Neher, E., Jahn, R. (1999) Inhibition of SNARE complex assembly differentially affects kinetic components of exocytosis, *Cell* **99**, 713–722.

Yang, B., Gonzalez, L., Prekeris, R., Steegmaier, M., Advani, R.J., Scheller, R.H. (1999) SNARE interactions are not selective, *J. Biol. Chem.* **274**, 5649–5653.

Yang, C., Coker, K.J., Kim, J.K., Mora, S., Thurmond, D.C., Davis, A.C., Yang, B., Williamson, R.A., Shulman, G.I., Pessin, J.E. (2001) Syntaxin 4 heterozygous knockout mice develop muscle insulin resistance, *J. Clin. Invest.* **107**, 1311–1318.

Zimmerberg, J., McLaughlin, S. (2004) Membrane curvature: How BAR domains bend bilayers, *Curr. Biol.* **14**, R250–R252.

Membrane Transport

Caroline Engvall and Per Lundahl
Uppsala University, Uppsala, Sweden

1 Cell Membranes 143
1.1 Structure 143
1.1.1 Measurements Indicating the Membrane Structure 145

2 Transport Across Cell Membranes 146
2.1 Partition-diffusion 147
2.1.1 Theory and Examples 147
2.1.2 Measurements 148
2.2 Diffusion through Porins 149
2.3 Diffusion through Ion Channels 150
2.3.1 Chloride Channels 150
2.3.2 Na^+, Ca^{2+}, and K^+ Channels 150
2.4 Facilitated Diffusion 151
2.4.1 Examples 151
2.4.2 Measurements 153
2.5 Active Transport 153
2.5.1 Examples 153
2.5.2 Measurements 155
2.6 Endo-, Exo-, and Transcytosis 155

3 Transport Across Epithelia 156
3.1 Intestinal Epithelia 156
3.2 The Blood-brain Barrier 156

4 Solute Affinities for Membrane Proteins 157

5 Prediction and Modeling of Drug Absorption Across Membranes 157

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

6 Conclusions and Visions 158

Bibliography 158
Books and Reviews 158
Primary Literature 159

Keywords

Amphiphilic Molecule
A molecule with both polar and nonpolar elements.

Cell Membrane or Plasma Membraneff
The boundary layer of a cell, composed of a mosaic of proteins in phospholipid bilayers, which is supplemented with cholesterol in animal cells.

Hydrophilic Molecule or Atom
A highly water-soluble polar molecule or atom.

Hydrophobic Molecule or Atom
A nonpolar molecule or atom that is highly soluble in organic solvents.

Liposome
A body consisting of an artificial lipid bilayer (or bilayers) enclosing an aqueous compartment (or aqueous layers).

Membrane Protein
An integral membrane protein is inserted into a lipid bilayer, and is called a transmembrane protein if exposed on both membrane faces, whereas a peripheral membrane protein is bound to one of the membrane surfaces.

Permeability ($cm^3\ s^{-1}$) of a Membrane for a Solute
The ratio between the flow across the membrane (mol s^{-1}) and the concentration difference (mol cm^{-3}).

Phospholipid Bilayer
Thin lamellae composed of two layers of phospholipid molecules oriented to contain a hydrophobic, oil-like interior and expose hydrophilic structures toward the aqueous phases.

Phospholipid Molecule
An amphiphilic molecule with two hydrophobic chains and a polar headgroup containing a phosphate group and choline, ethanolamine, serine, inositol, or a similar moiety.

Proteoliposome
Liposome containing one or more membrane proteins inserted into the bilayers.

Small hydrophobic or amphiphilic solutes, such as oxygen or drugs, partition into the lipid bilayer of a cell membrane, diffuse within it, and partition out from it. More polar molecules, for example, glucose, require the help of a membrane protein to traverse the membrane. Porins contain pores of passage, whereas facilitative transporters change conformation to let the substrate through the membrane. Ion channels allow regulated passage of anions or cations. These modes of passive transport are thermally driven toward lower substrate concentration. Active transport proteins concentrate a solute by the use of adenosine triphosphate (ATP) hydrolysis or by means of an ion gradient whereby ions passing through the transporter drive the uphill transport of the substrate. Xenobiotics are expelled from cells by active transporters, for example, P-glycoprotein. Endo- or exocytosis transports compounds across the cell membrane in vesicles budding off from or fusing with the membranes, respectively. Partitioning and transport are analyzed with cells or cell models.

1 Cell Membranes

1.1 Structure

A membrane (schematically shown in Fig. 1) separates the cell from its surroundings, yet allows contact across the barrier, and membranes form organelles within cells. The thin, flexible membranes are composed of membrane proteins imbedded in lipid bilayers essentially impermeable to ions and large molecules. Protons, small molecules such as water (H_2O) and oxygen (O_2), and larger amphiphilic molecules, such as anesthetics and drugs can pass the bilayers slowly, whereas transmembrane proteins control the selective uptake of nutrients and the export of metabolic products, regulate the balance of ions and solutes between the exterior and interior of the cell or transmit signals. Certain membranes collect or transform energy.

Phosphatidylcholine, -ethanolamine, and -serine are major lipid components of mammalian cell membranes (Fig. 2A), and contain two fatty acyl chains of 12 to 24 carbons, one saturated and one unsaturated, esterified to carbons 1 and 2, respectively, of a glycerol backbone, whereas carbon 3 is connected to a phosphate group, which in turn carries an alcohol (X in Fig. 2A) to form a polar headgroup. Other lipid components are the phosphatidylcholine analogue sphingomyelin, which contains amide bonds instead of ester bonds, and cholesterol (Fig. 2B), with four fused sterol rings bearing a single hydroxyl group on ring A and a branched hydrophobic tail on ring D. The phospholipid/cholesterol bilayer of an animal cell membrane is approximately 4.5-nm thick and exposes the polar phospholipid headgroups on each face of the

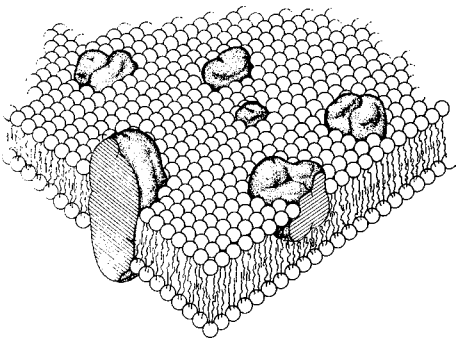

Fig. 1 The fluid mosaic model of the lipid-protein structure of a biological membrane. Reprinted, with permission, from Singer, S.J., Nicolson, G.L. (1972). The fluid mosaic model of the structure of cell membranes. Cell membranes are viewed as two-dimensional solutions of oriented globular proteins and lipids, *Science* **175**, 720–731. Copyright (1972) American Association for the Advancement of Science.

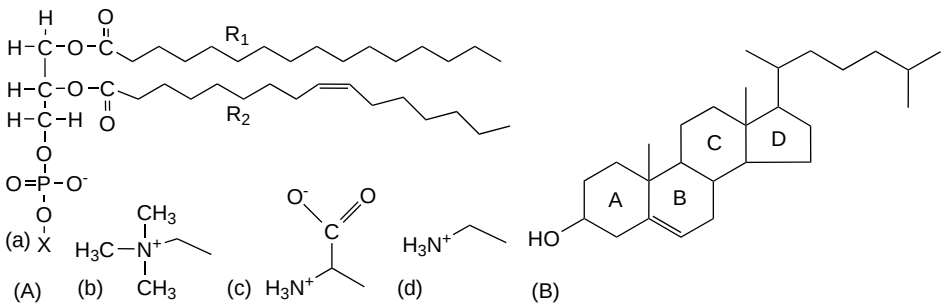

Fig. 2 (A) (a) Chemical structure of a glycerophospholipid with two 16-carbon chains where R_1 is saturated and R_2 is unsaturated with a double bond. X can be, for example, (b) choline, (c) serine, or (d) ethanolamine. (B) Cholesterol. The four fused rings are denoted A to D.

membrane. The overall structure of a lipid bilayer resembles the structure of a double layer of crystalline phospholipid, such as that determined by X-ray diffraction in Fig. 3, except that the lipids rotate, move their chains, and diffuse laterally in the liquid-crystalline phase, which solidifies to a gel-like consistence when the temperature is decreased below a transition temperature that depends mainly on the degree of unsaturation. Refinement of X-ray and neutron scattering data show the mobility of the various parts of the phospholipid molecules in the different regions of the membrane. They move in and out from their average position within certain limits. The fluidity of the membrane is modulated by cholesterol molecules, each one with its hydroxyl group in one of the two outer polar membrane regions and its rigid sterol skeleton in one half of the hydrophobic oil-like liquid in between the polar surfaces.

Integral or transmembrane proteins (which may be monomeric, dimeric, or oligomeric) are inserted into the phospholipid bilayer. Most mammalian integral membrane proteins are designed as a hydrophobic bunch of α-helices, each one containing about 20 amino acid residues and connected by hydrophilic loops. An individual three-dimensional structure allows a specific function for each protein, such as signal transduction, nutrient transport, or transport of xenobiotics across the membrane. Other membrane

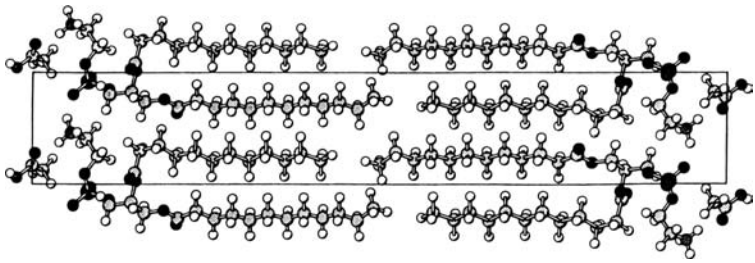

Fig. 3 The structure of crystalline dilauroylphosphatidylethanolamine acetic acid. The acetic acid molecules are seen outside the phospholipid molecules. The crystallographic unit cell is shown. Reprinted with permission from M. Karplus and the copyright owner from Biological membranes: a molecular perspective from computation and experiment, Eds. K.M. Merz, B. Roux, An empirical potential energy function for phospholipids: Criteria for parameter optimization and applications, by M. Schlenkrich, J. Brickmann, A.D. MacKerell Jr., M. Karplus, p. 31–81, Figure 14C, Birkhäuser, Boston. Copyright (1996) Springer.

proteins of entirely different structures (peripheral membrane proteins) are coupled to the membrane surfaces one by one or in chains forming cytoskeletal fibrous networks.

1.1.1 Measurements Indicating the Membrane Structure

The osmotic properties of plant cells were studied during the last half of the nineteenth century. Overton suggested (in 1899–1900) that the osmotic barrier of the cell consisted of a membrane "impregnated" with lipids of properties similar to cholesterol esters and phospholipids. The lipid–protein structure of cell membranes was debated during the first seven decades of the twentieth century. In 1925, Gorter and Grendel used a Langmuir trough to determine the surface area of a monolayer of the lipids extracted from one red blood cell. Calculations involving two errors that fortuitously cancelled each other indicated that the lipids formed a bilayer in the cell membrane. In 1972, Singer and Nicolson proposed, on the basis of thermodynamic considerations, that the hydrophilic and hydrophobic properties of membrane components and experimental data on cell membranes were consistent with a "fluid mosaic" membrane model with proteins inserted into and floating around in a phospholipid bilayer (Fig. 1). One of the observations supporting this model is the random spreading of the Rh_0(D) integral membrane protein molecules of the human red blood cell over the membrane as revealed by electron microscopy following ferritin staining of the Rh_0(D) proteins by using antibodies. This view has essentially persisted since then, although peripheral membrane proteins are associated with the membrane, some integral membrane proteins are coupled to cytoskeletal networks of proteins at the membrane surface or in the cytoplasm, and many integral membrane proteins are dimers or oligomers of one or more subunits. In 1975, Henderson and Unwin described the three-dimensional structure of the small light-driven proton pump, bacteriorhodopsin, according to electron diffraction data collected from crystalline patches, called purple membranes, of this transmembrane protein in

Halobacterium halobium. The first detailed X-ray crystallographic three-dimensional structure of a transmembrane protein, the big photosynthetic reaction center, which is involved in photosynthesis in *Rhodopseudomonas viridis*, was presented in 1985 by Deisenhofer, Huber, and Michel. The development of genetic techniques and the high stability and availability of bacterial membrane proteins has transferred the emphasis of the crystallographic membrane research to such proteins. These show close resemblance to human transmembrane proteins in the few cases that, so far, have been more closely studied. One example is the similarity between the *Escherichia coli* lactose permease, a membrane protein that mediates lactose transport across the inner bacterial membrane, and the human facilitated glucose transporter GLUT1.

2 Transport Across Cell Membranes

Solutes enter or leave a cell (Fig. 4) by partitioning into and diffusing across the bilayer, by protein-mediated transport or by endo- or exocytosis. Small amphiphilic molecules, for example, most orally administered drugs, may use the nonspecific partition-diffusion transport (Fig. 4a,b), which slows down with decreasing temperature but cannot be stopped. Small hydrophilic solutes diffuse through passages between the cells in cell layers; this is called the paracellular route (not illustrated). Ions and polar molecules partition only weakly into the hydrophobic region of the membrane and are helped by transmembrane proteins to pass the membranes at a high rate. Such transport proteins (Fig. 4c) either provide nonselective pores or offer specific binding, passage, and release of the substrate. Depending on the type of organism or organ, the transport is either passive, toward a lower solute concentration, or active, that is, concentrating the solute by use of an energy source, for example, adenosine triphosphate (ATP). Alternatively, a gradient of protons or other ions in one direction can be utilized to drive a cotransport of the solute toward higher concentration and the ion toward lower concentration.

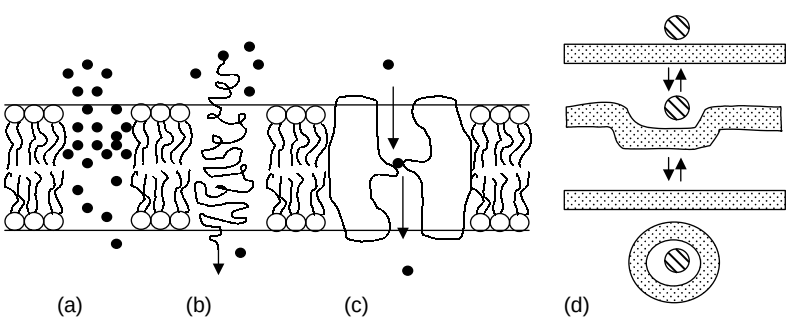

Fig. 4 Schematic illustrations of the permeation of a solute across a cell membrane by (a,b) partition–diffusion through the lipid bilayer (for clarity the lipid molecules are not drawn where the solute resides), where (a) exemplifies the concentration gradient formed, and (b) depicts the Brownian motion of a single solute molecule, (c) diffusion through a pore or channel of a membrane protein or facilitated diffusion mediated by a substrate-binding membrane protein, and (d) endo- or exocytosis.

Active multisubstrate transporters, such as the P-glycoprotein, expel xenobiotics from cells and cell organelles. Endo- and exocytosis are involved in the import and export, respectively, of certain molecules, for example, lipoproteins (Fig. 4d). The research on the many essential functions of membrane lipids and proteins has been hampered by methodological problems due to the insolubility of most membrane components in water.

2.1 Partition-diffusion

The molecules or atoms in gases or liquids move and collide randomly because of their thermal energy. Molecules that are nonrandomly distributed will move from a higher concentration toward a lower concentration because a collision is more probable where the molecules are closer to each other. This is diffusion, which also occurs in the semiliquid interior of a biological membrane. Solute diffusion in and across lipid bilayers is sometimes called simple diffusion, partition-diffusion, or nonfacilitated diffusion, to distinguish the process from facilitated diffusion; a thermally driven solute transport by a membrane protein toward a lower concentration.

2.1.1 Theory and Examples

The diffusional flux, J, (mol s^{-1}), of molecules from a region of high concentration toward regions of lower concentration can be expressed by Fick's first law of diffusion,

$$J = -D\frac{dc}{dx} \qquad (1)$$

with D, the diffusion coefficient (cm^2 s^{-1}); and dc/dx, the concentration gradient, that is, the derivative of the solute concentration with regard to the distance across the membrane. In the case of diffusion, across a membrane forming a hydrophobic barrier, Equation 1 can be written

$$J = \frac{A \times D_m \times \Delta C_m}{d} \qquad (2)$$

with A, the area (cm^2) of the membrane; d, the thickness (about 3×10^{-7} cm) of the hydrophobic region; D_m, the diffusion coefficient of the solute in the hydrophobic region; and ΔC_m, the concentration difference of the solute across the hydrophobic region.

A solute partitions between the hydrophobic region of a membrane and the water (Fig. 4a). At equilibrium, the solute concentrations within and outside the membrane balance the energy change as the solute crosses the boundary. The partition coefficient, K, is defined as the ratio between the solute concentrations in the nonpolar and aqueous phases. Therefore,

$$\Delta C_m = K \times \Delta C_{aq} \qquad (3)$$

with ΔC_{aq}, the concentration difference between the water on each side of the membrane.

The permeation of a small, hydrophilic molecule is slow because of weak partitioning into the membrane, creating a shallow concentration gradient within the membrane. More hydrophobic compounds partition more strongly into the membrane, but may suffer from a low aqueous solubility. Amphiphilic compounds thus require a certain balance between the amount and distribution of hydrophilic and hydrophobic structural features in order to permeate a membrane quickly.

A correlation between the partition and permeability coefficients for many solutes has been observed for a long time. In

the partition–diffusion model, K and D_m determine the rate of solute migration across a membrane. The permeability coefficient, P_s (cm s^{-1}), for the solute S is given by

$$P_s = \frac{D_m \times K}{d} \quad (4)$$

and the permeability, $P_s \times A$, equals the flux J (mol s^{-1}) of S divided by the concentration difference, ΔC_{aq} (mol cm^{-3}). For example, the permeability coefficient for glucose is 2×10^{-10} cm s^{-1} and that for O_2 is 2×10^{-3} cm s^{-1}. D_m is approximately 10^{-8} cm^2 s^{-1} for glucose and 0.5×10^{-5} cm^2 s^{-1} for O_2. Equation 4 thus requires that the membrane–water partition coefficient of oxygen is 10^4-fold that of glucose.

A solute that enters or leaves the membrane interacts with the electric charges and dipole moments of the polar headgroup region of the lipid bilayers, of oriented water molecules, of the carbonyl oxygens of the fatty acyl chains and of membrane proteins. Upon steady state transport, the solute concentration, therefore, varies nonlinearly across the membrane and also laterally, which complicates the situation described by Equations 2–4 and affects different transported compounds in different ways. Very small molecules may diffuse at an enhanced speed in the membrane by jumping through empty spaces between the lipid molecules.

The partition coefficient, K, can be expressed in the free energy change, $-\Delta G$:

$$K = e^{\frac{-\Delta G}{RT}} \quad (5)$$

with R, the general gas constant and T, the temperature. ΔG can be described by a hydrophobic-effect term and four electrostatic terms: the Born energy of transfer of an ion or dipole into a hydrophobic solvent; the image energy owing to the thinness of the membrane; the dipole-potential energy caused by the polar headgroup region; and the energy change upon the formation or breakage of hydrogen bonds and the binding or release of hydrating water. The partition–diffusion model works for hydrated ions, although cations may pass thin bilayers through transient pores and H^+ may be transferred along water strands in transient defects of the bilayer.

The equilibrium ratio between the concentrations of an ionic compound on each side of a membrane is directly affected by the transmembrane potential. A 10-fold concentration ratio of a univalent ion balances a 59-mV potential difference and a 100-fold ratio corresponds to twice that potential difference.

2.1.2 Measurements

The partitioning of solutes into lipid bilayers or cell membranes can be determined by chromatography on liposomes, proteoliposomes, membrane vesicles or cells adsorbed to or entrapped in gel beads packed into columns. This is called immobilized-liposome chromatography or immobilized-biomembrane chromatography (ILC, IBC). The chromatographic retardations of the solutes divided by the lipid amount in the column provide partition data, which correlate with octanol-water (P_{oct}) partition data, as do solute partition data obtained by pH titration of the solute in a liposome suspension, or by optical surface-plasmon resonance analysis of solute binding to liposomes coupled to a planar surface. These techniques reveal details of solute interactions with membranes. The partitioning of drugs into lipid bilayers decreases in the order neutral, positively charged, and negatively charged drugs compared to partitioning into octanol. Cholesterol in the bilayers

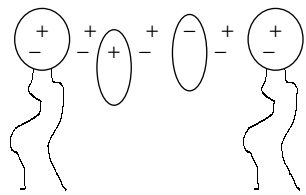

Fig. 5 Proposed distribution of positively and negatively charged amphiphilic solute molecules at the interface between the polar and nonpolar regions of the membrane. The lower half of the solute molecule is thought to be predominantly hydrophobic. Electrostatic effects cause different distributions of the two charged solutes, and thereby stronger or weaker hydrophobic effect.

decreases the partitioning in proportion to its amount. Membrane proteins enhance or decrease the partitioning by electrostatic interactions in combination with hydrophobic partitioning; and positively charged drugs partition relatively strongly into the hydrophobic region of the bilayer, whereas negatively charged drugs tend to dissociate more rapidly from liposome surfaces (Fig. 5). The simpler model of drug partitioning into a layer of $(oxyethylene)_{23}$ dodecyl ether molecules adsorbed to a stationary phase to mimic the membrane structure upon micellar liquid chromatography shows strikingly high correlations with ILC/IBC partitioning data obtained for a set of drugs on liposomes, red blood cell membrane vesicles, red blood cells, and ghosts.

A number of methods for permeability measurements are available. The permeabilities of cell membranes for solutes determined by observations of osmotic swelling or shrinking of cells by microscopy or weighing correlate rectilinearly with the oil–water partitioning or with P_{oct} of the osmotically active solutes. The permeability of liposomes for protons has been determined by measuring the pH-dependent fluorescence of pyranine, a cation-sensitive electrode has been used for potassium ions, and light-scattering analysis of osmotic swelling by use of a stop-flow apparatus has been used for water, glycerol, and urea. The liposomal and proteoliposomal permeabilities for radioactively labeled glucose, tyrosine, and tryptophan have been estimated by incubation followed by size-exclusion chromatographic separation monitored by flow-scintillation detection. The permeability coefficients for α-substituted *p*-toluic acids and glucose have been determined by incorporation into liposomes, removal of the free compound by size-exclusion chromatography, collection of the compound released during incubation by ultrafiltration and determination of the amount of the transported compound by high performance liquid chromatography (HPLC).

In the pharmaceutical industry, the use of sealed monolayers of cultured epithelial tumor cells, for example, Caco-2 cells, has achieved a widespread practical application in the determination of permeabilities. The drawbacks are inconsistency among permeability values determined in different laboratories and the time-consuming culturing of the cell layers.

2.2 Diffusion through Porins

Pore-forming proteins, called porins, provide openings that allow diffusional passage of polar molecules of moderate size. There are 10 families of aquaporins that let water pass rapidly through membranes. Such a protein, in red blood cells is

a tetramer of M_r 28 000-subunits, each containing six transmembrane α-helices forming a barrel that leads to a water-selective pore, which does not allow ion passage. The pore contains a hydrophobic section with three widely separated hydrophilic sites. Aquaporins in kidney cell membranes are involved in the reabsorption of 150 l of water per day.

In the outer membrane of gram-negative bacteria, for example, *E. coli*, porins provide openings that allow diffusional passage of polar molecules of M_r up to about 600. These porins are trimers of hollow cylinders, β-barrels, with a hydrophobic belt on the outside in contact with the fatty acyl chains of the surrounding phospholipids. The pores are hydrophilic and water-filled with protruding amino acid side chains that limit the size of the solutes that can pass by diffusion toward lower concentration. Similar size-selective porins are also found in the outer membrane of mitochondria.

The efflux of a variety of antibacterial components, including protein toxins, from *E. coli* and other gram-negative bacteria takes place by interaction between inner-membrane proteins and an outer-membrane TolC protein that spans both the membrane and the periplasmic space with a 14-nm tunnel, offering a transport mechanism similar to that of the porins, which span the outer membrane only. However, the tunnel can be sealed and unsealed to regulate the transport.

2.3 Diffusion through Ion Channels

2.3.1 Chloride Channels

Chloride channels allow the passage of different anions, of which chloride is physiologically the most common. The channels are divided into three classes, the dimeric chloride channels (ClC), the cystic fibrosis transmembrane conductance regulator (CFTR) and the ligand-gated GABA and glycine receptors.

The chloride/hydrogen carbonate (anion) exchange protein of human red cells has been studied for decades and the sequence was deduced in 1985. A bacterial ClC has been crystallized and shows a complex pattern of 18 tilted α-helices. The ClC in the electric fish *Torpedo* contains two gated barrels that open and close independently to give three conductance levels, zero (both closed), medium (one open), and high (both open). The same behavior is shared by the muscle chloride channel expressed in oocytes from the African tree frog, *Xenopus laevis*.

The special CFTR type of chloride channel that regulates salt and water transport across epithelial tissue is an ATP binding cassette (ABC) transport protein activated by phosphorylation of a cytoplasmic domain by protein kinase A. Cystic fibrosis is caused by decreased CFTR activity in the lung and secretory diarrhea is caused by increased CFTR activity in the gut.

2.3.2 Na^+, Ca^{2+}, and K^+ Channels

Voltage-dependent channel proteins that allow passage of Na^+, Ca^{2+}, or K^+ are tetramers of identical or similar subunits. The three-dimensional structures of potassium channel domains have recently been solved by X-ray crystallography and NMR spectroscopy. In the known K^+ channels, the "pore-forming region" of each subunit joins with its neighbors in the tetramer to constitute a channel. In *Streptomyces lividans*, the short K^+-ion selective part of the channel is lined with carbonyl oxygen atoms and accommodates two non-hydrated K^+ ions at a time, whereas Na^+ ions are too small to fit. A wider cavity of

the K^+ pathway is surrounded by helix-end dipoles.

2.4 Facilitated Diffusion

Many solutes are transported passively and selectively by so-called facilitated transporters. Such transmembrane proteins seem to consist of multi-α-helical hydrophobic regions with connecting hydrophilic loops, and the various transporters form large families of related membrane proteins from different organisms and organs with various substrate specificities. As for passage through porins and ion channels, the driving process is thermal diffusion toward lower concentration. However, the facilitated transport involves recognition and binding of the substrate, with a strictly limited range of allowed substrates for each protein. Facilitated diffusion reaches a saturation rate at high substrate concentration and can be competitively inhibited. The basic kinetic equation resembles the Michaelis–Menten equation for the rate of enzymatic reactions. The flux, J, depends hyperbolically on the substrate concentration, C:

$$J = \frac{J_{\max} \times C}{K_m + C} \qquad (6)$$

with $J_{\max}$, the maximal flux and K_m, the substrate concentration at half the maximal flux, equivalent to the Michaelis–Menten constant.

2.4.1 Examples

Detailed studies of the kinetics of transport of certain solutes have been carried through in red blood cells and model systems, for example, with GLUT1, which transports D-glucose and other monohexoses, particularly in red blood cells, epithelial cells, and the blood-brain barrier. Low expression of GLUT1 or truncation of the protein, in two infants, has been shown to lead to a low glucose concentration in the cerebrospinal fluid and delayed development and seizures; owing to insufficient transport of glucose across the blood-brain barrier. The single subunit of this oligomeric protein probably contains 12 transmembrane α-helices. GLUT1's basic functional unit is the dimer, in which both the subunits switch 100 to 1000 times per second between two alternative conformations in a synchronized way, such that one subunit can bind a substrate molecule while the other can release another substrate molecule (Fig. 6). The thermal energy seems to be too low to drive the oscillations between the two states in a single, separate monomer, but the coupled mechanism of the conformational switches in the dimer may provide a much lower activation energy. Another mechanism involving evaporation and condensation of water in the glucose-binding spaces has been proposed. The three-dimensional structure of the protein is unknown, but a model structure has been designed by combining biochemical and mutagenesis data for GLUT1 with established similar data for a related active transporter, the lactose permease from *E. coli*. The packing of the 12 helices was deduced mainly from site-specific mutagenesis experiments. The GLUT1 model indicated the existence of a hydrophilic pore with two deeply embedded glucose-binding sites.

The blood glucose level is partly regulated by insulin. In muscle and fat cells, the GLUT4 glucose transporter, which is closely related to GLUT1, can be recruited from internal membrane systems to the cell membrane to enhance the rate of glucose transport. Insulin binds to a receptor in the cell membrane and

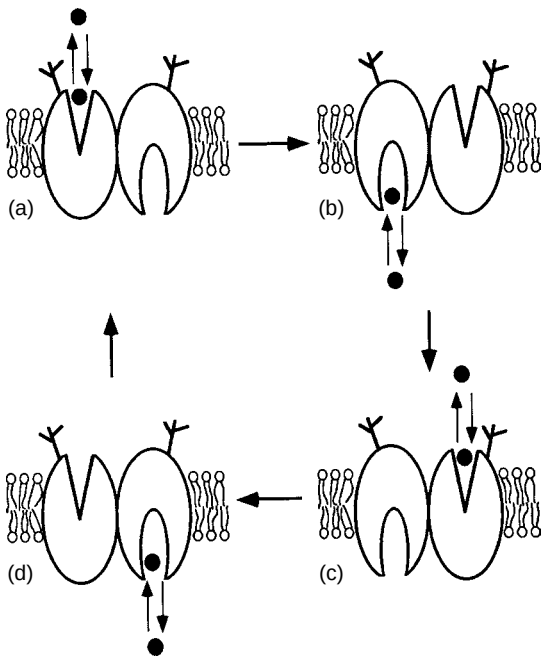

Fig. 6 The dimeric alternating-conformation model of facilitated glucose transport by GLUT1. Each subunit is shown with its oligosaccharide and, in each moment, a single substrate-binding site. The subunits are coupled to each other in the way that the change from outward- to inward-facing binding site of one subunit occurs together with a change in the opposite direction for the other subunit. (a) Glucose binds to the outward-facing site of one subunit (left), the conformation switches, and (b) the glucose molecule is released. (c) Glucose binds to the other subunit (right), the conformation switches, and (d) the glucose molecule is released. Next, the binding in (a) is repeated, and so on. Also, in the absence of glucose, the conformation alternates between the two conformations. The competitive inhibitor cytochalasin B binds only to the inward-facing opening and the number of cytochalasin B-binding sites is one per dimer in the native protein.

triggers the migration of vesicular internal membranes containing GLUT4 to the cell membrane, with which the vesicles then fuse. Too low insulin concentrations, or failure of the insulin receptor to convey the insulin signal to the vesicle migration system may cause diabetes of type 1 in children, or type 2 in adults or elderly persons, respectively. The response to insulin is also modulated in other ways, for example, by the lack of a membrane protein called FAT/CD36, whereby fatty acid uptake in various tissues decreases strongly.

The transport of cholesterol and phospholipids from the high-density lipoprotein (HDL) into the brush border membranes of the small intestine is mediated by receptor proteins. HDL binds to proteins in the membranes for the transfer of the lipids. HDL also absorbs cholesterol diffusing from cell membranes or transported out of the cell by facilitated diffusion, thereby protecting against cardiovascular diseases. Low-density lipoprotein (LDL) transfers cholesterol into cells by receptor-mediated endocytosis (Sect. 2.6).

Certain "equilibrative" nucleoside transporters are very similar to the GLUTs. Two cDNAs from placenta encoding two such transporters have been expressed. These proteins are predicted to contain 11 hydrophobic transmembrane α-helices. One of the transporters was inhibited by nitrobenzylthioinosine and facilitated the transport of adenosine and chemotherapeutic nucleosides.

Nucleotide sugars are transported across the membranes of the endoplasmic reticulum and the Golgi compartments in eukaryotic cells in exchange for nucleoside monophosphates to serve in the glycosylation of proteins and lipids.

Among amino acid transporters, the system L for large, neutral amino acids

is an unusual, facilitated system, which appears as a heterodimer containing a cell-surface glycoprotein. Transport of the essential amino acid tryptophan and of tyrosine, both of which are precursors of signal substances in the brain, is mediated by the recently discovered TAT protein.

2.4.2 Measurements

Facilitated diffusion can be analyzed by the use of cells, cell membrane vesicles, or proteoliposomes, which are incubated with radioactively labeled substrates in most types of analyses. This closely resembles permeability measurements discussed briefly in Sect. 2.1.2, although the facilitated diffusion usually is much more rapid and allows the use of inhibitors. Both these circumstances make the determinations relatively easy. Human red blood cells or *X. laevis* oocytes have often been used. The oocytes allow expression of membrane proteins that they normally do not contain. The transport is stopped by the use of an inhibitor, if available, and the cells and so on that contain the transported substrate are separated from free substrate by centrifugation, filtration, or size-exclusion chromatography. The radioactive substrate may be added outside the cells and so on, for determination of inward transport, or loaded inside before washing to determine outward transport. For equilibrium exchange measurements, the substrate is present on both sides of the membrane at the same concentration, whereas the radioactive tracer is added to one side only.

2.5 Active Transport

In contrast to the passive facilitated diffusion, which uses thermal energy only, active transport of certain substances, for example, sodium and calcium ions, amino acids, and sugars, against their concentration gradients requires chemical energy provided by hydrolysis of ATP; light energy, as in the purple membrane (primary active transport), or energy supplied indirectly by the electromotive force of an ion gradient (secondary active transport). In the latter case, the concentrated substrate is cotransported with an ion. The two transported entities can either migrate in the same direction through a transporter called symport, or in opposite directions through an antiport.

2.5.1 Examples

A primary active transporter is the P-type (Na^+, K^+)-ATPase, which transports sodium ions in one direction, out from the cell, and potassium ions in the opposite direction, both against the concentration differences, by use of energy obtained by hydrolysis of ATP. Similarly, as for facilitated diffusion, a Michaelis–Menten constant, K_m, can be assigned to each of the ions and to each of the components, ATP, adenosine diphosphate (ADP), and phosphate ions. The protein is composed of α/β-heterodimers of an α-subunit containing 10 transmembrane α-helices with the longest connecting loop between helices 4 and 5, and a β-subunit with a single transmembrane helix and a relatively long periplasmic polypeptide chain with three Cys–Cys cross-links. The α-subunit seems to alternate between two distinctly different conformations.

The (Na^+, K^+)-ATPases form one group of cation transporters together with the closely related Ca^{2+}-ATPases (Fig. 7). One of these autophosphorylated, P-type, ATPases of M_r 110 000 pumps Ca^{2+} released in skeletal muscles back into the sarcoplasmic reticulum. This protein has

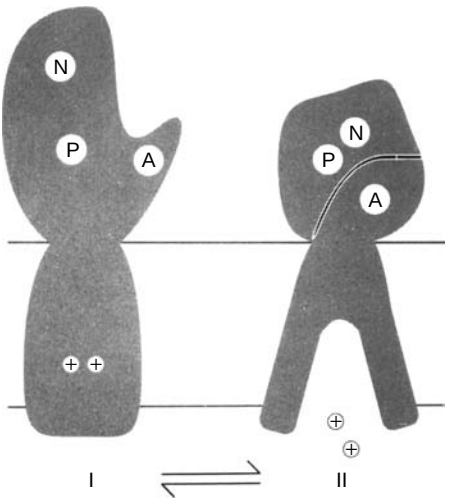

Fig. 7 Mechanism of a P-type ATPase. A schematic drawing of the conformational changes in the α-subunit of the sarcoplasmic reticulum Ca^{2+}-ATPase upon ion transport. This transporter has been crystallized and the structure was solved. It is thought to be closely similar to Na^+, K^+-ATPases. P, phosphorylation domain, A, actuator domain, and N, nucleotide binding domain. To the left (I) the protein is open to the cytoplasm but closed to the A domain. In (II), the protein is open to the membrane but closed to the cytoplasm. The change from I to II involves motion of the N and P domains toward each other and a rotation of the A domain. Reprinted with permission from J.H. Kaplan and the copyright owner. From Kaplan, J.H. (2002) Biochemistry of Na,K-ATPase, *Annual Reviews of Biochemistry*. **71**, 511–535. Copyright 2002 by Annual Reviews, www.annualreviews.org.

been determined by X-ray diffraction analysis of protein crystals to consist of a single subunit with 10 transmembrane α-helices. The phosphorylation and adenosine-binding sites reside on separate cytoplasmic domains.

Another type of primary active transporters, the ABC transporters, belong to a family of proteins that translocate various substrates across cell membranes. They contain two transmembrane domains forming a pathway for the substrate and two ABCs, which cause a conformational change of the pathway, effecting translocation of the substrate, by use of the energy dissipated upon hydrolysis of ATP to form ADP.

There are several ABC multidrug-resistance proteins, for example, the P-glycoprotein and MRP1 (multidrug resistance protein 1). The P-glycoprotein, a membrane protein with M_r of about 170 000, works as a dimer of two homologous subunits with six transmembrane helices and one ATP binding domain per subunit. The protein transports many substances, such as xenobiotics, which have in common that they are quite hydrophobic and, therefore, accumulate in the membrane, have an M_r of 300 to 2000, and are neutral or positively charged at pH 7. At least four transport and regulatory binding sites exist, which are able to switch between high and low affinities. Only one of the ABCs hydrolyzes ATP at a time and ADP release may be rate limiting. These transporters cause problems in cancer therapy. The cancer cells become resistant to certain drugs upon overexpression of the P-glycoprotein, or other such transporters, which extrude the drugs from the cells. These proteins are also widely distributed in normal tissues, for example, in liver, intestine, and kidney.

The crystal structure has been determined for another ABC transporter, the BtuCD protein in *E. coli*, which mediates vitamin B_{12} uptake through a

transmembrane dimer containing, in all, 20 helices. ATP binding to each of the cassettes and ATP hydrolysis cause an opening of the vitamin B_{12} pathway toward the cytoplasm.

The lactose permease from *E. coli* belongs to the secondary active transporters and is a functionally monomeric H^+/lactose symport. One neutral and five charged amino acid side chains in transmembrane helices cannot be replaced in order to preserve the transport as was shown by genetic-engineering replacement of all amino acid residues, one by one, by a cysteine residue. The expression of the protein is regulated by the presence of glucose and other sugars via a mechanism involving cyclic adenosine monophosphate (cAMP). Two-dimensional crystals have been formed of both lactose permease with cytochrome b_{562} introduced in a cytoplasmic loop and of a transmembrane subunit of the *E. coli* glucose transporter. Trimers and dimers, respectively, were observed by electron microscopy techniques.

Two-dimensional subunit crystals have also been prepared of another type of sugar transporter, the phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS), which transports hexoses and releases them as phosphorylated sugars that cannot leave the cell. The PTS transporter is preferentially used for sugar transport into the cell, and the amount of other sugar transporters is regulated accordingly.

Glucose can be actively taken up in the small intestine by a secondary Na^+/glucose symport. A Na^+ gradient created by a Na^+, K^+-ATPase drives the accumulation of glucose by the Na^+/glucose transporter. Another member of this Na^+/solute symport family is the high affinity choline transporter, which pumps choline into the presynaptic terminals of cholinergic neurons and thereby regulates the rate of acetylcholine synthesis. Changes in the function of cholinergic neurons are thought to be involved in Alzheimer's disease, Down's syndrome, Parkinson's disease, and schizophrenia.

2.5.2 Measurements

The primary active transport of nutrients and metabolites is determined similarly as passive transport, commonly by use of radioactive tracers. The ATP hydrolysis can be monitored by use of coupled reactions, for example, with pyruvate kinase and lactate dehydrogenase, or by use of separation followed by phosphorus analysis. Ion transport may be determined by the use of an ion-sensitive electrode, and the transmembrane potential and current can be monitored in the case of large cells or planar membranes. Fluorescent and light-absorbing probes sensitive to various parameters, such as pH, are useful tools.

2.6 Endo-, Exo-, and Transcytosis

Endocytosis imports molecules into the cell from the surrounding medium and exocytosis exports molecules from the cytoplasm out to the surroundings by budding off vesicles containing the molecules from the cell membrane or fusing such vesicles with the membrane (Fig. 4d). Receptor-mediated endocytosis is encountered in many cases, for example, upon internalization of LDL lipoproteins bound to LDL receptors in pits coated with the protein clathrin. The binding is followed

by an invagination of the cell membrane followed by the budding off of a membrane vesicle within the cell. In another example, receptors activated by the epidermal growth factor are internalized through clathrin-coated pits. Endothelial cells allow transcytosis of proteins by so-called cargo-vesicles and in transendothelial channels formed by such vesicles. These processes are mediated by protein machines. Exocytosis in eukaryotic cells involves an "exocyst," a complex of eight proteins that enables the secretory vesicle to dock on and fuse with the cell membrane.

3 Transport Across Epithelia

3.1 Intestinal Epithelia

The gut absorbs water, nutrients, and certain ions while excluding bacteria and unwanted molecules. Active and passive transporters allow the uptake of sugars, amino acids, nucleosides, and small peptides. For example, glucose is actively absorbed into the epithelial cell by the Na^+/glucose symport and passes on into the blood by facilitated diffusion mediated by GLUT5. The gaps between the apical parts of the epithelial enterocytes are sealed by tight junctions, allowing the passage of small molecules only. The elimination of xenobiotics, such as drugs, is handled by P-glycoproteins that use ATP to drive expulsion of the foreign compounds that have generally entered the cells by partition–diffusion. Drugs are also metabolized by enzymes, for example, cytochrome P450 proteins. Cell lines, such as Caco-2, are cultured to form monolayers that resemble the intestinal epithelia, and P-glycoprotein expulsion of drugs can be observed as a lower permeation from the apical to the basolateral side (corresponding to transport from the gut to the blood), than in the opposite direction.

3.2 The Blood-brain Barrier

The blood-brain barrier (BBB) consists of the endothelial cells lining the capillary blood vessels in the brain and separates the blood from the extracellular matrix around the astrocytes. Tight junctions of high electrical resistance between the cells limit the paracellular flow, and the amount of endocytotic vesicles is low in the cells, which restricts the transcellular flow. Certain molecules are allowed into the brain by facilitated diffusion via membrane proteins, notably the brain fuel D-glucose, which enters through GLUT1, L-ascorbic acid (vitamin C), and certain amino acids and similar molecules (L-3,4-dihydroxyphenylalanine, L-DOPA). Vitamin C is transported into the brain in the choroid plexus by Na^+-dependent active transporters. Dehydroascorbic acid can be transported by GLUT1 to be reduced in the brain, but this pathway is physiologically unimportant. Even certain macromolecules are believed to enter the brain via a receptor-mediated mechanism, and *Streptococcus pneumoniae* and six more bacterial pathogens can interact with the endothelial cells in an unknown way that allows the bacteria to cross the barrier. The P-glycoprotein is present in substantial amounts in the endothelial cells and tries to eject back into the blood unwanted lipophilic molecules that have diffused into the brain or just into endothelial cells. The embryonic and postnatal development of the BBB is

still unclear, and the use of cultured endothelial cells to mimic the physiological properties of the BBB is haunted by difficulties.

4
Solute Affinities for Membrane Proteins

Biospecific interaction between a solute, for example, an inhibitor, and a membrane protein in cells, membrane vesicles, proteoliposomes or other lipid environments can be quantified by immobilized-biomembrane affinity chromatography. Entrapment in gel bead cavities and covalent binding of phospholipid analogues are examples of the immobilization methods used. The retardation of the front of an eluting large-volume sample of the solute reflects the solute affinity for the protein and the solute concentration. In simple cases, a series of runs at different solute concentrations allow calculation of the equilibrium constant and the number of binding sites, provided that the number of binding sites matches the affinity. The glucose transporter from human red blood cells and a recombinant P-glycoprotein showed constant interactant affinities for months in the columns. The solute retardation is caused both by partitioning into the membrane and by binding to active sites of the membrane proteins. Both effects can be measured. A competitor for the solute–protein interaction in the system will decrease the retention volume of the solute to a degree determined by its affinity for the protein. Ligand binding to membrane proteins can also be analyzed by surface-plasmon resonance, and, in some cases, by analysis of changes in the intrinsic fluorescence of tryptophan side chains or other spectral properties upon ligand binding to the protein. A number of related techniques exist, but the fact that most membrane proteins require a lipid environment for activity limits the choice.

5
Prediction and Modeling of Drug Absorption Across Membranes

Some basic requirements for oral administration of new drug candidates are sufficient solubility in water, reasonable partitioning into membranes, and a fair diffusion coefficient. The partitioning is commonly modeled in the pharmaceutical industry by octanol/water partitioning, expressed as the logarithm of the partition coefficient, log P_{oct}. The advantages of log P_{oct} are the vast amount of available data and the extensive experience of the applicability of these data. Lipinski introduced the "rule of five," which can be used as a filter in early drug development. This rule predicts that poor absorption or permeability properties become likely when the drug molecule has more than five H-bond donors, more than 10 H-bond acceptors, when the M_r exceeds 500 and when the log P_{oct} is larger than 5.

Methods for experimental assessment of partitioning into membranes and permeability across membranes are required, although computer calculations are becoming increasingly popular for theoretical analysis. Computational methods for determination of absorption or permeability properties by statistical analysis often use molecular descriptors, such as the polar surface area of the drug and the number of H-bond donors/acceptors.

Molecular dynamics simulations are still at a primitive stage. For example,

computer modeling indicates that the diffusion coefficients of O_2 and H_2O are highest in the center of the membrane, 0.9×10^{-5} and 1.4×10^{-5} $cm^2\ s^{-1}$, respectively. An analog (M_r 315) of the drug nifedipine, obtained by replacing NO_2 by CH_3, had a diffusion coefficient of 1.3×10^{-6} $cm^2\ s^{-1}$ all across the hydrophobic region of the membrane. Furthermore, pharmacokinetic problems tend to override the shortcomings of the computer predictions of drug absorption. Absorption is a complex process dependent on factors such as the permeability and solubility of the drug, and physiological factors such as pH, intestinal enzymology and motility, so the absorption is therefore not so easily predicted from the structure alone or from computer simulations of drug-membrane interaction.

6 Conclusions and Visions

The permeability of lipid bilayers or biological membranes for a given solute is generally difficult to determine experimentally and can only be estimated approximately from the chemical structure of the solute or by the use of partition data with either octanol–water or membrane–water. These approaches do not truly mimic the interactions between solutes and membranes. Partition analyses with water–membrane systems do not reveal the detailed solute distribution within the membrane that determines the gradient in solute concentration and thereby the solute flux across the membrane. The solute permeability of membranes may, in the future, be determined by use of sensitive detection in miniaturized devices. Calculations by use of the detailed structures of the solute and the membrane may become more accurate after advances in quantum chemical or physical–chemical methods. In the future, the transport kinetics and transport mechanisms of transmembrane proteins may be evaluated by small-scale permeability assays and advances in crystallization technology, or single-protein molecule diffraction analyses may allow determination of the three-dimensional structures of membrane proteins.

See also Intracellular Fatty Acid Binding Proteins and Fatty Acid Transport; Membrane Traffic: Vesicle Budding and Fusion.

Bibliography

Books and Reviews

Barrett, M.P., Walmsley, A.R., Gould, G.W. (1999) Structure and function of facilitative sugar transporters, *Curr. Opin. Cell Biol.* **11**, 496–502.

Burton, P.S., Goodwin, J.T., Vidmar, T.J., Amore, B.M. (2002) Predicting drug absorption: how nature made it a difficult problem, *J. Pharmacol. Exp. Ther.* **303**, 889–895.

Deamer, D.W., Kleinzeller, A., Fambrough, D.M. (Eds.) (1999) *Current Topics in Membranes*, Vol. 48, *Membrane Permeability – 100 Years since Ernest Overton*, Academic Press, San Diego, CA.

Gerardy-Schahn, R., Oelmann, S., Bakker, H. (2001) Nucleotide sugar transporters: biological and functional aspects, *Biochimie* **83**, 775–782.

Jentsch, T.J., Stein, V., Weinreich, F., Zdebik, A.A. (2002) Molecular structure and physiological function of chloride channels, *Physiol. Rev.* **82**, 503–568.

Kaback, H.R., Sahin-Tóth, M., Weinglass, A.B. (2001) The kamikaze approach to membrane transport, *Nat. Rev. Mol. Cell Biol.* **2**, 610–620.

Kaplan, J.H. (2002) Biochemistry of Na,K-ATPase, *Annu. Rev. Biochem.* **71**, 511–535.

Lin, G. (2002) Insights of high-density lipoprotein apolipoprotein-mediated lipid efflux from cells, *Biochem. Biophys. Res. Commun.* **291**, 727–731.

Mannhold, R., Kubinyi, H., Folkers, G. (Eds.) (2002) *Methods and Principles in Medicinal Chemistry,* Vol. 15, *Drug-Membrane Interactions – Analysis, Drug Distribution, Modeling,* Wiley-VCH, Weinheim.

Merz, K.M. Jr., Roux, B. (Eds.) (1996) *Biological Membranes: A Molecular Perspective from Computation and Experiment,* Birkhäuser, Boston, MA.

Rubin, L.L., Staddon, J.M. (1999) The cell biology of the blood-brain barrier, *Annu. Rev. Neurosci.* **22**, 11–28.

Sauna, Z.E., Smith, M.M., Müller, M., Kerr, K.M., Ambudkar, S.V. (2001) The mechanism of action of multidrug-resistance-linked P-glycoprotein, *J. Bioenerg. Biomembr.* **33**, 481–491.

Simionescu, M., Gafencu, A., Antohe, F. (2002) Transcytosis of plasma macromolecules in endothelial cells: a cell biological survey, *Microsc. Res. Tech.* **57**, 269–288.

Sorkin, A., von Zastrow, M. (2002) Signal transduction and endocytosis: close encounters of many kinds, *Nat. Rev. Mol. Cell Biol.* **3**, 600–614.

Stein, W.D. (1990) *Channels, Carriers, and Pumps – An Introduction to Membrane Transport,* Academic Press, San Diego, CA.

Primary Literature

Appleman, J.R., Lienhard, G.E. (1985) Rapid kinetics of the glucose transporter from human erythrocytes. Detection and measurement of a half-turnover of the purified transporter, *J. Biol. Chem.* **260**, 4575–4578.

Artursson, P. (1990) Epithelial transport of drugs in cell culture. I: A model for studying the passive diffusion of drugs over intestinal absorptive (Caco-2) cells, *J. Pharm. Sci.* **79**, 476–482.

Artursson, P., Karlsson, J. (1991) Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells, *Biochem. Biophys. Res. Commun.* **175**, 880–885.

Avdeef, A., Box, K.J., Comer, J.E.A., Hibbert, C., Tam, K.Y. (1998) pH-Metric logP 10. Determination of liposomal membrane-water partition coefficients of ionizable drugs, *Pharm. Res.* **15**, 209–215.

Baird, C.L., Courtenay, E.S., Myszka, D.G. (2002) Surface plasmon resonance characterization of drug/liposome interactions, *Anal. Biochem.* **310**, 93–99.

Beigi, F., Yang, Q., Lundahl, P. (1995) Immobilized-liposome chromatographic analysis of drug partitioning into lipid bilayers, *J. Chromatogr. A* **704**, 315–321.

Cheng, A., van Hoek, A.N., Yeager, M., Verkman, A.S., Mitra, A.K. (1997) Three-dimensional organization of a human water channel, *Nature* **387**, 627–630.

Danelian, E., Karlén, A., Karlsson, R., Winiwarter, S., Hansson, A., Löfås, S., Lennernäs, H., Hämäläinen, M.D. (2000) SPR biosensor studies of the direct interaction between 27 drugs and a liposome surface: correlation with fraction absorbed in humans, *J. Med. Chem.* **43**, 2083–2086.

Doyle, D.A., Morais Cabral, J., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T., MacKinnon, R. (1998) The structure of the potassium channel: molecular basis of K^+ conduction and selectivity, *Science* **280**, 69–77.

Gottschalk, I., Gustavsson, P.-E., Ersson, B., Lundahl, P. (2003) Improved lectin-mediated immobilization of human red blood cells in superporous agarose beads, *J. Chromatogr. B* **784**, 203–208.

Gottschalk, I., Lundqvist, A., Zeng, C.-M., Lagerquist Hägglund, C., Zuo, S.-S., Brekkan, E., Eaker, D., Lundahl, P. (2000) Conversion between two cytochalasin B-binding states of the human GLUT1 glucose transporter, *Eur. J. Biochem.* **267**, 6875–6882.

Griffiths, M., Yao, S.Y.M., Abidi, F., Phillips, S.E.V., Cass, C.E., Young, J.D., Baldwin, S.A. (1997) Molecular cloning and characterization of a nitrobenzylthioinosine-insensitive (*ei*) equilibrative nucleoside transporter from human placenta, *Biochem. J.* **328**, 739–743.

Griffiths, M., Beaumont, N., Yao, S.Y.M., Sundaram, M., Boumah, C.E., Davies, A., Kwong, F.Y.P., Coe, I., Cass, C.E., Young, J.D., Baldwin, S.A. (1997) Cloning of a human nucleoside transporter implicated in the cellular uptake of adenosine and chemotherapeutic drugs, *Nat. Med.* **3**, 89–93.

Hajri, T., Han, X.X., Bonen, A., Abumrad, N.A. (2002) Defective fatty acid uptake modulates

insulin responsiveness and metabolic responses to diet in CD36-null mice, *J. Clin. Invest.* **109**, 1381–1389.

Hamill, S., Cloherty, E.K., Carruthers, A. (1999) The human erythrocyte sugar transporter presents two sugar import sites, *Biochemistry* **38**, 16974–16983.

Harik, S.I., Hall, A.K., Richey, P., Andersson, L., Lundahl, P., Perry, G. (1993) Ontogeny of the erythroid/HepG2-type glucose transporter (GLUT-1) in the rat nervous system, *Dev. Brain. Res.* **72**, 41–49.

Hitchcock, P.B., Mason, R., Thomas, K.M., Shipley, G.G. (1974) Structural chemistry of 1,2 dilauroyl-DL-phosphatidylethanolamine: molecular conformation and intermolecular packing of phospholipids, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3036–3040.

Koronakis, V., Sharff, A., Koronakis, E., Luisi, B., Hughes, C. (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export, *Nature* **405**, 914–919.

Kreusch, A., Pfaffinger, P.J., Stevens, C.F., Choe, S. (1998) Crystal structure of the tetramerization domain of the *Shaker* potassium channel, *Nature* **392**, 945–948.

Lagerquist, C., Beigi, F., Karlén, A., Lennernäs, H., Lundahl, P. (2001) Effects of cholesterol and model transmembrane proteins on drug partitioning into lipid bilayers as analysed by immobilized-liposome chromatography, *J. Pharm. Pharmacol.* **53**, 1477–1487.

Lipinski, C.A. (2000) Drug-like properties and the causes of poor solubility and poor permeability, *J. Pharmacol. Toxicol. Methods* **44**, 235–249.

Locher, K.P., Lee, A.T., Rees, D.C. (2002) The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism, *Science* **296**, 1091–1098.

Lux, S.E., John, K.M., Kopito, R.R., Lodish, H.F. (1989) Cloning and characterization of band 3, the human erythrocyte anion-exchange protein (AE1), *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9089–9093.

Martin, C., Berridge, G., Higgins, C.F., Mistry, P., Charlton, P., Callaghan, R. (2000) Communication between multiple drug binding sites on P-glycoprotein, *Mol. Pharmacol.* **58**, 624–632.

Mastroberardino, L., Spindler, B., Pfeiffer, R., Skelly, P.J., Loffing, J., Shoemaker, C.B., Verrey, F. (1998) Amino-acid transport by heterodimers of 4F2hc/CD98 and members of a permease family, *Nature* **395**, 288–291.

Molero-Monfort, M., Martín-Biosca, Y., Sagrado, S., Villanueva-Camañas, R.M., Medina-Hernández, M.J. (2000) Micellar liquid chromatography for prediction of drug transport, *J. Chromatogr. A* **870**, 1–11.

Mueckler, M., Caruso, C., Baldwin, S.A., Panico, M., Blench, I., Morris, H.R., Allard, W.J., Lienhard, G.E., Lodish, H.F. (1985) Sequence and structure of a human glucose transporter, *Science* **229**, 941–945.

Naren, A.P., Cormet-Boyaka, E., Fu, J., Villain, M., Blalock, J.E., Quick, M.W., Kirk, K.L. (1999) CFTR chloride channel regulation by an interdomain interaction, *Science* **286**, 544–548.

Noda, M., Shimizu, S., Tanabe, T., Takai, T., Kayano, T., Ikeda, T., Takahashi, H., Nakayama, H., Kanaoka, Y., Minamino, N., Kangawa, K., Matsuo, H., Raftery, M.A., Hirose, T., Inayama, S., Hayashida, H., Miyata, T., Numa, S. (1984) Primary structure of *Electrophorus electricus* sodium channel deduced from cDNA sequence, *Nature* **312**, 121–127.

Okuda, T., Okamura, M., Kaitsuka, C., Haga, T., Gurwitz, D. (2002) Single nucleotide polymorphism of the human high affinity choline transporter alters transport rate, *J. Biol. Chem.* **277**, 45315–45322.

Papazian, D.M., Schwarz, T.L., Tempel, B.L., Jan, Y.N., Jan, Y.L. (1987) Cloning of genomic and complementary DNA from *Shaker*, a putative potassium channel gene from *Drosophila*, *Science* **237**, 749–753.

Paula, S., Volkov, A.G., Van Hoek, A.N., Haines, T.H., Deamer, D.W. (1996) Permeation of protons, potassium ions, and small polar molecules through phospholipid bilayers as a function of membrane thickness, *Biophys. J.* **70**, 339–348.

Pessino, A., Hebert, D.N., Woon, C.W., Harrison, S.A., Clancy, B.M., Buxton, J.M., Carruthers, A., Czech, M.P. (1991) Evidence that functional erythrocyte-type glucose transporters are oligomers, *J. Biol. Chem.* **266**, 20213–20217.

Ramsden, J.J. (1993) Partition coefficients of drugs in bilayer lipid membranes, *Experientia* **49**, 688–692.

Saviane, C., Conti, F., Pusch, M. (1999) The muscle chloride channel ClC-1 has a double-barreled appearance that is differentially affected in dominant and recessive myotonia, *J. Gen. Physiol.* **113**, 457–467.

Schoenmakers, R.G., Stehouwer, M.C., Tukker, J.J. (1999) Structure-transport relationship for the intestinal small-peptide carrier: is the carbonyl group of the peptide bond relevant for transport? *Pharm. Res.* **16**, 62–68.

Seidner, G., Garcia Alvarez, M., Yeh, J.-I., O'Driscoll, K.R., Klepper, J., Stump, T.S., Wang, D., Spinner, N.B., Birnbaum, M.J., De Vivo, D.C. (1998) GLUT-1 deficiency syndrome caused by haploinsufficiency of the blood-brain barrier hexose carrier, *Nat. Genet.* **18**, 188–191.

Singer, S.J., Nicolson, G.L. (1972) The fluid mosaic model of the structure of cell membranes. Cell membranes are viewed as two-dimensional solutions of oriented globular proteins and lipids, *Science* **175**, 720–731.

Sui, H., Han, B.-G., Lee, J.K., Walian, P., Jap, B.K. (2001) Structural basis of water-specific transport through the AQP1 water channel, *Nature* **414**, 872–878.

Sultzman, L.A., Carruthers, A. (1999) Stop-flow analysis of cooperative interactions between GLUT1 sugar import and export sites, *Biochemistry* **38**, 6640–6650.

Suzuki, K., Kono, T. (1980) Evidence that insulin causes translocation of glucose transport activity to the plasma membrane from an intracellular storage site, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2542–2545.

Tanabe, T., Takeshima, H., Mikami, A., Flockerzi, V., Takahashi, H., Kangawa, K., Kojima, M., Matsuo, H., Hirose, T., Numa, S. (1987) Primary structure of the receptor for calcium channel blockers from skeletal muscle, *Nature* **328**, 313–318.

Toyoshima, C., Nakasako, M., Nomura, H., Ogawa, H. (2000) Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution, *Nature* **405**, 647–655.

Tsukaguchi, H., Tokui, T., Mackenzie, B., Berger, U.V., Chen, X.-Z., Wang, Y., Brubaker, R.F., Hediger, M.A. (1999) A family of mammalian Na^+-dependent L-ascorbic acid transporters, *Nature* **399**, 70–75.

Walz, T., Hirai, T., Murata, K., Heymann, J.B., Mitsuoka, K., Fujiyoshi, Y., Smith, B.L., Agre, P., Engel, A. (1997) The three-dimensional structure of aquaporin-1, *Nature* **387**, 624–627.

Wang, Y., Mackenzie, B., Tsukaguchi, H., Weremowicz, S., Morton, C.C., Hediger, M.A. (2000) Human vitamin C (L-ascorbic acid) transporter SVCT1, *Biochem. Biophys. Res. Commun.* **267**, 488–494.

Werder, M., Han, C.-H., Wehrli, E., Bimmler, D., Schulthess, G., Hauser, H. (2001) Role of scavenger receptors SR-BI and CD36 in selective sterol uptake in the small intestine, *Biochemistry* **40**, 11643–11650.

Widdas, W.F. (1998) The glucose transporter of human erythrocytes – working hypothesis for its functional mechanism, *Exp. Physiol.* **83**, 187–194.

Winiwarter, S., Bonham, N.M., Ax, F., Hallberg, A., Lennernäs, H., Karlén, A. (1998) Correlation of human jejunal permeability (in vivo) of drugs with experimentally and theoretically derived parameters. A multivariate data analysis approach, *J. Med. Chem.* **41**, 4939–4949.

Xiang, T.-X., Xu, Y.-H., Anderson, B.D. (1998) The barrier domain for solute permeation varies with lipid bilayer phase structure, *J. Membr. Biol.* **165**, 77–90.

Zhang, J.-Z., Ismail-Beigi, F. (1998) Activation of Glut1 glucose transporter in human erythrocytes, *Arch. Biochem. Biophys.* **356**, 86–92.

Zhang, Y., Leonessa, F., Clarke, R., Wainer, I.W. (2000) Development of an immobilized P-glycoprotein stationary phase for on-line liquid chromatographic determination of drug-binding affinities, *J. Chromatogr. B* **739**, 33–37.

Zhuang, J., Gutknecht, R., Flükiger, K., Hasler, L., Erni, B., Engel, A. (1999) Purification and electron microscopic characterization of the membrane subunit ($IICB^{Glc}$) of the *Escherichia coli* glucose transporter, *Arch. Biochem. Biophys.* **372**, 89–96.

Zhuang, J., Privé, G.G., Werner, G.E., Ringler, P., Kaback, H.R., Engel, A. (1999) Two-dimensional crystallization of *Escherichia coli* lactose permease, *J. Struct. Biol.* **125**, 63–75.

Zuniga, F.A., Shi, G., Haller, J.F., Rubashkin, A., Flynn, D.R., Iserovich, P., Fischbarg, J. (2001) A three-dimensional model of the human facilitative glucose transporter Glut1, *J. Biol. Chem.* **276**, 44970–44975.

Österberg, T., Svensson, M., Lundahl, P. (2001) Chromatographic retention of drug molecules on immobilised liposomes prepared from egg phospholipids and from chemically pure phospholipids, *Eur. J. Pharm. Sci.* **12**, 427–439.

Memory: *see* Immunologic Memory

Metabolic Basis of Cellular Energy

Meredith F. Ross and Michael P. Murphy
MRC Dunn Human Nutrition Unit, Cambridge, UK

1 Cellular Energetics: Displacement from Equilibrium, Coupling, and the Role of ATP 166
1.1 Coupling Drives Energetically Unfavorable Reactions 166
1.2 ATP Synthesis and Hydrolysis as a Coupling Reaction 169
1.3 Energy is Stored in the Displacement from Equilibrium of the ATP Hydrolysis Reaction 169

2 Energy Sources that Displace the ATP Hydrolysis Reaction from Equilibrium 170
2.1 Carbohydrate, Fat, and Protein as Energy Sources 170
2.2 Glycolysis 171
2.3 Mitochondria, the TCA Cycle and β Oxidation 171

3 Oxidation of NADH and $FADH_2$ by Mitochondria 173
3.1 The Mitochondrial Respiratory Chain 175
3.2 Energy Storage by the Protonmotive Force (Δp) 177
3.3 The Protonmotive Force Drives ATP Synthesis and Transport 180
3.4 Respiratory Control and Uncoupling 182
3.5 Alternative Uses of the Protonmotive Force 182

4 Overview of the Metabolic Basis of Cellular Energetics 184

Bibliography 185
Books and Reviews 185
Primary Literature 185

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

Keywords

Adenosine Triphosphate (ATP)
The "energy currency" of the cell. The energy derived from metabolism can be stored by synthesizing ATP from ADP and P_i, displacing the ATP hydrolysis reaction from equilibrium by many orders of magnitude. The energy stored in this displacement from equilibrium can then be used to do work as ATP is hydrolyzed back to ADP and P_i. Thus, the synthesis and hydrolysis of ATP link catabolic reactions with anabolic and work reactions in metabolism, and ATP acts as a short-term energy store.

ATP Synthase
The enzyme complex that catalyzes the synthesis of ATP from ADP and P_i. The ATP synthase makes use of the energy stored in the protonmotive force across the mitochondrial inner membrane. Protons move through the complex from the intermembrane space back into the matrix, catalyzing rotation of the central stalk (γ subunit) with respect to the catalytic α and β subunits, which synthesize ATP from ADP and P_i.

Glycolysis
The sequence of enzyme-catalyzed reactions in the cytosol that is the first stage in breaking down glucose to release its energy. The end product of glycolysis is two molecules of pyruvate per glucose, along with a small amount of ATP and reduced nicotinamide adenine dinucleotide (NADH). Pyruvate and electrons from NADH may be transported into mitochondria to be used as substrates for the TCA cycle and the respiratory chain, respectively.

Mitochondria
The central organelles of energy metabolism. Mitochondria are responsible for producing most of the cell's adenosine triphosphate (ATP). The tricarboxylic acid (TCA) cycle and β oxidation of fatty acids take place in mitochondria, as does electron transport by the respiratory chain, the process that drives ATP synthesis by the establishment of a protonmotive force across the mitochondrial inner membrane.

β Oxidation
The process by which fatty acids are broken down in mitochondria. Fatty acids are converted to fatty acyl CoA before entering the β oxidation cycle, which involves the sequential shortening of the fatty acyl chain by two carbons at a time, with the release of acetyl CoA and the concomitant reduction of the electron carriers NAD^+ and FAD to NADH and $FADH_2$, respectively. The acetyl CoA molecules enter the TCA cycle, while the NADH and $FADH_2$ are oxidized by the respiratory chain.

Oxidative Phosphorylation
The process by which the oxidation of carbohydrates and fatty acids by oxygen is coupled to the phosphorylation of adenosine diphosphate (ADP) to ATP. Oxidative phosphorylation occurs in mitochondria when the NADH and $FADH_2$ produced by the

TCA cycle and β oxidation are oxidized by the respiratory chain. The redox energy is conserved as a protonmotive force across the mitochondrial inner membrane, which is used to drive ATP synthesis by the ATP synthase in the mitochondrial inner membrane.

Protonmotive Force
A measure (in mV) of the proton electrochemical potential gradient across the mitochondrial inner membrane. The protonmotive force is generated by the pumping of protons by the mitochondrial respiratory chain, which in turn is driven by the release of redox energy from electrons as they pass through the chain. The protonmotive force is composed of a membrane potential ($\Delta\psi$) and a pH gradient (ΔpH) component. It is used by the ATP synthase to synthesize ATP and by the adenine nucleotide transporter (ANT) and phosphate (P_i) carrier to transport ATP, ADP, and P_i across the mitochondrial inner membrane.

Respiratory Chain
A series of protein complexes and mobile electron carriers in the mitochondrial inner membrane that translocates electrons down a gradient of increasing redox potential. The respiratory chain oxidizes the carriers NADH and $FADH_2$ and passes the electrons down the chain toward oxygen, using the redox energy differences to pump protons across the mitochondrial inner membrane. This enables the redox energy to be stored as a protonmotive force, which is then used to synthesize ATP from ADP and P_i.

Tricarboxylic Acid (TCA) Cycle
Also known as the *Krebs cycle* or the *citric acid cycle*. The TCA cycle is a sequence of enzyme-catalyzed reactions that occur in the mitochondrial matrix. Acetyl CoA (produced by the oxidative decarboxylation of pyruvate) acts as the entry point for the TCA cycle by fusing its two-carbon acetyl group with the four-carbon oxaloacetate to form the six-carbon citrate. Citrate is then sequentially oxidized by the various enzymes of the TCA cycle, losing two carbons as CO_2 to regenerate oxaloacetate while producing NADH and reduced flavin adenine dinucleotide ($FADH_2$), and also generating GTP by substrate level phosphorylation. Many other small molecules connect with the TCA cycle at various points, making the TCA cycle central for most of metabolism. The NADH and $FADH_2$ produced are oxidized by the respiratory chain to generate ATP by oxidative phosphorylation.

All cell processes require energy to carry out work, whether it be synthesis, replication, secretion, movement, or signaling. Consequently, a continuous supply of energy to the cell is essential for life. Ultimately, the energy for mammalian cells is provided by the breakdown of food. In order to make this chemical energy available to the cell in a form that can do many different kinds of work, a common intermediate is required. This energy intermediate is the displacement from equilibrium of the

$ATP \rightleftharpoons ADP + P_i$ reaction. Within our cells, this reaction is displaced from equilibrium by increasing the ATP concentration relative to that of ADP by several orders of magnitude, making approximately 50 $kJ.mol^{-1}$ of energy available from the hydrolysis of ATP to do work. Maintenance of this displacement from equilibrium requires the continued expenditure of energy supplied by food. Carbohydrate derivatives and fatty acids are broken down and oxidized by a series of reactions within mitochondria. Released electrons are transferred to NAD^+ and FAD, reducing them to NADH and $FADH_2$. In turn, these electron carriers are oxidized by the respiratory chain in the mitochondrial inner membrane, using the oxygen we breathe as the final electron acceptor. The action of the respiratory chain pumps protons across the mitochondrial inner membrane so that the redox potential energy is stored as an electrochemical potential gradient, or protonmotive force, across the mitochondrial inner membrane. The protonmotive force is then used to drive the synthesis of ATP from ADP and P_i and to export the ATP from the mitochondria into the cytoplasm. This process, called *oxidative phosphorylation*, maintains the displacement from equilibrium of the $ATP \rightleftharpoons ADP + P_i$ reaction and is central to cellular energetics.

1 Cellular Energetics: Displacement from Equilibrium, Coupling, and the Role of ATP

Living things are distinguished by the displacement from equilibrium of a myriad biochemical reactions. Indeed, one way of describing death is as a return to equilibrium. Why is displacement from equilibrium a necessity for life? Consider how potential energy is stored in a hydroelectric dam or in a tightly coiled spring. In a similar way, displacement from equilibrium provides a means of energy storage within the cell; indeed, it is only by being displaced from equilibrium that biochemical reactions can store energy that can be used to carry out cellular work such as movement, replication, growth, and repair.

Living things require large amounts of energy to maintain the displacement from equilibrium of chemical reactions within the cell. In mammals, the energy source is food, and the chemical energy stored in food is released through the reactions of metabolism. Metabolism comprises a series of catabolic reactions that break down food, releasing precursor molecules and energy for maintaining the displacement from equilibrium that drives anabolic reactions and work. The energy released by the catabolic reactions of metabolism must be available in a form that can do the work of the cell. A central concept underlying this process, and all of energy metabolism, is "coupling," by which the energy released from food drives otherwise unfavorable reactions. How this energy coupling occurs is the focus of this chapter.

1.1 Coupling Drives Energetically Unfavorable Reactions

By coupling together favorable and unfavorable chemical reactions, it is possible to use the energy stored in food to drive

the cellular processes that comprise life. Thermodynamics indicates whether or not a reaction will occur spontaneously and gives the energy change associated with the reaction. In biology, the Gibbs free energy change (ΔG; $kJ.mol^{-1}$) is used. For spontaneous reactions, there is a release of free energy, so $\Delta G < 0$. In contrast, reactions that do not occur spontaneously have a ΔG greater than zero and require energy input to take place. In biological systems, there are many such processes, including synthesis, movement, replication, or signaling, that are thermodynamically unfavorable and cannot occur without an energy input. Fortunately, such unfavorable reactions can be driven by coupling them to a reaction with a favorable (negative) ΔG. This is illustrated below where the unfavorable reaction 1 is driven by reaction 2.

Reaction 1	$A \rightarrow B$	$\Delta G_1 > 0$
Reaction 2	$C \rightarrow D$	$\Delta G_2 < 0$
Overall	$A + C \rightarrow B + D$	$\Delta G_1 + \Delta G_2 < 0$

This process, by which an unfavorable reaction is made possible by coupling it to a favorable reaction, is central to the bioenergetics of living organisms. The energy flowing through our cells and bodies (derived from favorable catabolic reactions) is coupled to drive unfavorable

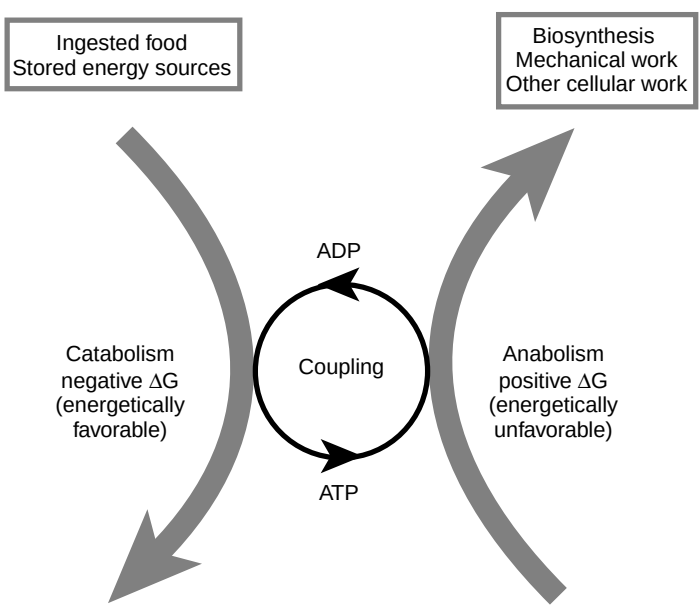

Fig. 1 The concept of coupling in mammalian energy metabolism. Energy is released by the breakdown of carbohydrate, fat, and protein, either from ingested food or from stored energy sources. Released energy is ''coupled'' to the otherwise unfavorable formation of adenosine triphosphate (ATP) by the addition of inorganic phosphate (P_i) to adenosine diphosphate (ADP). In turn, ATP hydrolysis to ADP and P_i (a highly favorable reaction) is coupled to unfavorable anabolic reactions, enabling them to take place. The Gibbs free energy change (ΔG) is used here to indicate whether a reaction is favorable (spontaneous) or unfavorable (requiring energy input through coupling).

reactions (anabolic reactions and work) (Fig. 1). In principle, many different reactions could be coupled together in this manner. However, biology has evolved to exploit a few general reactions that are used to couple most catabolic and anabolic reactions, and the most important of these is the synthesis and hydrolysis of adenosine triphosphate (ATP) (Fig. 2):

$$ATP \rightleftharpoons ADP + P_i$$

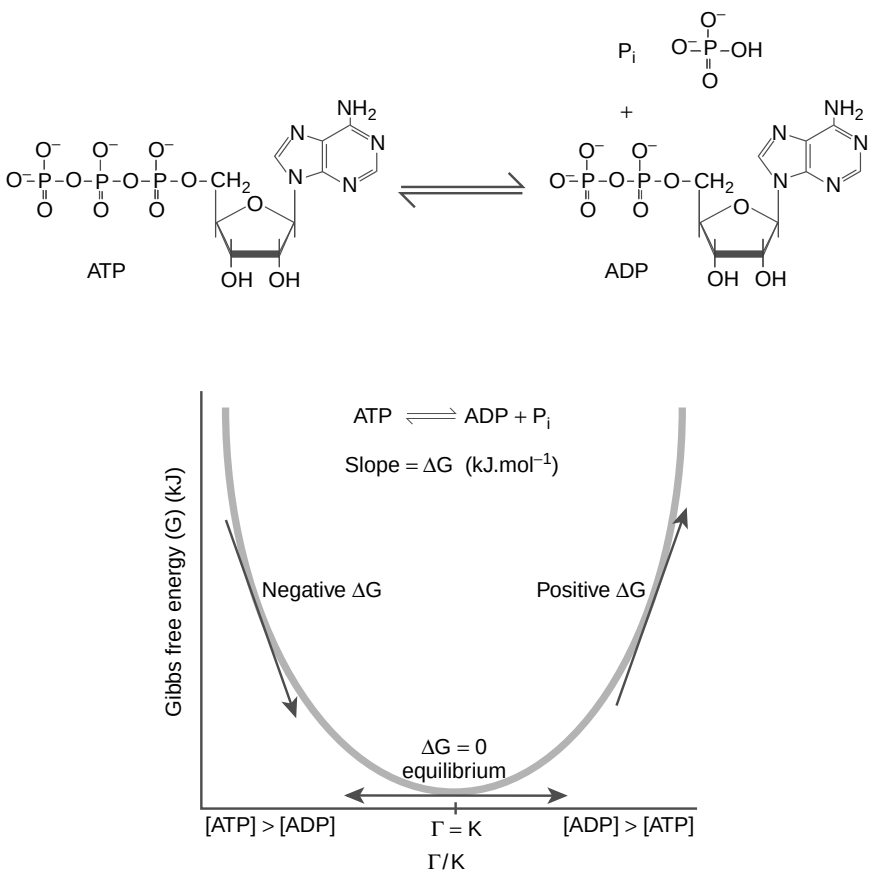

Fig. 2 Energy is stored in the displacement of ATP $\rightleftharpoons$ ADP + P_i from equilibrium. Energy from catabolic reactions is used to drive the synthesis of ATP from ADP and P_i, causing the ATP/ADP ratio to be many orders of magnitude greater than its equilibrium value (put another way, the mass action ratio (Γ) for the hydrolysis reaction is many times lower than the equilibrium constant (K)). Under these conditions, ATP hydrolysis becomes a highly favorable reaction and has a negative ΔG (indicated in the graph as the negative slope of the curve under conditions of high ATP concentration). When the ATP/ADP ratio is lower than its equilibrium value, ATP hydrolysis is not favored ($\Delta G > 0$); rather, the reverse reaction is energetically favorable. At equilibrium, the ATP/ADP ratio is equal to its equilibrium value (the mass action ratio equals the equilibrium constant), the $\Delta G = 0$, and neither the forward nor the reverse reactions are favored.

1.2 ATP Synthesis and Hydrolysis as a Coupling Reaction

The energy made available from the catabolism of food is funneled into a few forms that can be used in a wide range of anabolic and work reactions. The main reaction "chosen" by evolution for this purpose is the $ADP + P_i \rightleftharpoons ATP$ reaction. The energetically favorable reactions of catabolism are coupled to the synthesis of ATP from adenosine diphosphate (ADP) and P_i, and, in turn, the energy released by hydrolysis of ATP to ADP is coupled to work and anabolic reactions (Figs. 1 and 2). There are a number of reasons why this reaction plays a central role in biology. The equilibrium constant of the reaction enables it to straddle both the highly exergonic catabolic reactions and the highly endergonic anabolic reactions. Furthermore, coupling requires that the reaction be easily combined with many others within the active sites of enzymes, and phosphate transfer or ATP hydrolysis can be coupled to a wide range of reactions. Finally, while the hydrolysis of ATP at the biological concentrations of ATP, ADP, and P_i must be thermodynamically favorable in order to act as an energy store, it must not occur without the action of enzymes capable of harnessing the released energy for useful functions. The phosphodiester bonds in ATP fulfill this role as they are kinetically stable *in vivo*, but the activation energy for hydrolysis of the phosphodiester bond is easily decreased by appropriate arrangements within the active sites of enzymes to stabilize the pentagonal bipyramidal transition states. For these and other reasons, the $ATP \rightleftharpoons ADP + P_i$ reaction is central to bioenergetics.

1.3 Energy is Stored in the Displacement from Equilibrium of the ATP Hydrolysis Reaction

The ATP hydrolysis reaction ($ATP \rightleftharpoons ADP + P_i$) couples catabolic and anabolic reactions by storing energy in its displacement from equilibrium. As is the case for all chemical reactions, the amount of energy stored is determined by the mass action ratio (Γ; a measure of the relative concentrations of reactants and products) and its relationship to the equilibrium constant (K; a measure of the relative concentrations of reactants and products *at equilibrium*).

For ATP hydrolysis the mass action ratio is

$$\Gamma = [ADP][P_i]/[ATP]$$

The equilibrium constant is the special case of this reaction in which the reactants and products are at equilibrium ($K \sim 10^5$ M at pH 7). To see how energy is stored in this reaction, ΔG can be calculated from the following equation:

$$\Delta G = RT \ln(\Gamma/K)$$

This equation shows clearly the relationship between the displacement from equilibrium and the amount of energy stored, with $\Delta G = 0$ when the reaction is at equilibrium. This equation can be easily rearranged to give the more usual textbook presentation:

$$\Delta G = RT \ln\Gamma - RT \ln K$$

or, as $-RT \ln K$ is often written as ΔG^o,

$$\Delta G = \Delta G^o + RT \ln\Gamma$$

This equation shows that ΔG^o is merely another way of stating the equilibrium constant, and illustrates how misleading

are discussions of cellular thermodynamics based on ΔG^o. It also emphasizes that the relevant concentrations of the reactants and products must be used in calculating ΔG.

In vivo, catabolic reactions hold the cytoplasmic concentrations of ATP $\sim$ 5 mM and ADP $\sim$ 100 to 500 μM, corresponding to a displacement from equilibrium of 8 to 10 orders of magnitude. The role of this displacement from equilibrium in energy storage is further illustrated in Fig. 2, which shows ΔG as a function of Γ/K. The energy stored *in vivo* in this reaction is about -50 kJ.mol^{-1}. Note that this does not refer to the y-axis scale, or the energy change that would occur if the system were able to move from the *in vivo* position to that of $K = \Gamma$. Instead, the relevant ΔG is the slope of the curve, which is about -50 kJ.mol^{-1}. This is the amount of work made possible when 1 mol of ATP is hydrolyzed without affecting the overall concentrations of ATP, ADP, and P_i. Equally, the ΔG for synthesis of 1 mol of ATP under these conditions is $+50$ kJ.mol^{-1}; the energy change is the same as for hydrolysis, but the sign is opposite. *In vivo*, other reactions are constantly maintaining a ΔG of about -50 kJ.mol^{-1} for ATP hydrolysis by synthesizing ATP through reactions that require an energy input of about 50 kJ.mol^{-1}.

A corollary of this continual synthesis and breakdown of ATP is that ATP molecules within the cell (5 mM; approximately 100 g in the average male human) are turned over very rapidly. Indeed, it is estimated that a resting human will hydrolyze the equivalent of his or her own body weight of ATP every day. Furthermore, the turnover of ATP is increased greatly during periods of intense activity: a marathon runner may hydrolyze up to 10 g of ATP per second. Were we to rely on our ATP stores for energy, they would run out within about 2 min. So, it is clear that ATP acts as a go-between rather than as an energy store, ferrying energy from a primary source to drive a wide range of otherwise unfavorable reactions.

2 Energy Sources that Displace the ATP Hydrolysis Reaction from Equilibrium

2.1 Carbohydrate, Fat, and Protein as Energy Sources

The energy stored in the displacement from equilibrium of the ATP $\rightleftharpoons$ ADP + P_i reaction can act only as a transient energy intermediary. Other energy sources must be used to maintain the displacement from equilibrium of this reaction as ATP is used to do work. For mammals, this energy comes from food, specifically carbohydrate, fat, and protein. In the short term, carbohydrate and fat from our last meal circulates in the bloodstream for a few hours and can be used to provide energy; however, as we do not eat continually we have to store energy within our bodies. Fat in the form of triacylglycerides in adipose tissue is the most energy-dense store ($\sim$39 kJ.g^{-1}) and is consequently used for long-term energy storage. Carbohydrate ($\sim$17 kJ.g^{-1}) in the form of liver glycogen can be broken down to maintain glucose in the blood for 12 to 24 h; in the longer term, blood glucose levels are maintained by gluconeogenesis. Dietary protein is broken down to amino acids and those not required for protein synthesis are metabolized by pathways similar to those used for carbohydrates. During extended starvation, amino acids can be released from the body's protein

to provide energy. All of these energy stores are broken down by sequential catabolic reactions that use the stored chemical potential energy to displace the ATP hydrolysis reaction from equilibrium. How these occur is considered next (Fig. 3).

2.2 Glycolysis

When glucose enters the cell, it is converted to two molecules of pyruvate by a sequence of linked enzymatic reactions termed *glycolysis*. Two reactions in this sequence, catalyzed by phosphoglycerate kinase and pyruvate kinase, are coupled to the synthesis of ATP from ADP and P_i and so contribute to the displacement from equilibrium of the ATP hydrolysis reaction (although only a fraction of the chemical potential energy stored in glucose is released at this stage). During glycolysis, electrons are removed from the carbohydrate intermediates and are transferred to the electron carrier NAD^+, reducing it to NADH (Fig. 4). However, to maintain a flux through the glycolytic pathway, it is necessary to reoxidize the NADH continually to NAD^+. In mammals, this can occur in the absence of oxygen through the reduction of pyruvate by NADH; this reaction produces lactic acid and is catalyzed by lactate dehydrogenase. In the presence of oxygen, it is possible to extract far more ATP from glucose, but, for this, both the pyruvate and the electrons from NADH must be transported into mitochondria.

2.3 Mitochondria, the TCA Cycle and β Oxidation

The mitochondrion is the organelle at the heart of energy metabolism and ATP synthesis in eukaryotic cells. Most eukaryotic cells contain hundreds of mitochondria, which are filamentous organelles comprising a matrix surrounded by an invaginated inner membrane and a semipermeable outer membrane. Mitochondria arose from free-living bacteria that set up home within the cytoplasm of the protoeukaryotic cell, a process called *endosymbiosis*. These organelles break down fats and carbohydrate-derived metabolites using oxygen to maximize the amount of energy released and then use this energy to synthesize ATP.

To extract energy from carbohydrates, the pyruvate produced by glycolysis is taken up into the mitochondrial matrix. There, the three-carbon pyruvate is converted to a two-carbon acetyl group, which is carried by Coenzyme A as acetyl CoA. Acetyl CoA then reacts with the four-carbon oxaloacetate; together they generate the six-carbon citrate. In this way, carbohydrates enter the tricarboxylic acid (TCA) cycle (Fig. 3). In the course of the TCA cycle, oxaloacetate is reformed from citrate through a series of reactions that split off CO_2 and oxidize cycle intermediates, with the concomitant reduction of the electron carriers NAD^+ and FAD to NADH and $FADH_2$, respectively (Fig. 4). A single GTP molecule is also produced by each "rotation" of the TCA cycle; GTP is readily converted to ATP through the action of nucleoside diphosphate kinase. The oxaloacetate is then available to accept a further acetyl group from acetyl CoA and thus keeps the TCA cycle turning. When amino acids derived from the breakdown of proteins are used as energy sources, they are broken down to derivatives that are also fed into the TCA cycle.

Mitochondria are also the site of oxidation of the fatty acids derived from triacylglyceride, the storage form of fat. Within the cytoplasm, fatty acids are first

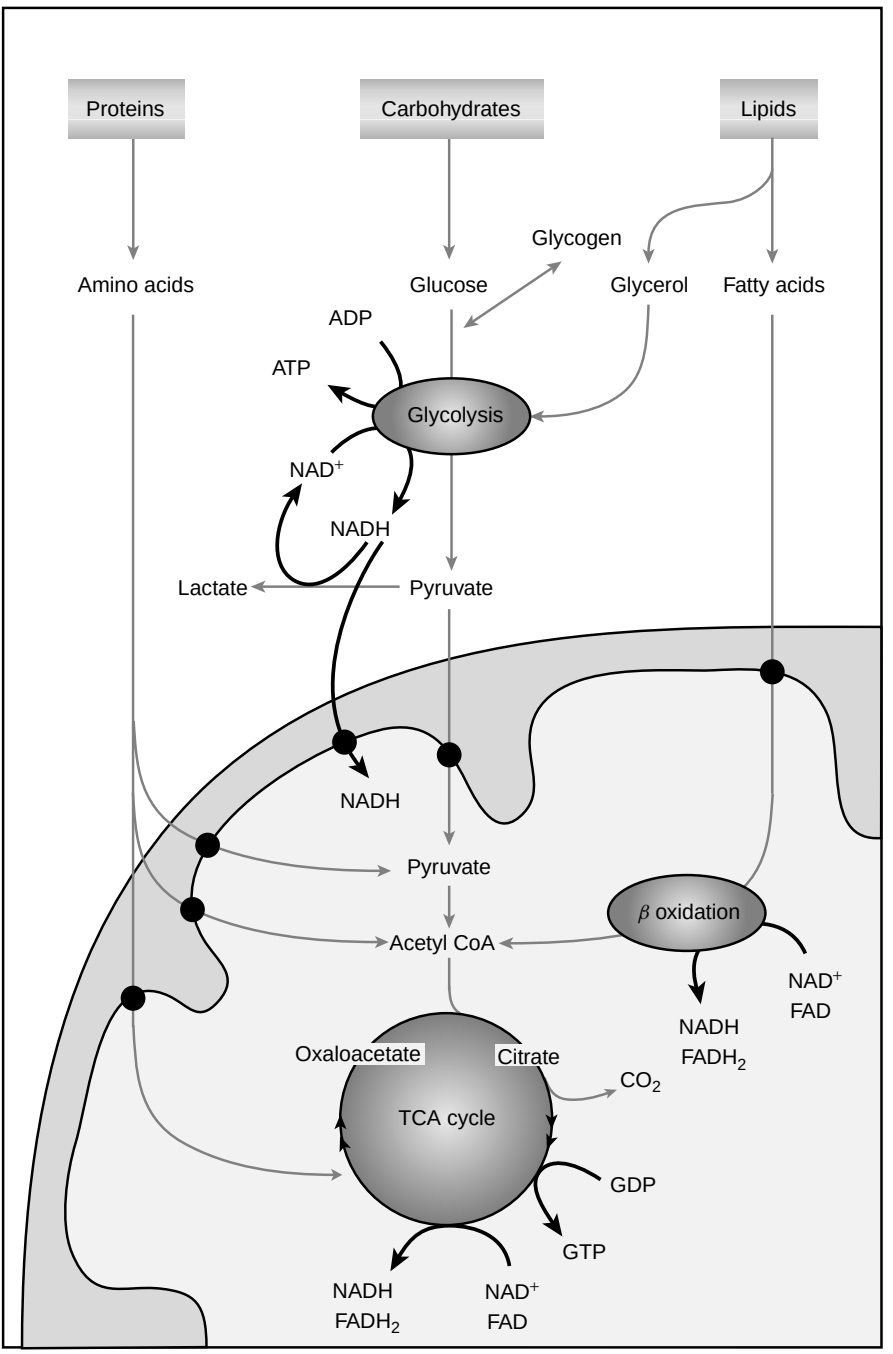
Proteins
Carbohydrates
Lipids
Glycogen
Amino acids
Glucose
Glycerol
Fatty acids
ADP
ATP
Glycolysis
NAD+
NADH
Lactate
Pyruvate
NADH
Pyruvate
β oxidation
Acetyl CoA
NAD+
FAD
NADH
FADH2
Oxaloacetate
Citrate
CO2
TCA cycle
GDP
GTP
NADH
FADH2
NAD+
FAD

linked to CoA to form acyl CoA and then converted to acyl carnitine. This is taken up into mitochondria through the inner membrane in exchange for carnitine. The acyl carnitine then reacts with CoA to form an acyl CoA within mitochondria. Typically, fatty acids are 16 to 18 carbons in length and within mitochondria they are broken down and oxidized by a process called *β oxidation*. During β oxidation, the acyl CoA goes through a continuous cycle of degradation and oxidation. In each cycle, two carbon units are removed from the acyl CoA to form acetyl CoA, and electrons are removed from the shortening fatty acid to reduce NAD^+ and FAD to NADH and $FADH_2$. This cycling persists until the fatty acid has been completely converted to acetyl CoA, which is then fed into the TCA cycle.

The endpoint of the metabolism of carbohydrates, fats, and protein within mitochondria is the formation of CO_2, NADH, and $FADH_2$. The CO_2 will diffuse out of the cells and be expired through the lungs; NADH and $FADH_2$ will be reoxidized. While the major function of the breakdown of fats and carbohydrates by mitochondria is to provide chemical energy to displace the ATP hydrolysis reaction from equilibrium, at this point only a fraction of the potential energy available has been used to make ATP (through glycolysis and by the formation of GTP). It is the reoxidation of NADH and $FADH_2$ that supplies the majority of the chemical potential energy for ATP synthesis. The means of coupling these two processes is discussed in the following sections.

3 Oxidation of NADH and $FADH_2$ by Mitochondria

Through the action of the TCA cycle and β oxidation within the mitochondrial matrix, there is an accumulation of electrons in the reduced carriers, NADH and $FADH_2$. The NADH pool within the matrix is soluble, and, typically, the ratio of NAD^+/NADH within mitochondria is about 5 to 10. The NADH built up in the cytoplasm by glycolysis can also be transported indirectly to the mitochondrial matrix through the malate–aspartate shuttle, or can donate electrons to the α-glycerophosphate

Fig. 3 Metabolic sources of the reducing equivalents NADH and $FADH_2$, and ATP. A small amount of NAD^+ is reduced to NADH with each round of glycolysis. This NADH can either be reoxidized to enable glycolysis to continue, or imported indirectly into mitochondria where it donates electrons to the mitochondrial respiratory chain. Pyruvate, the product of glycolysis, is transported to mitochondria where it is converted to acetyl CoA and enters the TCA cycle. The sequential oxidation of citrate enables the reduction of NAD^+ and FAD to NADH and $FADH_2$. Each "rotation" of the TCA cycle also produces a single GTP molecule (energetically equivalent to ATP). Triacylglyceride lipids are broken down into glycerol and fatty acids. Glycerol feeds into glycolysis, while fatty acids are transported to mitochondria and undergo multiple cycles of β oxidation, with each cycle yielding acetyl CoA (which feeds into the TCA cycle) and reducing NAD^+ and FAD to NADH and $FADH_2$. Proteins are metabolized through breakdown into amino acid monomers and the subsequent conversion of amino acids to pyruvate, acetyl CoA or TCA cycle intermediates. Black circles indicate either direct transport into mitochondria (e.g. the pyruvate transporter), or net transport through several molecular interconversions (e.g. the malate–aspartate shuttle for the "transport" of NADH). In summary, the catabolic reactions of metabolism act to reduce the electron carriers NAD^+ and FAD to NADH and $FADH_2$, and produce a small amount of ATP.

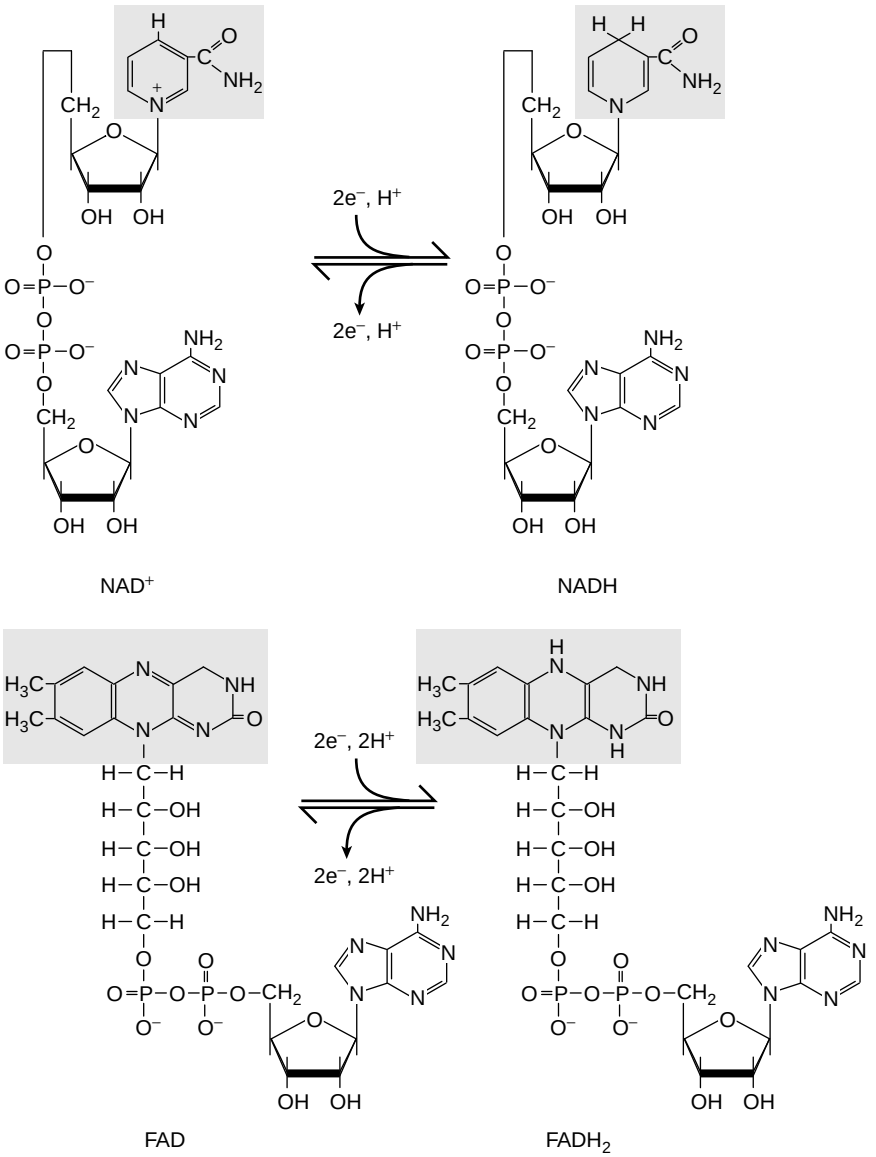

Fig. 4 Reduction of the electron carriers NAD^+ and FAD to their reduced forms NADH and $FADH_2$. Both carriers accept two electrons in conversion to the reduced form.

dehydrogenase on the outer surface of the mitochondrial inner membrane (Fig. 5). In contrast, the $FADH_2$ groups are not free in solution but are bound within the active sites of a number of different proteins, such as succinate dehydrogenase (for electrons derived from the TCA cycle), or on the electron transfer flavoprotein (ETF) (for electrons derived from β oxidation).

The reduction potential (E_h) of the NAD^+/NADH couple can be derived from the midpoint potential of the couple (E_m, ~ -320 mV) as follows:

$$E_h = E_m + \frac{RT}{2F} \ln\left(\frac{[NAD^+]}{[NADH]}\right)$$

$$E_h \sim -290 \text{ mV}$$

As the energy stored in this redox couple will be released by the reduction of oxygen, it should be compared with the E_h value for the O_2/water couple, which is about +790 mV for air-saturated water. This gives a ΔE_h of 1080 mV for the oxidation of NADH by O_2, corresponding to a ΔG of -208 kJ.mol^{-1}. In comparison, $FADH_2$ passes electrons directly to the ubiquinone/ubiquinol couple, which has a somewhat higher E_h ($\sim +10$ mV) than NADH, and thus a lower ΔE_h ($\sim$780 mV) and less negative ΔG (~ -150 kJ.mol^{-1}) for electron transfer to O_2.

The redox energy stored by NADH and $FADH_2$ is ultimately released on reaction with oxygen. However, were this to be done by direct reaction with O_2, the energy would be dissipated uncontrollably as heat and could not be harnessed for ATP synthesis. Instead, the release of the redox potential energy stored in NADH and $FADH_2$ is controlled by the action of the mitochondrial respiratory chain.

3.1 The Mitochondrial Respiratory Chain

The respiratory chain is a sequence of electron carriers embedded within the mitochondrial inner membrane that catalyzes electron movement from NADH and $FADH_2$ to oxygen (Fig. 5). The first component of the respiratory chain is complex I, a massive membrane protein complex of about 980 kDa. This protein takes electrons from NADH and passes them to a flavin mononucleotide within complex I. From there, the electrons pass through a series of iron–sulfur centers and are then used to reduce ubiquinone to ubiquinol; these are the two redox forms of Coenzyme Q (CoQ), a small, mobile electron carrier within the mitochondrial inner membrane, which is not a protein but a hydrophobic organic molecule.

A number of other respiratory complexes also pass electrons to CoQ. Complex II, or succinate dehydrogenase, catalyzes the oxidation of succinate to fumarate in the TCA cycle. The electrons initially reduce a bound FAD within complex II; the resultant $FADH_2$ passes the electrons through a series of iron–sulfur centers to the CoQ binding site within the membrane where they reduce ubiquinone to ubiquinol. α-Glycerophosphate dehydrogenase (GPDH) is an enzyme on the outer surface of the mitochondrial inner membrane that is expressed in some tissues. This enzyme oxidizes α-glycerophosphate to dihydroxyacetone phosphate and uses the electrons to reduce a bound FAD within the enzyme, which then reduces ubiquinone to ubiquinol. The oxidation of fatty acids by β oxidation leads to the reduction of an FAD in a small mobile protein within the mitochondrial matrix called *electron transfer flavoprotein* (ETF). This electron carrier is reoxidized by the respiratory chain through an inner membrane protein called *ETF–ubiquinone oxidoreductase* (ETF-OR). This enzyme reduces an internal FAD and subsequently reduces ubiquinone to ubiquinol, possibly *via* an iron–sulfur center. Therefore, these, and other, enzymes catalyze electron flow into the CoQ pool. *In vivo*, the CoQ pool is generally about 70 to 80% in the reduced (ubiquinol) form.

The fate of ubiquinol is oxidation by complex III (also known as the cytochrome bc_1 complex), the next component of the respiratory chain. Complex III takes two electrons from the ubiquinol of the CoQ pool and passes them through a branching pathway where one electron goes through the two b-type cytochromes before being passed back to reduce ubiquinone. The other electron reduces an iron–sulfur

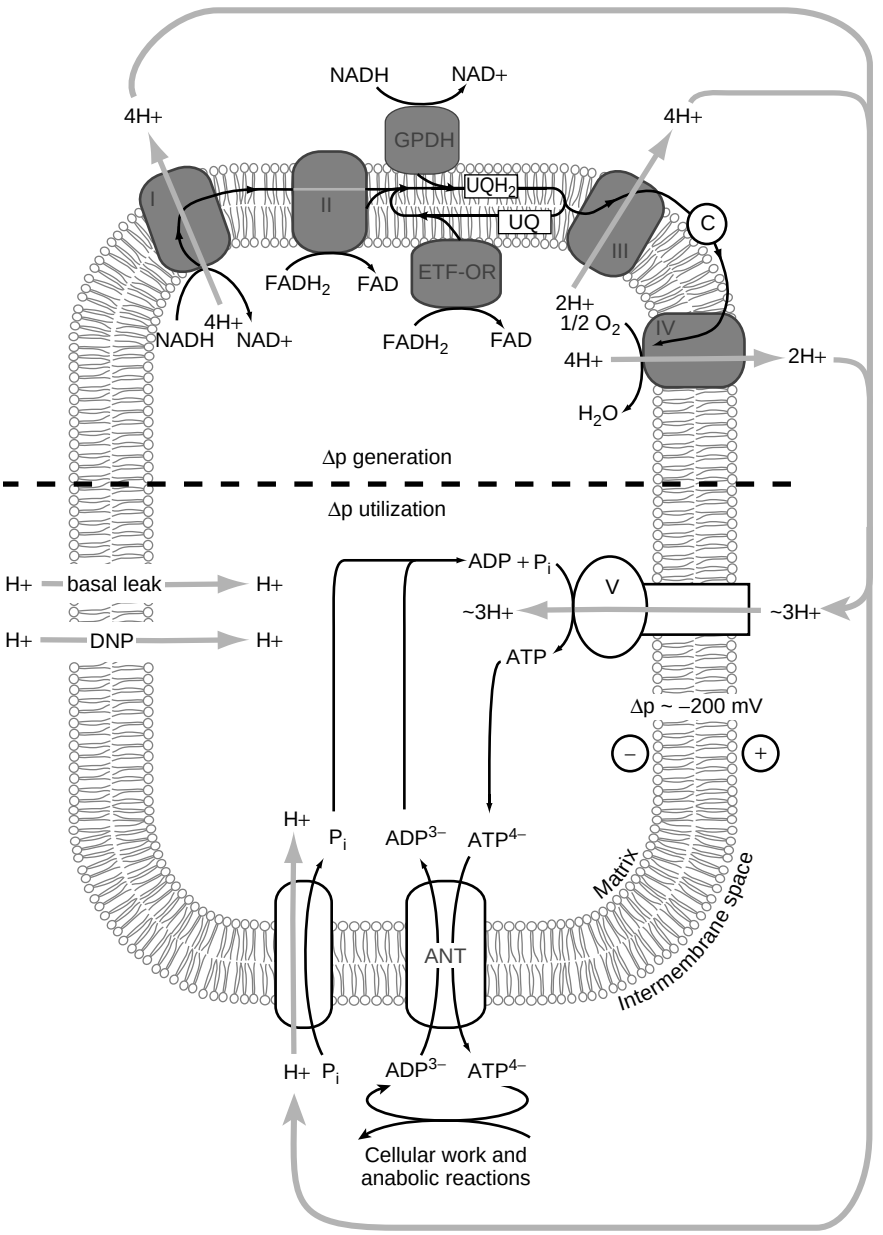

center, called the *Rieske iron-sulfur center*, and from there reduces a cytochrome c_1. This cytochrome is the last electron carrier in the complex and passes its electron to cytochrome c, a small protein that acts as a mobile electron carrier on the cytoplasmic surface of the mitochondrial inner membrane. The enzymes sulfite oxidase and NADH-cytochrome b_5 reductase (via cytochrome b_5) can also pass electrons to cytochrome c, but the flux through these reactions *in vivo* is thought to be small.

Cytochrome c carries electrons to complex IV (or cytochrome oxidase), the final stage of their journey toward oxygen. It is at cytochrome oxidase that about 95% of the oxygen we breathe is reduced to water. The electron from cytochrome c is initially passed to the heme *a* site within cytochrome oxidase and then to the Cu_a site. From there, the electrons pass to the heme a_3 site, which also contains the Cu_b center. Here the terminal electron acceptor of the respiratory chain, O_2, binds and is reduced to water.

So far, we have outlined the pathway taken by electrons as they move through the respiratory chain. In the next section, the mechanism of coupling this movement to ATP synthesis is explained.

3.2 Energy Storage by the Protonmotive Force (Δp)

The respiratory chain provides a complex pathway for the gradual movement of electrons from the low potential of NADH (~ -290 mV) to the high potential of oxygen ($\sim +790$ mV). This pathway enables the gradual release of potential energy and is best illustrated by Fig. 6, which shows the gradual increase in reduction potential (E_h) as electrons pass through the respiratory chain. Three significant changes in E_h are apparent within the respiratory chain, at complexes I, III, and IV; the large increases in E_h – which signify large drops in the potential energy of the electrons – suggest that at these three points the redox energy may be conserved as an intermediate that can be used to make ATP.

The mechanism of coupling of electron transport and ATP synthesis was one of the major challenges to biochemistry

Fig. 5 NADH and $FADH_2$ are oxidized by the mitochondrial respiratory chain, giving rise to the protonmotive force and ATP synthesis. NADH and $FADH_2$ donate electrons to the respiratory chain through a number of inner membrane proteins and multiprotein complexes: complex I (mitochondrial NADH), α-glycerophosphate dehydrogenase (GPDH; cytosolic NADH), complex II (TCA cycle-derived $FADH_2$) and electron transfer flavoprotein–ubiquinone oxidoreductase (ETF-OR; β oxidation-derived $FADH_2$). Each of these proximal electron acceptors feeds electrons to CoQ, reducing ubiquinone (UQ) to ubiquinol (UQH_2). Electrons are subsequently passed through complex III, cytochrome c, and complex IV, finally reducing O_2 to H_2O. Complexes I, III, and IV all derive sufficient energy from the drop in redox potential of electrons to catalyze the pumping of protons out of the matrix and into the cytoplasm. This generates a protonmotive force (Δp) across the mitochondrial inner membrane. Driven by Δp, protons flow back into the matrix through the ATP synthase (complex V), catalyzing the formation of ATP from ADP and P_i. Some protons can also return to the matrix by mechanisms that do not involve passage through the ATP synthase, such as the basal leak through the inner membrane in all mitochondria, or via uncouplers such as 2,4-dinitrophenol (DNP) that render the inner membrane permeable to protons. Newly synthesized ATP^{4-} is transported out of the mitochondria and into the cytosol by exchange with ADP^{3-}, which in turn is converted to ATP. The P_i required for ATP synthesis is transported into the matrix in symport with a proton.

in the 1970s and early 1980s. The two processes are now known to be linked by "chemiosmotic coupling," a concept developed by Peter Mitchell in the early 1960s and established over the subsequent two decades. The energy available as the electrons pass through complexes I, III, and IV is used to pump protons out of the matrix across the mitochondrial inner membrane. For every pair of electrons that passes through the respiratory chain, 4 H^+ are translocated at complex I, 4 H^+

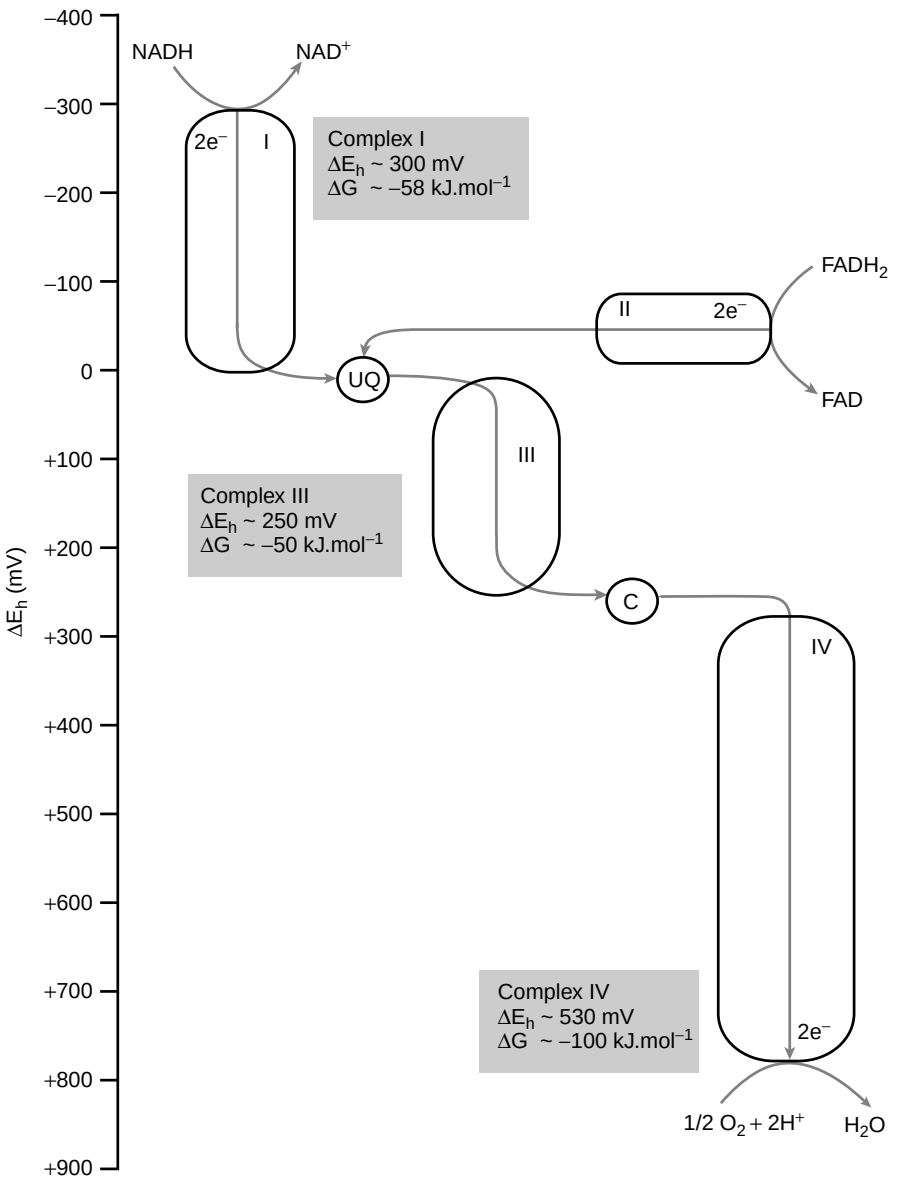

at complex III, and 2 H^+ at complex IV (although the movement of electrons with respect to the membrane at complexes III and IV means that the charges translocated by complexes I, III, and IV are 4, 2, and 4, respectively). The reduction of O_2 to water requires 4 e^- , so a pair of electrons corresponds to the reduction of a single O atom. As the two electrons go through the respiratory chain from NADH, 10 protons and charges are pumped across the inner membrane, while for electrons coming from $FADH_2$, six protons and charges are translocated.

Proton pumping builds up a charge gradient, or membrane potential, across the inner membrane ($\Delta\psi$ in mV, negative inside) (Fig. 5). In addition, proton translocation also leads to a proton concentration gradient across the inner membrane, which is usually expressed as a pH gradient (ΔpH, basic inside). Both the membrane potential and the pH gradient raise the potential energy of protons outside the mitochondrial inner membrane relative to those inside. The contribution of the two components to this potential energy difference is conveniently expressed as the protonmotive force. For this, the pH gradient is multiplied by $-2.303RT/F$ to express its value in mV; this can be added to the membrane potential to give the protonmotive force (Δp):

$$\Delta p = \Delta\psi - (2.303RT/F)\Delta pH$$

The advantage of this formulation is that the relative contributions of $\Delta\psi$ and ΔpH are clear. In isolated mitochondria, the maximum $\Delta\psi$ is about 170 to 190 mV and the pH gradient can be up to -0.5 to -1 pH unit (basic inside, hence the negative sign in the equation). At 37 °C, $2.303RT/F = 61$ mV, so the maximum Δp is typically around 180 to 200 mV. In cells, the average values tend to be lower because of the utilization of Δp to synthesize ATP, as well as limiting electron and O_2 supply.

To maintain the Δp across the mitochondrial inner membrane, the membrane has to be topologically intact and the phospholipid bilayer must be relatively impermeant to protons. A further implication of the closed membrane is that mechanisms must exist for the transport of polar metabolites across the membrane, such as pyruvate, ATP, ADP, and P_i, among many others. As this movement has to occur without dissipating the Δp, the metabolite carriers in the mitochondrial inner membrane must be able to catalyze transport without rendering the membrane permeable to protons. In summary, the redox energy stored by NADH

Fig. 6 Sequential increases in redox potential (E_h) of respiratory chain components are coupled to proton pumping. NADH, a major electron donor to the mitochondrial respiratory chain, has a low E_h (~ -290 mV) under conditions found in the matrix. After oxidation of NADH by complex I, electrons are passed through a series of electron carriers in the complex (not shown) before reducing the mobile electron carrier ubiquinone ($E_h \sim +10$ mV). The large increase in E_h of respiratory components between NADH and ubiquinone – and hence the drop in redox potential energy of the transported electrons – provides sufficient energy for translocation of $4H^+$ (per $2e^-$) by complex I into the cytosol. Similarly, increases in E_h between ubiquinol and oxidized cytochrome c ($E_h \sim +260$ mV), and between reduced cytochrome c and O_2 ($E_h \sim +790$ mV), provide sufficient energy for proton translocation at complexes III ($4H^+$) and IV ($2H^+$) , respectively. ΔE_h values shown for complexes I, III, and IV have been calculated from the increases in redox potential between NADH and ubiquinone, ubiquinol and oxidized cytochrome c, and reduced cytochrome c and O_2, respectively.

and $FADH_2$ is conserved as a protonmotive force that comprises a $\Delta\psi$ and a ΔpH across the mitochondrial inner membrane. This protonmotive force can now be used for the synthesis and transport of ATP.

3.3 The Protonmotive Force Drives ATP Synthesis and Transport

The energy stored in Δp is used to drive the synthesis of ATP through the action of the ATP synthase. This protein complex is also found in the mitochondrial inner membrane and catalyzes the movement of protons down their proton electrochemical gradient from the cytoplasm, back into the mitochondrial matrix. The drop in potential energy as the protons pass through the complex provides the energy for ATP synthesis by the synthase. This remarkable enzyme has a hydrophobic domain (the F_O domain) embedded in the lipid bilayer, which facilitates proton movement through the lipid bilayer, and a connected F_1 component in the mitochondrial matrix where the ATP is made (Fig. 7). The F_O subcomplex comprises a ring of 10 to 14 identical hydrophobic proteins called *subunit c*. This ring is attached to a stator arm that links together the F_O and F_1 subcomplexes. As a proton passes through the membrane, it forces the rotation of F_O relative to the stator arm, with the concomitant rotation of a second protein subunit, γ. The γ subunit is attached at its base to F_O; the remainder of the protein is inserted into the center of the F_1 complex, which comprises 3 α and 3 β subunits arranged alternately like the segments of an orange. This assembly of 3 α and 3 β subunits is held in place by the stator arm, so proton movement drives the rotation of F_O and the γ subunit relative to F_1. The rotation of the γ subunit against the α and β subunits forces the three catalytic sites at the interfaces of the α and β subunits to change shape, with each site switching between three conformations that alternately bind ADP and P_i, synthesize ATP, and finally release ATP. The Δp is used to drive the rotation of the γ subunit relative to the α and β subunits at about 50 to 100 Hz *in vivo*; thus, Δp is converted to mechanical energy and then back to chemical energy in the displacement from equilibrium of the ATP hydrolysis reaction. Each complete rotation of the γ subunit within the ATP synthase enables the synthesis of three ATP molecules. This is associated with the complete rotation of F_O. It is likely that the number of protons translocated for each complete F_O rotation equals the number of c subunits; since there are thought to be at least 10 subunit c components in mammals, this suggests that 3.3 or more H^+ are transported per ATP synthesized. However, direct measurement of the number of H^+ translocated per ATP synthesized puts the ratio at about 3.

The ATP synthase makes ATP in the mitochondrial matrix, but, as most ATP is consumed in the cytoplasm, ATP must be exported from the mitochondrion. In addition, P_i and ADP must be imported into mitochondria from the cytoplasm. Much of the anion transport across the mitochondrial inner membrane is accomplished by a family of carriers of about 30 to 35 kDa. Among these are the adenine nucleotide translocator (ANT) and the phosphate carrier (Fig. 5). The ANT catalyzes exchange of ATP^{4-} in the mitochondrial matrix with ADP^{3-} in the cytoplasm, hence transport involves the net movement of a positive charge into the mitochondrion, and ATP export and ADP

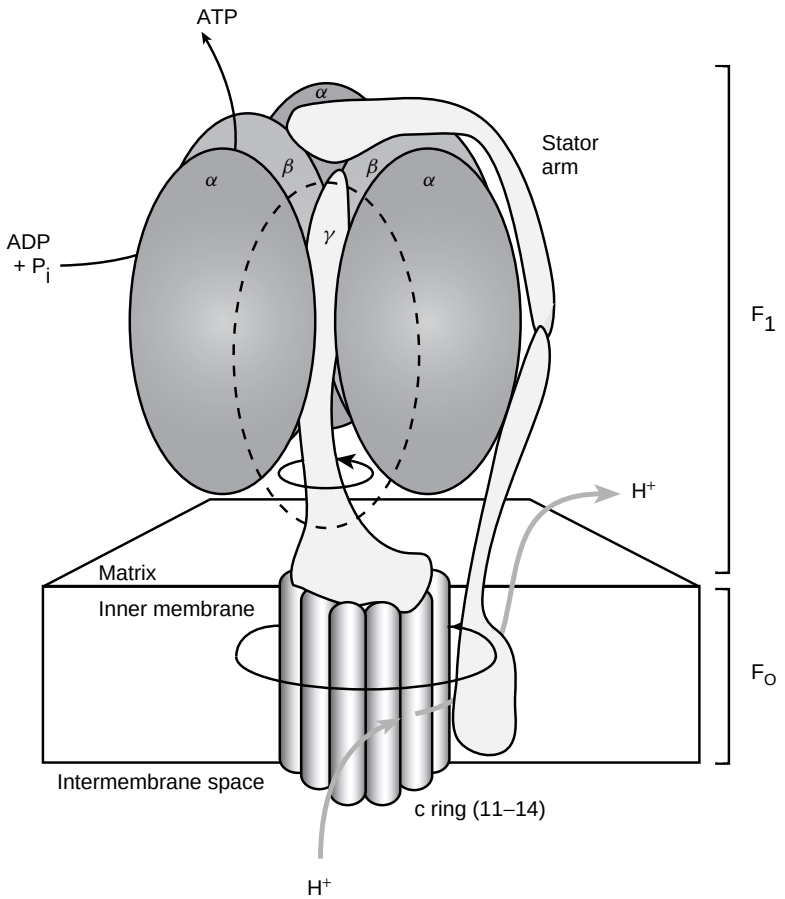

Fig. 7 The mitochondrial ATP synthase. To synthesize ATP, the ATP synthase harnesses the energy stored in the Δp across the mitochondrial inner membrane. Protons move down their electrochemical gradient from the intermembrane space via the ring of c subunits (F_O subcomplex) embedded in the membrane, catalyzing the rotation of both the F_O subcomplex and the neighboring γ subunit (part of the F_1 subcomplex). Rotation of the γ subunit relative to the static $\alpha_3\beta_3$ subunits (held in place by the stator arm) alters the shape of the three catalytic sites at interfaces between α and β subunits, promoting sequential ADP and P_i binding, ATP synthesis, and finally the release of ATP from the enzyme.

import is driven by the membrane potential component of Δp. A consequence of this is a higher ATP/ADP ratio in the cytoplasm than in the mitochondrial matrix. So, the displacement from equilibrium of the ATP hydrolysis reaction is driven by Δp acting through both the ATP synthase and the ANT. The phosphate carrier takes up the P_i required for ATP synthesis in electroneutral symport with a proton; hence, P_i uptake is driven by the ΔpH component of Δp. In sum, the import of ADP and P_i and the export of ATP is associated with the movement

of one proton through the mitochondrial inner membrane. So, if it takes approximately three protons to synthesize an ATP molecule in the matrix, it takes about four protons to synthesize ATP in the cytoplasm.

A final consideration is the number of oxygen atoms that must be consumed to synthesize an ATP molecule and export it to the cytoplasm. As the oxidation of NADH or $FADH_2$ pumps 10 and 6 protons respectively, the theoretical P/O ratios are 10/4 for NADH and 6/4 for $FADH_2$. The true values will be less because of inefficiencies such as leakage of protons through the mitochondrial inner membrane.

3.4 Respiratory Control and Uncoupling

As mitochondria make ATP, they consume Δp, enabling them to sense and respond to the rate of ATP consumption. ATP hydrolysis increases the ADP concentration, thereby elevating the rate of ATP synthesis. Greater ATP synthase activity involves the consumption of Δp, which is then matched by a rise in respiration rate, as the respiratory chain responds to the decreased Δp by increasing electron flow and proton pumping to maintain Δp. This phenomenon is known as *respiratory control* and enables mitochondria to match ATP synthesis to ATP consumption, maintaining the displacement from equilibrium of the ATP hydrolysis reaction under conditions of variable ATP demands.

The link between respiration rate and Δp is also indicated by the phenomenon of uncoupling. To maintain Δp, the mitochondrial inner membrane is relatively impermeable to protons. Uncouplers counteract this by making the inner membrane permeable to protons. Uncouplers are typically lipophilic weak acids in which the conjugate base is also lipophilic. An example is 2,4-dinitrophenol (DNP; Fig. 5) where the charge on the phenolate anion is delocalized and, consequently, both the anion and its conjugate acid are lipid soluble. This enables DNP to partition into the phospholipid bilayer of the mitochondrial inner membrane, where it can shuttle across the membrane, picking up a proton on the intermembrane space side and depositing it in the matrix. Thus, uncouplers catalyze proton movement through the lipid bilayer; in the presence of an uncoupler, the mitochondrial inner membrane is effectively permeant to protons and consequently cannot maintain a Δp. In response, the mitochondria respire rapidly to pump out protons in a vain attempt to maintain a Δp, which is instead rapidly dissipated as heat (Fig. 8).

3.5 Alternative Uses of the Protonmotive Force

While the principal function of Δp is the synthesis and export of ATP from the mitochondrion, there are other processes that exploit Δp to carry out work. These include metabolite transporters that catalyze movement across the mitochondrial inner membrane by coupling to charge or proton translocation (analogous to ADP/ATP antiport and P_i transport). The movement of calcium across the mitochondrial inner membrane is also driven by Δp. There is a calcium uniporter in the mitochondrial inner membrane that enables calcium uptake into mitochondria. As only the calcium dication is translocated, uptake is driven solely by the membrane potential. In addition, there is a separate

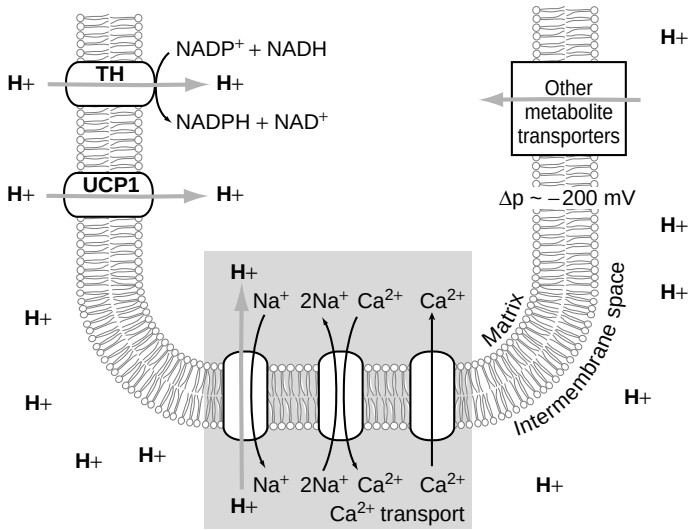

Fig. 8 Alternative fates of Δp. While the major function of the Δp is to generate ATP, proton movement down the electrochemical potential gradient across the mitochondrial inner membrane can be coupled to other activities. Such Δp-dependent processes include the action of the transhydrogenase (TH), which exploits Δp to drive reduction of $NADP^+$ to NADPH, at the expense of NADH; thermogenesis in brown adipose tissue by uncoupling protein 1 (UCP1), which dissipates the Δp as heat by increasing the proton permeance of the inner membrane; the activity of an Na^+-H^+ antiporter and Ca^{2+} uniporter, which (in concert with a $2Na^+/Ca^{2+}$ antiporter) exploit the ΔpH and $\Delta\psi$ components of the Δp, respectively, to catalyze mitochondrial uptake of calcium ions; and the transport of a number of other metabolites.

exchange activity involving electroneutral exchange of Ca^{2+} for $2H^+$ or $2Na^+$. In conjunction with the Ca^{2+} uniporter, this exchange can lead to the continual cycling of calcium across the mitochondrial inner membrane as cytosolic calcium increases. The overall effect is that increased cytoplasmic calcium concentration leads to a proportionate increase in calcium concentration in the mitochondrial matrix. As changes in cytosolic calcium are important signals that activate various aspects of metabolism, this mechanism enables such signals to be relayed to the mitochondrial matrix. The changes in matrix calcium activate dehydrogenases, thereby increasing electron supply to the respiratory chain and elevating ATP synthesis. The reason for this is thought to be that calcium signals are often associated with increased ATP turnover; Δp-dependent Ca^{2+} uptake enables mitochondria to upregulate ATP synthesis to match this increased demand.

A further use of Δp is in thermogenesis by brown adipose tissue. Under certain conditions, mitochondria can become uncoupled; that is, the energy stored in Δp can be dissipated as heat by providing a pathway for protons to flow back through the inner membrane to the matrix without carrying out work. Certain

molecules, such as DNP, can act as uncouplers and dissipate Δp as heat (Fig. 5). Indeed, historically, DNP was used as a slimming agent as it led to the rapid breakdown of fat stores. Physiologically regulated uncoupling is used by mitochondria in brown adipose tissue mitochondria to generate heat. Mitochondria in brown adipose tissue contain uncoupling protein 1 (UCP1), which catalyzes proton movement through the mitochondrial inner membrane. When this protein is activated, it partially uncouples mitochondria, increasing the rate of respiration such that the fat stored in brown adipose tissue is broken down and oxidized by the mitochondria. However, instead of the energy in Δp being used to make ATP, it is instead dissipated as heat. This activity enables brown adipose tissue to be a site of nonshivering thermogenesis in neonates and in small mammals exposed to the cold.

A final example of the use of Δp is the transhydrogenase (TH) enzyme in the mitochondrial inner membrane. Within the mitochondrial matrix, the $NADH/NAD^+$ ratio is about 0.1 to 0.2, while that of the $NADPH/NADP^+$ couple is approximately 99% reduced. This is thought to be because NADPH is required to maintain the mitochondrial glutathione pool and thioredoxin in a reduced state, through NADPH-dependent glutathione reductase and thioredoxin reductase. A reduced glutathione pool and reduced thioredoxin are important in protecting mitochondria from oxidative damage. One of the ways the high $NADPH/NADP^+$ ratio is maintained is through the transhydrogenase, which couples proton movement down the Δp gradient to drive electrons from NADH to $NADP^+$ and thus maintains the disparity in redox states.

4 Overview of the Metabolic Basis of Cellular Energetics

A central feature of cellular bioenergetics is the displacement from equilibrium of the ATP hydrolysis reaction. This displacement is at the heart of metabolism where it acts as a coupling reaction between catabolism and anabolic or work reactions. While there is continual turnover of ATP, the energy stored in this displacement from equilibrium is usually held at about $-50\ kJ.mol^{-1}$ through the expenditure of energy from fat, carbohydrate, or protein. As these molecules are broken down, the chemical energy is converted to redox potential energy in the reduced electron carriers NADH and $FADH_2$. This redox potential energy is gradually released as electrons pass through the mitochondrial respiratory chain and is conserved at three regions in the respiratory chain by pumping protons across the mitochondrial inner membrane. This builds up a Δp that is then used to drive the synthesis of ATP in the mitochondrion and its subsequent transport to the cytosol. Thus, there is a linked series of energy conversions, from chemical energy to redox energy and protonmotive force, followed by a return to chemical energy, which is stored in the displacement from equilibrium of the ATP hydrolysis reaction. These energy interconversions allow ATP to drive the many otherwise unfavorable reactions that are necessary for life.

See also Adipocytes; Intracellular Fatty Acid Binding Proteins in Metabolic Regulation; Oncology, Molecular.

Bibliography

Books and Reviews

Brown, G.C. (2000) *The Energy of Life*, Flamingo, London.

Harold, F.M. (1986) *The Vital Force: A Study Of Bioenergetics*, W. H. Freeman, New York.

Nicholls, D.G., Ferguson, S.J. (2002) *Bioenergetics 3*, Academic Press, London.

Rich, P. (2003) Chemiosmotic coupling: the cost of living, *Nature* **421**, 583.

Scheffler, I.E. (1999) *Mitochondria*, Wiley-Liss, New York.

Primary Literature

Saraste, M. (1999) Oxidative phosphorylation at the *fin de siecle*, *Science* **283**, 1488–1493.

Wallace, D.C. (1999) Mitochondrial diseases in man and mouse, *Science* **283**, 1482–1488.

Metabonomics and Metabolomics

David J. Grainger[1] *and Jeremy K. Nicholson*[2]
[1] *Department of Medicine, University of Cambridge, Cambridge, UK*
[2] *Department of Biological Chemistry, Imperial College, London, UK*

1 What is Metabonomics? 189

2 Methods for Generating a Metabonomic Profile 192
2.1 Criteria for Judging Metabonomic Profiling Methods 192
2.2 Nuclear Magnetic Resonance (NMR) Spectroscopy 193
2.3 Chromatography and Mass Spectroscopy (LC-MS and GC-MS) 194
2.4 Nanosensors 195
2.5 Other Approaches 195

3 Methods for Interpreting Metabolic Profiles 196
3.1 The Problems of Interrogating Very Large Datasets 196
3.2 Conventional Statistical Approaches (LDA) 197
3.3 Projection Methods (PCA and PLS) 198
3.4 Genetic Computing 199
3.5 Other Approaches 199
3.6 Data Preprocessing and Data Filters 199

4 Applications of Metabonomics 200
4.1 Probing Normal Human Metabolism 200
4.2 Probing the Pathophysiology of Human Disease 201
4.3 Investigating and Validating Animal Models of Disease 201
4.4 Clinical Diagnosis of Disease 202
4.5 Selection of Subjects for Clinical Trials 202
4.6 Monitoring Efficacy of Therapeutic Interventions 202
4.7 Toxicology 203
4.8 Predicting Future Disease Risk 203

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

5 Future Prospects and Challenges 203

Bibliography 204
Books and Reviews 204
Primary Literature 205

Keywords

Chromatography
A range of methods designed to separate the molecular components of a complex mixture on the basis of their physical properties.

Mass Spectrometry
An analytical chemistry technique that can be used to identify molecules by fragmenting them into ions, which are then separated on the basis of mass.

Nuclear Magnetic Resonance
The context-dependent resonance of various atomic nuclei (such as the hydrogen nucleus) when placed in a magnetic field, which can be used to obtain structural information about the molecular components of a solution.

Regression Modeling
Mathematical methods designed to predict one (or more) variables from the values of many other measured variables.

Metabonomics is the study of systemic biochemical profiles and regulation of function in whole organisms by analyzing biofluids and tissues. Like genomics (the study of the complete repertoire of genes in an organism) and proteomics (the study of the protein complement of a tissue or cell), metabonomics can provide a holistic overview of the current physiological status of an organism, and its response to external stressors. Here we review the technological approaches to generating metabolic profiles, highlighting the advantages and disadvantages of each methodology, as well as the various strategies for extracting useful conclusions from the very large datasets that can be generated by such profiling. Metabonomics can be applied to a wide range of biological applications, including predictive toxicology, probing the physiology of disease both in animal models and in man, and to make clinically useful diagnoses of disease. With examples of each of these applications, we illustrate the potential of metabonomics to contribute to our understanding of complex biological systems in a post-genomic era where we understand many of the components of living systems, but few of their dynamic interactions.

1
What is Metabonomics?

The suffix "omics" is now routinely applied in many fields to the holistic study of an entire system, as opposed to a reductionist description of each of its parts independently. Thus, metabonomics is the name given to the holistic study of metabolic systems in living organisms. In principle, a metabolic profile is therefore a simple list of all the low molecular weight metabolites (such as sugars, amino acids, and lipids) present in a biological system, together with the concentration of each metabolite present (the generation of such profiles is one of the definitions in use for metabo*l*omics). Clearly, such a profile is analogous to a genomic profile (a list of all the genes composing an organism, perhaps with their levels of expression also) or a proteomic profile (a list of the proteins in an organism).

Like genomics and proteomics, however, metabo*n*omics is also much more than a simple list. The metabolic profile of a particular biological sample is just a snapshot of a complex, dynamic network that reflects the physiological activity of the organism. Enzymes are rapidly interconverting metabolites; new compounds are being absorbed from the environment; waste products are being excreted. The science of metabonomics, therefore, is not only about capturing metabolic profiles (described in Sect. 2 below) but also about extracting an understanding of the underlying biological system from the resulting dataset (described in Sect. 3). In particular, metabonomics is a global metabolic regulation approach based on understanding complex system behavior, designed to reveal the response of an organism to an external stressor or stimulus.

The relationship between metabonomics and other systems biology disciplines is illustrated in Fig. 1. Genetic information (coupled with the pre-existing levels of various other proteins) is the major determinant of the mRNA (or transcriptomic) profile, which in turn is a key determinant of the proteomic profile. Proteins (in the form of enzymes) are an important determinant of the metabolic profile, which, in turn, feeds back to modulate gene expression patterns. This "homeostatic loop" is then modulated by external inputs from the environment. The major input of low molecular weight compounds from the diet can have a significant effect on the metabonomic profile, which subsequently affects the gene expression and protein profiles of the organism. Other environmental factors can also affect the metabolic profile (for example, the amount of light determines the rate of vitamin D formation), as well as directly altering gene expression (the amount of exercise modulates skeletal muscle gene expression and ultimately protein content). This position at the interface between genetic and environmental determinants makes metabolic profiling a uniquely powerful tool for probing the dynamic physiological status of an organism.

Compared with genomic and proteomic profiles, the metabonomic profile is also more dynamic, reflecting the current physiological status of the organism as well as its future behavior. Polymorphisms, in particular genes (a component of the genomic profile), may allow the risk of developing a disease to be estimated, but they cannot determine whether the organism is suffering from that disease at a given point in time.

To properly understand the metabolic network of a multicellular organism would require continuous measurement of the

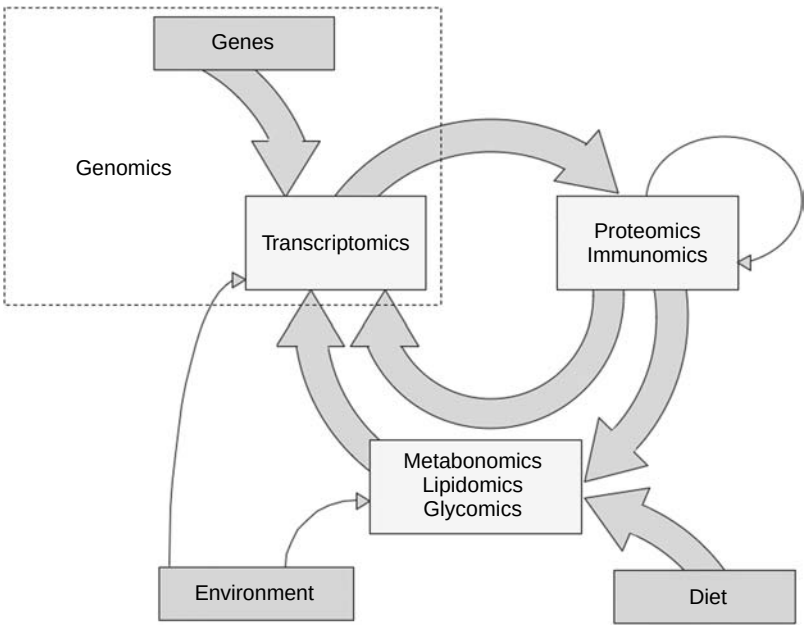

Fig. 1 Relationship between the ''omics.'' Much of the variation in protein levels (measured in proteomics) is due to variations in the expression of the mRNA encoding that protein (measured in transcriptomics, a subset of genomics). In turn, the variation in the levels of metabolites (measured in metabonomics) is determined by the levels of various enzymes (proteins). Metabolites can then feedback and regulate gene expression patterns, closing a ''homeostatic loop.'' The environment interacts with this homeostatic loop primarily through the influence of diet on the metabolic profile, although other environmental factors can also have a direct effect on both metabolism and gene expression (for example, exposure to UV light directly affects vitamin D3 levels, and exercise can affect gene expression patterns).

levels of all the metabolites present in all of the cells and tissues that compose the organism. Such an ''ideal'' metabolic profile is unlikely to be practicable attainable (at least in the near future). As a result, practical metabonomics involves selection of both a sample of the whole organism (for example, a blood specimen) and a time point (or series of time points) at which to make the observations. Clearly, the extent to which one can hope to understand the metabolism of the organism as a whole from a (potentially poorly representative) sample is unclear and care should be taken in drawing broad conclusions from limited measurements. Similarly, available analytical chemistry techniques (such as spectroscopy or chromatography) do not allow accurate measurements of the levels of every low molecular weight compound present in a given biological specimen. These practical limitations are illustrated in Fig. 2.

Even with the current practical limitations, metabonomics is a powerful new tool for studying complex biological systems. It has already been used successfully to monitor the physiological response to xenobiotics (such as new pharmaceuticals under development), and its use in

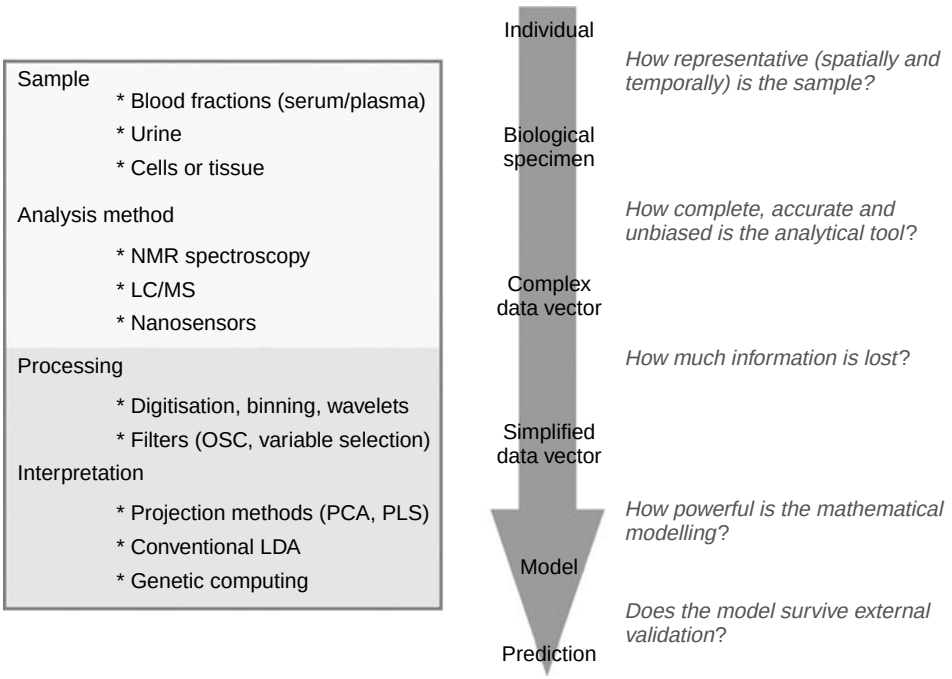

Fig. 2 A generic "omics" experiment. The aim of "omics" biology is to make measurements on individuals and then deduce predictive rules about the organism. In order to go from individual observations to a prediction, various steps must be followed: a sample must be taken on which many measurements are made to generate a complex data vector. Various mathematical modeling tools are then used to build a regression model which can be used to make a prediction that can be validated. Examples of each step in a typical metabonomics experiment are shown in the left panel. Various assumptions made at each step are listed in the right panel.

toxicology is increasing. Metabonomics can also be used to diagnose the presence of diseases, both for those where in-born errors of metabolism are responsible for the symptoms, and also in diseases where metabolic disregulation is less obviously involved in the pathogenesis of the disease. Using a metabolic profile to diagnose disease is not just useful in the clinic: it can also provide important new information about the physiological processes that are misregulated in the disease, which might ultimately assist in the search for new treatments.

The "homeostatic loop" illustrated in Fig. 1 also emphasizes the importance of integrating the genetic, protein, and metabolic profiles if we are to maximize our understanding of the organism. The boundaries between the "omics" represent technical limits in our methodologies for making measurements rather than being any useful dividing line between the applications of the information that has been generated. Fortunately, the profiles generated can be merged (at least in principle, although the bioinformatics challenges in doing so are significant) yielding a composite

(or "multi-omics") profile of the organism. Any question partially addressed by the separate genomic, proteomic, and metabonomic profiles will likely be more fully answered through the careful construction and analysis of such "multi-omics" profiles.

2 Methods for Generating a Metabonomic Profile

2.1 Criteria for Judging Metabonomic Profiling Methods

A range of different analytical chemistry approaches have been used to generate metabonomic profiles. Unfortunately, none of the profiles generated even approximate to the "ideal" profile, and it is necessary to make an informed choice of analytical tool depending on various trade offs. There are three important criteria that contribute to the "ideal" profile:

1. *Completeness*. If we assume that the "ideal" profile consists of a list of all the different low molecular weight compounds present in a biological sample, then any profiling method can be judged on the fraction of all the metabolites present that contribute to the profile. In absolute terms, this can be difficult to assess, since without a "gold standard" complete profile it is impossible to know what components have been missed. In practice, however, applying multiple analytical approaches to the same sample soon throws up examples of components missed by the other techniques.

Two factors contribute to the completeness of a profile. Firstly, the general sensitivity limit of the technique sets the threshold below which no components are detected. While an "ideal" profile would include even components present as just a single molecule, practicality suggests that components present at such low levels are unlikely to have a biologically significant effect, and that a profile with sensitivity threshold in the pM or nM range would be adequate for all but the most demanding applications. Unfortunately, several techniques are more insensitive still, and will miss components that are biologically relevant. It is worth noting, however, that sensitivity can be a double-edged sword – many of the low abundance metabolic components are derived from symbiotic or xenobiotic organisms (such as gut microflora) not directly related to mammalian metabolism. It is unclear whether gathering such enlarged metabolic datasets using highly sensitive analytical approaches is useful – indeed, it may simply make the task of extracting a meaningful picture from the resulting dataset more difficult (see Sect. 3).

The second factor is the invisibility of particular compounds, or more likely classes of compounds, to a particular analytical technique. An obvious example would be the inability of non-volatile components to contribute to a gas chromatograph. Less obvious might be the inability of a technique to separate or distinguish components with closely related structures, such that a single "entry" in the profile is in fact a composite measure of two or more compounds.

2. *Bias*. The "ideal" profile is not merely a list of the components present in the sample but also an indicator of the relative amounts of each component. Thus, if certain components are more readily detected than others (on a molar basis) the analytical technique will display a bias, suggesting some components are present at relatively higher levels than they actually are.
3. *Cost*. The resource implications of any data-gathering exercise must be properly considered as part of the scientific experimental design. This is particularly true of any nonselective, high data density "omics" experiment, where the amount of understanding about the system which is ultimately gained will likely depend on how many different (that is, uncorrelated) components of the system are analyzed, rather than whether any particular class of components (be it metabolites, proteins or genes) has been exhaustively investigated. Consequently, the resource implications of selecting any given analytical methodology rightly forms a part of the experimental design: does the additional information gained by adding a particular technique to the portfolio of analyses to be performed on a given sample add sufficient uncorrelated variables to justify the resources employed, or could the same resources generate more uncorrelated information through application of a different technique?

No technique currently available is complete, unbiased, and inexpensive, but each has certain advantages for particular applications. Application of multiple techniques to the same sample may improve completeness and reduce bias, but only at increased resource implication.

2.2 Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR (nuclear magnetic resonance) spectroscopy has been widely used to generate metabonomic profiles, particularly of serum and urine samples. The primary advantages of NMR spectroscopy are the intrinsic reproducibility of the generated spectrum and the complete lack of bias. Across the information-dense region of the spectrum the coefficient of variation between replicate measures made on different days is below 1%. Reproducibility of this nature allows even small differences between profiles to be interpreted as significant, increasing the power of the experiment, particularly when relatively small numbers of profiles are being compared.

The NMR spectrum depends on the context-dependent resonance of hydrogen nuclei within the various molecules that compose the biological sample. As a result, any molecular structure containing at least one hydrogen nucleus is in principle represented within the spectrum. Since all biological molecules fall into this category, essentially every metabolite can contribute to an NMR-derived metabonomic profile. Furthermore, the intensity of the signal due to each hydrogen nucleus is of the same strength irrespective of its molecular context. Consequently, there is absolutely no bias in the estimated relative amounts of each of the metabolites detected.

Despite this lack of bias, NMR spectroscopy cannot be considered to generate a complete metabonomic profile because of the inherent insensitivity of the approach. Although the high reproducibility

allows even very small peaks to be distinguished from baseline noise, these peaks still represent relatively abundant molecular components of the sample. Although the absolute sensitivity cutoff varies depending on the particular implementation of the technique (for example, the use of cryoprobes significantly improves the sensitivity with nanogram quantities of compounds detected), nevertheless, very low abundance molecules are difficult to detect. Since many biological molecules of importance (such as vitamins and signaling molecules like prostaglandins or cyclicAMP) rarely achieve concentrations above the nM range, they are effectively absent from NMR-derived metabonomic datasets.

Although NMR spectroscopy is the tool of choice for most chemical structure determination problems, it also struggles to distinguish certain classes of closely related molecular structures. For example, it is difficult to study complex mixtures of fatty acids of different chain lengths by NMR spectroscopy, because the signals from the different molecular structures are overlaid in the resulting spectrum. Although ever more complex NMR-based approaches have been devised to aid separation and unique identification of given molecular structures (such as 2D-TOCSY or various heteronuclear NMR approaches), these approaches are resource intensive and may still be less informative than mass spectrometry for certain molecular classes.

2.3 Chromatography and Mass Spectroscopy (LC-MS and GC-MS)

Chromatography followed by mass spectrometry is the other major analytical tool that has been used to generate metabonomic profiles. Both liquid chromatography (LC) and gas chromatography (GC) have been used. GC may offer superior resolving power, at least for certain classes of molecules, but it is limited by the lack of volatility of many metabolites. Although this can be overcome to some extent by covalent modification of the sample prior to chromatography, a range of important metabolites still fail to enter the gas phase and hence do not contribute to the resulting metabonomic profile. As a result, the effective size limit for GC-MS is about 700 Da, whereas much larger molecules can contribute to the NMR-, and to some extent, to the LC-MS-derived profiles.

The major advantage of GC-MS and LC-MS is the sensitivity of the technique, which can detect component compounds down to the likely limit of biological relevance (and, indeed, may be too sensitive in some cases, populating the metabolic dataset with large numbers of minor contaminants of the sample, which are not the products of mammalian metabolism). As a result, components such as vitamins and signaling intermediates, which were invisible to the NMR spectrometer, now contribute to the GC-MS-derived metabonomic profile.

The combination of chromatography and mass spectroscopy (MS) can also allow unambiguous assignment of structure to the components of the biological sample in a way that is difficult, though not impossible, with NMR spectroscopy. Comparison of fragmentation patterns with databases, allows deconvolution of peaks with overlapping retention times, and as many as 1000 different molecular components to be unambiguously identified from a single complex biological fluid. This is particularly true when comparing different members of the same homologous series (such as fatty acids of

different chain lengths). However, LC-MS and GC-MS usually poorly resolve structural isomers that are readily distinguished by NMR spectroscopy. It is also important to note that, unlike NMR spectroscopy, mass spectrometry cannot be used to obtain the structure of unknown components contributing the profile – if the fragmentation pattern is not among the database of "known" metabolites, no assignment is possible.

Furthermore, MS-based detection is inherently biased: some molecules do not ionize, or only ionize poorly, under any given set of conditions and as a result are either completely invisible or detected only weakly. Consequently, an MS-derived metabonomic profile may be highly biased, and little information can be gained from the relative signals due to different components in the biological sample. However, it is still possible to compare the levels of the same component across different samples in a reliable and reproducible way. Again, variations in the implementation of the technique (such as using both positive and negative ionization modes and varying the cone voltage) can alleviate, though not eliminate, this problem at the cost of increased resource implication.

2.4 Nanosensors

More recently, it has become clear that various designs of the nanosensor can also be applied to generating metabonomic profiles. Nanosensors coated with various hydrophobic coatings adsorb a wide range of metabolites differentially, allowing a profile to be generated that is related in a complex fashion to the molecular composition of the fluid under study. Although to date, no metabonomic profile that has been generated using nanosensor technology has been reported in the scientific literature, it seems likely that such datasets will appear imminently.

Nanosensor-derived metabonomic profiles will presumably have the advantage of low cost (being rapid and high throughput, and not requiring capital intensive reading equipment). However, the complex nature of the relationship between the resulting data vector and the molecular composition of the sample will likely preclude any straightforward listing of the component molecules or estimation of the relative amounts. Such a profile will be neither substantially complete nor unbiased.

Yet, at least for clinical diagnostic applications, such profiles are not without utility. Although it may be difficult (or indeed impossible) to understand precisely which molecular components contribute to the systematic difference in profiles between two groups of interest (such as diseased individuals versus healthy individuals), nevertheless, the very presence of a systematic difference may be diagnostically useful.

2.5 Other Approaches

Many other analytical chemistry techniques can be applied to complex biological fluids to generate a metabonomic profile, although (like nanosensors) such profiles cannot usually be translated into a list of molecular components with associated relative concentrations. For example, infrared spectroscopy provides a low-cost approach to generating a metabonomic profile, which may be useful in some circumstances. It may be possible, depending on the nature of the biological

sample and the particular molecular components of interest, to generate a limited list of molecular components from an infrared spectrum.

Ultimately, however, the utility of a metabonomic profile obtained with any given technique depends on the application. Many studies (particularly aimed at diagnostic applications) may not require the immense structural detail that can be generated using the resource-intensive NMR and MS techniques. Equally, attempts to identify biomarkers associated with a particular phenotype will require a metabonomic profile that more nearly corresponds to the theoretical "ideal" profile. There remains, therefore, a considerable amount of trial and error in the selection of the analytical toolkit best suited to answering a particular question.

3 Methods for Interpreting Metabolic Profiles

3.1 The Problems of Interrogating Very Large Datasets

Much of the power of the metabonomic approach stems from the generation of very large datasets, which are essentially unselected in terms of the contributing components (that is, there was no preexisting hypothesis governing the selection of variables to be measured). As a consequence, special techniques are required to handle the resulting datasets, since extracting meaningful conclusions from datasets with millions of datapoints is a daunting exercise.

The basic aim in interpreting a metabonomic dataset is no different from any conventional multivariate analysis. The value of a dependent (or *Y*-) variable is estimated from a collection of measured *X*-variables (Fig. 3). For example, one might wish to estimate blood pressure from a range of physiological (or metabonomic) measures. While the *Y*-variable to be estimated (or "modeled") may be continuous (like blood pressure), it may equally be a discrete classification variable (which divides each observation into two or more groups, such as the presence and absence of a particular disease).

The problems associated with analyzing very large datasets are basically twofold. Firstly, a metabonomic profile might typically be composed of thousands of datapoints for each individual, yet such a profile may only have been generated from a relatively small number of individuals (tens or at most hundreds per group). Such datasets are typically described as "short and fat," having many more variables than observations (illustrated in Fig. 3). Analyzing such short and fat datasets using conventional multivariate statistics is dangerous, because it becomes increasingly likely that you can construct a model that correctly describes the phenotypic classification of the individuals by chance alone as the number of variables exceeds the number of observations.

Secondly, many of the variables that compose the metabolic profile may be highly correlated with each other. This is a particular problem with spectroscopic data, where neighboring variables are integrals of a continuous spectrum, forcing a relationship between nearby spectral regions. Conventional statistical approaches to multivariate analysis assume that all the predictor variables are independent, and the widespread colinearity of metabonomic profile variables renders the conventional multivariate models error prone.

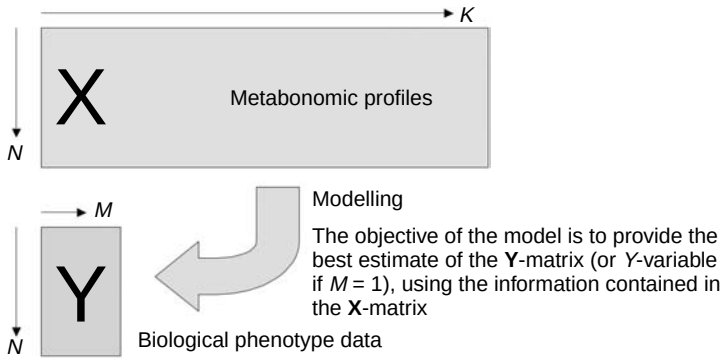

Fig. 3 The principle of regression modeling. The principle of regression modeling is to predict a range of features about a complex system (such as a biological organism) from a collection of unrelated measurements. In metabonomics, this consists of predicting phenotype or behavior (a **Y**-matrix of *M* features of *N* individuals) from a metabonomic dataset consisting of *K* measured metabolic variables from the same *N* individuals. Unlike conventional regression modeling, metabonomics datasets are typically "short and fat" with many more measurements per observation than individuals ($K \gg N$). The **Y**-matrix to be predicted might be a single continuous variable ($M = 1$), or even a categorical description (diseased or healthy, for example), whereupon the modeling is usually termed *a discriminant analysis*. Figure reproduced with permission from www.graingerlab.org.

3.2 Conventional Statistical Approaches (LDA)

As noted above, a conventional multivariate model (in which a dependent variable (Y) is predicted from a matrix of independent variables) is based on the assumption that the number of X-variables is less than the number of observations (n), and that all the X-variables are uncorrelated. In most experiments, a series of raw metabonomic profiles violates both assumptions and as a general rule conventional multivariate statistics should not be applied to metabonomic datasets.

However, various preprocessing steps can convert the raw metabonomic dataset into a form amenable to conventional statistical analysis. For example, there exists a range of prefilter algorithms (see Sect. 3.6 below) that allow the number of X-variables to be substantially reduced, for example, by retaining only those X-variables that are most significantly correlated with the dependent variable Y. Such variable-selection algorithms may also incorporate rules to eliminate intercorrelated X-variables. After variable selection, the resulting dataset can be analyzed by conventional multivariate statistics, such as linear discriminant analysis (LDA).

Although such variable-selection approaches circumvent the limitations of conventional multivariate statistics, they can limit the power of the analysis to identify associations between the **X**-matrix and the dependent Y-variable, since much of the information in the **X**-matrix has been discarded. In general, therefore, the other approaches below (which have been developed specifically for the analysis of very

large, or megavariate, datasets) tend to be more powerful.

3.3 Projection Methods (PCA and PLS)

Instead of selecting a subset of variables from the **X**-matrix, it is possible instead to combine the *X*-variables in linear combinations to generate a smaller number of composite variables. This approach (termed *projection*) deals effectively with both of the limitations to megavariate analysis: variable number is reduced, and intercorrelation is minimized.

The principle of projection is illustrated in Fig. 4. Here, a complex 3-dimensional object is represented by a simpler 2-dimensional shadow. In the left panel, the axis of projection is chosen such that the 2-D shadow poorly retains the information of the original object, whereas in the right panel the optimum projection is chosen, which retains most of the information encoded in the original 3-D object. The mathematical algorithms that underlie projection methods such as principal component analysis (PCA) or projection to latent structures using partial least squares (PLS) work by selecting the best projections. A major difference between PCA and PLS is that PLS is a supervised method, which means that the dependent *Y*-variable is used in the process of locating the best projection of the data. Application of these algorithms to high-dimensional datasets (such as metabonomic profiles) can yield just a handful of composite variables (principal components), which are simple linear combinations of the original matrix of *X*-variables. Unlike simpler variable-selection techniques, projection retains as much as possible of the information in the original **X**-matrix, while reducing the dimensionality of the dataset to manageable proportions.

One potential problem with projection methods is overfitting the model. With a sufficiently large number of *X*-variables, it will always be possible to generate combinations of the *X*-variables that predict the dependent variable very well. As a result, it is essential to include a robust external model validation step in the analysis protocol. One example of a useful validation step is scrambling the dependent variable and demonstrating that the **X**-matrix predicts the real dependent variable better than the scrambled variable. An alternative approach is to use the generated model to make predictions about an external dataset not used during model generation. Exhaustive validation is essential for models built using supervised techniques, such as PLS, where the dependent variable was

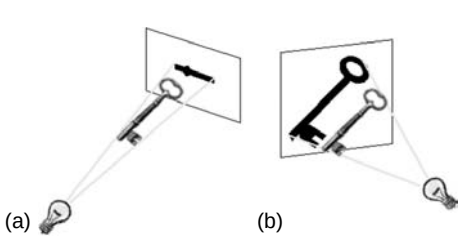

Fig. 4 The principle of projection. A complex high-dimensional object can be represented by a simpler, lower-dimensional model by projection. This is illustrated by the 2-D shadow of 3-D object such as key. However, depending on the particular projection selected, the 2-D model may be poorly representative of the original 3-D object (panel A). The aim of projection modeling tools such as PCA or PLS is to select the optimum lower-dimensional representation of the original complex object (right panel). Figure reproduced with permission from www.graingerlab.org.

used during the construction of the principal components.

3.4 Genetic Computing

Another approach to building models that optimally describe a dependent *Y*-variable from a very large **X**-matrix is to generate a large pool of random models (that are linear or nonlinear combinations of the *X*-variables), a few of which will be acceptable but most of which will be poor, and then apply an evolutionary algorithm to recombine the models and select for improved description of the dependent variable. The evolution is continued through a number of generations, and models can emerge that explain a good proportion of the variation in the *Y*-variable.

This approach can offer a number of advantages over projection-based methods. In particular, the ability to easily include rules that combine *X*-variables in nonlinear ways can be useful when modeling biological systems that have inherent nonlinearities. Another advantage of the evolved models is that they are generally easier to interpret than the optimized projection models. Combinations of simple rules can be more intuitively obvious than lists of principal components

One disadvantage of genetic computing approaches, however, can be "premature convergence" whereby the pool of models undergoing evolution rapidly converges on a local maximum of fitness and poorly explores the entire model space. Improvements are continually being made to the basic genetic computing algorithms, and recent advances such as multiobjective fitness functions can alleviate the premature convergence problem.

3.5 Other Approaches

The explosion of high data density analytical techniques in genomics and proteomics, as well as metabonomics, has stimulated the development and refinement of a wide range of other bioinformatics tools to assist in the interpretation of very large datasets. Hierarchical cluster analysis (HCA) is particularly popular for the analysis of gene expression datasets, but it is considerably less powerful than both projection methods and genetic computing algorithms, while offering few advantages. As a result, HCA has not been extensively used to interpret metabonomic datasets.

Another approach, which has been used in metabonomics, is neural network analysis. The neural net is set up with the *X*-variables as inputs and the dependent *Y*-variable as output, with a network of nodes in between. The mathematical function applied at each node is then iteratively varied during a learning phase to optimize the successful prediction of the *Y*-variable. Neural nets can be useful for performing discriminant analysis (that is, classifying observations into two or three groups on the basis of their **X**-matrix values) and hence used for clinical diagnostic purposes, but the models they generate are inherently difficult to interpret, because a large number of structurally different models can yield almost identical predictive power.

3.6 Data Preprocessing and Data Filters

Model building with very large datasets may be best performed in two or even more steps. Rather than directly applying one or more of the model building tools

described above to the raw **X**-matrix, it may be more powerful to perform a preprocessing step first. In fact, as with the model building tools themselves, the object of most of the preprocessing steps is to reduce dimensionality. However, empirical observation suggests that more powerful models can be generated if the dimensionality is reduced in stepwise fashion, perhaps using more than one tool, rather than in a single leap.

Consequently, the most straightforward preprocessing step would be to apply one of the model building tools twice in succession. For example, with projection methods, it is possible (and sometimes useful) to perform an initial principal component analysis reducing the dimensionality of the raw **X**-matrix from, say 8096 variables to tens or hundreds of principal components, and then treat the resulting matrix of principal components as the **X**-matrix for a second round of PCA to reduce the dimensionality down to two or three components, which can be more readily interpreted. The extent to which such hierarchical PCA improves the model compared with application of a single round of PCA depends very much on the particular data structure.

When dealing with metabonomic profiles derived from continuous spectra (such as an NMR spectrum), the neighboring variables (derived from integration of contiguous regions of the spectrum) can be very highly correlated. The shorter the interval along the abscissa that composes each variable, the higher is the degree of local intercorrelation. This can be reduced by integrating over wider intervals (or by averaging neighboring integrals, a process termed *binning*). This effectively reduces the dimensionality of the spectral data, while retaining much of the important information encoded in the spectrum.

Binning relies on the fact that the most highly intercorrelated variables are related to each other by their position along the abscissa in linear fashion. This may not be true for all types of data; in some cases, the most intercorrelated variables may fall cyclically (for example, in the untransformed free-induction decay signal from an NMR spectrometer). For analyzing such datasets, wavelet transformation, rather than binning, will provide the most efficient dimensionality reduction prior to model construction.

Finally, noise filters can be applied to the dataset prior to model building, and for NMR-derived metabonomic datasets this has been shown to significantly improve the performance of projection-based modeling tools. Noise filters (such as orthogonal signal correction; OSC) aim to remove variation in the **X**-matrix, which is uncorrelated with the dependent *Y*-variable, simplifying the dataset for subsequent modeling. It is important to remember that noise filters like OSC are therefore supervised methods, and exhaustive validation of models built on the filtered dataset will be required irrespective of the nature of the modeling tool subsequently used.

4 Applications of Metabonomics

4.1 Probing Normal Human Metabolism

The most obvious application for metabonomics is to aid understanding of normal human metabolism. It is possible to build up a detailed picture of metabolic pathways by taking a time series of metabolic profiles and constructing quantitative models describing the metabolic flux through

various pathways. Although the properties of key metabolic pathways (such as glycolysis or tricarboxylic acid cycle) have been extensively investigated for years, metabonomics can still reveal important aspects of metabolism, which had previously gone unnoticed. In particular, the interaction between endogenous metabolic pathways and the products of symbiotic bacterial metabolism have been extensively investigated using metabonomics.

4.2 Probing the Pathophysiology of Human Disease

Metabonomics is particularly powerful for analyzing the metabolic changes associated with the development of a particular disease state. By comparing metabonomic profiles taken from diseased individuals with profiles from healthy control individuals, it is possible to identify the systematic differences associated with the presence of the disease. What remains challenging, however, is determining which (if any) of these metabolic changes are causes of the disease pathology, as opposed to direct or even indirect consequences of the disease progression.

For example, metabonomics has been used to investigate the metabolic changes associated with the development of osteoporosis. Serum samples from women with low bone mineral density and from healthy control women were subjected to ^{1}H-NMR analysis, and the resultant profiles compared using the projection method PLS-DA following an OSC prefilter. One of the most important metabolites responsible for the separation of the diseased and healthy individuals was the amino acid proline. Women with pathological low bone mineral density had lower levels of proline in their serum; an observation that was later confirmed used conventional biochemical assays. While this does not prove that low serum proline is responsible for the lower bone density, it is a plausible hypothesis for further investigation: proline is a key constituent of the collagen component of bone matrix, and inadequate proline supplies for collagen biosynthesis could represent an entirely novel pathophysiological mechanism in the development of osteoporosis.

4.3 Investigating and Validating Animal Models of Disease

Animal models of human diseases are important tools in scientific and pharmaceutical research and development. The development of genetic manipulation techniques has allowed good models of many monogenic disorders (such as muscular dystrophy) to be developed, confident in the knowledge that the underlying cause of the disease is similar in the animal as in man. In contrast, for complex polygenic disorders such as atherosclerosis or Alzheimer's disease, it is unclear to what extent any given animal model mimics the molecular mechanisms underlying disease susceptibility in man, even if the phenotype of the animals faithfully mirrors the human disease.

Metabonomics offers an opportunity to address this question. Metabolic profiles of animal models can be compared directly with the profiles from human sufferers, allowing a comparison of the physiological perturbations accompanying disease development in the two species. Perhaps more powerfully, it should be possible to track the metabolic trajectory of both animals and humans as the disease progresses, with similar trajectories increasing the confidence in the likely validity

of the model. Studies of this nature are currently underway for a range of animal models of disease, but to date none have been published.

4.4 Clinical Diagnosis of Disease

In many senses, using metabonomics to perform clinical diagnoses is more straightforward than its application to biomarker identification and pathophysiological analysis. Providing a clinically useful diagnosis of a disease on the basis of a serum sample only requires the identification of a robust metabolic signature that always accompanies the disease and is rarely, if ever, present in healthy control individuals. It is not necessary to be able to identify any of the molecular components contributing to the disease-associated metabolic signature.

For example, projection analysis using PLS-DA of ^{1}H-NMR-derived metabonomic profiles of serum samples from individuals with coronary artery disease and healthy control individuals demonstrated clear separation of the two groups. After application of the OSC prefilter, it was possible to predict the disease status of individuals with at least 90% sensitivity and specificity. This metabonomics diagnostic test outperforms all existing noninvasive tests for coronary heart disease by a considerable margin. Although such a test is potentially useful in the clinic (since, at present, the gold-standard diagnostic test for heart disease is an invasive angiography procedure that is expensive and carries a small risk to the patient), it does not readily identify the particular molecular species responsible for separating the two groups. Much of the discriminatory power of the test falls in the region of the NMR spectrum due to lipid components (unsurprisingly, given our knowledge of the mechanisms underlying heart disease) but NMR poorly resolves these closely related lipid structures and considerable further work will be required before the precise molecular basis for the success of the test is known.

4.5 Selection of Subjects for Clinical Trials

At present, a substantial impediment to the testing of new therapeutics for certain diseases is the ability to identify potential sufferers ahead of time. For example, to test a drug proposed to reduce the incidence of myocardial infarction requires the recruitment of subjects at high risk of suffering a myocardial infarction during the trial. Unfortunately, current methods of identifying such subjects are poor, and trials of this nature can require the study of thousands of individuals for three years or more to accumulate sufficient myocardial infarction events for the impact of the drug to be detectable. Although metabonomics-based diagnostics have yet to be used in such applications, it is likely that their widespread adoption will rapidly follow the first successful demonstration of such use.

4.6 Monitoring Efficacy of Therapeutic Interventions

One of the most exciting applications of megavariate diagnostics, whether based on metabonomic, genomic, or proteomic profiles, is the prospect of personalized therapeutic interventions. At present, many drugs are used on broad swathes of the population (for example, statins to lower circulating LDL cholesterol) without any clear indication as to whether they are equally effective in all individuals. Pilot

studies already indicate that it is possible to predict the response of an individual to statin therapy from their metabonomic profile measured prior to beginning therapy. Such "pharmacometabonomics" could be extended to optimize the dose and delivery route of a wide range of drugs for each individual, with likely improvements in the efficacy of treatment.

4.7 Toxicology

Perhaps the most mature application of metabonomics is the application of metabolic profiling to toxicology. By studying the metabolic response to a range of model toxins (with organ-specific toxicity), it has been possible to identify metabolic signatures associated with damage to a particular organ. Extensive studies following the metabolic trajectory of animals treated with these model toxins have been published, although commercial considerations mean that few studies of clinically relevant pharmaceutical compositions have reached the public domain.

While metabonomics may help improve the predictivity of animal toxicology studies (which are notoriously difficult to interpret using conventional physiological and histological end points), perhaps the most exciting possibility is the use of metabonomics to perform early stage toxicology directly in man. Because of the sensitivity of metabonomics to detect minute perturbations in the metabolic signature, it may be possible to get an indication of the mode of toxicity of novel chemical entities given in man at doses well below those at which any irreversible damage might occur. The ability to perform meaningful toxicology in man should improve the safety of our medications, and at the same time reduce the number of promising pharmaceutical compounds dropped at a relatively late stage in development because of adverse, and possibly species-specific, side effects observed in the animal models currently used for toxicology.

4.8 Predicting Future Disease Risk

If a metabonomic profile can be used to diagnose the presence of an existing disease (such as coronary heart disease or osteoporosis), there is no reason in principle why it cannot be used to predict future disease susceptibility in the same way that genomic profiles are currently being used. To provide a useful indicator of future disease risk, there must be a component of the dynamic metabolic profile that is temporarily stable on a timescale of years and that is variable between individuals. We have already shown that such a stable interperson variance component exists in NMR-derived metabonomic profiles, and it will be interesting to see to what extent this stable element of the metabolic profile predicts the risk of a range of important diseases.

5 Future Prospects and Challenges

The discipline of metabonomics is expanding rapidly. Although selection from among the many combinations of analytical chemistry and mathematical modeling approaches for any given applications remains empirical, nevertheless, the number of metabonomics studies is growing quickly. Successful examples of the use of metabonomics to answer a broad array of scientific and practical questions are now plentiful in the literature (see

Sect. 4), and application to an ever-broader array of problems seems inevitable. Indeed, it is hard to think of a problem that would not be better addressed by a combination of high data density "omics" approaches than by hypothesis-driven reductionist experimentation. It seems that only a deep-seated unease among much of the scientific community for such exploratory, holistic approaches relegates the typical metabonomics experiment to "second-class status" and hampers an even more rapid expansion.

One of the challenges facing metabonomics over the coming years, therefore, is to better understand which of the many experimental approaches is optimum for a given use. Doing so will likely require the careful analysis of a single sample set by multiple different analytical chemistry techniques, and then each resulting dataset be interpreted using a range of mathematical modeling tools. In this way, the comparative power of the different approaches will begin to emerge.

Another important challenge is the integration of metabonomics datasets with the large profiles generated by other high data density techniques such as genomics or proteomics. In a sense, the division of "omics" science along the lines of the analytical techniques needed to make the measurements is entirely arbitrary. Ultimately, it should prove powerful to combine the profiles obtained from multiple different measurement approaches (whether gene expression, protein levels, or metabolite profiles) into a single "multi-omics descriptor." While there remains considerable debate as to exactly how this amalgamation should be performed, it is widely acknowledged that such a system-wide profile is likely to prove more powerful than a metabonomic or genomic profile alone for many applications. Only such a system-wide profile can allow a complete understanding of such a complex system as a biological organism.

See also Adipocytes.

Bibliography

Books and Reviews

Breiman, L. (2001) Statistical modelling: the two cultures, *Stat. Sci.* **16**(3), 199–231.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wold, S. *Multi-and Megavariate Data Analysis: Principles and Applications*, Umetrics Academy, Umea, Sweden, pp 1–525.

German, J.B., Roberts, M.A., Watkins, S.M. (2003) Personal metabolomics as a next generation nutritional assessment, *J. Nutr.* **133**(12), 4260–4266.

Holmes, E., Antti, H. (2002) Chemometric contributions to the evolution of metabonomics: mathematical solutions to characterising and interpreting complex biological NMR spectra, *The Analyst* **127**(12), 1549–1557.

Kell, D.B. (2004) Metabolomics and systems biology: making sense of the soup, *Curr. Opin. Microbiol.* **7**(3), 296–307.

Kell, D.B., Oliver, S.G. (2004) Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era, *BioEssays* **26**(1), 99–105.

Lindon, J.C., Holmes, E., Bollard, M.E., Stanley, E.G., Nicholson, J.K. (2004) Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis, *Biomarkers* **9**(1), 1–31.

Moolenaar, S.H., Engelke, U.F., Wevers, R.A. (2003) Proton nuclear magnetic resonance spectroscopy of body fluids in the field of inborn errors of metabolism, *Ann. Clin. Biochem.* **40**(Pt 1), 16–24.

Nicholson, J.K., Wilson, I.D. (2003) Understanding 'global' systems biology: metabonomics and the continuum of metabolism, *Nat. Rev. Drug Discov.* **2**(8), 668–676.

Weckwerth, W. (2003) Metabolomics in systems biology, *Annu. Rev. Plant Biol.* **54**, 669–689.

Primary Literature

Beckwith-Hall, B.M., Brindle, J.T., Barton, R.H., Coen, M., Holmes, E., Nicholson, J.K., Antti, H. (2002) Application of orthogonal signal correction to minimise the effects of physical and biological variation in high resolution 1H NMR spectra of biofluids, *The Analyst* **127**(10), 1283–1288.

Blake, G.J., Otvos, J.D., Rifai, N., Ridker, P.M. (2002) Low-density lipoprotein particle concentration and size as determined by nuclear magnetic resonance spectroscopy as predictors of cardiovascular disease in women, *Circulation* **106**(15), 1930–1937.

Boersma, M.G., Solyanikova, I.P., Van Berkel, W.J., Vervoort, J., Golovleva, L.A., Rietjens, I.M. (2001) 19F NMR metabolomics for the elucidation of microbial degradation pathways of fluorophenols, *J. Ind. Microbiol. Biotechnol.* **26**(1–2), 22–34.

Brindle, J.T., Nicholson, J.K., Schofield, P.M., Grainger, D.J., Holmes, E. (2003) Application of chemometrics to 1H NMR spectroscopic data to investigate a relationship between human serum metabolic profiles and hypertension, *The Analyst* **128**(1), 32–36.

Brindle, J.T., Antti, H., Holmes, E., Tranter, G., Nicholson, J.K., Bethell, H.W., Clarke, S., Schofield, P.M., McKilligin, E., Mosedale, D.E., Grainger, D.J. (2002) Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics, *Nat. Med.* **8**(12), 1439–1444; Epub 2002 Nov 25. Erratum in: *Nat. Med.* 2003 **9**(4), 477.

Bundy, J.G., Ramlov, H., Holmstrup, M. (2003) Multivariate metabolic profiling using 1H nuclear magnetic resonance spectroscopy of freeze-tolerant and freeze-intolerant earthworms exposed to frost, *Cryo. Letters* **24**(6), 347–358.

Burns, S.P., Woolf, D.A., Leonard, J.V., Iles, R.A. (1992) Investigation of urea cycle enzyme disorders by 1H-NMR spectroscopy, *Clin. Chim. Acta.* **209**(1–2), 47–60.

Bundy, J.G., Spurgeon, D.J., Svendsen, C., Hankard, P.K., Osborn, D., Lindon, J.C., Nicholson, J.K. (2002) Earthworm species of the genus Eisenia can be phenotypically differentiated by metabolic profiling, *FEBS Lett.* **521**(1–3), 115–120.

Choi, Y.H., Kim, H.K., Hazekamp, A., Erkelens, C., Lefeber, A.W., Verpoorte, R. (2004) Metabolomic differentiation of cannabis sativa cultivars using 1H NMR spectroscopy and principal component analysis, *J. Nat. Prod.* **67**(6), 953–957.

Coen, M., Lenz, E.M., Nicholson, J.K., Wilson, I.D., Pognan, F., Lindon, J.C. (2003) An integrated metabonomic investigation of acetaminophen toxicity in the mouse using NMR spectroscopy, *Chem. Res. Toxicol.* **16**(3), 295–303.

Eads, C.D., Furnish, C.M., Noda, I., Juhlin, K.D., Cooper, D.A., Morrall, S.W. (2004) Molecular factor analysis applied to collections of NMR spectra, *Anal. Chem.* **76**(7), 1982–1990.

Engelke, U.F., Liebrand-van Sambeek, M.L., de Jong, J.G., Leroy, J.G., Morava, E., Smeitink, J.A., Wevers, R.A. (2004) N-acetylated metabolites in urine: proton nuclear magnetic resonance spectroscopic study on patients with inborn errors of metabolism, *Clin. Chem.* **50**(1), 58–66.

Freedman, D.S., Otvos, J.D., Jeyarajah, E.J., Shalaurova, I., Cupples, L.A., Parise, H., D'Agostino, R.B., Wilson, P.W., Schaefer, E.J. (2004) Sex and age differences in lipoprotein subclasses measured by nuclear magnetic resonance spectroscopy: the Framingham study, *Clin. Chem.* **50**(7), 1189–1200.

Gavaghan, C.L., Holmes, E., Lenz, E., Wilson, I.D., Nicholson, J.K. (2000) An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse, *FEBS Lett.* **484**(3), 169–174.

Gavaghan, C.L., Nicholson, J.K., Connor, S.C., Wilson, I.D., Wright, B., Holmes, E. (2001) Directly coupled high-performance liquid chromatography and nuclear magnetic resonance spectroscopic with chemometric studies on metabolic variation in Sprague-Dawley rats, *Anal. Biochem.* **291**(2), 245–252.

Griffin, J.L. (2004) Metabolic profiles to define the genome: can we hear the phenotypes? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**(1446), 857–871.

Griffin, J.L., Cemal, C.K., Pook, M.A. (2004) Defining a metabolic phenotype in the brain of a transgenic mouse model of spinocerebellar ataxia 3, *Physiol. Genomics* **16**(3), 334–340.

Griffin, J.L., Troke, J., Walker, L.A., Shore, R.F., Lindon, J.C., Nicholson, J.K. (2000) The biochemical profile of rat testicular tissue as measured by magic angle spinning 1H NMR spectroscopy, *FEBS Lett.* **486**(3), 225–229.

Griffin, J.L., Williams, H.J., Sang, E., Clarke, K., Rae, C., Nicholson, J.K. (2001) Metabolic profiling of genetic disorders: a multitissue (1)H nuclear magnetic resonance spectroscopic and pattern recognition study into dystrophic tissue, *Anal. Biochem.* **293**(1), 16–21.

Hammad, S.M., Powell-Braxton, L., Otvos, J.D., Eldridge, L., Won, W., Lyons, T.J. (2003) Lipoprotein subclass profiles of hyperlipidemic diabetic mice measured by nuclear magnetic resonance spectroscopy, *Metabolism* **52**(7), 916–921.

Harrigan, G.G., LaPlante, R.H., Cosma, G.N., Cockerell, G., Goodacre, R., Maddox, J.F., Luyendyk, J.P., Ganey, P.E., Roth, R.A. (2004) Application of high-throughput Fourier-transform infrared spectroscopy in toxicology studies: contribution to a study on the development of an animal model for idiosyncratic toxicity, *Toxicol. Lett.* **146**(3), 197–205.

Hirabayashi, Y., Matsumoto, Y., Matsumoto, M., Toida, T., Iida, N., Matsubara, T., Kanzaki, T., Yokota, M., Ishizuka, I. (1990) Isolation and characterization of major urinary amino acid O-glycosides and a dipeptide O-glycoside from a new lysosomal storage disorder (Kanzaki disease). Excessive excretion of serine-and threonine-linked glycan in the patient urine, *J. Biol. Chem.* **265**(3), 1693–1701.

Holmes, E., Foxall, P.J., Spraul, M., Farrant, R.D., Nicholson, J.K., Lindon, J.C. (1997) 750 MHz 1H NMR spectroscopy characterisation of the complex metabolic pattern of urine from patients with inborn errors of metabolism: 2-hydroxyglutaric aciduria and maple syrup urine disease, *J. Pharm. Biomed. Anal.* **15**(11), 1647–1659.

Jonsson, P., Gullberg, J., Nordstrom, A., Kusano, M., Kowalczyk, M., Sjostrom, M., Moritz, T. (2004) A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS, *Anal. Chem.* **76**(6), 1738–1745.

Joshi, L., Van Eck, J.M., Mayo, K., Di Silvestro, R., Blake Nieto, M.E., Ganapathi, T., Haridas, V., Gutterman, J.U., Arntzen, C.J. (2002) Metabolomics of plant saponins: bioprospecting triterpene glycoside diversity with respect to mammalian cell targets, *OMICS* **6**(3), 235–246.

Kell, D.B. (2002) Metabolomics and machine learning: explanatory analysis of complex metabolome data using genetic programming to produce simple, robust rules, *Mol. Biol. Rep.* **29**(1–2), 237–241.

Ketchum, R.E., Rithner, C.D., Qiu, D., Kim, Y.S., Williams, R.M., Croteau, R.B. (2003) Taxus metabolomics: methyl jasmonate preferentially induces production of taxoids oxygenated at C-13 in Taxus x media cell cultures, *Phytochemistry* **62**(6), 901–909.

Keun, H.C., Beckonert, O., Griffin, J.L., Richter, C., Moskau, D., Lindon, J.C., Nicholson, J.K. (2002) Cryogenic probe 13 C NMR spectroscopy of urine for metabonomic studies, *Anal. Chem.* **74**(17), 4588–4593.

Keun, H.C., Ebbels, T.M., Bollard, M.E., Beckonert, O., Antti, H., Holmes, E., Lindon, J.C., Nicholson, J.K. (2004) Geometric trajectory analysis of metabolic responses to toxicity can define treatment specific profiles, *Chem. Res. Toxicol.* **17**(5), 579–587.

Khandelwal, P., Beyer, C.E., Lin, Q., Schechter, L.E., Bach, A.C., II. (2004) Studying rat brain neurochemistry using nanoprobe NMR spectroscopy: a metabonomics approach, *Anal. Chem.* **76**(14), 4123–4127.

Kikuchi, J., Shinozaki, K., Hirayama, T. (2004) Stable isotope labeling of Arabidopsis thaliana for an NMR-based Metabolomics approach, *Plant Cell. Physiol.* **45**(8), 1099–1104.

Kleno, T.G., Kiehr, B., Baunsgaard, D., Sidelmann, U.G. (2004) Combination of 'omics' data to investigate the mechanism(s) of hydrazine-induced hepatotoxicity in rats and to identify potential biomarkers, *Biomarkers* **9**(2), 116–138.

Kraus, W.E., Houmard, J.A., Duscha, B.D., Knetzger, K.J., Wharton, M.B., McCartney, J.S., Bales, C.W., Henes, S., Samsa, G.P., Otvos, J.D., Kulkarni, K.R., Slentz, C.A. (2002) Effects of the amount and intensity of exercise on plasma lipoproteins, *N. Engl. J. Med.* **347**(19), 1483–1492.

Lenz, E.M., Bright, J., Wilson, I.D., Morgan, S.R., Nash, A.F. (2003) 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects, *J. Pharm. Biomed. Anal.* **33**(5), 1103–1115.

Li, Z., Lamon-Fava, S., Otvos, J., Lichtenstein, A.H., Velez-Carrasco, W., McNamara, J.R., Ordovas, J.M., Schaefer, E.J. (2004) Fish

consumption shifts lipoprotein subfractions to a less atherogenic pattern in humans, *J. Nutr.* **134**(7), 1724–1728.

Mitchell, S., Holmes, E., Carmichael, P. (2002) Metabonics and medicine: the biochemical oracle, *Biologist (London)* **49**(5), 217–221.

Moolenaar, S.H., Engelke, U.F., Abeling, N.G., Mandel, H., Duran, M., Wevers, R.A. (2001) Prolidase deficiency diagnosed by 1H NMR spectroscopy of urine, *J. Inherit. Metab. Dis.* **24**(8), 843–850.

Moolenaar, S.H., Gohlich-Ratmann, G., Engelke, U.F., Spraul, M., Humpfer, E., Dvortsak, P., Voit, T., Hoffmann, G.F., Brautigam, C., van Kuilenburg, A.B., van Gennip, A., Vreken, P., Wevers, R.A. (2001) Beta-ureidopropionase deficiency: a novel inborn error of metabolism discovered using NMR spectroscopy on urine, *Magn. Reson. Med.* **46**(5), 1014–1017.

Mortishire-Smith, R.J., Skiles, G.L., Lawrence, J.W., Spence, S., Nicholls, A.W., Johnson, B.A., Nicholson, J.K. (2004) Use of metabonomics to identify impaired fatty acid metabolism as the mechanism of a drug-induced toxicity, *Chem. Res. Toxicol.* **17**(2), 165–173.

Nikiforova, V.J., Gakiere, B., Kempa, S., Adamik, M., Willmitzer, L., Hesse, H., Hoefgen, R. (2004) Towards dissecting nutrient metabolism in plants: a systems biology case study on sulphur metabolism, *J. Exp. Bot.* **55**(404), 1861–1870. Epub 2004 Jun 18.

Ohdoi, C., Nyhan, W.L., Kuhara, T. (2003) Chemical diagnosis of Lesch-Nyhan syndrome using gas chromatography-mass spectrometry detection, *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* **792**(1), 123–130.

Ohse, M., Matsuo, M., Ishida, A., Kuhara, T. (2002) Screening and diagnosis of beta-ureidopropionase deficiency by gas chromatographic/mass spectrometric analysis of urine, *J. Mass. Spectrom.* **37**(9), 954–962.

Ott, K.H., Aranibar, N., Singh, B., Stockton, G.W. (2003) Metabonomics classifies pathways affected by bioactive compounds, Artificial neural network classification of NMR spectra of plant extracts. *Phytochemistry* **62**(6), 971–985.

Otvos, J.D., Jeyarajah, E.J., Bennett, D.W., Krauss, R.M. (1992) Development of a proton nuclear magnetic resonance spectroscopic method for determining plasma lipoprotein concentrations and subspecies distributions from a single, rapid measurement, *Clin. Chem.* **38**(9), 1632–1638.

Pham-Tuan, H., Kaskavelis, L., Daykin, C.A., Janssen, H.G. (2003) Method development in high-performance liquid chromatography for high-throughput profiling and metabonomic studies of biofluid samples, *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* **789**(2), 283–301.

Plumb, R.S., Stumpf, C.L., Gorenstein, M.V., Castro-Perez, J.M., Dear, G.J., Anthony, M., Sweatman, B.C., Connor, S.C., Haselden, J.N. (2002) Metabonomics: the use of electrospray mass spectrometry coupled to reversed-phase liquid chromatography shows potential for the screening of rat urine in drug development, *Rapid Commun. Mass Spectrom.* **16**(20), 1991–1996.

Purohit, P.V., Rocke, D.M., Viant, M.R., Woodruff, D.L. (2004) Discrimination models using variance-stabilizing transformation of metabolomic NMR data, *OMICS* **Summer; 8**(2), 118–130.

Raamsdonk, L.M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M.C., Berden, J.A., Brindle, K.M., Kell, D.B., Rowland, J.J., Westerhoff, H.V., van Dam, K., Oliver, S.G. (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations, *Nat. Biotechnol.* **19**(1), 45–50.

Robertson, D.G., Reily, M.D., Albassam, M., Dethloff, L.A. (2001) Metabonomic assessment of vasculitis in rats, *Cardiovasc. Toxicol.* **1**(1), 7–19.

Sato, S., Soga, T., Nishioka, T., Tomita, M. (2004) Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection, *Plant J.* **40**(1), 151–163.

Slim, R.M., Robertson, D.G., Albassam, M., Reily, M.D., Robosky, L., Dethloff, L.A. (2002) Effect of dexamethasone on the metabonomics profile associated with phosphodiesterase inhibitor-induced vascular lesions in rats, *Toxicol. Appl. Pharmacol.* **183**(2), 108–109.

Soedamah-Muthu, S.S., Chang, Y.F., Otvos, J., Evans, R.W., Orchard, T.J. (2003) Pittsburgh epidemiology of diabetes complications study. Lipoprotein subclass measurements by nuclear magnetic resonance spectroscopy improve the prediction of coronary artery

disease in Type 1 diabetes. A prospective report from the Pittsburgh epidemiology of diabetes complications study, *Diabetologia* **46**(5), 674–682.

Tang, H., Wang, Y., Nicholson, J.K., Lindon, J.C. (2004) Use of relaxation-edited one-dimensional and two dimensional nuclear magnetic resonance spectroscopy to improve detection of small metabolites in blood plasma, *Anal. Biochem.* **325**(2), 260–272.

Tate, A.R., Foxall, P.J., Holmes, E., Moka, D., Spraul, M., Nicholson, J.K., Lindon, J.C. (2000) Distinction between normal and renal cell carcinoma kidney cortical biopsy samples using pattern recognition of (1)H magic angle spinning (MAS) NMR spectra, *NMR Biomed.* **13**(2), 64–71.

Van, Q.N., Chmurny, G.N., Veenstra, T.D. (2003) The depletion of protein signals in metabonomics analysis with the WET-CPMG pulse sequence, *Biochem. Biophys. Res. Commun.* **301**(4), 952–959.

Verhoeckx, K.C., Bijlsma, S., Jespersen, S., Ramaker, R., Verheij, E.R., Witkamp, R.F., Van Der Greef, J., Rodenburg, R.J. (2004) Characterization of anti-inflammatory compounds using transcriptomics, proteomics, and metabolomics in combination with multivariate data analysis, *Int. Immunopharmacol.* **4**(12), 1499–1514.

Viant, M.R. (2003) Improved methods for the acquisition and interpretation of NMR metabolomic data, *Biochem. Biophys. Res. Commun.* **310**(3), 943–948.

Wang, Y., Bollard, M.E., Keun, H., Antti, H., Beckonert, O., Ebbels, T.M., Lindon, J.C., Holmes, E., Tang, H., Nicholson, J.K. (2003) Spectral editing and pattern recognition methods applied to high-resolution magic-angle spinning 1H nuclear magnetic resonance spectroscopy of liver tissues, *Anal. Biochem.* **323**(1), 26–32.

Watkins, S.M., Reifsnyder, P.R., Pan, H.J., German, J.B., Leiter, E.H. (2002) Lipid metabolome-wide effects of the PPARgamma agonist rosiglitazone, *J. Lipid. Res.* **43**(11), 1809–1817.

Wevers, R.A., Engelke, U., Heerschap, A. (1994) High-resolution 1H-NMR spectroscopy of blood plasma for metabolic studies, *Clin. Chem.* **40**(7 Pt 1), 1245–1250.

Wevers, R.A., Engelke, U.F., Moolenaar, S.H., Brautigam, C., de Jong, J.G., Duran, R., de Abreu, R.A., van Gennip, A.H. (1999) 1H-NMR spectroscopy of body fluids: inborn errors of purine and pyrimidine metabolism, *Clin. Chem.* **45**(4), 539–548.

Xu, J., Chang, V., Joseph, S.B., Trujillo, C., Bassilian, S., Saad, M.F., Lee, W.N., Kurland, I.J. (2004) Peroxisomal proliferator-activated receptor alpha deficiency diminishes insulin-responsiveness of gluconeogenic/glycolytic/pentose gene expression and substrate cycle flux, *Endocrinology* **145**(3), 1087–1095.

Yu, H.H., Ginsburg, G.S., O'Toole, M.L., Otvos, J.D., Douglas, P.S., Rifai, N. (1999) Acute changes in serum lipids and lipoprotein subclasses in triathletes as assessed by proton nuclear magnetic resonance spectroscopy, *Arterioscler. Thromb. Vasc. Biol.* **19**(8), 1945–1949.

Metalloenzymes

Walther R. Ellis
Utah State University, Logan, UT, USA

1 Occurrence 211
1.1 Discovery 211
1.2 Biological Importance 211
1.3 Metal Ion Bioavailability 212
1.4 Active Site Assembly 212

2 Active Site Characterization 213
2.1 Metal Content 213
2.2 Physical Methods 214
2.3 Metal/Ligand Substitution 215

3 Representative Metalloenzymes of Medical Importance 215
3.1 Functional Roles of Metal Cofactors 215
3.2 Mammalian Iron Enzymes 216
3.3 Mammalian Copper Enzymes 219
3.4 Mammalian Zinc Enzymes 221
3.5 Mammalian Cytochrome *c* Oxidase 224
3.6 Mammalian Enzymes Containing Less Common Metals 226
3.7 Bacterial Metalloenzymes and Their Roles in Infection 229

4 Perspectives 231

Bibliography 232
Books and Reviews 232
Primary Literature 232

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

Keywords

Cytochrome P-450
Refers to a family of heme monooxygenases, present in certain pseudomonads and most mammalian cell types, that catalyze the oxidation of a wide variety of structurally diverse compounds.

Electrophilic Catalysis
Refers to the electrostatic stabilization of a negative charge that develops in the transition states of certain reactions.

Metalloprotease
An enzyme, typically containing Zn, that catalyzes the hydrolysis of peptide bonds.

Oxidase
An enzyme that catalyzes an oxidation using O_2 as the electron acceptor; O atoms from dioxygen are not incorporated into the product of the oxidation.

Oxygenase
An enzyme that catalyzes the reaction of O_2 with an organic substrate in which oxygen atoms (one in the case of monooxygenases; two in the case of dioxygenases) from dioxygen are incorporated into the product.

Peroxidase
An enzyme that catalyzes the oxidation of an organic/inorganic substrate by hydrogen peroxide.

Superoxide Dismutase
A metalloenzyme that catalyzes the disproportionation of superoxide, forming hydrogen peroxide and O_2.

Urease
A nickel enzyme that hydrolyzes urea, forming ammonia and carbamate.

Metalloenzymes, which comprise approximately one-third of the known enzymes, require stoichiometric quantities of metal ions as cofactors, typically transition metal ions, for their catalytic activities. The roles of metal ions in enzyme active sites (aside from structure maintenance) include electron transfer, oxygen atom transfer, formation of coordinated hydroxide, electrophilic catalysis, as well as substrate binding. Metalloenzymes catalyze numerous reactions of physiological importance, including mitochondrial O_2 reduction, peptide bond cleavage, hydrocarbon hydroxylation, destruction of O_2^- and H_2O_2, and hydration of CO_2.

1 Occurrence

1.1 Discovery

The study of metalloenzymes has its roots in investigations that took place more than a century ago. In 1897, Gabriel Bertrand, working on laccase (a polyphenol oxidase), suggested for the first time that a metal ion was essential for the catalytic activity of an enzyme. At this time, the nature of enzymes was a matter of vociferous debate – Are enzymes protein-based catalysts, or is the catalysis traceable to low-level contaminants? In 1926, James Sumner presented a pivotal result: crystallization of jack bean urease, demonstration that urease is a protein, and that the dissolved crystals catalyze the hydrolysis of urea. Sumner, who failed to detect any metal ions in his urease preparations, expressed the view that metal ions are unlikely to play important roles in the enzymatic catalysis of biological reactions. It is ironic that 49 years after Sumner's pioneering crystallization of an enzyme, urease was subsequently found to contain nickel ions, which are essential for enzyme activity!

Prior to the development of modern methods for trace element analysis and spectroscopic instrumentation, many metalloenzymes (e.g. blue copper oxidases) were initially isolated simply because they were a colored component of a cell or tissue homogenate. During the last 40 years, numerous nonchromophoric metalloenzymes have been isolated as well. Our current understanding of metalloenzymes is largely based on enzymes containing iron, copper, or zinc. Numerous examples of enzymes containing other metals (Mg, Ca, Mn, Co, Ni, V, Mo, W) are also known; however, the structural and mechanistic information currently available for these is less detailed.

1.2 Biological Importance

Metalloenzymes play key roles in many processes central to human physiology, including the biosynthesis of DNA and certain amino acids, steroid metabolism, destruction of superoxide and hydrogen peroxide, biosynthesis of leukotrienes and prostaglandins, carbon dioxide hydration, neurotransmitter metabolism, digestion, collagen biosynthesis, and, of course, respiration. The latter could be viewed as a process of global bioenergetic importance, one that complements photosynthesis – the dioxygen that is evolved (via the Mn_4 oxygen-evolving complex of Photosystem II) by photosynthetic organisms is consumed by aerobic microbes and animals (Fe, Cu cytochrome oxidases catalyze the reduction of O_2 to H_2O).

Within the last 25 years, it has become widely appreciated that metal ions play important catalytic roles that frequently cannot be matched by protein side chains or organic prosthetic groups. The most significant observation of this type concerns the microbial fixation of molecular nitrogen, a very inert molecule:

$$N_2 + 6H^+ + 6e^- \longrightarrow 2NH_3 \qquad (1)$$

This reaction is catalyzed by nitrogenase, whose active site consists of a dissociable Mo- and Fe-containing cofactor. The catalysis of such multielectron redox reactions is frequently carried out by metalloenzymes containing multimetal centers (i.e. metal clusters).

1.3 Metal Ion Bioavailability

Only a small number of the metallic elements appear to be utilized in biology. Of these, Na^+, K^+, Ca^{2+}, and Mg^{2+} are considered macrominerals, or bulk elements; high concentrations of these ions are needed for osmotic homeostasis, neuromuscular transmission, and biomineralization (e.g. bone formation). As indicated in Table 1, other essential metals are present in trace quantities in humans and most other organisms. Even the most prominent biologically active transition metals (iron, zinc, and copper) are trace elements.

The selection of metal ions for incorporation into metalloenzymes is strongly influenced by bioavailability – a given element must be abundant in the environment and must be present in an extractable form. A striking exception to this generalization involves the nearly universal requirement for iron. Organisms have evolved selective uptake mechanisms for this element, the most abundant transition metal in the earth's crust, which forms insoluble ferric hydroxides in the presence of O_2. The incorporation of metal ions into metalloenzymes is also influenced by other, chemically oriented, parameters such as ionic radius, charge, preferred coordination geometry, ligand substitution and redox kinetics, aqueous solution chemistry, and thermodynamic stability.

Tab. 1 Essential metal composition of a 70-kg human adult[a].

Element	*Composition (wt%)*
Ca	1.4
Na	0.63
K	0.26
Mg	4×10^{-2}
Fe	5×10^{-3}
Zn	3×10^{-3}
Cu	1×10^{-4}
Mn	2×10^{-5}
Mo	2×10^{-5}
Ni	4×10^{-6}
Cr	4×10^{-6}
V	3×10^{-6}
Co	2×10^{-6}

[a] Note: O,C,H,N,P,&S constitute 97.6 wt%.

1.4 Active Site Assembly

Incorporation of a metal ion, a posttranslational biosynthetic event, requires that the folding of the polypeptide chain permit several side chains to congregate to form an appropriate metal-binding site. The primary metal-binding amino acid side chains are imidazole (His), carboxylate (Asp and Glu), thiol (Cys), thioether (Met), and hydroxyl (Ser, Thr, and Tyr). Less frequently, indole (Trp), guanidinium (Arg), and amide (Asn and Gln) groups are used. Backbone carbonyl groups can also participate in metal binding. The side chain functional groups must usually be deprotonated in order for a donor atom (O, N, or S) to form a metal–ligand bond.

Some metal ions coordinate to their binding sites in apo-metalloenzymes as simple aqua ions. For example, Zn^{2+} binds to apo-carbonic anhydrase in a multidentate ligand reaction: the metal ion sheds coordinated water molecules as it binds to the active site of the apoenzyme, which could be viewed as an elaborate chelating agent. However, redox-active metal ions (e.g. Cu, Fe, Mn, Mo) pose special problems by virtue of their reactivities with O_2^- and H_2O_2. For example, it has become clear during the last decade that pools of soluble copper ions are not used in the physiological activation of eukaryotic

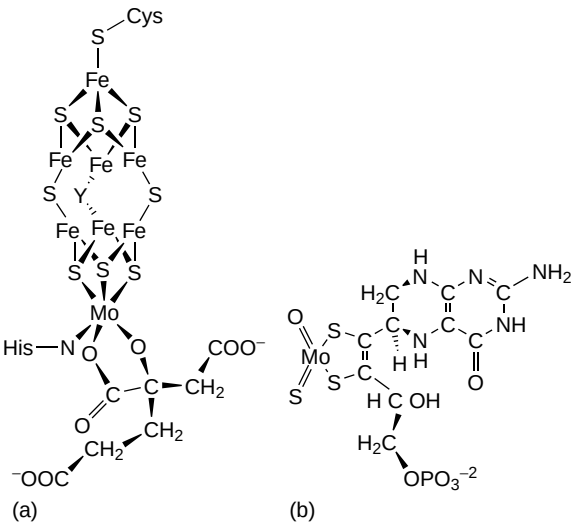

Fig. 1 Molybdenum cofactors: (a) Structure of the FeMo cofactor of *Azotobacter vinelandii* nitrogenase. The cofactor is bound to the enzyme via the indicated cysteine and histidine residues. Homocitrate is bound to the molybdenum center. (b) Proposed structure of the oxidized molybdopterin cofactor (Mo-co) of mammalian molybdenum enzymes.

copper enzymes, such as Cu,Zn superoxide dismutase and cytochrome *c* oxidase. Copper ions are known Fenton reagents ($OH^{\bullet}$ generators), and their sequestration is a straightforward way of circumventing unwanted redox reactions. Intracellular copper delivery is mediated by "chaperone" proteins that form complexes with their apoenzyme targets prior to metal-ion exchange. Emerging research results indicate that chaperones likely exist for other redox-active metals as well.

The biosyntheses of some metalloenzymes of physiological importance requires the prior synthesis of a specialized organometallic complex. Such species include vitamin B_{12} (a cobalt corrin), hemes, and molybdenum cofactors (Fig. 1). Molybdoenzymes contain unusual metal-containing cofactors that have been shown to be traceable to the participation of accessory biosynthetic genes. The FeMo cofactor (FeMoCo) of nitrogenase also illustrates a more complex phenomenon – the use of metal clusters as cofactors in enzymes. Such aggregates typically utilize sulfide or oxide (hydroxide) ions as bridges between metal centers. Sometimes (e.g. the iron–sulfur cluster of aconitase) the cluster assembles in a stepwise fashion after apoenzyme biosynthesis. In other cases (e.g. the nitrogenase FeMo cofactor), accessory proteins, "molecular scaffolds", are needed for complete cluster assembly; the cluster is subsequently transferred to the apoenzyme to complete the biosynthesis of the mature metalloenzyme.

2 Active Site Characterization

2.1 Metal Content

Traditional methods of trace element analysis have relied on the use of techniques best suited for quantitative, rather than qualitative, analysis such as complexometric titrations, colorimetric procedures, or atomic absorption spectrometry. Determinations of the metal contents of putative metalloenzymes have frequently

been in error, owing to impure enzyme preparations, cofactor loss during purification, or inadequate analytical techniques. The most suitable methods for qualitative metal analysis include X-ray fluorescence spectrometry, neutron activation analysis, and atomic emission plasma spectrometry. Many metalloenzymes contain more than one kind of metal ion; hence, both qualitative and quantitative metal analyses should be done whenever possible.

A key indicator of metal-activated enzyme turnover is the ability to abolish the activity by adding a metal chelator (usually EDTA) or a small molecule (e.g. CO, CN^-, N_3^-, carboxylic acids) that is known to bind to metal ions in proteins and coordination compounds. Further proof of the nature of the native metal ion often requires tedious trial-and-error studies to determine which metal gives the greatest enzyme activity.

2.2 Physical Methods

Going beyond the determination of metal content, to demonstrate that a particular metal ion is present at the active site of a metalloenzyme, requires spectroscopic and/or X-ray crystallographic experimentation. Detailed structures, derived from X-ray crystallographic or nuclear magnetic resonance analyses, are available for less than 5% of the known metalloenzymes. Other physical methods have therefore been used in obtaining information about physicochemical properties these enzymes, such as hydrodynamic radius, substrate-binding constants, or types of ligands bound to the metal(s).

Extended X-ray absorption fine structure (EXAFS) spectroscopy can provide information about a primary metal coordination environment that is comparable with that obtained from X-ray crystallography – numbers and types of ligands, metal–ligand bond lengths, and bond angles. Other spectroscopic methods (e.g. infrared, Raman, electron spin resonance, fluorescence, electronic absorption), while much less informative, can nonetheless yield valuable information about the ability of an active-site metal ion to bind exogenous ligands (e.g. inhibitors) and about the likely oxidation state(s) in cases involving redox-active metals. In cases involving redox-active metal ions, determining the metal-ion oxidation states is an important prelude to formulation of enzyme turnover mechanisms. In the past, controversies arose because of proposed metalloenzyme mechanisms that invoked unusual oxidation states (e.g. Cu(III)) that subsequently were shown to be incorrect. Sometimes, metal–ligand bonds (especially involving Cys, O_2, S^{2-}, O^{2-} ligands) have such a high degree of charge transfer that assignments of metal-ion oxidation states are difficult.

Redox metalloenzymes usually undergo pronounced color changes during enzyme turnover, making optical spectroscopic studies particularly valuable. This first became evident with the early twentieth-century publications by Michaelis and Menten of observations on the turnover of horseradish peroxidase, a heme enzyme. Changes in the visible region of the spectrum during turnover (due to formation of transient intermediates) led them to formulate the Michaelis–Menten model of enzyme action.

Determinations of redox potentials of resting forms of redox enzymes yields valuable information that can be used to set constraints on proposed enzyme mechanisms. In some cases, notably the heme enzymes horseradish peroxidase and yeast

cytochrome *c* peroxidase, electrochemical methods have been used to generate unstable intermediates (ferryl hemes) for spectroscopic study and determination of their redox potentials.

2.3 Metal/Ligand Substitution

Metalloenzymes frequently contain metal cofactors that are spectroscopically useless. For example, zinc enzymes are diamagnetic and colorless – electron spin resonance or electronic absorption spectroscopy are therefore uninformative. However, information about the environment of the metal-binding site can be obtained by replacing the Zn^{2+} ion with a paramagnetic, chromophoric metal-ion probe. Co^{2+} readily replaces Zn^{2+} – they prefer similar coordination environments, have comparable ionic radii, and display similar reactivity patterns. The resultant cobalt-substituted enzyme can be characterized with regard to intermediates formed during enzymatic turnover. Furthermore, the use of metal derivatives can also indicate whether the metal binding site is rigid or flexible.

Obtaining the structure of a metalloenzyme is only the first step in developing an understanding of its physiological function. The production of recombinant derivatives affords opportunities to address additional issues, such as endogenous ligand exchange at the active-site metal(s) and participation of nearby amino acid residues in substrate binding/release, proton transfer, and/or electron transfer. Fast kinetic techniques (stopped-flow spectrophotometry, freeze-quench trapping of intermediates, flash photolysis, temperature-jump relaxation) are typically used to map out the individual steps that characterize the turnover of a metalloenzyme. In many cases, individual steps are too fast to accurately measure. One commonly used application of such techniques is the study of absorption spectra of transient enzyme intermediates generated by photolyzing carbon monoxide adducts of reduced heme enzymes in the presence of O_2. What tends to be overlooked in interpretations of structure and reactivity data is the importance of protein fluctuations (particularly in substrate binding and product release) – metalloenzymes are much more than metal ions in disguise.

3 Representative Metalloenzymes of Medical Importance

3.1 Functional Roles of Metal Cofactors

There are two principal types of metal-dependent enzyme. Metalloenzymes contain at least one tightly bound metal ion, at the active site, that is required for activity. Metal-activated enzymes, on the other hand, generally lose catalytic activity during purification because their affinity for the required metal is rather low. Mg^{2+}-, K^{+}-, and (most) Ca^{2+}-dependent enzymes are metal-activated and will not be discussed further. Na^{+}-dependency has yet to be unequivocally demonstrated for an enzyme.

Examples of metalloenzymes are known for each class of enzyme designated by the International Union of Biochemistry and are as follows: oxidoreductases, transferases, hydrolases, lyases, isomerases, and ligases. The following properties of metal ions make them well suited as catalysts for these types of reactions:

1. Metal ions bind at least three (usually four or more) ligands, thereby promoting the organization of protein structure.
2. With the notable exception of zinc, all of the other known transition metal cofactors can potentially exist in two or more oxidation states; metalloenzyme catalysis of oxidation-reduction (redox) reactions is thus quite common.
3. Metal-ion cofactors are electrophilic and can serve as effective Lewis acids for binding and activating substrates.

Two types of metal-ion reactivity are of fundamental importance in considering the wide range of activities displayed by metalloenzymes: changes in oxidation state and changes in bound ligands. Numerous examples of electron-transfer metalloproteins (blue copper proteins, iron–sulfur proteins, cytochromes) are known; the metal coordination spheres of these electron carriers do not change during electron transfer. Metal ions possessing static coordination spheres could alternatively play purely structural roles, as do Ca^{2+} in calmodulin, horseradish peroxidase, or stromelysin.

Certain metals (e.g. Zn^{2+} in carboxypeptidase A) cannot undergo redox reactions within the constraints imposed by nature; instead, acid–base catalysis is the *raison d'être* for the metal in the enzyme. Lastly, both the metal-ion oxidation state and coordination sphere could change during enzymatic turnover. Regardless, it is clear that metal ions that directly participate in enzymatic catalysis must possess dynamic coordination environments.

3.2
Mammalian Iron Enzymes

With few exceptions (certain lactobacilli), all forms of life require iron as a biocatalyst. However, only small amounts are needed – a healthy 70-kg human adult possesses just 4 g of the metal. Most of the body's iron is found in hemoglobin and ferritin; taken together, the body's iron enzymes contain less than 300 mg of iron. The common oxidation states are 2+ and 3+; higher oxidation states are known to be operative during the turnover of some iron enzymes. The widespread occurrence of iron–porphyrin prosthetic groups, or hemes, in biology has prompted the subdivision of iron enzymes into nonheme iron and heme classes.

Nonheme iron enzymes, in turn, could be classified as mononuclear iron, binuclear iron, or iron–sulfur enzymes. Table 2 lists some of these enzymes that have been found in humans, together with the reactions catalyzed and metal contents. Iron can evidently be used to catalyze both redox and acid–base reactions. Unfortunately, the active site structures of the intermediates present during turnover are not yet well understood for many of these enzymes.

A representation, partly derived from an X-ray crystallographic analysis, of the active site of pig heart mitochondrial aconitase is displayed in Fig. 2. This iron–sulfur enzyme plays a key role in the Krebs cycle, where it catalyzes the isomerization, via a *cis*-aconitate intermediate, of citrate to isocitrate. Spectroscopic and X-ray diffraction experiments have convincingly shown that the substrate directly binds to the iron–sulfur cluster during enzyme turnover. Replacement of the Ser residue depicted in the figure with Ala does not alter binding of citrate, and X-ray structures indicate that the isomerization likely proceeds via a 180° inversion of *cis*-aconitate about the $C_\alpha - C_\beta$ bond.

Several nonheme iron enzymes play important roles in the metabolism of the

Tab. 2 Representative human nonheme iron enzymes.

Enzyme	***Reaction Catalyzed***
Aconitase (Fe_4S_4)	Citrate $\rightleftharpoons$ isocitrate
Homogentisate dioxygenase (Fe)	Homogentisate $\longrightarrow$ maleylacetoacetate
Lipoxygenase (Fe)	Unsaturated fatty acids $\longrightarrow$ conjugated fatty acid hydroperoxides
Phenylalanine hydroxylase[a] (Fe)	L-Phe $\longrightarrow$ L-Tyr
Purple acid phosphatases (Fe_2)	Phosphate monoester hydrolysis
Tryptophan hydroxylase[a] (Fe)	L-Trp $\longrightarrow$ 5-hydroxy-L-Trp

[a]Also uses a tetrahydrobiopterin as an active-site cofactor.

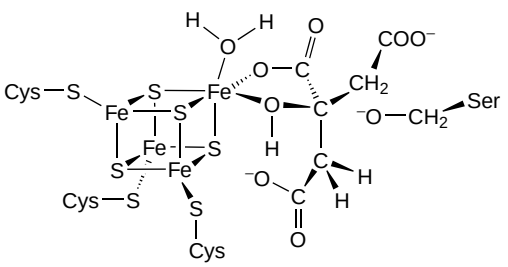

Fig. 2 Proposed structure of citrate bound to the active-site iron–sulfur cluster of pig heart mitochondrial aconitase.

aromatic amino acids phenylalanine, tyrosine, and tryptophan. These enzymes use nonheme iron and tetrahydrobiopterin (as redox cofactors) and O_2 in catalyzing regiospecific hydroxylations of these amino acids. Disfunctions in these enzymatic activities have been correlated with severe disorders, including phenylketonuria and psychiatric conditions attributed to altered serotonergic/dopaminergic neurotransmission. The mechanisms of these enzymes are believed to involve novel Fe^{2+}-O-O-pterin intermediates that heterolytically cleave to produce highly reactive Fe^{4+}=O species that subsequently oxidize substrates bound nearby.

Heme proteins figure prominently in biochemistry – their functions include electron transfer, O_2 transport, O_2 activation, and H_2O_2 activation. The latter two functions are associated with several major subclasses of oxidoreductases: oxidases, oxygenases, peroxidases, and catalases. While such activities can also be displayed by enzymes containing nonheme iron, copper, manganese, or even vanadium, it has long been evident that the use of heme cofactors in dioxygen and peroxide metabolism overshadows other alternatives in eukaryotes. Representative heme enzymes of relevance to human physiology are set out in Table 3.

All heme enzymes share one notable feature: during turnover, an oxidized intermediate (oxoferryl iron) is formed from the reaction of peroxide bound to Fe^{3+}. Two-electron reduction of dioxygen yields peroxide; not surprisingly, oxoferryl species can also be generated using O_2 and additional reducing equivalents. Catalases are unique in that H_2O_2 functions as both donor and acceptor of electrons.

Cytochromes P-450 constitute a family of heme enzymes that comprise

Tab. 3 Representative human heme enzymes.

Enzyme	***Reaction Catalyzed***
Catalase	$2H_2O_2 \longrightarrow 2H_2O + O_2$
Cytochrome *c* oxidase[a]	$O_2 + 4H^+ + 4e^+ \longrightarrow 2H_2O$; "proton pumping"
Cytochrome *c* peroxidase	$H_2O_2 + 2H^+ + 2e^- \longrightarrow 2H_2O$
Cytochromes[b] P-450	$RH + 2e^- + O_2 + 2H^+ \longrightarrow ROH + H_2O$
Lactoperoxidase	$H_2O_2 + I^- \longrightarrow H_2O + IO^-$; others
Myeloperoxidase	$H_2O_2 + Cl^- \longrightarrow OCl^- + H_2O$
Nitric oxide synthase[c]	L-Arg $+ 2O_2 + 2H^+ + 3e^- \longrightarrow$ L-citrulline $+ 2H_2O + {}^{\bullet}NO$
Prostaglandin H synthase	Arachidonic acid $+ O_2 \longrightarrow PGH_2$
Thromboxane synthase	$PGH_2 \longrightarrow$ thromboxane A_2 + malondialdehyde + hydroxyheptadecatrienoic acid
Thyroid peroxidase	L-Tyr $+ I^- + H_2O_2 \longrightarrow H_2O + OH^-$ + monoiodotyrosine; others
Tryptophan oxygenase	L-Trp $+ O_2 \longrightarrow$ formylkynurinine

[a]Also contains active-site Cu_B, and noncatalytic metals (heme *a* and Cu_A redox centers, Na^+, Mg^{2+}, Zn^{2+}).
[b]A broad family of heme enzymes.
[c]Also uses a tetrahydrobiopterin as an active-site cofactor.

as many as 200 members in humans alone. Their roles in drug and steroid metabolism, as well as the activation of potential cancer-causing aromatic hydrocarbons, have led to their present status as the most intensely studied enzyme family. A particularly fascinating aspect of cytochrome P-450 biochemistry involves genetic (transcriptional) regulation – potential substrates can also serve as inducers *in vivo*.

Although the mammalian microsomal P-450's are of great clinical interest; they have proven difficult to purify and characterize in detail. Studies using a soluble cytochrome P-450 from *Pseudomonas putida* have led to the formulation of a detailed mechanism of P-450 enzyme action that remains the "benchmark" against which mechanistic features of other cytochromes P-450 are still compared. The availability of a crystal structure of the cytochrome P-450_{cam}–camphor complex affords an unusual opportunity to inspect the active site of a metalloenzyme–substrate complex (Fig. 3). What is evident from this structure is the clear origin of the regiospecific hydroxylation of the camphor substrate. The metal cofactor (heme) produces the oxidizing moiety, while the substrate-binding pocket restricts the orientation of the substrate vis-à-vis the heme iron and hence dictates *where* the substrate is hydroxylated. Recent studies on cytochrome P-450 mutants, having single mutations in the substrate-binding pocket, reinforce this view.

As indicated in Fig. 4, the ferric "resting" enzyme is inactive. Reduction of the heme to the ferrous state is required prior to O_2 binding; the necessary reducing equivalents taken up during enzyme turnover are supplied, via a series of electron carriers, by NADH. The key enzyme intermediate contains iron, in a formally 4+ oxidation state, and a porphyrin radical cation – this oxoferryl species oxidizes the bound substrate to form the

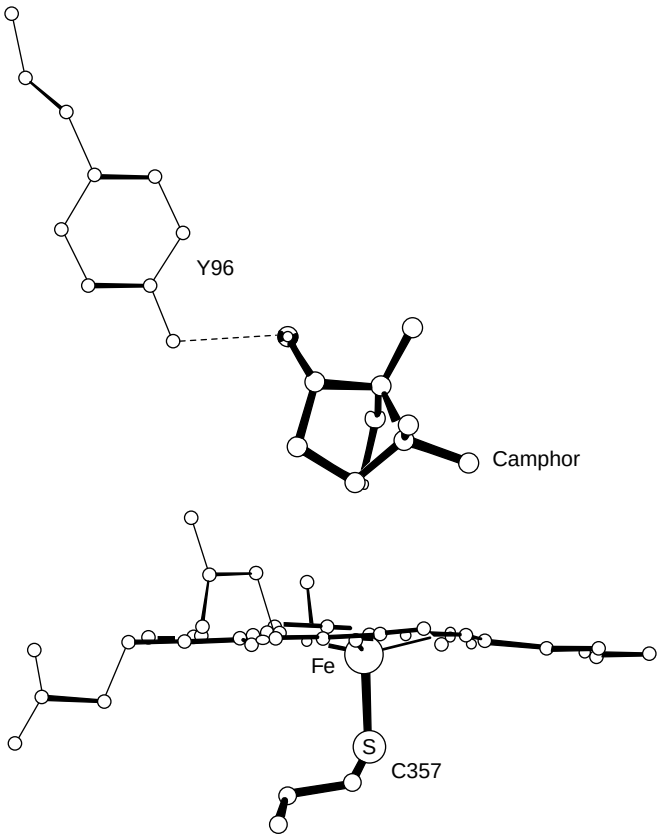

Fig. 3 Structure of the *Pseudomonas putida* cytochrome P-450$_{cam}$ active site. The substrate, camphor, is shown complexed to the enzyme, but does not coordinate to the iron.

corresponding alcohol product. The other oxygen atom from dioxygen dissociates as water; hence, all cytochromes P-450 are monooxygenases.

3.3 Mammalian Copper Enzymes

Copper ions have been found in dioxygen carriers, electron carriers, oxidases, superoxide dismutases, and oxygenases. Thus, copper and iron could be viewed as "cousins" insofar as their biological roles are concerned. Copper ions in biology evidently use just two oxidation states, 1+ and 2+; there is no compelling evidence for Cu^{3+} in any enzyme. Thus, catalysis of multielectron redox reactions requires more than one copper ion, or an additional redox unit – a tetrahydrobiopterin or 6-hydroxydopa quinone, for example.

Human copper enzymes (Table 4) are receiving increasing attention as a result of a growing awareness of their linkages to pathological conditions such as diabetes, cardiovascular disorders, and Alzheimer's disease. Serum amine oxidases catalyze the oxidative deamination of primary amines via a ping-pong mechanism composed of oxidative and reductive

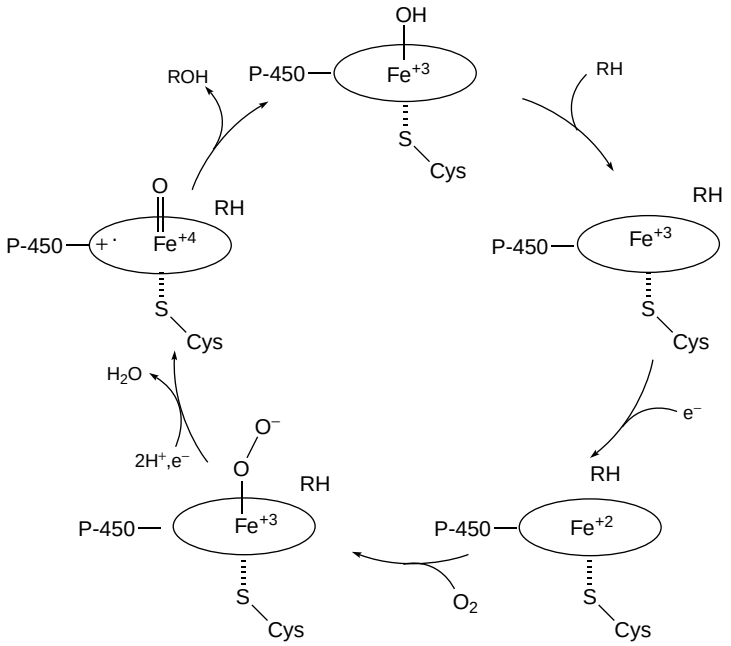

Fig. 4 Cytochrome P-450$_{cam}$ catalytic cycle. The substrate is designated as RH. Electrons are delivered to the heme in two discrete steps, the second of which is believed to limit the enzyme turnover rate.

Tab. 4 Representative human copper enzymes.

Enzyme	*Reaction Catalyzed*
Ceruloplasmin	Fe(II) $\longrightarrow$ Fe(III); other oxidations; oxidase activity
Dopamine β-hydroxylase	Dopamine $\longrightarrow$ norepinephrine
Hephaestin	Fe(II) $\longrightarrow$ Fe(III); oxidase activity
Peptidylglycine α-amidating monooxygenase	$\text{P-NH-C-COO}^- + O_2 + 2H^+ + 2e^- \longrightarrow \text{P-NH}_2 + \text{H(CO)COO}^- + H_2O$
Serum amine oxidase[b] B	$RCH_2NH_2^a + O_2 + H_2O \longrightarrow RCHO + H_2O_2 + NH_3$
Superoxide dismutase[c]	$2O_2^- + 2H^+ \longrightarrow O_2 + H_2O_2$

[a]Arylalkylamine substrates include serotonin and dopamine.
[b]Also uses topa quinone as a cofactor.
[c]The role of the active-site zinc is structural, rather than catalytic.

half-reactions (Eqs. 2a and 2b):

$$E_{ox} + RCH_2NH_2 \longrightarrow E_{red}\text{-}NH_2 + RCHO \quad (2a)$$

$$E_{red}\text{-}NH_2 + O_2 + H_2O \longrightarrow E_{ox} + NH_3 + H_2O_2 \quad (2b)$$

These enzymes are dimeric, having one copper ion and one covalently

bound cofactor, topa quinone (2,4,5-trihydroxyphenylalaninequinone) per monomer. Topa quinone is posttranslationally produced from a tyrosine residue during a multistep process involving the nearby copper ion and O_2. The Cu^{2+} active-site geometry in the resting enzyme is approximately square pyramidal, consisting of three histidine and two water ligands. In the reductive half-reaction (2a), a covalent bond (Schiff base complex) is formed between the substrate and topa quinone; chromophoric adducts of this type have been trapped and studied in detail. Hydrolysis of this adduct generates the aldehyde product and an E_{red}-NH_2, containing an aminoquinol form of topa quinone. Mechanistic details of the oxidative half-reaction (2b) are uncertain, and there is no unambiguous proof for redox cycling between Cu^{2+} and Cu^{1+} during enzyme turnover.

Also of intense interest is Cu,Zn superoxide dismutase (SOD), a catalyst for the disproportionation (Eq. 3) of the toxic superoxide anion:

$$2O_2^- + 2H^+ \longrightarrow O_2 + H_2O_2 \qquad (3)$$

X-ray crystal structures of the oxidized bovine erythrocyte enzyme reveal a binuclear active site: four histidine residues and a water molecule coordinate to the Cu^{2+}, while the Zn^{2+} is coordinated to three histidine and one aspartic acid. A bridging imidazolate, provided by His-61, holds the metal ions 6 Å apart. The role of the zinc appears to be structural; replacement of the Zn^{2+} by Co^{2+}, a redox-inactive spectroscopic probe, has been used to study active site changes that occur upon reduction of Cu^{2+} to Cu^{1+}. The copper bond to the imidazolate bridge is cleaved upon reduction, while the Co^{2+} bond to the bridge is retained. Subsequent X-ray crystal structures of the reduced native enzyme and mutants have confirmed this view. These observations have been incorporated into a catalytic mechanism that is illustrated in Fig. 5.

The accepted Cu,Zn SOD mechanism involves two phases: reduction of Cu^{2+} to Cu^{1+} (accompanied by liberation of O_2) and then oxidation of Cu^{1+} to Cu^{2+} (releasing hydrogen peroxide). During turnover, His-61 breaks and then reforms its bond to the active-site copper ion (there are no significant changes in Zn^{2+} geometry); these steps account for the ability of superoxide to act as both an oxidant and reductant during enzyme turnover. Situated on the rim of a deep channel, positively charged Arg-141 facilitates the catalysis by electrostatically steering and docking superoxide anions to the copper ion. Superoxide binding occurs at specific rates of ca. 2×10^9 $M^{-1}s^{-1}$, very near the limit of diffusion control.

3.4 Mammalian Zinc Enzymes

In 1939, Mann and Keilin discovered zinc in carbonic anhydrase, an enzyme whose function is to catalyze the hydration of carbon dioxide to the more soluble bicarbonate anion. Since that time, hundreds of zinc enzymes have been discovered; X-ray structures for dozens of these are now available. Table 5 lists some of the known human zinc enzymes, together with their activities. Divalent zinc plays a particularly conspicuous role in hydrolytic and group transfer biochemistry. Zinc is an excellent choice for such tasks because it is redox-inactive, easily exchanges ligands, and forms four-, five-, and six-coordinate complexes of comparable stability. Catalytic zinc sites always contain several endogenous (i.e. protein-derived) ligands and coordinated water.

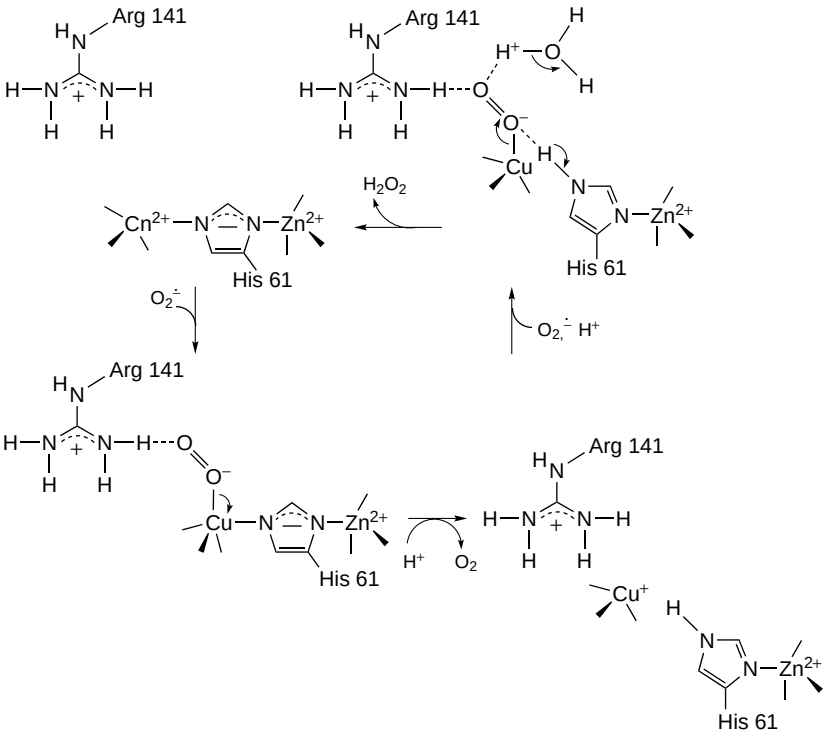

Fig. 5 Proposed mechanism of action of bovine erythrocyte Cu,Zn superoxide dismutase. The copper ion is involved in the catalysis, whereas the zinc plays a purely structural role.

Tab. 5 Representative human zinc enzymes.

Enzyme	*Reaction Catalyzed*
Alcohol dehydrogenase	$RR'CHOH + NAD^+ \longrightarrow RR'C{=}O + NADH + H^+$
Carbonic anhydrase	$CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$
Carboxypeptidases	C-terminal peptide bond hydrolysis
Glutaminyl cyclase	N-terminal Gln $\longrightarrow$ N-terminal pyroGlu + NH_3
Phospholipase C	Phospholipid hydrolysis
Ribonuclease A	RNA digestion (products contain 3′-phosphate termini)
RNA polymerase I	RNA biosynthesis (transcriptional elongation)
Stromelysin[a]	Peptide bond hydrolysis (ECM proteins)

[a]Also contains three noncatalytic Ca^{2+} ions.

The most important reactivity feature of divalent zinc is its Lewis acidity – it can polarize bound ligands and increase their susceptibility to attack by external nucleophiles or even increase the attacking power of a coordinated base. Water molecules coordinated to Zn^{2+} can deprotonate to form coordinated hydroxide

at neutral pH. Alternatively, displacement of the coordinated water by a substrate could lead to electrophilic catalysis.

Human carbonic anhydrase II catalyzes the *reversible* hydration of CO_2 in several distinct phases, where A represents a proton-transfer relay in which a proton is believed to reach His-64 before one transfers to the solvent:

$$EZNOH^- + CO_2 \longleftrightarrow EZnHCO_3^- \quad (4a)$$

$$EZnHCO_3^- + H_2O \longleftrightarrow EznH_2O + HCO_3^- \quad (4b)$$

$$EznH_2O + A \longleftrightarrow EznOH^- + AH^+ \quad (4c)$$

The kinetics of the carbonic anhydrase reaction increase with increasing pH, and display a sigmoidal pH-activity profile (apparent pK_a of ca. 7.0). It is now agreed that this reflects the deprotonation of coordinated water (by Thr-199) to produce a zinc-bound hydroxide ion, as indicated in Fig. 6. Thr-199 may also stabilize the transition state of the reaction through a hydrogen bond, and proline substitution at this position reduces catalytic efficiency by a factor of 3000. Further support for the proposed cycle in the figure comes from X-ray crystal structures of bicarbonate complexes of the Zn^{2+} and Co^{2+} enzymes. Bicarbonate is a monodentate ligand in both structures. However, water is also bound to the metal in the Co^{2+} enzyme–substrate complex, indicating that a five-coordinate metal intermediate in the native enzyme in solution is possible.

Carboxypeptidase A has also been the subject of intense mechanistic scrutiny. Glu-270 appears to abstract a proton from

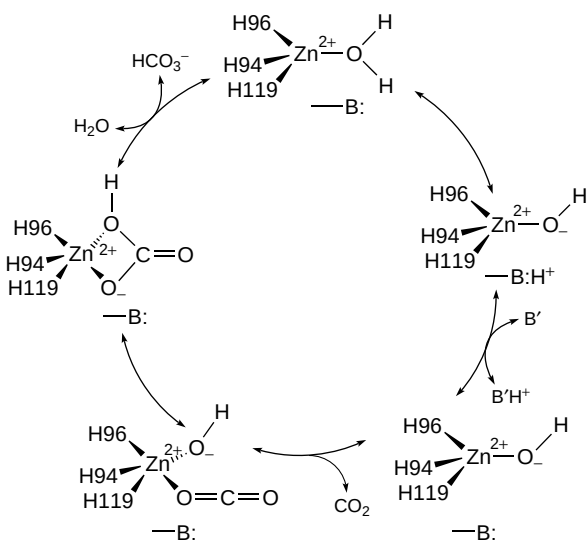

Fig. 6 Proposed mechanism for human carbonic anhydrase II. Basic side chains (B & B′) assist the zinc ion in forming coordinated hydroxide, which attacks coordinated CO_2 to form an anhydride intermediate. Residue B, likely Thr-199, performs a "gatekeeper" function.

the Zn^{2+}-coordinated water to produce the nucleophile needed to attack a backbone carbonyl of the substrate. A nearby side chain, Arg-127, activates the carbonyl group of the scissile peptide bond, facilitating the attack of coordinated hydroxide to produce a tetrahedral intermediate. This intermediate then coordinates to the zinc to produce a five-coordinate complex, which subsequently collapses to produce the products (two peptide fragments).

Many other zinc-catalyzed hydrolytic reactions are believed to proceed via the use of coordinated hydroxide. Matrix metalloproteases (MMPs) constitute a large family of mononuclear zinc endoproteases that are involved in the degradation of collagen, fibronectin, laminin, and other components of the extracellular matrix of connective tissue. They are important medicinal targets because uncontrolled MMP activity can lead to pathological conditions that include osteoarthritis, rheumatoid arthritis, and tumor invasion. Most MMPs share the consensus sequence HEXXHXXGXXH, in which the Zn^{2+} is coordinated to the three histidine residues. Substrates (which may include peptide loops of other enzyme molecules), inhibitors, and H_2O/OH^- molecules, can provide the fourth (or possibly a fifth) ligand.

Other human zinc enzymes contain more complex active sites involving more than one metal ion in the catalysis. This is thought to occur in renal dipeptidase-catalyzed reactions, for example. For this enzyme, a step involving binding of the dipeptide substrate by one of the two active-site zinc ions has been suggested on the basis of an X-ray structure, while the second zinc ion generates the coordinated hydroxide that catalyzes the hydrolytic reaction. Alkaline phosphatases contain two Zn^{2+} and one Mg^{2+} arranged in an active-site cluster whose molecular mechanism of action remains uncertain.

3.5 Mammalian Cytochrome *c* Oxidase

The redox chemistry catalyzed by heme enzymes includes respiratory dioxygen reduction. This reaction is catalyzed by cytochrome *c* oxidase, a complex membrane enzyme, consisting of 13 different polypeptides, that contains heme and copper centers at the dioxygen binding site, for the purpose of vectorially translocating protons across the inner mitochondrial membrane. The resulting proton gradient is then used by ATP synthase to catalyze ATP synthesis from ADP and phosphate. The bovine heart enzyme is the best studied of all of the cytochrome *c* oxidases, having been the subject of intense study in numerous laboratories for more than 70 years. For decades, there was a great deal of controversy concerning even the most basic information – the enzyme's metal content. Purification is extremely difficult, because of the multisubunit nature of this enzyme, and this has contributed to conflicting experimental results. During the last decade, Shinya Yoshikawa and coworkers have successfully crystallized the bovine heart enzyme and produced medium/low-resolution (2.3–2.9 Å) X-ray structures of four forms of the enzyme, including azide- and carbon monoxide-bound derivatives, while Hartmut Michel and coworkers produced X-ray structures of a bacterial cyrochrome *c* oxidase from *Paracoccus denitrificans*. These structures have clarified many controversial issues, and are guiding ongoing work on enzyme mutants and refinements of spectroscopic assignments.

Cytochrome *c* oxidase contains an unusual type of solvent-extractable heme cofactor, designated an *a*-type heme in contrast to the more familiar *b*-type heme found in hemoglobin and heme peroxidases (Fig. 7). Redox-active copper ions are also present, as well as sodium, zinc, and magnesium ions that appear to serve structural roles. Two of the redox cofactors, designated heme *a* and Cu_A, are believed to serve as electron-transfer agents, linking electron flow between the physiological reductant, cytochrome *c*, and the active site of the enzyme. Discrete cytochrome c^{red}-binding events are needed to transfer electrons to the fully oxidized enzyme, enabling it to catalyze the four-electron reduction of one O_2 molecule to two water molecules. Figure 8 displays a simplified view of the active site, which contains heme a_3 and Cu_B separated by ca. 5 Å. The X-ray structures revealed a surprising cross-link between Tyr-244 and His-240 that is of unknown origin, but is reminiscent of other cross-links found in copper enzymes like galactose oxidase. Another peculiar feature of the X-ray structures is the lack of an obvious channel for passage

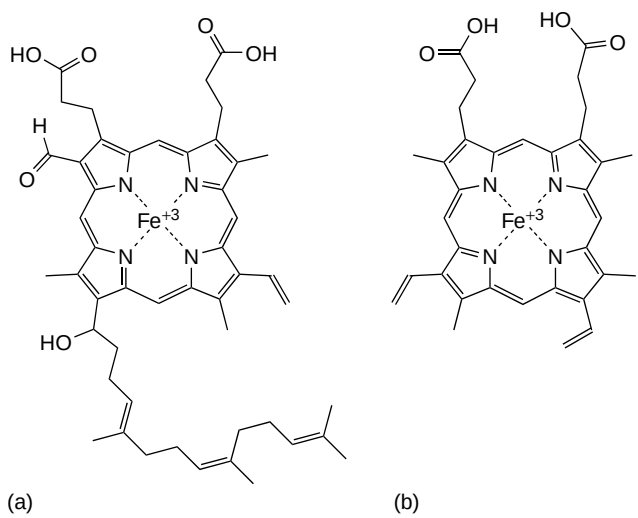

Fig. 7 Structures of active-site heme groups found in human metalloenzymes: (a) heme a_3 and (b) heme *b*.

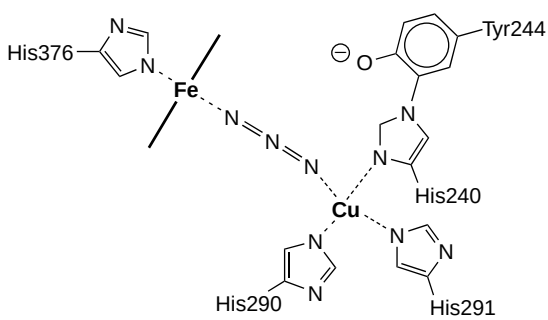

Fig. 8 Simplified cartoon of the binuclear active site of bovine cytochrome *c* oxidase, showing azide ion bound to both heme a_3 and Cu_B. Tyr-244 and His-240 are cross-linked, a result of a posttranslational self-processing event.

of O_2 to the binuclear active site during enzyme turnover, reinforcing the importance of protein fluctuations in the mechanism of action of cytochrome *c* oxidase.

While it is generally agreed that this heme a_3-Cu_B site is the active site for O_2 reduction, the details of this process remain unclear. Small molecules like azide and cyanide stabilize the active-site Fe(III)-Cu(II) oxidation states, locking the enzyme into a conformation that cannot react with O_2, and this is the molecular origin of the toxicities of these anions (the result is heart failure once the heart mitochondria begin to run out of ATP). CO, on the other hand, stabilizes the active-site Fe(II)-Cu(I) oxidation states, making another enzyme conformation amenable to study, and provides ready access to unstable intermediates when the CO adduct is photolyzed at low temperature in the presence of dissolved O_2. Resonance Raman spectroscopy has demonstrated that an early enzyme intermediate contains O_2 bound to reduced $Fe_{a3}{}^{2+}$. The X-ray structures also suggest that reduced Cu_B weakly interacts with bound O_2. Electron transfers from both metals to the bound O_2 would then produce a peroxo intermediate, long thought to be bridging Fe(III)-O-O-Cu(II). However, the X-ray structures raise the possibility of a hydrogen bond between protonated Tyr-244 and bound O_2, suggesting a peroxo intermediate that does not bind to Cu_B. Once a third electron is delivered to O_2, the O−O bond breaks, producing water (after protonation) and an oxoferryl heme a_3. Where does the third electron come from – the cross-linked Tyr-244, $Cu_B{}^{1+}$, or re-reduced heme a_3 (produced by fast electron transfer from heme *a*)? The fourth electron reduces the ferryl heme back to the ferric oxidation state, with the oxo group forming coordinated hydroxide that dissociates as water after protonation.

The mechanism of proton pumping by cytochrome *c* oxidase is even less clear because physical information, such as conserved amino acid residues or X-ray structures, cannot currently distinguish pumped protons from other dissociable protons. Hence, there has been much speculation regarding the proton pumping mechanism, particularly mechanistic proposals that link proton-translocation events to specific electron transfers.

3.6 Mammalian Enzymes Containing Less Common Metals

Much less is known about the active-site structural motifs and mechanisms of enzymes that contain metals other than iron, copper, or zinc. While most of these "unusual" enzymes (Table 6) are typically found in microbes and plants, there are known mammalian examples that play physiologically important roles.

Cobalt is found in vitamin B_{12}-dependent enzymes, and thus can play a role in radical reactions, as well as hydrolytic chemistry discussed earlier. There are two main roles for vitamin B_{12} (cyanocobalamin) in humans, both involving the formation of alkyl derivatives as enzyme intermediates. When cobalamin binds to its domain in nascent methionine synthase, it acquires an axial histidine ligand, which replaces benzimidazole, leaving a sixth coordination site available. The B_{12} domain of this complex enzyme must rapidly and efficiently interact with three other domains containing methyltetrahydrofolate, homocysteine, and *S*-adenosylmethionine, and methyl group transfer to form the methionine

Tab. 6 Representative human enzymes containing unusual metals.

Enzyme	*Cofactor*	*Reaction Catalyzed*
Arginase	2Mn	L-Arg + H_2O ⟶ L-ornithine + urea
Methionine aminopeptidase	2Mn	Hydrolysis of N-terminal Met residue of nascent proteins
Methylmalonyl-CoA mutase	1AdoCbl[a]	Methylmalonyl-CoA ⟶ succinyl-CoA
Secreted Phospholipases A_2	1Ca	Glycerophospholipid hydrolysis
Sulfite oxidase	1Mo-co, 1 heme/subunit	$SO_3^{2-} + H_2O \longrightarrow SO_4^{2-} + 2H^+ + 2e^-$
Superoxide dismutase	1Mn	$2O_2^- + 2H^+ \longrightarrow O_2 + H_2O_2$
Xanthine oxidase[b]	1Mo-co, $2Fe_2S_2$, 1FAD/subunit	Xanthine ⟶ uric acid

[a]5-Adenosylcobalamin.
[b]Can catalyze other reactions: O_2^- generation from O_2, and NO and nitrite production from nitrate during anoxia.

and tetrahydrofolate products is catalyzed by an intermediary methylcobalamin complex (Co^{3+}). During catalysis, the cobalt cycles between Co^{3+} and Co^{1+}. A minor Co^{2+} form can appear as a result of oxidative processes, which must be reduced back to Co^{1+} to reactivate the enzyme.

The 5-deoxyadenosyl adduct of cobalamin is the most abundant derivative of vitamin B_{12} in the liver, and is required for the catabolism of some branched-chain amino acids (using methylmalonyl-CoA mutase to generate succinyl-CoA from methylmalonyl-CoA). Deficiencies in this vitamin B_{12}, while rare, include neurologic disorders attributed to demyelination of nervous tissue. Succinyl-CoA formation occurs via an enzyme-catalyzed isomerization, rather than the methyl transfer process discussed above:

$$\text{CoA-SCOCH(CH}_3\text{)COOH} \longleftrightarrow \text{CoA-SCOCH}_2\text{CH}_2\text{COOH} \tag{5}$$

The CoA-SCO- is transferred intramolecularly and the hydrogen that is transferred does not exchange with water during enzyme turnover.

Three oxidation states (2+, 3+, and 4+) are commonly observed for manganese cofactors in biology, making them well suited for the catalysis of many of the reactions discussed in the section on iron enzymes. Human manganese superoxide dismutase (SOD) contains a manganese coordinated to three histidine, one aspartic acid, and one water molecule in a trigonal bipyramidal geometry. Catalysis is thought to proceed via the following scheme:

$$\text{Mn(III)SOD} + O_2^- + H^+ \longrightarrow \text{Mn(II)SOD(H}^+\text{)} + O_2 \tag{6a}$$

$$\text{Mn(II)SOD(H}^+\text{)} + O_2^- + H^+ \longrightarrow \text{Mn(III)SOD} + H_2O_2 \tag{6b}$$

$$\text{Mn(II)SOD(H}^+\text{)} + O_2^- \longrightarrow \text{Mn(III)}(O_2^{2-})\text{SOD(H}^+\text{)} \tag{6c}$$

$$\text{Mn(III)}(O_2^{2-})\text{SOD(H}^+\text{)} + H^+ \longrightarrow \text{Mn(III)SOD} + H_2O_2 \tag{6d}$$

Steps (6c) and (6d) represent the formation and dissociation of a product-inhibited complex. Gln-143 is hydrogen-bonded to the coordinated water and this residue that is essential for the high activity of the enzyme (minimizing steps (6c) and (6d), which involve a presumed, intermediary peroxo complex).

Liver arginase I contains a binuclear manganese center (ca. 3.3 Å Mn–Mn distance), catalyzing the hydrolysis of L-arginine to form L-ornithine and urea. This reaction is the final cytosolic step of the urea cycle. It is generally agreed that both Mn^{2+} ions orient and polarize a bridging hydroxide, which attacks the guanidinium carbon of the arginine substrate (held in place by Asp-277 and His-141). An isozyme, arginase II, is found in nonhepatic tissues (e.g. gastrointestinal smooth muscle, penile corpus cavernosum smooth muscle), where it facilitates the regulation of NO biosynthesis by attenuating the levels of L-arginine available to nitric oxide synthase (a heme enzyme):

$$\text{Arginase I/II : L-arginine} \longrightarrow \text{L-ornithine} + \text{urea} \quad (7)$$

$$\text{NO synthase : L-arginine} \longrightarrow \text{L-citrulline} + \text{NO} \quad (8)$$

The latter process proceeds via an N^{ω}-hydroxy-L-arginine intermediate, which is itself a modest, competitive inhibitor of arginase.

Molybdenum is a moderately important element – it is found in nitrogenase, a variety of bacterial enzymes, and several mammalian enzymes (sulfite oxidase, aldehyde oxidase, and xanthine oxidase). Molybdoenzymes are large and typically contain other cofactors as well. Rather surprisingly, since oxides (e.g. MoO_4^{2-}) are found in aqueous solutions, the molybdenum centers are usually in sulfur-rich environments (see Fig. 1). Most molybdoenzymes hydroxylate their substrates, yet are not classified as oxygenases because they transfer oxo ligands that are *not* derived from O_2.

Mitochondrial sulfite oxidase is responsible for the oxidation of sulfite (SO_3^{2-}) to sulfate (SO_4^{2-}) and has been studied in detail. This enzyme catalyzes the final reaction in the catabolism of sulfur-containing amino acids. Each enzyme monomer has domains for the molybdenum-pterin cofactor (Mo-co), bound to the polypeptide via a cysteine ligand, and for a cytochrome b_5 (an electron-transfer center, represented as iron in the scheme below). The overall reaction can be viewed in terms of the following events:

$$\text{Mo(VI),Fe(II)} + \text{cyt } c^{\text{ox}} \longrightarrow \text{Mo(VI),Fe(III)} + \text{cyt } c^{\text{red}} \quad (9a)$$

$$\text{Mo(VI),Fe(III)} + SO_3^{2-} \longrightarrow \text{Mo(IV),Fe(III)} + SO_4^{2-} + 2H^+ \quad (9b)$$

$$\text{Mo(IV),Fe(III)} \longrightarrow \text{Mo(V),Fe(II)} \quad (9c)$$

$$\text{Mo(V),Fe(II)} + \text{cyt } c^{\text{ox}} \longrightarrow \text{Mo(V),Fe(III)} + \text{cyt } c^{\text{red}} \quad (9d)$$

$$\text{Mo(V),Fe(III)} \longrightarrow \text{Mo(VI),Fe(II)} \quad (9e)$$

In step (9b), sulfite reacts with an oxo ligand coordinated to Mo(VI), forming sulfate coordinated to Mo(IV), and hydroxide (derived from water) coordinates to the molybdenum site after departure of sulfate. The oxo ligand (deprotonated hydroxide) forms in response to formation of Mo^{6+}.

3.7 Bacterial Metalloenzymes and Their Roles in Infection

Understanding molecular mechanisms underlying microbial virulence is critically important to the development of effective antimicrobial compounds and remedies for intoxication (e.g. botulism), particularly caused by food- and water-borne bacteria. Bacteria that must survive the acidic gastric environment (e.g. *Helicobacter pylori, Klebsiella aerogenes*) produce enzymes that provide a source of nitrogen for growth and help neutralize the pH immediately surrounding the cell. *Klebsiella aerogenes* urease has been the most intensively studied of these, and contains two Ni^{2+} ions in the enzyme active site, separated by ca. 3.6 Å. Assembly of the binuclear Ni^{2+} active site is complex, requiring a variety of accessory proteins (including at least one nickel chaperone, UreE) *in vivo*. These ions are bridged by a lysine residue that has been carbamylated as a result of a posttranslational self-processing reaction with CO_2. *In vitro* reconstitution of the apoenzyme in the absence of CO_2 does not generate any urease activity. The accepted mechanism comprises a step involving urea binding to Ni-1 (carbonyl oxygen coordination), and hydrogen bonding to His-320. Hydroxide coordinated to Ni-2 attacks the carbonyl carbon to form a tetrahedral intermediate. His-320 transfers a proton to an amide nitrogen of the urea intermediate, which then collapses into ammonia and carbamate (H_2NCOOH). The carbamate product is itself unstable, and spontaneously decomposes to form CO_2 and a second ammonia molecule. The net result of the urease-catalyzed and uncatalyzed reactions is the following:

$$CO(NH_2)_2 + H_2O \longrightarrow 2NH_3 + CO_2 \quad (10)$$

One host defense strategy involves the production of IgA molecules that bind to both bacterial surface antigens and mucin, trapping the invading bacterial cells in the mucin. Some strains of pathogenic bacteria (particularly streptococci) secrete a variety of zinc metalloproteases, including enzymes (IgA1 proteases) that cleave human IgA1 in the hinge region. These proteases all possess one Zn^{2+} ion coordinated to two histidine and one glutamic acid

Tab. 7 Representative bacterial metalloenzymes of medical importance.

Enzyme	*Cofactor*	*Reaction Catalyzed*
Botulinum toxin	1Zn	Peptide bond hydrolysis (long substrates[a], specific cleavages)
Lethal factor[b]	1Zn	Peptide bond hydrolysis (mitogen-activated protein kinases)
Phospholipase[c] C	3Zn	Hydrolysis of specific phospholipids
Pneumococcal IgA1 protease	1Zn	Peptide bond hydrolysis (hinge region of human IgA1)
Tetanus toxin	1Zn	Peptide bond hydrolysis (VAMP/synaptobrevin)
Urease	2Ni/subunit	$CO(NH_2)_2 + H_2O \longrightarrow NH_3 + H_2NCOOH$

[a] SNAP-25, VAMP (synaptobrevin), syntaxin.
[b] *Bacillus anthracis.*
[c] *Bacillus cereus.*

residue, with the remaining site occupied by hydroxide.

The causative agent of anthrax, *Bacillus anthracis*, produces a variety of secreted toxins, one of which (lethal toxin) is the primary cause of death in infected animals and is considered a key virulence factor. Lethal toxin consists of protective antigen and lethal factor, which is a zinc metalloprotease that cleaves members of themitogen-activated protein-kinase–kinase (MAPKK) family near their amino-termini. The X-ray structure of lethal factor shows that the Zn^{2+} is coordinated to His-686, His-690, Glu-735 and a water/hydroxide molecule in an approximately tetrahedral geometry. Glu-687 is positioned ca. 3.5 Å from the coordinated water to act as a proton acceptor and facilitate the formation of coordinated hydroxide. The active site of lethal factor is contiguous with a 40-Å long groove that serves as the substrate-binding pocket. The specificity for MAPKK substrates arises in primarily from this binding site, rather than the zinc site.

Botulism and tetanus are caused by Gram-positive, spore-forming bacteria (*Clostridium botulinum* and *Clostridium tetani* respectively) that require near-anaerobic conditions for growth. These organisms secrete dangerous toxins that are frequently fatal (death results from intoxication rather than infection *per se*). The toxins responsible for the symptoms are zinc metalloproteases that cleave proteins that are associated with neurotransmitter release. Strains of toxigenic *C. botulinum* produce a variety of antigenically distinct botulinum toxins, designated A, B, C1, D, E, F, and G. Botulinum toxins A, E, and C1 cleave synaptosomal-associated protein (SNAP-25) at different sites, the others (and tetanus toxin) cleave vessicle-associated membrane protein (VAMP, also known as synaptobrevin), and C1 additionally cleaves syntaxin. Types A, B, and E are the principal causes of botulism in humans. The catalytic domains contain the usual [HExxH + E] motif displayed by many other zinc endoproteases. Mature botulinum toxins are very large, composed of a heavy chain (ca. 100 kDa) and a light chain (ca. 50 kDa) that are held together by a disulfide bond. The heavy chain binds to the presynaptic membrane and helps to translocate the light chain across the membrane into the cytosol, where the light chain cleaves a VAMP/SNAP-25 target protein. Botulinum toxins, particularly type A toxins, are considered the most potent toxins known. Once a botulinum toxin A molecule enters the neuronal cytosol, it is estimated that a single enzyme molecule is sufficient to intoxicate a synapse.

The zinc active site is deeply buried in botulinum toxins, and is accessible to the solvent via a ca. 12 Å × 15 Å × 35 Å channel. The clostridial neurotoxins are atypical among zinc metalloproteases because of their stringent substrate specificities (substrates must be at least 16 amino acids in length, and cleavage is site-specific). The catalytic mechanism of the all of these toxins appears to involve formation of zinc-coordinated hydroxide, polarization of a carbonyl (amide) group of the substrate by a nearby amino acid residue, hydrogen bonding of an additional nearby residue to the amide nitrogen, and collapse to form the product peptides. Botulinum toxin A has been shown to form a dimer in aqueous solution at concentrations above 75 nM (the enzyme is a monomer in crystals studied to date). Endopeptidase activity is inhibited by 75% when the dimer

forms, suggesting that substrate access to the channel is sterically restricted in the dimer.

4 Perspectives

In addition to the metals mentioned in this article, roles for other metals as enzyme cofactors will undoubtedly be elucidated in the near future. One fascinating, and perplexing, feature of biological systems concerns the rationale for the selection of a particular metal cofactor for a given biological task. Consider the superoxide dismutases. Four different kinds of enzyme active sites have been discovered in various prokaryotes and eukaryotes, containing either mononuclear iron, mononuclear manganese, a binuclear Cu,Zn dimer, or mononuclear nickel (in some bacteria). Why do all forms of life not use just one type of cofactor? Why are specialized kinds of human superoxide dismutases needed (an intracellular manganese enzyme, an intracellular Cu,Zn enzyme, and an extracellular Cu,Zn enzyme)?

As noted earlier in this article, many redox enzymes utilize O_2 or H_2O_2 as the source of oxidizing equivalents needed to hydroxylate a substrate. In particular, iron-peroxo or iron-oxo intermediates have been proposed for most iron enzymes discussed in this chapter. What keeps these kinds of enzymes from inactivating (self-destructing) during enzymatic turnover, a result of repeated oxidation of nearby amino acid residues or production of radicals that tunnel and carry out redox transformations elsewhere in the enzyme? Both hemoglobin and myoglobin function as O_2 carriers, but can also react with H_2O_2 and organic substrates in peroxidase-like fashion. Unlike peroxidases, the globins rapidly lose enzyme activity during turnover because their Compound I intermediates are unstable, ultimately destroying the heme chromophores, and forming protein oligomers. What are the structural constraints for a good heme peroxidase? Similar stability problems are posed by other broad classes of redox enzymes.

New, clinically important, roles for small-molecule interactions with metalloenzymes continue to emerge. For example, the gases NO and CO, both vertebrate hormones, are generated by metalloenzymes. Furthermore, some target cell responses to these gases are the result of binding to other metalloenzymes (e.g. NO binding to the heme enzyme guanylyl cyclase causes a large conformational change that results in a ca. 300-fold increase in activity). Many therapeutic drugs target metalloenzymes. Prominent examples include nonsteroidal anti-inflammatory drugs, such as aspirin or indomethacin. These drugs block prostaglandin production by inhibiting a heme enzyme, prostaglandin H synthase.

Many clinically important roles for enzyme oligomerization and folding pathways involving metalloproteins are only beginning to be understood. For example, mutations in the structural gene SOD1, encoding a Cu,Zn superoxide dismutase, figure prominently in current research on the molecular origins of amylotropic lateral sclerosis (ALS). Some mutant enzymes cannot tightly bind active-site metal ions, their apo-forms are partially unfolded, and are hence more likely to form insoluble aggregates.

Finally, it is important to note that copper ions have been shown to play key roles in posttranslational self-processing,

leading to the production of new, covalently bound redox centers (thus far only known to be derived from endogenous tyrosine residues in the active site). Additional cases of this phenomenon are to be expected, with important mechanistic implications.

See also Bioinorganic Chemistry; Heme Enzymes; Hemoglobin.

Bibliography

Books and Reviews

Bertini, I., Gray, H.B., Lippard, S.J., Valentine, J.S., (Eds.) (1994) *Bioinorganic Chemistry*, University Science Press, Mill Valley, CA.

Crichton, R.R. (Ed.) (2001) *Inorganic Biochemistry of Iron Metabolism: from Molecular Mechanisms to Clinical Consequences*, 2nd edition, Wiley, New York.

da Silva, J.J.R.F., Williams, R.J.P. (2001) *The Biological Chemistry of the Elements*, 2nd edition, Oxford University Press, New York.

Dunford, H.B. (1999) *Heme Peroxidases*, Wiley-VCH, New York.

Eichhorn, G., Marzilli, L., Series (Eds.) (1979–1996) *Advances in Inorganic Biochemistry*, Vols. 1–11, Elsevier/Prentice Hall/VCH, New York.

Hooper, N.M. (Ed.) (1996) *Zinc Metalloproteases in Health and Disease*, Taylor & France, Bristol, PA.

Messerschmidt, A., Bode, M., Cygler, W. (Eds.) (2004) *Handbook of Metalloproteins*, Vol. 3, Wiley, New York.

Messerschmidt, A., Huber, R., Wieghardt, K., Poulos, T., (Eds.) (2001) *Handbook on Metalloproteins*, Vols. 1–2, Wiley, New York.

Ortiz de Montellano, P.R. (Ed.) (1995) *Cytochrome P-450: Structure, Mechanism, and Biochemistry*, 2nd edition, Plenum, New York.

Reedijk, J., Bouwman, E., (Eds.) (1999) *Bioinorganic Catalysis*, 2nd Edition, Dekker, New York.

Sigel, H., Sigel, A., Series (Eds.), (1974–2004) *Metal Ions Biological Systems*, Vols. 1–42, Dekker, New York.

Primary Literature

Andersen, O.A., Stokka, A.J., Flatmark, T., Hough, E. (2003) 2.0 Å resolution crystal structures of the ternary complexes of human phenylalanine hydroxylase catalytic domain with tetrahydrobiopterin and 3-(2-thienyl)-L-alanine or L-norleucine: substrate specificity and molecular motions related to substrate binding, *J. Mol. Biol.* **333**, 747–757.

Arnesano, F., Banci, L., Bertini, I., Ciofi-Baffoni, S., Molteni, E., Huffman, D.L., O'Halloran, T.V. (2002) Metallochaperones and metal-transporting ATPases: A comparative analysis of sequences and structures, *Genome Res.* **12**, 255–171.

Banerjee, R., Ragsdale, S.W. (2003) The many faces of vitamin B_{12}: catalysis by cobalamin-dependent enzymes, *Annu. Rev. Biochem.* **72**, 209–247.

Bartnikas, T.B., Gitlin, J.D. (2003) Mechanisms of biosynthesis of copper/zinc superoxide dismutase, *J. Biol. Chem.* **278**, 33602–33608.

Berry, C.E., Hare, J.M. (2004) Xanthine oxidoreductase and cardiovascular disease: molecular mechanisms and pathophysiological implications, *J. Physiol.* **555.3**, 589–606.

Borbulevych, O.Y., Jankun, J., Selman, S.H., Skrzypczak-Jankun, E. (2004) Lipoxygenase interactions with natural flavonoid, quercetin, reveal a complex with protocatechuic acid in its X-ray structure at 2.1 Å resolution, *Proteins* **54**, 13–19.

Bratton, M., Mills, D., Castleden, C.K., Hosler, J., Meunier, B. (2003) Disease-related mutations in cytochrome *c* oxidase studied in yeast and bacterial models, *Eur. J. Biochem.* **270**, 1222–1230.

Cai, S., Singh, B.R. (2001) A correlation between differential structural features and the degree of endopeptidase activity of type A botulinum neurotoxin in aqueous solution, *Biochemistry* **40**, 4693–4702.

Cama, E., Colleluori, D.M., Emig, F.A., Shin, H., Kim, S.W., Kim, N.N., Traish, A.M., Ash, D.E., Christianson, D.W. (2003) Human arginase II: crystal structure and physiological role in male and female sexual arousal, *Biochemistry* **42**, 8445–8451.

Cama, E., Pethe, S., Boucher, J.-L., Han, S., Emig, F.A., Ash, D.E., Viola, R.E., Mansuy, D.,

Christianson, D.W. (2004) Inhibitor coordination interactions in the binuclear manganese cluster of arginase, *Biochemistry* **43**, 8987–8999.

Choi, E.-Y., Stockert, A.L., Leimkühler, S., Hille, R. (2004) Studies on the mechanism of action of xanthine oxidase, *J. Inorg. Biochem.* **98**, 841–848.

Chopra, A.P., Boone, S.A., Liang, X., Duesbery, N.S. (2003) Anthrax lethal factor proteolysis and inactivation of MAPK kinase, *J. Biol. Chem.* **278**, 9402–9406.

Cioni, P., Pesce, A., Morozzo della Rocca, B., Castelli, S., Falconi, M., Parrilli, L., Bolognesi, M., Strambini, G., Desidero, A. (2003) Active-site copper and zinc ions modulate the quaternary structure of prokaryotic Cu,Zn superoxide dismutase, *J. Mol. Biol.* **326**, 1351–1360.

Collins, C.M., D'Orazio, S.E. (1993) Bacterial ureases: structure, regulation of expression and role in pathogenesis, *Mol. Microbiol.* **9**, 907–913.

Cooper, C.E. (1999) Nitric oxide and iron proteins, *Biochim. Biophys. Acta* **1411**, 290–309.

Costas, M., Mehn, M.P., Jensen, M.P., Que, Jr., L. (2004) Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates, *Chem. Rev.* **104**, 939–986.

DeMatteis, G., Agostinelli, E., Mondovì, B., Morpurgo, L. (1999) The metal function in the reactions of bovine serum amine oxidase with substrates and hydrazine inhibitors, *J. Biol. Inorg. Chem.* **4**, 348–353.

Duda, D., Govindasamy, L., Agbandje-McKenna, M., Tu, C., Silverman, D.N., McKenna, R. (2003) The refined atomic structure of carbonic anhydrase II at 1.05 Å resolution: implications of chemical rescue of proton transfer, *Acta Cryst.* **D59**, 93–104.

Elam, J.S., Thomas, S.T., Holloway, S.P., Taylor, A.B., Hart, P.J. (2002) Copper chaperones, *Adv. Protein Chem.* **60**, 151–219.

Erlandsen, H., Patch, M.G., Gamez, A., Straub, M., Stevens, R.C. (2003) Structural studies on phenylalanine hydroxylase and implications toward understanding and treating phenylketonuria, *Pediatrics* **112**, 1557–1565.

Eswaramoorthy, S., Kumaran, D., Keller, J., Swaminathan, S. (2004) Role of metals in the biological activity of *Clostridium botulinum* neurotoxins, *Biochemistry* **43**, 2209–2216.

Fei, M.J., Yamashita, E., Inoue, N., Yao, M., Yamaguchi, H., Tsukihara, T., Shinazawa-Itoh, K., Nakashima, R., Yoshikawa, S. (2000) X-ray structure of azide-bound fully oxidized cytochrome *c* oxidase from bovine heart at 2.9 Å resolution, *Acta Cryst.* **D56**, 529–535.

Garrett, R.M., Johnson, J., Graf, T.N., Feigenbaum, A., Rajagopalan, K.V. (1998) Human sulfite oxidase R160Q: identification of the mutation in a sulfite oxidase-deficient patient and expression and characterization of the mutant enzyme, *Proc. Natl. Acad. Sci. USA* **95**, 6394–6398.

Hamza, I., Gitlin, J.D. (2002) Copper chaperones for cytochrome *c* oxidase and human disease, *J. Bioenerget. Biomembr.* **34**, 381–388.

Hanson, M.A., Stevens, R.C. (2000) Cocrystal structure of synaptobrevin–II bound to botulinum neurotoxin type B at 2.0 Å resolution, *Nat. Struct. Biol.* **7**, 687–692.

Jabri, E., Karplus, P.A. (1996) Structures of the *Klebsiella aerogenes* Urease Apoenzyme and two active-site mutants, *Biochemistry* **35**, 10616–10626.

Linzie, S.D., Thevis, M., Ngo, K., Whitelegge, J., Loo., J.A., Abu- Omar, M.M. (2003) Posttranslational hydroxylation of human phenylalanine hydroxylase is a novel example of enzyme self-repair within the second coordination sphere of catalytic iron, *J. Amer. Chem. Soc.* **125**, 4710–4711.

Kisker, C., Schindelin, H., Rees, D.C. (1997) Molybdenum-cofactor-containing enzymes: structure and mechanism, *Annu. Rev. Biochem.* **66**, 233–267.

Kisker, C., Schindelin, H., Pacheco, A., Wehbi, W.A., Garrett, R.M., Rajagopalan, K.V., Enemark, J.H., Rees, D.C. (1997) Molecular basis of sulfite oxidase deficiency from the structure of sulfite oxidase, *Cell* **91**, 973–983.

Kolberg, M., Strand, K.R., Graff, P., Andersson, K.K. (2004) Structure, function, and mechanism of ribonucleotide reductases, *Biochim. Biophys. Acta* **1699**, 1–34.

Kuchar, J., Hausinger, R.P. (2004) Biosynthesis of metal sites, *Chem. Rev.* **104**, 509–525.

Kuwabara, Y., Nishino, T., Okamoto, K., Matsumura, T., Eger, B.T., Pai, E.F., Nishino, T. (2003) Unique amino acids cluster for switching from the dehydrogenase to oxidase form of xanthine oxidoreductase, *Proc. Natl. Acad. Sci. USA* **100**, 8170–8175.

Lalli, G., Bohnert, S., Deinhardt, K., Verastegui, C., Schiavo, G. (2003) The journey of tetanus and botulinum neurotoxins in neurons, *Trends Microbiol.* **11**, 431–437.

Lauble, H., Kennedy, M.C., Emptage, M.H., Beinert, H., Stout, C.D. (1996) The reaction of fluorocitrate with aconitase and the crystal structure of the enzyme-inhibitor complex, *Proc. Natl. Acad. Sci. USA* **93**, 13699–13703.

Lecerof, D., Fodje, M., Hansson, A., Hansson, M., Al-Karadaghi, S. (2000) Structural and mechanistic basis of porphyrin metallation by ferrochelatase, *J. Mol. Biol.* **297**, 221–232.

Maaß, A., Scholz, J., Moser, A. (2003) Modeled ligand-protein complexes elucidate the origin of substrate specificity and provide insight into catalytic mechanisms of phenylalanine hydroxylase and tyrosine hydroxylase, *Eur. J. Biochem.* **270**, 1065–1075.

Maksos, K., Bode, W. (2003) Structural basis of matrix metalloproteinases and tissue inhibitors of metalloproteinases, *Mol. Biotechnol.* **25**, 241–266.

Martelin, E., Lapatto, R., Raivio, K.O. (2002) Regulation of xanthine oxidoreductase by intracellular iron, *Am. J. Physiol. Cell Physiol.* **283**, C1722–C1728.

Montecucco, C., Schiavo, G. (1995) Structure and function of tetanus and botulinum neurotoxins, *Q. Rev. Biophys.* **4**, 423–472.

Moriwaki, Y., Yamamoto, T., Higashino, K. (1997) Distribution and pathophysiologic role of molybdenum-containing enzymes, *Histol. Histopathol.* **12**, 513–524.

Nitanai, Y., Satow, Y., Adachi, H., Tsujimoto, M. (2002) Crystal structure of human renal dipeptidase involved in β-lactam hydrolysis, *J. Mol. Biol.* **321**, 177–184.

Noor, R., Mittal, S., Iqbal, J. (2002) Superoxide dismutase – applications and relevance to human diseases, *Med. Sci. Monit.* **8**, RA10–RA15.

Offioni, M.R., Memmi, G., Maggi, T., Chiavolini, D., Iannelli, F., Pozzi, G. (2003) Pneumococcal zinc metalloprotease zmpc cleaves human matrix metalloproteinase 9 and is a virulence factor in experimental pneumonia, *Mol. Microbiol.* **49**, 795–805.

Oliver, C.F., Modi, S., Sutcliffe, M.J., Primrose, W.U., Lian, L.-Y., Roberts, G.C.K. (1997) A single mutation in cytochrome P450 BM3 changes substrate orientation in a catalytic intermediate and the regiospecificity of hydroxylation, *Biochemistry* **36**, 1567–1572.

Pan, Y.H., Yu, B.-Z., Singer, A.G., Ghomashchi, F., Lambeau, G., Gelb, M.H., Jain, M.K., Bahnson, B.J. (2002) Crystal structure of human group X secreted phospholipase A_2, *J. Biol. Chem.* **277**, 29086–29093.

Pannifer, A.D., Wong, T.Y., Schwarzenbacher, R., Renatus, M., Petosa, C., Bienkowska, J., Lacy, D.B., Collier, R.J., Park, S., Leppla, R.S., Hanna, P., Liddington, R.C. (2001) Crystal structure of the anthrax lethal factor, *Nature* **414**, 229–233.

Pearson, M.A., Park, I.-S., Schaller, R.A., Michel, L.O., Karplus, P.A., Hausinger, R.P. (2000) Kinetic and structural characterization of urease active site variants, *Biochemistry* **39**, 8575–8584.

Petersen, S.V., Oury, T.D., Valnickova, S., Thøgersen, I.B., Crapo, J.D., Enghild, J.J. (2003) The dual nature of human extracellular superoxide dismutase: one sequence and two structures, *Proc. Natl. Acad. Sci. USA* **100**, 13875–13880.

Pinakoulaki, E., Pfitzner, U., Ludwig, B., Varotsis, C. (2002) The role of the cross-link His-Tyr in the functional properties of the binuclear center in cytochrome *c* oxidase, *J. Biol. Chem.* **277**, 13563–13568.

Pylypenko, O., Schlichting, I. (2004) Structural aspects of ligand binding to and electron transfer in bacterial and fungal P450s, *Annu. Rev. Biochem.* **73**, 991–1018.

Rosetto, O., Caccin, P., Rigoni, M., Tonello, F., Bortoletto, N., Stevens, R.C., Montecucco, C. (2001) Active-site mutagenesis of tetanus neurotoxin implicates TYR-375 and Glu-271 in metalloproteolytic activity, *Toxicon* **39**, 1151–1159.

Segelke, B., Knapp, M., Kadkhodayan, S., Balhorn, R., Rupp, B. (2004) Crystal structure of *Clostridium botulinum* neurotoxin protease in a product-bound state: evidence for noncanonical zinc protease Activity, *Proc. Natl. Acad. Sci. USA* **101**, 6888–6893.

Skrzypczak-Jankun, E., Bross, R.A., Carroll, R.T., Dunham, W.R., Funk, Jr, M.O. (2001) Three-dimensional structure of a purple lipoxygenase, *J. Amer. Chem. Soc.* **123**, 10814–10820.

Song, H.K., Mulrooney, S.B., Huber, R., Hausinger, R.P. (2001) Crystal structure of *Klebsiella aerogenes* UreE, a nickel-binding metallochaperone for urease activation, *J. Biol. Chem.* **276**, 49359–49364.

Su, Q., Klinman, J.P. (1998) Probing the mechanism of proton coupled electron

transfer to dioxygen: the oxidative half-reaction of bovine serum amine oxidase, *Biochemistry* **37**, 12413–12525.

Swaminathan, S., Eswaramoorthy, S. (2000) Structural analysis of the catalytic and binding sites of *Clostridium botulinum* neurotoxin B, *Nat. Struct. Biol.* **7**, 693–700.

Thunnissen, M.M.G.M., Nordlund, P., Haeggström, J.Z. (2001) Crystal structure of human leukotriene A_4 hydrolase, a bifunctional enzyme in inflammation, *Nat. Struct. Biol.* **8**, 131–135.

Valentine, J.S., Hart, J.P. (2003) Misfolded CuZnSOD and amylotrophic lateral sclerosis, *Proc. Natl. Acad. Sci. USA* **100**, 3617–3622.

Vallee, B.L., Auld, D.S. (1993) New perspectives on zinc biochemistry: cocatalytic sites in multizinc enzymes, *Biochemistry* **32**, 6493–6500.

Wang, J., Sheppard, G.S., Lou, P., Kawai, M., Park, C., Egan, D.A., Schneider, A., Bouska, J., Lesniewski, R., Henkin, J. (2003) Physiologically relevant metal cofactor for methionine aminopeptidase-2 is manganese, *Biochemistry* **42**, 5035–5042.

Wang, L., Erlandsen, H., Haavik, J., Knappskog, P.M., Stevens, R.C. (2002) Three-dimensional structure of human tryptophan hydroxylase and its implications for the biosynthesis of the neurotransmitters serotonin and melatonin, *Biochemistry* **41**, 12569–12574.

Williams, P.A., Cosme, J., Vinković, D.M., Ward, A., Angove, H.C., Day, P.J., Vonrhein, C., Tickle, I.J., Jhoti, H. (2004) Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone, *Science* **305**, 683–686.

Wilmot, C.M. (2003) Oxygen activation in a copper-containing amine oxidase, *Biochem. Soc. Trans.* **31**, 493–496.

Wilmot, C.M., Hadju, J., McPherson, M.J., Knowles, P.F., Phillips, S.E.V. (1999) Visualization of dioxygen bound to copper during enzyme catalysis, *Science* **286**, 1724–1728.

Wilson, H.L., Rajagopalan, K.V. (2004) The role of tyrosine 343 in substrate binding and catalysis by human sulfite oxidase, *J. Biol. Chem.* **279**, 15105–15113.

Yoshikawa, S. (1999) X-ray structure and reaction mechanism of bovine heart cytochrome *c* oxidase, *Biochem. Soc. Trans.* **27**, 351–362.

Methanogens and the Archaebacteria, Molecular Biology of

Madeline E. Rasche[1] *and James G. Ferry*[2]
[1] *University of Florida, Gainesville, FL 32611, USA*
[2] *The Pennsylvania State University, University Park, PA 16801, USA*

1 Introduction 239

2 Biochemistry of Methanogenesis 239
2.1 Steps Common to the Two Major Pathways for Methanogenesis 241
2.2 Reactions Involved in the Synthesis of CH_3-H_4 MPT 242
2.2.1 Carbon Dioxide Reduction Pathway 242
2.2.2 Acetate Fermentation Pathway 243

3 Molecular Biology 244
3.1 General Features 244
3.2 Transcription 245
3.2.1 RNA Polymerase 246
3.2.2 Promoters and Terminators 246
3.2.3 Messenger RNA 247
3.3 Translation 248
3.3.1 Transfer RNA 248
3.3.2 Initiation and Elongation Factors 248
3.3.3 Archaeal Ribosomes 249
3.4 Complete Genome Sequences 249
3.5 Use of Methanoarchaeal Genomes to Identify Coenzyme Biosynthesis Genes 251
3.6 Genetic Systems 252
3.6.1 Methanococcus 252
3.6.2 Methanosarcina 254
3.7 Pyrrolysine 256

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

Bibliography 257
Books and Reviews 257
Primary Literature 257

Keywords

Domain
A phylogenetic taxon that is higher than the level of kingdom and is based on the classification of organisms according to similarities in 16S ribosomal RNA sequences. The current phylogenetic classification scheme divides organisms into three domains: the Archaea, Bacteria, and Eucarya.

Domain Archaea (Formerly Archaebacteria)
The domain composed of prokaryotic microorganisms with archaeal-type ribosomal RNA. Archaeal cellular membranes typically contain lipids with glycerol diether or diglycerol tetraether linkages. The domain Archaea consists of three phenotypic groups (methane producers, extreme halophiles, and sulfur-dependent extreme thermophiles), but it is formally divided into four kingdoms: the Euryarchaeota, Crenarchaeota, Korarchaeota, and Nanoarchaeota.

Domain Bacteria (Formerly Eubacteria)
The domain composed of prokaryotic microorganisms with bacterial-type ribosomal RNA. Bacterial cellular membranes contain lipids with glycerol diester linkages.

Domain Eucarya
The domain that includes all eukaryotic organisms, characterized by eucaryotic-type 16S ribosomal RNA. Eucaryal cells contain a defined nucleus and glycerol diester cell membrane lipids.

Prokaryote
A cellular organism that lacks a true nucleus.

Although members of the domain Archaea are prokaryotic in cell structure, 16S rRNA sequencing reveals that they form a phylogenetically coherent group, which is separate from both Bacteria (Eubacteria, domain Bacteria) and eucaryotes (Eucarya, domain Eucarya). The Archaea are further distinguished by the ether-lipid composition of cytoplasmic membranes and cell wall components not found in either the Bacteria or Eucarya domains. Like in Bacteria, transcription and translation in Archaea take place in the same cellular compartment. Unlike Bacteria, Archaea

possess a complex DNA-dependent RNA polymerase and genes with promoter elements similar to the TATA box of eucaryotes. Archaea produce and utilize homologs to eukaryotic transcription-initiation factors, which are not found in the Bacteria domain. The methane-producing Archaea (methanoarchaea) are the most diverse and extensively studied group of the Archaea. The strictly anaerobic methanoarchaea are terminal organisms in the anaerobic phase of the global carbon cycle.

1 Introduction

For nearly half a century, life was classified at the highest level by two taxons, prokaryotes and eucaryotes, on the basis of phenotypic cellular characteristics that included the presence or absence of nuclear membranes, organelles, or cell walls. It was not until 1990 that a molecular approach, based on comparisons of 16S ribosomal RNA sequences, revealed that all life forms are divided into three primary divisions (domains) of life: the Eucarya, encompassing all eucaryotes, the Bacteria (or Eubacteria), including ''typical'' Eubacteria, and the Archaea (or Archaebacteria). The Archaea are comprised of four kingdoms: Euryarchaeota, Crenarchaeota, Korarchaeota, and the recently described Nanoarchaeota. In evolutionary terms, the two prokaryotic domains are more genetically distinct from each other than one of the groups, the Archaea, is from the Eucarya. The largest, most diverse, and well-studied representatives of the Archaea are the methane-producing anaerobes (methanoarchaea), which provided the first clues leading to the discovery of the three-domain concept. The methanoarchaea are members of the Euryarchaeota. Research on the methanoarchaea over the past two decades has made the greatest impact on our understanding of the ecology, physiology, and biochemistry of the Archaea. More recently, genomic sequencing, development of genetic systems, and investigations into the molecular biology of the methanoarchaea has significantly expanded our knowledge of the Archaea domain.

2 Biochemistry of Methanogenesis

The Italian physicist Alessandro Volta is most often credited with the discovery of biological methane (natural gas) production, methanogenesis, more than 125 years ago. In performing the ''Volta experiment,'' he disturbed the sediment of lake Como in northern Italy with a pole releasing the trapped methane bubbles. Lighting the gas produced a flame that he called ''combustible air.'' Figure 1 shows a modern-day Volta experiment in which gas from the sediment of a freshwater pond is collected in an inverted funnel submerged in the water column and ignited on release by tipping the funnel. The process of methanogenesis requires a consortium of at least three interacting metabolic groups of obligately anaerobic microbes (Fig. 2). The fermentative group decomposes cellulose and other complex molecules to volatile carboxylic

Fig. 1 A modern-day Volta experiment (see color plate p. xxxiv).

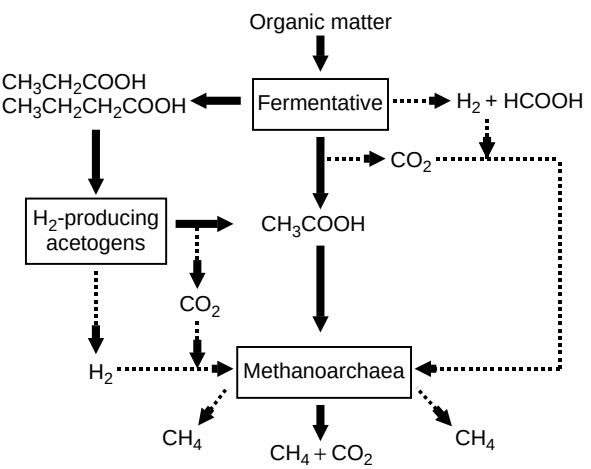

Fig. 2 A scheme showing carbon flow through the three major metabolic groups of anaerobes involved in the conversion of complex organic matter to methane in freshwater environments.

acids and hydrogen gas. Only acetate, carbon dioxide, hydrogen, and formate are the major substrates for the methanoarchaea; thus, the hydrogen-producing acetogenic group is necessary to further metabolize butyrate and propionate to substrates for the methanoarchaea. The methanoarchaea employ two separate pathways in which methane derives from either the methyl group of acetate (reaction 1) or the reduction of carbon dioxide with electrons from hydrogen or formate (reactions 2a and 2b). At least two-thirds of the total methane produced derives from the acetate and

the remainder from the reduction of carbon dioxide.

$$CH_3COO^- + H^+ \longrightarrow CH_4 + CO_2 \quad (1)$$

$$CO_2 + 4H_2 \longrightarrow CH_4 + 2H_2O \quad (2a)$$

$$4HCO_2H \longrightarrow 3CO_2 + CH_4 + 2H_2O \quad (2b)$$

2.1 Steps Common to the Two Major Pathways for Methanogenesis

The steps shown in Fig. 3 are common to the two major pathways for methane production in nature, carbon dioxide reduction and fermentation of acetate. The two pathways differ primarily in the manner by which a methyl group is generated and passed to the cofactor tetrahydromethanopterin (H_4MPT) to form CH_3-H_4MPT. Conversion of CH_3-H_4MPT to methane is essentially the same in both pathways. Step 1 in Fig. 3 is catalyzed by N^5-methyltetrahydromethanopterin:coenzyme M methyltransferase, which has been recently reviewed. The methyltransferase is an integral membrane-bound complex that also functions to generate a sodium-ion gradient across the membrane coupled to the methyl transfer reaction. The enzyme as characterized from *Methanothermobacter marburgensis* and *Methanosarcina mazei* contains a corrinoid cofactor, cobalamine (5′-hydroxybenzimidazolyl cobamide), which accepts the methyl group from CH_3-H_4MPT in the first of the two partial reactions catalyzed by the enzyme. The second partial reaction involves transfer of the methyl group from CH_3-cobalamine to coenzyme M (HS-CoM) producing CH_3-S-CoM. The methyltransferase complex contains eight nonidentical subunits (MtrA-H). MtrA contains the cobalamine cofactor protruding into the cytoplasm that undergoes methylation and demethylation. It is proposed that methylation is catalyzed by MtrH and demethylation by MtrE, which also translocates sodium ions. The mechanism of translocation is thought to involve conformational changes induced in MtrA by methylation and demethylation that are transmitted to MtrE, which drives the

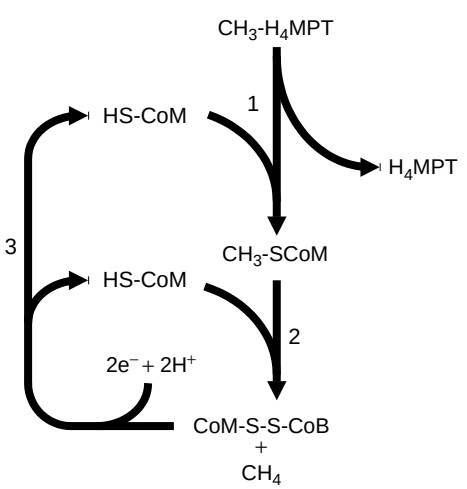

Fig. 3 Steps common to the two major pathways for methanogenesis.

translocation. Growth of the methanoarchaea utilizing either pathway is dependent on sodium and is proposed to use sodium gradients for energetic functions.

The methyl-coenzyme M-reductase catalyzes step 2 in Fig. 3. The electron donor is coenzyme B (CoB) and the heterodisulfide CoM-S-S-CoB is produced in addition to methane. *Methanothermobacter marburgensis* and *Methanothermobacter thermoautotrophicus* contain two isozymes (MCRI and MCRII) with native molecular masses of approximately 300 kDa composed of three different subunits with molecular masses of 65 (α), 46 (β), and 30–35 (γ) kDa in an $\alpha_2\beta_2\gamma_2$ configuration. The crystal structure of the MCRI isozyme from *M. marburgensis* shows coenzyme F_{430} (F_{430}) positioned at the bottom of a narrow channel in two independent active sites separated by approximately 50 Å. The two active sites are formed by residues from the $\alpha\alpha'\beta\gamma$ or $\alpha'\alpha\beta'\gamma'$ subunits indicating that two trimers are necessary to form the active sites. Structures, complexed with either HS-CoM plus HS-CoB or CoM-S-S-CoB, have been the basis for a proposed reaction mechanism. The relative positions of CoM, CoB, and F_{430} in the crystal structure suggests a nucleophilic attack of $[F_{430}]$Ni(I) on CH_3-S-CoM which forms the $[F_{430}]$Ni(III)-CH_3 intermediate. In the next step, $[F_{430}]$Ni(III) oxidizes HS-CoM producing the thiyl radical •S-CoM and $[F_{430}]$Ni(II)-CH_3. In the final step, protonolysis of $[F_{430}]$Ni(II)-CH_3 releases methane and the thiyl radical is coupled to $^-$S-CoB to form the heterodisulfide (CoB-S-S-CoM) accompanied by a one-electron reduction of Ni(II) to the active redox state. The binding of CoM induces specific conformational changes that ensures entry of CH_3-S-CoM adjacent to F_{430} and before entry of CoB in the narrow active-site channel.

Heterodisulfide reductase catalyzes the final step in Fig. 3 releasing the active sulfhydryl forms of the cofactors. The enzyme is comprised of two subunits (HdrDE). The smaller HdrE is a *b*-type cytochrome containing low-spin and high-spin hemes. HdrE accepts electrons from methanophenazine, a component of the membrane-bound electron transport chain. HdrE donates electrons to HdrD, which contains two Fe_4-S_4 centers, one of which is the proposed active site together with a redox-active disulfide of two cysteines. On the other hand, it has been proposed from kinetic experiments with the heterodisulfide reductase from *Methanosarcina thermophila* that the low-potential heme is involved in reduction of the heterodisulfide. In both pathways, heterodisulfide reduction is coupled to formation of an electrochemical proton gradient that drives ATP synthesis catalyzed by an A_1A_0-type ATP synthase. It is hypothesized that a proton gradient is generated across the membrane in a mechanism in which the protons are translocated upon electron transfer from methanophenazine to the heterodisulfide.

2.2 Reactions Involved in the Synthesis of CH_3-H_4 MPT

2.2.1 Carbon Dioxide Reduction Pathway

In most methanogenic environments, the six electrons needed for carbon dioxide reduction to CH_3-H_4MPT derive from the oxidation of either hydrogen gas or formate as shown in reactions (3–5).

$$H_2 + 2\ \text{ferredoxin}^{ox} \longrightarrow 2\ \text{ferredoxin}^{red} + 2H^+ \quad (3)$$

$$2H_2 + 2F_{420} \longrightarrow 2F_{420}H_2 \quad (4)$$

$$2HCOOH + 2F_{420} \longrightarrow 2F_{420}H_2 + 2CO_2 \quad (5)$$

Reaction (3), catalyzed by the Ech hydrogenase, is endergonic given the low partial pressures of hydrogen gas in the native environment; however, it is proposed that the Ech hydrogenase utilizes the electrochemical proton gradient to drive the reaction. Coenzyme F_{420}-dependent hydrogenases (reaction 4) and a formate dehydrogenase (reaction 5) have been characterized. Coenzyme F_{420} is an obligate two-electron carrier involving hydride transfer. Six reactions are required for reduction of carbon dioxide to CH_3-H_4MPT (reactions 6–10).

$$CO_2 + MF + 2\ \text{ferredoxin}^{red} + 2H^+ \longrightarrow 2\ \text{ferredoxin}^{ox} + \text{formyl-MF} + H_2O \quad (6)$$

$$\text{formyl-MF} + H_4MPT \longrightarrow \text{5-formyl-}H_4MPT + MF \quad (7)$$

$$\text{5-formyl-}H_4MPT + H^+ \longrightarrow \text{5,10-methenyl-}H_4MPT^+ + H_2O \quad (8)$$

$$\text{5,10-methenyl-}H_4MPT^+ + F_{420}H_2 \longrightarrow \text{5,10-methylene-}H_4MPT + F_{420} + H^+ \quad (9a)$$

$$\text{5,10-methenyl-}H_4MPT^+ + H_2 \longrightarrow \text{5,10-methylene-}H_4MPT + H^+ \quad (9b)$$

$$\text{5,10-methylene-}H_4MPT + F_{420}H_2 \longrightarrow \text{5-methyl-}H_4MPT + F_{420} \quad (10)$$

In the first reaction (reaction 6), carbon dioxide is attached to methanofuran (MF) as the carbamate and reduced to the formyl level by formylmethanofuran dehydrogenase with ferredoxin as the electron donor. Reaction (7) is catalyzed by formylmethanofuran:tetrahydromethanopterin formyltransferase. Reaction (8) is catalyzed by N^5,N^{10}-methenyltetrahydromethanopterin cyclohydrolase comprised of two identical subunits with no prosthetic groups. Two mechanistically distinct N^5,N^{10}-methylenetetrahydromethanopterin dehydrogenases catalyze reaction (9), one that oxidizes reduced coenzyme F_{420} (reaction 9a) and the other that oxidizes H_2 (reaction 9b). The F_{420}-dependent enzyme is composed of one type of subunit, either as a hexamer or octamer, with no detectable prosthetic group. The other dehydrogenase is essentially a hydrogenase that contains no metals in contrast to all known mechanisms for hydrogenases. The enzyme catalyses the stereospecific transfer of a hydride ion from hydrogen gas into the *pro-R* position of the methylene carbon involving a pentacoordinated carbonium ion (CH_3^+=H_4MPT) transition state intermediate in which the *pro-R* hydrogen bond is protonated and exchanges with water. N^5,N^{10}-Methylenetetrahydromethanopterin reductase catalyses reaction (10). The enzyme is F_{420}-dependent, contains one subunit with no discernible prosthetic groups, and exhibits a ternary complex kinetic mechanism suggesting direct hydride transfer.

2.2.2 Acetate Fermentation Pathway

Most of the methane produced in nature derives from the methyl group of acetate (reaction 1). In *M. thermophila*, CH_3-H_4MPT is synthesized by reactions (11–13).

$$CH_3COO^- + ATP \longrightarrow CH_3CO_2PO_3^{-2} + ADP \quad (11)$$

$$CH_3CO_2PO_3^{-2} + HS\text{-}CoA \longrightarrow CH_3COSCoA + Pi \quad (12)$$

$$CH_3COSCoA + H_4SPT + H_2O + \text{ferredoxin}^{ox} \longrightarrow CH_3 - H_4SPT + \text{ferredoxin}^{red} + CO_2 + HS\text{-}CoA \quad (13)$$

Acetate kinase (reaction 11) and phosphotransacetylase (reaction 12) function together to convert acetate to acetyl-CoA. The crystal structure of acetate kinase from *M. thermophila* reveals an α_2 homodimer with overall features suggesting that the enzyme belongs to the Sugar Kinase/Hsc70/Actin superfamily of phosphotransferases and is possibly the original member of this superfamily. Kinetic and biochemical analyses of site-specific replacement variants indicates a direct in-line mechanism in which the transition state phosphate is stabilized by Arg91 and Arg241. The CO dehydrogenase/acetyl-CoA synthase (Cdh) five-subunit complex cleaves the C−C and C−S bonds in the acetyl moiety of acetyl-CoA, transfers the methyl group to H_4SPT, and oxidizes the carbonyl group to CO_2 (reaction 13). The complex is divisible into three enzyme components. The nickel/iron-sulfur (Ni/Fe-S) component contains the CdhA and CdhB subunits. The corrinoid/iron-sulfur (Co/Fe-S) component contains the CdhD and CdhE subunits. The CdhC subunit constitutes the third component. Three metal clusters (A, B, and C) contained in the complex have properties nearly identical to three metal clusters in the well-characterized clostridial synthase for which functions have been assigned. Thus, cluster A is the proposed site for cleavage of the C−C and/or C−S bonds of acetyl-CoA, whereas the carbonyl group is thought to be oxidized on cluster C. Cluster B is a conventional Fe_4-S_4 center proposed to shuttle electrons from cluster C to the ferredoxin. Dissociation of CdhC from the complex results in loss of an EPR signal diagnostic of cluster A; a result identifying the location of cluster A in CdhC. Finally, the methyl group is transferred to the corrinoid cofactor III of the Co/Fe-S component in which H_4SPT is methylated.

3 Molecular Biology

Over the last decade, two significant advances in molecular biology have greatly facilitated our progress towards a more complete understanding of the biochemistry and physiology of methanoarchaea. First, the sequencing of complete methanoarchaeal genomes has revolutionized approaches to cloning methanoarchaeal genes and characterizing the functions of unknown genes. Second, the development of improved genetics tools has enhanced the ability of researchers to investigate methanogenic processes using genetics approaches. A third exciting discovery in methanoarchaeal genetics has been the expansion of the genetic code to include pyrrolysine, the twenty-second genetically encoded amino acid. This section of the review will briefly summarize the general features of methanoarchaeal genetics and then focus on recent advances in the molecular biology of the methanoarchaea.

3.1 General Features

Methanoarchaea and other Archaea share some genetic features in common with

both Bacteria and Eucarya. Like Bacteria, Archaea lack a membrane-enclosed nucleus and have a relatively small genome (1.66 to 5.75 Mbp) with circular chromosomal DNA and extrachromosomal elements. Some methanoarchaea, particularly members of the family Methanosarcinaceae, possess operons as well as bacterial-like regulatory mechanisms for gene expression. Furthermore, messenger RNA molecules contain a ribosome-binding site that is similar to the Shine–Dalgarno sequence of Bacteria. However, although Archaeal and Bacterial ribosomes are similar in size (70 S), rooted trees comparing small subunit rRNA sequences indicate that Archaea and Eucarya may share a common lineage separate from that of Bacteria. As predicted from this relationship, some components of archaeal replication, transcription, and translation are more similar to those of Eucarya. For example, Archaea possess homologs of eucaryal histones as well as a single complex RNA polymerase that resembles eucaryotic RNA polymerase II in subunit composition. Archaeal transcripts tend to lack introns, but the promoters of methanoarchaea and other Archaea contain a Box A sequence that is similar to the TATA box of Eucarya. Homologs of the eucaryal transcription factors TFIIB, the TATA-binding protein, and TFIIS have been identified in Archaeal genomes. In both Archaea and Eucarya, translation is initiated using methionine rather than formyl-methionine, which is common in Bacteria.

Histone-like proteins purified from *Methanothermus fervidus* (Hmf) have been characterized, and these proteins share amino acid sequence similarity with a protein family that includes eucaryotic histone proteins (H2A, H2B, H3, and H4). X-ray crystallography of the recombinant forms of two methanoarchaeal histone homologs (HmfA and HmfB) shows that the three-dimensional structure of both proteins resembles the characteristic histone fold of eucaryal core histone proteins. HmfA and HmfB may form homodimers, heterodimers, and tetramers in solution, but have not been observed to associate into octamers, which are typical of eukaryotic nucleosomes. The crystal structure of the histone-like protein from *Methanopyrus kandleri* (Hmk) has also been determined. Hmk is twice as long as HmfA, HmfB, and eucaryal histones, and contains two histone folds per monomer. The precise function of methanoarchaeal histone-like proteins has not yet been determined, but tetramers of Hmf interact with DNA and may play roles in regulation of gene expression or protection of DNA against heat denaturation.

3.2 Transcription

Gene expression in Archaea is said to resemble Bacteria in form and Eucarya in function. Similarities between Archaea and Bacteria include circular chromosomal DNA, the occurrence of transcription and translation in the same cellular compartment, the organization of some genes into operons, the absence of introns in protein-encoding genes, minimal post-transcriptional modification of mRNA, similarity in ribosome size, and the presence of a Shine–Dalgarno-like sequence to direct translation initiation. Interestingly, there are also a number of similarities between eucaryal and archaeal transcription and translation including the complexity of DNA-dependent RNA polymerases, promoters that contain TATA-box-like elements, transcription factors, elongation factors II that can be ribosylated with ADP,

introns in certain rRNA and tRNA genes, and translation beginning with methionine rather than formylated methionine.

3.2.1 RNA Polymerase

Bacteria contain a single DNA-dependent RNA polymerase composed of four different subunits with the stoichiometry of $\alpha_2\beta\beta'\sigma$. In contrast, Eucarya have three different nuclear RNA polymerases (I, II, and III), which function in rRNA synthesis (I), mRNA synthesis (II), and tRNA and small RNA synthesis (III). Each of the eucaryal RNA polymerases contains two large subunits and 8–12 smaller subunits. The large subunits A and B correspond to the β' and β subunits respectively of the bacterial RNA polymerase. Archaeal species possess a single DNA-dependent RNA polymerase, composed of about 12 subunits which may include B, A′, A″, D, E, F, G, H, I, K, L, and N. In some Euryarchaeota, subunit B appears to be divided into two fragments, B′ and B″, which respectively correspond to the C-terminal and N-terminal halves of subunit B. The deduced amino acid sequences of Archaeal A′ and A″ components have similarity to eucaryal subunit A and bacterial subunit β', while B (or B′ and B″) are similar to eucaryal B and bacterial β. Archaeal subunit H bears homology to a eucaryal subunit (RPB5) that is present in all three eucaryal RNA polymerases but has no counterpart in the bacterial holoenzyme.

3.2.2 Promoters and Terminators

The general features of methanoarchaeal promoters are illustrated in Fig. 4. A consensus sequence for the Archaeal promoter sequence called Box A has been determined from analysis of numerous rRNA, tRNA, and protein-encoding genes. Box A consists of a stretch of

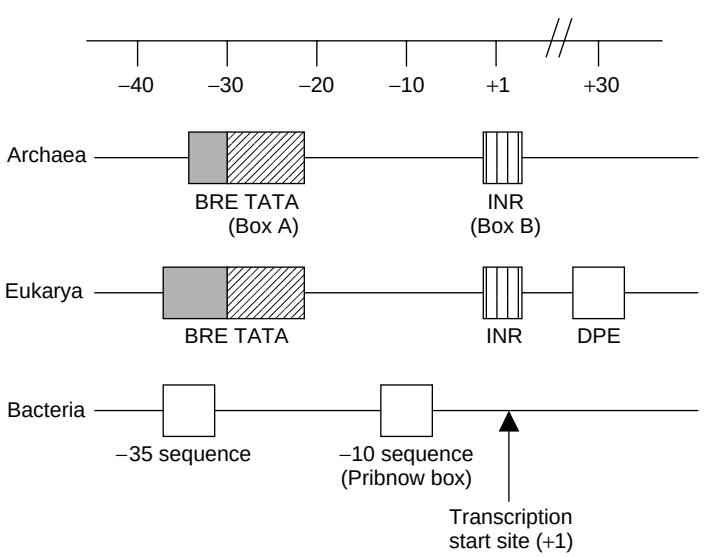

Fig. 4 Components of archaeal, eucaryal, and bacterial promoters. Archaeal and eucaryal promoters typically contain a TATA box sequence, a BRE, and an INR in the vicinity of the transcription start site. Eucarya may have a downstream promoter element (DPE). Bacterial promoters typically consist of a −35 and −10 consensus sequence.

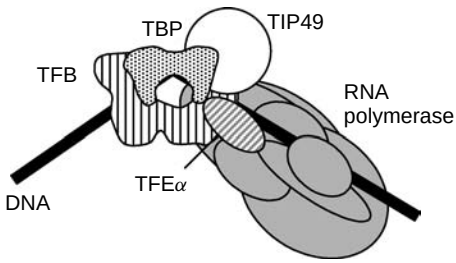

Fig. 5 Proposed model for the archaeal preinitiation complex. The TBP and TFB bind to DNA and recruit the complex archaeal RNA polymerase. Transcription activator factor TFE α is present in all sequenced archaeal genomes. A TBP-interacting protein homolog (TIP49) is also found in the genomes of *M. kandleri* and several hyperthermophilic sulfur-dependent archaea. Adapted from.

six nucleotides TTTA (T/A) A with similarity to the eukaryotic TATA box and is centered approximately 27 nucleotides upstream of the transcription start site. When the Archaeal consensus promoter sequence is replaced with the eukaryotic TATA box, transcription efficiency decreases to 13% of its original activity. Most methanoarchaea genes contain a second consensus sequence called box B or the initiator element (INR), which is located around the transcription start site. In the methanoarchaea, the INR consensus sequence (ATGC) transcription begins at the G position. In addition, a pair of adenines positioned approximately 33 and 34 nucleotides upstream of the transcription start site functions as a recognition element for transcription factor B (TFB) and is called the BRE (B recognition element) sequence. In Eucarya, the role of the TFB homolog (TFIIB) is to cooperate with transcription factor TFIID to assist in the binding between the TATA box and RNA polymerase II. The subunit of TFIID that specifically interacts with the eukaryotic TATA box is called the *TATA-binding protein* (TBP), and a homolog of TBP functions in Archaea. An archaeal homolog of transcription factor TFIIE has been shown to stimulate transcription from *M. thermoautotrophicus* promoters in an *in vitro* system. A proposed model for the archaeal preinitiation complex is shown in Fig. 5.

Two different potential termination signals have been found in Archaea. The first is an oligo-T sequence or a T-rich stretch of polypyrimidines. The second consists of short inverted repeats that may form stem-loop structures to direct termination. In thermophiles, transcription termination appears to be relatively inefficient, usually occurring at different sites within polypyrimidine stretches 4–30 nucleotides long.

3.2.3 Messenger RNA

Archaea contain polycistronic transcriptional units that resemble bacterial operons. Unlike most bacterial operons, however, Archaeal transcriptional units may contain internal promoters and terminators. In some cases, transcripts of different lengths are produced from the same operon, consistent with mRNA processing. No introns have been found in Archaeal or bacterial mRNAs that encode proteins; however, some Archaeal rRNA and tRNA transcripts contain sequences that are subsequently removed by posttranscriptional modification. At present, there is no evidence for posttranscriptional modification of mRNA in Archaea. RNAs with short poly(A) tails (about 12 nucleotides in length) have been found in Archaea, but they have not been experimentally identified as mRNA. The 5′ cap structures found in eucaryal mRNAs are not found in

Archaeal mRNAs. Ribosome-binding sites complementary to the 3′ terminus of the 16S Archaeal rRNA and similar to bacterial Shine–Dalgarno sequences are present in methanoarchaea mRNAs.

3.3 Translation

Archaea use the same universal genetic code as Bacteria and Eucarya. Codon usage generally depends on the GC content of the organisms, which varies from 26 to 71% in Archaea; thus GC-rich Archaea tend to prefer codons with G or C in the third position, although exceptions have been observed. The most common translation initiation codon is AUG, but GUG and UUG are also used. The ribosomes and other translational machinery used in the different domains exhibit distinctive qualities unique to each of the three lineages. Eucaryal translation differs from bacterial translation in several ways. Translation in Eucarya is posttranscriptional; mRNAs generally encode only a single gene; mRNAs are modified with a 5′ cap (7-methylguanosine) and a polyadenylated tail; the ribosome recognizes the 5′ cap structure of mRNA rather than a Shine–Dalgarno sequence; translation is initiated with methionine rather than formylated methionine; and initiation requires up to 10 additional protein factors. Translation in Archaea has some bacterial and some eucaryal features. Since no nuclear membrane is present, Archaeal transcription necessarily occurs in the same compartment as translation; mRNAs may be polycistronic, and posttranscriptional modification of mRNAs is not observed. Shine–Dalgarno sequences are utilized in some Archaea, but are unnecessary in others. As in Eucarya, translation is initiated with nonformylated methionine and uses a diphtheria-sensitive elongation factor (EF-2).

3.3.1 Transfer RNA

The tRNAs of Archaea contain several modified nucleotides present in Bacteria and Eucarya (pseudouridine, 2′-*O*-methylcytidine, 1-methylguanosine, and N^2N^2-dimethylguanosine), but lack ribothymidine and 7-methylguanosine, which are present in Bacteria and Eucarya. Archaeal tRNAs contain additional modified nucleotides unique to the Archaeal domain. Several aspects of Archaeal tRNA synthetases differ from the model systems described in Bacteria, including different mechanisms for charging asparagine, cysteine, glutamine, and lysine tRNAs.

3.3.2 Initiation and Elongation Factors

The number of protein factors required for archaeal initiation is unknown; however, it has been proposed that Archaea may possess initiation factors similar to those in Eucarya because hypusine, an unusual amino acid previously found only in eukaryotic initiation factor 5A, has been found in a *Sulfolobus* protein with sequence similarity to eukaryotic initiation factor 5A. The recombinant *M. jannaschii* initiation factor 5A homolog has been crystallized and contains a conserved lysine at position 40 that may be converted into hypusine *in vivo*. Archaeal elongation factors EF-1α and EF-2 contain regions that are highly conserved among elongation factors from all three domains. These conserved amino acid stretches are involved in GTP binding and hydrolysis. Interestingly, as in the eucaryal EF-2, one of the histidines at the C-terminus of the Archaeal EF-2 is posttranslationally modified to form

diphthamide, which makes it susceptible to ADP-ribosylation by the diphtheria toxin. Phylogenetic trees based on comparisons of EF-2 homologs are consistent with the clustering of Archaeal sequences on a branch separate from branches leading to eucarya or Bacteria; however, phylogenies based on EF-1α sequences are less consistent.

3.3.3 Archaeal Ribosomes

Archaeal ribosomes bear superficial similarity to those of Bacteria. Both Archaeal and bacterial ribosomes have a sedimentation coefficient of 70S and are composed of two subunits with sedimentation coefficients of 50S and 30S respectively. Archaeal rRNAs (23S, 16S, and 5S) are similar in size to bacterial rRNAs, but differ substantially in sequence. In terms of overall structure, Archaeal ribosomes appear intermediate in shape between eucaryal and bacterial ribosomes. For example, the small ribosomal subunits of Eucarya possess several bulges and gaps, plus a "bill" that is not present in bacterial ribosomes (Fig. 6). All Archaeal ribosomes possess a bill in their 30S ribosomes; however, only ribosomes from sulfur-dependent hyperthermophiles possess the other eukaryotic features.

3.4 Complete Genome Sequences

Sequencing of the complete genomes of five representative methanoarchaea has led to genome-level comparisons among organisms of different species and domains and simplified strategies for determining the function of unknown methanoarchaea genes. The first Archaeal genome to be sequenced was that of the hyperthermophilic methanoarchaeon *Methanococcus jannaschii*. The genome consists of a 1.66-Mbp chromosome with 1682 predicted protein-encoding genes, a large circular extrachromosomal element, and a small circular extrachromosomal element. Initially, putative cellular functions could be assigned to only 38% of the predicted genes. Genome sequencing has confirmed the similarity of Archaeal and eucaryal proteins involved in the processes of replication, transcription, and translation. For example, the *M. jannaschii* genome contains homologs of eucaryal histones, RNA polymerase

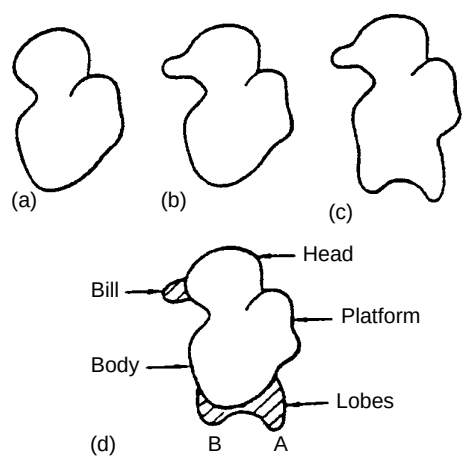

Fig. 6 Diagrammatic representation of small ribosomal subunits from (a), Bacteria, (b) Archaea, and (c) Eucarya. The striped regions (d) show the archaeal and eucaryal bill regions and the lobes (A and B). [From J. A. Lake, E. Henderson, M. W. Clark, and A. T. Matheson, *Proc. Natl. Acad. Sci. U.S.A.* **79**: 5948–5952 (1982).]

subunits, transcription-initiation factors, selected ribosomal subunits, and some aminoacyl–tRNA synthetases. However, homologs of the aminoacyl–tRNA synthetases for glutamine, asparagine, lysine, and cysteine are notably absent. The genome contains a tRNA for selenocysteine, which is encoded by the UGA stop codon, as well as four genes with internal UGA codons that are predicted to encode selenocysteine.

The genome sequence of *M. jannaschii* has provided an invaluable resource for cloning methanoarchaeal genes for expression in *Escherichia coli*. In many cases, the thermostable nature of recombinant *M. jannaschii* proteins provides a convenient means of protein purification by heating cell extracts to denature thermolabile *E. coli* proteins. Problems involving codon-usage bias have been overcome by expressing genes in *E. coli* strains that contain rare codon tRNAs. To a large extent, this strategy has obviated the need for growing large quantities of strictly anaerobic methanoarchaea for use in protein purification. The approach has been particularly successful for identifying the genes encoding low abundance proteins, such as enzymes involved in biosynthetic processes.

Sequencing the complete genome of *M. thermoautotrophicus* ΔH provided the first opportunity to make genomic comparisons of methanoarchaea with other Archaea, Bacteria, and Eucarya. Forty-six percent of the predicted proteins have been assigned putative functions based on homology to known enzymes, 28% have homology to genes of unknown function, and 27% lack significant homology to any known sequence. Substantial divergence of *M. thermoautotrophicus* from *M. jannaschii* has been postulated on the basis of a low degree of conservation in gene organization and the observation that less than 20% of the *M. thermoautotrophicus* proteins show greater than 50% identity with *M. jannaschii* proteins. Homologs of two-component regulatory systems and heat shock-70 response proteins are abundant in *M. thermoautotrophicus* but absent in *M. jannaschii*. As in *M. jannaschii*, tRNA synthetases for asparagine, glutamine, cysteine, and lysine are unrecognizable, but interestingly, the translation machinery for selenocysteine is lacking in *M. thermoautotrophicus*. The genome contains homologs of the eucaryal TATA-binding proteins, TFIIB and TFIIS.

Methanopyrus kandleri is a hyperthermophilic methanoarchaea that grows at temperatures up to 110 °C. Phylogenies based on rRNA sequencing have placed this methanoarchaeon in a deeply branching position separate from all other known methanoarchaea. Whole genome sequence comparisons have provided a different perspective on this evolutionary model. Although *M. kandleri* diverges from other methanoarchaea according to 16S rRNA phylogenies, it clusters with *M. jannaschii* and *M. thermoautotrophicus* when the genomes are compared on the basis of either conserved gene order and content or concatenated alignments of ribosomal proteins. In contrast to other methanoarchaea, the *M. kandleri* genome contains a high proportion of negatively charged amino acids that may assist in adapting proteins to the high intracellular salt concentrations of this organism. In addition, *M. kandleri* has relatively few putative regulatory and signaling proteins and less evidence for the lateral transfer of genes from Bacteria.

In contrast, the genomes of *Methanosarcina acetivorans* C2A and *Methanosarcina mazei* Go1 reflect the wide metabolic diversity of the *Methanosarcinaceae* family. The 5.75-Mbp genome of *M. acetivorans*

C2A is three times larger than the three previously sequenced methanoarchaea genomes and shows evidence of extensive lateral gene transfer. Similarly, in the 4.1-Mbp genome of *M. mazei* Go1 greater than 30% of the open-reading frames are more closely related to bacterial homologs than Archaeal homologs. In *M. acetivorans* C2A, predicted coding regions account for 74% of the genome, substantially lower than the gene densities of *M. jannaschii*, *M. thermoautotrophicus*, and *M. kandleri* that all exceed 90%. The genome contains multiple copies of numerous genes; including genes for methanogenesis related enzymes, putative corrinoid proteins, and methyltransferases, TATA-binding proteins, and homologs of bacterial regulator proteins such as sensory transduction histidine kinases. *M. acetivorans* contains both class I and class II lysyl–tRNA synthetases. One of these may be involved in translating a UAG amber codon into pyrrolysine.

3.5 Use of Methanoarchaeal Genomes to Identify Coenzyme Biosynthesis Genes

One field that has reaped tremendous benefit from the sequencing of methanoarchaeal genomes is functional genomics, including the identification of genes involved in the biosynthesis of specialized methanoarchaeal coenzymes. Progress has been made in the elucidation of the pathways of methanoarchaeal coenzyme biosynthesis. However, prior to genome sequencing progress in identifying the biosynthetic genes was limited, due in part to the low abundance of the enzymes and the lack of a genetics system.

Published genome sequences have opened up new strategies for identifying coenzyme biosynthesis genes. For example, the complete *Methanococcus jannaschii* genome sequence enabled use of a secondary structure analysis program to identify a proposed tetrahydromethanopterin (H_4MPT) biosynthesis gene encoding an enzyme that catalyzes a reaction similar in mechanism to that of dihydropteroate synthase. The secondary structure analysis program predicted that *M. jannaschii* gene *MJ0301* encodes a protein that would demonstrate secondary structure features similar those of dihydropteroate synthase. When *MJ0301* was expressed in *Escherichia coli*, the resulting protein catalyzed the predicted reaction of H_4MPT biosynthesis. *MJ0301* was the first H_4MPT biosynthesis gene to be identified.

A second effective strategy for identifying genes involved in coenzyme biosynthesis has been to propose a mechanism for the enzymatic reaction, and then search for genes that show homology to proteins that catalyze a similar reaction. For example, the formation of the α-keto acid (S)-sulfopyruvate during coenzyme M biosynthesis is proposed to occur by oxidation of (S)-sulfolactate. This mechanism resembles that of malate dehydrogenase, and the *M. jannaschii* genome contains only two genes annotated as malate dehydrogenase genes (*MJ1425* and *MJ0490*). Only the recombinant *MJ1425* gene product efficiently catalyzes the reaction of (S)-sulfolactate dehydrogenase, implicating a role for the gene *MJ1425* in coenzyme M biosynthesis. The *MJ1425* protein also catalyzes the reduction of α-ketoglutaric acid to (S)-hydroxyglutaric acid, a component of the coenzyme H_4MPT. Using mechanistic predictions, at least 16 genes involved in coenzyme, amino acid, or nucleotide biosynthesis have been assigned biochemical functions.

A third strategy to facilitate the assignment of coenzyme biosynthesis genes has been to take advantage of clustered genes in Bacteria that synthesize analogs of coenzymes. The methylotrophic bacterium *Methylobacterium extorquens* AM1 synthesizes an analog of H_4MPT and contains a cluster of genes that encode proteins involved in the H_4MPT-dependent oxidation of methanol. Some of these genes code for proteins with similarity to uncharacterized archaeal proteins and have been tested as candidates for H_4MPT biosynthesis genes. One of the key enzymes of H_4MPT biosynthesis is ribofuranosylaminobenzene 5′-phosphate synthase (RFAP), which catalyzes the first committed step of the pathway. The N-terminal sequence of purified RFAP synthase from the methanoarchaeon *M. thermophila* has been used to determine the corresponding gene in the genomes of *M. jannaschii* and *M. thermoautotrophicus*. Fortuitously, an RFAP synthase gene homolog also occurs among the clustered bacterial genes required for H_4MPT-dependent metabolism in *M. extorquens*. Expression in *E. coli* of the putative RFAP synthase genes from *M. thermoautotrophicus* (*MTH0830*) and *M. extorquens* (*orf4*) has produced enzymes with RFAP synthase activity.

The highly developed genetics system of *M. extorquens* has recently been used to identify seven additional genes of H_4MPT biosynthesis. Using an enzymatic assay to measure levels of H_4MPT coenzymes in *M. extorquens* cells, it has been shown that deletion mutants lacking the RFAP synthase gene are incapable of producing H_4MPT. Using this assay as a screen, seven additional genes implicated in H_4MPT biosynthesis have been found in *M. extorquens*. Six of these genes cluster together with *orf4* and encode Archaeal like proteins of unknown function. The seventh gene (*dmrA*) codes for an analog of bacterial dihydrofolate reductase that catalyzes the NADPH-dependent reduction of dihydromethanopterin to H_4MPT. These studies demonstrate how bacterial model systems can be combined with complete genome sequences to assist in determining the functions of uncharacterized genes.

3.6 Genetic Systems

Recent advances in the development of genetic systems have opened up the possibility of creating mutants to investigate the physiology, biochemistry, and regulation of gene expression in the methanoarchaea. Genetic tools for creating random and site-directed mutants, complementation strains, and gene reporters have been developed in combination with improved plating techniques, transformation procedures, and selectable markers. Model genetics systems have been developed for *Methanococcus maripaludis* and *Methanosarcina* species, and some of the genetics tools are effective in other methanoarchaea of the same genus.

3.6.1 Methanococcus

One of the challenges of developing a genetics system for methanoarchaea has been the lack of suitable antibiotics and antibiotics resistance markers. Because of fundamental differences between the ribosomal proteins of Bacteria and Archaea, typical bacterial antibiotics are often ineffective as inhibitors of methanoarchaeal protein synthesis. The first antibiotic to be used in a methanoarchaeal vector was puromycin, which inhibits the growth of *Methanococcus* and *Methanosarcina* species at concentrations less than $5\ \mu g\ mL^{-1}$.

Methanococcal integration vectors have been created using the puromycin acetyltransferase gene (*pac*) from *Streptomyces alboniger* under the control of the promoter and terminator for the methyl coenzyme M methylreductase gene. Vectors for transposon insertional mutagenesis are also available. Transformation procedures in *Methanococcus* species include polyethylene glycol treatment of *M. maripaludis* cells, which produces over 100 transformants per μg of plasmid, and transformation of *Methanococcus voltae* protoplasts by electroporation.

A shuttle vector (pDLT44) allowing plasmid replication in both *M. maripaludis* JJ and *E. coli* has been developed on the basis of the methanococcal plasmid pURB500. From pDLT44, an expression vector has been created by adding the strong methanococcal histone promoter and a ribosomal binding site. The sensitivity of *Methanococcus maripaludis* to neomycin at concentrations between 100 and 250 μg ml^{-1} and the use of neomycin resistance genes has expanded the range of methanococcal genetics tools and allowed for deletion of multiple genes in a single organism and complementation using expression vectors.

Studies on the regulation of gene expression in *M. maripaludis* have focused on processes involving nitrogen metabolism and formate utilization. Nitrogen fixation, the reduction of atmospheric nitrogen to ammonia, is highly regulated in methanoarchaea and has provided an excellent model system for identifying regulatory components in *M. maripaludis*. As in Bacteria, nitrogenase activity and the expression of nitrogen fixation (*nif*) genes in methanoarchaea decrease when fixed nitrogen is available as ammonia. Growth of *M. maripaludis* under nitrogen-fixing conditions requires molybdenum (but not vanadium) and is inhibited by tungsten. These data are consistent with phylogenetic evidence that the *M. maripaludis* nitrogenase is related to the class of bacterial nitrogenases that contain iron-molybdenum cofactor. Nitrogenase is encoded by the structural genes *nifH, nifD*, and *nifK*, and accessory genes are required for assembly of the enzyme into an active form. The order of the *nif* genes in methanoarchaea (*nifHDKENX*) is the same as the order found in Bacteria, with the interesting exception that two homologs of the regulatory gene *glnB* are inserted between the *nifH* and *nifD* genes. The *glnB* homologs also occur in the *nif* gene clusters of *Methanosarcina barkeri*. In Bacteria, the *glnB* gene product is involved in regulation of nitrogen metabolism and can be uridylylated at a conserved tyrosine residue in the T-loop.

In contrast to the bacterial *nif* genes, the eight *nif* genes of *M. maripaludis* form a single operon that can be transcribed as one large transcript (or up to ten smaller transcripts depending on the growth conditions). Transposon insertion mutagenesis indicates that all of the genes except *nifX* are essential for nitrogen fixation activity. The function of *nifX* in both Bacteria and Archaea is unknown, but a role in either cofactor biosynthesis or regulation has been proposed.

Transcription of the *nif* genes of *M. maripaludis* is repressed by ammonia. Two sets of palindromic sequences have been found between the *nifH* promoter and the transcriptional start site. Directed mutagenesis of the *nifH* promoter region fused to a *lacZ* reporter gene indicates that the first palindrome, but not the second, is involved in the regulation of *nif* gene transcription by ammonia. A similar palindrome (GGAA-N6-TTCC) is also found in the promoter region of several other genes

involved in nitrogen metabolism, including genes for glutamine synthetase (*glnA*), an ammonia-dependent NAD synthetase homolog, and an operon containing a putative ammonia transporter gene and a *glnB* homolog. Genes controlled by this operator may constitute a nitrogen metabolism regulon under the control of a single repressor protein. The Archaeal *nif* repressor protein is a negative regulatory element, in contrast to the bacterial regulatory proteins NifA and NtrC, which are positive regulators (activators) of *nif* expression. When alanine is the sole nitrogen source, alanine represses transcription of both the *nif* operon and the *glnA* gene to an intermediate level. Both the first and the second set of palindromic sequences in the *nifH* promoter region are needed for alanine repression of *nif* gene expression.

The availability of ammonia and alanine also regulates the transcription of *glnA*. The *glnA* promoter has only one operator for binding the nitrogen repressor protein, but there appear to be three TATA boxes and three transcriptional start sites. The first promoter provides a low, constitutive level of *glnA* expression, while the other two promoters are regulated through an operator that resembles the first palindromic sequence of the *nifH* promoter.

Nitrogenases from both bacterial and archaeal sources demonstrate the phenomenon of ammonia switch-off, which is a posttranslational inactivation of nitrogenase in the presence of ammonia. In *Rhodospirillum rubrum*, ammonia addition leads to the reversible ADP-ribosylation of the *nifH* gene product. However, in *M. maripaludis* ammonia switch-off occurs without ADP-ribosylation. Deletion mutagenesis and genetic complementation studies show that the process requires the two novel *glnB* homologs found in the methanoarchaeal *nif* operon. Although bacterial GlnB homologs can be modified by uridylylation of a conserved tyrosine in the T-loop, the Archaeal homologs might not become uridylylated because the methanoarchaeal homologs lack the conserved T-loop.

The ability of *M. maripaludis* to grow on either formate or H_2/CO_2 provides a second system for studying gene regulation in methanoarchaea. The genome of *M. maripaludis* has two clusters of formate dehydrogenase genes. The neomycin and puromycin resistance genes have been used to construct individual deletion strains and double mutants lacking both of the *fdhA* genes. Both genes are required for maximal growth rate on formate, and deletion of both *fdhA* genes eliminates the ability to grow on formate. Promoter–*lacZ* fusions show that H_2 represses expression of formate operons, but formate concentration has no effect on expression.

3.6.2 Methanosarcina

The metabolic diversity of *Methanosarcina* species and the higher abundance of regulated genes and operons make this genus an attractive target for genetic analysis. Early technical obstacles to developing a genetics system for *Methanosarcina* species included the tendency of the cells to grow in clumps called *sarcina*, the low transformation efficiency, a paucity of selectable markers, and a lack of suitable vectors for mutagenesis and expression. Progress in establishing a facile genetics system began with the observation that *Methanosarcina* cells grow as single cells in a moderate salt (marine) medium, which enables the plating of single colonies. A second breakthrough in *Methanosarcina* transformation has been the use of liposomes to deliver DNA into protoplasts. This procedure increases the transformation efficiency of

M. acetivorans by up to five orders of magnitude over polyethylene glycol-mediated transformation and electroporation. Nine out of eleven *Methanosarcina* strains tested could be transformed using the liposome-mediated technology.

Shuttle vectors for use in *E. coli* and *Methanosarcina* species have been developed by modifying plasmid pC2A from *M. acetivorans* to contain a puromycin resistance gene, a multiple cloning site, and *lacZα* for blue–white screening. Because neomycin apparently does not inhibit *Methanosarcina* strains, pseudomonic acid has been used to develop a second selectable marker system. Pseudomonic acid is a structural analog of isoleucine and inhibits the charging of isoleucyl tRNA synthetases. The resistance gene codes for a pseudomonic acid–resistant isoleucyl tRNA synthetase created by mutagenesis of *ileS* from *M. barkeri*.

A plasmid that can be used for random *in vivo* transposon mutagenesis of *Methanosarcina* is now available. The suicide vector contains the *Himar1* transposase under the control of the methyl coenzyme M methylreductase promoter and mini-*Himar1* elements. Using transposons, a mutation conferring resistance to the methanogenesis inhibitor bromoethanesulfonic acid (BES) has been localized to a putative membrane transport protein that may be needed for BES uptake. Random transposon mutagenesis has also shown that the resistance of *M. acetivorans* to fluoroacetic acid involves the acetate kinase–phosphotransacetylase operon needed for growth on acetate.

Directed mutagenesis of proline biosynthesis genes in *M. acetivorans* has been accomplished by replacing wild-type genes (*proA*, *proB*, or *proC*) with the genes disrupted by insertion of puromycin or pseudomonic acid resistance genes. Liposomes containing the mutated linearized DNA are used to transform *M. acetivorans*, and replacement of the wild-type gene by homologous recombination produces proline auxotrophs. These mutants can be complemented with plasmids containing the wild-type genes.

This suite of genetic tools has been used to investigate the physiological role of the Ech hydrogenase (Ech) of *M. barkeri* Fusaro. Ech catalyzes the reversible reduction of ferredoxin by H_2, and its association with the cell membrane has led to a proposed role in (i) energy conservation, (ii) the endergonic reduction of CO_2 to CHO-methanofuran during hydrogenotrophic growth, or (iii) steps in the synthesis of acetyl-CoA and pyruvate for anabolism. A *M. barkeri* Fusaro mutant has been created in which a puromycin resistance gene cassette replaces the entire *echABCDEF* operon. As anticipated, cell extracts from the *ech* deletion mutant lack ferredoxin-dependent hydrogenase activity. The mutant grows normally on methanol, but is unable to grow on acetate, H_2/CO_2, or H_2/CO_2/methanol. The mutant produces methane from methanol, but not from acetate or H_2/CO_2. Bypassing the first step of methanogenesis from H_2/CO_2 by providing H_2 and formaldehyde results in methane production, suggesting that the mutant is defective in the first step of hydrogenotrophic methanogenesis. The mutant is also unable to convert CO and H_2O to CO_2 and H_2, indicating that the oxidative half-reaction of methanogenesis from acetate is also nonfunctional. The mutant can be complemented for all phenotypes by inserting a functional copy of the *echABCDEF* operon in a neutral chromosomal location. These data are consistent with a

role for Ech in the first step of methanogenesis from H_2/CO_2 and in methane production from acetate, which may both require a ferredoxin-dependent hydrogenase. An additional role for Ech in synthesizing pyruvate for anabolic pathways has been proposed because exogenous pyruvate restores the ability of the deletion mutant to grow on H_2/CO_2/methanol, consistent with a role for Ech in pyruvate biosynthesis. The experiments described demonstrate the usefulness of the improved genetic tools for functional genomics analysis and for genetics studies that address fundamental questions about methanoarchaeal physiology and biochemistry.

3.7 Pyrrolysine

An exciting recent discovery in methanoarchaeal genetics has been the expansion of the genetic code through characterization of pyrrolysine, the twenty-second genetically encoded amino acid. This molecule consists of a substituted pyrroline-5-carboxylate attached by an amide bond to the epsilon amino group of lysine (See Fig. 7). Like selenocysteine, pyrrolysine is encoded by read-through of a stop codon (in this case, UAG rather than UGA). The N-terminal sequence of the purified monomethylamine methyltransferase from *M. barkeri* MS corresponds to the gene *mtmB1*. Although the molecular mass of the purified protein (MtmB) is 50 kDa, the *mtmB1* gene contains an apparent in-frame UAG stop codon that would truncate the protein of at 23 kDa. Internal sequencing of MtmB and mass spectrometric analysis of peptide fragments suggest that lysine occurs at the position corresponding to the UAG codon. The x-ray crystal structures of the monomethylamine methyltransferase show electron density consistent with a lysine linked to a 4-substituted-pyrroline-5-carboxylate moiety at the UAG position. The substitution at the 4-position may be an amine or a hydroxyl group. On the basis of its position in the protein active site, the novel amino acid has been proposed to play a specific role in transferring methyl groups during catalysis.

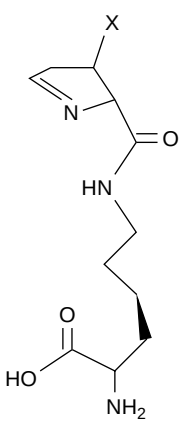

Fig. 7 Structure of pyrrolysine.

The mechanism of tRNA charging with pyrrolysine is currently being elucidated, and there is at present some controversy about the enzymes involved in the process. One investigation implicates the *pylTSBCD*L gene cluster located near one of the methyltransferase gene clusters of *M. barkeri* Fusaro. The *pylT* gene codes for a tRNA with a CUA anticodon capable of base pairing with the UAG codon. The protein encoded by *pylS* shares some homology with the core catalytic domains of class II aminoacyl–tRNA synthetases. Recombinant *M. barkeri* MS *pylS* expressed with an N-terminal six-histidine tag (His_6) in *E. coli* has been tested for lysine aminoacyl–tRNA synthetase activity. When combined with [^{14}C]lysine and

tRNA isolated from *M. barkeri*, His_6-PylS appears to catalyze the lysyl–tRNA synthetase reaction with an apparent K_m of 2.2 μM and an apparent k_{cat} of 1.6 min^{-1}. The K_m is comparable to the values found for other aminoacyl–tRNA synthetases. However, the k_{cat} value is between 12 and 5000 times lower than the rate obtained for other aminoacyl–tRNA synthetases. Given the low K_m for lysine, it is conceivable that the amber codon tRNA (PylT) is first charged with lysine and later modified to become pyrrolysyl-tRNA. A similar mechanism has been proposed for the biosynthesis of asparaginyl-tRNA, glutamyl-tRNA, and selenocysteinyl-tRNA. Because the proposed amber suppressor lysyl–tRNA synthetase (PylS) does not strongly resemble the typical class I or class II lysyl–tRNA synthetases, PylS is proposed to be a new class of lysyl–tRNA synthetase. In contrast, a second study presents contradictory evidence that His_6-PylS from *M. barkeri* Fusaro and *M. barkeri* MS cannot charge the amber suppressor tRNA (PylT) with lysine. Instead, it is proposed that PylT can be charged only if both the class I and class II lysyl–tRNA synthetases of *M. barkeri* are present in the reaction. Neither of the synthetases alone can catalyze the tRNA synthetase reaction. Further experiments will undoubtedly help resolve this inconsistency and provide further details on the genetic processes involved in incorporating pyrrolysine into proteins.

See also Bacterial Cell Culture Methods; Bacterial Growth and Division; Microbial Development; Organic Cofactors as Coenzymes; Plasmids.

Bibliography

Books and Reviews

Deppenmeier, U. (2002b) The unique biochemistry of methanogenesis, *Prog. Nucleic Acid Res. Mol. Biol.* **71**, 223–283.

Ferry, J.G. (1993) *Methanogenesis*, Chapman & Hall, New York, p. 536.

Ferry, J.G. (2003) One-carbon Metabolism in Methanogenic Anaerobes, in: Ljungdahl, L.G., Adams, M.W., Barton, L.L., Ferry, J.G., Johnson, M.K. (Eds) *Biochemistry and Physiology of Anaerobic Bacteria*, Springer-Verlag, New York, pp. 143–156.

Gottschalk, G., Thauer, R.K. (2001) The Na^+ translocating methyltransferase complex from methanogenic archaea, *Biochim. Biophys. Acta* **1505**, 28–36.

Hickey, A.J., Conway de Macario, E., Macario, A.J. (2002) Transcription in the archaea: basal factors, regulation, and stress-gene expression, *Crit. Rev. Biochem. Mol. Biol.* **37**, 537–599.

Lange, M., Ahring, B.K. (2001) A comprehensive study into the molecular methodology and molecular biology of methanogenic Archaea, *FEMS Microbiol. Rev.* **25**, 553–571.

Reeve, J.N., Nolling, J., Morgan, R.M., Smith, D.R. (1997) Methanogenesis: genes, genomes, and who's on first? *J. Bacteriol.* **179**, 5975–5986.

Soppa, J. (1999b) Transcription initiation in Archaea: facts, factors and future aspects, *Mol. Microbiol.* **31**, 1295–1305.

Thomm, M. (1996) Archaeal transcription factors and their role in transcription initiation, *FEMS Microbiol. Rev.* **18**, 159–171.

Primary Literature

Bartig, D., Lemkemeier, K., Frank, J., Lottspeich, F., Klink, F. (1992) The archaebacterial hypusine-containing protein. Structural features suggest common ancestry with eukaryotic translation initiation factor 5A, *Eur. J. Biochem.* **204**, 751–758.

Bartoschek, S., Vorholt, J.A., Thauer, R.K., Geierstanger, B.H., Griesinger, C. (2000) *N*-carboxymethanofuran (carbamate) formation from methanofuran and CO_2 in methanogenic archaea. Thermodynamics and kinetics of the spontaneous reaction, *Eur. J. Biochem.* **267**, 3130–3138.

Bechard, M.E., Chhatwal, S., Garcia, R.E., Rasche, M.E. (2003) Application of a colorimetric assay to identify putative ribofuranosylaminobenzene 5′-phosphate synthase genes expressed with activity in *Escherichia coli*, *Biol. Proced Online* **5**, 69–77.

Beifuss, U., Tietze, M., Baumer, S., Deppenmeier, U. (2000) Methanophenazine: structure, total synthesis, and function of a new cofactor from methanogenic Archaea, *Angew. Chem., Int. Ed.* **39**, 2470–2473.

Berghofer, Y., Klein, A. (1995) Insertional mutations in the hydrogenase *Vhc* and *Frc* operons encoding selenium-free hydrogenases in *Methanococcus voltae*, *Appl. Environ. Microbiol.* **61**, 1770–1775.

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D. et al. (1996) Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii, *Science* **273**, 1058–1073.

Buss, K.A., Cooper, D.R., Ingram-Smith, C., Ferry, J.G., Sanders, D.A., Hasson, M.S. (2001) Urkinase: structure of acetate kinase, a member of the ASKHA superfamily of phosphotransferases, *J. Bacteriol.* **183**, 680–686.

Buurman, G., Shima, S., Thauer, R.K. (2000) The metal-free hydrogenase from methanogenic Archaea: evidence for a bound cofactor, *FEBS Lett.* **485**, 200–204.

Caccamo, M.A., Malone, C.M., Rasche, M.E. (2004) Biochemical characterization of dihydromethanopterin reductase, a tetrahydromethanopterin biosynthesis enzyme in *Methylobacterium extorquens* AM1, *J. Bacteriol.* (in Press).

Chistoserdova, L., Vorholt, J.A., Thauer, R.K., Lidstrom, M.E. (1998) C_1 transfer enzymes and coenzymes linking methylotrophic bacteria and methanogenic Archaea, *Science* **281**, 99–102.

Cohen-Kupiec, R., Blank, C., Leigh, J.A. (1997) Transcriptional regulation in Archaea: in vivo demonstration of a repressor binding site in a methanogen, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1316–1320.

Cohen-Kupiec, R., Marx, C.J., Leigh, J.A. (1999) Function and regulation of glnA in the methanogenic archaeon *Methanococcus maripaludis*, *J. Bacteriol.* **181**, 256–261.

Creti, R., Ceccarelli, E., Bocchetta, M., Sanangelantoni, A.M., Tiboni, O., Palm, P., Cammarano, P. (1994) Evolution of translational elongation factor (EF) sequences: reliability of global phylogenies inferred from EF-1 alpha(Tu) and EF-2(G) proteins, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3255–3259.

Creti, R., Sterpetti, P., Bocchetta, M., Ceccarelli, E., Cammarano, P. (1995) Chromosomal organization and nucleotide sequence of the fus-gene encoding elongation factor 2 (EF-2) of the hyperthermophilic archaeum *Pyrococcus woesei*, *FEMS Microbiol. Lett.* **126**, 85–90.

Darcy, T.J., Hausner, W., Awery, D.E., Edwards, A.M., Thomm, M., Reeve, J.N. (1999) *Methanobacterium thermoautotrophicum* RNA polymerase and transcription in vitro, *J. Bacteriol.* **181**, 4424–4429.

Decanniere, K., Babu, A.M., Sandman, K., Reeve, J.N., Heinemann, U. (2000) Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon methanothermus fervidus, *J. Mol. Biol.* **303**, 35–47.

Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wiezer, A., Baumer, S., Jacobi, C. et al. (2002) The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and Archaea, *J. Mol. Microbiol. Biotechnol.* **4**, 453–461.

Deppenmeier, U., Muller, V., Gottschalk, G. (1996) Pathways of energy conservation in methanogenic Archaea, *Arch. Microbiol.* **165**, 149–163.

Duin, E.C., Madadi-Kahkesh, S., Hedderich, R., Clay, M.D., Johnson, M.K. (2002) Heterodisulfide reductase from methanothermobacter marburgensis contains an active-site [4Fe-4S] cluster that is directly involved in mediating heterodisulfide reduction, *FEBS Lett.* **512**, 263–268.

Ermler, U., Grabarse, W., Shima, S., Goubeaud, M., Thauer, R.K. (1997) Crystal structure of methyl-coenzyme M reductase: the key enzyme of biological methane formation, *Science* **278**, 1457–1462.

Fahrner, R.L., Ferry Cascio, D., Lake, J.A., Slesarev, A. (2001) An ancestral nuclear protein assembly: crystal structure of the Methanopyrus kandleri histone, *Protein Sci.* **10**, 2002–2007.

Ferry, J.G. (1997) Methane: small molecule, big impact [comment], *Science* **278**, 1413–1414.

Galagan, J.E., Nusbaum, C., Roy, A., Endrizzi, M.G., Macdonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D. et al. (2002) The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity, *Genome Res.* **12**, 532–542.

Gernhardt, P., Possot, O., Foglino, M., Sibold, L., Klein, A. (1990) Construction of an integration vector for use in the archaebacterium *Methanococcus voltae* and expression of a eubacterial resistance gene, *Mol. Gen. Genet.* **221**, 273–279.

Grabarse, W., Mahlert, F., Duin, E.C., Goubeaud, M., Shima, S., Thauer, R.K., Lamzin, V., Ermler, U. (2001) On the mechanism of biological methane formation: structural evidence for conformational changes in methyl-coenzyme M reductase upon substrate binding, *J. Mol. Biol.* **309**, 315–330.

Graupner, M., Xu, H., White, R.H. (2000) Identification of an archaeal 2-hydroxy acid dehydrogenase catalyzing reactions involved in coenzyme biosynthesis in methanoarchaea, *J. Bacteriol.* **182**, 3688–3692.

Hanzelka, B.L., Darcy, T.J., Reeve, J.N. (2001) TFE, an archaeal transcription factor in methanobacterium thermoautotrophicum related to eucaryal transcription factor TFIIEalpha, *J. Bacteriol.* **183**, 1813–1818.

Hao, B., Gong, W., Ferguson, T.K., James, C.M., Krzycki, J.A., Chan, M.K. (2002) A new UAG-encoded residue in the structure of a methanogen methyltransferase, *Science* **296**, 1462–1466.

Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., Stetter, K.O. (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont, *Nature* **417**, 63–67.

Ide, T., Baumer, S., Deppenmeier, U. (1999) Energy conservation by the H_2:heterodisulfide oxidoreductase from *Methanosarcina mazei* Go1: identification of two proton-translocating segments, *J. Bacteriol.* **181**, 4076–4080.

Jarrell, K.F., Bayley, D.P., Florian, V., Klein, A. (1996) Isolation and characterization of insertional mutations in flagellin genes in the archaeon *Methanococcus voltae*, *Mol. Microbiol.* **20**, 657–666.

Kessler, P.S., Blank, C., Leigh, J.A. (1998) The nif gene operon of the methanogenic archaeon *Methanococcus maripaludis*, *J. Bacteriol.* **180**, 1504–1511.

Kessler, P.S., Daniel, C., Leigh, J.A. (2001) Ammonia switch-off of nitrogen fixation in the methanogenic archaeon *Methanococcus maripaludis*: mechanistic features and requirement for the novel GlnB homologues, NifI(1) and NifI(2), *J. Bacteriol.* **183**, 882–889.

Kessler, P.S., Leigh, J.A. (1999) Genetics of nitrogen regulation in *Methanococcus maripaludis*, *Genetics* **152**, 1343–1351.

Klein, A.R., Fernandez, V.M., Thauer, R.K. (1995a) H_2-forming N^5,N^{10}-methylenetetrahydromethanopterin dehydrogenase: mechanism of H_2 formation analyzed using hydrogen isotopes, *FEBS Lett.* **368**, 203–206.

Klein, A.R., Hartmann, G.C., Thauer, R.K. (1995b) Hydrogen isotope effects in the reactions catalyzed by H_2-forming N^5,N^{10}-methylenetetrahydromethanopterin dehydrogenase from methanogenic archaea, *Eur. J. Biochem.* **233**, 372–376.

Lake, J.A., Henderson, E., Clark, M.W., Matheson, A.T. (1982) Mapping evolution with ribosome structure: intralineage constancy and interlineage variation, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5948–5952.

Langer, D.H.J., Thuriaux, P., Zillig, W. (1995) Transcription in Archaea: similarity to that in eucarya, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5768–5772.

Leigh, J.A. (2000) Nitrogen fixation in methanogens: the archaeal perspective, *Curr. Issues Mol. Biol.* **2**, 125–131.

Lie, T.J., Leigh, J.A. (2002) Regulatory response of *Methanococcus maripaludis* to alanine, an intermediate nitrogen source, *J. Bacteriol.* **184**, 5301–5306.

Macario, A.J., Lange, M., Ahring, B.K., De Macario, E.C. (1999) Stress genes and proteins in the archaea, *Microbiol. Mol. Biol. Rev.* **63**, 923–967.

Madadi-Kahkesh, S., Duin, E.C., Heim, S., Albracht, S.P.J., Johnson, M.K., Hedderich, R. (2001) A paramagnetic species with unique EPR characteristics in the active site of heterodisulfide reductase from methanogenic archaea, *Eur. J. Biochem.* **268**, 2566–2577.

Marc, F., Sandman, K., Lurz, R., Reeve, J.N. (2002) Archaeal histone tetramerization determines DNA affinity and the direction of DNA supercoiling, *J. Biologic. Chem.* **277**, 30879–30886.

Metcalf, W.W., Zhang, J.K., Apolinario, E., Sowers, K.R., Wolfe, R.S. (1997) A genetic system for Archaea of the genus *Methanosarcina*:

liposome-mediated transformation and construction of shuttle vectors, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2626–2631.

Meuer, J., Kuettner, H.C., Zhang, J.K., Hedderich, R., Metcalf, W.W. (2002) Genetic analysis of the archaeon *Methanosarcina barkeri* fusaro reveals a central role for Ech hydrogenase and ferredoxin in methanogenesis and carbon fixation, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5632–5637.

Muller, V., Ruppert, C., Lemker, T. (1999) Structure and function of the A_1A_0-ATPases from methanogenic Archaea, *J. Bioenerg. Biomembr.* **31**, 15–27.

Murakami, E., Deppenmeier, U., Ragsdale, S.W. (2000) Characterization of the intramolecular electron transfer pathway from 2-hydroxyphenazine to the heterodisulfide reductase from *Methanosarcina thermophila*, *J. Biol. Chem.* **276**, 2432–2439.

Murakami, E., Ragsdale, S.W. (2000) Evidence for intersubunit communication during acetyl-CoA cleavage by the multienzyme CO dehydrogenase/acetyl-CoA synthase complex from *Methanosarcina thermophila*. Evidence that the beta subunit catalyzes C—C and C—S bond cleavage, *J. Biol. Chem.* **275**, 4699–4707.

Ouhammouch, M., Dewhurst, R.E., Hausner, W., Thomm, M., Geiduschek, E.P. (2003) Activation of archaeal transcription by recruitment of the TATA-binding protein, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5097–5102.

Polycarpo, C., Ambrogelly, A., Ruan, B., Tumbula-Hansen, D., Ataide, S.F., Ishitani, R., Yokoyama, S., Nureki, O., Ibba, M., Soll, D. (2003) Activation of the pyrrolysine suppressor tRNA requires formation of a ternary complex with class I and class II lysyl-tRNA synthetases, *Mol. Cell* **12**, 287–294.

Praetorius-Ibba, M., Ibba, M. (2003) Aminoacyl-tRNA synthesis in archaea: different but not unique, *Mol. Microbiol.* **48**, 631–637.

Qureshi, S.A.B.P., Rowlands, T., Khoo, B., Jackson, S.P. (1995) Cloning and functional analysis of the TATA binding protein from sulfolobus shibatae, *Nucl. Acids Res.* **23**, 1775–1781.

Rasche, M.E., Wyles, S.A., Rosenzvaig, M. (2004) Characterization of two methanopterin biosynthesis mutants of *Methylobacterium extorquens* AM1 using a tetrahydromethanopterin bioassay, *J. Bacteriol.* **186**, (in Press).

Reeve, J.N. (1993) Structure and Organization of Genes, in: Ferry, J.G. (Eds) *Methanogenesis: Ecology, Physiology, Biochemistry & Genetics*, Chapman & Hall, New York, pp. 493–527.

Rospert, S., Linder, D., Ellermann, J., Thauer, R.K. (1990) Two genetically distinct methyl-coenzyme M reductases in *Methanobacterium thermoautotrophicum* strain Marburg and delta H, *Eur. J. Biochem.* **194**, 871–877.

Scott, J.W., Rasche, M.E. (2002) Purification, overproduction, and partial characterization of beta-RFAP synthase, a key enzyme in the methanopterin biosynthesis pathway, *J. Bacteriol.* **184**, 4442–4448.

Shima, S., Warkentin, E., Grabarse, W., Sordel, M., Wicke, M., Thauer, R.K., Ermler, U. (2000) Structure of coenzyme F420 dependent methylenetetrahydromethanopterin reductase from two methanogenic archaea, *J. Mol. Biol.* **300**, 935–950.

Shima, S., Warkentin, E., Thauer, R.K., Ermler, U. (2002) Structure and function of enzymes involved in the methanogenic pathway utilizing carbon dioxide and molecular hydrogen, *J. Biosci. Bioeng.* **93**, 519–530.

Slesarev, A.I., Mezhevaya, K.V., Makarova, K.S., Polushin, N.N., Shcherbinina, O.V., Shakhova, V.V., Belova, G.I., Aravind, L., Natale, D.A., Rogozin, I.B. et al. (2002) The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4644–4649.

Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K. et al. (1997) Complete genome sequence of methanobacterium thermoautotrophicum deltaH: functional analysis and comparative genomics, *J. Bacteriol.* **179**, 7135–7155.

Soppa, J. (1999a) Normalized nucleotide frequencies allow the definition of archaeal promoter elements for different archaeal groups and reveal base-specific TFB contacts upstream of the TATA box, *Mol. Microbiol.* **31**, 1589–1592.

Srinivasan, G., James, C.M., Krzycki, J.A. (2002) Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA, *Science* **296**, 1459–1462.

Tang, T.H., Rozhdestvensky, T.S., d'Orval, B.C., Bortolin, M.L., Huber, H., Charpentier, B., Branlant, C., Bachellerie, J.P., Brosius, J., Huttenhofer, A. (2002) RNomics in Archaea

reveals a further link between splicing of archaeal introns and rRNA processing, *Nucl. Acids. Res.* **30**, 921–930.

Thiru, A., Hodach, M., Eloranta, J.J., Kostourou, V., Weinzierl, R.O., Matthews, S. (1999) RNA polymerase subunit H features a beta-ribbon motif within a novel fold that is present in archaea and eukaryotes, *J. Mol. Biol.* **287**, 753–760.

Tumbula, D.L., Whitman, W.B. (1999) Genetics of *Methanococcus*: possibilities for functional genomics in Archaea, *Mol. Microbiol.* **33**, 1–7.

Watanabe, H., Gojobori, T., Miura, K., Watanabea, H. (1997) Bacterial features in the genome of *Methanococcus jannaschii* in terms of gene composition and biased base composition in ORFs and their surrounding regions, *Gene* **205**, 7–18.

Wich, G., Hummel, H., Jarsch, M., Bar, U., Bock, A. (1986a) Transcription signals for stable RNA genes in methanococcus, *Nucl. Acids. Res.* **14**, 2459–2479.

Wich, G., Sibold, L., Bock, A. (1986b) Genes for tRNA and their putative expression signals in methanococcus, *Systemat. Appl. Microbiol.* **7**, 18–25.

Woese, C.R., Kandler, O., Wheelis, M.L. (1990) Towards a natural system of organisms. Proposal for the domains archaea, bacteria, and eucarya, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576–4579.

Wood, G.E., Haydock, A.K., Leigh, J.A. (2003) Function and regulation of the formate dehydrogenase genes of the methanogenic archaeon *Methanococcus maripaludis*, *J. Bacteriol.* **185**, 2548–2554.

Xu, H., Aurora, R., Rose, G.D., White, R.H. (1999) Identifying two ancient enzymes in Archaea using predicted secondary structure alignment, *Nat. Struct. Biol.* **6**, 750–754.

Zhang, J.K., Pritchett, M.A., Lampe, D.J., Robertson, H.M., Metcalf, W.W. (2000) In vivo transposon mutagenesis of the methanogenic archaeon *Methanosarcina acetivorans* C2A using a modified version of the insect mariner-family transposable element Himar1, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9665–9670.

Zhang, J.K., White, A.K., Kuettner, H.C., Boccazzi, P., Metcalf, W.W. (2002) Directed mutagenesis and plasmid-based complementation in the methanogenic archaeon *Methanosarcina acetivorans* C2A demonstrated by genetic analysis of proline biosynthesis, *J. Bacteriol.* **184**, 1449–1454.

Zillig, W., Palm, P., Reiter, W.-D., Gropp, F., Puhler, G., Klenk, H.-P. (1988) Comparative evaluation of gene expression in archaebacteria, *Eur. J. Biochem.* **173**, 473–482.

Methods and Model Organisms in Embryogenomics: *see* Principles and Applications of Embryogenomics

Microarray-Based Technology: Basic Principles, Advantages and Limitations

Rumiana Bakalova[1], *Ashraf Ewis*[1,2], *and Yoshinobu Baba*[1,3,4]
[1] *National Institute of Advanced Industrial Science and Technology, Takamatsu, Japan*
[2] *El-Minia University, El-Minia, Egypt*
[3] *University of Tokushima, Shomachi, Tokushima, Japan*
[4] *Nagoya University, Nagoya, Japan*

1 Introduction in Functional Genomics and Proteomics 265

2 Microarray Technique – Historical Background and Basic Principles 267

3 Array Platforms and Fabrication 269
3.1 cDNA-based (Spotted) Arrays – Advantages and Drawbacks 270
3.2 Oligonucleotide-based Arrays (GeneChips and Codelink) – Advantages and Drawbacks 270
3.3 Fabrication of Spotted Arrays 271

4 Design of Microarray Experiment – Crucial Points 275
4.1 Choice of Reference (Control) Source 275
4.2 Labeling and Hybridization 276

5 Data Analysis, Validation, and Interpretation 278
5.1 Normalization of Microarray Data 278
5.2 Statistical Data Analysis 279
5.3 Microarray Data Validation 279
5.4 Data Mining and Interpretation 280

6 Major Technical Restrictions 280

7 Application of Microarrays in Molecular Cell Biology and Molecular Medicine 281

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

Bibliography 282
Books and Reviews 282
Primary Literature 283

Keywords

Genomics
Analysis of gene expression of cell, tissue, or organ under given conditions.

Proteomics
Analysis of proteins of cell, tissue, or organ under given conditions.

Pharmacogenetics
Analysis of variability in drug responses attributed to hereditary factors in different populations.

Pharmacogenomics
Analysis of genome and its products (RNA and proteins) related to drug response.

Messenger RNA (mRNA)
The class of cellular RNA that undergoes extensive editing and that contains the protein-coding sequences of genes and functions as informational intermediate between DNA and protein.

Complementary DNA (cDNA)
The DNA version of the messenger RNA that is produced after a PCR reaction using reverse transcriptase (RT) enzyme (RT-PCR).

Gene Expression
The cellular process by which the genetic information flows from gene to messenger RNA, and finally to the protein.

DNA Microarrays
DNA (from bacterial clones or synthesized oligonucleotides) attached to a surface in an ordered, predetermined fashion at extremely high density; allow the monitoring of gene expression for thousands of genes simultaneously in a single hybridization experiment based on two principles of nucleic acid hybridization: (1) DNA and RNA (isolated from tissue specimens or cell lines) specifically interact with their complementary sequence attached on the array surface, and (2) this interaction happens in proportion to the amount of mRNA in the mixture.

Spotted DNA Arrays
Composed of cDNA clones arranged robotically on a nonporous surface (usually glass microscope slide or nylon membranes).

Oligonucleotide-based DNA Arrays
Composed of short (~25 bp) or long (~60–70 bp) oligodeoxynucleotides synthesized *in situ* directly onto a glass wafer, using a modification of semiconductor photolithography technology.

In April 2003, the 50th anniversary year of the discovery of the double-helical structure of DNA, a high-quality and comprehensive sequencing of the human genome was completely accomplished, and a revolution in molecular cell biology and molecular medicine had begun. Many complex questions arise (e.g. how genes contribute to normal human development, individual variability, and common diseases) and they need complex answers. It was necessary to develop techniques that permit complex answers. Microarray technology is a revolution technology that allows the simultaneous assessment of the transcription of tens of thousands of genes rapidly, as well as of their relative expression between normal and injured cells. There is widespread hope that microarrays will significantly impact on our ability to explore the genetic changes associated with etiology and development of many diseases, and to discover new biomarkers for disease diagnosis and prognosis prediction, and new therapeutic tools.

The present article provides an overview of microarray technology with accents on its recent advantages and limitations. The first chapter is focused on the basic principles of microarrays, array platforms and fabrication, and advantages and restrictions of cDNA-based (spotted) and oligonucleotide-based arrays (GeneChips and Codelink). The second chapter describes the crucial points in microarray study design (e.g. choice of reference source, sample preparation and preservation, labeling, and hybridization). Finally, a brief description of microarray data normalization, mining, and validation is given at the end of the review, transferring to several basic web sites and software with detailed explanation of this most underappreciated challenge facing researchers working on microarray projects.

1
Introduction in Functional Genomics and Proteomics

All mammalian cells have an identical copy of one and the same genome. However, the human organism consists of different types of cells, realizing different functions. The cell variety is a result of the expression (transcription) of different genes from the identical genome, and subsequent translation of their genetic information into different types and quantity of proteins. Therefore, the status of contemporary genetics might be succinctly summarized as follows: many genes, few functions.

Only 1 to 2% of the human genome codes expressed genes via messenger RNA (mRNA), which is translated into protein. The rest of the genome consists

of repeated sequences, regulatory regions, or unique noncoding sequences with unknown function.

The Human Genome Project (HGP) was begun in 1990, and its goal is the complete mapping and understanding of all the genes of human beings. Two groups, The International Human Genome Sequencing Consortium and Celera Genomics, simultaneously reported the draft sequencing of human genome in February 2001 (www.genome.gov). In April 2003, the 50th anniversary year of the discovery of the double-helical structure of DNA, a high-quality and comprehensive sequencing of the human genome was completely accomplished. HGP covers about 99% of the human genome's gene-containing regions with a sequence accuracy of 99.99%. The number of genes in the human genome was revealed to be approximately 30 000, which is considerably less than the initial estimation. At present, all of the initial objectives of the HGP have been achieved at least two years ahead of expectation, and a revolution in biological research has begun.

Each gene can be present in different variants, called *polymorphisms* (mostly single-nucleotide polymorphisms). To date, 3×10^6 gene polymorphisms are described – a number that increases daily and is estimated to be greater than 11×10^6 for the whole human genome. Each individual is composed of a pair of inherited combinations of variants for each of the 30 000 genes, explaining the enormous genetic diversity. Only a fraction of these gene polymorphisms is important to human health and disease, and to find the important ones and to predict the diseases is a major challenge for the next decades. The discovery of microarray technique (allowing relatively inexpensive study of all genes expressed in the genome/transcriptome in parallel) is a great promise in this field.

Before proceeding to the description of the principles and design of microarray studies, let us introduce the major definitions in molecular cell biology and molecular medicine as functional genomics, proteomics, pharmacogenetics, and pharmacogenomics (Table 1).

Not all of the genes in human genome are used simultaneously. Depending on the developmental stage, age of the individual, cell type, organ, and environmental factors, a different set of genes is transcribed into mRNA. The mRNA expression analysis under defined conditions is called *functional genomics*. It allows a comparison of different sets of genes used in different conditions – in norm and pathology. Most studies that apply microarrays are functional genomic studies.

Not all of the transcribed genes result in proteins. Moreover, practically all proteins are modified after the first assembly of amino acids. It has been established that a protein derived from the same gene can be

Tab. 1 Major definitions in molecular cell biology and molecular medicine.

Term	*Definition*
Genomics	Analysis of gene expression of cell, tissue, or organ under given conditions
Proteomics	Analysis of proteins of cell, tissue, or organ under given conditions
Pharmacogenetics	Analysis of variability in drug responses attributed to hereditary factors in different populations
Pharmacogenomics	Analysis of genome and its products (RNA and proteins) related to drug response

altered in 10 to 20 different splits and 3D forms. Some proteins interact directly with DNA, leading to gene silencing. Almost all of the proteins interact with other proteins within biochemical pathways and form protein networks. The analysis of the proteins of a cell, tissue, or organ under given conditions is called *proteomics*. Recently, techniques similar to microarrays have been introduced in proteomic studies.

There is a long gap from the genotype to phenotype. If the genotype would automatically lead to a specific condition, so-called phenotype, all identical twins would have exactly the same homeostasis and diseases. Although they considerably resemble each other, they also differ in many ways. Thus, a 100% match is not there between genotype and phenotype because of environment–gene interactions. On the other hand, many conditions are a result of pathogenic mechanisms, and despite the different genetic makeup of the individuals, the same phenotype arises, resulting in development of one and the same disease (for example, hypertension, asthma, diabetes, etc.). In these so-called complex diseases, a constellation of different susceptibility and disease-modifying genes need to be present. Research has therefore failed to identify specific disease genes, for example, an ''asthma gene,'' a ''hypertension gene,'' a ''diabetes genes,'' and so on. To understand the functional aspects of the disease better and to bridge the long gap between genotype and phenotype, it is necessary to combine functional genomic and proteomic analyses.

For complex questions (e.g. how genes contribute to normal human development, individual variability, and common diseases), there might be complex answers. Thus, it is necessary to apply techniques that permit complex answers. Therefore, novel techniques are necessary to screen thousands of genes more rapidly and to generate new hypotheses. This is the role of high-throughput technology, like microarrays. Although this technology is new, it has already resulted in significant research advances and its use is spreading rapidly. As a hypothesis-generating approach, high-throughput methods can lead to the identification of a set of potentially interesting genes associated with a certain condition, so-called candidate genes. Microarray techniques, however, will not replace the classical hypothesis-driven research. Identified candidate genes have to be validated for their function and relevance by classical approaches. It is necessary to have in mind that each method of investigation has its limits and must be interpreted properly. For example, although several genes in a particular metabolic pathway may show changes at the mRNA level, the actual function of that pathway may not be affected if none of these genes is a rate-limiting factor. Therefore, these high-throughput genome-level technologies must be interpreted within the appropriate biological context.

This review summarizes the current state of microarray technology and presents examples of its application for translation of genome from laboratory to clinic. All about microarrays can be found in www.arrayit.com.

2 Microarray Technique – Historical Background and Basic Principles

DNA microarrays had first been described by Patrick Brown's Lab in 1995. Since that original report, the number of microarray publications has increased exponentially, from less than 150 publications in the

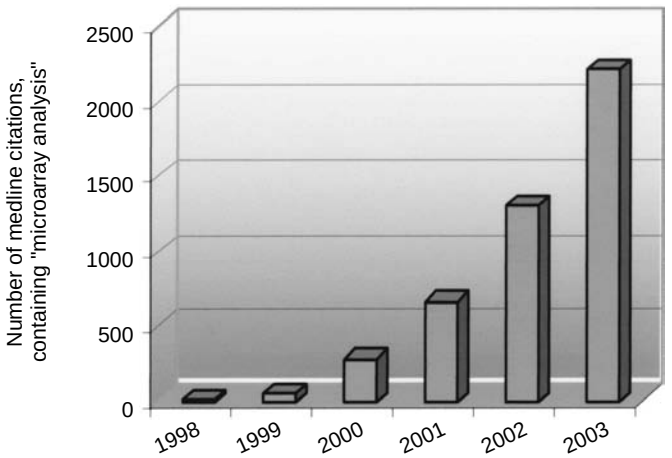

Fig. 1 Pubmed citations, containing "microarray analysis."

first 4 years (1995–1999) to about 2000 publications in 2003 (Fig. 1).

DNA microarrays – also called *gene chips* – consist of DNA attached to a surface in an ordered, predetermined fashion at extremely high density. More than 20 000 genes (almost a half of expressed human genes) can be arrayed on a surface with the size of a glass microscope slide. Microarrays allow the monitoring of gene expression for thousands of genes simultaneously in a single hybridization experiment because of two basic principles of nucleic acid hybridization: (1) DNA and RNA specifically interact with their complementary sequence attached on the array surface, and (2) this binding happens in proportion to the amount of mRNA (termed as *mRNA abundance*) in a mixture.

Although microarray technology has been originally designed for depositing DNA onto solid support for gene expression profiling (expression arrays), it has been extended to DNA copy number analysis (comparative genomic hybridization arrays), and, recently, a similar technology has been pioneered for protein and antibody spotting (Panorama Antibody Microarray Cell Signaling, www.sigma-aldrich.com).

Figure 2 provides an overview of a typical DNA microarray experiment from the selection of bacterial clones for manufacturing of the microarrays and nucleic acid extraction from tissue specimens or cell lines to *in silico* analysis, mining, and validation of the microarray data obtained. Optional steps, such as laser capture microdissection (LCM) and RNA amplification are also indicated.

The aim of the typical microarray experiment is to measure the amount of a given mRNA species (transcribed genes) of a tissue or cell type. For this purpose, the mRNA (isolated from the analyzed sample, e.g. drug-treated patients or cells – case sample, or untreated patients or cells – control sample) is usually transformed by reverse transcription into complementary DNA (cDNA), which is more stable. cDNA (or mRNA itself) from the sample is applied to the array surface and is further hybridized for several hours to the microarray. Each specific mRNA sequence (corresponding to the expressed gene) is able to bind only to a single complementary

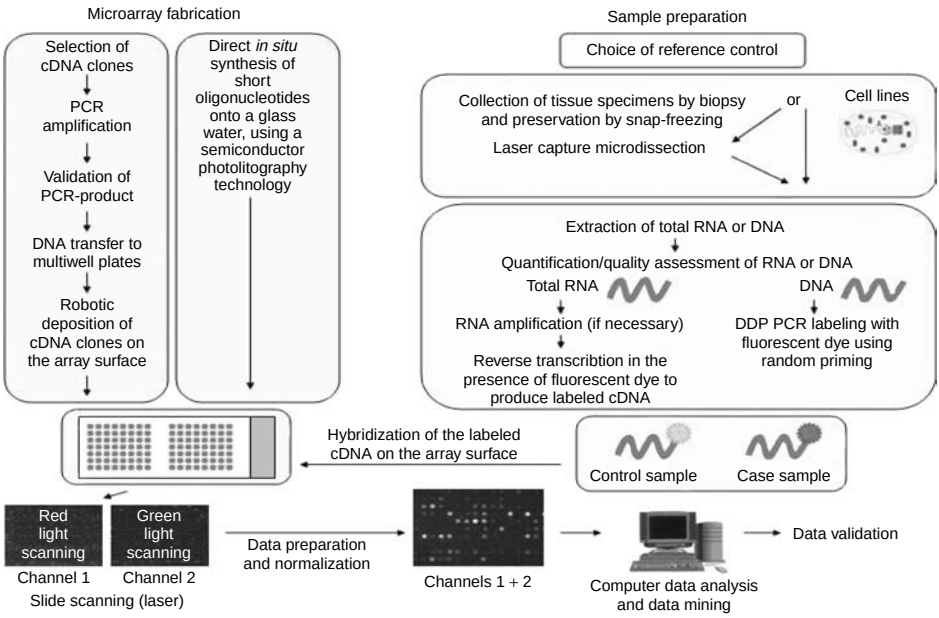

Fig. 2 Microarray study design.

oligonucleotide sequence attached on the array surface. If the mRNA sequence matches with the oligonucleotide sequence on a specific spot of the microarray, hybridization occurs. Direct information of how much of each gene is expressed in the cells is obtained by measurement of the amount of cDNA (corresponding to mRNA or expressed gene) bound to each spot on the array. The more abundant a given cDNA (or mRNA) is in a sample, the more it binds to its immobilized complement will occur on the array surface. The control and case samples are labeled with different fluorophores for detection. On each of the up to 450 000 spots, different binding intensities occur, depending on the concentration of the different cDNAs (equal to mRNAs or expressed genes) in the case and/or control samples tested. Thus, by comparison of the hybridization of cDNA from a control sample with that from a case sample, the relative amounts of all the genes present on the array and expressed in both samples can be measured.

In contrast to the conventional Northern blot, which analyses 1, 2, or up to 20 mRNAs, a microarray technique allows the simultaneous analysis of the expression levels of hundreds, thousands, or even tens of thousands of genes in a single experiment. The latest chips carry up to 450 000 spots for the analysis of more than 20 000 genes and gene sequences on a small glass slide (1.2 × 1.2 cm).

3 Array Platforms and Fabrication

Production of arrays begins with the selection of the probes to be printed on the array surface. These are often chosen directly from gene databases (e.g. GenBank, dbESt, and UniGene – all available on http://www.ncbi.nlm.nih.gov). Two main

methods are used to generate microarrays, and, currently, there are two main types of microarrays: cDNA arrays (also called *spotted arrays*) and oligonucleotide arrays (also called *GeneChips*).

3.1 cDNA-based (Spotted) Arrays – Advantages and Drawbacks

Spotted arrays had been originally developed in the Patrick Brown's Laboratory at Stanford University. They are composed of cDNA clones (selected from public databases and PCR-amplified) robotically arranged on a nonporous surface. A typical cDNA array is printed onto 30 × 15 mm glass microscope slide by a computer-controlled robotic cantilever arm; each spot is about 50 to 150 μm in diameter. cDNA clone sets of 15 000 to 20 000 mouse and 40 000 human genes are available for in-house arrays in addition to ready-made commercial chip sets. About 80% of all human genes are identified (the number increases daily) and can be queried on about two slides printed at high density. Other flexible and porous surfaces such as nylon membranes can be used as alternatives to glass. This approach allows the construction of arrays spotted with up to 80 000 spots.

The spotted arrays possess several advantages:

1. Affordability – they can be manufactured in-house.
2. Versatility – project-specific arrays (e.g. cancer arrays, immunoarrays, chromosome-specific arrays) can be custom designed.
3. Applicability to the simultaneous analysis of two different biological samples (e.g. treated vs untreated patient samples or cell lines, normal tissue vs malignant tissue, tumors of different stages vs a common internal reference), using two or more different fluorophores.
4. Finally, they provide the opportunity for discovery of novel genes since expressed sequence tags of unknown function can be spotted on the DNA arrays in addition to well-characterized genes.

The major restriction of spotted arrays is their inability to measure absolute levels of gene expression in biological samples, a measurement that GeneChips can provide. Several drawbacks are related to clone collections selected for spotting:

1. Clone collections must be sequence verified, a process that can be costly and time consuming.
2. Clone annotation needs to be updated regularly as new information becomes available in public databases.
3. Clones must be periodically reamplified by polymerase chain reaction (PCR), as the spotting process consumes DNA.

3.2 Oligonucleotide-based Arrays (GeneChips and Codelink) – Advantages and Drawbacks

Affymetrix Inc. (Santa Clara, CA, USA) owns a registered trademark – GeneChip®, which refers to its high-density oligonucleotide-based DNA arrays. However, in some articles appearing in professional journals, popular magazines, and on the WorldWideWeb, the term "gene chip(s)" has been applied, generally referring to microarray technology. Affymetrix GeneChips® are manufactured by synthesizing of *short oligonucleotides* (about 25 bp) *in situ* directly onto a glass surface, using a modification of semiconductor photolithography technology.

These arrays provide the advantage of an integrated platform, in which analytical tools are provided to the user in addition to the arrays themselves. The choice of microarray platform remains largely a user's personal preference, although it is often directed by budget considerations and local access and availability. These privileges are the reason for Affymetrix ChenChips® to be widely used in cancer research. The main restrictions of Affymetrix GeneChips® are in the impossibility to compare two biological samples directly on the same array and in their high cost that can be prohibitive to small research laboratories.

Agilent Technology (Palo Alto, CA, USA) offers arrays based on *longer oligonucleotides*. They are manufactured *in situ* by use of ink-jet printing technology. Many investigators use longer (60–70 bp) *oligonucleotide arrays* (e.g. Codelink arrays, Amersham Bioscience, Piscataway, NJ, USA) that are made in a way similar to cDNA arrays. However, these longer oligonucleotide arrays have an advantage that they do not require PCR-amplification and clone-insert purification before arraying. Several technical issues are recurrent challenges, as the optimization of slide surface and oligonucleotide attachment, and the design of the most specific oligonucleotide probe(s) for each gene. Cross-platform comparisons often yield significantly nonoverlapping results, and the source of these discrepancies is unknown.

Other techniques for measurement of gene expression that involve large-scale sequencing (such as serial analysis of gene expression – SAGE, and massively parallel signature sequencing – MPSS) can be more sensitive. These methods involve highly efficient identification of the expressed genes by sequencing of short fragments and counting of cDNA clones that correspond to mRNA expressed in a particular cell or tissue. However, neither technique is as rapid as a microarray, and, therefore, they are less likely to be useful in clinical practice.

This review will focus mainly on spotted (cDNA) microarrays. Excellent reviews of Affymetrix technology are available in Bibliography.

3.3 Fabrication of Spotted Arrays

The spotted microarrays are most commonly fabricated on poly-L-lysine-coated microscope slides (or surfaces of similar dimension). Other surface coatings that are also available include various silane derivatives, acrylamide gels, nitrocellulose, and nylon. Some of the microarray surfaces require the use of modified oligonucleotides or PCR-products for efficient attachment, while others rely solely on molecular charge to maintain binding of nucleic acid to the surface. The DNA fragments are cross-linked by UV to the matrix. After fixation, residual amines on the slide surface react with succinic anhydride to reduce the positive charge at the surface.

Several commercial robotic machines are currently used for fabrication of spotted arrays, for example, MicroGrid (BioRobotics, Cambridge, UK), GMS-417 (Genetic Microsystems, Woburn, MA, USA), OmniGrid (GeneMachines, San Carlos, CA, USA), and PixSys PA series (Cartesian Technologies, Irvine, CA, USA).

Several techniques are applied to replace DNA probe on the slide surface: Pin-to-Pin, Pin-and-Ring, Quill or Slit Pin, and Jetting (Piezo Electric, Inc Jet, Bubble Jet).

The *pin-to-pin technology* (Fig. 3) has the longest history of use as a spotting tool. Briefly, a series of pins is aligned on a

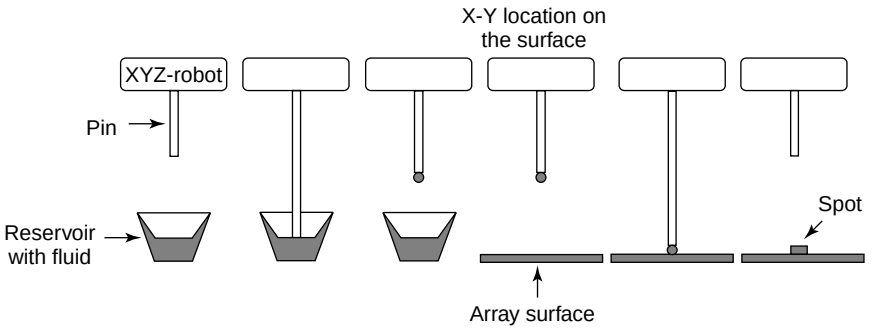

Fig. 3 Pin-to-pin technology.

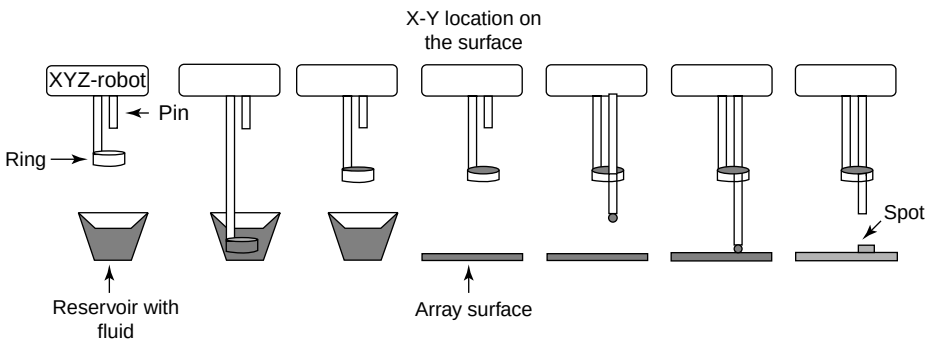

Fig. 4 Pin-and-ring technology.

head such that all the pins can be dipped into fluid reservoirs simultaneously (e.g. 96 pins, set in an array head, each pin at 9-mm center-to-center, for use with standard 96-well microplates). The pin-head module is attached to a 3D (x–y–z) robotic arm to move "into" and "out" of the wells, and to make contact with the spotting (array) surface. After dipping into the wells, the pins are lifted, and a small aliquot of fluid hangs on the tip of each pin by surface tension. These aliquots are transferred onto a slide surface through ensuring a pin–surface contact.

For fabrication of spotted microarrays on the microscopic slide surface, it is practically impossible to use more than 12 to 16 pins simultaneously, and the rate of the spotting is relatively slow. A crucial pint is the evaporation of the fluid aliquots during the time of transportation of the sample from the wells to the slide surface. This can provoke differences in the amount of fluid spotted and can introduce undesired variations into the arrays. Pin-to-pin alignment and consistency of pin dimensions are also both crucial points because of the necessity to place very small spots very close to each other, and to make at least a semiquantitative analysis of the hybridization to each spot.

The *pin-and-ring technology* (Fig. 4) has been specifically developed for fabrication of spotted DNA microarrays containing highly consistent elements. The key mechanical component consists of an open ring, which is oriented parallel to the wells (fluid reservoirs) and slide surface. The ring is fixed to a vertical rod, running perpendicular to the wells and slide surface.

A vertical pin is centered on the ring. Both the ring rod and the pin are attached to control devices so that each part can be moved separately in the z-axis. Both parts are kept in constant relation to each other in the $x-y$ plane. The spotting procedure includes several steps: the ring is dipped into the well (containing DNA solution) – the ring is small enough to fit easily into the well of either 96- or 384-well microplates; when the ring is lifted, it takes an aliquot of sample, which is held in the center of the ring by surface tension; the pin-ring device moves to any desired location in the $x-y$ plane and selects the place for spotting on the underlying surface; the pin is moved down through the ring; when the pin passes through the ring, a portion of the solution is transferred from the interior ring meniscus to the bottom of the pin – a pendant drop is formed on the pin tip; the pin continues to move down until the solution on the pin tip makes contact with the slide surface; the pin is then lifted; the combined forces of gravity and surface tension deposit the solution on the slide surface as a spot.

The pin-driving process can be repeated many times so that a very large number of similar spots can be created from a single moving ring. This consistency of spotting is the most important feature of the system. The spot size is dependent on pin dimensions and varies from 50 to 500 μm in diameter. It can be controlled by available hardware.

Special mechanisms have been designed to achieve a balance between the need to move the pin in z-axis rapidly, and to minimize the extent of contact between the pin and the slide surface. Thus, this technique is applicable for a wide variety of surface substrates – from solid glass to soft membranes and gels. The system appears to function well with a wide variety of fluids, ranging from volatile organic solvents to viscose aqueous solutions.

The *quills or slit pins* (Fig. 5) have been one of the most widely used devices for fabrication of spotted arrays. A quill is a solid pin with a slit at the end. The standard quill design looks very much like two tines of a fork, with a capillary between them. There are several variations of quills, but they all have the same principle of orientation. The quill is dipped into a fluid reservoir and the fluid is drawn up into the slit by capillary tension. The quill is moved to a desired position (in the $x-y$ plane) on the slide surface, then toward the slide surface (z-axis), and upon impact, the drop of fluid is placed on the surface. The quill

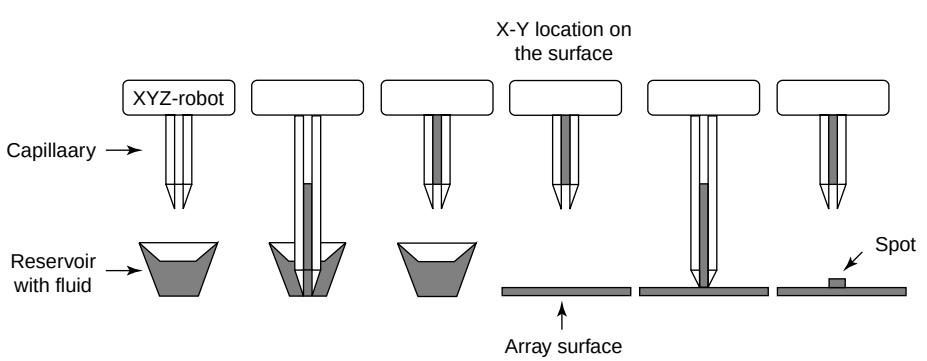

Fig. 5 Quill or slip pin technology.

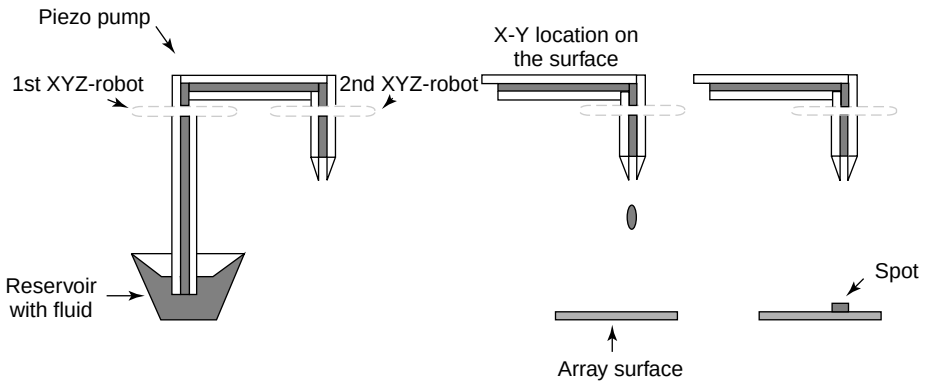

Fig. 6 Jetting technology.

is then withdrawn, moved to a subsequent location, and the process is repeated.

One of the advantages of the quill approach is that many quills can be attached to a single spotting head. One can spot 16 or more probes onto a surface simultaneously and without having to go back to the fluid reservoirs after each spotting. Therefore, this method enables rapid spotting. The quill has a relatively small void volume, and a little material is wasted. However, once the volume of the fluid in the quill becomes low, the amount of the spotted fluid begins to diminish with each subsequent spotting; this can compromise the quality of the arrays. The quill is also not highly effective for spotting of viscous solutions as a result of their difficult movement "into" and "out of" the capillary. The restriction of this technology is also due to the fact that it should be employed in a sterile environment because dust particles in the air or on the spotting surface can cause plugging of the quill. This characteristic of the quill limits its use to surfaces that are nonfibrous (as gels and membranes).

The *jetting technologies* (Fig. 6) had been originally developed for ink printing, and, nowadays, they are applied in microarray fabrication. They are based on syringes or piezo pumps.

The syringes offer a possibility of extremely fast jetting rates and exact repeatable volume deposition. However, the minimum volume they can dispense is about two orders of magnitude greater than other spotting techniques described above.

The piezo pumps are unidirectional pumps attached to flexible capillary tubing. The end of the tubing needs to be transferred from fluid to fluid by a 3D (x–y–z) robot. A second 3D (x–y–z) robot moves the other end of the tubing to a location above the surface onto which a drop of fluid has to be deposit. The pumps work with supply and wash solutions.

The piezo pumps offer a high speed and an ability to operate in any environment. However, there are several limitations: they cannot work with fluids containing long molecules, as well as with viscous fluids, because the glass heads easily become clogged; the spotting head is made of fragile glass that needs frequent repairs; this method of deposition often gives artifacts known as "satellite" spots that can compromise the quantitative analysis; a large volume of fluid is necessary to operate, and the void volume of fluid

is also high (100–500 μL). All these limitations make this technique less usable for microarray spotting.

4 Design of Microarray Experiment – Crucial Points

The most important factor for effective and successful microarray-based research projects is the careful experimental design, requiring a careful choice of the reference (control) source and unification of the experimental conditions. Even small variations in the conditions can induce significant changes in gene expression. For microarray analyses of human tissues, these variations include tissue-preservation methods, postmortem interval, dissection methods, and RNA quality. Frozen tissues usually provide the best material, although ethanol fixation can be also used. It is necessary to keep in mind that the differences in gene expression between samples may be due to differences in the genetic background (ethnicity). Other general sources of variability include cDNA amplification methods, probe labeling, hybridization conditions, and washing procedures. However, when these issues are taken into account in the experimental design and proper technique is applied, microarray experiments provide reliable data on gene expression.

4.1 Choice of Reference (Control) Source

One of the most critical decisions involves the choice of reference (control) sample when using dual-channel DNA arrays.

Several limitations have to be borne in mind during the design of a microarray-based project, especially in the clinic.

For example, for studies in cancer etiology and progression, the better choice is to compare a tumor (case) sample with a healthy (control) sample directly. It is ideal to obtain the healthy tissue from the same patient (case–control study). However, this possibility is often problematic and restricts the feasibility of this type of design. In this case, it is better to choose

Tab. 2 Labeling protocols for microarray experiments (according to Macgregor, P.F. (2003)).

Amount of total RNA	*Labeling mode*	*Protocols and kits*
>10 μg total RNA	Direct labeling (no amplification required)	Bittner, M. (2003) www.microarrays.ca
4–10 μg total RNA	Indirect labeling (no amplification required)	Randolph, J. et al. (1997) www.microarrays.ca
<10 μg total RNA	Amplification required before labeling:	
	-PCR-amplification	Vernon, S. et al. (2000) Livesey, F. et al. (2000) BD Atlas SMART Fluorescent Probe Amplification
	-Linear amplification	Eberwine, J. et al. (1992) Wang, E. et al. (2000) Arcturus RiboAmp Ambion MessageAmp aRNA

a comparison between tissue specimens obtained from the same patient before and after chemotherapy. This design is also limited by the impossibility to obtain sufficient amount of high-quality necrosis-free tumor tissue after chemotherapy. Because of these reasons, there is now a global trend to use a common internal reference in large microarray projects as the investigation of gene expression profiles of tumor specimens obtained from many different patients. This design increases the challenge to obtain global expression profiling across a large sample cohort, and it is already successfully applied by several groups. Commercially available reference RNA is often a good choice because of wide-gene-expression representation (e.g. Stratagene – La Jolla, CA, USA, and BD Bioscience Clontech – Palo Alto, CA, USA) and virtually unlimited availability. Finally, this approach allows easier comparison of microarray data between different research groups, thus providing the possibility of more global studies.

4.2 Labeling and Hybridization

Once the choice of reference source is made and the number of replicates decided, the actual labeling protocol depends predominantly on the amount of sample RNA (Table 2).

In studies using cell lines, RNA availability is usually not a problem. However, great care should be taken to ensure maximum reproducibility and stability in cell culture conditions, especially if RNA has to be isolated from cells during a long time interval.

In studies using human tumor tissues, RNA availability can be problematic because of the impossibility to obtain a sufficient amount of sample tissues every time.

Despite the amount of sample RNA, its high quality (degradation-free RNA) is crucial for successful microarray experiments. Assessment of RNA quality has been greatly facilitated by the use of microcapillary-based devices (e.g. Agilent Bioanalyzer, Agilent Technologies, Palo Alto, CA, USA), which can evaluate RNA quality and quantity using a little amount (~200 ng) of total RNA. More traditional techniques (such as agarose gel electrophoresis and spectroscopy) can also be used when RNA quantity is not limiting.

There are basically two labeling techniques, depending on the amount of RNA prepared for microarray analysis: direct and indirect labeling.

If the amount of total RNA available is in the range of 10 to 20 μg per experiment, direct labeling is the simplest choice. Direct labeling involves direct incorporation of Cy3- or Cy5-labeled dCTP during reverse transcription of RNA to cDNA. Usually 10 μg of total RNA are sufficient to obtain high-quality microarray signals using direct labeling.

If the amount of total RNA is slightly lower (in the range of 4–10 μg per experiment), indirect labeling can be used. In this case, the labeling proceeds in two steps: (1) a modified nonfluorescent nucleotide (e.g. amino-allyl dUTP) is incorporated into cDNA during the reverse transcription step, and (2) subsequent coupling to fluorophores is performed in a separate reaction.

Detailed protocols for direct and indirect labeling can be found in www.microarrays.ca.

If the amount of the sample RNA is only submicrograms (e.g. from limiting clinical samples, such as fine needle biopsies or LCM specimens), RNA amplification is

Tab. 3 Microarray data analysis software (according to Macgregor, P.F. (2003)).

Software package	*Company*	*Web site*
Acuity[R]	Axon Instruments (commercial)	www.axon.com/GN_Acuity.html
Array-Pro Analyzer	Media Cybernetics (commercial)	www.mediacy.com/arraypro.htm
ArrayStat[TM]	Imaging Research (commercial)	www.imagingresearch.com/products/AST.asp
DNA-Chip Analyser	Wing Wong and Cheng Li of the Harvard School of Public Health (free)	www.dchip.org/
F-Scan	National Institute of Health, Bethesda, MD (free)	http://abs.cit.nih.gov/fscan/
Gene Cluster	Whitehead Institute for Genomic Research (free)	www.genome.wi.mit.edu/cancer/software/software.html
GeneLinker[TM] Gold	Molecular Mining Corporation (commercial)	http://microarray.genelinker.com/products.htm#GeneLinkerGold
GeneSight[TM]	Biodiscovery (commercial)	www.biodiscovery.com/genesight.asp
GeneSpring	Silicon Genetics (commercial)	www.silicongenetics.com/cgi/SiG.cgi/Products/GeneSpring/index.smf
Genesis	Institute for Biomedical Engineering, Graz University of Technology (IBMT-TUG) (free)	http://genome.tugraz.at/Software/GenesisCenter.html
GeneTraffic[TM]	Iobion Informatics, Stratagene (commercial)	www.iobion.com
J-express Pro	MolMine (commercial)	www.molmine.com/frameset/frm_jexpress.htm
MAExplorer	National Cancer Institute, Laboratory of Experimental and Computational Biology (free)	www.lecb.ncifcrf.gov/MAExplorer/
Partek Software Suites	Partek (commercial)	www.partek.com/html/products/products.html
TIGR Array Viewer	The Institute for Genomic Research (free)	www.tigr.org/software
Xpression NTI[TM]	Informax Inc. (commercial)	www.informaxinc.com/solutions/expression/index.html

required before to proceed to the labeling. This strategy now enables the molecular profiling of biological samples down to few cells. Two amplification approaches have been used and commercial kits are available: PCR-based amplification (e.g. from BD Bioscience Clontech) and linear amplification (e.g. from Arcturus, Mountain View, CA, USA). In the case of necessity of RNA amplification, microarray results need more extensive validation. Moreover, the use of RNA amplification significantly increases the overall time and cost of molecular profiling of the samples.

After fluorophore incorporation, the fluorescent cDNA sample (probe) is hybridized to the microarray (target) at a temperature determined by the hybridization buffer used, in the presence of nonspecific competitors (e.g. yeast tRNA and salmon sperm DNA). The microarray is then washed under stringent conditions to remove nonspecific binding and is air-dried. The dried array is then scanned using a confocal laser scanner and the images obtained are quantified using image acquisition software.

The labeling and hybridization procedures, described above, are typical for the expression arrays.

A protocol very similar to the one used for expression arrays can be used for comparative genomic hybridization (CGH) arrays. In CGH arrays, the genetic material hybridized to the array is not RNA but genomic DNA, and the purpose is to investigate changes in gene copy numbers in tumor samples. Genomic DNA is extracted from tumor samples or normal tissue, digested with a restriction enzyme, such as EcoR1, and fluorescently labeled with Cy3 or Cy5 using random priming. Labeled DNAs are cohybridized to the microarray in the presence of blocking agent, usually Cot-1 DNA, yeast tRNA, and poly(dA-dT). After hybridization and stringent washes, the arrays are scanned and quantified in a process essentially identical to that described for expression arrays. CGH array experiments can be carried out using the same spotted arrays applied for expression analysis, thus providing the opportunity of a direct comparison on the same genes between gene copy numbers and expression levels. Alternatively, BAC arrays can be used, either manufactured in-house or commercially available, for example, from Spectral Genomics (Houston, TX, USA).

5
Data Analysis, Validation, and Interpretation

Perhaps the most underappreciated challenge facing researchers working on microarray projects is the extremely large quantity of data generated by this technology and the choice of an appropriate and effective database and data analysis system. In the early days of microarray technology, the basic data analysis was carried out using standard spread-sheet software, such as Microsoft Excell, or programs developed in-house. At present, there are a number of companies offering software for microarray data analysis. This is either offered free of charge on the Internet by academic institutions or it is commercially available (Table 3).

5.1
Normalization of Microarray Data

The first step in microarray data processing and analysis is the normalization. Normalization is used to combine the data from multiple microarray experiments into a single data set

for analysis. It is absolutely necessary to account for and to minimize the artifacts and experimental variations in the calculation of gene expression ratios. The normalization is a complex and often controversial issue in microarray data analysis, and a discussion of the different strategies is beyond the scope of this review. Briefly, there are two main approaches to normalization: (1) *intensity-dependent normalization*, in which a normalization factor is calculated on the basis of all spots on the array, and (2) *intensity-independent normalization*, which uses normalization function to equalize both channels (as Locally Weighted Scatter Plot Smoother, LOWESS – http://stat-www.berkeley.edu /users/terry/zarry/Html/). The results obtained after normalization are generally displayed as normalized expression ratios, where each row represents one spot on the array and each column represents one hybridization experiment. If some genes are represented by several spots on the array, it is possible to combine and average the expression ratios obtained for these different spots.

5.2 Statistical Data Analysis

In traditional data analysis, there are few variables and great number of replicates for each data-point. In microarray data analysis, there are many variables and a small number of replicates for each data-point. Microarray data analysis is typically carried out by one of two approaches – supervised or unsupervised (or a combination of both).

In the *unsupervised statistical analysis*, the data are analyzed without *a priori* assumption about the identity, characteristics, or clinical attributes of the specimens analyzed. Typical examples of unsupervised analysis in microarray data are 2D hierarchical clustering (software is now available – http://rana.lbl.gov), K-means clustering, self-organization maps (SOM), and binary tree–structured vector quantization (BTSVQ). Unsupervised clustering of microarray data is widely used in tumor classification, for instance, in leukemia, lung cancer, ovarian cancer, colon cancer, prostate cancer, or breast cancer.

The *supervised statistical analysis* of microarray data integrates some of the biological and clinical attributes available for the specimens analyzed. Typical examples of supervised analysis are significance analysis of microarrays (SAM) and prediction analysis of microarrays (PAM). Supervised data analysis is widely used for identification of genes capable of stratifying tumor samples based on one particular attribute, such as clinical outcome or response to treatment.

Detailed explanation of microarray data normalization and statistical analysis can be found in several books and reviews cited in Bibliography.

5.3 Microarray Data Validation

Microarray data analysis is a multistep process that uses sophisticated statistical analysis tools and is prone to errors and sometimes to overinterpretations. In addition, quality issue in the raw microarray data, for example, high background or low reproducibility, as well as nonlinearity of microarray signal will significantly impact on the final results. Finally, with spotted arrays, misidentification of probes can also lead to errors. Validation of gene expression data by conventional techniques is

essential. Usually, the validation of microarray data is carried out first at the RNA level by Northern blot, semiquantitative PCR, real-time quantitative (RT)-PCR, or *in situ* hybridization (ISH) on tissue sections or high-density microarrays (TMA). The microarray data can be also validated on the protein level by immunoblot analysis. Owing to the nonquantitative nature of immunoblotting, validation by other methods is required. An excellent example of such an extensive validation is presented in the study of Beer and colleagues, where the authors validated their microarray results both at the RNA level by Northern blot and at the protein level by IHC on tissue microarrays.

The development of high-throughput screening and validation technologies is an important priority because the number of genes identified through microarray studies rises exponentially.

5.4 Data Mining and Interpretation

Once gene data sets are identified and validated, data mining and interpretation is the ultimate challenge in microarray studies. It is not a trivial task to determine which of the hundreds of potentially differentially expressed genes identified are biologically relevant to the problem being investigated. Several bioinformatics tools and public databases provide information on the biological function, cellular localization, and gene ontology of the selected gene of interest: GenBank (www.ncbi.nlm.nih.gov), UniGene (http://ncbi.nlm.nih.gov/UniGene), PubMed (www.pubmed.com), US National Center for Biotechnology Information (NCBI) Databases (www.ncbi.nlm.nih.gov), Stanford Online Universal Resource for Clones and ESTs (SOURCE) (http://source.stanford.edu), SWISS SPOT (http://ca.expasy.org/sprot/). Another computer application of interest is GenMAPP, available online at www.genmapp.org. GenMAPP allows a visualization of gene expression data such as fold changes in expression ratios on maps representing biological pathways. Similarly, Onto-Express (www.openchannelfoundation.org/projects/Onto-Express/) correlates the identified genes with their functional and biological characteristics. Finally, literature searches can be greatly facilitated by using Internet text-mining tools, such as MedMiner, available at http://discover.nci.nih.gov/textmining/filters.html.

However, it is important to keep in mind that the complexity of the cellular machinery is extraordinary, and it is impossible to provide a biological explanation for every one of the genes identified by microarray analysis only with one simple click on the computer mouse. In addition, incorrect gene annotation in databases is a significant problem. Even when the genes identified in microarray analysis are correctly annotated, fully identified, and have a known function, many of these genes will belong to several biological pathways involved both in normal cellular processes and pathogenesis. In this context, downstream functional analysis using classical molecular biology approaches will remain paramount for the thorough interpretation of microarray findings.

6 Major Technical Restrictions

In order to perform genome-wide expression profiling, it is necessary to have *genome-wide microarrays*. In the case of

mouse and human, such whole genome arrays are on the horizon, although not yet available. This situation is likely to be resolved in the next two to three years, given the maturity of the genome and EST-sequencing projects in the mouse and humans. Certainly, the number of genes that can be investigated using microarrays increases all the time. It is likely that improvements in deposition or synthesis of DNA probes will allow a suitable density to be achieved, avoiding the requirement of hybridizing multiple arrays in order to achieve full genome coverage. The advent of spotted oligonucleotide arrays and related technologies also opens up the possibility of exon-specific probes for the detection of tissue-specific (or cell-specific) splice variants.

Perhaps one of the basic restrictions in the use of microarrays is *the limiting amount of RNA* available from standard tissue or cell specimens. The embryonic dissections used in developmental biology are a good example. A whole 7.5-dpc mouse embryo yields ~0.5 μg of total RNA. Usually, the conventional microarray hybridization analysis requires 5 to 100 μg of total RNA. It is clear that specific embryonic tissues or organs, especially from early stages, will yield only a fraction of the required target RNA. Embryo-sorting techniques have been employed, but target amplification appears to be the most widely applicable solution. A number of amplification protocols have been described that allow the generation of sufficient target from small amounts of staring RNA. It is not clear which protocol offers the "best" solution of the problem of limiting RNA because different scientists have different criteria for the assessment, including simplicity, reliability, and cost.

The last category of potential technical limitations of microarray analysis is closely related to *the possibility for identification of those incredibly rare transcripts* that encode those critical regulators of cellular homeostasis and development. A combination of variations in transcript abundance and cellular heterogeneity makes it difficult to perform expression profiling in standard tissue explants in a useful fashion. Subdissection and cell sorting are usually necessary; this makes labeling and hybridization practically impossible.

7 Application of Microarrays in Molecular Cell Biology and Molecular Medicine

What can be measured by microarrays?

Microarray technology can be used for three main applications:

1. Gene expression profiling – mRNA extracted from a biological sample is applied to the microarray. The result reveals the level of expression of tens of thousands of genes in the sample. This result is known as a gene expression "profile" or "signature."
2. Genotyping – DNA, extracted from a biological sample, is amplified by a PCR and applied to the microarray. The genotype for hundreds or thousands of genetic markers across the genome can be determined in a single experiment. This approach has considerable potential in disease risk assessment, both in research and clinical practice.
3. DNA sequencing – DNA extracted from a biological sample is amplified and applied to specific "sequencing" microarrays. Thousands of base pairs

of DNA can be screened on a single microarray for polymorphisms in specific genes whose sequence is already known. This greatly increases the scope for precise molecular diagnosis in single gene and genetically complex diseases.

In the medical practice, the precise diagnosis and effective treatment of disease depend on the ability of specialists to recognize the recurring constellations of clinical signs and symptoms that permit meaningful classification of the diseases. Unfortunately, the clinical signs and symptoms or even blood and pathological investigations are poor predictors of the clinical outcome or response to therapy. In the postgenomic era, studying of gene expression profiling will ultimately determine the biological behavior of both normal and injured tissues and cells, which may provide insights into disease mechanisms and help to identify novel candidates for therapeutic intervention. Gene profiling may also identify markers useful in diagnostic and prognostic purposes. Since thousands of genes are simultaneously and quantitatively analyzed on microarrays, this technology provides a resolution and precision not previously possible. Therefore, molecular phenotyping of diseases through gene expression profiling and sequencing may offer diagnostic, prognostic, and mechanistic insights that improve management of human diseases. Microarray technology has also a dramatic impact in the fundamental biological studies, providing an effective way to assess globally the transcriptional effects of specific genetic and pharmacological interventions, thus rapidly identifying possible up- and downstream effectors or alternative signaling pathways.

Bibliography

Books and Reviews

Beheshti, B., Park, P.C., Braude, I., Squire, J.A. (2002) Microarray CGH, *Methods Mol. Biol.* **204**, 191–207.

Bilban, M., Buehler, L.K., Head, S., Desoye, G., Ouaranta, V. (2002) Normalizing DNA microarray data, *Curr. Issues Mol. Biol.* **4**, 57–64.

Bittner, M. (2003) Fluorescent labeling of first-strand cDNA using reverse transcriptase, in: Botwell, D., Sambrook, J. (Eds.) *DNA Microarrays: a Molecular Cloning Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 178–186.

Cook, S.A., Rosenzweig, A. (2002) DNA microarrays: Implications for cardiovascular medicine, *Circ. Res.* **91**, 559–564.

Eyster, K.M., Lindahl, R. (2001) Molecular medicine: a primer for clinicians. Part XII: DNA microarrays and their application to clinical medicine, *S D J Med.* **54**, 57–61.

Geschwind, D.H. (2000) Mice, microarrays, and the genetic diversity of the brain, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10676–10678.

Geschwind, D.H. (2003) DNA microarrays: translation of the genome from laboratory to clinic, *Lancet Neurol.* **2**, 275–282.

Geschwind, D.H., Gregg, J. (Eds.) (2002) *Microarrays for the Neurosciences*, MIT Press, Cambridge, MA.

Goldsmith, Z.G., Dhanasekaran, N. (2004) The microevolution: applications and impacts of microarray technology on molecular biology and medicine, *Int. J. Mol. Med.* **13**, 483–495.

Greenberg, S.A. (2002) Bioinformatics in microarray expression analysis technology, in: Warrington, J.A., Todd, R., Wong, D. (Eds.) *Microarray and Cancer Research*, BioTechniques Press, Westborough, MA, pp. 25–42.

Hardiman, G. (2004) Microarray platforms – comparisons and contrasts, *Pharmacogenomics* **5**, 487–502.

Harkin, D.P. (2000) Uncovering functionally relevant signaling pathways using microarray-based expression profiling, *Oncologist* **5**, 501–507.

Harrington, C.A., Rosenow, C., Retief, J. (2000) Monitoring gene expression using DNA microarrays, *Curr. Opin. Microbiol.* **3**, 285–291.

Hartigan, J.A. (1975) *Clustering Algorithms*, John Wiley and Sons, New York.

Jain, K.K. (2004) Application of biochips: from diagnostics to personalized medicine, *Curr. Opin. Drug. Discov. Devel.* **7**, 287–289.

Joos, L., Eryuksel, E., Brutsche, M.H. (2003) Functional genomics and gene microarrays – the use in research and clinical medicine, *Swiss Med. Wklv.* **133**, 31–38.

Kohane, I.S., Kho, A., Butte, A.J. (2002) *Microarrays for an Integrative Genomics*, MIT Press, Cambridge, MA.

Leung, Y.F., Cavalieri, D. (2003) Fundamentals of cDNA microarray data analysis, *Trends Genet.* **19**, 649–659.

Liefers, G.J., Tollenaar, R.A. (2002) Cancer genetics and their application to individualized medicine, *Eur. J. Cancer* **38**, 872–879.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., Lockhart, D.J. (1999) High density synthetic oligonucleotide arrays, *Nat. Genet.* **21**(Suppl. 1), 20–24.

Macgregor, P.F. (2003) Gene expression in cancer: the application of microarrays, *Expert Rev. Mol. Diagn.* **3**, 185–200.

Marcotte, E.R., Srivastava, L.K., Ouirion, R. (2003) cDNA microarray and proteomic approaches in the study of brain diseases: focus on schizophrenia and Alzheimer's disease, *Pharmacol. Ther.* **100**, 63–74.

Mengel, M., Kreipe, H., von Wasielewski, R. (2003) Rapid and large-scale transition of new tumor biomarkers to clinical biopsy material by innovative tissue microarray systems, *Appl. Immunohistochem. Mol. Morphol.* **11**, 261–268.

Nadon, R., Shoemaker, J. (2002) Statistical issues with microarrays: processing and analysis, *Trends Genet.* **18**, 265–271.

Noordewier, M.O., Warren, P.V. (2001) Gene expression microarrays and the integration of biological knowledge, *Trends Biotechnol.* **19**, 412–415.

Panda, S., Sato, T.K., Hampton, G.M., Hogenesch, J.B. (2003) An array of insights: application of DNA chip technology in the stuffy of cell biology, *Trends Cell Biol.* **13**, 151–156.

Petrocoin, E.F., Hackett, J.L., Lesko, L.J., Puri, R.K., Gutman, S.I., Chumakov, K., Woodcock, J., Feigal, D.W., Zoon, K.C., Sistare, F. (2002) Medical applications of microarray technologies: a regulatory science perspective, *Nat. Genet.* **32**(Suppl. 2), 474–479.

Quackenbush, J. (2002) Microarray data normalization and transformation, *Nat. Genet.* **32**(Suppl. 2), 496–501.

Smith, L., Greenfield, A. (2003) DNA microarrays and development, *Hum. Mol. Genet.* **12**, R1–R8.

Tineke, C.M.T., van der Pouw Kraan, T.C., Kasperkovitz, P.V., Verbeet, N., Verweij, L. (2004) Genomics in the immune system, *Clin. Immunol.* **111**, 175–185.

Van Deerlin, V., Ginsberg, S.D., Lee, V.M., Trojanowski, J.G. (2002) The use of fixed human postmortem brain tissue to study mRNA expression in neurodegenerative diseases: application of microdissection and amplification, in: Geschwind, D.H., Gregg, J. (Eds.) *Microarrays for the Neuroscience: an Essential Guide*, MIT Press, Cambridge, MA, pp. 201–236.

Velculescu, V.E., Vogelstein, B., Kinzler, K.W. (2000) Analyzing uncharted transcriptomes with SAGE, *Trends Genet.* **16**, 423–425.

Warrington, J.A., Todd, R., Wong, D. (Eds.) (2002) *Microarrays in Cancer Research*, BioTechniques Press, Westborough, MA.

Weinstein, J.N., Scherf, U., Lee, J.K., Nishizuka, S., Gwadry, F., Bussey, A.K., Kim, S., Smith, L.H., Tanabe, L., Richman, S., Alexander, J., Kouros-Mehr, H., Maunakea, A., Reinhold, W.C. (2002) The bioinformatics of microarray gene expression profiling, *Cytometry* **47**, 46–49.

Primary Literature

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson, J., Jr., Lu, L., Lewis, D.B., Tibshirani, R., Sherlok, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson, W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O., Stuardt, L.M. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling, *Nature* **403**, 503–511.

Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D., Levine, A.J. (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissue probed by oligonucleotide arrays, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6745–6750.

Bayani, J., Brenton, J.D., Macgregor, P.F., Beheshti, B., Albert, M., Nallainathan, D., Karaskova, J., Rosen, B., Murphy, J., Laframboise, S., Zanke, B., Squire, J.A. (2002) Parallel analysis of sporadic primary ovarian carcinomas by spectral karyotyping, comparative genomic hybridization and expression microarrays, *Cancer Res.* **62**, 3466–3476.

Beer, D.G., Kardia, S.L., Huang, C.C., Giordano, T.J., Levin, A.M., Misek, D.E., Lin, L., Chen, G., Gharib, T.G., Thomas, D.G., Lizyness, M.L., Kuick, R., Hayasaka, S., Taylor, J.M., Iannattoni, M.D., Orringer, M.B., Hanash, S. (2002) Gene-expression profiles predict survival of patients with lung adenocarcinoma, *Nat. Med.* **8**, 816–824.

Bhattacharjee, A., Richards, W.G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E.J., Lander, E.S., Wong, W., Johnson, B.E., Golub, T.R., Sugarbaker, D.J., Meyerson, M. (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13790–13795.

Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., BenDor, A., Sampas, N., Dougherty, E., Wang, E., Marincola, F., Gooden, C., Lueders, J., Glatfelter, A., Pollock, P., Carpten, J., Gillanders, E., Leja, D., Dietrich, K., Beaudry, C., Berens, M., Alberts, D., Sondak, V. (2000) Molecular classification of cutaneous malignant melanoma by gene expression profiling, *Nature* **406**, 536–540.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roht, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridge, R.B., Kirchner, J., Fearon, K., Mao, J., Corcoran, K. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays, *Nat. Biotechnol.* **18**, 630–634.

Brutsche, M.N., Brutsche, I.C., Wood, P., Brass, A., Morrison, N., Rattay, M., Mogulkoc, N., Simler, N., Craven, M., Custovic, A., Egan, J.J., Woodcock, A. (2001) Apoptosis signals in atopy and asthma measured with cDNA arrays, *Clin. Exp. Immunol.* **124**, 181–187.

Chiang, L.W., Grenier, J.M., Ettwiller, L., Jenkins, L.P., Ficenec, D., Martin, J., Jin, F., DiStefano, P.S., Wood, A. (2001) An orchestrated gene expression component of neuronal programmed cell death revealed by cDNA array analysis, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2814–2819.

Dahlquist, K.D., Salomonis, N., Vranizan, K., Lawlor, S.C., Conklin, B.R. (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways, *Nat. Genet.* **31**, 19–20.

Datson, N.A., Van der Perk, J., DeKloet, E.R., Vreugdenhil, E. (2001) Expression profile of 30,000 genes in rat hippocampus using SAGE, *Hippocampus* **11**, 430–444.

De Risi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A., Trent, J.M. (1996) Use of a cDNA microarray to analyze gene expression patterns in human cancer, *Nat. Genet.* **14**, 457–460.

De Vos, J., Thykjaer, T., Tarte, K., Ensslen, M., Raynaud, P., Requirand, G., Pellet, F., Pantesco, V., Reme, T., Jourdan, M., Rossi, J.F., Orntoft, T., Klein, B. (2002) Comparison of gene expression profiling between malignant and normal plasma cells with oligonucleotide arrays, *Oncogene* **21**, 6848–6857.

Dhanasekaran, S.M., Barrette, T.R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K.J., Rubin, M.A., Chinnaiyan, A.M. (2001) Delineation of prognostic biomarkers in prostate cancer, *Nature* **412**, 822–826.

Dixon, A.K., Richardson, P.J., Lee, K., Carter, N.P., Freeman, T.C. (1998) Expression profiling of single cells using 3′ and amplification (TPEA) PCR, *Nucleic Acids Res.* **26**, 4426–4431.

Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M., Coleman, P. (1992) Analysis of gene expression in single live neurons, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 3010–3014.

Euer, N., Schwirzke, M., Evtimova, V., Burtscher, H., Jarsch, M., Tarin, D., Weidle, U.H. (2002) Identification of genes associated with metastasis of mammary carcinoma in metastatic versus non-metastatic cell lines, *Anticancer Res.* **22**, 733–740.

Furlong, E.E., Andersen, E.C., Null, B., White, K.P., Scott, M.P. (2001) Patterns of gene expression during Drosophila mesoderm development, *Science* **293**, 1629–1633.

Garber, M.E., Troyanskaya, O.G., Schluens, K., Petersen, S., Thaesler, Z., Pacyna-Gengelbach, M., van de Rijn, M., Rosen, G.D., Peron, C.M., Whyte, R.I., Altman, R.B., Brown, P.O., Botstein, D., Petersen, I. (2001) Diversity of gene expression in adenocarcinoma of the lung, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13784–13789.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesiroy, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science* **286**, 531–537.

Haab, B.B., Dunham, M.J., Brown, P.O. (2001) Protein microarrays for parallel detection and quantitation of specific proteins and antibodies in complex solutions, *Genome Biol.* **2**, Research2004.

Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O.P., Wilfond, B., Borg, A., Trent, J. (2001) Gene-expression profiles in hereditary breast cancer, *N. Engl. J. Med.* **344**, 539–548.

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., Kobayashi, S., Davis, C., Dai, H., He, Y.D., Stephaniants, S.B., Cavet, G., Walker, W.L., West, A., Coffey, E., Shoemaker, D.D., Stoughton, R., Blanchard, A.P., Friend, S.H., Linsley, P.S. (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer, *Nat. Biotechnol.* **19**, 342–347.

Iscove, N.N., Barbara, M., Gu, M., Gibson, M., Modi, C., Winegarden, N. (2002) Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA, *Nat. Biotechnol.* **20**, 940–943.

Jazaeri, A.A., Yee, C.J., Sotiriou, C., Brantley, K.R., Boyd, J., Liu, E.T. (2002) Gene expression profiles of BRCA1-linked, BRCA2-linked and sporadic ovarian cancers, *J. Natl. Cancer Inst.* **94**, 990–1000.

Kacharmina, J.E., Crino, P.B., Eberwine, J. (1999) Preparation of cDNA from single cells and subcellular regions, *Methods Enzymol.* **303**, 3–18.

Karsten, S.L., Geschwind, D.H. (2002) Gene expression analysis using cDNA microarrays, *Curr. Protoc. Neurosci.* Unit 4.25.

Karsten, S.L., Van Deerlin, V.M., Sabatti, C., Gill, L.H., Geschwind, D.H. (2002) An evaluation of tyramide signal amplification and archived fixed and frozen tissue in microarray gene expression analysis, *Nucleic Acids Res.* **30**, E4.

Livesey, F.J., Furukawa, T., Steffen, M.A., Church, G.M., Cepko, C.M. (2000) Microarray analysis of the transcriptional network controlled by the photoreceptor homeobox gene Crx, *Curr. Biol.* **10**, 301–310.

Lukasiuk, K., Pitkannen, A. (2004) Large-scale analysis of gene expression in epilepsy research: is synthesis already possible? *Neurochem. Res.* **29**, 1169–1178.

Luo, H., Salunga, R.C., Guo, H., Bittner, A., Joy, K.C., Galindo, J.E., Xiao, H., Rogers, K.E., Jackson, M.R., Erlander, M.G. (1999) Gene expression profiles of laser-captured adjacent neuronal subtypes, *Nat. Med.* **5**, 117–122.

Notterman, D.A., Alon, U., Sierk, A.J., Levine, A.J. (2001) Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma and normal tissue examined by oligonucleotide arrays, *Cancer Res.* **61**, 3124–3130.

Novoradoyskaya, N., Whitfield, M.L., Basehore, L.S., Novoradovsky, A., Pesich, R., Usary, J., Karaca, M., Wong, W.K., Aprelikova, O., Fero, M., Perou, C.M., Botstein, D., Braman, J. (2004) Universal reference RNA as a standard for microarray experiments, *BMC Genomics* **5**, 20.

Perou, C.M., Jeffrey, S.S., Van de Rijn, M., Rees, C.A., Eisen, M.B., Ross, D.T., Pergamenschikov, A., Williams, C.F., Zhu, S.X., Lee, J.C., Lashkari, D., Shalon, D., Brown, P.O., Botstein, D. (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9212–9217.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S.X., Lonning, P.E., Borresen-Dale, A.L., Brown, P.O., Botstein, D. (2000) Molecular portraits of human breast tumors, *Nature* **406**, 747–752.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., Daikee, S.H., Ljung, B.M.,

Gray, J.W., Albertson, D.G. (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays, *Nat. Genet.* **20**, 207–211.

Puskas, L.G., Zvara, A., Hackler, L., Jr., Van Hummelen, P. (2002) RNA amplification results in reproducible microarray data with slight ratio bias, *Biotechniques* **32**, 1330–1334.

Randolph, J.B., Waggoner, A.S. (1997) Stability, specificity and fluorescence brightness of multiply-labeled fluorescent DNA probes, *Nucleic Acids Res.* **25**, 2923–2929.

Rose, S.D. (1998) Application of a novel microarraying system in genomics research and drug discovery, *J. Assoc. Lab. Automat.* **3**, 53–57.

Sandberg, R., Yasuda, R., Pankratz, D.G., Carter, T.A., Del Rio, J.A., Wodicka, L., Mayford, M., Lockhart, D.J., Barlow, C. (2000) Regional and strain-specific gene expression mapping in the adult mouse brain, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11038–11043.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science* **270**, 467–470.

Selaru, F.M., Zou, T., Xu, Y., Shustova, V., Yin, J., Mori, Y., Sato, F., Wang, S., Olaru, A., Shibata, D., Greenwald, B.D., Krasna, M.J., Abraham, J.M., Meltzer, S.J. (2002) Global gene expression profiling in Barrett's esophagus and esophageal cancer: a comparative analysis using cDNA microarrays, *Oncogene* **21**, 475–478.

Snijders, A.M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A.K., Huey, B., Kimura, K., Law, S., Myambo, K., Palmer, J., Ylstra, B., Yue, J.P., Gray, J.W., Jain, A.N., Pinkel, D., Albertson, D.G. (2001) Assembly of microarrays for genome-wide measurement of DNA copy number, *Nat. Genet.* **29**, 263–264.

Sugita, M., Geraci, M., Gao, B., Powell, R.L., Hirsch, F.R., Johnson, G., Lapadat, R., Gabrielson, E., Bremnes, R., Bunn, P.A., Franklin, W.A. (2002) Combined use of oligonucleotide and tissue microarrays identifies cancer/testis antigens as biomarkers in lung carcinoma, *Cancer Res.* **62**, 3971–3979.

Sultan, M., Wigle, D.A., Cumbaa, C.A., Maziarz, M., Glasgow, J., Tsao, M.S., Jurisica, I. (2002) Binary tree-structured vector quantization approach to clustering and visualizing microarray data, *Bioinformatics* **18**(Suppl.), S111–S119.

Tabuchi, M., Ueda, M., Kaji, N., Yamasaki, Y., Nagasaki, Y., Yoshikawa, K., Kataoka, K., Baba, Y. (2004) Nanospheres for DNA separation chips, *Nat. Biotechnol.* **22**, 337–340.

Takahashi, Y., Ishii, Y., Nagata, T., Ikarashi, M., Ishikawa, K., Asai, S. (2003) Clinical application of oligonucleotide probe array for full-length gene sequencing of TP53 in colon cancer, *Oncology* **64**, 54–60.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, O., Kitareewan, S., Dmitrovsky, E., Lander, E.S., Golub, T.R. (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2907–2912.

Tanabe, L., Scherf, U., Smith, L.H., Lee, J.K., Hunter, L., Weinstein, J.N. (1999) MedMiner: an Internet text-mining tool for biomedical information, with application to gene expression profiling, *Biotechniques* **27**, 1210–1214.

Tanabe, L., Scherf, U., Smith, L.H., Lee, J.K., Hunter, L., Weinstein, J.N. (1999) MedMiner: an Internet text-mining tool for biomedical information, with application to gene expression profiling, *Biotechniques* **27**, 1216–1217.

Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., Church, G.M. (1999) Systematic determination of genetic network architecture, *Nat. Genet.* **22**, 281–285.

The International Human Genome Sequencing Consortium (2001) The sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Unami, A., Shinohara, Y., Kajimoto, K., Baba, Y. (2004) Comparison of gene expression profiles between white and brown adipose tissues of rats by microarray analysis, *Biochem. Pharmacol.* **67**, 555–564.

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W. (1995) Serial analysis of gene expression, *Science* **270**, 484–487.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G. et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Vernon, S.D., Unger, E.R., Rajeevan, M., Dimulescu, I.M., Nisenbaum, R., Campbell, C.E.

(2000) Reproducibility of alternative probe synthesis approaches for gene expression profiling with arrays, *J. Mol. Diagn.* **2**, 124–127.

Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R. et al. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome, *Science* **280**, 1077–1082.

Wang, E., Miller, L.D., Ohnmachr, G.A., Liu, E.T., Marincola, F.M. (2000) High-fidelity mRNA amplification for gene profiling, *Nat. Biotechnol.* **18**, 457–459.

Whitney, L.W., Becker, K.G., Tresser, N.J., Caballero-Ramos, C.I., Munson, P.J., Prabhu, V.V., Trent, J.M., McFarland, H.F., Biddison, W.E. (1999) Analysis of gene expression in multiple sclerosis lesions using cDNA microarrays, *Ann. Neurol.* **46**, 425–428.

Wigle, D.A., Jurisica, I., Radulovich, N., Pintilie, M., Rossant, J., Liu, N., Lu, C., Woodgett, J., Seiden, I., Johnston, M., Keshavjee, S., Darling, G., Winton, T., Breitkreutz, B.J., Jorgenson, P., Tyers, M., Shepherd, F.A., Tsao, M.S. (2002) Molecular profiling of non-small cell lung cancer and correlation with disease-free survival, *Cancer Res.* **62**, 3005–3008.

Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., Speed, T.P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation, *Nucleic Acids Res.* **30**, e15.

Yue, H., Eastman, P.S., Wang, B.B., Minor, J., Doctolero, M.H., Nuttall, R.L., Stack, R., Becker, J.W., Montgomery, J.R., Vainer, M., Johnston, R. (2001) An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression, *Nucleic Acids Res.* **29**, E41–E41.

Yuen, T., Wurmbach, E., Pfeffer, R.L., Ebersole, B.J., Sealfon, S.C. (2002) Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays, *Nucleic Acids Res.* **30**, E48.

Microarray Technology: *see* Microarray-Based Technology: Basic Principles, Advantages and Limitations

Microbial Development

Paul Robert Fisher
La Trobe University, VIC, Australia

1 Introduction 292

2 Bacteria 292
2.1 Cell Cycle Control and Morphogenesis in *Caulobacter crescentus* 292
2.1.1 CtrA – A Master Regulator of Cell Cycle Progression and Polar Morphogenesis 293
2.1.2 Morphogenesis in the Swarmer Cell – Flagellar Biogenesis 296
2.1.3 Phosphorelays Regulating the Cell Cycle and Morphogenesis 296
2.1.4 Dynamic Subcellular Localization of Phosphorelay Proteins Controlling Cell Division and Morphogenesis 298
2.2 Sporulation and Differentiation in *Bacillus subtilis* 299
2.2.1 Spo0A and the Phosphorelay in Endospore Formation 301
2.2.2 Mother–Daughter Communication in Endospore Formation – A Spatiotemporally Regulated Sigma Factor Cascade 304
2.3 Differentiation and Morphogenesis in *Myxococcus xanthus* 307
2.3.1 Initiation of Development by Starvation and Extracellular Signals 308
2.3.2 C-signal Transduction and Sporulation After Aggregation 310

3 Eukarya 312
3.1 Cell Cycle Control and Differentiation in *Saccharomyces cerevisiae* 312
3.1.1 Pseudohyphal Growth 314
3.1.2 Meiosis and Sporulation 314
3.1.3 One Locus, Three Cell Types 316
3.1.4 Gametogenesis and Mating 317
3.1.5 Silent Copies of the Mating-type Alleles and Mating-type Switching 319
3.2 Differentiation and Morphogenesis in *Dictyostelium discoideum* 321
3.2.1 The Transition from Growth to Development 323
3.2.2 Chemotactic Aggregation 326
3.2.3 Postaggregative Development and Cell Type Choice 331

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

3.2.4 Culmination and Terminal Differentiation 334
3.2.5 Spore Maturation and Germination 337

4 Conclusion 338

Acknowledgments 339

Bibliography 339
Books and Reviews 339
Primary Literature 340

Keywords

Differentiation
A stable change in global patterns of gene expression usually producing a recognizably different type of cell.

Morphogenesis
The creation or development of an organism's or cell's morphology, structure, and form.

Development
A genetically programmed combination of differentiation and morphogenesis.

Cell Cycle
The orderly progression of events during cell proliferation through phases of growth, DNA replication, chromosome segregation, and cell division (cytokinesis).

Protein Kinase
An enzyme that phosphorylates itself and/or other proteins. Major categories of protein kinase are named according to the amino acid that is specifically phosphorylated, for example, tyrosine protein kinases phosphorylate target proteins on tyrosine residues.

Sigma Factor
A subunit of bacterial RNA polymerase that directs binding of the polymerase to specific DNA sequences (promoters), resulting in the transcription of downstream genes. Sigma factors form a family of polypeptides, each of which recognizes a different promoter sequence and so directs transcription of a different set of genes.

Transcription Factor
A DNA-binding protein that is not part of the RNA polymerase, but which binds to specific DNA sequences and regulates the expression of nearby genes by activating or inhibiting transcription.

Signal Transduction
The process whereby extracellular or intracellular, physical or chemical signals elicit a series of intracellular biochemical events in which each event either activates or inhibits the next.

Holdfast
A specialized attachment organ (of a multicellular organism) or organelle (of a unicellular organism).

Flagellum
A long, thin cellular appendage that is responsible for swimming motility – in the case of bacteria a rigid, helical, hollow filament that functions as a propellor attached to a rotary motor in the cell membrane through a semiflexible coupling.

Pilus
A long, fiberlike appendage on a bacterial cell that is thinner than a flagellum.

Pseudopod or Pseudopodium
A region of an amoeboid cell that does not contain membrane-bounded organelles and is extended outward during amoeboid movement.

Spore
A specialized, dormant, resistant cell formed as a means of dispersal from and survival of harsh conditions.

Endospore
A spore formed from a daughter cell within the membrane of a mother cell in certain bacterial species (notably of the genera *Bacillus* and *Clostridium*).

Sorus
A cluster or mass of spores.

Development is not only restricted to large multicellular animals and plants but is also a feature of the life cycles of microorganisms. The environment of a microorganism is rapidly changeable so that both eukaryotic and prokaryotic microbes have evolved developmental strategies to enhance their nutrient-scavenging abilities or in more extreme circumstances to form dormant, resistant cells (spores). The spores can survive under harsh conditions that do not support growth and can be dispersed to other environments that do. The endospores of pathogenic *Clostridium* species are amongst the most resistant cells on earth, and their destruction (or removal) is the key criterion for successful microbiological sterilization. Developmental programs in microbial pathogens are initiated in response to infection of the host and play important roles in pathogenesis. However, the best-understood examples of microbial development are not pathogens, but free-living microbes. The general principles underlying microbial development are being elucidated by their study.

1
Introduction

Biologists do not usually think of microbial model organisms when asking questions about the fundamental mechanisms of development. Yet, even in the most complex of metazoans, development involves cells doing things in an ordered way through space and time – growing, dividing, adhering to one another and to the substratum, differentiating, moving, secreting, and, above all, communicating – and the fundamental mechanisms of what cells do evolved very early in the history of life. So it should not be surprising to learn that the molecular mechanisms at the heart of these cellular processes in development are frequently similar.

At the same time, every organism in the tree of life has faced its own unique problems in using these primeval molecular mechanisms to meet the needs of its "chosen" life style. Each has accordingly evolved its own developmental programs using its own unique selection of molecular techniques to carry them out. So we are not surprised either that organisms belonging to different phylogenetic lineages have often opted for different molecular mechanisms to perform similar developmental functions or that in some lineages, particular molecular mechanisms have been elaborated upon, used, and reused in many different biological roles, while others are not used at all.

An example of these principles is to be found in the universal use of protein kinase cascades in signal transduction pathways controlling development. Yet there are differences. The histidine protein kinases are used in developmental signaling pathways in all of the major bacterial groups and in many eukaryotic lineages including the fungi, the slime moulds, and the plants, but not, it seems, in metazoan animals. No one knows why the use of histidine kinases was abandoned in the animal lineage, but retained in the next most closely related eukaryotic lineages. The bacteria, on the other hand, appear to make little, if any, use of the serine/threonine protein kinases so ubiquitously found in eukaryotic organisms. Differences like this tell us that there is nothing about one class of kinase compared to the other that makes its use essential for particular signaling purposes. They also identify classes of molecules that could be potentially used as targets for drug or antibiotic therapy. Histidine kinase signaling pathways participate in bacterial pathogenesis and are therefore potential antibiotic targets.

What then might be the general principles to be gleaned from our current understanding of development in eukaryotic and bacterial microbes? To attempt to answer this question I have opted in the following sections to discuss microbial development by describing what we know about developmental pathways in five well-studied systems – three bacterial and two eukaryotic. The emphasis is on the signaling cascades that control cellular activities in development. Having examined each of these five systems in some detail, I conclude by noting general features that seem to be shared by all.

2
Bacteria

2.1
Cell Cycle Control and Morphogenesis in *Caulobacter crescentus*

Bacteria are usually thought of as exhibiting one of only a limited number of cellular morphologies – rods, cocci, or

spirals – that is monotonously present in every cell. *Caulobacter crescentus* is different. An aquatic bacterium, common in freshwater environments characterized by low and patchy nutrient levels, *Caulobacter's* unique cell cycle involves polarized differentiation and cytofission to produce two morphologically distinct cells – a non-dividing, motile, swarmer cell, and a sessile, stalked cell that is able to progress through cell division (Fig. 1). The swarmer cell is able to disperse by both random and chemotactic swimming motility to nutrient-rich microhabitats where it differentiates, losing its polar flagella and pili to replace them with a stalk and holdfast. Through these, it attaches to the substratum and begins progression through the cell cycle just as the mother cell did from which it was derived. It is this cell cycle, with its two dramatically different cell types, that has made *Caulobacter* a favorite organism for studying the mechanisms of cell cycle control, polarized differentiation and morphogenesis in bacteria.

2.1.1 CtrA – A Master Regulator of Cell Cycle Progression and Polar Morphogenesis

At the heart of the molecular machinery controlling *Caulobacter* morphogenesis is a protein called CtrA, a DNA-binding protein that functions as a transcriptional activator and repressor to regulate the expression of genes involved in DNA synthesis and cytofission, flagellar motility, and chemotaxis. CtrA is the final response regulator in a multicomponent phosphorelay system of the histidine protein kinase type that is universally found in bacteria and archaea, as well as in some eukaryotes (e.g. *Dictyostelium discoideum*). The phosphorylation cascade in such systems is initiated by an autophosphorylating histidine protein kinase, which is activated (or inhibited) by reception of a signal in the form of a ligand binding to its sensor domain. The phosphate is initially attached to a conserved histidine residue in the transmitter domain of the kinase, from where it is subsequently transferred (sometimes via multiple phosphotransfers) to an aspartate residue in the receiver domain of the same or a different protein – the response regulator that mediates the final response. In multicomponent systems, multiple phosphotransfers proceed from the primary histidine to the final aspartate via alternating histidine and aspartate residues in transmitter and receiver domains respectively of what can be distinct polypeptides.

The sequencing of the entire genome of *C. crescentus* made possible a comparative microarray analysis of gene expression in a wild type and temperature-sensitive *CtrA* loss-of-function mutant. Nearly 20% of the ca. 3700 genes in *Caulobacter* are cell cycle regulated and the microarray studies revealed that about one-third of these are controlled by CtrA, either directly or indirectly. Many of the direct targets were found by coimmunoprecipitation of DNA fragments that had been cross-linked to CtrA. This approach identified 55 promoters (regulating 95 genes) that bind and are presumably controlled by the active, phosphorylated form of CtrA (CtrA∼P). The binding site is a conserved nonameric DNA sequence motif that is found not only in CtrA-regulated promoters but also at five sites within the chromosomal origin of replication, *ori*. Binding of CtrA∼P to *ori* prevents the initiation of DNA synthesis and entry to S-phase. CtrA∼P also represses expression of the early cell division gene *ftsZ*, the product of which forms a ring that circumscribes the cell at the division site and assembles the cellular division proteins there at the inner face of the cytoplasmic membrane.

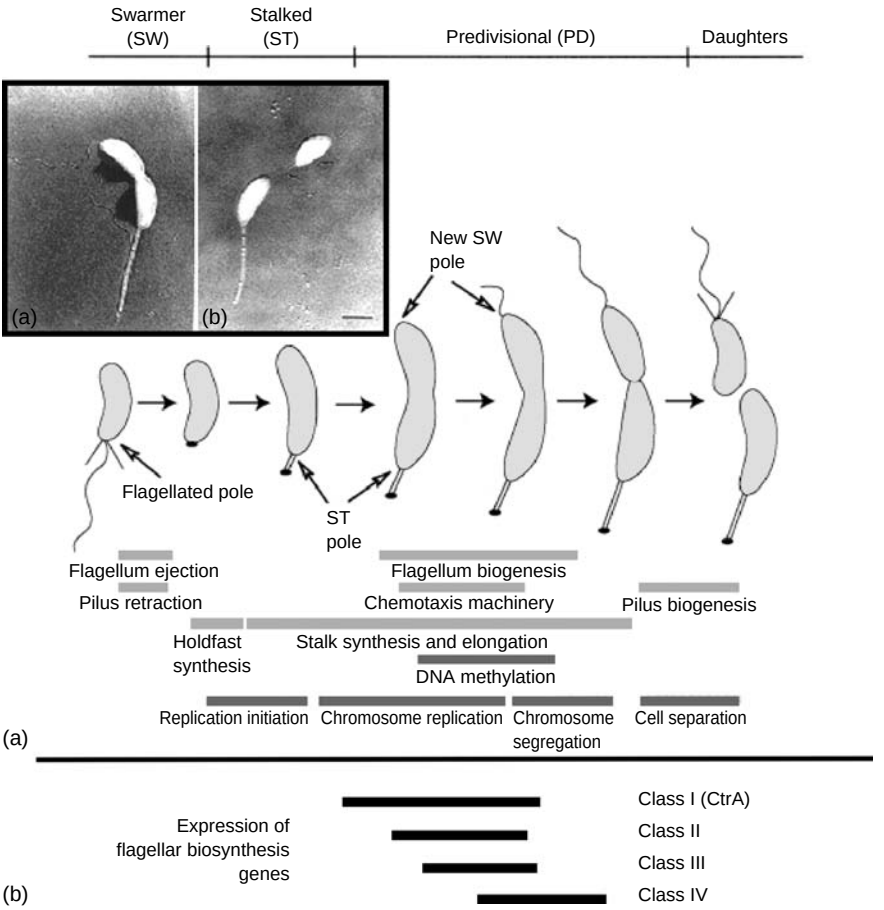

Fig. 1 The cell cycle and flagellar morphogenesis in *Caulobacter*. (a) The motile swarmer cell on the left loses its flagellum, retracts its polar pili and differentiates into a stalked cell, which enters the S-phase and progresses through the cell division cycle. The differentiation of two distinct daughter cells begins in the predivisional cell (inset, panel a) with distinct morphogenetic events occurring at the stalked (ST) and swarmer (SW) poles. After cytofission (inset, panel b), the stalked daughter cell is able to resume cell cycle progression, whereas the swarmer cell is arrested in growth phase (G1) until such time as it has differentiated into a stalked cell. The bars show the timing of various morphogenetic events during the cell cycle. (b) In the predivisional cell, a cascade of sequentially dependent inductions of flagellar gene expression and assembly events result in biogenesis of the flagellar apparatus at the swarmer pole. Horizontal bars show the timing of expression of the four sequentially induced classes of flagellar biosynthesis genes. Panel (a) is modified from Fig. 1 of Jacobs-Wagner, C. (2004) Regulatory proteins with a sense of direction: cell cycle signalling network in *Caulobacter, Mol. Microbiol.* **51**, 7–13. Panel (b) contains information presented in Fig. 1 of England & Gober (2001) Cell cycle control of morphogenesis in *Caulobacter Curr. Opin. Microbiol.* **4**, 674–680. Inset image courtesy of Jeanne S. Poindexter, Barnard College, Columbia University.

In keeping with its regulatory roles, the levels of CtrA in *Caulobacter* oscillate in synchrony with the cell cycle and differ in stalked and swarmer cells (Fig. 2). The levels of the phosphorylated form of CtrA change in concert with the levels of the protein itself. After accumulating in the late predivisional cell, CtrA∼P remains at high levels in swarmer cells where it represses cell cycle progression at the point of entry into S-phase and also induces the gene expression cascades involved in swarmer cell differentiation (e.g. those involved in pilus and flagellar assembly). In the stalked cell at the onset of replication, CtrA∼P levels are low, the protein having been both dephosphorylated and also degraded by ClpXP proteases. These two mechanisms of inactivating CtrA are redundant – mutant *ctrA* alleles that either are resistant to proteolysis or to inactivation by dephosphorylation are not lethal. The mutants progress through the cell cycle normally because in each case the alternative CtrA inactivation mechanism (dephosphorylation or degradation) is still present and is active at the appropriate cell cycle stage. As expected, a CtrA mutant that is both constitutively active and protease-resistant causes cell cycle arrest at the G1–S transition.

The disappearance and inactivation of CtrA at the end of G1 relieves its repression of replication, allowing the cell to enter S-phase and initiate DNA synthesis. Also relieved is the autoinhibitory repression of *ctrA* expression by high levels of CtrA∼P bound to the low affinity *ctrA* promoter, P1. The *ctrA* expression that results is autocatalytic because of CtrA∼P-mediated induction through a second, high affinity promoter, P2. Consequently, CtrA∼P levels rise again to high levels in the late predivisional cell, where they prevent

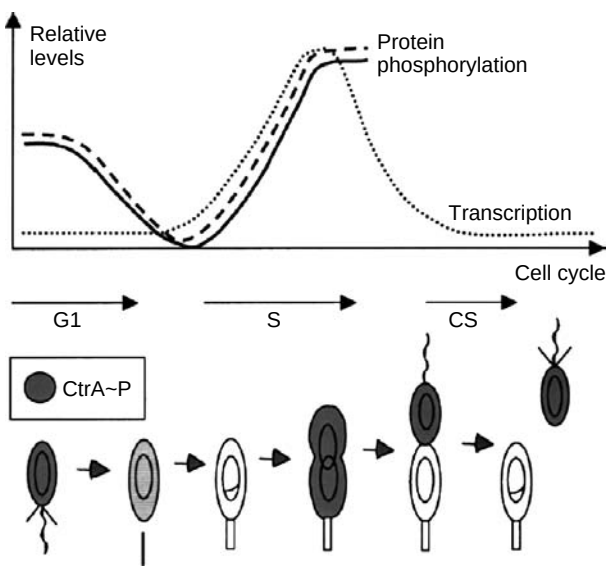

Fig. 2 Schematic representation of the regulatory mechanisms that modulate the level of CtrA∼P (in grey) during the cell cycle. From Fig. 2 of Ausmees, N., Jacobs-Wagner, C. (2003) Spatial and temporal control of differentiation and cell cycle progression in *Caulobacter crescentus*, *Annu. Rev. Microbiol.* **57**, 225–247.

initiation of another round of replication and again repress *ctrA* expression through the P1 promoter site.

2.1.2 Morphogenesis in the Swarmer Cell – Flagellar Biogenesis

Having accumulated in the predivisional cell, CtrA∼P persists in the swarmer cell, but is absent from the stalked cell. In the swarmer compartment of the predivisional cell, it initiates a gene expression cascade whose outcome is the differentiated swarmer cell and whose most obvious and best-understood feature is flagellar biogenesis. The ca. 50 gene products required for flagellar assembly in *Caulobacter* belong to four classes: Class I, the master regulator (CtrA); Class II, including a σ^{54} transcription factor (FlbD) and the first flagellar components to be assembled (the MS ring – FliF; the switch – FliG,M,N; the flagellum-specific secretion system – FlhH, FliI,J,O,P,Q,R); Class III, polypeptides forming outer parts of the basal body and hook; Class IV, the flagellar filament (Fig. 1).

Once initiated by CtrA∼P, the assembly of the flagellar apparatus proceeds in three major stages defined by the onset of supply of the requisite polypeptides encoded by the Class II, III, and IV genes. Within each stage, the assembly of the parts progresses in a sequentially dependent manner – the incorporation of each polypeptide depending upon the assembly of those before it. The start of each major assembly stage is defined by a checkpoint at which the synthesis of polypeptides for that stage can begin only if the preceding stage has been completed. Thus, the Class II genes are induced in the predivisional cell by CtrA∼P, the accumulation of which constitutes the Class I/II checkpoint.

The product of one of the Class II genes, *FlbD*, in its active phosphorylated form, in turn induces the Class III and IV genes in the swarmer pole of the predivisional cell. However, FlbD activity also depends upon two other Class II proteins – FlbE and FliX. FlbE was thought to be the kinase that phosphorylates FlbD, but recent evidence suggests that it activates FlbD by some other means. FliX functions as the checkpoint sensor that detects proper assembly of the inner parts of the flagellar basal body – FliX interacts directly with FlbD at its N terminus and only permits its activation when this Class II/III checkpoint has been reached. Active FlbD∼P not only induces the Class III and IV genes but it also feeds back to repress early Class II genes.

Although both Class III and Class IV genes are induced by FlbD∼P, the otherwise very stable Class IV mRNAs are targeted for degradation by the binding to them of the FlbT protein. This protein is present throughout the cell cycle, but its activity is inhibited by completion of the assembly of the flagellar motor, complete with hook. This constitutes the Class III/IV checkpoint. Only when this checkpoint has been reached are the Class IV mRNAs stabilized so that flagellin synthesis and completion of the entire flagellar apparatus can proceed.

2.1.3 Phosphorelays Regulating the Cell Cycle and Morphogenesis

CtrA is active only when phosphorylated at the conserved aspartate residue in its receiver domain (Asp51), and its activation by this means is controlled by at least three converging phosphorelay cascades (Fig. 3). The first of these involves CckA, a histidine protein kinase that is responsible for CtrA activation during the cell cycle. Consistent with its upstream role as a CtrA

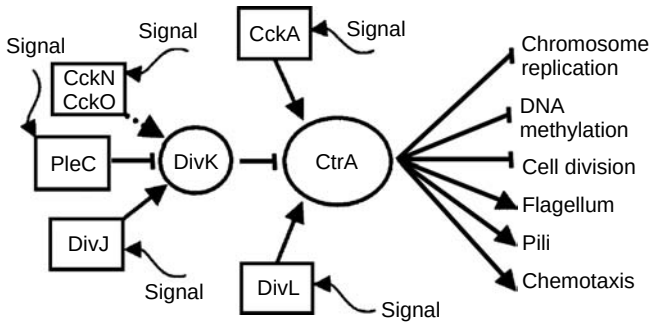

Fig. 3 Model of the signal transduction network controlling the activity of the master response regulator, CtrA. Response regulators and histidine kinases are shown as circles and boxes respectively. Barred ends indicate inhibition and arrowheads indicate activation. The dotted arrow indicates that the functions of the indicated upstream kinases are unknown. The molecular nature of the indicated signals that initiate the phosphorelays shown are unknown. Modified from Fig. 3 of Ausmees, N., Jacobs-Wagner, C. (2003) Spatial and temporal control of differentiation and cell cycle progression in *Caulobacter crescentus*, *Annu. Rev. Microbiol.* **57**, 225–247.

activator, CckA is essential for viability and temperature-sensitive mutants of this protein exhibit similar or identical phenotypes to temperature-sensitive CtrA mutants. However, CckA is rendered dispensable by expression of a constitutively active CtrA derivative (CtrAD51E) that does not require phosphorylation for activity. Because CckA contains its own response regulator receiver domain at its C terminus, it is anticipated that CckA passes phosphates to CtrA via an as yet unidentified intermediate histidine phosphotransferase in a multicomponent phosphorelay.

The second phosphorelay pathway regulating CtrA activity is mediated by DivL, a sensory kinase whose interactions with CtrA have been demonstrated genetically. Although it possesses all the conserved regions of the histidine kinase superfamily, DivL is unusual in possessing an autophosphorylatable tyrosine residue (Tyr550) instead of histidine. A cold sensitive *divL* allele is genetically suppressed (i.e. the wild-type phenotype is restored) by a specific *ctrA* allele (called *sokA*). This phenomenon of suppression is a classical genetic indicator of interactions between the corresponding encoded proteins. It is supported in this case by *in vitro* biochemical evidence that DivL can phosphorylate CtrA.

DivK is a response regulator protein that directly or indirectly controls the levels of active, phosphorylated CtrA, probably by accelerating both its dephosphorylation and its degradation by ClpX proteases. The cold-sensitive *divK341* allele causes cell cycle arrest at the G1–S transition – initiation of replication is blocked, genes involved in replication fail to be induced, and the ClpX-mediated degradation of CtrA does not occur. The upstream histidine kinases that regulate DivK include the nonessential proteins PleC and DivJ. *In vitro*, they can catalyze both the phosphorylation and the dephosphorylation of the response regulator DivK. However, their

in vivo actions appear to be opposed – DivJ phosphorylates and PleC dephosphorylates DivK so that its phosphorylation levels are reduced in a DivJ mutant, but increased in a PleC mutant. Two additional cytoplasmic histidine kinases of unknown function, CckN and CckO, have been identified in yeast two-hybrid screens as interacting partners of DivK.

The response regulator DivK and its upstream histidine kinases constitute the third phosphorelay cascade, leading to CtrA and regulating its activity. Just as the *sokA* allele of *ctrA* suppresses a conditional *divL* allele, it also suppresses mutant alleles of *divK, divJ*, and *pleC*. This provides a genetic demonstration of interactions between CtrA and the proteins encoded by these three genes. DivJ and PleC also interact with another response regulator protein, PleD, whose phosphorylated form activates proteolytic degradation of FliF, leading to loss of the flagella, transition to the stalked form, and entry to S-phase. The phosphorylated form of PleD is an active di-guanylate cyclase that synthesizes cyclic di-guanosine monophosphate (c-di-GMP) from 2 GTP molecules. In some other bacteria, this small molecule acts as an allosteric regulator activating cellulose synthase, but its downstream targets in *Caulobacter* are not known.

2.1.4 Dynamic Subcellular Localization of Phosphorelay Proteins Controlling Cell Division and Morphogenesis

Fusing the gene encoding jellyfish green fluorescent protein (GFP) with that for specific *Caulobacter* signaling proteins made it possible to use fluorescence microscopy to study their dynamic localization in living cells during the cell cycle (Fig. 4). Each of

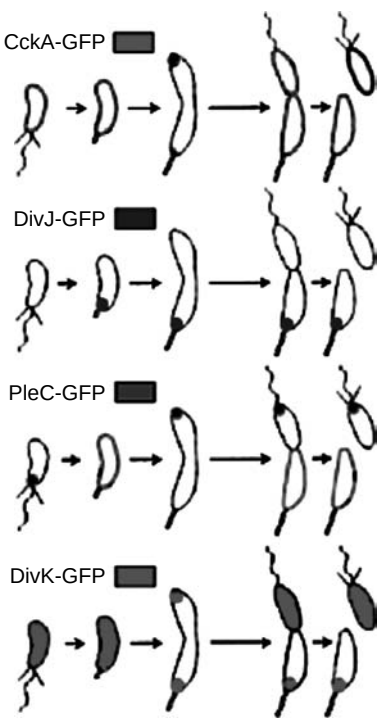

Fig. 4 Localization of some morphogenetic signaling proteins during the *Caulobacter* cell cycle. CckA and PleC are membrane-bound and cycle as shown between being dispersed throughout the membrane and concentrated at the indicated poles. DivJ, when present, is always concentrated at the stalked (ST) pole. DivK cycles between a diffuse cytoplasmic localization in swarmer cells and concentration at the indicated poles in predivisional and stalked cells. The length of the arrows indicates the time spent between each stage of differentiation. From Fig. 2 of Jacobs-Wagner, C. (2004) Regulatory proteins with a sense of direction: cell cycle signalling network in *Caulobacter, Mol. Microbiol.* **51**, 7–13. (See color plate. p. xxix)

these signaling proteins exhibits a characteristic pattern of localization that reflects its functions in morphogenesis and cell division. From being dispersed throughout the inner, cytoplasmic membrane, CckA translocates in the predivisional cell to the poles, particularly the swarmer pole where it can phosphorylate CtrA and initiate flagellar assembly. It returns to its original diffuse membrane distribution by the time cell division is complete. DivK translocates from the cytoplasm in the swarmer cell to both poles in the predivisional cell where it is regulated by either DivJ (at the stalked pole) or PleC (at the swarmer pole). By the time division is complete, DivK has been released from its polar location in the swarmer cell to return to its original distribution. However, it remains localized at the stalked pole in the stalked cell. At the stalked pole DivJ would activate DivK and so promote dephosphorylation and degradation of CtrA, favoring progression through the cell cycle. During early differentiation of the stalked cell, PleC is released from the newly forming stalked pole and moves to the opposite pole where it would keep DivK inactive and CtrA active in inducing flagellar biogenesis.

What controls these very striking localization patterns? For the most part, the answer to this question is not yet known, but some elements of the mechanisms involved have emerged. Firstly, correct polar localization of DivK depends upon phosphorylation of its conserved receiver aspartate and recruitment by DivJ at the stalked pole. Secondly, DivK release from the swarmer pole requires PleC, presumably because of the PleC-mediated dephosphorylation of DivK. Thirdly, the release of PleC itself from the future stalked pole requires PleC's own catalytic activity, while its localization to the swarmer pole requires a protein with similar localization patterns called *PodJ*. Fourthly, a 61-residue sequence in the cytoplasmic linker region of the sensor domain of DivJ has been shown to be necessary for its proper localization. Finally, the localization of all of these proteins to the correct pole depends on MreB, a bacterial homolog of the eukaryotic cytoskeletal protein, actin. MreB forms a filament that oscillates during the cell cycle between a spiral form running the length of the cell and a medial ring. The ring could be an intermediate in a reversal of polarity of the spiral as the flagellated swarmer pole matures into a stalked pole.

Clearly, although much remains to be elucidated, the correct localization of signaling proteins in *Caulobacter* morphogenesis both controls and is itself dependent upon the phosphorelay cascades in which these proteins participate.

2.2 Sporulation and Differentiation in *Bacillus subtilis*

Caulobacter is not the only bacterium that differentiates into two cell types. Under conditions of high population density and starvation, many Gram-positive bacteria belonging most notably to the genera *Bacillus* (aerobic) and *Clostridium* (anaerobic) can differentiate to form an endospore enclosed within a mother cell, which subsequently perishes. The endospore is an extremely heat- and dessication-resistant quiescent cell that is of medical significance because some endospore-forming bacteria of both genera are human pathogens (e.g. *Bacillus anthracis*, causative agent of anthrax, and *Clostridium tetani*, causative agent of tetanus). It is the extreme resistance properties of bacterial endospores that determine the minimum

requirements for successful sterilization of culture media, equipment, and surgical instruments. Phylogenetic studies have shown that the important signaling interactions and central mechanisms for sporulation are conserved amongst all endospore-forming bacteria. What seems to change the most in different lineages is the sensor domains of the receptor kinases that transmit relevant environmental information to the cells. This implies that different environmental signals relevant to endospore formation are monitored by the different species – as might be expected given the very varied habitats in which they live, including oil wells, salt lakes, soil, or infected mammalian hosts. Endospore formation has been most intensively studied in the saprophytic soil bacterium, *Bacillus subtilis*.

When *B. subtilis* cells reach a stationary phase, they enter a so-called transition state in which they must choose between several possible survival strategies – endospore formation, nutrient scavenging, and transformation. Nutrient scavenging is the strategy of first choice and involves expression of genes for chemotaxis toward nutrient sources, secretion of antibiotics that inhibit competitors, and the secretion of extracellular enzymes, such as proteases, that degrade environmental macromolecules into their constituents for uptake as nutrients. If the adverse conditions that elicit the nutrient scavenging strategy persist, but do not worsen significantly, 1–10% of the cells will opt for the transformation strategy. This involves the cell differentiating to become competent for the uptake of exogenous DNA (released into the environment by the lysis of other cells). This can lead to the formation of recombinant cells with new combinations of genetic information that could enhance survival. Because endospore formation is a complex energy-intensive process, the cell only commits itself to this path as a last resort in response to integrated information from multiple environmental and physiological signals. These signals converge on the phosphorylated form of a master regulator protein, Spo0A~P, whose levels determine whether the cell enters the sporulation pathway or not. At intermediate Spo0A~P levels, only the nutrient scavenging and (in some cells) transformation strategies are pursued. When

Fig. 5 The key stages of the sporulation cycle in *Bacillus subtilis* (from Fig. 1 of Errington, J. (2003) Regulation of endospore formation in *Bacillus subtilis, Nat. Rev. Microbiol.* **1**, 117–126). The inset (from "Todar's Online Textbook of Bacteriology", http://www.textbookofbacteriology.net/ Kenneth Todar, 2002) shows electron micrographs of sporulating cells at each of the major stages. The formation of endospores is a complex and highly regulated form of development in a relatively simple (procaryotic) cell. In all *Bacillus* species studied, the process of spore formation is similar, and can be divided into seven defined stages (0–VI). The vegetative cell (Stage 0) begins spore development when the DNA coils along the central axis of the cell as an "axial filament" (Stage I). The DNA then separates and one chromosome becomes enclosed in plasma membrane to form a protoplast (Stage II). The protoplast is then engulfed by the mother cell membrane to form a intermediate structure called a *forespore* (Stage III). Between the two membranes, the core (cell) wall, cortex, and spore coats are synthesized (Stage IV). As water is removed from the spore and as it matures, it becomes increasingly heat resistant and more refractile (Stage V). The mature spore is eventually liberated by lysis of the mother cell. The entire process takes place over a period of 6 to 7 h and requires the temporal regulation of more than 50 unique genes. Inset image and description of the stages were kindly provided by Kenneth Todar.

Spo0A~P reach high-enough levels, cells will abandon the nutrient scavenging strategy, forego the transformation strategy, and sporulate (see Fig. 5).

2.2.1 Spo0A and the Phosphorelay in Endospore Formation

In Fig. 6, which shows the basic outline of the sporulation phosphorelay mechanism, a sensor histidine-kinase autophosphorylates the conserved histidine in its transmitter domain and then passes this phosphate to an aspartate in the receiver domain of Spo0F. The phosphate is then passed via a histidine in Spo0B to an aspartate residue in the receiver domain of Spo0A, the final response regulator. Although only one sensor kinase is shown in Fig. 6, there are in fact five different sensor kinases (KinA to KinE) in *B. subtilis* that send signals into this phosphorelay (Fig. 7). The two most important of these, in order of significance, are KinA and KinB. The other kinases on their own seem to be able only to generate intermediate levels of Spo0A~P, sufficient for the nutrient scavenging and competence pathways, but insufficient to induce sporulation. Although it is clear that sporulation is initiated primarily by starvation and high cell density, the molecular identities of the signals to which the sensor kinases respond are unknown.

Spo0A~P is an active DNA-binding protein that can function either as a positive or a negative regulator of transcription and has been shown in recent microarray studies to induce, directly or indirectly, the transcription of 349 genes (66 of them in combination with the sigma factor σ^F).

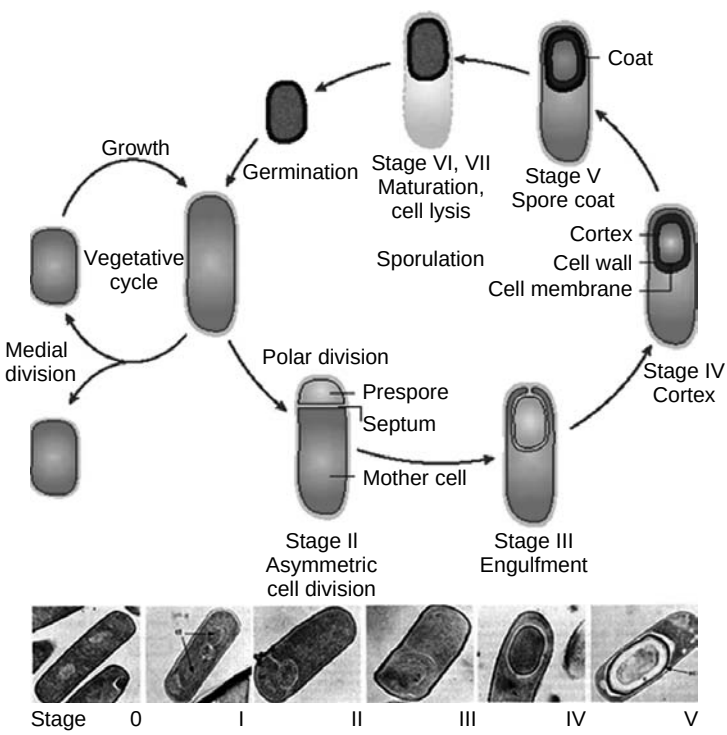

The *spoIIA*, *spoIIG*, and *spoIIE* promoters are amongst those that are directly induced by Spo0A~P, and the corresponding gene products play essential roles in Stage II of sporulation (see later). These genes are also induced indirectly by Spo0A~P via SinI. The synthesis of SinI is induced by Spo0A~P, and, once made, SinI interacts with SinR, preventing it from repressing its target genes.

SpoA~P also directly or indirectly represses expression of as many as 242 different genes, including *abrB*, which encodes a transcriptional repressor. AbrB is responsible for repressing stationary phase and sporulation functions during

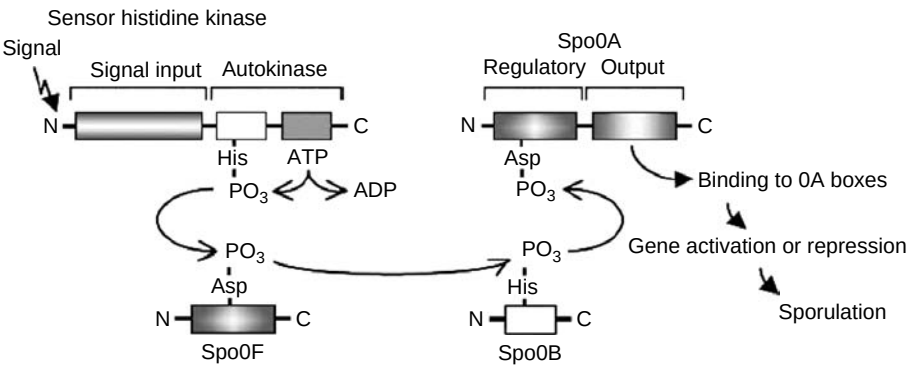

Fig. 6 The *Bacillus subtilis* sporulation phosphorelay (from Fig. 1 of Stephenson, K., Hoch, J.A. (2002) Evolution of signalling in the sporulation phosphorelay, *Mol. Microbiol.* **46**, 297–304).

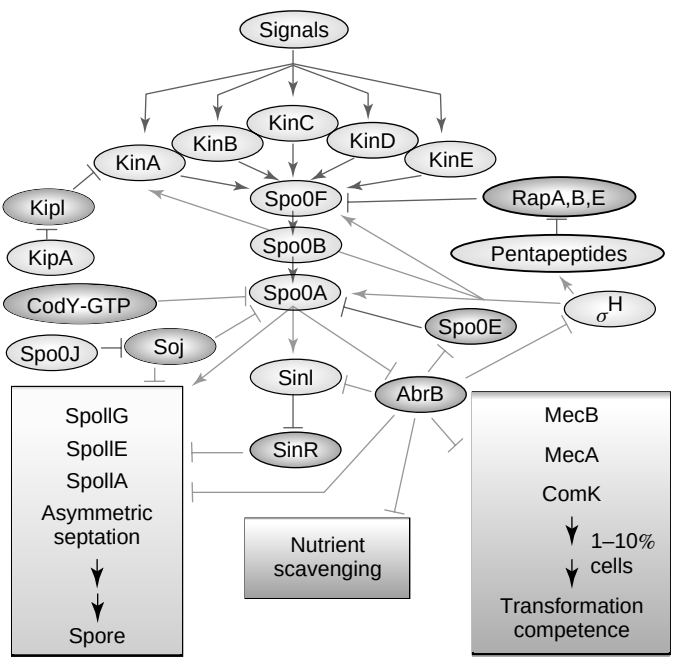

the growth phase. Its targets are upregulated by Spo0A~P indirectly by removal of AbrB-mediated repression. One of these target genes, *spo0H*, encodes the sigma factor σ^H, which targets RNA polymerase to at least 49 promoters controlling at least 87 early sporulation genes. These include the *kinA* promoter and secondary, highly active promoters for *spo0F* and *spo0A*. By this means, a positive feedback is established that results in a new, higher level of Spo0A~P, committing the cell to sporulation.

The phosphorelay leading to Spo0A~P accumulation in stationary phase cells is subject not only to the positive controls described but also to negative controls that prevent the cell from committing to sporulation inappropriately or prematurely. There are several such negative control mechanisms.

The first mechanism for limiting Spo0A~P accumulation is mediated by cellular GTP levels, which serve as the main indicator of the status of intermediary metabolism. During growth, GTP levels are high (1–3 mM), but they drop significantly upon entry to the stationary phase. Two proteins have been implicated in sensing GTP levels – CodY (early stages) and Obg (later stages). CodY in its GTP-bound form is a repressor of stationary phase/sporulation genes including *spo0A*.

The second mechanism for restraining Spo0A~P accumulation is by the action of Spo0E, a phosphatase that specifically dephosphorylates Spo0A~P. Without Spo0E, Spo0A would become maximally phosphorylated even at low levels of activity of the phosphorelay. Spo0E expression is subject to AbrB-mediated repression.

The third mechanism for restricting Spo0A phosphorylation is mediated by pentapeptide pheromone signals. The precursors of the pentapeptides are small polypeptides of about 40 amino acids that are secreted and processed proteolytically outside the cell, then imported via specific transport systems (products of the *opp* and *app* operons) into the cell where they inhibit the Rap family phosphatases RapA, B, and E. The extracellular concentrations of the signaling pentapeptides depend upon the density of stationary phase cells, so that only at high cell density do they prevent the dephosphorylation of Spo0F and thereby allow unrestrained phosphorelay to produce maximum Spo0A~P levels.

Fig. 7 Regulation of the phosphorelay and the initiation of sporulation in *Bacillus subtilis*. During the growth phase, AbrB represses genes involved in the three stationary phase survival strategies (boxed) – sporulation, nutrient scavenging, and transformation. At the onset of stationary phase, multiple signals result in activity of the phosphorelay, producing intermediate levels of phosphorylated Spo0A. These levels of Spo0A~P are sufficient to support the nutrient scavenging and, eventually, in some cells the transformation strategy for survival. Under more extreme circumstances of nutritional stress and high density, several positive feedbacks can combine with the lifting of several negative restraints to elicit production of high levels of Spo0A~P. At these high levels, the expression of sporulation-specific genes is induced and the cell progresses down the pathway of endospore formation. Green ellipses indicate proteins whose activity stimulates the phosphorelay, while red ellipses indicate proteins whose activity inhibits it. Arrowheads indicate stimulatory interactions and barred ends indicate inhibitory interactions. Blue arrows stand for interactions at the transcriptional level (gene induction or repression) and red arrows stand for interactions at the protein level (activation or inhibition). MekR, MekB, and ComK are AbrB-repressible proteins involved in the development of transformation competence. Other proteins are described in the text. (See color plate. p. xxx)

The fourth mechanism for limiting the activity of the sporulation phosphorelay is inhibition of the KinA kinase by the KipI protein, which is itself antagonized by KipA. The expression of KipI and KipA is controlled by carbon and nitrogen sources, so that phosphorelay is inhibited by the availability of nutrients and the cell persists with the nutrient scavenging strategy rather than sporulate.

The fifth mechanism by which Spo0A~P levels are limited is by Soj-mediated repression of expression of *spo0A*. Soj also represses the *spoIIA,E* and *G* operons, which are targets for Spo0A~P induction and its repressive activity is antagonized by Spo0J. During growth, the interaction between Spo0J and Soj serves what appears to be a completely different function in partitioning chromosomes into daughter cells. However, these disparate roles may be intimately connected – during sporulation Spo0J is required to localize Soj at the poles, probably in association with the chromosomal origin of replication *oriC*. Key *spo* promoters elsewhere may be liberated from Soj repression by this means, which could thus constitute a final checkpoint, delaying progression of sporulation until the chromosome has been replicated and the *oriC* regions have segregated.

2.2.2 Mother–Daughter Communication in Endospore Formation – A Spatiotemporally Regulated Sigma Factor Cascade

Once the cell has embarked on endospore formation as a result of the accumulation of high enough levels of Spo0A~P, a gene expression cascade is induced, which is coupled to morphogenetic events such as septation to form the forespore and its subsequent engulfment by the mother cell. This gene expression cascade is orchestrated by timed, serially dependent expression of specific sigma factors, each of which associates in turn with the RNA polymerase core enzyme, targeting it to a specific set of promoters (Fig. 8). The first of the sigma factors to be induced are σ^E and σ^F, whose genes belong to the *spoIIG* and *spoIIA* operons respectively (Fig. 7). Both proteins are synthesized prior to septum formation in an inactive form, and both operons also encode proteins that regulate the activities of their cognate sigma factors. Only after septum formation do these two sigma factors become active – σ^E in the mother cell and σ^F in the forespore.

Septum formation is thus a critical event in orchestrating other events in spore formation, and only when the septum has formed is the cell irrevocably committed to sporulation. The formation of the septum is a specialized cell division event that uses much of the same cellular machinery as normal cell fission, including the FtsZ ring, which forms, in this case, not at the middle of the cell, but near one end. FtsZ is one of the proteins whose synthesis is induced directly by Spo0A~P and the resulting rise in its cellular concentration is necessary for ring formation during sporulation. During sporulation, FtsZ initially forms two rings located near the cell poles, one of which (usually the larger) is subsequently "selected" for septum formation. SpoIIE, a protein phosphatase also necessary for σ^F activation (see later), is required for correct placement of the ring. There, as in normal cell division, FtsZ acts as a scaffold for recruitment of other cell division proteins to the division site and possibly as a GTP-powered contractile structure.

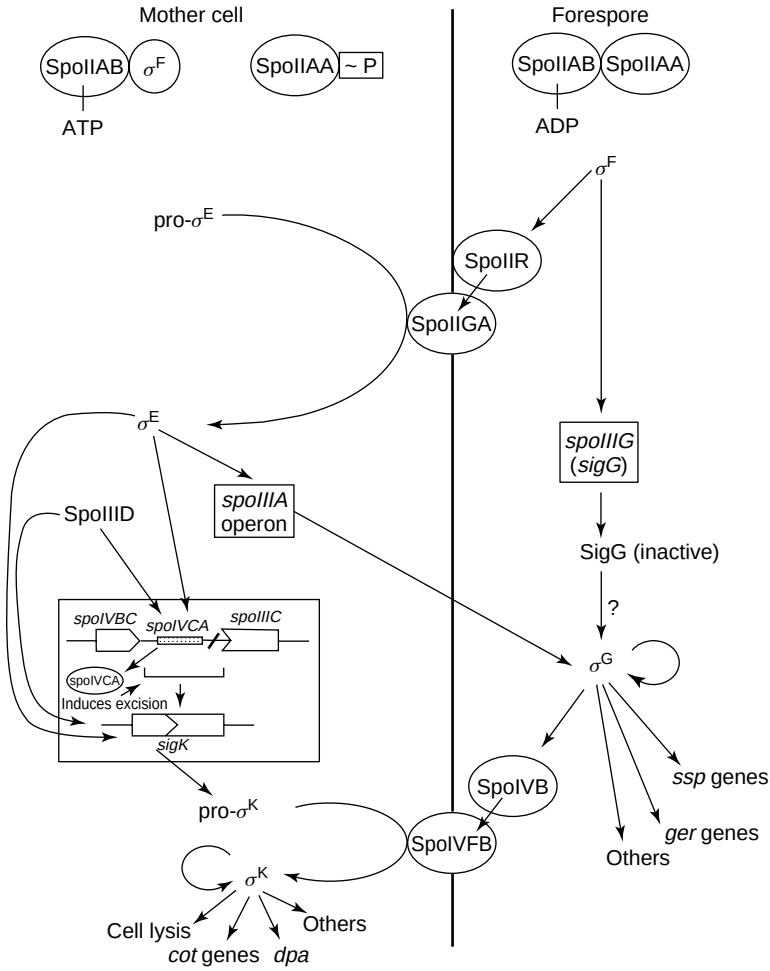

Fig. 8 The mechanisms involved in the regulation of sigma factors in the compartmentalized sporulating cell. The symbol σ and appropriate letter indicates the sigma factors, and precursor proteins of these sigma factors are indicated as pro-σ and the appropriate letter. The status of sigma F (σ^F) in each compartment post septation is indicated at the top of the figure. A vertical line represents the asymmetric septum, which divides the mother cell and the forespore. Arrows indicate activation. Circular arrows indicate autoregulation. Those proteins associated with the membrane are illustrated as embedded in the septum. Unknown regulatory mechanisms are indicated with a question mark. From Fig. 2 of Phillips, Z.E.V., Strauch, M.A. (2002) *Bacillus subtilis* sporulation and stationary phase gene expression, *Cell. Mol. Life Sci.* **59**, 392–402.

With the FtsZ ring in place, septation proceeds by a process of membrane invagination and initial synthesis of some cell wall material, which is later removed (this lends support to the idea that spore formation evolved as a modified form of normal cell division). Septation is complete before the entire forespore chromosome has been

segregated to the spore compartment. During chromosome segregation, the chromosomal origin of replication, *oriC,* is translocated and anchored to the poles by a mechanism that is not understood, but involves several proteins – DivIVA, Spo0J, Soj, and RacA. However, only about 30% of the chromosome, centered around *oriC,* is initially located within the forespore. The rest is subsequently transported there from the mother cell over a period of 10 to 15 min by a DNA translocase, SpoIIIE. It is during this brief period when the forespore is bereft of about two-third of the genes, while the mother cell has two copies of them, that some of the critical asymmetries between mother and forespore are established. These asymmetries cause the two cells to embark on different programs of gene expression, as shown in Fig. 8.

One such asymmetry arises from the location of the SpoIIA operon on the opposite side of the chromosome from *oriC*. SpoIIAB is an antisigma factor (and protein kinase), which, in the mother cell, is bound to σ^F, keeping it inactive. However, SpoIIAB is labile so that its concentration drops rapidly in the forespore in the period before delivery of the SpoIIA operon to that compartment. This would result in activation of σ^F only in the forespore. The second, and perhaps the most important asymmetry, is that the SpoIIE protein phosphatase, which is associated with the division machinery, is delivered into the forespore after septation by a SpoIIIE-dependent mechanism. There it dephosphorylates the protein SpoIIAA in opposition to the kinase activity of SpoIIAB. Dephosphorylated SpoIIAA displaces σ^F from its complex with SpoIIAB, so that the sigma factor is free to associate with RNA polymerase. The combined effects of increased SpoIIE and decreased SpoIIAB concentrations in the forespore thus result in the activation of σ^F-mediated gene transcription there, but not in the mother cell.

Amongst the genes whose transcription is directed by σ^F is *spoIIR* whose gene product is secreted into the intermembrane space of the septum bounded by the forespore and mother cell membranes. There it activates the SpoIIGA protease, an integral membrane protein located in the septal membrane of the mother cell. On the mother cell side of the membrane, the activated SpoIIGA proteolytically cleaves pro-σ^E to release active σ^E and so initiates σ^E-dependent gene expression in the mother cell.

Another of the σ^F-dependent genes expressed in the forespore is *spoIIIG*, which encodes σ^G, the next sigma factor to be deployed in the forespore. However, σ^G is also synthesized in an inactive form. Its activation depends upon the proteins encoded in the *spoIIIA* operon transcribed only in the mother cell by σ^E-directed RNA polymerase. The SpoIIIA proteins are initially located in the mother cell septal membrane and, after engulfment, are found in the topologically equivalent outer forespore membrane. How, from this location, they activate σ^G is unknown, but having been thus activated, σ^G directs the transcription of a suite of genes in the forespore involved in spore maturation and germination. Amongst the σ^G-dependent genes is *spoIVB* whose protein product in the forespore somehow sends a signal to activate the SpoIVFB protease that resides in the mother cell septal membrane in an inhibitory complex with SpoIVFA and BofA.

The role of the activated SpoIVFB protease in the mother cell membrane is to activate the final mother cell–specific sigma factor, σ^K, encoded by *sigK*. Expression of *sigK* depends upon σ^E in two ways. Firstly, *sigK* is transcribed from a σ^E-driven promoter. Secondly, *sigK* in many

Bacillus strains is interrupted by a 48-kbp element called *skin* that is excised precisely by a site-specific recombinase (SpoIVCA) whose expression also depends on σ^E (as well as on SpoIIID). SpoIVCA is encoded within the *skin* element itself. Transcription in the mother cell from σ^K-driven promoters results in the assembly of the outer spore coat and production of dipicolinic acid, which is accumulated as its calcium salt by the maturing spore to very high concentrations and is responsible for many of its essential resistance properties. Some of the σ^K-dependent gene products are also responsible for the eventual lysis of the mother cell – a form of programmed cell death – that occurs after spore maturation is complete.

2.3 Differentiation and Morphogenesis in *Myxococcus xanthus*

Caulobacter differentiation occurs as part of the growth cycle and does not depend upon the presence of neighboring cells, whereas sporulation in *Bacillus* is initiated under starvation conditions at high cell densities when growth is no longer an option. The social bacterium *Myxococcus xanthus* takes this cooperativity one step further in its foraging behavior and fruiting body formation. *M. xanthus* is one species of the myxobacteria, gliding bacteria that live in soil environments and, under conditions of starvation and high density, aggregate together to form true multicellular structures that contain resistant, dormant spores (Fig. 9). *Myxococcus* cells exhibit two different forms of gliding motility that can be separated genetically and are mediated by completely different mechanisms. Adventurous or A motility is exhibited by isolated cells, while social or S motility requires close association between cells and is exhibited only in groups. The mechanism of A motility is still not understood. S motility is mediated by extension and retraction of Type IV pili and also requires the lipopolysaccharide

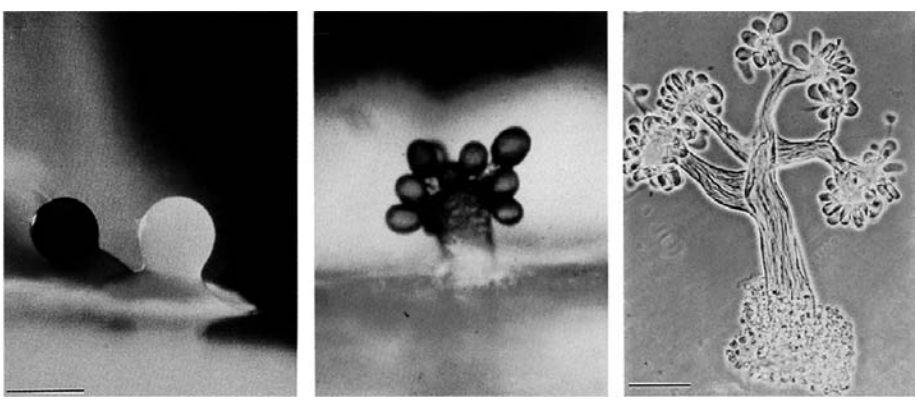

(a) (b) (c)

Fig. 9 Fruiting bodies of myxobacteria. (A) *Myxococcus fulvus*. Phase contrast. Bar, 50 mm. (B) *Stigmatella aurantiaca*. Phase contrast. Fruiting body is about 150 mm tall. (C) *Chondromyces crocatus*. Slide mount, phase contrast. Bar, 100 mm. *Myxococcus xanthus* fruiting bodies are similar in morphology (but not pigmentation) to those of *M. fulvus*. From Fig. 1 of Dworkin, M. (1996) Recent advances in the social and developmental biology of the myxobacteria, *Microbiol. Rev.* **60**, 70–102.

O-antigen on the cell surface as well as polysaccharide/protein appendages called *fibrils*. During growth, S motility allows *M. xanthus* cells to migrate together in groups and prey on other bacteria by secreting enzymes that lyse and digest the macromolecules of their victims. This so-called "wolf pack" behavior supports faster growth because of the cooperative effect of many cells secreting the necessary lytic enzymes. Under starvation conditions, S motility allows myxobacteria to aggregate and form fruiting bodies, within which a proportion of the cells differentiate into round myxospores. Mutants deficient in S motility are therefore unable to aggregate and sporulate.

2.3.1 Initiation of Development by Starvation and Extracellular Signals

The primary signal that initiates development in *M. xanthus* is starvation. As in other bacteria, this elicits the so-called *stringent response*, whereby nutritional stress is sensed by the stalling of protein synthesis and accompanying activation of RelA, the ribosome-associated (p)ppGpp synthase. The resulting increase in (p)ppGpp concentrations regulates transcription, leading to decreased rRNA synthesis, cessation of growth, and increased expression of early developmental genes. Amongst these are genes encoding proteins whose activity leads to production of extracellular developmental signals (morphogens) that accumulate and thus convey cell density information to the developing myxobacteria.

The extracellular signals controlling early development were originally discovered genetically by testing the ability of pairs of nondeveloping mutants to restore one another's ability to develop normally when mixed together. This phenomenon was called *extracellular complementation* by analogy with genetic complementation in which two different mutant genomes are mixed within a cell (by appropriate genetic crosses) and tested for restoration of normal behavior. Extracellular complementation tests of pairwise combinations of mutants revealed five distinct groups, A to E. Whereas mixtures of cells from different mutants within the same group were unable to form fruiting bodies, mixtures of mutants from different groups could undergo normal development. These results indicated that mutants within each extracellular complementation group were deficient in the production of a specific extracellular factor or signal that could still be supplied by mutants from a different group.

The mutants within each group have been found to express developmentally regulated genes normally up to a characteristic time in development. Development was arrested at this time point, which therefore represented the time at which the extracellular signal missing in that group would normally have been deployed. Thus, the order of action in development of the five extracellular signals was found to be B, A, D, E, and C (Fig. 10).

The mutant genes associated with the extracellular complementation groups in *M. xanthus* have been identified and, in some cases, provide clues to the identity of the corresponding signal. The genes in the A group are *asgA*, *B*, *C*, and *D*. Two of these (*asgA* and *asgD*) encode histidine protein kinases with both kinase and response regulator domains. AsgB is a putative transcription factor with a helix-turn-helix DNA-binding motif near its C terminus, while AsgC is the RNA polymerase sigma factor σ^{70} (also known as SigA or RpoD). AsgA and/or D could directly or indirectly sense the accumulation of the alarm molecule (p)ppGpp and activate AsgB-

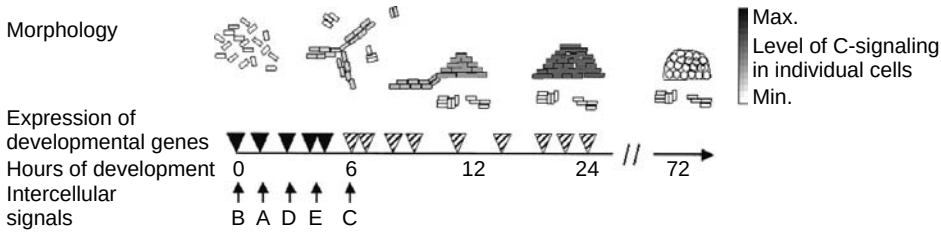

Fig. 10 Multicellular morphogenesis in *Myxococcus xanthus*. On the developmental timeline, the black triangles indicate the expression time of developmental genes expressed in all cells before the initiation of C-signaling; the hatched triangles indicate the expression time of C-signal-dependent genes, all of which are only expressed in sporulating cells. The time of action of the A to E signals is indicated by the arrows below the timeline. Cellular arrangements during the different stages of fruiting body formation are shown above the timeline. The grayscale coloring of cells according to the intensity profile shown on the right indicates the level of C-signaling in individual cells during the different morphological stages. At 72 h, cells inside the fruiting bodies have matured to spores, and the cells remaining outside the fruiting bodies are the peripheral rods. From Fig. 1 of Søgaard-Anderson et al. (2003) Coupling gene expression and multicellular morphogenesis during fruiting body formation in *Myxococcus xanthus*, *Mol. Microbiol.* **48**, 1–8.

and AsgC-directed transcription of genes encoding the extracellular proteases that produce the A-signal. These extracellular proteases are secreted and hydrolyze cell surface proteins whose constituent amino acids accordingly increase locally in concentration to an average of about 25 μM. This is insufficient to support growth, but sufficient for a subset of six amino acids (tryptophan, proline, phenylalanine, tyrosine, leucine, and isoleucine) to jointly constitute the A-signal. Proteins involved in transducing the A-signal include SasS, a histidine kinase, and SasR, its cognate response regulator.

Apart from the A-signal, the best understood of the five extracellular *M. xanthus* developmental signals is the C-signal. The processed product of the only C group gene, *csgA*, the authentic C-signal appears to be the 17-kD C-terminal portion of the 25-kDa CsgA protein. CsgA is secreted by developing *M. xanthus* cells and processed by an extracellular protease to produce the active C-signal. The full length CsgA protein shares homology with short chain alcohol dehydrogenases (SCADs), its N-terminal region containing the NAD^+-binding motif and the C-terminal portion containing the catalytic domain. The absence of the NAD^+-binding site from the active C-signal suggests that the protein's putative enzymatic activity does not play a role in its C-signaling capabilities. Exogenous addition of both the active, N-terminally truncated form, and the full-length form of CsgA are able to rescue development in C-group mutants, the full-length form doing so by first being clipped by extracellular proteases, as in normal development.

The three remaining extracellular signals (B, D, and E) are less well characterized. The only gene in the B group is *bsgA*, which encodes a protein sharing homology with the *Escherichia coli* Lon protease. The nature of the B-signal is still not understood, but its generation requires the BsgA protein and so may depend, directly or indirectly, on a proteolytic event. As is the case with A-signal transduction, a

histidine kinase (SpdS) and response regulator (SpdR) pair are involved in B-signal transduction. The only gene in the D group is *dsgA*, which encodes an initiation factor for translation, IF3. Point mutations in *dsgA* reduce the efficiency of sporulation, but do not completely prevent it, while disruption of the gene is lethal. Presumably, the generation of the D-signal depends more sensitively than other cellular functions on some aspect of IF3 activity in translation initiation. The generation of the E-signal depends upon two *esg* genes, which encode the E1α and E1β subunits of branched-chain keto acid dehydrogenase. This enzyme produces coenzyme A-linked forms of the fatty acids isovalerate, methyl butyrate, and isobutyrate, suggesting the possibility that these short-chain-branched fatty acids might constitute the E-signal. Consistent with this, the defect in E-group mutants can be rescued by exogenous supply of isovalerate.

2.3.2 C-signal Transduction and Sporulation After Aggregation

Four of the five *Myxococcus* morphogenetic signals are involved in early development – the first 6 h after the onset of starvation, during which time there is little overt morphological change in the lawn of starving bacteria. After six h of development, the cells begin to aggregate to form mounds that will develop into mature fruiting bodies. This change and the associated shifts in gene expression patterns are initiated by the C-signal. Although C-signal can rescue development in *csgA* mutants when supplied exogenously in solution, transmission of the C-signal between cells during normal development depends on direct end-to-end cell contacts because CsgA (full length and clipped) is membrane-bound and (presumably) localized at the poles. Such end-to-end contacts between cells are more frequent in aggregating streams of cells and most frequent within mounds, so that the aggregation process results in an ordered increase in the level of C-signaling and provides a link between it and the stage of morphogenesis that has been reached.

C-signal transduction pathways lead from an unidentified receptor(s) to three different responses in the cells that are separated in space and time – C-signal amplification, aggregation, and sporulation in mounds (Fig. 11).

The first of the three C-signaling pathways leads to heightened induction of transcription of *csgA* and so to an increase in the level of C-signal. The proteins encoded by the *act* operon, ActA to D mediate this response. ActA is a histidine kinase and ActB is its cognate response regulator that activates *csgA* transcription, while ActC and D somehow control the correct developmental timing of *csgA* expression. A consequence of this positive feedback is that the level of C-signaling continues to increase once it has been initiated.

The second of the C-signal-mediated responses is aggregation and, as noted above, it also feeds back to increase the level of C-signaling by increasing the frequency of end-to-end contacts between cells. The earliest recognized component in the transduction pathway leading to aggregation is the activation (possibly by phosphorylation) of FruA, a DNA-binding response regulator whose expression was induced earlier in development by the A- and E-signals. FruA, in turn, activates the Frz chemotaxis-like signaling system and this leads to aggregation. The Frz proteins include homologs of the major chemotaxis-signaling proteins in *E. coli*. FrzCD is a methyl-accepting chemotaxis protein (MCP) homolog whose role is presumed to be analogous to the MCPs

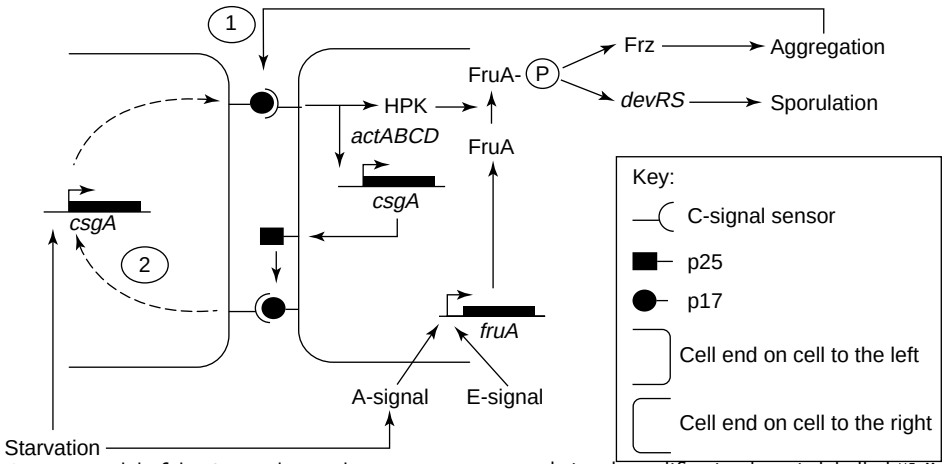

Fig. 11 Model of the C-signal transduction pathway. The schematic illustrates the C-signal transmission between two neighboring cells. For simplicity, the components in the pathway are only shown in the cell on the right. These components are also present in the cell on the left. In this cell, only *csgA* is shown to illustrate the signal amplification loop labelled "2." The second signal amplification loop is labelled "1." HPK indicates the hypothetical FruA histidine protein kinase. Modified from Fig. 2 of Søgaard-Anderson et al. (2003) Coupling gene expression and multicellular morphogenesis during fruiting body formation in *Myxococcus xanthus*, *Mol. Microbiol.* **48**, 1–8.

of *E. coli*, which function as receptors and transmembrane signal transducers. However, FrzCD differs from them in being a soluble, cytoplasmic protein, implying that the signals to which it directly responds must be intracellular, having either entered the cell or been produced in response to an extracellular signal. Cytoplasmic MCPs are known in other bacteria as well, *Rhodobacter sphaeroides*, for example. FrzE is a hybrid histidine kinase with catalytic and response regulator domains related respectively to the CheA and CheY chemotaxis proteins of *E. coli*. A second chemotaxis-like signaling system, the Dif system, also contributes to S motility and aggregation in response to lipidic (dioleoyl phosphatidyl ethanolamine) signals released by neighboring cells. It includes DifA, a transmembrane MCP whose extracellular domain seems too small to specifically bind an extracellular water-soluble ligand. The transmembrane domains of DifA may interact directly with specific lipid ligands or indirectly via other macromolecules in the membrane such as the polysaccharide fibrils that adorn the surface of *M. xanthus* cells. Both the fibrils and the lipid seem to be required for aggregation based on S motility. Other proteins of the Dif signaling system controlling S motility are DifE (a CheA homolog) and DifD (a CheY homolog).

The third C-signal transduction pathway involves FruA-mediated induction of the genetic program, leading to sporulation. Phosphorylation of FruA by an unknown upstream histidine kinase leads to induction of FruA∼P-dependent genes such as *devTRS* and repression of genes such as *dofA* whose expression depends upon the unphosphorylated FruA. It is this shift in gene expression at high C-signaling levels that leads to sporulation.

The three C-signaling pathways indicated are able to control the timing of

morphogenesis because they differ in the threshold levels of C-signaling activity required for their activation. Thus, the early, low levels of *csgA* expression are sufficient to activate further expression via the Act proteins. By 6 h of development, the C-signaling levels have reached the threshold for eliciting aggregation. As the cells enter aggregation streams and then form mounds, the level of C-signaling through end-to-end contacts increases to the point where for some of the cells within the mounds sporulation is induced. Other cells within the mounds ultimately undergo autolysis in a form of programmed cell death, while the peripheral rods, cells that failed to enter the aggregates, embark on a completely different program of gene expression. Thus, in *M. xanthus,* it is contact-mediated C-signaling that provides the necessary checkpoints, allowing the extent of morphogenesis to regulate gene expression.

3
Eukarya

3.1
Cell Cycle Control and Differentiation in *Saccharomyces cerevisiae*

Eukaryotic microorganisms, like their prokaryotic neighbors, face fluctuating levels of nutrients and food supplies in the environment, and, like them, have evolved survival strategies to deal with these uncertainties. The baker's yeast *Saccharomyces cerevisiae* has a choice of different developmental pathways when faced with declining nutrient levels (Fig. 12).

At low nitrogen levels in the presence of an abundant, fermentable carbon source, the normally unicellular, diploid yeast switch to a filamentous form of growth during which the cells elongate, bud at the pole opposite the last site of cytokinesis, and fail to separate from daughter cells after budding. This results in the generation of multicellular pseudohyphae, so called because of their macroscopic similarity to the true hyphae of filamentous fungi or molds, which are composed of multinucleated (coenocytic) tubes of continuous cytoplasm. The pseudohyphal growth form of *Saccharomyces* can be considered a nutrient scavenging mode – the elongated form has a more favorable surface-to-volume ratio for nutrient uptake and the filamentous growth allows penetration of the substratum for potentially better access to nutrient sources.

When both nitrogen and fermentable carbon sources are lacking, diploid yeast cells undergo meiotic divisions to produce a packet (ascus) containing four haploid progeny called *ascospores* – two of each of the mating types **a** and **α**. The process of meiosis provides an opportunity for genetic recombination between the members of each pair of chromosomes in the diploid. Just as is the case for *B. subtilis* cells that become transformation competent, new combinations of genetic information could arise from this process that might better suit their bearers to altered environmental circumstances. The ascospores are resistant to dessication, heat, freezing, and some chemical agents (although not as resistant as bacterial endospores) and can be dispersed in that form to environments more suited to growth. They thus represent a potent survival strategy for yeast cells that have fallen on hard times.

Germination of the ascospores produces haploid yeast cells that are able to grow vegetatively just as the diploid parent did from which they were derived. However, if a nearby cell is of the alternative mating

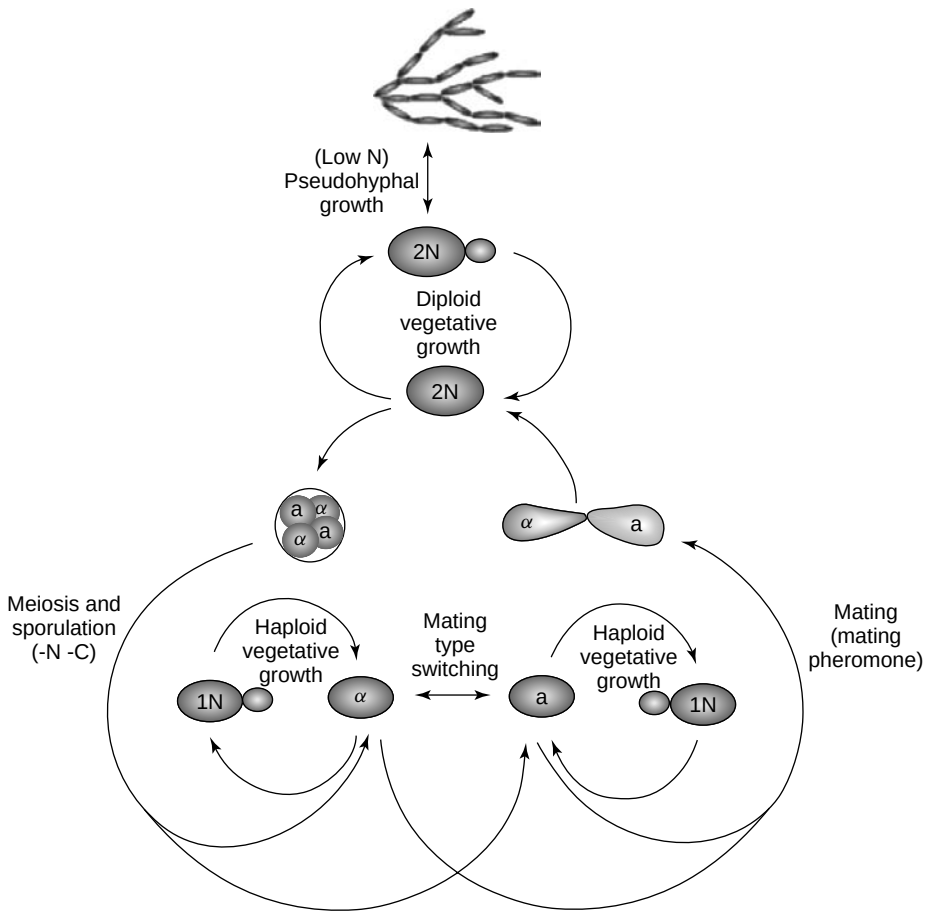

Fig. 12 Cell differentiation during the life cycle of the budding yeast *S. cerevisiae*. In the presence of abundant nutrients, yeast cells can proliferate as either haploid or diploid vegetative cells via a mitotic cell cycle. However, they can differentiate in several ways in response to both internal and environmental cues. Diploid cells, when starved for nutrients, can progress down either of the two pathways of differentiation. Cells starved for nitrogen in the presence of a fermentable carbon source can differentiate into a filamentous form known as *pseudohyphae*, which are capable of invading the substratum. Alternatively, cells starved for nitrogen on a poor carbon source can undergo meiosis and sporulation to generate haploid cells. Haploid cells of both mating types (a and α) can proliferate via a mitotic cell cycle during which they can convert to the other mating type via a programmed DNA rearrangement known as mating-type switching. In addition, when exposed to peptide mating pheromone secreted by cells of the other mating type, haploid cells can differentiate into gametes capable of mating to form a zygote competent to give rise to diploid vegetative cells. From Fig. 1 of Wittenburg, C., La Valle, R. (2003) Cell-cycle-regulatory elements and the control of cell differentiation in the budding yeast, *Bioessays* **25**, 856–867.

type, each of the two haploids responds to a pheromone released by its partner by differentiating into an elongated, pear-shaped gamete called a "shmoo". During shmooing, the cells elongate toward one another in a chemotaxis-like process, meet, and fuse to form a new diploid zygote. Additionally, during growth, yeast mother cells that have budded at least once undergo a process called *mating type switching*, whereby copies of the alternate mating type alleles are placed into the mating-type expression site on chromosome III. Thus, even a solitary ascospore that has by misfortune (or by the design of an experimenter!) found itself without sexually compatible neighbors will generate them during growth and go on to produce diploids homozygous at all except the expressed mating-type loci.

3.1.1 Pseudohyphal Growth

Of the two alternative differentiation pathways open to a starving diploid yeast cell, pseudohyphal growth is the less well studied. However, it is clear that one of the fundamental features of this form of growth is an alteration of the cell cycle – the G_1 phase of the cell cycle (following mitotic division (M phase) and preceding DNA replication (S-phase)) is shortened, while the G_2-phase (following S and preceding M) is lengthened. It is the prolonged period of polar cell growth in G_2 that generates the elongated morphology of the cell prior to unipolar budding. This shifting of the cell cycle results from modulation of the two major cell cycle checkpoints controlling entry to the S- and M-phases respectively. As in other eukaryotic cells, these checkpoints are controlled by cyclin-dependent protein kinases (CDKs), which consist of a catalytic (the kinase) and a regulatory subunit (cyclin). In *S. cerevisiae*, a single catalytic subunit, Cdc28, is used in conjunction with several different cyclins, which are employed at different stages of the cell cycle. Three G_1 cyclins (Cln1–3) control passage through G_1 and entry to S, while four mitotic or B-type cyclins (Clb1–4) control entry to and passage through mitosis.

The pathways that induce pseudohyphal growth do so by activating the G_1 cyclins (firstly Cln3, which in turn induces synthesis of Cln1 and Cln2) and inhibiting the mitotic cyclins so that passage through G_1 and entry to S is accelerated, while entry to mitosis at the end of G_2 is delayed (Fig. 13). The two best-studied pathways involve a MAP kinase cascade or cAMP-activated protein kinase A. In the former case, the pathway contains most of the elements involved in mating pheromone responses (see later), but leads to phosphorylation of a different MAP kinase – Kss1. Both pathways lead to activation of transcription factors that would induce expression of G_1 cyclins. In addition to these two pathways, the transcription factor Xbp1 and the cyclin-inhibiting tyrosine kinase Swe1 play roles in filamentation, although precisely how they are integrated into the signaling pathways is not clear.

3.1.2 Meiosis and Sporulation

S. cerevisiae zygotes enter meiosis and sporulation when deprived of both a nitrogen source and a fermentable carbon source. The entry to meiosis depends both upon repression of the G_1 cyclins so that the mitotic cycle is arrested in G_1 and upon induction of the transcription factor Ime1, which elicits transcription of meiosis-specific genes (Fig. 14). Both are a response to starvation, but Ime1 is expressed only in diploid cells heterozygous at the mating type locus (i.e. *MAT***a**/**α**). This is

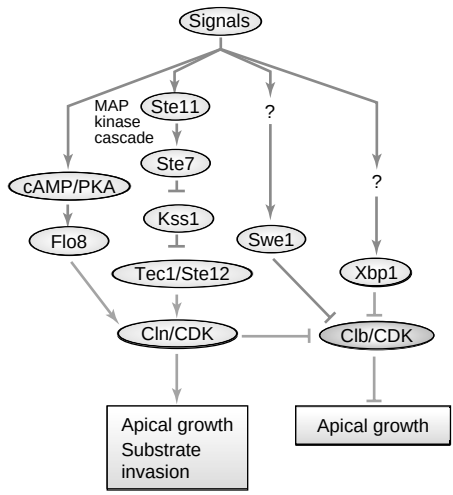

Fig. 13 Proposed pathways regulating cyclin-dependent kinases during pseudohyphal differentiation in *Saccharomyces cerevisiae*. From the nutritional signals that elicit filamentation, the pathways proceed via activation of a MAP kinase phosphorylation cascade, cAMP-activated protein kinase (PKA), Swe1 tyrosine protein kinase (a Cln/CDK inhibitor), or the stress-induced transcription factor Xbp1. Ste11 is the yeast equivalent of mammalian MAPKKK (MAPKK kinase), Ste7 the equivalent of MAPKK (MAPK kinase) and Kss1 is a MAP kinase (MAPK). Flo8 and Tek1/Ste12 are transcription factors. Unphosphorylated Kss1 inhibits Ste12 and its phosphorylation relieves this inhibition. Flo8 and Tek1/Ste12 are transcription factors. The pathways activating Swe1 and Xbp1 are uncertain and both may be coupled to the MAP kinase pathway. A variety of other proteins whose activities influence CDK activities also have corresponding effects on filamentation. Green ellipses indicate proteins whose activity stimulates apical growth, while red ellipses indicate proteins whose activity inhibits it. Arrowheads indicate stimulatory interactions and barred ends indicate inhibitory interactions. Blue arrows stand for interactions at the transcriptional level (gene induction or repression) and red arrows stand for interactions at the protein level (activation or inhibition). (See color plate. p. xxix)

because, in haploid cells, the transcriptional regulator Rme1 is active, both in repressing *IME*1 expression and in inducing G_1 cyclin expression. *IME*1 expression is also repressed by the G_1 CDKs containing cyclins Cln1 or Cln2 (Cln1–2/CDKs). In addition, Cln1–2/CDKs inhibit the nuclear localization of Ime1 with the result that it cannot interact with target DNA sequences. Accordingly, only starving, diploid cells embark upon meiosis.

During mitotic growth of both haploids and diploids, *IME*1 expression is repressed partly by Cln1–2/CDKs and partly by the transcriptional regulator SBF (which also induces the G_1 cyclins). Under these same conditions, the Ime1 polypeptide is also inactivated by virtue of being excluded from the nucleus and being inhibited from interacting with its transcriptional coactivator Ume6. Thus, mitotic growth and meiotic sporulation are mutually exclusive.

Amongst the genes whose expression is induced by Ime1/Ume6 upon entry into meiosis is *IME*2, which encodes a meiosis-specific protein kinase. *Ime2* phosphorylates and so promotes the proteolytic destruction of Sic1, an inhibitor of mitotic CDKs containing the cyclins Clb5 and Clb6 (Clb5–6/CDKs). At the same time, Ime2 activates Ndt80, the major transcriptional activator of *CLB*1–6 expression. As a result of these two activities of Ime1, the B-type cyclins are activated and elicit progression through the events of meiosis, including DNA replication, synapsis/recombination, and meiotic division. Along the way, Swe1 functions as a checkpoint kinase to inhibit Clb1–4/CDKs and meiotic progression until chromosome synapsis and recombination events are completed properly.

Swe1 also plays a checkpoint role in mitosis, restricting progression until

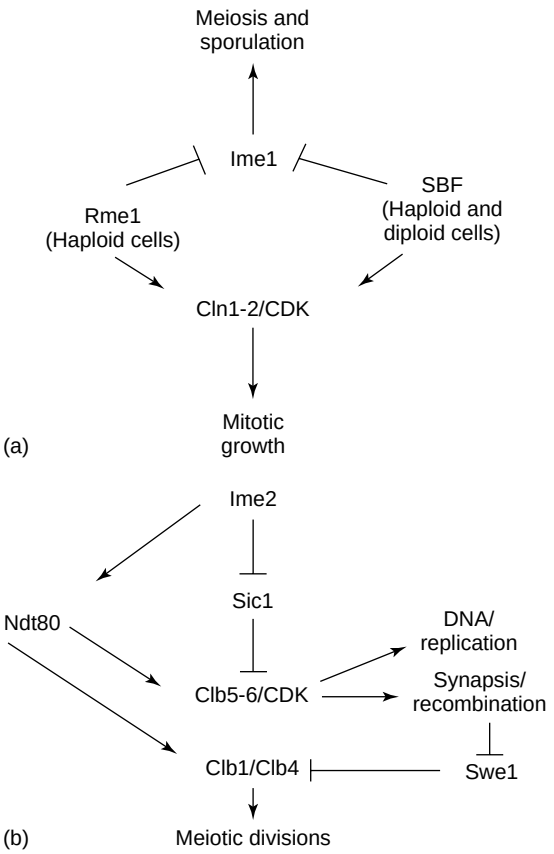

Fig. 14 Regulation of cyclin-dependent protein kinase (CDK) during meiosis and sporulation. Diploid cells undergo meiosis and sporulation when starved for nitrogen and a fermentable carbon source, but grow vegetatively when both nitrogen and carbon sources are abundant. A: Ime1, the primary inducer of meiosis-specific genes, and G1 cyclins, which promote entry into the mitotic cell cycle are subject to differential regulation by the Rme1, a haploid-specific repressor of meiosis, and SBF, the primary G1-specific transcriptional activator. B: The primary repressor of B type cyclin/CDK activity during meiosis, Sic1, and the major transcriptional activator of CLB genes, Ndt80, are differentially regulated by the meiosis-specific protein kinase Ime2. Clb-associated CDKs are also subject to inhibition by the Swe1 protein kinase until synapsis and recombination are complete. From Fig. 2 of Wittenburg, C., La Valle, R. (2003) Cell-cycle-regulatory elements and the control of cell differentiation in the budding yeast, *Bioessays* **25**, 856–867.

the actin cytoskeleton is properly assembled in preparation for budding and, as noted above, it monitors filamentation signals during pseudohyphal growth. How it monitors these disparate signals is not known and is one example of the gaps in our understanding of yeast cell differentiation pathways that remain to be filled.

3.1.3 One Locus, Three Cell Types

The result of meiotic sporulation is an ascus containing four haploid ascospores, two of each mating type (**a** and **α**), derived from one parental diploid cell. When it germinates, each haploid cell can begin vegetative growth and mitotic proliferation. However, the presence of mating pheromones from cells of the opposite mating type will induce further differentiation – gametogenesis or shmooing – as a prelude to mating. For this to occur, the cells of each mating type must synthesize and release its corresponding pheromone and must also express a receptor for the pheromone of the opposite mating type. In other words, vegetatively growing yeast cells can be one of three distinct cell types – diploid, haploid **a**, and haploid **α**. The different patterns of gene expression in these three cell types result from the expression of

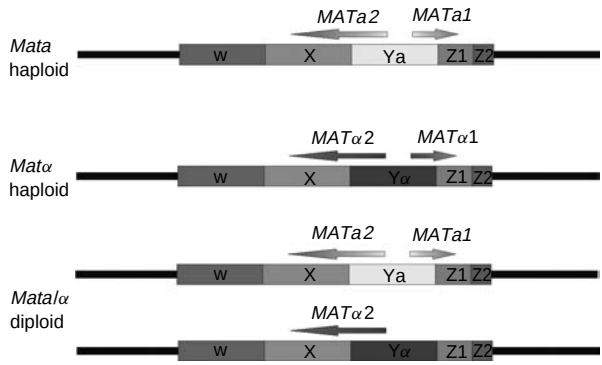

Fig. 15 The mating-type locus in *Saccharomyces cerevisiae*. The locus on Chromosome III can be divided into five regions on the basis of their presence in the two mating type alleles, *MAT*a and *MAT*α, and in the two silent copies on the left (*HML*) and right (*HMR*) arms of Chromosome III. The two *MAT* alleles differ only in the approximately 700 bp region Y, which thus has two alleles Ya and Yα. Each of the Y-regions contains divergent promoters for two transcripts, as indicated by the arrows. Most of the coding sequence of the encoded polypeptides is contained within the Y-region. The *MAT*α1 transcript is repressed in the *MAT*a/α diploid by the Mata1/Matα2 heterodimer. The X and Z1 regions are shared with *HMR*, while the W and Z2 regions are shared with *HML*. These provide regions of sequence identity that promote the strand alignment and annealing involved in the gene conversions that lead to mating type switching.

mating type–specific transcription factors encoded by the mating type alleles in residence at the mating-type (*MAT*) locus on chromosome III.

The genetic configuration of the mating-type locus in each of these three cell types is shown in Fig. 15. As described in the previous section, in haploid cells, the transcriptional repressor Rme prevents expression of Ime and diploid-specific genes. However, *RME* expression is itself repressed in the diploid by a heterodimer of the Mat**a**1 corepressor and the Mat**α**2 homeodomain DNA-binding protein. Cells of the two haploid mating types owe their differences to the activities of the *MAT***α** gene products in combination with the Mcm1 polypeptide – Mat**α**2/Mcm1 represses **a**-specific genes, while Mat**α**1/Mcm1 induces the expression of **α**-specific genes. In a haploid **a** cell, neither of the *MAT***α** gene products is present so that the constitutive state prevails – **a**-specific genes are expressed and **α**-specific genes are not. Mat**a**2 has no known function, while Mat**a**1 exerts its function only in the diploid state where it can combine with Mat**α**2.

3.1.4 Gametogenesis and Mating

The process of gametogenesis leading to mating is induced in haploid **a** and **α** cells by pheromones secreted by neighboring cells of the opposite mating type. The pheromones bind to their cognate receptors (Ste2 for the **α**-factor, Ste3 for the **a**-factor) of the ubiquitous eukaryotic GPCR

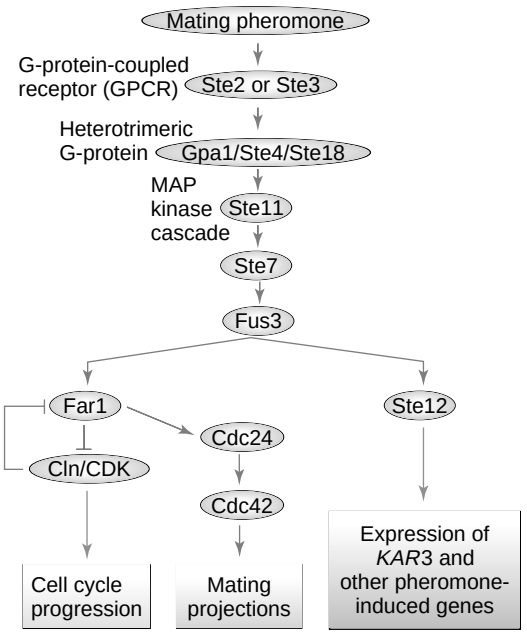

Fig. 16 Signal transduction pathways controlling gametogenesis in *Saccharomyces*. Binding of the pheromone to its cognate receptor (Ste2 = α factor receptor, Ste3 = a factor receptor) causes a conformational change and dissociation of the heterotrimeric G-protein (Gpa1 = α subunit, Ste4 = β subunit, Ste18 = γ subunit). The Ste4/Ste18 ($\beta\gamma$) heterodimer activates Ste20 a serine/threonine protein kinase, which initiates the MAP kinase phosphorylation cascade. Ste11 (MAPKKK), Ste7 (MAPKK) and Fus3 (MAPK) are each in their turn phosphorylated by the preceding kinase and thereby activated. At the end of the phosphorylation cascade, Fus3 phosphorylates and activates the Cln/CDK inhibitor Far1 as well as the transcription factor Ste12. As described in the text, the former causes cell cycle arrest in G1 and formation of mating projections, while the latter causes pheromone-inducible gene expression. Arrowheads indicate stimulatory interactions and barred ends indicate inhibitory interactions. Blue arrows stand for interactions at the transcriptional level (gene induction or repression) and red arrows stand for interactions at the protein level (activation or inhibition). (See color plate. p. xxxi)

(G-protein-coupled receptor) family and thereby elicit dissociation of the α and $\beta\gamma$ subunits of the associated G-protein. The $G_{\beta\gamma}$ heterodimer released in this way then interacts with and activates the Ste20 protein kinase, which phosphorylates Ste11, a MAPKK kinase (MAPKKK). The ensuing phosphorylation cascade culminates in activation of the MAP kinase Fus3, which, in turn, phosphorylates and activates two target proteins – the transcription factor Ste12 and the CDK inhibitor Far1 (Fig. 16).

Ste12~P initiates the characteristic pheromone-dependent changes in gene expression that accompany gametogenesis. For example, pheromone treatment induces a 20-fold increase in the expression of *KAR*3, which encodes a kinesin-like motor protein required for nuclear fusion. At the same time, Far1 that has been phosphorylated by Fus3 inhibits the G_1 CDKs – both Cln3/CDK-mediated induction of Cln1,2 expression and Cln1–2/CDK activities are suppressed, so that the cell cycle arrests in G1-phase. In addition, Far1~P binds to and activates Cdc24, a GTP exchange factor (GEF) for the small GTP-binding protein Cdc42. This promotes release of GDP and binding of GTP to Cdc42, activating its role in actin filament recruitment to and promotion of the mating projections that lend shmoos their characteristic elongated morphology. Cdc42 also plays an important role in actin recruitment during budding of vegetatively growing cells.

As is the case with so many critical developmental events, Far1 activation is coupled to the cell cycle and is autocatalytic. Thus, *FAR*1 expression is limited to G1

and is induced severalfold by pheromone signaling. Furthermore, the Far1 protein is also phosphorylated by Cln/CDK, an event that promotes its ubiquitination and proteolytic degradation. By inhibiting Cln/CDK activity, Far1 thus facilitates its own stabilization in G1. After G1, when Cln/CDK levels are high, the protein becomes labile because of its CDK-mediated phosphorylation. This pattern of regulation means that Far1-mediated gametogenesis and mating are restricted to G1. Mutants that lack Far1 still exhibit normal pheromone-induced gene expression because of Ste12 activation, but gametogenesis and mating are impaired. Restricting mating to G1 coordinates it with nuclear events, ensures that each mating partner contributes a single spindle polar body (SPB) for participation in nuclear fusion, and guarantees correct stable ploidy after fusion of nuclei in the zygote. Furthermore, the formation of the mating projection in a shmoo may be effective only in G1 when the cell is not budding.

3.1.5 Silent Copies of the Mating-type Alleles and Mating-type Switching

In homothallic yeast strains, mating can occur in a clonal population of cells derived from a single haploid ascospore. This occurs because of the phenomenon of mating-type switching whereby experienced mother cells (i.e. those that have budded at least once) switch mating types before budding again. Mating-type switching involves copying the mating-type allele present at one of the silent loci *HML* or *HMR* into the mating-type locus *MAT* to replace the existing alleles at that site

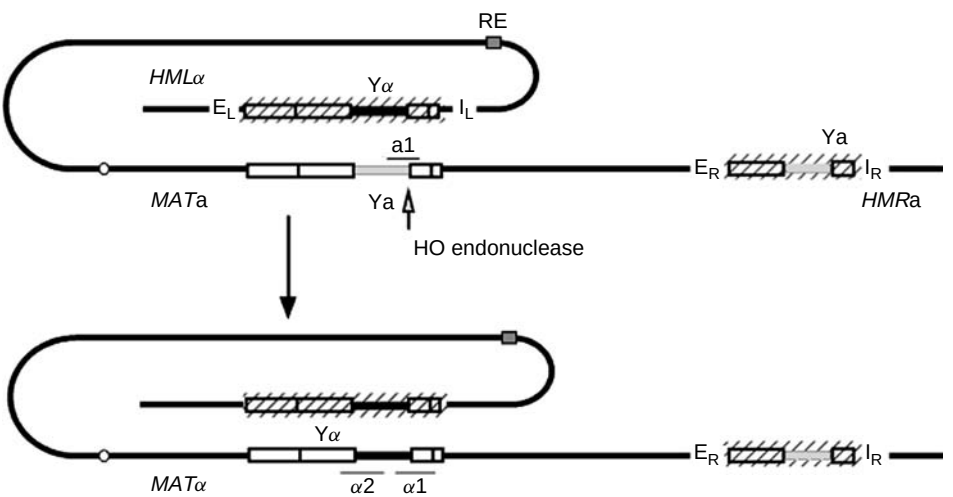

Fig. 17 Mating-type switching at the *MAT*a locus. The HO endonuclease creates a double-strand break at the Y/Z1 border of the *MAT* locus and initiates gene conversion by using the silent donor. The two donor loci (*HML* and *HMR*) are maintained in a transcriptionally inactive chromatin structure (indicated by diagonal stripes). Other shared regions of homology are indicated. The preference for the *HML* donor in *MAT*a cells is mediated by the recombination enhancer (RE) region located ~17-kb centromere-proximal to *HML*. The positions of *HML*, RE, the centromere (○), *MAT*, and *HMR* are indicated. Figure kindly provided by Prof. J.E. Haber, Brandeis University (http://www.bio.brandeis.edu/haberlab/jehsite/donorprf.html).

(Fig. 17). Unlike *MAT, HML* and *HMR* are each flanked by a pair of related, but distant silencer regions – *HML*-E, *HML*-I, *HMR*-E, *HMR*-I – sequences, which are binding sites for a complex of proteins that collaborate to produce a transcriptionally silent region of about 3 kb of highly ordered heterochromatin between the E and the I silencer sequences. These proteins include Sir1–4 (silent information regulator proteins), DNA-binding proteins Rap1 and Abf1, proteins of the origin recognition complex (ORC), which assemble at DNA replication origins, protein *trans*-acetylases and deacetylases, and chromatin assembly factors.

Each of the four silencer regions is able to act as an origin of replication when placed into an extrachromosomal plasmid, but does not function as such normally. Thus, although the ORC proteins in this case are needed for silencing, they do not normally initiate replication at these sites. The Sir proteins do not bind directly to the DNA, but bind indirectly via other DNA-binding proteins in the complexes. The Sir proteins also play other roles in the cells – in gene silencing at telomeres, in maintaining mitotic chromatin structure, in limiting rates of recombination, and in stabilizing rDNA. These functions also have repercussions for cell longevity – Sir proteins promote longevity in yeast cells and Sir2-like proteins (sirtuins – NAD^+-dependent deacetylases) do the same in animals. These and many other proteins are required for or affect silencing, enhancing, or weakening it, and their roles are not all understood.

Because of the transcriptional silencing of the *HML* and *HMR* loci, only the mating-type alleles at the *MAT* locus are expressed and thus control the gene expression patterns associated with the haploid mating type or with diploidy (see earlier). The switching of mating types that occurs in homothallic yeast strains is initiated by the activity of HO, a site-specific endonuclease that recognizes a 24-bp sequence at the border of the Y-region with the Z-region in *MAT* (Fig. 15). The silencing of *HML* and *HMR* also protects them from recognition and cleavage by HO, so that only the *MAT* locus is cut to leave a 4-bp, 3′-overhang on each side of the cleavage site. The 3′ ends are resistant *in vivo* to exonucleolytic digestion, but the 5′-ends are processed by 5′ to 3′ exonuclease activity, which includes a trio of interacting proteins (Rad50, Mre11, and Xrs2). This yields long, single stranded 3′ tails both sides of the cut site, one of which extends through the Y-region and the other of which extends through the Z-region. The latter invades the homologous region of *HML* or *HMR*, as the case may be, where it is extended across the HO cut site through the Y-region using the silent alleles as template, replicated and subsequently used to displace the remaining 3′ tail at the *MAT* locus. The nonreciprocal nature of this gene conversion means that a copy of the silent alleles has now replaced the original alleles at the *MAT* locus.

Because there are two silent loci containing mating-type alleles, the process of mating-type conversion involves a selection of either *HML* or *HMR* as the donor locus. This selection is not random – *MAT***a** cells almost always use *HML* as the donor and *MAT*$\boldsymbol{\alpha}$ cells use *HMR*. Since the great majority of strains carry $\boldsymbol{\alpha}$ alleles at *HML* and **a** alleles at *HMR*, this results in high rates of mating-type switching. This donor selectivity results from activation of the left (*HML*) end of chromosome III for mating-type donation (and for recombination) in *MAT***a** cells and its inactivation in *MAT*$\boldsymbol{\alpha}$ cells

(making *HMR* the only donor available). A 270-bp sequence called *RE* (recombination enhancer) that is responsible for donor selectivity has been identified on the left arm of chromosome III, 17 kb to the right of *HML*. This 270-bp sequence contains four regions (A to D) that are highly conserved amongst different *Saccharomyces* species.

Region C of *RE* includes a consensus binding site, an operator sequence, to which the Mat**α**2/Mcm1 repressor binds. At other such sites in the genome, this represses the expression of nearby **a**-specific genes, but, in this case, in conjunction with the Tup1 and Ssn6 proteins, it inactivates *HML* selection, suppresses other recombinational events in the region, and elicits the assembly of a chromatin structure with tightly phased nucleosomes. In *MAT***a** cells, with Mat**α**2 absent, Mcm1 binds to the same site and activates the region for recombination and *HML* selection. Thus, a 2-bp mutation in the operator eliminates the ability of Mcm1 to bind to it and abolishes the preference for *HML* in *MAT***a** cells. Furthermore, a mutant Mcm1 with reduced ability to bind to the operator also impairs *HML* preference in *MAT***a** cells.

Regions A and D, as well as another nearby sequence region (designated E), have been found to contain sites for *in vivo* binding of two transcriptional activators – the forkhead proteins Fkh1 and Fkh2 and the associated protein Ndd1. As is the case for the Mcm1 site in Region C, the Fkh1/Fkh2 binding sites in Regions A, D, and E do not function in transcription regulation, but in activation of *HML* donor preference.

The process of mating-type switching is one of the best-understood examples of a developmentally regulated DNA rearrangement that functions in the regulatory cascades involved in cellular differentiation. A prokaryotic example involving excision of an intervening sequence (*skin*) was described in the earlier section on *Bacillus* sporulation.

3.2 Differentiation and Morphogenesis in *Dictyostelium discoideum*

Yeast differentiation is fundamentally a unicellular process – even the pseudohyphae, although multicellular, do not contain specialized cells communicating with one another and serving different functions in the whole multicellular assemblage. In this respect, the social amoeba (or cellular slime mould) *Dictyostelium discoideum* is different. In the wild, *Dictyostelium* amoebae lead a predatory lifestyle, actively hunting bacteria chemotactically (bacterially secreted folate and other pterins serving as the attractants) and feeding on them by phagocytosis. Like the prokaryotic myxobacteria, *Dictyostelium* amoebae, when starved at high density, will differentiate and aggregate chemotactically (toward secreted cAMP signals) to eventually form a multicellular fruiting body within which the cells specialize to form different tissues with different functions (Fig. 18). The choices by cells as to which differentiation pathway to take have already been made at the mound stage of development when they first differentiate into recognizably different cell types – prespores and several subclasses of prestalk cells – as revealed by differential expression of cell-type-specific genes. The different cell types have been marked by expression of reporter genes such as *lacZ* under the control of cell-type-specific promoters and their developmental fates mapped (Fig. 18). Although cells in the mounds and later stages of development

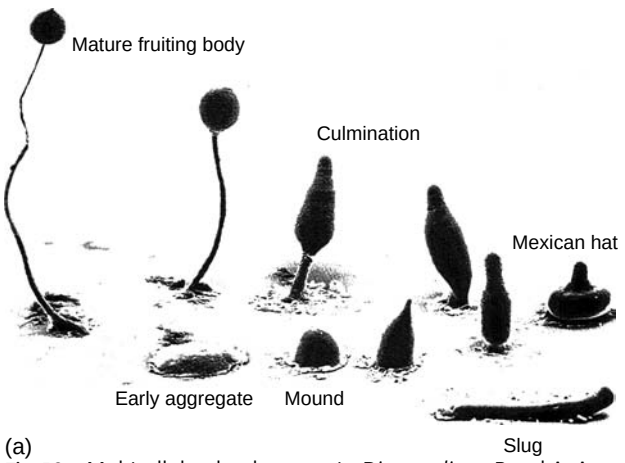

(a)

Fig. 18 Multicellular development in *Dictyostelium*. Panel A. A montage of scanning electron micrographs of stages in the *Dictyostelium* life cycle. Successive developmental stages are shown proceeding anticlockwise from the early aggregate formed by chemotactic aggregation of starving cells. The mature fruiting body is approximately 2 mm high. The original image was kindly provided by M.J. Grimson and R.L. Blanton, Biological Sciences Electron Microscopy Laboratory, Texas Tech University. Panel B. Diagrammatic representation of culmination where, for the sake of clarity, the band of pstB cells that will form the outer basal disc (see Panel C) is not shown. The *ecmA* promoter can be divided into two parts, a proximal part (the *ecmA* region) that directs expression predominantly in the cells within the tip (ie. in the pstA cells), and a distal part (the *ecmO* region) that directs expression in cells in the back of the prestalk region (the pstO cells) and in a subset (pstO:ALC cells) of the anterior-like cells (ALC). The whole *ecmA* promoter (the *ecmAO* promoter) directs expression in all these cell subtypes and has been termed the *pstAO population*. From Plate 6 of Maeda et al.(1997) *Dictyostelium discoideum – A Model Organism for Cell and Developmental Biology*. Universal Academy Press Inc., Tokyo, Japan. Image kindly provided by Prof. J.G. Williams, University of Dundee. Panel C. PstB cell behavior at culmination. The pstB cells are defined by selective staining with neutral red and because they express the *ecmB* gene at a high level relative to the *ecmA* gene. They have a complex movement pattern during slug migration. In this representation, for the sake of simplicity, separate pstA and pstO populations are not shown, but the behavior of the entire pstAO population is represented. From Plate 7 of Maeda et al. (1997). Image kindly provided by Prof. J.G. Williams, University of Dundee. (See color plate. p.xxxii)

are predestined to differentiate according to the developmental fate map shown in Fig. 18, they are not irrevocably committed to do so. In fact, right up until the final stages of differentiation with the deposition of the spore or stalk cell wall, they are able to transdifferentiate into the other cell types or even dedifferentiate and return to the vegetative state under appropriate conditions. This

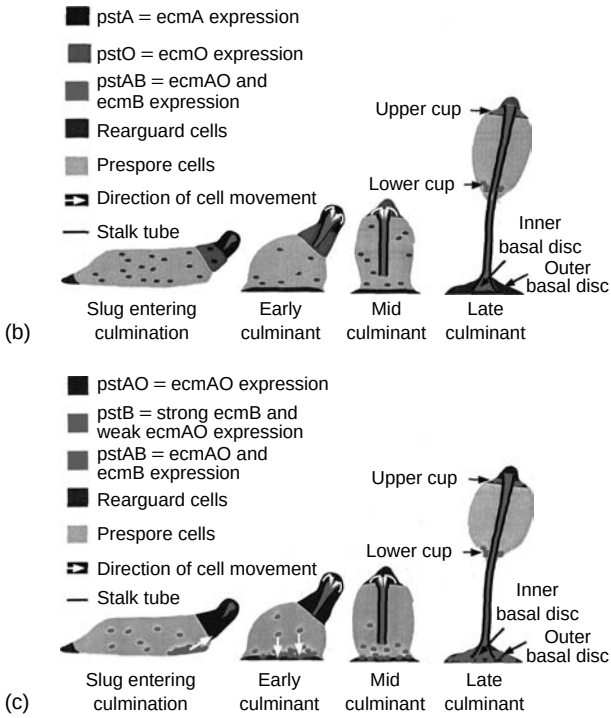

Fig. 18 (*Continued*)

developmental plasticity is important in the changeable soil environment for a fragile organism that might easily suffer mechanical injury with tissue loss during development.

Unlike the myxobacteria, *Dictyostelium* interposes between the aggregation stage and the final formation of the fruiting body, a multicellular migratory stage – the "slug", so called because of its superficial resemblance to garden slugs. The slug is phototactic and thermotactic – behaviors which are controlled by a specialized region in its tip and designed to take the organism to the soil surface, the most advantageous place from which to disperse the spores of the fruiting body. The slug contains several different tissues and cell types, distinguishable on the basis of subtle morphological differences and different patterns of gene expression (Fig. 18). Intercellular signals and associated intracellular signaling cascades control the initial transition from growth to development, the subsequent process of chemotactic aggregation, the formation of the slug and its motile behavior, the morphogenetic processes that transform the slug into a fruiting body as well as the spatial patterns, and proportions of its cell types. Not surprisingly, *Dictyostelium* devotes more of its ca. 12 000 genes to signal transduction than any other well-studied microorganism.

3.2.1 The Transition from Growth to Development

Like the prokaryotes *Bacillus* and *Myxococcus* and the eukaryote *Saccharomyces*, *Dictyostelium* differentiation is elicited in

response to two major signals – shortage of nutrients and signals from neighbors. At the onset of starvation at high cell density, polysomes are disassembled (and later reassembled), the expression of growth-specific genes (such as those encoding UMP- and GMP-synthetases) is repressed, and the earliest developmentally induced genes are expressed. Although the precise molecular nature of the starvation signal is not understood, at least one signal has been characterized that informs the cells of their density in a growing culture. That signal is PSF (prestarvation factor), a 65-kD protease-sensitive and heat-labile glycoprotein that is secreted into the medium. Its concentration is proportional to the cell density and reaches the threshold for eliciting development about four generations prior to the end of the exponential phase of growth. PSF induces some of the first genes to be expressed during development when added to growing cells that would otherwise not express them. These include *carA* (encoding the major cAMP receptor used during chemotactic aggregation), *pdsA* (which encodes a secreted cAMP phosphodiesterase needed during aggregation), and members of the discoidin gene family (possibly used during aggregation for cell–substratum interactions). However, this response is inhibited by the presence of bacteria, so that the PSF response effectively measures the ratio of amoebal to bacterial cell density. When exhaustion of the bacterial food supply becomes imminent, the first events in the growth to development transition are induced.

The PSF receptor has not been identified, but several intracellular signaling proteins have been identified as playing roles in the PSF signal transduction pathway (Fig. 19). YakA is a protein kinase whose absence results in cells that fail to aggregate and fail to repress growth-specific genes at the onset of starvation. YakA is a member of the family of Dyrk (Dual specificity Yak-related kinases) found in a variety of eukaryotes, including *Saccharomyces, Drosophila* and mammals, where their functions in the cell cycle and development, when known, appear to be analogous to those of YakA in *Dictyostelium*. The use of a temperature-sensitive YakA mutant allowed demonstration that it plays essential roles throughout development, downstream of G-protein-coupled receptors. PufA is a member of the pumilio/FBF family of proteins that bind to specific sequences at the 3′ ends of target RNA transcripts, preventing their translation. In this way, the *Drosophila* pumilio protein, together with the nanos protein, binds to nanos response elements (NREs) in the mRNA for the hunchback protein, inhibiting its expression. PufA in *Dictyostelium* binds to similar sequence elements found near the 3′ end of the coding sequence of the mRNA of PKA (protein kinase A, cAMP-dependent protein kinase). PufA was found by virtue of the fact that in a YakA-deficient mutant, disrupting the *pufA* gene restores starvation-induced PKA synthesis and development. PufA is expressed during growth, but disappears after a couple of hours of starvation. While present and active, it keeps PKA expression suppressed at the translational level even after transcription of the *pkaC* gene has been initiated in the earliest stages of development.

Intracellular cAMP signals, through PKA, act as master signals throughout *Dictyostelium* development and, like so many key developmental signals, are autocatalytic in nature. The genes whose transcription is induced by PKA-activated

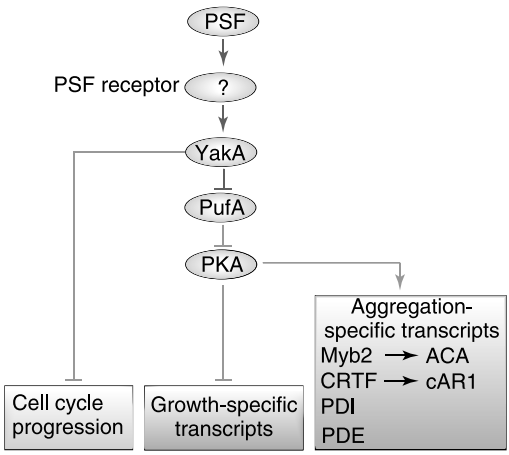

Fig. 19 The Prestarvation Factor (PSF) signaling pathway for initiating *Dictyostelium* development. PSF binds to an unknown receptor presumably belonging to the GPCR (G-protein-coupled Receptor) superfamily and elicits a signaling cascade that activates the protein kinase YakA, which, in turn, phosphorylates and inhibits the RNA-binding protein PufA. In its active, nonphosphorylated form, PufA binds to the 3′ end of cAMP-dependent protein kinase (PKA) mRNA preventing its translation. Once made and activated, PKA induces downstream genes directly or indirectly by phosphorylation of target proteins, ultimately regulating specific transcription factors such as Myb2 or CRTF (cAMP responsive transcription factor). Growth phase-specific transcripts are repressed and aggregation-specific transcripts are induced. These include those required for synthesizing and secreting cAMP such as adenylyl cyclase A (ACA) and for sensing and responding to extracellular cAMP signals, such as the cAMP receptor (cAR1). This establishes an autoactivatory feedback loop for induction of early development. YakA also phosphorylates other targets to arrest cell cycle progression at the phase shift point PS where cells exit from the cell cycle and enter differentiation. Arrowheads indicate stimulatory interactions and barred ends indicate inhibitory interactions. Blue arrows stand for interactions affecting protein expression (at the transcriptional or posttranscriptional level), and red arrows stand for interactions affecting protein activity (activation or inhibition). (See color plate. p. xxxiii)

transcription factors include those encoding the major adenylyl cyclase involved in aggregation (ACA), the cAMP receptor (cAR1), the extracellular cAMP phosphodiesterase (PDE) whose activity keeps background extracellular cAMP levels low during aggregation, and the phosphodiesterase inhibitor (PDI) that prevents premature destruction of the extracellular cAMP signal in the earliest stages of aggregation. One of the responses to extracellular cAMP is activation of ACA to produce more cAMP, which is secreted as the extracellular attractant during aggregation, but which also functions as an intracellular signal activating PKA. PKA signaling also represses growth-specific transcripts. Two of the transcription factors that participate in these signaling pathways downstream of PKA are Myb2 and CRTF (cAMP response transcription factor).

Another action of PSF signaling through YakA is to cause cell cycle arrest. Conditional YakA overexpression causes growth arrest, while growth-specific genes including *PufA* continue to be expressed in YakA-deficient mutants. As is the case with some rapidly growing embryonic cells in metazoa, the *Dictyostelium* cell cycle includes only a very short or nonexistent G1 phase. Experiments with synchronously dividing cell populations have revealed that starving cells progress about two-thirds of the way through G2 before they exit the cell cycle and differentiate. The cells that are closest to this point at the onset of starvation differentiate first, are presumably the first to start extracellular cAMP signaling, and thereby become the cells at the centers of aggregation. During subsequent differentiation in the aggregates, they preferentially differentiate into prespore cells

and ultimately into spores. The later arrivals, initially at the periphery of the aggregates, preferentially become prestalk cells, entering the tip and ultimately forming the stalk.

There are other proteins involved in the growth-to-development transition whose roles are not yet completely clear, including GdtA (a large, integral membrane protein) and AmiB, which appears to act upstream of Myb2 (as mutant phenotypes can be suppressed by ectopic expression of Myb2). Clearly, there is more to be learned about how development is initiated in *Dictyostelium*, but a solid foundation for future work has been laid.

3.2.2 Chemotactic Aggregation

PSF informs the cells of imminent depletion of the bacterial food supply and induces early gene expression in preparation for aggregation. Once the food supply has been exhausted and starvation has begun, PSF production declines and a second secreted protein called CMF (conditioned medium factor) assumes the role of a cell-density-dependent signal. CMF is an 80-kD glycoprotein whose presence is required above a threshold concentration for the onset of aggregation. Starved mutant cells lacking CMF are unable to respond to extracellular cAMP signals with the normal Ca^{2+} influx, cGMP synthesis, or cAMP synthesis, and they cannot aggregate unless CMF is supplied exogenously. However, they become responsive to cAMP within 10 s of exposure to CMF, suggesting that CMF's permissory action in aggregation is to regulate cAMP signal transduction. CMF also induces expression of a number of developmentally regulated genes including those encoding gp80 (a cell–cell adhesion molecule involved in aggregation) and SP70 (or CotB, a protein expressed in prespore cells in aggregates). The details of the CMF signaling pathways remain unknown, but there appear to be at least two CMF receptors with affinities in the nanomolar range – a G-protein-coupled receptor regulating cAMP responsiveness and a 50-kD transmembrane protein with two or three predicted transmembrane domains (CMFR1) mediating CMF-induced gene expression.

The first hours of differentiation in *Dictyostelium* render the cell competent to begin synthesizing, secreting, and responding to the chemoattractant cAMP. The cAMP signals are detected by cAMP receptors belonging to the ubiquitous eukaryotic GPCR family with their seven transmembrane domains, extracellular ligand-binding regions and intracellular signaling domains that couple to heterotrimeric G-proteins. There are 4 cAMP receptors in *Dictyostelium*, cAR1-4 encoded by the genes *carA-D*, two with high affinities (cAR1,3 – in the nM range) and two with lower affinities (cAR2,4 – in the μM range). The major receptor involved in aggregation is cAR1. Binding of extracellular cAMP to cAR1 elicits three kinds of responses:

1. Synthesis and secretion of cAMP so that the chemoattractant signal is relayed from one cell to the next.
2. Chemotactic movement toward the source of cAMP so that the cells aggregate.
3. Expression of aggregation-specific genes including those required for cAMP responses so that once it has been initiated, aggregation is an autocatalytic process.

During aggregation, the cAMP signal is pulsatile with a period of about 6 min and it spreads through the field of aggregation-competent amoebae as waves

that emanate from the aggregation center. This occurs because the cAMP signal elicits synthesis and secretion of further cAMP by neighboring cells, while at the same time extracellular phosphodiesterase degrades the cAMP left behind in the back of the passing wave. The signal transduction pathway involved in cAMP signal relay is shown in Fig. 20. Binding of cAMP to cAR1 elicits a conformational change in the receptor and dissociation of the heterotrimeric G-protein $G\alpha 2\beta\gamma$ from the receptor to release the $\alpha 2$ subunit and the $\beta\gamma$ heterodimer. The $\alpha 2$ subunit stimulates some of the downstream events involved in chemotaxis, such as cGMP synthesis, while the $\beta\gamma$ heterodimer stimulates phosphatidylinositol-3 kinases (PI3K1-3) to phosphorylate phosphatidyl inositol bisphosphate (PIP2), converting it to PIP3. PIP3 recruits the protein CRAC (cytosolic regulator of adenylyl cyclase) through its PH (pleckstrin homology) domain to the membrane from where

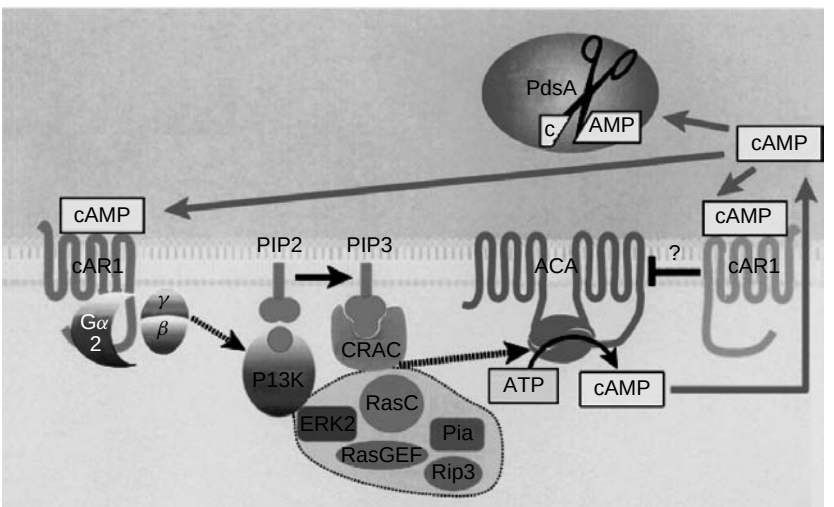

Fig. 20 The regulation of adenylyl cyclase A. Binding of cAMP to the serpentine receptor cAR1 induces dissociation of the heterotrimeric G-protein, G2, into its α and $\beta\gamma$ subunits. The $\beta\gamma$ subunits induce activation of phosphatidylinositol-3 kinase (PI3K) in a manner that is not yet understood. PI3K converts phosphatidyl inositol bisphosphate (PIP2) into phosphatidyl inositol trisphosphate (PIP3). PIP3 binds to the pleckstrin homology domain of CRAC, the cytosolic regulator of adenylyl cyclase and thereby recruits CRAC from the cytosol to the plasma membrane, where it can participate in adenylyl cyclase A (ACA) activation. A set of proteins, including the MAP kinase, ERK2, the small G-protein, RasC, the guanine nucleotide exchange factor, RasGEF, the Ras Interacting Protein (Rip3), and a novel protein, pianissimo (Pia) are also required for ACA activation. The interactions of these proteins with each other and with CRAC and ACA have not yet been clarified. cAMP produced by ACA is rapidly secreted to further activate ACA in a positive feedback loop. Binding of cAMP to cAR1 blocks ACA activation via a negative feedback loop, that is little understood. Extracellular cAMP is degraded by the phosphodiesterase PdsA, which terminates both loops and returns cells to the basal excitable state. From Fig. 2 of Saran, S., Meima, M., Alvarez-Curto, E., Weening, K.E., Rozen, D.E., Schaap, P. (2002) cAMP signaling in *Dictyostelium, J. Muscle Res. Cell Motil.* **23**, 793–802.

it activates ACA (adenylyl cyclase A). Other proteins that participate in this process include ERK2 (extracellular signal regulated kinase 2), RasC, Aimless (a guanine nucleotide exchange factor that activates Ras) and Rip3 (Ras interacting protein 3). The recruitment of CRAC to the plasma membrane occurs locally in the cell in the region of highest cAMP concentration, closest to the cAMP source. This would be the leading edge of an aggregating cell. However, the adenylyl cyclase is localized at the opposite end of the cell where synthesis and secretion of cAMP relays the attractant signal to cells further out from the aggregation center. Clearly, information must pass from the leading edge to the rear of the cell to activate ACA, but the nature of this information transfer is unknown.

Amoeboid motility involves different cytoskeletal events in different parts of the cell – pseudopod activation and extension at the leading edge, retraction in the rear. Chemotactic signals activate both of these processes in such a way as to translate a shallow extracellular attractant gradient into steep intracellular signaling and response gradients – a fast, local pseudopodium activation signal at the leading edge and a slower, global pseudopodium inhibition signal elsewhere. This spatially polarized response is not due to gradients of chemoreceptor or associated G-proteins, as fluorescently tagged forms of these are distributed uniformly or close to uniformly around the cell surface. Instead, it is the activation of PI3 kinases at the upgradient or leading edge, which exhibits the first dramatic asymmetry in chemotactic signaling events. Their action is opposed by that of PTEN, a phosphatase, which converts PIP3 (phosphatidyl inositol triphosphate) back to PIP2. After a chemotactic stimulus, the PI3 kinases phosphorylate PIP2 to form PIP3 in the membrane at the leading edge of the cell. At the same time PTEN, which is initially distributed in the cell membrane uniformly over the entire cell surface, is displaced from the front leaving it localized in the rear portions of the cell. This accentuates further the PIP3 gradient in the membranes created by the PI3 kinases. The phenotypes of null mutants show that PTEN and all three PI3 kinases in *Dictyostelium* contribute to chemotactic orientation – PI3K2 to the greatest extent, followed by PI3K1 and PI3K3. The activation of PI3 kinases involves their dynamic translocation from the cytoplasm to the membrane by an unknown recruitment mechanism that depends upon their N-terminal domains.

Once recruited to the membrane, the PI3 kinases are activated by a Ras protein through their Ras-binding domains to synthesize PIP3. This results in the recruitment of several PH domain proteins to the site including not just CRAC (see above) but also the docking protein PhDA and Akt/PKB protein kinase. PhDA acts at the leading edge to facilitate the assembly there of the necessary proteins for pseudopodial extension including F-actin. PhDA-null mutants are defective in actin assembly at the leading edge. Other proteins implicated in F-actin assembly and pseudopodium extension include the Arp2/3 complex SCAR and WASP (Wiskott-Aldrich Syndrome Protein). Amongst the target proteins for the Akt/PKB kinase is PAKa, itself a protein kinase homologous to the mammalian p21-activate kinases (PAKs). Surprisingly, in response to an attractant stimulus, PAKa translocates to the posterior of the cell where it associates with the actomyosin cytoskeleton. PAKa activity in the rear cortex of the cell inhibits myosin heavy chain kinase, causing dephosphorylation of the

myosin heavy chain by uncharacterized phosphatases. This activates myosin assembly into bipolar thick filaments, its association with actin, and its participation in actomyosin contractions. Cortical contraction of the rear of the cell is inimical to pseudopodium extension. Thus, the relocation of PAKa to the rear after its PKB-mediated phosphorylation at the leading edge provides a means whereby localized pseudopodium activation (in the front) can be coupled to and accompanied by a pseudopodium inhibition signal elsewhere (in the rear).

PTEN and PAKa are not the only signaling molecules whose site of action in chemotaxis is in the rear cortex. Attractant binding also elicits an influx of extracellular Ca^{2+} that occurs primarily in the rear cortical regions of the cell via both a G-protein-dependent and G-protein-independent pathway. A number of the actin-binding proteins that regulate its assembly into a cross-linked network of filaments are Ca^{2+}-regulated. In particular, the actin cross-linking protein, α-actinin, is Ca^{2+}-inhibited, while the actin filament severing and capping protein, severin, is Ca^{2+}-activated. The combined effect of elevated Ca^{2+} on these proteins would be a shortening of the actin filaments and a reduction in the extent of their cross-linking. This could facilitate the myosin-mediated sliding of actin filaments required for contraction of the rear cortex. However, prevention of the Ca^{2+} influx either with Ca^{2+} channel blockers or by disruption of the *iplA* (IP3-receptor-like protein) gene has no obvious effect on chemotaxis, so that the possible role for Ca^{2+} fluxes in chemotaxis remains elusive.

Another small molecule whose concentrations increase in response to an attractant stimulus (a second messenger) is cGMP. In this case, there is strong genetic evidence in the form of mutant phenotypes that cGMP plays an important role in actomyosin-mediated contractile events that inhibit pseudopodium extension in the rear cortex. The intracellular concentrations of cGMP are controlled by two guanylyl cyclases (GCA and sGC) and three phosphodiesterases (PDE3, 5, and 6), one of which (PDE6) prefers cAMP over cGMP as a substrate. The guanylyl cyclases are activated either directly or indirectly by the $\alpha 2$ subunit that has been liberated from its heterotrimeric G-protein-receptor complex by cAMP binding to the receptor. Their action is countered by the phosphodiesterases, one of which (PDE5) is itself activated by cGMP at an allosteric site separate from the catalytic site. cGMP is a small highly diffusible molecule highly suited to carrying a global pseudopodium inhibition signal to all parts of the cell upon attractant stimulation. It binds to two proteins (GbpD and GbpC) whose combined activities initiate a phosphorylation cascade terminating at the phosphorylation of the regulatory light chain (RLC) of myosin II (Fig. 21). RLC phosphorylation by MLCKA (myosin light chain kinase A) and other unidentified kinases activates both the association of myosin with the actin cytoskeleton in the rear cortex and its catalytic activity in contraction.

Chemoattractant stimulation of *Dictyostelium* amoebae has been reported to elicit a number of other molecular events whose roles in chemotaxis are unclear or controversial. These include K^+ efflux and H^+ efflux, phospholipase C-mediated release of inositol 1,4,5-triphosphate, and an associated release of Ca^{2+} from the endoplasmic reticulum into the cytoplasm. In addition, all of the described intracellular signaling processes that mediate both cAMP signal relay and chemotaxis are transient in the face of a constant extracellular

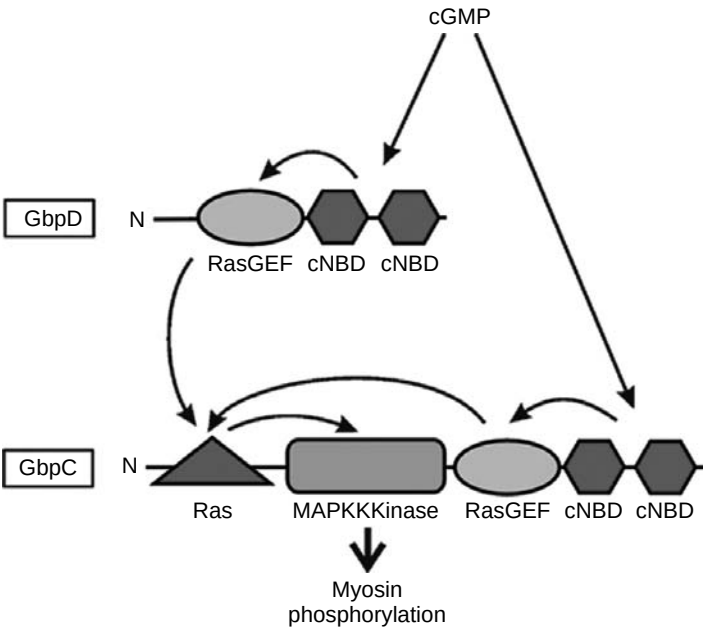

Fig. 21 Model of the function of GbpC and GbpD. The cyclic nucleotide–binding domains (cNBD) of the cGMP-binding proteins GbpC and GbpD bind cGMP, leading to the activation of the Ras guanine exchange factor domains (RasGEF), which activates the small G-protein domain (Ras) of GbpC. This results in the activation of the kinase domain of GbpC (MAPKKK), which stimulates a phosphorylation cascade eventually leading to the phosphorylation of myosin II and chemotaxis. Modified from Fig. 4 of Bosgraaf, L., van Haastert, P.J.M. (2002) A model for cGMP signal transduction in *Dictyostelium* in perspective of 25 years of cGMP research, *J. Muscle Res. Cell Motil.* **23**, 781–791.

cAMP stimulus. This is a result of adaptation processes that are poorly understood and include, but are not limited to, phosphorylation of the intracellular C-terminal domain of the cAMP receptor. The phosphorylated receptor has reduced affinity for cAMP. Adaptation not only renders the responses transient but also ensures that cells respond to relative changes in attractant concentration rather than to the absolute levels of the attractant.

The final response to pulsatile cAMP signaling during aggregation is induction of aggregation-specific gene expression. This begins with the early starvation-induced differentiation events as soon as intracellular cAMP and PKA levels are sufficient for PKA activation. However, with the arrival of each wave of extracellular cAMP, the ACA is activated and the intracellular levels of cAMP rise dramatically before most of it is secreted. The resulting pulsatile activation of PKA induces further expression of aggregation-related proteins including gp80 or contact sites A (csA), which mediates end-to-end contact of cells in aggregation streams. These contacts facilitate, but are not essential for, the natural aggregation process. Contact sites A was one of the earliest cell–cell adhesion molecules (CAMs) to be isolated from any organism and the methods used to identify

it were subsequently applied successfully to mammalian systems. During the course of aggregation, the expression of another adenylyl cyclase (ACB) is induced and, although it is not activated by extracellular cAMP signals, it contributes to basal cAMP synthesis and secretion rates and thus to the activation of PKA. Both ACA and ACB contribute to the supply of intracellular cAMP to activate PKA beyond aggregation into the later stages of development.

3.2.3 Postaggregative Development and Cell Type Choice

The result of chemotactic aggregation is that the cells crowd together into aggregation streams and then into aggregates that with time become ever more tightly packed. One consequence of this is the accumulation of a suite of extracellular protein factors that sense cell density and act to break up aggregate streams, thereby preventing aggregates from becoming too large. One of these factors is CF (counting factor), which is a high molecular weight complex (~450 kD) of several polypeptides including countin (40 kD) and CF50 (50 kD). Another of the secreted proteins affecting aggregate size is a homolog of countin called countin2 that is not part of CF. The ability of CF to promote the disintegration of aggregation streams has been attributed to the combined effects of subtle changes in cAMP signal relay, cell–cell adhesion, and actomyosin-mediated chemotactic motility.

A second consequence of cells crowding together into aggregates is that the extracellular cAMP levels experienced by the cells continue to rise during aggregation to reach micromolar concentrations in the mound and slug stages. At these concentrations, binding to the high-affinity receptors cAR1 and cAR3 is saturated so that even in the face of pulsatile cAMP signals, these receptors perceive a constant stimulus that induces expression of genes important for postaggregative development. These genes include GBF (G-box-binding factor) a transcription factor required for the expression of postaggregative genes, LagC (a cell surface signaling molecule), and RasD (a small GTPase of the Ras superfamily). GBF binds to a defined GC-rich sequence element in the promoters of these genes called the *G-box*, and is required for the transcription of both prespore-specific and prestalk-specific genes. GBF is activated by high cAMP concentrations through the cAR1 cAMP receptor in a G-protein-dependent manner. Extracellular cAMP thus acts as a cell density–sensing signal itself, coupling the progress of morphogenesis to expression of the appropriate genes.

The morphogenetic movements required for slug formation and migration as well as culmination are controlled by high concentration waves of cAMP that are relayed through the developing tissue in a manner analogous to the waves of cAMP that orchestrate the movements of aggregating cells earlier in development. These cAMP waves can be sensed by the low affinity cAMP receptors cAR2 and cAR4, which are expressed at these later stages of development.

Although high extracellular cAMP concentrations and active GBF are necessary for the expression of developmentally regulated genes during postaggregative development, they are not sufficient on their own to ensure proper morphogenesis and differentiation into the correct cell types in the correct spatiotemporal sequence. This requires additional information to inform cells of their appropriate developmental fate and position in the multicellular aggregate. This information takes two forms. The first is historical – cells that happened

to be in the mid-to-late G2-phase of the cell cycle at the time of starvation begin to differentiate and enter the aggregate first, occupying its central regions where they preferentially differentiate into prespores and ultimately spores. The later arrivals initially occupy the outer layers of the mound and preferentially locate (sort) to the tip as prestalk cells. The second type of information informing cells as to their correct developmental choices takes the form of coupled extracellular signals. These induce cells to choose either the prestalk or prespore differentiation pathways with correspondingly different locations in the aggregate and different fates in the final fruiting body (Fig. 18). They are also responsible for regulation of cell-type proportions, which remain relatively constant both in the face of major variations in slug size and after mechanical trauma to the slug, resulting in removal of parts of the tissue (such as might occur accidentally in the soil environment or be deliberately performed by an experimenter!). Slugs that have been injured in this way reorganize their tissues to form normal slugs with the correct proportions and locations of the different cell types.

The best understood of these extracellular signals is differentiation-inducing factor (DIF), a dichlorinated hexaphenone that is produced in the mound and later stages of development and induces cells to differentiate into stalk cells. In the presence of high extracellular cAMP concentrations (required both for prespore and prestalk differentiation), DIF will cause most cells in a suspension culture to enter the stalk differentiation pathway, as assayed by expression under the control of the prestalk cell-specific *ecmA* and *ecmB* promoters. DIF does this both by repressing prespore-specific gene expression and inducing prestalk-specific gene expression (although there are at least two genes expressed preferentially in prestalk cells that are not DIF-inducible, *tagB* and *carB*). The prestalk differentiation pathway leading to stalk cell formation is a form of programmed cell death that differs from, but shares some features with apoptotic and necrotic cell death pathways in mammalian cells. Interestingly enough, DIF has been found to induce cell death in mammalian cells as well, making it a potential antitumor agent.

Although the pathways involved in DIF biosynthesis in *Dictyostelium* have been elucidated, the signaling pathways involved in DIF-induced gene expression remain to be clarified – the only signaling molecule known to be involved is intracellular Ca^{2+}. DIF treatment elicits a slow prolonged elevation of intracellular free Ca^{2+} levels that temporally coincides with induction of *ecmB* expression, a marker for late prestalk differentiation. Artificially inducing a similar elevation of intracellular Ca^{2+} levels by pharmacological inhibition of ATP-driven Ca^{2+} pumps in the endoplasmic reticulum also causes *ecmB* expression. Conversely, inhibition of the DIF-induced Ca^{2+} elevation prevents *ecmB* expression. The induction of expression of an "early" prestalk gene, *ecmA*, by DIF precedes and is independent of the elevation of Ca^{2+} concentration. Thus, there are at least two different DIF signaling pathways – one involved in "early" prestalk gene expression that is independent of Ca^{2+}, and a Ca^{2+}-dependent pathway involved in "late" prestalk gene expression.

What is the relationship between the cell cycle–regulated and DIF-mediated choices of cell type? This was investigated by determining the DIF-sensitivity of cells starved at different stages of the cell cycle. Cells starved late in the cell cycle,

which preferentially differentiate into prespores were less sensitive to DIF than "stalk-loving" cells starved earlier in the cell cycle. Thus, the cell cycle–regulated preferences for particular cell types may be a result of differential DIF sensitivity. On the other hand, a mutant that is DIF-deficient (but not totally DIF-less) still shows relatively normal development, suggesting that cell-type choice is not completely DIF-dependent. It is possible that in normal development, cell-type preferences are initially based on the stage of the cell cycle at which cells entered starvation and that the role of DIF is to reinforce this and to generate correct proportions of the two cell types.

How would DIF allow cells to sense cell number in such a way as to produce correct cell-type proportions in the multicellular tissue of the slug? DIF is produced by prespore cells which are, however, comparatively resistant to its stalk-inductive activity. Prestalk cells are much more sensitive to DIF and also produce a DIF dechlorinase, which degrades it so that DIF concentrations in the slug are lower in the tip region. These properties can explain how DIF regulates cell-type proportions as follows. An overabundance of prespore cells would result in overproduction of DIF and underproduction of the dechlorinase so that DIF levels would rise. This would cause some cells to transdifferentiate from prespore into prestalk cells and thereby lead to a decrease in DIF levels. At some point, the decreased DIF levels would no longer be sufficient to induce further transdifferentiation, but would be sufficient to prevent the more DIF-sensitive prestalk cells from converting back into prespore cells. Conversely, if prestalk cells were overabundant, DIF levels would drop too low to prevent some of them from transdifferentiating into prespores and producing more DIF, until its levels rise sufficiently to stabilize the prestalk/prespore proportions. With DIF being responsible for regulating the proportions of prespore and prestalk cells in aggregates, their spatial separation in the slug results from the fact that prestalk (pstA) cells sort chemotactically to the tip. This differential sorting behavior of the different cell types couples the spatial arrangement of cells to their differentiation state.

The differentiation of the different cell types in *Dictyostelium* aggregates must require the activity of cell-type-specific transcription factors. Whereas GBF is a necessary transcription factor for both prestalk and prespore-specific genes, the transcription factors STATa and STATc regulate differentiation specifically in the stalk pathway. Both belong to a family of transcription factors that are activated by tyrosine phosphorylation. In mammalian cells, STAT phosphorylation is mediated by receptor tyrosine kinases (receptors with an intracellular tyrosine kinase catalytic domain) and Janus kinases (JAKs) coupled to ligand-induced receptor oligomerization. In *Dictyostelium*, no JAK homologs have been found, and the tyrosine kinases responsible for STAT phosphorylation remain unidentified. STATa induces some prestalk genes (e.g. *cudA*) and represses others (e.g. *ecmB*). The main defect in STATa deficient mutants is in precocious expression of late prestalk genes exemplified by the marker *ecmB*, so that they are expressed throughout the prestalk region. The phosphorylation of STATa is dependent on the cAR1 cAMP receptor (but independent of heterotrimeric G-proteins) and leads to its homodimerization, activation, and translocation to the nucleus. During normal development, STATa initially localizes to the nucleus in all cells in the mounds, but this is maintained

subsequently only in tip cells. Within minutes of exposing cells to exogenous DIF, STATc is phosphorylated and translocated to the nucleus. It represses *ecmA* and is responsible for the differential expression of *ecmA* in different subpopulations of prestalk cells. Like STATa, STATc can also function as an activator of gene expression, for example, in induction of *gapA* and *rtoA* expression in response to hyperosmotic stress.

The major roles of both STATa and STATc in prestalk differentiation appear to be as repressors. There must also be a positive transcriptional regulator responsible for DIF-induced prestalk gene expression. DimA is a bZIP/bRLZ family transcription factor that is required for DIF induction of prestalk genes – a DimA-null mutant produces, but is unresponsive to both endogenous and exogenous DIF.

Our understanding of the signaling molecules and pathways required for regulating cell-type choice and proportions in *Dictyostelium* is still very incomplete. Other proteins for which mutant phenotypes indicate roles in these pathways include the putative transcriptional regulators rZIP (which contains a RING-type Zn^{2+}-binding domain, a leucine zipper, and an SH3 binding motif) and Wariai (a homeobox protein preferentially expressed at the mound stage). Disruption of the genes encoding rZIP or Wariai cause, respectively, increases or decreases in the proportion of prespore cells. The upstream signaling pathways controlling cell-type proportions include MEKKα (a MAPKK kinase), whose absence causes a decrease in the proportion of prespore cells.

3.2.4 Culmination and Terminal Differentiation

Dictyostelium slugs migrate for a variable period of time determined by genotype and by environmental conditions – low humidity, high ionic strength, overhead light, low cell density prior to aggregation, and small slug size are amongst the circumstances that favor an earlier decision to cease migration and culminate. During development, *Dictyostelium* cells use as a source of energy proteins from vegetative growth and earlier developmental stages that are no longer required. Ammonia is released as a waste product of this process and is used as a morphogenetic signal. Many of the conditions that affect the decision to culminate do so by influencing the concentrations of free ammonia in and around the tip of the migrating slug – low NH_3 levels induce culmination. Overhead light, for example, causes the slugs to spend more time with their tips lifted into the air as a result of phototactic responses. This allows gaseous NH_3 to escape more rapidly from the slug with a concomitant decrease in the concentrations in the tip.

Culmination proceeds by a series of morphogenetic movements in which pstA cells enter a stalk tube funnel in the central regions of the tip and differentiate into stalk cells as they move downward in the direction of the substratum relative to other cells in the aggregate. Terminal differentiation of stalk cells involves enlargement and vacuolization of the cell, laying down of a cellulosic cell wall and, ultimately, membrane breakdown and cell death. The beginning of this process is marked by expression of *ecmB* and other late prestalk genes as the cell enters the stalk funnel. The other cells ascend the nascent stalk as it forms, and the prespore cells amongst them undergo terminal differentiation to form a droplet of spores atop the stalk in the mature fruiting body (Fig. 18). The terminal differentiation of spores involves cell dehydration and shrinkage, the

production of a peculiar cytoskeletal structure (the actin rod) responsible for the elongated, slightly curved shape of mature spores, and the laying down of the spore wall – an extracellular matrix of cellulose and proteins whose incorporation into the coat involves extrusion of the contents of specialized vacuoles found in prespore cells.

As with all of the other major stages of *Dictyostelium* development, intracellular cAMP signals activating PKA are essential for terminal differentiation. Thus, precocious maturation of both stalk and spore cells occurs in response to treatment with the active, membrane permeant cAMP analogue 8-bromo-cAMP and in mutants that overexpress the catalytic subunit of PKA (PKAC), or are deficient in the inhibitory, regulatory subunit of PKA (PKAR) or which lack RegA – a cAMP phosphodiesterase that limits the accumulation of intracellular cAMP. Conversely, maturation of both cell types is inhibited by overexpression of a mutant form of the regulatory subunit that permanently inhibits the catalytic subunit because it is no longer responsive to cAMP. These results show that a further elevation of intracellular cAMP concentrations and correspondingly higher levels of PKA activity act as the signal to permit or initiate terminal differentiation and culmination.

What causes cAMP levels to increase at the time when culmination is initiated? Current evidence indicates that extracellular ammonia signals detected by the cell surface receptor DHKC act to inhibit the initiation of culmination (Fig. 22). DHKC null mutants exhibit accelerated morphogenesis like mutants deficient in RegA or PKAR. Conversely, culmination is delayed by overexpression of a DHKC mutant form that lacks the DHKC sensor domain and may be constitutively active in signaling. This arrest in development can be overcome either by addition of 8-bromo-cAMP or by a deficiency in the RegA cAMP phosphodiesterase. The precocious culmination of the DHKC null mutant cannot be rescued by ammonia (which normally delays culmination), presumably because the mutant is unable to sense the presence of ammonia. The decision to culminate is thus mediated in normal development by a drop in extracellular ammonia levels, which relieves the DHKC-mediated activation of RegA and thereby allows intracellular cAMP levels to rise.

DHKC is one of the five histidine protein kinases found in *Dictyostelium* – homologs of the histidine kinases ubiquitously found in bacterial signaling systems. The phosphotransfer from DHKC proceeds via the phosphoshuttle protein RdeA, which has been shown by genetic suppressor analysis to lie upstream of RegA. RegA is a homolog of the bacterial response regulators with an N-terminal response regulator domain and a C-terminal cAMP phosphodiesterase catalytic domain.

For culmination to proceed successfully, morphogenesis and gene expression in differentiating spores and stalk cells must be coordinated and correctly timed. This coordination is achieved by intercellular signals that mediate communication between cells and regulate the differentiation process (Fig. 22). Two such signals have been identified. They are secreted peptides called *spore differentiation factors 1 and 2* (SDF1, molecular mass 1100 Da and SDF2, molecular mass 1300 Da), which were identified on the basis of their ability to elicit spore encapsulation in an *in vitro* assay. The encapsulation response to SDF1 takes about 75 min and

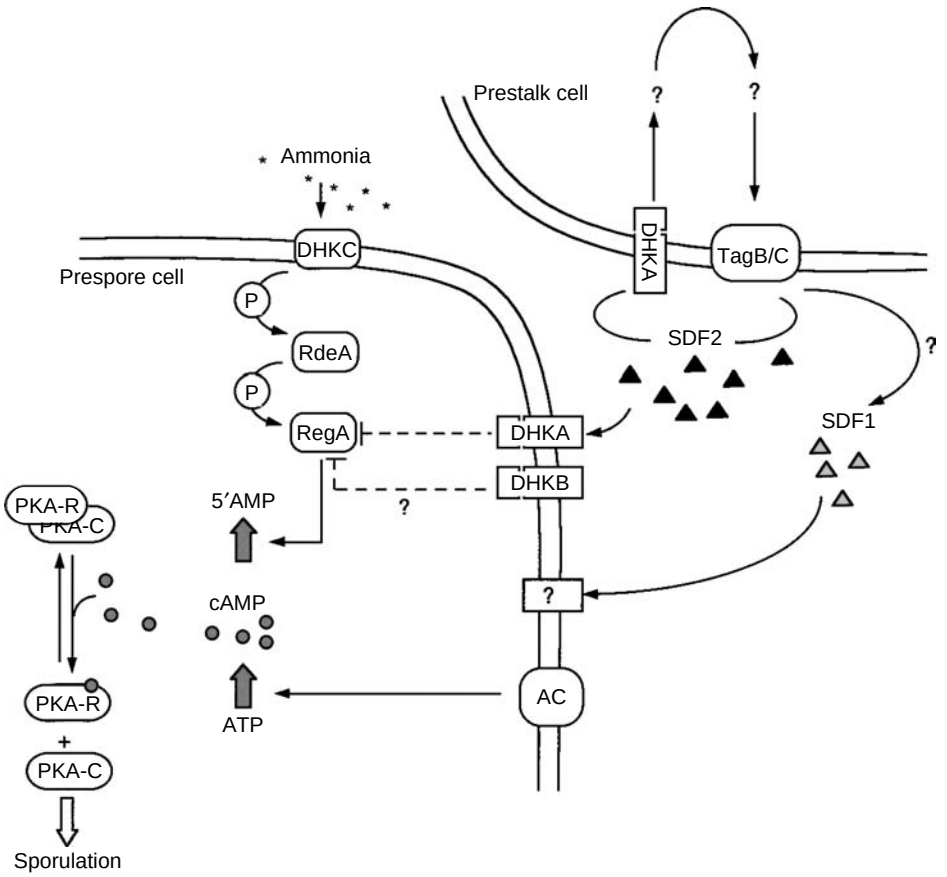

Fig. 22 Signaling pathways controlling sporulation. During culmination, protein kinase (PKA) is essential to trigger spore encapsulation. PKA activity is dependent on the level of intracellular cAMP (which controls the dissociation of the PKA-C and PKA-R subunits), which is the result of the balance between synthesis by the adenylyl cyclase (AC) and degradation by the hybrid protein phosphodiesterase/response regulator RegA. The pathway leading to the stimulation of RegA activity is proposed to be regulated by extracellular ammonia accumulation and involves a classical two component signaling phosphorelay, including the hybrid histidine kinase DHKC and the phosphodonor RdeA. Gene disruption of any of these components results in the interruption of the cascade initiated by DHKC and induces rapid sporulation. In addition to ammonia, two peptides, SDF1 and SDF2, released by the prestalk cells, have been proposed to activate intracellular pathways, the final output of which seems to be a modulation of PKA activity. SDF2 is probably processed by the protease/ABC transporter TagB/C and is probably a ligand for the histidine kinase DHKA. The downstream components of DHKA are so far unknown. DHKA shares partially redundant functions with a third histidine kinase, DHKB, a potential receptor for discadenine, an inhibitor of germination. From Fig. 5 of Aubry, L., Firtel, R. (1999) Integration of signaling networks that regulate *Dictyostelium* differentiation, *Annu. Rev. Cell Dev. Biol.* **15**, 469–517.

requires *de novo* protein synthesis, while prespore cells encapsulate in a matter of minutes in response to SDF2. SDF1 is produced at the onset of culmination when it induces the expression of culmination-specific genes, while SDF2 is made late in culmination when it elicits a wave of encapsulation that passes downward through the spore mass or sorus. SDF2 binds to the cell surface receptor DHKA, another of the *Dictyostelium* sensory histidine kinases. DHKA null mutants exhibit a specific defect in sporulation – they form fruiting bodies with long stalks and few spores. This phenotype can be rescued by 8-bromo-cAMP, by PKAC overexpression or by inactivation of either RegA or PKAR. The pathway downstream of DHKA connecting it to Reg A is unknown, but the upstream production of SDF2 has been shown genetically to involve the TagB and TagC membrane proteins expressed on prestalk cells. These proteins have a predicted protease domain, suggesting that their role in SDF2 production may be hydrolytic processing of a larger precursor polypeptide.

The final targets of signaling pathways in terminal differentiation are specific transcription factors, which are activated or inhibited as a result. In spore cell differentiation, these include Stalky and SrfA. Stalky null mutant prespore cells differentiate into stalk cells rather than spores. SrfA null mutants produce abnormal, nonviable spores. In stalk cell differentiation, one of the important transcription factors is STATa – STATa null mutants do not form stalks and fail to culminate.

3.2.5 Spore Maturation and Germination

Dictyostelium spore coats contain nine major glycoproteins (and a number of minor proteins) and the polysaccharide galuran, which are synthesized in prespore cells where they are stored in preformed complexes in specialized vesicles – prespore vesicles (PSVs). The PSV contents are the main markers for prespore cells. During encapsulation of the spores, the PSVs fuse with the plasma membrane extruding their contents, which then form the protein component of the spore coat. During subsequent spore maturation, cellulose is synthesized and extruded though the plasma membrane and the coat protein complexes bind to it via one member of the complex, SP96, which is a cellulose-binding protein. The final mature coat is 210 nm thick, and has a complex trilaminar structure with an outer protein layer, a double layer of cellulose fibrils, and an inner protein and galuran layer, with the galuran concentrated near the plasma membrane. Toward the end of spore maturation, the spore polysomes break down to single ribosomes and trehalose accumulates to high concentrations in the cytoplasm where it serves as a heat-shock protectant and as an energy source during later germination.

The spores are prevented from germinating prematurely in the sorus by extracellular accumulation of a specific germination inhibitor, discadenine (an adenine analogue), and of high concentrations of ammonium phosphate and other osmolytes. Germination is induced by removal of these (as would occur if the sorus were dispersed into the environment) in combination with a brief heat shock and in the presence of amino acids, indicating the existence of a food supply. Both discadenine and high osmolarity inhibit germination by maintaining high levels of cAMP and correspondingly high activities of PKA activity. The receptor for the discadenine signal that inhibits germination is DhkB, another of the *Dictyostelium* sensor histidine kinases that elicits high cAMP

levels by inhibiting the RegA phosphodiesterase. The osmosensor in spores is the third adenylyl cyclase of *Dictyostelium*, ACG, which is directly activated by high osmolarity to synthesize cAMP. Another signaling protein involved in germination is phospholipase C, which cleaves PIP2 to form diacyl glycerol and inositol 1,4,5-triphosphate (IP3). Diacyl glycerol in the membrane and IP3 in the cytoplasm both act as small signaling molecules (second messengers) in eukaryotes, respectively activating protein kinase C and the opening of Ca^{2+} channels in the endoplasmic reticulum. Phospholipase C–deficient mutants are unable to abort germination once it has been initiated, even in appropriate conditions that are inimical to survival of the emerging amoeba. The upstream signal for phospholipase C activation in spores and the downstream events that control abortion of germination are unknown. Spore germination is itself a developmental program that is poorly understood, but includes expression of two cellulases, which hydrolyze cellulose in the spore coat to allow emergence of the amoeba within.

4 Conclusion

The following general principles of microbial development may be gleaned from what is known about well-studied bacterial and eukaryotic systems.

1. Microbial development is usually initiated in response to nutritional stress, frequently in the form of a shortage of nitrogen combined with high cell density, which is sensed *via* secreted signaling molecules. Developmental programs presumably evolved independently in multiple lineages as a means of escaping the deadly consequences of starvation. The exception, *Caulobacter*, in which development is a constitutive part of the cell division cycle, lives in nutrient-poor aquatic environments, may be under constant nutritional stress, and so may receive little advantage by having development as an optional alternative to growth.
2. The signaling pathways that initiate development involve protein kinase cascades. Bacteria primarily use phosphorelays involving histidine protein kinases and response regulators – the so-called two-component signaling systems. Eukaryotes use tyrosine and serine/threonine protein kinases as well as histidine kinases.
3. Signaling pathways for development tend to converge upon a master regulator whose activity controls entry to major developmental stages, for example, CtrA∼P in *Caulobacter*; Spo0∼P in *B. subtilis*; PKA in *Dictyostelium*.
4. Signaling pathways for initiating development usually terminate with changes in the phosphorylation state and activity of transcriptional regulators that either induce or repress developmentally regulated genes.
5. Differentiation is coupled to the cell cycle and cells embark upon differentiation only after cell cycle arrest at a specific stage of the cell cycle. The particular cell cycle stage at which this occurs can differ in different organisms, for example, entry into gametogenesis in late G1 by yeast; onset of differentiation in middle-to-late G2 by *Dictyostelium*.
6. Development involves serially dependent, coordinate changes in gene expression whereby genes are induced in major groups under the control of common transcription factors, and

each major shift in gene expression is induced by components of the previous one, for example, sigma cascades in *Bacillus*; cAMP-elicited PKA signaling in *Dictyostelium* after the onset of starvation.

7. Transitions between the major morphogenetic stages in a developmental program involves checkpoints that couple morphogenesis to differentiation. This ensures that morphogenetic events and gene expression are coordinated, for example, septation in *Bacillus* endospore formation and the sigma factor cascade in the mother cell and the forespore; initiation of culmination by ammonia loss from the *Dictyostelium* slug when the tip is raised from the substratum for an extended time.
8. Differentiation of two or more cell types is coordinated by intercellular signals, for example, cross talk between the *Bacillus* forespore and mother cell; cAMP-, DIF- and SDF2-mediated signaling interactions between differentiating spores and stalk cells in *Dictyostelium*.

Although the details differ, these fundamental principles of development seem applicable to organisms from widely divergent groups. The reason no doubt is a combination of convergent evolution and the original presence in ancestral cells of the central mechanisms for signal transduction and gene regulation.

Acknowledgments

I am grateful to Profs. U. Jenal and J.G. Williams for helpful comments on sections of this manuscript.

See also Bacterial Cell Culture Methods; Bacterial Growth and Division; Bacterial Pathogenesis, Molecular Basis of; Bacteriorhodopsin, Molecular Biology of; *E. Coli* Genome; Methanogens and the Archaebacteria, Molecular Biology of.

Bibliography

Books and Reviews

Aubry, L., Firtel, R. (1999) Integration of signaling networks that regulate *Dictyostelium* differentiation, *Annu. Rev. Cell Dev. Biol.* **15**, 469–517.

Ausmees, N., Jacobs-Wagner, C. (2003) Spatial and temporal control of differentiation and cell cycle progression in *Caulobacter crescentus*, *Annu. Rev. Microbiol.* **57**, 225–247.

England, J.C., Gober, J.W. (2001) Cell cycle control of morphogenesis in *Caulobacter*, *Curr. Opin. Microbiol.* **4**, 674–680.

Errington, J. (2003) Regulation of endospore formation in *Bacillus subtilis*, *Nat. Rev. Microbiol.* **1**, 117–126.

Haber, J.E. (1998) Mating-type switching in *Saccharomyces cerevisiae*, *Annu. Rev. Genet.* **32**, 561–599.

Kaplan, H.B. (2003) Multicellular development and gliding motility in *Myxococcus xanthus*, *Curr. Opin. Microbiol.* **6**, 572–577.

Kessin, R.H. (2001) Dictyostelium, *Evolution, Cell Biology, and the Development of Multicellularity*, Cambridge University Press, Cambridge, UK.

Maeda, Y., Inouge, K., Takeuchi, I. (eds.) (1997) Dictyostelium *A Model Organism for Cell and Developmental Biology*, Universal Academy Press Inc., Tokyo, Japan.

Phillips, Z.E.V., Strauch, M.A. (2002) *Bacillus subtilis* sporulation and stationary phase gene expression, *Cell. Mol. Life Sci.* **59**, 392–402.

Søgaard-Andersen, L., Overgaard, M., Lobedanz, S., Ellehauge, E., Jelsbak, L., Rasmussen, A.A. (2003) Coupling gene expression and multicellular morphogenesis during fruiting body formation in *Myxococcus xanthus*, *Mol. Microbiol.* **48**, 1–8.

Wittenburg, C., La Valle, R. (2003) Cell-cycle-regulatory elements and the control of cell differentiation in the budding yeast, *Bioessays* **25**, 856–867.

Primary Literature

Anderson, P.E., Gober, J.W. (2000) FlbT, the post-transcriptional regulator of flagellin synthesis in *Caulobacter crescentus*, interacts with the 5′ untranslated region of flagellin mRNA, *Mol. Microbiol.* **38**, 41–52.

Baker, D.A., Kelly, J.M. (2004) Structure, function and evolution of microbial adenylyl and guanylyl cyclases, *Mol. Microbiol.* **52**, 1229–1242.

Bath, J., Wu, L.J., Errington, J., Wang, J.C. (2000) Role of *Bacillus subtilis* SpoIIIE in DNA transport across the mother cell-prespore division septum, *Science* **290**, 995–997.

Bosgraaf, L., van Haastert, P.J.M. (2002) A model for cGMP signal transduction in *Dictyostelium* in perspective of 25 years of cGMP research, *J. Muscle Res. Cell Motil.* **23**, 781–791.

Britton, R.A., Eichenberger, P., Gonzalez-Pastor, J.E., Fawcett, P., Monson, R., Losick, R., Grossman, A.D. (2002) Genome-wide analysis of the stationary-phase sigma factor (sigma-H) regulon of *Bacillus subtilis*, *J. Bacteriol.* **184**, 4881–4890.

Chung, C., Firtel, R.A. (2002) Signaling pathways at the leading edge of chemotaxing cells, *J. Muscle Res. Cell Motil.* **23**, 773–779.

Colomina, N., Gari, E., Gallego, C., Herrero, E., Aldea, M. (1999) G_1 cyclins block the Ime1 pathway to make mitosis and meiosis incompatible in budding yeast, *EMBO J.* **18**, 320–329.

Dirick, L., Bohm, T., Nasmyth, K. (1995) Roles and regulation of Cln-Cdc28 kinases at the start of the cell cycle of *Saccharomyces cerevisiae*, *EMBO J.* **14**, 4803–4813.

Dohlman, H.G., Thorner, J.W. (2001) Regulation of G protein-initiated signal transduction in yeast: paradigms and principles, *Annu. Rev. Biochem.* **70**, 703–754.

Dworkin, M. (1996) Recent advances in the social and developmental biology of the myxobacteria, *Microbiol. Rev.* **60**, 70–102.

Eichenberger, P., Jenson, S.T., Conlon, E.M., van Ooij, C., Silvaggi, J., Gonzalez-Pastor, J.E., Fujita, M., Ben-Yehuda, S., Stragier, P., Liu, J.S., Losick, R. (2003) The σ^E regulon and the identification of additional sporulation genes in *Bacillus subtilis*, *J. Mol. Biol.* **327**, 945–972.

Fabret, C., Feher, V.A., Hoch, J.A. (1999) Two component signal transduction in *Bacillus subtilis*: how one organism sees its world, *J. Bacteriol.* **181**, 1975–1983.

Fawcett, P., Eichenberger, P., Losick, R., Youngman, P. (2000) The transcriptional profile of early to middle sporulation in *Bacillus subtilis*, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8063–8068.

Fukuzawa, M., Abe, T., Williams, J.G. (2003) The *Dictyostelium* prestalk cell inducer DIF regulates nuclear accumulation of a STAT protein by controlling its rate of export from the nucleus, *Development* **130**, 797–804.

Fukuzawa, M., Araki, T., Adrian, I., Williams, J.G. (2001) Tyrosine phosphorylation-independent nuclear translocation of a *Dictyostelium* STAT in response to DIF signaling, *Mol. Cell* **7**, 779–788.

Gitai, Z., Dye, N., Shapiro, L. (2004) An actin-like gene can determine cell polarity in bacteria, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8643–8648.

Gomer, R., Gao, T., Tang, Y., Knecht, D., Titus, M.A. (2002) Cell motility mediates tissue size regulation in *Dictyostelium*, *J. Muscle Res. Cell Motil.* **23**, 809–815.

Gronewold, T.M.A., Kaiser, D. (2001) The act operon controls the level and time of C-signal production for *Myxococcus xanthus* development, *Mol. Microbiol.* **40**, 744–756.

Grossman, A.D. (1995) Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*, *Annu. Rev. Genet.* **29**, 477–508.

Grunenfelder, B., Rummel, G., Vohradsky, J., Roder, D., Langen, H., Jenal, U. (2001) Proteomic analysis of the bacterial cell cycle, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4681–4686.

Jacobs, C., Hung, D., Shapiro, L. (2001) Dynamic localization of a cytoplasmic signal transduction response regulator controls morphogenesis during the *Caulobacter* cell cycle, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4095–4100.

Jacobs, C., Domian, I.J., Maddock, J.R., Shapiro, L. (1999) Cell cycle-dependent polar localization of an essential bacterial histidine kinase that controls DNA replication and cell division, *Cell* **97**, 111–120.

Jacobs-Wagner, C. (2004) Regulatory proteins with a sense of direction: cell cycle signalling network in *Caulobacter*, *Mol. Microbiol.* **51**, 7–13.

Jenal, U., Stephens, C. (2002) The *Caulobacter* cell cycle: timing, spatial organization and checkpoints, *Curr. Opin. Microbiol.* **5**, 558–563.

Jiang, M., Grau, R., Perego, M. (2000) Differential processing of propeptide inhibitors of Rap phosphatases in *Bacillus subtilis*, *J. Bacteriol.* **182**, 303–310.

Jiang, M., Shao, W., Perego, M., Hoch, J.A. (2000) Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*, *Mol. Microbiol.* **38**, 535–542.

Julien, B., Kaiser, D., Garza, A. (2000) Spatial control of cell differentiation in *Myxococcus xanthus*, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9098–9103.

Kaiser, D. (2001) Building a multicellular organism, *Annu. Rev. Genet.* **35**, 103–123.

Kelly, A.J., Sackett, M.J., Din, N., Quardokus, E., Brun, Y.V. (1998) Cell cycle-dependent transcriptional and proteolytic regulation of FtsZ in *Caulobacter*, *Genes Dev.* **12**, 880–893.

Kim, S.K., Kaiser, D. (1990a) Cell alignment required in differentiation of *Myxococcus xanthus*, *Science* **249**, 926–928.

Kim, S.K., Kaiser, D. (1990b) C-factor: a cell-cell signalling protein required for fruiting body morphogenesis of *M. xanthus*, *Cell* **61**, 19–26.

Kirby, J.R., Zusman, D.R. (2003) Chemosensory regulation of developmental gene expression in *Myxococcus xanthus*, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2008–2013.

Laub, M.T., McAdams, H.H., Feldblyum, T., Fraser, C.M., Shapiro, L. (2000) Global analysis of the genetic network controlling a bacterial cell cycle, *Science* **290**, 2144–2148.

Lazazzera, B.A. (2000) Quorum sensing and starvation: signals for entry into the stationary phase, *Curr. Opin. Microbiol.* **3**, 177–182.

Lee, S., Parent, C.A., Insall, R., Firtel, R.A. (1999) A novel Ras-interacting protein required for chemotaxis and cyclic adenosine monophosphate relay in *Dictyostelium*, *Mol. Biol. Cell* **10**, 2829–2845.

Lobedanz, S., Søgaard-Andersen, L. (2003) Identification of the C-signal, a contact-dependent morphogen coordinating multiple developmental responses in *Myxococcus xanthus*, *Genes Dev.* **17**, 2151–2161.

Losick, R., Stragier, P. (1992) Crisscross regulation cell-type-specific gene expression during development in *B. subtilis*, *Nature* **355**, 601–604.

McBride, M.J., Weinberg, R.A., Zusman, D.R. (1989) 'Frizzy' aggregation genes of the gliding bacterium *Myxococcus xanthus* show sequence similarities to the chemotaxis genes of enteric bacteria, *Proc. Natl. Acad. Sci. U. S. A.* **86**, 424–428.

Montelone, B.A. (2002) Yeast Mating Type, *Nature Encyclopedia of Life Sciences*, Nature Publishing Group, London. http://www.els.net/ [doi:10.1038/npg.els.0000598].

Muir, R.E., Gober, J.W. (2001) Regulation of the late flagellar gene transcription and cell division by flagellum assembly in *Caulobacter crescentus*, *Mol. Microbiol.* **41**, 117–130.

Nebl, T., Kotsifas, M., Schaap, P., Fisher, P.R. (2002) Multiple signalling pathways connect chemoattractant receptors and calcium channels in *Dictyostelium*, *J. Muscle Res. Cell Motil.* **23**, 853–865.

Ogawa, M., Fujitani, S., Mao, X., Inouye, S., Komano, T. (1996) FruA, a putative transcription factor essential for the development of *Myxococcus xanthus*, *Mol. Microbiol.* **22**, 757–767.

Ohlson, K.L., Grimsley, J.K., Hoch, J.A. (1994) Deactivation of the sporulation transcription factor Spo0A by the Spo0E protein phosphatase, *Proc. Natl. Acad. Sci. U. S. A.* **91**, 1756–1760.

Parkinson, J.S., Kofold, E.C. (1992) Communication modules in bacterial signalling proteins, *Annu. Rev. Genet.* **26**, 71–112.

Paul, R., Weiser, S., Amiot, N.C., Chan, C., Schirmer, T., Giese, B., Jenal, U. (2004) Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel diguanylate cyclase output domain, *Genes Dev.* **18**, 715–727.

Peter, M., Herskowitz, I. (1994) Direct inhibition of the yeast cyclin-dependent kinase Cdc28-Cln by Far1, *Science* **265**, 1228–1231.

Peter, M., Gartner, A., Horecka, J., Ammerer, G., Herskowitz, I. (1993) FAR1 links the signal transduction pathway to the cell cycle machinery in yeast, *Cell* **73**, 747–760.

Quon, K.C., Marczynski, G.T., Shapiro, L. (1996) Cell cycle control by an essential bacterial two-component signal transduction protein, *Cell* **84**, 83–89.

Roberts, C.J., Nelson, B., Marton, M.J., Stoughton, R., Meyer, M.R., Bennett, H.A., He, Y.D. D., Dai, H.Y., Walker, W.L., Hughes, T.R., Tyers, M., Boone, C., Friend, S.H. (2000) Signalling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles, *Science* **287**, 873–880.

Rua, D., Tobe, B.T., Kron, S.J. (2001) Cell cycle control of yeast filamentous growth, *Curr. Opin. Microbiol.* **4**, 720–727.

Rudner, D.R., Losick, R. (2001) Morphological coupling in development: lessons from prokaryotes, *Dev. Cell* **1**, 733–742.

Saran, S., Meima, M., Alvarez-Curto, E., Weening, K.E., Rozen, D.E., Schaap, P. (2002) cAMP signaling in *Dictyostelium*, *J. Muscle Res. Cell Motil.* **23**, 793–802.

Sciochetti, S.A., Lane, T., Ohta, N., Newton, A. (2002) Protein sequences and cellular factors required for polar localization of a histidine kinase in *Caulobacter crescentus*, *J. Bacteriol.* **184**, 6037–6049.

Shimkets, L.J. (1999) Intercellular signalling during fruiting-body development of *Myxococcus xanthus*, *Annu. Rev. Microbiol.* **53**, 525–549.

Søgaard-Andersen, L., Slack, F., Kimsey, H., Kaiser, D. (1996) Intercellular C-signalling in *Myxococcus xanthus* involves a branched signal transduction pathway, *Genes Dev.* **10**, 740–754.

Sonnenshein, A.L. (2000) Control of sporulation initiation in *Bacillus subtilis*, *Curr. Opin. Microbiol.* **3**, 561–566.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., Futcher, B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization, *Mol. Biol. Cell* **9**, 3273–3297.

Spormann, A. (1999) Gliding motility in bacteria: insights from studies of *Myxococcus xanthus*, *Microbiol. Mol. Biol. Rev.* **63**, 621–641.

Stephenson, K., Hoch, J.A. (2002) Evolution of signalling in the sporulation phosphorelay, *Mol. Microbiol.* **46**, 297–304.

Strauch, M.A. (1993) Regulation of *Bacillus subtilis* gene expression during the transition from exponential growth to stationary phase, *Prog. Nucleic Acids Res. Mol. Biol.* **46**, 121–153.

Strauch, M.A., Trach, K.A., Day, J., Hoch, J.A. (1992) Spo0A activates and represses its own synthesis by binding at its dual promoters, *Biochemie* **74**, 619–626.

Sun, K., Coïc, E., Zhou, Z., Durrens, P., Haber, J.E. (2002) *Saccharomyces* forkhead protein Fkh1 regulates donor preference during mating-type switching through the recombination enhancer, *Genes Dev.* **16**, 2085–2096.

Takamatsu, H., Watabe, K. (2002) Assembly and genetics of spore protective structures, *Cell. Mol. Life Sci.* **59**, 434–444.

Thomason, P., Kay, R. (2000) Eukaryotic signal transduction via histidine-aspartate phosphorelay, *J. Cell Sci.* **113**, 3141–3150.

Thompson, C.R.L., Kay, R.R. (2000) Cell-fate choice in *Dictyostelium*: intrinsic biases modulate sensitivity to DIF signaling, *Dev. Biol.* **227**, 56–64.

Thompson, C.R.L., Fu, Q., Buhay, C., Kay, R.R., Shaulsky, G. (2004) A bZIP/bRLZ transcription factor required for DIF signaling in *Dictyostelium*, *Development* **131**, 513–523.

Van Es, S., Weening, K.E., Devreotes, P.N. (2001) The protein kinase YakA regulates G-protein-linked signaling responses during growth and development of *Dictyostelium*, *J. Biol. Chem.* **276**, 30761–30765.

Wheeler, R.T., Shapiro, L. (1999) Differential localization of two histidine kinases controlling bacterial cell differentiation, *Mol. Cell* **4**, 683–694.

Williams, H.P., Harwood, A.J. (2003) Cell polarity and *Dictyostelium* development, *Curr. Opin. Microbiol.* **6**, 621–627.

Wood, J.G., Rogina, B., Lavu, S., Howitz, K., Helfand, S.L., Tatar, M., Sinclair, D. (2004) Sirtuin activators mimic caloric restriction and delay ageing in metazoans, *Nature* **430**, 686–689.

Micronutrients, Trace Elements

Robert J. Cousins
University of Florida, Gainesville, FL, USA

1 Scope of Trace Element Micronutrients in Molecular Biology 345
1.1 Chemical Considerations 345
1.2 Nutritional Considerations 345
1.3 Classes of Involvement 345
1.4 Cellular Distributions 346

2 Structural/Catalytic/Regulatory Roles 346
2.1 Metalloenzymes and Metal-binding Domains 346
2.2 Metalloregulatory Proteins 347

3 Regulation of Gene Expression by Trace Element Micronutrients 347
3.1 Transcriptional Regulation 347
3.2 Translational Regulation 349

4 Applications and Perspectives 350

Bibliography 352
Books and Reviews 352
Primary Literature 353

Keywords

Iron regulatory Protein (IRP)
Binds in response to metal occupancy to a stem loop of nucleotides called an IRE in the 5′ untranslated region of a specific mRNA to influence its translation.

Iron-responsive Element (IRE)
A specific mRNA stem-loop structure that binds an iron regulatory protein.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

Metal-responsive Element (MRE)
Specific sequence of promoter region of gene that enhances transcription through a specific transcription factor (MRE-binding transcription factor) that requires metal occupancy for DNA binding.

Metal-responsive Transcription Factor (MTF)
In response to metal occupancy, translocates to the nucleus and binds to a specific MRE sequence of the DNA.

Metalloenzyme
An enzyme that requires one or more atoms of a metal for catalytic activity and/or for structure required for activity.

Trace Element Micronutrient
An essential nutrient metal required in the diet in small (μg or mg) amounts per day.

Zinc Finger
A zinc-binding domain of a protein that has an X-Cys-X_2-Cys-X_n-Cys-X_{2-5}-Cys-X (or His substituted for Cys) motif that, upon Zn(II) binding, forms a structure capable of binding DNA or RNA or interacting with another protein with a compatible zinc finger.

Trace element micronutrients is a term that refers to the metals (inorganic elements) in the diet that are essential for specific cellular processes and are required in only small amounts (mg or less) daily. In mammals, required dietary trace elements include Cu, Fe, Mn, Se, and Zn. They perform catalytic, regulatory, and/or structural roles related to gene expression. Zinc has a major role in maintaining conformations (i.e. zinc fingers) necessary for proteins to bind to specific DNA sequences (regulatory elements) or protein–protein interactions required for specific cellular functions. Zinc can also influence transcription through metal-binding proteins (metal-responsive transcription factors) that, during metal occupancy, initiate translocation to the nucleus for binding to specific DNA sequences (metal response elements) in the promoter/enhancer region of genes. Regulation of gene expression by iron can occur through metal-requiring, RNA-binding proteins. These metal-binding proteins influence translation of specific mRNAs via nucleotide sequences in the untranslated regions. Trace element micronutrients have potential applications in medicine and biotechnology either directly as signaling molecules for regulation of existing or chimeric genes or indirectly as therapeutic agents by targeting specific metal-binding sites in regulatory factors.

1 Scope of Trace Element Micronutrients in Molecular Biology

1.1 Chemical Considerations

Cellular constituents, particularly proteins, provide ligands (e.g. amine, thiolate, carboxyl, etc.) for metal binding, which follow approximately the Irving–Williams order of divalent ions: Cu > Zn ≥ Ni > Co > Fe > Mn > Mg > Ca. Cellular transport systems regulate intracellular trace element concentrations through homeostatic control of uptake and efflux. These processes limit the influences of natural abundance and thermodynamic constraints in determining which elements participate in specific cellular functions. If ligand concentrations and metal concentrations were not controlled within cells, abnormal binding of Cu would predominate for most ligands. Only Cu, Fe, and Zn will be discussed here as these nutrients have received the bulk of attention with respect to mammalian molecular biology.

1.2 Nutritional Considerations

Micronutrients as a general term describes both essential organic molecules (vitamins) and metals (trace elements). In contrast to vitamins, metals are utilized by all living organisms to sustain cellular processes. Required trace elements include Cu, Fe, Mn, Se, and Zn. In humans, these are required in the diet in μg to mg amounts per day to maintain balance with endogenous losses. The composite cellular and extracellular need for each trace element to satisfy the needs of binding sites as balanced against turnover and endogenous losses comprises the dietary requirement for the micronutrient.

Food abundance in the developed world has minimized the incidence of deficiencies of these trace elements. Genetic variance in trace element utilization and function suggest individual requirements must be considered. This is particularly true where intakes are marginally adequate through selective diets and/or consumption of highly processed/unfortified foods. Nutritional inadequacy in the developing world is prevalent. Iron and zinc are the trace elements most likely to be in limited supply. How such nutritional deficits influence gene regulation and epigenetic effects that influence human health are active areas of investigation.

1.3 Classes of Involvement

There are three classes of involvement for trace element micronutrients (metals) in molecular biology. These are catalytic, structural, and regulatory. Metalloenzymes, where metals provide a catalytic function, comprise the first class. Another class is structural, with metals providing interaction among various binding groups of specific motifs to facilitate conformations necessary to achieve unique opportunities for specific interaction. The third class is regulatory, as exemplified by the metal binding to specific trans-acting proteins, which then provide signaling to initiate, enhance, or inhibit transcription of genes through interaction with specific DNA sequences, that is, metal-responsive elements (MRE). Metal-responsive RNA-binding proteins provide another example of regulatory involvement. These proteins, while performing a regulatory function, act through structural changes. Collectively, these classes of biological function of trace

element micronutrients are such that they provide a vital essential link between the organism and its external nutrient supply.

Eukaryotic cells employ trace elements to differing extents. Overall, zinc has the greatest spectrum of gene-related involvement through catalytic, structural, and regulatory roles in all eukaryotes. Iron and copper have catalytic roles in lower eukaryotes and a complex array of regulatory roles through metal-responsive transcription factors or metal-binding elements of RNA binding proteins. These are directed at iron and copper uptake, trafficking, and homeostasis. Most such information has been obtained using yeast (a fungus). In higher eukaryotes, copper functions in catalytic roles, while iron has catalytic and structural roles, most related to oxygen utilization. Transcriptional regulation of genes in mammalian cells is limited to zinc. The teleological explanation most likely is that zinc is redox neutral with respect to the metal atom.

1.4 Cellular Distributions

Concentrations of trace elements in various cell types vary greatly but usually are no higher than the μg/g tissue range. Frequently, such concentrations are expressed on a protein, DNA, or dry weight basis. Cells differ with respect to which organelles trace elements are most abundantly distributed. However, metal occupancy at specific binding sites, rather than concentration, determines functional effects.

Lower eukaryotes vary widely in response to the trace element content of the environment. For example, metal-resistance genes with elaborate control mechanisms have evolved to deal with elevated levels of a particular metal present in the environment through regulated transport and compartmentalization. Mammalian cells within an integrative system are not subjected to such extremes, in part through powerful homeostatic mechanisms operative at both the cell and organ level. Therefore, the tissue metal concentration in cells of animals that consume diets deficient in that essential trace element or receive an excess of that element are frequently within normal limits. Functional deficiency may occur only after depletion of a critical pool or compartment of the trace element, which then affects specific binding sites. Unfortunately, our understanding of the cell biology of trace elements lags behind our understanding of roles of these micronutrients in molecular biology.

2 Structural/Catalytic/Regulatory Roles

2.1 Metalloenzymes and Metal-binding Domains

Characterization of individual metal-binding molecules has shown that stoichiometry is usually 1 to 4 metal atoms per binding molecule. There are exceptions, for example, ferritin (4500 atoms Fe/molecule) and metallothionein (7 atoms Zn or 12 atoms Cu/molecule). In addition, abundance of metal-binding molecules ranges from a few to thousands of molecules per cell. Trace elements are essential to the enzymology of molecular biology. For example, all of the RNA nucleotide transferases (RNA polymerases) are zinc-requiring metalloenzymes. Metalloenzyme systems are not uniformly susceptible to alteration in dietary trace element restriction or supplementation.

However, many manipulations in molecular biology use metal chelation (e.g. with EDTA) as a method of controlling activity of metal-requiring enzymes.

2.2 Metalloregulatory Proteins

Metalloproteins that act as DNA-binding transcription factors, trans-acting MRE-binding proteins, mRNA-binding proteins, and proteins exhibiting metal-dependent interactions with other proteins usually bind metals through thiolate groups from cysteine (Cys) and/or nitrogen from histidine (His). Very little is known about the ligand-exchange reactions that are necessary for metals to fulfill these regulatory roles. Progress has been made regarding how dietary abundance and transport mechanisms are necessary for trace elements to meet these needs.

3 Regulation of Gene Expression by Trace Element Micronutrients

Trace elements provide signals to systems that influence rates of either gene transcription or mRNA translation. They do not act directly. Rather, through recognition that requires specific coordination chemistry, metals bind to metalloregulatory proteins (sensors or receptors) and influence signaling pathways. This ultimately augments production of a specific protein. In the classical Newtonian sense, these micronutrients (metals) act as inducers (or repressors) of the system. Since the dietary supply provides the source of the inducer (or repressor), fluctuations in intake and/or physiologic factors influencing utilization or tissue distribution of the micronutrient are determinants of how effective the trace element will be in regulating the metalloregulatory system.

3.1 Transcriptional Regulation

Examples of transcriptional regulation by trace elements are found in all eukaryotic cells, but detailed descriptions of the cellular apparatus involved have emerged more rapidly from systems in lower eukaroytes. For example, in yeast, over 100 genes that influence trace element homeostasis and cellular functions, such as defense, have been shown to be regulated by 13 different metal-responsive transcription factors. Metal transporters and chaperones comprise the largest group of these metal-regulated genes.

Metallothionein has been the most widely studied mammalian gene that is transcriptionally regulated by trace elements. Mammalian metallothionein is tissue-specific, with high expression in liver, intestine, bone marrow, pancreas, and kidney, but expression can be detected in most tissues. Depletion of the animals' dietary zinc supply has a negative effect on expression, while reduced copper intake has little effect. Similarly, elevation of dietary zinc intake level provides a positive stimulus for expression, but changes in the copper intake level have little effect. This difference between copper and zinc is an example of how mammalian systems are not analogous to metal-produced responses with unicellular organisms. In the latter, copper is an important inducer of *metallothionein*. Studies with intact animals in which trace metals are injected can show regulation by multiple inducer trace metals. These responses could be a direct action of the trace element on a sensor protein or an indirect action via transcriptional regulation signaled by metal-stimulated release

of stress-related mediators or activation of oxidant-responsive factors. Therefore, an underlying question to answer in integrative systems is whether the trace metal is a direct or an indirect mediator of gene transcription. The body of evidence with many mammalian systems, including transfected cells, demonstrates that zinc and the toxic metal cadmium (not a nutrient) induce gene expression via transcription. The exact mechanism of how zinc enters mammalian cells, traverses the intracellular space, and enters the nucleus to initiate or repress transcription has been described in limited detail. Zinc homeostasis in mammals is controlled by genes regulating influx, efflux, and compartmentalization. Some appear to be nutritionally responsive. Some transporter proteins may have metal-sensing domains and provide a signal transduction function.

At least three components are considered necessary for transcriptional regulation by a trace element (Fig. 1). First, the responsive gene must have an MRE at one or more sites in regulatory regions. All mammalian DNA motifs responsible for metal regulation thus far identified have a consensus 13 to 15 nucleotide imperfect sequence. The consensus MRE sequence for the mammalian metallothionein promoter is CTCTGCRCNCGGCCC. The core of this consensus sequence is underlined. MREs are usually found in multiple copies of either orientation within the first few hundred bases of the promoter. They are most frequently upstream from the transcription initiation site. MREs are frequently located near other regulatory elements, perhaps for synergistic or differential regulatory purposes. A search of the genome reveals that a plethora of genes have MRE sequences within regulatory regions. The proportion of these that confer transcriptional activity is not known.

The second component needed for transcriptional regulation by trace element micronutrients is a trans-acting MRE-binding metalloregulatory protein(s) called a *metal-responsive transcription factor* (MTF). Such a metal-binding/sensing protein (depicted in Fig. 1) acts as a transcription factor to initiate or enhance transcription of an MRE-regulated gene. In mammalian cells, there are only two such factors, MTF-1 and MTF-2, currently identified. Only MTF-1 has been functionally characterized. Zinc binds to MTF-1 through unique (most likely tetrahedral) coordination chemistry. MTF-1 has six (C_2H_2-type) zinc-finger domains; three bind zinc ions more strongly than the other three. The more facile sites have the regulatory role and, upon occupancy, facilitate translocation to the nucleus and DNA binding to the MRE. This may constitute the mode by which zinc from the diet activates the MTF-1/MRE system. Evidence suggests MTF-1 DNA-binding activity is influenced by phosphorylation and the state of cellular redox. Shuttling of MTF-1 between cytoplasm and the nucleus could be a point of regulation. Similarly, the *MTF-1* gene is upregulated in zinc restriction, and the *MTF-2* gene is concomitantly downregulated. Multiple MTF proteins may yield complexes that produce negative as well as positive effects on transcription of a specific MRE-regulated gene. Furthermore, tissue-specificity of metal regulation requires that genes for the MTF protein are expressed to differing extents requisite to cell function and stage of development. Data are lacking on these points. Genome screening methods, including DNA microarray analysis, have shown that numerous genes are likely regulated by MTF-1. In mice, a null mutation of *MTF-1* causes death during fetal growth. In contrast, *MT* null mice

exhibit normal growth. This difference suggests MTF-1 regulates a gene(s) other than *MT* that is critical for the developing mouse embryo. Numerous zinc transporter genes are believed to be regulated by MTF and may be among these regulated genes.

The third component for this metal-responsive unit is the trace element (metal), that is, the inducer. Intracellular concentrations of zinc, while held relatively constant, vary sufficiently with physiologic state and dietary intake to influence metal availability for binding to the MTF protein. The source (a low molecular weight chelate or a chaperone protein, possibly of lower binding affinity) and the cellular location (nuclear, cytoplasmic, or both) of metal destined for binding to MTF proteins remain to be defined (Fig. 1). It is likely that the intracellular pools for these roles may constitute only a few atoms per cell.

3.2 Translational Regulation

The initial concept of trace element control of translation was developed from the acute induction of ferritin synthesis by iron, a process that does not alter transcription of the *ferritin* gene. Iron circulates in the plasma, bound to transferrin (Tf). This protein binds to transferrin receptor (TfR) located on the cell surface, undergoes endocytosis, and contributes iron to a low molecular weight iron pool. When this cellular pool is large, ferritin synthesis is stimulated via increases in translatable ferritin mRNA. In contrast, when the cellular iron pool is low, for example, during dietary iron deficiency, translation of TfR mRNA increases (Fig. 2).

The translational regulation of TfR and ferritin mRNAs and those of other iron-regulated proteins is controlled by two cytoplasmic iron regulatory proteins (IRP-1 and IRP-2). A unique secondary structure of specific mRNAs are nucleotide loops that constitute iron-responsive elements (IRE). IREs for different mRNAs are not identical. Ferritin, TfR, the ferroportin transporter, and erythroid δ-aminolevulinate synthase are among the proteins of iron metabolism whose mRNAs have IREs. RNA-binding activity of IRP molecules is strongly regulated by dietary iron. IRP high-affinity binding to TfR mRNA causes stabilization at the A/U rich region of the 3′ end of the mRNA by inhibiting degradation and stimulating its translation. This promotes cellular iron uptake when the iron supply is low. IRP-1 may bind and thus stabilize the mRNA for the iron transporter DMT-1. This would enhance synthesis of the transporter for increased apical iron transport by enterocytes when the dietary iron supply is low. Conversely, with an increased iron supply, IRP binding to ferritin and ferroportin mRNAs is reduced, and their translation increases. This facilitates iron storage. Of interest is that slight structural differences in the 3′ end of ferroportin mRNA may cause low iron to enhance translation in the enterocytes but repress translation in liver. Iron binding is correlated with the c-aconitase activity of IRP-1. This mitochondrial TCA cycle enzyme may alter cellular citrate levels and, through iron-binding properties of citrate, influence the low molecular weight iron pool. Mitochondrial succinate dehydrogenase iron protein mRNA is also regulated by IRP-1. Iron metabolism makes further use of the IRE-IRP system for control of heme synthesis in erythroid cells via regulation of δ-aminolevulinate synthase mRNA translation.

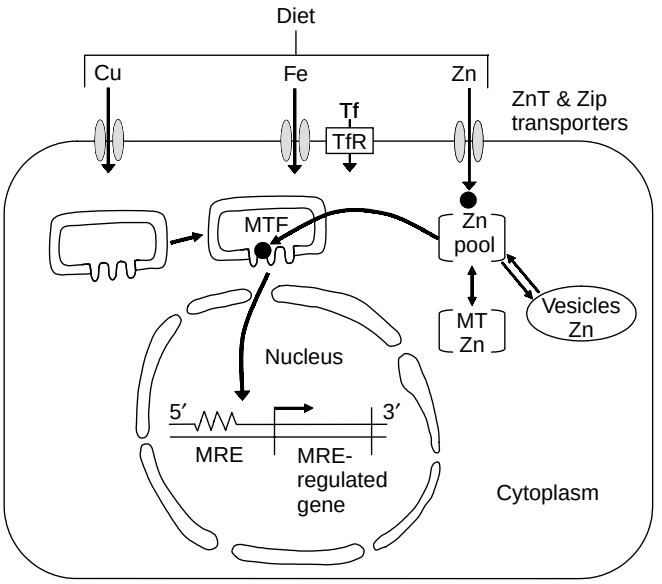

Fig. 1 Transcriptional regulation of gene expression by a trace element (metal) micronutrient. The diet provides copper, iron, and zinc for distribution to cells via the plasma. Of the three nutrients, only zinc provides a demonstrated stimulus for induction of transcription. Metallothionein gene expression by Zn provides an example of this mode of regulation. A requirement of this system is that the regulated gene(s) has a metal response element (MRE) sequence(s) in its promoter/enhancer region. The metal (●), following transport into the cell via ZnT and Zip transporters, interacts rapidly with a variety of ligands and/or is transported into vesicles, thus maintaining an extremely small pool. One of these ligands is an MRE-binding protein termed a *metal-responsive transcription factor* (MTF). Upon metal occupancy of specific zinc fingers, the MTF enters the nucleus for MRE binding. Zinc may influence the MTF translocation process and/or interaction with specific inhibitors. Phosphorylation and the cell redox state may influence these processes also. The rate of transcription is directly proportional to the dietary metal supply, the intracellular metal pool, and the extent of MTF/MRE binding. It is possible that some MRE-binding transcription factors act as negative rather than positive regulators of transcription.

4
Applications and Perspectives

The application of trace element micronutrients for control of biotechnological systems has been demonstrated. The dramatic growth of transgenic animals producing excess growth hormone through activation of chimeric metallothionein promoter-growth hormone constructs in response to zinc in the drinking water illustrated the potential of this approach two decades ago.

Chimeric constructs have been used to regulate plant genes by trace elements. The copper-responsive transcription factor

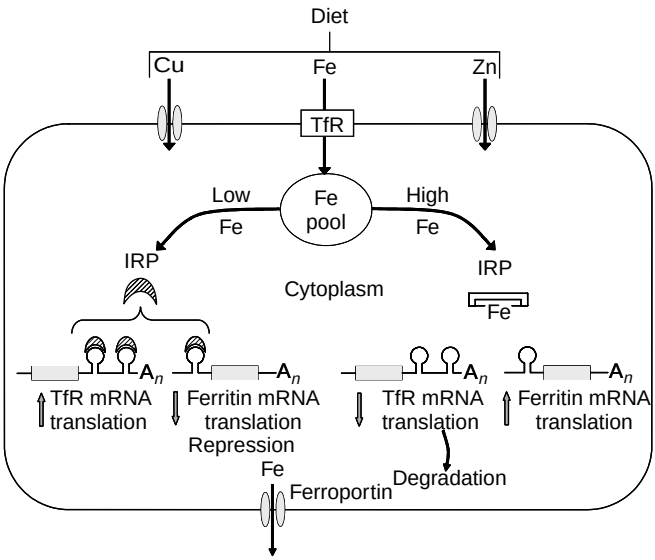

Fig. 2 Translational regulation of mRNA by a trace element (metal) micronutrient. The diet provides copper, iron, and zinc for distribution to cells via the plasma. Of the three nutrients, only iron provides a demonstrated stimulus for control of mRNA translation. Transferrin receptor mRNA (TfR mRNA) and ferritin mRNA are among the small number of mRNAs regulated by iron. The intracellular iron pool is reflective of dietary iron availability. Cellular iron is monitored by iron regulatory proteins (IRP). One IRP (IRP-1) has c-aconitase activity, which may help regulate the low molecular weight iron pool size via citrate availability. When cellular iron is low, less iron is bound to IRPs. This favors their binding to specific mRNAs via unique stem loops in the UTR of the mRNA (5′ and 3′ of the coding region for ferritin and TfR mRNAs, respectively). Binding protects TfR mRNA from degradation and increases its translation, thus transferrin receptor synthesis increases. Consequently, more iron enters the cell. Simultaneously, low iron also favors IRP binding to ferritin mRNA, which decreases translation of this mRNA. When cellular iron levels are high, iron binding to IRPs increases. This lowers the mRNA-binding affinity of IRPs, thus increasing ferritin mRNA translation and TfR mRNA degradation. As a result, ferritin synthesis increases, and TfR synthesis decreases. Erythroid δ-aminolevulinate synthase mRNA and succinate dehydrogenase mRNA translation decrease upon IRP binding with low iron in a fashion like mRNAs for H/L ferritin. Ferroportin mRNA translation in intestine is stimulated by low iron (like TfR mRNA) but, in liver, ferroportin mRNA translation is decreased. Since ferroportin is an iron export protein, this apparent difference stimulates intestinal absorption and hepatic retention.

from yeast (ACE1) was used to stimulate expression of a gene coding for a commercially important enzyme, via copper treatment of the plant. A DNA construct with the ACE1 binding site (MRE) was inserted into the promoter of the gene coding for the enzyme. Through copper treatment of the transgenic plant, production of the enzyme is increased.

A key to using trace elements for technological purposes is that those chosen must be nontoxic, for example, zinc, so as not to damage the organism but sufficiently bioavailable to promote robust expression of the transgene. The unique physiology, tissue-specific sequestration, and nutritional intake of each trace element would need to be considered in these strategies. A concern is that the metal-resistance genes of coexisting microbial systems could lead to overgrowth of endogenous microbial populations and induction of metal-inducible proteins, which may provide gratuitous changes in metabolic fates of xenobiotics. Similarly, efforts to improve micronutrient (iron and zinc) nutrition in the developing world through plants engineered to sequester more of these micronutrients could lead to overconsumption in humans and dysregulation of genes regulated by these nutrients.

The development of pharmaceuticals with actions related to metal chelation or metal donation is of major current interest. Many of these novel compounds will use trace element micronutrients as signaling systems for altering gene expression. Alternatively, a naturally occurring cellular component, for example, a zinc-finger transcription factor, could be the target of a specific trace element–containing therapeutic or sequestering agent. Similarly, engineering zinc-finger transcription factor delivered by viral vectors (gene therapy) to atypically regulate specific target genes by trace elements is possible.

See also Adipocytes.

Bibliography

Books and Reviews

Davis, S.R., Cousins, R.J. (2000) Metallothionein expression in animals: a physiological perspective on function, *J. Nutr.* **130**, 1085–1088.

Eisenstein, R.S. (2000) Iron regulatory proteins and the molecular control of mammalian iron metabolism, *Annu. Rev. Nutr.* **20**, 627–662.

Hambidge, M., Cousins, R.J., Costello, R.B. (Eds.) (2000) Proceedings of the international workshop on zinc and health: current status and future directions, *J. Nutr.* **130**, 1341S–1520S.

Hellman, N.E., Gitlin, J.D. (2002) Ceruloplasmin metabolism and function, *Annu. Rev. Nutr.* **22**, 439–458.

Institute of Medicine. (2001) Zinc, in: *Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc*, National Academy Press, Washington, DC, pp. 442–501.

King, J.C. (Ed.) (2003) 11th international symposium on trace elements in man and animals, *J. Nutr.* **133**, 1429S–1587S.

Liuzzi, J.P., Cousins, R.J. (2004) Mammalian zinc transporters, *Annu. Rev. Nutr.* **24**, 151–172.

Lutsenko, S., Petris, M.J. (2003) Function and regulation of the mammalian copper-transporting ATPases: insights from biochemical and cell biological approaches, *J. Membr. Biol.* **191**, 1–12.

Miret, S., Simpson, R.J., McKie, A.T. (2003) Physiology and molecular biology of dietary iron absorption, *Annu. Rev. Nutr.* **23**, 283–301.

Rutherford, J.C., Bird, A.J. (2004) Metal-responsive transcription factors that regulate

iron, zinc, and copper homeostasis in eukaryotic cells, *Eukaryot. Cell* **3**, 1–13.

Primary Literature

Beck, M.A., Levander, O.A., Handy, J. (2003) Selenium deficiency and viral infection, *J. Nutr.* **133**, 1463S–1467S.

Beguin, Y. (1992) The soluble transferrin receptor: biological aspects and clinical usefulness as quantitative measure of erythropoiesis, *Haematologica* **77**, 1–10.

Blalock, T.L., Dunn, M.A., Cousins, R.J. (1988) Metallothionein gene expression in rats: tissue-specific regulation by dietary copper and zinc, *J. Nutr.* **118**, 222–228.

Blanchard, R.K., Cousins, R.J. (1996) Differential display of intestinal mRNAs regulated by dietary zinc, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6863–6868.

Bremner, I., Morrison, J.N., Wood, A.M., Arthur, J.R. (1987) Effects of changes in dietary zinc, copper and selenium supply and of endotoxin administration on metallothionein I concentrations in blood cells and urine in the rat, *J. Nutr.* **117**, 1595–1602.

Bulteau, A.L., O'Neill, H.A., Kennedy, M.C., Ikeda-Saito, M., Isaya, G., Szweda, L.I. (2004) Frataxin acts as an iron chaperone protein to modulate mitochondrial aconitase activity, *Science* **305**, 242–245.

Burk, R.F., Hill, K.E., Motley, A.K. (2003) Selenoprotein metabolism and function: evidence for more than one function for selenoprotein P, *J. Nutr.* **133**, 1517S–1520S.

Chan, R.Y., Schulman, H.M., Ponka, P. (1993) Expression of ferrochelatase mRNA in erythroid and non-erythroid cells, *Biochem. J.* **292**, 343–349.

Cousins, R.J. (1985) Absorption, transport, and hepatic metabolism of copper and zinc: special reference to metallothionein and ceruloplasmin, *Physiol. Rev.* **65**, 238–309.

Cousins, R.J., Blanchard, R.K., Popp, M.P., Liu, L., Cao, J., Moore, J.B., Green, C.L. (2003) A global view of the selectivity of zinc deprivation and excess on genes expressed in human THP-1 mononuclear cells, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6952–6957.

Cousins, R.J., Leinart, A.S. (1988) Tissue-specific regulation of zinc metabolism and metallothionein genes by interleukin 1, *FASEB J.* **2**, 2884–2890.

Eisenstein, R.S., Garcia-Mayol, D., Pettingell, W., Munro, H.N. (1991) Regulation of ferritin and heme oxygenase synthesis in rat fibroblasts by different forms of iron, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 688–692.

Failla, M.L., Cousins, R.J. (1978) Zinc accumulation and metabolism in primary cultures of adult rat liver cells. Regulation by glucocorticoids, *Biochim. Biophys. Acta* **543**, 293–304.

Fraker, P.J., DePasquale-Jardieu, P., Zwickl, C.M., Luecke, R.W. (1978) Regeneration of T-cell helper function in zinc-deficient adult mice, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5660–5664.

Granick, S. (1946) Ferritin. IX. Increase of the protein apoferritin in the gastrointestinal mucosa as a direct response to iron feeding. The function of ferritin in the regulation of iron absorption, *J. Biol. Chem.* **164**, 737–746.

Grider, A., Bailey, L.B., Cousins, R.J. (1990) Erythrocyte metallothionein as an index of zinc status in humans, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1259–1262.

Hamza, I., Faisst, A., Prohaska, J., Chen, J., Gruss, P., Gitlin, J.D. (2001) The metallochaperone Atox1 plays a critical role in perinatal copper homeostasis, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6848–6852.

Hamza, I., Prohaska, J., Gitlin, J.D. (2003) Essential role for Atox1 in the copper-mediated intracellular trafficking of the Menkes ATPase, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1215–1220.

Harris, E.D. (1976) Copper-induced activation of aortic lysyl oxidase in vivo, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 371–374.

Harris, Z.L., Durley, A.P., Man, T.K., Gitlin, J.D. (1999) Targeted gene disruption reveals an essential role for ceruloplasmin in cellular iron efflux, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10812–10817.

Hershko, C., Cook, J.D., Finch, C.A. (1974) Storage iron kinetics. VI. The effect of inflammation on iron exchange in the rat, *Br. J. Haematol.* **28**, 67–75.

Huang, L., Gitschier, J. (1997) A novel gene involved in zinc transport is deficient in the lethal milk mouse, *Nat. Genet.* **17**, 292–297.

Kennedy, M.C., Mende-Mueller, L., Blondin, G.A., Beinert, H. (1992) Purification and characterization of cytosolic aconitase from beef liver and its relationship to the iron-responsive element binding protein, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 11730–11734.

Klotz, L.-O., Kröncke, K.-D., Buchczyk, D.P., Sies, H. (2003) Role of copper, zinc, selenium and tellurium in the cellular defense against oxidative and nitrosative stress, *J. Nutr.* **133**, 1448S–1451S.

Klug, M., Rhodes, D. (1987) Zinc fingers: a novel protein motif for nucleic acid recognition, *Trends Biochem. Sci.* **12**, 464–469.

Knutson, M., Wessling-Resnick, M. (2003) Iron metabolism in the reticuloendothelial system, *Crit. Rev. Biochem. Mol. Biol.* **38**, 61–88.

Kuo, Y.M., Zhou, B., Cosco, D., Gitschier, J. (2001) The copper transporter CTR1 provides an essential function in mammalian embryonic development, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6836–6841.

Lee, J., Prohaska, J.R., Thiele, D.J. (2001) Essential role for mammalian copper transporter Ctr1 in copper homeostasis and embryonic development, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6842–6847.

Lönnerdal, B. (2003) Genetically modified plants for improved trace element nutrition, *J. Nutr.* **133**, 1490S–1493S.

Lyons, T.J., Liu, H., Goto, J.J., Nersissian, A., Roe, J.A., Graden, J.A., Café, C., Ellerby, L.M., Bredesen, D.E., Gralla, E.B., Valentine, J.S. (1996) Mutations in copper-zinc superoxide dismutase that cause amyotrophic lateral sclerosis alter the zinc binding site and the redox behavior of the protein, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12240–12244.

Maret, W. (2004) Zinc and sulfur: a critical biological partnership, *Biochemistry* **43**, 3301–3309.

Mason, K.E. (1979) A conspectus of research on copper metabolism and requirements of man, *J. Nutr.* **109**, 1979–2066.

Mercer, J.F.B., Llanos, R.M. (2003) Molecular and cellular aspects of copper transport in developing mammals, *J. Nutr.* **133**, 1481S–1484S.

Michalska, A.E., Choo, K.H. (1993) Targeting and germ-line transmission of a null mutation at the metallothionein I and II loci in mouse, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8088–8092.

Nemeth, E., Rivera, S., Gabayan, V., Keller, C., Taudorf, S., Pedersen, B.K., Ganz, T. (2004) IL-6 mediates hypoferremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin, *J. Clin. Invest.* **113**, 1271–1276.

Owen, C.A. Jr. (1964) Distribution of copper in the rat, *Am. J. Physiol.* **207**, 446–448.

Osaki, S., Johnson, D.A. (1969) Mobilization of liver iron by ferroxidase (ceruloplasmin), *J. Biol. Chem.* **244**, 5757–5758.

Palmiter, R.D., Findley, S.D. (1995) Cloning and functional characterization of a mammalian zinc transporter that confers resistance to zinc, *EMBO J.* **14**, 639–649.

Pekarek, R.S., Burghen, G.A., Bartelloni, P.J., Calia, F.M., Bostian, K.A., Beisel, W.R. (1970) The effect of live attenuated Venezuelan equine encephalomyelitis virus vaccine on serum iron, zinc, and copper concentrations in man, *J. Lab. Clin. Med.* **76**, 293–303.

Pietrangelo, A. (2004) Hereditary hemochromatosis-a new look at an old disease, *N. Engl. J. Med.* **350**, 2383–2397.

Prasad, A.S. (1982) Clinical and Biochemical Spectrum of Zinc Deficiency in Human Subjects, in: Prasad, A.S. (Ed.) *Clinical, Biochemical and Nutritional Aspects of Trace Elements*, Alan R. Liss, New York, pp. 3–62.

Requena, J.R., Groth, D., Legname, G., Stadtman, E.R., Prusiner, S.B., Levine, R.L. (2001) Copper-catalyzed oxidation of the recombinant SHa(29–231) prion protein, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7170–7175.

Richards, M.P., Cousins, R.J. (1975) Mammalian zinc homeostasis: requirement for RNA and metallothionein synthesis, *Biochem. Biophys. Res. Commun.* **64**, 1215–1223.

Roesijadi, G., Bogumil, R., Vasak, M., Kagi, J.H. (1998) Modulation of DNA binding of a tramtrack zinc finger peptide by the metallothionein-thionein conjugate pair, *J. Biol. Chem.* **273**, 17425–17432.

Schroeder, J.J., Cousins, R.J. (1990) Interleukin 6 regulates metallothionein gene expression and zinc metabolism in hepatocyte monolayer cultures, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3137–3141.

Shim, H., Harris, Z.L. (2003) Genetic defects in copper metabolism, *J. Nutr.* **133**, 1527S–1531S.

Snyder, S.L., Walker, R.I. (1976) Inhibition of lethality in endotoxin-challenged mice treated with zinc chloride, *Infect. Immun.* **13**, 998–1000.

Squibb, K.S., Cousins, R.J., Feldman, S.L. (1977) Control of zinc-thionein synthesis in rat liver, *Biochem. J.* **164**, 223–228.

Suzuki, K.T., Imura, N., Kimura, M. (Eds.) (1993) *Metallothionein III: Biological Roles and Medical Implications*, Birkhäuser Verlag, Basel, Switzerland.

Tang, C.K., Chin, J., Harford, J.B., Klausner, R.D., Rouault, T.A. (1992) Iron regulates the activity of the iron-responsive element binding protein without changing its rate of synthesis or degradation, *J. Biol. Chem.* **267**, 24466–24470.

Valentine, J.S., Hart, P.J. (2003) Misfolded CuZnSOD and amyotrophic lateral sclerosis, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3617–3622.

Vallee, B.L., Galdes, A. (1984) The metallobiochemistry of zinc enzymes, *Adv. Enzymol. Relat. Areas Mol. Biol.* **56**, 283–430.

Vallee, B.L., Williams, R.J. (1968) Metalloenzymes: the entatic nature of their active sites, *Proc. Natl. Acad. Sci. U.S.A.* **59**, 498–505.

Vulpe, C., Levinson, B., Whitney, S., Packman, S., Gitschier, J. (1993) Isolation of a candidate gene for Menkes disease and evidence that it encodes a copper-transporting ATPase, *Nat. Genet.* **3**, 7–13.

Wang, K., Zhou, B., Kuo, Y.M., Zemansky, J., Gitschier, J. (2002) A novel member of a zinc transporter family is defective in acrodermatitis enteropathica, *Am. J. Hum. Genet.* **71**, 66–73.

Wang, Y., Wimmer, U., Lichtlen, P., Inderbitzin, D., Stieger, B., Meier, P.J., Hunziker, L., Stallmach, T., Forrer, R., Rulicke, T., Georgiev, O., Schaffner, W. (2004) Metal-responsive transcription factor-1 (MTF-1) is essential for embryonic liver development and heavy metal detoxification in the adult liver, *FASEB J.* **18**, 1071–1079.

Weinberg, E.D. (1984) Iron withholding: a defense against infection and neoplasia, *Physiol. Rev.* **64**, 65–102.

Zahringer, J., Baliga, B.S., Munro, H.N. (1976) Novel mechanism for translational control in regulation of ferritin synthesis by iron, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 857–861.

Mitochondrial Porins, Eukaryotic

Roland Benz
Biocenter of the University of Würzburg, Würzburg, Germany

1 Introduction 359

2 Reconstitution of Mitochondrial Porins (VDACs) in Model Membranes 361
2.1 Isolation and Purification of Mitochondrial Porins 361
2.2 Heterologous Expression of Eukaryotic Porins 362
2.3 Reconstitution Methods 362

3 Characterization of the Pore-forming Properties of Eukaryotic Porins 363
3.1 Single-channel Analysis of the Mitochondrial Pore in the Open State 363
3.2 Eukaryotic Porins are Voltage-gated 364
3.3 Single-channel Conductance of the Closed States 368
3.4 Selectivity of the Open and Closed States of Eukaryotic Porins 369

4 Inhibition of the Mitochondrial Pore 370

5 Structure of the Channel Formed by Eukaryotic Porins 372
5.1 Primary Structure of Eukaryotic Porins 372
5.2 Secondary Structure of Mitochondrial Porins 374
5.3 Structure of the Channel-forming Unit 375
5.4 Are Sterols Involved in the Formation of the Channel-forming Unit? 377
5.5 Electron Microscopic Analysis of Mitochondrial Porin 378

6 Conclusions 378

Acknowledgments 379

Bibliography 380
Books and Reviews 380
Primary Literature 380

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

Keywords

β-Barrel Cylinder
Membrane-spanning hollow cylinder formed by antiparallel amphipathic β-strands, typical in the structure of bacterial and mitochondrial outer membrane channels formed by porins.

Mitochondria
Cell organelle responsible for oxidative phosphorylation. Mitochondria are probably descendants of strictly aerobic gram-negative ancestors.

Mitochondrial (Eukaryotic Porin)
Channel-forming protein responsible for exchange of mitochondrial metabolites across the mitochondrial outer membrane.

Mitochondrial Inner Membrane
Surrounds the mitochondrial matrix space and contains the enzymes of the respiration chain and the H^+-ATPase. The mitochondrial inner membrane has many invaginations (cristae) to increase its surface and contains many carrier proteins responsible for specific substrate transport.

Mitochondrial Outer Membrane
Outer boundary membrane of mitochondria, forms with the mitochondrial inner membrane the intermembrane space that contains many different enzymes important for mitochondrial metabolism such as the mitochondrial creatine kinase.

Peripheral Kinases
Enzymes such as hexokinase and glycerokinase that bind to porin on the surface of mitochondria and utilize mitochondrial ATP.

Porin Family
Most genomes of eukaryotic cells contain more than one gene coding for mitochondrial (eukaryotic) porins. The porins form a family of highly homologous proteins that are posttranslationally imported into mitochondria.

VDAC
Voltage-dependent anion-selective channel, synonym for mitochondrial porin.

The mitochondrial outer membrane contains a channel that is responsible for the passage of hydrophilic metabolites across the membrane. The channel-forming protein known for many organisms, called *mitochondrial porin* or VDAC (voltage-dependent anion-selective channel), has a length of about 280 amino acids. The genomes of eukaryotic organisms contain several porin isoforms of not well-defined function. The primary structure of mitochondrial or eukaryotic porins is not particularly hydrophobic and secondary structure predictions suggest that a β-barrel cylinder typical for the secondary structure of bacterial porins also forms the mitochondrial channel. Kinases involved in mitochondrial metabolism such as hexokinase or glycerokinase bind to porin and play an important role in compartment formation in mitochondria. Mitochondrial porins are voltage-gated and switch into ion-permeable substates at voltages higher than 20 to 30 mV. The open channel has a small preference for anion over cations of the same aqueous mobility. The closed states are cation selective and impermeable for ATP and ADP. Mitochondrial porins seem to be involved in apoptosis and the release of cytochrome *c*. There exists emerging evidence that the cytoplasmic membrane of eukaryotic cells also contains porins of the eukaryotic porin family with a role in cellular metabolism that is not yet understood.

1 Introduction

The matrix space of mitochondria is, similar to the situation in gram-negative bacteria, surrounded by two membranes. The mitochondrial inner membrane is especially rich in proteins. Important for mitochondrial function are the enzymes of the respiration chain and oxidative phosphorylation and the many highly specific carriers for the substrates of mitochondria. In contrast to this, the mitochondrial outer membrane is far less substrate specific and acts as a molecular filter for hydrophilic solutes since it is freely permeable to a variety of small hydrophilic solutes up to a well-defined exclusion limit of about 3000 to 5000 Da. Molecules with higher molecular masses are retained. The well-defined exclusion limit indicates that the mitochondrial outer membranes contain defined passages for substrates. A considerable part of its permeability properties is caused by the presence of general diffusion pores, called *mitochondrial porins* or voltage-dependent anion-selective channels (VDACs). As early as 1976, a channel-forming component was discovered in mitochondria, but not in other membrane fractions, of *Paramecium aurelia*. Because of the voltage dependence and the anion selectivity of the channel, the name VDAC was used. It is obvious that the anion selectivity of mitochondrial porin was overestimated at the beginning because the selectivity ratio of the channel was estimated to be more than 7 for chloride over potassium ions, while more recent data suggest that it is close to 2 at neutral pH. Mitochondrial porins were isolated from a variety of mitochondria from different organisms and tissues, and there exists emerging evidence that similar proteins are also present in cytoplasmic membranes of

eukaryotic cells with yet unknown function. They form a family of slightly basic proteins with molecular masses around 30 kDa. The genes of many mitochondrial porins are known to date. The proteins are encoded in the nucleus without a leader sequence and are synthesized on cytoplasmic ribosomes. They are targeted to mitochondria as chaperone-bound species, recognized by the outer membrane general import complex (TOM) and then inserted into the outer membrane probably from the intermembrane space, similar to the situation for porins of the putative gram-negative ancestor of mitochondria.

The eukaryotic genomes may contain several genes and pseudogenes coding for porins. Mitochondria contain mostly isoform 1 (VDAC1, porin 31HL), but the role of the other one to two isoforms in eukaryotic cells was also studied. Experiments with water-soluble porin from *Neurospora crassa* and the composition of isolated mammalian porin suggest that the channel-forming unit is a complex composed of probably only one polypeptide chain and sterols. The amino acid composition exhibits a high polarity similar to bacterial porins, although the channel-forming unit is a membrane protein that contains many membrane-spanning segments. The exact structure of the pore is still a matter of debate because it has not been crystallized to date. Secondary structure predictions, electron microscopic analyzes of two-dimensional arrays and circular dichroism (CD) spectra of mitochondrial porins suggest that they contain large amounts of β-sheet structure with no indication of hydrophobic transmembrane α-helical regions, which makes it likely that bacterial and mitochondrial porins share common structural motifs. So far it is not clear if this has to do with the evolution of mitochondria or if it simply reflects that both membrane channels have water-soluble precursors.

Reconstitution experiments with planar lipid bilayers and liposomes define mitochondrial porins as pore-forming components, which have a high permeability for both ionic and neutral substrates. According to these studies, the mitochondrial outer membrane pore has a diameter of about 2 to 3 nm in the open state, which agrees well with the electron microscopic analysis of mitochondrial porin *in vitro* (i.e. in reconstituted systems) and *in vivo* (i.e. in the outer membrane The open channel appears to be wide and water filled. It is slightly anion selective for salts composed of cations and anions of the same aqueous mobility such as potassium and chloride, but may change selectivity when the anion has a smaller aqueous mobility than the cation. For membrane voltages larger than about 30 mV, the channel closes in distinct steps in several closed states. These closed states have a reduced permeability toward hydrophilic solutes and a completely different selectivity than the open state. In addition to increasing membrane potential, polyanions such as Konig's polyanion (a copolymer of molecular mass 10 kDa from methacrylate, maleate, and styrene in a 1 : 2 : 3 proportion) or dextran sulfate result in channel closure *in vivo* and *in vitro*. In particular, adenosine triphosphate (ATP) and adenosine diphosphate (ADP) are not permeable through the closed states of mitochondrial porin.

The role of the mitochondrial outer membrane in the physiology and metabolism of mitochondria was underestimated for a long time. It was described as a simple barrier that only prevented

the destruction of the mitochondrial inner membrane during osmotic swelling. More recent papers gave some insight into the function of the mitochondrial outer membrane and the role of the porin in mitochondrial metabolism. Porin is the hexokinase-binding protein and glycerokinase also binds to this protein. Cross-linking studies of the mitochondrial outer membrane of yeast revealed that the pore associates or interacts with a 14-kDa protein that was identified as glutathione transferase. Subfractionation of the mitochondrial membranes by sucrose density centrifugation allowed the isolation of contact sites between inner and outer membranes. These contact sites are especially enriched in hexokinase-binding capacity. A further interesting aspect of mitochondrial porin is its possible role in the mitochondrial permeability transition pore (PTP) and its part in cell signaling. Furthermore, there exists a strong indication that porin is involved in apoptosis because of its interaction with proapoptotic and antiapoptotic factors such as Bax and Bcl-2. Furthermore, it shows negative regulation by C-Raf kinase. This may indicate that mitochondrial porin pore plays a crucial role in the interaction of mitochondria with cellular components.

2 Reconstitution of Mitochondrial Porins (VDACs) in Model Membranes

2.1 Isolation and Purification of Mitochondrial Porins

Several different methods are well established for the successful isolation and purification of mitochondrial porins. All start from mitochondria that are obtained by differential centrifugation or density gradient centrifugation of homogenized cellular material. It has to be noted, however, that both preparations also contain other cellular material. For preparation of the mitochondrial outer membrane, the mitochondria are exposed to a swelling/shrinking procedure. Mitochondria are first exposed to 10 mM phosphate followed by addition of 0.45 M sucrose. Centrifugation at 10 000 *g* yields the mitoplasts in the pellet that also contain most of the contact sites between the inner and outer membrane. The outer membrane is in the supernatant. Porin can be isolated from the outer membrane following a standard protocol by using neutral detergents but not by ionic detergents because they destroy its channel-forming activity. A higher yield of porin may be obtained following an alternative method starting from whole mitochondria. The cell organelles are extracted with neutral detergents such as Genapol X-80, *N*,*N*-dimethyldodecylamine-*N*-oxide (LDAO) or Triton X-100. The extract is passed through a dry hydroxyapatite (HTP) column where most of the proteins bind. Porin appears just after the void volume of the column because it is deeply buried in lipid-detergent micelles and does not interact with the column material when Triton X-100 or Genapol X-80 are employed. The use of LDAO instead of Genapol or Triton obviously leads to smaller micelles that bind to HTP and allow a normal elution protocol. The use of ionic detergents such as cholate, desoxycholate, and sodium dodecylsulfate (SDS) during the isolation and purification procedure seems to destroy the structure of the channel-forming complex, and leads to its inactivity in reconstitution experiments.

2.2 Heterologous Expression of Eukaryotic Porins

It is possible to create a water-soluble form of mitochondrial porin from *N. crassa* and other organisms. Water-soluble porin is inactive in reconstitution experiments and does not form ion-permeable channels. However, when the porin is treated with Triton or Genapol in the presence of sterols, channel formation is regained with properties that cannot be distinguished from its detergent-solubilized form. It is possible that the detergents act as some kind of chaperone. Careful analyzes of the sterol requirements suggested that cholesterol leads to highest activation, whereas the addition of epicholesterol did not influence the low channel-forming ability of water-soluble porin. The possibility to refold mitochondrial porin allows an interesting insight into the structure–function relationship of porin mutants because the genes of many of the mitochondrial porins are known to date. For heterologous expression of mitochondrial porins in *Escherichia coli*, the porin-coding sequences have to be cloned in an expression vector that contains, for easy purification of the expressed protein, a DNA sequence coding for additional six histidines (His-tag) upstream of the polycloning site. This His-tag allows the purification of the overexpressed protein by Ni-NTA affinity chromatography. The expressed protein is found within inclusion bodies and needs to be solved in 1% SDS or a high concentration of urea. After chromatography across a Ni-NTA column in SDS or urea, the protein is renatured by an overnight dialysis against a 1% Genapol or Triton X-100 solution and by the addition of a cholesterol suspension in the same detergent.

2.3 Reconstitution Methods

The properties of the channels formed by mitochondrial porins (VDACs) were studied in different model membranes. Mitochondrial porins from mung bean and rat liver were investigated in lipid vesicles. The reconstitution of rat liver porin into the vesicles made them permeable for [^{14}C]sucrose, but not for [^{3}H]dextran, which suggested that a specific pathway and not a leak is introduced into the vesicles. The exclusion limit of the channel from the outer membrane of mung bean mitochondria was measured in similar experiments using radioactively labeled solutes of different molecular mass to be between 1 and 10 kDa, probably between 3 and 6 kDa. An alternative method was used for the study of porin from *N. crassa* mitochondria. Porin-containing liposomes were transferred into a hypertonic solution containing a test solute. First, the liposomes shrink and can only swell when the test solute is permeable through the channels. The results of experiments using carbohydrates and polyethylene glycols (PEG) of different molecular masses suggest that PEGs with molecular masses up to about 6.8 kDa are permeable through the mitochondrial pore. From these results, the diameter of the mitochondrial pore was calculated to be around 3 to 4 nm. It is noteworthy that this type of vesicle-swelling assay is not as precise as the swelling method used previously for bacterial porins. The reason for this is that the method used for the study of the channel diameter of bacterial porins is based on the relative rate of permeation of many different solutes through the channel, which depends on the channel diameter according to the Renkin equation.

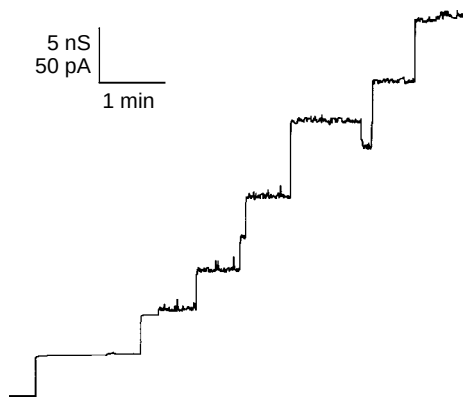

Fig. 1 Stepwise increase of the membrane current (given in pA) after the addition of *N. crassa* porin to a black lipid bilayer membrane given as a function of time. The aqueous phase contained 10 ng mL^{-1} protein and 1 M KCl. The membrane was formed from diphytanoyl phosphatidylcholine/*n*-decane. The voltage applied was 10 mV; $T = 25\,^{\circ}C$.

Many reconstitution studies with mitochondrial porins were performed using the planar lipid bilayer technique. The reconstitution is performed as follows. Purified porin is added at a low concentration (10 ng mL^{-1} to 1 μg mL^{-1}) to the aqueous phase bathing black lipid bilayer membranes formed according to the two different methods, which yield more or less similar results. After the addition of the porins to a preformed membrane, the membrane current starts increasing in a stepwise fashion after an initial lag of about 2 min (see Fig. 1 for the reconstitution of mitochondrial porin from *N. crassa*). This process indicates the insertion of ion-permeable channels into the membrane. The number of channels formed is dependent on time and the concentration of protein. In general, the membrane conductance increased for about 15 to 20 min. After that time, the membrane conductance (i.e. the current per unit voltage) increased at a much slower rate. When the rate of conductance increase was relatively slow (as compared with the initial one), it was shown for different mitochondrial porins that the membrane conductance was a linear function of the protein concentration up to porin concentrations of about 1 μg mL^{-1}. The results are consistent with the assumption that the protein samples contain preformed channels.

3 Characterization of the Pore-forming Properties of Eukaryotic Porins

3.1 Single-channel Analysis of the Mitochondrial Pore in the Open State

Mitochondrial porins from different eukaryotic cells form ion-permeable channels in lipid bilayer membranes – each step in Fig. 1 reflects the insertion of one conductive unit (i.e. of one channel into the membrane). These conductance steps are caused by the reconstitution of channels since they are not observed when only the detergents Triton X-100, Genapol X-80 or LDAO are added to the aqueous phase. Figure 1 demonstrates that most of the conductance steps are directed upwards and closing steps are only rarely observed at small transmembrane potentials of about 10 to 20 mV. The most frequent value for the single-channel conductance of mitochondrial porins in 1 M KCl (the conditions of Fig. 1) was about 4 to 5 nS (see Fig. 2 for a histogram of conductance

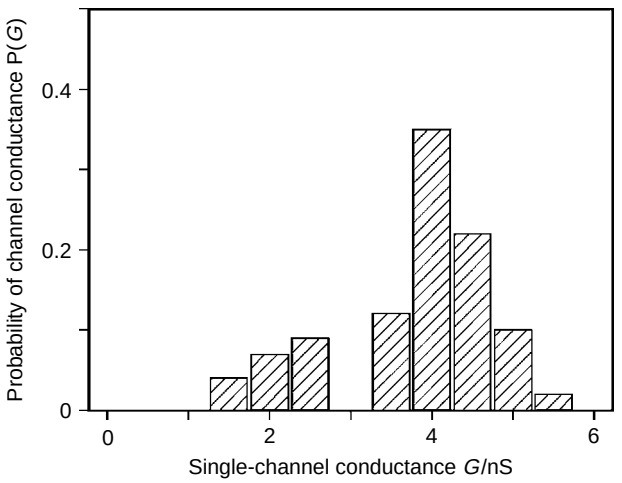

Fig. 2 Histogram of conductance fluctuations observed with membranes of diphytanoyl phosphatidylcholine/*n*-decane in the presence of *N. crassa* porin. *P*(*G*) is the probability for the occurrence of a conductance step with a certain single-channel conductance (given in nS). The aqueous phase contained 1 M KCl. The voltage applied was 10 mV. The mean value of all upward directed steps was 4.2 nS for the right-side maximum and 2.5 nS for the left-side maximum (in total 187 single events); $T = 25\,^{\circ}C$.

fluctuations observed with *N. crassa* porin). Only a limited number of smaller steps were observed under these conditions (about 25% of the total number of conductance steps). It should be noted that smaller steps are also found for other mitochondrial porins, including those from yeast, rat liver, and a number of cells from other tissues including mammalian cells. They may be explained as substates of the pore with the exception of porin from *Paramecium* mitochondria, where only small pores with a single-channel conductance around 2.5 nS were observed in 1 M KCl. This result may indicate that the mitochondrial outer membrane pore does not only switch to substates but that it may also exist in two different stable conformations. The channels formed by eukaryotic porins are wide and water filled. This is the result of measurements with salts composed of different anions and cations. Even large organic anions and cations such as $Tris^{+}$ and $HEPES^{-}$, were found to be permeable through the open state of mitochondrial porins (see Table 1). The single-channel conductance of these salts is a linear function of the specific conductivity of the bulk aqueous phase at small membrane potentials.

3.2
Eukaryotic Porins are Voltage-gated

The current recording shown in Fig. 1 demonstrates that closing events represent only a minor fraction of the recordings at 10 mV transmembrane potential. At larger voltages, beginning from about 20 to 25 mV, the number of closing events increases. They are always smaller than the initial on-steps, which indicate that

Tab. 1 Average single-channel conductance of porin 31HL from human T-lymphocytes in different salt solutions.

Salt	*c [M]*	*G [nS]*
KCl	0.01	0.05
	0.03	0.15
	0.1	0.45
	0.3	1.3
	1.0	4.3
	3.0	11
LiCl	1	3.2
K-acetate	1	1.5
Tris–HCl	0.5	1.5
Tris–HEPES	0.5	0.18
K-MES	0.5	0.70

The solutions contained 5–10 ng mL^{-1} porin 31HL and less than 0.1 μg mL^{-1} of the detergent NP-40; the pH was between 6.0 and 7.0. The membranes were made of diphytanoyl phosphatidylcholine/*n*-decane; $T = 25\,^{\circ}C$; $V_m = 10$ mV. *G* was determined by recording at least 70 conductance steps and averaging over the right-hand maximum (see Fig. 2). *c* is the concentration of the salt solution.

the pores switched to ion-permeable substates at voltages higher than 20 mV. An experiment of this type is shown in Fig. 3. Transmembrane potentials of +50 and −50 mV were applied to a membrane containing first three (upper trace) and then four (lower trace) mitochondrial pores from *Paramecium*. The arrow indicates the reconstitution of the fourth channel, which closes after about 0.8 s. The membrane conductance decays in defined steps to smaller values. Similar experiments were also performed with membranes containing many mitochondrial porins, such as shown in Fig. 4. Voltages of −40 and +40 mV were applied to a diphytanoyl phosphatidylcholine/*n*-decane membrane containing about 150 mitochondrial porin channels from *Paramecium*. In this case, the closing steps cannot be resolved because of the large number of pores in the membrane. The current decays in a single exponential function (see Fig. 4). The

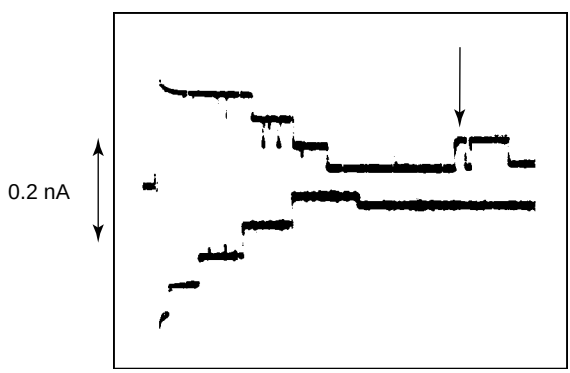

Fig. 3 Voltage-dependence of *Paramecium* porin. First a voltage of +50 mV was applied to a membrane containing three channels (upper trace). During the experiment, a fourth channel was reconstituted into the membrane (arrow), which closed after 0.8 s. Then, −50 mV was applied to the same membrane containing four channels (lower trace). The porin channels closed in defined steps. The membrane was formed from diphytanoyl phosphatidylcholine/*n*-decane. The aqueous phase contained 0.5 M Tris–HCl (pH 7.2); $T = 25\,^{\circ}C$.

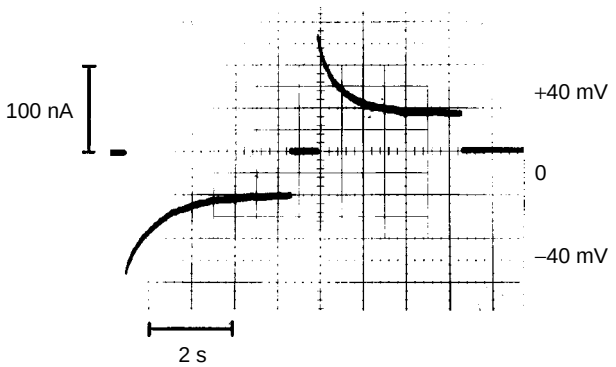

Fig. 4 Voltage dependence of *Paramecium* porin in a mutichannel experiment. About 150 porin pores were incorporated in a membrane from 1% diphytanoyl phosphatidylcholine/*n*-decane. The voltage across the membrane was switched to −40 mV (with respect to the cis-side, the side of the addition of 500 ng mL^{-1} protein) and then to +40 mV. The channels switched to substates of the open state in a single exponential curve. The aqueous phase contained 1 M KCl (pH 6); $T = 25\,^{\circ}C$.

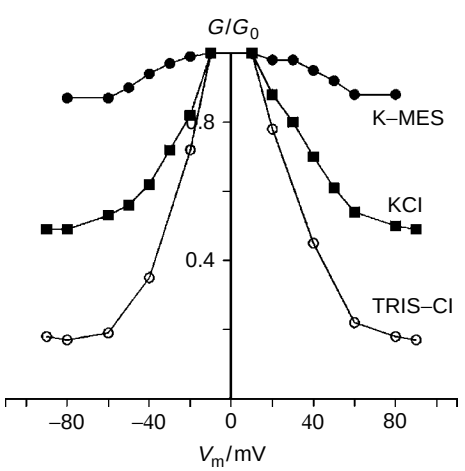

Fig. 5 Ratio of the conductance, G, at a given voltage, V_m, divided by the conductance, G_0, at 5 mV as a function of the voltage. The aqueous phase contained either 0.5 M KCl, 0.5 M K-MES or 0.5 M Tris–HCl (pH in all cases, 7.2). The cis-side contained 50 ng mL^{-1} *Paramecium* porin. The sign of the voltage is given with respect to the cis-side, the side of the addition of porin.

results of this and additional experiments at different voltages allowed the evaluation of the voltage dependence. The steady state conductance showed a bell-shaped curve as a function of the applied voltage when the conductance at a given voltage $G(V)$ divided by G_0 at zero potential is plotted as a function of membrane voltage. Figure 5 shows the results for *Paramecium* porin and three different salts (KCl, K-MES and Tris–HCl, all at pH 7.2). The results differ somewhat for different salts, although the voltage dependence is approximately the same. This has to do with the selectivity of the open channel as compared to that of the closed state (see below). The data

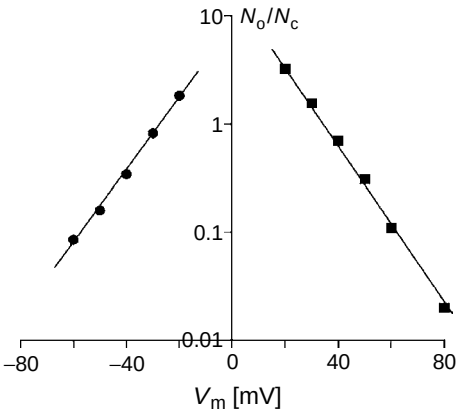

Fig. 6 Semilogarithmic plot of the ratio, N_o/N_c as a function of the transmembrane potential V_m. The data were taken from Fig. 5. The closed circles indicate the measurements with positive voltages, V_m, while the squares show measurements with negative potentials (with respect to the cis-side, the side of the addition of the protein). The slope of the straight lines is such that an e-fold change of N_o/N_c is produced by a change in V_m of 13 mV. The midpoint potential of the N_o/N_c distribution (i.e. $N_o = N_c$) is 35 mV for positive potentials (circles) and – 28 mV for negative potentials (squares). Only data for KCl are shown.

given in Fig. 5 could be analyzed as proposed previously, based on an earlier study of excitability-inducing material (EIM) by calculating the ratio of the number of open, N_o, to closed channels, N_c, from the data of the bell-shaped curves of Fig. 5:

$$N_o/N_c = (G - G_{min})/(G_0 - G) \qquad (1)$$

where G is in this equation the conductance at a given membrane potential V_m, G_0 and G_{min} are the conductance at zero voltage and very high potentials, respectively. The open to closed ratio of the channels, N_o/N_c, is given by:

$$N_o/N_c = \exp[-nF(V_m - V_0)/RT] \qquad (2)$$

where F (Faraday's constant), R (gas constant) and T (absolute temperature) are standard symbols, n is the number of gating charges moving through the entire transmembrane potential gradient for channel gating (i.e. a measure for the strength of the interaction between the electric field and the open channel), and V_0 is the potential at which 50% of the total number of channels are in the closed configuration (i.e. $N_o/N_c = 1$).

A semilogarithmic plot of the data given in Fig. 5 shows that they could be fitted to a straight line with a slope of 13 mV for an e-fold change of V_m (because $RT/F = 25$ mV). This result suggests that the number of charges involved in the gating process is approximately 2 in the case of *Paramecium* porin (see Fig. 6). The distribution of open and closed channels, N_o/N_c, follows a Boltzmann distribution. This means that (2) allows also the calculation of the energy difference for channel gating according to:

$$N_o/N_c = \exp[-W(V_m)/RT] \qquad (3)$$

$W(V_m)$ is the voltage-dependent energy difference of one mole channels between the open and the closed state. A comparison of (2) and (3) shows that $W(V_m) = nF(V_m - V_0)$. The energy needed for channel closure, nFV_0, calculated from the data of Figs. 5 and 6 is approximately 7.7 kJ mol^{-1}, an energy that is not very high. It is considerably below the energy of one hydrogen bond, which means that channel gating is a low-energy process.

The time constant τ of the single exponential relaxation process shown in Fig. 4 decreases with increasing voltage. Interestingly, the time constants could be fitted to a similar formalism as given in (2)

for the ratio N_o/N_c, which means that a semilogarithmic plot of the time constants versus voltage yields a straight line. The slope of this line corresponded to an e-fold decrease of the time constant τ for an increase of the voltage by about 13 mV under the conditions of Fig. 4. This result suggested again that the number of gating charges involved in channel gating is about 2, which agreed satisfactorily with the number of gating charges derived from the plot of N_o/N_c (see Fig. 5). The time constant of the switching of the pores from the "closed" to the "open" state could not be followed for mitochondrial porins within the time resolution of the experimental instrumentation (about 1 ms). Figure 7 shows an experiment in which the voltage across a lipid bilayer membrane containing about 30 yeast porins was first set to 70 mV. After the exponential decay of the membrane current, the voltage was decreased to 20 mV (arrow), but no relaxation was observed (see Fig. 7). This result indicated largely different reaction rates for the closing and opening of the mitochondrial porins.

All eukaryotic porins studied to date are voltage dependent. Nevertheless, the threshold voltage for complete channel closure is somewhat different for porins from different organisms. Whereas 35 mV is sufficient to decrease the initial conductance of yeast porin to 50%, about 90 mV has to be applied to membranes containing porin from rat brain to obtain the same effect. Semilogarithmic plots of the ratio N_o/N_c showed that both V_0 and n were different in these cases. These results indicated that the number of charges involved in the gating process varies for mitochondrial porins from different organisms.

3.3 Single-channel Conductance of the Closed States

The closed state of eukaryotic porins has a reduced permeability for ions. Single-channel conductance experiments allow the evaluation of the conductance of the closed state at higher membrane potentials of 30 or 40 mV. At these voltages, the open state of the channels has only a limited lifetime because of the voltage dependence of the open state. It is possible to evaluate the single-channel conductance of the closed state by subtracting the conductance of the closing events from those of the open state. Table 2 shows the results of these type of measurements obtained for three different salts and two types of porin, from yeast and VDAC1 from a human B-lymphocyte cell line (porin 31HL). The single-channel conductance

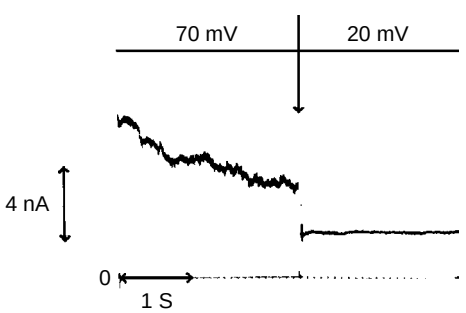

Fig. 7 Voltage dependence of the open and closed states of yeast porin. The voltage across a diphytanoyl phosphatidylcholine/*n*-decane membrane containing about 30 channels was first set to 70 mV at the cis-side of the membrane. After the relaxation of the membrane current (time constant about 1.5 s), the voltage was decreased to 20 mV (arrow). The opening of the channels was very fast (less than 50 ms) and could not be followed at the timescale of the experiment. The aqueous phase contained 1 M KCl and 20 ng mL^{-1} 1 yeast porin at the cis-side of the membrane; $T = 25\,^{\circ}C$.

Tab. 2 Average single-channel conductance of the open and closed states of yeast and human porins in different 0.5 M salt solutions.

Salt	*Open state [nS]*	*Closed state [nS]*
Yeast porin		
KCl	2.3	1.3
K-MES	0.95	0.65
Tris–HCl	1.5	0.30
Human porin (porin 31 HL)		
KCl	2.4	1.4
K-MES	0.70	0.65
Tris–HCl	1.5	0.30

The pH of the aqueous salt solutions was adjusted to 7.2. The protein concentration was between 5 and 10 ng/ml; $V_m = 30$ mV, $T = 25\,^{\circ}$C. The single-channel conductance of the closed state was calculated by subtracting the conductance of the closing events from the conductance of the initial opening of the pores.

of the closed state of the pore was considerably smaller for Tris–HCl than for K-MES, despite a similar aqueous mobility of K^+ and Cl^-. This result suggested that the closed state(s) of mitochondrial porins is cation selective.

3.4 Selectivity of the Open and Closed States of Eukaryotic Porins

The data in Table 1 suggests that ions move through the open state of mitochondrial porin in the same way they move in the bulk aqueous phase. Interestingly, the pores also exhibit certain specificity for charged solutes because the single-channel conductance in potassium acetate is somewhat smaller than that in LiCl despite the same aqueous mobility of lithium ions as compared to acetate (see Table 1). This means that the mitochondrial porin is selective although it is a wide and water-filled channel in the open state. Experiments with lipid bilayer membranes under zero-current conditions and an externally applied concentration gradient, c''/c', across the membrane allow the evaluation of the ionic selectivity. From the membrane potential, V_m, caused by the preferential movement of one sort of ions through the channel, the ratio P_{cation}/P_{anion} of the permeabilities for cations and anions can be calculated using the Goldman–Hodgkin–Katz equation.

Table 3 shows the zero-current membrane potentials and the permeability ratios for different mitochondrial and eukaryotic porins [porin 31HL (human), rat liver, yeast, and *Paramecium*] in potassium chloride, potassium acetate, and lithium chloride. It is obvious from the data of Table 3 that the asymmetry potential and

Tab. 3 Zero-current membrane potentials, V_m, of membranes from diphytanoyl phosphatidylcholine/*n*-decane in the presence of rat liver, yeast, and *Paramecium* mitochondrial porins measured for a 10-fold gradient of different salts[a].

Salt	V_m *[mV]*	P_{anion}/P_{cation}
Rat liver		
KCl (pH 6)	−11	1.7
LiCl (pH 6)	−24	3.4
K acetate(pH 7)	+14	0.50
Yeast		
KCl (pH 6)	−7	1.4
LiCl (pH 6)	−20	2.6
K acetate (pH 7)	+14	0.5
Paramecium		
KCl (pH 6)	−11	1.7
LiCl (pH 6)	−24	3.4
K acetate (pH 7)	+14	0.50

[a] V_m is defined as the potential of the dilute side (10 mM) relative to that of the concentrated side (100 mM); the temperature was 25 °C. P_{anion}/P_{cation} was calculated from the Goldman–Hodgkin–Katz equation.

the ion selectivity of mitochondrial porins are dependent on the combination of different anions and cations. The different porin channels are slightly anion selective (ratio $P_{cation}/P_{anion} = 2:1$) for potassium and chloride, which have approximately the same aqueous mobility. For the combination of the less mobile lithium ion (because of its larger hydration shell) with chloride, the anion selectivity increases. Since the ions move within the channel in a similar way as in bulk aqueous phase, the channel becomes cation selective for potassium acetate because of the smaller mobility of acetate as compared with potassium ions. This result represents another support for the assumption of the mitochondrial pore as a wide water-filled channel in the "open" state.

The open state of all mitochondrial porins characterized so far is slightly anion selective for salts composed of equally mobile cations and anions such as KCl (see above). The mitochondrial porins switch to closed states when the transmembrane voltage exceeds 20 to 30 mV. These states definitely have reduced ion permeability as the bell-shaped curves of Fig. 5 clearly indicate. The ion selectivity of the closed states cannot be measured using zero-current membrane potential measurements with pores in the closed state. However, the results of Fig. 5 and Table 2 suggest that the closed state is cation selective since for the combination K-MES (a mobile cation combined with a less mobile anion) the conductance in the open and in the closed states differs only little. The difference between open state and closed state is more substantial for Tris–HCl (a mobile anion combined with a less mobile cation). This result suggests indeed that the channel is cation selective in the closed state.

4 Inhibition of the Mitochondrial Pore

Konig and coworkers showed that an amphiphilic, synthetic polyanion (a copolymer of 10 kDa molecular mass of methacrylate, maleate, and styrene in a 1 : 2 : 3 proportion) inhibits different carriers in the mitochondrial inner membrane and the ATPase. It is possible that these effects have nothing to do with a direct interaction between the polyanion and inner membrane carriers. Moreover, it seems that the polyanion binds to mitochondrial porin and shifts its voltage dependence in a defined way, thus closing the channel when the membrane voltage has a negative sign at the cis-side, the side of the addition of the polyanion. For a positive potential at the cis-side, the channel is always in its open configuration even for voltages up to 120 mV and higher. The effect of the polyanion on mitochondrial porin from rat liver is demonstrated in Fig. 8. First the voltage dependence of reconstituted rat liver porin was studied and showed the typical bell-shaped curve. The polyanion was added at a concentration of 100 ng mL^{-1} polyanion to one side of the membrane (the cis-side). When the voltage had a negative sign on the cis-side, the membrane current started to decrease at smaller voltages than without the polyanion. A membrane potential of −10 mV was already sufficient to close the channels. Higher voltages (−40 mV; see Fig. 8) resulted in a more rapid closure of rat liver porin (relaxation time constant smaller than 100 ms), whereas +40 mV had no effect on porin-induced membrane conductance. Even voltages with magnitudes up to +120 mV did not result in any current decrease, which means that the polyanion stabilized the pore in the open state if the sign of the transmembrane

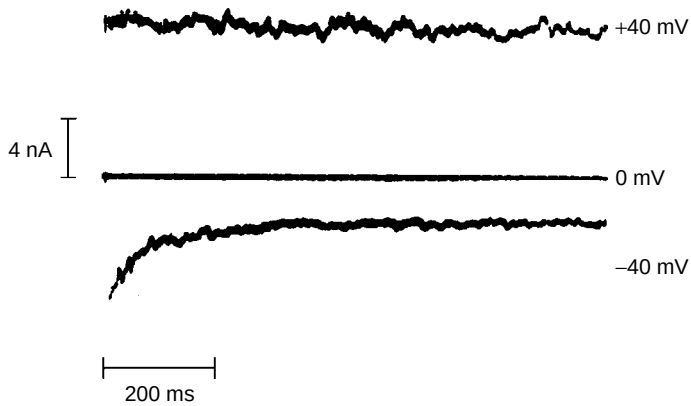

Fig. 8 Asymmetric current response of rat liver porin after application of membrane potentials of different polarity. The membrane was made of asolectin/*n*-decane. The concentration of the porin was 100 ng mL^{-1} in the 1 M KCl solution. The cis-side contained in addition 1μg mL^{-1} polyanion. A membrane potential of +40 mV (upper trace) and −40 mV (lower trace) was applied to the membrane (referred to as the *cis-side*) 40 min after the addition of the porin and 10 min after that of the polyanion. Note that the membrane current only decreased if the cis-side was negative; $T = 25\,^{\circ}C$.

potential was positive on the cis-side. The addition of the polyanion to both sides resulted in a bell-shaped curve because the channels closed at positive and negative potentials. However, the parameters of channel closure were completely different to those without polyanion, that is, n was much larger (about 4–5) and V_0 was smaller than 10 mV. The experimental data suggest that there does not exist a "sideness" for the interaction of the polyanion with porin, which means that the polyanion can interact with the gate(s) from both sides of the channel. Furthermore, the asymmetric effect of the polyanion on rat liver porin was stable during the whole lifetime of the membranes (up to several hours), indicating that the polyanion could not penetrate the channel.

The interaction of the polyanion with mitochondrial porins from different sources also allows us to measure the conductance of the closed state when the polyanion was added at a concentration of 100 ng mL^{-1} to the side of the membrane with negative polarity (the cis-side). The single-channel conductance was about half that of the open state measured at low voltage, which means that the closed channels are still permeable for ions, which makes a direct polyanion-induced block of the channel rather unlikely. Table 4 shows the effect of polyanion on the single-channel conductance obtained with salts other than KCl. The reduction of the single-channel conductance caused by the polyanion is especially large when a mobile anion (e.g. Cl^-) is combined with a less mobile cation (e.g. $Tris^+$). In this case, the single-channel conductance was reduced to less than 10% of the open state value. The effect of the polyanion on combinations of mobile cations with less mobile anions (e.g. on K-MES) was very small. These results are consistent with the assumption that the slightly anion-selective channel in the

Tab. 4 Average single-channel conductance of the open and the polyanion-induced closed state of rat liver mitochondrial porin in different 0.5 M salt solutions. (The pH of the aqueous salt solutions was adjusted to 7.2. The membrane voltage was 10 mV at the cis-side; $T = 25\,^{\circ}C$. The aqueous phase contained in the measurements of the closed state 0.1 μg mL^{-1} polyanion added to the cis-side).

Salt	*Open state [nS]*	*Closed state [nS]*
KCl	2.2	1.2
LiCl	1.8	0.40
K acetate	1.1	0.85
K-MES	0.88	0.74
Tris–HCl	1.5	0.25

open state becomes highly cation selective in the closed state (see also Fig. 5). This was also the result of selectivity measurements in the presence of polyanion. The ratio P_{cation}/P_{anion} for rat liver porin in selectivity measurements without polyanion is approximately 0.6 for KCl. For a polyanion concentration of 15 μg mL^{-1} on both sides of the membrane, P_{cation}/P_{anion} increases to about 10, which suggests that chloride has a very small permeability through the polyanion-mediated closed state of rat liver porin.

Channel blockage by the polyanion can be compared to the situation in whole mitochondria in which polyanions added to the outside (presumably the negative side of the outer membrane) are able to close the channel and inhibit the passage of anionic metabolites across the mitochondrial outer membrane, including ADP and ATP, thus blocking the mitochondrial metabolism. Other polyanions such as dextran sulfate have a similar effect on the voltage dependence of mitochondrial porin, but these compounds need a considerably larger concentration for channel closure than the polyanion. In addition to the different polyanions, the action of a protein component from mitochondria was studied in mitochondria and in reconstituted systems. This "modulator" exhibits a similar action on the porin channel as the polyanion, which means that it disturbs nucleotide exchange across the outer membrane and shifts the voltage dependence of the channel in a dose-dependent way. Although its action is known, it has not been possible to identify the modulator. It is presumably a polypeptide since it is susceptible to pronase, that is, its action on porin is destroyed by pronase treatment.

5 Structure of the Channel Formed by Eukaryotic Porins

5.1 Primary Structure of Eukaryotic Porins

The primary structure of eukaryotic porins from many organisms, including humans, mouse, fruit fly, plants, yeast, fungi, and *Dictyostelium*, is known to date. With the exception of porin 31HL (HVDAC1 from humans), all were derived from their cDNA sequences. Porin 31HL is known from its amino acid sequence. Despite many variations of single amino acids, all sequences are related to one another and have an approximate length of about 280 amino acids. Figure 9 shows a comparison of the amino acid sequences of human (HVDAC1, porin 31HL), *Drosophila melanogaster*, potato, *Saccharomyces cerevisiae* and *N. crassa* mitochondrial porins. Despite the clear homology between the different sequences, only 15 (out of about 280) amino acid residues are strictly

```
                                 ●   ●   ●
Porin 31HL   AVPPTYADLGKSARDVFTK-GYGFGLIKLDLKTKSENGLEFTSSGSANTET 50
   DM-VDAC   MAPPSYSDLGKQARDIFSK-GYNFGLWKLDLKTKTSSGIEFNTAGHSNQES 50
    P-VDAC MGKGPGLYTEIGKKARDLLYKD-YQ-SDHKFSITTYSPTGVVITSSGSKKGD- 50
   SC-VDAC   MSPPVYSDISRNINDLLNKDFYHATPAAFDVQTTTANGIKFSLKAKQPVKD 51
   NC-VDAC   MAVPAFSDIAKSANDLLNKDFYHLAAGTIEVKSNTPNNVAFKVTGK-STHD 50

                                          ●  ●           ●●●
Porin 31HL TKVTGSLETKYRWTEYGLTFTEKWNTDNTLGTEITVEDQLARGLKLTFDSSFS 103
   DM-VDAC GKVFGSLETKYKVKDYGLTLTEKWNTDNTLFTEVAVQDQLLEGLKLSLEGNFA 103
    P-VDAC -LFLADVNTQLKNKN--VTTDIKVDTNSNLFTTITV-DEAAPGLK--TILSF- 96
   SC-VDAC GPLSTNVEAKLNDKQTGLGLTQGWSNTNNLQTKLEFA-NLTPGLKNELITSLT 103
   NC-VDAC KVTSGALEGKFTDKPNGLTVTQTWNTANALETKVEMADNLAKGLKAEGIFSFL 103

                                                  ●               ●
Porin 31HL PNTGKKNAKIKTGYKREHINLGCDMDFDIAGPSIRGALVLGYEGWLAGYQMNF 156
   DM-VDAC PQSGNKNGKFKVAYGHENVKADSDVNIDLKGPLINASAVLGYQGWLAGYQTAF 156
    P-VDAC RVPDQRSGKLEVQYLHDYAGICTSVGLTAN-PIVNFSGVVGTNIIALGTDVSF 148
   SC-VDAC P-GVAKSAVLNTTFTQPFFTARGAFDLCLKSPTFVGDLTMAHEGIVGGAEFGY 155
   NC-VDAC PATNARGAKFNLHFKQSNFHGRAFFDL-LKGPTANIDAIVGHEGFLAGASAGY 155

Porin 31HL ETAKSRVTQSNFAVGYKTDEFQL-HTNVNDGTEFGGSIYQKVNKKLETAVNLA 208
   DM-VDAC DTQQSKLTTNNFALGYTTKDFVL-HTAVNDGQEFSGSIFQRTSDKLDVGVQLS 208
    P-VDAC DTKTGDFTKCNAGLSFTNADLVASLNLNNKGDNLTASYYHTVSPLTSTAVGAE 201
   SC-VDAC DISAGSISRYAMALSYFAKDYSLGAT-LNNEQITTVDFFQNVNAFLQVGAKAT 207
   NC-VDAC DVQKAAITGYSAAVGYHAPTYSAAITATDNLSVFSASYYHKVNSQVEAGSKAT 208

                 ●                 ●
Porin 31HL WT-AGNSNTR--FGIAAKYQIDPDACFSAKVNNSSLIGLGYTQTLKPGIKLTL 258
   DM-VDAC WA-SGTSNTK--FAIGAKYQLDDDASVRAKVNNASQVGLGYQQKLRDGVTLTL 258
    P-VDAC VNHSFSTNENI-ITVGTQHRLDPLTSVKARINNFGKASALLQHEWRPKSLFTV 253
   SC-VDAC MN-CKLPNSNVNIEFATRYLPDASSQVKAKVSDSGIVTLAYKQLLRPGVTLGV 259
   NC-VDAC WN-SKTGNT-VGLEVATKYRIDPVSFVKGKINDRGVAAIAYNVLLREGVTLGV 259

            ●           ● ●
Porin 31HL SALLDGKNVNAGGHKLGLGLEFQ                                281
   DM-VDAC STLVDGKNFNAGGHKIGVGLELE                                281
    P-VDAC SGEVDTKSVDK-GAKFGLALALK                                275
   SC-VDAC GSSFDALKLSEPVHKLGWSLSFD                                282
   NC-VDAC GASFDTQKLDQATHKVGTSFTFE                                282
```

Fig. 9 Comparison of the amino acid sequences of human (HVDAC1, porin 31HL), *D. melanogaster*, potato, *S. cerevisiae* and *N. crassa* mitochondrial porins using BLAST. Three or more identical amino acids are shaded. Points indicate identical amino acids in all five sequences.

conserved in all five porins (marked by points in Fig. 9). The most remarkable identity in all five porins is the GLK triplet near amino acid 90, which may play an essential, but unknown, role in eukaryotic porin function. However, it is not the place of ATP binding. Otherwise, the identities between the five porins are rather small in large stretches of the aligned porin sequences. These results suggest that the possible β-barrel structure of mitochondrial porins (see below) tolerates extensive amino acid variations without substantial alterations in the secondary structure and of the function of a voltage-dependent membrane channel, which was found for all the mitochondrial porins. Similarly, all porins form wide water-filled

channels with an average single-channel conductance of about 4 nS in 1 M KCl, that are slightly cation selective in the open state and highly cation selective in the voltage-dependent closed state.

No identity exists between primary sequences of mitochondrial porins and those of bacterial porins, especially of phototrophic bacteria, which are related to the ancestor of mitochondria. On the other hand, the structure and function of bacterial porins as channels are clearly similar to those of the mitochondrial porins, since they both contain predominantly antiparallel β-barrel stave structures, which form hollow cylinders in the outer membrane and they both have similar molecular masses. Hydropathicity plots indicate that the primary sequence of mitochondrial porins is predominantly hydrophilic (46–50% polar residues). This clearly distinguishes them from the ion channels in nerve and muscle tissue with their typical α-helical structure, although the p-elements of potassium and similar ion channels contain an approximately 20-amino acid stretch that lines the channel and is organized similar to a β-sheet structure. Eukaryotic porins contain only about a 19-amino acid long stretch at the N-terminal end that may form an α-helical region. The function of the amphipathic α-helical structure is still a matter of debate, and it is suggested that it has to do with channel gating; protein targeting or it is even a part of the channel itself.

5.2 Secondary Structure of Mitochondrial Porins

The secondary structure of most mitochondrial porins was analyzed using computer-based prediction programs. The simplest possibility of identifying amphipathic β-strands is the method proposed by Vogel and Jahnig. It is based on the observation that every second amino acid in a sided β-strand is either on average hydrophobic or hydrophilic. This means that one side of a β-strand is hydrophilic and faces the channel interior; the other side is hydrophobic and faces the lipid side chains. In addition to this simple approach, more sophisticated methods are also possible, such as secondary structure predictions based on a neural network-based predictor, which is suitable for finding β-sheets along protein sequences and is trained to recognize the topography of known bacterial β-strands. An example of this treatment is given in Fig. 10 for *N. crassa* porin. Within its primary sequence, 16 amphipathic β-strands (indicated by shaded boxes) can be recognized by this method. All studies of porin structure agree that the eukaryotic porins contain a high degree of antiparallel β-sheets as the main structural feature. The results of the secondary structure predictions are supported by CD measurements of eukaryotic porins. Native and recombinant porin from *N. crassa* shows, in these experiments, a high degree of β-sheet structure. In general, the recombinant porins seem to be a valuable model system for the study of eukaryotic porins by spectroscopic methods, in which high amounts of protein are needed. CD spectroscopy of root plastid porin was performed to determine the secondary structure of the protein under different conditions. It has a high degree of β-sheet structure in the nonionic detergent Genapol X-80 and in lipid vesicles. Although the presence of sterols is essential for reconstitution of eukaryotic porins in lipid bilayer membranes, the presence of sterols in the experiments with lipid vesicles does not change the CD spectra. The more polar detergent

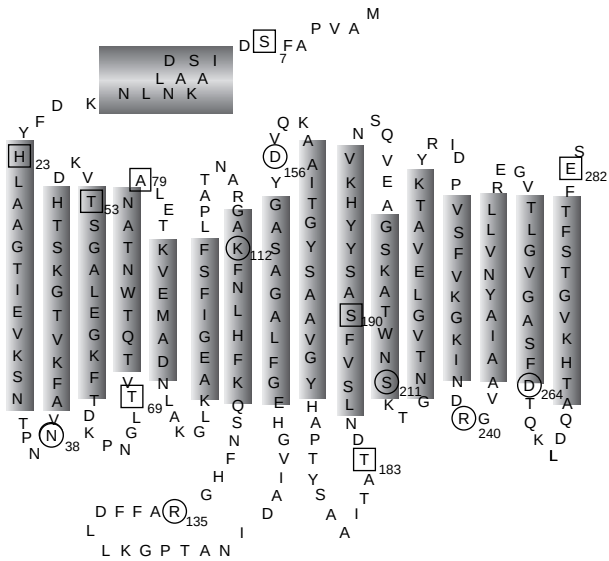

Fig. 10 Model of the arrangement of *N. crassa* porin in the membrane. The 16 putative β-strands are shaded in gray. Taken from Casadio, R., Jacoboni, I., Messina, A., De Pinto, V. (2002) A 3D model of the voltage-dependent anion channel (VDAC), *FEBS Lett.* **520**(1–3), 1–7 with permission.

SDS induces a large amount of α-helix structure in the protein. Similar results were obtained from experiments with two plant porin isoforms using attenuated total reflection Fourier-transform infrared spectroscopy (ATR-FTIR). Here, the β-sheet content of the eukaryotic porins was about 50%. Furthermore, the arrangement of the porins as detected by ATR-FTIR was very similar to that of OmpF of *E. coli*.

5.3 Structure of the Channel-forming Unit

Two different models of the channel-forming unit were derived from secondary structure predictions, site-directed mutagenesis, and deletion of part of the primary sequence of mitochondrial porins. One model assumes that the channel is composed of either 12 (yeast) or 13 (*N. crassa*) β-strands and the amphipathic N-terminal α-helix as part of the channel. This means that in the yeast model C- and N-termini of porin are on different sides of the membrane. It is, however, more likely that both are on the same side of the mitochondrial outer membrane because of the similarity to the situation in gram-negative bacteria, where they are at the periplasmic side of the outer membrane. According to the model, the amphipathic N-terminal α-helix resides in the lumen wall as part of channel gating. Its displacement by voltage should result in some sort of structural rearrangement of the pore, which leads to partial channel closure. Our own experiments with ΔN mutants of *N. crassa* porin (ΔN2–12porin and ΔN3–20porin) suggest that the N-terminus itself seems to be important for overall channel stability, rather than being directly involved in

gating. The gating characteristics of the ΔN channels were only slightly different from those of the wild-type channels, indicating that the structures responsible for the gating process are still intact in these mutants. This means that the N-terminus probably does not reside in the lumen wall of the mitochondrial outer membrane channel.

The second model assumes that the channel is formed by 16 antiparallel membrane-spanning β-strands, similar to the arrangement for general diffusion pores in the bacterial outer membranes. In this case, the amphipathic α-helix at the N-terminal end is localized within the surface of the membrane and exposed to the aqueous phase because it represents a major antigenic determinant of mitochondrial porins. These results suggest that the N-terminus is indeed localized in the surface of the membrane, most probably on the side of the intermembrane space of mitochondria; however, its possible role in channel gating remains open, but it may play a major role in channel stability (see below). Treatment of intact and broken bovine heart mitochondria with different proteases identified different cleavage sites of the pore-forming unit exposed to the aqueous phase and this was unchanged whether intact or broken mitochondria were exposed to the proteases. The cleavage sites are probably localized within surface-exposed loops of the protein. The C-terminal end may also be exposed to the surface, but it is not accessible to proteases because only a few amino acids may be localized outside the membrane. The C-terminus of *N. crassa* porin is clearly organized in antiparallel β-strands. The mutant that lacks 15 amino acids at the C-terminal (ΔC269–283porin) showed the same channel-forming activity as wild-type porin. It had a smaller single-channel conductance than wild-type protein, indicating a smaller channel size. However, its other biophysical properties, in particular, the voltage dependence, were identical to those of wild-type porin.

Figure 11 shows a model for *N. crassa* porin derived from neural network-based

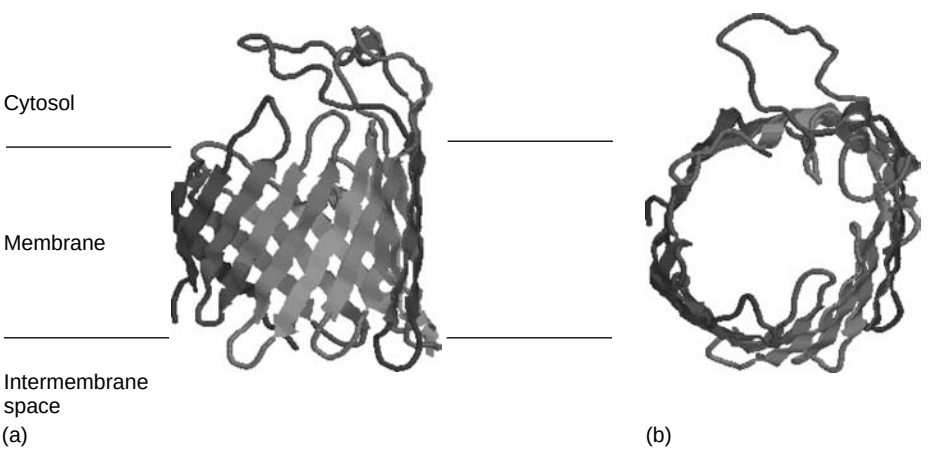

Fig. 11 Schematic view of the 3-D structure of the *N. crassa* porin. Note that the location of the N-terminus is not given. The coordinates were kindly provided by Rita Casadio and Vito De Pinto. (a) Side view. (b) View from the mitochondrial surface-exposed side.

predictors. The result of the prediction was compared to bacterial β-sheet proteins and used to align the primary sequence in coordinates of a three-dimensional (3-D) model. It is very similar to that proposed for yeast porin using the same formalism and has 16 antiparallel β-strands similar to the general diffusion porins of gram-negative bacteria. The model of *N. crassa* porin is shown in Fig. 11 ((a) from the side and (b) from the top). N- and C-termini are probably localized on the inner surface of the mitochondrial outer membrane (similar to the situation in gram-negative bacteria. The β-strands of *N. crassa* porin are tilted by about 40° to 60° similar to those of general diffusion pores. The loops connecting two β-strands are short on the internal surface of mitochondrial outer membrane and small at the opposite side. The hole is slightly oval and has a diameter of about 2.5 to 3 nm. The position of the amphipathic α-helical structure (not shown in Fig. 11) presumably represents a major problem for the validity of the model because it should somehow be involved in the stability of the pore as experiments with N-terminal deletion porins suggest. A simple, hollow β-stranded cylinder is presumably not stable in a membrane because of the lateral pressure in a lipid bilayer. It may result in a variation of its size as was shown for FhuA when the cork inside the β-barrel cylinder was removed. All known *β-stranded membrane channels* are somehow stabilized in their structure. External loop 3 stabilizes the bacterial porin channels. The α-hemolysin channel of *Staphylococcus aureus* is stabilized by the huge mushroom-like protrusion on the cis-side of the membrane and the receptors (for iron-siderophores or vitamin B12) are stabilized by the cork. A similar stabilizing element is also missing in both models for eukaryotic porins.

The model with 16 membrane-spanning β-strands appears to be more likely than that containing 13 (12) β-strands and the amphipathic N-terminus, although the latter model cannot be ruled out completely. The reasons for this can be summarized as follows:

- The porin channel is the primordial outer membrane channel and existed already when mitochondria evolved. Its overall structure may have been conserved, although there does not exist any homology between the primary structures of mitochondrial and bacterial porins. It has to be noted that homology between different bacterial porins can be very low, but the channel structure is well preserved.
- The amphipathic α-helix at the N-terminal end is the major antigenic determinant of eukaryotic porins, which means that it is probably not hidden inside the membrane.
- The results of experiments in which channel properties are changed by site-directed mutagenesis or by chemical modifications have to be taken with some care since the change of a single amino acid may induce a change that is related to a completely different region of the channel.

5.4 Are Sterols Involved in the Formation of the Channel-forming Unit?

Mitochondrial inner and outer membranes have different lipid compositions. The inner membrane contains cardiolipin, whereas sterols of different kind are exclusively found in the outer membrane dependent on the organism. Purified *N. crassa* porin contains the sterol ergosterol. Water-soluble mitochondrial porin from *N. crassa* that lost its associated

lipids and sterols needs sterols for functional reconstitution into lipid bilayer membranes. The detergent Triton X-100 plays a critical chaperone-like role in this process, similar to what has been found in the folding of bacterial porins. The role of different sterols was studied in the functional reconstitution of water-soluble porins from *Dictyostelium discoideum*, *Paramecium*, plant, and rat liver. The water-soluble porins regain their channel-forming ability after preincubation of the polypeptides with sterols in the presence of detergents. Some of the sterols can replace one another for channel formation, but influence channel gating. Interestingly, a sterol content of five cholesterol molecules per polypeptide unit was calculated in pig heart porin purified using the detergent LDAO. Sterols are probably not tightly bound to the polypeptide. The removal of sterols from the channel-forming complex may be responsible for the substantial loss of channel-forming activity during purification of mitochondrial porins.

5.5 Electron Microscopic Analysis of Mitochondrial Porin

Three-dimensional crystals of mitochondrial porin have not been obtained to date, which means that its tertiary folding structure is still not clear. However, it was possible to study 2-D crystals of fungal and recombinant human porin using electron microscopic analysis. Porin from *N. crassa* mitochondria crystallizes in periodic polymorphs with well-defined lattice parameters when outer membranes are subjected to phospholipase A2 treatment. According to the Fourier-filtered electron microscopic images of the crystalline arrays, the mitochondrial pore is a cylinder normal to the membrane plane with an outer diameter of about 3.8 nm for the hydrocarbon cylinder (the diameter of the polypeptide backbone is about 3.8 nm) and an inner diameter of about 2.5 nm. The crystalline arrays are composed of subunits that contain six porin channels. The lattice geometry responds to a number of external compounds. In particular, polyanion binds to the porin and leads to a substantial contraction of the lattice. Besides polyanion, apocytochrome *c* and the targeting sequence of an imported cytochrome oxidase subunit seem to bind to the periphery of the porin channel. More recently, 2-D crystals of *N. crassa* porin were analyzed using single-particle approaches. The results look promising, but it is not known yet if these approaches will lead to an improved resolution of porin structure. Human porin expressed in *E. coli* and reconstituted using detergent dialysis also forms 2-D crystals. The data derived from electron microscopic analysis of these crystals agree basically with those derived from 2-D crystals of *N. crassa* porin, but two neighboring molecules have opposite directions, which means that the 2-D crystals of human porin are not suitable for high-resolution imaging.

6 Conclusions

Eukaryotic porins form large water-filled channels in the "open" state with diameters around 2.5 to 3 nm that are slightly anion selective. The channels appear to be voltage-gated starting at about 20 to 30 mV and switch to ion-permeable closed states that are highly cation selective. Eukaryotic porins represent the major permeability pathway

of mitochondrial outer membranes, although porin-deficient mitochondria also exhibit oxidative phosphorylation. In addition to mitochondrial localization, there exists emerging evidence that other cellular compartments also contain eukaryotic porins with a yet unknown function. The first indication was the extramitochondrial location of porin 31HL, which is localized in the cytoplasmic membrane of a transformed human B-lymphocyte cell line. A similar channel is localized in astrocytic plasma membranes. Later, more evidence was found of such a location because a 36-kDa porin-like protein from rat brain copurified with the γ-aminobutyric acid/benzodiazepine receptor protein. It is possible that the channel from the astrocytic plasma membrane and the 36-kDa protein from rat brain relate to the "maxi" chloride channels present in the cell membrane of a number of eukaryotic cells. In excised patches, the channel has a maximum open probability at zero voltage, similar to the situation in lipid bilayer experiments.

The electrophysiological characterization of mitochondrial porin suggests that the permeability properties of the mitochondrial outer membrane could be voltage controlled. The generation of a voltage across the mitochondrial outer membrane is still open and it is not clear if the close apposition of inner and outer membranes *in situ* mitochondria is sufficient for electric coupling of both membranes and for channel gating. Furthermore, the exact mechanism of channel gating and the residues responsible for selectivity in the different states are completely open. In general, there exists some knowledge about the secondary structure of the eukaryotic porins. However, the precise 3-D structure, including the exact number of membrane-spanning β-strands and the diameter, are still unknown. Crystallographic data are badly needed to give clear answers to these questions and possibly also the localization of sterols in the channel-forming unit. Another open problem is the assembly of the porin channel. It is unlikely that it (i.e. the β-strands) forms inside a membrane since eukaryotic porins are polar proteins and are synthesized as water-soluble precursors outside membranes. This means that the channel is formed chaperone-promoted on the (inner?) surface of the mitochondrial outer membrane at a receptor site (Tom 20) from polypeptide and sterols. After formation, the channel inserts into the membrane using the general import pathway (TOM). So far it is not clear if it requires special insertion sites in the mitochondrial outer membrane or if the insertion process is similar to that in lipid bilayer membranes. In any case, it is likely that the channel is preformed before insertion. The preformation of channels in the aqueous phase or on the membrane surface is probably also the reason why lipid bilayer membranes represent a powerful tool for the study of mitochondrial and bacterial porins.

Acknowledgments

The generous supply of Figs. 10 and 11 by Rita Casadio and Vito De Pinto is gratefully acknowledged. I would like to thank Dieter Brdiczka, Vito De Pinto and Friedrich P. Thinnes for the fruitful and excellent collaboration during many years, and my collaborators Otto Ludwig, Birgit Popp, Elke Maier, and Angela Schmid for their excellent work with outer-membrane porins on which this review is largely based.

My own research was supported by the Deutsche Forschungsgemeinschaft (Project Be865/10) and by the Fonds der Chemischen Industrie.

See also Calcium Biochemistry.

Bibliography

Books and Reviews

Bauer, M.F., Hofmann, S., Neupert, W. (2002) Import of mitochondrial proteins, *Int. Rev. Neurobiol.* **53**, 57–90.

Bay, D.C., Court, D.A. (2002) Origami in the outer membrane: the transmembrane arrangement of mitochondrial porins, *Biochem. Cell. Biol.* **80**(5), 551–562.

Benz, R. (Ed.) (2004) *Bacterial and Eukaryotic Porins: Structure, Function, Mechanism*, VCH-Wiley, Weinheim, Germany.

Colombini, M. (2004) VDAC: the channel at the interface between mitochondria and the cytosol, *Mol. Cell Biochem.* **256–257**(1–2), 107–115.

De Pinto, V., Messina, A., Accardi, R., Aiello, R., Guarino, F., Tomasello, M.F., Tommasino, M., Tasco, G., Casadio, R., Benz, R., De Giorgi, F., Ichas, F., Baker, M., Lawen, A. (2003) New functions of an old protein: the eukaryotic porin or voltage dependent anion selective channel (VDAC), *Ital. J. Biochem.* **52**(1), 17–24.

Graham, B.H., Craigen, W.J. (2004) Genetic approaches to analyzing mitochondrial outer membrane permeability, *Curr. Top. Dev. Biol.* **59**, 87–118.

Palmieri, L., Runswick, M.J., Fiermonte, G., Walker, J.E., Palmieri, F. (2000) Yeast mitochondrial carriers: bacterial expression, biochemical identification and metabolic significance, *J. Bioenerg. Biomembr.* **32**(1), 67–77.

Shoshan-Barmatz, V., Gincel, D. (2003) The voltage-dependent anion channel: characterization, modulation, and role in mitochondrial function in cell life and death, *Cell. Biochem. Biophys.* **39**(3), 279–292.

Taylor, R.D., Pfanner, N. (2004) The protein import and assembly machinery of the mitochondrial outer membrane, *Biochim. Biophys. Acta* **1658**(1–2), 37–43.

Vyssokikh, M.Y., Brdiczka, D. (2003) The function of complexes between the outer mitochondrial membrane pore (VDAC) and the adenine nucleotide translocase in regulation of energy metabolism and apoptosis, *Acta Biochim. Pol.* **50**(2), 389–404.

Primary Literature

Aiello, R., Messina, A., Schiffler, B., Benz, R., Tasco, G., Casadio, R., De Pinto, V. (2004) Functional characterization of a second porin isoform in *Drosophila melanogaster*. DmPorin2 forms voltage-independent cation-selective pores, *J. Biol. Chem.* **279**(24), 25364–25373.

Albrecht, H., Goormaghtigh, E., Ruysschaert, J.M., Homble, F. (2000) Structure and orientation of two voltage-dependent anion-selective channel isoforms. An attenuated total reflection Fourier-transform infrared spectroscopy study, *J. Biol. Chem.* **275**(52), 40992–40999.

Albrecht, H., Wattiez, R., Ruysschaert, J.M., Homble, F. (2000) Purification and characterization of two voltage-dependent anion channel isoforms from plant seeds, *Plant. Physiol.* **124**(3), 1181–1190.

Aljamal, J.A., Genchi, G., De Pinto, V., Stefanazzi, L., Benz, R., Palmieri, F. (1993) Purification and characterization of porin from corn (*Zea mays* L.) mitochondria, *Plant Physiol.* **102**, 615–621.

Arora, K.K., Parry, D.M., Pedersen, P.L. (1992) Hexokinase receptors: preferential enzyme binding in normal cells to nonmitochondrial sites and in transformed cells to mitochondrial sites, *J. Bioenerg. Biomembr.* **24**(1), 47–53.

Babel, D., Walter, G., Gotz, H., Thinnes, F.P., Jurgens, L., Konig, U., Hilschmann, N. (1991) Studies on human porin. VI. Production and characterization of eight monoclonal mouse antibodies against the human VDAC "Porin 31HL" and their application for histotopological studies in human skeletal muscle, *Biol. Chem. Hoppe Seyler* **372**(12), 1027–1034.

Benz, R., Kottke, M., Brdiczka, D. (1990) The cationically selective state of the mitochondrial outer membrane pore: a study with intact mitochondria and reconstituted mitochondrial

porin, *Biochim. Biophys. Acta* **1022**(3), 311–318.

Benz, R., Schmid, A., Dihanich, M. (1989) Pores from mitochondrial outer membranes of yeast and a porin-deficient yeast mutant: a comparison, *J. Bioenerg. Biomembr.* **21**(4), 439–450.

Benz, R., Janko, K., Boos, W., Lauger, P. (1978) Formation of large, ion-permeable membrane channels by the matrix protein (porin) of *Escherichia coli*, *Biochim. Biophys. Acta* **511**(3), 305–319.

Benz, R., Wojtczak, L., Bosch, W., Brdiczka, D. (1988) Inhibition of adenine nucleotide transport through the mitochondrial porin by a synthetic polyanion, *FEBS Lett.* **231**(1), 75–80.

Benz, R., Maier, E., Thinnes, F.P., Gotz, H., Hilschmann, N. (1992) Studies on human porin. VII. The channel properties of the human B-lymphocyte membrane-derived "Porin 31HL" are similar to those of mitochondrial porins, *Biol. Chem. Hoppe Seyler* **373**(6), 295–303.

Blachly-Dyson, E., Peng, S., Colombini, M., Forte, M. (1990) Selectivity changes in site-directed mutants of the VDAC ion channel: structural implications, *Science* **247**(4947), 1233–1236.

Blumenthal, K., Kahn, A., Beja, O., Galun, E., Colombini, M., Breiman, A. (1993) Purification and characterization of the voltage-dependent anion-selective channel protein from wheat mitochondrial membranes, *Plant Physiol.* **101**, 579–587.

Brdiczka, D. (1994) Function of the outer mitochondrial compartment in regulation of energy metabolism, *Biochim. Biophys. Acta* **1187**(2), 264–269.

Bureau, M.H., Khrestchatisky, M., Heeren, M.A., Zambrowicz, E.B., Kim, H., Grisar, T.M., Colombini, M., Tobin, A.J., Olsen, R.W. (1992) Isolation and cloning of a voltage-dependent anion channel-like Mr 36,000 polypeptide from mammalian brain, *J. Biol. Chem.* **267**(12), 8679–8684.

Carbonara, F., Popp, B., Schmid, A., Iacobazzi, V., Genchi, G., Palmieri, F., Benz, R. (1996) The role of sterols in the functional reconstitution of water-soluble mitochondrial porins from plants, *J. Bioenerg. Biomembr.* **28**(2), 181–189.

Casadio, R., Jacoboni, I., Messina, A., De Pinto, V. (2002) A 3D model of the voltage-dependent anion channel (VDAC), *FEBS Lett.* **520**(1–3), 1–7.

Castellan, G.W. (1983) The Ionic Current in Solutions, *Physical Chemistry*, Addison-Wesley, Reading, MA, pp. 769–784.

Cheng, E.H., Sheiko, T.V., Fisher, J.K., Craigen, W.J., Korsmeyer, S.J. (2003) VDAC2 inhibits BAK activation and mitochondrial apoptosis, *Science* **301**(5632), 513–517.

Colombini, M. (1979) A candidate for the permeability pathway of the outer mitochondrial membrane, *Nature* **279**(5714), 643–645.

Colombini, M. (1983) Purification of VDAC (voltage-dependent anion-selective channel) from rat liver mitochondria, *J. Membr. Biol.* **74**(2), 115–121.

Colombini, M., Holden, M.J., Mangan, P.S. (1989) Modulation of the Mitochondrial Channel VDAC by a Variety of Agents, in: Nalecz, K.A., Nalecz, M.J., Wojtczak, L., Azzi, A. (Eds.) *Ion Carriers of Mitochondrial Membranes*, Springer, New York, pp. 215–224.

Colombini, M., Yeung, C.L., Tung, J., Konig, T. (1987) The mitochondrial outer membrane channel, VDAC, is regulated by a synthetic polyanion, *Biochim. Biophys. Acta* **905**(2), 279–286.

Cowan, S.W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R.A., Jansonius, J.N., Rosenbusch, J.P. (1992) Crystal structures explain functional properties of two E. coli porins, *Nature* **358**(6389), 727–733.

De Pinto, V., Benz, R., Palmieri, F. (1989) Interaction of non-classical detergents with the mitochondrial porin. A new purification procedure and characterization of the pore-forming unit, *Eur. J. Biochem.* **183**(1), 179–187.

De Pinto, V., Benz, R., Caggese, C., Palmieri, F. (1989) Characterization of the mitochondrial porin from *Drosophila melanogaster*, *Biochim. Biophys. Acta* **987**(1), 1–7.

De Pinto, V., Ludwig, O., Krause, J., Benz, R., Palmieri, F. (1987) Porin pores of mitochondrial outer membranes from high and low eukaryotic cells: biochemical and biophysical characterization, *Biochim. Biophys. Acta* **894**(2), 109–119.

De Pinto, V., Prezioso, G., Thinnes, F., Link, T.A., Palmieri, F. (1991) Peptide-specific antibodies and proteases as probes of the transmembrane topology of the bovine heart

mitochondrial porin, *Biochemistry* **30**(42), 10191–10200.

De Pinto, V., Messina, A., Schmid, A., Simonetti, S., Carnevale, F., Benz, R. (2000) Characterization of channel-forming activity in muscle biopsy from a porin-deficient human patient, *J. Bioenerg. Biomembr.* **32**(6), 585–593.

Dermietzel, R., Hwang, T.K., Buettner, R., Hofer, A., Dotzler, E., Kremer, M., Deutzmann, R., Thinnes, F.P., Fishman, G.I., Spray, D.C., Siemen, D. (1994) Cloning and in situ localization of a brain-derived porin that constitutes a large-conductance anion channel in astrocytic plasma membranes, *Proc. Natl. Acad. Sci. USA* **91**(2), 499–503.

Dihanich, M., Suda, K., Schatz, G. (1987) A yeast mutant lacking mitochondrial porin is respiratory-deficient, but can recover respiration with simultaneous accumulation of an 86-kD extramitochondrial protein, *EMBO J.* **6**(3), 723–728.

Dolder, M., Zeth, K., Tittmann, P., Gross, H., Welte, W., Wallimann, T. (1999) Crystallization of the human, mitochondrial voltage-dependent anion-selective channel in the presence of phospholipids, *J. Struct. Biol.* **127**(1), 64–71.

Ehrenstein, G., Lecar, H., Nossal, R. (1970) The nature of the negative resistance in bimolecular lipid membranes containing excitability-inducing material, *J. Gen. Physiol.* **55**(1), 119–133.

Fiek, C., Benz, R., Roos, N., Brdiczka, D. (1982) Evidence for identity between the hexokinase-binding protein and the mitochondrial porin in the outer membrane of rat liver mitochondria, *Biochim. Biophys. Acta* **688**(2), 429–440.

Freitag, H., Janes, M., Neupert, W. (1982) Biosynthesis of mitochondrial porin and insertion into the outer mitochondrial membrane of *Neurospora crassa*, *Eur. J. Biochem.* **126**(1), 197–202.

Freitag, H., Neupert, W., Benz, R. (1982) Purification and characterisation of a pore protein of the outer mitochondrial membrane from *Neurospora crassa*, *Eur. J. Biochem.* **123**(3), 629–636.

Godbole, A., Varghese, J., Sarin, A., Mathew, M.K. (2003) VDAC is a conserved element of death pathways in plant and animal systems, *Biochim. Biophys. Acta* **1642**(1–2), 87–96.

Guo, X.W., Mannella, C.A. (1993) Conformational change in the mitochondrial channel, VDAC, detected by electron cryo-microscopy, *Biophys. J.* **64**(2), 545–549.

Guo, X.W., Smith, P.R., Cognon, B., D'Arcangelis, D., Dolginova, E., Mannella, C.A. (1995) Molecular design of the voltage-dependent, anion-selective channel in the mitochondrial outer membrane, *J. Struct. Biol.* **114**(1), 41–59.

Heins, L., Mentzel, H., Schmid, A., Benz, R., Schmitz, U.K. (1994) Biochemical, molecular, and functional characterization of porin isoforms from potato mitochondria, *J. Biol. Chem.* **269**(42), 26402–26410.

Hoch, F.L. (1992) Cardiolipins and biomembrane function, *Biochim. Biophys. Acta* **1113**(1), 71–133.

Hovius, R., Lambrechts, H., Nicolay, K., de Kruijff, B. (1990) Improved methods to isolate and subfractionate rat liver mitochondria. Lipid composition of the inner and outer membrane, *Biochim. Biophys. Acta.* **1021**(2), 217–226.

Jacoboni, I., Martelli, P.L., Fariselli, P., De Pinto, V., Casadio, R. (2001) Prediction of the transmembrane regions of beta-barrel membrane proteins with a neural network-based predictor, *Protein Sci.* **10**(4), 779–787.

Kayser, H., Kratzin, H.D., Thinnes, F.P., Gotz, H., Schmidt, W.E., Eckart, K., Hilschmann, N. (1989) Identification of human porins. II. Characterization and primary structure of a 31 kDa porin from human B lymphocytes (Porin 31HL), *Biol. Chem. Hoppe Seyler* **370**(12), 1265–1278.

Kleene, R., Pfanner, N., Pfaller, R., Link, T.A., Sebald, W., Neupert, W., Tropschug, M. (1987) Mitochondrial porin of *Neurospora crassa*: cDNA cloning, in vitro expression and import into mitochondria, *EMBO J.* **6**(9), 2627–2633.

Konig, T., Stipani, I., Horvath, I., Palmieri, F. (1982) Inhibition of mitochondrial substrate anion translocators by a synthetic amphipathic polyanion, *J. Bioenerg. Biomembr.* **14**(5–6), 297–305.

Konig, T., Kocsis, B., Meszaros, L., Nahm, K., Zoltan, S., Horvath, I. (1977) Interaction of a synthetic polyanion with rat liver mitochondria, *Biochim. Biophys. Acta* **462**(2), 380–389.

Koppel, D.A., Kinnally, K.W., Masters, P., Forte, M., Blachly-Dyson, E., Mannella, C.A. (1998) Bacterial expression and characterization of the mitochondrial outer membrane channel.

Effects of N-terminal modifications, *J. Biol. Chem.* **273**(22), 13794–13800.

Krause, J., Hay, R., Kowollik, C., Brdiczka, D. (1986) Cross-linking analysis of yeast mitochondrial outer membrane, *Biochim. Biophys. Acta* **860**(3), 690–698.

Krimmer, T., Rapaport, D., Ryan, M.T., Meisinger, C., Kassenbrock, C.K., Blachly-Dyson, E., Forte, M., Douglas, M.G., Neupert, W., Nargang, F.E., Pfanner, N. (2001) Biogenesis of porin of the outer mitochondrial membrane involves an import pathway via receptors and the general import pore of the TOM complex, *J. Cell Biol.* **152**(2), 289–300.

Kunkele, K.P., Heins, S., Dembowski, M., Nargang, F.E., Benz, R., Thieffry, M., Walz, J., Lill, R., Nussberger, S., Neupert, W. (1998) The preprotein translocation channel of the outer membrane of mitochondria, *Cell* **93**(6), 1009–1019.

Le Mellay, V., Troppmair, J., Benz, R., Rapp, U.R. (2002) Negative regulation of mitochondrial VDAC channels by C-Raf kinase, *BMC Cell Biol.* **3**(1), 14.

Linden, M., Gellerfors, P., Nelson, B.D. (1982) Purification of a protein having pore forming activity from the rat liver mitochondrial outer membrane, *Biochem. J.* **208**(1), 77–82.

Liu, M.Y., Colombini, M. (1992) A soluble mitochondrial protein increases the voltage dependence of the mitochondrial channel, VDAC, *J. Bioenerg. Biomembr.* **24**(1), 41–46.

Ludwig, O., Benz, R., Schultz, J.E. (1989) Porin of *Paramecium* mitochondria isolation, characterization and ion selectivity of the closed state, *Biochim. Biophys. Acta* **978**(2), 319–327.

Ludwig, O., De Pinto, V., Palmieri, F., Benz, R. (1986) Pore formation by the mitochondrial porin of rat brain in lipid bilayer membranes, *Biochim. Biophys. Acta* **860**(2), 268–276.

Ludwig, O., Krause, J., Hay, R., Benz, R. (1988) Purification and characterization of the pore forming protein of yeast mitochondrial outer membrane, *Eur. Biophys. J.* **15**(5), 269–276.

MacKinnon, R. (2003) Potassium channels, *FEBS Lett.* **555**(1), 62–65.

Mannella, C.A. (1984) Phospholipase-induced crystallization of channels in mitochondrial outer membranes, *Science* **224**(4645), 165–166.

Mannella, C.A. (1998) Conformational changes in the mitochondrial channel protein, VDAC, and their functional implications, *J. Struct. Biol.* **121**(2), 207–218.

Mannella, C.A., Guo, X.W. (1990) Interaction between the VDAC channel and a polyanionic effector. An electron microscopic study, *Biophys. J.* **57**(1), 23–31.

Mannella, C.A., Forte, M., Colombini, M. (1992) Toward the molecular structure of the mitochondrial channel, VDAC, *J. Bioenerg. Biomembr.* **24**(1), 7–19.

Mannella, C.A., Guo, X.W., Dias, J. (1992) Binding of a synthetic targeting peptide to a mitochondrial channel protein, *J. Bioenerg. Biomembr.* **24**(1), 55–61.

Mannella, C.A., Ribeiro, A.J., Frank, J. (1987) Cytochrome c binds to lipid domains in arrays of mitochondrial outer membrane channels, *Biophys. J.* **51**(2), 221–226.

Mannella, C.A., Ribeiro, A., Cognon, B., D'Arcangelis, D. (1992) Structure of paracrystalline arrays on outer membranes of rat-liver and rat-heart mitochondria, *J. Struct. Biol.* **108**(3), 227–237.

Messina, A., Neri, M., Perosa, F., Caggese, C., Marino, M., Caizzi, R., De Pinto, V. (1996) Cloning and chromosomal localization of a cDNA encoding a mitochondrial porin from Drosophila melanogaster, *FEBS Lett.* **384**(1), 9–13.

Michejda, J., Guo, X.J., Lauquin, G.J. (1990) The respiration of cells and mitochondria of porin deficient yeast mutants is coupled, *Biochem. Biophys. Res. Commun.* **171**(1), 354–361.

Mihara, K., Sato, R. (1985) Molecular cloning and sequencing of cDNA for yeast porin, an outer mitochondrial membrane protein: a search for targeting signal in the primary structure, *EMBO J.* **4**(3), 769–774.

Nikaido, H. (1983) Proteins forming large channels from bacterial and mitochondrial outer membranes: porins and phage lambda receptor protein, *Methods Enzymol.* **97**, 85–100.

Nikaido, H., Rosenberg, E.Y. (1981) Effect on solute size on diffusion rates through the transmembrane pores of the outer membrane of *Escherichia coli*, *J. Gen. Physiol.* **77**(2), 121–135.

Ostlund, A.K., Gohring, U., Krause, J., Brdiczka, D. (1983) The binding of glycerol kinase to the outer membrane of rat liver mitochondria: its importance in metabolic regulation, *Biochem. Med.* **30**(2), 231–245.

Pedersen, P.L., Greenawalt, J.W., Reynafarje, B., Hullihen, J., Decker, G.L., Soper, J.W., Bustamente, E. (1978) Preparation and characterization of mitochondria and submitochondrial particles of rat liver and liver-derived tissues, *Methods Cell Biol.* **20**, 411–481.

Peng, S., Blachly-Dyson, E., Colombini, M., Forte, M. (1992) Determination of the number of polypeptide subunits in a functional VDAC channel from *Saccharomyces cerevisiae*, *J. Bioenerg. Biomembr.* **24**(1), 27–31.

Pfaff, E., Klingenberg, M., Heldt, H.W. (1965) Unspecific permeation and specific exchange of adenine nucleotides in liver mitochondria, *Biochim. Biophys. Acta* **104**(1), 312–315.

Pfaller, R., Freitag, H., Harmey, M.A., Benz, R., Neupert, W. (1985) A water-soluble form of porin from the mitochondrial outer membrane of *Neurospora crassa*. Properties and relationship to the biosynthetic precursor form, *J. Biol. Chem.* **260**(13), 8188–8193.

Popp, B., Schmid, A., Benz, R. (1995) Role of sterols in the functional reconstitution of water-soluble mitochondrial porins from different organisms, *Biochemistry* **34**(10), 3352–3361.

Popp, B., Court, D.A., Benz, R., Neupert, W., Lill, R. (1996) The role of the N and C termini of recombinant *Neurospora* mitochondrial porin in channel formation and voltage-dependent gating, *J. Biol. Chem.* **271**(23), 13593–13599.

Popp, B., Gebauer, S., Fischer, K., Flugge, U.I., Benz, R. (1997) Study of structure and function of recombinant pea root plastid porin by biophysical methods, *Biochemistry* **36**(10), 2844–2852.

Renkin, E.M. (1954) Filtration, diffusion, and molecular sieving through porous cellulose membranes, *J. Gen. Physiol.* **38**(2), 225–243.

Roos, N., Benz, R., Brdiczka, D. (1982) Identification and characterization of the pore-forming protein in the outer membrane of rat liver mitochondria, *Biochim. Biophys. Acta* **686**(2), 204–214.

Runke, G., Maier, E., O'Neil, J.D., Benz, R., Court, D.A. (2000) Functional characterization of the conserved "GLK" motif in mitochondrial porin from *Neurospora crassa*, *J. Bioenerg. Biomembr.* **32**(6), 563–570.

Schein, S.J., Colombini, M., Finkelstein, A. (1976) Reconstitution in planar lipid bilayers of a voltage-dependent anion-selective channel obtained from *Paramecium* mitochondria, *J. Membr. Biol.* **30**(2), 99–120.

Schmid, A., Kromer, S., Heldt, H.W., Benz, R. (1992) Identification of two general diffusion channels in the outer membrane of pea mitochondria, *Biochim. Biophys. Acta* **1112**(2), 174–180.

Sen, K., Nikaido, H. (1990) In vitro trimerization of OmpF porin secreted by spheroplasts of *Escherichia coli*, *Proc. Natl. Acad. Sci. U S A* **87**(2), 743–747.

Shao, L., Kinnally, K.W., Mannella, C.A. (1996) Circular dichroism studies of the mitochondrial channel, VDAC, from *Neurospora crassa*, *Biophys. J.* **71**(2), 778–786.

Shinohara, Y., Ishida, T., Hino, M., Yamazaki, N., Baba, Y., Terada, H. (2000) Characterization of porin isoforms expressed in tumor cells, *Eur. J. Biochem.* **267**(19), 6067–6073.

Smack, D.P., Colombini, M. (1985) Voltage-dependent channels found in the membrane fraction of corn zea-mays cultivar w-64an mitochondria, *Plant Physiol.* **79**(4), 1094–1097.

Song, J., Midson, C., Blachly-Dyson, E., Forte, M., Colombini, M. (1998) The topology of VDAC as probed by biotin modification, *J. Biol. Chem.* **273**(38), 24406–24413.

Song, L., Hobaugh, M.R., Shustak, C., Cheley, S., Bayley, H., Gouaux, J.E. (1996) Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore, *Science* **274**(5294), 1859–1866.

Sottocasa, G.L., Kuylenstierna, B., Ernster, L., Bergstrand, A. (1967) An electron-transport system associated with the outer membrane of liver mitochondria. A biochemical and morphological study, *J. Cell Biol.* **32**(2), 415–438.

Thinnes, F.P., Gotz, H., Kayser, H., Benz, R., Schmidt, W.E., Kratzin, H.D., Hilschmann, N. (1989) Identification of human porins. I. Purification of a porin from human B-lymphocytes (Porin 31HL) and the topochemical proof of its expression on the plasmalemma of the progenitor cell, *Biol. Chem. Hoppe Seyler* **370**(12), 1253–1264.

Troll, H., Malchow, D., Muller-Taubenberger, A., Humbel, B., Lottspeich, F., Ecke, M., Gerisch, G., Schmid, A., Benz, R. (1992) Purification, functional characterization, and cDNA sequencing of mitochondrial porin from *Dictyostelium discoideum*, *J. Biol. Chem.* **267**(29), 21072–21079.

Tsujimoto, Y., Shimizu, S. (2002) The voltage-dependent anion channel: an essential

player in apoptosis, *Biochimie* **84**(2–3), 187–193.

Verschoor, A., Tivol, W.F., Mannella, C.A. (2001) Single-particle approaches in the analysis of small 2D crystals of the mitochondrial channel VDAC, *J. Struct. Biol.* **133**(2–3), 254–265.

Vogel, H., Jahnig, F. (1986) Models for the structure of outer-membrane proteins of Escherichia coli derived from Raman spectroscopy and prediction methods, *J. Mol. Biol.* **190**(2), 191–199.

Zalman, L.S., Nikaido, H., Kagawa, Y. (1980) Mitochondrial outer membrane contains a protein producing nonspecific diffusion channels, *J. Biol. Chem.* **255**(5), 1771–1774.

Mobile Structures: Cilia and Flagella

Koji Yonekura[1,2,3] *and Keiichi Namba*[2,3]
[1] *University of California, San Francisco, CA*
[2] *Osaka University, Osaka, Japan*
[3] *Dynamic NanoMachine Project, Suita, Osaka, Japan*

1 Eukaryotic Flagella and Cilia 389
1.1 The Axoneme 390
1.1.1 Microtubule 390
1.1.2 Axonemal Dynein 391
1.1.3 The Radial Spoke and the Central Pair 392
1.1.4 Other Proteins Associated with the Axoneme Structure 392
1.2 The Basal Body 393
1.3 The Intraflagellar Transport 394
1.4 Perspectives 395

2 Bacterial Flagella 395
2.1 Component Structures 396
2.1.1 The Basal Body 396
2.1.2 The Motor 397
2.1.3 The Hook and the Hook-filament Junction 397
2.1.4 The Flagellar Filament 398
2.1.5 The Filament Cap 401
2.2 The Assembly Process of the Bacterial Flagellum 402
2.2.1 The Type III Protein Export Apparatus 403
2.2.2 Growth of the Flagellar Filament 403
2.3 Perspectives 406

Bibliography 408
Books and Reviews 408
Primary Literature 408

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

Keywords

Axoneme
The core structure of the eukaryotic flagella and cilia composed of nine outer doublets of microtubule and a central pair of single microtubules, called "9 + 2".

Bacterial Flagellum
A long filamentous propeller rotated by a rotary motor used for bacteria swimming, composed of ~25 different proteins. Driven by ion flows through the motor part embedded in the cytoplasmic membrane.

Cilium
A short hairlike organelle found in many cells with almost the same structure with the eukaryotic flagella. Working for the transport of fluids over epithelial cells by a whiplike stroke. Some work as a sensor.

Dynein
A processive motor protein complex. Axonemal dynein generates the force of the eukaryotic flagellar and ciliary bending.

Eukaryotic Flagellum
A long filamentous organelle used for protozoan and metazoan sperm swimming, composed of highly organized microtubules and ~250 associated proteins. Showing a wavelike oscillatory motion generated by the dynein motor within the filament using the energy of ATP hydrolysis.

Flagellin
A structural protein forming the bacterial flagellar filament.

Intraflagellar Transport (IFT)
A system for transport of eukaryotic flagellar and ciliary components to the tip or base. Necessary for their assembly, disassembly, and maintenance. Driven by a kinesin motor to the distal tip (anterograde IFT) and a cytoplasmic dynen motor to the base (retrograde IFT).

Microtubule
A cytoskeleton composed of the dimer of α- and β-tubulin. Forming the core of the axoneme. Also associated with various cell activities, for example, mitosis, and so on.

MotA/B, PomA/B
An ion-driven torque generation unit for the rotation of the bacterial flagellum. MotA/B is a proton channel found in *Escherichia coli, Salmonella*, and so on. PomA/B is a sodium channel in alkaliphilic *Bacillus* and *Vibrio* species.

Type III Protein Export Apparatus
A protein complex for export of bacterial flagellar axial proteins into the cell exterior or virulence factors into host cells, driven by the energy of ATP hydrolysis.

Both eukaryote and prokaryote have flagella with a similar overall look, and use them for swimming in liquids, but they are completely different from each other. Cilia look like short hairs and have almost the same structure and mechanism as the eukaryotic flagella. The long filamentous structure of the eukaryotic flagellum is composed of highly organized microtubule-based structure and huge numbers of its associate proteins, while that in the bacterial flagellum is made of a single protein called *flagellin*. The length and the diameter of the filament in eukaryote are more than 10 times larger than those in bacteria. The total number of component proteins involved in the structure of eukaryote is also ~10 times of the prokaryote. The eukaryotic flagellum makes wavelike movements generated by the dynein motor within the filament using the energy of ATP hydrolysis, while the bacterial flagellum is a rotary motor driven by ion flows through the motor part embedded in the membranes. They both grow at their distal end, but use quite different mechanism for transport of their components to the distal tip.

Although the eukaryotic and bacterial flagella are complex macromolecules, recent studies are beginning to elucidate the roles of component proteins, their complex structures and growth mechanisms.

1 Eukaryotic Flagella and Cilia

Eukaryotic flagellum is a highly conserved organelle throughout protozoan to mammalian sperm. Its length is diverged in 10 to 200 μm, and the diameter is ~250 nm or more. The flagella show a wavelike oscillatory motion for swimming in fluid environments. Cilia are hairlike projections extended from the surface of various cells. They are not only shorter versions of the flagella but also work as systems for sensory perception of environments. In this article, we will not further discuss on these nonmotile cilia. Motile cilia show a whiplike stroke at 0.1 to 0.2 s per cycle for fluid transport over epithelial cells. For example, there are huge numbers of cilia in our respiratory tract. They sweep layers of mucus to transport trapped dusts and bacteria by their harmonic motions, up to the mouth where they are drunk or spewed out.

The motions of flagella and cilia are generated by bending of their internal cytoskeletal structure called the *axoneme*. The axoneme is composed of highly organized microtubules and their associated proteins, which include a motor protein complex, axonemal dynein. It generates the force for axoneme bending by utilizing the energy of ATP hydrolysis. In the axoneme, there are nine outer doublet microtubules surrounding the two single microtubules at its center, which

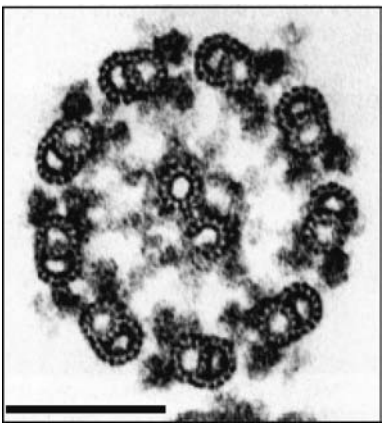

Fig. 1 Electron micrograph of a cross section of the flagellum from *Chlamydomonas*. Bar represents 100 nm (permission from Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2002) *Molecular Biology of the Cell*, 4th edition, Garland Science, New York).

has been preserved for almost all eukaryotic flagella and cilia (Figs. 1 and 2). This characteristic structure is often called "9 + 2". A sensory cilium is an exception, which lacks the two central microtubules, and so-called "9 + 0". The base structure known as the basal body also shows a complicated structure, composed of nine triplets of microtubules without the central pair (Fig. 3). The basal body is critically important for constructing the axoneme. *Chlamydomonas*, a green alga, has two motile flagella at the anterior end of the cell. Recent studies especially on this alga have explored its huge number of constituent molecules and have begun to reveal the growth mechanism of the flagella and cilia.

1.1 The Axoneme

The axoneme has a huge complex structure composed of ~250 different proteins, which include cytoskeletal proteins, motor proteins, molecular chaperons, calcium-binding proteins, protein kinases, and so on. However, more than half of them have not yet been characterized. In general, the component proteins are well conserved among the flagella of protozoan and metazoan sperm, except for some components. Figures 1 and 2 shows a typical electron micrograph and a schematic diagram of a cross section of the axoneme.

1.1.1 Microtubule

Microtubules form the core 9 + 2 structure in the axoneme. Nine outer doublet microtubules are arranged in a ring around the central two microtubules (Figs. 1 and 2). Each of the outer doublets is made of a complete tube called *A-tubule* and a partial tube called *B-tubule*, fused together. This characteristic design has been unchanged throughout protozoan to mammals. The structural unit of microtubule is a hetero dimer of α- and β-tubulin, which share ~40% sequence homology. α- and β-tubulin are evolutionally well-conserved molecules comprising ~450 residues. The atomic model of the dimer was first determined by electron crystallography on zinc-induced two-dimensional sheets, in which the dimers form a regular array.

α- and β-tubulin show very similar folds, having two β-sheets made of 4 and 6 strands, flanked by 12 α-helices. They have a typical Rossman fold for nucleotide binding at their N-terminus,

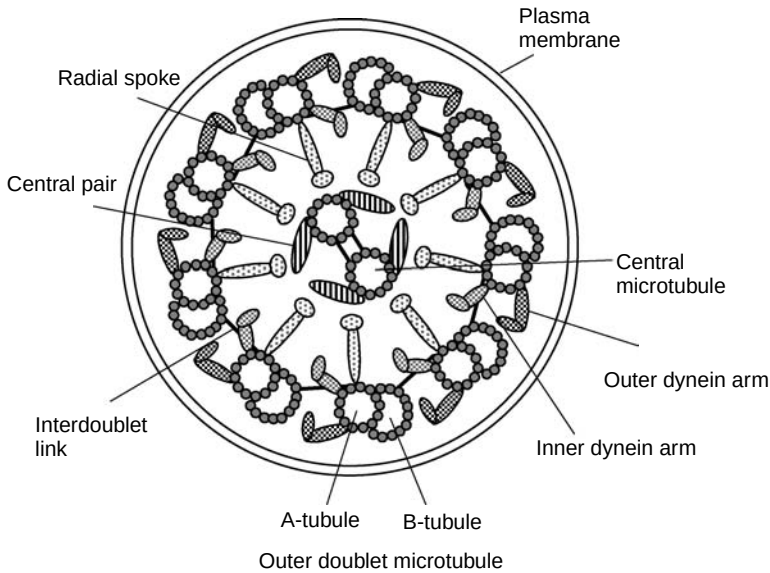

Fig. 2 Cartoon of the cross section of the flagellum or the cilium shown in Fig. 1.

where unexchangeable GTP is tightly bound in α-tubulin and exchangeable GTP or GDP is bound in β-tubulin. The dimers form a protofilament by connecting α-tubulin of one dimer to β-tubulin of another. The nucleotide binding sites of β-tubulin is placed at the interface of the dimers along the protofilament. A-tubule is a cylindrical hollow structure built of 13 protofilaments, while B-tubule is made of 11 protofilaments. Because all the protofilaments are oriented in parallel, one end of microtubule is occupied by all α-tubulin and the other end by all β-tubulin. The end made of α-tubulin is called the *minus end* and that made of β-tubulin is called the *plus end*. Microtubule assembles or disassembles ~100 times faster from the plus end than the minus end. Whether the β-tubulin molecules at the plus end bind GTP or GDP affects the assembly and disassembly; GTP promotes assembly, while GDP promotes disassembly. In general, the GTP-bound filament becomes straight and stiff, while the GDP-bound one tends to be curved and unstable.

In the zinc-induced two-dimensional sheets, the protofilaments are arranged in an antiparallel fashion. Hence, details of the lateral interactions between the protofilaments in microtubule are not yet clear, although the same kinds of molecules are known to make major lateral contacts, such as $\alpha-\alpha$ or $\beta-\beta$. α- and β-tubulin form a complete microtubule *in vitro*, while they never form a doublet microtubule, as found in the axoneme. This requires the basal body and other associated proteins described below.

1.1.2 Axonemal Dynein

Dynein is a microtubule-specific motor protein complex that consumes ATP for force generation. The dynein family is divided into two major groups: cytoplasmic dynein and axonemal dynein. Axonemal dynein is a huge molecular complex with a molecular mass of ~2 MDa and is involved

in the motility of eukaryotic flagella and cilia. In the axoneme, a pair of dynein "arms" with a dimension of ~25 nm are projected from A-tubule to form bridges with B-tubule of the adjacent outer doublet microtubule (Fig. 2). They are placed at a regular interval of ~24 nm along A-tubule. The arms projected inside and outside the ring of the microtubule doublet are called the *inner* and *outer dynein arms* respectively. When the motor is activated, the dynein molecules try to "walk" along the adjacent B-tubule just as myosin molecules slide along the actin filament on muscle contraction. However, owing to the interdoublet links (see below) among the outer doublets, the walking motion is converted to bending of the cilium or the flagellum.

The outer dynein arm consists of two heavy chains, three to five intermediate chains, and six light chains in metazoan sperm, and is made of three heavy chains, two intermediate chains, and eight light chains in *Chlamydomonas*. Dynein heavy chains are huge proteins with molecular masses of ~500 kDa, and act as the fundamental motor unit. They contain a globular head with ATPase activity and a microtubule-binding stalk. The globular head shows a ringlike structure made of six subdomains, each of which corresponds to AAA (ATPase Associated with various cellular activity) domains. Four of the six AAA domains have P-loop for ATP binding. There is a coiled-coil domain between the 4th and 5th AAA domains, which is thought to be essential for ATP-sensitive binding to the adjacent B-tubule. Domain between each AAA domain seems to be important for conformational changes necessary for the dynein power stroke. Dynein intermediate chains are 120- to 60-kDa proteins, and are presumably necessary for binding on A-tubule. Dynein light chains are 30- to 8-kDa proteins, and might be a regulator for activation of the outer dynein arm.

The inner dynein arm has more complicated molecular compositions than the outer dynein arm. It contains multiple molecular species with more heavy chains.

1.1.3 The Radial Spoke and the Central Pair

The radial spoke and the central pair (Fig. 2) are composed of a large number of proteins as well, and some of them include a domain that anchors cAMP-dependent protein kinase and a calcium binding domain. Therefore, the radial spoke presumably regulates the inner dynein arm through phosphorylation/dephosphorylation and calcium-dependent changes in the flagellar waveform. The central pair probably determines the plane of bending of the flagellum by sending signals to the radial spoke.

1.1.4 Other Proteins Associated with the Axoneme Structure

The interdoublet link (Fig. 2) known as the "nexin" linkage is thought to hold the outer doublet microtubules together and prevent their mutual sliding by the power stroke of the dynein arms.

Fibrous proteins called *tektin* presumably connect with the interdoublet links and support the ninefold symmetry architecture of the axoneme. Chaperons homologous to HSP70 are associated with this fiber. Also, another chaperon, HSP40, is identified as an axonemal component. These chaperons might help the folding of newly transported proteins through the intraflagellar transport described below.

Proteins related to determine a regular 24-nm interval of the outer dynein arms have also been identified and are called the *outer dynein arm docking complex*. They

are predicted to have a coiled-coil structure and may work as a molecular scale.

1.2 The Basal Body

In the root of the cilia and flagella, there is a cylindrical structure called *the basal body*. The basal body has another characteristic structure constructed from nine sets of triplet microtubules (Fig. 3). The triplets are tilted inward and are connected to the neighbors as well as to the radial spokes extended from the central core, forming a pinwheel-like structure. Each triplet comprises one complete microtubule (A-tubule) fused to two partial microtubules (B- and C-tubule). The basal body has the same structure as centrioles, which are embedded in the center of mitotic spindles known as *centrosomes*. Indeed, the basal body of sperms and flagella is converted to the centriole during fertilization in mammals and mitosis in *Chlamydomonas* respectively. The basal body and the centriole are duplicated from its origin so as to form a smaller daughter perpendicular to the origin. In *Chlamydomonas*, specialized microtubule bundles called *rootlet microtubules* are known to be involved in positioning the daughters, but little is known about the replication mechanism.

The basal body is essential for the assembly of the cilia and the flagella. It not only serves as a template for organization of the nine doublet microtubules of the axoneme but also incorporates huge numbers of proteins involved in the intraflagellar transport (see below). A recent electron tomographic study revealed various cross sections of the basal body along the central axis and transition zones to the axoneme (Fig. 4). The basal body starts from an amorphous disk at its proximal end (Fig. 4-1) and then continues to the pinwheel-like structure made of the nine triplets (2 and 3). It extends to ~40 nm in length, and then becomes the distal transitional fibers with striated structures, which end on the plasma membrane (4). There are stellate fibers in the transition zone associated with the loss of the C-tubule (5 and 6). At the end of the stellate fibers (7 and 8), the central pair of microtubules appears along with additional axonemal structures (9).

The basal body assembly also needs over 200 distinct proteins, whose roles are being characterized. A member of the tubulin

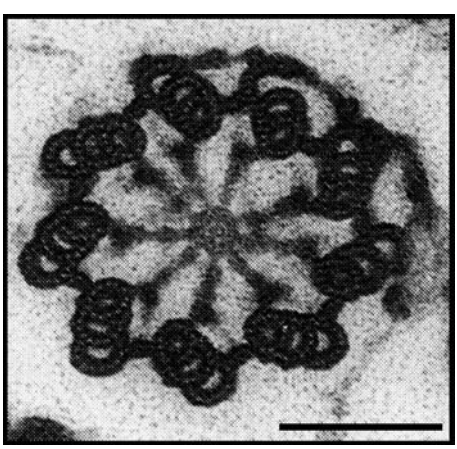

Fig. 3 Electron micrograph of a cross section of the basal body from a protozoan. Bar represents 100 nm (courtesy of Linck, R.W. and Woodrum, D.T. and permission from Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2002) *Molecular Biology of the Cell*, 4th edition, Garland Science, New York).

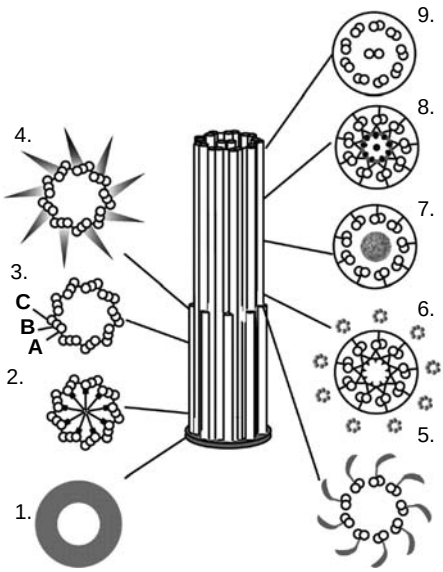

Fig. 4 Structural organization of the basal body, transition zone, and flagellum. Cross sections (1–9) are displayed from proximal to distal, respectively. Permission from O'Toole, E.T., Giddings, T.H., McIntosh, J.R., Dutcher, S.K. (2003) Three-dimensional organization of basal bodies from wild-type and δ-tubulin deletion strains of *Chlamydomonas reinhardtii*, *Mol. Biol. Cell.* **14**, 2999–3012.

super family, γ-tubulin, which is ~30% identical to α- and β-tubulin, is involved as an essential factor. Microtubules are nucleated at their minus end, but the concentrations of α- and β-tubulin inside the cells are below those required for spontaneous nucleation *in vitro*. Therefore, in all eukaryotes, the assembly of microtubules requires γ-tubulin. γ-tubulin is attached at the minus end of microtubules, forming a microtubule-organizing center (MTOC) together with other associated proteins. The basal body also includes other members of the tubulin super family, δ-, ε-, ζ-, and η-tubulin, which are probably required for construction of the basal body and formation of the axoneme.

1.3 The Intraflagellar Transport

The eukaryotic flagella and cilia grow at their distal tip. Components synthesized in the cytoplasm are transported through a system known as the *intraflagellar transport* (IFT). IFT was identified only a decade ago as particles rapidly moving (2–4 μm/s) within the flagella of *Chlamydomonas*. This movement has been recognized to be essential for the growth, maintenance, and shortening of the cilia and flagella in all eukaryotes.

IFT consists of bidirectional movement of large protein particles along the outer doublet microtubules. Particles called the *IFT particles* are transported to the distal end (anterograde IFT) at ~2 μm s^{-1} by a motor protein, kinesin II, and back to the cell (retrograde IFT) at ~4 μm s^{-1} by a cytoplasmic dynein. Kinesin II is a member of the kinesin super family, which, in general, proceeds toward the plus end of microtubule, but is unrelated to dynein, which proceeds toward the minus end. The IFT particles work as a raft to carry components of the axoneme and form large linear arrays up to several hundred nanometers long. They can be divided into two complexes known as "A and B". Complex A is composed of four to six subunits with an apparent molecular

mass of ~550 kDa, and complex B is made of at least nine subunits with a molecular mass of ~750 kDa. Complex A and B seem to be associated with retrograde IFT and anterograde IFT respectively.

IFT never stops and keeps moving at the same speed when the flagellum reaches its full length. Also, the IFT particles are remodeled at the base and the distal tip of the flagellum. Therefore, cells could regulate the assembly or disassembly of the flagellum by changing loads of the IFT particles at the base or the tip.

1.4 Perspectives

The eukaryotic flagella and cilia show a characteristic structure made of microtubules and ~250 associated proteins, half of which have not been characterized yet. Different from the bacterial flagella (see below), they have adopted a widely used system, the dynein motor for the force generation. Microtubule-dependent motors are also used for the process known as *IFT*, for growth, maintenance, and shortening of the cilia and flagella. Extensive studies, especially on the *Chlamydomonas* flagella, have just started to identify the roles of each protein and the growth mechanism. However, there are still too many unknowns. High-resolution structural information would be essential for further understanding of this system, and electron microscopy would be one of the most powerful tools to explore the structures and mechanisms of this complicated organelle.

2 Bacterial Flagella

Many bacteria swim by rotating helical flagellar filaments, which grow up to ~15 μm. The motor at the base of the filament drives the rotation of this helical propeller at hundreds of revolutions per second. In *Escherichia coli* and *Salmonella*, proton flow through the cytoplasmic membrane is utilized for the rotation. For chemotaxis and thermotaxis, the swimming pattern of bacteria alternates between "run" and "tumble"; a run lasts for a few seconds and a tumble for a fraction of second. During a run, the motor rotates counterclockwise (as it is viewed from outside the cell), and several flagellar filaments with a left-handed helical shape form a bundle and propel the cell. A tumble is caused by quick reversal of the motor to clockwise rotation, which produces a twisting force that transforms the left-handed helical form of the filament into a right-handed one, causing the bundle to fall apart rapidly. The separated filaments act in an uncoordinated way to generate forces that change the orientation of the cell. This leads to the tactic behavior, moving toward favorable or away from unfavorable conditions. Thus, the structure of the flagellar filament and its dynamic properties play an essential role in bacterial taxis.

The bacterial flagellum consists of several substructures. The motor structure called the *flagellar basal body* crosses both cytoplasmic and outer membrane, and continues to the extracellular structure called the *hook* and the *filament*. The largest of all is the filament, which is only ~20 nm in diameter, but grows up to a length around 15 μm by self-assembly of as many as 20 000 to 30 000 flagellin subunits. The assembly of the flagellum proceeds in a highly regulated fashion by the type III flagellar protein export apparatus, and flagellar axial proteins are always incorporated into its distal end, where capping proteins

help the assembly in most cases. The bacterial flagella are best understood in terms of the structure and function for those of *E. coli* and *Salmonella typhimurium*, because much of relevant work has been carried out using them. Overall structure of the bacterial flagellum is shown in Fig. 5.

2.1 Component Structures

2.1.1 The Basal Body

The basal body of the bacterial flagellum consists of the MS ring (also called *FliF ring*), the C ring, the rod, and the L–P ring.

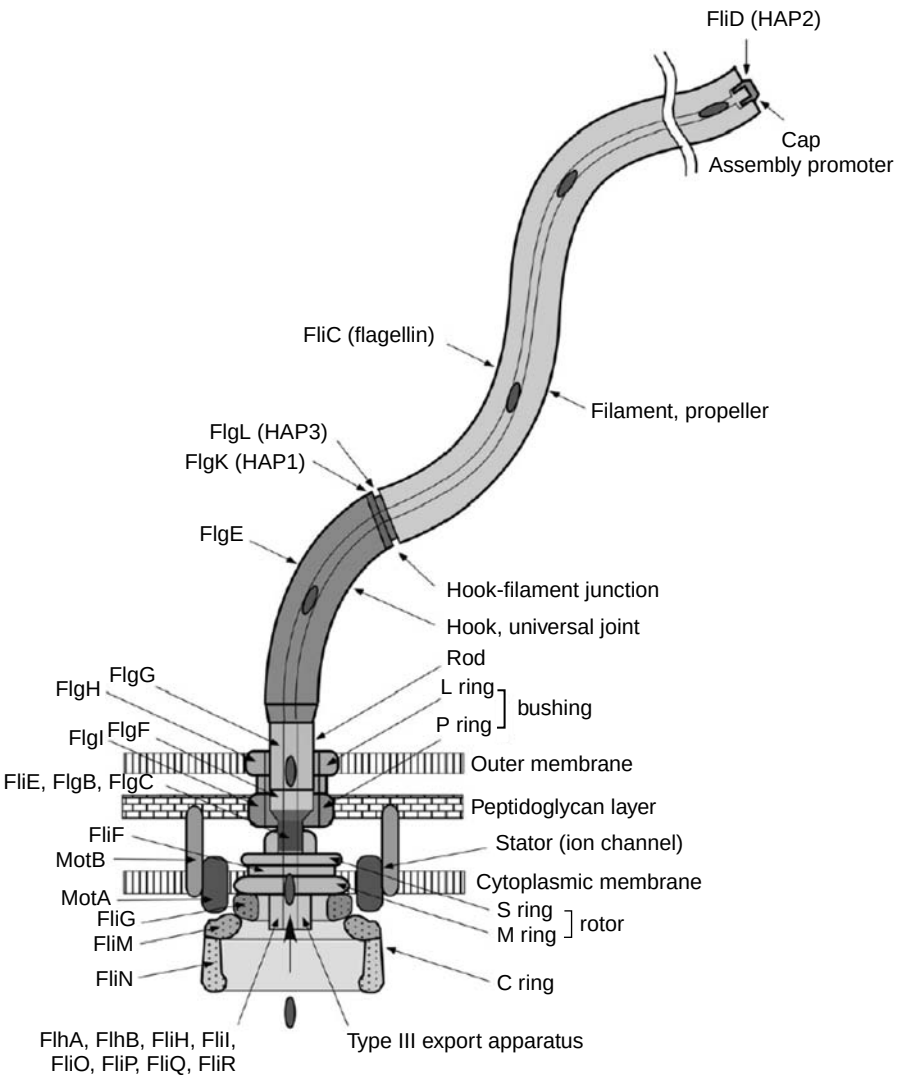

Fig. 5 Schematic diagrams showing the architecture of the bacterial flagellum. Reproduced with permission from Namba, K., Vonderviszt, F. (1997) Molecular architecture of bacterial flagellum, *Q. Rev. Biophys.* **30**, 1–65.

The MS ring with a diameter of ~250 Å is made of ~26 copies of FliF protein arranged cylindrically in the cytoplasmic membrane. "MS" means membrane and supramembrane. The C ring, a large cup-like structure attached on the cytoplasmic surface of the MS ring, is made of FliG, FliM, and FliN. "C" stands for cytoplasm. The rod, which traverses through the periplasmic space and the outer membrane, is made of FliE, FlgB, FlgC, FlgF, and FlgG proteins and is attached to the periplasmic face of the MS ring. The L–P ring is made of FlgH and FlgI, which are located within the outer membrane and the peptidoglycan layer respectively. L–P means "lipopolysaccharide" and "peptidoglycan." The MS ring is the base for self-assembly of all the flagellar proteins to form the flagellum and is partly involved in the type III flagellar protein export mechanism. The C ring is involved in torque generation as well as in regulation of its direction. The rod transmits the motor torque to the extracellular structures such as the hook and filament. The L–P ring is a bushing that mechanically supports the rapid rotation of the flagellum.

2.1.2 The Motor

Two cytoplasmic membrane proteins, MotA and MotB, are responsible for torque generation of the flagellar motor. MotA and MotB have four and one transmembrane-spanning helices respectively, and together form a proton channel. The MotA/B complex also functions as the stator by being attached to the peptidoglycan layer of the cell through a binding motif in the periplasmic domain of MotB. Electron micrographs of freeze-fractured samples showed that the stator complexes appear as a circular array of particles on the cytoplasmic membrane, surrounding the MS ring.

Some bacteria, such as alkaliphilic *Bacillus* and *Vibrio* species, normally have only one flagellum at their cellular pole, and use the electrochemical potential of sodium ions to generate the torque. The maximum rotational speed of the sodium-driven motor is ~1700 Hz, while that of the proton-driven motor is ~300 Hz. Four proteins, PomA (POlar flagellar Motility A), PomB, MotX, and MotY have been identified and are shown to be necessary for the rotation of the polar flagellar motor of *Vibrio*, such as *V. alginolyticus*, *V. parahaemolyticus*, and *V. cholerae*. PomA and PomB are homologous to MotA and MotB, while MotX and MotY are unique to the sodium-driven motor of *Vibrio*. Although the functions of MotX and MotY are not yet clear, a recent study showed that they are located in the outer membrane, and suggested that they may be essential for sodium recognition and for fast rotation of the polar flagellum.

For torque generation, MotA/B and PomA/B are thought to undergo a conformational change coupled to the proton flow and sodium flow, respectively while they interact with FliG, which is a component of the C ring. The X-ray crystal structures of the C-terminal and middle domains of FliG have revealed that important charged residues are lined up on one side of FliG so as to interact with a cytoplasmic loop of MotA or PomA. The functional unit of the motor complex is thought to be 4 MotA/2 MotB or 4 PomA/2 PomB.

2.1.3 The Hook and the Hook-filament Junction

The hook is a tubular structure, continuing directly from the rod into the outside of the cell. It is ~55 nm long and ~180 Å wide, and is made of ~120 copies of

FlgE, which is also called the *hook protein*. The hook and the filament are connected with a short segment composed of two proteins, FlgK (also called *hook-associated protein 1* or HAP1) and FlgL (HAP3). Because the hook shows a relatively high flexibility in bending, it probably works as a universal joint to transmit the torque to the filament in a wide range of its orientation, thereby allowing several filaments to form a bundle behind the cell during run as well as to rotate individually during tumble.

2.1.4 The Flagellar Filament

The flagellar filament is a helical assembly of a single protein, flagellin, with roughly 5.5 subunits per one turn of the 1-start helix. The filament can also be described as a tubular structure comprising 11 protofilaments, which are nearly longitudinal helical arrays of subunits. Left- and right-handed helical forms of the filament are produced by supercoiling caused by mixture of two distinct protofilament conformations, L- and R-type. When all of the 11 protofilaments are the same type, two types of straight filaments with distinct helical symmetries are formed: The longitudinal 11-start helix is left-handed in the L-type and right-handed in the R-type straight filament.

The structures at ~10 Å resolution by earlier studies using electron microscopy and X-ray fiber diffraction revealed the subunit shape, domain organization, and subunit packing in these two types of straight filaments. However, higher resolution was needed to understand the structural mechanisms of the filament formation and supercoiling in atomic detail. Flagellin has a strong tendency to polymerize into filaments and this has prevented its crystallization. By clipping off terminal chains for ~100 residues in total, a fragment named "F41" was crystallized, and its structure was solved at 2.0 Å resolution by X-ray crystallography. Recently, the complete atomic model of the R-type straight flagellar filament was built on a density map at ~4.0 Å resolution obtained by electron cryomicroscopy and helical image reconstruction.

Flagellin from *S. typhimurium* SJW1103 strain consists of 494 amino acid residues, and its molecular mass is 51.5 kDa. Flagellin in Fig. 6 is viewed perpendicular to the filament axis. The overall shape of flagellin looks like a Greek letter "Γ" with a vertical dimension of ~140 Å and a horizontal dimension of ~110 Å. Flagellin consists of four linearly connected domains labelled D0, D1, D2, and D3, which are arranged from inside to outside of the filament (Fig. 7a). The N-terminal chain starts from D0, going through D1, D2, and reaches D3, and then comes back through D2 and D1, and the C-terminal chain ends in D0. While all three domain connections are formed by pairs of short antiparallel chains, the one that connects domains D0 and D1 is longer than the other two, and therefore it is called the *spoke region*. The top and the bottom correspond to the distal and proximal side of the flagellum respectively. We also define the distal side "up" and proximal side "down," and describe the structure accordingly.

Domain D0 consists of the N- and C-terminal α-helices forming an α-helical coiled-coil structure (labelled ND0 and CD0 respectively in Fig. 6a). The spoke region consists of two chains (NS and CS), which connect to two α-helices in domain D1 (ND1a and CD1). Helix ND1a is followed by a loop connecting to the second, shorter α-helix (ND1b), which goes down, and the chain continues to two β-turns, a β-hairpin pointing down and an extended chain going up. The chain

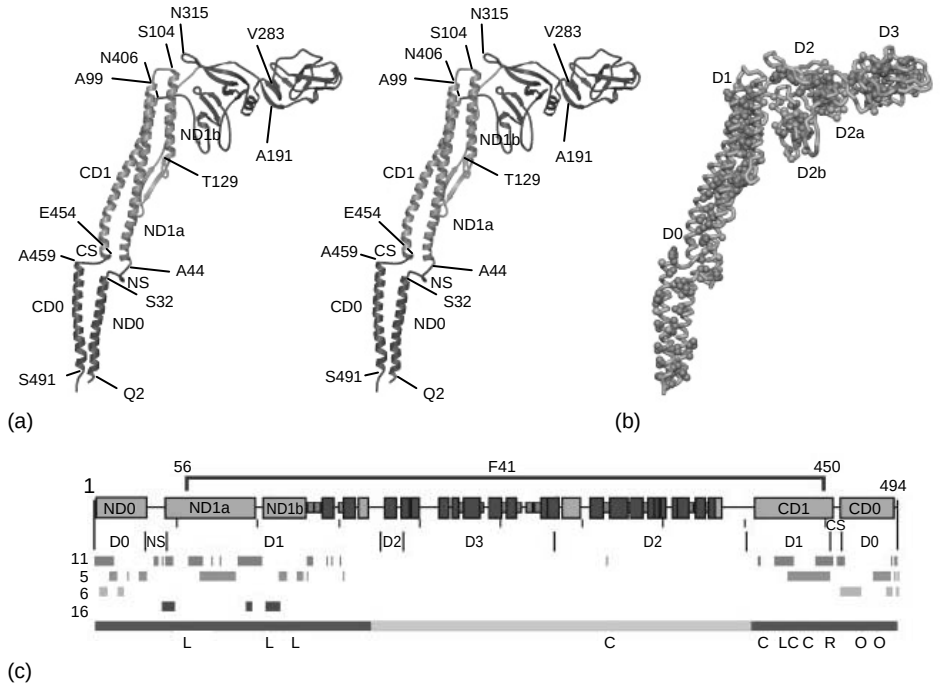

Fig. 6 The Cα backbone trace, hydrophobic side-chain distribution, and structural information of flagellin. (a) Stereo diagram of the Cα backbone. The chain is color coded as follows: residues 1 to 44, blue; 44 to 179, cyan; 179 to 406, green; 406 to 454, yellow; 454 to 494, red. (b) Distribution of hydrophobic side chains, mainly showing hydrophobic cores that define domains D0, D1, D2a, D2b, and D3. Side-chain atoms are color-coded: Ala and Met, yellow; Leu, Ile, and Val, orange; Phe, Tyr, and Pro, purple (carbon) and red (oxygen). (c) Position and region of various structural features in the amino acid sequence of flagellin. Shown are, from top to bottom, the atomic model of a major fragment of flagellin called F41 in blue; the secondary structure distribution with α-helix in yellow, β-structure in green, and β-turn in purple; tic mark at every 50th residue in blue; domains D0, D1, D2, and D3, and spoke regions NS and CS; the subunit contact regions along the 11-start in cyan, along the 5-start in orange, along the 6-start in pink, and along the 16-start in green; the well-conserved amino acid sequence in red and variable region in violet; point mutations that produce the filament of different supercoils. Letters at the bottom indicate the morphology of mutant filaments: L (F53V, D107E, R124A, R124S, G426A), L-type straight; R (A449V), R-type straight; C (D313Y, A414V, A427V, N433D) curly; O (Q472L, Q481L, Q481S) coiled. Permission from Yonekura, K., Maki-Yonekura, S., Namba, K. (2003) Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy, *Nature* **424**, 643–650. (See color plate. p. xxvii)

finally goes into the outer parts, domain D2 and D3, which are mainly composed of β-structures. The vertical dimensions of domain D0 and D1 are ~50 Å and 80 Å respectively. The extended chains in the spoke region are ~20 Å long.

The end-on view of the R-type straight filament from the distal end of the flagellum (Fig. 7a) shows the concentric double-tubular structure made of domains D0 and D1 in the densely packed filament core. Domains D2 and D3, which project

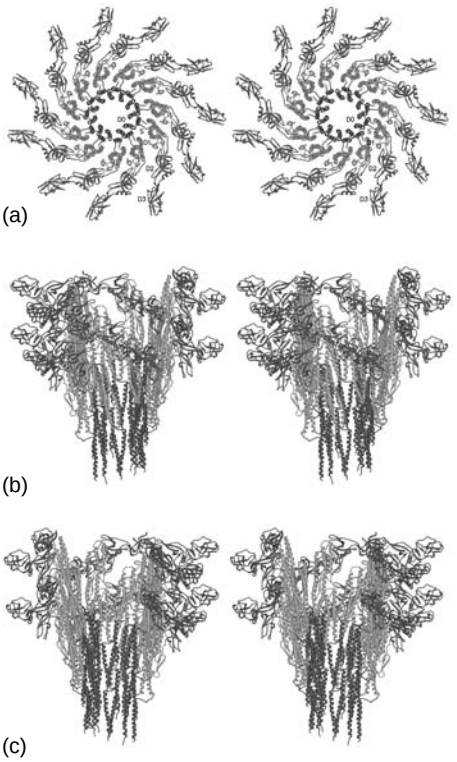

Fig. 7 Ribbon diagram of the Cα backbone of the filament model in stereo view. (a) End-on view from the distal end of the filament. Eleven subunits are displayed. (b) Side view from outside the filament. Three protofilaments on the far side have been removed for clarity. (c) Side view from inside the filament. Three protofilaments on the near side have been removed. Top and bottom of the side view images correspond to the distal and proximal ends of the filament respectively. The chain is color coded as in Fig. 6(a). Permission from Yonekura, K., Maki-Yonekura, S., Namba, K. (2003) Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy, *Nature* **424**, 643–650. (See color plate. p. xxiii)

out from the filament core, are relatively well separated from one another. The diameter of the filament is ~240 Å and that of the central channel is ~20 Å. The N- and C-terminal α-helices (ND0 and CD0) are radially arranged with ND0 outside and CD0 inside, exposed to the central channel. The chains connecting domains D0 and D1 (NS and CS) look exactly like radial spokes in this view. As shown in side views from outside and inside the filament (Figs. 7b and 7c), domains D0 and D1 are making intimate intersubunit interactions, both axially and laterally. As a polar assembly of flagellin having a largely asymmetric structure, the filament model shows the concaved feature at the distal end (top of Fig. 7c) and the pointed tip at the proximal end (bottom), which are observed even in electron micrographs of negatively stained filaments. The terminal chains of ~65 N-terminal and 45 C-terminal residues are unfolded in the monomeric form of flagellin in solution. It can be inferred from this structure that, in the absence of axial and lateral packing interactions of these terminal chains in the inner tube, the two-stranded α-helical coiled coil in domain D0 with relatively less extensive hydrophobic core and a pair of extended, rather flexible looking chains connecting to domain D1 would be highly unstable. In contrast, the upper two-thirds of domain D1 can form its compact tertiary structure in the monomeric form because of its extensive hydrophobic core formed by three α-helices and one β-hairpin. The outer parts of flagellin are variable so that bacteria can escape from the attacks of host immune system, while the core

part consisting of domains D0 and D1 are highly conserved (Fig. 6c).

The interactions within the inner and outer tubes are extensive, both along and between the protofilaments, while those between the inner and outer tubes are limited and are mainly in the spoke region. Most of the intersubunit interactions found within the outer tube are polar–polar or charge–polar, and contributions of hydrophobic interactions are relatively small, whereas those found within the inner tube and between the inner and outer tubes are mostly hydrophobic, contributing to the high stability of the filament structure.

There are heptad repeats of hydrophobic amino acid residues in the terminal regions of flagellin as well as other axial components, such as FlgE, HAP1, HAP3, and so on. In the filament structure, the terminal chains of neighboring subunits fold together into α-helical coiled coils. This interlocking organization would be the common motif by which the flagellar axial proteins form a continuous, mechanically stable tubular structure. The needle structures of the type III protein export system (see below) of pathogenic bacteria used to inject virulence proteins into host cells are genetically related to the bacterial flagellum, and are thought to have α-helical coiled coil as a common motif as well.

The atomic structure of the R-type straight filament has also revealed interactions of mutation sites responsible for polymorphic supercoiling of the filament. However, the true understanding of the mechanism of polymorphic supercoiling and its dynamic transition has to wait until the atomic model of the L-type straight filament becomes available.

2.1.5 The Filament Cap

The distal end of the bacterial flagellum shows a thin, flat plate structure as identified by electron microscopy. This flat plate is a part of the filament cap made of HAP2 subunits. HAP2 from *S. typhimurium* is composed of 466 amino acid residues, and its molecular mass is 49.8 kDa. HAP2 forms a decameric complex in solution. The decamer shows a pentameric feature in its end-on view and a bipolar structure with two thin plates on both edge connected by a pair of long, slightly tapered cylindrical walls in its side view. The connecting portion appears to be relatively flexible. Figure 8

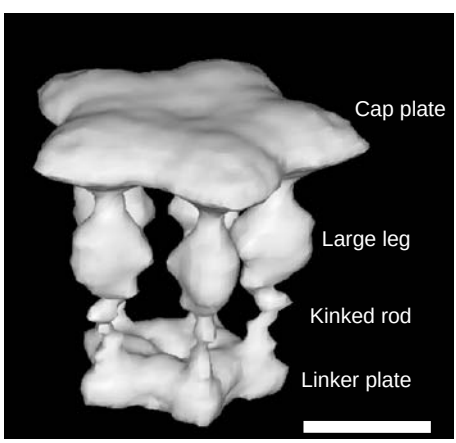

Fig. 8 Reconstructed three-dimensional image of the HAP2 pentamer. Bar represents 50 Å. Permission from Maki-Yonekura, S., Yonekura, K., Namba, K. (2003) Domain movements of HAP2 in the cap-filament complex formation and growth process of the bacterial flagellum, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15528–15533.

shows the three-dimensional structure of the HAP2 cap reconstructed from electron cryomicrographs. The HAP2 cap is composed of a pentagonal plate and five leg domains, whose lower part has no particular feature. This portion is made of both terminal regions of HAP2, because when ~40 N- and 50 C-terminal residues of HAP2 are cleaved off, the fragment does not form the decamer. These terminal chains are disordered in the pentameric or monomeric form of HAP2, suggesting that they also have the common structural motif shared by the flagellar axial proteins to form α-helical coiled coil in the inner core of the tubular structure. The heptad repeats are actually identified in the amino acid sequence of these terminal regions.

2.2 The Assembly Process of the Bacterial Flagellum

The assembly process of the bacterial flagellum starts from the formation of the MS ring complex in the cytoplasmic membrane. Then, the C ring is formed on the cytoplasmic face of the MS ring. The type III protein export apparatus, which is homologous to those that enable pathogenic bacteria to secret virulence factors, is presumably formed within the C ring. The export apparatus is composed of six membrane proteins and several cytoplasmic proteins (see below). The torque generation unit MotA/B assembles around the MS ring and is anchored to the peptidoglycan layer with the periplasmic domain of MotB. Each flagellum probably has about eight torque generation units.

The outward assembly of the flagellar axial structures proceeds at the distal end as it was initially found for the filament growth. The L–P ring assembly is an exception, where the subunit proteins FlgH and FlgI assemble around the rod in the outer membrane and peptidoglycan layer respectively. These proteins are exported by a general secretion system known as the *Sec pathway*. The formation of the P ring needs a periplasmic chaperone, FlgA. The type III protein export apparatus selectively binds and translocates flagellar axial proteins into the central channel of the growing flagellum by using the energy of ATP hydrolysis by FliI (see below). The axial proteins construct the rod, the hook, the hook-filament junction, and the long filament in this order. The proximal rod components FliE, FlgB, FlgC, and FlgF are connected with the periplasmic side of the MS ring, probably in this order, and the distal rod component FlgG follows to assemble. The rod assembly requires a help of the rod cap made of FlgJ, whose C-terminal domain has a muramidase activity to digest the peptidoglycan layer and to promote the rod growth through it. After the rod assembly is completed, the rod cap is somehow released, and FlgE assembles on the distal end of the rod in a helical manner to form the hook. A cap complex made of FlgD is attached at the distal end of the hook for its efficient assembly until it grows up to a length of ~55 nm. The length of the hook is controlled by FliK, one of the rod/hook-type export substrates, and FlhB, a member of the export apparatus (see below). A relatively short C-terminal region of FliK interacts with the cytoplasmic domain of FlhB at an appropriate timing to switch its conformation, resulting in the switching of substrate specificity of the export apparatus from the rod/hook type to the filament type (see below). This is how the hook growth stops, but the detailed mechanisms as to how the hook length information is transmitted to the cytoplasmic domains of the export apparatus is not yet well understood. Recent studies suggest

that FliK acts as a ruler, but an alternative hypothesis called the *measuring cup mechanism*, in which 120 copies of FlgE are measured by the C-ring, may also be valid.

Then, the hook cap falls off the tip of the hook, probably by the binding of FlgK (HAP1), and then FlgL (HAP3) and FliD (HAP2) are transported and bound at the distal end in this order. FlgK, FlgL, and FliD are also called hook-associated proteins, HAP1, HAP3, and HAP2 respectively, because the hook-HAP1-HAP3-HAP2 complex is formed momentarily before the initiation of flagellin assembly into the long helical filament. Flagellin, also called "FliC", travels a long way through the narrow tunnel, only ~20 Å in diameter, but up to 10 to 15 μm long, to the distal end of the growing flagellum, where it assembles with a helical symmetry into the mechanically stable filamentous structure to function as a helical propeller. The filament cap made of HAP2 is always attached at the distal end, promoting the efficient assembly process.

2.2.1 The Type III Protein Export Apparatus

Most of the flagellar axial proteins are exported by the type III protein export apparatus into the central channel and to the distal end. The flagellum-specific secretion apparatus is made of six membrane proteins (FlhA, FlhB, FliO, FliP, FliQ, and FliR) and two soluble proteins, FliH and FliI. FliI has an ATPase activity and acts as the engine of export.

In addition to those, there are a general chaperon, FliJ, and substrate-specific chaperons, FlgN, FliS, and FliT, which bind to the export proteins synthesized in the cytoplasm and prevent them from undesirable aggregation or digestion, and from premature assembly within the cell. The partners of the substrate-specific chaperons are as follows: FliS–FliC; FlgN–FlgK; FlgN–FlgL; and FliT–FlgD. The structure of a complex of FliS with a short fragment of flagellin was recently solved by X-ray crystallography. This fragment corresponds to the C-terminal ~40 residues, which forms helix CD0 in the filament (Figs. 6 and 7). In the crystal structure of the complex, this helix is bent in the middle and partly extended, so as to keep the hydrophobic residues away from solution. FliS would prevent flagellin from forming the filament in this manner.

The flagellar type III protein export apparatus also specifies the order of the export substrate proteins. A recent study shows that there are two clear modes; the rod-type (FliE, FlgB, FlgC, FlgF, and FlgG) and the hook-type (FlgD, FlgE, FliK) proteins are exported in one mode, and the filament-type proteins, including FlgK, FlgL, FliD, and FliC, are exported in the other. The export of rod/hook-type proteins is followed by the export of the filament-type, and the switching of the substrate specificity occurs at a well-defined point in the assembly process to control the hook length, which is important for the proper hook function as a universal joint.

2.2.2 Growth of the Flagellar Filament

The flagellar filament is extended from the cell. Possible ways to construct such structures are either that constituent molecules are inserted at the base or they are transported to the distal end and bound there. As mentioned above, the latter is adopted by the bacterial flagellar filament assembly, probably because the bonding between the basal body, the hook, and the flagellar filament must always be tight and stable against the mechanical stress from high-speed rotation and quick reversal of

the motor with the long helical propeller in viscous environments.

Flagellin subunits passing through the central channel of the flagellum polymerize just below the cap one after another to form the long helical filament. The cap stays attached at the distal end during the filament growth, and the simplest role of the cap would be to prevent flagellin from leaking out. This results in an increase in the local concentration of flagellin at the filament end above the critical concentration for polymerization. However, the cap does not completely cover the openings because HAP proteins that are occasionally transported pass through the capped end. Just in case the filament is broken in the middle, those transported HAP2 molecules can bind to the broken end of the filament and form a new cap again. Thus, the cap appears to act as a gatekeeper, which can be recovered when it is lost. The cap may also induce a conformational change of flagellin for efficient polymerization because flagellin must be unfolded to fit in the narrow channel during its transport. For insertion of flagellin subunits, the cap must move up by steps to make a room for the next flagellin monomer. Thus, the cap must play apparently contradictory roles; it must remain stably attached while permitting the insertion of flagellin subunits.

Structure analysis of the cap–filament complex by electron cryomicroscopy showed that the pentagonal plate of the cap is attached to the distal end of the filament at five positions through the anchor domains of the cap, which look just like five legs, where these anchor domains show significant deviation from the fivefold symmetry (Fig. 9). The five anchor domains are well separated from one another, forming five gaps with different sizes and shapes between the plate and the filament end (Fig. 9). One of the gaps (Fig. 9c-1) is distinctly larger than the other four, having an inverted "L" shape and the size equivalent to domain D1 of flagellin. Because of the axial stagger in the lateral subunit packing, the end of the filament is not flat, but has five indentations around its circumference, where one of them is a double indentation. Thus, this is most likely to be the site of flagellin assembly.

Inside the filament just below the cap plate, there is a cavity ~30 Å wider than the central channel of the filament. Folded flagellin molecules cannot pass through the narrow central channel with a diameter of ~20 Å. Hence, flagellin definitely needs to be considerably unfolded until it reaches the distal end. This cavity may play some role in helping flagellin to fold up, perhaps in a similar way to Anfinsen's cage in chaperonin. The cavity appears to have the right size to accommodate only one flagellin subunit at a time, allowing its refolding without aggregation with other subunits.

During the growth of the filament, a large number of flagellin molecules, together with a small number of the hook-associated proteins, HAP1, HAP2, and HAP3, are synthesized in the cell and selectively translocated by the export apparatus into the long, narrow central channel of the flagellum, through which they are transported to the distal end by diffusion. This narrow channel would prevent unfolded flagellar proteins from undesirable aggregation.

The inner surface of the channel consists of mainly polar amino acids with one positively charged residue. The polar nature of the surface may be advantageous for fast diffusion of unfolded proteins, because unfolded proteins have many hydrophobic side chains exposed, which would be trapped on the channel surface of hydrophobic nature. This would also

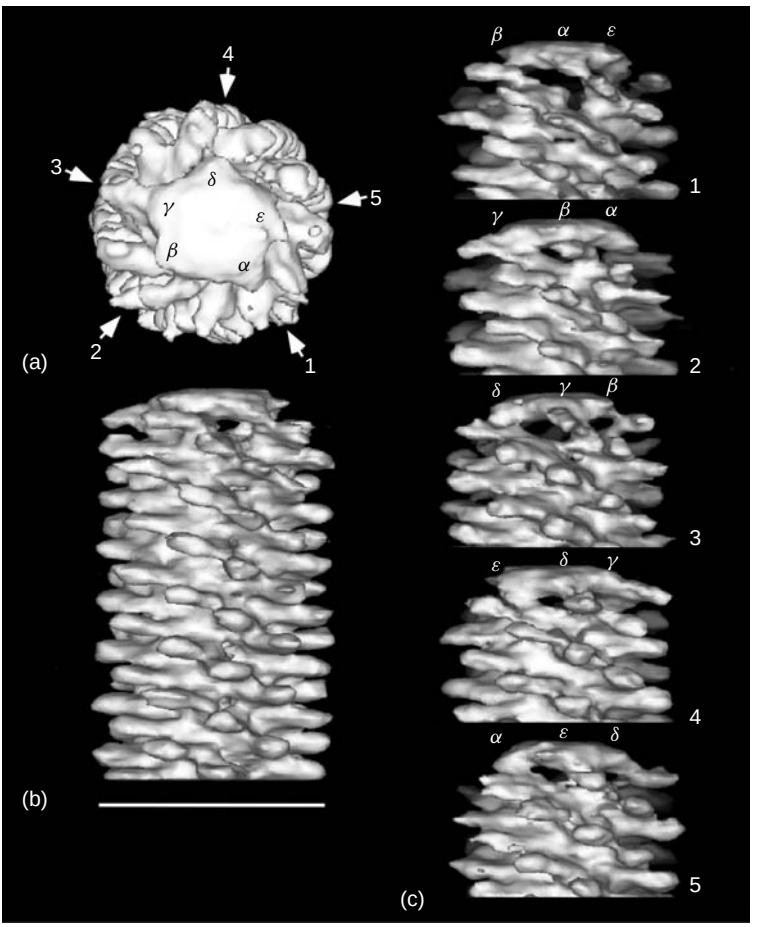

Fig. 9 Three-dimensional density map of the cap-filament complex and its central section. (a) End-on view from the top, showing a pentagonal shape of the plate domain of the cap. The five vertices of the pentagon are labeled with Greek letters, $\alpha - \varepsilon$, which are used to guide the orientation of the pentagon in the side views in (c). (b) Side view, showing a regular helical array of flagellin subunits and the plate domain of the cap. (c) Five side views, showing each of the five gaps between the plate and the filament end. The viewing directions are as indicated by arrows in (a), labeled with corresponding numbers. Greek letters labeling the vertices as in (a) also indicate the orientations of the pentagonal plate. Note that part of the gap in view one has a significantly larger axial extension than the other four gaps, and this is likely to be the site of assembly for the newly arriving flagellin subunit. Bar represents the diameter of the filament, 240 Å. Permission from Yonekura, K., Maki, S., Morgan, D.G., DeRosier, D.J., Vonderviszt, F., Imada, K., Namba, K. (2000) The bacterial flagellar cap as the rotary promoter of flagellin self-assembly, *Science* **290**, 2148–2152.

be a common feature of the type III protein export system including the needle complex of pathogenic bacteria.

Because of the symmetry mismatch between the helical subunit array of the filament with 11 protofilaments forming the tube and the pentameric annular structure of the cap, the five leg domains of the cap must have significantly different conformations from one another for interaction with flagellin. The flexible nature of the leg domain allows such polymorphic interactions, reducing conformational strains and permitting stable interactions between the two structures in spite of the symmetry mismatch. At the same time, the high conformational entropy of these cap anchor domains would counterbalance the strong binding of the cap to the filament as a whole for efficient insertion of flagellin.

Every insertion of a flagellin subunit is likely to force the cap to move into the next stable position energetically equivalent to the current position. A prediction from this structure and a previous discussion is that the cap moves up 4.7 Å and rotates 6.5° along the left-handed 5-start helix of the filament (Fig. 10). In terms of the cap conformation, the next position is approximately 65.5° away along the right-handed 1-start helix (arrangement of yellow subunits in Fig. 10). However, with permutation of the conformational states of the legs, which are 72° apart, a rotation of only 6.5° in the other direction would suffice. This would result in a roughly complete rotation of the cap by the assembly of 55 flagellin subunits. Upon every subunit incorporation, four legs of the cap rearrange their conformations and the last one changes its binding partner; namely, these legs walk along the helical steps of the filament end. These dynamic movements of the cap and its leg domains would be required for efficient promotion of the flagellin self-assembly.

The energy for the cap rotation with the conformational rearrangements is presumably supplied by the binding energy of newly incorporated flagellin subunit to domains of multiple subunits of flagellin and HAP2, where the binding occurs by concentration of flagellin within the channel. The energy to keep the effective concentration of flagellin at the distal end of the export channel above the critical concentration for its polymerization would therefore come from ATP hydrolysis by the type III protein export apparatus located at the cytoplasmic face of the flagellar basal body. The energy transduction network formed here is unique and interesting because the energy produced in the cytoplasmic space is transmitted quite a long distance (10 ~ 15 μm) through flagellin export to be used by the rotational movement of the cap at the end. The energy required for these movements would be relatively small because the process would involve a number of small conformational rearrangements in sequence, where the energy would propagate from one to the next.

2.3 Perspectives

The bacterial flagellum is a macromolecular complex made of ~25 different proteins and is used for bacterial swimming. Bacteria have evolved most of those various apparatuses for this specific purpose. Biochemical and genetic studies have identified functions of each protein and have played important roles in understanding the assembly process of the flagellum. Recent electron microscopy and X-ray crystallography have also been elucidating details of the components structures,

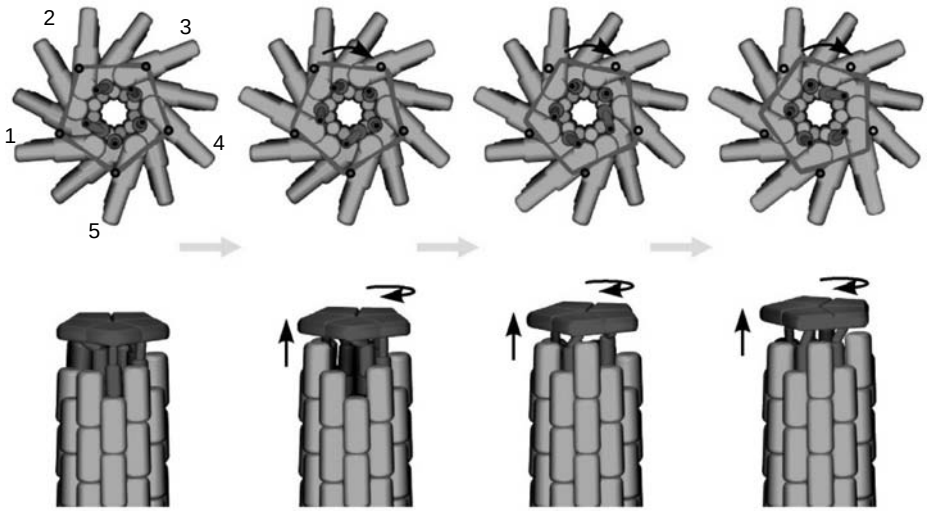

Fig. 10 Schematic diagram depicting the rotary cap mechanism promoting the flagellin assembly. This describes the rotation and axial translation of the cap plate and accompanied rearrangement of the legs upon every incorporation of a flagellin subunit (from left to right). Top view in the upper panel and an oblique view in the lower panel. In the upper panel, the cap plate is made transparent to show the different ways of leg domain binding, where black dots indicate the five positions of leg domain attachment to the plate. The plate and the black dots have strict fivefold symmetry, while the leg domains do not. In the lower panel, the outer domain of flagellin is removed for clarity. Subunits in yellow are newly incorporated flagellin molecules. Five open circles in the upper panel indicate the initial positions of the cap plate vertices as a reference for the cap rotation. The flagellin assembly proceeds along the 1-start helix, which is in the counterclockwise direction when viewed from the top, approximately at every 65.5° (360 × 2/11). This is also the angle of rotation after which the next binding site appears. However, because the legs of the cap are located every 72° (360/5), a 6.5° clockwise rotation with permutation of the leg conformations is sufficient to make the appropriate interactions between the leg and flagellin subunits. Numbers indicate the directions of views in Fig. 9(a). Animation at http://www.npn.jst.go.jp/yone.html. Permission from Yonekura, K., Maki, S., Morgan, D.G., DeRosier, D.J., Vonderviszt, F., Imada, K., Namba, K. (2000) The bacterial flagellar cap as the rotary promoter of flagellin self-assembly, *Science* **290**, 2148–2152. (See color plate. p. xxviii)

in particular, for the extracellular components. However, still little is known about molecular mechanisms of how each part of the flagellum functions. Obviously, we need high-resolution structures of not only the component proteins but also complexes made of many proteins. How the energy of ion flow is converted into the rotation of the flagellum would be the most important question to answer, and it still remains a mystery. High-resolution structure of the motor complex would be a starting point to solve the problem.

See also Bacterial Cell Culture Methods; Bacterial Growth and Division; Electron Microscopy in Cell Biology; Motor Proteins.

Bibliography

Books and Reviews

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2002) *Molecular Biology of the Cell*, 4th edition, Garland Science, New York.

Cole, D.G. (2003) The intraflagellar transport machinery of *Chlamydomonas reinhardtii*, *Traffic* **4**, 435–442.

Dutcher, S.K. (2003) Elucidation of basal body and centriole functions in *Chlamydomonas reinhardtii*, *Traffic* **4**, 443–451.

Inaba, K. (2003) Molecular architecture of the sperm flagella: molecules for motility and signaling, *Zoolog. Sci.* **20**, 1043–1056.

Minamino, T., Namba, K. (2004) Self-assembly and type III protein export of the bacterial flagellum, *J. Mol. Microbiol. Biotechnol.* **7**, 5–17.

Namba, K., Vonderviszt, F. (1997) Molecular architecture of bacterial flagellum, *Q. Rev. Biophys.* **30**, 1–65.

Snell, W.J., Pan, J., Wang, Q. (2004) Cilia and flagella revealed: from flagellar assembly in *Chlamydomonas* to human obesity disorders, *Cell* **117**, 693–697.

Yorimitsu, T., Homma, M. (2001) Na^+-driven flagellar motor of *Vibrio*, *Biochim. Biophys. Acta* **1505**, 82–93.

Primary Literature

Evdokimov, A.G., Phan, J., Tropea, J.E., Routzahn, K.M., Peters, H.K., Pokross, M., Waugh, D.S. (2003) Similar modes of polypeptide recognition by export chaperones in flagellar biosynthesis and type III secretion, *Nat. Struct. Biol.* **10**, 789–793.

Maki-Yonekura, S., Yonekura, K., Namba, K. (2003) Domain movements of HAP2 in the cap-filament complex formation and growth process of the bacterial flagellum, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15528–15533.

Nogales, E., Wolf, S.G., Downing, K.H. (1998) Structure of the $\alpha\beta$ tubulin dimer by electron crystallography, *Nature* **391**, 199–203.

O'Toole, E.T., Giddings, T.H., McIntosh, J.R., Dutcher, S.K. (2003) Three-dimensional organization of basal bodies from wild-type and δ-tubulin deletion strains of *Chlamydomonas reinhardtii*, *Mol. Biol. Cell.* **14**, 2999–3012.

Samatey, F.A., Imada, K., Nagashima, S., Vonderviszt, F., Kumasaka, T., Yamamoto, M., Namba, K. (2001) Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling, *Nature* **410**, 331–337.

Yonekura, K., Maki, S., Morgan, D.G., DeRosier, D.J., Vonderviszt, F., Imada, K., Namba, K. (2000) The bacterial flagellar cap as the rotary promoter of flagellin self-assembly, *Science* **290**, 2148–2152.

Yonekura, K., Maki-Yonekura, S., Namba, K. (2003) Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy, *Nature* **424**, 643–650.

Molecular Display Technologies

Ece Karatan, Zhaozhang Han, and Brian Kay
Argonne National Laboratory, Argonne, IL, USA

1 Display Methods 411
1.1 Phage display 411
1.1.1 Bacteriophage M13 411
1.1.2 Bacteriophage T7 414
1.1.3 Bacteriophage Lambda (λ) 414
1.2 Cell Surface Display 414
1.2.1 Bacterial Display 415
1.2.2 Yeast Display 415
1.3 In vitro Display 416
1.3.1 Ribosome Display 417
1.3.2 mRNA Display 419

2 Types of Libraries 420
2.1 Combinatorial Peptide Libraries 420
2.2 scFv and Fab Libraries and Fab Libraries 420
2.3 Antibody Mimetic Libraries 421
2.4 cDNA and Genomic DNA Libraries 422

3 Applications 422
3.1 Protein Function and Target Validation 422
3.1.1 Mapping Protein–protein Interactions with Peptides 422
3.1.2 Elucidation of Enzyme Function 423
3.1.3 Functional Knock-outs Using Intrabodies 424
3.2 Drug Discovery and Therapeutics 424
3.2.1 Identification of Tissue- and Tumor-specific Markers via Whole-cell and *in vivo* Panning 424
3.2.2 Screening of Small Molecule Inhibitors Using Surrogate Peptide Ligands 425
3.2.3 Zinc Fingers for Controlling Gene Expression 425

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

3.3 Protein Engineering and Industrial Applications 425
3.3.1 Molecular Evolution 425
3.3.2 Bacterial and Yeast Surface Display as Biocatalysts 426
3.3.3 Nanotechnology 426

4 Outlook and Perspectives 427

Acknowledgments 427

Bibliography 427
Books and Reviews 427
Primary Literature 427

Keywords

Antibody Mimetic
A protein engineered to act as an antibody such that it binds target molecules other than its natural binding partners.

Affinity Reagent
An antibody, antibody mimetic or a peptide that has an affinity for a molecule of interest.

Bacteriophage/phage
A virus that infects bacteria.

Binder
Same as "affinity reagent."

CDR
Complementarity Determining Regions. Highly variable regions of the antibody light and heavy chain variable domains that are responsible for antigen recognition.

Fab
Fragment of antigen binding of antibodies generated by either recombinant methods or limited proteolysis.

scFv
Single-chain Fragment of variable region. A short fragment of an antibody molecule that is engineered in such a way that it contains only the variable regions of the light and heavy chains in a single polypeptide.

Molecular display refers to the presentation or the display of molecules – mostly proteins and peptides – on the surface of a virus, a cell, or a molecular complex such as a ribosome or messenger RNA. The coding information of the displayed molecule is carried by the entity displaying it, thereby establishing a physical link between the properties (phenotype) and DNA sequence (genotype) of the displayed molecule. All molecular display techniques have *in vivo* or *in vitro* methods for enrichment by selective propagation of the molecule with the desired properties. This, along with the genotype–phenotype link, allows large numbers (up to 10^{14}) of variations to be surveyed for the desired characteristics. These properties have made molecular display technologies a very valuable tool for the study of molecular interactions, the isolation of affinity reagents with increased affinities and stabilities, and the development of therapeutics with tissue- and cell-targeting capabilities. This article introduces different display methods, the types of random libraries most commonly used to isolate binders, and some of the interesting and creative applications of molecular display technologies in answering a variety of biological questions.

1 Display Methods

1.1 Phage display

In 1985, Dr George Smith, a professor at the University of Missouri-Columbia, published an article in which he described an experiment where he inserted a segment of the gene encoding the *Eco*RI restriction enzyme into the genome of the M13 bacteriophage. Smith found that fusions of the *Eco*RI protein fragment with one of the minor coat proteins were tolerated by the virus, as they still produced infectious particles. Interestingly, these particles incorporated the chimeric protein, and they could be fractionated from nonrecombinant virus particles with an antibody that recognized the expressed portion of the *Eco*RI protein. With sequential rounds of affinity selection, termed *biopanning*, the recombinant particles that displayed the antigen could be selectively enriched from an excess of nonrecombinant particles. These results led to the concept of "phage display," in which coding regions for peptide or protein segments are inserted in a bacteriophage genome and the expressed peptide or protein segments are accessible for molecular interactions. The major benefit of phage display, as well as other display technologies we review here, is the physical linkage between the displayed entity (phenotype) and its genetic code (genotype), which permits a method for "bar-coding" and isolating rare peptides and proteins with desired properties from diverse libraries. In this section, we briefly discuss different molecular display methods that are in use today.

1.1.1 Bacteriophage M13

To date, bacteriophage M13 display has been the most extensively used display method. M13 phage particle are filamentous in shape (i.e. 9 nm wide, 900 nm long) and contain five types of coat proteins as well as a single-stranded DNA molecule (Fig. 1a). In the original experiments of

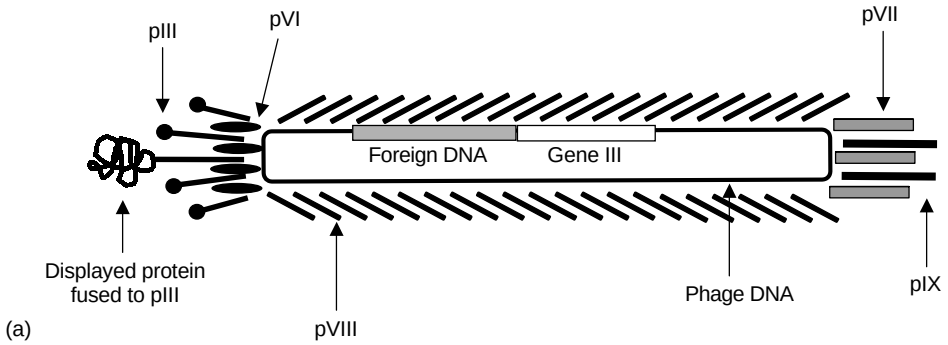

Fig. 1 Bacteriophage M13 display.
(a) Schematic representation of M13. (b) Affinity selection of phage-displayed libraries. The protein (pIII, pVI, pVII, pVIII, and pIX) and DNA (single-stranded) components of a prototypic bacteriophage M13 particle are shown. DNA fusions to the N-terminus of the coding region for mature protein III leads to the generation of a chimeric protein that is incorporated in the secreted viral particle. If the virus particle is assembled in an infected cell that carries a recombinant phagemid and a helper virus, display can be monovalent (shown); alternatively, use of a different vector can lead to pentavalent display (not shown). A population of virus particles displaying a library of antibody fragments, cDNA segments, combinatorial peptides, or engineered proteins can be fractionated by affinity selection. Phage are incubated in a microtiter plate well coated with a target protein and nonbinding particles are washed away, and retained particles are recovered by chemically denaturing the target. Infection of *E. coli* cells with the released particles amplifies the phage enriched for binding and the process is repeated two or more additional times. After limiting dilution of the final sample, individual clones are amplified and tested for binding to the target (and not a negative control protein) by an ELISA. DNA sequencing of the phage insert leads to identification of the affinity selected antibody fragment, cDNA segment, combinatorial peptide, or engineered protein.

Smith, protein fragments were displayed at the N-terminus of mature protein III. Fortuitously, the site chosen for display did not interfere with either virus morphogenesis or the infection process. Since then, research has shown that it is possible to display at the N-terminus of proteins III, VII, VIII, IX and at the C-terminus of proteins III, VI, and VIII. The display of proteins and peptides at the N-terminus of proteins III or VIII is most common.

Phage-displayed protein or peptide libraries are typically fractionated by affinity selection (Fig. 1b). A target protein is immobilized on a surface, incubated with an aliquot of viral particles, nonbinding particles are washed away, and bound particles are recovered. To release the bound particles without destroying their infectivity, it is possible to expose them to pH extremes (i.e. pH 2 or 12), protein denaturants (i.e. 6 M urea, 15 mM dithiothreitol), or proteases (i.e. trypsin, subtilase). Once the conditions for elution have been neutralized, the recovered phage can be used to infect *Escherichia coli* cells, thereby amplifying the selection output. Since bacteriophage M13 is not a lytic virus, infected cells secrete phage particles on a continual basis as they grow, yielding $\sim 10^{12}$ plaque-forming units (pfu) per milliliter in an overnight liquid culture.

The target proteins can be affixed to the surface of microtiter plate wells either directly by nonspecific adsorption or indirectly through a streptavidin-biotin

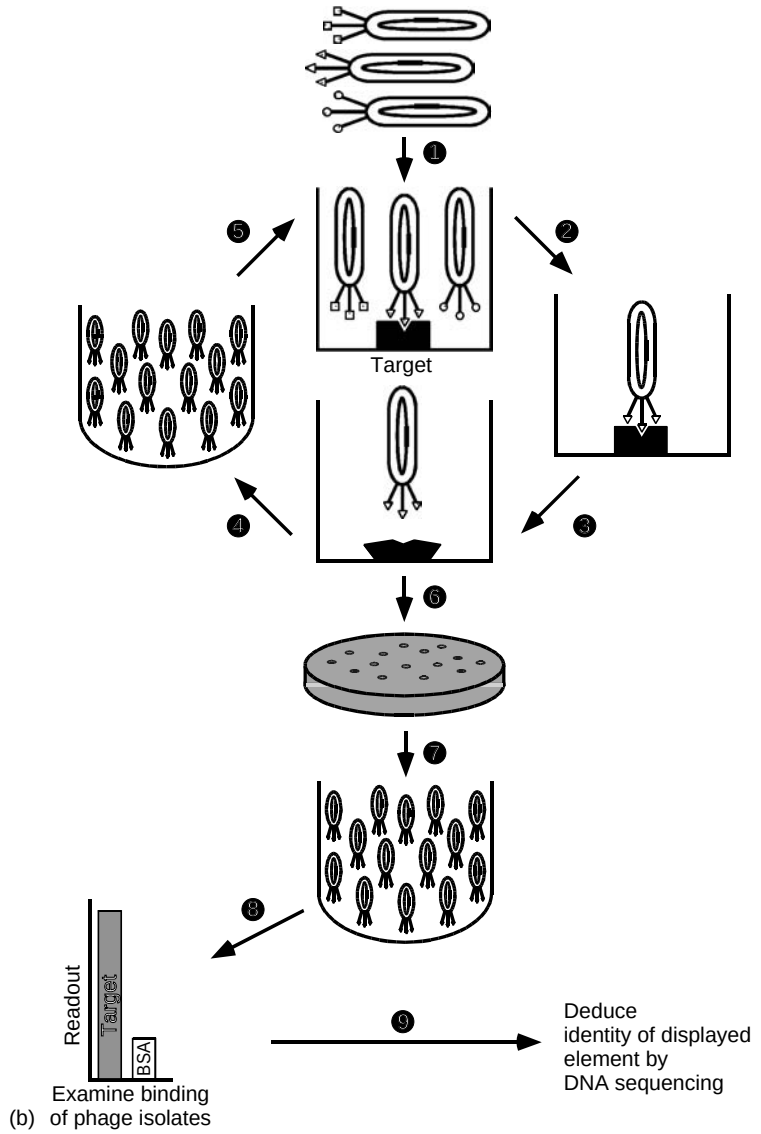

Fig. 1 (*Continued*)

linkage (i.e. the target protein is biotinylated and bound to immobilized streptavidin). Another option is to capture complexes formed between soluble biotinylated protein targets and phage on streptavidin-coated microtiter plate wells or magnetic beads. Finally, it is possible to select phages that bind to the surfaces of, or are internalized by, mammalian cells or tissues. In general, three rounds of affinity selection are sufficient to enrich binding phage particles from a library. After

confirming that the output phage clones actually bind to the target, often through an enzyme-linked immunosorbent assay (ELISA), sequencing the DNA insert of the individual phage clones reveals the coding sequence of the displayed peptide or protein.

Many of the proteins displayed on bacteriophage M13 adopt their native structure and activity, even though they are fused to a coat protein that is assembled into a viral particle. However, in some cases, the displayed protein interferes with the function of the coat protein, viral particle morphogenesis, or secretion of the viral particle by *E. coli*. In many of these situations, these limitations can be overcome by displaying the protein on a shortened version of the coat protein and providing the wild-type form of the protein, with the aid of a helper virus. Thus, instead of being displayed on all five copies of protein III (multivalent display), the protein can be displayed in a monovalent manner, with four copies of wild-type protein III available for attaching to the pili of *E. coli* cells and infecting them. M13 has been a versatile tool for the display of many different kinds of peptides and proteins, thereby leading to many interesting biomedical and biotechnical applications, as discussed below.

1.1.2 Bacteriophage T7

It has been possible to display proteins and peptides on a number of other bacteriophages as well. One popular virus is bacteriophage T7, in which it is possible to fuse peptides and proteins to the C-terminus of the 10B capsid protein. Vectors have been developed that direct display from 1 to 415 copies of polypeptides, 10 to 1000 amino acids in length, per virion. The low-valency display vectors are suitable for the selection of high-affinity ligands, while the higher valency display vectors enhance the recovery of low-affinity binders. Unlike the filamentous phage-display system, in which the displayed proteins need to be secreted into the periplasmic compartment of *E. coli* cells, T7 phage particles assemble in the cytoplasm, followed by lysis of the host cell. This characteristic overcomes the limited capacity of the protein secretion machinery of the *E. coli* cell and is more favorable for display of cytosolic proteins that carry reduced cysteine residues. The short life cycle (i.e. 2 h) for this lytic virus also makes it attractive for accelerating the screening process. Finally, since the protein of interest can be fused to the C-terminus of the major capsid protein, T7 phage is ideal for displaying segments of proteins encoded by cDNA fragments, which are typically biased toward containing C-terminal stop codons.

1.1.3 Bacteriophage Lambda (λ)

The λ display system has several attractive attributes. Compared to the T7 phage system, in which multivalent display of larger proteins has been reported to affect plaque formation, λ appears to tolerate display of proteins up to 300 amino acids in length on at least 90% of the copies of its capsid D protein. In addition, it is possible to display at either the N- or the C-termini of protein D. From a λ library displaying protein segments encoded by cDNA or genomic DNA fragments, one can identify interacting proteins for a variety of targets.

1.2 Cell Surface Display

In the recent years, cell surface display has emerged as an alternative to phage display. While cells can display thousands of copies of a peptide or protein, one of the biggest

advantages of cell display is that one can use fluorescence-activated cell sorting (FACS) to identify and quantitate positive clones during the selection process. In addition, the carbohydrates on the cell surface offer a largely inert surface in terms of nonspecific binding to target proteins, potentially leading to lower numbers of false positives. This section discusses various bacterial and yeast cell surface display techniques.

1.2.1 Bacterial Display

A variety of expression systems for the display of either short peptides or fully folded proteins on *E. coli* and, to a lesser extent, on gram-positive bacteria have been developed. Peptides or proteins have been targeted to the cell surface of bacteria by fusing them with protein components of the outer membrane, pilus, or flagella of *E. coli*. It has also been possible to display on the surface of gram-positive bacteria, *Staphylococcus carnosus* and *S. aureus*. Through FACS, bacteria that bind a particular fluorescent target protein can be enriched over 10 000-fold from a library, and with subsequent rounds of growth and cell sorting, they are eventually purified. Bacterial display has many applications, including vaccine development, mapping epitopes, bioremediation, biocatalysis, and biosensor development.

1.2.2 Yeast Display

Surface display of proteins on baker's yeast *Saccharomyces cerevisiae* was first reported in 1997 and, since then, it has been gaining increased use in the field of protein engineering. This system utilizes the cell surface receptor a-agglutinin, which mediates cell–cell adhesion during mating of the "a" and "α" haploid forms to produce a diploid yeast cell. The large subunit of a-agglutinin, Aga1p, is anchored to the yeast cell wall by a covalent β-glycan linkage, while the small subunit Aga2p is linked to Aga1p by two disulfide bonds (Fig. 2a). Proteins of interest can be displayed as C-terminal fusions to the Aga2p protein without disturbing assembly of the agglutinin protein heterodimer on the cell surface. To tightly regulate display, expression of the *Aga2p* gene fusion is directed by the GAL1-10 promoter, which is completely repressed by growth on glucose and induced by galactose. Thus, a yeast-display library can be propagated in the presence of glucose without expression of the fusion protein, which prevents loss of recombinant clones whose Aga2p fusion protein expression is toxic. A variety of proteins have been successfully displayed on yeast, including scFv's, single-chain T-cell receptors, Fabs, a single-chain form of the class II major histocompatibility complex product, and selectins.

Affinity selection of a yeast-display library is commonly performed by mixing the yeast cells displaying the protein of interest with a fluorescently labeled target, followed by sorting of the yeast-target complexes using FACS. The amount of Aga2p fusion protein displayed on cells can be monitored in parallel through detection of an epitope added to the C-terminus of the Aga2p fusion protein (Fig.2a). With FACS, the enrichment of yeast displaying an interacting protein is evident in real time during each selection step (Fig.2b). One less-expensive alternative to FACS is the use of streptavidin magnetic beads with biotinylated targets for fractionating yeast. To date, yeast display has been applied primarily to *in vitro* evolution studies in which 10^5 to 10^7 variants of a protein are generated and screened for an

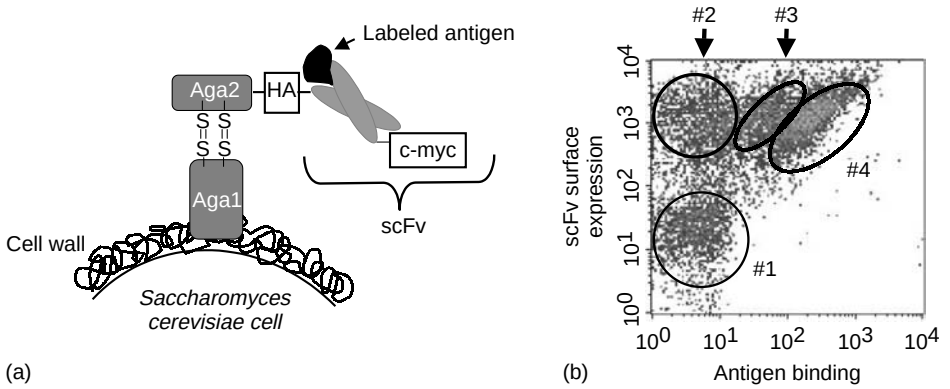

Fig. 2 The yeast display method. (a) Aga2-scFv fusion displayed on the surface of *S. cerevisiae*. The protein of interest (i.e. an scFv) is expressed on the yeast surface as a C-terminal fusion to Aga2. Aga2 is attached to Aga1 by disulfide bonds, which in turn is anchored in the yeast cell wall. An HA tag C-terminal to Aga2 and a c-myc tag C-terminal to the scFv allow determination of the fraction of cells expressing the recombinant clone and those expressing Aga2 only. The antigen is labeled so that the yeast, which bind the antigen, can be fractionated and quantified using FACS. (b) Bivariate flow cytometric analysis of yeast-displayed antibody fragments. A greatly enriched antigen-binding population shows the discrimination of clones with slight differences in affinities and nonbinding populations. Cells are double labeled with biotinylated antigen/streptavidin-phycoerythrin, and anti-c-myc/anti-mouse FITC labels. Subpopulation #4 expresses higher affinity scFv's and labels more brightly with antigen at a given scFv expression level. Subpopulation #3 expresses somewhat lower affinity scFv's, and subpopulation #2 does not detectably bind antigen. Subpopulation #1 is not expressing scFv on the cell surface. (Figure 2b was reproduced with permission from Feldhaus, M.J., Siegel, R.W. et al. (2003) Flow-cytometric isolation of human antibodies from a nonimmune *Saccharomyces cerevisiae* surface display library, *Nat. Biotechnol.* **13**(2), 163–170.)

optimal variant. Even though yeast cells are not efficiently transformed with exogenous DNA, such studies generally involve smaller libraries of variants, which are easily achievable using yeast. Recently, Feldhaus and coworkers reported the construction and characterization of the first nonimmune human scFv library displayed on yeast, with a diversity of 10^9 clones. By performing a considerable number of transformations, it was possible to overcome the low transformation efficiency of yeast. With this library, scFv's to many targets have been isolated and demonstrated to have affinities similar to those obtained from phage-display libraries. Moreover, the researchers were also able to amplify this library 10^{10} fold without loss of diversity, suggesting that the original library is essentially an unlimited source for scFv's.

1.3 In vitro Display

In vitro display technologies utilize cell-free systems, and are not limited by the biology of the virus or cells. *In vitro* libraries are typically 10^{12} to 10^{14}, orders of magnitude higher than the largest phage libraries available, which allows a much larger "sequence space" to be surveyed and tighter binders to be selected.

1.3.1 Ribosome Display

Ribosome display was first applied to the display of short peptides and later modified to display proteins and protein fragments. In this technique, a linear DNA segment, containing a promoter recognized by the T7 RNA polymerase, a ribosome binding site, the coding region for the peptide or protein, and a spacer region encoding a long unstructured stretch of amino acids, is transcribed and translated *in vitro* (Fig.3a). If the translation product lacks a stop codon, the nascent polypeptide chain remains tethered to the ribosome and

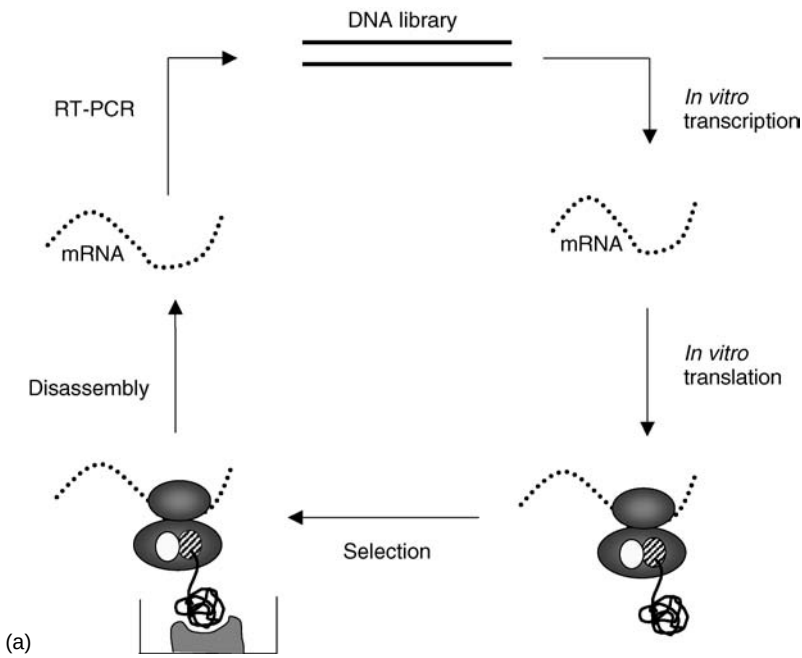

Fig. 3 Affinity selection using *in vitro* display techniques: DNA and mRNA are denoted by solid and broken lines, respectively. (a) Ribosome Display. A DNA library encoding proteins of interest is transcribed and translated *in vitro*. A long C-terminal spacer allows proteins to fold and lack of a stop codon prevents release of the protein. The mRNA-protein–ribosome complex is then used for selection on immobilized target. At the end of the selection step, the bound mRNA is released from the ribosome with EDTA, reverse transcribed, and the resulting cDNA is amplified by PCR and used either for another round of affinity selection or cloned and sequenced to determine its identity. (b) mRNA display. A DNA library encoding proteins of interest is transcribed *in vitro* and a short segment of DNA carrying a 5′ psoralen linker and a 3′ puromycin moiety (P) is hybridized to the mRNA. Photo-activation of the psoralen leads to covalent cross-linking of the mRNA and the DNA. This complex is then *in vitro* translated. The ribosomes stall at the RNA–DNA junction, allowing the puromycin to enter the A site of the ribosome (shaded circle) and accept the nascent polypeptide chain. Because of the covalent linkage of the mRNA and the protein, the ribosome can be dissociated at this point. Second-strand synthesis of the mRNA results in a DNA–RNA hybrid that is more stable and the selections are performed with this complex. At the end of the selection, the DNA is amplified by PCR and either used for additional rounds of selection or sequenced to yield the identity of the clone.

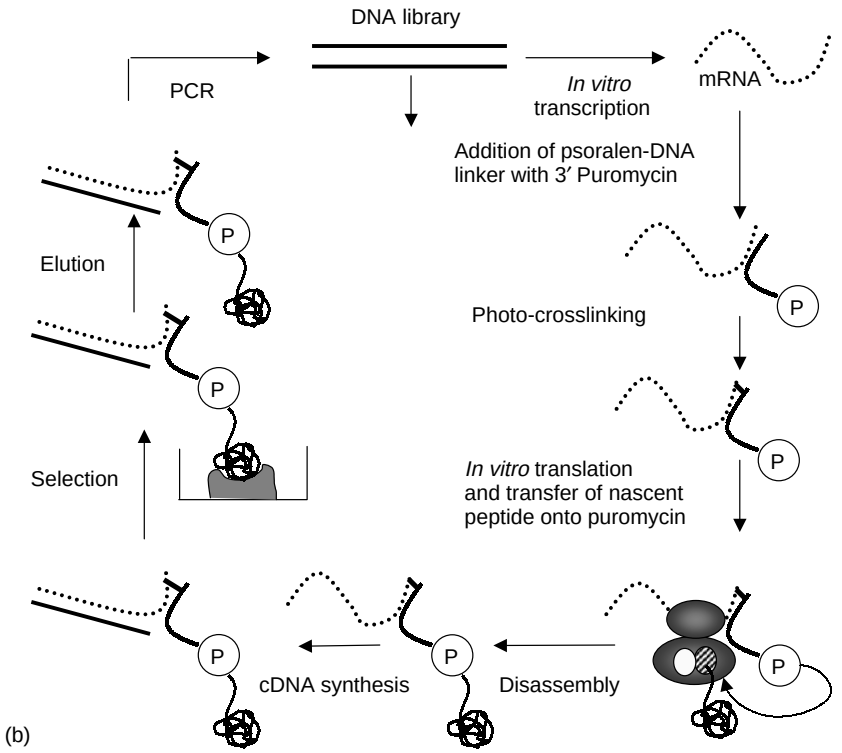

Fig. 3 (*Continued*)

mRNA, thereby creating a physical linkage between the genotype and phenotype. This RNA-protein complex, which is stabilized by high concentrations of magnesium and low temperatures, can be used in selection experiments. Sequences of interest are recovered by reverse transcribing the selected mRNA molecules, synthesizing the second DNA strand, and amplifying the resulting DNA by the polymerase chain reaction (PCR). Through additional rounds of *in vitro* transcription, translation and affinity selection, peptides or protein fragments with interesting properties are isolated from libraries.

Ribosome display has been mainly used for the display and selection of scFv's. By this technique, scFv's selected from libraries of $>10^{12}$ molecules generally have dissociation constants in the low nanomolar range. Since the PCR amplification step in the screening process can be modified to be mutagenic (i.e. 0.5–1%), this technique is well suited for molecular evolution experiments such as affinity maturation and stability engineering of scFv's. In one application, a low fidelity DNA polymerase was used for PCR amplification of scFv's selected with bovine insulin, and after six cycles of selection, scFv's were recovered with dissociation constants of 80 picomolar. In another study, scFv's that recognize the guanine quadruplex of human telomeric DNA sequences were isolated from a ribosome display library and demonstrated to have K_d values in the low nanomolar range. In immunofluorescence studies, it was possible to prove the existence of

the guanine quadruplex *in vivo,* for which there had previously been no conclusive evidence, using such antibody fragments.

Ribosome display has also been used to select for proteins with catalytic activity. A beta-lactamase could be enriched over an inactive mutant by ribosome display using a suicide substrate inhibitor. Even though the active enzyme made a covalent bond with the suicide substrate, chelating the magnesium ions releases its encoding mRNA for subsequent amplification and rescreening steps.

1.3.2 mRNA Display

mRNA display was first reported in 1997 by two groups working independently. This technique is similar to ribosome display, except that the nascent polypeptide chain is linked to its mRNA covalently by use of the aminoacyl-tRNA mimic puromycin. Puromycin can be attached to the 3′ end of the mRNA by hybridizing with single-stranded DNA containing psoralen and puromycin moieties, followed by photomediated cross-linking of the hybrid through the psoralen. This RNA–DNA partial duplex is then used for *in vitro* translation: the ribosome stalls at the RNA–DNA junction, which allows time for the puromycin to enter the ''A site'' of the ribosome and accept the nascent polypeptide chain, thereby creating a covalent linkage between the mRNA and the protein it encodes (Fig. 3b). To stabilize the mRNA component of the complex prior to selection experiments, it is reverse transcribed or replaced by double-stranded cDNA, which is covalently attached to the protein and the selection is performed with this complex. Since proteins are covalently linked to their coding sequence, affinity selection experiments can be performed under stringent conditions.

A variety of peptides and proteins have been displayed on mRNA and have been used to isolate high-affinity reagents that bind RNA, small molecules, and other proteins. RNA-binding peptides, which bound their target with low nanomolar affinity and high specificity, were isolated from combinatorial peptide libraries of 10^{12} to 10^{14} individual members. Some of these peptides were even able to differentiate between a single nucleotide change in the loop structure of the RNA. From a combinatorial peptide library containing 80 randomized amino acids, specific binders of ATP have been isolated with high affinity ($K_d = 100$ nM). From tissue-specific cDNA libraries, interaction partners for the antiapoptotic protein Bcl-X_L, which included known partners as well as previously unidentified proteins, have been isolated. Finally, using this technology, it was possible to isolate antibody mimetics, which were based on randomizing residues within the tenth fibronectin type III domain, to tumor necrosis factor-α (TNF-α) with K_d values in the picomolar range.

mRNA display also lends itself to unique applications such as incorporation of unnatural amino acids and chemical adducts into the synthesized protein. Attachment of a penicillin moiety to a fixed cysteine in a combinatorial peptide library allowed identification of novel inhibitors of the penicillin binding protein 2a, which were 100-fold more potent than penicillin alone in inhibiting synthesis of bacterial cell walls. The same research group also introduced a biotinylated amino acid into a displayed peptide by adding a biocytin-charged amber suppressor tRNA to the *in vitro* translation reaction. The ability to use unnatural amino acids overcomes the diversity limits imposed by the 20 natural amino acids and

can lead to the identification of peptides and proteins with unique characteristics.

2
Types of Libraries

Molecular display technologies allow the recovery and identification of molecules with desired properties without the need for any prior sequence information. This property has made it possible to survey very large libraries(i.e. 10^9–10^{14} unique sequences) for those rare members with the desirable biochemical traits. In this section, we briefly discuss the different kinds of libraries that are often used to accomplish this previously daunting task.

2.1
Combinatorial Peptide Libraries

One of the most common uses for bacteriophage M13 has been the display of combinatorial peptides. Oligonucleotides, which are fixed in length but degenerate in sequence, are inserted into one of the genes encoding one of the six capsid proteins, thereby generating a library of combinatorial peptides displayed on the surface of phage particles. While there are many codon schemes to choose from, the degenerate regions are commonly based on the NNK coding scheme: N is an equimolar mixture of A, C, G, and T, and K is an equimolar mixture of G and T nucleotides. In this scheme, 32 codons encode one stop codon (TAG), which is suppressed in certain *E. coli* strains, and all 20 amino acids, which are represented once (C, D, E, F, H, I, K, M, N, Q, W, Y), twice (A, G, P, V, T), or three times (L, R, S). These combinatorial peptide libraries, also called *random peptide libraries*, typically contain over one billion clones, with each viral particle displaying a different peptide sequence fused to protein III. A variety of combinatorial peptide libraries have been constructed over the years, of 6 to 42 amino acids in length, and with no or several fixed amino acids, such as cysteines. (Since the phage particles are secreted into an oxidizing environment, peptides with even numbers of cysteines will form intramolecular disulfide bonds, thereby conformationally constraining the peptides.) While such libraries are diverse enough to encode all possible six and shorter amino-acid-long peptide sequences, there is evidence for exclusion from phage-displayed combinatorial peptide libraries of certain peptide sequences that interfere with viral morphogenesis or secretion through the membrane.

2.2
scFv and Fab Libraries and Fab Libraries

Fabs or scFv's have been successfully displayed on the surface of bacteriophage. Creation of large repertoires of antibody fragments from antibody genes, as well as the development of high-throughput screening methods, reduce, and in some cases eliminate, the need to use animals for antibody production. Phage-displayed antibodies have been generated from a variety of animals (camel, chimpanzee, human, llama, mouse, and rabbit), and since each phage particle displays a single antibody fragment, they are monoclonal in nature. M13 phage display remains the most extensively used method for isolation of antibody fragments partly because it is well suited for the expression of immunoglobulin protein fragments; as the chimeric antibody-capsid proteins pass through the periplasm during viral morphogenesis, the disulfide bonds in the heavy

and light chains have an opportunity to form properly.

There are three routes for generating antibody display libraries. First, it is possible to immunize an animal with an immunogen, recover the circulating B-cells, extract mRNA, and clone portions of the immunoglobulin heavy and light chains into a display vector. In this manner, monoclonal antibodies can be rescued from immunized animals *in lieu* of generating hybridomas. It is possible to use this method to obtain neutralizing antibodies from human patients that recovered from exposure to infectious agents, such as the human immunodeficiency virus, respiratory syncytial virus, and herpes simplex virus, and *Helicobacter pylori*. Second, it is possible to generate a naïve antibody library by amplifying the variable regions of the heavy and light chains from a small number of B-cell donors and then cloning and displaying a large number of pair-wise combinations. From libraries containing over one billion clones, it is possible to generate antibodies to most antigens with nanomolar dissociation constants. Third, it has been possible to randomize the codons for the hypervariable regions within the heavy and light chain variable regions, thereby generating "semisynthetic" antibody libraries. All three approaches have been extremely productive in generating antibodies for basic research, biotechnology, and medical applications.

In the postgenomic era, antibody fragments can help to define novel gene function and to measure protein expression in pathological states. In addition to these new directions, antibody fragments have begun to show success in clinical diagnosis and as immunotherapies for cancer and other diseases. Radiolabeled antibodies, derived by display technologies, have great potential for cancer imaging and treatment. Antibodies have also been exploited for prodrug therapy, cancer treatment, and gene delivery by fusing them with different molecules such as toxins, enzymes, and viruses. The list of FDA approved antibody therapeutics against cancer, viral infection, and inflammatory disease is growing rapidly (http://www.fda.gov/cber/efoi/approve.htm).

2.3 Antibody Mimetic Libraries

Although scFv's and Fabs have been used extensively as affinity reagents, they sometimes suffer from poor yields when expressed in *E. coli* and can be unstable. These drawbacks have prompted a search for alternative scaffolds, which bind targets with comparable specificity and affinity. These mimetics generally have improved stabilities, better expression yields in *E. coli*, and, in some cases, are smaller, inherently fluorescent, or lack disulfide bonds. Mimetics without disulfide bonds can be used for *in vivo* inhibition of cytoplasmic targets because they adopt stable structures in the reducing environment of the cytoplasm. Although it remains to be determined if antibody mimetics can replace therapeutic antibodies, they offer great promise in research and diagnostics.

Antibody mimetic libraries can be constructed by randomizing residues that are exposed on the scaffold surface or by grafting the complementarity determining regions (CDR) of antibodies onto the scaffold. Some of these mimetics have been termed *affibodies, monobodies*, and *anticalins*, which are based on a three-helix bundle Protein A domain from *Staphylococcus aureus*, the tenth type III domain of human fibronectin, and

lipocalin, respectively. Affibodies, monobodies, and fluorobodies have been used to generate binders with nanomolar to picomolar affinities against protein targets, whereas anticalins have been mainly used for isolating binders with midnanomolar affinity for small molecules such as fluorescein and digoxigenin. Monobodies isolated against estrogen receptor complexed with different ligands have been employed as intracellular detectors for ligand-induced conformational changes of the target. Recently, repeat proteins such as ankyrins and leucine repeats have been explored as antibody mimetics where multiple randomized repeats are stacked together to create a binding surface for targets. The modular structure of this class of antibody mimetics may allow greater flexibility in designing binders with greater diversity and increased surface area for interactions with their targets. Ankyrins are also promising reagents for intracellular protein "knock-out" experiments because they do not require disulfide bond formation for their stability.

2.4 cDNA and Genomic DNA Libraries

Bacteriophages have also been used to display libraries of protein segments encoded by complementary DNA (cDNA) fragments. Libraries of plant, fungal, and dust mite cDNAs have been cloned and displayed on bacteriophage M13 for the purpose of identifying human allergens: the libraries are screened with serum IgE and phage-displaying reactive allergens are selected and subsequently identified by DNA sequencing. Bacteriophage λ and T7 have been shown effective in displaying protein fragments encoded by cDNA segments cloned at the C-terminus of their major capsid proteins. Libraries of cDNA inserts corresponding to mRNA purified from different cell types are now widely available and such libraries have been screened extensively for interacting with specific protein, RNA, DNA promoter, carbohydrate, or small chemical molecules.

3 Applications

3.1 Protein Function and Target Validation

3.1.1 Mapping Protein–protein Interactions with Peptides

Peptide ligands selected from a phage-displayed library appear to bind preferentially to "hot spots" on the surface of protein targets. A hot spot is defined as a small region on the surface of a protein, often composed of several hydrophobic, disordered amino acids, which binds other proteins or small molecules. Peptide ligands selected from a phage library tend to home in on hot spots and to mimic proteins that normally interact with the target protein. In fact, very often, the primary structures of the selected peptides resemble regions within the interacting proteins. This phenomenon has been termed an example of *convergent evolution*, where the biological and molecular evolution has yielded similar results. Consequently, phage display has proven very useful in mapping the protein–protein interactions of extracellular and intracellular proteins (Fig. 4). Several notable examples include defining the epitopes of monoclonal and polyclonal antibodies, discovering agonists and antagonists of hormone receptors, and defining the specificity of protein interaction modules

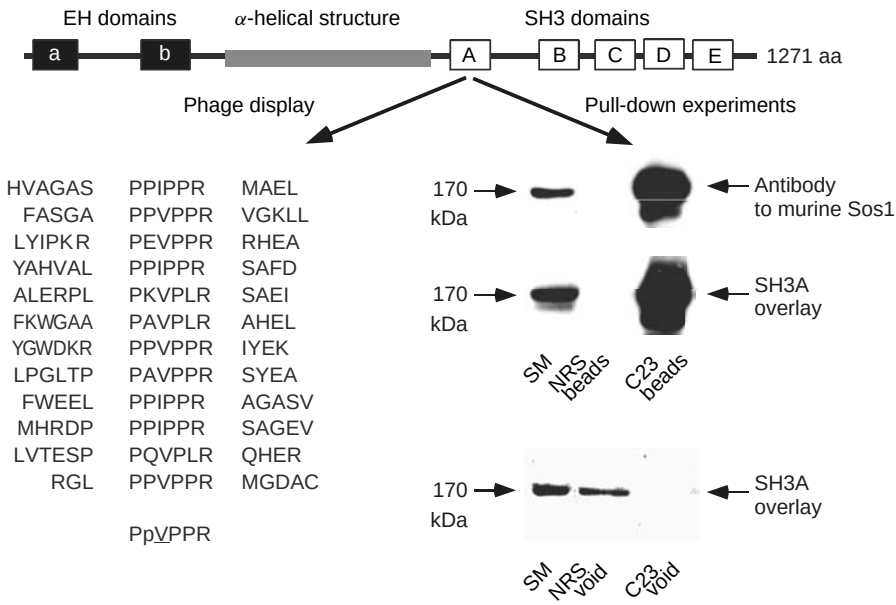

Fig. 4 Mapping protein–protein interactions. The intersectin protein consists of two Eps15 Homology (EH) domains, a central coiled-coil region, and five Src Homology 3 (SH3) domains. A protein fusion between the N-terminal SH3 domain (SH3A) and the glutathione-S-transferase (GST) can be used to affinity select phage-displayed combinatorial peptide libraries. The predicted amino acid sequences of selected phage are aligned to reveal the motif, PpVPPR, where p is commonly proline. Computer searches with the motif matched a sequence (PPVPPR) within the Son-of-Sevenless (Sos) protein, suggesting that it may interact with intersectin. Sos was demonstrated to interact with intersectin through pull-down and coimmunoprecipitation experiments. Sos is 170 kilodaltons (kDa) and reacts with C23, a commercial monoclonal antibody. SM and NRS void refer to starting material (rat brain extract) and nonreactive serum respectively.

present in signal transduction and endocytic proteins.

3.1.2 Elucidation of Enzyme Function

Peptides selected for binding to enzymes appear to preferentially bind at the active sites of the enzymes, even though they are not involved in protein–protein interactions. In a survey of ten different enzymes, peptide ligands were isolated for eight of the target proteins, and at least six were shown to inhibit the enzymes. Synthetic forms of the selected peptides acted as competitive inhibitors in activity assays. In addition, they were used in displacement assays to screen libraries of chemicals and natural products for those that prevented binding of the peptide ligands to the enzymes *in vitro*, and thus bind to the active site of the enzyme. Such peptides can also be expressed inside cells where they can bind to the enzyme and inhibit its *in vivo* activity; in some cases, inhibition leads to a cellular phenotype.

It is also possible to screen combinatorial peptide libraries to define the substrate specificity of enzymes. By tethering phage through an anchor sequence, the

substrate specificity of proteases can be defined: only those viruses that carry a cleavage site in the linker composed of combinatorial peptides are released. In this manner, "substrate phages" have been identified for a variety of proteases, such as subtilisin, factor Xa, furin, matrix metalloproteinases, urokinase, granzyme B, prostate-specific antigen, and HIV-1 protease. Substrate phages have also proven useful for defining the substrate specificity of various protein tyrosine kinases.

3.1.3 Functional Knock-outs Using Intrabodies

Over the past few years, several groups have expressed antibody fragments inside cells. These intracellularly expressed antibody fragments, also named "intrabodies," can provide a way of validating targets by blocking pathways in cells. While only a small fraction of phage-displayed scFv molecules can fold properly in the reducing environment of the cytoplasm, those which do fold are still able to bind to their targets and yield interesting biological results, such as arresting HIV infection, blocking apoptosis, disturbing proper embryo development, and arresting oncogenic transformation of cells in culture. Future experiments in functional genomics will rely heavily on antibodies or mimetics that fold efficiently inside cells.

3.2 Drug Discovery and Therapeutics

3.2.1 Identification of Tissue- and Tumor-specific Markers via Whole-cell and *in vivo* Panning

Phage-displayed peptide and scFv libraries have also been used successfully for affinity selections against target mixtures present on the surface of cultured mammalian cells or tissues. Because whole-cell or tissue panning bypasses the need to purify the antigen, this method is useful for generating affinity reagents against antigens that are hard to purify, such as transmembrane proteins. Using this procedure, scFv's and peptides have been isolated for cell surface receptors such as urokinase plasminogen activated receptor, thrombin receptor, melanocortin receptor, and cell types such as colorectal tumor cells, lymphatic vessels, and *Bacillus* spores. Cell- and tumor-specific peptide ligands and antibodies have great potential usefulness in profiling the type or the stage of cancer in a manner similar to expression profiling using DNA microarrays.

In working with intact cells, scientists have realized that phage particles that promote cellular internalization can be selected from a display library. After incubating the phage particles with intact cells, the unbound phage are washed away and the cells are lysed in order to recover the internalized phage. Internalizing scFv's, which were shown to recognize the ErbB2 antigen, have been identified for SKBR3 breast cancer cells, and when fused to liposomes, target delivery into ErbB2 expressing tumor cells. Peptides selective for internalization by fibroblast cells, laryngeal carcinoma cells and ECV304 endothelial cells, and the human urothelium have been reported.

Phage-displayed combinatorial peptide libraries have also been used in *in vivo* panning. In a series of ground-breaking papers, Renata Pasqualini and Erki Ruoslahti injected phage particles displaying combinatorial peptides into mice, dissected out various tissues, amplified the bound phage

particles in *E. coli*, and repeated the selection process several times. Individual clones, when tested, were confirmed to bind selectively to their respective tissue targets. Similar experiments have also been conducted in humans: phage particles have been injected into patients and peptides were identified that bind selectively to human vasculature. Such "organ homing" peptides can be linked to drugs, and adenovirus or adeno-associated virus, for the purpose of targeting the delivery of the drugs or gene therapy agents respectively.

3.2.2 Screening of Small Molecule Inhibitors Using Surrogate Peptide Ligands

Since peptide ligands selected from a phage-displayed combinatorial peptide library often bind at sites of protein–protein interactions or catalysis, assays for discovering small molecule inhibitors can be developed. For example, peptides chemically synthesized with fluorophores attached at one end can be incubated in microtiter plate wells with the target protein and compounds from natural and chemical libraries. If the compound does not bind to the target, the fluorescence polarization of the labeled peptide will become anisotropic because of binding of a small peptide to a larger target molecule. On the other hand, if the compound binds to the target, thereby blocking peptide binding, the fluorescence polarization of the peptide will be isotropic, because it is free in solution. High-end devices can scan 96-, 384-, 1536-, or 6144-well microtiter plates rapidly in a laboratory, thereby permitting research groups to survey libraries of thousands and thousands of compounds for potential inhibitors. This general approach has lead to the discovery of compounds with potential antimicrobial activity.

3.2.3 Zinc Fingers for Controlling Gene Expression

One fruitful avenue for designing affinity reagents has been with zinc fingers, which typically bind DNA. Multiple zinc fingers can be displayed on phage and can bind to specific DNA sequences, with each finger (30 amino acids) roughly binding three adjacent bases in a double-stranded DNA molecule. Thus, it has been possible to randomize residues within a phage-displayed zinc finger and select for variants that could bind to specific double-stranded DNA molecules. Through a concerted effort, zinc fingers that bind to many different DNA triplet sequences have been identified. Such fingers can be "stitched" together to make polydactyl fingers (i.e. 6 adjacent zinc fingers) that recognize unique 18-bp sequences with subnanomolar to femtomolar dissociation constants. One practical consequence of this work is that it is now possible to introduce engineered finger proteins into cells and induce or repress expression of specific genes. It will be interesting to follow how this approach matures.

3.3 Protein Engineering and Industrial Applications

3.3.1 Molecular Evolution

Molecular evolution refers to improvement of the desired characteristics of a molecule using techniques such as error-prone PCR or DNA shuffling and the appropriate selection protocols. This strategy is used extensively in protein engineering in situations where the mutations that might bring about a desired property cannot be predicted from existing biochemical or structural information, necessitating the generation and screening of random mutants. With the application of different display technologies, the number

of clones that can be screened at the end of a molecular evolution experiment has increased by orders of magnitude. This allows researchers to access larger numbers of permutations for the desired protein products. By enhancing the specificity, affinity, or stability of an scFv, one can generate a reagent with improved therapeutic value. Using yeast display and a combination of DNA shuffling and error-prone mutagenesis, the dissociation constant of an scFv that binds fluorescein was improved over three orders of magnitude to 48 femtomolar.

Through phage display and targeted mutagenesis, the dissociation constant of an anti-Vascular Endothelial Growth Factor (VEGF) antibody fragment was improved 120 fold over the parent molecule. Finally, by using ribosome display, an antihemagglutinin scFv was evolved to have increased stability in DTT. Some of the clones were able to bind their antigen when expressed in the reducing environment of the *E. coli* cytoplasm. These types of antibody fragments would be invaluable as reagents for protein ''knock-out'' and target validation studies.

Phage-displayed proteins are amenable to selection for their resistance to extreme conditions, such as denaturants, detergents, proteases, and elevated temperatures. Thus, one can select for protein variants that maintain their shape or function under such conditions. For example, from a phage library displaying mutated ribonuclease T1 molecules, it was possible to select variants more stable than the wild-type molecule. A screen of mutagenized *Bacillus subtilis* cold shock protein CspB led to the identification of variants that had higher thermal or chemical stability, depending on whether the selection was done at elevated temperature or in the presence of a denaturant, respectively. In another case, from a library of mutagenized alpha-amylase proteins, variants were discovered that hydrolyze starch at low pH conditions that favor industrial applications. Finally, a partially unfolded four-helix bundle protein, apocytochrome b(562), was phage displayed and challenged with a protease, to yield a stably folded form of the protein.

3.3.2 Bacterial and Yeast Surface Display as Biocatalysts

Ability to display proteins and enzymes on bacterial and yeast cell surfaces have proven to be valuable in industrial applications. For example, construction of a whole-cell biocatalyst with its sequential reaction has been achieved by the genetic immobilization of two amylolytic enzymes on the yeast cell surface. Organophosphorus hydrolase, a bacterial enzyme that has been shown to degrade a wide range of neurotoxic organophosphate nerve agents, was evolved and displayed on *E. coli* surface to hydrolyze methyl parathion, providing the possibility of creating biocatalysts that can degrade pesticides (i.e. diazinon and chlorpyrifos) and nerve agents (i.e. sarin and soman). *E. coli* and *Moraxella sp* (a bacterium known to survive in contaminated environments) have been successfully explored to display phytochelatins, which could be used for bioaccumulation of cadmium and mercury, providing a new method for treating heavy metal contamination. Finally, display of glycosyltransferases at the yeast cell surface through a cell wall anchored protein has shown potential as a simple *in vitro* method for oligosaccharide synthesis using the yeast intact cell as a biocatalyst.

3.3.3 Nanotechnology

Recently, several groups have screened phage-displayed combinatorial peptide

libraries for peptides that bind to inert surfaces. Even though such peptide-surface interactions are intrinsically weak, they are detected because of the avidity resulting from the pentavalent presentation of protein III on M13. Peptides have been discovered that bind selectively to polystyrene, semiconductor surfaces, silver nanoparticles, and ZnS or CdS nanocrystals. These peptides have great promise for engineering new nanostructures.

4 Outlook and Perspectives

Scientists in both basic science and biotechnology have used display technologies creatively. While the technologies have matured over the past few years, new applications appear on a regular basis, suggesting that they will continue to impact various fields of science in a significant manner for the next 10 years. In particular, it is anticipated that these technologies will contribute significantly in the emerging areas of nanotechnology and personalized medicine, where display technologies can be used for fabricating nanostructures and treating disease respectively.

Acknowledgments

The submitted manuscript has been created by The University of Chicago as Operator of Argonne National Laboratory ("Argonne"). Argonne, a US Department of Energy Office of Science laboratory, is operated under Contract No. W-31-109-Eng-38. The US Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.

See also Combinatorial Phage Antibody Libraries; DNA Libraries; Peptide and Non-Peptide Combinatorial Libraries.

Bibliography

Books and Reviews

Amstutz, P., Forrer, P., Zahnd, C., Pluckthun, A. (2001) *In vitro* display technologies: novel developments and applications, *Curr. Opin. Biotechnol.* **12**(4), 400–405.

Brown, K.C. (2000) New approaches for cell-specific targeting: identification of cell-selective peptides from combinatorial libraries, *Curr. Opin. Chem. Biol* **4**(1), 16–21.

Dower, W.J., Mattheakis, L.C. (2002) *In vitro* selection as a powerful tool for the applied evolution of proteins and peptides, *Curr. Opin. Chem. Biol.* **6**(3), 390–398.

Kay, B.K., Kasanov, J., Knight, S., Kurakin, A. (2000) Convergent evolution with combinatorial peptides, *FEBS Lett.* **480**(1), 55–62.

Kretzschmar, T., von Ruden, T. (2002) Antibody discovery: phage display, *Curr. Opin. Biotechnol.* **13**(6), 598–602.

Ladner, R.C., Ley, A.C. (2001) Novel frameworks as a source of high-affinity ligands, *Curr. Opin. Biotechnol.* **12**(4), 406–410.

Ruoslahti, E. (2002) Drug targeting to specific vascular sites, *Drug Discov. Today* **7**(22), 1138–1143.

Skerra, A. (2000) Engineered protein scaffolds for molecular recognition, *J. Mol. Recognit.* **13**(4), 167–187.

Primary Literature

Abe, H., Shimma, Y., Jigami, Y. (2003) *In vitro* oligosaccharide synthesis using intact yeast cells that display glycosyltransferases at the cell surface through cell wall-anchored protein Pir, *Glycobiology* **13**(2), 87–95.

Amstutz, P., Pelletier, J.N., Guggisberg, A., Jermutus, L., Cesaro-Tadic, S., Zahnd, C., Pluckthun, A. (2002) *In vitro* selection for catalytic activity with ribosome display, *J. Am. Chem. Soc.* **124**(32), 9396–9403.

Arkin, M.R., Randal, M., DeLano, W.L., et al. (2003) Binding of small molecules to an adaptive protein-protein interface, *Proc. Natl. Acad. Sci. U.S.A.* **100**(4), 1603–1608.

Bae, W., Mulchandani, A., Chen, W. (2002) Cell surface display of synthetic phytochelatins using ice nucleation protein for enhanced heavy metal bioaccumulation, *J. Inorg. Biochem.* **88**(2), 223–227.

Barrick, J.E., Takahashi, T.T., Ren, J., Xia, T., Roberts, R.W. (2001) Large libraries reveal diverse solutions to an RNA recognition problem, *Proc. Natl. Acad. Sci. U.S.A.* **98**(22), 12374–12378.

Beerli, R.R., Barbas, C.F. III. (2002) Engineering polydactyl zinc-finger transcription factors, *Nat. Biotechnol.* **20**(2), 135–141.

Benson, R.E., Gottlin, E.B., Christensen, D.J., Hamilton, P.T. (2003) Intracellular expression of Peptide fusions for demonstration of protein essentiality in bacteria, *Antimicrob. Agents Chemother.* **47**(9), 2875–2881.

Beste, G., Schmidt, F.S., Stibora, T., Skerra, A. (1999) Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold, *Proc. Natl. Acad. Sci. U.S.A.* **96**(5), 1898–1903.

Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P., Pluckthun, A. (2003) Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins, *J. Mol. Biol.* **332**(2), 489–503.

Boder, E.T., Midelfort, K.S., Wittrup, K.D. (2000) Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity, *Proc. Natl. Acad. Sci. U.S.A.* **97**(20), 10701–10705.

Boder, E.T., Wittrup, K.D. (1997) Yeast surface display for screening combinatorial polypeptide libraries, *Nat. Biotechnol.* **15**(6), 553–557.

Chen, Y., Wiesmann, C., Fuh, G., Li, B., Christinger, H.W., McKay, P., de Vos, A.M., Lowman, H.B. (1999) Selection and analysis of an optimized anti-VEGF antibody: crystal structure of an affinity-matured Fab in complex with antigen, *J. Mol. Biol.* **293**(4), 865–881.

Cho, C.M., Mulchandani, A., Chen, W. (2002) Bacterial cell surface display of organophosphorus hydrolase for selective screening of improved hydrolysis of organophosphate nerve agents, *Appl. Environ. Microbiol.* **68**(4), 2026–2030.

Choo, Y., Klug, A. (1994) Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11163–11167.

Christensen, D.J., Gottlin, E.B., Benson, R.E., Hamilton, P.T. (2001) Phage display for target-based antibacterial drug discovery, *Drug Discov. Today* **6**(14), 721–727.

Chu, R., Takei, J., Knowlton, J.R., Andrykovitch, M., Pei, W., Kajava, A.V., Steinbach, P.J., Ji, X., Bai, Y. (2002) Redesign of a four-helix bundle protein by phage display coupled with proteolysis and structural characterization by NMR and X-ray crystallography, *J. Mol. Biol.* **323**(2), 253–262.

Clackson, T., Wells, J.A. (1995) A hot spot of binding energy in a hormone-receptor interface, *Science* **267**, 383–386.

Crameri, R., Kodzius, R. (2001) The powerful combination of phage surface display of cDNA libraries and high throughput screening, *Comb. Chem. High Throughput Screen.* **4**(2), 145–155.

DeLano, W.L., Ultsch, M.H., de Vos, A.M., Wells, J.A. (2000) Convergent solutions to binding at a protein-protein interface, *Science* **287**(5456), 1279–1283.

Falke, D., Juliano, R.L. (2003) Selective gene regulation with designed transcription factors: implications for therapy, *Curr. Opin. Mol. Ther.* **5**(2), 161–166.

Feldhaus, M.J., Siegel, R.W., Opresko, L.K., et al. (2003) Flow-cytometric isolation of human antibodies from a nonimmune *Saccharomyces cerevisiae* surface display library, *Nat. Biotechnol.* **21**(2), 163–170.

Goldenberg, D.M. (2002) Targeted therapy of cancer with radiolabeled antibodies, *J. Nucl. Med.* **43**(5), 693–713.

Goodson, R.J., Doyle, M.V., Kaufman, S.E., Rosenberg, S. (1994) High-affinity urokinase receptor antagonists identified with bacteriophage peptide display, *Proc. Natl. Acad. Sci. U.S.A.* **91**(15), 7129–7133.

Hammond, P.W., Alpin, J., Rise, C.E., Wright, M., Kreider, B.L. (2001) *In vitro* selection and characterization of Bcl-X(L)-binding proteins

from a mix of tissue-specific mRNA display libraries, *J. Biol. Chem.* **276**(24), 20898–20906.
Hanes, J., Pluckthun, A. (1997) *In vitro* selection and evolution of functional proteins by using ribosome display, *Proc. Natl. Acad. Sci. U.S.A.* **94**(10), 4937–4942.
He, M., Taussig, M.J. (1997) Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for in vitro display and evolution of antibody combining sites, *Nucl. Acids Res.* **25**(24), 5132–5134.
Hoess, R.H. (2002) Bacteriophage lambda as a vehicle for peptide and protein display, *Curr. Pharm. Biotechnol.* **3**(1), 23–28.
Hyde-DeRuyscher, R., Paige, L.A., Christensen, D.J., et al. (2000) Detection of small-molecule enzyme inhibitors with peptides isolated from phage-displayed combinatorial peptide libraries, *Chem. Biol.* **7**(1), 17–25.
Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J., Pluckthun, A. (2001) Tailoring in vitro evolution for protein affinity or stability, *Proc. Natl. Acad. Sci. U.S.A.* **98**(1), 75–80.
Keefe, A.D., Szostak, J.W. (2001) Functional proteins from a random-sequence library, *Nature* **410**(6829), 715–718.
Koide, A., Bailey, C.W., Huang, X., Koide, S. (1998) The fibronectin type III domain as a scaffold for novel binding proteins, *J. Mol. Biol.* **284**(4), 1141–1151.
Kretzschmar, T., von Ruden, T. (2002) Antibody discovery: phage display, *Curr. Opin. Biotechnol.* **13**(6), 598–602.
Lee, S.Y., Choi, J.H., Xu, Z. (2003) Microbial cell-surface display, *Trends Biotechnol.* **21**(1), 45–52.
Li, S., Millward, S., Roberts, R. (2002) *In vitro* selection of mRNA display libraries containing an unnatural amino acid, *J. Am. Chem. Soc.* **124**(34), 9972–9973.
Li, S., Roberts, R.W. (2003) A novel strategy for in vitro selection of peptide-drug conjugates, *Chem. Biol.* **10**(3), 233–239.
Martin, A., Sieber, V., Schmid, F.X. (2001) *In vitro* selection of highly stabilized protein variants with optimized surface, *J. Mol. Biol.* **309**(3), 717–726.
Mattheakis, L.C., Bhatt, R.R., Dower, W.J. (1994) An *in vitro* polysome display system for identifying ligands from very large peptide libraries, *Proc. Natl. Acad. Sci. U.S.A.* **91**(19), 9022–9026.
Murai, T., Ueda, M., Shibasaki, Y., Kamasawa, N., Osumi, M., Imanaka, T., Tanaka, A. (1999) Development of an arming yeast strain for efficient utilization of starch by co-display of sequential amylolytic enzymes on the cell surface, *Appl. Microbiol. Biotechnol.* **51**(1), 65–70.
Nemoto, N., Miyamoto-Sato, E., Husimi, Y., Yanagawa, H. (1997) *In vitro* virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro, *FEBS Lett.* **414**(2), 405–408.
Nord, K., Nilsson, J., Nilsson, B., Uhlen, M., Nygren, P.A. (1995) A combinatorial library of an alpha-helical bacterial receptor domain, *Protein Eng.* **8**(6), 601–608.
Pasqualini, R., Ruoslahti, E. (1996) Organ targeting in vivo using phage display peptide libraries, *Nature* **380**, 364–366.
Roberts, R.W., Szostak, J.W. (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins, *Proc. Natl. Acad. Sci. U.S.A.* **94**(23), 12297–12302.
Rodi, D.J., Makowski, L., Kay, B.K. (2002) One from column A and two from column B: the benefits of phage display in molecular-recognition studies, *Curr. Opin. Chem. Biol.* **6**(1), 92–96.
Sarikaya, M., Tamerler, C., Jen, A.K., Schulten, K., Baneyx, F. (2003) Molecular biomimetics: nanotechnology through biology, *Nat. Mater.* **2**(9), 577–585.
Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H.J., Pluckthun, A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei, *Proc. Natl. Acad. Sci. U.S.A.* **98**(15), 8572–8577.
Sieber, V., Pluckthun, A., Schmid, F.X. (1998) Selecting proteins with improved stability by a phage-based method, *Nat. Biotechnol.* **16**(10), 955–960.
Smith, G.P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface, *Science* **228**(4705), 1315–1317.
Stemmer, W.P. (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution, *Proc. Natl. Acad. Sci. U.S.A.* **91**(22), 10747–10751.
Stumpp, M.T., Forrer, P., Binz, H.K., Pluckthun, A. (2003) Designing repeat proteins: modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family, *J. Mol. Biol.* **332**(2), 471–487.

Tao, J., Wendler, P., Connelly, G., et al. (2000) Drug target validation: lethal infection blocked by inducible peptide, *Proc. Natl. Acad. Sci. U.S.A.* **97**(2), 783–786.

Tong, A.H., Drees, B., Nardelli, G., et al. (2002) A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules, *Science* **295**(5553), 321–324.

Verhaert, R.M., Beekwilder, J., Olsthoorn, R., van Duin, J., Quax, W.J. (2002) Phage display selects for amylases with improved low pH starch-binding, *J. Biotechnol.* **96**(1), 103–118.

Winter, G., Griffiths, A.D., Hawkins, R.E., Hoogenboom, H.R. (1994) Making antibodies by phage display technology, *Annu. Rev. Immunol.* **12**, 433–455.

Xu, L., Aha, P., Gu, K., et al. (2002) Directed evolution of high-affinity antibody mimics using mRNA display, *Chem. Biol.* **9**(8), 933–942.

Yamabhai, M., Hoffman, N.G., Hardison, N.L., McPherson, P.S., Castagnoli, L., Cesareni, G., Kay, B.K. (1998) Intersectin, a novel adaptor protein with two Eps15 homology and five Src homology 3 domains, *J. Biol. Chem.* **273**(47), 31401–31407.

Molecular Immunology: *see* Immunology

Molecular Mediators: Cytokines

Jean-Marc Cavaillon
Institut Pasteur, Paris, France

1 Historical Record 434

2 A Universal Language of Cells 435
2.1 Definitions 435
2.2 Families 437

3 Receptors 438
3.1 Families 438
3.2 Signaling 439
3.3 Soluble Receptors 440

4 Functions 441
4.1 Hematopoiesis 441
4.2 Immune Response 441
4.2.1 Innate Immunity 441
4.2.2 Adaptive Immunity 442
4.2.3 Th1 and Th2 Cytokines 443
4.3 Cell Survival and Cell Death 443
4.4 Link with the Central and Peripheral Nervous System 444
4.5 Reproduction 444
4.6 Inflammation 445

5 Life without Cytokines (What Knockout Mice Tell Us) 447

6 Cytokine Synthesis 449
6.1 Homeostasis 449
6.2 Induction 449
6.3 Parameters that Affect Production 450
6.4 Measurements 450

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

7 The Cytokine Network 451

8 Individual Heterogeneity 452

9 Cytokines and Infections 453
9.1 Half Angel–Half Devil 453
9.2 The Strategies of Microbes against Cytokines 454

10 Cytokines and Diseases 454
10.1 Autoimmune Diseases 455
10.2 Allergy 455
10.3 Sepsis 455
10.4 Chronic Inflammatory Diseases 455
10.5 Cancer 456
10.6 Hepatitis C 456
10.7 Multiple Sclerosis 456
10.8 Transplantation 456
10.9 Neutropenia 457

11 Conclusion 457

Bibliography 457
Books and Reviews 457
Primary Literature 457

Keywords

Cytokines
Constitutively released or produced upon cell activation, cytokines are soluble mediators – sometimes membrane bound – binding to receptors on target cells and inducing, modulating, or inhibiting cellular functions. Most cytokines are produced by immune cells and act on immune cells, but cytokines are produced by numerous cell types and act on a great variety of cells, accordingly displaying a large spectrum of activities. They are characterized by their redundancy, their synergistic action, and their regulatory loops. Those produced constitutively contribute to the tissue and systemic homeostasis. The cytokine are elements of a universal language used by cells to communicate.

Chemokines
A subfamily of cytokines with structural similarities and sharing the property to recruit cells within the extravascular space of the tissues (chemotaxis), either at homeostasis or during inflammatory processes.

Interferons (IFN)
A subfamily of cytokines sharing an antiviral bioactivity. Interferons are characterized by different subgroups (type I and II), depending on their origin. In addition to their common property, some (e.g. IFNγ) are involved at different stages of the onset and further development of the immune responses, or play a role during gestation in mammals (e.g. IFNτ).

Interleukin (IL-)
A convenient name to classify cytokines. The numbers have been given as a reflection of their discovery through the years. This term does not define a family, since none are biochemically or biologically related. However, the most recent interleukins have been described after gene banks and are thus related to previous well-known interleukins. When the word was created (1979), there was only IL-1 and IL-2.

Receptors
The receptors of cytokines allow the cells to respond to the signal delivered by cytokines. They can involve 1, 2, or 3 different chains. Most chains are constituted with one extracellular domain, one transmembrane domain and one intracellular domain, except the chemokine receptors that are constituted by 7 transmembrane domains.

Cytokines are proteins or glycoproteins produced by cells and acting on other cells that display on their surface the specific cytokine receptor. Cytokines are words used by the cells to communicate. Alone or within a determined sequence, these mediators lead the responding cells to modify its function (e.g. secretion, proliferation, induction, inhibition, enhanced/reduced function, migration, apoptosis, etc.). Despite cytokines having been mainly discovered by immunologists as a product of cells of the leukocyte lineages, they are now recognized as elements of a universal language used by most cells of any other lineages. Accordingly, they are essential for many events through the life (reproductive tissue remodeling, embryogenesis, steady state and adaptive hematopoiesis, surveillance and maintenance of the tissue structure, functions by the immune system, cell survival, and cell death, etc). They also allow a dialogue with the central nervous system and some cytokines may modify different behaviors (fever, anorexia, sleep, etc). They usually act within their vicinity, but they can also act via an endocrine fashion. Their productions are tightly controlled within a complex network of positive and negative loops. They are a prerequisite to control infection by invasive microorganisms, but their exacerbated production can be deleterious at the local or systemic levels. Their half angel–half devil aspect has rendered them difficult to use with therapeutic purposes, although some recombinant cytokines or some cytokine-neutralizing strategies have been successfully used in different pathological conditions.

1
Historical Record

The history of cytokines can be divided in four main steps through the years. The first step corresponds to the identification of biological activities of factors present in cell culture supernatants or within the blood stream (late 40s–early 70s). During the second step (late 70s–early 80s), the purification and the biochemical characterization of these factors allowed to reach the stage of defined molecular entities. During the third step (mid-80s–mid-90s), the cloning of the coding and noncoding sequences, the production of recombinant cytokines, as well as the production of genetically manipulated mice extended the understanding of the complexity of the cytokine-mediated language. Since the late 90s (the fourth step), the discovery of new cytokines has been relying on the availability of different genome sequences and the different gene banks. In the later case, bioactivities of newly identified cytokines remained to be discovered. Probably, the first reported bioactivity was the induction of fever and the discovery of an endogenous pyrogen in the late 40s, early 50s. The purification of human endogenous pyrogen was published in 1977 by Charles Dinarello. This allowed the identification of endogenous pyrogen as interleukin-1 (IL-1). It definitively established that fever was due to endogenous factor(s) and not directly induced by microbial products. After the cloning of IL-1 in 1984, this was confirmed the following year, when injection of recombinant IL-1 was shown to induce fever in rabbits. Later, genes encoding different IL-1 homologues have been identified by means of their significant sequence similarities to IL-1, but further studies are required to know whether some members of the IL-1 family (IL-1F5–10) are pyrogenic. The other earlier-described bioactivity was reported in 1957 by Isaacs and Lindenmann. They showed that resistance to virus development could be conferred to a cell by another virus-infected cell. This bioactivity interfering with virus spreading was named "interferon." In the 60s, numerous bioactivities had been detected in supernatants of lymphocytes. Accordingly, identified as nonantibody mediators, D. Dumonde et al. proposed in 1969, the word "lymphokines." In the following years, the word "monokine" was employed to characterize other bioactivities found in monocyte/macrophage supernatants. In the 70s, the most famous lymphokine was the "macrophage migration inhibitory factor" (MIF). Owing its name to its *in vitro* property, MIF had been first identified in 1966 as a T-lymphocyte-derived product associated with delayed-type hypersensitivity. In the absence of cell lines producing high amounts of MIF, its biochemical characterization was delayed. The cloning of this molecule occurred in 1989, long after its discovery. In the meantime, MIF-like activity was reported in cultures of fibroblasts, and S. Cohen et al. found a MIF-like activity in the supernatants of virus-loaded cell lines. Then, it appeared evident that lymphokines were incorrectly named because many cell types other than lymphocytes could produce similar bioactivities. And, in 1974, S. Cohen coined the name "cytokine." Indeed, MIF is a good example of a cytokine produced by many cell types, including macrophages, dendritic leukocytes, endothelial cells, epithelial cells, and fibroblasts. Most fascinating was the rediscovery of MIF as a product of the pituitary gland, further reinforcing the concept of cytokine and illustrating that, sometimes,

the border between cytokines and hormones may be difficult to draw.

Names of factors were traditionally reflecting their various biologic activities identified in different laboratories. This approach resulted in an imprecise and redundant nomenclature. In 1979, during the 2nd international Lymphokine Workshop in Ermatingen (Switzerland), it was decided that the name "interleukin" (IL-) be given to two different characterized factors. Under the new term IL-1 were gathered numerous names or acronyms defined according to one of the properties of this cytokine: osteoclast activating factor (OAF), lymphocyte activating factor (LAF), B-cell activating factor (BAF), epidermal thymocyte activating factor (ETAF), T-cell replacing factor III (TRFIII), leukocyte endogenous mediator, endogenous pyrogen, mitogenic protein, catabolin, hemopoietin-1, and so on. The name interleukin (between leukocytes) was obviously not the best choice since these factors were not restricted to leukocytes, but it had been mainly chosen by immunologists too self-absorbed, who were considering the immune system working independently of the other systems. When interviewed about it, Stanley Cohen specified: "It is a common misconception that the term interleukin was accepted for adoption by the nomenclature committee. When it was proposed, both Byron Waksman and I, as Co-chairs, felt it would turn out to be too restrictive (even then) and there was general agreement by the committee that it would be premature to refer to mediators as interleukins. This view was conveyed to the proposers. However, shortly thereafter, an editorial appeared in Journal of Immunology stating that the term 'interleukin' had been discussed and considered by the committee (true) and that henceforth the term would be used in scientific communications by the scientists writing the editorial. This was a true-true-unrelated kind of position to take, but it caught on in the general scientific community."

The last neologism, "chemokine," was created in 1992 to gather under one name a large family of small biochemically related cytokines, all possessing chemoattractant properties.

The availability of the first cloned molecules allowed major progresses, not only in fundamental research but also for clinical applications. While growth hormone was the first recombinant molecule used in humans in 1980, interferon-alpha was the first cytokine injected in humans in 1981, and IL-2 the second in 1985. Indeed, in 1984, 10 000 L of supernatants of activated Jurkat cell were required to prepare 30 mg of natural IL-2; the following year, 10 L of supernatants of recombinant *Escherichia coli* expressing the *IL-2* gene were sufficient to prepare 1 g of recombinant IL-2.

2 A Universal Language of Cells

2.1 Definitions

Cytokines have been defined by immunologists as cooperative factors produced by activated lymphocytes, dendritic leukocytes, macrophages that induce, modulate, favor, increase, inhibit cell proliferation, cell differentiation, or cell activities such as antibody production. The onset of the messages delivered by these soluble (or membrane-bound) molecules relies on the binding to their specific receptors displayed on the surface of the immune cells, a process leading to an intracellular

signaling cascade. However, far from being limited to the leukocyte lineages, this is a universal language of cells of the whole body, independently of their lineages, their nature, and their localization. This is probably the main difference between cytokines and other soluble molecules such as neuromediators mainly produced by the brain and the neurons, or hormones mainly produced by the endocrine system (Table 1).

Like hormones, cytokines can act in an endocrine fashion, but their actions mainly occur within the cellular microenvironment in a paracrine fashion. However, since some cytokines can be expressed on the cell surface of the producer cell, cytokines can act in a juxtacrine fashion within a close contact between cells. Finally, cytokines can act as an autocrine factor for a cell that both produces a cytokine and displays its receptor (Fig. 1).

Cell interactions involve more than one cytokine. One can consider each cytokine as a word, and the full sentence that will finally emerge is the reflection of the delivery of different cytokine-mediated signals. Together, these signalings lead to orientate the activities of the target cell. Thus, the order of delivery of cytokines has major consequences on the nature of the message. Furthermore, the concentration of cytokines, their localization, the nature of the target cell, and the timing are important parameters that affect the exact nature of the cytokine-initiated process.

Tab. 1 Comparison between hormones and cytokines.

	Sources	***Targets***	***Activities***	***Action***
Hormones	Secreted by a unique cell lineage	Specificity rather limited to one single type of target cell	Single action	Endocrine
Cytokines	Produced by many cell types	Numerous target cells	Wide spectrum of activities (redundancy)	Paracrine Juxtacrine Autocrine Endocrine

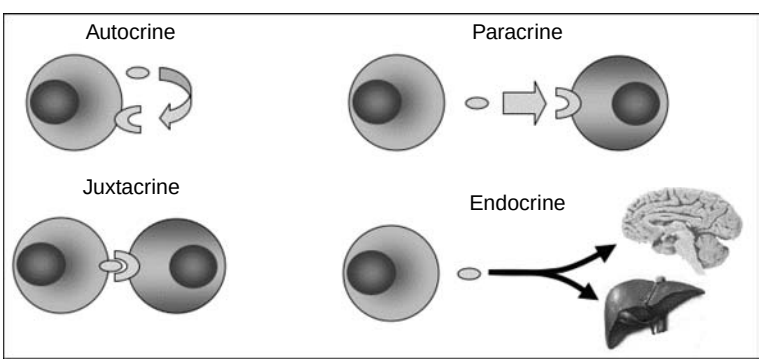

Fig. 1 Mode of action of cytokines.

In most cases, constitutive production of cytokines is limited and most cytokines are produced upon activation of cells. However, highly sensitive readout assays have allowed to determine a constitutive production at homeostasis in numerous tissues and cells.

2.2 Families

The classification of cytokines can be achieved according to (1) a common bioactivity: antiviral activity (interferons), chemoattraction (chemokines), hematopoiesis (colony-stimulating factors, CSF); (2) a common biochemical structure, suggesting a common ancestral molecule (chemokines, tumor necrosis factor (TNF) family); or (3) the sharing of a common chain of a receptor (gp130 cytokine family) (Table 2). Different subfamilies of interferons are known. Twenty-six genes of IFNα have been identified, of which at least five are pseudo genes. Interferons-alpha are produced by leukocytes and interferon-beta is produced by fibroblasts. IFNγ produced during immune response is a potent macrophage activator. IFNω is induced by virus, IFNτ is produced by trophoblast, and IFNκ is expressed by keratinocytes. Among chemokines, four different subfamilies can be identified on the basis of the highly conserved presence of the first two cysteine residues, which are either separated or are not separated by other amino acids: the CCL, CXCL, CX3CL, and the CL

Tab. 2 Families of cytokines.

Interferons	IFNα, IFNβ, IFNγ, IFNδ, IFNκ, IFNτ, IFNλ
Interleukins[1]	IL-1α, IL-1β, IL-1Ra, IL-18, IL-1F5-10 | IL-2, IL-15, IL-21
	IL-10, IL-19, IL-20, IL-22, IL-26 | IL-12, IL-23, IL-27
	IL-17A-E, IL-25 | IL-28, IL-29 IL-3, IL-4, IL-5, IL-6, IL-7
	IL-8 (CXCL8), IL-9, IL-11, IL-13, IL-14, IL-16
Hematopoietic factors	M-CSF, G-CSF, GM-CSF, Stem cell factor
Chemokines	CCL1 . . . CCL28 ; CXCL1 . . . CXCL16 ; XCL1, XCL2 ; CX3CL1
TNF superfamily	TNF, Ltα, Ltβ, NGF, CD27L, CD30L, CD40L, CD137L, APRIL, BAFF, FasL, GITRL, LIGHT, OPGL TRAIL, RANKL, TWEAK, VEGI.
Growth factors	TGF$\beta_{1,2,3}$
Gp130 cytokine family	IL-6, IL-11, CNTF, LIF, Oncostatin-M, Cardiotrophin-1

[1] Biochemically related interleukins are framed.
APRIL: apoptosis-inducing ligand; BAFF: B cell-activating factor; CCL: chemokine CC ligand; CNTF: ciliary neurotrophic factor; CSF: colony-stimulating factor; CXCL: chemokine CXC ligand; GITRL: glucocorticoid-induced TNF receptor family related gene ligand; IL-1F: IL-1 family; LIF: leukemia inhibitory factor; LIGHT : lymphotoxins, inducible expression, competes with HSVglycoprotein D for HVEM, a receptor expressed on T-lymphocyte; Lt: lymphotoxin; OPGL: osteoprotegerin ligand; RANKL: receptor activator of NF-kappaB ligand; TGFβ : transforming growth factor-β; TRAIL: TNF-related apoptosis-inducing ligand; TWEAK: TNF-like weak inducer of apoptosis; VEGI: vascular endothelial cell growth inhibitor.

chemokines. The assigned numbers of interleukins only reflect the timing of their discovery. Thus, some interleukins are also hematopoietic factors (e.g. IL-3 previously named multi-CSF), chemokines (IL-8, also named CXCL8), or even interferon (IL-28 and IL-29 are identical to IFNλ). The first interleukins have been discovered as bioactive factors, and there are no similarities in their amino acid sequences. However, there are some homologies in tertiary and quaternary structures and similar helical structures are found. The more recently discovered interleukins have been identified in gene banks according to sequence homology to previously identified interleukins. The TNF ligand superfamily gathers 18 distinct members that include soluble molecules with biochemical homology as well as membrane-bound ones. There are few other cytokines that can be displayed at the plasma membrane of the producing cells: IL-1α, IL-10, IL-15, and IFNγ. The TNF ligand superfamily also includes a growth factor (i.e. nerve growth factor). Usually, growth factors have a rather narrow spectrum of activity, are produced constitutively, are not acting on hematopoietic cells, and thus are distinct from cytokines. One exception is the transforming growth factor-β family (TGFβ), particularly because at least one member, namely TGFβ1, displays anti-inflammatory properties.

Localization of genes on chromosomes is known, and often genes of a family are gathered on the same chromosome (e.g. IL-1α, IL-1β, IL-1 receptor antagonist on human chromosome 2; most CXCL chemokines on human chromosome 4; most CCL chemokines on human chromosome 17). Most cytokines are glycosylated, their molecular weight ranging between 7 and 30 kDa. Most cytokines are monomers (e.g. IL-1, IL-2, IL-3, IL-4...), and few are homodimers (e.g. IL-8) or homotrimers (e.g. TNF). Some are heterodimers like IL-12, IL-23, and IL-27: IL-12 and IL-23 sharing the same p40 chain. Some are heterotrimers like lymphotoxin-β (Ltβ) that is the association of Ltα and Ltβ chains.

3 Receptors

3.1 Families

Most cytokine receptors had been characterized once their ligands had been identified, but there were few receptors, called *orphan receptors*, that were known before their ligands (e.g. c-kit, the receptor of stem cell factor; CD40, the receptor of CD40L; IL-1 receptor-related protein-1, part of the IL-18 receptor, etc). In most cases, receptors are the association of more than one transmembrane chain. Usually, one chain is essential for the binding (α-chain), while the second chain is required to initiate the intracellular signaling (β- or γ-chain). Some chains are common for receptors of few cytokines; this is the case of the gp130 chain of the receptors of IL-6, IL-11, CNTF, LIF, Oncostatin-M, and Cardiotrophin-1, or the gamma chain of IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21. Cytokine receptors are characterized by domains derived from ancestral common genes. In terms of evolution, it would be interesting to decipher which mutations in genes coding for cytokines or for cytokine receptors were first fixed. A common domain, called *hemopoietin receptor superfamily domain* characterized by the presence of four cysteine residues and a conserved sequence tryptophan-serine-x-tryptophan-serine, defines a family of numerous cytokine receptors (Fig. 2).

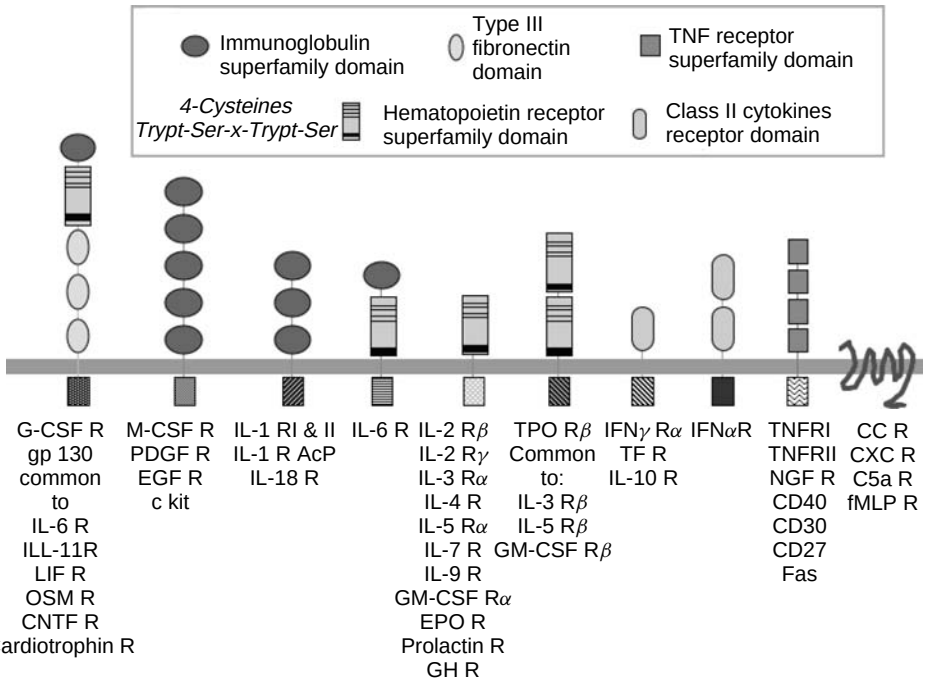

Fig. 2 Families of cytokine receptors.

Most interestingly, this family of receptors also includes receptors for hormones (e.g. growth hormone, prolactin) or the hematopoietic factor (e.g. erythropoietin). This domain can be duplicated (the common β chain of IL-3R, IL-5R, and GM-CSFR) or be associated with other domains like an immunoglobulin-like domain (IL-6Rα) or a fibronectin type III–like domain (gp130, G-CSFR). Another important family is the TNF receptor family. The activation by TNF is consecutive to the bridging of similar chains of the TNF receptor by the cytokine. Each cytokine receptor is highly specific for its ligand, except the TNFR p55 and p75 that bind both TNF and Ltα, and the chemokine receptors that can bind various chemokines. Chemokine receptors, constituted by a 7-transmembrane chain, are similar to receptors for other chemoattractant molecules such as the anaphylatoxin C5a or the bacterial fMet-Leu-Phe peptide. Cytokine receptors are often internalized after ligation to their ligands, and their expression can be induced, down- or upregulated.

3.2 Signaling

Upon binding of the cytokine to the extracellular domain of receptors, the intracellular receptor domains become associated with a variety of signaling molecules. A cascade of phosphorylation of various adaptor proteins occurs following interaction of the intracellular domain of the receptors with cytoplasmic tyrosine kinases (e.g. Janus kinases, Jak). Phosphorylation of latent cytoplasmic transcriptional activators (e.g.

signal transducers and activators of transcription, STAT) allows their dimerization and their translocation into the nucleus where they bind to specific sequences present in promoters of certain genes. Other pathways involving the mitogen-activated protein kinase (MAPK) and other kinases lead to the activation of various transcription factors (e.g. c-fos/c-jun, NF-κB, etc). Some signaling pathways are shared with other receptors. This is the case of the IL-1 and IL-18 receptors of which the intracellular domain is similar to the toll-like receptors (Toll–IL-1 Receptor, TIR domain). Thus, similar adaptors (e.g. MyD88, IRAK, etc) are involved after interaction of the respective receptors with their ligands, either members of the IL-1 family or pathogen-associated molecular patterns (PAMPs) such as endotoxin (lipopolysaccharide). Intracellular domain of some receptors (TNFR p55; Fas; TRAIL R) possesses a death domain that initiates a specific signaling cascade, leading to apoptosis.

3.3 Soluble Receptors

Receptors are integral part of the cytokine network (*cf* Sect. 7): not only can their membrane expression be modulated but they also exist as soluble molecules. Indeed, most cytokine receptors, except the chemokine receptors, can be shed from the cell surface after enzymatic cleavage following or not following a neosynthesis. As shown in Fig. 3, depending upon the nature of the receptor, its soluble form can behave as an inhibitor (e.g. soluble IL-1R, soluble TNFR) and can prevent the cytokine from reaching the membrane form. It can also be a carrier, protecting the cytokine within a complex, and increasing

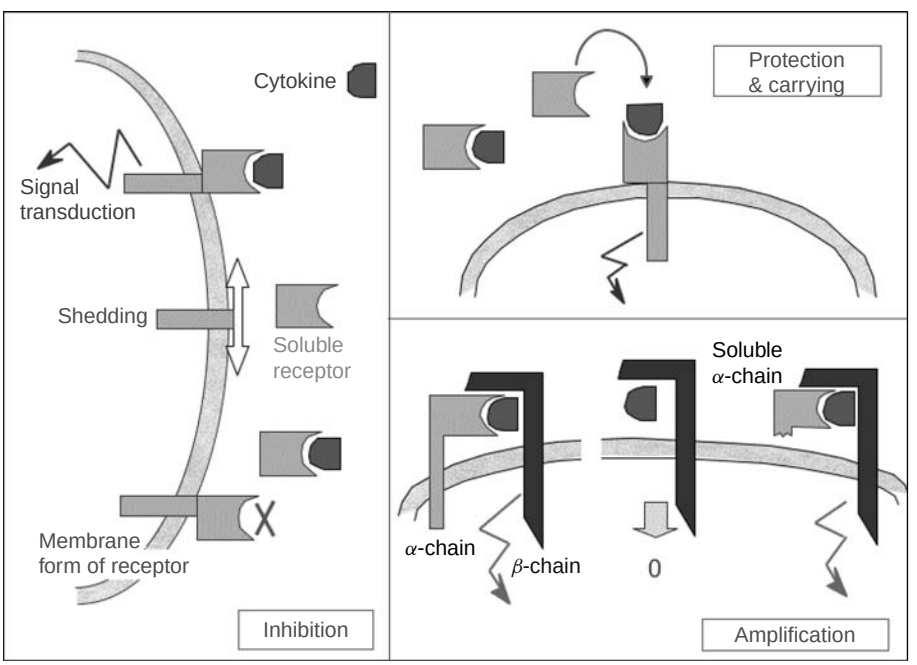

Fig. 3 Different properties of soluble receptors.

its half-life in the plasma or other biological fluids. Then, the dissociation of the cytokine-soluble receptor complex can allow the release of the cytokine close to a membrane receptor (e.g. TNFR). Lastly, in the case of a receptor constituted by one α-chain that binds the cytokine and one β-chain that recognizes the bound cytokine, the soluble α-chain receptor can enhance the responsiveness to the cytokine or can even allow the responsiveness of cells missing the α-chain, but expressing the β-chain. One example of the latter case is provided by the soluble IL-6R that, together with IL-6, controls the switch of leukocyte recruitment during inflammation.

4 Functions

4.1 Hematopoiesis

Production of leukocytes within the bone marrow is a very active and efficient process (4×10^8 white cells/leukocytes per hour in humans). Numerous cytokines, expressed at homeostasis in human bone marrow contribute to hematopoiesis, while others (like GM-CSF) are only active during infection to further increase the renewal of available leukocytes. Maturation, proliferation, and differentiation of bone marrow progenitor cells are under the control of different cytokines acting concomitantly. For example, IL-1 alone is unable to allow the differentiation of bone marrow precursors cells, but it acts in synergy with other hematopoietic cytokines. IL-3 and stem cell factor (also named *c-kit ligand*) are cytokines with a wide spectrum of activity, acting on pluripotent hematopoietic stem cells. GM-CSF favors the differentiation of myeloid progenitors cells, while M-CSF or G-CSF are specific of monocyte/macrophage and granulocyte lineages respectively. IL-6, IL-11, and mainly thrombopoietin favor thrombocytopoiesis, leading to megakaryocyte development and platelet production. Interleukin-5 has been identified as a major regulator of eosinophil development and function. Interleukin-7 is the main cytokine of lymphopoiesis required for the development of B, alpha-beta, and gamma-delta T-lymphocytes. Most of these hematopoietic cytokines also act as activators of mature cells and as antiapoptotic ligands.

4.2 Immune Response

4.2.1 Innate Immunity

Innate immunity is characterized by a response localized within the site of infection (Fig. 4). Epithelial cells and leukocytes residing in the extravascular space of the tissue (mast cells, resident macrophages, and immature dendritic leukocytes), and endothelial cells of the microvessels are triggered by microorganisms and microbial-derived products and release inflammatory cytokines (e.g. IL-1, TNF, IL-12, IL-18, etc). These cytokines further activate the endothelial cells of the postcapillary venules, leading to the increased adherence of circulating leukocytes and to coagulation process. Other cytokines, such as IFNγ, are produced within an amplificatory loop. Local production of chemokines contributes to the recruitment of new leukocytes from the blood compartment to further prevent the infectious process. Cytokines contribute to enhance the antimicrobial activities of newly recruited neutrophils and monocytes/macrophages (e.g. production of free radicals, enhanced phagocytosis), and the

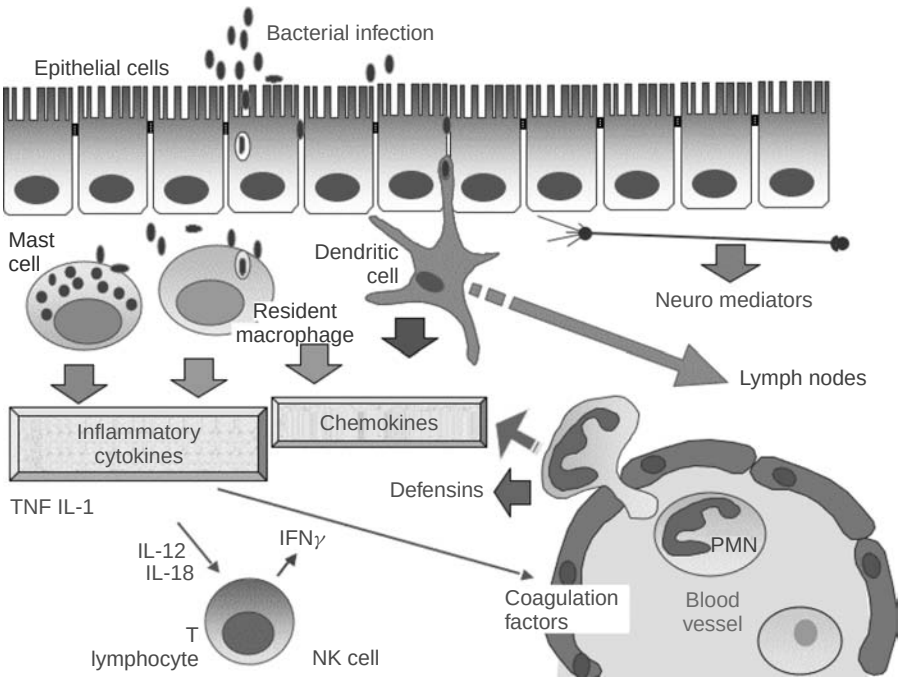

Fig. 4 Cytokines and soluble mediators as coordinators of regulated processes taking place in a bacteria-loaded sites.

antiviral activity of natural killer (NK) cells. Human leukocytes release defensins, a kind of natural antibiotics that contribute to limit bacterial dissemination. Neuromediators help control the inflammatory process. In addition to local responsiveness, a systemic response occurs, which includes fever, a reflection of the presence of pyrogenic cytokines within the plasma (e.g. IL-1, TNF, IL-6). Hyperthermia contributes to the anti-infectious process by enhancing some immune cell activities, and by reducing bacterial growth or viral replication. Finally, the production of hematopoietic cytokines (e.g. GM-CSF) further increases the number of leukocytes otherwise known to contribute to the clearance of the microorganisms and to the initiation of the healing process.

4.2.2 Adaptive Immunity

Adaptive immunity is characterized by processes relying on unique properties of the secondary lymphoid organs. Briefly, in secondary lymphoid organs such as the peripheral lymph nodes, professional antigen-presenting cells (dendritic leukocytes and mononuclear phagocytes), after processing of the native antigen, display on their membranes antigenic peptides and interact with T-lymphocytes, expressing a T-cell receptor able to specifically recognize these peptides. B-lymphocytes recognize native antigen. As a result of the cross talks between antigen-presenting cells, T- and B-lymphocytes, T-lymphocytes proliferate and differentiate as primed effector or memory T-cells, and B-lymphocytes proliferate and differentiate as antibody-secreting plasma

cells. All these events are controlled by cytokines. For example, within the secondary lymphoid organs, IL-2 and IL-4 favor T-cell proliferation, IL-12 activates cytotoxic T-cells, IL-2 and IL-4 allow B-cell proliferation, IL-4, IL-5, IL-6, IL-13, and TGFβ influence plasma cell differentiation and favor different immunoglobulin classes synthesis. Within microorganisms-loaded tissues, the effector T-lymphocytes, once reactivated could release IFNγ and GM-CSF that activate professional phagocytes conferring them microbicidal functions.

4.2.3 Th1 and Th2 Cytokines

In 1986, Tim Mosmann et al. described two types of murine helper CD4$^+$ T-cell (Th) clones defined according to the profile of released cytokines. Th1 cells produce IL-2, IFNγ, and Ltα, and Th2 cells produce IL-4, IL-5, IL-6, IL-10, and IL-13. Both subpopulations produce IL-3 and GM-CSF. Their differentiation from a Th0 precursor occurs, depending upon the cytokine environment (Fig. 5). Th1 and Th2 cells control each other: IFNγ derived from Th1 neutralizes Th2 cells, and IL-4 and IL-10 inhibit Th1 cells. Th1 cells are involved in cellular immunity and activate macrophages, Th2 cells are involved in humoral immunity and activate B-lymphocytes. Once stably differentiated as long-lived clones, some chemokine receptors allow to distinguish between these subpopulations: CXCR3 is mainly expressed on Th0 and Th1 cells, while CCR3 and CCR4 are expressed on Th2 cells. The Th1/Th2 dichotomy also exists in humans, as nicely illustrated by the early detection of Th2 cytokine mRNA expressed in response to an intradermal allergen challenge (immediate hypersensitivity) and the delayed detection of Th1 cytokine mRNA after tuberculin injection (delayed-type hypersensitivity). However, in most clinical settings, this dichotomy is not that obvious and mixed patterns are observed. More recently, other T-cell subsets with immunosuppressive properties have been characterized. Th3-type cells are a unique T-cell subset that primarily secretes TGFβ, provides help for IgA switch, and has suppressive properties for Th1 and other immune cell. Type-1 regulatory T-cells (Tr1) inhibit Th1 and Th2-type immune responses through the secretion of IL-10. They are also able to secrete TGFβ.

4.3 Cell Survival and Cell Death

Not only can cytokines be maturation and differentiation factors but they can

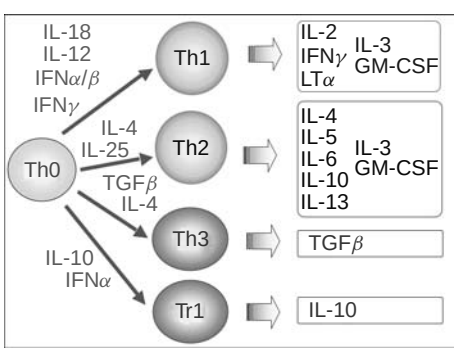

Fig. 5 T-cell subpopulations and the nature of environmental cytokines required for their differentiation.

also contribute to cell survival. For example, IL-2, IL-4, IL-6, IL-7, and IL-15 have all been shown to inhibit resting T-cell death *in vitro*. B-cell survival factors include IL-4 and B cell-activating factor (BAFF) of the TNF family. Plasma cell longevity depends on combination of IL-5, IL-6, stromal cell-derived factor-1α (SDF-1/CXCL12), and TNFα. IL-15 can sustain NK cell survival in the absence of serum. Indeed, *in vitro* serum starvation often leads to apoptosis that can be prevented by cytokines like TGFβ (e.g. survival of epithelial cell). Indeed, these properties reflect an antiapoptotic mechanism initiated by cytokines like the suppression of caspase activity or the maintenance or induction of Bcl-2 protein expression. RANKL suppresses apoptosis of primary cultured endothelial cells; SDF-1/CXCL12 enhances the survival of myeloid progenitor cells *in vitro*; stem cell factor (SCF) regulates the survival of cellular lineages by suppressing apoptosis. In addition to SCF, most hematopoietic factors allow survival of hematopoietic progenitors. In addition to sustain cell survival, cytokines can allow cell growth. For example, this is the case of IL-2 for T-cells, or IL-8 and some other chemokines that favor angiogenesis. Cytokines can also be growth factors for tumor cells; this is, for example, the case of IL-6 for malignant myeloma plasma cells. In contrast, certain cytokines can induce apoptosis of a variety of tumor cells and normal cells. Indeed, TNF was first identified in 1984 as a cytokine with antitumor effects, and other members of the TNF superfamily can induce proliferation, survival, as well as cell death: specifically, TNFα, FasL, TRAIL, and VEGI. The intracytoplasmic portion of their receptors contains a "death domain." Upon activation by their ligands, the death domain recruits intracytoplasmic adaptors expressing a death domain: TRAF (TNF receptor-associated factor) and FADD (Fas-associated death domain), which in turn recruit the proform of caspase-8. Autoactivation of caspase-8 leads to the subsequent activation of caspase-3, a proapoptotic enzyme.

4.4 Link with the Central and Peripheral Nervous System

A cross talk exists between immune leukocytes and the central nervous system. Proinflammatory cytokines, particularly IL-1 and TNF, induce fever, anorexia, and slow wave sleep. Part of their activities is conveyed by the vagal nerve. Fever involves local production of IL-6 and prostaglandins E2 (PGE2). Certain chemokines are also pyrogenic (e.g. IL-8, MIP-1α, MIP-1β, RANTES), but do not induce PGE2. IL-1 and TNF also induce a neuroendocrine loop in response to their signal, hypothalamus produces a corticotropin-releasing factor (CRF), and CRF induces the release of adrenocorticotropin hormone (ACTH) by pituitary gland. ACTH induces the release of glucocorticoid by adrenals. Glucocorticoids inhibit the production of most cytokines and antagonize many of their activities. Vagal nerve releases acetylcholine that represses inflammatory cytokine production. In addition, other neural and neuronal mediators, adrenaline, vasoactive intestinal peptide (VIP), pituitary adenylate cyclase-activating polypeptide (PACAP) are inhibitors of IL-1 and TNF production, while noradrenaline and substance P are potentiators.

4.5 Reproduction

There is compelling evidence that cytokines are involved in spermatogenesis, ovogenesis, and gestation. Cytokines play

an important regulatory role in the development and normal function of the testis. They are produced by Leydig cells, Sertoli cells, and germinal cells. Proinflammatory cytokines, including IL-1 and IL-6, have direct effects on spermatogenic cell differentiation and testicular steroidogenesis. Stem cell factor and leukemia inhibitory factor, cytokines normally involved in haematopoiesis, also play a role in spermatogenesis. Anti-inflammatory cytokines of the TGFβ family are involved in testicular development. TNFα, a secretory product of round spermatids, increases endogenous androgen receptors expression in primary cultures of Sertoli cells. Given the requirement of testosterone for spermatogenesis and the importance of androgen receptors in mediating Sertoli cell responsiveness to testosterone, the stimulation of androgen receptors expression by TNFα may represent an important regulatory mechanism required to maintain efficient spermatogenesis. M-CSF is the principal growth factor regulating macrophage populations in the testis, male accessory glands, ovary, and uterus. Both male and female M-CSF deficient (op/op) mice have fertility defects. Males have low spermatozoid number and libido as a consequence of dramatically reduced circulating testosterone. Females have extended estrous cycles and poor ovulation rates.

Normal ovarian tissue is rich in cytokines. Cytokines and chemokines are important in the physiology of ovarian function and of ovulation. Actual rupture of the follicle during ovulation may be dependent on tissue remodeling that shares some features with an acute inflammatory process. TNF and IL-1 are implicated in ovarian follicular development and atresia, ovulation, steroidogenesis, and corpus luteum function (including formation, development, and regression). Chemokines such as MCP-1 are also involved in luteolysis.

A great number of cytokines are produced by endometrium and placenta. Cytokines released at the fetomaternal interface play an important role in regulating embryo survival, controlling not only the maternal immune system but also angiogenesis and vascular remodeling. For example, LIF plays a role in the embryo pre-, peri-, and postimplantation, as does IFNτ in bovine and ovine. IFNτ is produced constitutively by embryonic trophectoderm during the period immediately prior to implantation. It acts on uterine epithelium to suppress transcription of the genes for estrogen receptor and oxytocin receptor. It blocks development of the uterine luteolytic mechanism. TGFβ is involved in regulating placental development and functions. The abortive influences of TNFα and IFNγ may terminate pregnancy during infection of uteroplacental unit. The slight dominance of proinflammatory cytokines in the fetal membranes and decidua suggest that inflammatory processes occur modestly with term labor, but much more robustly in preterm delivery, particularly in the presence of intrauterine infection.

4.6 Inflammation

"*Notae vero inflammationis sunt quatuor: rubor and tumor cum calore et dolore*" (Celsus, first century B.C.). Redness, swelling, heat, and pain are the four main parameters that characterize inflammation. They reflect the action of numerous inflammatory mediators and proinflammatory events that are orchestrated by inflammatory cytokines, mainly IL-1 and TNF, but also IL-12, IL-18, and IFNγ (Fig. 6). These cytokines are produced following

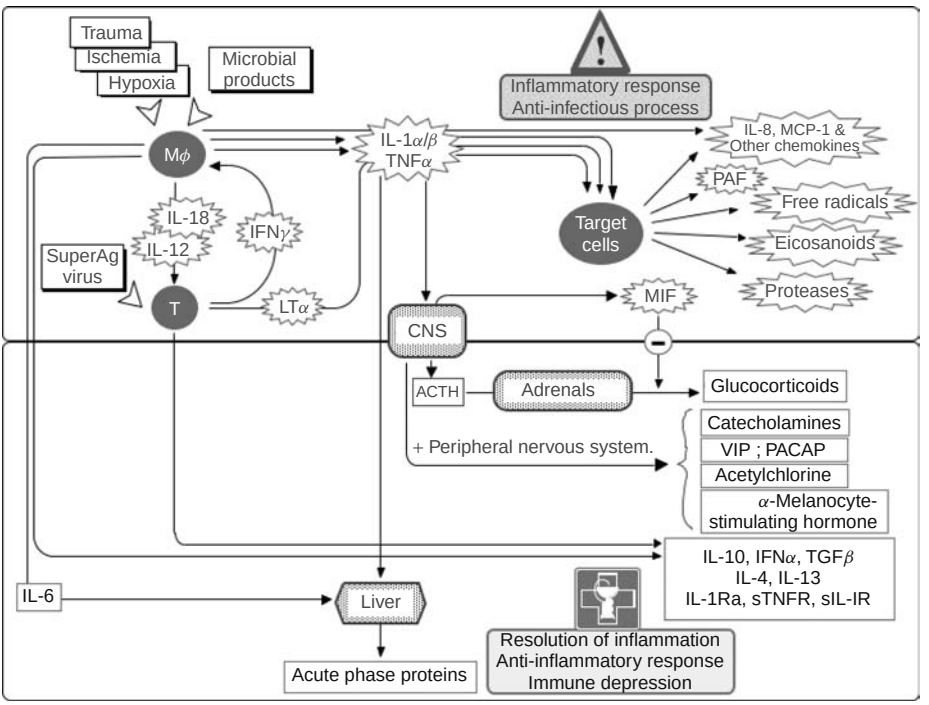

Fig. 6 Inflammation is the consequence of a cascade of events initiated by IL-1 and/or TNF. The action of these cytokines on various target cells leads to the release of numerous inflammatory mediators. Anti-inflammatory cytokines, the activation of the neuroendocrine pathway and the effects of glucocorticoids as well as the enhanced production of acute-phase proteins control negatively the inflammatory process (ACTH: adrenocorticotropic hormone; CNS: central nervous system; PACAP: pituitary adenylate cyclase-activating polypeptide; PAF: platelet activating factor; VIP: vasoactive intestinal peptide).

tissue or systemic steady state disruption like tissue loading by pathogens, ischemia, hypoxia, or trauma. Acting on target cells, they induce a cascade of mediators, including other cytokines and chemokines, proteases, lipid mediators (e.g. prostaglandins, platelet activating factor), free radicals (e.g. superoxide anion, nitric oxide) that contribute to the inflammatory process and tissue injury. Acting on endothelial cells, especially those of the postcapillary venules, they increase vascular permeability and plasma transudation, they increase circulating leukocyte adherence and margination, and they initiate the coagulation process by inducing tissue factor expression. Chemokines contribute to the recruitment of leukocytes that further maintain the inflammatory process. Some cytokines enhance the production of IL-1 and TNF (e.g. IFNγ, GM-CSF, IL-3), while others inhibit it (e.g. IL-4, IL-10, IL-13, TGFβ, IFNα). The latter contribute to attenuate the inflammatory process and to allow the tissue-repair process to occur. Other anti-inflammatory mediators include soluble IL-1 and TNF receptors, IL-1 receptor antagonist (IL-1Ra), some neuropeptides (e.g. adrenaline, acetylcholine), glucocorticoids induced after activation of the neuroendocrine loop by IL-1 and TNF,

and the acute-phase proteins produced by the hepatocytes in response to IL-1, TNF, and particularly IL-6. Acute-phase proteins contribute to eliminate debris due to cell lysis, to favor phagocytosis, to neutralize toxic mediators, and to inhibit proteases. Recently, IL-1Ra was recognized as one of the acute-phase proteins. In vivo, in the endotoxin shock model, most of these molecules have been shown to be protective.

5 Life without Cytokines (What Knockout Mice Tell Us)

Deletion of genes encoding for a cytokine or a cytokine receptor has led to the generation of many different knockout (K0) mice that further allowed to decipher the contribution of cytokines to different physiological events. In most cases, viable, fertile, and clinically healthy mice could be obtained. In few cases, the deletion of the genes led to death *in utero*, illustrating a major role of certain cytokines during embryogenesis. This was the case for SDF1 (CXCL12) and its receptor CXCR4, LIF, Cardiotrophin-1, and the gp130 chain of their receptors (see Table 3). Deletions of certain signaling molecules (e.g. STAT3) are also lethal during embryonic life. Other deletions reveal the role of certain cytokines for organogenesis during embryonic life. This is the case of lymphotoxin-α, lymphotoxin-β, and LtβR for the development of lymph nodes and Peyer's patches, and RANKL for the development of lymph nodes and mammary glands.

Tab. 3 Cytokine or cytokine receptor KO mice reveal relative contribution of cytokines to life processes.

Biological events	*Deleted gene*	*Consequences*
Embryogenesis	LIF	Absence of LIF in female prevents implantation of blastocytes
	Cardiotrophin-1 & gp130 receptor chain	Heart development blocked; death *in utero* on day 16
	SDF-1 & CXCR4	Defects in heart and brain development, intestinal vascularization, and B-cell hematopoiesis. Death *in utero*
Organogenesis	Ltα, Ltβ, LtβR	Absence of lymph nodes and Peyer's patches
	TNFα	Altered spleen and lymph node organization
	RANKL	Absence of lymph nodes, mammary gland defects
	IL-8R	Splenomegaly, lymphoadenopathy, increased circulating neutrophils
Hematopoiesis	G-CSF	Chronic neutropenia
	Stem cell factor	Absence of mast cells
	M-CSF	Altered function of monocytes and osteoclast
	IL-7	Reduced number of lymphocytes

(*continued overleaf*)

Tab. 3 (*Continued*)

Biological events	*Deleted gene*	*Consequences*
Leukocyte differentiation	Ltα, Ltβ IL-15 γ-chain	Absence of follicular dendritic cells Absence of NK and NK-T cells Impaired B, T and NK cell development
Osteoclastogenesis	RANK & RANKL	Defect in the differentiation of hematopoietic osteoclast progenitor
Pulmonary homeostasis	GM-CSF & β-chain of GM-CSF R	Accumulation of surfactant lipids and proteins in the alveolar space. Lymphoid hyperplasia.
Gut homeostasis	IL-10	Intestinal mucosal hyperplasia, inflammatory cell infiltration, MHC Class II expression on colonic epithelium.
	IL-2	Colitis
	IL-15	Reduced intestinal epithelial B and T lymphocytes
Fever	IL-6	No pyrogenic response
Acute-phase response	IL-6	Profound alteration of acute-phase proteins production
Inflammation	TGFβ	Systemic lethal inflammation (day 24)
	IL-1-Ra	Chronic arthropathy, spontaneous dermatosis, exacerbated delayed-type hypersensitivity
	IL-5	Reduced eosinophilia
Antibody production	IL-4, IL-6, IL-13, CD40L	Altered antibody production
Immune function	IL-2, IL-4, IL-6, IL-12, IL-13, IL-15, IL-18, TNF, IFNγ, Stem cell factor, CSF, Chemokines, and their receptors.	Altered anti-infectious response

Organization of hematopoietic compartments are under the control of numerous cytokines and absence of TNFα, or the γ-chain of the receptor of IL-3, -5, -7, -9, -15, and -21 have profound perturbations on tissue organization and leukocyte differentiation. This is also true for the deletion of chemokines that contribute to tissue colonization at homeostasis. While studies of KO mice often comforted the knowledge, there were some surprises. One of them was the discovery that the absence of the hematopoietic cytokine GM-CSF did not alter hematopoiesis (mice had normal numbers of peripheral leukocytes, bone marrow progenitors, and tissue hematopoietic populations), but led to a profound alteration of the lung status, demonstrating a role of GM-CSF in pulmonary homeostasis. Of major interest was the demonstration that the deletion of anti-inflammatory cytokines (IL-10, TGFβ, IL-1Ra) was deleterious at homeostasis in the absence of experimental exogenously delivered inflammatory signals, establishing their crucial role to prevent

potential ongoing inflammatory processes at homeostasis. As expected, numerous deletions had major effects on the quality of the anti-infectious process. For example, in the absence of IL-6, mice produce lower levels of IgG antibodies against stomatitis vesicular virus, display a lower cytotoxic T-cell activity during vaccinia virus infection, leading to an increased number of virus in lungs, and failed to control *Listeria monocytogenes* load in the tissues they reached. In contrast, the same deficiencies may appear beneficial in certain experimental models of septic shock in response to endotoxin (e.g. TNFα, TNFR p55, IL-1β converting enzyme).

In humans, natural mutations occur that may affect cytokines, signaling molecules of receptors. In the latter case, the most famous mutation leads to the deletion of the γ-chain of the receptor of IL-3, -5, -7, -9, -15, and -21 and is associated with a severe combined immunodeficiency. This is for this clinical setting that occurred in 2000, the first successful gene therapy.

6 Cytokine Synthesis

6.1 Homeostasis

Owing to highly sensitive techniques (RT-PCR, *in situ* hybridization, ELISpot), it has been possible to show the presence of cytokine mRNA in various types of cells or to demonstrate the presence of cytokine-producing cells in the absence of activation. For example, IL-6 is spontaneously produced by 0.5% of bone marrow cells, 0.1% of spleen cells, and 0.01% of mesenteric lymph node cells. IL-6 is also produced in the absence of exogenous stimuli by enterocytes, eosinophils, neutrophils, epidermal cells, smooth muscle cells, bone marrow stromal fibroblasts, anterior pituitary cells, trophoblast cells, and so on. Furthermore, at homeostasis, certain cells contain cytokines. This is particularly the case for keratinocytes that contain large amounts of IL-1α. Mast cells contain a large panel of preformed cytokines (e.g. IL-1, IL-4, IL-5, IL-6, IL-13, TNF, etc), and MIF is preformed in numerous leukocytes. Furthermore, the presence of cytokines in biological fluids has been reported in absence of any infection or inflammatory diseases: IL-1α and IL-8 are found in sweat; IL-1α, IL-1β, IL-6, IL-8, GM-CSF, and TGFβ have been reported in tears; IL-2, IL-8, TNFα, TGFβ, and soluble TNFR are present in saliva; IL-2, IL-6, IL-8, IL-10, IL-12, TNF, and soluble IL-2R and IL-6R have been detected in seminal fluid; and a great number of different cytokines have been identified in human colostrum. In the later case, all of these cytokines probably act on the oropharyngeal and gut-associated lymphoid tissue of the newborn and favor the development and maturation of the immune system and may protect the newborn against invasive microorganisms delivered by the oral route. Finally, certain chemokines are present in tissues where they contribute to leukocyte recruitment at homeostasis (e.g. CCL17 & CCL25 in thymus, CCL21 & CXCL13 in lymph nodes, and Peyer's patches, CXCL12 in numerous tissues).

6.2 Induction

The production of cytokines by professional antigen-presenting leukocytes and

activated T-cells occurs during the cellular cooperation required to initiate the adaptive immunity. This production is representative of the dialogue between immune leukocytes. In contrast, innate immunity is associated with the production of cytokines in response to exogenous activators, mainly the "pathogen associated molecular patterns" (PAMPs). Following their interaction with the "pattern recognition receptors" (PRR), various intracellular signaling cascades lead to the synthesis and the release of cytokines. The best-known PRRs are the toll-like receptors (TLR) that share with IL-1R and IL-18R a homologous Toll-IL-1R (TIR) domain. Ten different TLRs have been identified so far in humans that recognize bacterial, viral, parasitic, or fungal conserved structures expressed on the surface or within the microbes (e.g. DNA, single- or double-strand RNA). Endotoxin of Gram-negative bacteria (lipopolysaccharide, LPS) is among the most potent PAMPs and picograms of this molecule are sufficient to induce the production of a whole panel of cytokines by macrophages. LPS is trapped by the CD14 molecule on the cell surface and triggers the cell via a complex MD2/TLR4. In the case of Gram-positive bacteria, some of these bacteria release exotoxins, also known as superantigens that trigger the production of cytokines by both T-lymphocytes and macrophages.

6.3 Parameters that Affect Production

In addition to the genetic polymorphism (see Sect. 8), there are numerous parameters that influence the level of the production of cytokines. Of course, *in vitro* experiments require careful sampling of the cells. The use of medium and agents free of endotoxin is very important. There are probably numerous wrong statements that suggest the capacity of a given molecule to induce the production of cytokines that are consecutive to their contamination by endotoxin, as it has been shown for heat shock proteins. *Ex vivo* studies of cytokine production by human cells is influenced by age, nutrition status, the use of drugs, alcohol, and smoking habit. Seasonal and circadian rhythms, physical exercise, altitude exposure, microgravity, social and psychological stress, physical stress, and gender affect cytokine production. In the later cases, neuromediators and sexual hormones modify cellular production of cytokines.

6.4 Measurements

The measurement of cytokines in humans can be performed in natural biological fluids, in induced biological fluids (e.g. broncho-alveolar or peritoneal lavages), on tissue biopsies, and with blood leukocytes. The detection of cytokines in biological fluids may represent the "tip of the iceberg," and can be detected in the case of exacerbated production since, once produced, cytokines are efficiently trapped by their specific receptors on environmental cells. Cytokine mRNA analysis can be achieved by Northern blots, *in situ* hybridization or reverse transcriptase polymerase chain reaction (RT-PCR). Cytokine-producing cells can be monitored by ELISpot or by flow cytometry. Biological assays are not performed anymore and, nowadays, the measurement of cytokines is mainly achieved by enzyme-linked immunosorbent assays (ELISA). Cytokines constitute a tightly regulated network (see below). Thus, information concerning cytokines should be achieved through the analysis

of simultaneous quantification of several cytokines. Indeed, it is the cytokine milieu that influences the cellular response rather than the action of a single cytokine. Techniques recently available such as microarrays combined to real-time PCR or multiplex immunological detection of cytokines might give precious information in a near future.

7 The Cytokine Network

As illustrated in Fig. 7, the interactions between cytokine-producing cells and target cells lead to define a cytokine network. Once produced in the context of a specific immune response or following microbial activation, cytokines act on target cells and induce synthesis of new cytokines. The induction by IL-1 of the release of IL-2 by lymphocytes, IL-6 by fibroblasts, G-CSF by macrophages, and IL-8 by endothelial cells are examples of cytokine cascades. Cytokines can induce their own synthesis, like IL-1, or further enhance the productions, leading to amplificatory loops, like IFNγ. Another key word to define this network is synergy. Owing to the great redundancy of cytokines, two cytokines can share similar activities. When acting together on the same target cell, the effect will be far greater than the only additive effect of each individual cytokine. For example, the production of complement factor C3 by endothelial cells in response to TNF and IL-1 is far higher than that induced by these cytokines alone. The synergy may be the consequence of the induction of cytokine receptors, allowing cells to respond to a second cytokine they would not do normally because they lack the specific receptor. The network also involves negative loops as a consequence of the action of cytokines, blocking the production or the effects of others. IL-10 is an example of a cytokine blocking the production and the action of other cytokines. However, the lineage of its target cell may lead to opposite observations. For example, IL-10 represses LPS-induced IL-8

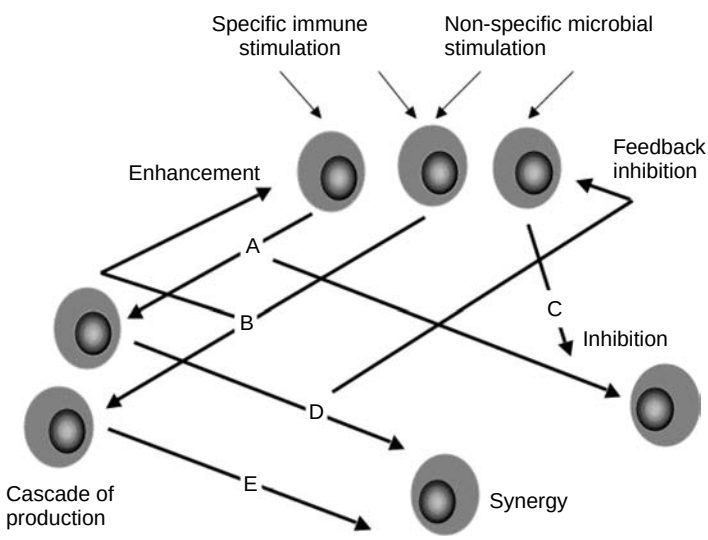

Fig. 7 The cytokine network (see Sect. 7).

production by monocytes, but enhances that produced by endothelial cells.

8 Individual Heterogeneity

Among genetic predisposition to diseases, premature death due to uncontrolled invasion and growth by microorganisms has the higher relative risk linked to heritability. The individual heterogeneity reflects in part genetic polymorphisms of cytokines. Three levels can be dissociated (Fig. 8) as follows:

1. A genetic polymorphism exists for the receptors of PAMPs. These sensors are the first essential elements in initiating an intracellular signaling cascade that leads to cytokine production. Certain single nucleotide polymorphisms (SNP) or mutations can directly modify the intensity of the responsiveness and influence disease susceptibility, as recently shown by the existence of rare mutations of TLR4 among patients with meningococcal infection.
2. SNP or mutations can affect cytokine genes, particularly when expressed within gene promoters or other regulatory sequences. Accordingly, high, intermediate, or low producers are reported when assessing the levels of cytokines produced in response to an activator. Correlation between TNF genotypes and levels of released TNF in response to LPS has been shown.
3. Finally, depending upon the donors, target cells can react intensively, moderately, or weakly to a given amount of a cytokine. This has been nicely demonstrated with genetically distinct endothelial cells, which express various levels of adhesion molecules in response to similar amounts of IL-1 or TNF. In all these situations, the genetic polymorphism could also reflect SNP or mutations among the genes of the adaptors and signaling molecules involved in the signaling cascades.

As illustrated in Table 4, numerous cytokine gene polymorphisms have been associated with the occurrence or severity of diseases. The reported SNP or mutations can be associated with a protection against the disease or in contrast with a higher susceptibility. More recently, similar gene polymorphisms have been associated with the efficiency or the lack of efficiency of certain therapeutic approaches, particularly when the treatment targets cytokines.

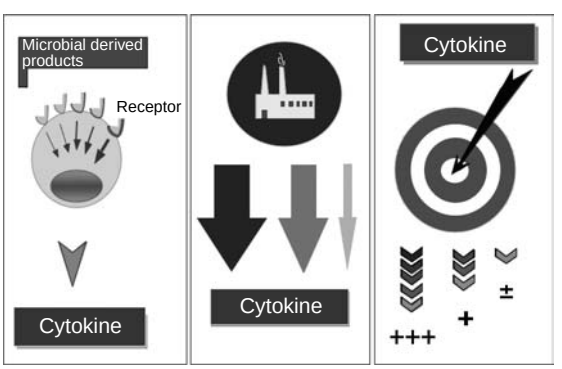

Fig. 8 Individual heterogeneity for cytokine production and responsiveness.

Tab. 4 Some examples of cytokine or cytokine receptor gene polymorphisms associated with diseases.

Disease	*Gene polymorphism*
Acute renal failure mortality	TNF, IL-10
Allergy	RANTES promoter; MIF promoter; IL-13 promoter IL-16 promoter; IL-1 gene complex; IL-18; IL-12B promoter
Asthma	IL4R α-chain; IL-10; IL-15; IL-18; TGFβ1 promoter; Eotaxin
Alzheimer	IL-1α, TNFR2, IL-6
Breast cancer	IL-6
Chronic periodontitis	IL-10 promoter, IL-1B +3953 and TNF-A-308 allele 2 positive
Coronary disease	IL-1Ra
Crohn's disease	IL-10, MCP-1 (CCL2); IL-16 promoter; TNFR1
HIV resistance	CCR5 deletion
Idiopathic pulmonary fibrosis	TNF-alpha (-308 A) allele
Infectious nephropathy	low CXCR1 expression
Infertility	functional mutations in the LIF gene
Lupus nephritis	MCP-1 (CCL2) promoter
Multiple sclerosis	microsatellite allele of TNF gene; IL-2 promoter
Parkinson's disease	promoter region of IL-8
Rheumatoid arthritis	noncoding region of IFNγ gene; TGFβ1; MIF promoter; IL-1 gene cluster; IL-4
Schizophrenia	IL-1 gene complex
Sepsis mortality	TNF
Systemic lupus erythematosus	IL-1α; TNFα
Sudden infant death syndrome	IL-10
Susceptibility to cerebral malaria	TNF
Transplantation	TNF, IL-10

9 Cytokines and Infections

9.1 Half Angel–Half Devil

As mentioned in Sect. 4.2.1, cytokines play a major role in innate immunity during the early processes that occur in sites of microorganisms' invasion. During viral infection, the antiviral properties of interferons have been established for more than half a century. After interaction of IFNα or IFNβ with its receptor, the activation of 2–5 A synthetase leads to the production of an endoribonuclease that degrades viral RNA, while activation of protein kinase P1 contributes to the capacity of the cell to inhibit viral protein synthesis. During bacterial infection, the beneficial effects of proinflammatory cytokines have been widely demonstrated. The injection of cytokines such as IL-1, IL-6, IL-12, IL-18, TNF, G-CSF, M-CSF, GM-CSF, and IFNγ before the injection of bacteria have been shown to promote faster clearance of bacteria, and to enhance survival. The use of cytokine-neutralizing antibodies or antagonists molecules further demonstrated that endogenous cytokines produced in the course of infections play an essential role in preventing the microorganisms to establish themselves, or to grow, and their neutralization was highly deleterious. Finally, mice rendered deficient for the expression of a cytokine or a cytokine

receptor display enhanced "susceptibility" to infection such as an enhanced microbial load in tissues, and reduced survival.

Despite their undoubtedly beneficial effects to clear microorganisms, some cytokines can also contribute to death in some clinical situations like the septic shock in response to endotoxin injection or after a lethal bacterial infection. Antibodies against TNF protect mice against lethal injection of LPS and baboons against lethal injection of *E. coli*. The ambivalent role of cytokines is illustrated in TNF receptor p55 knockout mice: these mice are more sensitive to *L. monocytogenes* infection than normal mice, but they are more resistant to toxic shock induced by endotoxin or Staphylococcal enterotoxin B. Furthermore, the synergy already described as a key word to define cytokine network is also valid for lethality and none lethal dose of TNF, injected with non-lethal dose of IL-1 or IFNγ leads to lethality in mice.

9.2
The Strategies of Microbes against Cytokines

The fact that pathogenic microorganisms have developed numerous subterfuges to counteract the action of cytokines further illustrates that cytokines are crucial in the fight against the infectious process. Hijacking of genes for cytokine receptors has been perpetrated by virus to elaborated soluble molecules that neutralize cytokines of the host (e.g. IFNγR and TNFR homologs of Myxoma virus, IL-1R homolog of Vaccinia virus). Another hijacking concerns the gene of IL-10, allowing viruses (Epstein–Barr virus, equine herpesvirus, parapoxvirus orf virus) to create an immunosuppressive environment. Another strategy has been developed by Vaccinia virus to prevent the maturation of biologically active IL-1 and IL-18. This virus produces an inhibitor (serpine) of caspase-1, the enzyme required for the cleavage of the proforms of these cytokines. Other mechanisms include the production of viral factors that downregulate cytokine mRNA expression (e.g. HTLV, adenovirus), the use of cytokines to further enhance the viral replication (e.g. HIV), the use of chemokine receptors to enter the cell (e.g. HIV), or the synthesis of viral chemokines that can behave as antagonists (e.g. Kaposi' sarcoma associated virus, human Herpesvirus 8, Steakth virus, etc.). Bacteria also display strategies to block cytokine production (e.g. *Yersinia enterocolitica*, enteropathogenic *E. coli*) and to use them as growth factors (*Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Acinetobacter*).

10
Cytokines and Diseases

As the discovery of new cytokines were occurring, great hopes to cure diseases were mentioned, often followed by great disappointments. For example, while TNF was discovered for its antitumor effect, it was soon demonstrated that it was one of the most potent proinflammatory cytokines, obviously limiting its therapeutic use. One unique cytokine is rarely the unique key effector underlying the pathogenic process, and the complexity of the cytokine network has rendered difficult their use as therapeutic tools or as therapeutic target. In addition, the identification of an overexpressed cytokine at one peculiar stage of the disease process does not allow discriminating whether this cytokine is a marker or an actor. Nevertheless, there exist certain clinical situations in which the use of recombinant cytokines or the use of

neutralizing antibodies has allowed significant therapeutic progresses. Few examples of diseases for which the involvement of cytokines is well established and clinical applications have been considered are summarized below.

10.1 Autoimmune Diseases

Among autoimmune diseases, systemic lupus erythematosus has been associated with an exacerbated production of IL-10. Considering the properties of IL-10, which activates B-lymphocytes and inhibits T-lymphocytes, its overproduction in lupus could explain the immune anomalies of this disease, which is characterized by anti-self antibodies production and by deficient T-responses. Thus, the use of IL-10 antagonists may be beneficial in the management of systemic lupus erythematosus, as suggested by a preliminary report. Human lupus patients also have elevated blood levels of BAFF that support differentiation of selected B-cells into mature long-lived B-cells and may be critical in generating deleterious autoimmune responses. Indeed, overexpression of BAFF in mice may lead to systemic lupus erythematosus–like disease. Thus, BAFF could be another potential cytokine to neutralize in systemic lupus erythematosus.

10.2 Allergy

Allergen specific IgE antibodies, mast cell degranulation releasing inflammatory mediators such as histamine, and influx of eosinophils in airway mucosa and airway lumen are a hallmark of atopy and allergic asthma. Thus, Th2 cytokines remain important candidates for a role in the pathogenesis of atopy and allergic asthma: IL-4 and IL-13 are required for IgE production, IL-9 is involved in mast cell growth, and IL-5 is required for eosinopoiesis. Humanized monoclonal antibodies (hMAbs) against IL-5, anti-IL-4, a recombinant soluble human IL-4 receptor, anti-IL-9, CCR3 antagonists (which block eosinophil chemotaxis), and CXCR2 antagonists (which block neutrophil and monocyte chemotaxis) have been developed as possible therapeutic interventions.

10.3 Sepsis

Animal models of sepsis have clearly shown that TNF, IL-1, or MIF are involved in sepsis-related lethality, and antibodies against these cytokines are highly protective. Unfortunately, anti-TNF antibodies and IL-1Ra have been used in numerous placebo double-blinded controlled assays in humans without any success. This failure may reflect that antibodies in animal models and in humans are not used with similar timings, and that mice, often used in experimental models, being 10^5 less sensitive to Gram negative bacteria–derived endotoxin than humans, are not the most relevant laboratory animals to model the human settings!

10.4 Chronic Inflammatory Diseases

Fortunately for companies that had developed these anti-TNF antibodies, successful treatments have been achieved for Crohn's disease and rheumatoid arthritis. The significant improvement of patients allowed the marketing of these antibodies. However, in none of these studies,

the successful improvement concerned all patients. For rheumatoid arthritis, the first successful treatment was reported in 1994. In responding patients, clinical improvements were associated with biological changes, reflecting the attenuation of the inflammatory process. There was no clinical response in 30% of the cases. While the effect of a single injection is long lasting (few months), it has to be repeated. As a consequence, the occurrence of microbial pathogenic processes (tuberculosis, *Pneumocystis*- triggered pneumonia, histoplasmosis, listeriosis, sepsis) has been deplored. Other approaches have been proposed to neutralize TNF with recombinant soluble TNF receptors. IL-1Ra has also been approved by the Food and Drug Administration (FDA) for treatment of rheumatoid arthritis. For this disease, the targeting of IL-17 has also been suggested to be of potential value. For Crohn's disease, the anti-TNF treatment was approved in 1998, and clinical response was still observed 12 weeks after the infusion. A remission could be obtained for half of the patients after 4 doses every 8 weeks.

10.5 Cancer

IL-2 has been extensively studied in cancer patients, either alone, or associated with chemotherapy, with activated NK cells (Lymphokine-activated killer cells, LAK), with tumor-infiltrating lymphocytes (TIL), with IFNα, and, more recently, with tumor-pulsed dendritic cells. In some specific cases (melanoma, metastatic renal cancer), success remained modest. The major achievement has been obtained with IFNα in hairy cell leukemia and chronic myelogenous leukemia.

10.6 Hepatitis C

IFNα has been successful in a subset of hepatitis C virus–infected patients. More recently, once associated to ribavirin, a polymerized form of IFNα, the half-life of which is prolonged, has been shown to lead to sustained virological response rates of $>50\%$ in chronic hepatitic C patients.

10.7 Multiple Sclerosis

In 1993, interferon-β was approved in the United States for relapsing–remitting multiple sclerosis. The use IFNβ and glatiramer acetate for the treatment of multiple sclerosis has, to some extent, changed the course of the disease. The annual relapse rate of patients treated with these drugs is lower than that in placebo-treated patients, and more treated patients remain relapse-free as compared to untreated patients.

10.8 Transplantation

The alpha-chain of the IL-2R is a specific peptide against which monoclonal antibodies have been raised, with the aim of blunting the immune response by means of inhibiting proliferation and inducing apoptosis in primed lymphocytes. One of such antibodies has proved to be effective in reducing the episodes of acute rejection after kidney and pancreas transplantation. The use of this antibody was associated with a significant reduction in the incidence of any treated rejection episodes after kidney transplantation in the two major randomized European and US studies.

10.9 Neutropenia

Following chemotherapy for leukemia, neutropenia renders the patients more susceptible to develop infection. Treatment of patients with G-CSF and GM-CSF accelerates the recovery of normal neutrophil counts and significantly reduces the occurrence of documented infections.

11 Conclusion

Despite the huge number of molecules that belong to the cytokine family, the complexity of the cytokine network, the great number of parameters that may affect the biological properties of cytokines, tremendous amount of work has allowed to decipher the language of cells. Despite the difficulty, new promising use of cytokines or cytokine antagonists can be expected to master certain diseases.

See also Cytokines: Interleukins; Growth Factors; Programmed Cell Death.

Bibliography

Books and Reviews

Aggarwal, B.B. (1998) *Human Cytokines: Handbook for Basic and Clinical Research*, Blackwell, Oxford, UK.

Bona, C.A., Revillard, J-P. (2000) *Cytokines and Cytokine Receptors: Physiology and Pathological Disorders*, Harwood Academic Publishers, Amsterdam, Netherlands.

Cavaillon, J.M. (1994) Cytokines and macrophages, *Biomed. Pharmacother.* **48**, 445–453.

Cavaillon, J-M. (2001) Pro- versus anti-inflammatory cytokines: myth or reality, *Cell. Mol. Biol.* **47**, 695–702.

Fitzgerald, K.A. (2001) *Cytokine Facts Book*, Academic Press, San Diego, CA.

Giamila, F., Durum, S.K. (2003) *Cytokine Knockouts*, Contemporary immunology, 2nd edition, Humana Press, Totowa, NJ.

Kotb, M., Calandra, T. (2003) *Cytokines and Chemokines in Infectious Diseases Handbook*, Humana Press, Totowa, NJ.

Oppenheim, J.J., Feldmann, M., Durum, S.K. (2000) *Cytokine Reference: A Compendium of Cytokines and Other Mediators of Host Defense*, Academic Press, London, San Diego, CA.

Santamaria, P. (2003) *Cytokines and Chemokines in Autoimmune Disease*, Landes Bioscience/Eurekah.com, Georgetown, TX.

Thomson, A.W., Lotze, M.T. (2003) *The Cytokine Handbook*, 4th edition, Academic Press, Amsterdam, Netherlands.

Primary Literature

Auron, P.E., Webb, A.C., Rosenwasser, L.J., Mucci, S.F., Rich, A., Wolff, S.M., Dinarello, C.A. (1984) Nucleotide sequence of human monocyte interleukin 1 precursor cDNA, *Proc. Natl. Acad. Sci. USA* **81**, 7907–7911.

Bartfeld, H., Atoynatan, T. (1969) Cytophilic nature of migration inhibitory factor associated with delayed hypersensitivity, *Proc. Soc. Exp. Biol. Med.* **130**, 497–501.

Bazan, J.F. (1990) Structural design and molecular evolution of a cytokine receptor superfamily, *Proc. Natl. Acad. Sci. USA.* **87**, 6934–6938.

Bender, J.R., Sadeghi, M.M., Watson, C., Pfau, S., Pardi, R. (1994) Heterogeneous activation thresholds to cytokines in genetically distinct endothelial cells: evidence for diverse transcriptional responses, *Proc. Natl. Acad. Sci. U S A.* **91**, 3994–3998.

Bernhagen, J., Calandra, T., Mitchell, R.A., Martin, S.B., Tracey, K.J., Voelter, W., Manogue, K.R., Cerami, A., Bvcala, R. (1993) MIF is a pituitary-derived cytokine that potentiates lethal endotoxaemia, *Nature* **365**, 756–759.

Bigazzi, P.E., Yoshida, T., Ward, P.A., Cohen, S. (1975) Production of lymphokine-like factors (cytokines) by simian virus 40-infected and simian virus 40-transformed cells, *Am. J. Pathol.* **80**, 69–78.

Burdin, N., Peronne, C., Banchereau, J., Rousset, F. (1993) Epstein-Barr virus transformation induces B lymphocytes to produce human interleukin 10, *J. Exp. Med.* **177**, 295–304.

Carswell, E.A., Old, L.J., Kassel, R.L., Green, S., Fiore, N., Williamson, B. (1975) An endotoxin-induced serum factor that causes necrosis of tumors, *Proc. Natl. Acad. Sci. USA* **72**, 3666–3670.

Cavazzana-Calvo, M., Hacein-Bey, S., de Saint Basile, G., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J.L., Bousso, P., Deist, F.L., Fischer, A. (2000) Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease, *Science* **288**, 669–672.

Chai, Z., Gatti, S., Toniatti, C., Poli, V., Bartfai, T. (1996) Interleukin (IL)-6 gene expression in the central nervous system is necessary for fever response to lipopolysaccharide or IL-1 beta: a study on IL-6-deficient mice, *J. Exp. Med.* **183**, 311–316.

Coffman, R.L., Ohara, J., Bond, M.W., Carty, J., Zlotnik, A., Paul, W.E. (1986) B cell stimulatory factor-1 enhances the IgE response of lipopolysaccharide-activated B cells, *J. Immunol.* **136**, 4538–4541.

Cohen, S., Bigazzi, P.E., Yoshida, T. (1974) Similarities of T cell function in cell-mediated immunity and antibody production, *Cell. Immunol.* **12**, 150–159.

DeForge, L.E., Kenney, J.S., Jones, M.L., Warren, J.S., Remick, D.G. (1992) Biphasic production of IL-8 in lipopolysaccharide (LPS)-stimulated human whole blood. Separation of LPS- and cytokine-stimulated components using anti-tumor necrosis factor and anti-IL-1 antibodies, *J. Immunol.* **148**, 2133–2141.

de Waal Malefyt, R., Figdor, C.G., Huijbens, R., Mohan-Peterson, S., Bennett, B., Culpepper, J., Dang, W., Zurawski, G., de Vries, J.E. (1993) Effects of IL-13 on phenotype, cytokine production, and cytotoxic function of human monocytes. Comparison with IL-4 and modulation by IFN-gamma or IL-10, *J. Immunol.* **151**, 6370–6381.

Dinarello, C.A., Renfer, L., Wolff, S.M. (1977) Human leukocytic pyrogen: purification and development of a radioimmunoassay, *Proc. Natl. Acad. Sci. U S A.* **74**, 4624–4627.

Doherty, G.M., Lange, J.R., Langstein, H.N., Alexander, H.R., Buresh, C.M., Norton, J.A. (1992) Evidence for IFN-gamma as a mediator of the lethality of endotoxin and tumor necrosis factor-alpha, *J. Immunol.* **149**, 1666–1670.

Dranoff, G., Crawford, A.D., Sadelain, M., Ream, B., Rashid, A., Bronson, R.T., Dickersin, G.R., Bachurski, C.J., Mark, E.L., Whitsett, J.A. (1994) Involvement of granulocyte-macrophage colony-stimulating factor in pulmonary homeostasis, *Science* **264**, 713–716.

Dumonde, D.C., Wolstencroft, R.A., Panayi, G.S., Matthew, M., Morley, J., Howson, W.T. (1969) "Lymphokines": non-antibody mediators of cellular immunity generated by lymphocyte activation, *Nature* **224**, 38–44.

Echtenacher, B., Männel, D.N., Hultner, L. (1996) Critical protective role of mast cells in a model of acute septic peritonitis, *Nature* **381**, 75–77.

Elliott, M.J., Maini, R.N., Feldmann, M., Kalden, J.R., Antoni, C., Smolen, J.S., Leeb, B., Breedveld, F.C., MacFarlane, J.D., Bijl, H. (1994) Randomised double-blind comparison of chimeric monoclonal antibody to tumour necrosis factor alpha (cA2) versus placebo in rheumatoid arthritis, *Lancet* **344**, 1105–1110.

Fernandez-Botran, R. (1991) Soluble cytokine receptors: their role in immunoregulation, *FASEB J.* **5**, 2567–2574.

Frendeus, B., Godaly, G., Hang, L., Karpman, D., Lundstedt, A.C., Svanborg, C. (2000) Interleukin 8 receptor deficiency confers susceptibility to acute experimental pyelonephritis and may have a human counterpart, *J. Exp. Med.* **192**, 881–890.

Gabay, C., Smith, M.F., Eidlen, D., Arend, W.P. (1997) Interleukin 1 receptor antagonist (IL-1Ra) is an acute-phase protein, *J. Clin. Invest.* **99**, 2930–2940.

Gerard, C., Bruyns, C., Marchant, A., Abramowicz, D., Vandenabeele, P., Delvaux, A., Fiers, N., Goldman, M., Velu, T. (1993) Interleukin 10 reduces the release of tumor necrosis factor and prevents lethality in experimental endotoxemia, *J. Exp. Med.* **177**, 547–550.

Hapel, A.J., Lee, J.C., Farrar, W.L., Ihle, J.N. (1981) Establishment of continuous cultures of thy1.2+, Lyt1+, 2-T cells with purified interleukin 3, *Cell* **25**, 179–186.

Horai, R., Saijo, S., Tanioka, H., Nakae, S., Sudo, K., Okahara, A., Ikuse, T., Asano, M., Iwakura, Y. (2000) Development of chronic inflammatory arthropathy resembling rheumatoid arthritis in interleukin 1 receptor antagonist-deficient mice, *J. Exp. Med.* **191**, 313–320.

Howard, M., Farrar, J., Hilfiker, M., Johnson, B., Takatsu, K., Hamaoka, T., Paul, W.E. (1982) Identification of a T cell-derived B cell growth

factor distinct from interleukin 2, *J. Exp. Med.* **155**, 914–923.

Hurst, S.M., Wilkinson, T.S., McLoughlin, R.M., Jones, S., Horiuchi, S., Yamamoto, N., Rose-John, S., Fuller, G.M., Topley, N., Jones, S.A. (2001) IL-6 and its soluble receptor orchestrate a temporal switch in the pattern of leukocyte recruitment seen during acute inflammation, *Immunity* **14**, 705–714.

Isaacs, A., Lindenmann, J. (1957) Virus interference I. The interferon, *Proc. Roy. Soc. B.* **147**, 258–267.

Jacob, C.O., Fronek, Z., Lewis, G.D., Koo, M., Hansen, J.A., McDevitt, H.O. (1990) Heritable major histocompatibility complex class II-associated differences in production of tumor necrosis factor alpha: relevance to genetic predisposition to systemic lupus erythematosus, *Proc. Natl. Acad. Sci. USA.* **87**, 1233–1237.

Kopf, M., Baumann, H., Freer, G., Freudenberg, M., Lamers, M., Kishimoto, T., Zinkernagel, R., Bluethmann, H., Kohler, G. (1994) Impaired immune and acute-phase responses in interleukin-6-deficient mice, *Nature* **368**, 339–342.

Kuhn, R., Lohler, J., Rennick, D., Rajewsky, K., Muller, W. (1993) Interleukin-10-deficient mice develop chronic enterocolitis, *Cell* **75**, 263–274.

Llorente, L., Zou, W., Levy, Y., Richaud-Patin, Y., Wijdenes, J., Alcocer-Varela, J., Morel-Fourrier, B., Brouet, J.C., Alarcon-Segovia, D., Galanaud, P. (1995) Role of interleukin 10 in the B lymphocyte hyperactivity and autoantibody production of human systemic lupus erythematosus, *J. Exp. Med.* **181**, 839–844.

Moore, K.W., Vieira, P., Fiorentino, D.F., Trounstine, M.L., Khan, T.A., Mosmann, T.R. (1990) Homology of cytokine synthesis inhibitory factor (IL-10) to the Epstein-Barr virus gene BCRFI, *Science* **248**, 1230–1234.

Mosmann, T.R., Cherwinski, H., Bond, M.W., Giedlin, M.A., Coffman, R.L. (1986) Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins, *J. Immunol.* **136**, 2348–2357.

Nakano, Y., Onozuka, K., Terada, Y., Shinomiya, H., Nakano, M. (1990) Protective effect of recombinant tumor necrosis factor-alpha in murine salmonellosis, *J. Immunol.* **144**, 1935–1941.

Ohlsson, K., Bjork, P., Bergenfeldt, M., Hageman, R., Thompson, R.C. (1990) Interleukin-1 receptor antagonist reduces mortality from endotoxin shock, *Nature* **348**, 550–552.

Pfeffer, K., Matsuyama, T., Kundig, T.M., Wakeham, A., Kishihara, K., Shahinian, A., Wiegmann, K., Ohashi, P.S., Kronke, M., Mak, T.W. (1993) Mice deficient for the 55 kDa tumor necrosis factor receptor are resistant to endotoxic shock, yet succumb to L. monocytogenes infection, *Cell* **73**, 457–467.

Pociot, F., Briant, L., Jongeneel, C.V., Mölvig, J., Worsaae, H., Abbal, M., Thomsen, M., Nerup, J., Cambon-Thomsen, A. (1993) Association of tumor necrosis factor and class II major histocompatibility complex alleles with the secretion of TNFα and TNFβ by human mononuclear cells: a possible link to insulin-dependent diabetes mellitus, *Eur. J. Immunol.* **23**, 224–231.

Porat, R., Clark, B.D., Wolff, S.M., Dinarello, C.A. (1991) Enhancement of growth of virulent strains of Escherichia coli by interleukin-1, *Science* **254**, 430–432.

Rosenberg, S.A., Lotze, M.T., Muul, L.M., Leitman, S., Chang, A.E., Ettinghausen, S.E., Matory, Y.L., Skibber, J.M., Shiloni, E., Vetto, J.T. (1985) Observations on the systemic administration of autologous lymphokine-activated killer cells and recombinant interleukin-2 to patients with metastatic cancer, *N. Engl. J. Med.* **313**, 1485–1492.

Sallusto, F., Lenig, D., Mackay, C.R., Lanzavecchia, A. (1998) Flexible programs of chemokine receptor expression on human polarized T helper 1 and 2 lymphocytes, *J. Exp. Med.* **187**, 875–883.

Shalaby, M.R., Waage, A., Aarden, L., Espevik, T. (1989) Endotoxin, tumor necrosis factor-alpha and interleukin 1 induce interleukin 6 production in vivo, *Clin. Immunol. Immunopathol.* **53**, 488–498.

Shirai, A., Holmes, K., Klinman, D. (1993) Detection and quantitation of cells secreting IL-6 under physiologic conditions in BALB/c mice, *J. Immunol.* **150**, 793–799.

Shull, M.M., Ormsby, I., Kier, A.B., Pawlowski, S., Diebold, R.J., Yin, M., Allen, R., Sidman, C., Proetzel, G., Calvin, D. (1992) Targeted disruption of the mouse transforming growth factor-beta 1 gene results in multifocal inflammatory disease, *Nature* **359**, 693–699.

Smith, C.A., Davis, T., Wignall, J.M., Din, W.S., Farrah, T., Upton, C., McFadden, G., Goodwin, R.G. (1991) T2 open reading frame from the Shope fibroma virus encodes a soluble form of the TNF receptor, *Biochem. Biophys. Res. Commun.* **176**, 335–342.

Stuber, F., Petersen, M., Bokelmann, F., Schade, U. (1996) A genomic polymorphism within the tumor necrosis factor locus influences plasma tumor necrosis factor-alpha concentrations and outcome of patients with severe sepsis, *Crit. Care Med.* **24**, 381–384.

Supajatura, V., Ushio, H., Nakao, A., Akira, S., Okumura, K., Ra, C., Ogawa, H. (2002) Differential responses of mast cell Toll-like receptors 2 and 4 in allergy and innate immunity, *J. Clin. Invest.* **109**, 1351–1359.

Tracey, K.J., Fong, Y., Hesse, D.G., Manogue, K.R., Lee, A.T., Kuo, G.C., Lowry, S.F., Cerami, A. (1987) Anti-cachectin/TNF monoclonal antibodies prevent septic shock during lethal bacteraemia, *Nature* **330**, 662–664.

Tsicopoulos, A., Hamid, Q., Haczku, A., Jacobson, M.R., Durham, S.R., North, J., Barkans, J., Corrigan, C.J., Meng, O., Moqbel, R. (1994) Kinetics of cell infiltration and cytokine messenger RNA expression after intradermal challenge with allergen and tuberculin in the same atopic individuals, *J. Allergy Clin. Immunol.* **94**, 764–772.

Upton, C., Mossman, K., McFadden, G. (1992) Encoding of a homolog of the IFN-gamma receptor by myxoma virus, *Science* **258**, 1369–1372.

van Dullemen, H.M., van Deventer, S.J., Hommes, D.W., Bijl, H.A., Jansen, J., Tytgat, G.N., Woody, J. (1995) Treatment of Crohn's disease with anti-tumor necrosis factor chimeric monoclonal antibody (cA2), *Gastroenterology* **109**, 129–135.

Wang, H., Yu, M., Ochani, M., Amella, C.A., Tanovic, M., Susarla, S., Li, J.H., Wang, H., Yang, H., Ulloa, L., Al-Abed, Y., Czura, C.J., Tracey, K.J. (2003) Nicotinic acetylcholine receptor alpha7 subunit is an essential regulator of inflammation, *Nature* **421**, 384–388.

Yoshimura, T., Matsushima, K., Oppenheim, J.J., Leonard, E.J. (1987) Neutrophil chemotactic factor produced by lipopolysaccharide (LPS)-stimulated human blood mononuclear leukocytes: partial characterization and separation from interleukin 1 (IL 1), *J. Immunol.* **139**, 788–793.

Molecular Motors in Plant Cells

Anireooy S.N. Reddy
Colorado State University, Fort Collins, CO

1 Microtubule-based Motors 463
1.1 Kinesins 463
1.1.1 Phylogenetic Analysis 469
1.2 Dyneins 475

2 Actin-based Motors 477
2.1 Myosins 477
2.1.2 Phylogenetic Analysis 478

3 Cellular Roles of Motors 481
3.1 Cell Division 481
3.2 Cell Polarity and Morphogenesis 484
3.3 Cytoplasmic Streaming 485
3.4 Microtubule Dynamics and Organization 485
3.5 Intercellular Transport 486
3.6 Other Functions 486

4 Regulation of Motors 487
4.1 Calcium/Calmodulin 487
4.2 Protein Phosphorylation 488

5 Concluding Remarks 488

Acknowledgment 489

Bibliography 489
Books and Reviews 489
Primary Literature 489

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

Keywords

Cytoskeleton
A dynamic network of protein tubules (microtubules) and filaments (microfilaments and intermediate filaments) in all eukaryotic cells that is necessary for cell shape, cell motility, and spatial and temporal organization of molecules, and organelles within the cells.

Microtubules
Long hollow cylindrical tubes with an external diameter of 24 nm that are formed by polymerization of α- and β-tubulin monomers.

Microfilaments
Filaments of about 7-nm diameter that are formed by polymerization of actin.

Kinesins
A family of motor proteins in eukaryotes that can hydrolyze ATP and use the derived energy to either transport organelles, vesicles, and RNA protein complexes on microtubules or regulate microtubule dynamics.

Myosins
A family of motor proteins that can hydrolyze ATP and move on microfilaments. These motors function in muscle contraction, cell division, and vesicle transport.

Preprophase Band
A narrow band of interconnected microtubules and microfilaments encircling the nucleus that appears in cells preparing to enter mitosis and marks the location of future cell wall.

Phragmoplast
A plant-specific structure made of two discs of microtubules (and microfilaments) that is formed between two daughter nuclei at late telophase. It constructs the new cell wall between the daughter nuclei.

Trichome
A hair-like branched or unbranched structure on the surface of a leaf or stem. These structures may be unicellular or multicellular.

Molecular motors regulate diverse cellular functions, including the organization and dynamics of the microtubule (MT) and actin cytoskeleton, cytoplasmic streaming, cell polarity, cell growth, morphogenesis, chromosome segregation, and transport of vesicles, organelles, and macromolecular complexes. In eukaryotes, there are three families of molecular motors: the kinesins, the dyneins, and the myosins. All three types of motors use ATP to move along filamentous structures. Kinesins and dyneins move on MTs, whereas the myosins translocate on actin filaments. Using

biochemical, cell biological, molecular, and genetic approaches, several molecular motors have been identified in plants, and functions of some of these are beginning to be understood. Recent completion of genome sequence of several eukaryotes ranging from yeast, a simple unicellular eukaryote, to highly evolved multicellular organisms, including humans and flowering plants, has permitted comparative analysis of three families of motors in various species. Such analysis across phylogenetically divergent species has yielded some interesting insights into evolutionary and functional relationships among motor proteins from various organisms. Systematic analyses of the recently completed Arabidopsis genome sequence with the conserved motor domain of kinesins, myosins, and dyneins revealed the presence of 78 molecular motors (61 kinesins and 17 myosins) in this organism. Of the two families of MT-based motors, dyneins are absent in Arabidopsis. Surprisingly, Arabidopsis has the largest number of kinesins as compared to other multicellular organisms including humans, suggesting that the kinesin superfamily is expanded considerably in plants. Also, all plant myosins belong to two novel classes. Although the identification of molecular motors in plants is progressing rapidly due to genome and EST-sequencing projects, only a few plant motors have been characterized in any detail and the functions of many these motors are not known. Nevertheless, it is becoming obvious that plants contain novel families of kinesins and myosins with unique functions and regulatory mechanisms. In this article, the focus is primarily on plant molecular motors, their function, and regulation with emphasis on model plant Arabidopsis.

1 Microtubule-based Motors

The motors that move on MTs (microtubules) belong to two different families: kinesins that move either toward plus end or minus end of the MTs, and the dyneins, which move toward the minus end of MTs.

1.1 Kinesins

Members of the kinesin superfamily are known or implicated to play important roles in many fundamental cellular and developmental processes, including intracellular transport of vesicles and organelles, mitotic and meiotic spindle formation and elongation, chromosome segregation, germplasm aggregation, MT organization, and dynamics and intraflagellar transport. A large number of kinesins have been identified in plants and animals (Table 1). All members of the kinesin superfamily have a highly conserved catalytic region of about 350 amino acid residues, known as the *motor domain*, with ATP- and MT-binding sites. The motor domain in kinesins is located either in the N-terminus, C-terminus, or in the middle of the protein. In addition to the motor domain, most kinesins have a stalk region that forms an alpha-helical coiled-coil region, which aids in dimerization, and a highly variable tail, which is thought to interact with a specific cargo.

Although molecular motors have been implicated in a variety of cellular processes

Tab. 1 The total number of kinesins and myosins in some of the completely sequenced organisms.

	S. cerevisiae	*S. pombe*	*C. elegans*	*D. melanogaster*	*H. sapiens*	*A. thaliana*
Kinesins	6	9	21	24	45	61
Myosins	5	5	17	13	40	17
Total	11	14	38	37	85	78

in plants, including spindle function, cytokinesis, cell polarity, morphogenesis, organelle and vesicle transport, and cytoplasmic streaming, until recently very little was known about their molecular identity. Plant kinesins have been identified in pollen tubes of tobacco and other plants using antibodies to animal kinesins. An MT-based motor (90 kD) from tobacco pollen tubes, which induced MT gliding in motility assays, binds organelles associated with MTs in the cortical regions of the pollen tube, suggesting its involvement in organelle transport. Two kinesins (125 kD and 120 kD) that showed plus-end motor activity in *in vitro* motility assays and MT-dependent ATPase and GTPase activity were isolated from tobacco phragmoplasts. Using various approaches, cDNAs for several kinesins have been characterized from plants. Thirteen of these are from Arabidopsis (KatA, KatB, KatC, KatD, AtKCBP, AtPAKRP1, AtPAKRP1L, AtPAKRP2, AtMKRP1, AtMKRP2, AtNCAK1, KCA1, KCA2), 15 from tobacco (TKRP125, KCBP, NCAK1, NCAK2, and TBK1 to TBK11), and 1 from potato (KCBP). A kat gene family (katA, katB, katC, katD, and katE) encoding kinesins in *Arabidopsis thaliana* was characterized using primers corresponding to conserved regions of the kinesin motor domain. Using a similar approach, 12 different kinesins that belong to seven subfamilies have been isolated from tobacco BY-2 cells. It is interesting that in tobacco BY-2 cells alone 15 different kinesins that belong to different subfamilies are expressed. KatA (89 kD) protein has been shown to be a minus-end-directed motor, and a similar protein has been found in carrot and tobacco. KatA (new name ATK1) is necessary for bipolar spindle assembly during male meiosis. The central region of KatD shares sequence similarity to the motor domain of kinesin and is followed by about 240 residues at the C-terminus. Phylogenetic analysis with the katD motor domain indicates that it belongs to the C-terminal family despite the fact that it has a 240 amino acid stretch following the motor domain. The amino-terminal region of katD showed sequence similarity to the calponin homology (CH) domain. KatD is a flower-specific motor protein, suggesting that it may function in transport or cytoskeleton organization in pollen. TKRP125 (tobacco kinesin–related polypeptide of 125 kD) is an N-terminal kinesin with strong similarity to members of the BimC (blocked in mitosis) subfamily. Two other members of BimC class kinesins (DcKRP120-1 and DcKRP120-2) were isolated from carrot suspension cells. DcKRP120-1 is a homolog of TKRP125. In Arabidopsis, there seem to be at least three TKRP-like proteins. Recently, three phragmoplasts-associated kinesin-related proteins (AtPAKRP1, AtPAKRP1L, and AtPAKRP2) have been isolated and characterized from Arabidopsis. AtPAKRP1 is an N-terminal kinesin and is most closely related to

XKlp2, an ungrouped kinesins from *Xenopus laevis*. AtPAKRP2 is also an N-terminal motor with no known homologs in non-plant systems. In phylogenetic analysis, AtPAKRP2 does not group with any of the existing families of kinesins. These motors have distinct functions in cytokinesis. The AtPAKRP1L that is closely related to AtPAKRP1 shows same localization as AtPAKRP1, suggesting that these two kinesins may have a similar role in cell division.

KCBP (kinesin-like calmodulin-binding protein), a novel kinesin with a C-terminal motor domain, was isolated from a number of plants as a calmodulin-binding protein. Although KCBP, like other kinesins, contains three distinct regions (a motor domain, a coiled-coil stalk, and a tail), it has two features that make it unique among members of the kinesin superfamily in eukaryotes. These include (1) a calmodulin-binding domain adjacent to the motor domain at the C-terminus and (2) a myosin tail homology and talin-like region in the N-terminal tail. KCBP binds calmodulin in a calcium-dependent manner at physiological calcium concentrations. KCBP is a minus-end directed MT motor and has two MT-binding domains, one located at the C-terminus and the second one located at the N-terminus. Unlike the MT-binding domain in the C-terminus, the N-terminal region of KCBP binds MTs, both in the presence and absence of ATP, indicating that the MT-binding domain in the N-terminus is insensitive to ATP. KCBP is highly conserved in phylogenetically divergent plant species, including dicots, monocots, gymnosperms, and algae. Recently, a calmodulin-binding C-terminal kinesin (kinesin C) was cloned from sea urchin. The calmodulin-binding domain of kinesin C shared 35% sequence identity with the calmodulin-binding domain in KCBP. The existence of calmodulin-binding kinesins in both plants and sea urchins suggests that the origin of this group of kinesins predates the divergence of plants and animals from a common ancestor, which is believed to have occurred about 1.5 billion years ago. If this was the case, there must have been insertion of some domains such as MyTH4 and talin-like regions in the tail of KCBP (or deletion of these domains in kinesin C) to acquire functional specialization of these kinesins. Alternatively, calmodulin-binding kinesins may have evolved independently in plant and animals after they diverged from a common ancestor. The amino-terminal tail and stalk regions of KCBPs from different plant systems are highly conserved and contain myosin tail homology (MyTH4) and talin-like regions that are not present in kinesin C. In phylogenetic analysis of kinesins, KCBPs and kinesin C are grouped with other known C-terminal kinesins. However, Arabidopsis KCBP together with its orthologs from other plants constitute a distinct group within the C-terminal subfamily of motors.

Recently, two Arabidopsis kinesins (AtMKRP1 and AtMKRP2) with mitochondrial targeting signals were found to be targeted to mitochondria, suggesting that these motors might function in this organelle. AtMKRP1 and AtMKRP2 belong to ungrouped kinesins. What functions these MKRPs perform in mitochondria remains to be seen. There are other Arabidopsis kinesins with targeting signals to other organelles such as chloroplasts. It would be interesting to see if any other Arabidopsis kinesins are targeted to organelles. Kinesins that are targeted to organelles have not been reported previously in the literature.

Mutations in the *HINKEL* (*HIK*) gene in Arabidopsis result in defective cytokinesis. These mutants have high frequency of incomplete cell walls and multinucleate cells. The *HIK* gene encodes a plant-specific N-terminal kinesin that plays an important role in cytokinesis. Two tobacco kinesins (NACK1 and NACK2) interact with NPK1, a mitogen-activated kinase kinase kinase, and stimulate the activity of the kinase. Orthologs of NACK1 and NACK2 have been identified in Arabidopsis and AtNACK1 was found to be identical to HIK. These kinesins are highly diverged N-terminal motors and do not group with any of the known kinesin subfamilies. The closest nonplant kinesin that shares significant similarity with the motor domain of NACKs is CENP-E. At least eight Arabidopsis kinesins are known to function in some aspect of cell division (see Sect. 3.1 on cell division).

Analysis of the recently completed Arabidopsis genome sequence with the conserved motor domain of kinesins has resulted in identification of 61 kinesins in the model plant. A corresponding cDNA or EST (expressed sequence tag) was found for most kinesins, suggesting that most kinesins in Arabidopsis are expressed and not likely to be pseudogenes. The genes encoding kinesins are distributed throughout the genome. Chromosome 3 has the highest number (18) of kinesin genes. In comparison, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, *Drosophila*, and *Homo sapiens* have 6, 9, 21, 24, and 45 kinesins respectively (Table 1). Surprisingly, the Arabidopsis genome contains the largest number of kinesins among all eukaryote genomes that have been sequenced. Arabidopsis has the highest percentage (0.24%) of the total number of genes as compared to *S. cerevisiae* and *S. pombe* with 0.1% and 0.17% respectively, *C. elegans* with 0.11%, and *Drosophila* with 0.18%. On the basis of the number of kinesins in Arabidopsis, the number of cellular processes that is regulated by kinesins in plants is likely to match or exceed the number of processes controlled by animal kinesins.

Only 13 of the 61 Arabidopsis kinesins have been reported in the literature, whereas the 3 AtKRP125 kinesins (AtKRP125a,b,&c; Table 2) show sequence similarity to a tobacco kinesin (NtKRP125) that was isolated from phragmoplasts of tobacco. AtKRP125b has 68% identity with NtKRP125 over the 1000 residues they have in common.

The number of known and predicted introns in kinesins ranges from 3 to 34 (Table 2). In the Arabidopsis genome, the number of introns ranges from 0 to 77 with an average of about 5. More than 85% of Arabidopsis genes have 10 or less introns, while the Arabidopsis kinesin genes have an average of 16.4 with only 10 genes having 10 or less introns. The number and locations of the introns need to be verified experimentally. It is likely that the predicted sizes of the proteins in Tables 2 and 3 may change somewhat as corresponding cDNAs are characterized. In a few cases in which the cDNAs were isolated recently, there are some differences between the predicted and actual size of the protein due to inaccurate predictions of introns and exons.

All Arabidopsis kinesins except three have a coiled-coil region, suggesting that they may function as homo or heterodimers (Fig. 1 and Table 2). Interestingly, some Arabidopsis kinesins have predicted domains that are not present nonplant kinesins (Fig. 1, Table 2). Six of the Arabidopsis kinesins have a calponin homology (CH) domain, which is an actin-binding domain present in the N-termini

Tab. 2 Kinesin-like proteins in arabidopsis.

	Gene code (published name)	*Motor location*	*#aa*	*Number of introns*	*Motility*	*Cellular localization*	*Other Domains*
1	**At2g37420 (AtKRP125a)**	N-terminal	1022	17	ND	ND	CC
2	**At2g36200 (AtKRP125b)**	N-terminal	1056	22	ND	ND	CC
3	**At2g28620 (AtKRP125c)**	N-terminal	1076	18	ND	ND	CC,
4	At3g45850	N-terminal	1058	21	ND	ND	CC
5	At4g05190	C-terminal	777	17	ND	ND	CC,
6	**At4g21270 (AtKatA)**	C-terminal	793	8	Minus[a]	Mitotic MT arrays	CC
7	**At4g27180 (AtKatB)**	C-terminal	745[d]	15	ND	ND	CC
8	**At5g54670 (AtKatC)**	C-terminal	754[d]	15	ND	ND	CC
9	At5g27550	N-terminal[b]	425	6	ND	ND	CC
10	At2g22610	Internal[c]	1068	18	ND	ND	CC
11	At1g72250	Internal[c]	1195	17	ND	ND	CC,
12	At3g10310	Internal[c]	897	16	ND	ND	CC,CH
13	At1g18410	Internal[c]	1162[d]	17	ND	ND	CC
14	At1g73860	Internal[c]	1050	17	ND	ND	CC
15	At1g63640	Internal[c]	1056	19	ND	ND	CC,CH
16	At5g41310	Internal[c]	967	19	ND	ND	CC,CH
17	**At5g27000 (AtKatD)**	Internal[c]	987	17	ND	ND	CC,CH
18	At1g09170	Internal[c]	1032[d]	18	ND	ND	CC,CH
19	At2g47500	Internal[c]	861	14	ND	ND	CC,CH
20	At3g44730	Internal[c]	767	14	ND	ND	CC
21	**At5g10470 (KCA1)**	N-terminal[b]	1273	22	ND	ND	CC
22	**At5g65460 (KCA2)**	N-terminal[b]	1264	22	ND	ND	CC
23	At5g27950	N-terminal[b]	640	8	ND	ND	CC
24	At1g55550	N-terminal[b]	887	9	ND	ND	CC
25	**At5g65930 (AtKCBP)**	C-terminal	1259	20	Minus	PPB, spindle, spindle poles, Phragmo-plast	CC, MyTH4 Talin-like, CBD, PEST
26	At1g20060	N-terminal	923[d]	21	ND	ND	
27	At3g10180	N-terminal	459	11	ND	ND	CC
28	At1g59540	N-terminal	823	17	ND	ND	CC
29	At5g06670	N-terminal	997	22	ND	ND	CC

(*continued overleaf*)

Tab. 2 (*Continued*)

	Gene code (published name)	*Motor location*	*#aa*	*Number of introns*	*Motility*	*Cellular localization*	*Other Domains*
30	At3g12020	N-terminal	956	22	ND	ND	CC
31	**At1g21730 (MKRP1)**	N-terminal	909[d] (890)	22	ND	Mito-chondria	CC
32	**At4g39050 (MKRP2)**	N-terminal	1121 (1055)	22	ND	Mito-chondria	CC
33	At2g21380	N-terminal	857	18	ND	ND	CC
34	**At1g18370 (HINKEL/ AtNACK1)**	N-terminal	1003[d] (974)	11	ND	Phragmo-plast equator	CC
35	**At3g43210 (ATNACK2/ TES)**	N-terminal	932	13	ND	Interzonal Mts, Phragmo-plast	CC,
36	At4g38950	N-terminal	834	11	ND	ND	CC
37	At2g21300	N-terminal	581	9	ND	ND	CC
38	At3g51150	N-terminal	968	13	ND	ND	CC
39	At5g66310	N-terminal	1037	13	ND	ND	CC
40	At5g42490	N-terminal	1087	9	ND	ND	
41	At4g24170	N-terminal	1263	14	ND	ND	CC
42	At3g16630	Internal	799	12	ND	ND	CC
43	At3g16060	Internal	706	12	ND	ND	CC
44	At5g02370	N-terminal	664	10	ND	ND	
45	At1g18550	N-terminal	703[d]	3	ND	ND	CC
46	At3g49650	N-terminal	813	14	ND	ND	CC
47	At5g23910	N-terminal	665	17	ND	ND	HhH1
48	At3g50240	N-terminal	1075	21	ND	ND	CC
49	At5g47820	N-terminal	1032	22	ND	ND	CC
50	At5g60930	N-terminal	1335[d]	24	ND	ND	CC
51	At3g63480	N-terminal	439	11	ND	ND	CC
52	At1g01950	N-terminal	885[d]	18	ND	ND	CC,ARM,
53	At1g12430	N-terminal	895[d]	16	ND	ND	CC,ARM
54	At3g54870	N-terminal	1070	20	ND	ND	CC, ARM
55	At3g19050	N-terminal	2756	34	ND	ND	CC,
56	At3g17360	N-terminal	2158[d]	30	ND	ND	CC,
57	At3g44050	N-terminal	1229	20	ND	ND	CC
58	At3g20150	N-terminal	1103[d]	16	ND	ND	CC
59	**At3g23670 (ATPAKRP1L)**	N-terminal	1268[d] (1313)	14	ND	Spindle midzone, Phragmo-plast	CC,
60	**At4g14150 (AtPAKRP1)**	N-terminal	1292[e] (1294)	24	ND	Spindle midzone, Phragmo-plast	CC

Tab. 2 (*Continued*)

	Gene code (published name)	Motor location	#aa	Number of introns	Motility	Cellular localization	Other Domains
61	**At4g14330 (AtPAKRP2)**	N-terminal	959 (869)	9	ND	Punctate staining of Phragmoplast	CC

[a] Motility was determined using in vitro assays.
[b] Although the motor is N-terminal, it groups into C-terminal family.
[c] Although the motor is internal, it groups into C-terminal family.
[d] Slight discrepancy in #aa between AtDB and MIPS predictions;
[e] NCBI predicted protein has 1662 amino acids.
Notes: The number of amino acids in parenthesis are based on the deduced number from cDNA. CC: coiled coil; CH: calponin homology domain; MyTH4: domain present in the tail region of some myosins; Talin-like: talin-like domain found in some myosins and band 4.1 superfamily; CBD: calmodulin binding domain; PEST: motif rich in proline, glutamine, serine, and threonine residues; ARM: Armadillo/beta-catenin-like repeats; HhH1: helix-hairpin-helix.
(Modified from Reddy, A.S.N., Day, I.S. (2001a) Analysis of the myosins encoded in the recently completed *Arabidopsis thaliana* genome sequence, *Genome Biol.* **2**, 24.21–24.17).

of spectrin-like proteins. The CH domain is a protein module of approximately 110 residues found in cytoskeletal and signal transduction proteins, either as two domains in tandem or as a single copy. Proteins with a tandem pair of CH domains cross-link F-actin, bundle actin, or connect intermediate filaments to cytoskeleton. Proteins with a single copy are involved in signal transduction. Perhaps the kinesins containing a CH domain bind actin and are involved in signal transduction or linking of actin and MTs. Some domains that are involved in protein–protein or protein–DNA interactions are also present in some kinesins (Fig. 1 and Table 2). These include armadillo repeats, a helix-hairpin-helix DNA-binding domain. KCBP has MyTH4 and talin-like domains present in some myosins, suggesting that it has domains of both MT- and actin-based motors. Such motors may be involved in cross talk between MT and actin cytoskeleton. KCBP also has been shown to have a calmodulin-binding domain.

1.1.1 Phylogenetic Analysis

In nonplant systems, nine subfamilies of kinesins with some ungrouped kinesins have been identified by phylogenetic analysis using the conserved motor domain. Three of the subfamilies (KHC, KRP85/95, and Unc104/KIF1) are involved in transport. Members from the other subfamilies (C-terminal, Kip3, MKLP1, BimC, chromokinesin/KIF4, and MCAK/KIF4) function in various processes associated with cell division. A phylogenetic analysis of the 61 Arabidopsis kinesins motor domain sequences with 113 other motor domain sequences from nonplant systems has revealed that seven of the nine families are represented in Arabidopsis (Figs. 2 and 3). However, several Arabidopsis kinesins do not fall into any of the nine subfamilies and are likely to represent

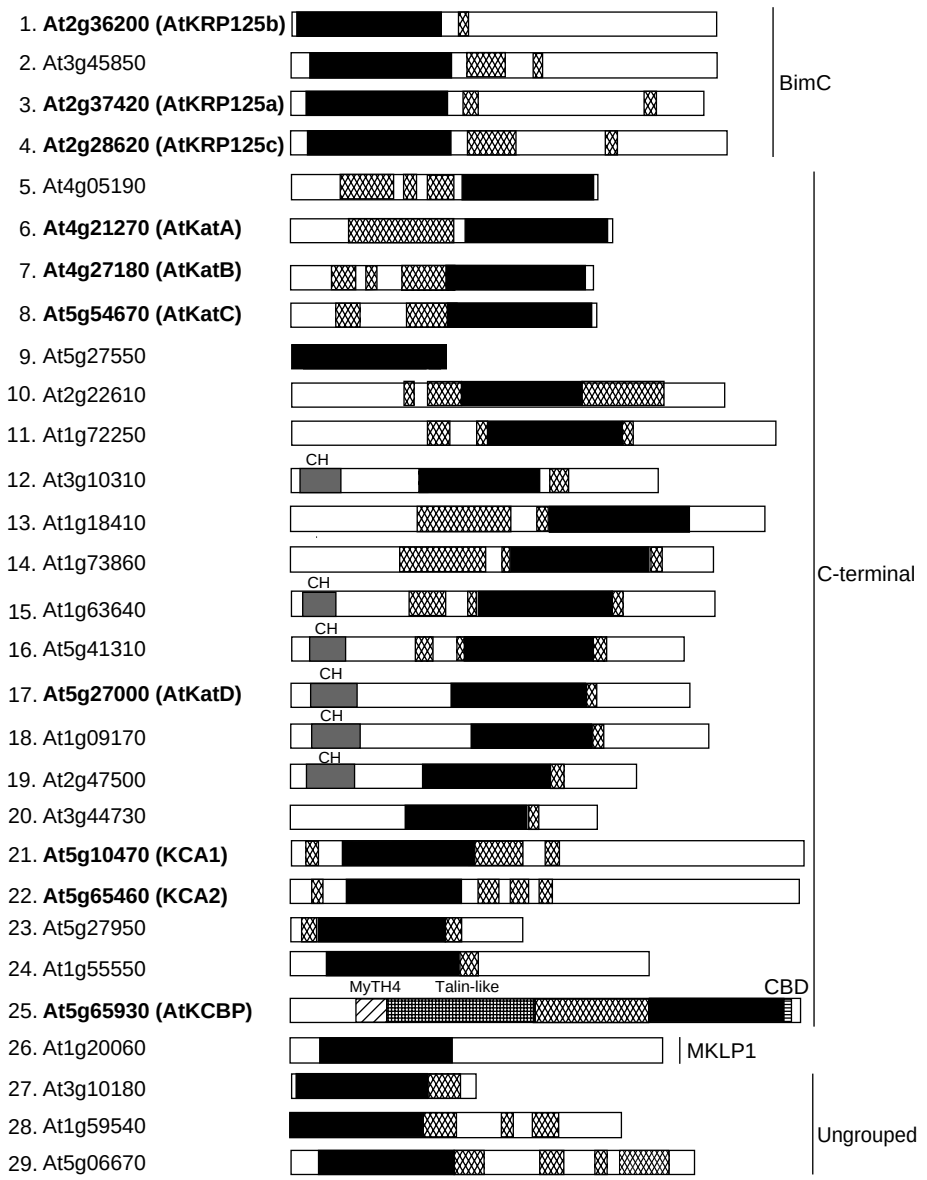

Fig. 1 Structural features of Arabidopsis kinesins. The deduced amino acid sequence of kinesins in the Arabidopsis database was analyzed for the presence of various domains. The kinesins are arranged according to their grouping in the phylogenetic tree in Fig. 3. Characterized kinesins are indicated in bold. ARM, Armadillo/beta-catenin-like repeat, CBD, calmodulin-binding domain; CC, coiled-coil region; CH, Calponin homology domain; MD, motor domain; MyTH4, myosin tail homology region; Talin-like, talin-like domain; HhH1, helix-hairpin-helix domain. Bar = 100 aa. (Modified from Reddy, A.S.N., Day, I.S. (2001b) Kinesins in the Arabidopsis genome: A comparative analysis among eukaryotes, *BMC Genomics* **2**, 2.1–2.13).

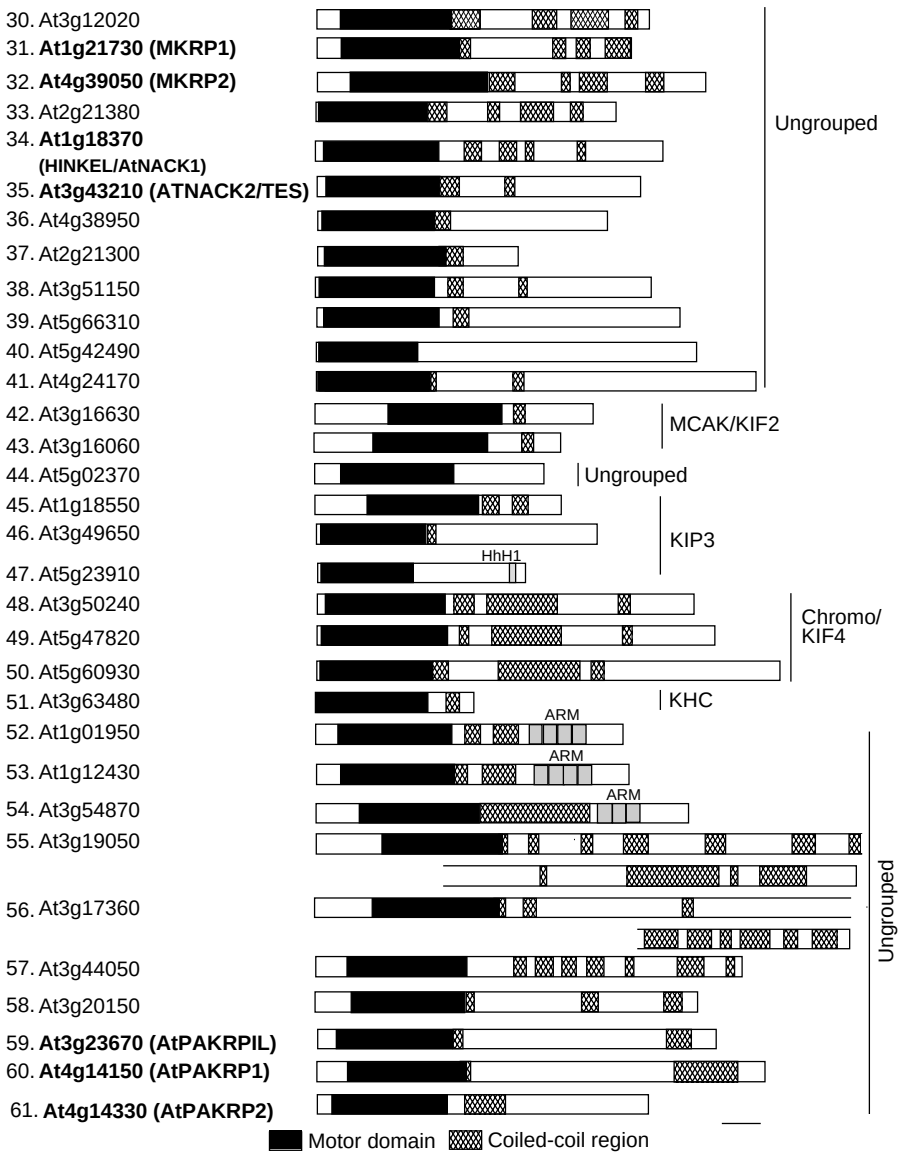

Fig. 1 (*Continued*)

additional subfamilies that are unique to plants (Figs. 2 and 3). Most of the Arabidopsis kinesins are more closely related to another Arabidopsis or another plant kinesin than to any other kinesin used in the comparison. The subfamilies that are involved in transport are underrepresented in Arabidopsis (Fig. 2). There are no members of the KRP85/95 or Unc104/KIF1 subfamilies in Arabidopsis (Figs. 2 and 3). No KHCs have been reported in plants. The phylogenetic tree indicates that there is possibly one KHC-type kinesin in Arabidopsis. At3g63480

Tab. 3 Myosins in arabidopsis.

Name	*# of aa*	*Gene code*	*EST/cDNA*	*Class*	*Domains*
1. **At ATM**	1166	AT3g19960 (ATM1)[a]	+	VIII	MD,CC,IQ
2. **At ATM2**	1111	AT5g54280	+	VIII	MD,CC,IQ
	1101[b]	ATM2/AtMYOS1[a]			
3. At VIIIA	1085	AT1g50360	+	VIII	MD,CC,IQ
4. At VIIIB	1126	AT4g27370		VIII	MD,CC,IQ
5. **At MYA1**	1520	AT1g17580	+	XI	MD,CC,IQ
	1599[c]	(AtMYA1)[a]			
6. **At MYA2**	1505[c]	AT5g43900	+	XI	MD,IQ
	1515	(AtMYA2)[a]			
7. At XIA	1730	AT1g04600		XI	MD,CC,IQ
8. At XIB	1519	AT1g04160	+	XI	MD,IQ
9. At XIC	1572	AT1g08730	+	XI	MD,CC,IQ
10. At XID	1611	AT2g33240	+	XI	MD,CC,IQ
11. At XIE	1529	AT1g54560		XI	MD,CC,IQ
12. At XIF	1556[d]	AT2g31900	+	XI	MD,IQ
13. At XIG	1502	AT2g20290		XI	MD,CC,IQ
14. At XIH	1452[d]	AT4g28710	+	XI	MD,CC,IQ
15. At XI-I	1374	AT4g33200	+	XI	MD,CC,IQ
16. **At XIJ**	1242	AT3g58160		XI	MD,CC,IQ
	963[b]	(AtMYOS3)[a]			
	998[b]	(AtMYA3)[a]			
17. At XIK	1544	AT5g20490		XI	MD,CC,IQ

[a] Name as reported in the literature.
[b] Number of amino acids previously reported for partial sequence.
[c] Number of amino acids predicted by NCBI.
[d] Edited for full-length sequence.
Notes: MD: motor domain; CC: Coiled-coil region; IQ: putative calmodulin-binding motif
(Modified from Reddy, A.S.N., Day, I.S. (2001a) Analysis of the myosins encoded in the recently completed *Arabidopsis thaliana* genome sequence, *Genome Biol.* **2**, 24.21–24.17).

falls into the KHC group with a closer relationship to KHCs found in fungi and, like fungal KHC, lacks binding site for kinesin light-chain proteins. However, biochemical and functional studies are needed to verify that At3g63480 codes for a KHC. One Arabidopsis kinesin groups with the MKLP1 subfamily (Fig. 3). Two internal motor Arabidopsis kinesins grouped with the MCAK/KIF2 subfamily (also called *the internal family*), whose members are involved in vesicle transport, chromosome movement, and MT catastrophe. Three Arabidopsis kinesins fall into a group with Kip3 subfamily members in which ScKip3 is involved in nuclear movement. Three Arabidopsis kinesins form a branch off of the chromokinesin/KIF4 subfamily members, some of which are involved in vesicle transport (HsKIF) and spindle organization and chromosome positioning (Xlklp1). Arabidopsis has many kinesins (24 out of 61) that did not fall into any of the known groups (Figs. 2 and 3). Members of BimC subfamily, which are N-terminal plus-end motors, are present in all five eukaryotic organisms that have been sequenced. The three NtKRP125-like

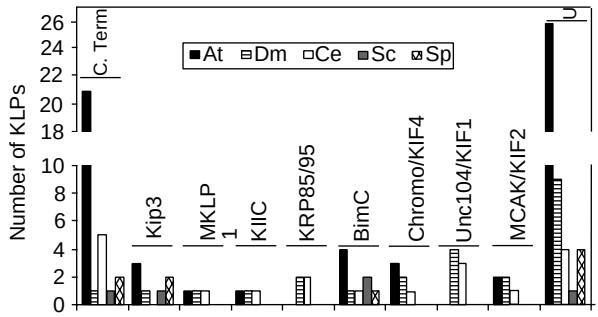

Fig. 2 The number of kinesins in each family in different organisms. C.Term, C-Terminal motor; Chromo/KIF4, chromokinesin/KIF4; U, ungrouped. At, *Arabidopsis thaliana*; Dm, *Drosophila melanogaster*; Ce, *Caenorhabditis elegans*; Sc, *Saccharomyces cerevisiae*; Sp, *Schizosaccharomyces pombe* (Adapted from Reddy, A.S.N., Day, I.S. (2001b) Kinesins in the Arabidopsis genome: A comparative analysis among eukaryotes, *BMC Genomics* **2**, 2.1–2.13).

Arabidopsis kinesins (AtKRP125a, b, and c) are grouped with the tobacco homolog in the BimC subfamily, which are involved in cross-linking and antiparallel sliding of MTs.

Twenty-one Arabidopsis kinesins were grouped into the C-terminal subfamily. This is an unusually large number compared to the other organisms. It is also unusual in that 11 kinesins in this group have internal motors and five have N-terminal motors. The internal kinesins have a motor domain that is closer to the C-terminus than the N-terminus, but each has some sequence C-terminal to the motor domain and could be called internal, depending on the parameters used to define an internal motor. It will be interesting to find out the direction of movement of these kinesins whose motor domains are either N-terminal or internal, but are most closely related to the C-terminal subfamily. Kinesins in the C-terminal subfamily translocate toward the minus end of MTs and have a conserved sequence at the neck/motor core junction. It was reported that the residues G and N residues at the neck/motor core junction are necessary for minus-end directed movement. Analysis of the neck/motor core junction of the 21 C-terminal class Arabidopsis kinesins showed conservation of these residues in most of the C-terminal Arabidopsis kinesins.

Several Arabidopsis kinesins show some similarity to other ungrouped kinesins (Figs. 2 and 3). The ungrouped centromeric proteins (HsCENPE and UmKin1) cluster with seven Arabidopsis kinesins. CENP-E binds to the kinetochore throughout mitosis and to MTs of the spindle midzone during late stages of mitosis. One Arabidopsis kinesin is paired with HsKid, an ungrouped kinesin. HsKid is a kinesin-like DNA-binding protein that is involved in spindle formation and the movements of chromosomes during mitosis and meiosis. Arabidopsis PAKRP1 along with five other Arabidopsis kinesins are grouped with XlKlp2. XlKlp2 is required for centrosome separation and maintenance of spindle bipolarity, whereas PAKRPs associates with the phragmoplast and is

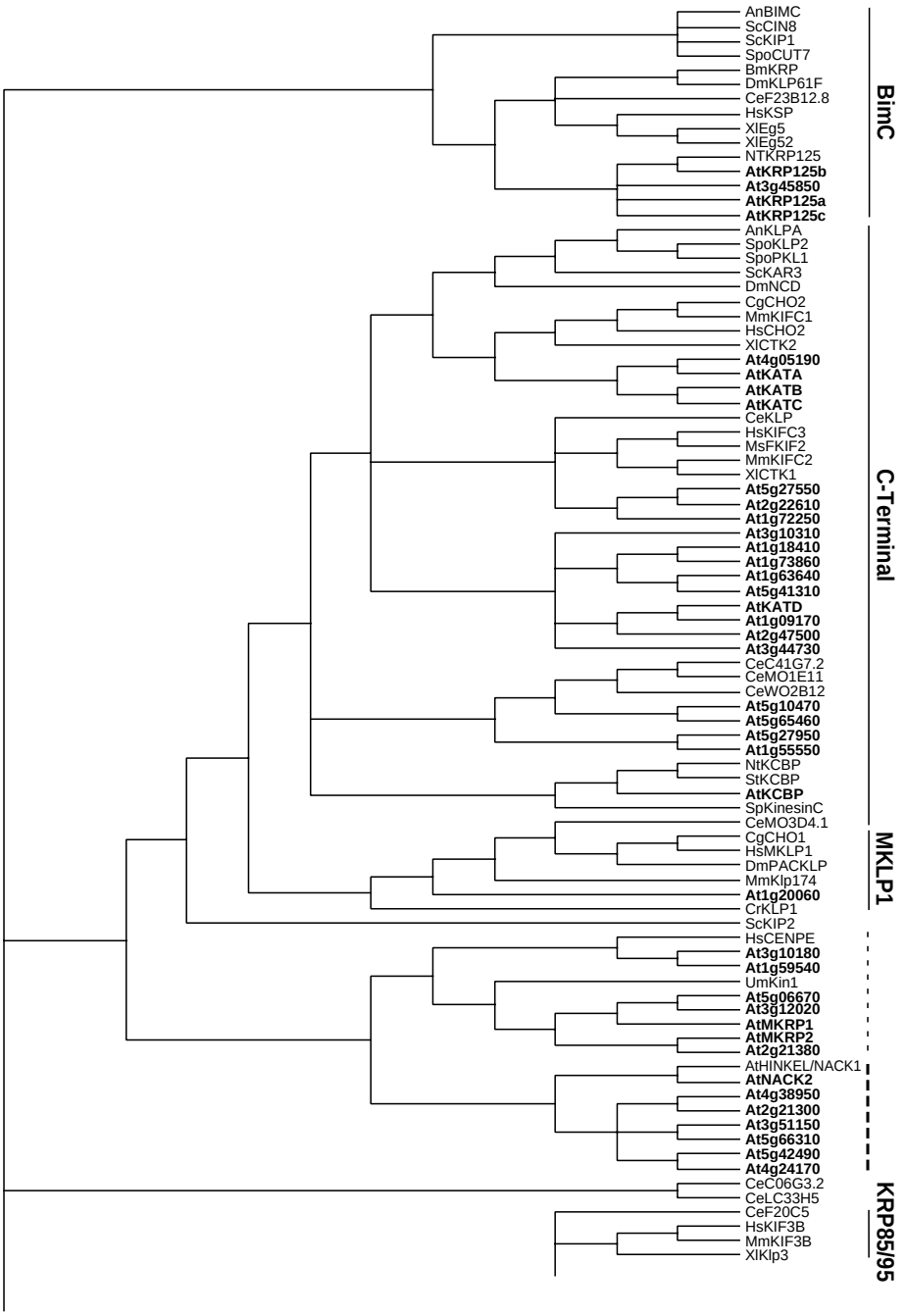

Fig. 3 (*Continued on page 476*)

expected to function in cytokinesis. Three Arabidopsis kinesins form a subgroup separate from any other kinesin, but share a branch with CeLF22F4 and DmNOD (Fig. 3). Eight other Arabidopsis kinesins form a subgroup separate from kinesins of any other organism. In many cases, a group of Arabidopsis kinesins forms a separate branch within the major subgroup in which they fall.

The large number of kinesins in plants is expected to control many diverse cellular processes including some plant-specific functions. There are many MT-associated processes that are unique to plants. For example, during cell division in plants, several plant-specific MT arrays such as the preprophase band and the phragmoplast are formed that are important in determining the future location of the cell wall and cell wall formation respectively. These unique processes are likely to require additional plant-specific motors. In addition, centrosomes play an important role in MT organization in animals, whereas plants have no well-defined centrosomes. Hence, MT organization and dynamics in plants may also require additional MT motors. The presence of unique organelles such as chloroplasts and MT-dependent processes such as cell wall synthesis may also warrant for additional motors. Also, cell-to-cell transport of macromolecules such as RNA and viruses through plasmodesmata may require MT-based motors. It is very likely that plant kinesins may participate in functions other than transport and MT dynamics and organization. Recently, it has been shown that two plant kinesins function as activators of a protein kinase. Since Arabidopsis does not have dyneins, some kinesins may be performing functions that are carried out by dyneins in other systems.

1.2 Dyneins

Dyneins move along MTs toward the minus ends. The motor domain in dyneins (~1000 amino acids) is much longer than kinesins (~340 amino acids) and does not share sequence similarity with the kinesin motor domain. Cytoplasmic dynein is a large multisubunit complex consisting of two heavy chains (~530 kDa each), three intermediate chains (74 kDa), and four light intermediate chains (~55 kDa). The central and C-terminal regions of dynein heavy chain, which are predicted to form globular structure, interact with MTs and contain motor activity, whereas the N-terminal region is thought to bind

Fig. 3 Phylogenetic tree of all Arabidopsis kinesins along with 113 nonplant kinesins. The tree was generated using the motor domain of kinesins, as described. Vertical dashes indicate ungrouped kinesins (light dashes – those grouped with other kinesins, bold dashes – exclusively Arabidopsis kinesins). The Arabidopsis kinesins are in bold. Kinesins from the following organisms were used: An, *Aspergillus nidulans*; Bm, *Bombyx mori*; Ce, *Caenorhabditis elegans*; Cf, *Cylindrotheca fusiformis*; Cg, *Cricetulus griseus*; Cr, *Chlamydomonas rheinhardtii;* Dd, *Dictyostelium discoideum*; Dm, *Drosophila melanogaster*; Gg, *Gallus gallus*; Hs, *Homo sapiens*; Lc, *Leishmania chagasi*; Lm, *Leishmania major;* Lp, *Loligo pealii;* Mm, *Mus musculus;* Ms, *Morone saxatilis*; Nc, *Neurospora crassa;* Nh, *Nectria haematococca*; N, *Nicotiana tabacum;* St, *Solanum tuberosum;* Rn, *Rattus norvegicus;* Sc, *Saccharomyces cerevisiae;* Sp, *Strongylocentrotus purpuratus;* Spo, *Schizosaccharomyces pombe*; Sr, *Syncephalastrum racemosum*; Um, *Ustilago maydis*; Xl, *Xenopus laevis.* (Modified from Reddy, A.S.N., Day, I.S. (2001b) Kinesins in the Arabidopsis genome: A comparative analysis among eukaryotes, *BMC Genomics* **2**, 2.1–2.13).

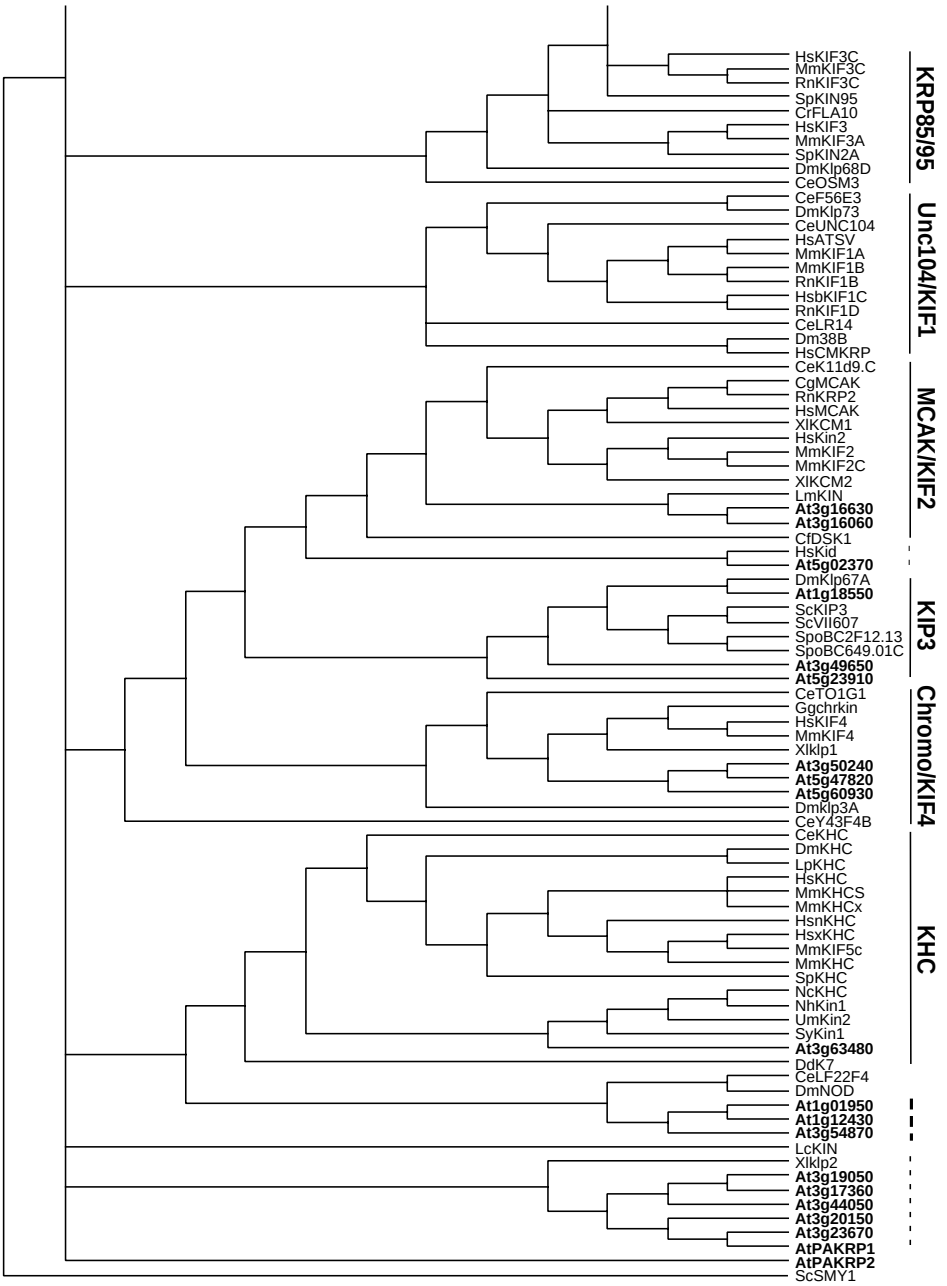

Fig. 3 (*Continued from page 474*)

cargo. The intermediate and light intermediate chains of cytoplasmic dynein associate with a protein complex called *dynactin*, which is made of 10 subunits. Although there is some immunological and biochemical data indicating the presence of dynein heavy chains in plants, sequences that are similar to dynein and dynactin subunits have not been found in the Arabidopsis genome, suggesting that they may be absent in flowering plants. The expanded family members of the kinesin superfamily may carry out the functions performed by dyneins in animals. Interestingly, dynein genes appear to be present in rice and tobacco.

2 Actin-based Motors

2.1 Myosins

Myosins, a diverse group of actin-based molecular motors, perform a broad array of cellular functions. A large number of myosins have been described in many eukaryotes. Most of these are identified on the basis of sequence similarity with the myosin motor domain. All myosin heavy chains have three common domains: a highly conserved motor domain (~850 amino acids) located at the amino-terminus in almost all myosins, which is followed by a neck region that binds to calmodulin or calmodulin-related proteins, and a C-terminal tail domain. The calmodulin-binding domain "IQ motif" in these myosins has a consensus sequence "IQXXXRGXXXR" (I, isoleucine; Q, glutamine; R, arginine; G, glycine). This motif is present in one to seven repeats, depending on the type of myosin. The only myosin that does not contain an "IQ motif" is myosin XIV. The tail region of different myosins varies considerably in length and structure. The tail is characteristic to each myosin with interesting domains that are also found in nonmotor proteins. The number of myosins in an organism varies among organisms. However, not all types of myosins have been identified in a single organism.

Although there is compelling indirect evidence for the role of actin and actin-based motors in various transport processes in plants, little is known about specific plant myosins that perform these functions. Myosins have been identified in plants both biochemically and at the molecular level. Immunological detection of myosins using animal myosin antibodies identified proteins of various sizes from different plants. Immunofluorescence studies localized myosin to the surface of organelles, and the vegetative nuclei and generative cells in pollen grains and tubes, to the active streaming lanes and cortical surface in pollen tubes and, more recently, to plasmodesmata in root tissues. Motility assays and ATPase assays using myosin-like proteins isolated from plants have also demonstrated the presence of myosins in plants. Using PCR-based approaches, a few myosins have been cloned from Arabidopsis and other plants.

A 170-kDa myosin has been purified from lily pollen and the antibodies against this myosin stained particles of various sizes and the apical cytoplasm of lily and tobacco pollen tubes, suggesting that several myosins are likely to be present in pollen tubes. The 170-kDa and 175-kDa myosins that translocate F-actin at a velocity of about 9 μm/s and 3 to 4μm-s respectively were found to be associated with calmodulin. In *in vitro* motility assays, the characean myosin translocated actin at 50 μm-s. In Arabidopsis, the expression of

several myosins in a given cell type implies that actin-based motors are involved in a wide range of cellular functions.

By analyzing the Arabidopsis genome sequence with the conserved motor domain of myosins, a total of 17 myosins were identified in Arabidopsis. Similar analysis in rice indicates that there are at least 14 myosins in this organism. Table 3 lists the Arabidopsis myosins by names as given in the phylogenetic tree. In comparison, *S. cerevisiae, S. pombe, C. elegans, Drosophila melanogaster* and *H. sapiens* have 5, 5, 17, 13, and 40 myosins respectively. Five of the 17 Arabidopsis myosins have been reported in the literature. All Arabidopsis myosins have three to six putative calmodulin-binding "IQ" motifs (Fig. 4). The IQ domains usually follow right after the motor domain, but are separated slightly from the motor domain in At XID, At XI-I, and At XIK. There are three or four IQ domains in Class VIII myosins and five or six in Class XI, except for At XIK, which has only four (Fig. 4). There are coiled-coil domains in all the myosins that differ in length and number. They often follow directly after the IQ domains, but, in some cases, there is intervening sequence. On the basis of the presence of the coiled-coil domains, it is likely that the Arabidopsis myosins either form dimers or interact with other proteins. The Class XI myosins are much longer than the Class VIII myosins with the difference being in the length of the C-terminal region following the conserved domains found in myosins. Myosins containing IQ domains are typically calmodulin-sensitive. However, the interaction with, or regulation by, calmodulin has not yet been demonstrated for Arabidopsis myosins. The information on myosin light chains is limited in plants. Calmodulin associates with two myosins from lily pollen and regulates their motor activity. Besides the motor, IQ, and coiled-coil domains, several other domains have been identified in myosins from classes other than the plant classes VIII and XI. These include SH3 domains (Src homology 3 domains), ankyrin repeats, MYTH4 (domain of unknown function found in a few classes of myosins), zinc-binding domain, plecstrin homology, FERM/talin (band 4.1/ezrin/radixin/moesin), GPA-rich domains, and a protein kinase domain. Plant myosins do not have any of these other domains.

2.1.2 Phylogenetic Analysis

Phylogenetic analysis using the conserved motor domain grouped all known myosins into 18 distinct classes. Phylogenetic analysis of the Arabidopsis myosins motor domain with nonplant and plant myosins revealed that all the Arabidopsis myosins and other plant myosins fall into only two groups, Class VIII and Class XI (Fig. 5). Four Arabidopsis myosins are in Class VIII and 13 in Class XI. These groups contain exclusively plant or algal myosins with no animal or fungal myosins, suggesting that plants have a unique set of myosins (Fig. 5). The motor domain in all cases is in the N-terminal region (Fig. 4). The motor domain starts at about 50 to 55 residues for the Class XI myosins, while the Class VIII myosins have a longer N-terminal region prior to the motor domain (99–159 residues). An algal (*Chara corallina*) myosin, Cc ccm, does group with the plant Class XI myosins, but is on a separate branch from any other Class XI myosin (Fig. 5). A phylogenetic tree that was constructed using the full-length sequences of Arabidopsis and other plant myosins showed similar grouping of plant myosins. Among the Class XI myosins, the similarity ranges from 40 to 85% (full length) and 61 to

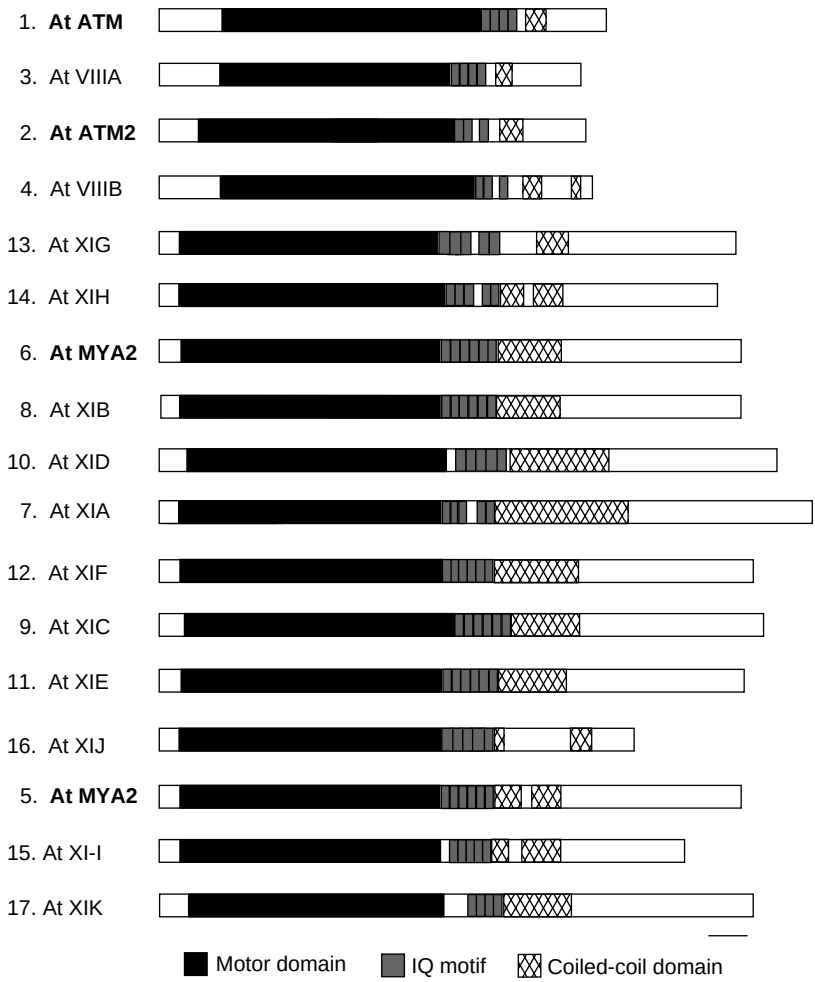

Fig. 4 Schematic diagram showing various predicted domains in Arabidopsis myosins. The Arabidopsis genome database was searched with the conserved motor domain of a myosin to identify Arabidopsis myosins. The deduced amino acid sequence of Arabidopsis myosins was analyzed for the presence of various domains. The numbers refer to the number in Table 1. The first four myosins are in Class VIII and the following 13 are in Class XI. The bar represents 100 amino acids. (Adopted from Reddy, A.S.N., Day, I.S. (2001a) Analysis of the myosins encoded in the recently completed *Arabidopsis thaliana* genome sequence, *Genome Biol.* **2**, 24.21–24.17).

91%(motor domain). The similarity between the Class VIII myosins ranges from 50 to 83% (full length) and 64 to 92% (motor domain). When Class VIII myosins are compared to Class XI myosins, the similarity only ranges from 22 to 29% (full length) and 35 to 42% (motor domain). Myosins have 131 highly conserved residues spread throughout the motor domain that define a core consensus sequence. Comparison of an alignment of Arabidopsis myosin motor domains to these conserved sequences

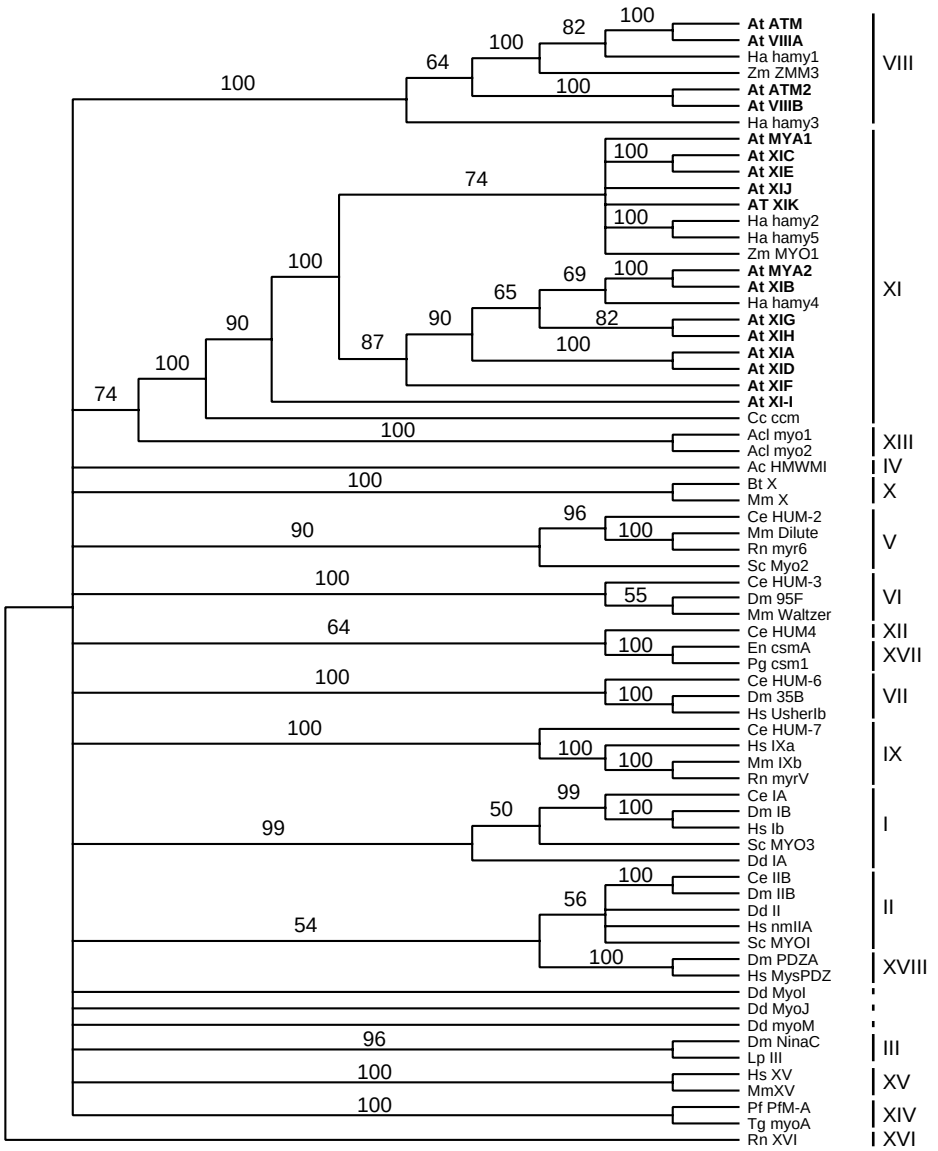

Fig. 5 Phylogenetic analysis of Arabidopsis myosins. A phylogenetic tree was generated as described earlier. The myosin groups are identified on the left in roman numerals. The Arabidopsis myosins are in bold. Myosins from the following organisms were used: Ac, *Acanthamoeba castellani;* Acl, *Acetabularia cliftoni*; Cc, *Chara corallina,* Ha, *Helianthus annuus;* Zm, *Zea mays;* Bt, *Bos taurus*; Mm, *Mus musculus;* Ce, *Caenorhabditis elegans;* Dm, *Drosophila melanogaster;* Rn, *Rattus norvegicus*; Sc, *Saccharomyces cerevisiae;* Hs, *Homo sapiens*; Dd, *Dictyostelium discoideum*; Lp, *Limulus polyphemus*; En, *Emericella nidulans*; Pg, *Pyricularia grisea*; Pf, *Plasmodium falciparum*; and Tg, *Toxoplasma gondii*. The number at the branches indicates the number of times the dichotomy was supported out of 100 bootstrap tries.

shows a great deal of conservation among them, suggesting that they are capable of motor function.

3 Cellular Roles of Motors

3.1 Cell Division

Although many of the dynamic processes of cell division such as formation of the spindle, congression of chromosomes at the equatorial region, and migration of chromatids during cell division are common to both animals and plants, there are significant differences between plant and animal cell division. For example, the dividing plant cells organize their cytoskeletal elements into unique structures that are not found in nonplant systems. Just prior to prophase, plant cortical MTs rearrange to form a band of MTs called the *preprophase band* (PPB), which accurately predicts the future location of the cell plate. Another distinctive feature of plant mitosis is the formation of a bipolar spindle in the absence of centrosomes. The structures involved in the organization of MTs are the centrosomes in animal cells and the spindle pole body (SPB) in fungal cells, which are lacking in plant cells. Finally, cytokinesis, the process that produces two daughter cells following the completion of nuclear division, in plants is accomplished quite differently from plants. Cytokinesis in plant cells occurs via the formation of a polysaccharide cell plate, which expands from the center to the periphery. Phragmoplast, a structure composed of two disks of parallel MTs, with their plus ends toward the equatorial zone, and actin filaments is involved in cell-plate formation. Vesicles carrying the cell plate materials are transported to the equatorial zone of the phragmoplast. The association of vesicles with phragmoplast MTs indicates that MT-based motors are likely to be involved in transporting the vesicles to the cell plate. The fusion of vesicles in the cell plate is mediated by phragmoplastin, a dynamin homolog, in plants. The plant spindle, which is formed in the absence of a well-defined centrosome, and several unique plant-specific MT arrays such as the preprophase band and the phragmoplast suggest that plants are likely to contain novel kinesins that are not present in animal cells.

In animals, six subfamilies of kinesins have been shown to be involved in some aspect of cell division. Several recent reports on kinesins expression and immuunolocalization indicate that at least nine plant kinesins (Kat A, KatB/C, TKRP125, AtKCBP, AtPAKRP1, AtPAKRP1L, AtPAKRP2, DcKRP120-2, HINKEL/AtNACK1, AtNACK2/TES) that belong to different families have a mitotic function. Some kinesins are expressed in a cell cycle–dependent manner. TKRP125, a member of BimC subfamily and NCAK from tobacco and two kinesins from Arabidopsis (KCBP, KatB/C) show a high level of expression in M-phase of the cell cycle. DcKRP120-1 and DcKRP120-2 are not detectable in nondividing cells. TKRP125 localizes to the cortical MTs in the S-phase, along MTs of PPB and perinuclear MTs in prophase, the equatorial plane of spindle MTs in metaphase and anaphase, and the phragmoplast MTs in telophase and cytokinesis. A carrot homolog of TKRP125 (DcKRP120-1) also localizes to the cortical MTs, the preprophase band, spindle, and the phragmoplast. Another member of BimC subfamily (DcKRP120-2) stained the spindle and phragmoplast, predominantly in the midline of the phragmoplast where plus ends of the MTs overlap. AtPAKRP1, a

nonmember of BimC, which resembles an ungrouped Xklp2 from Xenopus, also localizes to the midline of the phragmoplast. KatA, a C-terminal motor, localizes near the midzone at metaphase and anaphase and the phragmoplast in Arabidopsis and tobacco BY-2 suspension cells. Anti-KCBP antibodies stain the preprophase band, the spindle, and the phragmoplast. The KCBP antibody did not stain the cell plate region of the phragmoplast, but MT bundles on either side of the cell plate are strongly labeled (Fig. 6). Localization of KCBP in *Haemanthus* also showed localization to mitotic MTs arrays. In *Haemanthus,* anti-KCBP staining is seen almost exclusively at the spindle poles in late anaphase. The assembly of the acentriolar spindle in plants may involve convergence of MT minus ends leading to the formation of spindle poles. The fact that KCBP is a minus-end-directed motor and localizes to the spindle poles suggests that it may be involved in acentriolar spindle formation in plants. Between early and late anaphase, the localization of KCBP shifts toward the spindle pole, supporting a role for KCBP in the formation of a converging bipolar spindle. Localization of minus-end (KatA and KCBP) and plus-end (TKRP125) motors in spindle suggests their involvement in counterbalancing forces generated by plus- and minus-end motors to stabilize the spindle. Constitutive activation of KCBP during late prophase caused premature breakdown of nuclear envelope and arrest of cells at prometaphase, whereas activation of KCBP at late metaphase or anaphase did not affect the progression of anaphase, but caused aberrant phragmoplasts formation and delayed cytokinesis. This study suggests that KCBP is differentially active during the various phases of cell division. Its activity is downregulated

Fig. 6 Triple localization of nucleus, MTs, and KCBP in a cytokinetic cell of tobacco. Nuclei stained with DAPI (DAPI), MTs (MT) are localized with a monoclonal antibody to β-tubulin and KCBP (KCBP) was localized with affinity-purified anti-KCBP antibody. Secondary antibodies conjugated to FITC and Cy3 were used to detect MTs and KCBP respectively. (From Bowser, J., Reddy, A.S. (1997) Localization of a kinesin-like calmodulin-binding protein in dividing cells of Arabidopsis and tobacco, *Plant J.* **12**, 1429–1437).

in metaphase and telophase and upregulated in anaphase most likely by changes in cytosolic calcium level. Colocalization of some of the kinesins with mitotic MT arrays suggests that they may have a role in forming these arrays by bundling MTs. The motor domain of KCBP has been shown to bundle MTs.

Numerous forces are at play in the development, maintenance, and function of the phragmoplast. Golgi-derived vesicles are transported to the forming cell plate where the plus ends of phragmoplast MTs interdigitate. It is likely that MT motors are involved in forming the phragmoplast and powering the movement of the vesicles along the MTs. Seven plant kinesins (Kat A, TKRP125, AtKCBP, AtPAKRP1, AtPAKRP2, DcKRP120-2, HINKEL/AtNACK1) including plus- and minus-end motors and some plant-specific kinesins are associated with the phragmoplast and so are expected to function in cytokinesis. The localization patterns with various kinesins differed considerably. Immunolocalization studies as well as phragmoplast MT gliding assay in the presence of TKRP antibodies indicates its involvement in the organization of MT arrays especially in phragmoplast function. The polarity of MTs and microfilaments in the phragmoplast indicates that plus-end motors are likely to be involved in the transport of vesicles to the cell plate. Of the eight kinesins that localize to the phragmoplast, six are N-terminal motors (TKRP125, PAKRP1, PAKRP1L, DcKRP120-2, PAKRP2, HIK/NCAK) and two are C-terminal (KatA and KCBP). TKRP125 is a plus-end motor, and the other four N-terminal kinesins are likely plus-end motors. One or more of these are likely to be involved in vesicle transport to the cell plate. The punctate-staining pattern of PAKRP2 and its association with vesicles suggest that it is involved in transporting vesicles to the cell plate. It is unlikely that KCBP and KatA play a role in vesicle transport to the cell plate as these are known to be minus-end motors and the MTs are oriented with their plus ends facing the cell plate. The possible functions of minus-end motors are organization/stabilization of phragmoplast and/or recycling of Golgi vesicle membranes from the expanding cell plate. Functional analysis of TKRP125 indicates its role in cross-linking and sliding of antiparallel MTs in the phragmoplast. Immunofluorescence with AtPAKRP1 and AtPAKRP1L antibodies showed staining of interzonal MTs at late anaphase and later on its localization is restricted to the plus ends of interdigitating MTs in the phragmoplast. Functional studies with this kinesin indicate its involvement in maintaining the integrity of phragmoplast MTs. In order for the growth of the phragmoplast to be achieved, a rapid turnover of phragmoplast MTs must occur. The expansion of the phragmoplast from the center to the periphery involves disassembly of the MTs inside and assembly of MTs at the outside. Blocking of disassembly in the center by MT-stabilizing drugs causes cessation of phragmoplast expansion and formation of incomplete cell walls. Phragmoplast-associated kinesins may play a role in the MT assembly and disassembly associated with its growth. The loss of MTs from the center of the phragmoplast as it grows does not occur in *hik* mutants, suggesting that it controls directly or indirectly the dynamics of MTs in the phragmoplast. HIK does not appear to be involved in either vesicle transport or organization of MTs in the phragmoplast. Studies with tobacco NCAK1, a homolog of HINKEL, showed its localization to the equatorial zone of the phragmoplast, but not in the inner regions

of the phragmoplast where the cell plate matures. It appears that NCAK through a protein kinase, NPK1, regulates depolymerization of MTs in the center of the phragmoplast as it expands.

There is some evidence of the presence of a myosin in the phragmoplast, suggesting that actin-based motors may also be involved in the transport of vesicles on the phragmoplast. During cell division, the anti-Class VIII myosin staining remains confined to the transverse cell walls and is strongest in the newly formed cell wall. Immunogold electron microscopy showed labeling of Class VIII myosin associated with the plasma membrane and plasmodesmata. These studies suggest that Class VIII myosins may be involved in new cell wall formation and transport in the plasmodesmata.

3.2 Cell Polarity and Morphogenesis

Both actin and MT cytoskeleton have been implicated in cell polarity and morphogenesis. The first indication that a kinesin might play a role in morphogenesis came from studies with the *zwichel* mutant. In wild-type Arabidopsis, trichomes are unicellular with a stalk and three branches. In *zwichel* (*zwi*), mutants' trichomes are abnormal with a short stalk and one or two branches, depending on the severity of the allele (Fig. 7). Cloning of *ZWICHEL* has indicated that it is a calmodulin-binding kinesin, suggesting the requirement of this motor for expansion of the stalk and branching. Using paclitaxel, a known MT stabilizer, *zwichel* mutants were induced to form similar growth points indicative of branch formation in normal trichomes. Transient stabilization of MTs could compensate for KCBP/ZWI activity, suggesting that it may be involved in stabilization of MTs. The two MT-binding sites of KCBP could be involved in stabilization or possibly a KCBP/ZWI-interacting protein(s) could be responsible for the stabilization. Several alleles of *zwi* have been characterized. All *zwi* mutants grew normally with no apparent defects in cell division, except that they contain abnormal trichomes. The lack of a phenotype in other tissues suggests that either KCBP is nonessential or another motor

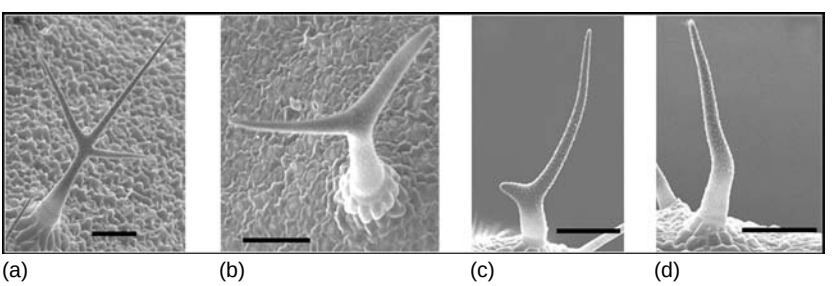

Fig. 7 Scanning electron micrographs of Arabidopsis trichomes of wild-type and *zwichel/kcbp* mutants. (a) Wild type; (b) *zwi*; (c) *zwi w2*; (d) *zwi9311 -11*. In *zwi w2,* most trichomes have two branches but the length of the second branch is varied. In the *zwi9311-11* mutant, about 40% of trichomes are unbranched and the rest have two branches. Seeds of *zwichel* mutants were kindly provided by Dr. David Oppenheimer and Dr. Martin Hülskamp. Scale bar = 100 μm (From Reddy, A.S.N., Day, I.S. (2000) The role of the cytoskeleton and a molecular motor in trichome morphogenesis, *Trends Plant Sci.* **5**, 503–505).

with overlapping functions may substitute for ZWI function in other tissues. In Arabidopsis, there are 61 kinesins including several C-terminal motors. Three extragenic suppressors (suppressor of *zwichel*-3; *suz1, suz2,* and *suz3*) that rescued trichome branch number defect in a *zwichel* mutant have been isolated. All three suppressors were found to be allele-specific, indicating direct interactions between these proteins. The *suz1 zwi-3* double mutants are male sterile due to a defect in pollen germination and pollen tube growth, suggesting a role for these genes in pollen germination and tube growth. At least four proteins (two calcium-binding proteins, a plant-specific protein kinase, and ANGUSTIFOLIA, a protein required for trichome branching) have been shown to interact with KCBP/ZWI. Inactivation of one of the myosins (*MYA2*) of myosin XI family in Arabidopsis also results in abnormal (unbranched) trichomes.

3.3 Cytoplasmic Streaming

Cytoplasmic streaming in plant cells is important in intracellular transport. The force that drives the cytoplasmic streaming appears to be dependent on actin and actin-based motors. Insight into the function of plant myosins in cytoplasmic streaming has been gained primarily by studies done with algae and pollen tubes. Characean cells exhibit a very rapid cytoplasmic streaming ($\sim$70 $\mu m\ s^{-1}$). Also, in pollen tubes, vesicles flow rapidly from the shank to the tip of the pollen tube where they fuse to support extensive tip growth of these cells. In addition to the bidirectional movement of vesicles, the generative cell and the nucleus move unidirectionally toward the tip. In *Chara,* an increase in Ca^{2+} concentration causes cytoplasmic streaming to stop. A myosin isolated from the alga *Chara corallina* was shown to be responsible for cytoplasmic streaming. The myosin was cloned and found to be a Class XI myosin related to the Arabidopsis MYA myosins. Using immunofluorescence, myosin was localized to vesicles, organelles, and generative cells and vegetative nuclei in grass pollen tubes. A myosin isolated from lily pollen has been shown to be responsible for cytoplasmic streaming in pollen tubes, and two myosins were identified in tobacco cell cultures are also thought to participate in cytoplasmic streaming. Myosin has been localized to vesicles, chloroplasts, and other organelles, indicating their involvement in transport of these organelles. The stop-and-go movements of Golgi stacks in plants also appear to be dependent on myosin. However, specific myosin(s) responsible for the movement of these organelles have not been identified. Biochemical properties and the localization data of two lily pollen myosins (175 kDa and 170 kDa) suggest that these myosins are at least partially responsible for cytoplasmic streaming, organelle transport, and calcium sensitivity of cytoplasmic streaming. It has been shown that the directional movement of mitochondria is dependent on F-Actin and myosin system. Unlike in animals, the movement of peroxisomes in plants is actin-based and actin polymerization appears to drive the movement of these organelles.

3.4 Microtubule Dynamics and Organization

Microtubules are highly dynamic structures with a half-life of an individual MT of about 10 min. Drugs that disrupt the dynamics of MTs have been shown to

severely hamper the MT-dependent process. Treatment of dividing cells with colchicine leads to disappearance of the mitotic spindle, suggesting that MTs in the spindle undergo continuous polymerization and depolymerization. Taxol, in contrast to colchicine, bind to MTs and stabilizes them. Addition of taxol to dividing cells leads to arrest of dividing cells in mitosis. The dynamics of the MT cytoskeleton during cell division in plant cells have been studied extensively. The rapid turnover of plant MTs suggests that the transition between MT arrays and change in orientations may involve depolymerization/repolymerization of MTs and/or movement of MTs. The dynamic instability of MTs is important for their normal function. For example, stabilization of MTs in elongating root hairs causes loss of growth directionality and promotes branching. The MTs of the phragmoplast have a t1/2 of 60 s. However, the mechanisms that regulate the formation and dynamic instability of MT arrays in plants are not understood. Several aspects of MT dynamics such as cross-linking, zippering, bundling, sliding, and stability of MTs is regulated by motor proteins. KCBP like Ncd in *Drosophila* can cross-link and zipper MTs to focus MTs to form spindle poles. Two distinct MT-binding sites and minus-motor activity enable these proteins to perform such a function. In addition, C-terminal motors have been shown to bundle MTs. Some motors are implicated in polymerization and depolymerization of cytoskeletal elements. There is at least one example where a kinesin has no motor activity, but controls MT dynamics. Plant kinesins such as NCAK1 appear to be involved in MT dynamics. Hence, the paradigm that motors bind cargo and move along cytoskeletal tracks may not explain the functions of all the motors.

3.5 Intercellular Transport

Plant cells are connected through plasmodesmata to form a continuous symplastic network. Recent studies show that macromolecules such as RNA and protein move from cell to cell through plasmodesmata. Cytoskeleton and molecular motors have been implicated in this regulated transport of macromolecules through plasmodesmata. The movement protein of viruses, which enables the movement of viruses from cell to cell colocalizes to MTs, suggesting a role for cytoskeleton and possibly molecular motors in the transport of protein from cell to cell through plasmodesmata. These studies hint at an interesting possibility that motors may play a role in transporting macromolecules from cell to cell and influence cell-to-cell communication. Immunolocalization studies have also detected myosin associated with plasmodesmata. A recent study using an antibody to a cloned Class VIII Arabidopsis myosin ATM1 (At ATM) localized this myosin to the plasmodesmata and the plasma membrane regions involved in the assembly of new cell walls. Earlier work suggested that actin was involved in regulation of plasmodesmal transport. Other studies using antibodies to animal myosins in root tissues of *Allium cepa*, *Zea mays*, and *Hordeum vulagare* have also indicated the presence of myosin in the plasmodesmata.

3.6 Other Functions

It appears that the certain plant motors may perform unexpected novel functions.

For example, two plant kinesins (NCAK1 and NCAK2) have been shown interact with NPK1, a MAP kinase–kinase–kinase, and stimulate its activity. Motors in plants may also be involved in regulating cell polarity, development, and differentiation by transporting the proteins or mRNA to localized regions within the cell. There is substantial indirect evidence indicating that asymmetric localization of mRNA (e.g ASH1 mRNA in yeast), protein (e.g. calmodulin in Drosophila photoreceptors, a MAPKKK in tobacco BY2 cells), and RNA/protein complexes in some eukaryotic cells is dependent on cytoskeletal motors.

4 Regulation of Motors

Although a large number of kinesins have been characterized from diverse organisms (www.blocks.fhcrc.org/~kinesin/), the mechanism that regulate the activity/function of MT-dependent motors is limited both in plants and animals.

4.1 Calcium/Calmodulin

Calcium is a key messenger in transducing many hormonal and environmental signals in plants. Calmodulin, a ubiquitous calcium-binding protein in eukaryotes, is highly conserved in eukaryotes and regulates many diverse cellular functions by modulating the activity of the proteins that interact with it. Upon binding calcium, calmodulin becomes activated via a conformational change and is able to regulate the activity of its target proteins. Cytoplasmic streaming of characean cells is regulated by the cytosolic calcium concentration. Elevated levels of calcium inhibit cytoplasmic streaming and the movement of organelles. Movement of pollen tube organelles along actin cables is inhibited by calcium, suggesting that the activity of actin-based motors is sensitive to calcium levels. The mechanisms by which calcium regulates the activity of these motors is just beginning to emerge. All myosins in Arabidopsis possess two or more putative calmodulin-binding motifs, suggesting that myosins in plants are likely to be regulated by calmodulin in response to changes in intracellular calcium. Two myosin heavy chains (170 and 175 kDa) in plants associate with calmodulin, indicating that calmodulin serves as a light chain for these myosins. Calcium inhibits the myosin motor activity as well as myosin-activated ATPase activity in plants. The inhibition of myosin activity in the presence of calcium appears to be due to dissociation of calmodulin from the heavy chain since calcium inhibition of motor activity can be restored by exogenous addition of calmodulin. The gene encoding 175-kDa myosin has been cloned. Sequence analysis indicates that it is a Class X1 myosin with six IQ motifs. This myosin has been shown to move processively on actin filaments.

KCBP and recently discovered kinesin C are the only kinesins among the kinesin superfamily that bind calmodulin. The binding of activated calmodulin to KCBP inhibits its interaction with MTs via the calmodulin-binding domain, resulting in the inhibition of MT-dependent ATPase activity and motility. These studies strongly suggest that the binding of calmodulin to KCBP affects MT-binding regions, resulting in inhibition of KCBP binding to MTs and dissociation of KCBP/MT complex. Crystal structure of

KCBP motor domain with its calmodulin-binding domain indicates that calmodulin binding to KCBP blocks MT-binding sites on the motor. The calmodulin effect on KCBP suggests that activated calmodulin can act as a molecular switch to down-regulate the activity of KCBP. On the basis of *in vitro* studies with KCBP, it is reasonable to speculate that spatial and temporal changes in free cytosolic calcium levels in response to signals are likely to regulate KCBP activity in the cell.

4.2 Protein Phosphorylation

The members of BimC family have a conserved "BimC Box" in which threonine residue is phosphorylated by cdc2-kinase during cell division. Targeting of the kinesins to spindle is regulated by phosphorylation. TKRP125, all three members of Arabidopsis TKRP-like proteins (AtKRPa, b, and c), and DcKRP120-1 and DcKRP120-2 that belong to BimC family have the conserved BimC box, suggesting that phosphorylation of these kinesins may be involved in targeting them to right cellular location. Recently, it was found that the tail region of KCBP also interacts with a protein kinase (KIPK, KCBP interacting protein kinase) in the yeast two-hybrid system. The association of KIPK with KCBP suggests regulation of KCBP or KCBP-associated proteins by phosphorylation and/or KCBP is involved in targeting KIPK to proper cellular location. The tobacco NPK1, a mitogen-activated protein kinase–kinase–kinase, NPK1, interacts with NACKs. This interaction results in activation and autophosphorylation of NPK1 and phosphorylation of NACK1. Although the significance of NACK1 phosphorylation is not known, NPK1 phosphorylation targets it to the equatorial zone of the phragmoplasts. A cyclin-dependent kinase (CDK) has been shown to interact with two Arabidopsis kinesins (KCA1 and KCA2). This interaction results in phosphorylation of kinesins, suggesting that the activity/function of these motors may be regulated by CDK.

5 Concluding Remarks

The complete sequencing of several phylogenetically diverse model organisms ranging from a unicellular eukaryote to highly complex multicellular animals and plant has allowed comparative analysis of molecular motors in eukaryotes. Plants contain a large number of motors. About 0.4% of the predicted genes in Arabidopsis encode molecular motors. The comparative analysis provides a framework for future functional studies with plant kinesins. The number of myosins in Arabidopsis is comparable to other multicellular organisms of similar genome size. However, plant myosins are unlike myosins from any other organism except algae. Nonplant myosins contain a variety of other known domains that are lacking in plants. Furthermore, Arabidopsis has a surprisingly large number of kinesins among the completed eukaryotic genomes. The role of molecular motors in many plant cellular processes is beginning to be unraveled. Although the functions of most of the plant motors remains to be determined, phylogenetic analysis and known domains in these proteins are providing clues to their function, which can be tested empirically. Many Arabidopsis kinesins do not fall into any known subfamilies of kinesins, and several Arabidopsis kinesins are not

present in other organisms and are likely to represent new plant-specific subfamilies. Analysis of expression of each motor and cellular localization using reporter fusions coupled with studies using the loss-of-function and gain-of-function mutants should help understand the functions of these motors. Identification of the proteins that are associated with motors is also important to our understanding of their roles in plants. Knockout mutants for almost all Arabidopsis motors are already available from SIGnAL (http://signal.salk.edu/). Because of the large number of motors and likely functional redundancy or overlap with other motors, it will be necessary to create not only single mutants but also double or triple mutants. Analysis of the function and regulation of plant motors will be an exciting area of research in plant cell biology and is bound to throw many surprises.

Acknowledgment

I thank members of my laboratory and all my collaborators for their contributions. I wish to thank Irene Day for preparing the figure and critically reading the manuscript. Research on motors in my laboratory is supported by grants from the National Science Foundation (MCB-9630782 and MCB-0079938).

See also E-Cell: Computer Simulation of the Cell; Metabolic Basis of Cellular Energy; Motor Proteins; Nucleic Acid and Protein Single Molecule Detection and Characterization.

Bibliography

Books and Reviews

Asada, T., Collings, D. (1997) Molecular motors in higher plants, *Trends Plant Sci.* **2**, 29–37.

Hepler, P.K., Valster, A., Molchan, T., Vos, J.W. (2002) Roles for kinesin and myosin during cytokinesis, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **357**, 761–766.

Hirokawa, N. (1998) Kinesin and dynein superfamily proteins and the mechanism of organelle transport, *Science* **279**, 519–526.

Liu, B., Lee, Y.R.J. (2001) Kinesin-related proteins in plant cytokinesis, *J. Plant Growth Regul.* **20**, 141–150.

Oppenheimer, D.G. (1998) Genetics of plant cell shape, *Curr. Opin. Plant Biol.* **1**, 520–524.

Reddy, A.S.N. (2001) Molecular motors and their functions in plants, *Int. Rev. Cytol. & Cell Biol.* **204**, 97–178.

Reddy, A.S.N., Day, I.S. (2000) The role of the cytoskeleton and a molecular motor in trichome morphogenesis, *Trends Plant Sci.* **5**, 503–505.

Smith, L.G. (1999) Divide and Conquer: cytokinesis in plant cells, *Curr. Opin. Plant Biol.* **2**, 447–453.

Vale, R.D. (2003) The molecular motor toolbox for intracellular transport, *Cell* **112**, 467–480.

Primary Literature

Arn, E.A., MacDonald, P.M. (1998) Motors driving mRNA localization: new insights from in vivo imaging, *Cell* **95**, 150–154.

Asada, T., Collings, D. (1997) Molecular motors in higher plants, *Trends Plant Sci.* **2**, 29–37.

Asada, T., Sonobe, S., Shibaoka, H. (1991) Microtubule translocation in the cytokinetic apparatus of cultured tobacco cells, *Nature* **350**, 238–241.

Baluska, F., Wojtaszek, P., Volkmann, D., Barlow, P. (2003) The architecture of polarized cell growth: the unique status of elongating plant cells, *Bioessays* **25**, 569–576.

Barroso, C., Chan, J., Allan, V., Doonan, J., Hussey, P., Lloyd, C. (2000) Two kinesin-related proteins associated with the cold-stable cytoskeleton of carrot cells: characterization of a novel kinesin, DcKRP120-2, *Plant J* **24**, 859–868.

Baskin, T.I., Cande, W.Z. (1990) The structure and function of the mitotic spindle in flowering plants, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **41**, 277–315.

Berg, J.S., Powell, B.C., Cheney, R.E. (2001) A millennial myosin census, *Mol. Biol. Cell* **12**, 780–794.

Bezanilla, M., Horton, A.C., Sevener, H.C., Quatrano, R.S. (2003) Phylogenetic analysis of new plant myosin sequences, *J. Mol. Evol.* **57**, 229–239.

Bowser, J., Reddy, A.S.N. (1997) Localization of a kinesin-like calmodulin-binding protein in dividing cells of Arabidopsis and tobacco, *Plant J.* **12**, 1429–1437.

Cai, G., Romagnoli, S., Moscatelli, A., Ovidi, E., Gambellini, G., Tiezzi, A., Cresti, M. (2000) Identification and characterization of a novel microtubule-based motor associated with membranous organelles in tobacco pollen tubes, *Plant Cell* **12**, 1719–1736.

Cyr, R.J., Palevitz, B.A. (1995) Organization of cortical microtubules in plant cells, *Curr. Opin. Cell Biol.* **7**, 65–71.

Day, I.S., Miller, C., Golovkin, M., Reddy, A.S.N. (2000) Interaction of a kinesin-like calmodulin-binding protein with a protein kinase, *J. Biol. Chem.* **275**, 13737–13745.

Deavours, B.E., Reddy, A.S.N., Walker, R.A.N. (1998) Ca^{2+}/calmodulin regulation of the Arabidopsis kinesin-like calmodulin-binding protein, *Cell Motil. Cytoskeleton* **40**, 408–416.

Desai, A., Verma, S., Mitchison, T.J., Walczak, C.E. (1999) Kin I kinesins are microtubule-destabilizing enzymes, *Cell* **96**, 69–78.

Euteneur, U., Jackson, W.T., McIntosh, J.R. (1982) Polarity of spindle microtubules in *Haemanthus* endosperm, *J. Cell Biol.* **94**, 644–653.

Folkers, U., Kirik, V., Schobinger, U., Falk, S., Krishnakumar, S., Pollock, M.A., Oppenheimer, D.G., Day, I.S., Reddy, A.S.N., Jurgens, G., Hulskamp, M. (2002) The cell morphogenesis gene *ANGUSTIFOLIA* encodes a CtBP/BARS-like protein and is involved in the control of the microtubule cytoskeleton, *EMBO J.* **21**, 1280–1288.

Goddard, R.H., Wick, S.M., Silflow, C.D., Snustad, D.P. (1994) Microtubule components of the plant cell cytoskeleton, *Plant Physiol.* **104**, 1–6.

Goldstein, L.S.B., Philip, A.V. (1999) The road less traveled: Emerging principles of kinesin motor utilization, *Annu. Rev. Cell Dev. Biol.* **15**, 141–183.

Gunning, B.E., Wick, S.M. (1985) Preprophase bands, phragmoplasts, and spatial control of cytokinesis, *J. Cell. Sci. Suppl.* **2**, 157–179.

Hayama, T., Shimmen, T., Tazawa, M. (1979) Participation of Ca^{2+} in cessation of cytoplasmic streaming induced by membrane excitation in *Characeae* internodal cells, *Protoplasma* **99**, 305–321.

Hepler, P.K., Hush, J.M. (1996) Behavior of microtubules in living plant cells, *Plant Physiol.* **112**, 455–461.

Heslop-Harrison, J., Heslop-Harrison, Y. (1989) Myosin associated with the surface of organelles, vegetative nuclei and generative cells in angiosperm pollen grains and tubes, *J. Cell Sci.* **94**, 319–325.

Heslop-Harrison, J., Heslop-Harrison, Y. (1990) Dynamic aspects of apical zonation in the angiosperm pollen tube, *Sex. Plant Reprod.* **3**, 187–194.

Hodge, T., Cope, M.J. (2000) A myosin family tree, *J. Cell Sci.* **113**, 3353–3354.

Holweg, C., Nick, P. (2004) Arabidopsis myosin XI mutant is defective in organelle movement and polar auxin transport, *Proc. Natl. Acad. Sci. U S A* **101**, 10488–10493.

Hush, J.M., Wadsworth, P., Callaham, D.A., Hepler, P.K. (1994) Quantification of microtubule dynamics in living plant cells using fluorescence redistribution after photobleaching, *J. Cell Sci.* **107**, 775–784.

Itoh, R., Fujiwara, M., Yoshida, S. (2001) Kinesin-related proteins with a mitochondrial targeting signal, *Plant Physiol.* **127**, 724–726.

Jiang, S., Ramachandran, S. (2004) Identification and molecular characterization of myosin gene family in Oryza sativa genome, *Plant Cell Physiol.* **45**, 590–599.

Kashiyama, T., Kimura, N., Mimura, T., Yamamoto, K. (2000) Cloning and characterization of a myosin from characean alga, the fastest motor protein in the world, *J. Biochem. (Tokyo)* **127**, 1065–1070.

Kikuyama, M., Tazawa, M. (1982) Ca^{2+} ion reversibly inhibits the cytoplasmic streaming of *Nitella*, *Protoplasma* **113**, 241–243.

Kim, A.J., Endow, S.A. (2000) A kinesin family tree, *J. Cell Sci.* **113**, 3681–3682.

Kinkema, M., Wang, H., Schiefelbein, J. (1994) Molecular analysis of the myosin gene family in Arabidopsis thaliana, *Plant Mol. Biol.* **26**, 1139–1153.

Knight, A.E., Kendrick-Jones, J. (1993) A myosin-like protein from a higher plant, *J. Mol. Biol.* **231**, 148–154.

Kohno, T., Shimmen, T. (1988) Mechanism of Ca^{2+} inhibition of cytoplasmic streaming in lily pollen tubes, *J. Cell Sci* **91**, 501–509.

Kohno, T., Ishikawa, R., Nagai, T., Kohama, K., Shimmen, T. (1992) Partial purification of myosin from lily pollen tubes by monitoring with an in vitro motility assay, *Protoplasma* **170**, 77–85.

Kohno, T., Okagaki, T., Kohama, K., Shimmen, T. (1991) Pollen tube extract supports the movement of actin filaments in vitro, *Protoplasma* **161**, 75–77.

Kong, L.J., Hanley-Bowdoin, L. (2002) A geminivirus replication protein interacts with a protein kinase and a motor protein that display different expression patterns during plant development and infection, *Plant Cell* **14**, 1817–1832.

Krishnakumar, S., Oppenheimer, D.G. (1999) Extragenic suppressors of the Arabidopsis zwi-3 mutation identify new genes that function in trichome branch formation and pollen tube growth, *Development* **126**, 3079–3088.

Lawrence, C.J., Malmberg, R.L., Muszynski, M.G., Dawe, R.K. (2002) Maximum likelihood methods reveal conservation of function among closely related kinesin families, *J. Mol. Evol.* **54**, 42–53.

Lawrence, C.J., Morris, N.R., Meagher, R.B., Dawe, R.K. (2001) Dyneins have run their course in plant lineage, *Traffic* **2**, 362–363.

Lazarowitz, S.G., Beachy, R.N. (1999) Viral movement proteins as probes for intracellular and intercellular trafficking in plants, *Plant Cell* **11**, 535–548.

Lee, Y-R.J., Liu, B. (2000) Identification of a phragmoplast-associated kinesin-related protein in higher plants, *Curr. Biol.* **10**, 797–800.

Lee, Y.R., Giang, H.M., Liu, B. (2001) A novel plant kinesin-related protein specifically associates with the phragmoplast organelles, *Plant Cell* **13**, 2427–2439.

Liu, B., Lee, Y.R., Pan, R. (2003) Identification of kinesin-related proteins in the phragmoplast, *Cell Biol. Int.* **27**, 227–228.

Liu, L., Zhou, J., Pesacreta, T.C. (2001) Maize myosins: diversity, localization, and function, *Cell Motil. Cytoskeleton* **48**, 130–148.

Lloyd, C., Hussey, P. (2001) Microtubule-associated proteins in plants–why we need a MAP, *Nat. Rev. Mol. Cell Biol.* **2**, 40–47.

Lloyd, C.W. (Ed.) (1991) *The Cytoskeletal Basis of Plant Growth and Form*, Academic Press, New York.

Marc, J., Granger, C.L., Brincat, J., Fisher, D.D., Kao, T., McCubbin, A.G., Cyr, R.J. (1998) A GFP-MAP4 reporter gene for visualizing cortical microtubule rearrangements in living epidermal cells, *Plant Cell* **10**, 1927–1940.

Marcus, A.I., Li, W., Ma, H., Cyr, R.J. (2003) A kinesin mutant with an atypical bipolar spindle undergoes normal mitosis, *Mol. Biol. Cell* **14**, 1717–1726.

Marcus, A.I., Ambrose, J.C., Blickley, L., Hancock, W.O., Cyr, R.J. (2002) Arabidopsis thaliana protein, ATK1, is a minus-end directed kinesin that exhibits non-processive movement, *Cell Motil. Cytoskeleton* **52**, 144–150.

Mas, P., Beachy, R.N. (2000) Role of microtubules in the intracellular distribution of tobacco mosaic virus movement protein, *Proc. Natl. Acad. Sci. U S A* **97**, 12345–12349.

Mathur, J., Chua, N.H. (2000) Microtubule stabilization leads to growth reorientation in Arabidopsis trichomes, *Plant Cell* **12**, 465–477.

Mathur, J., Mathur, N., Hulskamp, M. (2002) Simultaneous visualization of peroxisomes and cytoskeletal elements reveals actin and not microtubule-based peroxisome motility in plants, *Plant Physiol.* **128**, 1031–1045.

Mathur, J., Spielhofer, P., Kost, B., Chua, N. (1999) The actin cytoskeleton is required to elaborate and maintain spatial patterning during trichome cell morphogenesis in *Arabidopsis thaliana*, *Development* **126**, 5559–5568.

Matsui, K., Collings, D., Asada, T. (2001) Identification of a novel plant-specific kinesin-like protein that is highly expressed in interphase tobacco BY-2 cells, *Protoplasma* **215**, 105–115.

Mitsui, H., Hasezawa, S., Nagata, T., Takahashi, H. (1996) Cell cycle-dependent accumulation of a kinesin-like protein, KatB/C, in synchronized tobacco BY-2 cells, *Plant Mol. Biol.* **30**, 177–181.

Mitsui, H., Nakatani, K., Yamaguchi-Shinozaki, K., Shinozaki, K., Nishikawa, K., Takahashi, H. (1994) Sequencing and characterization of the kinesin-related genes *katB* and

katC of *Arabidopsis thaliana*, *Plant Mol. Biol.* **25**, 865–876.

Moscatelli, A., Del Casino, C., Lozzi, L., Cai, G., Scali, M., Tiezzi, A., Cresti, M. (1995) High molecular weight polypeptides related to dynein heavy chains in Nicotiana tabacum pollen tubes, *J. Cell Sci.* **108**, 1117–1125.

Nagai, R. (1993) Regulation of intracellular movements in plant cells by environmental stimuli, *Int. Rev. Cytol.* **145**, 251–310.

Narasimhulu, S.B., Reddy, A.S.N. (1998) Characterization of microtubule binding domains in the Arabidopsis kinesin-like calmodulin-binding protein, *Plant Cell* **10**, 957–965.

Narasimhulu, S.B., Kao, Y-L., Reddy, A.S.N. (1997) Interaction of Arabidopsis kinesin-like calmodulin-binding protein with tubulin subunits: Modulation by Ca^{2+}-calmodulin, *Plant J.* **12**, 1139–1149.

Nebenfuhr, A., Gallagher, L.A., Dunahay, T.G., Frohlick, J.A., Mazurkiewicz, A.M., Meehl, J.B., Staehelin, L.A. (1999) Stop-and-Go movements of plant Golgi stacks are mediated by the acto-myosin system, *Plant Physiol.* **121**, 1127–1142.

Nishihama, R., Ishikawa, M., Araki, S., Soyano, T., Asada, T., Machida, Y. (2001) The NPK1 mitogen-activated protein kinase kinase kinase is a regulator of cell-plate formation in plant cytokinesis, *Genes Dev.* **15**, 352–363.

Nishihama, R., Soyano, T., Ishikawa, M., Araki, S., Tanaka, H., Asada, T., Irie, K., Ito, M., Terada, M., Banno, H., Yamazaki, Y., Machida, Y. (2002) Expansion of the cell plate in plant cytokinesis requires a kinesin-like protein/MAPKKK complex, *Cell* **109**, 87–99.

Oppenheimer, D.G., Pollock, M.A., Vacik, J., Szymanski, D.B., Ericson, B., Feldmann, K., Marks, D. (1997) Essential role of a kinesin-like protein in Arabidopsis trichome morphogenesis, *Proc. Natl. Acad. Sci. USA* **94**, 6261–6266.

Preuss, M.L., Delmer, D.P., Liu, B. (2003) The cotton kinesin-like calmodulin-binding protein associates with cortical microtubules in cotton fibers, *Plant Physiol.* **132**, 154–160.

Reddy, A.S.N. (2001) Calcium: silver bullet in signaling, *Plant Sci.* **160**, 381–404.

Reddy, A.S.N., Day, I.S. (2001a) Analysis of the myosins encoded in the recently completed *Arabidopsis thaliana* genome sequence, *Genome Biol.* **2**, 24.21–24.17.

Reddy, A.S.N., Day, I.S. (2001b) Kinesins in the Arabidopsis genome: A comparative analysis among eukaryotes, *BMC Genomics* **2**, 2.1–2.13.

Reddy, A.S.N., Narasimhulu, S.B., Safadi, F., Golovkin, M. (1996a) A plant kinesin heavy chain-like protein is a calmodulin-binding protein, *Plant J.* **10**, 9–21.

Reddy, A.S.N., Safadi, F., Narasimhulu, S.B., Golovkin, M., Hu, X. (1996b) A novel plant calmodulin-binding protein with a kinesin heavy chain motor domain, *J. Biol. Chem.* **271**, 7052–7060.

Reddy, V.S., Reddy, A.S.N. (2002) The calmodulin-binding domain from a plant kinesin functions as a modular domain in conferring Ca^{2+}-CaM regulation to animal plus- and minus-end kinesins, *J. Biol. Chem.* **277**, 48058–48065.

Reddy, V.S., Reddy, A.S.N. (2004) Developmental and Cell-specific expression of KCBP/ZWICHEL is regulated by the intron and exon sequences of its upstream protein-coding gene, *Plant Mol. Biol.* **54**, 273–293.

Reddy, V.S., Day, I.S., Thomas, T., Reddy, A.S. (2004) KIC, a novel Ca^{2+} binding protein with one EF-hand motif, interacts with a microtubule motor protein and regulates trichome morphogenesis, *Plant Cell* **16**, 185–200.

Reichelt, S., Knight, A.E., Hodge, T.P., Baluska, F., Samaj, J., Volkmann, D., Kendrick-Jones, J. (1999) Characterization of the unconventional myosin VIII in plant cells and its localization at the post-cytokinetic cell wall, *Plant J.* **19**, 555–567.

Reilein, A.R., Rogers, S.L., Tuma, M.C., Gelfand, V.I. (2001) Regulation of molecular motor proteins, *Int. Rev. Cytol.* **204**, 179–238.

Rogers, G.C., Hart, C.L., Wedman, K.P., Scholey, J.M. (1999) Identification of kinesin-C, a calmodulin-binding carboxy-terminal kinesin in animal (*Strongylocentrotus purpuratus*) cells, *J. Mol. Biol.* **294**, 1–8.

Romagnoli, S., Cai, G., Cresti, M. (2003) In vitro assays demonstrate that pollen tube organelles use kinesin-related motor proteins to move along microtubules, *Plant Cell* **15**, 251–269.

Schwab, B., Mathur, J., Saedler, R., Schwarz, H., Frey, B., Scheidegger, C., Hulskamp, M. (2003) Regulation of cell expansion by the DISTORTED genes in Arabidopsis thaliana: actin controls the spatial organization of microtubules, *Mol. Genet. Genomics* **269**, 350–360.

Shimmen, T., Yokota, E. (2004) Cytoplasmic streaming in plants, *Curr. Opin. Cell. Biol.* **16**, 68–72.

Siddiqui, S.S. (2002) Metazoan Motor Models: Kinesin Superfamily in C. elegans, *Traffic* **3**, 20–28.

Smirnova, E., Reddy, A.S.N., Bowser, J., Bajer, A.S. (1998) A minus end-directed kinesin-like motor protein, KCBP, localizes to anaphase spindle poles in Haemanthus endosperm, *Cell Motil. Cytoskeleton* **41**, 271–280.

Smith, L.G. (2002) Plant cytokinesis: motoring to the finish, *Curr. Biol.* **12**, R206–R208.

Song, H., Golovkin, M., Reddy, A.S.N., Endow, S.A. (1997) *In vitro* motility of AtKCBP, a calmodulin-binding kinesin-like protein of Arabidopsis, *Proc. Natl. Acad. Sci. USA.* **94**, 322–327.

Staehelin, L.A., Hepler, P.K. (1996) Cytokinesis in higher plants, *Cell* **84**, 821–824.

Staiger, C.J., Lloyd, C.W. (1991) The Plant Cytoskeleton, *Curr. Opin. Cell Biol.* **3**, 33–42.

Strompen, G., El Kasmi, F., Richter, S., Lukowitz, W., Assaad, F.F., Jurgens, G., Mayer, U. (2002) The Arabidopsis *HINKEL* gene encodes a kinesin-related protein involved in cytokinesis and is expressed in a cell cycle-dependent manner, *Curr. Biol.* **12**, 153–158.

Sylvester, A.W. (2000) Division decisions and the spatial regulation of cytokinesis, *Curr. Opin. Plant Biol.* **3**, 58–66.

Szymanski, D.B., Marks, D.M., Wick, S.M. (1999) Organized F-actin is essential for normal trichome morphogenesis in Arabidopsis, *Plant Cell* **11**, 2331–2348.

Tamura, K., Nakatani, K., Mitsui, H., Ohashi, Y., Takahashi, H. (1999) Characterization of katD, a kinesin-like protein gene specifically expressed in floral tissues of Arabidopsis thaliana, *Gene* **230**, 23–32.

Tominaga, M., Kojima, H., Yokota, E., Orii, H., Nakamori, R., Katayama, E., Anson, M., Shimmen, T., Oiwa, K. (2003) Higher plant myosin XI moves processively on actin with 35 nm steps at high velocity, *EMBO J.* **22**, 1263–1272.

Van Gestel, K., Kohler, R.H., Verbelen, J.P. (2002) Plant mitochondria move on F-actin, but their positioning in the cortical cytoplasm depends on both F-actin and microtubules, *J. Exp. Bot.* **53**, 659–667.

Vanstraelen, M., Torres Acosta, J.A., De Veylder, L., Inze, D., Geelen, D. (2004) A plant-specific subclass of C-terminal kinesins contains a conserved a-type cyclin-dependent kinase site implicated in folding and dimerization, *Plant Physiol.* **135**, 1417–1429.

Vinogradova, M., Reddy, V.S., Reddy, A.S.N. and Fletterick, R.J. (2004) Crystal structure of KCBP: Regulation by calcium/calmodulin, *J. Biol. Chem.* **279**, 23504–23509.

Vos, J.W., Safadi, F., Reddy, A.S.N., Hepler, P.K. (2000) The Kinesin-like calmodulin binding protein is differentially involved in cell division, *Plant Cell* **12**, 979–990.

Wang, D.Y., Kumar, S., Hedges, S.B. (1999) Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi, *Proc. R. Soc. Lond. B. Biol. Sci.* **266**, 163–171.

Wang, Z., Pesacreta, T.C. (2004) A subclass of myosin XI is associated with mitochondria, plastids, and the molecular chaperone subunit TCP-1alpha in maize, *Cell Motil. Cytoskeleton* **57**, 218–232.

Wasteneys, G.O. (2002) Microtubule organization in the green kingdom: chaos or self-order? *J. Cell. Sci. Suppl* **115**, 1345–1354.

Williamson, R.E. (1976) *Actin and Motility in Plant Cells*, North-Holland Publications, Amsterdam, Netherlands.

Williamson, R.E. (1993) Organelle movements, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **44**, 181–202.

Wolenski, J.S. (1995) Regulation of calmodulin-binding myosins, *Trends Cell Biol.* **5**, 310–316.

Yamamoto, K., Hamada, S., Kashiyama, T. (1999) Myosins from plants, *Cell. Mol. Life Sci.* **56**, 227–232.

Yamashita, R.A., Sellers, J.R., Anderson, J.B. (2000) Identification and analysis of the myosin superfamily in Drosophila: a database approach, *J. Muscle Res. Cell Motil.* **21**, 491–505.

Yang, C.Y., Spielman, M., Coles, J.P., Li, Y., Ghelani, S., Bourdon, V., Brown, R.C., Lemmon, B.E., Scott, R.J., Dickinson, H.G. (2003) TETRASPORE encodes a kinesin required for male meiotic cytokinesis in Arabidopsis, *Plant J.* **34**, 229–240.

Yokota, E., Muto, S., Shimmen, T. (1999) Inhibitory regulation of higher-plant myosin by Ca^{2+} ions, *Plant Physiol.* **119**, 231–240.

Zhong, R., Burk, D.H., Morrison, W.H. 3rd, Ye, Z.H. (2002) A kinesin-like protein is essential for oriented deposition of cellulose microfibrils and cell wall strength, *Plant Cell* **14**, 3101–3117.

Molecular Nanotechnology: *see* Nanobiotechnology

Molecular Neurobiology, Single-Cell

Jennifer Spaethling, Emily Rozak, and James Eberwine
University of Pennsylvania, Philadelphia, PA, USA

1 Importance of Single-cell Molecular Experiments 497

2 Isolation Techniques 498
2.1 Microaspiration of Live Cells 498
2.2 Microaspiration of Fixed Cells 499
2.3 Laser Capture Microdissection 499

3 mRNA Profiling 499
3.1 *In situ* Hybridization 499
3.2 RT-PCR 500
3.3 qRT-PCR 501
3.4 aRNA Amplification 501
3.5 Microarrays 501
3.6 Experimental Examples of Profiling 502

4 mRNA Localization 503
4.1 Targeting and Trafficking 503
4.2 Translational Machinery 505

5 Protein Profiling 505
5.1 Immunohistochemistry 506
5.2 Protein Arrays 507
5.3 Mass Spectrometry 508

6 Protein Localization 508

7 Other Applications 508
7.1 Electrophysiological Measurements of Molecular Changes 508
7.2 Molecular Response to Cellular Input 509

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

8 Future Considerations 509

Acknowledgment 510

Bibliography 510
Books and Reviews 510
Primary Literature 510

Keywords

CNS
Central nervous system.

Dendrite
Postsynaptic neuronal processes branching from cell soma that receive information from presynaptic cells.

Expression Profile
Relative abundances of multiple RNAs.

Postsynaptic
Cell that receives a signal from a neuron.

Presynaptic
Cell that sends a signal to another neuron.

Proteomics
Analysis of multiple proteins simultaneously.

aRNA
Amplified antisense RNA.

RT-PCR
Reverse-transcribed PCR.

qRT-PCR
Quantitative reverse-transcribed PCR.

Synapse
Connection point between a presynaptic and postsynaptic neuron.

The brain is composed of hundreds of thousands of cell types, each making thousands of connections to other cells in order to communicate and to facilitate proper interactive functioning. One of the ways that cells respond to and propagate signals is to change the abundances of various signaling molecules within the cell. These molecules include RNA and proteins, which, through their unique abundances and degree of activation transmit accurate messages to responsive cells. To understand these changes, evaluation of molecular pathways in single cells is required. With the development of single-cell molecular biological analysis techniques, changes in RNA and protein abundances can be monitored. Using methods such as, aRNA amplification, qRT-PCR, IDAT and protein arrays it is possible to obtain these measurements. Such anatomical specificity has the potential to revolutionize molecular research by simplifying the complexities of a multicellular pathway to an experimentally approachable system.

1 Importance of Single-cell Molecular Experiments

With a system in charge of complex tasks like memory, senses, mood, and development, it is not surprising that the cellular composition of the brain is equally complex. Neurons and glia are the two dominant classes of cells in the central nervous system (CNS). Neurons are thought to be at the heart of all cell–cell communication; yet, recent data has highlighted the importance of glial cells in modulating neuronal function. The variety of neuronal and glial cell types and their connections allows for the distribution and integration of information involved in everyday activities. Over the last century, it has become clear that an understanding of the complexities of CNS functioning requires not only utilization of systems science but also the reductionist approach of characterization of the individual components of the system. The development of single-cell molecular techniques has enabled researchers to begin to differentiate between not only the activities of the different regions of the brain but also the different types of cells that make up each of these regions. It is estimated that there are hundreds of thousands of cell types in the brain, each expressing tens of thousands of mRNAs in differing abundances. These differences in the abundances of RNAs as well as selective gene expression in each cell enables the complexity whereby so few chemicals in the brain produce diverse cell-specific responses.

Prior to single-cell experimental procedures, experiments used whole brain or regional sections of cells to try to elucidate the molecular changes in response to outside stimuli. Whether such stimuli originate as a result of disease, behavior, or pharmacological stimulation, quantitation of the resulting cellular responses at the molecular level enables researchers to begin to decipher what underlies the subsequent response. Since an individual neuron can have thousands of connections (synapses) with other cells that feed and distribute information, it is likely that there are differences in expressed gene phenotypes in each cell depending upon the types and quantity of these connections.

Importantly, diseases such as Alzheimer's do not necessarily affect every cell in an affected brain region. Therefore, if a group of cells were evaluated to investigate the cellular response to a stimulus, it would be hard to separate the nonresponsive cells' molecular milieu from that of the responsive cells. Single-cell molecular methods now allow researchers to separate and study target genes or whole expression pathways in individual cells so that specific cellular phenotypes can be described.

2 Isolation Techniques

After determining and marking which cells are of interest in an experiment, single cells can be collected either from fixed-tissue samples or from live cells in culture.

2.1 Microaspiration of Live Cells

Live cells in primary cell culture, continuous cell culture, or slice culture can provide a source for single-cell RNA and protein analysis. These studies require the precise removal and collection of the cytoplasmic contents of individual cells. This precision is afforded through the use of electrophysiological sampling procedures, namely whole cell patch pipette harvesting (Fig. 1). In this procedure, a thin capillary tube is drawn to a very fine tip and used to impale a cell. The cytoplasmic contents (RNA and protein) can then be aspirated into the pipette and transferred to a microcentrifuge tube for further manipulation and analysis. This technique further allows for collection of electrophysiological data from the cell to be harvested. The marrying of analyses, electrophysiological and molecular, permits a detailed phenotypic portrait of the cell to be obtained. The analysis of the mRNA complement of the isolated cellular material requires amplification (discussed later in this chapter) of the RNA signal; consequently, the mRNA must initially be converted into cDNA. This is accomplished by reverse transcriptase and can be initiated either in the microcentrifuge tube or directly in the patch pipette. Very fine manipulations are possible in this approach, allowing not only for single or multiple cell collection but also for isolation of cellular subregions such as dendrites, axons, or cell bodies. Although this is a time-intensive technique, the fine control allows for very accurate analysis of the RNA content of single cells.

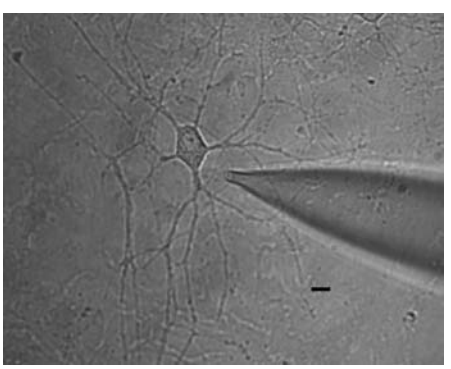

Fig. 1 Harvesting of a neuronal cell soma with a patch pipette – The manual harvesting of the cytoplasmic contents of a single rat hippocampal neuron in culture is shown. In this phase contrast image, the patch pipette used to harvest the cytoplasmic contents is in the foreground and the vast dendritic arborization associated with this neuron is clear. Aspiration through the patch pipette affords the negative pressure necessary to bring the cellular contents into the pipette. The black scale bar below the pipette corresponds to ~10 γm.

2.2 Microaspiration of Fixed Cells

Oftentimes it is difficult to obtain live cells from a particular brain region or from a particular species (e.g. human). Fixed-tissue sections offer the advantage of being able to examine cellular processes in many different cell types from essentially any organism. Fixation usually involves either chemical cross linking of proteins and nucleic acids or intracellular precipitation of these entities to preserve their localization and abundances during tissue processing. When collecting single cells from fixed-tissue sections, there are two commonly used techniques. The first technique is microaspiration, where micromanipulators and a microcontrolled vacuum are to remove the cellular constituents of the single cell. Often, the fixation protocol will dictate the success of this harvesting procedure. If the goal of these studies is to determine the RNA composition of an individual fixed cell, then as a prelude to nucleic acid amplification, cDNA synthesis is often performed *in situ* prior to microdissection of the identified cell.

2.3 Laser Capture Microdissection

The second method for single-cell collection is laser capture microdissection (LCM). This procedure uses an infrared laser beam to melt a transparent polymer transfer film (ethylene vinyl acetate polymer) so that it permeates the individual microscopically imaged cell. When the film is lifted, the desired cell is detached from all other surrounding cells and connective tissues. The film adheres to a transparent cap, which can subsequently be placed in a vial to perform any further desired manipulations. Although a laser is used, the energy emitted by the laser is thought to not interfere with the morphology of or the chemical composition of the isolated cell. It is important to note that issues of surrounding cellular contamination and method of fixation will impact the utility of LCM. With these caveats in mind, this method promises a fully animated, fast way to obtain a single cell sample, or many cells of a similar type depending upon the goals of the experiment. Additional newer methods for laser dissection have been developed that do not rely upon film transfer. These procedures use a low-power laser to separate the cell of interest from surrounding tissue, and a higher-powered beam transfers the cell to a vial. This LCM approach offers an advantage over the cap collection methodology of limiting the amount of surrounding-cell contamination.

3 mRNA Profiling

3.1 *In situ* Hybridization

There are two basic questions that can be asked when evaluating mRNA in single-cell experiments. The first is "is a mRNA present within the cell" and the second is "what is the quantity of the RNA." Table 1 summarizes the benefits and disadvantages of each mRNA profiling method. The second column refers to the quantitative nature of the amplification of the technique. The third column addresses whether the method measures the presence of individual mRNAs or the mRNA population. The fourth column refers to the amount of sample required for quantification. Lastly, the final column suggests how difficult the procedure is

Tab. 1 mRNA profiling.

	Quantitative	*Individual/population*	*Sample size*	*Simplicity*
In situ hybridization	No	Individual	Small	
RT-PCR	Semi	Individual	Small	
qRT-PCR	Semi	Individual	Small	
aRNA amplification	Yes	Population	Small	
Microarray	Yes	Population	Large	Advanced

to use. To evaluate the existence or coexistence of mRNAs in a single cell, a commonly employed technique is *in situ* hybridization (ISH). This technique uses labeled complementary RNA probes to bind to the RNA of interest. After permeabilizing the cell membrane, the probe is hybridized to the target RNA and, in turn, the RNA tag (Fluorescence, radioactivity, DAB, etc.) is visualized to determine the cellular location of the RNA. By using more than one type of fluorescence label, it is possible to determine the colocalization of multiple RNAs in a single cell. This technique can be used to localize RNAs to subregions of cells and, more recently, as a verification of other mRNA profiling techniques. However, ISH is quite difficult to quantitate; consequently, its use as a confirmation of RNA abundances determined by other procedures such as microarray analysis is limited. ISH is quite useful in differentiating between decreased mRNA levels due to decreased cell numbers versus decreased mRNA levels.

3.2 RT-PCR

Once a single cell has been harvested, there are many methods available to evaluate differences in mRNA levels due to variables such as disease or cellular stimulation. Since the quantity of mRNA in a single neuron or glia is quite small, often the first step in the evaluation process is amplifying the mRNA signal. This can be done by using a technique called *reverse-transcription polymerase chain reaction* (RT-PCR). After the mRNA is harvested from a cell, the first step in the amplification process is to anneal a specific primer to the mRNA of interest. Upon annealing, the reverse transcriptase will bind and, in the presence of dNTPs, create a complementary strand to the mRNA (cDNA). Next, the original mRNA is removed and a second primer binds to the cDNA strand to direct binding of a DNA polymerase that creates a copy of the first cDNA strand. This double-stranded cDNA is then used as a template for the iterative action of TAQ polymerase and the primers in the PCR amplification procedure. Such amplification is necessary since a large amount of the product is required for later quantitative analysis. During each round of PCR, the amount of DNA doubles, so in an average PCR reaction (27 rounds of amplification), the theoretical amplification yield is approximately 2^{27} or 10^8 cDNA copies (amplicons). Although this technique is used to quantify the mRNA abundance from a single cell, the PCR amplification is not linear and can lead to a skewing of amplicon abundance thereby producing aberrant results. mRNAs in the initial RNA pool that are present in lower

abundances are often not amplified to the same degree as more abundant mRNAs in the reaction mixture. This misrepresentative amplification is multiplied as the PCR amplification occurs. Also, secondary structure in the RNA or cDNA may lead to the TAQ polymerases falling off the strand, producing cDNAs that are not long enough to contain the second primer binding site, resulting in erroneously low quantitative results.

3.3 qRT-PCR

A convenient method that provides rapid RNA abundance measurements is quantitative RT-PCR (qRT-PCR) and is usually used to probe individual RNAs as opposed to the population of RNAs. Using the reverse-transcription method described above, cDNA is made from the original mRNA. During the PCR step, a quenched fluorescent probe is added, which binds selectively to the cDNA. When the polymerase moves along the strand, the quenching region of the probe is cleaved and the probe fluoresces. Using a spectrofluorimeter, the level of fluorescence is measured after each round of PCR. The number of rounds required to reach a threshold point can be compared between target RNAs in the sample. Quantitative results can only be achieved through quantitation during the linear phase of amplification prior to the point of PCR saturation. The amplified products corresponding to the RNA of interest are often normalized to a housekeeping gene for further comparison purposes and to limit variability. This technique limits the need for further quantification after amplification; however, many of the nonlinear problems described for RT-PCR also apply to this method.

3.4 aRNA Amplification

A second form of quantitative amplification is antisense RNA (aRNA) amplification. As originally described, this technique makes use of the fact that mRNA has a poly-A tail by using a complementary dT primer containing the T7 promoter to begin reverse transcription on the starting mRNA. Following the production of the first strand of cDNA, the original RNA is removed and the complement second strand cDNA is created, creating a functional T7 promoter. T7 RNA polymerase is then added to transcribe multiple antisense RNA copies from the new cDNA. Given that one round of RNA amplification produces up to 10 000-fold amplification, an important aspect of this technique is that only two rounds of amplification are required to obtain $\sim 10^6$ copies of the aRNA. This limits the problems of nonlinear amplification that are apparent from 25 to 40 rounds of RT-PCR.

3.5 Microarrays

Following a quantitative amplification protocol, microarrays are often used as a way to measure the abundance of thousands of genes simultaneously. This process usually begins with a solid support, either a glass slide or nylon membrane, with multiple oligonucleotides "printed" or deposited onto the surface in a square grid pattern. These grids can be "printed" by adding one nucleotide at a time until the complete sequence is present. Similarly, the grids can be made through direct deposition of cDNAs. The amplified RNA

or cDNA, made from the aRNA, is then radioactively or fluorescently labeled and hybridized to the microarray. Any immobilized oligonucleotides or cDNAs with the complementary sequence will bind the labeled aRNAs. After hybridization occurs, either fluorescence can be measured with a spectrofluorimeter or the radioactive signal is visualized using a phosphorimager. Since the pattern of DNAs immobilized on the membrane is known, the levels of fluorescence or radioactivity represent the quantity of RNA that is bound to each particular DNA on the array. There are two main benefits of using fluorescent dyes as opposed to radiolabeling; the first is that, instead of exposing the array to a phosphor screen for up to seven days, fluorescence can be measured immediately following hybridization. The second advantage is the possibility of using multiple dyes that fluoresce with different colors allowing for competitive hybridization comparison of RNAs profiles from different samples on the same array. Radioactively labeled probes have the advantage of being more sensitive; therefore, they are better able to detect a small amount of RNA.

3.6 Experimental Examples of Profiling

There are many areas of study where the techniques of RNA profiling can be used. One is in interpreting the molecular changes associated with disease state. For example, an area of important neurological disease research is understanding the genetic, genomic, and proteomic underpinnings of schizophrenia. The hope is that by targeting the molecular deficits in affected cells, researchers may be able to create chemicals that are therapeutically effective. For example, in the prefrontal cortex (PFC) of schizophrenia patients, there is a deficit of glutamate decarboxylase (GAD) mRNA which is necessary for the synthesis of the neurotransmitter GABA. Since the PFC is a brain region involved in working memory and is known to be impaired in schizophrenic individuals, understanding the pathways leading to these deficits might uncover novel treatments of this symptom. Single-cell mRNA profiles have shown that while some GABAergic neurons had normal expression other neurons exhibited expression far below control levels. Using molecular techniques, it is possible to investigate which subset of GABAergic neurons exhibits such deficits.

Such cellular specification highlights the selectivity of vulnerable cell types in schizophrenia. Further, recent experiments utilizing dual-label *in situ* hybridization have investigated the coincidence of GAD and parvalbumin (PV) or calretinin (CR). Since each type of GABAergic neuron in the PFC expresses only one of three calcium binding proteins, PV, CR or calbindin D-28, subdivision into three subtypes based on the presence of one of these proteins is possible. Using *in situ* hybridization, tissue samples from the PFC were labeled with probes for GAD and PV or CR, and direct correlations between GAD density and PV density were seen, whereas no correlation was seen between CR levels and GAD levels. Further insight into the relationship between GAD and PV RNA localization was achieved through dual labeling of cells with GAD and PV probes. In controls, GAD was often colocalized with PV, whereas, in schizophrenia patients, many cells expressing PV did not contain GAD mRNA. Other experiments have highlighted additional mRNAs that are downregulated in schizophrenia patients, such as those related to G-protein

subunits, glutamate receptors, synaptophysin, and other proteins involved in synaptic regulation.

Diagnosis of many neurological and psychiatric diseases occurs postmortem through examination of the morphology of brain tissue. More definitive diagnoses have been troublesome because of the relative lack of biomarkers for many diseases. By using well-prepared postmortem brain samples from patients, genomics and proteomics methodologies coupled with single-cell resolution may highlight the molecules and pathways involved in diseases like Alzheimer's and Parkinson's. It is hoped that such data can be deciphered and used for diagnostic purposes and eventually for possible treatments. In recent experiments, postmortem tissue from Alzheimer's patients was used to compare mRNA profiles in affected (tangle bearing) and nonaffected cells. qRT-PCR was used to compare quantities of synaptophysin mRNA, in patients and controls. Given the downregulation of synaptophysin in patient tissue samples, this RNA may serve as a new Alzheimer's biomarker.

A hallmark of Alzheimer's is the presence of intracellular protein aggregates called *neurofibrillary tangles* (NFT). It was hypothesized that differences in mRNA profiles existed in affected versus nonaffected cells. To investigate this, aRNA amplification from single tangle-bearing and nontangle bearing cells was used to generate probes that were used to screen microarrays. Large decreases in both amyloid-ß precursor as well as many dopamine (DA) receptor mRNA abundances were observed in the tangle-bearing neurons. Since these are all known to be involved in neurodegenerative diseases, this finding may be relevant to the etiology or propagation of the disease. Interestingly, lower levels of DA receptor binding (functional DA receptor) have previously been observed in Alzheimer's tissue; consequently, there appears to be a correlation between the functional and molecular studies. When working with a neurodegenerative disorder, it is important to note that decreases in RNA abundance may result from a decrease in cell number rather than from regulation of selected genes. Consequently, data showing that the RNA abundance for α1-ACT, a protein related to inflammation, is increased in Alzheimer's disease in single cells strongly suggests a link between inflammatory pathways and the disease state.

4 mRNA Localization

4.1 Targeting and Trafficking

Not only can there be variations in mRNA levels between cells but also within regions of a single cell. In neurons, subcellular regions such as dendrites, axons, and the cell soma can have different abundances of mRNA present. Similar nucleic amplification techniques can be used for analysis of cell subregion mRNA as are used for single-cell expression profiling studies. For example, individual dendrites can be manually harvested from a live neuron using a patch pipette and aspiration. The mRNA in the dendrite can be linearly amplified to quantifiable amounts using the same aRNA amplification techniques mentioned earlier. In recent studies, investigators have shown the possible differential distribution of large numbers of selective mRNAs in dendrites as compared to the cell soma. Since synapses on the dendrite are the place where the cell initially receives information from

presynaptic cells, it is reasonable to assume that the identity and abundances of dendritically localized mRNAs may be involved in the initial phase of some postsynaptic responses. This hypothesis is supported by the identification of some of the RNAs that are dendritically – targeted, such as those encoding NMDA, AMPA, and kainate classes of glutamate receptors, all of which are involved in postsynaptic responsiveness. Comparative analysis was performed on RNA that originated in the dendrites versus that of the soma. Differences in not only presence but also in quantity of some receptor subunits were observed between the two groups.

These data suggest several questions including the following: how do these differences in subcellular localization occur? Is the localization driven by differences in sequences such as the 3′-untranslated region (3′UTR) or are there RNA secondary or tertiary structural differences that help to target the mRNAs? Part of the answer comes from work investigating the somatodendritically localized microtubule-associated protein 2 (MAP2) mRNA. Normally, this RNA can be found in dendrites as well as in cell bodies (Fig. 2); however, if the 3′-UTR is removed, leaving just the coding and 5′-UTR, then the mRNA does not leave the cell soma. Also, if the 3′-UTR of the MAP2 RNA is fused to a nondendritically localized sequence, then the recombinant RNA is directed to the dendrites. It has now been shown that the 3′-UTR of many RNAs are necessary and sufficient for dendritic targeting. As our knowledge of RNA targeting sequences advances, the primary sequence as well as secondary and tertiary RNA structure will almost certainly be shown to influence RNA movement.

Advances in cell biological methodology are permitting experiments that monitor the real-time transport of RNAs into the dendrites to become feasible. An older method used pulse-chase experiments where RNA is made with incorporated radioactively labeled uridine (pulse) followed by chasing with nonlabeled uridine (chase). At different times after the chase, the movement of labeled RNA into different cellular regions can be observed. Also, visualization of the labeled RNA is done postfixing of cells, meaning that the observations are not made in real time. Further, the labeling used in pulse-chase experiments has been shown to change the stability of the mRNA. A vital dye, SYTO-14, fluoresces when it binds to nucleotides, making it possible to visualize RNA movement in live cells. RNAs were found to move at a rate of $0.1\ \mu m\ s^{-1}$, which is similar to the rate of transport when cargoes are moved along microtubules.

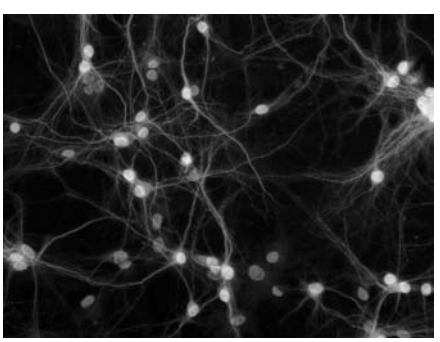

Fig. 2 Immunohistochemical localization of Map2 protein to the somatodendritic region of neurons – A rabbit polyclonal antibody raised against Map2 was applied to neurons of rat hippocampal neurons in primary culture. This first antibody was localized with a goat antirabbit secondary antibody that was fluorescein labeled (Green color). Green staining was observed in the fine hairlike structures that radiate from the cell soma. The nucleus is stained with DAPI. Glial cells in the primary cell culture system have a strong DAPI staining but no Map2 staining consistent with the absence of Map2 protein form glial cells (see color plate p. xxv).

Other types of tags can be used to visualize movement of mRNA such as a GFP-MS2 fusion protein that combines a fluorescent protein with the MS2 RNA binding protein. This is transfected into cells, and, when a fusion RNA construct is transfected into the same cells, the fusion protein binds to the engineered MS2 RNA sequence. Another *in vivo* marker of RNA is the molecular beacon probe, which is an antisense DNA oligonucleotide that is attached to a fluorophore and quencher. The structure forms a hairpin and binds with itself, quenching the fluorophore. When hybridization with the target RNA occurs, the hairpin is opened and fluorescence is emitted, allowing for very low background since only bound probes will fluoresce. This and modifications of the molecular beacon strategy are being used by several investigators to assess RNA trafficking in cells.

Different levels of mRNAs in dendrites as opposed to the cell soma have also been linked to developmental changes. For example, mRNA abundances have been monitored in dendritic growth cones through different stages of development. Amplification of RNA in growth cones at various stages of development followed by cDNA array analysis has identified RNAs that are present in the dendrite at each time point. As a function of developmental time, increasing numbers of the RNAs appeared in increased concentrations; others stayed in the soma for the entire time course of the experiment.

4.2 Translational Machinery

Even if there are different levels of mRNAs in dendrites, it is likely that much of actual cellular signaling takes place through the action of proteins. It is this rationale that suggests that dendritically localized mRNA may be translated into functioning proteins to facilitate dendritic-specific responses. Several years ago, in 1965, the ribosomal machinery necessary for protein translation was shown to exist in neuronal dendrites and to not be restricted to the cell soma. These data strongly suggested that localized translation could be a link to rapid changes in protein composition induced by synaptic modulation. This first observation of dendritically localized translational machinery was followed by several groups showing that polyribosomes localized preferentially near dendritic spines, the sites where synaptic connections with presynaptic neurons occurs. In further support of the dendritic protein synthesis hypothesis, other components of the translational machinery, including Golgi and endoplasmic reticulum, have been localized in dendrites. Currently, none of the intact neuron studies designed to prove dendritic protein synthesis have conclusively shown that protein synthesis can occur in dendrites. However, dendritic protein synthesis was confirmed using techniques to transfect isolated dendrites with exogenous RNAs encoding epitopes that when translated are immunologically or fluorescently detectable. These and subsequent studies have highlighted the complexities of dendritic protein synthesis and have further confirmed that integral membrane proteins can be synthesized and inserted into cellular membranes in dendrites.

5 Protein Profiling

While identifying changes in mRNA levels can give researchers a beginning point suggesting molecular causes of diseases or signaling pathway responses, it is usually

Tab. 2 Protein profiling.

	Quantitative	*Individual/population*	*Sample size*	*Simplicity*
Immunohistochemistry	No	Individual	Small	
Protein array	Yes	Population	Large	Advanced
IDAT	Yes	Individual	Small	
Mass spectrophotometry	Yes	Population	Large	Advanced

the protein abundances and their cellular localization that provides the most functionally useful information. Given that translational control of gene expression is a well-established fact, changes in mRNA levels do not necessarily correspond to changes in protein levels. RNA abundance differences suggest protein abundance differences but this should be viewed as hypothesis-generating data rather than as functionally causative biological changes. Measurement of cellular quantities of proteins and their posttranslational modifications will provide a more accurate picture of cellular responses to stimulation. Table 2 summarizes the different types of protein profiling techniques and the benefits and disadvantages of each method. Quantitative refers to whether the technique addresses just the presence of a protein or its cellular abundance. Individual/Population addresses whether specific antibodies are used or if the method is designed to measure many of the proteins present within the sample. Sample size refers to how much sample protein is required for the method. Finally, simplicity suggests how difficult the procedure is to use.

5.1 Immunohistochemistry

Understanding which proteins are present in a cell and at what concentrations can help researchers predict what the cellular response to stimulation will be or the results of aberrant stimulation such as occurs as a result of disease. Protein profiling in single cells is considerably more difficult than mRNA profiling because of the complexities of amino acid chemistry. However, methods for the evaluation of protein profile in a single cell are developing. The simplest way is to study the presence or absence of a certain protein using immunohistochemical techniques. In this methodology, antibodies that will bind to a particular protein are applied to a thin fixed-tissue section. The primary antibodies bind to the target protein and are detected with a secondary antibody that can be fluorescently tagged. The tissue is then microscopically analyzed and the presence of the protein visualized. By using different colors of fluorescence or other detection schemes, multiple proteins can be observed simultaneously. This process enables researchers to determine colocalization; if the proteins are colocalized, possible interacting proteins can be identified.

A recent example of the application of this methodology is shown in the examination of Lewy bodies (LB), which are intraneuronal protein aggregates involved in many neurodegenerative diseases such as Parkinson's and Diffuse Lewy Body Disease (DLBD). A major protein involved in LBs is known to be α-synuclein (α-syn), which is thought to interact with other proteins that have been linked to

these diseases. In other neurodegenerative diseases such as Alzheimer's, protein aggregates made of the protein tau are known as *Neurofibrillary Tangles* (NFT). Given the existence of aggregates in both diseases, it is reasonable to ask whether tau is localized to LBs. To address this question, immunohistochemistry was used to check for colocalization of α-syn and tau in LBs. Using a green fluorescent labeled (Cy3) antibody to detect one antigen and a red fluorescent labeled (Cy5) antibody to detect the second antigen, coexistence of the antigens was observed in the protein aggregates as the overlap of the fluorescence signal (yellow fluorescence). Eighty percent of the DLBD cases were found to exhibit coexistence of tau and α-syn in LBs. This high level of concurrency makes these proteins good candidates for biochemical analysis of protein interactions.

5.2 Protein Arrays

Other protein profiling techniques can provide quantitative analysis. Through the use of similar principles as microarrays, protein arrays or protein chips use protein binding as a way to separate and quantify proteins from single cells. There are two main ways for the initial protein targets to be attached to the array membrane. The first is to directly build the proteins on the arrays using a printing technique. The other way is to attach cDNAs to the membrane and run individual transcription and translation reactions, resulting in specific protein targets on the array. The latter technique eliminates worries of protein degradation and removes the need for protein purification before application onto traditional arrays. Protein arrays work on the principle of protein–protein interactions. Similar to coimmunoprecipitation, where antibodies are attached to beads and proteins that interact with the targeted protein are precipitated out of solution, protein microarrays allow for an initial screen of antibody–protein or protein–protein interactions. By applying the proteins from a single cell to known proteins on an array, it is possible to determine which proteins readily interact. Using fluorescently labeled probe proteins from cells, the amount of binding, and therefore the quantity of protein, can be measured. This method is useful when comparing protein profiles from disease and control samples. Similar methods use immobilized antibodies in place of target proteins to probe for labeled proteins.

The latest method in single-cell protein arrays uses what are called *nanoarrays*. These miniature spots use one-thousandth of the space that a traditional microarray spot occupies. This requires a much smaller probe and target protein samples which, in theory, makes it ideal for single-cell work. The current limiting factor to this method is that of detection sensitivity. A new technique called *Immuno-Detection Amplified by T7* RNA polymerase (IDAT) uses antibodies modified to contain a covalently attached double stranded DNA with an integral T7 RNA polymerase promoter. The anitbody binds to the antigen and this interaction is detected and quantified by T7 RNA amplification of the anitbody-attached DNA. The antibody is connected to a double-stranded oligonucleotide that contains a T7 promoter. This allows for specific amplification of the targeted RNA using the linear and quantifiable T7 RNA polymerase methodology.

5.3 Mass Spectrometry

Following protein array hybridization or traditional, two-dimensional gel electrophoresis to separate proteins, mass spectrometry (MS) can be used to definitively identify the proteins and their respective concentrations in the cells. Unlike fluorescently labeled proteins, MS techniques require no labeling, which eliminates any worries of labeling induced artifacts. Using only the proteins' masses and peptide fragment masses, MS separates and identifies proteins on the basis of time of flight (TOF) of fragments through the mass spectrometer. This procedure has been used successfully to identify and quantify down to 10 proteins from a single interacting target protein. In neurons, experiments using MALDI-MS have been performed to analyze the protein profiles of dendrites. Indeed, the advent of single-cell MALDI promises to identify the most-abundant proteins in the single-cell soma and potentially in isolated regions of the neuron.

6 Protein Localization

As previously emphasized, dendrites are the first areas of the cell to receive intercellular messages. Such intercellular communication causes overall changes in the cells that facilitate and result in complex brain function. The plasticity of synapses, or the ability of a neuron to adapt to different types of stimulation, is thought to be caused by changes in the abundance of locally functional proteins. It was historically thought that these changes occurred by transport of proteins from the soma; however, given the existence of dendritic protein synthesis, it is possible that differentially stimulated regions of a dendrite might respond in a subdendritic translationally selective manner. Experiments have been performed to address this issue in which isolated dendrites (free of the cell soma) were transfected with green fluorescent protein (GFP) mRNA so that local translation could be measured, *in vivo*, though observation of fluorescence. Interestingly, dendrites that were stimulated with the metabotropic glutamate receptor agonist, DHPG (a known stimulator of protein synthesis), exhibited an increase in fluorescence that could be blocked with a translational inhibitor, anisomycin. These experiments showed that the new GFP protein was being translated locally in the isolated dendrites. Further studies showed that the newly synthesized GFP was translated in hot spots that correlated with the location of ribosomes, which previously were shown to be preferentially located near areas of synaptic input. Also, data has been generated that shows that localized dendritic translational rates not only varied from those of the cell body but also within the dendrite. These heterogeneous and specialized translational activities of dendrites give insight into the signaling processes, which are likely involved in modulating the complex biology of intercellular communication within the brain.

7 Other Applications

7.1 Electrophysiological Measurements of Molecular Changes

The hippocampus is a brain region linked to learning and memory. It is also a region where neurogenesis, the generation of

new neurons, occurs. Experiments have been performed that connect these two facts, in an effort to understand the cellular events at the foundation of memory. Since new neurons are constantly being created in the hippocampus, there is always a subgroup of cells that are going through various stages of development. Investigators have determined the electrophysiological responses of neurons in each stage of development. The cells can be divided into five electrophysiologically distinct groups and single-cell RT-PCR was used to determine the developmental state of the neuron through the identification of stage-specific neuronal markers. Results showed that as development progresses, the hyperpolarity of the resting membrane potential of the cells increases and the input resistance of the cells decreases. The logical extension of these experiments is to identify the exact molecules and signaling pathways leading to these endpoints. This can be envisioned through the use of single-cell aRNA and microarray experiments coupled with proteomics.

7.2 Molecular Response to Cellular Input

One of the difficult aspects of psychiatric illness is the time course required for drug treatment to be effective. Most antipsychotic drugs take at least three weeks to begin to show their clinical efficacy. Drugs such as haloperidol are used for the treatment of diseases such as hyperactivity and mania and provide relief to the patient only if given chronically. Looking at gene expression profiles of cells that have been exposed to a drug can provide insight into the mechanisms resulting in the antipsychotic efficacy time-lag. In the rat model system, RNA was extracted from single cells in two different brain regions using LCM after administration of haloperidol for 1, 10, or 21 days. Radiolabeled aRNA was used to probe macroarrays containing 93 cDNAs of hypothesized relevance. After chronic administration, many of the mRNAs showed a steady baseline level or an increased expression level. At the intermediate time points, there were varying patterns. Some of the mRNAs spiked and later returned to a baseline level while others dropped far below the starting levels only to increase with chronic treatment. The complexities seen in the different patterns of time-dependent mRNA profiling suggest several avenues for dissection of the underlying reasons for the delayed efficacy in antipsychotic treatment.

8 Future Considerations

The last decade has seen an increase in the awareness of the importance of single-cell molecular analysis. The coordinate changes in mRNA abundance (transcription and degradation), protein abundance, posttranslational protein modifications, and the subcellular localization of RNA and proteins underlie all aspects of cellular functioning. As we progress into the twenty-first century, it is clear that we now need to not only quantitate the previously described issues *in vitro* but also to do so *in vivo*. Live cell imaging of biological processes such as transcription, translation, and subcellular movement will provide significant insight into normal cell functioning and subsequently into cell functioning as modified by disease state. To achieve these goals, it will be necessary for scientists in different disciplines to interact in an unprecedented manner. For instance, chemists making new fluorescent probes or other types of

molecular detectors, physicists helping to understand the folding dynamics of RNA and protein, and biologists defining the quantifiable correlates of cellular functioning will facilitate an understanding of *in vivo* biology.

Acknowledgment

The authors gratefully acknowledge Kevin Miyashiro for supplying Fig. 2 of this manuscript. This work was funded by NIH grants to J. Eberwine.

See also Brain Development; Electron Microscopy in Cell Biology; Neuron Chemistry; Pain Transduction: Gating and Modulation of Ion Channels.

Bibliography

Books and Reviews

Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M., Coleman, P. (1992) Analysis of gene expression in single live neurons, *Proc. Natl. Acad. Sci. U.S.A.* **89**(7), 3010–3014.

Eberwine, J., Miyashiro, K., Kacharmina, J.E., Job, C. (2001) Local translation of classes of mRNAs that are targeted to neuronal dendrites, *Proc. Natl. Acad. Sci. U.S.A.* **98**(13), 7080–7085.

Eberwine, J., Kacharmina, J.E., Andrews, C., Miyashiro, K., McIntosh, T., Becker, K., Barrett, T., Hinkle, D., Dent, G., Marciano, P. (2001) mRNA expression analysis of tissue sections and single cells, *J. Neurosci.* **21**(21), 8310–8314.

Ginsberg, S.D., Elarova, I., Ruben, M., Tan, F., Counts, S.E., Eberwine, J.H., Trojanowski, J.Q., Hemby, S.E., Mufson, E.J., Che, S. (2004) Single-cell gene expression analysis: implications for neurodegenerative and neuropsychiatric disorders, *Neurochem. Res.* **29**(6), 1053–1064.

Grant, S.G., Blackstock, W.P. (2001) Proteomics in neuroscience: from protein to network, *J. Neurosci.* **21**(21), 8315–8318.

Hinkle, D., Glanzer, J., Sarabi, A., Pajunen, T., Zielinski, J., Belt, B., Miyashiro, K., McIntosh, T., Eberwine, J. (2004) Single neurons as experimental systems in molecular biology, *Prog. Neurobiol.* **72**(2), 129–142.

Job, C., Eberwine, J. (2001) Localization and translation of mRNA in dendrites and axons, *Nat. Rev. Neurosci.* **2**(12), 889–898.

Kelz, M., Therianos, S., Dent, G., Coleman, P., Eberwine, J. (2002) *Single Cell Gene Expression Analysis: Insight into Alzheimers Disease and Other Neurological Disorders.* Science's SAGE KE.

Primary Literature

Aakalu, G., Smith, W.B., Nguyen, N., Jiang, C., Schuman, E.M. (2001) Dynamic visualization of local protein synthesis in hippocampal neurons, *Neuron* **30**(2), 489–502.

Ambrogini, P., Lattanzi, D., Ciuffoli, S., Agostini, D., Bertini, L., Stocchi, V., Santi, S., Cuppini, R. (2004) Morpho-functional characterization of neuronal cells at different stages of maturation in granule cell layer of adult rat dentate gyrus, *Brain Res.* **1017**(1–2), 21–31.

Angenstein, F., Greenough, W.T., Weiler, I.J. (1998) Metabotropic glutamate receptor-initiated translocation of protein kinase p90rsk to polyribosomes: a possible factor regulating synaptic protein synthesis, *Proc. Natl. Acad. Sci. U.S.A.* **95**(25), 15078–15083.

Bassell, G.J., Zhang, H., Byrd, A.L., Femino, A.M., Singer, R.H., Taneja, K.L., Lifshitz, L.M., Herman, I.M., Kosik, K.S. (1998) Sorting of beta-actin mRNA and protein to neurites and growth cones in culture, *J. Neurosci.* **18**(1), 251–265.

Bassell, G.J., Singer, R.H. (2001) Neuronal RNA localization and the cytoskeleton, *Results Probl. Cell Differ.* **34**, 41–56.

Becker, A.J., Chen, J., Paus, S., Normann, S., Beck, H., Elger, C.E., Wiestler, O.D., Blumcke, I. (2002) Transcriptional profiling in human epilepsy: expression array and single cell real-time qRT-PCR analysis reveal distinct

cellular gene regulation, *Neuroreport* **13**(10), 1327–1333.

Behar, L., Marx, R., Sadot, E., Barg, J., Ginzburg, I. (1995) cis-acting signals and trans-acting proteins are involved in tau mRNA targeting into neurites of differentiating neuronal cells, *Int. J. Dev. Neurosci.* **13**(2), 113–127.

Benson, D.L. (1997) Dendritic compartmentation of NMDA receptor mRNA in cultured hippocampal neurons, *Neuroreport* **8**(4), 823–828.

Bernhardt, R., Matus, A. (1984) Light and electron microscopic studies of the distribution of microtubule-associated protein 2 in rat brain: a difference between dendritic and axonal cytoskeletons, *J. Comp. Neurol.* **226**(2), 203–221.

Betsuyaku, T., Senior, R.M. (2004) Laser capture microdissection and mRNA characterization of mouse airway epithelium: methodological considerations, *Micron* **35**(4), 229–234.

Blichenberg, A., Schwanke, B., Rehbein, M., Garner, C.C., Richter, D., Kindler, S. (1999) Identification of a cis-acting dendritic targeting element in MAP2 mRNAs, *J. Neurosci.* **19**(20), 8818–8829.

Blichenberg, A., Rehbein, M., Muller, R., Garner, C.C., Richter, D., Kindler, S. (2001) Identification of a cis-acting dendritic targeting element in the mRNA encoding the alpha subunit of Ca2+/calmodulin-dependent protein kinase II, *Eur. J. Neurosci.* **13**(10), 1881–1888.

Bodian, D. (1965) A suggestive relationship of nerve cell RNA with specific synaptic sites, *Proc. Natl. Acad. Sci. U.S.A.* **53**, 418–425.

Chow, N., Cox, C., Callahan, L.M., Weimer, J.M., Guo, L., Coleman, P.D. (1998) Expression profiles of multiple genes in single neurons of Alzheimer's disease, *Proc. Natl. Acad. Sci. U.S.A.* **95**(16), 9620–9625.

Coss, R.G., Perkel, D.H. (1985) The function of dendritic spines: a review of theoretical issues, *Behav. Neural Biol.* **44**(2), 151–185.

Crair, M.C. (1999) Neuronal activity during development: permissive or instructive? *Curr. Opin. Neurobiol.* **9**(1), 88–93.

Crino, P.B., Eberwine, J. (1996) Molecular characterization of the dendritic growth cone: regulated mRNA transport and local protein synthesis, *Neuron* **17**(6), 1173–1187.

Crino, P., Khodakhah, K., Becker, K., Ginsberg, S., Hemby, S., Eberwine, J. (1998) Presence and phosphorylation of transcription factors in developing dendrites, *Proc. Natl. Acad. Sci. U.S.A.* **95**(5), 2313–2318.

Curran, S., McKay, J.A., McLeod, H.L., Murray, G.I. (2000) Laser capture microscopy, *Mol. Pathol.* **53**(2), 64–68.

Davis, L., Banker, G.A., Steward, O. (1987) Selective dendritic transport of RNA in hippocampal neurons in culture, *Nature* **330**(6147), 477–479.

de Moor, C.H., Richter, J.D. (2001) Translational control in vertebrate development, *Int. Rev. Cytol.* **203**, 567–608.

Denk, W., Strickler, J.H., Webb, W.W. (1990) Two-photon laser scanning fluorescence microscopy, *Science* **248**(4951), 73–76.

Eberwine, J., Cao, Y., Nair, S., Miyashiro, K., Mackler, S., Finnell, R., Surmeier, J., Dichter, M. (1995) Mechanisms of neuronal plasticity as analyzed at the single cell level, *Prog. Brain. Res.* **105**, 117–126.

Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A., Liotta, L.A. (1996) Laser capture microdissection, *Science* **274**(5289), 998–1001.

Estee Kacharmina, J., Crino, P.B., Eberwine, J. (1999) Preparation of cDNA from single cells and subcellular regions, *Methods Enzymol.* **303**, 3–18.

Estee Kacharmina, J., Job, C., Crino, P., Eberwine, J. (2000) Stimulation of glutamate receptor protein synthesis and membrane insertion within isolated neuronal dendrites, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11545–11550.

Evanko, D.S., Zhang, Q., Zorec, R., Haydon, P.G. (2004) Defining pathways of loss and secretion of chemical messengers from astrocytes, *Glia* **47**(3), 233–240.

Fasulo, W.H., Hemby, S.E. (2003) Time-dependent changes in gene expression profiles of midbrain dopamine neurons following haloperidol administration, *J. Neurochem.* **87**(1), 205–219.

Femino, A.M., Fay, F.S., Fogarty, K., Singer, R.H. (1998) Visualization of single RNA transcripts in situ, *Science* **280**(5363), 585–590.

Gardiol, A., Racca, C., Triller, A. (1999) Dendritic and postsynaptic protein synthetic machinery, *J. Neurosci.* **19**(1), 168–179.

Garner, C.C., Tucker, R.P., Matus, A. (1988) Selective localization of messenger RNA for cytoskeletal protein MAP2 in dendrites, *Nature* **336**(6200), 674–677.

Gazzaley, A.H., Benson, D.L., Huntley, G.W., Morrison, J.H. (1997) Differential subcellular

regulation of NMDAR1 protein and mRNA in dendrites of dentate gyrus granule cells after perforant path transection, *J. Neurosci.* **17**(6), 2006–2017.

Geiger, J.R., Melcher, T., Koh, D.S., Sakmann, B., Seeburg, P.H., Jonas, P., Monyer, H. (1995) Relative abundance of subunit mRNAs determines gating and Ca2+ permeability of AMPA receptors in principal neurons and interneurons in rat CNS, *Neuron* **15**(1), 193–204.

Grant, S.G., Husi, H. (2001) Proteomics of multiprotein complexes: answering fundamental questions in neuroscience, *Trends Biotechnol.* **19**(Suppl. 10), S49–S54.

Gutala, R.V., Reddy, P.H. (2004) The use of real-time PCR analysis in a gene expression study of Alzheimer's disease post-mortem brains, *J. Neurosci. Methods* **132**(1), 101–107.

Haab, B.B. (2003) Methods and applications of antibody microarrays in cancer research, *Proteomics* **3**(11), 2116–2122.

Hashimoto, T., Volk, D.W., Eggan, S.M., Mirnics, K., Pierri, J.N., Sun, Z., Sampson, A.R., Lewis, D.A. (2003) Gene expression deficits in a subclass of GABA neurons in the prefrontal cortex of subjects with schizophrenia, *J. Neurosci.* **23**(15), 6315–6326.

Haydon, P.G. (2003) Biological near-field microscopy, *Methods Enzymol.* **360**, 501–508.

Hemby, S.E., Ginsberg, S.D., Brunk, B., Arnold, S.E., Trojanowski, J.Q., Eberwine, J.H. (2002) Gene expression profile for schizophrenia: discrete neuron transcription patterns in the entorhinal cortex, *Arch. Gen. Psychiatry* **59**(7), 631–640.

Huang, E.P. (1999) Synaptic plasticity: regulated translation in dendrites, *Curr. Biol.* **9**, R168–R170.

Huber, K.M., Kayser, M.S., Bear, M.F. (2000) Role for rapid dendritic protein synthesis in hippocampal mGluR-dependent long-term depression, *Science* **288**(5469), 1254–1257.

Husi, H., Grant, S.G. (2001) Proteomics of the nervous system, *Trends Neurosci.* **24**(5), 259–266.

Ishizawa, T., Mattila, P., Davies, P., Wang, D., Dickson, D.W. (2003) Colocalization of tau and alpha-synuclein epitopes in Lewy bodies, *J. Neuropathol. Exp. Neurol.* **62**(4), 389–397.

Jakoi, E.R., Severt, W.L. (2000) Disruption of mRNA-RNP formation and sorting to dendritic synapses by antisense oligodeoxynucleotides, *Methods Enzymol.* **313**, 456–466.

Job, C., Eberwine, J. (2001) Identification of sites for exponential translation in living dendrites, *Proc. Natl. Acad. Sci. U.S.A.* **98**(23), 13037–13042.

Johansen, F.F., Lambolez, B., Audinat, E., Bochet, P., Rossier, J. (1995) Single cell RT-PCR proceeds without the risk of genomic DNA amplification, *Neurochem. Int.* **26**(3), 239–243.

Kamme, F., Salunga, R., Yu, J., Tran, D.T., Zhu, J., Luo, L., Bittner, A., Guo, H.Q., Miller, N., Wan, J., Erlander, M. (2003) Single-cell microarray analysis in hippocampus CA1: demonstration and validation of cellular heterogeneity, *J. Neurosci.* **23**(9), 3607–3615.

Kanai, Y., Hirokawa, N. (1995) Sorting mechanisms of tau and MAP2 in neurons: suppressed axonal transit of MAP2 and locally regulated microtubule binding, *Neuron* **14**(2), 421–432.

Knowles, R.B., Sabry, J.H., Martone, M.E., Deerinck, T.J., Ellisman, M.H., Bassell, G.J., Kosik, K.S. (1996) Translocation of RNA granules in living neurons, *J. Neurosci.* **16**(24), 7812–7820.

Lambolez, B., Audinat, E., Bochet, P., Crepel, F., Rossier, J. (1992) AMPA receptor subunits expressed by single Purkinje cells, *Neuron* **9**(2), 247–258.

Lynch, M., Mosher, C., Huff, J., Nettikadan, S., Johnson, J., Henderson, E. (2004) Functional protein nanoarrays for biomarker profiling, *Proteomics* **4**(6), 1695–1702.

Mackler, S.A., Brooks, B.P., Eberwine, J.H. (1992) Stimulus-induced coordinate changes in mRNA abundance in single postsynaptic hippocampal CA1 neurons, *Neuron* **9**(3), 539–548.

Mackler, S.A., Eberwine, J.H. (1993) Diversity of glutamate receptor subunit mRNA expression within live hippocampal CA1 neurons, *Mol. Pharmacol.* **44**(2), 308–315.

Mazumder, B., Seshadri, V., Fox, P.L. (2003) Translational control by the 3'-UTR: the ends specify the means, *Trends Biochem. Sci.* **28**(2), 91–98.

Mirnics, K., Middleton, F.A., Marquez, A., Lewis, D.A., Levitt, P. (2000) Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex, *Neuron* **28**(1), 53–67.

Miyashiro, K., Dichter, M., Eberwine, J. (1994) On the nature and differential distribution of

mRNAs in hippocampal neurites: implications for neuronal functioning, *Proc. Natl. Acad. Sci. U.S.A.* **91**(23), 10800–10804.

Miyashiro, K.Y., Beckel-Mitchener, A., Purk, T.P., Becker, K.G., Barret, T., Liu, L., Carbonetto, S., Weiler, I.J., Greenough, W.T., Eberwine, J. (2003) RNA cargoes associating with FMRP reveal deficits in cellular functioning in Fmr1 null mice, *Neuron* **37**(3), 417–431.

Mori, Y., Imaizumi, K., Katayama, T., Yoneda, T., Tohyama, M. (2000) Two cis-acting elements in the 3′ untranslated region of alpha-CaMKII regulate its dendritic targeting. [see comments], *Nat. Neurosci.* **3**(11), 1079–1084.

Morozov, V.N., Morozova, T.Y., Johnson, K.L., Naylor, S. (2003) Parallel determination of multiple protein metabolite interactions using cell extract, protein microarrays and mass spectrometric detection, *Rapid. Commun. Mass Spectrom.* **17**(21), 2430–2438.

Nitin, N., Santangelo, P.J., Kim, G., Nie, S., Bao, G. (2004) Peptide-linked molecular beacons for efficient delivery and rapid mRNA detection in living cells, *Nucleic Acids Res.* **32**(6), e58.

Pascual, O., Haydon, P.G. (2003) Synaptic inhibition mediated by glia, *Neuron* **40**(5), 873–875.

Pierce, J.P., van Leyen, K., McCarthy, J.B. (2000) Translocation machinery for synthesis of integral membrane and secretory proteins in dendritic spines, *Nat. Neurosci.* **3**(4), 311–313.

Racca, C., Gardiol, A., Triller, A. (1998) Cell-specific dendritic localization of glycine receptor alpha subunit messenger RNAs, *Neuroscience* **84**(4), 997–1012.

Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S., Rosen, B., Lau, A.Y., Walter, J.C., LaBaer, J. (2004) Self-assembling protein microarrays, *Science* **305**(5680), 86–90.

Rubakhin, S.S., Garden, R.W., Fuller, R.R., Sweedler, J.V. (2000) Measuring the peptides in individual organelles with mass spectrometry, *Nat. Biotechnol.* **18**(2), 172–175.

Rusakov, D.A., Richter-Levin, G., Stewart, M.G., Bliss, T.V. (1997) Reduction in spine density associated with long-term potentiation in the dentate gyrus suggests a spine fusion-and-branching model of potentiation, *Hippocampus* **7**(5), 489–500.

Service, R.F. (2001) Proteomics. Searching for recipes for protein chips, *Science* **294**(5549), 2080–2082.

Shi, S., Hayashi, Y., Esteban, J.A., Malinow, R. (2001) Subunit-specific rules governing AMPA receptor trafficking to synapses in hippocampal pyramidal neurons, *Cell* **105**(3), 331–343.

Sokol, D.L., Zhang, X., Lu, P., Gewirtz, A.M. (1998) Real time detection of DNA. RNA hybridization in living cells, *Proc. Natl. Acad. Sci. U.S.A.* **95**(20), 11538–11543.

Steward, O., Levy, W.B. (1982) Preferential localization of polyribosomes under the base of dendritic spines in granule cells of the dentate gyrus, *J. Neurosci.* **2**(3), 284–291.

Surmeier, D.J., Eberwine, J., Wilson, C.J., Cao, Y., Stefani, A., Kitai, S.T. (1992) Dopamine receptor subtypes colocalize in rat striatonigral neurons, *Proc. Natl. Acad. Sci. U.S.A.* **89**(21), 10178–10182.

Tecott, L.H., Barchas, J.D., Eberwine, J.H. (1988) In situ transcription: specific synthesis of complementary DNA in fixed tissue sections, *Science* **240**(4859), 1661–1664.

Tiedge, H., Brosius, J. (1996) Translational machinery in dendrites of hippocampal neurons in culture, *J. Neurosci.* **16**(22), 7171–7181.

Tongiorgi, E., Righi, M., Cattaneo, A. (1998) A non-radioactive in situ hybridization method that does not require RNAse-free conditions, *J. Neurosci. Methods* **85**(2), 129–139.

Torre, E.R., Steward, O. (1996) Protein synthesis within dendrites: glycosylation of newly synthesized proteins in dendrites of hippocampal neurons in culture, *J. Neurosci.* **16**(19), 5967–5978.

Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D., Eberwine, J.H. (1990) Amplified RNA synthesized from limited quantities of heterogeneous cDNA, *Proc. Natl. Acad. Sci. U.S.A.* **87**(5), 1663–1667.

Weiler, I.J., Irwin, S.A., Klintsova, A.Y., Spencer, C.M., Brazelton, A.D., Miyashiro, K., Comery, T.A., Patel, B., Eberwine, J., Greenough, W.T. (1997) Fragile X mental retardation protein is translated near synapses in response to neurotransmitter activation, *Proc. Natl. Acad. Sci. U.S.A.* **94**(10), 5395–5400.

Zhang, H.T., Kacharmina, J.E., Miyashiro, K., Greene, M.I., Eberwine, J. (2001) Protein

quantification from complex protein mixtures using a proteomics methodology with single-cell resolution, *Proc. Natl. Acad. Sci. U.S.A.* **98**(10), 5497–5502.

Molecular Systematics and Evolution

Jeffrey H. Schwartz
University of Pittsburgh, Pittsburgh, PA, USA

1 **The Beginning of Molecular Systematics 518**

2 **The Molecular Assumption 519**

3 **DNA Hybridization 522**

4 **Mitochondrial DNA 526**

5 **DNA Sequences 527**

6 **"Evo-devo" 531**

7 **Positional Information and Shape 533**

8 **"Mutation" 534**

9 **Toward a Theory of Evolutionary Change 535**

10 **Molecules and Systematics: Looking Toward the Future 536**

Bibliography 538
Books and Reviews 538
Primary Sources 538

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

Keywords

African Apes
Chimpanzees, bonobos, and gorillas.

Anthropoidea
The so-called higher primates; the subordinal rank that subsumes New and Old World monkeys plus the hominoids.

Clade
A group of related organisms; a group of organisms united by common ancestry.

Derived Feature
A relative character state determined by how restrictively shared a feature under study is.

Hominids
Humans and their fossil relatives.

Hominoids
The group of anthropoid primates that includes gibbons and siamangs, chimpanzees, gorillas, orangutans, and humans.

Homology
Similarity due to common ancestry.

Large-bodied Hominoids
Orangutans, humans, and African apes.

Lesser (Small-bodied) Hominoids
The gibbons and siamangs.

Monophyletic Group
A clade.

Outgroup
A taxon that is outside of (a sister taxon of) the focal group of phylogenetic reconstruction.

Phylogeny Reconstruction
The generation of a theories of evolutionary relationships among taxa.

Primitive Feature
A relative character state determined by how broadly shared a feature under study is.

Primitive Retention
A feature that is shared by the members of a group by inheritance from a common ancestor.

Sister Taxon
A taxon that is the closest relative of another.

Systematics
The study of the pattern and relationships of organisms, including the identification of species and delineation of groups of taxa.

Taxic
Relating to taxa.

Taxon (Plural: Taxa)
Any taxonomic rank, for example, species, genus, family, order, and class.

Studies that rely on genetic and molecular information to address evolutionary questions fall roughly into two different categories: the reconstruction of the evolutionary relationships of organisms (including the times of divergence of groups or lineages), and the formation and emergence of morphological novelties that distinguish or characterize different organisms. The former endeavor is sometimes referred to as *molecular systematics*, and when it is applied to primates and especially the relatedness of humans and apes, ''molecular anthropology.'' The debate over the regularity of molecular change, resulting in a molecular clock, lies within the realm of molecular systematics, and especially that of molecular anthropology. In the past decade, in particular, the primary focus of molecular systematists has been DNA sequences, both nuclear and mitochondrial. Throughout, the assumption has been that the degree of similarity reflects the degree of evolutionary relatedness, because differences accrue once lineages have diverged from a common ancestor. The popularity of molecular systematics in recent years is also predicated on the notion of ''a law of large numbers,'' that is, the thousands of bases that produce DNA sequences. The question is whether these assumptions are sustainable.

The rather newly defined discipline of evolution and development, or ''evo-devo'' as it has been nicknamed, is less involved in the interpretation of molecular data for purposes of reconstructing evolutionary relationships than it is in trying to identify molecular elements – whether they be transcription factors or other kinds or classes of proteins, as well as ''genes'' themselves – that are relevant to, or participate in, the processes of development of structure and form. Here, what is important is not a particular RNA, DNA, or amino acid sequence, but the signal transduction pathways or sequences of communication between different molecules in the regulation of development and the origin of structure. Sometimes, hypotheses of when and at what level within groups of organisms specific features emerged are generated as a result of overlaying this information on a presumed theory of relationships of the organisms under consideration. It appears that insights from developmental genetics will prove to be more useful than sequence data alone for systematic and phylogenetic inquiry.

This entry will attempt to summarize the main aspects of these fields of inquiry, including their underlying assumptions, and suggest possible avenues for future research that might bring these disciplines together in light of new ways of thinking about evolution.

1
The Beginning of Molecular Systematics

The first inquiries into systematics and phylogenetic reconstruction by investigation of the "blood relationship" of organisms can be traced to a few individuals, of whom the best known is George H. F. Nuttall. In the research that preceded his monograph of 1904, *Blood Immunity and Blood Relationship: A Demonstration of Certain Blood-Relationships Amongst Animals by Means of the Precipitin Test for Blood*, Nuttall sought to demonstrate that the degree of similarity between animals in their blood serum proteins was a reflection of their evolutionary closeness. The idea was simple: produce an antiserum (antibodies or antigens) to blood serum proteins of one animal, mix the antiserum with blood serum proteins of another animal, and measure the amount of precipitin that settled out. The more profound the precipitation, the greater the similarity (because of a greater number of antibody–antigen binding sites), and, consequently, the closer the evolutionary relationship between the two organisms being compared.

The strength of the antiserum played an important role in determining just how many species might belong to the same group. A weak antiserum would provoke reactivity in the sera of only the most closely related of organisms, while a stronger solution would produce a precipitate when combined with the sera of a greater number of related animals. The more distantly related the groups being compared, the less the reactivity between serum and antiserum. In addition to strength of reactivity, Nuttall thought that the reaction rate was also an indicator of closeness of relatedness. In the end, he decided that "the zoological relationships between animals are best demonstrated by means of powerful antisera." Consequently, he concluded, "if we accept *the degree of blood reaction as an index of the degree of blood relationship* within the Anthropoidea [the so-called higher primates, New and Old World monkeys plus the hominoids, which includes gibbons, chimpanzees, gorillas, orangutans, and humans], then we find that the Old World apes are more closely allied to man than are the New World apes, and this is exactly in accordance with the opinion expressed by Darwin."

Although Nuttall believed that his approach accurately revealed the evolutionary relationships of animals, with the exception of reference in 1922 by the paleontologist, W. K. Gregory, to Nuttall having demonstrated "the anthropoid heritage of man," his work went largely unacknowledged until the 1960s, when various publications claimed to have demonstrated the evolutionary relationships of organisms through study of elements of their biochemistry.

2
The Molecular Assumption

Perhaps the most influential of these publications was by Emile Zuckerkandl and Linus Pauling on fetal and adult hemoglobin, which included a comparison of sequences between human, gorilla, horse, and fish. Although one might question the comparison of gill- with lung-bearing animals, the endeavor demonstrated that the fish was more dissimilar to humans than to the horse, while the gorilla was the most similar to humans in hemoglobin sequence. Since this pattern of taxic distance mirrored the morphologically accepted scheme of evolutionary relatedness of these animals, Zuckerkandl and Pauling suggested that their observations "can be understood at once if it is *assumed* [emphasis added] that in the course of time the hemoglobin-chain genes duplicate, [and] that the descendants of the duplicate genes 'mutate away' from each other." From this assumption, they felt justified in proposing the following: "[O]ver-all similarity must be an expression of evolutionary history," with descendants "mutating away" and becoming "gradually more different from each other." In other words, the more time that elapses after a lineage's divergence, the greater the molecular difference its succession of descendants would accumulate. Consequently, evolutionary closeness became synonymous with molecular difference. Also assumed in this nexus of assumption is the notion of constant and gradual unilinear change.

There is, however, nothing in Zuckerkandl and Pauling's work that justifies the notion that "overall similarity" is "an expression of evolutionary history." They merely assumed such a correspondence, just as they assumed that difference also meant change and the same kind of change, which, in turn, was achieved at a constant and gradual rate. However, the discovery of a hundred nucleotide differences between two taxa does not translate into the rate at which presumed substitutions occurred. One needs additional information, such as a suggestion of the time at which these differences began to accumulate. And then one has to assume that the rate of substitution over this period of time was gradual and that differences did not arise during a few concentrated phases of replacement.

Nevertheless, Zuckerkandl and Pauling's effort resulted in the "molecular assumption" – continual molecular change and its continual accumulation over evolutionary time – which would thereafter become the foundation of molecular systematics. As Adalgisa Caccone and Jeffrey Powell would write in 1989: "Virtually all molecular phylogenetic studies. . .have a major underlying assumption: the genetic similarity or difference among taxa is an indication of phylogenetic relatedness. Lineages that diverged more recently will be genetically more similar to one another than will be lineages with more ancient splits." If one embraces the assumption, the rest may follow logically, but not necessarily because of biological demonstration. One of the major extensions of the "molecular assumption" is the notion of a "molecular clock," the existence of which was promoted initially by Vincent Sarich and Allan Wilson.

Beginning with their first major paper in 1966, Sarich and Wilson sought to elucidate the evolutionary relationships of primates, and especially of humans and the large-bodied apes, using the technique of microcomplement fixation (MC'F), which requires only minute amounts of serum and antiserum for study of immunological

reactivity. Their molecule of choice was the larger blood serum protein, albumin. The degree of reactivity achieved between albumin and anti-albumin was translated into an "index of dissimilarity," which was subsequently referred to as *immunological distance* (ID). The closer an ID value was to 1.0, the greater was the overall molecular similarity and thus, given an assumed equivalence between ID and evolutionary relationship, the presumption of closeness of relatedness between the organisms under study. As an apparent check on the validity of this approach, Sarich and Wilson conceived of the "test of reciprocity." For example, if chimpanzee serum was cross-reacted with antihuman serum (produced by injecting another animal, for instance, a rabbit or chicken, with human serum) the first time around, human serum was cross-reacted with antichimpanzee serum the next time to see if the resultant ID values were reasonably similar. Predictably, the test of reciprocity usually confirmed the initial immunological finding.

The supposed evolutionary arrangement of the primates that Sarich and Wilson achieved was more consistent with the general pattern of primate relationships based on comparative morphology than it was in disagreement. Because of this concordance, they concluded that "the MC'F data are in qualitative agreement with the anatomical evidence, on the basis of which the apes, Old World monkeys, New World monkeys, prosimians, and nonprimates are placed in taxa which form a series of decreasing genetic relationship to man." In 1967, in their second article on this topic, Sarich and Wilson argued that molecular change must have occurred at a constant rate among all major groups of primates. In the third and last article of this series, which was also published in 1967, they concluded that the small amount of difference they detected between hominoids in their albumin reflected little molecular change and that this, in turn, implied that little time had elapsed since the separation of the hominoid lineages. Since, as they claimed, molecular change ticked away at a constant rate, IDs represented a "molecular clock" that could reveal the times at which the various hominoids – actually any species – diverged and went off on their own evolutionary paths.

After calibrating the molecular clock on the basis of paleontologists' interpretation, based on fossils, of when the Old World monkey and hominoid lineages might have separated, Sarich and Wilson calculated divergence dates of about 10 million years ago for the gibbon lineage, eight million years ago for the orangutan lineage, and a mere five million years ago for the human and African ape lineages. The major implication of this calculation was that fossils older than five million years, such as *Ramapithecus* (now referred to the genus *Sivapithecus*) from c. 12- to 14-million-year old deposits in Indo-Pakistan, could not, as paleontologists concluded from study of morphology, be hominid.

However, while molecular systematists were in general agreement on the premise of the molecular assumption – that degrees of similarity reflect degrees of evolutionary closeness – not all of them embraced the notion of a molecular clock that "ticked" at a constant rate. Among the most vocal dissenters was Morris Goodman, who in the early 1960s was the leading proponent of Nuttall's work. Although Goodman went beyond claiming that molecular similarity indicated a very close relationship between humans and the African apes to advocating that these three hominoids should also be placed in the same taxonomic family, Hominidae

(a position that for years caused many comparative primate anatomists and paleontologists to be skeptical of molecular systematics in general), he nonetheless tried to accommodate the paleontologists' identification of pre-five-million-year old fossils as being hominid.

In contrast to Sarich (who in 1971 was the first molecular systematist to reject morphology as being phylogenetically informative), Goodman accepted the paleontologically derived date of at least 14 million years as the time of divergence between hominids and their potential ape relatives. As a consequence, though, Goodman then had to interpret the immunological and biochemical data in a way that would make them compatible with a deep timescale of hominoid diversification. He did so by constructing a selectionist argument to explain the apparent acceleration or deceleration in rates of molecular change that *de facto* must have occurred, given the paleontologically established dates for the earliest appearances of each of the large-bodied hominoids. Thus, for instance, if the various hominoid lineages had indeed originated between 18 and 14 million years ago, the extreme similarity between the various extant hominoids in their albumin had to have resulted from a slowdown in the rate of molecular change of that particular blood serum protein.

Goodman proposed that the inferred deceleration in rate of molecular change in large-bodied hominoid albumin was related to the fact that, although they are similar to other anthropoid primates in developing hemochorial placentation (the most intimate of placental modes in the approximation of maternal and fetal blood systems), large-bodied hominoids differ from other anthropoids in having very long gestation periods. He argued that in animals with hemochorial placentation but fairly short gestation periods, there would be enough slack in the system to allow unimpeded molecular change to occur and accrue. The fetus would be born before the mother's immune system could produce antibodies to it. A large-bodied hominoid's gestation period, however, would be long enough not only for maternal-fetal immunological incompatibility to build up but also for maternally produced antibodies to diffuse through the placenta, with deleterious consequences to the fetus.

The dilemma which then had to be confronted was: How could an animal have both hemochorial placentation and a long gestation period? Goodman's answer was that natural selection must have acted to reduce the possibility of immune responses from the mother toward her fetus's proteins. Consequently, there had to have been a slowdown of molecular change that was commensurate with a prolongation of the gestation period. Although albumin had been the basis of Sarich and Wilson's constant-rate molecular clock, Goodman used the same molecule as the exemplar of how the molecular clock could run at different rates.

Concurrent with the various approaches – immunological reactivity as well as gel-column electrophoretic separation – that were brought to bear during the 1960s and 1970s on the determination of evolutionary relationships was an increased interest in protein sequences. The first major study was Zuckerkandl and Pauling's 1962 analysis of hemoglobin, and, for some years thereafter, hemoglobin was the most intensely studied molecule. Beginning in 1973, however, a research group headed by A. E. Romero-Herrera analyzed myoglobin sequences across a number of primarily mammalian taxa. As more

taxa were added to the study, the resultant, most "parsimonious" arrangement of the hominoids placed the gibbon rather than the orangutan closer to a human-African clade. Only by arguing for a more complicated scheme of myoglobin "evolution" in hominoids – including, for instance, the condition that there had to have been various "back-mutations" to a state similar to the presumed unchanged state – could one arrive at the more commonly accepted arrangement of divergences: gibbon first, then the orangutan, and then a human-African ape group.

But while protein sequencing was seen by some, especially Zuckerkandl as late as 1987, as the molecular level on which to focus for purposes of determining evolutionary relationships (primarily because amino acid sequences are less subject to the problems with DNA, such as insertions, deletions, and back-mutations), there was a growing consensus that the best systematic information lay at an even deeper "genetic" level. Since base differences at the third position of a codon (a nucleotide triplet) did not yield different resultant amino acids, the concern of DNA sequence advocates was that demonstration of similarity in protein sequences could produce "false" phylogenies because the underlying nucleotide sequences themselves may be different. Consequently, it became imperative to get to the level of DNA – nuclear DNA – itself, especially because the general belief or at least expectation (which was based on bacteria and inferred for metazoans) was the existence of a direct correlation between specific DNA sequences and specific genes. And, as genes were conceptualized, there was also supposed to be correspondence between one or perhaps a few genes and a specific feature or structure. In addition, because DNA sequences were composed of thousands upon thousands of bases, the apparent massive scale of the comparison had the appeal of providing an overwhelming amount of phylogenetically relevant information.

3
DNA Hybridization

Although DNA sequencing was possible by the late 1970s, it was an expensive and laborious procedure, which made the endeavor itself, much less the comparison of DNA sequences for phylogenetic purposes, prohibitive. There was, however, another way in which DNA sequence information could be achieved: DNA–DNA hybridization.

The theory behind using DNA–DNA hybridization (or just DNA hybridization) as an approximation of overall similarity in DNA sequences between different taxa lay in the fact that if the two strands that form the helical structure of nuclear DNA were dissociated (which could be accomplished by subjecting it to heat), their complementary bases would rebond and the two strands would reform their original helical organization. However, it is not only two cleaved but also originally helically arranged DNA strands from the same individual that will reassociate (reanneal). Analogous to the geneticist Theodosius Dobzhansky's discovery earlier in the twentieth century with chromosomes of different lengths from different species of fruit fly, any two strands of DNA, from different individuals of even different species, will attempt to anneal.

As in Dobzhansky's fruit-fly experiments, in which the larger chromosome would loop and fold so that whichever of its loci were also present in the shorter one

would match up, single strands of DNA (derived from heat splitting or melting of a double helix) from different organisms would attempt to recombine or hybridize at complementary base positions. The more complementarity there was between hybridized DNA strands, the greater was the intensity of heat needed to break down the bonds holding them together. Consequently, it seemed logical to conclude that the heat (ΔT) it took to disassociate a DNA–DNA hybrid (that is, the more thermally stable the hybrid was), the more similar the annealed strands were in their nucleotide sequences. From this apparent demonstration, one could then invoke the molecular assumption to explain how higher ΔT's, reflecting greater molecular similarity, were also a reflection of closer evolutionary relatedness. As logical, though, as this thought experiment might appear to be, it relies on an assessment of overall similarity, which does not identify the bases or sequence positions that underlie differences in ΔT's. Conversely, similar ΔT's in different taxa might not be the result of the same regions hybridizing.

One of the first applications of the technique of DNA hybridization to evolutionary questions was published in 1968 by R. J. Britten and D. E. Kohne, who were primarily concerned with learning more about the genome. They discovered that the genomes of higher organisms – but not of bacteria or viruses – contained "hundreds of thousands of copies of DNA sequences." However, not only did repeated DNA sequences represent a considerable portion of a genome, Britten and Kohne also found that they were "trivial and permanently inert." Only a small fraction of a genome was composed of unrepeated, unique, single-copy DNA sequences, which apparently constituted the active or functional elements of that genome.

Britten and Kohne also speculated on pathways or mechanisms that might produce genomic change, which they argued must be considered on two different levels: change in nucleotide sequence, and the origin of new families of nucleotide sequences. As Zuckerkandl and Pauling had assumed for differences in hemoglobin, Britten and Kohne proposed that changes in nucleotide sequences – which would be identified as point mutations – occur slowly over time. However, this model would explain only the "divergence of preexisting families" of nucleotide sequences, not the introduction of new families of sequences, as was assumed would eventually happen in the gradual-accumulation-of-change model that dominated molecular systematics. The introduction of new families of sequences, Britten and Kohne alternatively suggested, must "result from relatively sudden events," called *saltatory replications*. Accordingly, "saltatory replications of genes or gene fragments occurring at infrequent intervals during geologic history might have profound and perhaps delayed results on the course of evolution." Although Britten and Kohne's distinctions between types of molecular change were not at the time embraced by molecular systematists, their suggestion of "saltatory replication" would certainly appear to have been borne out with the subsequent identification of vertebrate regulatory genes involved in segmentation – *Hox-a, Hox-b, Hox-c*, and *Hox-d* – as replicates of the orthologous *Antennapedia* gene identified in fruit flies.

One of the early attempts at using DNA hybridization to reconstruct relationships among the primates was published in 1976 by Raoul Benveniste and George Todaro, whose primary focus was actually on the distribution among mammals of the type C viral gene, the presence of which, they

discovered, was particularly characteristic of animals found in Asia, including the orangutan and the gibbon, but which, unexpectedly, was also present in humans. After reviewing these data, Benveniste and Todaro turned to DNA hybridization for sorting out the relationships of various primates. Relying on previous studies, they estimated "the effect of mismatched base-pairs on thermal stability" as being "between 0.7 and 1.7 °C per 1% altered pairs." Since, depending on the animal, nuclear DNA consists of anywhere from 10 to 100 million nucleotide pairs, every 1% of difference in nucleotide sequence between DNA hybrids would result in a drop of about 1.0 °C of the heat necessary to melt the hybrid molecule.

Although Benveniste and Todaro calculated a 1.1% sequence difference between humans and both African apes and a 2.4% difference between humans and the orangutan, their published ranges of ΔT's actually demonstrate overlap between all large-bodied hominoids (although one does not know the identity of nucleotide difference and similarity). Taking into consideration the fact that presumed-to-be-homologous DNA sequences are not the same length (that is, they do not consist of the same number of nucleotides) in all animals, including the large-bodied hominoids (see below for human and orangutan), one also does not know from DNA hybridization experiments the details of how a short sequence from one species aligns with a longer sequence of another species.

Probably because DNA hybridization seemingly provided a way of getting closer to the supposedly most "informative" genetic level, questions about what, exactly, similarity in ΔT's meant were not explored. But other questions were.

In the 1980s, Charles Sibley and Jon Ahlquist reviewed earlier DNA hybridization experiments and concluded that a major problem with them all was that only a limited number of cross hybridizations had been tested. This, Sibley and Ahlquist claimed, was the reason previous studies had such difficulty in refining certain evolutionary relationships, particularly those between humans and the African apes. Although there was not particular morphological support for this theory of relationship, molecular systematists often grouped these three hominoids together. However, the level of resolution in the majority of the molecular work was never fine enough to determine which two of the three hominoids were the most similar (and thus presumably the most closely related).

Sibley and Ahlquist felt another source of error in these earlier studies resulted from using all of an animal's DNA, repeated as well as single-copy, to form hybrids. They argued that it was important to remove all repeated DNA – since it was redundant – and to work only with single-copy DNA.

Before Sibley and Ahlquist decided to tackle the evolutionary relationships of the large-bodied hominoids, they had spent years applying their DNA hybridization technique to the phylogeny of birds, through which they also developed the notion of a uniform average rate of genomic change (UAR). UARs, they believed, characterized the nature of molecular, and thus evolutionary, change in all taxa. The notion of a UAR also seemed to negate the need for debates over constant versus irregularly running molecular clocks, which were directed at individual molecules. As they saw it, the entire genome could change at a uniform average rate, even though the rates

at which individual molecules change may be quite different.

Sibley and Ahlquist argued that the power of DNA hybridization in "discovering" the phylogenetic relationships of organisms came from sampling the entire genome of an animal. They claimed that, since an organism's genome is composed of millions of nucleotides, the "law of large numbers" provides the checks and balances necessary to rule out "false" similarities or parallelisms. Accordingly, since DNA hybrid strands will only link up along those stretches of sequences that are complementary, there should be no question about homology. Conversely, sequences that did not line up were not homologous. At least that was the argument.

Sibley and Ahlquist's melting temperatures for their DNA hybrids produced numbers, which they then converted into phylogenetic distances using a procedure called *average linkage*. "Average linkage" begins "by clustering the closest pair or pairs of taxa," after which "one links the taxa which have the smallest average distance to any existing cluster," and on it goes until "all taxa are linked." The underlying assumption, which permits the linking of the most similar pairs, is that the DNA hybridization technique "measures the net divergence between the homologous nucleotide sequences of the species being compared."

Many aspects of Sibley and Ahlquist's phylogenetic reconstructions of birds were consistent with theories of relatedness derived from study of morphology. However, there were some significant differences, for instance, with regard not only to the broader phylogenetic relationships of, but also to the details of relatedness among, the flycatchers. By the 1980s, however, it was common practice, when molecular and the morphological phylogenies were in discord, to opt for the former. As Caccone and Powel argued in 1989, once the molecular assumption is accepted, overall similarity becomes the yardstick for determining closeness of relatedness. Indeed, whether it is a presumed albumin clock or a UAR, the assumption of ongoing molecular change validates the molecular assumption, which, in turn, demands that a molecular phylogeny is correct in its entirety, even if other sources of information contradict it.

Sibley and Ahlquist's "law of large numbers" and UAR appealed to most molecular systematists. This appeal was probably the primary reason paleontologists finally succumbed to the notion that molecular phylogenies had greater authority in deciphering evolutionary relationships when morphologically based theories were in conflict with them. But not all molecular systematists were convinced. Among them was Alan Templeton, who objected to the use of DNA hybridization because it did not allow one to determine the polarity of the similarity, which is necessary to address the question: "Is similarity due to distant or recent common ancestry?" For, only by studying the actual sequences of nucleotides can one determine the identity and from this attempt to infer the significance of similarity and dissimilarity. Sibley and Ahlquist countered, however, that there was "no reason to expect data derived from base sequences to improve on those from amino acid sequences, which have produced contradictory results." In their support, they also quoted A. E. Friday (who had collaborated with Romero–Herrera on myoglobin sequencing): "Phylogenetic conclusions derived from a study of nucleotide sequences will be subject to the same suspicions as those derived from amino acid sequences."

4
Mitochondrial DNA

As the field of molecular systematics was expanding its sphere of investigation during the 1970s, DNA located outside the nucleus, in a cell's mitochondria, was attracting attention. In a study published in 1979, W. Brown, M. George, and A. Wilson noted that there was more difference between humans and a sampling of Old World monkeys in their mitochondrial (mt) DNA than there was among these same primates in their nuclear DNA. By assuming that the divergence between the human and Old World monkey lineages occurred over 20 million years ago (based on morphological studies of the primate fossil record), Brown et al. concluded that the differences between these humans and Old World monkeys could be explained if nucleotide substitutions occurred at a slower pace –5 to 10 times slower – in nuclear than in mtDNA (in other words, if mtDNA "evolved" 5 to 10 times faster than nuclear DNA). On the basis of this interpretation, Brown et al. calculated that mtDNA data are the most accurate for studying evolutionary events that occurred "within the past 3–10 million years."

In subsequent publications, they proceeded to analyze the mtDNA of the large-bodied hominoids, which, largely because of Sarich and Wilson's molecular clock, were assumed to have diverged within this time period and which, because of this assumption, were thus amenable to such analysis. Although their data showed the fewest differences (interpreted as substitutions) between humans and chimps, Brown et al. concluded that humans were related to an African ape group. Given the molecular assumption, however, the data should have yielded a human–chimpanzee sister group, as Maryellen Ruvolo and a number of collaborators concluded in 1991 from their analysis of mtDNA.

In addition to the fact that mtDNA is single-, not double-stranded, the allure of using mtDNA was and continues to be twofold. One assumption is that, unlike nuclear DNA, mtDNA is supposedly inherited only through the maternal line and thus not subject to the complexities that occur through recombination of maternal and paternal DNA. The second assumption is the existence of a "hotspot" in the D-looped configuration of mtDNA and that this is the primary site of molecular activity. This region is called the *hypervariable zone* and its study has been interpreted as providing evidence of evolutionary change. Although crucial to the use of mtDNA for purposes of phylogenetic reconstruction, these assumptions are probably incorrect.

The introduction of paternal mtDNA into fertilized eggs has been reported in the literature from time to time. For example, in 1992, Allan Wilson and coworkers demonstrated that this happens in mice, as part of the sperm's tail, which does contain mtDNA, penetrates the ovum's membrane. More recently, John Maynard-Smith and coworkers calculated that there had to have been a recombination of maternal and paternal mtDNA during human evolution, and concluded that this possibility should cause systematists to reconsider the seemingly inviolable "fact" that humans and chimpanzees are closely related.

As Erika Hagelberg reviewed at great length, there is increasing indication that mtDNA is not as exempt from paternal inheritance and recombination as was initially believed. With regard to so-called hypervariable sites being regions of preferentially high mutation rates, which therefore lend themselves to phylogenetic analysis,

Hagelberg pointed out that "there is no direct evidence of hypervariability," although "most researchers believe that anomalous patterns of DNA substitution are best explained by mutation." Indeed, she writes, "because the notion of hypervariability fits with the received view of mtDNA clonality [maternal inheritance only], anomalies are seldom questioned." Hagelberg gives an interesting example. Depending on which subject's mtDNA is used, one can reconstruct "our most-recent female common ancestor. . .[as having] lived just 6000 years ago, a date more consistent with Biblical Eve than Mitochondrial Eve." As she cautions: "The picture is far from simple, and it is clear that extreme care must be taken in the interpretation of mtDNA phylogenetic trees in the face of possible recombination. . .[T]here are enough unexplained patterns in mtDNA data to warrant reassessment of the conclusions of many mtDNA studies."

5
DNA Sequences

In spite of Sibley and Ahlquist's and Friday's warnings about nucleotide sequence data not being any more reliable than protein sequence data for reconstructing phylogenetic relationships, the increasing ease with which DNA could be sequenced could not be ignored by molecular systematists. Nuclear as well as mtDNA were sequenced. Regardless of the genetic "level" under scrutiny, the interpretation of similarity or difference in DNA sequences under comparison was still predicated on the molecular assumption. The weight of DNA sequence data as being key to deciphering evolutionary relationships assumed special significance in some areas of molecular systematics because of the supposed information content of nuclear DNA in general. As Elizabeth Bruce and Francisco Ayala wrote as early as 1979 in an article on blood serum proteins: "Information macromolecules – that is, nucleic acids and proteins – document evolutionary history. . .[Thus] degrees of similarity in such macromolecules reflect, on the whole, degrees of phylogenetic propinquity."

Almost coincident with the rise in popularity of comparing DNA sequences for purposes of inferring evolutionary relationships came the cautionary notes. Of particular importance in this regard is the question of alignment, which, in 1991, J. A. Lake was among the first to address. As mentioned above, not all DNA sequences chosen for comparison, if homologous (e.g. representing the same "gene" or segment) are the same length. Therefore, decisions must be made with regard to how to subdivide the shorter sequence in order to align its nucleotides with those of the longer sequence. Typically, the alignment of compared sequences is presented in the literature without justification of the assumptions and decisions that produced the alignment, which, in turn, was then used as the basis for the phylogenetic analysis. But, as Lake warned in the title of his article, "[t]he order of sequence alignment can bias the selection of tree topology." In addition to assumptions that inform the decision to break up a short sequence so that its bases align with those of a longer sequence is the issue of whether, for one or another sequence, bases may have been added or inserted, removed or deleted, or one base substituted for another.

In *what it means to be 98% chimpanzee*, Jonathan Marks provides an example of these problems with DNA sequences from a human and from an orangutan to show three different ways in which

their bases (C,T,G,A) can be aligned and the interpretive consequences. Even beforehand, the assumption must be made that the 40 bases in the human sequence actually have homologous counterparts in the 54 bases in the orangutan sequence.

```
HUMAN CCTCCGCCGCGCCG      CTCCGC  GCCGCCGGGCA                CGGCC              CCGC
ORANG CC                GTCGCCTCCGCCACGCCGCGCCACCGGGCCGGGCCGGCCCGGCCCGCCCCGC

HUMAN          CCTCCGCCGCGCCGCT            CCGCGCCGCCGGGCACGGCCCCGC
ORANG CCGTCGCCTCCGCCACGCCGCGCCACCGGGCCGGGCCG GCCCGGCCCGCCCCGC

HUMAN CCTCCGCCGCGCCG      CT CCGCGCCGCCGGG  CAC    GGCC                          CCGC
ORANG                 CCGTCGCCTCCGCCACGCCGCGCCACCGGGCCGG GCCGGCCCGGCCCGCCCCGC
```

As Marks comments:

"Tabulate the differences. The top one invokes five gaps and six base substitutions; the middle has only two gaps but nine base substitutions. And the bottom one has five gaps and only three base substitutions. The three pairs of sequences differ in the assumptions about which base in one species corresponds to which base in the other. While we might, by Occam's Razor, choose the alignment that invokes the fewest inferred hypothetical evolutionary events, we still have to decide whether a gap "equals" a substitution. Does the bottom one win because it has a total of only eight differences? Or might the middle one win because a gap should be considered rare and thereby "worth," say, five base substitutions?

The problem is that we cannot know which is "right," and the one we choose will contain implicit information about what evolutionary events have occurred, which will in turn affect the amount of similarity we tally. How similar is this stretch of DNA between human and orangutan? There may be seven differences or there may be eleven differences, depending on how we decide the bases correspond to each other across the species – and that is, of course, assuming that a one-base gap is also equivalent to a five-base gap and to a base substitution.

In a more general sense, however, the problem of taking *quantitative* estimates of difference between entities that differ in *quality* is prevalent throughout the genetic comparison of human and ape. The comparison of DNA sequences presupposes that there are corresponding, homologous sequences in both species, which of course there must be if such a comparison is actually being undertaken. But other measurements have shown that a chimpanzee cell has 10% more DNA than a human cell. (this doesn't mean anything functionally, since most DNA is functionless.) But how do you work that information into the comparison, or into the 99.44% similarity [between human and chimp]?" [comment added].

These concerns have not, however, been widely appreciated by molecular systematists, especially molecular anthropologists, who not only portray the analysis of DNA sequences as neutral and objective but also use the assumption of relatedness to inform the way in which they analyze

the sequences they have aligned according to certain assumptions. Exemplary in this regard is the multiple DNA sequence analysis Maryellen Ruvolo and collaborators published in 1997, in which they sought to resolve the supposed dilemma of to which African ape humans are more closely related. In order to pursue this question, they assumed first that the orangutan was the sister taxon of a clade or evolutionary group consisting of humans, the chimpanzee, and the gorilla. Consequently, the differences in the orangutan had to be considered primitive relative to any similarities that were delineated between humans and one or the other of the African apes.

With the ever-growing popularity of the parsimony-based phylogenetic computer program PAUP (phylogenetic analysis using parsimony), it is common practice to "root" a phylogenetic analysis in a taxon that is chosen as the primitive outgroup – that is, the taxon that diverged earlier than the others – prior to the analysis taking place. Rooting parsimony or any of the other available clustering analyses (for example, nearest-neighbor joining or maximum likelihood, which are essentially similar to Sibley and Ahlquist's linking technique) in a particular taxon may be necessary for the algorithm to "work." Nevertheless, this procedure artificially determines character polarity since, by definition, the outgroup (the taxon in which the tree is rooted) is defined from the outset as being primitive in its entirety. In turn, the taxa to which the outgroup is the supposed primitive sister taxon are predetermined as being derived in whatever ways they differ from it.

The widespread use of this algorithm-based approach to analyzing nuclear and mtDNA as well as protein sequences presents its own set of problems and assumptions. Consider the molecular assumption: Since molecular change is supposedly continually occurring and being accumulated as a lineage proceeds through time, the degree of molecular similarity reflects the antiquity or recency of lineage divergence. Accordingly, each lineage accumulates it own unique array of molecular changes, which should make a lineage more distinctive (that is, different) the longer it is in existence. Although tautological, this assumption explains why more recently diverged taxa are more similar than more anciently divergent lineages. On the other hand, in order to root an algorithm for purposes of generating presumed phylogenetic relationships, one must assume that the taxon chosen as the earlier-divergent outgroup is totally primitive relative to the taxa to which it is supposed to be the sister taxon. Yet, it is the molecular assumption that validates the use of overall similarity as the key to resolving phylogenetic relationships by contrasting it with the unique differences that earlier-divergent lineages accumulated along their own, unique evolutionary trajectories. Clearly, both assumptions cannot be correct at the same time. Either the earlier divergent-taxa or lineages did not change, but remained primitive (which is the logical extension of identifying a taxon as the outgroup in which to root a computer analysis), or they did change by accumulating their own suites of molecular difference (the basis of the molecular assumption), in which case they are at least in some aspects derived (and uniquely so, for that matter, because of their unique molecular histories) and not primitive relative to the taxa to which they are being compared.

In the realm of morphological systematics, according to Hennigian or cladistic

principles, overall similarity is not *de facto* a clue to evolutionary relatedness. Similarity must be sorted out into features that reflect a hierarchy of inheritance: primitive features from ancient ancestors, and derived features from recent ancestors.

Since the pattern of life is one of a hierarchy of nested sets of smaller and smaller clades (groups of related taxa), that which is considered primitive versus that which is considered derived depends on the level in the hierarchy of nested clades one is investigating. Primitive features – features retained in descendants – do not elucidate the relationships of these taxa. Only derived features can. It is also important conceptually to recognize that a derived feature at one level of the hierarchy is a primitive retention at another. There is no theoretical reason why this approach to systematics cannot be applied to molecular data. The major difficulty is that molecular systematists would have to sample and compare a wide range of taxa. This is the only way in which relative primitiveness and derivedness can be determined. It cannot be justified by *a priori* assumptions of directionality, as underlies the molecular assumption, or by choosing an outgroup on the basis of its presumed evolutionary relationship to other taxa. However, even from the beginning, it is also crucial to realize that shared similarity does not translate directly into a demonstration of relatedness. Taxa may be similar, not because they inherited changes that distinguished a recent common ancestor, but because they share primitive retentions, that is, features that have not changed in a succession of ancestors.

Nevertheless, it is becoming increasingly popular in the literature for molecular studies on the relatedness of taxa to be identified as being "cladistic." One argument is that nucleotide bases – C, G, T, A – represent alternative character states. On one level this may appear logical, but it is actually misleading since none of them represents a character. Phylogenetically relevant alternative molecular character states would be better represented by comparison, for example, of arrangements of "gene" sequences with regard to *cis* and *trans* elements, patterns of introns and exons and of methylation of transposons or other elements, and pathways of molecular communication. Another argument in support of molecular studies being cladistic derives from the claim that molecular similarity is equivalent to synapomorphy; that is, shared similarity represents shared derived character states. This conclusion is, of course, only a restatement of the molecular assumption: the most recently diverged taxa share more recently accumulated (equate with derived) molecular states. Thus, the supposedly shared derived molecular states are delineated *a posteriori*; in other words, after the algorithm of choice has clustered taxa on the basis of their greater or lesser degrees of similarity (depending on the algorithm), typically after rooting the tree in a particular taxon (which, as pointed out above, at once defines it as being primitive and the taxa being compared to it as derived in their shared similarities). This, however, is not how a cladistic analysis proceeds. The endeavor of hypothesizing primitive versus derived character states occurs prior to hypothesizing relationships – which is the only way in which such a methodology can actually be employed.

The assumption of continual molecular change – whether through point mutations affecting nuclear or mtDNA, or altering amino acids in protein sequences – is also of interest. Recall that this idea was initially framed by Zuckerkandl and Pauling

as a way of explaining their data: "overall similarity *must* [emphasis added] be an expression of evolutionary history," with descendants "mutating away" from each other, becoming "gradually more different from each other." It is this assumption that proposes that earlier-divergent taxa accumulate their own molecular differences, while the most recently divergent taxa are similar because of the longer shared history of accumulated molecular changes and shorter time of independent molecular change. The existence of molecular clocks and UAR is predicated on this notion. Nevertheless, it must be recognized that this is an extrapolation – an explanation of how something might come to be. It has not been demonstrated.

The contradiction is that while constant molecular change is predicted through the molecular assumption as occurring during gametogenesis, or in some way as to be passed on to offspring, in molecular biology, it is well known that the only source of constant molecular change is ultraviolet radiation, which produces a mutation rate of 10^{-8}–10^{-9}. But the other element of UV-derived mutation is that it is random, with the potential of affecting either somatic or sex cells and also with regard to the molecule that is affected. Thus, while there might appear to be concordance between the reality of the physical world in which there is a relatively constant UV-provoked mutation and the concept of a constantly "evolving" molecular world, this is an illusion.

The notion of constant and accumulative mutation affecting sex cells is of further interest because it also contradicts the basic tendency of cells to remain in homeostasis. As seen, for example, in the roles of heat shock proteins (HSP) – maintaining cell membrane physical states through lipid transport, eliminating reading errors that occur during transcription or translation, DNA repair, chaperoning other proteins, and ensuring proper folding of proteins as they emerge from the ribosomes – the basic propensity of a cell is to resist change. Intuitively, this should make sense inasmuch as unabated molecular change would undermine the integrity of cell function, as would also be the case with a constant accumulation of point mutations, and more probably lead to the death of organisms than to change.

6
"Evo-devo"

In 1975, Mary-Claire King and Allan Wilson surveyed all available data on blood serum proteins, as well as the results of DNA hybridization, with regard to humans and chimpanzees. Although their publication is cited as having demonstrated the relatedness of these two hominoids, this was not their intention. As they stated: "the only two species which have been compared by all of these methods are chimpanzees. . .and humans," and thus "a good opportunity is. . .presented for finding out whether the molecular and organismal estimates of distance agree." The result was that humans and chimpanzees differed in their genetic makeup only by about one percent. King and Wilson concluded that "all the biochemical methods agree in showing that the genetic distance between humans and the chimpanzee is probably too small to account for their substantial organismal differences." In order to explain how humans and chimpanzees could be virtually identical in their genes but markedly different animals, King and Wilson suggested that humans and chimpanzees must be

different in those genes that regulate development.

Since then, studies on the regulation of development have expanded exponentially, not only with regard to distinguishing between regulatory and structural genes, but especially with regard to the array of molecules that induce gene transcription and communicate via signal transduction pathways to produce structure. Interestingly, those animals that have been studied in depth – such as the fruit fly, zebra fish, frog, chick, mouse, human – demonstrate a commonality of "homeotic genes" (which contribute to segmental patterning and segment identify). In turn, through their protein products (transcription factors), homeotic genes control or at least affect gene expression. Time and time again during development, the same proteins (e.g. various growth factors, trans-inducing and bone-modifying proteins) and regulatory genes (and their products) are coopted to produce what in adult organisms are different morphologies.

In 1994, Lewis Wolpert summarized the situation: "During development, differences are generated between cells in the embryo that then lead to spatial organization (pattern formation), changes in form, and the generation of different cell types. Genes control development by controlling cell behavior." But one should not be too gene-centric in envisioning the emergence of form from genes and gene products alone. For, while there might be genetic regulation of some cells' activities, the results (e.g. cellular asymmetry, cell membrane elasticity or rigidity, compressive forces) might produce physical or mechanical responses, which may not themselves be genetically based, but which, nonetheless, greatly affect cell geometry and ultimately organismal form. "Cell behavior," to return to Wolpert, has many different levels of meaning.

In the 1970s, Søren Løvtrup argued that one must recognize the importance of epigenesis in development: especially that changes in properties of the fertilized egg can alter the chronology and spatial organization of patterns of cellular diversification. Since the larger clades of multicellular organisms possess the same kinds of cells, as well as the same chemical substances that form the immediate environment of the cells, variation in the spatial and chronological organization of cellular differentiation must be at least one of the keys to the emergence of evolutionary novelty. For instance, whether a cell divides symmetrically or asymmetrically (which can be affected even by the positions of the chromosomes relative to the center or periphery of the cell) can greatly impact the spatial relationships of cells and, therefore, eventually have an effect on organismal shape. As Pere Alberch has emphasized, the development of organismal form and structure is also a function of the physical and mechanical properties of cells' sizes, shapes, and spatial relationships.

The application to evolutionary questions of discoveries in the regulation of development has given rise to the field of "evo-devo" (evolution and development), in which the interrelationship between the "genetics" and the "epigenetics" of development has become a primary focus. Indeed, Løvtrup's concern with the influence on metazoan development of differences in cellular differentiation during gastrulation appears to be even more germane to an understanding of the emergence of form and its conservation across taxa as well as of the emergence of differences in form, whether their expression constitutes variation (individual differences) or diversity (species differences).

7
Positional Information and Shape

One of the ongoing questions in developmental biology is how cells acquire positional information, not only in terms of entire structures themselves (e.g. where limbs will grow) but also with regard to how cells acquire information to contribute to regional shapes of a structure (e.g. the different segments of a limb).

In invertebrates, wing (e.g. as in a fruit fly) and limb (e.g. as in the brine shrimp-like crustacean, *Artemia*) positioning involves activation of the regulatory genes *nubbin* and *apterous*. In vertebrates, various regulatory genes, especially *Hox* genes, and also *dlx* (distal-less), are known to be involved in segmentation and limb positioning. In fish and tetrapods, the *Hoxd11-13* genes are expressed along the posterior margin of the enlarging limb bud; however, in tetrapods, this homeodomain expands anteriorly across the distal (lower) end of the limb bud. Additionally, in vertebrates, *Hox* genes are not only recruited in the formation of a segmented trunk but also through regional activation, in the segmentation of the hindbrain. Regional activation of *sonic hedgehog*, however, in part, underlies forebrain segmentation.

Eye development is also of interest in this regard. For although what used to be thought of as "master-control genes," such as *Pax-6*, were found to participate in signal transduction pathways leading to eye formation in invertebrates (typically multilensed, rigid) as well as in vertebrates (single lensed, deformable), there is at least one element that vertebrates have in common: Even though there are differences among vertebrates so far studied with regard to when in ontogeny and how often and in how many different regions the *Rx* gene (which is also recruited in fruit-fly eye development) is activated, it is always expressed in the vertebrate forebrain. Thus, at one level, one can hypothesize that the last common ancestor of vertebrates was characterized by early activation of the *Rx* gene in the presumptive forebrain and that differences between taxa are the result of differences in other aspects of *Rx* gene expression: for example, the different proteins in the lenses of amphibians and mammals may be due in part to down- or upstream effects of the *Rx* gene being expressed later on in development in the amphibian (frogs) retina, whereas in mammals (mice), the *Rx* gene is expressed early on in the presumptive eye itself.

In addition to considering morphological differences in light of differences in regional (as well as in overlapping) domains of regulatory gene expression, it is becoming increasingly clear that differences in fields of molecular gradients (morphogenetic fields) also play a role. As C. Owen Lovejoy, M. J. Cohn, and T. D. White hypothesized in 1999 in their discussion of the evolution of human pelvic form, "if a particular PI [positional information] gradient were to span *n* cell diameters, and those cells defined the ultimate anteroposterior dimension of the presumptive ilium (superoinferior in the adult human), then a slight increase in the steepness of its slope would cause that signal to span fewer cells, 'distorting' the presumptive anlagen and substantially altering downstream adult morphology." In other words, although it would seem to be a process involving myriad steps, "the transformation of the common ancestral pelvis [in its entirety] into that of early hominids may have been as 'simple' as a slight modification of a gradient" [comment added]. Thus, in addition to differences in gene expression and pathways of molecular communication,

as well as to the physical and mechanical consequences of cellular organization, morphological novelty in metazoans (and presumably plants as well) may also be affected by altering the domains of morphogenetic fields.

But what is the source of differences in gene or molecular gradient expression? Lovejoy et al. suggest that one need not seek the answer in mutation, which is a position that Sean Carroll has recently also strongly argued.

8
"Mutation"

The concept of "mutation" is about as slippery as that of a "gene." It means different things to different researchers, and, interestingly, the differing concepts seem to "work" in their disparate intellectual contexts. With regard to mutation, the "textbook" notions of preceding decades included point mutation, gene duplication, and chromosomal rearrangement. The latter was basic to the earlier experimental studies and theoretical considerations of the fruit-fly geneticist, Theodosius Dobzhansky. Dobzhansky's emphasis on chromosomal rearrangement as a potential source of evolutionary novelty was subsequently adopted by the developmental biologist, Richard Goldschmidt, in his theory of systemic mutation, which he argued would lead to the abrupt appearance of novel form. Unfortunately, Goldschmidt is best remembered, and consequently criticized, for suggesting that such novelties would emerge in individuals he identified as "hopeful monsters."

One of Goldschmidt's major theoretical thrusts, however, was distinguishing between what he identified as micromutation (leading to variation and microevolution) and macromutation (leading to the origin of species or evolution). The small mutations that fruit-fly population geneticists, such as Thomas Hunt Morgan, inferred lay behind small phenotypic changes, Goldschmidt identified as micromutations, which, he argued, led only to the survival of species, not to their origin. The latter required a larger source of genetic disturbance, and for that he turned to chromosomal rearrangement. The logic is understandable: If the chromosome theory was correct (that, indeed, units of heredity or genes were contained on chromosomes – as Morgan presented it, like beads in a necklace), then manipulating them on a grand scale (producing a systemic mutation) might yield evolutionarily significant novelty, that is, new species. A major problem with Goldschmidt's theory, however, was that he did not provide a mechanism by which more than one individual would be the bearer of the novelty and, thus, of the systemic mutation underlying it.

Point mutations, commonly the result of UV radiation, are random with regard to affecting somatic or sex cells. In addition, if they do not interfere with cell function, such point mutations are not a reliable source of potential genetic and subsequently morphological novelty. Indeed, it appears that point mutations do not often cause any noticeable effect, either genetically or phenotypically. Gene duplication – as seen, for instance, in the emergence of *Hoxa-d* – does sometimes occur, but knockout experiments have demonstrated that duplication typically reflects redundancy, not a source of phenotypic novelty.

It may be true that manipulation of levels of thyroxin or retinoic acid during ontogeny can affect the size of an organism, or the shapes of some of its features, just

as a mother's diet can affect methylation in its fetus or fetuses, and thus aspects of her progeny's postnatal growth. However, these disturbances only affect an individual during its lifetime, and should not be expected to be repeated across generations and under the influence of fluctuating environmental stimuli (e.g. diet, temperature, amount of daylight). The problems, then, that still must be addressed are as follows: How does a genetic or cellular change remain "fixed" or constant, and how do many individuals come to bear it?

It is important to realize that there is a difference between change at the genetic level and what is perceived as phenotypic change. Common in the literature on the genetics of evolution is the mistake of conflating the two as constituting macromutation – a misconception that derives from the confusion Dobzhansky introduced with regard to the terms micromutation and macromutation when, in 1941, he sought to discredit Goldschmidt. Nevertheless, especially with the increasing awareness from molecular biology that there are not "genes for" features, we must be vigilant in making a distinction between what appears morphologically to have been the result of a macromutation (e.g. developing feathers instead of scales) and the underlying genetic–epigenetic interactive pathway.

9
Toward a Theory of Evolutionary Change

The question at hand, then, is the articulation of a mechanism that can first provide the potential for genetic novelty. Building on my original theory for the sudden appearance of morphological novelty (through the silent spread of recessive "mutations"), the molecular biologist, Bruno Maresca, and I have proposed that the opportunity for genetic novelty may lie in overstressing cells to the extent that their HSPs can no longer maintain genetic homeostasis; that is, they cannot fulfill their roles as chaperones, respond to the needs of the cell membrane, and, perhaps most importantly, for this discussion, properly fold other proteins and assist in DNA repair. Although first identified in heat shock experiments, HSPs can be affected by a variety of stresses, including diet (saturated vs unsaturated fatty acids), wind, aridity, and cold. Since most multicellular plants and animals possess HSPs, the theory is more widely applicable than metazoan-centric Darwinian and neo-Darwinians models of evolutionary change, the latter of which relies on unwarranted extrapolations from fruit-fly population genetics.

Since most multicellular organisms have a window of heat shock response, they can "adapt" to normal fluctuations in their environmental circumstances. If environmental change exceeds this window (as when seasons change), most organisms can "reset" it, often in less than two months. Until this window is reset, the stress induces an increase in HSP production. If, however, there is a spike in environmental stress that exceeds an organism's ability to reset its HSP response, HSP function will fail, and opportunities for introducing genetic novelty will emerge – especially as a result of improper protein folding and inefficient DNA repair. In the former situation, improperly folded proteins may, for example, no longer recognize (or be recognized by) promoter or enhancer regions to which they would normally bind, but they may now be capable of binding to different sites. An obvious result could be the activation or deactivation of a

"gene" or "genes," and, thus, the creation of one or more new developmentally significant signal transduction pathways. With regard to inefficient DNA repair, genetic novelty of a different sort can be introduced, with obvious potential consequences. In both cases, however, the fact that the environmental stress will be at least regional (if not global), the circumstances exist for more, perhaps many more, than one individual of a species to be affected (not, however, necessarily in the same way).

However, while it might be tempting to extrapolate immediately from these possible sources of genetic novelty to the emergence of evolutionarily relevant morphological novelty, one must be cautious. First, the effects of extreme environmental spikes on HSPs must be actualized during gametogenesis. If they are not, offspring will not be affected. Second, if the effects do not kill the individuals that inherit any of these genetically based novelties, they will probably not be expressed; that is, these genetic novelties will be in the "recessive" state. Consequently, there will not be an immediate phenotypic reflection of these genetic changes. In the recessive state, however, they can spread "silently" throughout the population, until it becomes sufficiently saturated with heterozygotes that homozygotes for the genetic novelty will be produced. If the resultant phenotypic expression – cellular or greater – does not kill its bearers, they may continue to reproduce themselves, as heterozygotes will also contribute to the numbers of individuals bearing the phenotypic novelty. Thus, the spread of a genetic basis for potential phenotypic novelty may take numerous generations before there is any statistical possibility of the phenotype being expressed. In other words, there will be a temporal disjunction between the disruption of cellular and genetic homeostasis, and what will be seen as the abrupt or sudden appearance of phenotypic novelty, and in some number of individuals. In addition, one must bear in mind that, during periods of "silently spreading" genetic novelty, there could be other environmental spikes that would contribute to the pool of potential for genetic novelty and also then phenotypic novelty (however defined). Superficially, this process – or at least the sudden appearance of phenotypic novelty – may seem macromutational, but, clearly, it is not, at least in the original sense of Goldschmidt or even that of Dobzhansky. Indeed, something as simple as slight changes in protein folding could have major cascading morphological effects.

10
Molecules and Systematics: Looking Toward the Future

It may be widely believed, and even true at some level, that, as Sean Carroll has recently reiterated, "genomes diverge as a function of time." However, the observation that genomes may be different (in whatever ways difference, and similarity, may be identified and defined) does not in and of itself provide clues to how this difference was achieved. No doubt, some difference is due to the rare and random effects of UV radiation. In addition, genomic difference may be due to failures in DNA repair. There may be something intuitively appealing about Sibley and Ahlquist's the "law of large numbers" – the idea that organisms are closely related because they share "lots" of their genome. However, as Jonathan Marks points out, humans share

about 25% of their genome with bananas. Essentially, there is nothing in an observation of genomic difference or similarity that directly translates into the "molecular assumption" and, consequently, a theory of evolutionary relatedness.

Can, then, molecular information be useful in systematics and phylogenetic reconstruction?

The answer is yes, but it will have to be at the level of cell biology and pathways of molecular communication. As King and Wilson came close to predicting many decades ago, it is not through the study of molecular or genomic similarity of organisms that we will come to understand their biology, but through the investigation of those elements that underlie the development of their biology. This makes sense. For, if something as simple as the inactivation or deletion of a transcriptional enhancer can result in a more caudal repositioning of the sacrum, or if the expansion of a morphogenetic gradient can transform in its entirety a narrow pelvic girdle with tall, thin ilial blades into a broad, deep, and squat structure, then it is by seeking to identify the similarity or difference in these molecular events that we may more profitably explore the molecular basis of morphology and, consequently, the evolutionary relationships of complex organisms.

This is, perhaps, a timely occasion to both question and expand our perceptions of what is or will be evolutionarily revealing at the molecular level. There has been a steady increase in the number of studies that demonstrate virtual molecular identity between taxa that are morphologically very different and then express amazement at this apparent contradiction. As Sean Carroll pointed out with regard to the importance placed on the human and chimpanzee-genome projects – especially since so much money has been poured into them in the hope that forthcoming comparisons will instantaneously provide answers to any questions – demonstrating molecular similarity does not translate into deciphering the pathways that make these organisms so different in hard- and soft-tissue anatomy, physiology, reproductive biology, cognitive abilities, and behavior. Here, the "law of large numbers" fails to be enlightening. For, in contrast to the bacterial world, in the metazoan world, a one-to-one correspondence between a "gene" (a sequence of nucleotides bound by start and stop codons) and a "gene product" (a protein or amino acid sequence) is not there. In multicellular animals, RNA essentially directs the "show," for example, in reading select bases and splicing specific introns, as it composes different proteins from the same stretches of DNA. The surprise "The International Chimpanzee Chromosome 22 Consortium" had at finding upon comparing human chromosome 21 with its apparent orthologue in the chimpanzee, chromosome 22 – not only that these hominoids differ by 83% in their amino acid sequences but also that this large difference is produced from very similar DNA sequences – should serve as a lesson: While there may be appeal to the "law of large numbers" that comparison of chromosomes and especially of entire genomes purportedly represents, in the end, this molecular level may not be the evolutionarily informative hotspot everyone has been seeking.

See also Gene Mapping and Chromosome Evolution by Fluorescence–Activated Chromosome Sorting; Genetic Intelligence, Evolution of; Genetic Variation and Molecular Evolution; Immunoassays.

Bibliography

Books and Reviews

Frontiers in biology: development, *Science* (1994) **266**, 561–614.

Gerhart, J.C., Kirschner, M.W. (1997) *Cells, Embryos, and Evolution: Toward a Cellular and Developmental Understanding of Phenotypic Variation and Evolutionary Adaptability*, Blackwell Science, New York.

Marks, J. (2003) *What it Means to be 98% Chimpanzee*, (revised paperback edition). University of California Press, Berkeley, CA.

Müller, G.B., Newman, S.A. (Eds.) (2000) *Origination of Organismal Form*, MIT Press, Cambridge, MA.

Raff, R.A. (1996) *The Shape of Life: Genes, Development, and the Evolution of Animal Form*, University of Chicago Press, Chicago, IL.

Schwartz, J.H. (1987) *The Red Ape: Orangutans and Humans Origins*, Basic Books, New York; in press.

Schwartz, J.H. (1999) *Sudden Origins: Fossils, Genes, and the Emergence of Species*, John Wiley, New York.

Primary Sources

Averoff, M., Cohen, S.M. (1997) Evolutionary origin of insect wings from ancestral gills, *Nature* **385**, 627–630.

Awadella, P., Eyre-Walker, A., Maynard Smith, J. (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA, *Science* **286**, 2524–2525.

Baker, R.H., Xiaobo, Y., DeSalle, R. (1998) Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees, *Mol. Phylogenet. Evol.* **9**, 427–436.

Benveniste, R.E., Todaro, G.J. (1976) Evolution of type C viral genes: evidence for an Asian origin of man, *Nature* **261**, 101–8.

Britten, R. (2002) Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13633–13635.

Britten, R.J., Kohne, D.E. (1968) Repeated sequences in DNA, *Science* **161**, 529–40.

Brown, W.M., George, M. Jr., Wilson, A.C. (1979) Rapid evolution of mitochondrial DNA, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1967–71.

Brown, W.M., Prager, E.M., Wang, A., Wilson, A.C. (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution, *J. Mol. Evol.* **18**, 225–39.

Bruce, E.J., Ayala, F.J. (1979) Phylogenetic relationships between man and the apes: electrophoretic evidence, *Evolution* **33**, 1040–1056.

Caccone, A., Powell, J.R. (1989) DNA divergence among hominoids, *Evolution* **43**, 925–942.

Carroll, S.R. (2003) Genetics and the making of *Homo sapiens*, *Nature* **422**, 849–857.

Cohn, M.J., Patel, K., Krumlauf, R., Wilkinson, D.G., Clarke, J.D.W., Tickle, C. (1997) Hox9 genes and vertebrate limb specification, *Nature* **387**, 97–101.

Crockford, S.J. (2003) Thyroid rhythm phenotypes and hominid evolution: a new paradigm implicates pulsatile hormone selection in speciation and adaptation changes, *Comp. Biochem. Physiol., A* **135**, 105–129.

Cronin, J.E., Sarich, V.M. (1980) Tupaiid and Archonta phylogeny: the Macromolecular Evidence, in: Luckett, W.P. (Ed.) *Comparative Biology and Evolutionary Relationships of Tree Shrews*, Plenum Press, New York, pp. 293–312.

Czelusniak, J., Goodman, M., Moncrief, N.D., Kehoe, S.M. (1990) Maximum parsimony approach to construction of evolutionary trees from aligned homologous sequences, *Methods Enzymol.* **183**, 601–615.

Dobzhansky, T. (1935) *Drosophila miranda*, a new species, *Genetics* **20**, 377–391.

Dobzhansky, T. (1941) *Genetics and the Origin of Species*, 2nd edition, Columbia University Press, New York.

Felsenstein, J. (1988) Phylogenies from molecular sequences: inference and reliability, *Annu. Rev. Genet.* **22**, 521–565.

Ferris, S.D., Wilson, A.C., Brown, W.M. (1981) Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2431–2436.

Figdor, M., Stern, C. (1993) Segmental organization of embryonic diencephalon, *Nature* **363**, 630–634.

Goldschmidt, R. (1940) *The Material Basis for Evolution*, Yale University Press, New Haven, CT, (reprinted 1982).

Goodman, M. (1962) Immunochemistry of the primates and primate evolution, *Ann. N.Y. Acad. Sci.* **102**, 219–234.

Goodman, M., Tashian, R.E. (Eds.) (1976) *Molecular Anthropology*, Plenum Press, New York.

Goodman, M., Braunitzer, G., Stangl, A., Schrank, B. (1983) Evidence on human origins from haemoglobins of African apes, *Nature* **303**, 546–48.

Goodman, M., Olson, C.B., Beeber, J.E., Czelusniak, J. (1982) New perspectives in the molecular biological analysis of mammalian phylogeny, *Acta Zoolo. Fennica* **169**, 1–73.

Graur, D., Martin, W. (2004) Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision, *Trends Genet.* **20**, 80–86.

Gregory, W.K. (1922) *The Origin and Evolution of the Human Dentition*, Williams and Wilkins, Baltimore, MD.

Hagelberg, E. (2003) Recombination or mutation rate heterogeneity? Implications for mitochondrial Eve, *Trends Genet.* **19**, 84–90.

Hasegawa, M., Yano, T. (1984) Maximum likelihood method of phylogenetic inference from DNA sequence data, *Bull. Biometri. Soc. Jpn.* **5**, 1–7.

Hasegawa, M., Kishino, H., Yano, T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA, *J. Mol. Evol.* **22**, 160–174.

Hedges, S.B., Kumar, S. (2002) Vertebrate genomes compared, *Science* **297**, 1283–1285.

Hedges, S.B., Kumar, S. (2003) Genomic clocks and evolutionary timescales, *Trends Genet.* **19**, 200–206.

Hedges, S.B., Kumar, S., Tamura, K., Stoneking, M. (1992) Human origins and analysis of mitochondrial DNA sequences, *Science* **255**, 737–739.

Horai, S., Hayasaka, K., Kondo, R., Tsugane, K., Takahata, N. (1995) Recent African origin of modern humans revealed by complete sequence of hominid mitochondrial DNAs, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 532–536.

Hunt, P., Gulisano, M., Cook, M., Sham, M.H., Faiella, A., Wilkinson, D., Boncinelli, E., Krumlauf, R. (1991) A distinct Hox code for the branchial region of the vertebrae head, *Nature* **353**, 861–864.

King, M.-C., Wilson, A.C. (1975) Evolution at two levels in humans and chimpanzees, *Science* **188**, 107–88.

Lake, J.A. (1991) The order of sequence alignment can bias the selection of tree topology, *Mol. Biol. Evol.* **8**, 378–385.

Lovejoy, C.O., Cohn, M.J., White, T.D. (1999) Morphological analysis of the mammalian postcranium: a developmental perspective, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13247–13252.

Lovejoy, C.O., McCollum, M.A., Reno, P.I., Rosenman, B.A. (2003) Developmental biology and human evolution, *Annu. Rev. Anthropol.* **32**, 85–109.

Løvtrup, S. (1974) *Epigenetics: A Treatise on Theoretical Biology*, Wiley, New York.

Lowenstein, J.M., Molleson, T., Washburn, S.L. (1982) Piltdown jaw confirmed as orang, *Nature* **299**, 294.

Lowenstein, J.M., Sarich, V.M., Richardson, B.J. (1981) Albumin systematics of the extinct mammoth and Tasmanian wolf, *Nature* **291**, 409–411.

Lumsden, A., Krumlauf, R. (1996) Patterning the vertebrate neuraxis, *Science* **274**, 1109–1115.

Maresca, B., Schwartz, J.H. Environmental change and stress protein concentration as a source of morphological novelty: sudden origins, a general theory on a mechanism of evolution, in manuscript.

Marshall, C.R. (1991) Statistical tests and bootstrapping: assessing the reliability of phylogenetics based on distance data, *Mol. Biol. Evol.* **8**, 386–391.

Mathers, P., Grinberg, A., Mahon, K., Jamrich, M. (1997) The Rx homeobox gene is essential for vertebrate eye development, *Nature* **387**, 604–607.

Morgan, T.H. (1916) *A Critique of the Theory of Evolution*, Princeton University press, Princeton, NJ.

Muragaki, Y., Mundlos, S., Upton, J., Olsen, B.R. (1996) Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13, *Science* **272**, 548–551.

Nuttall, G.H.F. (1904) *Blood Immunity and Blood Relationship*, Cambridge University Press, Cambridge, MA.

O'hUigin, C.O., Satta, Y., Takahata, N., Klein, J. (2002) Contribution of homoplasy and of ancestral polymorphism to the evolution of genes in anthropoid primates, *Mol. Biol. Evol.* **19**, 1501–1513.

Osborn, J.F. (1978) Morphogenetic Gradients: Fields Versus Clones, in: Butler, P.M., Joysey, K.A. (Eds.) *Development, Function, and*

Evolution of Teeth, Academic Press, New York, pp. 171–201.

Oster, G., Alberch, P. (1982) Evolution and bifurcation of developmental programs, *Evolution* **36**, 444–459.

Pennisi, E. (1999) Genetic study shakes up out of Africa theory, *Science* **283**, 1828.

Romero-Herrera, A.E., Lehmann, H., Castillo, O., Joysey, K.A., Friday, A.E. (1976) Myoglobin of the orangutan as a phylogenetic enigma, *Nature* **261**, 162–64.

Romero-Herrera, A.E., Lehmann, H., Joysey, K.A., Friday, A.E. (1978) On the evolution of myoglobin, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **283**, 61–183.

Rubenstein, J.L.R., Martinex, S., Shimamura, K., Puelles, L. (1994) The embryonic vertebrate forebrain: the prosomeric model, *Science* **266**, 578–580.

Ruvolo, M. (1997) Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets, *Mol. Biol. Evol.* **14**, 248–265.

Ruvolo, M., Disotell, T.R., Allard, M.W., Brown, W.M., Honeycutt, R.L. (1991) Resolution of the African hominoid trichotomy by use of a mitochondrial gene sequence, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1570–1574.

Samollow, P.B., Cherry, L.M., Whitte, S.M., Rogers, J. (1996) Interspecific variation at the Y-linked *RPS4Y* locus in hominoids: implications for phylogeny, *Am. J. Phys. Anthropol.* **101**, 333–343.

Sarich, V.M. (1971) A Molecular Approach to the Question of Human Origins, in: Dolhinow, P., Sarich, V.M. (Eds.) *Background for Man*, Little, Brown, Boston, MA, pp. 60–81.

Sarich, V.M., Wilson, A.C. (1966) Quantitative immunochemistry and the evolution of primate albumins: micro-complement fixation, *Science* **154**, 1563–66.

Sarich, V.M., Wilson, A.C. (1967a) Rates of albumin evolution in primates, *Proc. Natl. Acad. Sci. U.S.A.* **58**, 142–148.

Sarich, V.M., Wilson, A.C. (1967b) Immunological time scale for hominid evolution, *Science* **158**, 1200–1203.

Schwartz, J.H. (2001) A review of the systematics and taxonomy of Hominoidea: history, morphology, molecules, and fossils, *Ludus Vitalis* **IX**, 15–45.

Shubin, N., Alberch, P. (1986) A morphogenetic approach to the origin and basic organization of the tetrapod limb, *Evol. Biol.* **20**, 319–387.

Shubin, N., Tabin, C., Carroll, S. (1997) Fossils, genes and the evolution of animal limbs, *Nature* **388**, 639–648.

Sibley, C.G., Ahlquist, J.E. (1983) Phylogeny and Classification of Birds Based on the Data of DNA-DNA Hybridization, in: Johnston, R.F. (Ed.) *Current Ornithology*, Vol. 1, Plenum Press, New York, pp. 245–92.

Sibley, C.G., Ahlquist, J.E. (1984) The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization, *J. Mol. Evol.* **20**, 2–15.

Sordino, P., van der Hoeven, F., Duboule, D. (1995) Hox gene expression in teleost fins and the origin of vertebrate digits, *Nature* **375**, 678–681.

Stauffer, R.L., Walker, A., Ryder, O.A., Lyons-Weiler, M., Hedges, S.B. (2001) Human and ape molecular clocks and constraints on paleontological hypotheses, *J. Hered.* **92**, 469–474.

Summerbell, D. (1981) Evidence for regulation of growth, size and pattern in the developing chick limb bud, *J. Embryol. Exp. Morphol.* **65**, 129–150.

The International Chimpanzee Chromosome 22 Consortium. (2004) DNA sequence and comparative analysis of chimpanzee chromosome 22, *Nature* **429**, 382–388.

Tickle, C. (1992) A tool for transgenesis, *Nature* **358**, 188–189.

Wolpert, L. (1980) Positional Information and Pattern Formation in Limb Development, in: Pratt, R.M., Christiansen, R.L. (Eds.) *Current Research Trends in Prenatal Craniofacial Development*, Elsevier/North Holland, New York, pp. 89–101.

Zuckerkandl, E. (1987) On the molecular evolutionary clock, *J. Mol. Evol.* **26**, 34–46.

Zuckerkandl, E., Pauling, L. (1962) Molecular Disease, Evolution, and Genic Heterogeneity, in: Kasha, M., Pullman, B. (Eds.) *Horizons in Biochemistry*, Academic Press, New York, pp. 189–225.

Motor Neuron Diseases: Cellular and Animal Models

Georg Haase
Institut de Neurobiologie de la Méditerranée, Marseille, France

1 Human Motor Neuron Diseases 542

2 Genetics of Motor Neuron Diseases 543
2.1 SOD1 Mutations 544
2.2 Alsin Mutations 544

3 Culture Models of Motor Neuron Degeneration 544
3.1 Neuronal Cell Lines 544
3.2 Motor Neuron Cultures 545
3.3 Spinal Cord Slice Cultures 545

4 Animal Models of Motor Neuron Degeneration 547
4.1 Axotomy Models 547
4.2 Transgenic Mutant SOD1 Mice and Rats 547
4.3 *pmn* Mice 549
4.4 Mutant VEGF Mice 550

5 Therapeutic Approaches to Motor Neuron Diseases 550
5.1 Studies in Animal Models 550
5.2 Future Concepts 552

6 Conclusions and Perspectives 552

Bibliography 553
Books and Reviews 553
Primary Literature 553

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

Keywords

Amyotrophic Lateral Sclerosis (ALS)
A neurodegenerative disease involving the selective degeneration of motor neurons in cerebral cortex, brainstem, and spinal cord. ALS was first described in 1869 by the French neuropathologist Jean-Marie Charcot.

Chaperone
A protein that assists in the folding or assembly of another protein, without becoming part of the completed structure.

Guanine Nucleotide Exchange Factor (GEF)
A protein that facilitates the exchange of GDP for GTP in a GTP-binding protein.

Superoxide Dismutase 1 (SOD1)
An enzyme that catalyzes the conversion of superoxide anion to water and hydrogen peroxide.

Motor neuron diseases are human disorders characterized by the progressive degeneration and death of motor neurons and the denervation of skeletal muscles. Their course is always fatal and there are currently no really effective therapies. While most adult motor neuron diseases appear sporadically, others are inherited, because of genetic mutations in Superoxide Dismutase 1, Alsin, or Dynactin, or associated with risk factors such as genetic polymorphisms in Neurofilaments or Vascular Endothelial Growth Factor (VEGF). Multiple molecular mechanisms might therefore cause – or contribute to – disease. Some of the proposed disease mechanisms might operate cell-autonomously in motor neurons, for example, by perturbing energy metabolism, endosome trafficking, or axonal transport, whereas others seem to act indirectly, involving nonneuronal cells and diffusible factors such as nitric oxide and excess glutamate. Relevant cellular and animal models are now available to test these hypotheses and to investigate new therapeutic strategies.

1 Human Motor Neuron Diseases

Motor neuron diseases represent a spectrum of disorders in which neurons controlling voluntary movement progressively degenerate while other types of neurons are spared. Degeneration of upper motor neurons in cerebral cortex results in spasticity and hyperreflexia. Degeneration of lower motor neurons in brainstem and spinal cord leads to muscle paralysis

and atrophy. According to these clinical and pathological criteria, motor neuron diseases can be further classified: Amyotrophic lateral sclerosis (ALS, Charcot disease or Lou Gehrig's disease) affects both upper and lower motor neurons. An ALS form with predominant brainstem symptoms is called progressive bulbar palsy. Progressive muscular atrophies represent conditions in which only lower motor neurons degenerate, whereas primary lateral sclerosis involves only upper motor neurons. ALS is the most frequent motor neuron disease in the adult, with a lifetime risk of 1 in 2000 and a mean age of onset around 55 years. ALS is inevitably fatal but its course is variable between patients. Death is most often caused by failure of respiratory muscles appearing within one to five years after diagnosis.

Histopathological studies have documented a loss of motor neurons in the cervical and lumbar spinal cord of ALS patients. The number of myelinated large caliber axons in the corticospinal tract and in the ventral roots was also found to be reduced and signs of Wallerian degeneration and atrophy have been noted. Other microscopic abnormalities such as reactive gliosis, neuronal inclusion bodies, axonal spheroids, and the loss of dendrites have also been described. In skeletal muscle, denervated fibers appear atrophic and angulated. Grouping of muscle fiber types positive for esterase, ATPase, or myosin isoforms indicates muscle reinnervation from the remaining motor neurons.

2 Genetics of Motor Neuron Diseases

While most ALS cases appear spontaneously, some (~10%) are familial (FALS). These can show autosomal dominant, recessive, or X-linked inheritance. Genetic studies in FALS (familial amyotrophic lateral sclerosis) are rendered difficult by the late onset of disease, its incomplete penetrance, and the short survival of affected family members. Nevertheless, several FALS genes have been identified (Table 1) and additional FALS loci have been mapped to chromosomes 9q21-22, 15q15-22, 16, 18, and 20.

Tab. 1 Gene mutations responsible for motor neuron diseases.

Gene	*Disease*	*Locus*	*Inheritance*	*Mutation type*
SOD1	ALS1	21q22.1	Dominant[a]	Missense
Alsin	Juvenile ALS (ALS2) progressive lateral sclerosis	2q33	Recessive	Frameshift
Tau	ALS/dementia/ Parkinson complex	17q21	Dominant	Missense or intronic
Dynactin (p150 subunit)		2p13	Dominant	Missense
Androgen receptor	Spinal and bulbar muscular atrophy (SBMA)	Xq12	Recessive	CAG triplet expansion
Survival motor neuron (SMN)	SMA	5q13	Recessive	Deletion or missense

[a]With the exception of SOD1 D90A.

2.1 SOD1 Mutations

About 10 to 20% of familial ALS cases (1–2% of total) are linked to mutations in the superoxide dismutase 1 (SOD1) gene. The *SOD1* gene is ubiquitously expressed and encodes a dimeric enzyme containing 153 amino acids, one copper, and one zinc ion per monomer. The normal function of this enzyme is to catalyze the conversion of superoxide to hydrogen peroxide. Over 100 different missense mutations scattered throughout the protein have now been shown to produce motor neuron disease. All SOD1 mutations except one (D90A) show autosomal dominant transmission. It has been proposed that SOD1 mutations cause motor neuron degeneration by catalyzing aberrant free radical chemistry, by perturbing energy homeostasis, by initiating protein aggregation, by causing excessive glutamate levels, or through a combination of these mechanisms.

2.2 Alsin Mutations

Mutations in a gene termed *ALS2 or Alsin*, located on chromosome 2q33-35, have been detected in familial ALS cases with juvenile onset and slow disease progression. The gene mutations are autosomal-recessively inherited and most of them represent small deletions. Computer algorithms have predicted that the Alsin protein contains guanine nucleotide exchange factor (GEF) domains for small GTPases. Small GTPases are molecular switches that cycle between an inactive GDP-bound and an active GTP-bound state; they are activated by GEFs, which stimulate the exchange of GDP for GTP. The 183-kDa alsin protein acts as guanine nucleotide exchange factor for the small GTPase Rab5, localizes to early endosomes, and enhances their fusion. Reported alsin mutations are predicted to result in a failure of Rab5 activation, suggesting that normal endosome fusion and trafficking are essential for motor neuron maintenance.

3 Culture Models of Motor Neuron Degeneration

In an attempt to unravel disease mechanisms and to identify or validate new therapeutic candidates, various motor neuron culture models have been developed. These include neuronal cell lines, primary motor neurons, and spinal cord slices.

3.1 Neuronal Cell Lines

Neuronal cell lines, derived from tumors or *in vitro* transformed cells, only rarely resemble motor neurons. One of these exceptions is the NSC-34 cell line, which was generated by fusion of mouse neuroblastoma cells with embryonic spinal motor neurons. Like motor neurons, NSC-34 cells display a multipolar phenotype and express choline acetyltransferase (ChAT), different ion channels, and neurofilaments. These cells can also generate action potentials and induce acetylcholine receptor clusters on cocultured myotubes. NSC-34 cells transfected with mutant SOD1 plasmids display various abnormalities including mitochondrial dysfunction, increased cytochrome c release, and decreased cell survival following oxidative stress. Gene expression profiling and proteome analysis revealed underlying molecular changes such as downregulation of neurofilaments

and glutathion S-transferases and upregulation of proteins involved in NO metabolism (argininosuccinate synthase, argininosuccinate lyase, and neuronal NO synthase). Some of these changes could be confirmed *in vivo* in individual spinal cord motor neurons.

3.2 Motor Neuron Cultures

Motor neurons are notoriously difficult to purify and to maintain in primary culture. Current methods rely on three properties of these cells: their characteristic size, their expression of specific surface markers, and their dependence on trophic factors for survival and neurite outgrowth. Since motor neurons are larger in size and lower in density than most other types of neurons and glial cells, they can be enriched by differential centrifugation. As a further purification step, immuno-panning techniques for chicken and rat spinal motor neurons have been developed using monoclonal antibodies that recognize specific cell surface antigens such as SC1, BEN, or the p75 neurotrophin receptor. Spinal motor neurons can also be isolated using antibody-coated microbeads and subsequent magnetic cell sorting. Once purified, motor neurons are seeded in culture dishes coated with appropriate substrates such as polyornithine and laminin. While many motor neurons die during the first hours of culture, the remaining cells develop typical morphologies and are able to survive for several weeks when appropriate trophic factors are added. These procedures are currently limited to embryonic motor neurons from spinal cord and brainstem and not yet available for isolating cortical motor neurons.

One of the first motor neuron culture systems relevant to human ALS involved microinjection of mutant and wild-type SOD1 plasmid expression vectors into cultured motor neurons. In this system, abnormal protein aggregates appeared specifically in mutant SOD1-expressing cells. Newer studies suggest that cell death of SOD1 expressing motor neurons might also involve external triggers: In normal embryonic motor neurons, activation of the cell surface receptor Fas triggers death via two parallel pathways, the classical FADD/caspase-8 cascade and a second cascade involving Daxx, ASK1, p38, and transcriptional upregulation of neuronal NO synthase; see Fig. 1. Interestingly, motor neurons from transgenic mice for the SOD1 mutations G37R, G85R, or G93A are much more sensitive to Fas- or NO-triggered cell death than motor neurons expressing wild-type human SOD1. Cerebellar neurons, DRG neurons, or astrocytes do not display such increased sensitivity, reflecting the selective vulnerability of motor neurons in living mutant SOD1 mice. Despite their inherent technical difficulties, mutant SOD1 motor neuron cultures thus emerge as a promising tool to screen pharmacological compounds for their effects on disease-related molecular targets or cellular phenotypes.

3.3 Spinal Cord Slice Cultures

In an attempt to study motor neurons in their normal cellular environment, several groups have prepared spinal cord slice cultures from rat and mouse. These cultures can be maintained for up to 2 weeks (adult mouse) or even several months (perinatal rat). Individual motor neurons in living slice cultures are identified by their large size, the presence of action potentials, or retrograde DiI labeling prior to culturing. In fixed slices, motor neuron cell bodies

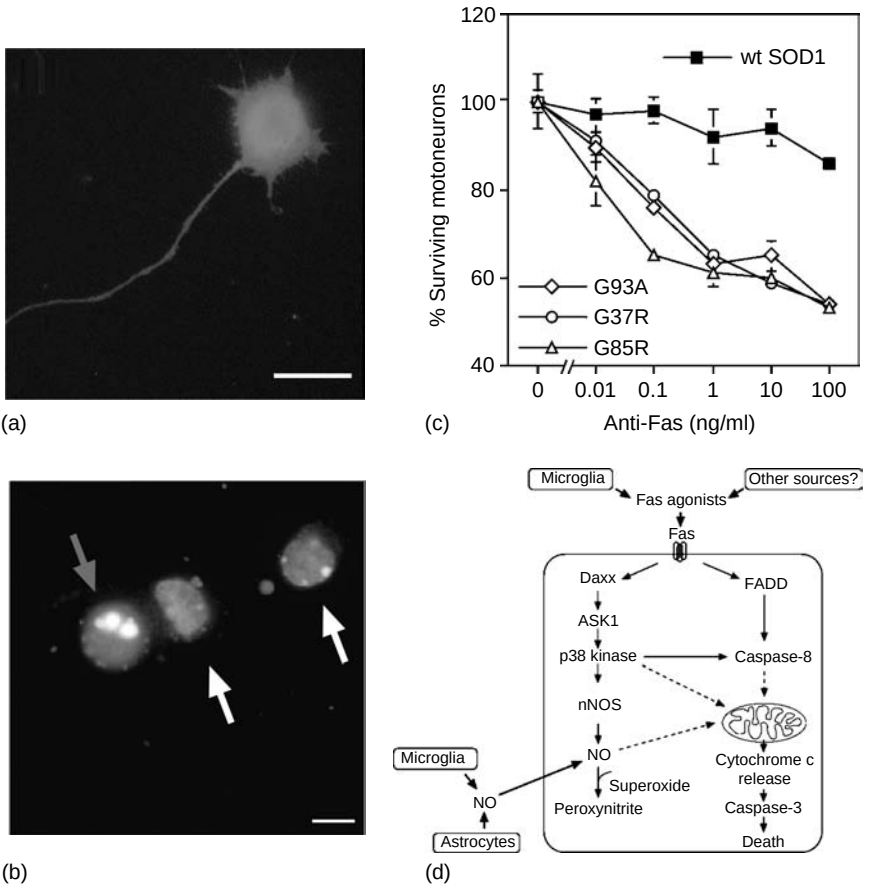

Fig. 1 Isolated motor neurons can be used to study mutant SOD1-linked motor neuron death. (a) Human superoxide SOD1 is expressed in cultured embryonic motor neurons from transgenic SOD1 G93A mice. (b) Treatment of SOD1 G85R motor neuron cultures with agonistic antibodies to the cell surface receptor Fas leads to increased apoptosis, as detected by staining with DAPI (in blue) and immunolabeling for activated caspase 3 (in red). An apoptotic motor neuron with strong caspase-3 activation and nuclear condensation c (red arrow) can be distinguished from healthy motor neurons displaying only weak caspase-3 activation and normal chromatin structure (white arrows). Scale bars: 25 µm. (c) Motor neurons from transgenic mutant SOD1 G93A, G85R, and G37R mice show higher susceptibility to Fas-triggered cell death than motor neurons from mice expressing wild-type SOD1. (d) Model of Fas-triggered motor neuron death. Cell death involves the classical FADD/caspase-8 pathway and a parallel pathway leading from Daxx, ASK1, and p38 activation to transcriptional upregulation of neuronal NO synthase (nNOS) and NO production. The presence of mutant SOD1 sensitizes motor neurons to Fas agonists and NO. Potential sources for these cell death triggers are astrocytes and microglia. (See color plate p. xxvi); see Fig. 2.

can be identified and counted by Nissl staining or immunocytochemical labeling for the neurotransmitter-synthesizing enzyme ChAT, the transcription factors Islet 1/2, or the surface receptor Ret. Slice culture systems have also been developed for the study of cortical motor neurons.

In the context of ALS research, organotypic culture systems have been used to show that chronic inhibition of glutamate uptake can cause slow degeneration of motor neurons. In this model, antiexcitotoxic agents such as CNQX, riluzole, and topiramate, nitric oxide synthase inhibitors, and neurotrophic factors such as IGF-1 and GDNF protected against motor neuron degeneration. When organotypic slice cultures were derived from transgenic mutant SOD1 mice and exposed to excitotoxic stimuli, mutant motor neuron death could be blocked with a cyclo-oxygenase-2 inhibitor. In another spinal cord slice model, chronic administration of malonate, an inhibitor of mitochondrial electron transport, was found to induce motor neuron death, which could be prevented by a number of antiexcitotoxic agents, antioxidants, and caspase inhibitors.

4 Animal Models of Motor Neuron Degeneration

4.1 Axotomy Models

In neonate animals, motor neuron death can be experimentally induced by a peripheral nerve lesion, since developing motor neurons depend for their survival on contact with their target muscle and their axonal environment. In the most widely used models, the sciatic, facial, or hypoglossal nerve is crushed or sectioned, and the survival of the lesioned motor neurons is eventually after prior retrograde labeling. The extent of motor neuron death depends on the age of the animal and the type of lesion. When a sciatic nerve lesion is performed in 1-day old rats, 90 to 100% of the corresponding motor neurons are lost. When the same type of lesion is performed at 4 or 5 days of age, between 80 and 100% of the motor neurons survive. Lesioning the peripheral nerve by section or close to the cell body both results in slower target reinnervation and reduced motor neuron survival, as compared to lesioning the nerve by section or distally.

Axotomy-induced motor neuron death has many apoptotic features. Morphological studies on lesioned motor neurons have provided evidence for DNA fragmentation, nuclear condensation, and mitochondrial disruption. Pharmacological inhibition of caspases or mitochondrial pore proteins or overexpression of apoptosis-inhibitory proteins (IAPs) reduced the extent of axotomy-induced motor neuron death. Finally, genetic studies in transgenic mice showed that overexpression of the anti-apoptotic gene *Bcl-2* or deletion of the proapoptotic gene *Bax* protected against axotomy-induced cell death. It has also been recognized that apoptotic death of axotomized motor neurons involves activation of cell surface receptors (TNF-R and Fas) and intracellular production of nitric oxide.

4.2 Transgenic Mutant SOD1 Mice and Rats

One breakthrough in ALS research has been the generation of transgenic mice and rats expressing human ALS-linked SOD1 mutations. Several transgenic lines show clinical, histopathological, and electrophysiological signs closely resembling

human motor neuron disease. Transgenic mice for the SOD1 mutants G37R, G85R, or G93A display progressive muscle atrophy and paresis, gait instability, and premature death. Similar symptoms were observed in transgenic rats expressing the SOD1 mutants G93A or H46R. In these animals, disease severity depends on the type of mutation, the number of transgenes, and the genetic background. Affected mice display a reduction in the number of spinal motor neuron cell bodies and motor axons and, to a more variable extent, astrogliosis. Other histopathological abnormalities include a fragmented Golgi apparatus, enlarged mitochondria, and various types of inclusions. Electrophysiological studies in mutant mice revealed the presence of fasciculations and fibrillations, and a decreased number of functional motor units.

Some of the molecular consequences of mutant SOD1 expression in mouse spinal cord have also been recognized: sequential activation of caspases, upregulation of apoptotic proteins such as Bad and Bax, release of cytochrome c from mitochondria, and deregulation of the cyclin-dependent kinase cdk5. How the SOD1 mutants cause these changes is still obscure. For the following reasons, altered SOD1 enzymatic activity does not seem to be causal: first, only some disease-causing SOD1 mutations are enzymatically inactive (e.g. G85R), while others retain activity (G93A or G37R). Second, no motor neuron disease is observed when SOD1 activity in mice is lowered (by knocking out the *SOD1* gene or the *CCS* gene required for SOD1 copper loading) or when SOD1 activity is elevated (by transgenic expression of wild-type SOD1). Finally, crossing transgenic mutant SOD1 mice with SOD1 knockout mice, transgenic mice overexpressing wildtype SOD1, or CCS knockout mice, has no influence on disease.

The cellular basis of mutant SOD1 toxicity also remains enigmatic: In transgenic mice, pathology is only observed when mutant SOD1 expression is ubiquitous in all cell types but not when it is restricted to astrocytes or neurons (Fig. 2). To rule out insufficient levels of transgene expression, chimeric mice have been generated that carry mixtures of wild-type and mutant

Fig. 2 Motor neuron degeneration in transgenic ALS mice is influenced by nonneuronal cells. The consequences of mutant SOD1 expression in different cell types are schematically illustrated on spinal cord cross sections. Genotypes of transgenic mice are indicated. Motor neurons are depicted as multipolar cells and glial cells as round cells. Mutant SOD1 expression in these cell types is indicated in gray or black, normal (endogenous) SOD1 expression is shown in white. (a) Normal situation. (b) In transgenic mice, ubiquitous expression of mutant (m) SOD1 causes typical motor neuron disease. (c, d) No disease is observed in transgenic mice expressing mutant SOD1, specifically in glial cells (c) or in neurons (d). (e) Crossing of Thy1: SOD1 mice with transgenic SOD1 G93A mice elevates mutant SOD1 levels (in black) in motor neurons above the ubiquitous expression levels but does not exacerbate disease. (f) Chimeric mice that carry a mixture of wild-type (wt) and mutant SOD1-expressing cells in their spinal cord show later disease onset and longer lifespan than their mutant littermates. These improvements correlate with the proportion of wild-type cells within individual spinal cords. Ubiquitin inclusions are found not only in mutant but also in wild-type motor neurons. Interestingly, mutant SOD1 motor neurons in the chimeras seem to be protected by wild-type nonneuronal cells in their environment. Taken together, these data suggest that motor neuron degeneration is the result of a complex interplay between motor neurons and surrounding nonneuronal cells.

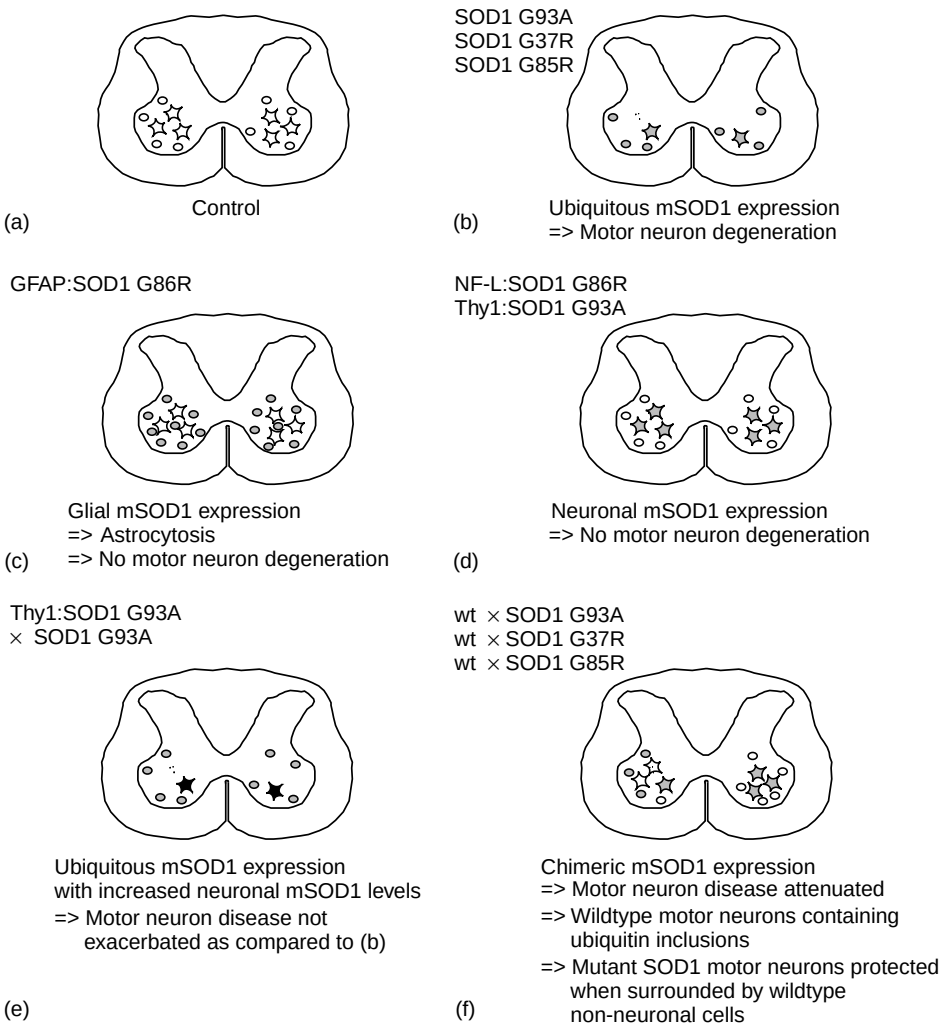

SOD1 expressing cells in the spinal cord. Interestingly, in these chimeras, degenerative signs such as ubiquitin inclusions are found not only in mutant SOD1 expressing motor neurons but also in wild-type motor neurons. Moreover, survival of mutant motor neurons seems to be enhanced by the presence of nonneuronal cells. These observations suggest that motor neuron degeneration can be triggered by non-cell-autonomous mechanisms.

4.3 *pmn* Mice

Mice with progressive motor neuronopathy (*pmn*) represent another popular motor neuron disease model. Homozygous *pmn* mice develop first neuromuscular symptoms at 2 weeks of age, i.e. much earlier than all reported mutant SOD1 mice. The *pmn* disease manifests with hindlimb atrophy and paresis and rapidly extends to

the forelimbs and other muscle groups. Affected *pmn* mice die around 40 days of age, most probably from respiratory failure. They display fibrillations, a loss of axons in motor nerves and ventral roots and a reduced number of motor neurons in some brainstem nuclei. The genetic defect was identified as a point mutation in the *Tbce* gene, which encodes a chaperone protein for tubulin folding and microtubule assembly. The *pmn* mutation results in decreased Tbce protein stability, diminished concentrations of tubulins in peripheral nerves, and progressive loss of microtubules. Interestingly, Tbce gene mutations have also been discovered in human patients with Sanjad Sakati/Kenny Caffey disease, a rare hereditary syndrome comprising hypoparathyroidism, mental retardation, facial dysmorphism, osteosclerosis, and fulgurant infections. It is thus conceivable that the dramatic human disease course obscures motor neuron disease symptoms or that the human Tbce mutations impair another Tbce protein function than the mouse mutation.

4.4 Mutant VEGF Mice

Vascular endothelial growth factor (VEGF) is a secreted 165 amino acid protein that controls the growth and permeability of blood vessels. Under conditions of low oxygen tension, VEGF is upregulated by specific transcription factors that act on a hypoxia-responsive element located in the VEGF promoter. Mice with a targeted deletion of this hypoxia-responsive element are unable to induce VEGF in response to hypoxia. Surprisingly, these mice manifest typical signs of motor neuron degeneration at an age of 5 to 7 months, with progressive muscle atrophy, impaired motor performance, and loss of motor axons and endplates. Interestingly, three polymorphisms in the human VEGF seem associated with reduced levels of plasma VEGF and an increased risk for ALS. These studies incriminate VEGF as an ALS-modifier gene. Whether VEGF acts directly as a trophic factor for motor neurons or more indirectly, by modifying spinal cord vasculature or perfusion, is currently under investigation.

5 Therapeutic Approaches to Motor Neuron Diseases

5.1 Studies in Animal Models

In the past decade, more than 100 different therapeutic approaches have been tested in animal models of motor neuron degeneration. Lesion models have been instrumental in validating neurotrophic factors as therapeutic candidates: CNTF, BDNF, NT-4/5, GDNF, and CT-1 all enhanced survival of motor neurons after axotomy. More recently, several pharmacological compounds such as deprenyl, a monoamino-oxidase B inhibitor, MK-801, a glutamate receptor antagonist, CGP3466B, a GAPDH antagonist, and riluzole were also shown to be neuroprotective in these models.

In mutant SOD1 mice, numerous compounds have been tested but only few have resulted in robust positive effects (Table 2). Following treatment with riluzole, currently the only FDA-approved drug for ALS, the mean survival of SOD1 G93A mice was improved by 11% but histological or functional effects could not be documented. Feeding the animals with creatine also increased survival and provided some motor neuron protection. Numerous ALS

Tab. 2 Selected therapeutic trials in mouse motor neuron disease models.

Treatment	*Model*	*Delivery*		*Outcome*		
		Begin[a]	*Mode*	*Lifespan increase [%]*	*Neuromuscular function*	*Motor neuron protection*
Riluzole	SOD1 G93A	Early	Oral	11	Not reported	Not reported
	pmn	Early	Oral	<10	Yes	Not reported
Creatine	SOD1 G93A	Early	Oral	18	Yes	Yes
Minocyclin	SOD1 G37R	Early	Oral	≈6	Yes	Yes
Minocyclin+ Riluzole+ Nimodipin	SOD1 G37R	Early	Oral	13	Yes	Yes
zVAD-fmk	SOD1 G93A	Early	Intrathecal	21	Yes	Yes
NT-3 Adeno	*pmn*	Early	i.m.	50	yes	Yes
GDNF AAV	SOD1 G93A	Early	i.m.	14	yes	Yes
CT-1Adeno	SOD1 G93A	Early	i.m	8	yes	Yes
	pmn	Early	i.m.	18	yes	Yes
IGF AAV	SOD1 G93A	Late	i.m.	18	yes	Yes

[a]Treatment begin is classified with respect to disease onset: "early" treatments are initiated in the presymptomatic period, "late" treatments are started at or after disease onset.

patients have since then taken creatine, but a recent placebo-controlled clinical trial disclosed any beneficial effect of oral creatine at 10 g per day. Encouraging results were obtained by intrathecal infusion of the broad caspase inhibitor zVAD-fmk. Recently, several studies have also reported positive effects of minocyclin, which might be related to the inhibitory actions of this tetracyclin analog on caspase activation, mitochondrial cytochrome c release, cdk5 deregulation, or microglial cell activation. When minocyclin was administered in combination with riluzole and the calcium channel blocker nimodipin, the therapeutic benefit was further enhanced.

Gene therapy strategies have been developed to overcome the side effects of systemically delivered neurotrophic factors. These strategies have been tested since 1992 in *pmn* mice and since 2001 in transgenic mutant SOD1 mice (Table 2). In *pmn* mice, the implantation of genetically engineered tumor cells or subcutaneous implantation of encapsulated fibroblasts secreting CNTF reduced degeneration of phrenic and facial motor neurons. When injected intramuscularly in newborn *pmn* mice, adenovirus vectors encoding the neurotrophic factors NT-3, CNTF, and CT-1 were shown to improve neuromuscular function, reduce motor neuron degeneration, and increase the animal's lifespan. Recently, a nonviral gene transfer method, intramuscular injection and electroporation of plasmid DNA, proved to be as efficient as the adenovirus-based strategy to attenuate disease symptoms.

In transgenic mutant SOD1 mice, adenovirus, AAV, and lentivirus vectors coding for CT-1, GDNF, and more recently IGF-1 and VEGF, administered by intramuscular injection, resulted in therapeutic benefit. Some of these therapies were also capable to slow down motor neuron degeneration when started late, after the onset of

disease. In order to translate these encouraging findings into clinical therapies, a number of fundamental questions now needs to be addressed: Are the observed effects due to release of the neurotrophic factors into the bloodstream or due to retrograde vector transport from muscles to motor neuron cell bodies? What are the best factors, alone or in combination? What is the safest and most efficient vector system?

5.2 Future Concepts

Since motor axons degenerate weeks to months before cell bodies in human ALS and in corresponding mouse models, therapies aimed at axonal protection appear particularly attractive. This concept is further illustrated by studies in the mouse mutant *Wallerian degeneration slow (Wld^S, Ola)*: In normal mice, a peripheral nerve lesion leads to axonal disintegration and reactive proliferation of Schwann cells in the distal nerve stump, a process called *Wallerian degeneration*. In *Wld^S* mice, however, the distal nerve stump remains intact following a lesion and remains able to conduct action potentials during several weeks. The *Wld^S* mutation was identified as a triplication on chromosome 4 that gives rise to a fusion protein between a truncated ubiquitination factor (Ube4b) and an enzyme involved in NAD synthesis (nicotinamide mononucleotide adenylyltransferase, Nmnat). Interestingly, Wld^S protects axons not only against lesion but also in various other paradigms such as Taxol- and Vincristin-induced neurotoxicity and in the inherited motor neuron disease *pmn*. Because of this broad therapeutic potential, Wld^S and its molecular effectors have become the focus of intense research efforts.

Another innovative approach consists in knocking down the expression of pathological SOD1 alleles by using the RNA interference (RNAi) technique. RNAi has been discovered as a natural gene-silencing mechanism in which double-stranded RNA triggers the degradation of homologous messenger RNAs with extraordinary sequence specificity. The mediators of mRNA degradation are small, 21- to 23-nucleotide long, RNAs (siRNAs). Since almost most ALS-linked SOD1 mutations represent single nucleotide substitutions, it is possible to design RNA sequences that specifically silence mutant SOD1 alleles without affecting expression of the wild-type allele. These RNAs are transfected as double-stranded oligonucleotides into the target cell or expressed as small hairpin RNAs from plasmid DNA vectors or lentiviral vectors. The appropriate target cell and the level of SOD1 gene silencing necessary to achieve therapeutic effects remain to be determined.

6 Conclusions and Perspectives

In the last decade, advances in genetics, molecular and cell biology, and gene therapy have boosted research on motor neuron diseases. Several gene mutations responsible for motor neuron diseases in humans (SOD1, Alsin) and mice (Dynein, Chaperones, etc.) have been identified. These discoveries led to new pathogenic hypotheses, implicating excessive glutamate levels, toxic protein aggregates, mitochondrial damage, or axonal dysfunction. It is becoming clear that multiple disease pathways – triggered in the motor neuron or in neighboring cells – can converge and

cause degeneration. Experimental therapies targeting these pathways at various levels are investigated with increasing success in relevant cellular and animal models. Translating these approaches into clinically relevant therapies represents the major challenge of the future.

See also Motor Neuron Diseases: Molecular Mechanism, Pathophysiology, and Treatments; Neuron Chemistry.

Bibliography

Books and Reviews

Beckman, J.S., Estevez, A.G., Crow, J.P., Barbeito, L. (2001) Superoxide dismutases and the death of motoneurons in ALS, *Trends Neurosci.* **24**, S15–S20.

Bruijn, L.I., Miller, T.M., Cleveland, D.W. (2004) Unraveling the mechanisms involved in motor neuron degeneration in ALS, *Annu. Rev. Neurosci.* **27**, 723–749.

Coleman, M.P., Perry, V.H. (2002) Axon pathology in neurological disease: a neglected therapeutic target, *Trends Neurosci.* **25**, 532–537.

Cleveland, D.W., Rothstein, J.D. (2001) From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS, *Nat. Rev. Neurosci.* **2**, 806–819.

Hand, C.K., Rouleau, G.A. (2002) Familial amyotrophic lateral sclerosis, *Muscle Nerve* **25**, 135–159.

Henderson, C.E., et al. (1995) in *Nerve Cell Culture: A Practical Approach*, Cohen J., Wilkin G. (Eds.) Oxford University Press, London, UK, pp. 69–81.

Hirano, A. (1982) in: Rowland L.P. (Ed.) *Human Motor Neuron Diseases*, Raven Press, New York, pp. 75–87.

Julien, J.P. (2001) Amyotrophic lateral sclerosis. unfolding the toxicity of the misfolded, *Cell* **104**, 581–591.

Nicole, S., Cifuentes-Diaz, C., Frugier, T., Melki, J. (2002) Spinal muscular atrophy: recent advances and future prospects, *Muscle Nerve* **26**, 4–13.

Orrell, R.W. (2000) Amyotrophic lateral sclerosis: copper/zinc superoxide dismutase (SOD1) gene mutations, *Neuromuscul. Disord.* **10**, 63–68.

Valentine, J.S., Hart, P.J. (2003) Misfolded CuZnSOD and amyotrophic lateral sclerosis, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3617–3622.

Primary Literature

Allen, S., et al. (2003) Analysis of the cytosolic proteome in a cell culture model of familial amyotrophic lateral sclerosis reveals alterations to the proteasome, antioxidant defenses, and nitric oxide synthetic pathways, *J. Biol. Chem.* **278**(8), 6371–6383.

Arce, V., et al. (1999) Cardiotrophin-1 requires LIFRbeta to promote survival of mouse motoneurons purified by a novel technique, *J. Neurosci. Res.* **55**(1), 119–126.

Bar, P.R. (2000) Motor neuron disease in vitro: the use of cultured motor neurons to study amyotrophic lateral sclerosis, *Eur. J. Pharmacol.* **405**(1–3), 285–295.

Beckman, J.S., et al. (2001) Superoxide dismutases and the death of motoneurons in ALS, *Trends Neurosci.* **24**(11 (Suppl.)), S15–S20.

Bommel, H., et al. (2002) Missense mutation in the tubulin-specific chaperone E (Tbce) gene in the mouse mutant progressive motor neuronopathy, a model of human motoneuron disease, *J. Cell. Biol.* **159**(4), 563–569.

Bordet, T., et al. (1999) Adenoviral cardiotrophin-1 gene transfer protects pmn mice from progressive motor neuronopathy, *J. Clin. Invest.* **104**(8), 1077–1085.

Bordet, T., et al. (2001) Protective effects of cardiotrophin-1 adenoviral gene transfer on neuromuscular degeneration in transgenic ALS mice, *Hum. Mol. Genet.* **10**(18), 1925–1933.

Bruijn, L.I., et al. (1997) ALS-linked SOD1 mutant G85R mediates damage to astrocytes and promotes rapidly progressive disease with SOD1-containing inclusions, *Neuron* **18**(2), 327–338.

Bruijn, L.I., et al. (1998) Aggregation and motor neuron toxicity of an ALS-linked SOD1 mutant independent from wild-type SOD1, *Science* **281**(5384), 1851–1854.

Carlin, K.P., et al. (2000) Dendritic L-type calcium currents in mouse spinal

motoneurons: implications for bistability, *Eur. J. Neurosci.* **12**(5), 1635–1646.

Casanovas, A., et al. (1996) Prevention by lamotrigine, MK-801 and N omega-nitro-L-arginine methyl ester of motoneuron cell death after neonatal axotomy, *Neuroscience* **71**(2), 313–325.

Cashman, N.R., et al. (1992) Neuroblastoma x spinal cord (NSC) hybrid cell lines resemble developing motor neurons, *Dev. Dyn.* **194**(3), 209–221.

Chan, Y.M., et al. (2003) Inhibition of caspases promotes long-term survival and reinnervation by axotomized spinal motoneurons of denervated muscle in newborn rats, *Exp. Neurol.* **181**(2), 190–203.

Charcot, J.-M., Joffroy, A. (1869) Deux cas d'atrophie musculaire progressive avec lésion de la substance grise et des fisceaux de la moelle épinière, *Arch. Physiol.* **2**, 354, 629, 744.

Chiu, A.Y., et al. (1995) Age-dependent penetrance of disease in a transgenic mouse model of familial amyotrophic lateral sclerosis, *Mol. Cell. Neurosci.* **6**(4), 349–362.

Chou, S.M. (1992) in: Smith R.A. (Ed.) *Handbook of Amyotrophic Lateral Sclerosis*, Marcel Dekker, New York, pp. 133–181.

Cifuentes-Diaz, C., et al. (2002) Neurofilament accumulation at the motor endplate and lack of axonal sprouting in a spinal muscular atrophy mouse model, *Hum. Mol. Genet.* **11**(12), 1439–1447.

Clement, A.M., et al. (2003) Wild-type nonneuronal cells extend survival of SOD1 mutant motor neurons in ALS mice, *Science* **302**(5642), 113–117.

Coleman, M.P., et al. (1998) An 85-kb tandem triplication in the slow Wallerian degeneration (Wlds) mouse, *Proc. Natl. Acad. Sci. U. S. A.* **95**(17), 9985–9990.

Conforti, L., et al. (2000) A Ufd2/D4Cole1e chimeric protein and overexpression of Rbp7 in the slow Wallerian degeneration (WldS) mouse, *Proc. Natl. Acad. Sci. U. S. A.* **97**(21), 11377–11382.

Connelly, C.A., et al. (2000) Metabolic activity of cultured rat brainstem, hippocampal and spinal cord slices, *J. Neurosci. Methods* **99**(1–2), 1–7.

Corse, A.M., et al. (1999) Preclinical testing of neuroprotective neurotrophic factors in a model of chronic motor neuron degeneration, *Neurobiol. Dis.* **6**(5), 335–346.

Dalcanto, M.C., Gurney, M.E. (1995) Neuropathological changes in two lines of mice carrying a transgene for mutant human Cu,Zn SOD, and in mice overexpressing wild type human SOD: A model of familial amyotrophic lateral sclerosis (FALS), *Brain Res.* **676**(1), 25–40.

de Bilbao, F., Dubois-Dauphin, M. (1996) Acute application of an interleukin-1 beta-converting enzyme-specific inhibitor delays axotomy-induced motoneurone death, *Neuroreport* **7**(18), 3051–3054.

de Bilbao, F., Dubois-Dauphin, M. (1996) Time course of axotomy-induced apoptotic cell death in facial motoneurons of neonatal wild type and bcl-2 transgenic mice, *Neuroscience* **71**(4), 1111–1119.

Deckwerth, T.L., et al. (1996) BAX is required for neuronal death after trophic factor deprivation and during development, *Neuron* **17**(3), 401–411.

Drachman, D.B., Rothstein, J.D. (2000) Inhibition of cyclooxygenase-2 protects motor neurons in an organotypic model of amyotrophic lateral sclerosis, *Ann. Neurol.* **48**(5), 792–795.

Dubois-Dauphin, M., et al. (1994) Neonatal motoneurons overexpressing the bcl-2 protooncogene in transgenic mice are protected from axotomy-induced cell death, *Proc. Natl. Acad. Sci. U. S. A.* **91**(8), 3309–3313.

Durham, H.D., et al. (1993) Evaluation of the spinal cord neuron X neuroblastoma hybrid cell line NSC-34 as a model for neurotoxicity testing, *Neurotoxicology* **14**(4), 387–395.

Durham, H.D., et al. (1997) Aggregation of mutant Cu/Zn superoxide dismutase proteins in a culture model of ALS, *J. Neuropathol. Exp. Neurol.* **56**(5), 523–530.

Ferri, A., et al. (2003) Inhibiting axon degeneration and synapse loss attenuates apoptosis and disease progression in a mouse model of motoneuron disease, *Curr. Biol.* **13**(8), 669–673.

Frey, D., et al. (2000) Early and selective loss of neuromuscular synapse subtypes with low sprouting competence in motoneuron diseases, *J. Neurosci.* **20**(7), 2534–2542.

Garcia, M.L., Cleveland, D.W. (2001) Going new places using an old MAP: tau, microtubules and human neurodegenerative disease, *Curr. Opin. Cell Biol.* **13**(1), 41–48.

Gong, Y.H., et al. (2000) Restricted expression of G86R Cu/Zn superoxide dismutase in

astrocytes results in astrocytosis but does not cause motoneuron degeneration, *J. Neurosci.* **20**(2), 660–665.

Greensmith, L., Vrbova, G. (1996) Motoneurone survival: A functional approach, *Trends Neurosci.* **19**(11), 450–455.

Groeneveld, G.J., et al. (2003) A randomized sequential trial of creatine in amyotrophic lateral sclerosis, *Ann. Neurol.* **53**(4), 437–445.

Guegan, C., et al. (2001) Recruitment of the mitochondrial-dependent apoptotic pathway in amyotrophic lateral sclerosis, *J. Neurosci.* **21**(17), 6569–6576.

Gurney, M.E., et al. (1994) Motor neuron degeneration in mice that express a human Cu,Zn superoxide dismutase mutation, *Science* **264**(5166), 1772–1775.

Gurney, M.E., et al. (1996) Benefit of vitamin E, riluzole, and gabapentin in a transgenic model of familiar amyotrophic lateral sclerosis, *Ann. Neurol.* **39**(2), 147–157.

Haase, G., et al. (1997) Gene therapy of murine motor neuron disease using adenoviral vectors for neurotrophic factors, *Nat. Med.* **3**(4), 429–436.

Haase, G., et al. (1999) Therapeutic benefit of CNTF in progressive motor neuronopathy depends on the route of delivery, *Ann. Neurol.* **45**(3), 296–304.

Hadano, S., et al. (2001) A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2, *Nat. Genet.* **29**(2), 166–173.

Haenggeli, C., Kato, A.C. (2002) Differential vulnerability of cranial motoneurons in mouse models with motor neuron degeneration, *Neurosci. Lett.* **335**(1), 39–43.

Hannon, G.J. (2002) RNA interference, *Nature* **418**(6894), 244–251.

Henderson, C.E., et al. (1994) GDNF: A potent survival factor for motoneurons present in peripheral nerve and muscle, *Science* **266**(5187), 1062–1064.

Hentati, A., et al. (1994) Linkage of recessive familial amyotrophic lateral sclerosis to chromosome 2q33-q35, *Nat. Genet.* **7**(3), 425–428.

Ho, T.W., et al. (2000) TGFbeta trophic factors differentially modulate motor axon outgrowth and protection from excitotoxicity, *Exp. Neurol.* **161**(2), 664–675.

Hori, N., et al. (2001) Intracellular activity of rat spinal cord motoneurons in slices, *J. Neurosci. Methods* **112**(2), 185–191.

Howland, D.S., et al. (2002) Focal loss of the glutamate transporter EAAT2 in a transgenic rat model of SOD1 mutant-mediated amyotrophic lateral sclerosis (ALS), *Proc. Natl. Acad. Sci. U. S. A.* **99**(3), 1604–1609.

Ince, P.G., et al. (1998) Amyotrophic lateral sclerosis: current issues in classification, pathogenesis and molecular pathology, *Neuropathol. Appl. Neurobiol.* **24**(2), 104–117.

Iwasaki, Y., Ikeda, K. (1999) Prevention by insulin-like growth factor-I and riluzole in motor neuron death after neonatal axotomy, *J. Neurol. Sci.* **169**(1–2), 148–155.

Iwasaki, Y., et al. (1996) Deprenyl and pergolide rescue spinal motor neurons from axotomy-induced neuronal death in the neonatal rat, *Neurol. Res.* **18**(2), 168–170.

Kaal, E.C., et al. (2000) Chronic mitochondrial inhibition induces selective motoneuron death in vitro: a new model for amyotrophic lateral sclerosis, *J. Neurochem.* **74**(3), 1158–1165.

Kashihara, Y., et al. (1987) Cell death of axotomized motoneurones in neonatal rats, and its prevention by peripheral reinnervation, *J. Physiol.* **386**, 135–148.

Kaspar, B.K., et al. (2003) Retrograde viral delivery of IGF-1 prolongs survival in a mouse ALS model, *Science* **301**(5634), 839–842.

Kennedy, W.R., et al. (1968) Progressive proximal spinal and bulbar muscular atrophy of late onset. A sex-linked recessive trait, *Neurology* **18**(7), 671–680.

Kennel, P., et al. (2000) Riluzole prolongs survival and delays muscle strength deterioration in mice with progressive motor neuronopathy (pmn), *J. Neurol. Sci.* **180**(1–2), 55–61.

Kennel, P.F., et al. (1996) Neuromuscular function impairment is not caused by motor neurone loss in FALS mice: an electromyographic study, *Neuroreport* **7**(8), 1427–1431.

Kennel, P.F., et al. (1996) Electromyographical and motor performance studies in the pmn mouse model of neurodegenerative disease, *Neurobiol. Dis.* **3**(2), 137–147.

Kirby, J., et al. (2002) Differential gene expression in a cell culture model of SOD1-related familial motor neurone disease, *Hum. Mol. Genet.* **11**(17), 2061–2075.

Klivenyi, P., et al. (1999) Neuroprotective effects of creatine in a transgenic animal model of amyotrophic lateral sclerosis, *Nat. Med.* **5**(3), 347–350.

Kong, J.M., Xu, Z.S. (1998) Massive mitochondrial degeneration in motor neurons triggers the onset of amyotrophic lateral sclerosis in mice expressing a mutant SOD1, *J. Neurosci.* **18**(9), 3241–3250.

Korinthenberg, R., et al. (1997) Congenital axonal neuropathy caused by deletions in the spinal muscular atrophy region, *Ann. Neurol.* **42**(3), 364–368.

Krassioukov, A.V., et al. (2002) An in vitro model of neurotrauma in organotypic spinal cord cultures from adult mice, *Brain Res. Brain Res. Protoc.* **10**(2), 60–68.

Kriz, J., et al. (2002) Minocycline slows disease progression in a mouse model of amyotrophic lateral sclerosis, *Neurobiol. Dis.* **10**(3), 268–278.

Kriz, J., et al. (2003) Efficient three-drug cocktail for disease induced by mutant superoxide dismutase, *Ann. Neurol.* **53**(4), 429–436.

Kunst, C.B., et al. (2000) Genetic mapping of a mouse modifier gene that can prevent ALS onset, *Genomics* **70**(2), 181–189.

La Spada, A.R., et al. (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy, *Nature* **352**(6330), 77–79.

Lambrechts, D., et al. (2003) VEGF is a modifier of amyotrophic lateral sclerosis in mice and humans and protects motoneurons against ischemic death, *Nat. Genet.* **34**(4), 383–394.

Lefebvre, S., et al. (1995) Identification and characterization of a spinal muscular atrophy-determining gene [see comments], *Cell* **80**(1), 155–165.

Leigh, P.N., et al. (1991) Ubiquitin-immunoreactive intraneuronal inclusions in amyotrophic lateral sclerosis. Morphology, distribution, and specificity, *Brain* **114**(Pt 2), 775–788.

Lesbordes, J.C., et al. (2002) In vivo electrotransfer of the cardiotrophin-1 gene into skeletal muscle slows down progression of motor neuron degeneration in pmn mice, *Hum. Mol. Genet.* **11**(14), 1615–1625.

Li, L., et al. (1998) Characterization of spinal motoneuron degeneration following different types of peripheral nerve injury in neonatal and adult mice, *J. Comp. Neurol.* **396**(2), 158–168.

Li, M., et al. (2000) Functional role of caspase-1 and caspase-3 in an ALS transgenic mouse model, *Science* **288**(5464), 335–339.

Li, M., et al. (2000) Functional role of caspase-1 and caspase-3 in an ALS transgenic mouse model, *Science* **288**(5464), 335–339.

Lino, M.M., et al. (2002) Accumulation of SOD1 mutants in postnatal motoneurons does not cause motoneuron pathology or motoneuron disease, *J. Neurosci.* **22**(12), 4825–4832.

Liu, R., et al. (2002) Increased mitochondrial antioxidative activity or decreased oxygen free radical propagation prevent mutant SOD1-mediated motor neuron cell death and increase amyotrophic lateral sclerosis-like transgenic mouse survival, *J. Neurochem.* **80**(3), 488–500.

Lowrie, M.B., et al. (1982) Recovery of slow and fast muscles following nerve injury during early postnatal development in the rat, *J. Physiol.* **331**, 51–66.

Lyon, M.F., et al. (1993) A gene affecting Wallerian nerve degeneration maps distally on mouse chromosome 4, *Proc. Natl. Acad. Sci. U. S. A.* **90**(20), 9717–9720.

Mack, T.G., et al. (2001) Wallerian degeneration of injured axons and synapses is delayed by a Ube4b/Nmnat chimeric gene, *Nat. Neurosci.* **4**(12), 1199–1206.

Maragakis, N.J., et al. (2003) Topiramate protects against motor neuron degeneration in organotypic spinal cord cultures but not in G93A SOD1 transgenic mice, *Neurosci. Lett.* **338**(2), 107–110.

Mariotti, R., et al. (1997) Age-dependent induction of nitric oxide synthase activity in facial motoneurons after axotomy, *Exp. Neurol.* **145**(2 Pt 1), 361–370.

Martin, N., et al. (2002) A missense mutation in Tbce causes progressive motor neuronopathy in mice, *Nat. Genet.* **32**(3), 443–447.

Menzies, F.M., et al. (2002) Mitochondrial dysfunction in a cell culture model of familial amyotrophic lateral sclerosis, *Brain* **125**(Pt 7), 1522–1533.

Menzies, F.M., et al. (2002) Selective loss of neurofilament expression in Cu/Zn superoxide dismutase (SOD1) linked amyotrophic lateral sclerosis, *J. Neurochem.* **82**(5), 1118–1128.

Mourelatos, Z., et al. (1996) The Golgi apparatus of spinal cord motor neurons in transgenic mice expressing mutant Cu,Zn superoxide dismutase becomes fragmented in early, preclinical stages of the disease, *Proc. Natl. Acad. Sci. U. S. A.* **93**(11), 5472–5477.

Nagai, M., et al. (2001) Rats expressing human cytosolic copper-zinc superoxide dismutase transgenes with amyotrophic lateral sclerosis: associated mutations develop motor neuron disease, *J. Neurosci.* **21**(23), 9246–9254.

Nguyen, M.D., et al. (2001) Deregulation of Cdk5 in a mouse model of ALS: toxicity alleviated by perikaryal neurofilament inclusions, *Neuron* **30**(1), 135–147.

Oosthuyse, B., et al. (2001) Deletion of the hypoxia-response element in the vascular endothelial growth factor promoter causes motor neuron degeneration, *Nat. Genet.* **28**(2), 131–138.

Otomo, A., et al. (2003) ALS2, a novel guanine nucleotide exchange factor for the small GTPase Rab5, is implicated in endosomal dynamics, *Hum. Mol. Genet.* **12**(14), 1671–1687.

Parvari, R., et al. (2002) Mutation of TBCE causes hypoparathyroidism-retardation-dysmorphism and autosomal recessive Kenny-Caffey syndrome, *Nat. Genet.* **32**(3), 448–452.

Pasinelli, P., et al. (2000) Caspase-1 and -3 are sequentially activated in motor neuron death in Cu,Zn superoxide dismutase-mediated familial amyotrophic lateral sclerosis, *Proc. Natl. Acad. Sci. U. S. A.* **97**(25), 13901–13906.

Pennica, D., et al. (1996) Cardiotrophin-1, a cytokine present in embryonic muscle, supports long-term survival of spinal motoneurons, *Neuron* **17**(1), 63–74.

Perrelet, D., et al. (2000) IAP family proteins delay motoneuron cell death in vivo, *Eur. J. Neurosci.* **12**(6), 2059–2067.

Pramatarova, A., et al. (2001) Neuron-specific expression of mutant superoxide dismutase 1 in transgenic mice does not lead to motor impairment, *J. Neurosci.* **21**(10), 3369–3374.

Puls, I., et al. (2003) Mutant dynactin in motor neuron disease, *Nat. Genet.* **33**(4), 455–456.

Raivich, G., et al. (2002) Cytotoxic potential of proinflammatory cytokines: combined deletion of TNF receptors TNFR1 and TNFR2 prevents motoneuron cell death after facial axotomy in adult mouse, *Exp. Neurol.* **178**(2), 186–193.

Rakowicz, W.P., et al. (2002) Glial cell line-derived neurotrophic factor promotes the survival of early postnatal spinal motor neurons in the lateral and medial motor columns in slice culture, *J. Neurosci.* **22**(10), 3953–3962.

Raoul, C., et al. (1999) Programmed cell death of embryonic motoneurons triggered through the fas death receptor, *J. Cell. Biol.* **147**(5), 1049–1062.

Raoul, C., et al. (2002) Motoneuron death triggered by a specific pathway downstream of Fas. potentiation by ALS-linked SOD1 mutations, *Neuron* **35**(6), 1067–1083.

Reaume, A.G., et al. (1996) Motor neurons in Cu/Zn superoxide dismutase-deficient mice develop normally but exhibit enhanced cell death after axonal injury, *Nat. Genet.* **13**(1), 43–47.

Ripps, M.E., et al. (1995) Transgenic mice expressing an altered murine superoxide dismutase gene provide an animal model of amyotrophic lateral sclerosis, *Proc. Natl. Acad. Sci. U. S. A.* **92**(3), 689–693.

Rossiter, J.P., et al. (1996) Axotomy-induced apoptotic cell death of neonatal rat facial motoneurons: time course analysis and relation to NADPH-diaphorase activity, *Exp. Neurol.* **138**(1), 33–44.

Rothstein, J.D., et al. (1993) Chronic inhibition of glutamate uptake produces a model of slow neurotoxicity, *Proc. Natl. Acad. Sci. U. S. A.* **90**(14), 6591–6595.

Rothstein, J.D., Kuncl, R.W. (1995) Neuroprotective strategies in a model of chronic glutamate-mediated motor neuron toxicity, *J. Neurochem.* **65**(2), 643–651.

Sagot, Y., et al. (1995) Polymer encapsulated cell lines genetically engineered to release ciliary neurotrophic factor can slow down progressive motor neuronopathy in the mouse, *Eur. J. Neurosci.* **7**(6), 1313–1322.

Schmalbruch, H. (1984) Motoneuron death after sciatic nerve section in newborn rats, *J. Comp. Neurol.* **224**, 252–258.

Schmalbruch, H., et al. (1991) A new mouse mutant with progressive motor neuronopathy, *J. Neuropathol. Exp. Neurol.* **50**, 192–204.

Schmalbruch, H., Rosenthal, A. (1995) Neurotrophin-4/5 postpones the death of injured spinal motoneurons in newborn rats, *Brain Res.* **700**(1–2), 254–260.

Schnaar, R.I., Schaffner, A.E. (1981) Separation of cell types from embryonic chicken and rat spinal cord: characterization of motoneuron-enriched fractions, *J. Neurosci.* **1**(2), 204–217.

Sendtner, M., et al. (1990) Ciliary neurotrophic factor prevents the degeneration of motor neurons after axotomy, *Nature* **345**(6274), 440–441.

Sendtner, M., et al. (1992) Brain-derived neurotrophic factor prevents the death of motoneurons in newborn rats after nerve section, *Nature* **360**, 757–759.

Sendtner, M., et al. (1992) Ciliary neurotrophic factor prevents degeneration of motor

neurons in mouse mutant progressive motor neuronopathy, *Nature* **358**(6386), 502–504.

Subramaniam, J.R., et al. (2002) Mutant SOD1 causes motor neuron disease independent of copper chaperone-mediated copper loading, *Nat. Neurosci.* **5**(4), 301–307.

Tandan, R., Bradley, W.G. (1985) Amyotrophic lateral sclerosis: Part 1. Clinical features, pathology, and ethical issues in management, *Ann. Neurol.* **18**(3), 271–280.

Terrado, J., et al. (2000) Soluble TNF receptors partially protect injured motoneurons in the postnatal CNS, *Eur. J. Neurosci.* **12**(9), 3443–3447.

Tian, G., et al. (1996) Pathway leading to correctly folded beta-tubulin, *Cell* **86**(2), 287–296.

Tu, P.H., et al. (1996) Transgenic mice carrying a human mutant superoxide dismutase transgene develop neuronal cytoskeletal pathology resembling human amyotrophic lateral sclerosis lesions, *Proc. Natl. Acad. Sci. U. S. A.* **93**(7), 3155–3160.

Ugolini, G., et al. (2003) Fas/tumor necrosis factor receptor death signaling is required for axotomy-induced death of motoneurons in vivo, *J. Neurosci.* **23**(24), 8526–8531.

Van Westerlaak, M.G., et al. (2001) Chronic mitochondrial inhibition induces glutamate-mediated corticomotoneuron death in an organotypic culture model, *Exp. Neurol.* **167**(2), 393–400.

Vanderluit, J.L., et al. (2003) In vivo application of mitochondrial pore inhibitors blocks the induction of apoptosis in axotomized neonatal facial motoneurons, *Cell. Death Differ.* **10**(9), 969–976.

Vukosavic, S., et al. (1999) Bax and Bcl-2 interaction in a transgenic mouse model of familial amyotrophic lateral sclerosis, *J. Neurochem.* **73**(6), 2460–2468.

Waldmeier, P.C., et al. (2000) Neurorescuing effects of the GAPDH ligand CGP 3466B, *J. Neural. Transm. Suppl.* **60**, 197–214.

Wang, L.J., et al. (2002) Neuroprotective effects of glial cell line-derived neurotrophic factor mediated by an adeno-associated virus vector in a transgenic animal model of amyotrophic lateral sclerosis, *J. Neurosci.* **22**(16), 6920–6928.

Wang, M.S., et al. (2001) The WldS protein protects against axonal degeneration: a model of gene therapy for peripheral neuropathy, *Ann. Neurol.* **50**(6), 773–779.

Wang, M.S., et al. (2002) WldS mice are resistant to paclitaxel (taxol) neuropathy, *Ann. Neurol.* **52**(4), 442–447.

Wong, P.C., et al. (1995) An adverse property of a familial ALS-linked SOD1 mutation causes motor neuron disease characterized by vacuolar degeneration of mitochondria, *Neuron* **14**(6), 1105–1116.

Xia, X.G., et al. (2003) An enhanced U6 promoter for synthesis of short hairpin RNA, *Nucleic Acids Res.* **31**(17), e100.

Yan, Q., et al. (1992) Brain-derived neurotrophic factor rescues spinal motor neurons from axotomy-induced cell death, *Nature* **360**(6406), 753–755.

Yang, Y., et al. (2001) The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis, *Nat. Genet.* **29**(2), 160–165.

Zhu, S., et al. (2002) Minocycline inhibits cytochrome c release and delays progression of amyotrophic lateral sclerosis in mice, *Nature* **417**(6884), 74–78.

Motor Neuron Diseases: Molecular Mechanism, Pathophysiology, and Treatments

Michael Sendtner
Institute of Clinical Neurobiology, Wuerzburg, Germany

1 Neuronal Cell Death: Implications for Motoneuron Disease 560

2 Motor Neurons: Anatomical Aspects 561

3 Clinical Appearance and Pathology 562
3.1 Inherited forms of MND: the Identification of Gene Defects, Animal Models, and Disease Mechanisms 562
3.2 Basic Pathological Findings of ALS 565

4 Sporadic MND: The Pathophysiological Mechanisms Are Still Obscure 566
4.1 The Autoimmune Hypothesis of Motoneuron Disease 567
4.2 Toxic Lesions of Motor Neurons 567

5 Oxidative Stress: Is Motoneuron Degeneration in Familial ALS Due to Loss or Gain of Function of the Enzyme Superoxide Dismutase? 568

6 Disintegration of Axonal Transport: Implications for MND 569

7 Identification of Neurotrophic Factors and Cytokines that can Regulate Survival and Function of Motor Neurons 571

8 Future Prospects for Treatment of MND 571

Bibliography 573
Books and Reviews 573
Primary Literature 573

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

Keywords

Amyotrophic Lateral Sclerosis (ALS)
A clinical syndrome defined by degenerative changes of lower and upper motor neurons.

Motor Neuron Disease (MND)
A variety of sporadic and inherited diseases characterized by muscle weakness due to primary degeneration of motor neurons of the spinal cord, brain stem motor nuclei, and populations of central motor neurons.

Neurotrophic Factors
Polypeptide molecules that support the survival of specific populations of neuronal cells, such as spinal motor neurons in culture and *in vivo*, via specific signal-transducing receptors.

Spinal Muscular Atrophy (SMA)
A group of disorders that are characterized by degeneration of lower (spinal and bulbar) motor neurons.

The term "motor neuron disease" refers to a variety of devastating neurological disorders characterized by a generalized degeneration of motoneurons. The common clinical manifestations of such disorders are progressive muscle weakness and often death by respiratory failure. Inherited and sporadic forms occur. Further clinical distinctions are made according to the onset of the disorder and whether the disease involves only spinal and bulbar motor neurons (spinal muscular atrophy), or both the lower and upper motor neurons. The most common forms of motor neuron disease are spinal muscular atrophy with an incidence of 1 : 6000 to 10 000 newborns, and sporadic amyotrophic lateral sclerosis with a prevalence of 1 : 100 000, which corresponds to an incidence of 1 : 10 000. About 90% of all cases of amyotrophic lateral sclerosis are sporadic. This disorder affects mostly adults, has a typical clinical appearance, is noninherited, still incurable, and invariably fatal. Recent years have seen the identification of a variety of disease mechanisms, which all lead to the degeneration of motor neurons. Thus, pharmacological approaches that can interfere with degeneration and cell death can now be developed but are complicated by the fact that such therapeutic strategies might work only in subpopulations and not in all patients with this disease.

1
Neuronal Cell Death: Implications for Motoneuron Disease

Motor neurons were among the first populations of neurons that were seen to undergo cell death to a significant extent during development. Such observations, made in the first part of the twentieth century, led to the concept of programmed cell death in the nervous system and to the proposal that target-derived neurotrophic

factors play a key role in the maintenance of motor neuron survival and function during the development and in the adult. In the meantime, a variety of such neurotrophic factors have been characterized on the molecular and functional levels, and genetic analysis of patients with motor neuron disease has revealed that only very few patients show mutations or deletions in the genes for these factors. In such cases, it has been found that mutations in the genes for CNTF (ciliary neurotrophic factor) and LIF (Leukaemia inhibitory factor) might act as modifiers, but not as the cause for the disease. Both in patients and in a mouse model of familial ALS, homozygous inactivation of the CNTF gene leads to earlier disease onset. However, the cause of the disease might lie in other functional defects that are not related to the expression of neurotrophic factors or their effects on motor neurons. Studies with mouse models have shown that overexpression of mutant forms of Superoxide dismutase-I (SOD-I) leads to motor neuron disease and that mutations in genes that are relevant for axonal transport and maintenance of axonal processes seem to be more relevant than mechanisms that regulate survival and/or cell death of the motor neuron cell bodies. This indicates that the cell death of motor neuron cell bodies in motor neuron disease might be a secondary phenomenon that occurs as a consequence of axonal defects. On the other hand, a mouse model in which the gene for bcl-2, a key regulator of cellular survival, has been inactivated shows postnatal degeneration of motor neurons. Moreover, cell culture models have revealed that motor neurons derived from SOD-I transgenic mice are much more sensitive to pro-apoptotic stimuli such as Fas ligand and others. In addition, the induction of cell death after binding of excess glutamate to AMPA receptors is also discussed as a cause of motor neuron cell death in motor neuron disease (MND).

Movement of higher organisms requires nerve cells for coordination. The anatomy and physiology of the motor neurons innervating the skeletal musculature is highly conserved from lower to higher vertebrate species – much more than in other parts of the nervous system that are responsible for higher brain functions which are, for example, defective in Alzheimer's disease and other form of dementia. Thus, motor neurons appear as an ideal subject to study neurodegenerative mechanisms in cell culture models, animal models, and to compare these findings with clinical data.

2 Motor Neurons: Anatomical Aspects

The skeletal musculature is innervated by motor neurons located either in the ventral part (the ventrolateral horn) of the spinal cord or in the brain stem motor nuclei. The cell bodies of these motor neurons reside within the central nervous system, and the axons project into the nerves of the peripheral nervous system to the muscle, which they innervate. Many of these axons reach considerable length. For example, the cell bodies of motor neurons innervating the foot muscles are located in the lumbar spinal cord; thus, the total length of these cells can exceed 1 m in humans. This specific property identifies motor neurons as being among the most highly organized cell types of higher organisms.

The transmitter synthesized by motor neurons and released at the motor end plate, the specific synapses in skeletal

muscle, is acetylcholine. Cholinergic neurotransmission to muscle cells is highly conserved from lower organisms to man.

Spinal motor neurons are in contact with many other cell types, including other neurons, in particular, interneurons of the spinal cord, astrocytes, oligodendrocytes in the central nervous system, and Schwann and muscle cells in the periphery. They receive synaptic inputs both from proprioceptive sensory neurons and from spinal interneurons. These interneurons receive input from upper motor neurons. The upper motor neurons can also project directly to spinal motor neurons. Interneurons in the spinal cord are responsible for spinal reflexes, and also coordinate patterns for coordinated motor neuron innervation, which are responsible for movement. Neurons of the cortical spinal tract (upper motor neurons), projecting to spinal and brain stem motor neurons either directly or via interneurons, are important for voluntary movement. The cell bodies of these upper motor neurons are located in the brain in specific regions of the precentral cortex. These cells are also affected by degenerative changes in various forms of human MND, in particular, in amyotrophic lateral sclerosis (ALS).

3
Clinical Appearance and Pathology

3.1
Inherited forms of MND: the Identification of Gene Defects, Animal Models, and Disease Mechanisms

In the case of inherited MND, more than 50 different forms have been distinguished on the basis of clinical parameters. The underlying gene defects for the major forms of MND in childhood, classical SMA, and spinal muscular atrophy with respiratory distress (SMARD) have been identified, and the gene defects for three out of five forms of familial amyotrophic lateral sclerosis are known. Furthermore, an increased size of a polymorphic triplet (CRG repeat) in the coding region of the androgen receptor gene on the X chromosome has been identified as the cause of X-linked spinal muscular atrophy. This defect leads to progressive muscle weakness and atrophy, as well as to gynecomastia and reduced fertility in affected males. Although only about 10% of adult onset ALS is familial, this group appears to be heterogeneous. The first identified gene defect that accounts for about 10 to 20% of familial ALS are point mutations in the gene for Cu^{2+}/Cn^{2+}-dependent superoxide dismutase (SOD-I). More than 50 different mutations in this gene have been identified most of them leading to autosomal dominant ALS, but there are also mutations in this gene that cause an autosomal recessive form of the disease. Clinically, there seems to be no clear correlation between disease onset or severity with specific mutations in the *SOD-I* gene, suggesting that modifiers exist.

However, some types of mutations in the *SOD-I* gene are prone to cause severe and rapid course of the disease, for example, the A4V mutation, whereas others, that is, E21G, G37R, D90A, G93C, G93V, I104F, L144S, and I151T are generally associated with survival times of more than 10 to 15 years. The D90A mutation is generally inherited as a recessive trait, and this mutation exists also as a polymorphism in the Scandinavian population. There are studies suggesting that about 100 000 heterozygous carriers for the D90A polymorphism exist in Finland alone.

The gene for another autosomal recessive form of ALS, which is coupled to mutations in the Chr2q33 region, has been identified as a mutation in *Alsin2*. This gene encodes for a GTP exchange factor, which resembles other RanGEF or RhoGEF molecules. This finding appears interesting, as such factors appear as the major signal mediators for extracellular regulators of axon growth and possibly also for axon maintenance. The gene for another rare autosomal dominant form of juvenile ALS (ALS 4) has been recently identified on chromosome 9q34 as missense mutation (T3I, L389S, and R2136H) in the senataxin gene. This gene encodes for a protein of more than 300 kD. The function of this protein is still unknown, but it contains a DNA/RNA helicase domain with homology to human IGHMBP2 that has been shown to be mutant in a specific form of spinal muscular atrophy with respiratory distress (SMARD 1). Interestingly, mutations in senataxin lead to an early form of ALS and patients normally fall ill at an age >25 years. This suggests that some forms of SMA and juvenile ALS are pathophysiologically related. Rapid progress is currently made in the identification of other forms of familial ALS.

A major breakthrough in understanding the cause of motor neuron disease in children was the identification of the underlying gene defect in classical SMA. Familial forms of lower motor neuron degeneration are among the most common genetic causes of death in children, leading to muscle weakness and respiratory failure within a few years after birth in the most severe forms. According to their severity, the autosomal inherited disorders are classified as spinal muscular atrophy type I (most severe form, also named Werdnig–Hoffmann disease), type II and type III (Kugelberg–Welander disease). All three forms map to chromosome 5 (5q13). The responsible gene defect for all three forms of SMA type is located in the survival motoneuron (*SMN*1) gene, which is positioned within a duplicated region on human chromosome 5q13. Homozygous mutation or lack of the telomeric *SMN* gene is associated with SMA, whereas mutations within the centromeric copy are not responsible for SMA. The number of centromeric copies of the *SMN* gene varies, and various lines of evidence suggest that the number of centromeric *SMN* copies (*SMN2*) and the levels of *SMN*2 gene expression correlate with the severity of the disease. In general, patients with less severe forms (SMA type III, Kugelberg–Welander) have more copies of the centromeric *SMN*2 gene than patients with type I SMA (Werdnig–Hoffmann's disease).

The *SMN* gene is highly conserved among higher organisms. Orthologous genes have been identified in mouse, *drosophila*, *Caenorhabditis elegans*, and the fission yeast *Schizosaccharomyces pombe*. The *SMN* gene duplication is an event that took place just after the divergence of rodents and primates. In humans, both *SMN* copies differ in that full-length protein is predominantly produced from the telomeric *SMN* gene copy. The expression of the truncated SMN2 protein is caused by a C > T exchange at position +6 in exon 7, which leads to skipping of exon 7 encoded regions in about 70% of the protein product derived from the *SMN*2 gene. This specific gene constellation has not been observed in primates, indicating that differential splicing for exon 7 with consequent development of SMA is restricted to humans. Point mutations in the telomeric *SMN* gene are only observed in 5% of

SMA patients. Hot spots of mutational events are found in exon 6 and the 5′ part of exon 7. These regions code for the *SMN* homodimerization domain.

In the past two years, many efforts have been made both in the identification of SMN-associated proteins and the analysis of cellular functions of SMN and SMN-associated proteins. The *SMN* gene is ubiquitously expressed and encodes for a protein of 294 amino acids, which is found both in the nucleus within specific structures called *gemini of coiled bodies* (gems) and in the cytoplasm, both in the cell body and in axons. A group of interacting proteins, which are colocalized in the nucleus includes gemin-2, gemin-3/gp103, a putative DEAD box RNA helicase, gemin-4, which interacts with gemin-3, but not directly with SMN, profilin, fibrillarin, and spliceosomal U snRNPs of the Sm class. The so-called SMN complex also includes other proteins called gemin-5 – 7. The SMN complex is involved in the biogenesis of spliceosomal U snRNPs. U snRNPs are involved in pre-mRNA splicing in the nucleus. Antibodies raised against gemin-2 and SMN interfere with binding of SMN to U snRNAs. This indicates that the SMN complex is essential for the assembly of U snRNPs. Indeed, gene knockout of *Smn* and *gemin-2* in mice is embryonic lethal at early stages of development before blastocysts implant in the uterus. In contrast to humans, the *SMN* gene is not duplicated in mice.

U snRNP assembly is essential for all types of cells and not only for motor neurons. Therefore, the specificity of a disease process for motor neurons remains obscure. The question of why reduced amounts of ubiquitously expressed SMN protein specifically cause degeneration of motor neurons without affecting other somatic cell types is still not fully answered. One explanation could be that SMN, in addition to its reported function, exhibits motor neuron–specific functions that are not relevant for other cells. It is not clear whether all SMN protein is bound to gemin-2 *in vivo*, and evidence from immunohistochemistry indicates that Smn – but not other proteins of the Smn complex – is enriched in axon terminals. In this specific part of motor neurons, the Smn protein seems associated with the hnRNP R protein. The interaction of SMN with hnRNP R requires the exon 7 encoded domains. A complex of SMN and hnRNP R is associated with the 3′-untranslated region of the β-actin mRNA, and mouse models of SMA, which exhibit reduced levels of Smn protein show reduced accumulation of β-actin mRNA in axon terminals. This correlates with β-actin protein levels in axon terminals and reduced axon growth in cultured motor neurons from Smn-deficient mice. Surprisingly, survival of isolated motor neurons with reduced SMN levels is not altered in comparison to wild-type controls. This indicates that the degeneration of motoneurons in the mouse model for spinal muscular atrophy might be a consequence of axonal defects. Lack of motor neuron function and degeneration of motor endplates and terminal axon also affects the capacity of these motor neurons to take up neurotrophic factors. Retrograde transport of these ligands together with their receptors to the cell bodies is reduced, and this should result in degeneration of motor neurons, which is similar to that in mouse models that lack receptors for neurotrophic factors such as gp130, LIFR-β, or CNTFR-α.

In patients with sporadic motor neuron disease, mutations were identified in the

gene coding for heavy neurofilaments subunit. These mutations lead to alterations in the KSP-region, a domain with lysine-serine-proline-repeats, which is known to form the side-arm structures that cross-link the neurofilaments. This finding is particularly interesting in the light of neurofilaments pathology observed in ALS patients. However, such mutations are rare and only found in about 1% of patients with sporadic ALS.

Mutations were also found in the gene for dynactin, which points to a role by axonal transport in the pathophysiology of MND. Transgenic mice in which individual components of the axonal transport machinery are inactivated show classical signs of motor neuron disease. Also, the identification of the underlying gene defect in the mouse mutant progressive motoneuron disease (pmn) points to the relevance of axonal transport for the pathophysiology of MND. A point mutation in the tubulin-specific chaperone E gene is responsible for the motor neuron disease in this mouse model. Interestingly, disease onset and progression can significantly be modulated by pharmacological treatment with the neurotrophic factor ciliary neurotrophic factor (CNTF). This indicates that the role of neurotrophic factors is not restricted to the promotion of neuronal survival, but extends to axonal maintenance, regulation of microtubule synthesis, and regulation of axonal transport.

Interestingly, the androgen receptor, the Cu^{2+}/Cn^{2+}-dependent superoxide dismutase, and neurofilament genes are widely expressed and little is known about what makes the disease process specific to motor neurons. It can be speculated that generalized disorders might manifest themselves as MND, possibly because motor neurons are more vulnerable or sensitive to any of these defects than other cell types. Moreover, these examples demonstrate that different, unrelated gene defects have a similar clinical appearance resembling typical signs of MND indicating that MND cannot be considered to be a homogeneous disorder.

3.2 Basic Pathological Findings of ALS

As a characteristic feature of ALS (albeit with significant variation from case to case), loss of motor neurons occurs. The remaining cells appear shrunken, dark, with basophilic cytoplasm and pyknosis of the nuclei. Shrinkage of the cells, which precedes neuronal death, occurs in both upper and lower motor neurons. It has been suggested that the degeneration of lower motor neurons is a consequence of the loss of corticobulbar neurons. However, morphometric comparison of the loss of upper and lower motor neurons does not support such hypotheses. The morphological characteristics of motor neuron cell death such as disintegration of the Golgi complex, loss of Nissl structure, shrinkage, any particular pyknosis of the nuclei, very much resemble the changes occurring during apoptosis. However, experimental proof is poor that motor neuron cell death in all forms of MND is apoptotic. Several studies with postmortem ALS spinal cord have identified TUNEL-positive motor neurons. However, it is not clear whether classical apoptotic mechanisms are responsible for cell death, and when they are activated during the course of the disease. In particular, it remains to be demonstrated that apoptotic mechanisms are disturbed at early stages before first symptoms of the disease become clinically apparent.

Increasing evidence suggests that the disease process in ALS, the most common form of MND, might not be limited to

motor neurons. There is also evidence of abnormal rates of degeneration in sensory neurons, but there is little doubt that the motor system is the site of predominant manifestation of the disease. Impairment of axonal transport might underlie many of the pathological characteristics. Inclusion bodies of several types have been described in the perikarya of motor neurons, such as basophilic inclusion, which are thought to consist mainly of RNA, eosinophilic Bunina bodies that resemble autophagosomes, and hyaline-inclusion, originally thought to be characteristic of familial ALS. This points to a role of abnormal protein degradation via the proteasome system, and experimental evidence mainly from cultured neurons suggests that increased proteasome activity can interfere with the pathological mechanisms in some models of MND, such as cell lines overexpressing mutant forms of SOD-I. Hyaline inclusions consist of densely packed phosphorylated neurofilaments, and it is believed that an impairment of slow anterograde axonal transport is responsible for their accumulation in the perikarya of motor neurons. Thus, defects in axonal transport and in protein degradation could play together in the pathophysiology of MND. The concept of impaired axonal maintenance is further supported by the finding that more myelinated nerve fibers are detectable in proximal than in distal parts of motor nerves of ALS patients. Failure of motor neurons to maintain distal structures such as the endplates appears as an early stage in the disease process. Evidence of the degeneration of endplates in early stages of the disease has been derived both from electrophysiological (spontaneous muscle fiber activity) and morphological studies. The disease then progresses to degeneration of the axons in a dying-back fashion, finally leading to degeneration of the cell bodies in the spinal cord and brain stem. Clinical symptoms are expected to become apparent only after it has become impossible to compensate for the loss of functional motor neurons by collateral sprouting and rearrangement of motor units. This coincides with the stage at which motor neuron nerve terminals are affected. Thus, clinical symptoms can precede the degeneration of cell bodies.

4 Sporadic MND: The Pathophysiological Mechanisms Are Still Obscure

The majority of cases (>90%) of MND occur sporadically, precluding the analysis of underlying gene defects by classical molecular genetic analysis. Candidate gene approaches have been conducted in only a few places, and it is premature to state whether the identified mutations may be responsible for the development of the disease in general or only in small subsets of patients. For example, gene defects of heavy neurofilament subunits have been observed only in a few patients (see Sect. 3.1). However, these data do not tell us whether spontaneous gene defects are pathogenic, or whether epigenetic influences (such as intoxication, infection, autoimmune disorders, or any combination of such mechanisms) might be responsible for the disorder. Moreover, defects in the genes coding for neurotrophic factors or their receptors such as CNTF or CNTFR-α have been observed in healthy individuals and in patients with neurological disorders. Targeted disruption of such genes in mice leads to slow degeneration of motor neurons and changes in motor axons in the case of ciliary neurotrophic factor gene knockout. However, symptoms of muscle weakness are only

very mild. Mice with gene defects in the CNTFR-α or LIFR-β gene die around birth and thus cannot be considered as a model for human MND. The severe discrepancy between gene activation of ligands such as CNTF or LIF and for receptor components such as CNTFR-α or LIFR-β suggests that defects in the genes for neurotrophic factors are compensated for by other members of the same gene family. In this case, MND would only occur when such compensatory mechanisms fail or when several kinds of degenerative influences act together, leading to such severe cellular dysfunction of motor neurons that compensatory mechanisms are not sufficient to prevent functional deficits.

There have been suggestions that many pathological mechanisms, some of them listed in Sect. 3, may be responsible for MND, but convincing evidence that interference with any of them results in an effective treatment of MND is still lacking. The only effective drug currently used for treatment of MND, Riluzole, interferes with a variety of mechanisms including glutamate toxicity, function of ion channels, and possibly additional intracellular signaling pathways, which modulate the function and structural maintenance of motor neurons. The conclusion that can be drawn from this situation is that basic mechanisms of MND are far from being understood.

4.1 The Autoimmune Hypothesis of Motoneuron Disease

There are many reports of abnormal immunological findings in MND patients. One of the most intriguing is the toxicity of serum from ALS patients to cultures of isolated motor neurons. Moreover, abnormalities in the composition of subgroups of peripheral blood lymphocyte subsets have been described; for example, there is a decrease in mature T lymphocytes (CD3-positive cells) and a shift in the CD4/CD8-ratio in favor of CD8-positive cells, as well as an increase in natural killer cells (CD57 and CD16 positive cells). However, these findings have not been confirmed by recent studies. Another line of evidence has been derived from studies searching for autoantibodies directed at nerve antigens. Such studies have revealed antibodies against gangliosides (GM1), acetylcholine esterase, and L-type or P-type Ca^{2+}-channels. Such findings are in apparent contrast to many unsuccessful efforts to treat ALS by immunosuppressive strategies. Even total body lymph node irradiation was without effect. Some patients with high titers of anti-GM1-antibodies diagnosed as ALS exhibit peripheral nerve multifocal conduction block and are now considered to reflect a subgroup of chronic multifocal inflammatory neuropathy rather than degenerative MND. It now appears conceivable that in the course of MND epitopes on motor neurons that usually are hidden to the immune system become exposed, they allow secondary autoimmune mechanisms to come into play and add to the degenerative changes in motor neurons. Even if autoimmune mechanisms are not primary events, they could contribute to the clinical picture of MND by accelerating the functional impairment of the remaining motor neurons.

4.2 Toxic Lesions of Motor Neurons

Intoxication with ingredients of cycad plants is thought to be responsible for the ALS-Parkinsonism-dementia complex found in Chamorro populations in the

Western Pacific Islands. Since, however, the disease has been reported to occur up to several decades after exposure to the toxic agents, it is not clear whether the intoxication itself is responsible for the disease or whether it induces a long-lasting increase in the vulnerability of motor neurons to other agents. It has been suggested that *β-N*-methylamino-L-alanine (LBMAA) or the increased concentrations of glutamate and abnormal excitatory amino acids detectable in the cerebrospinal fluid of ALS patients could induce motor neuron degeneration via glutamate receptors. Glutamate toxicity via *N*-methyl-D-aspartate (NMDA) receptors is a common degenerative mechanism in other disorders of the nervous system such as stroke and CNS-trauma. However, clinical trials with NMDA-receptor antagonists such as dextromethorphan have not shown efficiency in ALS patients. However, there is evidence *in vitro* that AMPA receptors, in particular, Ca^{2+} influx through this receptor, are responsible for glutamate toxicity in motor neurons. EAAT2 variant transcripts have been found in the brain of ALS patients in combination with reduced EAAT2 protein levels. This is thought to cause elevated glutamate levels at the synapses that are located on the dendrites of bulbar and spinal motor neurons. Further evidence of glutamate toxicity in ALS is based on the efficacy of Riluzole, the only available drug for treatment of ALS up to now. This drug candidate is thought to act as a presynaptic inhibitor of glutamate release, but there is also evidence for other mechanisms of action. The role of glutamate toxicity in motor neurons has also been demonstrated after viral infection. Neuroadapted Sindbis virus mediates motor neuron death indirectly via an excitotoxic mechanism involving glutamate and calcium-permeable AMPA receptors.

5 Oxidative Stress: Is Motoneuron Degeneration in Familial ALS Due to Loss or Gain of Function of the Enzyme Superoxide Dismutase?

Mutations within the Cu/Zn-dependent superoxide dismutase (*SOD-I*) gene on chromosome 21 in about 2 to 5% of ALS patients has led to reinvestigations of all the observations and speculations that oxidative damage may play a pathogenic role in MND. SOD-I had already been the focus of interest in the context of the Alzheimer-like neurodegenerative changes associated with Down Syndrome. In the search for genes responsible for degenerative changes, patients with partial duplication of chromosome 21 translocated to other chromosomes were analyzed, and the 21q22 region has been found to be critical. Among the roughly 100 genes encoded in this region, the *SOD-I* gene was considered to be a good candidate, the duplication and consequent overexpression of which might be responsible for the degenerative changes observed in Down Syndrome: Indeed, overexpression of SOD-I in neural cell lines results in impaired neurotransmitter metabolism.

To obtain information on a possible role of increased SOD-I activity in neurodegenerative processes in more detail, transgenic mice overexpressing the human *SOD-I* gene under its homologous promoter were produced. These transgenic mice show pathological changes in motor neurons, such as withdrawal and destruction of terminal axons at the neuromuscular endplates and development of multiple small terminals in comparison to normal age-matched mice. Surprisingly, gene knockout of the *SOD-I* gene did not lead to MND, indicating that this enzyme is dispensable for maintenance of motor

neurons, and its function is probably fully compensated by other enzymes such as SOD-II.

In the meantime, more than 60 mutations have been reported in patients with autosomal dominant ALS (ALS-1). These are widely scattered over the gene with hot spots in exon 4 and 5. Mutations of the conserved glycine at position 93 are quite abundant, and the A4V mutation, which is in particular found in North American families, generally leads to a severe course of ALS. The D90A mutation is quite abundant in Sweden and Finland, and in contrast to the other mutations, it is inherited as a recessive trait. 1 to 2% of the population in Scandinavia is heterozygous for the D90A allele. In Finland, about 100 000 persons are estimated to be carriers of this mutation. In general, patients with the D90A homozygous mutation show a relatively mild form of ALS with an average survival time of more than 10 years. Surprisingly, the age of onset does not differ from very severe forms such as the A4V form, indicating that other factors determine the age of onset, which act independently from those that determine survival time. Enzyme activity of SOD-I is only reduced by about 50% in most patients with familial ALS, as expected for autosomal dominant inheritance, leaving one allele of the *SOD-I* gene intact. It was suggested that the mutations lead to decreased enzyme activity either by causing an allosteric effect on the active center of the enzyme or by interfering with the formation of dimers, which is the form in which the enzyme naturally occurs in cells.

Mice with transgenic overexpression of mutant SOD-I generally show protein aggregates and other alterations in motor neurons, which become clearly detectable when the disease is clinically apparent. However, it is not fully understood whether such aggregates are responsible for the disease. Alternatively, it has been suggested that the inability of mutant SOD-I to bind copper might lead to accumulation of protein complexes and toxicity of free copper in cells and thus to neurodegeneration. Acquisition of copper by SOD-I requires a specific copper chaperone, a member of the family of metallochaperones, which regulate intracellular trafficking of metal ions. However, inactivation of this chaperone in each of the three most frequently used mouse models for ALS indicates that the disease is unaffected by elimination of copper loading. This argues against the possibility that primary toxicity arises from aberrant copper within the neurons. Morphological changes in mitochondria within motor neurons were observed as an early feature both in SOD-I^{G37R} and SOD-I^{G93A}–transgenic mice. This correlates with reduced mitochondrial function at least in SOD-I^{G93A}. Interestingly, mutant SOD-I^{G93A} has been detected within spinal cord mitochondria, and recent studies have shown that the accumulation of mutant SOD-I in mitochondria is also found in other ALS-linked mutations, and that these changes lead to damage of mitochondria-bound proteins. Therefore, it has been suggested that motor neuron–specific damage of mitochondria might be the reason for the selective toxicity of mutant SOD-I in MND.

6 Disintegration of Axonal Transport: Implications for MND

Several mouse mutants have been identified or generated in which disturbance of components of the cytoskeleton, of microtubuli, or of components of the axonal transport machinery lead to typical signs

of MND. Recently, missense point mutations in the cytoplasmic dynein heavy chain have been identified as the cause for motor neuron disease in *Loa*- and *CIra1*-mutant mice. Dynein is part of the motor complex for protein and vesicle movement along microtubules in the minus-end direction and thus an important component for retrograde axonal transport. Interestingly, the defect does not only lead to reduced retrograde transport but also to progressive postnatal loss of motor neurons. Whereas motor neuron loss is not detectable at an age of 3 months, about 50% of the motor neurons are lost at 16 to 19 months. Mutations in the p150 subunit of the transporter protein dynactin have been identified in a family with autosomal dominant progressive MND. Dynactin binds both to microtubules and to dynein, and thus acts as an essential component of the axonal transport machinery, which mediates binding of transport vesicle and transport proteins to microtubules. Mice in which dynamitin, the p50 inhibitory subunit of the dynactin complex, is overexpressed also show late onset MND.

Other widely used mouse models are *Pmn* mice, which develop a rapidly progressive form of MND between the third and fifth postnatal weeks. Histologically, the mice exhibit atrophy and loss of motor axons and at a later stage loss of motor neuron cell bodies in the spinal cord and brain stem motor nuclei. In addition, neurons in the red nucleus degenerate. The axonal projections from this nucleus to the spinal motor neurons – the corticospinal tract – are functionally linked to the corticospinal tract in rodents and contribute significantly to the innervation of spinal motor neurons. It thus functionally resembles the upper motor neuron input to lower motor neurons in higher vertebrates. As the cause of the disease, a single amino acid exchange in the most C-terminal part of the tubulin-specific chaperone E (TBCE) protein was identified. The resulting protein appears less stable, but it could also be that translocation of the protein to axons, the function of the protein or interactions with other functionally relevant proteins are affected. These hypotheses need additional experimental data.

In summary, these data suggest that axonal transport plays an essential role in motor neuron maintenance, and future research on the pathophysiology of MND in humans will focus on the question whether defects in axonal transport are an early or late phenomenon in the disease.

The cytoskeleton of most animal cells contains a variety of elements also including thin actin filaments and intermediate filaments in addition to the microtubules. The β-actin mRNA is one among a small population of mRNAs, which are axonally transported. The functional relevance of axonal transport has been a point of debate for a long time. In contrast to dendritic transport, it has been argued that it is not clear whether the translocalized mRNAs are translated within the axonal compartment. Axonal transport of β-actin mRNA has been studied in much detail. A specific 3′-UTR region called *zip code* has been identified within β-actin mRNA that confers axonal localization. Proteins that bind to the zip-code region of β-actin mRNA have been named zip-code binding protein 1 (ZBP1) and ZBP2. It has been found that the formation of a particle involving β-actin mRNA and ZBP1 is necessary for β-actin translocation to growth cones. The translocation of the mRNA is increased by neurotrophic factors such as neurotrophin-3. In motoneurons, axonal terminals contain polyribosomes, which can be detected with specific antibodies against large subunit ribosomal RNA or

by electron microscopy. The SMN protein, which is deficient in spinal muscular atrophy, binds to hnRNP R in the axonal compartment, and it has been shown that a complex of SMN and hnRNP R also interacts with the zip-code domain of β-actin mRNA. Motor neurons that lack Smn show reduced accumulation of β-actin mRNA in axon terminals. This correlates with highly reduced levels of β-actin protein in the axon terminals and reduced axon growth of isolated motor neurons in cell culture. In contrast, overexpression of SMN and/or hnRNP R in neuronal cell lines enhances neurite outgrowth, thus indicating that the translocation of β-actin mRNA and the local translation of this protein in axon terminals play a major role in axon growth and maintenance. Also in other animal models such as zebrafish, knockdown of SMN primarily leads to axonal disturbance and not to motor neuron cell death. Thus, spinal muscular atrophy might be considered as a disease that is caused by defective axonal translocation of β-actin mRNA. It will be interesting to know whether similar mechanisms also account for subtypes of adult onset MND. First indications have been obtained by the identification of mutations in senataxin, a protein with a DNA-helicase domain that is mutated in a rare form of inherited amyotrophic lateral sclerosis.

7 Identification of Neurotrophic Factors and Cytokines that can Regulate Survival and Function of Motor Neurons

Survival of embryonic spinal motor neurons depends on neurotrophic factors obtained from skeletal muscle and glial cells. Such factors are taken up by motor neurons by means of specific receptors and transported to motor neuron cell bodies where they act on specific parameters. In this way, these factors regulate apoptosis, transmitter synthesis, and other aspects of motor neuron function. Developing embryos and cultured embryonic spinal motor neurons have been used to identify neurotrophic factors and cytokines capable of supporting the survival of developing motor neurons. Positive effects have been observed for members of the ciliary neurotrophic factor (CNTF) family, for brain-derived neurotrophic factor (BDNF), neurotrophin (NT)-3 and -4, insulin-like growth factor (IGF) and members of the glial-derived neurotrophic factor (GDNF) family. These factors differ in their tissue distribution and temporal expression patterns, and gene knockout experiments have indicated that their modes of action might overlap with other members of the same gene family and probably of others. Interestingly, neurotrophic factors can interfere with the pathophysiology of motor neuron disease caused by different mechanisms. Positive effects are observed with IGF-I in mice that overexpress SOD-I. Pmn mutant mice show improved survival and motor performance after treatment with CNTF. This indicates that these factors can interfere with distinct pathophysiological mechanisms that have been identified for MND. The molecular mechanisms of these actions appear to be important for development of therapy and could lead the way to the identification of new small molecules that are active on the same targets.

8 Future Prospects for Treatment of MND

A large number of clinical trials with a variety of drugs and therapeutic strategies

have been carried out in the past in the search for an effective treatment of MND. Among them were plasmapheresis (with the intention of removing toxic or autoaggressive molecules), drugs such as corticosteroids, corticotropin, cyclosporine A, and cyclophosphamide, which are known to be effective in autoimmune diseases. None of these trials were successful. Similar negative effects were observed in trials with growth hormone and thyrotropin-releasing factor.

A driving force behind some of the current therapeutic trials for the treatment of MND is the assumption that degeneration of motor neurons might be caused by mechanisms similar to those responsible for the death of other neurons of the central nervous system after glutamate exposure. Therefore, considerable attention has focused on clinical trials with glutamate antagonists and glutamate release inhibitors such as Riluzole and other drugs. Trials with free radical scavengers such as acetylcysteine, which are thought to reduce the generation of free radicals, a pathomechanism that has been implicated by mutations in the *SOD-I* gene in familial ALS, have not been successful so far.

Clinical trials with neurotrophic factors such as CNTF and BDNF have also failed. Among many potential reasons are, in particular, adverse pharmacokinetics. Thus, the factors are not available to the motor neuron cell bodies residing in the spinal cord. Alternatively, it could be that the receptors desensitize within a short time to neurotrophic factors, so that only short-time effects occur, which are not sufficient for a therapeutic effect of these drug candidates. Indeed, even promising trials with mouse models such as *Pmn* mice or SOD-I transgenic mice generally focus on treatment periods of only a few weeks, whereas treatment of patients would require positive drug effects over several years. Nevertheless, the positive effects of IGF-I, BDNF, and CNTF in various mouse models of motor neuron disease are a good basis for further research and the development of new therapeutic strategies that might help promote long-term effects in the degenerating motor neuron.

Another problem that needs to be solved is the apparent heterogeneity of pathomechanisms, in particular in familial and probably also sporadic ALS. It could well be that drug candidates that interfere with some pathomechanisms are without any effect in patients suffering from MND due to another defect. Therefore, the genetic and molecular basis of disease needs to be better understood, and techniques for genotyping and molecular diagnosis have to be developed so that the pathomechanism of MND can be characterized for individual patients, even in patients with sporadic ALS. Potential positive or negative effects for the various drug candidates can only be tested on this basis.

This problem seems to be easier for patients with various forms of spinal muscular atrophy. The molecular basis of classical autosomal recessive SMA, which is caused by a mutation of the *SMN* gene, is known and can be easily tested and characterized. Thus, prerequisites exist to develop drugs on a rational basis. The same is also true for other forms of childhood spinal muscular atrophy such as SMARD and for adult onset X-linked spinal muscular atrophy (Kennedy's disease), which is caused by a polyglutamine-expansion in the androgen receptor gene. Multiple lines of evidence indicate that this mutation leads to a gain of pathogenic function, which requires the translocation of the receptor after binding of testosterone. Thus, anti-androgens

have been suggested as drug candidates, and first trials with candidates such as leuprorelin showed promising results by reversing the behavioral and histopathological phenotypes in a mouse model. It is to be expected that clinical trials based on such results will be initiated soon, and new drug candidates will be developed with cell culture and mouse models, which can be developed once the molecular and genetic basis of the disease is known.

See also Motor Neuron Diseases: Cellular and Animal Models.

Bibliography

Books and Reviews

Bruijn, L.I., Miller, T.M., Cleveland, D.W. (2004) Unraveling the mechanisms involved in motor neuron degeneration in ALS, *Annu. Rev. Neurosci.* **27**, 723–749.

Chou, S.M., Norris, F.H. (1993) Amyotrophic lateral sclerosis: Lower motor neuron disease spreading to upper motor neurons, *Muscle Nerve* **16**, 864–869.

Cleveland, D.W., Rothstein, J.D. (2001) From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS, *Nat. Rev. Neurosci.* **2**, 806–819.

Fischbeck, K.H., Lieberman, A., Bailey, C.K., Abel, A., Merry, D.E. (1999) Androgen receptor mutation in Kennedy's disease, *Philos. Trans. R. Soc. Lond B Biol. Sci.* **354**, 1075–1078.

Jablonka, S., Wiese, S., Sendtner, M. (2004) Axonal defects in mouse models of motoneuron disease, *J. Neurobiol.* **58**, 272–286.

Katsuno, M., Adachi, H., Tanaka, F., Sobue, G. (2004) Spinal and bulbar muscular atrophy: ligand-dependent pathogenesis and therapeutic perspectives, *J. Mol. Med.* **82**, 298–307.

Smith, R.A. (1992) *Handbook of Amyotrophic Lateral Sclerosis*, Marcel Dekker, Inc., New York.

Thoenen, H., Sendtner, M. (2002) Neurotrophins: from enthusiastic expectations through sobering experiences to rational therapeutic approaches, *Nat. Neurosci.* **5**(Suppl.), 1046–1050.

Primary Literature

Bommel, H., Xie, G., Rossoll, W., Wiese, S., Jablonka, S., Boehm, T., Sendtner, M. (2002) Missense mutation in the tubulin-specific chaperone E (Tbce) gene in the mouse mutant progressive motor neuronopathy, a model of human motoneuron disease, *J. Cell Biol.* **159**, 563–569.

Chen, Y.Z., Bennett, C.L., Huynh, H.M., Blair, I.P., Puls, I., Irobi, J., Dierick, I., Abel, A., Kennerson, M.L., Rabin, B.A., Nicholson, G.A., Auer-Grumbach, M., Wagner, K., De Jonghe, P., Griffin, J.W., Fischbeck, K.H., Timmerman, V., Cornblath, D.R., Chance, P.F. (2004) DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4), *Am. J Hum. Genet.* **74**, 1128–1135.

Hafezparast, M., Klocke, R., Ruhrberg, C., Marquardt, A., Ahmad-Annuar, A., Bowen, S., Lalli, G., Witherden, A.S., Hummerich, H., Nicholson, S., Morgan, P.J., Oozageer, R., Priestley, J.V., Averill, S., King, V.R., Ball, S., Peters, J., Toda, T., Yamamoto, A., Hiraoka, Y., Augustin, M., Korthaus, D., Wattler, S., Wabnitz, P., Dickneite, C., Lampel, S., Boehme, F., Peraus, G., Popp, A., Rudelius, M., Schlegel, J., Fuchs, H., Hrabe, d.A., Schiavo, G., Shima, D.T., Russ, A.P., Stumm, G., Martin, J.E., Fisher, E.M. (2003) Mutations in dynein link motor neuron degeneration to defects in retrograde transport, *Science* **300**, 808–812.

Howland, D.S., Liu, J., She, Y., Goad, B., Maragakis, N.J., Kim, B., Erickson, J., Kulik, J., DeVito, L., Psaltis, G., DeGennaro, L.J., Cleveland, D.W., Rothstein, J.D. (2002) Focal loss of the glutamate transporter EAAT2 in a transgenic rat model of SOD1 mutant-mediated amyotrophic lateral sclerosis (ALS), *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1604–1609.

Kaspar, B.K., Llado, J., Sherkat, N., Rothstein, J.D., Gage, F.H. (2003) Retrograde viral delivery of IGF-1 prolongs survival in a mouse ALS model, *Science* **301**, 839–842.

Liu, J., Lillo, C., Jonsson, P.A., Velde, C.V., Ward, C.M., Miller, T.M., Subramaniam, J.R., Rothstein, J.D., Marklund, S., Andersen, P.M., Brannstrom, T., Gredal, O., Wong, P.C.,

Williams, D.S., Cleveland, D.W. (2004) Toxicity of familial ALS-linked SOD1 mutants from selective recruitment to spinal mitochondria, *Neuron* **43**, 5–17.

Nicole, S., Diaz, C.C., Frugier, T., Melki, J. (2002) Spinal muscular atrophy: recent advances and future prospects, *Muscle Nerve* **26**, 4–13.

Puls, I., Jonnakuty, C., LaMonte, B.H., Holzbaur, E.L., Tokito, M., Mann, E., Floeter, M.K., Bidus, K., Drayna, D., Oh, S.J., Brown, R.H. Jr., Ludlow, C.L., Fischbeck, K.H. (2003) Mutant dynactin in motor neuron disease, *Nat. Genet.* **33**, 455–456.

Rossoll, W., Jablonka, S., Andreassi, C., Kroning, A.K., Karle, K., Monani, U.R., Sendtner, M. (2003) Smn, the spinal muscular atrophy-determining gene product, modulates axon growth and localization of beta-actin mRNA in growth cones of motoneurons, *J. Cell Biol.* **163**, 801–812.

Subramaniam, J.R., Lyons, W.E., Liu, J., Bartnikas, T.B., Rothstein, J., Price, D.L., Cleveland, D.W., Gitlin, J.D., Wong, P.C. (2002) Mutant SOD1 causes motor neuron disease independent of copper chaperone-mediated copper loading, *Nat. Neurosci.* **5**, 301–307.

Xu, Z., Cork, L.C., Griffin, J.W., Cleveland, D.W. (1993) Increased expression of neurofilament subunit-L produces morphological alterations that resemble the pathology of human motor neuron disease, *Cell* **73**, 23–33.

Motor Proteins

Charles L. Asbury[1,3] *and Steven M. Block*[1,2]
[1] *Department of Biological Sciences, Stanford University, Stanford, California, USA*
[2] *Department of Applied Physics, Stanford University, Stanford, California, USA*
[3] *Present Address: Department of Physiology and Biophysics, University of Washington, Seattle, Washington, SA*

1 Introduction 577

2 Classic Molecular Motors 578
2.1 Muscle Myosin: Tilting Cross-bridges Drive Contraction 579
2.2 Seeing is Believing: Motility Assays Demonstrate Motor Activity 580
2.3 Kinesin: Intracellular Porter 580
2.4 Ciliary Dynein: The Dark Horse 583
2.5 Processivity Allows Kinesin to Work Alone 584
2.6 Rowers Versus Porters: Duty Ratio Makes a Difference 584
2.7 Molecular Tug-of-war: Applying Force to Individual Motors 586
2.8 Motors Move in Discrete Steps 586
2.9 Different Strokes: Variation Within and Across Motor Families 587
2.10 Fuel Economy and Energy Efficiency 588
2.11 Walk This Way: Processive Mechanoenzymes Move Hand-over-hand 588
2.12 Coordination is Required 589
2.13 The Kinesin Cycle and Working Stroke 590
2.14 Under the Hood, Motors are Still a Mystery 590

3 Nontraditional Molecular Motors 591
3.1 Nucleic Acid Enzymes 591
3.2 More Than a Motor: Multitasking by RNA Polymerase 591
3.3 RNA Polymerase Structure 592
3.4 What Causes Pauses? 592
3.5 ATP Synthase 593
3.6 The Rotary Motor of Bacterial Flagella 594

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

3.7 Polymers that Push and Pull 596
3.8 Microtubule Ends and Dynamic Instability 597
3.9 Motility Assays with Cytoskeletal Filaments 597

4 Conclusion 598

Bibliography 599
Books and Reviews 599
Primary Literature 599

Keywords

Allostery
Pertaining to or involving a change in conformation caused by the attachment of a ligand or substrate.

Mechanoenzyme
A catalytic enzyme that produces motion and force.

Processivity
A measure of the number of catalytic cycles an enzyme undergoes before detaching from the substrate. In the context of motors, processivity is proportional to the average distance over which the enzyme translocates on its filamentous substrate before detaching.

Substrate
A molecule on which an enzyme acts. In the context of molecular motors, either the fuel source (e.g. NTPs), or the force-generating partner (e.g. actin filament).

Working Stroke
A conformational change that occurs during a single round of catalysis, and which drives motion and force production. The working stroke length, or working distance, is the maximal distance associated with the stroke. The working distance can be different from the distance between consecutive attachment sites on the partner filament.

Cellular motions have fascinated biologists during the 400 years since the invention of the optical microscope first allowed them to be seen. Today, we know that motions underlying the most essential processes of life – such as cell division, energy transduction, muscle contraction, DNA replication, transcription, and translation – are generated by molecular motors. A molecular motor is a protein, or a complex of proteins and nucleic acids, that produces motion and force. For fuel,

many molecular motors consume nucleotide triphosphates, breaking an energy-rich phosphate bond to release chemical energy, and then converting this into mechanical work. Other motors tap electrochemical gradients that exist across membranes within bacteria, mitochondria, and chloroplasts. Motor proteins are Nature's nanomachines, and they often function with efficiency that far exceeds the best human-engineered machines.

1 Introduction

Producing motion and force is the primary role of the "classic" molecular motors, myosin, kinesin, and dynein. These *mechanoenzymes* all hydrolyze ATP as a source of energy and drive motion along protein filaments. Myosin generates motion along filamentous actin, and is well known for its role in muscle contraction. The seemingly simple act of flexing ones arm requires $\sim 10^{17}$ myosins working together to slide $\sim 10^{15}$ actin filaments toward one another. Kinesin and dynein move along microtubule filaments. An essential role of kinesin is to haul vesicles across neurons. This can be a 6-day haul since the longest neurons are more than a meter, and vesicle transport proceeds at only $2\,\mu m\,s^{-1}$. Dynein causes the beating of flagella and cilia, such as those lining the lungs, by sliding microtubules past one another.

Besides the classic motors, there are many "nontraditional" motor proteins. In some cases, motion and force production are byproducts rather than a primary function. The main role of DNA and RNA polymerases is to copy and transcribe the genetic code. In order to do so, they move along their nucleic acid templates, sometimes with amazing endurance. A single RNA polymerase molecule can transcribe all 2.5 million bases of the human dystrophin gene in 14 h, at roughly 50 bases per second. Another nontraditional motor is F_1F_0-ATP synthase, which is responsible for replenishing the entire pool of ATP in all cells using energy derived from metabolism. As this enzyme toils, it also spins – it is a rotary motor. A flow of protons causes a shaft within the motor to rotate continuously, and shaft rotation is then coupled to the synthesis of ATP from ADP and phosphate. A third type of nontraditional motor activity is driven by the cytoskeletal polymers, actin, and tubulin. In addition to their roles as structural cables and girders for maintaining cell shape, and as highways for motor proteins to move along, these polymers are themselves dynamic machines that produce force. The leading edges of macrophages and other crawling cells are pushed outward by polymerizing actin filaments. Microtubule depolymerization generates tension that pulls chromosomes apart prior to cell division.

This chapter is a survey of the main classes of molecular motors. It begins with a discussion of the classic mechanoenzymes, myosin, kinesin, and dynein. These motors have been the subject of biophysical research for decades, and our understanding of their function serves as a foundation for the study of other motors. The chapter then turns to nontraditional motors, focusing on a handful of key examples, including nucleic acid enzymes (RNA polymerase), rotary motors

(F_1F_0-ATP synthase, and the bacterial flagellar motor), and protein polymers (actin and tubulin). A central goal of research on motor proteins is to determine how underlying biochemical events, such as ATP hydrolysis, are coupled to mechanical action. Progress toward this goal is chronicled throughout the chapter through description of experiments with classic and nontraditional motors.

2 Classic Molecular Motors

Myosin, kinesin, and dynein are founding members of large families of proteins whose primary function is to generate motion and force. Owing to their structural resemblance, the motors of each family operate in a manner similar to the founding proteins. However, they drive a wide variety of different cellular motions beyond the stereotypical roles of muscle contraction, vesicle transport, and the beating of cilia. There are at least 15 classes of myosin (traditionally denoted with roman numerals I through XV), and only a handful are involved in muscle contraction. Some of the other types are implicated in vesicle budding, cytokinesis, and organelle transport along actin cables. Likewise, kinesin-like proteins and cytoplasmic dyneins are essential for the formation and positioning of the mitotic spindle, chromosome separation prior to cell division, and organelle

Fig. 1 Structures of some classic and nontraditional molecular motors. (a) Muscle myosin consists of two heads connected to a common coiled-coil tail. The heads bind actin and carry ATP hydrolysis activity. A rodlike portion of each head (light gray) functions as a lever-arm, tilting ~70° relative to the remainder of the head (dark gray) upon attachment to an actin filament. The tail promotes bundling of myosin molecules into thick filaments. (b) Kinesin also has two heads, connected through short polypeptides called *neck linkers* to a common coiled-coil stalk. The heads bind microtubules and carry ATP hydrolysis activity. A conformational change of the neck linkers may drive kinesin motion. The tail binds cargo. (c) Each dynein molecule consists of a donut-shaped head with two rodlike structures, the stem and stalk, emanating from the head. The head contains four ATP-binding sites, only one of which is catalytically active. The tip of the stalk binds microtubules. To drive motion, the head and stalk rotate relative to the stem, which attaches to cargo and can also bundle two or three dynein heads together. (d) RNAP is shaped like a claw, which opens to allow a DNA template to enter, and then wraps completely around the DNA during transcription. While transcribing, RNAP separates a portion of the DNA duplex called the *transcription bubble*, and maintains registration of a short section of hybrid RNA:DNA duplex. Nucleotides enter through a channel, leading to the active site, where they are incorporated into the nascent mRNA chain. (e) F_1F_0-ATP synthase consists of two rotary motors, connected to a common shaft, which act as a motor-generator pair. The F_0 portion taps a proton gradient across the inner mitochondrial membrane to drive spinning of the rotor (c_{12}) relative to the stator (ab_2). The F_1 portion sits directly above F_0, and contains a shaft ($\gamma\varepsilon$) that is rigidly fixed to the rotor of F_0, and also a ring $(\alpha\beta)_3$ that is rigidly fixed to the stator of F_0. Spinning of the shaft relative to the ring drives ATP synthesis. The isolated F_1 portion is known as F_1-ATPase because in the presence of ATP it will spin in reverse, catalyzing ATP hydrolysis. (e) Cross-sectional view of the rotary motor of bacterial flagella. The motor core contains a stack of rings (rotor), embedded in the multilayered cell wall that rotates as a single unit about an axis (dashed line) perpendicular to the surface of the bacteria. Rotation is driven by torque-generating units composed of MotA and MotB proteins. The MotA/B complex is anchored to fixed structures (peptidoglycan) within the cell wall. Protons flow through a channel within MotA/B, where protonation and deprotonation of MotB induces conformational changes in MotA, which attaches and detaches from the base of the rotor and drives its rotation.

movement along microtubules. The discussion here centers on the founding proteins, their functional properties, and some of the experiments that uncovered these properties. Particular attention is given to *in vitro* work with single motor molecules.

2.1 Muscle Myosin: Tilting Cross-bridges Drive Contraction

The motor activity of myosin was discovered more than 50 years ago. Electron microscopy revealed that muscle fibers consist of parallel thick and thin filaments that slide past one another during contraction. Tiny structures termed "cross-bridges," connecting laterally between the filaments, were suspected to drive filament sliding. In some images, the cross-bridges projected from the thick filaments at right angles, but in others, they were tilted, depending on tissue preparation conditions. These observations led to the theory, now well established, that cross-bridges drive filament sliding by cyclically attaching to the thin filaments, tilting, detaching, and untilting.

The thick filaments are now known to be bundles of myosin molecules. Each myosin consists of two identical 200-kDa polypeptides, plus two pairs of light chains (20 kDa). The heavy chains fold into twin globular heads connected to a common coiled-coil tail (Fig. 1a). Two light chains bind each head near the head–tail junction. Myosins bundle together by their long tails, and their heads project from the bundles, forming the cross-bridges that drive

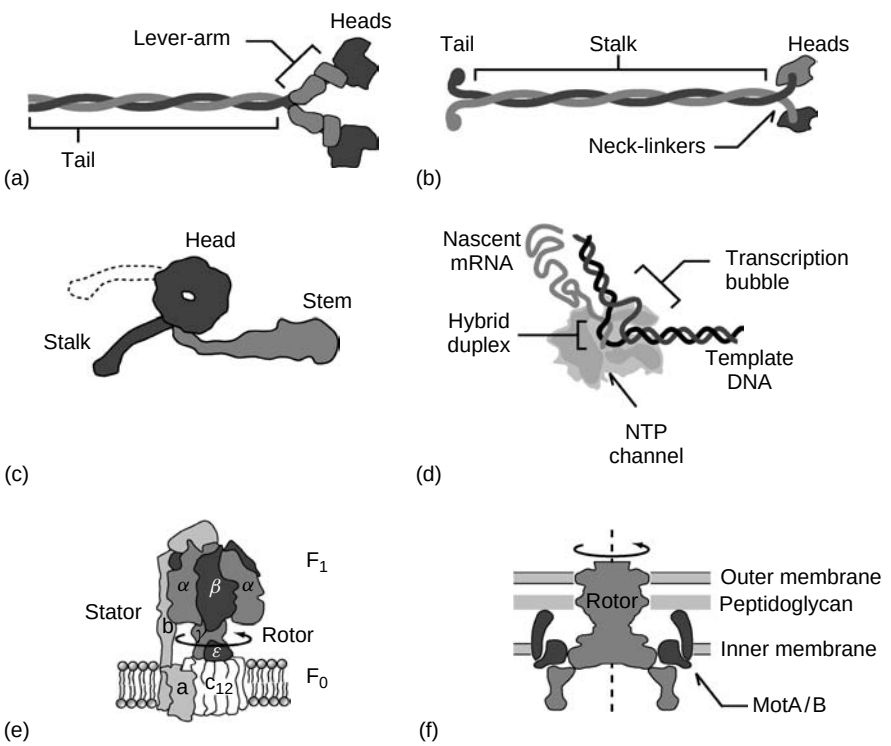

filament sliding. The heads can bind and hydrolyze ATP, and also carry a site that attaches to actin, the main component of the thin filaments, with ATP-dependent affinity. In high-resolution structures, a rodlike portion of the myosin head is found in several different orientations relative to the remainder of the head. Tilting of this "lever-arm," not the entire head, probably drives filament sliding.

The thin filaments of muscle are composed mainly of actin. Actin is a roughly spherical protein that polymerizes into a ropelike structure, with two strands, called *protofilaments*, that twist around one another. The filaments are polar, with a "plus" or "barbed" end, and a "minus" or "pointed" end, which are structurally different. During muscle contraction, the thick filaments slide toward the plus ends of the thin filaments.

2.2 Seeing is Believing: Motility Assays Demonstrate Motor Activity

A wealth of biochemical and structural information supports the tilting cross-bridge model of muscle contraction. However, the most compelling evidence for myosin motility comes from direct observation of motion generated *in vitro*. Myosin and actin are too small to see in an optical microscope. So, *in vitro* motility assays depend on various labeling schemes to render the motion visible. In the earliest assays, micron-sized beads were coated with myosin, and the beads were then observed to move along actin cables in the cytoplasm of the alga *Nitella*, and later along purified actin filaments bound to a glass surface. In an alternate strategy, actin filaments were made visible by fluorescent labeling, and gliding of these labeled filaments on myosin-coated glass surfaces was observed in a fluorescence microscope (Fig. 2a). These important experiments established beyond doubt that actin and myosin alone, without any additional components from muscle cells, were sufficient to generate motion and force. Consistent with the rotating cross-bridge theory, ATP was required for the motility, and the myosins moved toward the plus ends of the actin filaments. Filaments glided *in vitro* at 6000 nm s^{-1} (Table 1), similar to the speed at which thick and thin filaments slide past one another during muscle contraction. As discussed below, these two basic tests – the "bead assay" and the "gliding filament assay" – have been adapted and refined to study a variety of other motors in addition to myosin (Fig. 2b through d).

2.3 Kinesin: Intracellular Porter

Motility assays were instrumental in the discovery of kinesin. Observations of the squid giant axon suggested the existence of motors that consume ATP and haul vesicles at speeds of 1 to 2 μm s^{-1} along the dense array of microtubule filaments within the axon. A putative motor was first isolated by locking the vesicles onto the microtubules using a nonhydrolyzable ATP analog, AMPPNP, followed by purification of the microtubules, and then release of the motor with ATP. A gliding filament assay identical to that developed for myosin confirmed that the purified protein, kinesin, was indeed a motor: glass surfaces coated with kinesin supported the ATP-dependent gliding of microtubules. The gliding velocity, 800 nm s^{-1} (Table 1), closely matched the speed of vesicle transport.

Additional assays, using microtubules marked to reveal their intrinsic polarity, revealed the direction of kinesin-driven

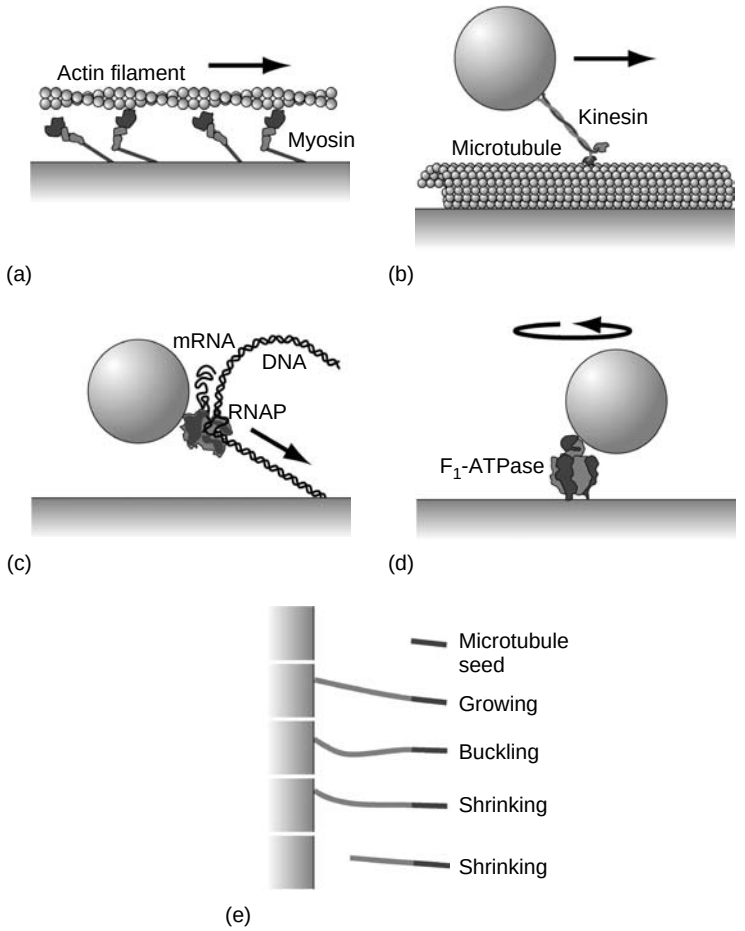

Fig. 2 Motility assays adapted for various molecular motors. (a) In the gliding filament assay, coverslip-bound motors drive filaments to move in a direction parallel to the filament long axis. Here, many myosin heads are shown interacting with a single actin filament. (b) In the bead motility assay, motors are attached to microscopic glass or plastic beads, which are pulled by the motors along coverslip-bound filaments. Kinesin is shown moving along a microtubule. (c) The tethered particle assay was developed to study the motion of nucleic acid enzymes along DNA filaments. Here, a microscopic bead is attached to RNA polymerase, which is transcribing a DNA filament that is attached at one end to the coverslip. Transcription results in a shortening (or lengthening) of the tether (depending on which end of the DNA is surface-bound). (d) In the rotation assay for F_1-ATPase, the motor is attached to a coverslip, and the orientation of the shaft is marked by off-axis attachment of a microscopic bead or filament. Spinning of the shaft causes the bead or filament to rotate. (e) Force generation by polymerization has been demonstrated by growing dynamic microtubule extensions from coverslip-bound seeds. The microtubules continue to elongate even after their growing ends encounter a barricade, which generates enough compressive force to buckle the filaments.

Tab. 1 Properties of selected molecular motors.

Protein Machine	*Molecular Weight (kDa)*	*Force-generating Partner*	*Energy Source*	*Maximum Speed (nm s^{-1})*[a]	*Maximum Force (pN)*[a]	*Step Size (nm)*[a]	*Processivity (cycles)*
Kinesin, native heterotetramer	340	Microtubule	ATP	1800			
Kinesin, truncated active homodimer	90	Microtubule	ATP	800	6	8	100
Myosin II, native heterohexamer	500	Actin filament	ATP	6000	1.5		
Myosin II, active HMM fragment	110	Actin filament	ATP	8000		6	1
Dynein, inner arm subspecies c	500	Microtubule	ATP	700	1.1	$8n$ ($n = 1, 2, 3 \ldots$)	~10
RNA polymerase, *E. coli* core enzyme	380	dsDNA	NTPs	5	27	0.34	>10 000
F_1F_0-ATP synthase	540	n/a	Protonmotive				n/a
F_1-ATPase, active rotary motor	350	n/a	ATP	150 Hz	40 pN nm	120°	n/a
Bacterial flagellar motor, basal body	9500	n/a	Protonmotive	300 Hz	4600 pN nm		n/a
Microtubule, growing	110 (tubulin dimer)	n/a	Binding	~50	4		n/a
Microtubule, depolymerizing		n/a	GTP	~500			
Actin filament, growing	40 (G-actin monomer)	n/a	Binding	~15			n/a

[a]Different units apply to the rotary motors, F_1F_0-ATP synthase, F_1-ATPase, and the bacterial flagellar motor. For these, the maximum rotation rate, torque, and angular step size are reported in units of Hz, pN nm, and degrees, as noted.

motion. Microtubules are rigid, tube-shaped polymers, composed of tubulin proteins arranged in a lattice, resembling a miniature drinking straw. Like actin filaments, microtubules have two structurally distinct ends, called ''plus'' and ''minus''. Polarity-marked filaments driven by kinesin glided with their minus ends leading, implying that kinesin was moving toward the plus ends.

Kinesin and myosin share many structural and functional similarities. Like myosin, each kinesin molecule consists of two identical polypeptides that form twin heads connected to a common coiled-coil stalk (Fig. 1b). The motor activity is carried by the heads. Each head hydrolyzes ATP and attaches to microtubule filaments with nucleotide-dependent affinity. But unlike myosin, the tail of kinesin does not cause bundling, it binds the motor to its cargo. Furthermore, atomic structures of kinesin heads are devoid of any rodlike structure resembling the lever-arm of myosin. Lacking a lever, kinesin's working stroke is likely to be very different from that of myosin.

2.4 Ciliary Dynein: The Dark Horse

Comparatively little functional information is available for the third classic motor, dynein, even though its activity was discovered around the same time as that of myosin. Electron micrographs revealed lateral connections between the parallel microtubules in cilia and flagella, similar to the myosin cross-bridges in muscle. The cross-bridges in cilia and flagella are dynein motors that drive bending motions by sliding microtubules past one another.

Dynein is larger and structurally more complex than kinesin or myosin, but it has many features common to all the classic motors. Depending on the source, dynein consists of one, two, or three large (500 kDa) polypeptides. Each of these forms a donut-shaped head, with two rodlike structures emanating from it, the ''stem'' and the ''stalk'' (Fig. 1c). The stem functions similar to the tails of kinesin and myosin, bundling the heads together, and also anchoring them tightly to their cargo. The tip of the stalk binds microtubules in an ATP-dependent manner, like the microtubule binding site within each of kinesin's heads. No atomic resolution structures are available for dynein, but electron microscopy of single dynein particles revealed a conformational change akin to the lever-arm tilting of myosin: under different nucleotide conditions, the stem adopts two different orientations relative to the head and stalk. Thus, dynein may move its stem-bound cargo by cyclically attaching via the stalk to a microtubule, rotating the stalk and head, detaching from the microtubule, and then unrotating. Dynein supports microtubule gliding and bead motion in *in vitro* assays. The direction of dynein-driven motion is toward the minus end of the microtubule, opposite that of kinesin-driven motion.

Dynein's complex structure contains a number of features with unknown functional significance. Dynein motors from different sources have different numbers of heads. Each donut-shaped head consists of six different subdomains arranged in a hexameric ring. Four of these subdomains bind ATP, but only one catalyzes hydrolysis. Nucleotide binding, but not hydrolysis, at one of the other subdomains is essential for motor activity. Uncovering the reasons for this complexity and elucidating dynein's mechanism of action are important frontiers for future research.

2.5 Processivity Allows Kinesin to Work Alone

Soon after its discovery, kinesin was found to possess a tenacity that set it apart from myosin and dynein. Kinesin is highly *processive*, staying attached to the microtubule as it undergoes many catalytic cycles, and translocating over relatively long distances before detaching. This processivity was first demonstrated when gliding filaments or moving beads were found to move long distances (1–2 μm), even when the surface density of motors on the slide or bead was extremely low, ensuring that single kinesin–microtubule interactions were very likely. Several independent lines of evidence now provide very strong evidence of kinesin's processivity (see Fig. 3).

The processivity of kinesin probably evolved as a means to conserve cellular resources. Kinesin's role of transporting vesicles across neurons is a critical task that must be accomplished repeatedly and with high reliability in order for these cells to function. The longest neurons contain millions of vesicles that each take weeks to make the journey from one end to the other. Kinesin's high processivity allows this Herculean task to be completed by just a few motors bound to each vesicle. In principle, the job could also be accomplished by nonprocessive motors, but many more motors would be required. The cell, in turn, would have to devote more energy and resources into producing these additional molecules.

Apart from its biological significance, kinesin's processivity has been a great advantage for experimentalists, allowing the first studies of single motor molecules. Motility assays for myosin relied on hundreds of motors acting together because the tiny tilting motions, or *working strokes*, of the individual heads are too small to see in a conventional optical microscope. However, owing to their high degree of processivity, kinesins generate hundreds of working strokes during each encounter with a microtubule moving distances of ~1 μm. The summation of many strokes renders the motion of individual kinesin motors easily visible.

Results from single molecule motility assays revealed a number of insights about how kinesin moves. The motion of kinesin-driven beads *in vitro* was not random over the microtubule surface, but appeared to follow a path parallel to the protofilaments. Gliding filament assays supplied strong evidence for protofilament tracking, when abnormal microtubules with helical protofilaments were shown to rotate about their long axis as they moved. In bead assays, engineered kinesin proteins with only one head failed to generate highly processive motion, indicating that two heads are, in fact, better than one.

2.6 Rowers Versus Porters: Duty Ratio Makes a Difference

The head domains of myosin and kinesin differ markedly in their duty ratio, the fraction of time during each biochemical cycle that they remain attached to their partner filament. Myosin possesses a low duty ratio that allows groups of molecules to work together efficiently, like the rowers of a large canoe, while kinesin has a high ratio, befitting its role as a lone porter. A myosin head has high affinity for actin just after hydrolysis, when the nucleotide-binding pocket contains either ADP and phosphate, or ADP alone, and low affinity when the pocket is empty or contains ATP. The timing of transitions between these states ensures that the high-affinity

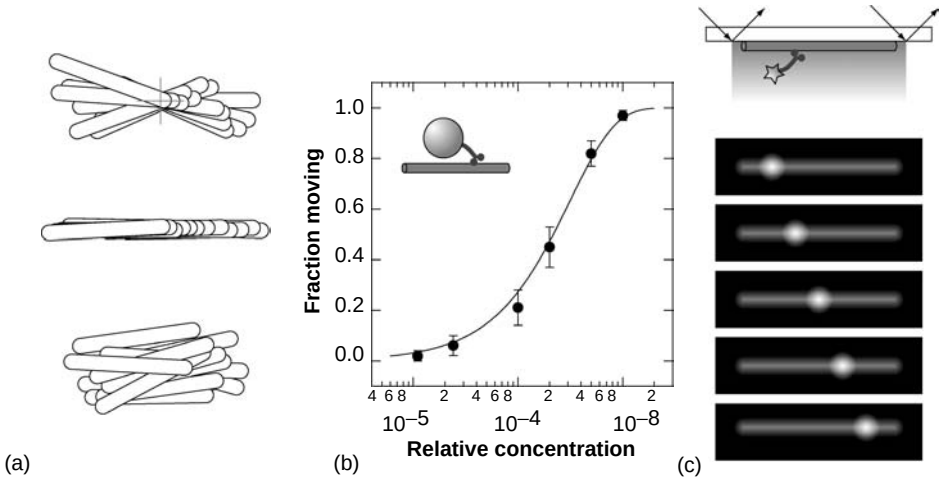

Fig. 3 Strong evidence for the processivity of kinesin. The first evidence for processivity was that kinesin-driven motion persisted *in vitro* even when the surface density of motors was extremely low. (a) Nodal point pivoting in the gliding filament assay also indicated processivity. Microtubules gliding on surfaces decorated sparsely with kinesin rotated erratically about a fixed location, even as they moved through this nodal point. When the trailing end of the microtubule reached the nodal point, it dissociated from the surface and diffused back into solution. A single motor at the nodal point presumably drove the motion (top). Negligible rotation occurred at high motor densities, when multiple motor–filament interactions constrained the filament orientation (middle). Both types of motion were distinct from thermal motion of free filaments in the absence of motor (bottom). [Adapted from Howard, J., Hudspeth, A.J., Vale, R.D. (1989) Movement of microtubules by single kinesin molecules, *Nature* **342**, 154–158.] (b) In the kinesin bead assay, the fraction of moving beads, f, decreased gradually as the relative motor concentration, C, was lowered, as expected if one molecule is sufficient to produce movement. The curve shows a one parameter (λ) fit to Poisson statistics, $f = 1 - \exp(-\lambda C)$. [Adapted from Svoboda, K., Block, S.M. (1994) Force and velocity measured for single kinesin molecules, *Cell* **77**, 773–784.] In the low-density regime, moving beads continued to translocate at normal speeds over distances that were independent of motor concentration (data not shown). (c) A third method used a microscope capable of imaging single fluorophores bound to kinesin (upper panel). The movement of labeled motors along coverslip-bound filaments was directly observed (shown schematically in the lower five panels). [Adapted from Vale, R.D., Funatsu, T., Pierce, D.W., Romberg, L., Harada, Y., Yanagida, T. (1996) Direct observation of single kinesin molecules moving along microtubules, *Nature* **380**, 451–453.] Labeling the motor by fusion to green fluorescent protein avoided chemical modification with reactive dyes, which can damage the motors, and ensured that every motor was labeled with the same number of fluorophores.

states represent <2% of the total cycle time. This low duty ratio is an adaptation that allows the myosin heads to avoid interfering with each other when many are acting on the same actin filament. They detach very quickly after undergoing a working stroke, so the speed of filament sliding is not limited by the hydrolysis rate of the individual heads. In contrast, kinesin heads have a duty ratio >50%, which partially explains the processivity of the motor. Even if the cycles of the two heads were completely uncorrelated, their high duty ratio would ensure that, on

average, at least one was always bound to the microtubule.

2.7 Molecular Tug-of-war: Applying Force to Individual Motors

With the development of single molecule assays, it became possible to directly measure the forces generated by individual motors. One method for measuring force production by kinesin was to attach a microtubule filament to a flexible glass fiber, and then hold the fiber near a surface sparsely coated with kinesin. As individual kinesins on the surface bound and moved along the microtubule, they pulled against the glass fiber and caused it to bend. By measuring the amount of bending, the maximum force against which a kinesin motor could move was estimated to be 5 or 6 pN.

Another method for applying force to individual kinesin molecules, which gave a similar estimate of the stall force and also led to a number of other discoveries, was to use an optical trap. An optical trap is made by focusing a laser through the objective lens of a high-magnification microscope, creating a very bright light spot at the specimen. The focused light traps small objects such as micron-sized beads. When the trapped object is moved away from the center of focus, it feels a restoring force pulling it back that is proportional to the distance from the center, as if the trap was a stretched spring pulling on the object. To apply force to kinesin, an optical trap was used to grab beads with single kinesin molecules attached, and to place them near microtubules stuck onto to a glass surface. When the kinesin began moving along the microtubule, it pulled the bead from the trap center, and the trap supplied a restoring force that placed tension on the kinesin. As the bead was pulled gradually away from the trap center, the force increased and the motor speed decreased, halting when the force reached 6 pN.

2.8 Motors Move in Discrete Steps

Kinesin molecules move discontinuously over the microtubule surface, advancing in discrete 8-nm increments and dwelling at well-defined positions between advancements (Fig. 4). Two key innovations allowed the first observation of steps in the motion of kinesin-driven beads. First, tension supplied by an optical trap suppressed the random, thermally driven ("Brownian") motion that would otherwise dominate. Second, the bead position was measured with very high spatial and temporal resolution by monitoring the distribution of scattered light with a photodetector. The 8-nm step size (Fig. 4b) matches the spacing of tubulin dimers in the microtubule lattice. Similar experiments with dynein, which is processive under some conditions, suggest that it also moves stepwise, advancing by multiples of 8 nm.

Optical trapping has been applied to measure the motion of single myosin motors. A different technique than that used for kinesin was required because the interactions between a myosin molecule and an actin filament are fleeting, lasting only a few tens of milliseconds. To resolve these quick attachments, an assay was developed in which an actin filament with beads attached at both ends was suspended between two optical traps and held near a surface sparsely coated with myosin (Fig. 4c). Thermal motion of the beads decreased when a motor became attached to the filament, and the average position of the beads shifted abruptly, due to the tilting of the myosin head, by 6 nm. This tiny

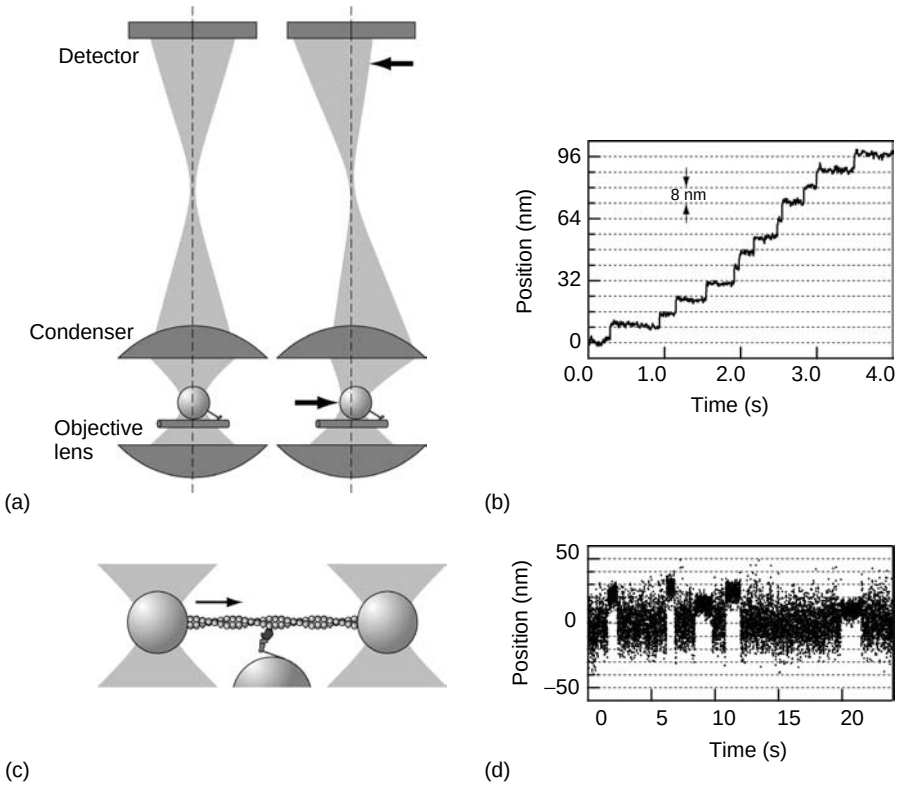

Fig. 4 Measuring the discrete motions of kinesin and myosin molecules using optical traps. Both methods use microscopic beads as handles to apply force, and as markers for the position of the motor or filament. (a) An optical trap applies tension during kinesin-driven movement of a bead along a coverslip-bound microtubule, and this tension reduces thermal motion of the bead so that individual 8-nm steps can be resolved. Bead position is detected with very high spatial resolution by monitoring the distribution of scattered light with a photodetector. (b) Example trace showing 8-nm steps generated by a single kinesin molecule. (c) The three-bead assay developed for measuring working strokes of muscle myosin. An actin filament is pulled taut between two microscopic beads held in optical traps. Binding of a single myosin head to the filament reduces thermal motion of the beads, and also induces a working stroke in the myosin, which causes the beads to deflect by 5 to 15 nm. (d) Example trace showing interactions between a single myosin head and an actin filament. [Data reprinted with permission from Lister, I., Schmitz, S., Walker, M., Trinick, J., Buss, F., Veigel, C., Kendrick-Jones, J. (2004) A monomeric myosin VI with a large working stroke, *EMBO J.* **23**, 1729–1738.]

distance is the maximum sliding distance that a myosin molecule can generate during a single interaction with actin during muscle contraction. More than a million of these interactions are evidently required just to lift a finger.

2.9 Different Strokes: Variation Within and Across Motor Families

In accordance with their diverse roles, motors of the myosin, kinesin, and dynein

families often have important differences in the way they move. The cellular role of type V myosin, for example, is more similar to that of kinesin than to muscle myosin. Myosin V acts alone or in small numbers to transport vesicles and organelles along actin cables. It was therefore not surprising to find that this myosin, like kinesin, exhibits processive movement, taking several steps along an actin filament before detaching. The step size of myosin V, 36 nm, is much larger than that of kinesin, and matches the wide spacing of binding sites that occur every half-period in the actin helix. The structure of myosin V explains how it can generate such large movements. Each head contains a lever-arm that is 24 nm long, three times longer than the lever-arm of muscle myosin.

Diversity within motor families invites comparison, which can illuminate important aspects of motor function. Strong evidence supporting the tilting lever-arm model came from comparisons of gliding speeds and stroke lengths generated *in vitro* by myosin-family motors with different lever-arm lengths. The speeds and stroke lengths varied in proportion with the lever-arm length, as predicted by the model. Structural differences between kinesin and Ncd, a related motor that moves toward the opposite end of microtubules, suggested that a "gearbox" region just outside the head domain controls the direction of motion of these motors. This hypothesis was confirmed by testing chimeric motors, made by swapping the gearbox regions of the two proteins, in gliding filament assays.

2.10 Fuel Economy and Energy Efficiency

How many ATP molecules does a motor require to generate a working stroke or step? Is this "fuel economy" always the same? For kinesin, the coupling ratio – the number of ATPs consumed per step – has been measured by comparing the stepping rate in single molecule assays to the rate of ATP hydrolysis. Over a wide range of ATP concentrations and loads, one ATP is hydrolyzed per 8-nm step. This tight, 1 : 1 coupling implies that the energy efficiency of kinesin can be very high. ATP hydrolysis under physiological conditions is worth ~80 pN nm (or $2 \cdot 10^{-23}$ kcal). When kinesin generates 8-nm steps under 5 pN of load, it produces as much as 40 pN nm (10^{-23} kcal), or 50% of the total chemical energy available. So kinesin is more than twice as efficient as the best man-made gasoline engines, which are 24% efficient at full power, and typically achieve only 10 to 15% on the road.

The energy efficiency of cytoplasmic dynein, 10%, is considerably lower than that of kinesin. However, dynein's complex structure may act like an automobile transmission, allowing it to maximize fuel economy. Near dynein's stall force, 1 pN, the motor takes 8-nm steps. At <0.4 pN, however, it seems to advance in larger increments of 24 or 32 nm. Thus, dynein can apparently shift into high gear when carrying a light load. Assuming that the coupling ratio under both conditions is equal, the larger step size will result in proportionally better fuel economy.

2.11 Walk This Way: Processive Mechanoenzymes Move Hand-over-hand

The fact that single kinesin molecules generate hundreds of steps, even under load, together with the earlier findings of motion parallel to the protofilaments and

the requirement for two heads, suggested that it might walk – or waddle – from one tubulin dimer to the next. An attractive hypothesis was that the twin heads each take turns, alternately detaching and moving past one another, in a "hand-over-hand" motion resembling that of a person swinging along monkey bars. The same model was also thought to apply to the processive myosin-family motor, myosin V.

Single molecule experiments with myosin V and kinesin have confirmed that both walk hand-over-hand. First, the stride length of myosin V was measured by labeling one of its two heads with a fluorophore, and tracking the label with nanometer resolution. This is like watching a person walking across a field on a moonless night with a flashlight attached to one foot: The person is invisible, but the light moves visibly with every other step. The heads of myosin V took turns making strides that were twice as long as the distance moved by the tail, evidence that strongly supported a hand-over-hand model. Next, optical trapping experiments showed that some kinesin molecules limp along the microtubule, exhibiting a difference in the timing of every other step. Limping implied that kinesin switched between two different configurations as it stepped. The most severe limpers were mutants in which one head hydrolyzed ATP more slowly than the other, arguing for a mechanism in which the two heads swap both mechanical and catalytic activities with each step. Finally, the stride length for one of kinesin's two heads was measured to be 16 nm, double the step size. Taken together, these results make a very strong case for a hand-over-hand mechanism for kinesin and myosin V.

2.12 Coordination is Required

To fully account for the hand-over-hand walking of kinesin and myosin V, coordination between the two heads is essential, and it may be achieved through mechanical tension between the heads. The mechanical cycle of both motors includes a transient state in which both heads are attached to the filament. The motors are probably stretched in this doubly attached state, owing to the relatively large distance between the heads, and the resulting intermolecular strain between the heads could bias their kinetics so that the trailing head nearly always detaches before the leading head. In support of this hypothesis, external loads have been shown to strongly affect the rate of detachment of single myosin V heads, with forward loads accelerating detachment and backward loads slowing detachment.

Coordinated action at distinct sites is a universal requirement for all motors, not just those with high processivity. Automobile engines rely on the carefully timed actions of pistons, valves, and spark plugs. Mechanoenzymes also have critical interacting parts. Consider the mechanochemical cycle of myosin: Within a few milliseconds after attachment of a myosin head to actin, small motions occur in the nucleotide-binding pocket, allowing phosphate release. This triggers lever-arm tilting, which moves the actin filament. Subsequent ADP release triggers a second, smaller motion in certain myosin types (e.g. myosin I and single myosin V heads), but it is not clear if this additional stroke occurs for muscle myosin. ATP binding triggers detachment from actin, and hydrolysis "primes" the motor for the next cycle. For other motors, the specifics and timing are different than for myosin,

but this coordinated action is a universal requirement.

2.13 The Kinesin Cycle and Working Stroke

The mechanochemical cycle of kinesin is not as thoroughly understood as that of myosin. Because two heads are involved in the stepping process, the kinesin cycle is necessarily more complex than that of muscle myosin. The specific conformational changes driving motion are poorly defined, and there is uncertainty about which biochemical events are associated with motion. A working hypothesis for kinesin is that all the mechanical action is associated with just one biochemical event. In support of this "one-stroke" hypothesis, reaction schemes with just one force-dependent rate can account for force-velocity and [ATP]-velocity curves measured in single molecule assays. These schemes predict a working stroke after ATP binding, possibly upon ATP hydrolysis. By contrast, a working stroke concomitant with ATP binding is suggested by kinetic measurements, showing that ADP release from one head is stimulated by binding of a nonhydrolyzable ATP analog (e.g. AMPPNP) to the other head.

A conformational change in the structure of kinesin has been discovered that may drive its motion. This putative working stroke is quite different from either the tilting lever-arm of myosin, or the stem reorientation of dynein. In single-headed kinesin constructs, a 15-amino acid peptide known as the "neck linker," which connects each kinesin head to the coiled-coil stalk, undergoes a nucleotide-dependent transition. In the presence of ADP, or when no nucleotide is present, the neck linker is disordered. In this state, it acts like a flexible tether, pivoting about a point on the backside of the head. When the head is attached to a microtubule in the presence of ATP analogs (AMPPNP, ADP-AlF_4^-), the neck linker "zips" onto to the surface of the head, and its end points toward the microtubule plus end. In the full two-headed motor, zipping of the neck linker on one head could drive stepping by moving the stalk, and therefore the other head, toward the next attachment site on the microtubule lattice.

2.14 Under the Hood, Motors are Still a Mystery

Even for the best-understood motors, where high-resolution structures are available, and major mechanical steps can be identified with specific biochemical transitions, there are very fundamental questions that remain unanswered. In particular, the atomic-scale motions that transduce small chemical events in the nucleotide pocket and convert them into larger motions elsewhere are poorly understood. Important residues have been identified by comparing sequences of related motors and their precursors (e.g. myosin, kinesin, and G-proteins), and subdomains that move relative to others have been suggested by structural comparison. However, ultimately, these static structures cannot elucidate the timing of movements and the cause-and-effect relationships between the various parts. A myriad of physical and biological techniques will no doubt be essential in this effort, but single molecule techniques that simultaneously record motion and biochemical changes (e.g. through fluorescence) seem particularly valuable in this regard.

3 Nontraditional Molecular Motors

There are many other protein machines that generate force and motion, but which do not fit the classic motor paradigm. A growing number are being studied using *in vitro* assays, like those discussed above, which allow the mechanical output of single motors to be measured.

3.1 Nucleic Acid Enzymes

Some of the most important processes of life are carried out by nucleic acid enzymes, many of which are processive motors that move along DNA. Every human cell stores genetic information in the form of 23 strands of DNA, totaling three billion base pairs, and measuring 1 m in total length. All 23 strands are copied by DNA polymerase enzymes, untangled by topoisomerase enzymes, and packaged into chromosomes by condensins, before cell division. The genetic information contained in the DNA is transcribed into mRNA by RNA polymerase (RNAP), and translated into protein by the ribosome. Each of these nucleic acid enzymes is a protein machine capable of generating force and motion, and each is fascinating in its own right. A discussion of all of them is beyond the scope of this chapter, which will focus on one important example, RNAP.

3.2 More Than a Motor: Multitasking by RNA Polymerase

Even though the size of an RNAP enzyme, by comparison of total molecular weight, is not so different from that of the classic mechanoenzymes, its function is much more complex. Motion is merely a by-product of the biological role of this protein machine, transcribing the genetic code. While moving along a DNA template, RNAP separates a short section of the DNA duplex, the "transcription bubble," and builds a copy of one strand by selecting complementary nucleotides from the surrounding solution and attaching them, one at a time, to the end of the nascent mRNA chain (Fig. 1d). Along the way, it must maintain registration of a short section of "hybrid" RNA:DNA duplex, and also respond to a number of different signals that control the initiation, termination, and elongation rate of transcription. Like other mechanoenzymes, RNAP derives energy from nucleotide hydrolysis, but, in this case, each nucleotide serves a dual role. After hydrolysis, the nucleotide becomes an information-containing subunit incorporated into the growing mRNA. (If automobile engines could make such efficient use of their exhaust, urban air quality would be much improved!)

Motion may not be its primary function, but RNAP is no slouch of a motor. Its motion can be directly observed by attaching a micron-sized bead, and recording bead motion as the enzyme transcribes a DNA template bound at one end to a glass surface (Fig. 2c). Using this "tethered particle" assay in conjunction with optical trapping, the motion of a single RNAP can be tracked with high spatial resolution, and the effect of applied load can be measured. RNAP is slow, moving in these assays at 10 to 15 bp s^{-1}, or just 5 nm s^{-1} (Table 1). This speed is roughly equivalent to the rate of human hair growth, and 160-fold slower than the speed of kinesin. However, RNAP is much more processive than kinesin. In cells, RNAP molecules synthesize mRNA chains of 10^4 (in

bacteria) to 10^6 (in mammals) nucleotides. *In vitro*, its processivity is reduced, but the enzyme typically moves across several thousand bases or more before detaching from the template. Movement continues, unhindered at 10 to 15 bp s^{-1}, even when backward loads as high as 27 pN are applied. A high stall force (>5-fold higher than that of kinesin) may be necessary for RNAP to function *in vivo*, perhaps allowing the enzyme to push through "road-blocks" formed by other DNA-binding proteins. RNAP is presumed to move in discrete steps corresponding to the distance between individual bases along the DNA helix, 0.34 nm. This distance is extraordinarily small (20 times smaller than kinesin's 8-nm steps), and steps of this size have not yet been directly observed. However, optical trapping technology is rapidly advancing, and such tiny motions may soon be resolvable.

3.3 RNA Polymerase Structure

RNAP is shaped like a claw. The claw opens to allow a DNA strand to enter, and during transcription it closes, wrapping completely around the DNA. Besides the DNA entry and exit channels, there is also a channel through which the newly synthesized RNA exits, and a pore for nucleotide entry (Fig. 1d). Inside the closed structure, the enzyme makes numerous contacts with the hybrid duplex and the DNA. It is unknown, which parts of RNAP are responsible for generating motion and force production, but many candidate features are apparent in the high-resolution structures. For example, a "bridge helix" located near the site of nucleotide addition may undergo a conformational change that pushes the enzyme to the next site. Determining which portions of RNAP are responsible for its motion is a great challenge for future research.

3.4 What Causes Pauses?

The motion of RNAP along the DNA template is interrupted by pauses, lasting from a few seconds to many minutes. The short-duration pauses (those with lifetimes of seconds) are very frequent, and may result from the enzyme encountering a few GC-rich base pairs of DNA that are tougher-than-average to separate. Occasionally, the enzyme pauses for a much longer duration (20 s to >30 min). These infrequent but long-lived pauses may occur for two reasons, both of which illustrate the sophisticated behavior that RNAP is capable of. First, long pauses can be induced when RNAP encounters a specific sequence in the DNA template it is transcribing. Such sequence-dependent pauses are an important mechanism for regulation of gene expression. By relieving these long pauses, a cell can greatly alter the rate of expression of a pause-containing gene. In some cases, these sequence-dependent pauses occur when the nascent mRNA chain folds into a hairpin structure, which then interacts directly with the RNAP enzyme. In other cases, these pauses occur when the RNAP transcribes a slippery AT-rich sequence, resulting in an unstable RNA:DNA hybrid duplex that allows the enzyme to backtrack. There is a second class of long-duration pauses that are not sequence-dependent, and these probably occur when the enzyme makes a copying error, misincorporating a noncomplementary nucleotide into the mRNA chain. These pauses are also associated with backtracking, and may reflect a

"proofreading" activity whereby the enzyme slides backward and then (with the help of accessory factors) cleaves a short section from the end of the mRNA, removing the mistake before resuming elongation. Determining the reasons why RNAP pauses is critical for understanding how gene transcription is controlled. Single molecule experiments will be useful in this effort because the motion of unsynchronized molecules can be followed with very high resolution, in real time.

3.5 ATP Synthase

A rotary machine, F_1F_0-ATP synthase, is at the heart of energy metabolism in plants, animals, and photosynthetic bacteria, where its role is to replenish the cellular store of ATP. The importance of this job is obvious when one considers that a human body contains 100 g of ATP (0.25 moles), each molecule of which gets hydrolyzed 400 times a day to power various cellular tasks. An army of ATP synthase enzymes performs the $\sim 10^{26}$ synthesis reactions required to regenerate the spent ATP. To do so, the enzymes tap into an electrochemical gradient, the protonmotive force, that exists across the inner mitochondrial membrane of animal cells, or across the thylakoid and plasma membranes of plants and eubacteria, respectively.

ATP synthase consists of two separate rotary motors, F_0 and F_1, that work together as a motor–generator pair. Normally, F_0 is the driving motor of the pair. As protons flow through it, down the electrochemical gradient, a portion of F_0 spins like a water wheel. The spinning wheel of F_0 supplies torque that rotates a shaft within the other motor, F_1, causing it to regenerate ATP from ADP and phosphate. The motor–generator pair can also operate in reverse. In this case, ATP hydrolysis by the F_1 motor causes the shaft to rotate backward, which supplies a torque that spins the wheel of F_0 and pumps protons back up the electrochemical gradient.

F_1F_0-ATP synthase is comprised of eight different types of protein subunits (Fig. 1e). The water wheel, or "rotor" portion of F_0 is a ring of 12 identical subunits, c_{12}, that spin in the plane of the membrane relative to a "stator" composed of three other subunits, ab_2. F_1, is a donut-shaped structure made of three pairs of proteins, $(\alpha\beta)_3$, plus two additional proteins, γ and ε, which form the shaft that fits into the center of the donut. Clockwise (CW) rotation of the $\gamma\varepsilon$ shaft (as seen from the F_0 or membrane side) causes ATP synthesis to occur sequentially at three catalytic sites located symmetrically around the $(\alpha\beta)_3$ ring. Because isolated F_1 can function in reverse, catalyzing ATP hydrolysis and counter clockwise (CCW) rotation of the shaft, it is often referred to as F_1-ATPase. In the full, F_1F_0-ATP synthase enzyme, F_1 sits directly above F_0, with the $\gamma\varepsilon$ shaft of F_1 making a rigid connection to the c_{12} ring of F_0, and with the ab_2 stator of F_0 connecting to the $(\alpha\beta)_3$ donut of F_1. Normally, spinning of c_{12} drives rotation of $\gamma\varepsilon$ within $(\alpha\beta)_3$ and hence ATP synthesis.

The ring-shaped structures within F_1F_0-ATP synthase provided the first clues that rotation might be important to its function. As for the classic mechanoenzymes, proof of motion came when an *in vitro* assay was developed, allowing the rotation to be directly observed. F_1-ATPase molecules were attached sparsely to a surface, and the orientations of their $\gamma\varepsilon$ shafts were marked by attaching micron-long, fluorescent-labeled actin filaments. In the presence of ATP, the filaments rotated CCW, indicating shaft rotation. At very

low concentrations of ATP, the filaments rotated in discrete steps, dwelling at well-defined orientations in between rapid, 120° reorientations. Rotation rates were one-third of the rate of ATP hydrolysis, implying that each 120° reorientation corresponds to hydrolysis of a single ATP. The filaments used to mark shaft orientation also supplied a drag force acting against the rotation. By calculating the drag on the filaments, F_1 was estimated to deliver 40 pN nm of torque during each 120° reorientation, giving 80 pN nm of mechanical work output per ATP hydrolysis. This is 100% of the available chemical energy, making F_1-ATPase one of the most efficient motors known.

F_1-ATPase is a relative newcomer to the molecular motor scene, but it is quickly becoming one of the best-understood examples of mechanochemical coupling. A series of experiments with single F_1 molecules has revealed that each 120° step occurs in two phases, or substeps, that correspond to particular biochemical transitions in the ATP synthesis reactions occurring at each of the three catalytic sites. Substeps were observed by marking the shaft orientation with 40-nm gold particles (Fig. 2d), rather than the much larger filaments used in earlier work. The small beads resulted in a lower drag force acting on the motor, allowing full speed rotation at 160 revolutions per second (Hz) (Table 1). Capturing the motion with a high-speed video camera showed a substep of 80 to 90° followed by one of 30 to 40°, underlying each of the 120° steps previously observed. Finally, simultaneous observation of shaft orientation and binding and release of a fluorescent nucleotide allowed these biochemical steps to be temporally correlated with the substeps. ATP binding to one of the catalytic sites (site 0) is concurrent with the 80 to 90° substep, and the remaining 30 to 40° substep requires hydrolysis at the site that previously bound ATP (site −1), and release of ADP from the remaining site (−2).

It is unknown whether the events seen during ATP hydrolysis by F_1 are simply the reverse of those occurring during synthesis, but several experiments confirm, at least, that the $\gamma\varepsilon$ shaft rotates in the opposite direction during synthesis. In one experiment, magnetic beads attached to the shaft were used to drive CW rotation in F_1-ATPase, in the presence of ADP and phosphate, and a luciferin–luciferase system that emits a photon upon reacting with ATP was used to verify synthesis. In another experiment, individual, fluorescent-labeled F_1F_0-ATP synthase complexes were embedded in liposomes. A pH difference was created across the membrane by rapid dilution, and rotation was recorded by fluorescence resonance energy transfer (FRET).

3.6 The Rotary Motor of Bacterial Flagella

Bacterial flagella are very different from the flagella and cilia of eukaryotic cells. Flagellated bacteria, such as *Escherichia coli*, swim by rotating a set of four corkscrew-shaped filaments (four, on average) that extend from the cell surface out into the surrounding medium. At the base of each filament is a large protein machine that drives filament rotation. This rotary motor, like F_1F_0-ATP synthase, is powered by the protonmotive force, but it is much larger and structurally more complex than F_1F_0-ATP synthase.

Bacteria swim to find food. They control their swimming behavior by altering the direction of rotation of their flagellar motors. When all four motors rotate CCW

(as seen by an observer outside the cell), the cell swims steadily, or "runs," in a relatively straight line parallel to its long axis. When one or more motors rotates CW, the cell "tumbles," erratically moving in place and reorienting itself. The motors switch from CCW to CW at random, so that typical swimming involves runs that last ~1 s, interspersed with tumbles that last a few milliseconds. When the bacteria senses rising nutrient concentrations, it lengthens the runs by increasing the probability of CCW rotation. In this way, the cell moves, on average, toward the food.

The core of the bacterial flagellar motor is a stack of ring-shaped structures, 45 nm in diameter, embedded in the multilayered cell envelope (see Fig 1f). The rings are composed of 20 different types of proteins, but they are all thought to rotate together as a single unit, the "rotor." Rotation of the stack of rings is driven by a circular array of ≤16 "studs" that surround the base of the stack, and which are anchored to the framework of the cell wall. Each stud is composed of two MotA proteins (32 kDa), and one MotB protein (34 kDa). No high-resolution structures are available for the MotA/MotB complex, but both proteins span the cytoplasmic membrane, forming a transmembrane proton channel. MotB has a proton-acceptor site, and MotA contains a site that interacts with the base of the rotor. Protonation and deprotonation of MotB is thought to cause conformational changes in MotA, which probably binds and unbinds from the rotor, driving rotation and torque generation. The minimal torque-generating unit may be composed of two studs (i.e. four MotA subunits plus two MotB subunits). More detailed descriptions of the structure can be found in review articles cited in the bibliography.

Individual flagellar motors can be studied by attaching bacteria to a glass surface by one flagellum. The tail wags the dog in this tethered cell assay – with the flagellum anchored, rotation of the motor causes the whole cell body to spin. To spin the entire cell, the motor must overcome a large viscous drag, so it spins relatively slowly in this assay, at 10 Hz. But it produces an impressive 4600 pN nm of torque. When mutant cells lacking MotB are tethered, they are paralyzed and do not spin. Amazingly, these paralyzed cells can be "resurrected" by expression of MotB from an inducible gene. Resurrected cells begin to rotate within several minutes after induction of MotB expression, and their speed increases in a series of discrete jumps. Each jump in speed represents the incorporation of one additional torque-generating unit into the motor. As many as eight jumps can be seen, implying that the maximum number of torque generators is eight.

Each torque generator is itself a processive, high-duty ratio motor that moves along the surface of the rotor without detaching. The best evidence for processivity is that tethered cells in the resurrection experiment with just one torque-generating unit spin relatively smoothly, and do not freely undergo rotational Brownian motion. Evidently, a single unit is sufficient to prevent the motor from slipping, and each unit remains attached to the rotor during most, or all, of its mechanical cycle. There are twice as many studs (16) as torque-generating units (8), and one hypothesis is that each unit is a coordinated pair of studs, possibly moving in a hand-over-hand motion like that of

kinesin or myosin V. A single unit is expected to move stepwise over the surface of the rotor, perhaps taking ~26 steps per revolution (the approximate number of subunits composing the base of the rotor), but such steps have not been directly observed. Dividing the maximum torque (4600 pN) by the number of torque-generating units (8), and by their distance from the axis of rotation (20 nm), shows that each unit generates considerable force, up to 29 pN, which is comparable to the stall force of RNAP. Their speed of motion over the rotor surface is also quite high. When the viscous load is minimal, the motor can rotate as fast as 300 Hz (Table 1). This translates into motion of the torque-generators at 38 000 nm s^{-1} over the rotor surface, which is >6-fold faster than muscle myosin, and similar to the speed of the fastest myosins (e.g. type XI, responsible for cytoplasmic streaming in algae).

The speed of rotation is proportional to the protonmotive force, as shown by wiring a cell to an external voltage source and watching an inert marker on the motor. To apply voltage, the cell body was drawn halfway into a micropipette, and the membrane permeabilized by chemical treatment. Estimates of the proton flux through the motor suggest that the motor is tightly coupled. Roughly 1200 protons flow through the motor during each complete revolution. By attaching a variety of different-sized latex beads to the filaments and adjusting the viscosity of the surrounding fluid, torque-speed relations have been measured over a wide range of speeds. Forward rotation under assisting torques, and backward rotation under torques above stall (>4600 pN), has been explored by using rotating electric fields or optical traps to apply torque in the tethered cell assay.

3.7 Polymers that Push and Pull

Actin filaments and microtubules are not just static polymers. In addition to their roles as structural cables and girders for maintaining cell shape, and as superhighways for mechanoenzymes to move along, these polymers are also dynamic machines that can produce force. In living cells, the cytoskeletal polymers are in a constant state of flux, and their growth and shrinkage is harnessed to drive many organelle and whole-cell movements. Crawling cells have a dense array of polymerizing actin filaments beneath their leading edge that pushes outward on the plasma membrane and causes protrusion. Similarly, the bacterial pathogen, *Listeria monocytogenes* is pushed by actin polymerization. The bacteria move in graceful arcs through the cytoplasm of a host cell, leaving "comet tails" of polymerized actin in their wake. During mitosis, chromosomes are pushed and pulled by dynamic microtubules whose ends are linked to specialized sites on the chromosomes, the kinetochores. Just before cell division, kinetochore-attached microtubules depolymerize, generating tension that pulls sister chromatids apart.

Both actin and microtubule filaments are composed of protein subunits arranged in a regular lattice. The monomeric form of actin, "G-actin," is a roughly spherical protein, 5 nm in diameter (45 kDa). Like a LEGO block, the surface of an actin monomer has several sites for attachment to other monomers. Each also has a cleft that binds an ATP molecule. Monomers assemble into a ropelike structure, "F-actin," with two strands, called *protofilaments*, that wind around each other with a helical period of 72 nm. The building blocks for microtubules are made

of tubulin, a molecule that consists of two nearly identical 50-kDa proteins fused tightly to form a dimer, 8 nm in length. Each dimer has two sites that bind GTP. The dimers assemble into a hollow, tube-shaped structure, 25 nm in diameter, with 13 protofilaments that run parallel to the long axis of the tube. Both types of filaments have fast-growing "plus" ends, and slow-growing "minus" ends.

Nucleotide hydrolysis supplies energy that makes actin and tubulin polymers very dynamic. Actin monomers in solution bind ATP, and have high affinity for one another. After polymerization, the ATP is hydrolyzed and phosphate is released, leaving ADP trapped in the binding clefts of the monomers within a filament. The ADP-containing monomers have reduced affinity, so hydrolysis destabilizes the actin filament, promoting depolymerization. GTP hydrolysis by tubulin has a similar effect. Each tubulin dimer binds two molecules of GTP, one of which is hydrolyzed upon incorporation of the dimer into a microtubule filament. The GDP-containing tubulin dimers have reduced affinity for one another, which destabilizes the lattice and promotes depolymerization. Without hydrolysis, both polymers would simply grow until equilibrium was reached, when the subunit pool was spent. Hydrolysis keeps the filaments out of equilibrium, allowing coexistence of growing and depolymerizing filaments.

3.8 Microtubule Ends and Dynamic Instability

The dynamic behavior of microtubules can be directly observed *in vitro*. In the presence of GTP and pure tubulin, microtubule growth is interrupted by periods of rapid depolymerization. This "dynamic instability" requires GTP hydrolysis. Growth rates are normal in the presence of the nonhydrolyzable GTP analog, GMPCPP, but the growth is uninterrupted. Subunit addition and removal occurs only at the ends of the filaments, which adopt different structures, depending on whether they are in a state of growth or depolymerization. The protofilaments that extend from growing ends are straight, forming sheets that are sometimes hundreds of subunits long. In contrast, the protofilaments at depolymerizing ends become highly curved, peeling away from the lattice. The ends of growing filaments are temporarily stabilized by a "cap" of GTP-containing subunits in which hydrolysis has not yet taken place. The transition between growth and depolymerization, called "catastrophe," is probably triggered when hydrolysis of the cap occurs before more GTP-containing subunits are added.

The curvature of protofilaments at the ends of depolymerizing microtubules, and the fact that the products of depolymerization are often curved oligomers, suggests a structural basis for the coupling of nucleotide hydrolysis and polymerization. Before hydrolysis, the GTP-containing dimers are probably straight, fitting snugly into the growing microtubule lattice. If the GDP-containing subunits are naturally curved, then they would be strained when trapped within the lattice. In this way, energy from hydrolysis could be stored within the lattice as mechanical strain.

3.9 Motility Assays with Cytoskeletal Filaments

Several *in vitro* experiments show that polymerization of pure actin or tubulin, without any additional proteins, can generate pushing force and do mechanical work. Polymerizing actin filaments inside

liposomes causes distension of the liposomes, demonstrating that the growing filaments can push outward on the lipid bilayer. Likewise, microtubule filaments grown inside a small chamber can push against the chamber walls with enough force to buckle themselves (Fig. 1e). By analyzing the shapes of buckled filaments, the maximum pushing force of a single microtubule has been estimated at 4 pN.

These important experiments prove that growing filaments can push against an object, but they are incomplete models for the polymer-driven motility that occurs in cells. In cells, a variety of accessory proteins provide spatial and temporal control of filament dynamics, and couple the ends of growing and shrinking filaments to other structures to apply force. *Listeria* promote spatially localized actin polymerization with a nucleation factor, ActA. Likewise, kinetochores contain a host of proteins that modulate microtubule dynamics and maintain attachment to microtubule ends. Understanding the mechanisms of these accessory factors will be key to understanding how cells harness cytoskeletal filaments to produce motion and force.

In vitro motility assays that reconstitute force generation using dynamic filaments coupled to accessory proteins provide more realistic models for filament-based motility in cells. Shrinking microtubules can pull against microscopic beads when the beads are coated with proteins that maintain attachment to the depolymerizing filament ends. This motion is similar to the way chromosomes are pulled apart before cell division, and also to the way the mitotic spindle is positioned inside asymmetrically dividing yeast cells. In a reconstituted assay that closely mimics the motion of *Listeria*, beads coated with ActA protein are pushed around by actin polymerization. The beads follow curved trajectories and leave comet tails of polymerized actin in their wake, just like the bacteria.

4 Conclusion

Motion is fundamental to life. Everyone is familiar with the macroscopic motion of muscle contraction. There are also exquisite motions taking place at the level of cells and molecules. The cells in our immune system crawl around our bodies and engulf invading bacteria. Cilia in our lungs beat to remove inhaled debris. In all these cases, the motion is generated by tiny protein machines, the molecular motors. Molecular motors are ubiquitous, and the list of known motors is growing. Besides the classic motors, myosin, kinesin, and dynein, and the cytoskeletal polymers, filamentous actin and microtubules, there are also protein machines at the heart of energy metabolism, and reading the genetic code. Studies of molecular motors, particularly *in vitro* work with single molecules, have revealed fascinating details about how they convert chemical energy into mechanical work.

While motors are arguably the most machinelike of the biological molecules, they are certainly not the only things inside living cells that remind us of man-made apparatus. The action at a distance that occurs within an allosteric enzyme, for example, is reminiscent of the push rods or levers inside an internal combustion engine. The large, multienzyme complexes that cells use to carry out sequences of reactions remind us of assembly lines. However, molecular

motors are an important special case because the motions they produce are large enough to be directly measured. The study of motor proteins offers rare, direct access to address general questions about how a protein's structure dictates it's dynamics and function.

See also Nucleic Acid and Protein Single Molecule Detection and Characterization.

Bibliography

Books and Reviews

Berg, H.C. (2003) The rotary motor of bacterial flagella, *Annu. Rev. Biochem.* **72**, 19–54.

Block, S.M. (1995) Nanometres and piconewtons: the macromolecular mechanics of kinesin, *Trends Cell Biol.* **5**, 169–175.

Cameron, L.A., Giardini, P.A., Soo, F.S., Theriot, J.A. (2000) Secrets of actin-based motility revealed by a bacterial pathogen, *Nat. Rev. Mol. Cell Biol.* **1**, 110–119.

Howard, J. (2001) *Mechanics of Motor Proteins and the Cytoskeleton*, Sinauer Associates, Publishers, Sunderland, MA.

Inoue, S., Salmon, E.D. (1995) Force generation by microtubule assembly/disassembly in mitosis and related movements, *Mol. Biol. Cell* **6**, 1619–1640.

Salmon, E.D. (1995) VE-DIC light microscopy and the discovery of kinesin, *Trends Cell Biol.* **5**, 154–158.

Schliwa, M. (2003) *Molecular Motors*, Wiley-VCH, Weinheim.

Schliwa, M., Woehlke, G. (2003) Molecular motors, *Nature* **422**, 759–765.

Spudich, J.A. (2001) The myosin swinging cross-bridge model, *Nat. Rev. Mol. Cell Biol.* **2**, 387–392.

Vale, R.D., Milligan, R.A. (2000) The way things move: looking under the hood of molecular motor proteins, *Science* **288**, 88–95.

Yoshida, M., Muneyuki, E., Hisabori, T. (2001) ATP synthase–a marvellous rotary engine of the cell, *Nat. Rev. Mol. Cell Biol.* **2**, 669–677.

Primary Literature

Asbury, C.L., Fehr, A.N., Block, S.M. (2003) Kinesin moves by an asymmetric hand-over-hand mechanism, *Science* **302**, 2130–2134.

Berliner, E., Young, E.C., Anderson, K., Mahtani, H.K., Gelles, J. (1995) Failure of a single-headed kinesin to track parallel to microtubule protofilaments, *Nature* **373**, 718–721.

Block, S.M., Berg, H.C. (1984) Successive incorporation of force-generating units in the bacterial rotary motor, *Nature* **309**, 470–472.

Burgess, S.A., Walker, M.L., Sakakibara, H., Knight, P.J., Oiwa, K. (2003) Dynein structure and power stroke, *Nature* **421**, 715–718.

Case, R.B., Pierce, D.W., Hom-Booher, N., Hart, C.L., Vale, R.D. (1997) The directional preference of kinesin motors is specified by an element outside of the motor catalytic domain, *Cell* **90**, 959–966.

Diez, M., Zimmermann, B., Borsch, M., Konig, M., Schweinberger, E., Steigmiller, S., Reuter, R., Felekyan, S., Kudryavtsev, V., Seidel, C.A., Graber, P. (2004) Proton-powered subunit rotation in single membrane-bound F0F1-ATP synthase, *Nat. Struct. Mol. Biol.* **11**, 135–141.

Dogterom, M., Yurke, B. (1997) Measurement of the force-velocity relation for growing microtubules, *Science* **278**, 856–860.

Finer, J.T., Simmons, R.M., Spudich, J.A. (1994) Single myosin molecule mechanics: piconewton forces and nanometre steps, *Nature* **368**, 113–119.

Hancock, W.O., Howard, J. (1998) Processivity of the motor protein kinesin requires two heads, *J. Cell Biol.* **140**, 1395–1405.

Henningsen, U., Schliwa, M. (1997) Reversal in the direction of movement of a molecular motor, *Nature* **389**, 93–96.

Howard, J., Hudspeth, A.J., Vale, R.D. (1989) Movement of microtubules by single kinesin molecules, *Nature* **342**, 154–158.

Hua, W., Young, E.C., Fleming, M.L., Gelles, J. (1997) Coupling of kinesin steps to ATP hydrolysis, *Nature* **388**, 390–393.

Ishijima, A., Kojima, H., Funatsu, T., Tokunaga, M., Higuchi, H., Tanaka, H., Yanagida, T. (1998) Simultaneous observation of individual ATPase and mechanical events by a single myosin molecule during interaction with actin, *Cell* **92**, 161–171.

Lang, M.J., Fordyce, P.M., Block, S.M. (2003) Combined optical trapping and single-molecule fluorescence, *J. Biol.* **2**, 6.

Lister, I., Schmitz, S., Walker, M., Trinick, J., Buss, F., Veigel, C., Kendrick-Jones, J. (2004) A monomeric myosin VI with a large working stroke, *EMBO J.* **23**, 1729–1738.

Loisel, T.P., Boujemaa, R., Pantaloni, D., Carlier, M.F. (1999) Reconstitution of actin-based motility of Listeria and Shigella using pure proteins, *Nature* **401**, 613–616.

Lombillo, V.A., Stewart, R.J., McIntosh, J.R. (1995) Minus-end-directed motion of kinesin-coated microspheres driven by microtubule depolymerization, *Nature* **373**, 161–164.

Mallik, R., Carter, B.C., Lex, S.A., King, S.J., Gross, S.P. (2004) Cytoplasmic dynein functions as a gear in response to load, *Nature* **427**, 649–652.

Mehta, A.D., Rock, R.S., Rief, M., Spudich, J.A., Mooseker, M.S., Cheney, R.E. (1999) Myosin-V is a processive actin-based motor, *Nature* **400**, 590–593.

Meyhofer, E., Howard, J. (1995) The force generated by a single kinesin molecule against an elastic load, *Proc. Natl. Acad. Sci. U S A* **92**, 574–578.

Nishizaka, T., Oiwa, K., Noji, H., Kimura, S., Muneyuki, E., Yoshida, M., Kinosita, K. Jr. (2004) Chemomechanical coupling in F1-ATPase revealed by simultaneous observation of nucleotide kinetics and rotation, *Nat. Struct. Mol. Biol.* **11**, 142–148.

Noji, H., Yasuda, R., Yoshida, M., Kinosita, K. Jr. (1997) Direct observation of the rotation of F1-ATPase, *Nature* **386**, 299–302.

Ray, S., Meyhofer, E., Milligan, R.A., Howard, J. (1993) Kinesin follows the microtubule's protofilament axis, *J. Cell Biol.* **121**, 1083–1093.

Rice, S., Lin, A.W., Safer, D., Hart, C.L., Naber, N., Carragher, B.O., Cain, S.M., Pechatnikova, E., Wilson-Kubalek, E.M., Whittaker, M., Pate, E., Cooke, R., Taylor, E.W., Milligan, R.A., Vale, R.D. (1999) A structural change in the kinesin motor protein that drives motility, *Nature* **402**, 778–784.

Ryu, W.S., Berry, R.M., Berg, H.C. (2000) Torque-generating units of the flagellar motor of Escherichia coli have a high duty ratio, *Nature* **403**, 444–447.

Sakakibara, H., Kojima, H., Sakai, Y., Katayama, E., Oiwa, K. (1999) Inner-arm dynein c of Chlamydomonas flagella is a single-headed processive motor, *Nature* **400**, 586–590.

Schafer, D.A., Gelles, J., Sheetz, M.P., Landick, R. (1991) Transcription by single molecules of RNA polymerase observed by light microscopy, *Nature* **352**, 444–448.

Schnitzer, M.J., Block, S.M. (1997) Kinesin hydrolyses one ATP per 8-nm step, *Nature* **388**, 386–390.

Shaevitz, J.W., Abbondanzieri, E.A., Landick, R., Block, S.M. (2003) Backtracking by single RNA polymerase molecules observed at near-base-pair resolution, *Nature* **426**, 684–687.

Svoboda, K., Block, S.M. (1994) Force and velocity measured for single kinesin molecules, *Cell* **77**, 773–784.

Svoboda, K., Schmidt, C.F., Schnapp, B.J., Block, S.M. (1993) Direct observation of kinesin stepping by optical trapping interferometry, *Nature* **365**, 721–727.

Vale, R.D., Reese, T.S., Sheetz, M.P. (1985) Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility, *Cell* **42**, 39–50.

Vale, R.D., Funatsu, T., Pierce, D.W., Romberg, L., Harada, Y., Yanagida, T. (1996) Direct observation of single kinesin molecules moving along microtubules, *Nature* **380**, 451–453.

Veigel, C., Wang, F., Bartoo, M.L., Sellers, J.R., Molloy, J.E. (2002) The gated gait of the processive molecular motor, myosin V, *Nat. Cell Biol.* **4**, 59–65.

Yasuda, R., Noji, H., Kinosita, K. Jr., Yoshida, M. (1998) F1-ATPase is a highly efficient molecular motor that rotates with discrete 120 degree steps, *Cell* **93**, 1117–1124.

Yasuda, R., Noji, H., Yoshida, M., Kinosita, K. Jr., Itoh, H. (2001) Resolution of distinct rotational substeps by submillisecond kinetic analysis of F1-ATPase, *Nature* **410**, 898–904.

Yildiz, A., Tomishige, M., Vale, R.D., Selvin, P.R. (2004) Kinesin walks hand-over-hand, *Science* **303**, 676–678.

Yildiz, A., Forkey, J.N., McKinney, S.A., Ha, T., Goldman, Y.E., Selvin, P.R. (2003) Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization, *Science* **300**, 2061–2065.

Yin, H., Wang, M.D., Svoboda, K., Landick, R., Block, S.M., Gelles, J. (1995) Transcription against an applied force, *Science* **270**, 1653–1657.

Mucosal Vaccination

W. Olszewska and Peter J. M. Openshaw
Imperial College, Paddington, London, UK

1 Introduction 603

2 Goals of Mucosal Vaccination 603

3 Benefits of Mucosal Vaccination 606
3.1 Main Features of the Common Mucosal Immune System 606
3.2 Distinctive Characteristics of Mucosal Immunity 607
3.3 Multivalent Mucosal Vaccines 608
3.4 Edible Vaccines 608
3.5 Overcoming Preexisting Immunity or Tolerance 609

4 Challenges for Mucosal Immunization 609
4.1 Mucosal Delivery Systems 609
4.1.1 Live Bacterial Vectors 609
4.1.2 Virosomes 610
4.1.3 Mucosal DNA Vaccines 611
4.2 Mucosal Adjuvants 611
4.2.1 Biodegradable Polymeric Particles 611
4.2.2 Bacterial Toxins 612
4.2.3 CpG Oligodinucleotides 613
4.2.4 Cytokines and Chemokines 613
4.2.5 Saponins 613
4.2.6 Immune Stimulating Complexes (ISCOMS) 614
4.2.7 MF59 615

5 Vaccination via the Respiratory Tract 615
5.1 Applications of Nasal Vaccination 615

6 Oral Vaccines 617

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

7 Conclusions 617

Acknowledgment 617

Bibliography 618
Books and Reviews 618
Primary Literature 618

Keywords

Adjuvant
(Latin: *adjuvere* to help) substances added to vaccines to increase immune responses.

Antigen
Foreign substance recognized by immune system.

B Cells
Lymphocytes that make immunoglobulin (antibody); can also present antigens.

T Cells
Lymphocytes that require the thymus for development.

Th1 Cells
Typically make interferon (IFN)-gamma and tumor necrosis factor (TNF) and activate macrophages to kill intracellular pathogens and switch on IgG1 production from B cells.

Th2 Cells
Make cytokines such as IL-4, IL-5, and IL-13, which promote growth, activity, and survival of eosinophils and switch B cells to make IgE.

Chemokines
Chemoattractive proteins made by many cell types that cause cell recruitment to inflamed tissues.

Cytokines
Small proteins produced by T cells that act as signals to other cells of the immune system or structural cells (e.g. IL-1, IL-3, IL-4, IL-5, IL-10, and IFN).

Lymphocytes
White blood cells that concentrate in lymph nodes, spleen, and so on.

Vaccine
A substance administered to trigger memory immune responses that protect against disease.

The immune system faces a dilemma. It must tolerate benign commensal organisms and antigens present in food and air, while rapidly mounting vigorous responses to the plethora of pathogens that enter via mucosal surfaces. The existing mucosal vaccines against poliomyelitis, influenza, and measles were developed on a largely empirical basis against transient self-limiting infections that themselves induce lifelong immunity.

In this article, we review the successes and failures of mucosal immunization and the opportunities for exploiting developments in immunology to create new and effective mucosal vaccines. The challenge now is to exploit new information about immunoregulation to develop effective vaccines against more subtle agents that circumvent lifelong immunity, recur, or persist.

1 Introduction

The mucous membranes lining the lung, gut, and urogenital tract present unique problems to the host defense systems. They have specialized functions that necessitate close contact with the environment, yet most common infections enter by these routes. Immune defenses at mucosal surfaces therefore have to defend against pathogens while being tolerant of nonthreatening substances in food and inhaled air. Both innate and acquired mechanisms of immunity are involved in preventing microbial entrance and spread. Mechanisms of acquired immunity combine humoral (production of antigen-specific secretory IgA, S-IgA) and cell-mediated immune responses (Fig. 1).

Mucosal immunization has the potential to induce protective immunity against infectious diseases or to elicit antigen-specific tolerance. The mucosal immune system has to be appropriately activated to achieve effective protection against colonization and invasion by infectious agents at mucosal surfaces. Although administration of vaccines directly to mucosal sites has many advantages, mucosal immunity can be also achieved by other routes of antigen delivery. However, in this review we focus primarily on the mucosal route of vaccination.

Because of the low absorption efficiency of mucosally delivered vaccines, almost all current vaccines are administered parenterally. In addition, suboptimal immune responses are frequently induced by mucosal immunization and the use of mucosal adjuvants is required. As a result, development of successful mucosal vaccines depends largely on improvements to mucosal delivery systems and on the discovery of new and effective mucosal adjuvants (Fig 2).

2 Goals of Mucosal Vaccination

Most environmental pathogens enter the body through the mucosal membranes of the intestinal, respiratory, or genital tract. Some only replicate in the mucosa, but others use the mucosa to gain entry before dissemination to other sites. For purely mucosal infections (e.g. those caused by most common cold agents and infective diarrheal organisms), prevention of initial

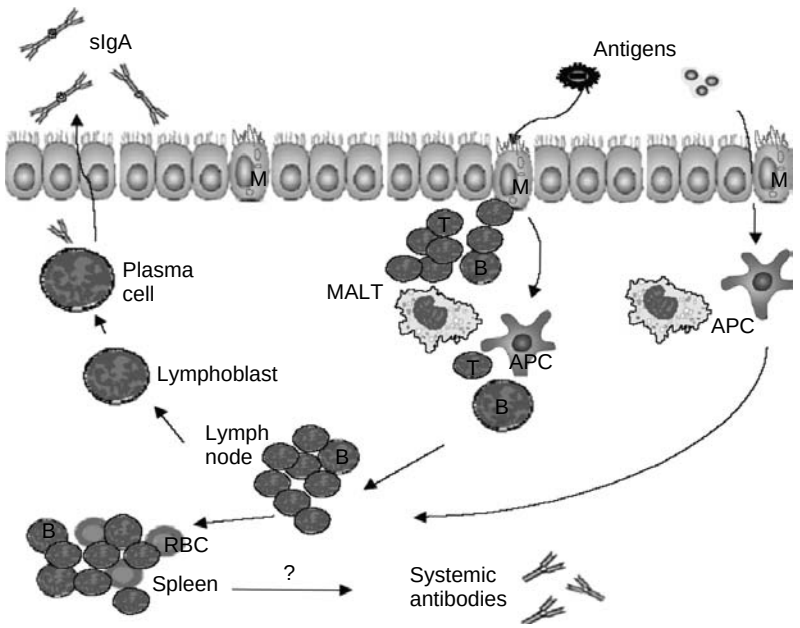

Fig. 1 Schematic representation of cells involved in immune reactions after mucosal challenge. APC, antigen-presenting cell; M, microfold epithelial cell; T, T cell; B, B cell; RBC, red blood cell.

colonization is an essential but difficult aim of vaccination; when dissemination or toxin production is essential to disease pathogenesis, prevention of surface invasion may not be necessary.

Mucosal vaccination is an attractive administration route for mass vaccination, as it does not require trained medical personnel and does not involve needles and syringes These factors affect the costs of vaccination and the risk of transmitting blood-borne infections (HIV or HepB) making mucosal vaccine particularly attractive for use in developing countries (Table 1).

New or improved vaccines are needed for a wide range of mucosal infections, including respiratory tract infections caused by *Mycobacterium* spp, *Mycoplasma pneumoniae*, influenza virus, rhinovirus, coronavirus, adenovirus, human metapneumovirus, and respiratory syncytial virus (RSV); urogenital tract infections caused by *Chlamydia*, HIV, *Neisseria gonorrhoeae*, *Treponema pallidum*, and herpes simplex virus (HSV); and gastrointestinal infections caused by *Escherichia coli*, *Salmonella*, *Shigella* spp, *Helicobacter pylori*, *Vibrio cholerae*, *Campylobacter jejuni*, *Clostridium difficile*, and rotaviruses. The challenge is to design vaccine preparations that induce neutralizing immunity to the pathogen or their toxins, preventing their attachment to mucosal surfaces, tissue invasion, and spread (Fig 3).

Induction of specific secretory IgA (S-IgA) is an essential aim of mucosal vaccination. Locally produced S-IgA is considered to be among the most important protective humoral immune factors and constitutes over 80% of all antibodies produced in mucosa. In humans, S-IgA

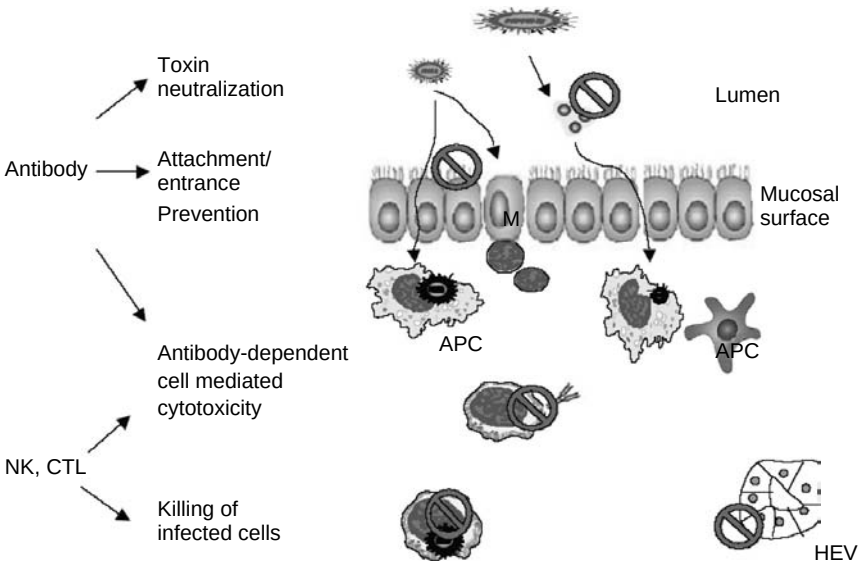

Fig. 2 Goals of mucosal immunization. Successful mucosal immunization may induce humoral and cellular immune responses. Possible roles of both are shown. M, microfold epithelial cell; HEV, high endothelial venule; APC, antigen-presenting cell; the Ø symbol indicates possible sites of immune intervention.

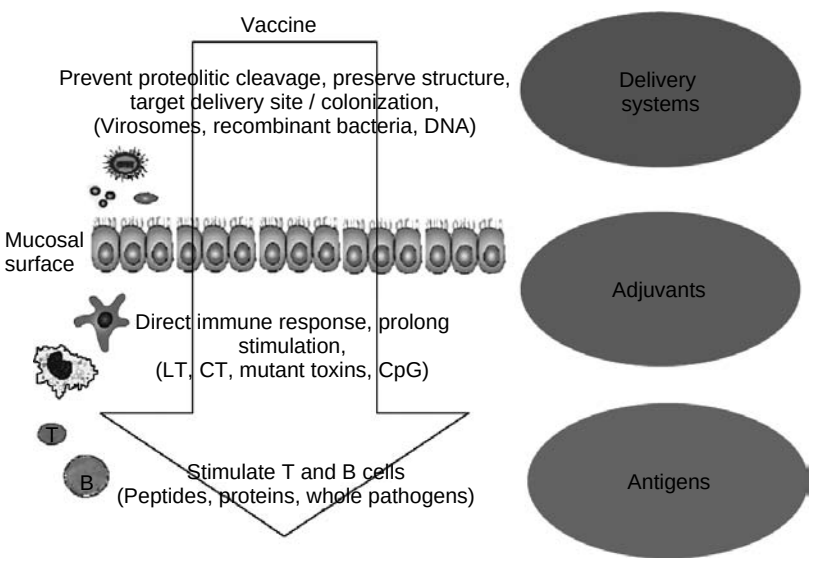

Fig. 3 Challenges for mucosal vaccination: schematic representation of key issues in mucosal vaccine development.

Tab. 1 Comparison of mucosal and parenteral delivery of vaccines.

Vaccination	*Mucosal*	*Parenteral*
Administration	Easy, self administration	Need of trained personnel
Risks	Nonaccurate dosing	Infection (HIV, Hepatitis B)
Antigen dose	Usually high	Low
Formulation	Need of mucosal adjuvant	Mixed with Alum
Immune responses	Stimulates mucosal immunity but may induce tolerance	Rarely induces mucosal responses but good systemic antibody and T cell responses
Use in human	Only few vaccines	Most of the vaccines

antibodies are usually dimeric whereas serum IgA is mainly monomeric. Antibodies secreted by mature plasma cells in various compartments of the common mucosal immune system are able to interact with invading pathogens, inhibit their attachment, and form immune complexes with potentially harmful molecules (immune exclusion). S-IgA may neutralize viruses and directly participate in antibody-dependent cell-mediated cytotoxicity (ADCC) in collaboration with macrophages and lymphocytes. However, natural killer cells and specific cytotoxic T cells are also key defenses in preventing infection with intracellular pathogens (Fig. 2). In testing vaccine efficacy, both antibody levels and cellular immunity should therefore be monitored.

3 Benefits of Mucosal Vaccination

3.1 Main Features of the Common Mucosal Immune System

A key issue in mucosal immunity is the necessity to distinguish between normal antigens, which are not harmful, and those that belong to dangerous pathogens. The discrimination between hazardous and nonhazardous antigens is determined by activation by pattern-recognition receptors (including Toll-like receptors), which may be displayed in intestinal crypts or in the cytoplasm of mucosal cells, and are therefore inaccessible to nonhazardous commensals. The anatomical placement of pattern-recognition receptors in inaccessible sites appears to be an important factor in allowing discrimination between hazardous and nonhazardous antigens.

The mucosal immune system is highly adapted toward tolerance, the breakdown of which can result in disease (e.g. celiac disease). Mucosal exposure frequently induces regulatory T cells, of which there are several distinct types. First, Th3 cells produce transforming growth factor beta (TGF-beta); these are frequently induced by repeated feeding of low doses of antigen orally. Second, Tr1 cells are a subset of CD4 T cells, which produce IL-10, the production of which is promoted by IL-15 and Type 1 interferon. Third, CD4+ CD25+ cells are potent regulators of autoreactivity *in vivo* and can also be induced by tolerogenic oral feeding. Fourth, CD8 T cells often have a regulatory role in the gut, but their exact functions have not been clearly defined. In addition,

there is mounting evidence that gamma delta T cells may, under some conditions, also promote a tolerogenic state after oral feeding. All these subsets are readily induced by oral administration of antigen, but induction must be avoided if mucosal vaccination is to be successful.

3.2 Distinctive Characteristics of Mucosal Immunity

In mammals, immune responses at mucosal surfaces are provided by a defence system known as *mucosa-associated lymphoid tissue* (MALT). It comprises an integrated network of cells and molecules anatomically grouped and functionally divided into sites of antigen uptake and effector function. Mucosal immune responses are triggered in specialized zones that sample foreign material from the epithelial surface. The main inductive site of the upper respiratory tract is in the nasal-associated lymphoid tissue (NALT), and in the gastrointestinal tract it is the Peyer's patch (PP), the appendix, and the solitary lymphoid nodules, collectively called the *gut-associated lymphoid tissue* (GALT).

The respiratory epithelium contains four different cell types: alveolar macrophages, dendritic cells, M cells, and intraepithelial lymphocytes. The latter are relatively scarce in the respiratory tract. M cells are associated with lymphoid structures and differ in morphology from absorptive cells by their short microvilli, small cytoplasmic vesicles, and few lysosomes. The M cells selectively absorb antigen by endocytosis or pinocytosis and direct it to professional APC (macrophages, dendritic cells, B lymphocytes) or directly process and present antigen to T cells (Fig. 1). In the gut, specialized M cells are thought to provide the main route by which complex antigens gain access to the immune system. Lymphocytes primed in the Peyer's patches express $\alpha_4\beta_7$ integrin, which binds to the mucosal adressin cell-adhesion molecule 1 (MAD-CAM1), which is expressed at high levels by the vascular endothelium in mucosal surfaces. Gut-derived T cells also express the chemokine receptor CCR9, which allows them to respond to the chemokine CCL25, which is expressed selectively by small bowel epithelial cells. By contrast, T cells that are primed peripherally typically display the $\alpha_4\beta_1$ integrin and CCR4 and so do not migrate or respond in mucosal sites. This selective expression explains why mucosal vaccination is often required to protect against mucosal infections, and why peripheral administration of vaccine antigens is often ineffective against mucosal infections. The establishment and maintenance of the immune response requires close cooperation between many cell types. Activated T cells help B cells to develop into plasma cells, a process that depends on the nature of the antigen and on the cytokines produced by T helper cells. TGF-β, IL-10, and IL-4 promote the switch from IgM to IgA production from stimulated B cells. Precursors of mucosal plasma cells derive from lymphoepithelial structures, mature in the regional lymph nodes, and enter the circulation via the thoracic duct. Then they can seed the lamina propria of distant mucosal sites (e.g. intestines, respiratory tract, genital tract, salivary gland, etc.), where they differentiate. In this way, antibodies can appear at different mucosal sites elsewhere within the common mucosal immune system. Natural killer cells may be also present in mucosal tissues and, together with

CTLs, are important for eliminating virus-infected cells.

There is strong evidence that various mucosal sites of the immune system can communicate. For instance, within 24 h of adoptive lymphocyte transfer, mesenteric lymph node (MLN) cells are found in recipient gut, cervix and vagina, uterus, and mammary glands. The ability of cells to migrate between mucosal zones has been extensively investigated. Thus, intranasal immunization may induce specific immunity in intestine, and boosted levels of IgA have been reported in the intestine after intranasal (rather than oral) administration. In these experiments, the relatively poor induction of gut immunity by oral administration could be due to degradation of antigen and dilution in the contents of the gastrointestinal tract.

Several examples show that intranasal immunization can result in efficient immunity at distant mucosal sites. For instance, intranasal administration of a recombinant HIV envelope protein formulated in CTB-associated GM1 lipid vesicles enhances mucosal IgA antibody responses in nasal and gut tissues. Administration of a DNA vaccine for herpes simplex induces antigen-specific cellular and secretory IgA responses in the gut, vagina, and oral cavity after intradermal, intraperitoneal, intravaginal, intranasal, or oral immunization. Taking advantage of the functional integrity of the common mucosal system, vaccines easily administrated into the nose may in the future include those against herpes simplex virus or HIV. For instance, recombinant adenovirus used to deliver antigens derived from HSV and HIV has been demonstrated to induce protective immunity in mice vaginal mucosa, particularly when cholera toxin was included in the vaccine preparation.

3.3 Multivalent Mucosal Vaccines

Delivery systems that allow simultaneous delivery of several antigens derived from one or more pathogens may be used to design a multivalent vaccine. Such a vaccine potentially may reduce the number of administrations needed and increase protection in the populations studied.

For example, RSV and human parainfluenza represent two of the most important viral agents of pediatric respiratory tract disease worldwide. Studies are under way that use recombinant bovine/human parainfluenza virus type 3 (rB/HPIV3), with bovine F and HN genes replaced with their HPIV3 counterparts and expressing the major surface antigens of respiratory syncytial virus. Both recombinant viruses were shown to replicate efficiently in the respiratory tract of hamsters and induced serum antibody titers similar to those induced by RSV or HPIV3 infection. Immunization of hamsters with rB/HPIV3-G1, rB/HPIV3-F1, or a combination of both viruses resulted in high level of resistance to challenge with RSV or HPIV3 28 days later. Schmidt et al. described a similar strategy for intranasal immunization against RSV subgroups A and B and human parainfluenza virus type 3 by using a live cDNA-derived vaccine in monkeys.

3.4 Edible Vaccines

Transgenic plant technology allows the production of large quantities of transgenic antigen from pathogenic microorganisms at low cost and would have particular benefits in underresourced areas. The first edible vaccine was obtained by inserting the gene encoding hepatitis B surface

antigen into tobacco plants, resulting in an antigen very similar to that obtained from recombinant yeast. Similar methods have been used to make vaccines containing genes of *E. coli* heat-labile enterotoxin B (LT-B), rabies virus glycoprotein, F protein of RSV, Norwalk virus capsid protein, and *V. cholerae* (CT-B) in potatoes, tomatoes, and other plants. Materials from recombinant plants have been shown to be immunogenic in feeding experiments in animals, inducing systemic and mucosal responses, and promising clinical trials show that specific anti-LT-B responses can be demonstrated in humans given transgenic potatoes expressing LT-B. The possibility of using transgenic plants for the production of specific immunoglobulins for passive immunotherapy is also being explored, but the commercial viability of these technologies is unproven.

Oral administration could theoretically induce tolerance to pathogens, resulting in serious adverse consequences. Fear of this effect has limited the clinical trials of oral vaccination against hepatitis B with edible potatoes expressing hepatitis surface antigens to preimmune individuals.

3.5 Overcoming Preexisting Immunity or Tolerance

Immunity to vaccinia potentially limits the utility of vaccinia vectors in people previously immunized against smallpox. However, the studies of Belyakov et al. suggest that modified vaccinia Ankara (MVA) expressing HIV gp160 could successfully immunize via the mucosal route and induce potent specific systemic humoral and CTL responses in vaccinia-immune mice. It therefore seems possible that immune responses at mucosal sites may not be prevented by prior systemic immunization and that this mucosal naiveté may be exploited in generating immune responses in previously vaccinated individuals.

Another challenge to immunization against RSV and measles is the inefficiency of vaccination in the presence of specific maternal antibody. It appears that mucosal immunization may overcome this difficulty. Mutwiri et al. concluded from studying immune responses in neonatal lambs that enteric immunization with a human adenovirus vector may be an effective approach for inducing both mucosal and systemic immune responses in neonates. Similarly, in the cotton rat model of measles mucosal vaccination with vesicular stomatitis virus expressing the hemagglutinin of measles virus induces seroconversion in the presence of maternal antibodies and leads to protection against measles challenge. On the basis of experience with highly effective live polio vaccine given at birth, it has been proposed that mucosal immunization may be safer and more effective than any other route of vaccination in young children.

4 Challenges for Mucosal Immunization

4.1 Mucosal Delivery Systems

4.1.1 Live Bacterial Vectors

Commensal flora as expression vectors
Genetic manipulation of normal surface bacteria is attractive in that no pathogen is deployed in vaccine production and the bacteria themselves can exhibit probiotic properties. However, the risk is that such bacteria will be tolerated and thus no immune response will be generated. Considerable work has been done

with *Lactococcus lactis*, some strains of *Lactobaccillus*, *Staphylococcus carnosus*, and *Streptococcus gordonii*. These bacteria are noninvasive or commensal organisms in healthy people. Internal adjuvanticity is provided by peptidoglycan, and lactobacilli tagged with green fluorescent protein are actively taken up by APC after intranasal administration, suggesting that antigens would be processed and presented. Recombinant antigens that have been expressed in commensal bacteria include the V3 domain of HIV-1 gp120, fusion and hemagglutinin from measles virus, fragment B of diphtheria toxin, peptides from *Plasmodium falciparum*, and epitopes from RSV.

Pathogens as expression vectors Many bacterial pathogens naturally invade via mucosal surfaces and induce strong mucosal immunity. Attenuated *Salmonella* has been extensively studied as a vector both for oral and intranasal delivery. Interestingly, i.n. vaccination induces better systemic and local immune responses to inserted antigens (e.g. hepatitis B core antigen) than oral or rectal immunization. Insertion of antigens from *Helicobacter pylori* into *Salmonella*, for instance, results in induction of specific CD4+ T cells producing IFN (interferon) and IL-10 in a mouse model, as well as good protection after just two i.n. doses of vaccine.

Attenuated strains of *Shigella* are also promising live bacterial vectors. In animal models, they were shown to induce specific serum antibody responses (IgG and IgA) to inserted antigens after i.n. immunization.

Bacille Calmette-Guérin (BCG) is the first attenuated vaccine introduced to humans and still is the most widely used vaccine in the world, since it is the only one available against tuberculosis. Since BCG induces cellular immunity and can accommodate foreign epitopes, it may work as a successful vector when specific cellular responses are needed. The safety of the vaccine given at birth suggests its relevance for immunizations is required very early in life. Candidates would therefore include measles and RSV vaccine. Indeed, recombinant BCG producing measles nucleoprotein provided protection against virus challenge in intranasally immunized infant rhesus macaques. Other examples of heterologous antigens expressed in BCG for mucosal vaccination include HIV, human papilloma virus, *Schistosoma haematobium*, *Plasmodium yoelii*, and *Toxoplasma gondi*.

Recently, attenuated *Bordetella pertussis* emerged as a bacterial vector for i.n. vaccination. Deletion of genes coding for pertussis toxin diminished the virulence of the bacteria but preserved their ability to colonize mucosal sites. Interestingly, the immunogenicity to the mayor antigen, filamentous hemagglutinin (FHA) was increased, as well as that to heterologous antigen fused to FHA. Additionally, significant protection was observed in a mouse model of *Schistosoma mansoni* infection following single i.n. immunization.

4.1.2 Virosomes

Cusi et al. investigated the efficacy of a vaccine composed of the RSV fusion protein associated with influenza virosomes (IRIV), which was administered intranasally together with *E. coli* heat-labile toxin (LT). After an intramuscular "priming" with influenza virus vaccine, mice were i.n. immunized with RSV-F/IRIV + LT or with RSV-F + LT or IRIV + LT. The results showed that mice immunized with RSV-F + LT developed Th2 type responses and that virosomal delivery greatly potentiated immune responses in animals. All mice immunized with RSV-F/IRIV + LT developed a balanced Th1/Th2 cytokine

profile with mucosal IgA and high levels of serum IgG. More importantly, histological analysis of lung tissue of RSV-challenged mice did not reveal vaccine-enhanced pulmonary eosinophilia.

4.1.3 Mucosal DNA Vaccines

The immunogenic potential of DNA vaccination has been extensively tested in animal models. DNA vaccination has many theoretical advantages, including ease of production, stability of vaccine preparations, and the ability to induce both antibody and cell-mediated immune responses. The drawbacks are that success in animal models is poorly predictive of outcome in human studies and that long-term follow-up studies are necessary to assess the risk of vaccine involvement in autoimmune diseases or immune disregulation.

DNA is not usually taken up in non-degraded form from mucosal surfaces. However, DNA immunization in mice and chickens has been shown to induce antibodies to influenza nucleoprotein and to trigger both mucosal and systemic cellular protective immune responses. In a recent study, Sasaki et al. described immune responses induced by intranasally and intramuscularly delivered DNA encoding HIV-1 proteins in mice. Both routes produced similar levels of cell-mediated immunity, but intranasal immunization induced higher levels of intestinal S-IgA than intramuscular immunization, and the adjuvant QS21 enhanced both intranasal and intramuscular immune responses.

Encapsulation of DNA into microparticles prevents degradation of DNA and enhances immunogenicity; therefore, it is an attractive delivery system for a mucosal DNA vaccine. Such a vaccine, designed using rotavirus VP6 DNA encapsulated in PLG microparticles, was shown to be effective in inducing systemic and mucosal immunity after oral administration to mice.

4.2 Mucosal Adjuvants

Many substances are known to have adjuvant properties; however, the majority are used for parenteral immunization and only a few are suitable for the mucosal route. Alum, the only universally licensed adjuvant for use in humans, is not a suitable adjuvant for mucosal immunization. The choice of an appropriate adjuvant for mucosal vaccination is very important to its success. Many mucosal adjuvants are based on bacterial toxins and their derivatives, CpG-containing DNA, and various cytokines and chemokines. Since most antigens are poorly immunogenic when introduced via the mucosal route if no adjuvant is added, induction of tolerance is likely.

4.2.1 Biodegradable Polymeric Particles

Micro- and nanoparticles may be used to encapsulate vaccine antigens. Many such polymers have the advantages of being biodegradable and at the same time of protecting the antigen from premature degradation. Slow degradation of polymers controls the release of entrapped antigen; therefore, such vaccines can induce immune responses over prolonged periods of time. Poly(D,L-lactide-co-glycolide) (PLG) polymer is probably the most extensively studied polymer because of its safety record in humans. The hydrolysis of PLG polymer to release encapsulated antigen is controlled by the polymer composition and its molecular weight. PLG particles may be suitable mucosal vaccine carriers, as shown by using fimbriae from

B. pertussis encapsulated in PLG particles for oral delivery to mice. The technique was able to induce serum and mucosal antibody responses in mice as well as to protect against live bacterial challenges. Baras et al. demonstrated recently the feasibility of long-term *in vivo* release of antigens from PLG particles. They showed that a single nasal or oral immunization of glutathione S-transferase from *S. mansoni* in PLG microparticles could induce a long-lasting antigen-specific antibody response in mice, with a peak at 9 to 10 weeks after immunization.

4.2.2 Bacterial Toxins

Bacterial toxins have been used for a long time as adjuvants in experimental models, and some chemically detoxified toxins have been employed to prevent bacterial infectious diseases (e.g. formalin inactivation of *Corynebacterium diphtheriae* or *Clostridium tetani* exotoxins). Although bacterial toxins possess excellent adjuvant properties, they have been prohibited from wide use as adjuvants in humans because of their high toxicity. Two bacterial toxins have found particular attention for application as mucosal adjuvants: cholera toxin (CT) produced by *V. cholerae* and heat-labile enterotoxin (LT) produced by *E. coli* have similar structures, since both contain subunits A and B and the toxicity originates from the A subunit, which catalyzes ADP-ribosylation of the stimulatory GTP-binding proteins on the surface of the epithelial cells and raises the intracellular levels of cAMP. This then leads to secretion of water and electrolytes into the mucosal lumen. Today it is possible to obtain detoxified derivatives by mutagenesis of the toxin genes. With this technology, the genes are modified in such a way as to encode different amino acid(s) that are no longer able to function in enzymatic activity. Such inactivated substances are safe and in the future could replace toxoids in existing vaccines as well as be used as mucosal adjuvants in new vaccination strategies.

Several mutant toxins have already been described; among them are mutations of the heat-labile toxin of *E. coli*: LTK63, LTR72, and LTG192. All of them have been studied in detail for their ability to induce systemic and local immune responses to coadministered antigen. LTR72 was the strongest adjuvant, as compared with the fully nontoxic LTK63 mutant, and showed only 0.6% of the enzymatic activity of the wild-type LT. LTK63 was shown to be 100 000-fold less toxic but 20 times less effective than LT. Moreover, CD4+ lymphocytes from animals immunized with ovalbumin together with LTR72 exhibited very strong proliferative responses, similar to those induced by wild type LT. LTK63 was also tested in the murine measles model, and mucosal coimmunization with the mutant and a synthetic peptide representing a CTL epitope from measles N protein was very effective in *in vivo* priming of peptide-specific and measles virus–specific CTL responses. Similarly, the addition of LTK63 to peptide RSV vaccine induced strong CTL responses and protection after intranasal administration into a mouse. Compared with administration of oral influenza vaccine alone, coadministration of vaccine with LTG192 provided enhanced protection from infection in the upper and lower respiratory tract, equivalent to and at similar doses as that obtained with wild-type LT. The mutant toxin augmented virus-specific IgG and IgA responses in serum, lung, and nasal washes and also the numbers of virus-specific antibody-forming cells in spleen, lung, and Peyer's patches in a manner comparable to that of wild-type LT.

Studies conducted in experimental animals have given promising results with respect to immune responses developed after intranasal vaccination by a number of delivery systems. Cholera toxin B given intranasally with measles virus stimulated systemic neutralizing antibodies, and intranasal immunization with a chimeric synthetic peptide containing two T-helper epitopes (MVF: aa 288–302) and one B-cell epitope (MVF: aa 404–414) produced both systemic and mucosal antibody responses that conferred protection against encephalitis after infection with neuroadapted measles virus.

A novel nontoxic form of chimeric mucosal adjuvant that combines the A subunit of mutant cholera toxin E112K with the pentameric B subunit of heat-labile enterotoxin from enterotoxigenic *E. coli* was constructed. Nasal immunization of mice with tetanus toxoid (TT) plus the new adjuvant elicited significant TT-specific immunoglobulin A responses in mucosal compartments and induced high serum immunoglobulin G and immunoglobulin A anti-TT antibody responses.

The suitability of CT and LT and their mutants in human vaccinology remains to be determined. Several clinical trials did not report side effects; however, at least one influenza vaccine coadministered with LT had to be withdrawn from study due to several instances of facial paresis.

4.2.3 CpG Oligodinucleotides

Synthetic oligodeoxynucleotides (ODN) that contain unmethylated CpG motifs (CpG ODN) are also novel candidates as adjuvants for mucosal immunization. Initially, it was reported that these motifs could induce *in vitro* production of IL-6 and IFNγ by CD4+ T cells, IL-6 and IL-12 by B cells, and IFNγ by NK cells. Such properties led to the use of CpG ODNs as an adjuvant in several experimental models and indeed, work published so far supports the view that Th1-type responses dominate after CpG coadministration with an immunogen. However, some authors indicate that these responses are very much dependent on the age of the primed animals, the route of antigen delivery, or the nature of the antigen. The potential of CpG motifs as adjuvants for delivery via mucosal surfaces is particularly promising, as can be judged from several recent publications.

4.2.4 Cytokines and Chemokines

Cytokines can be used as mucosal adjuvants, either added directly to a vaccine preparation or encoded in DNA. Recent advancements in cytokine applications are summarized in Table 2. As an example, combination of IL-1 and IL-12 or IL-18 and GM-CSF creates a potent stimulating environment for mixed Th1/Th2 immune responses with the effect of IFNγ, IgA, and induction of CTL responses. IL-2 and IL-6, coexpressed in recombinant bacteria, significantly increase specific antibody in serum as well as in mucosal IgA subsequent to i.n. inoculation in mice. Likewise, some chemokines are being explored for use to potentiate mucosal immunity, as reported by Lillard et al. and Eo et al. Particularly promising is RANTES, which is a chemoattractant for monocytes, T cells, and NK cells, and has strong ability to induce Th1 responses, particularly CTLs. Moreover, nasal coadministration of RANTES with a protein antigen was demonstrated to augment Th1 and Th2 local and systemic immune responses.

4.2.5 Saponins

Saponins purified from the bark of the tree *Quillaja saponaria molina* and their

Tab. 2 Examples of cytokines and chemokines coadministered with mucosal vaccines.

Cytokine/ chemokine	*Delivery system*	*Model*	*Route of delivery*	*Effect*
GM-CSF	Adenovector	Mouse	Intranasal	Increased IgG2a and IgG1, elevated levels of IFN-γ, TNF-α, and IL-10
	Plasmid DNA	Mouse	Rectal/ vaginal	Increased serum IgG, enhanced mucosal and fecal IgA
IL-2	Plasmid DNA-encapsulated in PLG microparticles	Mouse	Oral	Enhanced CTL responses
	Recombinant *L. lacti*	Mouse	Intranasal	Increased antibody titers
	Liposomes *s*	Mouse	Intranasal	Increased *P. aeruginosa* polysaccharide-specific pulmonary plasma cells, reduced mortality from pneumonia
IL-4	Plasmid DNA	Mouse	Rectal/ vaginal	Increased antibody levels; decreased CTL activity
IL-6	Recombinant *L. lactis*	Mouse	Intranasal	Increased antibody titers
IL-12	Liposomes	Mouse	Oral	Shift to IgG2a and IgG3, decreased IgE Abs, enhanced serum IFN-γ
	Plasmid DNA	Mouse	Rectal/ vaginal	Decreased antibody levels, enhanced CTL activity
IL-1alpha, IL-12, and IL-18	Peptide	Mouse	Intranasal	Specific anti-HIV IgA in saliva, fecal extracts, and vagina
RANTES	Peptide	Mouse	Intranasal	Augmented Th1 and Th2 responses
IL-10	Plasmid DNA	Mouse	Intranasal	Diminished Ag-induced delayed type hypersensitivity, production of Th1 cytokines
CCR7 ligands	Plasmid DNA	Mouse	Intranasal, intragastric	CD4+ T helper cell proliferation and CD8+ T cell-mediated CTL activity, serum IgG

derivatives are being used experimentally for mucosal immunization. Quil A can be incorporated into more potent adjuvant systems, such as ISCOMs (see below). Onjisaponins were shown in studies of Nagai et al. to be safe and potent adjuvants for intranasal inoculation together with influenza and DTP vaccines. Experiments in mice showed that the use of adjuvant significantly increased serum IgG and nasal IgA as well as inhibited proliferation of mouse-adapted influenza virus in BAL of infected mice.

4.2.6 Immune Stimulating Complexes (ISCOMS)

ISCOMs (immune stimulating complexes) consist of cholesterol, phospholipids, viral proteins, and glycosides of the adjuvant Quil A. For several viruses, it has been shown that the incorporation of viral proteins into the ISCOM structure can

dramatically enhance their immunogenicity. Strong B-cell and T-cell responses are usually observed together with induction of CTL responses, which are normally not evoked by nonreplicating vaccine preparations. ISCOM-based influenza vaccines are currently being evaluated in clinical trials, and initial studies have shown that individuals immunized with ISCOM preparations developed virus-specific CTL responses in addition to strong antibody responses.

Intranasal delivery of inactivated influenza vaccine plus the ISCOMATRIXØ (IMX) adjuvant was able to induce serum hemagglutination inhibition (HAI) titers in mice better than those obtained with nonadjuvanted vaccine delivered subcutaneously. Furthermore, the IMX-adjuvanted vaccine delivered intranasally induced mucosal IgA responses in the lung, nasal passages, and large intestine, together with high levels of serum IgA.

4.2.7 MF59

MF59 is an adjuvant approved for use in humans and elicits higher antibody titers than alum when used in combination with a variety of recombinant and natural subunit antigens. MF59 is an oil-in-water emulsion that contains squalene (a metabolite of cholesterol), polysorbate 80 (a surfactant soluble in water), and sorbitan trioleate (a surfactant soluble in oil). Although the mechanisms responsible for the adjuvant action of MF59 are not fully understood, enhancement of humoral immune response to parenteral influenza vaccine has been shown in humans, and mucosal immune responses to intranasally administered influenza vaccine were evoked in mice. Thus, it appears that MF59 may be a promising formulation for future mucosal vaccines.

5 Vaccination via the Respiratory Tract

The first point of contact for inhaled pathogens is usually the nasal mucosa. Many infections are initiated at the mucosal surface and so it would appear desirable to induce neutralizing antibodies and specific cellular responses at the site of pathogen entry. The large, highly vascularized surface (150 cm^2) has the potential of very efficient absorption of delivered vaccine, and the presence of immune cells in this area enables the initiation of immune reactions. Nasal epithelium absorbs principally soluble antigens, so use of a suitable delivery system is likely to contribute to achieving protective responses after immunization via the nasal route.

The technique of intranasal immunization is very important if large doses are administered. Some vaccine may be swallowed, leading to oral delivery, or, in anesthetized animals, a proportion of the vaccine may reach the lung. The epithelial deposition of antigens in the respiratory mucosa depends on size of particles (aerodynamic diameter); therefore, in designing a vaccine for delivery via the respiratory route particle size is a critical issue.

5.1 Applications of Nasal Vaccination

Many preclinical studies have been conducted on vaccines administered via the respiratory tract. These are mostly delivered into nasal mucosa, but some trials involving deep lung deposition have also been described. As mentioned above, the functional integrity of mucosal system allows induction of immune responses in sites distant from the immunization site. It seems obvious, however, that intranasal

immunization is particularly appropriate for prevention of those infectious diseases acquired by inhalation. Therefore, we focus here on recent developments of nasal vaccines against some respiratory viral diseases.

The protective efficacy of mucosal immunization with purified RSV fusion protein (F) or chimeric FG glycoprotein was shown in animal models when the vaccines were coadministered with CT adjuvant. Complete protection was also demonstrated in mice immunized with a synthetic peptide (residues 174–187) of the G protein mixed with CT, even though the peptide failed to induce a detectable level of secretory IgA. Effective mucosal immunization against RSV was reported after using F protein and a genetically detoxified toxin CT-E29H or CTL peptide from the M2 protein together with LTK63 as an adjuvant.

Recently, live, attenuated, cold-adapted viral vaccines have been developed as alternatives to inactivated vaccines. Thus, it is possible to produce an organism that replicates efficiently at 25 to 28 °C, that is, the temperature of the nasal passage, but not at 37 °C (the temperature of the lungs). Intranasal administration of cold-adapted vaccines usually induces good immune responses – including local IgA responses and secretory IgA antibodies, which can provide protection against pathogens that infect mucosal sites – although the magnitude of the serum antibody response depends on the extent of virus replication. Furthermore, live vaccines can induce CTL responses or can prime CTL responses induced during natural infection.

Several vaccines against influenza are licensed for use in humans. Among them, few are administered via the intranasal route. The trivalent Nasalflu Berna vaccine consists of influenza virosomes formulated from inactivated influenza surface glycoproteins, combined with lecithin and heat-labile toxin of *E. coli*. This vaccine has been reported to induce high levels of influenza-specific hemagglutination inhibition IgG and IgA in nasal mucosa and in the saliva. In clinical trials, 85% efficacy was reported in adults and nearly 90% efficacy in children. Although no significant adverse reactions were observed in initial studies, 43 instances of Bell's palsy were encountered among the first season's vaccine recipients (totaling more than 100 000 people). This resulted in suspension of sales and the launch of a detailed investigation of potential side effects.

Another human trial involving influenza vaccine compared intranasal and intramuscular trivalent whole virus vaccines in elderly people, and showed that the intranasal route was significantly more effective in inducing mucosal IgA response. Additionally, combined vaccination involving intramuscular inactivated and intranasal cold-adapted influenza vaccines had significantly increased efficacy in an elderly population. A different human study involving MF59-adjuvanted or nonadjuvanted subunit influenza vaccines indicated that in these trials immune responses, including mucosal IgA production, were not influenced by the presence of adjuvant.

Interesting results were obtained from a mouse study using a peptide mucosal influenza vaccine. A retro–inverso analog, encompassing the protective B-cell epitope sequence from hemagglutinin (HA) (91–108) conjugated to ovalbumin and coadministered with cholera toxin, produced strong systemic (serum IgG) and mucosal (lung IgA) antibody responses that protected against intranasal

challenge with a lethal dose of influenza virus. The half-life of the retro–inverso analog in the presence of lung homogenate proteases was at least 700 times greater than that of the parent L-peptide. These results demonstrated that peptido-mimetic analogs with high resistance to proteolytic degradation are very effective immunogens for intranasal administration.

Existing measles vaccines have been successful in young children when administered as an aerosol or via the intranasal route. However, results from one study did not support these findings and implied that further investigations of the immunization protocol were required.

6 Oral Vaccines

Oral administration of infectious nonpathogenic agents is an ideal method of vaccination, a principal aim being to induce specific mucosal IgA immune responses. An outstanding example of a successful oral vaccine is the Sabin polio vaccine, which induces both local and systemic immune responses and provides good protection against poliovirus infection. Other current oral vaccines include killed whole-cell B subunit and live–attenuated cholera vaccines, live–attenuated typhoid vaccine, and live adenovirus vaccine. Oral immunization with live–attenuated vaccine against adenovirus successfully eliminated frequent epidemics at trainee camps and was in routine use for 25 years in US army recruits. Rotashield was a highly successful live oral vaccine giving good protection against infectious diarrhea, but was withdrawn because of possible links with intussusception. The decision to withdraw vaccination resulted in the reemergence of adenoviral infections and work is now in progress to bring back this effective and safe vaccine.

Oral administration of nonliving antigens can induce oral tolerance to systemic and local autoantigens or allergens, thereby protecting against DTH. This approach could potentially be used to reduce inflammatory reactions in chronic infections, autoimmune disorders, or allergies. However, oral tolerance could hinder successful oral vaccine development, and in practice oral vaccines have proved difficult to develop; therefore, relatively few have been licensed for human use.

7 Conclusions

An important key to the development of novel mucosal vaccines is to understand how mucosal adjuvants can lead to the induction of protective responses (particularly T-cell immunity and local antibody production), while avoiding oral tolerance, induction of regulatory T cells, or immunopathogenic immunity. Genetically engineered live vaccines designed to exploit newly understood immune mechanisms of tolerance and immunoregulation offer fresh hope in the race to develop new mucosal vaccines.

Acknowledgment

This work was supported by the Wellcome Trust, UK (programme grant 054797/Z/98/Z) and European Union grant 'Impressuvac' (QLRT-PL1999-01044).

See also Autoantibodies and Autoimmunity; Dendritic Cells; Human and Veterinary Classical Vaccines against Bacterial Diseases; Immune Defence, Cell Mediated; Immunologic Memory; Immunology.

Bibliography

Books and Reviews

Del Giudice, G., Pizza, M., Rappuoli, R. (1998) Molecular basis of vaccination, *Mol. Aspects Med.* **19**, 1–70.

Eriksson, K., Holmgren, J. (2002) Recent advances in mucosal vaccines and adjuvants, *Curr. Opin. Immunol.* **14**, 666–672.

McCluskie, M.J., Davis, H.L. (1999) Novel strategies using DNA for the induction of mucosal immunity, *Crit. Rev. Immunol.* **19**, 303–329.

Mielcarek, N., Alonso, S., Locht, C. (2001) Nasal vaccination using live bacterial vectors, *Adv. Drug Deliv. Rev.* **51**, 55–69.

Walker, R.I. (1994) New strategies for using mucosal vaccination to achieve more effective immunization, *Vaccine* **12**, 387–400.

Primary Literature

Barchfeld, G.L., Hessler, A.L., Chen, M., Pizza, M., Rappuoli, R., Van Nest, G.A. (1999) The adjuvants MF59 and LT-K63 enhance the mucosal and systemic immunogenicity of subunit influenza vaccine administered intranasally in mice, *Vaccine* **17**, 695–704.

Bastien, N., Trudel, M., Simard, C. (1999) Complete protection of mice from respiratory syncytial virus infection following mucosal delivery of synthetic peptide vaccines, *Vaccine* **17**, 832–836.

Beck, M., Smerdel, S., Dedic, I., Delimar, N., Rajninger-Miholic, M., Juzbasic, M., Manhalter, T., Vlatkovic, R., Borcic, B., Mihajic, Z. (1986) Immune response to Edmonston-Zagreb measles virus strain in monovalent and combined MMR vaccine, *Dev. Biol. Stand.* **65**, 95–100.

Belshe, R.B., Gruber, W.C. (2000) Prevention of otitis media in children with live attenuated influenza vaccine given intranasally, *Pediatr. Infect. Dis. J.* **19**, S66–S71.

Belshe, R.B., Mendelman, P.M., Treanor, J., King, J., Gruber, W.C., Piedra, P., Bernstein, D.I., Hayden, F.G., Kotloff, K., Zangwill, K., Iacuzio, D., Wolff, M. (1998) The efficacy of live attenuated, cold-adapted, trivalent, intranasal influenzavirus vaccine in children, *N. Engl. J. Med.* **338**, 1405–1412.

Belshe, R.B., Gruber, W.C., Mendelman, P.M., Mehta, H.B., Mahmood, K., Reisinger, K., Treanor, J., Zangwill, K., Hayden, F.G., Bernstein, D.I., Kotloff, K., King, J., Piedra, P.A., Block, S.L., Yan, L., Wolff, M. (2000) Correlates of immune protection induced by live, attenuated, cold-adapted, trivalent, intranasal influenza virus vaccine, *J. Infect. Dis.* **181**, 1133–1137.

Belyakov, I.M., Moss, B., Strober, W., Berzofsky, J.A. (1999) Mucosal vaccination overcomes the barrier to recombinant vaccinia immunization caused by preexisting poxvirus immunity, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 4512–4517.

Bradney, C.P., Sempowski, G.D., Liao, H.X., Haynes, B.F., Staats, H.F. (2002) Cytokines as adjuvants for the induction of anti-human immunodeficiency virus peptide immunoglobulin G (IgG) and IgA antibodies in serum and mucosal secretions after nasal immunization, *J. Virol.* **76**, 517–524.

Brazolot Millan, C.L., Weeratna, R., Krieg, A.M., Siegrist, C.A., Davis, H.L. (1998) CpG DNA can induce strong Th1 humoral and cell-mediated immune responses against hepatitis B surface antigen in young mice, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 15553–15558.

Broide, D., Raz, E. (1999) DNA-Based immunization for asthma, *Int. Arch. Allergy Immunol.* **118**, 453–456.

Chang, E.J., Zangwill, K.M., Lee, H., Ward, J.I. (2002) Lack of association between rotavirus infection and intussusception: implications for use of attenuated rotavirus vaccines, *Pediatr. Infect. Dis. J* **21**, 97–102.

Chun, S., Daheshia, M., Lee, S., Eo, S.K., Rouse, B.T. (1999) Distribution fate and mechanism of immune modulation following mucosal delivery of plasmid DNA encoding IL-10, *J. Immunol.* **163**, 2393–2402.

Coulter, A., Harris, R., Davis, R., Drane, D., Cox, J., Ryan, D., Sutton, P., Rockman, S., Pearse, M. (2003) Intranasal vaccination with ISCOMATRIX((R)) adjuvanted influenza vaccine, *Vaccine* **21**, 946–949.

Cusi, M.G., Zurbriggen, R., Correale, P., Valassina, M., Terrosi, C., Pergola, L., Valensin, P.E.,
Gluck, R. (2002) Influenza virosomes are an efficient delivery system for respiratory syncytial virus-F antigen inducing humoral and cell-mediated immunity, *Vaccine* **20**, 3436–3442.

Davis, H.L., Weeratna, R., Waldschmidt, T.J., Tygrett, L., Schorr, J., Krieg, A.M., Weeratna, R. (1998) CpG DNA is a potent enhancer of specific immunity in mice immunized with recombinant hepatitis B surface antigen, *J. Immunol.* **160**, 870–876.

Fromen-Romano, C., Drevet, P., Robert, A., Menez, A., Leonetti, M. (1999) Recombinant Staphylococcus strains as live vectors for the induction of neutralizing anti-diphtheria toxin antisera, *Infect. Immun.* **67**, 5007–5011.

Gallichan, W.S., Rosenthal, K.L. (1995) Specific secretory immune responses in the female genital tract following intranasal immunization with a recombinant adenovirus expressing glycoprotein B of herpes simplex virus, *Vaccine* **13**, 1589–1595.

Giuliani, M.M., Del Giudice, G., Giannelli, V., Dougan, G., Douce, G., Rappuoli, R., Pizza, M. (1998) Mucosal adjuvanticity and immunogenicity of LTR72, a novel mutant of Escherichia coli heat-labile enterotoxin with partial knockout of ADP-ribosyltransferase activity, *J. Exp. Med.* **187**, 1123–1132.

Glueck, R. (2001) Pre-clinical and clinical investigation of the safety of a novel adjuvant for intranasal immunization, *Vaccine* **20**, S42–S44.

Hathaway, L.J., Obeid, O.E., Steward, M.W. (1998) Protection against measles virus-induced encephalitis by antibodies from mice immunized intranasally with a synthetic peptide immunogen, *Vaccine* **16**, 135–141.

Hordnes, K., Tynning, T., Brown, T.A., Haneberg, B., Jonsson, R. (1997) Nasal immunization with group B streptococci can induce high levels of specific IgA antibodies in cervicovaginal secretions of mice, *Vaccine* **15**, 1244–1251.

Hvalbye, B.K., Aaberge, I.S., Lovik, M., Haneberg, B. (1999) Intranasal immunization with heat-inactivated Streptococcus pneumoniae protects mice against systemic pneumococcal infection, *Infect. Immun.* **67**, 4320–4325.

Jabbar, I.A., Fernando, G.J., Saunders, N., Aldovini, A., Young, R., Malcolm, K., Frazer, I.H. (2000) Immune responses induced by BCG recombinant for human papillomavirus L1 and E7 proteins, *Vaccine* **18**, 2444–2453.

Kato, H., Bukawa, H., Hagiwara, E., Xin, K.Q., Hamajima, K., Kawamoto, S., Sugiyama, M., Sugiyama, M., Noda, E., Nishizaki, M., Okuda, K. (2000) Rectal and vaginal immunization with a macromolecular multicomponent peptide vaccine candidate for HIV-1 infection induces HIV-specific protective immune responses, *Vaccine* **18**, 1151–1160.

Klinman, D.M., Yi, A.K., Beaucage, S.L., Conover, J., Krieg, A.M. (1996) CpG motifs present in bacteria DNA rapidly induce lymphocytes to secrete interleukin 6, interleukin 12, and interferon gamma, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 2879–2883.

Kovarik, J., Bozzotti, P., Love-Homan, L., Pihlgren, M., Davis, H.L., Lambert, P.H., Krieg, A.M., Siegrist, C.A. (1999) CpG oligodeoxynucleotides can circumvent the Th2 polarization of neonatal responses to vaccines but may fail to fully redirect Th2 responses established by neonatal priming, *J. Immunol.* **162**, 1611–1617.

Kremer, L., Dupre, L., Riveau, G., Capron, A., Locht, C. (1998) Systemic and mucosal immune responses after intranasal administration of recombinant Mycobacterium bovis bacillus Calmette-Guerin expressing glutathione S-transferase from Schistosoma haematobium, *Infect. Immun.* **66**, 5669–5676.

Kweon, M.N., Yamamoto, M., Watanabe, F., Tamura, S., Van Ginkel, F.W., Miyauchi, A., Takagi, H., Takeda, Y., Hamabata, T., Fujihashi, K., McGhee, J.R., Kiyono, H. (2002) A nontoxic chimeric enterotoxin adjuvant induces protective immunity in both mucosal and systemic compartments with reduced IgE antibodies, *J. Infect. Dis.* **186**, 1261–1269.

Lee, S.W., Sung, Y.C. (1998) Immuno-stimulatory effects of bacterial-derived plasmids depend on the nature of the antigen in intramuscular DNA inoculations, *Immunology* **94**, 285–289.

Leung, N.J., Aldovini, A., Young, R., Jarvis, M.A., Smith, J.M., Meyer, D., Anderson, D.E., Carlos, M.P., Gardner, M.B., Torres, J.V. (2000) The kinetics of specific immune responses in rhesus monkeys inoculated with

live recombinant BCG expressing SIV Gag, Pol, Env, and Nef proteins, *Virology* **268**, 94–103.

Lillard, J.W. Jr, Boyaka, P.N., Taub, D.D., McGhee, J.R. (2001) RANTES potentiates antigen-specific mucosal immune responses, *J. Immunol.* **166**, 162–169.

Lu, H., Xing, Z., Brunham, R.C. (2002) GM-CSF transgene-based adjuvant allows the establishment of protective mucosal immunity following vaccination with inactivated Chlamydia trachomatis, *J. Immunol.* **169**, 6324–6331.

Marinaro, M., Boyaka, P.N., Finkelman, F.D., Kiyono, H., Jackson, R.J., Jirillo, E., McGhee, J.R. (1997) Oral but not parenteral interleukin (IL)-12 redirects T helper 2 (Th2)-type responses to an oral vaccine without altering mucosal IgA responses, *J. Exp. Med.* **185**, 415–427.

Mason, H.S., Lam, D.M., Arntzen, C.J. (1992) Expression of hepatitis B surface antigen in transgenic plants, *Proc. Natl. Acad. Sci. U. S. A.* **89**, 11745–11749.

Mason, H.S., Ball, J.M., Shi, J.J., Jiang, X., Estes, M.K., Arntzen, C.J. (1996) Expression of Norwalk virus capsid protein in transgenic tobacco and potato and its oral immunogenicity in mice, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 5335–5340.

Matsumoto, S., Yukitake, H., Kanbara, H., Yamada, T. (1999) Long-lasting protective immunity against rodent malaria parasite infection at the blood stage by recombinant BCG secreting merozoite surface protein-1, *Vaccine* **18**, 832–834.

McCluskie, M.J., Davis, H.L. (1998) CpG DNA is a potent enhancer of systemic and mucosal immune responses against hepatitis B surface antigen with intranasal administration to mice, *J. Immunol.* **161**, 4463–4466.

McDermott, M.R., Bienenstock, J. (1979) Evidence for a common mucosal immunologic system. I. Migration of B immunoblasts into intestinal, respiratory, and genital tissues, *J. Immunol.* **122**, 1892–1898.

Medaglini, D., Oggioni, M.R., Pozzi, G. (1998) Vaginal immunization with recombinant gram-positive bacteria, *Am. J. Reprod. Immunol.* **39**, 199–208.

Medaglini, D., Pozzi, G., King, T.P., Fischetti, V.A. (1995) Mucosal and systemic immune responses to a recombinant protein expressed on the surface of the oral commensal bacterium Streptococcus gordonii after oral colonization, *Proc. Natl. Acad. Sci. U. S. A.* **92**, 6868–6872.

Mielcarek, N., Riveau, G., Remoue, F., Antoine, R., Capron, A., Locht, C. (1998) Homologous and heterologous protection after single intranasal administration of live attenuated recombinant Bordetella pertussis, *Nat. Biotechnol.* **16**, 454–457.

Minutello, M., Senatore, F., Cecchinelli, G., Bianchi, M., Andreani, T., Podda, A., Crovari, P. (1999) Safety and immunogenicity of an inactivated subunit influenza virus vaccine combined with MF59 adjuvant emulsion in elderly subjects, immunized for three consecutive influenza seasons, *Vaccine* **17**, 99–104.

Morein, B., Sundquist, B., Hoglund, S., Dalsgaard, K., Osterhaus, A. (1984) Iscom, a novel structure for antigenic presentation of membrane proteins from enveloped viruses, *Nature (London)* **308**, 457–460.

Muller, C.P., Beauverger, P., Schneider, F., Jung, G., Brons, N.H. (1995) Cholera toxin B stimulates systemic neutralizing antibodies after intranasal co-immunization with measles virus, *J. Gen. Virol.* **76**, 1371–1380.

Niewiesk, S. (2001) Studying experimental measles virus vaccines in the presence of maternal antibodies in the cotton rat model (Sigmodon hispidus), *Vaccine* **19**, 2250–2253.

Olszewska, W., Partidos, C.D., Steward, M.W. (2000) Antipeptide antibody responses following intranasal immunization: effectiveness of mucosal adjuvants, *Infect. Immun.* **68**, 4923–4929.

Olszewska, W., Erume, J., Ripley, J., Steward, M.W., Partidos, C.D. (2001) Immune responses and protection induced by mucosal and systemic immunisation with recombinant measles nucleoprotein in a mouse model of measles virus-induced encephalitis, *Arch. Virol.* **146**, 293–302.

Partidos, C.D., Vohra, P., Steward, M.W. (1996) Induction of measles virus-specific cytotoxic T-cell responses after intranasal immunization with synthetic peptides, *Immunology* **87**, 179–185.

Pizza, M., Domenighini, M., Hol, W., Giannelli, V., Fontana, M.R., Giuliani, M.M., Magagnoli, C., Peppoloni, S., Manetti, R., Rappuoli, R. (1994) Probing the structure-activity

relationship of Escherichia coli LT-A by site-directed mutagenesis, *Mol. Microbiol.* **14**, 51–60.

Pozzi, G., Oggioni, M.R., Manganelli, R., Medaglini, D., Fischetti, V.A., Fenoglio, D., Valle, M.T., Kunkl, A., Manca, F. (1994) Human T-helper cell recognition of an immunodominant epitope of HIV-1 gp120 expressed on the surface of Streptococcus gordonii, *Vaccine* **12**, 1071–1077.

Robinson, K., Chamberlain, L.M., Schofield, K.M., Wells, J.M., Le Page, R.W. (1997) Oral vaccination of mice against tetanus with recombinant Lactococcus lactis, *Nat. Biotechnol.* **15**, 653–657.

Sabin, A.B., Flores, A.A., Fernandez, d.C., Albrecht, P., Sever, J.L., Shekarchi, I. (1984) Successful immunization of infants with and without maternal antibody by aerosolized measles vaccine. II. Vaccine comparisons and evidence for multiple antibody response, *JAMA* **251**, 2363–2371.

Sabin, A.B., Flores, A.A., Fernandez, d.C., Sever, J.L., Madden, D.L., Shekarchi, I., Albrecht, P. (1983) Successful immunization of children with and without maternal antibody by aerosolized measles vaccine. I. Different results with undiluted human diploid cell and chick embryo fibroblast vaccines, *JAMA* **249**, 2651–2662.

Sasaki, S., Sumino, K., Hamajima, K., Fukushima, J., Ishii, N., Kawamoto, S., Mohri, H., Kensil, C.R., Okuda, K. (1998) Induction of systemic and mucosal immune responses to human immunodeficiency virus type 1 by a DNA vaccine formulated with QS-21 saponin adjuvant via intramuscular and intranasal routes, *J. Virol.* **72**, 4931–4939.

Schmidt, A.C., McAuliffe, J.M., Murphy, B.R., Collins, P.L. (2001) Recombinant bovine/human parainfluenza virus type 3 (B/HPIV3) expressing the respiratory syncytial virus (RSV) G and F proteins can be used to achieve simultaneous mucosal immunization against RSV and HPIV3, *J. Virol.* **75**, 4594–4603.

Schmidt, A.C., Wenzke, D.R., McAuliffe, J.M., St Claire, M., Elkins, W.R., Murphy, B.R., Collins, P.L. (2002) Mucosal Immunization of Rhesus Monkeys against Respiratory Syncytial Virus Subgroups A and B and Human Parainfluenza Virus Type 3 by Using a Live cDNA-Derived Vaccine Based on a Host Range-Attenuated Bovine Parainfluenza Virus Type 3 Vector Backbone. *J. Virol.* **76**, 1089–1099.

Shroff, K.E., Marcucci-Borges, L.A., de Bruin, S.J., Winter, L.A., Tiberio, L., Pachuk, C., Snyder, L.A., Satishchandran, C., Ciccarelli, R.B., Higgins, T.J. (1999) Induction of HSV-gD2 specific CD4(+) cells in Peyer's patches and mucosal antibody responses in mice following DNA immunization by both parenteral and mucosal administration, *Vaccine* **18**, 222–230.

Simasathien, S., Migasena, S., Bellini, W., Samakoses, R., Pitisuttitham, P., Bupodom, W., Heath, J., Anderson, L., Bennett, J. (1997) Measles vaccination of Thai infants by intranasal and subcutaneous routes: possible interference from respiratory infections, *Vaccine* **15**, 329–334.

Simmons, C.P., Hussell, T., Sparer, T., Walzl, G., Openshaw, P., Dougan, G. (2001) Mucosal delivery of a respiratory syncytial virus CTL peptide with enterotoxin-based adjuvants elicits protective, immunopathogenic, and immunoregulatory antiviral CD8+ T cell responses, *J. Immunol.* **166**, 1106–1113.

Simonsen, L., Morens, D., Elixhauser, A., Gerber, M., Van Raden, M., Blackwelder, W. (2001) Effect of rotavirus vaccination programme on trends in admission of infants to hospital for intussusception, *Lancet* **358**, 1224–1229.

Staats, H.F., Ennis, F.A. Jr (1999) IL-1 is an effective adjuvant for mucosal and systemic immune responses when coadministered with protein immunogens, *J. Immunol.* **162**, 6141–6147.

Staats, H.F., Bradney, C.P., Gwinn, W.M., Jackson, S.S., Sempowski, G.D., Liao, H.X., Letvin, N.L., Haynes, B.F. (2001) Cytokine requirements for induction of systemic and mucosal CTL after nasal immunization, *J. Immunol.* **167**, 5386–5394.

Steidler, L., Robinson, K., Chamberlain, L., Schofield, K.M., Remaut, E., Le Page, R.W., Wells, J.M. (1998) Mucosal delivery of murine interleukin-2 (IL-2) and IL-6 by recombinant strains of Lactococcus lactis coexpressing antigen and cytokine, *Infect. Immun.* **66**, 3183–3189.

Tacket, C.O., Mason, H.S., Losonsky, G., Clements, J.D., Levine, M.M., Arntzen, C.J. (1998) Immunogenicity in humans of a recombinant bacterial antigen delivered in a transgenic potato, *Nat. Med.* **4**, 607–609.

Tebbey, P.W., Scheuer, C.A., Peek, J.A., Zhu, D., LaPierre, N.A., Green, B.A., Phillips, E.D., Ibraghimov, A.R., Eldridge, J.H., Hancock, G.E. (2000) Effective mucosal immunization against respiratory syncytial virus using purified F protein and a genetically detoxified cholera holotoxin, CT-E29H, *Vaccine* **18**, 2723–2734.

Ulmer, J.B., Donnelly, J.J., Parker, S.E., Rhodes, G.H., Felgner, P.L., Dwarki, V.J., Gromkowski, S.H., Deck, R.R., DeWitt, C.M., Friedman, A. (1993) Heterologous protection against influenza by injection of DNA encoding a viral protein, *Science* **259**, 1745–1749.

Walsh, E.E. (1993) Mucosal immunization with a subunit respiratory syncytial virus vaccine in mice, *Vaccine* **11**, 1135–1138.

Walsh, E.E. (1994) Humoral, mucosal, and cellular immune response to topical immunization with a subunit respiratory syncytial virus vaccine, *J. Infect. Dis.* **170**, 345–350.

Welter, J., Taylor, J., Tartaglia, J., Paoletti, E., Stephensen, C.B. (1999) Mucosal vaccination with recombinant poxvirus vaccines protects ferrets against symptomatic CDV infection, *Vaccine* **17**, 308–318.

Wierzbicki, A., Kiszka, I., Kaneko, H., Kmieciak, D., Wasik, T.J., Gzyl, J., Kaneko, Y., Kozbor, D. (2002) Immunization strategies to augment oral vaccination with DNA and viral vectors expressing HIV envelope glycoprotein, *Vaccine* **20**, 1295–1307.

Wigdorovitz, A., Carrillo, C., Dus Santos, M.J., Trono, K., Peralta, A., Gomez, M.C., Rios, R.D., Franzone, P.M., Sadir, A.M., Escribano, J.M., Borca, M.V. (1999) Induction of a protective antibody response to foot and mouth disease virus in mice following oral or parenteral immunization with alfalfa transgenic plants expressing the viral structural protein VP1, *Virology* **255**, 347–353.

Yusibov, V., Hooper, D.C., Spitsin, S.V., Fleysh, N., Kean, R.B., Mikheeva, T., Deka, D., Karasev, A., Cox, S., Randall, J., Koprowski, H. (2002) Expression in plants and immunogenicity of plant virus-based experimental rabies vaccine, *Vaccine* **20**, 3155–3164.

Zhu, Y.D., Fennelly, G., Miller, C., Tarara, R., Saxe, I., Bloom, B., McChesney, M. (1997) Recombinant bacille Calmette-Guerin expressing the measles virus nucleoprotein protects infant rhesus macaques from measles virus pneumonia, *J. Infect. Dis.* **176**, 1445–1453.

Mucoviscidosis (Cystic Fibrosis), Molecular Cell Biology of

Gerd Döring[1] *and Felix Ratjen*[2]
[1] *Universitätsklinikum Tübingen, Germany*
[2] *University of Essen, Germany*

1 Epidemiology and Diagnosis 625
1.1 Incidence and Prevalence 625
1.2 Diagnosis and Screening 625
1.3 Prognosis 626

2 Genetics and Functions of CFTR 626
2.1 CFTR, Mutations, Genotype-phenotype Relations and Modifier Genes 626
2.2 CFTR Function and Structure 627
2.3 CFTR Trafficking and Tissue Location 628

3 Basic Defect-pathology Relations 629
3.1 Respiratory Tract 629
3.2 Other Organs 630

4 Immunology of Chronic Respiratory Infections 632
4.1 Type III Hypersensitivity Reaction 632
4.2 Cytokines, Growth Hormones, and Nitric Oxide (NO) 633

5 Causative and Symptomatic Therapy Strategies 634
5.1 Gene Replacement Therapy 634
5.2 Pharmacological Treatment of CFTR 634
5.3 Prevention of Infection 635
5.4 Treatment of Infection 636
5.5 Anti-inflammatory and Mucolytic Drugs 636
5.6 Transplantation 637
5.7 Treatment of Pancreatic Insufficiency, Nutrition, and Liver Disease 637
5.8 Animal Models 637

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 8
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30550-5

Bibliography 638
Books and Reviews 638
Primary Literature 638

Keywords

Airway Surface Liquid (ASL)
Fluid lining the apical surface of epithelial cells. ASL is depleted in the respiratory tract of CF patients. This increases mucus viscosity and impairs cilia function, resulting in defective mucociliary clearance of bacterial pathogens from CF airways.

Chaperones
Chaperones enable misfolded proteins to be maintained in a conformation that is competent for folding, its subsequent release from the quality control apparatus, and delivery to the cell surface. Certain chemical chaperones, such as phenylbutyrate, CPX, genisteine, and glycerol can increase F508del folding yield in cell culture systems.

Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)
Chloride conducting channel in apical cell membranes. Mutations in CFTR cause the disease cystic fibrosis.

DF508 Mutation
Deletion of phenylalanine in position 508 (F508del) of the CFTR protein; the most common mutation causing CF worldwide. The mutation results in misfolding of the CFTR protein leading to premature intracellular degradation.

Pseudomonas aeruginosa
Most common pathogen in CF lung disease. *Pseudomonas aeruginosa* infection is associated with clinical deterioration and massive endobronchial inflammation in most patients with cystic fibrosis.

Cystic fibrosis (CF) is the most common autosomal recessive disorder in Caucasians, affecting ~1 : 2500 children, with a carrier frequency of 1 : 25. The causative gene, named *CF transmembrane conductance regulator* (*CFTR*), encodes a chloride channel in epithelial cells. Abnormal transport of chloride and sodium ions affects water movement across epithelia, leading to pathophysiological consequences in various organs including the respiratory, gastrointestinal and reproductive tract, the pancreas, and liver. The CF phenotype is rather heterogeneous due to many different mutations in CFTR and the influence of modifier genes. Chronic bacterial lung infections stimulate inflammatory defense mechanisms, leading to extensive tissue remodeling. The resulting emphysema and fibrosis mainly determine the reduced life expectancy in individuals with CF. Owing to improved symptomatic

treatment strategies, including better nutrition and antibiotic therapy, the prognosis of CF individuals has considerably improved and many children now reach adult life. Research is focused on the development of pharmacological drugs correcting ion channel dysfunction, anti-inflammatory drugs and vaccines to prevent airway infections. The causative gene replacement therapy has not yet been successfully applied in CF patients.

1 Epidemiology and Diagnosis

1.1 Incidence and Prevalence

CF (cystic fibrosis) is the most common fatal inherited disease in the Caucasian population, affecting about 1 : 2500 children, with a carrier frequency of 1 : 25. Case-finding studies resulted in incidence rates between 1 : 7700 in Sweden and 1 : 2500 in the United Kingdom. In the United States, an incidence rate of ~1 : 3500 live white births has been estimated and, based on registry data, the prevalence rate in the year 2000 was 8.9 CF patients in 100 000 individuals. Similar prevalence rates have been calculated in Germany (6.9/100 000), Italy (6.5/100 000), and United Kingdom (15/100 000). In other racial groups, CF is much less frequent. Estimates from non-Caucasian populations from the United States give incidence rates of 1 : 14 000 for black Americans and 1 : 11 500 for Hispanic births.

1.2 Diagnosis and Screening

CF is diagnosed on clinical symptoms including persistent cough and diarrhea caused by pancreatic insufficiency. The single most useful diagnostic procedure is the sweat test with chloride concentrations $>60\ \mathrm{mmol\ L^{-1}}$ in typical cases of CF. Generally, the diagnosis is confirmed by genotyping of the most common CFTR mutations, which vary between different geographic regions. Since CFTR mutations have been identified in persons with clinical conditions such as pancreatitis, infertility due to congenital bilateral absence of the vas deferens (CBAVD), allergic bronchopulmonary aspergillosis, disseminated bronchiectasis, and diffuse panbronchiolitis, it is generally thought that genotype alone is an insufficient basis for the diagnosis of CF.

Fifty percent of all CF patients in the United States are diagnosed by the age of 6 months and 90% by the age of 8 years. Neonatal screening has been recommended because of data revealing that screened patients have better growth and weight gain than controls and that early diagnosis decreases the rate of hospitalization and may lead to an improved lung function in the first 10 years of life. Neonatal screening programs are generally based on a two-step approach with immune reactive trypsinogen in dried blood spots and confirmation by mutation analysis, including multiple CFTR alleles in positive cases. Indeed, neonatal screening has now been established in several countries. It is important to note that clinical benefits from neonatal screening are dependent on adequate treatment, and the failure to provide an adequate standard of care after a diagnosis of CF has been made is probably

a critical factor in outcome. Furthermore, early genetic counseling gives parents the choice of avoiding the birth of another child with CF.

Population screening for CF can be carried out to identify CF carriers or patients, whereas heterozygote screening allows carrier couples to improve family planning. Whether CF carrier screening reduces the number of CF births on a population level is still questionable. Recent guidelines recommend prenatal screening for CF in the United States.

1.3 Prognosis

Prognosis of CF has improved dramatically in some but not all countries as a result of better care and therapy and most children now reach adult life. In the United States, the median age has risen from 8.4 years in 1969 to 14.3 years in 1998 and the median survival rose from 14 years in 1969 to 32.3 years in 1998. Similar improvements in survival have occurred in Europe but significant differences persist because of differences in treatment strategies, access to specialized CF centers, and socioeconomic status. It is generally agreed that patients treated in a CF center with specialized medical care have a better survival. Since chronic lung disease contributes mostly to morbidity and mortality in CF, optimal antibiotic therapy may influence prognosis considerably. Patients with class IV or V mutations associated with residual CFTR function have better clinical outcomes compared to those with class II mutations (including F508del, see below). However, for most patients, prognosis cannot be determined on the basis of genotype analysis.

2 Genetics and Functions of CFTR

2.1 CFTR, Mutations, Genotype-phenotype Relations and Modifier Genes

CF is caused by mutations in a 230-kb gene on chromosome 7, encoding a 1480 amino acid polypeptide named *cystic fibrosis transmembrane conductance regulator* (*CFTR*). Over 1200 mutations and sequence variants have been described to date and reported to the Cystic Fibrosis Genetic Analysis Consortium. Most of these mutations are rare and only 4 mutations occur in a frequency of more than 1%. CFTR mutations are grouped into five classes: defective synthesis (I), defective processing (II), defective regulation (III), defective conductance (IV), and partially defective production or processing (V). Class I–III mutations are more common and are associated with pancreatic insufficiency. Patients with the rarer class IV–V mutations often are pancreatic sufficient. The most common mutation worldwide is a class II mutation caused by a deletion of phenylalanine in position 508 (F508del) of the CFTR protein, leading to misfolding. Of the 43 849 CF chromosomes tested, 66% are F508del. Linking mutations to the severity of lung disease has been unsuccessful, and patients who are homozygous for the F508del mutation exhibit a wide spectrum in the rate of development and severity of lung disease.

Because F508del homozygous patients differ largely in phenotype and in chloride conductance, it is believed that environmental factors and/or genes other than CFTR modify the development, progression, and disease severity of CF. At present, a number of candidate genes including genes that regulate aspects of innate

lung defense and inflammatory cascades have been suggested, which await confirmation in large CF population studies. Modifier genes are now being identified using single nucleotide polymorphisms in genome-wide searches and complemented by proteomic approaches. Modifying genes are likely to affect many different aspects of the CF phenotype, which therefore remains difficult to predict from CFTR mutation data alone.

2.2 CFTR Function and Structure

CFTR functions as a chloride channel in apical membranes. The primary structure of CFTR indicated that it belongs to a family of transmembrane proteins called *ATP-binding cassette* (*ABC*) *transporters*. ABC transporters (or traffic ATPases) form a large family of proteins responsible for the translocation of a variety of compounds across membranes of both prokaryotes and eukaryotes. CFTR is composed of five domains: two membrane-spanning domains (MSDs), two nucleotide binding domains (NBDs), and a regulatory (R) domain (Fig. 1). The F508del mutation occurs in the DNA sequence that codes for the NBD1. In wild-type CFTR, an extracellular glycosylation site is present and the NBDs and R domain are located on the intracellular side of the membrane.

CFTR channel activity is controlled by the balance of kinase and phosphatase activity within the cell and by cellular ATP levels. Activation of a cAMP-dependent protein kinase (PK) (such as PKA and PKC) causes the phosphorylation of multiple serine residues within the R domain. Once the R domain is phosphorylated, channel gating is regulated by a cycle of ATP binding and hydrolysis at the NBDs, which form a head-to-tail dimer. The two NBDs are nonequivalent in their interactions with nucleotides. Whereas NBD1 acts as a site of stable nucleotide interaction, NBD2 constitutes a site of fast turnover. Finally, protein phosphatases dephosphorylate the regulatory domain and return the channel to its quiescent state. It has been proposed that energy from ATP hydrolysis by the conserved NBDs causes conformational changes in the MSDs, which are coupled with the opening and closing of the pore. The

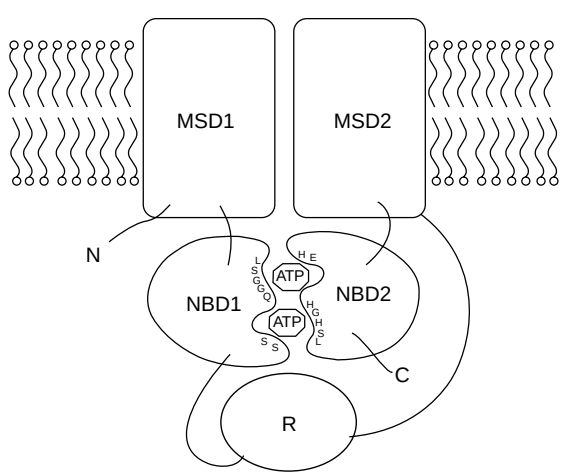

Fig. 1 Schematic model of CFTR. MSD: membrane-spanning domain, NBD: nucleotide binding domain; R: regulatory domain. Also indicated is a putative nucleotide binding domain. (From Lewis, H.A. et al. (2004) Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator, *EMBO J.* **23**, 282–93).

sixth transmembrane segment seems to play a dominant role in determining the pore properties of CFTR. The CFTR pore has a deep wide intracellular vestibule but a shallow wide extracellular vestibule. The resolution of the crystal structure of mouse NBD1 confirms some of these notions (Fig. 2). This NBD1 differs from typical ABC domains in having added regulatory segments, a foreshortened subdomain interconnection, and unusual nucleotide conformation and undetectable ATPase activity.

The complex secondary structure of the protein suggests that CFTR possesses other functions in addition to being a chloride channel. Amongst others, CFTR has been described as a regulator of other apical membrane conductance pathways through interactions with the amiloride-sensitive epithelial sodium channel and the outwardly rectifying chloride channel. Whether the activity of the epithelial sodium channel (ENaC) is inversely related to the activity of CFTR or whether its activity increases with CFTR activity is not clear at present. ENaC is also regulated by the serum and glucocorticoid-dependent kinase (SGK1), which is increased in CF lung tissue because of inflammatory processes. F508del CFTR has defects in both channel gating and endoplasmic reticulum-to-plasma membrane processing. Recent data suggest that CFTR also transports HCO_3^- as well as glutathione or regulates HCO_3^- transport through epithelial cells.

2.3 CFTR Trafficking and Tissue Location

F508del is retained in the endoplasmic reticulum (ER) because of improper folding and, therefore, will not reach the Golgi

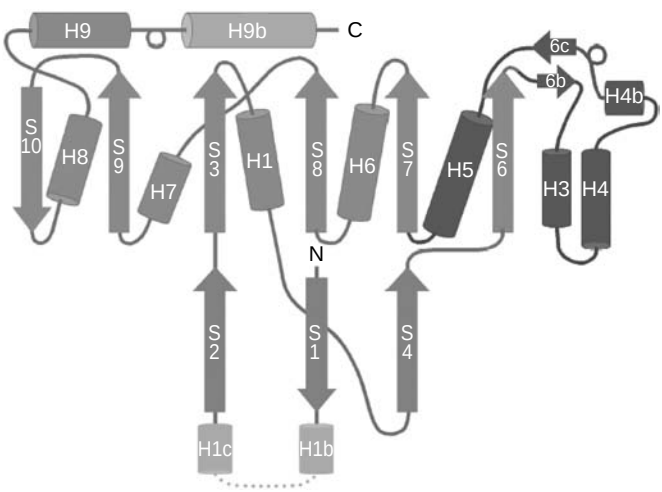

Fig. 2 Topology diagram of mNBD1. The F1-type ATP-binding core subdomain is shown in gold, the ABC α-subdomain in cyan, and the ABC ß-subdomain in green. Regions of mNBD1 that are different from previous ABC structures are shown in gray. Circles indicate the positions of 3_{10} helices. (From Lewis, H.A. et al. (2004) Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator. *EMBO J.*, **23**, 282–293) (see color plate p. xxiv).

complex for the acquisition of complex N-linked oligosaccharide chains, important in the quality control pathway to prevent the development of improperly folded glycoproteins. Thus, the F508del mutation is primarily a protein-folding defect with altered protein trafficking or kinesis secondary to altered folding. Although some F508del can be delivered to the plasma membrane, channel gating is profoundly defective, because PKA-dependent phosphorylation of the R domain activates F508del CFTR to a much lesser degree than wild-type CFTR. Like other membrane glycoproteins, CFTR interacts with chaperone proteins, which prevent premature and inappropriate folding interactions during biosynthesis and increase the overall yield of correctly folded product. In this context, hsp70, hsc70, and calnexin have been identified. CFTR has also been shown to form a macromolecular complex in apical membranes with the ß2 adrenergic receptor and the ezrin-/radixin-/moesin-binding phosphoprotein 50. The assembly of the complex is regulated by PKA-dependent phosphorylation and deletion of the regulatory domain of CFTR abolishes PKA regulation of complex assembly.

In general, CFTR is found in tissues that are clinically affected by CF although low levels also occur elsewhere. The most common site is in the apical membrane of epithelial cells that line exocrine ducts or airways, and this is consistent with the proposed chloride channel function. By immunohistochemistry, wild type but not F508del CFTR was detected at the luminal membrane of crypt colonocytes, sweat glands, and respiratory epithelial cells. Both B and T lymphocytes express CFTR, reveal abnormal chloride transport, although this seems to have little functional importance. No important functional abnormalities have been shown in the heart or the placenta, both sites of CFTR expression, and the electrolytes of ocular humor, breast milk, and seminal fluid are not significantly altered in CF.

3 Basic Defect-pathology Relations

CF leads to pathologic changes in organs that express CFTR; therefore, secretory cells, sinuses, lungs, pancreas, liver, and reproductive tract are involved. The most dramatic changes are observed in CF airways where the basic defect causes mucus retention, chronic bacterial infection, and inflammation.

3.1 Respiratory Tract

Lung infections with *Pseudomonas aeruginosa* constitute a predominant disease phenotype in CF patients. Infections with *Staphylococcus aureus* and *Haemophilus influenzae* are also frequent. Several hypotheses have been offered to explain the failure of mucosal defense and the high prevalence of *P. aeruginosa* in the CF lung. It has been proposed that *P. aeruginosa* binds to CF airway epithelial cell membranes in higher density than to respective cells from normal individuals owing to an increased *P. aeruginosa* asialo-GM1 receptor density. The higher bacterial number would then lead to infection in CF airways. Alternatively, wild-type CFTR (but not mutated CFTR) has also been shown to be a receptor for *P. aeruginosa*, which mediates bacterial cell internalization and *P. aeruginosa* killing. In CF airways, therefore, *P. aeruginosa* would not be eradicated intracellularly

and could multiply and cause infection. Other studies, however, reveal that both *P. aeruginosa* and *S. aureus* are located in the mucus layer on respiratory epithelial cells rather than directly on cell membranes and that no difference in location and number of adhering bacteria is visible regardless of whether normal or CF primary respiratory cells are used or infected CF lung tissue is investigated for *P. aeruginosa* or *S. aureus* adhesion. Additionally, on the basis of the assumption of an increased sodium chloride concentration due to a defective CFTR channel on the luminal side of the respiratory epithelium, it has furthermore been suggested that salt-sensitive cationic antimicrobial peptides (defensins) are inactivated in the airway surface liquid (ASL) of CF patients, which would lead to bacterial multiplication and subsequent infection. However, not all defensins are salt sensitive and it has been difficult to prove that the ASL in CF is indeed hypertonic.

In contrast, most *in vivo* data reveal that the ASL from normal and CF individuals is isotonic. This notion has been explained by a mechanism whereby the airway epithelium reduces its volume in order to reach isotonicity (Fig. 3).

Owing to defective chloride secretion of the CF respiratory epithelial cells, sodium and therefore also water are excessively reabsorbed from the periciliary liquid, which reduced the volume of the ASL from ~20 to 6 μm. Consequently, mucus viscosity is increased and cilia function is impaired, which leads to a defect in mucociliary clearance of bacterial pathogens from CF airways. Bacteria invading the CF lung are trapped in this viscous mucus layer on top of respiratory epithelial cells, where they encounter microaerophilic or anaerobic growth conditions due to abnormal oxygen consumption of the CF cell. The physical conditions in the viscous matrix trigger a switch of *S. aureus* and *P. aeruginosa* from nonmucoid to mucoid cell types, the predominant bacterial phenotype in CF lungs. Apparently, the size of the mucoid bacterial microcolonies impairs phagocytosis and eradication of the pathogens by phagocytis cells and thus is thought to be one important factor in the pathogenicity of lung disease. The anaerobic viscous environment on the respiratory epithelial mucus layer, which is also present in large mucus (sputum) plugs later in the course of the infection, is the main reason for the chronicity of the bacterial lung infection and inflammation in CF lungs (see below).

3.2 Other Organs

Water retention as a consequence of defective chloride secretion and excessive sodium absorption is also the explanation for the pathophysiologic events in the pancreas of CF patients, leading to exocrine pancreatic insufficiency in approximately 90% of CF patients. Without sufficient fluid and HCO_3^-, digestive proenzymes are retained in pancreatic ducts and prematurely activated, which ultimately leads to tissue destruction and fibrosis. As a consequence of increasing Langerhans cell destruction, insulin secretion is decreased and delayed in older CF patients, leading to increasing prevalence of both type I and II diabetes. The consequences of CFTR mutations also lead to abnormal liver function in at least one-third of patients and biliary cirrhosis in less than 10% of the patients. Ninety-eight percent of males with CF are infertile, with aspermia secondary to atretic or absent vasa deferentia and dilated or absent seminal vesicles. Sexual potency

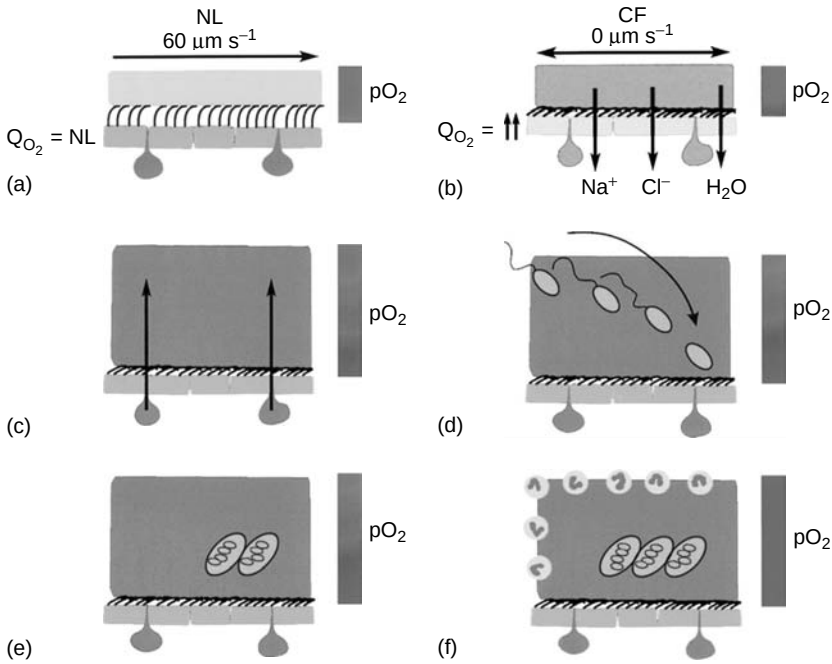

Fig. 3 Schematic model of the pathogenic events hypothesized to lead to chronic *P. aeruginosa* infection in airways of CF patients. (a) On normal airway epithelia, a thin mucus layer (light green) resides atop the PCL (clear). The presence of the low-viscosity PCL facilitates efficient mucociliary clearance (denoted by vector). A normal rate of epithelial O_2 consumption (Q_{O2}; left) produces no O_2 gradients within this thin ASL (denoted by red bar). (b–f) CF airway epithelia. (b) Excessive CF volume depletion (denoted by vertical arrows) removes the PCL, mucus becomes adherent to epithelial surfaces, and mucus transport slows/stops (bidirectional vector). The raised O_2 consumption (left) associated with accelerated CF ion transport does not generate gradients in thin films of ASL. (c) Persistent mucus hypersecretion (denoted as mucus secretory gland/goblet cell units; dark green) with time increases the height of luminal mucus masses/plugs. The raised CF epithelial Q_{O2} generates steep hypoxic gradients (blue color in bar) in thickened mucus masses. (d) *P. aeruginosa* bacteria deposited on mucus surfaces penetrate actively and/or passively (due to mucus turbulence) into hypoxic zones within the mucus masses. (e) *P. aeruginosa* adapts to hypoxic niches within mucus masses with increased alginate formation and the creation of macrocolonies. (f) Macrocolonies resist secondary defenses, including neutrophils, setting the stage for chronic infection. The presence of increased macrocolony density and, to a lesser extent neutrophils, render the now mucopurulent mass hypoxic (blue bar) (From Worlitzsch, D., Tarran, R., Ulrich, M. et al(2002) Effects of reduced mucus oxygen concentration in airway Pseudomonas infections of cystic fibrosis patients, *J. Clin. Invest.* **109**, 317–325) (see color plate p. xxiv).

as well as spermatogenesis are normal, and CF males have become fathers with techniques such as microscopic epididymal sperm aspiration and intracytoplasmatic sperm injection. Female reproductive function is normal, although cervical mucus has been reported to be dehydrated, which may impair fertility.

4 Immunology of Chronic Respiratory Infections

The inability of the immune defense system to keep the airways free of bacteria led to speculations that the then unknown genetic abnormality causes an immunological defect. With the knowledge that the CF gene codes for an epithelial ion-transport protein and data showing essentially normal systemic immune function in CF, the immunology of CF was then regarded no differently from the immunology of any other chronic infection. However, currently, CF lung disease is thought to result from a defect in innate immune response that leads to chronic lung infection (see above) and chronic inflammation. The question of whether the consequences of the basic defect provoke abnormal respiratory inflammation without prior bacterial stimuli has not been resolved as yet.

4.1 Type III Hypersensitivity Reaction

CF patients respond with the production of antibodies to a large number of *P. aeruginosa* antigens and antigens of other microbial pathogens. Antibody production provokes hypergammaglobulinaemia in infected CF patients, and leads to immune complexes, which are detectable in patients' sputa, bronchial secretions, or serum samples. Immune complexes stimulate the major phagocytic cell entering the endobronchial airways, the polymorphonuclear leukocytes (neutrophils). High immune complex levels (and antibody titers) correlate with poor clinical status of the patients, suggesting that neutrophil stimulation is responsible for tissue damage or remodeling in CF airways. Nevertheless, some of the antipseudomonal antibodies are nonopsonic or even blocking phagocytosis, particularly antibodies directed against *P. aeruginosa* polysaccharide components. This phenomenon has not yet been fully explained and might be related to a subsequent loss of avidity in antigen binding during the chronic course of the infection. During this type III hypersensitivity reaction, neutrophils are attracted from the blood vessels to the site of infection by chemoattractants, includingleukotriene B_4 (LTB_4) and interleukin (IL)-8.

As mentioned above, bacterial microcolony formation triggered by the anaerobic viscous environment on the respiratory epithelial mucus layer and in the large mucus (sputum) plugs (Fig. 4) is one reason for the chronicity of the bacterial lung infection. In addition, other factors may be involved. The highly viscous mucus may affect the motility of neutrophils, resulting in retarded neutrophil transmigration of the mucus layer overlaying the epithelial cells in CF airways, which could give bacteria a temporal advantage to form their protecting microcolony polysaccharide coat. Owing to the anaerobic environment in the viscous mucus, antimicrobial activity of neutrophils related to the production of reactive oxygen species is impaired and bacteria with intrinsic or acquired resistance to antimicrobial peptides (defensins) may survive phagocytosis under these circumstances. All three conditions may be present at the same time in infected CF mucus and lead to chronicity of the infection characterized by large volumes of necrotic neutrophils and accumulation of polymerized DNA strands that form highly viscous sputum plugs (Fig. 4).

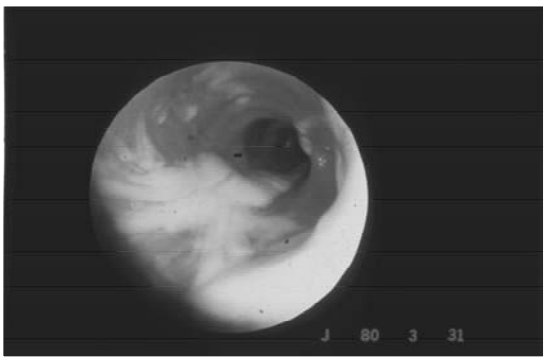

Fig. 4 Mucus plugs obstructing CF airways consisting of mostly necrotic neutrophils and bacterial microcolonies. By courtesy of Dr. Uschi Weber, Wangen im Allgäu, Germany.

The anaerobic nature of the mucus plugs obstructing airways of CF patients implies that lysosomal enzymes for neutrophils such as elastase rather than reactive oxygen species play a major role in the pathophysiology of endobronchial inflammation. Neutrophil elastase cleaves a variety of substrates including the endogenous serine proteinase inhibitors α_1-proteinase inhibitor (α_1-PI) and secretory leukocyte proteinase inhibitor (SLPI), the structural components or airways and alveoli, elastin, and proteins of the innate and specific immune systems such as immunoglobulines, immune complexes, complement components, and cell surface receptors on neutrophils or lymphocytes. Since neutrophil-mediated phagocytosis is dependent on intact complement receptors (CR1, CR3), complement components (C3b, C3bi) and intact, opsonic antibodies, opsonophagocytosis is heavily impaired in plugs in the CF lung. On the other hand, neutrophil elastase may act as a regulatory enzyme in chronic inflammation by lowering signal transduction between inflammatory molecules and effector cells. Thus, the impairment of opsonophagocytosis and T cell function by neutrophil serine proteases may have beneficial and deleterious consequences for the host: temporal downregulation of inflammation on the one hand and allowance of bacterial survival on the other hand. Similarly, stimulation of airway gland secretion by neutrophil serine proteases may be seen as beneficial and deleterious for the host. Mucus hypersecretion may be regarded as beneficial, because it removes bacterial pathogens from close contact with airway epithelial cells into the airway lumen, and deleterious because airway obstruction leads to lung function impairment.

Besides neutrophil serine proteases, macrophage-derived metalloproteinase and metalloproteinases from neutrophils including neutrophil gelatinase (type IV collagenase) may contribute to tissue remodeling in CF airways, particularly in the alveoli.

Despite the wealth of information on the neutrophil-dominated pathophysiology of CF lung disease, little is known about how endobronchial infection/inflammation impacts on fibrosis and emphysema in the alveolar space of CF patients.

4.2 Cytokines, Growth Hormones, and Nitric Oxide (NO)

The inflammatory system communicates by a large repertoire of chemical

messengers called *cytokines* produced by epithelial cells, T lymphocytes and phagocytic cells. In bronchoalveolar lavage (BAL) samples of CF patients, increased proinflammatory cytokine levels including TNFα, IL-6, and IL-8 have been measured. Persistent high cytokine levels may have deleterious consequences. For example, TNFα may cause cachexia and osteoporosis, and IL-8, a chemotactic attractant for neutrophils, may lead to a "self-perpetuating" inflammatory process. Furthermore, BAL fluids from CF patients contain significantly less IL-10 than BAL from normal control individuals, suggesting an imbalance between upregulating and downregulating cytokines during chronic inflammation. Besides cytokines, growth factors such as insulinlike growth factor 1 (IGF-1), which can be produced by many cell types including lymphocytes and alveolar macrophages may contribute to lung pathology in CF patients. One of the consequences of enhanced IGF-1 production is fibrosis.

NO originates from the biotransformation of L-arginine to L-citrulline by enzymes called *NO synthases* (*NOS*). Although one would expect that NO production is increased in CF due to chronic lung inflammation, this is not the case and NO levels in exhalations of CF patients are low. This phenomenon has not been completely resolved and attributed in part to low levels of the inducible isoform of NOS (iNOS) in the respiratory epithelium of CF patients and consumption of NO in airway secretions in CF airways during inflammation. Since NO has multiple functions (e.g. bronchodilation, phagocytic killing of bacteria, stimulation of ciliary function), low NO concentrations in airways of CF patients may be of pathophysiologic relevance.

5 Causative and Symptomatic Therapy Strategies

5.1 Gene Replacement Therapy

The ultimate curative treatment for CF is to restore CFTR function by transfecting CF cells with wild-type CFTR. Since the initiation of the first clinical trials for CF lung disease using recombinant adenovirus in the early 1990s, the field has encountered numerous obstacles including (1) difficulties with vector production, (2) inefficient delivery of vector, and (3) vector-induced inflammation. Over 30 trials on CFTR gene therapy, mainly using adenoviral vectors or liposomes have been published and have established the proof of principle for gene transfer to the lung. However, efficiency is generally low. Viral vectors have encountered the problem of host response to the vector that either results in mucosal inflammation or inefficiency of gene transfer on repeat dosing. Despite these obstacles, several research groups are currently involved in preclinical and clinical studies to enhance the efficiency and selectivity of gene delivery to the lung.

5.2 Pharmacological Treatment of CFTR

The intracellular production, trafficking or activation of CFTR are possible targets of therapeutic interventions. Class I mutations lead to decreased production of CFTR mRNA and gentamicin has been shown to partially overcome this problem. The frequency of the F508del allele among CF patients and the possibility that chloride channel activity can be restored in these

patients suggest strategies for CF treatment based on increasing the efficiency of folding and intracellular processing of this mutant. Loss of this amino acid decreases the efficiency of folding of NBF1 *in vitro* and appears to increase the probability that nascent CFTR chains become committed to degradation by a pathway that involves at least in part the ubiquitin/proteosome pathway. Certain chemical chaperones, such as phenylbutyrate, CPX, genisteine, and glycerol can increase F508del folding yield, possibly by stabilizing an early intermediate in CFTR folding, maintaining it in a conformation that is competent for folding and its subsequent release from the quality control apparatus and delivery to the cell surface. Moreover, 4-phenylbutyric acid has also been shown *in vitro* to increase production of mature CFTR and chloride transport at the cell surface, most likely by a mechanism involving upregulation at the transcriptional level and modulation of protein-folding steps. Other agents enhance the activity of F508del CFTR chloride channels at the cell surface. The best studied of these agents is the flavonoid genistein, which interacts directly with NBD2, to stabilize the open-channel configuration. Agents termed *pharmacological chaperones* (e.g. MPB compounds) potentiate the activity of mutant chloride channels present at the cell surface and rescue the cell surface expression of the same mutants. High-throughput screening technology is currently being used in the search for further compounds that either rescue the cell surface expression of mutant CFTR (termed *CFTR correctors*) or enhance the activity of mutant chloride channels present at the cell surface (termed *CFTR potentiators*). Vitamin C has been shown to regulate CFTR-mediated Cl secretion in epithelia and, therefore, represents a potential drug in CF. Furthermore, curcumin, a Ca-adenosine triphosphatase pump inhibitor, corrected defective nasal potential difference in F508del CFTR mice.

Whether treatment with chaperones will result in sufficient concentrations of CFTR on the cell surface of epithelial cells to improve CF related pathology is unclear at present. Furthermore, mutated CFTR that reaches the cell membrane may fail to function normally. Additional targets of CFTR pharmacotherapy include activators of chloride secretion such as UTP or inhibitors of sodium absorption such as amiloride. Both drugs have a very short half-life, limiting their efficacy, but newer compounds with an improved pharmacokinetic profile are being tested in CF trials. A double-blind phase II inhalation study of the UTP analog INS37217 in patients with mild CF lung disease was safe and resulted in significantly better FEV1, FEF 25 – to 75%, and FVC in the treatment group compared to patients receiving placebo. INS37217 is believed to enhance the lung's mucosal hydration and mucociliary clearance mechanisms by activating an alternative ion channel that acts in the same way as the defective ion channel in moving salt and water to the surface of the airways. Similarly, hypertonic saline is mechanistically active and early trials show beneficial results. Thus, the notion that getting water to the surface of the respiratory epithelium of CF patients will correct the pathological consequences of the basic defect becomes more and more realistic.

5.3 Prevention of Infection

Prevention of bacterial lung infection is regarded as the primary aim for

CF. Epidemiological studies suggest that transmission of *P. aeruginosa* and other pathogens occurs either by direct patient-to-patient contacts or from various environmental bacterial reservoirs. Improved hygienic measures and separation regimes have therefore been implemented in several CF centers to limit cross-infection. Prevention of *P. aeruginosa* lung infection by active vaccination may also represent a suitable alternative strategy in CF. Whereas studies using *P. aeruginosa* lipopolysaccharide preparations were unsuccessful, a small open vaccine trial using a polysaccharide-exotoxin A conjugate as antigen showed promising results. Currently, phase III multicenter vaccine trials using a *P. aeruginosa* flagella vaccine and the polysaccharide-exotoxin A conjugate vaccine are underway in Europe.

5.4 Treatment of Infection

Improved antibiotic therapy strategies against respiratory tract infections are considered as the main reason for the increased life expectancy of CF patients that has been achieved over the last decades. Antibiotics are administered by the intravenous, aerosol, or oral route, and treatment strategies include regular courses independent of the clinical status of the patient or restricted to episodes of acute exacerbations. Decisions about optimal antibiotic therapy regimes are difficult to take since eradication of the bacteria is not normally achieved. Owing to poor penetration of antibiotics into the anaerobic sputum plugs and the rapid development of mutator strains that display increased resistance against antimicrobial drugs, even with intensive antibiotic regimens, mucoid *P. aeruginosa* cannot be eradicated. A major improvement in the strategy to fight pulmonary *P. aeruginosa* infection in CF patients is early antibiotic therapy. Anti-staphylococcal treatment for at least two weeks results in an eradication rate of ~75% and only a low rate of patients harbor *S. aureus* for more than six months thereafter. In the early phase of *P. aeruginosa* colonization, antibiotic therapy may avoid the shift to chronic mucoid *P. aeruginosa* infection. The combination of inhaled colistin with oral ciprofloxacin, inhaled colistin or inhaled tobramycin alone has been used to treat early *P. aeruginosa* colonization. Long-term follow-up of patients treated with inhaled tobramycin following initial colonization with *P. aeruginosa* demonstrated that treatment not only postponed chronic infection but also led to eradication of the bacteria in treated patients. This treatment regime will most probably reduce the incidence of chronic airway infection with *P. aeruginosa* in CF patients.

5.5 Anti-inflammatory and Mucolytic Drugs

Since chronic inflammation is the cause of tissue remodeling and destruction in CF airways, the administration of anti-inflammatory drugs to other treatment arms is a logical strategy. High-dose prednisone was found to reduce the decline of lung function for the first two years in *P. aeruginosa* positive patients, but serious side effects such as growth retardation, glucose intolerance, and cataracts were found in treated patients. Therefore, the risk/benefit ratio does not favor long-term prednisone therapy in CF patients. Treatment with inhaled steroids did not result in clear benefits for pulmonary function in CF patients. High-dose ibuprofen slowed the lung function decline in CF patients, but the beneficial effect was concentrated

to a subgroup of patients aged 5 to 13 years. Increasing antiproteinase levels in the lung by supplementation with suitable inhibitors has been proposed. However, neither aerosolized recombinant human secretory leukocyte proteinase inhibitor (rSLPI) nor aerosolized α_1-PI has been tested in large numbers of CF patients. A multicenter trail with a small neutrophil elastase inhibitor, EPI-hNE4 is currently carried out in CF patients. Recombinant human DNase has been found to reduce sputum viscosity, improve pulmonary function, and reduce the number of pulmonary exacerbation in CF patients with moderate and mild lung disease. Lipoxins are anti-inflammatory lipid mediators that modulate neutrophilic inflammation. Furthermore, administration of a metabolically stable lipoxin analog to infected and inflamed murine lungs suppressed neutrophilic inflammation, decreased pulmonary bacterial burden, and attenuated disease severity, suggesting that lipoxins have therapeutic potential in CF.

5.6 Transplantation

Double lung or heart lung transplantation is a therapeutic option for CF patients with end-stage lung disease. Overall survival of lung transplant patients is poorer compared to other organ transplantation with a three-year survival of approximately 60% in CF patients. Generally, survival is better for adults than for children, but individual centers have reported a survival benefit through lung transplantation in children.

5.7 Treatment of Pancreatic Insufficiency, Nutrition, and Liver Disease

CF patients with poor nutritional status are more prone to chest infections and patients with normal fat absorption have a better pulmonary prognosis. Poor nutritional status has been linked to worse prognosis of CF patients, a finding that has prompted an aggressive approach of maintaining normal weight in CF patients. The introduction of acid-resistant microspheric pancreatic enzyme preparations in the 1970s has greatly improved but not normalized the weight of CF patients, and a European consensus was reached on optimized nutrition in CF. Unfortunately, both oral caloric supplements and enteral feeding by either nasogastric tube or gastrostomy tube have been introduced clinically prior to performing adequate scientific studies. Ursodeoxycholic acid has been shown to normalize elevated liver enzyme levels, but its long-term effect on the evolution of liver disease remains largely unproven. Liver disease is a life-limiting factor in only a few patients, but liver transplantation has been successfully performed in individual CF patients with advanced liver and limited pulmonary disease.

5.8 Animal Models

The possibility to directly modify CFTR in experimental animals by molecular techniques allows insight into the physiologic functions of CFTR and its malfunction in the course of CF. Several CFTR mouse models have been constructed. However, owing to differences in gland distribution between man and mouse, the most important trait of CF, chronic pulmonary infection, has not been reproduced in these animal models. One mouse strain has recently been shown to exhibit pathologic alterations in all organs typically affected in CF, but the pulmonary phenotype lacks chronic infection and inflammation. However, on

the basis of the hypothesis that mutations in the CFTR gene result in increased airway sodium absorption, mice were generated with airway-specific overexpression of ENaC. Indeed, these mice revealed ASL volume depletion, increased mucus production, delayed mucus transport, and mucus adhesion to airway surfaces. Furthermore, defective mucus transport caused a severe spontaneous lung disease, sharing features with CF, including mucus obstruction, goblet cell metaplasia, neutrophilic inflammation, and poorer bacterial clearance. However, the chronic type of infection as seen in CF patients was not observed. Other approaches to study particularly CF lung pathophysiology are based on human tracheal xenografts in immunodeficient severe combined immunodeficiency (SCID) mice and nude mice.

See also Calcium Biochemistry; Endocrinology, Molecular; Neuron Chemistry.

Bibliography

Books and Reviews

Boucher, R.C. (2004) New concepts of the pathogenesis of cystic fibrosis lung disease, *Eur. Respir. J.* **23**, 146–158.

Davis, P.B., Drumm, M., Konstan, M.W. (1996) Cystic fibrosis, *Am. J. Respir. Crit. Care Med.* **154**, 1229–1256.

Döring, G., Hoiby, N. for the Consensus Study Group. (2004) Early intervention and prevention of lung disease in cystic fibrosis: a European consensus, *J Cystic Fibrosis* **3**, 67–91.

Döring, G., Knight, R., Bellon, G. (2000) Immunology of Cystic Fibrosis, in: Hodson, M.E., Geddes, D. (Eds.) *Cystic Fibrosis*, Arnold, London, UK, pp. 109–141.

Driskell, R.A., Engelhardt, J.F. (2003) Current status of gene therapy for inherited lung diseases, *Annu. Rev. Physiol.* **65**, 585–612.

Gibson, R.L., Burns, J.L., Ramsey, B.W. (2003) Pathophysiology and management of pulmonary infections in cystic fibrosis, *Am. J. Respir. Crit. Care Med.* **168**, 918–951.

Pilewski, J.M., Frizzell, R.A. (1999) Role of CFTR in airway disease, *Physiol. Rev.* **79**(Suppl. 1), S215–S255.

Quinton, P.M. (1999) Physiological basis of cystic fibrosis: a historical perspective, *Physiol. Rev.* **79**(Suppl. 1), S3–S22.

Sinaasappel, M., Stern, M., Littlewood, J. et al (2002) Nutrition in patients with cystic fibrosis: a European consensus, *J. Cystic Fibrosis* **2**, 51–75.

Verkman, A.S. (2004) Drug discovery in academia, *Am. J. Physiol.* **286**, C465–C474.

Primary Literature

Boucher, R., Cotton, C.U., Gatzy, J.T. et al (1988) Evidence for reduced Cl− and increased Na+ permeability in cystic fibrosis human primary cell cultures, *J. Physiol.* **404**, 77–103.

Cai, Z., Scott-Ward, T.S., Sheppard, D.N. (2003) Voltage-dependent gating of the cystic fibrosis transmembrane conductance regulator Cl− channel, *J. Gen. Physiol.* **122**, 605–620.

Chappe, V., Hinkson, D.A., Howell, L.D., Evagelidis, A., Liao, J., Chang, X.B., Riordan, J.R., Hanrahan, J.W. (2004) Stimulatory and inhibitory protein kinase C consensus sequences regulate the cystic fibrosis transmembrane conductance regulator, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 390–395.

Cheng, S.H., Gregory, R.J., Marshall, J. et al (1990) Defective intracellular traffic and processing of CFTR is the molecular basis of most cystic fibrosis, *Cell* **63**, 827–834.

Clarke, L.L., Grubb, B.R., Yankaskas, J.R., Cotton, C.U., McKenzie, A., Boucher, R.C. (1994) Relationship of a non-cystic fibrosis transmembrane conductance regulator-mediated chloride conductance to organ-level disease in Cftr(−/−) mice, *Proc. Natl. Acad. Sci. U. S. A.* **91**, 479–483.

Dormer, R.L., Derand, R., McNeilly, C.M., Mettey, Y., Bulteau-Pignoux, L., Metaye, T., Vierfond, J.M., Gray, M.A., Galietta, L.J., Morris, M.R., Pereira, M.M., Doull, I.J., Becq, F., McPherson, M.A. (2001) Correction of

delF508-CFTR activity with benzo(c)quinolizinium compounds through facilitation of its processing in cystic fibrosis airway cells. et al, *J. Cell Sci.* **114**, 4073–4081.

Durie, P.R., Kent, G., Phillips, M.J., Ackerley, C.A. (2004) Characteristic multiorgan pathology of cystic fibrosis in a long-living cystic fibrosis transmembrane regulator knockout murine model, *Am. J. Pathol.* **164**, 1481–1493.

Egan, M.E., Pearson, M., Weiner, S.A., Rajendran, V., Rubin, D., Glockner-Pagel, J., Canny, S., Du, K., Lukacs, G.L., Caplan, M.J. (2004) Curcumin, a major constituent of turmeric, corrects cystic fibrosis defects, *Science* **304**, 600–602.

Farrell, P.M., Kosorok, M.R., Laxova, A. et al (1997) Nutritional benefits of neonatal screening for cystic fibrosis, *N. Engl. J. Med.* **337**, 963–969.

Fischer, H., Schwarzer, C., Illek, B. (2004) Vitamin C controls the cystic fibrosis transmembrane conductance regulator chloride channel, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3691–3696.

Galietta, L.V., Jayaraman, S., Verkman, A.S. (2001) Cell-based assay for high-throughput quantitative screening of CFTR chloride transport agonists, *Am. J. Physiol. Cell Physiol.* **281**, C1734–C1742.

Hung, L.W., Wang, I.X., Nikaido, K., Liu, P.Q., Ames, G.F., Kim, S.H. (1998) Crystal structure of the ATP-binding subunit of an ABC transporter, *Nature* **396**, 703–707.

Huang, P., Lazarowski, E.R., Tarran, R., Milgram, S.L., Boucher, R.C., Stutts, M.J. (2001) Compartmentalized autocrine signaling to cystic fibrosis transmembrane conductance regulator at the apical membrane of airway epithelial cells, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14120–14125.

Hwang, T.C., Sheppard, D.N. (1999) Molecular pharmacology of the CFTR Cl− channel, *Trends Pharmacol. Sci.* **20**, 448–453.

Karp, C.L., Flick, L.M., Park, K.W., Softic, S., Greer, T.M., Keledjian, R., Yang, R., Uddin, J., Guggino, W.B., Atabani, S.F., Belkaid, Y., Xu, Y., Whitsett, J.A., Accurso, F.J., Wills-Karp, M., Petasis, N.A. (2004) Defective lipoxin-mediated anti-inflammatory activity in the cystic fibrosis airway, *Nat. Immunol.* **5**, 388–392.

Knowles, M.R., Clarke, L.L., Boucher, R.C. (1991) Activation by extracellular nucleotides of chloride secretion in the airway epithelia of patients with cystic fibrosis, *N. Engl. J. Med.* **325**, 533–538.

Ko, S.B., Zeng, W., Dorwart, M.R., Luo, X., Kim, K.H., Millen, L., Goto, H., Naruse, S., Soyombo, A., Thomas, P.J., Muallem, S. (2004) Gating of CFTR by the STAS domain of SLC26 transporters, *Nat. Cell Biol.* **6**, 343–350.

Kogan, I., Ramjeesingh, M., Li, C., Kidd, J.F., Wang, Y., Leslie, E.M., Cole, S.P., Bear, C.E. (2003) CFTR directly mediates nucleotide-regulated glutathione flux, *EMBO J.* **22**, 1981–1989.

Konstan, M.W., Byard, P.J., Hoppel, C.L. et al (1995) Effect of high-dose ibuprofen in patients with cystic fibrosis, *N. Engl. J. Med.* **332**, 848–854.

Kopelman, H., Durie, P., Gaskin, K. et al (1985) Pancreatic fluid secretion and protein hyperconcentration in cystic fibrosis, *N. Engl. J. Med.* **312**, 329–334.

Lewis, H.A. et al. (2004) Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator, *EMBO J.* **23**, 282–293.

Locher, K.P., Lee, A.T., Rees, D.C. (2002) The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism, *Science* **296**, 1091–1098.

Mall, M., Grubb, B.R., Harkema, J.R., O'Neal, W.K., Boucher, R.C. (2004) Increased airway epithelial Na+ absorption produces cystic fibrosis-like lung disease in mice, *Nat. Med.* **10**, 487–493.

Mall, M., Kreda, S.M., Mengos, A., Jensen, T.J., Hirtz, S., Seydewitz, H.H., Yankaskas, J., Kunzelmann, K., Riordan, J.R., Boucher, R.C. (2004) The DeltaF508 mutation results in loss of CFTR function and mature protein in native human colon, *Gastroenterology* **126**, 32–41.

Matsui, H., Grubb, B.R., Tarran, R. et al (1998) Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease, *Cell* **95**, 1005–1015.

Morral, N., Bertranpetit, J., Estivill, X., Nunes, V., Casals, T., Gimenez, J., Reis, A., Varon-Mateeva, R., Macek, M. Jr., Kalaydjieva, L. et al (1994) The origin of the major cystic fibrosis mutation (delta F508) in European populations, *Nat. Genet.* **7**, 169–175.

Naren, A.P., Cobb, B., Li, C., Roy, K., Nelson, D., Heda, G.D., Liao, J., Kirk, K.L., Sorscher, E.J.,

Hanrahan, J., Clancy, J.P. (2003) A macromolecular complex of beta 2 adrenergic receptor, CFTR, and ezrin/radixin/moesin-binding phosphoprotein 50 is regulated by PKA, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 342–346.

Quinton, P.M. (2001) The neglected ion: HCO3−, *Nat. Med.* **7**, 292–293.

Ramsey, B.W., Pepe, M.S., Quan, J.M. et al (1999) Intermittent administration of inhaled tobramycin in patients with cystic fibrosis, *N. Engl. J. Med.* **340**, 23–30.

Ratjen, F., Döring, G., Nikolaizik, W. (2001) Eradication of *Pseudomonas aeruginosa* with inhaled tobramycin in patients with cystic fibrosis, *Lancet* **358**, 983–984.

Reddy, M.M., Light, M.J., Quinton, P.M. (1999) Activation of the epithelial Na+ channel (ENaC) requires CFTR Cl− channel function, *Nature* **402**, 301–304.

Riordan, J.R., Rommens, J.M., Kerem, B.S. et al (1989) Identification of the Cystic Fibrosis Gene: Cloning and characterization of complementary DNA, *Science* **245**, 1066–1073.

Rommens, J.M., Iannuzzi, M.C., Kerem, B.S. et al (1989) Identification of the Cystic Fibrosis Gene: Chromosome walking and jumping, *Science* **245**, 1059–1065.

Rosenberg, M.F., Kamis, A.B., Callaghan, R., Higgins, C.F., Ford, R.C. (2003) Three-dimensional structures of the mammalian multidrug resistance P-glycoprotein demonstrate major conformational changes in the transmembrane domains upon nucleotide binding, *J. Biol. Chem.* **278**, 8294–8299.

Rosenstein, B.J., Cutting, G.R. (1998) The diagnosis of cystic fibrosis: a consensus statement, *J. Pediatr.* **132**, 589–595.

Schwiebert, E.M., Morales, M.M., Devidas, S. et al (1998) Chloride channel and chloride conductance regulator domains of CFTR, the cystic fibrosis transmembrane conductance regulator, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2674–2689.

Shcheynikov, N., Kim, K.H., Kim, K.M., Dorwart, M.R., Ko, S.B., Goto, H., Naruse, S., Thomas, P.J., Muallem, S. (2004) Dynamic control of cystic fibrosis transmembrane conductance regulator Cl(−)/HCO3(−) selectivity by external Cl(−), *J. Biol. Chem.* **279**, 21857–21865.

Teem, J.L., Berger, H.A., Ostedgaard, L.S., Rich, D.P., Tsui, L.C., Welsh, M.J. (1993) Identification of revertants for the cystic fibrosis delta F508 mutation using STE6-CFTR chimeras in yeast, *Cell* **73**, 335–346.

Ulrich, M., Herbert, S., Berger, J., Bellon, G., Louis, D., Münker, G., Döring, G. (1998) Localization of *Staphylococcus aureus* in infected airways of patients with cystic fibrosis and in a cell culture model of *S. aureus* adherence, *Am. J. Respir. Cell Mol. Biol.* **19**, 83–91.

Varga, K., Jurkuvenaite, A., Wakefield, J., Hong, J.S., Guimbellot, J.S., Venglarik, C.J., Niraj, A., Mazur, M., Sorscher, E.J., Collawn, J.F., Bebok, Z. (2004) Efficient intracellular processing of the endogenous cystic fibrosis transmembrane conductance regulator in epithelial cell lines, *J. Biol. Chem.* **279**, 22578–22584.

Wagner, C.A., Ott, M., Klingel, K., Beck, S., Melzig, J., Friedrich, B., Wild, K.N., Broer, S., Moschen, I., Albers, A., Waldegger, S., Tummler, B., Egan, M.E., Geibel, J.P., Kandolf, R., Lang, F. (2001) Effects of the serine/threonine kinase SGK1 on the epithelial Na(+) channel (ENaC) and CFTR: implications for cystic fibrosis, *Cell. Physiol. Biochem.* **11**, 209–218.

Wang, F., Zeltwanger, S., Hu, S., Hwang, T.C. (2000) Deletion of phenylalanine 508 causes attenuated phosphorylation-dependent activation of CFTR chloride channels, *J. Physiol.* **524**, 637–648.

Wooldridge, J.L., Deutsch, G.H., Sontag, M.K., Osberg, I., Chase, D.R., Silkoff, P.E., Wagener, J.S., Abman, S.H., Accurso, F.J. (2004) NO pathway in CF and non-CF children, *Pediatr. Pulmonol.* **37**, 338–350.

Worlitzsch, D., Tarran, R., Ulrich, M. et al (2002) Effects of reduced mucus oxygen concentration in airway Pseudomonas infections of cystic fibrosis patients, *J. Clin. Invest.* **109**, 317–325.

Zhang, X.M., Wang, X.T., Yue, H., Leung, S.W., Thibodeau, P.H., Thomas, P.J., Guggino, S.E. (2003) Organic solutes rescue the functional defect in delta F508 cystic fibrosis transmembrane conductance regulator, *J. Biol. Chem.* **278**, 51232–51242.

Zielinski, J., Rozmahel, R., Bozon, D. Kerem, B., Grzelczak, Z., Riordan, J.R., Rommens, J., Tsui, L.C. (1991) Genomic DNA sequence of the cystic fibrosis transmembrane conductance regulator, *Genomics* **10**, 241–248.

Mutagenesis, Malignancy and Genome Instability

Garth R. Anderson, Daniel L. Stoler, and Jeremy D. Bartos
Roswell Park Cancer Institute, Buffalo, NY, USA

1 The Need for Genomic Instability 3
1.1 The Conceptual Framework 3
1.2 Multiple Forms of Genomic Damage 4

2 When Does Genomic Instability Begin? 7
2.1 Onset Early in Tumor Progression 7
2.2 The Leukemia Exception 8

3 The Maintenance of Genomic Integrity 9
3.1 DNA Damage Repair 9
3.2 High-fidelity DNA Replication 10
3.3 Chromosomal Segregation during Mitosis 10
3.4 Destabilized Chromosomes: Telomeres 11
3.5 When All Else Fails: Apoptosis. And without it, Trouble 11

4 In the Clinic 12
4.1 Heterogeneous Tumor Cell Populations and the Therapeutic Target Problem 12
4.2 Genomic Instability as a Clinical Prognostic Tool 13

5 A Difficult Future 14

Bibliography 14
Books and Reviews 14
Primary Literature 15

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

Keywords

Aneuploidy
A cell containing other than the normal diploid number of chromosomes.

Arbitrarily primed PCR and Inter-(Simple Sequence Repeat) PCR
Sampling methods to monitor genomic damage, in which polymerase chain reaction products are compared for tumor and corresponding normal tissue DNA. The former method utilizes arbitrary PCR primer pairs to copy the sampled set, while the latter method utilizes a single PCR primer to copy sequences (200–2000 bp) between relatively closely spaced repeat sites.

Intrachromosomal Instability
A form of genomic instability in which alterations occur within individual chromosomes, but not particularly to repeat sequences.

Microsatellite Instability
A form of genomic instability arising from defects in the DNA mismatch repair system, leading particularly to alterations of small numbers of bases within DNA-repeat sequences.

Spectral Karyotype (SKY)
A technique by which individual chromosomes are labeled with distinctive fluorescent probes, enabling the direct viewing of both the number of chromosomes in a cell and the translocations between chromosomes.

Genomic instability is a widespread and essential feature of the common adult-onset cancers. For solid tumors to arise, several key genes must become mutated (to generate growth, facilitate invasion, enlist nutrient supply, evade the immune response, avoid apoptosis, and so on) and this degree of mutation requires abnormally high mutation rates. To achieve such rates, genes whose normal role is to maintain genomic integrity become mutated, leading to ongoing genomic instability. Exogenous mutagens (carcinogens) additionally contribute to overall genomic damage, particularly in tobacco smokers. Genomic destabilization slowly leads to cancer as cells with mutations in genes advantageous for tumor development become selected, expand their populations, and evolve further. A considerable majority of human genes are involved in multicellular coordination and are not essential for maintaining the viability and reproductive potential of individual cells; thus, a window exists wherein an appropriate degree of genomic instability leads to tumor progression, as somatic evolution selects cells driven by self-interest. As with Darwinian evolution, such somatic evolution inherently leads to genomic diversity,

seen clinically as tumor heterogeneity. Although genomic instability and the ensuing genomic heterogeneity render therapeutic targeting difficult, measurement of the degree and form of genomic instability has considerable potential as a clinical prognostic tool.

1
The Need for Genomic Instability

1.1
The Conceptual Framework

Cancer. What a strange disease! For in biology we are so used to seeing order, stepwise differentiation and development, and precise control. Yet, with the common adult-onset solid tumors, we see imprecision, unpredictable behavior, evasion of targeted therapies, disorganized growth, and chaos. And death. But why is this so? What is the nature of this disease, and why does it not behave like other life? Or does it, and we just need to recognize that cancer has more in common with the random nature of evolutionary biology than it does with the fine coordination of organismic biology governed by precise processes of molecular biology?

To comprehend these cancers, we must first recognize the basic nature of multicellular organisms such as ourselves. Going from a unicellular organism to a multicellular organism was a huge leap in evolutionary history. Single-celled organisms have to cope with the formidable tasks of creating and replicating genetic information, controlling its expression, metabolically capturing energy to synthesize new components, and then coordinating the whole process to replicate their entire selves. These cells have the essential genes and systems to allow self-replication, but their only concern about their neighbors is to outcompete them. In contrast, for complex multicellular organisms like ourselves, with a wide variety of highly differentiated specialized tissues, spatial and temporal control and coordination issues come to predominate. Over the course of our evolutionary history, genes evolved, which enable cells to communicate with one another, responding to signals from other cells in a highly specific manner, while also enabling those same cells to respond to other programs locked in during differentiation. Our cells must cooperate, proliferating only when and where appropriate, and willingly dying when and where appropriate; the necessary genes to effect these behaviors evolved. Further gene-based regulatory mechanisms arose, which cause specific cells to adhere only to appropriate partners, and move only in appropriate manners.

The intricate, evolved coordination process generating this complex cellular behavior is contained within the information stored in each duplicated copy of our genome, present in each of our hundred trillion or so cells, and this coordination process continues, as most of these cells are themselves turning over, at widely varied rates. The complexities of this coordination process are far more intricate than those fundamentally required for unicellular life, and this is manifested within our genomes. For the unicellular eukaryote yeast, only about 4000 genes are essential for its cellular growth and replication (Saccharomyces Genome Database). In contrast, humans contain approximately

10 times as many genes; some of these additional genes generate more complex cellular structures and functions, but most are utilized to coordinate multicellular growth and organization, giving rise to differentiated cells and tissues signaling and responding to one another.

What happens when the quality of a cell's genetic information deteriorates? When damage is severe and extensive, essential gene functions will be lost and the cell dies. But when genetic damage is less severe, and is ongoing over a long period of time, cancer can slowly arise. For if genetic damage occurs at a rate insufficient to kill a cell, a population of mutant cells has time to develop, expand, and acquire additional mutations. With most human genes involved in multicellular coordination, it becomes substantially more likely that a coordination gene will be lost than a gene essential for cellular survival. If both alleles of a gene eventually become lost or mutated, less coordinated growth then occurs. And these daughter cells in turn evolve further, and expand their populations. Selection will occur again and again for cells with greater abilities to proliferate, spread, and evade attack in the form of host defenses or therapeutic treatments. Different daughter cells are free to evolve in different directions; unlike classical Darwinian evolution where natural selection is stringent and competition fierce, the human body is a hospitable environment for many of the successful diverging lineages. This somatically evolved, genomically heterogeneous cell population becomes what we see clinically presenting in the patient as a cancer.

Human evolution occurred in spite of this same basic problem inherent in all multicellular organisms, namely, where loss of genomic stability will lead to cells evolving toward unicellular behavior patterns, with disastrous consequences. We therefore contain within each cell a series of redundant safeguards necessary to maintain the quality of our genetic information for many decades, enabling us to survive as individuals to a point where our progeny are fully capable of surviving and reproducing on their own. If a single gene were sufficient to start the process of somatic evolution eventually leading to malignancy, then somewhere within our large population of cells that individual gene loss would inevitably occur, and we would never be able to survive to adulthood. Redundant safeguards to ensure replication fidelity and efficient damage repair therefore exist to preserve our genomic integrity. Additional apoptotic safeguards destroy cells with badly damaged genomes, reducing the probability of developing cancer to a tolerable level. In spite of this, the disease manages to eventually develop clinically in about one-fourth of us, and is present in nearly all the elderly at a microscopic level. Low mutation rates acting over a long period of time provide a powerful means for even the most stringent safeguards to be overcome by natural selection and cell lineage expansion.

1.2 Multiple Forms of Genomic Damage

Unlike leukemias where a single translocation can complete the two basic steps needed to produce this malignancy, solid tumors require several mutational events before a normal cell becomes an uncontrollably proliferating invasive cancer cell; as described above, this slowly develops through an ongoing process of cell population expansion and natural selection of advantageous mutations. To a large degree, this arises from the mutation of

genes whose role is for the maintenance of genomic integrity, yielding the genomic instability that underlies tumor progression; the continuing genomic instability of cultured tumor cells conclusively establishes a genetic component. Extrinsic carcinogens such as tobacco smoke can initiate the process by mutating genes for the preservation of genomic integrity, and expedite the process by directly generating numerous additional mutational events. Genomic instability and ensuing tumor progression thus become the combined result of endogenous (genetic) and exogenous factors.

Genome instability in cancer is seen in multiple forms. Roughly 10% of adult-onset solid tumors exhibit abundant single or oligobase events seen as microsatellite instability. As discussed below, this arises from defects in DNA mismatch repair during DNA replication. Since repeat sequences are inherently more likely to generate small replication errors because of slipped mispairing during replication, the phenomenon is most readily seen in repetitive microsatellite sequences. A small number of genes such as *TGFβ-RII* contain coding repeat sequences, and in tumors arising by microsatellite instability, mutations in such genes become selected for. These tumors have very few other mutations within nonrepeat coding sequences.

Genome instability in cancer is more often seen on a larger scale, in the form of chromosomal instability or aneuploidy. Entire chromosomes are missing or present in extra copies, arising from inappropriate chromosomal segregation during mitosis generating aneuploidy. Translocations generating recombinant and/or truncated chromosomes additionally occur, and multiple recombination events can arise within a single chromosome. Spectral karyotypic analysis of multiple cells from a single tumor reveals differences between cells, as chromosomal instability continues during tumor growth and genomic heterogeneity develops (Fig. 1).

Intrachromosomal instability adds a third fundamental element to the genomic complexity of solid tumors, in the form of numerous insertions, deletions, inversions, amplifications, and other smaller events. The estimated number of such events is increasing as our ability to detect and resolve them improves, with current rough estimates of around ten thousand events per tumor cell. The failure to detect genomic damage in the forms of microsatellite instability or chromosomal instability should not be taken to mean there is no genomic instability at all in a tumor.

Inherent advantages exist for aneuploidy to cooperate with intrachromosomal instability during tumor progression. In order to remove both alleles of a key regulatory gene through aneuploidy alone, both copies of the chromosome will have to be lost. But along with the gene of interest, other genes essential for cellular viability inevitably become lost, bringing the somatic evolutionary process of tumor progression to an abrupt end. On the other hand, mechanisms, which alone generate small events, also have problems; the probability of hitting both alleles of a given gene become very low as the event size shrinks. But when a process generating small events cooperates with a process generating chromosomal instability, critical genes can readily be mutated and the normal allele lost, still without killing the cell (Fig. 2). The evolutionary process of tumor progression is then free to slowly unfold.

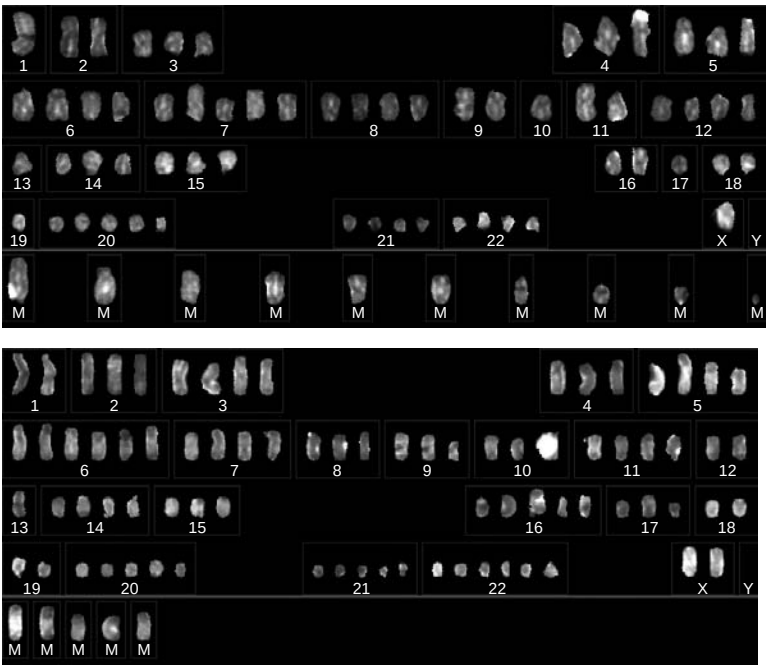

Colon 5

Cell carrying 92 chromosomes with various translocations and derivatives

Cell carrying 84 chromosomes with various translocations and derivatives

Spectral karyotypes of two cells from a single tumor

Fig. 1 Genomic instability gives rise to genomic heterogeneity within each tumor. Two cells from the same tumor, colon 5, differ extensively at the genomic level when analyzed by spectral karyotyping. Chromosomes designated M represent a variety of different translocations. Image is from Bartos et al. (2004).

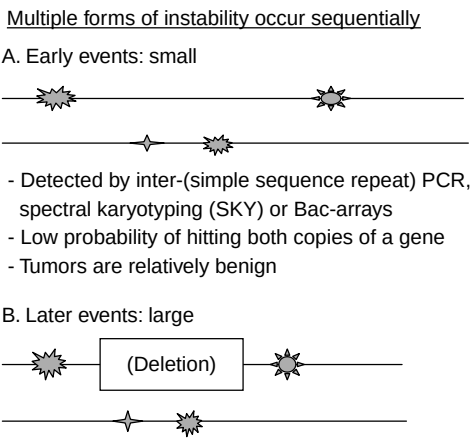

Fig. 2 Multiple forms of genomic instability arise during tumor progression. Inherent advantages exist for small-event instability to cooperate with large-event instability.

This article focuses on genomic instability at the genetic level, where cells acquire permanent heritable changes. An additional level of genomic instability occurs at the epigenetic level, where changes in DNA methylation patterns, particularly of CpG islands, can block the expression of otherwise normal genes. Clinically, this is a significant source of microsatellite instability, when the mismatch repair gene *MLH1* is silenced epigenetically. Complicating the study of epigenetic events in cancer is the finding by Smiraglia and colleagues that DNA methylation patterns of cell lines grown in culture exhibit major differences from the patterns of tumors themselves, presumably reflecting the silencing of genes superfluous to the artificial culture environment.

2 When Does Genomic Instability Begin?

2.1 Onset Early in Tumor Progression

Lawrence Loeb was first to clearly point out the need for genomic destabilization to permit the several requisite steps to convert a normal cell into a malignant tumor cell. This became more evident upon the seminal demonstration by Land et al. that mutation of two cooperating oncogenes alone was insufficient to produce malignancy in mice, and the observations by White, Vogelstein, and others that several events in addition to suppressor gene loss occurred during human colorectal tumor progression. On the basis of known normal mutation rates and the need for more than two mutational events, somatic evolution to malignancy had to somehow be facilitated by decreasing genomic stability.

However, that viewpoint is still not universally accepted. Rubin, Duesberg, and others postulated that normal mutation rates applied to a large, rapidly proliferating cell population, such as the stem cells in colonic crypts, would be likely to generate a rare cell, with exceptional bad luck, containing all the necessary mutations to produce malignancy. This cell would then expand its population inexorably, generating the malignant tumor. By this reasoning, tumors should be essentially homogeneous at the genomic level, but recent spectral karyotyping analyses of multiple tumor cells clearly show this not to be the case.

Verification of the Loeb model has come from multiple directions, centering on demonstration of extensive genomic damage within tumor cells, the finding of extensive genomic heterogeneity within tumors, and tissue culture assays of genomic instability using tumor cell lines showing substantially elevated gene amplification rates. Since genomic instability describes an ongoing process, it is important that these separate forms of evidence exist. Extensive genomic damage occurring in a single burst, as can be seen clinically when induced by radiation therapy or chemotherapy, does not establish that instability exists; in contrast, true genomic instability generates progressively accumulating genomic alterations and genomic diversity within a tumor.

Measuring genomic damage is a complex task, reflecting the diversity of forms of damage that may be present. Flow cytometry measurement of DNA content per cell provides a rough measurement of overall aneuploidy, but karyotyping is essential for identification of specific chromosomal gains and losses. The advent of spectral karyotyping has greatly simplified such analyses, and while also revealing

more complex chromosomal translocation events. Comparative genomic hybridization to entire metaphase chromosomes reveals amplifications and deletions of chromosomes and chromosomal fragments, ranging in size down to a few megabases. These may be integrated within other chromosomes, or may be existing freely in the form of double minutes. Substantially higher resolution is now achievable with the development of ordered Bac-arrays, in which chromosomal regions as small as 150 kb may be analyzed. Finer analyses down to single base size may be achieved through genomic sampling, in which several smaller regions distributed throughout the genome are studied in detail. Arbitrarily primed PCR (polymerase chain reaction) utilizes multiple primers with arbitrary sequences to amplify a small set of broadly distributed products. Inter-(simple sequence repeat) PCR, inter-SSR PCR, utilizes a single primer, anchored within repeat sequences at their 3′ ends facing outward, to amplify those non-repeat sequences between two repetitive elements and generates roughly forty different genomic products per reaction. Both of these techniques provide means of readily sampling the genome to estimate overall damage levels. Direct DNA sequencing of the entire tumor genome is not currently practical, and inevitably becomes complicated by genomic heterogeneity within the tumor cell population. But eventual comparisons of sequence data for individual tumor cell genomes can be expected to yield fascinating data.

Examination of premalignant colorectal adenomas demonstrated that extensive genomic damage begins early in tumor progression. Inter-(simple sequence repeat) PCR analysis shows several thousand genomic events have occurred in early adenomas, a level similar to that seen in frank carcinomas. Such results do not mean that further genomic damage does not accumulate during progression, however. In the inter-SSR PCR analyses, DNA from million cell specimens were used, making later events not present in a majority of cells of the specimen undetectable by this method. Karyotypic analyses of early adenomas, including spectral karyotypes, confirm early extensive genomic damage in the forms of aneuploidy and translocations. Recent work by Pretlow and colleagues, utilizing arbitrarily primed PCR indicates that certain forms of genomic instability begin even prior to the adenoma stage, and can be detected within approximately one-fourth of aberrant crypt foci. These observations are highly important in establishing that genomic instability begins early, putting it in the proper position to facilitate the somatic evolutionary process of tumor progression. Early reports proposing that genomic instability might simply represent one of the numerous phenotypic differences between malignant and normal cells, being a consequence and not a cause of malignancy, are no longer tenable.

2.2 The Leukemia Exception

Acute leukemias, some chronic leukemias, and a few other uncommon childhood/adolescent cancers present exceptions to this model, affecting cell types when they are particularly vulnerable to tumor development. If a normally proliferating cell type in a nonlimiting environment uncontrollably increases its proliferation rate, or loses the ability to differentiate, mature, and ultimately die, the population of this cell type will expand dramatically. To generate these tumors, the number of mutated genes can be as few as two,

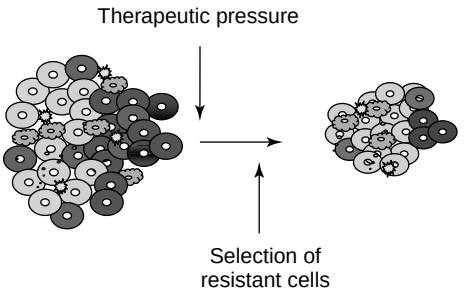

Fig. 3 Extensive genomic heterogeneity within a tumor underlies a considerable part of the eventual failure of pharmacologic therapies for the common, genomically unstable cancers. When therapeutic agents are first applied, susceptible cells are destroyed. But eventually, resistant cells are able to repopulate and further expand the tumor cell population.

which can be generated in a single step by a translocation event that simultaneously rearranges two genes. Such tumors arise rapidly, but due to their genomic homogeneity they also generally respond well to treatment; chronic myelogenous leukemia and its response to Glivec is the classical example of a drug targeted to the translocation event, but childhood acute leukemias also respond well to more conventional therapies. This picture can abruptly change, as such a tumor progresses through blast crisis, giving rise to a genomically unstable heterogeneous tumor cell population capable of evading therapeutic interventions, and in doing so, leukemias can come to resemble the adult-onset solid tumors (Fig. 3).

3 The Maintenance of Genomic Integrity

3.1 DNA Damage Repair

DNA repair is a versatile process evolved to cope with diverse types of DNA damage, which naturally occur. Single base damage elicits base excision repair, while dinucleotide damage (as occurs with UV-induced thymidine dimmers) utilizes nucleotide excision repair. Both of these processes exploit the genetic information present in the undamaged DNA strand. Larger events where breakage occurs can exploit the genetic information present in a sister chromatid or homologous chromosome, through homologous recombinational repair, and still maintain full integrity of the genetic information. While this is feasible in a diploid cell or in a haploid cell in cell cycle phase G2 using the sister chromatid, when allelic losses have occurred or the sister chromatid has not yet been synthesized, the process can fail because of the absence of an intact template. When homologous sequences have been lost or when breakage is extensive, repairs of multiple double-stranded DNA breakage events can still utilize nonhomologous end joining, but with the consequential loss of genomic integrity at the translocated repair sites. Further problems arise if chromosomal fragments are not reintegrated into whole chromosomes; either the genetic information is lost or the small fragment circularizes, with the risk of over-replication and amplification.

In cancer, examples are known for defects in each of these systems, as well as germ-line defects giving rise to hereditary cancer syndromes. The first such case identified was the causal relationship between mismatch repair defects and hereditary nonpolyposis colorectal cancer. Although germ-line defects are found in

the family cancer syndrome, with mutations predominantly in hMLH1 and hMSH2, microsatellite unstable sporadic colon cancers (about 10% of sporadic colon cancers) also exhibit defects in this same system, either arising from somatic mutations or epigenetic silencing. Identification of *BRCA1* and *BRCA2* as the major genes behind familial breast cancer, and the subsequent elucidation of the roles of these genes in homologous recombinational repair, reinforced the concept of defective repair genes underlying both genomic instability and the subsequent cancer arising after decades of somatic evolution.

3.2 High-fidelity DNA Replication

DNA replication is the point where genomic integrity inherently becomes most vulnerable, and most malignancies correspondingly arise out of the relatively rapidly proliferating cell populations such as those lining the gut or in the basal layer of the skin. During replication, the three billion base pairs of the human genome must be copied accurately, completely, and only once, during the few hours that constitute S-phase. This may be occurring in an environment contaminated with carcinogens, reactive oxygen species, or other agents that may damage DNA. And once the DNA is replicated, it must be accurately segregated to daughter cells during mitosis. In spite of these difficulties, and the fact that billions of cells endure these constraints over seven decades, most of us complete our lives never having developed a malignancy.

How is such precision achieved? DNA replication involves over two hundred separate gene products, in a most intricate orchestration. Aside from the highly functional gene products themselves, two keys are the ability to monitor the quality of the replication product, repairing damage immediately, and the ability to temporarily halt replication through checkpoint activation, waiting to resume until the more major damage is repaired. And as a final safeguard, the cellular suicide process of apoptosis is ready to eliminate any cell with irreparable damage to its genomic integrity.

While DNA repair draws most attention as a foundation for genomic instability and tumor progression, uncorrected errors introduced during DNA replication are equally serious and are known to underlie some familial cancers. Mismatched repair errors, in particular, give rise to hereditary nonpolyposis colorectal cancers and endometrial cancers. Errors arising during normal DNA replication are in general within the capability of repair systems to process, but high-level error rates have the potential to exceed normal repair capacities.

3.3 Chromosomal Segregation during Mitosis

The first known form of genomic instability in cancer was aneuploidy, ranging from extra or missing copies of single chromosomes up through tetraploid or higher states, where entire additional copies of the complete genome are present. These phenomena arise out of inaccurate segregation of daughter chromosomes during mitosis. One common diagnostic hallmark of solid tumors is the presence of multipolar mitoses leading to inappropriate segregation.

Centrosomal anomalies typically underlie such inaccurate segregation; these in turn can be a consequence of oncogene activation as shown by Duensing et al. Thus, a stimulus for overproliferation can

promote further tumor progression by giving rise to this particular form of chromosomal instability in turn leading to the loss of suppressor or other growth inhibitory genes, extra copies of growth-promoting genes, and the like. Qualitative changes in these genes, however, do not arise out of aneuploidy; for this reason there is a selective advantage during tumor progression to bring intrachromosomal instability into play. Apparent centrosomal abnormalities can also arise from improper separation of sister chromatids during mitosis, as is seen with separate defects.

3.4 Destabilized Chromosomes: Telomeres

Chromosomal ends are capped with unique structures, telomeres, which are recognized as end structures, and thus are not subjected to ligation-mediated repair. Human telomeres contain eight-base DNA-repeat sequences, totaling around 10 kb in length in young individuals. Telomeres shorten in somatic cells as they undergo repeated rounds of replication as part of the normal maturation process, since the telomerase that generates these cap structures is only active during embryonic development and in germ cell generation.

When telomeres are short or absent, chromosomes become very sticky at their ends as repair is attempted through ligation. This generates a dicentric chromosome, with two centromeres. Alternatively, exonuclease processing may play a significant role in generating dicentric chromosomes. During mitosis, dicentric chromosomes are broken randomly as the daughter cells separate. In turn, these breakpoints are highly unstable, fusing into other chromosomes or generating small circular chromosomal structures. This entire process is known as *bridge-breakage-fusion*, and represents one means of producing translocations and deletions. Amplifications can also arise as circular structures overreplicate and randomly reintegrate.

Telomere deficiency is able to promote tumor development in mice, and human cells with severely shortened telomeres are associated with tumor formation. Since telomerase is reactivated in many tumor cells as part of the acquisition of cellular immortality, telomere-mediated instability is likely to play a larger role relatively early in tumor progression. Attempts at treating cancer by interfering with this reactivated telomerase may be counterproductive as genomic instability is driven higher, although this would relate to the relative rate of increase in genomic instability compared to the relative rate at which the cells then progress to senescence.

3.5 When All Else Fails: Apoptosis. And without it, Trouble

Maintenance of genomic integrity includes the fail-safe mechanism of apoptosis to cause the self-destruction of cells, which in spite of their several DNA damage repair systems and replication checkpoints, still manage to contain significant unrepaired or incompletely repaired damage. Apoptosis utilizes sensors to detect DNA damage; these particularly rely on the ataxia telangiectasia mutated (ATM) protein family to sense broken DNA and trigger a kinase cascade leading through capase activation to cell destruction. Antiapoptotic mechanisms exist to suppress uncalled-for suicide, utilizing in particular the Bcl2 system. Overexpression of Bcl2

can promote hematologic malignancies simply by preventing proper numerical control of specific cell populations.

For solid tumors accumulating extensive DNA damage during their years of progression, defects in the apoptotic pathway are widely seen and play their own crucial role. As genomic damage accumulates, these cells survive and progress further, instead of altruistically eliminating themselves.

4 In the Clinic

4.1 Heterogeneous Tumor Cell Populations and the Therapeutic Target Problem

Medical oncology has achieved notable curative success in the treatment of acute childhood leukemias responding to agents targeting proliferating cells, and excellent if not curative responses have been achieved in a few adult tumors such as chronic myelogenous leukemia, where the etiologic Philadelphia Chromosome translocation provides a specific target in the recombined Abl kinase.

However, medical oncologists have achieved far less success in the treatment of adult-onset solid tumors, where death may be delayed but cure is rarely achieved. The same genome instability necessary for tumor progression to occur has an additional negative consequence as genomic diversity permeates the tumor cell population. Genetic events frequently seen in a tumor type are not present in all cells of the tumor, unless the event occurred very, very early in progression and all tumor cells have descended from a cell with that event. Apc evidently represents such a target for colorectal tumor progression, but even here alternative mutations occurring later in progression may create alternatives to the initiating event. And targeting a totally lost gene, such as with adenovirus cytocidal to p53-negative cells, is proving to be even more difficult than targeting a specific alteration.

This problem can be further illustrated with two antibody-based therapies. Herceptin is used in the treatment of breast cancer, directed against the frequently amplified Her2-neu homolog of the epidermal growth factor receptor, which confers a proliferative advantage. Therapeutic failure occurs after several months, largely as a subpopulation of existing resistant tumor cells become selected for and repopulates the tumor (Fig. 3). Antiangiogenic therapies have had considerable theoretical appeal, in that they are directed against the genomically stable endothelial cells of the tumor vasculature. Experimental studies of mouse models with two peptides (angiostatin and endostatin) were highly successful, but have so far yielded disappointing results in human clinical trials. This has been followed by new approaches using monoclonal antibodies targeted at the major angiogenic factor, vascular endothelial growth factor and its receptor. But even here therapeutic failure occurs within a few months. Heterogeneous tumor cells produce several different angiogenic factors, and the elimination of one system soon selects for cells utilizing other factors.

It now appears reasonable to conclude that therapeutic rehabilitation of adult-onset solid tumor cells is likely to be an exercise in futility. And based on results seen up to the present with even the newest classes of tumor cell targeted agents, the same could relate to therapies, even combination therapies, targeted at specifically killing the tumor cells. For within each such tumor there is a genomically diverse cell population,

exploiting a variety of pathways, and with a broad spectrum of genomic defects in individual cells. Common events will be present in various lineages, akin to the branches of a tree, and various recurring means of effecting proliferation and evading host defenses will be selected, but diversity and heterogeneity will almost always be present to some extent.

So how can cancer ever be cured? We must never forget that currently almost half of all cancers are indeed cured, and those successes are principally the result of surgical resection. If the patient is physically separated from the heterogeneous tumor cell population, the heterogeneity arising out of tumor cell genome instability no longer matters. Surgery, although relatively costly and associated with its own morbidities and mortalities, is the most effective approach to curing cancer, and efforts directed at improving surgical capabilities have genuine potential. Improvements in diagnosis and detection through imaging, molecular diagnostics, and patient-friendly screening approaches, can significantly improve the ability of the surgeon to cure cancer patients. This, combined with proven prevention strategies directed at tobacco and other environmental factors, should not be neglected in favor of glamorous but marginally effective new therapies.

4.2 Genomic Instability as a Clinical Prognostic Tool

Clinical application of genomic instability for purposes of diagnosis and prognosis is constrained by the limitations of assaying a biopsy specimen *in vitro*, and generally at a single point in time. Thus, what is being measured is more accurately genomic damage, but even from this, meaningful clues often can be obtained. Tumors with more extensive damage typically will reflect more rapidly evolving populations, when comparisons are made on a stage-for-stage basis and patients with prior DNA damaging therapies are excluded. Multiple biopsies can reveal genomic heterogeneity within the tumor cell population, which in turn further reflect the likelihood that cells resistant to any given therapy are already present.

The form of genomic instability present also has major clinical implications, in terms of genomic diversity. Microsatellite instability generates approximately 600 000 events, but since few genes contain coding repetitive sequences, relatively little damage to genetic information occurs. And such tumors generally have a good prognosis. With hematologic malignancies, acute childhood diseases driven by a single translocation provide a homogeneous target, and also respond to therapy. But once aneuploidy and intrachromosomal damage becomes extensive, as in blast crisis, therapeutic outcome becomes poor.

The selection of a clinically practicable assay methodology is not a trivial matter. A few approaches have demonstrated their practicability for providing genome-wide information in a clinical setting. Spectral karyotyping, in which chromosomes are painted with specific probes, requires short-term cell culture but reveals considerable information in terms of aneuploidy and chromosomal rearrangements. The cost and expertise level required for this technique presently make it of rather limited general clinical utility.

Array-based approaches comparing tumor DNA with normal DNA effectively reveal amplifications and deletions, in a higher resolution form of comparative genomic hybridization. Such approaches can exploit conventional expression arrays, but with nuclear DNA copies substituting for

cDNAs copied from mRNA. Alternatively, specialized arrays utilizing ordered bacterial artificial chromosomes (Bacs) can representatively scan the entire genome, at a resolving power of about 150 kb. Future improvements in resolution, exploiting single nucleotide polymorphisms, can be anticipated. Current analyses of array data are not yet practical for widespread clinical application due to cost and limitations in data evaluation, although this should be surmounted in the near future.

An alternative approach is the application of genomic sampling methodologies, arbitrarily primed PCR and its variant, inter-(simple sequence repeat) PCR. In these approaches, small pieces of the genome are amplified by PCR, and these genomic samples prepared from tumor DNA are compared with genomic samples prepared from the patient's normal tissue DNA. An electrophoretic comparison of the PCR products reveals what fractions of the bands are altered, providing a quantitative estimate of the degree of genomic damage.

These methodologies have been applied as a prognostic tool for breast cancer, ulcerative colitis, and colorectal cancer, along with Barrett's esophagus. Genomic instability quantitated by inter- (simple sequence repeat) PCR has further been shown to be capable of distinguishing between thyroid tumors and benign lesions, which should be of substantial utility in dealing with relatively common thyroid nodules of indeterminate cytology, which heretofore have been surgically removed as a precautionary measure.

5
A Difficult Future

Genomic destabilization is inherent in the process of development of the common adult-onset cancers. This same instability gives rise to genomic and cellular heterogeneity within the tumor as it presents in the patient; such heterogeneity generates serious problems for nonsurgical therapeutic approaches.

Yet, all is not hopeless. Most cancer cures are currently achieved with surgery, and there remains considerable room for improvement in this area. Promising routes for the future include improving early detection, particularly through the use of proteomic technologies, in turn enabling more timely surgical intervention before metastasis has occurred. Identification of at-risk individuals with familial cancer genes, and definition of tumor specific markers and tests, both fall in the area of early detection. Targeting the genomically stable endothelial cells of the tumor vasculature has its own inherent advantages over targeting the genomically unstable and genomically heterogeneous tumor cell population. And basic prevention measures such as eliminating tobacco use must not be taken lightly.

The cancer problem will be solved neither quickly nor easily. But as the nature of the disease becomes better appreciated, research efforts will hopefully come to focus more on truly productive areas.

See also Epigenetic Mechanisms in Tumorigenesis; Intracellular Signaling in Cancer; Oncology, Molecular.

Bibliography

Books and Reviews

Anderson, G.R., Stoler, D.L., Brenner, B.M. (2001) Cancer: the evolved consequence

of a destabilized genome, *BioEssays* **23**, 1037–1046.

Lengauer, C., Kinzler, K.W., Vogelstein, B. (1998) Genetic instabilities in human cancers, *Nature* **396**, 643–649.

Lindahl, T. (1996) *Genetic Instability in Cancer*, Cold Spring Harbor Press, New York.

Loeb, L.A. (2001) A mutator phenotype in cancer, *Cancer Res.* **61**, 3230–3239.

Moses, R.E. (2001) DNA damage processing defects and disease, *Annu. Rev. Genomics Hum. Genet.* **2**, 41–68.

Nyberg, K.A., Michelson, R.J., Putnam, C.W., Weinert, T.A. (2002) Toward maintaining the genome: DNA damage and replication checkpoints, *Annu. Rev. Genet.* **36**, 617–656.

Shiloh, Y. (2003) ATM and related protein kinases: safeguarding genomic integrity, *Nat. Rev. Cancer* **3**, 155–178.

Van Brabant, A.J., Stan, R., Ellis, N.A. (2000) DNA helicases, genomic instability and human genetic disease, *Annu. Rev. Genomics Hum. Genet.* **1**, 409–459.

Primary Literature

Almasan, A., Linke, S.P., Paulson, T.G., Huang, L.C., Wahl, G.M. (1995) Genetic instability as a consequence of inappropriate entry into and progression through S-phase, *Cancer Metast. Rev.* **14**, 59–73.

Anderson, G.R., Brenner, B.M., Swede, H., Chen, N., Henry, W.M., Conroy, J.M., Karpenko, M.J., Issa, J.P., Bartos, J.D., Brunelle, J.K., Jahreis, G.P., Kahlenberg, M.S., Basik, M., Sait, S., Rodriguez-Bigas, M.A., Nowak, N.J., Petrelli, N.J., Shows, T.B., Stoler, D.L. (2001) Intrachromosomal genomic instability in human sporadic colorectal cancer measured by genome wide allelotyping and inter-(simple sequence repeat) PCR, *Cancer Res.* **61**, 8274–8283.

Basik, M., Stoler, D.L., Kontzoglou, K.C., Rodriguez-Bigas, M.A., Petrelli, N.J., Anderson, G.R. (1997) Genomic instability in sporadic colorectal cancer quantitated by inter-simple sequence repeat PCR analysis, *Genes Chromosomes Cancer* **18**, 19–29.

Ball, C.A., Jin, H., Sherlock, G., Weng, S., Matese, J.C., Andrada, R., Binkley, G., Dolinski, K., Dwight, S.S., Harris, M.A., Issel-Tarver, L., Schroeder, M., Botstein, D., Cherry, J.M. (2001) Saccharomyces genome database provides tools to survey gene expression and functional analysis data, *Nucleic Acids Res.* **29**, 80–81.

Bartos, J.D., Stoler, D.L., Matsui, S.-I., Swede, H., Willmott, L.J., Sait, S.N., Petrelli, N.J., Anderson, G.R. (2004) Genomic heterogeneity and instability in colorectal cancer: spectral karyotyping, glutathione transferase-M1 and Ras, *Mutation Res.* in press.

Bertram, J.S. (2000) The molecular biology of cancer, *Mol. Aspects Med.* **21**, 167–223.

Cahill, D.P., Lengauer, C., Yu, J., Riggins, G.J., Willson, J.K., Markowitz, S.D., Kinzler, K.W., Vogelstein, B. (1998) Mutations of mitotic checkpoint genes in human cancers, *Nature* **392**, 300–303.

Chen, R., Rabinovitch, P.S., Crispin, D.A., Emond, M.J., Koprowicz, K.M., Bronner, M.P., Brentnall, T.A. (2003) DNA fingerprinting abnormalities can distinguish ulcerative colitis patients with dysplasia and cancer from those who are dysplasia/cancer-free, *Am. J. Pathol.* **162**, 665–672.

Cheng, K.C., Loeb, L.A. (1993) Genomic instability and tumor progression: mechanistic considerations, *Adv. Cancer Res.* **60**, 121–156.

Cheung, A.M., Hande, M.P., Jalali, F., Tsao, M.S., Skinnider, B., Hirao, A., McPherson, J.P., Karaskova, J., Suzuki, A., Wakeham, A., You-Ten, A., Elia, A., Squire, J., Bristow, R., Hakem, R., Mak, T.W. (2002) Loss of Brca2 and p53 synergistically promotes genomic instability and deregulation of t-cell apoptosis, *Cancer Res.* **62**, 6194–6204.

Coleman, W.B., Tsongalis, G.J. (1999) The role of genomic instability in human carcinogenesis, *Anticancer Res.* **19**, 4645–4664.

Da Costa, L.T., Lengauer, C. (2002) Exploring and exploiting instability, *Cancer Biol. Ther.* **1**, 212–225.

Davidson, J.F., Guo, H.H., Loeb, L.A. (2002) Endogenous mutagenesis and cancer, *Mutat. Res.* **509**, 17–21.

Duensing, S., Munger, K. (2002) Human papillomaviruses and centrosome duplication errors: modeling the origins of genomic instability, *Oncogene* **21**, 6241–6248.

Duker, N.J. (2002) Chromosome breakage syndromes and cancer, *Am. J. Med. Genet.* **115**, 125–129.

Feldser, D.M., Hackett, J.A., Greider, C.W. (2003) Telomere dysfunction and the initiation of genome instability, *Nat. Rev. Cancer* **3**, 623–627.

Fishel, R., Lescoe, M.K., Rao, M.R., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M., Kolodner, R. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer, *Cell* **75**, 1027–1038.

Garkavtsev, I.V., Kley, N., Grigorian, I.A., Gudkov, A.V. (2001) The Bloom syndrome protein interacts and cooperates with p53 in regulation of transcription and cell growth control, *Oncogene* **20**, 8276–8280.

Gatenby, R.A., Frieden, B.R. (2002) Application of information theory and extreme physical information to carcinogenesis, *Cancer Res.* **62**, 3675–3684.

Gaudet, F., Hodgson, J.G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J.W., Leonhardt, H., Jaenisch, R. (2003) Induction of tumors in mice by genomic hypomethylation, *Science* **300**, 489–492.

Haigis, K.M., Caya, J.G., Reichelderfer, M., Dove, W.F. (2002) Intestinal adenomas can develop with a stable karyotype and stable microsatellites, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8927–8931.

Heinen, C.D., Schmutte, C., Fishel, R. (2002) DNA repair and tumorigenesis: lessons from hereditary cancer syndromes, *Cancer Biol. Ther.* **1**, 477–485.

Hermsen, M., Postma, C., Baak, J., Weiss, M., Rapallo, A., Sciutto, A., Roemen, G., Arends, J.W., Williams, R., Giaretti, W., De Goeij, A., Meijer, G. (2002) Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability, *Gastroenterology* **123**, 1109–1119.

Hoglund, M., Gisselsson, D., Hansen, G.B., Sall, T., Mitelman, F. (2003) Ovarian carcinoma develops through multiple modes of chromosomal evolution, *Cancer Res.* **63**, 3378–3385.

Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D., Perucho, M. (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis, *Nature* **363**, 558–561.

Jackson, A.L., Loeb, L.A. (1998a) On the origin of multiple mutations in human cancers, *Semin. Cancer Biol.* **8**, 421–429.

Jackson, A.L., Loeb, L.A. (1998b) The mutation rate and cancer, *Genetics* **148**, 1483–1490.

Jackson, A.L., Loeb, L.A. (2001) The contribution of endogenous sources of DNA damage to the multiple mutations in cancer, *Mutat. Res.* **477**, 7–21.

Kahlenberg, M.S., Stoler, D.L., Basik, M., Petrelli, N.J., Rodriguez-Bigas, M., Anderson, G.R. (1996) p53 tumor suppressor gene status and the degree of genomic instability in sporadic colorectal cancers, *J. Natl. Cancer Inst.* **88**, 1665–1670.

Khong, H.T., Restifo, N.P. (2002) Natural selection of tumor variants in the generation of "tumor escape" phenotypes, *Nat. Immunol.* **3**, 999–1005.

Kolodner, R.D., Putnam, C.D., Myung, K. (2002) Maintenance of genome stability in Saccharomyces cerevisiae, *Science* **297**, 552–557.

Kucherlapati, M., Yang, K., Kuraguchi, M., Zhao, J., Lia, M., Heyer, J., Kane, M.F., Fan, K., Russell, R., Brown, A.M., Kneitz, B., Edelmann, W., Kolodner, R.D., Lipkin, M., Kucherlapati, R. (2002) Haploinsufficiency of Flap endonuclease (Fen1) leads to rapid tumor progression, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9924–9929.

Kunkel, T.A. (2003) Considering the cancer consequences of altered DNA polymerase function, *Cancer Cell* **3**, 105–110.

Kuraguchi, M., Yang, K., Wong, E., Avdievich, E., Fan, K., Kolodner, R.D., Lipkin, M., Brown, A.M., Kucherlapati, R., Edelmann, W. (2001) The distinct spectra of tumor-associated Apc mutations in mismatch repair-deficient Apc1638N mice define the roles of MSH3 and MSHK6 in DNA repair and intestinal tumorigenesis, *Cancer Res.* **61**, 7934–7942.

Lengauer, C., Kinzler, K.W., Vogelstein, B. (1997) Genetic instability in colorectal cancers, *Nature* **386**, 623–627.

Levitt, N.C., Hickson, I.D. (2002) Caretaker tumour suppressor genes that defend genome integrity, *Trends Mol. Med.* **8**, 179–186.

Loeb, L.A. (2001) A mutator phenotype in cancer, *Cancer Res.* **61**, 3230–3239.

Loeb, L.A., Christians, F.C. (1996) Multiple mutations in human cancers, *Mutat. Res.* **350**, 279–286.

Loeb, K.R., Loeb, L.A. (1999) Genetic instability and the mutator phenotype. Studies in ulcerative colitis, *Am. J. Pathol.* **154**, 1621–1626.

Loeb, L.A., Loeb, K.R., Anderson, J.P. (2003) Multiple mutations and cancer, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 776–781.

Luo, L., Li, B., Pretlow, T.P. (2003) DNA alterations in human aberrant crypt foci and colon cancers by random primed polymerase chain reaction, *Cancer Res.* **63**, 6166–6169.

Marchetti, M.A., Kumar, S., Hartsuiker, E., Maftahi, M., Carr, A.M., Freyer, G.A., Burhans, W.C., Huberman, J.A. (2002) A single unbranched S-phase DNA damage and replication fork blockage checkpoint pathway, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7472–7477.

Myung, K., Kolodner, R.D. (2003) Induction of genome instability by DNA damage in *Saccharomyces cerevisiae*, *DNA Repair (Amst.)* **2**, 243–258.

Nowak, M.A., Komarova, N.L., Sengupta, A., Jallepalli, P.V., Shih, I.e.M., Vogelstein, B., Lengauer, C. (2002) The role of chromosomal instability in tumor initiation, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16226–16231.

O'Sullivan, J.N., Bronner, M.P., Brentnall, T.A., Finley, J.C., Shen, W.T., Emerson, S., Emond, M.J., Gollahon, K.A., Moskovitz, A.H., Crispin, D.A., Potter, J.D., Rabinovitch, P.S. (2002) Chromosomal instability in ulcerative colitis is related to telomere shortening, *Nat. Genet.* **32**, 280–284.

Perucho, M., Welsh, J., Peinado, M.A., Ionov, Y., McClelland, M. (1995) Fingerprinting of DNA and RNA by arbitrarily printed polymerase chain reaction: applications in cancer research, *Methods Enzymol.* **254**, 275–290.

Pihan, G.A., Wallace, J., Zhou, Y., Doxsey, S.J. (2003) Centrosome abnormalities and chromosome instability occur together in pre-invasive carcinomas, *Cancer Res.* **63**, 1398–1404.

Powell, S.N., Willers, H., Xia, F. (2002) BRCA2 keeps Rad51 in line. High-fidelity homologous recombination prevents breast and ovarian cancer? *Mol. Cell* **10**, 1262–1263.

Przybytkowski, E., Girouard, S., Allard, B., Lamarre, L., Basik, M. (2003) Widespread bi-modal intrachromosomal genomic instability in sporadic breast cancers associated with 13q allelic imbalance, *Cancer Res.* **63**, 4588–4593.

Reid, T.M., Fry, M., Loeb, L.A. (1991) Endogenous mutations and cancer, *Princess Takamatsu Symp.* **22**, 221–229.

Rouse, J., Jackson, S.P. (2002) Interfaces between the detection, signaling, and repair of DNA damage, *Science* **297**, 547–551.

Rowley, J.D. (2001) Chromosome translocations: dangerous liaisons revisited, *Nat. Rev. Cancer* **1**, 245–250.

Schmutte, C., Fishel, R. (1999) Genomic instability: first step to carcinogenesis, *Anticancer Res.* **19**, 4665–4696.

Shih, I.M., Zhou, W., Goodman, S.N., Lengauer, C., Kinzler, K.W., Vogelstein, B. (2001) Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis, *Cancer Res.* **61**, 818–822.

Sieber, O.M., Heinimann, K., Gorman, P., Lamlum, H., Crabtree, M., Simpson, C.A., Davies, D., Neale, K., Hodgson, S.V., Roylance, R.R., Phillips, R.K., Bodmer, W.F., Tomlinson, I.P. (2002) Analysis of chromosomal instability in human colorectal adenomas with two mutational hits at APC, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16910–16915.

Snijders, A.M., Fridlyand, J., Mans, D.A., Segraves, R., Jain, A.N., Pinkel, D., Albertson, D.G. (2003) Shaping of tumor and drug-resistant genomes by instability and selection, *Oncogene* **22**, 4370–4379.

Stewart, G., Elledge, S.J. (2002) The two faces of BRCA2, a FANCtastic discovery, *Mol. Cell* **10**, 2–4.

Stoler, D.L., Chen, N., Basik, M., Kahlenberg, M.S., Rodriguez-Bigas, M.A., Petrelli, N.J., Anderson, G.R. (1999) The onset and extent of genomic instability in sporadic colorectal tumor progression, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 15121–15126.

Tauchi, H., Matsuura, S., Kobayashi, J., Sakamoto, S., Komatsu, K. (2002) Nijmegen breakage syndrome gene, NBS1, and molecular links to factors for genome stability, *Oncogene* **21**, 8967–8980.

Tlsty, T.D., Margolin, B.H., Lum, K. (1989) Differences in the rates of gene amplification in nontumorigenic and tumorigenic cell-lines as measured by Luria-Delbruck fluctuation analysis, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9441–9445.

Wang, X., Zou, L., Zheng, H., Wei, Q., Elledge, S.J., Li, L. (2003) Genomic instability and endoreduplication triggered by RAD17 deletion, *Genes Dev.* **17**, 965–970.

Wei, K., Clark, A.B., Wong, E., Kane, M.F., Mazur, D.J., Parris, T., Kolas, N.K., Russell, R., Hou, H. Jr., Kneitz, B., Yang, G., Kunkel, T.A., Kolodner, R.D., Cohen, P.E., Edelmann, W. (2003) Inactivation of exonuclease 1

in mice results in DNA mismatch repair defects, increased cancer susceptibility, and male and female sterility, *Genes Dev*. **17**, 603–614.

Windle, B., Draper, B.W., Yin, Y.X., Ogorman, S., Wahl, G.M. (1991) A central role for chromosome breakage in gene amplification, deletion formation, and amplicon integration, *Genes Dev*. **5**, 160–174.

Zou, L., Cortez, D., Elledge, S.J. (2002) Regulation of ATR substrate selection by Rad17-dependent loading of Rad9 complexes onto chromatin, *Genes Dev*. **16**, 198–208.

Nanobiotechnology

Christof M. Niemeyer
Universität Dortmund, Dortmund, Germany

1 Concepts 21
1.1 General Concept 21
1.2 Topics of Nanobiotechnology 21
1.3 Inorganic Nanoparticles Meet Biomolecules 22

2 Protein-functionalized Nanoparticles 24
2.1 Protein-directed Organization of Particles 24
2.2 Genetic Approaches to Particle Organization 25
2.3 Nanoparticle-protein Conjugates as Biolabels 26
2.4 Nanoparticles Modified with Functional Proteins 28

3 DNA-based Nanoparticle Assemblies 30
3.1 DNA-directed Oligomerization of Nanoparticles 30
3.2 Individual Nanoscaled Particle Assemblies 32
3.3 DNA–protein Conjugates as Tools in Nanofabrication 33
3.4 DNA Biotemplating 36
3.5 DNA–protein Interaction for Biotemplating 38
3.6 DNA-nanoparticle Conjugates at Tools in Bioanalytics 40

4 Magnetic Nanoparticle Conjugates 43
4.1 Magnetic Nanoparticles in Biomedical Sciences 43
4.2 Magnetic Switching of Nanoparticle Devices 44

5 Other Applications 45
5.1 Transfection 45
5.2 Nanoparticles as Model Systems 45

6 Conclusions and Perspectives 46

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

Bibliography 47
Books and Reviews 47
Primary Literature 47

Keywords

Atomic Force Microscopy
Atomic force microscopy (AFM) is the key analytical tool in nanobiotechnology. Initially developed for materials science, AFM is nowadays routinely applied to study biomolecular conformation and also reactivity.

Biotemplating
Utilization of the topographic and electrostatic properties of biological macromolecules as a template to grow inorganic structures by deposition of metals and metal ions. Examples of biomolecular templates include grids and hollow cages formed by protein-building blocks and nucleic acid–based architecture.

Directed Assembly
Organization of components mediated by molecular recognition systems. In Nanobiotechnology, often protein and DNA fragments are used to either direct the assembly of nanoparticles in homogeneous solution or to mediate the immobilization of the particles on a solid substrate.

Nanoparticles
Nanometer-scaled particles from metal or semiconductor materials. The latter are often referred to as ''quantum dots'' (QDs). Nanoparticles reveal strongly size-dependent properties, such as the wavelength of light emission from luminescent QDs.

Supramolecular Structure
An aggregate comprised of various molecular or colloidal components held together by noncovalent interactions, such as hydrogen bonding, Van-der-Waals or Coulomb forces.

Nanobiotechnology is a novel field of science, currently emerging at the intersection of materials research, nanosciences, and molecular biotechnology. This highly interdisciplinary field is closely associated to both the physical and chemical properties of organic and inorganic nanoparticles, as well as to the various aspects of molecular cloning, recombinant DNA and protein technology, and immunology. Evolutionary optimized biomolecules, such as nucleic acids, proteins, and supramolecular complexes thereof, are utilized in the fabrication of nanostructured and mesoscopic architecture from organic and inorganic nanoparticles and vice

versa; technical nanometer-scaled devices are used to exploit structure and reactivity of biological systems. This chapter, in particular, is focused on a key issue of nanobiotechnology, concerning the utilization of biomolecules to functionalize inorganic nanoparticles.

1 Concepts

1.1 General Concept

"Nanobiotechnology" is a novel field of research at the crossroads of biotechnology and nanosciences. Each of these two disciplines themselves are cross sections from various research fields, including fundamental sciences such as physics, chemistry, and biology, as well as engineering and computational sciences. Although nanobiotechnology has often been considered as one of the key technologies for the twenty-first century, its contents and applications are still very much in the process of definition and configuration. Despite the early stage of development, an increasing number of contributions from very diverse directions is claiming the label "nanobiotechnology." Driven by the current hype, for example, even the dispension of nanoliter volumes of aqueous solutions containing biomolecules has been considered as a topic of nanobiotechnology. We will use a more stringent definition here: The use of (individual) biomolecules for technical applications involving devices with at least two dimensions in the submicrometer regime, and/or vice versa, the use of technical nanometer-scaled devices to exploit biological structure and reactivity.

1.2 Topics of Nanobiotechnology

Natural evolution has led to incredibly functional assemblages of proteins, nucleic acids, and other macromolecules to perform complicated tasks that are still daunting for us to try to emulate. As an example, the 20-nm ribosome particle is a most effective supramolecular nanomachine, which spontaneously self-assembles from its more than 50 individual protein and nucleic acid building blocks, thereby impressively demonstrating the power of biologically programmed molecular recognition. Starting with the discovery of the DNA double helix structure, biology has meanwhile grown from a purely descriptive and phenomenological discipline to a molecular science. In view of the revolutionary developments in molecular biosciences, it seems particularly challenging to fuse biotechnology with materials science. The combination of these two disciplines will allow us to take advantage of the evolutionary improved biological components for generating new smart materials, and vice versa, to apply today's advanced materials and physicochemical techniques to solve biological problems. A first systematic and comprehensive survey on this novel interdisciplinary field has recently been published, covering contributions from bioorganic and bioinorganic chemistry, molecular biology, materials science, bioanalytics, and microsystems technology (Table 1).

Tab. 1 Research topics of nanobiotechnology.

I Interphase systems	
1	Biocompatible inorganic devices
2	Microcontact-printing for biosensor applications
3	Dip-pen nanolithography
4	Cell-nanostructure interactions
5	Networks of neuronal cells on silicon substrates
II Protein-based nanostructures	
6	S-layers
7	Nanopore technology
8	Genetic particle organization
9	Microbial nanoparticle production
10	Magnetosoms
11	Bacteriorhodopsin technology
14	Polymer-protein hybrid materials for fuel cells
12	Biomolecular motors
13	Motor proteins and inorganic nanostructures
14	Nanobioelectronic devices
III DNA-based nanostructures	
15	DNA-protein nanostructures
16	Biotemplated mineralization
17	Mineralization in nanostructured biocompartments
18	DNA-nanoparticle conjugates
19	DNA nanostructures for mechanics and computing
IV Nanoanalytics	
20	Quantum dot biolabeling
21	Protein-nanoparticle conjugates in histochemistry
22	Analysis of biomolecular structure by AFM
23	Force spectroscopy
24	Biofunctionalized nanoparticles for SERS and SPR

Owing to the broad-range topics rapidly being developed in the area of nanobiotechnology, it is impossible to cover the entire field, and thus this chapter rather is focused on a key issue of nanobiotechnology, concerning the utilization of biomolecules to functionalize inorganic nanoparticles. The research of hybrid components comprised of biomolecules and inorganic nanoparticles is currently attracting much attention because of the broad spectrum of applications ranging from life sciences to materials and nanosciences.

1.3 Inorganic Nanoparticles Meet Biomolecules

The motivation is based on the unique properties of the two classes of building blocks. Inorganic nanoparticles possess strongly size-dependent optical, electrical, magnetical, and electrochemical properties, while biomolecules, such as proteins and nucleic acids, reveal perfect binding properties and biochemical functionality, optimized within billions of years of evolution. Since they are more readily available by synthetic chemical means and are more conveniently to handle than proteins, nucleic acids have, in the early stages, preferentially been used for the functionalization of nanoparticles to provide highly specific binding sites. Proteins have been employed as well to extend the binding capabilities of inorganic nanoparticles; however, instead of to simply utilize their binding properties, a particular challenge of nanobiotechnology concerns the combination of the intrinsic functionality of many proteins, for instance, optic, catalytic, mechanic or switching capabilities, with the distinguished properties of nanoparticles.

Inorganic nanoparticles are particularly attractive building blocks for the generation of larger superstructures. Nanoparticles can be readily prepared from various materials in large quantities by relatively simple methods, and their dimensions can be controlled from one to about several hundred nanometers,

with a fairly narrow size distribution. Most often, the particles are comprised of metals, metal oxides, and semiconductor materials, such as Ag_2S, CdS, CdSe, CdTe, and TiO_2. Semiconductor nanoparticles are often referred to as "*quantum dots*" (QDs). The nanoparticles have highly interesting optical, electronic, and catalytic properties, which are very different from those of the corresponding bulk materials and which often depend strongly on the particle's size in a highly predictable way. Examples include the wavelength of light emission from semiconductor nanocrystals, which can be utilized for biolabeling or lasers as well as for the external magnetic field required to switch a magnetized particle, which is highly relevant in magnetotactic bacteria and hard disk drives. Moreover, nanoparticles are obtainable as highly perfect nanocrystals, useful as building blocks for the assembly of larger two- and three-dimensional superstructures. Such assemblies have a huge potential as transducers in biosensing as well as in materials and nanosciences.

Both biomolecules and nanoparticles meet at the same length scale (Fig. 1). Biomolecular components have typical size dimensions in the range of about 5 to 200 nm. Owing to their intrinsic functionality, they are potent building blocks with an appropriate size to bridge the gap between the submicrometer dimensions reachable by classical top–down engineering, for example, photolithography and microcontact printing, and the dimensions addressable by conventional bottom-up approaches, such as chemical synthesis and supramolecular self-assembly.

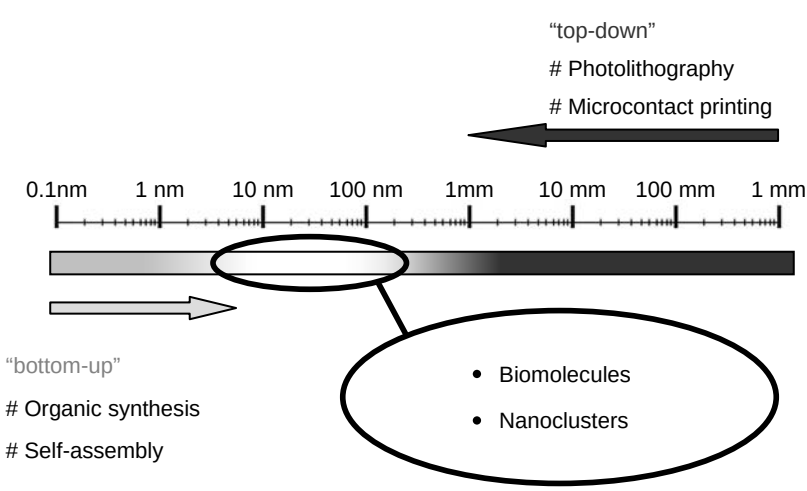

Fig. 1 A gap currently exists in the engineering of small-scale devices. While conventional "top–down" processes hardly allow to fabricate structures smaller than about 200 to 100 nm, the limits of "bottom-up" processes are in the range of about 2 to 5 nm. Owing to their own dimensions, biomolecules, such as proteins and nucleic acids, and inorganic nanoparticles comprised of metal and semiconductor materials, appear to be suited for addressing that gap. Reprinted with permission from Niemeyer, C.M. (2001) Nanoparticles, Proteins, and Nucleic Acids: Biotechnology Meets Materials Science, *Angew. Chem. Int. Ed.* **40**, 4128–4158.

2
Protein-functionalized Nanoparticles

2.1
Protein-directed Organization of Particles

Two- and three-dimensional arrays of metal and semiconductor nanoparticles can be fabricated by means of biomolecular recognition moieties (Fig. 2). Initially, two sets of nanoparticles to be organized are functionalized with individual recognition groups, which are either directly complementary to each other, or are complementary to a molecular linker

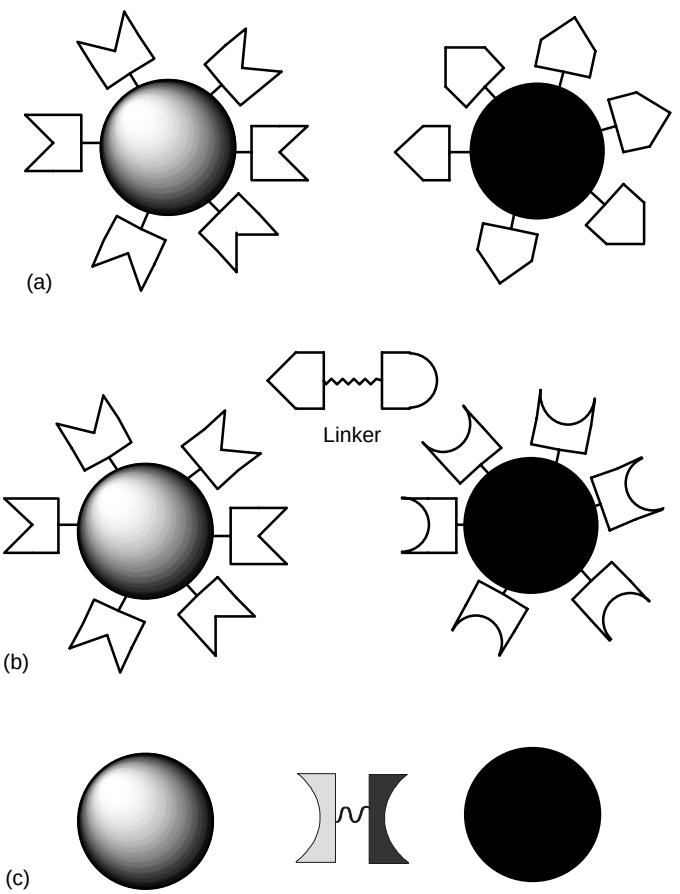

Fig. 2 Schematic drawing of general concepts for solution–phase coupling using biomolecular recognition elements. (a) Two sets of nanoparticles are functionalized with an individual recognition group, which are complementary to each other. (b) The particle-bound recognition groups are not complementary to each other, but can be bridged through a bispecific linker molecule. (c) A bivalent linker, which directly recognizes the surfaces of the nanoparticles is used for oligomerization. Reprinted with permission from Niemeyer, C.M. (2001) Nanoparticles, Proteins, and Nucleic Acids: Biotechnology Meets Materials Science, *Angew. Chem. Int. Ed.* **40**, 4128–4158.

(Fig. 2a and b respectively). Driven by the biospecific interaction, the two particle batches assemble to extended nanoparticle networks, which, in some cases, can grow to the size of macroscopic materials. The major motivation for this approach is based on the unique properties of biomolecular linkers, which promise new applications, such as tunable and switchable materials. In particular, the tremendous recognition capabilities of biomolecules, and their potential to be addressed and manipulated through biochemical procedures using designed tools, such as enzymes, should provide the basis for entirely novel routes to rationally construct advanced materials.

According to the generalized scheme depicted in (Fig. 2b), one example took advantage of the specific interaction between antibodies and low molecular weight organic compounds (haptens). The hapten groups were used to cross-link gold and silver nanoparticles, previously coated with antibody molecules which specifically recognized either the D-biotin or the dinitrophenyl (DNP) group. To allow for the cross-linking of the IgG-functionalized nanoparticles, bivalent linkers were used, which contained two hapten groups in terminal positions. Either the monospecific or bispecific linkers were used, allowing for the directed assembly of homo-oligomeric or heterodimeric aggregates respectively. However, the degree of organization was not as high as expected, indicating that further approaches will require more sophisticated engineering of the antibody–antigen interface. Another example of employing protein-based recognition to organize inorganic nanoparticles was based on the unique interaction between the biotin-binding protein streptavidin (STV) with its low molecular weight ligand D-biotin. The STV-biotin interaction was used to organize gold colloids functionalized by chemisorptive coupling of a disulfide-bridged bis-biotin compound. The particle cross-linking was achieved by the addition of STV as a linker. This process was monitored by spectroscopy and dynamic light scattering (DLS), allowing to determine that the average hydrodynamic radius of all particles in solution rapidly increased upon STV-directed assembly. transmission electron microscopy (TEM) images revealed networks containing on average about 20 interconnected particles, which are separated by about 5 nm, correlating with the diameter of an STV molecule. Recently, an artificial peptide-based linker system was designed to guide supramolecular organization of gold nanoparticles.

2.2 Genetic Approaches to Particle Organization

The power of directed evolution methodology was recently applied to *de novo* generate protein linker units, which directly recognize distinct surfaces of semiconductor materials, thereby avoiding the necessity of previous chemical functionalization (Fig. 2c). Phage-display was used to select 12-mer peptides with binding specificity for distinct semiconductor surfaces. On the basis of a combinatorial library of random 12-mer peptides, comprised of about 10^9 different peptides, phage clones were selected for their specific binding capabilities to five different single-crystal semiconductors, GaAs(100), GaAs(111)A, GaAs(111)B, InP(100), and Si(100). Specific peptide binding was found that is selective for the crystal composition (for example, binding to GaAs, but not to

Si) and crystal face (for example, binding to GaAs(100), but not to GaAs(111). A particular clone was even capable of differentiating between two zinc-blende crystal structures, GaAs and AlAs, which have almost identical lattice constants of 5.66 Å and 5.65 Å respectively. The basis for this selectivity, whether it is chemical, structural, or electronic, is still under investigation. Sequence analysis of various clones has revealed a significantly enhanced portion of amino acids with side chains carrying a Lewis base or Lewis acid group, suggesting that these are important for specific recognition of the substrate. The peptides evolved may be applied to the generation of bispecific linker molecules with binding capabilities for two different semiconductor or metal surfaces (Fig. 2c), thereby paving the way to novel bottom-up approaches for fabricating complex nanoparticle heterostructures.

The genetic selection approach was also used to fabricate highly ordered composite materials from genetically engineered M13 bacteriophage and ZnS nanocrystals. The bacteriophage were coupled with ZnS solution precursors and spontaneously evolved a self-supporting hybrid film material that was ordered at the nanoscale and at the micrometer scale into approximately 70-μm domains, which were continuous over a centimeter length scale. Directed evolution of bacterial surface proteins has been employed to generate proteins with specific recognition properties for gold nanoparticles. Besides opening up novel ways for generating unconventional materials, this work also has implications on the understanding of the fundamental processes of biomineralization, in which protein components control the formation of materials by specifically interacting with growing crystallites.

2.3 Nanoparticle-protein Conjugates as Biolabels

Gold nanoparticles, functionalized with proteins have long been used as tools in biosciences. For instance, antibody molecules adsorbed to 10–40-nm colloidal gold are routinely used in histology, allowing for the biospecific labeling of distinguished regions of tissue samples and subsequent TEM analysis. More advanced, small gold clusters of 0.8 or 1.4 nm diameter can be used for the site-specific labeling of biological macromolecules. These labels have a number of advantages over colloidal probes, including better resolution, stability, and uniformity. Moreover, their small size improves the penetration and more quantitative labeling of antigenic sites. The 1.4-nm particles are the smallest size of gold clusters, which can be seen directly in the conventional electron microscope, allowing for a spatial resolution of about 7 nm when covalently attached to antibody fragments. The clusters visibility can be improved by a wet-chemical silver enhancement step for use in electron or light microscopy for histological purposes, or to detect less than picogram amounts of antigens in immunoblots. The silver enhancement can also applied for the electrical sensing of biological binding events, for instance, via the short-circuiting of microelectrodes (Fig. 3). Gold nanoparticles are immobilized in the gaps of microelectrodes via biospecific interaction such as immunosorption of antibodies. The colloids serve as catalytic cores for the reductive deposition of a conducting layer of silver, which short circuits the two electrodes. The resulting decrease in ohmic resistance is used as a positive signal for the sensing of the biospecific interaction to be detected.

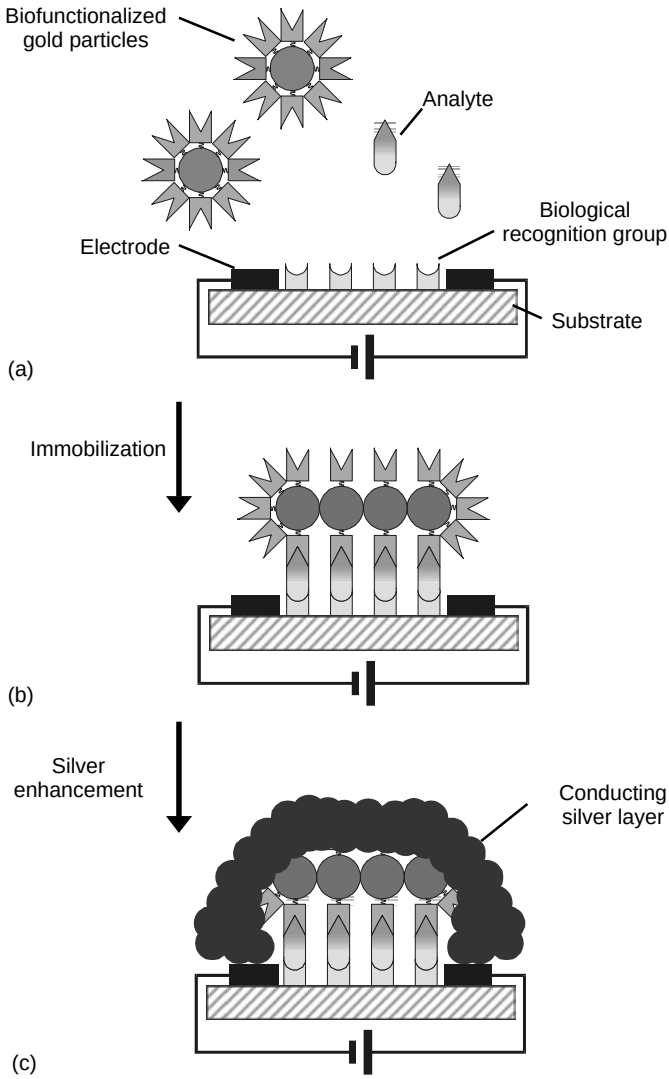

Fig. 3 Electrical sensing of biorecognition events. Receptor groups, such as antibodies or DNA oligomers, immobilized in the gap between two microelectrodes, are used as a capture agent to specifically bind complementary target molecules. In a sandwich-type assay, the captured analyte is tagged with colloidal gold through a second biological recognition unit. Subsequent reductive deposition of silver leads to the formation of a conducting metal layer, which shortcuts the two electrodes. Reprinted with permission from Niemeyer, C.M. (2001) Nanoparticles, Proteins, and Nucleic Acids: Biotechnology Meets Materials Science, *Angew. Chem. Int. Ed.* **40**, 4128–4158.

Gold nanoparticles can also be applied to enhance detection limits in surface plasmon resonance (SPR)-based real-time biospecific interaction analysis. The dramatic enhancement of SPR biosensing using colloidal gold was initially observed in a sandwich immunoassay in which gold nanoparticles were coupled to a secondary antibody. The latter was used to detect a primary antibody bound through specific immunosorption to the antigen immobilized on the gold sensor surface. The immunosorptive binding of the colloidal gold to the sensor surface led to a large shift in plasmon angle, a broadened plasmon resonance, and an increase in minimum reflectance, thereby allowing for picomolar detection of the antigen. Similarly, an about 1000-fold improvement in sensitivity was obtained in nucleic acid hybridization analysis, using a colloidal gold/oligonucleotide conjugate as a probe.

Semiconductor nanoparticles, often referred to as quantum dots (QDs) are powerful fluorescent probes, which can be used for the labeling of biological components. The QDs have several advantages over conventional fluorescent dyes. Their fluorescence absorption and emission is conveniently tunable by their size and material composition, and the emission peaks have a narrow spectral linewidth (Fig. 4). Typically, emission widths are 20–30 nm (full-width at half-maximum, FWHM), which is only one-third of the emission linewidth of a conventional organic dye. The high quantum yields often range from 35 to 50% for CdSe/ZnS core/shell nanoparticles. Moreover, QDs are about 100-fold as stable against photobleaching as an organic dye, and they often reveal a long fluorescence lifetime of several hundreds of nanoseconds. This allows for time-delayed fluorescence measurements, which can be used to suppress the autofluorescence of biological matrices. As demonstrated in numerous examples, QDs are versatile tools for biolabeling, such as immunostaining and fluorescence *in situ* hybridization (FISH).

2.4 Nanoparticles Modified with Functional Proteins

In addition to simply utilizing the recognition properties of proteins, a particular challenge of current research concerns the coupling of the intrinsic functionality of many proteins, for instance, optic, catalytic, mechanic, or switching capabilities, with the distinguished properties of nanoparticles. Initial steps toward this end concerned the optical interaction between QDs and chemisorbed bovine serum albumin and maltose-binding protein-fluorescent dye conjugates, enabling for the self-assembly of bioconjugates that function as a specific chemical sensor for saccharides. Other studies concerned the employment of chaperonin proteins as switchable nanocages for the specific encapsulation and the ATP-triggered release of QDs, and the "nanowiring" of enzymes with electrodes by specific incorporation of 1.4-nm gold nanoparticles into the active site of glucose oxidase redox enzymes. Moreover, mechanically active hybrids comprised of nanostructured inorganic components, and biomolecules were constructed taking advantage of motor proteins. This was achieved from a recombinant F_1-ATPase, immobilized at 80-nm diameter nickel posts and coupled with a nanopropeller of ca. 1000-nm length. The ATPase molecule, which is capable of producing about 100 pN*nm

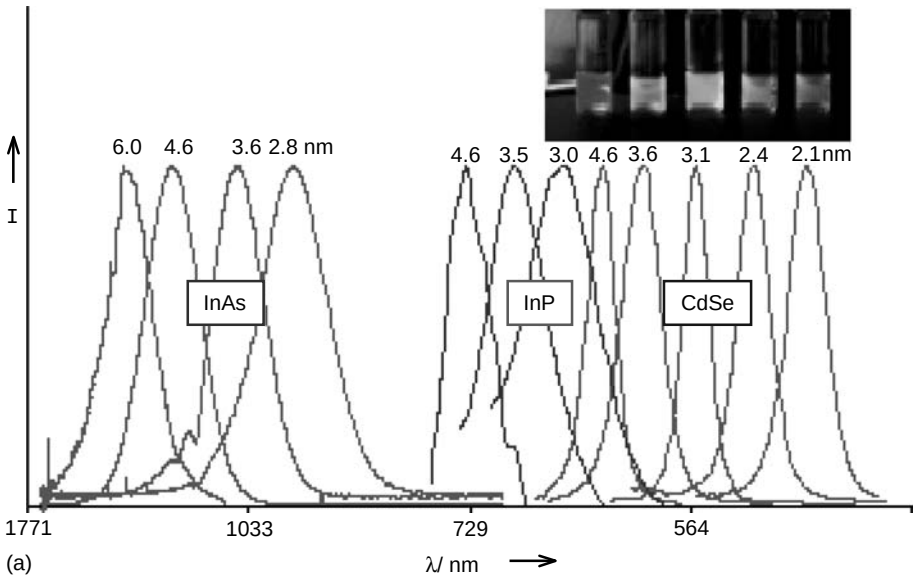

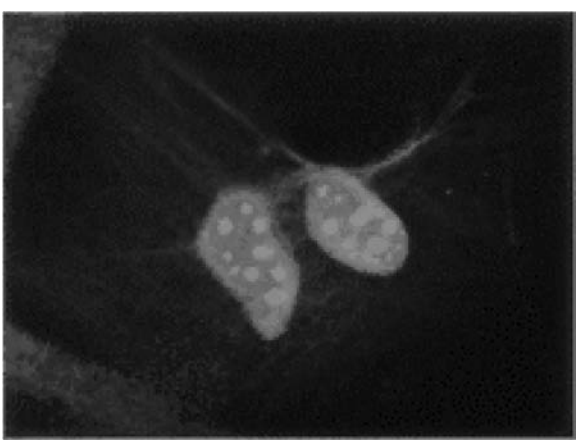

(b)

Fig. 4 (a) Fluorescence absorption and emission spectra of semiconductor quantum dots; (b) Cross section of dual-labeled mouse fibroblasts. The actin fibers are stained red. The nonspecific labeling of the nuclear membrane by both the red and the green probes results in a yellow color. Reprinted with permission from Bruchez, Jr., M., Moronne, M., Gin, P., Weiss, S., Alivisatos, A.P. (1998) Semiconductor nanocrystals as fluorescent biological labels, *Science* **281**, 2013–2015.

of rotary torque by hydrolysis of ATP, rotates the propeller with a velocity of about 5 rounds/s. These reports illustrate how biological systems can be tuned by nanostructured inorganic components and vice versa, how the properties of inorganic particles can be extended by biomolecular mechanisms, and thus these studies have added new facets to nanobiotechnology.

3
DNA-based Nanoparticle Assemblies

DNA is a particularly powerful construction material in nanosciences. The enormous specificity of Watson–Crick base pairing allows the convenient programming of artificial DNA receptor moieties. The power of DNA as a molecular tool is enhanced by our ability to synthesize virtually any DNA sequence by automated methods, and to amplify any DNA sequence from microscopic to macroscopic quantities by PCR. Another very attractive feature of DNA is the great mechanical rigidity of short double helices, so that they behave effectively like a rigid rod spacer between two tethered functional molecular components on both ends. Moreover, DNA displays a relatively high physicochemical stability, and nature provides us with a variety of highly specific enzymes, which allow for processing of the DNA material with atomic precision and accuracy on the angstrom level. No other (polymeric) material offers these advantages, which are ideal for molecular constructions in the range from about 5 nm up to the micrometer scale. Consequently, DNA has already extensively been used for constructions on the nanometer length scale, and recent advances indicate its applicability for fabricating nanomechanical devices. The following sections will focus on the use of DNA as a construction material for the fabrication of nanoparticular assemblies.

3.1
DNA-directed Oligomerization of Nanoparticles

Following the generalized scheme in (Fig. 2b), DNA hybridization has been used to generate repetitive nanocluster materials (Fig. 5). Two noncomplementary oligonucleotides were coupled in separate reactions with 13-nm gold particles via thiol adsorption. A DNA duplex molecule containing a double-stranded region and two cohesive single-stranded ends, which are complementary to the particle-bound DNA, was used as a linker. The addition of the linker to a mixture of the two oligonucleotide-modified nanoparticles led to the aggregation and slow precipitation of a macroscopic DNA-colloid material. This process is reversible. Since the nanoparticles contained multiple DNA molecules, the oligomerized aggregates were three-dimensionally linked, as judged from the TEM analysis. Images of two-dimensional, single-layer aggregates reveal close-packed assemblies of the colloids with uniform particle separations of about 6 nm, corresponding to the length of the DNA linker duplex. This self-assembly can also be used for the generation of binary networks. Owing to the specificity of Watson–Crick basepairing, only heterodimeric "A-B" composites with alternating particle sizes are formed (Fig. 5). For example, two types of gold clusters of either 40 or 5 nm diameters were assembled by DNA hybridization. In the case of an excess of one particle, satellite-like aggregate structures were generated and characterized by TEM. Similarly, oligonucleotide functionalized CdSe/ZnS QDs and gold nanoparticles were organized to form binary nanoparticle networks. TEM analysis revealed that the hybrid metal/semiconductor assemblies exhibited the A-B-structure expected, and fluorescence and electronic absorption spectroscopy indicated cooperative optical and electronic phenomena within the network

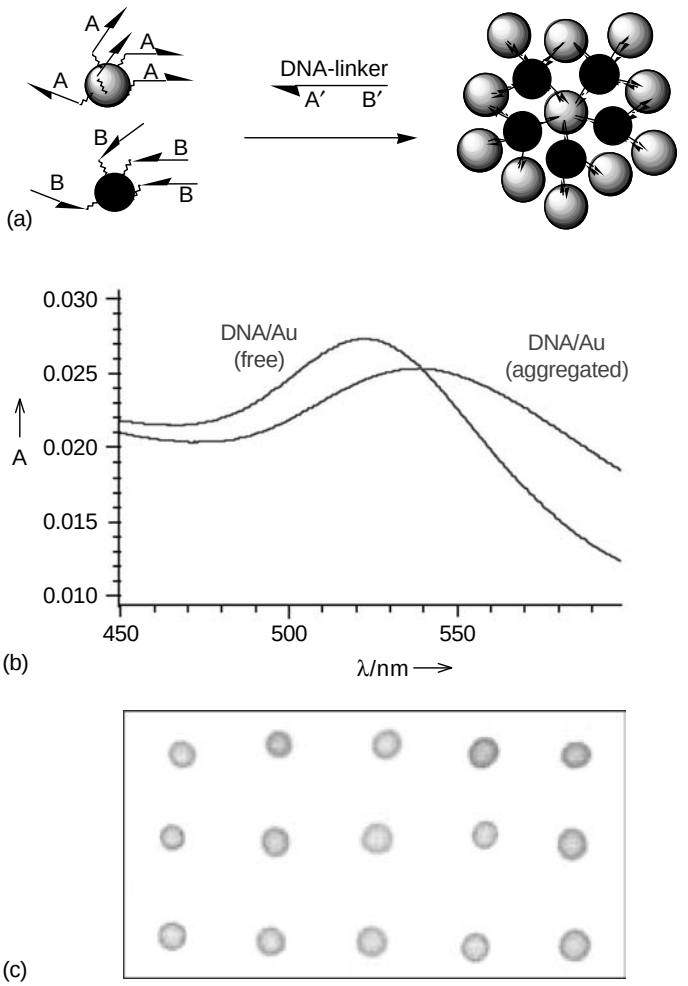

Fig. 5 Assembly of gold colloids using DNA hybridization. (a) Two batches of gold particles, are derivatized with noncomplementary oligonucleotides, either via 5′- or 3′-thiol groups. The nanoparticles are oligomerized using a single-stranded nucleic acid linker molecule. Note that exclusively heterodimeric "A-B" linkages are present within network. The network formation leads to a characteristic change in the plasmon absorption (b). The color change can be used to macroscopically detect DNA hybridization using an assay in which DNA/Au conjugates and a sample with potential target DNA is spotted on a hydrophobic membrane (c). Red spots indicate the absence of fully matched DNA, while blue spots are indicative for the presence of complementary target DNA. Data in (c) are reprinted from Storhoff, J.J., Mirkin, C. A. (1999) Programmed Materials Synthesis with DNA. *Chem. Rev.* **99**, 1849–1862. Copyright 1999 American Chemical Society. (See color plate p. xxvi).

materials. Recently, oligofunctional DNA-gold nanoparticle conjugates were developed. These highly functionalized conjugates contained up to seven different DNA oligonucleotide sequences, which were orthogonally addressable and revealed an efficient reactivity comparable to that of the analogous monofunctional conjugates.

Studies of DNA-linked gold nanoparticle assemblies concerned the influence of the DNA spacer length on the optical and electrical, as well as the melting properties of these networks. The experiments provided evidence that the linker length kinetically controls the size of the aggregates, and that the optical properties of the nanoparticle assemblies are governed by aggregate size. In contrast, it was found that the electrical properties of dried nanoparticle aggregates are not influenced by the linker length. Although SAXS clearly indicated the linker length–dependent distances between particles in solution, the networks collapse upon drying, thereby forming bulk materials comprised of nanoparticles covered with an insulating film of DNA. These materials show semiconductor properties not influenced by the linker lengths. The strongly cooperative melting effect of the nanoparticle networks results from two key factors: the presence of multiple DNA linkers between each pair of nanoparticles and a decrease in the melting temperature as DNA strands melt due to a concomitant reduction in local salt concentration. This mechanism, originating from short-range duplex-to-duplex interactions, is independent of DNA base sequences and should be universal for any type of nanostructured probe that is heavily functionalized with oligonucleotides.

3.2 Individual Nanoscaled Particle Assemblies

The oligomeric DNA-nanoparticle hybrid materials (see Sect. 3.1) have promising electronic and optical properties, potentially useful for applications in materials science, and also have a tremendous impact on immediate diagnostic applications, in particular, DNA microarray–based nucleic acid analyses (see Sect. 3.6). To even increase the enormous potential of the DNA–nanoparticle compounds, it is important to overcome limitations, which may result from the lack of stoichiometric control during the assembly process. To control the architecture of nanomaterials, spatially defined arrangements of molecular devices are required. For example, to organize metal and semiconductor nanocrystals into ultrasmall electronic devices, one may consider a linear aggregate of several individual components (Fig. 6). As previously demonstrated for proteins (see Sect 3.3), gold nanoclusters containing a single nucleic acid moiety have been rationally assembled into stoichiometrically defined nanoscale aggregates (Fig. 6b). To this end, well-defined monoadducts were synthesized from commercially available 1.4-nm gold clusters containing a single reactive maleimido group and thiolated 18-mer oligonucleotides. Subsequent to purification, these conjugates allowed the rational construction of well-defined "nanocrystaline molecules" by DNA-directed assembly using a single-stranded template, which contained the complementary sequence stretches (Fig. 6b). Depending on the template, the DNA-nanocluster conjugates were assembled to generate the head-to-head and head-to-tail homodimeric target molecules **1a** and **1b** in approximately 70% purity, as determined by TEM analysis. More recently, multiple gold nanocrystal aggregates were generated by DNA-directed assembly. The preparations, containing up to three nanoclusters of different sizes organized in multiple fashions, were characterized by TEM. The nanocrystal molecules have a high flexibility when the DNA backbone is nicked (e.g. **1c**), while an unnicked double helix (e.g. **1d**) significantly lowers the flexibility. UV/VIS absorbance measurements indicated changes in the

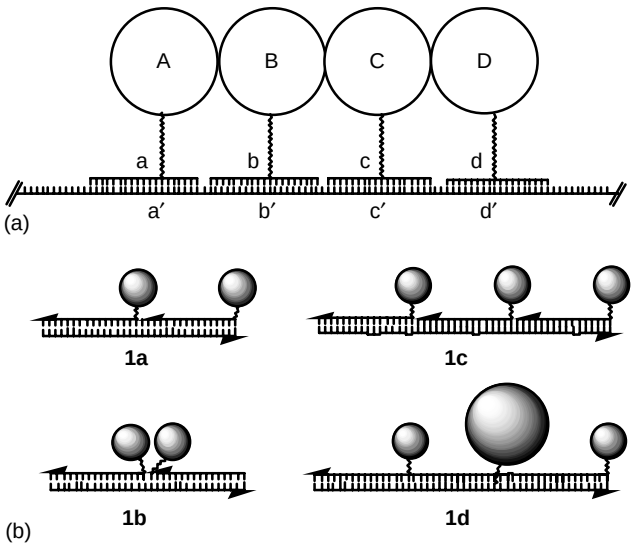

Fig. 6 (a) Schematic representation of DNA-directed assembly of four different nanoscaled building blocks to form a stoichiometrically and spatially defined supramolecular aggregate. (b) Assembly of nanocrystal molecules using DNA hybridization. Conjugates from gold particles (represented as shaded spheres) and 3′- or 5′-thiolated oligonucleotides allowed the fabrication of head-to-head (**1a**) or head-to-tail (**1b**) homodimers. A template containing the complementary sequence in triplicate affects the formation of the trimer **1c**. Aggregate **1d** reveals a limited flexibility owing to the unnicked double helical backbone. Note that, according to the scheme (A), the DNA-directed assembly has been applied to the organization of proteins (see Fig. 7). Reprinted with permission from Niemeyer, C.M. (2001) Nanoparticles, Proteins, and Nucleic Acids: Biotechnology Meets Materials Science, *Angew. Chem. Int. Ed.* **40**, 4128–4158.

spectral properties of the nanoparticles as a consequence of the supramolecular organization.

3.3 DNA–protein Conjugates as Tools in Nanofabrication

The initial demonstration of using DNA as a framework for the precise spatial arrangement of molecular components was carried out with covalent conjugates of single-stranded DNA oligomers and the streptavidin (STV) protein. Using the STV-DNA conjugates **2** (Fig. 7) as model systems, a variety of essential basic studies on the DNA-directed assembly of macromolecules were carried out, which have been reviewed earlier. The covalent DNA-STV conjugates are also convenient as versatile molecular adapters in the nanofabrication of supramolecular assemblies. The covalently attached oligonucleotide moiety supplements the STV's four native biotin-binding sites with a specific recognition domain for a complementary nucleic acid sequence. This bispecificity

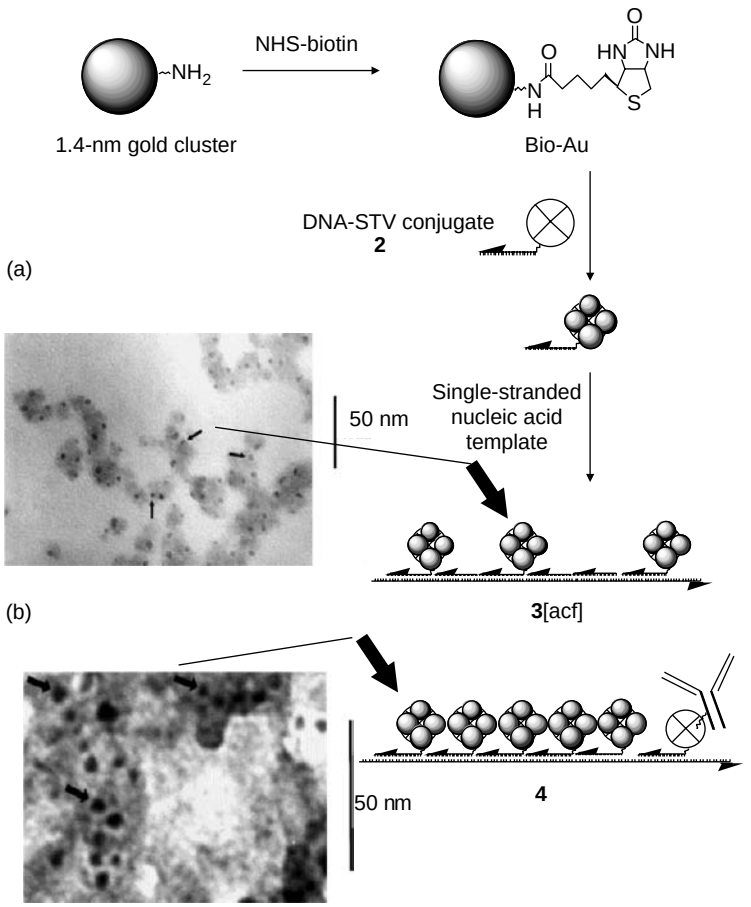

Fig. 7 (a) Schematic representation of DNA-directed assembly of biotinylated gold nanoclusters using covalent DNA-STV conjugates **2** as molecular adapters. The nanocluster-loaded STV conjugates self-assemble in the presence of a single-stranded nucleic acid carrier molecule, containing complementary sequence stretches, to form biometallic aggregates **3**. (b) TEM images of the biometallic aggregate 3[acf] (the letters in brackets indicate for the protein components bound to the template) and an antibody-containing biometallic construct **4** (c). The latter was fabricated from gold-labeled conjugates **2** and a conjugate of **2** and a biotinylated immunoglobulin, previously coupled in separate reactions. Note that the four gold clusters bound to **2** cannot be optically resolved by TEM. Reprinted with permission from Niemeyer, C.M. (2001) Nanoparticles, Proteins, and Nucleic Acids: Biotechnology Meets Materials Science, *Angew. Chem. Int. Ed.* **40**, 4128–4158.

allows to use the DNA-STV conjugates as adapters to assemble basically any biotinylated compound along a nucleic acid template. For example, the strong biotin–STV interaction and the specific nucleic acid hybridization capabilities of the DNA-STV conjugates **2** were utilized to organize gold nanoclusters (Fig. 7). To this end, 1.4-nm gold clusters containing a single amino-substituent were derivatized with a biotin group, and, subsequently, the biotin moiety was used to organize the nanoclusters into the tetrahedral superstructure, defined by the biotin-binding sites of the STV. Subsequently, the nanocluster-loaded proteins self-assemble in the presence of a complementary single-stranded nucleic acid carrier molecule, thereby generating novel biometallic nanostructures, such as **3** (Fig. 7). Since the DNA-STV conjugates can be used like a molecular construction kit, this approach even allows the combined assembly of inorganic and biological components to fabricate functional biometallic aggregates, such as **4**, containing an IgG molecule (Fig. 7). The antibody can be used for targeting the biometallic nanostructures toward certain tissues or other surfaces. This example demonstrates how protein–ligand interaction and DNA hybridization can be used for the nanoconstruction of novel inorganic/bioorganic hybrid systems, and similar approaches should even allow the fabrication of highly complex supramolecular structures. It should be noted, at this point, that Seeman and coworkers have impressively demonstrated the power of DNA in the rational construction of complex molecular framework. For instance, they synthesized the "truncated octahedron," a DNA polyhedron containing 24 individual oligonucleotide arms at its vertices, which can, in principle, be used as a framework for the selective spatial positioning of 24 different proteins, inorganic nanoclusters, and/or other functional molecular devices.

Even today, very little is known on the manipulation and tailoring of oligomeric nanoparticle networks (see Sect. 3.1), for instance, on ways to influence the structure and topography of the DNA hybrid materials subsequent to their formation by self-assembly. Within the context of basic studies on DNA-linked nanoparticle networks, the oligomeric aggregates **5** (Fig. 8) generated from bioorganic streptavidin (STV) particles and bis-biotinylated dsDNA are suitable model systems to gain insight into the properties of complex particle networks. The STV functions as a 5-nm model particle, which can realize only a limited number of interconnects to other particles within the network. Either one, two, three, or four biotinylated DNA fragments can be conjugated with the STV by means of the high affinity STV–biotin interaction. This simplifies the complexity of the supramolecular particle networks, and thus, allows for the convenient analysis and modeling of effects occurring from variations, for example, in the assembly and immobilization parameters. In addition, the size of the dsDNA linker fragments, which are typically about 30 to 170 nm in lengths, allows for convenient direct observation by (atomic force microscopy) AFM (Fig. 8). These findings are convenient for both disciplines involved: The use of the AFM allows to gain insights into biomolecular structure and assembly dynamics, and, in turn, the well-defined and clearly distinguishable DNA-STV nanostructures can be used as soft-materials calibration standards for the detailed study of tip convolution effects, and the influences of measuring parameters in AFM analyses.

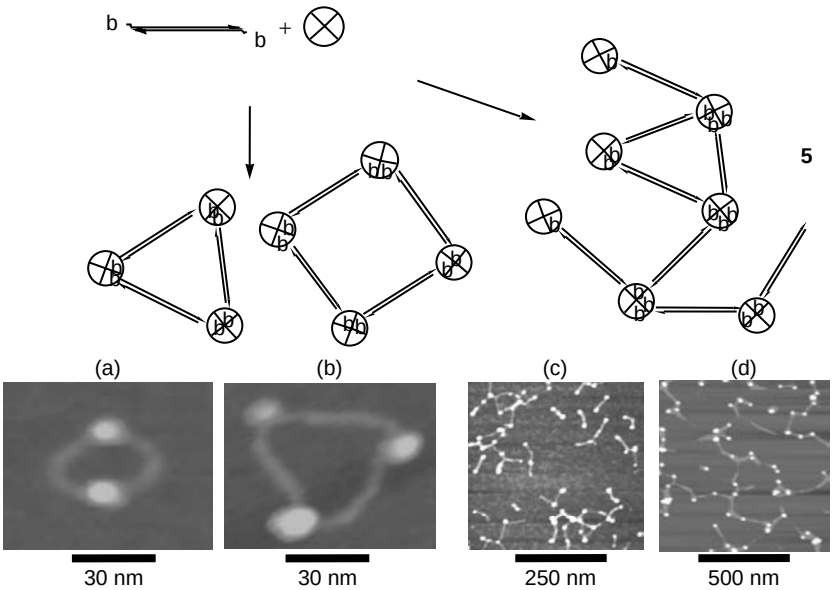

Fig. 8 DNA-protein **5** conjugates as model systems for DNA-interconnected nanoparticles. The synthesis of the networks was achieved by using double-stranded DNA (dsDNA) as spaces. The dsDNA fragments contain two binding sites attached to the two 5′-ends of the double-stranded DNA (b = binding site). The binding sites are either biotinyl groups allowing cross-linking of the biotin-binding protein streptavidin as a model nanoparticle (a-c), or else the binding sites are thiol-groups allowing to connect 5-nm gold colloids (d). The assembly of the two-components not only leads to the formation of oligomeric networks (c, d) but also individual particle aggregates are formed, which can be isolated by electrophoretic or chromatographic methods (a, b). The images at the bottom were obtained by AFM analysis. Reprinted with permission from Niemeyer, C.M. (2001) Nanoparticles, Proteins, and Nucleic Acids: Biotechnology Meets Materials Science, *Angew. Chem. Int. Ed.* **40**, 4128 – 4158.

Mirkin et al. demonstrated that the oligomerization of DNA-coated gold nanoparticles, depicted in Fig. 5, can as well be mediated by conjugates comprised of oligonucleotides and the STV protein. Interestingly, the most thermodynamically stable oligomeric networks are only formed upon heating of the kinetically controlled adducts initially produced. Additional topics of fundamental research on DNA/protein–nanoparticle interaction concern the investigation of whether and how DNA-functionalized particles can be manipulated by DNA-modifying enzymes.

3.4 DNA Biotemplating

The electrostatic and topographic properties of biological macromolecules and supramolecular complexes comprised thereof can be used for the synthesis and assembly of organic and inorganic components. In this "biological templating" approach, regular two-dimensional lattices of bacterial cell surface proteins, hollow biomolecular compartments, such as virus particles, as well as nanometer- and micrometer-sized nucleic

acid components have already been exploited for the generation of nanoparticles and supramolecular architecture. Pioneering work on the DNA-templated generation of metal and semiconductor nanoparticle arrays was carried out by Coffer and coworkers. They used the negatively charged phosphate backbone of the DNA double helix to accumulate Cd^{2+} ions, which were subsequently treated with Na_2S to form CdS nanoparticles. The templated growth of palladium clusters was conducted by using λ-DNA, leading to the formation of DNA-associated 1-D and 2-D arrays of well-separated 3- to 5-nm palladium clusters, which were dependent on the initial nucleation of the metal clusters. This was achieved by activating the DNA with palladium acetate and subsequent reduction using a sodium citrate/lactic acid/dimethylamine borane solution. In the DNA-templated fabrication of mesoscale structures from preformed, positively charged 3-nm CdS particles with a thiocholine modified surface, the particles were found to be assembled in a quasi one-dimensional dense packing to form chainlike structures. The electrostatic assembly of lysine-passivated gold particles on DNA or intercalator-modfied gold particles on DNA and polylysine was also reported.

An example of DNA-biotemplating concerned the use of a 16-μm λ-DNA molecule containing two cohesive ends to bridge the about 12- to 16-μm distance between two gold microelectrodes, prepared by standard photolithography (Fig. 9). Subsequent to functionalization of the electrodes with individual capture oligonucleotides,

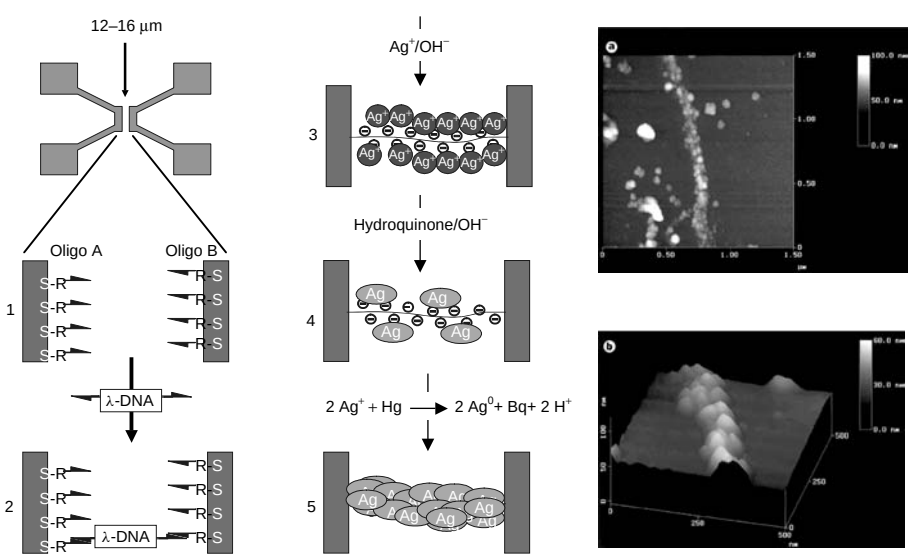

Fig. 9 DNA-templated synthesis of a conductive silver wire. A 16-μm λ-DNA molecule was used to bridge the gap between two gold microelectrodes. The sodium ions bound to the DNA's phosphate backbone were exchanged with Ag^+ ions, and the latter were chemically reduced by hydroquinone to form small silver aggregates. Further reductive deposition of silver led to the formation of a silver nanowire. AFM images revealed the granular structure of the wire. Reprinted with permission from Braun, E., Eichen, Y., Sivan, U., Ben-Yoseph, G. (1998) DNA-templated assembly and electrode attachment of a conducting silver wire, *Nature* **391**, 775–778.

the DNA fragment was allowed to hybridize. Successful interconnection of the electrodes was confirmed by optical microscopy of the fluorescently labeled λ-DNA. Next, the sodium ions bound to the phosphate backbone of the λ-DNA were exchanged with Ag^+ ions, and the latter were chemically reduced by hydroquinone. The small silver aggregates formed along the DNA backbone were then used as catalysts for further reductive deposition of silver, eventually leading to the formation of a silver nanowire. This micrometer-sized element with a typical width of 100 nm had a granular morphology, as determined by AFM (Fig. 9). Two terminal electrical measurements of the silver nanowire revealed nonlinear, history-dependent I-V curves, possibly a result of polarization or corrosion of the individual 30 to 50 nm silver grains that comprise the wires. These examples give initial impressions on how DNA can be utilized as a template to synthesize nanometric and mesoscopic structures. However, the studies have emphasized the importance of fundamental research on how the interaction between DNA and the various binders, such as metal and organic cations, influences the structure and topology of the nucleic acid components involved.

3.5 DNA–protein Interaction for Biotemplating

The examples of Sect. 3.4 clearly indicate the necessity for generating more complex biomolecular architecture, which can be used for the generation of inorganic components by means of biotemplating. An important step in this direction was recently reported, termed *biomolecular lithography*, which took advantage of the sequence-specific binding of conjugates comprised of ssDNA and recA protein. In molecular lithography, the DNA–protein complex is treated with silver ions, which are reduced by aldehyde groups, previously generated within the dsDNA target, thereby forming small silver grains to be subsequently utilized for the wet chemical deposition of gold. This procedure leads to the formation of a conductive wire, which contains an isolating gap, precisely at the position where RecA was bound (Fig. 10a). Therefore, in the molecular lithography, the information encoded in the DNA molecules has replaced the masks used in conventional lithography, while the RecA protein serves as the resist. This approach should work with high resolution over a broad range of length scales from nanometers to many micrometers. Moreover, molecular lithography enables the generation of branch points (three-way junctions) within linear DNA fragments by employing a dsDNA fragment containing a single-stranded end in the initial RecA polymerization step (Fig. 10b). The homologous recombination between RecA conjugate and double-stranded DNA can as well be utilized for the sequence-specific positioning of molecular objects (Fig. 10c). For this, the ssDNA used for polymerization of the RecA monomers is modified with molecular entities, which are functional devices by themselves, or else allow for subsequent binding as such. For instance, gold nanoclusters have been selectively positioned by this approach. Because of the broad scope of molecular lithography, a wide variety of applications might be foreseen.

Semisynthetic DNA–protein nanostructures have been constructed to serve as templates for generating inorganic structures by means of metallization. To this end, oligomeric DNA-STV conjugates **5**, (see Fig. 8), were effectively transformed

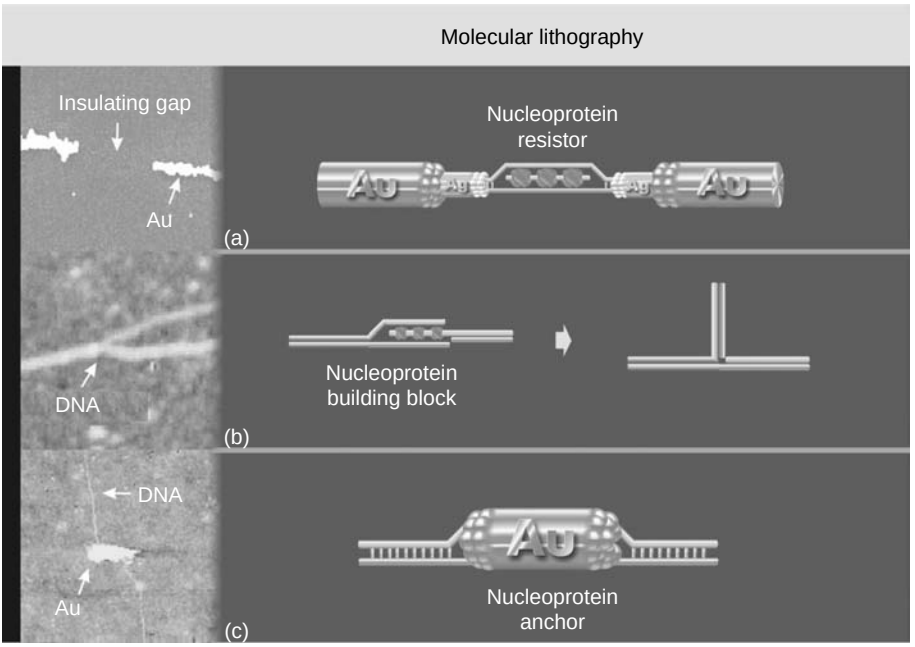

Fig. 10 Molecular lithography. Reprinted with permission from Niemeyer C.M. (2002) Nanotechnology: Tools for the biomolecular engineer, *Science* **297**, 62–63. Copyright 2002 American Association for the Advancement of Science.

into well-defined supramolecular DNA-STV nanocircles **6** by thermal treatment (Fig. 11). Because of their ready availability and well-defined stoichiometry and structure, the nanocircles **6** form the basis of a supramolecular construction kit for generating well-defined DNA-based conjugates. For example, functionalization of **6** with biotinylated haptens allows one to generate hapten conjugates **7**, applicable as reagents in a novel competitive immuno-PCR (cIPCR) assay for the ultra sensitive detection of low-molecular weight analytes. With respect to biomolecular nanotechnology, **5** and **6** can also be used as soft materials topography standardization reagents for AFM. Since the two different biopolymers, DNA and proteins, occur in a highly characteristic, well-defined composition and supramolecular structure, this allows one for direct comparison of, for instance, the deformation properties of the two biopolymers depending on the AFM measurement modes applied.

The conjugates **5** and **6** have also been used as building blocks for generating complex biomolecular nanostructures. As indicated in Fig. 12, networks **5a** and **5b** were generated from the covalent oligonucleotide-STV conjugates **2a** and **2b** respectively, and bis-biotinylated ds-DNA (see AFM images in Figs. 11 and 12). The single-stranded oligonucleotide moieties can be used for further functionalization of **5** by hybridization with complementary oligonucleotide-tagged macromolecules. For instance, oligomers **5a** were transformed into the corresponding

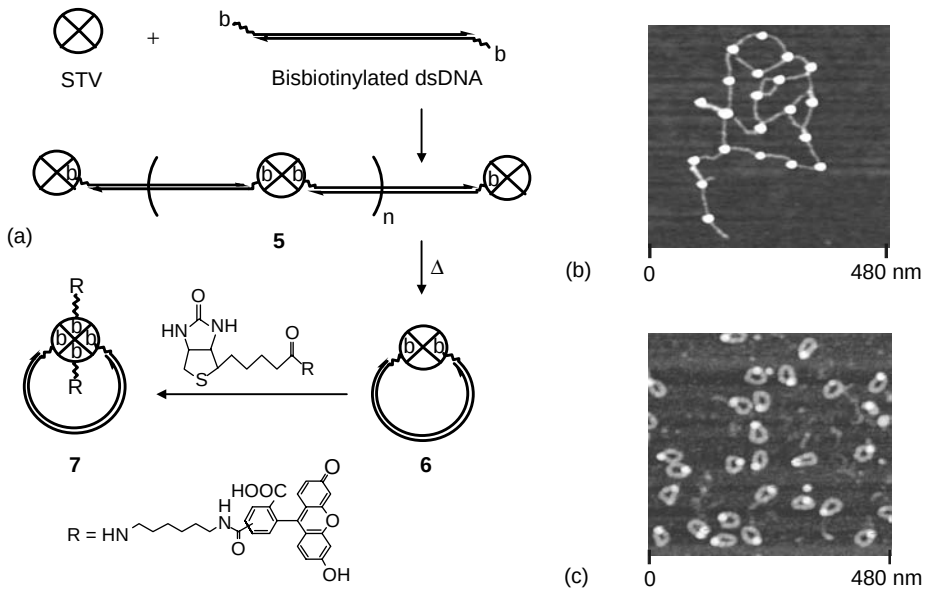

Fig. 11 Nanostructured DNA and STV conjugates. (a) Schematic representation of the self-assembly of oligomeric DNA-STV conjugates **5** from 5′,5′-bis-biotinylated DNA and STV. Note that the schematic structure of **5** is simplified since a portion of the STV molecules function as tri- and tetravalent linker molecules between adjacent DNA fragments (see AFM image b). The supramolecular networks of **5** can be disrupted by thermal treatment, leading to the efficient formation of DNA-STV nanocircles **6** (see AFM image c). The nanocircles **6** can be functionalized by the coupling of biotinylated hapten groups, such as fluorescein (Fsc) to yield nanocircle **7**. For simplification, complementary DNA strands are drawn as parallel lines. The 3′-ends are indicated by the arrow heads. Reprinted with permission from Niemeyer, C.M. et al. (2000) Supramolecular Nanocircles Consisting of Streptavidin and DNA, *Angew. Chem. Int. Ed.* **39**, 3055–3059.

nanocircles **6a**, which can hybridize with analogous circles **6b**, containing a complementary oligonucleotide sequence, to form the nanocircle dimers **8**. The AFM image of the dimeric conjugate **8** indicates the predicted structure (Fig. 12).

3.6 DNA-nanoparticle Conjugates at Tools in Bioanalytics

While protein-coated gold colloids have long been used in bioanalytical techniques, applications of DNA-functionalized gold particles have just been introduced a few years ago by Mirkin and co-workers. The DNA-directed nanoparticle aggregation depicted in Fig. 5 can be used for simple and cheap sensors in biomedical diagnostics, for example, for the detection of nucleic acids from pathogenic organisms. Many analytical applications have already been reported allowing for the detection of nucleic acids in homogenous solutions. In addition, the use of DNA-functionalized nanoparticles in the heterogeneous nucleic acid hybridization with capture oligonucleotides attached to solid supports have recently attracted much attention. Many groups have reported on the DNA-directed

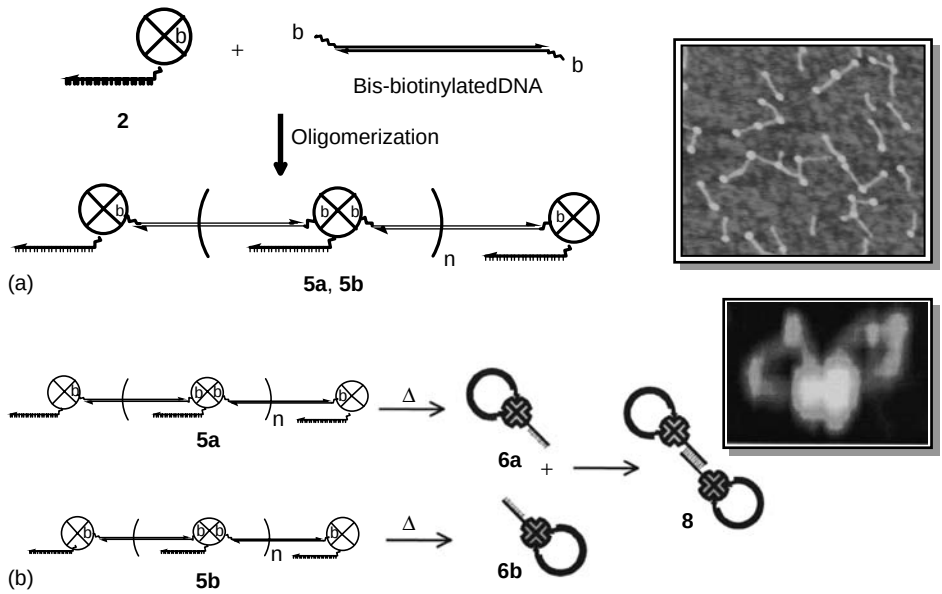

Fig. 12 Construction of supramolecular DNA and STV conjugates. (a.) Networks **5a** and **5b** are generated from covalent oligonucleotide-STV conjugates **2a** and **2b** respectively, containing complementary oligonucleotides. The inset shows a typical AFM image of **5a**. (b) Oligomers **5a**, **b** are transformed into the corresponding nanocircles **6a**, **b**, which hybridize with each other to form the nanocircle dimers **8**. The inset shows a typical AFM imgage of dimer **8**. Reprinted with permission from Niemeyer, C. M., Semisynthetic DNA-Protein Conjugates in C. M. Niemeyer, C. A. Mirkin (Eds.): *NanoBiotechnology: Concepts, Methods, and Applications*, Wiley-VCH, Weinheim 2004, pp. 227–243.

immobilization of gold nanoparticles to form supramolecular surface architecture, and this approach was also adopted to the specific sorting of DNA-functionalized QDs. Particles of different morphology, rods and sticks, have also been assembled by means of DNA hybridization. The specific nucleic acid–mediated immobilization of gold nanoparticles can be utilized for the topographic labeling of surface bound DNA targets. This readily allows for the highly sensitive scanometric detection of nucleic acids in DNA-chip analyses by means of the gold particle promoted reduction of silver ions (Fig. 13), allowing for an about 100-fold increased sensitivity compared to conventional fluorescent DNA detection. In a recent continuation, this approach was adopted to the array-based electronic detection of nucleic acid analytes, using similar setup, as depicted in Fig. 3.

The size-selective light scattering of 50- and 100-nm diameter gold nanoparticles has been used for the two-color labeling of oligonucleotide arrays. An alternative for the simultaneous determination of multiple DNA targets, the use of Raman-active dye-labeled gold nanoparticles, has recently been explored. The gold nanoparticles facilitate the formation of a silver coating that acts as a surface-enhanced Raman scattering promoter for the dye-labeled particles that have been captured by

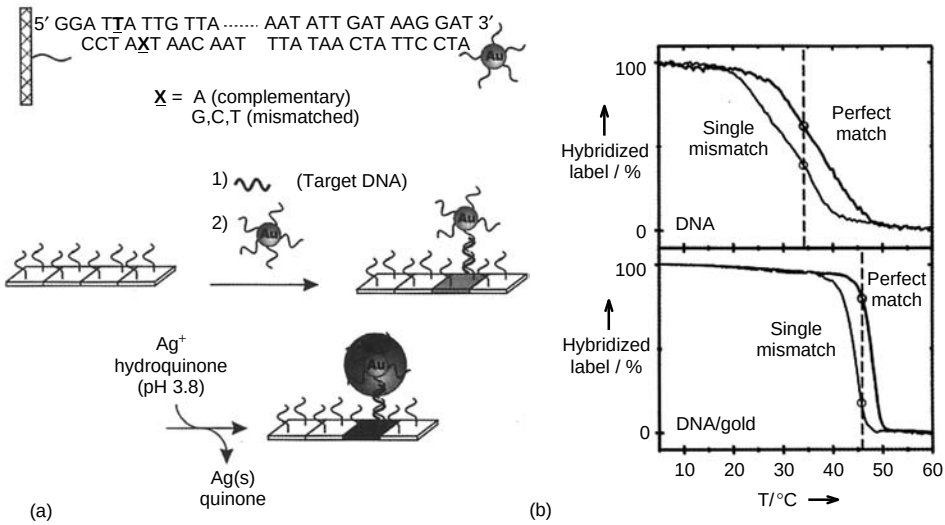

Fig. 13 Scanometric detection of nucleic acids in DNA-chip analyses. Capture oligonucleotides immobilized on glass slides are used for specific binding of target nucleic acids. Oligonucleotide-functionalized gold nanoparticles are employed as probes in solid-phase DNA hybridization detection. Subsequent to a silver enhancement step, the immobilization of the colloidal gold probe is detected by scanning the glass substrate with a conventional flat-bed scanner. Despite its simplicity, this technique allows for an about 100-fold increased sensitivity compared to fluorescent DNA detection. Moreover, owing to the extraordinary sharp melting behavior of the immobilized DNA–nanoparticle networks (see melting curves at the right), this method allows for single-mismatch detection. Reprinted with permission from Taton, T.A., Mirkin, C.A., Letsinger, R.L. (2000) Scanometric DNA array detection with nanoparticle probes, *Science* **289**, 1757–1760.

target molecules and an underlying chip in microarray format. This approach provides the high-sensitivity and high-selectivity attributes of gray-scale scanometric detection, but adds multiplexing and rationing capabilities because a very large number of probes can be designed on the basis of the concept of using a Raman tag as a narrow-band spectroscopic fingerprint. An electrochemical coding technology has been reported for the simultaneous detection of multiple DNA targets. Three encoding inorganic nanocrystal tracers (ZnS, CdS, PbS) were used to differentiate the signals of three DNA targets in connection to stripping-voltammetric measurements of the heavy metal dissolution products. These products yielded well-defined and resolved stripping peaks at the mercury-coated glassy-carbon electrode, thereby enabling femtomole detection limits. The use of colloidal gold nanoparticles also allows for the signal enhancement in the DNA hybridization detection using quartz crystal microbalance, angle-dependent light scattering, and SPR (see Sect. 2.3). DNA-functionalized CdS nanoparticles were used for the photoelectrochemical detection of nucleic acid hybridization. To this end, an array of QDs was assembled at the surface of an electrode using the specific hybridization of complementary DNA fragments. Upon irradiation, a photocurrent was generated, which was proportional to the amount of target DNA.

DNA-tagged proteins have been immobilized at DNA-coated gold nanoparticles by specific nucleic acid hybridization. This approach has several advantages over the conventional methods to adsorb proteins to colloidal gold. The biofunctionalized nanoparticles not only retain the undisturbed recognition properties of the proteins immobilized but they also reveal an extraordinary stability, which even allows for regeneration of the DNA-coated gold particles. The bioinorganic hybrid components were used as reagents in sandwich immunoassays for the detection of proteins using the gold particle promoted silver development. This allowed the spatially addressable detection of fmol-amounts of chip-immobilized antigens. By utilizing oligonucleotide-modified gold nanoparticles encoded with sequences that act as biobarcodes, one can screen for multiple target polyvalent proteins simultaneously in one solution. This concept was demonstrated with two types of detection formats, a homogeneous assay and one based on oligonucleotide microarrays. With such an approach, one can prepare a large number of barcodes from synthetically accessible oligonucleotides. For instance, a 12-mer sequence offers 4^{12} possible barcodes.

4 Magnetic Nanoparticle Conjugates

4.1 Magnetic Nanoparticles in Biomedical Sciences

Nanoparticles comprised of superparamagnetic materials, for example, iron oxide, have already been developed in the early 1990s by Weissleder and coworkers. The particles are typically 5–10 nm in diameter, they possess a half-live of about 90 min in blood, and the major potential applications are as intravenous contrast agents for the lymphatic system, bone marrow contrast agent, long-half-life perfusion agent for brain and heart, and as a magnetic moiety in organ-targeted superparamagnetic contrast agents for magnetic resonance imaging (MRI). With respect to the latter type of applications, in particular, attachment of biomolecular recognition groups to the surface of the inorganic nanoparticles greatly enhances the specificity of targeting. For instance, dextran coated superparamagnetic iron oxide particles were derivatized with a peptide sequence from the HIV-tat protein to improve intracellular magnetic labeling of different target cells. The particles were internalized into lymphocytes over 100-fold more efficiently than nonmodified particles. Labeled cells are highly magnetic and can be retained on magnetic separation columns. This method has great potential for in vivo tracking of magnetically labeled cells by MRI and for recovering intracellularly labeled cells from organs.

Paramagnetic nanoparticles are also being investigated for the treatment of cancer. Biocompatible dextran or silan-coated superparamagnetic magnetite nanoparticles can be incorporated into malignant cells by endocytosis. The intracellular magnetic fluids are excited with AC magnetic fields, leading to an increase in the local temperature, thereby inducing hyperthermia. It has been demonstrated that magnetic fluid hyperthermia (MFH) affects mammary carcinoma cells *in vitro* and *in vivo*.

With respect to biosensing, superparamagnetic iron oxide nanoparticles have been developed that can be used as magnetic relaxation switches to detect molecular interactions in the reversible

self-assembly of disperse magnetic particles into stable nanoassemblies. Monodisperse magnetic nanoparticles conjugated with complementary oligonucleotide sequences self-assemble into stable magnetic nanoassemblies, resulting in a decrease of the spin–spin relaxation times of neighboring water protons. When these nanoassemblies are treated with a DNA- cleaving agent, such as restriction enzymes, the nanoparticles become dispersed, switching the T2 of the solution back to original values. As shown for four different types of molecular interactions (DNA–DNA, protein–protein, protein–small molecule, and enzyme reactions), magnetic relaxation measurements can be used to detect these interactions with high efficiency and sensitivity. Since the magnetic changes are detectable in turbid media,, for example, whole-cell lysates without protein purification, magnetic nanosensors might be used in a variety of biological applications, such as in homogeneous assays, miniaturized microfluidic systems, as affinity ligands for the rapid and high-throughput magnetic readouts of arrays, as probes for magnetic force microscopy, and for *in vivo* imaging.

4.2 Magnetic Switching of Nanoparticle Devices

Magnetic fields have also been used for the remote electronic control of DNA hybridization of gold nanoparticle–DNA molecular beacon conjugates. DNA beacons are ssDNA molecules, which form an intramolecular hairpin-loop structure. Both ends are chemically modified with a fluorophor and a quencher dye, the latter of which, due to close spatial proximity, effectively quenches the fluorescence. Upon hybridization with nucleic acid targets, containing a sequence stretch complementary to the loop region of the beacon, the quencher is spatially removed from the fluorophor, leading to a strong enhancement in fluorescence. A novel class of molecular beacons has been introduced, in which the organic quencher dye was replaced by a gold metal cluster. The quenching efficiency was determined to be about 99.5%, indicating that the fluorescent signal of the beacon increases about 200-fold upon hybridization with complementary target. This type of gold-oligomer-dye hybrid beacons allowed one to detect single base mutations more efficiently than with conventional molecular beacons. Similar gold nanoparticle–DNA molecular beacon conjugates have been exploited for the remote electronic control of DNA hybridization. Inductive coupling of a radio frequency magnetic field (RFMF) with a frequency of 1 GHz to the 1.4-nm metal nanocrystal, which functions as an antenna in the DNA constructs, the local temperature of the bound DNA is increased, thereby inducing denaturation while leaving surrounding molecules relatively unaffected. Since inductive heating has already been applied to the treatment of cancer cells with magnetic field induced excitation, the use of gold nanocrystal–DNA conjugates should allow extension of this concept. For instance, complex operations, such as gene regulation, biomolecular assembly, and enzymatic activity, of distinct portions of nucleic acids or proteins might be controlled, while the rest of the molecule and neighboring species would remain unaffected. Moreover, because the addressing is not optical, this technology would even be applicable in highly scattering media.

The magnetic switching of biomolecular functionality has been demonstrated in a couple of examples using micrometer-sized superparamagnetic iron

oxide particles was demonstrated by Willner and colleagues. As an example, magnetic-field-stimulated "ON" and "OFF" biochemiluminescence was accomplished by electrocatalyzed reduction of naphthoquinone-functionalized magnetic particles in the presence of a biocatalytic peroxidase/luminol system. The dual biosensing by magnetocontrolled bioelectrocatalytic analysis of glucose and lactate by the enzymes glucose oxidase and lactate dehydrogenase was conducted by using ferrocene-functionalized electrodes and enzyme-modified iron oxide magnetic particles. Wang et al. have demonstrated the magnetic triggering of the electrical DNA detection, realized through a magnetic collection of assemblies, formed by specific hybridization of oligonucleotide-gold nanoparticles with microbeads containing complementary oligonucleotides. The assemblies are subjected to silver development using hydrochinon, and the resulting conglomerate is positioned on an electrode by an external magnetic field. The chronopotentiometric hybridization signals allowed the detection of small amounts of target DNA. The development of tiny nanometer-sized functionalized magnetic particles will even increase the scope of applications in nanobiotechnology and biomedical diagnostics.

5 Other Applications

5.1 Transfection

It is well established to use gold particles as carriers for the delivery of double-stranded DNA in the so-called *gene gun* technology. In this method, plasmid DNA of a typical length of several 1000 base pairs is adsorbed to gold or tungsten colloids, which have a typical size of about 500 nm to several micrometers. The DNA- or RNA-coated particles are loaded into a gun-like device in which a low-pressure helium pulse delivers the particles into virtually any target cell or tissue. As an advantage of this technique, one does not have to remove cells from tissue in order to transform cells. The particles penetrate the cells and release the DNA, which is diffused into the nucleus and, for instance, incorporated into the chromosomes of the organism. Since its development in the mid-1980s, the gene gun technology has led to innumerous examples demonstrating the tremendous potential of this method for "biolistic" transfection of organisms as well as for DNA immunization. The latter technique, also known as *DNA vaccination* or *genetic immunization*, is a vaccine technology that can be used to stimulate protective immunity against many infectious pathogens, malignancies, and autoimmune disorders in animal models. Plasmid DNA encoding a polypeptide antigen is introduced into host cells, for instance, via gene gun particle bombardment, where it serves as a template for the high-efficiency translation of its antigen. This leads to an immune response, which is mediated by the cellular and/or humoral immune system and is specific for the plasmid-encoded antigen.

5.2 Nanoparticles as Model Systems

Two general aspects of metal and semiconductor nanoparticles make them suitable model systems for the study of basic biological phenomena. On the one hand, their size matches with the macromolecular components employed in living systems. Proteins, nucleic acid fragments and their

supramolecular complexes, such as the nucleosome, the various complexes involved in DNA replication and transformation, and the ribosome, have typical dimensions in the range of 2 to 200 nm. On the other hand, nanoparticles of inorganic materials are the basic building blocks in biomineralization, a fundamental biological process in which nature chemically generates form by means of genetic instructions. It is thus not surprising that the current advances in the study of metal and semiconductor nanoparticles has led to their consideration as model systems for both, the study of the interactions between proteins and nucleic acids, and also for the understanding of the basic principles of biomineralization. This type of applications have been recently reviewed elsewhere.

6 Conclusions and Perspectives

Although the chemical coupling of biomolecules and inorganic materials can be achieved with a variety of methods, there is still a great demand for mild and selective coupling techniques, which allow for the preparation of thermodynamically stable, kinetically inert, and stoichiometrically well-defined bioconjugate hybrid nanoparticles. The first steps to using these hybrid compounds as building blocks for the bottom-up assembly of sophisticated nanostructured architecture have already been achieved using protein- or nucleic acid–based recognition elements. Directed evolution, already applied to tailor protein linkers, should be applicable to generate novel nucleic acid–based linkers. It seems safe to predict that the tremendous advances in the development of artificial nucleic acid receptors, aptamers selected to specifically recognize any target structure, will soon be applied to the *de novo* generation of nucleic acid receptors, capable of recognizing inorganic surfaces. Consequently, combinations of the two fundamental biological systems, nucleic acids and proteins, are very promising to synergistically cooperate and allow for novel functions and applications. Another crossover of biotechnology and materials science concerns the utilization of the chemical and topographical properties of biological components for templating the spatial assembly of organic and inorganic components to form nanometer- and micrometer-sized elements. Although commercial applications are rare at current state, the interdisciplinary work has a great potential for generating advanced materials, which might lead to novel devices for sensing, signal transduction, catalysis, as well as for new biocompatible materials and interfaces currently being developed for biomedical sciences and tissue engineering. In addition to long-term perspectives, today's nanoparticles are already used in bioanalytical applications or as model systems to solve biological problems.

The future development of the joint venture of biotechnology and material sciences will profit from the rapid current advances of chemistry as a central science. In addition, the current genome and proteome research will also be beneficial for this field, since it provides data, which will allow us to produce even more suitable biocomponents. This might open the door to nanosensors, catalytic and light-harvesting devices, ultra fast molecular switches and transistors, supramolecular mediators between electrical and living systems and other bio- and optoelectronic parts. Clearly, we are still far off from the fabled autonomous nanomachines

that heal wounds and perform surgery in a living organism. However, at the beginning of this new century, one should remember the dramatic development total organic synthesis has undergone from the early 1900s to the race for "molecular summits" in the late 1990s. A similar rate of progress in supramolecular sciences at the interface of biotechnology and materials research promises plenty of excitement from future developments.

See also Motor Proteins; Nucleic Acid and Protein Single Molecule Detection and Characterization; Optimization of Proteins for Molecular and Biomolecular Electronic Devices.

Bibliography

Books and Reviews

Kreuter, J. (1992) in: Donbrow, M. (Ed.) *Microcapsules and Nanoparticles in Medicine and Pharmacy*, CRC Press, Boca Raton, FL.

Mirkin, C.A. (2000) Programming the assembly of two- and three-dimensional architectures with DNA and nanoscale inorganic building blocks, *Inorg. Chem.* **39**, 2258–2272.

Niemeyer, C.M. (2001a) Bioorganic applications of semisynthetic DNA-protein conjugates, *Chem. Eur. J.* **7**, 3188–3195.

Niemeyer, C.M. (2001b) Nanoparticles, proteins, and nucleic acids: biotechnology meets materials science, *Angew. Chem. Int. Ed.* **40**, 4128–4158.

Niemeyer, C.M. (2002a) Nanotechnology: tools for the biomolecular engineer, *Science* **297**, 62–63.

Niemeyer, C.M. (2003a) Functional hybrid devices of proteins and inorganic nanoparticles, *Angew. Chem. Int. Ed.* **42**, 5796–5800.

Niemeyer, C.M., Adler, M. (2002b) Nanomechanical devices based on DNA, *Angew. Chem. Int. Ed.* **41**, 3779–3783.

Niemeyer, C.M., Mirkin, C.A. (Eds.) (2004) *NanoBiotechnology: Concepts, Methods and Applications*, Wiley-VCH, Weinheim, Germany.

Schmid, G. (1994) *Clusters and Colloids*, VCH, Weinheim, Germany.

Seeman, N.C. (2003) DNA in a material world, *Nature* **421**, 427–431.

Shipway, A.N., Katz, E., Willner, I. (2000) Nanoparticle arrays on surfaces for electronic, optical, and sensor applications, *Chem. Phys. Chem.* **1**, 19–52.

Sleytr, U.B., Messner, P., Pum, D., Sara, M. (1999) Kristalline Zelloberflächen-Schichten prokaryotischer Organismen (S-Schichten): von der supramolekularen Zellstruktur zur Biomimetik und Nanotechnologie, *Angew. Chem. Int. Ed.* **38**, 1034–1054.

Storhoff, J.J., Mirkin, C.A. (1999) Programmed materials synthesis with DNA, *Chem. Rev.* **99**, 1849–1862.

Willner, I., Katz, E. (2003) Magnetic control of electrocatalytic and bioelectrocatalytic processes, *Angew. Chem. Int. Ed.* **42**, 4576–4588.

Primary Literature

Alivisatos, A.P., Johnsson, K.P., Peng, X., Wilson, T.E., Loweth, C.J., Bruchez, Jr., M.P., Schultz, P.G. (1996) Organization of 'nanocrystal molecules' using DNA, *Nature* **382**, 609–611.

Braun, E., Eichen, Y., Sivan, U., Ben-Yoseph, G. (1998) DNA-templated assembly and electrode attachment of a conducting silver wire, *Nature* **391**, 775–778.

Brown, S. (2001) Protein-mediated particle assembly, *Nano Lett.* **1**, 391–394.

Bruchez, Jr., M., Moronne, M., Gin, P., Weiss, S., Alivisatos, A.P. (1998) Semiconductor nanocrystals as fluorescent biological labels, *Science* **281**, 2013–2015.

Cao, Y.C., Jin, R., Mirkin, C.A. (2002) Nanoparticles with Raman spectroscopic fingerprints for DNA and RNA detection, *Science* **297**, 1536–1540.

Chan, W.C.W., Nie, S.M. (1998) Quantum dot bioconjugates for ultrasensitive nonisotopic detection, *Science* **281**, 2016–2018.

Ciacchi, L.C., Pompe, W., De Vita, A. (2001) Initial nucleation of platinum clusters after reduction of K2PtCl4 in aqueous solution: a first principles study, *J. Am. Chem. Soc.* **123**, 7371–7380.

Coffer, J.L. (1997) Approaches for generating mesoscale patterns of semiconductor nanoclusters, *J. Cluster Sci.* **8**, 159–179.

Coffer, J.L., Bigham, S.R., Pinizzotto, R.F., Yang, H. (1992) Formation of CdS cluster using calf thymus DNA, *Nanotechnology* **3**, 69–76.

Connolly, S., Fitzmaurice, D. (1999) Programmed assembly of gold nanocrystals in aqueous solution, *Adv. Mater.* **11**, 1202–1205.

Douglas, T., Young, M. (1998) Host-guest encapsulation of materials by assembled virus protein cages, *Nature* **393**, 152–155.

Dubertret, B., Calame, M., Libchaber, A.J. (2001) Single-mismatch detection using gold-quenched fluorescent oligonucleotides, *Nat. Biotechnol.* **19**, 365–370.

Dujardin, E., Hsin, L.B., Wang, C.R.C., Mann, S. (2001) DNA-driven self-assembly of gold nanorods, *Chem. Commun.* 1264–1265.

Elghanian, R., Storhoff, J.J., Mucic, R.C., Letsinger, R.L., Mirkin, C.A. (1997) Selective colorimetric detection of polynucleotides based on the distance-dependent optical properties of gold nanoparticles, *Science* **277**, 1078–1081.

Gao, S., Chi, L.F., Lenhert, S., Anczykowsky, B., Niemeyer, C.M., Adler, M., Fuchs, H. (2001) High quality mapping of DNA-protein complex by dynamic scanning force microscopy, *Chem. Phys. Chem.* **2**, 384–388.

Gerion, D., Parak, W.J., Williamson, S.C., Zanchet, D., Micheel, C.M., Alivisatos, A.P. (2002) Sorting fluorescent nanocrystals with DNA, *J. Am. Chem. Soc.* **124**, 7070–7074.

Gurunathan, S., Klinman, D.M., Seder, R.A. (2000) DNA vaccines: immunology, application, and optimization, *Annu. Rev. Immunol.* **18**, 927–974.

Hainfeld, J.F., Furuya, F.R. (1992) A 1.4-nm gold cluster covalently attached to antibodies improves immunolabeling, *J. Histochem. Cytochem.* **40**, 177–184.

Hamad-Schifferli, K., Schwartz, J.J., Santos, A.T., Zhang, S., Jacobson, J.M. (2002) Remote electronic control of DNA hybridization through inductive coupling to an attached metal nanocrystal antenna, *Nature* **415**, 152–155.

He, L., Musick, M.D., Nicewarner, S.R., Salinas, F.G., Benkovic, S.J., Natan, M.J., Keating, C.D. (2000) Colloidal Au-enhanced surface plasmon resonance for ultrasensitive detection of DNA hybridization, *J. Am. Chem. Soc.* **122**, 9071–9077.

Ishii, D., Kinbara, K., Ishida, Y., Ishii, N., Okochi, M., Yohda, M., Aida, T. (2003) Chaperonin-mediated stabilization and ATP-triggered release of semiconductor nanoparticles, *Nature* **423**, 628–632.

Jaiswal, J.K., Mattoussi, H., Mauro, J.M., Simon, S.M. (2003) Long-term multiple color imaging of live cells using quantum dot bioconjugates, *Nat. Biotechnol.* **21**, 47–51.

Jin, R., Wu, G., Li, Z., Mirkin, C.A., Schatz, G.C. (2003) What controls the melting properties of DNA-linked gold nanoparticle assemblies? *J. Am. Chem. Soc.* **125**, 1643–1654.

Jordan, A., Scholz, R., Wust, P., Fähling, H. (1999) Cancer treatment with AC magnetic field induced excitation of biocompatible superparamagnetic nanoparticles, *J. Magn. Magn. Mater.* **201**, 413–419.

Josephson, L., Perez, J.M., Weissleder, R. (2001) Magnetic nanosensors for the detection of oligonucleotide sequences, *Angew. Chem. Int. Ed.* **40**, 3204–3206.

Keren, K., Krueger, M., Gilad, R., Ben-Yoseph, G., Sivan, U., Braun, E. (2002) Sequence-specific molecular lithography on single DNA molecules, *Science* **297**, 72–75.

Kumar, A., Pattarkine, M., Bhadbhade, M., Mandale, A.B., Ganesh, K.N., Datar, S.S., Dharmadhikari, C.V., Sastry, M. (2001) Linear superclusters of colloidal gold particles by electrostatic assembly on DNA templates, *Adv. Mater.* **13**, 341–344.

Lee, S.W., Mao, C., Flynn, C.E., Belcher, A.M. (2002) Ordering of quantum dots using genetically engineered viruses, *Science* **296**, 892–895.

Lin, M.T., Pulkkinen, L., Uitto, J., Yoon, K. (2000) The gene gun: current applications in cutaneous gene therapy, *Int. J. Dermatol.* **39**, 161–170.

Loweth, C.J., Caldwell, W.B., Peng, X., Alivisatos, A.P., Schultz, P.G. (1999) DNA als Gerüst zur Bildung von Aggregaten aus Gold-Nanokristallen, *Angew. Chem. Int. Ed.* **38**, 1808–1812.

Lyon, L.A., Musick, M.D., Natan, M.J. (1998) Colloidal Au-enhanced surface plasmon resonance immunosensing, *Anal. Chem.* **70**, 5177–5183.

Mandeville, J.B., Moore, J., Chesler, D.A., Garrido, L., Weissleder, R., Weisskoff, R.M. (1997) Dynamic liver imaging with iron oxide agents: effects of size and biodistribution on contrast, *Magn. Reson. Med.* **37**, 885–890.

Mbindyo, J.K.N., Reiss, B.D., Martin, B.R., Keating, C.D., Natan, M.J., Mallouk, T.E. (2001) DNA-directed assembly of gold nanowires on complementary surfaces, *Adv. Mater.* **13**, 249–254.

Medintz, I.L., Clapp, A.R., Mattoussi, H., Goldman, E.R., Fisher, B., Mauro, J.M. (2003) Self-assembled nanoscale biosensors based on quantum dot FRET donors, *Nat. Mater.* **2**, 630–638.

Mirkin, C.A., Letsinger, R.L., Mucic, R.C., Storhoff, J.J. (1996) A DNA-based method for rationally assembling nanoparticles into macroscopic materials, *Nature* **382**, 607–609.

Mitchell, G.P., Mirkin, C.A., Letsinger, R.L. (1999) Programmed assembly of DNA functionalized quantum dots, *J. Am. Chem. Soc.* **121**, 8122–8123.

Nam, J.M., Park, S.J., Mirkin, C.A. (2002) Biobarcodes based on oligonucleotide-modified nanoparticles, *J. Am. Chem. Soc.* **124**, 3820–3821.

Niemeyer, C.M., Ceyhan, B. (2001c) DNA-directed functionalization of colloidal gold with proteins, *Angew. Chem. Int. Ed.* **40**, 3685–3688.

Niemeyer, C.M., Bürger, W., Peplies, J. (1998) Covalent DNA-streptavidin conjugates as building blocks for the fabrication of novel biometallic nanostructures, *Angew. Chem. Int. Ed. Engl.* **37**, 2265–2268.

Niemeyer, C.M., Ceyhan, B., Hazarika, P. (2003b) Oligofunctional DNA-gold nanoparticle conjugates, *Angew. Chem. Int. Ed.* **42**, 5766–5770.

Niemeyer, C.M., Wacker, R., Adler, M. (2001d) Hapten-functionalized DNA-streptavidin nanocircles as supramolecular reagents in a novel competitive immuno-PCR assay, *Angew. Chem. Int. Ed.* **40**, 3169–3172.

Niemeyer, C.M., Adler, M., Gao, S., Chi, L.F. (2000) Supramolecular nanocircles consisting of streptavidin and DNA., *Angew. Chem. Int. Ed.* **39**, 3055–3059.

Niemeyer, C.M., Adler, M., Gao, S., Chi, L.F. (2002) Supramolecular DNA-Streptavidin nanocircles with a covalently attached oligonucleotide moiety, *J. Biomol. Struct. Dyn.* **20**, 223–230.

Niemeyer, C.M., Sano, T., Smith, C.L., Cantor, C.R. (1994) Oligonucleotide-directed self-assembly of proteins: Semisynthetic DNA-streptavidin hybrid molecules as connectors for the generation of macroscopic arrays and the construction of supramolecular bioconjugates, *Nucleic Acids Res.* **22**, 5530–5539.

Niemeyer, C.M., Adler, M., Pignataro, B., Lenhert, S., Gao, S., Chi, L.F., Fuchs, H., Blohm, D. (1999) Self-assembly of DNA-streptavidin nanostructures and their use as reagents in immuno-PCR, *Nucleic Acids Res.* **27**, 4553–4561.

Park, S.J., Taton, T.A., Mirkin, C.A. (2002) Array-based electrical detection of DNA with nanoparticle probes, *Science* **295**, 1503–1506.

Park, S-J., Lazarides, A.A., Mirkin, C.A., Letsinger, R.L. (2001) Directed assembly of periodic materials from protein and oligonucleotide-modified nanoparticle building blocks, *Angew. Chem. Int. Ed.* **40**, 2909–2912.

Park, S-J., Lazarides, A.A., Mirkin, C.A., Brazis, P.W., Kannewurf, C.R., Letsinger, R.L. (2000) The electrical properties of gold nanoparticle assemblies linked by DNA, *Angew. Chem. Int. Ed.* **39**, 3845–3848.

Patolsky, F., Weizmann, Y., Lioubashevski, O., Willner, I. (2002) Au-nanoparticle nanowires based on DNA and polylysine templates, *Angew. Chem. Int. Ed.* **41**, 2323–2327.

Perez, J.M., Josephson, L., O'Loughlin, T., Hogemann, D., Weissleder, R. (2002a) Magnetic relaxation switches capable of sensing molecular interactions, *Nat. Biotechnol.* **20**, 816–820.

Perez, J.M., O'Loughin, T., Simeone, F.J., Weissleder, R., Josephson, L. (2002b) DNA-based magnetic nanoparticle assembly acts as a magnetic relaxation nanoswitch allowing screening of DNA-cleaving agents, *J. Am. Chem. Soc.* **124**, 2856–2857.

Pignataro, B., Chi, L.F., Gao, S., Anczykowsky, B., Niemeyer, C.M., Adler, M., Fuchs, H. (2002) Dynamic scanning force microscopy study of self-assembled DNA-protein oligomers, *Appl. Phys. A* **74**, 447–452.

Richter, J., Adler, M., Niemeyer, C.M. (2003) Monte Carlo simulation of the assembly of bis-Biotinylated DNA and streptavidin, *Chem. Phys. Chem.* **4**, 79–83.

Ryadnov, M.G., Ceyhan, B., Niemeyer, C.M., Woolfson, D.N. (2003) "Belt and braces": a peptide-based linker system of de novo design, *J. Am. Chem. Soc.* **125**, 9388–9394.

Safer, D.E., Hainfeld, J., Wall, J.S., Reardon, J.E. (1982) Biospecific labeling with undecagold: visualization of the biotin-binding site on avidin, *Science* **218**, 290–291.

Shenton, W., Davies, S.A., Mann, S. (1999) Directed self-assembly of nanoparticles into macroscopic materials using antibody-antigen recognition, *Adv. Mater.* **11**, 449–452.

Soong, R.K., Bachand, G.D., Neves, H.P., Olkhovets, A.G., Craighead, H.G., Montemagno, C.D. (2000) Powering an inorganic nanodevice with a biomolecular motor, *Science* **290**, 1555–1558.

Souza, G.R., Miller, J.H. (2001) Oligonucleotide detection using angle-dependent light scattering and fractal dimension analysis of gold-DNA aggregates, *J. Am. Chem. Soc.* **123**, 6734–6735.

Storhoff, J.J., Lazarides, A.A., Mucic, R.C., Mirkin, C.A., Letsinger, R.L., Schatz, G.C. (2000) What controls the optical properties of DNA-linked gold nanoparticle assemblies? *J. Am. Chem. Soc.* **122**, 4640–4650.

Taton, T.A., Lu, G., Mirkin, C.A. (2001) Two-color labeling of oligonucleotide arrays via size-selective scattering of nanoparticle probes, *J. Am. Chem. Soc.* **123**, 5164–5165.

Taton, T.A., Mirkin, C.A., Letsinger, R.L. (2000) Scanometric DNA array detection with nanoparticle probes, *Science* **289**, 1757–1760.

Torimoto, T., Yamashita, M., Kuwabata, S., Sakata, T., Mori, H., Yoneyama, H. (1999) Fabrication of CdS nanoparticle chains along DNA double strands, *J. Phys. Chem. B* **103**, 8799–8803.

Tyagi, S., Kramer, F.R. (1996) Molecular beacons: probes that fluoresce upon hybridization, *Nat. Biotechnol.* **14**, 303–308.

Velev, O.D., Kaler, E.W. (1999) In situ assembly of colloidal particles into miniaturized biosensors, *Langmuir* **15**, 3693–3698.

Wang, J., Liu, G., Merkoci, A. (2003) Electrochemical coding technology for simultaneous detection of multiple DNA targets, *J. Am. Chem. Soc.* **125**, 3214–3215.

Wang, J., Xu, D., Polsky, R. (2002a) Magnetically-induced solid-state electrochemical detection of DNA hybridization, *J. Am. Chem. Soc.* **124**, 4208–4209.

Wang, Q., Lin, T., Tang, L., Johnson, J.E., Finn, M.G. (2002b) Icosahedral virus particles as addressable nanoscale building blocks, *Angew. Chem. Int. Ed.* **41**, 459–462.

Weissleder, R., Elizondo, G., Wittenberg, J., Lee, A.S., Josephson, L., Brady, T.J. (1990) Ultrasmall superparamagnetic iron oxide: an intravenous contrast agent for assessing lymph nodes with MR imaging, *Radiology* **175**, 494–498.

Whaley, S.R., English, D.S., Hu, E.L., Barbara, P.F., Belcher, A.M. (2000) Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly, *Nature* **405**, 665–666.

Willner, I., Patolsky, F., Wasserman, J. (2001) Photoelectrochemistry with controlled DNA-cross-linked CdS nanoparticle arrays, *Angew. Chem. Int. Ed.* **40**, 1861–1864.

Wu, X., Liu, H., Liu, J., Haley, K.N., Treadway, J.A., Larson, J.P., Ge, N., Peale, F., Bruchez, M.P. (2003) Immunofluorescent labeling of cancer marker Her2 and other cellular targets with semiconductor quantum dots, *Nat. Biotechnol.* **21**, 41–46.

Xiao, Y., Patolsky, F., Katz, E., Hainfeld, J.F., Willner, I. (2003) "Plugging into Enzymes": nanowiring of redox enzymes by a gold nanoparticle, *Science* **299**, 1877–1881.

Zhao, M., Kircher, M.F., Josephson, L., Weissleder, R. (2002) Differential conjugation of tat peptide to superparamagnetic nanoparticles and its effect on cellular uptake, *Bioconjug. Chem.* **13**, 840–844.

Nanoelectronic Devieces, Proteins for: *see* Optimization of Proteins for Molecular and Biomolecular Electronic Devices

Nanotechnology in Biology: *see* Nanobiotechnology

Nematode (*C. elegans*), Molecular Biology of

Andrew D. Chisholm
University of California, Santa Cruz, California

1 Overview of *C. elegans* Biology 54
1.1 *C. elegans* as a Biological Model 54
1.2 Taxonomy, Evolution, and Ecology 55
1.3 Sexual System, Life Cycle, and Developmental Stages 56
1.4 Genetics and Nomenclature 56

2 *C. elegans* Genomics 57
2.1 Genomes and Informatics 57
2.2 Protein-coding Genes and the Proteome 58
2.3 The Noncoding RNA World 58
2.4 Whole-genome Analyses of Gene Expression 59
2.5 Transposons and Repetitive Elements 59

3 *C. elegans* Chromosome Structure 60
3.1 The Chromosomes 60
3.2 Transgenes and Extrachromosomal Arrays 60
3.3 Meiosis and Recombination 61

4 The Nucleus and Gene Expression 61
4.1 Chromatin and Transcription 61
4.2 Splicing and mRNA Processing 62
4.3 Translation 62
4.4 mRNA Surveillance and Nonsense-mediated Decay 63
4.5 RNA Interference 63

Bibliography 64
Books and Reviews 64
Primary Literature 65

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

Keywords

Holocentric, Holokinetic
A chromosome in which kinetochore forms along the length of the chromosome.

Systemic RNAi
Non-cell-autonomous action of an interfering RNA.

The nematode worm *Caenorhabditis elegans* is a major model organism for studies in genetics, cell biology, neurobiology, and development. The use of *C. elegans* as a model organism began in the 1950s. The cellular basis of the development and neuroanatomy of *C. elegans* have been described in detail. Studies of *C. elegans* have elucidated the genetic basis of apoptotic cell death, axon guidance, aging, and RNA interference. The *C. elegans* genome is one of the smallest known for any animal and was the first animal genome to be completely sequenced. The number of genes in the worm genome is comparable to that of other larger animals, suggesting that although *C. elegans* is small, it has comparable biological complexity to other animals. The functions of most *C. elegans* genes are not known. *C. elegans* is well suited to both forward and reverse genetic manipulations of the genome. Postgenomic studies are continuing to explore the biological complexity of *C. elegans*.

1 Overview of *C. elegans* Biology

1.1 *C. elegans* as a Biological Model

For over forty years, the nematode *Caenorhabditis elegans* (*C. elegans*) has been used as a model for studies of animal development and behavior. Systematic genetic and anatomical analysis of *C. elegans* was begun by Brenner and his group in 1965. Currently, several hundred laboratories around the world use *C. elegans* for studies into the basic mechanisms of animal development, physiology, cell biology, and behavior.

Many advantages of *C. elegans* as a biological model result from its small size and cell number. Adult hermaphrodite animals are 1-mm long, transparent, and contain 959 somatic cells. The pattern of cell divisions, or cell lineage, is largely invariant between individuals. This has allowed the complete *C. elegans* cell lineage (zygote to adult) to be reconstructed directly from observations of development in living animals. The anatomy of *C. elegans*, including the synaptic connectivity of the nervous system, has been completely reconstructed from electron micrographs of serial sections. *C. elegans* remains the only animal for which the development and ultrastructure have been so completely described.

The reproductive biology of *C. elegans* is highly suited for genetic analysis. *C. elegans* reproduces sexually as a self-fertilizing hermaphrodite, allowing populations to be maintained without crossing. However, crossing is essential for genetic manipulations, and fortunately *C. elegans* males can cross-fertilize hermaphrodites. The generation time is short (3 days), and laboratory culture requirements are relatively simple. *C. elegans* stocks can be preserved cryogenically.

The genome of *C. elegans* is one of the smallest known for an animal (100 million base pairs, Mbp) and it was the first animal genome to be fully sequenced, in 1998. The *C. elegans* genome is estimated to contain over 21 000 genes, 40% of which have recognizable homologs in humans. The functions of most *C. elegans* genes are not known.

Studies of *C. elegans* have yielded insight into many areas of biology. The pioneering analysis of *C. elegans* development and programmed cell death by Brenner, Sulston, and Horvitz, was recognized by the award of the 2002 Nobel Prize in Physiology or Medicine (see Nobel lectures by Horvitz and Sulston). *C. elegans* studies have also played key roles in advancing our understanding of neuronal axon guidance, of the genetic control of aging, and in the development of RNA interference.

Detailed information on the biology of *C. elegans* is available in two monographs: *The Nematode Caenorhabditis elegans*, edited by W.B. Wood, and *C. elegans II*, edited by D. Riddle and others. These books should be consulted for a more detailed treatment of the topics in this article. More current descriptions of *C. elegans* biology are in preparation as an online supplement to Wormbase. Two books of methods related to *C. elegans* work have been published (see books edited by Hope, and by Shakes and Epstein).

1.2 Taxonomy, Evolution, and Ecology

C. elegans is a member of the phylum *Nematoda* (roundworms). Nematodes comprise one of the five most successful animal phyla in terms of number of organisms and number of species. The nematode body plan is highly conserved and *C. elegans* is representative of the group. *C. elegans* is a member of the order Rhabditida, composed mostly of nonparasitic terrestrial species. Phylogenetic analysis suggests that *C. elegans* is closely related to parasitic nematodes of the order Strongylida, and thus *C. elegans* studies may have particular relevance to these nematodes.

Several other species within the genus *Caenorhabditis* are the subject of active investigation. *Caenorhabditis briggsae* (*C. briggsae)* is morphologically hard to distinguish from *C. elegans*; in the older literature *briggsae* and *elegans* were often mistaken for one another. Remarkably, despite their similar morphology the *C. briggsae* and *C. elegans* genomes are quite divergent; the extent of divergence has been described as being comparable to that between the human and mouse genomes. Because there is almost no fossil record for nematodes, and because rates of gene divergence appear to differ between animal groups, the time of divergence of the two species is hard to estimate.

The genomes of three additional *Caenorhabditis* species are being sequenced (*C. remanei, C. japonica,* and *Caenorhabditis* sp. CB5161). The comparative genomic analysis of these closely related species is expected to be a rich resource for understanding the structure and function of the genome. Genomic or

EST sequencing projects are in progress for several parasitic nematodes.

C. elegans is typically found in soil, especially in mushroom compost. *C. elegans* is a cosmopolitan species; isolates have been found throughout the temperate northern hemisphere. The ''natural'' food of *C. elegans* is not known, but may include soil bacteria, yeast, and other microorganisms such as cellular slime molds. Different wild isolates display frequent sequence polymorphisms among themselves and relative to the reference wild type N2. As a result of genome sequence information, mutations isolated in the N2 background can usually be rapidly mapped with respect to single-nucleotide polymorphisms (SNPs) between N2 and wild isolates such as the Hawaiian strain CB4856.

1.3 Sexual System, Life Cycle, and Developmental Stages

C. elegans reproduces sexually. There are two sexes, a self-fertilizing hermaphrodite and a cross-fertilizing male. (A sexual system involving males and hermaphrodites is known as *androdioecy*.) Hermaphrodites make sperm during larval development and later switch to making oocytes as adults. Hermaphrodites make about 300 sperm, so the self-progeny brood size of a single hermaphrodite is about 300. Males make ~2000 sperm and so can sire many more cross progeny.

Although androdioecy is also found in other nematode species, including *C. briggsae*, the more common and presumably ancestral sexual system is gonochoristic (male–female), as found in *C. remanei* and all other members of the *Caenorhabditis* genus. Hermaphrodites may be thought of as modified females: they have a female soma and have developed a way to modulate the sex of their gametes, allowing self fertilization. Single-gene mutations are known, that prevent formation of sperm in hermaphrodites, essentially returning them to an obligate male–female system; *C. elegans* can be converted to a variety of sex-determination systems by genetic manipulations.

Fertilization occurs in the hermaphrodite uterus. Fertilized eggs are laid after about 2 hours of development; embryogenesis is complete about 11 hours later at 20 °C. The egg hatches into the first larval stage (L1). Larval development takes 2 days, and consists of the four larval stages L1 to L4. Between each larval stage the old cuticle is shed (molting) and a new cuticle is formed. The L4 molts into an adult, which begins to lay eggs a few hours afterwards. From egg to egg at 20 °C takes about 3.5 days.

The sequence of cell divisions in *C. elegans* development is largely invariant from individual to individual. This, together with their small size and transparency, allowed the entire cell lineage (zygote to adult) to be reconstructed from observations of living animals. Each cell in an animal can thus be named both according to its lineal ancestry and according to its differentiated fate. A comprehensive online guide to *C. elegans* anatomy is at www.wormatlas.org.

1.4 Genetics and Nomenclature

Modern *C. elegans* genetic analysis began with Brenner's choice of the Bristol strain of *C. elegans* as his reference wild type, which he designated N2. Most subsequent genetic work has been in the N2 strain background or its derivatives. Genetic drift from the original stock of N2 has been minimized by cryogenic

preservation of stocks. A public collection of wild type and mutant stocks is maintained by the (CGC) *Caenorhabditis* Genetics Center (http://biosci.umn.edu/CGC/CGChomepage.htm).

All *C. elegans* genes are given a "gene model" name by Wormbase. For example, cosmid clone K03D10 contains multiple genes, either experimentally confirmed or only predicted, and numbered as K03D10.1, K03D10.2, and so on. Different splice variants are indicated by subscripts, for example, K03D10.1a, K03D10.1b.

Subsets of genes have been assigned official "three-letter" names by the *Caenorhabditis* Genetics Center. Three-letter names may indicate the mutant phenotype of the gene or the nature of the predicted gene product. Genes that were first identified by their mutant phenotype are named after it, for example, "dumpy" (*dpy*), "blistered" (*bli*), uncoordinated (*unc*), and variably abnormal (*vab*). Many genes mutate to similar (but not identical) phenotypes, and so are put in the same general class with different numbers (*dpy-1, dpy-2,* etc.). Alternatively, some genes are identified by their sequence similarity to a gene in another organism. For example, *kal-1* is similar in sequence to the gene mutated in the X-linked form of Kallmann syndrome.

Mutations have unique names consisting of the laboratory code and a number. For example, the first mutant to be found in Brenner's genetic screens (in October 1967) had a dumpy phenotype. The name of the mutation causing the mutant phenotype is *e1*. The name of the gene affected by this mutation is *dpy-1*. Thus, the full name of the mutation is *dpy-1(e1)*. The *e* prefix indicates the laboratory in which the mutation was isolated. The wild-type (N2) version of a gene is written as *dpy-1(+)*.

All *C. elegans* strain names have a two-letter or three-letter lab code (e.g. MT for the Horvitz laboratory at MIT) and a number. Each strain has a unique name to avoid confusion. The genotype is the complete listing of all the mutations in the strain. For example, a very useful strain has the name MT465 and bears mutations on chromosomes I, II, and III. The full genotype of MT465 is *dpy-5(e61) I*; *bli-2(e768) II*; *unc-32(e189) III*. The mutations *e61, e768,* and *e189* cause dumpy, blistered, and uncoordinated phenotypes respectively.

C. elegans nomenclature distinguishes the gene name, the mutant phenotype, and the protein (or RNA) product. Genes and mutations are written in lower-case italics (*kal-1, vab-1*), while phenotypes are in Roman typeface with initial letters capitalized (e.g. Vab, Unc). Protein products are written in upper-case (e.g. KAL-1, PTP-3A, PTP-3B).

A more detailed account of *C. elegans* nomenclature is available online at the CGC web site.

2 *C. elegans* Genomics

2.1 Genomes and Informatics

The *C. elegans* genome was the first animal genome to be sequenced fully. An overview of *C. elegans* genomic features was published in 1998 when most of the sequence had been completed. The last sequence gaps of the genome were completed in late 2002. However, analysis and annotation of the sequence is a continuing process that will occupy researchers for many years. The genomic sequence

information and annotations are available online. Wormbase (www.wormbase.org) is a public repository of genomic, genetic, and anatomical information. The genomic sequence can also be searched at the University of California Santa Cruz genome browser (genome.ucsc.edu) and at several other online sites. The completion of a draft *C. briggsae* genome sequence in 2003 allows a global comparison with the now-finished *C. elegans* genome, on which the discussion below is based.

2.2 Protein-coding Genes and the Proteome

The *C. elegans* genome is estimated to contain 19 775 protein-coding genes, of which 2376 display alternative splicing, resulting in a predicted proteome of 22 151 proteins (Wormpep release 127, July 2004). 75% of these genes are supported by partial or complete experimental (EST) data; the remainder are predicted but not yet confirmed. Systematic projects are underway to generate a complete set of functional open reading frames (the ORFeome) and to test all known *C. elegans* proteins in two-hybrid interaction screens. One notable feature of the *C. elegans* proteome is the large number (>700) of seven-pass transmembrane receptors, many of which are thought to function as olfactory receptors. This gene family has expanded almost twofold in *C. elegans* relative to *C. briggsae*, for reasons unknown.

The distribution of genes within the worm genome exhibits several nonrandom features on various scales. In addition to being organized into operons (see below), genes of the same family are often locally clustered, probably as a result of tandem duplications in evolution. Where genes of different families are coexpressed in the same tissue, they also show local clustering into groups of 2 to 5 genes. Genes involved in reproduction show longer-range clustering. Interestingly, the latter clusters are not conserved in *C. briggsae*, suggesting that they have recently evolved. Finally, more conserved genes with essential functions tend to be found in the centers of chromosomes, whereas more diverged genes are enriched in chromosomal arms. Such chromosome-wide patterns of gene distribution have not so far been observed in other organisms.

2.3 The Noncoding RNA World

The *C. elegans* genome contains approximately 890 noncoding (nc) RNA genes, including approximately 600 tRNAs. Ribosomal RNAs are generated from two tandem arrays: an array of ~100 to 150 copies of the 18S, 5.6S, and 26S genes on chromosome I, and an array of 100 copies of the 5S repeat (which includes the SL1 trans-spliced leader) on chromosome V. Most other conserved ncRNAs are present in the *C. elegans* genome, including snRNAs U1, U2, U5, and SRP. Some ncRNAs have not yet been found; these may truly be absent (e.g. U12 snRNAs) or may be hard to detect computationally (e.g. telomerase RNAs). Mutations in U1 snRNAs have recently been identified as informational suppressors of mutations affecting the conserved +1G position of introns.

The *C. elegans* genome contains numerous small noncoding RNAs, generally known as microRNAs or miRNAs. The first two animal miRNAs, *lin-4* and *let-7*, were discovered through genetic analysis of developmental control genes known as heterochronic genes. *lin-4* and *let-7* are short (22 nucleotide) RNAs processed from longer hairpin precursors by the endonuclease DCR-1/Dicer. Both RNAs form

duplex structures by complementary base pairing with target sequences in the 3′ UTRs of other transcripts. Despite their short sequence, *let-7* orthologs have been found in many other animals, suggesting that *let-7* function is conserved in animal evolution.

From computational and experimental data, the total number of miRNA genes in the genome has been estimated to be approximately 100 to 120. Almost all *C. elegans* miRNAs are conserved in *C. briggsae*; about 30% appear to be conserved in insects or vertebrates. An additional class of "tiny noncoding RNAs" consists of short RNAs that do not appear to be processed from a longer precursor. Experimental searches for such short RNAs have also revealed antisense RNAs from hundreds of genes, suggesting that gene regulation by antisense RNA may be widespread.

2.4 Whole-genome Analyses of Gene Expression

Several approaches have been used to analyze global patterns of gene expression in *C. elegans*. Systematic sequencing of cDNAs has generated a database of expressed sequence tags (ESTs). ESTs, as well as experimentally defining gene structures, have been used in genome-wide *in situ* hybridization screens of expression patterns (available online at nematode.lab.nig.ac.jp).

Whole-genome DNA microarrays have been used to profile global gene expression patterns across developmental stages, tissue types, physiological conditions, and in different mutant backgrounds. Such gene profiling approaches have provided critical insights into the transcriptional programs underlying the development of specific tissues, such as the germline, of organs such as the pharynx, and of metabolic processes. Data from several hundred such microarray experiments have been combined by clustering algorithms to develop a "gene expression topological map" that represents the correlation in gene expression differences between different conditions. Clusters on this gene expression map can represent either tissues (e.g. neurally expressed genes) or biochemical processes.

2.5 Transposons and Repetitive Elements

Approximately 16% of the *C. elegans* genome consists of repetitive elements. The largest repeat family consists of ~3000 copies of a 439-bp element, forming 1.3% of the genome. The slightly larger size of the *C. briggsae* genome (104 Mbp) is solely due to expansion of certain repeat families.

C. elegans contains at least six families of active transposons, designated Tc1 through Tc5, and Tc7. Tc1 and Tc3 are members of the *mariner* subfamily and their mechanisms of transposition have been studied in detail. In the reference wild-type Bristol N2, there are 32 copies of Tc1. Although most of its Tc1 elements are intact, the N2 strain actively represses transposition in the germline by an RNA interference-related process. Other wild isolates such as the so-called *Bergerac* strains contain hundreds of Tc1 elements and are active for transposition. Strains in which transposition is elevated allow transposon-mediated mutagenesis both in forward and reverse genetic approaches. Nonnative elements such as the insect transposon *Mos1* can transpose in *C. elegans* and have been successfully used in transposon-tagging mutageneses.

3
C. elegans Chromosome Structure

3.1
The Chromosomes

C. elegans has five autosomes (I-V) and a sex chromosome (X). Hermaphrodites have two X chromosomes and males have one; the wild-type karyotype is therefore 12 for hermaphrodites (XX) and 11 for males (XO). Cytologically, *C. elegans* chromosomes appear as uniform rods a few microns in length and cannot be distinguished without the use of sequence-specific probes. As discussed below, *C. elegans* chromosomes have delocalized centromeres; no specific DNA sequences have yet been shown to be required for *C. elegans* centromere activity.

C. elegans telomeres consist of several kilobases of tandem repeats of the sequence TTAGGC, similar to those of other organisms. These repeats are probably sufficient for telomeric function, as the subtelomeric sequences appear to be unique. The telomerase enzyme has not been defined in *C. elegans*; some proteins have been shown to pay roles in telomere maintenance or replication.

Unlike mammalian chromosomes, *C. elegans* chromosomes lack a defined centromere. Instead, a kinetochore assembles uniformly along the entire length of the chromosome during mitosis. Such chromosomes are termed *holocentric* or *holokinetic*. Holocentric chromosomes appear to be characteristic of nematodes, and are also found in some insects and plants. The holocentric nature of *C. elegans* chromosomes probably underlies the ability of *C. elegans* cells to segregate exogenous extrachromosomal DNA with some fidelity. Despite these differences, the mitotic kinetochore resembles that of other eukaryotes in ultrastructure. Several components of the *C. elegans* kinetochore have been identified.

3.2
Transgenes and Extrachromosomal Arrays

C. elegans cells can propagate extrachromosomal "arrays" of exogenous DNA (notated as *Ex*, e.g. *juEx600*). Such arrays are the predominant means of introducing transgenes into *C. elegans*. Arrays are generated by injecting DNA into the syncytial or cellular gonad of a parental hermaphrodite. Arrays form in a semi-random process in the germ cells that give rise to the progeny of the injected animal; a single parent can give rise to multiple transgenic progeny, each with a different version of the transgenic DNA. Such arrays are replicated and transmitted with variable fidelity in mitosis and meiosis. Although these observations suggest that there is no specific sequence requirement for a centromere, the low fidelity of array transmission could reflect a subtle requirement for specific sequences for full centromeric function. Alternatively, arrays may be transmitted poorly because they are much smaller than normal chromosomes.

Little is known of the structure of *C. elegans* extrachromosomal arrays – it is not known whether arrays are linear or circular DNAs. Southern blotting experiments suggest that most stable arrays contain several hundred copies of the injected DNA arranged in tandem repeats. However, many copies may be nonfunctional or otherwise rearranged, so the effective copy number of a transgene is likely to be lower. Extrachromosomal arrays are usually integrated into the genome by a random chemical or radiation mutagenesis to induce double-strand breaks, to form integrated

arrays (denoted *Is*, e.g. *juIs1*). An alternative transformation method that can yield low-copy or single-copy transgenes is to introduce DNA by ballistic transformation; such DNA is integrated nonhomologously into a chromosome. Although there have been reports of homologous recombination in *C. elegans*, such events are at present too inefficient for routine use in gene targeting or replacement.

3.3 Meiosis and Recombination

Cytologically, *C. elegans* meiosis resembles that of other organisms. The stages of meiosis correlate with the development of germline nuclei in the gonad and can be readily distinguished by cytology of gonadal preparations. *C. elegans* chromosomes display several unusual features in their genetic organization that relate to the control of recombination in meiosis. First, despite wide variation in physical size (13.7 Mbp to 20.9 Mbp), all chromosomes are of the same genetic length (~50 centiMorgans). This reflects a chromosome-wide interference process that ensures that each bivalent undergoes, on average, one and only one crossover event. Second, recombination rates are not uniform over the chromosomes. Ninety percent of crossovers occur in the terminal one-third portion of chromosomes: thus, genes in the center of each chromosome appear genetically "clustered" whereas genes on the flanking arms appear dispersed. The unequal distribution of crossovers reflects both cis-acting sequences (e.g. recombination occurs preferentially in introns) and trans-acting factors (e.g. *rec-1*). Third, specific chromosomal regions are required for homolog pairing and, consequently, for crossovers. Each chromosome has one end that acts as its "pairing center." The existence of these regions explains why *C. elegans* reciprocal translocation chromosomes can act as crossover suppressors. Instead of forming a more typical tetravalent structure that maximizes pairing using sequence identity, *C. elegans* translocation heterozygotes probably form two separate bivalents using the two pairing centers.

4 The Nucleus and Gene Expression

4.1 Chromatin and Transcription

Eukaryotic chromatin is organized into states that allow or prevent active transcription. These states can be stably inherited through cell divisions, or can be "erased" by chromatin-remodeling complexes. The organization of *C. elegans* chromatin plays a key role in the developmental distinction between somatic and germline tissue. The X chromosome is modified selectively, lacking the modified histones associated with active transcription, and being enriched in histones associated with silenced chromatin. The X chromosome is generally depleted of genes expressed in the germline. Proteins involved in the germline-specific repression of X-linked genes are also required for somatic repression of developmental control genes, indicating a common mechanism in the control of these chromatin states.

Specification of the silenced state of germline chromatin is an essential and early part of germline development. Two factors are known to play critical roles in germline specification: the zinc finger protein PIE-1 and the perinuclear organelles

known as P granules. PIE-1 appears to function to protect the early germline from the activity of a chromatin-remodeling complex related to the vertebrate NuRD complex. This NuRD-like complex is active in somatic cells and is required for determination of somatic cell fates; loss of NuRD function transforms somatic cells into germline-like cells.

The unequal dosage of X-linked genes between XX and XO animals is compensated for by the twofold repression of X-linked gene expression in XX hermaphrodites. The protein complex responsible for this chromosome-specific repression has been characterized and is related to protein complexes required for chromosome condensation during mitosis and meiosis.

4.2 Splicing and mRNA Processing

Most aspects of mRNA processing in *C. elegans* resemble those of other eukaryotes, with some significant differences. First, *C. elegans* introns are unusually small; several confirmed introns are less than 40 bp in length. *C. elegans* introns also lack a defined branch site or polypyrimidine tract. These differences suggest that *C. elegans* spliceosome assembly may differ from that in yeast or vertebrates.

Second, in addition to conventional (cis) splicing, *C. elegans* transcripts frequently undergo trans-splicing, in which a ''spliced leader'' RNA is spliced to the 5′ end of an mRNAs. The trans-splicing reaction resembles that of *cis*-splicing. The most common leader sequence is the 22 nt SL1 spliced leader, which is donated from a 100 nt precursor, itself transcribed from the 5S ribosomal RNA tandem repeats. The SL1 leader provides the trimethylguanosine cap. SL1 function is essential for viability. SL1 addition is often very close to the initiator methionine codon, so it may have some role in translation.

The third and most unusual aspect of *C. elegans* mRNA processing is that about 15% of genes are transcribed in polycistronic messages from clusters of two to eight adjacent genes known as operons. Transcription of an operon begins at the promoter of the 5′-most gene. More 3′ genes are transcribed into a polycistronic message that is processed by conventional 3′-end formation and trans-splicing to make monocistronic mRNAs. Notably, 3′ messages are trans-spliced to a second class of spliced leader, known as SL2. SL2 leaders are made by a dispersed family of 18 genes and, unlike SL1, show some sequence variation.

Most operons are highly conserved between *C. elegans* and *C. briggsae*. Operons may be present throughout the nematode phylum, and may be characteristic of animals that exhibit trans-splicing. Polycistronic organization presumably facilitates coordinate expression of the genes in an operon. Some operons consist of genes with biochemically related functions, so knowledge of the function of one gene in an operon can sometimes give clues about the functions of others. However, for most operons the functional relationship of the genes is not so clear-cut. Relative to the genome as a whole, operons are enriched in essential genes with functions in gene expression, RNA metabolism, or mitochondrial physiology.

4.3 Translation

Analysis of translation in *C. elegans* has focused largely on translational control of

gene expression. The cytoplasmic translational apparatus appears typical of eukaryotes; however, components of the mitochondrial translational apparatus are the smallest known for eukaryotes. For example, *C. elegans* mitochondrial tRNAs lack some of the normal structural features of tRNAs. The *C. elegans* nuclear genetic code is standard. *C. elegans* encodes a selenocysteine tRNA that recognizes UGA codons in certain contexts. *C. elegans* genes exhibit limited synonymous codon bias that correlates with expression levels. Codon bias is less pronounced in longer genes, for unknown reasons. *C. elegans* is the only animal in which amber suppressor mutations have been found, by using forward genetic analysis. Over 10 genes have been identified genetically as suppressors of amber (UAG) termination codons; most have been shown to encode $tRNA^{Trp}$ genes. Amber suppressor mutants are all at least semiviable, probably reflecting redundancy among the family of 12 $tRNA^{Trp}$ genes.

Several genes are known to be regulated at the level of mRNA translation. In all known cases, cis-acting sites in the 3′ UTR are required for proper regulation of translation. In the early (4-cell) embryo, the *glp-1* mRNA is ubiquitously expressed, yet is translated only in anterior blastomeres. During larval development, translation of the *lin-14* mRNA is downregulated by binding of the noncoding RNA *lin-4* to a repeated sequence in the *lin-14* 3′ UTR. The *lin-4* RNA is partly complementary to the *lin-14* repeats and forms duplex structures interrupted by bulges and loops. Formation of such hybrid structures prevents *lin-14* translation, by mechanisms that are still not well defined. The imperfect complementarity of *lin-4* and *lin-14* may be the reason why such dsRNAs are not subjected to degradation (RNA interference).

4.4 mRNA Surveillance and Nonsense-mediated Decay

All eukaryotic cells possess a system that degrades mRNAs with premature termination codons. This process, known as nonsense mediated decay (NMD) was identified in *C. elegans* by classical genetic analysis of a class of suppressor mutations known as *smg* mutations. Mutations in seven *smg* genes suppress the effects of premature stop codon alleles in a variety of genes. Such premature stop codons are distinguished from normal stop codons by their position relative to the last exon–exon junction, and subjected to rapid degradation. NMD may occur in the nucleus but is thought to happen mostly in the cytoplasm in concert with translation.

4.5 RNA Interference

RNA interference (RNAi) is a homology-based process by which endogenous mRNAs can be degraded by introduction of homologous double-stranded RNAs (dsRNAs). The importance of dsRNAs as triggers for RNAi was first revealed by work in *C. elegans*. Since its discovery in 1998, RNAi has become an essential part of the *C. elegans* repertoire of methods and has also yielded many unexpected insights into the mechanisms of gene expression and RNA metabolism. Double-stranded RNAs can be introduced into the animal by a variety of means: injection, soaking, feeding, or by transgenes; each method has its own advantages and disadvantages. Under optimal conditions RNAi can reliably

phenocopy the null phenotype of a gene; under nonoptimal conditions RNAi may be partly or completely ineffective in reducing gene function. Certain life stages appear to be more sensitive to RNAi than others. Some tissues, such as the nervous system, are refractory to RNAi; this limitation is being overcome by the use of mutant backgrounds that are more or less generally sensitized to RNAi. Nevertheless, at present, the sensitivity of any one gene to RNAi must be tested empirically and verified independently.

The general mechanism of RNAi is under intensive investigation in a variety of species and will not be summarized here. The *C. elegans* RNAi system is unusual in two key aspects. First, Bristol N2 *C. elegans* can ingest dsRNAs from its bacterial food ("feeding RNAi"). The mechanism of dsRNA uptake from food is not yet known. This highly convenient feature is polymorphic in *C. elegans*: the Hawaiian isolate CB4856 is sensitive to RNAi by injection, but not by feeding. It is not yet certain whether other nematode species, or other animals, are competent to take up dsRNA from food. Feeding RNAi, being highly scalable, has been critical in allowing genome-wide screens for gene function. Second, RNAi in *C. elegans* is often systemic: dsRNAs generated in one tissue (e.g. the intestine) can inhibit mRNAs in a different tissue (e.g. the germline), thereby implying the existence of a system for transferring dsRNA between cells. It is not known if all tissues are equally systemic, or whether certain tissues are more proficient in export or import of dsRNAs. Other animals, such as *Drosophila*, do not display systemic RNAi; the effects of RNAi in these animals are thus cell-autonomous. Genetic screens have begun to reveal the molecular requirements for systemic RNAi, which include the transporter-like protein SID-1.

One of the normal functions of the RNA interference mechanism is to provide a surveillance system that controls the level of transposition of endogenous transposons. The first clue to the normal role of RNAi came from the serendipitous discovery that mutants with elevated levels of transposition were also RNAi-defective. Double-stranded RNAs corresponding to transposon sequences could be generated by read-through of transposons from adjacent genes, in both sense and antisense directions. Some components of the RNAi mechanism have a dual function in the production of miRNAs such as *lin-4* and *let-7*, which are cleaved from precursor messages.

See also Nematodes, Neurobiology and Development of.

Bibliography

Books and Reviews

Blaxter, M. (1998) *Caenorhabditis elegans* is a nematode, *Science* **282**, 2041–2046.

Epstein, H.F., Shakes, D.C. (Eds.) (1995) *C. elegans: Modern Biological Analysis of an Organism*, Methods in Cell Biology Series, Academic Press, New York, NY, USA.

Grant, B.D., Wilkinson, H.A. (2003) Functional genomic maps in *Caenorhabditis elegans*, *Curr. Opin. Cell Biol.* **15**, 206–212.

Gupta, B.P., Sternberg, P.W. (2003) The draft genome sequence of the nematode *Caenorhabditis briggsae*, a companion to *C. elegans*, *Genome Biol.* **4**, 238.

Hope, I. (Ed.) (1999) *C. elegans: A practical approach*, Oxford University Press, Oxford, New York.

Horvitz, H.R. (2003) Worms, life, and death (Nobel lecture), *Chembiochem* **4**, 697–711.

Jorgensen, E.M., Mango, S.E. (2002) The art and design of genetic screens: *Caenorhabditis elegans*, *Nat. Rev. Genet.* **3**, 356–369.

Riddle, D.L., Blumenthal, T, Meyer, B.J., Priess, J.R. (Eds.) (1997) *C. elegans II*, Cold Spring Harbor Monograph Series 33, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.

Sulston, J.E. (2003) *Caenorhabditis elegans*: the cell lineage and beyond (Nobel lecture), *Chembiochem* **4**, 688–696.

Wood, W.B. (Ed) (1988) *The Nematode Caenorhabditis elegans*, Cold Spring Harbor Monograph Series *17*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.

Primary Literature

Ahmed, S., Hodgkin, J. (2000) MRT-2 checkpoint protein is required for germline immortality and telomere replication in C. elegans, *Nature* **403**, 159–164.

Ambros, V. (2001) microRNAs: tiny regulators with great potential, *Cell* **107**, 823–826.

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., Jewell, D. (2003) MicroRNAs and other tiny endogenous RNAs in C. elegans, *Curr. Biol.* **13**, 807–818.

Anderson, P. (1995) Mutagenesis, *Methods. Cell Biol.* **48**, 31–58.

Bargmann, C., Hodgkin, J. (2002) Accolade for elegans, *Cell* **111**, 759–762.

Barnes, T.M., Kohara, Y., Coulson, A., Hekimi, S. (1995) Meiotic recombination, noncoding DNA and genomic organization in Caenorhabditis elegans, *Genetics* **141**, 159–179.

Berezikov, E., Bargmann, C.I., Plasterk, R.H. (2004) Homologous gene targeting in Caenorhabditis elegans by biolistic transformation, *Nucleic Acids Res.* **32**, e40.

Bessereau, J.L., Wright, A., Williams, D.C., Schuske, K., Davis, M.W., Jorgensen, E.M. (2001) Mobilization of a Drosophila transposon in the Caenorhabditis elegans germ line, *Nature* **413**, 70–74.

Blaxter, M. (1998) Caenorhabditis elegans is a nematode, *Science* **282**, 2041–2046.

Blaxter, M.L., De Ley, P., Garey, J.R., Liu, L.X., Scheldeman, P., Vierstraete, A., Vanfleteren, J.R., Mackey, L.Y., Dorris, M., Frisse, L.M., Vida, J.T., Thomas, W.K. (1998) A molecular evolutionary framework for the phylum Nematoda, *Nature* **392**, 71–75.

Blumenthal, T., Gleason, K.S. (2003) Caenorhabditis elegans operons: form and function, *Nat. Rev. Genet.* **4**, 112–120.

Blumenthal, T., Evans, D., Link, C.D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W.L., Duke, K., Kiraly, M., Kim, S.K. (2002) A global analysis of Caenorhabditis elegans operons, *Nature* **417**, 851–854.

Brenner, S. (1974) The genetics of Caenorhabditis elegans, *Genetics* **77**, 71–94.

The C. elegans Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology, *Science* **282**, 2012–2018.

Dernburg, A.F. (2001) Here, there, and everywhere: kinetochore function on holocentric chromosomes, *J. Cell Biol.* **153**, F33–F38.

Eichmuller, S., Vezzoli, V., Bazzini, C., Ritter, M., Furst, J., Jakab, M., Ravasio, A., Chwatal, S., Dossena, S., Botta, G., Meyer, G., Maier, B., Valenti, G., Lang, F., Paulmichl, M. (2004) A new gene-finding tool: using the Caenorhabditis elegans operons for identifying functional partner proteins in human cells, *J. Biol. Chem.* **279**, 7136–7146.

Feinberg, E.H., Hunter, C.P. (2003) Transport of dsRNA into cells by the transmembrane protein SID-1, *Science* **301**, 1545–1547.

Finch, C.E., Ruvkun, G. (2001) The genetics of aging, *Annu. Rev. Genomics Hum. Genet.* **2**, 435–462.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans, *Nature* **391**, 806–811.

Fischer, S.E., Wienholds, E., Plasterk, R.H. (2003) Continuous exchange of sequence information between dispersed Tc1 transposons in the Caenorhabditis elegans genome, *Genetics* **164**, 127–134.

Ganko, E.W., Fielman, K.T., McDonald, J.F. (2001) Evolutionary history of Cer elements and their impact on the C. elegans genome, *Genome Res.* **11**, 2066–2074.

Gaudet, J., Mango, S.E. (2002) Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4, *Science* **295**, 821–825.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A.,

Ruvkun, G., Mello, C.C. (2001) Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing, *Cell* **106**, 23–34.

Hannon, G.J. (Ed.) (2003) *RNAi: A Guide to Gene Silencing*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Harris, T.W., Chen, N., Cunningham, F., Tello-Ruiz, M., Antoshechkin, I., Bastiani, C., Bieri, T., Blasiar, D., Bradnam, K., Chan, J., Chen, C.K., Chen, W.J., Davis, P., Kenny, E., Kishore, R., Lawson, D., Lee, R., Muller, H.M., Nakamura, C., Ozersky, P., Petcherski, A., Rogers, A., Sabo, A., Schwarz, E.M., Van Auken, K., Wang, Q., Durbin, R., Spieth, J., Sternberg, P.W., Stein, L.D. (2004) WormBase: A multi-species resource for nematode biology and genomics, *Nucleic Acids Res* **32**(Database issue), D411–D417.

Hillers, K.J., Villeneuve, A.M. (2003) Chromosome-wide control of meiotic crossing over in C. elegans, *Curr. Biol.* **13**, 1641–1647.

Hodgkin, J. (2002) Exploring the envelope. Systematic alteration in the sex-determination system of the nematode Caenorhabditis elegans, *Genetics* **162**, 767–780.

Hodgkin, J., Papp, A., Pulak, R., Ambros, V., Anderson, P. (1989) A new kind of informational suppression in the nematode Caenorhabditis elegans, *Genetics* **123**, 301–313.

Hubbard, E.J., Greenstein, D. (2000) The Caenorhabditis elegans gonad: a test tube for cell and developmental biology, *Dev. Dyn.* **218**, 2–22.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D.P., Zipperlen, P., Ahringer, J. (2003) Systematic functional analysis of the Caenorhabditis elegans genome using RNAi, *Nature* **421**, 231–237.

Kelly, W.G., Schaner, C.E., Dernburg, A.F., Lee, M.H., Kim, S.K., Villeneuve, A.M., Reinke, V. (2002) X-chromosome silencing in the germline of C. elegans, *Development* **129**, 479–492.

Kim, S.K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J.M., Eizinger, A., Wylie, B.N., Davidson, G.S. (2001) A gene expression map for Caenorhabditis elegans, *Science* **293**, 2087–2092.

Kiontke, K., Gavin, N.P., Raynes, Y., Roehrig, C., Piano, F., Fitch, D.H. (2004) Caenorhabditis phylogeny predicts convergence of hermaphroditism and extensive intron loss, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9003–9008.

Kuersten, S., Goodwin, E.B. (2003) The power of the 3′ UTR: translational control and development, *Nat. Rev. Genet.* **4**, 626–637.

Lee, R., Feinbaum, R., Ambros, V. (2004) A short history of a short RNA, *Cell* **116**, S89–S92, 81 p following S96.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., Goldberg, D.S., Li, N., Martinez, M., Rual, J.F., Lamesch, P., Xu, L., Tewari, M., Wong, S.L., Zhang, L.V., Berriz, G.F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H.W., Elewa, A., Baumgartner, B., Rose, D.J., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S.E., Saxton, W.M., Strome, S., Van Den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K.C., Harper, J.W., Cusick, M.E., Roth, F.P., Hill, D.E., Vidal, M. (2004) A map of the interactome network of the metazoan C. elegans, *Science* **303**, 540–543.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., Bartel, D.P. (2003) The microRNAs of Caenorhabditis elegans, *Genes Dev.* **17**, 991–1008.

Mango, S.E. (2001) Stop making nonSense: the C. elegans smg genes, *Trends Genet.* **17**, 646–653.

Marais, G., Duret, L. (2001) Synonymous codon usage, accuracy of translation, and gene length in Caenorhabditis elegans, *J. Mol. Evol.* **52**, 275–280.

McKim, K.S., Peters, K., Rose, A.M. (1993) Two types of sites required for meiotic chromosome pairing in Caenorhabditis elegans, *Genetics* **134**, 749–768.

Mello, C.C., Kramer, J.M., Stinchcomb, D., Ambros, V. (1991) Efficient gene transfer in C. elegans: extrachromosomal maintenance and integration of transforming sequences, *EMBO J.* **10**, 3959–3970.

Meyer, B.J. (2000) Sex in the wormcounting and compensating X-chromosome dose, *Trends Genet.* **16**, 247–253.

Miller, M.A., Cutter, A.D., Yamamoto, I., Ward, S., Greenstein, D. (2004) Clustered

organization of reproductive genes in the C. elegans genome, *Curr. Biol.* **14**, 1284–1290.

Murphy, C.T., McCarroll, S.A., Bargmann, C.I., Fraser, A., Kamath, R.S., Ahringer, J., Li, H., Kenyon, C. (2003) Genes that act downstream of DAF-16 to influence the lifespan of Caenorhabditis elegans, *Nature* **424**, 277–283.

Okimoto, R., Wolstenholme, D.R. (1990) A set of tRNAs that lack either the T psi C arm or the dihydrouridine arm: towards a minimal tRNA adaptor, *EMBO J.* **9**, 3405–3411.

Pasquinelli, A.E., Ruvkun, G. (2002) Control of developmental timing by micrornas and their targets, *Annu. Rev. Cell Dev. Biol.* **18**, 495–513.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., Ruvkun, G. (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA, *Nature* **408**, 86–89.

Reboul, J., Vaglio, P., Rual, J.F., Lamesch, P., Martinez, M., Armstrong, C.M., Li, S., Jacotot, L., Bertin, N., Janky, R., Moore, T., Hudson, J.R. Jr., Hartley, J.L., Brasch, M.A., Vandenhaute, J., Boulton, S., Endress, G.A., Jenna, S., Chevet, E., Papasotiropoulos, V., Tolias, P.P., Ptacek, J., Snyder, M., Huang, R., Chance, M.R., Lee, H., Doucette-Stamm, L., Hill, D.E., Vidal, M. (2003) C. elegans ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression, *Nat. Genet.* **34**, 35–41.

Reinke, V. (2002) Functional exploration of the C. elegans genome using DNA microarrays, *Nat. Genet.* **32** Suppl, 541–546.

Reinke, V., Smith, H.E., Nance, J., Wang, J., Van Doren, C., Begley, R., Jones, S.J., Davis, E.B., Scherer, S., Ward, S., Kim, S.K. (2000) A global profile of germline gene expression in C. elegans, *Mol. Cell.* **6**, 605–616.

Roy, P.J., Stuart, J.M., Lund, J., Kim, S.K. (2002) Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans, *Nature* **418**, 975–979.

Rugarli, E.I., Di Schiavi, E., Hilliard, M.A., Arbucci, S., Ghezzi, C., Facciolli, A., Coppola, G., Ballabio, A., Bazzicalupo, P. (2002) The Kallmann syndrome gene homolog in C. elegans is involved in epidermal morphogenesis and neurite branching, *Development* **129**, 1283–1294.

Shin, T.H., Mello, C.C. (2003) Chromatin regulation during C. elegans germline development, *Curr. Opin. Genet. Dev.* **13**, 455–462.

Sijen, T., Plasterk, R.H. (2003) Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi, *Nature* **426**, 310–314.

Simmer, F., Tijsterman, M., Parrish, S., Koushika, S.P., Nonet, M.L., Fire, A., Ahringer, J., Plasterk, R.H. (2002) Loss of the putative RNA-directed RNA polymerase RRF-3 makes C. elegans hypersensitive to RNAi, *Curr. Biol.* **12**, 1317–1319.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., Coulson, A., D'Eustachio, P., Fitch, D.H., Fulton, L.A., Fulton, R.E., Griffiths-Jones, S., Harris, T.W., Hillier, L.W., Kamath, R., Kuwabara, P.E., Mardis, E.R., Marra, M.A., Miner, T.L., Minx, P., Mullikin, J.C., Plumb, R.W., Rogers, J., Schein, J.E., Sohrmann, M., Spieth, J., Stajich, J.E., Wei, C., Willey, D., Wilson, R.K., Durbin, R., Waterston, R.H. (2003) The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics, *PLoS Biol.* **1**, E45.

Sulston, J.E., Schierenberg, E., White, J.G., Thomson, J.N. (1983) The embryonic cell lineage of the nematode Caenorhabditis elegans, *Dev. Biol.* **100**, 64–119.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., Mello, C.C. (1999) The rde-1 gene, RNA interference, and transposon silencing in C. elegans, *Cell* **99**, 123–132.

Tijsterman, M., May, R.C., Simmer, F., Okihara, K.L., Plasterk, R.H. (2004) Genes required for systemic RNA interference in Caenorhabditis elegans, *Curr. Biol.* **14**, 111–116.

Timmons, L. (2004) Endogenous inhibitors of RNA interference in Caenorhabditis elegans, *Bioessays* **26**, 715–718.

Timmons, L., Court, D.L., Fire, A. (2001) Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans, *Gene* **263**, 103–112.

Unhavaithaya, Y., Shin, T.H., Miliaras, N., Lee, J., Oyama, T., Mello, C.C. (2002) MEP-1 and a homolog of the NURD complex component Mi-2 act together to maintain germline-soma distinctions in C. elegans, *Cell* **111**, 991–1002.

Vastenhouw, N.L., Plasterk, R.H. (2004) RNAi protects the Caenorhabditis elegans germline against transposition, *Trends Genet.* **20**, 314–319.

Villeneuve, A.M. (1994) A cis-acting locus that promotes crossing over between X chromosomes in Caenorhabditis elegans, *Genetics* **136**, 887–902.

Vos, J.C., van Luenen, H.G., Plasterk, R.H. (1993) Characterization of the Caenorhabditis elegans Tc1 transposase in vivo and in vitro, *Genes Dev.* **7**, 1244–1253.

Wadsworth, W.G. (2002) Moving around in a worm: netrin UNC-6 and circumferential axon guidance in C. elegans, *Trends Neurosci.* **25**, 423–429.

Wicks, S.R., Yeh, R.T., Gish, W.R., Waterston, R.H., Plasterk, R.H. (2001) Rapid gene mapping in Caenorhabditis elegans using a high density polymorphism map, *Nat. Genet.* **28**, 160–164.

Wicky, C., Villeneuve, A.M., Lauper, N., Codourey, L., Tobler, H., Muller, F. (1996) Telomeric repeats (TTAGGC)n are sufficient for chromosome capping function in Caenorhabditis elegans, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 8983–8988.

Wilm, T., Demel, P., Koop, H.U., Schnabel, H., Schnabel, R. (1999) Ballistic transformation of Caenorhabditis elegans, *Gene* **229**, 31–35.

Zahler, A.M., Tuttle, J., Chisholm, A.D. (2004) Genetic suppression of intronic +1 G mutations by compensatory U1 snRNA changes in *C. elegans*, *Genetics* **167**, 1689–1696.

Zetka, M., Rose, A. (1995) The genetics of meiosis in Caenorhabditis elegans, *Trends Genet.* **11**, 27–31.

Nematodes, Neurobiology and Development of

Darren R. Brooks[1], *Ian A. Hope*[2], *and R. Elwyn Isaac*[2]
[1] *School of Environment and Life Sciences, University of Salford, Salford, UK*
[2] *School of Biology, University of Leeds, Leeds, UK*

1 Nematodes 71
1.1 The Nematoda and their Body Plan 71
1.2 The Model Organism *Caenorhabditis elegans* 73
1.3 Molecular Genetics and Postgenomic Biology of *C. elegans* 73

2 Nematode Neurobiology 76
2.1 Anatomy of the Nematode Nervous System 76
2.2 Nematode Neurotransmission 78
2.3 Behavioral Mutants of *C. elegans* 79
2.4 *Caenorhabditis elegans* as a Neurodegenerative Disease Model 81

3 Development in *C. elegans* 81
3.1 Normal Development 81
3.2 Manipulation of Development 82
3.3 Nervous System Development 84
3.4 Cell Fusion 85
3.5 Programmed Cell Death 86
3.6 Aging 87

Acknowledgment 87

Bibliography 87
Books and Reviews 87
Primary Literature 88

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

Keywords

Expressed Sequence Tags (ESTs)
cDNA clones that are derived from a cDNA library and partially sequenced to allow alignment with the genomic sequence.

Green Fluorescent Protein (GFP)
A naturally fluorescent protein in the jellyfish *Aequorea victoria*. Heterologous expression of the *gfp* gene is used extensively in many biological systems to address questions related to cell and molecular biology.

Microarray
A technique for analyzing simultaneous changes in expression of large numbers of genes that is based upon hybridization of fluorescently labelled mRNAs to DNA "spots," each representing a gene, fixed to a solid support such as a glass slide.

MicroRNAs
Short single-stranded RNAs (~22 nucleotides) that are able to bind imprecisely to the 3′-untranslated regions of genes expressed during development, so mediating key temporal control of development.

Neuropeptides
Short peptides synthesized in nerve cells for use as neurotransmitters at synapses, or as neurohormones.

RNA Interference (RNAi)
The use of double-stranded RNA to elicit posttranscriptional gene silencing. RNAi was discovered using *C. elegans* and it is now widely utilized across the microbial, animal, and plant kingdoms.

Nematodes are ubiquitous roundworms. Many species are parasitic, causing immense economic damage to crops and livestock. Around a third of the world's population suffer from nematode infections.

The free-living *Caenorhabditis elegans* is a powerful model for the study of nematode biology, including animal development and behavior, as recognized by the award of the Noble Prize in Physiology or Medicine (2002) to Sydney Brenner, John Sulston, and Robert Horvitz, eminent biologists who pioneered this research field. *Caenorhabditis elegans* is amenable to classical genetic analysis, allowing large numbers of genes required for normal development and behavior to be identified by generating mutations. A scientific landmark was accomplished in 1998 with the publication of the entire genome of *C. elegans*, demonstrating that the complete

sequencing of the much larger human genetic code was technically feasible. The primary repository of all *C. elegans* physical and genetic data is Wormbase (www.wormbase.org). The draft genome of the closely related *C. briggsae* has recently been published, and efforts are ongoing to sequence the genome of the filarial parasite *Brugia malayi*. These genome sequences, together with numerous nematode expressed sequence tag (EST) projects from animal and plant parasites (www.nematode.net and www.nematodes.org), and the integrated physical and genetic map of the free-living marine nematode *Pristionchus pacificus*, will allow a more detailed understanding of the biology of the Nematoda. Such studies will present opportunities for the design of novel strategies for the control and therapy of parasitic nematodes.

1 Nematodes

1.1 The Nematoda and their Body Plan

Nematodes are probably the most ubiquitous phyla on earth. It is estimated that at least 40 000 species of nematode exist, of which about 20 000 have been described. Approximately 30% of described species are parasites of vertebrates, the remainder are either plant or invertebrate parasites, or free living. With one notable exception, *C. elegans*, parasitic species are the most heavily studied because of their economic and medical importance.

Nematodes are pseudocoelomic animals and their classification, historically, divided them into two groups, the Phasmidia (Secernentia), which are mainly marine, and the Aphasmidia (Adenophorea), which are predominantly terrestrial. With increasing genome sequence data becoming available, phylogenetic analyses segregate nematodes into five major clades and hence, the earlier classification of Phasmidia and Aphasmidia is now outdated. Parasitism appears to have evolved independently within each clade.

The typical nematode has a vermiform cylindrical body, tapered at both ends, with a pointed tail and blunt head (Fig. 1). The body wall is covered in a collagenous cuticle, comprised of scleroproteins, lipids, and carbohydrates. The cuticle may be smooth, or may bear a number of striations, ridges, or protuberances known as *alae* and *annular ridges*. Beneath the cuticle is a cellular, or syncytial hypodermis and a layer of longitudinal muscle cells. The muscle cells are uninucleated and each cell sends out at least one process towards a medial nerve. Nematode locomotion is constrained by the lack of circular muscles and by the rigidity of the cuticle. Many move in a sinusoidal manner, alternately contracting and relaxing opposed sections of the body. The nervous system is situated just below the hypodermis and consists of the nerve ring (brain) which encircles the pharynx, a complex head sensory system, and ventral and dorsal nerve cords that extend posterior from the nerve ring and run through almost the entire length of the body.

The alimentary system consists of a terminal mouth leading to a buccal cavity. The buccal cavity is lined with cuticle that can be specialized into plates, ridges, or, in

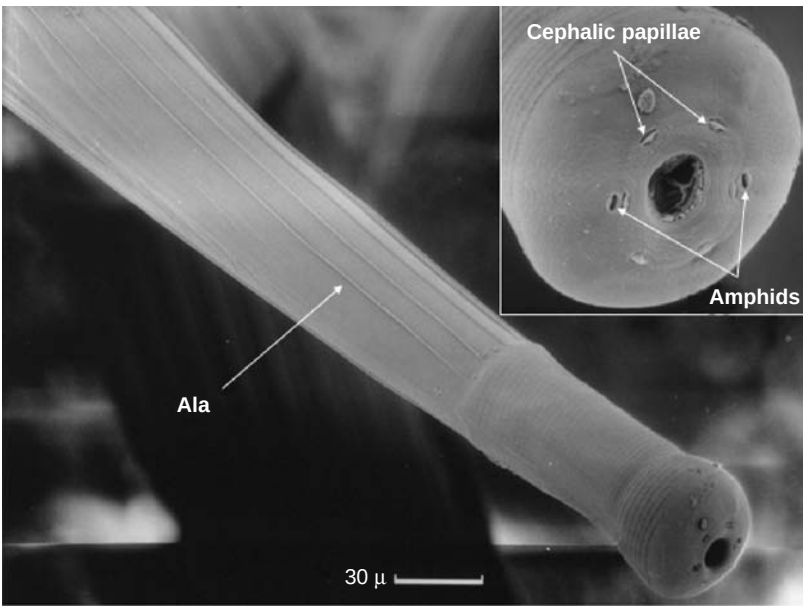

Fig. 1 Scanning electronmicrograph of an adult *Nematodirus battus*, a trichostrongyle intestinal parasite of ruminants. The image of the head (inset) shows the mouth surrounded by sensory organs.

the case of some carnivorous nematodes, teeth. The buccal cavity then leads into a muscular pharynx, an intestine, which is not lined with muscle, the rectum, and finally, a ventral subterminal anus. The gonads lie in the pseudocoelomic body cavity, which is bounded by the hypodermis and the intestine wall. The fluid filling the pseudocoelom provides a hydrostatic skeleton. Respiratory and circulatory organs are absent from nematodes.

The life cycle of parasitic nematodes may involve a free-living stage, or an intermediate host, as well as the primary host. The success of these nematodes is due to their ability to sense and to adapt to a new environment by changes in their structure, physiology, and behavior. Free-living nematodes also monitor and respond to changes in their surroundings, allowing them to survive shortages of food or other harsh environmental conditions. Sense organs are either chemosensory or mechanosensory and are found predominantly in the head region surrounding the mouth and in structures termed *amphids*, which are large paired chemosensory organs lying laterally on either side of the mouth. Other sense organs are found in lateral positions and those present in the posterior region of male nematodes are involved in the control of copulation. One group of nematodes also possess phasmids. These organs are also chemosensory and make contact with the external surroundings via pores on each side of the tail.

Nematodes follow a distinct developmental program: the embryo, four larval stages, and the adult organism. The cuticle is shed at the end of each larval stage. Many free-living and insect-parasitic nematodes also have a modified L3 larval

stage, termed the *dauer stage*, which is an adaptation to survive harsh conditions and hence, permit dispersal.

The majority of nematode species contain both sexes, but several, most notably, *C. elegans,* are hermaphroditic, with males arising only rarely.

1.2 The Model Organism *Caenorhabditis elegans*

Caenorhabditis elegans is a protandrous hermaphroditic nematode that inhabits soil and feeds mainly on bacteria. In the early 1960s, Sydney Brenner suggested that this organism be used as a model for metazoan development and, indeed, it has many features that make it ideal for the task. Its size allows it to be handled like a standard laboratory microorganism. The life cycle is relatively short (approximately 3.5 days at 20 °C) and the brood size large (>300). One sex of the species is hermaphroditic with two X chromosomes. Self-fertilization of the hermaphrodite allows individuals homozygous for recessive mutations to be generated from heterozygotes, a characteristic of immense value in genetic screens. The males of the species, with a single X chromosome, arise spontaneously in the population because of X chromosome nondisjunction, and they are vital in genetic experiments to permit combinations of different mutations. Interestingly, recent work has shown that some XX larvae generated by cross-fertilization lose one X chromosome when grown on specific metabolites. The same does not occur with XX larvae generated by self-fertilization, suggesting that sexual reproduction increases postembryonic developmental plasticity, allowing better adaptation of the species to environmental changes.

At all stages of its development, *C. elegans* is transparent, allowing continuous observation of individual cells. This advantage has led to a comprehensive map of all the developmental cell lineages in the animal. Moreover, since the cell lineage is almost entirely invariant in all animals, the properties of individual cells can be studied within the context of the whole developmental process. An individual cell can be ablated with a laser beam and the developmental mechanisms responsible for that cell's fate inferred from the consequences. Despite the invariance of the cell lineages, cell-to-cell communication appears to play a vital role in *C. elegans*' development. Interestingly, molecules implicated in cell–cell communication have been found to contain homology to conserved vertebrate signalling molecules.

Caenorhabditis elegans responds to starvation and to overcrowding by diverting second stage (L2) larvae to dauer larvae that are quiescent and resistant to harsh environmental conditions. In the presence of fresh bacteria, dauer larvae develop into fourth stage (L4) larvae and continue development to the normal adult stage. High levels of a *C. elegans* pheromone and low levels of an *Escherichia coli* food signal serve as two chemosensory cues to trigger the switch to dauer larvae. These signals are sensed by the amphids that comprise 12 sensory neurons, 8 of which end with nonmotile cilia that are in contact with the external environment via the amphidial pore.

1.3 Molecular Genetics and Postgenomic Biology of *C. elegans*

Classical forward genetic analysis of *C. elegans* is carried out by subjecting nematodes to a mutagen, such as ethyl methane

sulphonate (EMS), which induces random mutations in the germ line. Subsequent generations are screened for visible phenotypes. Using such approaches, Sydney Brenner and colleagues identified hundreds of mutants that included phenotypes showing movement and morphology defects. Genes, thereby defined through these mutations, are placed on a genetic map by performing genetic crosses and by analyzing the segregation of the phenotype with other known markers. Genes are molecularly cloned with reference to the physical map. Initially, cosmids are selected that span the desired genetic locus. The phenotype is then rescued by microinjection of the mutant with a series of overlapping cosmids. This should permit identification of a candidate gene and allow functional complementation to be carried out with a plasmid encoding the endogenous protein, hence proving unambiguously the genotype–phenotype relationship. Forward genetic approaches have defined about 10% of the predicted *C. elegans* genes, and when combined with emerging technologies, elucidation of the biological role of the remaining genes may be possible.

With the completion of the *C. elegans* genome sequence, techniques were developed to isolate mutants carrying mutations in desired genes – an approach termed *reverse genetics*. Chemical mutagenesis (EMS) is used to generate banks of mutants that are subsequently screened by the polymerase chain reaction (PCR) to detect a deletion in the gene of interest. The *C. elegans* gene knockout consortiums (http://celeganskoconsortium.omrf.org and http://shigen.lab.nig.ac.jp/c.elegans) are utilizing such methodology to isolate mutants for all predicted genes. Alternatively, the genome of *C. elegans* contains a number of endogenous *Tc* transposons that integrate randomly in the genome. These *Tc* transposons have been used to generate libraries of mutants that contain a transposon insertion in any given gene. PCR screens are again employed to isolate a mutant in a specific gene.

Forward and reverse genetic approaches are time consuming and the initial method of choice to analyze gene function in *C. elegans* is now RNA interference (RNAi). The double-stranded RNA (dsRNA) can be introduced by microinjection, overnight soaking, or by feeding nematodes on RNAse-deficient bacteria induced to express the dsRNA from a T7 promoter-based plasmid. The latter two approaches are amenable to genome-scale analyses. The ability of dsRNA to spread throughout the worm body, known as *systemic RNAi*, requires the presence of the transmembrane protein SID-1. Neuronal expressed genes appear less susceptible to RNAi, although, mutants are now available with increased sensitivity to dsRNA. RNA interference has transformed gene analysis and allowed genome-wide screens to be carried out by feeding and soaking. This RNAi-generated phenome dataset reveals functions for 14% of the genome, with the most common phenotype being embryonic lethality. Evolutionary conserved genes are more likely to result in a RNAi phenotype than genes that have evolved more recently. RNAi appears to be ubiquitous and hence, unsurprisingly, it is now reported effective as a gene silencing mechanism in both animal and plant parasitic nematodes. Although RNAi is extremely powerful, a thorough description of the biological function of a gene and its genetic pathway will still necessitate further analyses using a gene knockout mutant and enhancer and/or suppressor screens.

The metazoan developmental program requires precise coordination of the temporal and tissue-specific expression of large numbers of genes to correctly generate the morphological structures of the body. Unusually, for a eukaryote, *C. elegans* has a high proportion of genes that are found in operons and, hence, are transcribed polycistronically. Unlike prokaryotic operons, genes found within *C. elegans* operons are generally not functionally related. *Caenorhabditis elegans* is ideally suited for analyses of the spatio/temporal control of gene expression since germ-line transformation with promoter–reporter gene constructs is relatively easy. Furthermore, because *C. elegans* is transparent throughout development and the entire cell lineage has been determined, reporter gene activity can be studied precisely in all developmental stages. Our laboratory (IAH) is currently involved in analyzing large numbers of these expression patterns (http://bgypc086.leeds.ac.uk).

Gene expression patterns are generated by fusion of the genomic region immediately upstream of a gene (promoter) to a reporter gene (*lac*Z or *gfp*). The *lac*Z gene encodes the enzyme β-galactosidase, and its sites of expression are readily visualized with a histochemical stain that produces a blue precipitate wherever gene product is generated. The major advantage of using a *gfp* reporter gene is that expression patterns can be visualized in living animals. A significant proportion of such transgenes fail to drive reporter gene expression, which may reflect a higher proportion of pseudogenes present within the genome than initially envisaged. Analyses of the expression patterns of homologous *C. elegans* and *C. briggsae* promoters in both species show that the upstream regulatory elements are well conserved. Interesting differences do however exist; the *lin-48* gene of *C. elegans*, but not *C. briggsae*, is expressed in the excretory duct cell and this difference correlates with the more posterior position of the excretory pore in *C. elegans*. Promoters from several parasitic nematodes have been shown to drive reporter gene expression in *C. elegans* in a manner that shows good spatial agreement, though not so good temporal regulation, with the *C. elegans* homolog. This may reflect an increased rate of evolution upon the temporal element of nematode promoter sequences.

A large library (~190 000) of expressed sequence tags has been generated for *C. elegans*. A significant proportion of these ESTs show no homology to sequences deposited in databases. A thorough understanding of the biology of *C. elegans*, or any other organism, will require elucidation of the biological roles that these novel proteins, identified from EST databases, play. *In situ* hybridization studies, also using EST sequences, have determined approximately 7000 mRNA expression patterns for *C. elegans* genes (http://nematode.lab.nig.ac.jp).

Many open reading frames have been amplified from cDNAs and cloned using bacteriophage λ-based recombination technology (Gateway™ cloning). The resulting library of ORFeomes permits subsequent ease of manipulation into plasmids specialized for addressing different biological questions. This technology has also been used to clone PCR-amplified cDNAs for about 70% of predicted genes for which there is no EST evidence, generating a set of ORF sequence tags (OSTs). This dataset has raised the total number of experimentally confirmed genes in *C. elegans* to ~17 500 and indicated that computer predictions for gene splicing are ~70% accurate.

Global changes in gene expression, analyzed using microarray technology, have provided a *C. elegans* gene expression map that clusters genes with well-coordinated expression throughout development (http://cmgm.stanford.edu/~kimlab). A recent refinement has shown that it is possible to analyze gene expression using microarray technology in specific cells by selectively expressing **green fluorescent protein** (GFP) in the desired cells and, then, culturing these cells following extraction from the embryo with fluorescence activated cell-sorting (FACS). The identification of genes involved in pharyngeal organogenesis combined microarray, bioinformatic, and molecular genetic approaches and demonstrated that a forkhead box A transcription factor regulated this gene network.

The developmental profiling of protein expression is required to fully understand the complexity of biological processes since many events are mediated by post-translational protein modifications and, hence, not detectable by global gene expression analyses. To this end, technical improvements in two-dimensional protein gel analysis and identification by mass spectrometry (proteomics) have allowed *C. elegans*' stage-specific and temperature-regulated proteins to be identified.

Protein-interaction maps, utilizing yeast two-hybrid technology, are also being constructed as a means of understanding the proteome in the context of specific biological processes. The combination of interactome and phenome (RNAi) approaches is extremely powerful and has allowed identification of candidates involved in the response to DNA damage, including the orthologue of a gene frequently mutated in chronic lymphocytic leukemia. Other protein–protein interactions have been studied for vulval development, the germ line, and the proteasome (http://vidal.dfci.harvard.edu) and in total, the interactome map for *C. elegans* currently contains about 5500 interactions.

A complete understanding of molecular interactions requires knowledge of the three-dimensional structures of the interacting proteins. To this end, a *C. elegans* structural genomics initiative is underway, with the aim of expressing and purifying all predicted proteins and then producing high quality crystals suitable for X-ray diffraction (http://sgce.cbse.uab.edu).

2 Nematode Neurobiology

2.1 Anatomy of the Nematode Nervous System

Many nematode researchers study nematode neurobiology in order to understand more complex neural systems and their emergent behaviors. Other researchers are interested in the design of antinematode drugs, and the nervous system has been the focus for many of these studies. Almost all that is known about the neurobiology of nematodes comes from work done on *C. elegans* and *Ascaris suum*, an intestinal parasite of pigs. As the nervous systems of *C. elegans* and *A. suum* are well conserved, it is likely that the results of these studies can readily be extrapolated to many other nematodes.

The nervous system of *C. elegans* has been described completely at the cellular level and contains 118 classes of neuron comprising sensory neurons, motor neurons, and interneurons (http://www.wormatlas.org). The adult hermaphrodite contains 302 neurons and

56 supporting cells (socket and glial-like sheath cells). Owing to differences in the number of sex-specific neurons required for mating, the adult male contains 381 neurons and 92 supporting cells. Electron micrographs were used to determine the position and connectivity of every neuron within the animal and the resulting "wiring diagram" reveals that *C. elegans* contains about 5000 chemical synapses, 600 gap junctions, and 2000 neuromuscular junctions. The neuronal pattern is constant from animal to animal and individual neurons have been assigned functions following their removal by laser ablation.

Most neuronal cell bodies are in the head region, surrounding the pharynx, which itself contains 20 neurons that are physically separated from the rest of the nervous system by a basement membrane. The remaining cell bodies are in the ventral, lateral, or posterior regions of the tail. A single process (axon) usually emanates from the cell body, and synaptic contact can be formed with multiple adjacent processes (dendrites). The presynaptic region thickens and appears electron dense by electron microscopy, indicative of the accumulation of synaptic vesicles. Dendrites receive multiple synaptic inputs, though the lack of electron-dense material, indicative of a postsynaptic region in other organisms, hinders assignment. The pharynx is surrounded circumferentially by the nerve ring, the brain of the worm, which contains most of the interneurons, and synaptic connections. The nerve ring receives and transmits information via processes within the dorsal, ventral, and two lateral nerve cords that extend along the entire length of the body. The nerve cords are interconnected by commissures. Neurons of the head region are organized into an anterior ganglion and five ganglia that are posterior to the nerve ring. The tail region contains a preanal ganglion and four ganglia that are posterior to the anus.

Many of the sensory neurons project processes to the amphids in the nose tip, giving direct access to the environment and allowing chemosensation. A single amphid neuron is able to respond to multiple chemicals, suggesting that multiple receptors exist. Individual chemosensory neurons appear to sense distinct chemicals, with some overlap. Other chemosensory neurons project into the pseudocoelomic fluid. All chemosensory cells are surrounded by a sheath cell whose secretions are vital for chemoreception. Temperature is monitored by a thermosensory amphid neuron, allowing detection of very small temperature changes (0.05 °C). The thermosensory neuron has an enlarged membranous area, perhaps indicating that thermal changes may be detected via changes in membrane fluidity. Six touch-sensory neurons are present in the head and tail region, with processes embedded within the hypodermis to allow fine response to mechanosensation. Touch cells have specialized microtubules with 15 protofilaments, as opposed to the usual 11 protofilaments, which possibly confers greater rigidity toward mechanical stresses.

The motor neurons innervate body-wall, pharyngeal, egg-laying, and defecation muscles. In addition, the male has further motor neurons that control mating behavior; contact, turning, vulva location, spicule insertion, and sperm transfer. These behaviors require the activity of ray sensory neurons, hook and postcloacal sensilla, and spicule sensory neurons. Body-wall and defecation neuromuscular junctions are unique since the muscle cells project neuronal-like processes, termed *muscle arms*, to the motor neuron and synapses are made on the muscle arm.

The body-wall muscles receive synaptic input from excitatory and inhibitory motor neurons of the ventral and dorsal nerve cords, allowing out-of-phase contractions and, hence, sinusoidal movement.

Electrophysiological analyses of individual neurons allows assignment of the physiochemical properties of a nerve cell relative to time and activity and, hence, assesses the contribution any one cell can make to the neuronal circuit and animal behavior. Most work has been carried out with the porcine parasite, *A. suum*, since it can grow to 30 cm in length, aiding dissection of nervous tissues and the visualization of identified neurons for patch-clamp recordings. The large number of *C. elegans* behavioral mutants is an important resource for electrophysiology studies, however, the small size of this model nematode has hindered progress. Most work has been carried out with neuromuscular junctions of the pharynx and a small number of sensory and locomotory neurons. Recent advances with primary culture techniques for *C. elegans* neurons and muscle cells and the use of fluorescent indicators of membrane potential and calcium dynamics should ensure that exciting progress is made in this area.

2.2 Nematode Neurotransmission

Completion of the *C. elegans* genome-sequencing project has revealed that the basic machinery of neurotransmission exhibits much similarity with mammals, though genes associated with gap junctions and chemosensory receptors appear to have evolved independently. There are numerous types of potassium channels and several calcium and chloride channels, but no voltage gated sodium channels. Remarkably, about 5% of *C. elegans* genes encode G protein-coupled receptors. Most of these are designated as orphan receptors since the ligands, and hence functions, are unknown, though many are likely to be chemosensory receptors.

The major neurotransmitter of excitatory motor neurons is acetylcholine, while GABA (γ-aminobutyric acid) is the main neurotransmitter of inhibitory motorneurons. Acetylcholine results in cation channels becoming permeable and a number of antinematode drugs target the cholinergic system of nematodes. Organophosphorous anticholinesterases prevent inactivation of acetylcholine, while the imidazothiazoles, such as levamisole, activate nicotinic acetylcholine receptors at nerve–muscle junctions. Muscle cells also possess a muscarinic-like acetylcholine receptor, explaining why nicotinic receptors are not essential for the survival of *C. elegans*. γ-Aminobutyric acid stimulates opening of Cl^- channels, and in *C. elegans*, the GABAergic system is associated with three distinct behaviors. Most GABA neurons are involved in regulating the waves of contraction that run along the worm's body during locomotion. In response to cholinergic excitatory neurons causing muscle contraction, the GABAergic neurons inhibit muscle contractions on the opposite side of the body, initiating the characteristic sinusoidal pattern of movement. Animals with defects in these neurons have dorsal and ventral contractions occurring simultaneously, causing the animal to shrink. Other GABAergic neurons are involved in limiting the motion of the head during foraging. In contrast with the inhibitory function of other GABAergic neurons, two excitatory GABAergic neurons stimulate the final, expulsive step of defecation.

Glutamate is required for pharyngeal pumping and more recently, it has been

associated with long-term memory in *C. elegans*. Glutamate activates a Cl^- channel that is sensitive to ivermectins, a class of extremely potent drugs that reduce motor activity in nematodes. *Caenorhabditis elegans* also contains AMPA and NMDA receptor classes that respond to glutamate.

Biogenic amines found in nematodes include dopamine and serotonin, which have been localized to small subsets of neurons. Serotonin is found in the hermaphrodite-specific neurons (HSN), which control egg laying in *C. elegans*, and is present in two neurosecretory cells in the pharynx of *C. elegans* and *A. suum*. Serotonin has also been implicated in the regulation of locomotion and carbohydrate metabolism in *A. suum*, and the coordination of male copulatory behavior in free-living nematodes.

Dopamine appears to be the major catecholamine present in many nematode species, inhibiting locomotion and egg laying in *C. elegans*. Octopamine antagonizes the effect of serotonin and causes defects in movement and reduction in egg laying, pharyngeal pumping, and defecation. Noradrenalin has been identified in some species of nematode, but appears to be absent in *C. elegans*.

Immunocytochemistry has revealed peptidergic neurons in the pharyngeal nerves and the nerve cords, and a family of neuropeptides with Arg-Phe-NH_2 at the C-terminus have been isolated from *A. suum*, *C. elegans*, and *Panagrellus redivivus*. More than 60% of *A. suum* neurons are immunoreactive for Phe-Met-Arg-Phe-NH_2-related peptides (FaRPs), and a number of FaRPs have been shown to have potent effects on the inhibitory motorneurons of *A. suum*. *Caenorhabditis elegans* genome predictions indicate that 92 genes encode neuropeptides. This rich diversity includes 23 *flp* genes that encode multiple (~60) FaRPs, 37 insulin-related genes, and 32 *nlp* (neuropeptide-like protein) genes. Mutation of the *flp-1* gene causes defects in locomotion, mechanosensation, chemosensation, and egg laying.

2.3 Behavioral Mutants of *C. elegans*

The cell lineage and connectivity of all 302 neurons present in *C. elegans* hermaphrodites have been described in detail. This information, coupled with the powerful genetics and the availability of large numbers of behavioral mutants affecting locomotion, sensory input, survival, and reproduction has made *C. elegans* an excellent model for studying the genetic pathways that determine the development and functioning of the nervous system. Such studies have made major contributions to the understanding of nervous systems of nematodes and higher organisms; examples being the identification of the first GABA transporter (*unc-47*), the first glutamate transporter (*eat-4*), the first odorant receptor (*odr-10*), and the first member of the netrin family of axon-guidance molecules (*unc-6*).

Survival encompasses such behaviors as foraging, feeding, defecation, and dauer larva formation. Feeding is achieved through pharyngeal muscle contractions, allowing a grinder in the terminal bulb of the pharynx to break up bacteria before entry into the intestine. Laser ablation has shown that three neurons are essential for pharyngeal pumping, and by taking recordings from dissected pharyngeal muscles (electropharyngeograms), the M3 motor neurons were shown to inhibit pumping by controlling the time of muscle relaxation in response to the neurotransmitter glutamate. The *eat-4* mutant shows delayed

pharyngeal relaxation due to a reduction in the release of glutamate into neuronal terminals.

Many egg-laying defective mutants have been described, some of which fail to synthesize serotonin in the HSN (hermaphrodite-specific neurons) cells, while other mutants have abnormalities in HSN morphology and migration. Detailed studies of egg-laying defective mutants have led to the elucidation of the genetic pathway for the development of the HSN cells and includes the identification of genes that control HSN migration, axonal outgrowth, and the biosynthesis of serotonin.

Mutations (dauer-constitutive or *daf-c* and dauer-defective or *daf-d*) that affect the induction of dauer larvae by pheromone stimulation have led to the construction of a genetic pathway controlling dauer formation. Temperature-sensitive *daf-c* mutants enter the dauer stage even under favorable growing conditions. In contrast, *daf-d* mutants block dauer formation under strong dauer-inducing conditions. Most of the *daf-d* genes are required for normal sensory cilia structure in the amphid sensillia. Laser ablation experiments have identified two amphid neurons that repress dauer formation in the absence of pheromone and which are derepressed in the presence of pheromone. Characterization of the protein products encoded by *daf* genes has provided molecular and biochemical understanding of the signal transduction pathway for dauer formation. For example, d*af-1* and *daf-4* genes encode cell-surface receptors belonging to the serine/threonine protein kinase family, and the DAF-4 receptor has been shown to bind members of the bone morphogenetic protein family.

Caenorhabditis elegans NPR-1 is a receptor for FaRP neuropeptides encoded by the *flp-18* and *flp-21* genes and is structurally related to the neuropeptide Y receptor family. A single amino acid change in NPR-1 determines whether the animal exhibits solitary (Val215) or social (Phe215) feeding behavior. Nematodes carrying an *npr-1* gene deletion are social feeders, indicating that *npr-1* functions to suppress social feeding. Social feeding is induced by stressful conditions (e.g. overcrowding) activating mechano- and chemosensory nociceptive neurons, resulting in the opening of a cation channel. Social feeding is also induced by activation of a cyclic GMP-gated ion channel present in three neurons that contact the pseudocoelomic fluid. The activation of NPR-1 in these neurons again leads to suppression of social feeding behavior. The interplay between the two neuronal systems, therefore, results in complex regulation of the feeding behavior of *C. elegans*.

The complex mating behavior exhibited by *C. elegans* males is also being genetically dissected. Many mutants have been isolated that affect male development and, hence, the capacity to copulate. The desire to copulate is determined by the presence of hermaphrodites and physiological indicators of nutritional and reproductive status, including serotonin. Mutation of the serotonin reuptake transporter, *mod-5*, which functions to remove serotonin from the synaptic cleft, reduces mate searching. Interestingly, inhibitors of the human serotonin reuptake transporter (e.g. the antidepressant Prozac) also decrease libido, indicating that it may be possible to understand certain mammalian motivational behaviors using worm genetics.

2.4 *Caenorhabditis elegans* as a Neurodegenerative Disease Model

Since the *C. elegans* nervous system contains the basic neuronal machinery and its genome encodes homologs of human neuronal disease genes, it is studied as a model system for neurodegenerative diseases. Dissecting genetic pathways involved in neuronal disease in *C. elegans* may offer new insights into the roles of crucial gene products and may possibly lead to novel strategies for neuroprotection in humans. Transgenic nematodes that model neurodegeneration might also be screened against chemical banks in the search for novel pharmaceutical compounds that may alleviate the disease symptoms.

Caenorhabditis elegans has a single gene, *apl-1*, that is homologous to human amyloid precursor protein. The nematode gene lacks the neurotoxic β-amyloid peptide fragment, and hence, modeling Alzheimer's disease has involved intracellular expression of human β-amyloid in *C. elegans*. Studies have shown that the endoplasmic reticulum heat shock protein, HSP70, is protective against β-amyloid plaque formation in this transgenic model. Presenilins are components of the γ-secretase complex that cleave the amyloid precursor protein and release β-amyloid. The *C. elegans* presenilin homolgue, *sel-12*, cleaves members of the Notch transmembrane receptor family, and genetic screens have identified candidate components of the γ-secretase complex. Large-scale presenilin inhibitor screens are underway to induce egg-laying defective *C. elegans* that phenocopy the *sel-12* mutant and, hence, provide lead compounds to combat Alzheimer's disease.

Polyglutamine (polyQ) disorders, such as Huntington's disease, are characterized by expansion within exons of the CAG triplet, resulting in intracellular protein aggregation and neurodegeneration. The N-terminal polyQ fragment of human huntingtin protein has been expressed in *C. elegans*. A genetic screen for enhancers of neurodysfunction identified *pqe-1* as a candidate for neuroprotection, most likely through the prevention of cellular proteins becoming aggregated with polyQ proteins.

Caenorhabditis elegans is also being exploited as a model of other neurodegenerative diseases, including disorders such as Parkinson's disease, mucolipidosis (*cup-5*), and necrotic-like neuronal cell death (*mec-4*).

3 Development in *C. elegans*

3.1 Normal Development

Caenorhabditis elegans is the only metazoan for which a developmental program has been described completely at the cellular level. A collection of time-lapse movies showing various stages of *C. elegans* development can be accessed via http://www.bio.unc.edu/faculty/goldstein/lab/movies.html. The embryonic development of *C. elegans* takes place over 14 h at 20 °C and occurs in two distinct stages. Following fertilization in the *spermatheca*, the oocyte completes meiotic division, the gamete pronuclei are fused, the egg shell is formed, and mitotic cell division is initiated. A period of rapid cell division then occurs to generate a 550-cell ovoid with little overt differentiation. It is during this period that the eggs are generally laid, though during adverse conditions (e.g. starvation), the hermaphrodite can retain eggs in the uterus. The second stage

consists of differentiation and morphogenesis with little additional cell division. If the eggs are still inside the uterus due to adverse conditions, the embryos will continue to develop and hatch, resulting in the characteristic "bag-of-worms" phenotype.

Caenorhabditis elegans has a largely invariant cell lineage. The initial unequal cell divisions generate six founder cells, AB, C, D, E, MS, and P_4. These founder cells give rise to cell lineages that divide synchronously and equally (daughter cells of equal size) with characteristic division rates. Some founder cells produce a single tissue; E gives rise to all of and only the intestine, whereas other founder cells produce many different tissue types.

Gastrulation starts 100 min after first cleavage when the embryo consists of 28 cells. The two daughter cells of the founder E move from the ventral surface to the embryo interior where they will form the intestine. The MS descendants, which form part of the pharynx and the myoderm, and the germ-line progenitor, P_4, follow the E lineage through the ventral cleft. Gastrulation is complete after cells of the D lineage and the descendants of C, which generate the body-wall muscle, complete their movements. Movement of neuroblast cells within the ectoderm, mediated by ephrin signalling events, ensures that the ventral cleft closes after 230 to 290 min of development. Epidermal cells are born about 240 min after the first cleavage, and epidermal enclosure of the embryo, a process termed *epiboly*, is initiated. The dorsal epidermal cells intercalate (290–340 min) and ventral epidermal enclosure becomes complete (310–360 min). Epidermal enclosure is mediated by ephrins, α- and β-catenins, cadherins, and semaphorins.

Morphogenesis occurs due to contractions of the circumferentially orientated actin microfilament bundles within the epidermis, causing cells to change shape and the embryo to increase in length, approximately fourfold (about 360–600 min). Muscle and nerve cells start to differentiate and the embryo begins to twitch. After elongation is complete, the hypodermis secretes the cuticle (about 650 min), ensuring that the characteristic shape of the worm is maintained.

Postembryonic growth results in an approximate fourfold increase in length, with the cuticle being shed at the end of each larval stage. Two members of an unusually large family of nuclear hormone receptors (*nhr-23* and *nhr-25*) and a novel member of the angiotensin-converting enzyme family that lacks a metallopeptidase active site (*acn-1*) are essential for these moults (Fig. 2). During larval development, 32 blast cells continue to divide and the number of somatic nuclei increases from 558 to 959 in the hermaphrodite and from 560 to 1031 in the male. The final cell divisions to occur are those of the lateral hypodermis (seam cells) during the L4 larval stage. The somatic cells generated postembryonically contribute mainly to the sexually dimorphic structures such as the gonads, the vulva of the hermaphrodite, and the male tail. The postembryonic cell lineage and migrations have been completely described.

3.2 Manipulation of Development

Individual cells in a developing animal can be destroyed by a tightly focused microlaser beam and the effect on development ascertained. Such experiments support the interpretation made from the invariant cell lineages that *C. elegans* development was

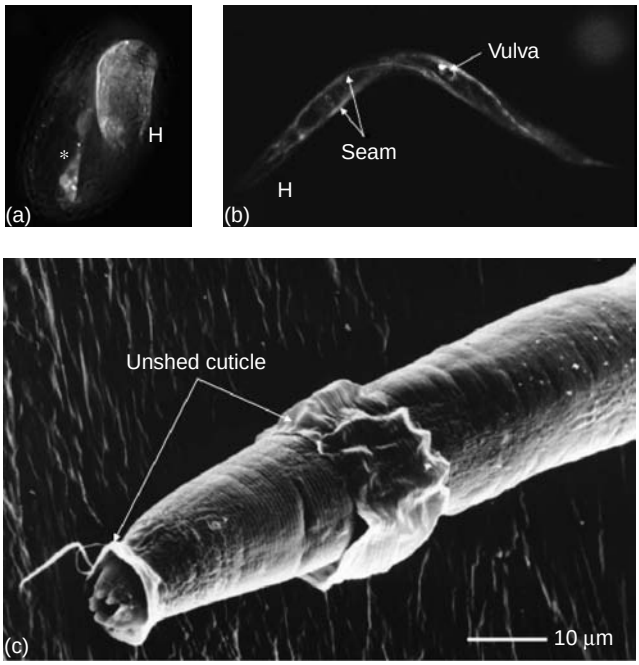

Fig. 2 Developmental expression and RNAi phenotype of *C. elegans acn-1*, a novel member of the angiotensin-converting enzyme family. Fluorescent images of transgenic *C. elegans (H; head)* expressing *acn-1::gfp* show reporter gene activity in (a) the lateral hypodermal seam cells (individual seam cell is highlighted with an asterisk) of a late-stage embryo and (b) in the seam and vulva cells of an L4 hermaphrodite. (c) Scanning electronmicrograph showing the head region of an adult *C. elegans* that has failed to shed cuticle during larval moults because of *acn-1 (RNAi)*.

largely cell autonomous; remaining cells failed to compensate for the structures that would have derived from the destroyed cell(s). For example, ablation of muscle precursors from the C or D lineages leads to larvae with gaps in the longitudinal bands of muscle.

However, there are several examples of nonautonomous development. In the developing gonad, the cells Z1.ppp and Z4.aaa are a cell pair for which the primary fate is to be an anchor cell and the secondary fate is to be a ventral uterine precursor cell. If either cell in this pair is destroyed by laser ablation, the remaining cell always adopts the primary fate. This suggests that the two cells are in ''competition'' for the primary fate, with the ''loser'' adopting the secondary fate. These two cells have equivalent developmental potential and are said to belong to an ''equivalence group''. Further laser ablation experiments have revealed other equivalence groups of postembryonic development that do not show natural variation during normal *C. elegans* development. During vulva development, Wnt and EGF signalling events distinguish the fate of the six cells of the ventral epidermis, which

form an equivalence group, known as the *vulva precursor cells*. The anchor cell of the gonad then provides the primary inductive signal (EGF-like protein LIN-3) for vulva formation by directing P6.p to adopt the primary fate, generating eight great-grandprogeny that form the center of the vulva (vulF and vulE cells). The flanking vulva precursor cells, P5.p and P7.p, adopt the secondary fate, forming seven progeny (vulA, B1, B2, C, and D cells). This is ensured as a consequence of P6.p lateral signalling, via the LIN-12 receptor, inhibiting P5.p and P7.p and hence, preventing these from adopting the primary fate. An inhibitory signal from the hypodermal syncytium, hyp-7, directs the two daughters of P3.p, P4.p and P8.p to adopt the tertiary fate and fuse with the hypodermis.

Early embryonic development also depends on many inductive interactions and AB.a and AB.p cells, that generate about two thirds of all embryonic cells, have equivalent developmental potential. Signaling from P_2 to AB.p is responsible for specifying the cell fates of the AB.p lineage.

Many genes involved in the cell lineages of *C. elegans* have been identified by mutations that affect the development of particular structures such as the vulva. One of the better characterized developmental mutations in *C. elegans* affects *lin-12*. This gene is best understood for its role in cell fate decision of the gonadal cells, Z1.ppp and Z4.aaa, as discussed above. Genetic analysis of *lin-12* mutants showed that the *lin-12* gene product specified the ventral uterine precursor cell fate of Z1.ppp or Z4.aaa. This analysis further suggested that *lin*-12 function is part of a receiving mechanism for a signal sent out by the presumptive anchor cell that enables expression of the secondary fate. The *lin-12* gene was isolated and found to encode a transmembrane protein with a sequence similarity to the Notch protein, which is involved in cell fate decisions in *Drosophila*.

Heterochronic genes coordinate the temporal sequence of events during development. Mutations in the heterochronic genes cause precocious or retarded development of specific cell lineages. Mutants in *lin-4, lin-14, lin-28, lin-29, lin-41,* and *lin-57* (*hbl-1*) have allowed the heterochronic gene pathway controlling the larval to adult switch to be characterized. The *lin-4* gene is expressed during L1 stage and it encodes a 22 nucleotide noncoding RNA, known as a *microRNA*, which represses *lin-14* and *lin-28* by imprecise base-pairing to their respective mRNAs. A second microRNA, *let-7*, is expressed during the L3 stage and it represses *lin-41* and *lin-57* (*hbl-1*). It is likely that many more microRNAs will be characterized with key functions during development. For example, the microRNA *lys-6* has recently been shown to control neuronal patterning by repression of *cog-1* during left/right axis formation in chemosensory neurons.

3.3 Nervous System Development

The AB cell lineage generates nearly all the neuronal cells of *C. elegans* (291 out of 302) and most (222) are formed during embryogenesis. A further 72 develop during the first larval stage and the final eight hermaphrodite neurons form during the L2 stage. Male-specific neurons continue to develop during the L3 larval stage. Neurons of the same cell type arise from different neuroblast sublineages that are reiterated. Most neuronal cells are born into their approximate final position and pioneer axonal processes emerge. The growth cone, present on

the tip of the axon, is able to detect and interpret guidance cues (netrins, slit, ephrins, and semaphorins) secreted from the ectodermal cells through modifications to its cytoskeleton. There are numerous axon pathfinding mutants, many of which map to transcription factors that are likely to regulate localized expression of the guidance molecules. A growth cone-independent mechanism also appears to operate in the migration of the proximal trajectory of the M2 motorneuron through the muscle cells of the pharynx. This unusual occurrence may be a consequence of pharyngeal development, which occurs via morphogenesis of a primordium of undifferentiated cells and without cell divisions. The pioneer axons guide subsequent axonal processes to their destinations. Nerve cell development is complete following onset of neurotransmitter expression. Most postembryonic ventral motor neurons express acetylcholine or GABA shortly after axonal migration. In contrast, expression of FMRF-related peptides in the VC cells, which arise during the first larval stage, does not occur until late in the L4 stage.

The nervous system of *C. elegans* exhibits plasticity during development. During embryogenesis, the GABAergic DD motor neurons innervate the ventral muscles and then, at late L1 stage, these rewire, losing connections with the ventral muscles and forming synapses with the dorsal muscles. During dauer formation, synaptic remodeling of some of the amphid chemosensory neurons also occurs. Some of the neurons develop larger processes, while the sensory endings of others retract within the amphidial pore. Neuromuscular changes occur during development of the adult male. For example, GABA release from the AVL and DVB neurons causes contraction of the enteric sphincter muscle in the L4 male and relaxation of it in the adult male.

3.4 Cell Fusion

Cell fusion is critical to *C. elegans* embryonic and postembryonic development and in the adult hermaphrodite, 300 of the 959 somatic cell nuclei are present within large multinucleate cells known as *syncytia*. During embryo elongation, cell fusion occurs between most of the dorsal cells and a number of the ventral cells to form the hypodermal syncytia. During each larval stage, the lateral hypodermal cells (seam cells) divide and one daughter cell remains in the seam, while the other daughter cell migrates and fuses with the hypodermal syncytia. Additional ventral epidermal cells also fuse with the hypodermal syncytia during larval development, with the net result being that the hypodermis of the adult comprises of eight syncytia (186 nuclei) and only three unfused cells. The seam cells also fuse during the L4-adult moult to generate the two lateral hypodermal syncytia of adults. The overall effect of all these cell fusions is to generate a tissue, and importantly, to also restrict the migratory paths of the remaining nonfusing cells.

Cell fusion events are also critical during development of the vulva. Cell fusions that occur during the L1 and L3 larval stages serve to affect cell fate determination of the ventral epidermal cells and vulva precursor cells. During the early L4 larval stage, cell fusions occur between migrating vulva primordial cells (precursors of the vulA and vulC cells). At the late L4 stage, intratoroidal cell fusions occur between the sister cells of the remaining five multicellular stacks to form a syncytial ring. In the final stage, the gonadal anchor

cell penetrates the vulva apex and fuses with the uterus.

Additional cell fusions occur during development of the pharynx, the excretory glands, and also the male tail. Genetic screens have identified EFF-1 as a key player in embryonic and postembryonic epithelial cell fusions. EFF-1 is an ectodomain protein attached to the cell surface by a single transmembrane domain. The functional basis of EFF-1 in cell fusion is probably related to the presence of a 22–amino acid putative fusion peptide sequence and/or a consensus phospholipase A2 active site within the large ectodomain.

3.5 Programmed Cell Death

During normal development of an adult *C. elegans* hermaphrodite, 131 out of 1090 somatic nuclei and large numbers of germ cells die by programmed cell death. During his Nobel Prize studies on the invariant cell lineage, John Sulston reported that most of the somatic programmed cell deaths occur during embryogenesis (113/135). A small number of programmed cell deaths also occur during the L2 stage of hermaphrodite development and in the male larval stages. Most programmed cell deaths involve neuronal and neuron-associated cells and hypodermal cells. A few muscle cells and sisters of the pharyngeal gland cells also undergo programmed cell deaths. For instance, the HSN neurons, required for egg laying, are destroyed by programmed cell death in males. Likewise, the CEM neurons, the male sensory neurons used in mating, undergo programmed cell death in hermaphrodites shortly after they arise.

Programmed cell death follows a characteristic pathway that is evolutionarily conserved. It involves a change in cell morphology, disintegration of the nucleus, and finally, engulfment of the dead cell by its healthy neighbors. Genetic analysis of *C. elegans* cell death mutants, pioneered by the Nobel Prize studies of Robert Horvitz, has identified a number of genes implicated in programmed cell death, with *ced-3* and *ced-4* being the most central. Normal developmental cell death is abolished in mutants lacking either of these genes, resulting in them having 12% more cells and a nervous system about one third larger than wild-type *C. elegans*. The *ced-3* gene encodes a cysteine protease from the caspase family, while CED-4 has homology to the mammalian apoptotic protease-activating factor, Apaf-1.

The activity of the gene *ced*-9 appears to be crucial in preventing cell death in inappropriate cells. The CED-9 protein is localized to the mitochondrial membrane and it shows sequence homology and functional similarity to the protein product of the human proto-oncogene *bcl-2*. Programmed cell death is suppressed when CED-9 binds CED-4, whereas cell death is initiated when CED-9 interacts with EGL-1, a BH3-domain-only member of the *bcl-2* family. The latter interaction releases CED-4, allowing it to activate the caspase CED-3, the key executioner of programmed cell death.

A recent genetic screen has also identified the *icd-1* gene as a suppressor of programmed cell death. ICD-1 is the β-subunit of the nascent polypeptide-associated complex and its expression suppresses cell death by a *ced-4* mediated, *ced-3* independent pathway, demonstrating that the core programmed cell death machinery can have distinct roles. This study also raises the possibility that other caspase family members may have crucial roles during programmed cell death.

3.6 Aging

There is huge divergence in life spans of species within Nematoda, with many free-living nematodes existing for weeks, whereas parasitic species can live for years (e.g. the filarid *Loa loa* may live up to 20 years). This disparity in longevity is a consequence of genetic differences that have evolved to reflect species ecology. Significant life-span differences can also be found within the same species of nematode and this is, in part, likely to reflect environmental variations. One particularly important factor for an ectothermic organism, such as a nematode, is temperature. For example, the *C. elegans* hermaphrodite has a maximum life span of 32 days at 16 °C and this decreases to 11 days at 25.5 °C. Another major influence is the availability of food, with significant life-span increases observed through dietary restriction. Recent laser ablation studies using *C. elegans* have led to the identification of specific gustatory and olfactory neurons that influence longevity. *C. elegans* mutants have been isolated with altered life span, and the dissection of genetic pathways of aging has begun. The *daf-2*/insulin receptor pathway is central to longevity; decreased signalling increases life span up to threefold. The *daf-2* mutant phenotype is suppressed by the forkhead transcription factor *daf-16*. Combining transcriptome and genome-wide RNAi analyses in *daf-2* and *daf-16* mutant backgrounds has identified target genes that modulate *C. elegans*' life span. Additional genetic screens have shown that *C. elegans*' mitochondrial proteins, including components of the energy generating electron transport chain, regulate longevity in a *daf-16* independent manner. One unifying hypothesis is that mutants with enhanced life spans use an anaerobic fermentative mitochondrial metabolic pathway, shown to operate in dauers, that results in a significant reduction in damaging free-radical oxygen species.

Acknowledgment

We wish to thank Darryl MacGregor for his contribution, Adrian Hick, Mike Winter, and Donald Lee (School of Biology, University of Leeds) for permission to use the electromicrograph in Fig. 1, Gareth Howell (Wellcome Trust Bioimaging Facility, University of Leeds) for assistance with Fig. 2.

Bibliography

Books and Reviews

Bargmann, C.I. (1998) Neurobiology of the *Caenorhabditis elegans* genome, *Science* **282**, 2028–2033.

Brownlee, D., Holden-Dye, L., Walker, R. (2000) The range and biological activity of FMRFamide-related peptides and classical neurotransmitters in nematodes, *Ad. Parasitol.* **45**, 109–180.

Chin-Sang, I.D., Chisholm, A.D. (2000) Form of the worm: genetics of epidermal morphogenesis in *C. elegans*, *Trends Genet.* **16**, 544–551.

Driscoll, M., Gerstein, B. (2003) Dying for a cause: invertebrate genetics takes on human neurodegeneration, *Nat. Genet. Rev.* **4**, 181–194.

Hobert, O. (2003) Behavioral plasticity in *C. elegans*: paradigms, circuits, genes, *J. Neurobiol.* **54**, 203–223.

Hope, I.A. (Ed.) (1999) *C. elegans: a practical approach*, Oxford University Press, Oxford.

Jorgensen, E.M., Mango, S.E. (2002) The art and design of genetic screens: *Caenorhabditis elegans*, *Nat. Genet. Rev.* **3**, 356–369.

Lee, D.L. (Ed.) (2002) *The Biology of Nematodes*, Taylor & Francis, London & New York.

Liu, Q.A., Hengartner, M.O. (1999) The molecular mechanism of programmed cell death in *C. elegans*, *Ann. N. Y. Acad. Sci.* **887**, 92–104.

Rankin, C.H. (2002) From gene to identified neuron to behaviour in *Caenorhabditis elegans*, *Nat. Rev. Genet.* **3**, 622–630.

Riddle, D.L., Blumenthal, T., Meyer, B.J., Priess, J.R. (Eds.) (1997) *C. elegans II*, Cold Spring Harbor Laboratory Press, New York.

Shemer, G., Podbilewicz, B. (2003) The story of cell fusion: big lessons from little worms, *BioEssays* **25**, 672–682.

The *C. elegans* sequencing consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology, *Science* **282**, 2012–2018.

Primary Literature

Aboobaker, A.A., Blaxter, M.L. (2003) Use of RNA interference to investigate gene function in the human filarial nematode parasite *Brugia malayi*, *Mol. Biochem. Parasitol.* **129**, 41–51.

Alcedo, J., Kenyon, C. (2004) Regulation of *C. elegans* longevity by specific gustatory and olfactory neurons, *Neuron* **41**, 45–55.

Blaxter, M.L., de Ley, P., Garey, J.R., Liu, L.X., Scheldeman, P., Vierstraete, A., Vanfleteren, J.R., Mackey, L.Y., Dorris, M., Frisse, L.M., Vida, J.T., Thomas, W.K. (1998) A molecular evolutionary framework for the phylum Nematoda, *Nature* **392**, 71–75.

Bloss, T.A., Witze, E.S., Rothman, J.H. (2003) Suppression of CED-3-independent apoptosis by mitochondrial βNAC in *Caenorhabditis elegans*, *Nature* **424**, 1066–1071.

Blumenthal, T., Evans, D., Link, C.D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W.L., Duke, K., Kiraly, M., Kim, S.K. (2002) A global analysis of *Caenorhabditis elegans* operons, *Nature* **417**, 851–854.

Brooks, D.R., Appleford, P.J., Murray, L., Isaac, R.E. (2003) An essential role in molting and morphogenesis of *Caenorhabditis elegans* for ACN-1, a novel member of the angiotensin-converting enzyme family that lacks a metallopeptidase active site, *J. Biol. Chem.* **278**, 52340–52346.

Boulton, S.J., Gartner, A., Reboul, J., Vaglio, P., Dyson, N., Hill, D.E., Vidal, M. (2002) Combined functional genomic maps of the *C. elegans* DNA damage response, *Science* **295**, 127–131.

Coates, J.C., de Bono, M. (2002) Antagonistic pathways in neurons exposed to body fluid regulate social feeding in *Caenorhabditis elegans*, *Nature* **419**, 925–929.

de Bono, M., Bargmann, C.I. (1998) Natural variation in a neuropeptide Y receptor homolog modifies social behaviour and food response in *C. elegans*, *Cell* **94**, 679–689.

Faber, P.W., Voisine, C., King, D.C., Bates, E.A., Hart, A.C. (2002) Glutamine/proline-rich PQE-1 proteins protect *Caenorhabditis elegans* neurons from huntingtin, *Proc. Natl. Acad. Sci. USA* **99**, 17131–17136.

Fares, H., Greenwald, I. (2001) Regulation of endocytosis by CUP-5, the *Caenorhabditis elegans* mucolipin-1 homolog, *Nat. Genet.* **28**, 64–68.

Feinberg, E.H., Hunter, C.P. (2003) Transport of dsRNA into cells by the transmembrane protein SID-1, *Science* **301**, 1545–1547.

Fonte, V., Kapulkin, V., Taft, A., Fluet, A., Friedman, D., Link, C.D. (2002) Interaction of intracellular beta amyloid peptide with chaperone proteins, *Proc. Natl. Acad. Sci. USA* **99**, 9439–9444.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*, *Nature* **391**, 806–811.

Gaudet, J., Mango, S.E. (2002) Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4, *Science* **295**, 821–825.

Herndon, L.A., Schmeissner, P.J., Dudaronek, J.M., Brown, P.A., Listner, K.M., Sakano, Y., Paupard, M.C., Hall, D.H., Driscoll, M. (2002) Stochastic and genetic factors influence tissue-specific decline in ageing *C. elegans*, *Nature* **419**, 808–814.

Hill, A.A., Hunter, C.P., Tsung, B.T., Tucker-Kellogg, G., Brown, E.L. (2000) A genomic analysis of gene expression in *C. elegans*, *Science* **290**, 809–812.

Jiang, M., Ryu, J., Kiraly, M., Duke, K., Reinke, V., Kim, S.K. (2001) Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*, *Proc. Natl. Acad. Sci. USA* **98**, 218–223.

Johnston, Jr., R.J., Hobert, O. (2003) A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*, *Nature* **426**, 845–849.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, G., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D.P., Zipperlen, P., Ahringer, J. (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi, *Nature* **421**, 231–237.

Kennedy, S., Wang, D., Ruvkun, G. (2004) A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*, *Nature* **427**, 645–649.

Kim, S.K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J.M., Eizinger, A., Wylie, B.N., Davidson, G.S. (2001) A gene expression map for *Caenorhabditis elegans*, *Science* **293**, 2087–2092.

Kraemer, B.C., Zhang, B., Levernez, J.B., Thomas, J.H., Trojanowski, J.Q., Schellenberg, G.D. (2003) Neurodegeneration and defective neurotransmission in a *Caenorhabditis elegans* model of tauopathy, *Proc. Natl. Acad. Sci. USA* **100**, 9980–9985.

Lee, R.C., Feinbaum, R.L., Ambros, V. (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*, *Cell* **75**, 843–853.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.-O., Han, J.-D.J., Chesneau, A., Hao, T., Goldberg, D.S., Li, N., Martinez, M., Rual, J.-F., Lamesch, P. et al. (2004) A map of the interactome network of the metazoan *C. elegans*, *Science* **303**, 540–543.

Link, C.D. (1995) Expression of human beta-amyloid peptide in transgenic *Caenorhabditis elegans*, *Proc. Natl. Acad. Sci. USA* **92**, 9368–9372.

Lund, J., Tedesco, P., Duke, K., Wang, J., Kim, S.K., Johnson, T.E. (2002) Transcriptional profile of aging in *C. elegans*, *Curr. Biol.* **12**, 1566–1573.

Maeda, I., Kohara, Y., Yamamoto, M., Sugimoto, A. (2001) Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi, *Curr. Biol.* **11**, 171–176.

Mohler, W.A., Shemer, G., del Campo, J.J., Valansi, C., Opoku-Serebuoh, E., Scranton, V., Assaf, N., White, J.G., Podbilewicz, B. (2002) The type I membrane protein EFF-1 is essential for developmental cell fusion, *Dev. Cell* **2**, 355–362.

Mounsey, A., Bauer, P., Hope, I.A. (2002) Evidence suggesting that a fifth of annotated *Caenorhabditis elegans* genes may be pseudogenes, *Genome Res.* **12**, 770–775.

Murphy, C.T., McCarroll, S.A., Bargmann, C.L., Fraser, A., Kamath, R.S., Ahringer, J., Li, H., Kenyon, C. (2003) Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*, *Nature* **424**, 277–283.

Nelson, L.S., Rosoff, M.L., Li, C. (1998) Disruption of a neuropeptide gene, *flp-1*, causes multiple behavioral defects in *Caenorhabditis elegans*, *Science* **281**, 1686–1690.

Nollen, E.A.A., Garcia, S.M., van Haaften, G., Kim, S., Chavez, A., Morimoto, R.I., Plasterk, R.H.A. (2003) Genome-wide RNA interference screen identifies previously undescribed regulators of polyglutamine aggregation, *Proc. Natl. Acad. Sci. USA* **101**, 6403–6408.

Prahlad, V., Pilgrim, D., Goodwin, E.B. (2003) Roles for mating and environment in *C. elegans* sex determination, *Science* **302**, 1046–1049.

Ranganathan, R., Sawin, E.R., Trent, C., Horvitz, H.R. (2001) Mutations in the *Caenorhabditis elegans* serotonin reuptake transporter MOD-5 reveal serotonin-dependent and -independent activities of fluoxetine, *J. Neurosci.* **21**, 5871–5884.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., Ruvkun, G. (2000) The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*, *Nature* **403**, 901–906.

Reinke, V., Smith, H.E., Nance, J., Wang, J., van Doren, C., Begley, R., Jones, S.J.M., Davis, E.B., Scherer, S., Ward, S., Kim, S.K. (2000) A global profile of germline gene expression in *C. elegans*, *Mol. Cell* **6**, 605–616.

Rogers, C., Reale, V., Kim, K., Chatwin, H., Li, C., Evans, P., de Bono, M. (2003) Inhibition of *Caenorhabditis elegans* social feeding by FMRFamide-related peptide activation of NPR-1, *Nat. Neurosci.* **6**, 1178–1185.

Sharma-Kishore, R., White, J.G., Southgate, E., Podbilewicz, B. (1999) Formation of the vulva in *Caenorhabditis elegans*: a paradigm for organogenesis, *Development* **126**, 691–699.

Shen, K., Fetter, R.D., Bargmann, C.I. (2004) Synaptic specificity is generated by the synaptic guidepost protein SYG-2 and its receptor, SYG-1, *Cell* **116**, 869–881.

Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H.A., Fire, A. (2001) On the role of RNA

amplification in dsRNA-triggered gene silencing, *Cell* **107**, 465–476.

Simmer, F., Tijsterman, M., Parrish, S., Koushika, S.P., Nonet, M.L., Fire, A., Ahringer, J., Plasterk, R.H.A. (2002) Loss of the putative RNA-directed polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi, *Curr. Biol.* **12**, 1317–1319.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., Coulson, A., D'Eustachio, P., Fitch, D.H.A., Fulton, L.A., Fulton, R.E. et al. (2003) The genome of *Caenorhabditis briggsae*: a platform for comparative genomics, *PLoS Biol.* **1**, 166–192.

Syntichaki, P., Xu, K., Driscoll, M., Tavernarakis, N. (2002) Specific aspartyl and calpain proteases are required for neurodegeneration in *C. elegans*, *Nature* **419**, 939–944.

Timmons, L., Court, D.L., Fire, A. (2001) Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*, *Gene* **263**, 103–112.

Wang, J., Kim, S.K. (2003) Global analysis of dauer gene expression in *Caenorhabditis elegans*, *Development* **130**, 1621–1634.

Wang, X., Chamberlin, H.M. (2004) Evolutionary innovation of the excretory system in *Caenorhabditis elegans*, *Nat. Genet.* **36**, 231–232.

Winston, W.M., Molodowitch, C., Hunter, C.P. (2002) Systemic RNAi in *C. elegans* requires the putative transmembrane protein SID-1, *Science* **295**, 2456–2459.

Wittenburg, N., Elmer, S., Lakowski, B., Röhrig, S., Rudolph, C., Baumeister, R. (2000) Presenilin is required for proper morphology and function of neurons in *C. elegans*, *Nature* **406**, 306–309.

Xu, K., Tavernarakis, N., Driscoll, M. (2001) Necrotic cell death in *C. elegans* requires the function of calreticulin and regulators of Ca^{2+} release from the endoplasmic reticulum, *Neuron* **31**, 957–971.

Zhang, Y., Ma, C., Delohery, T., Nasipak, B., Foat, B.C., Bounoutas, A., Bussemaker, H.J., Kim, S.K., Chalfie, M. (2002) Identification of genes expressed in *C. elegans* touch receptor neurons, *Nature* **418**, 331–335.

Neoplastic Disease Diagnosis, DNA in

Mukesh Verma, Jacob Kagan, and Sudhir Srivastava
Cancer Biomarkers Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD, USA

1 Introduction: Current Status in Diagnostics for Early Cancer Detection 95

2 Examples of Diagnostic Markers for Preneoplasia 96

3 Application of Biomarkers in Cancer Detection 99
3.1 Case for Early Detection 100
3.1.1 Detection Leads to Better Surveillance 100
3.1.2 Early Detection Leads to Improved Survival 101

4 Technology Advancement in Diagnostics 102
4.1 Pattern Analysis as a Marker 102
4.1.1 Surface-enhanced Laser Desorption/Ionization Mass Spectrometry (SELDI) 103
4.1.2 Two-dimensional Gel Electrophoresis 104
4.1.3 Immunoassays 104
4.1.4 Protein Arrays 104
4.1.5 Mass Spectrometry 105
4.1.6 Matrix-assisted Laser Desorption/Ionization (MALDI)-time-of-flight (TOF) Mass Spectrometry (MS) 105
4.1.7 Laser Capture Microdissection 106
4.2 Target Specific Detection 106
4.2.1 Tumor Antigen Markers: Techniques to Detect Autoantibodies 106
4.2.2 Techniques for Detecting Glycoproteins 106
4.2.3 Activity-based Protein Profiling 108
4.3 Other Technologies 109
4.3.1 Genomics 109
4.3.2 Epigenomics 109
4.3.3 RNA-based Assays 111

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

4.3.4 Posttranslational Modifications 111
4.3.5 Imaging 112

5 Challenges in Cancer Diagnostics: Validation of Biomarkers 113
5.1 Biomarkers as Surrogate End Points 113
5.2 Biomarker Validation 114

6 Summary 115

Bibliography 116
Books and Reviews 116
Primary Literature 116

Keywords

Biomarkers
A biological marker, or biomarker, is a phenotypic parameter (e.g. a substance, structure, or process) that is measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention. In cancer research and detection, a biomarker refers to a substance or process that is indicative of the presence of cancer in the body. It might be either a molecule secreted by a malignancy itself, or it can be a specific response of the body to the presence of cancer.

Cluster Analysis
A technique for grouping a collection of objects into subsets or clusters such that those within each cluster are more closely related to one another than are objects assigned to different clusters.

DNA Hypermethylation
An increased level of DNA methylation in a DNA sample at either an individual CpG dinucleotide or a group of dinucleotides relative to a reference DNA sample, usually derived from a normal tissue.

Early Detection
Detection of the disease (cancer) before it metastasizes and there are options of preventing it.

Epigenetics
Epigenetics is the study of processes that establish metastable (i.e. somatically heritable) states of gene expression without altering the DNA sequence. In large part, the field of epigenetics addresses processes that affect *DNA methylation* and/or *chromatin structure* to establish such metastable, but somatically heritable, states of

gene expression. Thus, epigenetic phenomena can be defined as changes in the genetic material (specifically the genomic DNA and chromatin) that alter the regulation of gene expression in a manner that is somatically heritable (and sometimes heritable through the germline) but that is nonmutational, that is, does not involve changes in the DNA sequence.

Mass Spectrometry
A field that, in its biological applications, uses sophisticated analytical devices to determine the precise molecular weights (mass) of proteins and nucleic acids, as well as the amino acid sequence of protein molecules.

Proteomics
The characterization and quantification of proteins and protein systems. Proteomics methods allow for the comparison of patterns of proteins isolated from bodily fluids or cells, in normal versus diseased subjects.

SELDI-TOF
Surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) is a method for profiling a population of proteins in a sample according to the size and net electrical charge of the individual proteins. The position of an individual protein in the spectrum produced corresponds to its ''time of flight'' because the small proteins fly faster and the large proteins fly more slowly.

Sensitivity and Specificity
Sensitivity is the true-positive test results (individuals with disease who test positive for that disease) expressed as a percentage of all tested individuals who have that disease (total of true positives and false negatives). When applied to cancer, sensitivity is the chance that a person with cancer has a positive test. (http://www.nci.nih.gov/cancerinfo/pdq/screening/overview). In the context of biomarker testing, the *clinical sensitivity* of a biomarker refers to the proportion of case subjects (individuals with confirmed disease) who test positive for the biomarker assay, *Analytical Sensitivity* refers to the sensitivity of an assay that is used to detect a biomarker of disease, that is, the ability of the assay to detect a biomarker given that it is present. *Clinical Sensitivity* is a composite of the biomarker prevalence in a tumor, the efficiency of transfer of the marker to the bodily fluid or tissue being tested, and the analytical sensitivity of the assay.
Specificity is the true-negative results (tested individuals without disease who test negative for that disease) expressed as a percentage of all tested individuals who do not have that disease (total of true negatives and false positives). When applied to cancer, specificity is the chance that a person without cancer has a negative test (http://www.nci.nih.gov/cancerinfo/pdq/screening/overview). In the context of biomarker testing, *specificity* refers to the proportion of control subjects (individuals without disease) who test negative for the biomarker assay.

Abbreviations

ABP:	Activity-based Protein Profiling
APC:	Adenomatous Polyposis Coli
ATM:	Ataxia Telangiectasia
CAD:	Computer-assisted Diagnosis
CARET:	Carotene and Retinol Efficiency Trial
CEA:	Carcinoma Embryonic Antigen
CyP-A:	Cyclophilin A
DCC:	Deleted in Colon Cancer
2DGE:	Two-dimensional Gel Electrophoresis
DHPLC:	Denaturing High-pressure Liquid Chromatography
Gd-DTPA:	Gadolinium-diethyl-enetriamine Penta Acetic Acid
HDAC:	Histone Deacetylase
HBV:	Hepatitis B Virus
HCV:	Hepatitis C Virus
HCC:	Hepatocellular Carcinoma
ILK:	Integrin-linked Kinase
IP:	Imaging Probe
IRSH:	Invasiveness Related Serine Hydrolase
HNPCC:	Hereditary Nonpolyposis Colorectal Cancer
IP:	Imaging Probe
LCM:	Laser Capture Microdissection
LCMS:	Liquid Chromatography Mass Spectrometry
LOH:	Loss of Heterozygosity
MCA:	Methyl Cytosine Amplification
MCC:	Mutated in Colorectal Cancer
MALDI:	Matrix-associated Laser Desorption/Ionization
MIF:	Macrophage Migration Inhibitory Factor
MMP9:	Matrix Metalloproteinase
MMR:	Mismatch Repair
MRI:	Magnetic Resonance Imaging
MS:	Mass Spectrometry
MTAP:	Multitarget DNA Assay Panel
N-nonyl-DGJ:	*N*-nonyl-deoxygalactojirimycin
NSCL:	Non Small Cell Lung Carcinoma
PCNA:	Proliferating Cell Nuclear Antigen
PSA:	Prostate Specific Antigen
PTK:	Protein Tyrosine Kinase
PTM:	Posttranslational Modifications
RLGS:	Restriction Landmark Genomic Screening
RSR:	Relative Survival Rate
RLGS:	Restriction Landmark Genomic Screening
SAGE:	Serial Analysis of Gene Expression

SEBs:	Surrogate Endpoint Biomarkers
SEER	Surveillance, Epidemiology, and End Results
SELDITOF:	Surface-enhanced Laser Desorption Ionization Time of Flight
SH:	Serine Hydrolase
TMS:	Tandem Mass Spectrometer
TSA:	Trichostatin A
WECARE:	Women's Environment, Cancer and Radiation Epidemiology.

Cancer development involves several stages: initiation, development, and progression. Biomarkers can be used to detect cancer in different stages. The desirable property and utility of a biomarker lies in its ability to provide an early indication of disease progression. Biomarkers should be easy to detect; measurable across populations; and suitable for detection at an early stage, identification of high-risk individuals, early detection of recurrence, and as an intermediate endpoint in chemoprevention. Recent technological advances have helped develop noninvasive, sensitive, and specific biomarkers in small clinical samples to detect preneoplasia.

1 Introduction: Current Status in Diagnostics for Early Cancer Detection

Cancer is a heterogeneous disease in most respects, including its cellularity, different genetic alterations, and diverse clinical behaviors. Alterations in gene sequence, expression levels, and protein structure and function have been associated with every type of cancer. These changes can be used to detect cancer, determine prognosis, and monitor disease progression and therapeutic response. One of the major challenges is to identify these molecular changes and incorporate the information into the clinical setting. Perhaps a major challenge in cancer medicine is to recognize the "black sheep of the family"; that is, those tumors in the family that are clinically distinct but blend in by histology. In breast cancer, for example, infiltrating breast carcinoma appears homogeneous histologically but heterogeneous clinically (at least five distinct tumor subtypes were identified by biomarker and gene expression profiling).

Several cancers originate in the epithelium that lines the internal surfaces of organs throughout the body. Although these are readily treatable provided they are diagnosed in preinvasive stages, early lesions often evade detection before they become invasive. Specifically, the nuclei become enlarged (from 5–10 μm to 20 μm), crowed and hyperchromatic, as detected by histological examination. However, light-scattering microscopy now offers a biopsy-free means to measure the size distribution and chromatin content of epithelial cell nuclei as an indicator of preinvasive neoplasia.

Detection of mutator phenotype may provide a snapshot of subsequent changes. Normal mutation rates are insufficient to account for the multiple mutations found in different cancers. Cancers must exhibit a mutator phenotype early during evolution. The mutator phenotype

hypothesis was originally postulated for mutations in genes involved in DNA replication (especially fidelity in DNA replication) and DNA repair. Currently, other genes that affect chromosome segregation, damage surveillance (e.g. checkpoint control) and cellular response (e.g. apoptosis) are also included. The mutator phenotype arising from mutations in genetic stability genes can have diverse manifestations, such as point mutations, microsatellite instability, and loss of heterozygosity.

During the long time interval associated with cancer development, cancer cells acquire the capacity to divide, invade, and metastasize. According to mutator hypothesis, these phenotypes result from mutations in genes that maintain gene stability in normal cells. Thus, mutations in several genes (and whole genome) will be affected because of mutations in the genetic stability genes. Some of the resulting mutations will exhibit selective advantages, allowing the mutated cells to expand and achieve clonal dominance.

It has been proposed that genetic instability in human cancers can be divided into two categories; microsatellite instability, which is usually equated with DNA polymerase errors, and chromosomal instability, which can result from errors in chromosome partitioning. In the case of endometrial and colorectal cancers, most tumors exhibit either chromosomal instability or microsatellite instability, but not both. It is emphasized here that chromosomal instability and microsatellite instability are well characterized as a result of facile detection by current methods, PCR and karyotyping, respectively. Conversely, detection of random mutations, particularly point mutations, is more difficult, and this has limited the search for evidence of a mutator phenotype at the single nucleotide level.

It remains to be determined whether the acquisition of a mutator phenotype underlies tumor progression and when the enhancement in mutation rate occurs during the life of tumors. In order for a mutator phenotype to be necessary for tumor progression, it would have to occur early and result in the accumulation of large numbers of random mutations; the existence of multiple mutations in cancer cells has important implications in stratification, tumor evolution, and treatment. A functional classification of cancer markers is presented in Table 1.

The application of biomarkers in clinical practice has been limited to single entity quantitation and to small panels of tests that identify changes in proteins, the presence or absence of which signifies an important event. The ongoing discovery of new biomarkers lends to the belief that no single marker will provide sufficient information for detection and diagnosis. Therefore, a multiparameter, multimarker approach to biomarker development and validation is needed. Efforts should be made to take advantage of the number of tumor samples now available in tissue banks to identify new molecular markers and to assess existing ones more fully. A list of biomarkers for cancer diagnosis is presented in Table 2.

2 Examples of Diagnostic Markers for Preneoplasia

Two major classes of genes – oncogenes and suppressor genes – are involved in carcinogenesis. Oncogenes are activated or deactivated genes whose products normally promote cell growth. Products

Tab. 1 Functional classification of cancer diagnostic markers.

Function	*Marker*	*Tumor type*
Angiogenesis	Thrombospondin-1 (TSP-1), TIMP-3	Several
Apoptosis		
Cell cycle	p15, p16, p73, DAP kinase	Acute myeloid, lymphoid leukemias, lymphomas, lung
DNA repair	BRCA1, glutathione S-transferase, hMLH1, LKB1, methylguanine methyl transferase (MGMT)	Breast, colon, gastric, endometrial
Invasive/metastasis suppressor	CD-44, E-cadherin, TIMP-3	Brain, colon, renal
Oncogenes		
Tumor antigen	Thrombospondin-1 (TSP-1), TIMP-3	Several
Tumor suppressor	APC, AR, ARF/INK4A, HIC1, p15, p16, p73, p53, RARβ, Rb, VHL	Several

include peptide growth factors, growth factor receptors, signal transduction factors, tyrosine kinase, and transcription regulatory proteins. Several oncogenes have been mapped within the human karyotype. Tumor suppressor genes normally regulate cell growth. These genes typically produce nuclear regulatory proteins, which interact with other regulatory proteins to alter their activity. The frequent involvement of a limited number of oncogenes, for example, *c-myc* and *ras*, in a wide variety of tumor scenarios suggests, nevertheless, that the number of the most critical genes needed for the transformation process may be limited. This suggests that we might be able to develop common approaches to detect these cancers.

Genes not related to the oncogenes or suppressor genes have been implicated in carcinogenesis. These advances have fueled optimism that abnormal genes, whether acquired or inherited, may be used to assess the risk of cancer or to detect existing cancer early even in its preclinical stages. Knowledge of the biology of tumor progression, therefore, allows us to identify specific tests useful for early detection or screening. Molecular probes, for instance, could detect altered DNA shed into the feces (colorectal cancer), into the sputum (lung cancer), and in exfoliated cells in bladder washings (bladder cancer). Correlation of the molecular alterations with demographic data, risk factors, environmental exposure, family history, and dietary history may provide important information on the etiology of cancer.

Molecular genetic alterations could also contribute toward the assessment of risk. Risk assessment is the search for risk factors that provide the earliest evidence for the risk of cancer in persons not diagnosed with the disease. Biomarkers that are predictive of risk usually trigger more aggressive interventions and surveillance. The earliest risk factors are probably inherited genetic defects, which are well demonstrated in the case of colorectal cancer. Adenomatous polyposis coli (*APC*) and hereditary nonpolyposis colorectal cancer (HNPCC) are associated with germline mutations in the *APC* gene and mismatch repair genes, respectively. Large bowel cancers often arising at a young age are found

Tab. 2 Tumor types and biomarkers. This is a partial list of biomarkers that are abnormally regulated in preneoplasia and have the potential to be used in cancer diagnostics.

Tumor type	*Gene*
Bone	Cathepsin, urokinase plasminogen activator
Breast	BRCA1, BS106, BU101, CDH1, Cyclin D1, DAPK, GSTP1, Her-2/neu, p16, mammaglobin, TIMP-3
Brain	MGMT, p16, $p14^{ARF}$, TIMP-3
Bladder	APC, DAPK, p16
Cervical	GSTP1, FHIT, MGMT, p16, RARbeta
Colon	APC, BAT-26, Beta catenin, DAPK, DCC, galactin-3, hMLH1, K-ras, L-DNA, MGMT, MCC, p16, $p14^{ARF}$, TIMP-3
Esophagus	APC, CDH1, GSTP1, $p14^{ARF}$, p16
Glioma	MGMT
Head and neck	DAPK, MGMT, p16, PAI-2
Kidney	APC, CDH1, DAPK1, GSTP1, MGMT, p16, $p14^{ARF}$, TIMP-3
Leukemia	CALC1, CDH1, DAPK1, MGMT, p15, p73
Liver	AFP, APC, complement-3 fragment, DCP, GP73, GSTP1, p16
Lymphoma	DAPK, MGMT, p16, p15, p73
Lung	Annexin I and II, CyP-A, DAPK, E-cadherin, GSTP1, FHIT, MIF, MGMT, p16, $p14^{ARF}$, p38, RARbeta, RASSF1A, TIMP-3
Oral	DCC, CADH1
Ovary	Alpha heptaglobin, BRCA1, CA125, DAPK, kallikrein-6, LPA, MMP, osteopontin, prostasin, p16
Pancreas	APC, CDKN2A, GSTP1, MGMT, p16
Prostate	B23, GSTP1, kallikrein-2, PSA, PSMA, p27
Stomach	APC, DAPK, HMLH1, MGMT, p16
Uterus	hMLH1, p16, $p14^{ARF}$

Note: AFP: alpha fetoprotein; APC: adenomatous polyposis coli; CALC1: calcitonin; CDH1: E-cathedrin; CyP-A: cyclophilin A; DAPK: death-associated protein kinase; DCC: deleted in colon cancer; DCP: des-gamma carboxy-prothrombin; FHIT: fragile histidine triad; GP73: glycoprotein 73; GSTP1: glutathione S-transferase P1; HMLH1: human mut L homolog; L-DNA: long DNA; LPA: lipophosphotidic acid; MCC: mutated in colorectal cancer; MGMT: O6-methylguanine-DNA methyltransferase; MIF: macrophage migration inhibitory factor; MMP: metalloproteinase; PAI-2: plasminogen activator inhibitor-2; PSA: prostate-specific antigen; PSMA: prostate-specific membrane antigen; RARbeta: retinoic acid receptor beta; RASSF1A: RAS association domain family protein 1A.

in almost 100% of individuals who inherit *APC* gene mutations.

Few diagnostic markers are available currently. Examples of such markers include prostate-specific antigen (PSA) for prostate cancer and serum carcinoma embryonic antigen (CEA) for colorectal cancer. About one-third of patients with an elevated PSA level undergo further unnecessary medical procedures, however, because they do not have a malignant form of prostate cancer. New molecular markers are needed that can identify patients who are at high risk for developing life-threatening cancers and patients who harbor aggressive tumors that require immediate treatment. A comprehensive list of biomarkers with potential in cancer diagnostics is presented in Table 2.

In colon cancer, for example, delineation of various stages during tumor progression offers a window of opportunity

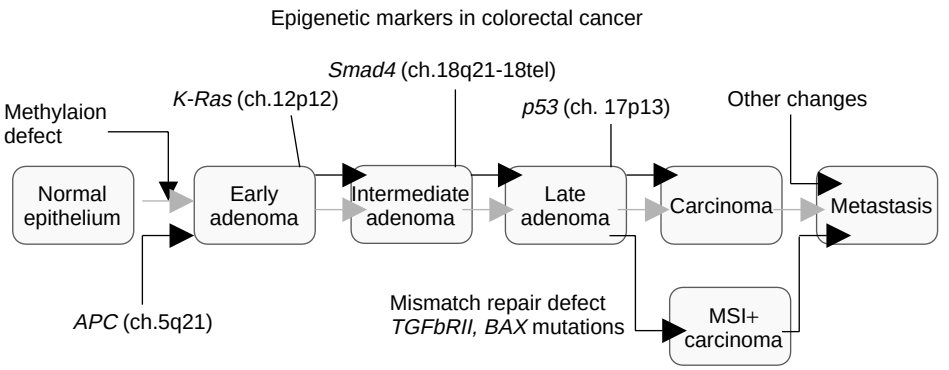

Fig. 1 Epigenetic markers in colon cancer. A general model of colorectal cancer is presented. A number of epigenetic markers express during early stages of the cancer development. These markers can be used for cancer diagnostics. Since epigenetic changes are reversible, these markers could be the potential therapeutic targets also.

to intervene in the process by detecting stage-specific molecular changes (such as changes in gene expression and accumulation of mutations) (Fig. 1). To date we know of mutations and loss of heterozygosity (LOH) which affect oncogenes such as *K-ras,* and several tumor suppressor genes including *p53, MCC* (mutated in colorectal cancer), *APC,* and *DCC* (deleted in colon cancer).

3 Application of Biomarkers in Cancer Detection

DNA released from cells that regularly slough into the stool can be used to identify cancer or precancerous polyps in the colon. A multi-target DNA assay panel (*MTAP*) was developed, which shows a remarkably higher sensitivity and specificity than traditional fecal blood tests. The *MTAP* targets point mutations at any of 15 mutational hot spots on *K-ras, p53,* and *APC* genes; Bat-26, a microsatellite instability marker and L-DNA have been identified. Using this test a sensitivity of 91% and a specificity of 100% were observed. Thus, an assay of altered DNA from exfoliated cells holds promise as a stool-screening approach for colorectal neoplasia.

Recently, Campa et al. identified two differentially expressed proteins, macrophage migration inhibitory factor (MIF) and cyclophilin A (CyP-A), in lung cancer specimens using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry(MALDI-TOF MS). Cyp-A is a member of the immunophilin family of proteins, typically studied for their binding of various immunosuppressive drugs, most notably cyclosporine A, and their role in cellular signaling pathways. This is the first demonstration of the utilization of MALDI-TOF MS in screening lung fingerprints and identifying novel markers.

Because of the multifactorial nature of cancer, it is very likely that a combination of several markers will be necessary to effectively detect and diagnose cancer. To look for such "fingerprints" of cancer, it will require not only high-throughput genomic or proteomic profiling but also sophisticated bioinformatics

tools for complex data analysis and pattern recognition. Taking advantage of the recent development in surface-enhanced laser deionization/ionization (SELDI) and of the ProteinChip technology, Petricoin's group was able to simultaneously analyze the protein profiles of 169 serum samples from patients with or without breast cancer. The software package ProPeak allows evaluation of each mass peak according to its collective contribution toward the maximal separation of cancer patients from the noncancer controls. These two advances led to the identification of three discriminatory biomarkers that, in combination, achieved both high sensitivity (93%) and high specificity (91%) in detecting breast cancer patients from the noncancer controls. Using the same technology, Petricoin's group has identified differentially expressed protein profiles specific for early stages of ovarian and prostate cancer with very high sensitivity and specificity.

3.1 Case for Early Detection

The Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute (NCI) is an authoritative source of information on cancer incidence and survival in the United States (www.seer.cancer.gov). The SEER program currently collects and publishes cancer incidence and survival data from 11 population-based cancer registries and three supplemental registries covering approximately 14% of the US population. The SEER registries routinely collect data on patient demographics, primary tumor site, morphology, stage at diagnosis, first course of treatment, and follow-up for vital status. The information collected in the program above is very useful in biomarker discovery and validation.

The Gail Model, developed at the NCI, is a statistical tool to help estimate a woman's risk of breast cancer based on a number of predisposing factors. A study of the reliability of mammographic densities was completed recently, and this characteristic has been found to be reliable enough to be formally included in the model.

3.1.1 Detection Leads to Better Surveillance

We have described below, as an example, that the ovarian cancer patients with *BRCA* mutations have better survival than noncarriers. Women who inherit mutations to the *BRCA1* or *BRCA2* gene have a greater risk of developing ovarian cancer. However, until recently, researchers could not predict whether women with these mutations would fare better, the same, or worse than other women with ovarian cancer. This issue was examined via the medical records of almost 900 newly diagnosed ovarian cancer patients with inherited *BRCA* mutations and their progress was followed for five years. On average, the patients with *BRCA* mutations survived about 20 months longer than women without the mutations. This survival advantage was not due to earlier detection, but seems to reflect a difference in the pattern of disease in women with and without *BRCA* mutations. The mechanism of this phenomenon is not known. Further follow-up in this study should reveal whether the longer survival of *BRCA* mutation carriers is a long-term phenomenon or is limited to the first few years of the disease. Thus, *BRCA* mutation has been used as a useful biomarker in ovarian cancer studies.

In a separate multicenter study, mutations in the ataxia telangiectasia (ATM) gene was used to screen a large population for breast cancer. Ataxia telangiectasia is an autosomal recessive disorder

characterized by a progressive cerebellar ataxia, occulo-cutaneous telangiectasias, immune deficiency, chromosomal instability, hypersensitivity to ionizing radiation, and an increased incidence of cancer. ATM shares significant sequence similarity with a family of large serine–threonine kinases that play a crucial role in cell-cycle regulation. ATM functions as a protein kinase, phosphorylating a large and growing list of substrates, including BRCA1, in response to the presence of DNA double-stranded breaks. Although ATM carriers do not display any specific phenotype, they had a significant increased risk of death due to breast cancer. Generally, ATM carriers are more sensitive to the lethal effects of the ionization radiations. To investigate the relationship between breast cancer, radiation exposure, and genetic variation in the ATM gene, the WECARE (Women's Environment, Cancer and Radiation Epidemiology) study was conducted where multiple centers (2100 subjects) participated. Results indicated novel mutations, detected by denaturing high-pressure liquid chromatography (DHPLC), in the ATM gene identified both in patients with antaxia telangiectasia and in patients with unilateral or bilateral breast cancer. These results point to the importance of screening techniques with the sensitivity to detect both missense and truncating mutations. In addition, the importance of standardizing the lab conditions cannot be minimized in a multicenter study of this nature, and in the type of quality control assessments described in this study, which will reduce bias and laboratory errors.

3.1.2 Early Detection Leads to Improved Survival

An analysis of temporal trends for esophageal cancer patients diagnosed in 1961–1996 in Sweden showed improvement in relative survival rates (RSRs), but data were not available on stage at diagnosis or on cancer-directed treatment. In another study, the temporal changes in RSRs by histology for esophageal cancer patients in the United States were reported. In the NCI's SEER program, data were obtained from nine locations. Patients with esophageal cancer as the first or only reportable cancer were identified, excluding those ascertained only by death certificate or autopsy. Using a computer program that includes mortality rates for the general US population (by age, sex, and race), RSRs at 1, 3, and 5 years after diagnosis in 1975–1979, 1980–1984, 1985–1989, 1990–1994, and 1995–1998 were obtained for squamous cell carcinoma and adenocarcinoma of the esophagus (defined by ICD-O-2 morphology codes). Confidence limits on RSRs were estimated as $\pm$ twice the SE ($p = 0.05$). RSRs increased over time for both squamous cell and adenocarcinoma, beginning earlier than in Sweden. The overall increases from 1975–1979 to 1995–1998 were similar for the two histologic groups, but the larger sample sizes for squamous cell carcinoma in the earlier years resulted in statistical significance for more comparisons than for adenocarcinoma. Age-standardized incidence rates for adenocarcinoma of the esophagus in the SEER program have surpassed the declining rates for squamous cell carcinoma. For squamous cell carcinoma patients, RSRs were statistically, significantly higher in 1980–1984 versus 1975–1979 and in 1985–1989 versus 1980–1984 (except for the 5-year RSR); thereafter, increases were smaller. RSRs were slightly higher for adenocarcinoma than for squamous cell carcinoma, but differences were not statistically significant.

Another example in this area is that of pancreatic cancer. It is known that about 10% of all cancers of pancreatic cancer are hereditary. However, although mutations have been found in the tumors of patients with nonhereditary pancreatic cancers, the major genes responsible for inherited cases are yet to be discovered. Investigators searched for suspicious mutations in the *BRCA2* gene of pancreatic patients who had at least three relatives with this disease. They discovered mutations likely to be involved in tumor development in 17% of patients in this study. Further research is needed to define the risk of pancreatic cancer in patients with these mutations better. Additional genetic as well as environmental factors may influence this risk.

More information is needed on a variety of topics relevant to cancer diagnostic biomarkers. Identification of risk factors and potential confounders for different markers is another. Both sources and extent of variability (intra and interindividual), which can have both genetic and environmental sources, and laboratory variability, which has both random and systematic components, are important. Understanding the natural history of biomarkers, analogous to the natural history of disease, will help us to use biomarkers effectively in epidemiological studies.

4 Technology Advancement in Diagnostics

The last 20 years have seen a molecular revolution influencing all aspects of oncology: from the *in vitro* modeling of genetic and epigenetic mechanisms of oncogenesis and tumor propagation to *in vivo* animal models of malignancy and from the clinical stratification of patients following "biomarkers" to the actual targeting of molecular defects using drugs specifically designed for that purpose. The transgression of novel findings in the field of molecular oncology can take virtually years from bench to bedside. *BRCA1* and *BRCA2*, for example, were discovered in the early eighties but their clinical implication could only be achieved several years later.

In the past few years, a number of technologies and assays have emerged that can be used for cancer detection. In the following section, we have discussed those techniques that are being used or have the potential to be used in a clinical set up for early cancer detection and risk assessment. The major categories of these technologies are genomics, epigenomics, proteomics, gene expression profiling (RNA-based techniques), and imaging.

Two major categories in the technology advancement in diagnosis are:

1. Patterns Analysis as a Marker
2. Target Specific Detection.

4.1 Pattern Analysis as a Marker

Techniques to identify alterations in protein structure or function that are associated with cancer include antibody-based detection methods, measurement of enzymatic activity, two-dimensional gel analysis (2DGE), and surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI TOF MS). By combining features of systemic and local disease markers into reference standards for the diagnosis of new disease entities, functional proteomics has the potential to redefine illness. A brief description of selected proteomic techniques and their

implications for the molecular diagnosis of cancer is given in the following section.

4.1.1 Surface-enhanced Laser Desorption/Ionization Mass Spectrometry (SELDI)

Implication of affinity MS to the identification and measurement of cancer-associated biomarkers was achieved by Ciphergen Biosystems, Inc. (Fremont, CA) to develop SELDI ProteinChip® MS technology. In protein analysis, SELDI has several advantages over other existing technologies such as liquid chromatography-MS (LC-MS), 2-D gel electrophoresis coupled-MS (2DGE-MS), enzyme-linked immunosorbent assay ELISA, and fluorescent-based binding assays for high-throughput screening because of its versatility, ease of use, speed, and cost. SELDI-MS is rapid, reproducible, highly sensitive (i.e. its detection limit is in the femtomolar range), and it is readily adaptable to a diagnostic format. Its major limitations are the difficulty in identifying proteins detected in mass spectra and the limited ability to detect most proteins in complex samples. SELDI-TOF-MS has been utilized for discovering proteomic patterns in blood and offers a novel approach for early cancer detection. Proteomic signatures emanate from the cell as cancer develops, and the emanating proteins generate biomarker profiles that end up in the effluent blood surrounding the cells. The concept is that the collection of proteins flowing into the tissue can be modified as it leaves the tissue because of enzymatic events from the cancer cell or the host, or proteins could be added to the perfusing blood as biomarkers. The serum proteome is a complex of thousands of proteins and peptides and the smaller posttranslationally modified proteins that are cleaved, glycosylated, phosphorylated, or tagged with a lipid, provide a wealth of diagnostic information about the pathologic state of this tissue.

In a SELDI pattern, the cancer diagnostic proteins are identified without knowing their function. One challenge of this technology is to identify the proteins derived in the cancer cell from those derived from the host cell. This is a low-resolution system that can process one drop of raw serum on a microchip that binds proteins. A profile of differentially expressed proteins is generated. After validating the protein patterns, an algorithm is developed to find the subset of proteins that are diagnostic of the cancer compared to the noncancer state. The genetic algorithm takes two sets of training data and finds the subset pattern that can detect the difference between the two training populations. That pattern is mapped in *n*-dimensional space as a series of clusters and an unknown sample can thus be compared to the cluster to determine if it is a cancer or normal pattern.

SELDI analysis of serum from prostate, ovarian, and bladder cancer patients and a comparison with normal serum identified distinct disease-specific protein profiling. Results have been encouraging for identifying cancer in men with PSA levels between 4 and 10 mg ml^{-1}, which is a level that has no specificity for that test. The use of the multiparametric proteomic patterns application method was able to identify men who subsequently developed prostate cancer. Results of investigations for ovarian cancer showed that the algorithm could identify patterns that were subsequently found to be cancerous. The specificity and sensitivity were very high in these studies.

The hope is that if cancer researchers can shift the diagnosis of ovarian cancer from stage III, where it currently exists, to stage I, many lives will be saved. Moving to a clinical testing situation will demand a more advanced and reliable system than

is now available. Therefore, current efforts are being made to move from the SELDI-TOF-MS to the QSTAR technology, which can offer much higher resolution, quicker response times, and assessment of a wider molecular weight range of proteins. In preliminary tests with the QSTAR, results have been very encouraging.

4.1.2 Two-dimensional Gel Electrophoresis

Two-dimensional (2-D) gel electrophoresis is the most widely used tool for separating proteins in expression proteomics. In 2-D gel electrophoresis, proteins are separated first in one dimension on the basis of isoelectric point and then in a second dimension on the basis of molecular mass. In some cases, 2-D gel electrophoresis may evaluate whole cell or tissue protein extracts. The pH gradient employed in the first dimension can also vary, depending on the application. Using very narrow, immobilized pH gradients for the first-dimension separation increases resolving power and can help detect low-abundance proteins. Radioactive or fluorescent labeling or silver staining offers visualization of hundreds of proteins. Differences between the samples can be compared and relative quantities determined by quantifying the ratios of spot intensities in independent 2-D gels. MALDI-TOF mass spectrometry (MS) allows the analysis and identification of very small amounts of protein isolated from the gel. These advances have combined to make 2-D electrophoresis a more attractive option for the analysis of complex protein mixtures.

4.1.3 Immunoassays

Immunoassays are the fastest-growing technologies for the analysis of biomolecules and are used widely for diagnostic purposes. The development of a competitive binding assay using radioisotopes and enzyme-labeled probes later was a major milestone in antibody-based immunoassays. These developments paved the way for an enormous expansion in application of these technologies, particularly in biomedical research and in clinical chemistry. These assays are limited, however, to the analysis of a few thousand assays per day, even with the adaptation of the microtiter plate format. Furthermore, the presence of only known analytes that can be screened for becomes a serious bias in proteome analysis, where unknown gene products are significant. A new technology, gene and protein microarrays, is further advancing the ability to perform tens of thousands of assays in parallel. The first analytical breakthrough came with the use of oligonucleotides and cDNA microarrays; however, the analysis of the proteome requires a different approach and the development of protein microarrays using protein chips.

4.1.4 Protein Arrays

Protein arrays are composed of multiple recombinant proteins that are arranged in an orderly fashion on a small surface. The protein and antibody arrays are mostly miniaturized immunoassays for multiple proteins and antibodies, respectively. They also can be used to detect protein–protein, protein–ligand, protein–DNA, and protein–RNA interactions, however. Their production could be automated using either pin-based or liquid microdispenser robots. Protein and antibody microarrays will be the approach of choice to close the information gap between genomics and proteomics in the development of new markers for the early detection, diagnosis, and classification of cancer. Ordered microarrays of proteins, protein ligands, or antibodies directed at well-defined

antigens are the best platform for the high-throughput, precise analysis of quantitative differences between the proteomic maps of healthy and cancer-stricken individuals. This emerging technology will be applied to identify differences at the cellular level, to compare normal cells and cancer cells from the same organ, and to compare body fluids such as plasma and urine from healthy donors and cancer patients. Advantages of this assay are its reproducibility, scalability, and quantitation.

4.1.5 Mass Spectrometry

Direct qualitative and quantitative comparisons of complex protein mixtures have been greatly facilitated by the development of a high-throughput mass spectrometry (MS). This method employs a chemical group or label made in two isotopic forms: heavy and light. These labels, isotope-coded affinity tags, combine with all of the cysteine residues in a protein mixture. The heavy reagent is added to one sample and the light reagent is added to another sample. These samples are combined, and the resulting protein digestion and peptides are analyzed via MS. The isotopic substitutions do not affect the behavior of the peptides during separation. Thus, peptides derived from the two different samples enter the spectrometer at the same time. The MS measures the relative abundance of heavy and light peptide forms in each sample and identifies each peptide by generating and analyzing the peptide fingerprints. In this manner, a global view of the protein abundance in cells or tissues in two different states of health emerges.

4.1.6 Matrix-assisted Laser Desorption/Ionization (MALDI)-time-of-flight (TOF) Mass Spectrometry (MS)

Matrix-assisted laser desorption/ionization (MALDI)-time-of-flight (MALDI-TOF) uses a nitrogen laser. MALDI-TOF samples are irradiated after coating the sample with an energy-absorbing matrix. Molecules from the surface are dissolved and accelerated, and their TOF is measured to determine molecular weight, mass, and electrical charge.

Liquid chromatograph (LC)–tandem mass spectrometer (TMS), which is an tandem mass spectrometer (MS-MS) instrument combined with LC to measure molecular species disassociated from a matrix, is utilized in determination of sequences of peptides.

MALDI-TOF and LC-TMS technologies are applied to develop images from protein samples. There are several advantages of the MALDI technologies: they are a very high-throughput discovery tool, can perform direct analysis of tissues, have high-molecular specificity, can be used for simple protocols, and result in multiple images at different molecular weights. These technologies can be used for protein identification and in the study of PTMs. The highest sensitivity is in the molecular weight range below 50 kDa. Although it currently is possible to analyze proteins up to molecular weight 30 KDa, there is a decline in accuracy and specificity above 50 kDa. There is a requirement of further processing for protein identification. Images developed from these methods can illustrate the distribution of a specific molecular weight molecule (i.e. protein) in a tissue slice. The range of molecular weights using these methods is between 1000 and 100 000 m/z. The MALDI-TOF results are very reproducible. The profiles produced by these methods may be used to follow the progression of a disease, but not to classify the disease. These methods can identify protein patterns to be used as biomarkers or can be used to determine the efficacy of drug therapy with biopsy

material. A finding that may be applicable to both areas is that different proteins exist in different regions of a tumor.

Another improved technology, MALDI-QSTAR (QqTOF) is used to determine concentration of drugs in tissue. The QSTAR is a quadrapole TOF MS-MS machine and can be very specific for very low concentrations of a protein. This technology can help show the positive correlation between LC-MS analysis, which can take days to analyze, and MALDI imaging, which can be completed within minutes, in drug analysis studies.

4.1.7 Laser Capture Microdissection

The most tedious job in proteomics is sample preparation, as some proteins are insoluble in a heterogeneous mixture of cells. An important advance in sample preparation is the development of l**aser capture microdissection(**LCM). Although manual microdissection can achieve good precision, it is time consuming, labor intensive, and requires a high degree of manual dexterity. The LCM system allows one-step procurement of selected human cell populations from a section of complex, heterogeneous tissue. LCM has been used successfully to obtain pure populations of cancer cells from both frozen and paraffin-embedded tissues, stained or unstained, for the molecular analysis of macromolecules. LCM is capable of isolating single cells, making it possible to procure pure populations of neoplastic cells from lesions less than 1 mm in diameter without encroaching upon adjacent nonneoplastic cells. Using this technique, matched normal epithelium, stroma, benign, preneoplastic neoplasia, and cancer cells from the same specimen have been isolated successfully.

4.2 Target Specific Detection

Following technologies are covered under target-specific detection.

4.2.1 Tumor Antigen Markers: Techniques to Detect Autoantibodies

The identification of tumor antigens that elicit an antibody-based immune response may be useful in cancer screening and eventually in immunotherapy. The contribution of PTMs to the detection and identification of cancer biomarkers is significant. One approach is to rely on the immune system to help identify proteins that are synthesized by tumor cells that may be immunogenic. It may be possible to detect antibodies against the proteins that can be used as biomarkers. This is a simple approach that can be very productive.

There are three proteins that are immunogenic in cancer – beta-tubulin3 in neuroblastoma, annexin I and II in lung cancer, and calreticulin in liver cancer – that have been investigated to determine whether they can be used for early detection. For example, there appears to be some promise for using antibodies to annexin in the early detection of lung cancer. Studies on serum from patients in the carotene and retinol efficacy trial (CARET) indicated that 3 of 10 patients who progressed to lung cancer had antibodies to annexin a year before diagnosis.

4.2.2 Techniques for Detecting Glycoproteins

The global study of proteins has many unique difficulties that set it apart from comprehensive studies of genes and transcripts. First, the behavior of proteins is determined by the tertiary structure of the molecule. Any assay based on protein

binding depends on maintaining the native conformation of the protein. This limits the systems that can be used to capture protein targets in affinity-based assays. Second, the detection of low-abundant proteins poses a particular challenge, especially given that the dynamic ranges of proteins in biological systems can reach parts per million or greater. An amplification system analogous to the polymerase chain reaction is yet to be developed for protein studies. Furthermore, the behavior of proteins may or may not be governed quantitatively. Protein regulation often is based not on synthesis and degradation but on reversible modifications such as acetylation, for example. Compounding the difficulty is the fact that RNA splicing can produce highly homologous splice variants that differ in function. Nonetheless, protein science has advanced to the point that some of these hurdles can be overcome.

Protein profiles change because of infection with viruses associated with cancer. For example, individuals chronically infected with the hepatitis B or C virus (HBV, HCV) are at high risk for developing hepatocellular carcinoma (HCC), with disease progression occurring relentlessly over many years. The diagnosis of HCC usually occurs at late stages in the disease, when there are few effective treatment options and the prognosis is very poor. The latency period (up to 20 years), together with clearly identified at-risk populations, provides opportunities for earlier detection that will allow more timely and effective treatment of this devastating cancer. 2DGE was utilized to test the hypothesis that changes in the amount of certain serum polypeptides, or changes in their post-translational modifications, can be used to predict the onset of HCC. Advances in the standardization of 2DGE, together with automation of critical steps such as image analysis, permit the reproducible resolution of thousands of polypeptides per run. Serum polypeptides from individuals at different stages on the disease continuum are being resolved by 2DGE to identify those that change with disease progression. Polypeptides found by this method can be characterized further by MS. In addition, investigators are exploring the potential for changes in the glycan structure of certain polypeptides to serve as a marker for disease progression.

The proteomic approach is expected to facilitate identification of the best indicators and biomarkers of the progression of HCC. Information also may be obtained about the pathobiology of the disease process. It has been demonstrated that *N*-nonyl-deoxynojirimycin (*N*-nonyl-DNJ), a modified glycan, contains an antiviral activity designated to a function other than that of an impact on glycoprocessing. This idea was tested by experiments showing that *N*-nonyl-deoxygalactojirimycin (*N*-nonyl-DGJ), an alkyl derivative of galactose with no impact on glycoprocessing, retains anti-HBV activity. The data suggest that *N*-nonyl-DGJ exerts its antiviral action at a point before the virus is encapsulated and may prevent the proper infectivity of the HBV pregenomic RNA.

Tumor markers shed into the serum have been sought as a means to identify individuals at risk for developing cancer. An alteration in the oligosaccharides associated with glycoproteins is one of the many molecular changes that accompany malignant transformations. The methodology used to assess these alterations, however, has been hampered by several factors, including insensitive techniques that require large sample amounts, poor quantitation, and the lack of accurate information about the relative quantitation of

all of the oligosaccharides that are present. As a result, the systematic and thorough structural analysis of all of the oligosaccharides associated with particular proteins from the tissues of cancer patients has not been carried out. Since specific alterations in oligosaccharides typically are used in histology studies to differentiate between cancerous and noncancerous cells, the potential for oligosaccharides to serve as biomarkers in noninvasive methodologies remains a promising, but unexploited, area. Attempts are being made to explore oligosaccharide structure and disease-state relationships using a new robust technology platform that employs fluorescent high-pressure liquid chromatography (HPLC) analysis of sugars to quantitate every oligosaccharide present (abundance of >0.1% in the total glycan pool) in protein samples. Some of the markers that have been identified to date, such as fucosylated glycoform of alpha fetoprotein, have altered glycosylation patterns in the disease state.

4.2.3 Activity-based Protein Profiling

The activity-based protein profiling (ABP) technology is utilized to directly measure protein functions in complex proteomes. The goal of this approach is to develop activity-based chemical probes (ABPs). Regarding enzymatic activity and the role of PTMs, there are three significant enzyme types that have become the focus of research in this area: proteases (zymogens, endogenous inhibitors), kinases (autoinhibitory domains), and phosphatases (autoinhibitory domains, endogenous inhibitors).

The challenge of ABP is in identifying active sites on specific enzymes that guide their chemical activity and finding chemical reactive groups that label and bind the active sites in complex proteomes. The first enzyme class targeted with this approach was the serine hydrolases (SHs), a large and diverse enzyme family involved in a variety of pathophysiologic processes, including blood clotting, angiogenesis, neural plasticity, peptide hormone processing, T-cell immunity, reproduction, cancer, emphysema, and plant-parasite interactions.

Inhibition of SHs can be achieved by the ABP fluorophosphonate (FP) reagents. A study of breast and melanoma cancer cell lines indicated that it is possible, using cluster analysis, to characterize cell lines based on enzymes by subcellular distribution and tumor of origin. In this study, one cluster each of breast and melanoma cells was identifiable. A third cluster formed that had one breast line, one melanoma line, and one line of unknown origin and shared the property of being invasive. This suggested that cell lines may cluster by invasive properties rather than by tumor of origin. This indicated the existence of a dedifferentiation phenotype.

In melanoma cell lines, sialyl acetylesterase (SAE) actively reacted with the ABP and caused PTMs. Another example was that invasiveness-related serine hydrolase (IRSH), an integral membrane enzyme upregulated in invasive cancer cells from several tumor types, is upregulated in invasive cancer lines independent of tumor origin. A number of documentations indicate that mRNA expression in a disease like cancer does not correlate with protein expression mainly because protein expression is regulated not by changes in the level of expression but by its modifications. Some proteins that are regulated posttranslationally have a message level different from activity level, which is indicative of posttranslational regulation. Whether any of the SHs upregulated in invasive cancers is required for their aggressive properties

is presently unknown. A global analysis of SH activities provides proteomic information of sufficient quantity and quality to permit the phenotypic classification of cancer cells; invasive and noninvasive cancers exhibit nearly orthogonal serine hydrolase activity profiles; secreted and membrane enzymes are better predictors of cancer phenotype than are cytosolic enzymes. The information obtained from cell-line system will be useful in identifying markers from tumor. It is also emphasized here that the assays described above can be used by any clinical laboratory for the rapid and affordable screening of a variety of tumors.

4.3 Other Technologies

Some other technologies to detect markers are also utilized for cancer diagnosis.

4.3.1 Genomics

DNA-based markers include mutations, loss of heterozygosity (LOH), microsatellite instability, DNA hypermethylation, mitochondrial DNA mutations, and detection of cancer-associated viral and bacterial DNA. Lung cancer is used as an example here (Fig. 2). The development of lung cancer requires repeated insults of carcinogens, mainly from tobacco smoke, into the lungs over a long period of time. The carcinogenesis process takes a few decades and results in the accumulation of multiple molecular abnormalities in cells, which is the basis for malignant transformation and tumor progression. There is evidence that genetic abnormalities occur in the early carcinogenic process in the lungs of chronic smokers and that certain abnormalities may persist for many years after smoking cessation. Genetic and molecular alterations, such as mutations in the *p53* tumor suppressor gene and *K-ras* protooncogene, promoter hypermethylation of the *p16* tumor suppressor gene, and LOH in multiple critical chromosome regions have been identified frequently in early stage lung cancers. Investigators are exploring the possibility of using these alterations as biomarkers in the early detection of and in risk assessment for lung cancer. With the completion of human genome mapping and advances in high-throughput technologies, the discovery of molecular alterations in the carcinogenic process is accelerating. A substantial effort is now underway to conduct large-scale cooperative discoveries and validations of biomarkers for early cancer diagnosis. It is anticipated that molecular marker-based novel diagnostic strategies will be developed and merged into clinical practice to augment current, inefficient tools for diagnosing patients with early stage lung cancer. Different examples of biomarkers used to detect a variety of cancers are discussed in Section 5.1.

4.3.2 Epigenomics

The development of a malignant phenotype in a given cell is accompanied and often preceded by an accumulation of DNA lesions. Until recently, it was firmly believed that this change was exclusively genetic, that is, affecting the actual base pairing of the double stranded DNA (e.g. base pair mutations, deletions, LOH, microsatellite instability, and so on.). However, changes in the structure of the DNA and its scaffolding proteins may, without disrupting the base pairing of the DNA, interfere with the normal physiologic functions of the DNA. These changes are referred to as "epigenetic" and include mechanisms like nuclear compartmentalization, stable feedback loops, differential replication timing, heritable chromatin structures such as histone acetylation and

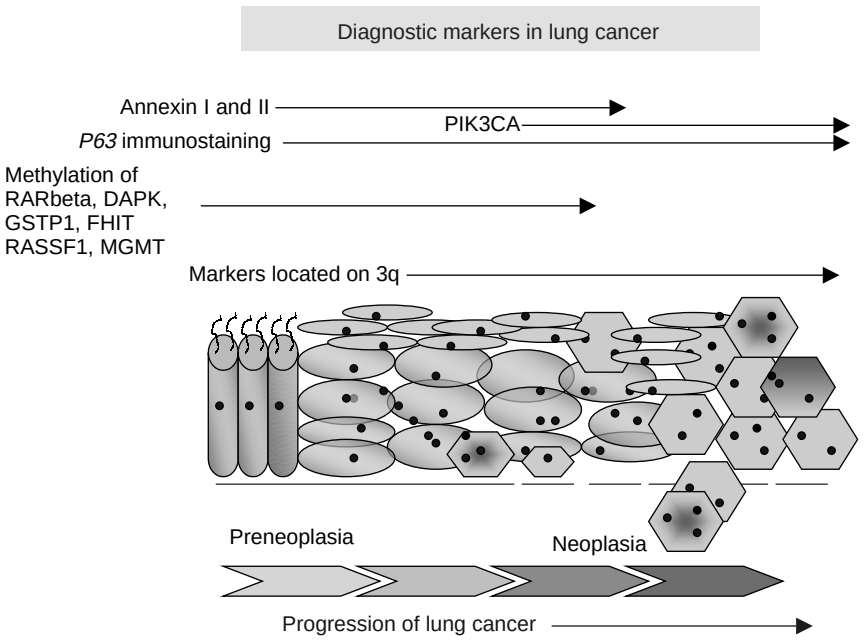

Fig. 2 Potential biomarkers to predict lung cancer. Epithelial carcinogenesis is thought to be a multistep process involving sequential activation of oncogenes as well as the inactivation of tumor suppressor genes in a clonal population of cells. These genetic changes generate concomitant phenotypic changes in the tumor cells that promote survival and proliferation. A number of genetic alterations in oncogenes and tumor suppressor genes have been reported in a variety of tumor types. Because conventional histology is unable to predict which lesions are likely to progress to cancer, the development of molecular-based approaches to identify predictive genetic changes would greatly improve the potential for early detection, prognostication, and intervention. In the figure, markers expressed in different stages of tumor development are shown. As the cancer progresses, the shape of cells changes and many cells become multinucleated.

foremost DNA methylation. Epigenetics is the study of processes that establish metastable (i.e. somatically heritable) states of gene expression without altering the DNA sequence. A number of epigenetic markers can be utilized for cancer diagnostics. As an example, we have presented potential epigenetics biomarkers in colon cancer development (Fig. 1). In large part, the field of epigenetics addresses processes that affect DNA methylation and/or chromatin structure to establish metastable, but somatically heritable, states of gene expression. Thus, epigenetic phenomena can be defined as changes in the genetic material (specifically the genomic DNA and chromatin) that alter the regulation of gene expression in a manner that is somatically heritable (and sometimes heritable through the germline) but is nonmutational, that is, does not involve changes in the DNA sequence.

In the normal mammalian genome, methylation occurs only at the cytosines 5′ to guanosines at the CpG dinucleotides. The CpG dinucleotides have been depleted progressively from the eukaryotic genome over evolution via spontaneous

deamination. The remaining CpGs have a very high frequency of methylation-facilitating chromatin arrangements to render most of the genome late replicating and repressed for the transcription of repeated regions. CpG islands, small stretches of DNA, do contain the expected frequency of CpGs, however. The methylation-protected regions generally are located in the promoter regions. In a normal human, up to 90% of CpGs are methylated.

It has been suggested that the lack of methylation might be a prerequisite for active transcription. Only two exceptions to this rule have been reported: fully methylated CpG islands in the silenced allele for specific imprinted autosomal genes and multiple silenced genes on the inactivated X-chromosomes of females. Another emerging concept indicates that deacetylation of histone is the primary step that results in the recruitment of methyltransferase to the CpG islands and in hypermethylation of the promoter. To develop cancer-prevention strategies, it will be essential to identify factors that regulate the activity of deacetylation. Extensive research is needed in this area, as targets for chemoprevention could emerge from such studies. Furthermore, epigenetic regulation occurs early in the development of cancer and provides an opportunity to develop intervention approaches that prohibit further cancer development.

Concurrent aberrant methylation patterns have been evaluated as tools in the management of patients with cancer. The predictive value, the therapeutic manipulation, and the prognostic significance of aberrantly methylated gene loci have been tested in hematological as well as in solid neoplasia of adults and children. Importantly, methylation does not only hold the potential for being "just another" biomarker, but also, as it can be reverted chemically, for being a phenomenon that holds great promise for therapeutic exploitation.

Technologies used to detect epigenetic alterations include methylation-specific PCR, MethyLight DNA analysis, restriction landmark genomic screening (RLGS), methyl cytosine amplification (MCA), and immunoassays to detect histone acetyl transferase. Nanochips with multiple epigenetic marker genes are also being developed in academic and private institutes.

4.3.3 RNA-based Assays

There are various ways of detecting cancer cells by analyzing RNA. Studies of differential mRNA expression can be informative, but mRNA levels do not always correlate with protein levels. Proteins often are subject to proteolytic cleavage or to posttranslational modifications, such as phosphorylation or glycosylation. Cancer-detection strategies that target expressed proteins are becoming increasingly popular because proteomic approaches characterize the modified or unmodified proteins that are involved in cancer progression. It is critical to preserve samples in appropriate condition – for example, extremely low-storage temperature – for RNA analysis. Such requirements make RNA-based assays somewhat difficult to apply in clinical settings.

4.3.4 Posttranslational Modifications

Posttranslational modifications (PTMs) are an integral part of gene regulation. PTMs of proteins, not detected through RNA analysis, may occur at different stages of tumor development, which is indicative of early or late events of transformation. During PTM, a protein is modified either in the endoplasmic reticulum or the Golgi

complex by undergoing phosphorylation, acetylation, glycosylation, palmitoylation, myristilation, ubiquitination, farensylation, methylation, or sialylation. Specific genes are turned on or off at the onset of initiation, development, and progression of diseases such as cancer. PTMs have been reported for the key steps in tumor progression, such as cell cycle check point and apoptosis. There is a need for research to identify PTM markers at an early stage of cancer development and for developing prevention strategies.

4.3.5 Imaging

Imaging is a powerful tool for cancer diagnosis. Mammography is currently regarded as the most-effective method for early detection of breast cancer. Nevertheless, it is estimated that up to 30% of breast lesions are missed during the screening process. In addition, when radiologists classify mammograms as suspicious, and biopsies are performed to confirm malignancy, only about 10 to 30% of mammographically identified suspicious lesions are positive. The use of magnetic resonance imaging (MRI) offers a great opportunity to develop alternative ways to diagnose breast cancer. The image resolution is substantially higher and true three-dimensional identification is present. Cancers are shown to enhance relative to other tissue following the administration of different agents such as gadolinium-diethyl-enetriamine pentaacetic acid (Gd-DTPA).

The sentinel node technique is widely employed in cases of breast cancer diagnosis. Different detector systems have been used, but several problems have occurred in clinical practice. One such problem is the difficulty in accurately detecting the sentinel node within the auxiliary soft tissue. This problem is even more pronounced in cases of abdominal and thoracic tumors. The innovative imaging probe (IP) allows primary tumors and secondary lesions, if appropriately radiolabeled, to be visualized on a monitor. The IP is a small, portable, high-resolution gamma camera used in radioguided surgery and can be applied optimally for minimally invasive surgery in breast cancer treatment. The collection of more data would improve identification techniques for carcinoma and parenchyma, fibroadenoma, and benign proliferative changes.

Another powerful technology is functional imaging that can address as yet unresolvable questions about cancer biology in both basic sciences and clinics. Combining functional imaging with anatomical criteria of the disease can be beneficial in the management of cancer. Functional CT describes the use of existing technologies and conventional contrast agents to capture physiological parameters that reflect the vasculature of tumors and other tissues. The technique is readily incorporated into routine conventional CT examinations. In tumors, the physiological parameters obtained provide an *in vivo* marker of angiogenesis. In addition to serving as a research tool, functional CT has clinical applications in tumor diagnosis, staging, risk stratification, and therapy monitoring, including the characterization of pulmonary nodules, detection of occult hepatic metastases, grading of cerebral glioma, and monitoring of antiangiogenesis drugs. Since software is available for the development of multisided CT systems, functional CT is poised to have a significant impact on the imaging of patients with cancer.

Computer-assisted diagnosis (CAD) has been applied to mammography and chest radiography to improve diagnostic

accuracy and has the potential to improve the accuracy of MRI for detecting nodal metastases in women with newly diagnosed breast cancer. Clinical researchers should design future trials to incorporate plans for collecting and analyzing molecular markers.

5 Challenges in Cancer Diagnostics: Validation of Biomarkers

Our ability to predict the biology of premalignant lesions has been limited by controversial clinical terms, inaccurate and subjective assessments because of the lack of well-defined criteria for grading interexaminer variability, and, most importantly, the lack of genotypic-/phenotypic-based biomarkers. It is somewhat paradoxical that the spectacular achievements made in the last few years in defining molecular genetic alterations in cancer have not translated into effective strategies for screening and early diagnosis. Methods other than nonrandomized clinical trials must be established to evaluate and validate biomarkers for clinical application during the early stages of investigation. Validation of new biomarkers by traditional clinical trials will be unfeasible owing to cost, time required for follow-up, rapidly changing technologies, variations in treatments, and the large number of biomarkers expected to be discovered.

5.1 Biomarkers as Surrogate End Points

Biomarkers can also be used as surrogate endpoints; however, it is necessary that the following criteria are satisfied: (1) modulation or elimination of the risk marker must correlate with a decrease in cancer incidence; (2) the marker must be modulated by chemopreventive agents; and (3) the marker must be a determinant of outcome. Risk markers are usually used as surrogate outcomes in order to detect the effect of a prevention intervention more rapidly than waiting for the definitive outcome. Biomarkers for colon cancer are most likely to be found in stool specimens.

Utilization of surrogate endpoint biomarkers (SEBs) addresses many of the limitations seen in large chemoprevention trials. However, there always has been a question of whether the surrogate endpoint is a valid step along the pathway to carcinogenesis. In this regard, three types of SEBs that are being used in chemoprevention research are: (1) *Pathologic Biomarkers* are premalignancies, such as colon polyps, that are known to progress to cancer in some cases; (2) *Drug Effect Biomarkers* are dynamic endpoints of a direct drug-to-target effect, such as aspirin, which inhibits the production of prostaglandins. (3) *Biological Biomarkers* are proteins, genes, or quantitative cellular processes that are used to predict cancer diagnosis or risk. These can also be divided into *Process Biomarkers*, such as those involved in pathway-derived carcinogenesis or regulatory endpoints, and *Screening Biomarkers*, which are derived from proteomic assessment.

Another major challenge is to make diagnostic assays high throughput, easy to adopt, and applicable to multiple samples. With the advancement of knowledge in the biomarker field, it is possible to achieve these goals. Furthermore, the knowledge gained in diagnostics will be useful in approaches, such as prevention, which relies on interventions during early phases of carcinogenesis to reduce the incidence of invasive cancer and, ultimately, to reduce cancer-related morbidity and mortality.

5.2 Biomarker Validation

For biomarker validation, information on quantifying biomarker endpoints is essential. There should be strict laboratory standards and procedures for completing assays. Validation is very difficult unless newer technologies are used to increase statistical power. One newer approach using technology is to develop pattern recognition with quantitative computed imaging techniques. There is a need for reliable biomarker panels for use in large population studies that meet the following specifications: (1) serum is the preferred sample for the assay; (2) the assay should not be expensive; (3) there must be an easy method for quality control; (4) there must be an algorithm because proteins, which are multiclonal, will vary among different types of cancer.

Novel biomarkers must be validated prior to general use. Sensitivity, specificity, and predictive value must be determined through the use of body fluids, paired tumors, and surrounding tissue from a wide variety of cancers. A large number of samples from individuals with known characteristics should be processed to minimize the problems of confounding and to avoid forged associations. Prior to field testing, it should be established that the biomarker is truly in the path of pathogenesis and not simply the result of an adaptive response. Case-control studies on stored samples should be employed to test biomarker efficiency. While the emerging technologies show great promise, care must be taken to define and establish references or baseline profiles from a variety of tissues (including normal tissues) or body fluids.

Extensive animal studies may help to refine human testing prior to screening. The biomarker assay should be reproducible to avoid mislabeling individuals as false positives or false negatives and to provide sufficient time before clinical diagnosis. Sample sources such as urine and stools for DNA make it important to evaluate techniques and methods for extraction. A biomarker should be tested in the population intended for ultimate use because tests for early detection often require a higher specificity than do tests in clinical diagnosis. With regard to early detection and prevention strategies, biological samples should be screened for a cluster or pattern of markers and for substantial accuracy in performing risk assessment. The positive predictive value of a test can be assessed and labeling of individuals as either false positives or false negatives can be minimized by using a panel of biomarkers.

Despite extensive efforts to improve the diagnosis and treatment of patients with lung cancer over the past three decades, the overall survival rate of patients with the disease remains dismal. Because lung cancer takes a few decades to develop, early diagnosis of the disease or identification of truly high-risk populations may provide an opportunity to successfully cure or prevent the disease. Recent advances in understanding the biological basis of lung tumorigenesis have provided new tools for detecting malignant cells and the process of malignant transformation and progression. Along with the identification of molecular abnormalities in early lung tumorigenesis, advanced molecular analytic technologies have emerged that may facilitate the development of rapid and effective methods for early diagnosis and risk assessment.

The use of functional imaging to develop "functional" approaches to diagnosis and therapy needs further investigation. Many methodologies exist for the acquisition of

primary data on the efficacy of imaging technologies. A form of economic cost-effectiveness modeling called iterative decision analysis can be used to set research and service priorities. Cancer clinicians need to take an increased role in functional imaging research because they have considerable expertise in the development and use of treatments that modify cell and tissue function.

Microscopic examination of tissue detects a disease after it has developed and is important in forecasting tumor behavior and prognosis. Additional procedures, however, are essential for early detection.

With regard to the use of biomarkers as risk factors or for detecting preneoplastic lesions, several criteria must be met as described below:

- The variance of the detection tests and the intra and interlaboratory variance must be known.
- The biomarker and its assay must provide acceptable predictive accuracy for risk or for the presence of cancer.
- The biomarker must be differentially expressed in normal, premalignant, and in tumor tissue.
- The assay of the biomarker should be simple enough to be accommodated in clinical settings.

In cancer prevention, intervention is achieved through the use of pharmaceuticals, vitamins and minerals, or other chemicals that reduce cancer risk. Choosing the targets for chemoprevention research is of the utmost importance. There should also be evidence of a nongenetic component (i.e. acquired mutations) to the carcinogenesis process in the target disease. Another important issue is the ability to obtain tissue to study the disease, and there must be some amount of preclinical (e.g. mechanism-of-action, *in vitro* studies, and preclinical *in vivo* data) evidence to support the research.

A number of limitations exist in chemoprevention trials. Some studies indicate a screen-to-accrual rate of 5 to 10%. A significant barrier in the process is the number of participants needed to reach a cancer endpoint to a trial. Large numbers of participants are needed for a significant period of time (10–20 years) if cancer is used as an endpoint. In addition, the design of chemoprevention trials is very complex.

6 Summary

Early cancer detection is one of the best defenses against cancer. Emerging technological advancements in proteomics and genomics are allowing us to extract larger amounts of information from smaller and smaller samplings. In the future, single-cell analysis and nanotechnology may allow us to study cancer heterogeneity and the effects of complex cellular interactions precisely. The early results in microarray-based gene expression analysis are painting a detailed portrait of cancer that illustrates the individuality of each tumor and allows familial relationships to be recognized through the identification of cell types sharing common expression patterns. As proteomic and genomic tools continue to develop, it will be essential to integrate information from several analyses into a common framework to develop a more complete understanding of cancer diagnostics and its implications in cancer-risk assessment and therapeutics. Efforts are continuing to identify new markers that may be useful in cancer diagnosis. Once good markers of risk are identified, it may eventually be possible to perform risk-based screening in which the frequency of

screening is a function of the patient's risk based on the family history and other risk factors.

See also Oncogenes.

Bibliography

Books and Reviews

Bestor, T.H. (2003) Unanswered Questions about the Role of Promoter Methylation in Carcinogenesis, in: Verma, M., Dunn, B.K., Umar, A. (Eds.) *Epigenetics in Cancer Prevention: Early Detection and Risk Assessment*, New York Academy Press, New York, pp. 22–27.

Srivastava, S., Henson, D.E., Gazdar, A. (Eds.) (1999) *Molecular Pathology of Early Cancer*, IOS Press, Amsterdam.

Verma, M. (Eds.) (2002) *Molecular Signaturs of Infectious Agents*, IOS Press, The Netherlands.

Verma, M., Dunn, B.K., Umar, A. (Eds.) (2003) *Epigenetics in Cancer Prevention: Early Detection and Risk Assessment*, New York Academy Press, New York.

Primary Literature

Adam, B.L., Qu, Y., Davis, J.W. et al. (2002a) Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men, *Cancer Res.* **62**, 3609–3614.

Adam, B.L., Vlahou, A., Semmes, O.J. et al. (2001) Proteomic approaches to biomarker discovery in prostate and bladder cancers, *Proteomics* **1**, 1264–1270.

Adam, G.C., Sorensen, E.J., Cravatt, B.F. (2002b) Trifunctional chemical probes for the consolidated detection and identification of enzyme activities from complex proteomes, *Mol. Cell Proteomics* **1**, 828–835.

Ahlquist, D.A., Shuber, A.P. (2002) Stool screening for colorectal cancer: evolution from occult blood to molecular markers, *Clin. Chim. Acta.* **315**, 157–168.

Ahlquist, D.A., Skoletsky, J.E., Boyant, K.A. et al. (2000) Colorectal cancer screening by detection of altered human DNA in stool: feasibility of a multitarget assay system, *Gastroenterology* **119**, 1219–1122.

Alaiya, A.A., Franzen, B., Auer, G. et al. (2002) Cancer proteomics: from identification of novel markers to creation of artificial learning models for tumor classification, *Electrophoresis* **21**, 1210–1217.

Ardekani, A.M., Liotta, L.A., Petricoin, E.F., 3rd. (2002) Clinical potential of proteomics in the diagnosis of ovarian cancer, *Expert. Rev. Mol. Diagn.* **2**, 312–320.

Assimakopoulos, D., Kolettas, E., Patrikakos, G. et al. (2002) The role of CD44 in the development and prognosis of head and neck squamous cell carcinomas, *Histol. Histopathol.* **17**, 1269–1281.

Banks, R.E., Dunn., M.J., Forbes, M.A. (1999) The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis–preliminary findings, *Electrophoresis* **29**, 689–700.

Baylin, S., Bestor, T.H. (2002) Altered methylation patterns in cancer cell genomes: cause or consequence? *Cancer Cell* **1**, 299–305.

Belinsky, S.A., Nikula, K.J., Palmisano, W.A. et al. (1998) Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11891–11896.

Bernstein, J., Teraoka, S., Haile, R. et al. (2003) Designing and implementing quality control for multi-center screening of mutations in the ATM gene among women with breast cancer, *Hum. Mutat.* **21**, 542–550.

Brichory, F.M., Misek, D.E., Yim, A.M. et al. (2001) An immune response manifested by the common occurrence of annexins I and II autoantibodies and high circulating levels of IL-6 in lung cancer, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9824–9829.

Cahill, D.J. (2001) Protein and antibody arrays and their medical applications, *J. Immunol. Methods* **250**, 81–91.

Campa, M.J., Wang, M.Z., Howard, B. et al. (2003) Protein expression profiling identifies macrophage migration inhibitory factor and cyclophilin A as a potential molecular targets in non small cell lung cancer, *Cancer Res.* **63**, 1652–1656.

Champliaud, M.F., Virtanen, I., Tiger, C.F. et al. (2000) Post-translational modifications and beta/gamma chain associations of human

laminin alpha1 and laminin alpha5 chains: purification of laminin-3 from placenta, *Exp. Cell Res.* **259**, 326–335.

Chang, F., Steelman, L.S., McCubrey, J.A. (2002) Raf-induced cell cycle progression in human TF-1 hematopoietic cells, *Cell Cycle* **1**, 220–226.

Chapman, J.D., Bradley, J.D., Eary, J.F. et al. (2003) Molecular (functional) imaging for radiotherapy applications: an RTOG symposium, *Int. J. Radiat. Oncol., Biol., Phys.* **55**, 294–301.

Chim, C.S., Liang, R., Tam, C.Y. et al. (2001) Methylation of p15 and p16genes in acute promyelocytic leukemia: potential diagnostic and prognostic significance, *J. Clin. Oncol.* **19**, 2033–2040.

Chong, B.E., Hamler, R.L., Lubman, D.M. et al. (2001) Differential screening and mass mapping of proteins from premalignant and cancer cell lines using nonporous reversed-phase HPLC coupled with mass spectrometric analysis, *Anal. Chem.* **73**, 1219–1227.

Dwek, R.A., Butters, T.D., Platt, F.M. et al. (2002) Targeting glycosylation as a therapeutic approach, *Nat. Rev. Drug Discov.* **1**, 65.

Feinberg, A.P. (2001) Cancer epigenetics takes center stage, *Proc. Natl. Acad. Sci. USA.* **98**, 392–394.

Feras, E.R., Stephens, D.J., Walters, C.E. (1999) The role of cholesterol in the biosynthesis of beta-amyloid, *NeuroReport* **10**, 1699–1705.

Fung, E.T., Wright, G.L., Dalmasso, E.A. (2000) Proteomic strategies for biomarker identification: progress and challenges, *Curr. Opin. Mol. Ther.* **2**, 643–650.

Gharib, T.G., Chen, G., Wang, H. et al. (2002) Proteomic analysis of cytokeratin isoforms uncovers association with survival in lung adenocarcinoma, *Neoplasia* **4**, 440–448.

Gruner, D., Sturm, I., Hemmati, P. et al. (2003) Multigene analysis of Rb pathway and apoptosis control in esophageal squamous cell carcinoma identifies patients with good prognosis, *Int. J. Cancer* **103**, 445–454.

Herbert, B. (1999) Advances in protein solubilisation for two-dimensional electrophoresis, *Electrophoresis* **20**, 660–663.

Ho, W.T., Lam, P.W. (2003) Clinical performance of computer-assisted detection (CAD system in detecting carcinoma in breasts of different densities, *Clin. Radiol.* **58**, 133–136.

Issaq, H.J., Veenstra, T.D., Conrads, T.P. et al. (2002) The SELDI-TOF MS approach to proteomics: protein profiling and biomarker identification, *Biochem. Biophys. Res. Commun.* **292**, 587–592.

Jones, P.A. (2002) DNA methylation and cancer, *Oncogene* **21**, 5358–5360.

Jones, C., Nonni, A.V., Fulford, L. et al. (2001) CGH analysis of ductal carcinoma of the breast with basaloid/myoepithelial cell differentiation, *Br. J. Cancer* **85**, 422–427.

Jones, M.B., Krutzsch, H., Shu, H., Zhao, Y., Liotta, L.A., Kohn, E.C., and Petricoin, E.F., 3rd (2002) Proteomic analysis and identification of new biomarkers and therapeutic targets for invasive ovarian cancer, *Proteomics* **2**, 76–84.

Jonsson, H., Larsson, L.G., Lenner, P. (2003) Detection of breast cancer with mammography in the first screening round in relation to expected incidence in different age groups, *Acta Oncol.* **42**, 22–29.

Kiekhaefer, C.M., Grass, J.A., Johnson, K.D. et al. (2002) Hematopoietic-specific activators establish an overlapping pattern of histone acetylation and methylation within a mammalian chromatin domain, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14309–14314.

Kornblith, A.B., Herndon, J.E., Silverman, L.R. (2002) Impact of azacytidine on the quality of life of patients with myelodysplastic syndrome treated in a randomized phase III clinical trial: a cancer and leukemia group B study, *J. Clin. Oncol.* **20**, 2441–2452.

Li, J., Zhang, Z., Rosenzweig, J. et al. (2003) Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer, *Clin. Chem.* **48**, 1296–1304.

Liotta, L.A., Petricoin, E.F. 3rd, Ardekani, A.M. et al. (2003) General keynote: proteomic patterns in sera serve as biomarkers of ovarian cancer, *Gynecol. Oncol.* **88**, S25–S28.

Loeb, L.A., Loeb, K.R., Anderson, J.P. (2003) Multiple mutations and cancer, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 776–781.

Mao, L., Lee, J.S., Kurie, J.M. et al. (1997) Clonal genetic alterations in the lung of current and former smokers, *J. Natl. Cancer Inst.* **89**, 857–862.

Marceau, N., Loranger, A., Gilbert, S. et al. (2001) Keratin-mediated resistance to stress and apoptosis in simple epithelial cells in relation to health and disease, *Biochem. Cell Biol.* **79**, 543–555.

Mattu, T.S., Royle, L., Langridge, J. et al. (2000) O-glycan analysis of natural human neutrophil gelatinase B using a combination of normal

phase-HPLC and online tandem mass spectrometry: implications for the domain organization of the enzyme, *Biochemistry* **39**, 15695–15704.

Mehta, A., Carrouee, S., Conyers, B. et al. (2001) Inhibition of hepatitis B virus DNA replication by imino sugars without the inhibition of the DNA polymerase: therapeutic implications, *Hepatology* **33**, 1488–1594.

Negm, R.S., Verma, M., Srivastava, S. (2002) The promise of biomarkers in cancer screening and detection, *Trends Mol. Med.* **8**, 288–293.

Pandey, A., Blagoev, B., Kratchmarova, I. et al. (2002) Cloning of a novel phosphotyrosine binding domain containing molecule, Odin, involved in signaling by receptor tyrosine kinases, *Oncogene* **21**, 8029–8036.

Pawluczyk, O., Augustine, B.J., Yaffe, M.J. et al. (2003) A volumetric method for estimation of breast density on digitized screen-film mammograms, *Med. Phys.* **30**, 352–364.

Pepe, M.S., Etzioni, R., Feng, Z. et al. (2001) Phases of biomarker development for early detection of cancer, *J. Natl. Cancer Inst.* **93**, 1054–1061.

Petricoin, E.F., Liotta, L.A. (2002) Proteomic analysis at the bedside: early detection of cancer, *Trends Biotechnol.* **20**, S30–S34.

Petricoin, E.F., Ardekani, A.M., Hitt, B.A. et al. (2002) Use of proteomic patterns in serum to identify ovarian cancer, *Lancet* **359**, 572–577.

Polednak, A.P. (2003) Trends in survival for both histologic types of esophageal cancer in U.S. surveillance, epidemiology and end results areas, *Int. J. Cancer* **105**, 98–100.

Schwetzer, B., Kingsmore, S. (2002) Measuring proteins on microarrays, *Curr. Opin. Biotechnol.* **13**, 14–19.

Sidransky, D. (2002) Emerging molecular markers of cancer, *Nat. Rev. Cancer* **2**, 210–219.

Silverman, L.R., Demakos, E.P., Peterson, B.L. (2002) Randomized controlled trial of azacytidine in patients with the myelodysplasia syndrome: a study of the cancer and leukemia group B, *J. Clin. Oncol.* **20**, 2429–2440.

Smith, D.L., Evans, C.A., Pierce, A. et al. (2002) Changes in the proteome associated with the action of bcr-abl tyrosine kinase are not related to transcriptional regulation, *Mol. Cell Proteomics* **1**, 876–884.

Sorlie, T., Perou, C.M., Tibshirani, R. et al. (2002) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10867–10874.

Srinivas, P.R., Verma, M., Zhao, Y. et al. (2002) Proteomics for cancer biomarker discovery, *Clin. Chem.* **48**, 1160–1169.

Steel, L.S., Mattu, T., Heibersteit, H. et al. (2001) A proteomic approach for the discovery of early detection markers of hepatocellular carcinoma, *Dis. Markers* **17**, 179–189.

Takamatsu S., Oguri S., Minowa M.T. et al. (1999) Unusually high expression of N-acetylglucosaminyltransferase-IVa in human choriocarcinoma cell lines: a possible enzymatic basis of the formation of abnormal biantennary sugar chain, *Cancer Res.* **59**, 3949–3953.

Takamochi, K., Ogura, T., Suzuki, K. et al. (2001) Loss of heterozygosity on chromosomes 9q and 16p in atypical adenomatous hyperplasia concomitant with adenocarcinoma of the lung, *Am. J. Pathol.* **159**, 1941–1948.

Verma, M., Srivastava, S. (2002) Epigenetics in cancer: implications for early detection and prevention, *Lancet Oncol.* **3**, 755–763.

Verma, M., Wright, G.L., Jr., Hanash, S.M. et al. (2001) Proteomic approaches within the NCI early detection research network for the discovery and identification of cancer biomarkers, *Ann. N. Y. Acad. Sci.* **945**, 103–115.

Wistuba, I.I., Lam, S., Behrens, C. et al. (1997) Molecular damage in the bronchial epithelial of current and former smokers, *J. Natl. Cancer Inst.* **89**, 1366–1373.

Xu, B., O'Donnell, A.H., Kim, S.T. et al. (2002) Phosphorylation of serine 1387 in Brca1 is specifically required for the Atm-mediated S-phase checkpoint after ionizing irradiation, *Cancer Res.* **62**, 4588–4591.

Yip, T.T., Van de Water, J., Gerswin, M.E. (1996) Cryptic antigenic determinants on the extracellular pyruvate dehydrogenase complex/mimeotope found in primary biliary cirrhosis. A probe by affinity mass spectrometry, *J. Biol. Chem.* **272**, 32825–32833.

Zhao, L., Xu, S., Peterson, C. et al. (2002) Purification and characterization of a 95-kDa protein–carcinoembryonic antigen-related cell adhesion molecule 8–from normal human granulocytes, *J. Immunol. Methods* **270**, 27–35.

Neural Tube Defects, Molecular Biology of

Andrew J. Copp
Institute of Child Health, University College London, London, UK

1 Neurulation: A Multiphase Embryonic Event 122
1.1 Rostro-caudal Sequence of Neurulation Events 123

2 Causation of Spina Bifida and Other NTDs 124
2.1 Genetic Factors 124
2.2 Environmental Factors 125

3 Developmental Mechanisms of Spina Bifida and Other NTDs 126
3.1 Initiation of Neurulation: Role of Planar Cell Polarity Signaling 127
3.2 Regulation of Neural Plate Bending: Mechanisms of Spina Bifida 129
3.3 Molecular Regulation of Cranial Neurulation 131
3.3.1 Cranial Mesenchyme 131
3.3.2 Actin Cytoskeleton 133
3.3.3 Neural Crest Emigration 133
3.3.4 Programmed Cell Death 133
3.3.5 Neuronal Differentiation 133
3.3.6 Cell Proliferation 134

4 Prevention of NTDs by Folic Acid and Other Vitamins 134
4.1 Animal Models of Folate-preventable NTDs 134
4.2 Folate-resistant NTDs and Inositol as an Adjunct Therapy 135

5 Conclusions 135

Bibliography 136
Books and Reviews 136
Primary Literature 136

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

Keywords

Anencephaly
Exposure of the interior of the brain to the outside, with absence of the skull vault, resulting from failure of neural tube closure in the future brain.

Closed spina bifida (or dysraphism)
Skin-covered spinal lesions in which the lowest portion of the spinal cord fails to become separated from adjacent tissues during embryonic development, leading to spinal cord tethering inside the vertebral canal.

Craniorachischisis
The most severe NTD in which most of the brain and the entire spinal cord remain open, as a result of failure of the initiating event of neurulation (Closure 1).

Exencephaly
An early developmental form of anencephaly, in which the persistently open cranial neural folds appear everted and ''overgrown.'' Later in gestation, degeneration of the exposed neural tissue results in the ''flat skull'' appearance of anencephaly.

Hydrocephalus
Increase in cerebrospinal fluid pressure within the cavities of the brain, often associated with spina bifida.

Myelomeningocele
Alternative term for open spina bifida. The spinal cord is open (myelo-), and the meninges emerge from the defect in a cystic lesion (meningocele). An open spinal cord without a covering meningeal sac is called *myelocele*.

Neural plate
A layer of ectodermal cells on the back of the early embryo that undergoes folding and fusion to produce the neural tube, the precursor of the entire brain and spinal cord.

Neural tube defects
The group of birth defects comprising spina bifida, anencephaly, and craniorachischisis that result when the neural tube fails to complete closure during embryogenesis.

Neural tube
The tubular structure formed in the early embryo by folding and fusion of the neural plate in the dorsal midline to produce the precursor of the brain and spinal cord.

Neuropore
A transient "holelike" opening in the neural tube at which neural tube closure is undergoing completion.

Neurulation
The embryonic process by which the neural tube is formed. In higher vertebrates, neurulation comprises primary events, characterized by neural folding, and secondary neurulation in which the caudal neural tube forms by hollowing of a cord of cells to create a tubular structure.

Open spina bifida (spina bifida aperta)
Exposure of the interior of the spinal cord to the outside environment. This definition encompasses the pathological terms myelocele and myelomeningocele.

Spina bifida is usually grouped together with the closely related conditions anencephaly and craniorachischisis, to form the category of congenital malformations known as *neural tube defects* (NTDs). NTDs affect approximately 1 in every 1000 pregnancies, equating to a worldwide prevalence of around 120 000 pregnancies with NTDs per year. Hence, NTDs represent one of the commonest, most severe birth defects in humans. The term NTDs reflects the origin of the malformations. In normal development, the *neural plate*, the precursor of the brain and spinal cord in the early embryo, undergoes folding and fusion to form the *neural tube*. If neural tube closure fails in the developing spine, the result is open spina bifida (also called *myelomeningocele or myelocele*; Fig. 1(a)–(d)), a defect in which the interior of the spinal cord remains open to the outside, rather than being covered by the bony vertebrae and skin. The exposed nervous tissue degenerates before birth so that babies born with open spina bifida are often paralysed and lack sensation below the level of the lesion. They may exhibit urinary and faecal incontinence, and have associated defects such as *hydrocephalus* (Fig. 1(e)). If the neural tube fails to close in the embryonic brain, the resulting birth defect is initially described as *exencephaly* (Fig. 1(b)), in which the neural folds appear everted. Later in gestation, degeneration of the neural tissue converts the defect to *anencephaly*, in which the interior of the skull is exposed owing to loss of the open brain and failure of the skull roof to develop. Fetuses with anencephaly do not survive beyond birth, and are often stillborn or die during pregnancy. If the entire neural tube fails to close, from brain to low spine, this is referred to as *craniorachischisis* (Fig. 1(f)–(h)), a uniformly lethal condition. Closed spina bifida (also called *spina bifida occulta*) represents the mild end of the NTD spectrum. The lowest part of the spinal cord fails to separate from adjacent tissues during development, a defect that can produce weakness of the legs or urinary incontinence. Surgical intervention may be required in order to free the spinal cord within the vertebral canal, relieving pressure on the spinal nerve roots.

1
Neurulation: A Multiphase Embryonic Event

NTDs (neural tube defects) result from failure of neurulation, the embryonic event responsible for formation of the neural tube. Neurulation comprises a series of coordinated cellular behavior, beginning with differentiation of neurectoderm from lateral ectoderm. Concomitantly, the neural plate narrows medio-laterally and lengthens rostro-caudally by the process of convergent extension, with neuroepithelial and underlying mesodermal cells moving toward the midline and intercalating along the rostro-caudal axis. Subsequently, neural folds form at the edges of the neural plate, which then elevate, with bending at

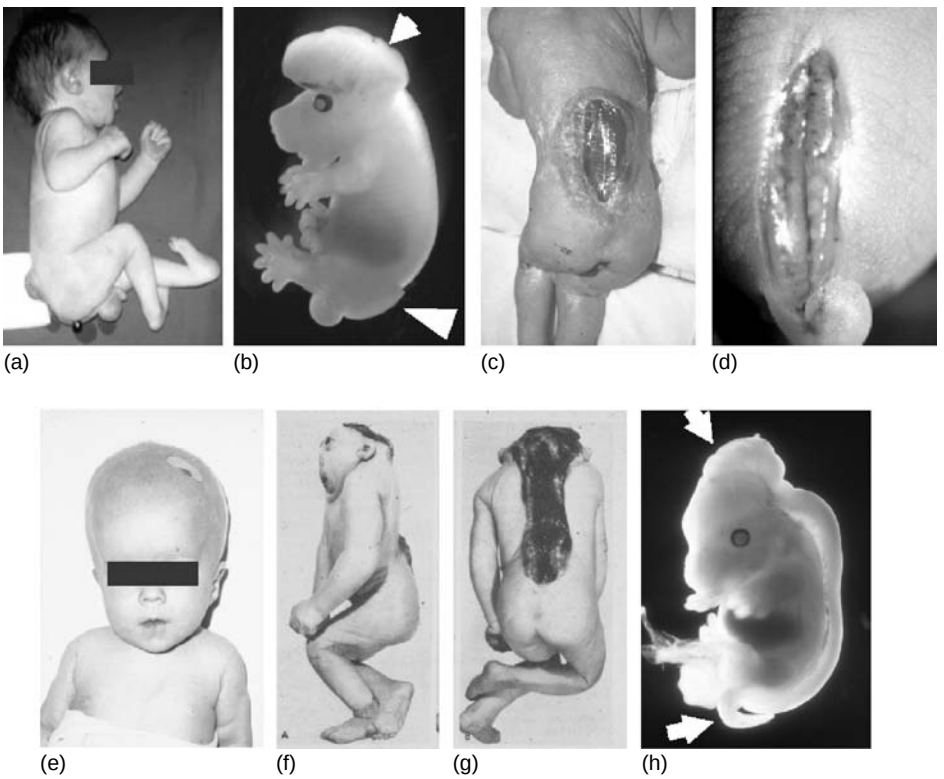

Fig. 1 The spectrum of NTDs in humans and mice. (a) Myelomeningocele in a human neonate; (b) exencephaly (arrow) and open spina bifida (arrowhead) in an E15.5 mouse embryo homozygous for the *curly tail* mutation; (c) myelocele in a human neonate; (d) open spina bifida in an E18.5 mouse fetus. Note the close similarity between these mouse and human NTDs; (e) hydrocephalus, with marked expansion of the cranial vault; (f, g) craniorachischisis in a human fetus from sideview (f) and back view (g). Most of the brain and spine are open; (h) craniorachischisis in an E15.5 mouse embryo homozygous for a mutation of *Celsr1*. The neural tube is open between the arrows, affecting most of the brain and the entire spine, as in the human defect. Parts (b, h) modified from Copp, A.J., Greene, N.D.E., Murdoch, J.N. (2003) *Nat. Rev. Genet.* **4**, 784–793; Part (c) from Copp, A.J. (1999) *Encyclopaedia of Life Sciences/www.els.net* Nature Publishing Group, London, UK; Part (d) from Stiefel, D., Shibata, T., Meuli, M., Duffy, P., Copp, A.J. (2003) *J. Neurosurg. (Spine)* **99**, 206–213.

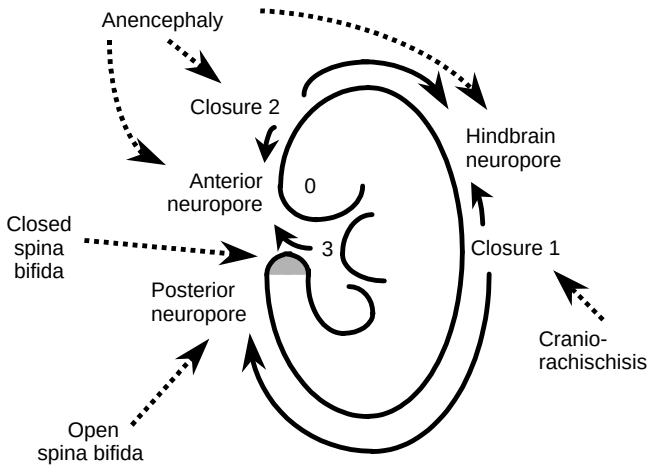

Fig. 2 Schematic summary of the successive initiation events of mouse neurulation (Closures 1, 2, and 3), the direction of spread of closure from the initiation sites (solid arrows), and the sites of completion of closure (anterior, hindbrain, and posterior neuropores). The tail bud region (red shading) is the site of secondary neurulation, which follows immediately after closure of the posterior neuropore. The main types of NTD that result from failure of the different components of neurulation are indicated by the dotted arrows. Modified from Copp, A.J., Bernfield, M. (1994) *Curr. Opin. Pediatr.* **6**, 624–631. (See color plate p. xxv).

precise dorso-ventral locations, until the tips of the neural folds come into contact in the dorsal midline. The apposed neural folds adhere and then fuse, with subsequent remodeling of the epithelia, to create the neural tube and covering surface ectoderm.

1.1 Rostro-caudal Sequence of Neurulation Events

In the mouse, neural tube closure is initiated at the hindbrain/cervical boundary (Closure 1), and then proceeds concurrently in the future brain and spinal regions (Fig. 2). Brain closure is initiated independently at the forebrain/midbrain boundary (Closure 2) and at the extreme rostral end of the forebrain (Closure 3). Closure between these initiation sites occurs simultaneously, culminating in the completion of cranial neurulation at the anterior and hindbrain neuropores. The spread of closure caudally along the spinal axis leads to progressive closure of the posterior neuropore, marking the end of primary neurulation.

While the neurulation process has been studied most closely in the mouse, human neurulation appears quite similar. Closures 1 and 3 certainly occur in human embryos, whereas there is disagreement between authors about the occurrence of an event comparable to mouse Closure 2. In fact, even in the mouse, Closure 2 is a variable event, with some inbred strains exhibiting the closure rather caudally, within the midbrain, whereas other strains undergo Closure 2 more rostrally, within the

forebrain. In these latter strains, Closures 2 and 3 may be located extremely close together, making the two events difficult to distinguish from each other. This variant of mouse Closure 2 resembles the human situation, where a distinct Closure 2 event has been inconsistently observed.

2 Causation of Spina Bifida and Other NTDs

NTDs have a multifactorial aetiology, with both genetic and environmental influences playing an important role.

2.1 Genetic Factors

The genetic component of NTD etiology is revealed by the dramatically increased risk among women who have had a previous NTD pregnancy. Compared with the average background prevalence of 0.5 to 2 per 1000 pregnancies, the recurrence risk rises after one affected pregnancy to around 1 in 25 and after two affected pregnancies to about 1 in 10. These average figures relate to live born or therapeutically aborted NTD cases, and may underestimate the true recurrence rates. Significant numbers of NTDs are lost as spontaneous abortions early in pregnancy, suggesting that the true incidence, and perhaps also the recurrence rate, of NTDs may be higher than commonly quoted. The finding that the risk of an NTD pregnancy is also increased in first-degree relatives of an index case, with risk declining in second- and third-degree relatives is further evidence of a significant genetic component to the etiology of NTDs.

In the mouse, NTDs occur as a result of mutations in more than 80 different genes. This suggests that normal functioning of a large number of genes is necessary to ensure normal neural tube development. It predicts, moreover, that human NTDs are unlikely to result from the action of just a single gene, but ultimately may prove to be etiologically related to many genes. Hence, NTDs are probably heterogeneous both morphologically, in terms of the varying level of the body axis that is primarily affected, and also etiologically in terms of their precise genetic causation.

Despite the evidence for a strong genetic etiology, no major causative genes have yet been identified for human NTDs. The strongest link is to the gene encoding 5,10-methylene tetrahydrofolate reductase (MTHFR), an enzyme involved in intracellular folate metabolism. A polymorphic, thermolabile variant of MTHFR (caused by the C677T mutation) has a higher frequency among NTD cases and their families than among normal controls in Dutch and Irish populations. Although a link with NTDs is more difficult to demonstrate in other human populations, a meta-analysis of the worldwide data showed an overall significant association between MTHFR and NTDs. The C677T mutation of MTHFR appears to contribute to the overall risk of NTDs, particularly when combined with low vitamin B_{12} status or with homozygosity for the A66G variant of the enzyme methionine synthase reductase. Hence, genetic variations in folate metabolism may predispose to NTDs.

Several of the genes implicated in mouse NTDs have been examined in human NTD cases, but to date, only rare mutations have been identified. There are several possible reasons for this negative finding. First, most of the mouse NTD–causing mutations act only in homozygous form, while heterozygotes are usually normal. This contrasts with human genetic disease

where haplo-insufficiency is the commonest cause of clinical phenotype. With a few exceptions, such as cystic fibrosis, homozygotes for specific genetic defects are usually rare and contribute little to the overall burden of genetically determined human disease. Second, many of the mouse NTD mutants are embryonic lethals. Given the extremely high level of pregnancy wastage in humans, it is probable that most disadvantaged human embryos fail to survive to late gestation. Third, it seems likely that gene–gene interactions, rather than single-gene effects, are most likely to play a major role in the etiology of human NTDs.

Gene–gene interactions have been well documented in the mouse. For example, while the *curly tail* (*ct*) mutation and the unrelated *splotch* (Sp^{2H}) mutation each cause spina bifida in homozygotes, mice doubly heterozygous for both mutations ($ct/+$; $Sp^{2H}/+$) also have NTDs. Embryos of genotype ct/ct; $Sp^{2H}/+$ exhibit particularly severe spina bifida. Translated to the human situation, this suggests that NTDs may result most commonly from the coming together in a single individual of two or more different genetic variants, which provides the interaction necessary to disturb neurulation. If many genes are required for successful neurulation, as in the mouse, a large number of possible combinations of genetic variants might be capable of producing NTDs. This could yield a "sporadic" occurrence of NTDs at a relatively high frequency, as is observed in all human populations.

2.2 Environmental Factors

Epidemiological studies point to a likely role for nongenetic factors, as well as genetic influences, in the etiology of NTDs. For example, the relatively high prevalence of defects among pregnancies in low socioeconomic groupings has led to the suggestion that nutritional factors may be of significance. This idea receives support from the finding of a preventive effect of folic acid supplementation on the occurrence and recurrence of NTDs.

As with the genetic aspects of NTD etiology, rather few environmental factors have been shown definitively to cause human NTDs. It is now widely accepted that exposure to the anticonvulsant sodium valproate during the early weeks of pregnancy, when the neural tube is closing, raises the risk of spina bifida to around 1.5 to 3%, a 15- to 30-fold increase in prevalence compared to background incidence. An elevated risk of NTDs has also been reported with carbamazepine, another anticonvulsant. Apart from these drug associations, the only other well-defined environmental association relates to the uterine milieu associated with maternal diabetes mellitus. The frequency of NTDs (like many other birth defects) is increased severalfold in pregnancies of diabetic mothers.

A variety of exogenous agents, when applied to rodent embryos, exert teratogenic effects with the production of abnormalities including NTDs. The majority of NTD-causing agents are drugs, chemicals, or other substances extraneous to pregnancy. For example, administration of sodium valproate either to the pregnant female or to the embryo *in vitro* produces NTDs although the principal defect produced in rodent embryos is exencephaly, not spina bifida as in humans. Other well-known NTD-producing teratogens include metals such as arsenic and cadmium, drugs including cytochalasins and alkaloids, and physical agents including radiation and

hyperthermia. Apart from exogenous teratogenic agents, a number of molecules derived from metabolism cause NTDs, particularly in the context of maternal disease condition. Elevated concentrations of glucose and ketone bodies, as found in diabetes mellitus, increase the risk of NTDs, a finding that parallels the predisposition to NTDs in human diabetic pregnancy. Moreover, an excess or deficiency of retinol (vitamin A), and particularly its active metabolite retinoic acid, are also potent causes of NTDs. Retinoic acid is produced endogenously in the embryo and is essential for a variety of developmental processes. An imbalance in signaling via retinoic acid receptors leads to NTDs, among other defects.

As with gene–gene interactions, environmental influences on nervous system development seem most likely to operate in the context of varying genetic susceptibility. This idea has been particularly well illustrated by studies in which different inbred mouse strains, each representing a unique genetic background, were found to exhibit varying susceptibility to cranial NTDs caused by teratogenic agents, including valproic acid. The genes that mediate this strain-specific variation in NTD predisposition are not yet known, whereas the developmental basis of the strain variation has been traced to a polymorphism in the precise pattern of cranial neurulation. Inbred strains differ in the position along the rostro-caudal axis of the developing brain at which the Closure 2 closure event occurs. While the majority of strains exhibit Closure 2 at the midbrain/forebrain boundary, others have this event occurring more caudally within the midbrain (e.g. DBA/2 strain), or more rostrally within the forebrain (e.g. NZW strain). Experimental studies have shown that strains with a caudal Closure 2 point are relatively resistant to exencephaly (Fig. 3), whether caused by teratogenic or genetic influences, whereas strains with a rostral Closure 2 point are relatively prone to exencephaly. Hence, the precise position of Closure 2, as dictated by the genetic background, sets the risk level of a strain in its interaction with either teratogenic or genetic factors that can cause cranial NTDs.

Teratogenic agents can also interact directly with the single-gene effects to modulate the frequency or severity of NTDs. For example, the low natural frequency of exencephaly in the *curly tail* strain is elevated by the administration of all-trans-RA although, interestingly, the frequency of spina bifida in this strain is actually reduced by low-dose RA. This example demonstrates that environmental agents can exert sharply contrasting effects, either exacerbating or ameliorating, upon interaction with a genetic predisposition.

3 Developmental Mechanisms of Spina Bifida and Other NTDs

Of the more than 80 genes demonstrated to be necessary for successful neurulation, only a few have been studied in sufficient detail to reveal how the genetic requirement is mediated developmentally. Recent studies have led to an improved understanding of two aspects of neurulation: the initial closure event (Closure 1), whose failure leads to the severe NTD, craniorachischisis, and the progressive closure of the spinal neural tube, which yields spina bifida when defective. In both cases, specific signaling pathways have been shown to be critical for successful neurulation. More diverse cellular events appear crucial for cranial neural tube closure,

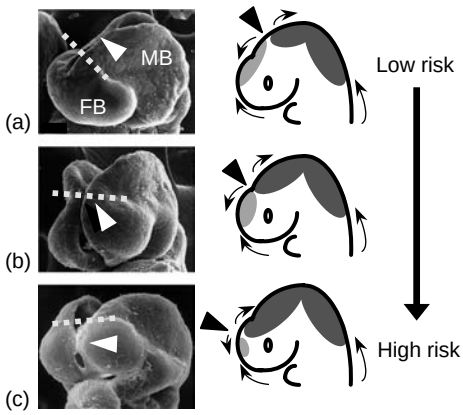

Fig. 3 Variations in position of Closure 2, in the future brain of different inbred mouse strains, in relation to the risk of exencephaly and anencephaly. Left panel shows scanning electron micrographs of embryonic brains from the left frontal view. Dotted lines indicate the forebrain/midbrain (FB/MB) boundary. Arrowheads point to the site of initiation of Closure, which is located caudal to the FB/MB boundary in the DBA/2 strain (a), at the FB/MB boundary in the CD1 strain (b), and rostral to the FB/MB boundary in the NZW strain (c). Middle panel summarizes diagrammatically this variation in Closure 2 morphology. Red shading indicates open neural folds caudal to the site of Closure 2, whereas green shading indicates the anterior neuropore, rostral to Closure 2. Arrows indicate direction of neural tube closure. Right panel shows an increasing risk of exencephaly correlated with a progressively more rostral position of Closure 2. The reason for this correlation is that the midbrain region, which lies at the apex of the cranial flexure, is mechanically the least favorable brain region for neural tube closure, and is aided in its closure by a caudal Closure 2 position (a) but hampered in its closure by a rostral Closure 2 position (c). Modified from Fleming, A., Copp, A.J. (2000) *Hum. Mol. Genet.* **9**, 575–581. (See color plate p. xxiii).

whose failure results in the development of anencephaly.

3.1
Initiation of Neurulation: Role of Planar Cell Polarity Signaling

The initial closure event (Closure 1) occurs when the neural folds fuse at the level of the future hindbrain/cervical boundary (Fig. 2). If Closure 1 fails, the embryo develops with an extensively open neural tube, from midbrain to low spine (craniorachischisis, Fig. 1(f)–(h)), although the forebrain is closed, as Closure 2 occurs normally in such embryos.

Recently, the cell movements of convergent extension, which occur from the stage of gastrulation onward, have been found to be essential for initiation of neural tube closure. Convergent extension requires signaling via the so-called planar cell polarity pathway, a noncanonical Wnt signaling pathway in which extracellular signals are transduced via cell-surface frizzled receptors and intracellular proteins of the dishevelled family (Fig. 4). Downstream events in this pathway include the activation of cytoskeleton-related proteins such as RhoA and the Jun-N-terminal kinases (JNKs).

In *Xenopus*, misexpression of dishevelled, or its interacting partner protein strabismus, prevents convergent extension, yielding embryos with an abnormally short, broad neural plate that fails to complete neural tube closure. This defect in *Xenopus* closely resembles the NTDs

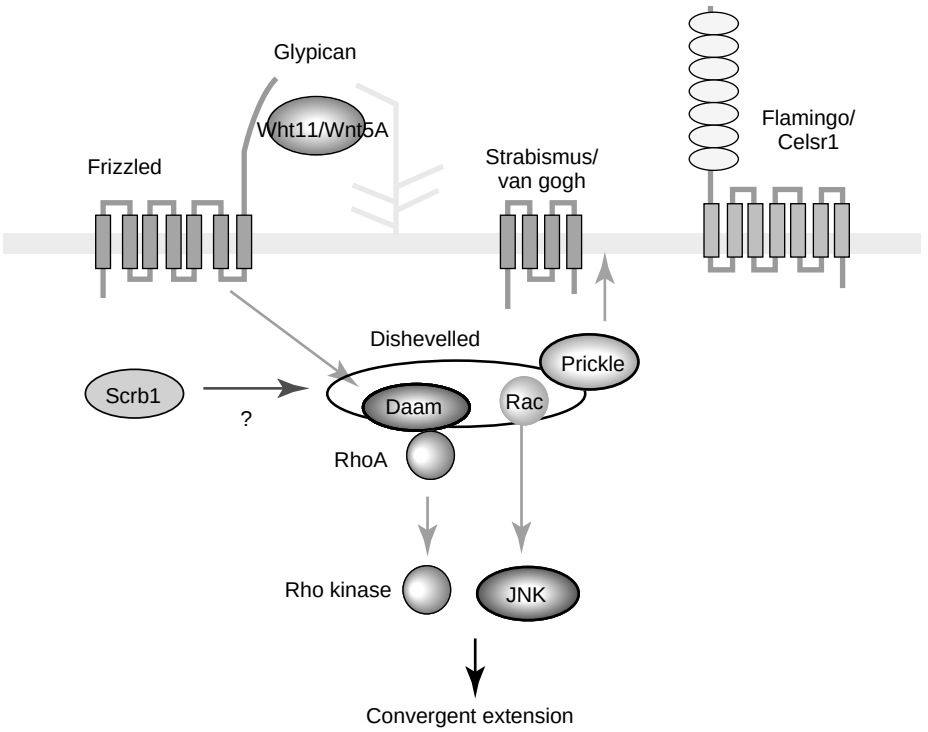

Fig. 4 The noncanonical Wnt (planar cell polarity) signaling pathway, which has been implicated in the regulation of convergent extension cell movements and initiation of neural tube closure (Closure 1). Mutations in the *strabismus/van gogh* (*Vangl2*), *Flamingo/Celsr1*, *Scrb1*, and *Dishevelled* genes all disturb Closure 1 in the mouse embryo, leading to the severe NTD craniorachischisis. Downstream signaling is thought to be via Rho kinase and/or jun-N-terminal (JNK) kinase. Modified from Ueno, N., Greene, N.D.E. (2003) *Birth Defects Res. (Part C)* **69**, 318–324.

in mouse embryos homozygous for the *loop-tail, circletail,* and *crash* mouse mutants, and in embryos lacking function of both *dishevelled1* and *dishevelled2* genes. In each of the mouse mutants, the neural plate is abnormally broad and has a non-bending region intervening between the neural folds, in contrast to the well-defined median hinge point of normal embryos. Although the neural folds elevate relatively normally, they are located too far apart to achieve closure (Fig. 5). Hence, convergent extension is required to establish a neural plate of width that is compatible with the medial bending essential for the initiation of closure.

The critical role of planar cell polarity in mammalian convergent extension and initiation of neurulation has become clear from recent genetic findings. All of the mouse mutants with failure of Closure 1 have proven to harbor mutations in components of the planar cell polarity pathway: *loop-tail* mice are mutant for the *strabismus/Vangl2* gene, *circletail* mice have mutations in *Scrb1,* and *crash* mice are mutant for *Celsr1,* a homologue of the *Drosophila* gene *flamingo*. It is striking,

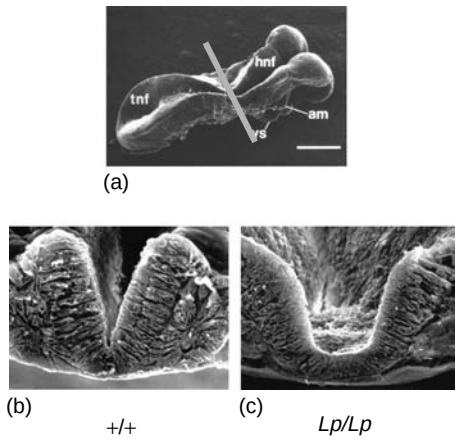

Fig. 5 Initiation of neurulation (Closure 1) in the mouse embryo, as studied by scanning electron microscopy. (a) Whole E8.5 embryo demonstrating the approaching neural folds approximately half way along the body axis, indicating the incipient Closure 1 event. This event occurs at the boundary of the hindbrain and cervical regions. (b, c) Sections transverse to the body axis as indicated by the red line in (a). The wild type embryo (b) shows "Mode 1" morphology with bending at a compact midline hinge point, and straight lateral neural folds. The homozygous *loop-tail* (*Lp/Lp*) embryo (c) shows a disturbance of the midline, which is broader than normal, without a compact bend in the neural plate. Although the neural folds elevate, laterally, they are not able to appose and fuse in the midline, leading to craniorachischisis in mutant embryos. Scale bar represents 0.25 mm (a) and 0.03 mm (b, c). Modified from Copp, A.J., Brook, F.A., Estibeiro, J.P., Shum, A.S.W., Cockroft, D.L. (1990) *Prog. Neurobiol.* **35**, 363–403; Greene, N.D.E., Gerrelli, D., Van Straaten, H.W.M., Copp, A.J. (1998) *Mech. Dev.* **73**, 59–72. (See color plate p. xxiv).

therefore, that the only group of mouse mutants that develop craniorachischisis all affect the same signaling pathway, suggesting a crucial role for the establishment of planar cell polarity in ensuring the successful initiation of neural tube closure.

3.2 Regulation of Neural Plate Bending: Mechanisms of Spina Bifida

In birds and mammals, bending of the neuroepithelium occurs at two specific sites in the neural plate: at the ventral midline (the median hinge point, MHP) and at paired dorsolateral bending sites near the junction of the neural plate and surface ectoderm (the dorsolateral hinge points, DLHPs). MHP bending creates the "neural groove," with a V-shaped cross-section, while DLHP bending creates longitudinal furrows that bring the lateral aspects of the neural plate toward each other in the dorsal midline.

Closure in the spinal region of the mouse embryo involves the progressive "zippering" of the neural tube down the spinal axis, until the posterior neuropore completes its closure at the future upper sacral level at E10. As the upper spine is forming, at embryonic day (E) 8.5, the closing neural folds exhibit an MHP but no DLHPs. By early E9.5, the neural plate begins to bend at DLHPs in addition to the MHP and, at the lowest spinal levels (at E10), the MHP disappears and the neural plate bends only at DLHPs. Hence, there is a complete switch from closure based on midline bending to closure dominated by dorsolateral bending during the course of mouse spinal neurulation (Fig. 6(a)–(d)).

Dorso-ventral molecular interactions have been implicated in the regulation of neural plate bending along the spinal region of the mouse embryo. A concentration-dependent influence of Shh, originating from the notochord and

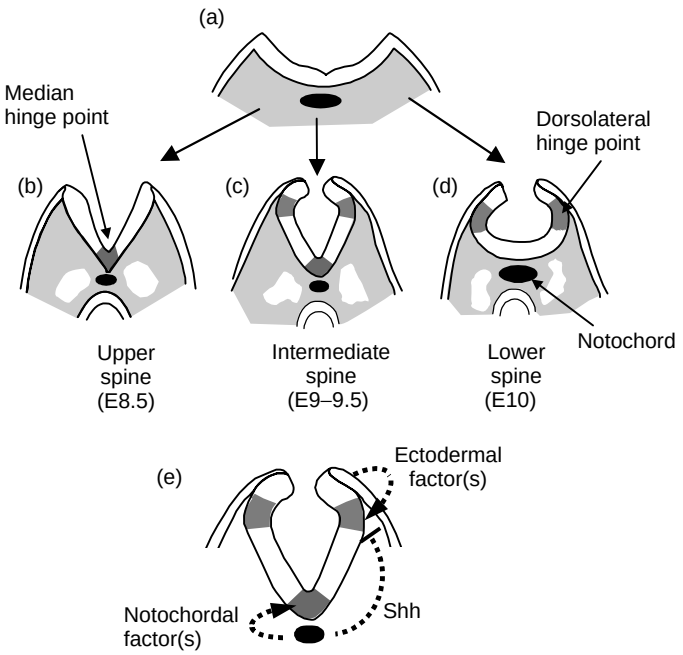

Fig. 6 Changing morphology of neurulation as neural tube closure progresses along the spinal region of the mouse embryo. The neural folds elevate from an initially flat morphology (a). In the upper spine (b), the neural plate bends only in the midline (Mode 1), at the median hinge point (MHP; see also Fig. 5(b)). In the intermediate spinal axis (c), both midline and dorsolateral bending are observed (Mode 2), whereas in the low spine (d), midline bending is lost and only dorsolateral hinge points (DLHPs) are observed (Mode 3). This changing pattern of neurulation is regulated by dorso-ventral patterning molecules (e). Notochordal factor(s) induce the MHP, while Shh, a secreted notochordal factor, inhibits DLHP formation. Factor(s) from the surface ectoderm induce DHLPs. As neurulation progresses along the spinal axis, the strength of notochordal signaling diminishes, with consequent loss of MHP-mediated bending and the emergence of DLHP-mediated neural tube closure. Modified from Ybot-Gonzalez, P., Cogram, P., Gerrelli, D., Copp, A.J. (2002) *Development* **129**, 2507–2517.

floor plate, patterns neuronal differentiation events along the dorso-ventral axis of the neural tube. BMPs (bone morphogenetic proteins) and Wnts, from the dorsal ectoderm, oppose Shh influence and promote dorsal patterning events. The MHP is induced by signals from the notochord, as demonstrated by experiments in which notochordal development was suppressed, with prevention of MHP formation. Shh, emanating from the notochord, may participate in the induction of the MHP, although midline bending is not completely abolished in Shh null embryos, suggesting a role for other, as yet unknown, notochordal inductive factors. Interestingly, although suppression of the notochord prevents MHP formation, the

neural tube still closes, suggesting the existence of a "default" mechanism of neural tube closure that operates in the absence of Shh influence (Fig. 6(e)).

An explanation for this observation has come from experimental studies with cultured mouse embryos, which show that Shh emanating from the notochord serves to inhibit DLHP formation in the upper spine of normal embryos. This inhibitory Shh influence counteracts a positive effect of the surface ectoderm that induces bending of the dorsal neural plate; in the absence of surface ectoderm, DLHPs are not formed. BMPs are possible mediators of this ectodermal influence, although this is speculative at present. As the wave of spinal neurulation passes down the body axis, the strength of Shh signaling from the underlying notochord lessens, enabling DLHP formation to "break through" at lower spinal levels. DLHP formation in the lower spine can be inhibited by local release of Shh peptide, demonstrating the negative regulation of dorsolateral bending by Shh. In the absence of Shh, as in notochordless embryos or in embryos genetically null for Shh, DLHP formation is de-inhibited and ensures closure of the neural tube at all levels of the body axis.

Hence, the precise pattern of bending of the neural plate during spinal neurulation is dictated by mutually inhibitory signals from the notochord and surface ectoderm. This conclusion is supported by the finding that mutations that increase the strength of Shh signaling (e.g. knockout of the *Patched1* gene) inhibit neural tube closure, presumably owing to an increased strength of Shh-mediated inhibition of DLHP formation, which is incompatible with the completion of closure of the posterior neuropore.

3.3 Molecular Regulation of Cranial Neurulation

Cranial closure is more susceptible to disturbance than spinal closure in the mouse, as evidenced by the increased occurrence of exencephaly compared to spina bifida among knockouts and mutants with NTD phenotypes. Similarly, more teratogenic agents induce mouse cranial NTDs than spinal NTD. Studies of gene knockout mice with cranial NTDs suggest several cell and molecular processes that appear crucial for the successful closure of the cranial neural folds. These include expansion of the cranial mesenchyme, function of the actin cytoskeleton, emigration of the cranial neural crest, and regulated balance between cell proliferation and cell death (Fig. 7). Knockout mice in which these events are disrupted have cranial but not spinal NTDs, suggesting that spinal neurulation is relatively independent of these mechanisms.

3.3.1 Cranial Mesenchyme

As the cranial neural folds elevate, there is a marked expansion of the cranial mesenchyme, with cell proliferation and a striking increase in extracellular space. This causes the neural folds to take on a bulging (biconvex) appearance, particularly in the midbrain. Mice with loss of function of the *Twist* or *Cart1* genes have cranial NTDs in which the principal defect appears to be a reduction in the proliferation and expansion of the cranial mesenchyme. Both genes are expressed in the mesenchyme, and chimeric analysis in the *Twist* mutant demonstrates that exencephaly develops only if there is a high proportion of *Twist*$^{-/-}$ cells in the cranial mesenchyme.

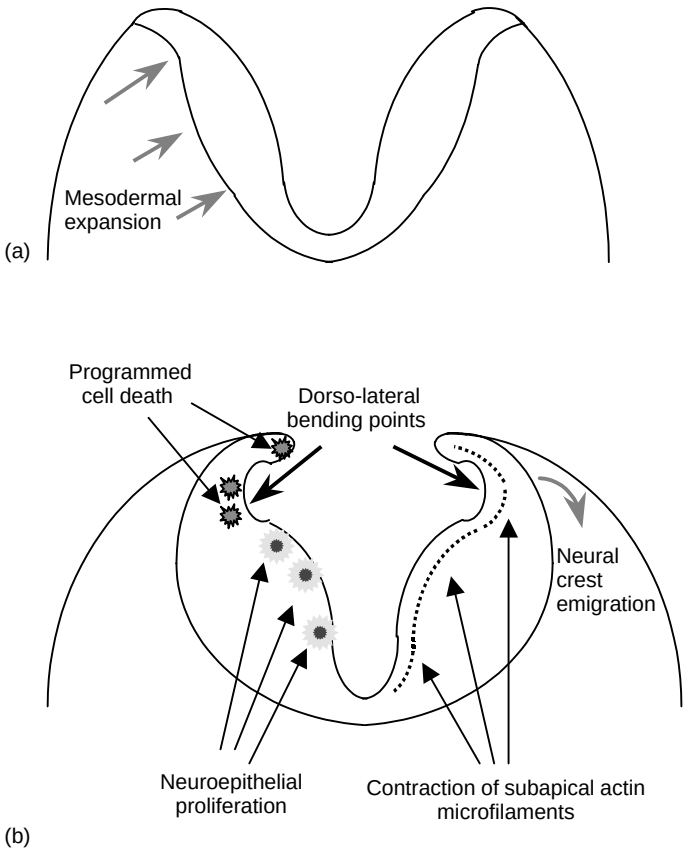

Fig. 7 Summary of the cellular and morphogenetic processes that are required for successful neural tube closure in the future brain. Cranial neural folds in the midbrain of the mouse embryo elevate initially with a biconvex morphology (a). Subsequently, dorsolateral bending occurs, causing the tips of the neural folds to converge on the midline (b), ensuring fusion and completion of cranial neurulation. Mesodermal expansion appears to be the most important morphogenetic factor in the initial elevation of the neural folds (a). However, the later stages of cranial neurulation require a number of additional processes including a functional actin cytoskeleton, successful initiation of neural crest emigration, precisely regulated programmed cell death and continuation of regulated neuroepithelial cell proliferation, with postponement of neuronal differentiation until after closure is complete (b). Defects in each of these processes are seen in one or more of the 80 gene knockout mice with NTDs. In each case, cranial neural tube closure fails and embryos develop exencephaly/anencephaly. Modified from Copp, A.J., Greene, N.D.E., Murdoch, J.N. (2003) *Nat. Rev. Genet.* **4**, 784–793.

3.3.2 Actin Cytoskeleton

In the second phase of cranial neural tube closure, the edges of the biconvex neural folds "flip around" and adopt a biconcave morphology, approaching the dorsal midline where fusion occurs to complete brain closure. This second neurulation phase is highly dependent on the actin cytoskeleton. Mice lacking the filamentous actin-associated protein shroom develop exencephaly, revealing a dependence of cranial neurulation on the actin microfilaments. This finding is consistent with *in vitro* studies in which rodent embryos are treated with cytochalasins, drugs that disassemble actin microfilaments. Cranial neural tube closure is inhibited by cytochalasins, whereas spinal neurulation proceeds to completion in the presence of these drugs. Hence, both genetic and nongenetic interventions demonstrate that the requirement for actin microfilaments is specific to cranial neural tube closure.

3.3.3 Neural Crest Emigration

The transition from biconvex to biconcave neural plate morphology also seems dependent on the initiation of cranial neural crest migration. In the midbrain and hindbrain, neural crest cells detach from the apices of the neural folds and start migration well in advance of neural tube closure, in contrast to the spinal region where neural crest emigration follows completion of neural tube closure. Mice overexpressing the gap junction protein connexin 43 exhibit marked cranial neural crest migration defects, and also develop exencephaly. Similarly, neural crest emigration and dorsolateral bending of the cranial neural folds are delayed in cultured rat embryos treated with chondroitinase ABC, to digest preexisting chondroitin sulphate. These findings suggest that neural crest emigration is a prerequisite for cranial neural tube closure, perhaps by increasing the flexibility of the lateral neural plate.

3.3.4 Programmed Cell Death

There is growing evidence to implicate a precise regulation of programmed cell death (apoptosis) in cranial neurulation. Dying cells occur within the dorsolateral regions of the cranial neural plate and at the tips of the fusing neural folds. Moreover, studies of knockout mice demonstrate that either an increase or decrease in the amount of this apoptosis is detrimental for cranial neurulation. Exencephaly develops in knockouts such as the *Apaf-1*, *caspase 9*, and *p*53 genes, which have reduced apoptotic cell death, and in knockouts of *AP-2*, *bcl10*, and *Tulp1*, which show increased apoptosis in the cranial neural folds. The precise role played by cell death in cranial neurulation is not clear. Dorsolateral cell death may synergize with neural crest cell emigration, to "loosen" the cranial neuroepithelium, thereby allowing the conversion from a biconvex to biconcave neural fold morphology. Additionally, apoptosis at the tips of the neural folds may play a quite different role. Once the apposing neural folds have contacted and adhered, midline epithelial remodeling is needed to break the continuity between neuroepithelium and surface ectoderm on each side, enabling the neural tube roof and its covering ectoderm to be formed at the midline. Inhibition of apoptosis using the peptide Zvad-fmk produces spinal NTDs in the chick embryo, probably by preventing this dorsal midline remodeling. Hence, apoptosis may play a dual role in cranial neurulation.

3.3.5 Neuronal Differentiation

The onset of neuronal differentiation occurs only after neural tube closure is

complete, during normal development, and this temporal sequence appears essential for successful cranial neurulation. Gene knockouts such as *RBP-Jκ*, *Hes1*, and *Numb*, which affect the Notch and related signaling pathways, exhibit failure of cranial neurulation in association with the premature onset of neuronal differentiation in the neural plate. The early appearance of differentiating neuroepithelial cells in these mutants might interfere with neurulation by rendering the neural plate mechanically inflexible, by interfering with the release of neural crest cells, or by inhibiting adhesion of the neural folds.

3.3.6 Cell Proliferation

Excessive cell proliferation has been implicated in some NTD mutants, such as mice with overexpression of Notch3, which produces increased numbers of neural progenitor cells in association with exencephaly. Neuroepithelial "overgrowth" has been suggested in relation to the exencephalic brains of other mutants but, when analyzed in detail, a normal or even lengthened cell cycle has generally been observed in exencephalic brains. It remains to be determined whether a disturbance of cell proliferation *per se* is a common cause of cranial NTDs.

4 Prevention of NTDs by Folic Acid and Other Vitamins

Much interest was stimulated by the finding that supplementation with folic acid can prevent a proportion of NTDs, estimated at around 70%, in the UK Medical Research Council Vitamin Research Group randomized clinical trial of NTD recurrence. Folic acid prevented 40 to 85% in a nonrandomized supplementation study in China, and there were no cases of NTD among 2394 pregnancies that received a folic acid–containing multivitamin supplement, compared with six out of 2310 unsupplemented pregnancies, in a randomized controlled trial of NTD occurrence in Hungary. Fortification of bread flour with folic acid in the United States has been associated with a 19% reduction in the prevalence of NTD. The variation in the proportion of cases prevented by folic acid in the various studies may relate to the differing doses of folic acid used. Indeed, NTD risk appears related to the log (serum folate level) in humans. Nevertheless, reports of successive NTD pregnancies in a single family despite high-dose folate intake suggest that a proportion of NTD cases may be resistant to folate, perhaps reflecting heterogeneity of etiology, with some subtypes of NTD being sensitive to folic acid whereas others are folate resistant.

4.1 Animal Models of Folate-preventable NTDs

Supporting evidence for the existence of folic acid–sensitive and folic acid–resistant NTDs comes from studies of mouse genetic mutant strains with NTDs. In some mutants, NTDs can be prevented by folic acid (e.g. *splotch, crooked tail, Cart1, Cited2*). Moreover, exencephaly has been reported in mice lacking the folate transport protein encoded by the gene *Folbp1*. This gene knockout causes lethality around the time of neural tube closure, hampering an assessment of whether *Folbp1* is essential for neural tube closure. However, treatment of the pregnant dams with folic acid rescues survival in these mice, and a proportion of the surviving homozygous knockouts exhibit exencephaly later in gestation. It

is interesting that spina bifida is not part of the *Folbp1* knockout phenotype, even though folic acid appears to prevent both spina bifida and anencephaly among human NTDs. Moreover, analysis of the folate receptors α and β have revealed no association with the NTD phenotype in a series of human pregnancies.

Other genes that are involved in folate transport and metabolism have not yet been associated with neural tube closure defects in mice. For example, MTHFR, which has been implicated as a risk factor in human NTDs (see Section 4.1), does not cause NTDs when inactivated in mice.

4.2 Folate-resistant NTDs and Inositol as an Adjunct Therapy

While there is ample evidence from experimental studies to support a preventive effect of folic acid on NTDs, some mouse NTD models have proven insensitive to folic acid therapy, for example, the *curly tail, axial defects,* and *ephrin-A5* mutants. *Curly tail* is the best-characterized mouse model of folate-resistant NTD, with information available on the embryonic mechanism of neural tube closure failure. Recently, *Ghl3* was identified as a strong candidate for the gene mutated in *curly tail*. The spinal NTDs in this mutant are not altered in frequency by administration of folic acid, folinic acid (5-formyl tetrahydrofolate), Pregnavite Forte F (a multivitamin preparation active in preventing human NTDs), or methionine. Interestingly, however, a significant reduction in frequency of spinal NTD, with a particular reduction in the incidence of the most severe defect lumbosacral spina bifida, is seen after maternal administration of *myo*-inositol and *D-chiro*-inositol. *Curly tail* embryos are also at increased risk of developing exencephaly under conditions of inositol deficiency, compared to nonmutant embryos.

The preventive effect of inositol on spinal NTD in *curly tail* mice is mediated via intracellular signaling through the inositol phospholipid cycle, by stimulation of protein kinase C (PKC) activity. A role for PKC is demonstrated by the ability of the PKC agonist, 12-O-tetradecanoylphorbol-13-acetate (TPA), to substitute for *myo*-inositol in preventing the NTD. On the other hand, the broad spectrum inhibitor of PKC action, GF109203X, blocks prevention both by *myo*-inositol and TPA. Recent studies have identified PKCβI, γ or ζ, but not other isoforms of PKC, as essential for mediating the preventive effect of inositol on NTDs. The PKCβI isoform is required for stimulation of cell proliferation in the hindgut endoderm of the caudal region of *curly tail* embryos, which counteracts the genetically determined cell proliferation defect, and rescues the phenotype.

Hence, the pathway that mediates the preventive action of *myo*-inositol on NTD in the *curly tail* mouse is becoming understood. The finding of inositol as an alternative to folic acid for prevention of mouse NTDs raises the possibility of a novel future adjunct therapy to folic acid for use in human pregnancy.

5 Conclusions

Spina bifida is a member of the group of congenital malformations known as NTDs. A combination of studies of humans and animal models is contributing to our increasing knowledge of the etiology of NTDs, with identification of key genetic and nongenetic influences, and a growing appreciation of the types of interaction that

promote or ameliorate the risk of NTDs. Study of mouse models is shedding light on the developmental mechanisms that underlie the different categories of NTDs, with marked variations in mechanism along the body axis. Primary prevention of NTDs is now a reality with introduction of folic acid, which has a preventive effect on a proportion of NTD cases. Folate-resistant NTDs require other strategies for prevention, and inositol appears promising in studies of one-mouse model.

See also Brain Development.

Bibliography

Books and Reviews

Bock, G., Marsh, J. (Eds.) (1994) *Neural Tube Defects*, Ciba Foundation Symposium 181. John Wiley & Sons, Chichester, UK.

Copp, A.J., Brook, F.A., Estibeiro, J.P., Shum, A.S.W., Cockroft, D.L. (1990) The embryonic development of mammalian neural tube defects, *Prog. Neurobiol.* **35**, 363–403.

Copp, A.J., Greene, N.D.E., Murdoch, J.N. (2003) The genetic basis of mammalian neurulation, *Nat. Rev. Genet.* **4**, 784–793.

Harding, B.N., Copp, A.J. (2002) Congenital Malformations, in: Graham, D.l., Lantos, P.L. (Eds.) *Greenfield's Neuropathology*, 7th edition, Arnold, London, pp. 357–483.

Schoenwolf, G.C., Smith, J.L. (1990) Mechanisms of neurulation: Traditional viewpoint and recent advances, *Development* **109**, 243–270.

Primary Literature

Barber, R.C., Shaw, G.M., Lammer, E.J., Greer, K.A., Biela, T.A., Lacey, S.W., Wasserman, C.R., Finnell, R.H. (1998) Lack of association between mutations in the folate receptor-α gene and spina bifida, *Am. J. Med. Genet.* **76**, 310–317.

Berry, R.J., Li, Z., Erickson, J.D., Li, S., Moore, C.A., Wang, H., Mulinare, J., Zhao, P., Wong, L.Y.C., Gindler, J., Hong, S.X., Correa, A., China-US Collaborative Project Neu. (1999) Prevention of neural-tube defects with folic acid in China, *N. Engl. J. Med.* **341**, 1485–1490.

Botto, L.D., Moore, C.A., Khoury, M.J., Erickson, J.D. (1999) Neural-tube defects, *N. Engl. J. Med.* **341**, 1509–1519.

Carter, C.O. (1974) Clues to the aetiology of neural tube malformations, *Dev. Med. Child Neurol.* **16**(()Suppl. 32), 3–15.

Cavalli, P., Copp, A.J. (2002) Inositol and folate-resistant neural tube defects, *J. Med. Genet.* **39**, e5.

Chen, W.-H., Morriss-Kay, G.M., Copp, A.J. (1994) Prevention of spinal neural tube defects in the curly tail mouse mutant by a specific effect of retinoic acid, *Dev. Dyn.* **199**, 93–102.

Chen, Z.T., Karaplis, A.C., Ackerman, S.L., Pogribny, I.P., Melnyk, S., Lussier-Cacan, S., Chen, M.F., Pai, A., John, S.W.M., Smith, R.S., Bottiglieri, T., Bagley, P., Selhub, J., Rudnicki, M.A., James, S.J., Rozen, R. (2001) Mice deficient in methylenetetrahydrofolate reductase exhibit hyperhomocysteinemia and decreased methylation capacity, with neuropathology and aortic lipid deposition, *Hum. Mol. Genet.* **10**, 433–443.

Chen, Z.-F., Behringer, R.R. (1995) *Twist* is required in head mesenchyme for cranial neural tube morphogenesis, *Genes Dev.* **9**, 686–699.

Cockroft, D.L., Brook, F.A., Copp, A.J. (1992) Inositol deficiency increases the susceptibility to neural tube defects of genetically predisposed (curly tail) mouse embryos in vitro, *Teratology* **45**, 223–232.

Cogram, P., Hynes, A., Dunlevy, L.P.E., Greene, N.D.E., Copp, A.J. (2004) Specific isoforms of protein kinase C are essential for prevention of folate-resistant neural tube defects by inositol, *Hum. Mol. Genet.* **13**, 7–14.

Cogram, P., Tesh, S., Tesh, J., Wade, A., Allan, G., Greene, N.D.E., Copp, A.J. (2002) D-*chiro*-inositol is more effective than *myo*-inositol in preventing folate-resistant mouse neural tube defects, *Hum. Reprod.* **17**, 2451–2458.

Copp, A.J., Greene, N.D.E., Murdoch, J.N. (2003) Dishevelled: linking convergent extension with neural tube closure, *Trends Neurosci.* **26**, 453–455.

Creasy, M.R., Alberman, E.D. (1976) Congenital malformations of the central nervous system

in spontaneous abortions, *J. Med. Genet.* **13**, 9–16.

Curtin, J.A., Quint, E., Tsipouri, V., Arkell, R.M., Cattanach, B., Copp, A.J., Fisher, E.M., Nolan, P.M., Steel, K.P., Brown, S.D.M., Gray, I.C., Murdoch, J.N. (2003) Mutation of *Celsr1* disrupts planar polarity of inner ear hair cells and causes severe neural tube defects in the mouse, *Curr. Biol.* **13**, 1–20.

Czeizel, A.E., Dudás, I. (1992) Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation, *N. Engl. J. Med.* **327**, 1832–1835.

Erickson, C.A., Weston, J.A. (1983) AN SEM analysis of neural crest migration in the mouse, *J. Embryol. Exp. Morphol.* **74**, 97–118.

Estibeiro, J.P., Brook, F.A., Copp, A.J. (1993) Interaction between splotch (*Sp*) and curly tail (*ct*) mouse mutants in the embryonic development of neural tube defects, *Development* **119**, 113–121.

Finnell, R.H., Spiegelstein, O., Wlodarczyk, B., Triplett, A., Pogribny, I.P., Melnyk, S., James, J.S. (2002) DNA methylation in Folbp1 knockout mice supplemented with folic acid during gestation, *J. Nutr.* **132**, 2457S–2461S.

Fleming, A., Copp, A.J. (2000) A genetic risk factor for mouse neural tube defects: defining the embryonic basis, *Hum. Mol. Genet.* **9**, 575–581.

Greene, N.D.E., Copp, A.J. (1997) Inositol prevents folate-resistant neural tube defects in the mouse, *Nat. Med.* **3**, 60–66.

Greene, N.D.E., Copp, A.J. (2002) Animal Models of Folate-Related Neural Tube Defects, in: Massaro, E.J., Rogers, J.M. (Eds.) *Folate and Human Development*, Humana Press, Totowa, NJ, pp. 1–26.

Greene, N.D.E., Gerrelli, D., Van Straaten, H.W.M., Copp, A.J. (1998) Abnormalities of floor plate, notochord and somite differentiation in the *loop-tail* (*Lp*) mouse: a model of severe neural tube defects, *Mech. Dev.* **73**, 59–72.

Hall, J.L., Harris, M.J., Juriloff, D.M. (1997) Effect of multifactorial genetic liability to exencephaly on the teratogenic effect of valproic acid in mice, *Teratology* **55**, 306–313.

Hildebrand, J.D., Soriano, P. (1999) Shroom, a PDZ domain-containing actin-binding protein, is required for neural tube morphogenesis in mice, *Cell* **99**, 485–497.

Holmes, L.B., Harvey, E.A., Coull, B.A., Huntington, K.B., Khoshbin, S., Hayes, A.M., Ryan, L.M. (2001) The teratogenicity of anticonvulsant drugs, *N. Engl. J. Med.* **344**, 1132–1138.

Honein, M.A., Paulozzi, L.J., Mathews, T.J., Erickson, J.D., Wong, L.Y.C. (2001) Impact of folic acid fortification of the US food supply on the occurrence of neural tube defects, *JAMA* **285**, 2981–2986.

Jessell, T.M. (2000) Neuronal specification in the spinal cord: inductive signals and transcriptional codes, *Nat. Rev. Genet.* **1**, 20–29.

Juriloff, D.M., Harris, M.J., Tom, C., MacDonald, K.B. (1991) Normal mouse strains differ in the site of initiation of closure of the cranial neural tube, *Teratology* **44**, 225–233.

Kucera, J. (1971) Rate and type of congenital anomalies among offspring of diabetic women, *J. Reprod. Med.* **7**, 61–70.

Lardelli, M., Williams, R., Mitsiadis, T., Lendahl, U. (1996) Expression of the Notch 3 intracellular domain in mouse central nervous system progenitor cells is lethal and leads to disturbed neural tube development, *Mech. Dev.* **59**, 177–190.

Lee, K.J., Jessell, T.M. (1999) The specification of dorsal cell fates in the vertebrate central nervous system, *Annu. Rev. Neurosci.* **22**, 261–294.

Morriss, G.M., Solursh, M. (1978) Regional differences in mesenchymal cell morphology and glycosaminoglycans in early neural-fold stage rat embryos, *J. Embryol. Exp. Morphol.* **46**, 37–52.

Morriss-Kay, G., Tan, S.-S. (1987) Mapping cranial neural crest cell migration pathways in mammalian embryos, *Trends Genet.* **3**, 257–261.

Morriss-Kay, G.M. (1981) Growth and development of pattern in the cranial neural epithelium of rat embryos during neurulation, *J. Embryol. Exp. Morphol.* **65**(Suppl.), 225–241.

Morriss-Kay, G.M., Tuckett, F. (1985) The role of microfilaments in cranial neurulation in rat embryos: effects of short-term exposure to cytochalasin D, *J. Embryol. Exp. Morphol.* **88**, 333–348.

Murdoch, J.N., Doudney, K., Paternotte, C., Copp, A.J., Stanier, P. (2001) Severe neural tube defects in the *loop-tail* mouse result from mutation of *Lpp1*, a novel gene involved in

floor plate specification, *Hum. Mol. Genet.* **10**, 2593–2601.

Murdoch, J.N., Henderson, D.J., Doudney, K., Gaston-Massuet, C., Phillips, H.M., Paternotte, C., Arkell, R., Stanier, P., Copp, A.J. (2003) Disruption of *scribble* (*Scrb1*) causes severe neural tube defects in the *circletail* mouse, *Hum. Mol. Genet.* **12**, 87–98.

Nakatsu, T., Uwabe, C., Shiota, K. (2000) Neural tube closure in humans initiates at multiple sites: evidence from human embryos and implications for the pathogenesis of neural tube defects, *Anat. Embryol.* **201**, 455–466.

O'Rahilly, R., Müller, F. (2002) The two sites of fusion of the neural folds and the two neuropores in the human embryo, *Teratology* **65**, 162–170.

Piedrahita, J.A., Oetama, B., Bennett, G.D., Van Waes, J., Kamen, B.A., Richardson, J., Lacey, S.W., Anderson, R.G.W., Finnell, R.H. (1999) Mice lacking the folic acid-binding protein Folbp1 are defective in early embryonic development, *Nat. Genet.* **23**, 228–232.

Sadler, T.W. (1980) Effects of maternal diabetes on early embryogenesis: II. Hyperglycemia-induced exencephaly, *Teratology* **21**, 349–356.

Schluter, G. (1973) Ultrastructural observations on cell necrosis during formation of the neural tube in mouse embryos, *Z. Anat. Entwickl.-Gesch.* **141**, 251–264.

Seller, M.J. (1981) Recurrence risks for neural tube defects in a genetic counselling clinic population, *J. Med. Genet.* **18**, 245–248.

Seller, M.J., Embury, S., Polani, P.E., Adinolfi, M. (1979) Neural tube defects in curly-tail mice. II. Effect of maternal administration of vitamin A., *Proc. R. Soc. Lond. B* **206**, 95–107.

Shenefelt, R.E. (1972) Morphogenesis of malformations in hamster caused by retinoic acid: relation to dose and stage at treatment, *Teratology* **5**, 103–108.

Shum, A.S.W., Copp, A.J. (1996) Regional differences in morphogenesis of the neuroepithelium suggest multiple mechanisms of spinal neurulation in the mouse, *Anat. Embryol.* **194**, 65–73.

Ting, S.B., Wilanowski, T., Auden, A., Hall, M., Voss, A.K., Thomas, T., Parekh, V., Cunningham, J.M., Jane, S.M. (2003) Inositol and folate-resistant neural tube defects in mice lacking the epithelial-specific factor Grhl-3, *Nat. Med.* **9**, 1513–1519.

Van der Put, N.M.J., Eskes, T.K.A.B., Blom, H.J. (1997) Is the common 677C→T mutation in the methylenetetrahydrofolate reductase gene a risk factor for neural tube defects? A meta-analysis, *Q. J. Med.* **90**, 111–115.

Van Straaten, H.W.M., Copp, A.J. (2001) Curly tail: a 50-year history of the mouse spina bifida model, *Anat. Embryol.* **203**, 225–237.

Wald, N., Sneddon, J., Densem, J., Frost, C., Stone, R., MRC Vitamin Study Res Group. (1991) Prevention of neural tube defects: results of the medical research council vitamin study, *Lancet* **338**, 131–137.

Wald, N.J., Law, M.R., Morris, J.K., Wald, D.S. (2001) Quantifying the effect of folic acid, *Lancet* **358**, 2069–2073.

Weil, M., Jacobson, M.D., Raff, M.C. (1997) Is programmed cell death required for neural tube closure, *Curr. Biol.* **7**, 281–284.

Wilson, A., Platt, R., Wu, Q., Leclerc, D., Christensen, B., Yang, H., Gravel, R.A., Rozen, R. (1999) A common variant in methionine synthase reductase combined with low cobalamin (vitamin B12) increases risk for spina bifida, *Mol. Genet. Metab.* **67**, 317–323.

Wilson, D.B. (1980) Cellular proliferation in the exencephalic brain of the mouse embryo, *Brain Res.* **195**, 139–148.

Ybot-Gonzalez, P., Cogram, P., Gerrelli, D., Copp, A.J. (2002) Sonic hedgehog and the molecular regulation of neural tube closure, *Development* **129**, 2507–2517.

Ybot-Gonzalez, P., Copp, A.J. (1999) Bending of the neural plate during mouse spinal neurulation is independent of actin microfilaments, *Dev. Dyn.* **215**, 273–283.

Zhao, Q., Behringer, R.R., De Crombrugghe, B. (1996) Prenatal folic acid treatment suppresses acrania and meroanencephaly in mice mutant for the Cart1 homeobox gene, *Nat. Genet.* **13**, 275–283.

Neuron Chemistry

Bonnie L. Firestein
The State University of New Jersey, Piscataway, NJ, USA

1 Parts of the Neuron 141
1.1 Functional Domains of the Neuron 141
1.2 The Synapse 142

2 The Presynaptic Component of a Synapse 143
2.1 Neurotransmitters 143
2.1.1 Classes of Neurotransmitters 143
2.1.2 The Main Classical Neurotransmitters in the Mammalian CNS 144
2.2 Mechanisms of Neurotransmitter Release 146
2.2.1 Uptake of Neurotransmitters into Synaptic Vesicles 146
2.2.2 Docking and Priming 147
2.2.3 Triggering and Exocytosis 147
2.2.4 Endocytosis and Vesicle Recycling 148

3 The Postsynaptic Component of a Synapse 149
3.1 The Postsynaptic Density 149
3.2 Dendritic Spines 149
3.3 Excitatory Synapses: Glutamate Receptors 149
3.3.1 NMDA Receptors 150
3.3.2 AMPA Receptors 150
3.3.3 Kainate and δ Receptors 151
3.3.4 Metabotropic Glutamate Receptors 151
3.4 Inhibitory Synapses: GABA Receptors 152
3.4.1 $GABA_A$ Receptors 152
3.4.2 $GABA_B$ Receptors 152
3.4.3 $GABA_C$ Receptors 152

4 Changes in Neuronal Morphology in Response to Signals 153
4.1 Signals that Regulate Dendrite Patterning 153

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

4.1.1 Electrical Activity 153
4.1.2 Neurotrophins 153
4.1.3 Extracellular Signaling Proteins 153
4.1.4 Intracellular Signaling Molecules 154
4.2 Signals that Regulate Axons 155
4.2.1 Electrical Activity 155
4.2.2 Extracellular Cues: Netrins 155
4.2.3 Extracellular Cues: Semaphorins 156
4.2.4 Extracellular Cues: Ephrins 157
4.2.5 Extracellular Cues: Slits 157
4.2.6 Extracellular Cues: CPG15 157
4.3 Intracellular Signaling Molecules 158

5 Perspectives 158

Acknowledgments 158

Bibliography 158
Books and Reviews 158
Primary Literature 158

Keywords

Axon
The long narrow part of a neuron that carries action potentials away from the cell body and finally transmits signals to other neurons via synaptic connections. The axon forms the presynaptic component of many synapses.

Dendrite
The part of a neuron that is a cytoplasmic extension of the cell body. Groups of dendrites often have a main trunk, for example, the apical dendrite, and many branches. This structure is called a *dendritic tree*. Dendrites form the postsynaptic component of most synapses.

PDZ Domain
A protein–protein interaction domain named for homologous regions in Postsynaptic protein 95, *Drosophila* discs-large, and zona-occludens 1.

Postsynaptic Neuron
The neuron on the receiving side of information flow across a synapse.

Postsynaptic Density
An electron dense region found immediately underneath the postsynaptic plasma membrane.

Presynaptic Neuron
The neuron on the transmitting side of information flow across a synapse.

Synapse
The specialized zone of contact at which one neuron communicates with another.

Synaptic Cleft
The small extracellular region between sites of signaling between neurons.

Synaptic Terminal or Bouton
The end of an axon that contains neurotransmitters stored in vesicles, which are released when the vesicles fuse with the membrane.

The brain consists of billions of neurons, cells that have specialized morphologies for precise communication. This communication takes place at synapses, which can be chemical or electrical. The focus here is chemical synapses, sites where one neuron, the presynaptic neuron, sends chemical signals called *neurotransmitters* that bind to receptors on the second, or postsynaptic, neuron. This binding then transduces signals to the postsynaptic neuron, allowing communication to be propagated.

To understand how the brain functions and how signaling processes go awry during pathophysiological conditions, the players in this signaling must be identified. Recently, neurobiologists have uncovered the identities of proteins that play a major role in initiating release of neurotransmitter from the presynaptic neuron. They have also elucidated the role of scaffolding proteins in setting up signaling networks in the postsynaptic neuron. Although by no means comprehensive, this article summarizes currently accepted theories that form the basis for our understanding of neurotransmitter-mediated neuronal signaling.

1 Parts of the Neuron

1.1 Functional Domains of the Neuron

The neuron is a polarized cell; that is, it has specialized domains that allow for information flow. The largest part of the neuron is the cell body or soma. The soma contains organelles necessary for basic cellular function. For example, the nucleus, responsible for gene expression, resides in the soma. In addition, unlike other parts of the neuron, if the soma is damaged, the neuron cannot recover and will die. Besides playing a role in basic function, the soma can serve to receive synaptic signals from other neurons.

On most neurons, there are multiple branches that often derive from the soma. These branches are known as *dendrites*. The majority of synapses occur on dendrites, which typically form the postsynaptic component of synapses. The extent

of dendritic branching can be influenced by factors such as learning, which increases branching, or alcohol exposure or disease, which often decrease the number of dendrites. This has important physiological implications. Neurons that have a greater number of branches have more sites for receiving information from other neurons. In fact, in conditions such as mental retardation or Alzheimer's disease, there are fewer branches on neurons in the hippocampus and cortex, areas important for learning and memory.

A neuron also must contain a site where it can send signals to other neurons. The part of a neuron that is specialized for carrying signals long distances from the cell body and storing and releasing neurotransmitters from its terminals is called the *axon* (Fig. 1). It is a single extension, or neurite, from the cell body that is much longer and thinner than the dendrites. Like dendrites, axons may have branches, or collaterals, which generally occur at a distance from the soma. In addition, when neurons are damaged, axons may "sprout" or branch to form new synapses and replace those destroyed. Neurotransmitters are stored in vesicles that are present in the tips of axons. These tips swell and are called *synaptic boutons* or *terminals*. When the neuron transmits signals to its postsynaptic partner, the vesicles fuse with the plasma membrane and the neurotransmitter is dumped into the synaptic cleft.

A majority of axons, especially those with longer diameters, are myelinated; that is, they are covered with a lipid and proteinaceous sheath. The myelin serves as an electrical insulator. There are small areas of myelinated axons that do not have myelin that are called *Nodes of Ranvier*. The myelin and the Nodes of Ranvier work together to increase the speed that electrical impulses travel along axons. Thus, heavily myelinated axons conduct electric impulses faster than unmyelinated axons. For example, neurons involved in the sensation of pain have unmyelinated axons, while neurons involved in tactile sensations, such as vibration and two point discrimination, are myelinated. This results in a time delay between pain sensation and other touch sensations associated with a stimulus.

1.2 The Synapse

Neuronal communication occurs at the synapse. Synapses consist of a presynaptic component, which is the axon terminal, the synaptic cleft, which is typically

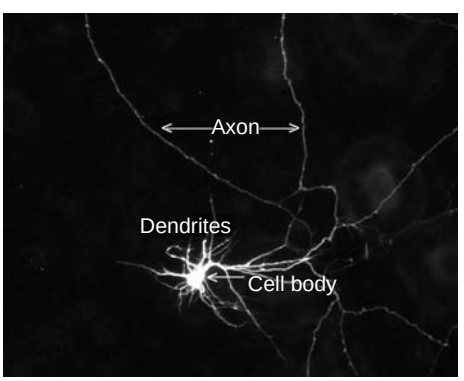

Fig. 1 Functional domains of a neuron. Morphology of a single hippocampal neuron is shown by expression of a fluorescent protein inside of the neuron. Dendrites, which contain sites for receiving information, are relatively short protrusions from the cell body, or soma. There is a single axon that extends from within the figure to outside of the figure and reenters the field shown. The axon is the longest protrusion from the cell body. This axon releases neurotransmitter at its end, which is not visible in the figure.

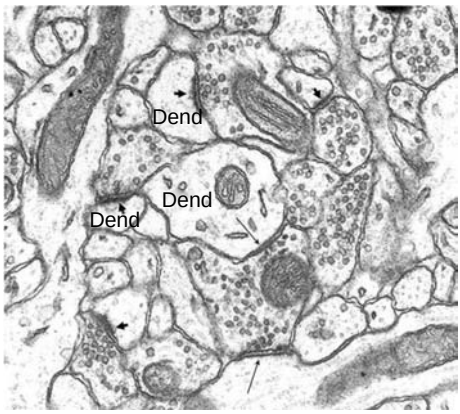

Fig. 2 Electron micrograph of mammalian central nervous system synapses. Note that axons contain small dark circles. These are vesicles filled with neurotransmitters. Some dendrites are denoted as dend. Thick arrows indicate asymmetric excitatory synapses. Note the dark electron rich areas just beneath the postsynaptic membrane. These areas are postsynaptic densities. Thin arrows point to symmetric (GABA) synapses. Note the absence of thick postsynaptic densities at these synapses. Electron micrograph is courtesy of Dr. Kristin Harris, Synapse Web, Medical College of Georgia, http://synapses.mcg.edu/

the small space between the two communicating neurons, and a postsynaptic component, which is typically a dendrite. As discussed above, neurotransmitter is released into the cleft, and it binds to receptors on the postsynaptic site. This binding triggers a signaling cascade in the postsynaptic neuron, which can lead to changes in the postsynaptic neuron or release of neurotransmitter from this postsynaptic neuron at other synapses sending signals to yet another neuron. Thus, depending on how synapses are arranged in the brain, signaling networks are established that lead to physiological responses.

Electron microscopists noticed that not all synapses are alike. There are two main forms. Synapses are classified as symmetric or asymmetric on the basis of the thickness of the postsynaptic density (Fig. 2). Inhibitory synapses (those that use GABA or glycine as a neurotransmitter) possess an axon terminal active zone and a postsynaptic region that are of equal thickness. Hence, they are symmetric synapses. Excitatory synapses of the CNS, the majority of which use glutamate as a neurotransmitter, have a thicker electron-dense region adjacent to the postsynaptic plasma membrane, making the thickness of the postsynaptic density greater than that of the presynaptic density. Hence, excitatory synapses are termed *asymmetric*.

2 The Presynaptic Component of a Synapse

2.1 Neurotransmitters

2.1.1 Classes of Neurotransmitters

There are two types of neurotransmitters that can be used for neuronal communication: classical and nonclassical. There are four main criteria by which neurotransmitters are judged to be classical (Table 1). First, the neurotransmitter must be synthesized in the neuron. Second, the neurotransmitter must be present in the presynaptic terminal and released in amounts sufficient to exert an effect on the postsynaptic cell. In general, neurotransmitters are produced in the cytoplasm of the axon terminal and are taken up into vesicles by specific transporters. Third, the neurotransmitter must be in a chemically identifiable form and, when administered exogenously, it must mimic the effects of

Tab. 1 Criteria for classical neurotransmitters.

Criteria for classical neurotransmitters
1. The neurotransmitter must be synthesized in the neuron.
2. The neurotransmitter must be present in the presynaptic terminal and released in amounts sufficient to exert an effect on the postsynaptic cell.
3. The neurotransmitter must be in a chemically identifiable form and, when administered exogenously, it must mimic the effects of the endogenous neurotransmitter.
4. There must be some type of removal from the synaptic cleft.

the endogenous neurotransmitter. Fourth, there must be some type of mechanism for removal from the synaptic cleft. This removal can occur in a number of ways. Neurotransmitter can diffuse away from the cleft, thus making it inaccessible to the postsynaptic receptors. Alternatively, some clefts contain specific enzymes that degrade the neurotransmitter. Additionally, some neurotransmitters have a system for reuptake into the presynaptic terminal. Finally, glial cells and neurons have neurotransmitter transporters in their plasma membranes, which allow for the intact neurotransmitter to be taken up by either of these cell types. These mechanisms allow for finite action by the neurotransmitter on the postsynaptic cell.

Nonclassical neurotransmitters do not adhere to all of the main criteria discussed above. These neurotransmitters fall into two general groups: peptide neurotransmitters and gaseous neurotransmitters (Tab. 2). Peptide neurotransmitters are usually synthesized in the cell body, where they are packaged into vesicles and transported to the axon terminal. Peptides are usually synthesized as larger inactive precursor molecules, which are then cleaved to produce smaller active molecules. Like classical neurotransmitters, peptides are released into the cleft and are inactivated by either diffusion or enzymatic cleavage. *Gaseous neurotransmitters*, also sometimes referred to as *neuromodulators*, include nitric oxide and carbon monoxide. These transmitters are synthesized in the cytosol of the postsynaptic neuron in which they can then act on enzymes or binding proteins to participate in signaling in the postsynaptic neuron. Additionally, since they are gases, these neurotransmitters can diffuse across the membrane and to the presynaptic neuron. This is sometimes referred to as *retrograde signaling*.

2.1.2 The Main Classical Neurotransmitters in the Mammalian CNS

Catecholamines The catecholamines are derived from tyrosine and include dopamine (DA), norepinephrine (NE or noradrenaline), and epinephrine (E or adrenaline) (Tab. 2). DA acts as an inhibitory neurotransmitter in the central nervous system (CNS). NE can acts as an excitatory neurotransmitter when it binds to α_1 or β_1 receptors and as an inhibitory neurotransmitter when it binds to α_2 or β_2 receptors. Dopaminergic neurons produce DA when tyrosine is converted to L-DOPA by tyrosine hydroxylase (TH), which is then converted to DA by L-aromatic amino acid decarboxylase (L-AADC). If dopamine-β-carboxylase (DBH) is present, all DA is converted to NE; if DBH is not present, the neuron produces DA as a neurotransmitter. Similarly, if the neuron contains phenylethanolamine *N*-methyltransferase (PMNT), all NE is converted to E. Catecholamines are cleared from the cleft by transporters and are degraded by enzymes, such as monoamine oxidase (MAO).

In the past 15 to 20 years, there has been great interest in DA and its role in the control of voluntary movement. Deficiencies in DA lead to tremors and rigidity, often manifesting as Parkinson's disease. There has been a major effort to cure this disease. Patients with Parkinson's disease are often given L-DOPA, which relieves the symptoms for a short time. However, a severe side effect often occurs – Huntington's disease–like symptoms, which are frequently due to excess DA or L-DOPA. Alternatively, there has been some success in both animal models and humans with gene therapy. Since Parkinson's disease occurs because of a deficiency in TH, strategies have included grafting of fetal tissue, which can differentiate, express TH, and produce DA, or of fibroblasts genetically engineered to express TH. These strategies are still being refined.

In contrast to Parkinson's disease, an excess of DA is thought to contribute to schizophrenia, which manifests as irrational thought, delusions, and hallucinations. However, this disease is thought to be multifactorial. Other neurotransmitters, such as glutamate and acetylcholine, may also contribute to the pathophysiology of this disease.

Serotonin Serotonin is produced when tryptophan is metabolized by tryptophan hydroxylase, yielding 5-Hydroxy L-Tryptophan (5-HTP) (Tab. 2). 5-HTP is then acted upon by L-AADC to yield 5-hydroxytryptophan (5-HTT), or serotonin. Once released into the cleft, where it can bind to receptors on both the pre- and postsynaptic neuron, it is taken up by the neuron by a high-affinity transporter, and degraded in the neuron by MAO. Serotonin has an inhibitory effect when it binds to 5-HT_1 receptors and an excitatory effect when it binds to 5-HT_2, 5-HT_3, and 5-HT_4 receptors.

Serotonin is known to play an important role in regulating mood, sleep, and appetite. In fact, low levels of serotonin have been linked to depression. Pharmaceuticals used to treat depression usually target serotonin uptake, receptors, or degradation. They work to either increase serotonin levels or stimulate the serotonin receptor. Norepinephrine is also thought to work in concert with serotonin in regulating mood, so some of these agents also work to block reuptake or degradation of NE.

Acetylcholine Unlike the catecholamines and serotonin, the excitatory neurotransmitter acetylcholine (Ach) is not synthesized from an amino acid. Instead, the enzyme choline acetyltransferase acteylates choline using acetyl-CoA to form Ach (Tab. 2). After Ach is released into the cleft, it can be degraded by acetylcholinesterase present in the cleft to produce choline and acetic acid. The choline is then taken up by the neuron, where it is recycled for production of more Ach.

Ach plays a major role in movement, learning, and memory. It is the sole neurotransmitter found at the neuromuscular junction, the site of communication between motor neurons and muscles. Furthermore, excess Ach can lead to muscle paralysis. Deficits in Ach are known to contribute to Alzheimer's disease. Cholinergic forebrain neurons degenerate in this disease, resulting in learning and memory deficits.

Glutamate Glutamate is the major excitatory neurotransmitter in the mammalian CNS (Tab. 2). Glutamate is thought to be synthesized via metabolic pathways, such

as glycolysis. It is then sequestered in synaptic vesicles through a transporter. Furthermore, after being released into the synaptic cleft, it is transported back into the neuron via high affinity transporters. In addition, glutamate can be taken up by glia, where it is recycled.

Recently, there has been a great deal of focus on the role of glutamate in learning and memory. The dissociation of specific glutamate receptors from signaling pathways has resulted in abnormal learning in transgenic mice. Furthermore, mice that overexpress certain glutamate receptors learn faster and are considered "smarter."

Excess glutamate is released from presynaptic terminals under pathophysiological conditions such as ischemia, anoxia, or stroke. Overstimulation of glutamate receptors results in neuronal death. Thus, a fine balance between understimulation and overstimulation of glutamate receptors must be maintained for normal functioning of the brain.

Gamma-aminobutyric acid Gamma-aminobutyric acid (GABA) is the main inhibitory neurotransmitter in the mammalian brain, and glycine is the main inhibitory neurotransmitter in the spinal cord. GABA is produced when α-ketoglutarate formed in Krebs cycle is transaminated to glutamate by the enzyme GABA-transaminase. Glutamate is then converted into GABA by a reaction catalyzed by glutamic acid decarboxylase (GAD). Thus, immunostaining for GAD distinguishes inhibitory neurons from excitatory neurons. GABA that is released into the synaptic cleft is taken up by both neurons and glia via high-affinity GABA transporters.

GABA is thought to play a role in the regulation of anxiety. Decreased activity of GABA is thought to lead to anxiety. Thus, antianxiety drugs, such as Valium, increase activity of GABA receptors. In addition, abnormalities in neurons that release GABA are thought to result in epilepsy.

2.2 Mechanisms of Neurotransmitter Release

Neurotransmitter release from synaptic vesicles is tightly regulated by calcium. It involves recycling of vesicles via the mechanisms of exocytosis and endocytosis. Specific stages of vesicle recycling include (1) reloading, (2) attachment or docking, (3) priming, (4) triggering/exocytosis, and (5) recycling/endocytosis. This section will serve as a brief overview of the process of the synaptic vesicle–recycling pathway (Fig. 3, Tab. 3).

2.2.1 Uptake of Neurotransmitters into Synaptic Vesicles

The cycle begins when synaptic vesicles take up neurotransmitter. This is accomplished using specialized neurotransmitter transporters that use ATP as an energy source. The interior of the vesicle is acidic, and an electrochemical gradient results, yielding energy for neurotransmitter

Tab. 2 Classification of neurotransmitters.

Classical neurotransmitters	*Nonclassical neurotransmitters*
Catecholamines:	Peptides
Dopamine	Nitric oxide
Norepinephrine	Carbon monoxide
Epinephrine	
Serotonin	
Acetylcholine	
Glutamate	
GABA	
Glycine	

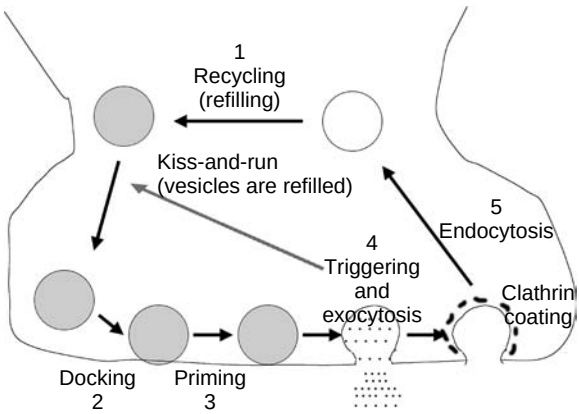

Fig. 3 Synaptic vesicle cycle. The cycle begins when synaptic vesicles take up or are filled with neurotransmitters. The "kiss-and-run" model does, the vesicles refill and enter the cycle.

uptake. Synaptic vesicles that are filled with neurotransmitters move to the active zone, a cytoskeletal matrix below the presynaptic plasma membrane specialized to transduce action potentials into neurotransmitter release.

2.2.2 Docking and Priming

Once at the active zone, vesicles attach to the presynaptic plasma membrane. This process is called *docking*. The vesicles contain a protein called VAMP (also called synaptobrevin, v-SNARE) that has a hydrophobic carboxyl terminus that is anchored to the membrane. VAMP forms a complex with two proteins on the plasma membrane, SNAP-25 and syntaxin (also called t-SNARES). Formation of this heterotrimeric complex, referred to as the *core complex*, causes a free energy change, which is thought to drive vesicle attachment to the membrane.

Once attached, the vesicles undergo an ATP-dependent process that primes them for calcium-dependent release. Recent studies suggest that the nsec1/Munc13 proteins mediate priming. When Munc13 is not present, no fusion-competent vesicles form, and therefore neurotransmission is blocked. Munc13 has two C2 regions, which bind calcium. It is thought that increases in diacylglycerol and calcium levels in the neuron activate Munc13 by binding to it. In addition, active zone proteins, such as RIM1, bind to Munc13 and regulate the dynamic use of neurotransmitters.

2.2.3 Triggering and Exocytosis

Calcium serves as a signal to trigger neurotransmitter release. There are three types of calcium channels in close proximity to vesicle release machinery – P/Q, N, and R – that play a role in increasing intracellular calcium to 10 to 20 μM at the active zone. Triggering occurs within 100 to 200 μs after this level of calcium is achieved. Furthermore, there must be a calcium sensor at the site of exocytosis. Recently, synaptotagmins have been identified as candidate calcium sensors. Synaptotagmin has two C2 domains that bind calcium, and electrophysiological studies have shown that

synaptotagmin is essential for only fast, calcium-triggered release.

Rab3A is a negative regulator of neurotransmitter release. Mice lacking rab3A release more neurotransmitters per calcium signal than wildtype mice. Rab3A resides on synaptic vesicles and binds to rabphilin and RIM1 in a GTP-dependent manner. Rabphilin is a soluble protein that is recruited to synaptic vesicles by rab3A. The mechanism by which these proteins work together to regulate release is still not understood.

2.2.4 Endocytosis and Vesicle Recycling

Once the vesicle is emptied of the neurotransmitter, it is endocytosed so that it may be refilled and reused. There are two main theories on how synaptic vesicles are recycled: (1) clathrin-mediated endocytosis and (2) "kiss and run" endocytosis. The role of a clathrin-mediated pathway is generally accepted; however, it has been proposed that a more direct pathway may predominate under certain conditions. It is thought that retrieval of the vesicle occurs by rapid closure of the fusion pore before the vesicle can collapse into the membrane. Hence, it has been termed "kiss and run." Current experiments to test this model include the use of dyes, such as FM 1–43 that label vesicles, and it has been reported that dye-labeled vesicles can recycle repeatedly before losing their dye.

What is known about endocytosis is that the SNARE complex (VAMP/synaptobrevin, SNAP-25 and syntaxin) disassembles and that endocytosis is regulated by calcium. In addition, there are reports that strongly support the idea that clathrin plays an important role in synaptic vesicle endocytosis. The presence of pits coated with clathrin has been observed in nerve terminals. The number of these pits is increased when neurons are fixed at the peak of endocytic activity. Furthermore, AP-2 and AP180, clathrin adaptor complexes, are concentrated in axon terminals, and when expression of these proteins is disrupted, there is an impairment in synaptic vesicle recycling.

After endocytosis, vesicles undergo several steps before they are available for reloading. The empty vesicles acidify by proton pump activity. They then move

Tab. 3 Proteins involved in synaptic vesicle recycling.

Step	*Proteins*
Uptake of neurotransmitters	Energy-dependent transporters
Docking	VAMP or synaptobrevin (t-SNARE)
	SNAP-25 (v-SNARE)
	Syntaxin (v-SNARE)
Priming	Munc13
	RIM1
Triggering and exocytosis	Calcium channels
	Synaptotagmin
	Rab3A
	Rabphilin
Endocytosis	Clathrin
	AP-2 and AP180

back to the interior of the nerve terminal. The vesicles then fuse with early endosomes, although this step does not seem to be essential. New vesicles bud off of the early endosomes, and then, as described in Sect. 2.2.1, neurotransmitters are taken up into vesicles by active transport.

3 The Postsynaptic Component of a Synapse

3.1 The Postsynaptic Density

For the brain to function, the postsynaptic component of a synapse must be poised to respond to neurotransmitter released from the presynaptic component. In order for the postsynaptic neuron to be prepared to receive signals, the proper proteins must be localized to specific sites on the postsynaptic membrane. These sites are referred to as *postsynaptic densities* (PSDs), which are the regions of the postsynaptic membrane directly opposing the presynaptic terminal. This density may be prominent, as in the case of asymmetric (excitatory) synapses, or of the same thickness as the presynaptic membrane, as in the case of symmetric (inhibitory) synapses. Until relatively recently, the proteins and structures that compose the PSD were unknown. Through the development of biochemical and molecular techniques, much has been learned about the components of PSD in the past 15 to 20 years. Here we focus on one class of proteins found in the PSD, the neurotransmitter receptors.

3.2 Dendritic Spines

In a number of neurons, usually excitatory neurons, the dendritic tree contains dendritic spines that protrude from the main dendritic shaft. These spines are not found on inhibitory neurons. Spines increase the surface area by which neurons can receive signals. Spines come in a number of shapes and sizes, and they contain organelles, such as smooth endoplasmic reticulum, which serves as a calcium store, and polyribosomes, which allow local protein translation. Ten percent of the surface area of spines is occupied by the PSD, and this is where signaling proteins, such as receptors and effectors, are localized. Furthermore, cytoskeletal elements, such as filamentous actin, are enriched in spines.

Recent studies suggest that dendritic spines are highly plastic structures. Changes in spine numbers and shape occur in response to application of neurotrophins, hormones, and glutamate receptor agonists, such as *N*-methyl-D-aspartate (NMDA). The regulation of spine structure is thought to occur by both activity-dependent and independent mechanisms.

3.3 Excitatory Synapses: Glutamate Receptors

As discussed in Sect. 2.1.2, glutamate is the major excitatory neurotransmitter, and glutamate receptors will therefore serve as an example of excitatory neurotransmitter receptors. Postsynaptic membranes of glutamatergic synapses must contain receptors that respond to glutamate. Like many families of neurotransmitter receptors, glutamate receptors can be classified into two main groups, ionotropic and metabotropic. Ionotropic glutamate receptors are ligand-gated ion channels and include NMDA, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionate (AMPA), kainate, and δ receptors. Metabotropic receptors are G-protein-coupled receptors of

which a subset activates phospholipase C and phosphoinositide turnover.

3.3.1 NMDA Receptors

NMDA receptors play a role in learning and memory as well as during pathophysiological states such as ischemia or stroke. Of all of the glutamate receptors, NMDA receptors are the most tightly associated with the PSD of excitatory synapses. NMDA receptors are linked to the cytoskeleton through interactions with a scaffold of linker proteins. The cytoplasmic termini of NMDA receptor subunits are responsible for these interactions.

NMDA receptors are heteromeric and are comprised of NR1 and NR2 subunits. There are four different NR2 subunits, all of which contain C-terminal sequences of ESDV or ESEV. These sequences are consistent with the consensus sequence (S/T)X(V/I/L), which bind to protein interaction domains called PDZ domains (named for the three proteins PSD-95, Discs-large, and zona-occludens 1). In fact, these C-terminal consensus sequences found in NR2 subunits mediate their interaction with PSD-95 (also termed SAP-90), a 95-kDa protein highly enriched in the PSD. PSD-95 is a member of the membrane-associated guanylate kinase (MAGUK) protein family. PSD-95 also binds neuronal nitric oxide synthase (nNOS), linking the NMDA receptor to nNOS, thus allowing calcium influx through the NMDA receptor to activate nNOS. Furthermore, PSD-95 also binds to proteins, such as MAP1A and CRIPT, which bind microtubules, and guanylate kinase associated protein (GKAP), which binds to Shank, linking the entire complex to the actin cytoskeleton. Hence, the binding of the NMDA receptor to PSD-95 plays a role in its signaling and anchoring to the cytoskeleton.

It appears that the binding of NMDA receptors to PSD-95 is not important in mediating its localization to the PSD. In fact, elimination of PSD-95 in mice by knockout technology did not result in a decrease in synaptic localization of NMDA receptors at the PSD. However, it could be argued that other MAGUK proteins could be involved in NMDA receptor localization, and recent reports suggest that this indeed could be the case.

The NR1 subunit also plays an important role in the localization of the NMDA receptors to the PSD. NR1 subunits have alternative splice variants of their carboxyl termini; however, a membrane proximal segment, termed C0, is common to all NR1 subunits. This C0 segment interacts with α-actinin, a protein that interacts with F-actin. Some splice variants of the NMDA receptor also contain a C1 exon segment in the carboxyl region that binds to neurofilament. Finally the cytosolic regions of both NR1 and NR2 subunits bind spectrin, an actin-binding protein. Thus, the NMDA receptor is tightly associated with the PSD and cytoskeleton via interaction with proteins that associate with microtubules and actin.

3.3.2 AMPA Receptors

Although not as tightly associated with the PSD, AMPA receptors are also concentrated at the PSD of excitatory synapses. Many synapses contain NMDA receptors but not AMPA receptors, especially early in development. These synapses are termed *silent synapses* because depolarization of the membrane by AMPA receptor function is often necessary for NMDA receptors to become functional themselves. Electrical activity can regulate the delivery of AMPA

receptors to the PSD, and it is thought that this may be an underlying mechanism for learning and memory.

AMPA receptors are heteromers of GluR1-4 subunits. Like NMDA receptors, the carboxyl termini of AMPA receptors interact with intracellular proteins, some of which are PDZ-containing proteins. However, most of these PDZ-containing proteins are different than those that bind NMDA receptors. Two multi-PDZ proteins that interact with GluR2 and GluR3 have been identified, ABP (for AMPA receptor binding protein) and GRIP (for glutamate receptor interacting protein). However, the function of the binding of these proteins to GluR2 and 3 is still not completely clear. Similarly, GluR2 and 3 have been shown to interact with the PDZ-containing protein PICK1 (protein that interacts with C-kinase 1), and it is thought that this interaction plays a role in linking AMPA receptors to protein kinase C.

Recent evidence suggests that PSD-95 plays a role in localizing AMPA receptors to synapses. PSD-95 also binds a protein called *stargazin*, and it has been shown that the concurrent binding of AMPA receptors and stargazin to PSD-95 is important for the delivery of AMPA receptors to the PSD. Thus, MAGUK proteins may play a role in the delivery of multiple receptors, such as NMDA and AMPA receptors, to the synapse.

Another family of PDZ-containing proteins, the LIN-10/Mint1-3 family, has been shown to play a role in regulating surface expression of AMPA receptors. The mechanism by which this family of proteins regulates receptor targeting is currently unknown.

3.3.3 Kainate and δ Receptors

Kainate receptors can be pre- or postsynaptic, and consist of heteromers of GluR5-7, KA1, and KA2 subunits. Like NMDA and AMPA receptor subunits, the GluR6 and KA2 subunits can bind to PSD-95. GluR6 binds to the first PDZ domain of PSD-95 and KA2 binds to the SH3 and GK domains. PSD-95 may serve to localize kainate receptors to the PSD; however, this has not yet been definitively shown. δ receptors are comprised of GluRδ subunits, of which GluRδ2 has been shown to bind to the PSD-95/MAGUK family member PSD-3/chapsyn-110. This interaction is thought to anchor GluRδ receptors to the PSD.

3.3.4 Metabotropic Glutamate Receptors

Metabotropic glutamate receptors are classified into three main groups: group 1 (includes mGluR1 and 5) activate phospholipase C, group 2 (includes mGluR2 and 3) and group 3 (includes mGluR4, 6, 7, and 8), both of which inhibit adenylate cyclase. Similar to the ionotropic glutamate receptors, mGluR targeting is mediated by the binding of intracellular carboxyl termini to intracellular partners.

mGluRs of group 1 are concentrated in a ring around the PSD. Recent studies have focused on the role of their binding partners in localization and signaling. It appears that these receptors bind to Homer via a PPXXF motif present in the carboxyl terminus. A subset of Homer proteins multimerize and are thought to crosslink group 1 mGluRs to other mGluRs or to other proteins containing Homer binding motifs, such as the IP_3 receptor. It is thought that the binding of Homer to mGluR1 and 5 is similar to the binding of PSD-95 to NMDA receptors in that it scaffolds the receptor with signaling proteins.

The carboxyl terminal tail of mGluR5 also contains a PDZ-binding motif. As

such, mGluR5 has been shown to bind to the PDZ domain of Shank. Since Shank binds to actin, this interaction may serve to anchor the mGluR5 to the PSD cytoskeleton. Since Shank is part of the NMDA receptor complex, it may serve to aid in crosstalk between NMDA and metabotropic glutamate receptors.

3.4 Inhibitory Synapses: GABA Receptors

The principal inhibitory receptor in the brain is the ionotropic GABA receptor, which is a class of ligand-gated ion channels made of subunits from at least six different types. GABA receptors are primarily found at the PSD of inhibitory synapses with a small subset found in extrasynaptic locations. Like the glutamate receptors, GABA receptors rely on interaction with intracellular proteins to help to anchor it at the PSD. However, unlike NMDA and AMPA receptors, GABA receptors interact with partners via their intracellular loops between transmembrane regions 3 and 4.

3.4.1 $GABA_A$ Receptors

Ionotropic $GABA_A$ receptors are found throughout the mammalian CNS. Activation of this receptor results in increased chloride conductance of neuronal membranes, which can be blocked by bicuculline. $GABA_A$ receptors are pentamers, and there are at least seven classes of $GABA_A$ receptor subunits – α, β, γ, δ, ε, π, and θ. The most common $GABA_A$ receptor in the mammalian CNS consists of two copies each of $\alpha 1$ and $\beta 2$ subunits associated with a $\gamma 2$ subunit. Benzodiazapines bind with high affinity to a subset of $GABA_A$ receptors, where they exert their sedative and anticonvulsive effects. Furthermore, barbiturates interact with $GABA_A$ receptors.

A $GABA_A$ receptor-associated protein (GABARAP) has been identified, and this protein binds to the intracellular loop. GABARAP has 30% identity with a light chain of some microtubule-associated proteins, suggesting that GABARAP may associate with microtubules. Indeed, GABARAP fractionates with microtubules. Thus, $GABA_A$ receptors are linked to the cytoskeleton of the PSD at inhibitory synapses.

3.4.2 $GABA_B$ Receptors

$GABA_B$ receptors are metabotropic receptors that inhibit adenylate cyclase activity and intracellular calcium levels. They are widely expressed throughout the mammalian CNS. They are heterodimers of $GABA_{B1}$ and $GABA_{B2}$ subunits, which interact through their carboxyl termini. $GABA_B$ receptors are the target of baclofen, which is used to treat spasticity. Thus, $GABA_B$ receptors may be a treatment for disorders such as epilepsy.

3.4.3 $GABA_C$ Receptors

Like the $GABA_A$ receptors, $GABA_C$ receptors are ligand-gated chloride channels. However, unlike $GABA_A$ receptors, $GABA_C$ receptors are not blocked by bicuculline and do not bind benzodiazepines. $GABA_C$ receptors contain ρ subunits of which two types (1 and 2) can be assembled into homomeric or heteromeric pentamers. These receptors play an important inhibitory role in the retina. MAP1B (microtubule associated protein 1B) binds to $GABA_C$ receptors, anchoring it to the neuronal cytoskeleton.

4 Changes in Neuronal Morphology in Response to Signals

4.1 Signals that Regulate Dendrite Patterning

4.1.1 Electrical Activity

Electrical activity plays a role in shaping synaptic connections in the nervous system during development as well as during experience and learning. It is thought that signals from incoming presynaptic terminals activate the appropriate postsynaptic partners, shaping the dendritic tree so that it will have the correct mature morphology. Recent evidence suggests that activity also "fine-tunes" already branched dendrites.

Blocking activity can result in decreased dendritic growth and branching. This has been demonstrated in the visual system where dark-rearing results in attenuation of the length and number of dendritic branches in stellate and lateral geniculate neurons. In the cerebellum, when granule cells fail to innervate Purkinje cells, their dendritic branches are less elaborate. Similarly, when cultured sympathetic or hippocampal neurons are depolarized using KCl, dendritic branching increases.

What do these results mean physiologically? When rats are reared in enriched environments, the dendritic arbors of cortical pyramidal cells are much more complex than those of rats reared in control environments. Furthermore, it has been shown that in some systems, activity does not affect dendritic arborization. Recently, a study of the dynamics of dendrite development in the optic tectum of frogs provided insight into apparent conflicts regarding whether activity is an important factor in determining dendrite morphology. It appears that there is a specific developmental time window ("critical period") when dendrite patterning can be regulated by electrical activity. Hence, for neurons that are plastic for a critical period during development, activity may influence dendrite number only during this critical period. For neurons that remain plastic into adulthood, for example in the hippocampus and cortex, activity (perhaps as a result of environment) may continue to shape the dendritic tree.

4.1.2 Neurotrophins

Originally, the neurotrophins (nerve growth factor or NGF, brain-derived growth factor or BDNF, Neurotrophin-3 or NT-3, and neurotrophin-4 or NT-4) were studied as survival factors for neurons. However, in the late 1980s, it was found that neurotrophins could regulate dendritic patterning in the peripheral nervous system. Similarly, soon after, it was discovered that neurotrophins could also regulate dendrite patterning in the developing neocortex. In recent years, BDNF has been shown to promote dendritic arborization in the developing retina and cerebellum and as a mediator of activity-dependent synaptic plasticity.

4.1.3 Extracellular Signaling Proteins

Although little is known about signaling molecules that play a role in dendrite patterning, studies performed during the past five years have led to the identification of proteins that regulate dendrite branching (Fig. 4). Semaphorin 3A, originally identified as a chemorepellent for cortical axons, acts to regulate dendrite patterning in these same neurons by acting as a chemoattractant for apical dendrites. Similarly, the ephrin family of ligands and receptors was originally identified as an axon guidance molecule in the visual system. Overexpression of one of the ephrin

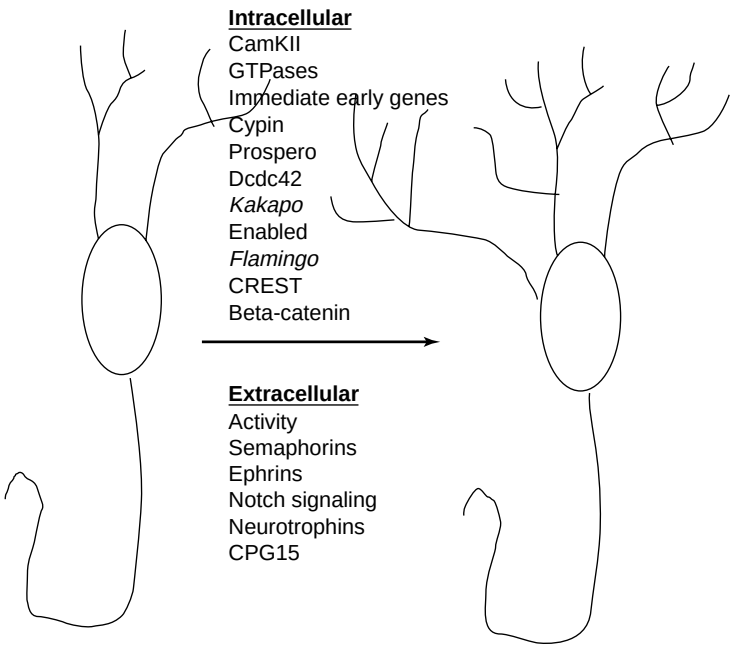

Fig. 4 Signaling proteins that regulate dendrite branching and morphology.

receptors, EphA3, results in decreased dendrite branching in the visual system. Recent studies have shown that other proteins, such as Notch, Wnt, beta-catenin, and the cell adhesion molecule L1, also regulate dendrite growth. Like EphA3, Notch is a receptor, and increasing amounts of Notch result in decreased dendritic growth. In contrast, but in parallel with Semaphorin 3A, L1 signaling positively affects dendrite number. Thus, it appears that regulators of axonal pathfinding also play a role in dendritic patterning (Fig. 4).

In addition to axon guidance molecules, hormones regulate dendritic arborization. Rats that have underactive thyroid glands have attenuated dendrite arborization while those with overactive thyroids have more complex arborization in hippocampal pyramidal neurons. Similarly, thyroid hormone increases dendritic arbors in neonatal rats, while estradiol increases dendrite number, and exposure to glucocorticoids decreases dendrite number. Thus, hormones may play an ongoing role in determining dendrite morphology.

4.1.4 Intracellular Signaling Molecules

The study of intracellular molecules that play a role in dendrite patterning is a relatively new field. However, recent studies indicate that intracellular regulators of dendrite arborization include calcium/calmodulin-dependent protein kinase II (CaMKII), the small GTPases, immediate early genes, cypin, and a number of novel genes (Fig. 4). The role of CaMKII in dendrite patterning has been studied in *Xenopus* tectal neurons. CaMKII negatively regulates dendritic growth and is thought to be a signal that halts growth. The role of small GTPases, including Rho, Rac, and Cdc42, appears to be more complex. These proteins regulate

the actin cytoskeleton, and together, they mediate the effects of extracellular signals. Similarly, activity increases the expression of immediate early genes, two of which are *Arc* (for activity regulated cytoskeletal-associated protein) and *Narp* (for neuronal activity-regulated pentraxin). Although there is no direct evidence that *Arc* mediates dendritic remodeling, its localization in dendrites makes it a prime candidate as an intracellular mediator. Stronger evidence exists for a role for *Narp* in dendrite patterning. *Narp* is induced by electrical activity, is secreted from neurons, and stimulates dendritic growth in cultured neurons. Cypin (for cytosolic PSD-95 interactor) is another protein that is upregulated by electrical activity. Cypin is a guanine deaminase that binds to tubulin heterodimers and promotes microtubule assembly. Cypin's enzymatic activity may play a role in dendrite patterning since its overexpression results in increased dendrite number. Moreover, knocking down cypin protein levels results in a decrease in dendrite number.

Genetic screens are underway to identify genes that control dendrite morphology. *Drosophila* screens have identified at least a dozen candidates that regulate sensory neuron dendrite patterning. These include the transcription factor Prospero, genes that regulate the cytoskeleton such as Dcdc42, *kakapo*, and enabled, and the membrane-spanning protein *flamingo*. Using a strategy called *transactivator trapping*, a calcium-responsive transactivator called CREST was very recently identified. CREST plays a role in dendritic patterning in the mammalian hippocampus and cortex. Mice with disrupted expression of CREST have deficits in calcium-dependent dendritic growth. Thus, the field of dendrite patterning is quickly evolving with the identification of new players that regulate dendrite number and morphology.

4.2 Signals that Regulate Axons

4.2.1 Electrical Activity

The role of electrical activity on the patterning of the visual system has been well studied. Dark-rearing or visual deprivation during visual system development results in a disruption of ocular dominance columns in layer 4 of the cortex. Similarly, if action potentials are blocked during the "critical period" when thalamic axons grow into the primary visual cortex, normal branching patterns of these axons are disrupted. Furthermore, the neurons project into areas where they normally do not innervate. Similarly, in barn owls, the auditory map is guided by vision, and hence electrical activity. If the owls are raised wearing prismatic spectacles that displace vision, their auditory neurons become tuned to sound locations that correspond to visually displaced sites. Thus, neuronal activity plays a major role in guiding axons in the visual and auditory systems (Fig. 5).

The role of electrical activity on axon guidance has been recently studied using cultured Xenopus spinal neurons. Brief periods of electrical stimulation increase netrin-1 induced attraction and converts myelin-associated glycoprotein repulsion to attraction. In contrast, electrical activity does not play a major role in axon guidance in mammalian motor neurons. Thus, electrical activity may serve as a modulator for axon guidance in particular sets of neurons.

4.2.2 Extracellular Cues: Netrins

It has been long established that neuronal migration and pathfinding is regulated

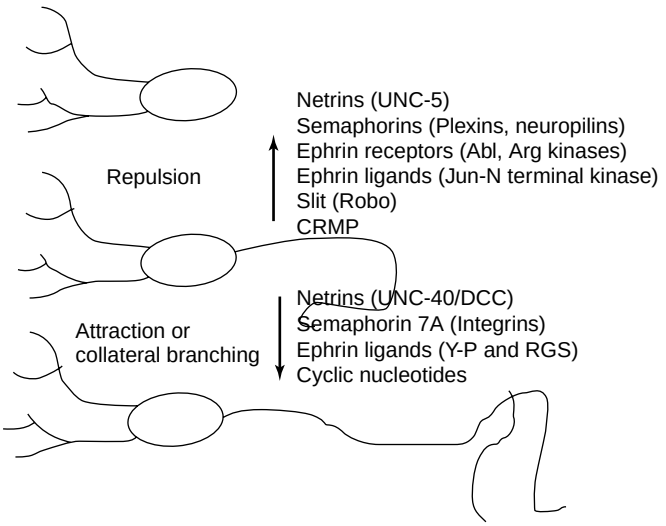

Fig. 5 Signaling proteins involved in axon guidance. This is a very simplified view of proteins that regulate axon attraction or repulsion. There is still much that is not known. Y-P is tyrosine phosphorylation and RGS is regulator of G-protein signaling. All other names are discussed in more detail in the text.

by extracellular cues. These cues include the netrins, semaphorins, ephrins, and Slits (Fig. 5). Netrins were first identified as playing a role in axon guidance in both the vertebrate and invertebrate nervous systems. They are diffusible substances, and the same netrin that can be attractive to some axons may be repulsive to others. Netrins bind to two classes of receptors UNC-40 and UNC-5. Netrin attraction appears to involve UNC-40 homologs (i.e. deleted in colorectal cancer, DCC), which signal through phospholipase C, mitogen-activated kinase, and the small GTPases Cdc42 and Rac. Repulsion appears to be mediated by UNC-5 family members either in combination with UNC-40 homologs (DCC) or alone. The cytosolic regions of UNC-5 homologs contain protein interaction domains, which may regulate intracellular signaling.

In addition to providing cues for axon pathfinding, netrins can also promote axon outgrowth. The nuclear factor of activated T cell (NFAT) transcription factors act downstream of DCC. Inhibition of NFAT activation or elimination of NFAT in mice results in defects in axonal outgrowth.

4.2.3 Extracellular Cues: Semaphorins

Some members of the semaphorin family are membrane-bound, and some are secreted. In *Drosophila*, the invertebrate semaphorin SemI was identified in genetic studies of axon fasciculation. In vertebrates, semaphorins were discovered in assays of growth cone collapse. Sema3A is a repulsive cue for NGF-responsive neurons of the dorsal root ganglia and is thought to prevent axons of these neurons from invading the ventral spinal cord during development. Plexins and neuropilins have been identified as receptors for the semaphorins. Neuropilins can associate with plexins and form functional receptors

for Semaphorin 3A. Met, a receptor tyrosine kinase, forms a complex with one of the plexins (B-1), signaling through small GTPases to alter actin dynamics, which then results in axonal repulsion. In contrast, a different plexin subtype (A-1) signals through a cytosolic protein called *collapsin response mediator protein*(CRMP). Recently, it has been shown that overexpression of CRMP results in multiple axon formation. CRMP binds to tubulin heterodimers and promotes microtubule assembly, a mechanism, which may underlie CRMP's role in axon guidance and outgrowth.

In contrast to Sema3A, the plexins, which primarily cause axon retraction or repulsion, semaphorin 7A (Sema7A) has been shown to promote axon outgrowth. In fact, Sema7A does not signal through neuropilins or plexins; it binds to integrins instead. Its signaling is mediated by mitogen-activated and focal adhesion kinases.

4.2.4 Extracellular Cues: Ephrins

It is thought that complementary expression and binding of ephrin receptors in retina to their ephrin ligands in tectal targets mediate retinal–tectal patterning. Unlike other axon guidance molecules, ephrins are not secreted. Ephrins of type A are GPI-linked and ephrins of type B are transmembrane proteins. Ephrin receptors for both types of ligands are transmembrane receptors with intracellular tyrosine kinase domains. Like other tyrosine kinase receptors, binding of Eph ligands to their receptors results in receptor phosphorylation and dimerization. Activation of these receptors results in inactivation of Ras and suppression of extracellular signal-regulated kinase (ERK). Furthermore, signaling of ephrin receptors through Abelson (Abl) and related (Arg) kinases results in axonal repulsion. In addition, the ephrin ligands themselves can be involved in reverse signaling. There are three possible pathways: (1) one involving tyrosine phosphorylation, (2) one involving Jun-N terminal kinase, and (3) one involving the regulator of G-protein signaling, inhibiting G-protein coupled receptor signaling. The first and third pathways are involved in axonal outgrowth, and the second pathway is involved in axonal repulsion (Fig. 5).

4.2.5 Extracellular Cues: Slits

Slit was originally identified in a genetic screen in *Drosophila* designed to identify proteins involved in embryonic patterning. More recently, it has been shown that both *Drosophila* and vertebrate slits are axon repellents. Roundabout (Robo) is the receptor that is responsible for the actions of slit (Fig. 5). Although potential players have been identified, genetic studies have shown that slit signaling is complex, and no concrete evidence exists to assign a definitive signaling pathway for slit in axonal repulsion. However, like the ephrin receptor signaling, Abl may play a role in Robo-mediated slit signaling. Enabled, an Abl substrate, and Capulet, an adenylate cyclase-associated protein, act downstream of Abl. There is evidence that small GTPases play a role in mediating slit-promoted axon repulsion. Furthermore, it is thought that slit can use DCC as a receptor in its signaling.

4.2.6 Extracellular Cues: CPG15

CPG15 is an extracellular molecule that is anchored to the plasma membrane by a glycosyl–phosphatidylinositol linkage. CPG15 mRNA is expressed during development, and the protein is expressed in areas of the brain that are developmentally

plastic, such as in the visual system. CPG15 mRNA is upregulated by electrical activity and BDNF and is responsible for regulating the growth of apposing dendritic and axonal arbors by decreasing rates of branch retractions of presynaptic retinal axons. It is still not known what protein functions as the CPG15 receptor.

4.3 Intracellular Signaling Molecules

CRMP, discussed above as a mediator of semaphoring signaling, causes the formation of multiple axons when overexpressed (Fig. 5). In addition, CRMP is present at high levels in axonal growth cones, and thus, may play a major role in determining neuronal morphology. Other intracellular signaling molecules that can regulate axon morphology include cyclic nucleotides, which convert repulsive growth cone responses to attractive ones, and calcium, which can regulate growth cone turning, and hence, axonal guidance (Fig. 5). It is clear that we are just beginning to make headway into understanding the mechanisms by which axonal morphology is determined.

5 Perspectives

Our understanding of how the brain functions is expanding exponentially. Recent advances have resulted in our understanding of how neurotransmitter vesicles are recycled, how signaling complexes are targeted to synaptic sites, and how neuronal morphology is determined. New molecular and biochemical methods will enable us to continue to identify and clarify details and mechanisms underlying neuronal function. This chapter is intended to build a basic foundation in neuron chemistry and to provide the background necessary to understand ongoing investigations of mechanisms that are critical for neuronal signaling.

Acknowledgments

I would like to thank Chris Rongo and members of the Firestein laboratory for valuable suggestions for this chapter. Bonnie L. Firestein is supported by grants from the National Science Foundation and New Jersey Commission on Spinal Cord Research.

See also Pain Transduction: Gating and Modulation of Ion Channels.

Bibliography

Books and Reviews

Cowan, W.M., Südhof, T.C., Stevens, C.F. (2001) *Synapses*, The Johns Hopkins University Press, Maryland.

Guan, K.L., Rao, Y. (2003) *Nat. Neurosci. Rev.* **4**, 941–956.

Flanagan, J.G., Vanderhaegen, P. (1998) The ephrins and Eph receptors in neural development, *Annu. Rev. Neurosci.* **21**, 309–345.

McAllister, A.K. (2000) *Cereb. Cortex* **10**, 963–973.

Rosenmund, C., Rettig, J., Brose, N. (2003) *Curr. Opin. Neurobiol.* **13**, 1–11.

Squire, L.R., Bloom, F.E., McConnell, S.K., Roberts, J.L., Spitzer, N., Zigmond, M.J. (2002) *Fundamental Neuroscience*, 2nd edition, Academic Press, New York.

Primary Literature

Aizawa, H., Hu, S.C., Bobb, K., Balakrishnan, K., Ince, G., Gurevich, I., Cowan, M.,

Ghosh, A. (2004) Dendrite development regulated by CREST, a calcium-regulated transcriptional activator, *Science* **303**, 197–202.

Akum, B.F., Chen, M., Gunderson, S.I., Riefler, G.M., Scerri-Hansen, M.M., Firestein, B.L. (2004) Cypin regulates dendrite patterning in hippocampal neurons by promoting microtubule assembly, *Nat. Neurosci.* **7**, 145–152.

Augustin, I., Rosenmund, C., Südhof, T., Brose, N. (1999) Munc13-1 is essential for fusion competence of glutamatergic synaptic vesicles, *Nature* **400**, 457–461.

Baker, R.E., Dijkhuizen, P.A., Van Relt, J., Verhaagen, J. (1998) Growth of pyramidal, but not non-pyramidal, dendrites in long-term organotypic explants of neonatal rat neocortex chronically exposed to neurotrophin-3, *Eur. J. Neurosci.* **10**, 1037–1044.

Bashaw, G.J., Kidd, T., Murray, D., Pawson, T., Goodman, C.S. (2000) Repulsive axon guidance: Abelson and Enabled play opposing roles downstream of the roundabout receptor, *Cell* **101**, 703–715.

Betz, A., Okamoto, M., Benseler, F., Brose, N. (1997) Direct interaction of the rat unc-13 homolog Munc13-1 with the N terminus of syntaxin, *J. Biol. Chem.* **272**, 2520–2526.

Betz, A., Thankur, P., Ashery, U., Rhee, J.S., Scheuss, V., Rosenmund, C., Rettig, J., Brose, N. (2001) Functional interaction of the active zone proteins Munc-13 and RIM1 in synaptic vesicle priming, *Neuron* **30**, 183–196.

Blahos, J. II, Wenthold, R.J. (1996) Relationship between N-methyl-D-aspartate receptor NR1 splice variants and NR2 subunits, *J. Biol. Chem.* **271**, 15669–15674.

Borges, S., Berry, M. (1976) Preferential orientation of stellate cell dendrites in the visual cortex of the dark-reared rat, *Brain Res.* **112**, 141–147.

Brainard, M.S., Knudsen, E.I. (1993) Experience-dependent plasticity in the inferior colliculus: a site for visual calibration of the neural representation of auditory space in the barn owl, *J. Neurosci.* **13**, 4589–4608.

Brenman, J.E., Chao, D.S., Gee, S.H., McGee, A.W., Craven, S.E., Santillano, D.R., Wu, Z., Huang, F., Xia, H., Peters, M.F., Froehner, S.C., Bredt, D.S. (1996) Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and alpha1-syntrophin mediated by PDZ domains, *Cell* **84**, 757–767.

Brenman, J.E., Topinka, J.R., Cooper, E.C., McGee, A.W., Rosen, J., Milroy, T., Ralston, H.J., Bredt, D.S. (1998) Localization of postsynaptic density-93 to dendritic microtubules and interaction with microtubule-associated protein 1A, *J. Neurosci.* **18**, 8805–9913.

Brose, K., Bland, K.S., Wang, K.H., Arnott, D., Henzel, W., Goodman, C.S., Tessier-Lavigne, M., Kidd, T. (1999) Slit proteins bind robo receptors and have an evolutionarily conserved role in repulsive axon guidance, *Cell* **96**, 795–806.

Bruckner, K., Pasquale, E.B., Klein, R. (1997) Tyrosine phosphorylation of transmembrane ligands for Eph receptors, *Science* **275**, 1640–1643.

Buck, K.B., Zheng, J.Q. (2002) Growth cone turning induced by direct local modification of microtubule dynamics, *J. Neurosci.* **22**, 9358–9367.

Cantallops, I., Haas, K., Cline, H.T. (2000) Postsynaptic CPG15 promotes synaptic maturation and presynaptic axon arbor elaboration in vivo, *Nat. Neurosci.* **3**, 1004–1011.

Catalano, S.M., Shatz, C.J. (1998) Activity-dependent cortical target selection by thalamic axons, *Science* **281**, 559–562.

Chan, S.S., Zheng, H., Su, M.W., Wilk, R., Killeen, M.T., Hedgecock, E.M., Culotti, J.G. (1996) UNC-40, a *C. elegans* homolog of DCC (Deleted in colorectal cancer) is required in motile cells responding to UNC-6 netrin clues, *Cell* **87**, 187–195.

Chen, L., Chetkovich, D.M., Petralia, R.S., Sweeney, N.T., Kawasaki, Y., Wenthold, R.J., Bredt, D.S., Nicoll, R.A. (2000) Stargazin regulates synaptic targeting of AMPA receptors by two distinct mechanisms, *Nature* **408**, 936–943.

Cheng, H.J., Flanagan, J.G. (1994) Identification and cloning of ELF-1, a developmentally expressed ligand for the Mek4 and Sek receptor tyrosine kinases, *Cell* **79**, 157–168.

Cho, K.O., Hunt, C.A., Kennedy, M.B. (1992) The rat brain postsynaptic density fraction contains a homolog of the *Drosophila* discs-large tumor suppressor protein, *Neuron* **9** 929–942.

Coleman, P.D., Rieen, A.H. (1968) Environmental effects on cortical dendritic fields. I. Rearing in the dark, *J. Anat.* **102**, 363–374.

Constantine-Paton, M., Caprianica, R.R. (1975) Central projection f the optic tract from translocated eyes in the leopard frog (Rana pipens), *Science* **189**, 480–482.

Corriveau, R.A., Shatz, C.J., Nedivi, E. (1999) Dynamic regulation of cpg15 during activity-dependent synaptic development in the mammalian visual system, *J. Neurosci.* **19**, 7999–8008.

Dailey, M.E., Smith, S.J. (1996) The dynamics of dendritic structure in developing hippocampal slices, *J. Neurosci.* **16**, 2983–2994.

de la Torre, J.R., Hopker, V.H., Ming, G.L., Poo, M.M., Tessier-Lavigne, M., Hemmati-Brivanlou, A., Holt, C.E. (1997) Turning of retinal growth cones in a netrin-1 gradient mediated by the netrin receptor DCC, *Neuron* **19**, 1211–1224.

Drescher, U., Kremoser, C., Handwerkerm, C., Loschinger, J., Noda, M., Bonhoeffer, F. (1995) *In vitro* guidance of retinal ganglion cell axons by RAGS, a 25 kDa tectal protein related to ligands for Eph receptor tyrosine kinases, *Cell* **82**, 359–370.

Elowe, S., Holland, S.J., Kulkarni, S., Pawson, T. (2001) Downregulation of the Ras-mitogen-activated protein kinase pathway by the EphB2 receptor tyrosine kinase is required for ephrin-induced neurite retraction, *Mol. Cell. Biol.* **21**, 7429–7441.

Fernandez-Chacon, R., Konigstorfer, A., Gerber, S.H., Garcia, J., Matos, M.F., Stevens, C.F., Brose, N., Rizo, J., Rosenmund, C., Südhof, T.C. (2001) Synaptotagmin 1 functions as a calcium regulator of release probability, *Nature* **410**, 41–49.

Forcet, C., Stein, E., Pays, L., Corset, V., Llambi, F., Tessier-Lavigne, M., Mehlen, P. (2002) Netrin-1 mediated axon outgrowth requires deleted in colorectal cancer-dependent MAPK activation, *Nature* **417**, 443–447.

Fukata, Y., Itoh, T.J., Kimura, T., Menager, C., Nishimura, T., Shiromizu, T., Watanabe, H., Inagaki, N., Iwamatsu, A., Hotani, H., Kaibuchi, K. (2002) $CRMP^{-2}$ binds to tubulin heterodimers to promote microtubule assembly, *Nat. Cell Biol.* **4**, 583–591.

Gao, F.B., Brenman, J.E., Jan, L.Y., Jan, N.Y. (1999) Genes regulating dendritic outgrowth, branching, and routing in *Drosophila*, *Genes Dev.* **13**, 2549–2561.

Geppert, M., Bolshakov, V.Y., Siegelbaum, S.A., Takei, K., De Camilli, P., Hammer, R.E., Südhof, T.C. (1994) The role of Rab3A in neurotransmitter release, *Nature* **369**, 493–497.

Giordano, S., Corso, S., Conrotto, P., Artigiani, S., Gilestro, G., Barberis, D., Tamagnone, L., Comoglio, P.M. (2002) The semaphorin 4D receptor controls invasive growth by bcoupling with Met, *Nat. Cell Biol.* **4**, 720–724.

Gitai, Z., Yu, T.W., Lundquist, E.A., Tessier-Lavgne, M., Bargmann, C.I. (2003) The netrin receptor Unc-40/DCC stimulates axon attraction and outgrowth through Enabled and, in parallel, Rac and Unc-115/Ablim, *Neuron* **37**, 53–65.

Glover, J.C. (2000) Development of specific connectivity between premotor neurons and motoneurons in the brain stem and spinal cord, *Physiol. Rev.* **80**, 615–647.

Goshima, Y., Nakamura, F., Strittmatter, P., Strittmatter, S.M. (1995) Collapsin-induced growth cone collapse mediated by an intracellular protein related to UNC-33, *Nature* **376**, 509–514.

Gould, E., Allan, M.D., McEwen, B.S. (1990) Dendritic spine density of adult hippocampal pyramidal cells is sensitive to thyroid hormone, *Brain Res.* **525**, 327–329.

Gould, E., Westlind-Danielsson, A., Frankfurt, M., McEwen, B.S. (1990) Sex differences and thyroid hormone sensitivity of hippocampal pyramidal cells, *J. Neurosci.* **10**, 996–1003.

Gould, E., Woolley, C.S., McEwen, B.S. (1991) The hippocampal formation: morphological changes induced by thyroid, gonadal, and adrenal hormones, *Psychoneuroendocrinology* **16**, 67–84.

Graef, I.A., Wang, F., Charron, F., Chen, L., Neilson, J., Tessier-Lavigne, M., Crabtree, G.R. (2003) Neurotrophins and netrins require Cacineurin/NFAT signaling to stimulate the outgrowth of embryonic axons. *Cell* **113**, 657–670.

Greenough, W.T., Volkmar, F.R. (1973) Pattern of dendritic branching in occipital cortex of rats reared in complex environments, *Exp. Neurol.* **40**, 491–504.

Hanley, J.G., Koulen, P., Bedford, F., Gordon-Weeks, P.R., Moss, S.J. (1999) The protein MAP-1B links GABA(C) receptors to the cytoskeleton at retinal synapses, *Nature* **397**, 66–69.

He, Z., Tessier-Lavigne, M. (1997) Neuropilin is a receptor for the axonal chemorepellent Semaphorin III, *Cell* **90**, 739–751.

Hedgecock, E.M., Culotti, J.G., Hall, D.H., Stern, B.D. Genetics of cell and axon migration in *Caenorhabdtis elegans*, *Development* **100**, 365–382.

Hedgecock, E.M., Culotti, J.G., Hall, D.H., Stern, B.D. The *unc-5, unc-6*, and *unc-40* genes guide circumferential migrations of pioneer axons and mesodermal cells on the epidermis in *C. elegans*, *Neuron* **4**, 61–85.

Hermann, K., Shatz, C.J. (1995) Blockade of action potential activity alters initial arborization of thalamic neurons within cortical layer 4, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11244–11248.

Hirbec, H., Francis, J.C., Lauri, S.E., Braithwaite, S.P., Coussen, F., Mulle, C., Dev, K.K., Coutinho, V., Meyer, G., Isaac, J.T., Collingridge, G.L., Henley, J.M., Couthino, V. (2003) Rapid and differential regulation of AMPA and kainate receptors at hippocampal mossy fibre synapses by PICK1 and GRIP, *Neuron* **37**, 625–638.

Holland, S.J., Gale, N.W., Mbamalu, G., Yancopoulos, G.D., Henkemeyer, M., Pawson, T. (1996) Bidirectional signaling through the EPH-family receptor Nuk and its transmembrane ligands, *Nature* **383**, 722–725.

Holloway, R.L. (1966) Dendritic branching: some preliminary results of training and complexity in rat visual cortex, *Brain Res.* **2**, 393–396.

Hubel, D.H., Wiesel, T.N. (1970) The period of susceptibility to the physiological effects of unilateral eye closure in kittens, *J. Physiol. (London)* **206**, 419–436.

Dong, H., O'Brien, R.J., Fung, E.T., Lanahan, A.A., Worley, P.F., Huganir, R.L. (1997) GRIP: a synaptic PDZ domain-containing protein that interacts with AMPA receptors, *Nature* **386**, 279–284.

Hong, K., Nishiyama, M., Henley, J., Tessier-Lavigne, M., Poo, M. (2000) Calcium signalling in the guidance of nerve growth by netrin-1, *Nature* **403**, 93–98.

Horch, H.W., Kruttgen, A., Portbury, S.D., Katz, L.C. (1999) Destabilization of cortical dendrites and spines by BDNF, *Neuron* **23**, 353–364.

Hu, H., Marton, T.F., Goodman, C.S. (2001) Plexin B mediates axon guidance in *Drosophila* by simultaneously inhibiting active Rac and enhancing RhoA signaling, *Neuron* **32**, 39–51.

Hutson, L.D., Chien, C.B. (2002) Pathfinding and error correction by retinal axons: the role of ashtray/robo2, *Neuron* **33**, 205–217.

Inagaki, N., Chihara, K., Arimura, N., Menager, C., Kawano, Y., Matsuo, N., Nishimura, T., Amano, M., Kaibuchi, K. (2001) $CRMP^{-2}$ induces axons in cultured hippocampal neurons, *Nat. Neurosci.* **4**, 781–782.

Jun, K., Piedras-Renteria, E.S., Smith, S.M., Wheeler, D.B., Lee, S.B., Lee, T.G., Chin, H., Adams, M.E., Sceller, R.H., Tsien, R.W., Shin, H.S. (1999) Ablation of P/Q-type Ca^{2+} channel currents, altered synaptic transmission, and progressive ataxia in mice lacking the alpha (1A) -subunit, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 15245–15250.

Kennedy, T.E., Serafini, T., de la Torre, J.R., Tessier-Lavigne, M. (1994) Netrins are diffusible chemotropic factors for commisural axons in the embryonic spinal cord, *Cell* **78**, 425–435.

Keino-Masu, K., Masu, M., Hinck, L., Leonardo, E.D., Chan, S.S., Culotti, J.G., Tessier-Lavigne, M. (1996) Deleted in Colorectal Cancer (DCC) encodes a netrin receptor, *Cell* **87**, 175–185.

Kidd, T., Blend, K.S., Goodman, C.S. (1999) Slit is the midline repellent for the robo receptor in *Drosophila*, *Cell* **96**, 785–794.

Kim, E., Cho, K.O., Rothschild, A., Sheng, M. (1996) Heteromultimerization and NMDA receptor-clustering activity of Chapsyn-110, a member of the PSD-95 family of proteins, *Neuron* **17**, 103–113.

Kim, E., Naisbitt, S., Hsueh, Y.P., Rao, A., Rothschild, A., Craig, A.M., Sheng, M. (1997) GKAP, a novel synaptic protein that interacts with the guanylate kinase-like domain of the PSD-95/SAP90 family of channel clustering molecules. *J. Cell Biol.* **136**, 669–678.

Kirsch, J., Wolters, I., Triller, A., Betz, H. (1993) Gephyrin antisense oligonucleotides prevent glycine receptor clustering in spinal neurons, *Nature* **366**, 745–748.

Kistner, U., Wenzel, B.M., Veh, R.W., Cases-Langhoff, C., Garner, A.M., Appeltauer, U., Voss, B., Gundelfinger, E.D., Garner, C.C. (1993) SAP90, a rat presynaptic protein related to the product of the Drosophila tumor suppressor gene dlg-A, *J. Biol. Chem.* **268**, 4580–4583.

Knudsen, E.I., Brainard, M.S. (1991) Visual instruction of the neural map of auditory space

in the developing optic tectum, *Science* **253**, 85–87.

Koleske, A.J., Gifford, A.M., Scott, M.L., Nee, M., Bronson, R.T., Miczek, K.A., Baltimore, D. (1998) Essential roles for the Abl and Arg tyrosine kinases in neurulation, *Neuron* **21**, 1259–1272.

Kolodkin, A.L., Mattes, D.J., Goodman, C.S. (1993) The semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules, *Cell* **75**, 1389–1399.

Kolodkin, A.L., Levengood, D.V., Rowe, E.G., Tai, Y.T., Giger, R.J., Ginty, D.D. (1997) Neuropilin is a semaphorin III receptor, *Cell* **90**, 753–762.

Kornau, H.C., Schenker, L.T., Kennedy, M.B., Seeburg, P.H. (1995) Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95, *Science* **269**, 1737–1740.

Kossel, A., Lowel, S., Bolz, J. (1995) Relationships between dendritic fields and functional architecture in striate cortex of normal and visually deprived rats, *J. Neurosci.* **15**, 3913–3926.

Leenders, A.G., da Silva, F.H., Ghijsen, W.E., Verhage, M. (2001) Rab3a is involved in transport of synaptic vesicles to the active zone in mouse brain nerve terminals, *Mol. Biol. Cell* **12**, 3095–3102.

Leonardo, E.D., Hinck, L., Masu, M., Keino-Masu, K., Ackerman, S.L., Tessier-Lavigne, M. (1997) Vertebrate homologs of *C. elegans* UNC-5 are candidates netrin receptors, *Nature* **386**, 833–838.

Li, H.S., Chen, J.H., Wu, W., Fagaly, T., Zhou, L., Yuan, W., Dupuis, S., Jiang, Z.H., Nash, W., Gick, C., Ornitz, D.M., Wu, J.Y., Rao, Y. (1999) Vertebrate slit, a secreted ligand for the transmembrane protein roundabout, is a repellent for olfactory bulb axons, *Cell* **96**, 807–818.

Li, X., Saint-Cyr-Proulx, E., Aktories, K., Lamarche-Vane, N. (2002) Rac1 and Cdc42 but not RhoA or Rho kinase activities are required for neurite outgrowth induced by the netrin-1 receptor DCC (deleted in colorectal cancer) in N1E-115 neuroblastoma cells, *J. Biol. Chem.* **277**, 15207–15214.

Li, Z., Van Aelst, L., Cline, H.T. (2000) Rho GTPases regulate distinct aspects of dendritic arbor growth in Xenopus central neurons *in vivo*, *Nat. Neurosci.* **3**, 217–225.

Leung-Hagesteijn, C., Spence, A.M., Stern, B.D., Zhou, Y., Su, M.W., Hedgecock, E.M., Culotti, J.G. (1992) UNC-5, a transmembrane protein with immunoglobulin and thrombospondin type 1 domains, guides cell and pioneer axon migrations in *C, Elegans. Cell* **71**, 289–299.

Lom, B., Cohen-Cory, S. (1999) Brain-derived neurotrophic factor differentially regulates retinal ganglion cell dendritic and axonal arborization *in vivo*, *J. Neurosci.* **19**, 9928–9938.

Luo, Y., Raible, D., Raper, J.A. (1993) Collapsin: a protein in brain that induces the collapse and paralysis of neuronal growth cones, *Cell* **75**, 217–227.

Lu, Q., Sun, E.E., Klein, R.S., Flanagan, J.G. (2001) Ephrin-B reverse signaling is mediated by a novel PDZ-RGS protein and selectively inhibits G-protein coupled chemoattraction, *Cell* **105**, 69–79.

McAllister, A.K., Lo, D.C., Katz, L.C. (1995) Neurotrophins regulate dendritic growth in developing visual cortex, *Neuron* **15**, 791–803.

McAllister, A.K., Katz, L.C., Lo, D.C. (1996) Neurotrophin regulation of cortical dendritic growth requires activity, *Neuron* **17**, 1057–1064.

McAllister, A.K., Katz, L.C., Lo, D.C. (1997) Opposing roles for endogenous BDNF and NT-3 in regulating cortical dendritic growth, *Neuron* **18**, 767–778.

Miao, H., Wei, B.R., Peehl, D.M., Li, Q., Alexandrou, T., Schelling, J.R., Rhim, J.S., Sedor, J.R., Burnett, E., Wang, B. (2001) Activation of EphA receptor tyrosine kinase inhibits the Ras/MAPK pathway, *Nat. Cell Biol.* **3**, 527–530.

Migaud, M., Charlesworth, P., Dempster, M., Webster, L.C., Watabe, A.M., Makhinson, M., He, Y., Ramsay, M.F., Morris, R.G., Morrison, J.H., O'Dell, T.J., Grant, S.G. (1998) Enhanced long-term potentiation and impaired learning in mice with mutant postsynaptic density-95 protein, *Nature* **396**, 433–439.

Ming, G., Henley, J., Tessier-Lavigne, M., Song, H., Poo, M. (2001) Electrical activity modulates growth cone guidance by diffusible factors, *Neuron* **29**, 441–452.

Missler, M., Zhang, W., Rohlmann, A., Kattenstroth, G., Hammer, R.E., Gottmann, K., Südhof, T.C. (2003) α-neurexins couple Ca^{2+}-channels to synaptic vesicle exocytosis, *Nature* **424**, 939–948.

Naisbitt, S., Kim, E., Tu, J.C., Xiao, B., Sala, C., Valtschanoff, J., Weinberg, R.J., Worley, P.F., Sheng, M. (1999) Shank, a novel family of postsynaptic density proteins that binds to the NMDA receptor/PSD-95/GKAP complex and cortactin, *Neuron* **23**, 569–582.

Naisbitt, S., Kim, E., Weinberg, R.J., Rao, A., Yang, F.C., Craig, A.M., Sheng, M. (1997) Characterization of guanylate kinase-associated protein, a postsynaptic density protein at excitatory synapses that interacts directly with postsynaptic density-95/synapse-associated protein 90, *J. Neurosci.* **17**, 5687–5696.

Nakagawa, S., Brennan, C., Johnson, K.G., Shewan, D., Harris, W.A., Holt, C.E. (2000) Ephrin-B regulates the ipsilateral routing of retinal axons at the optic chiasm, *Neuron* **25**, 599–610.

Nedivi, E., Wu, G.Y., Cline, H.T. (1998) Promotion of dendritic growth by CPG15, an activity-induced signaling molecule, *Science* **281**, 1863–1866.

Nicholson, J.L., Altman, J. (1972) The effects of early hypo- and hyperthyroidism on the development of the rat cerebellar cortex. I. Cell proliferation and differentiation, *Brain Res.* **44**, 13–23.

Niethammer, M., Valtschanoff, J.G., Kapoor, T.M., Allison, D.W., Weinberg, T.M., Craig, A.M., Sheng, M. (1998) CRIPT, a novel postsynaptic protein that binds to the third PDZ domain of PSD-95/SAP90, *Neuron* **20**, 693–707.

Nishiyama, M., Hoshino, A., Tsai, L., Henley, J.R., Goshima, Y., Tessier-Lavigne, M., Poo, M.M., Hong, K. (2003) Cyclic AMP/GMP-dependent modulation of Ca^{2+}-channels sets the polarity of nerve growth-cone turning, *Nature* **424**, 990–995.

Pasterkamp, R.J., Peschon, J.J., Sprigs, M.K., Kolodkin, A.L. (2003) Semaphorin 7A promotes axon outgrowth through integrins and MAPKs, *Nature* **424**, 398–405.

Polleaux, F., Giger, R.J., Ginty, D.D., Kolodkin, A.L., Ghosh, A. (1998) Patterning of cortical efferent projections by semaphoring-neuropilin interactions, *Science* **282**, 1904–1906.

Polleux, F., Morrow, T., Ghosh, A. (2000) Semaphorin 3A is a chemoattractant for cortical apical dendrites, *Nature Neurosci.* **404**, 567–573.

Purves, D., Snider, W.D., Voyvodic, J.T. (1988) Trophic regulation of nerve cell morphology and innervation in the autonomic nervous system, *Nature* **336**, 123–128.

Redmond, L., Oh, S.R., Weinmaster, G., Ghosh, A. (2000) Nuclear Notch1 signaling and the regulation of dendritic development, *Nature Neurosci.* **3**, 30–40.

Rhee, J.S., Betz, A., Pyott, S., Reim, K., Varoqueaux, F., Augustin, I., Hesse, D., Südhof, T., Takahashi, M., Rosenmund, C., Brose, N. (2002) ß-Phorbol ester and diacylglycerol-induced augmentation of transmitter release is mediated by Mun13s and not by PKC, *Cell* **108**, 121–133.

Rong, R., Ahn, J.Y., Huang, H., Nagata, E., Kalman, D., Kapp, J.A., Tu, J., Worley, P.F., Snyder, S.H., Ye, K. (2003) PI3 kinase enhancer-Homer complex couples mGluRI to PI3 kinase, preventing neuronal apoptosis, *Nat. Neurosci.* **6**, 1153–1161.

Rongo, C., Kaplan, J.M. (1999) CaMKII regulates the density of central glutamatergic synapses in vivo, *Nature* **402**, 195–199.

Rosenmund, C., Sigler, A., Augustin, I., Reim, K., Brose, N., Rhee, J.S. (2002) Differential control of vesicle priming and short-term plasticity by Munc13 isoforms, *Neuron* **33**, 411–425.

Rothberg, J.M., Hartley, D.A., Walther, Z., Artivanis-Tsakonas, S. (1988) slit: an EGF-homologous locus of *D. melangaster* involved in the development of the embryonic central nervous system, *Cell* **55**, 1047–1059.

Rothberg, J.M., Jacobs, J.R., Goodman, C.S., Artivanis-Tsakonas, S. (1990) slit: an extracellular protein necessary for development of midline glia and commisural axon pathways contains both EGF and LRR domains, *Genes Dev.* **4**, 2169–2187.

Ruchhoeft, M.L., Ohnuma, S., McNeill, L., Holt, C.E., Harris, W.A. (1999) The neuronal architecture of Xenopus retinal ganglion cells is sculpted by rho-family GTPases *in vivo*, *J. Neurosci.* **19**, 8454–8463.

Sattler, R., Xiong, Z., Lu, W.Y., Hafner, M., MacDonald, J.F., Tymianski, M. (1999) Specific coupling of NMDA receptor activation to nitric oxide neurotoxicity by PSD-95 protein, *Science* **284**, 1845–1848.

Schneggenburger, R., Neher, E. (2000) Intracellular calium dependence of transmitter release rates at a fast central synapse, *Nature* **406**, 889–893.

Schoch, S., Deak, F., Konigstorfer, A., Mozhayeva, M., Sara, Y., Südhof, T.C., Kavalali, E.T. (2001) SNARE funstion analyzed in

synaptobrevin/VAMP knockout mice, *Science* **294**, 1117–1122.

Segal, R.A., Pomeroy, S.L., Stiles, C.D. (1995) Axonal growth and fasciculation linked to differential expression of BDNF and NT3 receptors in developing cerebellar granule cells, *J. Neurosci.* **15**, 4970–4981.

Serafini, T., Kennedy, T.E., Galko, M.J., Mirzayan, C., Jessell, T.M., Tessier-Lavigne, M. (1994) The netrins define a family of axon outgrowth-promoting proteins homologous to *C. elegans* UNC-6, *Cell* **78**, 409–424.

Shewan, D., Dwivedy, A., Anderson, R., Holt, C. (2002) Age-related changes underlie switch in netrin-1 responsiveness as growth cones advance along visual pathway, *Nature Neurosci.* **5**, 955–962.

Snider, W.D. (1988) Nerve growth factor enhances dendritic arborization of sympathetic ganglion cells in developing mammals, *J. Neurosci.* **8**, 2628–2634.

Song, H., Ming, G., He, Z., Lehmann, M., McKerracher, L., Tessier-Lavigne, M., Poo, M. (1998) Conversion of neuronal growth cone responses from repulsion to attraction by cyclic nucleotides, *Science* **281**, 1515–1518.

Sorensen, J.B., Matti, U., Wei, S.H., Nehring, R.B., Voets, T., Ashery, U., Binz, T., Neher, E., Rettig, J. (2002) The SNARE protein SNAP-25 is linked to fast calcium triggering of exocytosis, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1627–1632.

Sperry, R.W. (1963) Chemoffinity in the orderly growth of nerve fiber patterns and connections, *Proc. Natl. Acad. Sci. U.S.A.* **50**, 703–710.

Srivastava, S., Osten, P., Vilim, F.S., Khatri, L., Inman, G., States, B., Daly, C., DeSouza, S., Abagyan, R., Valtschanoff, J.G., Weinberg, R.J., Ziff, E.B. (1998) Novel anchorage of GluR2/3 to the postsynaptic density by the AMPA receptor-binding protein ABP, *Neuron* **21**, 581–591.

Stevens, A., Jacobs, J.R. (2002) Integrins regulate responsiveness to slit repellent signals, *J. Neurosci.* **22**, 4448–4455.

Takahashi, T., Fournier, A., Nakamura, F., Wang, L.H., Murakami, Y., Kalb, R.G., Fujisawa, H., Strittmatter, S.M. (1999) Plexin-neuropilin-1 complexes form sunctional semaphorin-3A receptors, *Cell* **99**, 59–69.

Tamagnone, L., Artigiani, S., Chen, H., He, Z., Ming, G.I., Song, H., Chedotal, A., Winberg, M.L., Goodman, C.S., Poo, M., Tessier-Lavigne, M., Comoglio, P.M. (1999) Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates, *Cell* **99**, 71–80.

Threadgill, R., Bobb, K., Ghosh, A. (1997) Regulation of dendritic growth and remodeling by Rho, Rac, and Cdc42, *Neuron* **19**, 625–634, 583–592.

Tu, J.C., Xiao, B., Naisbitt, S., Yuan, J.P., Petralia, R.S., Brakeman, P., Doan, A., Aakalu, V.K., Lanahan, A.A., Sheng, M., Worley, P.F. (1999) Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of postsynaptic density proteins, *Neuron* **23**, 583–592.

Udin, S. (2000) CPG15 and the dynamics of retinotectal synapses, *Nat. Neurosci.* **3**, 971–972.

Varoqueaux, F., Sigler, A., Rhee, J.S., Brose, N., Enk, C., Reim, K., Rosenmund, C. (2002) Total arrest of spontaneous and evoked synaptic transmission but normal synaptogenesis in the absence of Munc-13-mediated vesicle priming, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9037–9042.

Volkmar, F.R., Greenough, W.T. (1972) Rearing complexity affects branching of dendrites in the visual cortex of the rat, *Science* **176**, 1445–1447.

Wang, H., Bedford, F.K., Brandon, N.J., Moss, S.J., Olsen, R.W. (1999) GABA(A)-receptor-associated protein links GABA(A) receptors and the cytoskeleton, *Nature* **397**, 69–72.

Wang, Y., Sugita, S., Südhof, T.C. (2000) The RIM/NIM family of neuronal C2 domain proteins. Interactions with Rab3 and a new class of Src homology 3 domain proteins, *J. Biol. Chem.* **275**, 20033–20044.

Weisel, T.N., Hubel, D.H. (1963) Effects of visual deprivation on morphology and physiology of cells in the cat's lateral geniculate body, *J. Neurophysiol.* **26**, 978–993.

Wills, Z., Emerson, M., Rusch, J., Bikoff, J., Baum, B., Perrimon, N., Van Vactor, D. (2002) A *Drosophila* homolog of cyclase-associated proteins collaborates with the Abl tyrosine kinase to control midline axon pathfinding, *Neuron* **36**, 611–622.

Wong, K., Ren, X.R., Huang, Y.Z., Xie, Y., Liu, G., Saito, H., Tang, H., Wen, L., Brady-Kalnay, S.M., Mei, L., Wu, J.Y., Xiong, W.C.,

Rao, Y. (2001) Signal transduction in neuronal migration: roles of GTPase activating proteins and the small GTPase Cdc42 in the Slit-Robo pathway, *Cell* **107**, 209–221.

Wooley, C.S., Gould, E., McEwen, B.S. (1990) Exposure to excess glucocorticoids alters dendritic morphology of adult hippocampal pyramidal neurons, *Brain Res.* **531**, 225–231.

Wu, W., Wong, K., Chen, J., Jiang, Z., Dupuis, S., Wu, J.Y., Rao, Y. (1999) Directional guidance of neuronal migration in the olfactory system by the protein Slit, *Nature* **400**, 331–336.

Wyszynski, M., Lin, J., Rao, A., Nigh, E., Beggs, A.H., Craig, A.M., Sheng, M. (1997) Competitive binding of alpha-actinin and calmodulin to the NMDA receptor, *Nature* **385**, 439–442.

Xia, J., Zhang, X., Staudinger, J., Huganir, R.L. (1999) Clustering of AMPA receptors by the synaptic PDZ domain-containing protein PICK1, *Neuron* **22**, 179–187.

Xu, Z., Lai, K.O., Zhou, H.M., Lin, S.C., Ip, N.Y. (2003) Ephrin-B1 reverse signaling activates JNK through a novel mechanism that is independent of tyrosine phosphorylation, *J. Biol. Chem.* **278**, 24767–24775.

Yu, T., Hao, J.C., Lim, W., Tessier-Lavigne, M., Bargmann, C.I. (2002) Shared receptors in axon guidance: SAX-3/Robo signals via UNC-34/Enabled and a netrin-independent UNC-40/DCC function *Nature Neurosci.* **5**, 1147–1154.

Yu, X., Malenka, R.C. (2003) Beta-catenin is critical for dendritic morphogenesis. *Nat. Neurosci* **6**, 1169–1177.

Zheng, J.Q. (2000) Turning of nerve growth cones induced by localized increases in intracellular calcium ions, *Nature* **403**, 89–93.

Neuropeptides, Invertebrate

Edward P. Masler
Nematology Laboratory Plant Sciences Institute, Beltsville, MD, USA

1 Background 169
1.1 Neurosecretion 169
1.2 Neuropeptides 169

2 Invertebrates 170
2.1 Cnidarians 170
2.2 Platyhelminthes 171
2.3 Nematodes 172
2.4 Arthropods 174
2.5 Molluscs 176

3 Discussion 177

Bibliography 177
Books and Reviews 177
Primary Literature 177

Keywords

Invertebrate
An animal that lacks an internal skeleton and maintains body integrity by a variety of strategies including hydrostatic pressure or a rigid exoskeleton.

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

Neuropeptide
This is a generic term for a peptide of as few as 3 to nearly 200 amino acids, produced in specialized neurons, released from neurosecretory cells and usually acting as a hormone.

Neurosecretion
This is the release of material (e.g. biogenic amines, peptides) from secretory granules located in neurosecretory cells in response to internal or external stimuli.

Precursor
Neuropeptides are typically synthesized as members of a large precursor polypeptide that is processed by specialized proteases and other enzymes resulting in production of biologically active peptides.

Receptor
Receptors are signal transduction proteins embedded in cell membranes that bind ligands (e.g. neuropeptides) and transmit signals from the exterior to the cytoplasm. Neuropeptide receptors are typically associated with G-proteins that bind guanine nucleotides and are referred to as GPCRs or G-protein coupled receptors.

Invertebrates present enormous diversity and occupy every ecological niche. They comprise microscopic worms and large squid, sedentary corals and migrating locusts. The 30 invertebrate phyla include animals with the relatively simple nerve net of the cnidarians and the complex central and peripheral nervous systems of insects and molluscs. Common to all invertebrates, regardless of size, habit or complexity, however, is a dependence upon neuropeptides for control of physiological processes. This feature is shared among all animals. Activities ranging from behavior and reproduction to energy metabolism and metamorphosis are under the control of one or more neuropeptides. The lower invertebrates appear to rely upon few families of neuropeptides that include large numbers of a variety of sequences, whereas higher invertebrates have many neuropeptide families with fewer numbers of different sequences within each family. In both cases, however, there is a complex neuropeptide biochemistry. Mechanisms of neuropeptide synthesis, secretion, action, and catabolism are similar to those described for vertebrates, but useful differences exist. These differences facilitate studies of biochemical evolution and provide unique experimental systems. It was, in fact, work on insect metamorphosis early in the twentieth century that led to the concept of neurosecretion. Research into invertebrate neuropeptides is as important to better understand the life processes of animals, in general, as it is to discover environmentally responsible means to control pests.

1 Background

1.1 Neurosecretion

Neurosecretion is the release of signal molecules – peptides, amines, or both – from specialized cells. The molecules effect specific responses after binding to receptors on the target cell. Molecules acting at a target in close proximity to the release site (e.g. from the efferent end of one neuron to the afferent end of the adjoining neuron, or at the membrane of an innervated cell) are considered neurotransmitters. If the release and target sites are distant from one another, then the signal molecule acts as a hormone. *Neuropeptide* is a generic term for regulatory peptides, as short as 3 to nearly 200 amino acids, secreted from neural cells, which act as neurohormones or neurotransmitters. They are produced in specialized neurons, stored in secretory granules and released in response to internal or external stimuli. Invertebrates use an enormous number and variety of neuropeptides to control most physiological processes.

One of the earliest demonstrations that factors from one tissue may affect other tissues not neurally connected was work by Berthold (1849) on castration in roosters. Baylis & Starling (1902) noted that material secreted from the small intestine of dogs stimulated pancreatic activity. They called the active factor *secretin* and coined the term *hormone*. However, neither the testes nor the intestines are neural organs, and it was actually research on an insect that experimentally demonstrated the merging of neural and endocrine activities. Kopec performed ligation and surgical experiments on the larvae of the gypsy moth *Lymantria dispar*, noting that the brain was required for larval development and molting. After further work, he concluded that the brain must be an organ of internal secretion. This was the first report that a neural organ could also act as an endocrine organ. During the same period, Banting & Best made their landmark discovery of insulin and Spiedel identified secretory cells within the spinal kord of skates.

Secretion of hormones by neural tissues is now a principle of neurobiology. The idea that neural tissues also function as endocrine tissues, however, was a point of scientific controversy when proposed in the early 1900s, and many decades would pass before wide acceptance of the neuroendocrine concept. Initial opposition to a revolutionary idea, from work on an invertebrate, eventually gave way to experimental proof from scores of laboratories, and the importance of neuropeptides as key messengers of neurosecretion was recognized.

Invertebrates comprise at least 30 phyla, two of which, Nematoda and Arthropoda, include most of the different animal species known. Neuropeptides are frequently members of families of related sequences, which provide biological variability, redundancy, and an enormously complex picture for the investigator. A complete cataloging of invertebrate neuropeptides and their actions is impractical and not of particular value to an introductory review. Rather, consideration of selected neuropeptides and their functions in representative phyla will be used to illustrate the complex and pervasive action of these molecules throughout the invertebrates.

1.2 Neuropeptides

Neuropeptides are synthesized as large inactive precursors, preproneuropeptides,

that are transported within the cell from the polysome, through the rough endoplasmic reticulum (RER), and on to the Golgi where processing takes place. An N-terminal hydrophobic leader sequence guides translocation through the RER membrane and is removed by an endoproteolytic signal peptidase. The proneuropeptide is transported along the ER to vesicles within the Golgi. Packaged with the proneuropeptide in the vesicle are processing enzymes, including both proteases and nonproteases, required for generation of active neuropeptide, ready for secretion. The processing of all propeptides is similar. Active peptide sequences within precursors are typically flanked by basic amino acids (e.g. K and/or R), which serve as primary cleavage signals for propeptide convertases (or proconvertases). Proconvertases are members of a family of subtilisin-like endoproteases. Those that process neuropeptides are specifically able to function in the acidic environment of the neurosecretory cell vesicle. After proconvertase action, carboxypeptidases trim the C-terminus. Many neuropeptides in invertebrates are further modified at the C-terminus, N-terminus, or both, to protect them against degradation by cellular proteases during transport to target receptors. Amidation requires a C-terminal G, which is converted to an amide by peptidylglycine α-amidating monooxygenase (PAM). PAM is a bifunctional enzyme that has both peptidylglycine α-hydroxylating monooxygenase (PHM) and peptidyl-α-hydroxyglycine α-amidating lyase (PHL) domains. In some invertebrates, amidation is done by separate PHL and PHM enzymes. At the N-terminus, a glutamic acid (E) residue is converted by glutaminyl cyclase into pyroglutamic acid (pE). Peptide processing is a generally conserved system among animals and between different tissues, but specific variations in enzyme complement provide flexibility in processing, depending, for example, upon the tissue or developmental stage requirements. In lower invertebrates, some additional processing proteases that recognize acidic rather than basic residue cleavage sites have been proposed.

2 Invertebrates

2.1 Cnidarians

The most primitive of the animals to have a nervous system are in the phylum Cnidaria, which includes classes Hydrozoa (*Hydra*), Scyphozoa and Cubozoa (jellyfishes), and Anthozoa (corals, anemones). Development of the Hydrozoa and Anthozoa is particularly well studied. The cnidarian nervous system is a simple nerve net lacking the arborization found in higher invertebrates. Nevertheless, the cnidarian nervous system and its control are biochemically quite complex. The chemical messengers seem to be nearly all peptides. In fact, typical neurotransmitters such as acetylcholine, catecholamines, and serotonin have not been found, and the nervous system is primarily peptidergic.

The *Hydra*, a small freshwater polyp known for its ability to regenerate a complete organism from a few cells, can be manipulated to produce a complete epithelial self devoid of a nervous system. Addition of a few neuronal stem cells results in the production of a normal, functioning nervous system. This remarkable plasticity is due to the action of regulatory peptides. In addition, the neurons of cnidarians are multifunctional, serving the

roles of sensory neurons, interneurons, motor neurons, and secretory neurons; roles fulfilled separately among the more specialized neural cells of higher animals. It may not be surprising that of all regulatory peptides in *Hydra*, 50% are believed to be neuropeptides.

Over 800 peptides have been isolated from *Hydra* through the Hydra Peptide Project and about 260 have been sequenced. Although many of the peptides are present in other organisms, many appear to be unique to *Hydra*. At least 60 unique neuropeptide sequences have been reported for anemones. Most cnidarian neuropeptides are approximately 3 to 10 residues long, amidated at the C-terminus and blocked by pE at N-terminus. In addition, the C-terminal sequences usually include a penultimate Arg (R) residue. Thus, groups of cnidarian neuropeptides are known as *RFamides*, *RNamides*, *RPamides*, *RWamides*, and so on, depending upon the terminal residue. All the RXamides have neuromuscular effects.

Although the *RXamide* structure predominates, other sequences are present. Among them are EQPGLWamide, known as metamorphosin A, isolated from the anemone *Anthopleura elegantissima* and Hym-355 (FPQSFLPRGamide) from *Hydra magnapapillata*. Metamorphosin A stimulates hydrazoan metamorphosis, and Hym-355 stimulates neuron differentiation. Analogs of vertebrate gonadotropin-releasing peptide hormones are also present in cnidarians. Thus, cnidarians have numerous neuropeptides that affect neuromuscular activity such as feeding but also have neuropeptides involved in controlling development.

Genes for many cnidarian neuropeptides have been cloned. One notable feature is the high copy number of peptides encoded within some genes. For example, in the sea pansy *Renilla kollikeri*, the precursor of the peptide pEGRFamide contains 36 copies of this peptide sequence and up to 38 copies of peptides have been observed in genes from *A. elegantissima*. These are some of the highest copy numbers reported for neuropeptides, and suggest highly efficient production of certain sequences. There are other cnidarian neuropeptide genes that code for fewer and for single copies of some peptides, and genes that code for more than one different peptide.

Another feature of cnidarian neuropeptide genes is that they code for cleavage sites in precursors that are bounded by acidic residues. This is in addition to the typical basic and dibasic cleavage site residues found in peptide precursors. To accommodate these unusual acidic cleavage sites, additional processing enzymes have been proposed including endoproteases, aminopeptidases, and a dipeptidyl aminopeptidase. Cnidarians appear to require neuropeptides for all neural transmission and for developmental signaling as well. They depend upon an efficient production system that features large copy numbers of active sequences in precursors and a processing pathway that includes proteases that recognize atypical processing sites. The relatively simple nervous system of cnidarians is biochemically highly complex. In this phylum housing the invertebrates with the most primitive of nervous systems, both the number of neuropeptides and their involvement in a variety of complex biochemical events illustrate that primitive is not simple.

2.2 Platyhelminthes

Flatworms comprise over 13 000 species, 10 000 of which are parasitic to animals.

The platyhelminth phylum comprises four classes. Turbellarians are free-living and aquatic (e.g. *Planaria)*, Monogenea that includes ectoparasites of fish and reptiles, Trematoda including endoparasites of vertebrates (e.g. *Schistosoma*), and the Cestodes (tapeworms). Flatworms range in size from the microscopic up to 30-m long tapeworms, and represent the early appearance of bilateral metazoans. Their nervous systems are ladderlike, with central and peripheral components, and unlike the cnidaria have a rudimentary brain.

Most of the information on platyhelminth neuropeptides has been obtained by immunoscreening, and immunoreactivity has been detected throughout the flatworm body including sense and reproductive organs, and alimentary tissues. However, few neuropeptides have been isolated and characterized. The six sequences available are members of two neuropeptide families. Two neuropeptide F sequences are analogous to the vertebrate neuropeptide Y, have conserved proline (P) residues in the N-terminal region of the peptides and characteristic x-R-x-R-y-amide C-terminal motifs. The y is a phenylalanine (F) in flatworm NPFs and a tyrosine (Y) in NPY.

The second family represented is the FMRFamides (more generally known as FaRPs for FMRFamide-related peptides). Four flatworm sequences are characterized from four species. YIRFamide, GYIRFamide, and RYIRFamide were isolated from nonparasitic turbellarians and GNFFRFamide was isolated from the parasitic cestode tapeworm *Moniezia expansa*. The invertebrate homolog, neuropeptide F, was first isolated from *M. expansa* and is expressed abundantly and widely within this animal. There is no gene information available for platyhelminth FMRFamides, but the presence of amidated C-terminals in all isolated sequences and precedent from lower and higher phyla suggest that processing should resemble the mechanisms found in other animals. A gene for *M. expansa* NPF has been characterized, it is highly expressed and codes for a single peptide. A cDNA containing the gene coding for PAM has been isolated from the parasitic trematode *Schistosoma mansoni*, and the expressed *S. mansoni* enzyme has different kinetics from vertebrate PAM.

2.3 Nematodes

Nematodes are unsegmented roundworms with longitudinal nerve chords and a nerve ring. The anatomy of the nervous system is no more complex than the platyhelminths but a greater number of distinct organs is present including mouthparts with a pharyngeal pump to facilitate feeding, an intestine that serves as a metabolic organ somewhat analogous to the liver, gonads, and sexually dimorphic features. Most nematodes are microscopic, but some species are quite large, measuring in meters. As a group, nematodes are second only to the insects in number of species, breadth of distribution, and amount of biomass. They occupy nearly every ecological niche and include both parasitic and nonparasitic species.

Parasitic nematodes infect vertebrates, invertebrates, and plants. Their importance to medicine and agriculture has made them subjects of extensive research efforts to discover ways to effectively control them. A nonparasitic free-living nematode *Caenorhabditis elegans* has become important as a molecular genetic model and tool for discovery of cellular mechanisms in both vertebrates and invertebrates. In fact, the *C. elegans*

genome was the first animal genome to be completely sequenced.

The neuropeptides predominantly studied in nematodes are members of the FaRP family of sequences. The neuromuscular effects of these peptides have made them a focus of veterinary medical research to discover treatments for parasitic disease of livestock, and the discovery of large numbers of FaRP genes in *C. elegans* has captured the attention of molecular biologists. Two species, then, have predominated nematode neuropeptide research: *Ascaris suum*, an important parasite of livestock, and the molecular model *C. elegans*.

The first neuropeptide isolated from a nematode was the FaRP KNEFIRFamide from *A. suum*, and the first FaRP gene cloned from nematodes was from *C. elegans*. To date, in fact, all neuropeptides isolated from nematodes have been FaRPs. These include 19 different sequences from *A. suum*, 1 FaRP from another animal-parasitic nematode, *Haemonchus contortus*, 5 from the free-living species *Panagrellus redivivus*, and 1 from *C. elegans*. Nematode FaRP genes contain coding sequences that predict large numbers of FaRP sequences. *C. elegans* has 22 FaRP-like peptide genes (*flp* genes) coding for approximately 60 different FaRP sequences, the plant-parasitic nematode *Globodera pallida* has 5 *gpflp* genes coding for 14 different sequences, and a gene cloned from *A. suum* (*afp-1*) codes for 6 different sequences. Around 70 different FaRPs have been identified in nematodes, with only a minority having been tested physiologically, usually in a muscle assay system using *A. suum*. All have neuromuscular effects, but not all of the effects are the same, leading to the hypothesis that a variety of FaRP receptors exist. This argument is supported by the large number of unique sequences identified.

Some intriguing questions arise from a survey of FaRP sequences in nematodes. In *C. elegans*, many of the *flp* genes are expressed throughout development. Nematodes have an egg stage, four larval stages, and adults. *C. elegans flp*s 1 to 12 are expressed in all 6 stages but *flp*s 8, 9, and 13 are expressed only during the egg and larval stages and not in the adult stage. This suggests developmental involvement of at least some FaRPs.

Among the many sequences identified in nematodes, KHEYLRFamide is present in all species examined and is the most abundant FaRP in *A. suum, H. contortus, C. elegans* and *P. redivivus*. In both *C. elegans* and *G. pallida,* the genes coding for KHEYLRFamide also code for multiple copies of the peptide. A second FaRP KSAYMRFamide is present in *A. suum, G. pallida, C. elegans* and *P. redivivus*, and the genes in *C. elegans* and *G. pallida* also code for multiple copies of this FaRP. The abundance of these two sequences and their appearance in all nematodes suggests that they have some fundamental importance to all species.

In concert with this shared importance, the wide variety of unique sequences observed suggests some species-specific or developmental stage-specific roles. In addition, it is interesting to note that among the complex variety of nematode FaRPs, only two (above), are widely shared and most FaRPs have been identified in either parasitic or nonparasitic species, but not both. Examples include SADPNFLRFamide and SDPNFLRFamide, which are present only in the free-living species *C. elegans* and *P. redivivus*, and AVPGVLRFamide, which is exclusive to the animal – (*A. suum*) and plant – (*G. pallida*) parasitic nematodes.

Can FaRP cocktails be used to sort nematode species for phylogenetic comparison?

Are there really a few *universal* FaRPs in nematodes? What are their functions among different species? Are there FaRPs associated with parasitism? As more species are examined, such questions may be answered. Levels of specific FaRPs in *C. elegans* can affect feeding and social behavior, and may also influence activation of specific receptors. FaRPs, then, do not solely affect neuromuscular activity.

Not all nematode neuropeptides are FaRPs. Immunochemical screens using antisera against vertebrate and invertebrate antigens provide circumstantial evidence for numerous other neuropeptides in nematodes. At least one report showed that *Ascaris* homogenates elicit responses in insects similar to those associated with the insect metabolic neuropeptide adipokinetic hormone. More directly, a screen of the *C. elegans* genome has revealed over 30 non-*flp* genes encoding more than 50 putative neuropeptides. Thus, nematodes must depend upon a large number and complex variety of neuropeptides for physiological regulation, and most have yet to be examined.

2.4 Arthropods

Studies of neuropeptides in arthropods constitute a much more comprehensive approach than in the lower invertebrates. Well-defined organs with specific physiological functions, a sophisticated nervous system, and physical size of many arthropods facilitate investigations with a variety of biochemical, physiological, molecular biological, and anatomical methods. This is especially true in the insects, where many families of neuropeptides have been discovered, and where member neuropeptides are categorized by physiological processes controlled as well as by sequence.

Developmental, metabolic, neuromuscular, behavioral, and other physiological activities have been examined. Developmental neuropeptides include the ecdysiotropins that control levels of steroid hormones necessary for molting and egg production. The allatotropins and allatostatins regulate titers of specific fatty acids, juvenile hormones, responsible for determining developmental age. Metabolic neuropeptides regulate lipid mobilization, energy production, diuresis, and heart rate. Some behavioral neuropeptides control highly orchestrated movements required for eclosion (adult emergence), while others regulate the production of exquisitely specific and potent pheromones, volatile chemicals released by females to attract males. Myotropic neuropeptides represent a varied group, including FaRPs, that control locomotion, flight, egg laying, and the other neuromuscular demands of highly complex organisms. The literature on these and the many other arthropod neuropeptides is voluminous, and numerous reviews are available. Here, two neuropeptide families are presented as examples of the status of neuropeptide research in insects.

Pheromones are semiochemicals released by a female to attract a mate. In lepidopterans, the subesophageal gland produces pheromones in response to a neuropeptide – pheromone biosynthesis-activating neuropeptide (PBAN). PBAN was initially isolated from the moth *Helicoverpa zea* and subsequently identified and prepared from a number of other species. PBANs have only been identified in moths and butterflies, although a homolog is predicted in *Drosophila*. PBANs are 33 to 34 amino acids long, with high sequence homologies (~70–90%), and have an amidated C-terminus with a FSPRLamide motif. The

C-terminal pentapeptide was found to be the shortest PBAN fragment to retain pheromonotropic activity (its active core), a discovery that stimulated research to produce novel insect control agents based upon neuropeptides. The approach consists of two fundamental strategies. First, since pheromonotropic activity is retained within a short peptide sequence, it is possible to synthesize analogs of the C-terminus, and prepare various iterations containing different amino acid substitutions. The peptides are screened for antagonistic activity, and these lead sequences are used to design more potent antagonists. Second, peptides are polar molecules that do not readily pass through the insect cuticle, and are also susceptible to proteolytic attack within the animal. To address these problems, cyclic versions of lead peptides are synthesized such that polarity is reduced and resistance to proteolysis is increased. Through this approach, antagonists have been identified and the cuticle successfully penetrated. Improvement in design and potency of the peptide mimics, needed before practical use can be realized, represent major challenges, but the recent molecular cloning of the PBAN receptor will undoubtedly be of significant help.

A surprising feature of the PBAN FSPRLamide C-terminus is that it is shared among other neuropeptides with different biological activities and in various insect species. Myotropic neuropeptides called *pyrokinins*, isolated from cockroaches and locusts, stimulate muscle contraction in the gut and oviduct. They were also found to affect cuticle pigmentation in moth larvae, pupariation in flies, and pheromone production in moths. Reciprocal experiments demonstrated that moth PBAN could stimulate muscle contractions in other insects. Subsequent experiments have established a pyrokinin/PBAN family in insects. Common to all the neuropeptides examined was the FxPRLamide motif. In PBAN, $x = \mathrm{S}$, and in pyrokinin, $x = \mathrm{T}$. Cloning of the PBAN gene revealed a precursor peptide containing sequences of five peptides each containing the FxPRL structure. One sequence was PBAN and the other four are diapause hormone-like peptides, with FGPRLamide at the C-terminus. Thus, the FxPRLamide motif is associated with developmental and myotropic, as well as behavioral, activities. Specificity of action probably comes from association of the entire molecule with the receptor (as opposed to the active core only) and the population of receptors present in target cells.

Sequencing of the *Drosophila melanogaster* genome greatly expands analysis of insect neuropeptides. As with *C. elegans*, screens have been initiated to discover neuropeptide genes and genes for neuropeptide receptors. To date, 44 genes for G-protein coupled receptors (GPCRs) have been detected including those predicted to bind neuropeptide Y homolog, gonadotropin-releasing hormone, allatostatins, and a number of other receptors with homologs in humans, cnidarians, and other insects. The 22 neuropeptide genes discovered include those for adipokinetic hormone, ecdysiotropin, PBAN, and FaRPs.

FaRPs in *Drosophila* are sorted into three classes on the basis of amino acid sequences: sulfakinins, myosuppressins, and FMRFamides. All have neuromuscular activities, but these vary depending upon the tissue. Single genes were cloned for each group, only one gene for each group was found in the genome screen, and these had been cloned. The single FMRFamide gene codes for a precursor containing a total of 11 peptides with 6

different sequences. This compares with the multiple genes and sequences found in the lower invertebrates (e.g. nematodes). In molluscs, FMRFamide is also coded by a single gene. *Drosophila* FMRFamide receptors (GPCRs) have been cloned and characterized. They express varying preferences for the different *Drosophila* FMRFamides, suggesting that additional FMRFamide receptors may also exist and offer a source of specificity.

2.5 Molluscs

Since its discovery in the clam *Macrocallista* nearly 40 years ago, FMRFamide and related FaRPs have been described in nearly every invertebrate examined, and have stimulated a great deal of invertebrate neurochemical research. Molluscs have single genes coding for FaRPs but exhibit a variety of FaRP production strategies. The snail *Lymnaea* has a single five-exon FaRP gene that is alternatively spliced, resulting in 12 FaRP sequences including FMRFamide: The bivalve *Mytilus* gene codes for 16 copies of FMRFamide and three other peptides with no alternative splicing. In molluscs with the most complex nervous systems (the cephalopods: e.g. *Loligo*, squid; *Sepia*, cuttlefish), the single gene codes for 14 FaRPs including FMRFamide, and in *Aplysia*, a marine snail that has become a widely used model in neurochemical research, the FaRP gene encodes 28 copies of FMRFamide. The presence of a single FaRP gene in molluscs, as in insects, is in sharp contrast with the multiple gene strategy in the lower invertebrates.

The biosynthetic machinery that results in neuropeptide variety includes multiple genes and multi-exon genes, alternative gene splicing, propeptide and enzyme sorting, and posttranslational processing. One of the most prominent model systems used to examine neuropeptideff biosynthesis has been provided by *Aplysia*. The *Aplysia* nervous system is more sophisticated than the lower invertebrates, but contains relatively few and large neurons compared with vertebrates. Electrophysiological studies of stereotyped behavior and anatomical identification of the involved neurons were accomplished and, early on, the roles of neurochemicals including neuropeptides were examined. A most critical behavior in the gastropod molluscs is egg laying, and all of the above experimental approaches came to be focused on egg laying in *Aplysia*.

A remarkable feature of *Aplysia* is that egg laying is controlled by two clusters of peptidergic neurons, so-called bag cell clusters, each containing hundreds of homogeneous neurons producing the protein precursor of egg-laying hormone (ELH). Few systems allow the study of large numbers of homogeneous neurons producing one common precursor, and *Aplysia* became a particularly useful model for the study of neuropeptide regulation. The ELH gene encodes a 271 amino acid precursor containing the sequences for ELH, bag cell proteins (BCP) and acidic peptide (AP). These sequences are cleaved from the precursor and differentially packaged and transported. ELH is 36, BCPs range from 5 to 19, and AP is 27 amino acids long. The specific function of AP is not known, but all of the peptides act to coordinate egg-laying behavior. The extensive posttranslational processing of ELH peptides has been thoroughly studied, and the processing enzymes, including proconvertases, carboxypeptidase E, and the alpha-amidating enzymes have been characterized and their genes cloned.

3 Discussion

Invertebrates represent an enormous component of animal life, not only in terms of sheer numbers and biomass but also in diversity and distribution. Their importance to ecological health, and challenges to animal and plant health, makes their study of practical as well as academic importance. The examination of invertebrates is also essential in an evolutionary sense, as for example, in research on biochemical evolution. Throughout the invertebrate phyla, neuropeptides control all of the most critical physiological processes. The simplest nervous systems are biochemically complex, relying upon numerous neuropeptides to coordinate fundamental functions such as muscle contraction.

More complex animals use increased varieties of different neuropeptides in a number of families to control fundamental and specialized activities. While some neuropeptides such as PBAN are found only in a limited number of species, others such as the FaRPs are universally distributed among all of the phyla. Although processing mechanisms vary in specifics between the lowest and highest phyla, they are otherwise remarkably conserved and used by all invertebrates to produce active neuropeptides of all sizes. The enormous diversity among invertebrates is bridged by neuropeptides and their production.

Finally, studies of neuropeptide biochemistry and molecular biology in invertebrates have an effect on studies of all other animals. From the discovery of the neuroendocrine phenomenon in insects, genetic and developmental models in nematodes and *Drosophila*, and the neuropeptide processing contributions of *Aplysia*, invertebrate neuropeptide research continues to have far-reaching effects.

Bibliography

Books and Reviews

Brusca, R.C., Brusca, G.J. (1990) *Invertebrates*, Sinauer Associates, Sunderland, MA.

Day, T.A., Maule, A.G. (1999) Parasitic peptides! The structure and function of neuropeptides in parasitic worms, *Peptides* **20**, 999–1019.

Grimmelikhuijzen, C.J., Hauser, F., Eriksen, K.K., Williamson, M. (1999) Invertebrate neurohormones and their receptors, *Results Probl. Cell Differ.* **26**, 339–362.

Halton, D.W., Shaw, C., Maule, A.G., Smart, D. (1994) Regulatory peptides in helminth parasites, *Adv. Parasitol.* **34**, 164–217.

Kaldany, R.R., Nambu, J.R., Scheller, R.H. (1985) Neuropeptides in identified *Aplysia* neurons, *Annu. Rev. Neurosci.* **8**, 431–455.

Mains, R.E., Eipper, B.A. (1999) Peptides, in: Siegel, G.J., Agranoff, B.W., Albers, R.W., Fisher, S.K., Uhler, M.D. (Eds.) *Basic Neurochemistry: Molecular, Cellular and Medical Aspects*, Lippincott, Williams & Williams, New York, pp. 363–382.

Masler, E.P., Kelly, T.J., Menn, J.J. (1993) Insect neuropeptides: discovery and application in insect management, *Arch. Insect Biochem. Physiol.* **22**, 87–111.

Scharrer, B. (1987) Neurosecretion: beginnings and new directions in neuropeptide research, *Annu. Rev. Neurosci.* **10**, 1–17.

Shaw, C. (1996) Neuropeptides and their evolution, *Parasitology* **113**, s35–s45.

Strand, F. (1999) *Neuropeptides: Regulators of Physiological Processes*, MIT Press, Cambridge, MA.

Primary Literature

Altstein, M., Ben-Aziz, O., Schefler, I., Zeltser, I., Gilon, C. (2000) Advances in the application of neuropeptides in insect control, *Crop Prot.* **19**, 547–555.

Anctil, M. (2000) Evidence for gonadotropin-releasing hormone-like peptides in a cnidarian nervous system, *Gen. Comp. Endocrinol.* **119**, 317–328.

Banting, F.G., Best, C.H. (1922) The discovery of insulin, *J. Lab. Clin. Med.* **5**, 256–271.

Baylis, W.M., Starling, E.H. (1902) The mechanism of pancreatic secretion, *J. Physiol.* **28**, 325–335.

Berthold, A.A. (1849) Transplantation der Hoden, *Arch. Anat. Physiol. Wissen. Med.* **16**, 42–46.

Bosch, T.C., Fujisawa, T. (2001) Polyps, peptides and patterning, *Bioessays* **23**, 420–427.

Cazzamali, G., Grimmelikhuijzen, C.J. (2002) Molecular cloning and functional expression of the first insect FMRFamide receptor, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12073–12078.

Choi, M.Y., Fuerst, E.J., Rafaeli, A., Jurenka, R. (2003) Identification of a G protein-coupled receptor for pheromone biosynthesis activating neuropeptide from pheromone glands of the moth *Helicoverpa zea*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9721–9726.

Chun, J.Y., Korner, J., Kreiner, T., Scheller, R.H., Axel, R. (1994) The function and differential sorting of a family of *Aplysia* prohormone processing enzymes, *Neuron* **12**, 831–844.

Cowden, C., Stretton, A.O.W., Davis, R.E. (1989) AF1, a sequenced bioactive neuropeptide isolated from the nematode *Ascaris suum*, *Neuron* **2**, 1465–1473.

Curry, W.J., Shaw, C., Johnston, C.F., Thirn, L., Buchanan, K.D. (1992) Neuropeptide F: primary structure from the turbellarian, *Artioposthia triangulata*, *Comp. Biochem. Physiol.* **101C**, 269–274.

Darmer, D., Hauser, F., Nothacker, H.P., Bosch, T.C., Williamson, M., Grimmelikhuijzen, C.J. (1998) Three different prohormones yield a variety of Hydra-RFamide (Arg-Phe-NH2) neuropeptides in *Hydra magnipapillata*, *Biochem. J.* **332**, 403–412.

Davenport, T.R.B., Eaves, L.A., Hayes, T.K., Lee, D.L., Isaac, R.E. (1994) The detection of AKH/HrTH-like peptides in *Ascaridia galli* and *Ascaris suum* using an insect hyperglycaemic bioassay, *Parasitology* **108**, 479–485.

Davis, R.E., Stretton, A.O. (2001) Structure-activity relationships of 18 endogenous neuropeptides on the motor nervous system of the nematode *Ascaris suum*, *Peptides* **22**, 7–23.

DesGroseillers, L. (1990) Molecular aspects of egg-laying behavior in *Aplysia californica*, *Behav. Genet.* **20**, 251–264.

Edison, A.S., Messinger, L.A., Stretton, A.O.W. (1997) afp-1: a gene encoding multiple transcripts of a new class of FMRFamide-like neuropeptides in the nematode *Ascaris suum*, *Peptides* **18**, 929–935.

Fan, X., Nagle, G.T. (1996) Molecular cloning of Aplysia neuronal cDNAs that encode carboxypeptidases related to mammalian prohormone processing enzymes, *DNA Cell Biol.* **15**, 937–945.

Fan, X., Spijker, S., Akalal, D.B., Nagle, G.T. (2000) Neuropeptide amidation: cloning of a bifunctional alpha-amidating enzyme from *Aplysia*, *Brain Res. Mol. Brain Res.* **82**, 25–34.

Favrel, P., Lelong, C., Mathieu, M. (1998) Structure of the cDNA encoding the precursor for the neuropeptide FMRFamide in the bivalve mollusc *Mytilus edulis*, *NeuroReport* **9**, 2961–2965.

Gade, G., Goldsworthy, G.J. (2003) Insect peptide hormones: a selective review of their physiology and potential application for pest control, *Pest Manag. Sci.* **59**, 1063–1075.

Garden, R.W., Shippy, S.A., Li, L., Moroz, T.P., Sweedler, J.V. (1998) Proteolytic processing of the *Aplysia* egg-laying hormone prohormone, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3972–3977.

Geary, T.G., Marks, N.J., Maule, A.G., Bowman, J.W., Alexander-Bowman, S.J., Day, T.A., Larsen, M.J., Kubiak, T.M., Davis, J.P., Thompson, D.P. (1999) Pharmacology of FMRFamide-related peptides in helminths, *Ann. N. Y. Acad. Sci.* **897**, 212–227.

Geary, T.G., Price, D.A., Bowman, J.W., Winterrowd, C.A., Mackenzie, C.D., Garrison, R.D., Williams, J.F., Friedman, A.R. (1992) Two FMRFamide-like peptides from the free-living nematode Panagrellus redivivus, *Peptides* **13**, 209–214.

Grimmelikhuijzen, C.J., Westfall, J.A. (1995) The nervous systems of cnidarians, *EXS* **72**, 7–24.

Grimmelikhuijzen, C.J., Leviev, I., Carstensen, K. (1996) Peptides in the nervous systems of cnidarians: structure, function, and biosynthesis, *Int. Rev. Cytol.* **167**, 37–89.

Grimmelikhuijzen, C.J., Darmer, D., Schmutzler, C., Carstensen, K., Moosler, A., Nothacker, H.P., Reinscheid, R.K., Vollert, H., Rinehart, K.L., McFarlane, I.D. (1992) The peptidergic nervous system of coelenterates, *Prog. Brain Res.* **92**, 137–148.

Gustafsson, M.K., Halton, D.W., Kreshchenko, N.D., Movsessian, S.O., Raikova, O.I., Reuter, M., Terenina, N.B. (2002) Neuropeptides in flatworms, *Peptides* **23**, 2053–2061.

Hewes, R.S., Taghert, P.H. (2001) Neuropeptides and neuropeptide receptors in the *Drosophila melanogaster* genome, *Genome Res.* **11**, 1126–1142.

Homberg, U. (2002) Neurotransmitters and neuropeptides in the brain of the locust, *Microsc. Res. Tech.* **56**, 189–209.

Johnston, R.N., Shaw, C., Halton, D.W., Verhaert, P., Baguna, J. (1995) GYIRFamide: a novel FMRFamide-related peptide (FaRP) from the triclad turbellarian, *Dugesia tigrina*, *Biochem. Biophys. Res. Commun.* **209**, 689–697.

Johnston, R.N., Shaw, C., Halton, D.W., Verhaert, P., Blair, K.L., Brennan, G.P., Price, D.A., Anderson, P.A.V. (1996) Isolation, localization, and bioactivity of the FMRFamide-related neuropeptides GYIRFamide and YIRFamide from the marine turbellarian *Bdelloura candida*, *J. Neurochem.* **67**, 814–821.

Keating, C.D., Holden-Dye, L., Thorndyke, M.C., Williams, R.G., Mallett, A., Walker, R.J. (1995) The FMRFamide-like neuropeptide AF2 is present in the parasitic nematode *Haemonchus contortus*, *Parasitology* **111**, 515–521.

Kimber, M.J., Fleming, C.C., Bjourson, A.J., Halton, D.W., Maule, A.G. (2001) FMRFamide-related peptides in potato cyst nematodes, *Mol. Biochem. Parasitol.* **116**, 199–208.

Kimber, M.J., Fleming, C.C., Prior, A., Jones, J.T., Halton, D.W., Maule, A.G. (2002) Localisation of *Globodera pallida* FMRFamide-related peptide encoding genes using in situ hybridisation, *Int. J. Parasitol.* **32**, 1095–1105.

Klowden, M.J. (2003) Contributions of insect research toward our understanding of neurosecretion, *Arch. Insect Biochem. Physiol.* **53**, 101–114.

Kopec, S. (1917) Experiments on metamorphosis in insects, *Bull. Int. Acad. Cracov. B* **57**, 57–60.

Kopec, S. (1922) Studies on the necessity of the brain for the inception of insect metamorphosis, *Biol. Bull.* **42**, 323–342.

Kubiak, T.M., Larsen, M.J., Nulf, S.C., Zantello, M.R., Burton, K.J., Bowman, J.W., Modric, T., Lowery, D.E. (2003) Differential activation of "social" and "solitary" variants of the *Caenorhabditis elegans* G protein-coupled receptor NPR-1 by its cognate ligand AF9, *J. Biol. Chem.* **278**, 33724–33729.

Leitz, T., Morand, K., Mann, M. (1994) Metamorphosin A: a novel peptide controlling development of the lower metazoan *Hydractinia echinata* (Coelenterata, Hydrozoa), *Dev. Biol.* **163**, 440–446.

Leviev, I., Williamson, M., Grimmelikhuijzen, C.J. (1997) Molecular cloning of a preprohormone from *Hydra magnipapillata* containing multiple copies of Hydra-L Wamide (Leu-Trp-NH2) neuropeptides: evidence for processing at ser and asn residues, *J. Neurochem.* **68**, 1319–1325.

Li, L., Garden, R.W., Floyd, P.D., Moroz, T.P., Gleeson, J.M., Sweedler, J.V., Pasa-Tolic, L., Smith, R.D. (1999) Egg-laying hormone peptides in the aplysiidae family, *J. Exp. Biol.* **202**, 2961–2973.

Ma, P.W.K., Knipple, D.C., Roelofs, W.L. (1994) Structural organization of the *Helicoverpa zea* gene encoding the precursor protein for pheromone biosynthesis-activating neuropeptide and other neuropeptides, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6506–6510.

Mair, G.R., Halton, D.W., Shaw, C., Maule, A.G. (2000) The neuropeptide F (NPF) encoding gene from the cestode, *Moniezia expansa*, *Parasitology* **120**, 71–77.

Mair, G.R., Niciu, M.J., Stewart, M.T., Brennan, G., Omar, H., Halton, D.W., Mains, R., Eipper, B.A., Maule, A.G., Day, T.A. (2004) A functionally atypical amidating enzyme from the human parasite *Schistosoma mansoni*, *FASEB J.* **18**, 114–121.

Marks, N.J., Maule, A.G., Li, C., Nelson, L.S., Thompson, D.P., Alexander-Bowman, S., Geary, T.G., Halton, D.W., Verhaert, P., Shaw, C. (1999) Isolation, pharmacology and organization of KPSFVRFamide: a neuropeptide from *Caenorhabditis elegans*, *Biochem. Biophys. Res. Commun.* **254**, 222–230.

Masler, E.P. (1996) Insect Neuropeptides, in: Meyers, R.A. (Ed.) *Encyclopedia of Molecular Biology and Molecular Medicine*, Vol. 4, VCH Publishers, Weinheim, FRG, pp. 188–198.

Maule, A.G., Halton, D.W., Shaw, C., Thirn, L., Johnston, C.F., Fairweather, I. (1991) Neuropeptide F: a novel parasitic flatworm regulatory peptide from *Moniezia expansa* (Cestoda: Cyclophyllidea), *Parasitology* **102**, 309–316.

Maule, A.G., Shaw, C., Halton, D.W., Brennan, G.P., Johnston, C.F., Moore, S. (1992) Neuropeptide F (*Moniezia expansa*): localization and characterization using specific antisera, *Parasitology* **105**, 505–512.

Maule, A.G., Shaw, C., Bowman, J.W., Halton, D.W., Thompson, D.P., Geary, T.G.,

Thim, L. (1994a) The FMRFamide-like neuropeptide AF2 (*Ascaris suum*) is present in the free-living nematode, *Panagrellus redivivus* (Nematoda, Rhabditida), *Parasitology* **109**, 351–356.

Maule, A.G., Shaw, C., Bowman, J.W., Halton, D.W., Thompson, D.P., Geary, T.G., Thim, L. (1994b) KSAYMRFamide: a novel FMRFamide-related heptapeptide from the free-living nematode, *Panagrellus redivivus*, which is myoactive in the parasitic nematode, Ascaris suum, *Biochem. Biophys. Res. Commun.* **200**, 973–980.

Meeusen, T., Mertens, I., Clynen, E., Baggerman, G., Nichols, R., Nachman, R.J., Huybrechts, R., De Loof, A., Schoofs, L. (2002) Identification in *Drosophila melanogaster* of the invertebrate G protein-coupled FMRFamide receptor, *Proc. Natl. Acad. Sci U.S.A.* **99**, 15363–15368.

Merte, J., Nichols, R. (2002) *Drosophila melanogaster* FMRFamide-containing peptides: redundant or diverse functions? *Peptides* **23**, 209–220.

Minobe, S., Koizumi, O., Sugiyama, T. (1995) Nerve cell differentiation in nerve-free tissue of epithelial hydra from precursor cells introduced by grafting. I. Tentacles and hypostome, *Dev. Biol.* **172**, 170–181.

Mitgutsch, C., Hauser, F., Grimmelikhuijzen, C.J. (1999) Expression and developmental regulation of the Hydra-RFamide and Hydra-LWamide preprohormone genes in *Hydra*: evidence for transient phases of head formation, *Dev. Biol.* **207**, 189–203.

Moffett, C.L., Beckett, A.M., Mousley, A., Geary, T.G., Marks, N.J., Halton, D.W., Thompson, D.P., Maule, A.G. (2003) The ovijector of *Ascaris suum*: multiple response types revealed by Caenorhabditis elegans FMRFamide-related peptides, *Int. J. Parasitol.* **33**, 859–876.

Nachman, R.J., Teal, P.E., Ujvary, I. (2001) Comparative topical pheromonotropic activity of insect pyrokinin/PBAN amphiphilic analogs incorporating different fatty and/or cholic acid components, *Peptides* **22**, 279–285.

Nagle, G.T., de Jong-Brink, M., Painter, S.D., Bergamin-Sassen, M.M., Blankenship, J.E., Kurosky, A. (1990) Delta-bag cell peptide from the egg-laying hormone precursor of *Aplysia*. Processing, primary structure, and biological activity, *J. Biol. Chem.* **265**, 22329–22335.

Nagle, G.T., Garcia, A.T., Knock, S.L., Gorham, E.L., Van Heumen, W.R., Kurosky, A. (1995) Molecular cloning, cDNA sequence, and localization of a prohormone convertase (PC2) from the *Aplysia* atrial gland, *DNA Cell. Biol.* **14**, 145–154.

Nassel, D.R. (2000) Functional roles of neuropeptides in the insect central nervous system, *Naturwissenschaften* **87**, 439–449.

Nathoo, A.N., Moeller, R.A., Westlund, B.A., Hart, A.C. (2001) Identification of neuropeptide-like protein gene families in *Caenorhabditis elegans* and other species, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14000–14005.

Nelson, L.S., Kim, K., Memmott, J.E., Li, C. (1998) FMRFamide-related gene family in the nematode, *Caenorhabditis elegans*, *Mol. Brain Res.* **58**, 103–111.

Newcomb, R., Scheller, R.H. (1987) Proteolytic processing of the *Aplysia* egg-laying hormone and R3-14 neuropeptide precursors, *J. Neurosci.* **7**, 854–863.

Newcomb, R., Fisher, J.M., Scheller, R.H. (1988) Processing of the egg-laying hormone (ELH) precursor in the bag cell neurons of *Aplysia*, *J. Biol. Chem.* **263**, 12514–12521.

Ouimet, T., Castellucci, V.F. (1997) Gene expression and regulation of *Aplysia californica* prohormone convertases aPC1B and aPC2, *J. Neurochem.* **68**, 1031–1040.

Predel, R., Eckert, M. (2000) Neurosecretion: peptidergic systems in insects, *Naturwissenschaften* **87**, 343–350.

Raina, A.K. (1993) Neuroendocrine control of sex pheromone biosynthesis in lepidoptera, *Annu. Rev. Entomol.* **38**, 329–349.

Raina, A.K., Kempe, T.G. (1992) Structure activity studies of PBAN of *Helicoverpa zea*, *Insect Biochem. Mol. Biol.* **22**, 221–225.

Reinscheid, R.K., Grimmelikhuijzen, C.J. (1994) Primary structure of the precursor for the anthozoan neuropeptide antho-RFamide from *Renilla kollikeri*: evidence for unusual processing enzymes, *J. Neurochem.* **62**, 1214–1222.

Rogers, C., Reale, V., Kim, K., Chatwin, H., Li, C., Evans, P., de Bono, M. (2003) Inhibition of *Caenorhabditis elegans* social feeding by FMRFamide-related peptide activation of NPR-1, *Nat. Neurosci.* **6**, 1178–1185.

Rosoff, M.L., Burglin, T.R., Li, C. (1992) Alternatively spliced transcripts of the flp-1 gene encode distinct FMRFamide-like

peptides in *Caenorhabditis elegans*, *J. Neurosci.* **12**, 2356–2361.

Rouille, Y., Duguay, S.J., Lund, K., Furuta, M., Gong, Q., Lipkind, G., Oliva, A.A. Jr, Chan, S.J., Steiner, D.F. (1995) Proteolytic processing mechanisms in the biosynthesis of neuroendocrine peptides: the subtilisin-like proprotein convertases, *Front. Neuroendocrinol.* **16**, 322–361.

Santama, N., Benjamin, P.R. (2000) Gene expression and function of FMRFamide-related neuropeptides in the snail *Lymnaea*, *Microsc. Res. Tech.* **49**, 547–556.

Sato, Y., Oguchi, M., Menjo, N., Imai, K., Saito, H., Ikeda, M., Isobe, M., Yamashita, O. (1993) Precursor polyprotein for multiple neuropeptides secreted from the suboesophageal ganglion of the silkworm *Bombyx mori*: characterization of the cDNA encoding the diapause hormone precursor and identification of additional peptides, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3251–3255.

Scheller, R.H., Jackson, J.F., McAllister, L.B., Schwartz, J.H., Kandel, E.R., Axel, R. (1982) A family of genes that codes for ELH, a neuropeptide eliciting a stereotyped pattern of behavior in *Aplysia*, *Cell* **28**, 707–719.

Schmutzler, C., Darmer, D., Diekhoff, D., Grimmelikhuijzen, C.J. (1992) Identification of a novel type of processing sites in the precursor for the sea anemone neuropeptide Antho-RFamide (<Glu-Gly-Arg-Phe-NH2) from *Anthopleura elegantissima*, *J. Biol. Chem.* **267**, 22534–22541.

Spiedel, C.C. (1919) Gland-cells of internal secretion in the spinal cord of the skates, *Carnegie Inst. Washington* **13**, 1.

Sweedler, J.V., Li, L., Floyd, P., Gilly, W. (2000) Mass spectrometric survey of peptides in cephalopods with an emphasis on the FMRFamide-related peptides, *J. Exp. Biol.* **203**, 3565–3573.

Takahashi, T., Koizumi, O., Ariura, Y., Romanovitch, A., Bosch, T.C., Kobayakawa, Y., Mohri, S., Bode, H.R., Yum, S., Hatta, M., Fujisawa, T. (2000) A novel neuropeptide, Hym-355, positively regulates neuron differentiation in *Hydra*, *Development* **127**, 997–1005.

Taussig, R., Scheller, R.H. (1986) The Aplysia FMRFamide gene encodes sequences related to mammalian brain peptides, *DNA* **5**, 453–461.

Vanden Broeck, J. (2001) Neuropeptides and their precursors in the fruitfly, *Drosophila melanogaster*, *Peptides* **22**, 241–254.

Yum, S., Takahashi, T., Hatta, M., Fujisawa, T. (1998) The structure and expression of a preprohormone of a neuropeptide, Hym-176 in *Hydra magnipapillata*, *FEBS Lett.* **439**, 31–34.

Zhang, T.Y., Sun, J.S., Zhang, L.B., Shen, J.L., Xu, W.H. (2004) Cloning and expression of the cDNA encoding the FXPRL family of peptides and a functional analysis of their effect on breaking pupal diapause in *Helicoverpa armigera*, *J. Insect Physiol.* **50**, 25–33.

Noncoding Tandemly Repeated DNA Sequences

Wolfgang Stephan
Department of Biology II, University of Munich, Munich, Germany

1 The C-value Paradox 185

2 Properties of Tandem Arrays 186
2.1 Microsatellite Sequences 186
2.2 Minisatellite Sequences 187
2.3 Satellite Sequences 187

3 Genetic Mechanisms of Copy Number Change 189

4 The Selfish DNA Hypothesis 192

5 Evolutionary Dynamics of Tandem Arrays 193
5.1 The General Model 193
5.2 Copy Number Dynamics of Tandem Arrays 194
5.3 Sequence Homogenization 196

6 Comparison of Theory and Data 196
6.1 Microsatellites 197
6.2 Minisatellites 197
6.3 Satellites 199

7 Population Processes 200

Bibliography 201
Books and Reviews 201
Primary Literature 201

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

Keywords

Copy Number Polymorphism
A hallmark of noncoding, tandemly repeated DNA sequences referring to the observation that the copy number of an array is characteristically variable within a population or species.

Mechanisms of Copy Number Variation
DNA turnover mechanisms such as gene amplification, deletion, and unequal crossing over, and evolutionary processes such as genetic drift and natural selection on array size.

Noncoding, Tandemly Repeated DNA
Localized arrays of nucleotide repeats found in eukaryote genomes and usually classified as microsatellites (small tandem arrays of very short repeat units), minisatellites (larger arrays of longer repeat units), and satellite DNA (very large arrays of typically long, complex repeat units).

Selfish DNA
DNA sequences such as tandemly repeated sequences that may be responsible for significant fitness losses in the organism but are maintained by their ability to replicate quickly within the genome.

Noncoding, tandemly repeated dna sequences form a substantial fraction of the genomes of eukaryotes. In bacteria, however, large tandem arrays are not present. Some of the variation in genome size between different eukaryotic species, which bears little relation to differences in organismal complexity or the numbers of protein-coding genes (the *c-value paradox*), appears to be due to these nongenic sequences. This class of DNA includes satellite DNA (very highly repetitive sequences), minisatellite DNA (moderately repetitive sequences), and microsatellite DNA (short tandem arrays). In many cases, tandem arrays of noncoding DNA sequences seem to be maintained solely by their ability to replicate quickly within the genome (the *selfish DNA* hypothesis). Their behavior can result in mutations that cause human genetic disorders (e.g. fragile X syndrome). Features of the organization of tandemly repeated DNA sequences in eukaryotic genomes reflect the genetic mechanisms and evolutionary pressures acting on selfish DNA. New data from the human genome project provide a first glimpse on centromeric satellite DNA, a hitherto unexplored bastion of our genomes, and allow us to test previously proposed models of the evolution of highly repetitive sequences.

1
The C-value Paradox

The amount of DNA in an unreplicated haploid genome, such as that of the sperm nucleus, is called the *genome size* or, alternatively, *C-value*, because it is usually constant in any one species. In contrast to this relative constancy of genome size within species, great variation in C-value has been found among eukaryotic species. C-values in eukaryotes range from 2.9×10^6 base pairs (bp) to 6.9×10^{11} bp. The most dramatic differences are observed among unicellular organisms. For example, between *Saccharomyces cerevisiae* (yeast) and *Amoeba dubia*, variation in genome size is roughly 80 000-fold. Within a given taxonomic group, unicellular protists show the greatest range in genome size (20 000-fold). In contrast, the three amniote classes (mammals, birds, and reptiles) vary much less (1.3- to 4-fold). Most extensively studied eukaryote groups, whether plant, animal, or microorganism, show much greater variation in C-value than the amniotes. Protozoa exhibit a 1000-fold and unicellular algae, a 3000-fold range, whereas most classes of multicellular animals and plants show variation of the order of 10- to 100-fold. Great variation in C-value is, therefore, the general rule in most eukaryote groups; the relative uniformity found in mammals, birds, and reptiles appears to be an exception.

Given the relative constancy of genome size within species, it is puzzling that eukaryotes exhibit huge interspecific differences in C-value that bear little relation to differences in organismal complexity, ploidy level, or the percentage of coding DNA. For example, several unicellular protozoans possess more DNA than mammals; and the newt *Triturus cristatus* has around six times as much DNA as humans. This lack of correspondence between genome size and genetic information content has become known in the literature as the *C-value paradox*. The C-value paradox is particularly apparent in comparisons of closely related species, which, by definition, show similar levels of organismal complexity. Among protists, bony fishes, amphibia, and flowering plants, many sibling species differ greatly in their C-values. For example, the ciliate *Paramecium caudatum* has about 45 times as much DNA as *P. aurelia*.

Since a species does not contain less DNA than the amount required for specifying its vital functions in inheritance, the observed variations in genome size must be due to DNA that is in excess to this requirement. Eukaryotes with well-characterized genomes show a 30-fold variation in the number of protein-coding genes, ranging from 2000 in the parasitic microsporidian *Encephalitozoan cuniculi* to approximately 60 000 in *Oryza sativa*. Although this value is probably too low, the variation in the number of protein-coding genes observed so far is much too small to account for the 80 000-fold differences in nuclear genome size. Likewise, the C-value paradox cannot be explained by the interspecific variation in gene length because the several-fold variation in the average size of protein-coding genes (largely caused by variation in intron length) is also far too small to account for the C-value differences between species. On the other hand, the (repetition) number of RNA-specifying genes (e.g. ribosomal RNA genes) and the haploid DNA content are positively correlated; and a positive correlation also exists between the number of regulatory sequences (e.g. for replication) and C-value. Since, however, RNA-specifying genes and regulatory elements constitute

only a minute fraction of the eukaryotic genome, they cannot explain the great variation in genome size. On the basis of these observations, it has been concluded that the interspecific differences in C-values are largely due to nongenic DNA. It has been estimated that the amount of nongenic DNA per haploid genome varies in eukaryotes from less than 1.0×10^6 bp to more than 1.0×10^{11} bp. Nongenic DNA per genome makes up anything from less than 30% to almost 100% of the genome.

Classical studies of the kinetics of DNA reassociation showed that the eukaryote genome can be divided roughly into four fractions: foldback DNA (i.e. palindromic DNA sequences), highly repetitive DNA, middle-repetitive DNA, and single-copy DNA. The proportion of the genome consisting of repetitive sequences varies widely between taxa. In yeast, the amount of repetitive DNA is small (20%); in mammals, up to 60% of the DNA is repetitive. In plants, the amount of repetitive DNA can exceed 80%, and higher values than that have been registered. A large fraction of the repetitive DNA is noncoding and organized in tandem arrays. Thus, at least some of the variation in C-value is due to these noncoding, tandemly repeated DNA sequences (transposable elements are another major factor). The C-value differences between species are therefore caused by a *combination* of genetic and evolutionary forces acting on these repetitive sequences. Other authors have suggested that a single major mechanism (in particular, deletion bias leading to DNA loss) can account for the species-specific differences in genome size. However, although the data are scarce at present, it is unlikely that any such a mechanism alone can explain the C-value paradox.

2 Properties of Tandem Arrays

Variation in the numbers and lengths of tandemly repeated units occurs on many different scales. According to these characteristics, three major classes of tandemly repeated, noncoding DNA sequences can be distinguished: micro-, mini-, and satellite DNAs.

2.1 Microsatellite Sequences

Arrays of short (2–6 bp) nucleotide repeats, found in every organism analyzed so far, are called *microsatellite loci*; at least 30 000 are present in the human genome, located in the euchromatin. Copy numbers are characteristically variable within a population, typically with mean array sizes on the order of 100 but with multiple array size classes distributed around the mean. The distribution of microsatellites across the genomes of extensively studied organisms (e.g. man, mouse, and rat) appears to be relatively uniform. However, a significant deficit of $(CA)_n$ arrays (i.e. arrays in which the dinucleotide is repeated n times) has been found on the X chromosomes of these species. Contrasting evidence has been reported from several *Drosophila* species. More $(CA)_n$ loci (irrespective of their degree of polymorphism) have been detected on the X chromosome of *D. melanogaster* than on the autosomes. All autosomes show approximately equal numbers except chromosome 4, which has no detectable stretches of $(CA)_n$. Another striking feature is the lack of such repeats in the heterochromatic regions around the centromere. The same overall chromosomal pattern was seen by *in situ* hybridization in other *Drosophila* species, including *D. simulans*, *D. hydei*,

and *D. virilis*. The abundance of different repeat types varies between species. For example, the dinucleotide repeats $(CA)_n$ are very frequent in mammals, but rare in plants; and the trinucleotide repeat array $(CTG)_n$ is overrepresented in *Drosophila*, but is less frequent in other species. In general, all types of microsatellite sequences (from di- to hexanucleotide repeats) are found in frequencies higher than predicted on the grounds of base composition in noncoding genomic regions across seven eukaryotic clades: *Saccharomyces cerevisiae, Caenorhabditis elegans, Schizosaccharomyces pombe, Mus, Drosophila*, plants and primates.

The discovery of the causative role of the expansion of microsatellites in several important human diseases (e.g. the fragile X syndrome of mental retardation) has revealed an unexpected practical significance to the studies of tandemly repeated noncoding DNA. The mutations causing neurological diseases such as the fragile X syndrome and myotonic dystrophy are preceded by *premutations* that lead to arrays of large size within the normal range of variation in the general population. The normal mechanisms operating on microsatellites (such as replication slippage; see Section 3) may play a role in the generation of premutation alleles by increasing array sizes by a few repeats at a time. At the onset of the full disease stage (when the length of the permutation alleles exceeds a certain threshold value), however, sudden large copy number changes occur.

2.2 Minisatellite Sequences

Minisatellite sequences are tandem arrays of longer (>15 bp) repeats, generally involving mean array lengths of 0.5 to 30 kilobase pairs (kb). They are found in euchromatic regions of the genome of vertebrates, fungi, and plants, and many minisatellites are also highly variable in array sizes. In contrast to hypervariable microsatellites, which are relatively uniformly dispersed throughout mammalian genomes (with the exception of X chromosome–autosome differences), the distribution of minisatellite loci across the genomes of some extensively studied species appears to be uneven. Hypervariable minisatellites are usually located in chromosomal regions of high recombination rates. In humans, most hypervariable loci have been localized to subtelomeric regions that have been found recombination proficient (particularly on metacentric chromosomes). In mice, the distribution of variable minisatellite loci seems to be more uniform, but the hypervariable minisatellites tend to be also located in recombination prone regions.

2.3 Satellite Sequences

Repeat units for satellites can be similar in length (2–30 bp) to micro- and minisatellites, or much larger (>100 bp). They are typically organized as large (up to 10^8 bp) clusters in the heterochromatic regions of chromosomes, near centromeres and telomeres, or on the Y chromosome. In plants, blocks of satellite sequences are also found interstitially (i.e. within euchromatic regions). Pulsed-field gel electrophoresis has shown that satellites are apparently not as variable in array size within populations as micro- and minisatellites (on a per-nucleotide basis). However, substantial differences in array size have frequently been observed between isolated populations and closely related species of both plants and animals. Satellite DNA families that are overabundant

in one species may be absent or in very low amounts in a sibling species. Species-specific satellite DNA clusters have been found in many organisms.

Another characteristic difference between satellite and minisatellite DNAs is the greater tendency for multimeric repeats to be found in satellites. The formation of higher-order repeats is particularly often seen in larger arrays (>200 kb). The best-studied example is the alpha satellite DNA family in primates. This consists of a basic repeat unit of 171 bp that is combined into various higher-order structures, ranging from dimers to (at least) 16-mers. The very highly repetitive satellite sequences can account for a large proportion of the DNA in the genome. Copy numbers of satellite DNA clusters range between 10^3 and 10^9. Arrays with shorter repeat units tend to have higher copy numbers. On average, the proportion of satellite DNA in the eukaryote genome is around 10%. Satellite DNA is virtually absent in yeast. But other species (including mammals) can possess more than 50% satellite DNA. For example, in the kangaroo rat *Dipodomys ordii*, 50 to 60% of the genome consists of three repeated sequences: AAG (repeated 2.4×10^9 times), TTAGGG (2.2×10^9 times), and ACACAGCGGG (1.2×10^9 times). The repeats within a tandem array are not completely identical. Interrepeat variability typically amounts to 5 to 10%. Another distinctive feature is related to their propensity to form higher-order structures: satellite repeats are often composed of subunits of simple sequences such as homopolymeric runs or short runs of di- or trinucleotide repeats. For example, the third satellite of the kangaroo rat contains both a homopolymeric run and a short stretch of dinucleotide repeats. Because of this simple nucleotide composition, satellite DNAs usually form one or more (*satellite*) bands in buoyant density gradients of genomic DNA, which are clearly distinguishable from the main band. This property led to the discovery of satellite DNA and to the name of this class of DNA. Satellite DNAs that cannot be separated from the main band by density gradients are known as *cryptic*.

The inability of current technology to determine a contiguous sequence for highly repetitive tandem arrays means that we know little about the sequence composition (such as the order of subfamilies) of satellite DNA clusters. However, high-resolution mapping of a human X-centromere border and sequencing of selected fragments recently provided a first glimpse into a large contiguous region of very highly repetitive DNA sequences of about 450 kb. This region between expressed sequences on the short arm of the X and the chromosome-specific alpha satellite array DXZ1 is satellite-rich. Repeat units of the DXZ1 array that forms the centromeric core are nearly homogeneous (98–99%), but homogeneity drops to around 70% toward the junction to the satellite-rich region (over a stretch of about 20 kb).

Micro-, mini-, and satellite DNAs comprise only a fraction of repeated, localized DNA sequences in eukaryotes. Interspersed DNA sequences, such as transposable elements, may also form clusters of repetitive DNA in certain parts of the genome after they have lost their ability to move. For instance, these nonfunctional transposable elements may be found as scrambled, clustered arrays in regions around the centromere at the boundary between heterochromatin and euchromatin. The individual sequence elements of these numerous arrays of moderately repeated sequences are usually <1000 bp in length. A cluster, which is several kilobase pairs

long, usually contains many different types of sequence, some of which may be repeated either directly or in inverted orientation. Different clusters within the same genome may share some, though not all, sequences, but the arrangement of shared sequences is not conserved between clusters. In this way, genetic and evolutionary mechanisms have created a considerable diversity of different types of repetitive DNA clusters that go beyond the three main tandem array classes discussed here.

3 Genetic Mechanisms of Copy Number Change

Several genetic mechanisms can affect the dynamics of tandem arrays: gene amplification by processes such as replication slippage and rolling circle amplification, unequal exchange, intrastrand exchange, and mutations by base substitutions (Figs. 1–3). However, the relative importance of these DNA turnover processes for micro-, mini-, and satellite DNAs, and the values of the parameters that describe them, are still uncertain.

In vitro studies suggest that strand slippage during DNA replication is the major cause of the observed length polymorphism of microsatellites between individuals within populations (Fig. 1). This process results in length changes of a few bases at a time. Dinucleotide repeats show a higher *in vitro* slippage rate than trinucleotide motifs; and those trinucleotides with the highest GC content have the lowest rates. In general, the *in vitro* data suggest that microsatellite arrays should be very highly variable in natural populations. The frequencies with which changes in array sizes occur at microsatellite loci are as high as a few percent per generation. These rates are much higher than normal mutation rates (due to base substitutions), but lower than expected from the *in vitro* results. Furthermore, there appears to be no relation between the occurrence of certain repeat types in eukaryotic genomes and their *in vitro* slippage properties. For example, $(GCC)_n$ repeats occur frequently in genomic DNA, but their *in vitro* slippage propensity is the poorest among all trinucleotides. These observations suggest that mechanisms other than slippage

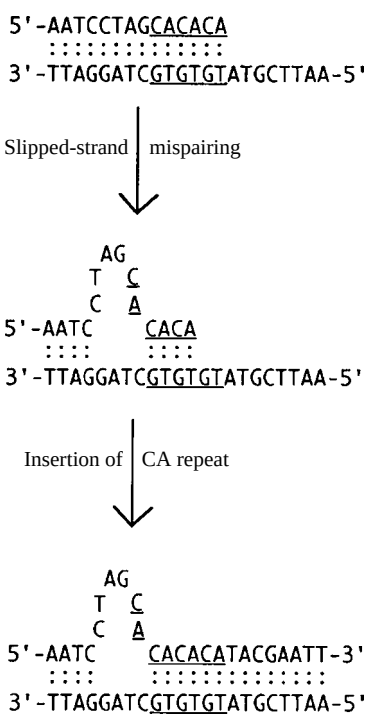

Fig. 1 Replication slippage: the newly synthesized strand (upper strand) is slipping in the 3′ → 5′ direction in the tandem array $(CA)_3$ (underlined). This leads to the insertion of an additional CA repeat as replication proceeds. Similarly, if the template strand (lower strand) slips in the 5′ → 3′ direction, a deletion of one or a few repeat units may occur.

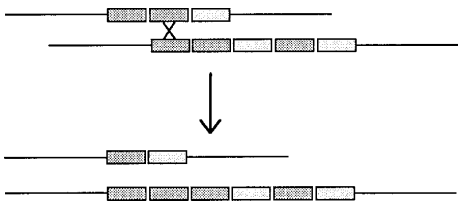

Fig. 2 Unequal exchange: unequal crossing over occurring between tandem arrays on homologous chromosomes. The boxes represent repeat units; heavy and light stippling indicates that the members of a tandem array may differ. Recombination is most likely to occur between repeats with the highest degree of similarity. Unequal exchange leads to the expansion of one array (lower one) and to the contraction of the other one. Furthermore, it results in more homogeneous arrays, a process called *concerted evolution*.

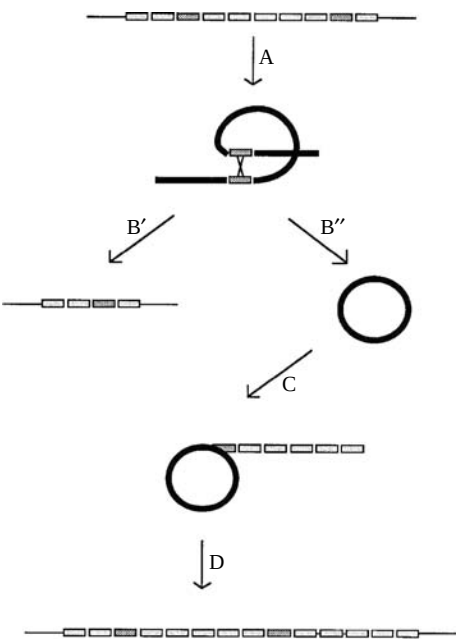

Fig. 3 Intrastrand deletion and rolling circle replication: pathway (A, B′), which includes the steps A and B′, results in a deletion of repeat units, whereas pathway (A, B″, C, D) leads to an amplification of repeats. In step A, pairing occurs between the two heavily stippled copies of the tandem array shown at the top (the other repeats are not shown). In step B′, this leads to the deletion of the repeats that are located between the two pairing repeat units (except one). In step B″, an extrachromosomal circular plasmid is created that carries the deleted repeats. Most of these plasmids are lost. However, some of them may undergo *rolling circle* replication. Step C results in the replication of all copies of this circle. In step D, these amplified repeats are reintegrated into the tandem array by recombination.

must contribute to the dynamics of simple repeats in organisms.

There is a relation between replication slippage of microsatellite sequences and defective DNA repair, since mutations in DNA repair genes can cause tract instability in yeast when errors resulting from slippage replication remain unrepaired. Similarly, a mutation in a gene of the same mismatch repair system in humans can lead to expansion of microsatellite tracts associated with human disorders such as colorectal cancers. These observations may suggest that replication slippage is the primary force involved in the expansion and contraction of microsatellite arrays, but that most of the length changes because of slippage are corrected by the DNA repair system. In addition, unequal exchanges could occur in microsatellite arrays that

are sufficiently large (Fig. 2). Total array size, however, appears to be an important constraint for the occurrence of unequal exchange. In contrast to replication slippage, unequal exchange is an interhelical event, involving DNA molecules from two different sister chromatids or homologs. This may place some constraints on the length of the sequences undergoing unequal exchange. Large changes in copy number, which are occasionally observed at microsatellite loci in population surveys (see Sect. 7), may result from unequal exchanges.

Several processes may cause the instability of arrays of minisatellites. At some well-studied human minisatellite loci, a bias toward gains of a few repeats has been found, with gains occurring preferentially at one end of a tandem array. New repeats on a given chromosome can originate either from the same chromosome or from its homologue. A gene conversion-like mechanism causing this net increase in copy number has been proposed. In addition, unequal exchange could occur between tandem arrays on sister chromatids or between arrays on homologous chromosomes. At present, however, there is no direct evidence for unequal exchanges at minisatellite loci. The frequencies with which changes in array sizes occur at minisatellite loci are similar to those of microsatellites and can be as high as a few percent per generation. At extremely variable loci, mutations to new length alleles occur at a rate high enough to measure it by direct observation in pedigree analysis. The apparent relative stability in populations of small minisatellite alleles provides indirect evidence that the expansion and/or contraction of minisatellite arrays may occur at a frequency dependent on array size, such that longer arrays exhibit a higher rate of copy number change.

The forces governing the dynamics of satellite DNA are more difficult to understand, mainly because the large sizes of satellite DNA clusters preclude direct experimental analysis, and also hinder the collection of meaningful population data. However, new advanced mapping techniques, in combination with genomic sequencing, have recently revealed some of the complexity of satellite DNA arrays and also advanced our knowledge about the mechanisms acting on these sequences. Since replication slippage involves only a short stretch of DNA at a time, it probably plays only a minor role in the amplification of the longer satellite repeat units (>100 bp). Slippage may, however, be important in the initial formation of satellite repeats that are generally composed of smaller subunits (see Sect. 2.3). The observation of extrachromosomal circular satellite DNA suggests the possibility that long satellite repeats may primarily be amplified by extrachromosomal rolling circle replication, followed by reinsertion into the genome (Fig. 3). The best evidence for the occurrence of unequal exchange comes from the analysis of the human DXZ1 alpha satellite. A large stretch of the DXZ1 satellite toward the middle of the array consists of homogeneous repeat units, but sequence similarity among repeats drops relatively quickly toward the edge of the array. Such a pattern can only be explained by unequal crossing over, not by any other mechanism of sequence homogenization such as gene conversion (see Sect. 6.3). The question of whether unequal exchange occurs predominantly between sister chromatids or between homologues is, however, still open. If recombination occurs either between arrays on sister strands or between different homologues, it is also likely

to occur between members within a strand. This process, called *intrastrand exchange* or *intrastrand deletion*, is expected to result in deletion of members on that array (Fig. 3). In fact, the finding of extrachromosomal DNA circles containing satellite repeats is the best available evidence that intrastrand exchange occurs in large tandem arrays. Most of the extrachromosomal plasmids carrying copies of the array may be lost. However, some of these plasmids may undergo rolling circle replication, which amplifies a small section of the array before integration into the array.

4
The Selfish DNA Hypothesis

It has often been proposed that tandem-repetitive DNA sequences are functionally important for the host organism. The *Responder* satellite sequences of the *D. melanogaster* segregation distorter system seem to affect the fitness of their carriers. The *Responder* locus in natural populations of *D. melanogaster* consists of 20 to 2500 copies of a 120-bp repeat unit. In a competition experiment involving a mixed population of flies with 700 copies and flies with 20 copies, it was observed that the frequency of the flies containing 20 copies decreases with time. This may suggest that flies with 700 copies have a higher fitness than flies with 20 copies. However, there is a potential problem with this experiment in that the chromosomes carrying 20 copies contain a large deletion that includes a segment of DNA flanking the *Responder* satellite.

There is also evidence for biological effects of heterochromatin (which is rich in satellite sequences), such as the formation of heterochromatic genomic compartments important for proper chromosomal segregation in mitosis and meiosis. Satellite DNAs appear to be major constituents of functional centromeres, as has been shown in *Drosophila* and humans. Similarly, there appears to be an effect of minisatellites on the expression of genes that are located nearby. For example, the 14-bp minisatellite in the diabetes susceptibility locus *IDDM2* contains high affinity–binding sites for the transcription factor, Pur-1 and may stimulate insulin transcription. It has therefore been suggested that this result demonstrates a possible function of the insulin minisatellite.

But these properties of minisatellite and centromeric satellite DNA may not have a role in the origin and maintenance of tandemly repeated sequences: they could simply be a response to their presence in the genome, that is, consequences rather than causes. That the insulin repeat is found in much lower copy numbers in other primates (e.g. chimpanzees) and is absent in nonprimates seems to support this hypothesis, although it does not completely rule out the possibility that this repetitive structure is the result of a recent adaptive process. The large differences in amounts of satellite DNA sequences between closely related species (see Sect. 2.3), and the absence of measurable fitness effects of large deletions or duplications of heterochromatin in *Drosophila*, also suggest strongly that selection does not actively maintain the high copy numbers of many of these sequences.

It has therefore been proposed that in most cases, tandemly repeated noncoding DNA sequences are maintained solely by their ability to replicate quickly within the genome (the *selfish DNA* hypothesis). Far from conferring benefits, their behavior can result in a fitness loss to the host. Some human genetic diseases are known to be caused in this way, including mutations

due to the amplification of microsatellite sequences (see Sect. 2.1).

5 Evolutionary Dynamics of Tandem Arrays

Because of an obvious lack of functional significance and technical constraints, many aspects of the biology of tandemly repeated, noncoding DNA sequences cannot be explored by direct experimentation. In this section, some basic theoretical concepts that have been used to understand the properties of this class of DNA are introduced. It is shown that many features of the organization of tandemly repeated DNA sequences in eukaryotic genomes may be explained on the basis of the selfish DNA hypothesis (Sect. 4) and the genetic mechanisms outlined in Sect. 3.

5.1 The General Model

The state of a haploid genome with respect to a given class of tandemly repeated DNA sequences at a particular chromosomal location is characterized by i, the number of repeats in the array. The state of the population is described by the distribution of values of i among haploid sets, such that x_i is the frequency of chromosomes with array size i. The following factors affect the change of this distribution over time. Experimental evidence for these mechanisms is discussed in Sect. 3.

Amplification: A sudden increase in array size by one or more repeat units. Depending on array type (see Sect. 3), it may involve replication slippage (Fig. 1), gene conversion-like processes, rolling circle replication of extrachromosomal repeat units followed by reinsertion (Fig. 3), or other, unknown, mechanisms (e.g. in disease-related microsatellites).

Deletion: A sudden decrease in array length by one or more units, caused by replication slippage or intrastrand exchange (Fig. 3).

These two processes can be included in a model in which the probability that an array of length j generates an array of length i is $\alpha_j P_{ij}$, where α_j is the probability that a change of an array of length j occurs per generation, and P_{ij} is the conditional probability of the change from j to i. Although there is evidence that array size affects the rate of copy number change, assume here for simplicity that all α_j are identical.

Unequal exchange: This can occur either between homologous chromosomes or between sister chromatids. Only the first category is considered here (Fig. 2). Let γ_{jk} be the probability of an exchange between an array of j and an array of k copies, and Q_{ijk} be the conditional probability that an exchange produces a daughter chromosome with array size i from parental chromosomes with copy numbers j and k ($0 < i < j + k$). There is an equal chance of production of a daughter with $j + k - i$ copies. The probability of the production of a daughter chromosome with i (or $j + k - i$) copies is then $\gamma_{jk} Q_{ijk}$. Again, for simplicity, it is assumed that all γ_{jk} are identical ($= \gamma$).

Natural selection: According to the selfish DNA hypothesis, selection is usually assumed to be nonexistent or very weak (i.e. fitness declines very slightly) until array sizes reach a relatively high value. Above a certain threshold value Ω, however, fitness is assumed to drop rapidly to zero with increasing numbers of repeats for individuals carrying more than Ω copies. The

threshold effects observed in the dynamics of disease-related microsatellite loci (see Sect. 2.1) seem to suggest that such abrupt changes in fitness may occur. Base substitutions in the sequence are generally considered to be selectively neutral.

Genetic drift: As a result of finite effective population size N_e, the frequencies of haploid genomes with different array sizes are subject to multinomial sampling during reproduction in the same way as a conventional multiple-allele system with allele frequencies x_i.

Mutation: Mutational differences in sequence (occurring at rate μ) among members of an array may affect their probability of undergoing strand pairing during recombinational processes. Recombinational steps are believed to be important in several DNA turnover processes that operate on tandem arrays (e.g. unequal exchange, rolling circle replication, gene conversion-like amplification). Mutational differences among the repeat units also have a strong impact on the rate of strand slippage during replication because these differences, which appear as mismatches in strands that are reannealed out of register, can be more easily detected by the DNA repair system when they are more frequent.

5.2 Copy Number Dynamics of Tandem Arrays

The qualitative behavior of this model can best be studied by considering the persistence times of tandem arrays. In this approach, it is assumed that a certain distribution of array sizes exists at a given initial time in the population, generated by a sudden amplification event. In the absence of further amplifications, these tandem arrays are lost from the population owing to the joint action of DNA turnover processes, genetic drift, and selection. A discussion of unequal exchange as a typical DNA turnover process that can affect copy number is now given; but other processes with similar properties can be substituted or added, as will be shown. Natural selection is modeled such that fitness of an individual in the population decreases slowly with increasing copy number and declines rapidly to zero when copy number is larger than a threshold value Ω. This selection scheme is consistent with the selfish DNA hypothesis. Unequal exchange can both expand and contract a tandem array with equal probability and by equal amounts; hence, on average, mean copy number does not change from generation to generation. However, unequal exchange, in conjunction with genetic drift and selection, leads to a decrease in mean copy number over time until the population has reached a state in which it is fixed for chromosomes with array size 1, and unequal exchanges are no longer possible. The average time to loss of a tandem array is inversely proportional to γ the frequency of unequal exchanges per generation, if γ is sufficiently small ($4N_e\gamma < 1$); furthermore, it increases with Ω, and decreases with increasing effective population size N_e. However, if $4N_e\gamma$ exceeds 1, the expected time to loss of a tandem array from the population increases with increasing population size. This behavior of the model is summarized in Fig. 4.

Four predictions that can be compared with the data (see Sect. 6) follow immediately from these results:

1. Longer tandem arrays persist in regions of low rates of unequal exchanges.
2. More variable arrays are less conserved over evolutionary time.

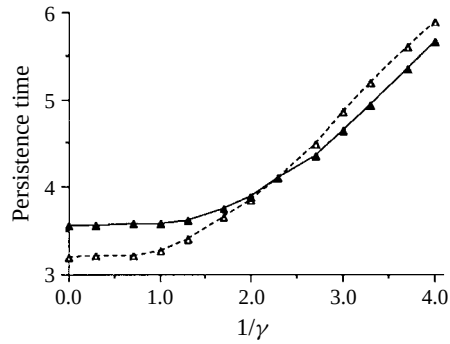

Fig. 4 Persistence time of tandem arrays versus inverse recombination rate. The mean time to loss of tandemly repeated DNA sequences (in generations) under the joint action of unequal exchange, genetic drift, and selection on array size is plotted against the inverse recombination rate in a double-logarithmic fashion. It is assumed that amplification processes are not operating and that the mean copy number of the population is initially 50. The results have been generated by computer simulations for two population sizes, $N_e = 50$ (open triangles) and $N_e = 200$ (solid triangles), and a maximum array size of $\Omega = 10\,000$. Note the dramatic increase in persistence time as the recombination rate decreases. Furthermore, note that the effect of N_e on persistence time is reversed for small and large values of γ. Because of the logarithmic scale of the y axis, the persistence times between the two populations differ for both small and large values of roughly by a factor 2. That means effective population size, in addition to recombination rate, can have a dramatic effect on the persistence of tandem arrays in populations. In contrast, the effects of initial copy number and selection (Ω) are less pronounced.

3. Tandem arrays are larger and persist longer when selective constraints on array size are weak (i.e. larger Ω).
4. Tandem arrays are larger and persist longer in populations whose effective size N_e is small.

Predictions 1 and 4 may not be intuitively obvious and need additional explanation. In regions of low rates of crossing over, arrays with high copy numbers are not frequently generated by unequal exchange; thus, selection is less effective in eliminating from the population individuals whose copy numbers are too high. As a result, mean array size decreases slowly, which explains prediction 1. Indeed, a detailed analysis of the model indicates that the effect of unequal exchange on the persistence of arrays is particularly pronounced in regions of very low rates of crossing over, suggesting that array sizes of tandemly repeated sequences should be much larger in regions of reduced rates of exchange than in those of intermediate or high rates of crossing over. Prediction 4 is a consequence of the interaction of genetic drift and selection on array length. In populations with small effective sizes, random fluctuations of the frequencies of alleles containing tandem arrays due to genetic drift are more dramatic than in populations with large effective sizes. Therefore, in small populations, drift is stronger and can overcome the effects of selection pressure on array size. This reduces the efficiency with which alleles whose copy numbers are too high are removed from the population by selection, and increases the persistence times of tandem arrays.

The foregoing results are not limited to unequal exchange, which is thought to act primarily on larger tandem arrays such as minisatellites and satellite DNAs. Any other mechanism that is symmetrical with respect to copy number change can be

substituted for unequal exchange such that the results are essentially unchanged. For example, replication slippage (operating on microsatellites) and unequal exchange would have qualitatively similar effects on copy number dynamics.

When recurrent amplification processes are included in the model, an equilibrium array size will result as a balance between amplification (leading to a net increase in copy number) and unequal exchange, genetic drift, and selection (causing a decrease). A particularly interesting prediction of this model that may pertain to satellite DNA is the following: as a result of amplification in combination with selection against high copy numbers, unequal exchange, and genetic drift, mean array size is high only when $4N_e\gamma$ is less than 1 and is highest when selection against arrays is weakest, as might be the case for organisms with long developmental times. Thus, satellite sequences should accumulate in chromosomal regions of very low rates of unequal exchanges, and more slowly developing species should possess more repetitive DNA.

5.3 Sequence Homogenization

Computer simulations of the general model for tandem arrays have explored the relation between copy number change and sequence variability between members of a tandem array. In these models, mutation was allowed to alter the sequence of repeat copies by base substitution. This has consequences for the probability that copies can recombine in unequal exchanges, and strands reanneal during replication slippage. In other words, the exchange rate and amplification rate were assumed to be dependent on sequence similarity. As suggested by experimental evidence from several different *in vivo* and *in vitro* systems, it was assumed that a short stretch of perfect sequence identity rather than long segments of overall homology is required for a successful turnover process to occur. The outcomes of these simulations are remarkable:

1. Tandem arrays of repetitive sequences can be generated *de novo*, starting from a random DNA sequence.
2. Short simple repeat sequences emerged when the sequence-dependent DNA turnover processes occurred frequently; in contrast, arrays with longer repeat units (including higher-order structures such as multimeric repeats) were generated when the rates of amplification and exchange were low. A related phenomenon is that the basic repeat units often exhibited an internal substructure featuring runs of homopolymeric nucleotides and/or short stretches of di- or trinucleotide repeats.
3. A positive correlation between the rate of unequal exchange γ and sequence homogeneity between repeat units within a tandem array was observed. Remarkably, however, sequence homogeneity remained relatively high and did not drop below a certain threshold level until γ was reduced dramatically relative to the mutation rate μ.
4. Weak selective constraints (i.e. large Ω values) led on average to larger repeat units and increased the tendency to form higher-order structures.

6 Comparison of Theory and Data

This section compares the theoretical results from Sect. 5 with the properties of tandem arrays.

6.1 Microsatellites

Microsatellites have very short and homogeneous repeats and a very small total array size. They are distributed in euchromatic regions of the genome in a relatively uniform manner and have been found near and within gene regions. The theoretical and experimental analyses suggest that these properties are consistent with high rates of DNA turnover mechanisms (in particular, replication slippage) and strong selective constraints on total array size (small values of Ω). As a consequence of these high turnover rates, most microsatellite arrays are highly variable within populations. Furthermore, as expected, there appears to be an inverse correlation between the degree of intraspecific variability of copy number and interspecific conservation of microsatellite loci, such that the more variable loci exist only over a narrower phylogenetic range. However, some simple repeat loci do not follow the predictions of the proposed model in that they show fewer interspecific differences in copy number than expected on the basis of the degree of intraspecific variation. In these cases, functional explanations may have to be invoked to describe their evolutionary dynamics.

Conflicting evidence has been reported from a number of well-studied species with respect to the abundance of microsatellite loci on X chromosomes versus autosomes (see Sect. 2.1). There is a significant lack of highly polymorphic $(CA)_n$ loci on the X chromosomes in several species such as man, mouse, and rat; on the other hand, in *Drosophila*, the total number of $(CA)_n$ microsatellites (poly- and monomorphic ones) on the X chromosome is slightly higher. This uneven distribution in *Drosophila* is consistent with the theoretical model discussed earlier. Since the rate of copy number change for *Drosophila* microsatellites is generally very low and the effective population size of the X chromosome in *Drosophila* is smaller than that of the autosomes, one should expect that the forces for removal of repetitive DNA are weaker on the X chromosome and that therefore more loci persist. A smaller effective population size increases the effect of genetic drift relative to selection and thus alleviates the constraints imposed by selection on total array size. As a consequence, the removal of repetitive DNA is less efficient. Thus, it is not necessary to invoke hypotheses about differences in the genetic mechanisms between the X chromosome and the autosomes or functional explanations to understand the uneven distribution of the number of microsatellite loci. In contrast, the deficit of highly polymorphic microsatellites on the X chromosome in the three mammalian species is probably not compatible with this model. For a population undergoing replication slippage (by single steps) and genetic drift, it has been shown that the persistence time of microsatellites does not depend on effective population size in a linear fashion. Thus, the 30% deficit of hypervariable microsatellites is better explained by a lower rate of replication slippage, owing to the fact that the X chromosome is present less often than the autosomes in the male germ line, which is thought to be more mutation prone in mammals.

6.2 Minisatellites

The salient properties of minisatellites may also be explained by the proposed model. Minisatellite loci are found in the euchromatin. Many of them are

highly polymorphic, though not, on a per-nucleotide basis, as polymorphic as microsatellites. Their repeat units are longer, and their total array sizes are much larger than those of microsatellites. These properties are consistent with the predictions of the model if smaller rates of DNA turnover are assumed. Indeed, the per-base-pair turnover rates that have been estimated by direct observation in pedigree analysis are substantially lower than those of microsatellites. The relatively large array sizes of minisatellites indicate that selective constraints on total array size are weaker than for microsatellites, possibly because minisatellites do not appear to be located as close to gene regions as microsatellites.

It is unknown at present why highly polymorphic minisatellite loci tend to cluster in subtelomeric regions of the human genome. However, since the genetic map is extended in these portions of chromosomes, recombinational steps may generally occur at a higher rate in these regions. Recombinational steps are involved in processes such as gene conversion-like amplification, unequal exchange, and intrastrand exchange, which are thought to operate on minisatellites (see Sect. 2.). According to the proposed model, increased rates of DNA turnover should lead to a high degree of copy number variation and may therefore explain the overabundance of hypervariable minisatellites in these regions.

Because of a lack of exchange of flanking markers, unequal exchange between homologues is probably not the predominant mechanism of copy number change at minisatellite loci. However, there is indirect evidence for the occurrence of unequal exchange between sister chromatids. In accordance with the predictions of the model, it has been found for a sample of human minisatellites with different degrees of allelic variability in repeat copy number that a significantly negative correlation exists between copy number variability and interrepeat sequence variability (see Fig. 5). In other words, the most highly polymorphic loci (with respect to copy number variation) show the highest degree of sequence homogeneity between repeat units. This homogeneity in tandem arrays is difficult to explain by events that are restricted to certain parts of arrays, such as the gene conversion-like amplification mechanism mentioned in Sect. 3. Instead, long-range forces such as unequal exchange that act over a large distance of an array must be invoked. Furthermore, as expected from the theoretical analysis, minisatellites that are highly polymorphic within species are evolutionarily unstable. For example,

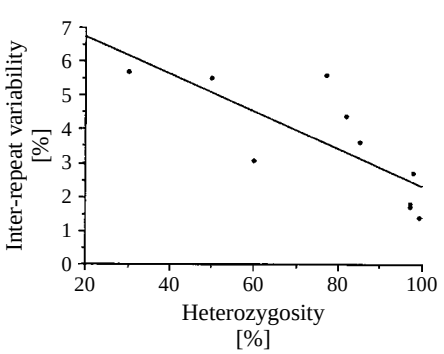

Fig. 5 Interrepeat variability versus allelic heterozygosity for human minisatellite loci: allelic heterozygosity caused by copy number variation versus sequence variation between members of a tandem array. The negative correlation between allelic heterozygosity and sequence heterogeneity is highly significant ($P < 0.01$). [Reproduced from Stephan, W., Cho, S. (1994) Possible role of natural selection in the formation of tandem-repetitive noncoding DNA, *Genetics* **136**, 333 – 341, with permission.]

comparisons between humans, great apes, and monkeys have shown that minisatellite arrays that are highly polymorphic in humans (e.g. minisatellite locus *MS32*) are sometimes very short and relatively monomorphic in great apes and monkeys (3–4 repeats for *MS32*).

6.3 Satellites

The interpretation of the properties of satellite DNA is the greatest challenge for the proposed model because this data set is relatively complex, and despite a long history of research and some new results (see Sect. 2.3), not as comprehensive as those of micro- and minisatellites. Satellite DNA can form short repeat units that are similar in length to micro- and minisatellite repeats and also very large ones (>100 bp). In any case, their total array sizes are much larger than those of micro- and minisatellites. In contrast to the two latter classes of DNAs, satellites are located in heterochromatic regions of chromosomes. These properties of satellite DNA are clearly compatible with weak selective constraints on array size, probably because satellites are located in heterochromatin that is devoid of functional genes (for reasons given shortly). Furthermore, the features of satellite DNAs are compatible with low rates of DNA turnover mechanisms such as unequal exchange and rolling circle amplification that produce copy number variation and sequence homogeneity in these tandem arrays. The short repeat length of some satellites may raise a caveat to this interpretation, but computer simulations of the general model have demonstrated that repeat unit lengths of the simulated arrays may vary greatly when the rate of DNA turnover is low relative to the nucleotide substitution rate μ. This interpretation is also consistent with the absence of microsatellites from heterochromatin (see Sect. 2.1). Tandem arrays of very short repeat units that are located in heterochromatin, often juxtaposed to satellites with longer repeat units (e.g. in humans and *Drosophila*), are much longer than microsatellites (Sect. 2.3).

Direct evidence for low rates of DNA turnover processes in heterochromatic regions near centromeres and telomeres has been found for meiotic recombination. Contraction of the genetic map suggests that meiotic recombination is strongly suppressed in regions near the centromeric and/or telomeric heterochromatin in several well-studied species, including human, *Drosophila*, and yeast. The causes for this reduction are unknown, although it may be selectively advantageous for recombination to be suppressed near centromeres. Since recombinational steps are involved in essentially all DNA turnover mechanisms that alter copy number of satellite arrays, it is possible that there is a general reduction of the rates at which these processes occur in heterochromatin. As a consequence, unequal exchange between homologues as well as between sister chromatids, intrastrand exchange, and rolling circle amplification may be less frequent than in euchromatin. It therefore seems likely that suppression of recombination (or, in general, of recombinational processes) is the primary cause for the accumulation of satellite sequences.

This hypothesis is consistent with the observation that satellite sequences show a greater tendency to higher-order structure (multimeric repeats) than minisatellites (Sect. 2.3). Computer simulations of the general model suggest that higher-order structures are preferentially found in regions of low rates of unequal exchange

and weak selective constraints on total array size. Additional evidence that rates of DNA turnover in satellites are low comes from the observation that satellites, in sharp contrast to micro- and minisatellites, show little variation in copy number within populations, relative to their large array sizes (Sect. 2.3). However, substantial differences in array size are often observed between closely related species of animals and plants, indicating that processes causing expansion or contraction of satellite arrays must occur at a significant rate on an evolutionary timescale. Evidence that a minimal level of DNA turnover occurs in satellite DNAs also comes from the observation of a high degree of sequence homogeneity in the middle of satellite arrays and much lower homogeneity at the edge. The simulations indicate that unequal exchange is a strong long-range homogenization force that may operate over the entire array length. However, rates of exchange cannot be too low relative to the rate of point mutations; otherwise, it would not be possible to maintain the observed degree of sequence homogeneity in satellites, which is as high as 98 to 99% in the core of the DXZ1 satellite of humans (see Sect. 2.3).

In sum, the great majority of tandem arrays of noncoding DNA studied so far follows the predictions of the proposed model that was based on the selfish DNA hypothesis and included a suite of DNA turnover processes. A purely neutral model in which fitness of an individual is not affected by the amount of tandemly repeated DNA would have accounted for most of the properties of micro, mini, and satellite DNA. However, some of the observations, such as the distinctly different chromosomal locations of these classes of DNA in euchromatin versus heterochromatin and the effects of population size on the higher abundance of microsatellites on X chromosomes versus autosomes in *Drosophila*, cannot readily be explained without invoking the idea of selection acting against high copy number. Thus, it appears that tandem arrays are subject to generally weak selective constraints on array size, as predicted by the selfish DNA hypothesis.

7
Population Processes

Noncoding, tandemly repeated DNA provides population biologists with the great opportunity to study processes such as migration and population differentiation, population structure, and genetic relatedness at the molecular level. Most prominent is the use of minisatellite loci as evidence in legal proceedings, usually in the determination of paternity or individual identification in forensic medicine. Population genetic models of tandem arrays generally assume that they are nonfunctional, that is, neutral or slightly deleterious. Base substitutions in the sequence are considered to be selectively neutral, but there may be natural selection against increased array size. In other words, these models are similar to those described in Sect. 5. Alternatively, it is assumed that point mutations prevent microsatellite arrays from growing too long. This assumption is based on the notion that the slippage rate depends on the homogeneity of the repeats.

The frequency distributions of length alleles at microsatellite loci in most populations fit a *stepwise mutation* model quite well. In such a model, arrays increase or decrease in repeat number

by one or a few repeats, with occasional increases by larger steps. The fit of the population data to this model suggests that mutational steps usually involve only very few repeat units, as with slippage replication. This confirms the results of the *in vitro* studies of microsatellites reported in Sect. 3.

In contrast to microsatellites, the length distributions of minisatellite alleles do not fit a stepwise mutation model. This suggests that events that alter copy number occasionally involve many repeat units, as can happen with unequal exchange. On the other hand, the population data are not consistent with a model in which every new array that appears in the population is unique, as would be expected if events such as unequal exchanges with unconstrained misalignment were frequent. For satellite DNAs, only a very limited amount of population data is available at present; therefore, no efforts have been made thus far to analyze the length distributions of satellite alleles within populations.

See also Fragile Sites; Fragile X-linked Mental Retardation; Gene Distribution in the Human Genome; Genetic Analysis of Populations.

Bibliography

Books and Reviews

Charlesworth, B., Sniegowski, P., Stephan, W. (1994) The evolutionary dynamics of repetitive DNA in eukaryotes, *Nature* **371**, 215–220.

John, B., Miklos, G.L.G. (1988) *The Eukaryote Genome in Development and Evolution*, Allen & Unwin, London.

Li, W.-H. (1997) *Molecular Evolution*, Sinauer, Sunderland.

Orgel, L.E., Crick, F.H.C. (1980) Selfish DNA: the ultimate parasite, *Nature* **284**, 604–607.

Schlötterer, C. (2000) Evolutionary dynamics of microsatellite DNA, *Chromosoma* **109**, 365–371.

Ugarkovic, D., Plohl, M. (2002) Variation in satellite DNA profiles – causes and effects, *EMBO J.* **21**, 5955–5959.

Primary Literature

Charlesworth, B., Langley, C.H., Stephan, W. (1986) The evolution of restricted recombination and the accumulation of repeated DNA sequences, *Genetics* **112**, 947–962.

Cummings, C.J., Zoghbi, H.Y. (2000) Trinucleotide repeats: mechanisms and pathophysiology, *Annu. Rev. Genomics Hum. Genet.* **1**, 281–328.

Deka, R., Chakraborty, R., Ferrell, R.E. (1991) A population genetic study of six VNTR loci in three ethnically defined populations, *Genomics* **11**, 83–92.

Di Rienzo, A., Peterson, A.C., Garza, J.C., Valdes, A.M., Slatkin, M., Freimer, N.B. (1994) Mutational processes of simple sequence repeat loci in human populations, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3166–3170.

Dib, C., Faure, S. et al. (1996) A comprehensive map of the human genome based on 5,264 microsatellites, *Nature* **380**, 152–154. (12 additional coauthors)

Glenn, T.C., Stephan, W., Dessauer, H.C., Braun, M.J. (1996) Allelic diversity in alligator microsatellite loci is negatively correlated with GC content of flanking sequences and evolutionary conservation of PCR amplifiability, *Mol. Biol. Evol.* **13**, 1151–1154.

Gregory, T.R. (2003) Is small indel bias a determinant of genome size, *Trends Genet.* **19**, 485–488.

Henikoff, S., Ahmad, K., Malik, H.S. (2001) The centromere paradox: stable inheritance with rapidly evolving DNA, *Science* **293**, 1098–1102.

Jeffreys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neil, D.L., Armour, J.A.L. (1994) Complex gene conversion events in germline mutation at human minisatellites, *Nat. Genet.* **6**, 136–145.

Kennedy, G.C., German, M.S., Rutter, W.J. (1995) The minisatellite in the diabetes susceptibility locus *IDDM2* regulates insulin transcription, *Nat. Genet.* **9**, 293–298.

Kruglyak, S., Durret, R.T., Schug, M., Aquadro, C.F. (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10774–10778.

Li, Y.-C., Korol, A.B., Fahima, T., Beiles, A., Nevo, E. (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review, *Mol. Ecol.* **11**, 2453–2465.

Lynch, M., Conery, J.S. (2003) The origins of genome complexity, *Science* **302**, 1401–1404.

Metzgar, D., Bytof, J., Wills, C. (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA, *Genome Res.* **10**, 72–80.

Okumura, K., Kiyama, R., Oishi, M. (1987) Sequence analyses of extrachromosomal *Sau*3A and related family DNA: analysis of recombination in the excision event, *Nucleic Acids Res.* **39**, 1–30.

Payseur, B.A., Nachman, M.W. (2000) Microsatellite variation and recombination rate in the human genome, *Genetics* **156**, 1285–1298.

Petrov, D.A. (2002) Mutational equilibrium model of genome size evolution, *Theor. Pop. Biol.* **61**, 533–546.

Queller, D.C., Strassmann, J.E., Hughes, C.R. (1993) Microsatellites and kinship, *Trends Ecol. Evol.* **8**, 285–288.

Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K., Willard, H.F. (2001) Genomic and genetic definition of a functional human centromere, *Science* **294**, 109–115.

Smith, G.P. (1976) Evolution of repeated DNA sequences by unequal crossover, *Science* **191**, 528–535.

Stephan, W. (1986) Recombination and the evolution of satellite DNA, *Genet. Res.* **47**, 167–174.

Stephan, W. (1989) Tandem-repetitive noncoding DNA: forms and forces, *Mol. Biol. Evol.* **6**, 198–212.

Stephan, W., Cho, S. (1994) Possible role of natural selection in the formation of tandem-repetitive noncoding DNA, *Genetics* **136**, 333–341.

Stephan, W., Kim, Y. (1998) Persistence of microsatellite arrays in finite populations, *Mol. Biol. Evol.* **15**, 1332–1336.

Thomas, C.A. (1971) The genetic organization of chromosomes, *Annu. Rev. Genet.* **5**, 237–256.

Wu, C.-I., True, J.R., Johnson, N. (1989) Fitness reduction associated with the deletion of a satellite DNA array, *Nature* **341**, 248–251.

Nuclear Transfer for Cloning Animals

Andras Dinnyes[1], *X. Cindy Tian*[2], *Jie Xu*[3], *and Björn Oback*[4]
[1] *Hungarian Academy of Sciences and Szent Istvan University, Godollo, Hungary; Agricultural Biotechnology Center, Godollo, Hungary*
[2] *University of Connecticut, Storrs, CT*
[3] *Evergen Biotechnologies, Inc., CT*
[4] *AgResearch, Hamilton, New Zealand*

1 Introduction and Brief Historical Overview 206

2 Key Elements and Critical Aspects of the Technology 209
2.1 Nuclear Reprogramming and Epigenetics 209
2.1.1 DNA Methylation 209
2.1.2 Histone Acetylation 211
2.1.3 Imprinting and Imprinted Genes in Cloned Animals 211
2.2 Epigenetic Differences between Donor Cells 212
2.2.1 Embryonic Cloning 213
2.2.2 Somatic Cloning 215
2.3 Genetic Differences between Donor Cells 217
2.4 Choice of Nuclear Donor Cell Cycle Stage 223
2.4.1 The Donor Cell Cycle 223
2.4.2 Cell Cycle Coordination to Maintain Normal Ploidy 223
2.4.3 Cell Cycle Coordination to Promote Reprogramming 225
2.5 Selection and Enucleation of Recipient Oocytes 226
2.6 Nuclear Transfer Method 229
2.6.1 Nuclear Injection 229
2.6.2 Fusion 229
2.7 Artificial Activation 230
2.8 *In Vitro* Culture of Cloned Embryos 231
2.9 Transfer of Cloned Embryos 231
2.10 Pregnancy Monitoring and Progeny Production 232
2.11 Adult Clone Phenotypes and Transgenerational Effects 232

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

3 Applications of Nuclear Transfer in Different Species 233
3.1 Amphibians, Fish, and Birds 233
3.2 Laboratory Animals: Mouse, Rat, and Rabbit 234
3.3 Farm Animals: Sheep, Cattle, Goat, Pig, and Horses 235
3.4 Endangered Species and Companion Animals 237
3.5 Humans and Nonhuman Primates 238
3.6 Ethical and Animal Welfare Aspects of Nuclear Transfer Research 239

4 Future Perspectives of Nuclear Transfer 240

5 Concluding Remarks 241

Acknowledgments 241

Bibliography 241
Books and Reviews 241
Primary Literature 242

Keywords

Adult Stem Cells
Undifferentiated multipotent (able to give rise to a subset of cell lineages) cells capable, at the single cell level, of proliferation, self-renewal, and the production of differentiating daughter cells.

Animal Cloning
Production of genetically identical progeny from a founder; different techniques can be used (e.g. embryo disaggregation, embryo splitting, or nuclear transfer).

Cloning Efficiency
Proportion of all cloned embryos transferred into surrogate mothers that develop into viable offspring.

Cytoplast
An oocyte, zygote, or 2-cell blastomere whose nuclear DNA was removed or destroyed.

Embryonic Stem (ES) Cells
Immortal pluripotent (able to give rise to all cell types of the embryo proper) cells derived from placing inner cell mass blastomeres into *in vitro* culture.

Epigenetic
Changes in gene activity that do not entail changes in DNA sequence. These changes are stably maintained through DNA replication and include covalent modifications of DNA and DNA-binding proteins.

Mitochondrial Heteroplasmy
Presence of more than just the maternally inherited types of mitochondria within a cell.

Nuclear Reprogramming
The process of returning a differentiated nucleus to a totipotent stage.

Nuclear Transfer (NT)
Procedure in which the complete chromatin content from one cell (e.g. somatic donor) is moved to another (e.g. recipient oocyte), either by microinjection or membrane fusion.

Reproductive Cloning
Cloning of progeny from diploid cells; in humans, banned in most countries.

Serial Nuclear Transfer
Procedure in which the product of the first NT step is allowed to cleave and form pronuclei which are then each transferred into another enucleated recipient cell (usually a zygote or a 2-cell blastomere). The second NT step capitalizes on the better developmental potential of correctly fertilized zygotic cytoplasm.

Tetraploid Embryo Complementation
An indirect two-step cloning procedure, whereby cloned blastocysts, derived from nuclear transfer, are used to generate ES cells for injection into tetraploid host blastocysts. The cloned ES cells form the entire embryo proper, whereas the complementing tetraploid host cells form the placenta.

Telomere
Structures that "cap" the ends of mammalian chromosomes and prevent repair mechanisms that would result in end-to-end chromosomal fusions. In the absence of the telomerase enzyme, they progressively shorten as a consequence of cell division.

Therapeutic Cloning
Deriving stem cells from embryos created with somatic cells of a human patient in order to produce immunologically compatible replacement cells for the same patient.

Xenotransplantation
Transplantation of tissues or organs between different species.

Nuclear transfer (NT), also called *nuclear replacement*, or *nuclear transplantation* is the method of choice for animal cloning. NT is a complex technique, which involves the removal/destruction of the nuclear DNA from an oocyte or zygote (enucleation), and replacing it with nuclear material from an embryonic or somatic cell. The newly introduced DNA interacts with the recipient cytoplasmic environment, and following complex and little known changes in the nuclear DNA, in some cases, will develop into adult fertile animals. Nuclear transfer is an important procedure as it allows addressing fundamental questions on epigenetic "reprogramming." As a practical tool, it can be used for agricultural and medical applications, especially combined with transgenic and embryonic stem cell technologies.

1
Introduction and Brief Historical Overview

The word "cloning" comes from the Greek "κλον," meaning twig. Indeed, many plants and simple organisms can routinely reproduce asexually. In vertebrates, however, initial cloning experiments were not aimed to study reproduction, but to understand the differentiation process of the genetic material during development. The basic idea was not new, historic experiments in the 1930s by Spemann on salamanders paved the way for later studies to explore functional changes in the nuclear genome during differentiation. These pioneering experiments in Amphibia revealed many basic mechanisms, which are similar in mammals. Despite the fact that currently most of the attention is being paid to the mammalian systems, the progress could not have been achieved without the lessons that were learnt in frog experiments.

Starting in the 1950s, with two frog species, *Rana pipiens* and *Xenopus laevis*, it was shown that embryonic cell nuclei transplanted into enucleated eggs reacquire a totipotent status and can develop into normal adults. Experiments using differentiated cells taken from adults were less successful, and only supported development to the swimming tadpole stage. Nevertheless, these studies proved that the donor cells retained all genetic information to form heart, muscle, brain, and all the other cell types in a tadpole, demonstrating that the genome is entirely preserved throughout cell differentiation and growth. Studies in amphibians and then in mammals revealed that the potential of the nucleus to direct embryonic development decreases progressively with the developmental stage of the donor cell.

In mammals, the first experiments in the 1980s using embryo-derived blastomeres and sophisticated micromanipulation systems resulted in the birth of numerous animals. This created a new commercial interest in multiplying high-value farm animals through cloning. Owing to the results in amphibians, researchers were initially discouraged to use adult donor cells for cloning. This radically changed with the birth of a lamb ("Dolly," in 1996) cloned from an adult donor's mammary gland cell, showing that even such differentiated somatic cells maintain a remarkable flexibility, and this opened new possibilities to "copy" existing individuals. Success of producing embryos and progeny in various laboratory, farm and endangered

species followed, and, in 2004, good quality human embryos were obtained from adult somatic cells.

The possibility of "xeroxing" existing individuals captured the imagination of the media and the public. In the true sense of the meaning the new individuals are not "genetically identical clones" as cytoplasmic contributions in each may vary and some of the genetic processes, for example, random X chromosome inactivation patterns in female mammalian "clones" would differ (see details in Sect. 2.3.). Furthermore, the absence of any chromosomal rearrangements would have to be demonstrated. The procedures involve numerous, technically demanding steps (Fig. 1, see details in 2.1–2.11), and these are far from efficient. In mammals, only a small fraction (1–10%, depending on the species) of blastocysts derived from somatic donor nuclei develop to term, and even fewer reach adulthood.

True "clones" can be produced by methods other than NT as well: splitting embryos into two to four pieces can result in identical twins, triplets, or quadruplets; however, high number of identical animals cannot be obtained by this method. In mouse, from embryonic stem (ES) cells, identical genetically progeny can be obtained when diploid ES cells are mixed with tetraploid embryos. In that case, the diploid cells form the fetus

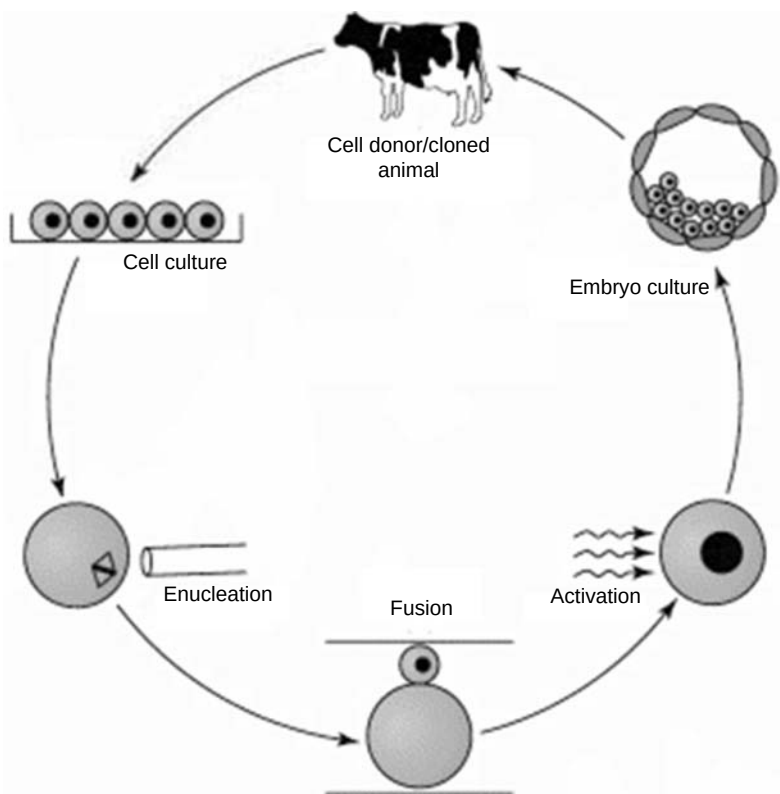

Fig. 1 Main steps of mammalian cloning by nuclear transfer. (adapted from Oback, B., Wells, D. N. (2003). *Cloning Stem Cells* **5**, 243–256).

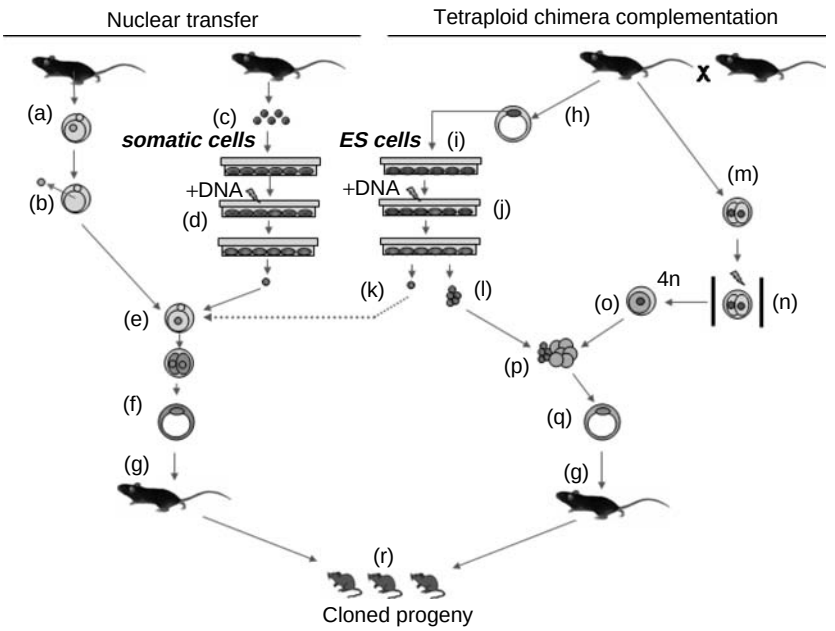

Fig. 2 Methods to produce clonal populations of mice (potentially carrying targeted genetic modifications by the first progeny generation); (a) flushed recipient oocyte (b) enucleation (c) nuclear DNA donor cells (d) gene addition, potentially gene targeting and selection of transgenic somatic cells (e) reconstructed and activated embryo (f) *in vitro* embryo culture (g) embryo transfer into pseudo-pregnant recipient (h) *in vivo* blastocyst washing (i) ES cell line establishment (j) gene addition, potentially gene targeting and selection of transgenic ES cells (k) ES cells can be used for nuclear transfer (l) ES cell cluster (m) two-cell stage *in vivo* embryo (n) fusion of the two blastomeres (o) tetraploid embryo created (p) diploid–tetraploid chimera embryo (q) *in vitro* embryo culture (r) cloned progeny.

and the tetraploid cells almost exclusively contribute to the placenta formation. This method is applied in mouse genetic manipulations (Fig. 2), and the outcome is similar to that of the nuclear transfer procedures. The present chapter will focus exclusively on the NT methods.

In the biomedical field, therapeutic cloning combined with the recent progress in human ES cell technology might offer new treatments for various diseases. Xenotransplantation in human medicine using pig organs would rely on cloning technology as well, increasing the research efforts in this species. The potential for cloning in agricultural applications is substantial, and cloned cattle were sold in several countries for breeding and production purposes. In some countries (e.g. Japan), these animals can enter the food chains, although in most other countries concerns about the safety of such animals might result in lengthy investigations before applications. Endangered breed and species preservation might also include cloning technology among the supporting technologies. Transgenic research also profited from the NT technology via the generation of transgenic and knockout farm and laboratory animals from somatic cells. The

future of the nuclear transfer technology depends on a major increase in efficiency, reliability, and safety of the technology.

This chapter discusses in detail the most relevant elements of the NT technology, including the state-of-the-art methodology, the limiting factors, ethical aspects, and the future of cloning.

2 Key Elements and Critical Aspects of the Technology

2.1 Nuclear Reprogramming and Epigenetics

The success of cloning adult animals from differentiated somatic cells has unequivocally disproved the idea that mammalian somatic cell differentiation is irreversible and terminally differentiated cells can no longer transdifferentiate into another cell type or become totipotent again.

The process of returning a differentiated somatic nucleus to a totipotent stage is termed *nuclear reprogramming*. Genes inactivated because of cell differentiation are subjected to reactivation, allowing the reconstructed, cloned embryos to support development and generation of all tissue types in the cloned individual. The inactivation of genes during cell differentiation is believed to involve epigenetic modifications of chromatin. Epigenetics is defined as nuclear inheritance, which is not based on differences in DNA sequences. It involves differential DNA methylation, posttranslational modifications of DNA-binding proteins (e.g. histone acetylation and methylation), and placement of chromatin-binding proteins (e.g multimeric complexes of polycomb- and trithorax-group proteins) to keep chromatin in an active or repressed configuration. These epigenetic signals are stably transmitted during cell division, but are reset in each generation in the gonads during fetal development. Most epigenetic signals are not inheritable from one generation to the next, but are stably maintained within the generation (with some exceptions of transgenerational epigenetic inheritance). In the following section, two of the best-studied epigenetic mechanisms are discussed.

2.1.1 DNA Methylation

In the mammalian genome, the cytosine residue (C) 5′ to a guanine residue (G) can be methylated by DNA methyl transferases and becomes ^{m}CpG (p = phosphate). The methylation pattern of DNA is maintained during cell replication (Fig. 3) by the primary DNA methyl transferase, Dnmt1, which is responsible for the transferring of the methyl group to the semi-methylated, newly replicated DNA (Fig. 3a). DNA elements that contain clusters of CpGs and have a range longer than 500 bp are called *CpG islands*. CpG islands are present in the promoter regions of most housekeeping genes and in imprinted genes, and are often associated with active transcription. DNA methylation plays a critical role in transcriptional repression, during tissue differentiation, allelic specific expression of imprinted genes, and silencing of parasitic retrotransposons. In natural reproduction, DNA methylation undergoes major changes during early development. Shortly after fertilization, a rapid loss of methylation occurs in the DNA of the paternal origin (sperm DNA), which occurs in the absence of transcription or DNA replication and is thus termed *active demethylation*. Thereafter, stepwise decreases in DNA methylation occur in the early embryos until the morula stage, as a result of the absence of functional Dnmt1.

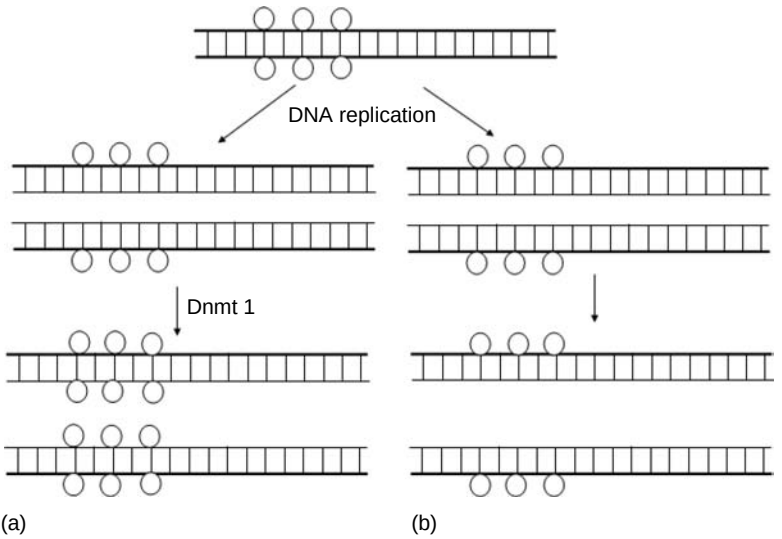

Fig. 3 Schematic illustrations of DNA methylation maintenance and passive demethylation. During DNA replication, the newly synthesized strand (thin line) of DNA will be methylated (ovals) by DNA methyl transferase, DNMT1, in the hemimethylated DNA according to the template of the mother strand (thick line, a). In the absence of functional DNMT1, the newly synthesized DNA strand will remain unmethylated (b). In the next cycle of DNA replication, half of the DNA will be completely unmethylated.

This renders the newly replicated DNA strand devoid of methylation, and the overall level of DNA methylation declines. This replication-dependent demethylation is referred to as *passive demethylation* (Fig. 3b).

In natural reproduction, relatively low levels of DNA methylation exist in the male and female gametes, which are further demethylated during early embryo development. With nuclear transplantation, the somatic nucleus carries the specific epigenetic modifications of its tissue type, which must be erased during nuclear reprogramming. Failure to erase and reestablish epigenetic marks will lead to a lack of embryo totipotency and affect further differentiation and development. Therefore, the levels of epigenetic modification existing in donor cells may affect their reprogrammability following NT. A discrepancy in reprogrammability has been observed in different cell types, which results in altered *in vitro* and *in vivo* development of cloned embryos. Furthermore, treating donor cells with pharmacological agents to remove some epigenetic marks prior to NT may improve the ability of the donor cells to be fully reprogrammed by the recipient karyoplast.

In bovine and other species, embryos cloned from somatic cells showed abnormally higher levels of DNA methylation than control embryos. In these studies, however, either the repetitive regions of the genome or the global methylation of the entire embryo were studied. The global methylation of cloned embryos also is mainly reflective of the methylation levels of the repetitive regions of the DNA. In naturally reproduced animals, repetitive regions of the DNA are normally hypermethylated, and it is unclear

whether hypermethylated repetitive DNA causes abnormal development of cloned embryos. The DNA methylation levels of specific imprinted genes, however, were found hypomethylated in cloned sheep and cloned mouse embryos. In contrast to DNA hypermethylation of cloned embryos, reduced genome-wide DNA methylation levels were found in aborted bovine cloned fetuses, implying that survivability of cloned cattle may be closely related to global methylation losses.

In summary, both global and gene-specific DNA hypermethylation and/or hypomethylation have been associated with cloned embryos and aborted cloned fetuses, suggesting that the reprogramming of epigenetic marks is aberrant using the current NT technology.

2.1.2 Histone Acetylation

Closely associated with DNA methylation is acetylation of nucleosomal histone molecules. Specifically, histone H4, H3, and H2B can be acetylated at numerous lysine residues especially, at lysine 8, 18, and 20, respectively. Acetylated histones in chromatin are also associated with increased gene expression. Evidence suggests that a threshold level of histone acetylation is required to unfold higher-order chromatin structures and facilitate gene transcription. In amphibians, failure of histone deacetylation results in block in embryonic development. In mouse, hyperacetylation is associated with strong zygotic genome activity.

There are two main epigenetic phenomena that are relatively well characterized. Both involve DNA methylation and histone acetylation changes. They are genomic imprinting and X chromosome inactivation (XCI, in females only; see further details below). Both have been active areas of study in somatically cloned animals because somatic cloning bypasses the natural process of parental specific erasure and reestablishment of epigenetic signals (occurs in the gonads). Cloning thus provides an excellent model to investigate whether and how erasure and reestablishment of epigenetic marks occur. Clones of the same donor provide unique experimental materials in that they are genetically identical, yet epigenetically different, at least so far as imprinted and X-linked genes are concerned. This animal model provides insights into epigenetic regulation that cannot be studied by models from natural reproduction, thus revealing unique features of epigenetics not possibly shown previously. A thorough understanding of epigenetic reprogramming in cloned animals will also improve the young and promising technology of cloning by revealing the ideal conditions for a complete reprogramming of the somatic nucleus.

2.1.3 Imprinting and Imprinted Genes in Cloned Animals

Earlier transplantation experiments in the mouse have established that both the maternal and paternal genomes are essential for normal embryonic development. The distinct maternal and paternal contributions are mediated by genetic imprinting, an epigenetic mechanism by which the monoallelic expression of these genes is dependent on whether they are inherited from the mother or the father. Most imprinted genes are involved in fetal growth regulation. To date, more than 40 imprinted genes have been identified, and they code for proteins as well as conserved untranslated RNAs. Examples are insulin-like growth factor 2 (*IGF2*), a paternally expressed fetal specific mitogen, and *H*19, a maternally expressed untranslated RNA

believed to be important in the regulation of *IGF2* imprinting.

Disturbance of normal imprinting (biallelic expression of imprinted genes) as seen in human patients with uniparental disomy (chromosomes or chromosomal regions derived from one parent) can result in severe developmental abnormalities. In mice, experimentally created disruptions in several imprinted genes/chromosomal regions result in embryonic or fetal lethality.

Expression of imprinted genes has been examined in a number of studies in cloned mouse embryos as well as cloned calves or mice. Cloned animals often show signs of imprinting disruptions/disturbance. Various degrees of abnormal levels of expression of several imprinted genes in fetuses and placentas have been observed in ES cell–cloned mice. However, when stem cells are used for nuclear transfer, it is possible that the widespread dysregulation of imprinted genes in these clones may have resulted from the use of stem cells whose DNA is almost totally demethylated. In mice cloned from somatic cells, it was found that the expression pattern of imprinted genes in the donor cells were faithfully transmitted to cloned animals. The levels of expression of imprinted genes, however, were abnormal. Others found striking disruptions in total transcript abundance, allele-specific expression of five imprinted genes examined compared to *in vivo* produced embryos. Only about 4% of cloned embryos recapitulated a blastocyst mode of expression for all five genes studied. Cloned embryos also exhibited extensive loss of allele-specific DNA methylation at the imprinting control regions of the *H*19 and *Snrpn* genes.

Since many of the developmental defects in cloned animals (see details below) are analogous to experimentally created imprinting disruptions in mice and naturally occurring imprinting diseases in humans, this suggests that imprinting disruptions are being introduced by nuclear transfer. Indeed, it was reported that the majority of deceased cloned full-term calves have disruption of imprinting of H19 (biallelic expression). In cloned fetuses derived from donor cells of a bovine interspecies hybrid (*Bos gaurus X B. taurus*), however, normal allelic expression of *Igf2* and *Gtl2* (Gene trap locus 2) in Day 40 bovine cloned fetuses and placentas was observed. Because cloned mice from genetically heterozygous F_1 mice survive better than those cloned from inbred mice, it is possible that this may also be true in the bovine and that the normal imprinting patterns found in this case might have reflected those from surviving clones of highly heterozygous background.

Expression abnormality of imprinted genes may be due to incomplete and aberrant reprogramming of epigenetic marks already existing in the donor cells.

2.2 Epigenetic Differences between Donor Cells

The first step in cloning and a major source of experimental variation is choosing a nuclear donor. On the basis of the donor cell epigenetic status, there are two broad categories of cloning: **embryonic cloning**, encompassing blastomeres, ES, and germ cells as donors, and **somatic cloning**, using somatic cells of varying differentiation status from fetal, newborn, or adult sources. Normally, epigenetic differences are very stable. Differentiated cells do not easily switch from one cell fate to another. It requires radical experimental manipulations, such as NT, to force them to change their cellular identity

and differentiation status. The success of cloning decreases as donor cells differentiate, thus embryonic cloning is generally much more efficient than somatic cloning. The following section describes, in order of increasing differentiation, the range of donor cell types that have been used for NT, and evaluates their usefulness for cloning.

2.2.1 Embryonic Cloning

Blastomeres After fertilization, the zygote cleaves into progressively smaller nucleated cells, called *blastomeres*. In mammals, individual blastomeres are phenotypically equivalent and totipotent (able to give rise to all embryonic and extraembryonic cell types), at least up to the 8-cell stage. Thereafter, the reduced size of the resulting embryo makes it technically difficult to assess their full developmental potential. Blastomeres were the first cells to be used in cloning amphibians. Similarly, all initial successes in mammalian cloning were achieved with nuclei from early-cleavage-stage embryos (8–16 cells) of sheep and cattle. Following NT, there is a progressive restriction in the developmental potency from the 1-cell stage onward, which is reflected in decreasing rates of cloned embryo development after NT of 2-, 4-, and 8-cell blastomeres into enucleated metaphase II (MII)–arrested oocytes, zygotes, or 2-cell blastomeres as recipients (Fig. 4). Cloning efficiencies with such

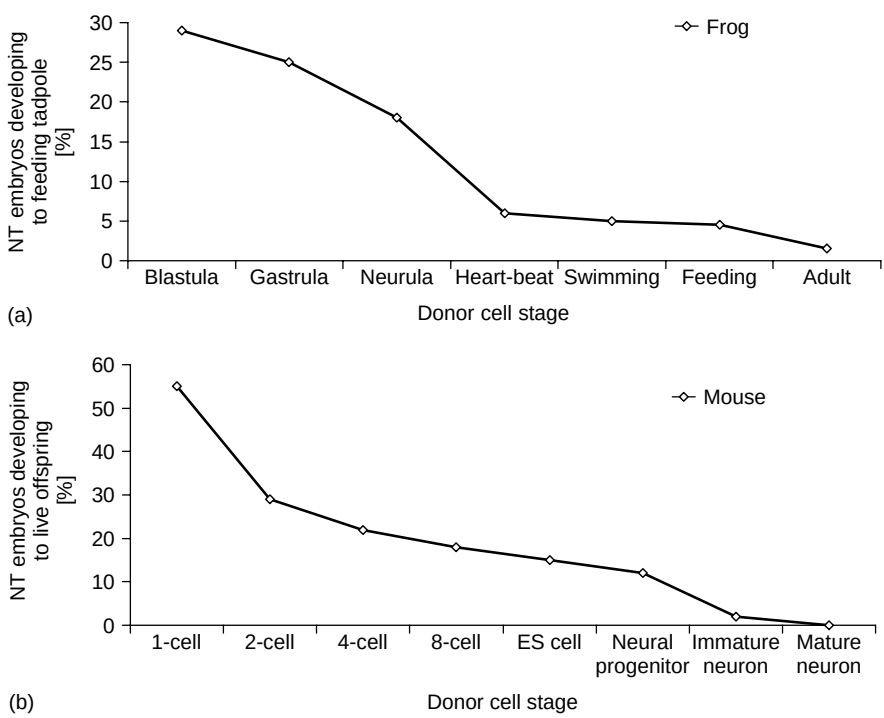

Fig. 4 Cloning efficiency declines with increasing donor cell differentiation in both frogs and mice. (a) Tadpoles advance from heartbeat to swimming to feeding stages before metamorphosis into adults. (b) Neural progenitors and immature neurons were isolated from the embryonic and mature neurons from the adult cerebral cortex, respectively.

early-cleavage-stage donor cells are much higher than with somatic cells. For example, metaphase-arrested 4-cell-stage donor nuclei are one order of magnitude more efficient in generating live mouse clones than metaphase-arrested fetal fibroblasts (43% vs 3%), using the same serial NT (see below) technique. Morula formation (8-cell stage in mouse, 32-cell stage in cattle/sheep) marks the beginning of cellular differentiation during preimplantation development, eventually leading to the foundation of the trophectoderm and inner cell mass (ICM) lineages from the outer and inner cells of the morula respectively. Donor cells derived from such later preimplantation stage embryos result in progressively lower cloning success. However, compared to somatic cells, bovine morula blastomeres still are about sixfold more efficient (28% vs 5%). Fresh ICM cells were first successfully used in sheep and cattle, soon followed by cloning from short-term bovine ICM cultures. In mouse, successful cloning from ICM and trophectoderm cells required serial NT and revealed no differences in cloning efficiencies between these two cell types (11% vs 8%). Although abnormal phenotypes are still observed with blastomere cloning (e.g. oversized fetuses and higher perinatal loss), the incidence and severity are greatly reduced. The high success rates with blastomere donors is likely due to blastomeres retaining an epigenotype that is more compatible with early embryonic development.

Embryonic stem cells ES cells are immortal pluripotent (able to give rise to all cell types of the embryo proper) cells derived from placing ICM blastomeres into *in vitro* culture. It is still unclear if ICM cells are homogeneously pluripotent, and if ES cells derive from a definable and functionally identical pluripotent ICM cell or subpopulation of cells that exists within the blastocyst prior to ES cell isolation and culture. *Bona fide* ES cells, defined as being capable of generating germ line chimaeras, are only available from a few mouse strains, and all attempts to derive such ES cells from livestock embryos have failed so far. However, long-term "ES-like cell" cultures in farm animal species show extensive *in vitro* differentiation ability or demonstrate somatic cell chimerism. Some of these cells resemble murine ES cells in both morphology and antigen profile and remain pluripotent in culture after 150 passages. Cloning efficiencies with murine ES cells are similarly remarkable as with blastomeres (10–20% vs 1–3% with fibroblasts, Sertoli, or cumulus cells) and probably for the same reasons. Like the ICM blastomeres they are derived from, ES cells retain an epigenotype that is more compatible with early embryonic development and does not need extensive reprogramming. Candidate gene expression profiling shows that all ES cell–derived blastocysts faithfully express key embryonic genes such as the transcription factor Oct4 and related sequences that are present in the pluripotent cells of the early embryo, whereas 38% of cumulus-cloned blastocysts fail to reactivate these genes.

The major problem with ES cell donors compared to blastomeres is their extreme epigenetic instability in long-term culture. DNA-methylation patterns and expression of imprinted genes vary widely between ES cell lines, between subclones of a given ES cell line, and even between individual cells of a given ES subclone after *in vitro* culture. Mice cloned from either single cells (via NT) or several cells (via tetraploid embryo complementation, see Fig. 2) of a particular ES subclone show

substantial variation in gene expression. This can be explained by epigenetic changes among daughter cells derived from the original single-cell clone. It makes it difficult to distinguish whether the observed widespread dysregulation of imprinted genes in ES cell–derived cloned mice is due to reprogramming errors or solely due to preexisting epigenomic errors in the ES donor cells.

Primordial germ cells In mammals, the germ cell precursors – the primordial germ cells (PGCs) – can be identified for the first time in the gastrulating embryo. PGCs have characteristic epigenomic modifications that distinguish them from all other cells. These include extensive genome-wide demethylation, leading to erasure of allele-specific methylation of imprinted loci and reactivation of the silent X chromosome. PGCs are probably the only mammalian cells where both parental genomes are functionally equivalent. The absence of imprints results in either biallelic expression or repression of specific imprinted genes. Contrary to previous expectations, cloning from imprint-free PGCs did not generate viable offspring, but resulted in developmental arrest at early postimplantation stages and abnormal placental development. This phenotype was consistent with a complete lack of gene dosage regulation for imprinted genes. Cloned embryos derived from meiotically arrested nongrowing female germ cells displayed similar dysregulation of gene expression and did not progress beyond the 8-cell stage. In frogs, PGC nuclear donors can support development to metamorphosis stages, consistent with the notion that global demethylation, imprinting, or inactivation of sex chromosomes is not evident in amphibian species.

Embryonic germ cells When PGCs are cultured *in vitro*, they transform into immortal pluripotent embryonic germ cells (EGCs). Their developmental potential is so similar to ES cells that their different origin is sometimes ignored and both are collectively referred to as *ES cells*, even though they differ in their expression of some imprinted genes. *Bona fide* EGCs are only available for mouse and humans, and no EGCs capable of generating germ line chimaeras have been reported from livestock animals. Previous studies have shown that they share a similar epigenotype with the PGCs they are derived from and undergo comparable epigenetic changes. Consequently, imprint-free EGCs are likely to result in the same problems previously encountered with cloning from PGCs. The epigenetic stability of EGCs compared to ES cells has not been thoroughly investigated and will depend on their DNA-methylation status at the time of isolation.

2.2.2 Somatic Cloning

The first four decades of NT experiments were dominated by the idea that cloning success depended on the donor cell being somehow embryonic in nature. This notion was supported by data from amphibian cloning, where NT with adult cells had only given limited success and never resulted in adult frogs cloned from adult somatic donor cells. However, previous preconceptions changed radically with the first successful cloning of sheep from an established embryonic cell line in 1995. This cell line was derived from culturing differentiated embryonic disk cells over several passages and did not resemble any of the previously described ES cell–like lines. This success suggested that cloning from more differentiated, even adult, somatic cells might be imminent. In 1996,

the birth of Dolly, cloned from a mammary gland cell, proved this prediction to be correct and triggered a renaissance of adult somatic nuclear transfer. Since then, numerous studies comparing various differentiated somatic donor cells types have been undertaken in frog, mouse, and cattle. As with embryonic blastomeres, the ability of somatic nuclei to promote normal development after NT decreases as donor cell differentiation increases.

Adult stem cells The mechanisms governing the segregation into a germ line and a soma line are commonly associated with restriction of cell potency in somatic cells, but this may not be universally true. Adult stem cells are defined as undifferentiated multipotent (able to give rise to a subset of cell lineages) cells capable, at the single cell level, of proliferation, self-renewal, and the production of differentiating daughter cells. They have been described for the adult bone marrow, nervous system, skin, muscle, and intestine where they are thought to be responsible for regenerating damage and maintaining homeostasis. Some adult stem cells may even be capable of differentiating across tissue lineage boundaries; however, studies proposing such apparent ''lineage plasticity'' remain controversial. Owing to the lack of exclusive phenotypic markers, adult stem cells are much more difficult to isolate and purify than ES cells; however, it is only a question of time until their applicability for cloning will be explored. Future NT experiments may also help to better define the epigenomic status of adult stem cells.

Progenitor cells Stem cells produce progenitor or precursor cells as a lineage progresses down its differentiation pathway. Progenitors are operationally defined as transiently proliferating oligopotent (able to give rise to a restricted subset of cell lineages) or unipotent (able to contribute only one mature cell type) cells that have passed the critical transition point of commitment to one or multiple lineages. Freshly isolated neural progenitors from the ventricular side of the cerebral cortex of mouse fetuses have resulted in considerably higher cloning efficiencies than any other somatic cell type in this species, (12% vs 2% with putative embryonic neurons and commonly used fibroblasts or cumulus cells), indicating that these cells may be more amenable to epigenetic reprogramming than their differentiated progeny.

Terminally differentiated cells In animals, more than 200 distinct cell types are plainly distinguishable by morphology, and more will probably be discovered when better molecular markers become available. Less than 10% of these have been tested as nuclear donors and many repeatedly failed to generate viable offspring. In frog, live offspring surviving into adulthood were obtained only with larval intestinal epithelial cells. In mouse, live offspring were obtained after NT with fibroblasts from various different tissues (skin, liver, muscle, and gonads), Sertoli, and cumulus cells. Most cell types repeatedly failed to develop after implantation, including macrophages, spleen, and brain cells. In cattle, live offspring were obtained from follicular and oviduct epithelial, mammary gland, skin, liver and muscle fibroblasts, and blood leukocytes.

One problem in somatic cloning is the lack of accurate donor cell identification. Presently, the most commonly used somatic cells are of the morphologically defined ''fibroblast-like'' phenotype.

Fibroblasts are dispersed in connective tissues throughout the body and are capable of extensive proliferation, tissue repair, and differentiation into a variety of other connective tissue cells. None of several molecular fibroblast markers, such as vimentin and FSP1, have been proven highly specific. Reliable molecular markers to distinguish between immature and mature fibroblasts or fibroblast lineages are rare. Lung fibroblasts, for example, fall into at least two subpopulations that differ in growth rate, morphology, and antigen expression. Clearly, cell lines isolated from whole fetuses and loosely termed fibroblasts on the basis of morphology are a population of mixed lineages. Leukocytes, cumulus, and Sertoli cells are better defined in terms of their origin, distinct morphology, and surface marker expression. The best-defined donor cells to date are genetically marked mature B and T lymphocytes and olfactory neurons in mouse. Fertile adult mice were derived from both these donor cell types, demonstrating unequivocally that the nuclei of terminally differentiated cells can be reprogrammed to a totipotent state. However, in both cases, cloning success required an indirect two-step procedure, whereby somatic cell NT first generates cloned blastocysts, from which ES cells are derived for injection into tetraploid host blastocysts in a subsequent round of cloning (tetraploid complementation). In this approach, the embryo is derived entirely from the ES cells, whereas the placenta is derived from the tetraploid host cells. The standard approach of directly transferring NT embryos into recipient females repeatedly failed, suggesting that cloning from terminally differentiated cells might be extremely inefficient.

Differentiated cells can remain mitotically active or postmitotic. There is no obvious correlation between natural exit from the cell cycle and cloning competence. For example, fibroblast and immature Sertoli cells are highly proliferative, whereas cumulus cells are a naturally quiescent subpopulation of ovarian granulosa cells. Cloning efficiencies for both cell types are similar.

For any given somatic cell type, the best source of cells for NT is still a matter of debate. For example, skin fibroblasts have been isolated from fetal, newborn, or adult sources with either no difference in cloning efficiencies or fetal sources achieving higher efficiencies than adult ones (13% vs 5%). For some agricultural applications, such as cloning of progeny-tested bulls, the source of donor cells will usually be restricted to easily accessible adult tissue. For transgenic applications, where the cells undergo time-consuming selection steps and extensive propagation *in vitro*, it may be beneficial to use cells from fetuses or calves since they have a longer proliferative life span before reaching replicative senescence. Cells after long-term culture or near the end of their replicative life span do not result in significantly reduced cloning efficiencies compared to fresh or short-term culture cells.

2.3 Genetic Differences between Donor Cells

Genomic non-equivalence Viable nuclear clones provide the most complete experimental evidence for genomic equivalence, that is, the concept that the DNA in every somatic cell nucleus of an individual is identical and contains all the information for the development of a new organism. Only in a few cases does cell specialization involve detectable loss of genetic material.

In mammals, red blood cells are an extreme example as they lose their entire nucleus. B and T lymphocytes eliminate a portion of their genome during differentiation; despite this, following NT, they can develop into blastocysts and term pregnancy. It is not known whether the very low frequency of further development from such cells is related to their genetic rearrangements.

Gene damage For cloning to work, the donor genome must probably be intact. It is assumed that in amphibians chromosomal abnormalities resulting from incomplete replication after NT account for all cases of early embryo losses. The same problem arises when donor cells have not completed DNA replication before NT. Continuous cell lines often show karyotypic alterations, in particular, for negative cell cycle regulators, and embryos reconstructed with spontaneously immortalized bovine mammary gland cells failed to develop to blastocysts, even though 86% showed a normal diploid karyotype. NT embryos derived from granulosa cells with high levels of aneusomy produce significantly more embryos with karyotypic errors than those from fibroblasts with fewer chromosomal anomalies or IVF-controls embryos. Evidently, such donor-derived chromosomal anomalies can substantially affect cloning success, emphasizing the need for rigorous prescreening of donor cells before NT. Higher aneuploidy has also been observed in cloned blastocysts derived from karyotypically normal granulosa cells, indicating that they can be introduced by the NT procedure itself.

Each nuclear gene is inevitably subject to a small risk of spontaneous mutation at an estimated rate of about 10^{-6} mutations per gene per cell division. Therefore, in the absence of mutagens, the chances of mutating both copies of an essential gene *in vivo* are extremely small. Theoretically, cells from some tissues may be more prone to suffer from DNA damage. Skin, for example, is more susceptible to UV-induced damage; however, ear skin fibroblasts from a 17-year old bull have been successfully used in cloning. Likewise, there is no evidence that accumulating mutations during aging and/or time *in vitro* affect cloning efficiencies. Cells after long-term culture or near the end of their replicative life span did not have significantly reduced cloning efficiencies compared to short-term culture cells.

The mutation rate in mitochondria is 10 to 20 times greater than that in nuclear DNA with some components mutating up to a 100 times more rapidly, and this may contribute to potential problems with mitochondrial heteroplasmy.

Mitochondrial heteroplasmy During normal development, genetic material outside the nucleus could in principle be inherited through both the cytoplasm of the recipient oocyte and the sperm. Mitochondrial DNA (mtDNA) in mammals, however, is exclusively maternal in origin. The sperm contributes up to 100 functional mitochondria at fertilization, which is less than 0.1% of the total number present in the egg cytoplasm. All paternal mitochondria are selectively destroyed during early cleavage stages. Destruction depends on the ubiquitination of sperm mitochondrial membrane proteins during spermatogenesis. This posttranslational modification marks them for subsequent proteolysis by the lysosomal machinery of the developing embryo. The exact evolutionary and developmental advantages of this elimination process are unknown.

Depending on NT method and donor cell type, a variable number of foreign mitochondria is initially introduced into the oocyte during cloning. Strictly speaking, cloned animals are therefore only genetic duplicates of the donor nuclear genome used for NT. At later stages some, but not all, cloned animals show low levels of mitochondrial heteroplasmy, that is, the presence of more than one type of mtDNA within a cell.

In principle, this observed heteroplasmy could result in pathologies if either the introduced mtDNA carries deleterious mutations or if the presence of intact foreign mtDNA per se somehow interferes with cell survival due to incompatibility between nuclear and different mitochondrial genomes.

In humans, more than 100-point mutations and numerous large-scale deletions have been associated with various diseases that frequently manifest themselves in tissues with high metabolic demands, such as brain and muscle. This clinical phenotype is inconsistent with the major phenotypes observed in clones. Conclusive evidence that the presence of intact foreign mtDNA alone is sufficient to impair development is also lacking. Even the assumption that most organisms are homoplasmic may be overestimated. Natural heteroplasmy occurs, indicating that the tolerance to multiple mitochondrial genotypes is higher than previously thought. Consequently, cross-subspecies and species cloning has proven to be a viable approach in cattle, sheep, and some Felidae. In conclusion, the link between heteroplasmy and cloned-offspring syndrome remains largely speculative.

Intriguingly, some incompatibility problems can even arise in homoplasmic animals. Nucleocytoplasmic hybrids between different mouse strains have shown alterations in the adult phenotype. Strikingly, these changes are transmittable through the germ line. Mismatches in the complex interactions between recipient mitochondrial genes and their complementary partner genes in the donor genome may account for the observed phenotype. These experiments emphasize the need for matching genetic background between donor cell and recipient oocyte in cloning experiments.

Genetic background Among cloned mammals so far, only mice offer the advantage of a defined genetic background. There are many inbred strains available and F1-crosses can readily be obtained. The highest cloning efficiencies were achieved with oocytes and donor cells of hybrid mice. Somatic nuclei from inbred mice (except the 129/Sv strain) seldom supported postimplantation development. Both postimplantation and postnatal viability improved dramatically when F1 ES cells, instead of inbred or out-crossed ES donors, were used. Remarkably, all mice derived from ES cells of several inbred strains died at birth from respiratory failure. Hybrid mouse embryos also perform much better in culture than inbred strains. The beneficial effect of hybrid vigor or heterosis is a well-known phenomenon in animal breeding. F1-animals are maximally heterozygous for all genetic loci, especially when they come from two inbred strains. Similarly, hybrid recipient oocytes have an advantage over homozygous oocytes in supporting *in vitro* development of cloned embryos, at least in certain culture media.

Telomeres Telomeres are protective structures (copies of a hexamer repeat sequence (5′TTAGGG3′ in human) and associated proteins) that "cap" the ends

of mammalian chromosomes and prevent repair mechanisms that would result in end-to-end chromosomal fusions. In the absence of telomerase (the primary enzyme that is capable of adding telomeric DNA to the end of chromosome) activity, they progressively shorten as a consequence of cell division. Telomere shortening correlates with cellular aging of human fibroblasts *in vitro*, and may play a role in the induction of replicative cell division arrest in the G_1 phase of cultured human cells, a state known as *senescence*. Some cells isolated from older individuals have a shorter replicative life span in culture and shorter telomeric DNA, which has been taken as evidence for a role of telomere shortening *in vivo*. However, the correlation between telomere length and aging is dependent on cell type, species, and culture conditions. Mouse cells have comparably long telomeres (≈60 kb), show telomerase activity in many tissues, and reach replicative senescence well before detectable telomere erosion (after 15–30 doublings in culture). Humans have shorter telomeres (≈12 kb), no telomerase activity in most tissues, and therefore telomere shortening may limit the longevity of human cells in culture. In cattle, average telomere length (≈20 kb) was shown to decrease with age and time in culture for both fibroblasts and very late passages of ES-like cells. Bovine ES-like cells show higher telomerase activity than fibroblasts, and, for both cell types, early passages show higher telomerase activity compared to late passage cells. More importantly, it has not been convincingly demonstrated whether cellular senescence occurs *in vivo* and whether such a process would then have anything to do with organismal aging. If this was the case, cloning from adult cells could result in offspring with shorter telomeres, premature onset of aging, and a reduced life span.

The dynamics of telomerase activity during early embryo development was extensively studied in cattle. Telomerase activity is found upregulated at the blastocyst stage in nuclear transfer systems similar to that of *in vitro* fertilized embryos. The upregulating of telomerase activity at the blastocyst stage may contribute to telomere length restoration in embryo development after nuclear transfer and could explain the normal telomere length found in some cloned animals, including cattle and mice.

In contrast, Dolly, the sheep, had shorter telomeres and calves cloned from a 10-year old bull had similar or even shorter telomeres than the donor animal, apparently depending on the cell type investigated. The latter report also describes an "adult" phenotype (e.g. wrinkled skin, thick bones, and rough hairs) of the cloned male calves, in contrast to clones from a 17-year old bull, which appeared phenotypically normal.

Donor cell effect on telomere length was described in cattle, using nuclei of donor cells derived from muscle, oviduct, mammary gland, and ear skin. All three types (normal, shorter, longer) of telomeres were found when different donor cell types were compared. Clones from muscle or ear skin cell origin had longer telomeres than the donor animal, but were within the normal range of same age controls. Clones from oviduct and mammary epithelia cell origin had telomeres that were shorter than those of controls, while clones derived from embryonic cell origins have telomeres longer than those of their age-matched controls. The time cells spend in culture might affect the process: senescent cultured bovine fibroblasts with shortened telomeres were used as nuclear donors in a

Lazarus-type experiment, whereby the resurrected donor cells, after NT, displayed even longer telomeres than age-matched controls and increased proliferative life span in culture. Similarly, in sheep, it has been shown that NT did extend the telomeres from donor nuclei derived from late-passage cells that are near senescent (maybe by repair mechanisms triggered); however, in fetuses generated from early passage donor cells, telomeres were not significantly different in length to those in the donor cells.

One difficulty in interpreting the conflicting data is the significant telomere length variation among and within individual cloned and control animals. Ideally, telomerase activity and telomere length should be compared between donor cells and cloned cells of the same type and age and cultured under exactly the same conditions. In summary, telomerase reactivation and telomere length adjustment can be accomplished after NT; however, potential effects on overall longevity of cloned animals still need to be determined in long-term studies.

X chromosome inactivation In mammals, males have one, while females have two copies of the X chromosomes. This creates a situation in which there is an unequal gene dosage among males and females. During evolution, this was solved by a process termed *X chromosome inactivation* (XCI), the random transcriptional silencing of one of the two X chromosomes in somatic cells of females during early development. XCI occurs by the process of epigenetic modification, the inactivated X chromosome is hypermethylated in DNA and hypoacetylated on histones. Proper XCI is essential to embryonic development. Inactivation of both X chromosomes in mouse embryos leads to embryonic lethality, and having more than one active X chromosome is deleterious to extraembryonic development and also causes early embryonic death in mice.

X chromosome inactivation occurs at late blastocyst stage in mice. Before fertilization, the egg carries an active X chromosome that is of maternal origin (X_a^m), while the sperm carries an inactive X of paternal origin (X_i^p). At the formation of the female zygote, both X chromosomes become active ($X_a^m X_a^p$). This state of activation persists through early blastocyst stage. During late blastocyst stage, the expression of the X-inactivation specific transcript (*Xist*) gene from one of the two X chromosomes, in a random fashion in the inner cell mass, leads to its inactivation (X_i). In extraembryonic tissues, however, the paternal X chromosome is preferentially inactivated in the mouse, resulting in imprinted XCI. Once established, the inactive state of a particular X chromosome is epigenetically inherited throughout all subsequent cell divisions.

In somatic cloning through nuclear transfer, the cloned zygotes receive one active (X_a) and one inactive (X_i) X chromosome from the donor cells (Fig. 5). This state of X inactivation is different from that in naturally fertilized female zygotes, in which both X are active. During nuclear reprogramming, it is unclear whether the inactivated X is reactivated or the pattern of XCI in the donor cell is maintained in the cloned animals. The first study of XCI reprogramming in cloned animals reported that epigenetic marks on the somatic X chromosomes in mice were completely erased and then appropriately reestablished by the nuclear reprogramming process, leading to normal random XCI in the cloned embryos. Using female donor cells with genetically marked X chromosomes, it was demonstrated in

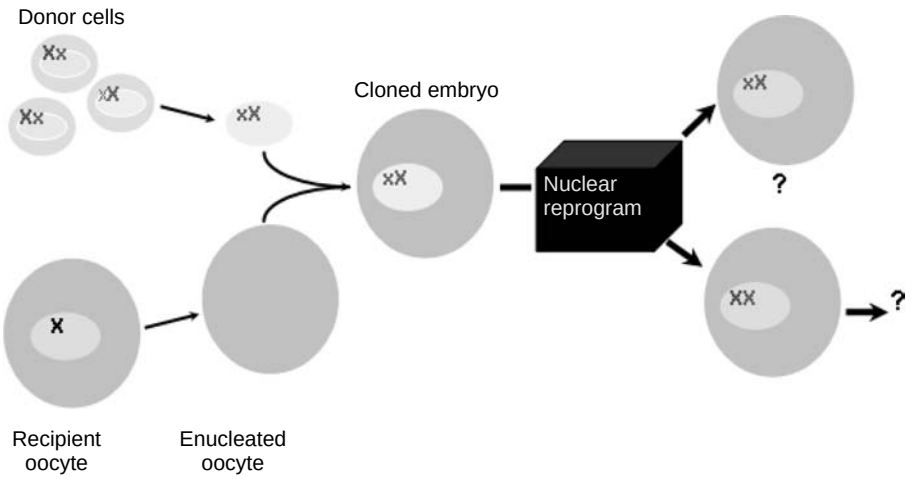

Fig. 5 Question of reprogramming of X chromosome inactivation in cloned embryos. During nuclear transfer, a somatic cell with preexisting active and inactive X chromosomes is transferred into an enucleated oocyte. After nuclear reprogramming, which is still largely a black box, it is unclear whether the inactivated X will become reactivated to result in two active X as in the naturally fertilized zygote, or the pattern of XCI of the somatic donor cell will be maintained in the cloned animal.

cloned mouse embryos that X chromosomes are initially reactivated in most or all blastocyst cells, and, subsequently, either randomly inactivated in the embryo proper or nonrandomly inactivated in the trophectoderm. The X_i in extraembryonic tissues was not always the paternal one (as during normal embryogenesis), but was determined by the status of X inactivation in the somatic donors, suggesting some functional equivalence between the imprinted paternal or maternal X and the inactive or active somatic X, respectively. Thus, the process of X inactivation was recapitulated at least in the few surviving blastocysts, female fetuses, and offspring.

In cattle studies on XCI by following the allele-specific expression of the X-linked housekeeping gene monoamine oxidase type A (*MAOA*), it was found that, as in the mouse, XCI in cattle somatic cells is also random and in placenta is paternally imprinted. In cloned calves that died shortly after birth, however, expression of both alleles of the *MaoA* gene was observed, suggesting aberrant XCI. Furthermore, inactivation of both alleles of several X-linked genes was also observed in organs of several deceased clones. Interestingly, the transducin (beta) like 1 (*TBL1*), a gene known to escape XCI in humans and mice, was expressed in all organs of these clones. Consistent with observations of aberrant XCI in internal organs, random XCI in the placenta of all deceased clones examined was found. Placental abnormalities have been reported in both live and deceased cloned calves.

The aberrant XCI in bovine clones may have resulted from incomplete erasure of the epigenetic marks on the X chromosomes of the somatic donor cells during nuclear reprogramming, which in turn may lead to only partial reactivation of the X_i or to abnormal methylation patterns of

the *Xist* gene in one or both X chromosomes prior to XCI. Upon differentiation, those epigenetic marks already present on the X chromosomes of the cloned embryo may interfere with the ones further imposed during XCI in clone development, ultimately leading to the observed aberrant expression patterns of X-linked genes. The interesting finding that *TBL1*, a gene that escapes XCI, was properly expressed in all clones may indicate that regions of the X chromosomes not subjected to XCI, thus not epigenetically modified, are less affected by events involved in nuclear reprogramming. In summary, reprogramming of the inactivated X chromosome in cloned female animals appears to be different in different species. In any case, faulty reactivation and subsequent nonrandom inactivation would not be a problem for male cells and those 50% of somatic female donor cells with a paternal X_i. Studies comparing the cloning competence of male and female somatic donors found no significant differences using male and female tail-tip fibroblasts, thus demonstrating that abnormal XCI is not the main limiting factor of NT.

2.4 Choice of Nuclear Donor Cell Cycle Stage

2.4.1 The Donor Cell Cycle

The cell division cycle has four distinct phases: G_1, S, G_2, and M. Following mitosis, the daughter cells enter G_1, during which the cells increase in size and respond to extracellular growth factors that regulate the synthesis of G_1 cyclins. The accumulation of active complexes between G_1 cyclins and cyclin dependent kinases (CDKs) results in progression past the restriction point (R-point) of the cell cycle. As cells transit the R-point and the later phase of G_1, they are committed to undergo another round of DNA replication in S-phase. This is followed by accumulation of active complexes between mitotic cyclins and their respective CDKs to pass another checkpoint in G_2 before cells enter mitosis, whereby the replicated DNA is segregated equally to the two daughter cells. This canonical cell cycle can be adjusted to meet special developmental requirements, as in the rapid cycles of cleavage blastomeres with very short gap phases, and the relative importance of different checkpoints may be altered.

In principle, cell cycle coordination between the nuclear donor cell and the enucleated recipient cell (cytoplast) can increase cloning efficiency for two reasons: (1) by maintaining normal chromosome constitution (ploidy) in the reconstructed NT embryo and (2) by promoting epigenetic reprogramming of the donor genome.

2.4.2 Cell Cycle Coordination to Maintain Normal Ploidy

There is ample evidence for the importance of donor-recipient cell cycle coordination to maintain normal ploidy. The ability of donor nuclei to maintain normal ploidy depends on the activity of cyclin B and Cdk 1 (formerly $p34^{cdc2}$) complexes (M-Cdk, formerly maturation promoting factor or MPF) in the cytoplast. In its active form, M-Cdk induces nuclear envelope breakdown, chromosome condensation, and assembly of a microtubule spindle. Metaphase II-(MII) arrested oocytes maintain a high M-Cdk activity that is stabilized by a calcium sensitive cytostatic factor (CSF). After activation, an increase in intracellular calcium inactivates CSF, enabling ubiquitin-dependant degradation of cyclin B through the anaphase-promoting complex which destroys M-Cdk activity.

Therefore, two types of cytoplast may be used for NT: either nonactivated (high M-Cdk) or activated (low M-Cdk).

NT into non-activated cytoplasts When an interphase donor nucleus is introduced into a nonactivated cytoplast, the high M-Cdk activity induces or maintains nuclear envelope disassembly, chromosome condensation, and spindle formation. Following activation, the nuclear envelope reforms around the donor chromatin, which then undergoes a round of DNA synthesis regardless of its previous cell cycle stage. From these general events, three situations can be distinguished, depending on the state of nuclear contents (e.g. amount of DNA) and cytoplasmic contents (e.g. cell cycle regulatory factors, organelles) of the donor cell:

1. *G_0/G_1-phase donor nuclei* G_0 and G_1 are often grouped together as G_0/G_1, although these phases are operationally defined differently. G_0 cells are thought to exit the cell division cycle and enter transient or permanent quiescence, whereas G_1 is the intermediate stage between M- and S-phase in growing cells. However, cells in both phases contain the same amount of chromatin (2N), and their chromosomes are not ready to segregate. As the DNA has not yet replicated, the genome must remain within the reconstructed embryo to stay diploid. Moreover, the organelles (e.g. nuclear envelope, endoplasmic reticulum, and Golgi apparatus) are in their intact interphase configuration; the centriole has not replicated; and M-Cdk is stably inactivated.
   Following NT, the donor chromatids prematurely condense and each attaches, apparently randomly, to a single pole of a newly formed spindle apparatus. After activation, nonhomologous chromosomes segregate and form multiple pseudopronuclei. In mouse, actin-depolymerizing drugs, such as cytochalasin B, are required to prevent cytokinesis and expulsion of a pseudopolar body containing chromatin, which would otherwise lead to nonviable aneuploid embryos. Expulsion of a pseudopolar body is related to the time interval between nucleus introduction and activation. When activation occurs within 5 to 30 min of reconstruction, no polar bodies are extruded and a single diploid nucleus results. However, when activation occurs at least 1 to 3 h later, the majority of NT embryos extrude a polar body. The claims that rabbit and livestock NT embryos do not undergo pseudopolar body extrusion may in part be related to the NT methodologies used in these studies, including the time interval between fusion and activation, and/or possible differences in spindle formation between species.
   Presumptive G_0 cells are obtained by depletion of essential mitogens and growth factors in the culture medium either through serum deprivation or culture to confluency. G_1 cells are obtained by different methods; for example, mitotic selection or drug-induced G_1-phase arrest. Both G_0 and defined G_1 cells have resulted in viable cloned animals.
2. *G_2/M-phase donor nuclei* In the case of G_2/M-phase cells, the donor chromosomes are duplicated (4N), sister chromatid cohesion is established to enable equal segregation; the chromatin is condensed, transcriptionally inactive, and ready to segregate. Moreover, M-phase cells have grown to their maximal

size, the nuclear envelope, endoplasmic reticulum, and Golgi apparatus are dissociated; the centriole has replicated and the mitotic spindle formed; and M-Cdk activity is highest. This is essentially the cell cycle status of the MII oocyte before enucleation and NT. The use of donor cells in M-phase may therefore be maximally compatible with MII cytoplasts.

Following NT, the double-stranded condensed chromatids align regularly on a spindle resembling a metaphase plate. Sister chromatids attach to opposite poles of the spindle and segregate orderly after activation in the absence of cytochalasin B. This is followed by expulsion of a pseudopolar body, removing the extra set of chromatin and resulting in a single, presumably diploid, pseudopronucleus in most reconstructed embryos. This phenomenon may not occur in early G_2 or if the sister chromatids fail to properly align on a spindle (if insufficient time is allowed between reconstruction and activation) or if there is uneven or random segregation. If segregation or expulsion does not occur, then following nuclear envelope reformation, DNA rereplication will compromise ploidy in the NT embryo.

Mitotic somatic donor nuclei cells have been reported to result in viable cloned mice and calves, respectively. G_2-phase somatic donors resulted in viable offspring in mouse, but not in cattle, probably due to difficulties of reliably synchronizing cell in G_2.

3. *S-phase donor nuclei* Depending upon the degree of replication, S-phase nuclei possess anything between 2 to 4N amounts of DNA, and the chromosomes are not yet ready to segregate. Following NT in a high M-Cdk environment, the chromatin of S-phase nuclei typically has a fragmented appearance and high incidence of chromosomal abnormalities. Following activation, the nucleus rereplicates its DNA, resulting in incorrect ploidy and developmental failure.

NT into activated cytoplasts In contrast to the situation above, interphase nuclei transferred to cytoplasts with basal levels of M-Cdk activity, following a sufficient interval after activation, do not undergo nuclear envelope breakdown, and it is the donor nucleus that controls DNA replication in accordance with its own cell cycle stage at the time of transfer. Thus, nuclei in G_1 or S-phase may initiate or continue replication, respectively, while those in G_2 are not induced to enter another round of DNA synthesis. Such preactivated cytoplasts have been termed *universal recipients* and are capable of coordinating the development of donor cells at any interphase stage of the cell cycle, at least in terms of maintaining normal ploidy. This has been especially important for cloning preimplantation-stage embryos, where most blastomeres are in S-phase (80–90%), and are therefore only compatible with transfer to cytoplasts low in M-Cdk. However, somatic cells transferred into cytoplasm low in MPF may not be optimal for reprogramming and embryo development.

2.4.3 Cell Cycle Coordination to Promote Reprogramming

In summary, somatic donor cells in presumptive G_0/G_1 and G_2/M-phase can result in offspring upon transfer into non-activated cytoplasts, whereas S-phase donors cannot. The question remains, however, whether any of the cell cycle

stages compatible with ploidy is superior for epigenetic reprogramming and overall cloning efficiency.

NT into non-activated cytoplasts The first reports describing the production of cloned mammals from differentiated cultured cells claimed that their success was due to the use of serum-starved cells (presumptive G_0-cells). Intriguingly, the failure of earlier studies to generate fertile cloned adults following NT with quiescent cells, for example, nondividing keratinocytes and intestinal epithelium cells in *Xenopus*, terminally differentiated erythrocytes in *Rana*, Day 15.5 male mouse fetal germ cells, and bovine cumulus cells may have been due to other limiting facets of the NT techniques. Shortly after the initial successful reports with quiescent cells, it became obvious that randomly picked donor cells from nonserum-starved cultures (presumptive G_1-cells) were also successfully reprogrammed and resulted in viable offspring following NT in cattle and mice. However, in both cases, the exact stage of the donor cell cycle at the time of NT remained unproven. Later studies directly compared better-defined cell populations (e.g. mitotically selected G_1-with serum-starved or confluent donors) to address the question whether there is an optimal donor cell cycle stage. From these studies, it appears that there is little difference between G_2/M and G_0/G_1, particularly with mice. It is less clear as to the efficiency of G_2/M-phases in livestock, and it is only a suggestion that G_0 is superior to G_1 in cattle.

NT into activated cytoplasts Preactivated cytoplasts with low M-Cdk result in very poor embryo development after NT with somatic cells in either G_0/G_1- or S-phase. It appears that nuclear envelope breakdown may be required in order to allow chromatin remodeling and expose the chromatin to licensing factors enabling initiation of DNA synthesis following NT. Preactivated telophase II cytoplasts have been successfully used with G_0/G_1-donor cells in goats, but these recipients may still contain enough residual M-Cdk activity to disassemble the nuclear envelope.

2.5 Selection and Enucleation of Recipient Oocytes

The oocyte cytoplasm contains factors that reprogramme the donor cell genome after NT. Since these factors have not yet been molecularly identified, intact oocytes are still needed for NT, and phenotypic selection of oocytes with high reprogramming potential is currently not possible. Comparing oocytes of different developmental stages and maturation treatments for their utility during NT has defined a subpopulation of recipient oocytes that currently works best for somatic NT procedures. This subpopulation consists of mature nonactivated MII-arrested oocytes derived from developmentally competent follicles of slaughtered adult animals. Metaphase I (MI) oocytes, which are also high in M-Cdk activity, have proven equally suitable recipients as MII, at least for NT experiments in frogs. In pigs, however, MI oocytes result in lower *in vitro* development rates, suggesting that high M-Cdk is necessary, but not sufficient for nuclear reprogramming. MI and MII oocytes can be sourced from prepubertal or adult animals and both types have supported development of cloned embryos into offspring. The maturation from MI to MII oocytes can occur *in*

vivo or *in vitro* and both *in vivo* and *in vitro* matured oocytes have been successfully used for cloning. Further direct comparative studies are needed to fully evaluate the influence of oocyte source and maturation method on cloning efficiency. As outlined above, for somatic donor cells, preactivated MII oocytes result in poorer development than nonactivated oocytes, presumably because maintenance of an intact donor nuclear envelope in a low M-Cdk environment does not allow efficient chromatin remodeling to occur. However, low M-CdK oocytes are suitable for blastomere cloning where most donors are in S-phase. Aging of oocytes lowers M-Cdk level, which has been shown to be beneficial in blastomer cloning. Fertilized oocytes (zygotes) and early-cleavage-stage embryos (2-cell blastomeres) have also been successfully used for early mammalian cloning experiments with embryonic, but not somatic, donor cells.

Oocyte availability often fluctuates and might limit the optimal use of cloning facilities. Cryopreserved stored oocytes have been used successfully for somatic cell NT in bovine.

Prior to NT, the genetic material of the recipient cell has to be removed or destroyed. This process is known as *enucleation*, irrespective of whether the genetic material is surrounded by a nuclear envelope or not, and the resulting enucleated recipient is called *a cytoplast*. Efficient enucleation should completely inactivate the recipient genome without compromising viability and reprogramming potential of the cytoplast. In newts and in frogs of the genus *Rana*, enucleation is done by lifting the egg pronucleus and surrounding cytoplasm out off the egg with a needle (''Porter's technique,'' 1939). The strength and elasticity of the extracellular matrix surrounding *Xenopus* eggs prevents using this method in this frog species. Instead, short exposure (less than 1 min) of the egg to ultraviolet (UV) irradiation inactivates the egg pronucleus, apparently without otherwise damaging the egg. In mammals, UV irradiation for similar time intervals, alone or in combination with the bisbenzimide DNA stain Hoechst 33342, compromises both oocyte membrane and intracellular components and is therefore not used for enucleation. Instead, a capillary glass needle is usually used to gently aspirate the chromosomes and some surrounding cytoplasm (the so-called *karyoplast*) without penetrating the plasma membrane (Fig. 6). In young MII oocytes, the chromosomes are typically, but not always, located underneath the first polar body. However, in aging MII oocytes, for example, after prolonged *in vitro* maturation, the polar body gets frequently displaced from its original position, making enucleation more difficult. Therefore, enucleation solely relying on the polar body as an indirect morphological landmark is less accurate than directly visualizing the chromosomes. Visualization can be done either by various transmitted light contrast methods, such as differential interference contrast after Normarski (DIC), Hoffmann modulation contrast (HMC), or polarized light (Pol-Scope), in species with transparent cytoplasm (e.g. mouse and rat) or by epifluorescence microscopy after staining the DNA with fluorochromes in species with opaque cytoplasm (e.g. cattle, sheep, and pig). Enucleation accuracy with such methods is close to 100%. The short-term exposure to transmitted or fluorescent light has no obvious detrimental effect on cloned embryo viability and survival. A slightly more invasive method than using a glass needle for aspiration is to manually cut the whole oocyte in half and

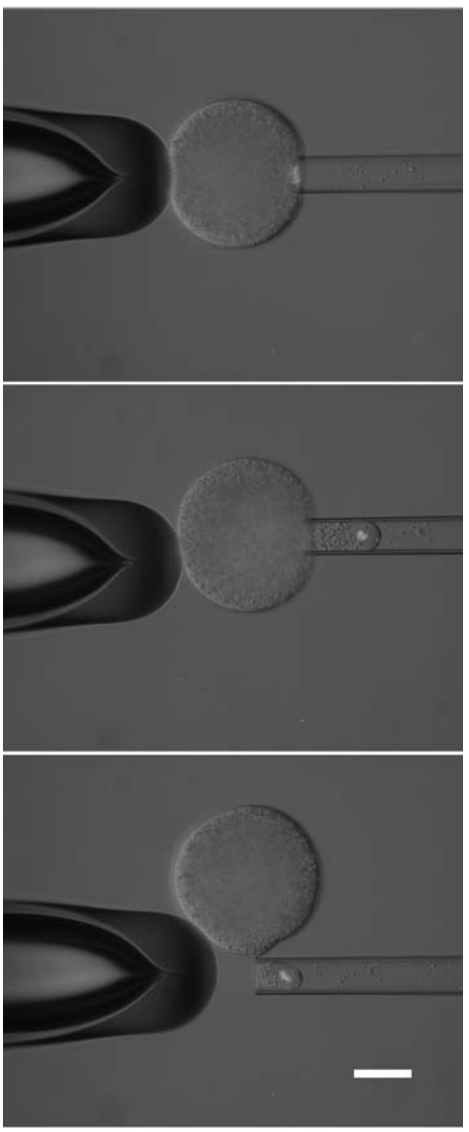

Fig. 6 Enucleation of bovine oocytes. DNA of the recipient egg is stained with a fluorescent dye (shown in blue) and mechanically removed with a fine glass needle. The plasma membrane is not penetrated during the procedure. Scale bar = 40 µm. (See color plate p. xxvii).

then UV-selecting and discarding those halves that contain the maternal chromosomes. Oocyte bisection does not use expensive micromanipulation equipment; however, it requires fusion of two half-cytoplasts in order to restore the original cytoplasmic volume, thereby wasting 50% of the original oocyte starting material. Furthermore, such cytoplasts contain two different genotypes of mitochondrial DNA, increasing the incidence of mitochondrial heteroplasmy.

An alternative method that avoids extensive micromanipulation or fluorescent light exposure is chemically assisted enucleation. This method has

been tried for more than a decade and aims to induce partial or full expulsion of the chromosomes in a pseudopolar body by using various combinations of microtubule-depolymerizing and oocyte-activating drugs. The main potential advantages are simplicity, ease of operation, and high throughput; however, so far it has only worked satisfactorily for mouse metaphase I oocytes. Irrespective of which enucleation method is employed, the quality and quantity of cytoplasmic components that are inevitably removed during the process may be critical. Depending on the method and skills of the operator, mechanical enucleation using a glass needle removes between 2 to 30% of oocytoplasm, containing not only the recipient DNA but also chromatin-associated proteins, the metaphase spindle complex, and its adjacent plasma membrane domain. The exact molecular composition of the karyoplast needs to be determined in order to better appreciate the potential developmental consequences of enucleation.

2.6 Nuclear Transfer Method

The step of introducing the donor cell DNA into the recipient cell is called *nuclear transfer* (NT), even though it is commonly not isolated nuclei that are transferred, but whole cells. The difference is probably not critical and both *bona fide* NT (by microinjection) and various whole-cell NT approaches (by fusion) have been successfully used.

2.6.1 Nuclear Injection

Intracellular injection of mechanically isolated nuclei was the first, and for many years the only, method of NT. It was part of the original amphibian cloning procedure by Briggs and King and was only adopted for mammalian NT in the 1990s, first in cattle and later in mice. For microinjection, the donor cell membrane is mechanically ruptured within the glass injection pipette and then injected within a few minutes of isolation. The amount of coinjected donor cytoplasmic components (cytoplasm, organelles) and culture medium is small compared to electrical fusion approaches. On the other hand, direct injection is technically more demanding and may require specialized equipment, such as a piezo-driven injection unit (a device that allows rapid movement over short distances), to minimize oocyte lysis during the procedure. Owing to these technical restrictions, NT of large donor cells (such as blastomeres, fibroblasts, or mitotically arrested donor cells) is typically done using fusion methods.

2.6.2 Fusion

Fusion methods are easier to master and by far the most widespread method of NT. Fusion was first achieved by exploiting the naturally fusiogenic properties of inactivated Sendai virus and later by using electrical pulses. In both cases, the plasma membranes of cytoplast and donor cell are brought in close proximity and fusion of the lipid bilayers is triggered either by catalytic viral fusion proteins or electrical pulses. In contrast to direct injection, during fusion, all cytoplasmic and plasma membrane contents of the donor cell, including organelles, centrosome, and cytosolic factors are introduced into the cytoplast. The consequences of this are not clear. Given that most oocytes are about 5- to 10-fold larger in diameter than the typical somatic donor, the donor contents get diluted by 1 to 2 orders of magnitude in the oocyte plasma

membrane and 2 to 3 orders of magnitude in the oocytoplasm, potentially minimizing their impact on subsequent development. Direct comparisons between piezo-actuated nuclear injection and electrical fusion in cattle found no significant differences in calving rates between the two methods.

Other variations of the injection/fusion methods have been explored, including injection of intact cells (rather than isolated nuclei) and fusion of transiently permeabilized cells containing artificially condensed chromatin; none of these approaches have resulted in significantly increased cloning efficiency.

2.7 Artificial Activation

Cloning subverts the sperm-mediated fertilization step that would normally lead to physiological activation of the oocyte. Since mammalian donor cells are unable to activate the recipient cytoplast, various artificial activation protocols have been employed that are meant to mimick the sperm-induced cellular events that typically occur during oocyte activation. These events include triggering of the inositol-1,4,5-trisphosphate signal transduction cascade that leads to characteristic transient intracellular calcium oscillations, initiation of the cortical granule reaction, degradation of M-Cdk, and formation of a new diploid embryo called a *zygote*.

There are a variety of agents that activate mammalian oocytes and the choice of which one to use for cloning has been largely empirical. Most current activation protocols work by sequentially administering an artificial stimulus (e.g. electric current, calcium ionophore, ethanol) followed by chemicals that are known to suppress M-Cdk reformation or activity (blockers of translation or broad-spectrum protein kinase-inhibitors, respectively). Comparative studies have so far not found any significant differences in cloning efficiency between different oocyte-activating agents. However, the origin, frequency, and amplitude of calcium oscillations are definitively altered in artificially activated eggs. Only a single calcium rise is induced by either an electric pulse or calcium ionophore, and, although this is sufficient to trigger cortical granule exocytosis and resumption of meiosis, repeated calcium oscillations may be more physiological and beneficial for later development. Likewise, downregulation of the inositol-1,4,5-trisphosphate receptor does not occur following *in vitro* oocyte activation. In addition, paternal messenger RNAs and early-sperm-derived transcripts might also be important in early embryonic development and are not delivered to the oocyte during NT. For these reasons, a zygote should theoretically make a much better recipient in terms of being properly fertilized; however, as noted under Sect. 2.4, it is not suitable for reprogramming somatic donors. One way around this dilemma is to use the zygote as a second recipient during serial NT. In this technique, the NT reconstruct is first allowed to form pronuclei and cleave before the pronucleus of each blastomere from the 2-cell NT embryo is microsurgically removed and transferred into an enucleated zygote during a second NT step. This way, the initial reprogramming occurs in the MII cytoplast, while the second NT step capitalizes on the better developmental potential of zygotic cytoplasm. The beneficial effect of serial NT has been demonstrated for cloning of mice and pigs. In summary, artificial activation and fertilization are functionally nonequivalent, and

one might expect different developmental consequences as a result of different activation protocols.

2.8 *In Vitro* Culture of Cloned Embryos

In particular in large mammals (e.g. cattle, horses, sheep), the transfer of early-cleavage-stage embryos into oviducts is technically more difficult and recipient animals are expensive. Therefore, cloned mammalian embryos are seldom transferred into recipient animals at the one-cell stage, but cultured *in vitro* for various periods of time, usually until reaching the blastocyst stage. In livestock, initial embryo culture in ligated sheep oviducts has now been virtually replaced by more defined culture methods using sequential media, where components change in accordance to the needs of the embryo, based on synthetic oviduct fluid (SOF). Studies in mice have shown that tailoring embryo culture medium to the requirements of the donor cell can also significantly improve development. Such refined *in vitro* culture systems allow selecting for NT reconstructs that are competent to complete preimplantation development and achieve blastocyst rates that are comparable to IVF embryos. However, this is unfortunately not a good predictor of later postimplantation survival after embryo transfer as most cloned embryos fail to develop into viable offspring.

2.9 Transfer of Cloned Embryos

In order to develop into viable cloned offspring, cloned embryos are transferred into appropriate surrogate mothers of the same or a closely related species. As noted under Sect. 2.8, most mammalian embryos are transferred into the uterus of recipient animals at the blastocyst stage, in particular, in large animals, where the high embryo wastage during preimplantation development would otherwise result in considerably increased cost for recipients. The number of embryos transferred varies between species. In ruminants, 1 to 4 embryos have been transferred in the past, but the trend is now toward single transfers. Since survival rates are comparable, this results in a 50 to 75% cost reduction. In pigs, pregnancy initiation requires a critical minimum signal from four embryos to the mother; consequently, the number of transferred embryo is much higher (>100). To minimize adverse *in vitro* conditions, some researchers have reduced the culture time to a minimum and transferred one-cell stage embryos into the surrogate oviduct. Goats and rabbits are similar to pigs in that respect, with 8 to 12 and 7 to 10 embryos at the 1-to-2-cell and 4-to-6-cell stage, respectively, transferred into oviducts. Transfer of early-cleavage-stage embryo circumvents potential *in vitro* culture artifacts. On the other hand, it does not allow using culture media tailored for the metabolic demands of the donor cell – a potential disadvantage, if the cloned embryos still exhibit somatic cell–like features.

Cryopreservation of cloned embryos would be beneficial to match the availability of transferable embryos and that of the recipients. Micromanipulated and *in vitro* cultured embryos are much more sensitive than flushed embryos; however, using advanced vitrification methods, cryopreservation of cloned bovine and sheep preimplantation-stage embryos has been successful. Pregnancy rates are reduced after such procedures, and further studies are needed prior to routine applications.

2.10 Pregnancy Monitoring and Progeny Production

After embryo transfer into appropriate foster mothers, pregnancy establishment and development throughout gestation are routinely monitored. A failure of the placenta to develop and function correctly is a common feature amongst somatic clones and has been documented for cattle, sheep, and mouse, but so far not for goat and pigs, indicating that species-specific problems with clones exist. Of particular concern are the losses in the second two-thirds of gestation; especially the occurrence of hydrops in cattle. Hydrops leads to acute excessive fluid accumulation in the allantoic (hydroalloantois), or rarely amniotic (hydroamnion), sac. It naturally occurs only in 0.02% of bovine pregnancies. Following artificial insemination or IVF, the incidence raises from 0.1% to up to 5%, respectively. In cloned cattle, however, up to 25% of pregnant cows at midgestation develop clinical hydrops. Cases are typically severe enough to kill the calves, and, in order to reduce the risk to the recipient cow, the standard practice is to terminate the pregnancy. Clones also often display parturition difficulties due to increased gestation period and birth weight and increased perinatal mortality due to respiratory distress is common in cattle, sheep, and mouse. To overcome problems owing to high birth weight, many clones are not delivered naturally, but by cesarean section.

2.11 Adult Clone Phenotypes and Transgenerational Effects

Somatic cell nuclear transfer is characterized by a series of developmental abnormalities collectively referred to as the *cloning-syndrome*. The cloning syndrome is defined as a predictable pattern of symptoms that occur together and characterize abnormalities caused by the cloning process. This encompasses higher rates of pregnancy loss, prolonged gestation, higher birth weight, higher rates of peri- and postnatal mortality, and specific adult phenotypes. Only about two-thirds of clones delivered at term survive at least until weaning. It has been claimed that at least these long-term survivors can be physiologically normal and apparently healthy, displaying normal behavior, growth rates, reproduction, and productivity. Equally, there are reports of various abnormal clone-associated phenotypes, including higher annual mortality rates in cattle, reduced maximal life span and obesity in mice, and compromised immune function in both species. The incidence of these anomalies varies according to the species, genotype, donor cell status, or specific aspects of the NT and culture protocols used, and it is not clear which of them could be eliminated by technical improvements.

Cloned offspring, produced by mating males and females cloned from the same cell line (namely, from male XY and phenotypically female XO ES cells), are normal and lack any obvious deleterious recessive genetic or epigenetic traits. This clearly implies that abnormal clone phenotypes are epigenetic in nature and can be corrected during gametogenesis. However, detailed molecular studies are required to exclude the possibility of subtle genetic alterations and to confirm whether all epigenetic modifications in clone-derived gametes are indeed restored to normal.

3 Applications of Nuclear Transfer in Different Species

Optimal conditions for NT vary among species. Comparative studies on reprogramming and organization of cellular elements and gene activation revealed factors potentially responsible for high variability in success rates among different species. Apart from species-specific aspects of nuclear reprogramming, in general good success is achieved in species where the level of general knowledge on embryology, especially on *in vitro* embryo production is advanced. The main goals and importance of basic research and practical applications varies among species as well. Nuclear transfer is succeeding in several new species every year, and the current status described in this section is likely to evolve rapidly.

3.1 Amphibians, Fish, and Birds

Experiments on *R. pipiens* and *X. laevis* frog species were the first to demonstrate the feasibility of cloning vertebrates, using different enucleation methods and various origin of nuclear material. Many of the mammalian experiments are simply verifying the existence of mechanisms first discovered in amphibians. However, there are several major differences in amphibian NT compared to that in mammals. The primary reason for developmental failure after NT may differ between amphibians and mammals. In Amphibia, the primary loss in somatic cell NT is during the first few cleavages as the majority of NT embryos with transplanted nuclei from differentiated or adult cells fail to divide, or undergo only a few irregular cleavages. In contrast, in mammals, the developmental loss tends to be greater after the initial cleavages. The explanation of this discrepancy might be found in the differences of the timing of the first cell cycle and the lack of imprinting in the amphibians.

In Amphibia, it is believed that nuclei from slow-dividing somatic cells cannot complete their chromosome replication in time (about 6 h needed) for the first cleavage of a recipient egg, which takes place according to the time schedule of the egg (at 1.5 h for the first cleavage in *Xenopus*). When the transplanted nuclei are starting mitosis with incompletely replicated chromosomes, chromosomal damage occurs and the defective embryos cannot survive. This might be the reason why failure in very early cleavages is the main source of losses in amphibians. When nuclei from differentiated or adult cells are used, a substantial proportion of NT embryos cleave abnormally and die as partial blastulae. However, when a second NT is performed (serial NT) with morphologically normal cells of partial blastulae into new enucleated eggs, many of the NT embryos develop relatively well in *Xenopus*, sometimes reaching the normal tadpole stage. These serial NT-derived tadpoles show the developmental potential of the original transplanted somatic cell nucleus, even though this potential was not realized by the first series of NT. The best explanation for this improvement in serial NT success might be due to the fact that in some cases the transplanted nucleus fails to complete chromosome segregation and the whole replicating transplanted nucleus moves into one of the first two daughter blastomeres. These blastomeres can complete the replication of its chromosomes before undergoing mitosis during the next cell cleavage and therefore the resulting

partial blastulae contain some normal blastomeres. Given a second chance during serial NT, these cells develop further normally. Besides serial NT, these normal looking cells from partial blastulae can be grafted to host embryos and demonstrate a wide range of developmental potential, including differentiation to normal muscle, notochord, epidermal, and other cells.

For reasons that are still not clear, nuclear transfer success declines rapidly with increasing donor age in *Rana,* and nuclear transplant development is much more successful in *Xenopus*. Despite decrease in success rates with differentiation, in some cases even nuclei of fully differentiated cells (intestinal epithelium of feeding larvae) were able to develop into normal, sexually mature males and females, thus establishing the principle that cell differentiation does not depend on losses or stable inactivation of genes. However, no such progeny have been obtained from donor cells of an adult frog, a major difference compared to mammals.

The connection between amphibian and mammalian experimental systems remains strong. Nuclei of differentiated adult cells of mice (thymocytes) and humans (white blood cells) can be to some extent reprogrammed by *Xenopus* oocytes, including the induction of pluripotency stem cell marker gene *Oct4* in the mammalian nuclei after injection into the oocyte nucleus. As amphibian oocytes contain molecules and conditions that can at least partially reprogram nuclei of adult mammalian cells, the large *Xenopus* oocytes, which have 4000 times the protein content of a mammalian egg and can be collected in big quantity (one female contains about 25 000 oocytes), provide a formidable tool for further research. Nuclear transfer has been achieved in other Amphibians, such as axolotl and *Pleurodeles*, as well.

Fish nuclear transfer has been achieved with transfer of embryonic cells into nonenucleated (generating triploid sterile progeny) or X-ray enucleated (generating cloned fertile progeny) unfertilized eggs in medaka (*Oryzias latipes*), and, to some extent, with embryonic and somatic cells in loach, cyprinidae, and, recently, in zebrafish, as well. The methods might open new possibilities for transgenic (TG) fish production and species preservation.

Attempts in birds for NT are intense, as it would present a potential method to produce TG chicken and other birds for medical and agricultural purposes, but the technical difficulties have so far prevented birth of cloned chicks.

3.2 Laboratory Animals: Mouse, Rat, and Rabbit

In mammals, there is a much longer time between the time of fertilization and the first cleavage division, therefore it is unlikely that incomplete chromosome replication is a problem like in Amphibia. However, mammals have imprinted genes not present in Amphibia, and their incomplete reprogramming can be a main reason for the high prevelance of later-stage losses, as discussed in details above. The presence of the placenta is another major difference, and in cloned mammals abnormal gene-expression patterns often reside prevalently in the placenta, rather than in the fetus itself. Placental cell lines often undergo DNA endoreplication or are multinucleated. Expression of placental genes in the adult could be deleterious, therefore they are strongly repressed in the genome. This repression is often not properly reversed during NT.

Other main differences of mammalian NT compared to Amphibia are due to the technical steps applied. In mammals, nuclear donors are often cultured cells, which raises the possibility of introducing noncontrolled epigenetic and genetic alterations. Mammalian NT enucleation removes maternal spindle proteins, which might be required to maintain ploidy through the initial cleavages. Failure to rupture the cell membrane of an injected donor cell might result in lack of cleavage, although recent studies have shown success with entire cells as well.

Mouse NT, despite many technical difficulties, is relatively well established in several laboratories around the world. A special aspect of mouse NT is the use of ES cell lines capable of generating germ line chimeras, which are not available in any other species. Several new technologies and molecular biological events are studied preferentially in mice, but the practical production of TG animals in this species is not likely to be dominated by the NT procedures in the near future, as the existing microinjection and ES cell technology–based methods work efficiently.

Rat NT is one of the biggest technological challenges, partially due to the high sensitivity of rat oocytes to spontaneous parthenogenetic activation. Recently, rat NT succeeded as well, using chemicals that block preactivation before NT. Use of TG rats in the pharmaceutical testing of new drugs might be one of the main activities in the future, and NT might represent a viable method to produce stable knockout and knockin animals in this important species.

Rabbit NT has been successful in producing progeny with fresh cumulus cells, but not with cultured cells. In this species, the first birth of progeny was partially due to a change in the "asynchrony" between the embryo recipient female's cycle and the NT-derived embryo age during embryo transfer, thus demonstrating the importance of species-specific technological adjustments. Rabbits are particularly important due to the possibility of using them as "mammary gland bioreactors" for secretion of TG proteins in the milk. Several TG human disease models could be developed in rabbit, which are not available in mouse, owing to anatomical and physiological differences (e.g. cystic fibrosis, atherosclerosis). Furthermore, placenta structure of the rabbit resembles that of the human, providing a chance for model studies on human reproduction.

3.3 Farm Animals: Sheep, Cattle, Goat, Pig, and Horses

Success rates of somatic cell NT are the highest in sheep and cattle, partially due to the high technological level of supporting embryo technologies.

Sheep was the first large animal species, where NT resulted in progeny from differentiated cells. Soon after, TG sheep carrying human blood factor genes have been generated via NT. Furthermore, the first gene-targeted animals other than mouse have been produced in this species. Sheep cloning efforts are continuing and are aimed at creating new TG animals for pharmaceutical protein production (in their milk or blood) or as models for human diseases (Fig. 7).

Cattle NT is the most efficient among all species studied and *in vitro* methods are used extensively. Recipient oocytes are usually collected from slaughterhouse ovaries and matured *in vitro*, thus reducing the expenses of the cloned embryo production. Embryos are cultured *in vitro* until blastocyst stage and then one or two embryos are transferred. In some cases,

Fig. 7 Lambs cloned by nuclear transfer with skin fibroblasts from a single founder individual of disease-resistant genotype. Producing identical copies of animals with such special genotypes is highly desirable for both basic research and commercial applications.

Fig. 8 Cloned bulls derived from adult skin fibroblasts of an elite crossbred steer. Clones from progeny-tested sires can transmit their elite genetics either through natural mating or increased semen production for artificial insemination. Cloning could be extremely useful in multiplying small numbers of genetically superior crossbred animals to maximize the benefits of both heterosis and greater uniformity within the clonal family.

around 25% of produced embryos go to term (Fig. 8). Animals have been produced from several cell types, including samples from steaks. Cryopreserved NT embryos have also been suitable for progeny production. There are potential applications of bovine NT in dairy, beef, and pharmaceutical industry. In dairy herds, cloned cows would greatly enhance progeny testing efficiency, and thus genetic progress. Also, rare or high-value animals, especially elite cows, can be reproduced this way. Even top genetic combinations achieved in F1 generation could be repeatable this way. In beef herds, cloned elite bulls or steers could be used instead of artificial insemination to distribute top genetics. Cloned bovine are allowed to enter the food chain in Japan, and are expected to be allowed in the US market in the near future. Furthermore, NT-produced cattle are tested for production of TG human proteins (including IgG antibodies) in their milk or blood.

Goat cloning has been successful, and herds of cloned TG goats have been established for human pharmaceutical protein production. There are also TG animals multiplied by NT, producing spider silk proteins, a source for a potentially high importance technical material in their milk. The use of *in vitro* matured oocytes has been successful in this species as well.

Porcine nuclear transfer progressed more slowly due to the difficulties with *in vitro* technologies and pregnancy establishment in this species. By now, the technology has been successful in many aspects, including production of progeny from slaughterhouse produced recipient oocytes and use of efficient activation methods and gene-targeted somatic cells. In order to establish pregnancies, often a very large number of embryos is transferred (over 100/recipient). The main focus of the porcine nuclear transfer was the production of TG animals suitable as

organ donors for humans (xenotransplantation). Despite the promising progress with the pig NT technology, many of the research projects have been aborted due to concerns of the presence of endogenous retroviruses in the porcine genome. Human disease models are also expected to be developed in this species, sometimes remarkably similar to humans in its physiology. In the long term, pig industry might profit from cloning top hogs, and distributing them to producers. A flattened pyramid in production systems and a greater flexibility of answering changing demands of the market could be achieved with NT. An important issue in these cases might be an increased demand for good management in order to avoid inbreeding due to the NT procedures.

In horses, numerous technological difficulties slowed down the first cloning attempts until recent success in Italy with adult somatic cells as nuclear donors. Interestingly, the embryo transfer recipient in that case was the nucleus donor as well, thus it gave birth to her own clone. Strict breeding regulations might reduce the practical applications in this species, especially in thoroughbreds. Furthermore, mule cloning has been successful, generating progeny from genetically nonfertile individuals in the United States. In the future, these animals might be used for mule racing competitions.

3.4 Endangered Species and Companion Animals

Endangered species preservation uses many reproductive techniques less expensive and much more efficient than present NT methods. However, in certain cases, NT offers advantages, such as creation of potential gene banks from any somatic tissues, including skin samples. Especially in field conditions, it would be relatively easy and noninvasive to obtain skin or hair samples, while gametes or embryos can be obtained only with much more difficulties. It would be especially beneficial in frog, fish, and bird species, where long-term preservation of female gametes and embryos has never been achieved.

In mammals, efforts to clone pandas, tigers, and other flagstaff species are gaining much attention. Saving the last remaining individual of an endangered bovine breed in New Zealand has been successful. Interspecies nuclear transfer would provide a major advantage by combining the availability of large number of bovine or other common species oocytes with that of somatic cells from a rare species. Experiments with bovine oocytes as recipients and donor cells from several species, including rats, sheep, pigs, monkeys, and humans resulted in further embryo development. However, so far only NT using donor cells and recipient oocytes from closely related species from the same genus (gaur or banteng and bovine; mouflon, argali (wild-sheep) and domestic sheep; African wildcat and domestic cat) resulted in pregnancies or progeny. Some of these animals died shortly after birth, potentially due to incompetence between the new nuclear material and the recipient oocyte mitochondria. A good demonstration of the potential of NT in species preservation is the case of a cloned African wildcat male, which was successful in breeding with wildcat females. Despite the male being heteroplasmic for wildcat and cat mitochondria, owing to the selective inheritance from only females the next

generation progeny contained only wildcat specific mitochondria.

Cloning efforts of extinct species, including Tasmanian tiger and mammoths have been started with much media attention, although owing to lack of good quality donor DNA and closely related recipient cytoplasts, the chances of success are very remote.

In companion animals, cat cloning has been successful. Commercial services for tissue collection, banking, and NT for recreating dead favorite pets are offered in several countries. Experimentation on creating TG nonallergenic cats via NT is also ongoing. In dog, no cloned progeny has been produced so far, despite intensive research efforts.

3.5 Humans and Nonhuman Primates

In humans, nuclear transfer has two potential applications. Both are using NT as starting steps, but the goals and final outcomes are profoundly different. Human "reproductive cloning" (i.e. production of a baby from a single individual's diploid cells) has been proposed and reported by some, but has never been confirmed scientifically. Adverse reaction of the public and the scientific community resulted in a strict ban of such attempts in most countries of the world.

The other approach is the "therapeutic cloning" (i.e. derivation of stem cells from embryos created with somatic cells of the patients), which would offer a chance to produce immunologically acceptable cells from and for patients. Therapeutic cloning is accepted and supported in many countries, although it is also often criticized and affected by the adverse reactions toward "reproductive cloning." Success in producing cloned human embryos followed by isolation of embryonic stem cells has been reported from South Korea, although the characterization and therapeutic potential of these cells needs further verification. Recent attempts on transfer of human somatic cells into enucleated animal oocytes (bovine and rabbit) resulted in further embryonic development; however, isolation of stable ES cell lines from the "hybrid" human-rabbit embryos was not successful. Earlier experiments on introduction of mitochondrial DNA from chimpanzee, pygmy chimpanzee, and gorilla into mitochondrial DNA-less cells were capable of restoring oxidative phosphorylation, while mitochondrial DNA from orangutan, lemur, and other monkeys failed to do so, thus demonstrating the importance of genetic compatibility for nuclear–mitochondrial interactions and the potential difficulties of NT between distant species.

In nonhuman primates, nuclear transfer would be useful in producing identical animals for research purposes, thus reducing the number of individuals required to obtain reliable data on treatment effects. Furthermore, rare animals with heritable diseases (retinitis pigmentosa, congenital hypothyroidism) could be multiplied with NT and used for developing therapies. Considering the high value of oocyte donor females and limited resources in these species, cloning attempts are rather expensive and are progressing slowly. Despite success in producing *in vitro* NT blastocysts with embryonic blastomeres, ES cells, and somatic cells, so far only those from embryonic blastomeres resulted in progeny. There are reports on producing identical "cloned" twins with the embryo splitting method, which is very different from the NT techniques. Efforts continue

to produce cloned offspring from ES and somatic cells.

3.6 Ethical and Animal Welfare Aspects of Nuclear Transfer Research

Birth of "Dolly," the first adult cell-derived sheep, initiated an intense debate on the potential of human cloning. The technical feasibility and potential impact of human reproductive cloning have often been misrepresented and exaggerated in the media and by some interest groups. From a technical point of view, the present state of the art in all species studied shows a low success rate and often serious health problems in the first generation of NT-derived animals. There is no scientific reason to believe that this would be different in humans, therefore, at present, there is no guarantee of producing only healthy babies by NT. Any human experimentation aimed to produce a progeny would be irresponsible and unethical at this stage. Despite this, several medical doctors and religious sects reported pregnancies and even birth of "healthy cloned babies." None of these reports have been scientifically confirmed, and they are most probably manipulations of the truth rather than nuclear materials. The public and legislative reactions have been very intense and mostly uniform, and have resulted in a total ban of human reproductive cloning in most countries of the world.

On the other hand, human therapeutic cloning has divided society much more and become a very complex issue; beyond the scientific arguments, the legal and societal acceptance of the technology in various countries is strongly affected by differences in culture, religion, and political standing. The spectrum of opinions varies from total refusal of the method (for example, by several religious groups, together with other forms of research involving destruction of human preimplantation stage embryos) to the strong urging of speeding up further research (for example, from some patient's interest groups, hoping for new cell therapeutic treatments in the future). Evolution of the societal and ethical judgment of therapeutic cloning is very fast and strongly connected to that of human embryonic stem cell research and its progress toward potential applications. South Korean scientists reported success in producing such embryos, and license for such experiments has been given to a research team in the United Kingdom in 2004, in accordance with a parliamentary legislative decisions from 2001.

One of the main ethical problems is related to the use of human oocytes for the generation of a human preimplantation stage embryo and then its destruction in order to isolate the embryonic stem cells. Legislation varies in different countries on the use of oocytes and embryos for research purposes. Often, one of the main distinctions is whether the oocytes and embryos have been produced specifically for research purposes, or are surplus ones from fertility treatments donated for research. The "Oviedo convention" ratified by many European countries forbids the first version and also any financial or other compensations for the donors of the biological materials. Regulation on experiments using human cells transferred into animal oocytes is not clear in most countries, neither is the status of such embryos clear from a legal point of view. As it was described in the endangered species section, it is not proven yet, whether embryos from two remotely related species would be able to develop into progeny, therefore it is not

clear whether theoretically these embryos could become human beings or not.

Society can resolve the ethical issues through open discussion and consensus reaching. Scientifically, there is also hope for resolving the ethically sensitive issues. In the future, research efforts aimed for the understanding of the nuclear reprogramming process will hopefully result in "oocyte-free" systems, which would be able to reprogram somatic cells into a "embryonic stem cell-like" status, thus would provide cells for therapy without the sacrifice of any human oocytes or embryos. Another solution would be the generation of ES or other kind of stem cells (most probably by genetic engineering), which are immunologically tolerated by any of the recipient patients during therapy. Until that stage, however, further studies and debates on therapeutic cloning can be expected.

Animal welfare of the NT animals is a complex issue. Beyond the usual measures to assure the well-being of experimental animals, the low efficiency of NT due to losses during pregnancy and after birth represents a welfare challenge. Improving the efficiency by reducing such losses would be very important, and technological changes must be evaluated in the light of the health status of the fetuses, recipient females, and progeny. In case of serious abnormalities during pregnancy or after birth, long-term suffering must be avoided by terminating the pregnancy or by humanly killing the suffering animals. There is a general hope that the successful outcome of some of the NT projects ultimately would result in reduction of the number of experimental animals needed. Furthermore, welfare issues are only associated with the NT-produced founder generation as subsequent generations derived from naturally mating cloned animals seem to be free of epigenetic errors and appear healthy.

The public attitude toward acceptance of NT-derived farm animals in the food chain, or cloning and genetic modification of companion animals in order to reproduce lost "friends" or to obtain nonallergenic cats varies and is likely to result in public debates in the future. Improving the information level of the public must be taken seriously by researchers, as frequent misinformation by media has led to many misbeliefs. An example for misleading information is the cause for the death of "Dolly," the cloned sheep who was put down due to an infectious pulmonary disease in 2003, not related to deadly NT side effects (only slight arthritis was detected postmortem), as often portrayed by the media.

4
Future Perspectives of Nuclear Transfer

Nuclear transfer is a fast developing technique, far from being optimized. Evolution in efficiency and applicability in new species is expected. However, it is not clear how much of technical development can be achieved by simply improving the present protocols. Revolutionary changes in our understanding of the reprogramming of the nuclear material, organization of chromatin, and epigenetic changes might be needed to have a proper control of the procedures, thus avoiding epigenetic aberrations in progeny.

Basic research in nuclear reprogramming will continue, especially due to connections to the ES research field. Practical application of NT in the near future is expected in the agriculture for enhanced food production and propagation of superior breeding stock; in medicine by using

animals as novel disease models, bioreactors, and organ donors produced by the combination of transgenesis/homologous recombination in somatic cell lines and NT; and in therapeutic cloning of human embryos. Methods with much improved efficiency would provide a good tool for endangered species preservation efforts as well. In case of industrial utilization, the importance of avoiding inbreeding with advanced breeding management will increase.

5 Concluding Remarks

The NT field is still in its infancy, concerning the complexity of the procedures involved and its low efficiency. Its initial effect by breaking the dogma of "irreversible cell determination" was very important, and affected the further development of stem cell and other research fields. After the production of Dolly, many scientists working on gene expression and chromatin structure got interested in using nuclear transfer as a research tool in studying epigenetic reprogramming, and now the underlying molecular mechanisms and contributing factors are getting revealed in several animal models.

The present fascination of the public shows how much the possibility of replicating existing individuals captured the imagination of many people; however, the true importance of NT will be decided only as time goes by, and by the continuing efforts of the scientific community.

Acknowledgments

The authors would like to thank Dr. Szilard Bodo for drawing Fig. 2 and Jacqui Forsyth for providing photographs of the cloned animals. A. Dinnyes has been supported by the Wellcome Trust, UK (International Senior Research Fellowship in Biomedical Science in Central Europe No. 070246), an EU FP6 Marie Curie Excellence Grant (FP6-509582), by the Hungarian National Office of Research and Technology (FEFA KMUFA BIO-00017/2002) and OTKA No. T046171 grants.

See also Developmental Cell Biology; Electroporation and Electrofusion; Living Organism (Animal) Patents; Mammalian Cell Culture Methods; Parthenogenetic Stem Cells.

Bibliography

Books and Reviews

Cibelli, J., Lanza, R., Campbell, K., West, M.D. (2002) *Principles of Cloning*, Academic Press, USA.

Clark, A.J., Burl, S., Denning, C., Dickinson, P. (2000) Gene targeting in livestock: a preview, *Transgenic Res.* **9**, 263–275.

Cummins, J.M. (2001) Mitochondria: potential roles in embryogenesis and nucleocytoplasmic transfer, *Hum. Reprod. Update* **7**, 217–228.

DiBerardino, M.A. (1997) *Genomic Potential of Differentiated Cells*, Columbia University Press, New York.

Dinnyes, A., De Sousa, P., King, T., Wilmut, I. (2002) Somatic cell nuclear transfer: recent progress and challenges, *Cloning Stem Cells* **4**, 81–90.

Fulka, J., Loi, P., Fulka, H., Ptak, G., Nagai, T. (2004) Nucleus transfer in mammals: noninvasive approaches for the preparation of cytoplasts, *Trends Biotechnol.* **22**, 279–283.

Gurdon, J.B. (1977) The croonian lecture, 1976. Egg cytoplasm and gene control in development, *Proc. R. Soc. Lond. B. Biol. Sci.* **198**, 211–247.

Gurdon, J.B. (1999) Genetic reprogramming following nuclear transplantation in Amphibia, *Semin. Cell Dev. Biol.* **10**, 239–243.

Gurdon, J.B., Byrne, J.A. (2003) The first half-century of nuclear transplantation, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 8048–8052.

Gurdon, J.B., Colman, A. (1999) The future of cloning, *Nature* **402**, 743–746.

Gurdon, J.B., Byrne, J.A., Simonsson, S. (2003) Nuclear reprogramming and stem cell creation, *Proc. Natl. Acad. Sci. U. S. A.* **100**(Suppl 1), 11819–11822.

Hochedlinger, K., Jaenisch, R. (2002) Nuclear transplantation: lessons from frogs and mice, *Curr. Opin. Cell Biol.* **14**, 741–748.

Jaenisch, R., Eggan, K., Humpherys, D., Rideout, W.M. 3rd, Hochedlinger, K. (2002) Nuclear cloning, stem cells, and genomic reprogramming, *Cloning Stem Cells* **4**, 389–396.

Lanza, R.P., Cibelli, J.B., West, M.D. (1999) Human therapeutic cloning, *Nat. Med.* **5**, 975–977.

Loi, P., Fulka, J., Ptak, G. (2003) Amphibian and mammal somatic-cell cloning: different species, common results? *Trends Biotechnol.* **21**, 471–473.

Oback, B., Wells, D. (2002) Donor cells for nuclear cloning: many are called, but few are chosen, *Cloning Stem Cells* **4**, 147–168.

Renard, J.P., Zhou, Q., LeBourhis, D., Chavatte-Palmer, P., Hue, I., Heyman, Y., Vignon, X. (2002) Nuclear transfer technologies: Between successes and doubts, *Theriogenology* **57**, 203–222.

Rideout, W.M. 3rd, Hochedlinger, K., Kyba, M., Daley, G.Q., Jaenisch, R. (2002) Correction of a genetic defect by nuclear transplantation and combined cell and gene therapy, *Cell* **109**, 17–27.

Solter, D. (2000) Mammalian cloning: advances and limitations, *Nat. Rev. Genet.* **1**, 199–207.

Spemann, H. (1938) *Embryonic Development and Induction*, Yale University Press, New Haven, CT.

Wilmut, I., Beaujean, N., de Sousa, P.A., Dinnyes, A., King, T.J., Paterson, L.A., Wells, D.N., Young, L.E. (2002) Somatic cell nuclear transfer, *Nature* **419**, 583–586.

Wolf, D., Zelinski-Wooten, M. (2001) *Assisted Fertlization and Nuclear Transfer in Mammals*, Humana Press, USA.

Primary Literature

Baguisi, A., Behboodi, E., Melican, D.T., Pollock, J.S., Destrempes, M.M., Cammuso, C., Williams, J.L., Nims, S.D., Porter, C.A., Midura, P., Palacios, M.J., Ayres, S.L., Denniston, R.S., Hayes, M.L., Ziomek, C.A., Meade, H.M., Godke, R.A., Gavin, W.G., Overstrom, E.W., Echelard, Y. (1999) Production of goats by somatic cell nuclear transfer, *Nat. Biotechnol.* **17**, 456–461.

Barnes, F.L., Collas, P., Powell, R., King, W.A., Westhusin, M., Shepherd, D. (1993) Influence of recipient oocyte cell cycle stage on DNA synthesis, nuclear envelope breakdown, chromosome constitution, and development in nuclear transplant bovine embryos, *Mol. Reprod. Dev.* **36**, 33–41.

Betthauser, J., Forsberg, E., Augenstein, M., Childs, L., Eilertsen, K., Enos, J., Forsythe, T., Golueke, P., Jurgella, G., Koppang, R., Lesmeister, T., Mallon, K., Mell, G., Misica, P., Pace, M., Pfister-Genskow, M., Strelchenko, N., Voelker, G., Watt, S., Thompson, S., Bishop, M. (2000) Production of cloned pigs from in vitro systems, *Nat. Biotechnol.* **18**, 1055–1059.

Betts, D., Bordignon, V., Hill, J., Winger, Q., Westhusin, M., Smith, L., King, W. (2001) Reprogramming of telomerase activity and rebuilding of telomere length in cloned cattle, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1077–1082.

Bordignon, V., Smith, L.C. (1998) Telophase enucleation: an improved method to prepare recipient cytoplasts for use in bovine nuclear transfer, *Mol. Reprod. Dev.* **49**, 29–36.

Bortvin, A., Eggan, K., Skaletsky, H., Akutsu, H., Berry, D.L., Yanagimachi, R., Page, D.C., Jaenisch, R. (2003) Incomplete reactivation of Oct4-related genes in mouse embryos cloned from somatic nuclei, *Development* **130**, 1673–1680.

Briggs, R., King, T.J. (1952) Transplantation of living nuclei from blastula cells into enucleated frogs' eggs, *Proc. Natl. Acad. Sci. U. S. A.* **38**, 455–463.

Bromhall, J.D. (1975) Nuclear transplantation in the rabbit egg, *Nature* **258**, 719–722.

Byrne, J.A., Simonsson, S., Western, P.S., Gurdon, J.B. (2003) Nuclei of adult mammalian somatic cells are directly reprogrammed to oct-4 stem cell gene expression by amphibian oocytes, *Curr. Biol.* **13**, 1206–1213.

Campbell, K.H., Loi, P., Otaegui, P.J., Wilmut, I. (1996) Cell cycle coordination in embryo cloning by nuclear transfer, *Rev. Reprod.* **1**, 40–46.

Campbell, K.H., McWhir, J., Ritchie, W.A., Wilmut, I. (1996) Sheep cloned by nuclear transfer from a cultured cell line, *Nature* **380**, 64–66.

Cezar, G.G., Bartolomei, M.S., Forsberg, E.J., First, N.L., Bishop, M.D., Eilertsen, K.J. (2003) Genome-wide epigenetic alterations in cloned bovine fetuses, *Biol. Reprod.* **68**, 1009–1014.

Chesne, P., Adenot, P.G., Viglietta, C., Baratte, M., Boulanger, L., Renard, J.P. (2002) Cloned rabbits produced by nuclear transfer from adult somatic cells, *Nat. Biotechnol.* **20**, 366–369.

Cibelli, J.B., Stice, S.L., Golueke, P.J., Kane, J.J., Jerry, J., Blackwell, C., Ponce de Leon, F.A., Robl, J.M. (1998) Cloned transgenic calves produced from nonquiescent fetal fibroblasts, *Science* **280**, 1256–1258.

Clark, A.J., Ferrier, P., Aslam, S., Burl, S., Denning, C., Wylie, D., Ross, A., de Sousa, P., Wilmut, I., Cui, W. (2003) Proliferative life span is conserved after nuclear transfer, *Nat. Cell Biol.* **5**, 535–538.

Collas, P., Robl, J.M. (1990) Factors affecting the efficiency of nuclear transplantation in the rabbit embryo, *Biol. Reprod.* **43**, 877–884.

Cui, W., Wylie, D., Aslam, S., Dinnyes, A., King, T., Wilmut, I., Clark, A.J. (2003) Telomerase-immortalized sheep fibroblasts can be reprogrammed by nuclear transfer to undergo early development, *Biol. Reprod.* **69**, 15–21.

Czolowska, R., Modlinski, J.A., Tarkowski, A.K. (1984) Behaviour of thymocyte nuclei in non-activated and activated mouse oocytes, *J. Cell Sci.* **69**, 19–34.

De Robertis, E.M., Gurdon, J.B. (1977) Gene activation in somatic nuclei after injection into amphibian oocytes, *Proc. Natl. Acad. Sci. U. S. A.* **74**, 2470–2474.

De Sousa, P.A., King, T., Harkness, L., Young, L.E., Walker, S.K., Wilmut, I. (2001) Evaluation of gestational deficiencies in cloned sheep fetuses and placentae, *Biol. Reprod.* **65**, 23–30.

Dean, W., Santos, F., Stojkovic, M., Zakhartchenko, V., Walter, J., Wolf, E., Reik, W. (2001) Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 13734–13738.

Dean, W., Bowden, L., Aitchison, A., Klose, J., Moore, T., Meneses, J.J., Reik, W., Feil, R. (1998) Altered imprinted gene methylation and expression in completely ES cell-derived mouse fetuses: association with aberrant phenotypes, *Development* **125**, 2273–2282.

Denning, C., Burl, S., Ainslie, A., Bracken, J., Dinnyes, A., Fletcher, J., King, T., Ritchie, M., Ritchie, W.A., Rollo, M., de Sousa, P., Travers, A., Wilmut, I., Clark, A.J. (2001) Deletion of the alpha(1,3)galactosyl transferase (GGTA1) gene and the prion protein (PrP) gene in sheep, *Nat. Biotechnol.* **19**, 559–562.

DiBerardino, M.A., King, T.J. (1967) Development and cellular differentiation of neural nuclear-transplants of known karyotype, *Dev. Biol.* **15**, 102–128.

Dindot, S.V., Farin, P.W., Farin, C.E., Romano, J., Walker, S., Long, C., Piedrahita, J.A. (2004) Epigenetic and genomic imprinting analysis in nuclear transfer derived Bos gaurus/Bos taurus hybrid fetuses, *Biol. Reprod.* **71**, 470–478.

Dinnyes, A., Dai, Y., Jiang, S., Yang, X. (2000) High developmental rates of vitrified bovine oocytes following parthenogenetic activation, in vitro fertilization, and somatic cell nuclear transfer, *Biol. Reprod.* **63**, 513–518.

Dinnyes, A., Dai, Y., Barber, M., Liu, L., Xu, J., Zhou, P., Yang, X. (2001) Development of cloned embryos from adult rabbit fibroblasts: effect of activation treatment and donor cell preparation, *Biol. Reprod.* **64**, 257–263.

Dominko, T., Mitalipova, M., Haley, B., Beyhan, Z., Memili, E., McKusick, B., First, N.L. (1999) Bovine oocyte cytoplasm supports development of embryos produced by nuclear transfer of somatic cell nuclei from various mammalian species, *Biol. Reprod.* **60**, 1496–1502.

Eggan, K., Akutsu, H., Hochedlinger, K., Rideout, W. 3rd, Yanagimachi, R., Jaenisch, R. (2000) X-Chromosome inactivation in cloned mouse embryos, *Science* **290**, 1578–1581.

Eggan, K., Akutsu, H., Loring, J., Jackson-Grusby, L., Klemm, M., Rideout, W.M. 3rd, Yanagimachi, R., Jaenisch, R. (2001) Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6209–6214.

Elsdale, T.R., Gurdon, J.B., Fischberg, M. (1960) A description of the technique for nuclear transplantation in Xenopus laevis, *J. Embryol. Exp. Morphol.* **8**, 437–444.

Enright, B.P., Kubota, C., Yang, X., Tian, X.C. (2003) Epigenetic characteristics and development of embryos cloned from donor

cells treated by trichostatin A or 5-aza-2′-deoxycytidine, *Biol. Reprod.* **69**, 896–901.

Evans, M.J., Kaufman, M.H. (1981) Establishment in culture of pluripotential cells from mouse embryos, *Nature* **292**, 154–156.

Evans, M.J., Gurer, C., Loike, J.D., Wilmut, I., Schnieke, A.E., Schon, E.A. (1999) Mitochondrial DNA genotypes in nuclear transfer-derived cloned sheep, *Nat. Genet.* **23**, 90–93.

Fankhauser, G. (1952) Nucleo-cytoplasmic relations in amphibian development, *Int. Rev. Cytol.* **1**, 165–193.

Fulka, J. Jr., Moor, R.M. (1993) Noninvasive chemical enucleation of mouse oocytes, *Mol. Reprod. Dev.* **34**, 427–430.

Galli, C., Duchi, R., Moor, R.M., Lazzari, G. (1999) Mammalian leukocytes contain all the genetic information necessary for the development of a new individual, *Cloning* **1**, 161–170.

Galli, C., Lagutina, I., Crotti, G., Colleoni, S., Turini, P., Ponderato, N., Duchi, R., Lazzari, G. (2003) A cloned horse born to its dam twin (vol 424, pg 635, 2003), *Nature* **425**, 680–680.

Gao, S., Czirr, E., Chung, Y.G., Han, Z., Latham, K.E. (2004) Genetic variation in oocyte phenotype revealed through parthenogenesis and cloning: correlation with differences in pronuclear epigenetic modification, *Biol. Reprod.* **70**, 1162–1170.

Gurdon, J.B. (1960) A description of the technique for nuclear transpantation in *Xenopus laevis*, *J. Embryol. Exp. Morphol.* **8**, 437–444.

Gurdon, J.B. (1962) The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles, *J. Embryol. Exp. Morphol.* **10**, 622–640.

Gurdon, J.B., Uehlinger, V. (1966) "Fertile" intestine nuclei, *Nature* **210**, 1240–1241.

Gurdon, J.B., Elsdale, T.R., Fischberg, M. (1958) Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei, *Nature* **182**, 64–65.

Gurdon, J.B., Laskey, R.A., Reeves, O.R. (1975) The developmental capacity of nuclei transplanted from keratinized skin cells of adult frogs, *J. Embryol. Exp. Morphol.* **34**, 93–112.

Hill, J.R., Roussel, A.J., Cibelli, J.B., Edwards, J.F., Hooper, N.L., Miller, M.W., Thompson, J.A., Looney, C.R., Westhusin, M.E., Robl, J.M., Stice, S.L. (1999) Clinical and pathologic features of cloned transgenic calves and fetuses (13 case studies), *Theriogenology* **51**, 1451–1465.

Hochedlinger, K., Jaenisch, R. (2002) Monoclonal mice generated by nuclear transfer from mature B and T donor cells, *Nature* **415**, 1035–1038.

Holliday, R. (1987) The inheritance of epigenetic defects, *Science* **238**, 163–170.

Huang, H., Ju, B., Lee, K.Y., Lin, S. (2003) Protocol for nuclear transfer in zebrafish, *Cloning Stem Cells* **5**, 333–337.

Humpherys, D., Eggan, K., Akutsu, H., Hochedlinger, K., Rideout, W.M. 3rd, Biniszkiewicz, D., Yanagimachi, R., Jaenisch, R. (2001) Epigenetic instability in ES cells and cloned mice, *Science* **293**, 95–97.

Humpherys, D., Eggan, K., Akutsu, H., Friedman, A., Hochedlinger, K., Yanagimachi, R., Lander, E.S., Golub, T.R., Jaenisch, R. (2002) Abnormal gene expression in cloned mice derived from embryonic stem cell and cumulus cell nuclei, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12889–12894.

Hutchison, C.A. 3rd, Newbold, J.E., Potter, S.S., Edgell, M.H. (1974) Maternal inheritance of mammalian mitochondrial DNA, *Nature* **251**, 536–538.

Inoue, K., Kohda, T., Lee, J., Ogonuki, N., Mochida, K., Noguchi, Y., Tanemura, K., Kaneko-Ishino, T., Ishino, F., Ogura, A. (2002) Faithful expression of imprinted genes in cloned mice, *Science* **295**, 297.

Jones, K.L., Hill, J., Shin, T.Y., Lui, L., Westhusin, M. (2001) DNA hypomethylation of karyoplasts for bovine nuclear transplantation, *Mol. Reprod. Dev.* **60**, 208–213.

Kang, Y.K., Koo, D.B., Park, J.S., Choi, Y.H., Chung, A.S., Lee, K.K., Han, Y.M. (2001) Aberrant methylation of donor genome in cloned bovine embryos, *Nat. Genet.* **28**, 173–177.

Kasinathan, P., Knott, J.G., Wang, Z., Jerry, D.J., Robl, J.M. (2001) Production of calves from G1 fibroblasts, *Nat. Biotechnol.* **19**, 1176–1178.

Kato, Y., Tani, T., Tsunoda, Y. (2000) Cloning of calves from various somatic cell types of male and female adult, newborn and fetal cows, *J. Reprod. Fertil.* **120**, 231–237.

Kato, Y., Tani, T., Sotomaru, Y., Kurokawa, K., Kato, J., Doguchi, H., Yasue, H., Tsunoda, Y. (1998) Eight calves cloned from somatic cells of a single adult, *Science* **282**, 2095–2098.

Keefer, C.L., Baldassarre, H., Keyston, R., Wang, B., Bhatia, B., Bilodeau, A.S., Zhou, J.F., Leduc, M., Downey, B.R., Lazaris, A., Karatzas, C.N. (2001) Generation of dwarf goat (Capra hircus) clones following nuclear transfer with transfected and nontransfected fetal fibroblasts and in vitro-matured oocytes, *Biol. Reprod.* **64**, 849–856.

King, T.J., Briggs, R. (1956) Serial transplantation of embryonic nuclei, *Cold Spring Harb Symp. Quant. Biol.* **21**, 271–290.

Kono, T., Kwon, O.Y., Watanabe, T., Nakahara, T. (1992) Development of mouse enucleated oocytes receiving a nucleus from different stages of the second cell cycle, *J. Reprod. Fertil.* **94**, 481–487.

Kubota, C., Yamakuchi, H., Todoroki, J., Mizoshita, K., Tabara, N., Barber, M., Yang, X. (2000) Six cloned calves produced from adult fibroblast cells after long-term culture, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 990–995.

Kubota, C., Yang, X., Dinnyes, A., Todoroki, J., Yamakuchi, H., Mizoshita, K., Inohae, S., Tabara, N. (1998) In vitro and in vivo survival of frozen-thawed bovine oocytes after IVF, nuclear transfer, and parthenogenetic activation, *Mol. Reprod. Dev.* **51**, 281–286.

Kuhholzer, B., Hawley, R.J., Lai, L., Kolber-Simonds, D., Prather, R.S. (2001) Clonal lines of transgenic fibroblast cells derived from the same fetus result in different development when used for nuclear transfer in pigs, *Biol. Reprod.* **64**, 1695–1698.

Kuroiwa, Y., Kasinathan, P., Choi, Y.J., Naeem, R., Tomizuka, K., Sullivan, E.J., Knott, J.G., Duteau, A., Goldsby, R.A., Osborne, B.A., Ishida, I., Robl, J.M. (2002) Cloned transchromosomic calves producing human immunoglobulin, *Nat. Biotechnol.* **20**, 889–894.

Kwon, O.Y., Kono, T. (1996) Production of identical sextuplet mice by transferring metaphase nuclei from four-cell embryos, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13010–13013.

Lai, L., Kolber-Simonds, D., Park, K.W., Cheong, H.T., Greenstein, J.L., Im, G.S., Samuel, M., Bonk, A., Rieke, A., Day, B.N., Murphy, C.N., Carter, D.B., Hawley, R.J., Prather, R.S. (2002) Production of alpha-1,3-galactosyltransferase knockout pigs by nuclear transfer cloning, *Science* **295**, 1089–1092.

Lanza, R.P., Cibelli, J.B., Diaz, F., C.T., M., Farin, P.W., Farin, C.E., Hammer, C.J., West, M.D., Damiani, P. (2000) Cloning of an endangered species (Bos gaurus) using interspecies nuclear transfer., *Cloning* **2**, 79–91.

Lanza, R.P., Cibelli, J.B., Blackwell, C., Cristofalo, V.J., Francis, M.K., Baerlocher, G.M., Mak, J., Schertzer, M., Chavez, E.A., Sawyer, N., Lansdorp, P.M., West, M.D. (2000) Extension of cell life-span and telomere length in animals cloned from senescent somatic cells, *Science* **288**, 665–669.

Lee, K.Y., Huang, H., Ju, B., Yang, Z., Lin, S. (2002) Cloned zebrafish by nuclear transfer from long-term-cultured cells, *Nat. Biotechnol.* **20**, 795–799.

Loi, P., Ptak, G., Barboni, B., Fulka, J. Jr., Cappai, P., Clinton, M. (2001) Genetic rescue of an endangered mammal by cross-species nuclear transfer using post-mortem somatic cells, *Nat. Biotechnol.* **19**, 962–964.

McCreath, K.J., Howcroft, J., Campbell, K.H., Colman, A., Schnieke, A.E., Kind, A.J. (2000) Production of gene-targeted sheep by nuclear transfer from cultured somatic cells, *Nature* **405**, 1066–1069.

McGrath, J., Solter, D. (1983) Nuclear transplantation in the mouse embryo by microsurgery and cell fusion, *Science* **220**, 1300–1302.

McGrath, J., Solter, D. (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes, *Cell* **37**, 179–183.

McGrath, J., Solter, D. (1984) Inability of mouse blastomere nuclei transferred to enucleated zygotes to support development in vitro, *Science* **226**, 1317–1319.

Meirelles, F.V., Bordignon, V., Watanabe, Y., Watanabe, M., Dayan, A., Lobo, R.B., Garcia, J.M., Smith, L.C. (2001) Complete replacement of the mitochondrial genotype in a Bos indicus calf reconstructed by nuclear transfer to a Bos taurus oocyte, *Genetics* **158**, 351–356.

Miyashita, N., Shiga, K., Yonai, M., Kaneyama, K., Kobayashi, S., Kojima, T., Goto, Y., Kishi, M., Aso, H., Suzuki, T., Sakaguchi, M., Nagai, T. (2002) Remarkable differences in telomere lengths among cloned cattle derived from different cell types, *Biol. Reprod.* **66**, 1649–1655.

Modlinski, J.A. (1978) Transfer of embryonic nuclei to fertilised mouse eggs and development of tetraploid blastocysts, *Nature* **273**, 466–467.

Modlinski, J.A. (1981) The fate of inner cell mass and trophectoderm nuclei transplanted to fertilized mouse eggs, *Nature* **292**, 342–343.

Munsie, M.J., Michalska, A.E., O'Brien, C.M., Trounson, A.O., Pera, M.F., Mountford, P.S. (2000) Isolation of pluripotent embryonic stem cells from reprogrammed adult mouse somatic cell nuclei, *Curr. Biol.* **10**, 989–992.

Ogonuki, N., Inoue, K., Yamamoto, Y., Noguchi, Y., Tanemura, K., Suzuki, O., Nakayama, H., Doi, K., Ohtomo, Y., Satoh, M., Nishida, A., Ogura, A. (2002) Early death of mice cloned from somatic cells, *Nat. Genet.* **30**, 253–254.

Ogura, A., Inoue, K., Takano, K., Wakayama, T., Yanagimachi, R. (2000) Birth of mice after nuclear transfer by electrofusion using tail tip cells, *Mol. Reprod. Dev.* **57**, 55–59.

Ogura, A., Inoue, K., Ogonuki, N., Noguchi, A., Takano, K., Nagano, R., Suzuki, O., Lee, J., Ishino, F., Matsuda, J. (2000) Production of male cloned mice from fresh, cultured, and cryopreserved immature Sertoli cells, *Biol. Reprod.* **62**, 1579–1584.

Onishi, A., Iwamoto, M., Akita, T., Mikawa, S., Takeda, K., Awata, T., Hanada, H., Perry, A.C. (2000) Pig cloning by microinjection of fetal fibroblast nuclei, *Science* **289**, 1188–1190.

Ono, Y., Shimozawa, N., Ito, M., Kono, T. (2001) Cloned mice from fetal fibroblast cells arrested at metaphase by a serial nuclear transfer, *Biol. Reprod.* **64**, 44–50.

Peura, T.T., Lewis, I.M., Trounson, A.O. (1998) The effect of recipient oocyte volume on nuclear transfer in cattle, *Mol. Reprod. Dev.* **50**, 185–191.

Polejaeva, I.A., Chen, S.H., Vaught, T.D., Page, R.L., Mullins, J., Ball, S., Dai, Y., Boone, J., Walker, S., Ayares, D.L., Colman, A., Campbell, K.H. (2000) Cloned pigs produced by nuclear transfer from adult somatic cells, *Nature* **407**, 86–90.

Prather, R.S., Barnes, F.L., Sims, M.M., Robl, J.M., Eyestone, W.H., First, N.L. (1987) Nuclear transplantation in the bovine embryo: assessment of donor nuclei and recipient oocyte, *Biol. Reprod.* **37**, 859–866.

Ptak, G., Clinton, M., Barboni, B., Muzzeddu, M., Cappai, P., Tischner, M., Loi, P. (2002) Preservation of the wild European mouflon: the first example of genetic management using a complete program of reproductive biotechnologies, *Biol. Reprod.* **66**, 796–801.

Rideout, W.M. 3rd, Eggan, K., Jaenisch, R. (2001) Nuclear cloning and epigenetic reprogramming of the genome, *Science* **293**, 1093–1098.

Rideout, W.M. 3rd, Wakayama, T., Wutz, A., Eggan, K., Jackson-Grusby, L., Dausman, J., Yanagimachi, R., Jaenisch, R. (2000) Generation of mice from wild-type and targeted ES cells by nuclear cloning, *Nat. Genet.* **24**, 109–110.

Robl, J.M., Gilligan, B., Critser, E.S., First, N.L. (1986) Nuclear transplantation in mouse embryos: assessment of recipient cell stage, *Biol. Reprod.* **34**, 733–739.

Schnieke, A.E., Kind, A.J., Ritchie, W.A., Mycock, K., Scott, A.R., Ritchie, M., Wilmut, I., Colman, A., Campbell, K.H. (1997) Human factor IX transgenic sheep produced by transfer of nuclei from transfected fetal fibroblasts, *Science* **278**, 2130–2133.

Shiels, P.G., Kind, A.J., Campbell, K.H., Waddington, D., Wilmut, I., Colman, A., Schnieke, A.E. (1999) Analysis of telomere lengths in cloned sheep, *Nature* **399**, 316–317.

Shin, T., Kraemer, D., Pryor, J., Liu, L., Rugila, J., Howe, L., Buck, S., Murphy, K., Lyons, L., Westhusin, M. (2002) A cat cloned by nuclear transplantation, *Nature* **415**, 859.

Sims, M., First, N.L. (1994) Production of calves by transfer of nuclei from cultured inner cell mass cells, *Proc. Natl. Acad. Sci. U. S. A.* **91**, 6143–6147.

Steinborn, R., Schinogl, P., Zakhartchenko, V., Achmann, R., Schernthaner, W., Stojkovic, M., Wolf, E., Muller, M., Brem, G. (2000) Mitochondrial DNA heteroplasmy in cloned cattle produced by fetal and adult cell cloning, *Nat. Genet.* **25**, 255–257.

Surani, M.A.H., Barton, S.C., Norris, M.L. (1984) Development of Reconstituted Mouse Eggs Suggests Imprinting of the Genome During Gametogenesis, *Nature* **308**, 548–550.

Tada, M., Tada, T., Lefebvre, L., Barton, S.C., Surani, M.A. (1997) Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells, *EMBO J.* **16**, 6510–6520.

Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., Tada, T. (2001) Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells, *Curr. Biol.* **11**, 1553–1558.

Takeda, K., Takahashi, S., Onishi, A., Goto, Y., Miyazawa, A., Imai, H. (1999) Dominant distribution of mitochondrial DNA from recipient oocytes in bovine embryos and offspring after nuclear transfer, *J. Reprod. Fertil.* **116**, 253–259.

Tamashiro, K.L., Wakayama, T., Blanchard, R.J., Blanchard, D.C., Yanagimachi, R. (2000) Postnatal growth and behavioral development of mice cloned from adult cumulus cells, *Biol. Reprod.* **63**, 328–334.

Tamashiro, K.L., Wakayama, T., Akutsu, H., Yamazaki, Y., Lachey, J.L., Wortman, M.D., Seeley, R.J., D'Alessio, D.A., Woods, S.C., Yanagimachi, R., Sakai, R.R. (2002) Cloned mice have an obese phenotype not transmitted to their offspring, *Nat. Med.* **8**, 262–267.

Tarkowski, A.K., Wroblewska, J. (1967) Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage, *J. Embryol. Exp. Morphol.* **18**, 155–180.

Tian, X.C., Xu, J., Yang, X. (2000) Normal telomere lengths found in cloned cattle, *Nat. Genet.* **26**, 272–273.

Tsunoda, Y., Yasui, T., Shioda, Y., Nakamura, K., Uchida, T., Sugie, T. (1987) Full-term development of mouse blastomere nuclei transplanted into enucleated two-cell embryos, *J. Exp. Zool.* **242**, 147–151.

Wakayama, T., Yanagimachi, R. (1999) Cloning of male mice from adult tail-tip cells, *Nat. Genet.* **22**, 127–128.

Wakayama, T., Tateno, H., Mombaerts, P., Yanagimachi, R. (2000) Nuclear transfer into mouse zygotes, *Nat. Genet.* **24**, 108–109.

Wakayama, T., Perry, A.C., Zuccotti, M., Johnson, K.R., Yanagimachi, R. (1998) Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei, *Nature* **394**, 369–374.

Wakayama, T., Rodriguez, I., Perry, A.C., Yanagimachi, R., Mombaerts, P. (1999) Mice cloned from embryonic stem cells, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 14984–14989.

Wakayama, T., Tabar, V., Rodriguez, I., Perry, A.C., Studer, L., Mombaerts, P. (2001) Differentiation of embryonic stem cell lines generated from adult somatic cells by nuclear transfer, *Science* **292**, 740–743.

Wakayama, T., Shinkai, Y., Tamashiro, K.L., Niida, H., Blanchard, D.C., Blanchard, R.J., Ogura, A., Tanemura, K., Tachibana, M., Perry, A.C., Colgan, D.F., Mombaerts, P., Yanagimachi, R. (2000) Cloning of mice to six generations, *Nature* **407**, 318–319.

Wells, D.N., Misica, P.M., Tervit, H.R. (1999) Production of cloned calves following nuclear transfer with cultured adult mural granulosa cells, *Biol. Reprod.* **60**, 996–1005.

Wells, D.N., Misica, P.M., Day, T.A., Tervit, H.R. (1997) Production of cloned lambs from an established embryonic cell line: a comparison between in vivo- and in vitro-matured cytoplasts, *Biol. Reprod.* **57**, 385–393.

Wells, D.N., Misica, P.M., Tervit, H.R., Vivanco, W.H. (1998) Adult somatic cell nuclear transfer is used to preserve the last surviving cow of the Enderby Island cattle breed, *Reprod. Fertil. Dev.* **10**, 369–378.

Wells, D.N., Laible, G., Tucker, F.C., Miller, A.L., Oliver, J.E., Xiang, T., Forsyth, J.T., Berg, M.C., Cockrem, K., L'Huillier, P.J., Tervit, H.R., Oback, B. (2003) Coordination between donor cell type and cell cycle stage improves nuclear cloning efficiency in cattle, *Theriogenology* **59**, 45–59.

Willadsen, S.M. (1986) Nuclear transplantation in sheep embryos, *Nature* **320**, 63–65.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., Campbell, K.H. (1997) Viable offspring derived from fetal and adult mammalian cells, *Nature* **385**, 810–813.

Woods, G.L., White, K.L., Vanderwall, D.K., Li, G.P., Aston, K.I., Bunch, T.D., Meerdo, L.N., Pate, B.J. (2003) A mule cloned from fetal cells by nuclear transfer, *Science* **301**, 1063–1063.

Wrenzycki, C., Wells, D., Herrmann, D., Miller, A., Oliver, J., Tervit, R., Niemann, H. (2001) Nuclear transfer protocol affects messenger rna expression patterns in cloned bovine blastocysts, *Biol. Reprod.* **65**, 309–317.

Xue, F., Tian, X.C., Du, F., Kubota, C., Taneja, M., Dinnyes, A., Dai, Y., Levine, H., Pereira, L.V., Yang, X. (2002) Aberrant patterns of X chromosome inactivation in bovine clones, *Nat. Genet.* **31**, 216–220.

Yamazaki, Y., Makino, H., Hamaguchi-Hamada, K., Hamada, S., Sugino, H., Kawase, E., Miyata, T., Ogawa, M., Yanagimachi, R., Yagi, T. (2001) Assessment of the developmental totipotency of neural cells in the cerebral cortex of mouse embryo by nuclear transfer, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14022–14026.

Young, L.E., Fernandes, K., McEvoy, T.G., Butterwith, S.C., Gutierrez, C.G., Carolan, C., Broadbent, P.J., Robinson, J.J., Wilmut, I., Sinclair, K.D. (2001) Epigenetic change in IGF2R is associated with fetal overgrowth after sheep embryo culture, *Nat. Genet.* **27**, 153–154.

Zhou, Q., Jouneau, A., Brochard, V., Adenot, P., Renard, J.P. (2001) Developmental potential

of mouse embryos reconstructed from metaphase embryonic stem cell nuclei, *Biol. Reprod.* **65**, 412–419.

Zhou, Q., Renard, J.P., Le Friec, G., Brochard, V., Beaujean, N., Cherifi, Y., Fraichard, A., Cozzi, J. (2003) Generation of fertile cloned rats by regulating oocyte activation, *Science* **302**, 1179–1179.

Nucleic Acid and Protein Sequence Analysis and Bioinformatics

Pietro Liò[1] *and Martin John Bishop*[2]
[1] *Computer Laboratory, University of Cambridge, Cambridge, UK*
[2] *MRC Rosalind Franklin Centre for Genomics Research Hinxton, Cambridge, UK*

1 Principles of Genome Sequence Analysis 251

2 Genome Segmentation Analysis 251

3 Detecting Genes 254
3.1 Analysis of Duplicated Genes 255
3.2 Phylogenetic Methods for Orthologous Genes 256

4 Detecting Regulatory Regions 257

5 Periodicities and Repeats 261

6 Bioinformatics Analysis of Protein Modifications 263

7 Genome Feature Classification: Gene Ontology 263

8 Limitations of Sequence Analysis 264

Bibliography 264
Books and Reviews 264
Primary Literature 265

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

Keywords

Markov Process
A mathematical model of infrequent changes of (discrete) states over time, in which future events occur by chance and depend only on the current state, and not on the history of how that state was reached. In molecular phylogenetics, the states of the process are the possible nucleotides or amino acids present at a given time and position in a sequence and state changes represent mutations in sequences.

Maximum Likelihood
The likelihood of a hypothesis is equal to the probability of observing the data if that hypothesis were correct. The statistical method of maximum likelihood chooses amongst hypotheses by selecting the one which maximizes the likelihood; that is, which renders the data the most plausible.

Phylogenetic Inferences
The statistical methodology based on the study of phylogenies and processes of evolution by the analysis of DNA or amino acid sequence data.

Phylogenetic Tree
The hierarchical relationships among organisms arising through evolution. The tree comprises a set of nodes linked together by branches. Terminal nodes typically represent known sequences from extant organisms. Internal nodes represent ancestral divergences into two (or more) genetically isolated groups; each internal node is attached to one branch representing evolution from its ancestor, and two (or more) branches representing its descendants. The lengths of the branches in the tree can represent the evolutionary distances that separate the nodes; the tree topology is the information on the order of relationships, without consideration of the branch lengths.

Bioinformatics is the application of computer technology to the management and use of molecular biology and genetic information. It is based on computer science and statistics; the current focus is on probabilistic models to analyze biological sequences at the genome level and other "omics" levels (e.g. transcriptomics and proteomics data). The major challenge in bioinformatics is the accurate prediction and modelling of the genomewide RNA:DNA, RNA:RNA, protein:DNA, protein:RNA, and protein:protein interaction patterns in the entire range of cellular and extracellular environments. This achievement would lead to the solution of the principles of genome-structure organization, the identification of the transcription initiation and termination and the RNA splicing pattern of any primary transcript in any tissue and the precise prediction of the folding pathway of any protein, the formation of quaternary and quinary (intermolecular) structures, and the design of effective drugs. Here, we summarize the state of the art and the potential of

sequence analysis. Since a single review cannot describe the ever-broadening subject that is genome analysis, we proceed by discussing methods for genome composition analysis, gene- and transcription-site identification, and analysis of repeats, and conclude with speculation on the future course of bioinformatics analyses.

1 Principles of Genome Sequence Analysis

Genome studies have become central to a wide range of biological areas. The term *genome* derives from Gene + Ome. A gene is an ordered sequence of nucleotides located in a particular position on a particular chromosome that encodes specific functional products (i.e. protein or RNA molecules). It contains upstream regulatory regions. The suffix "Ome" is, according to Merriam-Webster dictionary, from the Latin word for "mass". The genome, therefore, represents the whole genetic information content of an organism. Genome analysis involves the description of the general rules that allow an organism to utilize and combine the different types of information present in the genomic sequence, within its physiological and environmental contexts. There is no unique criterion for describing the genome of even the simplest organism. The situation can be well represented by the analogy from engineering drafting where blueprints of some complicated machinery are usually replaced with an exploded view of the same, with each component shown from many different angles or in holographic view. Similarly, location and abundance of coding regions, transcription signals, and repeats, are usually represented in the form of several color-coded wheels. Values that are more than plus or minus three standard deviations from average are shown as very dark lines in the diagram. Genome sequences have been investigated on the basis of several characteristics, for example, linguistics, percentage of shared homologous genes, and patterns associated with physicochemical properties of sequences, such as base stacking energy, propeller twist and curvature, and genomic features such as ancient repeats, gene family composition, and metabolic pathways. An important first step is to identify large regions with remarkably different sequence composition. These may result from lateral gene transfer in bacteria or from the existence of isochores (defined in Sect. 2) in higher vertebrate genomes. This analysis is generally termed *genome segmentation*.

2 Genome Segmentation Analysis

Genome segmentation methods are usually based on the assessment of the statistical significance of over- and underrepresented oligomers in complete genomes. The analysis of the bias of oligomers requires a model of how sequences are generated and the statistical distribution of the oligomer. The simplest composition analysis is the counting of guanines and cytosines. The G + C content is an interesting statistical property of a genome since there are large differences both within and between genomes. In bacteria, genomic islands with different G + C content with

respect to the neighboring regions are generally the result of lateral gene transfer and recombination events. In pathogenic strains, some of the G + C islands carry virulence genes that code for toxins, adhesins, invasins, or other virulence factors.

Vertebrate genomes are made up of long DNA segments, the isochores, which are homogeneous in GC content and differ in gene content. For example, the human genome can be subdivided into five families, L1, L2, H1, H2, and H3, which are characterized by increasing GC levels and gene concentrations; the two GC-poor families (L1 and L2) represent 30 and 33% of the genome, respectively, and three GC-rich families (H1, H2, and H3) represent 24, 7.5, and 4 to 5% of the genome, respectively. Gene concentration reaches up to a 20-fold higher level in H3 than in L1 isochores. The isochore organization of the avian genome contains an additional GC-richest H4 isochore family. It is known that isochores are mainly present in genomes of warm-blooded vertebrates, while only very few cold-blooded vertebrates have isochore-like genomes Since a 10 °C increase in *in vitro* temperature corresponds to a 5.7-fold increase in the rate of cytosine deamination, Fryxell and Zuckerkandl proposed that the deamination of methylcytosine and cytosine might be the cause of GC isochores.

In order to detect GC-patterns, sequences are usually coded as C, G = 1; A, T = −1. A G + C profile is analyzed using a sliding window of fixed size, usually 10 to 50 kb, and by computing the chi-square statistics of the G + C content of each window with respect to the average value of the whole genome. The average difference in G + C content between two adjacent windows gives an estimate of the composition's heterogeneity. The localization accuracy of this method is of the order of the length of the window: if the window is small, its G + C content is subject to strong fluctuations, whereas if the window is large, it may conceal the heterogeneity.

Nekrutenko and Li have developed a compositional heterogeneity index, H_{gc}, to quantify and compare the compositional differences within and between genome sequences. The authors considered a sequence that can be divided into n windows of length L and calculated the genomic average $GC\%$, GC_{av}, and the GC content of each window, (e.g. GC_1, $GC_2, \ldots, GC_n$). Then, they computed the average GC-content difference between two adjacent windows, as given in Eq. (1).

$$H_{gc} = \frac{\frac{1}{n}\sum_{i=2}^{n} |GC_i - GC_{i-1}|}{\sqrt{\frac{GC_{av}(1 - GC_{av})}{L}}} \qquad (1)$$

where the term at the denominator represents the standard error, s.e.; the s.e. reaches the maximum value at $GC_{av} = 0.5$. Note that, if the window is small, its GC content is subject to strong fluctuations, whereas if the window is large, it may conceal heterogeneity.

An oligomer appears with an unexpected frequency in a sequence if the number of its occurrences is significantly different from an estimator of the expected count under the considered model. Karlin and coworkers have shown that the relative frequencies of dinucleotides from DNA segments within a genome are more similar than sequences from different organisms and that related organisms have similar dinucleotide patterns. Local analysis of dinucleotide relative abundances has been used to detect gene-transfer events. A common assessment of dinucleotide bias

is through the odds ratio $\rho_{xy} = f_{xy}/f_x f_y$, where f_x denotes the frequency of the nucleotide X and f_{xy} is the frequency of the dinucleotide xy. In order to accommodate double-stranded DNA, the sequence is concatenated with its inverted complementary sequence. A measure of difference between two genome sequences A and B (from different organisms or from different regions of the same genome) is the absolute difference of the dinucleotide relative abundance, which is calculated as given in Eq. (2)

$$\delta(A, B) = \frac{1}{16} \sum_{xy} |\rho_{xy}(A) - \rho_{xy}(B)| \quad (2)$$

where the sum extends over all dinucleotides xy. Dinucleotide-odds ratio values reflect the chemistry of dinucleotide stacking energies and the species-specific properties of DNA modification, replication, and repair mechanisms. Because of the hydrophobic forces, the stacking energy of the base pairs on top of one another is the most important property contributing to DNA-helix stability. Trinucleotides are correlated to dinucleotide abundances and, in order to factor out all the lower-order biases, the odd ratio is computed, as given in Eq. (3)

$$\rho = \frac{f_{xyz}\, f_x\, f_y\, f_z}{f_{xy}\, f_{yz}\, f_{xnz}} \quad (3)$$

where f_{xyz} is the frequency of the trinucleotide xyz and n is A,C,T, or G. Trinucleotide frequencies also reflect the codon-usage bias. The codon-usage bias can be used to predict gene-expression levels in prokaryotic and low eukaryotic genomes: genes that deviate strongly in codon usage from the average gene but that are sufficiently similar to ribosomal protein genes and to translation and transcription processing factors. Let suppose that "g_1" is a gene with average codon frequencies, $a(x, y, z)$, for the codon triplets (x, y, z) normalized so that $\Sigma_{(xyz)=a} g_1(x, y, z) = 1$, where the sum extends over all codons (x, y, z) translated to amino acid a; the codon-usage bias of the gene "g_2" relative to the gene "g_1" is calculated by the formula given in Eq. (4).

$$C\left(\frac{g_2}{g_1}\right) = \sum_a p_a(g_1) \left[\sum_{((x,y,z)=a)} |g_2(x, y, z) - g_1(x, y, z)| \right] \quad (4)$$

where $p_a(g_1)$ are the average amino acid frequencies of the gene g_1.

The standard methodology for longer words uses Markov chains to obtain a theoretical expectation for the count of an oligomer. The order k of the Markov chain model means that the probability that a base occurs at a given position in the sequence depends on the k previous bases. The choice of the order of the Markov model depends on sequence length because of the data requirements in estimation. There are two kinds of statistics to compare the theoretical expectation with the real observed count. For common oligomers, a Gaussian approximation is appropriate and the statistic is simply the z-score. For rare oligomers, a compound Poisson approximation is generally used. More specialized methods, which are used in gene prediction, implement hidden Markov models (HMM).

A discrete Markov process is a mathematical model of infrequent changes of discrete states over time, where future events occur by chance and depend only on the current state, and not on the history of how the state was reached. A hidden Markov model is a finite sequence of states;

the Markov property assures that the next state to occupy depends only on the current state. Each state is associated with a probability distribution and transitions among the states are governed by a set of probabilities called *transition probabilities*. In a particular state, an outcome or observation can be generated according to the associated probability distribution. It is only the outcome, not the state, that is visible to an external observer, and therefore, states are "hidden" to the outside; hence the name hidden Markov Model. The states might correspond to columns of a DNA multiple alignment or to positions in the three-dimensional structure of a protein. HMM methods have a very rich theoretical structure and use simple modular objects that can be added in various architectures to describe complex models. Churchill proposed an early HMM for genome segmentation. The model consists of two (hidden) states (AT-rich and GC-rich regions), the transition probabilities between these states and the probability distributions of the four bases for each state. Nicolas and collaborators have proposed a new statistical segmentation method based on a hidden Markov model and have estimated parameters using maximum likelihood. Their model has several hidden states, and it enables different compositional classes to be separated, for example, noncoding RNAs, coding regions for each DNA strand (+,−), highly expressed coding regions, genes coding for hydrophobic proteins, and intergenic regions. Their approach is not based on sliding windows, and it enables different genome compositional classes to be separated without prior knowledge of their content, size, and localization.

An effective method for detecting specialized islands is based on wavelet analysis, which is similar to Fourier analysis. In short, a wavelet is an oscillation that decays quickly. A Fourier transform decomposes a function into a sum of sine and cosine functions with different frequencies. The result of a Fourier transform is a series of amplitudes of sines and cosines that sum up to the original function. Instead, wavelet methods use functions that are related to fractals in that the same shapes repeat at different orders of magnitude. The basic idea of wavelets is to decompose a signal, for example, a G + C profile, into several groups of coefficients, each group containing information about features of the G + C profile at a certain scale of sequence length. Coefficients at coarse scales capture gross and global features, for example, clusters of genes or single genes with different G + C content. Coefficients at fine scales contain the local details (few nucleotides) of the profile. Therefore, the original plot can be replaced by a more regular one generated using the coefficients at coarse scales.

3 Detecting Genes

The most important step in the analysis of a new genome is the identification of all its genes. It is important to understand the difference between prokaryotic and eukaryotic genes. In both cases, it is a unit of heredity; however, in eukaryotes this unit may include both the protein-coding region and RNA-coding region of a DNA sequence. In prokaryotes, a gene refers only to the protein-coding region because multiple genes may be expressed from a single RNA molecule (an operon). In some cases, a dozen human genes can span the same length of an entire bacterial chromosome. There are approximately 30 000 genes in the human genome, but it

is important to realize that there may be as many as 300 000 gene products. These are the result of alternative splicing and posttranslational modification.

The standard methodology for gene identification is to find a similar sequence from another organism. This can be done using the BLAST and FASTA families of programs to search all the entries in sequence databases. Since prokaryotic genomes tend to be gene-rich with frequent overlaps, an open problem is the improvement of the accuracy of gene start and termination prediction. "Glimmer" and the "GeneMark" family of programs are the most widely used tools for microbial gene prediction. Structural RNA genes can be detected using QRNA. All these programs are based on HMMs.

Once a putative gene has been identified, its function can be investigated using the COG database, which contains clusters of orthologous groups of proteins, or a pathway database, for example, Ecocyc, which describes biochemical pathways, reactions, and enzymes in *Escherichia coli*. Additional functional information is given by the analysis of cotranscribed genes, that is, genes organized into an operon structure. Although operons can be located through the analysis of their boundaries, that is, promoters on the 5′ end, and terminators on the 3′ end, such approaches can only be effective for organisms whose promoters and terminators are well known. An alternative method finds gene clusters where gene order and orientation is conserved in two or more genomes and estimates the likelihood that such conserved gene sets form operons. The sensitivity of this method is 30 to 50% for the *E. coli* genome. Note that, although positive selection for operons would result in similarly ordered sets of genes across phylogenetically distant genomes, the conservation of gene order between evolutionarily similar organisms may simply reflect the common ancestry of the species.

3.1 Analysis of Duplicated Genes

Genes exist predominantly as families with related structures and functions. In particular, eukaryotic genomes contain a large number of duplicated fragments. There are different extents of duplication: a small part of a gene, a module, an exon, a coding region, a full gene, a cluster, part of a chromosome, or a full chromosome. Whole genome duplications are widely believed to have played an important role in the evolution of the yeast, plant, and vertebrate genomes. The human genome contains roughly the same number of gene families as insects and nematodes, although we have more abundant gene variety in each family. The probability of gene preservation increases with the number of independent subfunctions in the regulatory regions (transcription sites) and in the coding regions (alternative splicing of exons) because of the greater number of ways in which gene duplicates can evolve. The recombination, loss, acquisition, shuffling, and duplication of regulatory sites and exons will increase rather than decrease the probability of duplicate gene fixation because each gene can now perform a function the other gene cannot, for example, the two genes being expressed in different tissues or developmental stages. The probability of fixation by differential subfunctionalization approaches zero in large populations because the long time to fixation magnifies the chances that secondary mutations will completely incapacitate one copy before joint preservation is complete. Therefore, it is more important in high eukaryotes than in bacteria

and low eukaryotes where neofunctionalisation, that is, that completely new genes arise, is a more frequent event.

3.2 Phylogenetic Methods for Orthologous Genes

Phylogenetic methodology allows one to recognize and exploit the statistical dependencies among duplicated genes. For example, there are cases in which gene functions are inferred from pairwise sequence similarities to genes whose functions have themselves been inferred from pairwise sequence similarities. Since paralogous genes (those believed to have arisen by gene duplication within a species) may not have the same function, the phylogenetic relationships among sequences allows each species to contribute its own quantity to all estimates. The most statistically robust approach is to consider the phylogenetic inference problem in a likelihood framework and to use accurate models of evolution. Maximum likelihood estimation chooses the hypothesis that maximizes the likelihood of the data, that is, which renders the data the most plausible. In phylogeny, the model of evolution relates the hypotheses, that is, the tree topologies, to the data, that is the sequences. The maximum likelihood method allows one to estimate evolutionary parameters, such as heterogeneity of rates of evolution among sites and among lineages, transition/transversion and nonsynonymous (amino acid changing; d_N)/synonymous (silent; d_S) substitutions ratios; therefore, important biological questions can be addressed and tested. A great attraction of the likelihood approach is the ability to perform robust statistical hypothesis. Therefore, when more than two genome sequences are available, it is better to use phylogenetic methods than simple sequence pairwise comparisons and it is also better to use maximum likelihood than other methods. Phylogenetic methods also make appropriate corrections for multiple hits. In pairwise comparisons of sequences, the multiple-hit corrections differ for each pair because the evolutionary separation of two human sequences is not the same as the evolutionary separation of human and mouse or mouse and rat. Moreover, contiguous regions might have different evolutionary separations.

Gu has developed a quantitative measure for testing the function divergence within a gene family. The method is based on measuring the decrease in mutation-rate correlation between gene clusters of a gene family; the HMM procedure allows one to identify the amino acid residues responsible for the functional divergence.

Tandemly duplicated genes are known to be subject to gene-conversion events that homogenize their sequences. There are several methods for testing gene conversion among duplicated genes and repeats. Most of these methods are derived from the test proposed by Sawyer. This test allows one to detect identical segments of DNA between clustered genes that may indicate gene-conversion events. Sawyer's gene-conversion test, when applied to pairs of sequences, measures the sum of the squared total lengths (SSL) of these fragments. If gene conversion has been operating, fragment lengths would be more varied and SSL would be larger than if variant sites were distributed at random along the sequence. Therefore, the same number of variant sites is then randomly permuted along the sequence many times and the SSL for each replication calculated and compared to the observed value.

4
Detecting Regulatory Regions

Even though every cell in an organism contains the same genetic material, each cell does not express the same set of genes. Therefore, one of the major problems facing genomic research today is to determine which genes are differentially expressed and under what conditions and how the expression of those genes is regulated. The first step in determining differential gene expression is the binding of sequence-specific DNA-binding proteins (i.e. transcription factors) to regulatory regions of the genes, that is, transcription factor binding sites (TFBSs). Therefore, identification of the regulatory elements is necessary for understanding mechanisms of cellular processes. A significant number of prokaryotic genes and DNA-binding sites are involved in regulatory responses to factors such as growth arrest, cell density, nutrient starvation, and stress. These responses may form part of a complex regulatory network designed to match cellular physiology with optimal growth and/or survival. Regulatory networks are also important in switching on virulence genes in human, animal, and plant pathogens. In eukaryotes, the regulation of gene expression is highly complex and often occurs through the coordinated action of multiple transcription factors. This combinatorial regulation has several advantages; it controls gene expression in response to a variety of signals from the environment and allows the use of a limited number of transcription factors to create many combinations of regulators. DNA-binding sites are shorter than genes and lack the sharp delimitations of start and termination signals. Genes coding for DNA-binding proteins are usually less conserved than genes coding for structural proteins. Moreover, regulatory proteins work in modules. Rajewsky and collaborators have demonstrated an inverse correlation between the rate of evolution of transcription factors and the number of genes that they regulate. Therefore, it may be easier to dissect a large gene network than a small one.

Information-theory-based methods have been successful in identifying bacterial regulatory sites. It is known that specific hydrogen bonds and van der Waals interactions allow a protein to discriminate, that is, to extract a certain amount of information from a DNA-binding site. From information theory, we know that one bit is the amount of information required to distinguish between two equally likely choices: if a protein picks one of the four bases, then it makes a two-bit choice, that is, the first choice is between purine and pyrimidine. Several authors have proposed a likelihood quantity to estimate the amount of information in multiple aligned sequences containing a binding site; they also derived an approximation for *p*-values and variance. Theory predicts that, because of mutations, the information content at a binding site is kept from being larger than is needed for the regulatory function and that the sequence pattern of a binding site is related to the number of genomewide occurrences and to the size of the genome. These methods have the problem of many false positives and an important limitation: they assume that each site interacts with a single protein independently. Additional information on regulatory sites may come from structural studies, for example, bacterial promoters have been found to be curved and rigid.

Although there is now a detailed understanding of regulatory mechanisms in unicellular organisms, such as bacteria and budding and fission yeast, the

extension of such results to multicellular organisms and, in particular, to the human genome presents considerable difficulties. In single-celled organisms, most regulatory elements are located within 200 to 500 bp of the 5′ end of open reading frames (ORFs). In multicellular organisms, however, regulatory elements may be found upstream or downstream of the gene, as well as in introns, and may be spread over tens or even hundreds of kilobases. Experimental evidence suggests that regulatory elements are more probable within a few thousand base pairs 5′ of the transcription start site, but current algorithms are able to locate only some of these regions.

The challenge is to allow biologists to automatically (1) retrieve genes and intergenic regions; (2) identify putative regulatory regions; (3) score sequences for known TFBSs; (4) identify candidate motifs for unknown binding sites; and (5) detect those statistically overrepresented sites that are characteristic for a gene set. In this context, the difficulty for bioinformatics is twofold. First, the approximate localization of regulatory sequences in large anonymous DNA sequences is required. Once those regions are located, the second task is the identification of individual transcriptional control elements and the correlation of a subset of such elements with transcriptional functions. The sequence motifs of individual cis-elements are usually too short and degenerate for confident detection. In most cases, the requirements for organization of cis-elements within these clusters are poorly understood.

How many sites do we expect to find in a genome? Cawley et al. have mapped the binding sites *in vivo* for three DNA-binding transcription factors, Sp1, cMyc, and p53, in a microarray incorporating sequences of human chromosomes 21 and 22. This mapping reveals an unexpectedly large number of TFBS, with a minimal estimate of 12 000 for Sp1, 25 000 for cMyc, and 1600 for p53 when extrapolated to the full genome. Only 22% of these TFBS regions are located at the 5′ termini of protein-coding genes, while 36% lie within or immediately 3′ to well-characterized genes and are significantly correlated with noncoding RNAs. Most methods for identifying regulatory sites rely upon the assumption that the nucleotides of binding sites exert independent effects on binding affinity. Bulyk et al. determined the effects of many possible sequence variations in binding affinities of mutants versus wild-type transcription factor bound to microarrays containing all possible central 3-bp triplet-binding sites. They found that the nucleotides of TFBSs cannot be treated independently. In higher organisms, it is more likely that multiple transcription factors bind to the same transcription control region, and to understand the regulatory content of eukaryotic genomes, it is necessary to consider the co-occurrence and spatial relationships of individual binding sites. Several works suggest that positional information, especially the relative spacing between transcription-factor binding sites, may represent a common organizing principle of transcription control regions. Chiang et al. used the genome sequences of four yeast species of the genus *Saccharomyces* to identify sequences potentially involved in multifactorial control of gene expression. They found 989 pairs of hexameric sequences that are jointly conserved in transcription regulatory regions and that also exhibited nonrandom relative spacing. The incorporation of word pairs to define sequence features yields more specific predictions of average expression profiles and more informative regression models

for genome-wide expression data than on considering sequence conservation alone. King et al. introduced a nonparametric representation of TFBSs that can model arbitrary dependencies between positions. As two parameters are varied, this representation smoothly interpolates between the empirical distribution of binding sites and the standard position-specific scoring matrix (PSSM).

TFBSs detection involves the search of DNA patterns that are overrepresented in the upstream region of a set of coregulated genes. Several computational algorithms have been developed to this end. These methods fall into three broad classes:

1. The word-enumeration approach compares the frequency counts of substrings in the upstream region to some reference set.
2. The probability-based models update a position-specific probability matrix, using local multiple alignment searches. In this approach, the model parameters are estimated using Expectation-maximization (EM) or Gibbs sampling methods.
3. Methods using microarray data.

The major limitation of type-1 methods is that it represents cis-elements using degenerate consensus sequences rather than the more general position specific scoring matrices. An undesirable feature of many methods is the use of arbitrary threshold parameters, such as a window size within which the cis-elements must occur. Then, the choice of a DNA background model that adequately reflects the properties of natural DNA is more problematic to construct. There are usually three techniques: an independent mononucleotide model estimated from the query sequence, a higher-order Markov model estimated from genomic DNA, and a locally varying mononucleotide model estimated from a sliding window within the query sequence.

Djordjevic and collaborators classified potential binding sites by means of the estimate of sequence-specific binding energy of a given transcription factor. The method also estimates the chemical potential of the factor that defines the threshold of binding. This results in a significant improvement in the number of expected false positives, particularly in the ubiquitous case of low-specificity factors.

Comparative genomics seems more promising than single species analysis. In particular, owing to the high similarity of both biology and sequence between human and mouse, the mouse genome is receiving considerable attention as a tool for cross-species comparisons. This strong similarity, however, has raised doubts regarding the general usefulness of human–mouse sequence comparison for distinguishing functionally conserved features against a background of recently evolved sequence.

Loots et al. have developed a computational tool, rVista, that combines clustering of predicted TFBSs and the analysis of interspecies sequence conservation to maximize the identification of functional sites. They analyzed the distribution of several TFBSs across 1 Mb of the well-annotated cytokine gene cluster (Hs5q31; Mm11), focusing on the distribution of all binding sites specific for the transcription factors AP-1, NFAT, and GATA-3. The exploitation of the orthologous human–mouse dataset resulted in the elimination of >95% of the approximately 58 000 binding sites predicted on analysis of the human sequence alone, whereas it identified 88% of the experimentally verified binding sites in this region.

Although comparative genomics of regions (e.g. phylogenetic footprinting)

across different species is a very promising approach, it has been unclear as to which species are most useful and how many genomes are sufficient for comparison. There is the belief that the comparison of just three well-chosen species that have similar physiology, genome size, and whose phylogenetic distances are neither too small nor too large, may allow to correctly identify most of TFBSs. Margulies et al. have identified sequences that are conserved across several multiple species, "Multispecies Conserved Sequences", and have found that approximately 70% bases reside within regions. They used the Bayes block aligner, which focuses on aligning highly conserved, ungapped blocks in which regulatory elements are most likely to reside. The alignment algorithm first produces data represented by a two-dimensional histogram reporting the probability that each pair of nucleotides from two subject sequences are located within a conserved block. The recursive sampling algorithm yields a representative sample of alignments that can be used to calculate the probability that any given base in the first sequence aligns with a specific base in the second sequence. To bias the output towards alignments with short blocks of high similarity, they used the PAM1 similarity matrix, that is, a substitution matrix based on the assumption of only one accepted mutation per 100 bp. In this matrix, the probability of a transition is three times that of a transversion.

Several methods for searching for specific regulatory sites using Markov and hidden Markov models, Bayesian clustering, and Machine learning have been proposed. Qin et al. have developed a Gibbs sampling-based Bayesian motif clustering (BMC) algorithm that show fewer errors than hierarchical and K-means clustering methods.

It is important to mention another important component of the identification of TFBSs: databases of TFBSs and TFs. TRANSFAC is a database about eukaryotic transcription regulating DNA sequence elements and the transcription factors binding to and acting through them. Database of transcriptional start sites (DBTSS) was originally constructed on the basis of a collection of experimentally determined TFBSs of human genes and now covers several species. The positions of single nucleotide polymorphisms (SNPs) in dbSNP were displayed on the upstream regions of contained human genes. The CORG – "Comparative regulatory genomics" – database contains conserved blocks from the upstream regions of orthologous genes. These blocks were computed on the basis of statistically significant local suboptimal alignments of 15 kb regions upstream of the translation start sites of a large number of pairs, 10 793 at April 2004, of orthologous genes. The resulting conserved blocks were annotated with expressed sequence tags (ESTs) that are good indicators for putative exons.

The use of both sequence and gene-expression information is producing important results. Bussemaker et al. have considered both yeast sequences and expression data and have outputs of statistically significant motifs: the regions upstream of genes contribute additively to the log-expression level of a gene. Recently, Conlon et al. have proposed a very interesting approach. They have first identified a number of sequence patterns upstream the yeast genes using Mdscan and have then tested for candidate transcription sites. Their algorithm is based on fitting a linear regression with gene expression as response and a score based on transcription-site candidates as covariates: $Y_g = \alpha + \beta_m S_{mg+}\varepsilon_g$, where Y_g is the

$\log_2$ expression value of the gene g, S_{mg} measures how well the upstream sequence of a gene g matches a motif m, in terms of both degree matching and number of sites, and ε_g is the gene-specific error term. The baseline expression α and the regression coefficient β_m are estimated from the data. Then, they performed stepwise regression starting with a model with intercept-only and adding/removing motifs at each step. The candidate motifs with a significant p-value (say for example, 0.01) for the simple linear regression coefficient β_m are retained and used by the stepwise regression procedure to fit a multiple regression model: $Y_g = \alpha + \sum_{m=1}^{M} \beta_m S_{mg} + \varepsilon_g$. Again, this stepwise regression begins with only the intercept term and at each step, adds the motif that gives the largest reduction in residual error. After adding each new transcription site, the model is checked to remove the ones whose effects have been sufficiently explained by the last added. Hopefully, this class of methods may be used to charting the regulatory circuits of all eukaryotic biochemical and signalling pathways and explaining how specific mutations serve to reprogram the "integrated circuit of the cell" so as to manifest cancer and other diseases.

5
Periodicities and Repeats

The studies of mechanical and structural properties of chromosomes through di- and trinucleotide models of properties, such as propeller twist, stacking energy, and curvature have shown that all bacterial chromosomes sequenced so far present a periodical pattern. Worning and collaborators have recently combined Fourier analysis, correlation, and noise reduction techniques to increase the signal to noise ratio and, therefore, detect weak periodical patterns in bacterial chromosomes. They found a periodicity of 11 bp in eubacterial genomes, while it is of 10 bp for archeal; for comparison, *Saccharomyces cerevisiae* has a dominant period of 10.2 bp. Mixtures of 10 and 11 bp periodicities provide evidence of lateral gene transfer between Archea and *Thermotoga maritima*. Audit and collaborators used wavelets to analyze bending properties of sequences and provided evidence for nucleosome patterns in human chromosomes.

The transitions from prokaryotes to unicellular eukaryotes and to multicellular eukaryotes are associated with an increase in genome size, in the organism's body size and a reduction in population size and density. The changes include gradual increases in gene number, resulting from the retention of duplicate genes, and more abrupt increases in the size and number of introns, low-complexity words, and mobile genetic elements. Simple repeats are rather rare in bacteria and in low eukaryotes and very abundant in high vertebrate genomes. Bacteria and human are at the two extremes, showing orders-of-magnitude difference in genome, body, and population size. The spreading of repeats across a genome occurred as nonadaptive when the organism size increased and the population size decreased. The selection process became less effective in determining genome size, allowing for nonadaptive processes, and more effective in controlling the phenotypic complexity. In vertebrates, if repeats are inserted in the regulatory or coding regions of genes, there is a high chance that metabolic or genetic redundancy will buffer the effect.

Repeats tend to have a GC and dinucleotide composition somewhat different from that of the DNA into which they

are inserted. Although the classification has become rather fictitious, for sake of simplicity, we can make a distinction between single-few copies, middle-, and highly repetitive DNA. Repeats in few copies are usually orthologous genes that may contain hidden repeats translated in the form of runs of amino acids and retroviruses inserted in the genome. For example, the human genome contains more than 50 copies of chemokine receptors that have high sequence similarity. Short-repetitive DNA may be placed in the highly and middle-repetitive categories. The first is formed by tandemly clustered "satellite" DNA of variable-length motifs (5–100 bp) and is present in large islands of up to 100 megabases and may constitute up to 40% of the genome. The middle-repetitive category can be either short islands of tandemly repeated microsatellites/minisatellites ("CA repeats," tri- and tetranucleotide repeats) or mobile genetic elements. Mobile elements include DNA transposons, short interspersed elements (SINEs) and long interspersed elements (LINEs), and processed pseudogenes.

Why should we be interested in repetitive DNA? Tandem repeats with 1 to 3 base motifs can differ in repeat number among individuals; therefore, they are used in population genetics and for the identification of individuals and forensic testing. It is known that trinucleotide repeats are involved in human neurodegenerative diseases (e.g. fragile X and Huntington's disease). DNA repeats increase DNA recombination events and have the potential to destroy (by insertion mutagenesis), to create (by generating functional retropseudogenes), and to empower (by giving old genes new promoters or regulatory signals). Moreover, the instability of short tandemly repeated DNA has been associated to cancer.

Particularly abundant in vertebrate genomes are strings of As and Ts that have several peculiar properties not shared by strings of Gs and Cs. *In vitro* studies have shown that such poly(A) and poly(T) sequences cannot be readily wound around the nucleosome, and this is most likely because these kinds of homopolymeric tracts have a straight and rigid structure.

The published sequence of the human genome, though yet incomplete, demonstrates that coding sequences comprise less than 5% of it, whereas repeating sequences account for at least 50% of the genome and probably even more. Many of the repeats can be found in mature mRNA and total cellular RNA. RNAs containing repetitive elements include Alu-containing mRNAs that amount to 5% of all known mRNAs.

Recently, bioinformatics has shown a link between microsatellite DNA and mobile elements. All mobile elements such as SINEs, LINEs, and processed pseudogenes contain A-rich regions of different lengths. In particular, the Alu elements, present exclusively in the primates, are the most abundant repeat elements in terms of copy number (1 090 000) and account for more than 10% of the human genome. They are typically 300 nucleotides in length and are predominantly present in the noncoding regions. Higher Alu densities were observed in chromosomes with a greater number of genes and vice versa. Alus have a dimeric structure and are ancestrally derived from the gene specifying 7SL RNA, an abundant cytoplasmic component of the signal recognition particle that mediates the translocation of secreted proteins across the endoplasmic reticulum. SINEs amplify using an RNA polymerase III-derived transcript as template, in a process termed *retroposition*. The

mobility is facilitated by a variable-length stretch of an A-rich region at the 3′ end of genomic SINE elements. Although each of the million plus Alu elements present in the human genome have poly(A) stretches, only a few are able to retropose. Therefore, the mere presence of an A stretch is not sufficient to confer on an Alu element the ability to retropose efficiently. The length of the A stretch seems to correlate positively with the mobility of the Alu.

6 Bioinformatics Analysis of Protein Modifications

Prediction and annotation of protein modifications (PMs) is a challenging bioinformatics research area because PMs have been shown to be important features for determining protein function and, since they are often involved in signalling, for systems/organismal biological studies. Main bioinformatics applications on PMs focus on prediction of the site of modifications and on database resources. Many PMs occur at specific, yet variable, motifs in the target proteins and most prediction methods use machine-learning techniques, such as support machines, artificial neural networks and HMM. Examples of prediction tools are: Ser, Thr, and Tyr phosphorylation sites in eukaryotic proteins (NetPhos), tyrosine sulphation sites (for example, Sulfinator), type O-glycosylation sites in mammalian proteins (NetOGlyc), GlcNAc-O, protein sorting signals (PSORT), mitochondrial and plastid targeting sequences (MITOPROT, Predotar, ChloroP), O-beta-GlcNAc attachment sites in eukaryotic protein sequences (YinOYang), and GPI Modification Site Prediction (big-PI Predictor, DGPI). The tendency is towards organism-specific tools, for example, glycosylation sites in *Dictyostelium discoideum* (DictyO-Glyc) and protease cleavage sites in picornaviral proteins (NetPicoRNA). The field is also moving toward prediction methods specific for a class of proteins, for example, specific kinase-specific phosphorylation-site prediction, caspase and propeptide cleavage sites, and tyrosine sulfation sites. Future challenges in prediction will focus on understanding the structural features of PM sites: modifying enzymes must recognize the native three-dimensional structure around an acceptor site, and therefore, structural motifs may be more conserved than sequence motifs. Further developments will be focussed on databases and on proteomics and mass spectrometry techniques. Another important target will be understanding variation in PMs associated with diseases and developmental states; for example, it is known that glycosylation of some proteins varies in alcoholism. What we mean is that we can predict only that a site can be modified but not how likely a modified site can turn into a nonmodified one during a disease or during the developmental stages.

7 Genome Feature Classification: Gene Ontology

The overload of genome data represents a tremendous challenge for organizing and transforming results into knowledge. The Gene Ontology (GO) Consortium was formed in 1998 to provide consistent annotation of genes and gene products, and thus aid cross-species comparisons. It started as collaboration between yeast, fruit fly, worm, and mouse databases, but has since evolved to include many other

projects. The GO consortium's aim is to develop and share structured vocabularies that can be used for cross-species annotation of gene products. Currently, the GO consortium provides ontologies for three independent biological domains: biological process, cellular component, and molecular function.

1. Biological process refers to a biological objective to which the gene or gene product contributes. Some examples of biological process terms are "protein biosynthesis," "meiosis," and "glucose metabolism."
2. Cellular component refers to the part of the cell where the gene product is contained. Examples include such terms as "ribosome," "proteasome," and "mitochondrial inner membrane."
3. Molecular function is defined as the biochemical activity of a gene product. Some examples of molecular function terms are "enzyme," "transporter," and "heat shock protein."

All three ontologies are accessible at http://www.geneontology.org/. The GO Consortium also hosts annotation projects for several model organisms, including fission and budding yeasts. Gene ontology terms and relationships have been represented in the form of directed acyclic graphs (DAG). Compared with a hierarchical tree, DAG has the advantage of allowing a more specific child term to have several parent terms. There are two types of child-to-parent relationships: "is a" and "part of." The "is a" relationship is used when a child represents a type of parent term, for example, cytosolic ribosome is a ribosome. The "part of" relationship is used when a child term is a component of the parent term, for example, cytosolic ribosome is part of cytosol. The one-to-many relationships between a gene product and ontology categories reflect the fact that a gene product may have several functions, may participate in different processes, and may interact with various proteins.

8
Limitations of Sequence Analysis

The genome is only a source of information. In order to function, it must be expressed. Differences between organisms are more related to gene expression and protein abundances than to DNA sequence. For example, chimp DNA is 98.7% identical with human DNA. If just the genes are compared, the similarity increases to 99.2%. Gene-expression levels in blood and liver of the human and chimp are quite similar but very different in the brain tissues. At the same time, many physiological variables show rhythmic changes along a time period, for example, a day, a week, a month, and a year, and it is also known that several acute diseases, such as renal colic attacks and cardiovascular diseases show a rhythmic variability as well. Therefore, future efforts will be focussed on integrating sequence data with gene expression and proteomics analysis.

See also Mass Spectrometry, High Speed DNA Fragment Sizing.

Bibliography

Books and Reviews

Baldi, P., Brunak, S. (2001) *Bioinformatics: the Machine Learning Approach*, 2nd edition. MIT Press, Boston, Mass., USA.

Durbin, R., Eddy, S.R., Krogh, A., Mitchinson, G. (1998) *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge, UK.

Eisenhaber, F., Eisenhaber, B., Maurer-Stroh, S. (2003) Prediction of Post-translational modifications from amino acid sequence: Problems, pitfalls, methodological hints. In Andrade, M.M. (Ed.) *Bioinformatics and Genomes: Current Perspectives*. Horizon Scientific Press, Wymondham, pp. 81–105.

Kanehisa, M. (2000) *Post-Genome Informatics*. Oxford University Press, Oxford, UK.

Swofford, D.L. (1996) Phylogenetic inference, in Hillis, D.M. (Ed.) *Molecular Systematics*, Sinauer Associates, pp. 407–514.

Warren, J., Ewens, J., Grant, G.R. (2001), *Statistical Methods in Bioinformatics: An Introduction*, Springer, New York.

Primary Literature

Audit, B., Vaillant, C., Arneodo, A., d'Aubenton-Carafa, Y., Thermes, C. (2002) Long-range correlations between DNA bending sites: relation to the structure and dynamics of nucleosomes. *J. Mol. Biol.* **316**, 903–918.

Bernardi, G. (2000) Isochores and the evolutionary genomics of vertebrates, *Gene* **241**, 3–17.

Besemer, J., Lomsadze, A., Borodovsky, M. (2001) GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions, *Nucleic Acids Res.* **29**, 2607–2618.

Blom, N., Gammeltoft, S., Brunak, S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites, *J Mol. Biol.* **294**(5), 1351–1362.

Bulyk, M.L., Johnson, P.L., Church, G.M. (2002) Nucleotides of transcription factor binding sites exert interdependent effects on the binding affinities of transcription factors, *Nucleic Acids Res.* **30**, 1255–1261.

Bussemaker, H.J., Li, H., Siggia, E.D. (2001) Regulatory element detection using correlation with expression, *Nat. Genet.* **27**, 167–171.

Campbell, A., Mrazek, J., Karlin, S. (1999) Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA, *Proc. Natl Acad. Sci. U.S.A.* **96**, 9184–9189.

Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., Wheeler, R., Wong, B., Drenkow, J., Yamanaka, M., Patel, S., Brubaker, S., Tammana, H., Helt, G., Struhl, K., Gingeras, T.R. (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs, *Cell* **116**, 499–509.

Chiang, D.Y., Moses, A.M., Kellis, M., Lander, E.S., Eisen, M.B. (2003) Phylogenetically and spatially conserved word pairs associated with gene-expression changes in yeasts, *Genome Biol.* **4**, R43.

Churchill, G.A. (1989) Stochastic models for heterogeneous DNA sequences, *Bull. Math. Biol.* **51**, 79–94.

Conlon, E., Liu, X.S., Lieb, J.D., Liu, J. (2003) Integrating sequence motif discovery and microarray analysis, *Proc. Natl Acad. Sci. USA* **100**, 3339–3344.

Delcher, A.L., Harmon, D., Kasif, S., White, O., Salzberg, S.L. (1999) Improved microbial gene identification with Glimmer, *Nucleic Acids Res.* **27**, 4636–4641.

Dermitazakis, E.T., Clark, A.G. (2001) Differential selection after duplication in mammalian developmental genes, *Mol. Biol. Evol.* **18**(4), 557–562.

Dieterich, C., Wang, H., Rateitschak, K., Luz, H., Vingron, M. (2003) CORG: a database for COmparative Regulatory, *Genomics Nucleic Acids Res.* **31**, 55–57.

Djordjevic, M., Sengupta, A.M., Shraiman, B.I. (2003) A biophysical approach to transcription factor binding site discovery, *Genome Res.* **13**, 2381–2390.

Down, T.A., Hubbard, T.J. (2002) Computational detection and location of transcription start sites in mammalian genomic DNA, *Genome Res.* **12**, 458–461.

Ellrott, K., Yang, C., Sladek, F.M., Jiang, T. (2002) Identifying transcription factor binding sites through Markov chain optimization. *Bioinformatics*, **18**(Suppl. 2), S1–S10.

Eyre-Walker, A., Hurst, L.D. (2001) The evolution of isochores, *Nat. Rev. Genet.* **2**, 549–555.

Frith, M.C., Hansen, U., Weng, Z. (2001) Detection of cis-element clusters in higher eukaryotic DNA, *Bioinformatics* **17**, 878–889.

Fryxell, K.J., Zuckerkandl, E. (2000) Cytosine deamination plays a primary role in the evolution of mammalian isochores, *Mol. Biol. Evol.* **17**, 1371–1383.

Ermolaeva, M.D., White, O., Salzberg, S.L. (2001) Prediction of operons in microbial genomes, *Nucleic Acids Res.* **29**, 1216–1221.

Ermolaeva, M.D., Hanif, G.K., White, O., Smith, H.O., Salzberg, S.L. (2000) Prediction of transcription terminators in bacterial genomes, *J. Mol. Biol.* **301**, 27–33.

Garavelli, J.S. (2003) The RESID database of protein modifications: 2003 developments, *Nucleic Acids Res.* **31**, 499–501. (PMID:12520062).

Gu, X. (2001) Maximum likelihood approach for gene family evolution under functional divergence, *Mol. Biol. Evol.* **18**, 453–464.

Karlin, S., Burge, C. (1995) Dinucleotide relative abundance extremes: a genomic signature, *Trends Genet.* **11**, 283–290.

Karlin, S., Campbell, A.M., Mrazek, J. (1998) Comparative DNA analysis across diverse genomes, *Ann. Rev. Genet.* **32**, 185–225.

Karp, P.D., Riley, M., Saier, M., Paulsen, I.T., Paley, S., Pellegrini-Toole, A. (2002) The Ecocyc database, *Nucleic Acids Res.* **30**, 56.

King, O.D., Roth, F.P. (2003) A non-parametric model for transcription factor binding sites, *Nucleic Acids Res.* **31**, e116.

Kiryutin, B., Galperin, M.Y., Fedorova, N.D., Koonin, E.V. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes, *Nucleic Acids Res.* **29**, 22–28.

Lynch, M., Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes, *Science* **290**, 1151–1155.

Lynch, M., Force, A. (2000) The probability of duplicate gene preservation by subfunctionalization, *Genetics* **154**, 459–473.

Liò, P., Vannucci, M. (2000) Finding pathogenicity islands and gene transfer events in genome data, *Bioinformatics* **16**, 932–940.

Liu, X., Clarke, N.D. (2002) Rationalization of gene regulation by a eukaryotic transcription factor: calculation of regulatory region occupancy from predicted binding affinities, *J. Mol. Biol.* **323**, 1–8.

Liu, J.S., Lawrence, C.E. (1999) Bayesian inference on biopolymer models, *Bioinformatics* **15**, 38–52.

Liu, J.S., Neuwald, A.F., Lawrence, C.E. (1995) Bayesian models for multiple local sequence alignment and Gibbs sampling strategies, *J. Am. Stat. Assoc.* **90**, 1156–1170.

Loots, G.G., Ovcharenko, I., Pachter, L., Dubchak, I., Rubin, E.M. (2002) rVISTA for comparative sequence-based discovery of functional transcription factor binding sites, *Genome Res.* **12**, 832–839.

Margulies, E.H., Blanchette, M., Haussler, D., Green, E.D. (2003) NISC comparative sequencing program. Identification and characterization of multi-species conserved sequences, *Genome Res.* **13**, 2507–2518.

McCue, L.A., Thompson, W., Carmack, C.S., Lawrence, C.E. (2002) Factors influencing the identification of transcription factor binding sites by cross-species comparison, *Genome Res.* **12**, 1523–1532.

McLysaght, A., Hokamp, K., Wolfe, K.H. (2002) Extensive genomic duplication during early chordate evolution, *Nat. Genet.* **31**, 128–129.

Messeguer, X., Escudero, R., Farre, D., Nunez, O., Martinez, J., Alba, M.M. (2002) PROMO: detection of known transcription regulatory elements using species-tailored searches, *Bioinformatics* **18**, 333–334.

Nakai, K. (2001) Review: prediction of in vivo fates of proteins in the era of genomics and proteomics, *J. Struct. Biol.* **134**, 103–116. (PMID:11551173).

Nekrutenko, A., Li, W.H. (2000) Assessment of compositional heterogeneity within and between eukaryotic genomes, *Genome Res.* **10**, 1986–1995.

Nelson, H.C., Finch, J.T., Luisi, B.F., Klug, A. (1987) The structure of an oligo(dA) oligo(dT) tract and its biological implications, *Nature* **330**, 221–226.

Nicolas, P., Bize, L., Muri, F., Hoebeke, M., Rodolphe, F.D., Ehrlich, S., Prum, B., Bessières, P. (2002) Mining *Bacillus subtilis* chromosome heterogeneities using hidden Markov models, *Nucleic Acids Res.* **30**, 1418–1426.

Nielsen, R., Yang, Z. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene, *Genetics* **148**, 929–936.

Pedersen, A.G., Jensen, L.J., Brunak, S., Staerfeldt, H.H., Ussery, D.W. (2000) A DNA structural atlas for Escherichia coli, *J. Mol. Biol.* **299**, 907–930.

Qin, Z.S., McCue, L.A., Thompson, W., Mayerhofer, L., Lawrence, C.E., Liu, J.S. (2003) Identification of co-regulated genes through

Bayesian clustering of predicted regulatory binding sites, *Nat. Biotechnol.* **21**, 435–439.

Rajewsky, N., Socci, N.D., Zapotocky, M., Siggia, E.D. (2002) The evolution of DNA regulatory regions for proteo-gamma bacteria by interspecies comparisons, *Genome Res.* **12**, 298–308.

Reinert, G., Schbath, S., Waterman, M.S. (2000) Probabilistic and statistical properties of words: an overview, *J. Comput. Biol.* **1–2**, 1–46.

Rivas, E., Klein, R.J., Jones, T.A., Eddy, S.R. (2001) Computational identification of noncoding RNAs in E. coli by comparative genomics, *Curr. Biol.* **11**, 1369–1373.

Roy-Engel, A.M., Salem, A.H., Oyeniran, O.O., Deininger, L., Hedges, D.J., Kilroy, G.E., Batzer, M.A., Deininger, P.L. (2002) Active Alu element ''A-tails'': size does matter, *Genome Res.* **12**, 1333–1344.

Sawyer, S. (1989) Statistical tests for detecting gene conversion, *Mol. Biol. Evol.* **6**, 526–538.

Schneider, T.D. (2000) Evolution of biological information, *Nucleic Acids Res.* **28**, 2794–2799.

Stormo, G.D. (2000) DNA binding sites: representation and discovery, *Bioinformatics* **16**, 16–23.

Stormo, G.D., Fields, D.S. (1998) Specificity, free energy and information content in protein-DNA interactions, *Trends Biochem. Sci.* **23**, 109–113.

Suzuki, Y., Yamashita, R., Sugano, S., Nakai, K. (2004) DBTSS, DataBase of Transcriptional Start Sites: progress report 2004, *Nucleic Acids Res.* 2004 Jan; **32** Database issue: D78–D81.

Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M. (1996) Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B. (Eds.) *Molecular Systematics* (2nd ed.), Sinauer Associates, Sunderland, Massachusetts, pp. 407–514.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Suzek, B.E., Ermolaeva, M.D., Schreiber, M., Salzberg, S.L. (2001) A probabilistic method for identifying start codons in bacterial genomes, *Bioinformatics* **17**, 1123–1130.

Thanka Christlet, T.H., Veluraja, K. (2001) Database analysis of O-glycosylation sites in proteins, *Biophys. J.* **80**(2), 952–960.

Wagner, A. (1998) The fate of duplicated genes: loss or diversification? *BioEssays* **20**, 785–788.

Wagner, A. (1999) Genes regulated cooperatively by one or more transcription factors and their identification in whole eukaryotic genomes, *Bioinformatics* **15**, 776–784.

Wagner, A. (2001) Birth and death of duplicated genes in completely sequenced eukaryotes, *Trends Genet.* **17**, 237–239.

Werner, T. (1999) Models for prediction and recognition of eukaryotic promoters, *Mamm. Genome* **10**, 168–175.

Werner, T. (2000) Identification and functional modelling of DNA sequence elements of transcription, *Brief Bioinform.* **1**, 372–380.

Whelan, S., Liò, P., Goldman, N. (2001) Molecular phylogenetics: State-of-art methods for looking into the past, *Trends Genet.* **17**, 262–272.

Wingender, E., Dietze, P., Karas, H., Knüppel, R. (1996) TRANSFAC: A database on transcription factors and their DNA binding sites, *Nucleic Acids Res.* **24**, 238–241.

Wiuf, C., Christensen, T., Hein, J. (2001) A simulation study of the reliability of recombination detection methods, *Mol. Biol. Evol.* **18**, 1929–1939.

Worning, P., Jensen, L.J., Nelson, K.E., Brunak, S., Ussery, D.W. (2000) Structural analysis of DNA sequence: evidence for lateral gene transfer in Thermotoga maritima, *Nucleic Acids Res.* **28**, 706–709.

Nucleic Acid and Protein Single Molecule Detection and Characterization

Karl Otto Greulich
Institute of Molecular Biotechnology, Jena, Germany

1 Single Molecule Research: A New Era of Molecular Cell Biology and Molecular Medicine 271
1.1 Books and Journal Issues on Single Molecules 271
1.2 Early Seminal Work and Some Recent Reviews 271

2 Techniques for Detection and Preparation 272
2.1 Conventional Far-Field Microscopy: Spatial Resolution is not Required but Temporal Resolution is High 272
2.2 Some Remarks on Dyes and Detectors 272
2.3 FRET: Foerster/Fluorescence Resonance Energy Transfer 273
2.4 TIRF: Total Internal Reflection Fluorescence-excitation 274
2.5 FCS: Fluorescence Correlation Spectroscopy 274
2.6 Near-Field Microscopy (Scanning Tunneling Microscopy, SNOM, AFM) 275
2.7 Manipulation with Nanometer Accuracy Laser Micromanipulation for Handling of Individual Macromolecules 276

3 Proteins 276
3.1 Spectroscopy on Single Biomolecules 276
3.2 Force Spectroscopy on Selectin, Titin, and von Willebrandt Factor 277
3.3 Single Molecule Immunology: Antibodies and Antigens 278
3.4 Molecular Motors 278
3.5 Visualizing the Action of Individual Dynein and Myosin V Molecules in Real Time 280
3.6 Single Molecule Enzyme Kinetics 280
3.7 Single Protein Molecules in Viruses and in Living Cells 281

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

4 RNA and DNA: Highly Individual Molecules 282
4.1 Individual DNA Molecules as Construction Material and Electric Wires 282
4.2 Stretching and Folding RNA, DNA, and Chromatin 283
4.3 DNA Melting and Zipping: Separating and Joining the Two Strands 283
4.4 Replication, Transcription DNA Repair and DNA Mapping 284
4.5 Single Molecule DNA Sequencing 285
4.6 DNA Analysis by Mapping and Restriction Fragment Sizing 286

5 Summary and Outlook 288

Acknowledgment 288

Bibliography 288
Books and Reviews 288
Primary Literature 289

Keywords

DNA

Proteins

Single Molecule Biochemistry
Studies on reactions of individual molecules with each other.

Single Molecule Biophysics
Testing and handling individual molecules by spectroscopy and by nanomechanical tools.

Techniques for Single Molecule Studies
Highly sensitive techniques for handling single molecules, such as atomic force microscope tips and optical tweezers, and techniques to observe single molecules, such as optical tweezers.

Single molecule research is progressing from pure single molecule detection (SMD), for example, by scanning microscopy or by high-resolution spectroscopy, toward true single molecule chemistry, biochemistry, or molecular biology. The last few years have seen a shift from small chromophores to (biological) macromolecules. Individual protein molecules vary in structural detail and thus may have different functional properties, even if they are, in bulk, assumed to represent the same

protein. The molecule with the most pronounced individuality, however, is DNA. Theoretically, one would need oligonucleotides only 17 bp in length to assign an own sequence to each human individual. The human genome has a length of 3 billion base pairs. Such an essentially infinite variability can only be studied on a single molecule basis.

Therefore, the present review article tries to give an overview on techniques for studying biophysics and biochemistry of proteins (in Sect. 3) and DNA (in Sect. 4) without aiming to be comprehensive. In most cases, recent examples of work from representative author groups are given. Patch clamp studies, for which Erwin Neher and Bert Sakmann won the 1991 Nobel prize, will not be treated in the present overview since this field would justify a separate review.

1 Single Molecule Research: A New Era of Molecular Cell Biology and Molecular Medicine

1.1 Books and Journal Issues on Single Molecules

It is certainly just a coincidence that the dawn of a new millennium is also the dawn of a new chemistry and biochemistry. The ultimate limits in sensitivity have been achieved. Single molecule chemistry and biochemistry are now a reality. The end of the last millennium has witnessed the development of techniques, which allowed detection of single molecules in a solution or at the surface of, for example, a microscope slide. By the end of 2002, in addition to a number of reviews, a few special issues of regular journals and two books on single molecules were available in literature. The first years of the new millennium now witness the development and use of techniques to directly measure and even *modify physical properties* of individual molecules and to directly *observe reactions* of such single molecules.

1.2 Early Seminal Work and Some Recent Reviews

Surprisingly, one of the first, if not the first, single molecule experiments was not on small chromophores but on an enzyme, and it even measured reactions: It was the study on a single β-galactosidase molecule held in a microdroplet. Later, individual molecules could be manipulated with the tips of scanning tunnel or force microscopes. These techniques did not yet allow spectroscopy, since they were monochromic. Color, and thus spectroscopy, in the sense as it is typically used by the analytical chemist was not yet widely used. This problem was alleviated with the advent of the near-field scanning optical microscope and then it was realized that conventional light microscopy was also suitable for single molecule research. Then it became possible to use wavelengths of an individual molecule known from bulk analytical or physical chemistry. The 1990s were then the decade of high-resolution single-molecule spectroscopy, pushed forward particularly by the groups of Orrit, Basche, and Moerner. This development has been recently

described in several reviews. One result of such studies showed that individual small molecules behave as predicted from bulk experiments. A different situation was expected for biomolecules: since theoretical knowledge on protein structures had predicted that all individual enzymes of a species would differ slightly but distinctly from each other, it was no surprise that they revealed different reaction rates. A review on work of the type just described is essentially a precursor of the present paper and a few paragraphs from there are reproduced here. A review on its uses in cell biology has been provided by Sako, one with emphasis on single biomolecular motors is available from Ishii, and an overview of single molecule enzyme kinetics has been given by Xie.

2
Techniques for Detection and Preparation

Conventional fluorescence microscopy is surprisingly suitable for single molecule research. However, most of the other techniques required had to be specifically developed or adapted to this task. Section 2 gives a glimpse of a number of these techniques. Some of the descriptions may be somewhat oversimplified or describe only special cases of applications. Therefore, in order to fully understand these techniques in detail, the reader is referred to the specific literature.

2.1
Conventional Far-Field Microscopy: Spatial Resolution is not Required but Temporal Resolution is High

One of the biggest surprises in single molecule research was certainly the finding that normal – though high-quality – fluorescence microscopes could be used for single molecule analysis. Single molecule detection was thought to contradict the Abbe criterion, which states that the resolution of a conventional far-field microscope is not better than 50% of the wavelength of the light used for illumination (0.5 μm for green light). This was too pessimistic since resolution was mixed up with reproducibility of localization. When the molecules were prepared with sufficiently small concentration onto a surface, they had an average distance of several micrometers from each other: they had no longer to be spatially resolved but had just to be visualized. Visualization of individual molecules is possible down to sizes of a few nanometers. The price one has to pay is that the molecules appear too large, but for many applications, this drawback is acceptable. The conventional fluorescence microscope has two significant advantages, for example, compared to the high-resolution near-field techniques described in Sect. 2.6. It is fast, that is, dynamics of molecules and reactions becomes experimentally accessible and preparation techniques can be easily adapted from bulk analytical methods. Even the imaging by lifetimes, not by intensity, of single molecules is possible.

2.2
Some Remarks on Dyes and Detectors

This may be one reason why conventional far-field microscopy has got its chance in single molecule analysis. Here, light is the carrier of information and comparably gentle interactions can be used to study individual molecules. A disadvantage is that only molecules with a significant intrinsic fluorescence yield or

nonfluorescing molecules after extrinsic labeling with a fluorescence dye can be investigated. Among biomolecules, NADH, flavines, and, in the ultraviolet, proteins occurring in biological cells can be studied directly. For nonfluorescing biological macromolecules, a large number of extrinsic fluorescing dyes, that is, dyes that have somehow to be attached to the macromolecule of interest are available. Many of them are known not to or only moderately to modify the macromolecule's function. Detecting single molecules by a fluorescence microscope is not so much a problem of the number of emitted photons. Since a molecule can, in principle, emit a light quantum every 10 ns, up to 100 million quanta might be expected from a single molecule. Two problems hamper this process. First, a dye molecule will decay after a few hundred up to a million cycles of excitation and emission into nonfluorescing derivative. Second, at dilutions required for single molecule analysis, more than 99.99% of the photons are not absorbed, but may be scattered. In principle, such noninformative scattered exciting light can be separated from the informative fluorescence light by optical filters. Since the latter are not ideal, a small fraction of exciting light reaches the detector, which should only detect fluorescence from the molecule under study. Part of these problems may be overcome by replacing classical fluorescence dyes by nanocrystals or quantum dots. These are nanometer-sized clusters of, for example, zinc sulfide–capped cadmium selenide. Their absorption and emission behavior is solely determined by their size: small clusters are blue, larger ones are red. The advantage of these quantum dots is their very good absorption and emission behavior, which alleviates to some extent the problem just described. A disadvantage is their bulkiness.

When a fluorescence microscope is equipped with a standard TV-CCD camera, one can obtain 25 images per second. This corresponds to a temporal resolution of 40 ms. Special cameras are much faster; millisecond resolution is easily available when one is willing to go a step beyond standard CCD recording. For low-level light, as in single molecule research, cooled CCD cameras integrating over several image frames can be used, but then the temporal resolution is decreased correspondingly. When extreme sensitivities at high temporal resolution are required, single photon-counting cameras are available. These are, in principle, arrays of microphoto multipliers, often with 250 × 250 pixel elements. In almost all cases, the reduced (as compared to standard CCD cameras with 700 × 500 pixels) image size is completely sufficient, since single molecule observation requires only parts of a full visual field in a light microscope. Color information can be added by using a true color camera. Usually, they do not have the sensitivity described above. More efficient is the use of optical filters: fast-switching filter wheels and computer programs for exactly overlaying the different colors that are available on the market.

2.3 FRET: Foerster/Fluorescence Resonance Energy Transfer

FRET (fluorescence resonance energy transfer) can increase the resolution beyond the diffraction limit of conventional microscopy. Here, a pair of fluorescent molecules is used to measure distances of the order of 10 nm. One molecule of this pair, the donor, absorbs light and usually reemits it as fluorescence at a

somewhat larger wavelength. If, however, the second molecule, the acceptor – with an absorption maximum close to the emission maximum of the donor – comes closer to the donor than the "Foerster" distance, the absorbed energy in the donor is transmitted to the acceptor, which finally emits light at an even larger (more red) wavelength. Thus, a red shift heralds the mutual approaching of the donor/acceptor pair to a distance closer than the Foerster distance, which, for example, is approximately 10 nm for the dye pair Cy3/Cy5. Since typical diameters of proteins are between 5 and 15 nm, FRET experiments with this donor/acceptor pair can be used to test whether a protein molecule is folded or denatured, that is, to study protein folding on a single molecule basis. Cy3/Cy5 or other donor/acceptor pairs can also be used to study protein–protein or protein–DNA interactions as well as RNA folding on a single molecule basis.

2.4 TIRF: Total Internal Reflection Fluorescence-excitation

It is quite obvious that ideal preparations for single molecule studies should be very pure. Unfortunately, one often obtains more impurities than molecules of interest. There are, however, techniques available, which allow to focus the view only on the wanted molecules: one of them is total internal reflection. Light falling from the glass side at a very flat angle onto a glass–water interface does not leave the glass but is reflected. However, when a molecule is located directly on top of the glass, the molecule can be excited to fluoresce. Molecules and impurities at a distance of more than a wavelength are not affected. Thus, it is possible to study single molecules in a thin layer above a microscope slide without illuminating all the impurity molecules in the solution. Conventional fluorescence microscopy and TIRF total internal reflection fluorescence microscopy in single molecule research have been compared by Paige et al.

2.5 FCS: Fluorescence Correlation Spectroscopy

While TIRF confines the view into a thin two-dimensional space, FCS, fluorescence correlation spectroscopy illuminates a diffraction-limited three-dimensional volume of interest just by focusing a laser into a solution containing the highly diluted molecules to be studied. FCS requires extreme dilution of the molecules and simultaneously extreme miniaturization of the observation volume to less than a femtoliter. Basically, in such experiments, the velocity of a molecule migrating through the tiny observation volume is measured. One detects a single molecule in a given volume, and after a preset time of waiting, one looks if it is still there. Slow particles will be found several times, fast particles will be found only two or three times. Mathematically, this is described by the correlation function

$$G(t) = \text{Sum over } I(t) * I(t+T),$$

where T is the waiting time.

$G(t)$ is related to the speed of the molecule that can be used to calculate the diffusion constant of the molecule. The latter can be used to estimate the molecular weight. Such experiments are often used to measure the binding of two individual

molecules. This is possible when the two binding partners differ by at least a factor of 8–10 in size, for example, the binding of an antigen to an antibody or the hybridization of a small DNA or RNA oligonucleotide to a long DNA or RNA molecule. Binding sites of single glucocorticoid molecules on cell membranes have been studied by Maier. In such experiments, the smaller binding partner is made fluorescent. One can see the small partner as a fast diffusing object and the complex of a small with a large partner as a slow object. The correlation analysis does not only allow to obtain the absolute migrating velocities but also the ratio of free to bound molecules, that is, the binding constants according to the mass action law. The aforementioned size ratio of 8–10 is sometimes a drawback of FCS. This has, however, been overcome by two color techniques, occasionally even combined with FRET.

2.6 Near-Field Microscopy (Scanning Tunneling Microscopy, SNOM, AFM)

Extremely high resolution can be achieved with near-field techniques. All the near-field microscopes work according to a common basic principle: A sort of ultra fine stylus scans an object line by line and subsequently assembles an image from the data it has obtained. This data assembly takes some time and therefore the temporal resolution of these microscopes is limited to seconds or below seconds (see, however, Sect. 3.5). In turn, the spatial resolution is better than with any other tool available to the analytical chemist or biochemist. The reason for this is that the stylus comes very close to the object under observation. It is the same principle as the stethoscope of a medical doctor: For example, acoustic waves of a beating heart with wavelengths of meters can be exploited to localize the heart with an accuracy of centimetres. Therefore, the optical version of the near-field microscopes, the SNOM (scanning near-field optical microscope) has also been also called *optical stethoscope*. It has a typical spatial resolution of 60 to 80 nm and can be used as a highly localized spectroscope. The second near-field microscope used in the analysis of single biomolecules is the atomic force microscope, AFM. Here, a mixture of van der Waals, electrostatic, and other forces are used to detect the molecule. Since these are all short-range forces, one has to get closer to the molecule under observation and the spatial resolution with the AFM is better than with the SNOM, for ideal (hard) objects in the subnanometer range. Even the combination of an AFM with a confocal laser scanning microscope is available. The near-field microscope with the best resolution is the scanning tunneling microscope, (STM), which, under favorable conditions, can detect submolecular structures. It needs electrically conducting objects and is therefore not generally suitable for single biomolecule studies. Interestingly, the most successful use of near-field microscopes is meanwhile not imaging, but manipulation. The tip of an atomic force microscope is used as an extremely fine manipulator for handling individual molecules and for measuring forces in macromolecules such as proteins, RNA, and DNA while stretching the latter. If one exerts such a force on molecules bound to each other, force-extension curves can be obtained, which show a break at some specific force. This force, however, is not sharply defined. In fact, one obtains a probability curve with a most probable force of a few hundred piconewtons and a full width at half maximum of 50–100 pN. AFM tips can now

be prepared with an individual atom as the active tip. Sections 3.2, 3.3, and 4.3 give a detailed account of AFM tip experiments with biomolecules.

2.7
Manipulation with Nanometer Accuracy Laser Micromanipulation for Handling of Individual Macromolecules

The tip of an AFM is best suited to apply and to measure forces from a few hundred piconewtons up to a nanonewton (for comparison: millions of such AFM tips would be required to lift 100 mg at the surface of the earth. If gentler handling or sensitive force measurement is required, optical tweezers are the tools of choice. Optical tweezers are made of a continuous (often infrared) laser of moderate power, which is focused into a microscope to the theoretical (diffraction) limit. Light pressure and gradient forces push and pull microscopic objects in the focus of the laser. By moving the focus, these objects can be moved along complex trajectories. Since only forces exerted by light are acting, no sticking of individual molecules to microtools such as AFM tips or disruption by the latter occurs. As with AFM tips, mechanical properties of biomolecules can be measured with optical tweezers. Optical tweezers are also used as stylus, similarly as in the near-field microscopy, to detect force landscapes on the surfaces of soft objects such as cells. This technique has been dubbed "Photonic Force Microscopy."

3
Proteins

For some biological macromolecules, the individual differences are not only just interesting for reasons of basic research. Such differences may be important for health and disease. For example, the malfunction of an enzyme may cause irregularities. Such a malfunction may develop by a sort of evolution and may only gradually turn into a true disease form. Early detection of such changes on the single molecule basis will probably allow early diagnostics and thus, may help prevent the outbreak of a disease.

3.1
Spectroscopy on Single Biomolecules

An obvious step toward single molecule research is just to increase the sensitivity down to the single molecule limit, even with Raman spectroscopy when surface enhancement is used. One protein molecule of particular interest is the green fluorescent protein GFP, since it's mRNA can be fused with the mRNA of another protein. After *in vitro* translation, the combined protein often preserves the function of the fused protein, but is now fluorescing and its action, for example, inside a living cell, can be studied. Therefore, the spectral properties of GFP are of special interest and have been studied correspondingly on a single molecule basis. An interesting class of proteins are light harvesting molecules that convert the energy of light into chemical energy. Complex energy conversion processes have now been understood on a single molecule basis. The most straightforward study of macromolecular structure is possible if these molecules change their spectroscopic behavior as a function of conformation. Such a molecule is the (nonbiological) MEH-PPV, a conjugated phenylene-vinylene polymer.Huser et al. could demonstrate that single MEH-PPV molecules have a conformational memory when cast from toluene, but not

when cast from chloroform, where they behave as multichromophore systems. Unfortunately, biomolecules do not reveal such an intrinsic structure–spectrum relationship and here dye molecules have to be attached for folding studies. FRET (see Sect. 2.3) may be the technique of choice. While a number of proteins, such as crystalline from the eye lens, spontaneously fold into their correct three-dimensional structure, many others need the support from partner proteins called *chaperones*. GroES and GroEL are such proteins, and their interaction is essential for protein folding in the bacteria *Escherichia coli*. Using Cy3 and Cy5 as FRET partners, Taguchi et al. could show that both molecules bind with kinetic constants known from bulk experiments. However, only the single molecule experiments could reveal that the two molecules remain in contact for approximately 3 s. This is the time a target protein has left in order to fold into its correct structure.

3.2
Force Spectroscopy on Selectin, Titin, and von Willebrandt Factor

An even more detailed insight into individual macromolecules is possible with force spectroscopy. Here, the tip of an atomic force microscope is used as a nanomechanical tool. One end of the molecule is attached to a surface, the other end to the AFM tip. Then the tip bound end is pulled and the tip bound end is pulled and the force that is required to pull the molecule apart is measured. Also, the force acting between two molecules can be studied. This technique has been applied to many molecules; one among them is DNA, which will be discussed later. One example is P-selectin-ligand binding, a molecule pair that mediates adhesive interactions in immune recognition and also in metastasis of several cancer types. Forces up to 175 pN, that is, one-fifth of a covalent bond, can act. The rate constant for unbinding may vary from $0.2\ s^{-1}$ to $2.8\ s^{-1}$. Such a pronounced difference explains different adhesion behavior of the cells that use P-selectin. A molecule with similar function is fibronectin. Its mechanical properties are critical to the elasticity of the extracellular matrix and connective tissue. Oberhauser et al. have found an extreme extensibility of this molecule. The molecule has a hierarchical force profile, indicating that it has different building blocks with different mechanical stability, particularly its FNIII region.

The von Willebrandt factor is a protein involved in many aspects of protein clotting in blood. Occasionally, it comes in ultralarge aggregates (ULVWF), particularly in catastrophic microangiopathic disorders. Using optical tweezers, Arya et al. could show that specific conformations, and not so much the size of the aggregates, is critical for the interaction of ULVWF with platelets – cells that mediate the generation of thrombi in blood.

Titin, a tandem array of fibronectin and a sort of immunoglobulin, is anchoring the motor protein myosin of muscle to elements of the skeleton. If a muscle is slightly overstretched, titin dampens the forces and reversibly unfolds. Only after strong over–stretching, the muscle is hurt. Owing to this importance, titin is among the molecules best investigated by single molecule force spectroscopy. Just to mention one work, Grama et al. have labelled titin with the fluorescing rhodamine, which, owing to self quenching when two rhodamines are approaching each other, act similarly as the donor/acceptor pairs described earlier. Surprisingly, chemical denaturation

of titin leaves the titin molecule intact, perhaps since acidified muscles need to keep their mechanical stability.

3.3 Single Molecule Immunology: Antibodies and Antigens

A particularly interesting field is single molecule immunology. The binding of an antibody to its antigen is a crucial step in the response of the immune system to infection. Therefore, knowledge on binding constants and forces between antigen and a given antibody gives information on the efficiency of the immune response of this antibody. Several techniques have so far been used to study this process. In a methodical work, Tetin et al. have shown that the equilibrium dissociation constant of an antibody against the test antigen digoxin yields the same value if fluorescence polarization or FCS is used, that is, both techniques are suited to determine such constants. A disadvantage of these techniques is that they are performed in solution, while typical antigen antibody reactions occur at surfaces. Such surface studies were performed by Seeger et al. and Li et al. With tapping mode atomic force microscopy, they monitored the binding constant of two different antigen/antibody pairs at near physiological conditions in molecular monolayer consisting of the antibodies. The binding constants were essentially determined by counting the ratio of antibodies with and without the bound antigen. This technique is obviously also suited to evaluate antibody microarrays. Since the Y-shaped antibody molecule has two binding sites for an antigen, it is interesting to know whether both act independently or cooperate. This question has been tackled by using an AFM simultaneously as a microscope and a tool to measure forces. The unbinding force for an antibody against serum albumin has been measured as 244 ± 22 pN and an action of both binding sites independent of each other has been detected. The binding to individually addressed antibodies has been measured by Ros et al.

An interesting technical approach has been presented by Jermutus et al. The authors coexpressed and cotranslated single-chain antibody Fc fragments with ribosomal mRNA, that is, they displayed them on these large molecules. Then they could study the antigen binding on a single molecule basis using FCS and determine kinetic constants of that process with good statistics.

3.4 Molecular Motors

Molecular motors are molecules generally converting chemical energy into motion. Some small organic molecules are known to act as motors. They will not be discussed here. (Bio-) molecular motors govern processes characteristic of the life of cells and organisms, such as generation of motion in eukaryotic cells, cell division, cell movement, segregation of genetic material, or positioning of organelles. Owing to their fundamental function biomolecular motors are used as therapeutic targets (e.g. taxol treatment in cancer chemotherapy.) Linear motors move along filamentous macromolecules or macromolecule assemblies, which act as a kind of track. Prominent examples are kinesins and dyneins, which belong to motor families with more than one hundred members generating motion along microtubules by hydrolyzing ATP as a fuel. Another motor is myosin in its numerous variants, well known for generating motion along

actin microfilaments and thus, contributing critically to muscle contraction.

The motor proteins kinesin, dynein, and myosin consist of an elongated stalklike tail and one or two globular heads binding ATP as well as microtubules respectively to the actin microfilaments. Upon ATP consumption, they undergo conformational transitions connected with walking along the filamentous tracks. Thereby, the angle between the tail and the head is changed causing, if bound, a stroke driving the head forward in relation to the track. After such a stroke, the head is released, swings back to its "starting" position, binds to a new site on the track, and is loaded by the new ATP. The process is ready for repetition. In principle, most of the related questions to motor function and regulation can be answered by bulk experiments comprising some or many motor molecules. For example, the force generated by a muscle filament can be measured macroscopically. The number of myosin molecules involved in this process can be determined and allows to calculate force generation per myosin molecule. But one problem arises immediately: The duty ratio, that is, the ratio of the active period of the motor to the whole molecular cycle. It represents the fraction of molecules of a collective, which is involved in force generation at any given instant. Depending on the model used, the duty ratio can be estimated in the range between 5 and 90% – resulting in a factor of 20 in uncertainty. Therefore, it is essential to measure these mechano–chemical processes on a single molecule basis. One way to measure single molecule mechanics is to use fine glass microtools, contributing, however, to inherent problems such as adhesion and perturbation of the tiny forces involved. Optical tweezers are a second tool for single molecule motility research, allowing evaluation of both techniques against each other. Owing to the high interest in these questions, a large number of single molecule experiments have been described in early literature by the groups of Block, Molloy, Schliwa, Sheetz, Simmons, Spudich, Warshaw, and Yanagida.

More recently, Nishiyama et al. studied the chemo-mechanical coupling of forward and backward stepping of kinesin. They formulated a model where the driving force for motion along the microtubule is Brownian motion and where the energy is primarily used to decrease entropy, that is, by giving this motion a direction. In similar experiments, Gross et al. have investigated dynein and the related molecule dynactin, which can move bidirectionally along microtubules. The authors postulated that dynactin enables dynein to participate efficiently in this bidirectional transport. Rock et al. reported that Myosin VI is a highly processive motor with very large step size. The maximum step size reported so far is 35 nm for higher plant myosin XI. Baker et al. provided the basis for understanding the ATP reaction, which provides the energy under minimal load. They found that the dissociation of one phosphate group is more closely related to the generation of motion than could be demonstrated before. The work of Peterman et al. put forth the idea that single molecule studies may be related to the multimolecule interaction in real life in kinesin. These authors compared the relative orientations of monomeric and dimeric kinesin and showed that both kinesin molecules have to bind with similar orientation in order to reveal optimal activity. Among others, such investigations predict that biomolecular motors are active to transport artificial loads, which can be regulated and tuned

under cell free conditions and that they are able to work in predetermined directions.

As for technical motors, rotatory molecular motors also exist. F1 ATPase is a representative member of this class of molecular motors and converts chemical energy of ATP into rotation. Interestingly, it can also perform the task in reverse: when rotation is induced by exerting force on the molecule, it converts the low-energy ADP into high-energy ATP, that is, it converts mechanical into chemical energy. In a typical experiment, a fluorescence-labeled actin filament is bound to one of the subunits of the F1 ATPase. The ATPase is fixed to a surface. The actin filament can be easily observed by fluorescence microscopy and manipulated by micromechanical tools or optical tweezers. If left free and supplied with ATP, the ATPase rotates. This can be observed by rotation of the actin filament. A study of this type is, for example, that of Kato-Yamada.

3.5 Visualizing the Action of Individual Dynein and Myosin V Molecules in Real Time

Probably the most exciting recent development in single molecule research results from the possibility to obtain movies from working individual molecules. This does not mean the somewhat indirect visualisation of molecule action as described in the previous section. Meanwhile, it is possible to have a direct look on the dynamics of individual biomolecular motors. Burgess et al. have organized a temporal series of electron microscopic snapshots of dynein c molecules into a sequence of events, which allows to reveal how dynein generates motion along microtubules. They could show, that the stalklike tail performs structural changes, which may be required to generate motion. Even more exciting are the results of Ando et al.; using a homemade atomic force microscope with a temporal resolution of 80 ms, that is, half the frequency of commercial TV video. These studies with myosin V directly reveal that the stalk of the myosin molecule is very flexible and that there are changes in lengths of this arm, indicating, similarly as the dynein data, that structural changes of this part of the molecule, and not so much changes in the relative orientation of the heads of these molecules, generate motion. Thus, the mechanisms developed in the more indirect experiments described in Sect. 3.4 may need some modification in order to allow for these recent observations. Figure 1 shows the sequence of events for the myosin.

3.6 Single Molecule Enzyme Kinetics

From conventional biochemistry (enzyme concentrations 10^{-9} mol ml^{-1} or

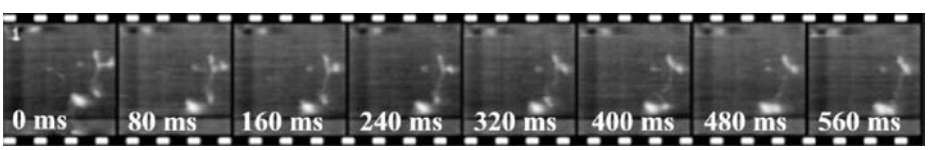

Fig. 1 Time series of myosin molecules in action (reproduced from Ando 2003). The motion of the Y-shaped myosin molecule is shown in steps of 80 ms. The molecule performs two types of motion: it slightly rotates as a whole and it's stalk changes length. These changes in shape are not completely consistent with the models derived from more conventional experiments and require that the latter should be modified accordingly.

10^{13}–10^{14} enzyme molecules), the step to the single molecule range involves an increase in sensitivity of at least 13 to 14 orders of magnitude. In recent years, enzyme kinetics has been developed down to the single molecule level. The conformational dynamics and cleaving mechanisms of enzymes have been directly observed at surfaces. The group of Xie has used the natural fluorescence of cholesterol oxidase, which is fluorescent in its oxidized form. As the reaction proceeds the enzyme oscillates between the oxidized and reduced states, that is, it blinks. Each blinking corresponds to one enzymatic turnover, which was followed with time. Also, fluctuation of reaction rates of one cholesterol oxidase molecule was observed. Earlier, Dovichi and coworkers had used a similar approach to confine single molecules of alkaline phosphatase and to measure the reaction products completely inside a capillary. A clinically interesting enzyme is lactate dehydrogenase (LDH), since it is used in the diagnostics of heart attacks. Xue and Yeung studied this reaction at selected substrate concentrations in a capillary tube equipped with a LIF detector and found different activities for individual molecules. As an alternative, the "droplet in substrate" technique, where the reaction proceeds at the surface of a highly viscous droplet containing a few LDH molecules, which was previously injected into a substrate solution allowed very detailed studies on this reaction. The conversion of nonfluorescing NAD+ into fluorescing NADH by individual molecules of LDH has been observed by this approach. Quantized reaction rates have been found. The rates were multiple integers of a basic reaction rate. This was interpreted as the reaction catalyzed by one, two, and so on, up to six enzyme molecules.

3.7 Single Protein Molecules in Viruses and in Living Cells

Single molecule research reaches even deeper into cell biology and thus life. A number of groups have been successful in imaging individual molecules on the surface of cells and even inside cells. An important step in such studies was the development of techniques to study the diffusion of individual molecules and the understanding of the influence of nearby surfaces close to diffusing molecules. The insertion of dye molecules into model membranes has been reported by Milhiet. Protein molecules, which usually reside in the cell membrane, have been incorporated into artificial membranes and their behavior could be studied there. Sub-wavelength resolution colocalization of individual dye molecules was reported by Heilemann. Also, on living cells the binding strength of single fibrinogen-integrin molecule pairs have been measured by allowing cells to get into contact and then measuring the acting forces. With optical tweezers, the binding strength was measured to range from 60 to 100 pN. Since the cells used in these studies, platelets, play a pivotal role in blood clotting, such information is of importance for understanding the basic features of the generation of heart attacks and may possibly give hints for strategies for the prevention of these catastrophic events.

Smith et al. have shown how the Phi 29 portal motor protein packs the 6.6 μm long double-stranded DNA molecule, which contains the genome of this virus against a force of more than 50 pN into the protein capsid of this virus. A corresponding theory on these processes is also available. But even the infection of individual

cells by individual adeno-associated virus particles has been directly observed by fluorescence microscopy. Comparing the reaction paths and speeds of many individual virus particles from the surface of the target cells into the nucleus it could be shown that there are preferential reaction paths and that the average infection time is approximately 12 min, however, with a large deviation among individual virus particles. Also, with the aim to study how the mono-cellular slime mold *Dictyostelium discoideum* reacts to chemotactic attraction, the binding of c-AMP molecules to different sites on the surface of these cells has been studied. The probability of individual association and dissociation events was greater for receptors at the anterior end of the cell, that is, the side into which the cells migrate. Particularly exciting is the possibility of "harvest" molecules, which are just passing a pore between cell nucleus and cytoplasm using an AFM tip. This has been accomplished by Oberleithner et al. with eggs from the frog *Xenopus laevis*. The motion of β-galactosidase molecules through 3T3 fibroblasts has been reported by Kues. This working group has also shown how to observe the motion of single molecules perpendicularly to the observation plane (the z-direction) of a microscope, a very important technique to observe single molecule signaling in a cell from its surface through the nuclear membrane into the cell nucleus or vice versa.

4 RNA and DNA: Highly Individual Molecules

4.1 Individual DNA Molecules as Construction Material and Electric Wires

Molecules have electric properties such as a resistance. For example, for single octanethiol molecules, a resistance of 900 ± 50 megohms was determined on the basis of more than 1000 individual molecules. Such properties are particularly interesting for DNA, of which a number of physical properties are known. This has led to the suggestion to use DNA as mechanical construction elements. Csaki et al. have shown how highly addressable single DNA molecules can be used as building blocks to construct molecular nanostructures. Two DNA molecules interact (hybridize), when their sequences are complementary. For example, if a DNA molecule has a sequence GAATCT at its end, it may bind to a sequence CTTAGA of a second DNA molecule. If two such molecules are brought into a specific geometric configuration, for example, with the tip of an AFM or with optical tweezers, complex nanostructures can be constructed with very high accuracy. The binding sequence with six DNA bases used in the example above has a length of only 1.8 nm, since the distance from "letter" to letter" of DNA is only 0.3 nm (3 Angstrom). Usually, in literature it is assumed that DNA can be directly used as an electric nanowire, since earlier experiments have found electric conductivity of DNA. This concept is, however, challenged by a paper of Gomez-Navarro et al. These authors have carried out two experiments: they have measured the charge transfer from an electrode to the single DNA molecule and they have measured the dielectric response, when the molecule has no contact to any electrodes at its two ends. The authors found high resistivity of DNA, comparable to mica, glass, or other electric insulators. DNA can even be used to generate motion. Such a nanomechanical device has been presented by Mao et al. Since the two structural variants of DNA, the B and the Z form, have different length

and are differently thick, the induction of a B-Z transition by changing the salt concentration may be used to induce motion in such a DNA construction network. Knots have been tied with DNA molecules and DNA molecules have been suggested to be used as molecular tweezers. Even when nanostructures are constructed from materials such as carbon nanotubes, single DNA molecules may be helpful as stabilizers of the structures.

4.2 Stretching and Folding RNA, DNA, and Chromatin

The step from the use of DNA as construction material to single molecule DNA biophysics is not large. While the B-Z transition described in the previous section can be induced physico-chemically, structural transitions in DNA can also be induced by exerting forces. In physiological environment, that is, at salt concentrations around 150 mM, DNA tends to collapse into a globular structure. With forces of the order of 10 to 20 pN, the DNA molecule can be extended into the normal structure with a length of about 0.3 nm per base pair. At forces between 20 and 60 pN, the DNA molecule is reversibly overstretched up to the 1.7 fold of its "natural" length and is occasionally termed S-DNA. Above 60 pN, irreversible overstretching will occur. Similar effects are seen with double-stranded RNA. Using the tip of an AFM, Bonin et al. have investigated its flexibility. They have found a conformational change similar to the DNA B-S transition, but overstretching is possible up to a factor of two.

The folding of single-stranded RNA is different and comparable rather to protein folding. Indeed, some single-stranded RNA molecules fold into structures with enzymatic activity (ribozymes). Using single molecule FRET, Russell et al. have shown that folding occurs in deep valleys of the folding landscape, that is, along very stable pathways. Interestingly, the temporal sequence of folding is very critical. If long-range contacts are made at the right time, they stabilize a structure, otherwise they may destabilize it. With similar techniques, Onoa et al. could show that the *Tetrahymena thermophilia* ribozyme has eight different folding barriers. Forces of 10 to 30 pN are required to overcome these barriers.

Similar techniques are used to study the stability of nucleosomes and of chromatin, the structures in which DNA is usually present in living cells. When a piece of chromatin consisting of DNA is wound around, the nucleosomes are stretched, in a first step 76 bp are released at low stretching force. When the force is increased, additional 80 bp are released in two stages. In a reverse experiment, Leuba et al. studied the assembly of nucleosomes and chromatin. They found a strong inhibition of assembly at forces as low as 10 pN, indicating that DNA needs its full flexibility to wrap around the nucleosomes.

4.3 DNA Melting and Zipping: Separating and Joining the Two Strands

A considerable wealth of information on the double-stranded DNA duplex can be obtained by pulling apart the two strands. To achieve that, one end of the molecule is attached to a surface. Of the other end, one strand is also fixed to some point while the second strand is attached to an AFM tip or a bead that may be manipulated by magnetic or optical tweezers. Force can then be used to separate the two strands of DNA. Stretched unwound DNA denatures locally even at comparably low forces of

2 pN. However, for complete separation of short double helices, 17–40 pN are required to overcome the energy barrier for rupture of 9–13 kcal mol^{-1}. Unzipping (or melting) can also be induced by high temperature. In combination with pulling the molecule, such a melting can be actively induced, and the DNA denaturation–renaturation kinetics can be studied. The *T*4 gene 32 protein gp32 specifically binds to single-stranded, that is, molten DNA. With single molecule force spectroscopy, it was shown that, surprisingly, gp32 does not reduce, as expected, the melting temperature, but affects the kinetics of this process. Danilowicz et al. found multiple metastable intermediates when lambda phage DNA molecules were unzipped. Unzipping changes pace in a sequence-dependent manner. Not only unzipping has been observed on a single molecule basis. Also, the rejoining of individual DNA strands can be observed on a single molecule basis. A single nanopore, made of alpha hemolysin, is required. By this approach, the binding of an octanucleotide to its complementary DNA strand has been observed with techniques derived from patch clamping. In spite of the complex environment, kinetics and binding constants for duplex formation were comparable to those of bulk experiments. This is an example, where single molecule experiments do not provide essentially new information, but confirm that bulk experiments provide useful data to understand the single molecule process.

4.4 Replication, Transcription DNA Repair and DNA Mapping

In Sects. 4.1 to 4.3, biophysical studies were reported. The step to biochemistry or molecular biology is represented by single molecule studies on enzymes working on DNA, particularly polymerases, enzymes involved in DNA replication and repair. A second sort of enzymes, endonucleases, will be discussed separately in Sect. 4.6. Occasionally, it is single molecule research that detects the true inherent complexity of polymerases, the enzyme class that synthesizes RNA from the DNA blueprint. Individual molecules of RNA polymerase reveal a surprisingly uniform kinetics, in contrast to some predictions derived from protein folding theories, from which significant differences were forecasted on the basis of structural heterogeneity. Obviously, evolution has selected, from the variety of possible structures, only those that reveal optimal function. Such a polymerase does not polymerize the RNA continuously, but pausing can be detected. When the template DNA molecule is stretched, this pausing can be modified and thus, studied in more detail. In a similar way, it has been shown that the rate of DNA replication slows down when the DNA molecule is stretched with a force of up to 20 pN. Above this force, replication completely ceases. In order to observe reactions combed DNA can also be used, that is, DNA stretched on the surface of a microscope slide. Topisomerases, that is, enzymes that are involved in the unwinding of DNA have been studied on a single molecule basis by Strick et al. and Dekker et al., nucleotide excision repair by Segers–Nolten. The question of how enzyme molecules find DNA and keep contact during catalysis is of major importance for maintenance and replication of the genome. The binding domain of Tc3 transposase, for example, when stained with the fluorescence dye tetramethyl rhodamine (TMR) fluctuates between a fluorescent and a quenched form, when

bound to DNA Daniel et al. Also, a so far unsolved or at least heavily disputed question asks for the mechanism of enzymes at work. Jiao et al. have detected, by time lapse atomic force microscopy, two modes of target recognition between the tumor suppressor protein p53 and a DNA fragment containing a recognition site for this protein: direct binding and initial nonspecific binding with subsequent one-dimensional diffusion of the protein along the DNA to the specific site. Even more important, polymerases or nucleases work on long DNA molecules, repeating a catalytic step many times. Polymerases have to add nucleotides step-by-step, nucleases have to remove or cut them. The question is then, for how many steps an enzyme molecule repeats this task, that is, how long does it stay attached to the individual DNA molecule. This has been discussed under the aspect of processivity. A number of these questions have been solved with ingenious interpretations of bulk experiments, but the final answer can only be obtained by direct observation of single molecule reactions. Harada et al. have shown in real time how a RNA polymerase molecule proceeds along a stretched DNA molecule for approximately 1 μm. In these studies, the DNA molecule remained unstained and the enzyme molecule was fluorescence labeled, so that one could observe a fluorescent spot running along an invisible linear track. In a similar study, however, with restriction endonucleases, the opposite strategy was employed: The DNA molecule was fluorescence labeled and the action of the enzymes was observed by the cuts they induced. Since, the cuts in Fig. 2 are temporarilly consecutive, (see time at the bottom of each figure) one can conclude that one enzyme molecule remains attached for several micrometers and induces sequentially all visible cuts.

4.5 Single Molecule DNA Sequencing

Probably, one of the most exciting potential applications of single molecule research is the sequencing of DNA. The idea was fostered by the early success to observe the enzymatic degradation of the molecule. Unfortunately, full single DNA molecule sequencing, after a decade of research and even after human genome sequencing has been completed, has not lived up to its promises. The principle idea is straightforward: Using an exonuclease, which digests DNA from one of its end base by base, a flow of single nucleotides through a detection volume is generated. By identification of the sequence of the single individual DNA bases flowing through this volume, the sequence of the DNA macromolecule, held in position by optical tweezers or an AFM tip, can be determined. Two problems have so far hampered this approach. As for all other optical detection techniques in single molecule research, fluorescence is needed. Unfortunately, at room temperature, DNA bases reveal a notoriously low fluorescence yield, four to five orders of magnitude lower than typical fluorescence dyes that have an yield of some ten percent. A few attempts to work in the cold at nitrogen or even helium temperatures have so far not been successful, since the digestion-enzyme reaction has to be performed at room temperature while detection has to be performed in the cold, and the sharp change of temperature has so far not yet been managed. The alternative, to attach sequence-specific fluorescence dyes, is exceedingly more difficult than in the examples described above, since for DNA sequencing, each DNA base has to be labeled and the overall yield of labeling has to be 99.9% in order to

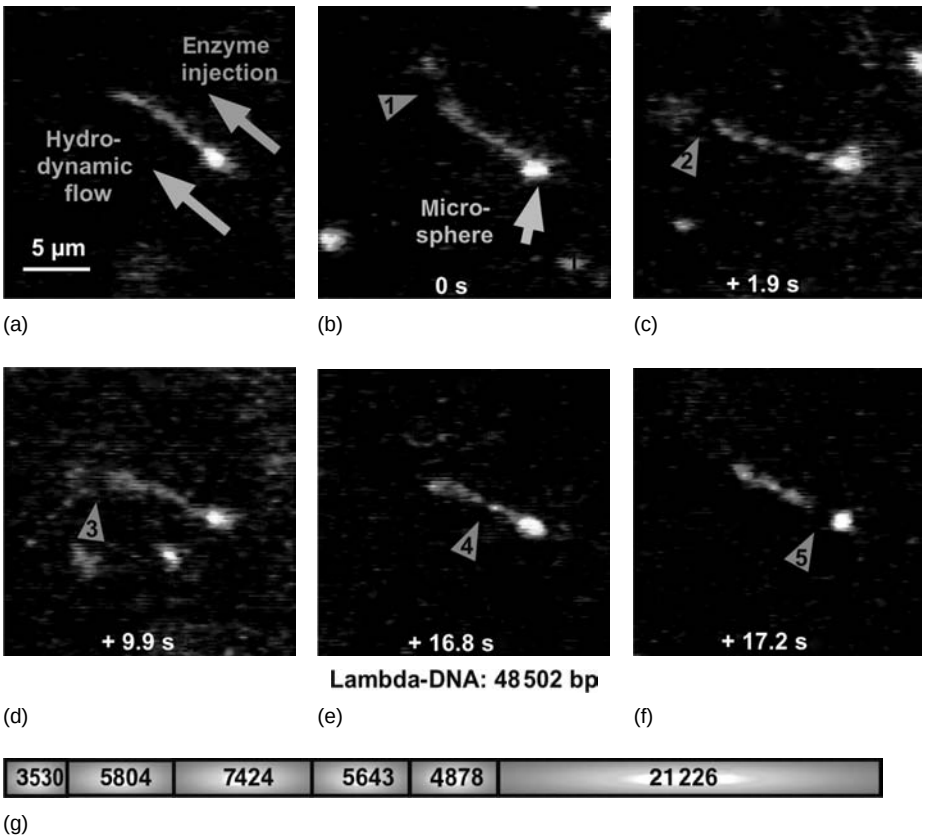

Fig. 2 Cutting of a single ''lambda phage'' DNA molecule, which is held by optical tweezers, into a sequence-specific pattern, The DNA molecule is stretched by hydrodynamic flow, otherwise it would collapse into a globular structure. The Eco R1 molecule attaches with highest probability ''downstream,'' which can be explained by a geometric shadow effect. Then it searches for it's first recognition sequence and cuts. It remains on the molecule and continues to search until all five restriction sites are consecutively cut. Cases in which the enzyme molecule binds not absolutely downstream, and cases in which the enzyme molecule falls off with the DNA fragment are also observed but represent the minority of events.

be competitive with conventional DNA sequencing. The labeling reaction has to be template directed, that is, to be a sort of DNA amplification. Once such a reaction is required, one can, however, easily amplify the single DNA molecule a million times and they perform conventional sequencing. The latter strategy indeed has been successful and is occasionally referred to as single molecule DNA sequencing. The original approach of direct sequencing, however, has remained just a concept.

4.6 DNA Analysis by Mapping and Restriction Fragment Sizing

In the mapping of individual DNA molecules, megabase-sized DNA mole-

cules are stretched and attached to the surface of a microscope slide and short kilobase-sized fluorescent DNA molecules are bound (hybridized) in a sequence-specific manner to the long target DNA. Such a mapping can also be combined with two-photon imaging. One will find a fingerprint-like sequence-dependent pattern of fluorescence spots. For the preparation of a large number of target molecules, molecular DNA combing has been developed, a technique in which many DNA molecules are stretched and paralleled by the receding meniscus of an evaporating microdroplet. Combed DNA can be used for mapping of a large number of parallel molecules. These authors could quantify the number of single gene copies on genomic DNA. A fluid flow technique, similar to the one described in the preceding section, has been used to characterize DNA fragments after application of ionizing radiation. The fragments are flowing away and can be quantified in a distant observation volume as fluorescent bursts. Anazawa et al. have presented a study on electrophoretic quantification of single DNA molecules, and Foquet et al. have performed fragment sizing in submicrometer sized fluidic channels. Fragment sizing can, however, also be used to get sequence information and thus to characterize individual DNA molecules. As in sequencing, the individual DNA molecule is held in position by optical tweezers and is stretched hydrodynamically or electrostatically. Then the molecule is cut into different fragments by sequence-specific restriction endonucleases, quite similar as it is done in bulk DNA typing to solve, for example, criminal cases or to perform paternity tests. The fragments have sizes that reflect the sequence of the individual DNA molecule.

Probably, the most detailed study uses nonspecifically labeled DNA and is a continuation of the basic study on sliding and hopping described in Sect. 4.4. Figure 3 shows, for DNA from the bacteriophage

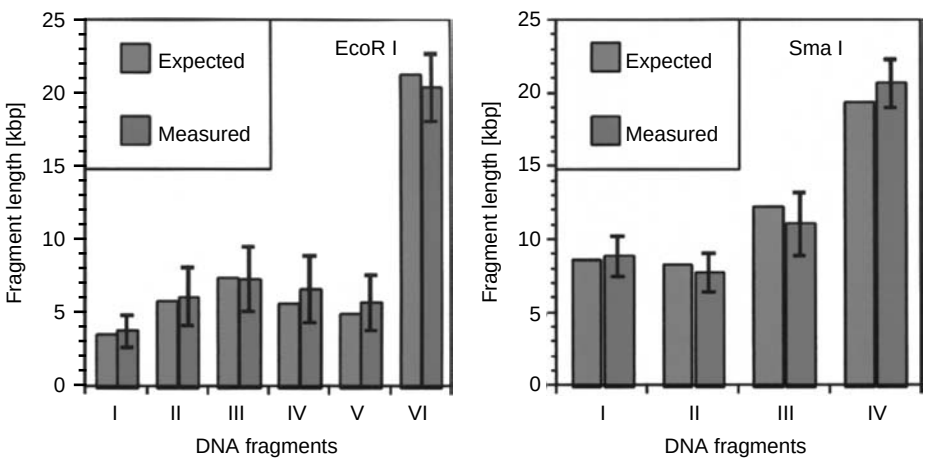

Fig. 3 Determination of the size of the restriction fragments: The fluorescence of the fragment is set in relation to the fluorescence of the whole molecule and then multiplied with the length of the latter, 48, 502 bp (red). The green columns are the values expected theoretically on the basis of the sequence of the molecule and the recognition sequences of the two enzyme molecules used, Eco R1 (left) and Sma 1 (right) (See color plate p. xxv).

Lambda, the sizes of the DNA fragments detected by this approach (green columns) as compared to the sizes expected from the sequence. The accuracy of individual fragment sizes is lower but comparable to those from conventional bulk experiments using 10–15 orders of magnitude more molecules. Individual DNA molecules can be safely characterized.

A particularly original strategy for DNA restriction uses unzipping of the double-stranded DNA molecule. If the two strands of a double-stranded DNA molecule are pulled apart with AFM tips, force is needed. At positions where restriction endonuclease molecules are bound to their recognition site on the respective DNA molecule, the force required abruptly increases and then decreases. This pattern of force peaks can be used to assemble a restriction pattern with an accuracy of approximately 25 bp.

5
Summary and Outlook

Single molecule techniques are now fully developed and used in the study of biological macromolecules. Since biomolecules have their own individuality, single molecule research is more than just a tremendous increase in sensitivity of 10 orders of magnitude. It opens the door to study the full width of variability in the living world.

Acknowledgment

This work was supported by the Stiftung Volkswagenwerk, Grant I 75099 and by the Fonds der Chemischen Industrie, Grant 161971. My particular thanks go to my colleague Eberhard Unger for help in the section on molecular motors.

See also Denaturation of DNA; Fluorescence; Fluorescence Correlation Spectroscopy (FCS); Motor Proteins; Nanobiotechnology.

Bibliography

Books and Reviews

Basche, T. (1999) Quantum optical experiments with single molecules (in German), *Phys. Bl.* **52**(5), 456–459.

Basche, T., Moerner, W.E., Orrit, M., Wild, U.P. (1997) *Single Molecule Optical Detection, Imaging and Spectroscopy*, VCH, Weinheim, Basel, Cambridge Tokyo, New York.

Erdmann, R., Enderlein, J., Seidel, C. (Eds) (1998) Single molecule detection: basics and applications in the life sciences, *Bioimaging* **6**(3).

Erdmann, R., Enderlein, J., Seidel, C. (Eds.) (1999) Single molecule detection and ultrasensitive analysis in the life sciences, *Cytometry* **36**(3), 162–264.

Erdmann, R., Seidel, C. (Eds) (1997) Single molecule detection: basics and applications in the life sciences, *Bioimaging* **5**(1), 1–300.

Greulich, K.O. (1999) *Micromanipulation by Light in Biology and Medicine: The Laser Microbeam and Optical Tweezers (Monography)*, Birkhäuser, Basel, Wien, Boston.

Greulich, K.O. (2000) Single Biomolecule Detection and Characterization, in: Meyers, R.A. (Ed.) *Encyclopedia of Analytical Chemistry*, John Wiley & Sons, Chichester, UK, pp 654–662.

Greulich, K.O., Klose, E., Neumann, H. (Eds) (1995) Single molecule detection, *Exp. Tech. Phys.* **41**(2), 157–311

Ishii, Y., Yanagida, T. (2000) Single molecule detection in the life sciences, *Single Mol.* **1**, 5–13.

Köhler, J. (2001) Optical spectroscopy of individual objects, *Nature* **88**, 514–521.

Moerner, W.E. (2002) A dozen years of single molecule spectroscopy in physics, chemistry

and biophysics, *J. Phys. Chem.* **106**(5), 910–927.

Sako, Y., Yanagida, T. (2003) Single molecule visualization in cell biology, *Nat. Cell Biol.* **5**, 1–5.

Tamarat, Ph., Maali, A., Lounis, B., Orrit, M. (2002) Ten years of single molecule spectroscopy, *J. Phys. Chem.* **104**(1), 1–16.

Xie, X.S. (2001) Single molecule approach to enzymology, *Single Mol.* **2**(4), 229–236.

Zander, C., Enderlein, J., Keller, R.A (Eds) (2002) *Single Molecule Detection in Solution*, Wiley-VCH, Berlin, Germany.

Primary Literature

Adelman, K., La Porta, A., Santangelo, T.J., Lis, J.T., Roberts, J.W., Wang, M.D. (2002) Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behaviour, *Proc. Natl. Acad. Sci. U.S.A.* **99**(21), 13538–13543.

Amediek, A., Haustein, E., Scherfeld, D., Schwille, P. (2002) Scanning dual-color cross correlation analysis for dynamic co-localization studies of immobile molecules, *Single Mol.* **3**, 201–210.

Anazawa, T., Matsunaga, H., Yeung, E.S. (2002) Electrophoretic quantitation of nucleic acids without amplification by single molecule imaging, *Anal. Chem.* **74**, 5033–5038.

Ando, T., Kodera, N., Naito, Y., Kinoshita, T., Furuta, K., Toyoshima, Y.Y. (2003) A high speed atomic force microscope for studying biological macromolecules in action, *Chem. Phys. Chem.* 1196–1202.

Anselmetti, D., Fritz, J., Smith, B., Fernandez-Busquets, X. (2000) Single molecule biophysics with atomic force microscopy, *Single Mol.* **1**, 53–558.

Arai, T., Yasuda, R., Akashi, K.I., Harada, Y., Miyata, H., Kinosita, K., Ito, H. (1999) Tying a molecular knot with optical tweezers, *Nature* **399**, 446–448.

Arya, M., Anvari, B., Romo, G.M., Cruz, M.A., Dong, J.-F., McIntire, L.V., Moake, J.L., Lopez, J.A. (2002) Ultralarge multimers of von Willebrand factor form spontaneous high-strength bond with the platelet glycoprotein Ib-IX complex: studies using optical tweezers, *Blood* **99**(11), 3971–3977.

Bach, H., Renn, A., Wild, U.P. (2000) Spectral imaging of single molecules, *Single Mol.* **1**, 73–78.

Baker, J.E., Brosseau, C., Jole, P.B., Warshaw, D.M. (2002) The biochemical kinetics underlying actin movement generated by one and many skeletal muscle myosin molecules, *Biophys. J.* **82**(4), 2134–2147.

Bensimon, D., Simon, A.J., Croquette, V., Bensimon, A. (1995) Stretching DNA with a receding meniscus: experiments and models, *Phys. Rev. Lett.* **74**, 4754–4757.

Bezanilla, M., Drake, B., Nudler, E., Kashlev, M., Hansma, P.K., Hansma, H.G. (1994) Motion and enzymatic degradation of DNA in the atomic force micrsocope, *Biophys. J.* **67**, 2454–2458.

Böhm, K.J., Stracke, R., Mühlig, P., Unger, E. (2001) Motor protein-driven unidirectional transport of micrometer-sized car goes across isopolar microtubule arrays, *Nanotechnology* **12**, 238–244.

Bonin, M., Zhu, R., Klaue, Y., Oberstrass, J., Oesterschulze, E., Nellen, W. (2002) Analysis of RNA flexibility by scanning force spectroscopy, *Nucleic Acids Res.* **30**(16), e81.

Brower-Toland, B.B., Smith, C.L., Yeh, R.C., Lis, J.T., Peterson, C.L., Wang, M.D. (2002) Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA, *Proc. Natl. Acad. Sci. U.S.A.* **99**(4), 1960–1965.

Constantino, C.J.L., Lemma, T., Antunes, P.A., Aroca, R. (2001) Single molecule detection using surface enhanced resonance Raman scattering and Langmuir – Blodgett monolayers, *Anal. Chem.* **73**(15), 3674–3678.

Csaki, A., Maubach, G., Born, D., Reichert, J., Fritzsche, W. (2002) DNA – based molecular nanotechnology, *Single Mol.* **3**(5–6), 275–280.

Cui, X.D., Primak, A., Zarate, X., Tomfohr, J., Sankey, O.F., Moore, A.L., Moore, T.A., Dust, D., Harris, G., Lindsay, S.M. (2001) Reproducible measurement of single-molecule conductivity, *Science* **294**, 571–574.

Daniel, D.C., Thompson, M., Woodbury, N.W. (2002) DNA-binding interactions and conformational fluctuations of Tc3 transposase DNA binding domain examined with single molecule fluorescence spectroscopy, *Biophys. J.* **82**(3), 1654–1666.

Danilowicz, C., Coljee, V.W., Bouzigues, C., Lubensky, D.K., Nelson, D.R., Prentiss, M. (2003) DNA unzipped under a constant force

exhibits multiple metastable intermediates, *Proc. Natl. Acad. Sci. U.S.A.* **100**(4), 1694–1699.

Dekker, N.H., Rybenkov, V.V., Duguet, M., Crisona, N.J., Cozzarrelli, N.R., Bensimon, D., Croquette, V. (2002) The mechanism of type IA topoisomerases, *Proc. Natl. Acad. Sci. U.S.A.* **99**(19), 12126–12131.

Dovichi, N.J., Martin, J.C., Jett, J.H., Trukula, M., Keller, R.A. (1984) *Anal. Chem.* **56**, 348–354.

Essevaz-Roulet, B., Bockelmann, U., Heslot, F. (1997) Mechanical separation of the complementary strands of DNA, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11935–11940.

Filippova, E.M., Monteleone, D.C., Trunk, J.G., Sutherland, B.M., Quake, S.R., Sutherland, J.C. (2003) Quantifying double-strand breaks and clustered damages in DNA by single-molecule laser fluorescence sizing, *Biophys. J.* **84**(2), 1281–1290.

Foquet, M., Korlach, J., Zipfel, W., Webb, W.W., Craighead, H.G. (2002) DNA fragment sizing by single molecule detection in submicrometer-sized closed fluidic channels, *Anal. Chem.* **74**, 1415–1422.

Forde, N.R., Izhaky, D., Woodcock, G.R., Wuite, G.J.L., Bustamante, C. (2002) Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by Escherichia coli RNA polymerase, *Proc. Natl. Acad. Sci. U.S.A.* **99**(18), 11682–11687.

Goel, A., Frank-Kamenetskii, M.D., Ellenberger, T., Herschbach, D. (2001) Tuning DNA strings: modulating the rate of DNA replication with mechanical tension, *Proc. Natl. Acad. Sci. U.S.A.* **98**(15), 8485–8489.

Gomez-Navarro, G., Moreno-Herrero, F., de Pablo, G.J., Colchero, J., Gomez-Herrero, C., Baro, A.M. (2002) Contactless experiments on individual DNA molecules show no evidence for molecular wire behavior, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8484–8487.

Grama, L., Somogy, B., Kellermayer, M.S.Z. (2001) Global configuration of single titin molecules observed through chain-associated rhodamine dimers, *Proc. Natl. Acad. Sci. U.S.A.* **98**(25), 14362–14367.

Grange, W., Strunz, T., Schumakovich, I., Güntherodt, H.J., Hegner, M. (2001) Molecular recognition and adhesion of individual DNA strands studied by dynamic force microscopy, *Single Mol.* **2**, 75–78.

Gross, S.P., Welte, M.A., Block, S.M., Wieschaus, E.F. (2002) Coordination of opposite polarity microtubule motors, *J. Cell Biol.* **156**(4), 715–7724.

Gueroui, N., Place, C., Freyssingeas, E., Berge, B. (2002) Observation by fluorescence microscopy of transcription on single combed DNA, *Proc. Natl. Acad. Sci. U.S.A.* **99**(9), 6005–6010.

Guthold, M., Erie, D.A. (2001) Single molecule study reveals a complex E. coli RNA polymerase, *Chem. Biol. Chem.* **2**, 167–170.

Ha, T., Ting, A.Y., Liang, J., Caldwell, W.B., Deniz, A.A., Chemla, D.S., Schultz, P.G., Weiss, S. (1999) Single molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 893–898.

Haab, B.B., Mathies, R. (1995) Single molecule fluorescence burst detection of DNA fragments separated by capillary electrophoresis, *Anal. Chem.* **67**, 3235–3260.

Hanley, W., McCarty, O., Jadhav, S., Tseng, Y., Wirtz, D., Konstanopoulos, K. (2003) Single molecule characterization of P-selectin/ligand binding, *J. Biol. Chem.* **278**, 10556–10561.

Harada, Y., Funatsu, T., Murakami, K., Nonoyama, Y., Ishihama, A., Yanagida, T. (1999) Single molecule imaging of RNA polymerase – DNA interactions in real time, *Biophys. J.* **76**, 709–715.

Heilemann, M., Herten, D.P., Heintzmann, R., Cremer, C., Müller, C., Tinnefeld, P., Weston, K.D., Wolfrum, J., Sauer, M. (2002) High resolution colocalization of single dye molecules by fluorescence lifetime imaging microscopy, *Anal. Chem.* **74**(14), 3511–3517.

Heymann, J.B., Möller, C., Müller, D.J. (2002) Sampling effects influence heights mea sured with AFM, *J. Microsc.* **207**(1), 43–51.

Hinterdorfer, P., Baumgartner, W., Gruber, H.J., Schilcher, K., Schindler, H.G. (1996) Detection and localization of individual antibody-antigen recognition events by atomic force microscopy, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 3477–3481.

Horton, M., Charras, G., Ballestrem, C., Lehenkari, P. (2000) Integration of atomic force and confocal microscopy, *Single Mol.* **1**, 135–1138.

Howorka, S., Movileanu, L., Braha, O., Bayley, H. (2001) Kinetics of duplex formation for individual DNA strands within a single protein nanopore, *Proc. Natl. Acad. Sci. U.S.A.* **98**(23), 12996–13001.

Huser, T., Yang, M., Rothberg, L.J. (2000) Single chain spectroscopy of conformational dependence of conjugated polymer photophysics, *Proc. Natl. Acad. Sci. U.S.A.* **97**(21), 11187–11191.

Jermutus, L., Kolly, R., Földes-Papp, Z., Hanes, J., Rigler, R., Plückthun, A. (2002) Ligand binding of a ribosome-displayed protein detected in solution at the single molecule level by fluorescence correlation spectroscopy, *Eur. Biophys. J.* **31**, 179–184.

Jiao, Y., Cherny, D., Heim, G., Jovin, T.M., Schäffer, T.E. (2001) Dynamic interactions of p53 with DNA in solution by time lapse atomic force microscopy, *J. Mol. Biol.* **314**, 233–243.

Kato-Yamada, Y., Noji, H., Yasuda, R., Kinosita, K. Jr., Yoshida, M. (1998) Direct observation of the rotation of the epsilon subunit in F1 – ATPase, *J. Biol. Chem.* **273**, 19375–19377.

Kinjo, M., Rigler, R. (1995) Ultrasensitive hybridization analysis using fluorescence correlation spectroscopy, *Nucleic Acids Res.* **23**, 1795–1799.

Kneipp, K., Wang, Y., Kneipp, H., Perelman, L.T., Itzkan, I., Dasari, R.R., Feld, M.S. (1997) Single molecule detection using surface enhanced Raman scattering (SERS), *Phys. Rev. Lett.* **78**, 1667–1670.

Koch, S.J., Shundrovsky, A., Jantzen, B.C., Wang, M.D. (2002) Probing Protein – DNA interactions by unzipping a single DNA double helix, *Biophys. J.* **83**(2), 1098–1105.

Kohl, T., Heinze, K.G., Kuhlemann, R., Koltermann, A., Schwille, P. (2002) A protease assay for two photon cross correlation and FRET analysis based solely on fluorescent proteins, *Proc. Natl. Acad. Sci. U.S.A.* **99**(19), 12161–12166.

König, K., Göhlert, A., Liehr, T., Loncarevic, I.F., Riemann, I. (2000) Two photon multicolor FISH: a technique to detect specific sequences on single DNA molecules in biological cells, *Single Mol.* **1**, 41–52.

Kues, T., Kubitschek, U. (2002) Single molecule motion perpendicular to the focal plane of a microscope: application to splicing factor dynamics within the cell nucleus, *Single Mol.* **3**, 218–224.

Kues, T., Peters, R., Kubitscheck, U. (2001) Visualization and tracking of single protein molecules in the cell nucleus, *Biophys. J.* **80**, 2954–2967.

Leuba, S.H., Karymov, M.A., Tomschik, M., Ramjit, R., Smith, P., Zlatanova, J. (2003) Assembly of single chromatin fibers depends on the tension in the DNA molecule: magnetic tweezers study, *Proc. Natl. Acad. Sci. U.S.A.* **100**(2), 495–500.

Li, L., Chen, S., Oh, S., Jiang, S. (2002) In situ single-molecules detection of antibody-antigen binding by tapping mode atomic force microscopy, *Anal. Chem.* **74**, 6017–6022.

Litvinov, R.I., Shuman, H., Bennett, J.S., Weisel, J.W. (2002) Binding strength and activation state of single fibrinogen-integrin pairs on living cells, *Proc. Natl. Acad. Sci. U.S.A.* **99**(11), 7426–7431.

Maier, C., Rünzler, D., Wagner, L., Grabner, G., Köhler, G., Luger, A. (2002) Evidence for specific glucocorticoid binding sites on the cell membrane by fluorescence correlation spectroscopy, *Single Mol.* **3**, 211–217.

Mao, C., Sun, W., Shen, Z., Seeman, N.C. (1999) A nanomechanical device based on the B-Z transition of DNA, *Nature* **397**, 144–146.

Milhiet, P.E., Giocondi, M.C., Baghdadi, O., Ronzon, F., Le Grimmellec, C., Roux, B. (2002) AFM detection of GPI protein insertion into DOPC/DPPC model membranes, *Single Mol.* **3**, 135–140.

Nishiyama, M., Higuchi, H., Yanagida, T. (2002) Chemomechanical coupling of the forward and backward steps of single kinesin molecules, *Nat.Cell Biol.* **4**, 790–795.

Oberhauser, A.F., Badilla-Fernandez, C., Carrion-Vasquez, M., Fernandez, J.M. (2003) The mechanical hierarchies of fibronectin observed with single molecule AFM, *J. Mol. Biol.* **319**(2), 433–447.

Oberleithner, H., Schäfer, C., Shahin, V., Schlune, A., Schillers, H., Reinhardt, J. (2001) Nuclear plug harvesting using atomic force microscopy, *Single Mol.* **2**, 117–120.

Onoa, B., Dumont, S., Liphardt, J., Smith, S.B., Tinoco, I., Bustamante, C. (2003) Identifying kinetic barriers to mechanical unfolding of the T. thermophila ribozyme, *Science* **299**, 1892–1895.

Paige, M.F., Bjerneld, E.J., Moerner, W.E. (2001) A comparison of through the objective total internal reflection microscopy and epifluorescence microscopy for single molecule research, *Single Mol.* **2**, 191–203.

Pant, K., Karpel, R.L., Williams, M.C. (2003) Kinetic regulation of single DNA molecule denaturation by T4 gene 32 protein structural domains, *J. Mol. Biol.* **327**(3), 571–578.

Peterman, E.G., Sosa, H., Goldstein, L.S.B., Moerner, W.E. (2001) Polarized fluorescence microscopy of individual and many kinesin motors bound to axonemal microtubules, *Biophys. J.* **81**, 2851–2863.

Pope, L.H., Davies, M.C., Laughton, C.A., Roberts, C.J., Tendler, S.J.B., Williams, P.M. (2001) Force induced melting of a short double helix, *Eur. Biophys. J.* **30**, 53–62.

Pralle, A., Florin, E.L., Stelzer, E., Hörber, H. (2000) Photonic force micrsocopy: a new tool providing new methods to study membranes at the molecular level, *Single Mol.* **1**, 129–133.

Purohit, P.P., Kondev, J., Phillips, R. (2003) Mechanics of DNA packaging in viruses, *Proc. Natl. Acad. Sci. U.S.A.* **100**(6), 3173–3178.

Rief, M., Clausen-Schaumann, H., Gaub, H.E. (1999) Sequence-dependent mechanics of single DNA molecules, *Nat. Struct. Biol.* **6**, 346–349.

Rock, R.S., Rice, S.E., Wlles, A.L., Purcell, T.J., Spudich, J.A., Sweeney, H.L. (2001) Myosin VI is a processive motor woth a large step size, *Proc. Natl. Acad. Sci. U.S.A.* **98**(24), 13655–13659.

Ros, R., Schwenger, F., Anselmetti, D., Kubon, M., Schäfer, R., Plückthun, A., Tiefenauer, L. (1998) Antigen binding forces of individually addressed single chain Fv antibody molecules, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7402–7405.

Rothman, B. (1961) Measurement of activity of single molecules of beta D galactosidase, *Proc. Natl. Acad. Sci. U.S.A.* **47**, 1981–1991.

Rowell, N.D. (2000) Tiny tweezers made of DNA, *Biophys. Int.* **9/10**, 86–87.

Russell, R., Zhuang, X., Babcock, H.P., Millett, I.S., Doniach, S., Chu, S., Herschlag, D. (2002) Exploring the folding landscape of structured RNA, *Proc. Natl. Acad. Sci. U.S.A.* **99**(1), 155–160.

Sako, Y., Uyemura, T. (2002) Total internal reflection microscopy for single-molecule imaging in living cells, *Cell Struct. Funct.* **27**, 357–365.

Schäfer, B., Gemeinhardt, H., Greulich, K.O. (2001) Direct microscopic observation of the time course of single molecule DNA restriction reactions, *Angew. Chem., Int. Ed. Engl.* **40**, 4663–4666.

Schäfer, B., Nasanshargal, B., Monajembashi, S., Gemeinhardt, H., Greulich, K.O., Uhl, V. (1999) Study of single molecule dynamics and reactions with classic light microscopy, *Cytometry* **36**(3), 209–216.

Schmidt, T., Schütz, G.J., Baumgartner, W., Gruber, H.J., Schindler, H. (1996) Imaging of single molecule diffusion, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2926–2929.

Schuster, J., Cichos, F., Wrachtrup, J., von Borszyskowski, C. (2000) Diffusion of single molecules close to surfaces, *Single Mol.* **1**, 299–305.

Seeger, S. (1998) Single molecule fluorescence: high performance molecular diagnosis and screening, *Bioforum* **4**, 179–185.

Segers-Nolten, G.M.J., Wyman, C., Wijgers, N., Vermeulen, W., Lenferink, A.T.M., Hoeijmakers, J.H.G., Greve, J., Otto, C. (2002) Scanning confocal microscopy for single molecule analysis of nucleotide excision repair complexes, *Nucleic Acids Res.* **30**(21), 4720–4727.

Seisenberger, G., Ried, M.U., Endreß, T., Büning, H., Hallek, M., Bräuchle, C. (2001) Real-time single molecules imaging of the infection pathway of an adeno-associate virus, *Science* **294**, 1929–1932.

Smith, D.E., Tans, S.J., Smith, S.B., Grimes, S., Anderson, D.L., Bustamante, C. (2001) The bacteriophage Phi29 portal motor can package DNA against a large internal force, *Nature* **413**, 748–752.

Stark, M., Stark, R.W., Heckl, W.M., Guckenberger, R. (2002) Inverting dynamic force microscopy: from signals to time resolved interaction forces, *Proc. Natl. Acad. Sci. U.S.A.* **99**(13), 8473–8478.

Strick, T.R., Croquette, V., Bensimon, D. (1998) Homologous pairing in stretched supercoiled DNA, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10579–10583.

Strick, T.R., Croquette, V., Bensimon, D. (2000) Single molecule analyis of DNA uncoiling by a type II topoisomerase, *Nature* **404**, 901–904.

Taguchi, H., Ueono, T., Tadakuma, H., Yoshida, M., Funatsu, T. (2001) Single molecule observation of protein – protein interactions in the chaperonin system, *Nat. Biotechnol.* **19**, 861–865.

Tetin, S.Y., Swift, K.M., Matayoshi, E.D. (2002) Measuring antibody affinity and performing immunoassay at the single molecule level, *Anal. Biochem.* **307**(1), 84–91.

Tietz, C., Gehrken, U., Jelezko, F., Wrachtrup, J. (2000) Polarization measurements of single

pigment protein complexes, *Single Mol.* **1**, 67–72.

Tinnefeld, P., Buschmann, V., Herten, D.P., Han, K.T., Sauer, M. (2000) Confocal fluorescence lifetime imaging micrsocopy (FLIM) at the single molecule level, *Single Mol.* **1**, 215–223.

Tomigawa, M., Kojima, H., Yokato, E., Orii, H., Nakamori, R., Katayama, E., Anson, M., Shimmen, T., Oiwa, K. (2003) Higher plant myosin XI moves processively on actin with 35 nm steps at high velocity, *EMBO J.* **22**(6), 1263–1272.

Ueda, V., Sako, Y., Tanaka, T., Devreotes, P., Yanagida, T. (2001) Single-molecule analysis of chemotactic signaling in *Dictyostelium* cells, *Science* **294**, 864–867.

Uhl, V., Pilarczyk, G., Greulich, K.O. (1998) Fluorescence microscopic observation of catalysis by single or few LDH 1 enzyme molecules, *Biol. Chem.* **379**, 1175–1180.

Vale, R.D., Milligan, R.A. (2000) The way things move: looking under the hood of molecular motor proteins, *Science* **288**, 88–95.

Wabuyele, M.B., Soper, S.A. (2001) PCR amplification and sequencing of single copy DNA molecules, *Single Mol.* **2**, 13–22.

Wang, Y.K., Huff, E., Schwartz, D.C. (1995) Optical mapping of site directed cleavages on single DNA molecules by the recA assisted restriction endonuclease technique, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 165–169.

Weidemann, T., Wachsmuth, M., Tewes, M., Rippe, K., Langowski, J. (2002) Analysis of ligand binding by two colour fluorescence cross correlation spectroscopy, *Single Mol.* **3**, 49–62.

Wintterlin, J., Völkering, S., Janssens, T.W., Zambelli, T., Ertl, G. (1997) Atomic and macroscopic reaction rates of a surface catalyzed reaction, *Science* **278**, 1931–1934.

Xie, X.S., Lu, H.P. (1999) Single molecule enzymology, *J. Biol. Chem.* **274**, 15967–15970.

Xue, Q., Yeung, E.S. (1995) Difference in the chemical reactivity of individual molecules of an enzyme reaction, *Nature* **373**, 681–683.

Zumbusch, A., Jung, G. (2000) Single molecule spectroscopy of the green fluorescent protein: a critical assessment., *Single Mol.* **1**, 261–270.

Nucleic Acid Chromatography

Douglas T. Gjerde[1], *Christopher P. Hanna*[1], *David Hornby*[2], *and Guenther Bonn*[3]
[1] *PhyNexus, Inc., San Jose, CA*
[2] *University of Sheffield, Sheffield, UK*
[3] *University of Innsbruck, Innsbruck, Austria*

1 General Overview 297
1.1 Historical Review and General Applications 297
1.2 General Principles 300
1.2.1 Overall System 300
1.2.2 Column Protection 302
1.2.3 Use of the Column Oven and Temperature Modes of DNA Chromatography 303
1.2.4 Detection 304

2 New Applications of DNA Chromatography 304
2.1 Analysis Using Enzymatic Site-specific Cleavage 304
2.2 DNA Methylation Analysis 304
2.3 Phosphorylation Analysis 306
2.4 Footprinting 308

3 Denaturing High Performance Liquid Chromatography (DHPLC) 310
3.1 The Basis of DHPLC 310
3.2 DHPLC and Its Practice 311
3.2.1 DNA Melting Fundamentals 311
3.2.5 Other DHPLC Recommendations 317
3.3 Review of DHPLC Development 318

4 RNA Chromatography 319
4.1 RNA Structure 319
4.2 Biological Extraction of RNA 321
4.3 Size-based RNA Separation 322
4.4 Separation of Cellular RNA Species 324

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

4.5 Separation of Messenger RNA from Ribosomal RNA 324
4.6 Analysis of Transfer RNA 325
4.7 Chromatography and Analysis of Synthetic Oligoribonucleotides 326

Bibliography 328
Books and Reviews 328
Primary Literature 329

Keywords

Autosampler
Automatically takes sample from the PCR plate, places the sample in the injector valve, and injects the sample into the mobile phase flowing to the column.

Chromatography
Literally "color writing," a term describing a separation technique that occurs on the basis of the difference of interaction of sample components with two phases: a stationary phase or a column and liquid mobile phase or eluent, sometimes called *HPLC* or *high performance liquid chromatography*.

Column
Typically a tube of sorts that contains the stationary phase or column packing of a chromatographic system, usually consisting of nonporous, polymer-based particles chemically bonded with a C18 chemical functional group.

Detector
One of the basic components of a DNA chromatographic system; responsible for providing the presence and amount of separated sample components as they pass through the flow cell.

Eluent
Another term for the mobile phase used to carry out a separation.

Gradient Elution
An HPLC procedure where the composition of the mobile phase increases in strength during the separation; a typical example is where the mobile phase starts out as water with a low concentration of acetonitrile and more acetonitrile is added continuously as the separation proceeds – increasing the amount of acetonitrile in the eluent makes it stronger, and sample bands leave the column sooner.

Fluorescence Detector
Device using excitation light to induce fluorescence of tagged nucleic acids as they elute from the column; advanced fluorescence detectors may excite and detect up to four different tags simultaneously.

Ion Pair Chromatography
A chromatographic method in which the sample ions are "paired" with the opposite charged agents, thus forming a neutral ion pair and allowing the sample to be retained on a nonpolar, reverse phase column.

Mass Spectrometry Detector
A chromatographic detector in which sample components are ionized and resolved on the basis of their mass to charge (m/z) ratio because of their behavior in electric and magnetic fields.

UV Detector
The most popular form of detector used in DNA chromatography based on the absorption of UV light by the nucleic acid fragments typically allows operation at any wavelength from 190 nm to near 380 nm, with the most popular setting at 260 nm (or 254 nm) for single wavelength detectors.

DNA chromatography is the high performance, automatic separation, and purification of nucleic acids by liquid chromatography. It is an analytical technique not widely understood nor widely used by the molecular biologist, but has tremendous power.

The needs in molecular biology science are changing rapidly. In order to understand and ultimately control the molecular basis of life, complex experiments must be designed and large numbers of samples must be analyzed. Stated differently, the molecular biologist must become more analytical. New analytical tools must be implemented, and analytical information must be generated and studied.

Because of the inherent analytical nature, DNA chromatography has the potential to become one of the leading analytical tools used by the molecular biologist. Fortunately, the technology and methods have been designed so that it is not necessary for a molecular biologist to become an analytical chemist in order to perform the work. However, it is necessary to understand some of the concepts that are the basis of the technology and to understand where and how it can be implemented. It is the intent of the authors that this article should help molecular biologists how to use DNA chromatography effectively in their research.

1 General Overview

1.1 Historical Review and General Applications

The knowledge base of molecular biology has been built upon the use of analytical tools. The success of any particular tool depends on how well it can perform its analytical function – for example, is the tool and method accurate, fast, and reproducible, and does it perform the task that is needed? One of the mainstay tools used in molecular biology is the

separation, quantification, and purification of nucleic acids.

Chromatography or high performance liquid chromatography (HPLC) is a column technique where samples are introduced to one end of the column and are separated as they are carried through the column by a fluid (the eluent) being pumped through it. The separation is monitored with a detector, and separated materials are collected if desired. A special variation of the chromatographic method is DNA chromatography, where large quantities of nucleic acid fragments can be introduced, separated, detected, purified, and collected through computer-controlled instrumentation. The collected DNA may be used for downstream processing techniques such as cloning, PCR, or sequencing.

With the introduction of high performance columns, modern DNA chromatography separations were achieved. This technology has been applied to a wide variety of applications, including quantitative PCR product analysis, cloning purification, RNA analysis and purification, viral diagnostics, short tandem repeat analysis, population analysis, primer extension, pooling of genomes to determine allele frequency, and many other applications.

Early nucleic acid separations have been described by J. Thompson in review articles. The fundamental technology leading to modern DNA chromatography was first described by Guenther Bonn, Christian Huber, and Peter Oefner in 1993. They showed a new nonporous, alkylated polymer column that provided rapid, high-resolution separations of both double-stranded and single-stranded DNA and were performed usually in less than 10 min, and, in many cases, single base pair resolution was achieved. Figure 1 shows an example of a separation of double-stranded DNA.

Ion pairing, reverse phase chromatography is the basis for most DNA chromatography technology and applications. An alkylammonium salt is added to the eluent, and this forms neutral ion pairs when a DNA sample is introduced into the HPLC instrument. An eluent gradient of an acetonitrile solvent separates the DNA fragments with the smaller fragments coming off the column first and then larger fragments eluting off the column and traveling through the detector.

Bonn and coworkers showed that the DNA separations were performed generally according to the size of the fragment, just as they do in gel electrophoresis. For double-stranded DNA, the particular sequence does not contribute to the retention of the fragment. Figure 2, taken from their work, demonstrates this with a plot of retention time versus fragment size for a number of different fragments. Various plasmids were digested with a number of different enzymes so that a mixture of DNA fragments was generated. Since the sequences of the plasmids are known, the effect of DNA sequence on their elution time can be shown. In the plot, all of the retention times of the various fragments fall on a line drawn through the data. This

Fig. 2 The capacity factors of various plasmid digest fragments are plotted according to size. The fragments have different sequences, but are separated according to their size using a DNASep® column 50 × 4.6 mm. Eluent and Gradient: A: 0.1 M TEAA; B: 0.1 M TEAA, 25% acetonitrile, 0 to 100% B in 30 min., 1.0 mL/min flow rate, UV detection at 254 nm (data from Huber, C.G., Oefner, P.J., Bonn, G.K. (1993) Rapid analysis of biopolymers on modified non-porous polystyrene-divinylbenzene particles, *Chromatographia* **37**, 653–658.)

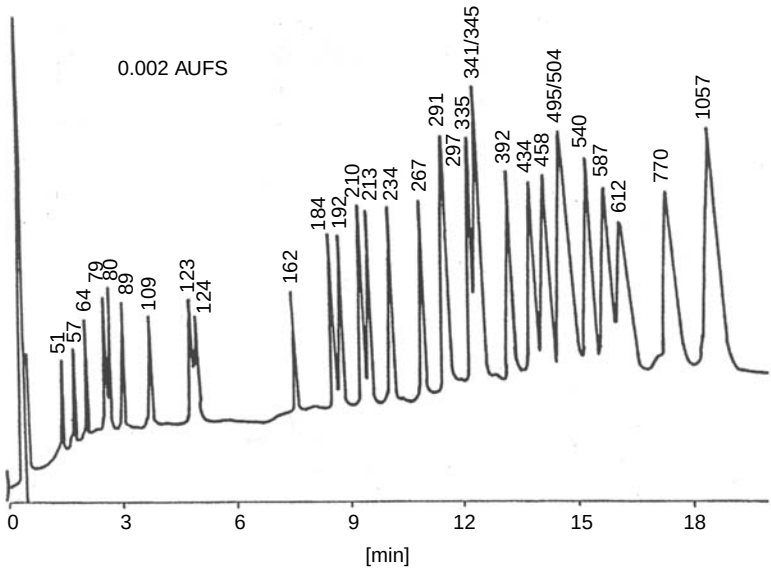

Fig. 1 High performance DNA chromatography separation of double-stranded DNA using a DNASep® column 50 × 4.6 mm. Sample was a mixture of Br322 HaeIII restriction digest and ΦX174 Hinc II restriction digest. Eluent and Gradient: A: 0.1 M TEAA; B: 0.1 M TEAA, 25% acetonitrile, 35 to 45% B in 2 min., 45 to 57% in 10 min., 57 to 61% in 4 min, 1.0 mL/min flow rate, UV detection at 254 nm, TEAA is triethylammonium acetate (Bonn, G., Huber, C., Oefner, P. (1996) *Nucleic Acid Separation on Alkylated Nonporous Polymer Beads*, U.S. Patent 5,585,236. with permission).

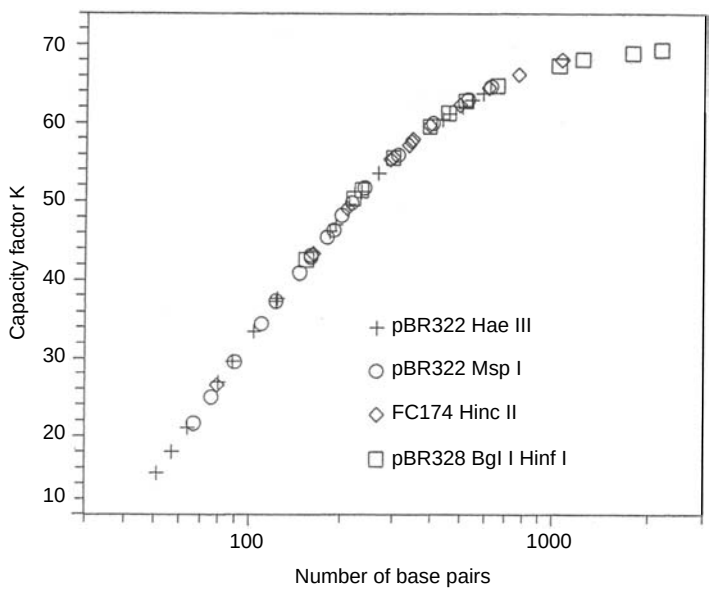

shows that the retention of DNA is dependent on fragment size, and is independent of fragment sequence. However, the selectivity of the separation of single-stranded DNA is based on both size and fragment sequence (i.e. some size reversals can occur in the separations).

A key development of modern DNA chromatography was performed by Gjerde, Haefele, and Togami, where they showed the detrimental effect of metal contamination and other contamination on the performance of the DNA separations. Without this development, reproducible and predictable separations of DNA could not have been performed and a commercial column could not have been developed. Furthermore, it was shown that a variety of polymer-based and silica-based columns could be used for DNA separations as long as they had good transport kinetics for the large nucleic acid biomolecules (i.e. the column packing should not trap the large biomolecules), and, most importantly, were free of metal contaminants or could be cleaned with an (chelating) additive in the eluent. Requirements of the column and system to be free of contamination are more critical for double-stranded DNA separation than for single-stranded DNA separations.

In 1996, Oefner and Underhill extended DNA chromatography with the invention of denaturing HPLC. By controlling the column temperature of DNA chromatography column, they demonstrated that DNA chromatography possessed unique properties, enabling the separation of double-stranded DNA based on its relative degree of helicity. Heteroduplex DNA has a lower melting point than homoduplex DNA. This technique is called *denaturing high performance liquid chromatography* (DHPLC). In short, they developed a method of DNA chromatography that is analogous to denaturing gradient gel electrophoresis (DGGE).

Many of the newest RNA and DNA applications of DNA chromatography have been performed by David Hornby and coworkers, and are described in Sect. 2 of this article.

In addition to these application developments, there have also been developments in separation media for DNA chromatography. Investigators have recently developed columns that are not made from packed beads (as in standard columns), but are rather made of a single polymer structure that extends the entire length of the column – thus giving them their common moniker of "monolith columns." The polymer structure possesses so-called "through-pores" that allow sample eluent to pass through, which are analogous to the spaces between the beads in a standard column. For the separation of biopolymers, Horvath and coworkers applied this type of polymeric column for the separation of proteins, while Huber and coworkers have shown these types of columns to provide excellent separation of nucleic acids for mass spectrometric detection. These columns have also been referred to in the section on DHPLC.

1.2 General Principles

1.2.1 Overall System

This section discusses the principles of DNA chromatography. While instrumentation is important, the key to the separation process is the chemistry that occurs within the column (the stationary phase) and eluent (the mobile phase).

DNA chromatography employs the following:

1. An efficient column with as high a resolving power as possible.

2. An eluting solvent or eluent that provides reasonable separation of the mixture.
3. A column/eluent system that attains equilibrium quickly so that peak broadening due to slow chromatography kinetic interaction rates is eliminated or minimized.
4. Elution conditions such that retention times are in a convenient range – neither too short nor too long.

One major difference between chromatography and electrophoresis is the use of gradient elution in chromatography. The content of the fluid (eluent) that is pumped through the column can be changed in real time as the separation is proceeding. This is called a *gradient process*, whereby the "strength" or "eluting power" of the eluent is increased as the separation is developed. Small fragments are eluted first and then the organic solvent is increased and larger and larger fragments are eluted. The time required for the entire separation depends on the "eluting program" that is employed. A rapid increase in the stronger eluent will elute the fragments faster. However, if a rapid program is used, a separation with a lower resolution of peaks could result. Conversely, a slow gradient process will produce the highest resolving conditions and greatest separation of peaks. However, this carries the penalty of a longer separation time. If the DNA chromatographic system is well designed and a high-efficiency column is used, then rapid gradients will produce suitable and useful separations. Conversely, electrophoresis must rely on the conditions selected for the entire separation. The resolving conditions are not changed once the run is initiated; thus, whatever conditions were first selected must be used for the entire process.

Another advantage of DNA chromatography is the ability to detect the DNA fragment without fluorescence detection. Tags need not be used and the fragments may be detected directly at the subnanogram level by UV automatic detection. Of course, fluorescence detection could also be used provided fluorescent tags are added to the DNA in the PCR amplification. Alternatively, a fluorescent tag can be added as an intercalating dye. Use of this detection method decreases the amount of DNA that can be detected by a factor of 10 to 100 (or even 1000 in some reported cases).

Figure 3 shows a block diagram of the general components of a DNA chromatography instrument. The components are as follows: a supply of eluents or buffers (also called the *mobile phase*), an eluent degasser to remove dissolved oxygen, carbon dioxide and nitrogen from the fluid, a high pressure pump (with pressure indicator) to deliver the eluent or mobile phase, an autosampler, including syringe and sample valve, for introducing sample into the eluent stream and onto the column, a column (also called the *stationary phase*) to separate the sample mixture into the individual components, an oven to contain the column and control the temperature of the nucleic acids, a detector to measure the peaks or bands of the nucleic acid fragments as they elute from the column, an optional fragment collector to collect any particular fragment of interest, and a data system for collecting and organizing the chromatograms. A computer controls the instrument through software that is used to develop and implement analytical methods for each set of samples. The software guides the user to calibrate the instrument and to develop the appropriate and optimum methods, depending on the analysis

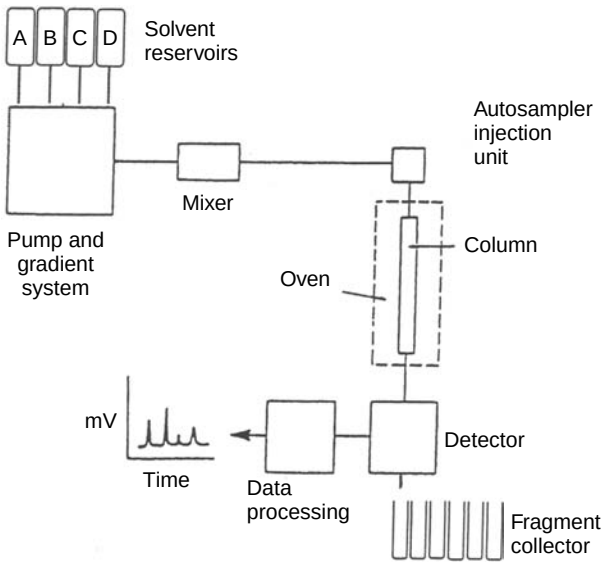

Fig. 3 Block diagram of a DNA Chromatograph.

being performed and whether fragments are to be collected.

Unlike electrophoresis gels that are used only once, the chromatographic column is used for several hundred or even thousands of injections before it is replaced. The eluent is pumped continuously through the column. Periodically, samples are injected and separated. A particular mix of eluent strength is used for each separation depending on the sample type (fragment sizes) and the desired separation.

The columns are packed with beads that provide the basis for the separations. Most column packings are beads approximately 2 μm in diameter, functionalized with a hydrophobic, neutral C18 alkyl group or they are functionalized with quaternary ammonium groups, which serve as the sites for an anion exchange process. A popular trademark name for this material is called DNASep® column and is supplied by Transgenomic, Inc., Omaha, NE. Another older type of column material is an anion exchanger available from a number of manufacturers, including Dionex, Inc. Sunnyvale, CA, and Tosoh Corporation of Japan. Although most of the recent work has been performed with neutral reverse phase materials, anion exchange is still used, especially for the separation of short, single-stranded DNA.

1.2.2 Column Protection

Column protection not only extends the useful life of the separation column, but proper protection of the separation column can also result in more reliable analytical results over the lifetime of the column. Dirty columns can show broad and, sometimes, even split peaks. The most common protection is the use of in-line filters located directly in front of the separation column. Material that would normally be trapped and would contaminate the separation column instead gets trapped by the filter. In-line filters are quite inexpensive and should

be changed frequently – at least every 2 to 3 weeks and sometimes much more often.

Another method to protect the system from metal contamination is to add small amounts (i.e. 0.1 mM) of tetrasodium EDTA to the mobile phase. This can be done without significant changes to the chromatography; however, this can affect removal of the solvent in nucleic acid material collected from the column to be used in further processing.

1.2.3 Use of the Column Oven and Temperature Modes of DNA Chromatography

The oven controls the temperature of the fluid entering the separation column. This includes not only the eluent but also the sample that is injected into the system. Temperature of the fluid can be thought of as an additional reagent used in the separation and analysis of nucleic acids. This is best described in the three modes of operation of analysis: nondenaturing mode, partially denaturing mode, and fully denaturing mode.

Nondenaturing mode The breaking of hydrogen bonds of double-stranded DNA through increasing the temperature is called *melting*. The temperature at which this occurs depends on the strength of the hydrogen bonding and the environment around the DNA. Higher salt and buffer content will raise the melting temperature. Also, DNA adsorbed onto a solid surface, such as column packing material, will require higher temperature to melt the DNA. Conversely, the presence of an organic solvent such as acetonitrile will lower the melting temperature. The effect is increased with increasing concentrations of the solvent. In the nondenaturing mode, high enough temperature is chosen to lower eluent viscosity (and therefore column back pressure), but not so high that denaturing occurs. The normal temperature of operation is 50 °C.

Partially denaturing mode The use of partially denaturing mode is discussed in the section on DHPLC. Mutations are detected through a heteroduplex detection process. Under this mode, differences in melting of double-stranded DNA are detected because increased melting decreases the retention time of the fragment. The temperature is chosen depending on the sequence and the melting of the fragment being analyzed. Typically, oven temperatures between 54 and 72 °C are used. The method has been called a *difference-detecting engine* because it detects the presence of a heteroduplex regardless of the sequence being studied.

Fully denaturing mode Single-stranded DNA and RNA may contain secondary structure due to complementary sequences within its own fragment. If portions of sequences are complementary, the fragments may fold back on themselves through the formation of intramolecular hydrogen bonds, giving several different possible structures for the same fragment. The presence of these secondary structures is uncertain and nonreproducible and will lead to peak broadening if they occur. Increasing the temperature can break up the hydrogen bonding and therefore reduce some or perhaps most of the secondary structure. An example of this behavior is the separation of an RNA ladder at 40 and 75 °C, where the chromatographic peaks of the RNA become sharper and more uniform with increasing oven temperature because of the reduction of RNA secondary structure.

1.2.4 Detection

One of the great strengths of DNA chromatography over capillary and slab gel electrophoresis is that UV detection can be used without any previous modification of the nucleic acid. No tagging is necessary for separation and purification of nucleic acids. Still, fluorescence has its strengths, and, for general R&D purposes, many researchers use UV and fluorescence detectors simultaneously. They are connected in series with UV detection first, followed by fluorescence detection. However, if only one detector were available for an instrument, the general UV detector would be the detector of choice.

Coupling to ESI-MS provides information about the size of the DNA fragment and allows a confident identification and characterization.

2 New Applications of DNA Chromatography

2.1 Analysis Using Enzymatic Site-specific Cleavage

Contemporary molecular biology has emerged through the integration of genetics, biochemistry, chemistry, and biophysics, and centers on the metabolism of nucleic acids. In experimental terms, the development of our understanding of nucleic acid synthesis, degradation, modification, and expression has relied heavily upon nucleic acid separation, isolation, and detection methods. By far, the most widely used method for nucleic acid separation is gel electrophoresis. However, gel electrophoresis presents problems of quantification and is difficult to use in a preparative mode.

One of the fundamental steps in the analysis and characterization of genomic DNA and in molecular cloning is the enzymatic, site-specific cleavage and detection of DNA using restriction endonucleases. Type II restriction endonucleases cleave DNA at specific sites that are usually palindromic and typically recognize between 4 and 8 bp. With careful choice of the eluent, DNA chromatography can be used to separate fragments generated from such reactions that are up to 1000 bp in length. An example of the use of DNA chromatography in the preparative separation of restriction fragments for molecular cloning has been described by Matin and Hornby. When the linear "chromosome" of bacteriophage ϕX174 is hydrolyzed to completion by HaeIII, a series of DNA fragments are produced ranging in size from 72 to 1353 bp. Separation of these fragments by DNA chromatography under nondenaturing conditions yields the chromatogram shown in Fig. 4.

Clearly, in view of the ease with which chromatograms can be analyzed quantitatively, DNA chromatography can be made in the analysis of the kinetic and chemical mechanisms of restriction enzymes and nucleases in general. This experimental approach has been widely used in the study of mechanistic diversity amongst restriction endonucleases, but has until now relied on electrophoretic separation and the use of tritiated plasmid DNA to provide accurate quantitative data; such experiments are greatly facilitated by DNA chromatography.

2.2 DNA Methylation Analysis

Following biological replication, DNA may become modified by methylation, extended by the action of telomerase, damaged by alkylation, interrupted by recombination, or mutated by erroneous

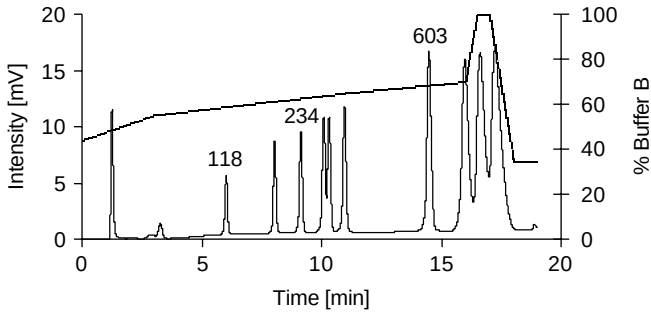

Fig. 4 Size-based separation of bacteriophage ϕX174 DNA following restriction digestion with HaeIII (the numbers above the peaks indicate fragment sizes in base pairs).

repair. This intrinsic plasticity of the genome contributes to the phenotype of a given cell or tissue and paves the way for evolutionary change. Therefore, there has been substantial interest shown in the molecular processes that lead to these and other DNA modifications. For example, the phenomenon of genomic imprinting and some aspects of developmental gene regulation are intimately associated with region-specific cytosine methylation.

Single nucleotide methods have recently been introduced in order to identify those specific cytosines that undergo methylation and demethylation *in vivo*. Such methods are generally based on the differential susceptibility of methylated and unmethylated cytosines to sodium bisulfite (Bis) mediated deamination. Following differential deamination, those bases that were originally methylated remain as Cs, and those that were unmethylated become Us. In the final stage of analysis, which typically involves PCR amplification, methylated C:G base pairs are retained in the PCR product, while unmethylated C:G base pairs are converted to T:A, and nucleotide sequence analysis provides a facile method of distinguishing the two.

Although deamination of cytosine proceeds at a significantly greater rate than deamination of the corresponding methylated form, the reproducibility of the method is often unsatisfactory and "statistical" (i.e. the same fragment must be analyzed more than once to establish a consensus). Still, with proper controls bisulfite sequencing has become the accepted approach to the robust analysis of the methylation status of a given genomic segment. DNA chromatography can provide a rapid means of evaluating the outcome of a deamination reaction. Thus, the deamination of a specific cytosine in the oligodeoxynucleotide homoduplex containing the sequence GAATTC (an *Eco*RI restriction site) can be assessed by the outcome of an *Eco*RI restriction endonuclease reaction as a function of the rate of a bisulfite deamination reaction. If bisulfite deamination is 100% successful, there will be no cleavage of the site since the sequence GAATTU is not a substrate for the enzyme; incomplete deamination is manifested by the appearance of two short duplices. Figure 5 shows the successful outcome of a bisulfite-mediated deamination reaction together with those species expected if deamination is incomplete. This quality control experiment, which is much easier to quantify by chromatography than electrophoresis, can lead

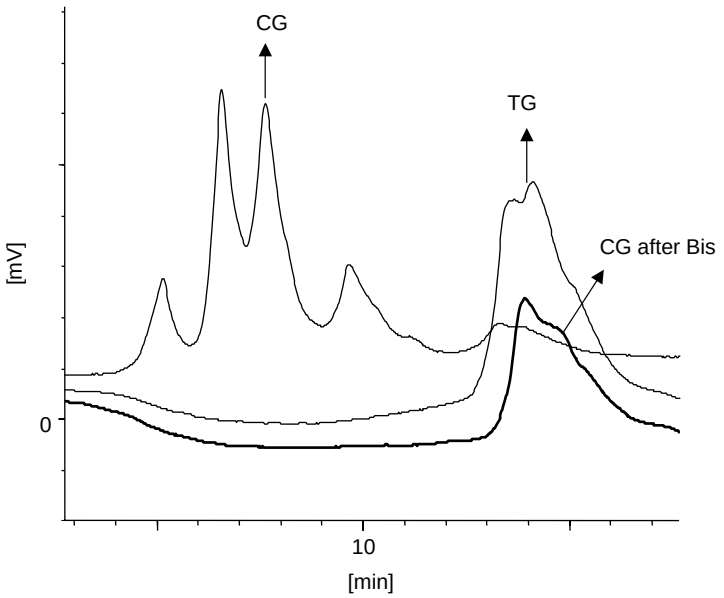

Fig. 5 Indirect monitoring of bisulfite (Bis)-mediated deamination of cytosine containing DNA duplices. An oligodeoxynucleotide duplex containing a unique, asymmetrically positioned *Eco*RI restriction site (GAATTC), was digested with an excess of *Eco*RI before (labeled CG) and after (CG after Bis) bisulfite deamination. As a control, a duplex, in which the *Eco*RI site was mutated to GAATTT (labeled TG) to mimic deamination, was also subjected to denaturing DNA chromatography. By linking restriction endonuclease susceptibility to deamination, this set of DNA duplices can be used to evaluate the outcome of any deamination reaction in a quality control manner. The deaminated strand is clearly distinguished from the cleaved species.

to a significant improvement in the reproducibility of bisulfite reactions when carried out in parallel with deamination reactions on PCR products. This can be important given that the variation in genomic DNA preparations and in the quality of the bisulfite reagents militate against reproducibility. DNA chromatography in conjunction with bisulfite deamination is well suited to the investigation of DNA methylation. Indeed, recent applications of DNA chromatography for the analysis of genomic DNA methylation have been reported that offer an alternative to conventional nucleotide sequencing.

2.3 Phosphorylation Analysis

DNA chromatography represents a generic platform for many aspects of nucleic acid enzymology as exemplified below by assays polynucleotide kinase. However, this approach is not confined to this reaction. Indeed, the ability to perform a variety of nonradioactive assays with throughput times of 10 min per sample in conjunction with in-line data analysis software represents a new direction for DNA chromatography. For example, polynucleotide kinase catalyzes

the transfer of a phosphate group from ATP to the 5′-OH of DNA, RNA, and nucleoside 3′-monophosphates. The enzymatic phosphorylation of 3′ hydroxyl groups of synthetic oligodeoxynucleotides has widespread application in experimental molecular biology. For example, the labeling of oligodeoxynucleotides as probes for identification of target sequences by hybridization is a key technique in molecular cloning. The addition of ^{32}P-labeled phosphate groups to oligodeoxynucleotide sequences using polynucleotide kinase is a common method for creating these labeled probes. Assessment of the proportion of probe that has been labeled and separation of the labeled probe from the unlabeled substrate duplex (or single strand) is normally not undertaken owing to the difficulty of carrying out such procedures.

This subtle ionic difference between phosphorylated and unphosphorylated DNA is readily resolved by DNA chromatography when the nucleic acid fragment is less than 50 bases in length. This can be demonstrated by separating the product from the substrate in a reaction containing an unphosphorylated oligodeoxynucleotide and polynucleotide kinase. Figure 6 shows the effects of phosphorylation of a single-stranded oligodeoxynucleotide upon retention time following DNA chromatography. Phosphorylation of the oligodeoxynucleotide reduces the retention time of the modified DNA, owing to the addition of the polar phosphate group, which diminishes the net hydrophobicity of the DNA. These data are readily corroborated by mass spectrometry. This particular application of DNA chromatography demonstrates not only the utility in molecular biology but also presents, for the first time, a rapid assay method for exploring the biochemical properties of polynucleotide kinases.

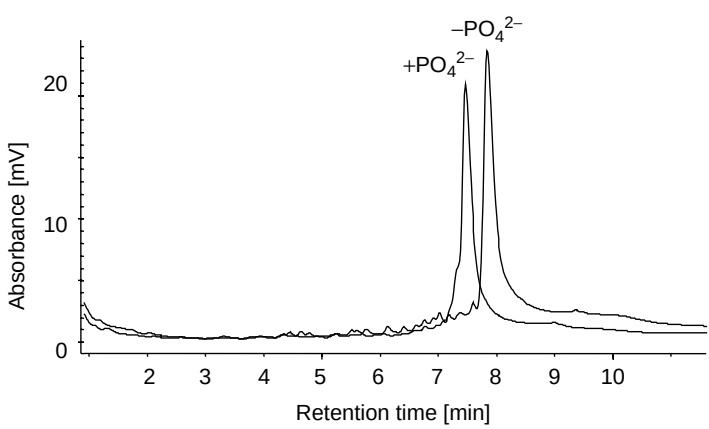

Fig. 6 The use of DNA chromatography in the resolution of differentially phosphorylated oligodeoxynucleotides under denaturing conditions. The difference in polarity between the two species is clearly sufficient to resolve the strands, which may be up to 50 to 100 nucleotides in length. Confirmation of the products was obtained by mass spectrometry, which is a powerful complementary technique to DNA chromatography since the eluted material is ideal for mass spectrometry.

2.4 Footprinting

The regulation of expression and the introduction of structural modifications to the genome rely on proteins that bind to specific DNA sequences or structures. The understanding of how such proteins recognize their binding sites in the midst of a plethora of similar sequences in the genome is an important step to understand such processes at the molecular level. Over 30 years ago, Galas and Schmitz introduced an elegant method called *footprinting* to determine the contacts between a specific DNA duplex and a protein.

Standard footprinting reactions involve the binding of a protein to radioactively labeled DNA, containing the sequence that the protein recognizes. This complex is then digested either enzymatically using DNase I or chemically using hydroxyl radicals. The regions of the DNA molecule covered by the bound protein are protected from digestion while the rest of the DNA backbone undergoes cleavage. The products of the reaction are then separated using electrophoresis; the extent of digestion can then be quantified by further analysis of the gel. A blank region on the autoradiograph is called the *footprint* and this corresponds to the site of the protein binding to the DNA.

Modifications to the standard footprinting reaction that allow the reaction products to be analyzed by DNA chromatography include the use of fluorescently labeled DNA. One of the strands of the duplex contains a 5′-fluorescent group introduced during the synthesis of the DNA. This facilitates the analysis of the cleavage products with the aid of a fluorescence detector at the appropriate wavelength(s). Analysis of the footprinting products is rapid with run-times of approximately 30 min for each sample. Direct quantification of the cleavage products is performed using peak evaluation software. The analysis of a protein–DNA interaction using hydroxyl radical footprinting is exemplified here by the interaction between the RuvA component of the bacterial resolvasome and a synthetic four-stranded Holliday junction.

Before discussing footprinting data, it is necessary to consider the choice of ion pair reagent and the methodology associated with "phasing" the footprint; that is, relating the chromatogram obtained in the presence of the protein to the nucleotide sequence of the DNA duplex. The use of tetrabutylammonium bromide (as the ion-pairing reagent) with DNA chromatography is essential for the (almost or near) size-dependent separation of fluorescently labeled DNA. This regime removes the influence of the hydrophobic fluorescent group and any residual sequence-specific effects. The sequencing reactions performed involve an express protocol for the generation of a "GA ladder". This method does not provide complete sequence information; instead, it produces cleavage at every G and A residue within the oligodeoxynucleotide strand. However, the Maxam–Gilbert sequencing protocol for complete sequence identification may also be performed, as the chemistries are very similar. Simultaneous nucleotide sequencing is required to allow complete identification and phasing of all the peaks in the footprinting reaction. The footprinting reaction can generate over 40 chromatographic peaks, and it is important to correctly identify each peak (which can often be difficult in the presence of an injection and salt peaks at the start of a chromatogram). Overlaying the GA sequence ladder allows complete characterization of the footprinting reaction once

the sequence of the GA ladder has been determined. Some small peaks arise in the sequencing reaction due to small amounts of cleavage at TA residues; however, these anomalous products can be minimized by optimizing the reaction conditions.

The Holliday junction and Holliday junction-RuvA complexes that had been subjected to hydroxyl radical cleavage were analyzed under denaturing conditions at 75 °C using DNA chromatography (Fig. 7). The chromatogram of the hydroxyl radical

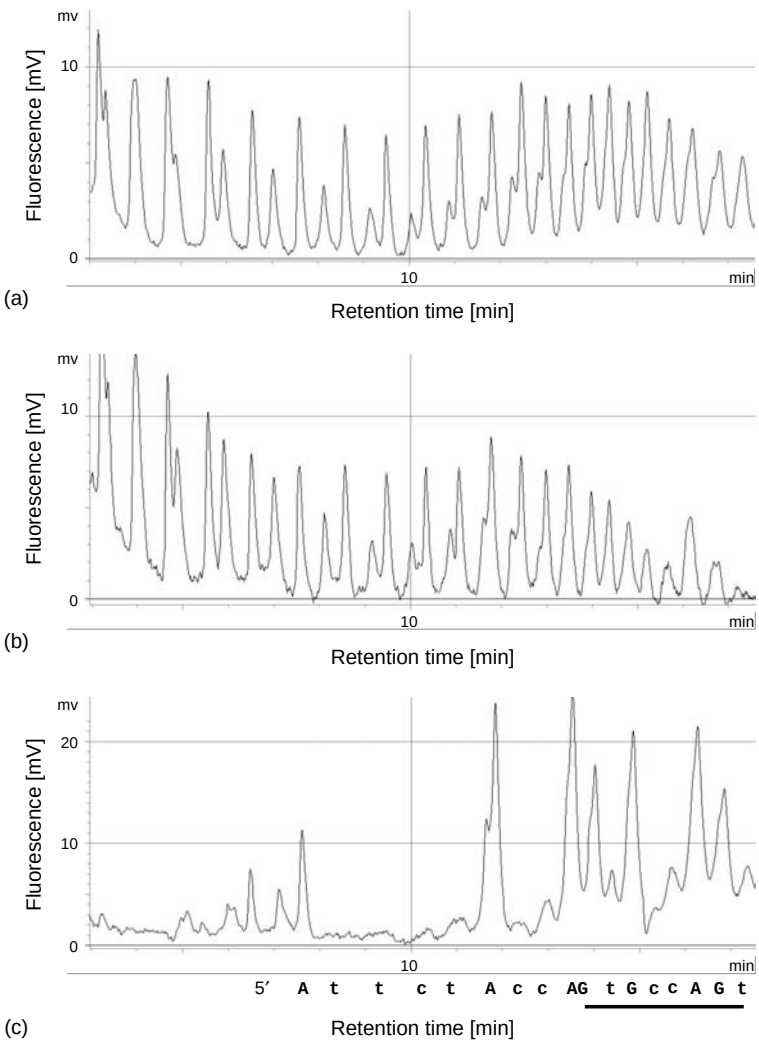

Fig. 7 The comparative analysis of a DNA four-way junction in (a) the absence of the bacterial recombination protein RuvA, (b) in the presence of the protein, and (c) a corresponding Maxam–Gilbert G + A reaction. This experiment shows that hydroxyl radical cleavage is inhibited by the presence of the RuvA protein (solid line underneath chromatogram c, which is an indirect measure of the specific bases that are in contact with the protein).

cleavage products from the Holliday junction in the presence of RuvA compares sharply with the cleavage products generated in the absence of the protein (compare Figs. 7a and 7b). Using peak analysis software, the direct quantification of the cleavage products can be achieved by determining the areas under the peaks in each chromatogram. This technique also facilitates the identification of the individual peaks by alignment with fragments from a GA sequencing ladder (see Fig. 7c). Previous DNase I footprinting experiments have shown that RuvA binds to the Holliday junction, producing a symmetrical footprint, extending approximately 13 bases on either side of the crossover point. Hydroxyl radical footprinting analysis of the RuvA-Holliday junction complex using this novel fluorescent-based analysis determined a more precise footprint, covering approximately 9 to 10 bases either side of the crossover point. The results from the hydroxyl radical footprinting experiment is in excellent agreement with crystal structures of the RuvA-Holliday junction, which show a tetrameric RuvA extending over 8 bases of the crossover point of the Holliday junction.

3 Denaturing High Performance Liquid Chromatography (DHPLC)

Denaturing HPLC, or what is often referred to as DHPLC, is the single most published-upon application of DNA chromatography. Its growth in popularity is superbly documented within reviews dedicated to the topic of DHPLC as a whole. DHPLC has been used for the analysis of hundreds, if not thousands, of different genes within the human and other genomes for their mutations.

The reason DHPLC has become so popular is because of the questions it answers. In short, DHPLC is capable of indicating whether a genetic sequence up to 1 kb in length carries any sequence alteration relative to some reference sequence. There are a number of reasons for gaining this type of information, such as the discovery of anonymous common sequence variants for performing genetic association studies, discovery of potential common functional variants within polygenes, or screening for detection of disease-causing mutations within inherited disorders. As demonstrated by the review articles cited above, all of these areas have contributed to the substantial growth in DHPLC's popularity and general acceptance.

This section on DHPLC begins by describing the basis of the technique's function and operation. This is followed by more detailed descriptions of the factors that must be attended to for successful DHPLC analyses (such as DNA melting behaviors and optimization of PCR conditions). A review is given of the milestones within the development of DHPLC, as well as introducing recent developments and special techniques within DHPLC.

3.1 The Basis of DHPLC

DHPLC is based on the detection of heteroduplex DNA in the presence of homoduplex DNA. A pair of primers is designed to generate a PCR product that is typically less than 1 kb in length, which spans the sequence region of interest. Individuals who are, for instance, heterozygous for a single nucleotide polymorphism (SNP) have a 1 : 1 ratio of wild-type and mutant DNA. Heating the PCR product to 95 °C and cooling slowly hybridizes the material

to form a mixture of heteroduplex and homoduplex species (Fig. 8). In cases where the individual possesses a mutation in both alleles, (a homozygous mutation), known wild-type PCR product is mixed with the sample PCR material and then the mixture is hybridized to form the heteroduplex and homoduplex species.

The heteroduplex contains a physical "bubble" of the two strands at the mutation site caused by a disruption in the hydrogen bonding at the site of the mismatch. The fundamental process governing the separation of heteroduplexes from homoduplexes in DHPLC is an increase in a heteroduplex fragment's overall single-strandedness or an increase in polarity of the fragment, resulting in reduced adsorption on the separation medium's surface. A larger proportion of the hydrophilic "core" of the DNA double helix is available for hydrogen bonding to the aqueous component of the eluent. The heteroduplex more easily desorbs from the column bead surface and is solvated by the eluent (as compared to its less single-stranded, homoduplex counterpart). At elevated temperature, this mismatch bubble becomes enlarged relative to the corresponding homoduplex. If the temperature is increased too much, all duplex species will completely melt into single strands and will not be separated from each other. Under ideal conditions (choice of primers, column temperature, fragment characteristics, etc.), four peaks will be detected – two each for each duplex (Fig. 9). More likely, only two or three peaks (either partially or fully resolved) will be detected, thus indicating the presence of a mutation.

3.2 DHPLC and Its Practice

3.2.1 DNA Melting Fundamentals

Successful practice of DHPLC lies in understanding and exploiting of DNA melting processes. DNA melting is a cooperative process. An often-invoked analogy for the description of DNA denaturation is a line of interconnected, buoyant pontoons floating on the surface of a body of water. If one of the pontoons is forced below the water's surface, then the others will also be forced below the surface as a result of their being interconnected. In other words, the buoyant bodies are "cooperatively" floating or submerged. In most cases, DNA denaturation occurs in much the same way. If the hydrogen bonding of two base pairs in the DNA double helix is disrupted as the result of any particular process (e.g. the presence of heat, chemical denaturants, mismatched base pairs, or combinations of any of these forces), the bases surrounding them are correspondingly disrupted

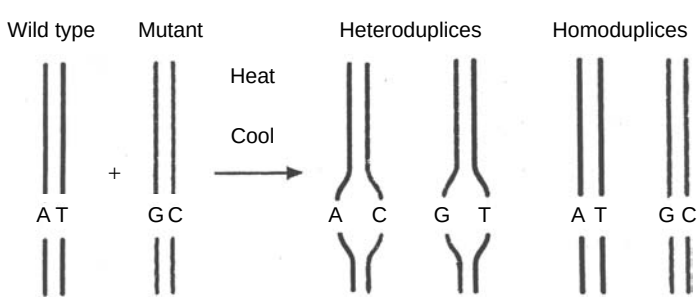

Fig. 8 Two heteroduplex species and two homoduplex species are formed when a mutation is present.

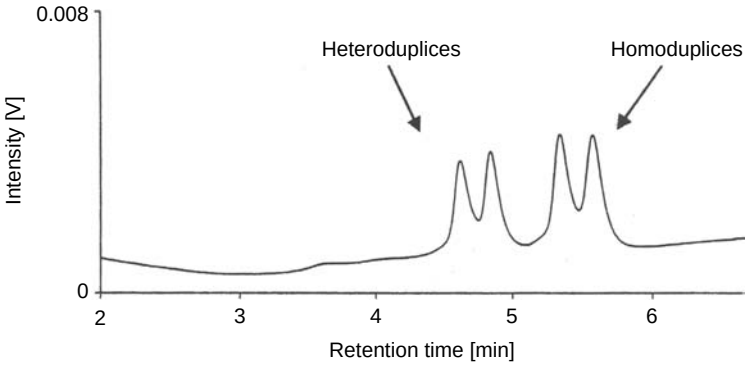

Fig. 9 Under ideal conditions, DHPLC will produce a four-peak pattern: two for each heteroduplex complex and two for each homoduplex. Usually, heteroduplices will not be completely resolved, and only a two- or three-peak pattern will be observed.

and the segment of DNA denatures as a whole. In other words, the DNA denatures "cooperatively." The segment in which cooperative melting occurs is referred to as a *domain* and is typically anywhere from 50 to 300 bases in length. Domains that are rich in G – C nucleotide hydrogen bonding melt at higher temperatures than domains that are rich in A – T hydrogen bonding. The detailed processes that govern the formation of domains, as well as the chemical processes involved in their melting, have been described in detail elsewhere.

As discussed, a normal (homoduplex) fragment is composed of domains that melt cooperatively at different temperatures. The mismatch within the heteroduplex will also act in a domain manner by depressing the melting temperature of the entire domain in which it resides and not just the sequence in direct contact with the mismatch site. Under the correct analytical separation conditions, the mismatch-containing domain within the heteroduplex will be considerably more denatured than its homoduplex counterpart. As stated above, this increase in single-strandedness for the heteroduplex means that its ability to remain on the surface of the resin is significantly reduced relative to the homoduplex. In other words, the destabilizing effect of the base mismatch is effectively exaggerated or "amplified" as a result of cooperative domain melting behavior. Collectively, this gives rise to separation patterns (chromatograms) that are largely more complex (either have more peaks or at least has a much broader peak) in cases where a mutation is present, compared to when a mutation is not present.

It is possible to select the DHPLC column oven temperatures without considering the effects of cooperative domain melting behaviors. If a homoduplex fragment peak's chromatographic retention time is plotted as a function of the column oven temperature, one will often observe a well-defined sigmoidal plot. The fragment may have more than one domain due to the different slopes of the inflection of the melt curve. However, a plot of this type usually shows a single curve extending over the melt region. The user may choose a run temperature indicated by the midpoint or multiple temperatures,

depending on the length and shape of the curve.

DNA melting predictions DHPLC temperatures can be selected by modeling the fragment's melting behavior in advance. The ability to predict a fragment's melting behavior *ab initio* has more than one benefit associated with it. For one, it is possible to predict a single set of analytical conditions for mutation detection. In a highly ambitious study that involved DHPLC mutation analysis of many hundreds of different fragment sequences, investigators determined that a single DHPLC analysis temperature provided 87% mutation detection sensitivity, with no attention being given to multidomain melting behaviors or PCR product design whatsoever.

The second benefit associated with melting prediction is that it provides the means necessary for obtaining virtually 100% mutation detection sensitivity. As noted above, a multidomain fragment can result in less reliable determination of temperature, since more than one melting domain can result in a bias to the predicted "optimum temperature." However, with *a priori* knowledge of more than one melting domain, it is possible to avoid improper selection of the analysis temperature. Shown in Fig. 10 are two plots for a particular sequence 300 bp long. The first plot shows "Helical fraction (%)" as a function of "Temperature." The second plot shows the melting "Temperature" as a function of base position, and is well known to practitioners of DGGE as a "melt map."

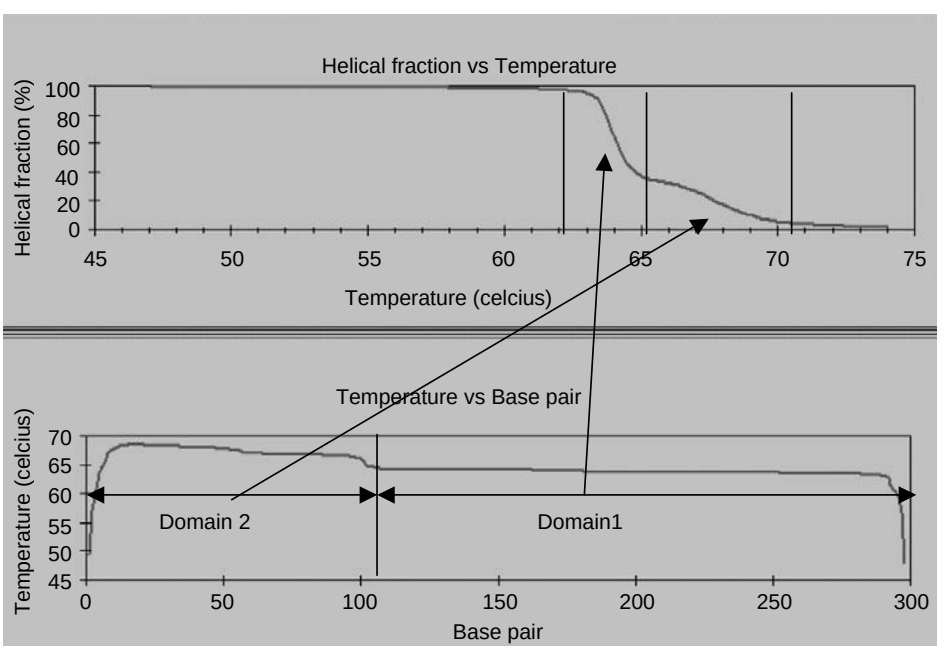

Fig. 10 Correlation of predicted thermal titration data (shown at the top of figure) and melt map features (shown at the bottom of the figure). Predictions performed with WAVEMaker® software, Transgenomic, Omaha, NE.

These algorithms have been optimized to predict melting behavior in solution states different from those commonly applied in DHPLC. For the purposes of DHPLC, these predictive algorithms must be adapted to reflect the presence of acetonitrile (an organic solvent) and triethylammonium acetate (a salt and ion-pairing reagent) in the separation environment (a nonpolar surface), since both of these components have significant effects on fragment melting behavior. An organic solvent will decrease the fragment melting temperature, a salt will increase temperature, and in the presence of a ion-pairing reagent, a nonpolar surface will increase the temperature. The extents to which these components influence fragment melting have been described in detail. DHPLC algorithms of this nature are available commercially through WAVEMaker® software (Transgenomic, Inc.) or via the Internet (http://www.insertion.stanford.edu/melt.html). Once these operations are performed, it is possible to accurately predict the analytical conditions required for complete DHPLC analyses: the analytical separation gradients are generated, and the analytical temperature is predicted in the manner described above. This general predictive approach to DHPLC analyses has been validated on numerous occasions. In summary, it has been shown that this approach can detect anywhere from 87% of mutations (single temperature selection with no fragment or PCR optimization whatsoever) to 100% of mutations (one or more temperatures selected with fragment optimization).

PCR optimization Temperature prediction alone will not always provide conditions that are fully optimized for the detection of every possible mutation. This is particularly the case when attempting to detect mutations located in high-melting domains within a fragment. This situation is addressed through PCR primer optimization. In the context of DHPLC, what is meant by PCR primer optimization is the selection of primer sequences that give rise to fragments whose melting characteristics are propitious with respect to the detection of mutations within a particular sequence. Of course, any primers that are optimized for DHPLC must also be appropriate for PCR amplification.

These points are well illustrated in the example shown in Fig. 11. These are melt maps for the same mutation, but with different primers selected. It is clear from map #1 that the mutation resides in a low-melting domain, and that the vast majority of the fragment (roughly 75%) is unmelted at the temperature that corresponds to the domain that possesses the mutation. In map #2, the mutation also resides in a low-melting domain, but a smaller proportion of the fragment (roughly 40%) is unmelted at the temperature that corresponds to the domain that possesses the mutation. In map #3, this proportion is reduced even further (roughly 15%). Important to note here is that the fragment size is identical in each case, and all that varies is the proportion of unmelted fragment at the mutation domain's melting temperature.

Figure 12 shows the consequence of these differences. The separation shown in Fig. 12(a) is not only sufficient for baseline resolution of the heteroduplices from the homoduplices but also for the separation of these duplices from one another. In Fig. 12(b), the resolution is clearly diminished, but still allows for highly reliable detection of the mutation. In Fig. 12(c), the heteroduplex is barely resolved from the homoduplex. There still is enough resolution for detection of the mutation, especially when using

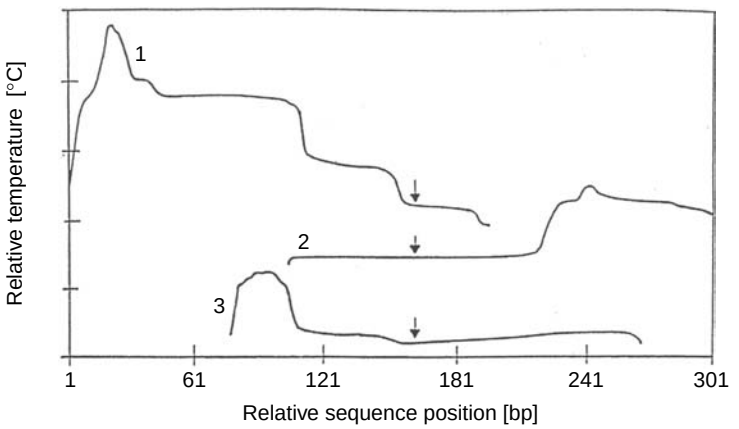

Fig. 11 Melt maps for the same p53 mutation with different primer locations selected relative to the mutation location (Gjerde, D.T., Taylor, P.D., Haefele, R.M. (2001) *Mutation Detection Method*, U.S. Patent 6,287,822 with permission).

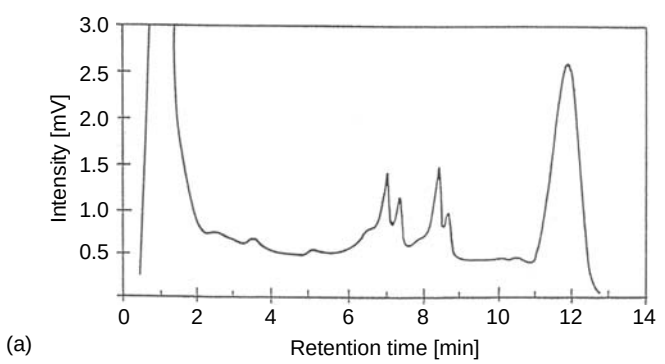

Fig. 12 (a) DHPLC chromatogram of fragment 1 of Fig. 4.5. (Gjerde, D.T., Taylor, P.D., Haefele, R.M. (2001) *Mutation Detection Method*, U.S. Patent 6,287,822 with permission). (b) DHPLC chromatogram of fragment 2 of Fig. 4.5. (from Gjerde, D.T., Taylor, P.D., Haefele, R.M. (2001) *Mutation Detection Method*, U.S. Patent 6,287,822 with permission). (c) DHPLC chromatogram of fragment 3 of Fig. 4.5. (from Gjerde, D.T., Taylor, P.D., Haefele, R.M. (2001) *Mutation Detection Method*, U.S. Patent 6,287,822 with permission).

computer analysis, the detection is clearly more difficult.

In all three cases, the mutation is situated in a low-melting domain, which is why it is successfully detected in all three cases. However, our degree of analytical confidence is always greatest when the resolution for any given mutation is at its maximum. Therefore, maximization of detection confidence is achieved through

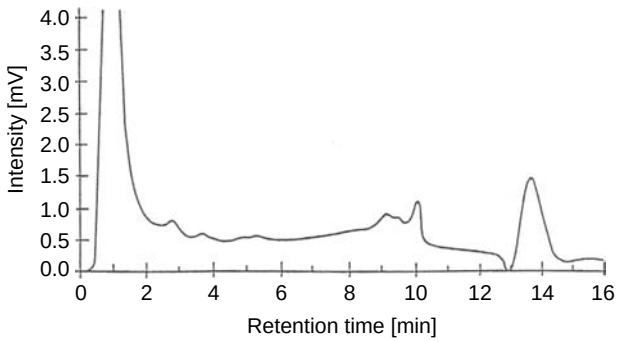

(b)

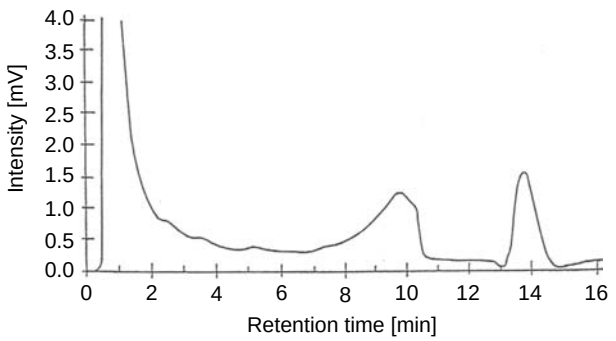

(c)

Fig. 12 (*Continued*)

proper primer optimization strategies that situate a sequence region for scanning in a low-melting domain with a high proportion of unmelted sequence adjacent to it, which is demonstrated here and elsewhere. The needs for the level of confidence must always be balanced against the detection productivity of the procedure. In other words, maintaining the absolute highest level of confidence at all times may necessitate analyzing numerous overlapping fragments for the presence of mutations, whereas fewer fragments can be scanned when the confidence level need not be at the absolute highest levels.

Even in cases where the confidence level can be relaxed, there are still circumstances where primer optimization does not yield conditions that allow reliable mutation detection. This is frequently the case when the region being scanned is in a high-melting domain, and no amount of primer optimization can render this domain more amenable to mutation detection (or, the scanned region is in a high-melting domain, and it is not possible to move the primer locations). A common means of overcoming this situation is the addition of a nontemplate GC-clamp to one of the primers used for amplification, which has been shown to work successfully in DHPLC. By doing so, one creates an additional high-melting domain adjacent to the region being analyzed for mutations. This general approach has been applied for over a decade in the context of DGGE, and is largely compulsory for the successful application of that technique.

However, in DHPLC, GC-clamps need only be applied in extreme circumstances such as those described above.

The detection of mutations by DHPLC requires the chromatographic separation of heteroduplices from homoduplices. The mutant sequence present in the template is carried through the PCR amplification process, the heteroduplices are subsequently formed through standard hybridization protocols, and the separation is commenced. While this is a generally straightforward procedure, one must always be on guard for the potential introduction of confounding artifacts brought about by the analytical procedure itself. One part of the DHPLC analytical technique where artifacts may be introduced is within the PCR process itself. PCR-induced mutations or amplification errors of sample can make detection of the sample mutations difficult or even impossible.

Differences in PCR fidelity, and how they present themselves in the DHPLC signal profiles, are shown in Fig. 13. For a given set of amplification conditions with a homozygous template, it is clear that the nonproofreading AmpliTaq® polymerase results in a significantly higher relative amount of background heteroduplices, whereas the proofreading *Pfu*® and *Pfu*Turbo® provide significantly reduced levels of this background. It is clear from these examples that the AmpliTaq®-derived signal could lead to misclassification of the sample being analyzed, whereas this is less likely for the higher fidelity results. Properly optimized amplifications can minimize the risk of misclassification with nonproofreading enzymes, while proofreading enzymes will always minimize the risk posed by the presence of background heteroduplices. The need for the lowest possible levels of background heteroduplices is most pronounced when attempting to detect mutant sequences at concentrations significantly lower than the 50% found in germline mutation detection situations.

3.2.5 Other DHPLC Recommendations

There is a strong need within DHPLC to have a system that is substantially free of any extraneous metal contamination. Any free transition metals in the

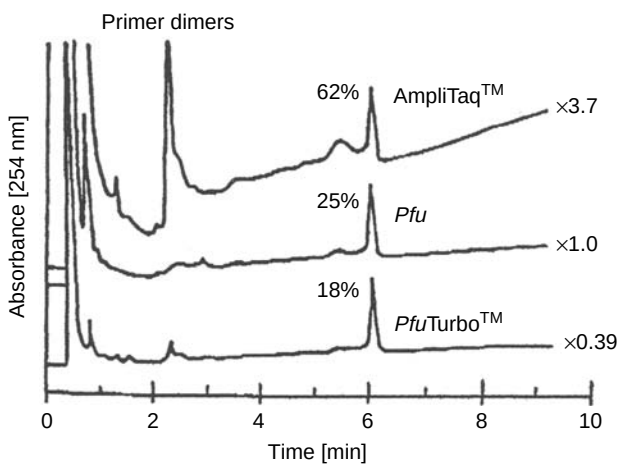

Fig. 13 Background heteroduplex signals generated as a result of PCR infidelities. The background signal is reduced upon application of a high-fidelity PCR polymerase.

system (such as iron) can compromise the analysis by interfering with efficient and precise interaction of the DNA fragment with the column material. While this has been observed for nondenaturing, size-based separation of DNA fragments with DNA chromatography, it is even more pronounced within DHPLC. Straightforward procedures are available for ensuring that metal contamination is eliminated, along with addressing other frequently asked questions (FAQs) regarding successful performance of DHPLC analyses.

3.3 Review of DHPLC Development

The first peer-reviewed publications that made use of DHPLC were in 1996. While these reports did not provide any significant information regarding DHPLC operation or optimization of its conditions, the power of the technique was clearly demonstrated by the speed and ease by which the analyses were performed for the PCR products generated.

The lack of detail regarding DHPLC was addressed in a number of publications in late 1996 and 1997. One paper dealt with the issue of DHPLC technique validation, while another paper mentions the critical issues surrounding fragment melting and their effects on heteroduplex retention by minimizing the stabilizing effects of G–C-rich regions.

With these beginnings for DHPLC, 1998 was a watershed year for the development of the technique as a whole. There was continued characterization of the technique through validation studies. In addition, there begins to be a much more detailed treatment of fragment melting behaviors with the publication and description of an internet-accessible algorithm that allows for prediction of DHPLC analysis temperature.

This work on elucidating the role of the various DHPLC analytical conditions continued in 1999. One paper gave a detailed description of melting processes and temperature predictions, and the principles behind this approach were applied to the analysis of *BRCA1* and *BRCA2* genes. Other publications in this same year made important empirical observations on the relationship between overall fragment melting patterns and the ability to detect mutations within DHPLC.

This enhanced understanding of DHPLC fundamentals and phenomena has led to the general acceptance of the technique, which is exemplified by the many publications that apply DHPLC as a matter of course to the analysis of many genes in many different contexts. This general degree of acceptance within the peer-reviewed literature began in 1999 and has continued to the present day. It is beyond the scope of this chapter to provide a complete listing of these publications.

The last two years have seen the extension of DHPLC's capabilities in new directions. One area is that of determining minority allele populations, which is relevant in various contexts (such as in detection of mutations within cancer cells present in a normal sequence "background"). The optimization of DHPLC conditions for this type of analysis has been described recently, as well as using DHPLC for the purification, collection, and subsequent amplification of mutant sequence that cannot even be detected within standard schemes.

There have also been developments in the nature of the chromatographic

separation media. The use of monolithic capillary column materials (as opposed to packed particles) has shown to be effective at improving separation performance and miniaturizing analyses. Furthermore, this DHPLC format has been used for performing multiplexed/"capillary array" separations, as well as being an effective separation means for interfacing to mass spectrometry. Developments within monolithic capillary columns for DHPLC have also been reviewed recently.

4 RNA Chromatography

4.1 RNA Structure

RNA molecules are central to the expression of genetic information in all living organisms. The product of transcription, messenger RNA (mRNA), carries the genetic instructions for decoding genetic information to the ribosome where protein synthesis takes place. In addition to these classical roles in cell biology, RNA is emerging as a key regulator of gene expression and has been shown to be capable of catalyzing a range of chemical reactions both *in vivo* and *in vitro*.

In contrast to double helical DNA, RNA molecules are much more structurally diverse. X-ray crystallography and multidimensional NMR have revealed the globular nature of a number of small biological RNA species, and it is likely that most mRNA molecules are even more complex in structure.

Our relatively slow accumulation of biochemical knowledge of RNA is largely a consequence of the inadequate methodologies available for its isolation, stabilization, purification, and characterization relative to RNA's biological partner, DNA. Indeed, while chemical synthesis of 100-mer oligodeoxynucleotides is routine, the synthesis of 30-mer oligoribonucleotides is still something of a synthetic challenge.

Our knowledge of RNA structure and function is largely extrapolated from structure and function studies on transfer RNA. However, with the development of high-yield enzymatic methods for the synthesis of RNA through the combination of gene cloning and *in vitro* transcription, the ability to carry out more rigorous biochemical and biophysical investigations of RNA is now becoming possible, and the need for both analytical and preparative chromatographic methods for RNA analysis are pressing.

One important aspect of translating human genomics into human physiology will involve the global analysis of whole cell populations of mRNA in disease and health, and the modulation of these populations during both natural physiological transitions and in response to drug administration. This aspect of RNA population biology is largely being addressed by microarraying technology and to a lesser degree by various RTPCR (reverse transcriptase polymerase chain reaction) approaches, including differential message display and serial analysis of gene expression. These forms of analysis of cellular transcriptomes typically produce high volumes of data and treat all RNAs as equivalent molecular species (similar in principle to two-dimensional SDSPAGE of proteins in proteome biology). The majority of DNA-DNA hybridization experiments can be considered as uniform Watson-Crick duplexes. The relationship between the function and structure of RNA lies somewhere between DNA and proteins. Therefore, hybridization-dependent

transcriptome approaches provide only a preliminary, and therefore limited, insight into the population biology of RNA at the cellular and tissue levels.

Here we present an overview of progress in adapting methods originally developed for the chromatography of DNA for the separation, stabilization, and analysis of RNA; this will be referred to as RNA chromatography. In the absence of an all embracing physical explanation for the separation of RNA species by RNA chromatography, we shall cautiously assume that the mechanism of separation is very similar under most conditions to that observed for DNA, but we shall add the caveat that there may well be different physical principles responsible for similar experimental outcomes.

RNA is a polymer of nucleotide units, which comprise of mainly pyrimidine (cytosine and uracil) and purine (adenine and guanine) bases, the sugar D-ribose and phosphoric acid. The nucleotides in RNA, as in DNA, are combined through phosphodiester linkages between the 5′ hydroxyl of one sugar and the 3′ hydroxyl of another. These polymer chains are not typically found as paired duplexes, but rather as complex forms that comprise paired segments and loops, which are often stabilized by metal ions. The most familiar structural element in an RNA molecule is probably the stem loop, in which a noncomplementary segment separates two complementary stretches of nucleotides. The resulting structure that often folds is a Watson–Crick-like duplex (the stem) closed off by an unpaired segment (the loop). Such stem loop modules are used routinely to display predicted secondary structures of complex RNA molecules: perhaps the most famous of which is the classical cloverleaf structure of tRNAs.

The conformation of the duplex segments (stems) in RNA is confined to the A form, owing to the steric consequences of the presence of a 2′ hydroxyl group on ribose ring (DNA, which is predominantly a B-form duplex *in vivo*, has a hydrogen atom at this position) and differences in sugar pucker. As an added structural consequence of the presence of the 2′ hydroxyl group, RNA can participate in an extended repertoire of intra- and intermolecular hydrogen bonding interactions.

One structural element that demonstrates the clear difference between RNA and DNA is the RNA pseudo knot, in which an antiparallel complementary duplex arises when the complementary sequences are separated by a sufficient length of chain. The structure takes the form of a handwritten letter, $\mathcal{L}$, the upper and lower complementary strands being linked by a pair of 3′ and 5′ loops.

This structural diversity, characteristic of biological RNA, therefore has significant consequences for the physical properties of these molecules when attempting to rationalize the chromatographic behavior of RNA. Chromatographic separation of RNA, using the same protocols as used in DNA chromatography will therefore be a function of parameters that include molecular weight (chain length), surface chemistry, and localized single- and double-stranded character. It is therefore likely that unless a given population of RNA molecules is completely free from secondary and tertiary interactions, by thorough thermal denaturation, for example, they will behave anomalously during chromatography compared to duplex DNA. These issues are addressed below after considering some practical aspects of RNA extraction and stabilization.

4.2 Biological Extraction of RNA

RNA is typically extracted from cells under strongly denaturing conditions since cells (and hands) contain ribonuclease enzymes, all of which will tend to disturb the original population of RNA species isolated from the cell after an unavoidable delay. All ribonucleases have a pair of histidine side chains in the active site that drive catalytic hydrolysis. These enzymes generally catalyze hydrolysis of the phosphodiester linkage by deploying suitably ionized histidine residues. Two types of strategies are generally used to overcome the deleterious effects of RNases during RNA preparation: addition of general protein denaturants such as urea and guanidine hydrochloride, or by addition of irreversible chemical modification (specific for ionized histidine side chains) agents such as diethyl pyrocarbonate. The key to the successful isolation of high quality RNA in good yield, however, remains largely a function of experimental experience.

When RNA from cell extraction is electrophoresed in agarose, the most abundant species are the two major rRNAs that are referred to as 16S and 23S in *Escherichia coli* (corresponding to 1542 and 2904 nucleotides, respectively) or 18S and 28S in most eukaryotes (corresponding to 1874 and 4718 nucleotides, respectively). The other major RNA molecules include a species-specific tRNA population comprising around 100 molecular species of 60 to 90 nucleotides in length, a 5S (120 nucleotides) rRNA species (5.8S, 158 nucleotides in mammals) and a small number of spliceosomal RNAs. In contrast, the population of mRNA molecules (the cellular transcriptome) is extremely diverse, ranging from several hundred to tens of thousands of nucleotides in length. Moreover, given that there appears to be around 40 000 separate genes in the human genome, with the potential for expression of a similar number of mRNAs in any given cell, the population of cellular mRNAs is clearly the most complex of the RNA classes. The protein products derived through translation of the transcriptome determine cellular phenotype. It is a result of this large complex population of mRNA molecules that a characteristic background smear is observed underlying the two dominant ribosomal species, when the products of an RNA extraction are submitted to agarose gel electrophoresis and then viewed using ethidium bromide fluorescence.

Given the complexity of the RNA population from even the simplest of organisms, it is critical that measures be taken to preserve that population prior to separation (or indeed any form of analysis). Since the solvent used to displace DNA during DNA chromatography and, as we shall see, RNA chromatography is acetonitrile, the integrity of RNA must be preserved in relatively high concentrations of the solvent and at elevated temperatures (75 °C). That this is the case was established by simply demonstrating that an RNA species exposed to a high concentration of acetonitrile retained its biological activity. Interestingly, when the effect of acetonitrile upon the activity of a typical ribonuclease enzyme was investigated at a range of solvent concentrations and temperatures, a marked effect upon the stability of the enzyme was observed. Thus, at 60 °C in 20% acetonitrile bovine, RNase is reversibly unfolded, whilst at 60% acetonitrile at the same temperature, the enzyme is unfolded, but in an irreversible manner. This observation suggests that RNA stabilization, or rather

elimination of the deleterious effects of contaminating RNases, will be afforded by the standard conditions employed in RNA chromatography.

In addition to the role of acetonitrile as a solvent in promoting RNA stability through RNase inactivation, the chromatographic separation of RNA isolated from tissues (as discussed below), eliminates RNase (as exemplified by bovine RNase) at low acetonitrile concentrations. Thus, as a consequence of the chromatography, RNase may be eliminated at least from the larger, more complex RNA species recovered from a biological extract.

4.3 Size-based RNA Separation

Figure 14 shows the application of RNA chromatography to effect the separation of a spectrum of RNA size standards that are commonly used in agarose gel electrophoresis. It appears at first glance that under similar denaturing conditions employed for the separation of single-stranded DNA, the RNA species are recovered in a size-dependent manner. However, as discussed earlier, the denaturation of double-stranded DNA (or indeed the chromatography of single-stranded DNA species) can produce anomalous retention times data. The additional peak in the chromatogram shown in the figure is a 451 nt RNA species that forms part of the human telomerase enzyme. The telomerase RNA elutes between the RNA size standards of 200 and 500 nt, somewhat earlier than would be ideally expected and with an asymmetrical appearance.

If one attempts to reconcile the chromatographic behavior of RNA (under these experimental conditions), with its structure and chemistry, there are a number of immediate inconsistencies. Size-based separation of double-stranded DNA is made possible by the structural uniformity and the shielded bases of the DNA double helix. Thus, the suppression of nucleotide sequence-specific hydrophobicity is achieved by forming an ion pair with the phosphate group and an ion-pairing

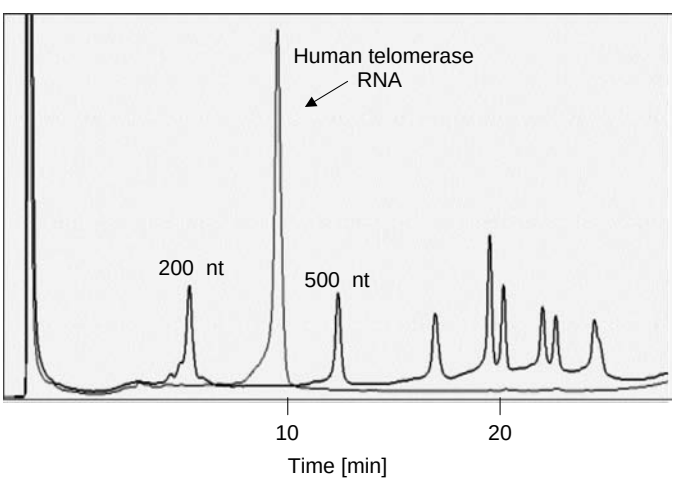

Fig. 14 Analytical separation of RNA size standards by RNA chromatography. The 451-nucleotide human telomerase transcript elutes between the 200- and 500-nt species as indicated.

reagent to establish the dominance of the tetraethyl moiety. The hydrophobic nature of the resulting molecule is responsible for the principal analyte–stationary phase interaction. Ion pairs of single-stranded DNA do not shield the more polar bases so separations of these molecules are not only strictly size dependent but are also based on polarity of the single strand.

Another example of the effect of the single-strand nature of a nucleic acid molecule on its retention in DNA chromatography is shown by the separation of homoduplex from heteroduplex DNA species. The increased single-strand character of a heteroduplex leads to a more polar molecule, which, in turn, reduces the retention time for the heteroduplex. This difference in retention time can be optimized by the manipulation of the temperature of the chromatography run.

Any DNA duplex can be considered as a negatively charged cylinder with a periodic distribution of hydrophobic patches. In some exceptional circumstances, as discussed above, polarity suppression is inadequate and this can prove useful, for example, in the enhanced retention of T-rich DNA following a bisulfite deamination and PCR in the analysis of DNA methylation. In the case of the separation of RNA species at elevated temperatures, it is instructive to note that electrophoresis of RNA under denaturing conditions (i.e. in the presence of urea or formamide gels) may not abolish all secondary and tertiary intramolecular interactions. Therefore, while some RNAs will be predominantly double stranded, some will be partly folded, and others still will be unfolded. In other words, it is impossible to predict what influence will be exerted by temperature and solvent parameters over the chromatographic behavior of complex RNA populations. Moreover, even solutions of single molecular species should not be assumed to be conformationally homogeneous. In the language of biological NMR spectroscopy, each RNA molecule will exhibit an ensemble of conformations defined by temperature, solvent chemistry, ionic strength and, in particular, the presence or absence of divalent cations. Ultimately, any given RNA species like a polypeptide chain will adopt its lowest free energy state(s), a process that is thermodynamically driven by the nucleotide sequence of that particular molecule.

Therefore, in considering the mechanism of size-based separation, a structure-specific factor must be considered. In addition, owing to the single-stranded nature of most RNA species (notwithstanding the formation of intramolecular double strands), there will be an unpredictable diversity in form in a biological RNA sample, making it very difficult to elaborate any rules for RNA chromatography. In fact, the separation of RNA in a size-based manner has much in common with gel filtration chromatography, perhaps in an analogous way that one can compare DNA chromatography in the size-based mode with SDS PAGE in the separation of polypeptides.

In the simplest description, there will be two modes of chromatography taking place when RNA is separated at, say, 75 °C. One of these components is equivalent to a denaturing mode (akin to the chromatography of single-stranded DNA) DNA chromatography, and the second will be a specialized form of hydrophobic interaction chromatography as the ion-pairing reagent produces a constellation of hydrophobic patches over the surface of folded and partially folded, complex RNA forms.

4.4 Separation of Cellular RNA Species

When a total RNA preparation is subjected to RNA chromatography, the resulting chromatogram is typified by that shown in Fig. 15. There are several distinctive features within the profile (obtained in this case at 75 °C) that are found in most biological preparations of this kind. The earliest eluting species include the population of tRNAs (and probably includes small nuclear RNAs), the middle section of the profile is dominated by the rRNA species, and finally underlying the entire chromatogram is a population of mRNAs centered on the later retention times. Corroboration of the above has been obtained in the case of the tRNA fraction by comparison with commercial tRNA preparations, for mRNA by RTPCR (see below) and for rRNA, by selectively enriching the polyadenylated mRNA population. Indeed, it is relatively easy to determine the retention time of any RNA species (from an organism whose genome sequence is available) by RTPCR, followed by nucleotide sequencing of the amplified product.

4.5 Separation of Messenger RNA from Ribosomal RNA

The expression of cellular phenotype is a direct result of the protein complement of a cell. The latter is, in turn, closely, but not inextricably, linked to the population of mRNA molecules produced by selective transcription of the genome in a given cell type. For this reason, there has been considerable interest in evaluating, both in a qualitative and quantitative manner, the products of transcription. The initial experiments in this field were gene specific, but more recently methods have been introduced that facilitate population-based experiments that often utilize microarray technology.

One of the key experimental procedures that forms the prelude to both of the above experiments is the systematic removal of rRNA from the mRNA (polyadenylated) fraction. Typically, the polyadenylated fraction is sequestered via an oligoT chromatography column or batch resin, the rRNA and tRNAs (together with the nonpolyadenylated mRNAs) are discarded and, after a suitable washing protocol, the

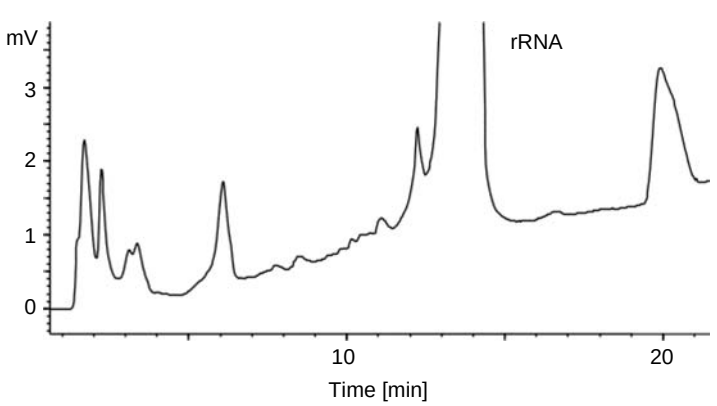

Fig. 15 RNA chromatography of a total cellular extract of RNA from tobacco plants. The large, broad peak eluting between 12 and 15 min is primarily rRNA together with mRNA.

mRNA fraction is concentrated and stored for subsequent experimentation. This procedure also forms the first step in the synthesis of cDNA from tissue samples and for the generation of template in a typical RTPCR experiment.

As can be seen in Fig. 16, the outcome of a typical mRNA purification can be readily followed by RNA chromatography. The process serves primarily to enrich for mRNA (or rather to deplete the rRNA) than to completely remove the excess rRNA. Indeed, at least two rounds of enrichment are usually required in order to remove the bulk of the non-polyadenylated RNA. In this particular experiment, RNA chromatography is used in an analytical mode for the evaluation of polyA mRNA purification. It is possible to apply the same principle in a preparative mode in order to produce a series of fractions that may facilitate the selective synthesis of cDNA populations for the production of size selected DNA libraries. In addition, one of the major drawbacks of polyA mRNA isolation is that some mRNAs are not polyadenylated and therefore will be excluded from any subsequent analysis. The use of RNA chromatography in a preparative mode therefore offers the potential for isolating at least a fraction of those mRNA species that do not co-elute with rRNA. This is clearly an important area for development, since the analysis of cell specific RNA populations in disease is becoming increasingly important in molecular medicine.

4.6 Analysis of Transfer RNA

Each cell contains a population (usually referred to as a *pool*) of tRNAs that meet the requirements of that particular cell's (or in the case of bacteria, that organism's) protein synthesis machinery. The range of sizes of tRNA molecules is particularly narrow, compared with mRNA, between around 60 and 100 nt, and a given cell typically contains around 100 species. While this molecular weight range is ideal for RNA chromatography, the close similarity of molecular sizes represents a problem for resolving individual components in a

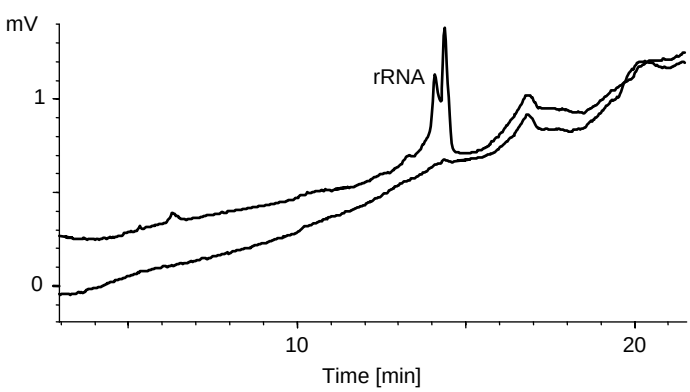

Fig. 16 Two cycles of polyA enrichment leads to the removal of the major contaminating rRNA species. The upper curve shows the RNA remaining (polyadenylated) after one round of enrichment on an oligo-dT resin and the lower curve after a second round.

typical cellular pool. This is readily seen in Fig. 17, where the total pool of tRNAs from *E. coli* has been subjected to RNA chromatography. Nevertheless, it is clearly possible to analyze fluctuations in tRNA pools by RNA chromatography when used in conjunction with a downstream procedure such as RTPCR.

It is also apparent that conformational differences between identical RNA molecules can be detected by RNA chromatography. In Fig. 18, the conformation of a single tRNA has been perturbed by changing the ionic composition and the temperature of the sample: the different elution profiles are a clear demonstration of the susceptibility of RNA chromatography to the conformation of the RNA molecule(s) under investigation. Therefore, there is a need to conduct a detailed systematic study of the temperature and solvent influences on the chromatography of complex RNA molecules: the tRNA population represents an ideal test bed for the development of a robust theoretical base for RNA chromatography.

4.7 Chromatography and Analysis of Synthetic Oligoribonucleotides

One of the major contributors to the rapid rate of progress in contemporary molecular biology has been the development of automated synthetic methods for the production of oligodeoxynucleotides. These relatively short (typically between 15–100 nt in length) oligomers are used in the PCR, numerous blotting procedures, transcription factor biochemistry, and, in some situations, have been shown to possess catalytic activity. The development of complementary protocols for RNA synthesis have been slower to emerge owing to problems arising through the presence of the additional 2′ hydroxyl group. Nevertheless, routine synthesis of 20- to 30-mer oligoribonucleotides is now possible. Such oligomers, as with DNA, have

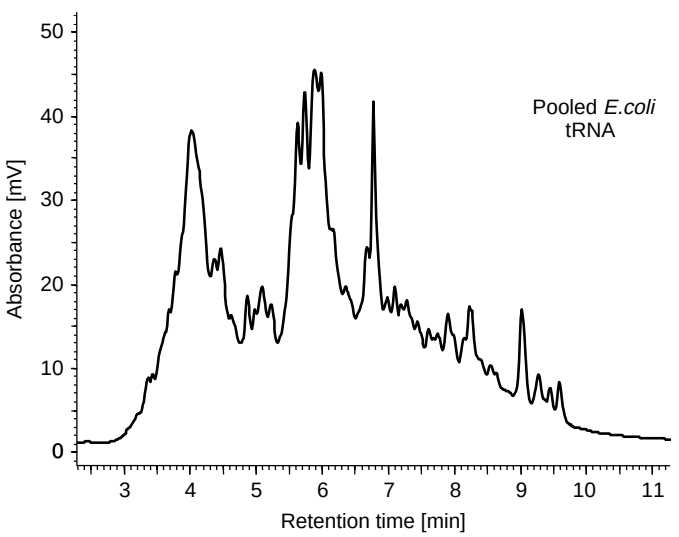

Fig. 17 RNA chromatography of the entire pool of tRNAs purified from *E. coli*. By coupling separation with RTPCR, it is possible to identify and quantify the individual tRNAs.

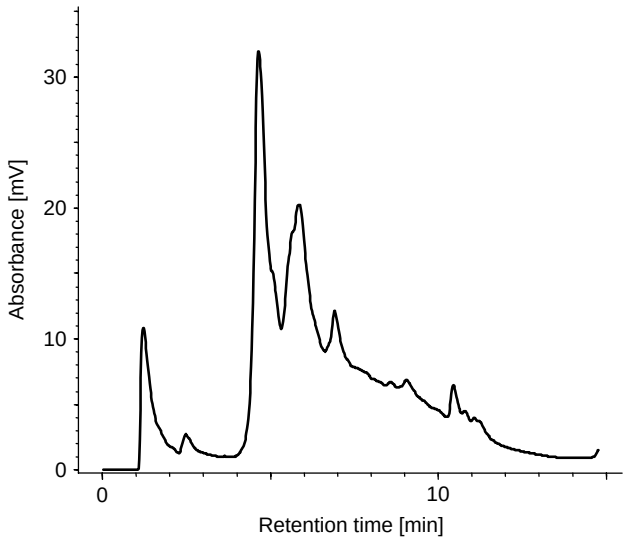

Fig. 18 Separation of the conformational species arising from preequilibration of an individual tRNA with Mg^{2+}. By systematically varying the buffer composition and temperature, RNA chromatography provides an insight into RNA conformation.

great value in experimental molecular biology, and are of particular importance in the study of catalysis mediated by RNA. Using the hairpin ribozyme as an example, RNA chromatography has been used as a general analytical tool in the study of the biochemical and conformational properties of synthetic RNA.

RNA is susceptible to acid–base hydrolysis, which provides a convenient means of optimizing conditions for the separation of oligoribonucleotides at single nucleotide resolution. The chromatogram shown in Fig. 19 is a 21 nucleotide synthetic oligoribonucleotides fluorescently labeled at its 5′ end, after exposure to 0.1 M sodium bicarbonate (pH 9 at 95 °C) for 30 min. The column was developed using acetonitrile (as described in the figure legend) and tetrabutylammonium bromide as the ion-pairing agent. The high resolution of the smaller cleavage products is apparent.

Base-catalyzed hydrolysis of RNA typically yields two products: a 2′phosphate moiety and a 2′-3′-cyclic phosphate intermediate, which undergoes further nucleophilic attack to yield the 3′-phosphate product. The 2′-3′-cyclic intermediate can be seen using 20% denaturing polyacrylamide sequencing gels (albeit with some with difficulty). However, with RNA chromatography, these two products are readily identified.

The information encoded by a typical genome is initially transcribed into mRNA and is subsequently translated into functional proteins, as discussed earlier. However, RNA also has the ability to catalyze biological reactions. This catalytic function is dependent on the three-dimensional shape of the RNA molecule. One technique that has been employed to obtain RNA structural information is "RNA footprinting." This method provides information on solvent accessibility within RNA molecules

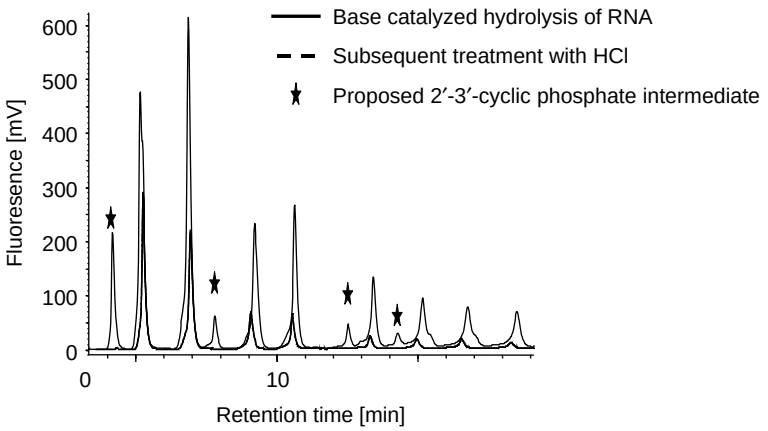

Fig. 19 The products of acid–base hydrolysis of a fluorescently end-labeled oligoribonucleotides. The 2′-3′-cyclic phosphate intermediate is indicated.

and can therefore be used to analyze secondary and tertiary interactions of small RNA structures. To differentiate between the internal and external regions of the folded RNA molecules, the solvent accessibility of the C5′- and also the generally quoted C4′-position of the ribose moiety can be monitored by the addition of an Fe (II)-EDTA complex together with hydrogen peroxide to the RNA in solution. The hydroxyl radicals generated primarily attack the C5′/C4′-position of the sugar, resulting in cleavage of the phosphodiester bond. The cleavage products can then be directly analyzed to identify those sites that show altered solvent accessibility.

Modifications to the standard footprinting reaction that allow the reaction products to be analyzed by RNA chromatography include the use of fluorescently labeled RNA. This allows the fluorescence-based detection of the cleavage products. The use of tetrabutylammonium bromide as the ion-pairing reagent is essential for the size-dependent separation of fluorescently labeled DNA. This regime particularly removes the influence of the hydrophobic fluorescent group and sequence-specific effects. Analysis of the footprinting products is rapid with run-times of approximately 30 min for each sample, with direct quantification of the cleavage products using proprietary software.

See also Denaturation of DNA; Footprinting Methods to Examine the Structure and Dynamics of Nucleic Acids; Footprinting Methods to Examine the Structure and Dynamics of Proteins; Genomic Imprinting, Molecular Genetics of.

Bibliography

Books and Reviews

Gjerde, D.T., Hanna, C.P., Hornby, D. (2002) *DNA Chromatography*, Wiley-VCH, Weinheim, Germany.

Heftmann, E. (Ed.) (1992) *Chromatography, Parts A and B*, 5th edition, Elsevier Science Publishers B.V., The Netherlands.

Huber, C.G., Oberacher, H. (2001) Analysis of nucleic acids by on-line liquid chromatography – mass spectrometry, *Mass Spectrom. Rev.* **20**, 310–343.

Poole, C.F., Poole, S.K. (1991) *Chromatography Today*, Elsevier Science Publishers B.V., The Netherlands.

Snyder, L.R., Kirkland, J.J. (1979) *Introduction to Modern Liquid Chromatography*, 2nd edition, Wiley, New York.

Thompson, J.A. (1986) A review of high performance liquid chromatography in nucleic acids research I. Historical perspectives, *Biochromatography* **1**, 16.

Thompson, J.A. (1987) A review of high performance liquid chromatography in nucleic acids research IV. Isolation, purification, and analysis of DNA restriction fragments, *Biochromatography* **2**, 4.

Xiao, W., Oefner, P.J. (2001) Denaturing high-performance liquid chromatography: A review, *Hum. Mutat.* **17**, 439–474.

Primary Literature

Azarani, A., Hecker, K.H. (2001) RNA analysis by ion-pair reversed phase high performance liquid chromatography, *Nucleic Acids Res.* **29**(2), E7.

Bahrami, A.R., Dickman, M.J., Matin, M.M., Ashby, J.R., Brown, P.E., Conroy, M.J., Fowler, G.J., Ro Sheikh, Q.I., Yeung, A.T., Hornby, D.P. (2002) Use of fluorescent DNA-intercalating dyes in the analysis of DNA via ion-pair reversed-phase denaturing high-performance liquid chromatography, *Anal. Biochem.* **309**, 248–252.

Belikov, S., Wieslander, L. (1995) Express protocol for generating G+A sequencing ladders, *Nucleic Acids Res.* **23**, 310.

Bird, A.P. (1986) CpG-rich islands and the function of DNA methylation, *Nature* **321**, 209–213.

Bonn, G., Huber, C., Oefner, P. (1996) *Nucleic Acid Separation on Alkylated Nonporous Polymer Beads*, U.S. Patent 5,585,236.

Breslauer, K.J., Frank, R., Blocker, H., Marky, L.A. (1986) Predicting DNA duplex stability from the base sequence, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3746–3750.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G.Q., Lander, E.S. (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes, *Nat. Genet.* **22**, 231–238.

Choy, Y.S., Dabora, S.L., Hall, F., Ramesh, V., Niida, Y., Franz, D., Kasprzyk-Obara, J., Reeve, M.P., Kwiatkowski, D.J. (1999) Superiority of denaturing high-performance liquid chromatography over single-stranded conformation and conformation-sensitive gel electrophoresis for mutation detection in TSC2, *Ann. Hum. Genet.* **63**, 383–391.

Devaney, J.M., Pettit, E.L., Kaler, S.G., Vallone, P.M., Butler, J.M., Marino, M.A. (2001) Genotyping of two mutations in the HFE gene using single-base extension and high-performance liquid chromatography, *Anal. Chem.* **73**, 620–624.

Dickman, M.J., Matin, M.M., Hornby, D.P. (2002) High-Throughput Analysis of Nucleic Acid Modification Reactions Using Ion-Pair Reverse-Phase High-Performance Liquid Chromatography, *Anal. Biochem.* **301**, 290–297.

Dickman, M.J., Conroy, M.J., Grasby, J.A., Hornby, D.P. (2002) RNA Footprinting Analysis using ion pair reverse phase Liquid Chromatography, *RNA* **8**, 247–251.

Doris, P.A., Chilton, B.S., Oefner, P.J., Hayward-Lester, A. (1998) Quantitative analysis of gene expression by ion-pair high-performance liquid chromatography, *J. Chromatogr.* **806**, 47–60.

Fedor, M.J. (2000) Structure and function of the hairpin ribozyme, *J. Mol. Biol.* **297**, 269–291.

Filipowicz, W. (2000) Imprinted expression of small nucleolar RNAs in brain: time for RNomics, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14035–14037.

Fixman, M., Freire, J.J. (1977) Theory of DNA melting curves, *Biopolymers* **16**, 2693–2704.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., Paul, C.L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1827–1831.

Galas, D.J., Schmitz, A. (1978) DNase footprinting: a simple method for the detection of protein-DNA binding specificity, *Nucleic Acids Res.* **5**, 3157–3170.

Gjerde, D.T., Taylor, P.D. (2000) *Chromatographic Method for Mutation Detection Using Mutation Site Specifically Acting Enzymes and Chemicals*, U.S. Patent 6,027,898.

Gjerde, D.T., Taylor, P.D. (2001) *System and Method for Automated Matched Ion Polynucleotide Chromatography*, PCT WO 01/46687.

Gjerde, D.T., Haefele, R.M., Togami, D.W. (1998) *System and Method for Performing Nucleic Acid Separations Using Liquid Chromatography*, U.S. Patent 5,772,889.

Gjerde, D.T., Haefele, R.M., Togami, D.W. (1999a) *Method for Performing Polynucleotide Separations Using Liquid Chromatography*, U.S. Patent 5,997,742.

Gjerde, D.T., Haefele, R.M., Togami, D.W. (1999b) *Process for Performing Polynucleotide Separations*, U.S. Patent 5,972,222.

Gjerde, D.T., Hanna, C.P., Hornby, D. (2002) *DNA Chromatography*, Wiley-VCH, Weinheim, Germany.

Gjerde, D.T., Taylor, P.D., Haefele, R.M. (2001) *Mutation Detection Method*, U.S. Patent 6,287,822.

Gjerde, D.T., Hornby, D.P., Hanna, C.P., Kuklin, A.I. (2002) *Method and System for RNA Analysis by Matched Ion Polynucleotide Chromatography*, U.S. Patent 6,475,388.

Gusev, I., Huang, X., Horvath, C. (1999) Capillary columns with in situ formed monolithic packing for micro high-performance liquid chromatography and capillary electrochromatography, *J. Chromatogr.* **855**, 273–290.

Hampel, K.J., Walter, N.G., Burke, J.M. (1998) The solvent-protected core of the hairpin ribozyme-substrate complex, *Biochemistry* **37**, 14672–14682.

Hargreaves, D., Rice, D.W., Sedelnikova, S.E., Artymiuk, P.J., Lloyd, R.G., Rafferty, J.B. (1998) Crystal structure of E. coli RuvA with bound DNA Holliday junction at 6 Åresolution, *Nat. Struct. Biol.* **6**, 441.

Hayward-Lester, A., Oefner, P.J., Sabatini, S., Doris, P.A. (1995) Accurate and absolute measurement of gene expression by single tube RT-PCR and HPLC, *Genome Res.* **5**, 494–499.

Hayward-Lester, A., Oefner, P.J., Sabatini, S., Kainer, D.B., Hinojos, C.A., Doris, P.A. (1998) Modeling and analysis of competitive RT-PCR, *Nucleic Acids Res.* **26**, 2511–2518.

Hecker, K.H., Turpie, B., Kuklin, A. (1999) Optimization of cloning efficacy by pre-cloning DNA fragment analysis, *BioTechniques* **26**, 216–218, 222.

Herman, J.G., Graff, J.R., Myohanen, S., Nelkin, B.D., Baylin, S.B. (1996) Methylation-specific PCR: A novel assay for methylation status of CpG islands, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9821–9826.

Hochleitner, E.O., Bakry, R., Huck, C.W., Flores, F., Stoggl, W.M., Stecher, G., Bonn, G.K. (2002) Analysis of isolectins on non-porous particles and monolithic polystyrene-divinylbenzene based sorbents and electrospray ionization mass spectrometry, *Int. J. Mass Spectrom.* **12239**, 1.

Hoogendoorn, B., Owen, M.J., Oefner, P.J., Williams, N., Austin, J., O'Donovan, M. (1999) Genotyping single nucleotide polymorphisms by primer extension and high performance liquid chromatography, *Hum. Genet.* **104**, 89–93.

Hoogendoorn, B., Norton, B., Kirov, G., Williams, N., Hamshere, M., Spurlock, G., Austin, J., Stephens, M., Buckland, P., Owen, M. (2000) Cheap, accurate and rapid allele frequency estimation of single nucleotide polymorphisms by primer extension and DHPLC in DNA pools, *Hum. Genet.* **107**, 488–493.

http://www.insertion.stanford.edu/ melt.html

Huber, C.G., Krajete, A. (1999) Analysis of nucleic acids by capillary ion-pair reversed-phase HPLC coupled to negative ion-electrospray ionization mass spectrometry, *Anal. Chem.* **71**, 3730–3739.

Huber, C.G., Oefner, P.J., Bonn, G.K. (1993) Rapid analysis of biopolymers on modified non-porous polystryrene-divinylbenzene particles, *Chromatographia* **37**, 653–658.

Huber, C.G., Oefner, P.J., Preuss, E., Bonn, G.K. (1993) High-resolution liquid chromatography of DNA fragments on highly cross-linked poly(styrene-divinylbenzene) particles, *Nucleic Acids Res.* **21**, 1061–1066.

Huber, C.G., Premstaller, A., Xiao, W., Oberacher, H., Bonn, G.K., Oefner, P.J. (2001) Mutation detection by capillary denaturing high-performance liquid chromatography using monolithic columns, *J. Biochem. Biophys. Methods* **47**, 5–19.

Jones, A.C., Austin, J., Hansen, N., Hoogendoorn, B., Oefner, P.J., Cheadle, J.P., O'Donovan, M.C. (1999) Optimal temperature selection for mutation detection by denaturing

HPLC and comparison to single-stranded conformation polymorphism and heteroduplex analysis, *Clin. Chem.* **45**, 1133–1140.

Jost, W., Unger, K., Schill, G. (1982) Reverse-phase ion-pair chromatography of polyvalent ions using oligonucleotides as model substances, *Anal. Biochem.* **119**, 214–223.

Kosaki, K., Yoshihashi, H., Ohashi, Y., Kosaki, R., Suzuki, T., Matsuo, N. (2001) Fluorescence-based DHPLC for allelic quantification by single-nucleotide primer extension, *J. Biochem. Biophys. Methods* **47**, 111–119.

Latham, J.A., Cech, T.R. (1989) Defining the inside and outside of a catalytic RNA molecule, *Science* **245**, 276–282.

Liang, P., Pardee, A.B. (1992) Differential display of eukaryotic messenger-RNA by means of the polymerase chain-reaction, *Science* **257**, 967–971.

Liu, W., James, C.D., Frederick, L., Alderete, B.E., Jenkins, R.B. (1997) PTEN/MMAC1 mutations and EGFR amplifications in glioblastomas, *Cancer Res.* **57**, 5254–5257.

Liu, W., Oefner, P.J., Qian, C., Odom, R.S., Francke, U. (1998) Denaturing HPLC-identified novel FBN1 mutations, polymorphisms and sequence variants in Marfan Syndrome and related connective tissue disorders, *Genet. Test.* **1**, 237–242.

Marino, M.A., Devaney, J.M., Smith, J.K., Girard, J.E. (1998) Sequencing using capillary electrophoresis of short tandem repeat alleles separated and purified by high performance liquid chromatography, *Electrophoresis* **19**, 108–118.

Matin, M.M., Hornby, D.P. (2000) A positive selection vector combining tetracycline resistance that eliminates the need for bacterial plating, *Anal. Biochem.* **278**, 46–51.

Matin, M.M., Baumer, A., Hornby, D.P. (2002) An Analytical Method for the Detection of Methylation Differences at Specific Chromosomal Loci Using Primer Extension and Ion Pair Reverse Phase HPLC, *Hum. Mutat.* **20**, 305–311.

Morasand, D., Poterszman, A. (1996) Getting into the major groove. Protein-RNA interactions, *Curr. Biol.* **6**, 530–532.

Narayanaswami, G., Taylor, P.D. (2001) Improved efficiency of mutation detection by high-performance liquid chromatography using modified primers and hybridization procedures, *Gen. Test.* **5**, 9–16.

O'Donovan, M.C., Oefner, P.J., Roberts, S.C., Austin, J., Hoogendoorn, B., Guy, C., Speight, G., Upadhyaya, M., Sommer, S.S., McGuffin, P. (1998) Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection, *Genomics* **52**, 44–49.

Oberacher, H., Huber, C.G. (2002) Capillary monoliths for the analysis of nucleic acids by high-performance liquid chromatography – electrospray ionization mass spectrometry, *Trends Anal. Chem.* **21**, 166–174.

Oberacher, H., Wellenzohn, B., Huber, C.G. (2002) Comparative sequencing of nucleic acids by liquid chromatography – tandem mass spectrometry, *Anal. Chem.* **74**, 211–218.

Oberacher, H., Parson, W., Muhlmann, R., Huber, C.G. (2001) Analysis of polymerase chain reaction products by on-line liquid chromatography – mass spectrometry for genotyping of polymorphic short tandem repeat loci, *Anal. Chem.* **73**, 5109–5115.

Oefner, P.J., Underhill, P.A. (1995) *Detection of Nucleic Acid Heteroduplex Molecules by Denaturing High-Performance Liquid Chromatography and Methods for Comparative Sequencing*, U. S. Patent 5,795,976.

Oefner, P.J., Huber, C.G., Umlauft, F., Berti, G.N., Stimpfl, E., Bonn, G.K. (1994) High-resolution liquid chromatography of fluorescent dye-labeled nucleic acids, *Anal. Biochem.* **15**, 39–46.

Ophoff, R.A., Terwindt, G.M., Vergouwe, M.N., van Eijk, R., Oefner, P.J., Hoffman, S.M.G., Lamerdin, J.E., Mohrenweiser, H.W., Bulman, D.E., Ferrari, M., Haan, J., Lindhout, D., van Ommen, G.J.B., Hofker, M.H., Ferrari, M.D., Frants, R.R. (1996) Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca^{2+} channel gene CACNL1A4, *Cell* **87**, 543–552.

Pingoud, A., Jeltsch, A. (2001) Structure and function of type II restriction endonucleases, *Nucleic Acids Res.* **29**, 3705–3727.

Poland, D. (1974) Recursion relation generation of probability profiles for specific-sequence macromolecules with long-range correlations, *Biopolymers* **13**, 1859–1871.

Premstaller, A., Oefner, P.J. (2002) Denaturing HPLC of Nucleic Acids, *LC-GC Europe* **15**, 410–422.

Premstaller, A., Oberacher, H., Huber, C.G. (2000) High-performance liquid chromatography – electrospray ionization mass spectrometry of single- and double-stranded nucleic acids using monolithic capillary columns, *Anal. Chem.* **72**, 4386–4393.

Premstaller, A., Xiao, W., Oberacher, H., O'Keefe, M., Stern, D., Willis, T., Huber, C.G., Oefner, P.J. (2001) Temperature-modulated array high-performance liquid chromatography, *Genome Res.* **11**, 1944–1951.

Reik, W., Walter, J. (2001) Genomic imprinting: parental influence on the genome, *Nat. Rev. Genet.* **2**, 21–32.

Sklar, J.L., Gjerde, D.T., Lamb, K.A., Hanna, C.P. (2001) *Method for Detection of DNA by Denaturing Matched Ion Polynucleotide Chromatography (DMIPC) and PCR*, U.S. Patent 6,218,153.

Skopek, T.R., Glaab, W.E., Monroe, J.J., Kort, K.L., Schaefer, W. (1999) Analysis of sequence alterations in a defined DNA region: Comparison of temperature-modulated heteroduplex analysis and denaturing gradient gel electrophoresis, *Mutat. Res.* **430**, 13–21.

Thompson, J.A. (1986) A review of high performance liquid chromatography in nucleic acids research III. Isolation purification, and analysis of supercoiled plasmid DNA, *Biochromatography* **1**, 68.

Thompson, J.A. (1987) A review of high performance liquid chromatography in nucleic acids research V. Nucleic acid affinity techniques in DNA and RNA research, *Biochromatography* **2**, 68.

Tullius, T.D. (1990) Hydroxyl radical footprinting – high resolution information about DNA protein contacts and application to lambda-repressor and cro protein, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5469–5473.

Underhill, P.A., Jin, L., Zemans, R., Oefner, P.J., Cavalli-Sforza, L.L. (1996) A pre-columbian Y chromosome-specific transition and its implications for human evolutionary history, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 196–200.

Underhill, P.A., Jin, L., Lin, A.A., Mehdi, S.Q., Jenkins, T., Vollrath, D., Davis, R.W., Cavalli-Sforza, L.L., Oefner, P.J. (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high performance liquid chromatography, *Genome Res.* **7**, 996–1005.

Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J.R., Mehdi, S.Q., Seielstad, M.T., Wells, R.S., Piazza, A., Davis, R.W., Feldman, M.W., Cavalli-Sforza, L.L., Oefner, P.J. (2000) Y chromosome sequence variation and the history of human populations, *Nat. Genet.* **26**, 358–361.

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W. (1995) Serial analysis of gene expression, *Science* **270**, 484–487.

Wagner, T., Stoppa-Lyonnet, D., Fleischmann, E., Muhr, D., Pages, S., Sandberg, T., Caux, V., Moeslinger, R., Langbauer, G., Borg, A., Oefner, P. (1999) Denaturing high-performance liquid chromatography detects reliably BRCA1 and BRCA2 mutations, *Genomics* **62**, 369–376.

Wyatt, J.R., Puglisi, J.D., Tinoco, I. Jr. (1990) RNA pseudoknots. Stability and loop size requirements, *J. Mol. Biol.* **214**, 455–470.

Xiao, W., Oefner, P.J. (2001) Denaturing High-Performance Liquid Chromatography: A Review, *Hum. Mut.* **17**, 439–474.

Nucleic Acid Hybrids, Formation and Structure of

James G. Wetmur
Department of Microbiology, New York, NY, USA

1 Introduction 334

2 Stability of DNA:DNA, DNA:RNA, and RNA:RNA Hybrids 335
2.1 Melting and Dissociation Temperatures Defined 335
2.2 Stability of Polynucleotide Duplexes (DNA:DNA, DNA:RNA, and RNA:RNA) 335
2.3 Stability of Oligonucleotide:Oligonucleotide and Polynucleotide:Oligonucleotide Duplexes 337

3 Formation of DNA:DNA, DNA:RNA, and RNA:RNA Hybrids 341
3.1 Hybridization with Polynucleotides 341
3.2 Hybridization with Oligonucleotides 345

4 Oligonucleotide Dissociation 346

Bibliography 347
Books and Reviews 347
Primary Literature 347

Keywords

Blot
A hybridization experiment in which the target nucleic acid is immobilized on a membrane (or filter). With electrophoretic separation of the target prior to immobilization, the blot is called a Southern blot if the target is DNA and a northern blot if the target is RNA.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

Dissociation Temperature
The temperature at which two strands of a hybrid are 50% dissociated after a specified time.

Hybrid
A partially or completely double-stranded (duplex) nucleic acid formed by interaction of two single-stranded nucleic acids, either DNA or RNA. Reassociation or renaturation may be used to describe hybridization of completely complementary DNA or RNA strands.

Melting Temperature
The temperature at which 50% of the base pairs of a polynucleotide:polynucleotide hybrid or 50% of the strands of oligonucleotide:polynucleotide or oligonucleotide:oligonucleotide hybrids have dissociated.

Polymerase Chain Reaction (PCR)
A method for geometric amplification of nucleic acids employing two sequence-specific oligonucleotides complementary to the DNA strands and a DNA polymerase. Each cycle includes a denaturation step, a hybridization (annealing) step, and a polymerization step.

Strand Separation Temperature
The temperature at which the two strands of a polynucleotide duplex are dissociated.

Hybridization is the formation of partially or completely double-stranded (duplex) nucleic acid (DNA:DNA, DNA:RNA, or RNA:RNA) by sequence-specific interaction of two complementary single-stranded nucleic acids. Reassociation or renaturation are often used to describe hybridization between completely complementary DNA or RNA strands. Solution hybridization is an integral part of the polymerase chain reaction. Hybridization is often carried out using a labeled probe in an attempt to detect the existence and quantity of complementary sequence in a complex nucleic acid mixture. The target is often immobilized, as in Southern and other blotting techniques. This chapter deals with the physical chemical aspects of hybridization, in particular, thermodynamics and kinetics. Melting temperatures, hybridization rates, and dissociation rates and temperatures are interrelated. Details are provided for calculation of these useful properties and examples of calculations are provided.

1 Introduction

DNA:DNA, DNA:RNA, and RNA:RNA hybrid duplexes are the products of hybridization reactions between complementary single-stranded molecules. Reassociation or renaturation are often used to describe hybrid formation between completely complementary DNA or RNA

strands. The reverse reaction is called strand separation or dissociation. The structural differences between deoxyribose and ribose in DNA and RNA strands are reflected in different forms of the antiparallel double helices and in quantitative differences in hybrid stability and the rate of hybrid formation. Nevertheless, they are qualitatively the same, and the basic concepts will be treated together.

2 Stability of DNA:DNA, DNA:RNA, and RNA:RNA Hybrids

2.1 Melting and Dissociation Temperatures Defined

The four *different* temperatures defined below are associated with nucleic acid melting:

1. t_m^∞, the melting temperature of a polynucleotide duplex, such as a long DNA molecule.
2. t_{ss}, the strand separation temperature of a polynucleotide duplex, also called the denaturation temperature in PCR.
3. t_m, the melting temperature where at least one strand of the duplex is an oligonucleotide. The definition of oligonucleotide is somewhat arbitrary. Oligonucleotides of length 80 are offered for sale. For the purposes of this chapter, oligonucleotides will be defined by their melting behavior, where the dissociation occurs without the formation of stable intermediates.
4. t_d, the dissociation temperature for an oligonucleotide bound to a polynucleotide on a solid support, such as a blot.

Lower case t implies °C, whereas upper case T implies K ($T = t + 273.16$).

2.2 Stability of Polynucleotide Duplexes (DNA:DNA, DNA:RNA, and RNA:RNA)

Figure 1 illustrates the definitions of t_m^∞ and t_{ss} for polynucleotide melting. t_m^∞ is the temperature where the fraction of

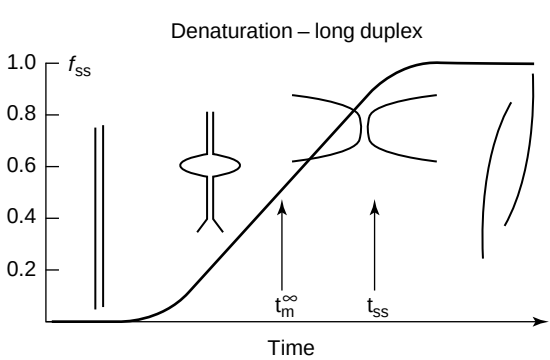

Fig. 1 Polynucleotide denaturation. f_{ss} = fraction single-stranded (nucleotides); t = temperature; t_m^∞ = melting temperature ($f_{ss} = 0.5$) for polynucleotides; t_{ss} = strand separation temperature. Steps in denaturation. (1) A duplex begins to denature reversibly at one end. End denaturation continues and internal loops are formed. At t_m^∞, the number of nucleotides in denatured ends and loops equals the number of nucleotides remaining base-paired. (2) Reversible denaturation continues until the strands are held together by the highest melting region. At t_{ss}, irreversible denaturation occurs. Superimposed on corresponding melting temperature profile (f_{ss} vs t) indicating t_m^∞ and t_{ss}.

single-stranded base pairs, f_{ss}, is 0.5, leading to molecules containing alternating duplex and denatured (loops or ends) regions. For natural nucleic acids, these loops may occur in the same places in many molecules, reflecting regions of lower melting temperature due to compositional heterogeneity. Because the reaction is intramolecular and equilibrium is rapidly achieved, t_m^∞ is independent of polynucleotide concentration and time, depending only on the composition of the polynucleotide duplex and the composition of the solvent. Strand separation occurs a few degrees above t_m^∞, at t_{ss}, and is not a rapidly reversible reaction. The denaturation temperature in PCR must be greater than t_{ss}.

For polynucleotides of length L containing only G:C and A:T or A:U base pairs, t_m^∞ (°C) may be determined from the empirical equation

$$t_m^\infty = A^* + 16.6\log_{10}\left[\frac{[\mathrm{SALT}]}{1.0 + 0.7[\mathrm{SALT}]}\right] + B^* \cdot (\%\ G + C) - \frac{500}{L}$$

where $[\mathrm{SALT}] = [Na^+] + [K^+] + [\mathrm{Tris}^+] + 4[Mg^{++}]^{0.5}$. A^* is 81.5, 67, or 78 °C and B^* is 0.41, 0.8, or 0.78 °C/(% G + C) for DNA:DNA, DNA:RNA and RNA:RNA hybrids, respectively. For example, t_m^∞ for a 50% G + C DNA:DNA hybrid of length 250 in 0.05 M KCl, 0.002 M $MgCl_2$ ([SALT] = 0.23) is given by $t_m^\infty = 81.5 - 11.7 + 0.41 \times 50 - 2 = 88.3$ °C. DNA:RNA hybrids adopt the A conformation of RNA:RNA hybrids rather than the B conformation of DNA:DNA hybrids but are always much less stable than RNA:RNA hybrids. However, DNA:RNA hybrids may be more or less stable than DNA:DNA hybrids, depending on % G + C. For example, in 1 M NaCl, t_m^∞ values for long DNA:RNA hybrids versus DNA:DNA duplexes are 103.2 °C versus 98.2 °C at 50% G + C and 87.2 °C versus 90.0 °C at 30% G + C.

Use of different solvents may affect A^* and/or B^*, and hence t_m^∞, with different effects for different polynucleotides. For example, A^* for DNA:DNA melting decreases 0.25 °C/% glycerol, 0.4 °C/% ethylene glycol, and 0.63 °C/% formamide. The effect of formamide on A^* is much less for DNA:RNA hybrids than for DNA duplexes, facilitating R-loop formation, where single-stranded RNA may displace a homologous DNA strand in a DNA duplex. In another example, base composition may be eliminated as a variable for DNA:DNA hybrids ($B^* = 0$) at 3.2 M Me_4NCl ($t_m^\infty = 94$ °C) or 2.4 M Et_4NCl ($t_m^\infty = 63$ °C) or 5.6 M betaine. Because betaine, unlike the quaternary ammonium compounds, is a zwitterion and does not neutralize the charge on DNA, the melting temperature may be varied from 75 °C down to 45 °C by varying KCl from 1 M to 0.01 M.

Introduction of modified bases also may affect A^* and/or B^*. Any substitution will affect either AT (or AU) or GC base pairs. If no charge change is introduced, the resulting change in t_m^∞ may be expressed as a change in A^*, ΔA^*, where

$$\Delta A^* = S(\%\ G + C)\, f_{AT}^M \quad \text{for A or T (U) modifications or}$$

$$\Delta A^* = S(\%\ G + C)\, f_{GC}^M \quad \text{for G or C modifications,}$$

where f_{AT}^M and f_{GC}^M are the fractional substitution of the modified base and S is characteristic of the nature of the substitution. For example, for 5-methyldeoxycytosine substitution for cytosine in duplex DNA, which may also occur in nature, $S = 0.12$.

Thus, for 100% substitution ($f_{GC}^{M} = 1$) in a 50% G + C DNA, t_m^{∞} is increased by 6 °C.

Formation of imperfectly matched hybrids also affects A*, t_m^{∞} being reduced approximately 1 °C for each percent mismatching. t_{ss} is proportionally reduced. Stringency in washing immobilized targets such as filters or blots (Southern, northern, slot) containing polynucleotide probes, a measure of selective dissociation of imperfectly matched hybrids, is increased as the temperature approaches t_m^{∞} in the wash solvent.

2.3 Stability of Oligonucleotide:Oligonucleotide and Polynucleotide:Oligonucleotide Duplexes

Figure. 2 illustrates the definition of t_m for the all-or-none melting of complementary oligonucleotides, with no denatured (frayed) ends or internal loops. t_m is the temperature at which the fraction of single strands, and hence single-stranded base pairs, f_{ss}, is 0.5. t_m is not t_d, the dissociation temperature important for washing of filters. t_d is discussed below in Sect. 4.

Because oligonucleotide melting is completely reversible, t_m, unlike t_m^{∞}, depends on oligonucleotide concentration as well as on nucleotide and solvent composition. Thus, for the same oligonucleotide in the same solvent, t_m determined spectrophotometrically at high concentration, t_m relevant for PCR priming at lower concentration, and t_m for filter hybridization at even lower concentration will be different. In a spectrophotometric determination of t_m, one oligonucleotide at molar strand concentration C_1 is often hybridized to a second complementary oligonucleotide at the same concentration. The molar concentration of the single-stranded fraction of the first oligonucleotide at t_m is $C_1/2$, which is equal to C in the calculations below. In the more common experiment, an oligonucleotide probe or primer is present at an excess concentration C_1 for saturation binding to a complementary region on a polynucleotide. The molar concentration, C, of the single-stranded oligonucleotide at t_m is C_1.

t_m for hybridization of excess oligonucleotide probe to a polynucleotide target is given by

$$t_m = \left[\frac{T^{\circ} \cdot \Delta H^{\circ}}{\Delta H^{\circ} - \Delta G^{\circ} + R \cdot T^{\circ} \ln [C]}\right] + 16.6 \log_{10} \frac{[SALT]}{1.0 + 0.7[SALT]} - 269.3$$

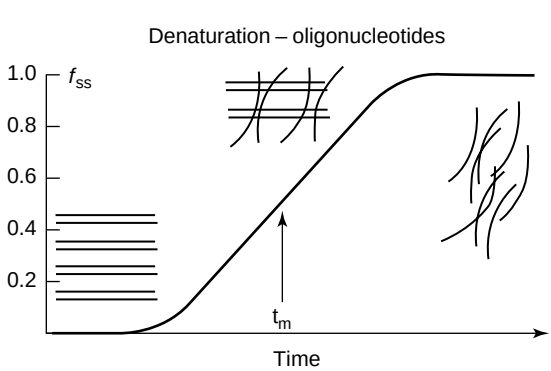

Fig. 2 Oligonucleotide denaturation. f_{ss} = fraction single-stranded (nucleotides and strands); t = temperature; t_m = melting temperature ($f_{ss} = 0.5$) for oligonucleotides. Steps in reversible denaturation with an all-or-none mechanism. At t_m, half of the duplexes have dissociated into single strands. Superimposed on corresponding melting temperature profile (f_{ss} vs t) indicating t_m.

where enthalpy $\Delta H^\circ =$

$$\Sigma_{nn}(N_{nn}\Delta H^\circ_{nn}) + \Delta H^\circ_e + \Delta H^\circ_i$$

and free energy $\Delta G^\circ =$

$$\Sigma_{nn}(N_{nn}\Delta G^\circ_{nn}) + \Delta G^\circ_e + \Delta G^\circ_i$$

and $T^\circ = 298.2^\circ K$ and R

$= 0.00199$ kcal/mol$^\circ$K

The first term in the t_m equation is T_m (K) in 1 M NaCl, the reference solvent. T is the reference temperature for the thermodynamic data tabulated below. The symbols i, nn, and e indicate contributions from initiation of base-pair formation, nearest-neighbors base pairs (stacking interactions), and other variables such as dangling ends, respectively. The use of nearest-neighbor thermodynamic functions is necessary because oligonucleotides with the same % G + C often have significantly different melting temperatures. Complete nearest-neighbor data sets are available for DNA:DNA, RNA:RNA, and DNA:RNA and chimeric RNA–DNA strands. The calculations below are illustrated for DNA:DNA, using as an example the parameters from Sugimoto et al. Several alternative parameter sets summarized in Owczarzy et al. give similar results. For the DNA:DNA nearest-neighbor values in the table below, $\Delta H^\circ_i = 0.6$ and $\Delta G^\circ_i = 3.3$ kcal mol^{-1} and

	AA	AT	TA	CA	GT	CT	GA	CG	GC	GG
or	TT			TG	AC	AG	TC			CC
$-\Delta H^\circ_{nn}$	8.0	5.6	6.6	8.2	9.4	6.6	8.8	11.8	10.5	10.9
$-\Delta G^\circ_{nn}$	1.47	1.07	1.11	1.94	1.80	1.71	1.79	3.15	2.63	2.43

For example, consider excess 1 μM (GGATAACG) binding to (CGTTATCC) at 1 M NaCl:

$$-\Delta H^\circ = 10.9 + 8.8 + 5.6 + 6.6 + 8.0 + 9.4 + 11.8 - 0.6 = 60.5 \text{ kcal/mol}$$

$$-\Delta G^\circ = 2.43 + 1.79 + 1.07 + 1.11 + 1.47 + 1.8 + 3.15 - 3.3 = 9.54 \text{ kcal/mol}$$

$$R \cdot T^\circ \ln[C] = -8.2 \text{ kcal/mol}$$

$$t_m = 32^\circ C \text{[in agreement with experiment]}$$

Figure 3 depicts various measurements of the shape of all-or-none oligonucleotide melting curves (shown at the right). The measurement of the width of the melting curve, ΔT, based on the slope at the midpoint, is illustrated in the top panel.

$$\Delta T = \frac{\varphi R T_m^2}{\Delta H^\circ}$$

The value of φ is 4 if the oligonucleotide is present in excess. Thus, for the 8-mer above, $T_m = 273.16 + 32 = 305$ K. $\Delta T = 12.5\,^\circ C$. Thus, even several degrees above t_m, there is a significant fraction of hybrid. In PCR, extension by DNA polymerase may be initiated above t_m because extended products no longer dissociate, allowing all templates to be initiated eventually. When the complementary sequence is present at the same concentration (e.g. in a spectrophotometer), the melting curve

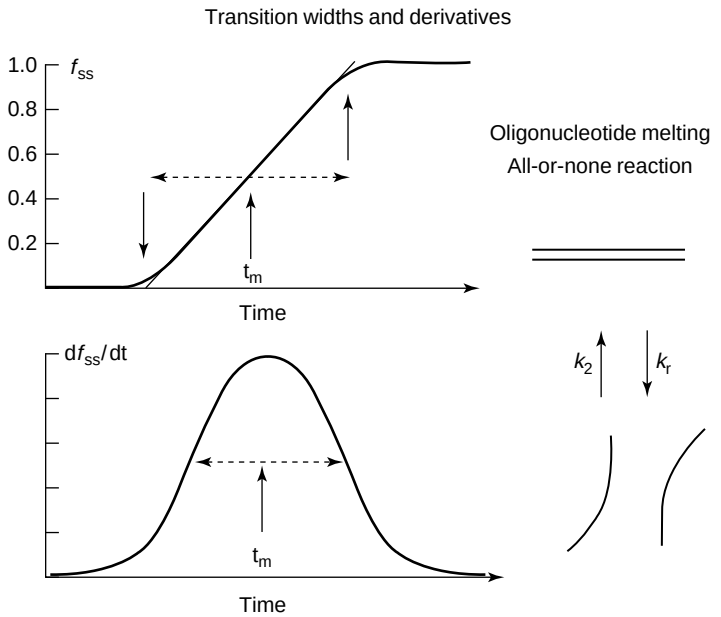

Fig. 3 Analysis of melting curves. f_{ss} = fraction single-stranded; t = temperature; t_m = melting temperature ($f_{ss} = 0.5$) for oligonucleotides; k_2 = second-order hybridization (reassociation, renaturation) rate constant; k_r = reverse first-order dissociation rate constant. Top: Melting temperature profile (f_{ss} 'vs' t) indicating t_m. The profile slope at the midpoint is used to create a straight line. The difference in t between the intercepts (noted by arrows) of this line at $f_{ss} = 0$ and $f_{ss} = 1$ is the transition width (dotted line). Right: Depiction of all-or-none oligonucleotide melting. Bottom: Derivative curve of df_{ss}/dt versus t also indicating t_m. The difference in t between the points where $f_{ss} = 0.5$ is the width at half-height of the derivative curve (dotted line).

is not completely symmetric about the melting temperature, and φ is increased to 6. The measurement of the width at half-height, $\Delta\ T_{1/2}$, for a derivative plot is illustrated in the bottom panel of Fig. 3.

$$\Delta T_{1/2} = \frac{-\varphi R T_m^2}{\Delta H^\circ}$$

The value of φ is 3.5 if the oligonucleotide is present in excess and 5.22 if the complementary sequence is present at the same concentration.

Because the nearest-neighbor thermodynamic functions have not been determined for most modified bases, the best available treatment is to calculate t_m in the absence of the modified bases (N/oligonucleotide) and to correct the result as follows:

$$\Delta t_m = -810000 \frac{NS}{\Delta H^\circ}$$

For example, substitution of 5-methyldeoxycytidine of deoxycytidine ($S = 0.12$) at three positions ($N = 3$) in the oligonucleotide example above (GGATAACG; $\Delta H^\circ = -60\,500$ cal mol^{-1}) gives $\Delta t_m = 5\,^\circ$C. An interesting modified base is 3-nitropyrrole, which interacts weakly but specifically with all four normal bases in

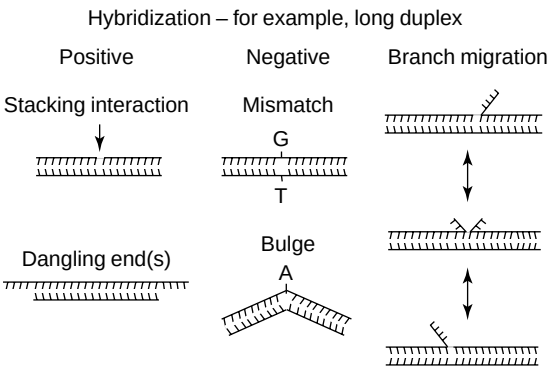

Fig. 4 Other contributions to the thermodynamics of oligonucleotide hybridization. Depictions of positive contributions. Top: Binding of an oligonucleotide adjacent to a duplex generates an additional stacking interaction (arrow). Bottom: Binding of an oligonucleotide to a longer strand generates 3′ and/or 5′ dangling ends. Depictions of negative contributions. Top: A GT base pair is used to illustrate a mismatch. Bottom: An inserted A in one of two complementary strands is used to indicate a bulge defect. Branch migration: The binding of two oligonucleotides with overlapping sequences to a complementary template generates a mixture of structures in which base-pairing by one and/or the other of the overlapping sequences interconvert by branch migration.

such a way that its inclusion in an oligonucleotide of length L leads to a melting temperature characteristic of an oligonucleotide of length $L-1$ but with a melting temperature almost independent of which base is pairing with the modified base. Oligonucleotides containing such a universal nucleotide have proven to be more useful than degenerate oligonucleotides for library screening and cloning by PCR when the genome comprising the library has not been sequenced.

Figure 4 illustrates other variables that may contribute to the thermodynamics of melting. There are two possible contributions that increase t_m, although the corrections are not large except for very short oligonucleotides. Very short oligonucleotides, 8 to 10 in length, have been used for array hybridization and as components of stacked primers for primer extension sequencing. Binding of an oligonucleotide adjacent to a duplex generates one additional nearest-neighbor (stacking) interaction, making an L-mer behave like an $L+1$-mer. The revised t_m may be calculated by adding the new nearest-neighbor contribution to the thermodynamic functions. Binding of an oligonucleotide to a longer strand generates 3′ and/or 5′ dangling ends. Their effect on t_m is strongly dependent on the orientation and sequence of the dangling end. Nevertheless, an estimated average contribution per dangling end may be taken to be $\Delta H_e^\circ = -5.0\,\text{kcal mol}^{-1}$ and $\Delta G_e^\circ = -1.0\,\text{kcal mol}^{-1}$, raising t_m for a 20-, 15- or 10-mer with one dangling end by an average 1, 1.5 or 3.0 °C, respectively. Filter hybridization typically involves two dangling ends. During the exponential amplification phase of PCR, one dangling end is formed upon primer-template binding, although PCR primers are typically too long for the dangling end to have a large effect.

Figure 4 depicts potential branch migration resulting from overlap (length O) of an oligonucleotide (length L) with a duplex. The total number of nearest-neighbor interactions is $L-O$, including the additional stacking interaction at the junction. The branch destabilizes the structure to

the extent of adding 0.8 kcal mol^{-1} to the free energy, lowering t_m for a 20-, 15- or 10-mer by an average 2.0, 2.6 or 3.6 °C, respectively. If the oligonucleotide overlap sequence is the same as the sequence in the adjacent duplex, branch migration can occur between all *O* possible structures. The branch migration stabilizes the structure to the extent of subtracting $RT_m \ln(O)$ from the free energy, canceling the destabilization introduced by the branch when the overlap, *O*, is about 4.

Figure 4 also illustrates variables that may contribute negatively to the thermodynamics of melting, including bulge defects and mismatches. Discrimination against these defects is a measure of the specificity of hybridization. A bulge defect may be estimated by calculating the thermodynamic functions without the bulge nucleotide and adding 2800 cal mol^{-1} to the free energy, lowering t_m for a 20-, 15-, or 10-mer by an average 7.0, 9.0, or 12 °C respectively. Mismatches may be approximated as follows. First calculate the thermodynamic parameters as if the mismatch had no effect. For example, if a GG mismatch occurs between G and G, one strand will contain GGG, with two GG stacking interactions and the other strand will contain CGC, with a CG and a GC stacking interaction. The entropy and free-energy contributions may be calculated by adding these four nearest-neighbor contributions and dividing the total by 2. Second, correct the thermodynamic functions depending on whether one of the mismatched bases is (1) a dG, (2) a dC, or (3) neither a dG or a dC. The enthalpies are corrected by adding 6700, 8700, or 9500 cal mol^{-1}, respectively. The free energies are corrected by adding 2900, 3700, and 4000 cal mol^{-1}, respectively. These corrections suggest that destabilization due to a mispair is of the order of the loss of two stacking interactions but with half the change in the temperature coefficient, although mismatches involving G residues are relatively more stable. A special case is a mismatch occurring at the end of an oligonucleotide, where only one stacking interaction can be lost. For internal mismatches, the average reduction in t_m for a 20-, 15-, or 10-mer would be 5, 8, or 12 °C respectively, if the mismatch contains a G, and 7, 10, or 16 °C respectively, if the mismatch contains a C, and 7, 10, or 17 °C respectively, if the mismatch contains neither a G or a C. These calculations are consistent with a decrease in t_m^∞ for polynucleotides of 1 °C/% mismatching.

3 Formation of DNA:DNA, DNA:RNA, and RNA:RNA Hybrids

3.1 Hybridization with Polynucleotides

Figure 5 illustrates a temperature profile for carrying out a denaturation and hybridization (renaturation, reassociation) reaction with polynucleotides. Denaturation requires a temperature above t_{ss}, the strand separation temperature. The hybridization reaction is initiated (time $t = 0$) by lowering the temperature to t_r, a hybridization (renaturation) temperature below t_m^∞, the polynucleotide melting temperature. Denaturation and renaturation of DNA also may be achieved by increasing the pH to approximately 13 followed by neutralization at t_r. The steps in the hybridization reaction are (1) a reversible nucleation step, in which the first 1 (or 2) correct base pairs are formed and (2) an irreversible zippering reaction, in which the subsequent base pairs are formed. Although duplexes containing fewer than 10 to 20 bp are usually unstable at t_r, the

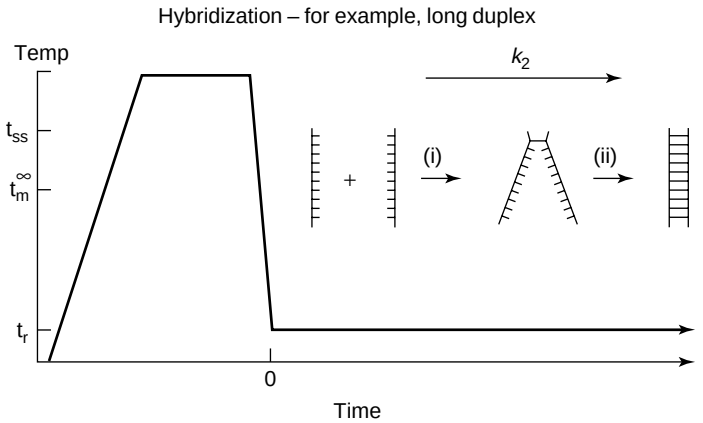

Fig. 5 Temperature profile for denaturation and hybridization as a function of time. t = temperature; t = time; t_{ss} = strand separation temperature; t_m^∞ = polynucleotide melting temperature; t_r = hybridization (renaturation, reassociation) temperature; k_2 = hybridization (renaturation, reassociation) rate constant. Denaturation occurs above t_{ss}. Hybridization is initiated ($t = 0$) and continued at t_r. Hybridization reaction steps: (1) nucleation; (2) zippering.

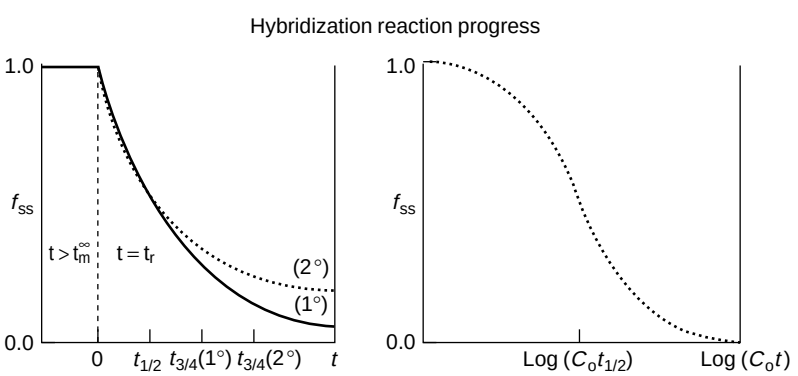

Fig. 6 Hybridization reaction measurements. f_{ss} = fraction of potential duplex remaining single-stranded; t_m^∞ = polynucleotide melting temperature; t_r = hybridization (renaturation, reassociation) temperature; C_o = total nucleotide concentration; t = time; $t_{1/2}$ = reaction half-time; $t_{3/4}$ = time to complete 3/4 of the reaction. Left panel: Reaction begins ($t = 0$) when t is decreased from $t > t_m^\infty$ to $t = t_r$. A second-order reaction (2°; dotted line) takes place if the complementary strands are present at equal concentration. A pseudo-first-order reaction (1°; solid line) takes place if one strand is present in excess over its complementary strand. Right panel: $C_o t$ plot of a second-order reaction.

rate-determining step is the formation of the second (or third) base pair. The overall hybridization reaction is characterized by the second-order rate constant, k_2.

Time course plots for the same hybridization reaction are illustrated in Fig. 6, including a $C_o t$ plot for a second-order reaction. Logarithmic $C_o t$

plots permit simultaneous display of reactions differing by orders of magnitude in half-time. Many reactions involve a labeled probe hybridizing to a complementary sequence in a complex target. If the concentrations of complementary probe and target sequences are unequal, the reactant in excess is termed the driver. A second-order reaction takes place if the complementary strands are present at equal concentration (e.g. DNA renaturation) or if the driver is double stranded and capable of reassociation. A pseudo-first-order reaction takes place if the driver is single stranded (e.g. mRNA) or cannot reassociate (e.g. immobilized on a solid support for a Southern blot). For all hybridization reactions, f_{ss} is the fraction of potential duplex remaining single stranded.

The concentration term used for polynucleotide hybridization is molar nucleotide concentration, C_o, and not strand concentration, C. For example, at 1 μg mL^{-1} double-stranded DNA, $C_o = 3$ μM. If one reactant is in excess, C_o is the driver concentration. When immobilized DNA is the driver, which is rarely the case, C_o is difficult to determine, being a measure of available target concentration. Furthermore, because a dilute probe would be used with the immobilized driver target, stirring may be necessary to maintain homogeneity.

Reciprocal rate plots are illustrated in Fig. 7 for the same pseudo-first-order and second-order reactions as depicted in the left panel of Fig. 6. These plots are described by the equations below.

Second Order: Half-time (seconds):

$$\frac{1}{f_{ss}} = k_2 \frac{C_o}{2} t + 1 \qquad t_{1/2} = \frac{2}{k_2 C_o}$$

Pseudo-First Order: Half-time:

$$f_{ss} = e^{-k_2 C_o t} \qquad t_{1/2} = \frac{\ln 2}{k_2 C_o}$$

For the second-order reaction, a plot of $1/f_{ss}$ versus t gives a slope equal to $k_2 C_o/2$. For the first-order reaction, a plot of ln $(1/f_{ss})$ versus t gives a slope equal to $k_2 C_o$.

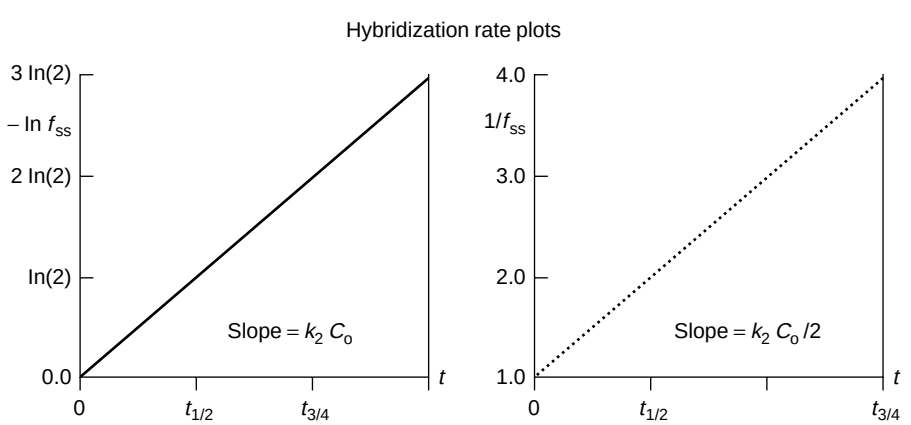

Fig. 7 Hybridization rate plots. f_{ss} = fraction of potential duplex remaining single-stranded; t = time; $t_{1/2}$ = reaction half-time; $t_{3/4}$ = time to complete 3/4 of the reaction. Left panel: Plot of $-\ln f_{ss}$ versus t for the pseudo-first-order reaction depicted in figure 6, left panel (solid line). Right panel: Reciprocal plot of $1/f_{ss}$ versus t for the second-order reaction depicted in figure 6, left panel (dotted line).

The second-order rate constant, k_2 in $M^{-1}\ sec^{-1}$, is given by

$$k_2 = \frac{k'_N\sqrt{L_s}}{N}$$

where L_s is the length of the shortest strand participating in duplex formation, N is the complexity or the total number of base pairs present in nonrepeating sequences and k'_N is the nucleation rate constant. The interpretations of N and L are illustrated in Fig. 8. The dependence of k_2 on the square root of L rather than on L suggests that nucleation sites become less available as the length of the single strands increases, consistent with an excluded volume effect.

The effects of temperature and salt on k'_N are illustrated in Fig. 9. The rate of hybridization reactions, and hence k'_N, is zero at t_m^∞. k_2 increases to a broad maximum at about 25 °C below t_m^∞. At lower temperatures, hybridization rates fall off slowly for homopolymers but precipitously for natural nucleic acids because of intramolecular base- pairing and consequent decreased availability of

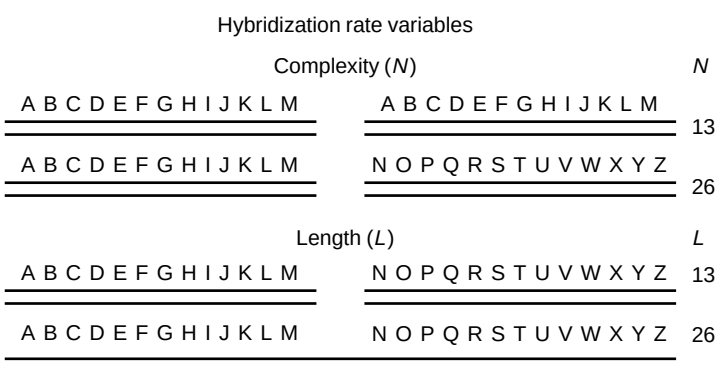

Fig. 8 Hybridization rate variables of the reactants. N = complexity, the number of nucleotide pairs of unique DNA; L = length, the average number of nucleotides in DNA single strands. Top panel depicts differences in N at constant L. Bottom panel depicts differences in L at constant N.

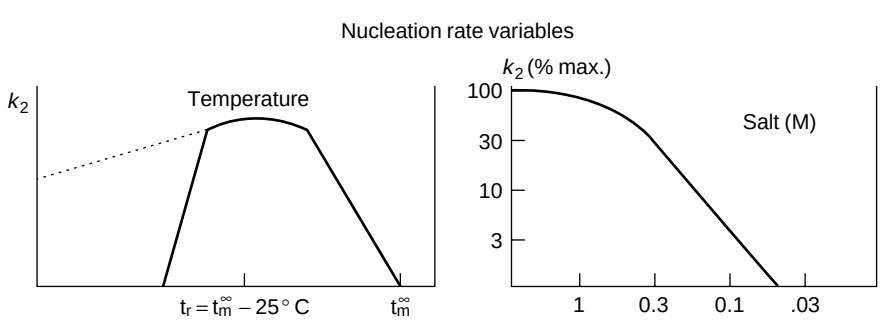

Fig. 9 Hybridization rate variables of the solvent. k_2 = hybridization (renaturation, reassociation) rate constant; t_m^∞ = polynucleotide melting temperature; t_r = hybridization (renaturation, reassociation) temperature. Homopolymers: dotted line; natural nucleic acids: solid line. Left panel: k_2 versus t_r. k_2 is 0 at t_m^∞. k_2 is maximum around $t_m^\infty - 25\,°C$. Right panel: k_2 versus salt (effective monovalent cation concentration).

nucleation sites. At optimal temperatures, k'_N for DNA:RNA hybrid formation is 20 to 50% slower than DNA:DNA hybrid formation, depending on the degree of RNA secondary structure. RNA:RNA hybridization rates are similarly reduced.

Although mismatched base pairs have a significant effect on melting temperature, mismatching of up to 10% of the bases has little effect on hybridization rates.

k'_N also depends on solvent viscosity and ionic strength. In solvents with glycerol, ethylene glycol, or formamide, k'_N is reduced because of increased solvent (microscopic) viscosity compared to 1.0 M NaCl at 70 °C. On the other hand, polymers, which increase the macroscopic viscosity, do not reduce hybridization rates. In fact, hybridization rates may be accelerated by use of dextran sulfate, which has the effect of concentrating the reactants. Another method for acceleration of hybridization is to concentrate the reactants at the interface of a phenol–water emulsion.

The dependence of k'_N on salt concentration is given by

$$k'_N = \{4.35 \log_{10}[\mathrm{SALT}] + 3.5\} \times 10^5 \quad \text{for } 0.2 \le [\mathrm{SALT}] \le 4.0$$

$$k'_N = \{4.35 \log_{10}(1 + 34[\mathrm{SALT}]^3)\} \times 10^5 \quad \text{for } [\mathrm{SALT}] < 0.2.$$

Thus, in typical hybridization buffers approximating 1 M NaCl, $k'_N = 3.5 \times 10^5\ \mathrm{M}^{-1}\ \mathrm{sec}^{-1}$.

Consider, as an example, the hybridization of a nick-translated 4-kb probe ($N = 4000$) to a blot. The probe would be the driver. Assume that the nick translation broke the single strands to 625 nucleotides ($L^{1/2} = 25$). Then, in 1 M NaCl at 70 °C, $k_2 = 2.2 \times 10^3\ \mathrm{M}^{-1}\ \mathrm{sec}^{-1}$. If the denatured probe is used at 10 ng mL^{-1} ($C_o = 30$ nM), $t_{1/2} = 2/k_2 C_o = 2/\{2.2 \times 10^3 \times 3 \times 10^{-8}\} = 3 \times 10^4$ s $= 8.5$ h. The half-times for hybridization may vary over many orders of magnitude depending linearly on the concentration and inversely on the complexity of the driver. For example, at 100 μg mL^{-1}, $t_{1/2}$ for the same nick-translated probe would be 3 s, whereas $t_{1/2}$ would be about 40 s, 1 h, and 30 days for λ ($N = 48\,000$), *E. coli* ($N = 4\,000\,000$), and human DNA ($N = 3\,000\,000\,000$) respectively.

3.2 Hybridization with Oligonucleotides

Hybridization with oligonucleotides usually takes place far below t_m^∞ (at or below t_m), so the decrease in k_2 near t_m^∞ is not a factor, and the temperature dependence of k_2 for oligonucleotides is small unless the target has interfering secondary structure. Oligonucleotide hybridization has the same salt dependence as polynucleotide hybridization. In fact, oligonucleotide hybridization may be described by the same equations. For oligonucleotide–polynucleotide reactions, the oligonucleotide is usually the driver leading to a pseudo-first-order reaction characterized by a rate constant with $N = L$ of the oligonucleotide. Oligonucleotide probe reactions are very fast. For example,

$$k_2 = \frac{k'_N}{\sqrt{L}} = 7 \cdot 10^4\ \mathrm{M}^{-1}\ \mathrm{sec}^{-1} \quad \text{for a 25-mer in 1 M NaCl at } t_m^\infty - 25\ ^\circ\mathrm{C}.$$

k_2 is reduced only 1.2-fold for every 10 °C decrease in hybridization temperature. At 1 nM ($C_o = 25$ nM), $t_{1/2} = \ln 2/(7 \times 10^4 \times 25 \times 10^{-9}) = 396$ s or 6.6 min. Note that $t_{1/2}$ increases proportional to the degeneracy of mixed oligonucleotide probes or primers.

In other examples, $t_{1/2}$ for a 1 μM primer in PCR buffer ($k'_N = 7 \times 10^4\ M^{-1}\ sec^{-1}$) is 3 s and, in 1 M NaCl at t_m for 1 μM GGATAACG, the oligonucleotide used in previous thermodynamic calculations, $t_{1/2} = 1.3$ s.

4 Oligonucleotide Dissociation

Figure 10 illustrates oligonucleotide reactions on solid supports. Capture reactions are simply hybridization reactions. However, when hybridized oligonucleotides on solid supports are washed to eliminate unhybridized oligonucleotides, both t_m and the rate characterized by k_2 decrease as C decreases, and the hybrid stability depends entirely on the dissociation rate constant, k_r. In Sect. 2.3, we determined the enthalpy of melting, $-\Delta H^\circ$, and t_m, which is a measure of the equilibrium constant for oligonucleotide hybridization, or the ratio of the hybridization and dissociation rate constants. In Sects 3.1 and 3.2, we determined k_2, which has a small activation energy. With this information, we may determine k_r. The activation energy for k_r, E_a, is approximately 4.0 kcal mol^{-1} added to $-\Delta H^\circ$. t_d is the dissociation temperature, which may defined as the temperature at which 50% of correctly matched hybrid is released in a specified length of time, t_{wash}. Because t_d depends on time and not on concentration, unlike $t_{1/2}$ for hybridization and t_m, t_d is the same whether using unique or degenerate probes. It is t_d may be estimated for extended washing of 14- to 20-mers as 2 °C for each dA:dT pair plus 4 °C for each dG:dC base pair. t_d may be calculated using the following:

$$t_d = \frac{T_m}{\{1 + R \cdot T_m \cdot [\ln (t_{wash}/t_{1/2})/E_a]\}} - 273.2$$

For example, for 1 μM GGATAACG with no dangling ends in 1 M NaCl, $E_a = -\Delta H + 4000 = 64\,500$ cal mol^{-1}, $T_m = 305$ K (Sec. 2.3), and $t_{1/2} = 1.3$ s (Sect. 3.2). $t_d = 10$ °C for 50% probe released with 1 h (3600 s washing. As another example, $t_d = 48$ °C for 6 h of washing for the arbitrary 16-mer GATCGTACCAAGGTTC with 2 dangling ends, in agreement with the 2 °C for each dA:dT pair plus 4 °C for each dG:dC base-pair rule. These calculations depend on all the elements of the physical chemistry of hybridization.

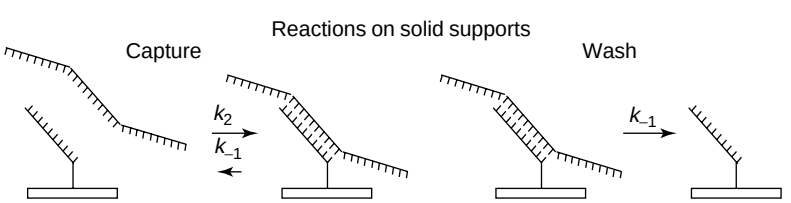

Fig. 10 Reactions on solid supports: hybridization (capture) and dissociation (washing). k_2 = second-order hybridization (reassociation, renaturation) rate constant; k_r = reverse first-order dissociation rate constant. Left: Capture of a polynucleotide by an oligonucleotide tethered to a solid support. Below the melting temperature, the hybridization rate is faster than the dissociation rate. Right: Dissociation of captured product. No hybridization occurs if unreacted polynucleotide is washed away. The reaction is governed by k_r.

See also Denaturation of DNA.

Bibliography

Books and Reviews

Bloomfield, V.A., Crothers, D.M., Tinoco, I. (2000) *Nucleic Acids: Structures, Properties, and Functions*, University Science Books, Sausalito, CA.

Bowtell, D., Sambrook, J. (2002) *DNA Microarrays: A Molecular Cloning Manual*, Cold Spring Harbor Laboratory Press, New York.

Ian, A.Darby. (2000) *In Situ Hybridization Protocols (Methods in Molecular Biology, 123)*, 2nd edition, Humana Press, Totowa, NJ.

Hames, B.D., Higgins, S.J. (1985) *Nucleic Acid Hybridisation, a Practical Approach*, IRL Press, Oxford, UK.

Joseph, S., David, W.R., Joe, S. (2001) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York.

Stephenson, F.H. (2003) *Calculations for Molecular Biology and Biotechnology: A Guide to Mathematics in the Laboratory*, Academic Press, Boston, MA.

Wetmur, J.G. (1991) DNA probes: applications of the principles of nucleic acid hybridization, *Crit. Rev. Biochem. Mol. Biol.* **26**, 227–259.

Primary Literature

Britten, R.J., Davidson, E.H. (1976) Studies on nucleic acid reassociation kinetics: empirical equations describing DNA reassociation, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 415–419.

Britten, R.J., Graham, D.E., Neufeld, B.R. (1974) Analysis of repeating DNA sequences by reassociation, *Methods Enzymol.* **29**, 363–418.

Camien, M.N., Warner, R.C. (1986) Denaturation of covalently closed circular DNA. Kinetics, comparison of several DNAs, mechanism and ionic effects, *J. Biol. Chem.* **261**, 6026–6033.

Chang, C.T., Hain, T.C., Hutton, J.R., Wetmur, J.G. (1974) Effects of microscopic and macroscopic viscosity on the rate of renaturation of DNA, *Biopolymers* **13**, 1847–1858.

Chen, J., Germer, S., Higuchi, R., Berkowitz, G., Godbold, J., Wetmur, J.G. (2002) Kinetic polymerase chain reaction on pooled DNA: a high-throughput, high-efficiency alternative in genetic epidemiological studies, *Cancer Epidemiol. Biomarkers Prev.* **11**, 131–136.

Dai, H., Meyer, M., Stepaniants, S., Ziman, M., Stoughton, R. (2002) Use of hybridization kinetics for differentiating specific from non-specific binding to oligonucleotide microarrays, *Nucleic Acids Res.* **30**, e86.

Demidov, V.V., Protozanova, E., Izvolsky, K.I., Price, C., Nielsen, P.E., Frank-Kamenetskii, M.D. (2002) Kinetics and mechanism of the DNA double helix invasion by pseudocomplementary peptide nucleic acids, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5953–5958.

Doktycz, M.J., Goldstein, R.F., Paner, T.M., Gallo, F.J., Benight, A.S. (1992) Studies of DNA dumbbells. I. Melting curves of 17 DNA dumbbells with different duplex stem sequences linked by T4 endloops: evaluation of the nearest-neighbor stacking interactions in DNA, *Biopolymers* **32**, 849–864.

Doktycz, M.J., Morris, M.D., Dormady, S.J., Beattie, K.L., Jacobson, K.B. (1995) Optical melting of 128 octamer DNA duplexes. Effects of base pair location and nearest neighbors on thermal stability, *J. Biol. Chem.* **270**, 8439–8445.

Flavell, R.A., Birfelder, E.J., Sanders, J.P., Borst, P. (1974) DNA-DNA hybridization on nitrocellulose filters. 1. General considerations and non-ideal kinetics, *Eur. J. Biochem.* **47**, 535–543.

Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T., Turner, D.H. (1986) Improved free-energy parameters for predictions of RNA duplex stability, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9373–9377.

Galau, G.A., Britten, R.J., Davidson, E.H. (1977) Studies on nucleic acid reassociation kinetics: rate of hybridization of excess RNA with DNA, compared to the rate of DNA renaturation, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 1020–1023.

Gonzalez, S., Wetmur, J.G. (2000) Holliday junction branch migration and resolution. RuvA, RuvB, and RuvC from the hyperthermophile thermotoga maritima, *Methods Mol. Biol.* **152**, 107–118.

Green, C., Tibbetts, C. (1981) Reassociation rate limited displacement of DNA strands by branch migration, *Nucleic Acids Res.* **9**, 1905–1918.

Hutton, J.R. (1977) Renaturation kinetics and thermal stability of DNA in aqueous solutions

of formamide and urea, *Nucleic Acids Res.* **4**, 3537–3555.

Hutton, J.R., Wetmur, J.G. (1972) Renaturation of DNA in the presence of ethidium bromide, *Biopolymers* **11**, 2337–2348.

Hutton, J.R., Wetmur, J.G. (1973a) Renaturation of bacteriophage phiX174 DNA-RNA hybrid: RNA length effect and nucleation rate constant, *J. Mol. Biol.* **77**, 495–500.

Hutton, J.R., Wetmur, J.G. (1973b) Effect of chemical modification on the rate of renaturation of deoxyribonucleic acid. Deaminated and glyoxalated deoxyribonucleic acid, *Biochemistry* **12**, 558–563.

Ikuta, S., Takagi, K., Wallace, R.B., Itakura, K. (1987) Dissociation kinetics of 19 base paired oligonucleotide-DNA duplexes containing different single mismatched base pairs, *Nucleic Acids Res.* **15**, 797–811.

Ivanov, I.B., Khrapko, K.R., Chernov, B.K., Khorlin, A.A., Lysov, Y., Florentiev, V.L., Mirzabekov, A.D. (1991) Hybridization properties of gel-immobilized oligonucleotides, *Nucleic Acids Symp. Ser* **24**, 189–190.

Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P., Gingeras, T.R. (2002) Large-scale transcriptional activity in chromosomes 21 and 22, *Science* **296**, 916–919.

Kennedy, G.C., Matsuzaki, H., Dong, S., Liu, W.M., Huang, J., Liu, G., Su, X., Cao, M., Chen, W., Zhang, J., Liu, W., Yang, G., Di, X., Ryder, T., He, Z., Surti, U., Phillips, M.S., Boyce-Jacino, M.T., Fodor, S.P., Jones, K.W. (2003) Large-scale genotyping of complex DNA, *Nat. Biotechnol.* **21**, 1233–1237.

Kierzek, R., Burkard, M.E., Turner, D.H. (1999) Thermodynamics of single mismatches in RNA duplexes, *Biochemistry* **38**, 14214–14223.

Kinberg-Calhoun, J., Wetmur, J.G. (1981) Circular, but not circularly permuted, deoxyribonucleic acid reacts slower than linear deoxyribonucleic acid with complementary linear deoxyribonucleic acid, *Biochemistry* **20**, 2645–2650.

Kohne, D.E., Levison, S.A., Byers, M.J. (1977) Room temperature method for increasing the rate of DNA reassociation by many thousandfold: the phenol emulsion reassociation technique, *Biochemistry* **16**, 5329–5341.

Kushon, S.A., Jordan, J.P., Seifert, J.L., Nielsen, H., Nielsen, P.E., Armitage, B.A. (2001) Effect of secondary structure on the thermodynamics and kinetics of PNA hybridization to DNA hairpins, *J. Am. Chem. Soc.* **123**, 10805–10813.

Lee, C.H., Wetmur, J.G. (1972a) Independence of length and temperature effects on the rate of helix formation between complementary ribopolymers, *Biopolymers* **11**, 549–561.

Lee, C.H., Wetmur, J.G. (1972b) On the kinetics of helix formation between complementary ribohomopolymers and deoxyribohomopolymers, *Biopolymers* **11**, 1485–1497.

Lima, W.F., Monia, B.P., Ecker, D.J., Freier, S.M. (1992) Implication of RNA structure on antisense oligonucleotide hybridization kinetics, *Biochemistry* **31**, 12055–12061.

Maskos, U., Southern, E.M. (1993) A study of oligonucleotide reassociation using large arrays of oligonucleotides synthesised on a glass support, *Nucleic Acids Res.* **21**, 4663–4669.

McCarthy, B.J., Church, R.B. (1970) The specificity of molecular hybridization reactions, *Annu. Rev. Biochem.* **39**, 131–150.

McTigue, P.M., Peterson, R.J., Kahn, J.D. (2004) Sequence-dependent thermodynamic parameters for locked nucleic acid (LNA)-DNA duplex formation, *Biochemistry* **43**, 5388–5405.

Miller, S.J., Wetmur, J.G. (1974) Determination of the rate of renaturation of modified DNA by fluorescence depolarization, *Biopolymers* **13**, 2545–2551.

Mir, K.U., Southern, E.M. (1999) Determining the influence of structure on hybridization using oligonucleotide arrays, *Nat. Biotechnol.* **17**, 788–792.

Morrison, L.E., Stols, L.M. (1993) Sensitive fluorescence-based thermodynamic and kinetic measurements of DNA hybridization in solution, *Biochemistry* **32**, 3095–3104.

Nakano, S., Kanzaki, T., Sugimoto, N. (2004) Influences of ribonucleotide on a duplex conformation and its thermal stability: study with the chimeric RNA-DNA strands, *J. Am. Chem. Soc.* **126**, 1088–1095.

Nguyen, H.K., Southern, E.M. (2000) Minimising the secondary structure of DNA targets by incorporation of a modified deoxynucleoside: implications for nucleic acid analysis by hybridisation, *Nucleic Acids Res.* **28**, 3904–3909.

Nichols, R., Andrews, P.C., Zhang, P., Bergstrom, D.E. (1994) A universal nucleoside for use at ambiguous sites in DNA primers, *Nature* **369**, 492–493.

Owczarzy, R., Vallone, P.M., Gallo, F.J., Paner, T.M., Lane, M.J., Benight, A.S. (1997) Predicting sequence-dependent melting stability of short duplex DNA oligomers, *Biopolymers* **44**, 217–239.

Panyutin, I.G., Hsieh, P. (1993) Formation of a single base mismatch impedes spontaneous DNA branch migration, *J. Mol. Biol.* **230**, 413–424.

Panyutin, I.G., Hsieh, P. (1994) The kinetics of spontaneous DNA branch migration, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2021–2025.

Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P., Fodor, S.P. (1994) Light-generated oligonucleotide arrays for rapid DNA sequence analysis, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5022–5026.

Pontius, B.W., Berg, P. (1991) Rapid renaturation of complementary DNA strands mediated by cationic detergents: a role for high-probability binding domains in enhancing the kinetics of molecular assembly processes, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8237–8241.

Quartin, R.S., Wetmur, J.G. (1989) Effect of ionic strength on the hybridization of oligodeoxynucleotides with reduced charge due to methylphosphonate linkages to unmodified oligodeoxynucleotides containing the complementary sequence, *Biochemistry* **28**, 1040–1047.

Quartin, R.S., Plewinska, M., Wetmur, J.G. (1989) Branch migration mediated DNA labeling and cloning, *Biochemistry* **28**, 8676–8682.

Rees, W.A., Yager, T.D., Korte, J., von Hippel, P.H. (1993) Betaine can eliminate the base pair composition dependence of DNA melting, *Biochemistry* **32**, 137–144.

Riccelli, P.V., Merante, F., Leung, K.T., Bortolin, S., Zastawny, R.L., Janeczko, R., Benight, A.S. (2001) Hybridization of single-stranded DNA targets to immobilized complementary DNA probes: comparison of hairpin versus linear capture probes, *Nucleic Acids Res.* **29**, 996–1004.

SantaLucia, J., Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1460–1465.

SantaLucia, J., Jr., Allawi, H.T., Seneviratne, P.A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **35**, 3555–3562.

Shchepinov, M.S., Case-Green, S.C., Southern, E.M. (1997) Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays, *Nucleic Acids Res.* **25**, 1155–1161.

Smith, M.J., Britten, R.J., Davidson, E.H. (1975) Studies on nucleic acid reassociation kinetics: reactivity of single-stranded tails in DNA-DNA renaturation, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4805–4809.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis, *J. Mol. Biol.* **98**, 503–517.

Southern, E.M. (1979) A preparative gel electrophoresis apparatus for large scale separations, *Anal. Biochem.* **100**, 304–318.

Southern, E.M., Maskos, U., Elder, J.K. (1992) Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models, *Genomics* **13**, 1008–1017.

Strider, W., Camien, M.N., Warner, R.C. (1981) Renaturation of denatured, covalently closed circular DNA, *J. Biol. Chem.* **256**, 7820–7829.

Sugimoto, N., Nakano, S., Yoneyama, M., Honda, K. (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes, *Nucleic Acids Res.* **24**, 4501–4505.

Sugimoto, N., Satoh, N., Yasuda, K., Nakano, S. (2001) Stabilization factors affecting duplex formation of peptide nucleic acid with DNA, *Biochemistry* **40**, 8444–8451.

Sugimoto, N., Katoh, M., Nakano, S., Ohmichi, T., Sasaki, M. (1994) RNA/DNA hybrid duplexes with identical nearest-neighbor base-pairs have identical stability, *FEBS Lett.* **354**, 74–78.

Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M., Sasaki, M. (1995) Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes, *Biochemistry* **34**, 11211–11216.

Szabo, P., Elder, R., Steffensen, D.M., Uhlenbeck, O.C. (1977) Quantitative in situ hybridization of ribosomal RNA species to polytene chromosomes of Drosophila melanogaster, *J. Mol. Biol.* **115**, 539–563.

Tsourkas, A., Behlke, M.A., Rose, S.D., Bao, G. (2003) Hybridization kinetics and thermodynamics of molecular beacons, *Nucleic Acids Res.* **31**, 1319–1330.

Tyagi, S., Kramer, F.R. (1996) Molecular beacons: probes that fluoresce upon hybridization, *Nat. Biotechnol.* **14**, 303–308.

Weinstock, P.H., Wetmur, J.G. (1990) Branch capture reactions: effect of recipient structure, *Nucleic Acids Res.* **18**, 4207–4213.

Wetmur, J.G. (1971) Excluded volume effects on the rate of renaturation of DNA, *Biopolymers* **10**, 601–613.

Wetmur, J.G. (1976) Hybridization and renaturation kinetics of nucleic acids, *Annu. Rev. Biophys. Bioeng.* **5**, 337–361.

Wetmur, J.G., Davidson, N. (1968) Kinetics of renaturation of DNA, *J. Mol. Biol.* **31**, 349–370.

Wetmur, J.G., Ruyechan, W.T., Douthart, R.J. (1981) Denaturation and renaturation of Penicillium chrysogenum mycophage double-stranded ribonucleic acid in tetraalkylammonium salt solutions, *Biochemistry* **20**, 2999–3002.

Wieder, R., Wetmur, J.G. (1982) Factors affecting the kinetics of DNA reassociation in phenol-water emulsion at high DNA concentrations, *Biopolymers* **21**, 665–677.

Williams, J.C., Case-Green, S.C., Mir, K.U., Southern, E.M. (1994) Studies of oligonucleotide interactions by hybridisation to arrays: the influence of dangling ends on duplex yield, *Nucleic Acids Res.* **22**, 1365–1367.

Wong, D.M., Weinstock, P.H., Wetmur, J.G. (1991) Branch capture reactions: displacers derived from asymmetric PCR, *Nucleic Acids Res.* **19**, 2251–2259.

Nucleic Acid Packaging of DNA Viruses

Philip Serwer
The University of Texas Health Science Center San Antonio, Texas

1 Biology of Bacteriophages: DNA Packaging 353
1.1 Microbe Hunting 355
1.2 Bacteriophage Genomics, Proteomics, and Informatics 356

2 Physical Chemistry of DNA Packaging 357
2.1 Energy of Compaction 357
2.2 DNA Packaging Motors 359

3 *In vitro* DNA Packaging: Complex System Biochemistry 360

Bibliography 361
Books and Reviews 361
Primary Literature 361

Keywords

Bacterial Lysis
Bursting of bacteria and the release of their contents.

Bacteriophage Plaque
Single bacteriophage particle-derived clearing in a continuous, gel-embedded, turbid slab of bacteria.

Bacterial Transduction
The use of bacteriophages to introduce new properties to bacteria.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

Bacteriophage Therapy
The use of bacteriophages for controlling/eliminating unwanted bacteria.

Capsid of a Bacteriophage
Protein shell with a cavity in which a bacteriophage genome is packaged during formation of a mature bacteriophage particle.

Concatemer
Linear polymer of a bacteriophage's double-stranded DNA genome.

Connector (Portal Ring)
Ring of (usually) 12 subunits through which a DNA molecule enters a capsid during DNA packaging.

DNA Packaging Motor
Terminase-containing, ATP-fueled complex that drives a DNA molecule into a bacteriophage capsid.

Lytic Bacteriophage
A bacteriophage that causes lysis of its host.

Lysogenic Bacteriophage
A bacteriophage that coinhabits its host.

Nanometry
Measurement of nanometer-sized movements.

Prophage
A bacteriophage genome that replicates with a host cell.

Procapsid
The bacteriophage capsid immediately after it is assembled and before it begins to package a DNA molecule.

Sodium Dodecylsulfate Polyacrylamide Gel Electrophoresis (SDSPAGE)
Gel electrophoresis of proteins whose surface charge density has been homogenized by boiling in sodium dodecylsulfate.

Tail of a Bacteriophage
External projection of a bacteriophage capsid that attaches the bacteriophage to host cells.

Terminase
A multimeric protein complex that forms at least an ATPase of a DNA packaging motor.

Transcription
The process of transferring information from a DNA genome to RNA.

Translation
The process of using information encoded in RNA sequence to synthesize a specific protein.

Bacteriophages are viruses whose hosts are bacteria. Bacteriophages are the most numerous and genomically diverse form of life in the environment. Many bacteriophage genes have either (1) no detectable homology to any other bacteriophage/cellular gene or (2) homology to extremely diverged bacteriophage genes of presumably similar function. When infecting a host bacterial cell, a double-stranded DNA bacteriophage produces proteins that package progeny DNA genomes in a protein container (capsid) that protects its genome when the bacteriophage is outside of a cell. The energy requirement for DNA packaging is approximately 0.5 kcal mol^{-1} of packaged nucleotide pair. The capsid and accessory proteins join to form an ATP-fueled motor that drives the DNA molecule into a cavity in the capsid. The motor packages the DNA molecule via end-first motion through a portal ring also called *the connector*. An accessory protein/ATPase is at least the spark plug of the motor. Bacteriophage DNA packaging motors are investigated as a prototype for understanding other biological motors. Testing for feedback control of the motor is an important current objective. Investigations of both bacteriophage DNA packaging and bacteriophage capsid assembly have produced *in vitro* systems that are complex. One strategy for analysis of such complex systems is observation at the level of single events (single-particle analysis). Single-particle analysis has already revealed surprising cooperativity among DNA packaging events and has also revealed a comparatively strong motor capable of generating about 40 to 70 pN of force while intermittently hesitating. Bacteriophage DNA packaging systems are useful for further developing procedures for single-particle analysis of complex biochemical systems.

1 Biology of Bacteriophages: DNA Packaging

Bacteria (and Archaea) are single cells without a nucleus that can live independently. Bacteria are found almost everywhere, even in environments of extreme high and low temperature. Bacteria have parasites that are smaller than cells. These parasites typically consist of a protein shell (capsid) that has an internal cavity in which a nucleic acid molecule is packaged. Although both RNA-containing and single-stranded DNA-containing bacteriophages have been isolated, the nucleic acid is usually (96% of current isolates) a double-stranded DNA molecule. These subcellular

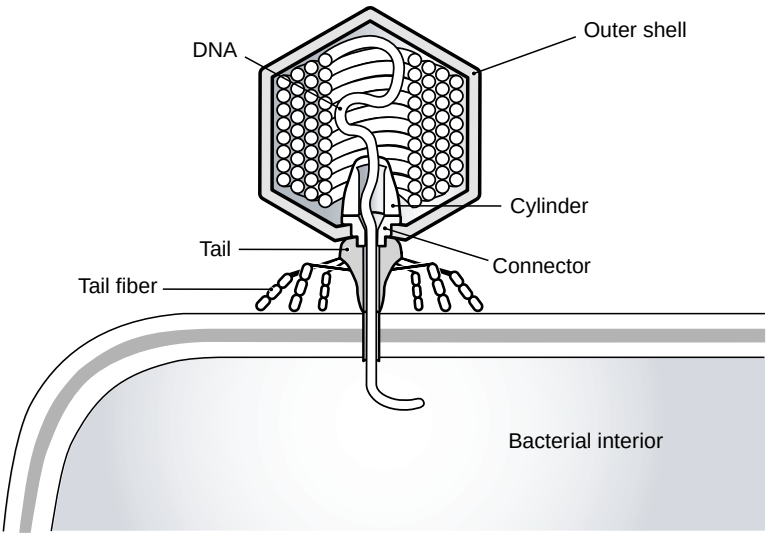

Fig. 1 Infection of a host cell (only partially shown) by a single particle of bacteriophage T7.

parasites are called *bacteriophages*. The capsid of a double-stranded DNA bacteriophage has an external projection (tail) that has fibers for attachment of the bacteriophage particle to a host cell. After attachment, a double-stranded DNA bacteriophage injects its DNA genome into the host cell. This process is illustrated in Fig. 1.

After the DNA genome is injected, it sometimes initiates production of progeny bacteriophages, typically over 100 per cell. The progeny bacteriophages are released when the cell bursts (lyses). This process is called *lytic infection*. Alternatively, the injected genome can become part of the cell's genome (prophage), sometimes by physically incorporating itself into a bacterial chromosome. In this latter case, the state of the bacteriophage chromosome is called *lysogenic* because of the capacity to reinitiate a lytic cycle at a later time. This latter life cycle is called *temperate*. Thus, bacteriophages can either destroy bacterial cells (lytic bacteriophage) or coinhabit bacterial cells (lysogenic bacteriophage). Lysogenic bacteriophages sometimes acquire exogenous genes and transfer them to the host genome. Investigators can make deliberate use of both lytic and gene-transferring functions.

A lytic infection produces progeny double-stranded DNA bacteriophages via the following route: (1) A nucleic acid-free capsid (procapsid) is assembled. (2) Then, the procapsid binds a DNA molecule. (3) Next, the procapsid draws the DNA molecule into its internal cavity through a 12-fold symmetric ring called *a connector*. This process is called *DNA packaging*. The driving force for DNA packaging is derived from a noncapsid accessory protein that attaches to the connector. This protein is the larger of two accessory proteins that are not initially part of the capsid, but attach to it for DNA packaging. The larger accessory protein (also called *the large terminase subunit*) is an ATPase that cleaves ATP to produce energy that is used to package the bacteriophage genome.

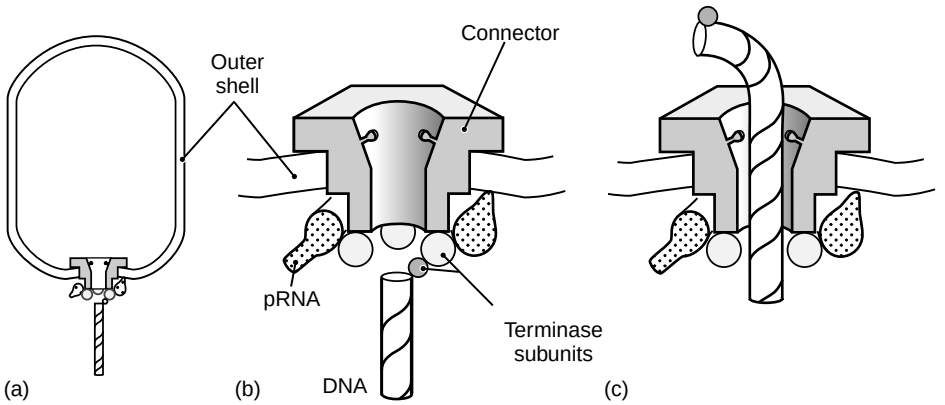

Fig. 2 The bacteriophage ϕ29 DNA packaging motor. (a) The entire ϕ29 procapsid that has initiated DNA packaging by binding a DNA molecule. (b) A magnified view of the DNA-binding region of the same procapsid with connector, terminase subunits, DNA molecule, and also a packaging RNA molecule (pRNA) that is part of the motor. (c) The same as the view of (b), but after the DNA molecule has started to enter the internal cavity of the capsid. The pRNA molecule has thus far been unique to bacteriophage ϕ29.

Typically, six large terminase molecules are present during DNA packaging. The process of DNA packaging is shown schematically in Fig. 2 for bacteriophage ϕ29. The capsid with accessory proteins is called the *DNA packaging motor*. DNA packaging motors are studied not only to understand bacteriophages but also to understand biological motors in general.

1.1 Microbe Hunting

Bacteriophages originate in the environment. So far, all double-stranded DNA bacteriophages have packaged their genome via a procapsid with an ATPase/large terminase subunit that formed at least the "spark plug" of the DNA packaging motor. The large terminase proteins from widely different bacteriophage genomes have either no or barely detectable sequence similarity. However, biochemical similarities among the characterized large terminases are explained with the assumption of descent from a single ancestral gene, followed by divergence. Thus, comparison at the level of secondary and tertiary structure is potentially useful for analysis of both the biochemical mechanisms and the evolution of DNA packaging genes. To fill the evolutionary tree of bacteriophage genes, a more comprehensive collection of bacteriophages is sought. This collection will also be useful for the practical purposes of (1) bacteriophage therapy, the use of bacteriophages for controlling/eliminating unwanted bacteria, and (2) bacterial transduction, the use of bacteriophages to introduce new properties to bacteria.

No doubt, researchers have much to do before they obtain a representative sample of environmental bacteriophages. Environmental water typically has 10s of billions of bacteriophages particles per liter and soil typically has 100s of billions of bacteriophages particles per pound. The number of strains is typically smaller, but is still considered very large, especially since very

few (none in some cases) of these bacteriophages are currently cultured. To compound this lack of knowledge, less than 1% of all strains of visible bacteria have been cultured. Thus, the art of microbe hunting is becoming important again, as it was 100–150 years ago. In the previous era of microbe hunts, success in culture was usually associated with success in identifying a component of bacterial nutrition. Perhaps, success in today's microbe culturing-based hunts will be associated with success in identifying other details of microbial life, for example, the need for either a solid surface or another niche in which a microbe is competitive. Some of the current hunts are performed by polymerase chain reaction amplification of environmental DNA, that is, without culturing the bacteriophage. Similarly, some hunts are performed by shotgun cloning/sequencing of environmental bacteriophage DNA, again without culturing.

1.2 Bacteriophage Genomics, Proteomics, and Informatics

In the 1960s–1980s, researchers investigated both bacteriophage genome organization and bacteriophage gene functions either by (1) mutating and mapping the various bacteriophage genes, or (2) by observing the biochemical changes that occurred during infection. The biochemical changes include production of messenger RNA and proteins, for example. Sometimes these two strategies were combined. For example, the various genes were deliberately given mutations that produced premature polypeptide chain termination when a mutant bacteriophage was grown in a nonpermissive host. The protein encoded by a mutated gene was deduced by observing which protein was shortened. The length of any given protein was determined by electrophoretic fractionation through a polyacrylamide gel, after homogenizing the electrical surface charge density of all proteins by boiling them in sodium dodecylsulfate(SDS) (SDS polyacrylamide gel electrophoresis, or SDSPAGE).

These early investigations of genetics and biochemistry identified the signals for both starting and stopping the production of both messenger RNA (transcription) and messenger RNA-encoded proteins (translation). Most sequenced bacteriophages are now initially characterized by computer-based recognition of these features. The current state of the art is that excellent predictions are made for large proteins and for major transcriptional start sites. Predictions for small proteins are often unreliable.

The expanded informatics also recognizes homology among the coding sequences of genes of both bacteriophages and prophages. Detection of conserved protein sequences is a means for determining whether a protein of significant function is, in fact, made. A steady stream of new bacteriophage gene sequences is helping the finding of additional genes. Nonetheless, about half of all prospective bacteriophage genes have no detected relative.

Furthermore, studies of homology have found extreme divergence among bacteriophages in the sequence of a gene for any given function. The gene for the large terminase subunit is an example. Evidence for lateral DNA transfer among bacteriophages has been found. The lateral transfer occurs for DNA segments that vary in span from a partial gene to a module that comprises several genes. The result is mosaicism in the pedigrees of bacteriophage

genomes. The implied common gene pool is accessed by multiple bacteriophages that do not have a known common host.

The variability and plasticity of any given bacteriophage gene is a potential subject for informatics-based correlation of biochemical characteristics with structural characteristics. For example, the informatics has identified a P-loop ATPase site near the N-terminus of the large terminase subunit of all studied double-stranded DNA bacteriophages. Genetic interference with the integrity of the P-loop causes either loss or slowing of DNA packaging activity for at least two bacteriophages (λ and T3). Thus, the assumption is made that this ATPase is the site of the "spark plug"of the DNA packaging motor. Furthermore, we expect differences among the large terminases because (1) the terminase is involved in the initiation of DNA packaging and (2) differences exist in the DNA substrate for packaging. Some substrates are DNA circles; some are linear polymers of the mature genome, and, therefore, longer than the mature genome (called *a concatemer*); others (including bacteriophage ϕ29 DNA) are mature, linear genomes with the smaller terminase protein covalently joined to the 5′ ends. Cleavage is part of initiation in the case of both circular and concatemeric genomes. Comparative informatics will hopefully show how the large terminase adapts to a change in DNA substrate for packaging.

Today, the DNA sequence of a newly isolated bacteriophage is determined directly by use of a combination of shotgun cloning/sequencing and direct sequencing of the bacteriophage DNA. The proteins produced by the bacteriophage can now be analyzed by a combination of SDSPAGE and mass spectroscopy. This technology is expected to reduce uncertainty about which prospective genes are actually expressed. Additionally, the occurrence of posttranslational changes, such as protein fusion and cleavage, will be directly analyzed when this type of protein analysis is combined with the DNA sequence.

2 Physical Chemistry of DNA Packaging

2.1 Energy of Compaction

While some investigators focus on bacteriophage biology, others focus on the physical chemistry. In fact, a primary reason for use of bacteriophages as a model system is the opportunity for interdisciplinary investigation. In the case of DNA packaging, one can calculate the energy needed to compact the DNA to the density found in the capsids of bacteriophages. This density is uniform $\pm 15\%$ among the various bacteriophages. The density of DNA packaging is measured primarily by use of low angle x-ray scattering, but also by use of buoyant density centrifugation in media that minimally dehydrate a packaged DNA molecule. These two measurements agree that the DNA double helix occupies about one half the volume of the cavity in which the DNA is packaged.

Given this number, theoreticians calculate the energy per mole of nucleotide pair needed for (1) overcoming electrostatic repulsion among negatively charged DNA segments, and (2) bending the double helix against elastic forces that originate in both electrostatic repulsion and base stacking. The packaging-associated entropic change makes a negligible contribution to the free energy of DNA compaction. Overcoming the electrostatic repulsion is the

primary energetic requirement for packaging. Experimentally, expulsion of the packaged DNA (without disassembly of the capsid) releases heat that is another indication of the free energy of packaging because the entropy of packaging is negligible. Experiment and theory agree on approximately 0.5 kcal mol^{-1} of packaged nucleotide pair. Both theory and experiment also agree that circular coiling of packaged DNA is the most probable conformation. However, neither theory nor experiment conclusively discriminates between unidirectional coiling (as with thread on a spool) and bidirectional coiling (as rope is sometimes stored on the deck of a ship). Bidirectional coiling would produce kinks. A drawing of the packaged DNA of bacteriophage T7 within its capsid is in Fig. 1. This drawing is based on cryoelectron microscopy of T7.

A third measure of the energy of DNA compaction was achieved by integration of the force times the distance interval of a DNA molecule during its packaging. But, how does one measure the force, especially when the force is expected to increase as more DNA becomes packaged? An answer has recently been provided via the development of procedures for (1) attaching to a large latex sphere both a capsid that is packaging a DNA molecule and the capsid-distal end of the DNA molecule (see Fig. 3a), (2) holding the capsid-distal sphere in an optical trap of known force constant, (3) holding the capsid-bound sphere in a pipette and (4) monitoring with better than 10-nm precision the position of the capsid-distal sphere during DNA packaging (Fig. 3a). This type of measurement is sometimes called *nanometry*.

As the DNA molecule is packaged, the capsid-distal latex sphere moves in the optical trap (Fig. 3b). As the capsid-distal latex sphere moves, the optically derived force on the sphere increases until the packaging is completely stalled.

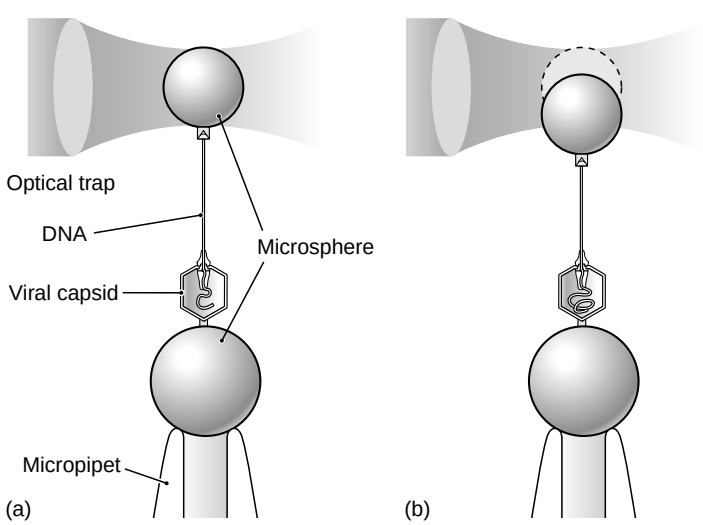

Fig. 3 Nanometry of a DNA molecule while it is being packaged in a bacteriophage capsid. (a) The DNA packaging motor after the DNA molecule begins to enter the capsid during nanometry. (b) The same DNA packaging motor after further DNA entry in the conditions of nanometry.

The relationship of force to packaging rate is measured. The stall force for bacteriophage ϕ29 DNA packaging motors varies between 40 and 70 pN. This stall force is larger than the stall forces for RNA polymerase and various eukaryotic motors that have been (partially) reconstructed *in vitro*. To obtain the force present throughout packaging, the packaging is allowed to proceed by feedback regulating the position of the pipet in Fig. 3 so that the optical trap exerts constant force. As DNA enters the capsid, the packaging rate is used to measure the force being used to raise the potential energy of the packaged portion of the DNA molecule. That is to say, this measurement supplies the force needed to calculate the energy of DNA compaction. This energy was 0.56 kcal mol^{-1} of nucleotide pair, in agreement with both the theory and the experiment described above.

The above agreement on the energy of compaction was found among three totally different strategies. This type of agreement is difficult to achieve unless all three strategies are basically producing accurate results. This conclusion is especially important for the nanometry because nanometry is a comparatively new discipline with not all potential sources of distortion well understood. Effects of the high density of surfaces are potentially important. But, the above data indicate that this does not change the measurements. In any case, the inside of a cell is very surface intense. So, perturbation by surfaces, if found, is not necessarily to be regarded to be an unproductive distortion (artifact).

2.2 DNA Packaging Motors

By analogy with the information needed to understand macroscopic motors (internal combustion engine, for example), the information needed to understand DNA packaging motors is the following: (1) a complete list of relevant variables that describe the motor's state (position of valves, current in spark plugs, and so on, in the case of an internal combustion engine) and (2) a description of how each of these variables changes as a function of time. In the case of a biological motor, one faces the following major problem: One normally works with an ensemble of motors. The ensemble is difficult, if not impossible, to synchronize. In any case, the motors are small enough so that thermal motion causes any synchrony to be lost.

One solution to the synchrony problem is to measure the time-dependence of state variables of a single motor. For example, the nanometry of the previous section has also yielded a plot of length of DNA packaged as a function of time. That is a beginning, but only a beginning because only one state variable (length of DNA packaged) is measured. Nonetheless, the nanometry yields the interesting observation that the rate of DNA packaging has the following two deviations from constancy as a function of time: (1) The rate smoothed for small fluctuations progressively decreases as a function of time. (2) The rate has a repeated small fluctuation that consists of a pause. That is, the rate sometimes becomes zero. The frequency of pauses increases dramatically as more DNA is packaged.

The first observation raises the question of whether the rate also undergoes a time-dependent decrease inside of an infected cell (*in vivo*) or, perhaps, in an *in vitro* system that accurately mimics the *in vivo* system. After all, the biology of survival leads to the conclusion that, if anything, the rate should increase once the DNA packaging motor has committed

to packaging a genome. Proposing an increase in the rate is equivalent to proposing feedback control for the DNA packaging motor, on the basis of the energetics discussed above. Complexity introduced by feedback control is the key differentiating factor between a process that is called *physical* and a process that is called *biological*. In the case of the motors of higher organisms (the motor that moves muscle, for example), the question of feedback control cannot be asked with current *in vitro* systems because these systems have been stripped of most of the components that they have *in vivo*. In the case of DNA packaging motors, this question has not yet been answered.

The observation of pauses raises the question of whether the pauses are part of the function of the motor or, alternatively, a trivial product of overcoming the resistance to packaging. The answer to this question is also not known. To further investigate the DNA packaging motor, more state variables must be measured in real time at the single-motor level while the motor is running. The best strategy appears to be the use of single-motor fluorescence microscopy, perhaps in conjunction with the nanometry of Fig. 3. Fluorescence microscopy has the capacity to observe either a single DNA molecule or its capsid (or both) during packaging. Importantly, fluorescence microscopy has the capacity, in theory, to measure several state variables simultaneously.

3
In vitro DNA Packaging: Complex System Biochemistry

The importance of the fidelity of *in vitro* systems is illustrated by examples from studies of the initiation of DNA packaging. Initiation of DNA packaging has not been discussed above because of its complexity and variability among bacteriophages. The first example is the finding that *in vivo* initiation requires transcription in the case of the related bacteriophages, T3 and T7. However, a purified T3 *in vitro* system has absolutely no dependence on transcription! Nonetheless, a much less purified *in vitro* system was developed and this latter system did have dependence on transcription. The added complexity of transcription presumably occurs as an adaptation to improve packaging success through control of packaging.

A second example is the finding in a very unpurified *in vitro* system that bacteriophage T7 DNA packaging initiation is cooperative among several capsids. This observation is not totally surprising because the capsids are packaging genomes that are all on a single DNA concatemer. However, purified *in vitro* systems have lost their specificity for concatemers and, therefore, would not have revealed the cooperativity. The use of an unpurified *in vitro* system, therefore, has some major advantages. Furthermore, the unpurified *in vitro* T7 system has the technical advantages that in this system (1) fluorescence microscopy can be used to observe both single DNA molecules and single bacteriophage capsids and (2) the system either reduces or eliminates photobleaching; photobleaching is potentially a major problem.

The point is well made in this case that the long-range interest of understanding biology does not coincide with the short-range interest of having neat *in vitro* systems. The long-term eventually arrives, though sometimes suddenly and after a long period of wondering where it is. Progress is not necessarily colinear with time.

See also Bacterial Growth and Division; Electron Microscopy in Cell Biology; Nucleic Acid Packaging of RNA Viruses.

Bibliography

Books and Reviews

Anderson, D., Grimes, S. (2004) The ϕ29 DNA Packaging Motor: Seeking the Mechanism, in: Catalano C.E. (Ed.) *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism*, Landes Publishing, Georgetown, TX; in press.

Bustamante, C., Bryant, Z., Smith, S.B. (2003) Ten years of tension: single-molecule DNA mechanics, *Nature* **421**, 423–427.

Catalano, C.E. (2000) The terminase enzyme from bacteriophage lambda: a DNA-packaging machine, *Cell. Mol. Life Sci.* **57**, 128–148.

Cerritelli, M., Conway, J.F., Cheng, N., Trus, B.L., Steven, A.C. (2003) Molecular Mechanisms in Bacteriophage T7 Procapsid Assembly, Maturation, and DNA Containment, in: Chiu, W., Johnson, J.E. (Eds.) *Advances in Protein Chemistry*, Vol. 64, Virus Structure, Academic Press, San Diego, CA, pp. 301–323.

Daniel, R. (2004) The soil metagenome – a rich resource for the discovery of novel natural products, *Curr. Opin. Biotechnol.* **15**, 199–204.

Fujisawa, H., Morita, M. (1997) Phage DNA packaging, *Genes Cells* **2**, 537–545.

Hendrix, R.W. (2002) Bacteriophages: evolution of the majority, *Theor. Popul. Biol.* **61**, 471–480.

Serwer, P. (2004) T3/T7 DNA Packaging, in: Catalano C.E. (Ed.) *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism*, Landes Publishing, Georgetown, TX; pp. 1–22.

Primary Literature

Cerritelli, M.E., Cheng, N., Rosenberg, A.H., McPherson, C.E., Booy, F.P., Steven, A.C. (1997) Encapsidated conformation of bacteriophage T7 DNA, *Cell* **91**, 271–280.

Serwer, P. (2003) Models of bacteriophage DNA packaging motors, *J. Struct. Biol.* **141**, 179–188.

Serwer, P., Hayes, S.J., Zaman, S., Lieman, K., Rolando, M., Hardies, S.C. (2004) Improved isolation of under sampled bacteriophages: finding of distant terminase genes, *Virology*; in press.

Smith, D.E., Tans, S.J., Smith, S.B., Grimes, S., Anderson, D.L., Bustamante, C. (2001) The bacteriophage straight ϕ29 portal motor can package DNA against a large internal force, *Nature* **413**, 748–752.

Tzil, S., Kindt, J.T., Gelbart, W.M., Ben-Shaul, A. (2003) Forces and pressures in DNA packaging and release from viral capsids, *Biophys. J.* **84**, 1616–1627.

Nucleic Acid Packaging of RNA Viruses

Markus Jaakko Pirttimaa and Dennis Henry Bamford
University of Helsinki, Helsinki, Finland

1 Background 364

2 Tobacco Mosaic Virus (TMV), the Classic Case 365

3 Small, Icosahedral, Positive-strand RNA Viruses 366

4 Double-stranded RNA Viruses 367
4.1 Bacteriophage ϕ6 367
4.2 Yeast Virus L-A 369

5 RNA Packaging Signaling in Animal Viruses 371
5.1 Sindbis Virus 371
5.2 Mouse Hepatitis Virus 372
5.3 Influenza Virus 372
5.4 Hepadnaviruses 372
5.5 Human Immunodeficiency Virus 373

6 Perspectives 374

Bibliography 375
Books and Reviews 375
Primary Literature 375

Keywords

Viral RNA-protein Interactions
RNA virus genomes specifically recognize and bind to proteins encoded by this RNA, eventually leading to assembly of infectious virions.

Virus Genome
It is a nucleic acid molecule. Chemically, the genome can be either ribonucleic or deoxyribonucleic acid (RNA or DNA). These molecules can exist in single- or

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

double-stranded form. Any of these four possible molecules can serve as a virus genome.

Virus Genome Packaging
The event in which the viral nucleic acid is placed inside the protective viral coat.

Viruses are obligate intracellular parasites composed of virus-specific nucleic acid and a protective coat. During virus assembly, the critical stage is the specific recognition of the viral genome with the maturing virus capsid leading to encapsidation (packaging) of the genome to the protective capsid. The formed and released viral particle is metastable, being primed to deliver the genome upon interaction with a new host cell. Viruses with an RNA genome either condense their nucleic acid concomitantly with the capsid assembly or produce empty capsids (procapsids), which translocate the viral genome into the particle in reaction requiring chemical energy in a form of NTPs. RNA viruses can have one to twelve genome segments. It is obvious that very delegate mechanisms have evolved to assure specific packaging of the entire genome either to a single or several particles.

1 Background

Viruses are intracellular parasites that are fully dependent on the metabolic apparatus of the cell. The viral genome has the information for its multiplication in the cell as well as for directing the synthesis of structural components of the virion. The virus genome can be either DNA or RNA, double-stranded (ds) or single-stranded (ss). The nucleic acid interacts with the viral structural proteins to form a nucleic acid–protein complex. This event is referred to as *nucleic acid packaging or encapsidation*, and the formed protein coat protecting the nucleic acid is the *virus capsid*. The capsid can be organized either helically or icosahedrally. In the helical nucleocapsid, an elongated rod-shaped structure, the coat proteins form a helix around the genome. The icosahedral capsid is a spherical shell-like structure enclosing the condensed viral genome (Fig. 1). In some viruses, a membrane additionally surrounds these structures, or is located underneath the capsid.

Double-stranded nucleic acids are antiparallel molecules. In the case of RNA, the strand that can serve as the messenger in protein synthesis is known as the (+) strand and the complementary strand as the (−) strand. There are RNA viruses that have either (+), (−), or dsRNA as their genome. The genome can be a single RNA molecule or segmented. The genome segments can reside in a single capsid or in different capsids (multipartite). The (+)-strand RNA genomes, upon introduction to the cell, can serve directly as messengers, whereas the (−) strand and dsRNA genomes must be transcribed to (+) strands in the cell. This (+) strand synthesis is carried out by virion-associated polymerases, either directly or via a DNA intermediate (retroviruses).

One of the key events during virion assembly is the encapsidation (packaging) of

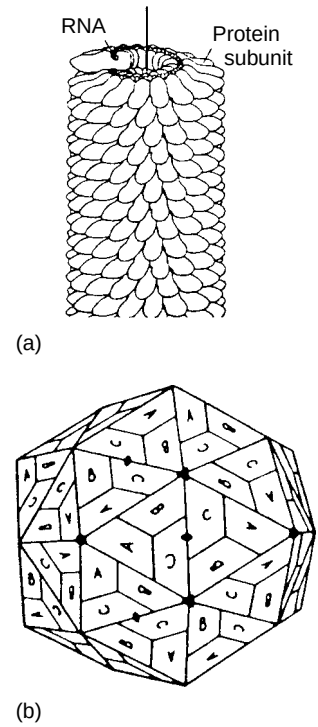

Fig. 1 Principal structural arrangements of virus capsids: (a) A helical capsid, with the coat proteins organized helically around the viral RNA (e.g. tobacco mosaic virus), (b) An icosahedral capsid in which the coat proteins (A, B, C) form a spherical structure enclosing a condensed viral RNA molecule (e.g. a plant (+)-strand RNA virus).

the viral genome into a protein capsid. On the basis of the great variety of genome strategies in RNA viruses, there evidently must also exist variations in the mechanisms used for RNA packaging. It is obvious that only very specific interactions between the RNA and the capsid protein(s) can produce the distinct detection of the viral RNA molecules in the cell and their subsequent packaging. The formation of a functionally metastable virion structure, owing to its responses to stimuli during the virus entry, results in virion disassembly and genome activation. Our understanding of the RNA packaging events relies mostly on a few model systems, but information from a great variety of viruses is rapidly accumulating. This article describes various examples of RNA virus systems about which more detailed packaging information is available.

2 Tobacco Mosaic Virus (TMV), the Classic Case

Tobacco Mosaic Virus (TMV) is a rod-shaped virus composed of a (+)-sense ssRNA molecule (6395 nucleotides) associated with some 2130 identical capsid protein molecules forming a helical structure (Fig. 1a). As long ago as 1955, an *in vitro* reconstitution of infectious virus particles from purified RNA and dissociated coat proteins was achieved. This self-assembly reaction of purified components allowed a detailed study of the packaging mechanism.

The nucleation of the TMV particle starts from the interaction between the genomic RNA and the coat-protein aggregate (the so-called 20S particle or *disk*). The disk is a two-layer cylindrical structure, each layer composed of a ring of 17 coat-protein molecules. An internal region of the RNA called the *origin of assembly* (OAS), located in the region between nucleotides 5443 and 5518, forms a bulged stem-loop structure that binds round the first turn of the disk. This leads to the melting of the base-paired stem and causes the disk to dislocate into a short helix (the "lock-washer" form) entrapping the RNA. The helix elongation continues towards both directions of the RNA molecule. The 5′ direction encapsidation is rapid, whereas

the 3′ direction encapsidation is slow. Furthermore in the 5′ directed reaction, RNA passes through the lumen of the growing rod as it is incorporated into the virion, and predominantly uses the disk subassembly of protein subunits as building blocks. The 3′ direction reaction uses smaller aggregates or single protein subunits. The assembly process is greatly dependent on the pH and ionic conditions, as is the disassembly upon infection. The genome length determines the length of the virion.

3 Small, Icosahedral, Positive-strand RNA Viruses

Small (+)-strand RNA viruses have been found infecting bacteria, plants, and animals. These viruses are among the simplest replication systems. High-resolution structural data, which are available for a number of such viruses, have greatly advanced our understanding of the assembly of these small viruses. The overall structural principles are surprisingly similar in ssRNA viruses infecting plants and animals (Fig. 1b), whereas a different arrangement is found in the corresponding bacterial viruses. In spite of the structural similarities of the capsids, the condensed RNA molecule inside the capsid can be organized differently. It can be either randomly organized (so that it does not contribute significantly to the diffraction pattern obtained from virus crystals) or parts of the RNA can be associated with the capsid proteins in a symmetric fashion (making part of the nucleic acid visible in the X-ray-derived electron density map).

Bacteriophage assembly initiation of ssRNA is probably the best understood among small, icosahedral (+)-strand viruses. These viruses (MS2, R17, etc.) have one coat protein that exists as a dimer. Initially, coat proteins bind specifically to an asymmetric RNA hairpin loop 21-nucleotides long (Fig. 2), which is located in the translation initiation region of the replicase gene. All unpaired nucleotides are necessary. There are two A residues in the loop that are base-specific. There is also an unpaired A in the stem which cannot be changed without loss of activity. The paired stem nucleotides, however, do not show any base specificity (compare also to yeast virus L-A packaging signal in Sect. 4.2: Fig. 5). Later in the infection, when the coat-protein concentration in the cell increases, the binding of the coat protein to this RNA loop leads to translational repression of the replicase synthesis and is most probably also responsible for the initiation of the genome packaging. The initial

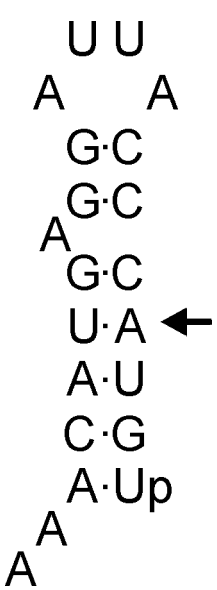

Fig. 2 The bacteriophage R17 coat protein–binding site just upstream of the replicase gene in the R17 genome; the arrow indicates the starting point of the replicase gene.

binding stimulates the binding of additional coat-protein dimers to nonspecific sequences, resulting in a highly cooperative assembly of coat proteins, forming a shell that simultaneously encloses the viral RNA.

An intact ssRNA bacteriophage capsid also contains a single copy of a maturase protein, which is necessary for infection. It has been shown to bind RNA, but capsid proteins can specifically package virus-specific RNA without this protein. In the model represented by ssRNA bacteriophages, the RNA genome is condensed to the capsid simultaneously with the protein shell assembly. Both picornaviruses as well as certain plant viruses (e.g. southern bean mosaic virus, tomato bushy stunt virus, turnip crinkle virus) assemble via discrete intermediate units.

In the turnip crinkle virus assembly model, three coat-protein dimers bind to a specific viral RNA sequence. The N-termini of these proteins are located around the threefold symmetry axis. Additional dimers are then added to this structure. A preformed capsid-packaging model is not yet, however, completely ruled out. Anyway, the packaging is very specific for the (+)-sense genomic RNA. The packaging signal is a 28 nt bulged hairpin loop in the 3′ end of the coat protein coding region, and the size of the packaged RNA is critical for stable assembly of the virions. The picornavirus genomes also contain a 5'-terminal covalently linked protein, but it is not the specificity factor for packaging.

Ordered ssRNA has been observed in the crystal structure of bean pod mottle virus, a comovirus. Some 20% of the RNA is icosahedrally oriented with no sequence specificity. Whether this RNA–capsid protein interaction has implications for RNA packaging remains to be seen. These viruses, however, can form stable empty capsids, indicating that correct capsid protein–protein interactions take place without the genomic RNA. On the other hand, a nodavirus (flock house virus) capsid has been shown to bind portions of the genomic ssRNA in a dsRNA form. About 10 bases in the dsRNA form bind to the coat proteins at the icosahedral twofold axes. The dsRNA is in the A form and is formed from the secondary structures of nonunique sequences. This RNA is an integral part of the capsid quaternary structure, thereby suggesting capsid condensation around the RNA. This model is also supported by the absence of reports of empty precursor particles in nodavirus-infected cells.

4 Double-stranded RNA Viruses

Since cellular polymerases are not capable of operating on dsRNA templates, all dsRNA viruses have a virion-associated, RNA-dependent RNA polymerase. Although in many cases dsRNA viruses have segmented genomes, many fungal viruses have only a single genome segment. The viruses with segmented genomes, in addition to having the specificity for selecting viral RNA for packaging, must have a mechanism ensuring that one copy of each genome segment is included in the virion. Depending on the virus, the number of segments can vary from 1 to 12. Here we concentrate on two dsRNA virus systems (ϕ6 and L-A), where *in vitro* assembly experiments have shed light on the RNA encapsidation reaction.

4.1 Bacteriophage ϕ6

Pseudomonas syringae virus ϕ6 is the type organism of the *Cystoviridae* family

of dsRNA bacteriophages. Like other members of the family, it has three genomic segments that are named according to their sizes (large, L, 6374 bp; medium, M, 4063 bp; and small, S, 2948 bp). These are packaged into a dodecahedral polymerase complex particle made of four viral proteins, which are encoded by the genome segment L. Protein P1 is the major component of the capsid, and forms the structural framework for the particle. It is responsible for the initial binding and recognition of the genomic (+)-strand ssRNAs. P2 is an RNA-dependent RNA polymerase and resides inside the particle at -fivefold symmetry positions. P4 is a nonspecific NTPase forming hexameric, ring-like structures that are located at the particle surface around the fivefold symmetry axes. It provides the energy for the genome packaging and acts also as a helicase. P7 is a small particle stabilizing protein and is needed for stable ssRNA packaging. An additional layer of protein P8 surrounds the polymerase complex particle, forming the nucleocapsid (NC). The nucleocapsid resides in a virus-specific membrane envelope.

Packaging of ϕ6 ssRNA has been studied by using an *in vitro* packaging assay. Empty polymerase complex particles (PCs) can be produced in *Escherichia coli* cells from a plasmid containing a cDNA copy of the L-segment. Purified particles can package, replicate and transcribe the viral RNA segments. The purified NC shell protein P8 can be assembled onto the dsRNA-filled particles, thus making them infectious to spheroplasts of *P. syringae* and leading to the production of infectious complete viruses.

The virus enters the cell via membrane fusion and endocytosis-type events, and loses the structural layers sequentially. In the cytoplasm, NC is activated by the removal of the protein P8 layer. This activation leads to the production of (+)-strand transcripts from all genomic segments (l^+, m^+, and s^+). The L-segment-encoded polymerase complex proteins are synthesized early in infection and self-assemble to empty dodecahedral particles. The synthesized (+)-strands serve also as precursor genomic strands, which are packaged to these polymerase complex particles. The packaging is specific for ϕ6 RNA and the packaging sites determine this specificity. These packaging signals extend for some 210 to 280 noncoding nucleotides from the 5′ end of each segment (Fig. 3). The 18 first 5′ nucleotides are conserved in all segments. In each segment, there is also a 10–12-nucleotide-long homologous sequence within the first 100 nucleotides from the terminus. This sequence forms a hairpin structure with a very stable tetraloop UACG. This is thought to be a ϕ6–specific signal in the packaging sites. In segment s^+, there is an extra copy of this tetraloop. Otherwise, the secondary structures of the packaging sites are very different, although some similar single-stranded regions between s^+ and m^+, and m^+ and l^+ segments can be observed.

Packaging of ϕ6 RNA is sequential. The polymerase complex recognizes the three ssRNA genomic precursor molecules in a segment-specific manner and translocates the RNA molecules into the interior of the particle so that the s^+ segment is packaged first, followed by m^+ and then l^+. The next step in the virus life cycle is the synthesis of the (−)-strands, which occurs inside the polymerase complex particle using the (+)-strands as the templates. The (−)-strand synthesis is coupled to the packaging in such a way that it does not occur before the packaging of l^+ segment has started. The particles with

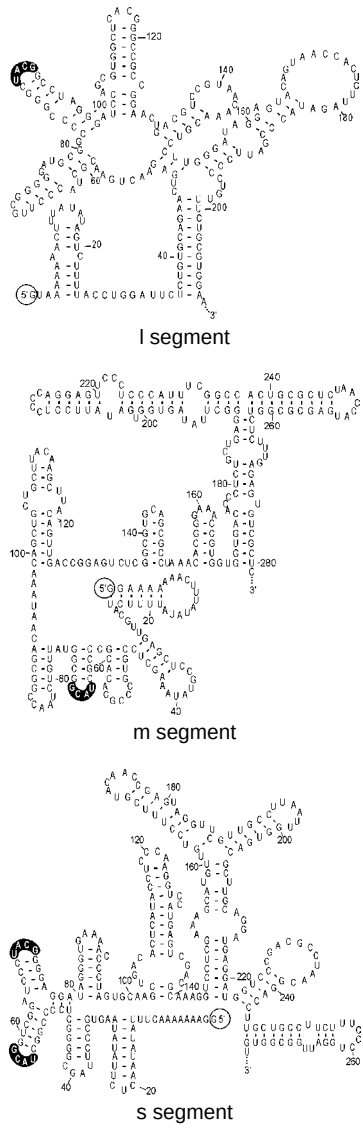

Fig. 3 Secondary structures of the packaging regions of bacteriophage ϕ6 (+)-strand genomic segments based on chemical and enzymatic probing and phylogenetic comparisons. Conserved secondary structure elements are highlighted in black.

synthesized dsRNA genomes move to (+)-strand synthesis mode.

The polymerase complex particle has to go through considerable conformational changes during packaging and particle maturation. The empty particle is smaller and more angular than the dsRNA-filled particle. Estimations of the inside volumes of different particles show that the particle has to expand during ssRNA packaging in order to have enough space for all the three segments. This could explain the sequential packaging: different segment-specific high-affinity binding sites are exposed on the particle surface upon particle expansion.

Energy for the packaging is provided by the NTP hydrolysis carried out by the P4 hexamers. One of these hexamers, however, is chemically distinct, being attached to the particle in a more stable manner. Mutant or detergent-treated particles containing reduced amounts of P4 (equivalent to occupancy of one vertex only) have shown to be efficient in RNA packaging and (−)-strand synthesis, but the (+)-strand synthesis in these particles is completely abolished. It can be concluded that the P4 hexamers are needed both in RNA packaging and (+)-strand synthesis, and the vertex used for packaging cannot be used in (+)-strand synthesis (mRNA exit).

Figure 4 depicts the NC activation, procapsid packaging, and NC maturation. It remains to be seen whether members of the reovirus family follow this general scheme for their RNA packaging. It is intriguing to address the specific RNA–protein interactions leading to the highly selective packaging of all the viral genomic RNA segments.

4.2 Yeast Virus L-A

Yeast dsRNA virus L-A is a degenerate virus, as are all viruses residing in fungi.

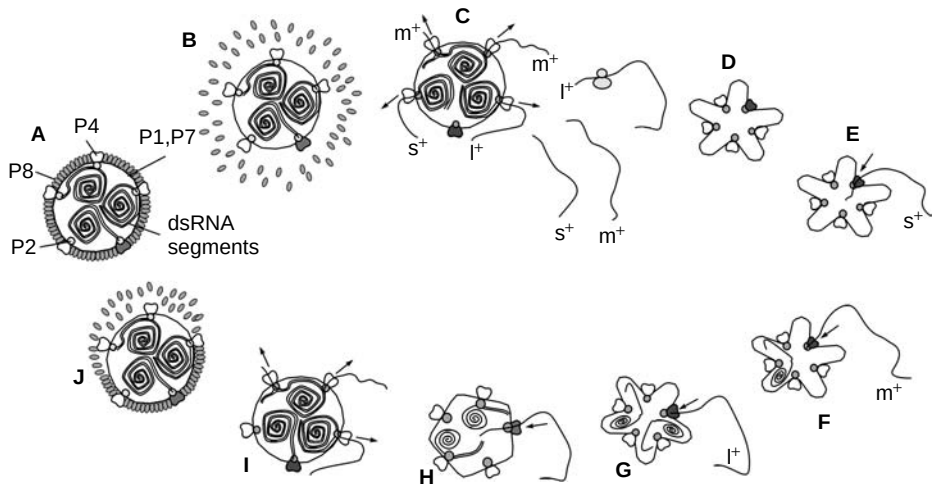

Fig. 4 Bacteriophage ϕ6 packaging and replication model. (a) Protein and RNA components of the nucleocapsid, (b) nucleocapsid activation by the removal of P8 shell, (c) transcription of (+)-strand genome segments, (d) assembly of empty polymerase complexes coded by the large genome segment, (e–g) sequential packaging of the genome segments through the packaging vertex, (h) initiation of (-)-strand synthesis, (i) (+)-strand synthesis, and (j) inhibition of (+)-strand synthesis by the assembly of P8 shell.

They exist intracellularly only, and no external virus particles have been observed. L-A is a parasite of *Saccharomyces cerevisiae*. The L-A system also supports the replication of several satellite dsRNAs in addition the viral genome. The L-A genome is a 4.6-kb long, single dsRNA molecule, whereas all the satellite dsRNAs are smaller. The satellites code for a secreted toxic protein and for immunity to that toxin.

The L-A genome codes for the major capsid protein (Gag) and for a fusion protein that has a Gag as its N-terminal and an RNA polymerase domain as its C-terminal portion (Gag–Pol). The Gag-Pol fusion protein is formed by a −1 ribosomal frameshift event. The frameshifting efficiency controls the formation of the Gag–Pol protein. *In vitro*, ~1.8% of the Gag transcripts are read to produce Gag–Pol fusion protein. The ratio of Gag to Gag–Pol production is very close to that estimated from the isolated virus particles; 120 Gag proteins, 2 Gag–Pol proteins (and one RNA molecule).

At about 400 nucleotides from the 3′ end of the (+)-strand, there resides a 24-base stem-loop structure with a five-nucleotide stem and an A bulged at the 5′ side of the stem. This structure and ten 5′ adjacent bases form the RNA packaging site (Fig. 5). The RNA packaging site binds to the Pol domain of the Gag–Pol fusion protein. The stem structure, but not the sequence of the stem, is necessary for the binding. The loop sequence, as well as the bulged A residue, however, are critical for the binding.

The packaging model involves the formation of an initiation complex where two Gag–Pol proteins dimerize and bind one (+)-strand RNA molecule. The Gag domain of the fusion protein primes the assembly of additional Gag proteins to form an icosahedral shell, condensing the RNA inside (Fig. 6). The possibility of packaging

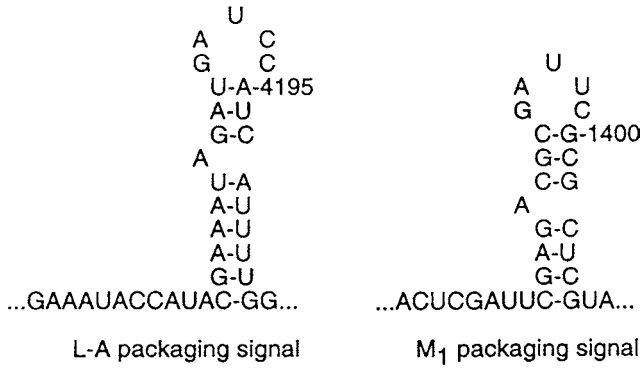

Fig. 5 Packaging signal of the L-A virus and that of its satellite RNA M_1. The structures are very similar and the loop sequences nearly identical. (Figure kindly provided by Dr. Reed B. Wickner, U.S. National Institutes of Health.)

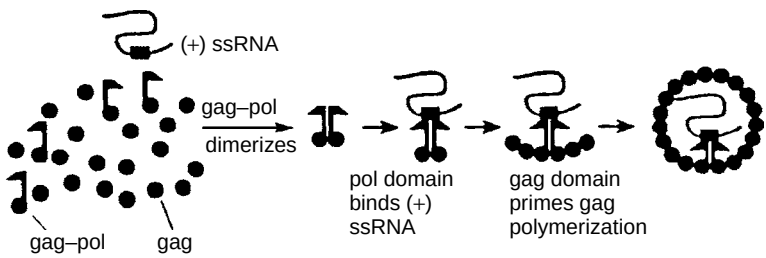

Fig. 6 RNA-packaging model of the L-A family of dsRNA viruses. (Figure kindly provided by Dr. Reed B. Wickner, U.S. National Institutes of Health.)

of a preformed capsid is, however, not completely ruled out.

Following packaging, the (−)-strand synthesis takes place inside the particle. The particle size is designed to give room for a single L-A dsRNA genome. If, however, the particle packages a short satellite RNA using a mechanism and a mode of signaling similar to those for the L-A genome, the RNA synthesis produces progeny RNA molecules that stay inside the particle as long as the RNA mass is equal to that of the L-A genome. This system results in particles with multiple dsRNA segments, the number depending on the size of the packaged molecule.

5 RNA Packaging Signaling in Animal Viruses

5.1 Sindbis Virus

Sindbis virus, a togavirus, has icosahedral symmetry. The assembly of the virion requires two steps. First, the genomic (+)-strand ssRNA interacts with the capsid protein leading to the formation of the nucleocapsid. Second, the nucleocapsid specifically interacts with the viral membrane proteins, which are translocated to the plasma membrane. This event results in the release (budding) of viruses. The genomic RNA is used to translate the nonstructural proteins from the 5′ portion

of the molecule. Structural proteins are translated from a subgenomic RNA (identical to one-third of the 3′ end of the genomic RNA). Only the genomic RNA, and not the subgenomic fragment, is packaged. The packaging signal is located at a 132-nucleotide region (between nucleotides 945 and 1076 in the Sindbis virus genome).

The binding site with predicted strong secondary structure lies within the coding region of a nonstructural protein nsP1. The binding of this region to the capsid protein is supposed to trigger protein–protein interactions between the capsid proteins leading to the nucleocapsid formation. The crystal structure of the Sindbis virus capsid protein reveals a dimer. However, monomeric capsid protein binds RNA. The RNA binding site is some 30-amino acids long around amino acid 100 of the protein, which is about 246-amino acids long. The binding site is located between two domains of this protein. The N-terminal domain is largely unstructured (flexible) and basic, a good candidate for an RNA interacting protein. The other domain forms the capsid structure.

5.2
Mouse Hepatitis Virus

Mouse hepatitis virus, a coronavirus, is an enveloped virus with a helical capsid. The genome is a positive-sense ssRNA molecule. Also, in this case there are subgenomic RNA species in addition to the genome-length species. Only the genome-length RNA is packaged. Deletion mapping has located the packaging signal to a 190-nucleotide internal sequence close to the 3′ end of gene 1. There are indications of a stem-loop RNA structure in this region. The nucleocapsid protein N has been shown to bind all viral mRNAs, whereas a transmembrane viral envelope protein M binds specifically to the packaging signal. This is the first known example of an RNA virus in which interaction between a viral envelope protein and RNA packaging signal determines the selective RNA packaging into viral particles.

5.3
Influenza Virus

The influenza virus, an orthomyxovirus, has eight different ssRNA genome segments of negative polarity, which are associated with four polypeptides forming a ribonucleoprotein core. This core exhibits RNA-dependent RNA polymerase activity and is enclosed into the viral envelope. Of the terminal untranslated 3′ and 5′ nucleotides of the genomic RNAs, about 15 are involved in forming a circle via a panhandle structure. Some 20 nucleotides at the termini of the segments are sufficient signals for transcription, replication, and packaging of the RNAs into the virus particles. It is assumed that the packaging of the genomic segments is not accurate, but more than eight segments are packaged. This somewhat random packaging leads to different assortments of genomic segments among the viral particles. Thus, with only a few particles having complete sets of genomic segments, specific infectivity should be low, as in fact has been observed. However, a more specific packaging system has not been ruled out.

5.4
Hepadnaviruses

In both the hepadna- and retroviruses, the packaged nucleic acid is positive-sense RNA, but the virus particle contains an RNA-dependent DNA polymerase (reverse

transcriptase), which converts the packaged RNA to DNA form.

Hepadnaviruses are small, enveloped icosahedral particles. Upon infection, the nucleocapsid is transported to the host cell nucleus. The 3.2-kb genome is converted into a covalently closed circular molecule, which is transcribed to a genome-length RNA molecule, as well as to two subgenomic RNAs. Only the genome-length (+)-strand RNA is encapsidated to the nucleocapsid along with the reverse transcriptase. At this stage, the packaged RNA molecule acts as a template for the synthesis of the complementary (−)-DNA strand, which is further synthesized to a partially duplex DNA molecule. The nucleocapsid acquires the virus-specific envelope from the endoplasmic reticulum, after which the viral particle is removed from the cell by the vesicular transport pathway.

The genome-length RNA in the hepadnavirus contains a region designated as "ε". It is a sequence about 90-nucleotides long, located near the 5′ end of the molecule, and it is necessary and sufficient for genome packaging. The same signal is also found at the 3′ end of the genome but does not have packaging activity at this position. The ε sequence forms a hairpin loop, which also binds the reverse transcriptase and is the initiation site for the reverse transcription. Even though the packaging is coupled to reverse transcription, these two events are independent. Nonfunctional reverse transcriptase does not prevent packaging. However, if there are no reverse transcriptase proteins available, only empty capsids are formed. The capsid (core) protein has a basic C-terminus that might have nonspecific nucleic acid binding functions. It is not, however, clear how ε, reverse transcriptase, and the capsid proteins interact to initiate the encapsidation reaction. An interesting feature in hepadnavirus packaging is a cis preference. When wild-type genomes are used to complement reverse transcriptase mutants, the progeny genomes packaged are preferentially of wild-type origin.

5.5 Human Immunodeficiency Virus

Human immunodeficiency virus (HIV) is an enveloped polyhedral retrovirus with a (+)-strand 9.2 kb RNA genome. As in other retroviruses, the viral particle contains two copies of the genomic RNA molecule. Upon entry into the host cell, the retroviral RNA genome is converted into a DNA molecule by the virion-associated reverse transcriptase. After the virus particle has entered the nucleus, the viral genome integrates with the host genome. The integrated provirus is transcribed by the cellular RNA polymerase II, producing full-length genome transcripts. These are spliced to subgenomic RNAs encoding viral regulatory proteins including Rev. Later in the infection, the accumulation of Rev and its binding to the Rev-responsive element (RRE) located at the 3′ end of the Env gene stabilizes these RNAs and splicing is avoided or limited. Rev contains a leucine-rich nuclear export signal, and together with cellular and possibly viral chaperone proteins the unspliced viral RNA genomes are exported from the nucleus to the cytoplasm.

Retroviral genomes code for polyprotein precursors (Gag, Pol, and Env). The virus capsid obtains the envelope by budding through the plasma membrane. The reverse transcriptase (Pol) is synthesized as a Gag–Pol fusion protein by a ribosomal frame shifting (see also Sect. 4.2 on yeast virus L-A). The Gag region produces the virion capsid proteins. After virus

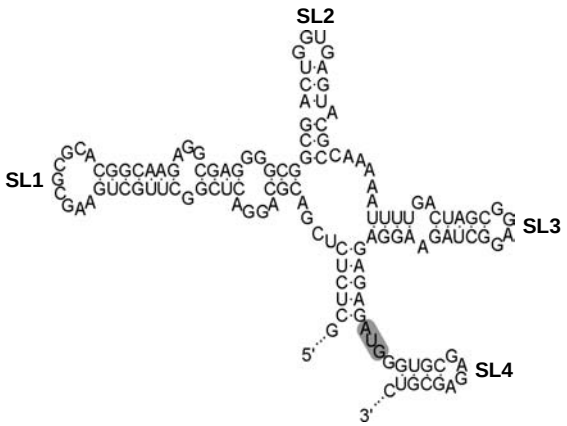

Fig. 7 Secondary structure model of the major packaging signal in human immunodeficiency virus 1. Stem loop 1 (SL1), putative dimerization initiation site. SL2, major splice donor. SL3, principal packaging determinant. Gag-initiation codon shown with gray background.

maturation, the reverse transcriptase is activated by a proteolytic cleavage, which leads to the binding of the reverse transcriptase to the RNA. It is assumed that the packaging of the reverse transcriptase is based on the Gag (capsid) domain of the fusion protein.

The packaging of retroviral RNA is highly specific, since only full-length RNAs are packaged. As in hepadnaviruses, the retrovirus RNA contains a packaging signal (ψ) near the 5′ end of the genomic RNA. Figure 7 shows the secondary structure model of this packaging region in HIV-1. Stem loop 3 contains the major packaging signal motif with the GGAG-terminal loop. This sequence has been shown to bind the zinc finger motif of the nucleocapsid (NC) region of the Gag polyprotein. Deletions in the NC zinc fingers will disrupt packaging; deletion of the stem loop 3 leads to only partial packaging defects, since the GGNG motif appears multiple times in lentiviral packaging regions. Stem loop 1 contains the dimer initiation site (DIS) with terminal palindromic loop (GCGCGC), which is able to interact with the sister genomic RNA. Stem loop 2 contains the splice donor site and has relatively little effect on packaging. There are also accessory packaging signals outside the major packaging region.

In HIV-2, the major packaging signal is in a region 5′ to the splice donor, and the packaging region is found in all RNA species. HIV-1 can also package HIV-2 RNA, but captures both spliced and unspliced RNA with equal efficiency. HIV-2 seems to package RNA in cis, cotranslationally, compensating for the presence of the packaging site in all RNA species, whereas HIV-1 is capable of picking up RNA in trans.

The hepadna- and retrovirus genomes with the RNA packaging sites indicated are depicted in Fig. 8.

6 Perspectives

Virus RNA packaging is a highly selective event, in which almost exclusively viral RNA is encapsidated. The selectivity is obtained by means of specific RNA–capsid protein interactions. These interactions may also have translational regulatory effects. There is a vast amount of information about DNA–protein interactions, but our understanding of RNA–protein

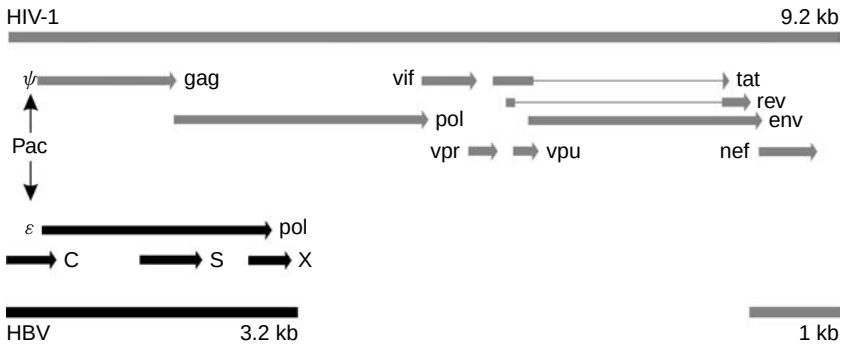

Fig. 8 Genome organizations of human immunodeficiency virus 1 (HIV-1) and human hepatitis B virus (HBV). The "Pac arrow" indicates localization of the encapsidation signals.

interactions is still relatively limited. Surely RNA viruses will be in the forefront in helping us to elucidate the interactions between the two most active cellular components, RNA and protein.

The development of recombinant DNA technologies (including reverse transcription) and the understanding of the packaging signaling in many virus systems have opened the opportunity to package not only the viral RNA but also heterologous RNA into viruses. These advances are of crucial importance in helping us to understand the packaging mechanisms.

It is presumed that RNA viruses will play an important role in the fast-developing field of RNA biology, including gene silencing. The ability to package and amplify desired RNA segments, either single- or double-stranded, is an asset that will certainly be utilized.

See also Bacterial Growth and Division; Electron Microscopy in Cell Biology; Nucleic Acid Packaging of DNA Viruses.

Bibliography

Books and Reviews

Butler, P.J.G. (1999) Self-assembly of tobacco mosaic virus: the role of an intermediate aggregate in generating both specificity and speed, *Phil. Trans. R. Soc. Lond. B* **354**, 537–550.

Klug, A. (1999) The tobacco mosaic virus particle: structure and assembly, *Phil. Trans. R. Soc. Lond. B* **354**, 531–535.

Lever, A.M.L. (2000) HIV RNA packaging and lentivirus-based vectors, *Adv. Pharmacol.* **48**, 1–28.

Poranen, M.M., Pirttimaa, M.J., Bamford, D.H. (2004), in: Viral Genome Packaging, Catalano, C. (Ed.) Landes Bioscience, Georgetown, TX.

Rossmann, M.G., Johnson, J.E. (1989) Icosahedral RNA virus structure, *Annu. Rev. Biochem.* **58**, 533–573.

Schlesinger, S., Makino, S., Linial, M.L. (1994) Cis-acting genomic elements and trans-acting proteins involved in the assembly of RNA viruses, *Semin. Virol.* **5**, 39–50.

Wickner, R.B. (1996) Prions and RNA viruses of Saccharomyces cerevisiae, *Annu. Rev. Genet.* **30**, 109–139.

Primary Literature

Bloomer, A.C., Champness, J.N., Bricogne, G., Staden, R., Klug, A. (1978) Protein disk of

tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits, *Nature* **276**, 362–368.

Chen, Z.G., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kamer, G., Shanks, M., Lomonossoff, G., Johnson, J.E. (1989) Protein-RNA interactions in an icosahedral virus at 3.0 A resolution, *Science* **245**, 154–159.

De Guzman, R.N., Wu, Z.R., Stalling, C.C., Pappalardo, L., Borer, P.N., Summers, M.F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element, *Science* **279**, 384–388.

Fosmire, J.A., Hwan, K., Makino, S. (1992) Identification and characterization of a coronavirus packaging signal, *J. Virol.* **66**, 3522–3530.

Frilander, M., Gottlieb, P., Strassman, J., Bamford, D.H., Mindich, L. (1992) Dependence of minus-strand synthesis on complete genomic packaging in the double-stranded RNA bacteriophage phi 6, *J. Virol.* **66**, 5013–5017.

Frolova, E., Frolov, I., Schlesinger, S. (1997) Packaging signals in alphaviruses, *J. Virol.* **71**, 248–258.

Fujimura, T., Wickner, R.B. (1988) Gene overlap results in a viral protein having an RNA binding domain and a major coat protein domain, *Cell* **55**, 663–671.

Fujimura, T., Esteban, R., Esteban, L.M., Wickner, R.B. (1990) Portable encapsidation signal of the L-A double-stranded RNA virus of S. cerevisiae, *Cell* **62**, 819–828.

Geigenmüller-Gnirke, U., Nitschko, H., Schlesinger, S. (1993) Deletion analysis of the capsid protein of Sindbis virus: Identification of the RNA binding region, *J. Virol.* **67**, 1620–1626.

Gottlieb, P., Qiao, X., Strassman, J., Frilander, M., Mindich, L. (1994) Identification of the packaging regions within the genomic RNA segments of bacteriophage ϕ6, *Virology* **200**, 42–47.

Gouet, P., Diprose, J.M., Grimes, J.M., Malby, R., Burroughs, J.N., Zientara, S., Stuart, D.I., Mertens, P.P. (1999) The highly ordered double-stranded RNA genome of bluetongue virus revealed by crystallography, *Cell* **97**, 481–490.

Ippolito, J.A., Steitz, T.A. (2000) The structure of the HIV-1 RRE high affinity rev binding site at 1.6 Å resolution, *J. Mol. Biol.* **295**, 711–717.

Junker-Niepmann, M., Bartenschlager, R., Schaller, H. (1990) A short cis-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA, *EMBO J.* **10**, 3389–3396.

Naitow, H., Tang, J., Canady, M., Wickner, R.B., Johnson, J.E. (2002) L-A virus at 3.4 Å resolution reveals particle architecture and mRNA decapping mechanism, *Nat. Struct. Biol.* **9**, 725–728.

Namba, K., Stubbs, G. (1986) Structure of tobacco mosaic virus at 3.6 Å resolution: implications for assembly, *Science* **231**, 1401–1406.

Namba, K., Pattanayek, R., Stubbs, G. (1989) Visualisation of protein-nucleic acid interactions in a virus. Refined structure of intact tobacco mosaic virus at 2.9 Å resolution by X-ray fiber diffraction, *J. Mol. Biol.* **208**, 307–325.

Narayanan, K., Makino, S. (2001) Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging, *J. Virol.* **75**, 9059–9067.

Olkkonen, V.M., Gottlieb, P., Strassman, J., Qiao, X., Bamford, D.H., Mindich, L. (1990) In vitro assembly of infectious nucleocapsids of bacteriophage phi6: formation of a recombinant double-stranded RNA virus, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9173–9177.

Pirttimaa, M.J., Bamford, D.H. (2000) RNA secondary structures of the bacteriophage phi6 packaging regions, *RNA* **6**, 880–889.

Pirttimaa, M.J., Paatero, A.O., Frilander, M.J., Bamford, D.H. (2002) Nonspecific nucleoside triphosphatase P4 of double-stranded RNA bacteriophage phi6 is required for single-stranded RNA packaging and transcription, *J. Virol.* **76**, 10122–10127.

Pollack, J.R., Ganem, D. (1993) An RNA stem-loop structure directs hepatitis B virus genomic RNA encapsidation, *J. Virol.* **67**, 3245–3263.

Poranen, M.M., Bamford, D.H. (1999) Packaging and replication regulation revealed by chimeric genome segments of double-stranded RNA bacteriophage phi6, *RNA* **5**, 446–454.

Poranen, M.M., Paatero, A.O., Tuma, R., Bamford, D.H. (2001) Self-assembly of a viral molecular machine from purified protein and RNA constituents, *Mol. Cell* **7**, 845–854.

Qu, F., Morris, T.J. (1997) Encapsidation of turnip crinkle virus is defined by a specific

packaging signal and RNA size, *J. Virol.* **71**, 1428–1435.

Tchatalbachev, S., Flick, R., Hobom, G. (2001) The packaging signal of influenza viral RNA molecules, *RNA* **7**, 979–989.

Tiley, L.S., Malim, M.H., Tewary, H.K., Stockley, P.G., Cullen, B.R. (1992) Identification of a high-affinity RNA binding site for the human immunodeficiency virus type 1 Rev protein, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 758–762.

Valegard, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J., Liljas, L. (1994) Crystal structure of an RNA bacteriophage coat protein-operator complex, *Nature* **371**, 623–626.

Zimmern, D. (1977) Nucleotide sequence at the origin for assembly on tobacco mosaic virus RNA, *Cell* **11**, 463–482.

Zimmern, D., Butler, P.J.G. (1977) The isolation of tobacco mosaic virus RNA fragments containing the origin for viral assembly, *Cell* **11**, 455–462.

Nucleic Acids (DNA) Damage and Repair

Vilhelm A. Bohr
National Institute on Aging, NIH, Baltimore, MD

1 Introduction 381
1.1 DNA Damage 381
1.2 DNA Base Modifications after Exogenous or Endogenous Exposures 382
1.3 Detection of DNA Damage 384
1.4 Consequences of DNA Damage 386
1.5 DNA Repair 387
1.1.1 Base Excision Repair (BER) 389
1.1.2 Nucleotide Excision Repair (NER) 390
1.1.3 Genomic Heterogeneity of NER 391
1.6 Mismatch Repair 391
1.7 Mitochondrial DNA Repair in Mammalian Cells 391
1.8 Age Related Changes in DNA Repair 392
1.9 Diseases with Defective DNA Repair 393
1.10 Perspectives 393

Bibliography 393
Books and Reviews 393
Primary Literature 393

Keywords

Base Excision Repair
DNA repair process that removes simple, monofunctional base lesions in DNA.

DNA Damage
Damage to any of the components of DNA: bases, sugar, and phosphate.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

DNA Repair
Enzymatic processes that survey and repair damage to DNA.

Nucleotide Excision Repair
DNA repair process that removes bulky DNA base lesions.

Abbreviations

NER	nucleotide excision repair
BER	base excision repair
MMR	mismatch repair
UV	ultraviolet
8-oxoG	8-hydroxyguanine
Adduct	DNA damage, chemical association with a DNA base
Lesion	DNA damage
Glycosylase	enzyme that cleaves the glycosylic bond between the DNA base and the sugar-phosphate backbone of DNA
ROS	reactive oxygen species
pol β	DNA polymerase β
CS	Cockaynes syndrome
WS	Werner syndrome
TCR	transcription coupled repair
GGR	general genome repair
mtDNA	mitochondrial DNA

All living organisms are constantly exposed to exterior agents and endogenous sources that interact with the macromolecules and cause substantive damage. This damage affects the function of all cellular macromolecules, including proteins, RNA, and DNA, and the damage accumulates over the life span of individuals. While RNA and proteins can be resynthesized, the genetic material, DNA, is the master molecule of life and it needs to be maintained intact for normal cellular functions to take place and to avoid death and disease. Thus, cellular steps have evolved to remove this damage from DNA. These processes are called DNA repair, and they involve a series of enzymatic processes that detect the damage and remove it. There are a number of DNA repair pathways and there are molecular interactions between them, and the DNA repair processes in general also connect with a number of other cellular processes. In this chapter, I will discuss the nature of the damage to the DNA and the various DNA repair processes that exist in higher cells. Examples will be given where specific protein functions are discussed in more detail, and references will be

provided so that the reader can seek other sources for more detail. A lack of DNA repair capacity can lead to cancers and to a number of other specific disorders, and some of these will be discussed. In particular, there is a connection between defective DNA repair and both the incidence of cancer and the aging process.

1 Introduction

DNA is the master molecule that serves as the blueprint for the formation of all proteins and enzymes in every organism. The proteins then generate all the other substances in our cells. Thus, it is essential for reproduction, growth, maintenance, and for sustaining normal living that the DNA remains intact so that the genetic code can be read correctly. The stability and quality of the DNA is a prerequisite for normal cellular functions and there is good evidence that damage to the DNA can lead to cellular dysfunction, cancer and other diseases, cell death, and aging. A major theory holds that much of the aging phenotype (changes that can be observed) is caused by the gradual accumulation of DNA damage over a life span. DNA damage occurs at a high frequency because of metabolic processes and environmental factors such as various exposures to chemicals and radiations and the intake of food and drugs. The prevention or repair of DNA damage is thus a major concern in biology and medicine.

Cells contain an elaborate defense system against damage to DNA. There are processes that prevent the formation of the damage and there are processes that remove the DNA damage once it has formed. This latter system is called DNA repair, and it involves a multitude of specific proteins that can be subdivided into a set of different molecular pathways. These molecular processes interact with each other and also have strong interactions with other cellular processes. For example, DNA repair pathways are directly linked to the transcription machinery. The DNA is constantly scanned and surveyed, and as soon as a DNA lesion is detected, the repair enzymes are directed to it so that it can get repaired. In recent years, we have gained much new insight into these DNA repair processes and how they interact with other processes, and this has lead to new understanding about disease processes and new ideas about therapeutic intervention against major diseases.

1.1 DNA Damage

The long, thin DNA molecules contain three components: nitrogen-rich bases, sugar groups, and phosphate groups. This basic structure is shown in Fig. 1. The composition of the four types of bases adenine, guanine, cytosine, and thymidine make up the genetic code. Damage to DNA can occur to any one of its components. The bases are the most reactive components and there is far more knowledge about changes to these than about changes to the sugar or phosphate components of DNA. Also, many chemicals or carcinogens will form adducts (lesions) with new chemical groups being attached to existing DNA bases. Figure 1 shows sites where DNA damage can be inflicted in DNA.

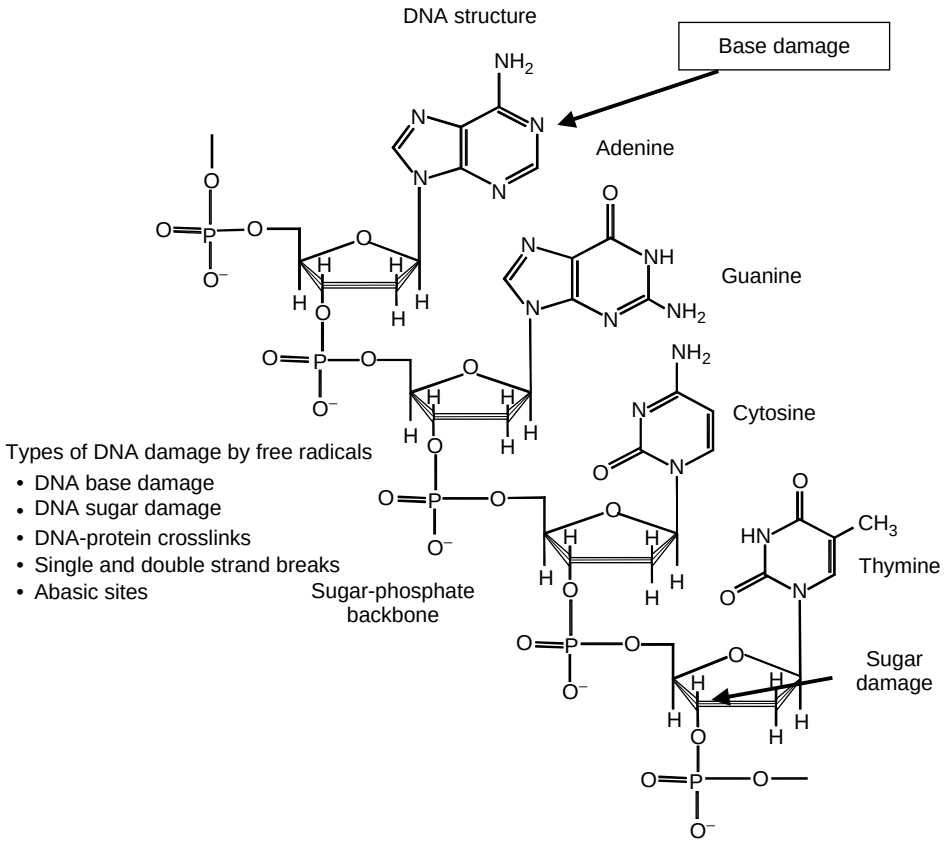

Fig. 1 Basic DNA structure and the types of damage that can be found in DNA.

1.2 DNA Base Modifications after Exogenous or Endogenous Exposures

Living organisms are constantly exposed to stress from environmental agents and from endogenous metabolic processes. For example, irradiation from various sources can directly damage bases in DNA. Ultraviolet irradiation from exposure to sunlight creates certain DNA lesions called photoproducts, shown in Fig. 2. Irradiation from γ-ray sources such as X-rays creates a large number of different kinds of lesions in DNA, including base modifications, sites with a loss of a base and breaks in the DNA strand. Since DNA contains two strands running in parallel but opposite directions, breaks can be either single stranded or double stranded. A large number of components of our food intake can directly damage the DNA. These include various carcinogens and chemicals that cause DNA damage, either by direct reactions with DNA or via metabolic modifications of the compounds that then make them reactive with DNA. For example, aromatic amines are found in variety of foods and are known to cause DNA damage and to be highly mutagenic. A number of poisons work by attacking the DNA and damaging it.

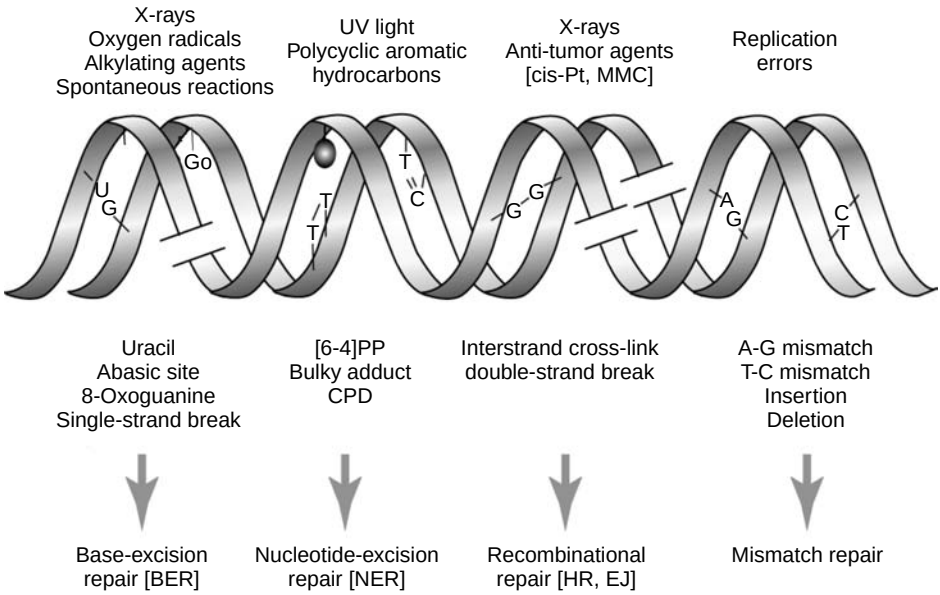

Fig. 2 Specific DNA lesions and some of the pathways that repair them.

An example of this is the poisonous gas nitrogen mustard that causes modification of DNA bases and which can also cross-link DNA bases on opposite DNA strands to generate so-called interstrand cross-links. The interstrand cross-links cause serious havoc in the cell by blocking the progression of polymerases that replicate or transcribe the DNA.

Some of the common lesions induced in the genomic DNA by different DNA damaging agents are shown in Fig. 2. This Figure also shows the various DNA repair processes that deal with the different lesions and this will be discussed later. UV irradiation generates two major types of lesions in DNA, the so-called photoproducts cyclobutane pyrimidine dimers and 6-4 photoproducts, while many DNA damaging agents induce a broad spectrum of lesions. The DNA repair pathway, nucleotide excision repair (NER), to be discussed later, removes the bulky and helix-distorting DNA adducts induced in the genomic DNA by physical and chemical agents. Chemotherapeutic agents often generate lesions varying from simple methylation of bases to complex bulky monoadducts and DNA-protein cross-links. Cisplatin is a classic therapeutic agent against cancers, and the most frequent lesion formed by this compound is the intrastrand adduct between adjacent purines (GG or AG). The major lesions induced by monofunctional alkylating agents such as methylmethanesulfonate include *N*-7-alkylguanine, O^6-alkylguanine, *N*-3-alkyladenine and O^4-alkylthymidine. Phosphotriesters can be formed by alkylation of phosphodiesters by alkylating agents. 4-Nitroquinoline 1-oxide (4-NQO) is a chemical carcinogen that forms several different types of DNA base lesions. Bleomycin, an antitumor drug, intercalates into DNA and generates free radicals resulting in single and double strand breaks, see Fig. 2. Acetylaminofluorene, a potent procarcinogen, is metabolically activated *in vivo* to form

N-hydroxy-2-acetylaminofluorene (N-OH-AAF). The aforementioned few examples illustrate the complexity of different DNA lesions that have to be dealt with to maintain the integrity of genomic DNA. Examples of some different categories of lesions are shown in Fig. 2. Some of these are more bulky than others and therefore affect the DNA structure more. Lesions can be categorized according to the degree of distortion that they afflict in the DNA, and this distortion appears to be important in the selection of the DNA repair pathway that is then chosen to remove the lesion.

Many different DNA base changes have been observed following oxidative stress, and these lesions are widely considered as instigators of the development of cancer, aging, and neurological degradation. After exposure to oxidative reagents or to oxidative stress, the resulting reactive oxygen species (ROS) attack proteins, lipids, and DNA. Since proteins and lipids are readily degraded and resynthesized, the most significant consequence of oxidative stress is thought to be the DNA modifications, which can become permanent via the formation of mutations and other types of genomic instability. The endogenous attack on DNA by ROS species generates a low steady-state level of DNA adducts that have been detected in DNA from human cells. Over 100 different oxidative base modifications in DNA have been measured in human cells, Fig. 3. The best known and most widely studied oxidative DNA base lesion is 8-hydroxyguanosine (8-oxoG), Fig. 3.

Oxidative DNA damage is thought to contribute to carcinogenesis and studies have shown that it accumulates in cancerous tissue. Notably, the cumulative risk of cancer increases dramatically with age in humans, and cancer can, in general terms, be regarded as a degenerative disease of old age. There is evidence for the accumulation of oxidative DNA damage with age, and if 8-oxoG remains in DNA, it can cause permanent alterations in the genetic code (mutations) as shown in Fig. 4. During replication, an adenine base is often mistakenly inserted into the opposite DNA strand by the DNA polymerase. In the next round of replication, the DNA polymerase now inserts a thymidine base opposite to the adenine. The result after two rounds of replication is the formation of a permanent mutation in DNA.

1.3 Detection of DNA Damage

DNA can be extracted from human cells or tissues and then analyzed chemically for its components. There are various methods to detect DNA modifications, including chemical analyses and chromatographical measurements of DNA. One issue, hotly debated, is the choice of method to purify DNA from cells or tissues. Many of the available extraction procedures introduce new DNA damage during the process of purification.

A number of enzymatic methods have been used to detect DNA damage. Here, specific enzymes or antibodies detect certain kinds of DNA base modifications and/or adducts formed in DNA. These enzymes become molecular "probes" for the damage and this can be a very sensitive approach. Furthermore, radioactive labeling procedures can be used as a simple and easy way to measure several modifications in DNA. In recent years, there has been a great deal of interest in detecting the oxidative DNA base modifications under various physiological conditions. Some of the methods used are shown in Fig. 5. This Figure shows some approaches for available analyses. High-pressure liquid

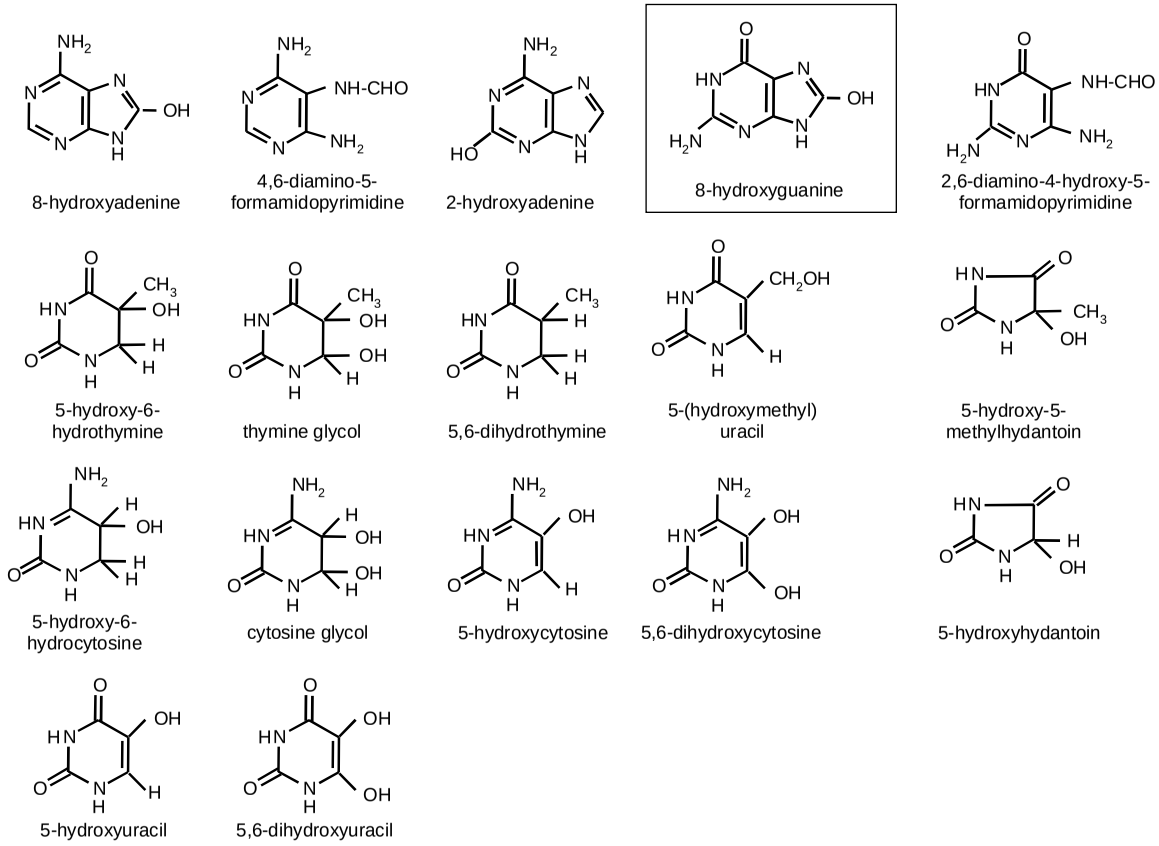

Fig. 3 Examples of oxidative DNA base modifications observed in human tissue. Modified from Olinski, R., Zastawny, T., Budzbon, J., Skokowski, J., Zegarski, W., Dizdaroglu, M. (1992) DNA base modifications in chromatin of human cancerous tissues, *FEBS Lett.* **309**, 193–198.

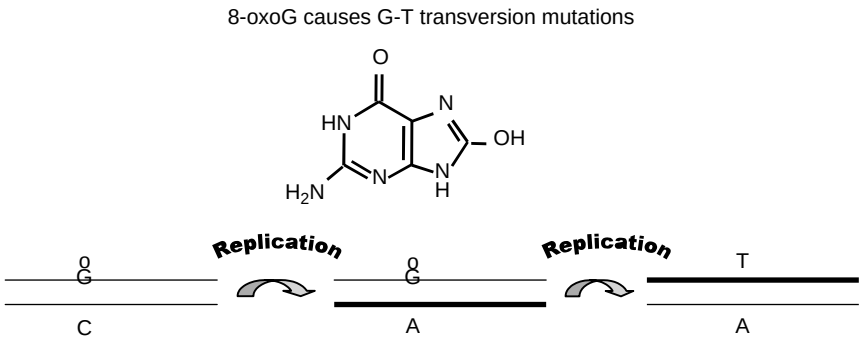

Fig. 4 8-oxoG lesions in DNA can cause the formation of lasting changes (mutations) in DNA.

Detection of oxidative DNA damage

- HPLC w. EC
  – 8 oxoguanosine
- Enzymatic
  – OGG1 and fpg enzymes cleave at certain base lesions
- Postlabeling
  – Some lesions are unidentified
- Gas Chromatography with MS
  – Detects many lesions, require much DNA
- LC/MSMS
  – More sensitive

Fig. 5 Some methods used to detect oxidative DNA base lesions. HPLC w. EC, high-pressure liquid chromatography with electrochemical detection. OGG1 (oxoguanine glycosylase) and fpg (fapy glycosylase) are glycosylases that detect base lesions in DNA. MS, mass spectrometry; LC/MSMS, liquid chromatography with mass spectrometry.

chromatography (HPLC) has been a useful way to detect the lesion 8-oxoG with a high level of sensitivity. More recently, mass spectrometry combined with liquid chromatography (LC/MS) has been used to detect this lesion with even higher sensitivity. Oxidative lesions other than 8-oxoG can be detected with gas chromatography or enzymatic probing, Fig. 5.

1.4 Consequences of DNA Damage

Figure 6 shows some of the cellular consequences of DNA damage. As mentioned above, DNA damage can be induced by external and internal sources. UV irradiation or ionizing irradiations are examples of exogenous sources of stress. Reactive oxygen species generated by the oxidative phosphorylation that occurs in mitochondria and thus via the cellular metabolism are examples of endogenous stress. Mutations in DNA can occur via replication of the damaged DNA whereby they become "fixed." Lesion bypass or replication errors (as illustrated in Fig 4) can give rise to other forms of genomic instability, both microscopically and macroscopically. A lesion in DNA can block transcription

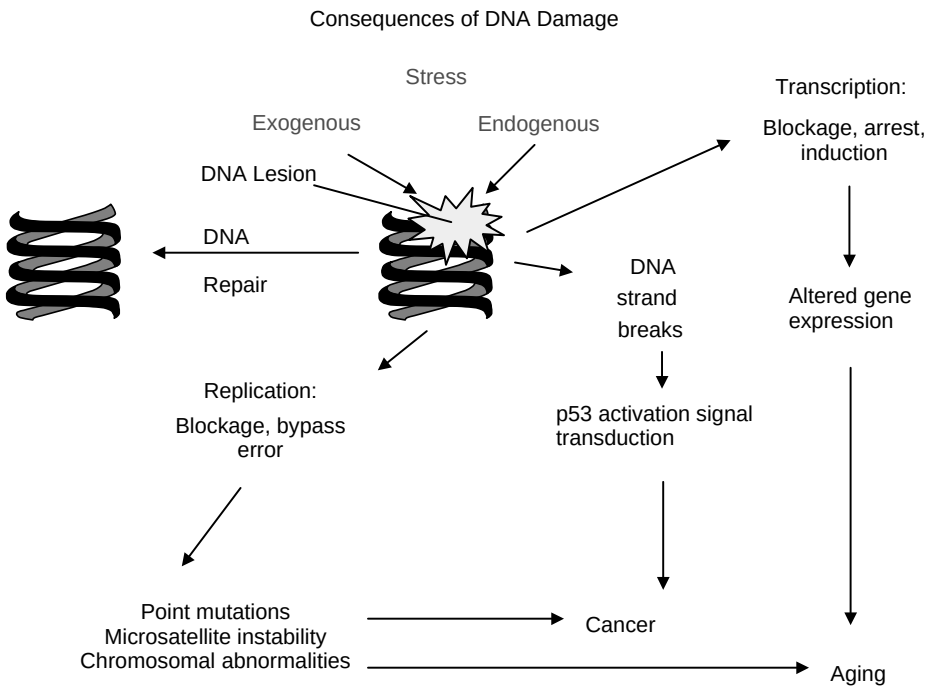

Fig. 6 Some consequences of DNA damage. Lesions in DNA can instigate a number of DNA related transactions, discussed in the text, that ultimately lead to genomic instability and other features of cancer and aging.

(conversion of DNA to RNA). It may truncate the transcript, or it may cause errors in the transcription. Alternatively, the DNA damage may induce new transcripts, and a number of genes have been shown to be inducible after various forms of cellular stress. These changes in transcription patterns that are caused by DNA damage may be part of the etiology of the malignant phenotype, as many changes in transcription have been reported in cancers. The modification of transcription caused by DNA damage may also underlie many of the changes associated with aging, where reductions, or in some cases increases, in transcriptional activity are well established. Lesions in DNA can also lead to the activation of the p53 tumor suppressor pathway leading to cell cycle arrest or apoptosis, or they can cause direct strand breaks in DNA, which again appear to be signals for several important enzymes and pathways.

It is estimated that there are several thousand DNA alterations in each cell in our organism per day caused by the endogenous and exogenous stresses, and were it not for an efficient DNA repair process, our genetic material would be destroyed by these processes over a normal human lifetime.

1.5 DNA Repair

Mammalian cells can make use of a variety of DNA repair pathways, and a schematic overview of these is listed in Fig. 7. Most studies over the years have

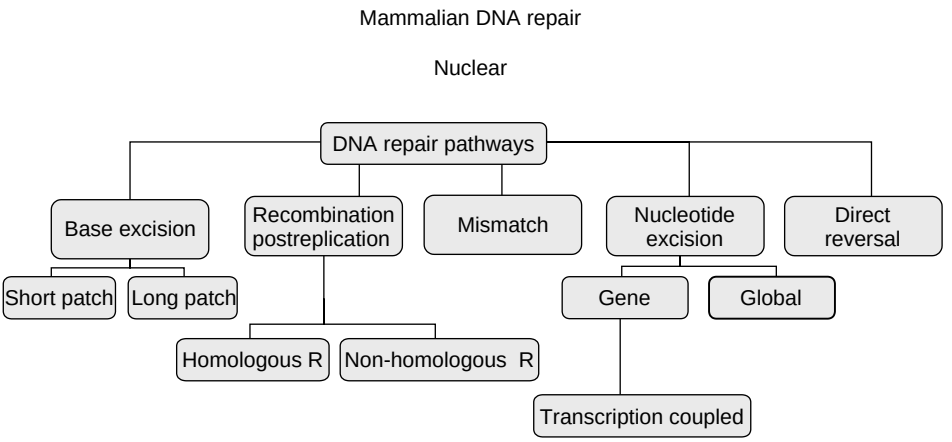

Fig. 7 Some of the major pathways of DNA repair in mammalian systems such as humans. These pathways can operate independently but are also interactive.

been based on the assumption that the DNA repair pathways listed were confined to the removal of specific lesions within certain categories. This is illustrated in Fig 2. For example, NER is the system that removes bulky lesions in DNA, such as pyrimidine dimers, cisplatin adducts, and other lesions that dramatically change the DNA structure. In contrast, base excision repair (BER) is the process responsible for the removal of simple lesions in DNA, exemplified by alkylation lesions such as methyl guanines that are considered to afflict only small structural changes in DNA and which may not represent major blocks to transcription and replication. This concept has changed somewhat in recent years as it has become evident that many of the DNA repair pathways listed in Figs. 2 and 7 are overlapping and share protein components. Thus, although it is useful to think of repair pathways as confined to the removal of different categories of DNA adducts, that distinction is becoming less certain.

It would be much too ambitious to review all the processes listed in Fig. 7. Two of the most predominant pathways are NER and BER, and these will be discussed in more detail. Mismatch repair (MMR) can protect against certain adducts in DNA and errors in DNA replication. Defects in the MMR system are closely associated with certain types of human cancers, and this pathway will also be briefly discussed. Recombinational DNA repair is a major pathway that removes certain complex lesions in DNA, for example DNA interstrand cross-links, shown in Fig. 2. This pathway of DNA repair and also direct reversal are not further described in the following. Examples of direct reversal of damage are the functions of the repair enzymes photolyase (repairs UV-induced pyrimidine dimmers) and methyl guanine methyl transferase (removes certain alkylation damage in DNA). The direct reversal processes are one-step repair that occur under certain, limited circumstances.

In general terms, the DNA repair process consists of a number of steps that act in concert to accomplish the complete repair of a DNA damage. This is shown in Fig. 8. The first step is the *recognition* of the DNA lesion, and this is carried

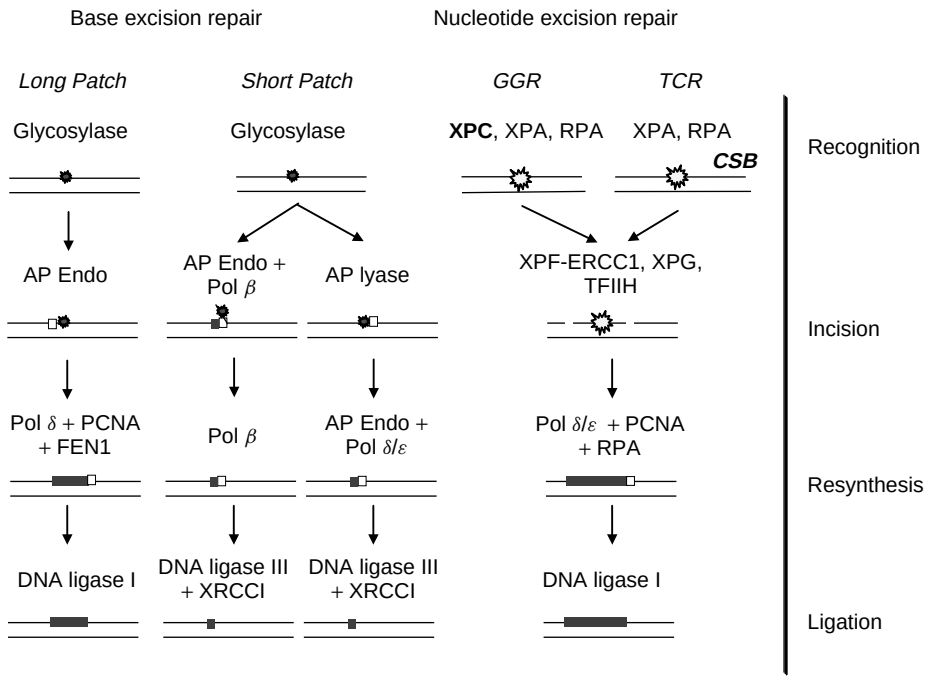

Fig. 8 The general steps in the major DNA repair pathways, and some of the proteins and steps involved in base excision repair (BER) and nucleotide excision repair (NER).

out by proteins that constantly survey the DNA for any unwanted modifications. The next step is called *incision,* and here enzymes cut into the DNA to remove the damaged DNA base. This step is complex and involves many proteins. When the damaged base has been removed, there is a step of *resynthesis,* where new DNA is generated in order to replace that which was removed in the incision step. This is accomplished by proteins called DNA polymerases, and there are several kinds of these in each mammalian cell. Since the DNA damage was in one strand of DNA, the other strand has the information required to copy itself. After the DNA synthesis, there is a *ligation* process in which the gaps in the DNA are sealed and a new intact double helix is formed.

1.1.1 Base Excision Repair (BER)

Base excision repair of oxidative DNA damage is initiated by DNA glycosylases, a class of enzymes that recognize and remove damaged bases from DNA by hydrolytic cleavage of the base-sugar bond leaving an abasic site (apurinic or apyrimidinic (AP site)). The process is shown in Fig. 8, where some of the proteins involved are indicated. Several DNA glycosylases that recognize and process oxidation products of purines and pyrimidines have been identified in human cells. The human homolog of the bacterial endonuclease III (hNTH1) glycosylase has been cloned and purified to homogeneity. Two human glycosylase genes, hOGG1 and hOGG2, have been cloned. The gene products recognize and process 8-oxoG in DNA. The glycosylases

primarily cut the glycosylic bond between the DNA base and the sugar, but many of these proteins also have an associated AP lyase activity, which then cleaves the DNA backbone and creates a nick in the DNA. There are also special enzymes, AP endonucleases, which cleave the DNA at abasic sites; the specifics of the cleavage depends upon the exact nature of the abasic site in the DNA. This results in a single-nucleotide gap that is then filled by a DNA polymerase using the intact DNA strand as a template, and the replaced gap is then sealed by DNA ligase. In mammalian cells, DNA polymerase β (pol β) is the major DNA polymerase involved in the single-nucleotide excision patch repair pathway. It was recently demonstrated that DNA pol β has an intrinsic AP lyase activity (cuts the DNA backbone like the AP endonuclease) that removes 5′-sugar-phosphate by β-elimination, and then pol β fills the single-nucleotide gap.

An alternative pathway, long patch BER, has been reported. In addition to a DNA glycosylase and AP endonuclease, this pathway involves the following proteins: flap endonuclease (FEN1), proliferating cell nuclear antigen (PCNA), DNA polymerase δ, and DNA ligase. Neither of these enzymes can remove a 5′ sugar phosphate and generate a one nucleotide gap. Instead, DNA polymerase first adds several nucleotides to the 3′ end of the nick and exposes the 5′ sugar phosphate as part of a single-stranded flap structure. This flap structure is recognized and excised by FEN1 and the DNA is finally ligated by DNA ligase. This process, shown in Fig. 8, is called long patch BER, and this type of DNA repair results in a 2 to 7 nucleotide – long newly incorporated DNA repair patch. Currently, there is much research activity in the BER area.

1.1.2 Nucleotide Excision Repair (NER)

Most of the understanding of this pathway has come through the study of the human disorder xeroderma pigmentosum (XP). There are seven genetically different types of this disease, designated A to G. XP proteins are designated according to the cell line in which they are mutated, for example, the XPA protein is the one that is mutated in XPA cells. The individuals afflicted with this condition suffer from high incidences of skin and internal cancers, hyper pigmentation, and premature aging. Cells from XP patients cannot incise their DNA at a site of a UV-induced pyrimidine dimmer photoproduct and are thus deficient in the incision process of NER. The enzymatic steps involved in NER are recognition of the lesion, incision of the DNA, excision of the damaged DNA template, resynthesis of new DNA based on the intact template, and ligation of the newly formed DNA repair patch into the reconstructed double helix. These steps and some of the proteins involved are listed in Fig. 8. The recognition step involves the XPC and XPA proteins as well as replication protein A (RPA). The incision involves the structure-specific endonucleases XPF-ERRC1 and XPG as well as the basic transcription factor TFIIH that contains the repair proteins XPD and XPB among its nine components. Interestingly, this represents a direct link between the DNA repair process and transcription since the TFIIH protein complex participates in both processes. The DNA incision process is a combination of DNA repair and transcription and illustrates the tight linkage between these processes. After incision, there is excision of the DNA strand containing the lesions, and then there is resynthesis performed by DNA polymerases followed by ligation, which

is performed by ligases. A number of recent reviews have discussed NER in much more detail.

1.1.3 Genomic Heterogeneity of NER

A major advance in the study of DNA repair has been the insight that NER and possibly other repair pathways operate with considerable heterogeneity over the mammalian genome. The NER pathway can thus be subdivided into different pathways relating to the functional and structural organization of the genome. Only about 1% of the genome is transcriptionally active, and a repair pathway entitled transcription-coupled repair (TCR) operates here. The remainder of the genome, the 99% that is inactive, has a separate pathway, general genome repair (GGR). Much work has been dedicated to the further delineation and clarification of these pathways. In the human premature aging syndrome, Cockayne syndrome (CS), TCR is deficient. CS is a rare and severe clinical condition in which the patients appear much older than their chronological age. It is a human premature aging disorder since many of the clinical signs and symptoms are those that we see in aging process in normal individuals.

1.6 Mismatch Repair

Unrepaired and misrepaired bases in DNA can arise by replication errors, spontaneous or induced base modifications, and during recombination. As for most of the other DNA repair pathways, the major understanding of this function comes from work in bacteria. Here, a process exists that recognizes the level of methylation in DNA to direct itself to the correct strand that needs repair. The proteins involved are called MutS, MutH, and MutL, and a hairpin structure is formed in DNA, which is then resolved by a DNA helicase that unwinds this structure. After that process, single-stranded DNA ends are degraded by an exonuclease, and then there is resynthesis of the DNA strand followed by ligation. Mispairings frequently transpire during DNA replication because of the limited fidelity of the DNA replicative machinery, and an incorrect base is incorporated. In human cells, a number of mismatches and aberrant DNA structures are recognized by a protein complex called MutSα, a dimer formation of the two proteins, MSH2 and MSH6. Another so-called heterodimer is the protein complex MutSβ, which also recognizes aberrant DNA structures and certain mismatches between DNA bases. The precise process as to how this pathway operates in higher organisms is not yet understood. Importantly, defects in the MMR pathway are associated with colon cancer and also with other types of cancer.

1.7 Mitochondrial DNA Repair in Mammalian Cells

Mitochondria are the energy stations inside the cells. Here, oxidative phosphorylation occurs, generating ATP. In this process, ROS are formed at high frequencies, and the mitochondrial DNA (mtDNA) is directly exposed. The mitochondrial DNA does not have a recognized chromatin structure and is thus particularly exposed to oxidative DNA base lesion formation.

There are about 1000 mitochondria per mammalian cell. Each mitochondrion has 4–5 DNA plasmids, each about 16-kb long. This means that about 2% of our total DNA is in the mitochondria. All mtDNA is transcribed whereas only about 1% of the nuclear DNA is transcribed, and thus, the mtDNA makes up a large

fraction of the total transcribed DNA in a mammalian cell.

MtDNA does not code for any DNA damage processing enzymes. It only carries the genes coding for the oxidative phosphorylation proteins and some tRNA's. Thus, all repair enzymes functioning in the mitochondria need to be transported into them. It was shown more than 20 years ago that mitochondria do not repair UV-induced pyrimidine dimers, and this observation provided the basis for the notion that there is no DNA repair capacity in mtDNA. Since then, there have been many observations indicating the presence of BER in mitochondria. BER enzymes have been identified and characterized and studies have shown that a number of simple monofunctional alkylating agents and oxidative DNA base lesions are efficiently removed from mtDNA. Other repair processes have also been detected. An important question remaining is whether mitochondria possess any capabilities to repair bulky lesions via the NER pathway.

Knowledge about mitochondrial DNA repair is limited since it has been very difficult to study this type of DNA repair. Experimental techniques and methods to study mtDNA repair have not been nearly as well developed as those for the study of nuclear DNA repair. Whereas *in vitro* DNA repair studies have been performed with great success in nuclear or whole cell extracts from cells, this kind of biochemical approach has only very recently become available for mtDNA. Recent advances suggest that mtDNA repair can now be studied using more sophisticated biochemical analysis, and this should provide great advances in the near future.

Oxidative phosphorylation in mitochondria results in the production of ROS. Several other processes contribute significantly to the pool of ROS. They include peroxisomal metabolism, enzymatic synthesis of nitric oxide, metabolism of phagocytic leukocytes, heat, ultraviolet light, various drugs such as those used in the treatment of HIV, ionizing radiation, and redox cycling compounds. Hydrogen peroxide, singlet oxygen, and hydroxyl radicals are among the ROS produced. The interactions between ROS and mtDNA result in oxidation of specific mtDNA bases. The most common oxidation products of mtDNA include 8-oxoG and thymine glycols (TG), and these base modifications have been detected in human cells. In addition, deamination of cytosine and incorporation of dUTP results in the formation of uracil in mtDNA. If unrepaired, this uracil can cause GC-AT transition mutations. Thus, insufficient mtDNA repair may result in mitochondrial dysfunction and thereby cause degenerative diseases, loss of energy formation, and pathophysiological processes leading to aging and cancer.

1.8 Age Related Changes in DNA Repair

There has been a great deal of interest in the question of whether DNA repair declines with age. Investigators have looked at many conditions of premature and normal aging and many different types of assays have been used. In general, there seems to be a growing consensus that DNA repair declines slightly with age. One study reported a 1% decline in DNA repair capacity per year with advancing age in individuals. Attempts to correlate DNA repair capacity of different organisms with maximum life span have been made. Hart and Setlow demonstrated a linear correlation between the logarithm of life span and the DNA repair capacity in

cells from different mammalian species, suggesting that higher DNA repair activity is associated with longer life span.

1.9 Diseases with Defective DNA Repair

Several human diseases are associated with defects in DNA repair. This clearly establishes DNA repair as an important function for avoiding human diseases and it suggests that therapies that enhance DNA repair may prevent disease, and thus there is great opportunity in the possibilities that lie in modulating DNA repair processes in humans. As mentioned earlier, the diseases of xeroderma pigmentosum are associated with a distinct defect in the DNA repair pathway of NER. The patients suffering from these conditions have high incidences of cancers, illustrating a link between defects in DNA repair and cancer. Another group of disorders that have defects in DNA repair are characterized by premature aging, that is, that the patients appear much older than their chronological age. These diseases include Cockayne syndrome, Werner syndrome, and other forms of progeria. Some diseases are characterized by a preponderance of genomic instability, likely also linked to DNA repair defects. These include Bloom syndrome, ataxia telangiectasia, and Nijmegen breakage syndrome. All of the above conditions are rare and this strengthens the notion that DNA repair is an essential process and if severely defective, life cannot be sustained.

1.10 Perspectives

DNA repair is a complex and fascinating process and it is very interactive with other DNA metabolic processes. It is tightly linked to the transcription process via the basal transcription factor TFIIH, which participates in both processes. It is also tightly linked with DNA replication and with a number of signal transduction pathways. An age-associated decline in DNA repair could explain why older individuals suffer from age-associated diseases and become highly susceptible to cancer. Increased insight into the mechanisms of DNA repair may also lead to therapeutic interventions in the future, for example by enhancing the DNA repair activity.

See also Genomic Responses to Oxidative Stress; Ionizing Radiation Damage to DNA.

Bibliography

Books and Reviews

Friedberg, E.C., Walker, G.C., Siede, W. (1995) *DNA Repair and Mutagenesis*, ASM Press, Washington, D.C.

Kraemer, K.H. (1999) Heritable Diseases with Increased Sensitivity to Cellular Injury, in: Freedberg, I.M., Eisen, A.Z., Wolff, K., Austen, K.F., Goldsmith, L.A., Katz, S.I., Fitzpatrick, T.B. (Eds.) *Dermatology in General Medicine*, McGraw Hill, New York, pp. 1848–1862.

Kraemer, K.H., Moriwaki, S., Tarone, R.E., Khan, S.G., Grossman, L. (1999) DNA Repair and Aging: Studies of Xeroderma Pigmentosum and Normal Cells, in: Bohr, V.A., Clark, B.F., Stevnsner, T. (Eds.) *Molecular Biology of Aging, Alfred Benzon Symposium 44*, Munksgaard, Copenhagen, Denmark, pp. 260–265.

Primary Literature

Balajee, A.S., Bohr, V.A. (2000) Genomic heterogeneity of nucleotide excision repair, *Gene* **250**, 15–30.

Bohr, V.A. (2001) Mitochondrial DNA repair, *Prog. Nucleic Acid Res. Mol. Biol.* **68**, 255–256.

Bohr, V.A. (2002) Repair of oxidative DNA damage in nuclear and mitochondrial DNA, and some changes with aging in mammalian cells(1,2), *Free Radical Biol. Med.* **32**, 804–812.

Bohr, V.A., Anson, R.M. (1995) DNA damage, mutation and fine structure DNA repair in aging, *Mutat. Res.* **338**, 25–34.

Buermeyer, A.B., Deschenes, S.M., Baker, S.M., Liskay, R.M. (1999) Mammalian DNA mismatch repair, *Annu. Rev. Genet.* **33**, 533–564.

Dizdaroglu, M., Jaruga, P., Birincioglu, M., Rodriguez, H. (2002) Free radical-induced damage to DNA: mechanisms and measurement, *Free Radical Biol. Med.* **32**, 1102–1115.

Fleck, O., Lehmann, E., Schar, P., Kohli, J. (1999) Involvement of nucleotide-excision repair in msh2 pms1-independent mismatch repair, *Nat. Genet.* **21**, 314–317.

Friedberg, E.C. (2001) How nucleotide excision repair protects against cancer, *Nat. Rev. Cancer* **1**, 22–33.

Hart, R.W., Setlow, R.B. (1974) Correlation between deoxyribonucleic acid excision repair and life span in a number of mammalian species, *Proc. Natl. Acad. Sci. USA* **71**, 2169–2173.

Hoeijmakers, J.H. (2001) Genome maintenance mechanisms for preventing cancer, *Nature* **411**, 366–374.

Jiricny, J., Nystrom-Lahti, M. (2000) Mismatch repair defects in cancer, *Curr. Opin. Genet. Dev.* **10**, 157–161.

Lindahl, T., Wood, R.D. (1999) Quality control by DNA repair, *Science* **286**, 1897–1905.

Nilsen, H., Krokan, H.E. (2001) Base excision repair in a network of defence and tolerance, *Carcinogenesis* **22**, 987–998.

Olinski, R., Zastawny, T., Budzbon, J., Skokowski, J., Zegarski, W., Dizdaroglu, M. (1992) DNA base modifications in chromatin of human cancerous tissues, *FEBS Lett.* **309**, 193–198.

Weindruch, R., Kayo, T., Lee, C.K., Prolla, T.A. (2002) Gene expression profiling of aging using DNA microarrays, *Mech. Ageing Dev.* **123**, 177–193.

Wilson, S.H., Kunkel, T.A. (2000) Passing the baton in base excision repair, *Nat. Struct. Biol.* **7**, 176–178.

Nucleic Acids (DNA) Sequencing, Transcriptional

Kazuhiro Shibata[1], *Masaki Izawa*[2], *and Yoshihide Hayashizaki*[1,3,4]
[1] *RIKEN Discovery and Research Institute, Wako, Japan*
[2] *Nippon Gene Tech. Co, Ltd., Arakawa, Japan*
[3] *RIKEN Genomic Sciences Center, Yokohama, Japan*
[4] *Institute of Basic Medical Sciences, University of Tsukuba, Japan*

1 Introduction 397

2 Components of Transcriptional Sequencing 398
2.1 Principle 398
2.2 Enzymes 399
2.2.1 Enzymological Strategies for Development of Novel Mutant Polymerases for Sequencing 399
2.2.2 Promoter Recognition and Binding of T7 RNA Polymerase to DNA Template 401
2.2.3 Initiation 401
2.2.4 Termination 401
2.2.5 Slippage 402
2.3 Composition and Methods for DNA Sequencing by TS 402
2.3.1 Mutant T7 RNA Polymerases 402
2.3.2 Polyamine 402
2.3.3 Pyrophosphatase 402
2.3.4 GMP as an Initiator and Modified NTPs as Substrates 402
2.3.5 DNA Templates 403

3 Remarkable Properties of Transcriptional Sequencing 404
3.1 Isothermal and Rapid Reaction 404
3.2 Wide Dynamic Range 404
3.3 Readability of GC-rich Templates 404

4 Detection of Transcriptional Sequencing Products 405
4.1 BODIPY-labeled Terminators for Auto Sequencers 406

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

4.2 Rhodamine-labeled Terminators for DNA Sequencing 406
4.3 3′-dNTP Terminators for Mass Spectroscopy 406

5 Standard Protocol for Template Preparation, Reaction, and Electrophoresis 407
5.1 Template Preparation, Reaction, and Purification 407
5.2 Direct Sequencing without Purification 408
5.3 Electrophoresis 408

6 Future Applications of TS 408
6.1 Modification of Substrate 408
6.2 Modification of Dye 408
6.3 Chemical Modification of Substrate for Mass Spectroscopy 409
6.4 Lab on a Chip 409

Acknowledgments 411

Bibliography 411
Books and Reviews 411
Primary Literature 411

Keywords

RNAP Mutant (in TS)
RNA polymerase mutants that rNTPs and 3′-dNTPs incorporate uniformly and efficiently.

MALDI-TOF-MS Sequencing
Mass spectroscopy of chain-terminated nucleotide molecules with MALDI-TOF-MS.

TS
Transcriptional Sequencing.

Transcriptional Sequencing (TS) is a novel DNA sequencing method based on the chain termination method using RNA Polymerase (RNAP). Transcriptional sequencing differs considerably from conventional cycle sequencing methods using DNA Polymerases (DNAPs) with 2′-deoxyribonucleotide triphosphates (2′-dNTPs) and 2′, 3′-dideoxyribonucleotide triphosphates (2′,3′-ddNTPs). The TS method employs a modified T7, T3, or SP6 RNAP with ribonucleotide triphosphates (rNTPs) and 3′-deoxyribonucleotide triphosphates (3′-dNTPs) as substrates and terminators, respectively, and the T7, T3, or SP6 promoter sequence is used as initiation site instead of the primer sequence. The characteristics of RNAP allow TS to be conducted under conditions that are substantially different from conventional cycle sequencing using DNAP. In TS, the reaction is performed isothermally (at 37 °C) and is stable

with high processivity. Transcriptional sequencing can be applied to various amounts of double-stranded DNA template because RNAP produces a large number of transcribed RNA molecules and the quantity of product is limited by the quantity of substrate rather than that of the template. Furthermore, in direct sequencing using TS, PCR products do not need to be purified before the sequencing reaction because RNAP requires neither 2′-dNTPs nor single-stranded PCR primers, but polymerization is primed using specific double-stranded promoters and rNTPs. This characteristic feature ensures that even highly GC-rich DNA templates can be sequenced effectively. Transcriptional sequencing is also applicable to MALDI-TOF-MS DNA sequencing owing to the stability of the RNA polymer against fragmentation caused by laser ablation. Thus, TS differs considerably from conventional cycle sequencing methods using thermostable DNA polymerases, but should complement such methods in numerous practical applications.

1 Introduction

There are two major families of enzymes responsible for polymerization of nucleic acids *in vivo*; DNA polymerases (DNAPs), which are responsible for DNA replication, and RNA polymerases (RNAPs), which are responsible for transcription. DNA polymerases produce double-stranded DNA (dsDNA) using a single-stranded DNA (ssDNA) template and consume deoxyribonucleotide triphosphates (dNTPs) as substrate. On the other hand, RNAPs produce RNA strands from a dsDNA template and consume ribonucleotide triphosphates (rNTPs) as substrate. Most modern sequencing kits utilize the Sanger method with DNAP and 2′,3′-dideoxyribonucleotide triphosphate (2′,3′-ddNTP) terminators bound to a dye, and in conjunction with development of DNAP mutants, such as mutated *Taq* DNAP (Thermosequenase, FS *Taq* etc.), and energy transfer dye, sequencing accuracy has been improved substantially and large amounts of data have been produced.

In principle, RNAPs can be also applied to the chain termination method, using 3′-deoxyribonucleotide triphosphates (3′-dNTPs) as terminators. Axelrod V.D. and Kramer F.R. reported the first trial of the chain termination method using T7 RNAP in 1985, but the technique was not developed further. On the other hand, cycle sequencing based on *Taq* DNAP has been widely used for sequencing kits. The reasons why sequencing with RNA and RNAP is not widely used are likely related to concerns regarding the instability and the strong secondary structure of RNA chain termination molecules, which would interfere with sequencing reaction and electrophoresis. However, in reality, both the reaction and subsequent electrophoresis are relatively simple to perform. Furthermore, transcriptional sequencing (TS) has numerous advantages over conventional cycle sequencing owing to the characteristic features of RNAP, such as the ability to sequence GC-rich regions that usually appear as sequencing gaps with DNAP-based sequencing kits, applicability to MALDI-TOF-MS (Matrix-assisted laser desorption/ionization time-of-flight

mass spectrometry) mass spectrometry for sequencing, which gives better stability of chain termination products than DNAP-based sequencing, and because there is no need for either large amounts of template molecules or purification of templates for direct sequencing.

Here, we describe the history, development, and characteristic features of "Transcriptional Sequencing" based on chain termination methods using modified RNAP.

2 Components of Transcriptional Sequencing

2.1 Principle

Transcriptional sequencing is a DNA sequencing method that uses RNAP. This method is based on the chain termination method, and uses rNTPs as substrate and 3′-dNTPs as terminators. Figure 1 shows a schematic comparison of conventional cycle sequencing and TS. With cycle sequencing, programmed repetitive changes in the reaction temperature are employed. One set of programmed temperature changes is called a *cycle*, and there are typically three steps in one cycle; annealing, elongation, and denaturation. DNA oligo-primers are allowed to anneal during the first step, DNAPs then incorporate 2′-deoxyribonucleotide triphosphates (2′-dNTPs) to elongate the primers until 2′, 3′-dNTPs are incorporated or until the reaction terminates spontaneously, and finally, the Sanger reaction products and primers are dissociated from template during the denaturation step. In contrast, there is only one step for the TS method; transcription (at 37 °C). Transcription is initiated from the promoter (T7, T3, or SP6) on dsDNA and transcripts elongate until RNAPs incorporate 3′-dNTPs or until the reaction terminates spontaneously. Reaction products are released

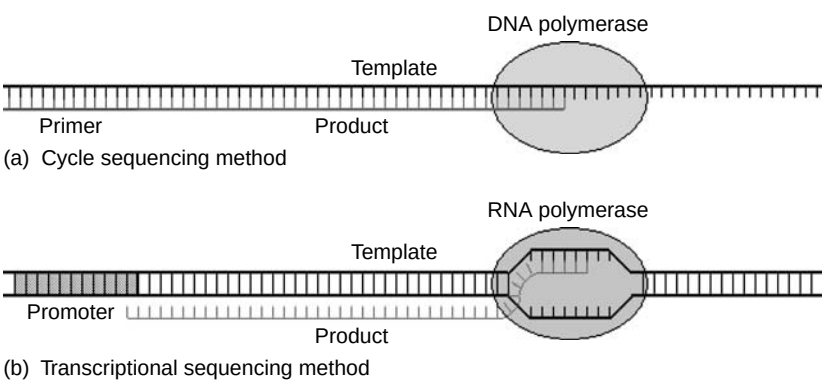

Fig. 1 Schematic comparison of TS and cycle sequencing methods. (a) Cycle sequencing method. Black line indicates template (ssDNA). Blue line indicates primer, which is elongated by DNAP to give the reaction product (red line). Elongation of product terminates either spontaneously or when DNAP incorporates a terminator (2′, 3′-deoxyribonucleotide triphosphate). (b) TS method. Black line indicates template (dsDNA). Gray zone indicates promoter region, which is the starting point of the transcript (red line). Elongation of product terminates either spontaneously or when RNAP incorporates a terminator (3′-deoxyribonucleic triphosphate). (See color plate p. xxiv).

when RNAP separates from the template. RNA molecules in the reaction mixture are separated by length and are detected by fluorescence or mass spectrometry.

2.2 Enzymes

Because of the low efficiency and nonuniform incorporation of substrates used by wild-type RNAP, mutants were designed in order to achieve accurate DNA sequences. We designed such mutants by carefully referring to the 3D structure of DNAP.

2.2.1 Enzymological Strategies for Development of Novel Mutant Polymerases for Sequencing

Cycle sequencing systems based on *Taq* DNAP have been developed in order to improve the readable length and accuracy of sequencing. In developing a mutant enzyme, information regarding the 3-D structure of DNAP is very useful. In particular, the common structure for DNAP indicates the amino acid residues that could be substituted to confer desirable properties on the polymerase. For cycle sequencing, mutant *Taq* DNAPs, such as Thermosequenase, FS *Taq* DNAP, and Therminator DNAP, have been developed to allow incorporation of fluorescence-labeled 2′,3′-ddNTP substrates efficiently and uniformly.

RNA Polymerases play a fundamental role in the transcription process in bacteriophages. In particular, T3, T7, and SP6 phages synthesize phage-encoded RNAPs after infecting the host cell. These RNAPs recognize their own phage-specific promoter sequences that are not encoded in the host genome sequence. The specificity and high processivity of these polymerases have led to the development of plasmid vectors that allow transcription of cloned DNA fragments downstream of the phage promoter. In addition, phage RNAPs always initiate transcription at a single and fixed position within the promoter region, and thus, these transcripts have a common 5′ end. This property of RNAPs allows us to apply this type of enzyme to chain termination DNA sequencing. The first trial to produce sequence data using the chain termination reaction with T7 RNAP was reported in 1985 (Axelrod V.D. and Kramer F.R.). The pattern obtained with T7 RNAP using internal radio-labeled nucleotide was quite similar to the analogous ladder pattern with various intensity bands using Klenow fragments. However, the intensity of the T7 RNAP ladders was much less uniform than that of the T7 DNAP. This property is undesirable for accurate sequencing and is clearly due to the strong preference of RNAP to incorporate natural nucleotides and terminators. Development of the T7 RNAP mutant that incorporates natural ribonucleotide substrates and terminators in an equal ratio was essential in order to establish TS for practical use.

For the above reasons, we planned a series of mutations to develop a T7 RNAP mutant. According to mutant characterization and three-dimensional structure analysis, the catalytic domain of T7 RNAP is separated into the T7 promoter binding region (674–752 aa), Mg^{2+} binding sites (Asp537, Asp812), and essential regions for polymerase activity (motif A, motif B, motif C). When we began developing the mutant RNAP for TS, two mutants were found to recognize the 2′ or 3′ hydroxy groups in nucleotide substrates of RNAP and DNAP (*Taq* polymerase), respectively. One of these is T7 RNAP Y639F. Tyr639 is located in the active site and has been reported to be the discrimination site for 2′-OH groups in T7 RNAP. The Y639F mutant retains both DNAP and RNAP activities and is unable to discriminate

between rNTPs and 2′-dNTPs. This mutant analysis suggests that the hydroxyl group in the tyrosine residue binds the 2′-site of dNTPs (hydrogen atom) within the pocket of RNAP in order to incorporate the 2′-dNTP substrates. Therefore, to develop a novel mutant that incorporates the 3′-dNTP, the amino acid that recognizes the 3′-OH of rNTPs should be mutated to a hydroxyl group. Mutation from phenylalanine to tyrosine or vice versa improves affinity for the 2′-OH group of 2′-dNTPs in the Y639F mutant and for the 3′-OH group of 2′,3′-ddNTPs in *Taq* polymerase or Thermosequenase, as described in the reverse compensation theory of 2′- and 3′-OH groups proposed by Tabor and Richardson.

The other mutant is T7 RNAP M635A. Originally, Met635 of T7 RNAP was thought to be the interaction site for the 3′-OH group, suggested both by homology analysis with Klenow fragments and by mutation analysis, where affinity for rNTPs decreased when Met635 was substituted with Arg635. However, there has been no direct proof that the M635A recognizes the 3′-OH group any more efficiently than Y639F. Thus, an RNAP mutant that incorporates 3′-dNTPs efficiently has not yet been reported, and we subsequently began developing an RNAP mutant for TS, designated transcriptional sequenase (TSase).

On the basis of the above information, we initially looked for a phenylalanine residue located in the nucleotide-binding pocket. To design the TSase mutant, we paid particular attention to the 3D structure of T7 RNAP in order to identify the catalytic domain for nucleotide binding.

In the catalytic domain, we identified Phe644, Phe646, Phe667, Phe782, Phe773, and Phe882 and we conducted site-directed mutagenesis in order to change Phe644, Phe646, Phe667, Phe782, and Phe773 to Tyr. All mutant RNAPs were expressed in *Escherichia coli*, and purified, and their catalytic activities for incorporation of 3′-dNTP were analyzed. The 3′-dNTP incorporation activity of F644Y and F677Y were 100-fold greater than that of wild-type polymerase. Figure 2 shows the sequence alignment of the mutation sites of T7 RNAP mutants in comparison with *E.coli* polymerase, *Taq*, and T7 DNAP. The pocket for nucleotide substrates of RNAP is formed

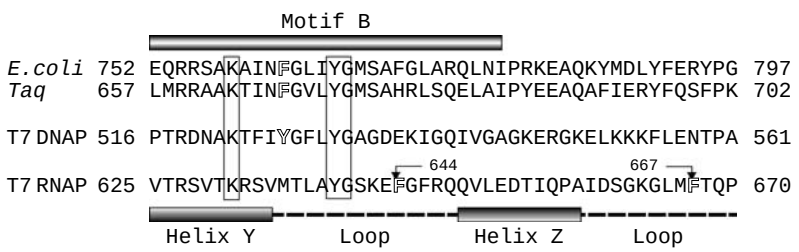

Fig. 2 Mutation sites of the T7 RNAP mutant at regions including motif B, and a comparison with several DNAPs. Sequences were taken from *E. coli* DNA polymerase I, *Taq* DNAP, T7 DNAP, and T7 RNAP. Strictly conserved positions in all polymerases are boxed, and the residues responsible for discriminating between 2′- and 3′-OH groups are indicated by white letters. The upper bar indicates motif B as one of the highly conserved regions, the lower bars and the dotted line indicate the helix and loop structure of T7 RNAP. The T7 RNAP mutants shown in this figure were designated F644Y and F667Y.

by the HelixY–loop–HelixZ–loop region. Phe644 and Phe667, which access the 3′-OH of nucleotide substrates in the polymerase domain, are located in this region.

We have designed T3 RNA polymerase mutants according to the above principles. In T3 RNA polymerase, Phe645 and Phe668 are located in this HelixY–loop–HelixZ–loop region, and the mutants F645Y and F668Y were found to incorporate 3′-dNTP much more rapidly than the wild-type polymerase.

Following the construction of the above mutants, Nippon Gene Tech. Co, Ltd. developed newly mutated RNAPs in order to improve the fidelity and efficiency of transcription. The features of the newly constructed mutants resulted in high S/N signals and were suitable for capillary-based DNA sequencers.

2.2.2 Promoter Recognition and Binding of T7 RNA Polymerase to DNA Template

Promoter sequences for T7 RNAP and related phage RNAPs consist of a 23-bp consensus sequence that extends from −17 to +6.

	−17 −10 −5 +1 +5
anti-sense strand	TAATACGACTCACTATAGGGAA
sense strand	ATTATGCTGAGTGATATCCCTT

This promoter sequence is similar to the truncated promoters for bacterial RNAP, having an upstream binding region of a TATA box element closer to the site of initiation. The phage promoter consists of two functional regions. One is the RNAP binding region located upstream from position −6 and the other is the initiation region for transcription located downstream from position −6. The roles of these two domains were demonstrated by experiments in which mutations at the binding domain decreased the affinity of the polymerase for the promoter, but had little effect on the initiation rate. Conversely, mutations in the initiation region had a significant effect on the kcat (Catalysis constant or turnover number) of the initiation rate. Mutations in this region vary the degree of abortive initiation, and thus, product yield decreases considerably. After polymerase binds to the promoter, both strands of DNA template are dissociated for the initiation of transcription.

2.2.3 Initiation

At the initiation of RNA synthesis by all RNAPs, an abortive transcription cycle usually proceeds before the elongation stage begins. RNAP can engage in multiple rounds of abortive initiation in which short RNA products are continuously synthesized and released. The frequency of this abortive initiation, the length of the RNA produced, and the efficiency with which the polymerase enters a stable elongation stage depends on numerous factors. During all stages of transcription, there exists a competition between incorporation of successive nucleotides (elongation) and dissociation of the RNAP–DNA–RNA complex.

2.2.4 Termination

The mechanism of T7 RNA polymerase termination has been less studied than the mechanism of transcription initiation. T7 RNAP does not terminate efficiently at the thr attenuator, which terminates *E. coli* polymerase with about 90% efficiency; T7 RNAP terminates at about 40% efficiency. Even at the T7-Tf terminator, termination efficiency is 66%. The most efficient

known terminator is the 7-bp sequence in the PTH (human prepro-parathyroid hormone) gene, which has an efficiency of above 90%.

2.2.5 Slippage

Transcriptional slippage, particularly in the TS procedure, is a phenomenon in which extra U residues are incorporated during the elongation. The slippage of TS frequently occurs in A or T homopolymeric regions, and particularly when there are continuous A or T regions of approximately 10 bp. Once slippage occurs, the sequencing pattern clearly becomes erroneous. The slippage described above is an intrinsic characteristic of wild-type RNAPs derived from the evolutional process. With regard to TS, such a characteristic is a disadvantage when sequencing the A or T homopolymeric sequences that preferentially appear in noncoding regions.

2.3 Composition and Methods for DNA Sequencing by TS

There are several available reagents that can be used to obtain accurate sequences with mutated RNAP. Here, we discuss those reagents briefly.

2.3.1 Mutant T7 RNA Polymerases

TS has a threshold for the enzymatic units of T7 RNAP necessary for numerous initiation cycles, some of which continue on to transcription. In order to increase the commercial viability of TS kits, the required amount of enzyme should be produced at low cost and at high purity. In this sense, the F644Y mutant is ideal when compared to F667Y because F644Y can be produced at high yield and with high purification efficiency.

2.3.2 Polyamine

In various *in vitro* enzyme reactions, polyamines such as spermine and spermidine have been widely used as counterions for nucleic acids. They have also been shown to stimulate RNA transcription by DNA-dependent RNAPs from phages. Therefore, we have attempted to optimize sequencing conditions by screening a variety of polyamines for increases in transcription rate, and have found that both spermidine and 1,8-diamino octane exhibited significant improvements in RNAP transcriptional activity.

2.3.3 Pyrophosphatase

Incorporation of each rNTP (or 2′-dNTP) into the polynucleotide chain by RNAP or DNAP releases an inorganic pyrophosphate (PPi) molecule. Inorganic pyrophosphate molecule degrades polynucleotide products by attacking the 3′-terminal internucleotide linkage. This phenomenon is called *pyrophosphorolysis* and it generates NTPs or dNTPs. This phenomenon is the reversal of polymerization. The enzyme inorganic pyrophosphatase (PPase) cleaves inhibitory pyrophosphate into monophosphate, which has less ''pyrophosphorolysis'' activity. Therefore, addition of PPase is necessary to ensure accurate sequence data, and it produces patterns with higher, sharper, and more uniform peaks (Fig. 3). The effects of PPase in improving the electropherogram can also be seen in chain termination using DNAP.

2.3.4 GMP as an Initiator and Modified NTPs as Substrates

Electropherograms are distorted by mobility shifts due to the secondary structure of chain termination products, particularly those of G–C stretches. To overcome this problem, we used the nucleotide analogs rITP or 7-deaza-GTP in place of

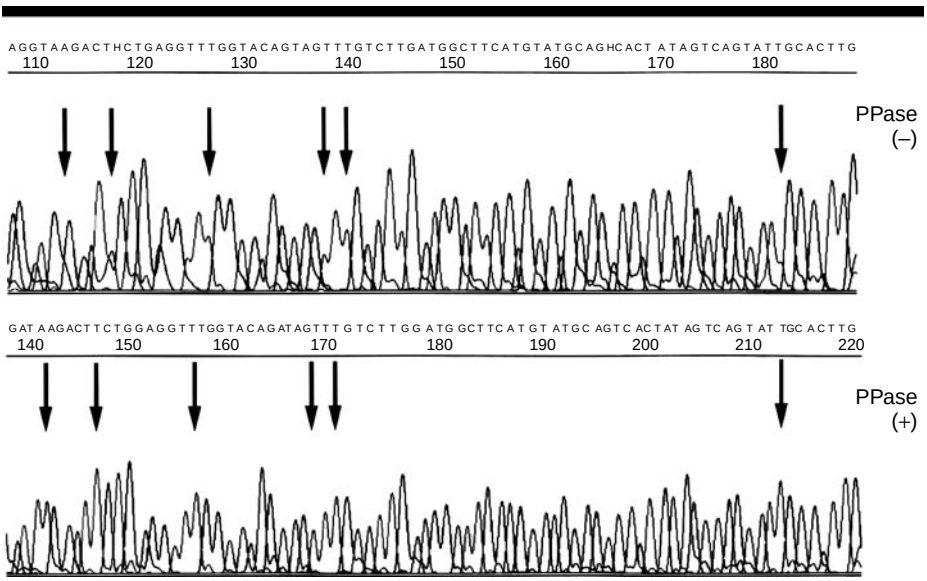

Fig. 3 Effect of PPase addition on transcriptional sequencing with F644Y T7 RNAP and four fluorescent-labeled terminators. Sequencing was carried out with unpurified human TSH-b cDNA PCR products as a template in the absence (–) or presence (+) of PPase. Arrows in both spectra are equivalent and indicate bases that are sensitive to pyrophosphorolysis.

rGTP. These nucleotide analogs proved to be very successful for accurate sequencing. Although these analogs can be substituted for GTP during chain elongation, they cannot be substituted for GTP during the initiation steps of transcription. Consequently, the use of rITP or 7-deaza-GTP would reduce the efficiency of transcription initiation. This problem can be overcome by addition of GMP to the reaction mixture. GMP is then incorporated first by RNAP, and then initiation efficiency increases dramatically.

2.3.5 DNA Templates

There are numerous widely used plasmid vectors that contain the corresponding T7, T3, or SP6 promoters. The most recent commercially available cloning vectors contain two different phage promoters flanking a multiple cloning site at both ends. Usually, one of phage promoters is the T7 RNAP promoter. In general, by digesting the template with a restriction enzyme that cleaves distal to the phage promoter on the other side, a run-off transcript of a defined size can be synthesized. These polymerases have been shown to be active on linear templates. However, supercoiled templates with slightly less activity of RNAP than linear templates can be also used in order to save the labor associated with preparation of linear templates.

Template preparation requires highly pure DNA that is free of ribonucleases. All RNase must be removed before using prepared DNA as a template for TS, and thus, most preparation protocols include an RNase treatment. In addition, T7 promoter-containing cloned or PCR-generated DNA can be used as a template.

Templates generated by polymerase chain reaction (PCR) having a promoter-tagged primer along with the sequence of interest allow production of RNA fragments for sequencing.

3 Remarkable Properties of Transcriptional Sequencing

3.1 Isothermal and Rapid Reaction

Because the transcription is an isothermal process, TS occurs at a constant temperature (37 °C) and does not require a thermal circler. Furthermore, the strong processivity of RNAP enables the reaction time to be greatly reduced. For standard reactions, the reaction time is 1 h, which is half that of standard cycle sequencing reactions.

3.2 Wide Dynamic Range

One of the most advantageous properties of TS is that it requires smaller amounts of DNA template when compared to conventional cycle sequencing. The amount of chain termination product is limited by the amount of template DNA in DNAP-dependent sequencing but by the amount of substrate in TS. In cycle sequencing, the chain termination products compete with primers for annealing with limited amounts of template, and the templates primarily anneal with the reaction products rather than the primer because of their size. In order to avoid this problem and increase reaction product, a large amount of primer might be added to reaction mixture. However, in those situations, misannealing of primer would frequently occur and the electropherogram pattern would be disordered. Therefore, number of molecules of sequence product cannot greatly exceed the number of molecules of template DNA. On the other hand, RNAP initiates transcription from a double-stranded promoter site, and one template DNA molecule can produce a large number of TS products. The TS reaction is terminated by the exhaustion of substrates. For this reason, for small amounts of template, better sequence quality has been obtained with TS rather than with cycle sequencing. In the current TS protocol, 10 ng of plasmid DNA can produce a high quality of electropherogram, while no signal can be detected using the conventional single dye terminator method.

3.3 Readability of GC-rich Templates

A remarkable feature of TS is that GC-rich templates can be easily sequenced. Usually, GC-rich templates are difficult to sequence by cycle sequencing methods because single-stranded DNA molecules with high GC contents tend to form secondary structures, such as stem loops, and, because DNAP uses ssDNA as a template, secondary structures of template DNA have a serious effect on elongation, resulting in sudden termination of polymerization. Such termination causes sudden stops in the sequence signals on electropherograms. On the other hand, dsDNA, whose secondary structure is less affected than that of ssDNA, is used as template in TS. In Fig. 4, the upper electropherogram (a) is the pattern for conventional cycle sequencing. The peak pattern is weakened and terminated in the middle of electrophoresis, where the content of GC is high. The lower electropherogram (Fig. 4b) is that produced for TS. No such effect is observed in this electropherogram,

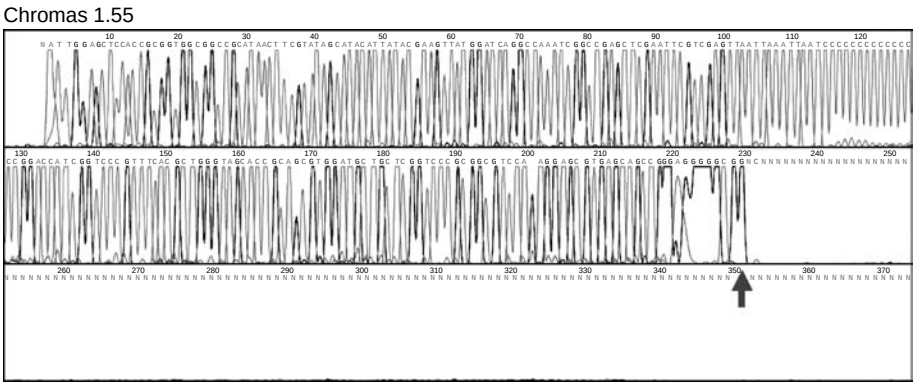

(a)

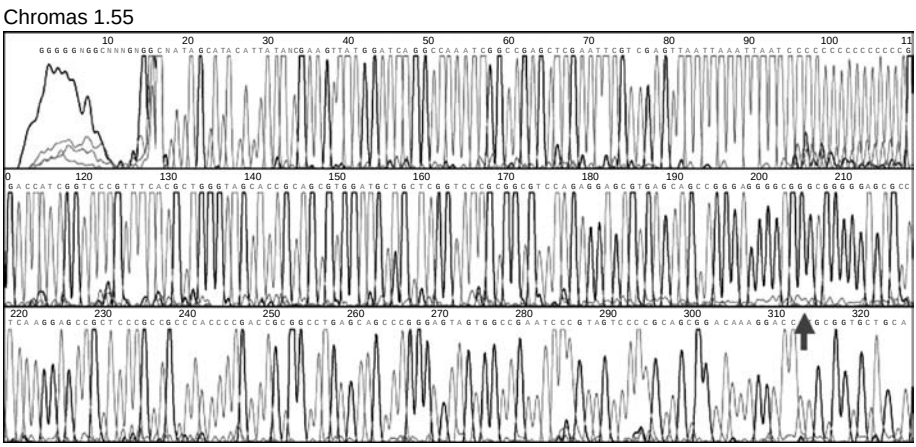

(b)

Fig. 4 Comparison of electropherograms
(a) Electropherogram using a Big Dye Terminator ver 1 kit. Peaks suddenly disappear at the point indicated by the arrow.
(b) Electropherogram using TS (CUGA® 7 sequencing kit) Arrow indicates the termination point shown in (a), and peaks continue beyond this point. Shibata K., et al., Practical application of transcriptional sequencing for GC-rich templates, *J. of Struct. Funct. Genomics.*, 2003, **4**(1), p. 35, by permission of Kluwer academic publishers.

thus, clearly demonstrating the applicability of TS to GC-rich regions.

This property is partly due to the origins of RNAP. Originally, RNAP evolved to initiate gene transcription from the 5′ site, where GC is preferentially located. GC-rich sequences, such as CpG Islands, were used for transcriptional control in the evolutional history of vertebrate genes. Therefore, it is very reasonable that GC-rich regions are more easily sequenced by TS than by cycle sequencing.

4 Detection of Transcriptional Sequencing Products

TS based on RI labeling was reported in 1985. Following these early reports,

however, none of the attempts to produce a clear TS pattern were successful until the four-colored fluorescence terminator appeared in 1995.

4.1 BODIPY-labeled Terminators for Auto Sequencers

Nippon Gene Tech. Co, Ltd. has developed fluorescent-labeled terminators using four colored BODIPY® fluorescent dyes. The synthetic terminators are a set of four fluorescent BODIPY-dye terminators. The four BODIPY dyes BODIPY FL, BODIPY R6G, BODIPY 564/570, and BODIPY 581/591 are bound with spacers to 7-deaza-3′-GTP, 7-deaza-3′-ATP, 3′-UTP, and 3′-CTP, respectively. The dye set and enzymes (T7 or T3 RNAP mutant) are now packaged together in one kit called the CUGA® sequencing kit, which is compatible with almost all sequencers currently in use. Chain termination RNA products can be detected in the same way as DNA ladders using conventional DNA sequencing kits, and the ladder pattern can be analyzed using the manufacturer-supplied basecaller.

4.2 Rhodamine-labeled Terminators for DNA Sequencing

Wako Pure Chemical Industries, Ltd. has synthesized fluorescent-labeled terminators. The synthetic terminators are a set of four fluorescent rhodamine-dye terminators. The four rhodamine dyes TMR, R6G, ROX, and R110 are bound with spacers to 3′-UTP, 3′-ATP, 3′-CTP, and 3′-GTP, respectively.

4.3 3′-dNTP Terminators for Mass Spectroscopy

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry has developed into a rapid, accurate, and sensitive method for the mass analysis of high molecular weight synthetic and biological polymers, including nucleotide acid molecules. This method enables rapid DNA sequencing using the chain termination method. This method is advantageous because it does not require separation matrices or fluorescent dyes, in contrast to conventional separation methods based on electrophoresis. However, despite these advantages, MALDI-TOF-MS cannot replace conventional electrophoresis, particularly for *de novo* sequencing, because of its short read length (less than 100 bp). This limitation is due to salt contamination and fragmentation of the polynucleotide by laser ionization.

Fragmentation occurs at the phosphodiester bond. In one hypothesis of DNA fragmentation, protonation of the nucleobase during the MALDI process is induced. The N-glycoside linkage at 1′-carbon becomes unstable, which is followed by base loss and cleavage of phosphodiester backbone. It has been shown that electron-withdrawing groups such as 2′-hydroxyl or 2′-fluoro groups stabilize the N-glycosidic bond, impede fragmentation, and thus, extend the accessible mass range. Because of the higher electron-withdrawing properties at the 2′-site of RNA, fragmentation of RNA polynucleotides is expected to reduce considerably. This property is advantageous for TS using MALDI-TOF-MS. In fact, TS using modified SP6 RNAP has been reported (Fig. 5), thus demonstrating the enormous potential for TS to apply MALDI-TOF-MS sequencing.

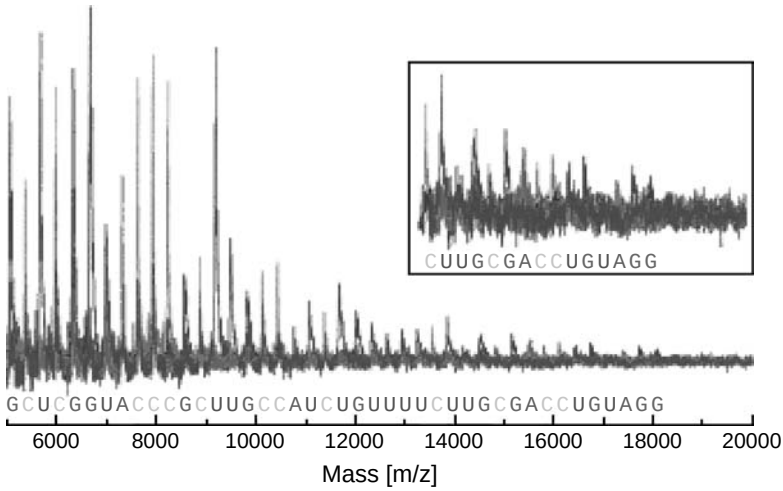

Fig. 5 Products of RNA chain termination analyzed by MALDI-TOF-MS. A nicked form of phage SP6 RNAP was used for chain termination reaction. MALDI-TOF-MS of reaction product allowed DNA sequence determination of 56 bases. Kwon Y, et al., DNA sequencing and genotyping by transcriptional synthesis of chain-terminated RNA ladders and MALDI-TOF mass spectrometry, *Nucleic Acids Res.*, 2001, **29**(3), e11, by permission of Oxford University Press.

5 Standard Protocol for Template Preparation, Reaction, and Electrophoresis

5.1 Template Preparation, Reaction, and Purification

RNase should be excluded in order to obtain high-quality sequences. For this reason, additional treatment is required when compared to conventional DNA sequencing methods. The procedure below is a rigid protocol for preparing DNA templates for TS.

1. Extract plasmid using standard alkaline lysis method.
2. Dissolve in 50 μL of TE (10 mM Tris-HCl (pH 8.0) and 1 mM EDTA).
3. Add 1 μL of RNase A solution (1 μg/μL) and store at room temperature for 30 min.
4. Add an equal volume of PEG solution (13% polyethylene glycol 6000 and 1.6 M NaCl), mix well, and then leave on ice for 20 min.
5. Centrifuge at 15 000 rpm for 20 min at 4 °C.
6. After removing the supernatant, rinse with 70% ethanol solution.
7. Dissolve pellets in 100 μL of TE and extract with phenol, phenol : chloroform, and then chloroform.
8. Add 200 μL of ethanol to the extracted solutions and leave tubes on ice for 20 min.
9. Centrifuge at 15 000 rpm for 20 min at 4 °C.
10. Dry for 10 min at room temperature.
11. Dissolve in 50 μL of TE.

Note: The RNase solutions supplied with commercially available plasmid preparation kits (ex. Qiaprep plasmid preparation kit from Qiagen Inc.) are not suitable for

preparation of plasmid DNA as a template for TS analysis.

For high-thoughtput sequencing, template preparation without RNase treatment is possible. Using the CUGA® kit, the reaction protocol is as described below.

1. Mix 4 μL of RNase-free plasmid DNA and 4 μl of 5x reaction buffer, 1 μL of 8 mM $MnCl_2$, 2 μL of NTP/Terminator mixture, and 8 μl of ddH_2O.
2. Add 1 μL of enzyme solution.
3. Add 1 μL of template.
4. Mix gently and incubate at 37 °C for 60 min.
5. Purify by EtOH precipitation.

5.2 Direct Sequencing without Purification

The most useful application of TS is direct sequencing. Direct sequencing is a method that allows PCR products to be sequenced directly after amplification from genomic DNA, BAC, plasmid DNA, or other DNA sources. In conventional dye terminator sequencing methods, PCR primers and 2′-dNTP substrates have to be removed in order to prevent mispriming and to preserve the ratio between the four 2′-dNTPs in subsequent chain termination reaction steps. On the other hand, in TS, PCR products amplified with T7 and T3 promoter-tagged primers do not need to be removed prior to sequencing because RNAP can recognize neither the single-stranded primers nor 2′-dNTPs. Therefore, PCR-direct sequencing can be achieved by addition of RNAP and appropriate reagents, without any purification steps after PCR amplification.

5.3 Electrophoresis

Although the RNase-free condition is theoretically essential, it seems that special treatment of the separation matrix and electrophoresis buffer to remove RNase are not required for TS. Sample loading on the gel slots and electrophoresis can be conducted in the same manner as for conventional DNA sequencing.

6 Future Applications of TS

6.1 Modification of Substrate

Numerous modified NTPs are incorporated by wild-type T7 RNAP. In particular, the Sp isomer of the [α-S]NTPs [nucleoside 5′-*O*-(1-thio-triphosphate)] is a good substrate for T7 RNAP. The Sp isomer contains a sulfur atom bound to the phosphate in the phosphodiester bond, and this sulfur atom is located on the side opposite the 2′-site of the sugar. This substrate set might make TS develop into a reaction system with higher performance. Moreover, if novel T7 mutants that incorporate the Sp isomer of the [α-S]NTPs at higher rates than normal NTPs could be designed, reaction speed and sensitivity would increase to levels above those in the current system.

6.2 Modification of Dye

Current TS dye sets for auto sequencers employ "single" dyes. The energy transfer mechanism has been used for dyes binding to terminators in cycle sequencing. These dyes have achieved better signal to noise ratios, resulting in more accurate and longer sequences. This technology has helped to realize the current genome projects (human, mouse, etc.) by substantially expanding the sequencing capacity because of its remarkable properties, for

example, readability of GC-rich regions. Energy transfer dyes for use with TS and substrate analogs will appear in the near future.

6.3 Chemical Modification of Substrate for Mass Spectroscopy

Reduction of fragmentation is important for yielding longer sequences with MALDI-TOF-MS. Chemical modifications of the nucleoside or backbone are able to reduce attacks to the N-glycosidic bond and block cleavage of the phosphodiester bond. Our group examined various nucleotide analogs for MALDI-TOF-MS. Fluorine at the 2′-position of ribose or 2′-arabinoribose highly reduces fragmentation, and moreover, backbone modification would be effective in blocking backbone cleavage. We investigated the effects of such changes using riboses, as shown in Fig. 6. S-RNA is a phosphorotioate oligonucleotide that is known to be a biostabilized nucleotide backbone owing to its nuclease resistance. We chemically synthesized 10, 20, and 30 mer oligos and used two types of matrix, 3-hydroxypicolinic acid (3-HPA) and trihydroxyacetophenone (THAP), in our experiments. The mass spectra of those oligos are shown in Fig. 7. These spectra suggest that phosphorothioate oligonucleotides are more resistant to signal drop-off than other native or chemical-modified ribonucleotides. The sulfur in the backbone is believed to prevent cleavage of the oligonucleotides.

6.4 Lab on a Chip

Because of miniaturization technologies, such as lithography, it may be possible to conduct experimental procedures on small chips. With this type of technology, once all of the samples are subjected to separation, the samples could not be recovered from a chip. Two basic conditions are preferred for such enzymatic reaction systems using lab-on-chip technology. One is that the reaction is isothermal, and the other is that the samples cannot be subjected to separation or purification steps, such as

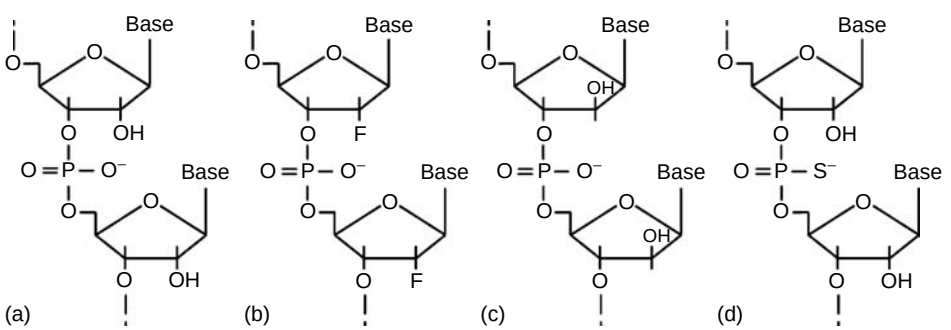

Fig. 6 Native RNA and modified RNAs used for experiments against fragmentation. (a) native RNA, (b) 2′-fluoro-RNA, (c) 2′-arabino-RNA, (d) S-RNA. RNA and S-RNA were purchased from GENSET OLIGOS (Kyoto, Japan). 2′-fluoro-RNA and 2′-arabino-RNA were purchased by TriLink BioTechnologies (San Diego, CA). Sequence of RNA and S-RNA oligos are d(GAUCUCAGCU), d(GAUCUCAGCUCUAAUGCGGU), d(GAUCUCAGCUCUAAUGCGGUUCGAUAAAUC) and d(GAUCUCAGCUCUAAUGCGGUUCGAUAAAUCCACUGUGAUA). Sequences of 2′-fluoro- and 2′-arabino-RNA were d(C)n (n = 10, 20, or 30, respectively).

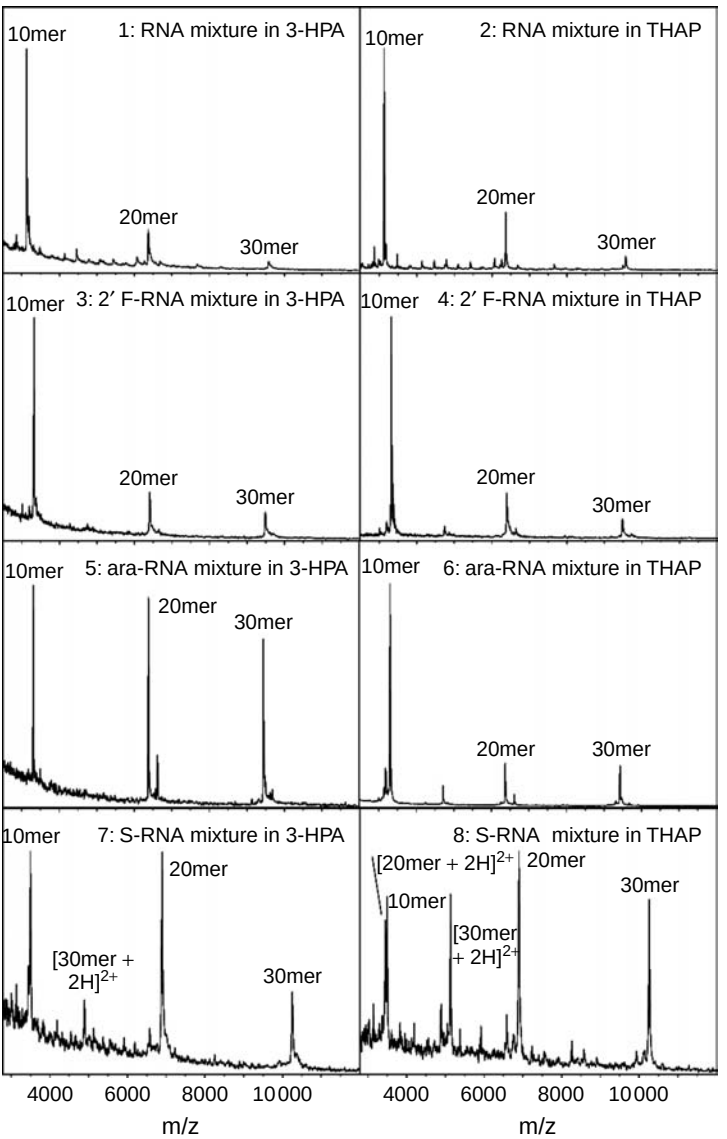

Fig. 7 UV-MALDI mass spectra of prepared oligos using 3-HPA (3-hydroxypicolinic acid) or THAP (trihydroxyacetophenone) matrixes.

centrifugation, ethanol precipitation, and phenol extraction. Thus, samples should be ready for sequencing in a detection system, such as fluorescence detection or mass spectrometry. TS meets these criteria perfectly, and we are hopeful that TS can be applied to lab-on-chip sequencing in the $1000 genome project.

Finally, as explained in this article, the principles of the chain termination are the same in both TS and cycle sequencing; however, the characteristic properties

differ substantially, but are complementary, between these two systems. This is primarily due to the enzymatic properties of DNAP and RNAP that developed during the evolutional process. Practical TS systems have been recently developed, and there are numerous possible applications in a wide variety of fields for these systems. We hope that this article aids in the understanding of TS and that further applications will be developed on the basis of TS methodology.

Acknowledgments

This study was supported by a Research Grant for the RIKEN Genome Explanation Research project from the Ministry of Education, Culture, Sports, Science, and Technology of the Japanese Government to Yoshihide Hayashizaki, a Research Grant for the National Project on Protein Structure and Functional Analysis from MEXT to Shigeyuki Yokoyama, and a Research Grant for Advanced and Innovational Research Programs in Life Science to Jun Kawai.

Bibliography

Books and Reviews

Brown, T.A. (1994) *DNA Sequencing: The Basics*, Oxford University Press, Oxford, UK.

Chamberlin, M.J. (1982) Bacterial DNA-dependent RNA Polymerases, *The Enzymes*, Vol. 15B, Academic Press, New York.

Enn, H.M. (1996) *Enzymology Primer for Recombinant DNA Technology*, Academic Press, California.

Primary Literature

Izawa, M., Sasaki, N., Watahiki, M., Ohara, E., Yoneda, Y., Muramatsu, M., Okazaki, Y., Hayashizaki, Y. (1998) Recognition sites of 3′-OH group by T7 RNA polymerase and its application to transcriptional sequencing, *J. Biol. Chem.* **273**(23), 14242–14246.

Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y., Hayashizaki, Y. (1998) Transcriptional sequencing: a method for DNA sequencing using RNA polymerase, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3455–3460.

Sanger, F., Nicklen, S., Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467.

Shibata, K., Izawa, M., Hayashizaki, Y., Watahiki, M. (2003) Practical application of transcriptional sequencing for GC-rich templates, *J. Struct. Funct. Genomics* **4**(1), 35–39.

Tabor, S., Richardson, C.C. (1995) A single residue in DNA polymerases of the Escherichia coli DNA polymerase I family is critical for distinguishing between deoxy-and dideoxyribonucleotides, *Proc. Natl. Acad. Sci. U.S.A.* **92**(14), 6339–6343.

Oligonucleotides

Olaf Heidenreich[1] *and Georg Sczakiel*[2]
[1] *Department of Molecular Biology, Institute for Cell Biology, University of Tübingen, Tübingen, Germany*
[2] *Institute for Molecular Medicine, University of Lübeck, Lübeck, Germany*

1 Introduction 415

2 Oligonucleotide Synthesis 415
2.1 Oligodeoxynucleotide Synthesis 416
2.2 Oligoribonucleotide Synthesis 417

3 Chemically Modified Oligonucleotides 419
3.1 Modifications of the Phosphate Linkage 419
3.1.1 Phosphorothioate Linkages 419
3.1.2 Methylphosphonate Linkages 420
3.1.3 Phosphoamidate Linkages 421
3.1.4 3′-3′ Linkages 421
3.2 Sugar Modifications 421
3.2.1 2′-Methoxy Oligonucleotides 421
3.2.2 2′-Fluoro and 2′-Amino Oligonucleotides 422
3.2.3 Morpholino Oligonucleotides 422
3.2.4 Peptide Nucleic Acids 422
3.3 Base Modifications 422
3.3.1 Pyrimidine Modifications 423
3.3.2 Purine Modifications 424
3.4 Nonnucleotidic Modifications 424
3.5 Fluorescent Oligonucleotides 424

4 Applications and Perspectives 425
4.1 General Considerations 425
4.2 Applications *in vitro* and in Cell Culture 425
4.2.1 Primer 426

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

4.3 Probes 428
4.4 Tools: Gene Function Analysis – Gene Validation 428

5 Applications *in vivo* and in Man ("Drugs") 429
5.1 Suppression of Pathological Gene Expression 429
5.2 Antisense Oligonucleotides 429
5.3 Ribozymes 430
5.4 Small RNAs – siRNA 431
5.5 CpG-Immunostimulatory Oligonucleotides 431

6 Prospects 431

Bibliography 432
Books and Reviews 432
Primary Literature 432

Keywords

Melting Temperature
The temperature at which nucleic acid double strands dissociate into single strands. The melting temperature is determined by properties of the oligonucleotide such as sequence or chemical modifications.

Oligodeoxynucleotide
An oligonucleotide consisting of 2′-deoxynucleotide moieties.

Oligoribonucleotide
An oligonucleotide consisting of 2′-hydroxynucleotide moieties.

Phosphoamidite
The building block for automated chemical synthesis of oligonucleotides.

Phosphodiester Linkage
The phosphate linkage connecting to nucleoside moieties within an oligo or polynucleotide.

Protection group
Groups that protect functional groups such as amino or hydroxyl groups against unwanted side reactions during the course of synthesis.

Oligonucleotides are important tools in biochemistry and in molecular and cellular biology. Moreover, oligonucleotides such as antisense oligonucleotides or small interfering RNAs hold promise as therapeutic agents to combat human disease. The main prerequisite for the wide use of oligonucleotides was the development of efficient and robust synthesis protocols, which allow for the production of oligonucleotides at a wide range of synthesis scales with very affordable costs. Furthermore, the chemical synthesis permits the site-specific introduction of chemical modifications such as phosphate, sugar or base derivatives, which affect the chemical and biological properties of oligonucleotides. Oligonucleotides are used in molecular biology for sequence detection and determination as well as in cell biology and molecular medicine for functional genomics, target validation, and drug development. This review describes recent advancements of oligonucleotide synthesis and discusses several common applications of oligonucleotides.

1 Introduction

Nucleotides can form homopolymers and heteropolymers with chain lengths ranging from as short as a dinucleotide or only a few nucleotides to extremely long chains that approach the length of long viral genomes or even whole genomes or chromosomes of lower and higher organisms. Depending on their sugar chemistry, nucleotides may form either ribonucleic acids (RNA), deoxyribonucleic acids (DNA) or they may be chemically modified. Oligomeric nucleic acids, that is, oligonucleotides, are of particular relevance and importance as tools and drugs in molecular biology and biomedicine. Oligonucleotides have been the subject of an extremely large variety of chemical modifications that are meant to confer favorable biochemical and biological properties. Improving their chemistry and adapting the chemistry to the specific requirements for their use is crucial to successful application.

The fields of research and development and the range of purposes for which oligonucleotides are used are extremely wide. Conversely, basic considerations and principles regarding the synthesis and the use of oligonucleotides are limited. Thus, it seems to be appropriate to describe fundamental issues of the synthesis and chemical characteristics of oligonucleotides and, subsequently, to introduce into the fields of their application in molecular medicine and related disciplines.

2 Oligonucleotide Synthesis

Oligonucleotides are commonly prepared by automated chemical synthesis on a solid support. For that, the nucleotide building block must contain a reactive phosphorus, allowing the formation of a 3′-5′ phosphodiester linkage. The phosphoamidite chemistry is currently the method of choice for the chemical synthesis of oligonucleotides. The phosphoamidite group contains such a reactive phosphorus linked

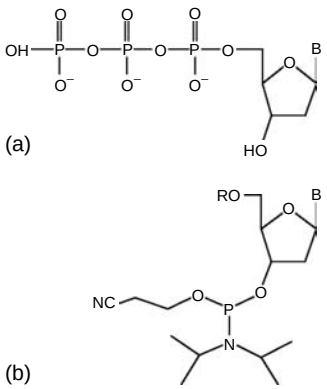

Fig. 1 Biological and chemical nucleotide monomers.
(a), 2′-deoxynucleoside triphosphate as a substrate for DNA polymerases;
(b), 2′-deoxynucleoside phosphoamidite as a building block for chemical oligonucleotide synthesis.

to the 3′-oxygen of the sugar moiety. This contrasts with the natural building blocks of nucleic acid synthesis, the nucleoside triphosphates, which carry the linkage-forming phosphorus on the sugar 5′-oxygen (Fig. 1). Consequently, chemical oligonucleotide synthesis proceeds in a 3′ to 5′ direction, whereas the biologic nucleic acid synthesis proceeds in the opposite direction. Other functional groups such as hydroxyl groups of the sugar and amino groups of the bases and also the phosphoamidate group itself need to be protected to exclude possible side reactions. Many companies offer oligodeoxynucleotide synthesis for very affordable prices. Oligoribonucleotide synthesis is considerably more expensive owing to more demanding synthesis protocols. The usual synthesis scales range from 10 nmol to 15 μmol corresponding to 50 μg to 10 mg of an oligonucleotide of 20 nucleotides in length. Oligonucleotide lengths of more than 100 nucleotides can be synthesized on a nucleic acid synthesizer.

An important feature of automated synthesis is the possible site-specific introduction of chemical modifications of the bases or the sugar-phosphate backbone. In conclusion, oligonucleotides can be synthesized for many different applications ranging from physicochemical over biological to medical applications.

2.1 Oligodeoxynucleotide Synthesis

The starting point of the synthesis is a protected nucleoside linked via its 3′-oxygen to a solid support (Fig. 2). Nowadays, nucleoside phosphoamidite chemistry is routinely used for this synthesis. The corresponding building blocks are shown in Fig. 2. The synthesis cycle consists of four steps: (1) Deprotection of the 5′-hydroxyl group, (2) coupling of nucleotide phosphoamidite to the 5′-hydroxyl group, (3) capping of unreacted 5′-hydroxyl groups, and (4) oxidation of the phosphorus. During the first step, the deprotection step, the DMT protection group of the 2′-deoxyribose 5′-oxygen is cleaved off to yield a free hydroxyl group. This hydroxyl group forms during the coupling step a linkage with the phosphoamidite group of the next building block, yielding a phosphite triester. The coupling step is with more than 98% of the 5′-hydroxyl groups reacting very efficient. However, there is always a small fraction of unreacted nucleosides. Since these may react during the following cycle with the next building block, leading to oligonucleotides with internal deletions, such unreacted groups are capped by acetylation. As a result, one obtains only prematurely terminated oligonucleotides but never oligonucleotides with internal deletions, which would be difficult to separate from the

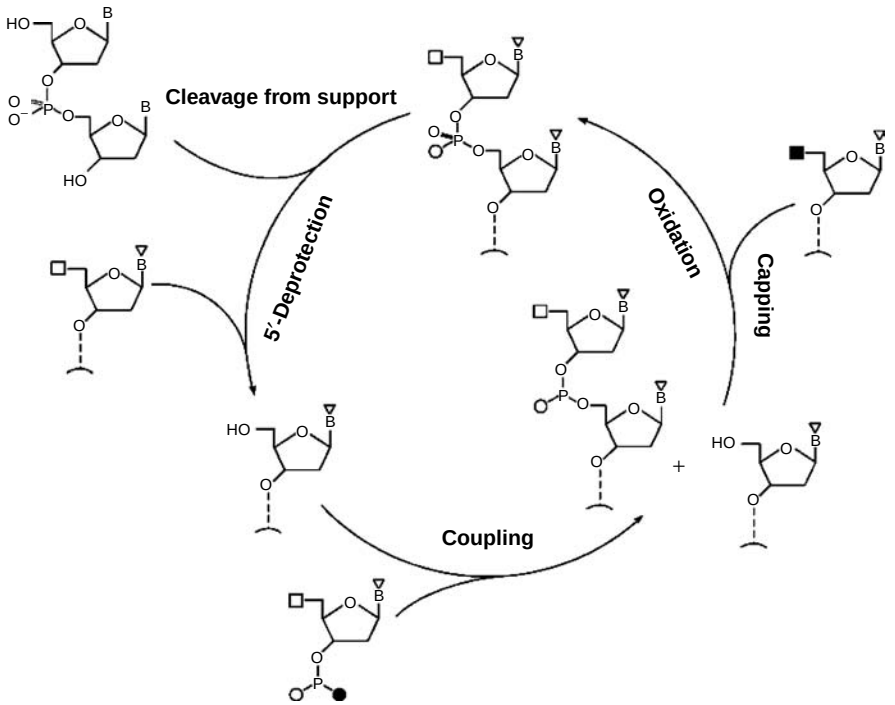

Fig. 2 Oligonucleotide synthesis scheme. The cycle for oligodeoxynucleotide synthesis is depicted. Step 1 initiates the synthesis with a nucleoside attached to a solid support and is not part of the repeating elongation cycles. B, base; white rectangle, 5′-protecting group; black rectangle, capping group; triangle, base protecting group; white circle, β-Cyanoethyl group, gray circle, amidite group. Oligoribonucleotide synthesis proceeds similarly with different building blocks containing an additional 2′-protecting group.

full-length oligonucleotide. Next, the phosphite linkage is oxidized to a phosphate linkage still carrying a protecting group at one nonbridging oxygen. The cycle is now completed; the oligonucleotide on the support is elongated by one nucleotide. After addition of the last building block, the oligonucleotide is cleaved from the solid support, and the base and phosphate protecting groups are removed. Oligonucleotides routinely contain a 5′-hydroxyl group. If a 5′-phosphate group is required, this can be introduced on the support as the final coupling step. Full-length oligonucleotides are then separated from prematurely terminated ones by either HPLC or denaturing polyacrylamide gel electrophoresis.

2.2 Oligoribonucleotide Synthesis

Like their oligodeoxynucleotide counterparts, oligoribonucleotides are synthesized using a phosphoamidite approach. The necessity of protecting the 2′-hydroxyl group of the ribose moiety results in a more demanding synthesis of the building blocks and also in a more complicated deprotection of the final oligonucleotide.

The conventional protection strategy consists of an acid-labile dimethoxytrityl

(DMT) group protecting the 5′-oxygen, and a silyl group blocking the 2′-oxygen of the ribose such as the *tert*-butyldimethylsilyl (TBDMS) or, more recently, the triisopropylsilyloxymethyl (TOM) group (Fig. 3). The TBDMS group is stable under acidic conditions and fairly stable under alkaline conditions necessary

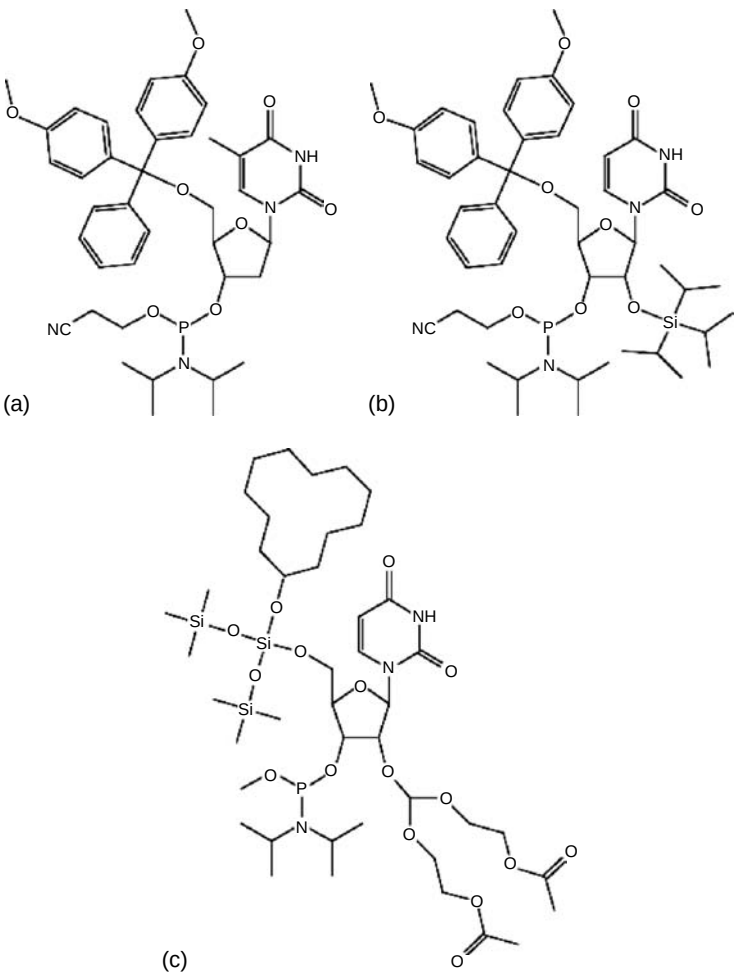

Fig. 3 Nucleoside phosphoamidite building blocks.
(a) Oligodeoxyribonucleotide building block containing a 5′-Dimethoxytrityl (DMT) protecting group and a phosphoamidite protected with a β-cyanoethyl group. (b) Oligoribonucleotide building block containing an additional 2′-O-triisopropylsilyloxymethyl (TOM) protecting group. (c) Alternative oligoribonucleotide building block containing a 2′-O-bis(2-acetoxyethoxy)methyl (ACE) protecting group, a 5′-O-Bis(trimethylsiloxy)cyclo-dodecyloxysilyl ether (DOD) protecting group for adenosine and cytidine (cyclooctyl for guanosine and uridine) and a phosphoamidite protected with methoxy group.

for the deprotection of the base functions and the phosphate linkages. However, a small fraction of the silyl groups may also be removed, resulting in strand cleavage of the oligoribonucleotide. For that reason, yields of full-length oligoribonucleotides often did not match those for oligodeoxyribonucleotides. The recent replacement of the TBDMS group by the TOM protecting group results in higher coupling yields; in combination with a modified deprotection protocol, this approach achieves full-length oligoribonucleotide yields comparable to oligodeoxyribonucleotide synthesis.

Alternatively, the 5′-oxygen can be protected by a silyl group, which can be easily removed during the synthesis cycle. In this case, an ACE group blocks the 2′-oxygen, which is stable under alkaline conditions (Fig. 3). After finishing the synthesis, cleaving the oligoribonucleotide from the solid support and deprotecting bases and phosphate linkages, the ACE group is cleaved off under mild acidic conditions. Like the previous TOM approach, this strategy achieves oligoribonucleotide yields approaching those for oligodeoxynucleotides.

Alternatively, oligoribonucleotides may be synthesized enzymatically by an *in vitro* transcription approach using DNA-dependent RNA polymerases from bacteriophages such as SP6, T3 or T7. This is still the method of choice for the synthesis of longer oligoribonucleotides. One limitation of this approach is the rather restricted set of modified nucleoside triphosphates accepted as substrates for the RNA polymerases. In addition, it is very demanding and often impossible to introduce site specifically modifications into an oligoribonucleotide.

3 Chemically Modified Oligonucleotides

3.1 Modifications of the Phosphate Linkage

The phosphate linkage is the point of attack for all nucleases. Therefore, particularly for antisense oligonucleotide applications, this linkage is a common target for stabilizing modifications.

3.1.1 Phosphorothioate Linkages

The most frequent phosphate modification is the replacement of a nonbridging oxygen by sulfur, yielding a phosphorothioate linkage (Fig. 4). This substitution creates a new chiral center in the oligonucleotide. Depending on the replaced oxygen, the phosphorothioate diester has either an Rp or an Sp configuration. Phosphorothioates are introduced during solid-phase synthesis by oxidizing the phosphite linkage with either molecular sulfur or a sulfur-transferring agent. Thus, for phosphorothioate oligonucleotide synthesis, the same phosphoamidite building blocks can be used as for the synthesis of unmodified oligonucleotides. Since the oxidation is not stereo selective, the resulting phosphorothioate oligonucleotide represents a diastereomeric mixture. In oligodeoxynucleotides, phosphorothioate groups allow for the site-specific attachment of reporter groups such as cross-linking agents of fluorescent dyes. Terminal phosphorothioates can be also introduced during chemical synthesis or afterwards by enzymatic manipulation of the purified oligonucleotide. Like internal phosphorothioate linkages, terminal phosphorothioate groups can be used for attaching dyes to the oligonucleotide. In addition, they permit coupling of oligonucleotides

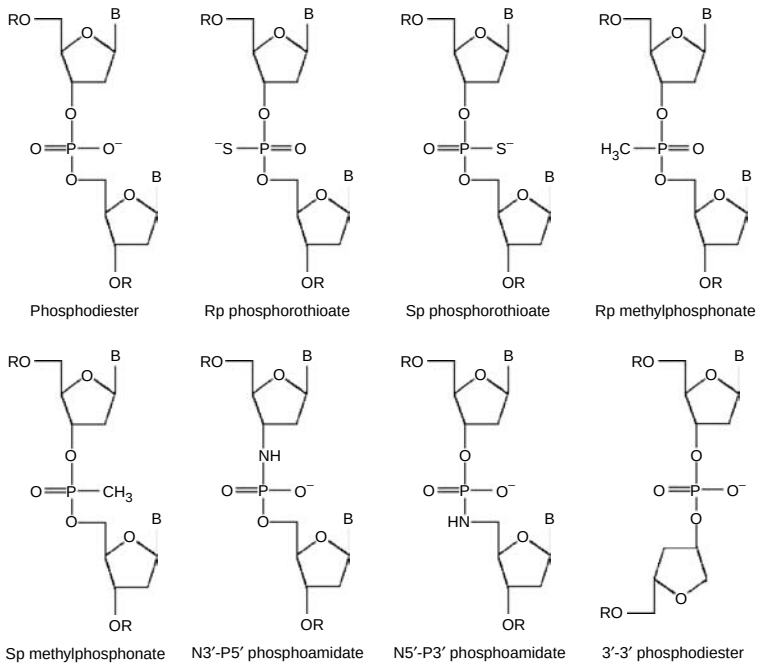

Fig. 4 Phosphate modification. Note that replacement of one nonbridging oxygen by a sulfur or a methyl group results in two different diastereomers.

to bromoacetyl agarose, for example, affinity purification of nucleic acid–binding proteins.

The phosphorothioate linkage is much less susceptible to nuclease-mediated degradation. Like unmodified oligodeoxynucleotides, phosphorothioate oligodeoxynucleotides induce RNase H-mediated degradation of complementary RNA sequences. These benefits are accompanied by some disadvantageous properties. One major disadvantage is a decreased affinity of phosphorothioate oligonucleotides for complementary sequences compared to their unmodified counterparts. Moreover, phosphorothioate oligonucleotides exhibit a strong unspecific binding to many proteins. In spite of these problems, phosphorothioate oligonucleotides are currently still the most popular antisense oligonucleotides.

3.1.2 Methylphosphonate Linkages

Substitution of one nonbridging oxygen by a methyl group yields a methylphosphonate linkage (Fig. 4). As in the case of phosphorothioate linkages, this replacement introduces an additional chiral center into the oligonucleotide backbone. To synthesize methylphosphonate oligodeoxynucleotides, special phosphoamidite building blocks already carrying the methyl group linked to the phosphorus are necessary. Like their phosphorothioate counterparts, methylphosphonate linkages reduce the affinity of an oligodeoxynucleotide to its complementary strand and increase its resistance against nucleases. However, methylphosphonate oligodeoxynucleotides are not as ‘‘sticky’’ as phosphorothioate oligonucleotides to proteins, and they do not support RNA degradation by RNase H.

3.1.3 Phosphoamidate Linkages

Phosphoamidate linkages, which should not be mixed up with the phosphoamidite building block, are characterized by nitrogen replacing either the bridging 3′- or the 5′-oxygen of the corresponding sugar part (Fig. 4). Despite their low chemical stability under acidic conditions, the 3′-*N*-phosphoamidates are particularly interesting for antisense approaches. These linkages are completely resistant against nucleases. Furthermore, in contrast to phosphorothioate or methylphosphonates, introduction of phosphoamidate linkages significantly increases the affinity to complementary nucleic acid strands. Like methylphosphonates, phosphoamidate oligonucleotides have a much lower tendency than phosphorothioate oligonucleotides to bind to proteins and do not induce degradation of complementary RNA strands by RNase H.

3.1.4 3′-3′ Linkages

Exonucleases degrade oligonucleotides from the 5′- or 3′-terminus. Their substrate is the 5′-3′-phosphodiester linkage. Introduction of a terminal 3′-3′-phosphodiester linkage confers resistance toward exonuclease action (Fig. 4). For instance, hammerhead ribozymes containing a 3′-terminal 3′-3′-phosphodiester linkage are efficiently protected against a 3′-exonuclease activity present in blood serum.

3.2 Sugar Modifications

3.2.1 2′-Methoxy Oligonucleotides

Methylation of the ribose 2′-oxygen is a naturally occurring modification found in rRNAs, tRNAs, and in the cap structure of mRNAs and snRNAs. This modification not only confers resistance to 2′-OH-dependent ribonucleases but also decreases the susceptibility toward other nucleases (Fig. 5). Since 2′-methoxyribose adopts a similar conformation as the unmodified ribose, 2′-methoxyoligonucleotides do not induce RNase H-mediated RNA degradation. However, 2′-methoxyoligonucleotide-RNA duplexes exhibit a higher thermal

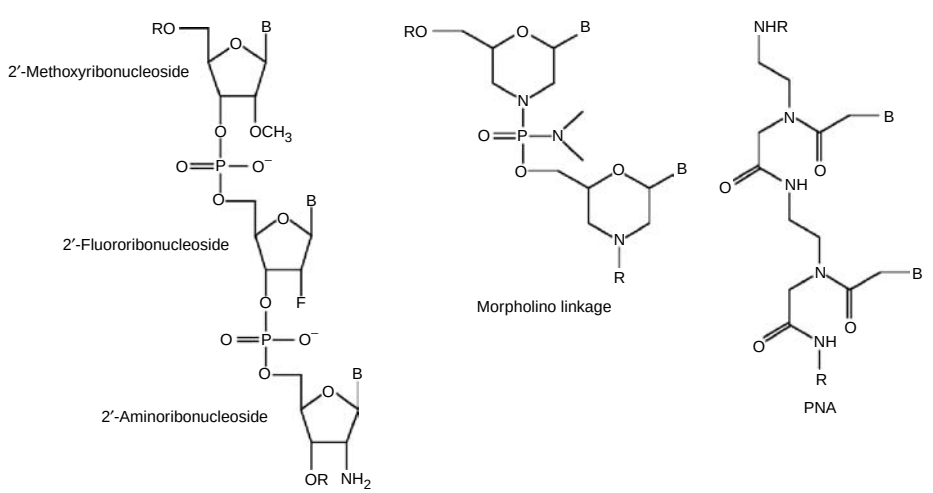

Fig. 5 Sugar and backbone modifications.

stability than oligodeoxynucleotide-RNA duplexes. 2′-Methoxyoligonucleotides are used for the isolation or depletion of RNA-binding proteins in cell lysates and for antisense applications. During the last decade, many other alkyl residues such as the methoxyethyl group or the allyl group have been introduced at the 2′-position to stabilize oligonucleotides against nuclease attack without compromising their affinity toward complementary RNA sequences.

3.2.2 2′-Fluoro and 2′-Amino Oligonucleotides

Introduction of fluorine or an amino group at the 2′-position yields 2′-fluoro or 2′-aminooligonucleotides, respectively (Fig. 5). Like the 2′-methoxy group, these modifications impart resistance toward 2′OH-dependent RNases such as RNase A. However, the affinity of oligonucleotides toward complementary RNA decreases in the order 2′-fluoro > 2′-OH (ribo) = 2′-methoxy > 2′-deoxy > 2′-amino and is paralleled by the tendency of the sugar to adopt a 3′-endo conformation. Like phosphorothioate linkages, 2′-amino groups can be used to introduce a reporter group into the sugar-phosphate backbone.

3.2.3 Morpholino Oligonucleotides

The following two examples describe modifications that replace the complete oligonucleotide sugar-phosphate backbone by new structures. The morpholino oligonucleotide contains hexameric morpholino rings and phosphodiamidate linkages (Fig. 5). The morpholino ring carries the base and replaces therefore the sugar moiety in conventional oligonucleotides. In contrast to the latter, the morpholino backbone is neutrally charged and has a lower tendency for protein binding. Such oligonucleotides are very resistant against nucleases and form duplexes with RNA of similar stability as oligodeoxynucleotides. Since such morpholino oligonucleotide-RNA duplexes are no substrate for RNase H, morpholino antisense oligonucleotides do not induce the degradation of their target RNA but block efficiently the initiation of translation when targeted against the translational start site. For that reason, morpholino antisense oligonucleotides are widely used to examine gene function in developmental model organisms such as zebra fish or Xenopus.

3.2.4 Peptide Nucleic Acids

The most radical backbone modification is its complete replacement by a peptide backbone (Fig. 5). Such peptide nucleic acids (PNAs) contain a neutral backbone and exhibit a very high affinity toward complementary nucleic acid sequences. They are completely resistant against both nuclease and proteinase-mediated decay but are less soluble in aqueous solvents than "normal" oligonucleotides. Unlike most other types of oligonucleotides, PNAs can invade into double-stranded DNA and may have the potential to modulate gene expression by interfering with the transcriptional process.

3.3 Base Modifications

In contrast to most modifications of the sugar-phosphate backbone, base modifications hardly affect the stability of an oligonucleotide. Instead, they often allow structure-function analyses by either substituting a functional group or by introducing new functions such as fluorescent labels or cross-linking groups. We discuss here only a small selection of base modifications amenable for automated synthesis (Fig. 6).

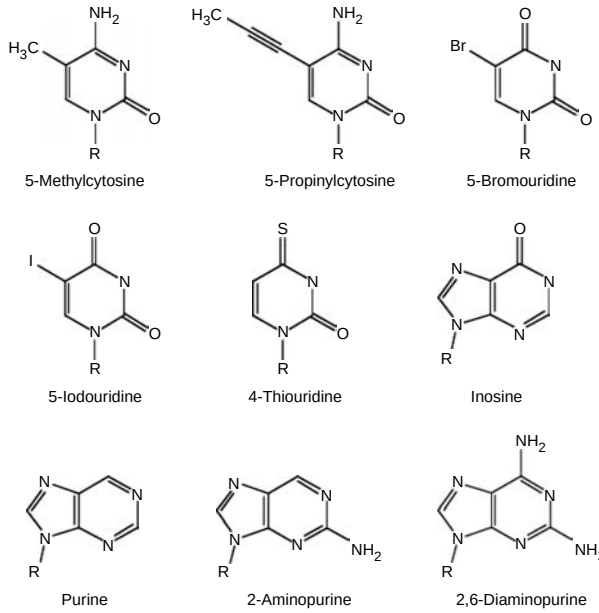

Fig. 6 Base modifications.

3.3.1 Pyrimidine Modifications

A prominent site for the introduction of modifications is the 5-position of the pyrimidine ring. Since this position is not involved in the formation of Watson–Crick base pairs, even quite bulky modifications can be introduced without significantly affecting the thermal stability of a nucleic acid double strand. A naturally occurring modified base is 5-methyl cytosine (Fig. 6). Oligonucleotides containing unmethylated CG motifs may stimulate immune responses, thereby negatively affecting the specificity of an antisense or an aptameric oligodeoxynucleotide. Replacement of the unmethylated cytosine by 5-methyl cytosine suppresses this immunostimulation and is thus frequently used to study such a possible mode of action. Photoaffinity labels such as 5-bromo or 5-iodo uridines can be used to crosslink an oligonucleotide to DNA- or RNA-binding proteins upon UV irradiation. In addition, cellular uptake and intracellular distribution of 5-bromo deoxyuridine-containing oligonucleotides can be examined with commercially available antibodies. Another 5-modification of the pyrimidine ring being of relevance for antisense oligonucleotides is the propinyl group. Compared to unmodified ones, oligodeoxynucleotides containing 5-propinyl cytidines and uridines show a higher affinity toward complementary RNA.

To investigate the interaction of bases with metal ions such as magnesium ions, oxygens may be replaced by sulfur. Such a replacement converts a ''hard'' center preferentially coordinating to magnesium ions to a ''soft'' center binding to calcium or cadmium but not to magnesium ions. For instance, 4-thio uridine has been applied to examine potential magnesium binding sites in the catalytic core region of the hammerhead ribozyme. Moreover, like

5-bromo uridine, 4-thio uridine is a photoactivatable cross linker to both nucleic acids and therewith interacting proteins.

3.3.2 Purine Modifications

Inosine is generated in the cell by deamination of adenosine by enzymatic adenosine deaminase activities. Inosine is frequently found in the +1 position of the tRNA anticodon loop where it can base-pair with uridine, cytosine, and adenosine. For this reason, it is used for the synthesis of primers containing degenerate base positions. Such primer can bind to DNA or RNA strands with incomplete sequence information or to a family of related sequences without a mismatch "penalty" reducing the thermal stability of such duplexes. Together with other modified purines such as 2,6-diamino purine or purine itself, it has been used to study protein–DNA interactions or ribozyme catalysis. The application of the fluorescent purine analogue 2-amino purine permitted the study of conformational changes associated with hammerhead ribozyme catalysis.

3.4 Nonnucleotidic Modifications

Automated oligonucleotide synthesis allows the introduction of nonnucleotidic modifications. In particular, alkyl chain groups such as the C3 or C6 chain in combination with terminal amino or sulfhydryl groups are frequently attached to the 5′- or 3′-terminus by solid-phase chemistry. After finishing the synthesis, the terminal functional group can be used to introduce further modifications or to attach the oligonucleotide to a matrix for affinity purification of, for instance, nucleic acid–binding protein complexes. In addition, carrier peptides such as antennapedia or tat peptides can be linked to the oligonucleotide via its sulfhydryl group. The resulting peptide–oligonucleotide conjugates show a significantly enhanced uptake and activity in cell culture when compared to the nonconjugated oligonucleotide alone.

Single-stranded loops are a preferred point of attack for nucleases. Replacement of such susceptible loops by an internal hexaethylene glycol linker may increase the stability of oligonucleotides hairpins in body fluids and in cell culture. For the hammerhead ribozyme, it has been shown that the replacement of its loop by such a linker hardly affected its catalytic efficiency.

3.5 Fluorescent Oligonucleotides

An important class of nonnucleotidic modifications consists of fluorochromes such as fluorescein or rhodamine dyes. Dependent on the type of phosphoamidite used, such fluorochromes can be internally as well as 5′-terminally attached. For 3′-labeled oligonucleotides, synthesis is performed on a solid support with the fluorescent dye already attached to it.

Fluorescently labeled oligonucleotides are used to study cellular oligonucleotide uptake and their intracellular distribution. Further applications are in DNA sequencing as fluorescent primers and in the analysis of intra- and intermolecular interactions by, for instance, fluorescence correlation spectroscopy. Combination of two different fluorochromes, of which the emitted light from the first dye excites the second dye, are applied in fluorescence resonance energy transfer studies. The same principle is used for real-time polymerase chain reaction (PCR), where the formation of a PCR product is followed on-line

by Fluorescence resonance energy transfer (FRET) analysis using fluorescently labeled oligodeoxynucleotide probes.

4 Applications and Perspectives

4.1 General Considerations

The application of single-stranded oligonucleotides in molecular biology is tightly related to the base-specific recognition of complementary nucleic acids usually via Watson–Crick base pairing. This means, whenever a sequence-specific detection or interaction between the oligomeric probe, tool, or drug and a nucleotide sequence of interest can be formed, this recognition step provides the signal of interest, confers specificity and selectivity, or is crucial for efficacy in biological systems. More recently, it became clear that also double-stranded RNA, including siRNA, may recognize its target RNA in a highly sequence-specific fashion, suggesting a step that involves nucleotide sequence recognition as well. Oligonucleotides represent a highly sensitive tool to qualitatively and quantitatively detect and to monitor nucleic acids in many kinds of experimental setup and environment. This includes, for example, solutions and buffers as well as structured biological material, including fixed cells and tissue slices. In the latter systems, fluorescently or radioactively labeled oligonucleotides are being used also for imaging purposes and to define the localization of specific nucleic acids within tissues, cells, subcellular structures, and compartments.

In the following, a number of such applications in molecular biology, cell biology, and molecular medicine will be described. In principle, the entity 'oligonucleotide' is the same in many fields of application in terms of length, chemistry, and biochemical function. However, the range of concepts and strategies in which oligonucleotides represent a key tool is extremely wide (Fig. 7). Here, selected key methodologies out of the large set of currently established protocols that include the use of oligonucleotides will be described in order to explain their role as a tool and drug in molecular biology.

With regard to the field of therapeutic approaches on the molecular level and the level of correction of gene expression, several classes of oligonucleotides are being used as lead compounds and as drugs in current molecular medicine. Among the potentially therapeutic oligonucleotides suited to suppress pathological and aberrant gene expression are groups such as antisense oligonucleotides, ribozymes, or double-stranded RNA as well as those inducing biological responses that are of therapeutic benefit in man, such as CpG-immunostimulatory oligonucleotides, decoys, or aptamers. To illustrate the role of oligonucleotides as drugs, some prominent groups will be discussed in chapter 5.

4.2 Applications *in vitro* and in Cell Culture

For most purposes, in biochemical reactions and for most experimental protocols commonly used in solution, *in vitro* oligonucleotides are unmodified. This means they can be obtained with little expense. However, a number of chemical modifications that improve the biochemical and biological characteristics of oligonucleotides can also be obtained from commercial suppliers at appropriate expense.

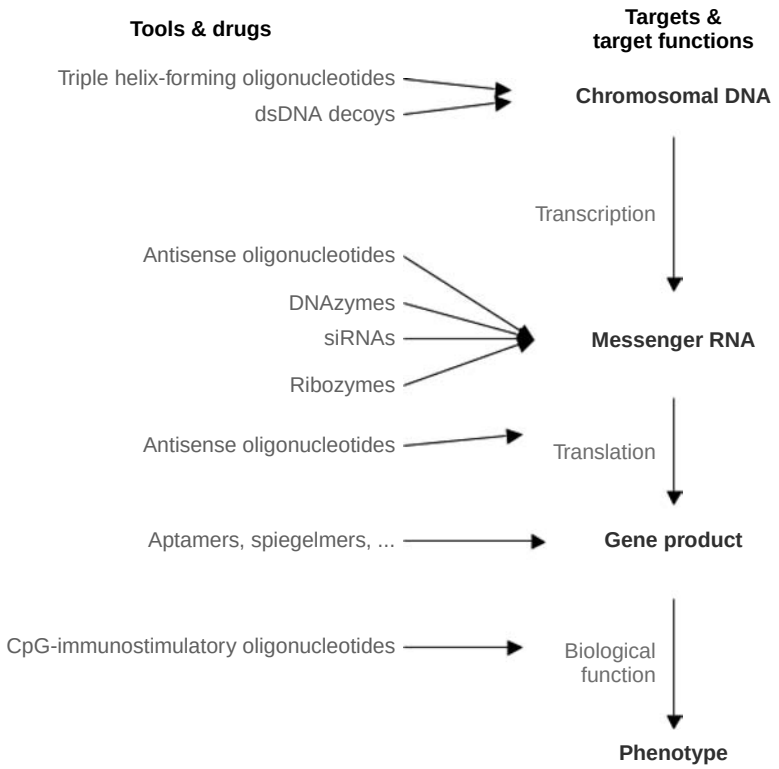

Fig. 7 Summary of modes of action of several classes of oligomeric tools and drugs acting on gene expression, gene function, and cellular processes.

4.2.1 Primer

In a number of extremely important and fundamental methods frequently applied in almost all fields of current biochemical, biological, and medical research, oligonucleotides serve as sequence-specific primers for enzymatic reactions *in vitro*. Such primes recognize their template in a sequence-specific manner, which confers "specificity" (Fig. 8). For example, oligonucleotides are commonly used as primers for directed reverse transcription of RNA into cDNA as an alternative to the unspecific oligo (dT)-primed synthesis of cDNA, or for the extension of the primer along a given template, which sheds light on position-specific characteristics of the substrate strands.

It is important to note that the binding efficiency of this step including sequence-selectivity is crucially influenced by the biochemical conditions of the reaction such as temperature, ionic conditions, and stoichiometry. A subsequent or combined enzymatic reaction usually in the use of a polymerase or domains thereof produces the signal that is used for detection or quantification of the templates that had been present in the original sample.

Primer extension reactions are also commonly used in the structural probing of RNA. In such analyses, structured RNA is modified and cleaved first in a sequence-

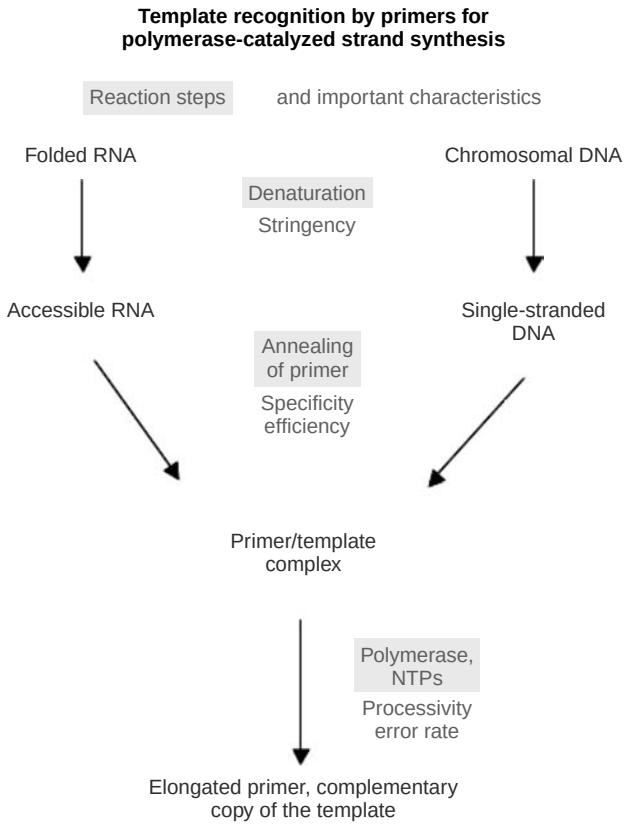

Fig. 8 Interactions between primers and their DNA or RNA templates. Key steps include the denaturation of the template and the binding of a primer to its template via base–base recognition followed by an enzymatic extension.

or structure-specific way followed by a primer extension reaction that terminates at the positions of cleavage.

Finally, and almost with an historical perspective to Sanger's work, it is worth noting that primers are crucial to the enzymatic sequence determination of DNA and RNA. Even though this kind of sequence determination is almost completely automated, oligomers as primers are still in highly frequent use.

PCR The polymerase chain reaction (PCR) makes use of two primers that direct their elongation by DNA polymerase in opposite directions such that their elongated products form two complementary single strands. They may pair to form a double-strand and increase the quantity of specific DNA double strands. However, the enzymatically synthesized double strands may also be denatured such that the primers present in excess in the reaction mixture can bind and form a new template for polymerase activity and the production of new duplexes.

At present, there are a number of academic and commercial computer

algorithms available that are held to design appropriate primers for PCR. Such programs are usually based on thermodynamic parameters and include components that align selected primer sequences with relevant sequence spaces and non-target sequences in order to avoid the amplification of irrelevant products that give rise to false signals and results.

In situ PCR Usually, a typical PCR reaction is carried out in solution in tubes that can be heated and cooled to specific temperatures in many repetitive cycles. Such reactions can also be performed in chambers that contain fixed cells or tissues that harbor the nucleic acid of interest, which may serve as templates for PCR reactions *in situ*. The necessary nucleotide triphosphates, buffer components, and enzymes are all able to access their "fixed" templates inside "fixed cells". By incorporating chemically modified nucleotides that can be recognized, for example, by antibodies (e.g. BrdU-directed antibodies in case of incorporation of bromodeoxyuridine, BrdU) or other high-affinity molecules, the *in situ* PCR reaction products can be visualized and monitored. This technique is clearly a qualitative method and requires minimal amounts of template nucleic acids to be detected. For example, in case of the detection of viral nucleic acids in infected cells and tissues, this method is very helpful to investigate the time course and subcellular localization of genomic and subgenomic viral nucleic acids.

4.3 Probes

In order to detect nucleic acids of interest, one usually exploits the specificity of Watson–Crick hydrogen bonding between two complementary strands. The strand to be detected can be recognized by a complementary oligonucleotide that contains a detectable label, for example, a radioisotope, a fluorescent moiety, or a substituent, which can be labeled subsequent to double-strand formation such as biotin or BrdU. Probes of oligomeric length offer certain advantages and disadvantages. The possibility to chemically synthesize oligonucleotides provides some of the most obvious advantages. Probes can be chemically modified such that they have improved properties, including increased binding stability (e.g. PNA, 2′-alkyl derivatives), increased resistance to nucleases usually present in biological environments (e.g. α-anomeric DNA, 2′-fluoro, 2′-alkyl), or improved discrimination between perfectly matching substrate strands and not perfectly matching strands.

In order to image nucleic acids and certain nucleotide sequence segments within chromosomal DNA, chemically modified and fluorescently labeled oligomeric nucleic acids are important. For example, 2′-*O*-allyl-modified riboprobes carrying a fluorescent label had been successfully used to determine the subcellular localization of small RNAs by image analysis of cells fixed *in situ*. And probably even more importantly, fluorescently labeled probes can be used for "fluorescence *in situ* hybridization" (FISH), a method routinely used to detect chromosomal aberrations in a diagnostic fashion.

4.4 Tools: Gene Function Analysis – Gene Validation

Oligomeric nucleic acids comprise a large set of functionally distinct subgroups of oligonucleotides. Prominent classes of oligomeric tools and drugs used in molecular biology include (i) short

double-stranded RNA most commonly termed "siRNA", micro RNA or "miRNA", (ii) antisense oligonucleotides, (iii) triple helix-forming oligonucleotides, (iv) catalytic RNA and catalytic DNA (DNAzymes), (v) immunostimulatory CpG-oligonucleotides, and (vi) aptamers. It seems to turn out that other subgroups of potential oligomeric drugs such as spiegelmers, intramers, or aptazymes are not equally established yet.

Predominantly antisense oligonucleotides, siRNA, and ribozymes are well suited to suppress the expression of specific genes in target cells and tissues of interest, thereby giving rise to knockdown and even knockout phenotypes of treated cells with respect to the expression of the target gene. This represents a crucial step in complex approaches to study the biological function of unknown genes, probably one of the major bottlenecks of research in the postgenome area. However, knowledge on the biological role of genes identified in genomewide mapping of genetic information is also crucial to the identification of new drug targets and the search for new drugs.

5 Applications *in vivo* and in Man ("Drugs")

5.1 Suppression of Pathological Gene Expression

A number of nucleic acid inhibitors are capable of suppressing gene expression at the posttranscriptional level in mammalian cells, including antisense nucleic acids, siRNA, and catalytic RNA. Those nucleic acids can be subdivided into at least two categories according to their mode of synthesis and delivery: oligomeric and usually synthetic ones on the one hand and long-chain molecules usually expressed endogenously after transfer of recombinant genes into target cells on the other hand. In the latter case, viral or nonviral vectors systems have to be used to deliver the therapeutic gene of interest into target cells *in vivo*. This mode of the application of recombinant genes encoding antisense RNA, siRNA, or ribozymes has to be assigned to the field of gene therapy. Recent drawbacks in this field of biomedical experimental therapy in man revealed the complexity of this kind of therapeutic concepts. Conversely, chemically synthesized oligonucleotides can be regarded as quasi-classical drugs for which there are clear schemes to test their toxicity and relevant pharmacological characteristics before delivering to patients along established lines. Thus, it is not surprising that in the field of antisense nucleic acids, by far the most clinical studies are being conducted in the use of antisense oligonucleotides.

5.2 Antisense Oligonucleotides

In general, the most frequently used chemistry for almost all classes of oligonucleotides that are used as a drug in clinical studies in man, in particular, antisense oligonucleotides, is the phosphorothioate chemistry. Even though this sort of first-generation chemistry has been substantially improved in terms of biochemical and biological parameters *in vitro* by second- and third-generation chemistry, it remains by far the most frequently used chemistry in the most advanced clinical studies. Accumulated experience from more than one decade of *in vivo* studies in the use of phosphorothioate-modified oligonucleotides revealed that side effects are mild and usually typical

for the sulfur modifications rather than specific sequences used. For this reason, it seems that this kind of chemistry will also be used for *in vivo* application in the near future.

In the beginning of clinical studies in the use of antisense oligonucleotides more than one decade ago, one usually addressed severe diseases such as, for example, cancer, chronic inflammation, or infection by highly pathogenic viruses. Most of such clinical trials were not successful and were terminated. If one can learn from those failures, then probably two major messages are important to note: Firstly, in technical terms it seems to be disadvantageous to deliver oligonucleotides systemically *in vivo*. This means that relatively high absolute doses have to be applied in order to achieve the necessary minimal concentrations at the site or tissue of action. This is expensive and for most cell types and tissues also inefficient. Secondly, if one aims to apply antisense oligonucleotides for therapeutic purposes, that is, to suppress the expression of a specific gene that is assumed to play a crucial role for the development of a specific disease, the pathomechanism has to be known. In case of cancer and other complex diseases, this is often not sufficiently the case. Choosing an inappropriate target or not all necessary targets that are related to the pathomechanism underlying a certain disease retrospectively often explains the lack of clinical success.

To avoid technical problems such as those mentioned above, one increasingly changes the way of the delivery of therapeutic oligonucleotides. For example, efficient body distribution and pharmacokinetics can be achieved by delivering formulated antisense oligonucleotides to the skin of patients from which the oligonucleotides are taken up and distributed systemically. Alternatively, local and topical application of antisense oligonucleotides represents a promising mode of delivery since amounts of the drug are minimal, and local concentrations are high.

In perspective to the near and midterm future, it is reasonable to assume that antisense oligonucleotides will remain the major class of inhibitors for drug development. In particular, some of the "second generation" chemistries of oligonucleotides showed very promising activity in cell culture and animal models, indicating the competitive characteristics of antisense drugs versus other classes of oligomeric nucleic acid drugs such as catalytic RNA and DNA and even double-stranded RNA.

5.3 Ribozymes

Initially, that is, more than 10 years ago, the concept of using ribozymes as tools to specifically recognize and hydrolyze a given target RNA and, hence, to suppress the expression of a specific gene was considered as highly elegant and promising. It turned out, however, that a number of criteria were not in favor of the clinical application of this class of potential drugs, including the costs of production and purification, the somewhat moderate biological activity *in vivo* when compared with antisense oligonucleotides, and more promising alternative concepts to deliver ribozymes almost in the field of somatic gene therapy. For example, the endogenous expression of ribozymes after transfer of recombinant genes expressing specific ribozymes cassettes in relevant tissues and cells seems to be promising.

5.4
Small RNAs – siRNA

Currently small RNAs that are formed by two short strands such as conventional siRNA or just a single RNA molecule forming "hairpin siRNA", snRNA, snoRNA, or miRNA represent the most promising class of potential oligonucleotide drugs. In principle, those RNAs can be synthesized by chemical means as well as endogenously by intracellular expression from appropriate templates to be delivered to target cells. In contrast to the application of chemically synthesized catalytic RNA meanwhile, a number of elegant concepts for the endogenous expression have been shown to be effective and to give rise to substantial biological effects.

On a more general level, one has to note that the specificity of such RNA compounds is still under question as the literature describes side effects such as induction of interferon or cosuppression of nontargeted genes controversially.

5.5
CpG-Immunostimulatory Oligonucleotides

Oligonucleotides harboring a 5′-CG-3′ dinucleotide termed *CpG* motif can exert immunostimulatory effects on subsets of immune cells in mammals, which are thought to support and even induce processes that are of therapeutic benefit in case of certain diseases such as cancer. It is attractive to assume that the concentration of CpG-carrying oligonucleotides necessary to induce significant biological effects *in vivo* is relatively low. Side effects seem to be negligible at those applied concentrations, and clinical studies in the use of CpG-immunostimulatory oligonucleotides are promising.

A number of other classes of oligomeric tools and drugs are known. For those, "proof of principle" has been shown *in vitro* or in cell culture; however, it seems that most of them do not play a major role in clinical drug development, which is the major field of interest in biomedicine and molecular medicine. Those classes include the above-mentioned aptamers, intramers, as well as the nuclease-resistant spiegelmers that are also developed as aptamer-like molecules. Further, within the context of diagnostic medicine, it is warranted to mention the imaging of proliferating tissues in man using radiolabeled pharmaceuticals (PET). The attempt to use oligonucleotides in PET is meant to provide a highly sequence-specific step in order to distinguish between malignant and nonmalignant cells.

6
Prospects

Oligonucleotides as drugs and tools in molecular medicine look back to approximately 25 years of research and development. The field of antisense nucleic acids and catalytic RNA has pioneered the application of oligonucleotides as pharmaceuticals. Later, CpG-immunostimulatory oligonucleotides have added to the clinical development of oligonucleotides. In the late 1990s, however, the field of oligomeric nucleic acids seemed to become less dynamic until the discovery of small double- and single-stranded RNAs that exert high biological activity and to represent a powerful class of regulators on the level of gene expression. With the characterization of siRNA and the various groups of small noncoding RNAs preferentially, oligoribonucleotides (RNA) moved into the focus of biomedical research, and it is not

difficult to predict that the field of "oligonucleotides" will have a great benefit and support by those recent findings.

See also Gene Targeting.

Bibliography

Books and Reviews

Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function, *Cell* **116**, 281–297.

Burgstaller, P., Jenne, A., Blind, M. (2002) Aptamers and aptazymes: accelerating small molecule drug discovery, *Curr. Opin. Drug. Discov. Devel.* **5**, 690–700.

Carola, C., Eckstein, F. (1999) Nucleic acid enzymes, *Curr. Opin. Chem. Biol.* **3**, 274–283.

Eckstein, F. (1985) Nucleoside phosphorothioates, *Annu. Rev. Biochem.* **54**, 367–402.

Eckstein, F. (1991) *Oligonucleotides and Analogues – A Practical Approach*, Oxford University Press, Oxford.

Emilsson, G.M., Breaker, R.R. (2002) Deoxyribozymes: new activities and new applications, *Cell. Mol. Life Sci.* **59**, 596–607.

Gewirtz, A.M. (1999) Oligonucleotide therapeutics: clothing the emperor, *Curr. Opin. Mol. Ther.* **1**, 297–306.

Gryaznov, S.M. (1999) Oligonucleotide N3′ → P5′ phosphoramidates as potential therapeutic agents, *Biochim. Biophys. Acta* **1489**, 131–140.

Knudsen, H., Nielsen, P.E. (1996) Antisense properties of duplex- and triplex-forming PNAs, *Nucleic Acids Res.* **24**, 494–500.

Krieg, A.M. (2003) CpG motifs: the active ingredient in bacterial extracts? *Nat. Med.* **9**, 831–835.

Michienzi, A., Rossi, J.J. (2001) Intracellular applications of ribozymes, *Methods Enzymol.* **341**, 581–596.

Micura, R. (2002) Small interfering RNAs and their chemical synthesis, *Angew. Chem. Int. Ed. Engl.* **41**, 2265–2269.

Miyagishi, M., Taira, K. (2004) RNAi expression vectors in mammalian cells, *Methods Mol. Biol.* **252**, 483–492.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H. (1992) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction, *Biotechnology* **24**, 17–27.

Nielsen, P.E. (2000) Antisense peptide nucleic acids, *Curr. Opin. Mol. Ther.* **2**, 282–287.

Opalinska, J.B., Gewirtz, A.M. (2002) Nucleic-acid therapeutics: basic principles and recent applications, *Nat. Rev. Drug. Discov.* **1**, 503–514.

Praseuth, D., Guieysse, A.L., Héléne, C. (1999) Triple helix formation and the antigene strategy for sequence-specific control of gene expression, *Biochim. Biophys. Acta* **1489**(1), 181–206.

Summerton, J. (1999) Morpholino antisense oligomers: the case for an RNase H-independent structural type, *Biochim. Biophys. Acta* **1489**, 141–158.

Sun, J.S., Helene, C. (2003) Oligonucleotides and derivatives as gene-specific control agents, *Nucleosides Nucleotides Nucleic Acids* **22**, 489–505.

Toulme, J.J. (2000) Aptamers: selected oligonucleotides for therapy, *Curr. Opin. Mol. Ther.* **2**, 318–324.

Tuschl, T. (2001) RNA interference and small interfering RNAs, *Chembiochem* **2**(2), 239–245.

Vater, A., Klussmann, S. (2003) Toward third-generation aptamers: Spiegelmers and their therapeutic prospects, *Curr. Opin. Drug Discov. Devel.* **6**, 253–261.

Verma, S., Eckstein, F. (1998) Modified oligonucleotides: synthesis and strategy for users, *Annu. Rev. Biochem.* **67**, 99–134.

Younes, C.K., Boisgard, R., Tavitian, B. (2002) Labelled oligonucleotides as radiopharmaceuticals: pitfalls, problems and perspectives, *Curr. Pharm. Des.* **8**, 1451–1466.

Primary Literature

Beigelman, L., McSwiggen, J.A., Draper, K.G., Gonzalez, C., Jensen, K., Karpeisky, A.M., Modak, A.S., Matulic-Adamic, J., DiRenzo, A.B., Haeberli, P., Sweedler, D., Tracz, D., Grimm, S., Wincott, F.E., Thackray, V.G., Usman, N. (1995) Chemical modification of hammerhead ribozymes. catalytic activity and nuclease resistance, *J. Biol. Chem.* **270**, 25702–25708.

Brown, D.A., Kang, S.H., Gryaznov, S.M., DeDionisio, L., Heidenreich, O., Sullivan, S.,

Xu, X., Nerenberg, M.I. (1994) Effect of phosphorothioate modification of oligodeoxynucleotides on specific protein binding, *J. Biol. Chem.* **269**, 26801–26805.

Carmo-Fonseca, M., Pepperkok, R., Sproat, B.S., Ansorge, W., Swanson, M.S., Lamond, A.I. (1991) In vivo detection of snRNP-rich organelles in the nuclei of mammalian cells, *EMBO J.* **10**, 1863–1873.

Dueholm, K.L., Egholm, M., Behrens, C., Christensen, L., Hansen, H.F., Vulpius, T., Petersen, K.H., Berg, R.H., Nielsen, P.E., Buchardt, O. (1994) Synthesis of peptide nucleic acid monomers containing the four natural nucleobases: thymine, cytosine, adenine, and guanine and their oligomerization, *J. Org. Chem.* **59**, 5767–5773.

Famulok, M., Blind, M., Mayer, G. (2001) Intramers as promising new tools in functional proteomics, *Chem. Biol.* **8**, 931–939.

Heidenreich, O., Gryaznov, S., Nerenberg, M. (1997) RNase H-independent antisense activity of oligonucleotide N3′ → P5′ phosphoramidates, *Nucleic Acids Res.* **25**, 776–780.

Miller, P.S., Reddy, M.P., Murakami, A., Blake, K.R., Lin, S.B., Agris, C.H. (1986) Solid-phase syntheses of oligodeoxyribonucleoside methylphosphonates, *Biochemistry* **25**, 5092–5097.

Nedbal, W., Tomakidi, P., Lehmann, M., Dörfer, C., Kohl, A., Sczakiel, G. (2002) Antisense-mediated inhibition of ICAM-1 expression: a therapeutic strategy against inflammation of human gingival tissue, *Antisense Nucleic Acid Drug Dev.* **12**, 71–78.

Pitsch, S., Weiss, P.A., Jenny, L., Stutz, A., Wu, X. (2001) Reliable chemical synthesis of oligoribonucleotides (RNA) with 2′-O-[(triisopropylsilyl)oxy]methyl(2′-O-tom)-protected phosphoramidites, *Helv. Chim. Acta* **84**, 3773–3795.

Robertson, M.P., Knudsen, S.M., Ellington, A.D. (2004) In vitro selection of ribozymes dependent on peptides for activity, *RNA* **10**, 114–127.

Sanger, F., Coulson, A.R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase, *J. Mol. Biol.* **94**, 441–448.

Scaringe, S.A., Francklyn, C., Usman, N. (1990) Chemical synthesis of biologically active oligoribonucleotides using beta-cyanoethyl protected ribonucleoside phosphoramidites, *Nucleic Acids Res.* **18**, 5433–5441.

Scaringe, S.A., Wincott, F.E., Caruthers, M.H. (1998) Novel RNA synthesis method using 5′-O-Silyl-2′-O-orthoester protecting groups, *J. Am. Chem. Soc.* **120**, 11820–11821.

Wagner, R.W., Matteucci, M.D., Lewis, J.G., Gutierrez, A.J., Moulds, C., Froehler, B.C. (1993) Antisense gene inhibition by oligonucleotides containing C-5 propyne pyrimidines, *Science* **260**, 1510–1513.

Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S., Usman, N. (1995) Synthesis, deprotection, analysis and purification of RNA and ribozymes, *Nucleic Acids Res.* **23**, 2677–2684.

Oncogenes

Zhong-Zong Pan and Andrew K. Godwin
Fox Chase Cancer Center, Philadelphia, PA

1 Introduction 437

2 Oncogenes and Molecular Mechanism of Cancer 438
2.1 Historical Perspective of Oncogenes 438
2.2 Molecular Mechanisms of Cancer 441
2.3 Models of Molecular Oncogenesis: the Two-hit Hypothesis and Multistep Tumorigenesis 445
2.4 Major Categories of Oncogenes 446
2.5 Mechanisms of Oncogene Perturbation 448
2.5.1 Viral Oncogenes 448
2.5.2 Cellular Oncogene Activation by Viral Integration 449
2.5.3 Chromosome Translocation 449
2.5.4 Gene Amplification 449
2.5.5 Genetic Sequence Mutation 449
2.5.6 Aberrant Gene Expression 449
2.5.7 Alterations Not at the Genetic Levels 449

3 Biochemical Properties and Cellular Functions of Oncogenes 450
3.1 Oncogenes Promoting Cell Cycle Progression and Cell Division 451
3.1.1 Oncogenes Derived from Cell Cycle Machinery 451
3.1.2 Oncoproteins Leading to Loss of Control of Cell Cycle Checkpoints 453
3.2 Growth Factors and Extracellular Mitogenic Signals 453
3.2.1 PDGF (Platelet-derived Growth Factor) and the *SIS* Oncogene 455
3.2.2 FGFs (Fibroblast Growth Factors) 455
3.2.3 The WNT Gene Family 455
3.2.4 EGF (Epidermal Growth Factor) and TGF-α 455
3.2.5 Hematopoietic Growth Factors 455
3.3 Oncogenes Derived from Cell Surface Receptors 456
3.3.1 Growth Factor Receptors 456

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

3.3.2 Cell Surface Receptors without Tyrosine Kinase Activity 461
3.3.3 Membrane-associated Nonreceptor Oncogenes 461
3.4 Oncogenes in Cytoplasmic Signal Transduction Pathways 462
3.4.1. RAS-RAF-MEK-MAPK (ERK) Pathway 462
3.4.2 The PI3K-AKT Pathway 464
3.4.3 Guanine Nucleotide Binding Proteins: G-Proteins 466
3.4.4 The JAK-STAT Signaling Pathway 468
3.4.5 Other Protein Serine-threonine Kinases 469
3.4.6 Antiapoptotic Proteins 469
3.4.7 Chaperone Proteins 470
3.4.8 Oncogenes Associated with Cancer Cell Adhesion and Motility 470
3.5 Transcription Factors and Nuclear Oncogenes 471
3.5.1 *FOS* and *JUN* Families 471
3.5.2 c-*MYC* 471
3.5.3 *REL* Oncogene: the NFκB Family Proteins 471
3.5.4 Estrogen Receptor and Other Steroid Hormone Receptors 472
3.5.5 Homeodomain-containing Proteins 472
3.5.6 *MYB* Family 472
3.5.7 *ETS* Gene Family 472

4 Identification of Oncogenes 473

5 Oncogenes and Clinical Implications 473
5.1 Oncogenes and Cancer Diagnosis and Prognosis 473
5.2 Oncogenes and Cancer Therapy 473
5.2.1 Oncogene-Targeted Chemotherapy 473
5.2.2 Oncogene-targeted Immunotherapy 474

6 Perspectives 476

Bibliography 476
Books and Reviews 476
Primary Literature 476

Keywords

Oncogene
Broadly defined, an oncogene may be considered a cellular or viral gene whose product in altered form and/or greater amount can cause malignant transformation. Oncogenes function in a dominant manner because they can transform cells despite expression of their normal counterparts. The classical oncogenes are more strictly

defined by functional analysis, that is, a gene that when introduced into a cell, typically NIH3T3 cells, can transform the cell with some attributes of malignancy. More recently, it has been shown that a combination of oncogenes (e.g. RAS family and MYC family) are required to induce one or more characteristics of neoplastic transformation when introduced into "truly normal" cells, such as primary mouse embryo fibroblasts (MEFs).

Oncoproteins
Oncoproteins are the protein products encoded by oncogenes.

Proto-oncogene
Proto-oncogenes are normal cellular genes that can be converted to active oncogenes.

Tumor suppressor gene
Tumor suppressors are those genes that encode for proteins that prevent normal cells from becoming cancer cells, and when their functions are lost, contribute to the genesis of cancer.

Oncogenes are cellular or viral genes whose products in altered form and/or greater amount can cause malignant transformation. The hallmarks of cancer cells include loss of control of cell growth and proliferation, loss of control of cell survival and apoptosis, metastasis, and gain of immortality. Oncoproteins, the protein products encoded by oncogenes, could disrupt the normal regulation of these processes at many levels in cell signaling and gene expression. Oncogenes could be divided into these major categories: growth factors and extracellular mitogenic signals, cell surface receptors, proteins associated with inner surface of the plasma membrane, cytoplasmic proteins in various signaling pathways, proteins associated with mitochondria, nuclear proteins, and transcription factors. Instead of being caused by a single gene, malignant transformation from a normal cell to cancer cell is the end-point phenomenon resulting from accumulated genetic and epigenetic alterations. The appropriate combinations or order of altered oncogenes and tumor suppressors are required for different stages of tumor progression. Clinically, oncogenes are important diagnosis biomarkers and therapy targets for cancer.

1 Introduction

Cancer cells arise as a result of a process known as malignant transformation. Over the course of a normal cell's life span, it might acquire multiple genetic and epigenetic changes. Some of these changes are directly related to malignant transformation, while most are not. However, in the rare cell in which the combination of changes and environmental conditions are just right, that once normal cell can become cancerous, displaying

phenotypic characteristics, which include loss of control of cell growth and proliferation, loss of control of cell survival and apoptosis, and metastasis (Fig. 1). In the human body, a variety of different kinds of cells constitute different tissues and organs, and these cells are maintained at homeostasis and restricted at the proper positions of the tissues or organs. The human genome is composed of about 2.9 billion nucleotide base pairs and about 30 000 to 40 000 protein encoding gene loci, cells of a particular cell type display a certain profile of expressed genes to fulfill their physiological functions as well as to regulate cell growth, proliferation, senescence, and apoptosis. Oncogenes and tumorsuppressor genes encode for two families of proteins that, when aberrantly expressed, are thought to contribute greatly to cancer. Tumorigenesis is a complex multistep process, including tumor initiation, tumor progression to form tumor mass (escape of many constraints), and metastasis. In addition to the conversion of a normal cell to tumor cell, changes in substratum, angiogenesis, escape from immune surveillance, metastatic spread, and seeding and survival in the secondary tumor sites are all critical steps for tumor cells to escape the normal regulations and become malignant. In this review, we will focus on the oncogenic genes in the cellular transformation process. First we will discuss oncogenes and molecular mechanism of cancer, followed by a description of the biochemical properties and cellular functions of several well-studied oncogenes. Then we will discuss identification of oncogenes, followed by the clinical implication of oncogenes in cancer diagnosis, prognosis, and therapy.

2 Oncogenes and Molecular Mechanism of Cancer

2.1 Historical Perspective of Oncogenes

Malignant transformation from a normal cell to cancer cell is the end-point phenomenon resulting from accumulated genetic and epigenetic alterations in the cell. The belief that genetic damage might be responsible for cancer arose from a variety of sources: the recognition of hereditary predisposition to cancer, the presence of damaged chromosomes in at least some cancer cells, the connection between susceptibility to cancer and impaired ability of cells to repair damaged DNA, and

Fig. 1 Malignant transformation from a normal cell to become cancer cells and formation of tumor mass. In normal tissue or organ, any particular cell type is maintained in the range of a certain number to maintain tissue homeostasis. A normal cell becomes a cancer cell when the cell contains dysfunctional oncogene(s), tumor suppressor gene(s), and those involved in telomere maintenance. Cancer cells display phenotypes of loss of control in cell growth and proliferation, loss of control in cell survival, or loss of control of cell senescence, and gain of immortality. Senescent cells or cells with gene mutation(s) are monitored by the cellular intrinsic checking programs and also by the body systemic immune surveillance mechanisms, and will be cleared in most cases by these mechanisms. Only under the circumstances that the cancer cell escapes these surveillance mechanisms will the cell continue to proliferate to form tumor mass (primary tumor) in the body. This could be a very long process, sometimes in decades. With time, other genetic mutations accumulate in the cancer cells and some of them gain the potential for metastasis.

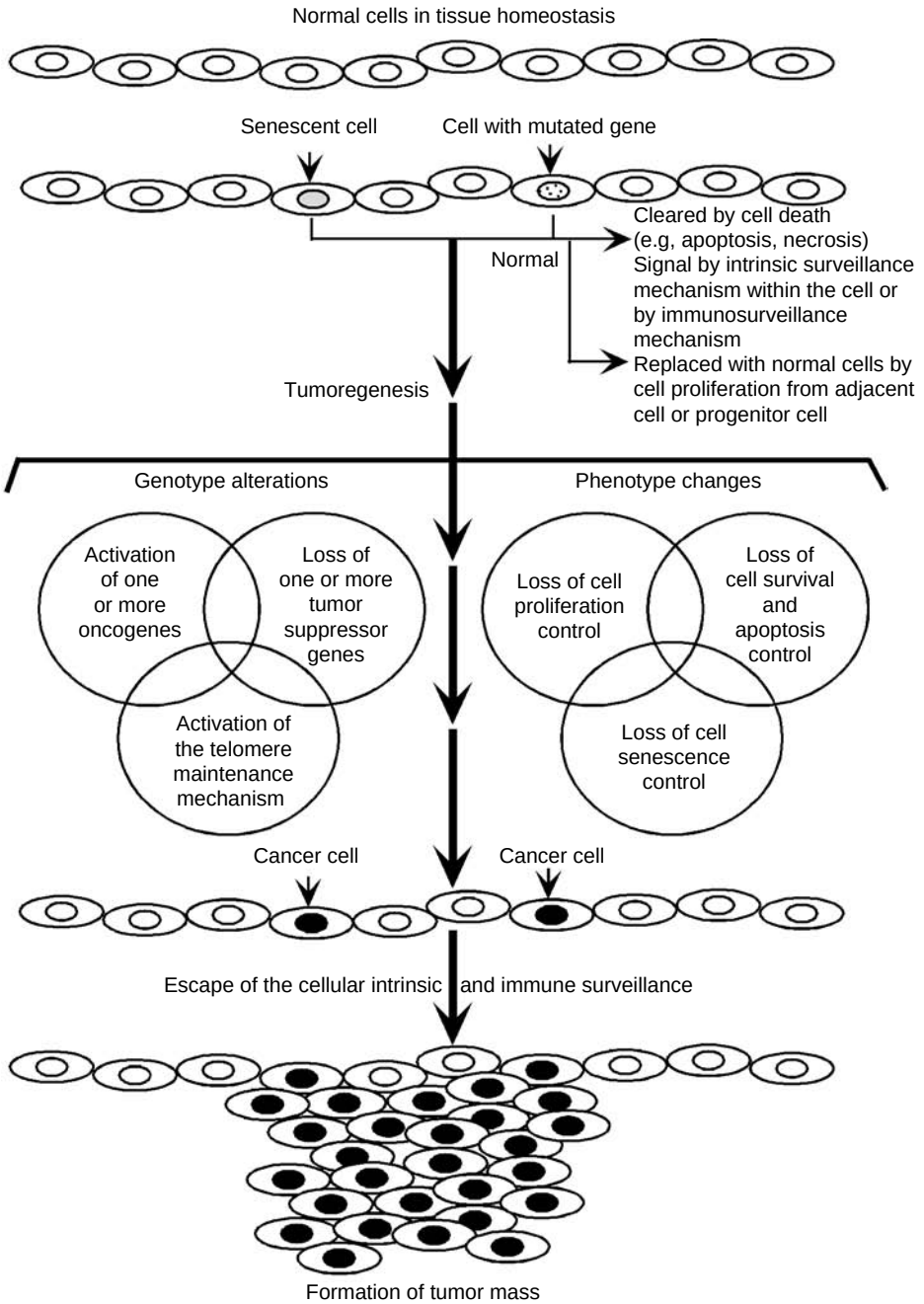
Normal cells in tissue homeostasis
Senescent cell
Cell with mutated gene
Normal
Cleared by cell death (e.g, apoptosis, necrosis)
Signal by intrinsic surveillance mechanism within the cell or by immunosurveillance mechanism
Replaced with normal cells by cell proliferation from adjacent cell or progenitor cell
Tumoregenesis
Genotype alterations
Phenotype changes
Activation of one or more oncogenes
Loss of one or more tumor suppressor genes
Activation of the telomere maintenance mechanism
Loss of cell proliferation control
Loss of cell survival and apoptosis control
Loss of cell senescence control
Cancer cell
Cancer cell
Escape of the cellular intrinsic and immune surveillance
Formation of tumor mass

evidence that relates the mutagenic potential of substance to their carcinogenicity. Nowadays we know that the vast majority of human cancers are caused by genetic or epigenetic changes in the genome of the cancer cells and only a very small fraction of cancer cases, such as cervical and liver, may be attributed to viral infection. However, the work on retroviruses in the earlier days of cancer research did open the door for the discovery of oncogenes. The first identified oncogene, *Src,* discovered in 1976 by the Nobel Laureate Harold Vamus and J. Michael Bishop, is the archetype of those viral oncogenes. It had been known since 1909 that Rous sarcoma virus (RSV) infection of chicken cells lead to tumors, but the *v-Src* gene as the cause of the malignant transformation was not identified until the 1970s. Studies by Varmus and Bishop further demonstrated that the *Src* gene was originally a cellular gene instead of a viral gene, but was picked up and integrated into the viral genome by an ancestor of the RSV (Fig. 2). Normal cellular genes that can be converted to active oncogenes are therefore referred to as proto-oncogenes. Other such viral oncogenes include *MYC* (avian myelocytomatosis virus), *RAS* (rat sarcoma virus), *ERBB1* (avian erythroblastosis virus), *ABL* (Abelson murine leukemia virus), as well as others.

In addition to virus-mediated proto-oncogene activation, many other mechanisms have been identified in cancer cells that can lead to activation of proto-oncogenes. At the genetic level, each gene

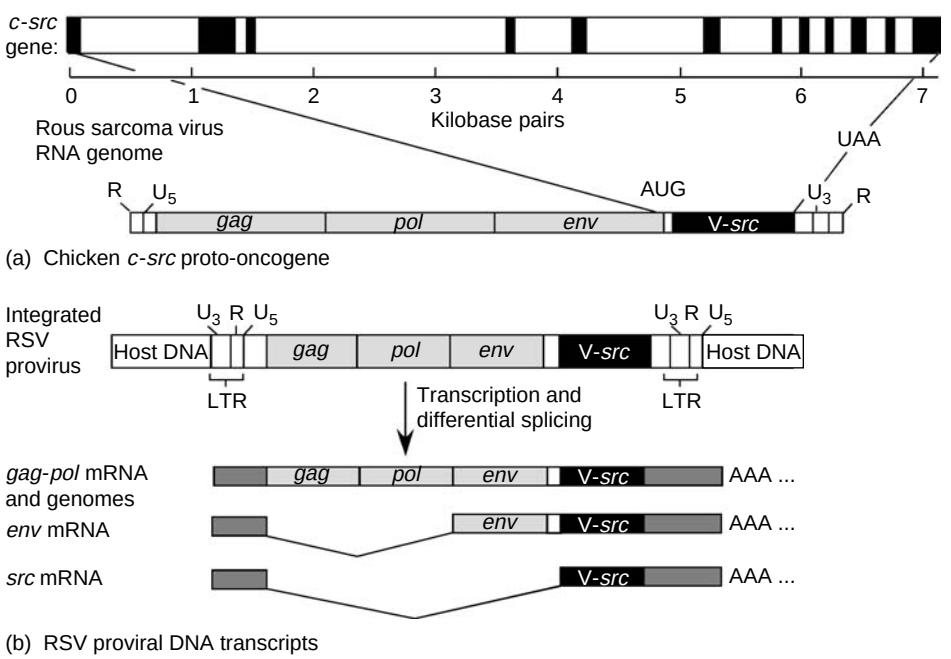

Fig. 2 The *c-SRC* proto-oncogene and the *v-SRC* oncogene. The cellular proto-oncogene *c-SRC* was picked up and integrated into the RSV genome as *v-SRC* (a), which encodes the v-SRC oncoprotein responsible for the malignant transformation of RSV (b).

has to be transcribed at the appropriate level within a certain time frame, and then translated into a functioning protein. Normally, the proto-oncogenes are latent or strictly regulated to perform their normal physiological functions in the cell life. The mechanisms involved in abnormal activation of proto-oncogenes include the following: virus-mediated, chromosome translocation, gene amplification, small deletions or insertions, point mutations, and aberrant gene expression caused by epigenetic alterations such as DNA methylation and histone deacetylation. In some cases, the activity of oncoproteins may be increased by other proteins regulating their stability or degradation. An example is that HSP 90 (heat-shock protein 90) may increase the stability of several oncoproteins such as HER-2, AKT, and KIT.

While many genes might be found altered in cancer cells and most of them are not related to malignancy, the alterations in oncogenes and tumor suppressor genes are most directly responsible for malignant transformation. The number of oncogenes and tumor suppressors identified to date is controversial and is dependent on the rigor of the definition. The classical oncogenes are usually defined by functional analysis, that is, an oncogene is a gene that when introduced into the cell can transform the cell with some attributes of malignancy. With the better understanding of the molecular mechanism of tumorigenesis, we know that the effect of a particular oncogene is dependent on the cellular context. In most cases of tumors, alterations in other oncogenes or tumor suppressor genes are required for a particular oncogene to exert its role in tumor pathogenesis. Therefore, an oncogene will be defined as a cellular or viral gene in its wild type or mutant form that can induce one or more characteristics of neoplastic transformation when introduced, either alone or in combination with another gene, into an appropriate cell type. Oncogenes function in a dominant manner because they transform cells despite expression of their normal counterparts. However, there are examples in which inherited mutations in oncogenes such as KIT, RET, and MET can predispose to cancer. In broad terms, tumor suppressors are those genes that encode for proteins that prevent normal cells from becoming cancer cells, and the loss of their functions contributes to the genesis of cancer.

2.2 Molecular Mechanisms of Cancer

There are two major features of cancer cells: one is the loss of control of cell growth and proliferation, resulting in loss of homeostasis and formation of tumor mass and the other is metastasis for tumor cells to migrate away from their normal location and invade into another tissue or organ. To maintain tissue homeostasis, a normal cell is strictly regulated for cell differentiation, cell growth and proliferation, cell survival and apoptosis, and cell senescence. Deregulation of proto-oncogenes and tumor suppressor genes might contribute to tumorigenesis by disrupting one or more of these processes (Fig. 1). In the process of tumorigenesis, cancer cells can acquire metastatic potential, that is, they migrate into the blood or lymphatic circulatory systems, circulate in the peripheral blood and arrest in a distant organ, extravasate from the vessels, dwell in the distant organ and form tumor masses. Oncogenes may promote metastasis by enhancing cancer cell motility and invasion, and by

promoting cancer cell survival and inhibiting apoptosis under adverse conditions in the process of metastasis. Changes in the environment such as extracellular matrix and angiogenesis are involved in the formation of primary tumors as well as the metastases; oncogenes are also involved directly or indirectly in these processes.

Cell growth and proliferation are regulated through cell cycles, and oncogenes could be derived from those proto-oncogenes that promote cell cycle progression or those that release the inhibitory functions of the negative regulatory molecules of cell cycle. The cell cycle for somatic cell division can be divided into four phases: the G_1 gap phase for preparation of DNA synthesis, the S-phase for DNA replication, the G_2 gap phase for preparation of mitosis, and the M-phase for separation of daughter chromosomes and completion of one cell cycle. The key proteins involved in cell cycle progression are highly conserved among different cell types in higher eukaryocytes, these include cyclins, CDKs (cyclin dependent kinases), CKIs (CDK inhibitors), RB (retinoblastoma protein), and the transcription factor E2F (E1A-inducible factor) (Fig. 3). CDKs are the engine proteins driving cell cycle progression, with CDKs-4, -6, and -2 working in G_1-phase, CDK2 working in S-phase, and CDK1 (CDC2) in G_2- and M-phases. Activation of the CDKs is positively regulated by cyclin binding, and synthesis of cyclins is strictly regulated during different phases of the cell cycle. G_0 and G_1 phases are the interface for the interaction between cell cycle machinery and the extracellular signals, including both mitogenic and inhibitory signals, cues from extracellular matrix, and cell–cell interactions. D-type cyclins are very unstable proteins with very short half-life, and their production is controlled at transcriptional level and regulated by mitogenic signaling. For example, cyclin D1 promoter contains binding sites for AP1, ER-E2/X, SRE, E2F4, E2F1, SP1, CRE, MAPK-c-ETS2-ETS. Synthesis of D-type cyclins awakes quiescent cells (G_0) to enter the G_1-phase of the cell cycle. In late G_1-phase of the cell cycle, there is a restriction point and continued mitogenic stimuli are required for the cycling cells to pass this point (Fig. 3). Once beyond this point, cell cycle progression becomes autonomous and cell cycle will be completed to give two daughter cells unless stalled by the checkpoint mechanisms. The main target of CDKs is the retinoblastoma protein encoded by tumor suppressor gene *RB*. In quiescent (G_0) cells, RB is hypophosphorylated and binds to E2F. The activated CDK4/6 then phosphorylates RB, resulting in dissociation of E2Fs from RB. E2Fs are transcription factors and their target genes include those essential for DNA synthesis in S-phase as well as cyclins (D, E, and A). Cyclins D and E activate CDK4/6 and CDK2 respectively, which further promote G_1-phase progression to pass the Restriction point. Cyclin A activates CDK2 to promote S-phase progression, and G_2- and M-phases are promoted by cyclin B/CDK1 and cyclin A/CDK1 respectively. Cyclin E/CDK2 peaks at the G_1–S transition, and is thought to help the cell transverse the restriction point. Cyclin E is rapidly degraded in S-phase, and cyclin A binds to CDK2 in S-phase for S–M phase transition.

Derangement of cell survival and apoptosis is another important feature contributing to tumor mass formation and tumor progression. Damaged or senescent cells and cells receiving death signals from the environment must be cleared to maintain the normal function and normal

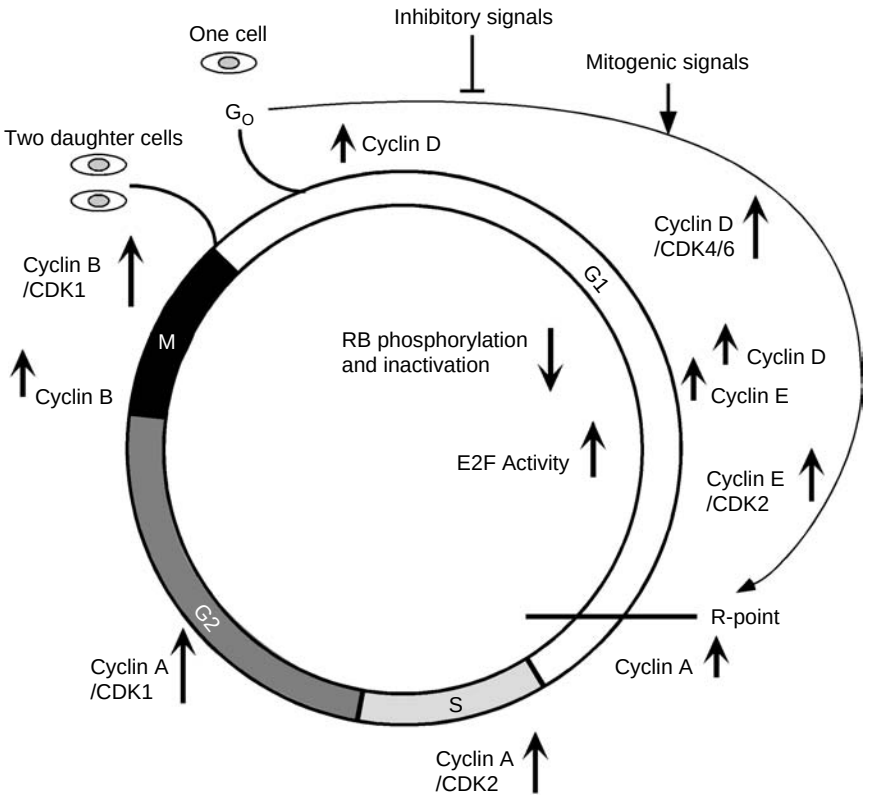

Fig. 3 Cell cycle and the key regulatory molecules. In somatic mitotic cell division, one cell produces two daughter cells after each cell cycle. Most of the differentiated cells in the body are in quiescent state (G_0), and they enter the cell cycle under the stimulation of mitogenic signals such as growth factors. The cell cycle is driven by the CDKs with different CDKs acting in different phases of the cell cycle, and CDKs are activated by cyclins. The target genes of E2F and the target genes regulated by cell signaling cascades initiated by the mitogenic signals are required for the progression in G_1-phase. Once beyond the R-point (restriction point) in late G_1-phase, cell cycle progression becomes autonomous, and no further mitogenic signals are required. E2F is negatively regulated by RB, and RB can be inactivated by CDK-mediated phosphorylation.

life span of a cell, and various oncogenes are known to block the execution of cell death pathways and/or activate the cell survival pathways to promote tumorigenicity (Fig. 1). There are two major pathways for the initiation of apoptosis, one is initiated from the death receptors (such as FAS and TNF receptors) on the plasma membrane and is largely independent of mitochondria; the other is mitochondria-mediated apoptotic pathway initiated by environmental stress signals or by DNA damage signals from the nucleus (Fig. 4). Under certain circumstances, mitochondria might also be involved in death receptor–mediated apoptotic pathway. Caspase-8 is the initiator caspase activated by death receptor, while caspase-9 is the initiator caspase activated by cytochrome C released from mitochondria.

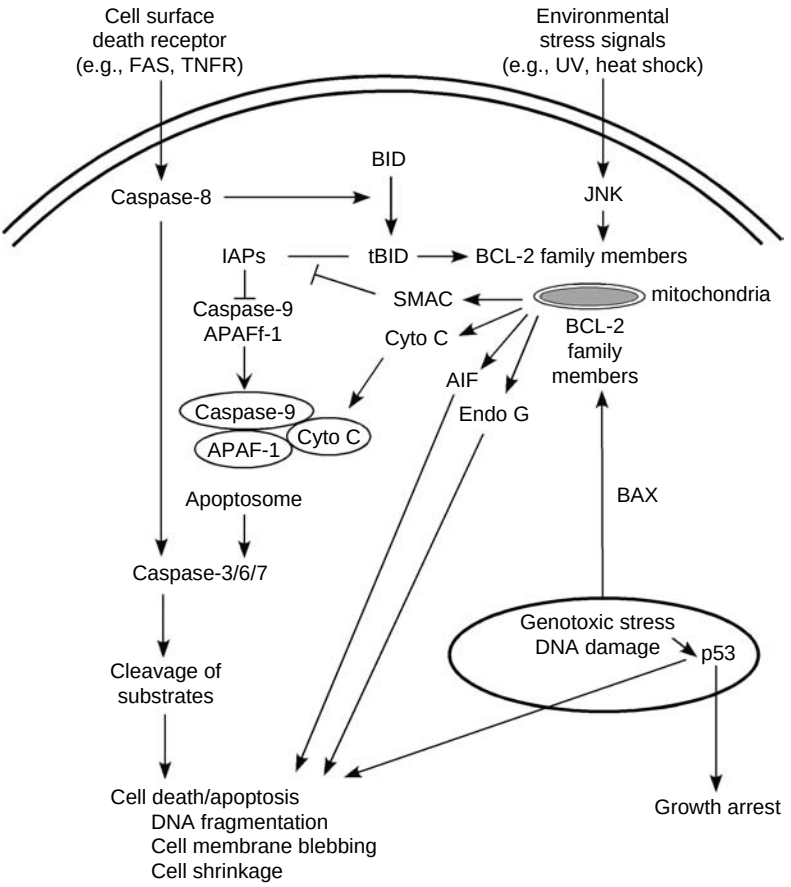

Fig. 4 Regulation of cell apoptosis. There are two major cellular apoptotic pathways, one is initiated from the cell surface in response to activation of the cell surface death receptor(s), and the other is mediated by dysfunctional mitochondria as a response to stress stimulations or nuclear chromatin DNA damage. In both pathways, caspases are involved, and the effector caspases for both pathways are caspases-3, -6, and -7. Many cytoplasmic and nuclear proteins are the substrates of the effector caspases, and cleavage of these substrates leads to apoptosis with cell membrane blebbing, cell shrinkage, nuclear chromatin DNA fragmentation, and so on.

The two pathways merge at the effector caspases-3, -6, and -7. Protein cleavage by effector caspases eventually leads to apoptosis, including cell membrane blebbing, chromosome DNA fragmentation, and cell shrinkage. In mitochondria-mediated apoptotic pathway, the BCL-2 (B-cell lymphoma/leukemia-2) family members play the critical role as gatekeeper in regulating the release of cytochrome C, SMAC (the second mitochondria-derived activator of caspase), and HTRA2 (high-temperature requirement factor). Cytochrome C together with APAF-1 bind to caspase-9 to form apoptosome, leading to activation of caspase-9, and SMAC

and HTRA2 bind IAPs and release their inhibition on effector caspases.

In the progression of tumorigenicity, cancer cells also acquire cell immortality. For differentiated somatic cells, the number or times a cell can proliferate is extremely limited because the length of telomere becomes shorter after each cell division, and eventually the telomere will become too short to permit further cell division. For cancer cells to continue to proliferate and form tumor mass or metastasis, some mechanism has to be adapted to overcome telomere shortening during each cell cycle. The mechanisms involved in telomere maintenance could be mediated either by increased telomerase activity, or by a process referred to as *alternative lengthening of telomeres* (ALT).

2.3 Models of Molecular Oncogenesis: the Two-hit Hypothesis and Multistep Tumorigenesis

One of the earliest models regarding molecular oncogenesis is the "two-hit hypothesis" proposed by Knudson on the basis of their studies of hereditary cancer. According to this model, for a cell carrying a germline mutation in one allele of tumor suppressor gene (the first hit) to become a cancer cell, the cell needs to acquire a second hit (somatic mutation in the remaining wild-type allele). This hypothesis led to the cloning of the first tumor suppressor gene, *RB*. Today we have a much better understanding of tumorigenesis; however, the two-hit hypothesis is still valid, although it is continually modified to reflect new knowledge. For example, it is clear that multiple genetic and epigenetic alterations must take place to cause a cell to become cancerous. Therefore, we can think in broader terms and consider the activation of an oncogene as the first hit, and the loss of tumor suppressor activity as the second hit, or vice versa. In addition to oncogenes and tumor suppressor genes, the telomere maintenance mechanism is required for the formation of tumor mass and might represent a distinct kind of hit on oncogenes.

As indicated above, tumorigenesis is a very complicated process consisting of multiple steps, including tumor initiation, tumor progression, and metastasis. The hit on different oncogenes or tumor suppressor genes might be required at different steps as exemplified in colon cancer (Fig. 5). The initiation and early stage of tumor progression is very slow, sometimes it takes years or decades for a malignant cell to form noticeable tumor mass (i.e. 1 cm diameter, or 1000 million cells). One major reason is that not all cells with active oncogene(s) will become cancerous, and not all cancer cells will escape the various surveillance mechanisms within the cell and in the body and survive to form tumor mass. In most carcinomas, cancer is not a disease caused by a single gene. It is more like a syndrome; within the cancer cells, the appropriate combinations or order of altered genes are required for different stages of tumor progression (Fig. 5), and, within the body, the immune surveillance needs to be either defective or compromised. For example, loss of *APC* (adenomatous polyposis coli) gene function is required for tumor initiation, while mutation of *RAS* alone or *p53* alone is not sufficient for tumor initiation, although both are required for tumor progression. For most, if not all, important cellular processes, the cell is regulated by the balance between positive and negative regulatory mechanisms at multiple levels. Therefore, the same kind of hit might occur at different steps of

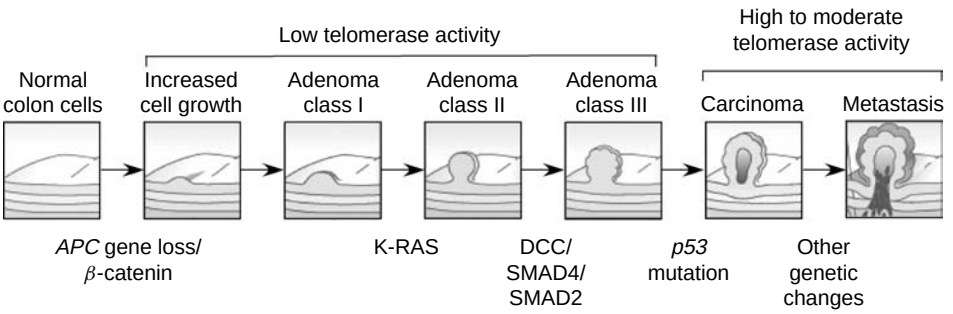

Fig. 5 Model of tumorigenesis: multistep carcinogenesis and the two-hit hypothesis. Colon cancer progression is a multistep process, involved with tumor initiation, tumor progression, and metastasis. The *K-RAS* oncogene and several tumor suppressor genes including *APC* and *p53* are involved in different stages of tumor progression. In tumor progression and metastasis, increased telomerase activity is involved in the maintenance of telomeres of the cancer cells.

tumorigenesis. For oncogenic transformation, the right combination of two or more oncogenes, such as *RAS* and *MYC*, might be considered as one kind of hit at different levels. In addition to RAS, telomerase activity is also a requirement for tumor progression and metastasis in colon cancer. While telomerase activity is not detectable in normal colon tissue, telomerase activity is low in polyps or adenomas and high to moderate in carcinomas and metastasis. For tumor suppressor genes, the serial loss of tumor suppressor genes *APC* and *p53* in hereditary colon cancer might be considered as another kind of hit at different stages of tumorigenesis. A similar situation is also found in lung cancer, that is, multiple oncogenes and tumor suppressor genes are involved. Finally, the kind of hit in some cases might function as a different hit in other circumstances, and some genes might be involved in more than one kind of hit. For example, MYC could act as either an oncogene to promote cell proliferation or a tumor suppressor to sense cells to apoptosis, depending on the particular circumstances, and the tumor suppressor gene p53 might also be involved in telomere maintenance (an event considered as oncogenic).

2.4 Major Categories of Oncogenes

Oncogenes can be classified into different categories based on different criteria. According to their origin, oncogenes could be divided into viral oncogenes and cellular oncogenes. Oncogenic viruses fall into two categories, the DNA viruses and retroviruses. As described above (see Sec. 2.1), retroviral oncogenes were originally derived from cellular proto-oncogenes. The oncogenic DNA viral oncogenes include Simian virus 40 (SV40) large T-antigen, polyomavirus middle T- and large T- antigens, papillomavirus E6 and E7 proteins, adenoviruses E1A and E1B proteins, herpesviruses such as Epstein Barr virus (EBV) LMP1 and other viral proteins, hepatitis B virus (HBV) X gene-encoded protein.

Cellular oncogenes can be further divided into different categories based on their subcellular locations, protein

structures, and their functions in cellular physiological processes. In this review, we will integrate these criteria and divide the oncogenes into these major categories: (1) Oncogenes in cell cycle progression and cell division; (2) growth factors and extracellular mitogenic signals; (3) oncogenic cell surface receptors; those include receptor tyrosine kinases such as mitogenic growth factor receptors, as well as non-tyrosine kinase receptors such as G-protein-coupled seven-transmembrane receptors and receptors for cytokines; (4) membrane-associated nonreceptor tyrosine kinases, adaptor proteins, and phospholipases; (5) oncogenes in cytoplasmic signal transduction pathways, including RAS-RAF-MEK-MAPK pathway, the PI3K-AKT pathway, guanine nucleotide–binding proteins (G-proteins), PKC (Protein kinase C) and protein serine-threonine kinases, BCL-2 family members and the antiapoptotic regulation; (6) transcription factors and nuclear oncogenes. A list of putative oncogenes is summarized in Table 1 (Table 1).

Tab. 1 List of oncogenes.

Cell cycle progression	
Cyclins	D-type cyclins
CDKs	*CDK4*
CDK activators	*CAK, CDC25A, CDC25B*
E2F	
Telomere Maintenance	*HTRT*
DNA viral oncogene	SV40 large T, polyomavirus middle T and large T, papillomavirus E6 and E7 genes, adenoviruses E1A and E1B genes
Growth factors and extracellular mitogenic signals	
PDGFs	*SIS/PDGF*β
FGFs	*FGF-3/INT-2, FGF4/HST, FGF-5, FGF-6*
WNT	*WNT-1, WNT-2, WNT-3*
EGFs	*EGF, TGF-*α
Cytokines	*IL-2, IL-3, GM-CSF, CSF-1*
Cell surface receptors	
Growth factor receptors	
EGF receptor family	*EGFR, ERBB2 (HER-2/neu)*
FGF receptor family	
PDGF receptor family	*PDGF*α and β receptors, *KIT*, *CSF*-1 receptor
Insulin receptor family	*IGF-IR*
MET/HGF receptor family	*MET*, *RON*, *c-SEA*
Neurotrophin receptor family	*TRKA*, *TRKB*, *TRKC*
VEGF receptor family	
Non-tyrosine kinase receptors	
Cytokine receptors	Receptors for IL-2, IL-3, GM-CSF, and erythropoietin
G-protein-coupled 7-TMR	*MAS*, thyrotropin receptor, serotonin 1c receptor
Plasma membrane–associated nonreceptor oncogenes	
Tyrosine kinases	
SRC family	*SRC*, *YES*, *FGR*, *LCK*, *LYN/SYN*, *FYN*, *HCK*, *BLK*, *YRK*, *LCK/TKI*
ABL family	*ABL*, *ARG*, *ABL/BCR*
FES family	*FES*, *FER*
JAK family	*JAK1*, *JAK2*, *TYK2*

(*continued overleaf*)

Tab. 1 (*Continued*)

ITK family	*ITK*, *BPK*, *TEC*	
SYK family	*SYK*, *ZAP*	
CSK family	*CSK/CYL*, *CTK*,	
Others	*FAK*, *BRK*, *FPS/FES/FER/TYK3*, *TKF*	
Adaptor proteins	*GRB2*, *SHC*, *CRK*, *NCK*	
Phospholipases	*PLC-γ*	
Oncogenes in cytoplasmic signal transduction		
RAS-RAF-MAPK pathway	*RAS* family	*H-RAS*, *K-RAS*, *N-RAS*, *R-RAS*
	RAF family	*RAF-1*, *A-RAF*, *B-RAF*
	MAPKs	*ERK1*, *ERK2*
PI3K-AKT pathway	PI3Ks	*PI3K/p85α*, *PI3K/p110α*
	AKT family	*AKT1*, *AKT2*
G-protein pathway		
	Heterotrimeric G-proteins	Gαi2, Gα12, Gα13, Gαz, Gαq, Gs
	GEFs	*SDC25*, *OST2*
	Phospholipases	*PLC-β*, *PLC-α*
	RAS family	See above
	RHO/RAC binding proteins	*BCR*, *VAV*, *DBL*, *ET2*, *NET1*, *TIAM1*, *TIM*
JAK-STAT pathway		
	JAK family	See above
	STAT family	*STAT3*, *STAT5*
Other protein serine-threonine kinases		
	PKC family	*PKC-ε*, *PKC-β1*, *PKC-γ*, *PKC-ζ*
Anti-apoptotic oncogenes		
	Bcl-2 family	*BCL-2*, *BCL-XL*,
	IAP family	c*IAP*1, c*IAP*2, *XIAP*, survivin, *LGALS*3
Chaperone proteins		*HSP*90
Nuclear transcription factors		
FOS family	*c-FOS*, *FOS-B*, *FRA-1*, *FRA-2*	
JUN family	c-*JUN*, *JUN*-B, *JUN*-D	
MYC family	c-*MYC*, N-*MYC*, L-*MYC*	
REL/NFκB family	*NFκB*1, *NFκB*2, *REL*, *REL*A, *REL*B	
Steroid hormone receptors	*ER*	
Homeodomain-containing oncogenes	*HOX*A9, *HOX*2.4, *HOX*7.1, *HOX*11,	
ETS family	*ETS*, *ELK*-1	
Others	*MYB* (*MBM*2), *β*-catenin	

2.5 Mechanisms of Oncogene Perturbation

2.5.1 Viral Oncogenes

Retroviral oncogenes are derived from cellular proto-oncogenes and their activation and function are similar to their cellular oncogene counterparts in the infected host cells. Viral infection will express the viral oncogenes in the host cell, leading to transformation. DNA viral oncogenes, including SV40 large T-antigen, polyomavirus middle and large T-antigens, papillomaviruses E6 and E7 proteins, adenoviruses E1A and E1B proteins, and EBV LMP1, promote tumorigenesis primarily by acting on RB and p53 tumor suppressors to suppress their surveillance role in cell cycle progression. Some viral oncoproteins may also activate signaling proteins

at the inner surface by protein–protein interaction as exemplified by association of T-antigen with SRC, SHC, and insulin-like growth factor (IGF-IR) and papillomavirus E5 oncoprotein association and activation of platelet-derived growth factor (PDGF) receptor and c-SRC. The *HBV X* gene-encoded protein may stimulate host-cell transcription to promote cell proliferation, leading to tumorigenicity.

2.5.2 Cellular Oncogene Activation by Viral Integration

Some viruses do not encode viral oncogenes, but they can integrate into the host-cell genome and alter cellular oncogene(s) and tumor suppressor gene(s) at or around the insertion site. These viruses include avian leucosis virus (ALV) and mouse mammary tumor virus (MMTV). The first example is the activation of c-*MYC* in Bursal lymphomas by ALV insertion.

2.5.3 Chromosome Translocation

There are 22 pairs of euchromosomes and two sex chromosomes in human somatic cells, and human proto-oncogenes have been identified in all of these chromosomes. In many carcinomas, cancer cells have abnormal karyotypes, such as translocations, duplications, deletions, and loss of chromosomes. Chromosomal translocations can put proto-oncogenes under control of the promoter/enhancer of another gene, or create recombinant, truncated, or deleted proto-oncogenes. In some of these circumstances, the proto-oncogenes can be activated. Examples include *ABL* activation on the *Philadelphia chromosome* in chronic myelogenous leukemia (CML), c-*MYC* translocation in Burkitt's lymphomas and plasmacytomas, and *BCL*-2 activation by chromosome translocation (t14 : 18) in follicular B-cell lymphomas. Other proto-oncogenes activated by chromosomal translocation include *BCL*-6, *HOS11, LYL*-1, *RHOM*-1, *RHOM*-2, *TAL*-1, and *TAL*-2.

2.5.4 Gene Amplification

Normally, all the genes on the euchromosomes and the X-chromosome in female have two copies in somatic cells. The genomic sequences of some genes may be amplified from two copies to 5–500 copies or even higher, typically generating more mRNA transcripts and higher protein levels of the oncogene, resulting in elevated activity of the oncogene. Examples include the amplification of N-*MYC* in neuroblastomas and *ERBB*-2 in breast cancer and many other cancer types.

2.5.5 Genetic Sequence Mutation

Simple genetic sequence mutation is also a major mechanism involved in the activation of many oncogenes. The mutation could be point mutation or deletion of a fragment of the gene.

2.5.6 Aberrant Gene Expression

Some genes may be expressed in some tissues for normal physiological processes or expressed during embryogenesis, but may be oncogenic when aberrantly expressed in the cells of a different tissue that normally do not express these genes. Aberrant expression is usually caused by epigenetic alteration(s) in the gene locus, such as demethylation of CpG islands in some genes, and chromatin hyperacytylation in the regulatory sequences.

2.5.7 Alterations Not at the Genetic Levels

The oncogenic activity of a gene may also be elevated by the regulation of other gene products, these include the stability of the

mRNA transcript, mRNA export from the nucleus to the cytoplasm, translation initiation and processing efficiency, protein stability and degradation, and downregulation of negative regulator(s) of a particular oncogene. For example, AKT can be stabilized by the chaperone protein HSP90, and AKT activity is elevated in cancer cells that have null mutation of its negative regulatory protein PTEN.

3 Biochemical Properties and Cellular Functions of Oncogenes

Normally, any cell type within an organ or tissue has to be maintained at a certain number (homeostasis). Whether a normal cell is to proliferate is usually regulated by the integrated mitogenic and inhibitory signals in the surrounding microenvironment (Fig. 6). Most of the extracellular signals trigger the signaling cascades through their receptors on cell surface, these include mitogenic (e.g. EGF (Epidermal Growth Factor)) or inhibitory (e.g. TGF-β) growth factors, cytokines, ligands for seven-transmembrane (7-TMR) receptors (e.g. some hormones such as bombesin). Signaling cascades can also be initiated from the cell surface by cell–cell interaction and cell-extracellular matrix (ECM) interaction. The execution of the integrated extracellular signals is carried out by the cell-signaling cascades and the newly synthesized gene products within the cell. Within the cells, the signaling pathways are very complicated, and there exist cross-regulations among different pathways at multitude levels that form a circuit network (Fig. 7). The alterations caused by these signaling

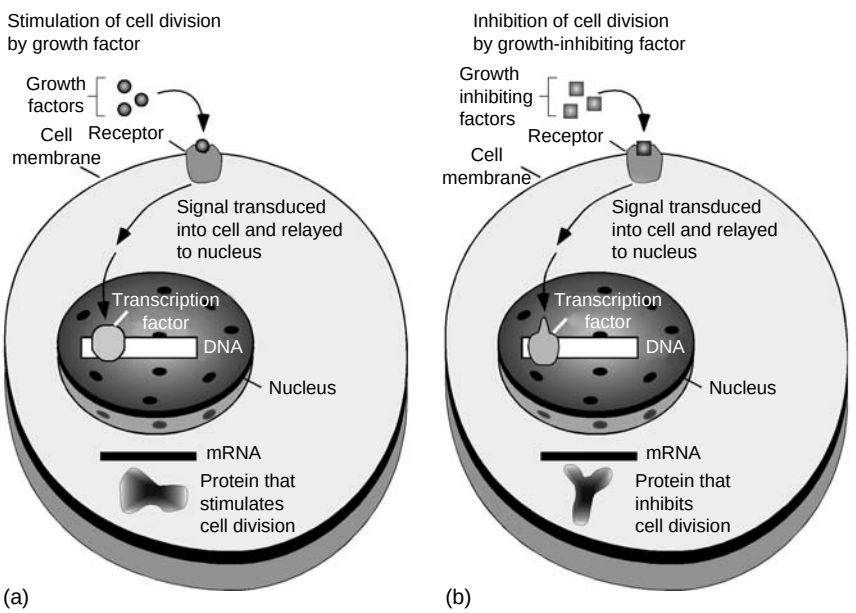

Fig. 6 Regulation of cell division by signals in the microenvironment. The impacts of extracellular signals such as growth factors (a) and growth inhibitory factors (b) on cell division are executed by the intracellular signaling, regulating protein activity and gene expression.

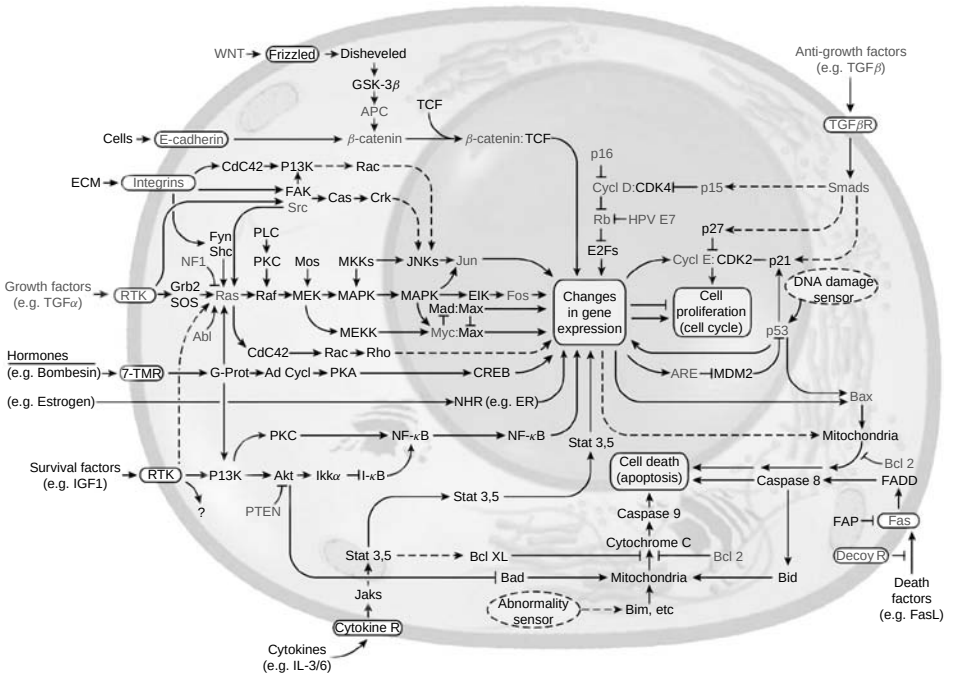

Fig. 7 The intracellular signaling circuit. Except steroid hormones such as estrogen, almost all extracellular signals are sensed by the cell surface receptors, which transmit the signals through the signal transduction cascades within the cell. Different signals or the same signals at different intensity may trigger common as well as distinct signaling pathways, leading to protein synthesis or activation. However, cross-regulations at many points of these different pathways exist, and many key regulatory molecules are shared among different pathways. The cell phenotype is regulated by the integrated signaling circuit in response to the complete signals in the environment. (Reproduced from Itanahan, D., Weinberg, R. A. (2000) The Hallmarks of Cancer, *Cell* **100**, 57–70. With permission of Cell Press).

molecules include qualitative and/or quantitative changes in proteins involved in cell signaling and the eventual outcome of cell phenotype. The qualitative changes could be mediated through protein phosphorylation by protein kinases (tyrosine and serine/threonine kinases) and dephosphorylation by phosphatases, as well as by other posttranslational modifications. The quantitative changes are mainly regulated at the transcriptional level mediated by the activation of transcriptional factors in the nucleus, but it can also be mediated through protein stability and degradation. In cancer cells, alteration(s) of an oncogene or several synergistic oncogenes could potentiate cells to proliferate without the requirement of normal surrounding signals.

3.1 Oncogenes Promoting Cell Cycle Progression and Cell Division

3.1.1 Oncogenes Derived from Cell Cycle Machinery

As described above (see Sec. 2.2), cyclins, CDKs, and E2F in the cell cycle machinery are positive regulatory proteins

for cell proliferation, and thus are candidate oncogenes. Of the cyclins, the D-type cyclins have been identified activated in cancers. Activation of CDK4 mediated by amplification or mutation (e.g. mutant CDK4 lacking the INK4 binding site) has been identified in tumors.

While CDKs are activated by cyclins, the activities of CDKs are also negatively regulated by the CKIs, including the INK4 ($p16^{INKA}$, $p15^{INKB}$, $p18^{INKC}$, and $p19^{INKD}$) that inhibit CyclinD/CDK4/6 and CIP/KIP ($p21^{WAF1/CIP1}$, $p27^{KIP1}$, and $p57^{KIP2}$) that inhibit all CDKs. The CKIs function as one of the cellular intrinsic surveillance mechanisms to make sure that the accurate genetic information is transmitted to the daughter cells. CKIs are regulated by growth-inhibitory factors, such as TGF-β, and by growth-arrest signals, such as DNA damage. When bound to inhibitory proteins such as INK4 ($p16^{INKA}$, $p15^{INKB}$, $p18^{INKC}$, and $p19^{INKD}$) and CIP/KIP ($p21^{WAF1/CIP1}$, $p27^{KIP1}$, and $p57^{KIP2}$), inactive CDK4 is phosphorylated at Tyr^{16} and Tyr^{17} by WEE1/MIK1/MYT1 and unphosphorylated at Thr^{172} (probably by p24KAP) (Fig. 8). Cyclins will compete with INK4 and CIP/KIP to bind to CDK, resulting in CDK phosphorylation at Thr^{172} by CAK (cyclin activated kinase) and dephosphorylation at Tyr^{16} and Tyr^{17} by CDC25 phosphatase, resulting in activation of CDKs. Therefore, *CAK* and *CDC25* could serve as proto-oncogenes to promote cell proliferation.

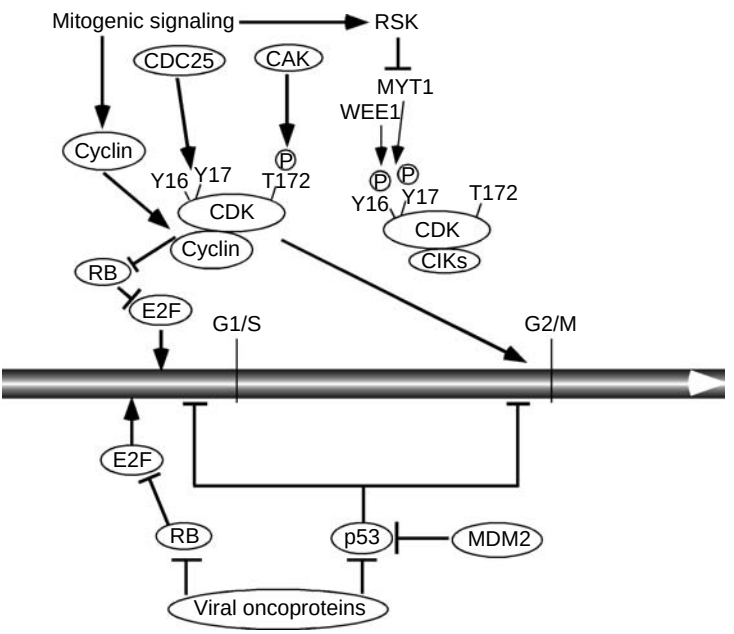

Fig. 8 Oncogenic proteins promoting cell cycle progression and cell division. Cellular oncoproteins E2F, cyclins, and CDKs promote cell cycle progression, and CDK activation is regulated by CDC25 and CAK. Viral oncoproteins may promote cell proliferation through inhibiting the negative regulatory proteins RB and p53.

3.1.2 Oncoproteins Leading to Loss of Control of Cell Cycle Checkpoints

Loss of chromosome integrity and accumulated genetic mutations caused by DNA damage are frequent phenomena found in cancer cells. In normal conditions, these cells cannot go through the cell cycle, instead the cells will either be arrested until the damaged DNA is fixed or the apoptotic pathway will be triggered in these cells. In addition to RB and CKIs, p53 is another principal regulatory molecule for the cell cycle checkpoints. Propagation of harmful genetic mutations and loss of chromosome integrity in cancer cells could be attributed to the defective function of these negative regulatory molecules, or to the gain of functions of oncogenes that suppress the function of these negative regulatory molecules. Examples of such oncogenes include MDM2 (mouse double minute 2) and some viral oncogenes (Fig. 8).

Under normal circumstances, p53 function is principally regulated at protein level through protein stability and degradation. MDM2 is a target gene of p53, MDM2 binds directly to p53, repressing its transcriptional activity and targeting p53 for ubiquitination and degradation. MDM2 in return is negatively regulated by p19/ARF. In response to DNA damage, p53 is upregulated through DNA-dependent protein kinase and ATM (ataxia telangiectasia mutated gene). The function of p53 in cell cycle arrest and apoptosis may be mediated by both transcription-dependent and -independent mechanisms. The target genes of p53 include $p21^{WAF1/CIP1}$, *GADD45* gene (growth arrest and DNA damage-inducible genes), *BAX, FAS, DR5,* and *IGFBP3*. $p21^{WAF1/CIP1}$ is involved in both G_1/S and G_2/M arrest by inhibiting CDKs to block cycle progression; $p21^{WAF1/CIP1}$ also binds to proliferating cell nuclear antigen (PCNA) to inhibit DNA replication. GADD45 expression is associated with G_1 and G_2/M arrest, and it also binds to PCNA to stimulate damaged DNA repair. p53 also can suppress the expression of some genes without p53-binding site; these include *RELA, IL-6, CYCLIN A,* and *PCNA*. MDM2 is a ubiquitin E3 ligase found highly expressed in many cancer types; its overexpression leads to degradation of p53 and thus suppresses the tumor suppressor functions of p53 in cell cycle arrest, apoptosis, and transcription activation of p53 target genes.

Many DNA viral oncogenic proteins promote tumorigenesis by inactivating the RB and p53 tumor suppressor genes. Viral oncoproteins interacting with and inactivating RB include adenoviral E1A, HPV E7, and SV40 large T-Ag. Viral oncoproteins interacting with p53 include HPV E6 (leading to p53 degradation by ubiquitin-dependent proteolysis), adenoviral E1B (repressing the transcriptional activity of p53), and SV40 T-Ag.

3.2 Growth Factors and Extracellular Mitogenic Signals

Growth factors are polypeptides that regulate cellular processes by binding to their specific receptors on target cells and trigger the intracellular signaling cascades by activating the receptors. Oncogenes derived from growth factors include those that are mitogens involved in cell proliferation. Unlike endocrine hormones, growth factors usually act on cells locally. The three types of growth factor action are autocrine, paracrine, and juxtacrine (Fig. 9). In autocrine regulation, the growth factor and its corresponding receptor are produced in

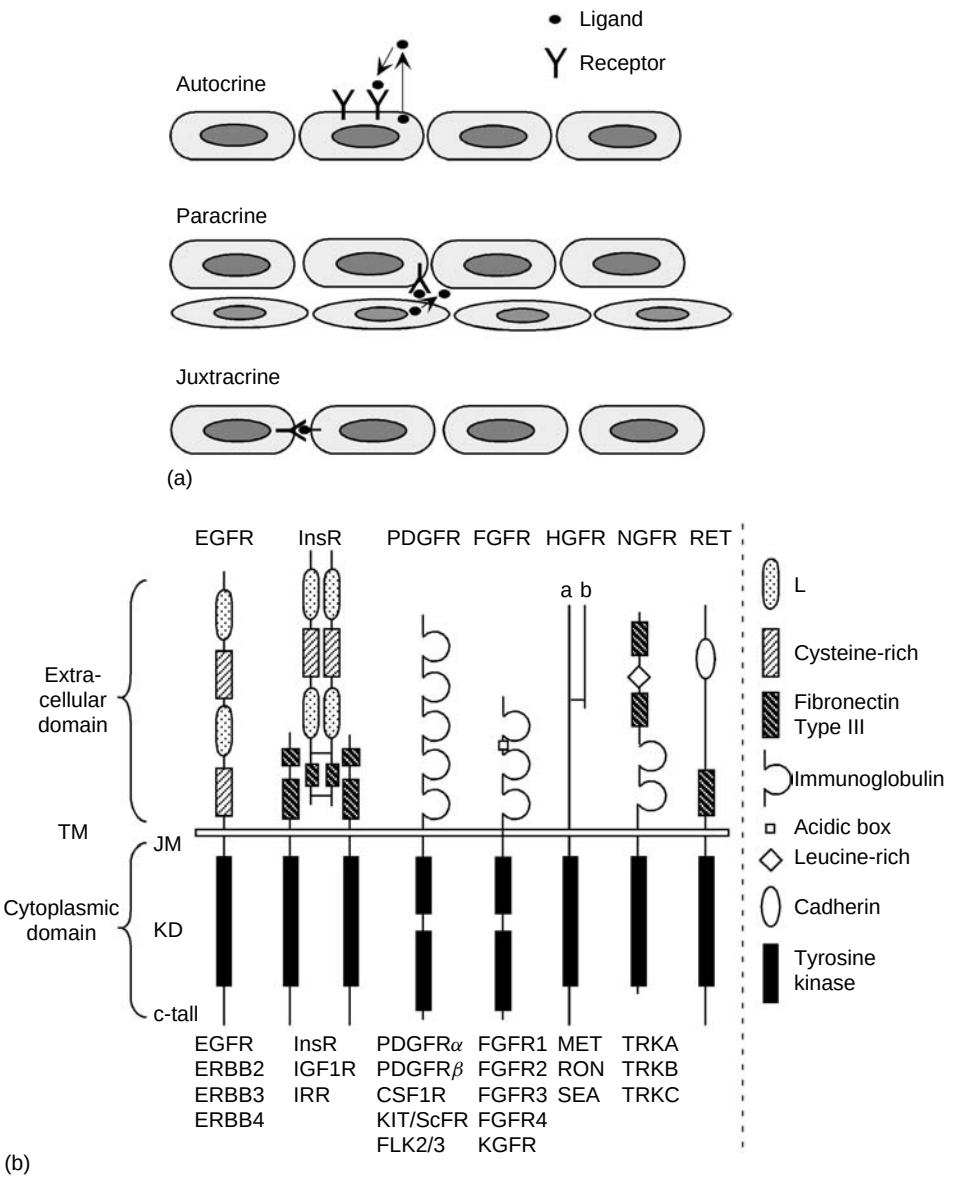

Fig. 9 Mitogenic signals and the receptor tyrosine kinases. The three types of growth factor action on cell surface receptors (a) and the structure features of representative growth factors (b), modified from Hubbard, S.R. (1999) Structural analysis of receptor tyrosine kinases, Prog. Biophys. Mol. Biol. **71**, 343–358. The extracellular domain is the site for growth factor ligand binding, and cytoplasmic domain possesses kinase domain and contains sites for binding by proximal signaling molecules.

the same cell, so that the growth factor can act on the cell from where it is produced. In paracrine regulation, growth factors are released from cells and act on cells in their immediate vicinity. An example of paracrine regulation is the regulation of KGF (keratinocyte growth factor) and KGF receptor; KGF receptor is expressed only in epithelial cells, activation of KGF receptor is regulated by the binding of its ligand KGF, which is produced by mesenchymal cells. Usually, extracellular matrix is involved in restraining the dissemination of the growth factors, and sometimes some matrix component is actively involved in the interaction between the growth factor and its receptor. A good example is the involvement of heparin in the interaction of FGFs (fibroblast growth factors) and FGF receptors (FGFRs). Juxtacrine regulation is a special type of paracrine action, in which growth factor is either a transmembrane protein or tightly restrained to the surface of the cell membrane, and cell–cell contact is required for the growth factor to exert its action.

3.2.1 PDGF (Platelet-derived Growth Factor) and the *SIS* Oncogene

PDGF contains two subunits, the A subunit and the B subunit, either in homodimer or heterodimer. The B subunit of PDGF is encoded by the *SIS* oncogene, originally found in simian sarcoma virus. While PDGF AA homodimer activates only the receptor for PDGFα, PDGF BB homodimer and AB heterodimer activate both PDGFR-α and PDGFRβ. Both PDGF AA and PDGF BB were shown having oncogenic activity when analyzed by *in vitro* experiments. PDGF and PDGF receptor may form autocrine regulation in sarcomas, lung carcinomas, astrocytomas, and acute myelogenous leukemia (AML).

3.2.2 FGFs (Fibroblast Growth Factors)

The FGF family comprises at lease nine members. Among the FGF family members, FGF-3/INT-2, FGF-4/HST/K-FGF, FGF-5, and FGF-6 have been identified as oncogenes. FGFs and their corresponding receptors can form autocrine regulation and may be involved in many cancer types, including colorectal, breast, ovarian, lung, and prostate carcinomas.

3.2.3 The WNT Gene Family

At least 10 members have been identified in the WNT family, and WNT-1, -2, -3, and -5 can be activated as oncogenes. The receptors for WNT are seven-transmembrane FZD (frizzled) proteins. The WNT-initiated signaling also releases the oncogenic function of β-catenin, as the loss of *APC* gene in colon cancer (Fig. 7). Deranged WNT signaling have been observed in colorectal, breast, gastric, cervical, and lung carcinomas.

3.2.4 EGF (Epidermal Growth Factor) and TGF-α

EGF and TGF-α (transforming growth factor-α) are ligands for EGF receptor (EGFR). Other EGF family members include heregulins. Expression of EGF and TGF-α are found in several cancer types, including breast and colorectal carcinomas.

3.2.5 Hematopoietic Growth Factors

More than 20 different hematopoietic growth factors have been identified, and at least 4 of them can function as oncogenes, including IL-2 (interleukin-2), IL-3, GM-CSF (granulocyte-macrophage colony stimulating factor), and colony stimulatory factor 1 (macrophage colony stimulating factor).

3.3 Oncogenes Derived from Cell Surface Receptors

3.3.1 Growth Factor Receptors

Growth factor receptors are RTKs (receptor tyrosine kinases) comprising an extracellular domain, the transmembrane domain, and the cytoplasmic portion composed of the juxtamembrane domain, the kinase domain, and the C-terminal region (Fig. 9). The extracellular domain is the site for specific ligand binding; the cytoplasmic domain has intrinsic kinase activity and binding sites for proximal signaling proteins. Upon ligand binding to the extracellular domain, monomeric RTKs dimerize as homo- or heterodimers and the tyrosine residues become autophosphorylated. The phosphorylated tyrosines provide docking sites for many SH2-domain-containing proteins, and the amino acid residues flanking the phosphorylated tyrosine residue may determine the specificity of binding. SH2-containing proteins could be protein kinases (SRC, ABL, PI3K/p85, PLC-γ), or adapter proteins without kinase activity (e.g. GRB2, CRK, SHC, NCK) (see Sec. 3.3.3). The major signaling pathways triggered by different RTKs are quite similar, including RAS-RAF-MEK-MAPK and PI3K-AKT pathways (see Sec. 3.4), and how different growth factors and their receptors are fine-tuned to achieve the specificity in signaling and cellular responses are still under intense investigation.

3.3.1.1 **Epidermal growth factor receptor family** The EGF receptor belongs to the ERBB family of four closely related cell membrane receptors: EGFR (HER1 or ERBB1), ERBB2 (HER2/neu), ERBB3 (HER3), and ERBB4 (HER4). EGFR and HER-2/neu are involved in many aspects of almost all cancer types, including lung, colorectal, prostate, breast, and ovarian cancer. These receptors are transmembrane glycoproteins that consist of an extracellular ligand binding domain, a transmembrane domain, and an intracellular domain with tyrosine kinase activity for signal transduction. Activation of ERFR occurs when a ligand, such as EGF, TGF-α, or amphiregulin, binds to its extracellular domain. This causes the receptor to dimerize with either another EGFR monomer or with another member of the ERBB family. Following receptor dimerization, activation of the intrinsic protein tyrosine kinase activity and tyrosine autophosphorylation occur. These events lead to the recruitment and phosphorylation of several intracellular substrates, leading to mitogenic signaling and other cellular activities. Receptors that lack kinase function, because of mutations at the ATP binding sites, do not display a full range of biochemical responses following ligand binding; this demonstrates that receptor tyrosine kinase activity is required in cellular signaling. A major signaling route of the ERBB family appears to be the RAS-RAF-MEK-ERK pathway. Another important pathway in ERBB receptor signaling is the one constituted by phosphatidylinositol 3-kinase and the downstream protein kinase AKT. After its activation, AKT transduces signals that regulate multiple biological processes including apoptosis, gene expression, and cellular proliferation. AKT is likely to send survival (antiapoptotic) signals by phosphorylating multiple targets, including the BCL-2 family member BAD (a proapoptotic factor), the cell-death pathway enzyme caspase-9, Forkhead transcription factors, and IκB kinase α. AKT also plays a prominent role in the regulation of cell cycle progression. Thus, EGFR signaling can lead to a variety of downstream reactions,

which are subject to complex regulatory mechanisms.

3.3.1.1.A. **EGFR and cancer** EGFR signaling impacts on many aspects of tumor biology. Activation of the EGFR has been shown to enhance processes responsible for tumor growth and progression, including the promotion of proliferation, angiogenesis, invasion/metastasis, and inhibition of apoptosis. The expression of EGFR in tumors has been correlated with disease progression, poor survival, poor response to therapy, and the development of resistance to cytotoxic agents. High levels of EGFR have been observed in a variety of tumors, including prostate, breast, gastric, colorectal, and ovarian. However, mechanisms other than EGFR expression might affect EGFR signaling. For example, genomic rearrangements in the *EGFR* can cause expression of modified receptors. One frequently occurring truncated form, EGFRvIII has previously been described by Wong and colleagues and is the result of a 267 amino acid in-frame deletion in the extracytoplasmic domain with the generation of a novel glycine at the fusion junction (Fig. 10). It has been shown that this altered receptor is present in a variety of cancers, including breast, lung, and ovarian carcinomas. The fact that this receptor is frequently found in tumors, but not present in normal tissue, makes it an attractive candidate for various antitumor strategies.

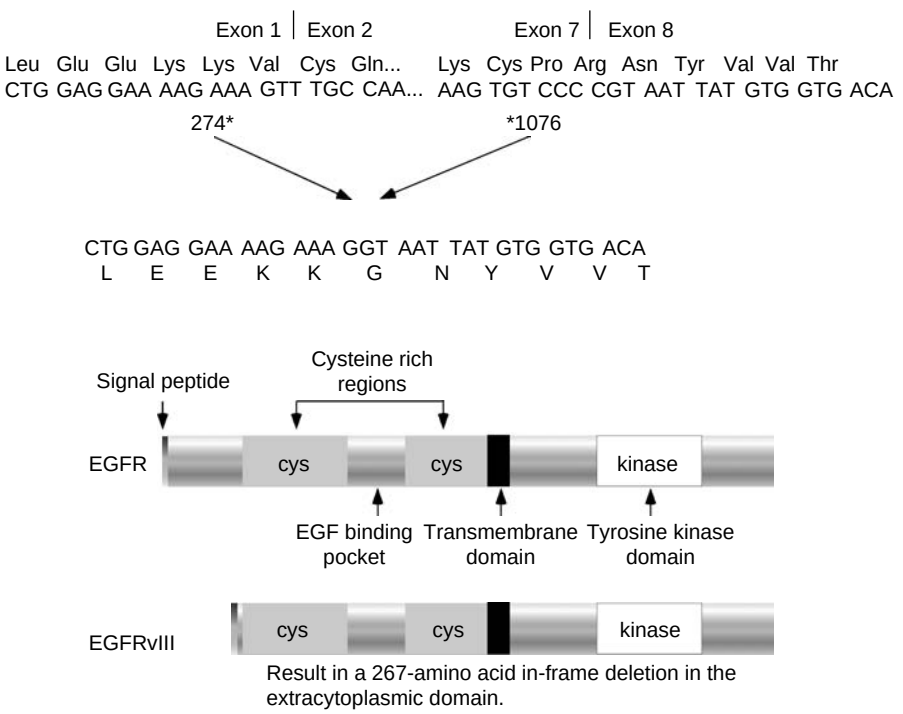

Fig. 10 Schematic representation of alternative spliced EGFR transcript, which encodes for the mutant EGFRvIII. EGFRvIII differs from EGFR in that 267 amino acid residues are absent from the extracytoplasmic domain.

3.3.1.1.B. **HER-2/neu and cancer** HER-2/neu has intrinsic kinase activity, and no ligand has been found for HER-2/neu. Overexpression of wild-type HER-2/neu or somatic mutations at the transmembrane (TM) region are believed to enhance and maintain the receptor dimerization. HER-2/neu can interact with other members of EGFR family, including EGFR, HER-3, and HER-4 to form heterodimer. Activated HER-2/neu can interact with many cellular signaling proteins, including GRB2, SHC, PLC-γ, c-SRC. GRB2 binding recruits SOS to the plasma membrane and activates the RAS-RAF-MEK-MAPK pathway (see Sec. 3.4). PI3K-AKT pathway is also activated by HER-2/neu, although it does not contain the p85/PI3K binding site, and, maybe, it is mediated through heterodimerization with HER-3.

HER-2/neu is overexpressed in many different types of human cancers, including breast, ovarian, lung, gastric, and oral cancers. Amplification of HER-2/neu or overexpression of HER-2/neu has been identified in about 20 to 30% of breast, ovarian, and gastric cancers. HER-2/neu is also frequently expressed in NSCLC (nonsmall cell lung cancer). Increased HER-2/neu expression can enhance the metastatic potential of human cancer cells, and repression of HER-2/neu suppresses the malignant phenotypes of HER-2/neu-overexpressing cancer cells.

3.3.1.2 **Fibroblast growth factor receptor family** Four members have been identified in the FGFR family. FGFRs are expressed by most cancer cells, and activation of FGFRs by mutation has been identified in several cancer types, including melanoma, bladder, and cervical carcinomas. Proteins binding to the cytoplasmic region of FGFRs include FRS2, GRB2/SOS, PLC-γ, CRK, and SRC. FRS2 is critical for FGFR-initiated activation of MAPKs and phosphatidylinositol-3 (PI-3) kinase activation in cell proliferation.

3.3.1.3 **Platelet-derived growth factor receptor family** The PDGF receptor family includes PDGF α and β receptors, stem cell factor receptor (KIT), macrophage-colony stimulating factor-1 receptor, and FLT3. Autocrine activation of PDGF receptor has been identified in many cancer types, including lung, gastric, prostate, glioblastoma, and astrocytomas. Activation of PDGF receptor are involved in various cellular responses, including cell growth, proliferation, and cell migration. The proximal proteins interacting with PDGF receptor and transmitting the signaling cascades include PI3K, CRK, PLC-γ, and SRC. Interaction of PDGF receptor with SRC, integrin and FAK (focal adhesion kinase) might be involved in lamellipodia formation and cell migration. STAT-1, -3, and -6 can be activated by PDGF receptor, independent of SRC activation. Protein tyrosine phosphatases (PTP) that are potentially associated with PDGF receptor and involved in downregulation of PDGF receptor include PTP-PEST, SHP-2, PTP-1B, low M(r) PTP.

KIT, a 145-kD transmembrane glycoprotein, is the receptor tyrosine kinase encoded by the c-*KIT* proto-oncogene and is the normal cellular homolog of the viral oncoprotein v-*Kit*. These kinases have an extracellular domain containing five immunoglobulinlike domains, a single transmembrane domain, and a cytoplasmic domain with a split kinase domain and hydrophilic kinase insert sequence. The juxtamembrane and kinase domains of these receptors are strongly conserved. c-KIT is normally expressed by hematopoietic progenitor cells, mast cells, germ cells, interstitial cells of Cajal, and also by

certain human tumors. Normal function of the c-Kit gene product, as demonstrated in mouse models, is essential for hematopoiesis, melanogenesis, gametogenesis, development of ICC (interstitial cell of Cajal), and mast cell growth and differentiation.

Two general mechanisms of c-KIT involvement in the development or maintenance of the malignant phenotype have been described. First is the acquisition of c-KIT mutations resulting in ligand-independent activation. Second is tumor cell expression of KIT with autocrine and/or paracrine stimulation of the receptor by its cognate ligand, steel factor (STL). c-KIT activating mutations have been described in the conserved juxtamembrane domain, extracellular domain, and in the kinase domain. Structural alterations of the juxtamembrane domain may be a common pathway for activation of type III receptor tyrosine kinases as similar mutations have been described for FLK2/FLT3 and PDGFR. Gastrointestinal stromal tumors (GISTs) are the most common mesenchymal malignancies of the GI tract and these tumors are characterized by the presence of constitutively activated KIT. The majority possesses gain of function mutations in c-*KIT* (Fig. 11), while a smaller percentage have a mutation in *PDGFR*.

KIT oncogenic activation results in autophosphorylation of critical tyrosine residues, which are "docking" sites for various signaling intermediates. These signaling intermediates are phosphorylated by the activated KIT proteins or by each other, and they transmit the KIT oncogenic signal to downstream targets. Various KIT oncoproteins differ in their structure and in their activation of specific downstream signaling pathways, depending on the mutation. KIT-mediated phosphorylation of signaling proteins such as AKT, STATs (Signal transducers and activators of transcription), MAPK, and JNK can influence proliferation, apoptosis, differentiation and/or cell adhesion, which may contribute to the malignancy of GIST. These key signaling proteins, and others downstream of KIT, have variable biological effects, depending on the cellular

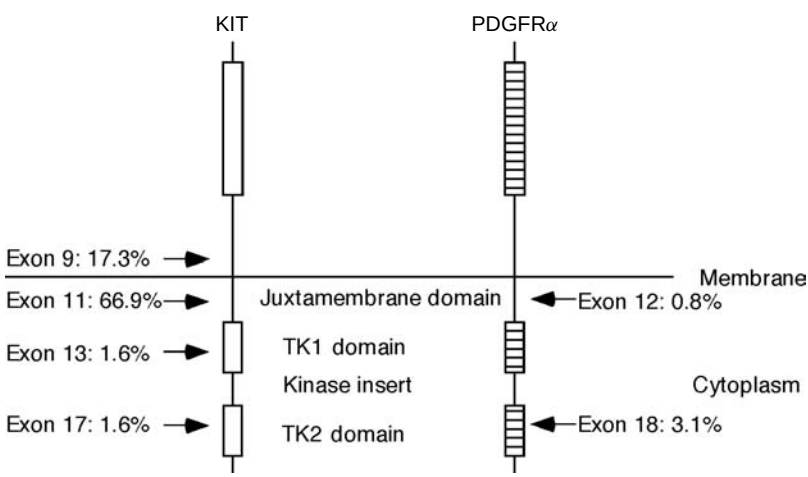

Fig. 11 Structure of KIT and PDGFR and the frequency of mutations observed in sporadic and familial gastrointestinal stromal tumors *(GISTs)*.

context. For example, RAS and MAPK activation is a key transforming event in some cancers, but many tyrosine kinase oncoproteins, including BCR-ABL, do not require this pathway.

3.3.1.4 **Insulin receptor family** The insulin-like growth factor (IGF) family, including insulin, IGF-I, and IGF-II, and their receptors and binding proteins (IGFBP) are involved in various normal cellular functions in development and in the body. Aberrant expressions of those proteins were found in many cancer types and they play an important role in tumorigenesis. IGF-I receptor (IGF-IR) is the principal receptor for IGF-I and IGF-II, IGF-II also binds to the isoform A of insulin receptor (IR-A) with high affinity. IGF-IR and its ligands are frequently overexpressed in many cancer types, including breast, colorectal, gastric, lung, ovarian, prostate, and pancreatic carcinomas. IGF-IR and its ligands IGF-I or IGF-II form an autocrine regulation, and the effects of IGF-I and IGF-II are regulated by IGF-binding proteins (IGFBPs) and IGFs may act synergistically with other mitogenic growth factors and steroids

Activation of IGF-IR may promote tumor progression by promoting cell growth and proliferation, inhibiting apoptosis, and inducing angiogenesis. Activation of IGF-IR may also promote metastasis by enhancing cell motility and regulating cell–cell and cell–matrix adhesion. The impact of IGF-IR on these processes has been assessed by evaluating the activation of several signaling pathways and the expression of many genes involved in these processes. Among the genes upregulated in response to activation of IGF-IR are vascular endothelial growth factor (VEGF) and COX-2. Activation of IGF-IR leads to the activation of several signaling pathways, including PI3K-AKT pathway and the RAS-RAF-MAPK pathway.

3.3.1.5 **MET/HGF receptor family** The MET family consists of three related proteins: MET, RON, and c-SEA. HGF/SF (Hepatocyte growth factor/scatter factor) is the ligand for MET, and many cell types, including epithelial cells, are regulated by HGF-MET interaction. Under physiological conditions, HGF is produced by mesenchymal cells and HGF-MET can form a paracrine loop regulation. Unlike most other RTKs, MET exists as a disulfide-linked heterodimer originated from the same precursor protein through proteolytic cleavage. The C-terminal end of MET has a multiple substrate docking site for the binding of adapter proteins, including GRB2, SHC, CRK/CRKL, GAB1. In addition to ligand-dependent activation, MET can also be transactivated (HGF-independent) by other membrane receptors, such as adhesive receptors (e.g. integrins and CD44), RON, and EGFR.

Increased MET expression has been found in many types of carcinomas, and dysregulation of MET is related to metastatic ability of tumor cells. The mechanisms of MET dysregulation include the following: (1) wt-MET and HGF-dependent: wt-MET is activated by HGF through autocrine or paracrine regulation; (2) HGF-independent MET activation that are caused by overexpression of wild type or mutant MET, gene rearrangement (e.g. *TPR-MET*), abnormal MET processing, defects of negative regulators, and truncation. TPR-MET encompasses the TPR leucine zipper domain and the MET cytoplasmic kinase domain, which results in constitutive activation of MET because of leucine zipper domain interactions.

3.3.1.6 **Neurotrophin receptor family** The TRK (tropomyosin receptor kinase, including TRKA, TRKB, and TRKC) family are receptors for neurotrophins, including nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3), and neurotrophin-4 (NT-4), and are normally expressed in neurons. While activation of TRK by neurotrophins is involved in tumors of both neuronal origin such as neuroblastoma and medulloblastoma and nonneuronal origin such as prostate and pancreatic cancer, TRK is constitutively activated in many nonneuronal cancer types caused by chromosome translocation. The *TRK* (*TRKA*) proto-oncogene encodes a 140-kDa protein composed of 790 amino acid residues. In colon cancer, the *TRK* oncogene is composed of sequences from nonmuscle tropomyosin gene and the *TRK* proto-oncogene (encoding amino acid residues 392–790). The nonmuscle tropomyosin sequence may contribute to the activation of TRK by allowing its kinase domain to fold into a constitutively active configuration. Proto-oncogene *TRK* can also be activated by genetic rearrangement with other genes, such as with the ribosomal large subunit protein *L7a* to form *TRK-2H* oncogene in breast carcinoma, with the *TPR* to form *TRK-T1* oncogene in papillary thyroid carcinomas, and with the *ETV6* to form the *ETV6-NTRK3* oncogene in human secretory breast carcinoma and several other cancer types. Activation of TRK may contribute to tumorigenesis by promoting cell proliferation and suppression of apoptosis, and these impacts are carried out by the activation of multiple signaling pathways, including PI3K-AKT pathway, RAS-RAF-MEK-MAPK pathway, and activation of PLC-γ.

3.3.2 Cell Surface Receptors without Tyrosine Kinase Activity

Some cell surface receptors do not have intrinsic kinase activity, but they may be oncogenic by activating the downstream targets to initiate the signaling cascades or induce expression of oncogenes. These include the cytokine receptors (e.g. receptors for IL-2, IL-3, and GM-CSF, receptor for erythropoietin) and the G-protein-coupled seven-transmembrane receptors (MAS, Thyrotropin receptor, Serotonin 1c receptor).

3.3.3 Membrane-associated Nonreceptor Oncogenes

3.3.3.1 **Membrane-associated nonreceptor tyrosine kinases** Some protein tyrosine kinases are not cell surface receptors, but they are associated with the inner surface of the plasma membrane. They are associated with the membrane by posttranslational modification such as myristilation in which myristic acid is added to the 2nd amino acid residue glycine after Met is removed, thus facilitating their interaction with cell surface receptors. These protein kinases also constitute a large group of oncogenes, including members of SRC and ABL subfamilies (see Table 1). The SRC subfamily includes SRC, YES, FGR, FYN, HCK, BLK, YRK, LCK (TKL), and LYN/SYN. The ABL subfamily includes ABL and ARG. ABL contains a SH3 domain at its N-terminus, followed by a SH2 domain, the kinase domain, and the C-terminal region. ABL proto-oncogene can be activated either by incorporation into retrovirus or by chromosome translocation. In Abelson leukemia virus, the N-terminus including the SH3 domain of c-*ABL* proto-oncogene is deleted during the integration into the viral genome to form the v-*ABL* oncogene. In Philadelphia

chromosome translocation, c-*ABL* is fused with *BCR* to form the oncogene *BCR/ABL*, encoding a recombinant protein BCR/ABL in which ABL is activated. Activation of ABL may activate multiple signaling pathways, including the JAK-STAT pathway.

3.3.3.2 **Membrane-associated adaptor proteins** Some SH2- and/or SH3-containing proteins may not have kinase domain, but may function as adaptor proteins to facilitate protein–protein interactions, and thus may also be oncogenic. These proteins include GRB2, CRK, SHC, and NCK.

GRB2 contains a SH2 domain and two SH3 domains, the SH2 domain can bind to phosphotyrosine on activated RTKs or proteins associated with RTKs (e.g. SHC). The SH3 domains bind to SOS, so that GRB2 and SOS always form a complex in the cells. SOS is the guanine nucleotide exchange factor for RAS (see Sec. 3.4.3). When recruited to the activated RTKs, the GRB2-SOS is activated to promote RAS GTP/GDP exchange, resulting in RAS activation in the RAS-RAF-MEK-MAPK pathway (see Sec. 3.4.1). In addition, GRB2 is also involved in cell transformation caused by polyomavirus middle T-antigen.

CRK family has four members, v-CRK, CRKI, CRKII, and CRKL. The SH2 domain of CRK family proteins is involved in association with cell surface receptors, while the SH3 has been identified to be associated with many signaling molecules including SOS, C3G (GEF for RAP1), ABL, and DOCK180.

3.3.3.3 **Membrane-associated phospholipase: PLC-γ (phospholipase C-γ)** While PLC-γ binds to and is activated by many RTKs such as EGFR and ERBB2, PLC-β is activated by G-proteins (see Sec. 3.4.3). Phospholipases can catalyze the hydrolysis of phosphatidylinositol 4,5-bisphosphate (PIP2) to generate diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP3). DAG activates PKC, and IP3 can stimulate Ca2+ release. PLC-γ is overexpressed in some cancer types, including breast and colon cancer. The signaling proteins activated by PLC may be involved in mitogenic signaling pathways in cell growth and cell invasion and contribute to tumor progression.

3.4 Oncogenes in Cytoplasmic Signal Transduction Pathways

3.4.1. RAS-RAF-MEK-MAPK (ERK) Pathway

Mitogen-activated protein kinase (MAPK) pathway and PI3K-AKT pathway are two major signaling pathways activated by cell surface receptors to regulate diverse cellular processes. With regard to oncogenesis, the MAPK pathway is involved in cell proliferation and cell survival, and PI3K-AKT is involved in cell growth and cell survival. The regulation of MAPKs is mediated by a conserved core module, which is composed of the MAPKs, MEK (MAPKK), MEKK (MAPKKK) (Fig. 12). ERK, JNK, and p38 are the major MAPKs in higher eukaryotes. The MEKK regulating ERK is RAF, and RAF is regulated by RAS. RAS is a small G-protein regulated by GRB2/SOS (see Sec. 3.4.3). Activated GTP-bound RAS recruits RAF to the membrane by directly interacting with RAF, resulting in RAF phosphorylation and activation. In addition to mitogenic stimulation activated RTKs, MAPK pathway is activated by several other mechanisms. Activation of heterotrimeric G-proteins can activate MAPK directly or

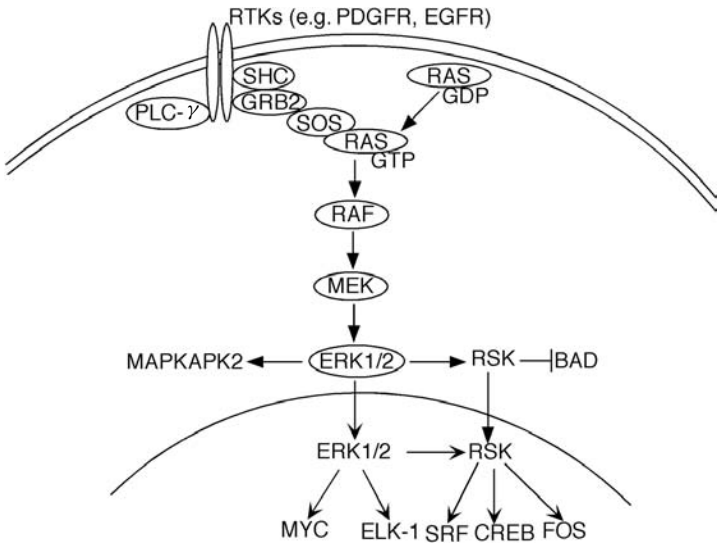

Fig. 12 The RAS-RAF-MEK-MAPK pathway. The small GTPase RAS is associated with cell membrane inner surface. Upon activation of RTKs, the guanosine nucleotide exchange factor SOS is activated, which, in turn, activates RAS. RAF-MEK-ERK is the major pathway for activated RAS, and can be cross regulated by many other pathways such as heterotrimeric G-proteins. The activated ERK and its substrate RSK family members regulate many immediate-early genes and the activity of many proteins involved in diverse cellular processes including cell survival and apoptosis.

indirectly through activation of PKC. Activated ERK can phosphorylate RSK1 to promote cell survival, and it can also activate many transcription factors to stimulate gene expression to promote cell proliferation.

3.4.1.A **RAS** There are three members of RAS: Ha-RAS (homologous to the Harvey murine sarcoma virus oncogene), K-RAS (homologous to the Kirsten murine sarcoma virus oncogene), and N-RAS (initially isolated from a neuroblastoma cell line). RAS is associated with the plasma membrane inner surface, which is mediated by posttranslational modifications and is essential for its function. The modifications include the farnesyl isoprenoid moiety added to the CAAX motif at the carboxyl terminus by farneysyltransferase, palmitoylation at cysteine residue(s) N-terminal to the CAAX motif, the removal of AAX from the CAAX motif by RCE1 cleavage followed by carboxymethylation of the modified cysteine residue in the CAAX motif. Mechanisms leading to RAS activation include gene mutation and gene amplification that have been identified in many cancer types. Some mutations lead to irreversible binding of GTP to RAS, including mutations at amino acid residues 12, 13, 59, 61, and 63. In certain circumstance, activations of RAS and MYC have a significant synergistic effect on tumorigenesis. Even in the case that there is no mutation in RAS, RAS may also contribute to tumorigenicity if its upstream regulatory proteins such as RTKs are activated.

3.4.1.B **RAF** There are three members in the RAF family, RAF-1, A-RAF, and B-RAF. B-RAF is implicated in tumor progression, but not in tumor initiation. The protein structure of RAF family members contain a kinase domain in its carboxyl-terminal half and a regulatory region in its amino-terminal half. In normal cells, RAF is activated by RAS in the RAS-RAF-MEK-MAPK pathway. AKT might be involved in RAF phosphorylation at Ser^{259}. RAF can become oncogenic when the N-terminal regulatory region is deleted, as observed in retroviruses and in in vitro functional analysis.

3.4.1.C **ERK and cell survival** MAPKs are activated in response to many mitogenic stimulations, and they are involved in numerous physiological functions, including cell survival, apoptosis, and metastasis. Activation of ERK may promote cell survival and suppress apoptosis. One of the pathways for ERK to promote survival is to activate RSKs; activated RSKs phosphorylate a number of proteins, including BAD and CREB that are involved in cell survival and inhibition of apoptosis. RSKs belong to the AGC protein serine/threonine kinase family, which includes PKA, PKB/AKT, PKC, and RSKs. RSKs and AKT share many common substrates, and many of these substrates are involved in cell survival.

ERK and immediate-early responsive genes in cell proliferation. MAPK activation also leads to the synthesis of immediate-early genes (50–100 in total, including many oncogenes such as FOS, c-MYC, JUN, and D-type cyclins) that are involved in cell proliferation. The synthesis of those immediate-early genes is very rapid and transient, and no new protein synthesis is required. For example, c-FOS is induced within 15 min of growth faction stimulation and returned to basal level within 2 h. Induction of FOS transcription is regulated by the binding of transcription factor complex, including SRF (serum response factor) and ELK1, to the SRE (serum response element), and binding of STAT to the SIE (SIS inducible element) in its promoter region. ELK1 is a substrate of ERK1/2, and SRF can be phosphorylated and activated by RSK, which is in turn phosphorylated and activated by ERK1/2. The induced FOS protein may be involved in the suppression of further FOS transcription by negative feedback autoregulation. The synthesis of immediate-early genes is regulated at both transcriptional and posttranscriptional levels, while the subsequent downregulation of immediate-early genes to basal level requires new protein synthesis.

3.4.2 The PI3K-AKT Pathway

AKT is involved in the tumorigenesis of many cancer types. There are three members in the AKT family, AKT1, AKT2, and AKT3. AKT family members have an N-terminal pleckstrin homology (PH) domain, a catalytic domain, and a C-terminal regulatory domain. AKT activation is regulated by PI3-K kinase, PDK1, MSK, and MAPKAP kinase-2 (Fig. 13). PI3K contains two subunits, the p110 catalytic subunit and the p85 regulatory subunit. AKT activation involves the phosphorylation of Thr^{308} in the catalytic domain and Ser^{473} in the C-terminal regulatory region. In response to RTK activation, PI3K is recruited to the RTK via the SH2 domain of p85 subunit and becomes activated. The activated PI3K catalyzes the generation of 3'-phosphoinositides, which, in turn, recruits PKB/AKT from the cytosol to the plasma membrane via the PH domain of AKT. The membrane-bound

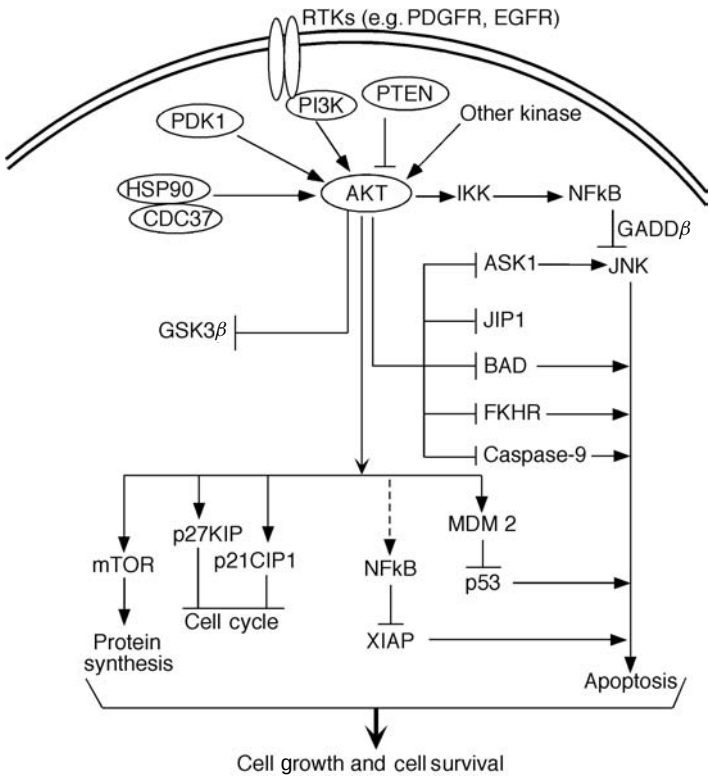

Fig. 13 The PI3K-AKT pathway. AKT activation is regulated by PI3-K kinase, PDK1, and other kinases such as MSK and MAPKAP kinase-2. HSP90/CDC37 promote AKT activity by stabilizing AKT. Activated AKT is involved in the regulation of many key molecules in cell growth and cell survival, including the IKK-IK-NFκB pathway, the stress-ASK-JNK pathway, the MDM2-p53 pathway, and many other proteins such as BAD and caspases-9.

AKT might change its conformation at this step followed by phosphorylation of AKT at Thr^{308} and Ser^{473}, which leads to the activation of AKT. The mechanism of AKT phosphorylation is still not completely understood. Among the protein kinases implicated in the phosphorylation of AKT are PDK1 (3'-phosphoinositide-dependent kinase) and MSK. While PI3K is the positive regulator for AKT activation, this process is negatively regulated by PTEN. PTEN is a lipid phosphatase that can dephosphorylate PtdIns(3,4,5)P3′ (phosphatidylinositol 3,4,5-trisphosphate). CTMP (Carboxy-terminal modulator protein, 27 kDa) is another negative regulator that can inhibit the phosphorylation of AKT at Ser^{473}. After being activated, AKT is dephosphorylated by phosphatases such as PP2A and its activity returns to basal level. ERK activation is also involved in AKT activation in some cell lines.

Deregulated PI3K-AKT signaling has been found in many cancer types, including glioblastoma, ovarian, breast, endometrial, hepatocellular, melanoma, digestive

tract, lung, renal-cell carcinoma, thyroid, lymphoid. The mechanisms involved in PH3K-AKT deregulated signaling include alterations in *PTEN* (mutations, allelic imbalance, inactivation), PI3K/p85α (mutation), PI3K/p110α (amplification), AKT (elevated kinase activity of AKT1, amplification and overexpression of AKT2), and RSK (amplification and overexpression). ERBB2 overexpression also leads to AKT activation and chemoresistance.

AKT might be involved in cell survival by multiple mechanisms, by phosphorylating and hence inactivating ASK1, BAD, caspase-9, and FKHR (a member of Forkhead family of TFs) (Fig. 13). BAD is a proapoptotic protein of the BCL-2 family (see Sec. 3.4.6). Phosphorylated BAD at serine^{136} by AKT or serine^{112} by RSKs can interact with 14-3-3 chaperone proteins and be sequestered from binding to BCL-XL, and thereby promoting cell survival. Phosphorylated FKHR cannot be translocated into the nucleus to activate its target genes of which some are proapoptotic proteins such as BIM and FAS ligand. AKT can also phosphorylate IKK (IκB kinase) and activate IKK-NFκB cell survival signaling. NFκB target genes include *IAP*s (c-*IAP*1, c-*IAP*2, and *XIAP*), and *GADD*45β. GADD45β overexpression can inhibit JNK activation, and XIAP can inhibit caspase activation. AKT can phosphorylate MDM2, which can translocate into the nucleus more efficiently and bind to p53, leading to E3 ubiquitin ligase-mediated degradation of p53. AKT promotes the transcriptional activity of ERα, independent of estrogen binding and resistant to tamoxifen.

3.4.3 Guanine Nucleotide Binding Proteins: G-Proteins

G-proteins can be divided into the heterotrimeric G-proteins and small G-proteins, both are regulated by GDP/GTP binding. The heterotrimeric G-proteins contain three subunits, and at least 16 different α-subunits, 4 β-subunits, and several γ-subunits have been identified. These G-proteins are distinguished by its α-subunit, which can be classified as Gs (stimulatory), Gi (inhibitory), Go, and Gq. Gs activates adenylyl cyclase, while Gi inhibits adenylyl cyclase, and Go is linked to the activation of ion channel. The α-subunit is about 40 kDa and it has the binding site for guanine nucleotide GTP/GDP, β- and γ-subunits are 37 kDa and 8 kDa respectively. In the inactive G-proteins, the α-subunit is bound to GDP and the three subunits form a tight complex. When the ligand binds to the receptor, GDP is dissociated from the α-subunit and GTP is recruited to bind the α-subunit. The GTP-bound α-subunit then dissociates from the β- and α-subunits, resulting in activation of G-proteins. The dissociated GTP-bound α-subunit may activate the adenylate cyclase to increase cAMP level, to activate PLCβ1 and PLCβ2, and to activate the RAS-RAF-MEK-MAPK pathway to promote cell proliferation (Fig. 14). The dissociated β- and α-subunit complex may activate PLCβ2. GTP hydrolysis is catalyzed by the intrinsic GTPase activity of the α-subunit. Examples of oncogenes related to this pathway include the Gq-protein α-subunit and PLCβ.

The RAS and RHO proteins are small G-proteins (small GDP/GTP binding proteins, 21-kDa) containing only one polypeptide. The activity of small G-proteins is determined by the ratio of GTP/GDP-bound forms in the cells, and the switch of these two forms is regulated by guanosine nucleotide exchange factors (GEFs), GTPase activating proteins (GAPs), and guanosine nucleotide dissociation inhibitor (GDIs). While GEFs promote GTP binding and activation of

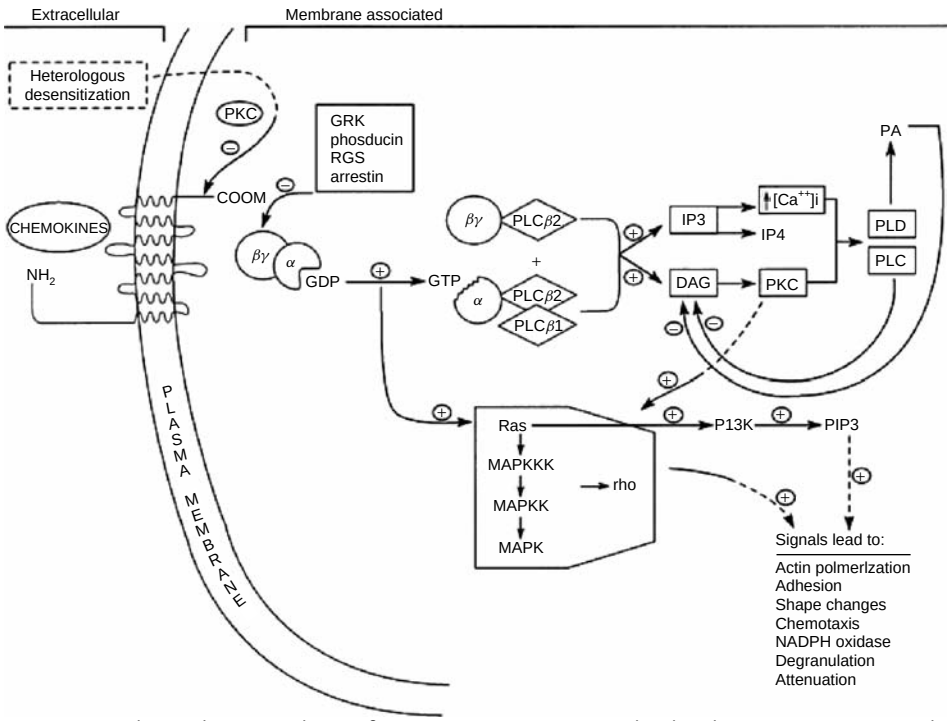

Fig. 14 Signal transduction pathway of G-protein-coupled receptors. The heterotrimeric G-protein-coupled receptors are seven-transmembrane protein receptors. Upon receptor activation by its ligand binding, the GTP-bound α-subunit dissociates from the β- and γ-subunits, resulting in activation of G-proteins. The dissociated GTP-bound α-subunit may activate the adenylate cyclase to increase cAMP level and to activate PLC-β1 and PLC-β2. The dissociated β- and α-subunit complex may activate PLCβ-2. RAS-RAF-MEK-MAPK pathway may also be regulated by activated G-proteins. (From Teresa Krakauer, Jan Vilcek, Joost J. Oppenheim. (2003). Fundamental Immunology, reprinted by permission of Lippincott Williams & Wilkins).

small G-proteins, GAPs inactivate small G-proteins by activating the intrinsic GTPase activity to hydrolyze GTP to GDP. In addition to SOS, as described above, the other GEFs for RAS activation include RASGRF1/2. RASGFP, CNRASGEF, and PLC-γ; and the GAPs for RAS inactivation include p120GAP, NF-1GAP, GAP1m, p135SYNGAP. RAS-mediated transformation involves RAF-MAPK pathway (see Sec. 3.4.1), but RAF activation is dispensable for RAS-mediated transformation in NIH 3T3 fibroblasts. This could be attributed to other RAS effector proteins that include PI3K, PLC-ε, RALGDS, RIN 1, AF-6, NORE1, and RASSF1. Activation of PI3K-AKT pathway by RAS may be mediated by interaction of p110 subunit with GTP-bound RAS.

The RHO family members are involved in cell invasion and motility and thus tumor metastasis. Cell migration and invasion involves dynamic changes in extracellular matrix, cell surface structures including focal adhesions, interaction with neighboring cells, and the intracellular motile machinery including actin stress fiber and myosin reorganization. A variety

of growth factors, motility factors, and extracellular matrix components have been shown to promote cell migration and invasion. The small GTPases of the RHO family have been shown to play pivotal roles in these signaling pathways, leading to metastasis. RHO-like GTPase family includes RHO (A, B, C), RAC (−1, 2, 3), CDC42 (CDC42Hs, G25K), and other members (RHO-6, 7, D, E, G, TC10, TTF). More than 30 GEFs and more than 20 GAPs that are involved in regulating RHO-like GTPases have been identified. Among the RHO-like family members, RHOA is upregulated in highly metastatic colon cancer, and overexpression of RHO C promotes cell motility and invasiveness in melanoma cells.

3.4.4 The JAK-STAT Signaling Pathway

JAK (Janus kinase) family proteins are associated with the receptors for interferons, cytokines, growth hormone, and the v-ABL oncoprotein. STATs are major effectors for JAK kinase–mediated signaling and gene expression. There are at least four members in the JAK kinase family, TYK2, JAK1, JAK2, and JAK3. Interferons have antitumor activities, and JAK family kinases are critical mediators for interferon-triggered signaling pathways. On the other hand, constitutive activation of JAKs and STATs are also involved in the malignant transformation of many types of cancers. In addition to STAT activation, other signaling pathways that may be activated by JAK include RAS-RAF-MEK-MAPK pathway, PI3K-AKT pathway, and activation of BCL-2 family members.

STATs (STAT 1–6) are transcription factors activated in response to cytokines as well as some growth factors. STATs function in transmitting signals from cell surface receptors to the cytoplasm and then translocating to the nucleus, where they act to regulate gene expression. STAT tyrosine phosphorylation is required for the biological function of STATs. This occurs when cytokines such as interleukin 6 (IL-6) and interferons or growth factors such as PDGF and EGF bind their respective receptors, which results in STAT protein recruitment to the inner surface of the plasma membrane in the vicinity of the cytoplasmic portion of the receptors. Tyrosine kinases that are known to phosphorylate STATs are non-RTKs such as SRC, JAK1, and JAK2. Several lines of evidence have implicated that some STAT family members, in particular, STAT3 and STAT5, are involved in malignant transformation and tumor cell survival. STAT3 involvement in oncogenesis is most thoroughly characterized. STAT3 is found constitutively tyrosine-phosphorylated (Y^{705}) and activated in many human cancers. This abnormal activation of STAT3 is prevalent in breast, lung, pancreatic, head and neck, brain, and prostate carcinomas. Recent studies have found that STAT3 is also constitutively activated in a significant fraction (74%) of malignant ovarian adenocarcinomas and ovarian cancer cell lines, but not in normal ovarian surface epithelial cells. Oncostatin M and IL-6, the JAK-STAT activators, have also been found to be overexpressed in ovarian cancer. Selective inhibition of JAK1/2, with a tyrphostin known as AG490, leads to enhanced apoptosis induced by cisplatin in ovarian cancer cells. Inhibition of STAT3 using antisense oligonucleotides and dominant negative STAT3 significantly suppresses the growth of ovarian cancer cells harboring constitutively active STAT3. Recently, Blaskovich and colleagues identified a small molecule from the National Cancer Institute Diversity Set that disrupts aberrant STAT3 signaling. This compound, JSI-124 or Cucurbitacin I, rapidly suppresses the levels

of phosphorylated STAT3 in SRC transformed cells. Future studies are likely to assess the therapeutic efficacy of this agent in a number of cancers that express constitutively active STAT3.

3.4.5 Other Protein Serine-threonine Kinases

3.4.5.A **PKC (Protein kinase C)** PKC belongs to the AGC family, which includes PKA, PKB/AKT, PKC, and RSKs. There are at least 10 members in the PKC family, the classical PKCs include PKC-α, PKC-β1, PKC-β2, and PKC-γ; the nonclassical PKCs include PKC-δ, PKC-ε, PKC-η, and PKC-θ; and atypical PKCs include PKC-ζ and PKC-λ. The classical PKCs require both DAG and Ca2+ for activation, while nonclassical PKCs only require DAG for activation. DAG and Ca2+ can be produced by PLCs that can be activated by G-proteins or RTKs. In addition, PKC is also activated by the tumor-promoting phorbol esters. Some of the PKC members, including PKC-β1, PKC-γ, PKC-ε, and PKC-ζ have been identified as being oncogenic.

3.4.5.B **MOS** MOS expression is restricted to germ cells and may be involved in meiotic cell cycle. MOS protein does not contain a regulatory region and may be active constitutively, thus abnormal overexpression of MOS is a very potent oncogene. Targets of MOS include MAPK, which may account for its transforming activity.

3.4.6 Antiapoptotic Proteins

3.4.6.A **BCL-2 protein family and oncogenesis** As described above (see Sec. 2.2), senescent or damaged cells have to be cleared to maintain tissue homeostasis, and mitochondria-mediated apoptosis is a major apoptotic mechanism. In response to environmental stress and genomic DNA damage signals, release of cytochrome C and SMAC from the mitochondria into the cytoplasm is essential for the execution of the caspase-mediated apoptotic processes. The BCL-2 family members function as gatekeepers for the exit of cytochrome C and SMAC, and therefore are key regulators for cell life or death. Some of the BCL-2 family proteins are antiapoptotic and oncogenic, while others are proapoptotic and suppression of their function might be oncogenic. The biochemical mechanism by which BCL-2 family members control cytochrome C release is not completely understood.

BCL-2 was cloned because of t(14;18) chromosome translocation in B lymphoma, which results in BCL-2 overexpression and therefore inhibition of apoptosis. Members of the BCL-2 family may possess up to four conserved α-helical domains, designated BH1, BH2, BH3, and BH4 (BH represents BCL-2 homology). The mammalian antiapoptotic BCL-2 family members (including BCL-2, BCL-X_L, BCL-W, MCL-1, A1/BFL-1, BOO/DIVA, and NR-13) possess all four BH domains, and they are found in the outer mitochondrial membrane, endoplasmic reticulum, or outer nuclear membrane. Aberrant activations of these proteins are oncogenic. Proapoptotic BCL-2 family members could be multidomain proteins (including BAX, BAK, and BOK/MTD) or BH3-only proteins (including BID, BAD, BIK/NBK, BLK, HRK, BIM/BOD, BNIP3, NIX, NOXA). Many of the proapoptotic molecules, especially the "BH3 domain only" members are localized to the cytosol or cytoskeleton and undergo posttranslational modification after a death signal, which allows them to

target and integrate into the mitochondrial membrane to initiate mitochondrial dysfunction. For example, BAD becomes dephosphorylated after death stimulus, resulting in the interaction of the BH3 domain of active BAD with the hydrophobic pocket of BCL-XL at the mitochondrial membrane. RSK and AKT can inactivate BAD by phosphorylating BAD at ser^{112} and ser^{136} to promote cell survival. Double knockout, but not single knockout, of BAK and BAX are resistant to apoptosis, suggesting that BAK and BAX might be the intermediate step essential for mitochondrial dysfunction.

3.4.6.B **Inhibitors of apoptosis protein (IAP) family** The IAP (inhibitors of apoptosis protein) proteins include c-IAP1, c-IAP2, XIAP, and survivin. They can inhibit the activation of caspases and apoptosis, and, therefore, may promote tumorigenesis when abnormally activated. The conserved BIR (baculovirus IAP repeat) domain of IAPs is involved in the antiapoptotic activity of IAPs. Survivin is expressed in almost all tumor types and it is involved in both antiapoptosis and cell cycle.

3.4.7 Chaperone Proteins

Heat-shock proteins (HSPs) are induced by environmental stress that might be involved in protein folding, stability, and degradation. They might be involved in tumorigenesis by affecting their client proteins, such as AKT, HER-2, and KIT, that are involved in cell survival or apoptosis. HSP90 has been shown to promote AKT activation by stabilizing AKT.

JNK is a MAPK activated by stress signals (e.g. UV, heat shock, osmotic shock, starvation) that initiate the mitochondrial protein release resulting in apoptosis. Proteins that inhibit JNK activation might be oncogenic, and these proteins include HSPs, JNK interacting proteins (JIPs), GST, EVI, and γ-synuclein. γ-Synuclein, also termed breast cancer–specific gene 1 (*BCSG1*), is aberrantly expressed in the vast majority (70%) of late-stage breast and ovarian cancers. Pan and colleagues have shown that γ-synuclein may promote cell survival and inhibit stress-induced and chemotherapeutic drug–induced apoptosis by suppressing JNK activation and promoting ERK activation. AKT is also activated by γ-synuclein and is involved in suppression of JNK activation and inhibition of apoptosis. The inhibition on JNK activation could be executed at different levels of JNK signaling pathway. JIPs could interact with upstream kinases, including upstream MKK7, MLK, RHO/GEF, and JNK phosphatase MKP-7. GSTs may function as suppressors to keep the JNK basal activity at a low level. JIP-1 can bind to JNK and block its translocation from cytoplasm into nucleus to activate its nuclear substrate such as c-JUN and ATF; EVI-1 is a transcription factor that blocks JNK function in the nucleus.

3.4.8 Oncogenes Associated with Cancer Cell Adhesion and Motility

One feature of cancer cells is anchorage-independent growth, and genetic or epigenetic alterations of oncoproteins involved in integrin and cadherin initiated signal transductions play significant roles in these processes. These include FAK, SRC, ABL, the small G-protein RHO families, as well as the MAPK and PI3K-AKT pathways. The other pathways include CAS/CRK and the low M(r) protein-tyrosine phosphatase (PTP). Cell migration and invasion involves dynamic changes in extracellular matrix, cell

surface structures, including focal adhesions, interaction with neighboring cells, and the intracellular motile machinery including actin stress fiber and myosin reorganization. The small GTPases of the RHO family have been shown to play pivotal roles in these signal transduction pathways leading to metastasis.

3.5 Transcription Factors and Nuclear Oncogenes

3.5.1 *FOS* and *JUN* Families

JUN family members include c-*JUN*, *JUN*-B, and *JUN*-D. *FOS* family members include c-*FOS*, *FOS*-B, *FRA*-1, and *FRA*-2. JUN family members and the FOS family members form the AP-1 transcription factor. Target genes of AP-1 include growth factors and phorbol ester inducible genes that are involved in cell proliferation and transformation. These target genes in turn can activate JUN and FOS by signaling cascades. FOS and JUN may become oncogenic by overexpression or alteration in protein structure.

3.5.2 c-*MYC*

There are three members identified in the *MYC* family, c-*MYC*, N-*MYC*, and L-*MYC*. N-*MYC* was originally identified in neuroblastoma cells. c-*MYC* is one of the immediate-early responsive genes, it is induced rapidly in the G_1-phase, and its expression is maintained at high levels throughout the cell cycle. Instead of forming homodimers, MYC form heterodimers with MAX to bind to DNA sequences and activate the transcription of MYC-target genes. In addition to forming MYC-MAX heterodimers, MAX can form MAX-MAX homodimers or MAX-MAD heterodimers. MAX-MAX or MAX-MAD dimers recognize the same sequence for binding (the E-box, with the consensus sequence of $CA^{T}/_{C}GTG$) as MYC-MAX, but function as repressors of transcription. Mechanisms of MYC activation include overexpression, gene amplification, and mutation. Different MYC members have been associated with many different cancer types. MYC may have synergistic role with Ha-RAS in tumorigenesis, as evidenced by tumor formation of rat embryo fibroblasts transfected with both Ha-RAS and MYC, but not transfected with Ha-RAS alone.

3.5.3 *REL* Oncogene: the NFκB Family Proteins

The *REL* oncogene belongs to The *NFκB* family, which includes *NF-κB*1 (p105/p50), *NF-κB*2 (p100/p52), *REL*, *RELA* (p65), and *RELB*. While mature NF-κB proteins are derived from their precursor proteins by proteolytic cleavage, the REL proteins are not posttranslationally processed. The REL/NF-κB proteins have a conserved N-terminal region for DNA binding and dimerization, and they can form homo- or heterodimers. Usually REL/NF-κB dimers are associated with an inhibitory subunit (IκB) that retains them in the cytoplasm. When cells are stimulated and the IKKs (IκB kinases) are activated, IKKs phosphorylate IκB, resulting in dissociation of REL/NF-κB from IκB and translocation of REL/NF-κB into the nucleus. In the nucleus, REL/NF-κB bind to the REL/NF-κB consensus sequences and activate the transcription of the target genes. The upstream kinases for IKKs phosphorylation include the AGC kinase such as PKB/AKT, PKC, and RSK1. The target genes of NFκB that are involved in cell survival include IAP proteins (c-IAP1, c-IAP2, and XIAP), and GADD45β. GADD45β overexpression can inhibit JNK activation, and XIAP (X-linked inhibitor of apoptosis) can inhibit caspase activation.

3.5.4 Estrogen Receptor and Other Steroid Hormone Receptors

The receptors for steroid hormones (e.g. estrogen, progesterone, glucocorticoids, and testosterone) are located in the cytoplasm, and are translocated into the nucleus when bound by the steroid hormones. The expression of the target genes of steroid hormone receptors may account for the oncogenic potential of these receptors. The receptors for steroid-related hormones, such as thyroid hormones and retinoid acids, may also be converted to oncogenes. Examples include the *ERBA* gene, which encodes the receptor for the thyroid hormone.

The great majority (80–90%) of ERs (estrogen receptor) are in the cytoplasm and nucleus. ER can either be activated by the binding of its ligand estrogen or by phosphorylation of ERα/ser^{167} by AKT or RSK1. Hence, there are two functions of ER, the estrogen-independent activation function 1 (AF-1) and the estrogen-dependent activation function 2 (AF-2). ERs mediate most of the biological effects by binding to estrogen response elements in the promoter region of target genes, and inhibition of ER by antiestrogen such as tamoxifen inhibits the growth of ER-positive breast cancers by reducing the expression of estrogen-regulated genes. The target genes of ERs include *BCL-2, EGFR, ERBB2*, macrophage inhibitory cytokine 1, *PR* (progesterone receptor), and *pS2*.

Although a great majority of ERs are in the cytoplasm and nucleus, a small fraction of ERs are present in cell surface and are involved in the nongenomic effect of ER. In ER-positive MCF-7 cells, estradiol (17-βE2) can rapidly ($\leq$10 min) activate AKT1 via ERBB2; this effect is dependent on ER and can be abrogated by the PI3K inhibitor or 4-hydroxytamoxifen. The activated ER/ERBB2 may lead to rapid phosphorylation of SHC and SHC/GRB2/SOS complex formation, as well as MAPK activation in MCF-7 cells. The nongenomic effect of ER could be mediated by the activation of PI3K-AKT and MAPK pathways.

3.5.5 Homeodomain-containing Proteins

Homeodomain-containing genes or *HOX* genes are transcription factors that are key regulators of cell proliferation and apoptosis for pattern formation in embryogenesis. These genes have a conserved homeodomain that can bind DNA sequences with a TAAT core site. Abnormal expression of some homeodomain-containing genes may contribute to tumorigenesis by deregulating cell proliferation and apoptosis.

3.5.6 *MYB* Family

The *MYB* family genes include v-*MYB*, c-*MYB*, A-*MYB*, and B-*MYB*. MYB family members are transcription factors that bind to DNA sequences (PyAACG/TG) as a monomer and are involved in the regulation of proliferation and differentiation of many different cell types. Deletion at either the N- or C-terminal has been found to convert MYB to its oncogene form. Oncogenic activation of c-MYB may be involved in malignant transformation of hematopoietic cells.

3.5.7 *ETS* Gene Family

The *ETS* family have at least 12 members, including *ETS*-1, *ETS*-2, *FLI*-1, *SPI*-1, *ERG*, and *ELK*-1. These genes may become oncogenic via viral integration, chromosomal translocation, or integration into the viral genome. ELK-1 is a target of the MAPK and activated ELK-1 is involved in the expression of the immediate-early gene expression. ETS-1 and ETS-2 can be activated

by growth factors and phorbol esters and the activated ETS-1/2 can cooperate with AP-1 (FOS/JUN) to regulate expression of target genes.

4 Identification of Oncogenes

As described above (see Sec. 2.1.), identification of oncogenes from transforming viruses plays an immeasurable role in the early days of cancer research and contributes significantly to the identification of many important oncogenes. Although very powerful, this approach is very limited because of the nature of methodology. The neoplastic properties of oncogenes are usually gain of function, and those of tumor suppressor genes are of loss of function. Therefore, approaches to identify differentially expressed genes between tumor cells and normal cells have been attempted, and several of these approaches arose along the way. These include the various differential gene expression display approaches, SAGE (serial analysis of gene expression), and DNA microarray. In the protein level, tissue-array and proteomics are in the development for the application of oncogene identification and validation. Genomic DNA transfer initially and gene transfer later into appropriate cell lines or in transgenic mice have been widely used to validate the oncogenic potentials of candidate oncogenes.

5 Oncogenes and Clinical Implications

5.1 Oncogenes and Cancer Diagnosis and Prognosis

Abnormal expression and/or activation of many of the oncogenes are important diagnostic biomarkers for cancer diagnosis and prognosis. Many of the oncogenes described in this review are applied for the diagnosis and prognosis of many different cancer types.

5.2 Oncogenes and Cancer Therapy

The most frequently used approaches in the current regimens for cancer therapy include surgery, radiotherapy, and chemotherapy. Although significant progress has been made in cancer therapy, cancer has been the second leading cause of death, second only to heart disease in the United States for many years. Therefore, better/improved diagnosis and therapy approaches are needed for the treatment and/or cure of cancer. Oncogene-based therapies have been and will continue to play a significant role in this respective. Below we will briefly discuss the potential of oncogene-targeted chemotherapy and immunotherapy.

5.2.1 Oncogene-Targeted Chemotherapy

The main targets of conventional chemotherapeutic drugs are microtubules and genomic DNA, these include the microtubule-interfering agents (e.g. vincristine, vinblastine, vinorelbine, paclitaxel, docetaxel), DNA-damaging agents such as topoisomerase inhibitors (e.g. etoposide, doxorubicin, irinotecan, and topotecan), alkylating agents (e.g. cisplatin, cyclophosphamide and ifosfamide), and some antimetabolites (e.g. gemcitabine). These agents kill the cells, cancer cells as well as normal cells, by activating the apoptotic pathways. As described in this review, the apoptotic pathways can be modified or overridden at many steps by oncogenes involved in the antiapoptotic or cell survival pathways.

Blocking the activity of one or multiple oncogenes could either sensitize cancer cells to conventional chemotherapy drugs, or kill the cancer cells by themselves.

Those potential therapeutic targets include the cell cycle regulatory proteins such as CDKs, receptor tyrosine kinases such as EGFR and HER-2/neu, farnesylated proteins such as RAS, nonreceptor cytosolic kinases such as RAF, MEK, PI3K, and AKT, nuclear oncoproteins such as MYC and ER. Among those, Tamoxifen against ER, and small-molecule EGFR tyrosine kinase inhibitors (EGFR-TKIs), such as ZD1839 (Iressa™) and OSI-774, are now in the clinic. The latter inhibit ATP binding to the tyrosine kinase domain of the receptor, thereby inhibiting tyrosine kinase activity and autophosphorylation, and subsequently blocking signal transduction from the EGFR.

The oncogene-based therapy that has garnished the most attention is imatinib mesylate (IM) (Gleevec™, STI-571). Imatinib mesylate is an oral 2-phenylaminopyrimidine derivative that acts as a selective inhibitor of the ABL, KIT, and PDGFR tyrosine kinases. Imatinib mesylate also inhibits the causative molecular translocation in CML, BCR-ABL. *In vitro* imatinib mesylate inhibits SLF-dependent human KIT expressing myeloid leukemia cell line and an SLF independent mast cell leukemia cell line, which expresses an activating juxtamembrane V560G mutation of c-KIT. In preclinical studies, imatinib has been shown to inhibit the activity of c-ABL, BCR-ABL, and PDGFRA/B. Several phase I/II clinical trials have demonstrated the efficacy of imatinib in the treatment of CML patients with the BCR-ABL translocation. Patients who progressed on interferon-α were not only shown to have clinical responses to IM, but some patients also demonstrated cytogenetic complete remissions. Current trials are evaluating the role of imatinib mesylate with chemotherapy.

Imatinib mesylate has also been shown to be a specific *in vitro* inhibitor of c-KIT phosphorylation in several tumor cell lines. It can rapidly inhibit c-KIT phosphorylation and tumor cell proliferation, without inducing apoptosis, in a human GIST cell line. The c-KIT target has been successfully exploited in phase I–III clinical trials of imatinib mesylate (STI571, Gleevec).

5.2.2 Oncogene-targeted Immunotherapy

Monoclonal antibody therapy targeted against a variety of epitopes, including oncogenes, has emerged as an important therapeutic modality for cancer. Antibodies may exert antitumor effects by inducing apoptosis, interfering with ligand–receptor interactions, or preventing the expression of proteins that are critical to the neoplastic phenotype. In addition, antibodies have been developed that target components of the tumor microenvironment, perturbing vital structures such as the formation of tumor-associated vasculature. Some antibodies target receptors whose ligands are growth factors, such as the EGF receptor. As indicated above, EGFR is overexpressed in many cancers, and therefore has prompted the development of a variety of agents targeted to the extracellular ligand-binding domain, the intracellular tyrosine kinase domain, the ligand, or to synthesis of the EGFR. These agents are being investigated as monotherapy as well as in combination with conventional therapies.

A number of monoclonal antibodies directed against the extracellular ligand-binding domain, which is thought to prevent ligand binding (e.g. IMC-C225

and ABX-EGF), have been developed. Another approach is provided by bispecific antibodies (e.g. MDX-447) that target the extracellular ligand-binding domain of the EGFR as well as epitopes on the surface of immune effector cells such as macrophage-activated killer cells. The aim is to encourage immune effector cell recruitment at the site of tumors, and hence, initiate destruction of the tumor cells and then stimulation of additional immune responses. Single-chain fragment variable (scFv) antibodies against the EGFR conjugated to toxins, such as pseudomonas endotoxin A (ETA), as well as to fungal and plant-derived toxins, have also been investigated. One of the most potent conjugates is the scFV-14e1-ETA-fusion toxin, which binds to EGFR and the mutant form, EGFRvIII, with equal affinity, but has 100-fold enhanced cytotoxicity against tumors expressing EGFRvIII compared to those expressing EGFR.

The most advanced anti-EGFR mAb in clinical development is cetuximab. Cetuximab is a genetically engineered chimeric antibody, which has been shown to block *in vitro* phosphorylation of the EGFR and induce receptor internalization, as occurs with binding of the natural ligand. This results in inhibition of cell growth and survival. It causes an increase in the expression of the cell cycle inhibitor $p27^{KIP1}$, resulting in the formation of inhibitory $p27^{KIP1}$-CDK2 complexes that prevent cells from exiting the G_1-phase of the cell cycle. Cetuximab has also been shown to induce apoptosis in some cell lines and to inhibit the production of angiogenic factors, *in vitro* and *in vivo*, as well as metastasis. Preclinical studies have also demonstrated that cetuximab can enhance the effects of cytotoxic agents and radiotherapy, for example, cetuximab, in combination with topotecan or irinotecan, increased survival of nude mice bearing human colon cancer xenografts. In recent clinical trials, cetuximab has shown promise in several solid tumors that express EGFR. For example, patients with cisplatin-resistant head and neck or colorectal cancers appear to respond upon subsequent treatment with cisplatin and cetuximab. The most common adverse event related to Cetuximab was acniform rash.

The most successful antibody therapy for solid tumors is trastuzumab, also known as Herceptin and rhuMab HER2. HER-2/neu (c-ERBB2) is a member of the EGF receptor family, has been targeted for therapy because it is overexpressed in ~25% of breast cancers, as well as other adenocarcinomas of the ovary, prostate, lung, and gastrointestinal tract. Trastuzumab is a humanized antibody derived from 4D5, a murine mAb, which recognizes an epitope on the extracellular domain of HER2/Neu. On the basis of a series of clinical trials, trastuzumab was approved by the FDA to treat women with metastatic breast cancer with HER2/neu overexpression, given either alone or in combination with paclitaxel. Herceptin also has activity in combination with vinorelbine, docetaxel, cisplatin, and the combination of gemcitabine and paclitaxel. The use of Herceptin for breast cancer is considered a therapeutic success. Although other adenocarcinomas may overexpress HER2/Neu, clinical trials have not documented consistent therapeutic successes in other cancers. Therapeutic antibodies have made the transition from concept to clinical reality over the past two decades. Many are now being tested as adjuvant or first-line therapies to assess their efficacy in improving or prolonging survival. Although new tumor-specific antigens are

being explored, oncogenes have proven to be the most effective molecular targets for therapy.

6 Perspectives

Carcinogenesis is a multistep process that involves alterations in many specific genes. The normal cell has multiple independent mechanisms that regulate its growth and differentiation and several separate events are required to override these control mechanisms. The fundamental mechanisms underlying the genetic basis of cancer are constantly being defined and involve alterations in proto-oncogenes, tumor suppressor genes, and mismatch repair (MMR) (also known as DNA repair genes, or ''caretaker'' genes). Malignancy appears, when these genes suffer regulatory perturbations and special types of mutational damage that can disrupt this balance, thus altering normal development and differentiation. However, the combination of changes both necessary and sufficient to induce malignant transformation varies greatly from tumor to tumor and from individual to individual. This type of heterogeneity is primarily the basis for the generally poor success rate in effectively treating most types of cancer. Because oncogenes are normally upregulated and constitutively activated in cancers, as opposed to tumor suppressors or *MMR* genes, they have become the primary target for the development of novel antitumor therapies. Current clinical trials with molecular targeted agents such as Herceptin and Gleevec give hope that by better understanding the role of oncogenes in the pathogenesis of cancer, a cure for many forms of the disease may not be that far behind.

See also Genetics and Molecular Biology of Lung Cancer; Intracellular Signaling in Cancer; Liver Cancer, Molecular Biology of; Mutagenesis, Malignancy and Genome Instability; Neoplastic Disease Diagnosis, DNA in; Oncology, Molecular.

Bibliography

Books and Reviews

Cooper, G.M. (1995) *Oncogenes*, Jones Sand Bartlett Publishers, Boston.

Hesketh, R. (1994) *The Oncogene Handbook*, Academic Press, London, San Diego.

Hesketh, R. (1997) *The Oncogene and Tumour Suppressor Gene Factsbook*, 2nd edition, Academic Press, San Diego.

Peters, G., Vousden, K.H. (1997) *Oncogenes and Tumour Suppressors*, IRL Press at Oxford University Press, Oxford, New York.

Rak, J. (2003) *Oncogene-Directed Therapies*, Human Press, Totowa, New Jersey.

Primary Literature

Adams, J.M., Cory, S. (2001) Life-or-death decisions by the Bcl-2 protein family, *Trends Biochem. Sci.* **26**, 61–66.

Adrain, C., Creagh, E.M., Martin, S.J. (2001) Apoptosis-associated release of Smac/DIABLO from mitochondria requires active caspases and is blocked by Bcl-2, *EMBO J.* **20**, 6627–6636.

Albanell, J., Baselga, J. (1999) Trastuzumab, a humanized anti-HER2 monoclonal antibody, for the treatment of breast cancer, *Drugs Today (Barc)* **35**, 931–946.

Alessi, D.R., Andjelkovic, M., Caudwell, B., Cron, P., Morrice, N., Cohen, P., Hemmings, B.A. (1996) Mechanism of activation of protein kinase B by insulin and IGF-1, *EMBO J.* **15**, 6541–6551.

Alroy, I., Yarden, Y. (1997) The ErbB signaling network in embryogenesis and oncogenesis: signal diversification through combinatorial

ligand-receptor interactions, *FEBS Lett.* **410**, 83–86.

Altieri, D.C. (2001) The molecular basis and potential role of survivin in cancer diagnosis and therapy, *Trends Mol. Med.* **7**, 542–547.

Altieri, D.C. (2003) Validating survivin as a cancer therapeutic target, *Nat. Rev. Cancer* **3**, 46–54.

Antoniades, H.N., Galanopoulos, T., Neville-Golden, J., O'Hara, C.J. (1992) Malignant epithelial cells in primary human lung carcinomas coexpress in vivo platelet-derived growth factor (PDGF) and PDGF receptor mRNAs and their protein products, *Proc. Natl. Acad. Sci. U. S. A.* **89**, 3942–3946.

Arbel, R., Rojansky, N., Klein, B.Y., Levitzki, R., Hartzstark, Z., Laufer, N., Ben-Bassat, H. (2003) Inhibition that targets protein kinases for the treatment of ovarian carcinoma, *Am. J. Obstet. Gynecol.* **188**, 1283–1290.

Arteaga, C.L. (2002a) Epidermal growth factor receptor dependence in human tumors: more than just expression?, *Oncologist* **7**(Suppl. 4), 31–39.

Arteaga, C.L. (2002b) Overview of epidermal growth factor receptor biology and its role as a therapeutic target in human neoplasia, *Semin. Oncol.* **29**, 3–9.

Ashcroft, M., Vousden, K.H. (2001) p53 Tumor Suppressor Protein, in: D.E. Fisher (Ed.) *Tumor Suppressor Genes in Human Cancer*, Human Press, Totowa, NJ, pp. 159–181.

Auersperg, N., Ziltener, H. (2002) Re: Reproductive hormone-induced STAT3-mediated interleukin 6 action in normal and malignant human ovarian surface epithelial cells, *J. Natl. Cancer Inst.* **21**, 1255–1256.

Baell, J.B., Huang, D.C. (2002) Prospects for targeting the Bcl-2 family of proteins to develop novel cytotoxic drugs, *Biochem. Pharmacol.* **64**, 851–863.

Barbacid, M. (1994) The Trk family of neurotrophin receptors, *J. Neurobiol.* **25**, 1386–1403.

Barbacid, M. (1995) Structural and functional properties of the TRK family of neurotrophin receptors, *Ann. N. Y. Acad. Sci.* **766**, 442–458.

Bartlett, J.M., Langdon, S.P., Simpson, B.J., Stewart, M., Katsaros, D., Sismondi, P., Love, S., Scott, W.N., Williams, A.R., Lessells, A.M., Macleod, K.G., Smyth, J.F., Miller, W.R. (1996) The prognostic value of epidermal growth factor receptor mRNA expression in primary ovarian cancer, *Br. J. Cancer* **73**, 301–306.

Baselga, J. (2000) New technologies in epidermal growth factor receptor-targeted cancer therapy, *Signal* **1**, 12–21.

Baselga, J. (2001a) Clinical trials of Herceptin(R) (trastuzumab), *Eur. J. Cancer* **37**(Suppl. 1), 18–24.

Baselga, J. (2001b) Clinical trials of Herceptin(trastuzumab), *Eur. J. Cancer* **37**(Suppl. 1), S18–S24.

Baselga, J. (2001c) Phase I and II clinical trials of trastuzumab, *Ann. Oncol.* **12**(Suppl. 1), S49–S55.

Baselga, J. (2002) Why the epidermal growth factor receptor? The rationale for cancer therapy, *Oncologist* **7**(Suppl. 4), 2–8.

Baselga, J., Norton, L., Masui, H., Pandiella, A., Coplan, K., Miller, W.H., Mendelsohn, J. Jr. (1993) Antitumor effects of doxorubicin in combination with anti-epidermal growth factor receptor monoclonal antibodies, *J. Natl. Cancer Inst.* **85**, 1327–1333.

Baselga, J., Trigo, J.M., Bourhis, J. et al. (2002) Cetuximab (C225) plus cisplatin/carboplatin is active in patients (pts) with recurrent/metastatic squamous cell carcinoma of the head and neck (SCCHN) progressing on a same dose and schedule platinum-based regimen, *Proc. Am. Soc. Clin. Oncol.* **21**, 226a.

Baselga, J., Tripathy, D., Mendelsohn, J., Baughman, S., Benz, C.C., Dantis, L., Sklarin, N.T., Seidman, A.D., Hudis, C.A., Moore, J. et al. (1996) Phase II study of weekly intravenous recombinant humanized anti-p185HER2 monoclonal antibody in patients with HER2/neu-overexpressing metastatic breast cancer, *J. Clin. Oncol.* **14**, 737–744.

Basilico, C., Newman, K.M., Curatola, A.M., Talarico, D., Mansukhani, A., Velcich, A., Delli-Bovi, P. (1989) Expression and activation of the K-fgf oncogene, *Ann. N. Y. Acad. Sci.* **567**, 95–103.

Beck, D., Gross, N., Brognara, C.B., Perruisseau, G. (1995) Expression of stem cell factor and its receptor by human neuroblastoma cells and tumors, *Blood* **86**, 3132–3138.

Berclaz, G., Altermatt, H.J., Rohrbach, V., Siragusa, A., Dreher, E., Smith, P.D. (2001) EGFR dependent expression of STAT3 (but not STAT1) in breast cancer, *Int. J. Oncol.* **19**, 1155–1160.

Berger, W., Setinek, U., Mohr, T., Kindas-Mugge, I., Vetterlein, M., Dekan, G., Eckersberger, F., Caldas, C., Micksche, M. (1999) Evidence for a role of FGF-2 and FGF receptors in

the proliferation of nonsmall cell lung cancer cells, *Int. J. Cancer* **83**, 415–423.

Bergmann, U., Funatomi, H., Yokoyama, M., Beger, H.G., Korc, M. (1995) Insulin-like growth factor I overexpression in human pancreatic cancer: evidence for autocrine and paracrine roles, *Cancer Res.* **55**, 2007–2011.

Besmer, P., Murphy, J.E., George, P.C., Qiu, F.H., Bergold, P.J., Lederman, L., Snyder, H.W., Brodeur, D., Zuckerman, E.E., Hardy, W.D. Jr. (1986) A new acute transforming feline retrovirus and relationship of its oncogene v-kit with the protein kinase gene family, *Nature* **320**, 415–421.

Bienz, M., Clevers, H. (2000) Linking colorectal cancer to Wnt signaling, *Cell* **103**, 311–320.

Bjorkoy, G., Overvatn, A., Diaz-Meco, M.T., Moscat, J., Johansen, T. (1995) Evidence for a bifurcation of the mitogenic signaling pathway activated by Ras and phosphatidylcholine-hydrolyzing phospholipase C, *J. Biol. Chem.* **270**, 21299–21306.

Blair, D.G., Athanasiou, M. (2000) Ets and retroviruses – transduction and activation of members of the Ets oncogene family in viral oncogenesis, *Oncogene* **19**, 6472–6481.

Blaskovich, M.A., Sun, J., Cantor, A., Turkson, J., Jove, R., Sebti, S.M. (2003) Discovery of JSI-124 (cucurbitacin I), a selective Janus kinase/signal transducer and activator of transcription 3 signaling pathway inhibitor with potent antitumor activity against human and murine cells in mice, *Cancer Res.* **63**, 1270–1279.

Boerner, J.L., Biscardi, J.S., Parsons, S.J. (2002) Overview of Oncogenesis, in: Alison, M. (Ed.) *Cancer Handbook*, Nature Publishing Group ©2002 Macmillan Publishers Ltd, London, UK. http://www.cancerhandbook.net.

Boldrini, L., Faviana, P., Gisfredi, S., Zucconi, Y., Di Quirico, D., Donati, V., Berti, P., Spisni, R., Galleri, D., Materazzi, G. et al. (2002) Evaluation of telomerase in the development and progression of colon cancer, *Int. J. Mol. Med.* **10**, 589–592.

Boon, K., Riggins, G.J. (2003) SAGE as a strategy to isolate cancer-related genes, *Methods Mol. Biol.* **222**, 463–479.

Borner, C. (2003) The Bcl-2 protein family: sensors and checkpoints for life-or-death decisions, *Mol. Immunol.* **39**, 615–647.

Borrello, M.G., Bongarzone, I., Pierotti, M.A., Luksch, R., Gasparini, M., Collini, P., Pilotti, S., Rizzetti, M.G., Mondellini, P., De Bernardi, B. et al. (1993) Trk and ret proto-oncogene expression in human neuroblastoma specimens: high frequency of trk expression in non-advanced stages, *Int. J. Cancer* **54**, 540–545.

Bouillet, P., Strasser, A. (2002) BH3-only proteins – evolutionarily conserved proapoptotic Bcl-2 family members essential for initiating programmed cell death, *J. Cell Sci.* **115**, 1567–1574.

Bowman, T., Jove, R. (1999) STAT Proteins and cancer, *Cancer Control* **6**, 615–619.

Bowman, T., Garcia, R., Turkson, J., Jove, R. (2000) STATs in oncogenesis, *Oncogene* **19**, 2474–2488.

Brabender, J., Danenberg, K.D., Metzger, R., Schneider, P.M., Park, J., Salonga, D., Holscher, A.H., Danenberg, P.V. (2001) Epidermal growth factor receptor and HER2-neu mRNA expression in non-small cell lung cancer is correlated with survival, *Clin. Cancer Res.* **7**, 1850–1855.

Brazil, D.P., Hemmings, B.A. (2001) Ten years of protein kinase B signalling: a hard Akt to follow, *Trends Biochem. Sci.* **26**, 657–664.

Broccoli, D., Godwin, A.K. (2002) Telomere length changes in human cancer, *Methods Mol. Med.* **68**, 271–278.

Broudy, V.C. (1997) Stem cell factor and hematopoiesis, *Blood* **90**, 1345–1364.

Brown, A.M. (2001) Wnt signaling in breast cancer: have we come full circle?, *Breast Cancer Res.* **3**, 351–355.

Bruening, W., Giasson, B.I., Klein-Szanto, A.J., Lee, V.M., Trojanowski, J.Q., Godwin, A.K. (2000) Synucleins are expressed in the majority of breast and ovarian carcinomas and in preneoplastic lesions of the ovary, *Cancer* **88**, 2154–2163.

Buchdunger, E., Cioffi, C.L., Law, N., Stover, D., Ohno-Jones, S., Druker, B.J., Lydon, N.B. (2000) Abl protein-tyrosine kinase inhibitor STI571 inhibits in vitro signal transduction mediated by c-kit and platelet-derived growth factor receptors, *J. Pharmacol. Exp. Ther.* **295**, 139–145.

Buchdunger, E., Zimmermann, J., Mett, H., Meyer, T., Muller, M., Druker, B.J., Lydon, N.B. (1996) Inhibition of the Abl protein-tyrosine kinase in vitro and in vivo by a 2-phenylaminopyrimidine derivative, *Cancer Res.* **56**, 100–104.

Burgering, B.M., Coffer, P.J. (1995) Protein kinase B (c-Akt) in phosphatidylinositol-3-OH kinase signal transduction, *Nature* **376**, 599–602.

Burke, W.M., Jin, X., Huang, M., Liu, R., Reynolds, R.K., Lin, J. (2001) Inhibition of constitutively active Stat3 suppresses growth of human ovarian and breast cancer cells, *Oncogene* **20**, 7925–7934.

Burstein, H.J., Kuter, I., Campos, S.M., Gelman, R.S., Tribou, L., Parker, L.M., Manola, J., Younger, J., Matulonis, U., Bunnell, C.A. et al. (2001) Clinical activity of trastuzumab and vinorelbine in women with HER2-overexpressing metastatic breast cancer, *J. Clin. Oncol.* **19**, 2722–2730.

Burtness, B.A., Li, Y., Flood, W. (2002) Phase III trial comparing cisplatin (C) + placebo (P) to C + anti-epidermal growth factor antibody (EGF-R) C225 in patients (pts) with metastatic/recurrent head & neck cancer (HNC), *Proc. Am. Soc. Clin. Oncol.* **21**, 226a.

Calo, V., Migliavacca, M., Bazan, V., Macaluso, M., Buscemi, M., Gebbia, N., Russo, A. (2003) STAT proteins: from normal control of cellular events to tumorigenesis, *J. Cell. Physiol.* **197**, 157–168.

Campbell, R.A., Bhat-Nakshatri, P., Patel, N.M., Constantinidou, D., Ali, S., Nakshatri, H. (2001) Phosphatidylinositol 3-kinase/AKT-mediated activation of estrogen receptor alpha: a new model for anti-estrogen resistance, *J. Biol. Chem.* **276**, 9817–9824.

Cardone, M.H., Roy, N., Stennicke, H.R., Salvesen, G.S., Franke, T.F., Stanbridge, E., Frisch, S., Reed, J.C. (1998) Regulation of cell death protease caspase-9 by phosphorylation, *Science* **282**, 1318–1321.

Carnero, A. (2002) Targeting the cell cycle for cancer therapy, *Br. J. Cancer* **87**, 129–133.

Carpenter, M., Buchsbaum, D.J. (2002) Statistical growth curve modeling of tumors treated with ERBITUX™ (IMC-C225) anti-EGFR antibody, gemcitabine, and radiation, *Proc. Am. Assoc. Cancer Res.* **43**, A2386.

Chai, J., Du, C., Wu, J.W., Kyin, S., Wang, X., Shi, Y. (2000) Structural and biochemical basis of apoptotic activation by Smac/DIABLO, *Nature* **406**, 855–862.

Chan, T.O., Rittenhouse, S.E., Tsichlis, P.N. (1999) AKT/PKB and other D3 phosphoinositide-regulated kinases: kinase activation by phosphoinositide-dependent phosphorylation, *Annu. Rev. Biochem.* **68**, 965–1014.

Chandler, L.A., Sosnowski, B.A., Greenlees, L., Aukerman, S.L., Baird, A., Pierce, G.F. (1999) Prevalent expression of fibroblast growth factor (FGF) receptors and FGF2 in human tumor cell lines, *Int. J. Cancer* **81**, 451–458.

Chellappan, S.P. (2001) Viral Oncoproteins as Probes as Tumor Suppressor Function, in: Fisher, D.E. (Ed.) *Tumor Suppressor Genes in Human Cancer*, Human Press, Totowa, NJ, pp. 29–47.

Chen, W.S., Lazar, C.S., Poenie, M., Tsien, R.Y., Gill, G.N., Rosenfeld, M.G. (1987) Requirement for intrinsic protein tyrosine kinase in the immediate and late actions of the EGF receptor, *Nature* **328**, 820–823.

Cheng, A.M., Saxton, T.M., Sakai, R., Kulkarni, S., Mbamalu, G., Vogel, W., Tortorice, C.G., Cardiff, R.D., Cross, J.C., Muller, W.J., Pawson, T. (1998) Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation, *Cell* **95**, 793–803.

Cheng, J.Q., Godwin, A.K., Bellacosa, A., Taguchi, T., Franke, T.F., Hamilton, T.C., Tsichlis, P.N., Testa, J.R. (1992) AKT2, a putative oncogene encoding a member of a subfamily of protein-serine/threonine kinases, is amplified in human ovarian carcinomas, *Proc. Natl. Acad. Sci. U. S. A.* **89**, 9267–9271.

Chiarugi, P., Cirri, P., Raugei, G., Manao, G., Taddei, L., Ramponi, G. (1996) Low M(r) phosphotyrosine protein phosphatase interacts with the PDGF receptor directly via its catalytic site, *Biochem. Biophys. Res. Commun.* **219**, 21–25.

Chiarugi, P., Cirri, P., Taddei, L., Giannoni, E., Camici, G., Manao, G., Raugei, G., Ramponi, G. (2000) The low M(r) protein-tyrosine phosphatase is involved in Rho-mediated cytoskeleton rearrangement after integrin and platelet-derived growth factor stimulation, *J. Biol. Chem.* **275**, 4640–4646.

Chiarugi, P., Cirri, P., Taddei, M.L., Talini, D., Doria, L., Fiaschi, T., Buricchi, F., Giannoni, E., Camici, G., Raugei, G., Ramponi, G. (2002) New perspectives in PDGF receptor downregulation: the main role of phosphotyrosine phosphatases, *J. Cell Sci.* **115**, 2219–2232.

Cho, S.Y., Klemke, R.L. (2000) Extracellular-regulated kinase activation and CAS/Crk coupling regulate cell migration and suppress apoptosis during invasion of the extracellular matrix, *J. Cell Biol.* **149**, 223–236.

Ciardiello, F., Bianco, R., Damiano, V., De Lorenzo, S., Pepe, S., De Placido, S., Fan, Z., Mendelsohn, J., Bianco, A.R., Tortora, G. (1999) Antitumor activity of sequential treatment with topotecan and anti-epidermal growth factor receptor monoclonal antibody C225, *Clin. Cancer Res.* **5**, 909–916.

Clark, A.S., West, K., Streicher, S., Dennis, P.A. (2002) Constitutive and inducible Akt activity promotes resistance to chemotherapy, trastuzumab, or tamoxifen in breast cancer cells, *Mol. Cancer Ther.* **1**, 707–717.

Clark, E.A., Golub, T.R., Lander, E.S., Hynes, R.O. (2000) Genomic analysis of metastasis reveals an essential role for RhoC [see comments], *Nature* **406**, 532–535.

Clarke, P.A., Hostein, I., Banerji, U., Stefano, F.D., Maloney, A., Walton, M., Judson, I., Workman, P. (2000) Gene expression profiling of human colon cancer cells following inhibition of signal transduction by 17-allylamino-17-demethoxygeldanamycin, an inhibitor of the hsp90 molecular chaperone, *Oncogene* **19**, 4125–4133.

Clements, W.M., Wang, J., Sarnaik, A., Kim, O.J., MacDonald, J., Fenoglio-Preiser, C., Groden, J., Lowy, A.M. (2002) beta-Catenin mutation is a frequent cause of Wnt pathway activation in gastric cancer, *Cancer Res.* **62**, 3503–3506.

Cobleigh, M.A., Vogel, C.L., Tripathy, D., Robert, N.J., Scholl, S., Fehrenbacher, L., Wolter, J.M., Paton, V., Shak, S., Lieberman, G., Slamon, D.J. (1999) Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease, *J. Clin. Oncol.* **17**, 2639–2648.

Cooper, G.M. (1995) *Oncogenes*, Jones and Bartlett Publishers, Boston, MA.

Corn, P.G., El-Deiry, W.S. (2002) Derangement of growth and differentiation control in oncogenesis, *BioEssays* **24**, 83–90.

Coulier, F., Martin-Zanca, D., Ernst, M., Barbacid, M. (1989) Mechanism of activation of the human trk oncogene, *Mol. Cell. Biol.* **9**, 15–23.

Cox, A.D., Brtva, T.R., Lowe, D.G., Der, C.J. (1994) R-Ras induces malignant, but not morphologic, transformation of NIH3T3 cells, *Oncogene* **9**, 3281–3288.

Dang, C.V., Lewis, B.C. (1997) Role of oncogenic transcription factor c-Myc in cell cycle regulation, apoptosis and metabolism, *J. Biomed. Sci.* **4**, 269–278.

Danial, N.N., Rothman, P. (2000) JAK-STAT signaling activated by Abl oncogenes, *Oncogene* **19**, 2523–2531.

Danial, N.N., Pernis, A., Rothman, P.B. (1995) Jak-STAT signaling induced by the v-abl oncogene, *Science* **269**, 1875–1877.

Danilkovitch-Miagkova, A., Zbar, B. (2002) Dysregulation of Met receptor tyrosine kinase activity in invasive tumors, *J. Clin. Invest.* **109**, 863–867.

Datta, S.R., Brunet, A., Greenberg, M.E. (1999) Cellular survival: a play in three akts, *Genes Dev.* **13**, 2905–2927.

Davis, R.J. (2000) Signal transduction by the JNK group of MAP kinases, *Cell* **103**, 239–252.

De Silva, C.M., Reid, R. (2003) Gastrointestinal stromal tumors (GIST): c-kit mutations, CD117 expression, differential diagnosis and targeted cancer therapy with imatinib, *Pathol. Oncol. Res.* **9**, 13–19.

De Smaele, E., Zazzeroni, F., Papa, S., Nguyen, D.U., Jin, R., Jones, J., Cong, R., Franzoso, G. (2001) Induction of gadd45beta by NF-kappaB downregulates pro-apoptotic JNK signalling, *Nature* **414**, 308–313.

DeMatteo, R.P., Lewis, J.J., Leung, D., Mudan, S.S., Woodruff, J.M., Brennan, M.F. (2000) Two hundred gastrointestinal stromal tumors: recurrence patterns and prognostic factors for survival, *Ann. Surg.* **231**, 51–58.

Demetri, G.D., von Mehren, M., Blanke, C.D., Van den Abbeele, A.D., Eisenberg, B., Roberts, P.J., Heinrich, M.C., Tuveson, D.A., Singer, S., Janicek, M. et al. (2002) Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors, *N. Engl. J. Med.* **347**, 472–480.

Dhulipal, P.D. (1997) Ets oncogene family, *Indian J. Exp. Biol.* **35**, 315–322.

Di Popolo, A., Memoli, A., Apicella, A., Tuccillo, C., di Palma, A., Ricchi, P., Acquaviva, A.M., Zarrilli, R. (2000) IGF-II/IGF-I receptor pathway up-regulates COX-2 mRNA expression and PGE2 synthesis in Caco-2 human colon carcinoma cells, *Oncogene* **19**, 5517–5524.

Dickson, C., Fantl, V. (1994) Fgf-3, an oncogene in murine breast cancer, *Cancer Treat. Res.* **71**, 331–343.

DiPaola, R.S., Kuczynski, W.I., Onodera, K., Ratajczak, M.Z., Hijiya, N., Moore, J., Gewirtz, A.M. (1997) Evidence for a functional kit receptor in melanoma, breast, and lung carcinoma cells, *Cancer Gene. Ther.* **4**, 176–182.

Dolci, S., Williams, D.E., Ernst, M.K., Resnick, J.L., Brannan, C.I., Lock, L.F., Lyman, S.D., Boswell, H.S., Donovan, P.J. (1991) Requirement for mast cell growth factor for primordial germ cell survival in culture, *Nature* **352**, 809–811.

Dong, J., Phelps, R.G., Qiao, R., Yao, S., Benard, O., Ronai, Z., Aaronson, S.A. (2003) BRAF oncogenic mutations correlate with progression rather than initiation of human melanoma, *Cancer Res.* **63**, 3883–3885.

Druker, B.J. (2002) STI571 (Gleevec) as a paradigm for cancer therapy, *Trends Mol. Med.* **8**, S14–S18.

Druker, B. J., Tamura, S., Buchdunger, E., Ohno, S., Segal, G.M., Fanning, S., Zimmermann, J., Lydon, N.B. (1996) Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells, *Nat. Med.* **2**, 561–566.

Druker, B.J., Talpaz, M., Resta, D.J., Peng, B., Buchdunger, E., Ford, J.M., Lydon, N.B., Kantarjian, H., Capdeville, R., Ohno-Jones, S., Sawyers, C.L. (2001) Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia, *N. Engl. J. Med.* **344**, 1031–1037.

Du, C., Fang, M., Li, Y., Li, L., Wang, X. (2000) Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation by eliminating IAP inhibition, *Cell* **102**, 33–42.

Duesberg, P.H., Vogt, P.K. (1970) Differences between the ribonucleic acids of transforming and nontransforming avian tumor viruses, *Proc. Natl. Acad. Sci. U. S. A.* **67**, 1673–1680.

Dunant, N.M., Senften, M., Ballmer-Hofer, K. (1996) Polyomavirus middle-T antigen associates with the kinase domain of Src-related tyrosine kinases, *J. Virol.* **70**, 1323–1330.

Ekstrand, A.J., Sugawa, N., James, C.D., Collins, V.P. (1992) Amplified and rearranged epidermal growth factor receptor genes in human glioblastomas reveal deletions of sequences encoding portions of the N- and/or C-terminal tails, *Proc. Natl. Acad. Sci. U. S. A.* **89**, 4309–4313.

Engelhardt, M., Drullinsky, P., Guillem, J., Moore, M.A. (1997) Telomerase and telomere length in the development and progression of premalignant lesions to colorectal cancer, *Clin. Cancer Res.* **3**, 1931–1941.

Esteller, M. (2003) Cancer epigenetics: DNA methylation and chromatin alterations in human cancer, *Adv. Exp. Med. Biol.* **532**, 39–49.

Etienne-Manneville, S., Hall, A. (2002) Rho GTPases in cell biology, *Nature* **420**, 629–635.

Evan, G.I., Vousden, K.H. (2001) Proliferation, cell cycle and apoptosis in cancer, *Nature* **411**, 342–348.

Evans, R.J., Wyllie, F.S., Wynford-Thomas, D., Kipling, D., Jones, C.J. (2003) A P53-dependent, telomere-independent proliferative life span barrier in human astrocytes consistent with the molecular genetics of glioma development, *Cancer Res.* **63**, 4854–4861.

Fan, M., Goodwin, M., Vu, T., Brantley-Finley, C., Gaarde, W.A., Chambers, T.C. (2000) Vinblastine-induced phosphorylation of Bcl-2 and Bcl-XL is mediated by JNK and occurs in parallel with inactivation of the Raf-1/MEK/ERK cascade, *J. Biol. Chem.* **275**, 29980–29985.

Fischer-Colbrie, J., Witt, A., Heinzl, H., Speiser, P., Czerwenka, K., Sevelda, P., Zeillinger, R. (1997) EGFR and steroid receptors in ovarian carcinoma: comparison with prognostic parameters and outcome of patients, *Anticancer Res.* **17**, 613–619.

Fleming, T.P., Matsui, T., Aaronson, S.A. (1992) Platelet-derived growth factor (PDGF) receptor activation in cell transformation and human malignancy, *Exp. Gerontol.* **27**, 523–532.

Ford, H.L., Pardee, A.B. (1999) Cancer and the cell cycle, *J. Cell. Biochem. Suppl.* **32**(33), 166–172.

Foss, B., Ulvestad, E., Bruserud, O. (2001) Platelet-derived growth factor (PDGF) in human acute myelogenous leukemia: PDGF receptor expression, endogenous PDGF release and responsiveness to exogenous PDGF isoforms by in vitro cultured acute myelogenous leukemia blasts, *Eur. J. Haematol.* **67**, 267–278.

Fretto, L.J., Snape, A.J., Tomlinson, J.E., Serogy, J.J., Wolf, D.L., LaRochelle, W.J., Giese, N.A. (1993) Mechanism of platelet-derived growth factor (PDGF) AA, AB, and BB binding to alpha and beta PDGF receptor, *J. Biol. Chem.* **268**, 3625–3631.

Frolov, A., Chahwan, S., Ochs, M., Arnoletti, J.P., Pan, Z.Z., Favorova, O., Fletcher, J., von Mehren, M., Eisenberg, B., Godwin, A.K. (2003) Response markers and the molecular mechanisms of action of Gleevec in gastrointestinal stromal tumors, *Mol. Cancer Ther.* **2**, 699–709.

Furge, K.A., Zhang, Y.W., Vande Woude, G.F. (2000) Met receptor tyrosine kinase: enhanced signaling through adapter proteins, *Oncogene* **19**, 5582–5589.

Galdemard, C., Brison, O., Lavialle, C. (1995) The proto-oncogene FGF-3 is constitutively expressed in tumorigenic, but not in non-tumorigenic, clones of a human colon carcinoma cell line, *Oncogene* **10**, 2331–2342.

Garbay, C., Liu, W.Q., Vidal, M., Roques, B.P. (2000) Inhibitors of Ras signal transduction as antitumor agents, *Biochem. Pharmacol.* **60**, 1165–1169.

Garcia, R., Bowman, T.L., Niu, G., Yu, H., Minton, S., Muro-Cacho, C.A., Cox, C.E., Falcone, R., Fairclough, R., Parsons, S. et al. (2001) Constitutive activation of Stat3 by the Src and JAK tyrosine kinases participates in growth regulation of human breast carcinoma cells, *Oncogene* **20**, 2499–2513.

George, D. (2003) Targeting PDGF receptors in cancer–rationales and proof of concept clinical trials, *Adv. Exp. Med. Biol.* **532**, 141–151.

Giles, R.H., van Es, J.H., Clevers, H. (2003) Caught up in a Wnt storm: Wnt signaling in cancer, *Biochim. Biophys. Acta* **1653**, 1–24.

Giri, D., Ropiquet, F., Ittmann, M. (1999) Alterations in expression of basic fibroblast growth factor (FGF) 2 and its receptor FGFR-1 in human prostate cancer, *Clin. Cancer Res.* **5**, 1063–1071.

Glover, H.R., Brewster, C.E., Dilworth, S.M. (1999) Association between src-kinases and the polyoma virus oncogene middle T-antigen requires PP2A and a specific sequence motif, *Oncogene* **18**, 4364–4370.

Godwin, A.K., Schultz, D.C., Hamilton, T.C., Knudson, A.G., Jr. (1997) Oncogenes and tumor-suppressor genes, in: Hoskins, W.J., Perez, C.A., Young, R.C. (Eds.) *Principles and Practice of Gynecologic Oncology*, Lippincott-Raven Publishers, Philadelphia, PA, pp. 107–148.

Golay, J., Basilico, L., Loffarelli, L., Songia, S., Broccoli, V., Introna, M. (1996) Regulation of hematopoietic cell proliferation and differentiation by the myb oncogene family of transcription factors, *Int. J. Clin. Lab. Res.* **26**, 24–32.

Grandis, J.R., Drenning, S.D., Zeng, Q., Watkins, S.C., Melhem, M.F., Endo, S., Johnson, D.E., Huang, L., He, Y., Kim, J.D. (2000) Constitutive activation of Stat3 signaling abrogates apoptosis in squamous cell carcinogenesis in vivo, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4227–4232.

Greco, A., Pierotti, M.A., Bongarzone, I., Pagliardini, S., Lanzi, C., Della Porta, G. (1992) TRK-T1 is a novel oncogene formed by the fusion of TPR and TRK genes in human papillary thyroid carcinomas, *Oncogene* **7**, 237–242.

Grobelny, J.V., Godwin, A.K., Broccoli, D. (2000) ALT-associated PML bodies are present in viable cells and are enriched in cells in the G(2)/M phase of the cell cycle, *J. Cell Sci.* **113**, 4577–4585.

Hadari, Y.R., Gotoh, N., Kouhara, H., Lax, I., Schlessinger, J. (2001) Critical role for the docking-protein FRS2 alpha in FGF receptor-mediated signal transduction pathways, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8578–8583.

Haines, D.S. (1997) The mdm2 proto-oncogene, *Leuk. Lymphoma* **26**, 227–238.

Hajitou, A., Deroanne, C., Noel, A., Collette, J., Nusgens, B., Foidart, J.M., Calberg-Bacq, C.M. (2000) Progression in MCF-7 breast cancer cell tumorigenicity: compared effect of FGF-3 and FGF-4, *Breast Cancer Res. Treat.* **60**, 15–28.

Hart, C.E., Forstrom, J.W., Kelly, J.D., Seifert, R.A., Smith, R.A., Ross, R., Murray, M.J., Bowen-Pope, D.F. (1988) Two classes of PDGF receptor recognize different isoforms of PDGF, *Science* **240**, 1529–1531.

Hayashi, Y., Ishii, N., Obayashi, C., Jinnai, K., Hanioka, K., Imai, Y., Itoh, H. (1995) Thymoma: tumour type related to expression of epidermal growth factor (EGF), EGF-receptor, p53, v-erb B and ras p21, *Virchows Arch.* **426**, 43–50.

Hayashi, Y., Widjono, Y.W., Ohta, K., Hanioka, K., Obayashi, C., Itoh, K., Imai, Y., Itoh, H. (1994) Expression of EGF, EGF-receptor, p53, v-erb B and ras p21 in colorectal neoplasms by immunostaining paraffin-embedded tissues, *Pathol. Int.* **44**, 124–130.

Hayward, W.S., Neel, B.G., Astrin, S.M. (1981) Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis, *Nature* **290**, 475–480.

Hegde, P., Qi, R., Gaspard, R., Abernathy, K., Dharap, S., Earle-Hughes, J., Gay, C., Nwokekeh, N.U., Chen, T., Saeed, A.I. et al. (2001) Identification of tumor markers in models of human colorectal cancer using a 19,200-element complementary DNA microarray, *Cancer Res.* **61**, 7792–7797.

Heinrich, M.C., Rubin, B.P., Longley, B.J., Fletcher, J.A. (2002b) Biology and genetic aspects of gastrointestinal stromal tumors: KIT activation and cytogenetic alterations, *Hum. Pathol.* **33**, 484–495.

Heinrich, M.C., Griffith, D.J., Druker, B.J., Wait, C.L., Ott, K.A., Zigler, A.J. (2000) Inhibition of c-kit receptor tyrosine kinase activity by STI 571, a selective tyrosine kinase inhibitor, *Blood* **96**, 925–932.

Heinrich, M.C., Croless, C.L., Banke, C., Demetri, G.D., Joensuu, H., von Mehren, M., McGreevey, L.S., Wait, C.L., Griffin, J.D., Chen, C.J. et al. (2002a) KIT mutational status predicts clinical response to STI571 in patients with metastatic gastrointestinal stromal tumors (GISTs), *Proc. Am. Soc. Clin. Oncol.* **21**, 2a.

Heinrich, M.C., Corless, C.L., Duensing, A., McGreevey, L., Chen, C.J., Joseph, N., Singer, S., Griffith, D.J., Haley, A., Town, A. et al. (2003) PDGFRA activating mutations in gastrointestinal stromal tumors, *Science* **299**, 708–710.

Herbst, R.S., Hong, W.K. (2002a) IMC-C225, an anti-epidermal growth factor receptor monoclonal antibody for treatment of head and neck cancer, *Semin. Oncol.* **29**, 18–30.

Herbst, R.S., Hong, W.K. (2002b) Targeted therapy against the epidermal growth factor receptor, *Semin. Oncol.* **29**, 1–2.

Herbst, R.S., Shin, D.M. (2002) Monoclonal antibodies to target epidermal growth factor receptor-positive tumors: a new paradigm for cancer therapy, *Cancer* **94**, 1593–1611.

Herr, I., Debatin, K.M. (2001) Cellular stress response and apoptosis in cancer therapy, *Blood* **98**, 2603–2614.

Hesketh, R. (1994) *The Oncogene Handbook*, Academic Press, London, San Diego.

Hesketh, R. (1997) *The Oncogene and Tumour Suppressor Gene Factsbook*, 2nd edition, Academic Press, San Diego, CA.

Hirota, S., Isozaki, K., Moriyama, Y., Hashimoto, K., Nishida, T., Ishiguro, S., Kawano, K., Hanada, M., Kurata, A., Takeda, M. et al. (1998) Gain-of-function mutations of c-*kit* in human gastrointestinal stromal tumors, *Science* **279**, 577–580.

Ho, A., Dowdy, S.F. (2002) Regulation of G(1) cell-cycle progression by oncogenes and tumor suppressor genes, *Curr Opin Genet Dev* **12**, 47–52.

Huang, M., Page, C., Reynolds, R.K., Lin, J. (2000) Constitutive activation of stat 3 oncogene product in human ovarian carcinoma cells, *Gynecol. Oncol.* **79**, 67–73.

Huang, S.M., Harari, P.M. (2000) Modulation of radiation response after epidermal growth factor receptor blockade in squamous cell carcinomas: inhibition of damage repair, cell cycle kinetics, and tumor angiogenesis, *Clin. Cancer Res.* **6**, 2166–2174.

Huang, Z. (2000) Bcl-2 family proteins as targets for anticancer drug design, *Oncogene* **19**, 6627–6631.

Hubbard, S.R. (1999) Structural analysis of receptor tyrosine kinases, *Prog. Biophys. Mol. Biol.* **71**, 343–358.

Huff, S.Y., Quilliam, L.A., Cox, A.D., Der, C.J. (1997) R-Ras is regulated by activators and effectors distinct from those that control Ras function, *Oncogene* **14**, 133–143.

Hung, M.C., Lau, Y.K. (1999) Basic science of HER-2/neu: a review, *Semin. Oncol.* **26**, 51–59.

Iida, S., Yoshida, T., Naito, K., Sakamoto, H., Katoh, O., Hirohashi, S., Sato, T., Onda, M., Sugimura, T., Terada, M. (1992) Human hst-2 (FGF-6) oncogene: cDNA cloning and characterization, *Oncogene* **7**, 303–309.

Ikeguchi, M., Makino, M., Kaibara, N. (2000) Telomerase activity and p53 gene mutation in familial polyposis coli, *Anticancer Res.* **20**, 3833–3837.

Irusta, P.M., DiMaio, D. (1998) A single amino acid substitution in a WW-like domain of diverse members of the PDGF receptor subfamily of tyrosine kinases causes constitutive receptor activation, *EMBO J.* **17**, 6912–6923.

Jaye, M., Schlessinger, J., Dionne, C.A. (1992) Fibroblast growth factor receptor tyrosine kinases: molecular analysis and signal transduction, *Biochim. Biophys. Acta* **1135**, 185–199.

Jeffers, M., Schmidt, L., Nakaigawa, N., Webb, C.P., Weirich, G., Kishida, T., Zbar, B., Vande Woude, G.F. (1997) Activating mutations for the met tyrosine kinase receptor in human

cancer, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 11445–11450.

Jemal, A., Murray, T., Samuels, A., Ghafoor, A., Ward, E., Thun, M.J. (2003) Cancer statistics, 2003, *CA Cancer J. Clin.* **53**, 5–26.

Ji, H., Liu, Y.E., Jia, T., Wang, M., Liu, J., Xiao, G., Joseph, B.K., Rosen, C., Shi, Y.E. (1997) Identification of a breast cancer-specific gene, BCSG1, by direct differential cDNA sequencing, *Cancer Res.* **57**, 759–764.

Jo, H., Zhang, H., Zhang, R., Liang, P. (1998) Cloning oncogenic ras-regulated genes by differential display, *Methods* **16**, 365–372.

Johnson, D.E., Williams, L.T. (1993) Structural and functional diversity in the FGF receptor multigene family, **60**, 1–41. *Adv. Cancer Res.*

Johnsson, A., Betsholtz, C., Heldin, C.H., Westermark, B. (1986) The phenotypic characteristics of simian sarcoma virus-transformed human fibroblasts suggest that the v-sis gene product acts solely as a PDGF receptor agonist in cell transformation, *EMBO J.* **5**, 1535–1541.

Jolly, C., Morimoto, R.I. (2000) Role of the heat shock response and molecular chaperones in oncogenesis and cell death, *J. Natl. Cancer Inst.* **92**, 1564–1572.

Jouanneau, J., Bertrand, S., Ducros, E., Longuet, M., Dore, J.F., Tavitian, A. (1987) Study of oncogenic potentialities of human melanoma: identification of N-ras oncogene after DNA transfer and tumour induction, *In Vivo* **1**, 119–124.

Juin, P., Hunt, A., Littlewood, T., Griffiths, B., Swigart, L.B., Korsmeyer, S., Evan, G. (2002) c-Myc functionally cooperates with Bax to induce apoptosis, *Mol. Cell. Biol.* **22**, 6158–6169.

Kabarowski, J.H., Allen, P.B., Wiedemann, L.M. (1994) A temperature sensitive p210 BCR-ABL mutant defines the primary consequences of BCR-ABL tyrosine kinase expression in growth factor dependent cells, *EMBO J.* **13**, 5887–5895.

Kanazawa, S., Soucek, L., Evan, G., Okamoto, T., Peterlin, B.M. (2003) c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis, *Oncogene* **22**, 5707–5711.

Kantarjian, H., Sawyers, C., Hochhaus, A., Guilhot, F., Schiffer, C., Gambacorti-Passerini, C., Niederwieser, D., Resta, D., Capdeville, R., Zoellner, U. et al. (2002) Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia, *N. Engl. J. Med.* **346**, 645–652.

Kaplan, D.R., Stephens, R.M. (1994) Neurotrophin signal transduction by the Trk receptor, *J. Neurobiol.* **25**, 1404–1417.

Kaplan, D.R., Hempstead, B.L., Martin-Zanca, D., Chao, M.V., Parada, L.F. (1991) The trk proto-oncogene product: a signal transducing receptor for nerve growth factor, *Science* **252**, 554–558.

Karna, E., Surazynski, A., Orlowski, K., Laszkiewicz, J., Puchalski, Z., Nawrat, P., Palka, J. (2002) Serum and tissue level of insulin-like growth factor-I (IGF-I) and IGF-I binding proteins as an index of pancreatitis and pancreatic cancer, *Int. J. Exp. Pathol.* **83**, 239–245.

Kassis, J., Moellinger, J., Lo, H., Greenberg, N.M., Kim, H.G., Wells, A. (1999) A role for phospholipase C-gamma-mediated signaling in tumor cell invasion, *Clin. Cancer Res.* **5**, 2251–2260.

Katoh, M., Kirikoshi, H., Terasaki, H., Shiokawa, K. (2001) WNT2B2 mRNA, up-regulated in primary gastric cancer, is a positive regulator of the WNT- beta-catenin-TCF signaling pathway, *Biochem. Biophys. Res. Commun.* **289**, 1093–1098.

Kazlauskas, A., Cooper, J.A. (1990) Phosphorylation of the PDGF receptor beta subunit creates a tight binding site for phosphatidylinositol 3 kinase, *EMBO J.* **9**, 3279–3286.

Keely, P.J., Rusyn, E.V., Cox, A.D., Parise, L.V. (1999) R-Ras signals through specific integrin alpha cytoplasmic domains to promote migration and invasion of breast epithelial cells, *J. Cell Biol.* **145**, 1077–1088.

Khalili, K., Del Valle, L., Wang, J.Y., Darbinian, N., Lassak, A., Safak, M., Reiss, K. (2003) T-antigen of human polyomavirus JC cooperates with IGF-IR signaling system in cerebellar tumors of the childhood-medulloblastomas, *Anticancer Res.* **23**, 2035–2041.

Kies, M.S., Arquette, M.A., Nabell, L. et al. (2002) Final report of the efficacy and safety of the anti-epidermal growth factor antibody Erbitux (IMC-C225), in combination with cisplatin in patients with recurrent squamous cell carcinoma of the head and neck (SCCHN) refractory to cisplatin containing chemotherapy, *Proc. Am. Soc. Clin. Oncol.* **21**, 232a.

Kim, E.S., Khuri, F.R., Herbst, R.S. (2001) Epidermal growth factor receptor biology (IMC-C225), *Curr. Opin. Oncol.* **13**, 506–513.

Kinzler, K.W., Vogelstein, B. (1996) Lessons from hereditary colorectal cancer, *Cell* **87**, 159–170.

Kiyoi, H., Towatari, M., Yokota, S., Hamaguchi, M., Ohno, R., Saito, H., Naoe, T. (1998) Internal tandem duplication of the FLT3 gene is a novel modality of elongation mutation which causes constitutive activation of the product, *Leukemia* **12**, 1333–1337.

Kiyoi, H., Naoe, T., Nakano, Y., Yokota, S., Minami, S., Miyawaki, S., Asou, N., Kuriyama, K., Jinnai, I., Shimazaki, C. et al. (1999) Prognostic implication of FLT3 and N-RAS gene mutations in acute myeloid leukemia, *Blood* **93**, 3074–3080.

Kiyokawa, E., Mochizuki, N., Kurata, T., Matsuda, M. (1997) Role of Crk oncogene product in physiologic signaling, *Crit. Rev. Oncog.* **8**, 329–342.

Klein, R., Jing, S.Q., Nanduri, V., O'Rourke, E., Barbacid, M. (1991) The trk proto-oncogene encodes a receptor for nerve growth factor, *Cell* **65**, 189–197.

Klominek, J., Baskin, B., Hauzenberger, D. (1998) Platelet-derived growth factor (PDGF) BB acts as a chemoattractant for human malignant mesothelioma cells via PDGF receptor beta-integrin alpha3beta1 interaction, *Clin. Exp. Metastasis* **16**, 529–539.

Knudson, A.G., Jr. (1985) Hereditary cancer, oncogenes, and antioncogenes, *Cancer Res.* **45**, 1437–1443.

Kochhar, K.S., Iyer, A.P. (1996) Hepatocyte growth factor induces activation of Nck and phospholipase C-gamma in lung carcinoma cells, *Cancer Lett.* **104**, 163–169.

Kolch, W. (2000) Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions, *Biochem. J.* **351**, 289–305.

Konecny, G.E., Arboleda, J., Slamon, D.J., Pegram, M. (2003) Inhibition of the HER-2/neu Oncogene, in: Rak, J. (Ed.) *Oncogene-Directed Therapies*, Human Press, Totowa, NJ, pp. 331–351.

Kouhara, H., Hadari, Y.R., Spivak-Kroizman, T., Schilling, J., Bar-Sagi, D., Lax, I., Schlessinger, J. (1997) A lipid-anchored Grb2-binding protein that links FGF-receptor activation to the Ras/MAPK signaling pathway, *Cell* **89**, 693–702.

Kozma, S.C., Redmond, S.M., Fu, X.C., Saurer, S.M., Groner, B., Hynes, N.E. (1988) Activation of the receptor kinase domain of the trk oncogene by recombination with two different cellular sequences, *EMBO J.* **7**, 147–154.

Kurokawa, H., Arteaga, C.L. (2003) ErbB (HER) receptors can abrogate antiestrogen action in human breast cancer by multiple signaling mechanisms, *Clin. Cancer Res.* **9**, 511S–515S.

Kwong, L., Wozniak, M.A., Collins, A.S., Wilson, S.D., Keely, P.J. (2003) R-Ras promotes focal adhesion formation through focal adhesion kinase and p130(Cas) by a novel mechanism that differs from integrins, *Mol. Cell. Biol.* **23**, 933–949.

Lahm, H., Amstad, P., Wyniger, J., Yilmaz, A., Fischer, J.R., Schreyer, M., Givel, J.C. (1994) Blockade of the insulin-like growth-factor-I receptor inhibits growth of human colorectal cancer cells: evidence of a functional IGF-II-mediated autocrine loop, *Int. J. Cancer* **58**, 452–459.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Lawlor, M.A., Alessi, D.R. (2001) PKB/Akt: a key mediator of cell proliferation, survival and insulin responses? *J. Cell Sci.* **114**, 2903–2910.

Lee, A.V., Schiff, R., Cui, X., Sachdev, D., Yee, D., Gilmore, A.P., Streuli, C.H., Oestereich, S., Hadsell, D.L. (2003a) New mechanisms of signal transduction inhibitor action: receptor tyrosine kinase down-regulation and blockade of signal transactivation, *Clin. Cancer Res.* **9**, 516S–523S.

Lee, J.H., Han, S.U., Cho, H., Jennings, B., Gerrard, B., Dean, M., Schmidt, L., Zbar, B., Vande Woude, G.F. (2000) A novel germ line juxtamembrane Met mutation in human gastric cancer, *Oncogene* **19**, 4947–4953.

Lee, J.Y., Eom, E.M., Kim, D.S., Ha-Lee, Y.M., Lee, D.H. (2003b) Analysis of gene expression profiles of gastric normal and cancer tissues by SAGE, *Genomics* **82**, 78–85.

Lee, M.P. (2003) Genome-wide analysis of epigenetics in cancer, *Ann. N. Y. Acad. Sci.* **983**, 101–109.

LeRoith, D., Roberts, C.T., Jr. (2003) The insulin-like growth factor system and cancer, *Cancer Lett.* **195**, 127–137.

Lim, S.J., Lopez-Berestein, G., Hung, M.C., Lupu, R., Tari, A.M. (2000) Grb2 downregulation leads to Akt inactivation in heregulin-stimulated and ErbB2-overexpressing breast cancer cells, *Oncogene* **19**, 6271–6276.

Lin, E.H., Abbruzzese, J.L. (2003) Clinical Evaluations of Agents Targeting EGFR in Cancer, in: Rak, J. (Ed.) *Oncogene-Directed Therapies*, Human Press, Totowa, NJ, pp. 313–330.

Lin, J., Tang, H., Jin, X., Jia, G., Hsieh, J.T. (2002) p53 regulates Stat3 phosphorylation and DNA binding activity in human prostate cancer cells expressing constitutively active Stat3, *Oncogene* **21**, 3082–3088.

Lowe, D.G., Capon, D.J., Delwart, E., Sakaguchi, A.Y., Naylor, S.L., Goeddel, D.V. (1987) Structure of the human and murine R-ras genes, novel genes closely related to ras proto-oncogenes, *Cell* **48**, 137–146.

Luttrell, L.M., Daaka, Y., Lefkowitz, R.J. (1999) Regulation of tyrosine kinase cascades by G-protein-coupled receptors, *Curr. Opin. Cell Biol.* **11**, 177–183.

Lux, M.L., Rubin, B.P., Biase, T.L., Chen, C.J., Maclure, T., Demetri, G., Xiao, S., Singer, S., Fletcher, C.D., Fletcher, J.A. (2000) KIT extracellular and kinase domain mutations in gastrointestinal stromal tumors, *Am. J. Pathol.* **156**, 791–795.

Madesh, M., Antonsson, B., Srinivasula, S.M., Alnemri, E.S., Hajnoczky, G. (2002) Rapid kinetics of tBid-induced cytochrome c and Smac/DIABLO release and mitochondrial depolarization, *J. Biol. Chem.* **277**, 5651–5659.

Madhani, H.D. (2001) Accounting for specificity in receptor tyrosine kinase signaling, *Cell* **106**, 9–11.

Markova, B., Herrlich, P., Ronnstrand, L., Bohmer, F.D. (2003) Identification of protein tyrosine phosphatases associating with the PDGF receptor, *Biochemistry* **42**, 2691–2699.

Martin-Zanca, D., Barbacid, M., Parada, L.F. (1990) Expression of the trk proto-oncogene is restricted to the sensory cranial and spinal ganglia of neural crest origin in mouse development, *Genes Dev.* **4**, 683–694.

Maundrell, K., Antonsson, B., Magnenat, E., Camps, M., Muda, M., Chabert, C., Gillieron, C., Boschert, U., Vial-Knecht, E., Martinou, J.C., Arkinstall, S. (1997) Bcl-2 undergoes phosphorylation by c-Jun N-terminal kinase/stress-activated protein kinases in the presence of the constitutively active GTP-binding protein Rac1, *J. Biol. Chem.* **272**, 25238–25242.

Mauro, L., Salerno, M., Morelli, C., Boterberg, T., Bracke, M.E., Surmacz, E. (2003) Role of the IGF-I receptor in the regulation of cell-cell adhesion: implications in cancer development and progression, *J. Cell. Physiol.* **194**, 108–116.

Maxwell, M., Naber, S.P., Wolfe, H.J., Galanopoulos, T., Hedley-Whyte, E.T., Black, P.M., Antoniades, H.N. (1990) Coexpression of platelet-derived growth factor (PDGF) and PDGF-receptor genes by primary human astrocytomas may contribute to their development and maintenance, *J. Clin. Invest.* **86**, 131–140.

Mendelsohn, J. (2002) Targeting the epidermal growth factor receptor for cancer therapy, *J. Clin. Oncol.* **20**, 1S–13S.

Mendelsohn, J., Baselga, J. (2000) The EGF receptor family as targets for cancer therapy, *Oncogene* **19**, 6550–6565.

Meyer, G.E., Shelden, E., Kim, B., Feldman, E.L. (2001) Insulin-like growth factor I stimulates motility in human neuroblastoma cells, *Oncogene* **20**, 7542–7550.

Meyers, M.B., Shen, W.P., Spengler, B.A., Ciccarone, V., O'Brien, J.P., Donner, D.B., Furth, M.E., Biedler, J.L. (1988) Increased epidermal growth factor receptor in multidrug-resistant human neuroblastoma cells, *J. Cell. Biochem.* **38**, 87–97.

Michalides, R.J. (1999) Cell cycle regulators: mechanisms and their role in aetiology, prognosis, and treatment of cancer, *J. Clin. Pathol.* **52**, 555–568.

Miettinen, M., Sarlomo-Rikala, M., Lasota, J. (1999b) Gastrointestinal stromal tumors: recent advances in understanding of their biology, *Hum. Pathol.* **30**, 1213–1220.

Miettinen, M., Monihan, J.M., Sarlomo-Rikala, M., Kovatich, A.J., Carr, N.J., Emory, T.S., Sobin, L.H. (1999a) Gastrointestinal stromal tumors/smooth muscle tumors (GISTs) primary in the omentum and mesentery: clinicopathologic and immunohistochemical study of 26 cases, *Am. J. Surg. Pathol.* **23**, 1109–1118.

Miknyoczki, S.J., Klein-Szanto, A.J., Ruggeri, B.A. (1996) Neurotrophin-Trk receptor interactions in neoplasia: a possible role in interstitial and perineural invasion in ductal pancreatic cancer, *Crit. Rev. Oncog.* **7**, 89–100.

Miller, M., Ginalski, K., Lesyng, B., Nakaigawa, N., Schmidt, L., Zbar, B. (2001) Structural basis of oncogenic activation caused by point mutations in the kinase domain of the MET proto-oncogene: modeling studies, *Proteins* **44**, 32–43.

Mizukami, Y., Nonomura, A., Noguchi, M., Taniya, T., Koyasaki, N., Saito, Y., Hashimoto, T., Matsubara, F., Yanaihara, N. (1991) Immunohistochemical study of oncogene product ras p21, c-myc and growth factor EGF in breast carcinomas, *Anticancer Res.* **11**, 1485–1494.

Moghal, N., Sternberg, P.W. (1999) Multiple positive and negative regulators of signaling by the EGF-receptor, *Curr. Opin. Cell Biol.* **11**, 190–196.

Mora, L.B., Buettner, R., Seigne, J., Diaz, J., Ahmad, N., Garcia, R., Bowman, T., Falcone, R., Fairclough, R., Cantor, A. et al. (2002) Constitutive activation of Stat3 in human prostate tumors and cell lines: direct inhibition of Stat3 signaling induces apoptosis of prostate cancer cells, *Cancer Res.* **62**, 6659–6666.

Moscat, J., Cornet, M.E., Diaz-Meco, M.T., Larrodera, P., Lopez-Alanon, D., Lopez-Barahona, M. (1989) Activation of phosphatidylcholine-specific phospholipase C in cell growth and oncogene transformation, *Biochem. Soc. Trans.* **17**, 988–991.

Moscatello, D.K., Holgado-Madruga, M., Godwin, A.K., Ramirez, G., Gunn, G., Zoltick, P.W., Biegel, J.A., Hayes, R.L., Wong, A.J. (1995) Frequent expression of a mutant epidermal growth factor receptor in multiple human tumors, *Cancer Res.* **55**, 5536–5539.

Mueller, P.R., Coleman, T.R., Kumagai, A., Dunphy, W.G. (1995) Myt1: a membrane-associated inhibitory kinase that phosphorylates Cdc2 on both threonine-14 and tyrosine-15, *Science* **270**, 86–90.

Murphy, C.A., Der, C.J. (2003) Ras-Mediated Deregulation of Gene Expression and Contribution to Oncogenesis, in: Rak, J. (Ed.) *Oncogene-Directed Therapies*, Human Press, Totowa, NJ, pp. 77–99.

Muthuswamy, S.K., Gilman, M., Brugge, J.S. (1999) Controlled dimerization of ErbB receptors provides evidence for differential signaling by homo- and heterodimers, *Mol. Cell. Biol.* **19**, 6845–6857.

Nakagawara, A. (2001) Trk receptor tyrosine kinases: a bridge between cancer and neural development, *Cancer Lett.* **169**, 107–114.

Nakaigawa, N., Weirich, G., Schmidt, L., Zbar, B. (2000) Tumorigenesis mediated by MET mutant M1268T is inhibited by dominant-negative Src, *Oncogene* **19**, 2996–3002.

Nakanishi, Y., Mulshine, J.L., Kasprzyk, P.G., Natale, R.B., Maneckjee, R., Avis, I., Treston, A.M., Gazdar, A.F., Minna, J.D., Cuttitta, F. (1988) Insulin-like growth factor-I can mediate autocrine proliferation of human small cell lung cancer cell lines in vitro, *J. Clin. Invest.* **82**, 354–359.

Natali, P.G., Nicotra, M.R., Sures, I., Santoro, E., Bigotti, A., Ullrich, A. (1992) Expression of c-kit receptor in normal and transformed human nonlymphoid tissues, *Cancer Res.* **52**, 6139–6143.

Needle, M.N. (2002) Safety experience with IMC-C225, an anti-epidermal growth factor receptor antibody, *Semin. Oncol.* **29**, 55–60.

Nesbit, C.E., Tersak, J.M., Prochownik, E.V. (1999) MYC oncogenes and human neoplastic disease, *Oncogene* **18**, 3004–3016.

Neudauer, C.L., McCarthy, J.B. (2003) Insulin-like growth factor I-stimulated melanoma cell migration requires phosphoinositide 3-kinase but not extracellular-regulated kinase activation, *Exp. Cell. Res.* **286**, 128–137.

Neumann, A.A., Reddel, R.R. (2002) Telomere maintenance and cancer – look, no telomerase, *Nat. Rev. Cancer* **2**, 879–884.

Nicholson, P.R., Empereur, S., Glover, H.R., Dilworth, S.M. (2001) ShcA tyrosine phosphorylation sites can replace ShcA binding in signalling by middle T-antigen, *EMBO J.* **20**, 6337–6346.

Nickerson, M.L., Weirich, G., Zbar, B., Schmidt, L.S. (2000) Signature-based analysis of MET proto-oncogene mutations using DHPLC, *Hum. Mutat.* **16**, 68–76.

Niikura, H., Sasano, H., Sato, S., Yajima, A. (1997) Expression of epidermal growth factor-related proteins and epidermal growth factor receptor in common epithelial ovarian tumors, *Int. J. Gynecol. Pathol.* **16**, 60–68.

Nistico, P., Mottolese, M., Cascioli, S., Benevolo, M., Del Bello, D., Di Modugno, F., Rubiu, O., Gentile, F.P., Botti, C., Venturo, I., Natali, P.G. (1999) Host immunosurveillance contributes to the control of erbB-2 overexpression in HLA-A2-breast-cancer patients, *Int. J. Cancer* **84**, 598–603.

Noh, D.Y., Lee, Y.H., Kim, S.S., Kim, Y.I., Ryu, S.H., Suh, P.G., Park, J.G. (1994) Elevated content of phospholipase C-gamma 1 in colorectal cancer tissues, *Cancer* **73**, 36–41.

Noh, D.Y., Kang, H.S., Kim, Y.C., Youn, Y.K., Oh, S.K., Choe, K.J., Park, I.A., Ryu, S.H., Suh, P.G. (1998) Expression of phospholipase

C-gamma 1 and its transcriptional regulators in breast cancer tissues, *Anticancer Res.* **18**, 2643–2648.

Nomura, M., Kaji, A., Ma, W.Y., Zhong, S., Liu, G., Bowden, G.T., Miyamoto, K.I., Dong, Z. (2001) Mitogen- and stress-activated protein kinase 1 mediates activation of Akt by ultraviolet B irradiation, *J. Biol. Chem.* **276**, 25558–25567.

Nonomura, A., Ohta, G., Nakanuma, Y., Izumi, R., Mizukami, Y., Matsubara, F., Hayashi, M., Watanabe, K., Takayanagi, N. (1988) Simultaneous detection of epidermal growth factor receptor (EGF-R), epidermal growth factor (EGF) and ras p21 in cholangiocarcinoma by an immunocytochemical method, *Liver* **8**, 157–166.

Noonberg, S.B., Benz, C.C. (2000) Tyrosine kinase inhibitors targeted to the epidermal growth factor receptor subfamily: role as anticancer agents, *Drugs* **59**, 753–767.

Normanno, N., Campiglio, M., De, L.A., Somenzi, G., Maiello, M., Ciardiello, F., Gianni, L., Salomon, D.S., Menard, S. (2002) Cooperative inhibitory effect of ZD1839 (Iressa) in combination with trastuzumab (Herceptin) on human breast cancer cell growth, *Ann. Oncol.* **13**, 65–72.

Ochs, M.F., Godwin, A.K. (2003) Microarrays in cancer: research and applications, *Biotechniques* **34**, 4–15.

Ogata, M., Naito, Z., Tanaka, S., Moriyama, Y., Asano, G. (2000) Overexpression and localization of heat shock proteins mRNA in pancreatic carcinoma, *J. Nippon Med. Sch.* **67**, 177–185.

Ohmichi, M., Decker, S.J., Pang, L., Saltiel, A.R. (1991) Phospholipase C-gamma 1 directly associates with the p70 trk oncogene product through its src homology domains, *J. Biol. Chem.* **266**, 14858–14861.

Okamoto, I., Kenyon, L.C., Emlet, D.R., Mori, T., Sasaki, J.-I., Hirosako, S., Ichikawa, Y., Kishi, H., Godwin, A.K., Yoshioka, M., Suga, M., Matsumoto, M., Wong, A.J. (2003) Expression of constitutively activated EGFRvIII in non-small cell lung cancer, *Cancer Sci.* **94**, 50–56.

Orr, F.W., Wang, H.H. (2001) Tumor cell interactions with the microvasculature: a rate-limiting step in metastasis, *Surg. Oncol. Clin. N. Am.* **10**, 357–381, ix–x.

Osada, H., Takahashi, T. (2002) Genetic alterations of multiple tumor suppressors and oncogenes in the carcinogenesis and progression of lung cancer, *Oncogene* **21**, 7421–7434.

Pan, Z.Z., Bruening, W., Giasson, B.I., Lee, V.M., Godwin, A.K. (2002a) Gamma-synuclein promotes cancer cell survival and inhibits stress- and chemotherapy drug-induced apoptosis by modulating MAPK pathways, *J. Biol. Chem.* **277**, 35050–35060.

Pan, Z.Z., Kronenberg, M.S., Huang, D.Y., Sumoy, L., Rogina, B., Lichtler, A.C., Upholt, W.B. (2002b) MSX2 expression in the apical ectoderm ridge is regulated by an MSX2 and Dlx5 binding site, *Biochem. Biophys. Res. Commun.* **290**, 955–961.

Panayotou, G., Bax, B., Gout, I., Federwisch, M., Wroblowski, B., Dhand, R., Fry, M.J., Blundell, T.L., Wollmer, A., Waterfield, M.D. (1992) Interaction of the p85 subunit of PI 3-kinase and its N-terminal SH2 domain with a PDGF receptor phosphorylation site: structural features and analysis of conformational changes, *EMBO J.* **11**, 4261–4272.

Park, J.I., Strock, C.J., Ball, D.W., Nelkin, B.D. (2003) The Ras/Raf/MEK/extracellular signal-regulated kinase pathway induces autocrine-paracrine growth inhibition via the leukemia inhibitory factor/JAK/STAT pathway, *Mol. Cell. Biol.* **23**, 543–554.

Park, W.S., Dong, S.M., Kim, S.Y., Na, E.Y., Shin, M.S., Pi, J.H., Kim, B.J., Bae, J.H., Hong, Y.K., Lee, K.S. et al. (1999) Somatic mutations in the kinase domain of the Met/hepatocyte growth factor receptor gene in childhood hepatocellular carcinomas, *Cancer Res.* **59**, 307–310.

Patapoutian, A., Reichardt, L.F. (2001) Trk receptors: mediators of neurotrophin action, *Curr. Opin. Neurobiol.* **11**, 272–280.

Patterson, H. (1992) Approaches to proto-oncogene and tumour suppressor gene identification, **28**, 258–263.

Paweletz, C.P., Charboneau, L., Bichsel, V.E., Simone, N.L., Chen, T., Gillespie, J.W., Emmert-Buck, M.R., Roth, M.J., Petricoin, I.E., Liotta, L.A. (2001) Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front, *Oncogene* **20**, 1981–1989.

Pawson, T., Gish, G.D. (1992) SH2 and SH3 domains: from structure to function, *Cell* **71**, 359–362.

Pegram, M.D., Lipton, A., Hayes, D.F., Weber, B.L., Baselga, J.M., Tripathy, D., Baly, D., Baughman, S.A., Twaddell, T., Glaspy, J.A., Slamon, D.J. (1998) Phase II study of receptor-enhanced chemosensitivity using recombinant humanized anti-p185HER2/neu monoclonal antibody plus cisplatin in patients with HER2/neu-overexpressing metastatic breast cancer refractory to chemotherapy treatment, *J. Clin. Oncol.* **16**, 2659–2671.

Peles, E., Levy, R.B., Or, E., Ullrich, A., Yarden, Y. (1991) Oncogenic forms of the neu/HER2 tyrosine kinase are permanently coupled to phospholipase C gamma, *EMBO J.* **10**, 2077–2086.

Peng, D., Fan, Z., Lu, Y., DeBlasio, T., Scher, H., Mendelsohn, J. (1996) Anti-epidermal growth factor receptor monoclonal antibody 225 up-regulates p27KIP1 and induces G1 arrest in prostatic cancer cell line DU145, *Cancer Res.* **56**, 3666–3669.

Perrotte, P., Matsumoto, T., Inoue, K., Kuniyasu, H., Eve, B.Y., Hicklin, D.J., Radinsky, R., Dinney, C.P. (1999) Anti-epidermal growth factor receptor antibody C225 inhibits angiogenesis in human transitional cell carcinoma growing orthotopically in nude mice, *Clin Cancer Res* **5**, 257–265.

Peters, G., Vousden, K.H. (1997) *Oncogenes and Tumour Suppressors*, IRL Press at Oxford University Press, Oxford; New York.

Petit, A.M., Rak, J., Hung, M.C., Rockwell, P., Goldstein, N., Fendly, B., Kerbel, R.S. (1997) Neutralizing antibodies against epidermal growth factor and ErbB-2/neu receptor tyrosine kinases down-regulate vascular endothelial growth factor production by tumor cells in vitro and in vivo: angiogenic implications for signal transduction therapy, of solid tumors, *Am. J. Pathol.* **151**, 1523–1530.

Plass, C., Soloway, P.D. (2002) DNA methylation, imprinting and cancer, *Eur. J. Hum. Genet.* **10**, 6–16.

Pompetti, F., Pilla, D., Giancola, R. (2003) Cancer therapy: switching off oncogenes, *Bioessays* **25**, 104–107.

Porter, A.C., Vaillancourt, R.R. (1998) Tyrosine kinase receptor-activated signal transduction pathways which lead to oncogenesis, *Oncogene* **17**, 1343–1352.

Prewett, M.C., Hooper, A.T., Bassi, R., Ellis, L.M., Waksal, H.W., Hicklin, D.J. (2002) Enhanced antitumor activity of anti-epidermal growth factor receptor monoclonal antibody IMC-C225 in combination with irinotecan (CPT-11) against human colorectal tumor xenografts, *Clin. Cancer Res.* **8**, 994–1003.

Ramljak, D., Coticchia, C.M., Nishanian, T.G., Saji, M., Ringel, M.D., Conzen, S.D., Dickson, R.B. (2003) Epidermal growth factor inhibition of c-Myc-mediated apoptosis through Akt and Erk involves Bcl-xL upregulation in mammary epithelial cells, *Exp. Cell Res.* **287**, 397–410.

Reddy, E.P., Korapati, A., Chaturvedi, P., Rane, S. (2000) IL-3 signaling and the role of Src kinases, JAKs and STATs: a covert liaison unveiled, *Oncogene* **19**, 2532–2547.

Reinmuth, N., Fan, F., Liu, W., Parikh, A.A., Stoeltzing, O., Jung, Y.D., Bucana, C.D., Radinsky, R., Gallick, G.E., Ellis, L.M. (2002) Impact of insulin-like growth factor receptor-I function on angiogenesis, growth, and metastasis of colon cancer, *Lab. Invest.* **82**, 1377–1389.

Ricotti, E., Fagioli, F., Garelli, E., Linari, C., Crescenzio, N., Horenstein, A.L., Pistamiglio, P., Vai, S., Berger, M., di Montezemolo, L.C. et al. (1998) c-*kit* is expressed in soft tissue sarcoma of neuroectodermic origin and its ligand prevents apoptosis of neoplastic cells, *Blood* **91**, 2397–2405.

Ridley, A.J. (2001) Rho GTPases and cell migration, *J. Cell Sci.* **114**, 2713–2722.

Riese, D.J., Stern, D.F. 2nd (1998) Specificity within the EGF family/ErbB receptor family signaling network, *Bioessays* **20**, 41–48.

Rincon-Arano, H., Rosales, R., Mora, N., Rodriguez-Castaneda, A., Rosales, C. (2003) R-Ras promotes tumor growth of cervical epithelial cells, *Cancer* **97**, 575–585.

Robertson, S.C., Tynan, J., Donoghue, D.J. (2000) RTK mutations and human syndromes: when good receptors turn bad, *Trends Genet.* **16**, 368.

Robinson, M.J., Cobb, M.H. (1997) Mitogen-activated protein kinase pathways, **9**, 180–186. *Curr. Opin. Cell Biol.*

Roccato, E., Miranda, C., Ranzi, V., Gishizki, M., Pierotti, M.A., Greco, A. (2002) Biological activity of the thyroid TRK-T3 oncogene requires signalling through Shc, *Br. J. Cancer* **87**, 645–653.

Rosenfeldt, H.M., Hobson, J.P., Maceyka, M., Olivera, A., Nava, V.E., Milstien, S., Spiegel, S. (2001) EDG-1 links the PDGF receptor to Src and focal adhesion kinase activation leading

to lamellipodia formation and cell migration, *FASEB J.* **15**, 2649–2659.

Ross, J.S., Fletcher, J.A. (1999) HER-2/neu (c-erb-B2) gene and protein in breast cancer, *Am. J. Clin. Pathol.* **112**, S53–S67.

Rousset, D., Agnes, F., Lachaume, P., Andre, C., Galibert, F. (1995) Molecular evolution of the genes encoding receptor tyrosine kinase with immunoglobulinlike domains, *J. Mol. Evol.* **41**, 421–429.

Rubin, B.P., Singer, S., Tsao, C., Duensing, A., Lux, M.L., Ruiz, R., Hibbard, M.K., Chen, C.J., Xiao, S., Tuveson, D.A. et al. (2001) KIT activation is a ubiquitous feature of gastrointestinal stromal tumors, *Cancer Res.* **61**, 8118–8121.

Sachsenmaier, C., Sadowski, H.B., Cooper, J.A. (1999) STAT activation by the PDGF receptor requires juxtamembrane phosphorylation sites but not Src tyrosine kinase activation, *Oncogene* **18**, 3583–3592.

Sako, Y., Minoghchi, S., Yanagida, T. (2000) Single-molecule imaging of EGFR signalling on the surface of living cells, *Nat. Cell Biol.* **2**, 168–172.

Salomon, D.S., Brandt, R., Ciardiello, F., Normanno, N. (1995) Epidermal growth factor-related peptides and their receptors in human malignancies, *Crit. Rev. Oncol. Hematol.* **19**, 183–232.

Saltz, L., Meropol, N.J., Loehrer, P.J. (2002) Single agent IMC-225 (Erbitux™) has activity in CPT-11-refractory colorectal cancer (CRC) that expresses the epidermal growth factor receptor (EGFR), *Proc. Am. Soc. Clin. Oncol.* **21**, 127a.

Sato, N., Maitra, A., Fukushima, N., van Heek, N.T., Matsubayashi, H., Iacobuzio-Donahue, C.A., Rosty, C., Goggins, M. (2003) Frequent hypomethylation of multiple genes overexpressed in pancreatic ductal adenocarcinoma, *Cancer Res.* **63**, 4158–4166.

Savarese, T.M., Campbell, C.L., McQuain, C., Mitchell, K., Guardiani, R., Quesenberry, P.J., Nelson, B.E. (2002) Coexpression of oncostatin M and its receptor and evidence for STAT3 activation in human ovarian carcinomas, *Ctyokine* **17**, 324–334.

Savelyeva, L., Schwab, M. (2001) Amplification of oncogenes revisited: from expression profiling to clinical application, *Cancer Lett.* **167**, 115–123.

Sawyers, C.L., Hochhaus, A., Feldman, E., Goldman, J.M., Miller, C.B., Ottmann, O.G., Schiffer, C.A., Talpaz, M., Guilhot, F., Deininger, M.W. et al. (2002) Imatinib induces hematologic and cytogenetic responses in patients with chronic myelogenous leukemia in myeloid blast crisis: results of a phase II study, *Blood* **99**, 3530–3539.

Schaffhausen, B. (1995) SH2 domain structure and function, *Biochim. Biophys. Acta* **1242**, 61–75.

Schmidt, M., Reiser, P., Hills, D., Gullick, W.J., Wels, W. (1998) Expression of an oncogenic mutant EGF receptor markedly increases the sensitivity of cells to an EGF-receptor-specific antibody-toxin, *Int. J. Cancer* **75**, 878–884.

Schmidt, M., Maurer-Gebhard, M., Groner, B., Kohler, G., Brochmann-Santos, G., Wels, W. (1999b) Suppression of metastasis formation by a recombinant single chain antibody-toxin targeted to full-length and oncogenic variant EGF receptors, *Oncogene* **18**, 1711–1721.

Schmidt, L., Junker, K., Nakaigawa, N., Kinjerski, T., Weirich, G., Miller, M., Lubensky, I., Neumann, H.P., Brauch, H., Decker, J. et al. (1999a) Novel mutations of the MET proto-oncogene in papillary renal carcinomas, *Oncogene* **18**, 2343–2350.

Schwab, M. (1998) Amplification of oncogenes in human cancer cells, *Bioessays* **20**, 473–479.

Schwartz, M.A. (1997) Integrins, oncogenes, and anchorage independence, *J. Cell Biol.* **139**, 575–578.

Sciacca, L., Mineo, R., Pandini, G., Murabito, A., Vigneri, R., Belfiore, A. (2002) In IGF-I receptor-deficient leiomyosarcoma cells autocrine IGF-II induces cell invasion and protection from apoptosis via the insulin receptor isoform A, *Oncogene* **21**, 8240–8250.

Sciacca, L., Costantino, A., Pandini, G., Mineo, R., Frasca, F., Scalia, P., Sbraccia, P., Goldfine, I.D., Vigneri, R., Belfiore, A. (1999) Insulin receptor activation by IGF-II in breast cancers: evidence for a new autocrine/paracrine mechanism, *Oncogene* **18**, 2471–2479.

Scorrano, L., Korsmeyer, S.J. (2003) Mechanisms of cytochrome c release by proapoptotic BCL-2 family members, *Biochem. Biophys. Res. Commun.* **304**, 437–444.

Seidal, T., Edvardsson, H. (1999) Expression of c-kit (CD117) and Ki67 provides information about the possible cell of origin and clinical

course of gastrointestinal stromal tumours, *Histopathology* **34**, 416–424.

Seki, A., Yoshinouchi, M., Seki, N., Kodama, J., Miyagi, Y., Kudo, T. (2000) Detection of c-erbB-2 and FGF-3 (INT-2) gene amplification in epithelial ovarian cancer, *Int. J. Oncol.* **17**, 103–106.

Self, A.J., Caron, E., Paterson, H.F., Hall, A. (2001) Analysis of R-Ras signalling pathways, *J. Cell Sci.* **114**, 1357–1366.

Shen, R., Su, Z.Z., Olsson, C.A., Fisher, P.B. (1995) Identification of the human prostatic carcinoma oncogene PTI-1 by rapid expression cloning and differential RNA display, *Proc. Natl. Acad. Sci. U. S. A.* **92**, 6778–6782. *Proc. Natl. Acad. Sci. U. S. A.*

Sherr, C.J. (2000a) Cell cycle control and cancer, *Harvey Lect.* **96**, 73–92.

Sherr, C.J. (2000b) The Pezcoller lecture: cancer cell cycles revisited, *Cancer Res.* **60**, 3689–3695.

Shi, Y. (2001) A structural view of mitochondria-mediated apoptosis, *Nat. Struct. Biol.* **8**, 394–401.

Silva, J.M., Dominguez, G., Gonzalez-Sancho, J.M., Garcia, J.M., Silva, J., Garcia-Andrade, C., Navarro, A., Munoz, A., Bonilla, F. (2002) Expression of thyroid hormone receptor/erbA genes is altered in human breast cancer, *Oncogene* **21**, 4307–4316.

Sinn, E., Muller, W., Pattengale, P., Tepler, I., Wallace, R., Leder, P. (1987) Coexpression of MMTV/v-Ha-ras and MMTV/c-myc genes in transgenic mice: synergistic action of oncogenes in vivo, *Cell* **49**, 465–475.

Slamon, D.J., Godolphin, W., Jones, L.A., Holt, J.A., Wong, S.G., Keith, D.E., Levin, W.J., Stuart, S.G., Udove, J., Ullrich, A. et al. (1989) Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer, *Science* **244**, 707–712.

Small, D., Levenstein, M., Kim, E., Carow, C., Amin, S., Rockwell, P., Witte, L., Burrow, C., Ratajczak, M.Z., Gewirtz, A.M. (1994) STK-1, the human homolog of Flk-2/Flt-3, is selectively expressed in CD34+ human bone marrow cells and is involved in the proliferation of early progenitor/stem cells, *Proc. Natl. Acad. Sci. U. S. A.* **91**, 459–463.

Smith, K., Bui, T.D., Poulsom, R., Kaklamanis, L., Williams, G., Harris, A.L. (1999) Up-regulation of macrophage wnt gene expression in adenoma-carcinoma progression of human colorectal cancer, *Br. J. Cancer* **81**, 496–502.

Solit, D.B., Basso, A.D., Olshen, A.B., Scher, H.I., Rosen, N. (2003) Inhibition of heat shock protein 90 function down-regulates Akt kinase and sensitizes tumors to Taxol, *Cancer Res.* **63**, 2139–2144.

Solit, D.B., Zheng, F.F., Drobnjak, M., Munster, P.N., Higgins, B., Verbel, D., Heller, G., Tong, W., Cordon-Cardo, C., Agus, D.B. et al. (2002) 17-Allylamino-17-demethoxygeldanmycin induces the degradation of androgen receptor and HER-2/neu and inhibits the growth of prostate cancer xenografts, *Clin. Cancer Res.* **8**, 986–993.

Song, J.I., Grandis, J.R. (2000) STAT signaling in head and neck cancer, *Oncogene* **19**, 2489–2495.

Song, L., Turkson, J., Karras, J.G., Jove, R., Haura, E.B. (2003) Activation of Stat3 by receptor tyrosine kinases and cytokines regulates survival in human non-small cell carcinoma cells, *Oncogene* **22**, 4150–4165.

Song, R.X., McPherson, R.A., Adam, L., Bao, Y., Shupnik, M., Kumar, R., Santen, R.J. (2002) Linkage of rapid estrogen action to MAPK activation by ERalpha-Shc association and Shc pathway activation, *Mol. Endocrinol.* **16**, 116–127.

Songyang, Z., Shoelson, S.E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W.G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R.J. et al. (1993) SH2 domains recognize specific phosphopeptide sequences, *Cell* **72**, 767–778.

Spaventi, R., Kamenjicki, E., Pecina, N., Grazio, S., Pavelic, J., Kusic, B., Cvrtila, D., Danilovic, Z., Spaventi, S. et al. (1994) Immunohistochemical detection of TGF-alpha, EGF-R, c-erbB-2, c-H-ras, c-myc, estrogen and progesterone in benign and malignant human breast lesions: a concomitant expression, *In Vivo* **8**, 183–189.

Stehelin, D., Varmus, H.E., Bishop, J.M., Vogt, P.K. (1976) DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA, *Nature* **260**, 170–173.

Stepanova, L., Harper, J.W. (2001) Retinoblastoma Protein in Growth Control and Differentiation, in: Fisher D.E. (Ed.) *Tumor Suppressor Genes in Human Cancer*, Human Press, Totowa, NJ, pp. 131–157.

Stoica, G.E., Franke, T.F., Wellstein, A., Czubayko, F., List, H.J., Reiter, R., Morgan, E., Martin, M.B., Stoica, A. (2003b) Estradiol rapidly activates Akt via the ErbB2 signaling pathway, *Mol. Endocrinol.* **17**, 818–830.

Stoica, G.E., Franke, T.F., Moroni, M., Mueller, S., Morgan, E., Iann, M.C., Winder, A.D., Reiter, R., Wellstein, A., Martin, M.B., Stoica, A. (2003a) Effect of estradiol on estrogen receptor-alpha gene expression and activity can be modulated by the ErbB2/PI 3-K/Akt pathway, *Oncogene* **22**, 6054–6067.

Sudol, M. (1998) From Src homology domains to other signaling modules: proposal of the 'protein recognition code', *Oncogene* **17**, 1469–1474.

Sun, M., Paciga, J.E., Feldman, R.I., Yuan, Z., Coppola, D., Lu, Y.Y., Shelley, S.A., Nicosia, S.V., Cheng, J.Q. (2001) Phosphatidylinositol-3-OH Kinase (PI3K)/AKT2, activated in breast cancer, regulates and is induced by estrogen receptor alpha (ERalpha) via interaction between ERalpha and PI3K, *Cancer Res.* **61**, 5985–5991.

Sunada, H., Magun, B.E., Mendelsohn, J., Mac Leod, C.L. (1986) Monoclonal antibody against epidermal growth factor receptor is internalized without stimulating receptor phosphorylation, *Proc. Natl. Acad. Sci. U. S. A.* **83**, 3825–3829.

Suprynowicz, F.A., Baege, A., Sunitha, I., Schlegel, R. (2002) c-Src activation by the E5 oncoprotein enables transformation independently of PDGF receptor activation, *Oncogene* **21**, 1695–1706.

Suzuki, J., Kaziro, Y., Koide, H. (1998) Synergistic action of R-Ras and IGF-1 on Bcl-xL expression and caspase-3 inhibition in BaF3 cells: R-Ras and IGF-1 control distinct anti-apoptotic kinase pathways, *FEBS Lett.* **437**, 112–116.

Syed, V., Ulinski, G., Mok, S.C., Ho, S.M. (2002) Reproductive hormone-induced, STAT3-mediated interleukin 6 action in normal and malignant human ovarian surface epithelial cells, *J. Natl. Cancer Inst.* **94**, 617–629.

Takai, Y., Sasaki, T., Matozaki, T. (2001) Small GTP-binding proteins, *Physiol. Rev.* **81**, 153–208.

Talarico, D., Ittmann, M.M., Bronson, R., Basilico, C. (1993) A retrovirus carrying the K-fgf oncogene induces diffuse meningeal tumors and soft-tissue fibrosarcomas, *Mol. Cell. Biol.* **13**, 1998–2010.

Talpaz, M., Silver, R.T., Druker, B. (2001) Gleevec (formerly ST1571): an active drug in patients with Ph+ chronic myeloid leukemia in accelerated phase – updated results of a phase II study, *Blood* **98**, 845a.

Tang, G., Minemoto, Y., Dibling, B., Purcell, N.H., Li, Z., Karin, M., Lin, A. (2001) Inhibition of JNK activation through NF-kappaB target genes, *Nature* **414**, 313–317.

Terasaki, H., Saitoh, T., Shiokawa, K., Katoh, M. (2002) Frizzled-10, up-regulated in primary colorectal cancer, is a positive regulator of the WNT - beta-catenin - TCF signaling pathway, *Int. J. Mol. Med.* **9**, 107–112.

Testa, J.R., Bellacosa, A. (2001) AKT plays a central role in tumorigenesis, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10983–10985.

Thormeyer, D., Baniahmad, A. (1999) The v-erbA oncogene (review), *Int. J. Mol. Med.* **4**, 351–358.

Tognon, C., Garnett, M., Kenward, E., Kay, R., Morrison, K., Sorensen, P.H. (2001) The chimeric protein tyrosine kinase ETV6-NTRK3 requires both Ras-Erk1/2 and PI3-kinase-Akt signaling for fibroblast transformation, *Cancer Res.* **61**, 8909–8916.

Tognon, C., Knezevich, S.R., Huntsman, D., Roskelley, C.D., Melnyk, N., Mathers, J.A., Becker, L., Carneiro, F., MacPherson, N., Horsman, D. et al. (2002) Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma, *Cancer Cell* **2**, 367–376.

Tomlinson, I.P., Roylance, R., Houlston, R.S. (2001) Two hits revisited again, *J. Med. Genet.* **38**, 81–85.

Tournier, C., Hess, P., Yang, D.D., Xu, J., Turner, T.K., Nimnual, A., Bar-Sagi, D., Jones, S.N., Flavell, R.A., Davis, R.J. (2000) Requirement of JNK for stress-induced activation of the cytochrome c-mediated death pathway, *Science* **288**, 870–874.

Tran, S.E., Holmstrom, T.H., Ahonen, M., Kahari, V.M., Eriksson, J.E. (2001) MAPK/ERK overrides the apoptotic signaling from Fas, TNF, and TRAIL receptors, *J. Biol. Chem.* **276**, 16484–16490.

Tsai, C.M., Chang, K.T., Perng, R.P., Mitsudomi, T., Chen, M.H., Kadoyama, C., Gazdar, A.F. (1993) Correlation of intrinsic chemoresistance of non-small-cell lung cancer cell lines with HER-2/neu gene expression but not with ras gene mutations, *J. Natl. Cancer Inst.* **85**, 897–901.

Tsai, M., Takeishi, T., Thompson, H., Langley, K.E., Zsebo, K.M., Metcalfe, D.D., Geissler, E.N., Galli, S.J. (1991) Induction of mast

cell proliferation, maturation, and heparin synthesis by the rat c-kit ligand, stem cell factor, *Proc. Natl. Acad. Sci. U. S. A.* **88**, 6382–6386.

Tsujimoto, Y. (2003) Cell death regulation by the Bcl-2 protein family in the mitochondria, *J. Cell. Physiol.* **195**, 158–167.

Turkson, J., Jove, R. (2000) STAT proteins: novel molecular targets for cancer drug discovery, *Oncogene* **19**, 6613–6626.

Turley, E.A. (1992) Molecular mechanisms of cell motility, *Cancer Met. Rev.* **11**, 1–3.

Turner, A.M., Zsebo, K.M., Martin, F., Jacobsen, F.W., Bennett, L.G., Broudy, V.C. (1992) Nonhematopoietic tumor cell lines express stem cell factor and display c-kit receptors, *Blood* **80**, 374–381.

Umekita, Y., Enokizono, N., Sagara, Y., Kuriwaki, K., Takasaki, T., Yoshida, A., Yoshida, H. (1992) Immunohistochemical studies on oncogene products (EGF-R, c-erbB-2) and growth factors (EGF, TGF-alpha) in human breast cancer: their relationship to oestrogen receptor status, histological grade, mitotic index and nodal status, *Virchows Arch. A Pathol. Anat. Histopathol.* **420**, 345–351.

Van Aelst, L., D'Souza-Schorey, C. (1997) Rho GTPases and signaling networks, *Genes Dev.* **11**, 2295–2322.

van Oosterom, A.T., Judson, I.R., Verweij, J., Stroobants, S., Dumez, H., Donato di Paola, E., Sciot, R., Van Glabbeke, M., Dimitrijevic, S., Nielsen, O.S. (2002) Update of phase I study of imatinib (STI571) in advanced soft tissue sarcomas and gastrointestinal stromal tumors: a report of the EORTC Soft Tissue and Bone Sarcoma Group, **38**(Suppl. 5), S83–S87. *Eur. J. Cancer*

Velicescu, M., Dubeau, L. (2002) p53: a key player in the telomere dynamics, *Cancer Biol. Ther.* **1**, 518–519.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Verhagen, A.M., Vaux, D.L. (2002) Cell death regulation by the mammalian IAP antagonist Diablo/Smac, *Apoptosis* **7**, 163–166.

Verma, A., Kambhampati, S., Parmar, S., Platanias, L.C. (2003) Jak family of kinases in cancer, *Cancer Metastasis Rev.* **22**, 423–434.

Verweij, J., van Oosterom, A., Blay, J.Y., Judson, I., Rodenhuis, S., van der Graaf, W., Radford, J., Le Cesne, A., Hogendoorn, P.C., di Paola, E.D. et al. (2003) Imatinib mesylate (STI-571 Glivec, Gleevec) is an active agent for gastrointestinal stromal tumours, but does not yield responses in other soft-tissue sarcomas that are unselected for a molecular target. Results from an EORTC Soft Tissue and Bone Sarcoma Group phase II study. *Eur. J. Cancer*, **39**, 2006–2011.

Vider, B.Z., Zimber, A., Chastre, E., Prevot, S., Gespach, C., Estlein, D., Wolloch, Y., Tronick, S.R., Gazit, A., Yaniv, A. (1996) Evidence for the involvement of the Wnt 2 gene in human colorectal cancer, *Oncogene* **12**, 153–158.

Vivanco, I., Sawyers, C.L. (2002) The phosphatidylinositol 3-Kinase AKT pathway in human cancer, *Nat. Rev. Cancer* **2**, 489–501.

von Gise, A., Lorenz, P., Wellbrock, C., Hemmings, B., Berberich-Siebelt, F., Rapp, U.R., Troppmair, J. (2001) Apoptosis suppression by Raf-1 and MEK1 requires MEK- and phosphatidylinositol 3-kinase-dependent signals, *Mol. Cell. Biol.* **21**, 2324–2336.

von Mehren, M. (2003) Recent advances in the management of gastrointestinal stromal tumors, *Curr. Oncol. Rep.* **5**, 288–294.

von Mehren, M., Adams, G.P., Weiner, L.M. (2003) Monoclonal antibody therapy for cancer, *Annu. Rev. Med.* **54**, 343–369.

Wallis, Y.L., Macdonald, F. (1999) Demystified . . . oncogenes, *Mol. Pathol.* **52**, 55–63.

Watson, C.J., Miller, W.R. (1995) Elevated levels of members of the STAT family of transcription factors in breast carcinoma nuclear extracts, *Br. J. Cancer* **71**, 840–844.

Weeraratna, A.T., Arnold, J.T., George, D.J., DeMarzo, A., Isaacs, J.T. (2000) Rational basis for Trk inhibition therapy for prostate cancer, *Prostate* **45**, 140–148.

Wei, M.C., Zong, W.X., Cheng, E.H., Lindsten, T., Panoutsakopoulou, V., Ross, A.J., Roth, K.A., MacGregor, G.R., Thompson, C.B., Korsmeyer, S.J. (2001) Proapoptotic BAX and BAK: a requisite gateway to mitochondrial dysfunction and death, *Science* **292**, 727–730.

Weinberg, R.A. (1994) Oncogenes and tumor suppressor genes, *CA Cancer J. Clin.* **44**, 160–170.

Weiner, D.B., Nordberg, J., Robinson, R., Nowell, P.C., Gazdar, A., Greene, M.I., Williams, W.V., Cohen, J.A., Kern, J.A. (1990) Expression of the neu gene-encoded protein

(P185neu) in human non-small cell carcinomas of the lung, *Cancer Res.* **50**, 421–425.

Wells, A. (2000) The epidermal growth factor receptor (EGFR) – a new target in cancer therapy, *Signal* **1**, 4–11.

Wong, A.J., Ruppert, J.M., Bigner, S.H., Grzeschik, C.H., Humphrey, P.A., Bigner, D.S., Vogelstein, B. (1992) Structural alterations of the epidermal growth factor receptor gene in human gliomas, *Proc. Natl. Acad. Sci. U. S. A.* **89**, 2965–2969.

Woodard, A.S., Garcia-Cardena, G., Leong, M., Madri, J.A., Sessa, W.C., Languino, L.R. (1998) The synergistic activity of alphavbeta3 integrin and PDGF receptor increases cell migration, *J. Cell Sci.* **111**, 469–478.

Woodburn, J.R. (1999) The epidermal growth factor receptor and its inhibition in cancer therapy, *Pharmacol. Ther.* **82**, 241–250.

Wosikowski, K., Schuurhuis, D., Kops, G.J., Saceda, M., Bates, S.E. (1997) Altered gene expression in drug-resistant human breast cancer cells, *Clin. Cancer Res.* **3**, 2405–2414.

Wu, G., Chai, J., Suber, T.L., Wu, J.W., Du, C., Wang, X., Shi, Y. (2000) Structural basis of IAP recognition by Smac/DIABLO, *Nature* **408**, 1008–1012.

Wu, X., Fan, Z., Masui, H., Rosen, N., Mendelsohn, J. (1995) Apoptosis induced by an anti-epidermal growth factor receptor monoclonal antibody in a human colorectal carcinoma cell line and its delay by insulin, *J. Clin. Invest.* **95**, 1897–1905.

Yamazaki, H., Fukui, Y., Ueyama, Y., Tamaoki, N., Kawamoto, T., Taniguchi, S., Shibuya, M. (1988) Amplification of the structurally and functionally altered epidermal growth factor receptor gene (c-erbB) in human brain tumors, *Mol. Cell. Biol.* **8**, 1816–1820.

Yang, Q.H., Church-Hajduk, R., Ren, J., Newton, M.L., Du, C. (2003) Omi/HtrA2 catalytic cleavage of inhibitor of apoptosis (IAP) irreversibly inactivates IAPs and facilitates caspase activity in apoptosis, *Genes Dev.* **17**, 1487–1496.

Yarden, Y., Kuang, W.J., Yang-Feng, T., Coussens, L., Munemitsu, S., Dull, T.J., Chen, E., Schlessinger, J., Francke, U., Ullrich, A. (1987) Human proto-oncogene c-*kit*: a new cell surface receptor tyrosine kinase for an unidentified ligand, *EMBO J.* **6**, 3341–3351.

Yin, L., Wu, L., Wesche, H., Arthur, C.D., White, J.M., Goeddel, D.V., Schreiber, R.D. (2001) Defective lymphotoxin-beta receptor-induced NF-kappaB transcriptional activity in NIK-deficient mice, *Science* **291**, 2162–2165.

You, M., Bose, H.R., Jr. (1998) Identification of v-Rel oncogene-induced inhibitor of apoptosis by differential display, *Methods* **16**, 373–385.

Yu, H., Rohan, T. (2000) Role of the insulin-like growth factor family in cancer development and progression, *J. Natl. Cancer Inst.* **92**, 1472–1489.

Yu, J., Liu, X.W., Kim, H.R. (2003) Platelet-derived growth factor (PDGF) receptor-alpha -activated c-Jun NH2-terminal kinase-1 is critical for PDGF-induced p21WAF1/CIP1 promoter activity independent of p53, *J. Biol. Chem.* **23**, 23.

Yu, J.C., Gutkind, J.S., Mahadevan, D., Li, W., Meyers, K.A., Pierce, J.H., Heidaran, M.A. (1994) Biological function of PDGF-induced PI-3 kinase activity: its role in alpha PDGF receptor-mediated mitogenic signaling, *J. Cell Biol.* **127**, 479–487.

Yu, Y., Feig, L.A. (2002) Involvement of R-Ras and Ral GTPases in estrogen-independent proliferation of breast cancer cells, *Oncogene* **21**, 7557–7568.

Yuan, Z.Q., Feldman, R.I., Sun, M., Olashaw, N.E., Coppola, D., Sussman, G.E., Shelley, S.A., Nicosia, S.V., Cheng, J.Q. (2002) Inhibition of JNK by cellular stress- and tumor necrosis factor alpha-induced AKT2 through activation of the NF kappa B pathway in human epithelial cells, *J. Biol. Chem.* **277**, 29973–29982.

Zaffaroni, N., Daidone, M.G. (2002) Survivin expression and resistance to anticancer treatments: perspectives for new therapeutic interventions, *Drug Resist. Updat.* **5**, 65–72.

Zhou, B.P., Hung, M.C. (2002) Novel targets of Akt, p21(Cipl/WAF1), and MDM2, *Semin. Oncol.* **29**, 62–70.

Zhu, G., Decker, S.J., Saltiel, A.R. (1992) Direct analysis of the binding of Src-homology 2 domains of phospholipase C to the activated epidermal growth factor receptor, *Proc. Natl. Acad. Sci. U. S. A.* **89**, 9559–9563.

Ziemiecki, A., Muller, R.G., Fu, X.C., Hynes N.E., Kozma, S. (1990) Oncogenic activation of the human trk proto-oncogene by recombination with the ribosomal large subunit protein L7a, *EMBO J.* **9**, 191–196.

Zimmermann, S., Moelling, K. (1999) Phosphorylation and regulation of Raf by Akt (protein kinase B), *Science* **286**, 1741–1744.

Zwick, E., Bange, J., Ullrich, A. (2001) Receptor tyrosine kinase signalling as a target for cancer intervention strategies, *Endocr. Relat. Cancer* **8**, 161–173.

Oncology, Molecular

Miguel H. Bronchud
Hospital General of Granollers, Barcelona, Spain

1 The Molecular and Clonal Evolution of a Cancer 501

2 Cancer as a Disease Process of Key Regulatory Pathways 507

3 Main Types of Molecular Markers of Cancer Cells: Genetic and Epigenetic Changes 515

4 The Era of 'Targeted Therapies' 520
4.1 Threats and Opportunities 520
4.1.1 Threats 520
4.1.2 Opportunities 526

Bibliography 544
Books and Reviews 544
Primary Literature 545

Keywords

Angiogenesis
This is the biological process that leads to the formation of new blood vessels, and involves the proliferation of endothelial cells and their morphological association into functional capillaries.

Apoptosis
This is the Biological process that describes the complex mechanisms of "programed cell death," eventually leading to endonuclease driven fragmentation of DNA and

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

chromatin by a number of Caspase-dependent or Caspase-independent biochemical events.

Binary State (of Regulatory Molecules)
Most, if not all, biochemical regulatory proteins can be present at any one time in an "on" or an "off" molecular form, with regard to one or more of their functions.

Carcinogenesis
This is now recognized to be a progressive, usually rather slow, multistep process involving sequential mutations, failure of adequate DNA repair, activation of oncogenes, loss of tumor suppressor function, epigenetic changes, and a gradual (or less frequently, rapid) transition from benign to malignant phenotypes.

Cellular Adhesion
Downregulation of 'adherens junctions' (AJ) assembly by mutations, hypermethylation, or transcriptional repression of the adhesion molecule E-Cadherin, has been described in several malignancies, and particularly in human diffuse gastric carcinomas, lobular breast cancers and familial gastric cancers. In many ways, E-cadherin behaves as a tumor suppressor gene, while beta-catenin behaves as a potentially potent oncogene.

Cell Cycle Checkpoints
Mammalian cells can remain quiescent (in Go phase of their cell cycle) for long periods of time, but the right combination of growth factors can induce them to enter G1 and then, if no impediment occurs, progress through S-phase (DNA replication), G2, and mitosis, and complete their cell division.

Chemoprevention
The treatment of carcinogenesis, its prevention, inhibition or reversal.

Drug Resistance
Surprisingly, conventional chemotherapy and radiotherapy work not because cancer cells are especially more "sensitive" to their cytotoxic effects than normal cells, but because normal cells can usually recover faster and more efficiently than cancer cells. Normal tissues virtually never develop resistance to chemotherapy, but human cancers are frequently chemo- or radio-resistant when they are diagnosed, or can become resistant following cytotoxic treatments, because of multiple pharmacological, cell kinetic or molecular mechanisms.

Growth Factors
They were defined in the 1970s, long before their molecular characterization and purification, as *substances other than nutrients required by cells in vitro for their survival, proliferation, and differentiation.*

Intraepithelial Neoplasia
Intraepithelial neoplasia (IEN) is a moderate to severe dysplasia that occurs on the carcinogenic pathway from normal tissue to malignant cancer.

Oncogenes
These are growth control genes present in the human genome (as ''proto-oncogenes''), as well as in the genomes of most, if not all, multicellular organisms. Incorrect expression or mutation of an oncogene usually results in ''Gain of Function'' and unregulated cell growth, as seen in malignant cells.

Mutations
The molecular changes in DNA associated with carcinogenesis are not very different from those associated with genetic variation and molecular evolution of organisms and species. These changes include local changes in DNA sequences, rearrangement of DNA segments within the genome, and (in the case of viral carcinogenesis) horizontal transfer of a DNA (or RNA) segment originating in another kind of organism.

Proteasomes
The average human cell contains about 30 000 proteasomes, each of which contains several protein-digesting proteases. Among other functions, they control the intracellular half-life of important regulatory proteins.

Stromal cells and Cancer
Stromal cells are all cells in a tumor that are not the cancer cells.

Tumor Antigens
These are molecules that contain the epitope determinants recognized by T cells that specifically target tumor cells.

Tumor Suppressor Genes
Unlike oncogenes, tumor suppressor genes (formerly ''antioncogenes'') display the function of suppressing cellular proliferation, and/or maintaining cellular differentiation, facilitating normal cellular adhesion mechanisms, stopping entry into the cell cycle (G1, S-phase, mitosis or G2) to allow DNA-repair mechanisms, maintaining normal cellular shape and cell contact inhibition mechanisms, and promoting normal differentiation and senescence.

Molecular Oncology can be defined as *that branch of medical science that looks at the cancer problem from a molecular point of view*. For several reasons, ''molecular oncology'' represents ''the heart of the matter'' of cancer.

As it would be quite an ambitious, and virtually impossible, task to pretend to cover the whole subject of ''Molecular Oncology'' in a single unipersonal review, what follows is a short personal view of what the author regards as the basic molecular understanding of cancer in the year 2004, and the more likely therapeutic implications of this new molecular knowledge.

We should not forget the huge prevalence of malignant diseases in the developed world. Cancers are probably responsible for one-third of deaths in men in industrialized countries and for almost one-fourth in women. At the beginning

of the twenty-first century, these diseases still cause a lot of human suffering, despite very significant progress in their early detection, improvements in radical treatments (local or systemic), and better medical control of iatrogenic side effects of chemotherapies.

In the same way as the invention and later refinement of the microscope eventually led to the discovery and classification of pathological microorganisms, the discovery of DNA, RNA, proteins, proteoglycans and other regulatory molecules (including lipids, arachidonic acid derivatives, steroids, and sex hormones) is rapidly leading to a better and deeper understanding of cancer, and remains the best promise to lead to better and more effective methods for cancer prevention, early detection, prognostic classifications, and cure.

This is because it is currently believed that cancer can be truly understood in terms of those molecular changes, genetic and epigenetic, that gradually lead a given cell clone, within a given tissue field, to develop a certain ''competitive advantage'' over neighboring cell clones, enabling the transformed cell clone to gradually displace its neighbors and grow into a neoplastic tissue. Finally, this ''new tissue'' can go on to transform itself into a malignant neoplasia, leading to invasion of local organs, or distant spread (generally through the bloodstream, the lymphatic system, or both). The growth of cancer cells at distant sites is called *metastasis*.

''Cancer'' is a state, whereas ''carcinogenesis'' is a process. Key to the multistep genetic nature of cancer is that carcinogenesis is ''progressive.'' In most epithelial tissues, progression means the sequential accumulation of somatic mutations, or even epigenetic changes (like abnormal DNA methylation patterns). In some cases of familial predisposition to cancer, some of these mutations are inherited.

Gradually, a given target tissue experiences a transition from normal histology, to proliferative and/or dysplastic changes, to so-called *intraepithelial neoplasia* (*IEN*), which can be early or severe, to superficial cancers *(in situ)*, and finally to invasive disease. In some instances, this process of malignant transformation may be aggressive and relatively rapid (e.g. in the presence of a DNA repair-deficient genotype, or an aggressive oncogenic virus, or a lethal combination of ''oncogenic hits''), but in general these changes occur over a long period of time, like 3 to 30 years.

The problem is that in the past two and a half decades there has been an ''explosion'' of information, rather than true knowledge, about the molecular aspects of cancer. More than 300 different genes and their respective protein products have been described as being directly or indirectly linked to cancer. The number of such 'cancer-related genes' is constantly increasing and the final figure could well be more than 1000. Although it seems reasonable to believe that some, or many, of these 'cancer-related' genes may represent the *consequence* rather than the *cause* of the cellular development of the cancer cell phenotype, the 'trees' are so many that there is a real risk of missing the ''wood.''

Indeed, some believe that, in the not too distant future, relevant information will pass from the Molecular Pathology Laboratory to the busy Cancer Clinical Units only with the help of clever computer programs. Before clinicians can make up their minds about the curability or incurability of any given cancer, and in order to decide which sequence and combination of drugs to use to treat a particular patient, they

will have to consult a computer program and the Molecular Pathology Laboratory. Although there is still no treatment for any of the major lethal cancers that is as effective as, say, antibiotics are against infections, there is vast accumulated knowledge of the fine regulatory mechanisms that are deranged in cancer cells, and this undoubtedly promises new therapeutic insights. For example, selective oral tyrosine-kinase inhibitors (Gleevec), for chronic myeloid leukemia, and specific monoclonal antibodies (Herceptin) for breast cancer or some lymphomas (Mabthera), have been introduced into routine clinical use.

In contrast to the situation twenty years ago, not only do we now know of many molecular targets to design new drugs for the chemoprevention or treatment of cancer but, paradoxically, we have an apparent excess of targets for our current resources of drug development worldwide. The Human Genome Project has completed its first basic human genome map well ahead of schedule, and it is likely to give us further insights and more potential targets. It is now estimated that the human genome contains some 35 000 genes, which is less than had originally been estimated by most researchers. Many of these genes are well characterized and their functions in various pathways are known. But the real function of the majority of these human genes remains inconclusive. In other words, the rate-limiting step in true progress against cancer is the amount of resources we can spend, and the optimization and coordination of this huge research process, rather than a shortage or lack of therapeutic targets.

Selecting the right targets for cancer therapy can make a big difference. If, for example, we were clever or lucky enough to correctly guess the right targets for the main human cancers, and if large multinational pharmaceutical companies agreed to focus their efforts and enormous resources on these right targets, then revolutionary new cancer treatments might become available for clinical testing within 5 to 10 years. But, if we got it wrong, or not enough importance was given to this war against cancer by politicians or business people, then it might take another 20 or 30 years, or even more.

1
The Molecular and Clonal Evolution of a Cancer

In view of the fact that the incidence of most of the common epithelial cancers is age related, the British epidemiologists Armitage and Doll had already calculated, in 1958, that for some of the common solid tumors the logarithm of cancer incidence should be linearly related to the logarithm of a person's age. If such an interpretation holds good, then we could deduce, from the slope of the death rate from cancer of the large intestine in relation to age (plotted logarithmically), that about 6 mutations are needed to produce a cancer of the large intestine.

This guess is extraordinary if one thinks that more or less the same conclusion was reached, based on molecular genetic knowledge, by the American scientists Fearon and Vogelstein in their classic work on colorectal tumorigenesis. The

'multistep mutations' theory of cancer was also supported by epidemiological evidence, and by the 'initiation and promotion' models of carcinogenesis. A characteristic feature of most forms of carcinogenesis is the long period that elapses between initial application of the carcinogen and the time the first cancers appear. It is necessary to apply coal tar repeatedly to the skin of a mouse for several months before any tumors are detectable. Similarly, most common human cancers can take 3 to 30 years or more to develop.

The chemical carcinogenesis experimental models also helped to identify at least two classes of carcinogenic compounds: the initiators and the promoters. For example, if a group of mice are fed a small amount of the carcinogen dimethylbenzanthracene (DMBA), this produces widespread irreversible alterations (presumably mutations) in the cells of each mouse. Subsequent irritation of the skin by painting it twice a week with croton oil (the "promoter") eventually results in the local appearance of tumors. These tumors will appear even if croton oil is not started until 16 weeks after the DMBA feeding, but no tumors arise if either DMBA or croton oil is given alone or if the order of the treatments is reversed. In several aspects, estrogens (in the case of breast cancer) and testosterone (in the case of prostate cancer) have also been regarded as potential tumor promoters.

Other insights into the "genetic" nature of tumorigenesis came from studies on viral carcinogenesis, and from seminal observations in the uncommon retinal cancers in children.

The discovery of tumor oncogenes and tumor suppressor genes almost twenty years ago opened the way to the study of the molecular epidemiology of cancer. It soon became apparent that in general more than one somatic mutational event is needed for malignant transformation, the possible exception being the uncommon hereditary retinoblastomas, already described by the "two-hits model" proposed by Knudson.

Later, it was found that certain carcinogens are linked to selective (though not entirely "specific") mutational events. For example, researchers have described molecular linkages between exposure to carcinogens and cancer types, in p53 mutational spectra of hepatocellular carcinoma, skin cancers, and lung cancer. In 1990, Fearon and Vogelstein proposed a molecular model for colorectal carcinogenesis, based on the sequential accumulation of genetic events in key regulatory genes along the sequence from adenoma to carcinoma.

More recently, in 1997, Kinzler and Vogelstein proposed the concept of two different types of carcinogenic genetic events: those involving "gatekeeper" genes, characterized by their control of net cellular proliferation, and those involving "caretaker" genes, associated with maintenance of genomic integrity. Examples of gatekeeper genes include APC and beta-catenin in colon epithelium, Rb in retinal epithelial cells, NF1 in Schwann cells, and VHL in kidney cells. Thus, it is proposed that an alteration in APC leads to a derangement of the cellular proliferation pathway that is important for maintaining a constant cell population, at least in colonic cells. The identification of other gatekeeper genes is anticipated, and some may be genes crucial to morphogenetic events of specific tissues. Unlike gatekeeper genes, caretaker genes generally maintain genomic stability and are not involved directly in the initiation of the neoplastic process, but their mutations enhance the probability of mutations in

other genes, including those in the gatekeeper class. Because multiple mutations are found in cancer cells, the existence of a "mutator phenotype" was suggested by Loeb in 1991 as being an important step in tumor development. Candidate mutator genes are involved in multiple cellular functions needed for maintaining genetic stability, such as DNA repair, DNA replication, chromosomal segregation, cell cycle control, and apoptosis.

Some individuals may be predisposed to cancer because of inherited mutations of some key genes. This has attracted considerable attention in recent times, particularly in relationship to genes having susceptibility for breast cancer and colon cancer.

The genetic alterations in oncogenes generally lead to an increased function of the protein, whereas, in general, tumor suppressor genes are inactivated during carcinogenesis with apparent loss-of-function of the protein. However, the mechanisms of activation or inactivation are multiple, and their precise effects on gain- or loss-of-function are incompletely understood. *K-ras* and *H-ras* genes are examples of oncogenes that are preferentially altered by point mutation (codons 12, 13 and 61), generating a protein with constant GTPase activity. The *c-myc* gene can be activated by chromosomal translocation (in some leukemias) or by gene amplification (in some solid tumors). The *p53* and *Rb* tumor suppressor genes are often knocked out by point mutation in one allele and by deletion (loss of heterozygosity) at the other. Others, like *p16*, have high rates of homozygous deletions or promoter hypermethylation. Some genetic defects are fairly characteristic for a given tissue type (most colorectal cancers have APC or beta-catenin mutations). But the "same players" are frequently involved in different tumors. Each human cancer can be regarded as a different molecular entity, with a different matrix of molecular targets, and it evolves with time (even as a result of systemic or local therapies).

However, it would be a mistake to believe that all common epithelial human cancers follow these clear-cut histological sequential patterns from adenoma to carcinoma. In fact, only a minority of the commonest type of breast cancers (infiltrating ductal carcinomas, or IDC) arise from ductal carcinoma *in situ*. Thus, the molecular changes leading to IDC, accounting for almost 70% of all breast cancers, can happen before the histological features associated with DCIS become evident. DCIS is characterized by a proliferation of malignant epithelial cells confined to the mammary ducts without light microscopic evidence of invasion through the basement membrane into the surrounding stroma; but IDC, by definition, show signs of invasion of stromal tissue, often with vascular and/or lymphatic vessel involvement. In other words, conventional histological and radiological techniques (e.g. bilateral mammograms) cannot detect with adequate precision the ongoing carcinogenic events in many cases of women at risk.

Patients with head and neck squamous cell carcinoma (HNSCC) often develop multiple (pre)malignant lesions, ranging from leucoplakia to other cancers. This led Slaughter et al, way back in 1953, to postulate the concept of "field cancerization." The incidence rate of second primary tumors following a first diagnosis of HNSCC is 10 to 35%, depending both on the location of the first primary tumor and the age of the patient. The carcinogens associated with HNSCC (alcohol and tobacco smoking) are thought to induce mucosal changes in the entire upper aerodigestive

tract (UADT), causing multiple genetic abnormalities in the whole tissue region. Similar arguments apply also to other tobacco-related cancers, like transitional cell carcinomas of the urogenital tract or bronchogenic carcinomas. An alternative theory for these observations is based on the premise that any transforming event is rare and that the multiple lesions arise because of the widespread migration of transformed cells through the whole UADT. However, most field changes appear to be induced by smoking, supporting the theory of carcinogen-induced field cancerization rather than field cancerization due to migrated transformed cells.

Other possible causes of "field carcinogenic events" can involve hormonal factors (e.g. changes in the ovaries, breasts or prostate), inflammation and hyperemia (increased proliferative and angiogenic activity in chronic cystitis, gastritis, esophagitis or colitis), chronic viral infections (e.g. Hepatitis B virus for hepatocarcinomas, Epstein-Barr virus for nasopharyngeal carcinomas or some lymphomas), aberrant methylation linked to aging, free-radical induced DNA damage (e.g. for cancers of the gastrointestinal tract), skin exposed to ultraviolet irradiation (e.g. actinic keratosis and squamous cell carcinomas), ionizing radiation-induced damage, or aberrant morphogenetic pathways. It is also possible that different carcinogenic pathways operate in different tissue fields belonging to the same organ. For example, adenocarcinomas of the right side of the colon are often associated with clinical and molecular characteristics different from those of adenocarcinomas of the colorectal region.

Even in breast cancer, the reported incidence of multicentric or multifocal lesions in areas away from the primary tumor in mastectomy specimens ranges from 9 to 75%, depending on the definition of multicentricity, the extent of tissue sampling and differences in the histological techniques of examination. Therefore, multifocality or multicentricity of breast cancers may in fact be a lot more common than is currently acknowledged.

In this context, the old "Field Cancerization" theory by Slaughter, and the more recent "Multi-step Carcinogenesis" model by Fearon and Vogelstein can now come together in a single model: Sequential Field Cancerization.

If it does require some seven sequential carcinogenic "genetic hits" in a single cellular clone for a malignant tumor to develop, it is mathematically more likely to occur in a tissue with a high background of genetic alterations in neighboring cellular clones, than in a tissue with a low background of such alterations, or with no detectable carcinogenic mutations at all (Figs. 1 and 2).

The probability of a single clone accumulating 7 independent but sequential genetic alterations leading to a malignant phenotype, without any similar events occurring in neighboring cells, would seem to be rather low. This simple conclusion, and our ability to measure "background carcinogenesis" in different parts of the body, might lead to several unexpected implications. Technology is just beginning to be sufficiently sensitive to start testing the hypothesis. One potential technical problem is that in premalignant tissue, the "signal" (e.g. relevant oncogenetic lesions) might be muted by the "noise" (normal genome of most of the cells in the tissue), until the premalignant clones have expanded enough to become more numerous locally than normal cells. However, it is only a matter of time before this goal is technically achievable.

A possible future objective is the development of a combined histological

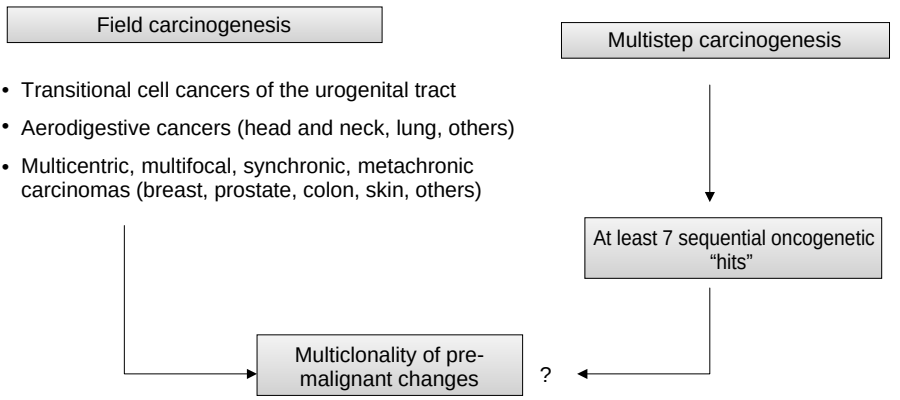

Fig. 1 Multiclonality of premalignant changes.

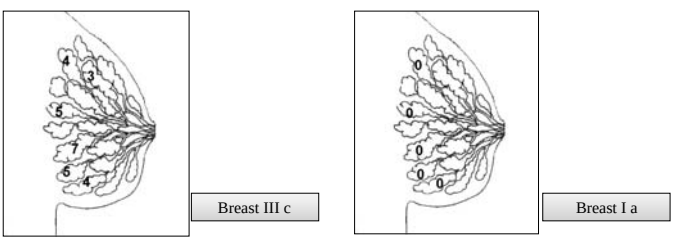

	Normal mucosa	Profilerative	Dysplastic
No mutations	I a	I b	I c
< 3 mutations	II a	II b	II c
> 3 mutations	III a	III b	III c

Fig. 2 Combined histological and molecular staging system of premalignant changes. Numbers inside the drawings refer to numbers of relevant gene mutations/deletions/amplifications identified in various regions of the breasts. Carcinogenic pathways may differ depending on the specific combination of genetic (and epigenetic) changes. (Reproduced from Bronchud, M. H. (2002) Is cancer really a "local" cellular clonal disease?, *Medical Hypothesis* **59**, 560–565. With Permission of Elsevier Science.)

and molecular staging system (Fig. 2). For example, after a follow-up of 5 to 10 years, one would expect more new cancers to develop in group IIIc of Fig. 2 (dysplastic changes and three or more significant mutations identified), than in group Ia (normal histology and no mutations identified).

There is a growing number of experimental findings in support of this complex concept. Among the first was the observation by Azadeh Stark and colleagues from various American institutions, that women with benign breast biopsies, demonstrating both HER-2/neu amplification and a proliferative histopathological diagnosis (either typical or atypical hyperplasia), may be at a substantially increased risk for subsequent breast cancer (up to more than sevenfold) compared to normal.

The clinical application of this concept and technology should help to classify patients into various relative risk groups early in the development of a malignant disease, allowing a tailor-made program for follow-up and screening, as well as more appropriate chemopreventive and therapeutic interventions. For example, a suitable combination of relevant biomarkers might help clinicians to identify smokers at high risk of developing lung cancers (approximately 10 to 15% of frequent smokers). Being confronted with a personal risk of cancer, rather than a general statistical risk, is a potent motivation to quit smoking and to undergo more frequent health checks (like high-resolution CT scans to detect isolated pulmonary nodules).

Some smokers may be protected because of genetic polymorphisms of enzymes involved in the molecular activation of precarcinogens present in tobacco, whereas others may be more vulnerable to the carcinogenic effects because of genetic defects in DNA-repair enzymes. Some molecular changes associated with aging and carcinogenesis might be epigenetic (e.g. promoter hypermethylation) rather than genetic. Even some pediatric malignancies might be secondary to abnormal morphogenetic events *in utero*.

It has been estimated that in the USA alone some 30% of people above the age of 60 can be found, by colonoscopy, to have adenomas of the colon, 70% or more of men above the age of 80 will have IEN of the prostate, 30% of people aged 60 or more have actinic keratosis on their skins, 20% of sexually active women above the age of 40 may have some degree of cervical IEN, at least 40% of heavy smokers can show metaplastic or dysplastic changes in their bronchial mucosa, and some 20% of women with dense mammograms and aged more than 50 may show atypical cells on ductal washings from the nipple or ultrasound-guided fine-needle aspirates.

The use of a battery of genetic or protein biomarkers relevant to each of the main cancer types may soon help us to better define individual cancer risks, and to measure background carcinogenesis in individual tissue samples. Perhaps, not too long from now, oncology units will be devoted to the treatment of carcinogenesis just as much as to the treatment of cancer.

One crucial difference between normal cells and transformed cells *in vivo*, which is difficult to prove experimentally but can be deduced, is that during cancer development, the effects of mutations are likely to be persistent, whereas most regulatory intercellular signals (particularly of "paracrine" nature) in normal cells are likely to be terminated by changes in cell position, production of signals, or modulatory effects on membrane receptor state. In normal tissues, the quantitative levels of signaling have been shown to be very important; signaling episodes, moreover, can have very different frequencies, and a specific cell can be influenced by a long-term signal while responding to a different signal of shorter duration. This does require that normal cells can sometimes remember a decision indefinitely, and can terminate their response to a specific signal, so that exposure to a new signal, or even the same signal, can subsequently produce a different molecular response.

The fact that cancer cells have lost, at least in part, this "position-dependent" obedience to signals, to the extent that they can even grow into distant metastases, makes them life threatening to their host organisms. Some sort of position dependence, however, is likely to remain even for most cancer cells, because their site of seeding and metastatic growth is not random, but depends on a number

of cell adhesions and microenvironmental predisposing factors.

2
Cancer as a Disease Process of Key Regulatory Pathways

The basic mechanisms controlling classical metabolic regulatory pathways have been known for many years. For example, the "citric acid cycle" (postulated by Krebs in 1937), the central role of ATP in energy-transfer cycles (postulated mainly by Lipmann in 1939–41), and Mitchell's intriguing hypothesis (1961) to explain the mechanism of oxidative and photosynthetic phosphorylation, to name but a few, have been part of biochemistry textbooks for decades. Some twenty years ago, the structure of DNA had already been known for over twenty years before that, and yet eminent scientists were pessimistic about real therapeutic progress in oncology.

In contrast, regulatory pathways involved in the complex regulation of cell growth, differentiation, senescence, and cell death are only recently being gradually understood. Although we are still largely unable to draw schematically precise cell-type-specific regulatory pathways, research efforts are intensifying, and our knowledge is rapidly expanding.

Growth and differentiation must be tightly controlled in any organism. Curious though it may seem, recent research suggests that fractal networks for transporting the materials essential for life, for example, nutrients, water, and oxygen, may be prime movers in determining shape and form (morphology) in Nature. A series of key papers by Geoffrey B. West, a theoretical physicist at Los Alamos National Laboratory, James H. Brown, a biology professor at the University of New Mexico, and Brian J. Enquist, a postdoctoral biologist at the nonprofit Santa Fe Institute, have reduced to just three general principles the allometric scaling laws, that are major determinants of most anatomical and physiological variables of organisms as they increase in body size.

In the year 2001, which was the hundredth anniversary of the institution of the Nobel Prizes, three of the leading investigators in the field of Cell Cycle Control were awarded the Nobel Prize for Physiology and Medicine. The American scientist Leland H. Hartwell of the Fred Hutchinson Cancer Research Center, and the British scientists R. Timothy Hunt and Sir Paul M. Nurse, both at the Imperial Cancer Research Fund in London, received this prestigious award for 'the confluence of two different approaches to learning about the molecular machinery regulating the cell cycle'.

While Hartwell was screening for temperature-sensitive mutants in the yeast *Saccharomyces cerevisiae* in the 1960s and 1970s, he identified a large series of genes that, in mutant form, arrested mitosis. This led him, and former postdoctoral researcher Ted Weinert, to hypothesize that 'checkpoints' regulate the sequence of events in mitosis (3). Even as Hartwell was uncovering these genes, Paul Nurse was identifying, in a different yeast (*Schizosaccharomyces pombe*), a protein kinase (CDC2) that is a key component of the maturation factors that drive cell division. Nurse then took a human cDNA library and eventually cloned a human CDC2, having observed the striking structural homology of these key regulatory molecules, that indicated a high degree of evolutionary conservation in different species.

Working at the Marine Biological Laboratory in Woods Hole, Mass., in the early-1980s, Hunt solved a crucial piece of

the cell cycle puzzle. In sea urchin oocytes, and later in frogs, Hunt identified proteins whose levels rose and fell throughout the cell cycle. He called them *cyclins*, and they were later found to be necessary in activating the ever-present kinases, like CDC2. Together a cyclin and a cyclin-dependent kinase (CDK) push the cells along the cell cycle regulatory checkpoints.

Previous work by others had already opened the path to these important findings. For instance, back in the early-1970s two scientists, Yoshio Masui and L. Dennis Smith, transferred the cytoplasm from activated frog oocytes to naïve oocytes arrested in meiosis, and identified a substance they called *MPF* (*maturation promotion factor*). Almost twenty years later, that substance proved to be none other than the CDK/Cyclin heterodimer.

In the year 2002 the Nobel Prize was awarded for seminal discoveries in another key area of biological relevance: the genetic regulation of organ development and programed cell death, or apoptosis. The winners of the Prize were the British scientists Sydney Brenner and John Sulston, and the American scientist Robert Horvitz. They discovered that specific genes control the cellular death program in the nematode worm *Caenorhabditis elegans*. This worm, approximately 1-mm long, has a short generation time and is transparent, which makes it possible to follow cell division directly under the microscope. The fertilized egg cell undergoes a series of cell divisions leading to cell differentiation and cell specialization, eventually producing the adult worm. In this organism, all cell divisions and differentiations are invariant that is, identical from individual to individual. This made it possible to gradually construct a precise cell lineage for all cell divisions. During the development of the organism, 1090 cells are generated, but precisely 131 of these cells are eliminated by programed cell death. This results in an adult nematode (the hermaphrodite), composed of 959 somatic cells. Again, most of the multiple genes controlling programed cell death have human equivalents (Apoptosis Pathways).

Thanks to the Human Genome Project, we now know that there are between 30 000 and 35 000 genes in the Human Genome, and 18 000 in the nematode worm. Both the Cell Cycle Control genes and the Programed Cell Death genes are crucial for carcinogenesis, and obvious targets for cancer therapy. In fact, optimal curative cancer therapies should include combinations of drugs that not only cause cell cycle arrest in tumor cells, but also lead to their apoptosis.

Recognizing that much of the cell's work is done not by individual proteins but by large macromolecular complexes, researchers are increasingly trying to map protein–protein interactions throughout the cell. The new term *Interactome* has been proposed to define this new type of "cartography" describing the main functional interactions of single proteins and protein macrocomplexes. For example, the map of the *C. elegans* interaction network, or interactome, links 2898 proteins (nodes) *via* 5460 interactions (edges).

It seems likely that, with time, different family patterns of "malignant interactomes" will be gradually identified, each with its unique "key pathogenetic changes" caused by the specific abnormal regulatory pathways implicated in its carcinogenesis process.

Over the past few years, researchers engaged in the nascent field of "interactomics" have been busy deconstructing these molecular machines, mapping protein–protein interactions in bacteria, yeast, fruit flies, nematodes, mice, and humans.

The resulting charts have exposed the functions of previously mysterious proteins, have helped drug companies *hone* their development efforts, and are laying the foundations for systems biology. Some proteins appear to be rather "promiscuous" in their interactions. Drugs targeting these so-called *sticky proteins* have the potential to produce troubling side effects. Pharmaceutical companies, therefore, would be better served by focusing on those molecules that appear more selective, and by limiting themselves only to the particular process they are targeting.

It must be said, however, that doubts about the validity of interactome information remain. Although "data rich," these studies produce lower-quality results than would a scientist who devotes years to a single protein. High throughput means that you do things fast, but the information can be "data-rich and analysis-poor." Again, as is true for most of Molecular Oncology today, "the trees may not allow us to see the woods."

Y2H data especially are open to question, because the technique essentially amounts to a genetic trick. It relies on artificial fusion constructs: often, mere fragments of the full-length protein are introduced into cells, forced into the nucleus, and overexpressed. False-positive results are inevitable.

The vast majority, 93%, of yeast proteins makes five or fewer connections, yet only one in five such proteins is essential.

Deletion of highly connected proteins (those joining 15 or more proteins) is three times more likely to be lethal.

Recently, Marc Vidal's team at the Dana-Farber Cancer Institute in Boston expanded this study by examining coexpression of genes linked by hubs. Their conclusion: all hubs are not created equal. Some, called *party hubs,* contain proteins that interact simultaneously, whereas *date hubs* contain proteins that interact at different times and locations.

The Dana–Farber group infers a modular architecture for the proteome, in which date and party hubs operate at different organizational levels. Party hubs function to assemble individual molecular complexes, or modules. These modules, in turn, are linked at a higher level, using date hubs. Thus, the date hub calmodulin links modules for cation homeostasis; budding, cell polarity, and filament formation; protein folding and stabilization; and the endoplasmic reticulum.

There can be little doubt that if this new branch of biological research (Interactomes), develops to the point of not only mapping the functional protein–protein interactions in any given key regulatory pathways like the TGF-beta pathway but also predicting, on the basis of genomic and/ or proteomic data for any given cancer tissue, the dynamic functional consequences of the intrinsic molecular changes observed in tumor specimens, and the manner of their response to external manipulation by drugs or ionizing radiation, then we can end up with truly useful tools for both drug development and clinical follow-up.

With so much disparate data available, the challenge for researchers is to integrate that information into a single "interaction network," and then to take decisions relating to research and development of new anticancer drugs. These could, for example, be based on new visualization tools such as "Cytoscape," jointly developed by the Institute for Systems Biology at the University of California at San Diego, and the Memorial Sloan-Kettering Cancer Center in New York. Cytoscape, a technology still undergoing development, incorporates interaction data with

other large-scale genomic information to build networks of cellular processes and pathways. Through the Systems Biology Markup Language (SBML), it can also communicate with other SBML-enabled programs to test mathematical simulations of cellular events.

Unraveling the twisted skeins of the relative contributions of altered regulatory pathways to the cancer phenotype of each individual cancer can be certainly annoying. Simplicity is elegant, but it can also be deceiving. In spite of the unquestionable success of many reductionistic approaches in science, we should never underestimate complexity, as often going against nonsense sensible arguments is just so much blowing against the wind.

However, at this stage of knowledge, it does seem likely, and it remains highly desirable, that most of these key regulatory cascades will converge into a number of key regulatory events: the DNA replication at multiple different sites in the genome, the coordinated expression or suppression of a battery of genes, the culmination of the developmental history and cell fate of any given clone, the progressive modification of the biological "clock-like mechanisms" that drive differentiation and aging, and the three-dimensional structure–function relationships of chromatin.

A very simplistic, and classic Cartesian, way to look at what happens in cancer cells is to describe the main functional implications of genetic events (point mutations, deletions, translocations, etc.) in binary terms (GF: gain-of-function; LF: loss-of-function *versus* N: normal function) in a table such as the one that I have called *'Matrix of Targets'*.

Each cancer may have, because of its own genetic and epigenetic backup and clonal history, a different Matrix of Targets. This adds to the complexity of potential curative therapies, but can eventually translate into therapeutic combinatorial models of several drugs or biological agents capable of stopping the growth and proliferation of cancer cells, reducing their invasive or metastatic properties, or even restoring their differentiation or apoptotic pathways.

Another way to look at things is to take the view that certain key genetic changes are the true and main pathogenetic mechanisms for any single cancer, because these key genetic changes (e.g. the abl-bcr fusion gene, associated with the Philadelphia chromosome in Chronic Myeloid Leukemia) lead to a much wider pattern of changes in gene expression, up or down, thereby consolidating the malignant phenotype: increased and unrestrained cellular proliferation, loss of contact inhibition, lack of normal differentiation or presence of abnormal differentiation, genomic instability, ectopic expression of genes, angiogenic activity, acquisition of the invasive phenotype, the ability of cells to spread and seed at distant sites, and so on.

These key "pathogenic genetic events" would drive the clonal evolution of cancer cells in a way similar to the "evolution genes" postulated by Werner Arber, elegantly described in the chapter on "Genetic Variation and Molecular Evolution" in this encyclopedia."

There is therefore a need, in Molecular Oncology, to follow very closely the rapidly accumulating knowledge on the molecular mechanisms that regulate gene expression.

Identifying genome-wide regulatory motifs is problematic because of their typically low consensus sequences. Grasping which transcription factors bind to such motifs is even more difficult, as numerous factors can bind to any one motif.

In two recent publications, Pugh and colleagues found that only 20% of yeast genes contained their TATA boxes, but they were all highly regulated in response to stresses such as heat and osmotic shock. Such genes generally use a multicomponent protein complex called *SAGA*, which is one of several complexes thought to bring the TBP to promoters. In contrast, genes without a recognizable TATA box appeared to use mainly another complex, TFIID, and were less involved in responding to stress. This could be an example of bipolar type of regulation.

Predicting gene expression from DNA sequence, and understanding the relative role of individual transcription factors in the development of normal tissues and in the oncodevelopment of cancers, are the clear goals of these technologies. However, these goals are still not within immediate reach.

Although probably no two cancers are 100% identical, in the same way as no two cancer patients are ever identical, the rational understanding of how altered regulatory pathways can lead to the development and consolidation of cancer cell clones, may provide us with some basic patterns of regulatory circuits associated with malignancy. To some extent these "biochemical maps" or "carcinogenic patterns" may be tissue dependent.

Most of these key regulatory processes, if not all, will have:

1. Upstream regulatory elements; for example, membrane-bound or membrane-associated.
2. Intermediate elements; cytosolic or bound to organelles, like mitochondria.
3. Downstream elements, transcription factors, DNA- or RNA-specific sequences.
4. Bottlenecks, Cross talk points and "Achilles heels" (Oncogene Addiction Hypothesis) of cancers.

New cancer therapies will not be directed merely towards hitting the DNA synthesizing machinery but also towards selective targets at or downstream from abnormally activated catalytic functions (thereby abrogating the abnormal stimulatory signals), or at or upstream from inactivated or missing catalytic functions (switching on parallel or alternative regulatory pathways).

In simplistic terms, the new therapeutic drugs should, for example, lock or inhibit activated proto-oncogenes in their inactive state, or somehow lock tumor suppressor proteins in their active state, which is the one normally responsible for inhibiting entry into S-phase or mitosis, or maintaining the normal differentiation and apoptotic pathways (Fig. 3). In other words, as regulatory proteins are usually present in normal cells in either an active form or an inactive form, and the chemical equilibrium between the two forms depends on the regulatory microenvironment of the cell, new targeted therapies should be able to correct, or at least limit, the malignant phenotype by restoring the normal balance between stimulatory and inhibitory signals within the cell.

This concept of regulatory pathways involved in cell behavior is the basis of a new biochemistry. Crucial alterations in this new biochemistry, perhaps as few as five distinct key molecular changes, may be enough to transform a normal human cell into a tumor cell. It appears that because of their higher complexity, human cells are more difficult to transform than rodent cells. Perhaps surprisingly, more simple organisms, such as *Drosophila*

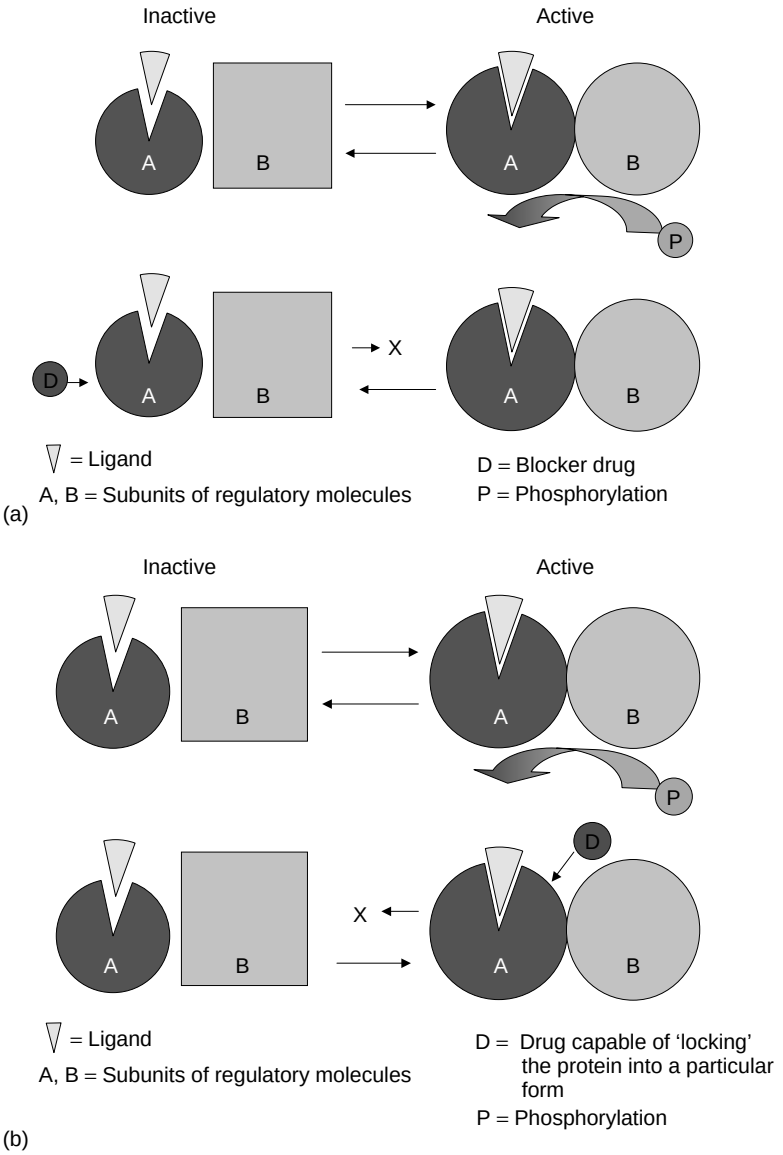

Fig. 3 Hypothetical new target-oriented therapies. (a) Oncogenes as targets. The new anticancer agents should lock the oncogenic protein in its *inactive form*, either by binding to the active site, or by binding to an allosteric regulatory site; (b) Tumor suppressors as targets. The new anticancer agents should do the opposite of what is described above. They should lock the tumor suppressor protein into its *active form*, that is, the one capable of inhibiting the cancer phenotype, by either binding to the relevant active site or to allosteric regulatory sites. (From M. H. Bronchud. (2004) *Principles of Molecular Oncology*, reprinted by permission of Humana Press.)

melanogaster flies, very rarely develop malignant cells, in spite of their ability to develop mutations in experimental systems. There has been no clear explanation for this apparent paradox.

Phosphate atoms are again involved in this "new biochemistry," because protein phosphorylation is the common end-result of many signal pathways. It was back in 1937 (the same year Krebs postulated the citric acid cycle) that C. Cori and G. Cori began their outstanding studies on glycogen phosphorylation (the main store of sugars in the liver of mammals). However, it was not until 1988 that Tonks obtained the first partial sequence of a tyrosine phosphatase. Today, it has been suggested that the human genome might contain as many as 2000 kinase genes and up to 1000 phosphatase genes.

A "normal" cell, under "normal" circumstances, probably decides nothing by itself (there is no cellular "free-will"), but merely obeys orders. Normal cells can signal each other in many different ways: through pores in the membranes (gap junctions or plasmodesmata), or because of specific membrane receptors that recognize soluble or bound ligands (endocrine, paracrine or autocrine mechanisms), or in some special cases (e.g. neurones) by synaptic transmission. However, a cancer cell (because of mutations, amplifications, translocations or deletions) makes mistakes (misinterpreting the normal orders or even making them up for the wrong reasons). And it cannot help doing so because its software has gone badly and progressively wrong. Cancer cells do not have exactly the same underlying chemistry as the normal cells of the body.

This makes them "strong," but also, paradoxically, makes them vulnerable to more specific and selective combinations of drugs.

Most of these, second, third, fourth, and so on, intracellular messengers are binary: they are either switched on or off (by phosphorylation-dephosphorylation, for example). At any one time a number of + and − signals travel from the cell membrane to the nucleus (and, perhaps, also backwards). These signals are irreversibly altered in cancer cells. The normal balance between the "on" and "off" signals is altered in cancer cells.

To reestablish the normal circuits, or to kill the cell by misleading it into apoptosis, we must learn to be clever, understand these key pathways, and develop drugs to block them or to activate alternative complementary pathways. Thus, cancer is a disease of regulatory pathways. There may not be "magic bullets" to kill all cancer cells. The signal transduction pathways involved in the proliferative response to various growth factors and mitogens are extremely complex and interactive. They do not act as simple linear cascades. The so-called *'cross talk'* is an invariable feature. The new biochemistry is essentially the network of sensing and signaling in cell homeostasis.

Several key questions remain:

1. How much selectivity can we expect to achieve with new drugs that will manipulate these complex signaling pathways?
2. Will these new drugs be cytostatic, or cytotoxic, or both?
3. How will these new drugs be obtained: serendipity; rational drug design; high-throughput screening; combinatorial chemistry; antisense technologies; gene vectors; antigrowth factors; monoclonal antibodies?
4. Will they imply chronic therapies or acute therapies?
5. Will they have new, and at times unexpected, toxicities?

6. Will some malignant clones still develop resistance to signal transduction therapy?

Some of these questions can already be answered by the available examples of targeted therapies in cancer. For example, it seems that most of the new drug entities acting on cell surface growth factor receptors, or cytosolic downstream regulatory elements, for example inhibitors of the EGFR (Epidermal Growth Factor Receptor) or associated tyrosine-kinase activities, are mainly "cytostatic" *in vivo* rather than "cytotoxic"; they are fairly nontoxic (though, at times, with new unpredicted toxicities, like acneiform skin rashes), and are best administered on a continuous long-term protocol.

A very competitive area at present is in fact represented by the pharmacology of epidermal growth factor receptor (EGFR) targeting. EGFR signaling pathways play a key role in the regulation of cell proliferation, survival, and differentiation. Monoclonal antibodies, mAbs, among which C225 (Cetuximab), and "specific" tyrosine- kinase inhibitors, TKIs, like ZD1839 (Iressa) have been developed, clearly differ in their mode of action at target level. The primary action mechanism of C225 (a chimeric mAb) is a competitive antagonism of EGFR. Independent of the phosphorylation status of the receptor, the EGFR-C225 complex is subsequently internalized. On the other hand, TKIs act on the cytosolic ATP binding domain of EGFR by inhibiting EGFR autophosphorylation. Depending on the nature of the TKI, EGFR inhibition can either be reversible, as with drugs like ZD1839 and OSI-774, or irreversible, as with PD 183 805.

There are still rather few published studies dedicated to examining in detail the various possible mechanisms of resistance to EGFR targeting, irrespective of the anti-EGFR drug considered. There is also some confusion as to the precise cytosolic and nuclear downstream effects following EGFR targeting, particularly when combined with conventional cytotoxic drugs. Moreover, although some consensual findings tend to suggest a link between the level of the EGFR protein target (as detected by IHC) and the intrinsic efficacy of the targeting drug, more pharmacodynamic studies *in vivo* are needed to establish convincing conclusions regarding EGFR levels and targeting efficacy on which clinical strategies can be based with confidence.

Some new, but not unexpected toxicities have already been described for many of these new targeted therapies. For example, the anti-VEGF monoclonal antibody Bevacizumab (Avastin), indicated in combination with intravenous (iv) 5-FU based chemotherapy as first line or second line treatment for patients with metastatic carcinoma of the colon or rectum, can result in the development of gastrointestinal perforation (2% of cases) and wound dehiscence, in some cases resulting in fatality. The appropriate interval between termination of Avastin and subsequent elective surgery required to avoid the risks of impaired wound healing/wound dehiscence has not yet been precisely determined. Moreover, this exciting new therapeutic molecule can also cause unexpected hemoptysis, and in early studies with squamous lung cancers, the incidence of serious or fatal hemoptysis reached up to 31%.

These considerations can lead to some methodological problems in the design of clinical trials. The primary endpoints in phase III randomized clinical studies should remain those reflecting survival of patients, but secondary end points should include measurable quality

of life, disease-related symptoms, performance status, and time to disease progression.

On the other hand, primary endpoints in phase I and II studies should include, besides toxicity and classical pharmacokinetics, data on relative efficacy and sophisticated *in vivo* pharmacodynamics, ranging from changes in genomics (gene expression profiling) and proteomics (protein expression, intracellular localization, and posttranslational modifications) of key targets and downstream effector molecules in tumors in response to the new drug. Additional surrogate endpoints to be considered in phase II studies include relevant *in vivo* dynamic changes on positive emission scanning (PET), new dynamic molecular imaging like dynamic enhanced magnetic resonance imaging (DCE-MRI), and other methods to quantify changes in the growth kinetics of tumors, and for the subset identification of cancers that respond (*versus* those that do not respond) to the new therapies.

3
Main Types of Molecular Markers of Cancer Cells: Genetic and Epigenetic Changes

There are different types of tumor markers:

1. *Genetic markers*: In both hereditary and nonhereditary tumors. These genetic changes can be detected by conventional DNA sequencing techniques, Real-Time Kinetic PCR-based techniques (RT-PCR), gene expression microarrays and DNA microarray-based sequence analysis, comparative genomic hybridization (CGH), fluorescence *in situ* hybridization (FISH), and a variety of new rapidly evolving molecular methods.
2. *Pharmacogenetic markers*: Pharmacogenetic markers are hereditary genetic polymorphisms that may influence the *in vivo* anticancer drug metabolism (activation of pro-drugs, or de-activation), or, perhaps from a more relevant clinical point of view, the likelihood of drug-related serious toxicities because of inherited metabolic polymorphisms. Thus, as an example, genetic variants in the UDP-glucuronosyltransferase 1A1 gene products can predict the risk of subsequent treatment-related severe neutropenia by irinotecan, a fairly active cytotoxic drug in colorectal cancers.
3. *Cellular and tissue markers*: Detected by conventional histopathology or immunohistochemistry, IHC, or by more sophisticated proteomics or genomics technologies; for example, Kodadek has described "protein-function arrays" and "protein-detecting-arrays", and Kononen et al. have described "tissue microarrays" that can allow the molecular study of several hundred tumor specimens in a single paraffin block.
4. *Circulating cancer markers*: The ones traditionally used in the clinic include a variety of glycoproteins, oncofetal antigens, hormones or pro-hormones, tissue-related secreted products, and so on. Some of these are (with the main cancer types associated to increased serum levels): CEA (colon cancer, lung cancer, breast cancer, etc.), CA125 (ovarian cancer), PSA (prostate cancers), Beta-2-Microglobulin (lymphoproliferative diseases), LDH (a nonspecific marker of tumor bulk and aggressive disease in Non-Hodgkin's Lymphomas and other cancers), CA19.9 (pancreatic cancers and other cancers), CA 15.3 (breast cancer and others), HCG-*beta* (nonseminomatous

germ cell cancers), AFP (hepatocarcinomas and nonseminomatous germ cell cancers), Immunoglobulin levels (multiple myelomas), and Thyroglobulin (Thyroid carcinomas). More recent markers detectable in serum or other body fluids (including urine, feces, sputum or bronchial washings) include DNA fragments of oncogenes, soluble growth factor receptors (like HER-2), cytokines (IL-6 for metastatic renal cell carcinoma), and other substances.

Some of these markers are already used routinely in clinical practice (e.g. several circulating cancer markers are useful for the diagnosis, prognosis, and follow-up of some cancer types), while others are being investigated as a source of important prognostic information or even as predictors of response to chemotherapy or radiotherapy (e.g. several cellular and tissue markers). Still others are being explored, in the context of genetic counseling, as potentially useful to screen for hereditary cancer predisposition, for example, BRCA-1 and BRCA-2 for breast and ovarian cancers; RERs and several MMR genes (hMSH2, hMLH1, hMSH6, hPMS1, hPMS2, etc.) for families with hereditary predisposition to colon cancer (HNPCC); and the APC gene in families with FAP (familial adenomatous polyposis).

One of the most peculiar types of molecular markers in human cancer cells is the increased expression of normal or abnormal ''Telomerases,'' and significant progress is being made in understanding how it works. In contrast, despite being implicated in many important regulatory pathways including DNA repair, protein ubiquination and cell cycle control, the exact mechanisms by which inactivation of the tumor suppressor gene BRCA1 or 2 might lead to malignant transformation, particularly in families with high risk of breast and ovarian cancers, remains speculative.

5. *Epigenetic markers*: I shall describe very briefly some of the described ''epigenetic'' markers. Unless genetic therapies develop to the point where it will truly be possible to perform *in vivo* some form of curative ''genetic surgery,'' for example, either replacing a missing tumor suppressor gene with a normal wild-type counterpart, or removing an activated oncogene, in every single cancer ''stem cell'', or cancer ''progenitor cell,'' and without unbearable side effects or deleterious long-term consequences, it is perhaps wiser and more relevant to consider as primary targets for therapy the epigenetic and protein elements associated with carcinogenesis, rather than just the ''genetic'' elements.

We should not forget that, ever since Paul Ehrlich described them in 1913, the classic biochemical targets for drug action are:

1. Inhibition of the active site of an enzyme (or competition with the natural ligand of a hormone or growth factor receptor);
2. Allosteric inhibition and false feedback inhibitors of enzymes or growth factor receptors;
3. Other inhibitors acting outside the active center (exoinhibitors);
4. Double blockade (e.g. where pairs of compounds are used together to block a single biochemical pathway).

The recent progress in the physical mapping of the Human Genome is already pointing toward a ''Post-Genome Era.'' Automated or semiautomated devices capable of reading thousands of genes are already

available. Genomics and proteomics are here to stay, but their routine use in the clinic obviously requires proof of efficacy and judicious use. Apart from technical problems inherent in these techniques, which time will solve, there are some general obstacles to realizing their promise:

1. the lack of genetic markers to cover 100% of all tumors;
2. the lack of knowledge about the functional implications of most genetic defects (the precise maps of these regulatory pathways are still under investigation, and may be to some extent tissue specific);
3. the widespread location and heterogeneity of many of the gene mutations (particularly in large genes);
4. the lack of knowledge about the prevalence of these genetic defects in an apparently healthy population (people with no detectable cancer or histologically malignant lesions; for example, presence of molecular lesions in endometriosis or in apparently normal oral or bronchial mucosa);
5. the lack of knowledge about the significance (in terms of risk of developing cancer) of each of these genetic lesions individually or in different combinations. For example, which combinations of genetic lesions are more relevant to neoplastic transformation in a given tissue?

Epigenetic markers of cancer are probably secondary and later events in most of carcinogenesis, but their importance had also been suspected for several decades. The relevance of these epigenetic changes, and particularly of changes involving the three-dimensional conformational structure of chromatin, making it more or less accessible to polymerases and Transcription Factors, have enabled these markers to gain popularity, and they can have significant therapeutic potential.

Methylation of DNA serves as an epigenetic method of modulating gene expression, and it has been demonstrated that hypermethylation silences a number of tumor suppressor genes including most of the CKIs. Epigenetic mechanisms of carcinogenesis are increasingly coming into focus. Recent findings indicate that aberrant methylation can also be detected in the smoking-damaged bronchial epithelium of cancer-free heavy smokers, suggesting it as an ideal candidate biomarker for lung cancer risk assessment, and potentially useful for the monitoring of chemoprevention trials. Of 12 lung cancer cell lines lacking a deletion or critical mutation in *Rb* or *CDKN2A*, which codes for the CKI p16(INK4A) protein, all were found to be methylated at exon 1 of the *CDKN2A* gene.

Other studies have also shown that underexpression of this gene *CDKN2A* is frequently seen in early preinvasive lesions of squamous cell carcinoma of the lung and cervix.

Some recent work in Korea, at the Samsung Biomedical Research Institute, suggests that tumor-specific methylation of the *p16* gene, and other lung cancer-related genes, like Ras association domain family 1A (*RASSF1A*), H-cadherin, and retinoic acid receptor *beta* (*RAR-beta*) genes may be suitable biomarkers, detectable by methylation-specific polymerase chain reaction (MSP), for the early detection of nonsmall cell lung cancers in bronchial lavage fluids of smokers.

While at least three functional DNA methyltransferases have been identified, the most abundant is DNMT1, which is mainly responsible for methylation during DNA replication. DNMT1 localizes at replication foci, at least in part by interacting with the proliferating cell

nuclear antigen known as PCNA, which is a subunit of DNA polimerase *delta* (that can also bind to the CKI p21 protein). The enzyme DNMT1 catalyzes the covalent addition of a methyl group from a donor S-adenosyl-methionine to the 5 position of cytosine, predominantly within the CpG dinucleotides in highly repeated transposable elements.

These elements are termed *parasitic* because of their resemblance to viral DNA and their ability to move between different chromosomal sites. During embryogenesis there is an initial generalized demethylation of DNA, followed by a specific adult pattern of methylation. While roughly half of the 5′ promoter proximal elements contain CpG islands, they are not usually methylated in normal tissues. But cancers display a particular pattern of methylation. Overall they are hypomethylated, but they do have specific regions of hypermethylation, often associated with regions rich in tumor suppressor genes.

Several molecular variations of deoxycytidine have been developed to inhibit DNMT1, each modified at position 5 of the pyrimidine ring, and these have been reviewed recently. Four agents have been employed clinically: 5-azacytidine (azacitidine), 5-aza-2′-deoxycytidine (decitabine), 1-*beta*-D-arabinofuranosyl-5-azacytosine (fazarabine) and dihydro-5-azacytidine (DHAC). These agents are not new. Some are almost 40 years old, but other analogs such as ara-C and gemcitabine (already in clinical use as anticancer agents) do not inhibit methylation of DNA.

Because methylation of CpG islands in promoter-proximal regions is uncommon in normal cells, normal gene expression should be largely unaffected by these agents. But, at least in some leukemic cells, decitabine has also been found to increase the expression of the multiple drug resistance MDR1 gene product, and this could be clinically counterproductive.

Some studies with these inhibitors go back to the 1980s, for example the EORTC studies with decitabine, and others have been completed more recently. In general, response rates to these agents in solid tumors have been rather low, but more promising results have been obtained in hematological malignancies. Toxicities have been similar to conventional cytotoxics: granulocytopenia, thrombocytopenia, alopecia, nausea, vomiting, and diarrhea. The reason why these drugs have not yet translated into better clinical outcomes, in spite of a sound biological basis for their activity in human malignancies, is not clear.

Acetylation and deacetylation of histones can also provide new insights into how to modify the gene expression patterns of normal and cancer cells.

Methylation of DNA has been shown to induce recruitment of histone deacetylase, which inhibits transcription, and agents have been devised that inhibit this deacetylase. The combination of a DNMT inhibitor and a histone deacetylase inhibitor might be a logical way to activate tumor suppressor gene expression (e.g. CKIs), and a number of preclinical studies are in progress.

Chromatin is no longer considered to be a passive scaffolding for nuclear DNA. At the heart of chromatin's structure is the nucleosome, a complex of DNA wound around an octamer containing two molecules each of histone proteins H2A, H2B, H3, and H4. Histones contain two distinct domains. These are: the protein's amino-terminal tails, which protrude from the nucleosome core, rich in charged amino acids such as lysine and arginine residues; and the globular histone core domain that is mainly responsible for the

histone : histone interactions involved in nucleosome formation.

A number of posttranslational modifications occur in the histone tail domain, including acetylation, phosphorylation, ribosylation, methylation, ubiquitinylation, and glycosylation. Of these, the most widely studied is the acetylation of histone tail domains on specific lysine residues.

The steady-state levels of histone acetylation in the cell are maintained by a delicate balance between the action of histone acetyltransferases (HATs), and histone deacetylases (HDACs).

A number of methodological breakthroughs have allowed a better understanding of this process. Tightly compacted nucleosomes forbid general transcription factors, such as TFIID and RNA polymerase II holoenzyme, from interacting with promoter sequences of specific genes. It is becoming clear that many transcriptional activators actually direct two different types of chromatin remodeling enzymes to specific promoters: an ATP-dependent remodeling enzyme and a histone acetyltransferase.

The field of chromatin research was revamped in 1996 when C.David Allis and colleagues published evidence that the yeast Gcn5p protein, already known to positively regulate gene transcription, was actually a HAT. This acetylation may induce a conformational change that weakens nucleosomal DNA : histone interactions, making the DNA more accessible to transcription factors. Many other transcriptional coactivators have also been recently found to contain HAT activity. Other studies have identified several HDACs (including the mammalian HDAC 1 to 8), as well as large multiprotein complexes, including the so-called *RbAp-46* and *48 (retinoblastoma and histone binding proteins)*, that can undo histone acetylation and modify gene expression.

Recently, Strahl and Allis have proposed the existence of a "histone code," whereby nonhistone proteins read the patterned display of various modifications on one or more histone tails, resulting in the regulation of downstream transcriptional events.

Considering all this new knowledge, it should come as no surprise that HATs and HDACs have also become potential targets for cancer therapy. They appear to act as both oncogenes and tumor suppressors, depending upon the genetic and epigenetic contexts. One interesting example is the variety of chromosome translocations, (often an undesired consequence of previous cytotoxic chemotherapy), that can affect the mixed lineage leukemia (MLL) locus. Thus, a MLL-CBP fusion protein maintains the HAC activity of CBP, and it can be highly oncogenic as a result of the specific targeting of the fusion protein to certain genes.

The field of HATs and HDACs modulating drugs is promising, but highly complex. To achieve selectivity, HATs must be highly targeted to specific genes, and this targeting depends on several structural protein motifs such as "bromo domains" and "PHD" fingers.

Allosteric inhibitors might be more selective and clinically useful than active site inhibitors. But traditionally, chemists have learnt how to develop suitable pharmacological compounds to interfere with active sites and substrate binding, rather than to interfere with the allosteric properties of a regulatory protein, or protein complex. Even modern automated or semi-automated "high-throughput" screens are usually designed for compounds affecting enzyme activity, rather than for those altering protein–protein interactions.

This is a major challenge for the whole field of new anticancer drugs development. Learning how to modify the allosteric properties of oncogenic proteins should become an important aspect of anticancer drug research.

4
The Era of 'Targeted Therapies'

4.1
Threats and Opportunities

Although the term "target-oriented therapies" has only recently been popularized in oncological literature, it must be acknowledged that hormone receptors (e.g. estrogen receptors in breast cancer, and testosterone receptors in prostate cancer) have been used as therapeutic targets by clinical oncologists for a rather long time.

In fact, even the antimetabolite 5-fluorouracil (5-FU), that is a common constituent of many conventional chemotherapy protocols (in 2002, for example, over 2 million patients with cancer were treated worldwide with this drug), was chemically synthesized back in 1957 by Heidleberger, but can also be regarded as a biochemically targeted molecule. This is because 5-FU's activity is based on the observation that uracil is preferentially used by cancer cells, and that fluorinated analogs of this base might "selectively" inhibit the functioning of DNA and RNA in cancer cells, by forming stable ternary complexes with the enzyme thymidylate synthase (TS), resulting in thymidine starvation and inhibition of DNA synthesis.

4.1.1 Threats

From the therapeutic research point of view, the real impact of Molecular Oncology is that it has provided us with many more potential targets for drug discovery than we can handle with our present ability to understand the real priorities, and our capacity to develop valid anticancer drugs in a relatively short period of time. As we move toward this new exciting era of cancer therapies ("Targeted Therapies") we face several key problems:

1. Current cytotoxics have a very limited curative potential, except in rare cancers.
2. Most current cytotoxics are relatively "old drugs" and their patents are running out. New anticancer drugs are needed in the short term.
3. There seem to be more "potential molecular targets" for therapy than true industrial capacity to develop new effective drugs in the short term. In general, once the relevant oncoprotein is identified and purified, gene cloning allows the production of sufficient quantities to allow the determination of its main molecular mechanisms (catalytic or regulatory) and its three-dimensional structure. Appropriate molecules (e.g. those developed by empirical methods like high-throughput screening or by rational drug design) can then be tested *in vitro* and in preclinical models to determine their activity and toxicity, pharmacokinetics and pharmacodynamics.
   More and more oncological units will be devoted to clinical testing of new drugs, and cancer research is likely to undergo a rapid growth, provided enough resources are made available. It could, perhaps, prove globally more rewarding to concentrate on normal regulatory proteins (e.g. upstream on the cell membrane, cytosolic or associated to organelles, like mitochondria or proteasomes, or at downstream "bottlenecks", or points of "cross talk") than

on mutated oncoproteins. The problem in this case, however, is that inhibition of normal downstream regulatory oncoproteins might prove more toxic than selective inhibition of mutated oncoproteins, or inhibition of upstream membrane-bound regulatory elements. Clinical oncologists, on the other hand, are well trained to cope with important toxicities, and highly effective methods to reduce nausea and vomiting, or drug-induced myelosuppression and serious infections have been developed over the past 15 years.

4. The average cost of bringing a new anticancer drug to market is still around 500 to 800 million US dollars, and the time from discovery to the market remains around 10 to 15 years. Both need to be considerably reduced.
5. New "target oriented" therapies will probably be more effective, and less toxic, but also more narrow in their indications, thereby limiting their potential markets.
6. New targeted therapies will require more sophisticated diagnostic tools than conventional cytotoxic chemotherapies, in order to correctly identify the right subset of patients most likely to benefit from such therapies. Conventional histopathology or immunohistochemistry (IHC) will not be enough to achieve this. This will add to overall costs.

   High levels of c-erbB-2 expression or c-erbB2 amplification must be used (by IHCS or FISH) to identify patients for whom Herceptin may be of benefit, but even in this group of patients only some 50% will show objective responses to the combination of Herceptin plus chemotherapy (30% if chemotherapy alone is used). The reasons for this are unclear. Theoretically, Herceptin, by interfering with signal transduction emanating from HER-2 (again, locking the oncogenic molecule into an inactive or less active form) should reduce the potential proliferative competitive advantage of cancer cells in comparison with normal cells.

   However, the additive or synergistic effects of cytotoxic drugs could be caused by as-yet-unknown effects of these drugs downstream on the signaling pathway. As a matter of fact, other inhibitors of EGFR (for example, inhibitors of the EGFR associated tyrosine-kinase activity), such as gefitinib, have not so far shown any significant survival benefit when used in the treatment of nonsmall cell lung cancer patients, in combination with platinum-based cytotoxics.

   Moreover, the relationship between tumor marker levels and sensitivity to new therapies may not be linear, or may depend on yet unknown circumstantial factors.

   For example, HER-2 inhibitors (in the case of Trastuzumab) can induce objective response rates in breast cancer patients, but often without clear evidence of a dose–response relationship in their response or survival, or in adverse events. The main predictor of response may be the degree of expression of the target molecule, as well as the intrinsic relevance of this target molecule to the cancer phenotype. Thus, response rates for single agent Trastuzumab in 111 assessable patients with metastatic breast cancer was 35% for patients whose tumors showed a 3+ staining of HER-2 by IHC, as compared to 0% for the case of tumors with only a 2+ staining. This suggests an interesting nonlinear relationship between marker expression and clinical response. The message, as

linguists have known for a long time, is that: 'the same word in a different context can give the final sentence a very different meaning'.

Thus, drug companies are well advised to present their phase II and III clinical studies with adequate and comprehensive biochemical (genomic or proteomic, or both) pharmacodynamic data to fully persuade health authorities and clinicians about the benefits of the new therapeutic agents, and the indications for their use.

7. The applications of pharmacodynamics are becoming increasingly important. We should be able to measure, in any given tumor, the direct and indirect genomic and proteomic effects of whatever changes a new drug is producing, or is meant to produce. We should also be able to put this usually complex type of information into some kind of model, so that it can be extrapolated to other tumors with different molecular compositions. Not an easy task, but one of the most exciting challenges ahead. Sound understanding of clinical and biological knowledge, as well as intuitive parallel thinking, good communication skills and an open collaboration within multidisciplinary teams can be really helpful. Paul Workman (2004) has suggested an "Audit Trail" for each new "target oriented" anticancer drug in development:
   - *Is the target expressed and is the pathway active?* It is important to understand the relationship among target expression, pathway activity, and response to a therapeutic agent.
   - *Are sufficient concentrations achieved in plasma, blood, and tissues?* Obviously, if for pharmacokinetic reasons adequate concentrations of the drug are not achieved in animal studies, or in patients in phase I studies, it is pointless to proceed forward with that particular drug structure.
   - *Is there activity on the desired molecular target?* For example it is common, when studying the activity of a particular kinase inhibitor, to assess the phosphorylation of a downstream substrate. Lack of activity on the target can be a reason for lack of therapeutic benefit, though it must be said that some of the new target-oriented therapies can lead to clinical benefit by apparently acting on targets (e.g. with similar "active sites") not previously foreseen or identified.
   - *Is there modulation on the desired molecular target?* It may be useful also to look at "off-target effects," since the *in situ* regulatory feedback circuits can lead to unexpected biochemical changes.
   - *How can we achieve the desired biological effect?* For example, which regulatory pathway (cell cycle control, apoptosis, necrosis, etc.) is to be predominantly interfered with?
   - *Is a clinical benefit seen (in objective tumor response, growth inhibition, time to disease progression, survival)?*
   - *Finally, but very importantly, is the molecular target really key in the pathogenesis of that particular cancer?* If so, as seems to be the case, for example, for the abl-bcr fused genes in CML, or the EWS gene translocations and rearrangements observed in many human sarcomas, then hitting the target *per se* would have beneficial therapeutic consequences. Otherwise, only one target

will not be enough, and multitargeting will be required for good clinical efficacy.

8. Some form of dynamic and efficient (''parallel or sequential'') coordination of projects between Diagnostics and Pharmacology, and *vice versa*, is also highly desirable. Clinicians are already faced with increasing numbers of difficult therapeutic choices, and it would be naive to believe that this increasingly complex decision-making process will be based exclusively on simple classical clinical or pathological characteristics such as the age of the patient, performance status, TNM stage (which describes the degree of local or distant growth and spread of the tumor), and conventional histopathological degrees of differentiation (SBR in breast cancer, or Gleason score in prostate cancers).

Let us look, for example, at the present clinically complex situation in two very common cancers:

1. *Colon cancer*: The introduction into clinical practice of 5 new drugs – the cytotoxics Irinotecan and Oxaliplatin, the oral cytostatic Capecitabine, and the monoclonal antibodies Cetuximab (an antagonist of EGFR), and Bevacizumab (best described as an inhibitor of angiogenesis that ''traps'' VEGF) – , in addition to the old classic antimetabolite ''5-Fluorouracil,'' has dramatically increased the number of possible permutations and combinations of all active drugs in metastatic colon cancer to the extraordinary figure of 720 possible choices.
2. *Breast cancer*: A special case of targeted therapies is that of endocrine therapies for cancer, which actually constitute the first historical example of targeted therapies. For example, inhibition of estrogen (in breast cancers), or testosterone (in prostate cancers) action or production can probably be regarded as the first examples of anticancer-targeted therapies. Beatson first observed tumor regression of breast cancer in some patients following oophorectomy in 1896, but it was not until some 60 years later that estradiol receptors were identified by following the localization of radioactive estradiol in the immature rat. How hormones influence growth and behavior of malignant tissues was not known in those days, and even today a lot of questions remain unanswered.

Approximately two-thirds of all breast cancer patients have estrogen-receptor positivity in their tumor, and about half of them respond to endocrine therapy, at least initially. The reasons why the other half or so of patients with estrogen-receptor-positive tumors do not respond to endocrine therapies are rather complex and not completely understood.

Somewhat similarly, at least 80% of prostate cancer sufferers initially respond to androgen withdrawal, but regretfully the mean treatment response duration in prostate cancer is only about two years. Both types of endocrine therapy have both cytostatic and cytotoxic properties for cancer cells. Perhaps surprisingly, neither estrogen receptors (ER) nor androgen receptors (AR) have been described as proto-oncogenes, even though, at least theoretically, their mutation into activated forms could lead to some competitive advantage to affected cells. The truth is that, in contrast to the large numbers of missense AR mutations that have been found in association with androgen insensitivity syndromes, there do not seem to be many AR mutations that

predispose to human prostate cancer, with the exception perhaps of a missense mutation present in the germline of 0.33% of the Finnish population and 1.91% of Finnish prostate cancer patients.

For many years, Tamoxifen, a drug that binds the estrogen receptor with high affinity and that globally acts as a partial agonist/antagonist, was the only oral endocrine therapy available for most breast cancer patients, with estrogen-receptor positive tumors. Now, however, the emergence of several new antiestrogens (the so-called *Aromatase Inhibitors*, for example, of which three different molecules have already been approved for marketing), and some clinical studies suggesting that LHRH-agonists may be equally, or more, active than conventional low-intensity chemotherapy (like CMF) as adjuvant therapy in young premenopausal women, has made decision-making algorithms significantly more complex for both clinical oncologists and breast cancer patients.

For example, as recently reviewed by Hortobagyi, with regard to conventional cytotoxic chemotherapy for breast cancer we have gradually learnt, over more than three decades of work, that the use of CMF (cyclophosphamide, methotrexate, and 5-FU) as "adjuvant chemotherapy," following primary breast surgery (to treat potential microscopic residual or metastatic disease), can reduce the odds of cancer recurrence by some 23.5%, and of cancer-related death by 17%. The use of anthracycline-containing regimens, though usually more toxic and associated with iatrogenic hair loss (alopecia) and potential cardiotoxicity, improves results by another 11 and 12% for odds of recurrence and death, respectively. Finally, the introduction of taxanes (that alter microtubular assembly in dividing cells, among many other biochemical changes) in addition to anthracyclines may result in a relative reduction in odds of recurrence, compared with no adjuvant systemic therapy, of some 40 to 60%, depending on age and hormonal status of the patient.

If tamoxifen, the partial agonist/antagonist of estrogen, is used alone as adjuvant therapy in postmenopausal women with estrogen-receptor postive breast cancers, a five years duration of treatment implies a reduction in the risk of local cancer relapse of close to 47%, and an impressive 26% reduction in the risk of cancer-related risks. A more recent large study, conducted by the National Cancer Institute of Canada, has shown that if five years of Tamoxifen, in a premenopausal estrogen-receptor-positive group of breast cancer patients, are followed by five years of an oral aromatase inhibitor called "Petrozole" (that effectively acts as a peripheral "antiestrogen"), the risk of local relapse of cancer is reduced by between 43 to 53%, depending on axillary lymph node status, with a 46% reduction in the risk of contralateral breast cancers, and a fairly significant increase in disease-free interval (and perhaps overall survival), as compared with the group of patients receiving tamoxifen alone.

To make matters even more complicated on the subject of hormonal treatments of breast cancers, a recent study with a different aromatase inhibitor (exemestane) has shown that three years of tamoxifen followed by two or three years of exemestane, produces better clinical results than the conventional five years of adjuvant Tamoxifen.

Bearing in mind these results and other considerations, even nonclinicians will easily understand why explaining the various treatment options to cancer patients and relatives, together with

adequate explanations on primary treatments, like the radical surgical approach (e.g. mastectomy), *versus* the conservative (e.g. ''tumorectomy plus radiotherapy'') approaches, or the new ''sentinel node'' technologies (e.g. to spare radical axillary lymph node dissection), tend to occupy increasing amounts of time during the initial clinical consultation, and are becoming one of the most difficult tasks for all clinical oncologists.

Conclusion on the threats Unless we find new and more efficient (faster and less expensive) ways to develop new anticancer drugs, and new and more rational (as well as more practical) criteria for using them in combination, the majority of health systems will simply be unable to cope with the expected ''avalanche'' of new medications.

Professionals will soon be in difficulty when deciding on which combination of drugs to use in any given patient, and in exactly which combination or sequence. Patients themselves (gradually better informed, and naturally keen to actively participate in the therapeutic decision process) will become more confused, and perhaps more vulnerable to biased information or misinformation; and, last but not least, many, if not most health providers may eventually be unable to pay for all new drugs and procedures, considering the ever-increasing costs.

If we also add the rapid demographic changes (that lead to an elderly population with many more cancers) in the developed world, and the increasing political pressure to control pharmacy costs, we come to the inevitable conclusion that dealing with the cancer problem is bound to get initially more complex and expensive, before it becomes rationally more manageable and less expensive.

Molecular Oncology has a lot to do with all this, both as a partial cause of the problem, and as a possible rational solution in the mid- to long-term.

More should probably also be spent on prevention/early detection, and research on ''chemoprevention.'' Thus, health authorities may soon ask themselves whether it makes sense to increase by 20to 40% the overall costs of treating, for example, metastatic colon cancer or advanced lung cancers, to gain only a year or less of overall survival, rather than adopting more aggressive policies to induce people to undergo more regular colonoscopies,, under sedation, from the age of 50 (say, three times in a lifetime, at 50, 60, and 70 years of age, in the absence of high risk factors) or to simply quit smoking.

The case of AIDS is a good example. Besides a better general knowledge and control of risk factors in the populations at risk, at least within the developed world, the introduction into the clinic, well over a decade ago, of ''expensive'' combinations of new antiretroviral oral therapies has dramatically increased the life expectancy of HIV-infected cases, from 3 to 5 years at first, to an almost normal life expectancy or, in any case, well over 15 years of life. In this case, chronic expensive therapies ''make sense,'' and are easily justifiable, even to the most rigorous and cost-conscious public health systems. Even under these favorable circumstances, given the fact that the HIV retrovirus in most infected patients often produces several mutant molecular forms each day, the problem of drug resistance remains. This justifies the molecular sequencing of the genome of HIV mutants and, based on the genetic variations detected, changing the individual medication to other suitable antiretroviral drugs, or putting the patient into new drug testing programs.

4.1.2 Opportunities

1. The activity of many of the cytotoxic agents used in clinical practice today was first detected by traditional screening methods, as in the case of murine leukemias (L1210 and P388) and murine solid tumors. But these cytotoxic drugs were essentially designed to interfere with DNA or chromosomal replication, because we were ignorant of the molecular basis of cancer.
2. As explained in this chapter, we are beginning to have enough information on the precise molecular mechanisms involved in the regulation of "cell behavior" (benign or malignant), and we shall soon be prepared to classify all the "players" (regulatory proteins) into "processes" (regulatory pathways in disease).
3. It is now possible not only to tell which are the key players in each pathway, but also how each pathway is organized (from upstream to downstream), and where we should intervene therapeutically to correct the malignant behavior of tumor cells (e.g. inhibiting angiogenesis, cell growth or the metastatic process), or to eradicate malignant cell clones (by facilitating senescence and/or apoptosis, or by stimulating the destruction of tumor cells).
4. In simplistic terms (see Fig. 3 for graphic explanation) the new therapeutic drugs should either:

   – block or inhibit activated proto-oncogenes in their inactive state, like Gleevec in the case of several Tyrosine Kinases, for example; or
   – considering the fact that most regulatory molecules have binary states (active/inactive), somehow lock or stabilize tumor suppressor proteins in their active state, which is the one normally responsible for inhibiting entry into S-phase or mitosis, and for maintaining the normal differentiation and apoptotic pathways.
5. In other words, as regulatory proteins are usually present in normal cells in either an active form or inactive form, and the chemical equilibrium between the two forms depends on the regulatory microenvironment of the cell, new targeted therapies should be able to correct, or at least limit, the malignant phenotype by restoring the normal balance between stimulatory and inhibitory signals within the cell.
6. Examples of these new types of "oncogene inhibitors" drugs are either already on the market (e.g. Gleevec or Herceptin), or in development (EGFR inhibitors, VEGF inhibitors).

   The case of tumor suppressor genes is more complex, because the ideal therapy in these cases, where the problem is usually a loss-of-function mutation of the promoter (e.g. a deletion or a hypermethylation), would be a genetic replacement of the wild type. However, because of mainly technical reasons (lack of suitable vector, inefficiency of the transformation systems, lack of stability of "transfected cells," danger of oncogenetic transformation by retroviruses, etc.), the truth is that the original research enthusiasm of the 1990s regarding the clinical therapeutic application of genetic therapy programs no longer exists, even if new approaches and new techniques are under active investigation. However, as I postulated in the Second Edition of "Principles of Molecular Oncology" (Humana Press, N.J., 2004) and also in this article, there is a possibility of developing low-molecular-weight

drugs capable of inhibiting the ''inactivation of tumor suppressor function,'' for example, inhibitors of Rb phosphorylation or activating ''cryptic'' tumor suppressor functions.
Because of the presence of either wild-type or mutant p53 protein in most human tumors, the restoration of endogenous p53 function in cancer cells by small-molecular-weight compounds (as shown by Bykov et al, 2002), or by the *in vivo* delivery of transducible, proteolytically stable p53C' d-peptides such as RI-TATp53C' (as recently shown by Snyder et al, 2004) have become promising new approaches to cancer therapy.

7. In this way, we can hope to eventually reduce the number of potential therapeutic targets from several hundreds to perhaps only twenty or thirty major primary targets. Most of these processes, if not all, will have upstream regulatory elements (e.g. membrane-bound or membrane-associated), intermediate elements (cytosolic or bound to organelles, like mitochondria) and downstream elements (transcription factors, DNA- or RNA-specific sequences). A critical and exciting phase will be the development of models based on the abnormal circuitry of regulatory pathways in cancer cells. This would help us to find the ''Achilles heel'' of the particular cancer (Weinstein, 2002), or the right combination of targets to tackle.

The main targets for cancer therapy identified so far can be summarized as follows:

Angiogenesis Biological process that leads to the formation of new blood vessels, and that involve the proliferation of endothelial cells and their morphological association into functional capillaries. The ''angiogenesis switch'' refers to the set of changes in gene expression and protein function that can change the natural local tissue balance between ''inducers and inhibitors of angiogenesis,'' for example, secretion of growth factors like FGF (Fibroblast Growth Factors) or VEGF (Vascular Endothelial Growth Factors), or changes in membrane receptors associated with the malignant invasive phenotype and growth of tumors.

Hypoxia-inducible factors (HIF), cytokines, cellular adhesion molecules (CAMS) like the Integrins family, or secreted proteases like the metalloproteinases (MMPs) can also alter the angiogenic phenotype of tissues.

A great variety of molecules produced by normal mammalian cells can inhibit angiogenesis, such as thrombospondin, angiostatin, endostatin, tumstatin, arresten, and vasostatin. Of the approximately 200 compounds with antiangiogenic activity described so far, several are already in clinical testing, but we are only really beginning to unravel the complexities of the subject, particularly in humans.

One of the most promising approaches is the use of humanized monoclonal antibodies with the capacity to inhibit binding of VEGF to its receptor, or other inhibitors of VEGFR activation. VEGF-A is a highly specific, selective mitogen for vascular endothelial cells *in vitro*, and is produced *in vivo* by many tissues.

For example, Bevacizumab, a humanized monoclonal antibody against VEGF, has already shown reasonably good results in clinical testing up to phase II and phase III studies. It has shown modest activity in breast cancer, but significant prolongation of the time to progression of disease in metastatic renal cancer and, when added to

cytotoxic chemotherapy, it significantly improved survival in patients with metastatic colorectal cancers.

Antiangiogenic therapies would no doubt be ideal adjuncts, or complementary long-term therapies, to primary surgical treatments or to chemo-radiotherapy programs. They should also reduce the problems of resistant phenotypes, considering that these new endothelial cells are normal, and have genomic stability.

"Angiolytic" substances such as Adherins and Integrins, which are inhibitors of the function of several cellular adhesion molecules, are also under investigation. In principle, they are capable of disrupting preexisting tumor blood vessels *in vivo*, thereby leading to interruption of the blood supply, and necrosis of the tumors.

Apoptosis Apoptosis is a biological process that describes the complex mechanisms of "programed cell death," eventually leading to endonuclease driven fragmentation of DNA and chromatin by a number of Caspase-dependent or Caspase-independent biochemical events. Although the mitochondrial release of cytochrome-C appears essential for at least some caspase-mediated events, mitochondrial-independent activation of caspases *via* activation of the Fas-FasLigand pathways (also known as CD95, or APO-1 pathways) is also reasonably well defined. The mitochondria are currently seen as an important convergence point for a number of pro-apoptotic or antiapoptotic members of the so-called *Bcl-2* family of proteins. However, the recent enthusiasm for "apoptosis" as the main, or only, mechanism of tumor cell death can be prudently questioned on several grounds. Except in cells of lymphoid origin, apoptosis is a relatively insignificant mode of cell death for most cells in solid tissues, and experimental conclusions based on measuring the extent of cell kill by cytotoxic agents, rather than on true effects on clonogenic survival of tumor cells, can be misleading. Thus, despite the seemingly strong case that cells die from cancer treatments (chemotherapy or radiotherapy) because of apoptosis chiefly controlled by wild-type p53, much current data do not appear to fit with this simple hypothesis. On the other hand, it seems reasonable to expect that unless new drugs are capable of inducing cell death (by necrosis or apoptosis), the therapeutic effects will be mainly "cytostatic" rather than "cytotoxic."

Clinicians, and patients of course, like to see tumors "shrink" or, ideally, to "disappear," rather than just stay "more or less the same," with "stabilization", that is, no clear tumor growth but no tumor size reduction either.

Binary state (of regulatory molecules) In ancient Chinese philosophy and medicine, the Universe we live in was clearly a model of the binary nature of things. For example, "cold" was considered to be a "Yin" pathogenic factor, and its nature is to slow things down. On the other hand, "heat or fire" was considered a "Yang" pathogenic factor, causing expansion and increased activity. In a similar sort of way, all biochemical regulatory proteins can be present at any one time in an "on" or "off" molecular form, with regard to one or more of their functions.

This "Yin" or "Yang" effect will depend, for the normal wild-type forms of regulatory proteins, on the regulatory microenvironment, and the possible conformational changes (usually allosteric) produced by a number of enzyme activities including phosphorylation/dephosphorylation, acetylation/deacetylation, farnesylation/

defarnesylation, methylation/demethylation, and so on.

Cell Cycle Checkpoints: Mammalian cells can remain quiescent (in Go phase of their cell cycle) for long periods of time, but the right combination of growth factors can induce them to enter G1 and then, if no impediment occurs, progress through S-phase (DNA replication), G2, and mitosis, and complete their cell division. Tumor doubling times can show tremendous variations. For example, some 60% of human breast cancers show tritriated thymidine incorporation rates and labeling indices of less than 4%, indicating a rather slow doubling time of tumor cells, but the rest of the breast cancers can show labeling indices between 4 and 41%. No doubt this reflects the molecular heterogeneity of breast cancers. The best and more selective anticancer drugs of the future, in this cell cycle setting, would have to tag specific cyclins (e.g. cyclin D1 and cyclin E, that can act as *de facto* potent oncogenes) for degradation. Recent research points towards a key role for Cyclin E in breast cancer, and Cyclin D1 in some non-Hodgkin's lymphomas and in colon cancer, where they act as potent oncogenes. Alternatively, marked cytostatic effects should be produced by new drugs designed to stimulate the activity of selective CKIs (Cyclin-dependent kinase inhibitors), or to lock crucial downstream elements (like the retinoblastoma protein complex) that act as *de facto* tumor suppressor genes, in order to enable them to maintain their active roles as cell cycle inhibitors.

Cell shape Surprisingly little systematic research has been published to establish the molecular links between "cell shape" and "cell behavior." Yet it makes sense to assume that both are related. For example, most differentiated epithelial cells adopt a rectangular or polygonal shape, and display a marked polarity (from their attachment to the basement membrane, to the opposite membrane surface on the lumen), as well as a solid cytoskeleton, parallel anchorage to neighboring cells, abundant cytoplasm and endoplasmic reticulum, and low, if any, migration capacity. They also display a low rate of proliferation (except at special sites like the intestinal crypts, where they change shape anyway). In general, as cells take a spindle shape (epithelial-mesenchymal transition, or *vice versa*), they become more potentially mobile and proliferative. Finally, blood cells tend to be round, with a high nuclear/cytoplasm ratio, and are both more mobile and more likely to undergo cell division if stimulated.

The so-called *blast cells*, in leukemias and lymphomas, and also in undifferentiated or poorly differentiated carcinomas, show similar changes in shape, nuclear/cytoplasmic ratio, mobility, and high- proliferative activity.

Downregulation of "adherens junctions" (AJ) assembly by mutations, hypermethylation, or transcriptional repression of the adhesion molecule E-Cadherin, have been described in several malignancies, and particularly in human diffuse gastric carcinomas, lobular breast cancers and familial gastric cancers. In many ways, E-cadherin behaves as a tumor suppressor gene, while *beta*-catenin behaves as a potentially potent oncogene. Thus, when the ratio of free cytoplasmic *beta*-catenin is too high, because (for example) of loss of E-cadherin, or because of APC mutations, an excess of *beta*-catenin is free to travel to the nucleus. This leads to the overexpression of a set of transcription factors (TCF/LEF), and of potent oncogenes such as cyclin D1 or c-myc.

When the intracellular levels of another important protein regulator (named *snail*) rise above a certain threshold, there is usually active repression of the transcription of the *E-cadherin* gene, and a promotion of the so-called *epithelial-mesenchymal transition* (*EMT*), that leads polygonal/epithelioid-shaped cells to convert into spindle/round type of cells.

Chemoprevention Chemoprevention is "the treatment of carcinogenesis, its prevention, inhibition or reversal." The term "chemoprevention" is controversial. Some may confuse the term "chemo" with "chemotherapy", and "prevention" may not be the best word to define "early detection" of cancer biomarkers. The subject is bound to grow very rapidly, both in terms of the identification, validation, and clinical relevance of cancer biomarkers, and also in terms of their impact on the quantitative estimation and prediction of individual human cancer risks.

In the relatively recent past, the possibility has been raised that pharmacological agents or nutritional modification might prevent the development of human cancer, or at least slow down its progression. The concept of chemoprevention is gaining support, although there are still significant experimental and conceptual problems associated with it. For example, if the study population is composed of normal or nearly normal subjects (such as normal volunteers, smokers without obvious disease, or people with a history of one isolated gastrointestinal polyp), very few side effects may be acceptable. In contrast, if the subjects are at high risk of cancer (e.g. have a history of familial polyposis, hereditary predisposition to breast or colon cancer, or second tumors) considerable side effects may be acceptable.

It remains possible that many of the target-oriented new drugs being developed will show clinically relevant efficacy in cancer chemoprevention studies, perhaps even superior to that shown in advanced cancers. The reason is that these target-oriented new drugs usually display tolerable or little toxicity, can be administered orally for long periods of time, and delay the growth of cell clones with a proliferative or angiogenic advantage over their normal counterparts (e.g. by interfering with growth factor signal transduction, or with the angiogenic process).

There is also significant scientific interest in certain metabolic or detoxification phenotypes that may lead to a higher cancer risk, such as aryl hydrocarbon hydroxylase activity (AHH), debrisoquine hydroxylation, and glutathione-S-transferase activity. For example, AHH activity depends on one subfamily of cytochrome P-450 microsomal enzymes that convert polycyclic aromatic hydrocarbons into carcinogenic intermediates. Case-control studies have yielded contradictory results regarding the association between AHH activity in lymphocytes and lung cancer risk. Why only a minority of heavy smokers actually develop lung cancer remains largely unknown.

Beta-carotene, the retinoids, folic acid, vitamins C and E, and tamoxifen have been reported so far in published phase III clinical chemoprevention trials. The retinoids have probably been studied the longest and, although their therapeutic index is somewhat narrow because of toxicities, they have achieved successful results in oral leukoplakia, actinic keratosis, prevention of skin cancer in xeroderma pigmentosum, and prevention of second primary tumors in squamous cell carcinoma of the head and neck. More recent, but rapidly growing, is the role of other agents, such

as cyclooxygenase-two (COX-2) inhibitors in chemoprevention. COX-2, a key enzyme for the production of prostaglandins from arachidonic acid, is overexpressed in colon carcinogenesis and other malignancies.

The role of tamoxifen in chemoprevention of breast cancer has recently reached the front pages of newspapers. A recent report from the American Society of Clinical Oncology (ASCO) lists several conclusions endorsed by the Health Services Research Committee and the ASCO Board of Directors. Tamoxifen, at 20 mg/d for 5 years, may be offered to reduce the risk to women who have a defined breast cancer risk of more than or equal to 1.66%. Risk/benefit models suggest that greatest clinical benefit with least side effects is derived from the use of Tamoxifen in younger premenopausal women who are less likely to suffer thromboembolic sequelae and uterine cancers. Data confirm a substantial reduction in breast cancer risk, but do not yet demonstrate, probably because of a relatively short follow-up, an overall health benefit or increased survival. Tamoxifen is not free from side effects but health-related quality of life is not adversely affected in most cases.

Drug resistance Conventional cytotoxic chemotherapy and radiotherapy were designed to hit the DNA synthetic machinery of rapidly proliferating cells, to alter the integrity of DNA in tumor cells, or to interfere with microtubule assembly and chromosomal movements during mitosis. With these rather nonspecific forms of therapy, some relatively uncommon cancers, like many pediatric tumors, some leukemias and lymphomas, germ cell cancers, and a few other malignancies can be totally cured. And in many other cases, although these therapies do not offer much hope of cure, they can nevertheless provide effective palliation of symptoms, and even (when used as adjuncts to primary surgical resection) increased overall survival, or at least disease-free interval. Although the judicious use of these agents, together with surgical treatments, remains the mainstay of Oncology Therapeutics, their success in curing or totally eradicating the majority of the most common cancers is limited, and usually inversely proportional to the total amount of cancer cells, among other important prognostic factors.

Surprisingly, conventional chemotherapy and radiotherapy work not because cancer cells are especially more ''sensitive'' to their cytotoxic effects than normal cells, but because normal cells usually can recover faster and more efficiently than cancer cells. Normal tissues virtually never develop resistance to chemotherapy, probably because they maintain the mechanisms that allow normal renewing cell populations, like the highly proliferating cells in the normal bone marrow or gastrointestinal tract, and are capable of monitoring and correctly repairing DNA damage, or induce the programed cell death (suicide) of cells with irreparable DNA damage. Unfortunately, many cancer cells that have lost these normal repair mechanisms become resistant to the action of cytotoxic treatments. In 1943, Luria and Delbruck observed that the bacterium *E.coli* developed resistance to bacterial viruses (bacteriophages), not by surviving exposure, but by expanding clones of bacteria that had spontaneously mutated to a type inherently resistant to phage infection. In other words, following a process of Darwinian selection, those clones inherently resistant to the killer phage gradually outgrow the rest of clones when exposed to the killer agent. On the basis of this model, Goldie and Coldman proposed, in 1979, a model that predicted that cancer

cells mutate to drug resistance at a rate intrinsic to the genetic instability of a particular tumor. A large number of potential drug-resistance mechanisms have already been described. Some refer to the existence of multidrug resistance proteins (like the so-called *P-glycoprotein*) on the cell membranes of some cancer cells, and even normal cells of certain tissue types. These proteins can actually actively pump cytotoxic agents out of cells. Others refer to specific mechanisms of resistance unique to each cytotoxic agent. But the predominating general opinion today is that the main mechanisms of resistance to current cytotoxic therapies are more general and, as already described above, more linked to the particular set or combination of genetic and epigenetic defects in DNA-repair pathways, cell cycle control pathways and apoptosis, than to the specific molecular structure of the cytotoxic drug in question.

Growth factors These were defined in the 1970s, long before their molecular characterization and purification, as substances other than nutrients required by cells *in vitro* for their survival, proliferation, and differentiation.

In fact, it was soon discovered that normal cells are simply unable to survive and grow *in vitro* except in the presence of serum, and fetal calf serum became the 'gold standard' critical component for culture medium. Only some transformed cells, or cancer-derived cell lines, as shown by seminal work by authors like Sporn and Todaro(1980) and Sporn and Roberts (1985), were capable of growing in serum-free medium, because these transformed cells either produce their own growth factors ("autocrine" stimulation), or had become "growth factor-independent" because of the activity of Oncogenes and the lack of activity of tumor suppressor genes.

Substantial evidence, derived from the use of recombinant growth factors *in vivo* (growth factors produced by molecular cloning, and expression of the gene product by a suitable recombinant vector), or from experiments with transgenic animals or site-directed mutagenesis, supports the notion that these substances serve similar purposes *in vivo*. Therefore, it is now widely accepted that differentiation, proliferation, and survival of most normal tissue types is controlled, at least in part, by a complex network of regulatory growth factors. These are usually glycoproteins, frequently of low molecular weight, that act at very low concentrations, in the picomolar range, by binding to specific cell membrane receptors. These receptors can be present in both high-affinity and low-affinity forms, they can be absolutely specific for one growth factor or be shared by two or more factors, and they can be membrane-bound, soluble or both. Some signaling systems, like the Wnt, depend also on the combined activity or modulation of membrane-bound "coreceptors." It has been suggested that HER-2 is also a "coreceptor."

Unlike conventional hormones, which are produced by specialized glands, most growth factors appear to be made systemically; stromal cells (e.g. fibroblasts or endothelial cells), inflammatory cells (e.g. macrophages), and activated T-lymphocytes are some of their main sources. Some perhaps surprising findings regarding the production and secretion of specific growth factors, point to the kidney as being the main physiological producer of erythropoietin (that stimulates the production of red blood cells), and the liver as the main physiological producer of thrombopoietin, also called *MGDF* (*megakaryocyte growth and differentiation factor*), a key regulatory molecule in the production of platelet precursors.

Most standard conventional chemotherapy protocols were designed based on the kinetics of recovery of the bone marrow in response to cytotoxic agents, because neutropenia and thrombocytopenia can easily become doselimiting. The successful introduction into clinical practice (in the mid-1980s) of recombinant human colony-stimulating factors, such as G-CSF (Granulocyte Colony-stimulating Factor, or Filgrastim) and GM-CSF (Granulocyte-Macrophage Colony-stimulating Factor, or Sargramostim), led to a marked decreased in iatrogenic morbidity and mortality.

The use of recombinant human colony-stimulating factors has, according to Chu and DeVita Jr (2002) revolutionized chemotherapy treatments, by playing an important role in decreasing the incidence of infections and the need for hospitalization, and in enabling the maintenance of the optimal dose intensity of chemotherapy treatments for chemosensitive malignancies.

Most human growth factors are ''pleiotropic.'' That is to say, they can modify several cell types and several key regulatory pathways at once, rather than being absolutely cell-type-specific, and display some degree of ''redundancy.'' For example, even in the absence (in experimental animal systems like gene knockout mice) of highly cell-type-specific growth factors like the erythropoietin gene (*EPO*) or the granulocyte colony-stimulating factor genes (*G-CSF*), these gene-deficient animals can still survive with background production of endogenous red blood cells (normally driven by *EPO*), and/or white blood cells (normally driven by *G-CSF*).

It would be over-optimistic to expect growth factor inhibitors alone to prove to be highly effective anticancer agents. On the other hand, the combination of growth factor inhibitors and conventional cytotoxics, or other inhibitors of key signal transduction pathways, might prove valid as human anticancer therapies.

Intraepithelial neoplasia Intraepithelial neoplasia (IEN) is characterized as a moderate to severe dysplasia that occurs on the carcinogenic pathway from normal tissue to malignant cancers. Accumulation of genetic and epigenetic lesions results in the typical phenotypic changes associated with cancer cells, including loss of cell cycle control points, defects in DNA repair, aberrant methylation patterns, and defects in the apoptosis pathways. Because this process of carcinogenesis often requires many years, and most of the common epithelial human cancers increase in frequency, at times almost exponentially, from the age of 45 to 55 to old age, identification of IEN by histopathological or molecular means (genomics or proteomics analysis) can offer a window of opportunity for intervention in the progression of the malignancy.

Clinical studies for the early detection, and primary or secondary prevention of cancer (chemoprevention) are already ongoing for bladder cancers, breast cancers, cervical cancers, colon cancers, endometrial cancers, esophageal cancers, lung cancers, ovarian cancers, prostate cancers, skin cancers, and so on.

Oncogenes These are growth control genes that are present in the human genome (as ''proto-oncogenes''), as well as in the genomes of most, if not all, multicellular organisms. Incorrect expression or mutation of an oncogene usually results in gain of function and unregulated cell growth, as seen in malignant cells. Many of them are part of the complex cell's signal transduction pathways, including cell membrane growth factor

receptors, or intracytoplasmic proteins. These signal proteins may be an intracytoplasmic piece of the receptor molecule, or another molecule activated just inside the membrane. Tyrosine kinase is an example of a signal molecule. The first signal may serve only as an intermediary, affecting a change in a second messenger. The G proteins, for example (belonging to the *Ras* oncogene family) are a group of intracytoplasmic second-messenger molecules. The intracellular cascade of signals eventually reaches the nucleus, leading to changes in gene expression and cell behavior. Of the over 100 oncogenes that have been identified so far, most are key players in these signal transduction pathways.

Mutations The molecular changes in DNA associated with carcinogenesis are not very different from those associated with genetic variation and molecular evolution of organisms and species, and include local changes in DNA sequences, rearrangement of DNA segments within the genome, and (in the case of viral carcinogenesis) horizontal transfer of a DNA (or RNA) segment originating in another kind of organism.

Basic definitions of these three different natural strategies that allow for genetic variations have been explained in detail in this same Encyclopedia by the Nobel Laureate Werner Arber (Biozentrum, University of Basel).

Point mutation, for example, means the replacement of a single correct nucleotide within a gene by an incorrect one, and can occur as a result of one of several mutagenic events. Mutagenic chemicals (e.g. from cigarette smoking) enter the cell and bind to DNA, forming a structure called a *DNA adduct* which, during replication of the DNA molecule, will lead to an incorrect nucleotide substitution.

Mutagenic chemicals can damage DNA in many other ways, such as by methylation of a nucleotide. Radiation can cause single strand breaks. In general, it is thought that most, but not all, carcinogens are mutagenic. In contrast, the so-called *Tumor Promoters* are substances that by themselves do not cause mutagenic events, but amplify, directly or indirectly, preexisting cellular clones with mutations.

There are at least 16 ways to reduce or abolish the function of a gene product: delete the entire gene, ensure loss of the relevant chromosome, delete part of the gene, disrupt the gene structure (by a translocation or an inversion), insert a sequence into the gene, inhibit or prevent transcription, promoter mutation reducing mRNA levels, decrease mRNA stability, inactivate donor splice sites (causing read-through into intron), inactivate donor or acceptor splice sites (causing exon to be skipped), activate cryptic splice sites, introduce a frameshift in translation, convert a codon into a stop codon, replace an essential aminoacid, prevent posttranscriptional processing, or prevent correct cellular localization of product. Mutation of a gene is not the only way to abolish its function (e.g. long-range chromatin alterations, abnormal methylation and/or imprinting). For example, in human neoplasms *p16* is silenced in at least three ways: homozygous deletion, methylation of the promoter, and point mutation. The first two represent the majority of inactivation events in most primary cancers. *P16* is a very common early event in cancer progression and is frequently seen in premalignant lesions. The importance of *p16* is probably similar to that of *p53*. Mutations in the *p53* gene have been found in

some 30% of human tumors, and wild-type *p53* has been reported to suppress tumorigenesis and promote apoptosis (236). The p53 protein is a potent transcription factor and may promote transcription of genes that are also involved in carcinogenesis and angiogenesis.

Loss-of-function mutations usually produce recessive phenotypes, so that as long as one allele remains normal there are no significant phenotypic changes. However, for a limited number of genes, a 50% reduction in the dosage of the gene can lead to phenotypic changes (dosage effect). Certain regulatory functions are inherently dosage-sensitive: for example, gene products which compete with each other to determine a developmental or metabolic switch, or that cooperate with each other in interactions with fixed stoichiometry, or those whose function depends on partial or variable occupancy of a receptor or DNA binding site.

Less frequently, mutations can lead to gain-of-function, rather than loss-of-function. For example, they can result in the ability to acquire a new substrate, overexpression of the gene product, the receptor being turned permanently 'on', the ion channel being inappropriately open, structurally abnormal multimers, chimeric gene, the ability to bind to new DNA sequences, or the ability to trap and inactivate important regulatory molecules. If a protein has several catalytic and allosteric domains (e.g. at a regulatory network bottleneck), destruction or loss-of-function of only one of these domains can allow others to be inappropriately activated.

It is, at least theoretically, possible that some carcinogenic events may include both loss of the natural function of the gene product, and gain of a function not normally associated with that particular gene product. For example, a truncated protein might be unable to perform the original function of the native protein, but could still interact functionally with other regulatory proteins by exposing the remaining protein domains.

Proteasomes The average human cell contains about 30 000 proteasomes, each of which contains several protein-digesting proteases. These complexes help regulate a variety of important cellular functions besides the cell cycle, such as the response to viral infection, the so-called *"stress responses,"* abnormal protein catabolism, neural and muscular degeneration, antigen processing, DNA repair, and even cellular differentiation. In fact, at least in principle, diseases can develop if the system is either overactive or underachieving. This research field is rapidly evolving, and a number of proteasome inhibitors are now under development. In fact, at least one of them (Bortezomib) has already been approved by the FDA for the clinical treatment of refractory multiple myelomas.

When the cell needs to destroy a protein, it usually marks it with a chain of small polypeptides called *ubiquitin*. The ubiquitin–proteasome pathway degrades 90% of all abnormal, misfolded proteins, and all of the short-lived regulatory proteins in the cell. These short-lived proteins, whose half-lives are less than three hours, account for 10 to 20% of all cellular proteins. As a result, proteasome inhibitors do not target individual cellular functions, but they usually affect a broad spectrum of functions.

The research area that has evolved fairly rapidly in recent years is that of the so-called *Ubiquitin-Proteasome Pathways*. This involves the processes of selective intracellular protein degradation. For example, the intracellular levels of cyclins, proteins

required for cell cycling, are strictly controlled, and seem to be crucial for the normal functioning of several regulatory cell cycle checkpoints. Indeed, the activity of cyclins appears to be mainly regulated by cyclin levels, rather than by intrinsic molecular changes. Cyclin-dependent Kinase Inhibitor(CKI) protein levels are also strictly controlled throughout the cell cycle, and their degradation is also dependent on the ubiquitin-proteasome pathways.

At least theoretically, selective degradation of cyclins (that often act as oncogenes) but selective inhibition of degradation of CKIs (like *p16* or *p27*, that act as tumor suppressor genes) would provide cancer clinicians with new and powerful tools to fight cancer cells.

The overall regulatory process is rather complex. For example, in normal proliferating cells, the *CKI p27* is actively degraded. The ubiquitin-conjugating enzyme Cdc34, which is a ubiquitin ligase (formed of at least four subunits: the F-box protein Skp2, Skp1, Cul1 and Roc1), and Cks1 are required for the transfer of ubiquitin to phosphorylated *p27*. In a self-amplification process, CDKs phosphorylate *p27* on Thr 187, allowing *p27* recognition by the Skp2 subunit of the ubiquitin ligase.

Stromal cells and cancer Stromal cells are all those cells in a tumor that are not cancer cells. They form the tumor microenvironment and, unlike tumor cells (which are usually monoclonal), they are polyclonal, and contain a mixture of mesenchymal cells (fibroblasts, for example), inflammatory cells (macrophages), immune cells (lymphocytes), antigen-presenting cells (dendritic cells), and other circumstantial cell types. The tumor microenvironment is extremely important for cancer growth, and often it represents the end-result of a perverse gradual mechanism of invasion and destruction by the mutated cancer cells (the ''evil terrorists'') of the tissues of the host (or ''victim'').

In other words, cancer cells often direct themselves the building of the stromal cell tissue microenvironment, in order to allow an even better growth advantage over normal cells, to evade any possible immune resistance, and to continue their invasion of normal structures.

One illustrative example of the sophisticated mechanisms that can lead to normal tissue destruction by cancer cells is the formation of osteolytic lesions in multiple myeloma. Multiple myelomas are a class of cancer cells derived from precursors of B-cell lymphocytes. They are usually diagnosed by detecting abnormally high-monoclonal serum levels of an immunoglobulin (heavy or light chains, or both, of any type), associated with an increase in plasma cells within the bone marrow, of different degrees of morphological differentiation. They can be localized to one site, but more often they are already disseminated in multiple sites at the time of clinical presentation. Although they are usually more frequent in old age, they can also affect young people. The associated symptoms are difficult to control clinically, because multiple myelomas can lead to multiple osteolytic lesions (''holes'' in the bones), generalized loss of bone density, pathological bone fractures, and insufficient hemopoiesis (blood production). These conditions produce a tendency to develop infections because of immune suppression (reduced levels of protective immunoglobulins and normal white blood cells), and bleeding problems (because of the reduced levels of platelets). Myelomas are usually treatable (but seldom curable), by a combination of chemotherapy, radiotherapy, and intravenous biphosphonates,

that are potent inhibitors of osteoclastic activation.

Some recent work has suggested that myeloma cancer cells modify the normal and dynamic "yin and yang," of bone formation (e.g. osteoblast function) and destruction (osteoclast function). The cancer cells achieve this by, among other mechanisms, secreting RANKL (receptor activator of nuclear factor-kB ligand), a potent stimulator of osteoclast function, and by increasing the degradation of 'Osteoprotegerin', the natural soluble antagonist of RANKL.

Furthermore, Tian and colleagues, by comparing different patterns of gene expression in 57 of approximately 10 000 genes in two subsets of myeloma patients (those with focal osteolytic lesions in bone, and those without), have recently demonstrated that some myeloma cells also secrete a soluble molecule called *DKK1*, that binds to an important class of growth factor "coreceptors" that can affect the Wnt-signaling pathway, resulting in decreased activity of osteoblasts. This might explain the loss of bone density in patients, and why normal osteoblasts in myeloma patients do not fill (with new bone) the "holes" made by myeloma cells themselves, or by their secreted proosteolytic substances.

A potential therapeutic approach to ameliorating the bone-specific morbidity of multiple myeloma is by the administration of a suitable antibody against DKK1. The development of new inhibitors of RANKL would also prove useful. In this way, the molecular approach would help toward the restoration of the normal microenvironment.

Tumor antigens They can be defined as molecules that contain the epitope determinants recognized by Tcells that specifically target tumor cells. For many years the underlying principle of tumor immunologists was that cells of the immune system can normally recognize tumors as distinct from the normal tissues (nonself), or as a danger to the organism, and eliminate them (concept of "immunosurveillance"). Although it has now been clearly shown that many, but probably not all, tumors (for example, melanomas) display "tumor antigens", as discrete molecular entities, various mechanisms have been implicated in the resistance of tumors to immune attack: modulation of NK (natural killer cells) receptors, secretion of inhibitory cytokines, recruitment of accessory cells that may thwart immune function, downregulation of molecules critical for antigen-presentation, costimulation of immune cells, and so on.

All these potential mechanisms, and others, can lead to 'tumor editing', so that it is no longer recognized by the normal immune system as 'foreign' and therefore dangerous.

Tumor suppressor genes Unlike oncogenes, Tumor suppressor genes (formerly "antioncogenes") display the function of suppressing cellular proliferation, and/or maintaining cellular differentiation, facilitating normal cellular adhesion mechanisms, stopping entry into the cell cycle (G1, S-phase, mitosis or G2) to allow DNA-repair mechanisms, maintaining normal cellular shape, cell contact inhibition mechanisms, and so on.

Tumor suppressor genes are either underexpressed in cancer cells, because of genetic deletions or promoter hypermethylation, or have undergone other forms of loss of function genetic or epigenetic changes. For example, abnormal *p53* function can be the result of the acquisition of

point mutations, posttranslation inactivation through binding to other regulatory proteins (such as the so-called *MDM2)*, enhanced degradation (e.g. by the activity of the E6 protein of carcinogenic human papilloma viruses), or other mechanisms like decreased translation of wild-type *p53* by the folate-dependent enzyme thymidylate synthase. In these cases, depending on the cellular context and microenvironment, cancer cells are unable to undergo cell cycle arrest and/or apoptosis in response to the DNA damage produced by cytotoxic chemotherapy or radiotherapy.

Reversion of the malignant phenotype was demonstrated, over three decades ago, by the classical ''cell fusion'' experiments by Henry Harris, at Oxford. The loss of particular chromosomes in hybrid normal-malignant cells, or the gain of normal chromosomes when fusing normal cells with particular types of malignant cells *in vitro*, led to the first experimental data to back the hypothesis of ''antioncogenes,'' or tumor suppressor genes. Later experiments in the 1980s and 1990s, with the help of gene transfer methods including modified retroviruses and adenovirus, have given more support to the idea that in many cases the malignant phenotype of cancer cells can be reversed, *in vitro* and *in vivo*, by replacing a key tumor suppressor gene. However, because of the technical difficulties and the potential dangers of gene therapy in patients, these findings have not yet led to clinically relevant forms of gene therapy.

An alternative approach, based on the binary state of most regulatory molecules, as proposed in ''Principles of Molecular Oncology'' (2004), is to develop drugs that can lock *in vitro* and '*in vivo*' a particular tumor suppressor protein into its active ''tumor suppressor function'' (Fig. 3).

The tumor suppressor paradigm characteristically calls for loss of both functional copies of the gene and indeed many, though not all, tumor suppressor genes must undergo biallelic inactivation in order to sustain a true loss-of-function effect.

An example is the PTEN tumor suppressor pathway.

In 1997, PTEN (phosphatase and tensin homolog deleted on chromosome10), was cloned and mapped to cytoband 10q23, a region of chromosome 10 that undergoes frequent somatic deletion in some malignant human tumors, in particular, in endometrial cancers (almost 50%), some brain tumors (e.g. 30–40% of glioblastoma multiforme, and to a lesser extent anaplastic astrocytomas), metastatic prostate cancers (20–60%) and malignant melanomas.

Discordance between the rate of ''loss of heterozygosity'' (LOH) and the rate of mutation of the second allele of the *PTEN* gene, has led some to suggest that a second tumor suppressor gene is located in the 10q23 region, but this difference could also due to other reasons: (1) technical inability to detect second mutational events (low sensitivity); (2) a gene dosage-effect where loss of one allele of *PTEN* may have a partial tumor promoting effect; c) epigenetic alterations in the *PTEN* gene, the mRNA, or corresponding protein; d) cooperation between loss of only one allele of *PTEN* and a genetic event in another not-yet-identified gene. The PTEN protein is also regulated by phosphorylation. In particular the serine/threonine kinase CK2, upregulated in many cancers, phosphorylates the PTEN C-terminal tail and reduces PTEN activity, raising the possibility that phosphorylation of PTEN by CK2 might be a mechanism contributing to the downregulation of PTEN in certain human cancers retaining a wild-type PTEN allele. In consequence, inhibitors of CK2

might prove useful therapeutic agents in some cancers.

Whatever be the reasons, loss of PTEN tumor suppressor protein leads to constitutive activation of the PIP3 (phosphatidylinositol-3,4,5-trisphosphate) pathway, with important downstream consequences, *via* the molecular participation of Oncogenic Protein Pathways (like Akt pathways and the m-TOR protein that derives its name from "mammalian target of rapamycin"). This process results in significant changes in cell cycle progression, survival, cell spreading and motility, and even in the angiogenic properties of tumor cells.

This PTEN/PI3K molecular system provides an example of an elegant "on-off switch" that has been evolved as "nutrient and growth factor response pathways"where the switch moves to "on" position when PI3 kinase (PI3K) deposits a phosphate group on the D3 position of the inositol ring, while it is turned "off" when the PTEN removes the phosphate group from the same position.

Another important example of a TS that can prove to be a valid target for cancer therapy is *p53*. Underactivity of *p53*, caused by loss-of-function mutations, encourages the growth of cancer, whereas its overactivity can accelerate the aging process and cell death. Following some elegant pioneering work by Snyder and colleagues (recently reviewed by David Lane in the New England Journal of Medicine), a synthetic 34-amino-acid peptide, made of *d*-amino acids, rather than the conventional *l*-amino acids, and called *RI-TATp53C*, has been shown to be a potent activator of DNA-binding and tumor suppressor functions in both wildtype and mutant *p53*. In preclinical models, this synthetic peptide, that cannot be degraded by proteases and is internalized into cancer cells by a lipid-raft-dependent pinocytosis mechanism, has been shown to achieve complete cures in a mouse model of terminal peritoneal lymphoma.

New cancer therapies will be directed not merely to hit the DNA synthetic machinery, but to hit selective targets at or downstream from abnormally activated catalytic functions (thereby abrogating the abnormal stimulatory signals), or at or upstream from inactivated or missing catalytic functions (switching on parallel or alternative regulatory pathways).

Multitargeting The "oncogene addiction hypothesis", first put forward by Weinstein, emphasizes the importance of oncogenes or tumor suppressor genes in the maintenance and initial development of any given cancer. In other words, some cancers are extremely dependent on only one or a few oncogenic changes, and may be inhibited by one or a few target-oriented drugs. The correction of a key oncogenetic lesion in a cancer that has undergone a multitude of other activating oncogenic events may nevertheless provide anticancer effects. Although the main clinical examples of such 'Achilles heels' of cancers are provided by relatively unusual tumors like chronic myeloid leukemia, and c-KIT mutants like GIST-tumor (both very sensitive to the therapeutic effects of Imatinib), the situation in most common solid tumors is probably more complex. For example, even in breast cancers that express, by IHC, huge numbers of HER-2 targets (tumors considered to be candidates for Herceptin must express positivity of HER-2 in over 80% of cells, or positivity for FISH with a chromosome amplified ratio of 17 or more), the use of Trastuzumab may lead to response rates of only around 30% when used alone,

and 50% when combined with conventional cytotoxic chemotherapy. The use of new HSP90 inhibitors can tackle several carcinogenic targets at once, the so-called *hallmark traits of cancer*, because HSP90 is a molecular chaperone that stabilizes several mutated and overexpressed ''client'' signaling molecules such as many oncogenic kinases or even the catalytic component of telomerase. Inhibition of HSP90 leads to incorrect folding and subsequent degradation of oncogenic client proteins through the ubiquitin-proteasome pathways, but this ''promiscuity'' of action of HSP90 inhibitors can also lead to less specificity, lower anticancer potency and more potential iatrogenic side effects.

Other multitarget agents currently undergoing clinical development include: (1) SU11248, an oral multitargeted kinase inhibitor with direct antitumor as well as antiangiogenic activity, *via* the targeting of the VEGF, platelet-derived growth factor, KIT, and FLT3 receptor tyrosine kinases; (2) PTK87/ZK 222584, an potent orally administered inhibitor of the VEGF-mediated Flt-1 and Kinase Domain-containing Receptor(KDR) family of tyrosine kinases; (3) BAY 43–9006, again an orally given drug that inhibits Raf-1, but also other membrane receptors and molecular targets; (4) CCI-779, a novel mammalian target of rapamycin (m-TOR) kinase inhibitor, that may be useful in tumors with *PTEN* mutations, or alterations in the Akt activity pathways, regulated by PI3-kinase and *PTEN* suppressor genes.

One obvious regulatory pathway with direct relevance to cancer therapeutics is Cell Cycle Control.

From an almost 'military' point of view, the 'heart of the matter' of tumor cell growth and survival is more likely to be in the nucleus of cancer cells (downstream), than on the cell membranes (upstream).

The role of cyclin D1 in circumventing the cell-cycle checkpoints has been known to be crucial for the pathogenesis of some nonHodgkins Lymphomas and also in colon cancer, and its relationship with intracellular catenin levels have recently come to light. In most cancers, a defect in the retinoblastoma checkpoint is what allows cells to divide uncontrollably and excessively. But usually, in colon cancer, this checkpoint is surprisingly free of defect, and for a long time it was unclear how colon cancer cells achieved a proliferative advantage over their normal counterparts. Then several researchers started to link the intracellular catenin levels with the tumor suppressor protein known as *adenomatous polyposis coli* (*APC*).

The search was on for the target of *beta*-catenin. Osamu Tetsu and Frank McCormick at the University of California in San Francisco published a key paper in 1999, linking intracellular catenin directly with increased expression of cyclin D1, thereby prompting cells to enter the S-phase.

High levels of cyclin D1 override the checkpoint in colonic epithelial cells, generating a population of cells that will not be totally transformed, but might acquire further mutations, for example in *Ras* or other oncogenes, to acquire the truly malignant phenotype. The two pathways, Ras and beta catenin-APC, somehow converge into the same apparent bottleneck activating the expression of cyclin D1.

Similarly, cyclin E has a pivotal role in transducing the mitogenic stimulus of several hormones, cytokines, and extracellular growth factors. Transient expression of cyclin E induces the formation and activation of CDK2 complexes and, as explained before, pRb is again a key substrate. It seems that pRb must be phosphorylated

by CDK4/cyclin D before being phosphorylated by CDK2/cyclin E.

Recent evidence suggests that levels of total cyclin E (and low-molecular-weight cyclin E) in tumor tissue, as measured by Western blot assays, correlate strongly with survival in patients with breast cancer.

In a retrospective analysis of 395 patients with breast cancer with a median follow-up of 6.4 years, high levels of truncated or total cyclin E correlate with poor disease-specific survival. Among patients with early (stage I) breast cancer, all those with low levels of cyclin E were alive at seven years, as compared with none of those with expression of high levels of cyclin E. If these findings are confirmed by others, then determination of cyclin E levels by an immunohistochemical assay might soon become routine in oncology units. Stimulation of the proliferation of breast cancer cells by estrogens and growth factors is also accompanied by increased cyclin E levels and the formation of active cyclin E–cdk2 complexes.

From the above arguments, it is obvious that selective inhibitors of cyclin E and D are good candidates for cancer therapy. Molecules designed to inhibit the binding of these cyclins to their corresponding pockets in CDKs (particularly CDK2 and CDK4) are being designed and developed, and some might soon be under preclinical development. Structural studies, for example, have focused on how RXL (cyclin-binding motif) and LXCXE (part of cyclin structure) contribute to substrate selection. The RXL motif in *p27* is thought to bind to a hydrophobic surface on cyclin A. This surface is conserved between cyclins A,B,D and E (lying opposite to the CDK2-binding site).

Cyclin-dependent Kinase Inhibitors (CKIs) can provide us with basic clues on how proliferation is normally controlled *in vivo*, on new classes of clinically relevant prognostic factors in cancer, and on how to develop molecular mimetics with high affinity for CDKs, and similarly potent inhibitory actions. Synthetic d-peptides (as opposed to naturally occurring l-peptides) are also being investigated in this context as potential allosteric pharmacological regulators of CKIs.

Molecular dynamics simulations can provide complementary information on the nature of these interactions. Owing to the large contact surface between both proteins, it is difficult to design a single small-molecular-weight drug capable of embracing all the interactions deduced from the crystal structures and the molecular dynamic simulations. However, increasing insight into the design of small molecules with sufficient selectivity to alter the T-loop conformation and distortion of the ATP binding site is gradually been achieved. Moreover, the development of short-chain synthetic peptides capable of inhibiting normal binding of p16 to CDK4 can allow a better understanding of the key structure, activity and allosteric properties of these complex molecular interactions.

Gene transfer experiments, for example transfection into tumor cells in culture of active CKI genes, has enabled several laboratories to test the principle of the induction of tumor regression by the activation, or transfer, of CKI activity A good illustration of this is that ectopic p18 expression inhibits growth and induces apoptosis of multiple myeloma (MM) cells. In about a quarter of human MM cell lines, cyclin D1 is overexpressed because the *cyclin D1* gene is brought under the influence of immunoglobulin heavy chain (IgH) enhancer and promoter elements by chromosomal translocation t(11;14)(q13;q32). The CKI p18 predominantly binds to CDK6 and, to a lesser extent, to

CDK4 and inactivates their kinase activity. Furthermore, p18 can efficiently block CDK6 phosphorylation by CAK/cyclin H1. Cell lines transfected with an inducible p18 expression vector not only exhibited growth inhibition, but also apoptosis.

Microinjection of CKI proteins has also confirmed other biological activities of CKIs. For example, injection of the CKI p21 and p27 in *Xenopus laevis* embryos blocked centrosome duplication. The centrosome nucleates the polymerization of microtubules and duplicates once per cell cycle starting at the G1-S transition, and plays a major role in organizing the poles of the mitotic spindle. The inhibition of normal centrosome function by p21 and p27 can also arrest cell growth, and lead to abnormal segregation of chromosomes.

Another unexpected property of CKI is the reported ability of p21 (WAF1/CIP1) to restore hormone responsiveness to estrogen-receptor (ER)-negative breast cancer cells. Again, the overexpression of p21 with a tetracycline-inducible gene transfer vector can make ER-negative, p21-negative breast cancer cells sensitive to growth inhibition by anti-estrogens. A strong positive association has been found between the expression of p21 and the presence of ER in tumor samples from 60 patients with breast cancer.

Besides gene therapy approaches, *in vivo* protein and peptide transduction systems are also being investigated by several laboratories. Recombinant proteins, like recombinant human colony-stimulating factors, have been used in cancer patients for some 15 years already. However, the inability of molecules larger than 600 daltons to cross the plasma membrane has restricted the pharmacological use of proteins to those which function outside the cell, for example, by binding to specific cell membrane receptors. Most tumor suppressors, on the other hand, are intracellular, and therefore cannot be administered to patients as recombinant proteins directly.

At first sight, the retinoblastoma (Rb) protein is the true bottleneck of most of these cell cycle–controlling pathways. In 1986-87 the tumor suppressor *Rb* gene, first identified because it mutated in hereditary retinoblastomas, was cloned by three different laboratories headed by eminent molecular biologists: Robert A. Weinberg and Thaddeus Dryja, William Benedict and Yuen-Kai Fung, and Wen-Hwa Lee. It is a rather large gene, over 200-kb long, and with 27 exons. It codifies for a nuclear protein of 928 amino acids, which is constitutively expressed during the cell cycle, but with characteristic cell cycle dependent varying degrees of phosphorylation.

Phosphorylation of pRb is carried out initially by D-type cyclins (in protein complexes with CDK4 or CDK6) followed later by cyclin E/CDK2. Virtually all human tumors tested so far show mutations that directly or indirectly alter the normal function of *Rb*. Nearly all tumor-derived *pRb* mutants have lost the ability to repress E2F-responsive genes, and reintroduction by gene transfer of wild-type pRb into Rb −/− tumor cells leads to restoration of E2F control and cell cycle arrest. There are at least six human *E2F* genes (*E2F*1 to *E2F*6) and their protein products bind to specific DNA sequences as heterodimers with either DP1 or DP2 proteins. It is now clear that binding to pRb converts the E2F family from transcriptional activators to potent transcriptional repressors. E2F6, unlike the other E2F family members, is an intrinsic transcriptional repressor and does not apparently interact with pRb family members.

Hyperphosphorylated pRb is unable to maintain the binding and inactivation of E2F, and free E2F molecules lead to transcription of multiple genes involved in the initiation of DNA synthesis and cellular proliferation.

Using gene expression profiling methods (Affymetrix GeneChips) it has been possible to demonstrate changes in the expression of over 200 genes, most of them involved in cell cycle control, by CDK phosphorylation of wild-type *pRb*. Some of these genes are also involved with DNA repair, or changes in chromatin structure. As expected, a significant fraction of Rb-repressed genes have promoters that are bound/regulated by E2F family members. However, targets were also identified that are distinct from genes known to be stimulated by overexpression of specific E2F proteins, suggesting that some of the multiple pRb effects are not directly related to E2F proteins.

The conformation and activity of pRb is rather complex, and not fully understood. It is widely accepted that the activity of this large tumor suppressor protein is largely dependent upon the phosphorylation status of at least 16 potential CDK phosphorylation sites. But how exactly the phosphorylation of any of these sites changes the full range of activities of pRb remains a matter of intense study. Low-molecular-weight drugs designed to block 'site-specific phosphorylation' of these sites are also under study, as they may ultimately arrest the growth of populations of tumor cells.

Yuen-Kai Fung's laboratory at the University of Southern California has identified all the CDK sites (Thr-356, Ser-807/Ser-811, and Thr-821), the phosphorylation of which drastically modifies the conformation of pRb. The so-called *m89 structural motif* (identified in the m89 mutant of pRb) has greatly enhanced growth suppressing activity, similar to a mutant with alanine substitutions at Ser-807/Ser-811. Moreover, this m89 region is part of a structural domain, p5, conserved antigenically and functionally between pRb and p53.

Rationally designed drugs capable of interacting with these key molecular sites may exploit the coordinated regulation of activity of these two tumor suppressors, or at least block the conformation of the tumor suppressor pRb into its active growth inhibitory hypophosphorylated structure (see also Fig. 3).

One of the seemingly tragic, and yet-unexplained, features of our human genome is that the key sensors that control many of the functional interactions between Rb and p53 are actually linked in the same gene, making them dually vulnerable to the same genetic attack. In general, mutational events that disable the Rb pathway and facilitate cell proliferation are counterbalanced by a p53-dependent response that eliminates, or at least inhibits, incipient cancer cells. Conversely, loss of p53 function enables cells that sustain oncogenic damage to survive and proliferate. Unfortunately for us, as explained before, the INK4A-ARF locus encodes, in the same genetic locus, two distinct tumor suppressor proteins that regulate both the Rb and the p53 pathways. ARF (p19) is a sensor of inappropriate proliferation brought about by loss of Rb. The product of the *p53* gene is part of the intrinsic mechanisms that monitor the cell cycle. For example, inducing the expression of p21, and mutations, deletions or underexpression of p19ARF facilitate MDM2's degradation of p53, leaving the cell unaware of the need to initiate cell death.

To complete the picture, reactivation of telomerase activity, normally suppressed in most normal human tissues, can dramatically contribute to the carcinogenesis process. Telomerase activity is needed to maintain chromosomal stability, particularly at the – tips of chromosomes where bits of DNA are otherwise lost during each round of DNA replication and cell division, because of the way DNA polymerase works and the need for "primers." If it were not for the activity of telomerase, many cancer types would eventually self destroy, because of the cumulative loss of DNA material at each round of cell division.

See also Cancer Chemotherapy, Theoretical Foundations of; Cancer of the Prostate: Molecular Genetics; Epigenetic Mechanisms in Tumorigenesis; Genetics and Molecular Biology of Lung Cancer; Intracellular Signaling in Cancer; Liver Cancer, Molecular Biology of; Mutagenesis, Malignancy and Genome Instability; Oncogenes.

Bibliography

Books and Reviews

Aiello, A., Tamborini, E., Frattini, M. et al. (2004) Genetic Markers in Sporadic Tumors, in: Bronchud, M.H., Foote, M.A., Giaccone, G., Olopade, O. and Workman, P. (Eds.) *Principles of Molecular Oncology*, 2nd edition, Humana Press, Totowa, NJ, pp. 73–150.

Baird, R., Workman, P. (2004) Emerging Molecular Therapies: Drugs Interfering with Signal Transduction Pathways, in: Bronchud, M.H., Foote, M.A., Giaccone, G., Olopade, O., Workman, P. (Eds.) *Principles of Molecular Oncology*, 2nd edition, Humana Press, Totowa, NJ, pp. 569–606.

Bronchud, M.H. (2004) Selecting the Right Targets for Cancer Chemotherapy, in: Bronchud, M.H., Foote, M.A., Giaccone, G., Olopade, O., Workman, P. (Eds.) *Principles of Molecular Oncology*, 2nd edition, Humana Press, Totowa, NJ, pp. 3–49.

Bronchud, M.H., Brizuela, L., Gyuris, J., Mansuri, M.M. (2004) Cyclin-dependent Kinases and their Regulators as Potential Targets for Anticancer Therapeutics, in: Bronchud, M.H., Foote, M.A., Giaccone, G., Olopade, O., Workman, P. (Eds.) *Principles of Molecular Oncology*, 2nd edition, Humana Press, Totowa, NJ, pp. 359–410.

Castillo, L., Etienne-Grimaldi, M.C., Fischel, J.L. et al. (2004) Pharmacological background of EGFR targeting, *Ann. Oncol.* **15**, 1007–1012.

Chabner, B.A., Boral, A.L., Multani, P. (1998) Translational research: walking the bridge between idea and cure – Seventeeth Bruce F. Cain Memorial Award lecture, *Cancer Res.* **58**, 4211–4216.

Chu, E., DeVita Jr., V.T. (2002) *Cancer Chemotherapy Drug Manual*, Jones and Bartlett Publishers, Boston, MA.

Conacci-Sorrel, M., Zhurinsky, J., Ben-Ze'ev, A. (2002) The Cadherin-Catenin adhesion system in signaling and cancer, *J. Clin. Invest.* **109**, 987–991.

Davies, A.J., Rohatiner, A.Z.S., Howell, S. et al. (2004) Tositumomab and Iodine I 131 tositumomab for recurrent indolent and transformed B-cell Non Hodgkin's Lymphoma, *J. Clin. Oncol.* **22**(8), 1469–1479.

Hahn, W.C., Weinberg, R.A. (2002) Modelling the molecular circuitry of cancer, *Nat. Rev. Cancer* **2**, 331–341.

Hahn, W.C., Weinberg, R.A. (2002) Rules for making human tumor cells, *N. Engl. J. Med.* **347**, 1593–1603.

Lane, D. (1998) The promise of molecular oncology, *Lancet* **351**(suppl II), SII17–SII20.

Lane, D. (2004) Curing cancer with p53, *N. Engl. J. Med.* **350**, 2711–2712.

Oved, S., Yarden, Y. (2002) Signal transduction: molecular ticket to enter cells, *Nature* **416**, 133–136.

Rich, T.A., Shepard, R.C., Mosley, S.T. (2004) Four decades of continuing innovation with Fluorouracil: current and future approaches to fluorouracil chemoradiation therapy, *J. Clin. Oncol.* **22**(11), 2214–2232.

Stratton, M.S., Stratton, S.P., Dionne, S.O. et al. (2004) Treatment of Carcinogenesis, in: Bronchud, M.H., Foote, M.A., Giaccone, G., Olopade, O., Workman, P. (Eds.) *Principles of Molecular Oncology*, 2nd edition, Humana Press, Totowa, NJ, 607–673.

Weinstein, I.B. (2002) Cancer: addiction to oncogenes- the Achilles heal of cancer, *Science* **297**, 63–64.

Primary Literature

Adams, J. (2002) Proteasome inhibition: a novel approach to cancer therapy, *Trends Mol. Med.* **8**, S49–S54.

Armitage, P., Doll, R. (1954) The age-distribution of cancer and a multistage theory of carcinogenesis, *Br. J. Cancer* **8**, 1–12.

Armitage, P., Doll, R. (1957) A two-stage theory of carcinogenesis in relation to the age distribution of human cancer, *Br. J. Cancer* **11**, 161–169.

Baltimore, D. (1976) Viruses, polymerases and cancer, *Science* **192**, 632–636.

Banerjee, D., Bertino, J.R. (2002) E2F and Cancer Chemotherapy, in: La Thangue, N.B., Bandara, L.R. (Eds.) *Targets for Cancer Chemotherapy: Transcription Factors and other Nuclear Proteins*, Humana Press, Totowa, NJ, 289–298.

Barrie, E., Eno-Amooquaye, E., Hardcastle, A. et al. (2003) High-throughput screening for the identification of small-molecule inhibitors of retinoblastoma protein phosphorylation in cells, *Anal. Biochem.* **320**, 66–74.

Basehoar, A.D. et al. (2004) Identification and distinct regulation of yeast TATA box-containing genes, *Cell* **116**, 699–709.

Bast, R.C., Ravdin, P., Hayes, D.F. et al. (2001) 2000 Update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology, *J. Clin. Oncol.* **19**, 1865–1878.

Bates, S., Vousden, K.H. (1996) P53 in signaling checkpoint arrest or apoptosis, *Curr. Opin. Genet. Dev.* **6**, 12–19.

Beer, M.A., Tavazoie, S. (2004) Predicting gene expression from sequence, *Cell* **117**, 185–198.

Bronchud, M.H., Scarffe, J.H., Thatcher, N. et al. (1987) Phase I/II study of recombinant human granulocyte colony-stimulating factor in patients receiving intensive chemotherapy for small cell lung cancer, *Br. J. Cancer* **56**, 809–813.

Brown, R., Strathdee, G. (2002) Epigenomics and epigenetic therapy of cancer, *Trends Mol. Med.* **8**, S43–S48.

Bykov, V.J.N., Issaeva, N., Shilov, A. et al. (2002) Restoration of tumor suppressor function to mutant p53 by a low molecular weight compound, *Nat. Med.* **8**, 282–288.

Carter, P. (2001) Improving the efficacy pf antibody-based cancer therapies, *Nat. Rev. Cancer* **1**, 118–129.

Chin, L., Pomerantz, J., Polsky, D. et al. (1997) Cooperative effects of INK4A and ras in melanoma susceptibility in vivo, *Genes Dev.* **11**, 2822–2834.

Coiffier, B., Lepage, E., Briere, J. et al. (2002) CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma, *N. Engl. J. Med.* **346**, 235–242.

Couch, D.B. (1996) Carcinogenesis: basic principles, *Drug Chem. Toxicol.* **19**, 133–148.

Dillman, R.O. (1997) Magic bullets at last! Finally – Approval of a monoclonal antibody for the treatment of cancer!!!, *Cancer Biother. Radiopharm.* **12**, 223–225.

Doll, R. (1996) Nature and nurture:possibilities for cancer control, *Carcinogenesis (Lond.)* **17**, 177–184.

Druker, B.J., Tamura, S., Buchdunger, E. et al. (1996) Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells, *Nat. Med.* **2**, 561–566.

Dy, G.K., Adjei, A.A. (2002) Novel targets for lung cancer therapy: Part I, *J. Clin. Oncol.* **20**, 2881–2894.

Dy, G.K., Adjei, A.A. (2002) Novel targets for lung cancer therapy: Part II, *J. Clin. Oncol.* **20**, 3016–3028.

Early BreastCancer Trialists' Collaborative Group (1998) Tamoxifen for early breast cancer:an overview of the randomised trials, *Lancet* **351**, 1451–1467.

El-Deiry, W.S. (1998) Regulation of p53 downstream genes, *Semin. Cancer Biol.* **8**, 345–357.

Elsayed, Y.A., Sausville, E.A. (2001) Selected novel anticancer treatments targeting cell signaling proteins, *The Oncologist* **6**, 517–537.

Fearon, E.R., Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis, *Cell* **61**, 759–767.

Fisher, J.C. (1958) Multiple-mutation theory of carcinogenesis, *Nature* **181**, 651–652.

Fisher, J.C., Hollomon, J.H. (1951) A hypothesis for the origin of cancer foci, *Cancer* **4**, 916–918.

Fox, E., Curt, G.A., Balis, F.M. (2002) Clinical trial design for target-based therapy, *Oncologist* **7**, 401–409.

Friedewald, W.F., Rous, P. (1944) The initiating and promoting elements in tumor production, *J. Exp. Med.* **80**, 101–126.

Gelmann, E.P. (2002) Molecular biology of the androgen receptor, *J. Clin. Oncol.* **20**, 3001–3015.

Goldie, J.H., Coldman, A.J. (1979) A mathematic model for relating the drug sensitivity of tumors to the spontaneous mutation rate, *Cancer Treat. Rep.* **63**, 1727–1733.

Goldman, J.M. (2000) Tyrosine-kinase inhibition in treatment of chronic myeloid leukaemia, *The Lancet* **355**, 1031–1032.

Goss, P.E., Ingle, J.N., Martino, S. et al. (2003) A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer, *N. Engl. J. Med.* **349**, 1793–1802.

Hancock, J.T. (1997) *Cell Signalling*, Addison Wesley Longman Limited, Harlow, England, UK.

Harbour, J., Dean, D. (2000) The Rb/E2F pathway: expanding roles and emerging paradigms, *Genes Dev.* **14**, 2393–2409.

Harris, C.C. (1996) The Walter Hubert Lecture. Molecular epidemiology of human cancer: insights from the mutational analysis of the p53 tumor suppressor gene, *Br. J. Cancer* **73**, 261–269.

Hernández-Bronchud, M. (1995) Growth factors and cancer, *Br. J. Hosp. Med.* **53**, 20–26.

Hortobagyi, G.N. (2004) What is the role of high-dose chemotherapy in the era of targeted therapies? *J. Clin. Oncol.* **22**(12), 2263–2266.

Huisinga, K.L., Pugh, B.F. (2004) A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae, *Mol. Cell.* **13**, 573–585.

Huron, D., Mizenina, O., Rosen, N., Moasser, M. (2002) A second generation inhibitor of Abl tyrosine kinase with picomolar biologic potency and selective activity against CML cells, *American Association for Cancer Research, 93rd Annual Meeting*, San Francisco CA: abstract 4198.

Innocenti, F., Undevia, S.D., Iyer, L. et al. (2004) Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan, *J. Clin. Oncol.* **22**(8), 1382–1388.

Jain, K.K. (2000) Applications of proteomics in oncology, *Pharmacogenomics* **1**, 385–393.

Johnstone, R.W. (2002) Histone-deacetylase inhibitors: novel drugs for the treatment of cancer, *Nat. Rev. Drug Discov.* **1**, 287–299.

Jonasson, J., Harris, H. (1977) The analysis of malignancy by cell fusion. (8) Evidence for the intervention of an extra-chromosomal element, *J. Cell Sci.* **24**, 255–263.

Kaelin, W.G. Jr. (2002) Cancer Chemotherapy Based on E2F and the Retinoblastoma Pathway, in: La Thangue, N.B., Bandara, L.R. (Eds.) *Targets for Cancer Chemotherapy: Transcription Factors and other Nuclear Proteins*, Humana Press, Totowa, NJ, 1–13.

Karin, M., Hunter, T. (1995) Transcriptional control by protein phosphorylation: signal transmission from the cell surface to the nucleus, *Curr. Biol.* **5**, 747–757.

Kim, H., Kwon, Y.M., Kim, J.S. et al. (2004) Tumor-specific methylation in bronchial lavage for the early detection of non-small-cell lung cancer, *J. Clin. Oncol.* **22**(12), 2363–2370.

Kinzler, K.W., Vogelstein, B. (1996) Lessons from hereditary colorectal cancer, *Cell* **87**, 159–170.

Kinzler, K.W., Vogelstein, B. (1997) Gatekeepers and caretakers, *Nature (London)* **386**, 761–763.

Knudson, A.G. (1973) Mutation and human cancer, *Adv. Cancer Res.* **17**, 317–352.

Kodadek, T. (2001) Protein microarrays: Prospects and problems, *Chem. Biol.* **8**, 105–115.

Li, S. et al. (2004) A map of the interactome network of the metazoan C *elegans*, *Science* **303**, 540–543.

Liggett, W.H., Sidransky, D. (1998) Role of the p16 tumor suppressor gene in cancer, *J. Clin. Oncol.* **16**, 1197–1206.

Lindor, N.M., Greene, M.H. (1998) The concise handbook of family cancer syndromes. Mayo Familial Cancer Program, *J. Natl. Cancer Inst.* **90**, 1039–1071.

Lindor, N.M., Burgart, L.J., Leontovich, O. et al. (2002) Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors, *J. Clin. Oncol.* **20**, 1043–1048.

Loeb, L. (1991) A Mutator phenotype may be required for multistage carcinogenesis, *Cancer Res.* **51**, 3075–3080.

Loeb, L.A. (1998) Cancer cells exhibit a mutator phenotype, *Adv. Cancer Res.* **72**, 25–56.

Lord, B.I., Bronchud, M.H., Owens, S. et al. (1989) The kinetics of human granulopoiesis following treatment with granulocyte colony-stimulating factor "in vivo", *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9499–9503.

Maloney, A., Workman, P. (2002) HSP90 as a new therapeutic target for cancer therapy: the story unfolds, *Expert Opin. Biol. Ther.* **2**, 3–24.

Mao, L., Hurban, R.H., Boyle, J.O. et al. (1994) Detection of oncogene mutations in sputum precedes diagnosis of lung caner, *Cancer Res.* **54**, 1634–1637.

Massagué, J., Polyak, K. (1995) Mammalian antiproliferative signals and their targets, *Curr. Opin. Genet. Dev.* **5**, 91–96.

Melnick, A., Licht, J.D. (1999) Deconstructing a disease: RARalpha, its fusion partners, and their roles in the pathogenesis of acute promyelocytic leukemia, *Blood* **93**, 3167–3215.

Mendelsohn, J. (2002) Targeting the Epidermal Growth Factor Receptor for cancer therapy, *J. Clin. Oncol.* **20**(Sept 15 Suppl.), 1s–13s.

Monks, A., Scudiero, D., Skehan, P. et al. (1991) Feasibility of a high-flux anticancer drug screen using a diverse panel of cultured human tumor cell lines, *J. Natl. Cancer Inst.* **83**, 757–766.

McCormick, F. (2001) Cancer gene therapy: fringe or cutting edge? *Nat. Rev. Cancer* **1**, 130–141.

Nordling, C.O. (1953) A new theory on the cancer-inducing mechanism, *Br. J. Cancer* **7**, 68–72.

Odom, D.T. et al. (2004) Control of pancreas and liver gene expression by HNF transcription factors, *Science* **303**, 1378–1381.

Opalinska, J.B., Gewirtz, A.M. (2002) Nucleic-acid therapeutics: basic principles and recent applications, *Nat. Rev. Drug Discov.* **1**, 503–514.

Palmer, H.G. (2001) Vitamin D promoted the differentiation of colon carcinoma cells by the induction of E-cadherin and the inhibition of beta-catenin signaling, *J. Cell Biol.* **154**, 369–387.

Paull, K.D., Shoemaker, R.H., Hodes, L. et al. (1989) Display and analysis of patterns of differential activity of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm, *J. Natl. Cancer Inst.* **81**, 1088–1092.

Perera, F.P. (1996) Molecular epidemiology: insights into cancer susceptibility, risk assessment, and prevention, *J. Natl. Cancer Inst.* **88**, 496–509.

Ponder, B. (1997) Genetic testing for cancer risk, *Science (Washington DC)* **278**, 1050–1054.

Rosenberg, S.A. (2001) Progress in human tumour immunology and immunotherapy, *Nature* **411**, 380–384.

Rubin, B.P., Schuetze, S.M., Eary, J.F. et al. (2002) Molecular targeting of platelet-derived growth factor B by imatinib mesylate in a patient with metastatic dermatofibrosarcoma protuberans, *J. Clin. Oncol.* **20**, 3586–3591.

Rusch, V., Klimstra, D., Linkov, I. et al. (1995) Aberrant expression of p53 or the epidermal growth factor receptor is frequent in early bronchial neoplasia, and coexpression precedes squamous cell carcinoma development, *Cancer Res.* **55**, 1365–1372.

Sansal, I., Sellers, W.R. (2004) The biology and clinical relevance of the PTEN tumor suppressor pathway, *J. Clin. Oncol.* **22**, 2954–2963.

Senior, K. (1999) Fingerprinting disease with protein chip arrays, *Mol. Med. Today* **5**, 326–327.

Sharp, P.A. (2001) RNA interference-2001, *Genes Dev.* **15**, 485–490.

Sharpless, N., Bardeesy, N., Lee, K.-H. et al. (2001) Loss of p16INK4a with retention of p19Arf predisposes mice to tumorigenesis, *Nature* **413**, 86–91.

Shields, P.G., Harris, C.C. (1991) Molecular epidemiology and the genetics of environmental cancer, *J. Am. Med. Assoc.* **266**, 681–687.

Slamon, D.J., Leyland-Jones, B., Shak, S. et al. (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2, *N. Engl. J. Med.* **344**, 783–792.

Snyder, E.L., Meade, B.R., Saenz, C.C., Dowdy, S.F. (2004) Treatment of terminal peritoneal carcinomatosis by a transducible p53-activating peptide, *PloS Biol.* **2**(2), 186–193.

Sporn, M.B., Roberts, A.B. (1985) Autocrine growth factors and cancer, *Nature* **313**, 745–747.

Sporn, M.B., Todaro, G.H. (1980) Autocrine secretion and malignant transformation of cells, *N. Engl. J. Med.* **303**, 878–880.

Stark, A., Hulka, B.S., Joens, S. et al. (2000) HER-2/neu amplification in benign breast disease and the risk of subsequent breastcancer, *J. Clin. Oncol.* **18**(2), 267–274.

Temin, H.M. (1971) The protovirus hypothesis: speculations on the significance of RNA-directed DNA synthesis for normal development and for carcinogenesis, *J. Natl. Cancer Inst.* **46**, iii–viii.

Tewari, M. et al. (2004) Systematic interactome mapping and genetic perturbation analysis of a C. elegans TGF-ß signaling network, *Mol. Cell.* **13**, 469–482.

Tian, E., Zhan, F., Walker, R., Rasmussen, E. et al. (2003) The role of the Wnt-signaling antagonist DKK1 in the development of osteolytic lesions in multiple myeloma, *N. Engl. J. Med.* **349**, 2483–2494.

Torrance, C.J., Jackson, P.E., Montgomery, E. et al. (2000) Combinatorial chemoprevention of intestinal neoplasia, *Nat. Med.* **6**, 1024–1028.

Visakorpi, T., Hyytinen, E., Koivisto, P. et al. (1995) In vivo amplification of the androgen receptor gene and progression of human prostate cancer, *Nat. Genet.* **9**, 401–406.

Walker, J.W., Hayward, N.K. (2002) P16INK4A and p14ARF tumour suppressors in melanoma: lessons from the mouse, *The Lancet* **359**, 7–8.

Welte, K., Gabrilove, J., Bronchud, M.H. et al. (1996) Filgrastim (r-metHuG-CSF): the first ten years, *Blood* **88**, 1907–1929.

Wiener, F., Klein, G., Harris, H. (1971) The analysis of malignancy by cell fusion (3). Hybrids between diploid fibroblasts and other tumour cells, *J. Cell. Sci.* **8**, 681–692.

Woodford-Richens, K.L., Rowan, A.J., Gorman, P. et al. (2001) SMAD4 mutations in colorectal cancer probably occur before chromosomal instability, but after divergence of the microsatellite instability pathway, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9719–9723.

Yuan, A., Yu, C.J., Luh, K.T., Kuo, S.H., Lee, Y.C., Yang, P.C. (2002) Aberrant p53 expression correlates with expression of vascular endothelial growth factor mRNA and interleuikin-8 mRNA and neoangiogenesis in non-small cell lung cancer, *J. Clin. Oncol.* **20**, 900–910.

Zhu, H., Snyder, M. (2001) Protein arrays and microarrays, *Curr. Opin. Chem. Biol.* **5**, 40–45.

Optimization of Proteins for Molecular and Biomolecular Electronic Devices

Kevin J. Wise[1], *Jason R. Hillebrecht*[1], *Jeremy F. Koscielecki*[1], *Jeffrey A. Stuart*[2], *and Robert R. Birge*[1,2]
[1] *University of Connecticut, Storrs, CT, USA*
[2] *Syracuse University, Syracuse, NY, USA*

1 Introduction 551
1.1 Bacteriorhodopsin as a Bioelectronic Material 551
1.2 Photocycle Overview and Characteristics 551

2 Optimization Methods for Proteins 553
2.1 Optimization Tools 553
2.2 Genetic Modification 554
2.2.1 Site-directed Mutagenesis 554
2.2.2 Semirandom and Random Mutagenesis 554
2.2.3 Directed Evolution and Organism Selection 557
2.3 Optimization Goals for Biomolecular Electronic Devices 558
2.4 Chemical Optimization of Proteins 559
2.4.1 Chemical Additives in Bacteriorhodopsin 559
2.4.2 Chromophore Analogs in Bacteriorhodopsin 559
2.5 Photophysical Optimization of Proteins 560
2.5.1 Holographic Memories: The M State in Bacteriorhodopsin 560
2.5.2 Three-dimensional Memories: the O State in Bacteriorhodopsin 561
2.6 Thermal Optimization of Proteins 561
2.6.1 Increasing the Thermostability of BR using *Thermus thermophilus* as a host 561
2.7 Electronic Optimization of Proteins 563
2.7.1 Oriented Bacteriorhodopsin Films and Layer-by-layer Assembly 563
2.7.2 Photovoltaic Signal Enhancement 564
2.8 Bacteriorhodopsin as a Sensor 564

3 Other Proteins as Biomolecular Devices 566

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

4 **Summary** 566

Bibliography 567
Books and Reviews 567
Primary Literature 567

Keywords

Bacteriorhodopsin
A naturally occurring trans-membrane protein found in the cellular membrane of *H. Salinarum*. Activated by the absorption of a photon, the protein undergoes a photocycle, with a number of spectroscopic active intermediates, as it pumps protons across the cell membrane to form chemical energy from sunlight.

Biomolecular Devices
A system that carries out a function that uses a biological molecule to carry out all or a portion of the device function. If the device is electronic or photonic, the term bioelectronic is often used, and if semiconductor technology is involved, the adjective hybrid is added.

Biophotonics
A device that uses or manipulates light to carry out function and uses a biological molecule as the active component in the reception or manipulation of the light.

Directed Evolution
The iterative process by which genetic modification and differential selection are used to optimize any genetically amenable parameter of a biological molecule.

Mutagenesis
A group of molecular biology techniques that can produce changes in the DNA sequence of a gene and in the structure or function of a protein for which it codes.

Protein Optimization
The process of altering the primary, secondary, or tertiary structure of a protein to enhance its function in a device application. Protein optimization can be carried out by using chemical or mutagenic methods, and if a chromophore is involved, *via* substitution of the native chromophore with a modified version.

Thermostability
The temperature at which a protein becomes irreversibly inactivated. Stabilizing interactions may be introduced into a protein *via* genetic modification to increase its temperature threshold.

Protein-based bioelectronic devices have emerged as practical alternatives to semiconductor-based technology. While nature does not optimize biological molecules for performance in device applications, present-day genetic engineering has enabled investigators to manipulate the speed, durability, and architecture of candidate molecules. The biophysical, chemical, electronic, and thermal parameters of certain genetically amenable molecules have been optimized *via* a technique known as *directed evolution*. Combining genetic modification with differential selection makes it possible to tailor biological macromolecules with a high degree of accuracy and speed. Genetically modified forms of bacteriorhodopsin have improved the performance of the protein in three-dimensional volumetric memories, real-time holographic media, protein-based semiconductor devices, and artificial retinas.

1 Introduction

Biomolecular electronics is broadly defined as the manipulation and study of biological molecules for use in photonic or electronic devices. Evolutionary pressure has created some biological molecules with many of the key properties necessary for device applications or to serve as templates for further optimization. A majority of the best candidates are proteins that transport ions, conduct current, transduce light or self-assemble while maintaining function. Genetic engineering provides a route to further optimization through specific or random changes to any amino acid in the protein. Biological molecules have shown comparative advantages in a number of architectures aimed at volumetric data storage, holographic memories and devices, photovoltaic receivers and converters, biosensors, and artificial retinas.

1.1 Bacteriorhodopsin as a Bioelectronic Material

Bacteriorhodopsin (BR) is a naturally occurring protein that has been optimized via natural selection to be a highly efficient light transducing protein with good thermal stability both *in vitro* and *in vivo* environments. Because this protein must function in the outer membrane of a salt-marsh archae, the protein is stable over a wide range of temperature and pH values that are encountered in extreme environments. Further advantage is derived from the fact that BR can be isolated from the native organism and produced easily and in large quantities. Because BR has been the subject of extensive device study for the past two decades, this review will focus on this protein. But the methods of optimizing this protein for device applications are applicable to any promising protein, and this chapter will conclude with a discussion of some other candidate proteins that are currently under study.

1.2 Photocycle Overview and Characteristics

Bacteriorhodopsin functions as a proton pump in the native organism, *Halobacterium salinarum*. The protein exists as a trimer in the native organism and together with membrane lipids forms

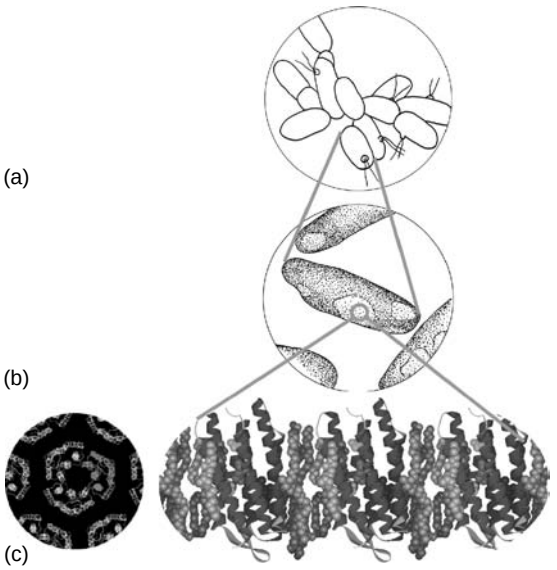

Fig. 1 Structure of bacteriorhodopsin (BR) in the purple membrane of Halobacterium salinarum. The membrane contains BR as the primary constituent, where the protein exists as a trimer in a hexagonal lattice.

a two-dimensional lattice known as the "purple membrane." Figure 1 shows the structure of the lattice and protein as it occurs in the native organism. The purple color of BR is a result of the absorbance of an embedded chromophore; in solution, retinal absorbs maximally at 380 nm, but, in the native protein environment, the complex absorbs maximally at 568 nm. The trimer lattice imparts a considerable amount of stability to BR and also provides an opportunity for the protein to self-assemble into layers. Although a native lipid matrix is not required for orientation (i.e. other proteins can also meet this goal without the existence of such a matrix), it makes the task easier.

The protein is composed of 248 amino acids that form seven membrane-spanning alpha helices. Embedded in the middle of this seven helical bundle is a light-absorbing moiety called retinal. Retinal is attached to the protein via a Schiff base linkage to lysine-216. Following absorption of a photon of light, retinal undergoes a conformational change from all-trans to 13-cis. This initial light-induced isomerization induces a series of electronic and conformational thermal transitions in the protein that proceed in the absence of light. Figure 2 shows the photocycle of BR. The main photocycle intermediates are labeled bR (the initial resting state of the protein), K, L, M, N, and O (note to reader: "BR" is used as an abbreviation to represent the protein, whereas "bR" is used to represent the resting state of the light-adapted form of this protein).

There is also a branched photocycle in BR, comprised of the P and Q states. The P state can be reached by illuminating the O state with red light ($\lambda_{max} = 640$ nm). The P state thermally decays to the Q state, which is stable for long periods of time. The Q state can then be illuminated with ultraviolet/blue light ($\lambda_{max} = 380$ nm) and the protein returns to the resting state, bR. These branched photocycle intermediates are stable for long periods of time (i.e. decades) and the protein will remain in the branched state indefinitely (P or Q) until illuminated with ultraviolet/blue

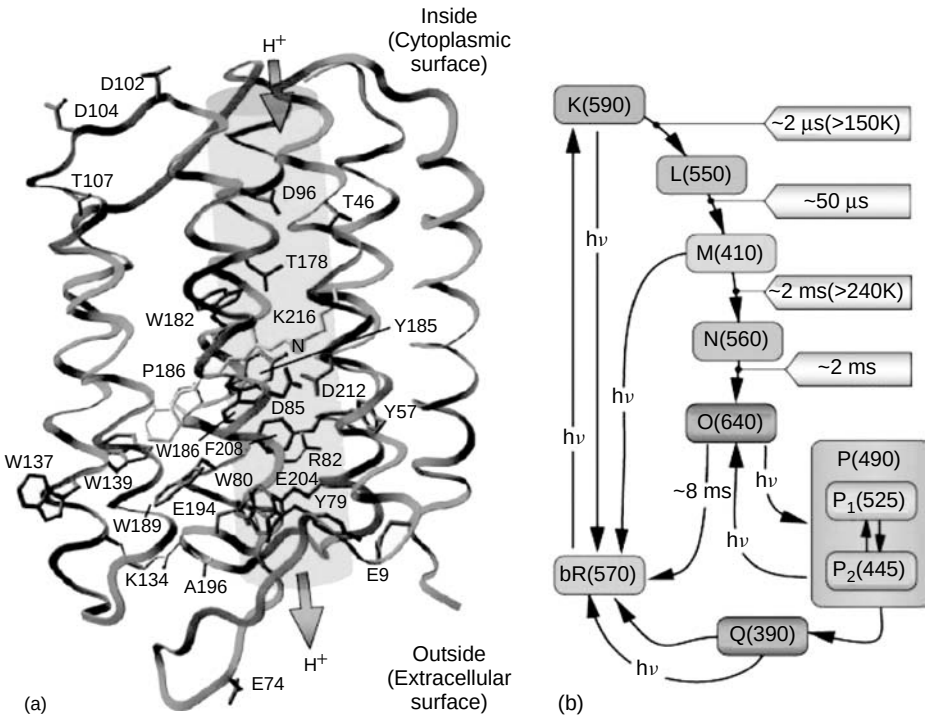

Fig. 2 Structure and photocycle of bacteriorhodopsin. Key amino acids are shown in the first panel. The second panel shows the main photocycle intermediates (bR, K, L, M, N, and O) as well as the branched photocycle intermediates (P_1, P_2, and Q).

light. The existence of this bistable system has provided promise for using BR as a template in photonic memory devices.

2 Optimization Methods for Proteins

Optimization of proteins for biomolecular electronic devices required the concerted effort of many molecular and chemical strategies. Before the insurgence of molecular techniques, traditional optimization of proteins included submitting the protein to various chemical and physical perturbations, including addition of chemicals, different solvents, and pH alteration. However, recent molecular tools have allowed for high-throughput examination of vast numbers of protein variants. Photoactive proteins can now be modified at the level of the protein, the chromophore, and the solvent environment. A unification of tools from all levels is employed to find suitable protein candidates for biomolecular electronic devices. Cross-method optimization provides the largest possibility of producing and detecting commercially viable protein variants.

2.1 Optimization Tools

A number of tools are now being used to generate photoactive protein mutants for various device applications. Mutagenesis

at various degrees of specificity is used to generate proteins that have been modified at the amino acid level. Proteins can be modified at any single amino acid or group of amino acids by mutagenesis tailored to suit the desired range of variability. The entire length of the protein can be modified randomly with no *a priori* knowledge of structure and function. Each technique can be selected depending on the desired level or specificity of variation desired. For example, a global genetic search of the entire protein may reveal a variant that can then be further studied using other optimization methods.

2.2 Genetic Modification

Table 1 summarizes the different levels of variability generation and optimization used on a protein. Mutagenesis techniques are used to generate all initial variability. Site-directed, semirandom, and random mutagenesis are used to generate genetic substitutions to one location, one region, or one gene, respectively. Directed evolution is used to iteratively generate and screen mutants that confer desired optimizations. Directed evolution type I and type II represent different levels of screening; type I involves screening on small volumes of isolated protein and type II involves screening on the actual organism expressing the protein. Table 1 shows the requirements and usefulness of each technique.

2.2.1 Site-directed Mutagenesis

Site-directed mutagenesis provides the ability to alter one specific amino acid at a time. This technique can also be used to mutate 3 to 5 contiguous amino acids. Mutation of a protein in site-directed fashion can be accomplished easily and quickly in the laboratory using any number of commercially available kits (e.g. QuikChange™ by Stratagene). Figure 3 shows the procedures and specificity provided by various mutagenesis strategies, and site-directed mutagenesis is shown in Fig. 3A. The entire process from designing mutant DNA to isolation of protein takes approximately six weeks, and most of this time (3–4 weeks) is spent cultivating *H. salinarum*, which grows at a much slower rate than commonly used *E. coli* laboratory strains.

This mutagenesis procedure can work for creating site-directed mutants in any protein given a suitable expression system (i.e. a plasmid containing the gene of interest and a method for expressing the protein) is in place. Using site-directed mutagenesis, over 100 mutants in BR have been constructed, many of which provide device optimization over wild-type (or naturally occurring) protein. Some of these mutants will be discussed later in this review. Site-directed mutagenesis is a commonly used tool in most molecular systems, and thus not limited to BR alone.

2.2.2 Semirandom and Random Mutagenesis

Site-directed mutagenesis is very useful when knowledge of the structure and function of the protein is known. However, this information is not always known. Without detailed crystal structures or higher-level theoretical calculations, it is hard to predict which amino changes will have a desired effect on the function of a protein. For this reason, more advanced mutational strategies have become popular. Semirandom and random mutagenesis are two methods to examine more global regions of a protein. An overview of these procedures is shown in Fig. 3. Semirandom mutagenesis (also called saturation mutagenesis)

Tab. 1 Generation of variability and optimization in proteins. Different mutagenesis techniques can be used to create any level of desired variability in the DNA sequence of photosensitive proteins. The specificity, variability generated, primary use, and requirements of each level is shown.

Technique	*Variability*					*Optimization*
	Site-directed mutagenesis	*Semirandom mutagenesis*	*Random mutagenesis*	*Directed evolution 1*	*Directed evolution 2*	*Chemical, photophysical, thermal, electronic*
Specificity	Single amino acid	Local region of amino acids	Global *bop* gene	Individual colony or protein	Individual *H. salinarum* cell	Protein or membrane patch
Variability Generated	Low	Medium	High	N/A	N/A	High
Primary use	Fine-tuning	Narrowing in on optimal mutation	New search	Colony/protein screening	Organism screening/selection	Material optimization
Requires	Modeling, *a priori* knowledge	Manageable segments	No *a prior* knowledge	Medium-throughput screening	High-throughput screening	High-level characterization

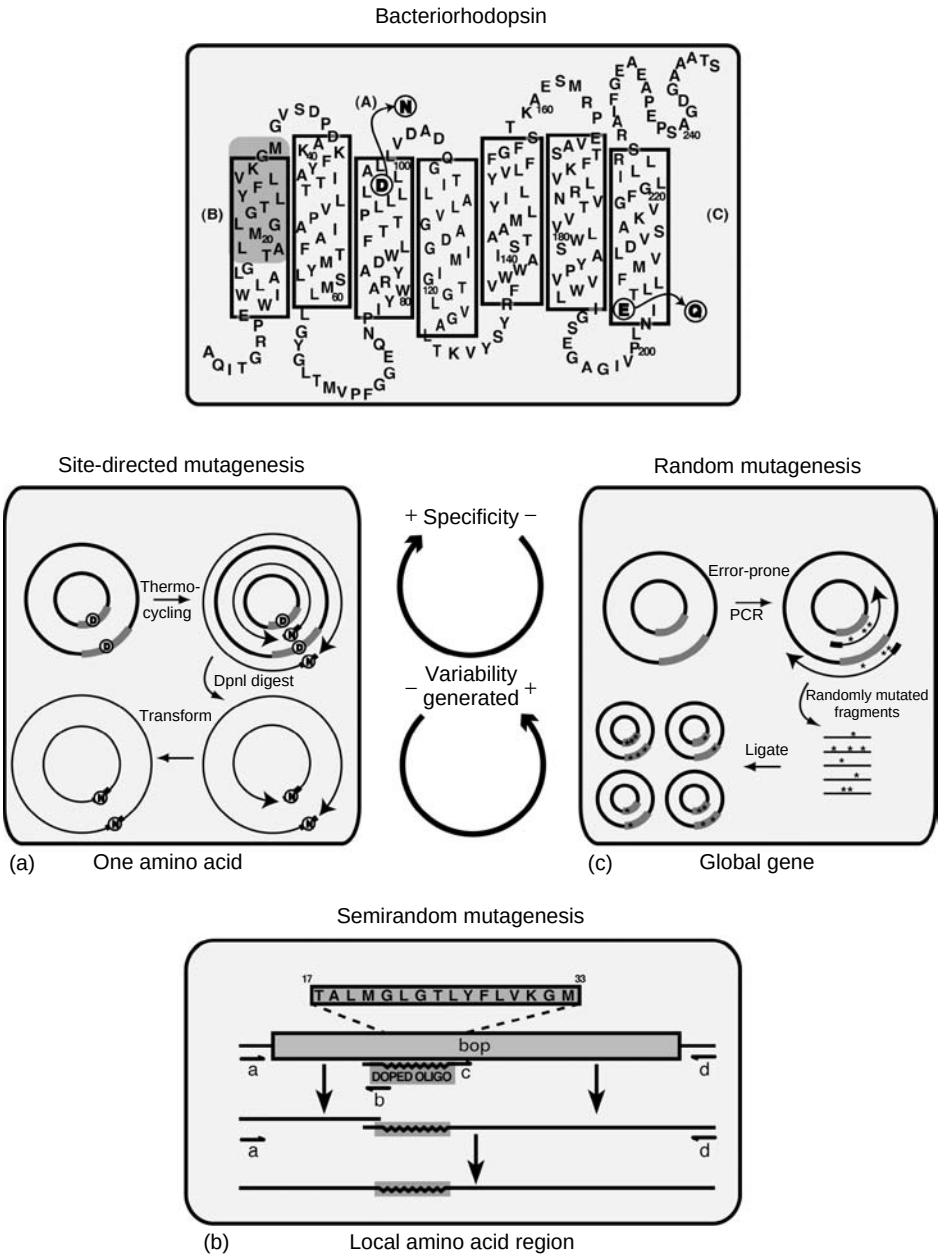

Fig. 3 Summary of mutagenesis techniques. Different mutagenesis techniques are selected based on the variability created and specificity of each technique. (a) Site-directed mutagenesis is used to modify one amino acid. (b) Semirandom mutagenesis is used to modify 10 to 20 amino acids. (c) Random mutagenesis is used to modify the entire length of the protein coding sequence.

involves mutation of a local region of 10 to 20 amino acids, while the remainder of the protein remains unaltered. For random mutagenesis, the entire protein region is affected; all amino acids are mutated with equal probability across the entire length of the protein.

Semirandom mutagenesis involves partitioning of the protein into short segments that are mutated independently. This approach was originally established by Sauer and co-workers and was also used to examine membrane protein folding and function. For example, BR contains 248 amino acids that were divided into 15 segments of ~15 amino acids each. Each segment contains a certain percentage of "doping" or probability that one amino acid is replaced by another amino acid. For example, a doping level of 20% means that there is an 80% chance that the wild-type nucleotide remains and a 6.7% chance this nucleotide is mutated to any other. Photophysical optimization of BR was accomplished using semirandom mutagenesis, and these results will be presented in Sect. 2.5.

Random mutagenesis is accomplished using many different molecular techniques, including error-prone PCR and the use of cell lines designed to introduce mutations due to errors in DNA replication mechanisms. With random mutagenesis, the entire gene is mutated with no *a priori* bias in sequence. The advantage of this technique is that no knowledge of protein structure and function is required.

The various mutagenesis techniques are most efficient when used concurrently. As shown in Fig. 3, an increase in specificity is achieved by moving from random to semirandom to site-specific mutagenesis. An increase in the amount of variability created is achieved by using the opposite order; the most variability can be created across the entire gene sequence by using random mutagenesis. For example, semirandom and random mutagenesis can be used to locate a region conferring an optimization. Then, site-directed mutagenesis can be used to dissect the particular residues responsible for a given optimization and examine their respective contributions to the overall effect.

2.2.3 Directed Evolution and Organism Selection

Directed evolution at the molecular level is the process whereby multiple iterations of mutation and selection result in proteins with optimized characteristics. Directed evolution can be used to optimize any protein, provided that an efficient method for generation and screening of mutants is in place and that the optimization is fundamentally possible (i.e. the character is able to optimized). Once these criteria are met, many different aspects of a protein can be optimized for use in biomolecular electronics devices. Many enzymes have already been optimized via directed evolution.

A typical round of optimization entails construction of new mutants, either by using semirandom or random mutagenesis, screening or selection of optimized variants, and identification of these variants. Mutants are then refined and tested using site-directed mutagenesis, and the process is repeated. Multiple rounds of this process result in directed evolution of an optimized characteristic. Typically 4 to 6 rounds of directed evolution produce significant optimizations, although it is possible to find a desired optimization in less.

Organisms that afford the ability to screen variants *in vivo* can be optimized more efficiently by elimination of time-consuming (and often expensive) protein isolation procedures. The efficiency of

the chosen screening method is the critical rate-limiting step in directed evolution procedures. However difficult *in vivo* screening and selection methods may be to devise, development of a successful method provides the ultimate streamlined tool in protein optimization. The ability to generate and screen proteins in the native organism avoids many *in vitro* steps and allows for a more high-throughput screen. The greatest chance of detecting an optimized protein is directly related to the number of proteins able to be screened in the shortest amount of time. For BR, rounds of semirandom mutagenesis have been coupled with site-directed mutagenesis to generate novel protein constructs. The protein has been optimized photophysically, and thermal optimization experiments are currently being conducted.

2.3 Optimization Goals for Biomolecular Electronic Devices

Despite the remarkable efficiency of proteins in the natural world, laboratory and device usefulness of proteins remain suboptimal. While nature has optimized proteins for speed, accuracy, and efficiency, these same attributes result in proteins that require significant modification if they are ever to be realized for device application materials.

There are many optimization goals for a particular protein for use in bioelectronic devices. Figure 4 shows the different ways that a protein such as BR is being used as a biomolecular photonic device. Four classes of optimization are chemical, photophysical, thermal, and electronic. Chemical modification involves the addition of additives and modification of the light-absorbing moieties in photosensitive proteins. Photophysical modification of proteins involves tailoring of the photocycle to meet specific photonic requirements. Thermal optimization involves increasing the denaturation temperature of proteins for increased stability in devices. Electronic optimization involves adjusting charge motion or net charge to enhance the photovoltaic properties and orientation capacity of the molecule for device applications.

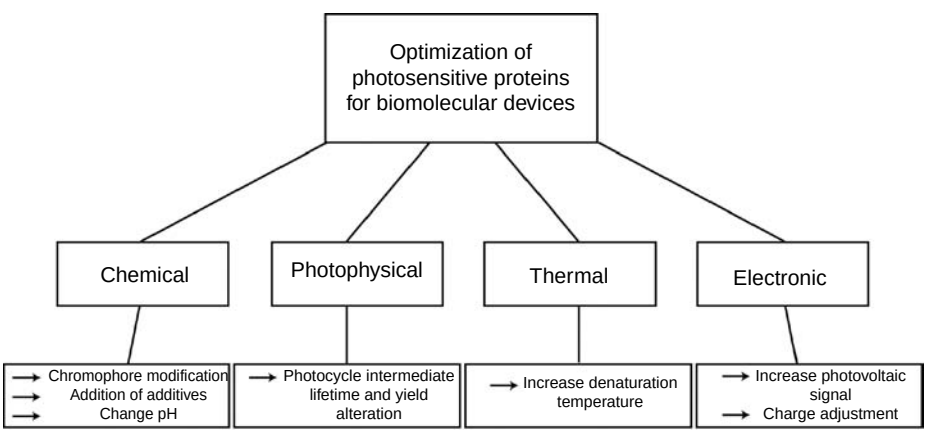

Fig. 4 An overview of the types of optimization targeted for use of a protein in biomolecular devices. The different types of optimization include chemical, photophysical, thermal, and electronic.

2.4 Chemical Optimization of Proteins

Some of the easiest changes to proteins in terms of cost and time requirements can be accomplished by changing the physical environment of the protein. Variations in pH, solvent, temperature, ionic strength, and addition of inorganic and organic chemicals can all result in altered protein structure and function. Chemical modification of any apoprotein moieties that may exist in the molecule can be used to alter the structure and function of those constituents.

2.4.1 Chemical Additives in Bacteriorhodopsin

The photocycle of BR can also be adjusted by the addition of chemical additives. Lowering the pH to less than 2.5 results in a color change from purple ($\lambda_{max} = 568$ nm) to blue ($\lambda_{max} = 605–610$ nm). This "blue membrane" results in abolishment of the photocycle. Other additives also have an impact on the photocycle of BR. Glycerol has been shown to inhibit the P $\rightarrow$ Q transition; because P $\rightarrow$ Q is a hydrolysis reaction, removal of water and addition of glycerol results in inhibition of Q state formation. Glycerol also results in longer lifetimes to main photocycle intermediate kinetics. These results have been important in understanding how BR functions in three-dimensional memories and in further elucidation of the P state.

BR has a number of naturally occurring cations present in the protein, consisting of monovalents and divalents such as calcium, magnesium, and sodium. Removal of these cations results in a shift from purple membrane to blue membrane. Replacement of these naturally occurring cations with synthesized cations known as "bolaforms" can restore the purple color of BR and also result in altered photokinetics. It is uncertain how these organic cations exert their effect on the photocycle of the protein, or where they bind in the protein, but this does not preclude their usefulness in biomolecular devices.

2.4.2 Chromophore Analogs in Bacteriorhodopsin

The retinal moiety of BR has been modified using standing organic chemistry techniques. A sample of two chromophore modifications to the retinal molecule in BR is shown in Fig. 5. Modification to the retinal molecule often results in altered photophysical kinetics. The BR variants 13-desmethyl BR and 4-keto BR were synthesized by removal of the C13 methyl group and by addition of a ketone group at C4, respectively. Studies were conducted to examine the relative conformations of the chromophores when inserted back into the protein. Whereas wild-type protein consists of a mixture of all-trans to 13-cis isomers of 50 : 50 in the dark, and 98 : 2 in the light, 13-desmethyl BR consists of a mixture of 15 : 85 in light or dark (i.e. 13-desmethyl BR does not "dark adapt" or "light adapt," or shift the relative proportions of retinal isomers when illuminated, as does wild-type protein). The other analog, 4-keto BR, results in a resting state absorption maximum of 510 nm. The M state in 4-keto BR has a vibronic structure in the M state not seen in wild type or 13-desmethyl BR. The photocycles of 13-desmethyl BR and 4-keto BR are therefore very different from the photocycle of wild-type BR and show that chromophore analogs can be used to alter the photochemistry of biomolecular BR-based systems. These results support the usefulness of modification to the apoprotein constituents of proteins

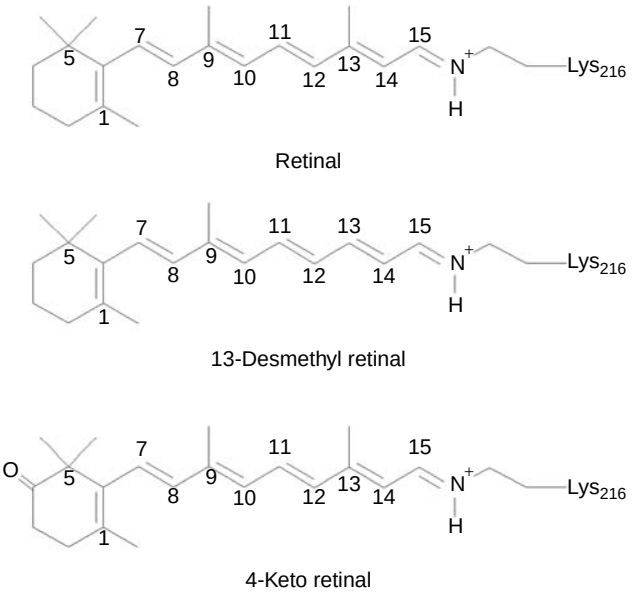

Fig. 5 Retinal isomers. Modification of the retinal molecule can be used to optimize BR for biomolecular device applications.

in addition to adjusting the amino acid structure to achieve desired photokinetic changes.

2.5 Photophysical Optimization of Proteins

Many proteins that absorb light undergo conformational and electronic changes. However, proteins with high cyclicity "reset" efficiently and, thus, following any initial changes induced by the absorbance of light, the protein is required to return to the original state. Mutations to any amino acid that has an effect on this normal functioning of the protein will adjust one or more photocycle intermediates as the protein attempts to return to the resting state.

Two bioelectronic devices have already been constructed based on the photocycle properties of BR, but many other devices are being designed. Two of these photonic devices, holographic memories and three-dimensional based memories, have been constructed using molecular variants of the protein that impact the M state and the O state (see Fig. 2).

2.5.1 Holographic Memories: The M State in Bacteriorhodopsin

The M state in BR is the most blue-shifted of all the photocycle intermediates. Holographic memory devices based on the M state take advantage of the fact that the refractive index of the M state is far removed from that of the bR resting state. A commercially available device, called the Fringemaker®, has been constructed based on a mutant variant that lengthens the M state. This mutant has a lengthened M state as a result of replacement of one aspartic acid residue to an asparagine residue at position 96 (also known as D96N). In naturally occurring protein, the lifetime of the M state is ~15 ms, but,

in this mutant variant, the lifetime of the M state reaches ~250 ms under standard conditions and pH 7. This mutant was constructed using site-directed mutagenesis. A number of long M state mutants have also been found using semirandom mutagenesis techniques, and these mutants are currently being investigated (data not shown).

2.5.2 Three-dimensional Memories: the O State in Bacteriorhodopsin

Because two distinct photocycles exist in BR (one composed of the bR, K, L, M, N, and O intermediates, and the other composed of the P and Q intermediates), information can be stored in the protein and controlled photonically (see Fig. 2). A three-dimensional memory storage device has been constructed based on the photocycle properties in BR. Significant advances to the protein for use in three-dimensional memory storage have come from site-directed and semirandom mutants. Protein mutants have been found that significantly increase the lifetime and yield of the O state. The lifetime of the O state has been increased to greater than 1 s (as compared to 20 ms in wild-type protein) using a combination of site-directed and semirandom mutagenesis techniques. Although this long lifetime is not optimal for use of BR in three-dimensional memory systems, the information provided by long O state mutants has provided valuable information about the nature of the O state and of the O to P transition.

2.6 Thermal Optimization of Proteins

Recent advancements in the field of biomolecular electronics have led to the development of device applications that utilize proteins, DNA, and a wide variety of synthetic macromolecules. While many of these macromolecules are stable in their native environments, most do not possess the degree of thermostability required for use in biomolecular electronic devices. Most organic macromolecules do not inherently possess the structural integrity to function at high temperatures, although a limited number of macromolecules thrive in extremely thermophilic environments.

Proteins with low to moderate thermostability can be engineered to function at extreme temperatures by amino acid sequence modification and computational modeling. Modeling and sequence comparisons can be used to visualize and predict regions of a protein that have poor structural integrity at high temperatures. By aligning the sequence of a thermostable protein to an analogous, mesophilic counterpart, the identity of residues likely to confer increased thermostability of the protein can be identified. Residue substitutions are then engineered into the protein to improve thermostability using the mutagenesis techniques discussed earlier.

A typical round of thermoselection is shown in Fig. 6. Pools of mutant proteins are screened for functional and structural integrity at high temperatures. Mutants that remain stable at the desired screening temperature are tested in device applications and subject to additional rounds of directed evolution for further optimization.

2.6.1 Increasing the Thermostability of BR using *Thermus thermophilus* as a host

In order to conduct directed evolution and optimization of proteins at higher temperatures, a suitable host organism must be chosen for expression of mutant proteins. For increasing the thermostability of BR, *Thermus thermophilus* (*T. thermophilus*) was chosen. *Thermus*

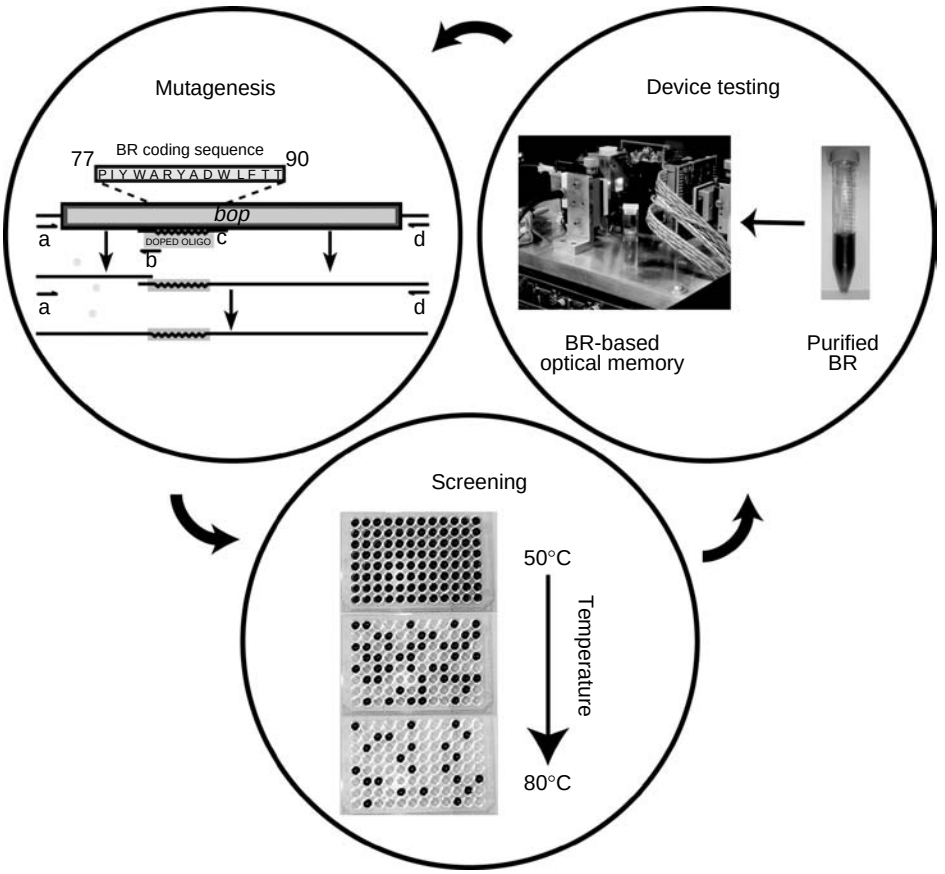

Fig. 6 A summary of a round of thermal optimization. Proteins are iteratively subjected to rounds of mutation, screening, and device testing.

thermophilus is an extremely thermophilic bacterium that grows in hot springs and industrial composts. This organism grows at temperatures in excess of 80 °C and is highly proficient at incorporating foreign DNA into the genome. These features have made *T. thermophilus* an attractive candidate to host directed evolution studies on mesophilic proteins. Recently, genetic tools were created for expression of heterologous proteins in *T. thermophilus*. These tools include thermostable antibiotic resistance selection and high copy number expression vectors. Such tools make *T. thermophilus* an efficient system to express large numbers of foreign mutant constructs at high temperatures.

Using the thermophilic expression vector described by Moreno et al., large numbers of BR variants are currently being expressed in *T. thermophilus*. The unique purple color of BR and absorbance band at 568 nm make the protein easily detectable in whole cell cultures of *T. thermophilus*. Mutant cultures are monitored relative to wild type to detect increased thermostability. Those mutants that retain structural integrity at higher temperatures

are isolated, tested for functional stability using UV-vis laser spectroscopy, and used as starting points for further rounds of mutagenesis and selection.

Bacteriorhodopsin variants that retain functional stability at the highest thermoselection temperatures are subjected to *beta* testing in biomolecular electronics devices. The performance of thermostable mutants in device applications determines whether additional optimization rounds are required.

In comparison to the work currently being conducted on BR, directed evolution has already resulted in optimization of thermal stability in a number of important enzymatic proteins. Thermally optimized enzymes are attractive agents for industrial fermentation systems, pharmaceutical treatments, and thermostable tools for genetic engineering. While these studies are significant, optimizing the thermostability of BR provides the first account of directed evolution being used to improve a membrane protein for device applications.

2.7 Electronic Optimization of Proteins

Electronic optimization of the charges and charge motion in a protein can assist in deposition and fabrication ability of the molecule for device applications. The initial photovoltaic signal present in photosensitive proteins can be enhanced in terms of both magnitude and strength of the signal by altering amino acids near the binding site of the chromophore.

2.7.1 Oriented Bacteriorhodopsin Films and Layer-by-layer Assembly

Orienting BR in a gel matrix or on a solid conductive support is achieved by a variety of deposition techniques. These techniques include electrophoretic sedimentation, Langmuir–Blodgett deposition, and immobilization onto planar lipid membranes or gel matrices. Electrophoretic sedimentation is a deposition technique used to orient an organic material between two conductive substrates by applying an external voltage. However, exposing most organics to external voltages can be damaging to both the structure and function of the macromolecule. The Langmuir–Blodgett technique involves the transfer of liquid monolayers of organic material to a solid substrate. While this technique can control the thickness and orientation of the organic components in a film, a high degree of precision is required to transfer a single monolayer of protein from liquid to solid phase. Electrostatic layer-by-layer assembly was developed out of a need for a nondestructive deposition technique that controls the thickness and orientation of an organic macromolecule onto a solid substrate with minimal expertise.

Recently, electrostatic layer-by-layer assembly was used to fabricate oriented thin films of BR (in the form of purple membrane fragments) without applying an external voltage or using a bulky polymer matrix. Layer-by-layer deposition uses electrostatic forces to orient a bilayer of protein and a polyelectrolyte (a polycation in the case of BR) to a solid substrate. In the case of BR, the cytoplasmic side of the protein is more negatively charged than the extracellular side in alkaline conditions. The dipolar nature of BR at high pH enhances its capacity to bind to a layer of polycations.

The substrate of choice is initially treated with potassium hydroxide to give it a layer of negatively charged ions on its surface. The negatively charged substrate is then dipped into a polycation solution to line it with a uniform layer of positive

charges. Upon removing the loosely attached polycation molecules with a series of washes, the substrate is dipped into an alkaline solution of BR. The cytoplasmic side of the protein attaches to the positively charged substrate, thus resulting in a uniform bilayer of polycation and protein. The layer-by-layer dipping procedure is quick (~12 min/layer), easily performed, and amenable to a number of different substrate shapes and orientations. Automating this procedure is possible when films with 100 to 200 bilayers of protein/polycation are needed.

There is a direct relationship between the number of deposited bilayers of protein/polycation and optical density of the thin film. Alternate layer-by-layer deposition proves that it is possible to both orient and control the number of bilayers assembled onto a solid support. This strategy can be used for other proteins in addition to BR and has a profound effect on the fabrication of thin films for biomolecular device applications.

2.7.2 Photovoltaic Signal Enhancement

Many photoactive proteins generate a fast, high intensity photoelectric signal upon light excitation. The photoelectric response of BR is a multicomponent signal that is triggered by the photoisomerization of a bound retinal molecule and a series of structural rearrangements of the protein. The ultrafast onset of the photovoltaic signal makes BR an appealing candidate for biomolecular field effect transistors and optoelectronic integrated circuits. A drawback to the fast photoelectric response of BR is the relatively long decay time of the signal (~4.5 ps). Utilizing BR as an efficient photodetecting element requires that the protein generate a signal with an ultrafast onset and decay time. Optimization of the photovoltaic response of the protein is achieved by employing a combination of mutagenesis techniques to alter residues in the retinal-binding pocket. Retinal analogs are also used to complement mutagenesis techniques in order to further optimize the photoelectric response of the protein.

The architecture of many biomolecular devices utilizes organic macromolecules, specifically proteins, in the form of a dried film. Dried films of protein are less bulky than liquid or gel matrices and resistant to fluctuation in pH and temperature. In the case of BR, thin film fabrication must preserve the structure of the protein while optimizing the photoelectric characteristics. To preserve the structure and function of BR, the protein is isolated and deposited onto the substrate in the form of purple membrane patches. Purple membrane patches are approximately 5-nm thick and are primarily composed of BR molecules arranged in a hexagonal lattice of trimers. This architecture allows BR to absorb light in all polarizations while stabilizing the protein over a wide range of temperature and pH values. Such characteristics make purple membrane films ideal for biomolecular device applications.

2.8 Bacteriorhodopsin as a Sensor

Bacteriorhodopsin has been long known to display an extreme sensitivity to environment conditions such as pH, relative humidity, and ionic strength. The effects of this sensitivity can be manifested as modulations to the photocycle or photovoltaic effect. Furthermore, various chemicals have well-defined and well-characterized effects on the protein. One such class of chemical modulators is organic amines, which act to prolong the M intermediate. The M state decays as a function of the deprotonation

of Asp 96, and any environmental condition that favors reduced proton mobility will prolong this state, including, for example, high pH or dehydration. Organic amines also facilitate prolonged M states; at first glance, it might be expected that this is a pH-related affect, although given sample preparations with identical pH, those treated with organic amines exhibit more pronounced effects. The O state, also characterized by proton transfer, is similarly sensitive to the presence of environmental additives. The mechanisms that cause these responses are not well understood.

Other classes of chemicals that modulate the photophysical properties of BR include alcohols (methanol, ethanol, propanol, and butanol) and anesthetics. Azides are also known to modulate the photophysical properties of BR. Sensitivity of the protein to alcohols is manifested in the form of modulations of the light-induced photocurrent of the protein and could be enhanced by chemical modification of the chromophore or genetic manipulation of the protein. One plausible mechanism for the alcohol-induced sensitivity of BR is through local dehydration effects that essentially mimic reduced relative humidity. Numerous studies have demonstrated the effect of anesthetics on BR response, which exhibit several general modes of action. Anesthetics modify ion transport through membrane channels and thus are of interest with respect to the nervous system. The concentration-induced effects observed in BR consist of a slight initial blue shift (569–567 nm) to a state with an accelerated M state. Higher anesthetic concentrations result in a larger blue shift (480 nm), a prolonged M state, and a loss of both proton pumping function and the purple membrane lattice structure. Some anesthetics will produce a 380-nm absorbing species. At lower anesthetic concentrations, most effects were found to be reversible. X-ray studies indicate that anesthetics bind at the lipid protein interface in the center of a BR trimer. Other studies have posited distinct binding sites. Generally speaking, therefore, chemical agents have several modes by which they can modulate protein response, including direct interaction with, or binding to, the protein, and/or indirect interaction by absorption into the lipids proximal to the protein.

As indicated earlier, various chemical species have been used in the past to modulate the response of the protein to light by employing chemicals that modulate the M and O states. Although these attempts were targeted at developing photonic devices, they are indicative of the ability of this protein to respond to and register the state of the immediate environment. Furthermore, as detailed above, a wide variety of external molecular stimuli have proven to modulate the response of the protein, indicating the potential role BR might play as a chemical sensor in a hybrid protein-semiconductor architecture. Given the ability to interrogate the protein with light, signals can be extracted that are a reflection of the environment experienced by the protein. Such signals might result from a modulation of the photocycle kinetics, the photovoltaic effect, or a combination of both. Thus, by extension, it is possible to use either the photocycle kinetics or photovoltaic responsivity to monitor the concentration of volatile species present in the environment. It can be concluded that the complexity of BR photophysics and sensitivity to external molecular stimuli combine to make this protein a strong candidate for the design of hybrid protein-semiconductor sensors in detection schemes for chemical agents that are deleterious to human health, including environmental toxins, industrial

pollutants, and chemical warfare agents. Additionally, the ability to implement multiple interrogation techniques gives BR a unique advantage over other more conventional materials.

3 Other Proteins as Biomolecular Devices

Discussion so far has been limited to the optimization of BR. This emphasis is appropriate from a historical perspective because the vast majority of bioelectronic devices have involved the use of BR and BR variants. Nevertheless, there are many other proteins that are currently under investigation for device applications. Greenbaum and Boxer have investigated the use of photosynthetic proteins and reaction centers for device applications. These proteins exhibit high photoconversion efficiencies and intrinsic diodelike characteristics that are useful for a number of applications.

One might anticipate that the visual pigment rhodopsin would find device application, but the use of this protein poses a major problem that will not be easily solved. When the protein absorbs a photon of light, the chromophore is isomerized from 11-*cis* to all-*trans* and is expelled from the protein. In vivo, a complex enzymatic process regenerates 11-*cis* retinal that is transported back to the rod outer segment and spontaneously recombines with the protein to reform an active visual pigment. The cycle requires many minutes and would be extremely difficult to implement in a device environment. However, there are sources of rhodopsin that provide binary responses without chromophore expulsion. Many invertebrate proteins operate in this fashion and this avenue is worthy of further study. Proteorhodopsin, a recently discovered light-transducing protein from oceanic bacteria, shows significant potential for device applications. This protein shares many of the positive characteristics of BR while providing significant differences in the photocycle that may provide a comparative advantage in photonic devices. What remains is to improve the thermal stabilities and inherent lifetimes of these proteins when used in device environments. Not one of the above mentioned proteins could be used in high-temperature electronic systems. Directed evolution will play an important role in optimizing the thermal stability of candidate proteins.

4 Summary

Nature has been optimizing proteins for roughly three billion years, and many of the characteristics required of proteins to function in biology are identical to those required for device applications. Thus, natural materials are logical candidates for electronic and photonic devices. Nevertheless, virtually all device environments pose unique problems that require further optimization of biological molecules for the device to function at full potential. This chapter has provided an overview of these methods and procedures using the protein bacteriorhodopsin (BR) as the template. A key observation, however, is that BR is only one of many proteins that has potential for device applications. Bioelectronics is in a very early stage of making contributions to electronics and photonics, and only a small number of architectures have been explored. Genetic engineering and directed evolution will prove crucial to the full exploration and optimization of materials for bioelectronic application.

See also Informatics (Computational Biology).

Bibliography

Books and Reviews

Hirai, T., Subramaniam, S. (2003) Structural insights into the mechanism of proton pumping by bacteriorhodopsin, *FEBS Lett.* **545**(1), 2–8.

Lanyi, J.K. (1999) Bacteriorhodopsin, *Int. Rev. Cytol.* **187**, 161–202.

Lanyi, J., Luecke, H. (2001) Bacteriorhodopsin, *Curr. Opin. Struct. Biol.* **11**, 415–419.

Lehmann, M., Wyss, M. (2001) Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution, *Curr. Opin. Biotechnol.* **12**, 371–375.

Wise, K.J., Birge, R.R. (2003) Biomolecular Photonics Based on Bacteriorhodopsin, in: Horspool, W., Lenci, F. (Eds.) *CRC Organic Handbook of Photochemistry and Photobiology*, 2nd edition, CRC Press, New York, Chapter 135.

Primary Literature

Arnold, F., Moore, J.C. (1997) Optimizing industrial enzymes by directed evolution, *Adv. Biochem. Eng. Biotechnol.* **58**, 1–14.

Arnold, F., Wintrode, P.L., Miyazaki, K., Gershenson, A. (2001) How enzymes adapt: lessons from directed evolution, *Trends Biochem. Sci.* **26**(2), 100–106.

Becker, R.S., Inuzuka, R., Balke, D.E. (1971) Comprehensive investigation of the spectroscopies and photochemistry of retinals. I. theoretical and experimental considerations of absorption spectra, *J. Am. Chem. Soc.* **93**, 38–42.

Beffa, T., Blanc, M., Lyon, P.F., Vogt, G., Marchiani, M., Fischer, J.L., Aragno, M. (1996) Isolation of *Thermus* strains from hot composts, *Appl. Environ. Microbiol.* **62**(5), 1723–1727.

Béjà, O., Spudich, E., Spudich, J., Leclerc, M., DeLong, E. (2001) Proteorhodopsin phototrophy in the ocean, *Nature* **411**, 786–789.

Béjà, O., Aravind, L., Koonin, E., Suzuki, M., Hadd, A., Nguyen, L., Joanovich, S., Gates, C., Feldman, R., Spudich, J., Spudich, E., DeLong, E. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea, *Science* **289**, 1902–1906.

Birge, R.R. (1990) Photophysics and molecular electronic applications of the rhodopsins, *Annu. Rev. Phys. Chem.* **41**, 683–733.

Birge, R.R. (1994) Three-dimensional optical memories, *Am. Sci.* **82**, 349–355.

Birge, R.R., Gillespie, N.B., Izaguirre, E.W., Kusnetzow, A., Lawrence, A.F., Singh, D., Song, Q.W., Schmidt, E., Stuart, J.A., Seetharaman, S., Wise, K.J. (1999) Biomolecular electronics: protein-based associative processors and volumetric memories, *J. Phys. Chem. B* **103**, 10746–10766.

Blatz, P.E., Johnson, R.H., Mohler, J.H., Al-Dilaimi, S.K., Dewhurst, S., Erickson, J.O. (1971) Effect of solution environment on the absorption maximum of schiff bases of retinal, *Photochem. Photobiol.* **13**, 237–245.

Boucher, F., Taneva, S.G., Elouatik, S., Dery, M., Messaoudi, S., Harvey-Girard, E., Beaudoin, N. (1996) Reversible inhibition of proton release activity and the anesthetic-induced acid-base equilibrium between the 480 and 570 Nm forms of bacteriorhodopsin, *Biophys. J.* **70**(2), 948–961.

Bowie, J.U., Reidhaar-Olson, J.F., Lim, W.A., Sauer, R.T. (1990) Deciphering the message in protein sequences: tolerance to amino acid substitutions, *Science* **247**(4948), 1306–1310.

Boxer, S.G., Stocker, J., Franzen, S., Salafsky, J. (1992) Re-engineering Photosynthetic Reaction Centers, in: Aviram, A. (Ed.) *Molecular Electronics- Science and Technology*, Vol. 262, American Institute of Physics, New York, pp. 226–241.

Bryl, K., Yoshihara, K. (2000) Can Purple Membrane Play the Role Of Biosensor Under Extreme Conditions, in: Der, A., Keszthelyi, L. (Eds.) *Bioelectric Applications of Photonic Pigments*, Vol. 335. Series 1: Life and Behavioural Sciences, IOS Press, Amsterdam, pp. 211–224.

Bryl, K., Varo, G., Drabent, R. (1991) The photocycle of bacteriorhodopsin immobilized in poly(vinyl alcohol) film, *FEBS Lett.* **285**(1), 66–70.

Chen, Z., Birge, R.R. (1993) Protein based artificial retinas, *Trends Biotech.* **11**, 292–300.

Dyukova, T.V., N.N. Vsevolodov. (1996) *Photochromic Compositions and Materials Containing Bacteriorhodopsin*, U. S. Patent No. 5,518,858, May 21, *1996*.

Friedrich, A., Prust, C., Hartsch, T., Henne, A., Averhoff, B. (2002) Molecular analysis of the natural transformation machinery and identification of pilus structures in the extremely thermophilic bacterium *Thermus thermophilus* strain Hb27, *Appl. Environ. Microbiol.* **68**(2), 745–755.

Gao, M.M., Boucher, F. (1998) The uncoupling of bacteriorhodopsin by high temperature and anaesthetics, *Toxicol Lett.* **100–101**, 393–396.

Gärtner, W. (2000) Invertebrate Visual Pigments, in: Stavenga, D.G., DeGrip, W.J., Pugh E.N., Jr. (Eds.) *Handbook of Biological Physics*, Vol. 3, Elsevier, New York.

Gillespie, N.B., Wise, K.J., Ren, L., Stuart, J.A., Marcy, D.L., Hillebrecht, J., Li, Q., Ramos, L., Jordan, K., Fyvie, S., Birge, R.R. (2002) Characterization of the branched-photocycle intermediates P and Q of bacteriorhodopsin, *J. Phys. Chem. B* **106**, 13352–13361.

Greenbaum, E. (1992) Kinetic studies of interfacial photocurrent in platanized chloroplasts, *J. Phys. Chem.* **96**, 514–516.

Greenbaum, E. (1990) Vectorial photocurrents and photoconductivity in metalized chloroplasts, *J. Phys. Chem.* **94**, 6151–6153.

Groma, G.I., Hebling, J., Ludwig, C., Kuhi, J. (1995) Charge displacement in bacteriorhodopsin during forward and backward Br-K phototransition, *Biophys. J.* **69**, 2060–2065.

Hamanaka, T., Nakagawa, T., Kito, Y., Nishimura, S., Uchida, I., Mashimo, T. (1998) Binding of volatile anesthetics to purple membranes studied by X-ray diffraction, *Toxicol. Lett.* **100–101**, 397–403.

Hampp, N., Juchem, T (2000) Fringemaker – the First Technical System Based on Bacteriorhodopsin, in: Dér, A., Keszthelyi, L. (Eds.) *Bioelectronic Applications of Photochromic Pigments*, Vol. 335, NATO Science Series I: Life and Behavioural Sciences, IOS Press, Szeged, Hungary, pp. 44–53.

He, J.-A., Samuelson, L., Li, L., Kumar, J., Tripathy, S.K. (1998) Oriented bacteriorhodopsin/polycation multilayers by electrostatic layer-by-layer assembly, *Langmuir* **14**, 1674–1679.

Hendler, R., Dracheva, S. (2001) Importance of lipids for bacteriorhodopsin structure, photocycle, and function, *Biochemistry. Biokhimiia.* **66**(11), 1311–1314.

Heyes, C., El-Sayed, M. (2002) The role of native lipids and lattice structure in bacteriorhodopsin protein conformation and stability as studied by temperature-dependent Fourier transform-infrared spectroscopy, *J. Biol. Chem.* **277**(33), 29427–29443.

Hong, F. (1994) Retinal Proteins in Photovoltaic Devices, in: Birge, R.R. (Ed.) *Molecular and Biomolecular Electronics*, Vol. 240, Advances in Chemistry, American Chemical Society, Washington, DC, pp. 527–560.

Hong, F.T. (1997) Molecular sensors based on the photoelectric effect of bacteriorhodopsin: origin of differential responsivity, *Mater. Sci. Eng. C* **4**, 267–385.

Juchem, T., Hampp, N. (2000) Interferometric system for non-destructive testing based on large diameter bacteriorhodopsin films, *Opt. Lasers Eng.* **34**(2), 87–100.

Jussila, T., Li, M., Tkachenko, N.V., Parkkinen, S., Li, B., Jiang, L., Lemmetyinen, H. (2002) Transient absorption and photovoltage study of self-assembled bacteriorhodopsin/polycation multilayer films, *Biosens. Bioelectron.* **17**, 509–515.

Keszthelyi, L. (1980) Orientation of membrane fragments by electric field, *Biochim. Biophys. Acta* **598**, 429–436.

Kotsuka, T., Akanuma, S., Tomuro, M., Yamagishi, A., Oshima, T. (1996) Further stabilization of 3-isopropylmalate dehydrogenase of an extreme thermophile, *Thermus thermophilus*, by a suppressor mutation method, *J. Bacteriol.* **178**, 723–727.

Krebs, M., Isenbarger, T. (2000) Structural determinants of purple membrane assembly, *Biochim. Biophys. Acta* **1460**, 15–26.

Kuchner, O., Arnold, F. (1997) Directed evolution of enzyme catalysts, *Trends Biotech.* **15**, 523–530.

Lee, I., Lee, J.W., Greenbaum, E. (1997) Biomolecular electronics: vectorial arrays of photosynthetic reaction centers, *Phys. Rev. Lett.* **79**, 3294–3297.

Lee, K.H., McIntosh, A.R., Boucher, F. (1991) The interaction between halogenated anaesthetics and bacteriorhodopsin in purple membranes as examined by intrinsic ultraviolet fluorescence, *Biochem. Cell Biol.* **69**(2–3), 178–184.

Lin, C.T., Mertz, C.J., Bitting, H.C., El-Sayed, M.A. (1992) Fluorescence anisotropy

studies of dibucaine.Hcl in micelles and bacteriorhodopsin, *J. Photochem. Photobiol. B* **13**(2), 169–185.

Lin, C.T., Chyan, Y.G., Kresheck, G.C., Bitting, H.C., el-Sayed, M.A. Jr. (1989) Interaction of dibucaine.Hcl local anesthetics with bacteriorhodopsin in purple membrane: a spectroscopic study, *Photochem. Photobiol.* **49**(5), 641–648.

Lo, K.M., Jones, S.S., Hackett, N.R., Khorana, H.G. (1984) Specific amino acid substitutions in bacterioopsin: replacement of a restriction fragment in the structural gene by synthetic DNA fragments containing altered codons, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2285–2289.

Lozier, R.H., Niederberger, W. (1977) The photochemical cycle of bacteriorhodopsin, *Fed. Proc.* **36**, 1805–1809.

Lozier, R.H., Bogomolni, R.A., Stoeckenius, W. (1975) Bacteriorhodopsin: a light-driven proton pump in halobacterium halobium, *Biophys. J.* **15**, 955–962.

Lukashev, E.P., Robertson, B. (1995) Bacteriorhodopsin retains its light-induced proton-pumping function after being heated to 140 °C, *Bioelectrochem. Bioenerg.* **37**, 157–160.

Misra, S., Govindjee, R., Ebrey, T.G., Chen, N., Ma, J.X., Crouch, R.K. (1997) Proton uptake and release are rate-limiting steps in the photocycle of the bacteriorhodopsin mutant E204q, *Biochemistry* **36**(16), 4875–4883.

Miyasaka, T., Koyama, K., Itoh, I. (1992) Quantum Conversion and image detection by a bacteriorhodopsin-based artifical photoreceptor, *Science* **255**, 342–344.

Moreno, R., Zafra, O., Cava, F., Berenguer, J. (2003) Development of a gene expression vector for *Thermus thermophilus* based on the promoter of the respiratory nitrate reductase, *Plasmid* **49**, 2–8.

Muller, D.J., Heymann, J.B., Oesterhelt, F., Moller, C., Gaub, H., Buldt, G., Engel, A. (2000) Atomic force microscopy of native purple membrane, *Biochim. Biophys. Acta* **1460**(1), 27–38.

Nakagawa, T., Hamanaka, T., Nishimura, S., Uruga, T., Kito, Y. (1994) The specific binding site of the volatile anesthetic diiodomethane to purple membrane by X-ray diffraction, *J. Mol. Biol.* **238**(3), 297–301.

Nakagawa, T., Hamanaka, T., Nishimura, S., Uchida, I., Mashimo, T., Kito, Y. (2000) The quantitative analysis of three action modes of volatile anesthetics on purple membrane, *Biochim. Biophys. Acta* **1468**(1–2), 139–49.

Nishimura, S., Mashimo, T., Hiraki, K., Hamanaka, T., Kito, Y., Yoshiya, I. (1985) Volatile anesthetics cause conformational changes of bacteriorhodopsin in purple membrane, *Biochim. Biophys. Acta* **818**(3), 421–424.

Oesterhelt, D., Stoeckenius, W. (1971) Rhodopsin-like protein from the purple membrane of Halobacterium halobium, *Nature (London), New. Biol.* **233**, 149–152.

Pande, C., Callender, R.H., Chang, C.H., Ebrey, T.G. (1986) Resonance Raman study of the pink membrane photochemically prepared from the deionized blue membrane of H. Halobium, *Biophys. J.* **50**, 545–549.

Pepe, I.M., Nicolini, C. (1996) Langmuir-blodgett films of photosensitive proteins, *J. Photochem. Photobiol. B: Biology.* **33**, 191–200.

Perozo, E., Hubbell, W.L. (1993) Voltage activation of reconstituted sodium channels: use of bacteriorhodopsin as a light-driven current source, *Biochemistry* **32**(39), 10471–10478.

Popp, A., Wolperdinger, M., Hampp, N., Bräuchle, C., Oesterhelt, D. (1993) Photochemical conversion of the O-intermediate to 9-Cis-retinal-containing products in bacteriorhodopsin films, *Biophys. J.* **65**, 1449–1459.

Radionov, A., Kaulen, A. (1996) Two bacteriorhodopsin M intermediates differening in accesibility of the schiff base for azide, *FEBS Lett.* **387**, 122–126.

Renthal, R., Brogley, L., Vila, J. (1988) Altered protein-chromophore interaction in dicyclohexylcarbodiimide-modified purple membrane sheets, *Biochim. Biophys. Acta* **935**(2), 109–114.

Sasabe, H., Furuno, T., Takimoto, K. (1989) Photovoltaics of photoactive protein/polypeptide Lb films, *Synth. Met.* **28**, C787–CC92.

Seta, P., Ormos, P., d'Epenoux, B., Gavach, C. (1980) Photocurrent response of bacteriorhodopsin adsorbed on bimolecular lipid membranes, *Biochim. Biophys. Acta* **591**(1), 37–52.

Shen, Y., Safinya, C.R., Liang, K.S., Ruppert, A.F., Rothschild, K.J. (1993) Stabilization of the membrane protein bacteriorhodopsin to 140 °C in two-dimensional films, *Nature* **366**, 48–50.

Simmeth, R., Rayfield, G.W. (1990) Evidence that the photoelectric response of bacteriorhodopsin occurs in less than 5 picoseconds, *Biophys. J.* **57**, 1099–1101.

Steinhoff, H., Pfeiffer, M., Rink, T., Burlon, O., Kurz, M., Riesle, J., Heuberger, E., Gerwert, K., Oesterhelt, D. (1999) Azide reduces the hydrophobic barrier of the bacteriorhodopsin proton channel, *Biophys. J.* **76**, 2702–2710.

Tabata, K., Kosuge, T., Nakahara, T., Hoshino, T. (1993) Physical map of the extremely thermophilic bacterium *Thermus thermophilus* Hb27 chromosome, *FEBS Lett.* **331**(1.2), 81–85.

Takao, K., Kikukawa, T., Araiso, T., Kamo, N. (1998) Azide accelerates the decay of M-intermediate of pharaonis phoborhodopsin, *Biophys. Chem.* **73**(1–2), 145–53.

Takei, H., Lewis, A., Chen, Z., Nebenzahl, I. (1992) Implementing receptive fields with excitatory and inhibitory optoelectrical responses of bacteriorhodopsin films, *Appl. Opt.* **30**, 500–509.

Tamakoshi, M., Nakano, Y., Kakizawa, S., Yamagishi, A., Oshima, T. (2001) Selection of stabilized 3-isopropylmalate vector system of an extreme thermophile, *Thermus thermophilus*, *Extremophiles* **5**, 17–22.

Tan, E.H.L., Govender, D.S.K., Birge, R.R. (1996) Large organic cations can replace Mg^{2+} and Ca^{2+} ions in bacteriorhodopsin and maintain proton pumping ability, *J. Am. Chem. Soc.* **118**, 2752–2753.

Taneva, S.G., Caaveiro, J.M., Petkanchin, I.B., Goni, F.M. (1995) Electrokinetic charge of the anesthetic-induced Br480 and Br380 spectral forms of bacteriorhodopsin, *Biochim. Biophys. Acta* **1236**(2), 331–337.

Uruga, T., Hamanaka, T., Kito, Y., Uchida, I., Nishimura, S., Mashimo, T. (1991) Effects of volatile anesthetics on bacteriorhodopsin in purple membrane, halobacterium halobium cells and reconstituted vesicles, *Biophys. Chem.* **41**(2), 157–168.

Wald, G. (1968) The molecular basis of visual excitation, *Nature* **219**, 800–808.

Wen, J., Chen, X., Bowie, J.U. (1996) Exploring the allowed sequence space of a membrane protein, *Nat. Struct. Biol.* **3**(2), 141–148.

Willner, I., Willner, B. (2001) Biomaterials integrated with electronic elements: en route to bioelectronics, *Trends Biotechnol.* **19**(6), 222–230.

Wise, K.J., Gillespie, N.B., Stuart, J.A., Krebs, M.P., Birge, R.R. (2002) Optimization of bacteriorhodopsin for bioelectronic devices, *Trends Biotechnol.* **20**(9), 387–394.

Xu, J., Stickrath, A.B., Bhattacharya, P., Nees, J., Varo, G., Hillebrecht, J.R., Ren, L., Birge, R.R. (2003) Direct measurement of the photoelectric response time of bacteriorhodospin via electro-optic sampling, *Biophys. J.* **85**(2), 1128–1134.

Zaks, A. (2001) Industrial biocatalysis, *Curr. Opin. Chem. Biol.* **5**, 130–136.

Organic Cofactors as Coenzymes

Donald B. McCormick
Department of Biochemistry, Emory University, Atlanta GA 30322-3050, USA

1 General History 572

2 Individual Coenzymes 573
2.1 Vitamin-derived Coenzymes 573
2.1.1 Thiamine Pyrophosphate 573
2.1.2 Flavocoenzymes 575
2.1.3 Pyridine Nucleotide Coenzymes 577
2.1.4 Pyridoxal 5′-phosphate and Pyridoxamine 5′-phosphate 579
2.1.5 Pterin Coenzymes 580
2.1.6 B_{12} Coenzymes 582
2.1.7 Biotinyl Functions 583
2.1.8 Phosphopantetheinyl Coenzymes 585

2.2 Other Coenzymes 586
2.2.1 Pyruvoyl Functions 586
2.2.2 Lipoyl Functions 586
2.2.3 Ubiquinones 586
2.2.4 Pyrroloquinoline Quinone and Quinoproteins 587
2.2.5 Methanogen Coenzymes 587

Bibliography 587
Books and Reviews 587

Keywords

Apoenzyme
A protein moiety of an enzyme that requires a coenzyme.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 9
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30551-3

Cofactor
A natural reactant, usually either a metal ion or coenzyme, required in an enzyme-catalyzed reaction.

Holoenzyme
A catalytically active enzyme constituted by coenzyme bound to apoenzyme.

Vitamin
An essential organic micronutrient that must be supplied exogenously and in many cases is the precursor to a metabolically derived coenzyme.

A coenzyme is an organic molecule that can bind within enzymes that require its function to catalyze a biochemical reaction. Hence, coenzymes are organic cofactors that augment the diversity of reactions that otherwise would be more limited to chemical properties of side chain substituents from amino acid residues within protein enzymes. Coenzymes bind to apoenzymes to generate catalytically competent holoenzymes. With tight binding, coenzymes may be referred to as prosthetic groups; with loose binding, they may be called *cosubstrates*.

1 General History

Most work on the chemical nature of coenzymes began in the 1930s, when elucidation of the structures of some vitamins in the B complex was being accomplished. In 1932, Auhagen had succeeded in dissociating a heat-stable component of yeast "carboxylase," which he called "cocarboxylase." By 1937, Lohmann and Schuster had determined its structure as thiamine pyrophosphate (TPP). Warburg and Christian were active in elucidating the nature of the oxidation–reduction coenzymes that contain riboflavin (FMN, FAD: see Sect. 2.1.2) and nicotinamide (NAD, NADP: see Sect. 2.1.3) during this same period. By the mid-1940s, the coenzyme forms of B_6 (PLP, PMP: see Sect. 2.1.4) were recognized by Gunsalus, Snell, and others. The discovery by Lipmann in 1947 of a coenzyme of acetylation (CoA), which contained pantothenate, led to the full structure elucidation by Baddiley et al. in 1953. Advances also made in the 1950s by a number of laboratories led to the recognition of coenzyme forms of folacin, and by the end of the decade, Lynen et al. had characterized covalent N^1-carboxybiotin and Barker and his associates had identified a coenzyme form of B_{12}. Prerequisite to determining structures of coenzymes were recognition and purification of enzyme systems that depend on the catalytic participation of these organic cofactors. Continued examinations of diverse enzymes are extending the list of newer coenzymes, some of which do not derive from vitamins.

2 Individual Coenzymes

2.1 Vitamin-derived Coenzymes

Many coenzymes are the metabolic result of converting ingested vitamins, especially those of the B complex, to forms suitable for binding and function in enzyme systems. Of the 13 vitamins presently known to be required in the diet of humans and most animals, at least eight (thiamine, riboflavin, niacin, vitamin B_6, folacin, vitamin B_{12}, biotin, and pantothenate) are simply the essential precursors for coenzymic forms made in the body. These coenzymes and their functions are listed in Table 1.

2.1.1 Thiamine Pyrophosphate

The pyrophosphate ester formed at the β-hydroxyethyl substituent in position 5 of the thiazole moiety of thiamine is the principle if not sole coenzyme derived from thiamine. TPP, as is the case with the parent vitamin, is relatively stable in acidic aqueous solutions, but sulfite causes cleavage at the methylene bridge to release the substituted pyrimidylsulfonate and the thiazole pyrophosphate. TPP is unstable in alkaline medium because the thiazole portion is subject to base attack at C-2. With hydroxyl ion, this leads to a pseudobase, thiazole ring opening, and some disulfide under mild oxidizing conditions. Also, the 4-amino on the pyrimidyl portion can attack as an intramolecular base to form the tricyclic amino adduct, which can be oxidized with ferricyanide to yield thiochrome pyrophosphate. The thiochrome-level compound is fluorescent (wavelengths of excitation and emission, 385 and 440 nm, respectively) and is easily detected.

TPP is formed in a number of tissues from thiamine, and a fraction is subsequently converted, especially in brain, to the triphosphate. Hydrolysis of the latter to the monophosphate and some further release of free vitamin occurs. Among the forms involved, TPP predominates in cells. About half the body stores are found in skeletal muscle, with much of the remainder in heart, liver, kidneys, and nervous tissue, including brain, which contains much of the triphosphate. Enzymes catalyzing the interconversions include thiaminokinase, which utilizes thiamine and adenosine triphosphate (ATP) to form TPP and adenosine monophosphate (5′-AMP). A phosphoryl transferase that seems to be adenylate kinase uses a second ATP to interconvert TPP to adenosine diphosphate (ADP) and thiamine triphosphate. A membrane-associated triphosphatase is found mainly in nervous tissue.

TPP functions as the Mg^{2+}-coordinated coenzyme for the so-called active aldehyde transfers in two general types of reaction involving attack of carbonyl metabolites by the carbanion generated at C-2 on the thiazole moiety of TPP (Fig. 1). First, in decarboxylation of α-keto acids, the condensation of the thiazole moiety of TPP with the α-carbonyl carbon on the acid leads to the loss of CO_2 and production of a resonance-stabilized carbanion. Protonation and release of aldehyde occur in fermentative organisms such as yeast, which have only the TPP-dependent decarboxylase, but reaction of the α-hydroxalkyl–TPP with lipoyl residues and ultimate conversion of acyl–CoA occur in higher eukaryotes, including humans, with multienzymic dehydrogenase complexes. A lyase that is required to shorten 3-methyl-branched fatty acids, for example, phytanic acid, by α-oxidation also requires TPP.

Tab. 1 Coenzymatic forms and functions of B vitamins.

Coenzyme forms	*Enzyme systems*	*Functions*
	Thiamine (B_1)	
Thiamine pyrophosphate (perhaps triphosphate)	α-Keto acid decarboxylases, transketolase	Decarboxylations of α-keto acids from metabolism of carbohydrates and amino acids; glycolaldehyde transfers; nerve membrane ion transport
	Riboflavin (B_2)	
Flavin mononucleotide, flavin–adenine dinucleotide (covalent and noncovalent)	Oxidases, dehydrogenases	Hydrogen atom transfers from substrates to oxygen and to cytochromes
	Niacin	
Nicotinamide adenine dinucleotide and the 2′-phosphate	Reductases, dehydrogenases	Hydride ion transfer
	Vitamin B_6	
Pyridoxal 5′-phosphate, pyridoxamine 5′-phosphate	Aminotransferases, amino acid decarboxylases, amino acid dehydratases, cystathionine synthetase, phosphorylase, 5-amino levulinate synthetase, etc.	Transaminations, decarboxylations, and side chain, cleavages of amino acids; breakdown of glycogen; biosynthesis of heme, phospholipid, etc.
	Folacin	
Tetrahydrofolate and γ-glutamyl conjugates	Transformylases, thymidylate synthetase, hydroxymethyltransferases, formiminotransferase, etc.	One-carbon activation and transfer as in purine and pyrimidine biosynthesis; catabolism of serine and histidine, etc.
	Vitamin B_{12}	
5′-Deoxyadenosylcobalamin	L-Methylmalonyl–CoA mutase	Isomerization of L-methylmalonyl–CoA to succinyl–CoA
Methylcobalamin	5-Methyltetrahydrofolate–homocysteine methyltransferase	Methyl group transfer as in methionine biosynthesis
	Biotin	
1′-*N*-Carboxybiotin as amide-linked to ε-lysyl residues	Acyl–CoA carboxylases and other bicarbonate-dependent carboxylases	CO_2 activation and transfer in formation of acids
	Pantothenic Acid	
CoA	Acyl transferases, etc.	Fatty acid metabolism, etc.
4′-Phosphopantetheine	Acyl carrier protein of fatty acid synthetase	Fatty acid biosynthesis

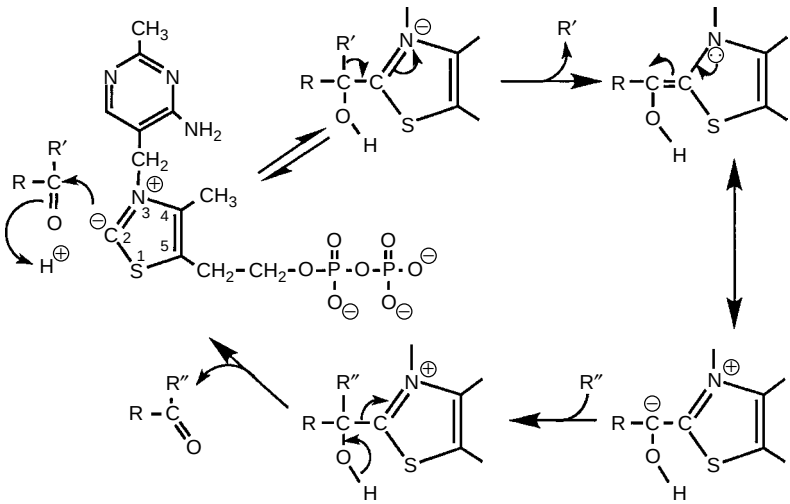

Fig. 1 Function of the thiazole moiety of TPP in α-keto acid decarboxylations (where R′ is a carboxylate lost as CO_2 and R″ is a proton generating an aldehyde) and in ketolations (where R′ is a carbohydrate moiety releasing a glycose and R″ is a different glycose generating a new carbohydrate).

The other general reaction involving TPP is the transformation of α-ketols (ketose phosphates). Although specialized phosphoketolases in certain bacteria and higher plants can split ketose phosphates to simpler, released products, the reaction of importance to humans and most animals is a transketolation. Transketolase is a TPP-dependent enzyme found in the cytosol of many tissues, especially liver and blood cells, where principal carbohydrate pathways exist. This enzyme catalyzes the reversible transfer of a glycoaldehyde moiety (α,β-dihydroxyethyl-TPP) from the first two carbons of a donor ketose phosphate to the aldehyde carbon of an aldose phosphate of the pentose phosphate pathway, which also supplies reduced nicotinamide adenine dinucleotide phosphate (NADPH) needed for biosynthetic reactions.

2.1.2 Flavocoenzymes

Structures of the natural flavocoenzymes, all containing a substituted tricyclic isoalloxazine nucleus derived from riboflavin, are given in Fig. 2. The monophosphate ester formed at the 5′-hydroxymethyl side chain substituent of riboflavin is a flavocoenzyme called flavin mononucleotide (FMN). It is the precursor to the more commonly occurring flavin adenine dinucleotide (FAD), which includes a pyrophosphoryl-linked 5′-AMP moiety extended from the phosphoryl function of FMN. In some less commonly encountered, but essential, forms of flavocoenzymes, there is covalent attachment of a peptidyl residue through an electronegative atom, usually in the 8α-position. Flavocoenzymes are relatively more stable in acid than in base and are photodecomposed, largely to lumichrome and lumiflavin, by cleavages at the ribityl side chain. In the natural oxidized (quinoid) forms, FMN and FAD are fluorescent (wavelengths of excitation and emittance, 450 and 530 nm, respectively) and are readily detected when unbound by protein; however, FAD is significantly quenched

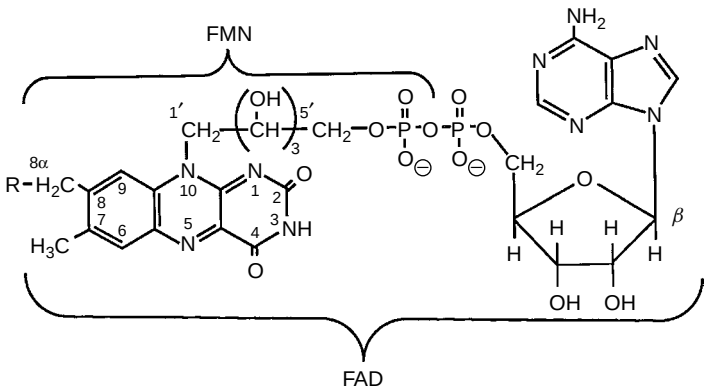

Fig. 2 Structure for principal flavocoenzymes where R = H for FAD or an electronegative atom (N, O, S) within a peptide side chain for covalently attached species.

by its intramolecular complex in solution. The observed oxidation–reduction potential near −0.2 V poises these coenzymes for electron transport, usually after (above) the more negative pyridine nucleotides and before (below) cytochromes.

Biosynthesis of flavocoenzymes occurs within the cellular cytoplasm of most tissues, but particularly in the small intestine, liver, heart, and kidney. The obligatory first step is the ATP-dependent phosphorylation of riboflavin catalyzed by Zn^{2+}-preferring flavokinase. The FMN product can be complexed with specific apoenzymes to form several functional flavoproteins, but the larger quantity is further converted to FAD in a second ATP-dependent reaction catalyzed by Mg^{2+}-preferring FAD synthetase. These coenzyme-forming steps are tightly regulated. FAD is the predominant flavocoenzyme present in tissues, where it is mainly complexed with numerous flavoprotein dehydrogenases and oxidases. Less than 10% of the FAD can also become covalently attached to specific amino acid residues of a few important apoenzymes. Examples include the 8α-*N*(3)-histidyl-FAD within succinate dehydrogenase and 8α-*S*-cysteinyl-FAD within monoamine oxidase, both of mitochondrial localization. Turnover of covalently attached flavocoenzymes requires intracellular proteolysis, and further degradation of the coenzymes involves a pyrophosphate cleavage of FAD to FMN and action by nonspecific phosphatases on the latter.

Flavocoenzymes participate in oxidation–reduction reactions in numerous metabolic pathways and in energy production via the respiratory chain. The redox functions of a flavocoenzyme (Fig. 3) include one-electron transfers during which the biologically encountered, neutral, oxidized quinone level of flavin is half-reduced to the radical semiquinone, which can exist within natural pH ranges as neutral or anionic species. A further electron can lead to a fully reduced hydroquinone. Additionally, a single-step, two-electron transfer from substrate to flavin can occur (Fig. 4). Such cases as hydride-ion transfer from reduced pyridine nucleotide or the carbanion generated by base abstraction of a substrate proton may lead to attack

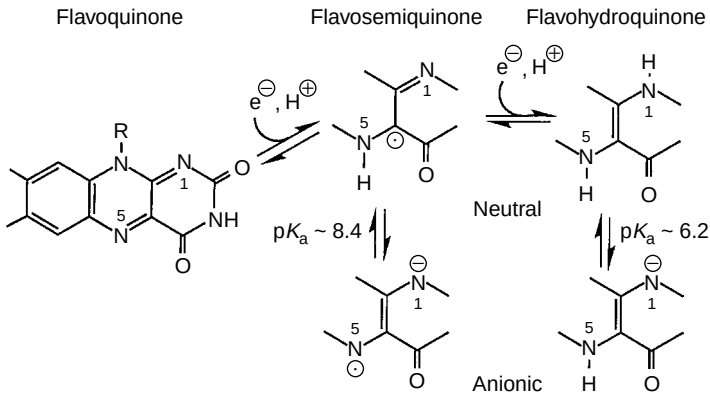

Fig. 3 Physiologically relevant redox states of flavocoenzymes with pK_a values estimated for interconversions of the free species.

Fig. 4 Reaction types encountered with flavoquinone coenzymes and natural nucleophiles (X^-).

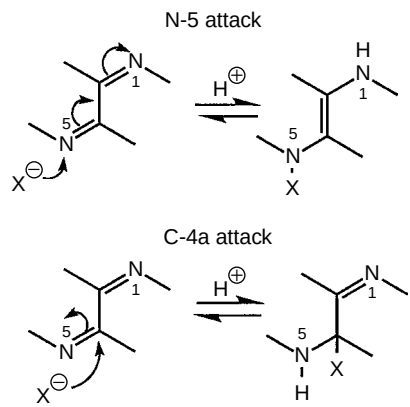

at the flavin at N-5; some nucleophiles such as the hydrogen peroxide anion add at C-4a.

There are flavoprotein-catalyzed dehydrogenations that are both pyridine nucleotide dependent and independent, reactions with sulfur-containing compounds, hydroxylations, oxidative decarboxylations, dioxygenations, and reduction of O_2 to hydrogen peroxide. The flavins are characterized by wide scope in their operation due to their intrinsic abilities to be varyingly potentiated as redox carriers upon differential binding to proteins, to participate in both one- and two-electron transfers, and, in reduced (1,5-dihydro) form, to react rapidly with oxygen.

2.1.3 Pyridine Nucleotide Coenzymes

Nicotinamide adenine dinucleotide and its phosphate (Fig. 5) are the two natural coenzymes derived from niacin. Both contain an N-1 substituted pyridine 3-carboxamide that is essential to function in redox reactions with a potential near −0.32 V. The oxidized coenzymes are labile to alkali, including nucleophilic addition at the para (4) position, whereas the reduced (1,4-dihydro) coenzymes are labile

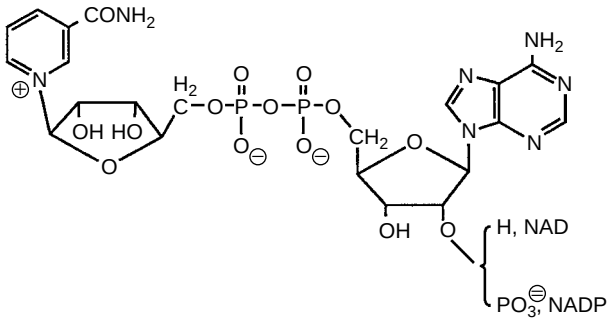

Fig. 5 Structures for pyridine nucleotide coenzymes.

to acid. Reduced coenzymes (NADH and NADPH) characteristically absorb light in the near-ultraviolet region (340 nm).

Converging pathways lead to the formation of NAD, a fraction of which is phosphorylated to NADP. Vitaminic precursors are converted to the coenzymes in blood cells, kidney, brain, and liver. Nicotinate and nicotinamide react with 5-phosphoribosyl-1α-pyrophosphate (PRPP) and nicotinic acid mononucleotide (NaMN) or nicotinamide mononucleotide (NMN), respectively. Additionally, in liver, quinolinate from catabolism of tryptophan is similarly converted, with concomitant decarboxylation to NaMN. Subsequent reactions with ATP to incorporate the 5′-adenylate portion yield NAD from NMN and deamido-NAD from NaMN. The deamido compound reacts with glutamine in a cytosolic-ATP-dependent step to yield NAD, glutamate, AMP, and pyrophosphate. In breakdown, NAD is hydrolyzed to NMN and the latter to nicotinamide, which, in turn, can be converted to nicotinate by a rather widespread microsomal deamidase. An NAD glycohydrolase (NADase) catalyzes hydrolysis of NAD to nicotinamide plus adenosine 5′-pyrophospho-5-ribose (ADPR). Some NAD glycohydrolases have the ability to transglycosidate (i.e. to transfer the ADPR moiety of NAD to acceptor macromolecules).

Numerous enzymes require the nicotinamide moiety within either NAD or NADP. Most of these oxidoreductases function as dehydrogenases and catalyze such diverse reactions as the conversion of alcohols (often sugars and polyols) to aldehydes or ketones, hemiacetals to lactones, aldehydes to acids, and certain amino acids to keto acids. The common mechanism of operation (generalized in Fig. 6) involves the stereospecific abstraction of a hydride ion from substrate, with para addition to one or the other side of C-4 in the pyridine ring of the nucleotide coenzyme. The second hydrogen of the substrate group oxidized is concomitantly removed as a proton and ultimately exchanged as a hydronium ion.

Most dehydrogenases utilizing NAD or NADP function reversibly. Glutamate dehydrogenase, for example, favors the oxidative direction, whereas others, such as glutathione reductase, catalyze preferential reduction. A further generality is that most NAD-dependent enzymes are involved in catabolic reactions, whereas NADP systems are more common to biosynthetic reactions. The additional function of NAD as a substrate for providing the ADPR moiety to modify macromolecules has

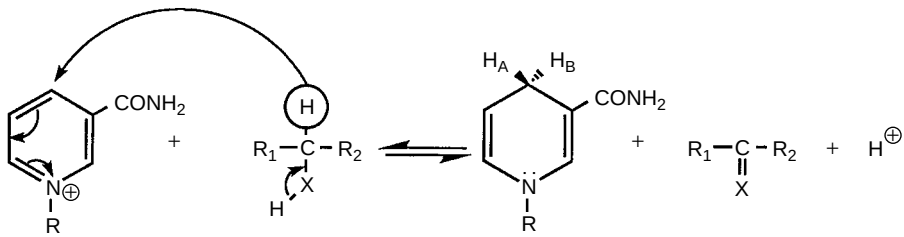

Fig. 6 The stereospecific hydride-ion transfer for operation of pyridine nucleotide coenzymes.

been recently more widely appreciated. ADP–ribosyl transferase catalyzes such a modification of the prokaryote elongation factor 2, thereby blocking translocation on ribosomes. Poly(ADPR) synthetases (polymerases) in eukaryotes catalyze a multiple addition of ADPR from NAD to form (ADPR) *n*-acceptor plus nicotinamide and hydrogen ion. This activity is found in mitochondria and bound to microsomes as well as in nuclei, where it affects the operation of DNA. This nonredox function of NAD probably accounts for the rapid turnover of NAD in human cells.

2.1.4 Pyridoxal 5′-phosphate and Pyridoxamine 5′-phosphate

Two of the three natural forms of vitamin B_6 (pyridoxine, pyridoxal, and pyridoxamine) can be phosphorylated to directly yield functional coenzymes; these are pyridoxal 5′-phosphate (PLP) and pyridoxamine 5′-phosphate (PMP). PLP is the predominant and more diversely functional coenzymic form, although PMP interconverts as coenzyme during transaminations. At physiological pH, the dianionic phosphates of these coenzymes exist as zwitterionic metaphenolate pyridinium compounds. They are very watersoluble, absorb light in the ultraviolet region, exhibit fluorescence, and, in general, are sensitive to light, particularly at alkaline pH. Both natural and synthetic carbonyl reagents (e.g. hydrazines and hydroxylamines) form Schiff bases with the 4-formyl function of PLP (and pyridoxal), thereby removing the coenzyme and inhibiting PLP-dependent reactions.

Each of the three vitamin-level compounds is phosphorylated in the cytosol by ATP-utilizing pyridoxal kinase, which, in mammalian tissues, prefers Zn^{2+}. Most cells of facultative and aerobic organisms contain a cytosolic, FMN-dependent pyridoxine (pyridoxamine) 5′-phosphate oxidase responsible for catalyzing the O_2-dependent conversion of pyridoxine 5′-phosphate (PNP) and PMP to PLP. During aminotransferase (transaminase)-catalyzed reactions, PLP and PMP interconvert with amino and keto functions of substrate–product participants. Release of free vitamin, mainly pyridoxal when physiological nonsaturating levels of vitamin are absorbed, occurs when the phosphates are hydrolyzed by nonspecific alkaline phosphatase located on the plasma membrane of cells.

PLP functions in numerous reactions that embrace the metabolism of proteins, carbohydrates, and lipids. Especially diverse are PLP-dependent enzymes that are involved in amino acid metabolism. By virtue of the ability of PLP to condense its 4-formyl substituent with an amine, usually the α-amino group of an amino acid, to form an azomethine (Schiff base) linkage, a conjugated double-bond system extending from the α-carbon of the amine (amino

acid) to the pyridinium nitrogen in PLP results in reduced electron density around the α-carbon. This reduction potentially weakens each of the bonds from the amine (amino acid) carbon to the adjoined functions (hydrogen, carboxyl, or side chain). A given apoenzyme then locks in a particular configuration of the coenzyme–substrate compound such that maximal overlap of the bond to be broken will occur with the resonant, coplanar, electron-withdrawing system of the coenzyme complex. These events are depicted in Fig. 7. Aminotransferases effect rupture of the α-hydrogen bond of an amino acid with ultimate formation of an α-keto acid and PMP; this reversible reaction provides an interface between amino acid metabolism and that for ketogenic and glucogenic reactions. Amino acid decarboxylases catalyze breakage of the α-carboxyl bond and lead to irreversible formation of amines, including several that are functional in nervous tissue (e.g. epinephrine, norepinephrine, serotonin, γ-aminobutyrate). The biosynthesis of heme depends on the early formation of δ-aminolevulinate from PLP-dependent condensation of glycine and succinyl–CoA followed by decarboxylation. There are many examples of enzymes, such as cysteine desulfhydrase and serine hydroxymethyltransferase, that affect the loss or transfer of amino acid side chains. PLP is the essential coenzyme for phosphorylase that catalyzes phosphorolysis of the α-1,4-linkages of glycogen. An important role in lipid metabolism is the PLP-dependent condensation of L-serine with palmitoyl–CoA to form 3-dehydrosphinganine, a precursor of sphingolipids.

2.1.5 Pterin Coenzymes

Among natural compounds with a pteridine nucleus, those most commonly encountered are derivatives of 2-amino-4-hydroxypteridines, which are trivially named pterins. Although a number of pterins when reduced to the 5,6,7,8-tetrahydro level function as coenzymes, the most generally utilized are poly-γ-glutamates of tetrahydrofolate (THF). These and their natural derivatives of tetrahydropteroylglutamates responsible for vectoring one-carbon units in different enzymic reactions are shown in abbreviated form in Fig. 8. All these pterins bear the substituent for transfer at N-5 or N-10 or are bridged between these basic centers. The number of glutamate residues varies, usually from one to seven, but a few to several glutamyls optimize binding of tetrahydrofolyl coenzymes to most enzymes requiring their function. Less commonly encountered,

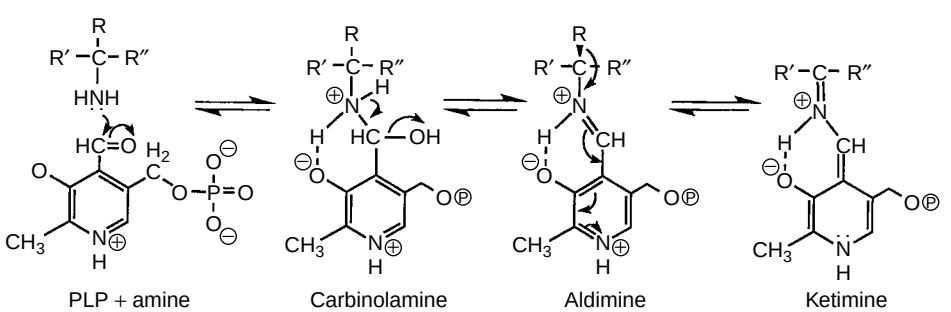

Fig. 7 Operation of PLP with a generalized amine.

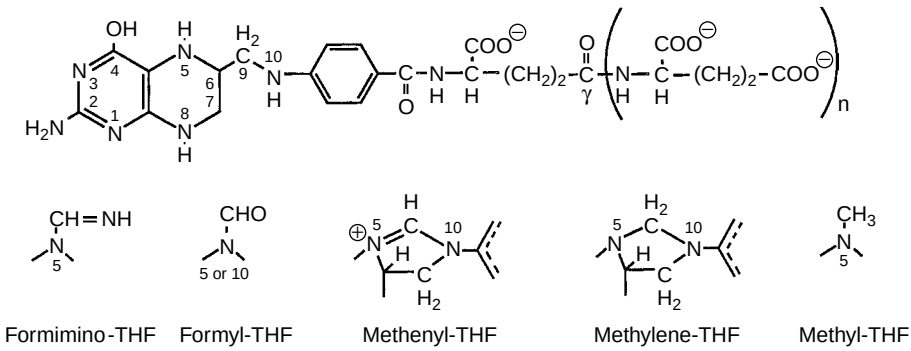

Fig. 8 Structures for the formimino, formyl, methenyl, methylene, and methyl derivatives of tetrahydropteroyl-L-glutamate.

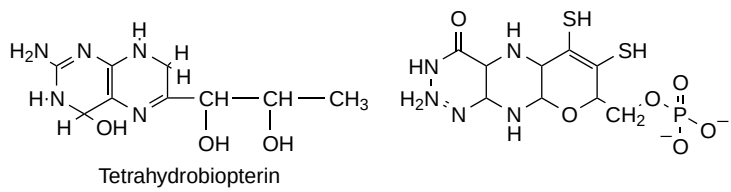

Fig. 9 Other pterin coenzymes: tetrahydrobiopterin of phenylalanine hydroxylase and the pterin cofactor of certain Mo/Fe flavoproteins.

but essential for some coenzymic roles of pterins, are the compounds shown in Fig. 9. Tetrahydrobiopterin cycles with its quinoid 7,8-dihydro form during O_2-dependent hydroxylation of such aromatic amino acids as in the conversion of phenylalanine to tyrosine. Tetrahydrobiopterin also serves as a one-electron donor in NO synthase catalysis. The recently elucidated molybdopyranopterin functions in some Mo/Fe flavoproteins such as xanthine dehydrogenase. Pterin coenzymes, most at the tetrahydro level, are sensitive to oxidation and have characteristic absorbance in ultraviolet light. Upon heating in aqueous media below pH 4, the pterin portion of the folyl-type coenzyme tends to split from the *p*-aminobenzoyl glutamate portion. The xanthine dehydrogenase pterin during isolation loses hydrogen from the side chain to generate a double bond between sulfur-bearing carbons.

The interconversions of folate with the initial coenzymic relatives, the tetrahydrofolyl polyglutamates, involve dihydrofolate reductase necessary for reducing the vitamin-level compound through 7,8-dihydro to 5,6,7,8-tetrahydro levels. This enzyme is the target of such inhibitory drugs as aminopterin and amethopterin (methotrexate). A similar dihydropterin reductase catalyzes reduction of dihydro- to tetrahydrobiopterin. Tetrahydrofolate is intracellularly trapped and extended to polyglutamate forms that operate with THF-dependent systems. In some cases (e.g. thymidylate synthetase), there is a redox change in tetrahydro to dihydro coenzyme, which is recycled by the

NADPH-dependent reductase. Turnover of coenzyme to folate at the monoglutamate level requires hydrolytic cleavage of the extra glutamyl residues. Cells of the small intestinal mucosa are especially rich in the γ-glutamyl peptidase ("conjugase") that degrades ingested natural folyl polyglutamates to the more readily absorbed folate.

Some of the major interconnections among one-carbon-bearing THF coenzymes and their metabolic origins and the roles are as follows:

1. Generation and utilization of formate.
2. *De novo* purine biosynthesis wherein glycinamide ribonucleotide and 5-amino-4-imidazole carboxamide ribonucleotide are transformylated by 5,10-methenyl-THF and 10-formyl-THF respectively.
3. Pyrimidine nucleotide biosynthesis, wherein deoxyuridylate and 5,10-methylene-THF form thymidylate and dihydrofolyl coenzyme.
4. Conversions of some amino acids, namely, *N*-formimino-L-glutamate (from histidine catabolism) with THF to L-glutamate and 5,10-methenyl-THF (via 5-formimino-THF), L-serine with THF to glycine and 5,10-methylene-THF, and L-homocysteine with 5-methyl-THF to L-methionine and regenerated THF.

2.1.6 B_{12} Coenzymes

The coenzyme B_{12} (CoB_{12}), known to function in most organisms, including humans, is 5′-deoxyadenosylcobalamin (Fig. 10). A second important coenzyme form is methylcobalamin (methyl-B_{12}), in which the methyl group replaces the deoxyadenosyl moiety of CoB_{12}. Some prokaryotes utilize other bases (e.g. adenine) in this position originally occupied by the cobalt-coordinated cyanide anion in cyanocobalamin, the initially isolated form of vitamin B_{12}. Coenzyme forms of B_{12} are light-absorbing, photolabile compounds that readily undergo photolysis to yield aquocobalamin (B_{12b}), in which H_2O is coordinated to cobalt in the corrin ring. Acid hydrolysis of CoB_{12} yields hydroxocobalamin (B_{12a}) with a coordinated hydroxyl ion.

The metabolic interconversions of vitamin B_{12} as the naturally occurring hydroxocobalamin (B_{12a}) with other vitamin- and coenzyme-level forms involves sequentially reduction of B_{12a} to the paramagnetic or radical B_{12r} and further to the very reactive B_{12s}. The latter reacts in enzyme-catalyzed nucleophilic displacements of tripolyphosphate from ATP to generate CoB_{12}, or of THF from 5-methyl-THF to generate methyl-B_{12}.

Seemingly CoB_{12}-dependent reactions proceed through a radical mechanism, and all but one (*Lactobacillus leichmanii* ribonucleotide reductase) involve a rearrangement of a vicinal group (X) and a hydrogen atom. This general mechanism is illustrated in Fig. 11. For the CoB_{12}-dependent mammalian enzyme, L-methyl-malonyl-CoA mutase, X is the CoA−S−CO−group, which moves with retention of configuration from the carboxyl-bearing carbon of LR-methylmalonyl–CoA to the carbon that lies beta to the carboxyl group in succinyl–CoA. This reaction is essential for funneling propionate to the tricarboxylic acid cycle. Without CoB_{12} (from vitamin B_{12}), more methylmalonate is excreted, but also the CoA ester competes with malonyl–CoA in normal fatty acid elongation to form instead abnormal, branched-chain fatty acids. Methyl-B_{12} is necessary in the transmethylase-catalyzed formation of L-methionine and regeneration of THF. Without this role, not only would there

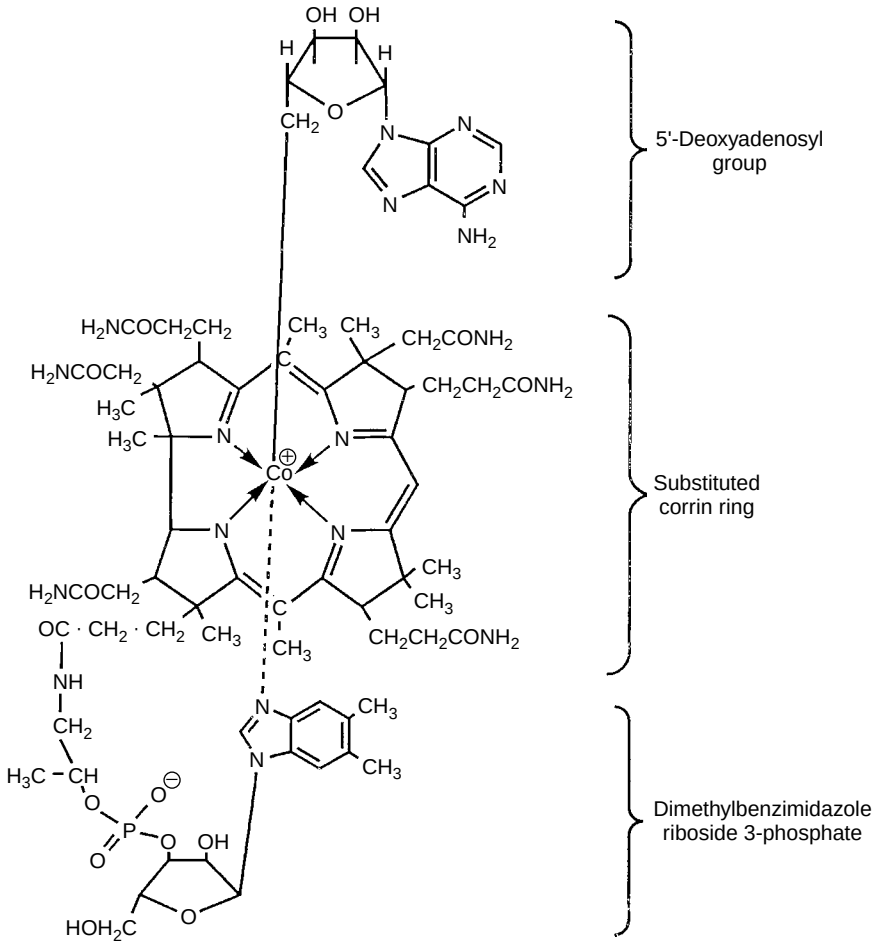

Fig. 10 Structure with identification of principal components for coenzyme B_{12}.

be impaired biosynthesis of the essential amino acid but increased exogenous supply of folate would be necessary to replenish THF, which would not otherwise be recovered from its 5-methyl derivative.

2.1.7 Biotinyl Functions

The coenzymic form of biotin occurs naturally only as the vitamin that is amide-linked to the ε-amino group of specific lysyl residues in carboxylases and transcarboxylase (Fig. 12). Though the two ureido nitrogens are essentially isoelectronic in the biotinyl moiety, the steric crowding of the thiolane side chain near N-3′ essentially prevents chemical additions, which therefore occur at N-1′.

Biotin is inserted into enzymes dependent on its operation by a holoenzyme synthetase that forms biotinyl 5′-adenylate as an intermediate from the vitamin and ATP. Usual proteolytic turnover of biotinylated enzymes release biocytin (ε-N-biotinyl-L-lysine), which is further hydrolyzed to

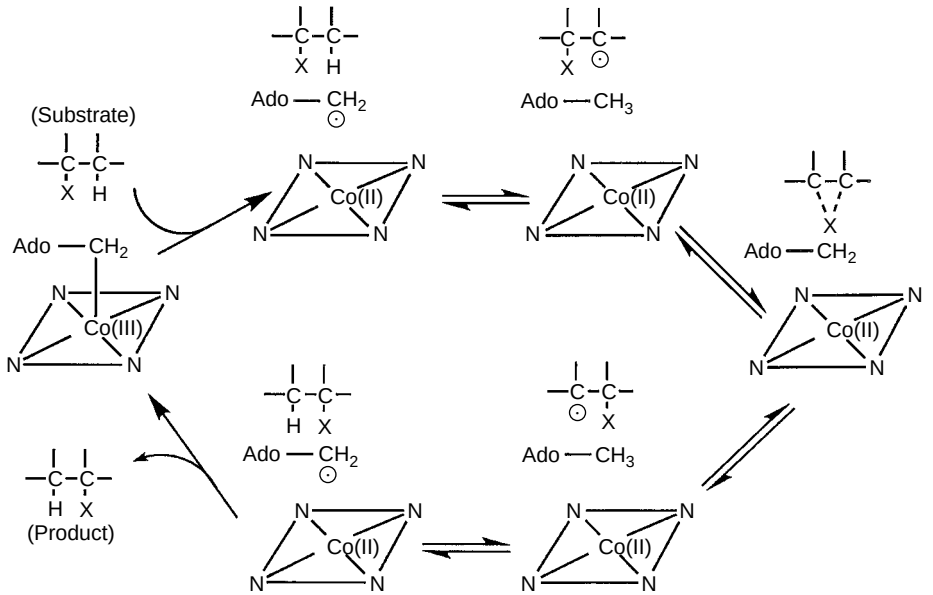

Fig. 11 Radical intermediates in Co-B_{12} enzyme reactions. Ado, adenosine.

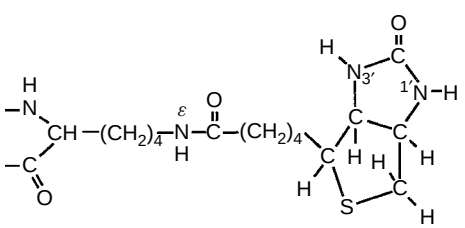

Fig. 12 Structure of the *d*-biotinyl-L-lysyl coenzyme moiety.

release the vitamin and amino acid in a reaction catalyzed by biotinidase (biocytinase, biotin amidohydrolase).

There are nine known biotin-dependent enzymes: six carboxylases, two decarboxylases, and a transcarboxylase. Of these, only four carboxylases are found in tissues of humans and other mammals. These carboxylases, named for their substrates, are (1) a cytosolic enzyme that converts acetyl–CoA to malonyl–CoA for fatty acid biosynthesis, (2) a mitochondrial enzyme that converts pyruvate to oxalacetate for citrate formation, (3) the enzyme for converting propionyl–CoA to D-methylmalonyl–CoA, and (4) the enzyme that carboxylates β-methylcrotonyl–CoA from L-leucine catabolism to form β-methylglutaconyl–CoA. Biotin-dependent carboxylases operate by a common mechanism (Fig. 13). This involves phosphorylation of bicarbonate by ATP to form carbonyl phosphate, followed by transfer of the carboxyl group from this electrophilic mixed-acid anhydride to the sterically less hindered and nucleophilically enhanced N-1′ of the biotinyl moiety. The resulting N(1)′-carboxybiotinyl enzyme can then exchange the carboxylate function with a reactive center in a substrate, typically at a carbon with incipient carbanion character.

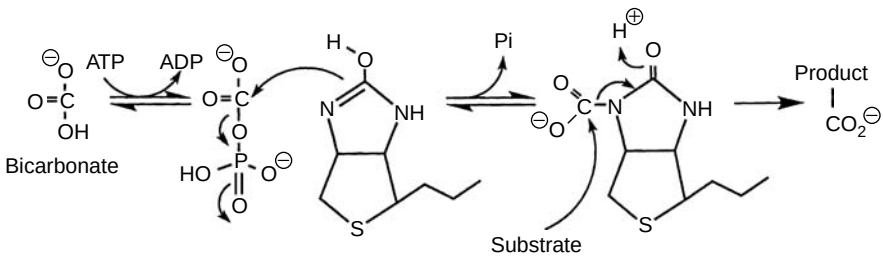

Fig. 13 Mechanism of carboxylations by biotin-dependent enzymes with the intermediacy of the putative carbonyl phosphate.

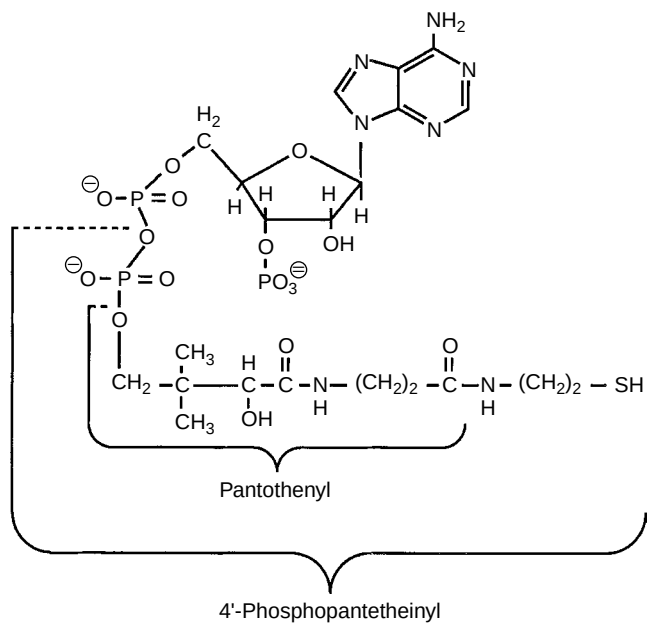

Fig. 14 Structure for coenzyme A with pantothenyl and 4′-phosphopantetheine as components.

2.1.8 Phosphopantetheinyl Coenzymes

The 4′-phosphopantetheinyl moiety, derived from the vitamin pantothenate and β-mercaptoethylamine, serves as a functional component within the structure of coenzyme A (Fig. 14) and as a prosthetic group covalently attached to a seryl residue of acyl carrier protein (ACP). Because of the thiol terminus with a pK_a near 9, phosphopantetheine and its coenzymic forms are readily oxidized to the catalytically inactive disulfides. CoA has a strong absorption maximum at 260 nm, attributable to the adenine moiety.

Within cells, synthesis of CoA and the 4′-phosphopantetheinyl moiety of ACP occurs via successive enzyme-catalyzed conversions from pantothenate to its 4′-phospho derivative to 4′-phosphopantothenyl-L-cysteine followed by decarboxylation to 4′-phosphopantetheine, which gains the AMP portion

to form dephospho–CoA. This latter is phosphorylated to form CoA. Following the ATP-driven steps in biosynthesis of CoA, formation of ACP occurs by transfer of the 4′-phosphopantetheinyl moiety of CoA, which binds via a phosphodiester link to apo-ACP in a reaction catalyzed by ACP holoprotein synthetase. About 80% of pantothenate in animal tissues is in CoA form, much of it as thioesters. The rest exists mainly as phosphopantetheine and phosphopantothenate. Cleavage enzymes catalyzing hydrolysis of the phosphate moieties (CoA → dephospho–CoA → 4′-phosphopantetheine → pantetheine) and release of β-mercaptoethylamine (cysteamine) and pantothenate from pantetheine operate during turnover and release of the vitamin, some of which is excreted in urine.

The myriad acyl thioesters of CoA are central to the metabolism of numerous compounds, especially lipids and the ultimate catabolic disposition of carbohydrates and ketogenic amino acids. The chemical properties of the thioester, which has a high group-transfer potential, permits facile acylations and hydrolysis; the ready formation of enolate ions and the carbanionlike character of the carbon α to the carbonyl facilitate condensation reactions. For example, acetyl–CoA, which is formed during metabolism of carbohydrates, fats, and some amino acids, can acetylate compounds such as choline and hexosamines to produce essential biochemicals; it can also condense with other metabolites such as oxalacetate to supply citrate, and it can lead to cholesterol. The reactive sulfhydryl termini of ACP provide exchange points for acetyl–CoA and malonyl–CoA. The ACP–S-malonyl thioester can chain-elongate during fatty acid biosynthesis in a synthetase complex.

2.2 Other Coenzymes

There are a growing number of coenzymes that are not formed from vitamins and hence can be biosynthesized by at least most of the organisms that require them. Sections 2.2.1 through 2.2.5 briefly describe several additional coenzyme types.

2.2.1 Pyruvoyl Functions

A simpler, but less frequently encountered variation on one way in which pyridoxal 5′-phosphate operates is provided by the pyruvoyl N-terminus of some enzymes. In these electrophilic centers, the amino function of a substrate amino acid condenses to form a ketimine, which facilitates loss of a carboxyl group from the attached amino acid moiety. For example, *S*-adenosylmethionine decarboxylases from mammals as well as from *Escherichia coli* use such a subsystem to form spermidine from putrescine and methionine.

2.2.2 Lipoyl Functions

α-Lipoic (thioctic) acid occurs naturally in amide linkage to the ε-amino group of lysyl residues within transacylases that are core protein subunits of α-keto acid dehydrogenase complexes. In such cases, the functional dithiolane ring is on an extended flexible arm. The lipoyl function mediates the transfer of the acyl group from α-hydroxylalkyl–TPP to CoA in a cyclic system that transiently generates the dihydrolipoyl residue.

2.2.3 Ubiquinones

The ubiquinones (coenzyme Qs) are a group of "ubiquitous" substituted benzoquinones with variable-length terpenoid chains. In eukaryotes, they are found

mainly in the mitochondrial inner membrane, where they function to accept electrons from several different dehydrogenases and relay them to the cytochrome system. In this process, the quinone function can be cyclically reduced and reoxidized. Evidence has been found for a function of CoQ in plasma membrane electron transport, which influences mammalian cell growth.

2.2.4 Pyrroloquinoline Quinone and Quinoproteins

Pyrroloquinoline quinone (PQQ) is a coenzyme originally isolated from bacteria possessing a methanol dehydrogenase. PQQ functions as a redox component of holoenzymes in which the coenzyme can be reduced in some cases by successive one-electron steps to the radical and dihydro forms and in other cases directly to the dihydro level by two-electron processes. More recent detailed examinations of several quinoproteins have elucidated other quinone/hydroquinone-related coenzymes as side chain-altered prosthetic groups. Examples include the 6-hydroxydopa (trihydroxyphenylalanyl) residue at the active site of serum amine oxidase and the tryptophan tryptophylquinone in a bacterial methylamine dehydrogenase.

2.2.5 Methanogen Coenzymes

A diverse group of coenzymes is required to convert carbon dioxide to methane. These rather novel compounds vary from a complex nickel-porphyrin–like F_{430} to the phenyl ether of methanofuran, the unusual pterin of tetrahydromethanopterin, and the deazaflavin of F_{420}, to the somewhat simpler sulfur-containing mercaptoheptanoylthreonine and coenzyme M. These are used by highly specialized Archaebacteria (known as *methanogens*) that can accomplish the reduction of CO_2 through formyl, methyl, methylene, and methyl stages of CH_4. The importance of the coenzymes that participate can be appreciated by the quantitative and global aspects of carbon cycling through methane-forming systems, which are not only free-living microbes but also are found in microbes in the human intestinal tract.

See also Enzymes, Energetics and Regulation in Biological Catalysis of; Vitamins, Structure and Function of.

Bibliography

Books and Reviews

Aurbach, G.D., McCormick, D.B. (Eds.) (1983–1991) *Vitamins and Hormones*, Vols. 40–46, Academic Press, New York.

Blakley, R.L., Benkovic, S.J. (1984) *Folates and Pterins*, Vol. 1,. Wiley, New York.

Blakley, R.L., Benkovic, S.J., Whitehead, V.M. (1984) *Folates and Pterins*, Vol. 3, Wiley, New York.

Chapman, S., Perham, R., Scrutton, N. (Eds.) (2002) *Flavins and Flavoproteins*, Rudolf Weber Agency for Scientific Publications, Berlin.

Curti, B., Ronchi, S., Zanetti, G. (Eds.) (1991) *Flavins and Flavoproteins*, de Gruyter, Berlin and New York.

Frey, P.A. (1988) Structure and function of coenzymes, in: Zubay, G. (Ed.) *Biochemistry*, Coord. Aw. Macmillan, New York.

Fukui, T., Kagamiyama, H., Soda, K., Wada, H. (Eds.) (1991) *Enzymes Dependent on Pyridoxal Phosphate and Other Carbonyl Compounds as Cofactors*, Pergamon Press, Oxford.

Iriarte, A., Kagan, H.M., Martinez-Carrion, M. (Eds.) (2000) *Biochemistry and Molecular Biology of Vitamin B_6 and PQQ-dependent Proteins*, Birkhauser-Verlag, Basel.

McCormick, D.B. (1997) Coenzymes, Biochemistry, in: Dulbecco, R. Editor-in-Chief, *Encyclopedia of Human biology*, Vol. 2, 2nd edition, Academic Press, San Diego, CA, pp. 847–964.

McCormick, D.B. (1999) Oxidation/reduction reactions, *Encyclopedia of Life Sciences*, Macmillan Ref. Ltd., London.

McCormick, D.B. (2001) Bioorganic mechanisms important to coenzyme function, in: Rucker, R.B., Suttie, J.W., McCormick, D.B., Macklin, L.J. (Eds.), *Handbook of Vitamins*, 3rd edition, Marcel Dekker, New York, pp. 199–212.

McCormick, D.B., Suttie, J.W., Wagner, C. (Eds.) (1997) Vitamins and Coenzymes, *Methods in Enzymology*, Vols. 279–282 (Parts I, J, K and L), Academic Press, San Diego, CA.

McCormick, D.B., Wright, L.D., Chytil, F., Suttie, J.W., Wagner, C. (Eds.) (1970–1997) Vitamins and Coenzymes, *Methods in Enzymology*, Parts A-L, Academic Press, New York.

McIntire, W.S. (1998) Newly discovered redox cofactors: possible nutritional, medical, and pharmacological relevance to higher animals, *Annu. Rev. Nutr.*, **18**, 145–177.

Rouviere, P.E., Wolfe, R.E. (1988) Novel biochemistry of methanogenesis (minireview), *J. Biol. Chem.* **263**, 7913–7916.

Origins of Life, Molecular Basis of

James P. Ferris
Department of Chemistry and NY Center for Studies on the Origins of Life, Rensselaer Polytechnic Institute, Troy, NY, USA

1 Introduction 3
1.1 Formation of the Solar System 3
1.2 The Nature of Life 4
1.3 The Earliest Life on Earth 4
1.4 Environments on the Early Earth 5

2 Reaction Conditions on the Primitive Earth 6

3 Possible Sources of Simple Organic Compounds 7
3.1 The Earth's Atmosphere 7
3.2 Interstellar Organic Molecules 7
3.3 Meteorites and Comets 8
3.4 Hydrothermal Systems 9

4 Formation of Complex Structures 10
4.1 Catalysis and Prebiotic Synthesis 11

5 Formation of Nucleic Acid Bases, Nucleosides, and Nucleotides 11
5.1 Nucleic Acid Bases 12
5.2 Nucleosides 12
5.2.1 Prebiotic Synthesis of Ribose 13
5.2.2 Prebiotic Synthesis of Nucleosides from Ribose and Bases 14
5.3 Nucleotides 14

6 Membranes and Vesicles 15

7 Chirality 17

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

8 Prebiotic Formation of Biopolymers 18
8.1 Polypeptides 18
8.2 Polynucleotides 19
8.3 Selectivity in Catalyzed Reactions 22
8.4 PNA as a Precursor to RNA 22

9 Polymer Reactions 23
9.1 Ribozymes 23
9.2 Template-directed Synthesis 23

10 Catalysis and the Origin of Life 24

Bibliography 24
Books and Reviews 24
Primary Literature 24

Keywords

Chiral Molecules
A molecules that is not superimposable on its mirror image. The two molecules (enantiomers) have identical physical properties with the exception of their interactions with other chiral molecules or with chiral radiation (e.g. plane or circularly polarized light). A racemic mixture contains equal amounts of each enantiomer.

Hydrothermal Systems
The circulation of seawater near the hot magma at a sea floor spreading to the center formed as a result of plate tectonics. This circulation results in heating of the water to 300–400 °C and the reduction of compounds in the seawater (e.g. $SO_4^{=}$ to S^{-}) and the deposition of metal sulfides and other reduced metal derivatives proximate to the spreading center.

Peptide Nucleic Acid (PNA)
A polymer composed of repeating β-amino acids with attached nucleic acid bases that are proposed to have been precursors to the RNA world. The nucleic-acid bases are attached to the polymer at the same distance apart as the bases in DNA and RNA so that PNA is able to form double helices with both these nucleic acids.

Ribozyme
An RNA that catalyzes reactions.

RNA World
A proposed early stage in the formation or evolution of primitive life in which RNA was the principal biopolymer. The RNA world is proposed to be the precursor to the DNA/protein world of contemporary life.

Template-directed Synthesis
The synthesis of a complementary RNA or DNA by condensation of mononucleotides on a preformed RNA or DNA template. The sequence of the synthesized chain is determined by Watson–Crick hydrogen bonding between the mononucleotides and the complementary bases on the template.

Vesicle or Liposome
An approximately spherical body with a wall that is composed of long chain hydrocarbons with polar terminal groups. It contains an aqueous phase inside which may contain organic and inorganic substances.

Research on the origins of life is an investigation of the reaction pathways that start from simple organic molecules and end with the complex macromolecular structures that organized into the first life. The understanding of this evolution of chemical complexity requires information from astronomy, geology, biology, chemistry, and physics. For example, astronomy, geology, and physics provide data on the formation and environment of the early Earth. Information on reaction processes that led to the first self-replicating system comes from the disciplines of chemistry and biology.

1 Introduction

1.1 Formation of the Solar System

The processes leading to the origins of life were initiated 13.7 billion years ago (bya) with the Big Bang. The Big Bang has been described as the simultaneous appearance of plasmas of primary particles everywhere in the Universe. Element formation began almost immediately with the appearance of the elements hydrogen, helium, lithium, and beryllium by the combination of protons and neutrons. The subsequent star formation resulted in a higher density and temperature of these elements, which enabled the fusion of three atoms of helium to form carbon. Once carbon formed, all the other major biogenic elements formed from it including N, O, and P up to ^{58}Fe occurred. Further fusion to the higher elements was energetically unfavorable and did not occur until stars used up all their H and He at which point the star collapsed catastrophically and then exploded into a supernova; this provided the energy to form the heavier elements as it blew them out into space during the explosion.

Some "carbon stars" with high abundances of carbon contribute large amounts of carbon into the interstellar medium when they explode. Interstellar dust clouds are composed of carbon and silicate dust. Chemistry takes place in these dust clouds,

driven mainly by cosmic rays, which enhances the chemical complexity of the carbon compounds. These dust clouds have lifetime of 10^7-10^8 years. They collapse due to the effect of external forces (e.g. nearby supernovae), and star and planet formation takes place. During the process of collapse, the bulk of the mass forms a protostar that gets sufficiently hot to initiate nuclear fusion. A protoplanetary disk forms around the protostar during the collapse process that is composed of mass that became a part of the star. Here a localized area of gravitational collapse occurred, which eventually led to the formation of the planets, asteroids, and comets. Further accretion of the planets still is taking place today but at a much slower rate than when the planets were formed.

Since there is a direct connection between the origin of life and the formation of stars and solar systems, research on the origins of life requires knowledge of interstellar and stellar processes. Thus, studies on the origins of life is a truly interdisciplinary area with important contributions by astrophysicists, planetary geologists, chemists, and biochemists working together to trace a coherent pathway to the first life starting from the Big Bang.

1.2 The Nature of Life

What is life? How would one recognize a very primitive form of our ancestors? There is no consensus as to the basic requirements of the first living system. Does it require metabolic machinery or will the requisite metabolites be available for direct use? Will this life require the equivalent of a cell membrane or could it have been an assemblage of polymers bound to a rock? The diversity of opinion in the scientific community is illustrated by an informal poll. Most scientists agree that the first life should be able to replicate and evolve, which implies the presence of informational biomolecules that can be copied. This may be possible for an assemblage of polymers bound to a rock. Robots or even astronauts may have difficulty recognizing primitive life forms on other planetary bodies. It will be even more difficult to recognize life that uses a biochemistry very different from the one present on Earth.

1.3 The Earliest Life on Earth

Contemporary life is the result of about 4 billion years of evolutionary development and undoubtedly differs appreciably from the first life on Earth. There is little hope of detecting the original life forms today because the environment of the Earth has changed and the inefficient early life forms could not have competed with the highly evolved life present on Earth today.

There is a small possibility that life exists elsewhere in our solar system. Water is of central importance to life on Earth and there are several bodies in our solar system where water has been detected. There is ample evidence of ice being present on Mars. Channels, characteristic of flowing water, and lakebeds have been identified here. Ice has also been detected below the surface of Mars. Thus, it is possible that simple life forms live below the surface of Mars today. The other possible locations for existence of life are Europa, Ganymede and Callisto, the moons of Jupiter. There is evidence for liquid water below the icy surfaces of these three moons, so there is also the potential for life there as well.

Fossils of microorganisms, believed to have been present on the Earth

3.4–3.5 billion years ago (bya), have been found in Africa and Australia. Exhaustive radioactive dating provides firm evidence for the age of the rocks in which these fossils were found. The morphology of these multicellular microfossils does not establish the biochemistry utilized by these life forms.

Indirect evidence suggests that life was present on Earth prior to the 3.5 bya microfossils. This evidence is in the carbon isotope ratios observed in old rock formations. When CO_2 is reduced to carbohydrates by photosynthesis, it generates a product that contains a higher $^{12}C/^{13}C$ ratio than was present before photosynthesis. This isotope signature is seen almost all the way back to 3.8 bya suggesting the presence of life forms that carried out photosynthesis.

1.4 Environments on the Early Earth

The possible presence of photosynthetic life on the Earth 3.8 bya indicates it was a complex life form and suggests that primitive life may have been present on Earth closer about 4 bya. These data are in conflict with lunar crater data that suggests that the moon and hence the Earth also underwent a massive bombardment 3.9 bya that might have killed all the life on Earth. If this is correct, then photosynthetic life must have arisen rapidly in 0.1–0.2 bya. Alternatively, if the 3.9-bya bombardment did not sterilize the Earth, then life may have arisen around 4.2–4.4 bya when the bombardments decreased in intensity and the Earth cooled enough for oceans to form, and was more hospitable for life.

The early rock record on the Earth is fragmentary but what is known is consistent about the presence of liquid water 3.8 bya. Sedimentary rocks, in the Isua Formation in Greenland, were present at that time, so it is likely that water was present. The absence of rocks formed before 3.8 bya and that had not undergone major thermal alteration by plate tectonics make it almost impossible to determine the nature of the primitive Earth in earlier times.

The geology of the Isua rock formation suggests that the temperature of the early Earth was generally in the 0–100 °C range. Less is known concerning the chemical composition of the atmosphere. Initially it was thought that the atmosphere was reducing, containing large amounts of methane, ammonia, and molecular hydrogen. Then it was recognized that methane and ammonia would have undergone rapid photochemical destruction by ultraviolet light and the hydrogen formed would have escaped from the top of the atmosphere. The probable absence of reduced organic compounds, like methane, is supported by measurements of the oxidation level of the Earth's crust and mantle from the present back to 3.8 bya. It was determined from the investigations of rocks of these former eras that the oxidation state of the crust and mantle is the same now as it was 3.8 bya. The high oxidation state of the crust today results in the emission of oxidized organics and inorganic compounds by volcanoes. The same oxidized gases might have been given off 3.9 bya because the oxidation state of the crust and mantle has not changed. Thus, it is unlikely that the reduced compounds necessary for the origin of life were present 3.8 bya. If life originated at an earlier time, say 4.2–4.4 bya, there exists the possibility that there was a reducing atmosphere at that time. If life originated then, it would have had to survive the increased impacts that took place 3.9 bya.

The current model for the primitive atmosphere is that it had large amounts of molecular nitrogen and carbon dioxide and much smaller amounts of reduced carbon and nitrogen compounds. This view is consistent with the predominantly carbon-dioxide atmospheres of Venus and Mars, the two planets closest to the Earth in the solar system. The molecular oxygen in the Earth's atmosphere is a direct result of photosynthesis, hence the presence of life.

No strong data exist for the chemical pathways leading to the origins of life. It has not been possible to do a meaningful chemical analysis of the fossilized microscopic life forms in ancient rocks because most of the organic material has been replaced or extensively metamorphosed. Since biological evolution proceeds from simpler to more complex life, it is assumed that the chemical evolution leading to the first life also proceeded from simpler molecules to more complex biological polymers.

The absence of data concerning the earliest life on Earth means that it will not be possible to prove that any laboratory model of the origins of life is an exact representation of the earliest life on Earth. Consequently, experiments on the origins of life will be considered to be successful only if systems can be developed that undergo replication, mutation, and catalysis, and may conceivably have been present on the primitive Earth. It should be noted that development of these replicating and mutating systems would be a major scientific accomplishment because chemical systems of this complexity have not been developed previously. Consequently, this research will not only provide insight into the origin of life but it will also generate new ways to synthesize complex organic structures using simple reactants.

2 Reaction Conditions on the Primitive Earth

The Earth underwent rapid change in the 4.5–4.0 bya time period. The crust, initially molten, solidified, and when it had cooled below 100 °C, the oceans formed. Internal heat was released through the crust by extensive volcanic eruptions, which discharged the gases trapped or chemically combined in the crust into the atmosphere. The atmosphere of the early Earth is generally believed to have consisted mainly of N_2, CO_2, and water vapor. Reduced gases, such as CH_4, NH_3, CO, and H_2, were present only if their rates of release from the crust exceeded their rates of photolysis by ultraviolet light. Oceans were probably slightly acidic owing to the carbonic acid formed by dissolution of atmospheric CO_2. Basic minerals and traces of NH_3 would have converted some of the carbonic acid to bicarbonate. The rate of precipitation of $CaCO_3$ and $MgCO_3$ determined how long the pH of the oceans was controlled by a carbonic acid–carbonate buffer system. The concentrations of metal ions in the oceans probably were different from those present today. The alkali metal ions, calcium and magnesium, were present in about the concentrations found today, and the absence of molecular oxygen in the atmosphere suggests that the soluble lower oxidation states of many of the transition metals were also present. In fact, the presence of Fe^{2+} is demonstrated by the banded iron formations in the ancient rock record. It is believed that these iron oxide deposits were caused mainly by the oxidation of Fe^{2+} to insoluble Fe^{3+} compounds by the molecular oxygen produced by photosynthesis.

Some of the insoluble minerals may have been different from those found

today. For example, sulfides and cyanide complexes may have formed in the absence of molecular oxygen. The mineral surfaces probably served as important catalysts for prebiotic chemical processes. The energy required to drive prebiotic processes came principally from solar radiation, but crustal radioactivity and internal heat from the Earth may have also been important. Solar energy manifested itself on the Earth in the form of UV and visible light, lightning discharges, and heat. The Earth's internal heat resulted in high crustal temperatures and frequent volcanic eruptions; magma injected into the crust under the sea generated hydrothermal systems.

The general reaction conditions, which are believed to have been present over much of the primitive Earth, are described above. Just as there exists special niches of unusual temperature, acidity, mineralogical, or atmospheric conditions on the present Earth, it is likely that similar niches were present on the primitive Earth. We cannot exclude the possibility that some important processes of chemical evolution were triggered in one of these special environments. For example, a heterotrophic form of life may have originated in one of the niches where in a photosynthetic autotrophy evolved. Once it had developed the capability to make its own nutrients, it may have been able to survive under the global conditions outside this special environment.

3 Possible Sources of Simple Organic Compounds

3.1 The Earth's Atmosphere

Stanley Miller's production of amino acids by passing an electric discharge through a mixture of methane, ammonia hydrogen, and water vapor in 1953 was a dramatic demonstration that it is possible to carry out laboratory studies of the origins of life. In this famous experiment, Miller generated hydrogen cyanide (HCN), aldehydes, ketones, and hydroxy acids, along with a large array of amino acids. It was recognized that the large proportions of methane, ammonia, and hydrogen present in the gas mixture he used were inconsistent with the rapid loss of these gases from the Earth's atmosphere. In studies of the action of electric discharges on gaseous mixtures that contain smaller percentages of reduced organics and N_2 gas, Miller observed that hydrogen cyanide was the principal reaction product. Nitrogen oxides (NO, NO_2) are the reaction products if no hydrogen or reduced carbon or nitrogen compounds were present in the sparked mixture. The lower yields of organic substances in these discharge experiments prompted the search for other sources of organic compounds.

3.2 Interstellar Organic Molecules

Interstellar organic molecules are believed to be formed by the action of cosmic rays and UV light on the carbonaceous material

Tab. 1 Some interstellar organic compounds.

Hydrocarbons
CH_4, C_2H_6, $CH_2{=}CH_2$, $HC{\equiv}CH$, $HC{\equiv}CC{\equiv}CH$, $CH_3C{\equiv}CH$, $CH_3C{\equiv}CC{\equiv}CH$
Carbon–Oxygen Compounds
$C{\equiv}O$, $H_2C{=}O$, $CH_3\underset{\underset{H}{\vert}}{C}{=}O$, $(CH_3)_2C{=}O$, $H\underset{\underset{OH}{\vert}}{C}{=}O$
Carbon–Nitrogen Compounds
$CH_3C{\equiv}N$, $HN{\equiv}C$, $HC{\equiv}N$, $HC{\equiv}CC{\equiv}N$, $HC{\equiv}CC{\equiv}CC{\equiv}N$, $H(C{\equiv}C)_5C{\equiv}N$

injected into space by the explosion of stars. The action of UV light is mainly effective in diffuse dust clouds while cosmic rays penetrate both dark and diffuse interstellar dust clouds. UV light is emitted by hot cores in dark clouds and is driven by the UV emission of a proximate protostar. The higher temperature of the protostar also warms the nearby dust grains and volatilizes the compounds condensed on them. Some of the volatile organic and inorganic molecules, radicals and ions identified by microwave spectroscopy are shown in Table 1.

3.3 Meteorites and Comets

The accretion of the Earth resulted from the impacts of smaller bodies on the Earth for several thousand years after it was formed. These bodies may have brought organics to the Earth's surface if they were traveling at a low velocity and thus were not pyrolyzed when the body on which they were carried hit the Earth's atmosphere.

Once the bulk of the planetesimals were "swept up" by the planets orbiting the Sun, the asteroids that orbited the Sun between Mars and Jupiter continued to serve as a source of meteorites. The asteroid belt exists because the gravitational perturbations of Jupiter are so great that it inhibits the coagulation of these smaller bodies into a planet. While the asteroids have defined orbits around the Sun, they tend to smash into each other–a process that results in their loss of dust and larger pieces with sufficient translational energy for these pieces to escape from their orbits and accelerate towards the Sun. Many of these bodies will impact with other planets, including the Earth, and bring organics to their surfaces. It is estimated that 80% of the original asteroid belt and an amount of material equivalent to a layer of material 0.2 to 1-km thick was deposited on the Earth as a result of these collisions. Accretion continues at a slower rate to this day with an estimated $30\,000 \pm 10\,000$ tons of dust per year reaching the Earth's surface.

Smaller meteorites survive impact with the Earth's atmosphere by the thermal ablation of their surface materials. This preserves the organics present below the surface of the meteorite from decomposing on impact with the atmosphere. These bodies may break into pieces when they hit the Earth's surface but the organics present are preserved. The much larger bodies survive the trip through the Earth's atmosphere but the impact is with such high energy that the body is almost totally pyrolyzed and the impacter gets so hot that any organics present are destroyed. The fate of each dust particle or larger body depends on its mass and velocity. Material originating from the asteroid belt usually has a lower velocity than those from comets and has a greater chance of survival. An impact on the atmosphere in Tunguska, Siberia (Russia) may have been due to cometary or meteorite material. The resultant shock wave blew down trees over a 2000 km^2 area.

Meteorites are our principal source of information about extraterrestrial material. Most meteorites do not contain much carbon but those that do, have as much as 2%. Analysis of the organic material present reveals that most of it is a polymeric matrix of organic material imbedded in an inorganic matrix. There is also a fraction of water-soluble organics, which accounts for less than 1% of the organic material present. This consists of diverse array of organics, many of which are reminiscent of biomolecules. These include

Tab. 2 Organic compounds detected in the Murchison meteorite[a,b].

Amino acids	++
Carboxylic acids	+++
Dicarboxylic acids	++
Hydroxy acids	++
Sulfonic acids	++++
Basic N-heterocycles	+
Purines & pyrimidines	+
Pyridine carboxylic acids	+
Amides	++
Amines	++
Alcohols	++
Aldehydes & ketones	++
Aliphatic hydrocarbons	++
Aromatic hydrocarbons	++
Sugar alcohols & acids	+

[a]++++ > 1000 ppm +++ > 100 ppm ++ > 10 ppm + > 1 ppm.
[b]Adapted from Pizzarello, S., Huang, Y., Becker, L., Poreda, R.J., Nieman, R.E., Cooper, G., Williams, M. (2001) The organic content of the Tagish lake meteorite, *Science* **293**, 2236–2240.

carboxylic and dicarboxylic acids amino acids, purines, pyrimidines, and simple sugar derivatives (Table 2).

There have been no direct analyses of comets although at the time of this writing there are several missions to comets are in progress. Spectroscopic analysis of the cometary coma, generated by the impact of the particles in the Sun's solar wind on the comet, revealed the presence of hydrocarbons and nitrile derivatives (Table 3). These findings give strong support to the hypothesis that comets contain reduced organics, some of which may have contributed to the inventory of organics on the Earth that led to the origins of life.

Tab. 3 Some carbon, hydrogen and nitrogen compounds in the comas of comets.

CH_4, C_2H_6, $HC{\equiv}CH$
$S{=}C{=}O$, $C{\equiv}O$, $H_2C{=}O$, CH_3OH
$HC{\equiv}N$, $HN{\equiv}C$, $CH_3C{\equiv}N$

3.4 Hydrothermal Systems

Marine hydrothermal systems with their spectacular "black smokers," "white smokers," and unique ecosystems may have provided yet another environment in which the biomolecules required for the origins of life were formed. The internal heat of the Earth drives these vast systems of seawater circulating through the Earth's crust. Temperature ranges as high as 350 °C and pressures up to 20–30 mPa (200–300 bar) occur as this hot water percolates through potential mineral catalysts.

The thermal gradient between the magma at the tectonic spreading center and the 2 °C sea water drives the flow of water and its dissolved compounds down to the magma where they are reduced by the iron in the magma at the high pressures and temperatures there. The reduced metal ions are vented together with the sulfide formed by the reduction of sulfate. The highly insoluble metal sulfides generate black and, sometimes white, precipitates in the vicinity of the vent as soon as they encounter the cold ocean water. It is possible that reduced organics and ammonia are formed from more oxidized precursors by the magma. Laboratory studies have shown the reduction of molecular nitrogen to ammonia when it was heated to 500–700 °C and pressures of 0.1 Gpa. This reaction is inhibited by water and is most efficient with a 1 : 1 ratio of water and reductant. Nitrogen oxides (NO_2^-, NO_3^-) are readily reduced to ammonia by iron sulfide minerals and hematite at 500 °C. Thus, hydrothermal systems could have been the

source of ammonia required for the formation of nitrogen-containing organics on the primitive Earth. It is not clear whether the rate of ammonia production exceeded its rate of photolytic destruction by solar UV light.

More complex structures are unlikely to survive the high temperatures and pressures present directly in the vents but they may be formed from the reduced metal ions and sulfides generated by the vents. But organics may also have been formed away from the high temperature areas at the vents in which they would have been stable. The temperature gradient that drives the vent process also drives the circulation of sea through the reduced forms of metals and their sulfides deposited in the vicinity of the vent. This "off axis" circulation may bring oxidized compounds in contact with the reducing conditions and catalytic minerals leading to the formation of reduced organic and inorganic compounds. These "off axis" compounds would not come in contact with the hot magma and consequently would not be destroyed by the high temperatures.

There have been several recent findings of the formation of reduced organics from oxidized precursors using this hydrothermal model. The laboratory synthesis of some simple organics from carbon dioxide and carbon monoxide has been reported in simulations of hydrothermal reaction conditions. Iron, iron sulfides together with hydrogen sulfide reduced carbon dioxide to methane thiol (CH_3SH). The reduction of CO at 100 °C with a mixture of iron and nickel sulfides in the presence of methane thiol resulted in the synthesis of acetic acid and its methyl thioester (CH_3COSCH_3). Reactions of CO with iron sulfide at 250 °C and high pressures give low yields (0.07%) of pyruvate. Dipeptides are formed in the reactions of amino acids with CO, a mixture of CO and a mixture of iron and nickel sulfides, hydrogen sulfide, or methane thiol at 100 °C. These data suggest the possibility of forming simple peptides near hydrothermal vents but it appears unlikely that polypeptides of sufficient lengths to be catalysts could have formed by this route because they are hydrolyzed under similar reaction conditions.

4
Formation of Complex Structures

None of the compounds formed from the sources described in Sect. 3 are biopolymers of the type present in life as we know it. The concentrations of these reactants leading to biopolymers probably were so low that bimolecular reactions did not occur in the solution phase. Rather, it is thought that the reactions took place on mineral surfaces or in complexes with metal ions that selectively adsorbed and catalyzed the reactions of these organics.

If the molecules discussed in Sect. 3 are indeed the limit of complexity, then additional reactions had to have occurred to generate the biopolymers essential for the first life. Life is based on biopolymers or the self-organization of more complex structures as is observed in the formation vesicles from lipid molecules. The formation of more complex structures may have been driven by UV light, discharges, radioactivity, shock waves, and thermal processes on the early Earth. All of these energy sources also have the potential to degrade complex organics, so there may have existed a critical balance between the rates of formation and degradation of the molecules required for the first life.

4.1
Catalysis and Prebiotic Synthesis

It is likely that catalysis had a central role in the origins of life just as it has a central role in life today. For example, a complex array of structures and enzymes are required for the biosynthesis of proteins and nucleic acids. Certainly a much simpler synthesis of these biopolymers occurred on the primitive Earth, but it is highly likely that mineral and/or metal ion catalysis was involved. A catalyst is essential since it can carry out the following functions:

1. Selectively bind low concentrations of reactant molecules from a complex mix of organic structures present in an aqueous solution on the primitive Earth.
2. Enhance the rate of the reaction of the reactants so that they can form bonds faster than they would be degraded.
3. Selectively form a restricted number of reaction products so that not all possible isomers are formed. This is especially important in the formation of biopolymers in which if it is was desired to make one molecule of a polypeptide containing 100 amino acids of defined sequence by a random method, it would be necessary to synthesize 10^{130} molecules. If two identical molecules of a 40 mer of an RNA containing four different bases were required, it would be necessary to synthesize 10^{48} molecules.

Most of the studies of the origins of life do not consider the role of catalysis in prebiotic synthesis. Bernal first proposed in 1947 that mineral catalysis might have had a role in prebiotic synthesis. In the absence of life, minerals and metal ions were the most likely catalysts on the primitive Earth. For example, divalent metal ions were required in the synthesis of nucleosides from purines and ribose. The uranyl ion (UO_2^{2-}) and lead ion (Pb^{2+}) catalyzed the oligomerization of 5′-activated nucleotides. Mg^{2+} is usually an essential cation in the reactions resulting in the formation of the phosphodiester bond present in oligonucleotides. Montmorillonite clay has a central role in the formation of oligopeptides from aminoacyl adenylates (Sect. 8.1) and in the formation of oligonucleotides (Sect. 8.2). The minerals illite and hydroxylapatite catalyze the oligomerization of negatively charged amino acids (Sect. 8.1). Hydrotalcite catalyzes the oligomerization of glycolaldehyde phosphate to the dimer erythrose-2,4-diphosphate as the major reaction product. Additional examples of catalysis are discussed in Sect. 8.

The current scenario for the origins of life proposes a series of chemical reactions yielding first simple products, then more complicated compounds. The reaction conditions for each stage of the process may very well have been quite different, and they probably required different catalysts. The chemical evolutionary process may have paralleled the changing reaction conditions on the primitive Earth that resulted from its cooling and the diminished flux of impacting comets and meteorites. The host of variables in the possible reaction conditions makes it almost impossible to define the precise conditions under which life originated on the early Earth.

5
Formation of Nucleic Acid Bases, Nucleosides, and Nucleotides

The central biopolymers of life today are proteins, DNA, and RNA. Amino acids, the building blocks of protein, are

found in meteorites, and formed directly in many prebiotic simulation reactions. The monomers for DNA and RNA are not found either in meteorites or in simulation experiments, so if they were present on the primitive Earth, they should have been synthesized by reactions that occurred here.

The prebiotic synthesis of RNA will be emphasized in this review. Proteins will only be discussed briefly because their ability to store information has not been established. If primitive life did not have the ability to maintain the information that led to its formation, it would not have evolved.

Many scientists in this field feel that RNA monomers are too complex to have formed spontaneously on the primitive Earth. They propose that simpler precursors to RNA formed first which directed the synthesis of RNA. While possible pre-RNAs have been proposed, no prebiotic synthesis of these pre-RNAs has been demonstrated (Sect. 8.4). In addition, unless the pre-RNAs were able to direct the synthesis of RNA monomers, the problem of formation of the RNA monomers still remained if the transition of pre-RNAs to RNAs took place using the pre-RNAs as templates. Consequently, the formation of RNA monomers and their polymerization will be emphasized.

5.1 Nucleic Acid Bases

Nucleic acid bases have been found in the Murchison meteorite, which fell in Murchison Australia in 1969, and other meteorites containing carbon compounds. Condensation reactions of HCN are another source of purine and pyrimidine bases. The condensation of HCN proceeds in a stepwise fashion in aqueous solution and generates a tetramer, diaminomaleonitrile (DAMN). Further reaction of DAMN with HCN generates a mixture of low molecular weight oligomers. Hydrolytic decomposition of the oligomer yields adenine, guanine, orotic acid, uracil, 4,5-dihydroxypyrimidine, and 5-hydroxyuracil. More recent studies with a solution of ammonium cyanide left at $-78\,^{\circ}C$ for 27 years generated a polymer, which on hydrolysis, resulted in the formation of adenine, guanine, hypoxanthine, xanthine, 2,6-diaminopurine, uracil, 5-aminouracil, 5-hydroxyuracil, orotic acid, 5-aminoorotic acid, and 4,5-dihydroxypyrimidine (Fig. 1).

Alternatively, photolysis of DAMN generates 4-aminoimidazole-5-carbonitrile (AICN). Reaction of AICN or its corresponding amide (AICA) with simple nitrile derivatives results in the formation of an array of purine bases including adenine, guanine, diaminopurine, hypoxanthine, and xanthine. The pyrimidines cytosine, uracil, and diaminopyrimidine are formed by addition reactions to cyanoacetylene. Heating formamide to $160\,^{\circ}C$ in the presence of mineral catalysts also forms adenine and cytosine.

Acid hydrolysis of the Murchison meteorite, which has 2–3% carbon, results in the release of adenine, hypoxanthine, guanine, xanthine, and uracil. The yields are low but there is the potential of delivery of substantial amounts of these compounds on the Earth from dust and meteorites that get broken off the asteroids (Sect. 3.3).

5.2 Nucleosides

There are two problems associated with the prebiotic synthesis of nucleosides. One is

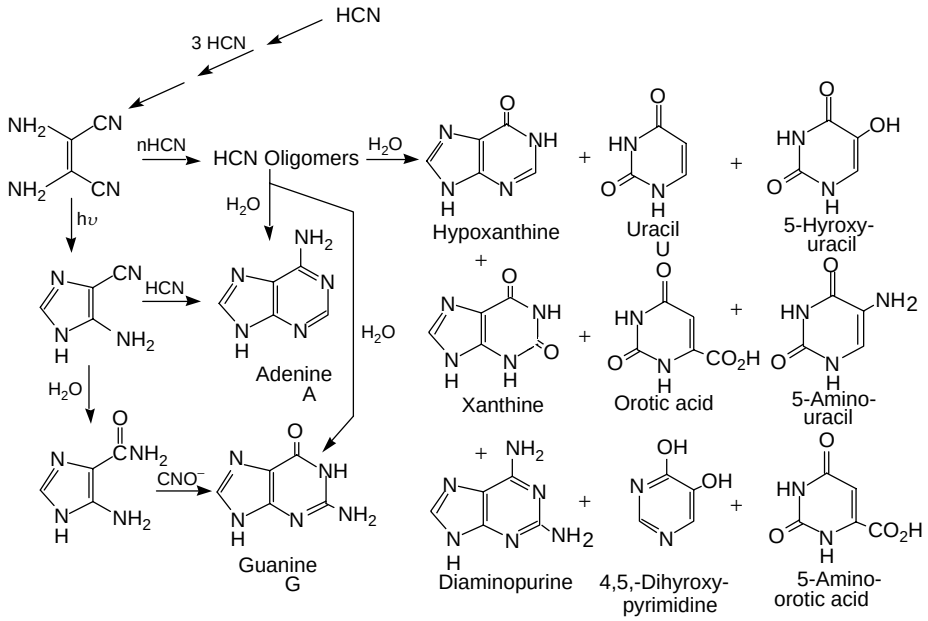

Fig. 1 Formation of purine and pyrimidine bases from HCN.

the formation of ribose and the second is the formation of nucleosides by the reaction of ribose and with a purine or pyrimidine base.

5.2.1 Prebiotic Synthesis of Ribose

Several routes have been proposed for prebiotic ribose synthesis. The first was based on the Formose reaction, the condensation of formaldehyde in basic solution catalyzed by divalent metal ions. This reaction, however, produces an array of more than 35 sugars, most in low yield. It is not clear how the small amount of ribose in this complex mixture was selected from all the other sugars present to form nucleosides.

Two improvements of this synthesis have been reported. One is the use of a magnesium ion–lead ion mixture that catalyzes the formation of all four pentoses in roughly equal amounts with little or no formation of any other sugars. This catalytic mixture decreases the number of products formed from over 20 to four. Interestingly, incubation of any one of these pentoses with the magnesium–lead catalyst results in its being converted to equal amounts of the four pentoses. In addition, these pentoses are formed by incubation of hexoses with the magnesium–lead catalyst.

A promising synthesis of ribose is from the reaction of formaldehyde with calcium hydroxide in the presence of borate. The calcium catalyzes the synthesis of sugars via the Formose reaction while the borate forms a complex with the hydroxyl groups of the ribose and ribose precursors that prevents its conversion to hexoses and higher molecular weight sugars as well as inhibiting the formation of brown decomposition products.

Ribose may also be produced by the reaction of glycolaldehyde phosphate with

formaldehyde (Fig. 2). Further reaction of glyceraldehyde phosphate with glycolaldehyde results in ribose-2,4-diphosphate. The overall yield is 33% in a one-pot reaction starting from glycolaldehyde phosphate and formaldehyde. The main problem with this scenario is the absence of a convincing prebiotic route to glycolaldehyde phosphate. Glycolaldehyde can be converted to glycolaldehyde phosphate by reaction with amidotriphosphate (NH_2Pi_3).

The Murchison meteorite contains sugar derivatives in which the aldehyde group of the sugar is either oxidized to an acid or reduced to an alcohol. One speculative route to these products is the formation of ribose in the asteroid and the subsequent oxidation and reduction of the ribose in the meteorite or its parent asteroid. If this speculation is correct, there may be meteorites in which this oxidation did not occur.

5.2.2 Prebiotic Synthesis of Nucleosides from Ribose and Bases

There has been limited success in the formation of nucleosides by the reaction of purine and pyrimidine bases with ribose. The dry-phase reaction of purines and ribose at 100 °C in the presence of Mg^{2+} or Ca^{2+} results in low yields of the natural β-anomers of the nucleotides (Table 4). No nucleosides were detected in comparable reactions using pyrimidine bases. This is one of the reasons cited for suggesting that compounds with simpler structures were the building blocks for the first genetic material. So far, it has not been possible to find a building block simpler than the RNA monomer that is formed by a plausible prebiotic reaction (Sect. 9).

5.3 Nucleotides

The addition of the phosphate grouping to nucleosides proceeds efficiently in the absence of water at 100 °C in the presence of urea and ammonium phosphate. The phosphorylation of uridine (Fig. 3) yields a mixture of nucleotides including 43% of 5′-UMP. The problem here is that a rather specific combination of reagents ($NH_4H_2PO_4$ and urea) is required for this phosphorylation to take place. Other

Tab. 4 Proposed prebiotic syntheses of purine nucleoside.

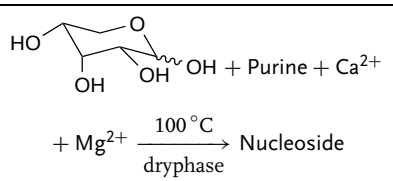

Component	Nucleoside (%)	
	α-nucleoside	β-nucleoside
Adenosine	1–2	2.3
Guanoisine		5
Xanthosine		2.1
Inosine	3	8

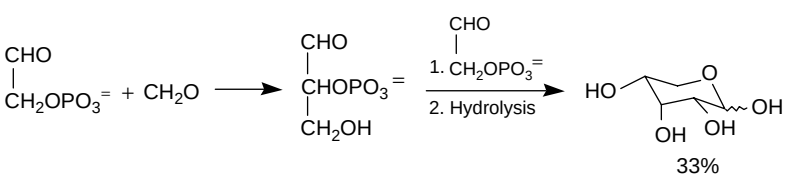

Fig. 2 A proposed prebiotic ribose synthesis.

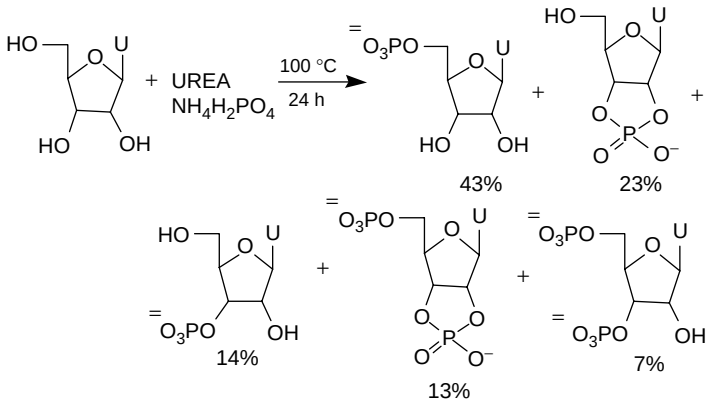

Fig. 3 Phosphorylation of uridine.

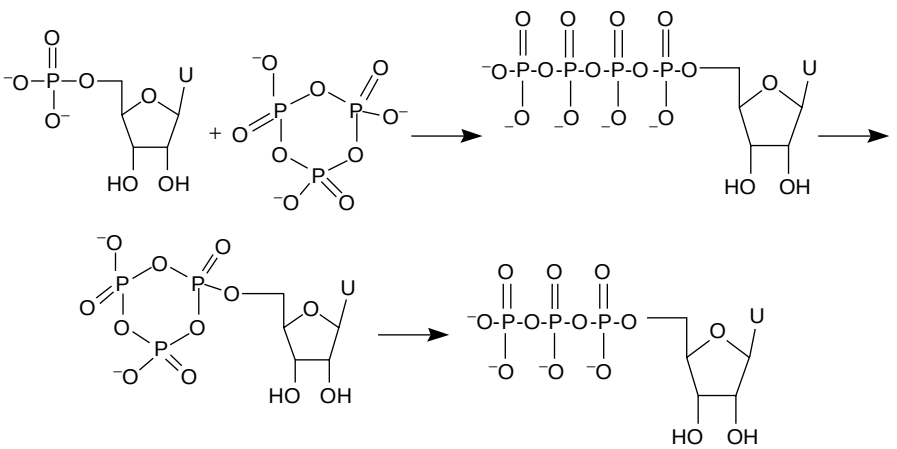

Fig. 4 A possible prebiotic route to ATP.

phosphate salts are much less effective in generating nucleotides from nucleosides.

Activated phosphates with pyrophosphate bonds may have been prepared by the reaction of 5′-nucleotides with trimetaphosphate (Fig. 4). The initial product is a tetraphosphate, which is hydrolyzed to the corresponding triphosphate. It is not clear whether 5′-nucleoside triphosphates were important in the first life or were the products of more sophisticated metabolic pathways that developed after life had evolved.

6 Membranes and Vesicles

Most scenarios for the first life propose the presence of an interacting group of biological polymers that is separated from the surrounding environment by a semipermeable membrane. The membrane is required to keep the interacting polymers and their metabolites in close proximity and to protect them from a hostile external environment. The membrane used in contemporary cells is made up of

phospholipids, but there are no convincing prebiotic syntheses of these materials. Linear fatty acids containing a minimum of 10–12 carbon atoms will form stable spherical vesicles. Here the pH is adjusted so that half the acids are in their anionic carboxylate form and the others are present as the acids (Fig. 5). Hydrogen bonding between the carboxylate and acid groups stabilizes these vesicles. Branched-chain fatty acids and hydrocarbons are formed in electric discharge reactions of methane and water. It is unlikely that branched acids would associate closely enough in a phospholipid to form a stable bilayer.

Linear carboxylic acids are formed in the reaction of carbon monoxide with iron at high temperatures and pressures under conditions reminiscent of those present in hydrothermal systems (Sect. 3.4). In studies on the formation of vesicles from linear carboxylic acids, it was observed that montmorillonite clay catalyzes vesicle formation. This is the same montmorillonite that catalyzes the formation of RNA from activated RNA monomers (Sect. 8.2). The vesicles formed catalytically trap some of the montmorillonite on the inside of the vesicles. When RNA is bound to the montmorillonite, it still catalyzes the formation of vesicles; a finding that suggests that this may have been a way of placing a catalytic system within a vesicle that has the capability of generating genetic material in a cell-like structure. It has not been determined at the time of writing whether it is possible to form RNA inside the vesicle by montmorillonite catalysis.

The compounds in the liquid phase within the vesicle are protected from dilution and interaction with reactive organic compounds in the environment. The problem is that the nutrients essential for life may not be able to diffuse into the vesicle and waste material may not be able to diffuse out. Vesicles formed from phospholipids containing carboxylic acids with ten carbon atoms are "leaky" in that they let materials go inside but they will also let them go outside as well. Vesicles containing phospholipids with 14 carbons are more selective in that they allow small compounds to pass through the cell wall but do not allow the polymers. For example, it was shown that vesicles containing phospholipids with 14 carbon atoms with encapsulated enzyme polynucleotide phosphorylase (PNPase) allowed the passage of adenosine diphosphate (ADP) into the vesicle and the PNPase catalyzed the conversion of ADP to oligomers of A. At the same time, the vesicles prevented the diffusion of a proteases enzyme through the bilayer that would destroy the PNPase. This experiment suggests that, if it were

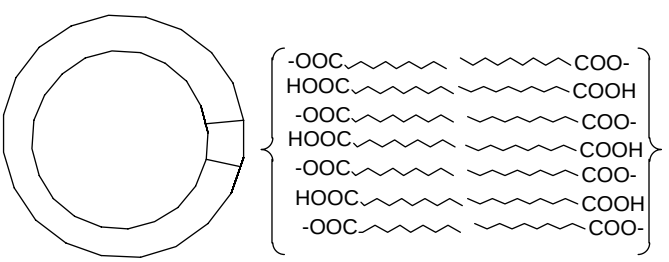

(a) Vesicle cross-section (b) Section of vesicle wall

Fig. 5 Schematic diagram of a vesicle.

possible for activated RNA monomers to diffuse through a vesicle, it would be possible to generate RNAs by clay mineral catalysis, which would either remain bound to the encapsulated montmorillonite or be too big to escape though the leaks in the vesicle.

7 Chirality

The molecules present in living systems consist of one of the two possible mirror-image forms called *enantiomers*. L-amino acids and D-nucleotides are the basic building blocks of proteins and nucleic acids respectively in contemporary life. The mirror-image D-amino acids and L-nucleotides occur rarely in contemporary life.

It is not known how the enantiomers observed in living systems were selected initially. The *frozen accident* theory proposes that a chance event resulted in the survival of the life bearing L-amino acids and D-nucleotides, and these amino acids and nucleotides have been associated with life ever since. The frozen accident theory assumes that life consisting of D-amino acids and L-nucleotides also existed at one time, and this chance event (a mutation?) resulted in the demise of this life form or the accelerated growth of the other. Scientists have been searching for an environmental explanation of the observed handedness that has investigated possible chiral forces as the source of the configuration of the amino acids and nucleotides of life. To date, there are no convincing experiments or explanations relating to this observation, and the origin of chirality remains an open question.

Initially it was believed that all the amino acids extracted from the Murchison meteorite contained equal amounts of both the D-, and L-enantiomers. The D- or L-enantiomers of amino acids interconvert very slowly (Fig. 6a). This interconversion is very small over time scale of hundreds of years but it is appreciable over thousands of years. So it was not surprising that a mixture of equal amounts of both D- and L-enantiomers were found in the Murchison meteorite that contains materials over 4 billion years old. Later, it was observed that some of the nonbiological amino acids in Murchison contained small excesses of the L-enantiomers. The particular amino acids had two substituents attached to the carbon atom bearing the amino and carboxyl groups (Fig. 6b, R and CH_3). This substitution would greatly inhibit the loss of their L-configuration for a time much longer than would a regular biological amino acid.

One of these amino acids, L-isovaline (Fig. 6b) catalyzes condensation of glycolaldehyde to form an excess of the D-tetrose, threose. Thus, it is possible that amino acids with an additional substituent at the carbon bearing the amino and carboxyl groups initiated the formation of

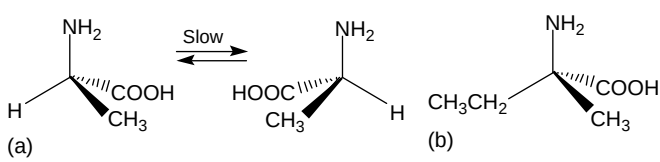

Fig. 6 (a) D-, L-enantiomers of alanine and (b) Isovaline, an amino acid that does not racemize.

D-sugars, the handedness of the sugar moiety in nucleic acids.

The observation of circularly polarized light in an area of star formation suggested that this might be the type of environment in which chiral selection occurred. Circularly polarized light can be either right- or left-handed. Laboratory studies have demonstrated that circularly polarized light will preferentially destroy one of the enantiomers in a D-, L-mixture faster than the other, thus leaving behind an excess of one of the enantiomers. This finding provides a possible explanation for the observation of an excess of one of the amino acid enantiomers in the Murchison meteorite. Since many of the theories and experiments to explain the origins of chirality put forth in the past 50 years have been shown to be incorrect, there is a reasonable chance that this proposal will not stand the test of time either.

In another approach to this problem, it was demonstrated that the montmorillonite catalysis of the reactions of D, L-activated mononucleotides generates more homochiral than heterochiral oligomers. This is not a claim of chiral selectivity. It does suggest the possibility of formation of oligomers composed mainly of the same but opposite handedness. This could have resulted in the formation of two different forms of life with opposite handednesses in their biopolymers. If this were the case, then it is assumed that a chance event led to one form predominating over the other.

8
Prebiotic Formation of Biopolymers

The interactions of biopolymers with themselves and with small molecules are an essential part of the biochemistry of living systems. The research outlined in Sect. 3 has uncovered potential prebiotic routes to monomers that may have been precursors to proteins and nucleic acids. Although it has not been possible to propose plausible, detailed chemical routes to all the necessary monomers, the findings suggest a number of possible pathways by which some of them may have formed on the primitive Earth. Since the question of the prebiotic formation of monomers has been solved in principle, if not in practice, there has been a shift in research focus to the problem of the prebiotic formation of biopolymers.

It is likely that on the primitive Earth, biopolymers formed in the presence of water, the principal liquid present. It is not possible to link amino acids to polypeptides or mononucleotides to nucleic acids in the presence of water. Additional energy in the form of activating groups or condensing agents is required. The activated monomers tend to hydrolyze in the presence of water, so the continuous formation of the high-energy compounds would have been essential for the formation of polypeptides and nucleic acids. An alternative scenario is for the polymerization to have occurred in the absence of water but that probably required an alternating wet and dry environment.

8.1
Polypeptides

Amino acids are formed in prebiotic simulation reactions and are found in carbonaceous meteorites, so it is reasonable to believe that they were present on the primitive Earth. Since polypeptides are not genetic (information storing) molecules, they may not have played a central role

in the first life because they are unable to reproduce themselves. There is the possibility that mineral or metal ion catalysts directed the continuous formation of polypeptides that in turn catalyzed the synthesis and/or replication of genetic polymers. If the mineral catalysts had sufficient selectivity, this could have resulted in the continuous formation of the same catalytic polypeptides.

Initial studies on prebiotic polymerization focused on the formation of polypeptides from amino acids. Most of these early studies invoked the use of heat to drive the polymerization, and anhydrous reaction conditions were used. The product mixtures from these reactions contain thermal polymers with molecular weights as high as 4000. Amide (peptide) and other bonds linked these amino acids together. The materials were shown to possess some weak catalytic activity.

Another synthesis of polypeptides involves the condensation of aminoacyl adenylates on montmorillonite to form the polypeptide adducts of 5′-adenylic acid (Fig. 7). This reaction proceeds in aqueous solution to form oligomers of up to 40 amino acids in length. Although it is difficult to envisage the formation of appreciable amounts of such a high-energy acyl anhydride in the presence of water, the experiment does suggest that other less reactive amino acid derivatives may condense to polypeptides on mineral surfaces.

Polypeptides are formed in aqueous bicarbonate buffer by the action of N',N'-carbonyldiimidazole (CDI) on amino acids. The reaction proceeds via the carboxyanhydride (Fig. 8) to give an 8 to 10 mer, which precipitates from solution. Since CDI is unstable in water, it should have been produced continually on the primitive Earth. These studies do show that condensing agents capable of generating N-carboxyanhydrides may have initiated the formation of polypeptides on the primitive Earth.

Reaction of glutamic acid with N',N'-carbonyldiimidazole is catalyzed by the clay mineral illite. Oligomers greater than 50 mers in length were obtained in a feeding reaction where a mixture of glutamate and CDI were added 50 times to the illite. Aspartic acid, another negatively charged amino acid, also formed polymers when a carbodiimide was used as the condensing agent and hydroxylapatite was the mineral surface on which the reaction took place.

8.2 Polynucleotides

The polymerization of mononucleotides to polynucleotides under plausible prebiotic

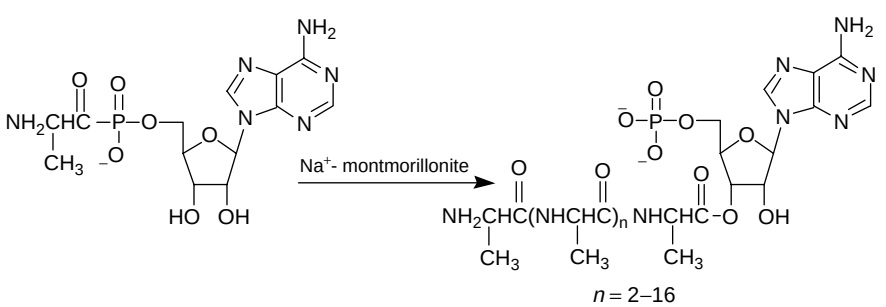

Fig. 7 Polymerization of aminoacyl adenylates on montmorillonite.

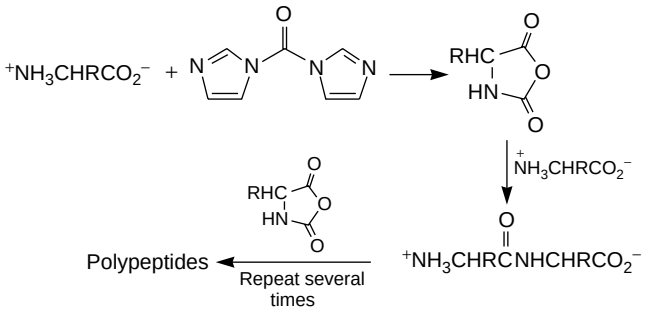

Fig. 8 Polypeptide formation in water using N',N'-carbonyl diimidazole carbonyl as the condensing agent.

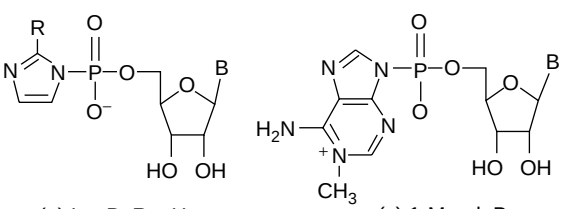

(a) ImpB, R = H
(b) 2-MeImpB, R = CH_3
(c) 1-MeadpB

Fig. 9 Activated RNA mononucleotides.

conditions has been reported. Initial studies in which the 2′,3′-cyclic phosphates of nucleotides were heated under anhydrous conditions in the presence of ethylenediamine resulted in the formation of polymers containing up to 13 monomer units (a 13 mer). This synthesis is inhibited by water, hence could have proceeded only in dry areas on the primitive Earth. In addition, there have been no prebiotic syntheses of the requisite diamines.

The oligomerization of the nucleoside 5′-phosphorimidazolides (5′-ImpN) (Fig. 9a) proceeds in aqueous solution in the presence of catalysts. Oligomers as long as the hexamers linked by 2′,5′-phosphodiester bonds were observed in reactions of 5′-ImpN in aqueous buffer catalyzed by Pb^{2+}. A major synthetic advance was made using uranyl ion (UO_2^{2+}) as the catalyst. 2′,5′-linked oligomers of adenylic acid as long as the hexadecamer were formed at room temperature in water. The reaction has been extended to the synthesis of oligomers of cytidylic and uridylic acid.

Syntheses of mainly 3′,5′-linked oligomers were observed in the reaction of the 5′-phosphorimidazolide of adenosine (Fig. 9a, B = adenine) in aqueous solution in the presence of the clay mineral montmorillonite. Oligomers as long as 10 mers were obtained which have a 2 : 1 ratio of 3′,5′ to 2′,5′ links. The oligomers formed in this reaction also contain some $A^{5'}$ppA groupings. Cyclic dimers are also formed. This reaction has also been extended to cytidylic and uridylic acids. The proportion of 2′,5′-linkages is greater with oligo(C)s and oligo(U)s than it is for oligo(A)s. The

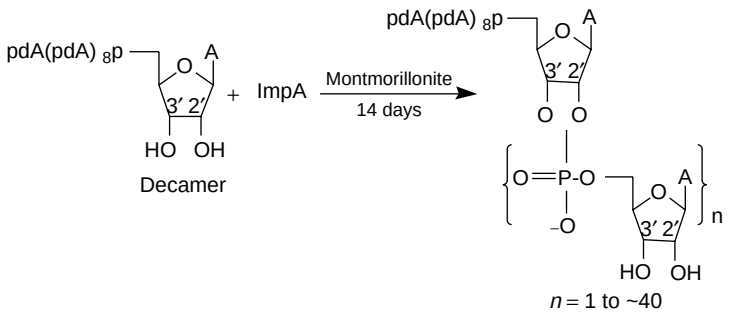

Fig. 10 Elongation reaction in which ImpA is added daily to a decameric oligomer.

oligo(C)s formed by the reaction of the 5′-phosphorimidazolide of cytidine (Fig. 9a, B = cytosine) on montmorillonite serve as templates for the formation of oligo(G)s (Sect. 9.2) showing that these short RNAs are capable of storing genetic information.

The RNA oligomers formed in these reactions were not long enough to initiate the processes of replication and catalysis expected in simple living systems, but this research constituted the first step in that direction. The next stage of this montmorillonite-catalyzed synthesis was the observation that it was possible to make RNA oligomers as long as 50 mers in a "feeding" reaction. In this approach, an activated monomer was added daily for 14 days to a decameric primer bound to the montmorillonite (Fig. 10). The decamer elongated to a 40 to 50 mer in the two-week time period.

The current status of this synthesis is the formation of 40–50 mers in one day using a 1-methyladenine activating group (Fig. 9c) in place of imidazole (Fig. 11). The advantage of this synthesis is that it provides a facile route to these longer oligomers so that they can be tested for their ability to bind and catalyze reactions; a property consistent with the proposed RNA world.

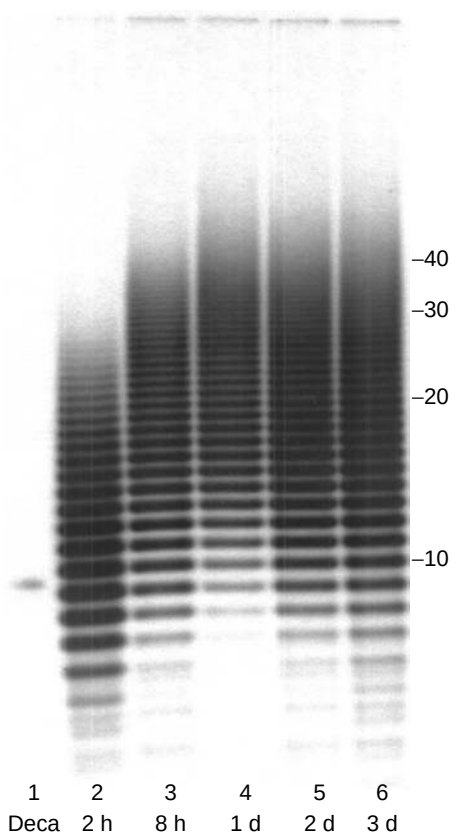

Fig. 11 Formation with time of oligomers formed from MeadpU (Figure 9c) on montmorillonite. Lane 1 10 mer marker; lanes 2, 3 h; 4, 8 h; 5, 1 day; 6, 2 days with additional MeadpA added; 7, 3 days with 2 additions of MeadpA added.

8.3 Selectivity in Catalyzed Reactions

Montmorillonite, like most catalysts, accomplishes the observed catalytic effect by lowering the activation energy of specific reaction pathways. Since the catalyst does not lower the activation energy of all reaction pathways, it produces a limited number of isomers of the products formed. It was recognized early in the studies on prebiotic synthesis that it would be difficult to form long polymers of amino acids or nucleotides because of the large number of possible isomers that would be formed in the random synthesis. Studies on the uncatalyzed synthesis of RNA oligomers from one-molar solutions of the imidazole-activated monomers (Fig. 9) yielded mainly dimers accompanied by trace amounts of higher molecular weight products. The successful synthesis of 40 mers using only a few milligrams of starting material and a montmorillonite catalyst clearly shows that the montmorillonite is generating a limited number of isomers. This was confirmed in a separate study of the formation dimers to pentamers using activated monomers of A and C. It was determined that in the pentamer fraction, there were four main products produced in yields of 4.3–13% while a random synthesis would have generated 512 products in 0.2% yield for each. Thus, catalysis solves the problem of producing too many isomers so that it is possible to make long oligomers.

8.4 PNA as a Precursor to RNA

The RNA world postulate suggests that RNA was the most important biopolymer in the first life on Earth. The RNA world is a generally accepted concept because it has been demonstrated that RNA has catalytic activity as well as the ability to store genetic information. The problem many scientists have with this scenario is that it assumes that RNA formed from RNA monomers that were formed by prebiotic processes on the primitive Earth. It has been noted previously that the complexity of the RNA structure and the failure, to date, to discover a plausible prebiotic pathway to RNA monomers has resulted in the search for alternative genetic molecules that have the potential for evolving into an RNA.

One of the frequently cited alternatives to RNA is peptide nucleic acid (PNA) (Fig. 12). The advantages of this structure are that it has no chiral centers and the distance between the bases is comparable to those in DNA and RNA. These factors make it possible for PNA to form double helical complexes with either D- and L-DNA or RNA. In addition, it is possible for PNA to serve as a template for the formation of RNA demonstrating that the transition from life based on PNA to RNA

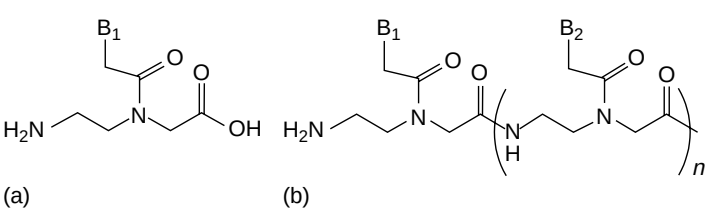

Fig. 12 PNA (a) monomer and (b) polymer.

life was possible if the synthesis of the RNA monomers had been accomplished.

There are major problems with PNA as a precursor to RNA. So far, there is no reported prebiotic synthesis of PNA monomers or polymers, and it degrades spontaneously in water. It appears unlikely that it will be possible to form polymers of PNA starting from the monomer because of competing cyclization reactions and RNA monomers are needed for the transition from PNA to RNA. The latter problem is the same complaint concerning the prebiotic synthesis of RNA so it will have to be assumed that a biosynthesis of RNA monomers occurred during the PNA world before the RNA world arose.

9 Polymer Reactions

The ultimate goal of research into the origins of life is the formation of polymers that replicate and catalyze enough reactions to maintain the viability of the reaction system that generated the first life. In the RNA world scenario, oligomers catalyze their own replication and their formation from simpler precursors in the absence of proteins. The replication process provides the means of information storage, which is not possible with proteins. The discovery of the limited catalytic action of RNA suggests that protein catalysis (enzymes) is not required. The absence of a requirement for protein makes the postulated RNA world much simpler than one requiring the concurrent prebiotic synthesis of both DNA and protein.

9.1 Ribozymes

Ribozymes, the RNAs that catalyze chemical reactions, are central to a scenario for primitive life on Earth based solely on RNA. The possible role of ribozymes in the RNA world was demonstrated by the catalytic elongation of a pentamer of cytidylic acid to a series of oligomers as long as the 30 mer by a ribozyme obtained from the protozoan Tetrahymena. This experimental finding suggests that oligomers produced in prebiotic processes described in Sect. 8.2 may have been elongated by the ribozymes of the RNA world. In addition, over twenty ribozymes that catalyze the reactions of biomolecules as well as structures that have no known connection with biological systems are known.

9.2 Template-directed Synthesis

The non-enzymatic template-directed synthesis of complementary RNA chains would have been an important process of information storage and RNA proliferation in the proposed RNA world. The synthesis of oligo(G)s (oligomers of guanylic acid) proceeds very efficiently on poly(C) templates with 2-methylImpG (Fig. 9, $R - CH_3$, base = G) as the starting material. The reaction occurs because the hydrogen bonding and stacking interactions of the guanine bases on the poly(C) template orient the monomer units for reaction. Oligo(A)s is formed less effectively in the reaction of ImpA on poly(U). One explanation for the low efficiency of oligo(A) formation is the weaker hydrogen-bonding interaction between A and U. No oligomers are formed in the reactions of the phosphorimidazolides of pyrimidine nucleosides on purine templates.

The failure of the pyrimidine nucleotide reactions is attributed to the virtual absence of stacking interactions between the pyrimidine nucleotide monomers bound to the poly-purine nucleotide chain. The

absence of the proximate association of pyrimidine monomers on the template means the monomers do not react to form phosphodiester links between the pyrimidine nucleotides. The failure to replicate pyrimidine monomers is also believed to explain why copolymers, containing equal amounts of purine and pyrimidine nucleotides, do not serve as templates for RNA synthesis. So far, the failure to discover conditions for the template-directed synthesis of oligomers containing both purine and pyrimidine bases is a major deficiency in the theory of the RNA world scenario.

10
Catalysis and the Origin of Life

Catalysis by substances such as minerals and metal ions on the primitive Earth was essential for the initiation of the first life on Earth. Studies have demonstrated that a clay mineral will catalyze the synthesis of an RNA containing up to 40 to 50 monomer units. The same activated monomers yield oligomers containing only two or three monomers units in the absence of the catalyst. The RNAs that form have a limited number of sequences and not all possible isomers are formed. This suggests that catalysis could result in the continual generation of an ensemble of the same structures that could have served as the starting point for the origins of life. It has been recently shown that the same montmorillonite that catalyzes the formation of RNA also catalyzes vesicle formation. Some of this catalyst with bound RNA is also encapsulated in the vesicle. This finding suggests that a few key catalysts may be all that was needed for the origin of life to occur. The search for catalysts that may have catalyzed the reactions of potential biomolecules on the primitive Earth should be emphasized in future studies on the origins of life.

See also Chirality in Biology; Genetic Analysis of Populations; Immunoassays; Paleontology, Molecular.

Bibliography

Books and Reviews

Brack, A. (Ed.) (1998) *The Molecular Origins of Life*, Cambridge University Press, Cambridge, UK.

Eschenmoser, A., Loewenthal, E. (1992) Chemistry of potentially pre-biological natural products. *Chem. Soc. Rev.* 1–16.

Ferris, J.P., Usher, D.A. (1988) Origins of Life, in: Zubay, G. (Ed.) *Biochemistry*, 2nd edition, Macmillan, New York, pp. 1191–1241.

Fry, I. (2000) *The Emergence of Life on Earth*, Rutgers University Press, New Brunswick, NJ.

Holland, H.D. (1984) *The Chemical Evolution of the Atmosphere and Oceans*, Princeton University Press, Princeton, NJ.

Palyi, G., Zucchi, C., Caglioti, L. (Eds.) (2002) *Fundamentals of Life*, Elsevier, Paris.

Schopf, J.W. (2002) *Life's Origins*, University of California Press, Berkeley, CA.

Stevenson, D.J. (1990) in: Newsom H.E., Jones J.H. (Eds.) *In Origin of the Earth*, Oxford University Press, New York.

Wilson, D.S., Szostak, J.W., Richardson, C.C., Abelson, J.N., Raetz, C.R.H., Thorner, J.W. (1999) *In vitro* selection of fundamental nucleic acids, *Annu. Rev. Biochem.*, **68**, 611–647.

Zubay, G. (2000) *Origins of Life on the Earth and in the Cosmos*, Academic Press, San Diego, CA.

Primary Literature

Böhler, C., Nielsen, P.E., Orgel, L.E. (1995) Template switching between PNA and RNA oligonucleotides, *Nature* **376**, 578–581.

Bonner, W.A., (1988) Origins of chiral homogeneity in nature, in: Eliel E.L.,

Wilen S.H. (Eds.) *Topics in Stereochemistry*, Wiley, New York, pp. 1–96.

Boynton, W.V. (2002) Distribution of hydrogen in the near surface of Mars: evidence for subsurface ice deposits, *Science* **297**, 81–85.

Brack, A., Ehler, K.W., Orgel, L.E. (1976) N′,N′-carbonyldiimidazole-induced diketopiperazine formation in aqueous solution in the presence of adenosine-5′-monophosphate, *J. Mol. Evol.* **8**, 307–310.

Brandes, J.A., Boctor, N.Z., Cody, G.D., Cooper, B.A. (1998) Abiotic nitrogen reduction on the early earth, *Nature* **395**, 365–367.

Braterman, P.S., Arrhenius, G., Pitsch, S., Bauder, A., Mueller, F., Eschenmoser, A. (1995) Preferential uptake of ammonium ions by zinc ferrocyanide, *Origin Life Evol. Biosphere* **25**, 531–538.

Canfield, D.E. (1998) A new model for Proterozoic ocean chemistry, *Nature* **396**, 450–453.

Cech, T.R. (1986) RNA as an enzyme, *Sci. Am.* **255**, 64–75.

Chakrabarti, A., Breaker, R.R., Joyce, G.F., Deamer, D.W. (1987) Production of RNA by a polymerase protein encapsulated within phospholipid vesicles, *J. Mol. Evol.* **39**, 555–559.

Cody, G.D., Boctor, N.Z., Filley, T.R., Hazen, R.M., Scott, J.H., Sharma, A., Yoder, H.S. Jr. (2000) Primordial carbonylated iron-sulfur compounds and the synthesis of pyruvate, *Science* **289**, 1337–1340.

Cooper, G., Kimmich, N., Belisle, W., Sarlnana, J., Brabham, K., Garrel, L. (2001) Carbonaceous meteorites as a source of sugar-related organic compounds for the early earth, *Nature* **414**, 879–883.

Cronin, J.R., Pizzarello, S. (1997) Enantomeric excesses in meteoritic amino acids, *Science* **275**, 951–955.

Deamer, D.W. (1997) The first living systems: a bioenergetic perspective, *Microbiol. Mol. Rev.* **61**, 239–261.

Decker, P., Schweer, H., Pohlmann, R. (1982) Biods X. Identification of formose sugars, presumable prebiotic metabolites, using capillary gas chromatography/gas chromatography-mass spectrometry of *n*-butoxime trifluoroacetates on OV-225., *J. Chromatogr.* **244**, 282–291.

Delano, J. (2001) Redox History of the Earth's Interior since ~3900 Ma: implications for prebiotic molecules, *Origins Life Evol. Biosphere* **31**, 311–341.

Ferris, J.P. (2002) Montmorillonite catalysis of 30–50 mer oligonucleotides: laboratory demonstration of potential steps in the origin of the RNA world., *Origins Life Evol. Biosphere* **32**, 311–332.

Ferris, J.P., Ertem, G. (1992) Oligomerization reactions of ribonucleotides on montmorillonite: reaction of the 5′-phosphorimidazolide of adenosine, *Science* **257**, 1387–1389.

Ferris, J.P., Hagan, W.J. Jr., (1984) HCN and chemical evolution: the possible role of cyano compounds in prebiotic synthesis, *Tetrahedron* **40**, 1093–1120.

Ferris, J.P., Hill, A.R. Jr., Liu, R., Orgel, L.E. (1996) Synthesis of long prebiotic oligomers on mineral surfaces, *Nature* **381**, 59–61.

Ferris, J.P., Joshi, P.C., Edelson, E.H., Lawless, J.G. (1978) HCN: A plausible source of purines, pyrimidines, and amino acids on the primitive earth, *J. Mol. Evol.* **11**, 293–311.

Fox, S. *The Emergence of Life*, Basic Books Inc., New York, p. 208.

Fuller, W.D., Sanchez, R.A., Orgel, L.E. (1972) Studies in prebiotic synthesis IV. Synthesis of purine nucleosides, *J. Mol. Biol.* **67**, 25–33.

Hanczyc, M.M., Fujikawa, S.M., Szostak, J.W. (2003) Experimental models of primitive cellular compartments: encapsulation, growth and division, *Science* **302**, 618–622.

Handschuh, G.J., Lohrmann, R., Orgel, L.E. (1973) The effect of Mg^{2+} and Ca^{2+} on urea-catalyzed phosphorylation reactions, *J. Mol. Evol.* **2**, 251–262.

Heinen, W., Lauwers, A.M. (1996) Organic sulfur compounds resulting from interaction of iron sulfide, hydrogen sulfide and carbon dioxide in an aqueous environment, *Origins Life Evol. Biosphere* **26**, 1–20.

Holm, N.G. (1992) Why are hydrothermal systems proposed as plausible environments for the origin of life? *Origins Life Evol. Biosphere* **22**, 5–14.

Huang, W., Ferris, J.P. (2003) Synthesis of 35–40 mers of RNA oligomers from unblocked monomers. A simple approach to the RNA world, *Chem. Commun.* 1458–1459.

Huber, C., Wachtershauser, G. (1997) Activated acetic acid by carbon fixation on (Fe,Ni)S under primordial conditions, *Science* **276**, 245–247.

Huber, C., Wachtershauser, G. (1998) Peptides by activation of amino acids with CO on

(Ni,Fe)S surfaces: implications for the origin of life, *Science* **281**, 670–672.

Inoue, T., Orgel, L.E. (1982) Oligomerization of (guanosine 5′-phosphor)-2-methylimidazolide on poly(C): an RNA polymerase model, *J. Mol. Biol.* **162**, 201–217.

Joshi, P.C., Pitsch, S., Ferris, J.P. (2000) Homochiral selection in the montmorillonite-catalyzed and uncatalyzed prebiotic synthesis of RNA, *Chem. Commun.* 2497–2498.

Joyce, G.F., Orgel, L.E. (1998) Prospects for Understanding the Origin of the RNA World, in: Gesteland R.F., Cech, T.R., Atkins, J.F. (Eds.) *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA*, Cold Spring Harbor Laboratory Press, Cold Spring, Harbor, NY, pp. 49–77.

Kanavarioti, A. (1997) Dimerization in highly concentrated solutions of phosphorimidazolide activated mononucleotides, *Origins Life Evol. Biosphere* **27**, 357–376.

Kaplan, R.W. (1971) The Problem of Chance in Formation of Protobionts by Random Aggregation of Macromolecules, in: Buvet R., Ponnamperuma C. (Eds.) *Molecular Evolution I. Chemical Evolution and the Origin of Life*, North Holland, Amsterdam, pp. 319–329.

Kvenvolden K., Lawless, J.G., Pering, K. Peterson, K., Flores, J., Ponnamperuma, C., Kaplan I.R., Moore, C. (1970), Evidence for extraterrestrial amino acids and hydrocarbons in the Murchison meteorite, *Nature*, **228**, 923–926.

Lohrmann, R. (1975) Formation of nucleoside 5′-polyphosphates from nucleotides and trimetaphosphate, *J. Mol. Evol.* **6**, 237–252.

Maher, K.A., Stevenson, D.J. (1988) Impact frustration of the origin of life, *Nature* **331**, 612–614.

Malin, M.C., Edgett, K.S. (2000) Evidence for recent groundwater seepage and surface runoff on Mars, *Science* **288**, 2330–2335.

Maurette, M. (1998) Carbonaceous meteorites and the origin of life, *Origins Life Evol. Biosphere* **28**, 385–412.

Miller, S.L. (1955) Production of some amino acids under possible primitive earth conditions, *J. Am. Chem. Soc.* **77**, 2351–2361.

Miyakawa, S., Ferris, J.P. (2003) Sequence- and Regioselectivity in the montmorillonite-catalyzed synthesis of RNA, *J. Am. Chem. Soc.* **125**, 8202–8208.

Miyakawa, S., Cleaves, H.J., Miller, S.L. (2002) The cold origins of life: B. Implications based on pyrimidines and purines produced from frozen ammonium cyanide solutions, *Origins Life Evol. Biosphere* **32**, 209–218.

Müller, D., Pitsch, S., Kittaka, A., Wagner, E., Wintner, C.E., Eschenmoser, A. (1990) Chemie von a-Aminonitrilen. Aldomerisierung von Glycolaldehyd-phosphat zu racemischen Hexose-2,4,6-triphosphaten und (in Gegenwart von Formaldehyd) racemischen Pentose-2,4-diphosphaten: *rac*-Allose-2,4,6-triphosphat und *rac*-Ribose-2,4-diphosphat sind die Reaktionshauptprodukte, *Helv. Chim. Acta* **73**, 1410–1468.

Nielsen, P.E., Egholm, M., Berg, R.H., Buchardt, O. (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide, *Science* **254**, 1497–1500.

Osterberg, R., Orgel, L.E., Lohrmann, R. (1973) Further studies of urea-catalyzed phosphorylation reactions, *J. Mol. Evol.* **2**, 231–234.

Paecht-Horowitz, M., Eirich, F.R. (1988) The polymerization of amino acid adenylates on sodium-montmorillonite with preadsorbed peptides, *Origins Life Evol. Biosphere* **18**, 359–387.

Palyi, G., Zucchi, C., Caglioti, L. (2002) *Fundamentals of Life*, Elsevier, Paris, p. 648.

Pitsch, S., Eschenmoser, A., Gedulin, B., Hui, S., Arrhenius, G. (1995) Mineral induced formation of sugar phosphates, *Origins Life Evol. Biosphere* **25**, 297–334.

Pizzarello, S. (2004) Chemical evolution and meteorites, an update, *Origins Life Evol. Biosphere* **34**, 25–34.

Pizzarello, S., Huang, Y., Becker, L., Poreda, R.J., Nieman, R.E., Cooper, G., Williams, M. (2001) The organic content of the Tagish lake meteorite, *Science* **293**, 2236–2240.

Prabahar, K.J., Ferris, J.P. (1997) Adenine derivatives as phosphate-activating groups for the regioselective formation of 3′,5′-linked oligoadenylates on montmorillonite: possible phosphate-activating groups for the prebiotic synthesis of RNA, *J. Am. Chem. Soc.* **119**, 4330–4337.

Ricardo, A., Carrigan, M.A., Olctt, A.N., Benner, S. (2004) Borate minerals stabilize ribose, *Science* **303**, 196.

Rubey, W.W. (1955) Development of the hydrosphere and atmosphere, *Geol. Soc. Am. Special Pap.* **62**, 631–650.

Sanchez, R.A., Ferris, J.P., Orgel, L.E. (1967) Synthesis of purine precursors and amino

acids from aqueous hydrogen cyanide, *J. Mol. Biol.* **30**, 223.

Sanchez, R., Ferris, J., Orgel, L. (1968) Studies in prebiotic synthesis IV. Conversion of 4-aminoimidazole-5-carbonitrile derivatives to purines, *J. Mol. Biol.* **38**, 121–128.

Sawai, H. (1976) Catalysis of internucleotide bond formation by divalent ions, *J. Am. Chem. Soc.* **98**, 7037–7039.

Sawai, H., Kuroda, K., Hojo, T. (1989) Uranyl ion as a highly effective catalyst for internucleotide bond formation, *Bull. Chem. Soc. Jpn.* **62**, 2018–2023.

Schidlowski, M. (2001) Carbon isotopes as biogeochemical recorders of life over 3.5 Ga of earth history: evolution of a concept, *Precambrian Res.* **106**, 117–134.

Schlesinger, G., Miller, S.L. (1983) Prebiotic synthesis in atmospheres containing CH_4, CO and CO_2, *J. Mol. Evol.* **19**, 376–382.

Sleeper, H.L., Lohrmann, R., Orgel, L.E. (1979) Template-directed synthesis of oligoadenylates catalyzed by Pb^{2+} ions, *J. Mol. Evol.* **13**, 203–214.

Spohn, T., Schubert, G. (2003) Oceans on the icy Galilean satellites of Jupiter? *Icarus* **161**, 456–467.

Stoks, P.G., Schwartz, A.W. (1981) Nitrogen-heterocyclic compounds in meteorites: significance and mechanisms of formation, *Geochim. Cosmochim. Acta* **45**, 563–569.

van Zulen, M.A., Lepland, A., Arrhenius, G. (2002) Reassessing the evidence for the earliest traces of life, *Science* **420**, 202.

Vasilyev, N.V. (1998) The Tunguska meteorite problem today., *Planet. Space Sci.* **46**, 129–150.

Verlander, M.S., Orgel, L.E. (1974) Analysis of high molecular weight material from the polymerization of adenosine cyclic 2′,3′-phosphate, *J. Mol. Evol.* **3**, 115–120.

Wilson, D.S., Szostak, J.W. (1999) In vitro selection of functional nucleic acids, *Annu. Rev. Biochem.* **68**, 611–647.

Zubay, G. (1998) Studies on the lead-catalyzed synthesis of aldopentoses., *Origins Life Evol. Biosphere* **28**, 13–26.

Pain Transduction: Gating and Modulation of Ion Channels

Peter A. McNaughton
University of Cambridge, Cambridge, UK

1 Ion Channels Gated by Noxious Stimuli 32
1.1 Ion Channels Gated by Noxious Heat 32
1.2 Ion Channels Gated by Noxious Cold 33
1.3 Ion Channels Gated by Acid 33
1.4 Ion Channels Gated by ATP 34
1.5 Ion Channels Gated by Mechanical Stimuli 35
1.6 Initiation of Action Potentials by Noxious Stimuli 35
1.7 Summary Diagram of a Nociceptive Terminal 36

2 Sensitization by Inflammatory Mediators 37

3 Short-term Sensitization: Mediators and Pathways 37
3.1 Bradykinin and the PKC Pathway 37
3.2 Prostaglandins and the PKA Pathway 39
3.3 Nerve Growth Factor 41
3.4 Direct Modulation of TRPV1 by Protons 42
3.5 Other Modulators of Nociceptor Sensitivity 42
3.5.1 ATP 43
3.5.2 Proteases 43
3.5.3 Bv8/Prokineticin 43
3.5.4 Glutamate 43
3.5.5 Norepinephrine 43

4 Long-term Sensitization 44
4.1 Gene Expression Regulated by NGF 44
4.2 Gene Expression Regulated by GDNF 45

5 Conclusion 46

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

Bibliography 46
Books and Reviews 46
Primary Literature 46

Keywords

Action Potential
A brief reversal of the neuronal membrane potential (i.e. a *depolarization*) that propagates along the nerve fiber. The universal currency by which neurons communicate over long distances.

Adaptation
The tendency for the effect of a stimulus on a sensory receptor to diminish with time. Adaptation means that a sensory receptor can function over a wide range of stimulus intensities without saturating.

Afferent Nerve
Nerve fiber that conveys sensory information to the central nervous system (i.e. which *affects* the behavior of the organism).

Depolarization
The normal negative resting potential of a neuron becomes more positive following a net influx of positive charge across the cell membrane.

Efferent Nerve Fiber
Nerve fiber that conveys a command to the muscles of an animal (i.e. which *effects* a motor action).

Hyperalgesia
A heightened pain state.

Ion Channel
Membrane protein with a central pore, which conducts ions across the cell membrane. The pore is usually selective for particular ions, for example Na^+ ions, which, because they are more concentrated outside the cell, will flow in and *depolarize* the cell membrane potential (i.e. make it more positive).

Neuropathic Pain
Long-lasting pain caused by nerve injury.

Nociceptor
A primary sensory neuron that is activated by painful stimuli (i.e. a receptor for noxious stimuli).

Primary Sensory Neuron
A neuron which directly detects sensory stimuli – examples include receptors for touch, pain (nociceptors), and light (photoreceptors).

Sensitization
The opposite of adaptation: the effect of a stimulus on a primary sensory neuron increases with time. Sensitization is a property only seen in nociceptors.

Transduction
The conversion of a sensory stimulus into an electrical signal in a sensory neuron.

The sensation of pain informs us that potentially or actually damaging stimuli (noxious stimuli) are impinging on our bodies. All animals, and even many single-celled organisms that could scarcely be dignified by the name animal, show some form of avoidance reaction in response to noxious stimuli. Pain has therefore been a feature of many forms of life since early in evolution, and we would expect pain to be a highly evolved and complex sensation, responsive to a multitude of harmful stimuli, either those present in the external environment or those generated within the organism itself.

The word ''pain'' causes some difficulties because it covers a multitude of quite separate concepts. Pain refers to the processes involved in detection of the physical effects of an injury (e.g. the pain of a fracture); to the perceptual and emotional reaction to an injury (the ''feeling'' of pain); and to similar sensations experienced in response to a purely emotional event (e.g. the pain of parting). Sherrington, in 1906, removed the emotional connotations of the word ''pain'' from the physical process of detection of a painful stimulus by coining the word ''nociceptor'' (a receptor for noxious stimuli) to refer to those primary sensory neurons that are activated by stimuli that we would regard as painful. The existence of nociceptors as a separate class of sensory neurons was doubted for many years, and it was proposed instead that the sensation of pain could result from the strong stimulation of sensory receptors responsible for detecting nonpainful stimuli, an idea which was given support by the discovery of second-order neurons in the deeper layers of the dorsal horn of the spinal cord which respond to a wide range of stimuli, both nonnoxious and noxious. However, Ed Perl in 1969 recorded directly from primary afferent nerve fibers and was able to demonstrate clearly that a separate class of primary sensory afferents responding only to high-threshold stimulation did indeed exist and had properties quite distinct from those of receptors responding to nonnoxious stimulation. Thus, the existence of a distinct class of nociceptive neurons, as postulated by Sherrington, was finally confirmed.

Nociceptors, like all somatic sensory neurons, are ''long'' neurons in which the sensory terminal must be depolarized and action potentials elicited in order to communicate to the distant central synaptic terminals. The generator current activated in nociceptive nerve terminals by all noxious stimuli must therefore be

inward, as it is in all somatic sensory neurons. In many other respects, however, nociceptive neurons differ from other sensory neurons, although most of the functional differences are readily explicable in terms of the tasks that nociceptors are designed to carry out. Most obviously, a painful stimulus is, almost by definition, large in magnitude, so the cellular amplification pathways that form a prominent part of (for instance) olfactory and visual sensory transduction are not often seen in nociceptive transduction – most stimuli, in fact, act directly to gate an ion channel. Our current understanding of ion channels important in nociception is discussed below. Secondly, while most sensory receptors respond with exquisite sensitivity to only a single sensory modality (for instance, a narrow band of wavelengths of light in the case of visual receptors), nociceptors are typically polymodal, responding to a wide range of stimuli, which may include strong mechanical stimuli, heat, extreme cold, and a range of noxious chemical stimuli. Polymodal behavior makes sense in the context of pain, where the organism will be more concerned with detecting a noxious stimulus and taking appropriate avoiding action, rather than with analyzing the precise nature of the stimulus. Finally, in all other sensory receptors, adaptation is seen in response to the maintained presentation of a stimulus, a property that enables the receptors to operate over a wide range of stimulus intensities. The response of nociceptors, in contrast, typically increases, or sensitizes, in response to a prolonged or intense stimulus, a property that ensures that a noxious stimulus is not ignored by the organism. Our current understanding of the molecular basis of sensitization is discussed below.

1 Ion Channels Gated by Noxious Stimuli

1.1 Ion Channels Gated by Noxious Heat

Many nociceptors, both unmyelinated and myelinated, respond to elevated temperatures with an action potential frequency that increases exponentially as the temperature increases above ca. 45 °C, which corresponds closely to the temperature at which a human observer would report a transition of the quality of the sensation from pleasant warmth to mild heat pain. An inward membrane current with similar properties was first recorded in our lab from isolated primary sensory neurons. The heat-sensitive neurons are small sensory neurons, as expected for nociceptors. The ion channel through which the current passes is activated at temperatures above 45 °C, turns on with a delay of ca. 35 ms following a temperature change, and is highly permeable to calcium and to monovalent cations. There is also a second neuronal population with a higher heat threshold of around 51 °C.

The heat-sensitive ion channel now thought to be responsible for the 45 °C threshold heat response was cloned in 1997 by the group of David Julius, by an expression cloning strategy using responsiveness to capsaicin, the active extract of chili peppers, as a probe. The 6-transmembrane (6-TM) channel that was obtained is a member of the TRP channel family, now named TRPV1 (VR1 originally). The channel passes an inward

current in response to temperatures above 45 °C, and is also activated by low pH. The sites of binding of capsaicin (to the intracellular loop between transmembrane domains 2 and 3) and protons (extracellular, adjacent to the channel mouth) have now been identified, but the molecular rearrangements involved in the gating of the channel by heat remain mysterious. In most aspects so far investigated, the properties of TRPV1 correspond closely to those seen in the corresponding population of nociceptive neurons (i.e. those with a heat threshold at ca. 45 °C). A second heat-sensitive channel named TRPV2 (VRL1 originally), showing substantial homology to TRPV1, was subsequently cloned. This channel is capsaicin-insensitive and has a higher heat threshold, and seems likely to be the channel expressed in the high-threshold population of nociceptors mentioned above.

Mice in which TRPV1 has been deleted are still sensitive to noxious heat, albeit at a slightly higher temperature. The explanation may be that they sense heat using TRPV2, but it is possible that other as-yet undiscovered noxious heat sensors also contribute. The most obvious defect in the TRPV1$^{-/-}$ mice was that the heat hyperalgesia observed following inflammation was almost completely abolished (discussed further below).

1.2
Ion Channels Gated by Noxious Cold

Most nociceptors respond to noxious cold temperatures (below around 20 °C), but not to innocuous cold. The ion channel likely to be responsible was identified by searching the human genome database for channels containing a 6TM motif plus the N-terminal domain ankyrin repeat characteristic of TRP channels. ANKTM1 (more recently renamed TRPA1) is activated by cooling below 17 °C, but is not activated by menthol, which potently activates TRPM8, the channel responsible for the sensation of non-noxious coolness. ANKTM1 is expressed in a subset of the neurons which also express TRPV1, the receptor for noxious heat, which may explain why noxious cold is described as "burning."

1.3
Ion Channels Gated by Acid

The first acid-sensing ion channel (ASIC, now called ASIC1) was cloned in a search for additional members of the amiloride-sensitive ion channel family, which includes a number of epithelial sodium channels. ASIC1 generates a transient inward current in response to even mild acidification, and is expressed in sensory neurons, suggesting a role in signaling the acidification associated with anoxia and metabolic overload. The main difficulty with this idea is the extremely transient nature of the ASIC1 response, and the powerful adaptation to even mild acidification, which makes a role in sustained signaling of acidotic pain seem improbable. ASIC1 is also widely expressed in neurons of the CNS, where its function is unknown. A second member of the same family, BNC1, is now recognized as acid-sensitive and has been renamed ASIC2. Like ASIC1, it is widely expressed in neurons, including small sensory neurons, but it generates a more sustained inward current in response to acidification, which may make it a more promising candidate for a physiological sensor of the pain associated with acidification. A third member of the family, ASIC3, was cloned from its homology with ASIC1, and is expressed mainly in small sensory neurones (and was initially named

DRASIC – DRG specific ASIC – in recognition of this). The response of ASIC3 to acidification is, like ASIC1, extremely transient, though there is some evidence for a sustained component activated at very low pH. Finally, the last member of the family as currently known, ASIC4, has strong homology with ASIC1 and is expressed in sensory neurons, but does not respond to acid and so seems unlikely to have a role in signaling acid pain.

The main alternative possible role for the ASIC family suggested to date is in mechanosensation, a suggestion that arises by analogy with the role of the related MEC family in mechanosensation in *Caenorhabditis elegans*. ASIC2a and ASIC3 mouse knockouts do show small alterations in nonnoxious mechanosensation, but the effects are not large, and the evidence for an involvement of the ASIC family in mechanosensation is not, at present, overwhelming. In support of this view, a recent study of mechanically gated ion currents in isolated sensory neurons from ASIC2 and ASIC3 knockout mice found that these currents were similar to wild type in both low-threshold neurons (presumed detectors of nonnoxious mechanical stimulation) and in a class of neurons responding only to high-threshold mechanical stimulation (presumed mechanical nociceptors).

1.4 Ion Channels Gated by ATP

It would be sensible for pain to be signaled by the release of the contents of damaged cells; in which case ATP is a prime candidate for activating nociceptive nerve terminals, as it is present at high concentrations within cells and is practically absent from the external medium. Injection of ATP does indeed cause pain. ATP is known to act at two main classes of membrane receptors: ion channels of the P2X family, and G-protein-coupled receptors of the P2Y family. The first are implicated in direct activation of nociceptors and are discussed here, while P2Y receptors are involved in sensitization and are discussed below.

Cloning of the first member of the P2X family opened up the way for homology cloning of seven other members, of which $P2X_3$ would appear to be of particular significance for pain transduction, as it is expressed only in nociceptors. ATP or its analogs rapidly activate an inward current through $P2X_3$, but the current is only transient, a property which is, as was the case with the ASIC family, not obviously consistent with the need to generate a sustained sensation of pain in response to ongoing cell damage. Nociceptors from the autonomic nervous system generate a sustained inward current in response to ATP, a property that can be explained by coexpression as a heteromeric channel of $P2X_2$, which generates a sustained inward current, and $P2X_3$, which generates a transient inward current. Most nociceptors from the somatic nervous system, however, express only $P2X_3$ and generate a transient inward current in response to ATP and its analogs. While these observations are internally self-consistent, they do not favor the idea that $P2X_3$ receptors might signal somatic pain in a sustained manner. Consistent with this, acute pain signaling was relatively normal in $P2X_3^{-/-}$ mice, although the detection of warm stimuli was affected, an observation that is still not fully explained. Interestingly, the main defect present in these mice was the absence of a normal bladder-emptying reflex, consistent with the idea that expression of $P2X_3$ ion channels in nociceptive neurons of the bladder is essential for

bladder-emptying when a state of painful fullness has been reached.

1.5 Ion Channels Gated by Mechanical Stimuli

The first stimulus most people would think of in the context of pain is a strong mechanical stimulus, but, ironically, the ion channels involved in the generation of this most intuitively obvious of the pain sensations have proven the most elusive. Recent studies have identified inward membrane currents activated in isolated DRG neurons by direct mechanical stimulation either with a probe or following stretching of the membrane by swelling in hypotonic solutions. Two populations of neurons can be distinguished: those with a low threshold, which are capsaicin-insensitive and therefore are presumably nonnociceptive, and those with a higher threshold, which are, in general, sensitive to capsaicin and are therefore presumably nociceptive neurons. These two populations are likely to correspond to the two main observed classes of *in vivo* mechanoreceptors, which detect nonnoxious and noxious stimuli, respectively.

The molecular identity of the mechanically gated ion channel is still unclear, but two main possibilities are currently being considered. As discussed above under "Ion channels gated by acid," the homology between the ASIC family and some members of the MEC family of ion channels, known to be mechanically gated in *C. Elegans*, has led to the proposal that ASICs form part of the mechanoreceptive complex, and there is some support for this proposal from gene knockout studies (see above). The mechanosensitive current is inhibited by calcium, a property that is reminiscent of the strong inhibition of ASIC-dependent currents by calcium, but, in other respects, the mechanosensitive currents do not obviously exhibit the properties of ion channels of the ASIC family. In particular, the amplitude of mechanosensitive currents is not correlated with that of acid-gated currents in DRG neurons; the mechanosensitive currents are not altered by acidification; and knockout of either or both of ASIC2 and ASIC3 has no effect on mechanically gated ion currents.

The second possibility is that a member of the TRP family may be involved. TRPV4 (originally called OTRPC4 or VR-OAC) generates an inward current in response to cell swelling, though the action appears to be via phosphorylation of TRPV4, rather than being a direct gating which would be more consistent with the known rapid response to noxious mechanical stimuli. Downregulation of *TRPV4* gene expression impaired the nociceptive response to injection of hypotonic solutions, and pressure sensation was reduced in TRPV4$^{-/-}$ mice. While these data are intriguing, they do not as yet provide unambiguous evidence for involvement of TRPV4 in mechanosensation, and other possibilities must remain open.

1.6 Initiation of Action Potentials by Noxious Stimuli

To be detected, all noxious stimuli must generate action potentials in the afferent nerve fiber. The first hint that this process may be different from that in other sensory neurons came when a voltage-sensitive sodium channel unique to nociceptive neurons was cloned in 1996 in John Wood's lab by a difference-cloning strategy. This channel, initially called SNS (*sensory neuron specific*) but subsequently renamed $Na_V1.8$, has two

main differences from most other voltage-dependent sodium channels: it is relatively resistant to tetrodotoxin, and the voltage at which it is first activated is more positive by about 20 mV than that of the widely expressed $Na_V1.2$ or $Na_V1.7$ channels. Significantly, the threshold of $Na_V1.8$ is modulated in the negative direction by inflammatory mediators such as prostaglandins, thereby making the neuron more excitable (see below).

A second TTX-insensitive sodium channel, $Na_V1.9$ (initially named NaN or SNS2), was subsequently identified. This channel, in spite of having a high homology to the other members of the Na_V family, seems not to be voltage-activated in the same way as other sodium channels, though it may generate a "window current" that contributes to maintenance of the resting potential. Surprisingly, this sodium channel has recently been shown to generate an inward current when directly activated by the binding of BDNF to TrkB receptors.

1.7 Summary Diagram of a Nociceptive Terminal

Figure 1 summarizes the ion channels currently known to be involved in excitation of the nociceptive nerve terminal. Typically, several different ion channels are expressed in a single nociceptor, thus conferring the property of polymodality commonly seen in nociceptors, but there is often, in practice, a more restricted expression than that shown in Fig. 1, leading

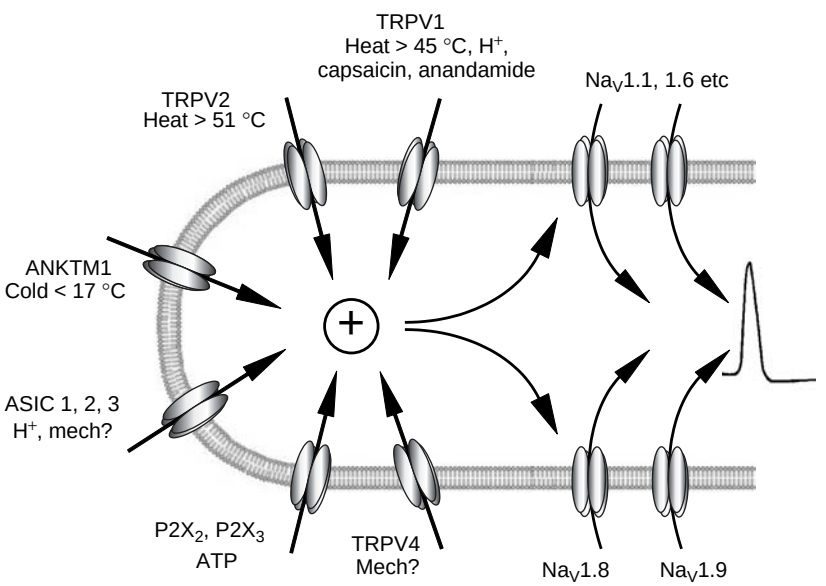

Fig. 1 Schematic diagram of ion channels activated by noxious stimuli in nociceptor nerve terminals. Ion channels responsible for activating inward current in response to noxious stimuli are shown at the left, along with the stimuli believed to be primarily responsible for their activation. Voltage-sensitive sodium channels activated by the consequent depolarization are shown diagrammatically on the right. Note that $Na_V1.9$ is not voltage-activated in the conventional sense, but, instead, appears to be activated by direct interaction with BDNF bound to TrkB.

to a somewhat greater degree of specificity in individual nerve terminals. An influx of current – usually carried mainly by Na^+ ions, though Ca^{2+} ions may also contribute – through any of the ion channels leads to depolarization and consequent activation of voltage-sensitive sodium channels, thus generating an action potential which propagates to the CNS. The sensitivity of the nociceptor to noxious stimuli can be modulated in two main ways: either by enhancing the generator current activated in response to a noxious stimulus, or by changing the threshold of the sodium channel so that the nociceptor is more readily excited. These possibilities are discussed below.

2 Sensitization by Inflammatory Mediators

Following tissue damage or inflammation, the sensitivity of nociceptive nerve terminals *in vivo* increases, a process known as *sensitization*. In the sensitized state, painful stimuli are enhanced (a phenomenon known as *hyperalgesia*) and ordinarily nonpainful stimuli are perceived as painful (*allodynia*). Sensitization is caused by the release of chemical mediators from surrounding damaged or inflamed tissue, because when nociceptive neurons are isolated they show only adaptation to repeated application of stimuli such as heat or capsaicin. The mediators capable of causing sensitization include prostaglandins, bradykinin, protons, ATP, and even nerve growth factor, and new physiologically important factors are being discovered at a high rate. The majority act via G-protein-coupled receptors and intracellular signaling pathways, though some may act directly (e.g. protons act directly on TRPV1). Sensitization has both short-term components, on the timescale of seconds and due to a local action of inflammatory mediators on nociceptive nerve terminals, and long-term components due to upregulation of gene transcription and a consequent change in the protein expression in the nerve terminal. The pathways involved are better understood in some cases than in others.

3 Short-term Sensitization: Mediators and Pathways

3.1 Bradykinin and the PKC Pathway

Bradykinin is a nonapepetide released by proteolysis of a precursor protein, kininogen, when proteolytic enzymes are released from damaged cells, and acts mainly at B2 receptors. Kininase I cleaves the C-terminal arginine to give des-arg^9 bradykinin, which is a potent activator of B1 receptors but largely ineffective at B2 receptors. Both B1 and B2 receptors are coupled to G_q and thence to PLCβ, leading to cleavage of PIP_2 and release of IP_3 and DAG. In 1996, we found that bradykinin potently enhances the heat-activated current, an observation that was made before the identification of TRPV1, but the downstream target is now identified as TRPV1 because a similar effect is seen in HEK293 cells transfected with TRPV1.

The following lines of evidence suggest that the pathway leading to sensitization of TRPV1 involves protein kinase C. PMA, a selective activator of PKC, has similar effects to bradykinin; staurosporine, a PKC inhibitor, blocks the effects of bradykinin; and phosphatase inhibitors prolong the effects of bradykinin. There are 11 isoforms of PKC, of which five (PKCβI, βII, γ, ε and

ζ) are expressed in neonatal DRG neurons, but only one isoform, PKCε, is involved in the sensitization of TRPV1 by bradykinin, because infusion of a constitutively active form of PKCε has similar effects in enhancing the heat-activated membrane current, and because a selective inhibitor of PKCε suppresses the action of bradykinin. Bradykinin causes a selective translocation of PKCε to the neuronal cell membrane, a sensitive indicator of PKCε activation. A later study showed that PKCα, which is expressed in more mature DRG neurons, but not in neonatal neurons, may also be able to activate TRPV1. The proposed pathway linking B2 receptor activation to TRPV1 activation is shown in Fig. 2(a).

Residues phosphorylated by PKCε have now been identified by site-directed mutagenesis. Serines 502 and 800 in rat TRPV1 are in the first intracellular loop and in the C-terminal tail respectively (see Fig. 2(b)), and replacement of both these residues by nonphosphorylatable alanines was found to completely abolish the enhancement of TRPV1-dependent current caused by activation of PKCε. All of the effects of PKCε on TRPV1 would therefore seem to be mediated by phosphorylation of these two residues.

A second mechanism for the action of bradykinin on TRPV1, independent of PKC activation, has also been proposed by the group of David Julius. Removal of PIP2 from the membrane, either by hydrolysis or with an antibody, reduced the action of bradykinin and NGF in expression systems, suggesting that PIP2 associates with and inhibits TRPV1, and that removal of PIP2 is the crucial factor in enhancing TRPV1 following activation of PLC. The relative importance of this second mechanism for the action of bradykinin in sensory neurons, as opposed to in expression systems, has yet to be investigated, but the observations that application of kinase inhibitors removes most or all of the enhancement caused by bradykinin

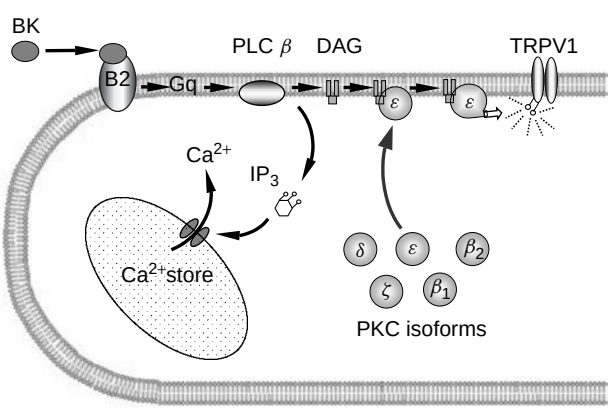

Fig. 2a Pathways leading to sensitization of TRPV1. Binding of bradykinin to the B2 receptor leads to activation of phospholipase C and release of IP_3 and diacylglycerol. Of the five PKC isoforms expressed in nociceptive neurons, only PKCε is translocated to the membrane by bradykinin. Activation of PKCε by bradykinin leads to phosphorylation of TRPV1 and enhancement of the membrane current gated by a heat stimulus.

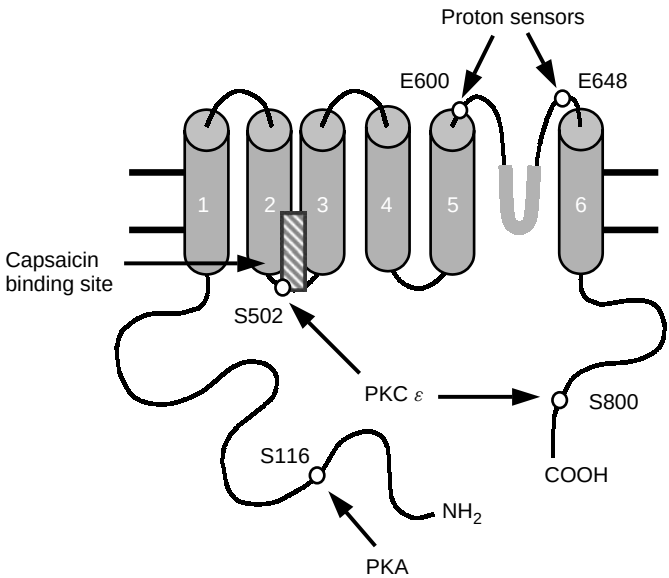

Fig. 2b Targets of sensitization on TRPV1 (numbering of sites for rat TRPV1). The binding site for activation by capsaicin is shown. Target sites include intracellular serines, which are phosphorylation sites for PKCε and PKA, and extracellular glutamates whose protonation also modulates gating (see text).

(see above) suggests that it plays a lesser role than the PKCε pathway shown in Fig. 2(a).

3.2
Prostaglandins and the PKA Pathway

Arachidonic acid is released from membrane lipids by the action of phospholipase A_2 in response to a wide range of damaging and stressful stimuli, and is the precursor for a vast range of proinflammatory mediators, including leukotrienes, lipoxins, thromboxanes, and prostaglandins. PGE_2 and PGI_2, which are both produced downstream of the cyclo-oxygenase (COX) enzymes, are particularly important mediators of hyperalgesia, and the inhibition of their production by COX inhibitors such as aspirin explains the potent analgesic effect of these inhibitors. When PGE_2 is applied to nociceptive neurons, the threshold of the TTX-insensitive sodium current shifts in the negative direction, leading to a decrease in the threshold for eliciting an action potential and, consequently, an increase in excitability in response to a wide range of noxious stimuli. Activation of adenylate cyclase with forskolin has a similar effect, suggesting that the cAMP/PKA pathway is important in this action of PGE_2. In agreement with this idea, it was found that the effect could be abolished by mutating five serine residues, all of which are consensus sites for PKA and are located on the first major intracellular loop of $Na_V1.8$. Thus, an important part of the action of PGE_2 can be attributed to phosphorylation of $Na_V1.8$ by PKA, as shown in Fig. 3.

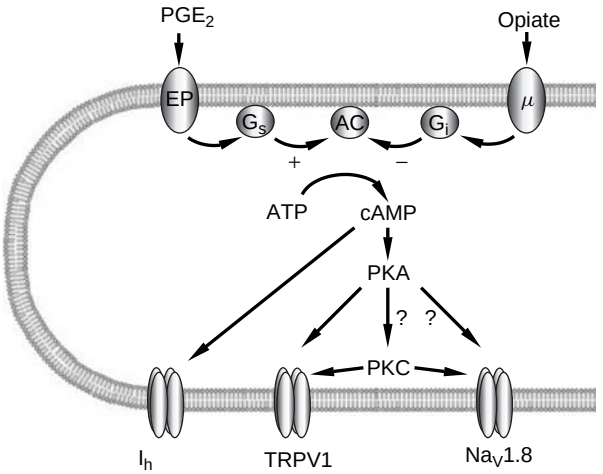

Fig. 3 Intracellular pathways and targets of the cAMP/PKA pathway causing nociceptor sensitization. An indicative proalgesic agent (PGE_2, binding to a receptor of the EP family) and an analgesic agent (an opiate, binding to the μ opiate receptor) are shown, but it seems likely that a number of other agents, both proalgesic and analgesic, will activate the same pathways.

An alternative possibility was suggested by studies in which PKA and PKC were each blocked by selective inhibitors. Activation of PKC alone increased the TTX-resistant sodium current (but apparently without the shift of threshold caused by PKA activation), and the increase was not blocked by PKA inhibitors, while the increase in current caused by PKA activation was blocked by PKC inhibitors. These experiments suggest that the pathway linking PGE_2 application to an increase in current through $Na_V1.8$ may be PKA → PKC → $Na_V1.8$. This result is not necessarily incompatible with the finding mentioned above, that is, that mutating consensus sequences for PKA abolishes modulation of $Na_V1.8$, as many potential phosphorylation sites are surrounded by adequate consensus sequences for both PKA and PKC. It seems likely that $Na_V1.8$ is targeted by both kinases, but the relative importance of each for the observed increase in TTX-insensitive sodium current and shift in excitation threshold has not been completely disentangled.

PKA also targets TRPV1. While phosphorylation by PKC directly enhances the heat or capsaicin-gated current, as described above, phosphorylation by PKA has the more subtle effect of reversing the desensitization caused by prolonged or repeated activation of TRPV1. Although more than one site is phosphorylated by PKA, the reversal of desensitization is abolished by mutating just a single serine, 116 in the rat TRPV1 sequence. The action of PKA on TRPV1 depends on anchoring by an A-kinase-anchoring protein (AKAP), suggesting the formation of a signaling complex in which AKAP anchors PKA to TRPV1 in order to achieve rapid and reproducible physiological modulation of the membrane current gated by heat (Fig. 3).

The cAMP produced by activation of adenylate cyclase may also have a direct effect in modulating the voltage sensitivity of I_h ion channels, which pass an inward current and are activated by hyperpolarization. Members of this family have a cyclic nucleotide binding site near the C-terminal, and binding of, for example, cAMP shifts the activation range in the positive direction, which has the effect of increasing the inward current and therefore of hastening depolarization to threshold for initiation of the next action potential. I_h plays a crucial role in many tissues, including heart, where it mediates the pacemaker potential, and neurones, where it determines the firing frequency by modulating the rate at which the membrane potential depolarizes between action potentials. Four channel subunits, HCN 1–4, have been cloned. I_h has been reported in nociceptors, where it may play an important role in modulating firing in response to agents that modulate cAMP, such as prostaglandins, sympathomimetic monoamines, or opiates.

Opiates have well-known and potent central analgesic actions, but it has recently been shown that they also have a peripheral analgesic action. Binding of agonist to the μ opiate receptor expressed on nociceptors activates G_i, thus causing an inhibition of cAMP production and consequently antagonizing the proalgesic actions of (for example) prostaglandins. A peripheral analgesic action for cannabinoids has also been suggested, but whether the cAMP pathway is involved in the same way as it is for opiates has yet to be established.

3.3 Nerve Growth Factor

Activation of mast cells by a stimulus such as bacterial lipopolysaccharide (LPS) causes release of nerve growth factor (NGF), which is a potent, though perhaps surprising, inflammatory mediator. NGF injection causes a rapid and long-lasting thermal hyperalgesia, and a mechanical hyperalgesia with a slower onset. NGF is involved in physiologically significant hyperalgesia *in vivo*, since removal of NGF reverses both the thermal and the mechanical hyperalgesia caused by an injection of LPS. There is general agreement that the long-term effects of NGF are due to upregulation of gene expression (see below), but the short-term effects, which can be seen *in vivo* within minutes of injection of NGF, must be due to activation of local signaling pathways within the nerve terminal. Rapid effects of NGF on TRPV1 *in vitro* have been characterized in isolated neurons using whole-cell patch clamp and calcium imaging, and show that an enhancement of the current passing through TRPV1 is induced within 2 min of exposure to NGF. TRPV1 gating is enhanced by NGF in around 40% of TRPV1-expressing neurons, a proportion that agrees with immunocytochemical experiments in which it was found that about 50% of nociceptive neurons express TrkA receptors for NGF.

The identity of the pathways activated by NGF and leading to sensitization of TRPV1 is currently the subject of some disagreement. Activation of TrkA receptors by NGF recruits many signaling molecules that can bind to the intracellular phosphorylated tyrosine residues within TrkA by means of Src homology (SH2) domains. Three proteins, in particular, have been identified on the basis of their specific binding to phosphorylated Trk receptors: Shc, which activates the Ras/MEK pathway; phospholipase C gamma-1 (PLCγ1), which cleaves PtdIns-4,5-P_2 to IP_3 and

DAG; and phosphatidylinositol-3-kinase (PI3K), which 3-phosphorylates PtdIns-4,5-P_2. Many other signaling molecules are possible secondary mediators of the physiological effects of TrkA activation. Mendell's group has proposed that PKA is an essential downstream element, as PKA inhibitors abolished the NGF-induced sensitization in patch clamp studies, while inhibitors of PKC and MAP kinase were without effect. Work from the group of David Julius has proposed a quite different mechanism, independent of phosphorylation: activation of PKCγ by TrkA, which is known to lead to breakdown of PIP2, may remove PIP2 from binding sites on TRPV1, releasing it from PIP2 mediated inhibition. This mechanism would unify the sensitizing actions of a number of agents, including bradykinin, NGF, ATP acting at P2Y receptors, and others, which have in common that they activate PLC. In our lab, we have used calcium imaging of intact neurons to investigate signaling pathways activated by TrkA, and our conclusions are again different: we find that wortmannin, a selective inhibitor of PI3 kinase, totally abolishes sensitization, while inhibition of PLC-γ is without effect, suggesting that PI3 kinase, and not PLCγ, is the crucial early step. MEK inhibition had no effect on NGF-induced sensitization, showing that the MEK/MAPK pathway is not involved, at least downstream of MEK, while ras inhibition had a small inhibitory effect on sensitization, consistent with a modulation of PI3K by ras, as shown in other studies. We have also found that inhibitors of PKC and CAM kinase abolish sensitization, suggesting an involvement of these kinases at a later stage, though the details of the later stages of the signaling pathway are currently unclear. The observation that kinase inhibitors abolish NGF-induced sensitization suggests that phosphorylation of TRPV1 is an essential final common pathway, as with sensitization in response to other inflammatory mediators discussed above, but the kinases involved, and the residues on TRPV1, which they phosphorylate, have yet to be identified.

3.4
Direct Modulation of TRPV1 by Protons

The acid produced by inflammation or during ischemic exercise, where pH levels may fall as low as 5.4, is an important contributor to pain. As discussed above, protons may cause pain directly by an action at either or both of TRPV1 or members of the ASIC family, but protons can, in addition, sensitize the response of TRPV1 to heat, without directly activating it, by shifting to lower temperatures the relationship between channel activation and heat, in much the same way as bradykinin does. The relevant residues have been identified as two glutamates near the mouth of the channel, one of which (E600) is involved in setting the sensitivity of the channel to heat, while a second (E648) is involved in direct activation of the channel. It is interesting to note that two actions on opposite sides of the membrane – protonation at an external site, and phosphorylation by PKC at an internal site – can have rather similar effects in terms of modulating the sensitivity of the TRPV1 ion channel to heat.

3.5
Other Modulators of Nociceptor Sensitivity

In a highly evolved system such as nociception, a wide variety of factors released by injury or inflammation would be expected to cause hyperalgesia, and the

few factors outlined above are only the best understood amongst many. The following list outlines what is known of some others, but this list is sure to be far from complete.

3.5.1 ATP

The best-characterized action of ATP is in activating P2X ion channels and thus directly exciting nociceptors, but a sensitizing action arising from an action at P2Y receptors may be more important *in vivo*. Through an action at the $P2Y_1$ G-protein-coupled receptors, ATP activates the same pathway as bradykinin, that is, PLCβ, releasing DAG and leading to activation of PKCε and consequent phosphorylation of TRPV1, as shown in Fig. 2a.

3.5.2 Proteases

Proteases released following cell damage activate a family of protease activated receptors (PARs), in which proteolysis of the extracellular N-terminal domain leads to unmasking of a tethered agonist peptide and consequent receptor activation. In intact animals, prolonged hyperalgesia results from injection of PAR activators, and PAR2 seems to be particularly involved in the generation of itch. Downstream pathways involve PLC, leading to intracellular calcium signals and consequent exocytosis of neuropeptides, and ERK activation may also play a role.

3.5.3 Bv8/Prokineticin

Bv8 is a small protein isolated from amphibian skin whose mammalian homologs are the prokineticin family. The receptors for these proteins, PKR1 and PKR2, are expressed in mammalian DRG neurons, and injection of Bv8 leads to a profound and long-lasting thermal and mechanical hyperalgesia in intact animals. Application of Bv8 to isolated DRG neurons causes calcium release from intracellular stores, suggesting that an activation of the PLC pathway may be the basis of the sensitization. While these observations are tantalizing, there is as yet no evidence that prokineticins are actually released during inflammation, nor what component of physiological hyperalgesia may be due to activation of PKRs.

3.5.4 Glutamate

The possibility that glutamate may play a role in hyperalgesia was raised by the observation that the metabotropic glutamate receptors mGluR1 and mGluR5 are expressed on C-fiber afferents, and that injection of agonists of these receptors causes thermal hyperalgesia. The observation that antagonists partially suppress the pain caused by formalin injection suggests a physiological role for glutamate in pain. Current understanding of the intracellular pathway is that PLC is activated by mGluR, but the main downstream effect is not to activate PKC, but, instead, to cause release of prostaglandins and TRPV1 sensitization via a PKA-dependent pathway.

3.5.5 Norepinephrine

Activation of β_2 receptors by norepinephrine (noradrenaline) or other adrenergic agonists causes a hyperalgesia which is reduced in PKCε KO mice or when PKCε is inhibited, suggesting an involvement of PKCε downstream of β_2 receptor activation. The target of PKCε when activated by norepinephrine has not been clearly established, and could be TRPV1, as is the case when bradykinin receptors are activated (see above, Fig. 2) and/or $Na_V1.8$. More recent work has suggested that the

ERK/MAPK pathway may, in addition, be an important mediator of sensitization caused by β_2 activation.

4 Long-term Sensitization

Pain is usually transient in nature, disappearing quickly when a mild noxious stimulus is removed, or more gradually as the inflammation caused by an injury dies away. Some forms of pain, however, notably those caused by nerve injury, can be much longer lasting, suggesting that long-term changes in gene expression underlie them. Long-lasting pain caused by nerve injury is called *neuropathic pain*, and its origin is currently the subject of much debate. Long-term changes in transmission to second-order neurons in the spinal cord are important, but phenotypic changes in the ability of primary sensory neurons to detect painful stimuli or to sensitize in response to inflammatory mediators also play a role.

Growth factors play a crucial role in the control of nociceptor phenotype. Neonatal neurons depend on NGF for survival, but in the adult animal, removal of NGF does not cause cell death but, instead, changes the expression of many proteins associated with the nociceptive phenotype. A reduction in NGF supply can be caused by nerve section, which removes the supply of target-derived NGF from innervated tissues. The supply of neurotrophins can also be increased in some circumstances – for instance, NGF is known to be released from mast cells and other sources during inflammation, and both NGF and GDNF are released from invading macrophages when nerve axons degenerate following nerve damage. Thus NGF and GDNF will bathe any surviving axons in the damaged nerve trunk in a neurotrophin-rich fluid.

4.1 Gene Expression Regulated by NGF

The expression of the neuropeptides substance P and CGRP was shown a number of years ago to be downregulated by NGF removal or, conversely, upregulated by addition of NGF, while other neuropeptides such as VIP and CCK are upregulated by NGF removal. One effect of an oversupply of NGF may, therefore, be to enhance peptidergic neurotransmission by substance P and CGRP at the first synapse, while deprivation of NGF, caused for instance by nerve section, would have the converse effect. TRPV1 expression is upregulated by NGF in a pathway that depends upon activation of ras and the MEK/MAPK cascade. Work in our own lab has shown that expression of the bradykinin B2 receptor is also upregulated by NGF.

One of the prominent physiological changes in nociceptors following nerve damage involves the generation of sustained action potential firing, and a neurotrophin-driven change in sodium channel expression may be a contributor to this. The expression of $Na_V1.8$ is upregulated by NGF and, conversely, is downregulated by nerve section. This last observation would not appear to offer a ready explanation of neuropathic pain, in which we would be looking for an upregulation of the expression of $Na_V1.8$ following nerve damage in order to explain repetitive activity, but other evidence suggests that in neuropathic pain states $Na_V1.8$ may, in fact, be upregulated in nerve trunks, if not in the neuronal cell body. A second channel, $Na_V1.3$, is not normally expressed in sensory neurons but appears after nerve damage, and has

characteristics that may make it a candidate for the originator of the repetitive firing seen following nerve damage.

A second possible contributor to the repetitive firing observed following nerve damage is the hyperpolarization-activated inward current channel, I_h, which contributes to action potential firing by regulating the rate at which the neuronal membrane depolarizes between action potentials (see above). Recently, it has been suggested that I_h may be functionally enhanced in neuropathic pain states, as a blocker of I_h, ZD7288, alleviated neuropathic pain. Surprisingly, I_h protein and mRNA expression was found to be downregulated in an experimental neuropathic pain state, but the magnitude of the I_h current nonetheless increased, perhaps because the voltage dependence of the current is shifted in the depolarizing direction.

4.2 Gene Expression Regulated by GDNF

Only around half of nociceptors express TrkA receptors for NGF, while the other half express Ret receptors for GDNF. Many of the nociceptor-specific proteins expressed by these two groups of neurons are different: the TrkA-expressing neurons also express the neuropeptides substance P and CGRP and terminate in outer layers of the dorsal horn (layer I and II outer), while the Ret-expressing neurons express surface receptors binding to the plant lectin IB4 and terminate in layer II (inner) of the dorsal horn. These two populations might be considered the "rods and cones" of the nociceptor system, but so far no very distinctive differences in their properties have been characterized, at least not to match those conferring differential wavelength selectivity and time course on the visual rod and cone systems.

Ability to bind IB4 provides a convenient method for distinguishing the two nociceptor populations, as it can be readily applied both to living and to fixed neurons. Several nociceptor-specific proteins are localized to the IB4-positive population: both the P2X3 receptor for ATP and the $Na_V1.9$ sodium channel are exclusively located in the $IB4^+$ neuronal population, and our group has identified B1 bradykinin receptors as also being exclusively located in $IB4^+$ neurons, unlike B2 receptors, which are expressed mainly in $TrkA^+$ neurons.

Ret forms functional receptors for members of the GDNF family, which, apart from GDNF itself, comprises the elegantly named members neurturin, artemin, and persephin, by combining with the GFRα receptor family, although there is some evidence that Ret may be active alone. By analogy with the action of NGF in upregulating gene expression in TrkA-expressing neurons, we might expect that GDNF and other members of this family would have a similar effect in upregulating the proteins expressed in the $IB4^+$ neurons, but, to date, this possibility has been little investigated. A recent report, however, shows that B1 expression is potently upregulated by GDNF. B1 receptors are normally expressed in only a very small fraction of neurons (around 2%), but, following exposure to GDNF functional B1 receptor expression, appear in around 20% of neurons. GDNF or related neurotrophins are therefore likely candidates to drive upregulation of B1 expression following injury or inflammation, leading to the switch in agonist responsiveness from a B2 to a B1 profile identified in *in vivo* experiments.

5
Conclusion

There has been an explosion of research into the cellular and molecular basis of pain transduction over the past few years. Many of the membrane ion channels responsible for the detection of painful stimuli have now been cloned, and we are making good progress in identifying the intracellular pathways responsible for modulating these ion channels and thus mediating the process of sensitization. Some areas are still mysterious and will no doubt form the subject of advances in understanding in future years. We still know little about the molecular basis for detection of strong mechanical stimuli (or, indeed, any mechanical stimuli). The means by which temperature-sensitive ion channels are gated still remains unknown at the molecular level. Much remains to be discovered about the pathways mediating sensitization – kinases seem to be important in most of the pathways investigated to date, but the details of how they are activated, and which are their critical target residues on the primary transducer ion channels, still remain, in many cases, to be elucidated.

See also Neuron Chemistry.

Bibliography

Books and Reviews

Campbell, J.N., Meyer, R.A. (1996) Cutaneous Nociceptors, in: Belmonte, C., Cervero, F. (Eds.) *Neurobiology of Nociceptors*, Oxford University Press, Oxford, pp. 117–145.

Caterina, M.J., Julius, D. (2001) The vanilloid receptor: a molecular gateway to the pain pathway, *Annu. Rev. Neurosci.* **24**, 487–517.

Cesare, P., McNaughton, P.A. (1997) Peripheral pain mechanisms, *Curr. Opin. Neurobiol.* **7**, 493–499.

Cesare, P., Moriondo, A., Vellani, V., McNaughton, P.A. (1999) Ion channels gated by heat, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 7658–7663.

Dray, A., Perkins, M. (1993) Bradykinin and inflammatory pain, *Trends Neurosci.* **16**, 99–104.

Sherrington, C.S. (1906) *The Integrative Action of the Nervous System*, Scribner, New York.

Snider, W.D., McMahon, S.B. (1998) Tackling pain at the source: New ideas about nociceptors, *Neuron* **20**, 629–632.

Primary Literature

Akopian, A.N., Sivilotti, L., Wood, J.N. (1996) A tetrodotoxin-resistant voltage-gated sodium channel expressed by sensory neurons, *Nature* **379**, 257–262.

Akopian, A.N., Chen, C-C., Ding, Y., Cesare, P., Wood, J.N. (2000) A new member of the acid-sensing ion channel family, *NeuroReport* **11**, 2217–2222.

Alessandri-Haber, N., Yeh, J.J., Boyd, A.E., Parada, C.A., Chen, X., Reichling, D.B., Levine, J.D. (2003) Hypotonicity induces TRPV4-mediated nociception in rat, *Neuron* **39**, 497–511.

Aley, K.O., Martin, A., McMahon, T., Mok, J., Levine, J.D., Messing, R.O. (2001) Nociceptor sensitization by extracellular signal-regulated kinases, *J. Neurosci.* **21**, 6933–6939.

Baloh, R.H., Enomoto, H., Johnson, E.M. Jr., Milbrandt, J. (2000) The GDNF family ligands and receptors – implications for neural development, *Curr. Opin. Neurobiol.* **10**, 103–110.

Belmonte, C., Giraldez, F. (1981) Responses of cat corneal sensory receptors to mechanical and thermal stimulation, *J. Physiol.* **321**, 355–368.

Bessou, P., Perl, E.R. (1969) Response of cutaneous sensory units with unmyelinated fibers to noxious stimuli, *J. Neurophysiol.* **32**, 1025–1043.

Bhave, G., Karim, F., Carlton, S.M., Gereau, R.W. (2001) Peripheral group I metabotropic glutamate receptors modulate nociception in mice, *Nat. Neurosci.* **4**, 417–423.

Bhave, G., Zhu, W., Wang, H., Brasier, D.J., Oxford, G.S., Gereau, R.W. (2002) cAMP-dependent protein kinase regulates desensitization of the capsaicin receptor (VR1) by direct phosphorylation, *Neuron* **35**, 721–731.

Bhave, G., Hu, H.J., Glauner, K.S., Zhu, W., Wang, H., Brasier, D.J., Oxford, G.S., Gereau, R.W. (2003) Protein kinase C phosphorylation sensitizes but does not activate the capsaicin receptor transient receptor potential vanilloid 1 (TRPV1), *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12480–12485.

Blum, R., Kafitz, K.W., Konnerth, A. (2002) Neurotrophin-evoked depolarization requires the sodium channel $Na_V1.9$, *Nature* **419**, 687–693.

Bonnington, J.K., McNaughton, P.A. (2003) Signalling pathways involved in the sensitisation of mouse nociceptive neurones by nerve growth factor, *J. Physiol.* **551**, 433–446.

Bron, R., Klesse, L.J., Shah, K., Parada, L.F., Winter, J. (2003) Activation of Ras is necessary and sufficient for upregulation of vanilloid receptor type 1 in sensory neurons by neurotrophic factors, *Mol. Cell Neurosci.* **22**, 118–132.

Calignano, A., La Rana, G., Giuffrida, A., Piomelli, D. (1998) Control of pain initiation by endogenous cannabinoids, *Nature* **394**, 277–281.

Caterina, M.J., Rosen, T.A., Tominaga, M., Brake, A.J., Julius, D. (1999) A capsaicin-receptor homologue with a high threshold for noxious heat, *Nature* **398**, 436–441.

Caterina, M.J., Schumacher, M.A., Tominaga, M., Rosen, T.A., Levine, J.D., Julius, D. (1997) The capsaicin receptor: a heat-activated ion channel in the pain pathway, *Nature* **389**, 816–824.

Caterina, M.J., Leffler, A., Malmberg, A.B., Martin, W.J., Trafton, J., Petersen-Zeitz, K.R., Koltzenburg, M., Basbaum, A.I., Julius, D. (2000) Impaired nociception and pain sensation in mice lacking the capsaicin receptor, *Science* **288**, 306–313.

Cesare, P., McNaughton, P.A. (1996) A novel heat-activated current in nociceptive neurons, and its sensitization by bradykinin, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 15435–15439.

Cesare, P., Dekker, L.V., Sardini, A., Parker, P.J., McNaughton, P.A. (1999a) Specific involvement of PKC-epsilon in sensitization of the neuronal response to painful heat, *Neuron* **23**, 617–624.

Chaplan, S.R., Guo, H.Q., Lee, D.H., Luo, L., Liu, C., Kuei, C., Velumian, A.A., Butler, M.P., Brown, S.M., Dubin, A.E. (2003) Neuronal hyperpolarization-activated pacemaker channels drive neuropathic pain, *J. Neurosci.* **23**, 1169–1178.

Chen, C.C., Akopian, A.N., Sivilotti, L., Colquhoun, D., Burnstock, G., Wood, J.N. (1995) A P2X purinoceptor expressed by a subset of sensory neurons, *Nature* **377**, 428–431.

Chuang, H.H., Prescott, E.D., Kong, H., Shields, S., Jordt, S.E., Basbaum, A.I., Chao, M.V., Julius, D. (2001) Bradykinin and nerve growth factor release the capsaicin receptor from PtdIns(4,5)P2-mediated inhibition, *Nature* **411**, 957–962.

Cockayne, D.A., Hamilton, S.G., Zhu, Q.M., Dunn, P.M., Zhong, Y., Novakovic, S., Malmberg, A.B., Cain, G., Berson, A., Kassotakis, L., Hedley, L., Lachnit, W.G., Burnstock, G., McMahon, S.B., Ford, A.P. (2000) Urinary bladder hyporeflexia and reduced pain-related behaviour in P2X3-deficient mice, *Nature* **407**, 1011–1015.

Davis, A.J., Perkins, M.N. (1994) Induction of B1 receptors in vivo in a model of persistent inflammatory mechanical hyperalgesia in the rat, *Neuropharmacology* **33**, 127–133.

Davis, J.B., Gray, J., Gunthorpe, M.J., Hatcher, J.P., Davey, P.T., Overend, P., Harries, M.H., Latcham, J., Clapham, C., Atkinson, K., Hughes, S.A., Rance, K., Grau, E., Harper, A.J., Pugh, P.L., Rogers, D.C., Bingham, S., Randall, A., Sheardown, S.A. (2000) Vanilloid receptor-1 is essential for inflammatory thermal hyperalgesia, *Nature* **405**, 183–187.

Dib-Hajj, S.D., Tyrrell, L., Black, J.A., Waxman, S.G. (1998) NaN, a novel voltage-gated Na channel, is expressed preferentially in peripheral sensory neurons and down-regulated after axotomy, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 8963–8968.

Ding, Y., Cesare, P., Drew, L., Nikitaki, D., Wood, J.N. (2000) ATP, P2X receptors and pain pathways, *J. Auton. Nerv. Syst.* **81**, 289–294.

Dirajlal, S., Pauers, L.E., Stucky, C.L. (2003) Differential response properties of IB4-positive and -negative unmyelinated sensory neurons to protons and capsaicin, *J. Neurophysiol.* **89**, 513–524.

Distler, C., Rathee, P.K., Lips, K.S., Obreja, O., Neuhuber, W., Kress, M. (2003) Fast Ca2+-induced potentiation of heat-activated ionic currents requires cAMP/PKA signaling and functional AKAP anchoring, *J. Neurophysiol.* **89**, 2499–2505.

Drew, L.J., Wood, J.N., Cesare, P. (2002) Distinct mechanosensitive properties of capsaicin-sensitive and - insensitive sensory neurons, *J. Neurosci.* **22**, RC228.

Drew, L.J., Rohrer, D.K., Price, M.P., Blaver, K., Cockayne, D.A., Cesare, P., Wood, J.N. (2004) ASIC2 and ASIC3 do not contribute to mechanically activated currents in mammalian sensory neurons, *J. Physiol.* **556**, 691–710.

England, S., Bevan, S., Docherty, R.J. (1996) PGE2 modulates the tetrodotoxin-resistant sodium current in neonatal rat dorsal root ganglion neurones via the cyclic AMP-protein kinase A cascade, *J. Physiol.* **495**, 429–440.

Fitzgerald, E.M., Okuse, K., Wood, J.N., Dolphin, A.C., Moss, S.J. (1999) cAMP-dependent phosphorylation of the tetrodotoxin-resistant voltage-dependent sodium channel SNS, *J. Physiol.* **516**, 433–446.

Fjell, J., Hjelmstrom, P., Hormuzdiar, W., Milenkovic, M., Aglieco, F., Tyrrell, L., Dib-Hajj, S., Waxman, S.G., Black, J.A. (2000) Localization of the tetrodotoxin-resistant sodium channel NaN in nociceptors, *NeuroReport* **11**, 199–202.

Gold, M.S., Levine, J.D. (1996) DAMGO inhibits prostaglandin E2-induced potentiation of a TTX-resistant Na+ current in rat sensory neurons in vitro, *Neurosci. Lett.* **212**, 83–86.

Gold, M.S., Levine, J.D., Correa, A.M. (1998) Modulation of TTX-R I(Na) by PKC and PKA and their role in PGE2-induced sensitization of rat sensory neurons in vitro, *J. Neurosci.* **18**, 10345–10355.

Gold, M.S., Reichling, D.B., Shuster, M.J., Levine, J.D. (1996) Hyperalgesic agents increase a tetrodotoxin-resistant Na+ current in nociceptors, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1108–1112.

Gold, M.S., Weinreich, D., Kim, C.S., Wang, R., Treanor, J., Porreca, F., Lai, J. (2003) Redistribution of Na(V)1.8 in uninjured axons enables neuropathic pain, *J. Neurosci.* **23**, 158–166.

Hamilton, S.G., McMahon, S.B. (2000) ATP as a peripheral mediator of pain, *J. Auton. Nerv. Syst.* **81**, 187–194.

Herzog, R.I., Cummins, T.R., Waxman, S.G. (2001) Persistent TTX-resistant Na+ current affects resting potential and response to depolarization in simulated spinal sensory neurons, *J. Neurophysiol.* **86**, 1351–1364.

Hu, H.J., Bhave, G., Gereau, R.W. (2002) Prostaglandin and protein kinase A-dependent modulation of vanilloid receptor function by metabotropic glutamate receptor 5: potential mechanism for thermal hyperalgesia, *J. Neurosci.* **22**, 7444–7452.

Immke, D.C., McCleskey, E.W. (2001) Lactate enhances the acid-sensing Na+ channel on ischemia-sensing neurons, *Nat. Neurosci.* **4**, 869–870.

Ingram, S.L., Williams, J.T. (1994) Opioid inhibition of I(h) via adenylyl cyclase, *Neuron* **13**, 179–186.

Ingram, S.L., Williams, J.T. (1996) Modulation of the hyperpolarisation-activated current (I(h)) by cyclic nucleotides in guinea-pig primary afferent neurons, *J. Physiol.* **492**, 97–106.

Ji, R.R., Kohno, T., Moore, K.A., Woolf, C.J. (2003) Central sensitization and LTP: do pain and memory share similar mechanisms? *Trends Neurosci.* **26**, 696–705.

Jordt, S.E., Julius, D. (2002) Molecular basis for species-specific sensitivity to ''hot'' chili peppers, *Cell* **108**, 421–430.

Jordt, S-E., Tominaga, M., Julius, D. (2000) Acid potentiation of the capsaicin receptor determined by a key extracellular site, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8134–8139.

Kaplan, D.R., Miller, F.D. (2000) Neurotrophin signal transduction in the nervous system, *Curr. Opin. Neurobiol.* **10**, 381–391.

Kaupp, U.B., Seifert, R. (2001) Molecular diversity of pacemaker ion channels, *Annu. Rev. Physiol* **63**, 235–257.

Kennedy, C., Assis, T.S., Currie, A.J., Rowan, E.G. (2003) Crossing the pain barrier: P2 receptors as targets for novel analgesics, *J. Physiol.* **553**, 683–694.

Khasar, S.G., McCarter, G., Levine, J.D. (1999b) Epinephrine produces a beta-adrenergic receptor-mediated mechanical hyperalgesia and in vitro sensitization of rat nociceptors, *J. Neurophysiol.* **81**, 1104–1112.

Khasar, S.G., Lin, Y-H., Martin, A., Dadgar, J., McMahon, T., Wang, D., Hundle, B., Aley, K.O., Isenberg, W., McCarter, G., Green, P.G., Hodge, C.W., Levine, J.D., Messing, R.O. (1999a) A novel nociceptor signaling pathway

revealed in protein kinase c epsilon mutant mice, *Neuron* **24**, 253–260.

Lai, J., Gold, M.S., Kim, C.S., Bian, D., Ossipov, M.H., Hunter, J.C., Porreca, F. (2002) Inhibition of neuropathic pain by decreased expression of the tetrodotoxin-resistant sodium channel, NaV1.8, *Pain* **95**, 143–152.

Lee, Y-J., Zachrisson, O., Tonge, D.A., McNaughton, P.A. (2002) Upregulation of bradykinin b2 receptor expression by neurotrophic factors and nerve injury in mouse sensory neurons, *Mol. Cell. Neurosci.* **19**, 186–200.

Lewin, G.R., Ritter, A.M., Mendell, L.M. (1993) Nerve growth factor-induced hyperalgesia in the neonatal and adult rat, *J. Neurosci.* **13**, 2136–2148.

Lewis, C., Neidhart, S., Holy, C., North, R.A., Buell, G., Surprenant, A. (1995) Coexpression of P2X2 and P2X3 receptor subunits can account for ATP-gated currents in sensory neurons, *Nature* **377**, 432–435.

Liedtke, W., Choe, Y., Marti-Renom, M.A., Bell, A.M., Denis, C.S., Sali, A., Hudspeth, A.J., Friedman, J.M., Heller, S. (2000) Vanilloid receptor-related osmotically activated channel (VR-OAC), a candidate vertebrate osmoreceptor, *Cell* **103**, 525–535.

Lindsay, R.M., Harmar, A.J. (1989) Nerve growth factor regulates expression of neuropeptide genes in adult sensory neurons, *Nature* **337**, 362–364.

Ma, Q-P., Woolf, C.J. (1997) The progressive tactile hyperalgesia induced by peripheral inflammation is nerve growth factor dependent, *NeuroReport* **8**, 807–810.

Marceau, F., Hess, J.F., Bachvarov, D.R. (1998) The B1 receptors for kinins, *Pharmacol. Rev.* **50**, 357–386.

McCarter, G.C., Reichling, D.B., Levine, J.D. (1999) Mechanical transduction by rat dorsal root ganglion neurons in vitro, *Neurosci. Lett.* **273**, 179–182.

McKemy, D.D., Neuhausser, W.M., Julius, D. (2002) Identification of a cold receptor reveals a general role for TRP channels in thermosensation, *Nature* **416**, 52–58.

McMahon, S.B., Bennett, D.L.H., Priestley, J.V., Shelton, D.L. (1995) The biological effects of endogenous nerve growth factor on adult sensory neurons revealed by a trkA-IgG fusion molecule, *Nat. Med.* **1**, 774–780.

Meyer, R.A., Campbell, J.N., Srinivasa, N.R. (1994) Peripheral Neural Mechanisms of Nociception, in: Melzack, R., Wall, P.D. (Eds.) *Textbook of Pain*, Churchill Livingstone, Edinburgh, Scotland, pp. 13–44.

Nagy, I., Rang, H. (1999) Noxious heat activates all capsaicin-sensitive and also a sub-population of capsaicin-insensitive dorsal root ganglion neurons, *Neuroscience* **88**, 995–997.

Naveilhan, P., ElShamy, W.M., Ernfors, P. (1997) Differential regulation of mRNAs for GDNF and its receptors Ret and GDNFR alpha after sciatic nerve lesion in the mouse, *Eur. J. Neurosci.* **9**, 1450–1460.

Negri, L., Lattanzi, R., Giannini, E., Metere, A., Colucci, M., Barra, D., Kreil, G., Melchiorri, P. (2002) Nociceptive sensitization by the secretory protein Bv8, *Br. J. Pharmacol.* **137**, 1147–1154.

Numazaki, M., Tominaga, T., Toyooka, H., Tominaga, M. (2002) Direct phosphorylation of capsaicin receptor VR1 by protein kinase Cepsilon and identification of two target serine residues, *J. Biol. Chem.* **277**, 13375–13378.

Olah, Z., Karai, L., Iadarola, M.J. (2002) Protein kinase C(alpha) is required for vanilloid receptor 1 activation. Evidence for multiple signaling pathways, *J. Biol. Chem.* **277**, 35752–35759.

Peier, A.M., Moqrich, A., Hergarden, A.C., Reeve, A.J., Andersson, D.A., Story, G.M., Earley, T.J., Dragoni, I., McIntyre, P., Bevan, S., Patapoutian, A. (2002) A TRP channel that senses cold stimuli and menthol, *Cell* **108**, 705–715.

Premkumar, L.S., Ahern, G.P. (2000) Induction of vanilloid receptor channel activity by protein kinase C, *Nature* **408**, 985–990.

Prescott, E.D., Julius, D. (2003) A modular PIP2 binding site as a determinant of capsaicin receptor sensitivity, *Science* **300**, 1284–1288.

Price, M.P., Snyder, P.M., Welsh, M.J. (1996) Cloning and expression of a novel human brain Na^+ channel, *J. Biol. Chem.* **271**, 7879–7882.

Price, M.P., McIlwrath, S.L., Xie, J., Cheng, C., Qiao, J., Tarr, D.E., Sluka, K.A., Brennan, T.J., Lewin, G.R., Welsh, M.J. (2001) The DRASIC cation channel contributes to the detection of cutaneous touch and acid stimuli in mice, *Neuron* **32**, 1071–1083.

Price, M.P., Lewin, G.R., McIlwrath, S.L., Cheng, C., Xie, J., Heppenstall, P.A., Stucky,

C.L., Mannsfeldt, A.G., Brennan, T.J., Drummond, H.A., Qiao, J., Benson, C.J., Tarr, D.E., Hrstka, R.F., Yang, B., Williamson, R.A., Welsh, M.J. (2000) The mammalian sodium channel BNC1 is required for normal touch sensation, *Nature* **407**, 1007–1011.

Rathee, P.K., Distler, C., Obreja, O., Neuhuber, W., Wang, G.K., Wang, S.Y., Nau, C., Kress, M. (2002) PKA/AKAP/VR-1 module: A common link of Gs-mediated signaling to thermal hyperalgesia, *J. Neurosci.* **22**, 4740–4745.

Regoli, D., Barabe, J. (1980) Pharmacology of bradykinin and related kinins, *Pharmacol. Rev.* **32**, 1–46.

Rodriguez-Viciana, P., Warne, P.H., Dhand, R., Vanhaesebroeck, B., Gout, I., Fry, M.J., Waterfield, M.D., Downward, J. (1994) Phosphatidylinositol-3-OH kinase as a direct target of Ras, *Nature* **370**, 527–532.

Shu, X., Mendell, L.M. (2001) Acute sensitization by NGF of the response of small-diameter sensory neurons to capsaicin, *J. Neurophysiol.* **86**, 2931–2938.

Simone, D.A., Kajander, K.C. (1996) Excitation of rat cutaneous nociceptors by noxious cold, *Neurosci. Lett.* **213**, 53–56.

Souslova, V., Cesare, P., Ding, Y., Akopian, A.N., Stanfa, L., Suzuki, R., Carpenter, K., Dickenson, A., Boyce, S., Hill, R., Nebenuis-Oosthuizen, D., Smith, A.J., Kidd, E.J., Wood, J.N. (2000) Warm-coding deficits and aberrant inflammatory pain in mice lacking P2X3 receptors, *Nature* **407**, 1015–1017.

Srinivasa, N.R., Meyer, R.A., Ringkamp, M., Campbell, J.N. (1999) Peripheral Neural Mechanisms of Nociception, in: Melzack, R., Wall, P.D. (Eds.) *Textbook of Pain*, Churchill Livingstone, Edinburgh, Scotland, pp. 11–45.

Steen, K.H., Reeh, P.W., Anton, F., Handwerker, H.O. (1992) Protons selectively induce lasting excitation and sensitization to mechanical stimulation of nociceptors in rat skin, in vitro, *J. Neurosci.* **12**, 86–95.

Stein, C., Schafer, M., Machelska, H. (2003) Attacking pain at its source: new perspectives on opioids, *Nat. Med.* **9**, 1003–1008.

Steinhoff, M., Neisius, U., Ikoma, A., Fartasch, M., Heyer, G., Skov, P.S., Luger, T.A., Schmelz, M. (2003) Proteinase-activated receptor-2 mediates itch: a novel pathway for pruritus in human skin, *J. Neurosci.* **23**, 6176–6180.

Steinhoff, M., Vergnolle, N., Young, S.H., Tognetto, M., Amadesi, S., Ennes, H.S., Trevisani, M., Hollenberg, M.D., Wallace, J.L., Caughey, G.H., Mitchell, S.E., Williams, L.M., Geppetti, P., Mayer, E.A., Bunnett, N.W. (2000) Agonists of proteinase-activated receptor 2 induce inflammation by a neurogenic mechanism, *Nat. Med.* **6**, 151–158.

Story, G.M., Peier, A.M., Reeve, A.J., Eid, S.R., Mosbacher, J., Hricik, T.R., Earley, T.J., Hergarden, A.C., Andersson, D.A., Hwang, S.W., McIntyre, P., Jegla, T., Bevan, S., Patapoutian, A. (2003) ANKTM1, a TRP-like channel expressed in nociceptive neurons, is activated by cold temperatures, *Cell* **112**, 819–829.

Strotmann, R., Harteneck, C., Nunnenmacher, K., Schultz, G., Plant, T.D. (2000) OTRPC4, a nonselective cation channel that confers sensitivity to extracellular osmolarity, *Nat. Cell Biol.* **2**, 695–702.

Stucky, C.L., Lewin, G.R. (1999) Isolectin B(4)-positive and -negative nociceptors are functionally distinct, *J. Neurosci.* **19**, 6497–6505.

Suzuki, M., Mizuno, A., Kodaira, K., Imai, M. (2003) Impaired pressure sensation in mice lacking TRPV4, *J. Biol. Chem.* **278**, 22664–22668.

Tominaga, M., Caterina, M.J., Malmberg, A.B., Rosen, T.A., Gilbert, H., Skinner, K., Raumann, B.E., Basbaum, A.I., Julius, D. (1998) The cloned capsaicin receptor integrates multiple pain-producing stimuli, *Neuron* **21**, 531–543.

Treede, R.D., Meyer, R.A., Raja, S.N., Campbell, J.N. (1995) Evidence for two different heat transduction mechanisms in nociceptive primary afferents innervating monkey skin, *J. Physiol.* **483**, 747–758.

Trupp, M., Ryden, M., Jornvall, H., Funakoshi, H., Timmusk, T., Arenas, E., Ibanez, C.F. (1995) Peripheral expression and biological activities of GDNF, a new neurotrophic factor for avian and mammalian peripheral neurons, *J. Cell Biol.* **130**, 137–148.

Vellani, V., Zachrisson, O., McNaughton, P.A. (2004). Functional bradykinin B1 receptors are expressed in nociceptive neurons and are upregulated by GDNF, *J. Physiol.* **560**, 391–402.

Vellani, V., Mapplebeck, S., Moriondo, A., Davis, J.B., McNaughton, P.A. (2001) Protein kinase C activation potentiates gating of the vanilloid receptor VR1 by capsaicin, protons, heat and anandamide, *J. Physiol* **534**, 813–825.

Verge, V.M.K., Richardson, P.M., Wiesenfeld-Hallin, Z., Hokfelt, T. (1995) Differential influence of nerve growth factor on neuropeptide expression in vivo: A novel role in peptide suppression in adult sensory neurons, *J. Neurosci.* **15**, 2081–2096.

Vergnolle, N., Ferazzini, M., D'Andrea, M.R., Buddenkotte, J., Steinhoff, M. (2003) Proteinase-activated receptors: novel signals for peripheral nerves, *Trends Neurosci.* **26**, 496–500.

Viana, F., De La, P.E., Pecson, B., Schmidt, R.F., Belmonte, C. (2001) Swelling-activated calcium signalling in cultured mouse primary sensory neurons, *Eur. J Neurosci.* **13**, 722–734.

Waldmann, R., Champigny, G., Bassilana, F., Heurteaux, C., Lazdunski, M. (1997b) A proton-gated cation channel involved in acid sensing, *Nature* **386**, 173–177.

Waldmann, R., Bassilana, F., De Weille, J., Champigny, G., Heurteaux, C., Lazdunski, M. (1997a) Molecular cloning of a non-inactivating proton-gated Na+ channel specific for sensory neurons, *J. Biol. Chem.* **272**, 20975–20978.

Waldmann, R., Champigny, G., Lingueglia, E., De, W.J., Heurteaux, C., Lazdunski, M. (1999) H+-gated cation channels, *Ann. N. Y. Acad. Sci.* **868**, 67–76.

Wall, P.D., Dubner, R. (1972) Somatosensory pathways, *Annu. Rev. Physiol* **34**, 315–336.

Wall, P.D., Waxman, S., Basbaum, A.I. (1974) Ongoing activity in peripheral nerve: injury discharge, *Exp. Neurol.* **45**, 576–589.

Waxman, S.G. (2001) Acquired channelopathies in nerve injury and MS, *Neurology* **56**, 1621–1627.

Waxman, S.G., Dib-Hajj, S., Cummins, T.R., Black, J.A. (1999) Sodium channels and pain, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 7635–7639.

Woolf, C.J., SafiehGarabedian, B., Ma, Q.P., Crilly, P., Winter, J. (1994) Nerve growth factor contributes to the generation of inflammatory sensory hypersensitivity, *Neuroscience* **62**, 327–331.

Xu, H., Zhao, H., Tian, W., Yoshida, K., Roullet, J.B., Cohen, D.M. (2003) Regulation of a transient receptor potential (TRP) channel by tyrosine phosphorylation. SRC family kinase-dependent tyrosine phosphorylation of TRPV4 on TYR-253 mediates its response to hypotonic stress, *J. Biol. Chem.* **278**, 11520–11527.

Paleontology, Molecular

Rob DeSalle and Mary G. Egan
Division of Invertebrates, American Museum of Natural History, New York, NY, USA

1 Definitions 54
1.1 Paleontology 54
1.2 Molecular Paleontology 55

2 Fossilization 55
2.1 Types of Fossilization 55
2.2 Mineralization 56
2.3 Amber, Mummies, Subfossils, and Museum Collections 57
2.4 Quality of Molecules from Ancient Sources: Proteins 58
2.5 Quality of Molecules from Ancient Sources: Nucleic Acids 59

3 Studies in Molecular Paleontology 61
3.1 Microscopy Studies 61
3.2 Chemical and Amino Acid Studies 62
3.3 Protein and Nucleic Acid Studies 62
3.4 Molecular Archaeology: Humans and Plants 64
3.5 Utility of Ancient DNA 64

4 Future of Ancient DNA Research: Authentication 65

Bibliography 68
Books and Reviews 68
Primary Literature 68

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

Keywords

Ancient DNA
Deoxyribonucleic acid isolated from ancient tissue sources; the most common method for isolation of DNA from ancient tissue sources is through the polymerase chain reaction.

Biomarkers
Chemical or biochemical structures that were once a part of the biomass of a living ancient organism; also known as molecular fossils.

Fossilization
The preservation of deceased organisms via mineralization, mummification, compression, or amberization.

Mineralization, Amberization, and Mummification
The process of fossilization that preserves materials for examination at the molecular level.

Paleontology
The study of the structure, composition, and behavior of dead organisms.

Recent developments in modern molecular biology have facilitated the expansion of the new field of molecular paleontology. This field comprises the isolation and characterization of proteins and nucleic acids from fossil material. The various kinds of fossil preservation are outlined and discussed with respect to the types of preservation most appropriate for producing fossils that will yield direct sequence information. The molecular techniques used to isolate nucleic acids and proteins, and to characterize the sequence information from these molecules are described. The biological importance of data obtained from fossils is also discussed.

1 Definitions

1.1 Paleontology

Paleontology is defined as the study of ancient organisms. Paleontological information has been recognized as an important aspect of evolutionary biology since before Darwin's time. Indeed, Darwin dedicated an entire chapter of *On the Origin of Species* to an examination of the fossil record. Darwin was preoccupied with the incompleteness of the fossil record and was concerned with the degree to which it supported his theory of natural selection. Several other examples of the influence of fossils on modern biology can

be cited, such as the discovery of entire groups of unusual organisms that have gone extinct (dinosaurs), the characterization of mass extinctions, and the use of the fossil record to calibrate divergence times in species comparisons. The utility of fossils for stratigraphic and chronologic correlations have also been repeatedly demonstrated.

1.2 Molecular Paleontology

The term *molecular paleontology*, at first was meant to include only the study of biological marker compounds, or biomarkers (organic molecules that were originally part of the biomass of an ancient organism, and have been preserved in rocks and molecules from fossils). The term has come to include the study of molecular information, both chemical and biochemical, potentially available from fossilized remains. (*Chemical fossil* is a broad term covering atomic and molecular fossils, including isotopic features.) More recently, molecular paleontology has been redefined to include historical information derived from comparisons of molecules of living organisms. A subbranch of molecular paleontology that concerns nonfossilized material from deceased organisms is known as *molecular archaeology*. Examination of recently deceased organisms at the molecular level is considered to be a part of molecular forensics. In the past, paleontological studies have provided information on ancient organisms only at the morphological level. With the development of gas chromatography, mass spectrometry, scanning and transmission electron microscopy, and modern molecular biology, the molecular aspects of fossils and subfossils have been explored. Three major kinds of molecular information can be recovered from fossils. The first is made of visual structural information using various forms of microscopy. Thin sections of microorganisms observed under microscopy have demonstrated the feasibility of this approach, and several fossil microscopic molecular structures have been observed this way. The second kind of molecular data consists of compositional information. Ancient atmospheric conditions as well as the chemistry of the ancient oceans have been examined using chemical composition information from fossils. Chemical and biochemical fossils have been examined for molecular structure by means of combined gas chromatography and mass spectrometry. Some information on the amino acid composition of tissues has been obtained from fossil and subfossil materials. Radioimmunoassay (RIA) techniques have also been used to examine the immunological distances of extinct organisms with their living relatives. The newest and perhaps most exciting class of molecular information obtained from fossils is the analysis of the actual genetic material of fossils, subfossils, and recently extinct organisms.

2 Fossilization

2.1 Types of Fossilization

The condition of the fossilized material is critical to the recovery of molecular information from any fossil. The depositional environment and the rock matrix where the fossil is deposited and the degree of decomposition of the fossil are critical factors in determining the state of the fossilized material. Fossilization of an organism is an extremely rare event; if it is going to occur,

it begins immediately after the death of the organism. In general, only a small portion of the organism is preserved. After death, an organism's body undergoes three types of decomposition. The first is biological and is accomplished by the decompositional activity of bacteria and by autolytic enzymatic processes. Bacteria are ubiquitous in environments with oxygen, so only organisms that are deposited into oxygen-free environments, hence able to escape bacterial decomposition, will have soft body parts preserved in any detail. The second type of decomposition is mechanical; it is facilitated by waves, wind, and other forms of abrasion. The final type of decomposition is chemical, which can proceed well after the fossilization process has begun. Ancient organisms can be divided into three categories: organisms that have been preserved through mineralization (mineralized fossils); organisms that have been preserved in amber (amber fossils); and organisms that have been preserved but not mineralized or amberized (subfossils and compression fossils).

2.2 Mineralization

Several types of mineralization can occur in the fossilization process, and each results in differences with respect to tissue preservation and the degrees of decomposition that occur in the organism. Mineralization usually produces fossils in which the molecules of the preserved organism have been replaced by minerals. Such fossils will be entirely devoid of organic molecules because all molecules in the organism will have been replaced or displaced by mineral matter. *Permineralization* is the impregnation of an organism with mineral materials, most frequently calcium carbonate and silica. Such a process produces a structure that is resistant to chemical change or chemical decomposition. *Metasomatism,* produced from the exchange of existing minerals for new minerals, yields a fossil that retains its physical appearance but undergoes a drastic change in chemical composition. In another process, the oxygen, hydrogen, and nitrogen molecules in plant remains are sometimes *leached* away, leaving a carbon residue that retains the original structures of the plant. *Incrustation,* in turn, produces fossilization, most frequently in more recent organismal material. When calcium-rich water flows over the remains of organisms they are covered with a *mineral film.* The organic material under the film then decomposes, leaving only a negative impression on the mineral layer. In the *distillation* process, water and gases are distilled away from the organism leaving on the rock a thin film of carbon that sometimes can give rich information on the structure of the distilled organism. Some tissues may be more resistant to the mineralization processes than others. Chitin and keratin are among the compounds found in organisms that are rather resistant to chemical decomposition and mineralization. Minerals such as apatite (teeth and bones of vertebrates) and calcium carbonate (shells of invertebrates) often form very resistant structures and parts of organisms. Chitinous exoskeletons of arthropods can be resistant to chemical decomposition by mineralization. Other proteins and nucleic acids (DNA and RNA) are much more susceptible to decomposition. Some depositional conditions produce fossils remarkably resistant to chemical decomposition and complete mineralization. The Solenhofen deposits in Bavaria and the Clarkia fossil beds of northern Idaho are excellent examples of this type of fossilization. Here,

organisms were deposited in an extremely low oxygen setting and rapidly buried in fine sediment. Because of the rapid burial and low oxygen availability, chemical and biological decomposition have not disturbed the organic material. During such a process, organisms are deposited in a very fine laminated sediment and preserved between two extremely thin sedimentary layers. The layers are usually found as shale deposits, and the organic material is usually preserved in extreme detail. In such deposits, it is actually possible to remove the remains of organisms from between the shale layers, and considerable ultrastructural preservation can be observed from these organisms.

2.3 Amber, Mummies, Subfossils, and Museum Collections

These forms of fossilization may leave remarkably well-preserved organismal remains as well as preserve organic material suitable for molecular studies. *Amberization* is caused by the flow of resinous material from several plant sources. The sites and ages of the major amber deposits on the face of the earth vary greatly. These deposits yield amber-preserved specimens from a wide range of animals, plants, protists, and bacteria. The common factor in the quality of the material found in amber is the thorough dehydration and protection of encased organic material. When the chemicals in amber (primarily diterpenoids) link together, trapped organic material dehydrates and becomes inert. (An excellent review of the biochemistry of amber formation was provided by Langenheim.) The ability of the flowing resin to completely encapsulate an organism and the bactericidal activity of terpenes in the resin suggest that amberization will produce highly protected fossil specimens. The most famous of these deposits are found from the Oligocene of Romania and the Baltic region and from the Miocene of the Dominican Republic, Mexico, Sicily, and the Apennines of Italy.

Natural desiccation (*mummification*) is yet another form of fossilization that can preserve ancient material in a form suitable for molecular analysis. This type of fossilization is extremely rare and involves almost complete preservation of the encased organism. An example of this kind of preservation from the Cretaceous is the anatosaurus, a dinosaur that was preserved complete with wrinkles in its skin. The specimen was preserved in this way as a result of extreme dehydration of the animal. Such preservation most likely occurs if the organism is deposited not in a submarine sedimentary basin but, rather, in unusual terrestrial environments. More recent examples of mummification are the ground sloth (*Mylodon*) and other smaller mammals of about 13 000 years ago from the Ultima Esperanza Cave in Chile, bog-preserved human remains of about 8000 years ago from the Windover Pond region of Florida, and animals of about 10 000 years ago preserved in California's La Brea tar pit. Permafrost-preserved organisms have also been used as a source of ancient material. The most famous is the woolly mammoth found in 40 000-year-old permafrost and the Neolithic human remains (named Oetzi) found in the Alps and believed to be 4000 to 5000 years old. Human cultural practices also result in mummification or preservation of organisms. The detailed preservation by the ancient Egyptians of the bodies of humans and some domestic animals is well known.

Other examples of more recently preserved remains exist. Nonmineralized (subfossil) remains of bones and teeth,

seeds, and even soil extracts have been used in ancient DNA studies. DNA from museum-preserved specimens of an incredibly diverse and large number also can be designated as ancient DNA. Teeth, bones, pieces of skins, feathers, alcohol-preserved specimens, and some pinned insects have all been used as sources of material for molecular studies. The Quagga study was the first study to extract DNA from museum specimens. Numerous studies on historical materials from excavations and museum collections representing varied taxa and tissues have followed. The continued importance of museum-derived collections should be emphasized as a source of ancient specimens for study. Not all museum material is suitable for molecular analysis, however, inasmuch as some of the preservation procedures used by museum curators are rather hydrophilic and produce massive damage to the specimen at the molecular level. Table 1 lists several sources of animal and plant material successfully yielding DNA from specimens dating from the Pleistocene and Holocene.

2.4 Quality of Molecules from Ancient Sources: Proteins

Tissues and molecules react very differently in living organisms and under

Tab. 1 Pleistocene and Holocene specimens yielding ancient DNA.

Preservation condition	*Material extracted*	*Taxon*	*Age of specimen [years before present]*
Dry remains			
	Animal bones and teeth	Ground sloth	13 000
		Cave bear	20 000–130 000
		Moa	3350
		Saber-toothed cat	14 000
		Horse	25 000
		Woolly rhinoceros	40 000–70 000
	Plant material	Maize	3400
		Wheat	1200
	Temperate cave sediments	Two species of moa, and 29 plant taxa	600–3000
	Parasites	Endogenous retrovirus-like element from woolly mammoth	13 000–26 000
	Coprolites	Shasta sloth coprolite recovering both sloth and ingested plant DNA	Late Pleistocene
Frozen remains			
	Siberian permafrost	Woolly mammoth	13 000–33 000
	Beringian permafrost	Brown bear	14 000–42 000
	Alpine permafrost	Human (Oetzi)	4000
	Permafrost soil	19 different plant taxa and mammoth, bison, and horse	10 000–400 000

conditions of fossilization. For this reason, different tissues, and molecules undergo varying degrees of degradation and breakdown as fossilization proceeds. Amino acids were among the first types of molecule examined in fossils. By continuously etching fossil clamshells, the existence of amino acids throughout the fossilized material was demonstrated. Contamination as an explanation for the discovery of fossil amino acids was ruled out on the basis of three observations. First, several mollusk shells from the same sedimentary deposit were examined for amino acid content and substantial qualitative and quantitative differences were observed from shell to shell. If there had been contamination from the surrounding sediment, the amino acid contents would have been similar. Second, in binding experiments the two most common amino acids found in fossils have little if any affinity to calcium carbonate, and therefore these substances (glycine and alanine) are highly unlikely to have been absorbed from groundwater. Finally, the sediment surrounding the fossils was examined and the amino acid content was determined to be minimal. The relative amounts of each amino acid in several fossils was calculated over geological time. From studies of this type, the conclusion emerged that relative abundances of amino acids prepared from specimens from the various geologic ages show similar but distinct patterns. The results of examining the amino acid content of amberized insect fossils, suggest that the degree of preservation of amino acids in amberized materials, though high, varies from specimen to specimen. Typically, in proteins that have been fossilized, the peptide bonds between amino acids will be rapidly degraded. Under special circumstances, especially in shells and skeletal remains, peptide bonds will remain intact and short peptides will be preserved. The position of proteins between (inter-) and within (intra-) crystalline structures in a fossil will affect the longevity of fossil molecules. Intracrystalline molecules would be expected to have greater survival potential as a result of the protective aspects of the surrounding crystal structure. The examination of fossils of several marine organisms with shell structures has revealed a rapid decline in the abundance of amino acids in the first few hundred thousand years after death. The degree of decomposition usually levels off at 2% surviving amino acids. Soluble and insoluble organic residues are also important factors in determining the degree of preservation of amino acids and peptide fragments in fossils.

2.5 Quality of Molecules from Ancient Sources: Nucleic Acids

Nucleic acids and DNA, in particular, would be expected to be much less well preserved in fossil materials than some proteins. DNA is primarily concentrated in soft tissues, which are highly susceptible and vulnerable to degradation during fossilization. The stability of the DNA double helix has been discussed at length by Lindahl. Basically there are three major modes of chemical and physical destabilization to the double helix: DNA hydrolysis, DNA oxidation, and nonenzymatic DNA methylation (Table 2). The stability of the nucleic acid phosphodiester bond that links one nucleotide to another in both deoxyribonucleotide strands and ribonucleotide strands is gained at the cost of a labile *N*-glycosyl bond. This bond connects the sugar ring of the nucleotide to

Tab. 2 Process of destabilization of the DNA double helix as discussed by Lindahl.

Process	*Effect*	*Target*	*Rate*
Hydrolysis			
I	Base sugar bonds destroyed	*N*-glycosyl bonds of G and A	Up to 10 000 purine bases removed in human cells per day
II	Deamination	Cytosine	Half-life of C residue in double-stranded DNA estimated at 30 000 years
III	Conversion of C to U	Cytosine	Not determined
Oxidation			
I	Hydroxylation of G to produce 8-hydroxyguanine	Guanine	Occurs 100–500 times per day in human cells; "indirect and imprecise"
II	Ring-saturated derivative of a pyrimidine	T and C	Not determined; "indirect and imprecise"
III	Major helical distortions	G and A	Not determined
Methylation	S-adenosyl-methionine (SAM)	Ring nitrogens; purines; G, A, T, C	600 alkylation events in the human cell per day

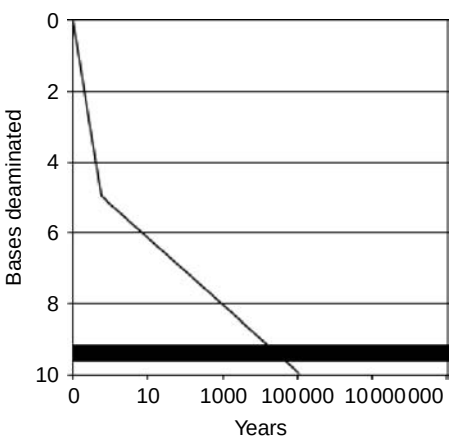

Fig. 1 Graph of time on a logarithmic scale versus the number of bases deaminated, also on a logarithmic scale. The numbers on the bases-deaminated axis refer to the log of 10 molecules deaminated. The figure was drawn by extrapolation from the published lower rate of deamination as 2000 bases per day. Stippled bar at the bottom indicates the cutoff for the genome size of humans (10^9 bases per haploid genome).

the base of the nucleotide. It was discovered by means of hydrolysis experiments that guanine (G) and adenine (A) are liberated almost 20 times more easily than thymine (T) and cytosine (C), indicating that depurination (loss of G and A) is much more common under hydrolysis than the loss of pyrimidines (T and C). The hydrolysis of the DNA double helix would result in lesions in single and double-stranded nucleic acids. The half-life of single-stranded DNA due to hydrolysis in solution at pH 7.4 and at 37 °C was calculated to be about 200 years. The half-life of double-stranded DNA under similar conditions due to hydrolysis was estimated at 30 000 years (Fig. 1). It has been estimated that this process of hydrolysis

causes the depurination of 2000–10 000 purine residues everyday in the typical human cell. The cell can tolerate such a high degree of depurination due to hydrolysis because enzymes exist in the living cell to repair the damaged regions of the double helix. In fossilized cells, repair enzymes would most definitely be inactive, with the result that any damage to the DNA in a fossilized cell does not get repaired. Active oxygen is an important source of damage to nucleic acids in growing cells. Degradative processes due to oxidation can cause oxidative damage to occur in several places along the DNA double helix (Table 2). The major product of this kind of damage to DNA is the conversion of guanine residues to 8-hydroxyguanine; indeed, this process and the hydrolytic deamination of cytosine to uracil are the two most prevalent mutagenic events in living cells. Two other products of oxidative damage to DNA are ring-saturated derivatives of pyrimidines and the production of major helical distortion. Estimates of the frequency of the most common oxidative damage (oxidation of guanine to 8-hydroxyguanine) are around 100–500 residues per day per human cell. These estimates are an order of magnitude less than that for hydrolysis. Again, repair enzymes exist to deal with the oxidative damage in a living cell, but such enzymes would be inactive in fossilized or preserved material. There are no estimates for the half-life of DNA due to oxidation. Other small molecules exist in cells that can cause damage to DNA. One of the best characterized is *S*-adenosylmethionine (SAM). This molecule causes the methylation of nucleic acids, and the activity of the molecule is analogous to a simple alkylating reaction. The main active sites of SAM are ring nitrogens and purine residues. The major products of such methylation are 7-methylguanine and 3-methyladenine. 3-Methyladenine is efficiently repaired by cellular enzymes while 7-methylguanine is not. The latter would be expected to permanently damage DNA. Estimates of the frequency of DNA methylation are on the order of 600 events per day per human cell. All these methods of DNA damage add up to the extrapolation that the half-life of DNA in fossilized material places a theoretical limit on the age of material that should yield DNA long enough to be retrieved in ancient DNA studies (Fig. 1). This limit, estimated by Lindahl at around 30 000 years, is based on extrapolations from experiments in which DNA was manipulated in aqueous solution and at specific pH and temperature values. Many of the fossilization processes discussed in this article are unique in that aqueous conditions are precluded and oxygen is excluded from the specimen at very early stages of preservation (as in amberization).

3 Studies in Molecular Paleontology

3.1 Microscopy Studies

The gross molecular structure of certain cellular organelles has been examined in several fossils. These studies are accomplished by thin sectioning of rocks containing putative microorganisms. The thin sections are then mounted on slides and observed under microscopy. Perhaps the oldest example is the discovery of bacterial filaments from 3.5-billion-year-old formations in western Australia that resemble the filaments from extant cyanobacteria. Other interesting findings from the study of ancient molecular-level structures are the presence of nuclear

membranes in billion-year-old fossils and the presence of colonial algal characteristics in microfossils from the gunflint chert in Michigan. Microscopic examination of amber-preserved fossils has revealed molecular-level information for these types of fossils. Transmission electron microscopy of amber-encased insects discloses astonishing detail at the cellular level. Nuclear membranes, endoplasmic reticula, and ribosomes are all visible in these ancient materials. In addition, chromatin can be detected in these micrographs. Scanning electron microscopy of amber-preserved insects also provides information at the molecular level, although the resolution is much lower.

3.2 Chemical and Amino Acid Studies

Studies of chemical fossils involve the characterization of biomarkers from a variety of microorganisms. Such biomarkers as pentacyclic triterpenoids, hopanes, sterols, pigment derivatives, and acyclic isoprenoids serve as signatures for unique fossil molecules and sediments. Isotopic composition of these biomarkers is a characteristic that is used to determine signature. For instance, *n*-alkanes can have even or odd numbers of carbon atoms. Even-number long-chain *n*-alkanes indicate an origin from plant waxes or algae. Odd-number *n*-alkanes indicate an environment high in salinity. An immediate application of this approach is to the petroleum industry, since *n*-alkanes and other biomarkers are the major components of petroleum. Abelson first showed in the 1950s that amino acids could be isolated from fossilized material. The fossils included representatives of mollusks of various ages, as well as dinosaurs, fish, and mammals from the Phanerozoic.

3.3 Protein and Nucleic Acid Studies

A few studies have reported the structural analysis of proteins from mineralized fossils. One of the earlier reports on recovery of proteins from fossils was the isolation of 80-million-year-old shell proteins from fossil mollusks. Although sequence information was not obtained from this material, the existence of intact proteins in these fossil shells was demonstrated by assaying the isolates on protein gels. Obtaining the actual sequences of fossil proteins is extremely difficult because a relatively large amount of starting material is needed for sequencing, and the amount of protein material in fossils is small. Despite this limitation, the N-terminal 15 amino acids from a small calcium binding protein has been sequenced from fossil oyster shells from the Holocene, Miocene, Cretaceous, and Jurassic epochs. The isolation and characterization of proteins and nucleic acids from ancient tissues for use in phylogenetic studies began with the pioneering work in Allan Wilson's lab at the University of California at Berkeley. Wilson and his colleagues, most notably Russel Higuchi, Ellen Prager, and Svante Pääbo worked on a variety of ancient tissue sources. Jerold Lowenstein, at the University of California at San Francisco, used the immunodiffusion and RIA techniques to examine the serum albumin of several extinct organisms. These techniques result in the establishment of immunological distances between taxa that can be used to infer relationships among organisms by means of distance algorithms. The organisms examined in these early studies were mastodon (*Mammut americanum*), woolly mammoth (*Mammuthus primigenius*), Steller's sea cow (*Hydrodamalis gigas*), Tasmanian wolf (*Thylacinus*

cynocephalus), and quagga (*Equus quagga*). The first report of DNA isolated and characterized from ancient tissue appeared in 1984 and concerned the isolation and cloning of DNA from the quagga. This pioneering study reported the isolation and characterization of a short fragment from the mitochondrial cytochrome *b* gene for this extinct mammal. DNA was isolated from a museum-preserved skin of a quagga that had been collected in the nineteenth century. The isolated DNA was cloned into a λ-vector and isolated by filter lift and hybridization techniques. Comparison of the DNA sequences from the quagga with sequences from other equids in a phylogenetic analysis revealed the close relatedness of the extinct quagga and Burchell's zebra. Publications followed concerning the isolation and characterization of nucleic acids from several animals such as the woolly mammoth, saber-toothed tiger, moa, Bronze Age rabbit, giant sloth, and marsupial wolf to name a few. One of the more interesting cases of phylogenetic inference obtained from ancient tissue sources is that of the marsupial wolf *Thylacinus*. The phylogenetic placement of this extinct marsupial species was determined using data from both immunoprecipitation and DNA sequencing studies. On the basis of morphological evidence (the remarkable similarity of dental and skeletal characters), some systematists have argued that this Australian genus is most closely related to another extinct group of South American carnivorous marsupials called the *borhyaenids*. Other systematists argue, that *Thylacinus* is most closely related to other Australian carnivorous marsupials (dasyurids). In an early study, Lowenstein and colleagues used immunoprecipitation RIA of serum albumin to demonstrate a close affinity of *Thylacinus* with the dasyurids, indicating that the Australian taxa were a monophyletic group and that the South American hypothesis was incorrect. Thomas and colleagues examined the DNA isolated from museum skins of *Thylacinus* for 219 bases of the mitochondrial cytochrome *b* gene and came to the same conclusion. More recently, much older material has been examined. The first study to break the million-year mark was one concerning the isolation and characterization of chloroplast DNA from fossilized magnolia leaves (17 million years old). This study approached the question of the relationship of the extinct magnolia species *Magnolia latahensis* to other closely related plants. Golenberg and colleagues used fossils from the Clarkia beds of Idaho as a source for *M. latahensis*. By means of polymerase chain reaction (PCR) techniques, they obtained an 820-bp long fragment from these ancient fossil plants and compared the sequences to several other plant species. The tree obtained from these analyses indicated that the extant species *Magnolia macrophyla* is the closest relative of *M. latahensis*. The isolation and characterization of DNA from Dominican amber-preserved insects, bacteria, and plants (30 million years old) and Lebanese amber-preserved insects (120 million years old) are examples of increase, by several orders of magnitude, of the limits of the age of ancient DNAs. Attempts to isolate DNA from Baltic amber (60–80 million years old), 80-million-year-old fossilized bone, 200-million-year-old fish fossils, and 400-million-year-old marine invertebrates have also been reported. A major challenge to the success reported in amber DNA studies has come from the work of Austin and colleagues. In a series of carefully planned and executed experiments, authentic ancient nucleic acids could not be isolated from several Dominican amber-preserved

specimens. The inability to isolate authentic ancient DNA from these specimens has curtailed the search for ancient DNA in these exquisitely preserved specimens and instead has focused the ancient DNA field on more recent specimens of age no greater than 100 000 years.

3.4 Molecular Archaeology: Humans and Plants

The primary source for most of the studies in ancient human DNA has consisted of bones or remains preserved in wet bogs, and natural and artificial mummies. Several studies have characterized DNA isolated from bone ranging from tens of years to thousands of years old (Table 3). Egyptian mummies and Amerindian populations were the initial major subjects of human ancient DNA studies. More recently, European and South Pacific populations have gained some attention from anthropologists studying ancient remains. Plant remains on the order of hundreds to thousands of years old have been examined. Examples of plant studies from archaeological studies concern the determination of patterns of domestication of certain cultivated plants such as maize and wheat.

3.5 Utility of Ancient DNA

Ancient DNA studies can be very useful in many aspects of modern biology. Systematists offer perhaps the best example of how these data can be useful. An amber termite study provides a useful example of the importance of nucleic acid information from fossils. Information on nearly 400 bases of the small ribosomal subunit (18S rDNA) was collected for a number of termites and their close relatives: mantids and cockroaches, and several outgroups. When a phylogenetic analysis using only the extant organisms is performed, the position

Tab. 3 Examples showing specimen ages of ancient DNA studies of humans.

Material extracted	*Locale*	*Age of specimen [years before present]*
Human mummies	Egypt	5000
	Andean area	6000
	Southwestern United States	2500
	Aleutian Islands	1400
	Australia	200
	Japan	900
Human bones	Oceanic and Amerindian peoples	3000
	Neanderthal	29 000
	Cro-Magnon	24 000
Human brains	Florida bogs	8000
Human parasites	*Mycobacterium tuberculosis* isolated from human mummies	2000–4000
	Yersinia pestis isolated from teeth of bubonic plague victims	400
	Clostridium from Andean mummy	1000
	Trypanosoma cruzi from Chilean mummy	4000

of cockroaches in the phylogenetic tree is ambiguous. When the extinct amber termite sequences are added to the analysis, however, the tree becomes resolved with respect to the placement of the cockroach. This result supports the contention, familiar to paleontologists, that fossil information is absolutely essential to an understanding of phylogeny in many groups. Two other examples involve the use of ancient DNA information to calibrate molecular clocks and the use of ancient DNA in population biology and conservation. Information about the sequences of extinct organisms is essential in attempts to more accurately determine absolute rates of divergence. If a fossil can be accurately dated and sequences from it can be directly compared to sequences from living organisms, more accurate rates for molecular clock estimates can be obtained. Population-level studies have already demonstrated the value of ancient DNA sequences in population, genetic, and taxonomic studies of wolves, tiger beetles, whales, and prairie chickens, to name a few. The ability to obtain information from extinct taxa and populations is essential in assessing relatedness among endangered species. Information about past genetic diversity has the potential to inform present-day conservation management decisions by providing context for biodiversity estimates and for assessing extinction risk. This information can also shed light on evolutionary processes (organismal and molecular) as well as human cultural practices and impacts.

4
Future of Ancient DNA Research: Authentication

The pioneering work of Higuchi and other members of the Wilson lab demonstrated that it was possible to obtain molecular information from long dead (and in some cases extinct) organisms. This possibility stimulated a field of research and the advent of PCR allowed for a dramatic expansion in the field. Early work focused on attempts to isolate DNA from various organisms, tissue types, and ages of specimens. For practical reasons, many studies relied on the use of high copy number mitochondrial gene regions; however, the information content of mtDNA is limited for studies addressing deep evolutionary questions, and more recent work has shown that it is possible to recover high copy nuclear gene regions such as microsatellites as well as single copy nuclear gene regions from ancient specimens. Ancient DNA research can be thought of as progressing along two nonexclusive paths. One is the application and refinement of ancient DNA extraction techniques in order to address a wide range of systematic, evolutionary, archaeological, and conservation biology questions. The other has focused on gaining an understanding of the molecules themselves by exploring questions such as what are the postmortem processes that degrade DNA, what are the theoretical temporal limits to DNA preservation and what preservation conditions are likely to protect DNA. These types of questions are informed by research in areas such as molecular biology, biochemistry, taphonomy, and forensics.

Given the degraded nature of ancient DNA, one question will remain: how reliable are the data obtained? Much of the recent work in ancient DNA research has focused on establishing means of authenticating ancient DNA sequences. One factor that bears on whether the data are reliable is the development and use of standardized procedures that prevent contamination of extracts with

modern (undergraded) DNA or PCR products and the monitoring of extracts for potential contamination. Several technical precautions are taken in the DNA isolation and PCR amplification stages.

1. Physical separation of labs (and equipment and supplies) where ancient DNA extractions are performed from labs where modern DNA extractions and PCR amplifications occur in order to minimize the potential for contamination of the ancient extract with high quality DNA or PCR product carryover; both of which could overwhelm a PCR of the ancient extract and lead to spurious results.
2. The use of disposable positive displacement pipette tips or plugged, coated pipette tips in all manipulations of DNA, buffers, and enzymes, as well as the use of bleach and UV irradiation of lab, equipment, and reagents.
3. Every PCR experiment involving ancient DNA should include positive and negative controls, run side by side with the experimental reactions to ensure the accuracy of the ancient DNA reaction. When attempting to amplify ancient tissue sources, a portion of the rock (in the case of mineralized fossils) or, in the case of amber, a portion of the amber that surrounds the fossil, should be treated as control material. The occurrence of PCR product from rock or amber without a fossil should be treated as information that the PCR reaction is faulty.
4. Direct and independent corroboration of results from one laboratory should be undertaken in other unrelated laboratories. The characterization of ancient nucleic acids currently relies on the polymerase chain reaction. Two problems specific to the PCR must be considered when examining ancient DNA. Since the *Taq* polymerase used in most PCR reactions preferentially inserts dATP into damaged regions of template DNA, ancient DNA sequences must be scrutinized for the unusual occurrence or frequency of deoxyadenines in amplified sequences. This is especially problematic when the starting template concentration is very low. Since errors introduced in early cycles of the PCR can become fixed, replicate PCR reactions should be performed to monitor sequence variation. The second problem concerns the ability of the PCR reaction to form chimeric molecules via PCR jumping, which occurs when damaged template DNA is degraded to sizes having only one of the priming sites present on the template, the other having been separated by damage. In these cases, the short templates can serve as primers on other pieces of template DNA, and *jumping* occurs across the damaged ends of the original templates. This process results in PCR products from template DNA that sometimes differ in size from the intended product. In addition, if some outside contaminant DNA is present in the PCR reaction, it is possible for a jump to occur from the intended target template to the contaminant. This scenario would produce a chimeric PCR product with sequences from the intended target on one end of the product and sequences from the contaminant on the other. Replication of PCR reactions, cloning of PCR products and replication of extractions can reveal these types of PCR-related problems.
5. Negative results should also be published, to let researchers know what has and what has not been attempted. The weight of the negative results can then

be used to assess the validity of any study that reports success at isolating ancient DNA. Reproducibility of results is a key aspect of reliability assessment. This is problematic for ancient DNA studies for which specimens are rare or unique, and in cases where even microhabitat-level variation in preservation conditions within a site can result in differences in the ability to recover DNA from contemporaneously buried specimens. Information on success rates for ancient DNA recovery from specimens of various ages and preservation conditions can also aid in deciding whether or not to undertake a study.

Another factor affecting whether the data are thought to be reliable is consideration of the likelihood of recovering authentic DNA from a given specimen. More detailed and sophisticated biochemical and chemical analysis of ancient tissues needs to be performed. Involvement of biochemists and chemists in the ancient DNA field to examine the nucleic acid content of ancient materials would be a positive move in this direction. A clearer yet incomplete picture is emerging of the factors that contribute to the postmortem preservation/degradation of DNA. Degradation of DNA over time does not appear to be a linear relationship. Studies have shown that the greater proportion of DNA damage occurs immediately following death and the rate of damage may decrease after that. Studies of contemporaneously buried humans in the same gravesite showed varying amounts of DNA recovered from bones, suggesting that microhabitat-level variation in taphonomic conditions can influence the effect of age on DNA recovery. Studies continue to explore the limits of age, starting materials, and preservation conditions that will yield recoverable DNA in an attempt to arrive at some general indicators of the likelihood of DNA recovery. Many studies continue to raise the upper limit of the age of specimens thought likely to yield authentic DNA (Table 3); however, extreme age of the specimens would suggest DNA recovery is unlikely. On the basis of detailed biochemical analyses of DNA degradation, Lindahl has expressed skepticism about the claims of several researchers on the isolation of DNA from materials older than 30 000 years. Those experiments were done in aqueous solution. Preservation conditions such as desiccation, cold temperature, and high salt may slow the degradation on DNA and allow for recovery of authentic DNA from specimens older than predicted on the basis of studies in aqueous solution. The presence of amino acids and extent of racemization have been used as proxies for indicating whether DNA recovery is likely. While there may be different factors affecting the preservation of proteins and DNA such as the protection of DNA sequestered in bones, teeth or amber, the continued presence of one macromolecule supports the possibility of recovery of another. An exploration of the extent of racemization has been suggested as a prerequisite before attempting ancient DNA extractions.

There is still much research to be done if general principles regarding the likelihood of DNA recovery are to be established. Another, perhaps the most important, factor affecting whether the data are thought to be reliable involves authentication of the resulting sequences. The verification procedures discussed above are indirect methods that minimize and monitor contamination, or infer the likelihood of authentic DNA recovery based on proxies such as age, condition, and the survival of other biomolecules. Detailed phylogenetic analyses provide a direct method of

authentication by identifying whether the resulting ancient DNA sequences are from the organism of interest and not the result of contamination or pseudogene amplification. The discovery of a genealogical relationship of DNA sequences from ancient sources with an extant, close relative is usually taken as verification of authenticity. Consequently, it is necessary to have as many exemplars or taxa as possible available for the phylogenetic analysis. The claim of the isolation and characterization of dinosaur DNA from mineralized fossils was open to question on the basis of these points concerning phylogenetic analysis. These studies highlight potential problems for ancient DNA work using mitochondrial sequences: (1) The presence of contaminating human sequences even in cases where standard ancient DNA lab precautions are observed. Research on organisms other than humans can minimize the impact of this potential problem through the use of species-specific primers in PCRs and by performing phylogenetic analyses on the resulting sequences to confirm that the study organism falls within an appropriate phylogenetic bracket. (2) The presence of copies of mitochondrial gene regions inserted into the nuclear genome. The dinosaur DNA claims and rebuttals showed that not only is human mitochondrial DNA contamination a problem, but that nuclear copies of human mitochondrial DNA can also be recovered. Comparisons of sequences based on Blast similarity scores alone may not be sufficient to rule out contamination owing to the potential for recovering nuclear insertions of mitochondrial sequences. These sequences may have accumulated additional substitutions since their insertion into the nuclear genome and might falsely be thought of as being sufficiently different from human DNA as to have come from the study organism. In fact, phylogenetic methods were used to disprove the authenticity of the dinosaur DNA claims.

See also Immunoassays; Origins of Life, Molecular Basis of.

Bibliography

Books and Reviews

Handt, O., Höss, M., Krings, M., Pääbo, S. (1994) Ancient DNA: methodological challenges, *Experientia* **50**, 524–529.

Herrmann, B., Hummel, S. (Eds.) (1994) *Ancient DNA. Recovery and Analysis of Genetic Material from Paleontological, Archaeological, Museum, Medical and Forensic Specimens*, Springer-Verlag, New York, Berlin, Heidelberg.

Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., Pääbo, S. (2001) Ancient DNA, *Nat. Rev. Genet.* **2**, 353–359.

Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds.) (1990) *PCR Protocols: A Guide to Methods and Applications*, Academic Press, New York.

O'Rourke, D., Hayes, M.G., Carlyle, S.W. (2000) Ancient DNA studies in physical anthropology, *Annu. Rev. Anthropol.* **29**, 217–242.

Pinna, G. (1985) *The Illustrated Encyclopedia of Fossils*, Facts on File, New York.

Poinar, G.O. (1994) The range of life in amber: significance and implications in DNA studies, *Experientia* **50**, 536–542.

Soltis, P.S., Soltis, D.E. (1993) Ancient DNA: prospects and limitations, *N. Z. J. Bot.* **31**, 203–209.

Primary Literature

Allard, M.W., Young, D., Huyen, Y. (1995) Detecting dinosaur DNA, *Science* **268**, 1192.

Austin, J.J., Ross, A.J., Smith, A.B., Fortey, R.A., Thomas, R.H. (1997) Problems of reproducibility–does geologically ancient DNA survive in amber-preserved insects? *Proc. R. Soc. London B* **264**, 467–474.

Bada, J.L., Wang, X.S., Hamilton, H. (1999) Preservation of key biomolecules in the

fossil record: current knowledge and future challenges, *Philos. Trans. R. Soc. London, Ser. B* **354**, 77–86.

Baron, H., Hummel, S., Herrmann, B. (1996) *Mycobacterium tuberculosis* complex DNA in ancient human bones, *J. Archaeol. Sci.* **23**, 667–671.

Bouzat, J.L., Lewin, H.A., Paige, K.N. (1998) The ghost of genetic diversity past: historical DNA analysis of the Greater Prairie Chicken, *Am. Nat.* **152**, 1–6.

Burger, J., Rosendahl, W., Loreille, O., Hemmer, H., Eriksson, T., Götherström, A., Hiller, J., Collins, M.J., Wess, T., Alt, K.W. (2004) Molecular phylogeny of the extinct cave lion *Panthera leo spelaea*, *Mol. Phylogenet. Evol.* **30**, 841–849.

Cano, R.J., Poinar, H., Poinar, G.O. (1992) Isolation and Partial Characterisation of DNA from the bee *Proplebeia dominicana* Apidae Hymenoptera in 25–40 million-year-old amber, *Med. Sci. Res.* **20**, 249–251.

Cano, J., Poinar, H., Pienazek, N.J., Acra, A., Poinar, G.O. (1993) Amplification and sequencing of DNA from a 120-135-million-year-old weevil, *Nature* **363**, 536–538.

Caramelli, D., Lalueza-Fox, C., Vernesi, C., Lari, M., Casoli, A., Mallegni, F., Chiarelli, B., Dupanloup, I., Bertranpetit, J., Barbujani, G., Bertorelle, G. (2003) Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6593–6597.

Cooper, A., Wayne, R. (1998) New uses for old DNA, *Curr. Opin. Biotechnol.* **9**, 49–53.

Cooper, A., Rhymer, J., James, H.F., Olson, S., McIntosh, C.E., Sorenson, M.D., Fleischer, R.C. (1996) Ancient DNA and island endemics, *Nature* **381**, 484.

Cooper, A., Mourer-Chauvire, C., Chambers, G.K., von Haeseler, A., Wilson, A.C., Pääbo, S. (1992) Independent origins of New Zealand moas and kiwis, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8741–8744.

DeSalle, R., Grimaldi, D. (1994) Very old DNA, *Curr. Opin. Genet. Dev.* **4**, 810–815.

DeSalle, R., Barcia, M., Wray, C. (1993) PCR jumping in clones of 30-million-year-old DNA fragments from amber-preserved termites (*Mastotermes electrodominicus*), *Experientia* **49**, 906–909.

DeSalle, R., Gatesy, J., Wheeler, W., Grimaldi, D. (1992) DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications, *Science* **257**, 1933–1936.

Drancourt, M., Aboudharam, G., Signoli, M., Dutour, O., Raoult, D. (1998) Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: an approach to the diagnosis of ancient septicemia, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12637–12640.

Gauthier, J., Kluge, A., Rowe, T. (1988) Amniote phylogeny and the importance of fossils, *Cladistics* **4**, 105–209.

Golenberg, E.M., Giannasi, D.E., Clegg, M.T., Smiley, S.J., Durbin, M., Henderson, D., Zurawski, G. (1990) Chloroplast DNA sequence from a Miocene *Magnolia* species, *Nature* **344**, 656–658.

Greenwood, A.D., Capelli, C., Possnert, G., Pääbo, S. (1999) Nuclear DNA sequences from late Pleistocene megafauna, *Mol. Biol. Evol.* **16**, 1466–1473.

Greenwood, A.D., Lee, F., Capelli, C., DeSalle, R., Tikhonov, A., Marx, P.A., MacPhee, R.D.E. (2001) Evolution of Endogenous Retrovirus-like Elements of the Woolly Mammoth (*Mammuthus primigenius*) and its relatives, *Mol. Biol. Evol.* **18**, 840–847.

Hagelberg, E., Clegg, J.B. (1991) Isolation and characterization of DNA from archaeological bone, *Proc. R. Soc. London B* **244**, 45–50.

Hagelberg, E., Clegg, J.B. (1993) Genetic polymorphisms in prehistoric Pacific islanders determined by analysis of ancient bone DNA, *Proc. R. Soc. London B* **252**, 163–170.

Hagelberg, E., Sykes, B., Hedges, R. (1989) Ancient bone DNA amplified, *Nature* **342**, 485.

Hagelberg, E., Bell, L.S., Allen, T., Boyde, A., Jones, S.J., Clegg, J.B. (1991) Analysis of ancient bone DNA: techniques and applications, *Philos. Trans. R. Soc. London, Ser. B* **333**, 399–407.

Handt, O., Krings, M., Ward, R.H., Pääbo, S. (1996) The retrieval of ancient human DNA sequences, *Am. J. Hum. Genet.* **59**, 368–376.

Handt, O., Richards, M., Trommsdorff, M., Kilger, C., Simanainen, J., Georgiev, O., Bauer, K., Stone, A., Hedges, R., Schaffner, W., Utermann, G., Sykes, B., Pääbo, S. (1994) Molecular genetic analyses of the Tyrolean Ice Man, *Science* **264**, 1775–1778.

Hansen, A.J., Willerslev, E., Wiuf, C., Mourier, T., Arctander, P. (2001) Statistical evidence for miscoding lesions in ancient DNA templates, *Mol. Biol. Evol.* **18**, 262–265.

Hedges, B., Schweitzer, M. (1995) Detecting dinosaur DNA, *Science* **268**, 1191–1192.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O.A., Wilson, A.C. (1984) DNA sequences from the quagga, an extinct member of the horse family, *Nature* **312**, 282–284.

Höss, M., Pääbo, S. (1993) DNA extraction from Pleistocene bones by a silica-based purification method, *Nucleic Acids Res.* **21**, 3913–3914.

Höss, M., Dilling, A., Currant, A., Pääbo, S. (1996) Molecular phylogeny of the extinct ground sloth *Mylodon darwinii*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 181–185.

Höss, M., Kohn, M., Pääbo, S., Knauer, F., Schröder, W. (1992) Excrement analysis by PCR, *Nature* **359**, 199.

Janczewski, D.N., Yuhki, N., Gilbert, D.A., Jefferson, G.T., O'Brien, S.J. (1992) Molecular phylogenetic inference from saber-toothed cat fossils of Rancho La Brea, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9769–9773.

Krings, M., Geisert, H., Schmitz, R.W., Krainitzki, H., Pääbo, S. (1999) DNA sequence of the mitochondrial hypervariable region II from the neandertal type specimen, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5581–5585.

Krings, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M., Pääbo, S. (1997) Neandertal DNA sequences and the origin of modern humans, *Cell* **90**, 19–30.

Leonard, J.A., Wayne, R.K., Cooper, A. (2000) Population genetics of ice age brown bears, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1651–1654.

Lindahl, T. (1993a) Instability and decay of the primary structure of DNA, *Nature* **362**, 709–715.

Lindahl, T. (1993b) Recovery of antediluvian DNA, *Nature* **365**, 700.

Loreille, O., Orlando, L., Patou-Mathis, M., Philippe, M., Taberlet, P., Hanni, C. (2001) Ancient DNA analysis reveals divergence of the cave bear, *Ursus spelaeus*, and brown bear, *Ursus arctos*, lineages, *Curr. Biol.* **11**, 200–203.

Oota, H., Saitou, N., Matsushita, T., Ueda, S. (1999) Molecular genetic analysis of remains of a 2,000-year-old human population in China and its relevance for the origin of the modern Japanese population, *Am. J. Hum. Genet.* **64**, 250–258.

Orlando, L., Leonard, J.A., Thenot, A., Laudet, V., Guerin, C., Hanni, C. (2003) Ancient DNA analysis reveals woolly rhino evolutionary relationships, *Mol. Phylogenet. Evol.* **28**, 485–499.

Pääbo, S. (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 1939–1943.

Poinar, H.N., Cano, R.J., Poinar, G.O. (1993) DNA from an extinct plant, *Nature* **363**, 677.

Poinar, H.N., Höss, M., Bada, J.L., Pääbo, S. (1996) Amino acid racemization and the preservation of ancient DNA, *Science* **272**, 864–866.

Poinar, H., Kuch, M., McDonald, G., Martin, P., Pääbo, S. (2003) Nuclear gene sequences from a late Pleistocene sloth coprolite, *Curr. Biol.* **13**, 1150–1152.

Roy, M.S., Geffen, E., Smith, D., Wayne, R.K. (1996) Molecular genetics of pre-1940 red wolves, *Conservation Biol.* **10**, 1413–1424.

Runnegar, B. (1991) Nucleic acid and protein clock, *Philos. Trans. R. Soc. London B* **333**, 391–397.

Stankiewicz, B., Poinar, H., Briggs, D., Evershed, R., Poinar, G. (1998) Chemical preservation of plants and insects in natural resins, *Proc. R. Soc. London B* **265**, 641–647.

Stone, A.C., Stoneking, M. (1999) Analysis of ancient DNA from a prehistoric Amerindian cemetery, *Philos. Trans. R. Soc. London B* **354**, 153–159.

Thomas, R.H., Schaffner, W., Wilson, A.C., Pääbo, S. (1989) DNA phylogeny of the extinct marsupial wolf, *Nature* **340**, 465–467.

Thuesen, I., Engberg, J. (1990) Recovery and analysis of human genetic material from mummified tissue and bone, *J. Archaeol. Sci.* **17**, 679–689.

Tuross, N. (1994) The biochemistry of ancient DNA in bone, *Experientia* **50**, 530–535.

Willerslev, E., Hansen, A.J., Binladen, J., Brand, T.B., Gilbert, M.T., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D.A., Cooper, A. (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments, *Science* **300**, 791–795.

Woodward, S.R. (1995) Detecting dinosaur DNA, response, *Science* **268**, 1194.

Woodward, S.R., Weyand, N.J., Bunnell, M. (1994) DNA sequence from Cretaceous period bone fragments, *Science* **266**, 1229–1232.

Zischler, H., Höss, M., Handt, O., von Haeseler, A., van der Kuyl, A.C., Goudsmit, J., Pääbo, S. (1995) Detecting dinosaur DNA, *Science* **268**, 1192–1193.

Parthenogenetic Stem Cells

Jason D. Hipp[1], *Josyf C. Mychaleckyj*[1], *Jose B. Cibelli*[2], *J. David Wininger*[1], *Kathleen A. Grant*[1], *and Kent E. Vrana*[3]
[1] *Wake Forest University School of Medicine, Winston-Salem, NC, USA*
[2] *Michigan State University, MI, USA*
[3] *Penn State University College of Medicine, Hershey, PA, USA*

1 Background 73
1.1 A Brief History of Parthenogenesis Research 73
1.2 Mammalian Parthenogenesis 73
1.3 Genomic Imprinting 74
1.4 Human Ovarian Teratomas 75

2 Parthenogenetic Stem Cells 76
2.1 Characteristics of Stem Cells 76
2.2 Primate Parthenogenetic Stem Cell Lines 77
2.3 Characteristics of Nonhuman Primate Parthenogenetic Stem Cells 77
2.4 Neuronal Differentiation 79

3 The Influence of Imprinting on Cellular Growth and Differentiation 79
3.1 Imprinting and Neuronal Phenotype 79
3.2 Imprinting and Parthenogenesis 80

4 The Promise of Parthenogenetic Stem Cells 81
4.1 Parthenogenetic Stem Cell Therapy 81
4.2 Primate Parthenogenetic Stem Cells are a Unique Tool to Study Epigenetic Regulation 82

Bibliography 83
Books and Reviews 83
Primary Literature 83

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

Keywords

Chimera
An individual organism composed of a mixed population of parthenogenetic and normally fertilized cells.

Genomic Imprinting
The differential expression of gene alleles, dependent on their parent of origin.

Parthenogenesis
The development of a new individual organism from an unfertilized ovum (an activated egg).

Parthenote
An embryo created by parthenogenesis.

Teratoma
A tumor composed of a variety of mixed tissues derived from two or more germ layers in normal embryonic development.

Uniparental Disomy
The inheritance of a diploid genome from a single diploid parent (the result of parthenogenesis).

One of the most exciting and contentious issues in biomedical science today is the potential use of stem cells as a therapeutic intervention for a wide range of diseases with pathologies that involve cellular destruction. Stem cell therapy is currently practiced, or under active research, to repair the effects of neuromuscular degeneration, autoimmune response, ischemic attack, malignant carcinoma, congenital defects, metabolic disorders, and alcohol-related pathologies. When stimulated in a proliferative environment, stem cells may also provide the raw material for large-scale *ex vivo* tissue regeneration. There are several sources of human stem cells for laboratory culture, but the primary source of pluripotent human embryonic stem cells is the inner cell mass of *in vitro* fertilized embryos. Despite the potential rewards to medicine, experimentation using cells from viable human fetuses has triggered global political and ethical controversy that has slowed down the pace of research. An alternative source of human embryonic stem cells that circumvents some of these issues would significantly enhance research in this area. Parthenogenetically derived stem cells may represent such a resource. Parthenogenesis involves activation of the oocyte without sperm, and produces a nonviable blastocyst. That is, the activated egg develops to the blastocyst (and so can be used to generate stem cells), but it will not produce a viable pregnancy in mammals. The unique genesis of the embryo, and its arrested development at blastocyst, may afford sufficient moral and political latitude to enable its acceptance

as an experimental entity. Research is ongoing to assess the genetic and phenotypic similarity of parthenogenetic stem cells (PSCs) to other stem cell types from mammalian models and to human stem cells derived from fertilized embryos.

1 Background

1.1 A Brief History of Parthenogenesis Research

Parthenogenesis refers to "the production of an embryo from a female gamete without the concurrence of a male gamete, and with or without eventual development into an adult." Parthenogenesis – Greek for virgin birth – is a common form of reproduction for insects such as aphids, flies, ants, and honeybees, but is also known to occur in vertebrates, including lizards, snakes, fish, birds, reptiles, and amphibians. Aristotle might have been one of the first to have recognized this form of reproduction, in fish and bees, as noted in his *Generation of Animals* (340 B.C.), but it was not until 1695 that van Leeuwenhoek documented this form of reproduction in aphids. Charles Bonnet described cyclical parthenogenesis in *Contemplation de la Nature* (1740), followed by Dzierzon's observation 100 years later that drone bees are derived from unfertilized eggs (1845). The term parthenogenesis was first used by Richard Owen in 1849. The first demonstration of *in vitro* artificially stimulated parthenogenesis was made by Jacques Loeb, who was able to activate oocytes from sea urchins and frogs by pricking them with a needle or by changing the ambient salt concentration. In the 1950s, parthenogenesis was classified into subtypes according to the sex of the resulting offspring. Arrhenotoky is parthogenesis where unfertilized ova produce males, while fertilized ova lead to females, as for example in ant and honeybee colonies; in thelytoky, the unfertilized ova become females and in deuterotoky, they may become offspring of either sex. Pincus demonstrated parthenogenetic activation of mammalian eggs in 1939, using temperature and chemical stimuli. To date, parthenogenetic activation of eggs has been accomplished in most mammals, including mice, goats, cows, monkeys, and humans. Plachot et al. were able to identify human parthenogenetic activation events by examining 800 human oocytes. They showed that 12 were spontaneously parthenogenetically activated, and of these, four had undergone normal cleavage. Although there have been no reports of naturally occurring human parthenotes, a human parthenogenetic chimera was recently recognized. The juvenile patient presented as developmentally delayed, with apparent sex reversal, and entirely parthenogenetic blood leukocytes. This finding confirmed the viability of chimeras in higher mammals, as suggested by successful laboratory experiments in mice over the previous two decades (as described in later sections).

1.2 Mammalian Parthenogenesis

There is no confirmed example of natural mammalian parthenogenetic reproduction, but mammalian oocytes can be artificially induced to undergo parthenogenesis *in vitro* by a two-step protocol involving

electroporation and/or treatment with a chemical agent (ionomycin, ethanol, or inositol 1,4,5-triphosphate) to elevate transient Ca^{2+} levels, followed by application of an inhibitor of protein synthesis (cycloheximide) or protein phosphorylation (6-dimethylaminopurine). Success rates and developmental survival time appear to be organism dependent. Mouse parthenotes are capable of developing past postimplantation stages *in vivo*; pig parthenotes have developed up to postactivation day 29 (limb bud stage, past the early heart beating stage); rabbit parthenotes until day 10 to 11, while primates (*Callithrix jacchus*) have only been shown to implant. The reason for this arrested development is believed to be due to genomic imprinting. In fertilized zygotes, maternal and paternal haploid genomes are epigenetically distinct, and both sets are required for successful development. That is, metastable chromosome modifications in the form of methylation, for example, exhibit unique patterns in sperm versus eggs. As a result, they carry unique patterns of gene expression into the embryo. Since all genetic material in parthenotes are of maternal origin, the paternal imprinting component is absent and this prevents proper development of extraembryonic placental tissues (from the trophectoderm) whose expression is regulated by the male genome. In most mammals, (including primates), just prior to ovulation, oocytes are arrested at the metaphase II stage. Cytogenetic microscopy shows the presence of a $2n$ polar body under the zona pellucida and a $2n$ protonucleus in the cytoplasm. After chemical activation to mimic the effects of sperm penetration on changes in cellular Ca^{2+} gradient, the cell does not complete meiosis II. Instead, the second polar body is never extruded, resulting in a diploid protonucleus derived from two sets of sister chromatids. These chromatids then begin to divide mitotically. On balance, then, the parthenote is related to the maternal oocyte by uniparental disomy (Fig. 1).

1.3 Genomic Imprinting

Genomic imprinting is an epigenetic regulatory process that imposes a parent-of-origin control over gene allele expression according to a dynamic program that is dependent on the stage of development and tissue type. A memory of the parental source of each disomic copy of imprinted genes in the genome is preserved, and this is coupled with a mechanism to effect mono-allelic differential expression of the gene copies based on the parent of origin, such as the paternally expressed gene, insulin-like growth factor 2 (IGF2). The time and tissue variability of this process is evident, even as early as oogenesis. The primary mechanism for imprinting is methylation of cytosine residues in DNA CpG di-nucleotides, but histone modifications such as acetylation and methylation cooperate in chromatin modification, reinforcing local genome silencing. It is believed that as many as 200 genes are imprinted in humans, although only about 50 imprinted genes have been identified in mice (with some conservation in humans). The evolutionary reason for the development of imprinting is not clear, but a number of theories have been posited. The "Conflict Theory" is one of the more popular and enduring explanations of why genes are imprinted. Empirically, genes that are paternally expressed are often those genes involved in fetal growth, while genes that are maternally expressed restrict growth. In the population genetic theory, the resource burden for carrying a fetus rests disproportionately on females

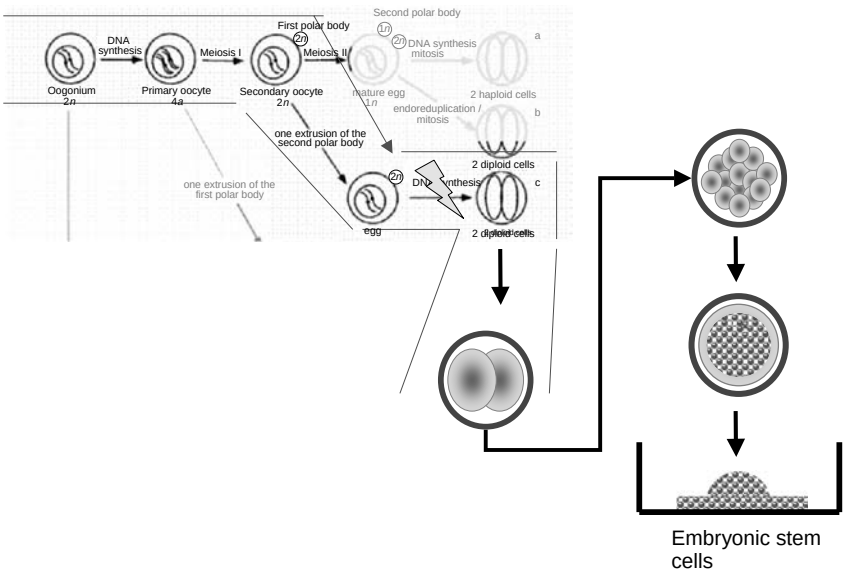

Fig. 1 The process by which stem cells are derived from parthenogenesis. An oocyte is arrested at Metaphase I, and is the artificially activated, bypassing meiosis II. The resulting diploid cells begin to mitotically divide to a morula, and further develop to a blastocyst. At this point, however, the blastocyst fails to form a functional placenta *in vivo*. The inner cell mass (ICM) is then harvested and plated and produces cultured stem cells.

of a species in eutherians (placental mammals) and metatherians (marsupials), and female mammals are known or presumed to be promiscuous. Hence, by carrying fetuses from different fathers, paternal evolutionary drive would seek to maximize resource benefits for his fetus. The counterpoint is that the mother is placed at risk, being drained of much needed nutrients and potentially carrying oversized fetuses; hence, she benefits from rationing the amount of her resources that a fetus can extract. Evidence from seed development reinforces the notion of an intragenomic parental conflict, whereby imprinting intercedes to differentially balance the selfish interests of each sex's genome during the development of the progeny.

It is believed that the parthenotes are not capable of developing to term because they fail to develop the necessary extraembryonic tissues – trophectoderm and primitive endoderm. In contrast, androgenotes (created by the fusion of two sperm nuclei, and of purely paternal origin) develop into a structure consisting of a trophoblast and yolk sac regulated by the male genome. These structures resemble hydatidiform moles (trophoblastic tissue only), which are formed when sperm fertilizes an enucleated egg.

1.4 Human Ovarian Teratomas

In the absence of natural primate parthenotes, human ovarian teratomas may represent one of the few clinical correlates available, albeit as a result of a different (oncogenic) transformation. There is controversy about the histogenesis of teratomas, with apparent subgroups based on gonadal versus extragonadal

site, and maturity of the tumor. Ovarian teratomas represent 10 to 15% of all ovarian tumors, and result from abnormal development of a primary oocyte. Most mature ovarian teratomas (dermoid cystic type predominantly) appear to be 46, XX diploid with a Barr body present, and are believed to be genetically equivalent to parthenotes. They may arise from tumorigenic events at meiosis I and/or II, genome reduplication in a mature ovum, or fusion of two ova, with clinical and genetic evidence to support all of these modes. Mature cystic teratomas are typically composed of adult-type tissues usually representing all three germ layers, including teeth, bone, cartilage, adipose, thyroid, and soft brain tissue. Rarely seen developed organs and tissues such as bowel, appendix, skull, vertebrae, limb buds, external genitalia, pancreas, prostate gland, and eyes will be observed. Ectodermal elements such as skin and its appendages predominate, but neuroectodermal components are also present. These teratomas may occasionally contain foci of immature or fetal tissues, and present with a complex mix of heterogeneous tumors. In contrast, the much rarer immature ovarian teratomas are more likely to have abnormal karyotypes reflecting the genome instability associated with malignant cells, and have variable percentages of immature embryonic-like tissues, together with mature components. Most of the immature tissues are neuroectodermal in this case, and their predominance directly reflects the histological grading of these cancers. The common form of treatment for immature teratomas consists of chemotherapy that not only kills the rapidly dividing cells but also induces a proportion of them to differentiate. The importance of this observation cannot be overstated; chemotherapy generally limits neoplastic growth by blocking the cell cycle, and its effects are similar to the differentiation induced in *in vitro* cultured parthenogenetic stem cells (PSCs) by the removal of growth factors, as described below. In very rare cases, pluripotent fetiform ovarian teratomas can organize differentiated tissues into a recognizable, but malformed, human fetus or homunculus, thereby showing that pseudonormal programmed development is possible from purely maternal tissues.

The relationship between the etiology and development of the different subgroups of ovarian teratomas and PSCs is not well understood, but there are obvious parallels worthy of further study. In mature ovarian teratomas, the benign nature of the tumor, the observation of a normal diploid karyotype, the involvement of meiotic failure at meiosis II, and the manifest pluripotency of the transformed oocytes have obvious analogy with PSCs. Immature ovarian teratomas, however, recapitulate the tendency of PSCs to commit to a neuronal lineage and the effect of withdrawal from the neoplastic cell cycle in promoting differentiation.

2 Parthenogenetic Stem Cells

2.1 Characteristics of Stem Cells

Stem cells are, by definition, cells that have the ability to replicate indefinitely, and can differentiate into multiple cell types. Embryonic stem cells are pluripotent and have the ability to become most, if not all, cell types in an organism. In contrast, adult stem cells have reduced plasticity, and are usually uni- or, at best, multipotent. They

are also more sensitive to *in vitro* culture conditions, and have shorter survival times than embryonic stem (ES) cells. Currently, there are several sources of stem cells: from *in vitro* fertilization, somatic cell nuclear transfer (cloning technologies), adult tissues, cord blood and germ cells.

2.2 Primate Parthenogenetic Stem Cell Lines

Our experiments in developing parthenote stem cell lines have used the cynomolgus macaque model (*Macaca fascicularis*). After harvesting oocytes from a female hyperovulating macaque, the cells undergo *in vitro* maturation to reach metaphase II. Eight days following oocyte activation with ionomycin and 6-dimethylaminopyridine, 14% of ova reach blastocyst stage, where they arrest (Fig. 1). Transplantation back to a uterus is unable to continue development at this point. The inner cell mass is isolated using immunosurgery; antibodies are added to bind to the trophectoderm, followed by complement proteins. This procedure destroys the trophectoderm leaving the inner cell mass, which harbors precursors to ES cells. These are transferred to grow on mitotically inactivated mouse feeder cells, to keep the stem cells undifferentiated and replicating (Fig. 1).

The creation of embryonic-like stem cells from oocytes is relatively inefficient. Only two primate (one monkey and the other human) PSC lines have been created, and as far as we are aware, only the cynomolgus macaque PSC line is still in culture. The process, described above, was used to generate a parthenogenetic blastocyst (Fig. 2a) and PSCs (Fig. 2b). A human PSC line was generated from 25 human oocytes, yielding 5 blastocysts, and 1 stem cell line that lasted for 2 passages. The difficulties in generating primate PSC lines are probably due to the complexities of higher-order genomic imprinting. However, when PSCs are differentiated into adult tissues, they appear fully functional. Characterization of the differentiated tissues from PSCs in comparison to analogous tissues derived from normal ES cells awaits further study.

2.3 Characteristics of Nonhuman Primate Parthenogenetic Stem Cells

In spite of the nonviability of nonhuman primate (NHP) parthenotes, the extracted stem cells seem to assume the morphology and functional behavior of human ES cells. Morphologically, PSCs possess a small cytoplasmic/nuclear ratio, numerous and prominent nucleoli, and cytoplasmic lipid bodies. Karyotype analysis reveals that the cells are fully diploid with a $2n = 40 + 2$ macaque chromosome complement. Functionally, NHP PSCs express the enzymes alkaline phosphatase and telomerase, showing a loss of the high telomerase activity two weeks after the initiation of a differentiation protocol; this corresponds to loss of pluripotency. Also, like human ES cells, they express the following traditional embryonic stem cell markers: POU domain, class 5, transcription factor 1 (Oct-4), Stage Specific Embryonic Antigen (SSEA-3/4), as well as Tumor Rejection Antigen (TRA 1-60/1-81); moreover, they are negative for SSEA-1 (present only in mouse ES cells, and not in human or primates ES cells). The NHP PSCs express major histocompatibility Class I and II antigens at levels comparable to normal peripheral blood leukocytes after treatment with Interferon-gamma (IFN-gamma). The treatment did not increase

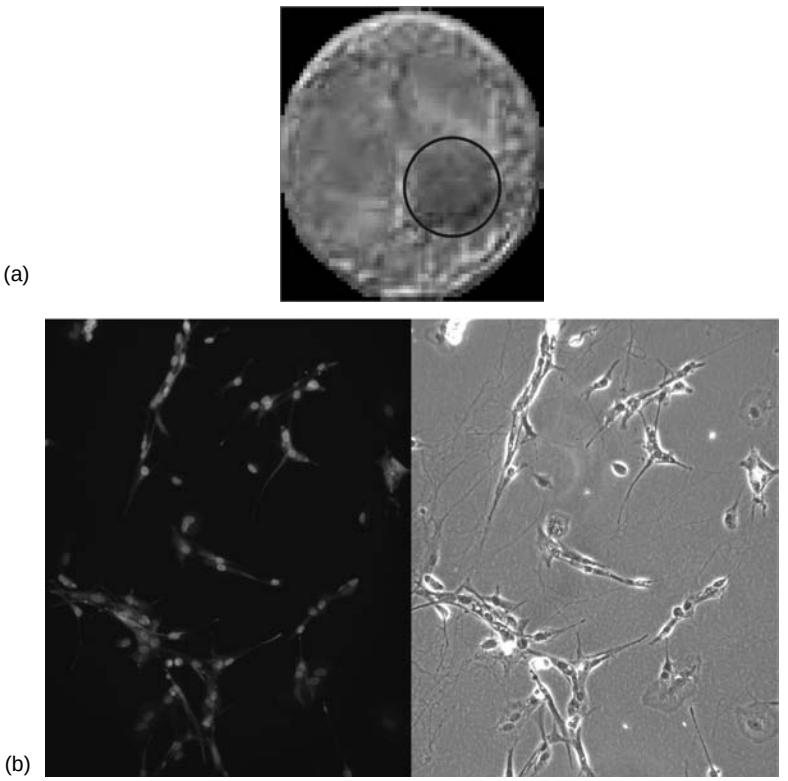

Fig. 2 (a) **Photomicrograph of a parthenogenetically derived blastocyst**. Circled in blue is the inner cell mass (ICM), the source of embryonic-like stem cells. The significance of this is that it is the first primate parthenote in which a stable stem cell line has bee derived. (b) **Photomicrograph of neural precursor cells derived from parthenogenetic stem cells**. Cells stained for nestin (marker for the developing nervous system) (top panel) accompanied by a phase contrast image of all cells (bottom panel). (See color plate p. xxiv.)

class II expression, as one would expect from a non-antigen-presenting cell.

The cells possess the fundamental characteristics of ES cells. They have embryonic-like replicative ability, and have been propagated *in vitro* in an undifferentiated state for up to 14 months. *In vitro* differentiation produces high percentages of dopaminergic and serotoninergic neurons, spontaneously contractile cardiomyocyte-like cells, smooth muscle, beating ciliated epithelia, adipocytes, and several types of epithelial cells. Resulting neurons have been shown to express a voltage-dependent sodium channel, a functional neuronal marker. This multipotency is recapitulated *in vivo*, since injection of NHP PSCs into immunocompromised mice induces formation of teratomas containing derivative tissues from the three germ layers (ectoderm, endoderm, and mesoderm), including cartilage, muscle, bone, neurons, skin, hair follicles, and intestinal epithelia (Fig. 3). The preference of these primate PSCs to differentiate into neuronal tissues has been noted by

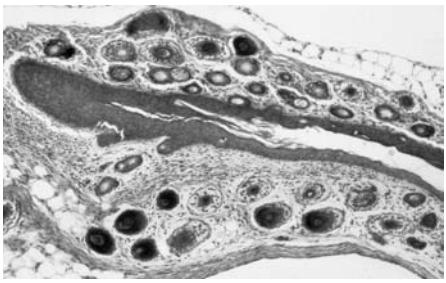

Fig. 3 Teratoma derived from injection into an immunocompromised (SCID) mouse. This H & E stained section demonstrates the *in vivo* differentiation capacity (pluripotency) of parthenogenetically derived stem cells (courtesy of Advanced Cell Technology).

several groups, but the reasons for the underlying preference are not well understood. One possible explanation is that it is a consequence of purely maternal genomic imprinting, and may represent the lack of epigenetic balance that would be conferred by paternally imprinted genes.

2.4 Neuronal Differentiation

In the course of experiments testing differentiation protocols for PSCs, we have identified an intermediate stage along the PSC differentiation trajectory that has unique properties (Fig. 2b). After plating NHP PSCs on a basement membranelike matrix and supplementing the medium with basic fibroblast growth factor (bFGF) and epidermal growth factor (EGF), the cells begin to commit to the neuronal lineage but halt at a transitory state. Cells in this intermediate state stain positive for the neuronal markers nestin and microtubule-associated protein-2. They are negative, however, for choline acetyl transferase, dopamine-beta-hydroxylase, and NeuN, and they do not express ion channels. These Nestin$^+$ precursors are highly replicative (5- to 8-fold increase over a 9-day period), and do not require a feeder layer to grow and remain undifferentiated.

Nestin$^+$ precursor cells can be induced to differentiate to neurons by removing bFGF and EGF, and adding ascorbic acid for a minimum of five days. The resulting differentiated cells represent the neuroepithelial lineage. Almost all of these neurons express TUJ1 (beta-tubulin III), and up to 25% of the TUJ1$^+$ cells coexpress tyrosine-hydroxylase. This latter enzyme marker is considered diagnostic for catecholaminergic neurons (dopamine, norepinephrine, and epinephrine). Furthermore, HPLC analysis of culture media following a depolarizing KCl-buffer identifies the release of the neurotransmitters dopamine and serotonin from the cells. After one week of differentiation, about half of the cells that have a neuronal morphology and begin to express voltage-dependent sodium channels that can be blocked by tetrodotoxin.

3 The Influence of Imprinting on Cellular Growth and Differentiation

3.1 Imprinting and Neuronal Phenotype

There are several developmental disorders that are caused by errors in imprinting, such as Silver–Russell syndrome, Beckwith–Wiedemann syndrome, Angelman's syndrome, and Prader–Willi syndrome. These disorders have extensive associated neurological and behavioral pathologies. Each of these syndromes results from structural disruption of a genome locus

that is subject to maternal or paternal imprinting. Of especial interest here is Prader–Willi Syndrome because it offers an *in vivo*, human clinical model representing only maternal allele(s) in the region of chromosome 15q11-13 (25% have two maternal alleles, 75% have one maternal allele and a deleted paternal allele). It appears that having two maternal 15q11-13 regions is responsible for the neurological and behavior abnormalities. This clinical example correlates with the previous parthenogenetic-chimera studies described below.

By creating mouse chimeras composed of PSCs and biparental ES cells, Keverne and colleagues observed that parthenogenetically derived cells are found predominantly in the neocortex, striatum, and, but not in the hypothalamus, septum, preoptic areas, and brainstem. These studies also showed that parthenogenetic chimeras have brains that are much bigger in size relative to their body weight, although they appear to be anatomically and functionally normal. The area of the brain that is significantly larger, however, is telencephalon. Other chimera studies have shown a restriction of parthenogenetically derived cells in muscle tissue and growth retardation that is proportional to the overall contribution of parthenogenetic cells. However, if the PSCs are cultured for a brief period of time *in vitro*, then used to create chimeras, no pre- or postnatal growth retardation is observed, and higher overall levels of PSCs are detected than in other studies. However, the cells distribute among tissues as per noncultured PSCs. Disruption of imprinting by the *in vitro* culturing step would contribute to this perturbation of cellular distribution.

From an evolutionary perspective, an increase in brain size is correlated with an increase in the complexity of imprinting. For example, the neocortices of prosimians (lower-order primates) are 54% of their total brain weight, while the neocortices of opossums are 22%, and the neocortices of basal insectivores occupy only 13%. This forebrain evolution also correlates with a switch from hormonal dependence to a more cognitive control of behavior. Taken together, these observations suggest that maternal imprinting has an effect on the development and function of the brain. Since the evolution of genetic imprinting coincides with an exponential growth of brain to body size, has evolution coupled its need for genetic diversity to forebrain evolution? Could maternal imprinting explain why PSCs have a tendency to form neurons?

3.2 Imprinting and Parthenogenesis

After the creation of PSCs, the cells and their tissue derivatives seem to assume the traditional phenotype and characteristics of the biparental stem cells. Since imprinting is temporal and tissue specific, and given that many imprinted genes are involved in fetal growth, one would expect some unusual growth characteristics of PSCs. This is true in the very early stages. For example, we have noticed that human PSCs do not spread out and grow like biparental stem cell colonies. Although the inner cell mass attaches to the mouse feeder cells, the cells do not begin dividing like true ES cells. The cell masses have survived in culture for up to several weeks but eventually die. However, in differentiating monkey PSCs into adult tissues, there is evidence that they appear fully functional.

It appears that lack of paternal imprinting does not affect cell phenotype, behavior, and function, once the inner cell

mass attaches to the feeder layer. The dangers of abnormal imprinting appear to be less severe than once thought. Confirming our observations, Szabo and coworkers have demonstrated the presence of mRNA from imprinted genes, where expression of mRNA should have been absent. Many of the stringent effects of imprinting are overcome when the oocytes are artificially activated with chemicals or electroporation. Imprinting appears to play a less significant role as a PSC differentiates, as confirmed by the diversity and functionality of cells generated from PSC. Imprinting may still be affecting the behavior of the cells, but it is only having a minor effect on the differentiation potential by altering PSC responses to developmental cues, such as ligands, growth factors, and so on.

In addition to the Conflict Theory, genomic imprinting might be a mechanism by which to prevent spontaneous births (via parthenogenesis). A possible reason why eutherians cannot reproduce parthenogenetically is that their sexual phenotype is determined by genetic dichotomy, while the phenotype of parthenogenetically reproducing species is influenced by environmental or behavior selection pressures. Since the species that commonly reproduce via parthenogenesis have equal-sized sex chromosomes, it makes it much more difficult to maintain genetic diversity. The ability to inactivate the X chromosome demonstrates how the complexity of genetic imprinting correlates with the complexity of the organism. This difference is best demonstrated by the insulin-like growth factor 2 receptor (Igf2r) gene being imprinted in the mouse and not in the humans. Many imprinted genes are differentially expressed in early development (fetal growth and function of the placenta), and then become biallelically expressed in the adult.

4 The Promise of Parthenogenetic Stem Cells

4.1 Parthenogenetic Stem Cell Therapy

The unprecedented research interest in embryonic and adult stem cells lies in their potential to replace diseased cells or regenerate organs that have been lost as a sequela to disease, or through a traumatic event. Research on PSCs is still in its infancy, with much that is poorly understood about the mechanisms underlying the fundamental properties of self-renewal and the cooperative cellular programs that regulate differentiation. It appears *prima facie* that the lack of paternal imprinting does not affect the self-renewal or pluripotency of the stem cells isolated from the inner cell mass of the parthenote, the cells that are matrilineally regulated. The pluripotency of the stem cells could offer therapy options for a variety of diseases, but the tendency of the PSCs to commit to the neuronal lineage under *in vitro* culture conditions suggests that interventions in neurodegenerative disorders such as Alzheimer's and Parkinson's diseases as well as alcohol-induced Wernicke–Korsakoff syndrome might be the most tractable therapy options.

A major advantage of PSCs over ES cells derived from a normally fertilized conceptus is their unique genome inheritance as a result of uniparental disomy, and the implications for immunocompatibility. Heterologous transplantation of adult stem cells, or ES cells derived from an anonymous stem cell line or from cord blood, requires immunological donor/recipient matching. Autologous transplantation using cord blood stem cells requires long-term storage,

is costly, and requires advanced planning. Although meiotic recombination may shuffle the allele combination in the parthenotes, relative to the somatic genome of the maternal parent, PSCs should be highly immunocompatible with the oocyte donor female, and offer a treatment option for women over the period that they retain an ovarian reserve, and the oocytes can be parthenogenetically activated. Alternatively, one could envisage generation and cryogenic storage of PSC from all women of childbearing age. Moreover, androgenote generation – derived from the fusion of two sperm nuclei in an enucleated egg – while currently undeveloped, provides the possibility of creating these types of cells for males.

Finally, as mentioned in the introduction to this chapter, PSCs offer a unique stem cell resource that literally does not require the creation of fertilized and viable embryos for research, nor the destruction of unused embryos from assisted reproductive technology. The hope is that this might abrogate many of the ethical concerns surrounding research on human ES cells and provide the necessary flexibility to allow research to continue. Parthenotes are not free from ethical controversy though, and are viewed by some in society as artificial entities that in some sense represent "tampering with nature." The proof that a human chimera is possible by natural events, and by existing medical practice in the treatment of human ovarian teratomas, has a direct bearing on the complex ethical status of human parthenotes. As discussed above, the spontaneous *in vivo* tumorigenic event creates a similar phenotypic entity for classes of teratomas, as compared with *in vitro* parthenogenetic activation. Since most mature ovarian teratomas appear to be composed of 46, XX diploid chromosomes with Barr Body present, the two cell types can also be argued to be cytogenetically equivalent. Despite their nonviability for development of pregnancies, the idea that parthenotes are artificial is not as axiomatic as might be thought. Furthermore, since there is *de facto* acceptance of experimentation using teratoma tumor tissue and surgical destruction after resection, this lends some legitimacy to experimentation on parthenotes. These contradictions await reconciliation in a comprehensive ethical framework.

4.2 Primate Parthenogenetic Stem Cells are a Unique Tool to Study Epigenetic Regulation

Primate PSCs serve as a unique discovery tool to study the epigenetic mechanisms involved in development and gene expression. The unilineally inherited diploid genome enables the detailed influence of maternal imprinting to be studied in isolation, and by sampling cells from time points along the path of differentiation, and with differing phenotypes, a multidimensional profile of imprinting by time and lineage will be possible. Further comparison with stem cells derived from embryos with a biparental origin will allow the maternal and paternal contributions to be disentangled at these data points, giving insight into their relative influence at different stages of cell maturity and lineage committal.

See also Mammalian Cell Culture Methods; Nuclear Transfer for Cloning Animals; Principles and Applications of Embryogenomics.

Bibliography

Books and Reviews

Aristotle. *Generation of Animals* (circa 340 BP), translated by Peck A.L. (1953) The Loeb Classical Library, Harvard University Press, Cambridge, MA.

Baroux, C., Spillane, C., Grossniklaus, U. (2002) Genomic imprinting during seed development, *Adv. Genet.* **46**, 165–214.

Dzierzon, J. (1845) Uber die Fortpflanzung der Bienen; Noch etwas uber die Befruchtung der Konigin, *Eichstatt. Bienen-Ztg* **1**, 113

Jaenisch, R., Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals, *Nat. Genet.* **33**(Suppl.), 245–254.

Monk, M. (1988) Genomic imprinting, *Genes Dev.* **2**, 921–925.

Owen, R. (1849) *On Parthenogenesis or the Successive Production of Procreating Individuals from a Single Ovum*, van Voorst, London.

Scully, R.E., Young, R.H., Clement, P.B. (1998) *Tumors of the Ovary, Maldeveloped Gonads, Fallopian Tube and Broad Ligament*, Armed Forces Institute of Pathology, Washington, DC, Atlas of Tumor Pathology, Third series, Fascicle 23.

Sleutels, F., Barlow, D.P. (2002) The Origins of Genomic Imprinting in Mammals, in: Dunlap, J.C., Wu, C.T. (Eds.) *Homology Effects*, Academic Press, San Diego, CA, pp. 119–154.

Suomalainen, E., Saura, A., Lokki, J. (1987) *Cytology and Evolution in Parthenogenesis*, CRC Press, Boca Raton, FL.

Primary Literature

Abbott, T.M., Hermann, W.J. Jr., Scully, R.E. (1984) Ovarian fetiform teratoma (homunculus) in a 9-year-old girl, *Int. J. Gynecol. Pathol.* **2**, 392–402.

Allen, N.D., Barton, S.C., Hilton, K., Norris, M.L., Surani, M.A. (1994) A functional analysis of imprinting in parthenogenetic embryonic stem cells, *Development* **120**, 1473.

Barlow, D.P., Stoger, R., Herrman, B.G., Saito, K., Schweifer, N. (1991) The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus, *Nature* **349**, 84–87.

Cibelli, J.B., Grant, K.A., Chapman, K.B., Cunniff, K., Worst, T., Green, H.L., Walker, S.J., Gutin, P.H., Vilner, L., Tabar, V., Dominko, T., Kane, J., Wettstein, P.J., Lanza, R.P., Studer, L., Vrana, K.E., West, M.D. (2002) *Science* **295**, 819.

Fundele, R., Norris, M.L., Barton, S.C., Reik, W., Surani, M.A. (1989) Systematic elimination of parthenogenetic cells in mouse chimeras, *Development* **106**(1), 29–35.

Kaufman, M.H., Barton, S.C., Surani, M.A.H. (1977) Normal postimplantation development of mouse parthenogenetic embryos to the forelimb bud stage, *Nature* **265**, 53–55.

Keverne, E.B. (1997) Genomic imprinting in the brain, *Curr. Opin. Neurobiol.* **4**, 463–468.

Keverne, E.B., Fundele, R., Narasimha, M., Barton, S.C., Surani, M.A. (1996) Genomic imprinting and the differential roles of parental genomes in brain development, *Brain Res. Dev.* **92**, 91–100.

Kumer, S.C., Vrana, K.E. (1996) Intricate regulation of tyrosine hydroxylase activity and gene expression, *J. Neurochem.* **67**(2), 443–462.

Lin, H., Lei, J., Wininger, D., Nguyen, M.T., Khanna, R., Hartmann, C., Yan, W.L., Huang, S.C. (2003) Multilineage potential of homozygous stem cells derived from metaphase II oocytes, *Stem Cells* **21**, 152–161.

Linder, D., McCaw, B.K., Hecht, F. (1975) Parthenogenic origin of benign ovarian teratomas, *N. Engl. J. Med.* **292**, 63–66.

Marshall, V.S., Wilton, L.J., Moore, H.D.M. (1998) Parthenogenetic activation of marmoset (Callithrix jacchus) oocytes and the development of marmoset parthenogenones In vitro and In vivo, *Biol. Reprod.* **59**, 1491–1497.

McGrath, J., Solter, D. (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes, *Cell* **37**, 179–183.

Nicholls, R.D., Saitoh, S., Horsthemke, B. (1998) Imprinting in Prader – Willi and Angelman syndromes, *Trends Genet.* **14**(5), 194–200.

Obata, Y., Kono, T. (2002) Maternal primary imprinting is established at a specific time for each gene throughout oocyte growth, *Biol. Chem.* **277**(7), 5285–5289.

Ozil, J.P. (1990) The parthenogenetic development of rabbit oocytes after repetitive pulsatile electrical stimulation, *Development* **109**(1), 117–127.

Plachot, M., Mandelbaum, J., de Grouhy, J. (1984) La parthenogenese in vitro dans l'espece humaine, *Ann. Genet.* **27**, 158.

Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R.C., Melton, D.A. (2002) "Stemness": transcriptional profiling of embryonic and adult stem cells, *Science* **298**, 597–600.

Sasaki, H., Jones, P.A., Chaillet, J.R., Ferguson-Smith, A.C., Barton, S.C., Reik, W., Surani, M.A. (1992) Parental imprinting: potentially active chromatin of the repressed maternal allele of the mouse insulin-like growth factor II (Igf2) gene, *Genes Dev.* **6**, 1843–1856.

Strain, L., Warner, J.P., Johnston, T., Bonthron, D.T. (1995) A human parthenogenetic chimaera, *Nat. Genet.* **11**, 164–169.

Surani, M.A., Barton, S.C. (1983) Development of gynogenetic eggs in the mouse: implications for parthenogenetic embryos, *Science* **222**, 1034–1036.

Surani, M.A.H., Barton, S.C., Norris, M.L. (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis, *Nature* **308**, 548–550.

Surani, M.A., Kothary, R., Allen, N.D., Singh, P.B., Fundele, R., Ferguson-Smith, A.C., Barton, S.C. (1990) Genome imprinting and development in the mouse, *Dev.* (Suppl. 1), 89–98.

Surti, U., Hoffner, L., Chakravarti, A., Ferrell, R.E. (1990) Genetics and biology of human ovarian teratomas. I. Cytogenetic analysis and mechanism of origin, *Am. J. Hum. Genet.* **47**, 635–643.

Szabo, P., Mann, J.R. (1994) Expression and methylation of imprinted genes during in vitro differentiation of mouse parthenogenetic and androgenetic embryonic stem cell lines, *Development* **120**, 1651–1660.

Vrana, K.E., Jason, D., Hipp, J.D., Ashley, M., Goss, A.M., McCool, B., Riddle, D., Walker, S.J., Wettstein, P., Studer, L., Tabar, V., Chapman, K., Vilner, L., West, M.D., Grant, K.A., Cibelli, J.B. (2003) Non-human primate parthenogenetic stem cells, *Proc. Natl. Acad. Sci. U.S.A.* **100**(Suppl. 1), 11911–11916.

Pathogens: Innate Immune Reponses

Stefan Ehlers[1] *and Silvia Bulfone-Paus*[2]
[1] *Department of Immunochemistry and Biochemical Microbiology, Parkallee, Borstel, Germany*
[2] *Department of Immunology and Cell Biology, Parkallee, Borstel, Germany*

1 Relationship between Innate and Adaptive Immunity 87

2 Functions of Barrier Organs 88

3 Activation of Innate Immunity: Sensing the Enemy 88
3.1 Pathogen-associated Molecular Patterns (PAMPs) 89
3.2 Host Cellular Sensors 93
3.2.1 Dendritic Cells 93
3.2.2 Mast Cells 94
3.2.3 NKT Cells 95
3.3 Innate Effector Functions 95
3.3.1 Antimicrobial Peptides and Proteins 96
3.3.2 Reactive Oxygen and Reactive Nitrogen Metabolites (ROI, RNI) 96

4 Translating Innate Immune Activation into Regulatory Circuits: Molecular Pathways Shaping Adaptive Immunity 97
4.1 TLR-initiated Signaling Cascades 97
4.2 Molecules Involved in Recruiting Effector Cells 98
4.2.1 Defensins 98
4.2.2 Chemokines 100
4.3 Molecules Involved in T- and B-Cell Differentiation 101
4.3.1 Th1-inducing Cytokines 101
4.3.2 Th2-inducing Cytokines 105

5 Practical Implications 106

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

Bibliography 107
Books and Reviews 107
Primary Literature 108

Keywords

pathogen-associated molecular pattern (PAMP)
Structural motif of a microbial constituent, for example, of a cell wall component or genetic material. These motifs are not really pathogen specific, but occur also on other microorganisms. They are, however, distinct from any motifs encountered in vertebrate hosts. PAMPs are recognized by host cells via pattern-recognition receptors, initiating a cascade of antimicrobial activities.

toll-like receptors (TLR)
A family of pattern-recognition receptors bearing strong homologies to Toll, a protein involved in dorsal-ventral patterning as well as antifungal responses in *Drosophila melanogaster*. TLR ligation results in intracellular signaling cascades that lead to the elaboration of antimicrobially active molecules as well as to the regulation of adaptive responses.

Innate Immunity
All immune responses executed by cells of the innate immune system, that is, granulocytes, monocytes, macrophages, natural killer cells, and mast cells. These responses comprise sensing microbial motifs, activating effector molecules, and alerting specific lymphocytes.

Adaptive Immunity
All immune responses executed by specific antigen-receptors bearing T and B cells. A generally accepted view of adaptive immunity and, in particular, effector T-cell development and function distinguishes between type 1 and type 2 responses. Infection with bacteria, viruses, and fungi, or stimulation of innate immune cells with components thereof, induces type 1 immunity, mediated by T helper 1 (Th1) or T cytolytic 1 (Tc1) cells. In this context, type-1 cytokines (i.e. Interleukin-2, IL-12, IL-15, IFN-γ, IFN-α, and IFN-β) are particularly involved in cell-mediated immunity to intracellular microbes. Cytotoxic T cells can kill infected cells, using perforin and granzymes. In contrast, infections by extracellular pathogens and multicellular parasites, or exposure to allergens, induce type 2 responses. Type-2 cytokines (i.e. Interleukin-4, IL-5, IL-6, IL-10, and IL-13) play a crucial role in promoting humoral immunity and/or immune deviation to a nonprotective response (tolerance).

Cytokines
Polypeptides enabling the communication between cells, mostly of the immune system. Cytokines are involved in the growth, differentiation, survival, and activation of cells bearing specific receptors.

Chemokines
Chemoattractive molecules (polypeptides with positionally conserved cystein residues) that induce migration of chemokine-receptor bearing cells along a gradient toward the site of their production.

Responses of macrophages, dendritic cells, granulocytes, and mast cells constitute the first line of immune defense against infectious agents. These cells have the capacity to recognize and distinguish between physicochemical motifs expressed on microbes and to tailor rapid antimicrobial defenses accordingly. Identification of microbial constituents involves host membrane-bound detectors, called *toll-like receptors*, and intracellular sensors, termed *Nod proteins*. Upon activation, the innate immune system arms a series of effector molecules capable of thwarting the growth or killing invading microorganisms, including oxygen and nitrogen metabolites. Innate immune cells also play a central role in instructing the adaptive immune system as to which type of specific response should develop to optimally combat infectious agents. To this end, innate cells translate the sensed danger or damage wrought by infectious agents into a language understood by specific lymphocytes, including membrane-bound costimulatory molecules and soluble mediators in the form of cytokines and chemokines.

1 Relationship between Innate and Adaptive Immunity

More than 98% of multicellular animal species cannot produce an adaptive immune response to a pathogen. Immunological memory is thus a privilege of a few vertebrate hosts, and its primary purpose is to sterilize primary infection and provide resistance to reinfection with pathogens. However, during the time it takes for clonal expansion of antigen receptor-bearing lymphocytes to occur, the primordial system of innate defense takes care of the host. Individual functions of innate immune cells are not only involved in mitigating the damaging insults of microbial or tumor growth but are also critical in directing the differentiation processes that, following antigen encounter, ultimately lead to T and B effector cells.

Full-fledged adaptive immunity, in turn, is the result of cells of the adaptive immune system calling upon cells of the innate immune system, either to provide professional phagocytes to engulf and destroy pathogens (type 1 immunity) or to release toxic mediators in order to establish environments that are inhospitable to

helminth or arthropod invaders (type 2 immunity). As a consequence, the innate and adaptive immune systems should be viewed as two interdependent parts of a single integrated immune system.

It is the purpose of this chapter (1) to delineate some of the mechanisms that trigger the activation of innate immune cells during infection; (2) to describe some of the engrained effector functions exerted by innate immune cells upon activation; and (3) to summarize the mediators and dissect the major molecular pathways regulating the differentiation of effector functions of adaptive immunity.

2 Functions of Barrier Organs

Infectious organisms first need to gain access to their preferred sites of multiplication by penetrating either the skin or mucosal surfaces in the lung, gut, or urogenital tract. Antimicrobial defenses within these barrier organs are not generally thought of as facets of "immunity," but are certainly innate and highly effective.

The skin provides an impenetrable barrier for most microorganisms as long as there are no erosions owing to injury or insect bites. Secretions from sebaceous glands, slightly acidic pH, and competition from resident members of the "commensal" skin flora make it difficult for many infectious agents to take hold. In addition, epidermal cells are a major source of antimicrobial peptides that can kill a variety of bacteria and fungi; in fact, the initial purification of human defensins was first successfully performed from shedded psoriatic skin.

The gut is lined with goblet cells that secrete mucus, which serves as a nutrient to the resident intestinal flora. The normal flora of the large intestine, for example, consists of approximately 10^{14} bacteria comprising more than 500 species, which provide severe competition for potentially invasive pathogens, particularly in terms of nutrient supply and attachment to the gut epithelium. Mucus is also the carrier of a number of antimicrobially active substances, secreted from epithelial cells and Paneth cells located in the crypts of the intestinal villi. These antimicrobial proteins effectively prevent bacterial adhesion and invasion into lower strata of the gut mucosa.

The upper airways are also covered with viscous mucus, which contains high concentrations of lysozyme and lactoferrin, as well as epithelial defensins. This combination of antimicrobials enables initial control of most infectious agents, leaving mucociliary clearance to accomplish further expulsion. The alveoli of the lower respiratory tract are bathed in lung surfactant. Surfactant proteins A, B, C, and D have multiple functions: on the one hand, they can opsonize bacteria and thereby enhance phagocytic uptake, on the other hand, they also modulate overall macrophage functions, thereby regulating excess inflammatory responses.

3 Activation of Innate Immunity: Sensing the Enemy

Once a pathogenic organism has penetrated the mucosal or epithelial barrier systems, defense strategies initiated by cells of the innate immune system take over. These are strategically placed on many mucosal surfaces (e.g. alveolar macrophages in the lung) and are primarily aimed at recognizing foreign structures and eliminating them. Since infectious pathogens

have often evolved to subvert these mechanisms, specialized cells of the innate immune system are also capable of presenting digested antigen to T cells in a context and environment appropriate for optimal effector cell development. Activation of innate immunity to provide this context may be conceptually divided into three stages (Fig. 1): (1) a phase of antigen sensing during which a combination of surface receptors detects the presence of nonself structures on invading microorganisms (''detection phase''); (2) a phase of translating this sensory information into a language understood by the cells of the adaptive immune system, for example, chemokines and cytokines (''transmission phase''); (3) a phase of mobilizing an effector armamentarium, capable of eliminating or at least restricting the multiplication of infectious agents (''effector phase''). This latter phase may proceed, to a limited extent, without the help of the adaptive immune system, but is consistently amplified by it to maximal potency.

3.1 Pathogen-associated Molecular Patterns (PAMPs)

The costimulatory signals required for differentiation of adaptive immune responses are exclusively turned on when cells of the innate defense system interact with pathogen-associated molecular

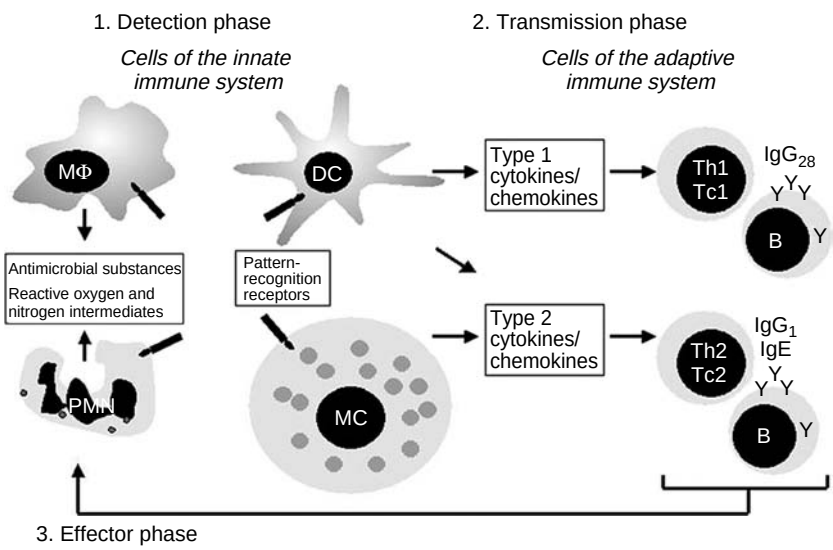

Fig. 1 Overall view of innate immune responses. Cells of the innate immune system (DC = dendritic cells; MC = mast cells) detect the presence of foreign structures, such as pathogen-associated molecular patterns, by surface-expressed pattern-recognition receptors (''detection phase''). Innate immune cells translate this sensory input into a language (cytokines, chemokines) understood by cells of the adaptive immune system, skewing the response into either T helper 1 (Th1)/T cytolytic 1 (Tc1) or Th2/Tc2-type immunity (''transmission phase''). At the same time, ligation of pattern-recognition receptors initiates an antimicrobial program within innate immune cells (''effector phase''). This program is further amplified and complemented by activating cytokines derived from differentiated adaptive immune cells.

patterns (PAMPs, that is, highly conserved structural motifs that are necessary for the survival of microorganisms).

Many pattern-recognition receptors primarily serve the purpose of enhancing phagocytosis. For example, the macrophage mannose receptor (recognizing terminal mannose and fucose residues on microbial cell walls) or the macrophage scavenger receptor (recognizing polyanionic ligands, such as double-stranded RNA, lipopolysaccharide, and lipoteichoic acid) are mainly involved in clearing pathogens from the site of invasion.

Discrimination between different classes of microorganisms, however, is afforded by the toll-like receptors 1 through 11. Since PAMPs have multiple effects on the signals generated by cells of the innate immune system and may therefore be useful as natural adjuvants in driving adaptive responses, they are described here in detail (Table 1). Surveillance mechanisms that sense and respond to microbial compounds in the cytosol are provided by the Nod proteins, a family of nucleotide-binding site, leucine-rich repeat proteins.

Peptidoglycan (*PG*) is a component of all bacterial cells. PG chains consist of alternating β-MurNAc(1 $\rightarrow$ 4)GlcNAc disaccharide units linked via β-(1 $\rightarrow$ 4) glycosidic bonds. Peptide side chains, which are

Tab. 1 Recognition of pathogen structures via surface or intracellular receptors on innate immune cells.

Pattern-recognition receptors on innate immune cells	***Cellular distribution of expression of pattern-recognition receptors***	***Pathogen-associated molecular patterns recognized by pattern-recognition receptors***
TLR1/TLR2	Monocytes, macrophages, DC	Triacylated lipopeptides
TLR2	Monocytes, macrophages, DC	Lipoproteins, lipoteichoic acid, glycolipids, modulin, arabinose-capped lipoarabinomannan, GPI-anchored molecules from parasites
TLR2/TLR6	Monocytes, macrophages, DC	Diacylated lipopeptides; zymosan
TLR3	DC, NK cells	Double-stranded RNA
TLR4	Macrophages, DC, and endothelial cells	Lipopolysaccharide
TLR5	Monocytes, immature DC, NK cells	Flagellin
TLR7 and TLR8	B cells, plasmacytoid precursor DC	Imidazoquinoline derivatives, single-stranded RNA
TLR8	Monocytes, NK and T cells	Imidazoquinoline derivatives, single-stranded RNA
TLR9	B cells, macrophages, NK cells, plasmacytoid precursor DC	Unmethylated CpG-motif oligonucleotides
TLR10	B cells, plasmacytoid precursor DC	Currently unknown
TLR11	Epithelial cells	Uropathogenic bacteria
NOD-1	All cells	D-Glu-mesoDAP
NOD-2	Bone marrow–derived cells	MurNAc-L-Ala-D-isoGLn

attached to the carboxyl group of muramic acid residues, are primarily composed of five amino acids (L-Ala-D-Glu-X-D-Ala-D-Ala, where X stands either for L-Lys or diaminopimelic acid). Certain variations both in peptide and glycan structure exist, depending on the microorganism. A core constituent of PG, muramyl dipeptide (MurNAc-L-Ala-D-isoGln) is recognized by the intracellular sensor protein, NOD2. The minimal structure recognized by NOD1 is the terminal two amino acids in the stem peptide of a peptidoglycan structure mainly found in Gram-negative bacteria: D-Glu-mesoDAP.

Lipopeptides and lipoproteins are components of both Gram-positive and Gram-negative cell walls and may be viewed as functional substituents of PG. A 19-kDa lipoprotein derived from *Mycobacterium tuberculosis* activates cells via TLR2. The diacylated mycoplasma macrophage-activating lipopeptide (MALP-) 2 requires the cooperation of TLR2 and TLR6 to initiate proinflammatory responses, whereas triacylated bacterial lipopeptides require dimerization of TLR2 with TLR1.

Lipoteichoic acid (*LTA*) is an amphiphilic, negatively charged glycolipid found in most Gram-positive bacteria. Appropriate purification of LTA from *Staphylococcus aureus*, or chemical synthesis of LTA, have revealed LTA to be a potent and exclusive TLR2 stimulus. Structural requirements include a lipid anchor and D-Ala substituents, as L-Ala derivatives reduce the activity by a factor of 100.

Lipopolysaccharide (*LPS*) is a major component of the outer membrane of Gram-negative bacteria and is composed of polysaccharides attached to a lipid portion (lipoid A). Lipoid A is responsible for all biological activities ascribed to LPS, and it is the spatial conformation of lipoid A – in part determined by the number and length of acyl chains – that determines biological activity on a given cell type. LPS recognition by TLR4 requires several accessory molecules. LPS is first bound to the LPS-binding protein LBP, which transfers LPS monomers to CD14. How exactly CD14 facilitates LPS recognition is unclear, as is the exact role of MD-2, another component of the LPS-recognition complex. Ultimately, LPS interacts with TLR4 and initiates both MyD88-dependent and -independent intracellular signaling cascades. Some forms of LPS (e.g. from *Porphyromonas gingivitidis* or *Leptospira interrogans*) seem to be recognized by TLR2. The structural requirements for this interaction are not clear, other than that this recognition is greatly enhanced in the presence of TLR1. Uropathogenic *Escherichia coli* contain another, not yet defined structural motif that is capable of activating TLR11 on macrophages and liver, kidney, and bladder epithelial cells.

Flagellin is the protein subunit of bacterial flagella. It is recognized by TLR5, possibly via highly conserved N- and C-terminal ends that form its hydrophobic core. TLR5 is expressed only on the basolateral, but not apical, surface of epithelium. Therefore, flagellin of invading pathogens, but not commensal bacteria, induces intestinal epithelial cells to mount an inflammatory response.

Unmethylated CpG DNA motifs are prevalent in bacterial, but not vertebrate, genomic DNA and are recognized by host cells bearing TLR9. TLR9 is not on the cell surface but is present in vesicles that fuse with lysosomes. Some CpG motifs are very powerful activators of NK cells and induce IFN-α production by DC2 cells; other motifs are potent B-cell activators, triggering cell cycle progression and antibody secretion. Nonspecific activation of innate responses afforded by CpG can increase

resistance in animal models of intracellular infection and sepsis. The ability of CpG to act as a potent adjuvant was confirmed in studies using both model antigens, such as hen egg lysozyme, ovalbumin, as well as tumor antigens and antigens from infectious agents, including hepatitis B, HIV, and influenza virus or *Leishmania major*. In fact, CpG is a stronger Th1-like adjuvant for inducing B- and T-cell responses than the gold standard, Freund's complete adjuvant, as measured by its ability to drive the differentiation of cytolytic T cells and IFN-γ-secreting cells. It should be noted that optimal CpG motifs for activation differ between human beings and mice.

Lipoarabinomannans (*LAMs*) are lipoglycans restricted to the genus *Mycobacterium*. The carbohydrate backbone is comprised of a D-mannan core and a D-arabinan domain. At its reducing end, the mannan core is terminated by a glycosylphosphatidylinositol anchor, while the arabinan domain is capped by either mannosyl (ManLAM) or phosphoinositide residues (PILAM). PILAMs prevalent in rapidly growing mycobacterial species (*M. smegmatis*) activate mammalian cells in a TLR2-dependent manner to secrete TNF and IL-12. In contrast, ManLAM is not an agonist for TLR2. A small mycobacterial glycolipid, phosphatidylinositol dimannoside (PIM) also activates cells in a TLR2-dependent manner.

Double-stranded RNA (*dsRNA*) is produced by most viruses during their infectious cycle. DsRNA induces protein kinase R activation via TLR3. TLR3 is expressed predominantly on myeloid DCs, making them the prime targets for initiating innate responses against viruses. TLR3-deficient mice also showed a reduced response to the viral RNA mimic, polyinosine-polycytidylic acid (poly I:C).

Single-stranded RNA (*ssRNA*), prepared from GU rich sequences in HIV-1 RNA, was found to ligate TLR7 and TLR8. During viral infection, ssRNA possibly reaches the endosome through receptor-mediated uptake of viral particles or by fusion events of budding viruses. Another class of TLR7 and TLR8 ligands is constituted by a series of small synthetic antivirally active molecules, the imidazoquinolines.

Glycosylphosphatidylinositol-anchored molecules (*GPI*) are the predominant antigens present on the plasma membrane of protozoa such as *Trypanosoma cruzi* and *Plasmodium falciparum*. GPI from *T. cruzi* are recognized by TLR2 and potently induce nitric oxide, TNF, and IL-12 in murine macrophages. The PI moiety is mostly composed of unsaturated fatty acids, which are essential for optimal biological activity.

FimH is a subunit protein of type 1 fimbrial adhesins present in many enterobacterial species, for example, *E. coli* or *Salmonella enterica*. It can bind directly to a GPI-anchored receptor present also on the surface of mast cells (CD 48). Ligation of CD 48 by FimH-expressing bacteria results in bacterial uptake into caveolar chambers, promotes bacterial survival inside the cytosol, and has been linked to the release of inflammatory mediators by mast cells.

Heat shock proteins (*HSPs*) are released from necrotic cells during tissue injury or lysis of infected cells, and can form stable complexes with peptides. Some of them, for example, HSP60, can be recognized by TLR4; others, for example, gp96, HSP90, and HSP70, interact with antigen-presenting cells via a common receptor, CD91. HSP-chaperoned peptides enter the macrophage/dendritic cells through CD91 and are processed and presented by the MHC class I and MHC class II

molecules, resulting in the stimulation of CD4 and CD8 T cells. HSP-DC interaction through CD91 leads to maturation of DCs and secretion of an array of proinflammatory cytokines. Immunization of mammals with HSP-peptide complexes elicits potent responses against the HSP-chaperoned peptides.

Parasite-derived glycans and lectins, for example, from *Schistosoma mansoni* eggs or filarial nematodes, trigger responses by TLR-independent mechanisms and are known for their potent induction of Th2 responses. One well-characterized lectin from *S. mansoni* eggs, termed *IPSE*, leads to cross-linking of IgE on basophils, which release IL-4 and IL-13, thus driving a strong Th2 response.

All currently known PAMPs are derived from prokaryotic, fungal, viral, or protozoan pathogens, and it is currently unknown which receptors are required for recognition of "patterns" or structural motifs expressed by multicellular eukaryotic parasites or allergens. Whether as yet undefined molecular patterns are at all necessary to skew adaptive immunity toward a type 2 pattern after encounter with these different types of insults remains to be determined. It is possible that a different type of signal, such as the absence of PAMPs – similar to the lack of self-MHC I for NK-cell activation – is a stimulus in itself for triggering type 2 immunity.

3.2 Host Cellular Sensors

At the immunological level, innate defenses rely mostly on granulocytes, mast cells (MC), macrophages, dendritic cells (DC),natural killer (NK), NKT, and $\gamma\delta$T cells. These cells serve as a bridge between PAMPs and antigen-specific cells of the adaptive immunity, translating the sensory input of pattern-recognition receptors into soluble mediators that communicate with T and B cells via specific cytokine/chemokine receptors. Although many cell types contribute in various ways to linking innate and adaptive responses, we shall focus in this chapter – for the purpose of clarity, and because the two prototypic immune responses (type 1 and type 2) are differentially regulated by them – on two pivotally involved cell types, DCs and MCs.

3.2.1 Dendritic Cells

The key cell type, and indeed the most potent type of any antigen-presenting cell, which is able both to sense a foreign insult and to orchestrate the cells of the adaptive immune system, is the dendritic cell. Like other cell types within the immune system, immature DCs are continuously produced from hematopoietic stem cells within the bone marrow. Flt-3 ligand and, to a lesser extent, granulocyte-macrophage colony-stimulating factor (GM-CSF) are the key growth and differentiation factors *in vivo*.

Immature DCs have long been considered "sentinels" of the immune system, since they are specialized, relatively long-lived phagocytic cells, which are resident in most tissues where they survey the local environment for pathogens. Pattern-recognition receptors, in particular, TLRs, allow DCs to recognize microbial antigens and mature into highly effective antigen-presenting cells, undergoing changes that enable them to activate pathogen-specific lymphocytes, which they encounter in the lymph nodes.

Maturing DCs lose the ability to take up and process antigen, but they upregulate surface expression of MHC class I and II expression, as well as of costimulatory and adhesion molecules such as CD86, CD80,

CD40, and CD54. Mature DCs express proinflammatory cytokines like IL-6, IL-12, IL-18, IL-23, and IL-27, and are able to detect the expression of cytokines by infected cells, thus integrating different signals of danger. Another consequence of imDC stimulation is the upregulation of CCR7 expression and downregulation of CCR5 expression, which allows DCs to migrate to regional lymph nodes. There, mature DCs prime naive T cells and, depending on the predominating maturation signals, differentially induce Th1 versus Th2 development. Bacterial CpG DNA, double-stranded viral RNA, LPS, as well as CD40 ligand and IFN-γ induce immature DCs to produce IL-12 and promote Th1 immune responses. In contrast, anti-inflammatory molecules such as IL-10, TGF-β, corticosteroids, and prostaglandin E_2 promote Th2 or regulatory T-cell immune responses.

The ability of DCs to prime greatly diverse immune responses is the result of different DC subsets and their unlimited plasticity. In addition to immature DCs, stem cells also give rise to two types of DC precursors (pre-DCs), monocytes (pre-DC1) and plasmacytoid cells (pre-DC2), during hematopoiesis. Functionally, immature DCs and pre-DCs differ since pre-DCs are more directly involved in innate immunity against microbes. Pre-DC1 phagocytose and kill various bacteria and fungi, while pre-DC2s play a major role in the early antiviral innate immune response by producing large amounts of IFN-α/β upon viral infection. Therefore, these cells were also named "natural interferon–producing cells" (NIPCs). Unlike other innate immune cells, such as neutrophils, eosinophils, and basophils, which die after performing their functions, pre-DC1 and pre-DC2 differentiate into DCs. When placed in culture with GM-CSF and IL-4, or after bacterial uptake, pre-DC1 differentiate into immature myeloid DCs. Upon CD40 ligand–induced activation, DC1 produce large amounts of IL-12 and prime Th1 and cytotoxic T-cell immune responses. In contrast, pre-DC2 differentiate into immature DC2 in the presence of IL-3 or CD40 ligand, and mature DC2 primarily induce Th2 immune responses. Interestingly, when immature DC2 cells become infected by a virus, they take on the typical morphology of mature DCs and upregulate costimulatory molecules, but acquire the distinct ability to differentiate naïve CD4+ T cells into producers of IFN-γ and IL-10.

3.2.2 Mast Cells

Mast cells (MC) are derived from CD34+ stem cells. They initiate their differentiation in the bone marrow under the influence of stem cell factor and interleukin-3 . Their final maturation takes place in connective tissues.

Within minutes following activation, mast cells extrude granule-associated substances, such as histamine. Within hours, mast-cell activation is followed by *de novo* synthesis of numerous cytokines (TNF-α, Interleukins 1–10, IL-12, IL-13, IL-15, IL-16, IL-18, IL-25, GM-CSF) and chemokines (CCL2-5, CCL8, CCL11, IL-8). Mast cells are long-lived and therefore able to respond repeatedly to the same stimulus.

In addition to their well-established central role in the pathogenesis of allergic disorders, MCs are now also appreciated as key effector cells in the induction of protective immune responses to bacteria. This notion is supported by several independent lines of evidence. MCs are found at the portals of entry of many infectious agents, that is, within mucosal surfaces of the gastrointestinal and respiratory tracts,

or in the skin in close proximity to peripheral nerves and are often associated with blood vessels. MCs are equipped with a large variety and number of receptors to detect bacteria or signs of bacterial infection including toll-like receptors (TLRs), CD48, and complement receptors. Furthermore, genetically MC-deficient animals, such as $Kit^{W}/Kit^{W-\nu}$ mice, exhibit insufficient host defense responses in various models of bacterial infection, for example, in cecal ligation and puncture-induced acute septic peritonitis. Finally, protection from bacterial infections in mice can be improved by enhancing MC numbers and/or function.

Mast cells may exert many different effects on inflammatory processes: (1) at early stages, mast cells can be activated via TLRs in the peripheral tissues through interaction with microbes; (2) the release of cytokines, such as IL-4 and IL-13, may influence DC maturation, and as a consequence, biases T-cell differentiation to a Th2/Tc2 phenotype; (3) mast cell–derived chemokines may recruit distinct subsets of inflammatory cells to the site amplifying a predominantly type-2 response; (4) activated mast cells can migrate to secondary lymphoid organs, where they can directly influence the development of T-cell immune responses via the release of cytokines and the expression of costimulatory molecules.

3.2.3 NKT Cells

NKT cells are a subset of T cells that express a heavily biased T-cell receptor repertoire as well as the NK-cell receptor CD161; they are deficient in immunological memory. In contrast to conventional T cells, NKT cells recognize glycolipid antigens presented by the nonpolymorphic MHC class I–like molecule CD1d. Although they are not strictly necessary for the generation of a polarized Th cell response, they can modulate the latter by secreting large amounts of IL-4, IL-10, IL-13, as well as IFN-γ or TNF upon T-cell receptor engagement.

A glycosphingolipid originally isolated from the sea sponge Agelas mauritianus, α-galactosylceramide (α-GalCer), is a potent inducer of NKT activity. Optimal stimulation of NKT cells requires direct contact between NKT cells and DC through CD40–CD40–ligand interactions and IL-12 production by DC. Most studies indicate that α-GalCer enhances Th2 polarization, but Th1 differentiation may also be affected.

In vivo activation of NKT cells with α-GalCer ameliorates the diseases caused by *Cryptococcus neoformans* or encephalomyocarditis virus and inhibits the development of the intrahepatocytic stages of *Plasmodium yoelii* and *Plasmodium berghei* in mice. α-GalCer potentiates protective immune responses induced by malaria vaccines.

3.3 Innate Effector Functions

In the first phase of inflammatory responses, chemotactic substances are produced that recruit granulocytes to the site of infection. Microorganisms covered with antibodies or activated complement factors are bound to granulocytes through specific receptors on the plasma membrane, inducing ingestion of the particle (phagocytosis). In addition, tissue macrophages, or monocytes differentiating into activated macrophages, are also capable of phagocytosing microbes via a range of specialized receptors.

The purpose of phagocytosis is to confine microbes in specialized compartments (phagosomes) where they can be exposed to a low pH and to very high

local concentrations of antimicrobial substances, stored in other organelles (such as lysosomes). The acid milieu in the phagolysosome is created by an ATPase-dependent proton pump. Antimicrobial activity can be mediated by effectors ranging from inorganic molecules to polypeptides, carbohydrates, and lipids. Granulocytes and macrophages dispose of a wide range of toxic mechanisms to fight invading microorganisms.

Upon recognition of their cognate ligands, TLRs induce the expression of a variety of host defense genes encoding for antimicrobial peptides and proteins, and for enzyme systems generating toxic oxygen and nitrogen intermediates.

3.3.1 Antimicrobial Peptides and Proteins

Antimicrobial peptides form a large and diverse family of structurally related molecules that have the ability to adopt a shape in which the hydrophobic amino acids are organized on one side of the molecule and the cationic hydrophilic amino acids on the other. For example, defensins are 2 to 6 kDa, cationic, microbicidal peptides that are classified on the basis of their size and pattern of disulfide bonding, into α, β, and θ (extent only in primates) categories. α-defensins are particularly abundant in neutrophils, whereas β-defensins are more often produced by epithelial cells lining, for instance, the skin, the lung, and the genitourinary tract. Defensins are produced constitutively or in response to microbial products. Defensins can kill or inactivate a wide spectrum of bacteria, fungi, or viruses by disrupting the microbial membrane.

Antimicrobial proteins are widely distributed in host defense cells and secretions. Within cells, they are stored in granules (e.g. lysozyme, lactoferrin, RNAse, granulysin, phospholipase A2). When the cells are appropriately stimulated, the granules fuse to internalized cell membranes enclosing ingested microbes, or to the external plasma membrane facing the pericellular space, a process called *degranulation*.

Some antimicrobial proteins are enzymes that lyse their target microbial structure. For example, lysozyme degrades peptidoglycan; similarly, phospholipase A2 destroys bacterial phospholipid membranes. Others disrupt microbial cell walls and membranes or inhibit microbial growth by nonenzymatic mechanisms. For example, lactoferrin chelates iron, thereby depriving microbes of an essential component for growth, and BPI (bacterial permeability increasing protein) binds to LPS of Gram-negative bacteria, first permeabilizing their outer membrane and subsequently destroying their inner membrane. Granulysin is a small protein expressed in the granules of NK cells and cytolytic T lymphocytes that also permeabilizes target membranes.

3.3.2 Reactive Oxygen and Reactive Nitrogen Metabolites (ROI, RNI)

The two most important oxygen-dependent pathways for generating antimicrobial effector molecules are via the phagocyte NADPH oxidase complex and the inducible nitric oxide synthase (iNOS). The former is typically expressed in neutrophilic granulocytes, the latter is most abundant in activated macrophages.

The oxidative burst of granulocytes and macrophages can be triggered by phagocytosis of pathogens, by certain chemoattractants, and by TLR-ligating structural motifs. Reactive oxygen metabolites, such as HOCl, cause the oxidation of thiol groups, heme prosthetic groups, and the chlorination of tyrosine residues, and the

hydroxal radicals incur DNA strand breaks and lipid peroxidation.

Prototypic inducers of iNOS are proinflammatory cytokines, in particular, interferon-gamma, and microbial products. The direct antimicrobial activity of nitric oxide likely involves mutation of DNA, inhibition of DNA repair and synthesis, S-nitrosylation, and peroxidation of membrane lipids. However, RNIs also have multiple immunomodulatory functions that may result in indirect antimicrobial effects.

4
Translating Innate Immune Activation into Regulatory Circuits: Molecular Pathways Shaping Adaptive Immunity

While innate immune cells can often keep invading microorganisms at bay for some time, pathogens have evolved mechanisms to subvert these strategies. TLR ligation on cells of the innate immune system, however, also triggers the expression of genes that will alert T and B cells whose action will then complement and augment the functions of innate defenses. Bridging the gap between innate and adaptive immunity is accomplished by inflammatory cytokines and chemokines that serve as costimulatory molecules to prime T- and B-cell differentiation.

4.1
TLR-initiated Signaling Cascades

After ligation of TLRs, receptor dimerization occurs, triggering recruitment of the adapter MyD88 to the receptor complex. Association of the TIR (Toll/IL-1R) domain of MyD88 with the TIR part of TLRs facilitates association of the death domain of MyD88 with the IL-1R-associated kinase (IRAK). IRAK then autophosphorylates, dissociates from the receptor complex, and interacts with TNF-receptor-associated factor (TRAF) 6. TRAF6 leads to activation of MAP kinases, and by recruiting ECSIT (evolutionarily conserved signaling intermediate in Toll pathways), activates the IKK complex (inhibitor of nuclear factor-κB-kinase complex), thereby releasing NF-κB from its inhibitor so that it can translocate to the nucleus and direct the transcription of inflammatory cytokines. This pathway is used by TLR1, TLR2, TLR4, TLR5, TLR6, TLR7, and TLR9. TIRAP (TIR-domain-containing adaptor protein) is also involved in the MyD88-dependent signaling pathway through TLR2 and TLR4. By contrast, TLR3- and TLR4-mediated activation of interferon-regulatory factor 3 (IRF) and the induction of IFN-β are observed in an MyD88-independent manner. A third TIR-domain-containing adaptor, TRIF, is essential for the MyD88-independent pathway that involves non-typical IKKs. Finally, a fourth adaptor, TRAM (TRIF-related adaptor molecule) is specific to the TLR4-mediated, MyD88-independent/TRIF-dependent pathway.

Further regulatory mechanisms and signaling pathways downstream of TLRs exist, such as the Rac1-PI3K-AKT pathway activated by TLR2. Much remains to be learnt concerning the bifurcation of intracellular pathways that differentially determine type 1 versus type 2 cytokine secretion during infection. In a mycobacterial infection model, TNF production was predominantly ERK-dependent, but independent of p38 MAP kinase activation, whereas IL-10 secretion was predominantly p-38-dependent, but independent of ERK signaling. The use of small molecules as inhibitors of these putatively divergent pathways is an attractive novel

option to selectively skew adaptive immune responses.

At present, there is little evidence that specific TLRs differentially induce Th1 versus Th2 responses. However, certain subsets of DCs respond differently to different PAMPs because they express different TLRs. Thus, plasmacytoid DCs express TLR7 and TLR9, whereas myeloid DCs express many of the remaining TLRs, in particular, TLR3, but not TLR7, and TLR9. This may provide a means of fine-tuning the adaptive response by targeting only certain specialized DCs.

TLR2-dependent activation of mast cells was reported to preferentially induce TNF, IL-4, IL-5, IL-6, and IL-13, whereas TLR4 ligation led to TNF, IL-1β, IL-6, and IL-13, but not IL-4 or IL-5 secretion. In the same study, TLR2-mediated mast cell activation led to degranulation, while TLR4-mediated activation did not.

All MyD88-dependent signals favor Th1 responses. In the absence of MyD88, no Th1 differentiation occurs, whereas Th2 development is seemingly unaffected by MyD88 deficiency. It is unclear whether Th2 differentiation is determined by specific, as yet unknown intracellular signals, or whether it is simply the default pathway. Surprisingly, human patients with a complete recessive IRAK-4 deficiency only seemed to suffer from pyogenic bacterial infections, but were apparently resistant against other ubiquitous microorganisms, such as *Mycobacterium avium*, the latter being clinically highly relevant in patients with hereditary defects in the IL-12/IFN-γ signaling pathways. The fact that significant protective immunity against common pathogens is afforded even in the absence of critical receptors and/or major signaling pathways indicates that there is substantial plasticity and room for compensatory mechanisms in immune cell activation.

There is now also evidence that stimulation via NOD2 can regulate the response to TLR ligation. Nuclear localization of all NF-κB subunits, and, in particular, cREL, was increased in cells deficient for NOD2 following stimulation with muramyl dipeptide, and total splenocytes of NOD2-deficient mice showed greatly increased levels of IL-12, IL-18, and IFN-γ subsequent to TLR2-ligation. Since patients with Crohn's disease (a chronic inflammatory bowel syndrome) have a mutation in NOD2, the lack of repression afforded by the mutant protein likely accounts for the excessive type 1 inflammatory responses present in these patients.

4.2 Molecules Involved in Recruiting Effector Cells

Correct and efficient priming of adaptive responses and, indeed, the overall type of immunity itself heavily depend on the rapid recruitment of professional antigen-presenting cells together with T and B lymphocytes. Thus, molecules coordinating cellular movement from the blood stream into either inflammatory sites, or sites of antigen presentation such as lymph nodes, play a critical role in directing the priming process (Fig. 2). Conversely, differential regulation of surface receptors capable of detecting gradients of chemoattractants is a crucial factor determining the cellular environment in which priming of immune responses takes place.

4.2.1 Defensins

Defensins are not only direct effector molecules of innate antimicrobial immunity but they can also enhance phagocytosis, induce the activation and degranulation of mast cells, and stimulate bronchial epithelial cells to augment

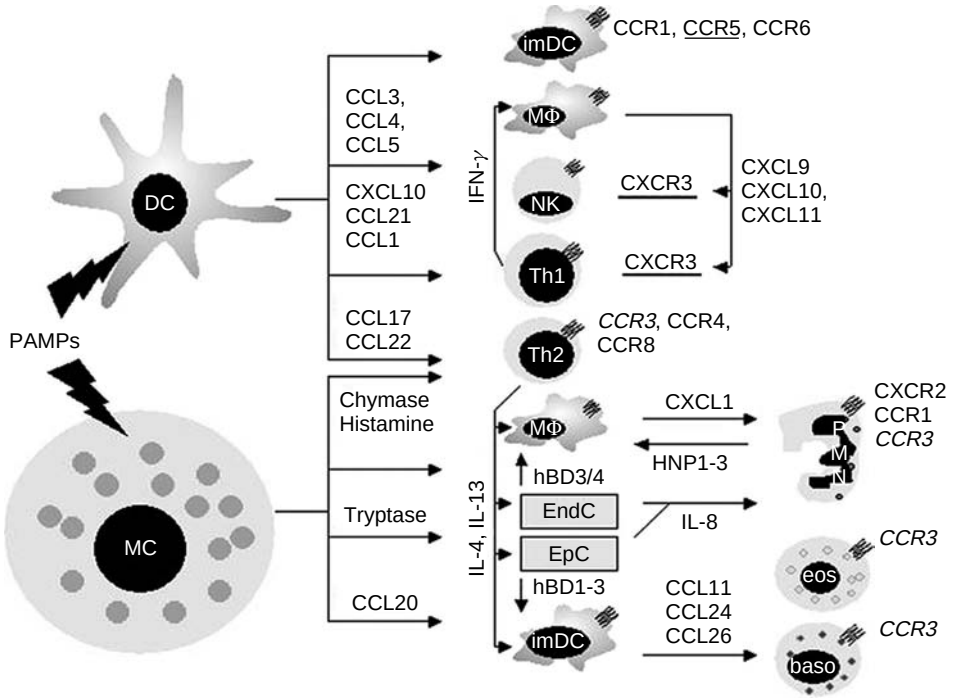

Fig. 2 Major chemokine pathways linking innate and adaptive immune responses. Cells of the innate immune system, triggered by structures on microbes, release chemokines, thereby recruiting different type of cells during the initial phase of the immune response. Type 1 responses are dominated by cells bearing CCR5 and CXCR3 (underlined) and are further amplified by IFN-γ derived from NK and Th1 cells upregulating expression of CXCR3-binding chemokines. Type 2 responses are dominated by cells bearing CCR3 (italics) and are further modulated by IL-4 and IL-13 derived from Th2 cells and mast cells. DC (dendritic cell), MC (mast cell), imDC (immature dendritic cell), MΦ (macrophage), NK (natural killer cell), PMN (polymorphonuclear granulocyte), eos (eosinophil granulocyte), baso (basophil granulocyte), EndC (endothelial cell), EpC (epithelial cell), PAMP (pathogen-associated molecular patterns).

IL-8 production. In this way, defensins are involved in recruiting and activating neutrophils into inflammatory sites. Because some defensins can enhance TNF and IL-1 production – while reducing IL-10 secretion – in monocytes, they may play a role in amplifying the local inflammatory response.

Human neutrophil-derived α-defensins (HNP) 1 and HNP2 are chemotactic for human monocytes and T cells. They are selectively chemotactic for resting CD4/CD45RA and CD8 T cells, whereas human β-defensin (hBD) 2 is chemotactic for immature DCs and CD45RO memory T cells by interacting with human CCR6. HBD 3 and hBD4 are also chemotactic for monocytes via an as yet unidentified receptor. Defensins may also be involved in the maturation of immature DCs either directly or indirectly by inducing the production of TNF and IL-1.

Defensins therefore appear to be important potentiators of T-cell priming.

Indeed, a mixture of HNP1-3, when simultaneously administered with ovalbumin intranasally into C57BL/6 mice, enhanced the production of OVA-specific serum IgG antibody and the generation of IFNγ, IL-5, IL-6, and IL-10 secreting OVA-specific CD4+ T cells. Intraperitoneal injection of HNP1-3 together with keyhole limpet hemocyanin or B-cell lymphoma idiotype Ag into mice not only augmented the levels of Ag-specific IgG but also enhanced the resistance of immunized mice to tumor challenge.

4.2.2 Chemokines

Chemokines are a superfamily of small, heparin-binding cytokines that induce directed migration of various types of leukocytes through interaction with a group of seven-transmembrane G-protein-coupled receptors. TLR activation induces the release of chemokines from tissue macrophages and resident DCs. For example, both TLR2 and TLR4 ligation result in the expression of CCL3, CCL4, and CCL5. TLR2 signaling, however, preferentially leads to secretion of IL-8, while TLR4 signaling leads to CXCL10 production. IL-8 recruits neutrophils to the site of chemokine production, whereas CXCL10 will guide activated T cells into infected tissues. Thus, the discrimination of the pathogen by TLRs and the subsequent production of a distinct subset of chemokines is one of the earliest points at which the immune system tailors its response to specific pathogens.

CCL3 and CCL4 are also crucial for the initial influx of CCR5+ NK cells into infected tissues, as demonstrated in murine models of cytomegalovirus and *Toxoplasma gondii* infection. These NK cells represent an important early source of IFN-γ, providing an environment favorable for Th1 differentiation. IFN-γ in turn induces the expression of the chemokines CXCL9, CXCL10, and CXCL11, instigating the recruitment of activated CXCR3+ T cells and NK cells into the tissue and amplifying the protective response. CXCL10 may also be induced in the absence of IFN-γ following TLR ligation and is also important for the initial influx of CXCR3+ NK cells and T cells.

In terms of modulating humoral immunity, CCL3 treatment stimulated strong antigen-specific serum IgG and IGM responses, while CCL4 administration promoted higher serum IgA and IgE responses, as well as higher titers of antigen-specific mucosal secretory IgA levels in mice. These chemokines may therefore differentially enhance serum and mucosal humoral immune responses.

Immature DCs express several chemokine receptors, including CCR1, CCR5, and CCR6. These are important for recruiting imDCs in response to early pathogen-induced release of CCL3, CCL4, and CCL20, and for keeping them in the tissue. After TLR-mediated activation, however, DCs downmodulate the expression of these chemokine receptors and upregulate the expression of CCR7, allowing them to respond to CCL19 and CCL21 expressed in lymphatic tissues. This serves as a stimulus for DCs to leave the tissue and ultimately migrate into the T-cell rich regions of lymph nodes where priming of T cells occurs. Generally speaking, Th1-dominated responses are associated with the recruitment of CCR5- or CXCR3-bearing cells, whereas Th2-dominated responses are more closely dependent on the influx of CCR3-expressing cells.

Mast cell granule–derived serine proteases also play a major role in recruiting cells to the site of infection. Chymase, a serine protease found in human mast cells,

induces monocyte and neutrophil migration directly. Murine mast cell protease-6, also known as *tryptase*, induces neutrophil extravasation indirectly by enhancing endothelial cell IL-8 production. In addition, histamine is a major mediator synthesized by mast cells. Vasodilation and inflammation are among the prime consequences of its being released from the granules in which it is stored. Histamine promotes the expression of Th2 polarizing (e.g. IL-6 and IL-10) cytokines from human lung macrophages. Histamine also inhibits IL-12 production and augments IL-10 secretion of maturing monocyte-derived DC, resulting in a change from Th1 to Th2 in their T-cell polarizing function. In addition, activated human plasmacytoid DCs respond to histamine by a marked downregulation of IFN-α and TNF.

Most importantly, mast cell granules contain preformed TNF, which is a major mediator of neutrophil influx into sites of mast cell activation. Upon appropriate stimulation, mast cells are also capable of releasing – without degranulation – CCL20, a potent chemoattractant for imDC and T cells.

4.3 Molecules Involved in T- and B-Cell Differentiation

Cytokines provide the most direct link between innate and adaptive immunity because they regulate the communication between dendritic cells, lymphocytes, and other host cells in the course of an immune response. The local milieu of cytokines determines the commitment of T cells to a Th1/Tc1 or Th2/Tc2 type, as well as the immunoglobulin class switch in B cells, and drives the overall immune response toward either tolerance or immunity (Fig. 3).

4.3.1 Th1-inducing Cytokines

Type I Interferons (*IFNs*) were among the first cytokines identified, originally because of their antiviral activity and their effects on the cells of the innate response. IFNs have recently been shown to play an important role in connecting innate and adaptive immunity by acting on T cells and dendritic cells. The term IFN-α/β defines a large group of closely related cytokines, which exert their activity by binding to a common receptor, the IFN-α/β receptor. These cytokines include IFN-α, -β, -δ, -τ, -ω, and -κ.

Although virtually any cell type can express IFN-α/β following viral infection, NIPCs produce 200 to 1000 times more IFN than other blood cells after microbial infection or following stimulation with bacterial components such as LPS and bacterial DNA. IFN-α/β produced by NIPCs and macrophages can enhance the efficiency of the innate immune response by activating natural killer and other host reactive cells.

Type I IFNs provide important signals for the differentiation and activation of DCs. In particular, IFN-α/β promotes the rapid differentiation of GM-CSF-treated human monocytes into DCs with enhanced T-cell stimulatory capacity. Importantly, IFN-α/β plus GM-CSF-matured DCs induced a potent primary antibody response when pulsed with antigen and injected into SCID mice reconstituted with autologous peripheral blood lymphocytes.

In addition, IFN-α/β treatment of imDCs induces phenotypical and functional maturation of these cells, as shown by their expression of DC markers and their migratory response to chemokines. The effects of IFN-α/β on DC maturation imply that they can exert an adjuvant activity. Indeed, experiments in mice revealed that (1) type I IFNs are extremely potent

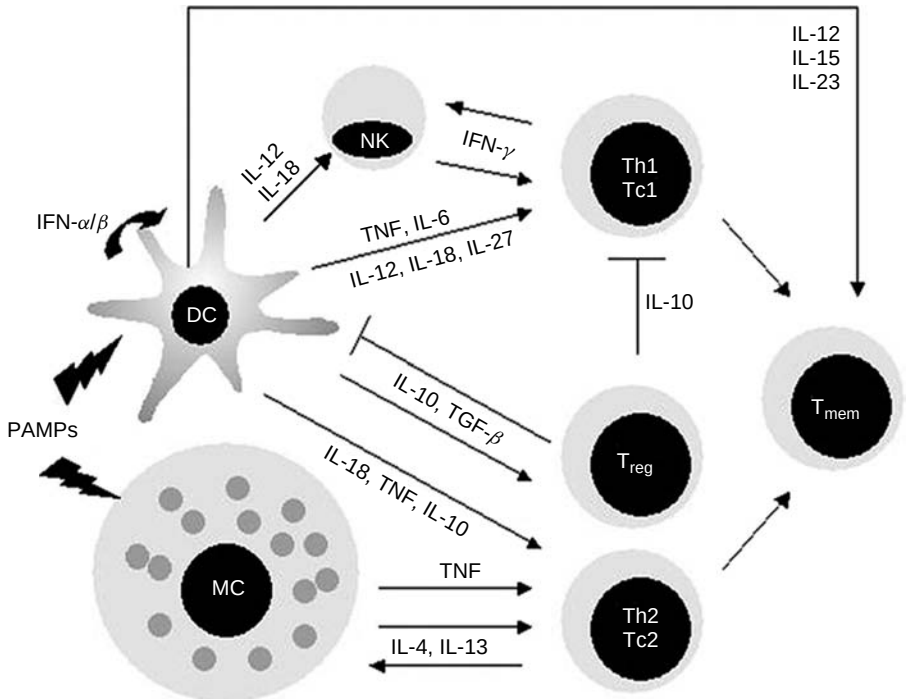

Fig. 3 Major cytokine pathways linking innate and adaptive immune responses. Cells of the immune system, triggered by structures on microbes, release cytokines that skew T-cell development into a T helper 1/T cytolytic 1 (Th1/Tc1) or a Th2/Tc2 pathway. Feedback loops involving IFN-γ (type 1 immunity) or IL-10, TGF-β, IL-4, and IL-13 (type 2 immunity) are indicated. DC (dendritic cell), MC (mast cell), T_{reg} (regulatory, CD4+CD25+ T cell), T_{mem} (memory T cell).

adjuvants when injected into immuncompetent mice with a reference antigen; (2) IFNs directly act on DCs; and (3) the natural adjuvant activity associated with infectious agents is mediated by IFN-α/β.

Interferon-gamma (*IFN-γ*) is a potent activator of the antimicrobial functions of phagocytes, and plays an important role in resistance to many pathogenic bacteria, fungi, and intracellular parasites. It is mainly produced by Th1 cells and natural killer cells, but other T-cell subsets such as NKT cells, CD8+ T cells, and $\gamma\delta$ T cells are also capable of secreting IFN-γ. Furthermore, macrophages, dendritic cells, and B cells are able to produce IFN-γ when appropriately stimulated. Both type I and type II interferons activate natural killer (NK) cells to kill virus-infected cells and release cytokines.

Interleukin-18 (*IL-18*) is a proinflammatory cytokine that belongs to the IL-1 cytokine family, owing to its structure, receptor family, and signal transduction pathways. Similar to IL-1β, IL-18 is synthesized as a precursor requiring caspase-1 for cleavage into an active IL-18 molecule. Produced mainly by antigen-presenting cells, it induces the production of Th1 cytokines, for example, IFN-γ, and enhances cell-mediated cytotoxicity in the presence of IL-12. Interestingly, IL-18, in the absence

of IL-12, induces the production of Th2-related cytokines from T and NK cells and basophils/mast cells. Therefore, IL-18 should be seen as a cytokine that enhances innate immunity and both Th1- and Th2-driven immune responses.

Interleukin-15 (*IL-15*) and IL-2 both utilize the IL-2Rβ and the IL-2Rγ chain as receptor structures; therefore, the two cytokines induce similar biological activities, such as stimulating NK, T, and B cells to proliferate, secrete cytokines, become cytotoxic, or produce antibodies. Nevertheless, there are major differences between these two cytokines in terms of their cellular origins and the mechanisms controlling their synthesis and secretion. In contrast to the predominantly T-cell-restricted expression of IL-2 transcripts, there is widespread constitutive expression of IL-15 mRNA in a variety of tissues including placenta, skeletal muscle, kidney, lung, heart, fibroblasts, epithelial cells, and activated monocytes. However, IL-15 protein was found to be produced only by a limited number of cells such as activated macrophages and epithelial cells.

Findings obtained from gene-targeted mice deficient in individual components of the IL-2/IL-15 system demonstrate distinct roles of IL-15 and IL-2 for the activation of the innate immune system. IL-15 is of pivotal importance for maintaining NK-cell homeostasis, for promoting the differentiation of $\gamma\delta$T cells, and for inducing CD8+ memory T cells. IL-15 also promotes monocyte differentiation into Langerhans' cells. Furthermore, some of the effects induced by IFN-α on the differentiation and/or activation of DCs seem to be mediated by IL-15, which is produced by DCs in response to IFN-α. In addition, IL-15 is a potent inhibitor of apoptosis. Phagocytosis and chemokine production of neutrophils are strongly enhanced by IL-15, and expression of IL-15 in macrophages is upregulated in response to numerous signals that trigger innate immune responses, such as LPS, mycobacteria, or *T. gondii*. Most importantly, exogenous IL-15 has been shown to improve host defense against and/or clearance of Salmonella, *P. falciparum*, and *C. neoformans*.

Interleukin-12 (*IL-12*) is a heterodimeric proinflammatory cytokine formed by a 35-kDa light chain (p35) and a 40-kDa heavy chain (p40), which is produced mainly by DCs and phagocytes (monocytes/macrophages and neutrophils) in response to pathogens during infection. In synergy with TNF and IL-18, IL-12 can induce the production of large amounts of IFN-γ by NK cells, a crucial event in the early phase of the innate immune response.

IL-12 acts not only on NK cells but also on NKT cells and T cells, inducing proliferation, cytotoxic activity, and the differentiation of Th1 cells. T cell and NK cells, in turn, enhance the production of IL-12 through IFN-γ, giving origin to a positive feedback during Th1 immune responses. Surprisingly, both IL-4 and IL-13 are also potent IL-12 inducers. Recent data argue against an absolute requirement for IL-12 for Th1 responses, since two new members of the family of IL-12p40-related heterodimeric cytokines, IL-23 and IL-27, are also involved in the maturation of Th1 responses.

Interleukin-23 (*IL-23*) is a heterodimer, comprising IL-12p40 and the recently cloned IL-23-specific p19 subunit, and binds to a receptor complex composed of the IL-12Rβ1 subunit and a private IL-23R. IL-12 and IL-23 both promote cellular immunity by inducing IFN-γ production and proliferative responses in target cells.

IL-23, which is produced predominantly by macrophages and dendritic cells, does differ from IL-12 in the T-cell subsets that it targets. Whereas IL-12 acts on naïve and effector CD4+ T cells, IL-23 preferentially acts on memory CD4+ T cells. Recent studies of IL-23 receptor expression and IL-23 overexpression in transgenic mice revealed that IL-23 plays a crucial role in autoimmune inflammation in the brain by a direct action on macrophage function.

Interleukin-27 (IL-27) is a new heterodimeric cytokine that consists of an IL-12p40-related protein, EBI-3 (Epstein–Barr virus-induced gene 3), and p28, an IL-12p35-related polypeptide. IL-27 is a strong inducer of IFN-γ production, particularly in synergy with IL-12 and IL-18. Mice deficient for EBI-3 have markedly decreased numbers of invariant NKT cells, which, in addition, produce significantly less IL-4 and IFN-γ compared to normal mice. Thus, EBI-3 also plays a critical regulatory role in the induction of Th2-type immune responses.

IL-27 is produced by activated antigen-presenting cells and is mostly involved in the Th1 commitment of naive CD4+ T cells, mediated by IL-27R-dependent upregulation of the major signal-transducing IL-12Rβ2 chain. In contrast, IL-23 (like IL-15) is critically important for the maintenance of immunological memory, since IL-23R is expressed on memory, but not on naïve T cells.

IL-27 is also involved in modulating inflammatory processes, since mice deficient in a component of the IL-27 receptor (WSX-1) suffer from a vastly enhanced tissue-destroying inflammatory response during infections with *T. cruzi* and *M. tuberculosis*.

Interleukin-6 (IL-6) is a potent differentiation factor for B and T cells. IL-6 produced upon TLR ligation in DCs is a necessary requirement for T-cell activation whenever CD4+CD25 regulatory T cells are present. Indeed, IL-6-deficient mice were severely compromised in their induction of OVA-specific immune responses as long as CD4+CD25+ cells were present, but not when the latter were removed. Therefore, innate immune recognition by TLRs seems to control the activation of adaptive immune responses by at least two distinct mechanisms: the induction of costimulatory molecules on DCs and the production of IL-6, which renders pathogen-specific T cells refractory to the suppressive activity of CD4+CD25+ regulatory T cells. Since release of suppression is necessary for efficient T-cell priming during vaccination, IL-6 is an attractive candidate for an immunopotentiating adjuvant. Indeed, when recombinant human IL-6 was used as an adjuvant in a subunit vaccine against tuberculosis consisting of the culture filtrate proteins of *M. tuberculosis* emulsified in the adjuvant dimethyl-dioctadecylammonium bromide, an increased Th1 response characterized by enhanced IFN-γ production and cell proliferation was observed. Moreover, a specific local IgA response to heterologous antigen was increased in the lungs of mice when genes for murine IL-6 were coexpressed in recombinant vaccinia virus and inoculated intranasally.

Granulocyte-macrophage colony-stimulating factor (GM-CSF) is by far the most widely used hematopoietic growth factor to augment immune responses. It was initially defined as survival and growth factor for hematopoietic progenitor cells, and as a differentiation and activating factor for granulocytic and monocytic cells. GM-CSF plays a crucial role *in vitro* for the differentiation of dendritic cells from precursor cells from bone marrow and peripheral blood.

GM-CSF is secreted by several different cell types, including fibroblasts, endothelial cells, and smooth muscle cells, following stimulation by proinflammatory cytokines such as IL-1 and TNF-α. LPS stimulates GM-CSF secretion from monocytes or macrophages, which could further potentiate the cellular response to LPS. It is known since the early seventies that endotoxin administration in mice raises the serum levels of GM-CSF. The injection of GM-CSF in mice leads to an increase in the numbers of neutrophils and cycling peritoneal macrophages. It was therefore proposed that GM-CSF blockade may have some anti-inflammatory effects and be employed for host protection.

The immunomodulatory effects, which include the activation and augmentation of many of the functions of neutrophils, monocytes, macrophages, and dendritic cells, make this cytokine a candidate for a natural vaccine adjuvant. Indeed, the use of soluble GM-CSF (fusion-protein), as well as gene-transduced, plasmid DNA, in vaccination-based therapies has proven to be relevant for the induction of cellular as well as humoral immune responses, therefore enhancing host defenses against a broad spectrum of invading organisms.

4.3.2 Th2-inducing Cytokines

Interleukin-4 (IL-4) is one of a number of cytokines produced by activated mast cells, and plays a predominant role in many immune and inflammatory reactions. Two biological activities have been defined in detail: IL-4 induces selective IgG1 and IgE isotype switching in B cells and promotes the differentiation of naïve CD4 T cells into a subset of T helper cells that express IL-4 as well as IL-5, IL-10, and IL-13. These cytokines and Ig isotypes are hallmarks of a Th2 response and play a protective role in immunity to extracellular pathogens.

Interleukin-13 (IL-13) belongs to the Th2 cytokine family and is structurally related to IL-4, belonging to the same α-helix protein family. The production of IL-4 and IL-13 is restricted to activated T lymphocytes (in particular, Th2 cells), mast cells, basophils, natural killer cells, and dendritic cells. Both cytokines are functionally active on a wide variety of cell types such as macrophages, NK cells, eosinophils, and mast cells. IL-4 and IL-13 act synergistically to promote DC maturation induced by bacterial products or proinflammatory mediators (such as TNF or CD40 ligand), as evaluated by upregulation of MHC class II and costimulatory molecules. Both cytokines show important immunosuppressive and anti-inflammatory activities, which include the inhibition of proinflammatory cytokines (like IL-1, IL-6, IL-10, IL-12, and TNF-α) and chemokines (like IL-8, CCL3-5, and CCL11).

In mice, a coordinate regulatory sequence element, situated in the intergenic region between the genes for IL-4 and IL-13, was found to be critical for the optimal expression of type 2 cytokines. Mice deficient in this regulator had normal, IL-4-producing mast cells, while their capacity to develop Th2 cells was completely absent *in vitro* and *in vivo*.

Initially, IL-13 was thought to be functionally redundant with IL-4 as a predominant anti-inflammatory factor secreted during type-2 T-cell responses. However, IL-13 possesses a number of additional properties that distinguish it from IL-4: for example, IL-13 plays a key protective role in the expulsion of helminths from the gut and, unlike IL-4, is the major contributor to granulomatous pathology in schistosomiasis.

Therapeutic administration of IL-13 inhibitors in mice successfully prevents allergic responses and Schistosoma

egg–induced lung pathology. Agonist IL-13-treatment in mice, by inducing mastocytosis, eosinophilia, IgE synthesis, and mucus production, protects against ectoparasites and gastrointestinal worms and suppresses inflammation induced by Th1 cytokines. Future studies are required to understand if this scenario holds true in humans.

Interleukin-10 (IL-10) was initially described as a mouse Th2 cell product, but is now known to be produced by a wide range of cell types, including T cells, B cells, keratinocytes, monocytes/macrophages, and dendritic cells. It activates B lymphocytes and NK cells and has an inhibitory function on T cell and macrophage activation. IL-10 not only reduces the production of proinflammatory cytokines such as TNF-α, IL-1, IL-6, IL-12, IFN-γ, and GM-CSF but also of chemokines like IL-8 and CCL3. Furthermore, IL-10 downregulates the production of molecules important in triggering specific immunity such as MHC class II and costimulatory molecules, CD86, therefore inhibiting antigen-presenting cell maturation.

Some bacterial pathogens, like mycobacteria, *Bordetella pertussis*, or *Yersinia enterocolitica*, have developed mechanisms for modulating cytokine production by host cells. These bacteria can induce the production of anti-inflammatory cytokines, such as IL-10 and TGF-β, which dampen the immune response. These immunosuppressive cytokines may also contribute to the generation of so-called regulatory T cells, which downregulate immune activation.

It is noteworthy that IL-10 also is a potent stimulator of NK cell activity. Cytotoxic lysis of infected cells may contribute to the clearance of pathogens and facilitate antigen acquisition from dead cells for cross-priming by activated dendritic cells, providing a link between the innate and the adaptive immune responses.

The transforming growth factor (*TGF*)-β, family comprises three closely related cytokines with potent immunoregulatory proteins that influence both positively and negatively the process of inflammation and repair. TGF-β regulates processes such as angiogenesis, chemotaxis, fibroblast proliferation, and the controlled synthesis and degradation of matrix proteins, for example, collagen and fibronectin.

Many cell types express TGF-β constitutively and produce more when stimulated. These include macrophages, T and B cells, platelets, and fibroblasts. Importantly, TGF-β is also produced by regulatory T cells, which downmodulate the function of Th1 and other immune cells. TGF-β inhibits immune functions of T and B cells by reducing TNF-α and IL-1 production in peripheral blood monocytes, and directly inhibits DC maturation. In a model of murine malaria, low amounts of active TGF-β production by splenic mononuclear cells were associated with a lethal outcome, and anti-TGF-β treatment transformed a normal resolving infection into a lethal one characterized by a hyperproduction of proinflammatory cytokines.

5 Practical Implications

Because TLRs play a central role in directing both innate and adaptive immune responses, reagents ligating TLRs or interfering with TLR ligation are currently the subject of intense investigation in both preventive and therapeutic medicine.

Inactivated pathogens and highly purified recombinant proteins or peptides without adjuvants do not elicit efficient

Th cell responses. Therefore, it is imperative to optimize strategies that will enhance antigen presentation by inducing a favorable costimulatory milieu. Biological adjuvants do this by activating TLRs, and thereby, stimulating innate immune responses.

Live carrier vaccines, in particular, have the advantage of bringing a variety of strongly activating PAMPs with them. For example, the widely used tuberculosis vaccine strain, *Mycobacterium bovis* BCG, is a powerful inducer of T-cell responses because it contains PG, LAM, lipoproteins, lipopeptides, and CpG motifs as TLR agonists, and because it can persist for some time within antigen-presenting cells, providing long-lasting T-cell priming. However, the particular combination of PAMPs in a given bacterial strain may also favor some undesired responses because of a significant cross-regulation of PAMPs interacting with different pattern-recognition receptors simultaneously. For instance, ligation of the mannose receptor (with ManLAM, present on certain mycobacterial strains) at the same time as stimulation of TLRs (e.g. TLR4 with LPS) reduces the production of IL-12 in DCs compared to DCs stimulated with the TLR agonist only. This may effectively diminish the adjuvant Th1 inducing potency of, for example, a BCG carrier vaccine. Similar mechanisms may be operative in other bacterial carrier systems, although the precise ligands and receptors involved are probably quite different. As a consequence, the combination of highly purified, recombinant molecules (PAMPs, chemokines, and cytokines) may prove to be superior to bacterial vaccine strains, because their individual dose and composition can be modified to fit the particular requirements of the vaccinee.

Conversely, interfering with TLR ligation and/or signaling may prove to be a worthwhile anti-inflammatory strategy. For example, septic shock continues to be fraught with a very high mortality rate, and attempts to modulate mediator cytokine secretion or function (for example, with anti-TNF antibodies) have not been successful. It appears straightforward and attractive to blunt the overzealous inflammatory response by inhibiting sensory input, for example, by blocking TLR ligation using synthetic compounds that bind to TLRs without activating them. Although far from eliminating the actual cause for the inflammatory response, this would interfere with a very proximal event in the signaling cascade leading to organ failure. Also, depending on whether signal transducers such as MyD88 or TRIF can be effectively blocked, this would provide a broader approach than neutralizing endotoxin (TLR4-ligand) or lipoteichoic acid (TLR2-ligand) only.

See also Antigen Presenting Cells (APCs); Autoantibodies and Autoimmunity; Cellular Interactions; Immune Defence, Cell Mediated; Immunology; Innate Immunity; Mucosal Vaccination.

Bibliography

Books and Reviews

Akira, S., Takeda, K. (2004) Toll-like receptor signaling, *Nat. Rev. Immunol.* **4**, 499–511.

Athman, R., Philpott, D. (2004) Innate immunity via Toll-like receptors and Nod proteins, *Curr. Opin. Microbiol.* **7**, 25–32.

Ehlers, S. (2004) Adaptive immunity in the absence of innate immune responses? The

un-Tolled truth of the silent invaders, *Eur. J. Immunol.* **34**, 1783–1788.

Ezekowitz, R.A.B., Hoffmann, J.A. (Eds.) (2002) *Innate Immunity*, Human Press, Totowa, NJ.

Janeway, C.A. Jr., Medzhitov, R. (2002) Innate immune recognition, *Annu. Rev. Immunol.* **20**, 197–216.

Kaufmann, S.H.E., Medzhitov, R., Gordon, S. (Eds.) (2004) *The Innate Immune Response to Infection*, ASM press, Washington, D.C.

Kopp, E., Medzhitov, R. (2003) Recognition of microbial infection by Toll-like receptors, *Curr. Opin. Immunol.* **15**, 396–401.

Oppenheim, J.J., Feldmann, M. (Eds.) (2001) *Cytokine Reference*, Academic Press, San Diego, CA.

Takeda, K., Kaisho, T., Akira, S. (2003) Toll-like receptors, *Annu. Rev. Immunol.* **21**, 335–376.

Primary Literature

Alexander, C., Rietschel E.T. (2001) Bacterial lipopolysaccharides and innate immunity, *J. Endotoxin Res.* **7**, 167–202.

Alexopoulou, L., Holt, A.C., Medzhitov, R., Flavell, R.A. (2001) Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3, *Nature* **413**, 732–738.

Aliberti, J., Reis e, Sousa., Schito, M., Hieny, S., Wells, T., Huffnagle, G.B., Sher, A. (2000) CCR5 provides a signal for microbial induced production of IL-12 by CD8 alpha+ dendritic cells, *Nat. Immunol.* **1**, 83–87.

Almeida, I.C., Camargo, M.M., Procopio, D.O., Silva, L.S., Mehlert, A., Travassos, L.R., Gazzinelli, R.T., Ferguson, M.A. (2000) Highly purified glycosylphosphatidylinositols from Trypanosoma cruzi are potent proinflammatory agents, *EMBO J.* **19**, 1476–1485.

Armitage, J.O. (1998) Emerging applications of recombinant human granulocyte-macrophage colony-stimulating factor, *Blood* **92**, 4491–4508.

Asselin-Paturel, C., Boonstra, A., Dalod, M., Durand, I., Yessaad, N., Dezutter-Dambuyant, C., Vicari, A., O'Garra, A., Biron, C., Briere, F., Trinchieri, G. (2001) Mouse type I IFN-producing cells are immature APCs with plasmacytoid morphology, *Nat. Immunol.* **2**, 1144–1150.

Banchereau, J., Steinman, R.M. (1998) Dendritic cells and the control of immunity, *Nature* **392**, 245–252.

Bancroft, A.J., Artis, D., Donaldson, D.D., Sypek, J.P., Grencis, R.K. (2000) Gastrointestinal nematode expulsion in IL-4 knockout mice is IL-13 dependent, *Eur. J. Immunol.* **30**, 2083–2091.

Befus, A.D., Mowat, C., Gilchrist, M., Hu, J., Solomon, S., Bateman, A. (1999) Neutrophil defensins induce histamine secretion from mast cells: mechanisms of action, *J. Immunol.* **163**, 947–953.

Bendelac, A., Rivera, M.N., Park, S.H., Roark, J.H. (1997) Mouse CD1-specific NK1 T cells: development, specificity, and function, *Annu. Rev. Immunol.* **15**, 535–562.

Bulfone-Paus, S., Ungureanu, D., Pohl, T., Lindner, G., Paus, R., Ruckert, R., Krause, H., Kunzendorf, U. (1997) Interleukin-15 protects from lethal apoptosis in vivo, *Nat. Med.* **3**, 1124–1128.

Campos, M.A., Almeida, I.C., Takeuchi, O., Akira, S., Valente, E.P., Procopio, D.O., Travassos, L.R., Smith, J.A., Golenbock, D.T., Gazzinelli, R.T. (2001) Activation of Toll-like receptor-2 by glycosylphosphatidylinositol anchors from a protozoan parasite, *J. Immunol.* **167**, 416–423.

Cella, M., Facchetti, F., Lanzavecchia, A., Colonna, M. (2000) Plasmacytoid dendritic cells activated by influenza virus and CD40L drive a potent TH1 polarization, *Nat. Immunol.* **1**, 305–310.

Cella, M., Jarrossay, D., Facchetti, F., Alebardi, O., Nakajima, H., Lanzavecchia, A., Colonna, M. (1999) Plasmacytoid monocytes migrate to inflamed lymph nodes and produce large amounts of type I interferon, *Nat. Med.* **5**, 919–923.

Chiaramonte, M.G., Schopf, L.R., Neben, T.Y., Cheever, A.W., Donaldson, D.D., Wynn, T.A. (1999) IL-13 is a key regulatory cytokine for Th2 cell-mediated pulmonary granuloma formation and IgE responses induced by Schistosoma mansoni eggs, *J. Immunol.* **162**, 920–930.

Cua, D.J., Sherlock, J., Chen, Y., Murphy, C.A., Joyce, B., Seymour, B., Lucian, L., To, W., Kwan, S., Churakova, T., Zurawski, S., Wiekowski, M., Lira, S.A., Gorman, D., Kastelein, R.A., Sedgwick, J.D. (2003) Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain, *Nature* **421**, 744–748.

Dalod, M., Salazar-Mather, T.P., Malmgaard, L., Lewis, C., Asselin-Paturel, C., Briere, F., Trinchieri, G., Biron, C.A. (2002) Interferon alpha/beta and interleukin 12 responses to viral infections: pathways regulating dendritic cell cytokine expression in vivo, *J. Exp. Med.* **195**, 517–528.

Deininger, S., Stadelmaier, A., Von Aulock, S., Morath, S., Schmidt, R.R., Hartung, T. (2003) Definition of structural prerequisites for lipoteichoic Acid-inducible cytokine induction by synthetic derivatives, *J. Immunol.* **170**, 4134–4138.

Dorman, S.E., Holland, S.M. (2000) Interferon-gamma and interleukin-12 pathway defects and human disease, *Cytokine Growth Factor Rev.* **11**, 321–333.

Echtenacher, B., Mannel, D.N., Hultner, L. (1996) Critical protective role of mast cells in a model of acute septic peritonitis, *Nature* **381**, 75–77.

Feger, F., Varadaradjalou, S., Gao, Z., Abraham, S.N., Arock, M. (2002) The role of mast cells in host defense and their subversion by bacterial pathogens, *Trends Immunol.* **23**, 151–158.

Fehniger, T.A., Caligiuri, M.A. (2001) Interleukin 15: biology and relevance to human disease, *Blood* **97**, 14–32.

Garcia, J.R., Krause, A., Schulz, S., Rodriguez-Jimenez, F.J., Kluver, E., Adermann, K., Forssmann, U., Frimpong-Boateng, A., Bals, R., Forssmann, W.G. (2001) Human beta-defensin 4: a novel inducible peptide with a specific salt-sensitive spectrum of antimicrobial activity, *FASEB J.* **15**, 1819–1821.

Gerosa, F., Baldani-Guerra, B., Nisii, C., Marchesini, V., Carra, G., Trinchieri, G. (2002) Reciprocal activating interaction between natural killer cells and dendritic cells, *J. Exp. Med.* **195**, 327–333.

Gonzalez-Aseguinolaza, G., Van Kaer, L., Bergmann, C.C., Wilson, J.M., Schmieg, J., Kronenberg, M., Nakayama, T., Taniguchi, M., Koezuka, Y., Tsuji, M. (2002) Natural killer T cell ligand alpha-galactosylceramide enhances protective immunity induced by malaria vaccines, *J. Exp. Med.* **195**, 617–624.

Govinden, R., Bhoola, K.D. (2003) Genealogy, expression, and cellular function of transforming growth factor-beta, *Pharmacol. Ther.* **98**, 257–265.

Gracie, J.A., Robertson, S.E., McInnes, I.B. (2003) Interleukin-18, *J. Leukoc. Biol.* **73**, 213–224.

Grouard, G., Rissoan, M.C., Filgueira, L., Durand, I., Banchereau, J., Liu, Y.J. (1997) The enigmatic plasmacytoid T cells develop into dendritic cells with interleukin (IL)-3 and CD40-ligand, *J. Exp. Med.* **185**, 1101–1111.

Hacker, H., Mischak, H., Miethke, T., Liptay, S., Schmid, R., Sparwasser, T., Heeg, K., Lipford, G.B., Wagner, H. (1998) CpG-DNA-specific activation of antigen-presenting cells requires stress kinase activity and is preceded by non-specific endocytosis and endosomal maturation, *EMBO J.* **17**, 6230–6240.

Hayashi, F., Smith, K.D., Ozinsky, A., Hawn, T.R., Yi, E.C., Goodlett, D.R., Eng, J.K., Akira, S., Underhill, D.M., Aderem, A. (2001) The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5, *Nature* **410**, 1099–1103.

Heil, F., Hemmi, H., Hochrein, H., Ampenberger, F., Kirschning, C., Akira, S., Lipford, G., Wagner, H., Bauer, S. (2004) Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8, *Science* **303**, 1526–1529.

Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., Matsumoto, M., Hoshino, K., Wagner, H., Takeda, K., Akira, S. (2000) A Toll-like receptor recognizes bacterial DNA, *Nature* **408**, 740–745.

Hickling, T.P., Clark, H., Malhotra, R., Sim, R.B. (2004) Collectins and their role in lung immunity, *J. Leukoc. Biol.* **75**, 27–33.

Horng, T., Barton, G.M., Medzhitov, R. (2001) TIRAP: an adapter molecule in the Toll signaling pathway, *Nat. Immunol.* **2**, 835–841.

Inaba, K., Inaba, M., Romani, N., Aya, H., Deguchi, M., Ikehara, S., Muramatsu, S., Steinman, R.M. (1992) Generation of large numbers of dendritic cells from mouse bone marrow cultures supplemented with granulocyte/macrophage colony-stimulating factor, *J. Exp. Med.* **176**, 1693–1702.

Jarrossay, D., Napolitani, G., Colonna, M., Sallusto, F., Lanzavecchia, A. (2001) Specialization and complementarity in microbial molecule recognition by human myeloid and plasmacytoid dendritic cells, *Eur. J. Immunol.* **31**, 3388–3393.

Kadowaki, N., Antonenko, S., Lau, J.Y., Liu, Y.J. (2000) Natural interferon alpha/beta-producing cells link innate and adaptive immunity, *J. Exp. Med.* **192**, 219–226.

Kadowaki, N., Ho, S., Antonenko, S., Malefyt, R.W., Kastelein, R.A., Bazan, F., Liu, Y.J. (2001) Subsets of human dendritic cell precursors express different toll-like receptors and respond to different microbial antigens, *J. Exp. Med.* **194**, 863–869.

Kawai, T., Adachi, O., Ogawa, T., Takeda, K., Akira, S. (1999) Unresponsiveness of MyD88-deficient mice to endotoxin, *Immunity* **11**, 115–122.

Kennedy, M.K., Glaccum, M., Brown, S.N., Butz, E.A., Viney, J.L., Embers, M., Matsuki, N., Charrier, K., Sedger, L., Willis, C.R., Brasel, K., Morrissey, P.J., Stocking, K., Schuh, J.C., Joyce, S., Peschon, J.J. (2000) Reversible defects in natural killer and memory CD8 T cell lineages in interleukin 15-deficient mice, *J. Exp. Med.* **191**, 771–780.

Khan, I.A., MacLean, J.A., Lee, F.S., Casciotti, L., DeHaan, E., Schwartzman, J.D., Luster, A.D. (2000) IP-10 is critical for effector T cell trafficking and host survival in Toxoplasma gondii infection, *Immunity* **12**, 483–494.

Kobayashi, H., Ishizuka, T., Okayama, Y. (2000) Human mast cells and basophils as sources of cytokines, *Clin. Exp. Allergy* **30**, 1205–1212.

Langenkamp, A., Messi, M., Lanzavecchia, A., Sallusto, F. (2000) Kinetics of dendritic cell activation: impact on priming of TH1, TH2 and nonpolarized T cells, *Nat. Immunol.* **1**, 311–316.

Le Bon, A., Tough, D.F. (2002) Links between innate and adaptive immunity via type I interferon, *Curr. Opin. Immunol.* **14**, 432–436.

Lehrer, R.I., Ganz, T. (2002) Defensins of vertebrate animals, *Curr. Opin. Immunol.* **14**, 96–102.

Lillard, J.W. Jr., Boyaka, P.N., Chertov, O., Oppenheim, J.J., McGhee, J.R. (1999) Mechanisms for induction of acquired host immunity by neutrophil peptide defensins, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 651–656.

Lillard, J.W. Jr., Singh, U.P., Boyaka, P.N., Singh, S., Taub, D.D., McGhee, J.R. (2003) MIP-1alpha and MIP-1beta differentially mediate mucosal and systemic adaptive immunity, *Blood* **101**, 807–814.

Liu, Y.J. (2001) Dendritic cell subsets and lineages, and their functions in innate and adaptive immunity, *Cell* **106**, 259–262.

Luster, A.D. (2002) The role of chemokines in linking innate and adaptive immunity, *Curr. Opin. Immunol.* **14**, 129–135.

Macatonia, S.E., Hosken, N.A., Litton, M., Vieira, P., Hsieh, C.S., Culpepper, J.A., Wysocka, M., Trinchieri, G., Murphy, K.M., O'Garra, A. (1995) Dendritic cells produce IL-12 and direct the development of Th1 cells from naive CD4+ T cells, *J. Immunol.* **154**, 5071–5079.

Marone, G., Galli, S.J., Kitamura, Y. (2002) Probing the roles of mast cells and basophils in natural and acquired immunity, physiology and disease, *Trends Immunol.* **23**, 425–427.

Maurer, M., Echtenacher, B., Hultner, L., Kollias, G., Mannel, D.N., Langley, K.E., Galli, S.J. (1998) The c-kit ligand, stem cell factor, can enhance innate immunity through effects on mast cells, *J. Exp. Med.* **188**, 2343–2348.

Mazzoni, A., Young, H.A., Spitzer, J.H., Visintin, A., Segal, D.M. (2001) Histamine regulates cytokine production in maturing dendritic cells, resulting in altered T cell polarization, *J. Clin. Invest.* **108**, 1865–1873.

Mohrs, M., Blankespoor, C.M., Wang, Z.E., Loots, G.G., Afzal, V., Hadeiba, H., Shinkai, K., Rubin, E.M., Locksley, R.M. (2001) Deletion of a coordinate regulator of type 2 cytokine expression in mice, *Nat. Immunol.* **2**, 842–847.

Moingeon, P., Haensler, J., Lindberg, A. (2001) Towards the rational design of Th1 adjuvants, *Vaccine* **19**, 4363–4372.

Moore, K.W., de Waal, M.R., Coffman, R.L., O'Garra, A. (2001) Interleukin-10 and the interleukin-10 receptor, *Annu. Rev. Immunol.* **19**, 683–765.

Morath, S., Stadelmaier, A., Geyer, A., Schmidt, R.R., Hartung, T. (2002) Synthetic lipoteichoic acid from Staphylococcus aureus is a potent stimulus of cytokine release, *J. Exp. Med.* **195**, 1635–1640.

Mosmann, T.R., Coffman, R.L. (1989) Heterogeneity of cytokine secretion patterns and functions of helper T cells, *Adv. Immunol.* **46**, 111–147.

Moustakas, A., Pardali, K., Gaal, A., Heldin, C.H. (2002) Mechanisms of TGF-beta signaling in regulation of cell growth and differentiation, *Immunol. Lett.* **82**, 85–91.

Murdoch, C., Finn, A. (2000) Chemokine receptors and their role in inflammation and infectious diseases, *Blood* **95**, 3032–3043.

Musso, T., Calosso, L., Zucca, M., Millesimo, M., Puliti, M., Bulfone-Paus, S., Merlino, C., Savoia, D., Cavallo, R., Ponzi, A.N., Badolato, R. (1998) Interleukin-15 activates proinflammatory and antimicrobial functions in polymorphonuclear cells, *Infect. Immun.* **66**, 2640–2647.

Muzio, M., Ni, J., Feng, P., Dixit, V.M. (1997) IRAK (Pelle) family member IRAK-2 and MyD88 as proximal mediators of IL-1 signaling, *Science* **278**, 1612–1615.

Muzio, M., Bosisio, D., Polentarutti, N., D'amico, G., Stoppacciaro, A., Mancinelli, R., van't Veer, C., Penton-Rol, G., Ruco, L.P., Allavena, P., Mantovani, A. (2000) Differential expression and regulation of toll-like receptors (TLR) in human leukocytes: selective expression of TLR3 in dendritic cells, *J. Immunol.* **164**, 5998–6004.

Nathan, C., Shiloh, M.U. (2000) Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8841–8848.

Nigou, J., Zelle-Rieser, C., Gilleron, M., Thurnher, M., Puzo, G. (2001) Mannosylated lipoarabinomannans inhibit IL-12 production by human dendritic cells: evidence for a negative signal delivered through the mannose receptor, *J. Immunol.* **166**, 7477–7485.

Okano, M., Satoskar, A.R., Nishizaki, K., Harn, D.A. Jr. (2001) Lacto-N-fucopentose III found on Schistosoma mansoni egg antigens functions as adjuvant for proteins by inducing Th2-type response, *J. Immunol.* **167**, 442–450.

Oppmann, B., Lesley, R., Blom, B., Timans, J.C., Xu, Y., Hunte, B., Vega, F., Yu, N., Wang, J., Singh, K., Zonin, F., Vaisberg, E., Churakova, T., Liu, M., Gorman, D., Wagner, J., Zurawski, S., Liu, Y., Abrams, J.S., Moore, K.W., Rennick, D., Waal-Malefyt, R., Hannum, C., Bazan, J.F., Kastelein, R.A. (2000) Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12, *Immunity* **13**, 715–725.

Ozinsky, A., Underhill, D.M., Fontenot, J.D., Hajjar, A.M., Smith, K.D., Wilson, C.B., Schroeder, L., Aderem, A. (2000) The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 13766–13771.

Parlato, S., Santini, S.M., Lapenta, C., Di Pucchio, T., Logozzi, M., Spada, M., Giammarioli, A.M., Malorni, W., Fais, S., Belardelli, F. (2001) Expression of CCR-7, MIP-3beta, and Th-1 chemokines in type I IFN-induced monocyte-derived dendritic cells: importance for the rapid acquisition of potent migratory and functional activities, *Blood* **98**, 3022–3029.

Pasare, C., Medzhitov, R. (2003) Toll pathway-dependent blockade of CD4+CD25+ T cell-mediated suppression by dendritic cells, *Science* **299**, 1033–1036.

Pflanz, S., Timans, J.C., Cheung, J., Rosales, R., Kanzler, H., Gilbert, J., Hibbert, L., Churakova, T., Travis, M., Vaisberg, E., Blumenschein, W.M., Mattson, J.D., Wagner, J.L., To, W., Zurawski, S., McClanahan, T.K., Gorman, D.M., Bazan, J.F., de Waal, M.R., Rennick, D., Kastelein, R.A. (2002) IL-27, a heterodimeric cytokine composed of EBI3 and p28 protein, induces proliferation of naive CD4(+) T cells, *Immunity* **16**, 779–790.

Raj, P.A., Dentino, A.R. (2002) Current status of defensins and their role in innate and adaptive immunity, *FEMS Microbiol. Lett.* **206**, 9–18.

Reiling, N., Holscher, C., Fehrenbach, A., Kroger, S., Kirschning, C.J., Goyert, S., Ehlers, S. (2002) Cutting edge: Toll-like receptor (TLR)2- and TLR4-mediated pathogen recognition in resistance to airborne infection with Mycobacterium tuberculosis, *J. Immunol.* **169**, 3480–3484.

Reis e Sousa, C, Hieny, S., Scharton-Kersten, T., Jankovic, D., Charest, H., Germain, R.N., Sher, A. (1997) In vivo microbial stimulation induces rapid CD40 ligand-independent production of interleukin 12 by dendritic cells and their redistribution to T cell areas, *J. Exp. Med.* **186**, 1819–1829.

Sallusto, F., Lanzavecchia, A. (2000) Understanding dendritic cell and T-lymphocyte traffic through the analysis of chemokine receptor expression, *Immunol. Rev.* **177**, 134–140.

Schnare, M., Barton, G.M., Holt, A.C., Takeda, K., Akira, S., Medzhitov, R. (2001) Toll-like receptors control activation of adaptive immune responses, *Nat. Immunol.* **2**, 947–950.

Schramm, G., Falcone, F.H., Gronow, A., Haisch, K., Mamat, U., Doenhoff, M.J., Oliveira, G., Galle, J., Dahinden, C.A., Haas, H. (2003) Molecular characterization of an Interleukin-4-inducing factor from

Schistosoma mansoni eggs, *J. Biol. Chem.* **278**, 18384–92.

Schromm, A.B., Lien, E., Henneke, P., Chow, J.C., Yoshimura, A., Heine, H., Latz, E., Monks, B.G., Schwartz, D.A., Miyake, K., Golenbock, D.T. (2001) Molecular genetic analysis of an endotoxin nonresponder mutant cell line: a point mutation in a conserved region of MD-2 abolishes endotoxin-induced signaling, *J. Exp. Med.* **194**, 79–88.

Shtrichman, R., Samuel, C.E. (2001) The role of gamma interferon in antimicrobial immunity, *Curr. Opin. Microbiol.* **4**, 251–259.

Sozzani, S., Allavena, P., Vecchi, A., Mantovani, A. (1999) The role of chemokines in the regulation of dendritic cell trafficking, *J. Leukoc. Biol.* **66**, 1–9.

Srivastava, P. (2002) Roles of heat-shock proteins in innate and adaptive immunity, *Nat. Rev. Immunol.* **2**, 185–194.

Supajatura, V., Ushio, H., Nakao, A., Akira, S., Okumura, K., Ra, C., Ogawa, H. (2002) Differential responses of mast cell Toll-like receptors 2 and 4 in allergy and innate immunity, *J. Clin. Invest.* **109**, 1351–1359.

Swain, S.L., Bradley, L.M., Croft, M., Tonkonogy, S., Atkins, G., Weinberg, A.D., Duncan, D.D., Hedrick, S.M., Dutton, R.W., Huston, G. (1991) Helper T-cell subsets: phenotype, function and the role of lymphokines in regulating their development, *Immunol. Rev.* **123**, 115–144.

Takeda, A., Hamano, S., Yamanaka, A., Hanada, T., Ishibashi, T., Mak, T.W., Yoshimura, A., Yoshida, H. (2003) Cutting edge: role of IL-27/WSX-1 signaling for induction of T-bet through activation of STAT1 during initial Th1 commitment, *J. Immunol.* **170**, 4886–4890.

Takeuchi, O., Hoshino, K., Kawai, T., Sanjo, H., Takada, H., Ogawa, T., Takeda, K., Akira, S. (1999) Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components, *Immunity* **11**, 443–451.

Takeuchi, O., Kawai, T., Muhlradt, P.F., Morr, M., Radolf, J.D., Zychlinsky, A., Takeda, K., Akira, S. (2001) Discrimination of bacterial lipoproteins by Toll-like receptor 6, *Int. Immunol.* **13**, 933–940.

Tani, K., Ogushi, F., Kido, H., Kawano, T., Kunori, Y., Kamimura, T., Cui, P., Sone, S. (2000) Chymase is a potent chemoattractant for human monocytes and neutrophils, *J. Leukoc. Biol.* **67**, 585–589.

Tani, K., Murphy, W.J., Chertov, O., Salcedo, R., Koh, C.Y., Utsunomiya, I., Funakoshi, S., Asai, O., Herrmann, S.H., Wang, J.M., Kwak, L.W., Oppenheim, J.J. (2000) Defensins act as potent adjuvants that promote cellular and humoral immune responses in mice to a lymphoma idiotype and carrier antigens, *Int. Immunol.* **12**, 691–700.

Territo, M.C., Ganz, T., Selsted, M.E., Lehrer, R. (1989) Monocyte-chemotactic activity of defensins from human neutrophils, *J. Clin. Invest.* **84**, 2017–2020.

Thoma-Uszynski, S., Stenger, S., Takeuchi, O., Ochoa, M.T., Engele, M., Sieling, P.A., Barnes, P.F., Rollinghoff, M., Bolcskei, P.L., Wagner, M., Akira, S., Norgard, M.V., Belisle, J.T., Godowski, P.J., Bloom, B.R., Modlin, R.L. (2001) Induction of direct antimicrobial activity through mammalian toll-like receptors, *Science* **291**, 1544–1547.

Van Wetering, S., Mannesse-Lazeroms, S.P., Dijkman, J.H., Hiemstra, P.S. (1997) Effect of neutrophil serine proteinases and defensins on lung epithelial cells: modulation of cytotoxicity and IL-8 production, *J. Leukoc. Biol.* **62**, 217–226.

Wang, C., Deng, L., Hong, M., Akkaraju, G.R., Inoue, J., Chen, Z.J. (2001) TAK1 is a ubiquitin-dependent kinase of MKK and IKK, *Nature* **412**, 346–351.

Watanabe, T., Kitani, A., Murray, P.J., Strober, W. (2004) NOD2 is a negative regulator of Toll-like receptor 2-mediated T helper type 1 responses, *Nat. Immunol.* **5**, 800–808.

Weighardt, H., Jusek, G., Mages, J., Lang, R., Hoebe, K., Beutler, B., Holzmann, B. (2004) Identification of a TLR4- and TRIF-dependent activation program of dendritic cells, *Eur. J. Immunol.* **34**, 558–564.

Wright, S.D., Ramos, R.A., Tobias, P.S., Ulevitch, R.J., Mathison, J.C. (1990) CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein, *Science* **249**, 1431–1433.

Wynn, T.A. (2003) IL-13 effector functions, *Annu. Rev. Immunol.* **21**, 425–456.

Yang, D., Biragyn, A., Kwak, L.W., Oppenheim, J.J. (2002) Mammalian defensins in immunity: more than just microbicidal, *Trends Immunol.* **23**, 291–296.

Yang, D., Chen, Q., Chertov, O., Oppenheim, J.J. (2000) Human neutrophil defensins selectively chemoattract naive T and immature dendritic cells, *J. Leukoc. Biol.* **68**, 9–14.

Yoshie, O., Imai, T., Nomiyama, H. (2001) Chemokines in immunity, *Adv. Immunol.* **78**, 57–110.

Zhang, D., Zhang, G., Hayden, M.S., Greenblatt, M.B., Bussey, C., Flavell, R.A., Ghosh, S. (2004) A toll-like receptor that prevents infection by uropathogenic bacteria, *Science* **303**, 1522–1526.

Pathophysiology of Stroke

Andreas Meisel, Konstantin Prass, Tilo Wolf, and Ulrich Dirnagl
Department of Experimental Neurology (Charitá) Humboldt University, Berlin, Germany

1 **Introduction 117**

2 **The Penumbra Concept 120**

3 **Excitotoxicity 120**

4 **Oxygen Free Radicals 121**

5 **Tissue Acidosis 123**

6 **Peri-infarct Depolarizations 124**

7 **Inflammation 124**

8 **Damage to the Blood–Brain barrier 126**

9 **Programmed Cell Death and Apoptosis 127**

10 **Ischemia-induced DNA Damage, DNA Repair, and p53 as Genotoxic Sensor 130**

11 **Epigenetic Mechanisms 131**

12 **Gene Expression 132**

13 **Cell Replacement 133**

14 **Endogenous Neuroprotection – Ischemic Tolerance 134**

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

15 Stroke-induced Immune Depression (SIDS) 137

16 Perspectives 139

Bibliography 140
Books and Reviews 140
Primary Literature 140

Keywords

Agnosia
It is the inability to recognize and identify objects or persons despite having knowledge of the characteristics of those objects or persons.

Amaurosis
It is a partial or complete loss of sight.

Apraxia
It is the inability to execute a voluntary motor movement despite being able to demonstrate normal muscle function.

Ataxia
It is an incoordination and lack of balance.

Dyslexia
It is a difficulty (in learning) to read.

Dysphasia
It is a group of speech disorders with an impairment of expression by speech, or an impairment of the comprehension of spoken language. More severe forms of dysphasia are called aphasia.

Excitotoxicity
It is an excessive exposure to glutamate or related compounds killing brain neurons by overstimulating them.

Hemianopia
It is a blindness or reduction in vision in one half of the visual field.

Hemiparesis
It is a weakness on one side of the body.

Ischemic Tolerance
Ischemic tolerance or ischemic preconditioning is an intrinsic adaptive response to a mild ischemic stimulus protecting from an otherwise "lethal" ischemic event.

Penumbra
It is an area of constrained blood flow with partially preserved energy metabolism lying between the lethally damaged core and the normal brain.

SAGE
Serial analysis of gene expression, a technique for profiling gene expression.

The current understanding of stroke pathophysiology is basically founded on experimental models. Among these, *in vitro* models with primary cultures of cerebral cells permit stroke pathophysiology to be examined on a molecular level, while animal models (mainly mouse and rat) are used to evaluate medical intervention. There is a highly complex sequence of events leading to the eventual ischemic cerebral damage that follows a well-defined spatiotemporal pattern. Overwhelming excitotoxicity leads to early necrotic cell death in what is to become the core of the infarction, while the tissue damage in the surrounding zone called *penumbra* happens on a longer timescale. The excitotoxic or inflammatory mechanisms are milder, bearing the biochemical hallmarks of apoptosis. The brain cells, challenged by such a large-scale assault, activate endogenous protective programs. These have been studied by experimentally inducing ischemic tolerance (i.e. ischemic preconditioning). Importantly, cerebral ischemia not only affects the brain but also impacts other systems. For example, stroke induces dramatic immunodepression through overactivation of the sympathetic nervous system. As a result, severe bacterial infections such as pneumonia occur. The complex signaling cascades not only decide over cell survival in the brain and the neurological deficit but also over mortality after stroke from extracerebral complications. Their ability to govern not only the maturation of the eventual infarction but also the immune system make them a promising target for intervention and the development of neuroprotective drugs.

1 Introduction

In the United States, more than 600 000 people per year suffer from strokes, and, among acutely hospitalized neurological patients, stroke patients make up the largest share – about 50%. Currently, there are about 4 million stroke survivors in the United States. Mortality from stroke is an estimated 25%, which makes it the third major cause of death in industrialized countries. Because of the high rate of severe permanent disability, stroke is a burden not only for the affected patients and their families but also for national economies: In the United States, it is estimated that the annual costs caused by strokes ranges between 30 and 40 billion dollars. In the United Kingdom, the cost of

one individual patient amounts to £30 000 over 5 years.

The term stroke accommodates a variety of different conditions. About 85% of all strokes are caused by cerebral ischemia due to vessel occlusion. Primary cerebral bleeding is rare in comparison – 15%. Of the ischemic strokes, 75% are caused by emboli, of either arterial or cardial origin, while microvascular occlusion, that is, hyalinosis or *in situ* thrombosis is responsible for 20% of cases. Hemodynamic ischemia, caused by stenoses of brain-supplying arteries, account for less than 5% of ischemic strokes.

The prospective risk of suffering an ischemic stroke is partially a function of social and behavioral (nutrition, tobacco use, stress) factors and long-standing disorders like hypertension, diabetes, disorders of cholesterol and lipid metabolism, and obesity. Atherosclerosis is not only the main underlying condition in ischemic stroke but also in coronary heart disease and peripheral vascular disease. Moreover, the genetic background of affected individuals is coming under increasing scientific scrutiny, as might be expected for familial stroke conditions like CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leucencephalopathy), which is characterized by relapsing subcortical ischemia. The condition is caused by a mutation in the *notch3* gene on chromosome 19. MELAS (mitochondrial encephalopathy, lactate acidosis, stroke-like episodes) is another rare familial stroke disorder, characterized by migraine, grand-mal seizures, and recurrent cortical infarctions. The condition is caused by a range of mutations in the mitochondrial genome, which explains its maternal inheritance. The severity of the phenotype depends mainly on the mutation locus and on the proportion of mutated mitochondrial genomes in the cerebral mosaic. However, family and twin studies suggest that genetic factors are likely to be involved in common types of ischemic stroke, not just in rare and well-defined familial stroke disorders.

A locus on chromosome 5q12 described for the Icelandic population seemed to increase stroke susceptibility significantly. The responsible gene has been identified as phosphodiesterase 4D (pde4d). The risk-conveying polymorphisms are located in the gene-regulatory part. This hints at a defect in the regulation of the encoded cAMP degrading protein. Phosphodiesterase 4D belongs to a group of proteins that are drug targets in the treatment of asthma, erectile dysfunction, and inflammation.

Symptoms of focal cerebral ischemia depend on the individual affected vessel with its typical supply territory. Table 1 summarizes the frequency of affection and the typical clinical syndromes for the main brain arteries. Stroke mortality is rated between 20 and 30%. Unfavorable prognostic factors are old age, initial coma, papillary asymmetry, coronary heart disease, and heart failure. Pyrexia and infections (pneumonia, in particular) deteriorate the prognosis, particularly during the initial phase (see below). Of the surviving patients, one-third improve within a week, 40% remain unchanged in their disability, and 20% deteriorate further during the first week.

A great deal of our knowledge about the pathophysiology of stroke comes from experimental research. The experimental animal models follow two main paradigms. One is the model of focal cerebral ischemia as a model of ischemic stroke. The other is the model of global cerebral ischemia as model of circulatory arrest. For obvious

Tab. 1 Territorial stroke syndromes.

Anterior territory		
Middle cerebral artery (MCA)	(≈60%)	Hemiparesis mainly arm, hemihypaesthesia, hemianopia, dysphasia or neglect
Anterior cerebral artery (ACA)	(4%)	Hemiparesis mainly leg, urinary incontinence, apraxia
Anterior choroidal artery (AchA)	(8%)	Hemiparesis, hemihypaesthesia, hemianopia
Posterior territory		
Vertebral/basilar artery (VA/BA)	(10%)	Vertigo, diplopia, bilateral hypaesthesia, and paresis, crossed syndromes, amaurosis, ataxia, headache, coma
Cerebellum posterior/anterior inferior cerebellar artery (PICA/AICA)	(7%)	Headache, ataxia, vertigo, gaze palsy, facial weakness, deafness
Posterior cerebral artery (PCA)	(9%)	Hemianopia, dyslexia, visual agnosia

reasons, we will focus mainly at studies of focal cerebral ischemia. The bulk of experimental studies has been carried out with rats and mice, although some primates have been used. In rodents, ischemia is mostly induced by intravascular occlusion of the MCA, using a monofilament just exceeding the critical vessel diameter. Depending on the duration of occlusion, we distinguish permanent from transient ischemia models. The latter are also used as models for spontaneous reperfusion or for the state after successful lysis therapy with recombinant tissue plasminogen activator (rtPA) in humans, respectively.

In models of cerebral ischemia, perfusion to the brain is impeded by some form of manipulation. Since the essence of ischemic stroke is vascular occlusion, it is not surprising that in these models the extent of the eventual infarction can be predicted quite accurately from the degree of reduction in regional cerebral blood flow (rCBF). Thus, if the rCBF is reduced to less than 25% of normal, the likelihood of infarction in a given volume of brain tissue is greater than 95%. In contrast, the likelihood of infarction is less than 5% if rCBF does not fall below 50% of normal. These thresholds have been established in comparative human PET and MRI studies and correspond with experimental data from animal models. Thus, the initial reduced rCBF determines the extent of the anticipated infarction. This, however, holds only with the provision that a spontaneous or therapeutic reperfusion does not take place.

We now look at ischemic infarction as the result of a complex and prolonged process of infarct maturation rather than a simple result of reduced regional perfusion. As brain tissue has a high demand for oxygen and glucose, a disruption of perfusion leads to substrate depletion within few minutes, while toxic metabolites accumulate. The ensuing cellular energy deficit leads to a collapse of the established ion gradients and the membrane potential. Neurons and glial cells depolarize. Depending on the extent and the duration of this energy deficiency, the cells will suffer not only a functional but also a structural breakdown. The highly complex sequence of events within the ischemic area follows a fairly well-defined

stereotypic spatiotemporal pattern, which we will discuss below in more detail. The concept of the ischemic *penumbra* is crucial to the understanding of these mechanisms. The cascade of ischemic damage begins with *excitotoxicity*, the formation of reactive *oxygen free radicals*, the increasing *tissue acidosis* and the occurrence of *periinfarct-depolarizations*. It is followed by the stages of *inflammation* and programmed cell death (*apoptosis*). This is associated with *DNA damage* that, in turn, induces *DNA repair programs*. Although the process is not yet fully understood, we know that chromatin remodeling, that is, *epigenetic mechanisms*, and the activation of transcription factors induce complex *gene programs*. These changes initiate the expression of destructive proteins involved in inflammation and apoptosis, as well as a host of protective genes that help repair the ischemic damage. It is the activation of these protective genes that builds up *ischemic tolerance*. Among the many newly discovered protective mechanisms, *endogenous* and *exogenous cell replacements* have met with a lot of interest. Apart from these autochthonous mechanisms of the brain tissue, there are other mechanisms on a systemic scale, which bear important clinical significance. For instance, the phenomenon of *stroke-induced immunodepression* can help understand why stroke patients are at such high risk of contracting serious bacterial infections. Neuroprotective treatment must be based on an understanding of these mechanisms.

2 The Penumbra Concept

In the ischemic brain, we commonly distinguish two tissue volumes – the core of the infarction and the surrounding zone, known as *ischemic penumbra* – the underperfused and metabolically compromised margin surrounding the irrevocably damaged core. Core and penumbra are characterized by two different kinds of cell death: necrosis and apoptosis (which is also called programmed cell death or delayed neuronal cell death). The severe perfusion deficit in the core causes a breakdown of metabolic processes, cellular energy supply, and ion homeostasis, which causes the cells to lose their integrity within minutes. Thus, acute necrosis of cell and tissue prevails in the core. In the penumbra, some residual perfusion is maintained by collateral vessels, which may be unable to maintain the full functional metabolism, but prevents immediate structural disintegration. However, over time, the alteration of cellular homeostasis causes more and more cells to die, and the volume of the infarction increases (Fig. 1). The penumbra has thus to be considered as tissue at risk during the maturation of the infarct. In this region, apoptosis and inflammatory signaling cascades play an important role. It may initially constitute 50% of the volume that will end up as infarction. The mechanisms that lead to delayed cell death within the penumbra are the subject of intense research, as they provide targets for a specific neuroprotective therapy in brain regions challenged by ischemia, but which are still viable.

3 Excitotoxicity

The depolarization of neurons and glia due to local energy deficit causes the activation of voltage-gated calcium channels and the release of excitatory amino

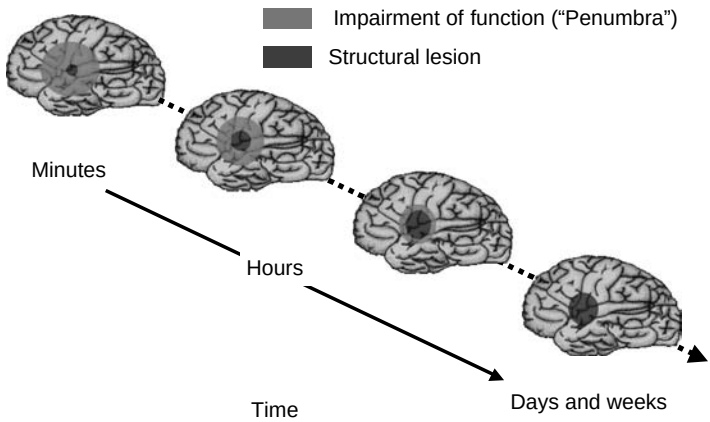

Fig. 1 Regression of the functional neurological deficit while the structural lesion grows. Early in the course of stroke, clinical symptoms mostly reflect an impairment of function (green) but not necessarily a structural lesion (blue). Over time, some areas either spontaneously, or because of therapy, recover function, which explains why symptoms in patients can regress, while the structural lesion actually grows (Reprinted from Dirnagl, U., Iadecola, C., Moskowitz, M.A. (1999) Pathobiology of ischaemic stroke: an integrated view, *Trends Neurosci.* **22**, 391–397, with permission from Elsevier) (See color plate p. xxiii).

acids into the extracellular space. Glutamate, in particular, which, under normal cellular energy conditions, would be immediately taken up presynaptically or through astrocytes, now remains in the extracellular space where it accumulates dramatically. Through the activation of NMDA and AMPA subtype glutamate receptors, the intracellular Ca^{2+} level rises (Fig. 2). Furthermore, metabotropic glutamate receptors are activated through the induction of phospholipase C (PLC) and inositol triphosphate (IP_3), and calcium is mobilized from intracellular stores. Additionally, the AMPA receptor induces an increase of intracellular Na^+ and Cl levels.

Altogether, the result is a massive disturbance of ion homeostasis, accompanied by passive water influx and cell edema. Ultimately, these massive changes in cell volume account for osmotic cell lysis. This lytic type of cell death, also referred to as *necrosis*, is primarily observed in the core of the infarction. Cells that escape this most dramatic form of disintegration, as they can not in the core but in the penumbra, excitotoxicity may be an initiator of molecular events that lead to apoptosis and inflammation.

4 Oxygen Free Radicals

As a consequence of ischemia and particularly of reperfusion, reactive oxygen free radicals such as superoxide, hydrogen peroxide, and hydroxyl radicals are generated. Nitric oxide is generated via the activation of the calcium calmodulin–dependent nitric oxide synthase (NOS); it reacts with superoxide radicals and forms thus the highly reactive peroxynitrite radical.

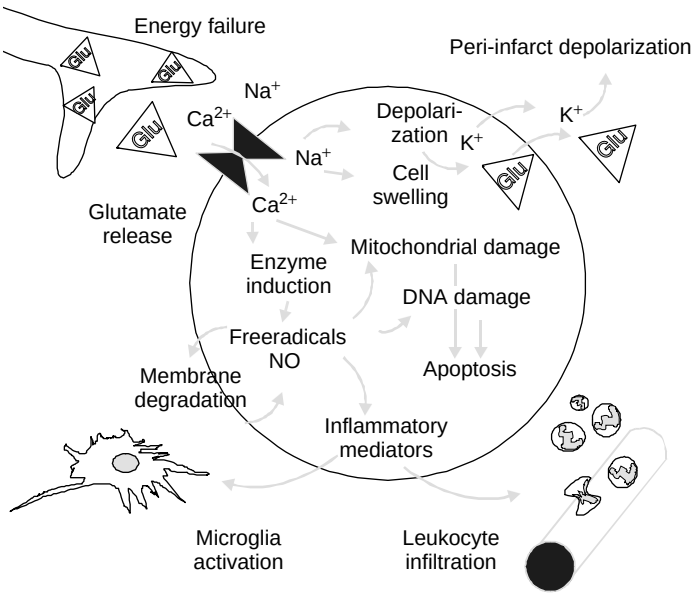

Fig. 2 Simplified overview of pathophysiological mechanisms in the focally ischemic brain. Energy failure leads to the depolarization of neurons. Activation of specific glutamate receptors dramatically increases intracellular Ca^{2+}, Na^{+}, Cl^{-} levels while K^{+} is released into the extracellular space. Diffusion of glutamate (Glu) and K^{+} in the extracellular space can propagate a series of spreading waves of depolarization (peri-infarct depolarizations). Water shifts to the intracellular space via osmotic gradients and cells swell (oedema). The universal intracellular messenger Ca^{2+} overactivates numerous enzyme systems (proteases, lipases, endonucleases, etc.). Free radicals are generated, which damage membranes (lipolysis), mitochondria, and DNA, in turn triggering caspase-mediated cell death (apoptosis). Free radicals also induce the formation of inflammatory mediators, which activate microglia and lead to the invasion of blood-borne inflammatory cells (leukocyte infiltration) via upregulation of endothelial adhesion molecules (Reprinted from Dirnagl, U., Iadecola, C., Moskowitz, M.A. (1999) Pathobiology of ischaemic stroke: an integrated view, *Trends Neurosci.* **22**, 391–397, with permission from Elsevier).

Further sources of oxygen free radicals in the damaged brain tissue are the breakdown products of the adenosine phosphates, which contribute to radical production via xanthine oxidase and the iron-catalyzed Haber–Weiss reaction. The many different radical species that are thus formed can react with virtually any cellular components (carbohydrates, amino acids, DNA, phospholipids) and damage them. The peroxidation of membrane lipids releases further radicals – and further glutamate (Fig. 2). Oxygen free radicals gain even more significance when new oxygen reaches the damaged tissue by virtue of reperfusion, or in the penumbra where oxygen supply has not ceased entirely.

Hypoxia itself as well as the elevated intracellular concentration of calcium ions and free radicals disrupt the function

of neuronal mitochondria. Consequently, a so-called mitochondrial permeability transition pore (MPT) in the mitochondrial membrane may form. Besides impeding ATP production through loss of mitochondrial potential, the MPT leads to mitochondrial swelling, a burst of free oxygen radicals, and the release of proapoptotic molecules. Thus, a vicious cycle of further disintegration is fueled. This vicious cycle is counterbalanced in part by antioxidative enzymes like the manganese superoxide dismutase (Mn-SOD) and the cytosolic forms of the copper–zinc superoxide dismutase (CuZn-SOD). These may prevent the breakdown of the mitochondrial membrane and, thus, the release of cytochrome C, which would be the trigger of apoptosis.

5
Tissue Acidosis

In the context of stroke pathophysiology, the proton balance is intimately linked with the glucose metabolism. With reduced oxygen availability, anaerobic glycolysis, as the only remaining source of ATP production, leads to tissue acidosis. It has long been assumed that this acidosis was one of the main noxious mechanisms in ischemic stroke. This so-called "lactate-acidosis hypothesis" is often quoted as explanation for the "glucose paradox" of cerebral ischemia. This paradox refers to the observation that excessive supply of glucose, the most important source of energy of the brain, during focal cerebral ischemia does not reduce tissue damage as one should think but, instead, augments it. However, by which mechanism this happens and, in fact, whether levels of acidosis reached in brain ischemia can generate brain tissue damage at all is still far from being clear. Possibly, the pH-dependent transition of Fe(III) to Fe(II) and the release of iron from molecular storages lead to a facilitation of the Haber–Weiss reaction that forms toxic free oxygen radical species (see above). Besides the production of different species of oxygen free radicals, acidosis also interferes with intracellular protein synthesis. However, the lactate-acidosis hypothesis is not unchallenged. Particularly the fact that acidosis blocks the NMDA receptor and thus has an antiexcitotoxic effect indicates the complexity of the role that acidosis plays in cerebral ischemia.

A similarly hotly debated topic are the findings on hyperglycemia during stroke. Experimental data from animal models show a detrimental effect of hyperglycemia during focal cerebral ischemia. Clinical data also suggest that hyperglycemia during the acute phase of stroke worsens prognosis. Persisting hyperglycemia beyond the acute phase is also an independent prognostic factor for larger infarct volume and poorer functional outcome in stroke patients. It is a matter of debate whether these observations are due to a causal relationship or if hyperglycemia is a mere epiphenomenon, possibly due to a stress reaction (i.e. an effect of concomitant catecholamine and glucocorticoid release proportional to initial tissue damage). The controversial role of an associated lactate acidosis has already been discussed. An alternative concept is the glucocorticoid hypothesis. It suggests that hyperglycemia enhances the release of glucocorticoids, which are also released during ischemia in the sense of a stress reaction. Glucocorticoids have been shown to have a direct cytotoxic effect and the blockade of glucocorticoid receptors does reduce the damaging effect of hyperglycemia. On the other hand, hyperglycemia and acidosis

have a protective effect *in vitro*, and some experimental studies demonstrate that hyperglycemia can also have a beneficial effect under special circumstances.

However, a recent study on humans seems to shed light on this controversy. Perfusion and diffusion MR imaging and MR spectroscopy shows that stroke patients with hyperglycemia have a higher lactate accumulation in the penumbra. While the perfusion–diffusion mismatch (as a measure of the size relation of penumbra and core) was similar in the normoglycemic and hyperglycemic groups, the eventual infarct size and disability scores were significantly higher in the hyperglycemic group. While it could still be argued that initially hyperglycemic patients were also likely to be long-standing or latent diabetics with an overall worse microvascular state, we take this study – which links in with much of the experimental animal data – as a strong indicator for the vulnerability of the penumbra to hyperglycemia. Indirectly, it also corroborates the lactacidosis theory, which, however, still lacks direct proof.

6
Peri-infarct Depolarizations

The anoxic as well as excitotoxic depolarization of neurons and glia increases the concentration of glutamate and potassium in the ischemic core. These diffuse into the penumbral zone where they induce the electrochemical depolarization of more neuronal and glial cells that are "just managing" under the circumstances of oligemia. Depolarization in this context is not to be confused with short-lived action potentials, but means sustained depolarization of the entire cell membrane, a breakdown of the membrane potential that has to be actively rectified. Depolarization also causes a further significant increase in extracellular glutamate and potassium (Fig. 2). Thus, waves of sustained electrochemical depolarization spread from the core of infarction into the penumbra, bearing similarity with Leao's spreading depression. These waves are called *peri-infarct depolarizations*. Peri-infarct depolarizations occur with a frequency of 1–4/h and have been detected by functional MRI and near infrared spectroscopy. Since active rectification of the membrane potential requires energy, these peri-infarct depolarizations contribute further to the metabolic compromise and each wave turns a larger area of the penumbra toward irrevocable infarction. In addition, peri-infarct depolarizations seem to compromise the impaired microcirculation within the penumbra even further. Peri-infarct depolarizations have been detected in patients with head trauma.

7
Inflammation

The early stage of inflammation, which starts a few hours after the onset of ischemia, is characterized by the expression of adhesion molecules in the vascular endothelium as well as on circulating leukocytes. Thus, leukocytes adhere to the endothelium and transmigrate from the blood into the brain parenchyma, which is of decisive importance for stroke-induced brain inflammation (Fig. 2).

Newly expressed endothelial adhesion molecules like ICAM-1 and VCAM-1 facilitate the interaction between endothelium and leucocytes as first step of transmigration of white blood cells from blood into the tissue. They interact with the β_2-integrins CD11b/CD18 (Mac-1) and CD11a/CD18

(LFA-1) that are, in turn, expressed on leucocytes. Neutrophil granulocytes accumulate in the capillaries of the penumbra and interfere with the already impaired microcirculation. Blocking this interaction with CD18, CD11, or ICAM-1 antibodies reduces not only the number of leucocytes but also the size of infarction. ICAM-1 knockouts have also smaller infarcts than wild-type mice. Next to the β_2-integrins, other integrins that are more specific for lymphocytes like VLA-4($\alpha_4\beta_1$) come into action, too. Consequently, blocking the α_4-component reduces also the volume of infarction.

A major part in inflammation is ascribed to the population of microglial cells. These cells are the primary immunoeffectors of the CNS and make up about 20% of the entire glial mass, about the same proportion as neurons. Activation of microglia is a striking feature in the penumbra. The cells change their shape from highly ramified cells to a more plump and amoeba-like form, and they proliferate. They also express a number of characteristic surface molecules (for instance, MHCI and MHCII) and are capable of antigen presentation. Similar to leukocytes, activated microglia cells are able to produce not only a variety of proinflammatory cytokines but also cytotoxic metabolites, especially oxygen free radicals (such as peroxynitrite and superoxide) and enzymes (for instance, cathepsins). In addition, microglia is phagocytically active. The inhibition of microglial activation turns out to be protective in experimental stroke models. On the other hand, activated microglia have also been shown to release anti-inflammatory cytokines (such as TGF-β1) and growth factors. Because of the Janus-faced nature of microglial products (destructive, for example, free radicals vs protective, for example, growth factors), the overall role of microglia in cerebral ischemia is not clear at present. It is very likely that microglia play different roles at different times, with protective or regenerative activities occurring days or even weeks after the onset of ischemia.

Activated leukocytes (granulocytes, monocytes/macrophages, lymphocytes) as well as neurons and glial cells (astrocytes, microglia) produce cytokines and chemokines. In particular, proinflammatory cytokines like TNFα, IL-1, and IL-6 play a major role as mediators of the inflammatory response. They are responsible for the transition from the early, excitotoxic phase to the inflammation phase. Their regulation depends on transcription factors like NFκB, which, in turn, are activated by oxygen free radicals. Cytokine-mRNA and also the protein of TNFα and IL-1 can be detected only few hours after induction of an experimental focal ischemia. Expression of IL-6 follows only after about 24 h. These cytokines are mainly released by microglial cells and macrophages. In stroke patients, it could be shown that the intrathecal concentration of IL-1 and IL-6 correlated with the infarct size. Cytokine receptor antagonists reduce the infarct volume in animal models. For example, blockade of TNFα, by TNF-binding proteins reduces brain injury after focal cerebral ischemia. The role of TNFα is not entirely clear, however, as mice lacking the TNFα-receptor (TNFR2) have larger infarctions. In part, these contrasting results may reflect a difference in signal transduction cascades activated by the two TNFα-receptors, TNFR1 and TNFR2. It is also arguable that an increased IL-6 induction may be responsible, since IL-6 knockouts do not have smaller infarctions than their litter mates. Further to the

induction of adhesion molecules, the above-mentioned cytokines also render the blood–brain barrier more permeable and induce prothrombotic functions of the endothelium.

Not only can the proinflammatory cytokines be observed, but anti-inflammatory cytokines such as TGF-1β or IL-10 can also be induced. By downregulating the inflammation, these cytokines have a protective effect in the context of cerebral ischemia. The neuroprotective action of TGF-β1 is well known. Interestingly, this cytokine seems to play an important part in immunological tolerance. As in ischemic tolerance (see below), it is possible to induce a tolerant state toward ischemia. This protection is induced by immunogenetic proteins (for instance, myelin basic protein, MBP) and mediated by a certain lymphocyte population expressing TGF-β1. In contrast to ischemic tolerance, immunological tolerance is longer lasting. Protection lasts for months, whereas in ischemic tolerance it lasts no longer than 7 days.

Apart from cytokines, there are two other proteins that dominate the inflammatory phase: the inducible nitric oxide synthase (iNOS) and cyclooxygenase 2 (COX2). Focal cerebral ischemia models have shown that iNOS is induced on the mRNA as well as on the protein level. Protein expression reaches its height after 24 h. By inhibiting the expression of iNOS, the size of experimental infarcts can be reduced by about 30%, even if treatment is undertaken only 24 h after the onset of ischemia. Nitric oxide (NO), which is produced in much greater amounts by iNOS than by the constitutional NO synthases, acts via the formation of peroxynitrite highly cytotoxic. NO triggers apoptosis not only through peroxynitrite but also via the induction of p53, which probably causes direct DNA damage. Cox-2 is mainly generated in the penumbra. The enzyme has a destructive effect on the penumbral tissue, mainly through producing oxygen free radicals and toxic prostanoids. This seems a likely hypothesis, as both genetic and pharmacological inhibition of Cox-2 have a protective effect. Cox-2 as well as iNOS are promising targets for the treatment of stroke patients because blocking them is still effective even 6 to 24 h after ischemia.

Not only the activated residual microglia but also monocytes from the circulating blood migrate into the affected brain tissue. This transmigration is mediated by a chemokine, the monocyte-attractant protein 1 (MCP-1), which is induced 1 to 2 days after ischemia and causes immigration of monocytes. Moreover, about a third of these transmigrated cells differentiate within 14 days post stroke into microglial cells – virtually indistinguishable from autochthonous microglia. A transdifferentiation into astrocytes has also been observed, but our group has not been able to reproduce this finding. Irrespective of that, the important role of astrocytes during inflammation has been recognized, producing both proinflammatory as well as neuroprotective factors such as erythropoietin (EPO), TGF-β_1, or metallothionein 2.

In general, much further study needs to be conducted before the complex role of inflammation in cerebral ischemia is sufficiently understood to be able to specifically block the damaging proteins and foster protective elements during the inflammation process in stroke.

8
Damage to the Blood–Brain barrier

The function of the blood–brain barrier (BBB) depends on the integrity of

its cellular matrix, which consists of endothelial cells, basal lamina, and astrocytes. Cerebral ischemia causes damage to this matrix and its intercellular signal exchange. Central to these damaging processes are proteases like cathepsins, plasminogen activators, and the matrix metalloproteinases (MMP). The action of the MMPs, a group of more than 20 zinc endopeptidases, has been extensively studied. Proteolytic destruction of the basal membrane by the MMPs permits the immigration of leukocytes and promotes vasogenic edema. MMP-2 and MMP-9 are expressed within 1 to 3 h after cerebral ischemia; the level of their expression correlates with the level of BBB breakdown, the risk of hemorrhagic transformation, and the extent of eventual neuronal damage. Pharmacological inhibition of the MMPs prevents this breakdown and reduces stroke volumes. Mice with target disruption of MMP-9 are partly protected from the effects of ischemia.

Of the plasminogen activators, tissue plasminogen activator (tPA) is best known. In the brain, tPA is expressed mainly in neurons, astrocytes, and microglia. Studies of mainly primary cortex neurons have shown that tPA acts also as neurotoxin, and experiments on tPA knockouts have produced conflicting data. Since recombinant tPA is the sole licensed drug for thrombolytic therapy, these findings reflect also on current therapeutic practice. tPA has furthermore been shown to induce MMP-9 and, thus, – at least indirectly – damaging the BBB. The same holds probably for the *recombinant* tPA. This may explain why such a high number of secondary hemorrhages have been observed after thrombolytic stroke therapy. Animal experiments showed that delayed rtPA treatment – mimicking real-life delays in treatment – lead to significantly greater damage to the BBB and a major increase in mortality. The pharmacological inhibition of MMPs can reverse this effect of rtPA. In addition, tPA enhances NMDA-mediated calcium signaling and subsequent excitotoxic neuronal injury. It is a matter of controversy whether this detrimental effect is due to a cleavage of the NR1 subunit of the NMDA receptor by tPA.

The example of rtPA demonstrates once again that in order to refine treatment of ischemia, a thorough understanding of the complex mechanisms is indispensable. The development of effective and low-risk thrombolytic treatment is, of course, still one of the main objectives of pharmacological stroke research. One promising approach is the development of novel plasminogen activators, such as recombinant *desmodus rotundus* salivary plasminogen activator (DSPA-α1; Desmoteplase). Experimental evidence indicates that Desmoteplase has pharmacological and toxicological properties superior to rtPA.

9
Programmed Cell Death and Apoptosis

Apoptosis follows a similar time schedule as inflammation. It sets in after a few hours and continues over days. It is the commonest form of cell damage in the penumbra, and is very distinct from the rapid neuronal and glial death in the adjacent ischemic core, where pannecrosis is complete within few hours. In necrosis, it is the cellular edema that leads to osmolysis (see above). The released cell contents are a potent inflammatory stimulus, as discussed above.

Caspases, a group of proteases, has a pivotal role in apoptosis. They facilitate the

"orderly" destruction of the cell, which is characterized by a condensation of both nucleus and the cell as a whole. The cell contents are fragmented into so-called apoptotic bodies. Endonucleases fragment the cell's DNA into typical strands of 180 base pair length. Other morphologic criteria of apoptosis are membrane blebbing and shrunk cytoplasm with morphologically intact organelles. While the terms "delayed neuronal death" (DND) and "programmed cell death" (PCD) are often used synonymously with "apoptosis," the two former may well occur without the typical features of the latter. Although most experimental studies show some biochemical evidence for apoptosis to have taken place, histological features are often missing to substantiate the claim. This has led some researchers to question apoptosis as an actual feature of ischemia. A number of reviews have dealt extensively with the significance of apoptosis in the pathophysiology of ischemia. This is why we shall focus here on some of its mechanistic aspects in this context.

Essentially, apoptosis can be triggered in two ways, by intrinsic (i.e. basically mitochondrial) or extrinsic activation. Binding to the Fas-receptor, for example, by TNFα, triggers the extrinsic pathway. Intrinsic triggers of apoptosis in the context of ischemia are elevated concentrations of intracellular calcium, reactive oxygen species, and glutamate as well as increased DNA damage (Fig. 2). Through damage to the mitochondrial membranes, both pathways directly or indirectly lead to the activation of caspases.

The caspases are a hierarchic group of at least 14 cysteine-dependent and aspartate-specific proteases. Their inactive precursor proteins exist in every cell and can be activated through cleavage. Apoptosis is essentially an active and in itself energy-dependent cellular deconstruction process. In cerebral ischemia, the caspases 1, 3, 8, and 9 are involved, of which, the caspases 8 and 9 are the initiators of a signal cascade that activates the so-called execution caspase 3. Caspase 1 is predominantly engaged in cytokine activation. Caspase 3 is of predominant importance not only in the context of cerebral ischemia, and its genetic or pharmacological inhibition is not only in this context neuroprotective. The substrates degraded by caspase 3 are DNA repair enzymes like poly-(ADP-ribose)-polymerase (PARP) and the DNA dependent protein kinase (DNA-PK). The latter is important for the repair of double strand breaks. PARP recognizes single-strand disruptions and mediates self-repair of the DNA through swift automodification with synthesis of highly negatively charged, long, and branched ADP-ribose molecules. This process utilizes ATP and thus taps the cellular energy pool. Although not universally accepted, this ATP consumption by PARP is widely believed to be in itself cytotoxic. PARP is also capable of inducing caspase-independent cell suicide via the so-called *apoptosis-inducing factor* (AIF; see below). This also helps in explaining the partial failure of caspase inhibitors in some models of cerebral ischemia. The genetic or pharmacological blockade of PARP has, however, been demonstrated to be neuroprotective in experimental stroke. Of significance is the cleavage by caspase 3 of the inhibitor of caspase-activated DNase (ICAD). ICAD is thus activated and causes the characteristic feature of "laddering" that has been described in all stroke models. The term laddering draws on the regular appearance of those DNA fragments as ladderlike bands on a gel electrophoretic separation. The length of the resulting DNA fragments

reflects the complex structure of eukaryotic genetic information in the chromatin. Chromatin is a structure, in which most of the DNA is enclosed by specific proteins, the histones (see epigenetics), leaving only few positions accessible for endonucleases. Complete apoptotic laddering leaves DNA-oligomers of about 140 to 180 base pairs in length, which do no longer contain any useful genetic information. Apart from destroying genetic information, caspases also degrade structure proteins of the nucleus and cell, such as laminin, actin, and gelsolin.

Cytochrome C is not only an essential part of the mitochondrial electron transfer chain and hence for the energy production in the cell but also a crucial mediator in the caspase activation cascade. On the extrinsic pathway of apoptosis, stimulation of the Fas-receptor leads to the formation of the "death-inducing signaling complex" (DISC). DISC activates caspase 8 from its precursor protein, which, in turn, activates Bid, a proapoptotic protein of the Bcl-2 family (see below). Bid then induces the liberation of cytochrome C from the mitochondrial membrane. Other members of the proapoptotic Bcl-2 family are Bad and Bax, important on the intrinsic pathway. Their activation leads to their translocation into the outer mitochondrial membrane and to the establishment of the mitochondrial permeability transition pore (MPTP), leading to the release of cytochrome C. Another way of its release from mitochondria is obviously direct membrane damage by free radicals. Cytochrome C, together with the cytosolic protein Apaf-1, forms the caspase precursor protein, and, using ATP, the so-called apoptosome. The apoptosome activates caspase 9 through cleavage from its precursor. Whether cytochrome C is eventually released depends, however, on the balance of proapoptotic members of the Bcl-2 family like Bid, Bax, Bad, and Bag, and their antiapoptotic counterparts Bcl-2 and Bcl-x_L. Prevailing expression of Bcl-2 and Bcl-x_L results in clearly smaller infarction after focal cerebral ischemia. Treatment with a Bcl-x_L fusion protein containing the TAT domain of the human immunodeficiency virus (HIV) for cell penetration acts also neuroprotectively in models of ischemia. Bcl-2 and Bcl-x_L seem to have a stabilizing effect on the mitochondrial membrane; they prevent the formation of MPTP, the ensuing cytochrome C liberation, and the formation of AIF. Not only the reinforcement of antiapoptotic factors but also the attenuation of proapoptotic proteins may be neuroprotective in cerebral ischemia. For instance, the phosphorylation of Bad leads to its sequestration by the protein 14-3-3 and thereby to its inactivation. The PI3K/Akt-pathway, utilized by many neurotrophic factors such as erythropoietin (see below: endogenous neuroprotection), plays an essential role in this. It has been reported that Akt has been activated during cerebral ischemia. Whether this has a functional impact in stroke remains unclear at the moment.

Other important protein groups involved in stroke are the members of the IAP-group (inhibitors of apoptosis proteins). The overexpression of the X-chromosome-linked IAP (XIAP) has been shown to be protective in stroke models. IAPs act either indirectly by inhibiting the apoptosome or directly inhibiting caspase 3, but can themselves be blocked by other proteins such as DIABLO (in mice, the human homologue is Smac) that is physiologically localized in mitochondria. DIABLO is released after focal ischemia and binds to XIAP. It is yet unclear what its function is during focal ischemia. The data discussed show, however, that apoptosis underlies

a complex regulation process that is controlled at many levels.

10
Ischemia-induced DNA Damage, DNA Repair, and p53 as Genotoxic Sensor

The emergence of DNA damage and its repair functions in the context of focal cerebral ischemia are particularly interesting. DNA damage is caused by oxygen free radicals (see above). Particularly, the interaction of hydroxyl radicals with DNA results in strand breaks and base alterations. Hydroxyl radicals have a very short half-life – with the exception of peroxynitrite, formed from NO and superoxide – and are unlikely to reach the nuclear DNA from outside the cell. They are more likely to be produced in close proximity to the nucleus as a homolytic cleavage product of peroxynitrite, which has the ability to diffuse over longer distances.

Under physiological circumstances, DNA lesions are generated with a frequency of about 10 000 events per cell each day. They are recognized and repaired efficiently, provided the repair enzymes are intact and not overloaded. In the wake of ischemia and reperfusion, the tissue is flooded with oxygen free radicals, causing abundant DNA damage. Alterations that are typically caused by free radicals have been found in the mouse model of transient focal ischemia, where a massive increase in DNA damage was observed 10 to 20 min after reperfusion. Foremost, 8-hydroxy-deoxy guanosine becomes much more prevalent in both nuclear as well as mitochondrial DNA after ischemia. These alterations are repaired partially within 4 to 6 h, and the increase in repair activity is proportional to the degree of DNA damage.

There are three basic DNA repair mechanisms – base excision repair (BER), nucleotide excision repair (NER), and what is known as mismatch repair. BER seems to be of particular significance in ischemia. During BER, DNA glycosylases excise damaged or modified bases. The thus generated purine and pyrimidine-free sites (AP-sites) are eliminated from the damaged DNA strand by AP-endonucleases (apurinic/apyrimidinic endonucleases) and replaced by DNA polymerase β and DNA ligase with a short strand of 1 to 4 nucleotides based on the complementary blueprint of the remaining undamaged strand.

Apart from oxidative DNA damage, deamination of cytosine may produce uracil. In BER, such alterations are sorted out by uracil-DNA-glycosylase (UDG). We found that knockout mice deficient in UDG have substantially larger infarctions than the wild-type controls.

DNA repair is neither inexhaustible nor infallible. Persisting mutations are the result of erroneous DNA repair. In models of transient focal ischemia, a dramatic increase in the frequency of mutations has been observed. Half of them were transitions and transversions. The other half were deletions and occasional insertions, mostly resulting in a frame shift (frame shift mutations) of the open reading frame (ORF). It seems that not only the exhaustion of the repair capacity but also an impaired proofreading ability in the exonuclease-negative DNA polymerase β expressed in the brain is to blame for these mutations. The major role of DNA repair is to ensure transcription of mRNA from intact genes. An obvious consequence of mutations is instability of the genetic information that may lead to faulty proteins, eventually resulting in

either cellular malfunction, disintegration, or spawning of tumors.

The maintenance of swift and effective DNA repair is therefore essential in postmitotic cells like neurons. In order to adapt to an increased demand, a cell may try to increase its repair efficiency. Apoptosis is another way of avoiding erroneous protein synthesis and the potential emergence of tumors (see above). Both mechanisms are found in focal cerebral ischemia, and it seems to be the extent of the damage that determines which scenario will prevail.

We have several reasons to believe that the tumor suppressor protein p53 represents the molecular switch on the crossroads of attempted repair and apoptosis. (1) p53 is mainly induced by DNA damage; (2) p53 can halt the cell cycle, and thus prevent the division of non-neuronal cells before complete repair; (3) p53 can induce apoptosis if the damage is extensive. In most models of ischemia, it could be shown that p53 had been induced. The fact that p53 knockouts have significantly lower infarct volumes accords with p53's proapoptotic function. However, it seems paradoxical that heterozygous $p53^{+/-}$ mice show even smaller infarctions. An attenuated expression of p53 seems to give better neuroprotection than a total lack of it, which hints at the dual function of the gene. It would be interesting to find out if longer-term neoplasias could be the downside of this short-term effect.

11 Epigenetic Mechanisms

Experimental work of the last years has shown that, apart from DNA repair, epigenetic mechanisms like DNA methylation and histone modification also play an important role in ischemic stroke. The transcriptional gene expression is not only regulated by transcription factors but also depends on epigenetic mechanisms that mainly modify the chromatin structure.

In the mammalian genome, DNA methylation is a covalent, postreplicative modification at the carbon 5 position of the pyrimidine ring of cytosine bases (m^5C); it typically occurs in CpG dinucleotides. DNA methylation is catalyzed by DNA methyltransferases (DNMT1, DNMT3a, DNMT3b). The main function of DNA methylation lies in the regulation of gene transcription. Methylated genes are barred from expression in a process called "gene silencing." First of all, many transcription activators are unable to bind to methylated loci at all. Furthermore, there is a group of methyl-CpG-binding domain proteins (MBD) that form a complex with histone deacetylases (HDAC), which, in turn, by deacetylation of histones, cause the chromatin structure to condense. Through this condensation, the gene-regulatory loci become inaccessible for transcription factors. Five proteins of the MBD family are known to date: MBD1 through 4 and MeCP2. HDACs deacetylate predominantly histones H3 and H4. Conversely, acetylated histones open up the chromatin structure, permitting the gene-regulatory regions (promoters, enhancers) to become active again, and, ultimately, enabling gene transcription through binding of transcriptional activator proteins. While the state of deacetylation or acetylation of a CpG hypomethylated promoter or enhancer region is comparatively volatile, DNA hypermethylation of these regulatory regions results in a more stable repression.

Initially, DNA methylation was thought to be of biological significance only for fundamental phenomena of cell biology like embryogenesis and differentiation,

in X-chromosome inactivation, genomic imprinting, and in several neoplastic diseases. In more recent years, its pathophysiological role in a number of further conditions has been recognized, such as Fragile-X syndrome, ICF (*I*mmunodeficiency, *C*entromere instability, *F*acial anomalies) syndrome, or Rett syndrome.

For the first time, we were able to detect the involvement of DNA methyltransferase 1 and DNA methylation in an acute disease, such as ischemic stroke. In addition, a complex pattern of an altered expression of MBD-family proteins has been described in a model of transient cerebral ischemia. The inhibition of DNA methylation had a protective effect, as had manipulations to the acetylation state of histones. Pharmacological inhibition of a histone acetyltransferase with trichostatin A (TSA) caused enhanced histone acetylation *in vitro* as well as *in vivo*, and proved to be neuroprotective. A similar effect was shown recently in a mouse model of Huntington's disease. Thus, the epigenetic mechanisms of gene regulation may provide yet another axis of neuroprotective therapeutic intervention.

12 Gene Expression

The role of transcription factors such as AP-1, HIF-1, and NF-κB in the context of cerebral ischemia has been recognized for much longer. They initiate the transcriptional activation of complex programs of gene expression. The whole hierarchical system of such programs has been unraveled over more than 15 years. Within few hours, the transcription of immediate early genes such as c-fos is activated. Some of these, in turn, form transcription factors such as AP-1. They induce a variety of genes that may have protective or destructive effects. Cells that survive the ensuing biochemical negotiation finally express genes that contribute to a structural reorganization and thus to greater plasticity. A great many of the genes involved have been subject to pathophysiological studies in stroke. Their functional significance has been studied either in descriptive ways, that is, by describing candidate genes on RNA or protein level or using transgenic technology. Especially knockout mice have been extremely useful for this approach. The genetic elimination of specific proteins in mice used for these experiments has been particularly enlightening for the untangling of the signaling cascades involved in and after stroke.

In recent years, the advent of genetic screening methods has allowed for the systematic investigation of complex gene expression programs that are initiated by cerebral ischemia. In contrast to candidate-gene research, they permit a more unbiased approach. They are based on the assumption that focal cerebral ischemia does incite complex changes of gene expression. Not just a single, specifically targeted gene is examined for its role in the scenario of stroke but – ideally – all genes. For this new research, paradigm terms have been coined like discovery science, genomics, transcriptomics, proteomics, and cellomics. So-called high-throughput technologies as well as a highly evolved bioinformatic methodology are a prerequisite for acquiring and managing the vast bulk of information involved. DNA micro arrays or SAGE technology facilitate the examination of global changes in gene expression on mRNA level (transcriptomics). Global changes on the level of protein synthesis can be described with a combination of two-dimensional gel electrophoresis and identification techniques

(proteomics). The first results of a number of studies using these techniques in the context of cerebral ischemia are now available.

What do these results imply? First of all, they confirm that the assumption of the existence of complex stroke-induced programs of gene expression is correct. Secondly, they demonstrate that they follow a robust temporal and spatial pattern. Not only do the expression patterns differ between core and penumbra but also within the penumbra there are many spatial and temporal differentiations. This has even given rise to a concept of "multiple penumbras." Thirdly, this approach has not only confirmed many findings of candidate gene research but also underpinned our pathophysiological understanding. Thus, a number of heat shock proteins, antioxidative enzymes, growth factors, and pro- as well as anti-inflammatory proteins have been detected. Not just transcription factors, but a variety of proteins have been found to play essential roles in the metabolism of RNA and DNA. Finally, proteins that contribute to structural and functional reorganization are thought to establish the molecular base of the functional healing potential after strokes. Not surprisingly, a number of hitherto unidentified genes have been brought forth by discovery science.

Although the results of these screening methods give a great boost to stroke research and open doors to new insights, a full functional characterization of the differentially expressed genes will take many more years. Our understanding of the emerging network of gene expression increasingly depends on not only biological but also mathematical models. However, there are things we do know already; for instance, that protein modifications are of great functional importance. Thus, the functional value of numerous proteins depends on their phosphorylation state. For example, the specific phosphorylation of the usually proapoptotic protein Bad prevents apoptosis in the ischemic penumbra. The differential function equals thus not simply differential gene expression, and genetic screening methods will never be sufficient to paint the full picture of stroke-induced molecular pathways. The same holds for proteomics, which, just like transcriptomics, neglects the functional importance of RNAs that do not code for proteins.

These molecular pathways, however, potentially lead to new, neuroprotective therapies and open up new perspectives for fundamental issues like neuronal plasticity and stem cell research.

13
Cell Replacement

In recent years, it has been shown in different models that ischemic brain injury induces neuron formation from neuronal stem cells. Such cells reside in the subgranular zone of the dentate gyrus, in the subventricular zone below the lateral ventricles, and periventricularly in the hippocampus. They can migrate into parts of ischemic lesions within the striatum, the dentate nucleus, and the hippocampal CA1 region. It is still a matter of debate as to what extent a similar neuroneogenesis also takes place within the penumbra in cortical regions. Neither is it entirely understood to what extent these migrating cells, which doubtlessly express neuronal markers, become functional. Observations so far indicate that it is not a very effective process and that most of the migrating cells perish. However, if these neuronal cells could be enabled or encouraged to

assume the function of damaged neurons, this would imply another new therapeutic paradigm. Indeed, intrathecal infusion of growth factors like FGF-2 and EGF after global cerebral ischemia did not only stimulate the migration and the survival of these cells but also resulted in improved functional results both on a cellular and on a neurobehavioral level. Astrocytes are newly formed in a similar fashion. For future therapeutic evaluation, it should be kept in mind not only that growth factors such as erythropoietin, FGF-2, EGF, VEGF, SCF, and BDNF have this beneficial effect but also that their expression depends on the stimulation of the NMDA receptor by glutamate. A pharmacological blockade of this receptor with the intention of inducing neuroprotection could interfere with this effect and thus be counterproductive. This is yet another example of an incomplete understanding of the signaling mechanisms, leading to a paradox in the interpretation of our neuroprotective data. Bearing this in mind, it is needless to say that the ongoing experimental studies of cell replacement are still far from clinical implementation. However, the available experimental observations are very exciting: For instance, neural and also haematopoietic pluripotent cells differentiate into neuronal and glial cells when injected into infarction sites. On the other hand, this potentially helpful strategy may have detrimental effects too. In a mouse model of stroke, for example, murine embryonic stem cells did not migrate into the damaged tissue, but produced highly malignant teratocarcinomas at the site of implantation, independent of whether or not they had been predifferentiated *in vitro* into neural progenitor cells. As far as safety is concerned, an endogenous pathway of cell regeneration may be more promising. Bone marrow–derived cells (BMDC) introduced into the brain possess a differentiation potential capable not only of forming microglia but also neurons. This apparent plasticity has been attributed to transdifferentiation. However, more recent data suggest, in the case of neurons, that BMDCs do not differentiate into but rather fuse with Purkinje cells in the brain. These observations suggest that cell fusion contributes to the maintenance of neurons. One potential use for haematopoietic cells migrating from blood into ischemic brain tissue could be the therapeutic transfer of protective gene products such as growth factors.

14
Endogenous Neuroprotection – Ischemic Tolerance

In the wake of ischemic stroke, not only cascades of destruction are induced but also endogenous protective mechanisms are activated. These mechanisms are currently studied intensively. Central to this research are models of ischemic preconditioning. The idea behind ischemic preconditioning is to achieve a protected state of a cell, tissue, or whole organism by subliminal exposure to a noxious stimulus (trigger), which is applied just below the level at which damage would occur. The stimulus induces a protective state against insults that would normally be lethal. Subliminal ischemia ('ischemic tolerance'), hypoxia, reactive oxygen free radicals, inflammation, and so on can serve as tolerance-inducing stimuli. After induction, there are two time windows within which ischemic tolerance can be observed. Early tolerance occurs after about 5 to 120 min following stimulation (classical preconditioning), while delayed preconditioning sets in after a latency of about

24 to 72 h. Seven days after induction, the state of protection is no longer detectable. Studies of patients who had transient ischemic attacks (TIA) before suffering a persistent stroke indicate that ischemic tolerance may also exist in humans. Subsequent to a TIA, brain infarcts were smaller in comparison with patients without preceding TIA. As a TIA, by definition, does not lead to infarction, it may thus be considered as a stimulus that induces tolerance.

Unlike severe ischemia, leading to stroke, a preconditioning stimulus induces exclusively protective, but not destructive, signaling cascades. There are basically three stages of protective signaling that can be distinguished in delayed preconditioning. In the *induction phase*, molecular sensors (mostly receptors, channels, and regulators) are activated by transcription factors. During the *transduction phase*, certain protein kinases, transcription factors, and paracrine and autocrine mediators, such as growth factors, amplify the signal and thereby prepare the *effector phase*. In this phase, proteins with a direct protective impact (antiapoptotic, anti-inflammatory, antioxidative) are switched on. These mechanisms are the molecular basis for ischemic tolerance and have been reviewed in detail elsewhere. Here, we want to give only a brief outline of the concept. On the basis of animal models in mice and rats as well as cell culture experiments, we were able to describe the following signaling cascade: In astrocytes, a tolerance-inducing hypoxic stimulus induces the transcription factor HIF-1 (hypoxia-inducible-factor-1). HIF-1 is one of the central sensors of hypoxia and induces the expression of erythropoietin (Epo). Erythropoietin is released by astrocytes and binds to the neuronal Epo receptor. Via the activation of a protein kinase cascade, Epo prevents apoptosis, as shown in Fig. 3. This neuroprotective signaling cascade mediates approximately 50% of the protection observed in our model. The remaining 50% of protection that it does not account for suggests the presence of further mechanisms in ischemic preconditioning. As a matter of fact, beside this paracrine cascade, which is mediated by astrocytes and neurons, a number of other mechanisms have been described. A host of antiexcitotoxic, anti-inflammatory and antiapoptotic mechanisms, including programs for regeneration and repair, have been implicated so far (Fig. 4).

Understanding these mechanisms may enable us in the future to induce or boost endogenous protection in patients as a novel strategy to safeguard the brain against hypoxic/ischemic damage. For example, Epo is a promising drug in stroke therapy. There is solid evidence that exogenously applied Epo-is neuroprotective *in vivo* and *in vitro* models of cerebral ischemia. Clinical phase I and II trials suggest that Epo is not only safe but also beneficial in the therapy for acute stroke. In addition, pharmacological inducers of ischemic tolerance in brain could become a promising tool in circumstances when ischemia can be anticipated, for example, before cardiac or carotid surgery. For example, desferrioxamine, an approved drug for the treatment of hemosiderosis, induces robust ischemic tolerance by inducing erythropoietin in an HIF-1-dependent manner.

Combined, the above-mentioned results demonstrate that brain cells are not only challenged by deleterious mechanisms but also activate innate programs to protect themselves from ischemia. However, cerebral ischemia does not only affect the brain

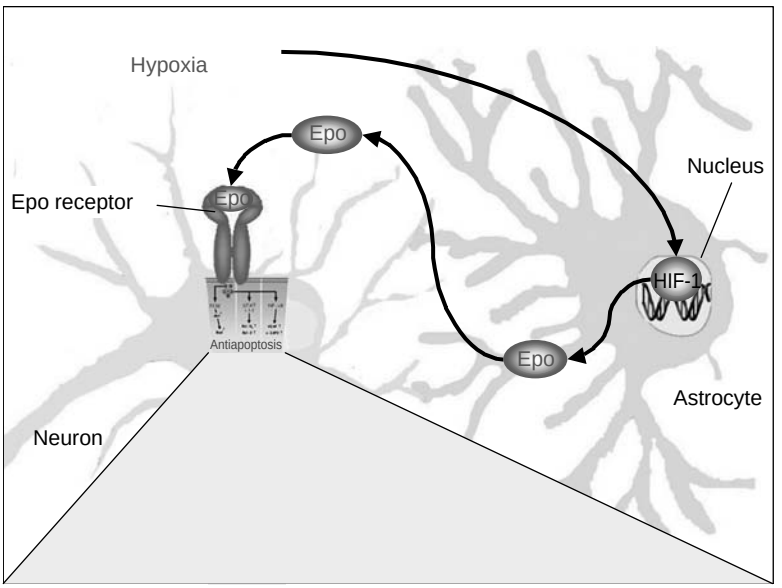
Hypoxia
Epo
Epo receptor
Epo
Nucleus
HIF-1
Antiapoptosis
Epo
Astrocyte
Neuron

(a)

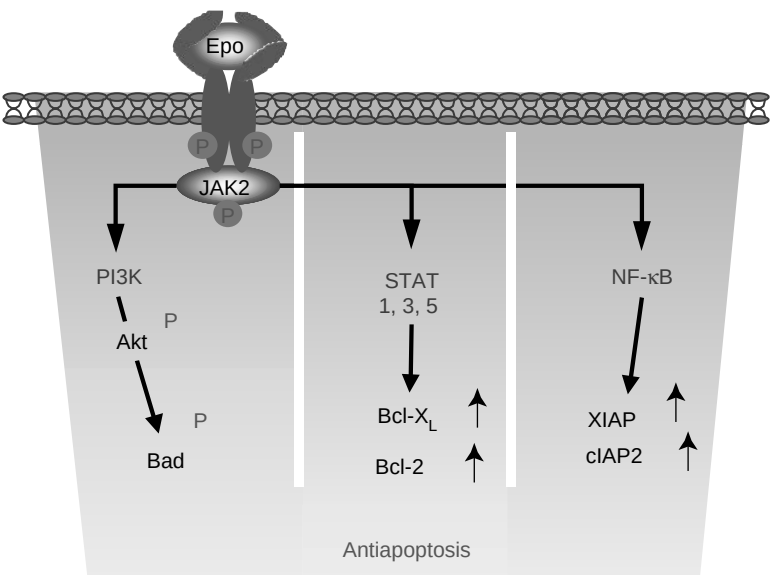
Epo
P
P
JAK2
P
PI3K
P
Akt
P
Bad
STAT
1, 3, 5
Bcl-X_L
Bcl-2
NF-κB
XIAP
cIAP2
Antiapoptosis

(b)

parenchyma but also other vital organs, in particular, the immune system.

15
Stroke-induced Immune Depression (SIDS)

Within three days after stroke, up to 61% of the stroke patients develop fever. Findings from animal experiments show that pyrexia of more than 1 °C does not only lead to a dramatic increase of the cytotoxic neurotransmitter glutamate but also to a significantly larger volume of the infarction. Clinical trials demonstrated that both mortality and morbidity rise with the occurrence of fever in stroke. The most common cause of fever in the acute course of stroke is infection. Infectious complications occur in 21 to 65% of stroke patients. Pneumonia is the most frequent

Fig. 3 Epo-mediated endogenous neuroprotection (a) The transcriptional activator protein HIF-1 functions as a global regulator of oxygen homeostasis that facilitates both oxygen delivery and adaptation to oxygen deprivation. HIF-1 consists of two subunits, HIF-1α and HIF-1β. The HIF-1ß subunit is constitutively expressed, whereas the HIF-1α subunit is precisely regulated by the cellular oxygen concentration. Most relevant is the regulation of HIF-1α steady state protein levels. Under normoxic conditions, HIF-1α is rapidly hydroxylated by prolyl hydroxylase domain containing protein (PHD), subsequently ubiquitinated and degraded through a proteasomal pathway. Since oxygen is rate limiting for PH activity, hypoxia inhibits HIF-1α degradation and HIF-1α can heterodimerize with HIF-1ß. After dimerization, HIF-1 enters the nucleus, binds to its recognition site 5′-RCGTG-3′ in the hypoxia response element, and recruits coactivators as well as basal transcription factors. This transcription complex initiates the gene expression that mediates response to hypoxia. More than 40 HIF-1 target genes are known. One of them is Epo, a glycoprotein hormone that regulates hematopoiesis by promoting cell survival, proliferation, and differentiation of immature erythroid cells. Although Epo is mainly expressed in adult kidney and fetal liver tissue, it is also found in the brain. In ischemic preconditioning, low oxygen tension (hypoxia) induces Epo in astrocytes, depending on HIF-1. Epo is a key mediator of paracrine neuroprotection from astrocytes to neurons, which acts via its cognate receptor. (b) Epo exerts its neuroprotective effect by preventing neuronal apoptosis. The intracellular molecular pathways of Epo-induced neuroprotection involve tyrosine phosphorylation. Rather than containing an intrinsic protein tyrosine kinase activity, the relatively short cytoplasmatic domain of the Epo receptor (EpoR) interacts with nonreceptor tyrosine kinases. A key molecule is the janus kinase 2 (JAK2), which is constitutively associated with EpoR. The first step of this intracellular signaling is Epo-induced homodimerization of EpoR. Homodimerization of EpoR induces a conformational change of the whole protein complex, which place JAK2 in close proximity to EpoR, thus promoting transtyrosine phosphorylation and activation of JAK2. JAK2 induces phosphorylation of cytosolic tyrosines in the EpoR. This allows docking to the EpoR cytosolic domain of several signaling proteins, including phosphatidyl-inositol 3-kinase (PI3K) and Signal transducer and activator of transcription proteins (STAT1, STAT3, STAT5). These proteins, in turn, become tyrosine phosphorylated and thus activated. PI3K suppresses apoptotic cell death through its downstream effector protein kinase B (PKB)/Akt. Activated Akt kinase suppresses apoptosis by phosphorylation of the proapoptotic Bcl2-family protein Bad. Phosphorylated Bad prevents apoptosis, since unphosphorylated Bad is capable of forming heterodimers with the antiapoptotic proteins Bcl-X_L or Bcl-2 and thereby antagonizes their antiapoptotic function. STATs exert their antiapoptotic effects by transcriptional activation of antiapoptotic proteins from the bcl-2 gene family, in particular, bcl-2 and bcl-X_L. Activation of transcription factor NF-κB, a key event for ischemic preconditioning, is activated by phosphorylation of its inhibitor subunit IkappaB. NF-κB induces the expression of the inhibitor of apoptosis proteins cIAP2 and XIAP. Activating these three signaling pathways, Epo induces a fast and long-lasting neuroprotection by upregulation of antiapoptotic as well as downregulation of proapoptotic proteins.

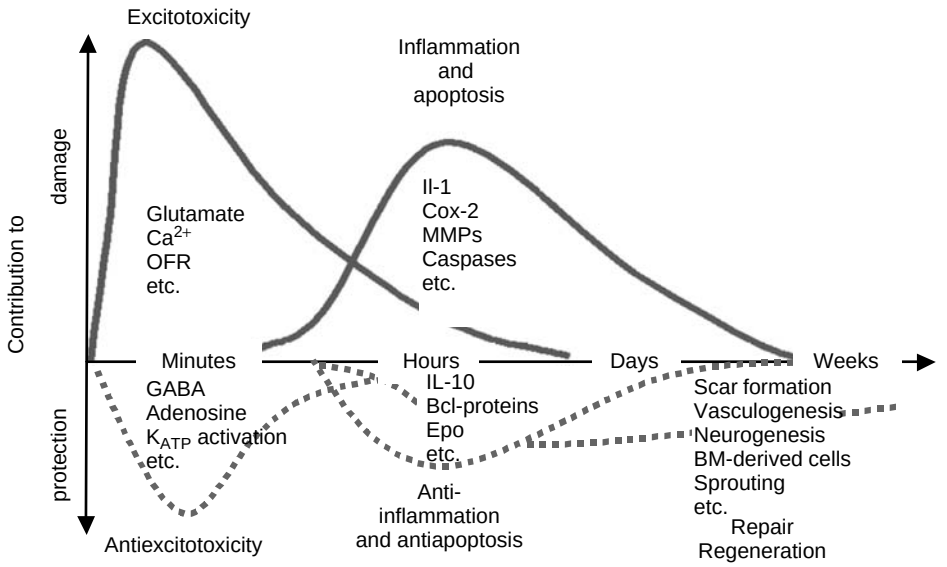

Fig. 4 Cell death–inducing and endogenous neuroprotective mechanisms in cerebral ischemia. Focal cerebral ischemia induces a complex series of mechanisms that damage brain cells. Brain tissue responds to most of the listed examples of noxious signals by inducing protective mechanisms. In infarcted tissue, destruction has overwhelmed protection, although tissue around the ischemic core might have been spared by restored substrate delivery (owing to delivery to collaterals, spontaneous or therapeutic reperfusion, or vasculogenesis) and cellular mechanisms of protection. In this cartoon, the *x*-axis reflects the evolution of the cascades over time and the *y*-axis illustrates the impact of each element of the destructive (top) and protective (bottom) cascades on the final outcome. The red dotted lines envelop major pathophysiological entities of tissue destruction in stroke (grouped for the acute mechanisms of excitotoxicity and the delayed mechanisms of inflammation and apoptosis) and the green broken lines envelop corresponding protective tissue responses. Abbreviations: BM, bone marrow; Cox-2, cyclooxygenase 2; Epo, erythropoietin; IL, interleukin; MMPs, matrix metalloproteinase; OFR, oxygen free radicals (Reprinted from Dirnagl, U., Simon, R.P., Hallenbeck, J.M. (2003) Ischemic tolerance and endogenous neuroprotection, *Trends Neurosci.* **26**, 248–254, with permission from Elsevier). (See color plate p. xxvi).

infection in stroke patients and a major mortality risk factor.

Stroke patients have a much higher risk of developing severe bacterial infections. It is obvious that factors such as immobilization, reduced bulbar reflexes, drowsiness, and, subsequently, a higher risk of aspiration promote pulmonary infections. Although these are risk factors for bacterial infection, they cannot fully explain the high risk of infection observed. A large meta-analysis showed that other factors had to be equally important. In particular, a reduced function of the immune system was postulated. Very recently, we were able to prove that immunodepression following stroke causes severe bacterial infections. In a mouse model, we demonstrated that, three days after stroke, severe bacterial infections (mostly pneumonia and sepsis) develop spontaneously. Stroke induces an overactivation of the sympathetic nervous system and leads to a rapid, severe, and sustained lymphopenia as well

as to a disturbed function of lymphocytes and monocytes. Interestingly, there is also evidence for an overactivation of the sympathetic nervous system within the first two days after stroke in humans. Furthermore, we were able to show that a disturbed secretion of interferon-γ from T cells and NK cells is central to these severe infections. Both a cellular reconstitution of the immune system and the application of interferon-γ prevented infections. In addition, pharmacological inhibition of the sympathetic nervous system with the β-receptor-blocker propranolol not only prevented infections but also dramatically reduced the high rate of mortality in this model. Our studies are the first to provide a mechanistic explanation for the clinical phenomenon of increased susceptibility to infections after stroke.

Approaches to prevent infection after stroke seem to be desirable. Preventive anti-infective therapy is expected to reduce a number of negative prognostic factors such as fever, infection-induced arterial hypotension, and the systemic release of proinflammatory cytokines. In addition, with the prevention of severe infections, an earlier mobilization and rehabilitation becomes feasible. In an experimental stroke model, we were able to demonstrate that preventive anti-infective therapy with an antibiotic after focal cerebral ischemia not only reduces mortality and the infarct sizes but also decreases the functional deficit. These experimental results lead to the initiation of a Phase IIb trial for preventive antibacterial short-term therapy in patients with acute ischemic media infarction (Preventive AN*ti-infective* THER*apy* I*n* S*troke*; PANTHERIS, Berlin). Other groups initiated similar trials (A. Chamorro, Spain; M. Hennerici, Germany; personal communications).

16 Perspectives

Owing to intensive experimental and clinical research, our pathophysiological understanding of cerebral ischemia has greatly improved over the last 15 years. Complex pathways involving excitotoxicity, oxidative and nitrosative stress, peri-infarct depolarizations, inflammation, and apoptosis-like mechanisms have been described in cell culture as well as animal models of cerebral ischemia. Despite impressive results in animal experiments on neuroprotection, the transfer of preclinical results into human stroke therapy has been so far unsuccessful. Possible reasons for this dilemma include problems with the models used, the design of the clinical trials as well as the drugs, and they have been discussed intensively. However, we believe that although we have often not been able to transfer our experimental results into human stroke therapy, some cautious optimism is justified. More and more experimental data demonstrate that the division of molecular mechanisms in "good" and "bad guys" is an improper simplification, and we need to gather more data in order to identify more suitable molecular targets. Since in stroke there are multiple pathways that cause cell death, it seems reasonable to develop combination therapies targeting all these mechanisms. In addition, the advances in understanding endogenous neuroprotective mechanisms provide new therapeutic options to improve tissue survival after stroke. Targeting systemic complications like severe bacterial infections in a preventive manner may also improve outcome. Thus, an ever more refined understanding of complex damaging cascades as well as the development of more complex therapeutic strategies will

continue to be the basis for more successful therapies of stroke in the future.

See also Brain Development.

Bibliography

Books and Reviews

Caplan, L.R. (2000) Caplan's stroke, *A Clinical Approach*, Butterworth-Heinemann, Boston, MA.

Chan, P.H. (2001) Reactive oxygen radicals in signaling and damage in the ischemic brain, *J. Cereb. Blood Flow Metab.* **21**, 2–14.

Dirnagl, U., Iadecola, C., Moskowitz, M.A. (1999) Pathobiology of ischaemic stroke: an integrated view, *Trends Neurosci.* **22**, 391–397.

Dirnagl, U., Simon, R.P., Hallenbeck, J.M. (2003) Ischemic tolerance and endogenous neuroprotection, *Trends Neurosci.* **26**, 248–254.

Emsley, H.C., Tyrrell, P.J. (2002) Inflammation and infection in clinical stroke, *J. Cereb. Blood Flow Metab.* **22**, 1399–1419.

Friedlander, R.M. (2003) Apoptosis and caspases in neurodegenerative diseases, *N. Engl. J. Med.* **348**, 1365–1375.

Kokaia, Z., Lindvall, O. (2003) Neurogenesis after ischaemic brain insults, *Curr. Opin. Neurobiol.* **13**, 127–132.

Lipton, P. (1999) Ischemic cell death in brain neurons, *Physiol. Rev.* **79**, 1431–1568.

Liu, P.K., Grossman, R.G., Hsu, C.Y., Robertson, C.S. (2001) Ischemic injury and faulty gene transcripts in the brain, *Trends Neurosci.* **24**, 581–588.

Lo, E.H., Dalkara, T., Moskowitz, M.A. (2003) Mechanisms, challenges and opportunities in stroke, *Nat. Rev. Neurosci.* **4**, 399–415.

Sharp, F.R., Bernaudin, M. (2004) HIF1 and oxygen sensing in the brain, *Nat. Rev. Neurosci.* **5**, 437–448.

Primary Literature

Arvidsson, A., Collin, T., Kirik, D., Kokaia, Z., Lindvall, O. (2002) Neuronal replacement from endogenous precursors in the adult brain after stroke, *Nat. Med.* **8**, 963–970.

Asahi, M., Wang, X., Mori, T., Sumii, T., Jung, J.C., Moskowitz, M.A., Fini, M.E., Lo, E.H. (2001) Effects of matrix metalloproteinase-9 gene knock-out on the proteolysis of blood-brain barrier and white matter components after cerebral ischemia, *J. Neurosci.* **21**, 7724–7732.

Azzimondi, G., Bassein, L., Nonino, F., Fiorani, L., Vignatelli, L., Re, G., D'Alessandro, R. (1995) Fever in acute stroke worsens prognosis. A prospective study, *Stroke* **26**, 2040–2043.

Back, T., Ginsberg, M.D., Dietrich, W.D., Watson, B.D. (1996) Induction of spreading depression in the ischemic hemisphere following experimental middle cerebral artery occlusion: effect on infarct morphology, *J. Cereb. Blood Flow Metab.* **16**, 202–213.

Baena, R.C., Busto, R., Dietrich, W.D., Globus, M.Y., Ginsberg, M.D. (1997) Hyperthermia delayed by 24 hours aggravates neuronal damage in rat hippocampus following global ischemia, *Neurology* **48**, 768–773.

Baird, T.A., Parsons, M.W., Phanh, T., Butcher, K.S., Desmond, P.M., Tress, B.M., Colman, P.G., Chambers, B.R., Davis, S.M. (2003) Persistent poststroke hyperglycemia is independently associated with infarct expansion and worse clinical outcome, *Stroke* **34**, 2208–2214.

Bak, S., Gaist, D., Sindrup, S.H., Skytthe, A., Christensen, K. (2002) Genetic liability in stroke: a long-term follow-up study of Danish twins, *Stroke* **33**, 769–774.

Becker, K., Kindrick, D., McCarron, R., Hallenbeck, J., Winn, R. (2003) Adoptive transfer of myelin basic protein-tolerized splenocytes to naive animals reduces infarct size: a role for lymphocytes in ischemic brain injury? *Stroke* **34**, 1809–1815.

Becker, K., Kindrick, D., Relton, J., Harlan, J., Winn, R. (2001) Antibody to the alpha4 integrin decreases infarct size in transient focal cerebral ischemia in rats, *Stroke* **32**, 206–211.

Becker, K.J., McCarron, R.M., Ruetzler, C., Laban, O., Sternberg, E., Flanders, K.C., Hallenbeck, J.M. (1997) Immunologic tolerance to myelin basic protein decreases stroke size after transient focal cerebral ischemia, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 10873–10878.

Bonita, R. (1992) Epidemiology of stroke, *Lancet* **339**, 342–344.

Boysen, G., Christensen, H. (2001) Stroke severity determines body temperature in acute stroke, *Stroke* **32**, 413–417.

Bruce, A.J., Boling, W., Kindy, M.S., Peschon, J., Kraemer, P.J., Carpenter, M.K., Holtsberg, F.W., Mattson, M.P. (1996) Altered neuronal and microglial responses to excitotoxic and ischemic brain injury in mice lacking TNF receptors, *Nat. Med.* **2**, 788–794.

Busch, E., Gyngell, M.L., Eis, M., Hoehn-Berlage, M., Hossmann, K.A. (1996) Potassium-induced cortical spreading depressions during focal cerebral ischemia in rats: contribution to lesion growth assessed by diffusion-weighted NMR and biochemical imaging, *J. Cereb. Blood Flow Metab.* **16**, 1090–1099.

Cao, G., Pei, W., Ge, H., Liang, Q., Luo, Y., Sharp, F.R., Lu, A., Ran, R., Graham, S.H., Chen, J. (2002) In vivo delivery of a Bcl-xL fusion protein containing the TAT protein transduction domain protects against ischemic brain injury and neuronal apoptosis, *J. Neurosci.* **22**, 5423–5431.

Castillo, J., Davalos, A., Marrugat, J., Noya, M. (1998) Timing for fever-related brain damage in acute ischemic stroke, *Stroke* **29**, 2455–2460.

Chan, P.H. (2001) Reactive oxygen radicals in signaling and damage in the ischemic brain, *J. Cereb. Blood Flow Metab.* **21**, 2–14.

Chen, H., Chopp, M., Zhang, R.L., Bodzin, G., Chen, Q., Rusche, J.R., Todd, R.F. 3rd. (1994) Anti-CD11b monoclonal antibody reduces ischemic cell damage after transient focal cerebral ischemia in rat, *Ann. Neurol.* **35**, 458–463.

Chopp, M., Chan, P.H., Hsu, C.Y., Cheung, M.E., Jacobs, T.P. (1996) DNA damage and repair in central nervous system injury: National Institute of Neurological Disorders and Stroke Workshop Summary, *Stroke* **27**, 363–369.

Crumrine, R.C., Thomas, A.L., Morgan, P.F. (1994) Attenuation of p53 expression protects against focal ischemic damage in transgenic mice, *J. Cereb. Blood Flow Metab.* **14**, 887–891.

Davenport, R.J., Dennis, M.S., Wellwood, I., Warlow, C.P. (1996) Complications after acute stroke, *Stroke* **27**, 415–420.

del Zoppo, G.J. (1995) Why do all drugs work in animals but none in stroke patients? 1. Drugs promoting cerebral blood flow, *J. Intern. Med.* **237**, 79–88.

del Zoppo, G.J., Schmid-Schonbein, G.W., Mori, E., Copeland, B.R., Chang, C.M. (1991) Polymorphonuclear leukocytes occlude capillaries following middle cerebral artery occlusion and reperfusion in baboons, *Stroke* **22**, 1276–1283.

Digicaylioglu, M., Lipton, S.A. (2001) Erythropoietin-mediated neuroprotection involves cross-talk between Jak2 and NF-kappaB signalling cascades, *Nature* **412**, 641–647.

Eglitis, M.A., Mezey, E. (1997) Hematopoietic cells differentiate into both microglia and macroglia in the brains of adult mice, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 4080–4085.

Ehrenreich, H., Hasselblatt, M., Dembowski, C., Cepek, L., Lewczuk, P., Stiefel, M., Rustenbeck, H.H., Breiter, N., Jacob, S., Knerlich, F., Bohn, M., Poser, W., Ruther, E., Kochen, M., Gefeller, O., Gleiter, C., Wessel, T.C., De Ryck, M., Itri, L., Prange, H., Cerami, A., Brines, M., Siren, A.L. (2002) Erythropoietin therapy for acute stroke is both safe and beneficial, *Mol. Med.* **8**, 495–505.

Eliasson, M.J., Sampei, K., Mandir, A.S., Hurn, P.D., Traystman, R.J., Bao, J., Pieper, A., Wang, Z.Q., Dawson, T.M., Snyder, S.H., Dawson, V.L. (1997) Poly(ADP-ribose) polymerase gene disruption renders mice resistant to cerebral ischemia, *Nat. Med.* **3**, 1089–1095.

Endres, M., Wang, Z.Q., Namura, S., Waeber, C., Moskowitz, M.A. (1997) Ischemic brain injury is mediated by the activation of poly(ADP-ribose)polymerase, *J. Cereb. Blood Flow Metab.* **17**, 1143–1151.

Endres, M., Fink, K., Zhu, J., Stagliano, N.E., Bondada, V., Geddes, J.W., Azuma, T., Mattson, M.P., Kwiatkowski, D.J., Moskowitz, M.A. (1999) Neuroprotective effects of gelsolin during murine stroke, *J. Clin. Invest.* **103**, 347–354.

Endres, M., Meisel, A., Biniszkiewicz, D., Namura, S., Prass, K., Ruscher, K., Lipski, A., Jaenisch, R., Moskowitz, M.A., Dirnagl, U. (2000) DNA methyltransferase contributes to delayed ischemic brain injury, *J. Neurosci.* **20**, 3175–3181.

Endres, M., Biniszkiewicz, D., Sobol, R.W., Harms, C., Ahmadi, M., Lipski, A., Katchanov, J., Mergenthaler, P., Dirnagl, U., Wilson, S.H., Meisel, A., Jaenisch, R. (2004) Increased postischemic brain injury in mice deficient in uracil-DNA glycosylase, *J. Clin. Invest.* **113**, 1711–1721.

Englander, E.W., Hu, Z., Sharma, A., Lee, H.M., Wu, Z.H., Greeley, G.H. (2002) Rat MYH, a glycosylase for repair of oxidatively damaged

DNA, has brain-specific isoforms that localize to neuronal mitochondria, *J. Neurochem.* **83**, 1471–1480.

Evans, A., Perez, I., Harraf, F., Melbourn, A., Steadman, J., Donaldson, N., Kalra, L. (2001) Can differences in management processes explain different outcomes between stroke unit and stroke-team care? *Lancet* **358**, 1586–1592.

Gasche, Y., Fujimura, M., Morita-Fujimura, Y., Copin, J.C., Kawase, M., Massengale, J., Chan, P.H. (1999) Early appearance of activated matrix metalloproteinase-9 after focal cerebral ischemia in mice: a possible role in blood-brain barrier dysfunction, *J. Cereb. Blood Flow Metab.* **19**, 1020–1028.

Ginsberg, M.D., Prado, R., Dietrich, W.D., Busto, R., Watson, B.D. (1987) Hyperglycemia reduces the extent of cerebral infarction in rats, *Stroke* **18**, 570–574.

Gladstone, D.J., Black, S.E., Hakim, A.M., Heart and Stroke Foundation of Ontario Centre of Excellence in Stroke Recovery. (2002) Toward wisdom from failure: lessons from neuroprotective stroke trials and new therapeutic directions, *Stroke* **33**, 2123–2136.

Gretarsdottir, S., Thorleifsson, G., Reynisdottir, S.T., Manolescu, A., Jonsdottir, S., Jonsdottir, T., Gudmundsdottir, T., Bjarnadottir, S.M., Einarsson, O.B., Gudjonsdottir, H.M., Hawkins, M., Gudmundsson, G., Gudmundsdottir, H., Andrason, H., Gudmundsdottir, A.S., Sigurdardottir, M., Chou, T.T., Nahmias, J., Goss, S., Sveinbjornsdottir, S., Valdimarsson, E.M., Jakobsson, F., Agnarsson, U., Gudnason, V., Thorgeirsson, G., Fingerle, J., Gurney, M., Gudbjartsson, D., Frigge, M.L., Kong, A., Stefansson, K., Gulcher, J.R. (2003) The gene encoding phosphodiesterase 4D confers risk of ischemic stroke, *Nat. Genet.* **35**, 131–138.

Hajat, C., Hajat, S., Sharma, P. (2000) Effects of poststroke pyrexia on stroke outcome: a meta-analysis of studies in patients, *Stroke* **31**, 410–414.

Hara, H., Friedlander, R.M., Gagliardini, V., Ayata, C., Fink, K., Huang, Z., Shimizu-Sasamata, M., Yuan, J., Moskowitz, M.A. (1997) Inhibition of interleukin 1beta converting enzyme family proteases reduces ischemic and excitotoxic neuronal damage, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 2007–2012.

Heiss, W.D., Kracht, L.W., Thiel, A., Grond, M., Pawlik, G. (2001) Penumbral probability thresholds of cortical flumazenil binding and blood flow predicting tissue outcome in patients with cerebral ischaemia, *Brain* **124**, 20–29.

Hilker, R., Poetter, C., Findeisen, N., Sobesky, J., Jacobs, A., Neveling, M., Heiss, W.D. (2003) Nosocomial pneumonia after acute stroke: implications for neurological intensive care medicine, *Stroke* **34**, 975–981.

Hoehn-Berlage, M. (1995) Diffusion-weighted NMR imaging: application to experimental focal cerebral ischemia, *NMR Biomed.* **8**, 345–358.

Iadecola, C., Zhang, F., Casey, R., Nagayama, M., Ross, M.E. (1997) Delayed reduction of ischemic brain injury and neurological deficits in mice lacking the inducible nitric oxide synthase gene, *J. Neurosci.* **17**, 9157–9164.

Iadecola, C., Niwa, K., Nogawa, S., Zhao, X., Nagayama, M., Araki, E., Morham, S., Ross, M.E. (2001) Reduced susceptibility to ischemic brain injury and N-methyl-D-aspartate-mediated neurotoxicity in cyclooxygenase-2-deficient mice, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1294–1299.

Jin, K., Mao, X.O., Sun, Y., Xie, L., Greenberg, D.A. (2002) Stem cell factor stimulates neurogenesis in vitro and in vivo, *J. Clin. Invest.* **110**, 311–319.

Jin, K., Mao, X.O., Eshoo, M.W., Nagayama, T., Minami, M., Simon, R.P., Greenberg, D.A. (2001) Microarray analysis of hippocampal gene expression in global cerebral ischemia, *Ann. Neurol.* **50**, 93–103.

Joutel, A., Corpechot, C., Ducros, A., Vahedi, K., Chabriat, H., Mouton, P., Alamowitch, S., Domenga, V., Cecillion, M., Marechal, E., Maciazek, J., Vayssiere, C., Cruaud, C., Cabanis, E.A., Ruchoux, M.M., Weissenbach, J., Bach, J.F., Bousser, M.G., Tournier-Lasserve, E. (1996) Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia, *Nature* **383**, 707–710.

Katzan, I.L., Cebul, R.D., Husak, S.H., Dawson, N.V., Baker, D.W. (2003) The effect of pneumonia on mortality among patients hospitalized for acute stroke, *Neurology* **60**, 620–625.

Keyvani, K., Witte, O.W., Paulus, W. (2002) Gene expression profiling in perilesional and contralateral areas after ischemia in rat brain, *J. Cereb. Blood Flow Metab.* **22**, 153–160.

Kilic, E., Dietz, G.P., Hermann, D.M., Bahr, M. (2002) Intravenous TAT-Bcl-Xl is protective

after middle cerebral artery occlusion in mice, *Ann. Neurol.* **52**, 617–622.

Lapchak, P.A., Chapman, D.F., Zivin, J.A. (2000) Metalloproteinase inhibition reduces thrombolytic (tissue plasminogen activator)-induced hemorrhage after thromboembolic stroke, *Stroke* **31**, 3034–3040.

Le, D.A., Wu, Y., Huang, Z., Matsushita, K., Plesnila, N., Augustinack, J.C., Hyman, B.T., Yuan, J., Kuida, K., Flavell, R.A., Moskowitz, M.A. (2003) Caspase activation and neuroprotection in caspase-3- deficient mice after in vivo cerebral ischemia and in vitro oxygen glucose deprivation, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 15188–15193.

Lindsberg, P.J., Carpen, O., Paetau, A., Karjalainen-Lindsberg, M.L., Kaste, M. (1996) Endothelial ICAM-1 expression associated with inflammatory cell response in human ischemic stroke, *Circulation* **94**, 939–945.

Linnik, M.D., Zahos, P., Geschwind, M.D., Federoff, H.J. (1995) Expression of bcl-2 from a defective herpes simplex virus-1 vector limits neuronal death in focal cerebral ischemia, *Stroke* **26**, 1670–1674.

Liu, P.K., Hsu, C.Y., Dizdaroglu, M., Floyd, R.A., Kow, Y.W., Karakaya, A., Rabow, L.E., Cui, J.K. (1996) Damage repair, and mutagenesis in nuclear genes after mouse forebrain ischemia-reperfusion, *J. Neurosci.* **16**, 6795–6806.

Lu, A., Tang, Y., Ran, R., Clark, J.F., Aronow, B.J., Sharp, F.R. (2003) Genomics of the periinfarction cortex after focal cerebral ischemia, *J. Cereb. Blood Flow Metab.* **23**, 786–810.

Martinou, J.C., Dubois-Dauphin, M., Staple, J.K., Rodriguez, P., Frankowski, H., Missotten, M., Albertini, P., Talabot, D., Catsicas, S., Pietra, C., Huarte, J. (1994) Overexpression of bcl-2 in transgenic mice protects neurons from naturally occurring cell death and experimental ischemia, *Neuron* **13**, 1017–1030.

Meisel, C., Prass, K., Braun, J., Victorov, I., Wolf, T., Megow, D., Halle, E., Volk, H.D., Dirnagl, U., Meisel, A. (2004) Preventive antibacterial treatment improves the general medical and neurological outcome in a mouse model of stroke, *Stroke* **35**, 2–6.

Moncayo, J., de Freitas, G.R., Bogousslavsky, J., Altieri, M., van Melle, G. (2000) Do transient ischemic attacks have a neuroprotective effect? *Neurology* **54**, 2089–2094.

Mori, E., del Zoppo, G.J., Chambers, J.D., Copeland, B.R., Arfors, K.E. (1992) Inhibition of polymorphonuclear leukocyte adherence suppresses no-reflow after focal cerebral ischemia in baboons, *Stroke* **23**, 712–718.

Nakatomi, H., Kuriu, T., Okabe, S., Yamamoto, S., Hatano, O., Kawahara, N., Tamura, A., Kirino, T., Nakafuku, M. (2002) Regeneration of hippocampal pyramidal neurons after ischemic brain injury by recruitment of endogenous neural progenitors, *Cell* **110**, 429–441.

Namura, S., Zhu, J., Fink, K., Endres, M., Srinivasan, A., Tomaselli, K.J., Yuan, J., Moskowitz, M.A. (1998) Activation and cleavage of caspase-3 in apoptosis induced by experimental cerebral ischemia, *J. Neurosci.* **18**, 3659–3668.

Nedergaard, M., Hansen, A.J. (1993) Characterization of cortical depolarizations evoked in focal cerebral ischemia, *J. Cereb. Blood Flow Metab.* **13**, 568–574.

Parsons, M.W., Barber, P.A., Desmond, P.M., Baird, T.A., Darby, D.G., Byrnes, G., Tress, B.M., Davis, S.M. (2002) Acute hyperglycemia adversely affects stroke outcome: a magnetic resonance imaging and spectroscopy study, *Ann. Neurol.* **52**, 20–28.

Perry, L., Love, C.P. (2001) Screening for dysphagia and aspiration in acute stroke: a systematic review, *Dysphagia* **16**, 7–18.

Pfefferkorn, T., Rosenberg, G.A. (2003) Closure of the blood-brain barrier by matrix metalloproteinase inhibition reduces rtPA-mediated mortality in cerebral ischemia with delayed reperfusion, *Stroke* **34**, 2025–2030.

Prass, K., Scharff, A., Ruscher, K., Lowl, D., Muselmann, C., Victorov, I., Kapinya, K., Dirnagl, U., Meisel, A. (2003) Hypoxia-induced stroke tolerance in the mouse is mediated by erythropoietin, *Stroke* **34**, 1981–1986.

Prass, K., Meisel, C., Hoflich, C., Braun, J., Halle, E., Wolf, T., Ruscher, K., Victorov, I.V., Priller, J., Dirnagl, U., Volk, H.D., Meisel, A. (2003) Stroke-induced immunodeficiency promotes spontaneous bacterial infections and is mediated by sympathetic activation reversal by poststroke T helper cell Type 1-like immunostimulation, *J. Exp. Med.* **198**, 725–736.

Priller, J., Flugel, A., Wehner, T., Boentert, M., Haas, C.A., Prinz, M., Fernandez-Klett, F., Prass, K., Bechmann, I., de Boer, B.A., Frotscher, M., Kreutzberg, G.W., Persons, D.A., Dirnagl, U. (2001) Targeting gene-modified hematopoietic cells to the central

nervous system: use of green fluorescent protein uncovers microglial engraftment, *Nat. Med.* **7**, 1356–1361.

Pulsinelli, W.A., Waldman, S., Rawlinson, D., Plum, F. (1982) Moderate hyperglycemia augments ischemic brain damage: a neuropathologic study in the rat, *Neurology* **32**, 1239–1246.

Raghavendra Rao, V.L., Bowen, K.K., Dhodda, V.K., Song, G., Franklin, J.L., Gavva, N.R., Dempsey, R.J. (2002) Gene expression analysis of spontaneously hypertensive rat cerebral cortex following transient focal cerebral ischemia, *J. Neurochem.* **83**, 1072–1086.

Reith, J., Jorgensen, H.S., Pedersen, P.M., Nakayama, H., Raaschou, H.O., Jeppesen, L.L., Olsen, T.S. (1996) Body temperature in acute stroke: relation to stroke severity, infarct size, mortality, and outcome, *Lancet* **347**, 422–425.

Rosenberg, G.A., Estrada, E.Y., Dencoff, J.E. (1998) Matrix metalloproteinases and TIMPs are associated with blood-brain barrier opening after reperfusion in rat brain, *Stroke* **29**, 2189–2195.

Ruscher, K., Isaev, N., Trendelenburg, G., Weih, M., Iurato, L., Meisel, A., Dirnagl, U. (1998) Induction of hypoxia inducible factor 1 by oxygen glucose deprivation is attenuated by hypoxic preconditioning in rat cultured neurons, *Neurosci. Lett.* **254**, 117–120.

Ruscher, K., Freyer, D., Karsch, M., Isaev, N., Megow, D., Sawitzki, B., Priller, J., Dirnagl, U., Meisel, A. (2002) Erythropoietin is a paracrine mediator of ischemic tolerance in the brain: evidence from an in vitro model, *J. Neurosci.* **22**, 10291–10301.

Sairanen, T., Ristimaki, A., Karjalainen-Lindsberg, M.L., Paetau, A., Kaste, M., Lindsberg, P.J. (1998) Cyclooxygenase-2 is induced globally in infarcted human brain, *Ann. Neurol.* **43**, 738–747.

Saito, A., Hayashi, T., Okuno, S., Ferrand-Drake, M., Chan, P.H. (2003) Interaction between XIAP and Smac/DIABLO in the mouse brain after transient focal cerebral ischemia, *J. Cereb. Blood Flow Metab.* **23**, 1010–1019.

Schurr, A., Payne, R.S., Miller, J.J., Tseng, M.T. (2001) Preischemic hyperglycemia-aggravated damage: evidence that lactate utilization is beneficial and glucose-induced corticosterone release is detrimental, *J. Neurosci. Res.* **66**, 782–789.

Seo, S.Y., Kim, E.Y., Kim, H., Gwag, B.J. (1999) Neuroprotective effect of high glucose against NMDA, free radical, and oxygen-glucose deprivation through enhanced mitochondrial potentials, *J. Neurosci.* **19**, 8849–8855.

Sharp, F.R., Lu, A., Tang, Y., Millhorn, D.E. (2000) Multiple molecular penumbras after focal cerebral ischemia, *J. Cereb. Blood Flow Metab.* **20**, 1011–1032.

Siren, A.L., Fratelli, M., Brines, M., Goemans, C., Casagrande, S., Lewczuk, P., Keenan, S., Gleiter, C., Pasquali, C., Capobianco, A., Mennini, T., Heumann, R., Cerami, A., Ehrenreich, H., Ghezzi, P. (2001) Erythropoietin prevents neuronal apoptosis after cerebral ischemia and metabolic stress, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4044–4049.

Soriano, S.G., Lipton, S.A., Wang, Y.F., Xia, M., Springer, T.A., Gutierrez-Ramos, J.C., Hickey, P.R. (1996) Intercellular adhesion molecule-1-deficient mice are less susceptible to cerebral ischemia-reperfusion injury, *Ann. Neurol.* **39**, 618–624.

Strong, A.J., Fabricius, M., Boutelle, M.G., Hibbins, S.J., Hopwood, S.E., Jones, R., Parkin, M.C., Lauritzen, M. (2002) Spreading and synchronous depressions of cortical activity in acutely injured human brain, *Stroke* **33**, 2738–2743.

Takeda, H., Spatz, M., Ruetzler, C., McCarron, R., Becker, K., Hallenbeck, J. (2002) Induction of mucosal tolerance to E-selectin prevents ischemic and hemorrhagic stroke in spontaneously hypertensive genetically stroke-prone rats, *Stroke* **33**, 2156–2163.

Tarkowski, E., Rosengren, L., Blomstrand, C., Wikkelso, C., Jensen, C., Ekholm, S., Tarkowski, A. (1995) Early intrathecal production of interleukin-6 predicts the size of brain lesion in stroke, *Stroke* **26**, 1393–1398.

Trapp, T., Korhonen, L., Besselmann, M., Martinez, R., Mercer, E.A., Lindholm, D. (2003) Transgenic mice overexpressing XIAP in neurons show better outcome after transient cerebral ischemia, *Mol. Cell. Neurosci.* **23**, 302–313.

Trendelenburg, G., Prass, K., Priller, J., Kapinya, K., Polley, A., Muselmann, C., Ruscher, K., Kannbley, U., Schmitt, A.O., Castell, S., Wiegand, F., Meisel, A., Rosenthal, A., Dirnagl, U. (2002) Serial analysis of gene expression identifies metallothionein-II as major neuroprotective gene in mouse focal cerebral ischemia, *J. Neurosci.* **22**, 5879–5888.

Wang, H., Zhan, Y., Xu, L., Feuerstein, G.Z., Wang, X. (2001) Use of suppression subtractive hybridization for differential gene expression in stroke: discovery of CD44 gene expression and localization in permanent focal stroke in rats, *Stroke* **32**, 1020–1027.

Wang, X., Lee, S.R., Arai, K., Lee, S.R., Tsuji, K., Rebeck, G.W., Lo, E.H. (2003) Lipoprotein receptor-mediated induction of matrix metalloproteinase by tissue plasminogen activator, *Nat. Med.* **9**, 1313–1317.

Wang, Y.F., Tsirka, S.E., Strickland, S., Stieg, P.E., Soriano, S.G., Lipton, S.A. (1998) Tissue plasminogen activator (tPA) increases neuronal damage after focal cerebral ischemia in wild-type and tPA-deficient mice, *Nat. Med.* **4**, 228–231.

Wass, C.T., Lanier, W.L., Hofer, R.E., Scheithauer, B.W., Andrews, A.G. (1995) Temperature changes of > or = 1 degree C alter functional neurologic outcome and histopathology in a canine model of complete cerebral ischemia, *Anesthesiology* **83**, 325–335.

Wehner, T., Boentert, M., Eyupoglu, I., Prass, K., Prinz, M., Klett, F.F., Heinze, M., Bechmann, I., Nitsch, R., Kirchhoff, F., Kettenmann, H., Dirnagl, U., Priller, J. (2003) Bone marrow-derived cells expressing green fluorescent protein under the control of the glial fibrillary acidic protein promoter do not differentiate into astrocytes in vitro and in vivo, *J. Neurosci.* **23**, 5004–5011.

Weih, M., Kallenberg, K., Bergk, A., Dirnagl, U., Harms, L., Wernecke, K.D., Einhaupl, K.M. (1999) Attenuated stroke severity after prodromal TIA: a role for ischemic tolerance in the brain? *Stroke* **30**, 1851–1854.

Yrjänheikki, J., Tikka, T., Keinanen, R., Goldsteins, G., Chan, P.H., Koistinaho, J. (1999) A tetracycline derivative, minocycline, reduces inflammation and protects against focal cerebral ischemia with a wide therapeutic window, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13496–13500.

Yu, S.W., Wang, H., Poitras, M.F., Coombs, C., Bowers, W.J., Federoff, H.J., Poirier, G.G., Dawson, T.M., Dawson, V.L. (2002) Mediation of poly(ADP-ribose) polymerase-1-dependent cell death by apoptosis-inducing factor, *Science* **297**, 259–263.

Peptide and Non-Peptide Combinatorial Libraries

Ruiwu Liu and Kit S. Lam
University of California Davis, Sacramento, CA, USA

1 Principles 149
1.1 Random Combinatorial Libraries 149
1.2 Synthetic Peptide Libraries, Synthesis and Screening 149
1.3 Synthetic Nonpeptide Oligomer or Small Molecule Libraries 151

2 Techniques 153
2.1 Biological Peptide Library Method 153
2.2 Spatially Addressable Parallel Library Method 156
2.2.1 Multipin Technology 156
2.2.2 SPOT Synthesis 156
2.2.3 NanoKan Technique 157
2.2.4 Chemical Microarray 157
2.3 Affinity Selection 159
2.4 Library Methods Requiring Deconvolution 160
2.5 OBOC Library Method 162
2.6 Nonpeptide Oligomer or Small Molecule Libraries 165
2.7 Encoding and Decoding for Nonpeptide Libraries 166

3 Application 169

4 Perspectives 171

Bibliography 172
Books and Reviews 172
Primary Literature 172

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

Keywords

Acceptor Molecule
A molecule, usually a macromolecule that has the potential of binding to specific peptides, or ligands.

Amino Acid
The smallest subunit of protein, having an amino, a carboxyl, and a side-chain functional group.

Combinatorial Peptide Library
A large collection of different peptides, with many possible combinations of amino acids joined together.

Combinatorial Nonpeptide Oligomer Library
A large collection of different oligomers, with many possible combinations of non-amino acid subunits joined together.

Combinatorial Small Molecule Library
A large collection of different small molecules, with many possible combinations of small, organic moieties joined together.

Motif
A pattern of amino acid sequences that interact specifically with an acceptor molecule.

Peptide
A polymeric compound formed by the condensation of two or more amino acids; here we refer to compound less than 20 amino acids long.

The field of combinatorial peptide chemistry was created when the first limited peptide library was synthesized with multipin technology in 1984. Since 1990, the field of combinatorial chemistry has advanced rapidly. It has emerged as a powerful tool in the study of many biological systems. In essence, a combinatorial peptide library (numbering 10^4–10^8 different molecules) is chemically synthesized and subjected to screening for biological activity. Such screening allows for the rapid discovery of peptide ligands with biological interest. The amino acid sequence of an "active" peptide is then determined. Other biologic peptide libraries, such as those involving filamentous phage-display peptides, can typically accommodate only the 20 natural amino acids. In contrast, synthetic peptide libraries have the potential to incorporate D-amino acids and other unnatural amino acids, as well as specific secondary structures or scaffolding structures that may enhance biologic activity. In addition to amino acids, biological subunits such as monosaccharides, nucleotides,

lipids, or even small organic moieties can be used. The advent of these peptide, nonpeptide oligomer, or small molecule library methods not only facilitates the drug discovery process but also provides important information for the fundamental understanding of molecular recognition.

1 Principles

1.1 Random Combinatorial Libraries

The traditional drug discovery process relies heavily on random screening of plant and animal extracts, fermentation broths, and synthetic compounds, individually or in mixtures. The process is often laborious, expensive, and time consuming. And even after an active mixture has been detected, a tedious process of purification and structure determination is still required. Once a drug lead has been identified, additional studies on the structure-activity relationship of individual synthetic analogues of the parent compound often are necessary for the optimization of the drug. The random combinatorial library synthesis and screening methods can speed up the drug discovery process. Notably, this method can accommodate much larger numbers of candidate molecules. Fig. 1 gives the number of possible permutations in a random peptide library when 20 eukaryotic amino acids are used in each coupling cycle. For example, there are 1.28×10^9 possible permutations for a random heptapeptide. The number of peptides present in a library ranges from a thousand to as many as 100 million, depending on the method used. In general, the methods of generating a combinatorial library can be categorized into two approaches: the synthetic approach and the biological approach.

1.2 Synthetic Peptide Libraries, Synthesis and Screening

Most of the synthetic peptide libraries are synthesized on a solid-phase support system, such as a polystyrene resin bead, using either the Fmoc/*t*-But (9-fluorenylmethoxycarbonyl/*tert*-butyl) or *t*-Boc/Bn (*tert*-butoxycarbonyl/benzyl) amino acid protection strategies. However, hydrofluoric acid (HF) is required to cleave off the side-chain protecting groups for the *t*-Boc/Bn chemistry. Since HF is highly toxic, requires a special apparatus, and partially degrades the polyethyleneglycol (PEG) chain on the resin, the Fmoc/*t*-But strategy is preferable for the synthesis

Number of permutations for a peptide

X	$20^1 =$	20
XX	$20^2 =$	400
XXX	$20^3 =$	8 000
XXXX	$20^4 =$	160 000
XXXXX	$20^5 =$	3 200 000
XXXXXX	$20^6 =$	64 000 000
XXXXXXX	$20^7 =$	1 280 000 000

Fig. 1 Number of possible permutations for a random peptide library constructed with 20 eukaryotic amino acids.

of peptide libraries. By using three- to four-fold excess of activated amino acids, the yield of each coupling cycle is usually in excess of 99%. Occasionally, double coupling may be needed if the first coupling cycle is not complete. In general, there are two methods of synthesizing a random library: the "split-mix synthesis" method (Fig. 2) and the "mixed synthesis" method. In 1991, Furka et al., Houghten et al. and Lam et al. described the "split-mix synthesis" method, in which the resin beads are split into several equal aliquots, followed by coupling of each activated amino acid separately with each individual resin aliquot. The reactions are driven to completion as monitored by the ninhydrin test, and then the resins are mixed and washed, and the N_α-amino groups are deprotected, washed, and split into several equal aliquots again. The process is repeated for multiple cycles until the desired peptide length has been attained. In the "mixed synthesis," all the N_α-protected amino acids, which are usually concentration adjusted according to predetermined reactive rates and are added concurrently to the resin. It has been claimed that almost equimolar ratio mixtures of individual peptides can be obtained with this method.

Peptide libraries can contain linear or cyclic structures. Linear peptides usually are very flexible. In certain cases, one may prefer a more rigid or constrained conformation, such as a cyclic structure, to enhance biologic activity. Cyclization of these synthetic peptide libraries can be accomplished by disulfide formation, β-lactam formation, thioether bond formation, alkylation, or metal chelation. For example, side chains of cysteine, homocysteine or penicillamine can be used for disulfide formation. Side chains

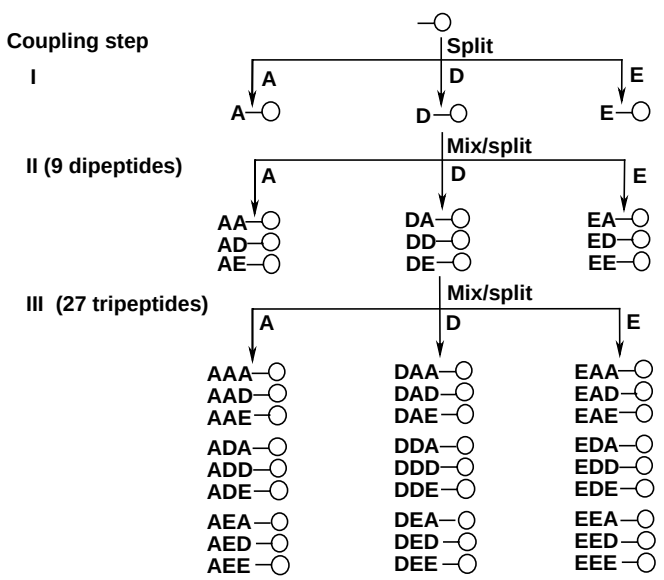

Fig. 2 Synthetic scheme of the "split-mix synthesis" method to generate a one-bead one-compound combinatorial library. A, D, and E are building blocks (in this case amino acids). The circles represent the resin beads.

of lysine or ornithine and aspartic acid or glutamic acid can be used for β-lactam formation. γ-carboxylglutamic acid side chains can be used for Ca^{2+} chelation, and histidine side chains for Ni^{2+} or Zn^{2+} chelation. Proline (L- or D-enantiomers) or other amino acids able to mimic turns (e.g. (3S,6S,9R)-2-Oxo-3-amino-7-thia-1-azabicyclo[4.3.0]nonane-9-carboxylic acid, a commercially available β-turn inducing amino acid) have also been placed in the middle of a peptide to promote a turn structure. When synthetic peptide chemistry is used, natural amino acids and their D-enantiomers, unnatural amino acids including non-α amino acids (e.g. β- or γ-amino acids), amino acid derivatives from side-chain derivatization, and *N*-methylated amino acids can be incorporated into the peptide libraries. In addition, amino acids with saccharide or nucleotide side chains can be used to introduce carbohydrate or nucleic acid features into the random libraries. Furthermore, compounds with reduced peptide bonds can be constructed. In some cases, an *N*-acetylated peptide library may be preferred.

An important feature for the effective use of a peptide library is the screening method. When the peptide library has been synthesized, the solid-phase bound peptides, such as the "one-bead one-compound" (OBOC) combinatorial library generated by "split-mix synthesis" (Fig. 2), can then be screened directly in an ultra high-throughput manner. Millions of peptides can be screened in parallel for a specific acceptor molecule (receptor, antibody, enzyme, virus, whole cell, etc.) within a day or two concurrently, and only small amount of acceptor is required. With various screening techniques, compound beads with a specific biological, chemical, or physical property can be rapidly identified. Direct binding assays have been commonly used to screen the solid-phase bound peptide library. These peptides are often covalently linked to the solid support. Detection is usually accomplished by an enzyme-linked assay, fluorescent assay, or radionuclide assay. Direct functional assays have also been developed including covalent modification of peptide beads by enzymes (e.g. phosphorylation by protein kinases with [γ-^{32}P]ATP) and proteolytic cleavage of bead-bound peptide by proteases.

In some peptide library methods, the peptides can also be cleaved off the solid support and screened by a solution-phase assay if a cleavable linker is used. The major advantage of this approach is that various standardized receptor-binding, biochemical, or cellular assays can easily be adapted for screening. Purified acceptor molecules are often not needed, and a biochemical or biologic end point may sometimes be measured even if the molecular target is unknown. However, the solution-phase assay, in general, is more time consuming, and some of the ligands may be insoluble in the assay buffer and therefore not detected.

1.3 Synthetic Nonpeptide Oligomer or Small Molecule Libraries

Since the advent of combinatorial peptide libraries, nonpeptide oligomer or small molecule libraries have emerged as an exciting alternative development in the combinatorial library field. This has drawn special attention from the pharmaceutical industry, since ligands isolated from such libraries are often smaller than peptides and may contain lipophilic, aromatic, or heterocyclic structures. Therefore, ligands

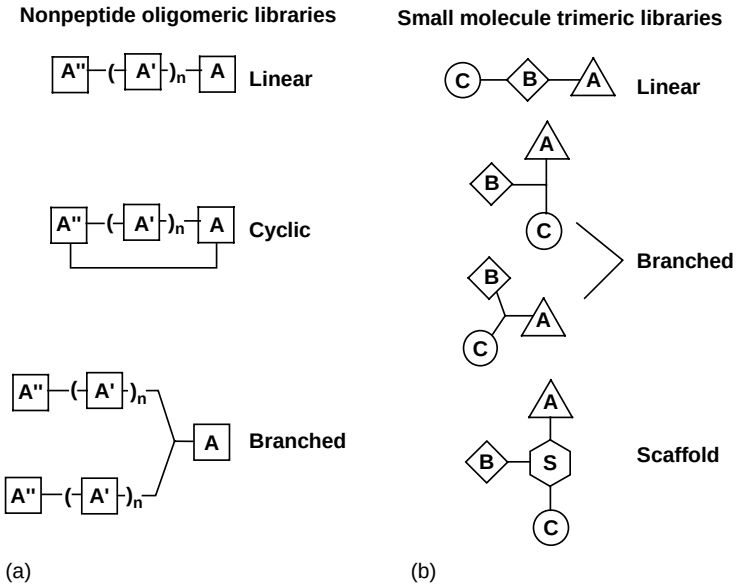

(c)

# Subunits/coupling step	# Permutations for a trimeric library		
20	20^3	=	8 000
50	50^3	=	125 000
100	100^3	=	1 000 000
500	500^3	=	125 000 000

Fig. 3 Nonpeptide oligomeric and small molecule combinatorial libraries. (a) and (b) represent different form of assembly of synthetic oligomeric and small molecule libraries and (c) number of possible permutations for a trimeric library. A, B, and C, building blocks; S, scaffold.

identified from such combinatorial libraries are more likely to be able to cross cell membranes, making them ideal candidates for drug development. Ideally, the small molecule library should have a molecular weight of 300 to 500 dalton. Usually, this translates to a trimeric library. Fig. 3 illustrates the basic principle of a nonpeptide oligomeric library (Fig. 3a) and a small molecule trimeric library (Fig. 3b). For the small molecule trimeric library, the building blocks can be assembled in three different ways: linear (similar to a peptide), branch, and via a scaffold. For the latter approach, depending on the size, rigidity, and chemical nature of the scaffold, the presentation of the randomized subunits can be vastly different. Fig. 3c shows the number of possible permutations available in a random trimeric library. For example, even with 100 possible subunits per coupling cycle, a trimeric library will yield a million possible permutations. Currently, a few thousand carboxylic acids, primary or secondary amines, alcohols, acid halids, sulfonyl halids, alkyl halids, isocyanates, or aldehydes are commercially available. They are all potential subunits for the generation of small molecule libraries.

Because of the high coupling efficiency of solid-phase peptide synthesis, synthetic peptide libraries were first used in the

synthetic combinatorial library field. For the nonpeptide library synthesis, amide bond formation has been commonly used because of its high efficiency of coupling. Over the past decade, many other reactions have also been successfully applied to the library format, such as urea formation, reductive alkylation, *N*- and *O*-alkylation, Wittig reaction, Mitsunobu reaction, 1,3-dipolar cycloaddition reaction, Diels–Alder reaction, Michael addition, Aldol condensation, Knoevenagel condensation, Stille reaction, Ugi reaction, Suzuki coupling, and Mannich reaction. Undoubtedly, more and more new efficient coupling strategies will continue to be developed for the library format in the future.

2 Techniques

In general, there are five methods of preparing and screening combinatorial libraries: (i) the biological peptide library method (e.g. filamentous phage), (ii) the spatially addressable parallel library method, (iii) the affinity selection method, (iv) combinatorial library methods requiring deconvolution, and (v) the combinatorial library method based on the OBOC concept. Although amino acids are used throughout the following discussion, other nonamino acid moieties can also be similarly used in the nonpeptide oligomer or small molecule libraries. Features of various library methods are compared in Tab. 1.

2.1 Biological Peptide Library Method

Biological peptide libraries include filamentous phage-display peptide libraries, plasmid peptide libraries, polysome peptide libraries, fimbria libraries, yeast-display libraries, bacterial flagella libraries, and other more recently developed *in vitro* selection methods.

The filamentous phage approach was first introduced by Parmley and Smith in 1988. It involves the insertion of a stretch of random deoxyoligonucleotide into the pIII gene of filamentous phage, resulting in the expression of a random oligopeptide at the amino terminus of the viral pIII coat protein. Each phage particle has on one end five copies of these grafted coat proteins. The phages that bind specifically to an acceptor molecule will then be enriched and isolated via several cycles of panning (e.g. binding to acceptor coated petri dish) and amplification. Finally, the amino acid sequence of the grafted peptide is determined by sequencing the pIII gene of the isolated phage. This very powerful method enables the researcher to routinely generate 10^8 to 10^9 different phages. Furthermore, the size of the grafted peptide or protein is not limited by the constraints of synthetic chemistry, as in the case of the synthetic peptide library. In addition, the library can be easily generated by simply growing the microorganisms. Moreover, the biological library method can take advantage of known protein folds (e.g. immunoglobulin fold, zinc-finger fold, or conotoxin fold) by grafting random oligopeptides on such tertiary folds. However, this biologic approach suffers from three major disadvantages:

1. Only the natural L-amino acids (20 eukaryotic amino acids) can be incorporated into these libraries, and incorporation of unnatural amino acids or other organic subunits into these libraries is generally not feasible.
2. Although simple disulfide cyclization is feasible, and longer peptides with

Tab. 1 Combinatorial library methods: A comparison.

Combinatorial library methods		*Size of library*	*Constrained structure*	*Binding assay*	*Solution assay*	*Screening approach*	*Cost/Resource*	*Multiple use*
Biological library[a]	Phage display	10^7–10^9	Simple disulfide or protein fold	Yes	No	Parallel	Inexpensive; library commercially available, and from many research laboratories	Library can be expanded, aliquoted, and stored frozen
Spatially addressable parallel library	Multipin technology	10^4	Yes	Yes	Yes	Parallel	Moderately expensive pin and crowns; commercially available	Pins can be used a few times without obvious lost of activity; solution peptides afford multiple assays
	SPOTs-membrane	10^4	Yes	Yes	No	Parallel	Moderately expensive; spot-synthesis equipment and custom- spot peptide membrane commercially available	Peptide spot membrane generally not recyclable for subsequent use
	NanoKan technology	10^4	Yes	No	Yes	Parallel	Very expensive equipment and supplies commercially available	Solution compounds can be used for multiple assays
	Chemical microarray	$<10^5$	Yes	Yes	No	Parallel	Moderately expensive; peptide chip or compound chip not yet commercially available	Replicates of compound chips can be made; compound chip generally not recyclable for subsequent use

Synthetic library methods requiring deconvolution	Iterative approach Positional scanning (PS-SCL)	10^6-10^8 10^6-10^8	Yes Yes	Yes Yes	Yes Yes	Convergent[b] Convergent	Not very expensive; library not commercially available; can be easily synthesized by experienced synthetic organic chemist	Positional scanning libraries, once made, can be aliquoted and used for multiple assays
OBOC library	On-bead screening Releasable assay	10^6-10^8 10^6-10^8	Yes Yes	Yes No	No Yes	Parallel Parallel	Not very expensive; library not commercially available; can be easily synthesized by experienced synthetic organic chemist	In principle, on-bead screening library can be recycled for subsequent screenings, but generally not done because library is not too expensive to make
Synthetic library method using affinity chromatography selection		10^2-10^5[c]	Yes	Yes	No	Parallel	Not very expensive; library not commercially available; can be easily synthesized by experienced synthetic organic chemist	Possibly

[a]In general, only the 20 eukaryotic L-amino acids are used in the biological libraries; unnatural amino acids or other organic building blocks can be used for synthetic combinatorial libraries.
[b]With the multistep iterative process, usually only one solution motif will be identified.
[c]In rare exception, a larger library may be used with some success.

specific protein folds (such as immunoglobulin folds) may be used, complicated bicyclic, compact scaffolding, branched structures, or molecules with special chemistry of cyclization are impossible.
3. The screening assays of the biological libraries are generally limited to the binding assays (e.g. panning) and some functional assays such as protease substrate determination.

Synthetic peptide libraries, on the other hand, can generally overcome many of the limitations listed.

2.2 Spatially Addressable Parallel Library Method

Sometimes a collection of peptides is synthesized on solid-phase support in a spatially addressable format. The amino acid sequence of each of these peptides is predetermined. The peptide collection is screened, depending on the library method used, either by a direct solid-phase binding assay or by a solution-phase assay. The positive peptide is then located and its chemical structure determined. The major drawback of this method is that only limited number of peptides can be synthesized and therefore the library is very small. There are four reported techniques based on this strategy: multipin technology, SPOT synthesis, NanoKan, and chemical microarray.

2.2.1 Multipin Technology

In the multipin method, pioneered by Geysen et al., the solid-phase support is a polyethylene pin in a 96-well format. The polyethylene pins are first "functionalized" with polyacrylic acid or polymethacrylic acid via irradiation. Each pin is inserted into an adapter that fits over a 96-well polypropylene plate with standard microtiter plate footprint. Each well of the microtiter plate then serves as a separate reaction vessel for the amino acid–coupling steps. In the original publication, analysis was performed by enzyme-linked immunosorbent assay (ELISA) with pin-bound peptides, but later modifications permitted peptides to be released by introducing enzymatic or chemically cleavable linkers and analyzed by solution-phase assay. The multipin system is now commercially available from Mimotopes (San Diego, CA) as Pepsets. Peptide loading of each pin has been increased significantly by introducing the so-called "SynPhase™ Lantern" or "SynPhase™ Crown" to fit into the tip of each pin.

2.2.2 SPOT Synthesis

Frank synthesized peptides onto cellulose membrane or paper (e.g. Whatman 540 paper) by applying the N_α-Fmoc amino acids and reagents directly onto the sheet in the form of a spot. In this SPOT-synthesis technique, different peptides are synthesized at different locations on a single sheet of cellulose paper. The volumes dispensed create a specific SPOT size determining both the scale of reaction and the absolute number of compounds that can be arranged on an area of a membrane. The compounds synthesized can be evaluated using conventional high-throughput screening techniques while still attached to the membrane or in solution following release of compounds from the membrane. SPOT synthesis is amenable to miniaturization and automation. Automatic machines, such as the Auto-SPOT Robot ASP222 (Intavis Inc., Cologne, Germany), for the synthesis of peptide spots are commercially available. Final analysis of the solid-phase bound peptide is usually

performed by an enzyme-linked colorimetric assay, autoradiographic detection of the bound or covalently linked radionuclide, or phosphorylation of protein kinase substrates. Using a larger spot and with releasable linker, the peptides can be released for solution-phase assays.

Cotton (another form of cellulose) and polystyrene-grafted polyethylene film segments have also been used as solid supports. Recently, polymeric membranes, including hydroxy-functionalized PEG acrylate polypropylene membranes and amino-functionalized ester free PEG methacrylamide polypropylene membranes, have been developed. These new membranes are chemically, mechanically, and thermally more stable and can be used for SPOT synthesis.

2.2.3 NanoKan Technique

In 1985, Houghten first reported the "tea bag" method for simultaneous multipeptide synthesis. In this method, batches of resin are sealed inside labeled, porous polypropylene packets. In 1995, the use of mini-baskets (e.g. NanoKan) to encapsulate radio-frequency microchips (R_f tagging) and a sample of resin beads for peptide synthesis was reported and became commercially available as IRORI NanoKan™ system (Discovery Partners International Inc, San Diego, CA). Each of these mini-baskets is scanned with an electronic reader prior to or right after each coupling cycle during the "split-mix" synthesis (Fig. 2). Therefore, the synthetic history of each compound in each mini-basket is recorded. At the end of the synthesis, resins in each basket contain a single compound. The compounds are cleaved from the beads inside each mini-basket and placed in a 96-well plate to form a spatially addressable compound library. By utilizing the power of the "split-mix" strategy, a few thousand to ten thousands individual compounds can be easily generated. The main advantages of this method are (i) it offers considerable synthetic flexibility, as one may use any resin beads in the synthesis, (ii) peptides or small molecule compounds can be produced in sufficient quantities (1–10 μmol) for purification and complete characterization if desired. However, the IRORI system is very expensive. A less-expensive encoding system, called *Encore* (Torviq Inc., Granger, IN), uses "necklace" color tags and multicolor reaction vessels to encode the synthetic history. This system also uses the "split-mix" synthesis strategy to generate compound libraries with limited number of compounds.

2.2.4 Chemical Microarray

Chemical microarrays, which predate DNA microarrays, are usually prepared by printing protein, peptide, carbohydrates, small molecule, and other biological molecules on a solid surface. In addition to glass slides, cellulose sheets, polymer-based membranes, or other materials have been used. Glass coated with polymers or dendrimers has also been reported. Chemical microarrays enable investigators to rapidly analyze, in parallel, molecular interactions between immobilized molecules and complex biological mixtures. Screening is accomplished by a binding assay with fluorescent-labeled acceptor molecules or by using a functional assay. They have been widely applied in the field of diagnostics and proteomics. For high-density microarrays (e.g. 2000–3000 compounds per cm^2), the radiographic detection method is probably not practical because the resolution of such detection methods is generally low. Solution-phase assays on microarrays is

difficult, but potentially can be developed. The immobilized molecules on chemical microarrays are generally addressable. Therefore, the chemical identity of each immobilized molecule is known. Many chemical methods for immobilizing these molecules onto solid support have been developed. The three general methods to prepare chemical microarrays are physical adsorption, *in situ* synthesis, and chemical ligation.

Physical adsorption Proteins and macromolecules can be readily immobilized onto a polystyrene surface and nitrocellulose or polyvinylidene difluoride (PVDF) membranes through physical adsorption (nonspecific adsorption). This approach has been used in standard ELISA, dot blot, and Western blot for many years. However, immobilization of small molecules, carbohydrates, or short peptides often requires covalent ligation to the solid surface, unless these molecules are first ligated to a macromolecule prior to adsorption onto the solid support.

In situ synthesis In general, there are three methods for *in situ* synthesis of chemical microarrays: (i) SPOT synthesis, (ii) light-directed parallel synthesis with photolithographic masking methods, and (iii) maskless light-directed synthesis using a digital micromirror array. SPOT synthesis, as described above, uses the standard solid-phase peptide synthesis method to generate low-density peptide arrays (e.g. 25 spots/cm^2). The latter two methods use advanced instruments and light-directed *in situ* synthesis to generate very high-density microarray. In 1991, Fodor et al. reported a photolithographic masking process in conjunction with an amino acid–coupling strategy, which uses photolabile protecting groups (e.g. 6-nitroveratryloxycarbonyl) to synthesize a minute quantity of a total of 1024 peptides on a single glass slide. Each peptide occupies an area of 50 μm × 50 μm. Combinatorial masking strategies are used to decrease the required number of photochemical coupling steps. The method was later improved by using a photolithographic masking method to pattern a polymeric photoresist film on the glass surface. In 1999, the development of a maskless array synthesizer to generate *in situ* oligonucleotide microarrays containing more than 76 000 features, each measuring 16 μm^2, was reported. Independently, Gao and coworkers have reported a similar maskless microarray synthesizer to generate oligonucleotide and peptide microarrays using common building blocks with acid labile protection groups. Deprotection of the amino acids is accomplished by the use of light to generate acid, which allows the growth of the oligomers *in situ*.

One major drawback of *in situ* synthesis is that the available chemistries for such synthesis are more limited. In addition, the efficiency of amino acid coupling using this photochemistry is reported to range between 85 and 95%. The synthetic yield and purity of final products would therefore be a major problem for the synthesis of longer peptide libraries (e.g. 10 to 15 mer). In contrast, chemical ligation of presynthesized products by a commercially available automatic arrayer onto the surface through a covalent bond is generally more efficient and less expensive because each compound needs to be synthesized only once and multiple replicates can be produced simply by spotting.

Chemical ligation For most of the peptide and small molecule chemical microarrays,

the compounds are synthesized as stock solutions prior to immobilization. In the past decade, several combinatorial synthesis techniques have been developed for parallel synthesis of peptides or small molecules. These include the Multipin/lantern system, multisyringe system and the 96 deep-well plate system. These methods are labor-intensive but can be facilitated by robotics. An alternative, more efficient approach is to synthesize large number of compounds using macrobeads (up to 0.1 μmol per bead), bead aggregates (1–10 μmol per aggregate, see below), or the IRORI Nanokans (1–10 μmol per nanokan).

Various chemical ligation methods have been developed to immobilize short peptides or small molecules to the solid support via covalent linkage. Schreiber's group linked thiol-containing compounds to maleimide-derivatized glass slides, via Michael addition, to form a microarray of small molecules. They also reported the formation of ether bonds between the chlorinated glass and hydroxy groups of small molecules to prepare microarrays. Recently, they used diazobenzilidene-functionalized glass slides for immobilization of compounds containing acidic protons, such as carboxylic acids, phenols, and sulfonamides. Rogers and coworkers described the use of disulfide bond formation to immobilize oligonucleotides on thiol-derivatized glass surfaces. The mild reaction of an aminooxy group with ketone and aldehyde has been widely used for chemical microarrays. We took advantage of this highly selective chemoselective ligation reaction to attach unprotected peptides covalently onto the glyoxylyl derivatized glass slide via oxime bond or thiazolidine ring formation. Very recently, we reported an efficient method to prepare glyoxylyl-coated glass surface by coupling acrylic acid to the amino glass surface, followed by oxidation with $NaIO_4/OsO_4$ to form a glyoxylyl group. Salisbury et al. reported a similar approach to make chemical microarray by immobilizing aminooxy-derivatized peptides to an aldehyde-modified glass surface. Houseman et al. reported the ligation of carbohydrates to a gold surface via Diels–Alder reaction. Lesaicherre et al. employed a different chemoselective ligation approach to immobilize C-terminal thioester-modified peptides onto an amino modified glass. This microarray was subsequently used for antibody-based fluorescence detection of kinase activity. Soellner et al. employed the Staudineger reaction to ligate the azide-modified peptide to phospinothioester-functionalized glass surface to form an amide bond, while Kohn et al. used this reaction to immobilize small molecule onto a phosphane-decorated glass slide. Kanoh et al. recently reported a "nonselective" immobilization method using the chemistry developed for photoaffinity labeling. In this method, they first functionalized the glass surface with a photoreactive group, such as diazirine. A small molecule was then spotted, and the coupling reaction was initiated by UV irradiation.

2.3 Affinity Selection

Affinity column chromatography has been applied successfully for combinatorial oligodeoxynucleotide library methods in which a large collection of oligodeoxynucleotides is allowed to pass through an acceptor-bound column (e.g. thrombin-Sepharose), and the bound oligodeoxynucleotides are eluted and amplified. After a

few more cycles of binding, elution, and enrichment, the structure of the oligonucleotide is finally determined. However, this affinity technique has been applied to a peptide library only with limited success. The reason is that peptides cannot be amplified or cloned. To have enough material for structure determination, the retrieved peptide generally must be pure and in sufficient quantity (at least a picomole, or 6×10^{11} molecules). Nonspecific binding becomes the major problem if a large library of peptides is used. There is no simple method of isolating a single peptide for sequencing. Despite its limitation, Zuckermann et al. were able to apply the affinity selection method to retrieve an antibody-specific binding peptide from a mixture of peptides in solution. Cantley et al. reported the successful application of the affinity selection method to identify peptide motifs for SH_2 domains and kinase domains of protein tyrosine kinases. Concurrent microsequencing of the retrieved peptide mixture was performed rather than sequencing individual peptides. Therefore, the data could be difficult to interpret and inconclusive, particularly in the presence of more than one motif. In general, because of high background from nonspecific binding, the affinity selection method, with rare exceptions, can be applied only to a relatively small peptide library (e.g. $<$10 000 peptides).

2.4 Library Methods Requiring Deconvolution

In this method, mixtures of peptides are first synthesized and then subjected to biological testing. On the basis of the test results, one may be able to deduce the chemical structure of the active ligand without need of further structure elucidation. However, in general, additional synthesis and testing are needed to reach a final solution. Examples of this library approach are the iterative approach, the positional scanning approach, the orthogonal partition approach, the recursive deconvolution approach, and the recently developed dual recursive deconvolution.

Fig. 4 shows how the amino acid at each position of the ligand can be defined by an iterative process that involves multistep synthesis and analysis. For example, in a combinatorial hexapeptide library, one of the 20 amino acids is incorporated into positions 1 and 2, while all 20 amino acids are incorporated (presumably in an equimolar ratio) into the remaining four residues. This results in the generation of 400 (or 20×20) peptide mixtures (step 1, Fig. 4). Each mixture contains 160 000 (or 20^4) different peptides. Each of these 400 peptide mixtures is then subjected to screening, from which the "active" amino acids for the first two positions can be defined. In step 2, the first two "active" residues are fixed, and one of the 20 amino acids is incorporated into position 3, while all 20 amino acids are incorporated into the remaining three residues. This results in the generation of 20 peptide mixtures. Each of these 20 mixtures is again screened, and the third set of "active" residues is then defined. The same iterative process is repeated for three more cycles until the last residue has been defined.

A synthetic combinatorial library constructed and screened by means of an iterative approach was first described by Geysen in the multipin system, where a peptide mimotope was defined for a monoclonal antibody. The pin-bound peptide library was screened by means of ELISA. Subsequently, Houghten et al. described a similar method termed as synthetic peptide combinatorial libraries (SPCLs), but the peptides were released into solution

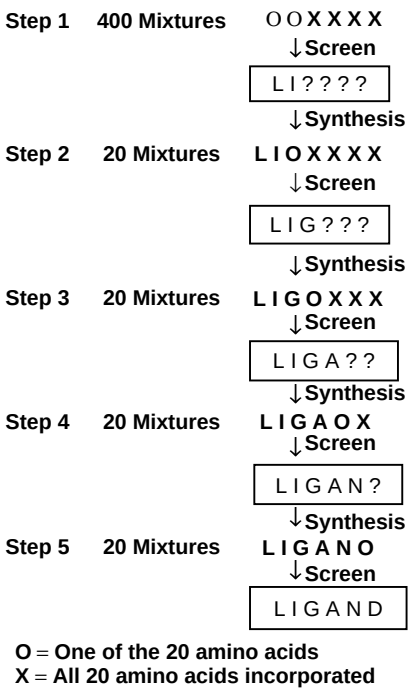

Fig. 4 Scheme of the iterative approach for combinatorial library method that requires deconvolution.

prior to screening. The libraries are prepared using 4-methylbenzhydrylamnine (MBHA) polystyrene resin and standard *t*-Boc/Bn chemistry. To ensure equimolar random mixtures, a "split-mix synthesis" method (see Fig. 2) is used. The peptide mixtures are deprotected and cleaved from the resin with the low/high hydrogen fluoride method. Peptide mixtures are then extracted with water or dilute acetic acid and lyophilized. The major advantages of the SPCL approach are that a large number (10^6–10^8) of peptides can be synthesized and screened. Also, many existing biological assays, including functional assays, can be adapted to this screening process (because it is a solution-phase assay). One concern is that not all the peptides in the library are extractable by water or dilute acetic acid.

Subsequently, Houghten et al. described the "positional scanning methodology" in which 20 mixtures of hexapeptides were synthesized with one of the 20 amino acids incorporated into one residue, while all 20 amino acids were incorporated into the rest of the five positions. By scanning these 120 mixtures, it is possible to define each active residue. However, this approach will work well for acceptor molecules for which no more than one motif is possible. If an acceptor has multiple motifs, the results will likely be scrambled and sometimes uninterpretable.

Houghten et al. also described a so-called "libraries from libraries method," in which the random peptide libraries are chemically modified, while the protected peptides were still attached to the solid phase. Examples of such chemical transformation are permethylation, perbenzylation, perallylation, and decarbonylation. The modified libraries are then released for solution-phase testing.

Deprez et al. introduced the orthogonal partition approach of deconvolution by synthesizing and screening multiple orthogonal sub-libraries. Erb et al. reported a recursive deconvolution strategy. In this approach, a "split-mix synthesis" method is used to generate the combinational library. During each step of synthesis, a portion of the resin from each reaction vial is set aside and catalogued. These set-aside resins are then used to generate sub-libraries at each iterative step in a more efficient manner.

Dual recursive deconvolution (DRED) was introduced by Fenniri et al. as a hybrid approach of deconvolution. In this method, a new (encoding) dimension was added to the classical deconvolution strategies. DRED operates through the

iterative identification of the first and last randomized positions of active members. The last position of the combinatorial library can be readily obtained from pool screening after the last step of the split-mix synthesis, while the first position need to be encoded by the unique vibrational fingerprint of the resin beads used. Once the first and last positions are identified, the remaining positions are subjected to the same deconvolution process.

2.5 OBOC Library Method

We first reported the OBOC concept in 1991. The peptide is synthesized on solid-phase beads using a "split-mix synthesis" method (Fig. 2). Since each bead is in contact with only one amino acid at a time during each coupling cycle and the reaction is driven to completion, each bead expresses a single peptide entity (Fig. 2). The peptide library, consisting of millions of beads, each with a unique peptide, is then subjected to mass screening. The reactive beads are identified and physically isolated, and each is subjected to microsequencing with Edman degradation chemistry. Each 80 to 100 μm bead contains approximately 100 pmol of peptide (6×10^{13} molecules) and is more than enough for microsequencing. If there is a predominant common motif, instead of sequencing individual bead, 10 to 50 "active" beads can be pooled for concurrent microsequencing. The methods of screening include both on-bead binding assays and solution-phase biological assays.

TentaGel resin (Rapp Polymere, Tubingen, Germany) is a good choice for OBOC peptide libraries because of its uniformity in size as well as its nonstickiness. It consists of hydroxyethyl-polystyrene beads onto which have been grafted 3000 to 4000 dalton PEG linkers. A functional group, for example, hydroxyl or amino, is located at the end of the linkers. This resin can be swollen in a wide range of solvents from water to toluene. The PEG linker is advantageous for many biological assays. For protease substrate screening, porous beads such as PEGA [bis(2-acrylamidoprop-1-yl) poly(ethylene glycol) cross-linked dimethyl acrylamide and mono-2-acrylamidoprop-1-yl [2-aminoprop-1-yl] poly(ethylene glycol)] beads (Calbiochem-Novabiochem, San Diego) are preferable.

For the on-bead binding assays, standard enzyme-linked colorimetric assays, fluorescent assays, or radionuclide assays can be used. The enzyme-linked colorimetric assay is the simplest because it adapts the standard Western blot technology and no special equipment is required. Fluorescent microscopy has been used successfully to screen bead libraries with fluorescent probes. However, the greatest potential using a fluorescent assay to analyze a library is to use a fluorescent-activated bead sorter (COPAS™ BIOBEAD, Union Biometrica, Inc, Somerville, MA). In a standard analysis, 30 000 beads can be sorted per hour.

The "on-bead" binding assay was modified to identify protein kinase substrate motifs by covalently modifying the peptide bead with [γ-^{32}P] ATP and a protein kinase. The phosphorylated peptide-bead library is washed thoroughly and immobilized on a glass plate with agar. The ^{32}P-labeled beads are then detected by autoradiography, and individual radiolabeled ("active") beads are isolated for microsequencing. Another screening approach is to use a color probe to screen a library, to take advantage of the ease with which positive beads may be identified by the color change. For example, we have identified small peptides that bind specifically to an indigo dye.

The OBOC library method has been successfully applied to identification of protease substrates and inhibitors. Highly porous PEGA resin was used for the peptide library construction because it allowed the enzyme to gain access to the bead interior. For protease substrate determination, OBOC peptide libraries with fluorescent dye incorporated at the C terminus and a quencher at the N terminus have been used. Peptide beads susceptible to protease cleavage will fluoresce. These fluorescent-labeled beads were isolated under a fluorescent microscope, or isolated by a fluorescent-activated bead sorter and subjected to microsequencing. This method is a powerful tool to investigate the protease activity and specificity since it gives information about the substrate sequence, the cleavage point, and the degree of cleavage. Meldal et al. reported a so-called ''one-bead two-compounds'' approach to directly identify protease inhibitors. This approach combines the OBOC concept and fluorogenic quench screening strategy. In this method, each bead contains a putative protease inhibitor along with a fluorescence-quenched substrate for the protease. Kodadek et al. recently reported the use of fluorescent-labeled proteins as probe to screen OBOC peptoid (*N*-substituted oligoglycine) libraries. In order to increase the signal-to-noise ratio, they later reported the use of quantum dots, rather than organic fluorescent dye, as a fluorophor to label the target protein.

The OBOC method has also been applied to the identification of cell surface binding ligands, using a whole-cell binding assay. In this assay, peptide-bead library is mixed with intact cells, and beads coated with a monolayer of cells are isolated for microsequencing.

To minimize nonspecific binding, blocking agents (e.g. bovine serum albumin or gelatin), salt (e.g. 0.26 M NaCl), and nonionic detergent (e.g. Tween-20) are routinely added to the screening buffer. Furthermore, with a dual color detection scheme in the presence or absence of a competing ligand, most of the false positive beads can be eliminated, thus increasing the efficiency of microsequencing, which is the rate-limiting step of the process.

Besides binding assays, the OBOC method has been adapted to solution-phase assays. Peptide bead libraries with dual orthogonal cleavable linkers are used, and peptides can be released sequentially with a two-step process into solution for biological testing. Conditions for the release of peptides from the resin beads are mild and nondenaturing. They involve neutralization in step 1, and alkali treatment followed by neutralization in step 2. In this two-step process, approximately 500 peptide beads are aliquoted into each well of a microtiter filtration plate (96-well format). Upon neutralization, the peptides are released from the first linker into solution for solution-phase assays (step 1). Peptide beads from the positive wells are then redistributed at the much lower density of one bead per well. Upon alkali treatment, peptides are released from the second linker. After neutralization, the released peptides are subjected to solution-phase assays. Thus, similar to the solution-phase combinatorial library approach described by Houghten, the screening can be adapted to many existing biological assays. Positives can be tracked to the originating bead, allowing the residual peptide on the bead to be analyzed by a microsequencer. Depending on the size and loading of each bead, 1 to 10 μM compound (in 100 μL assay volume) can be released for standard solution-phase biological assays. However, this approach suffers some disadvantages,

for example, it is not easy to retrieve minute amounts of soluble compounds (100 pmol) from one single 80 to 100 μm bead, and often there is not enough material from one single bead for biological assays. Schreiber et al. reported the use of macrobeads (500 μm) to develop releasable OBOC libraries in microtiter plates. However, the capacity of these macrobeads is still rather limited (<0.1 μmol/bead). Further increases in size of the bead lead to incomplete coupling because reagent diffusion will become a limiting factor. We have recently developed a "one-aggregate one-compound" method to address this problem. Bead aggregates were prepared by cross-linking the TentaGel resin beads with glutaraldehyde. Each bead aggregate contains some color beads with an orthogonal protecting group that can be chemically encoded during the library synthesis and retrieved for decoding after biological screening. The loading capacity of each bead aggregate ranges between 1 and 10 μmol, for example, 10 to 100-fold more than that of the macrobeads used by Schreiber's group. By exploiting the OBOC concept, this "one-aggregate one-compound" method is highly efficient and much more economical than the IRORI Nanokan® system. Such diverse solution-phase libraries can be either screened with the standard solution-phase assay or used to print multiple replicates of chemical microarrays (see Sect. 2.2.4).

A more elegant and potentially much more powerful approach to screen the OBOC libraries is the *in situ* solution-phase releasable assay, in which the compound-bead libraries are immobilized in a thin layer of agar. Compounds from each bead are then released to the vicinity of each bead for solution-phase assay in the semisolid. This method works particularly well for identifying antimicrobial and anticancer agents, since a zone of growth inhibition around the positive beads can easily be detected. We have used this *in situ* releasable assay to identify anticancer agents. The compound beads were immobilized in a thin layer of soft agar together with cancer cells. After being cleaved from the linker, compounds from each bead were released and diffused outward. A clear zone of growth inhibition around the positive bead was detected by adding 3-(4,5-dimethylthazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) to the culture plate, in which only the live cells were stained purple. Jayawickreme et al. reported a "cell-based lawn format" utilizing an *in situ* photocleavage method to release the compound. These *in situ* releasable solution-phase assays have great potential but will require further development before they can be reliably used for drug screening. For example, special solid supports need to be developed such that all compounds will diffuse freely out of all beads into the surrounding media.

The main advantages of the OBOC method are that a large number (10^6–10^8) of compounds can be synthesized and screened rapidly, and either one or a combination of both on-bead binding and solution-phase assays can be used. In addition, since it is a single-step process that does not require deconvolution, multiple peptide ligands with completely different motifs can often be identified in a single screen. That is, the OBOC method is a parallel approach. This is the opposite of the (convergent) iterative approach, where multistep synthesis and screening results in the emergence of only one, but not necessarily the best, solution motif.

Since the compound beads in an OBOC combinatorial library are spatially separable, an OBOC library can be considered

as a huge chemical microarray that is not addressable. Many of the detection methods developed for OBOC combinatorial libraries have also been used to screen chemical microarrays. These include colorimetric assays, fluorescent assays, radioactive assays, and whole-cell binding assays. Other assays that do not require protein labeling, such as surface plasmon resonance, have also been developed.

Positive compound beads from screening of OBOC library are physically isolated under the microscope. For peptide libraries that are comprised of eukaryotic amino acids, an automatic protein sequencer (e.g. Procise 494, Perkin-Elmer/Applied Biosystems) is routinely used via Edman chemistry to determine the amino acid sequence of the positive beads. The incorporation of unnatural amino acids into peptide libraries not only greatly increases the diversity of the library but also renders some of the peptide library members more resistant to proteolysis. We recently developed methods to sequence peptides containing many unnatural α-amino acids. However, Edman sequencing is time consuming and expensive, compared to mass spectrometry analysis. Youngquist et al. described a "ladder sequencing" method by using limited capping after each cycle of coupling followed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) to determine the full peptide sequence of single beads. Later, this method was modified with either replacement of the capping step, by the partial incorporation of a methionine residue at each coupling stage, or by the use of the dual capping groups and analysis of molecular ion redundancy to directly elucidate the structure by mass spectrometry analysis. Very recently, Pastor et al. have further improved this method by using an isotope-labeling strategy to terminate the N terminus and analyzing the generated ladders with ion-extraction mass spectrometry.

2.6 Nonpeptide Oligomer or Small Molecule Libraries

Within two years from the initial publications by Lam and Houghten on screening large synthetic peptide libraries in 1991, several groups reported on the application of similar approaches for the synthesis and screening of nonpeptide oligomeric libraries. The three general strategies for subunit assembly to form linear, cyclic, and branched oligomeric libraries are shown in Fig. 3a. In the case of the linear strategy with repeating subunits, the Chiron group described the "peptoid" approach with N-substituted glycines as subunits. Subsequently, they modified their synthetic method by using α-haloacetic acid submonomers and a variety of primary amines as subbuilding blocks.

Cho et al. described the oligocarbamate libraries using N-protected amino carbonates as the monomers. Because of the repetitive nature of the synthetic process, similar to the peptide library, both the peptoid and oligocarbamate strategies allow the synthesis of a longer linear oligomer. However, as indicated earlier, the main advantage of the nonpeptide library method is that it allows the synthesis and screening of a library of relatively rigid small molecules. These small molecules (e.g. trimeric structures with molecular weight of 300–500) are ideal candidates for drug development because they have improved physical properties for crossing the biological membranes. The subunit assembly strategies for such a library could

involve linear, branch, or scaffold linkage methods.

The scaffolding approaches of presenting several pharmacophores into space have been described by many groups, but Ellman and coworkers were the first to publish on the scaffolding approach in a library format. They described the assembly of a small molecule, 1,4-benzodiazepine derivative library with high yield. Since then, there have been numerous reports on the synthesis of different rigid small molecule libraries. Several thousands of nonpeptide libraries have been reported. These include hydantoin, benzopyran, tetrahydrofuran, acylpiperidine, β-turn mimetics, arylacetic acid, prostaglandin, purine, triazine, quinazoline, phenylamine-pyrimidine, phthalazine, and steroid libraries. Various chemical reactions have also been successfully applied to the synthesis of libraries, including C–C formation reaction. Undoubtedly, many more examples of applications of small molecule libraries will be published in the next few years.

2.7 Encoding and Decoding for Nonpeptide Libraries

If combinatorial nonpeptide or small molecule libraries that require deconvolution are used, structure determination with physicochemical methods is not needed because the structure is deduced from the synthetic algorithm. However, in the OBOC approach, one needs to isolate the individual "active" bead and determine the chemical structure either directly or via an encoding strategy. Since nonpeptide structures are not amenable to structure determination with Edman degradation, alternative strategies have been developed. Several research groups have developed various physical and chemical encoding methods to encode OBOC nonpeptide libraries. Chemical encoding is ideal for OBOC nonpeptide libraries since it can encode large libraries, for example, >10 000 members. Three encoding principles are illustrated in Fig. 5. Chemical tags are added to the bead during the synthetic steps so that the synthetic history of each compound bead in the chemical library can be recorded. These chemical codes can then be decoded by spectroscopic or chromatographic methods such as HPLC (high performance liquid chromatography), GC (gas chromatography), MS (mass spectrometry), fluorescence, IR (infrared), NMR (nuclear magnetic resonance) spectroscopy and electron capture. Several chemical encoding methods have been reported, including using oligonucleotide tags, peptide tags, fluorophenyl ether tags, secondary amine tags, and trityl-based mass tags. A separating anchor with differential protecting groups (Fig. 5a) or a growing chain utilizing separate branches (Fig. 5b) on the solid phase are used for the library synthesis. One protecting group is used for the synthesis of the testing ligand and the other for the coding sequence. For every cycle of the split-mix synthesis, coding subunits are coupled immediately after coupling ligand subunits but before the mixing of the resins. At the end of the library synthesis, each bead has a testing ligand and a corresponding coding sequence. The chemical structure of the isolated ligand is then determined by analyzing the structure of the coding sequence. With this sequential coding approach, the peptide tag can be determined by means of Edman degradation, and the oligonucleotide tag can be determined by polymerase chain reaction followed by DNA sequencing. One-step digital or binary coding systems have been developed.

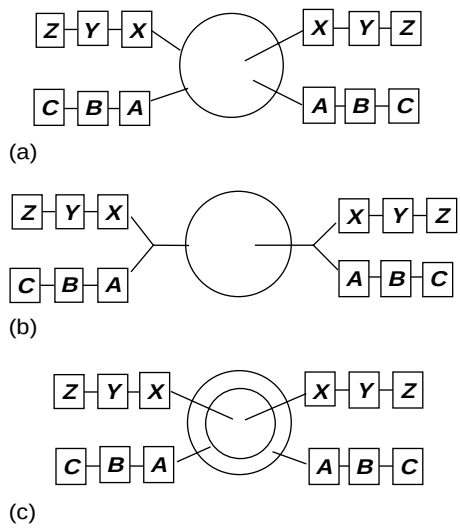

Fig. 5 Principle of encoding for nonpeptide combinatorial library. (a) Sequential coding; (b) Branch coding; (c) Spatial separation of testing and coding molecules. *A*, *B*, and *C* are building blocks; *X*, *Y*, and *Z* are coding tags; *X* encodes *A*, *Y* encodes *B*, and *Z* encodes *C*.

These use specially modified amino acids (analyzed by HPLC) or halophenol derivatives (analyzed by electron capture gas chromatography) that elute at different regions of the chromatogram. The elution range signifies the coding position (residue 1, 2, or 3, etc.), and the binary combination of position with specific peak in the range signifies the subunit used in that position. These methods have already been successfully applied to the synthesis and screening of nonpeptide libraries.

However, these methods have two major limitations: (i) the coding tags may negatively affect the binding of the ligand structure to the target acceptor molecule, and (ii) these chemical encoding methods require that the chemistry of adding the tag and synthesizing the library be orthogonal, resulting in nearly doubling the number of synthetic steps. To solve the problem of interference by the coding tag, we have developed a novel and robust peptide-based encoding strategy. In this method, the testing molecule presents on the bead surface, and the coding tag resides in the interior of the beads (Fig. 5c). This encoding method is highly efficient, as each of the building blocks is incorporated into the testing arm (bead surface) and the coding peptide backbone (bead interior) simultaneously (Fig. 6). This eliminates many extra synthetic steps of standard chemical encoding. The positive beads identified through screening can be isolated, and the peptide coding tags, which consist of α-amino acids with side chains derivatized by the building blocks, can be readily decoded by Edman microsequencing. We have successfully applied this encoding method to a number of peptidomimetic and small molecule libraries. This peptide-encoding method, however, is relatively slow because Edman chemistry for microsequencing is slow. Recently, we have further improved our encoding method by incorporating triple or quadruple cleavable coding arms in the bead interior so that the coding tag can be decoded with MS (Fig. 7). Prior to library synthesis, the inner core of each bead is derivatized with three or four different coding arms on a cleavable linker. Similar to the above-mentioned peptide-based encoding method, each of

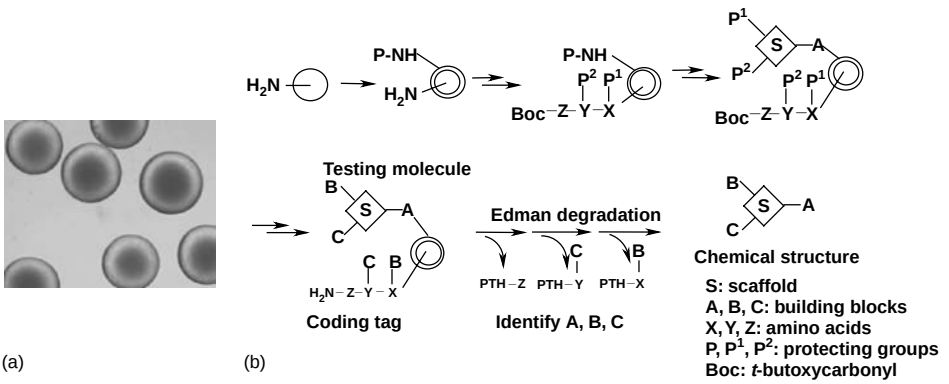

Fig. 6 (a) Photomicrograph of the topographically segregated bifunctional bead. Free amines at the inner core of each bead stained blue because they reacted with bromophenol blue. (b) General encoding and decoding scheme of peptide-encoded small molecule library (see color plate p. xxiii).

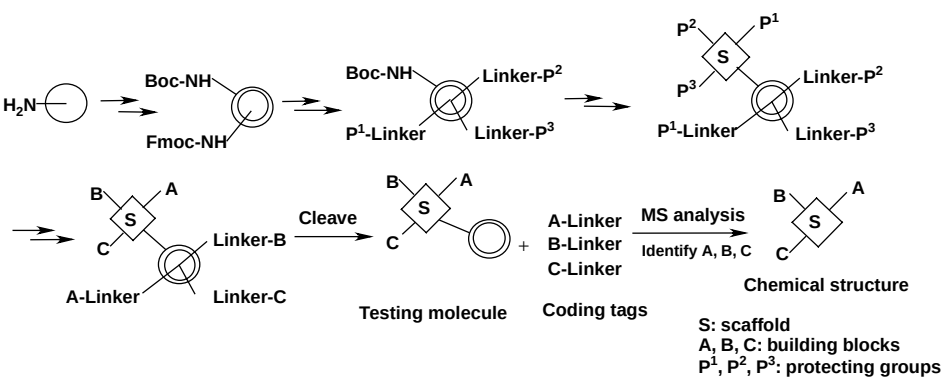

Fig. 7 General encoding and decoding strategy of the MS-encoded small molecule library.

these coding arms contains a functional group that is identical or related to the functional groups on the scaffold of the testing compound to be synthesized. Also, each building block will react with the testing and encoding arms simultaneously. Consequently, no extra synthetic steps for encoding are needed. After screening, the coding tags in the positive beads are released, followed by molecular mass determination using matrix-assisted laser desorption/ionization-Fourier transform mass spectrometry (MALDI-FTMS).

Very recently, we combined the bilayer bead concept with the "ladder-synthesis" method (see above), which has been used successfully in sequencing active beads isolated from small molecule libraries. In this method, the small molecule library is synthesized such that library compound is confined to the outer layer of the bead and its ladder fragment resides in the bead interior. Releasates of these compounds and ladder fragments from the positive bead are analyzed by matrix-assisted laser desorption/ionization mass

spectrometry (MALDI-MS)). On the basis of the mass difference between the library compound and its ladder fragments, the identity of each building block can be determined.

3 Application

Over the last decade, the synthetic combinatorial library approach has been applied successfully to various biologic systems. These include identification of ligands for antibodies (both continuous and discontinuous epitopes), streptavidin, avidin, opioid receptors, melanin-stimulating hormone (MSH) receptor, surface idiotype of B-cell lymphoma cell lines, bombesin receptor, molecules A2 and B7 of major histocompatibility complex (MHC) class I cytokine receptors, signal transduction adaptor molecules such as SH_2 and SH_3 domains, adhesion molecules such as gpllb/Illa, specific metal ions, double-stranded DNA, and organic dye molecules. In addition, peptides with antiprotease (including HIV protease), antiinflammatory, antibacterial, and catalytic activities have been discovered. Furthermore, substrate motifs for various protein kinases, proteases, or other post-translational modification enzymes have been elucidated. Peptide-based enzyme mimics (artificial enzymes) have been identified through screening combinatorial peptide libraries. We reported the use of a ^{32}P or ^{33}P phosphorylation assay and an autoradiographic method to identify specific and efficient peptide substrates for protein kinases. We have also described the use of a whole-cell binding assay in which bead libraries are mixed with live cells, to identify cell surface binding peptide ligands specific for prostate cancer, nonsmall cell lung cancer, and lymphoma cells. Tab. 2 summarizes some of the published applications of combinatorial library methods.

Combinatorial chemistry has also successfully applied to the field of material science. This includes the discovery of polymeric structures with specific physical, chemical, electrochemical, photochemical, or photoelectric properties.

Tab. 2 Biological applications of combinatorial library methods.

Methods	*Application examples*
Biological libraries	MHC class II molecule binding peptides Peptide substrate for tissue plasminogen activator Peptide substrates for HIV-1 protease Epitope mapping for B-cells Cell surface ligands Binding to the vasculature of various organs and tumors by injecting the phage library intravenously into live animals
Spatially addressable parallel library	Epitope mapping for monoclonal and polyclonal antibodies T-cell epitopes MHC class I molecule binding peptides

Tab. 2 (*continued*)

Methods	*Application examples*
	Structure-activity relationship of substance P Protein kinase substrates and inhibitors
Synthetic library methods requiring deconvolution	Mimotopes/epitopes for monoclonal or polyclonal antibodies T-cell epitopes Protease inhibitors Antibacterial agents Opioid receptor ligands Molecules with catalytic activity Molecules with hemolytic activity α-Adrenergic receptor ligand (from peptoid libraries) CAMP-dependent protein kinase motif DNA-binding peptides Tab-2-binding peptides
One-bead one-compound library method	Mimotopes/epitopes for monoclonal antibodies MHC class I binding motif SH_3 domain-binding peptides SH_2 domain-binding peptides GpIIb/IIIa-binding motif Melanin-stimulating hormone-receptor ligands Bombesin-receptor ligands Streptavidin Avidin Factor Xa inhibitor Protease substrates and inhibitors Antibacterial agent CAMP-dependent protein kinase motif $P60^{c-src}$ protein kinase motif Small organic dye-binding peptides Artificial receptors Anti-insulin monoclonal antibody recognizing a discontinuous epitope Surface idiotype of B-cell lymphoma cell lines Ligands for an anti-β-endorphin antibody Ligands for cell surface receptor of live cells Peptides with catalytic activity
Synthetic library method using affinity chromatography selection	Epitope mapping for monoclonal antibodies Protein kinase substrates and inhibitors SH_2 domain-binding peptides Kinase domains of protein tyrosine kinases

Note: Biological libraries are limited to peptide libraries with eukaryotic amino acids. The remaining four methods can be applied to peptide or nonpeptide libraries.

Screening of diverse random compound libraries for specific biological or physical properties often leads to the discovery of lead compounds that have low potency. On the basis of the structural motif of these primary leads or lead compounds determined by other means, secondary or focused libraries may be designed and screened under higher stringency. Compounds identified from such secondary screen often exhibit higher potency. This approach can be repeated several times to further optimize the compound. Although combinatorial chemistry can be applied to biological systems without any known structural data such as X-ray crystal structural of target proteins, structural data do help. For example, one may apply molecular modeling techniques to select appropriate scaffoldings and building blocks so that combinatorial libraries with better ''hit rates'' can be designed. The development of these combinatorial library methods not only facilitates the drug discovery process but also provides important information for the understanding of molecular recognition.

4 Perspectives

Tab. 1 compares five peptide library methods. In general, there are only four methods of generating and screening libraries of 10^6 to 10^8 peptides: biological peptide library, combinatorial library requiring deconvolution, the OBOC library, and the affinity selection method. The spatially addressable parallel library method, while very useful, can only generate limited number of compounds (e.g. less than 10 000). Each method has its advantages and disadvantages, and in every case the choice of methods will depend on the specific application and the resources and expertise available. Sometimes, it may be advantageous to combine some of these methods together. Because of the versatility of synthetic chemistry, enormous research efforts have been given to the development of totally nonpeptide small molecule libraries. New methods for library synthesis, via solid-phase synthesis methods or solution-phase synthesis methods, will continue to be developed. Microwave-assisted reactions will continue to be developed to improve the coupling yields. Novel building blocks and multifunctional scaffoldings will be developed and commercialized. Novel encoding methods for OBOC combinatorial libraries will continue to evolve. New solid supports with various composition, shape, and form will be developed for various combinatorial chemistry applications. As an alternative to synthetic chemistry, enzyme-catalyzed solid-phase biosynthetic methods may also be used. In fact, these biosynthetic methods show great potential in the generation of oligosaccharide or polyketide libraries whose chemical synthesis may otherwise be difficult.

The combinatorial library method represents a powerful technique for basic research and drug discovery. It has also proven to be extremely useful in material and catalyst development. Chemical microarray is a unique form of combinatorial library that is currently being exploited in the field of drug discovery, proteomics, and the development of novel diagnostics. In the next few years, highly sensitive analytical techniques will be developed to quantitate the binding of target proteins to such chemical microarrays and to determine the identity of the bound proteins to such chemical microarrays. We anticipate that combinatorial chemistry will continue

to have a major impact on the field of drug development and chemical biology in the years to come.

See also Combinatorial Phage Antibody Libraries; Molecular Display Technologies.

Bibliography

Books and Reviews

Barnes, C., Balasubramanian, S. (2000) Recent developments in the encoding and deconvolution of combinatorial libraries, *Curr. Opin. Chem. Biol.* **4**, 346–350.

Czarnik, A.W. (1997) Encoding methods for combinatorial chemistry, *Curr. Opin. Chem. Biol.* **1**, 60–66.

Dower, W.J., Mattheakis, L.C. (2002) In vitro selection as a powerful tool for the applied evolution of proteins and peptides, *Curr. Opin. Chem. Biol.* **6**, 390–398.

Fenniri H., (Ed.) (2000) *Combinatorial Chemistry: A practical approach*, Oxford University Press, Oxford, UK.

Lam, K.S., Lebl, M., Krchnak, V. (1997) The "one-bead-one-compound" combinatorial library method, *Chem. Rev.* **97**, 411–448.

Lam, K.S., Liu, R., Miyamoto, S., Lehman, A.L., Tuscano, J.M. (2003) Applications of one-bead one-compound combinatorial libraries and chemical microarrays in signal transduction research, *Acc. Chem. Res.* **36**, 370–377.

Meldal, M. (2002) The one-bead two-compound assay for solid phase screening of combinatorial libraries, *Biopolymers* **66**, 93–100.

Morales, G.A., Bunin, B.A., (Eds.) (2003) *Methods in Enzymology*, Vol. 369 (Combinatorial Chemistry, Part B) Elsevier Academic Press, San Diego, CA.

Westerlund-Wikstrom, B. (2000) Peptide display on bacterial flagella: Principles and applications, *Int. J. Med. Microbiol.* **290**, 223–230.

Xu, Q., Lam, K.S. (2003) Protein and chemical microarrays-powerful tools for proteomics, *J. Biomed. Biotechnol.* **5**, 257–266.

Primary Literature

Beck, Z.Q., Hervio, L., Dawson, P.E., Elder, J.H., Madison, E.L. (2000) Identification of efficiently cleaved substrates for HIV-1 protease using a phage display library and use in inhibitor development, *Virology* **274**, 391–401.

Benters, R., Niemeyer, C.M., Wohrle, D. (2001) Dendrimer-activated-solid supports for nucleic acid and protein microarrays, *Chembiochem* **2**, 686–694.

Blackwell, H.E., Perez, L., Stavenger, R.A., Tallarico, J.A., Eatough, E.C., Foley, M.A., Schreiber, S.L. (2001) A one-bead, one-stock solution approach to chemical genetics: part 1, *Chem. Biol.* **8**, 1167–1182.

Blondelle, S.E., Takahashi, E., Dinh, K.T., Houghten, R.A. (1995) The antimicrobial activity of hexapeptides derived from synthetic combinatorial libraries, *J. Appl. Bacteriol.* **78**, 39–46.

Boder, E.T., Wittrup, K.D. (1997) Yeast surface display for screening combinatorial polypeptide libraries, *Nat. Biotechnol.* **15**, 553–557.

Brown, B.B., Wagner, D.S., Geysen, H.M. (1995) A single-bead decode strategy using electrospray ionization mass spectrometry and a new photolabile linker – 3-amino-3-(2-nitrophenyl)propionic acid, *Mol. Divers.* **1**, 4–12.

Buchholz, C.J., Peng, K.W., Morling, F.J., Zhang, J., Cosset, F.L., Russell, S.J. (1998) In vivo selection of protease cleavage sites from retrovirus display libraries, *Nat. Biotechnol.* **16**, 951–954.

Clemons, P.A., Koehler, A.N., Wagner, B.K., Sprigings, T.G., Spring, D.R., King, R.W., Schreiber, S.L., Foley, M.A. (2001) A one-bead, one-stock solution approach to chemical genetics: part 2, *Chem. Biol.* **8**, 1183–1195.

Cwirla, S.E., Peters, E.A., Barrett, R.W., Dower, W.J. (1990) Peptides on phage – a vast library of peptides for identifying ligands, *Proc. Natl. Acad. Sci. USA* **87**, 6378–6382.

Demers, J.P. (1998) *Spatially Addressable Combinatorial Chemical Arrays in CD-ROM Format*, PCT Int. Appl. WO 9812559.

Deprez, B., Williard, X., Bourel, L., Coste, H., Hyafil, F., Tartar, A. (1995) Self-deciphering, orthogonal combinatorial libraries of soluble organic compounds: Discovery of a potent V2 vasopressin antagonist, publisher: ESCOM,

Maia, H.L.S. (Ed.) *Proceeding of 23rd European Peptide Symposium*, Leiden, Neth., 455–456.

DeRoock, I.B., Pennington, M.E., Sroka, T.C., Lam, K.S., Bowden, G.T., Bair, E.L., Cress, A.E. (2001) Synthetic peptides inhibit adhesion of human tumor cells to extracellular matrix proteins, *Cancer Res.* **61**, 3308–3313.

Devlin, J.J., Panganiban, L.C., Devlin, P.E. (1990) Random peptide libraries – a source of specific protein binding molecules, *Science* **249**, 404–406.

Dolle, R.E. (2000) Comprehensive survey of combinatorial library synthesis: 1999, *J. Comb. Chem.* **2**, 383–433.

Dolle, R.E. (2001) Comprehensive survey of combinatorial library synthesis: 2000, *J. Comb. Chem.* **3**, 477–517.

Dolle, R.E. (2002) Comprehensive survey of combinatorial library synthesis: 2001, *J. Comb. Chem.* **4**, 369–418.

Dolle, R.E. (2003) Comprehensive survey of combinatorial library synthesis: 2002, *J. Comb. Chem.* **5**, 693–753.

Dooley, C.T., Houghten, R.A. (1993) The use of positional scanning synthetic peptide combinatorial libraries for the rapid determination of opioid receptor ligands, *Life Sci.* **52**, 1509–1517.

Eichler, J., Beyermann, M., Bienert, M. (1989) Application of cellulose paper as support material in simultaneous solid phase peptide synthesis, *Coll. Czech. Chem. Commun.* **54**, 746–1752.

Eichler, J., Bienert, M., Stierandova, A., Lebl, M. (1991) Evaluation of cotton as a carrier for solid-phase peptide synthesis, *Pept. Res.* **4**, 296–307.

Enjalbal, C., Maux, D., Martinez, J., Combarieu, R., Aubagnac, J.L. (2001) Mass spectrometry and combinatorial chemistry: New approaches for direct support-bound compound identification, *Comb. Chem. High Throughput Screen.* **4**, 363–373.

Erb, E., Janda, K.D., Brenner, S. (1994) Recursive deconvolution of combinatorial chemical libraries, *Proc. Natl. Acad. Sci. USA* **91**, 11422–11426.

Falsey, J.R., Renil, M., Park, S., Li, S.J., Lam, K.S. (2001) Peptide and small molecule microarray for high throughput cell adhesion and functional assays, *Bioconjug. Chem.* **12**, 346–353.

Fenniri, H., Ding, L.H., Ribbe, A.E., Zyrianov, Y. (2001) Barcoded resins: A new concept for polymer-supported combinatorial library self-deconvolution, *J. Am. Chem. Soc.* **123**, 8151–8152.

Fodor, S.P.A., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., Solas, D. (1991) Light-directed, spatially addressable parallel chemical synthesis, *Science* **251**, 767–773.

Frank, R. (1992) Spot-synthesis – an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support, *Tetrahedron* **48**, 9217–9232.

Furka, A., Sebestyen, F., Asgedom, M., Dibo, G. (1991) General method for rapid synthesis of multicomponent peptide mixtures, *Int. J. Pept. Protein Res.* **37**, 487–493.

Geysen, H.M., Meloen, R.H., Barteling, S.J. (1984) Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid, *Proc. Natl. Acad. Sci. USA* **81**, 3998–4002.

Geysen, H.M., Rodda, S.J., Mason, T.J. (1986) A priori delineation of a peptide which mimics a discontinuous antigenic determinant, *Mol. Immunol.* **23**, 709–716.

Hergenrother, P.J., Depew, K.M., Schreiber, S.L. (2000) Small-molecule microarrays: covalent attachment and screening of alcohol-containing small molecules on glass slides, *J. Am. Chem. Soc* **122**, 7849–7850.

Hoffmann, C., Blechschmidt, D., Kruger, R., Karas, M., Griesinger, C. (2002) Mass spectrometric sequencing of individual peptides from combinatorial libraries via specific generation of chain-terminated sequences, *J. Comb. Chem.* **4**, 79–86.

Houghten, R.A. (1985) General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen–antibody interaction at the level of individual amino acids, *Proc. Natl. Acad. Sci. USA* **82**, 5131–5135.

Houghten, R.A., Appel, J.R., Blondelle, S.E., Cuervo, J.H., Dooley, C.T., Pinilla, C. (1992) The use of synthetic peptide combinatorial libraries for the identification of bioactive peptides, *Biotechniques* **13**, 412–421.

Houghten, R.A., Pinilla, C., Blondelle, S.E., Appel, J.R., Dooley, C.T., Cuervo, J.H. (1991) Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery, *Nature* **354**, 84–86.

Houseman, B.T., Huh, J.H., Kron, S.J., Mrksich, M. (2002) Peptide chips for the quantitative evaluation of protein kinase activity, *Nat. biotechnol.* **20**, 270–274.

Jayawickreme, C.K., Graminski, G.F., Quillan, J.M., Lerner, M.R. (1994) Creation and functional screening of a multi-use peptide library, *Proc. Natl. Acad. Sci. USA* **91**, 1614–1618.

Jayawickreme, C.K., Sauls, H., Bolio, N., Ruan, J., Moyer, M., Burkhart, W., Marron, B., Rimele, T., Shaffer, J. (1999) Use of a cell-based, lawn format assay to rapidly screen a 442,368 bead-based peptide library, *J. Pharmacol. Toxicol. Methods* **42**, 189–197.

Kanoh, N., Kumashiro, S., Simizu, S., Kondoh, Y., Hatakeyama, S., Tashiro, H., Osada, H. (2003) Immobilization of natural products on glass slides by using a photoaffinity reaction and the detection of protein-small-molecule interactions, *Angew. Chem. Int. Ed. Engl.* **42**, 5584–5587.

Karlsson, T., Songyang, Z., Landgren, E., Lavergne, C., Di Fiore, P.P., Anafi, M., Pawson, T., Cantley, L.C., Claesson-Welsh, L., Welsh, M. (1995) Molecular interactions of the Src homology 2 domain protein Shb with phosphotyrosine residues, tyrosine kinase receptors and Src homology 3 domain proteins, *Oncogene* **10**, 1475–1483.

Kawasaki, G. (1991) *Cell-free Synthesis and Isolation of Novel Genes and Polypeptides*, PCT WO 9105058.

Kieke, M.C., Shusta, E.V., Boder, E.T., Teyton, L., Wittrup, K.D., Kranz, D.M. (1999) Selection of functional T cell receptor mutants from a yeast surface-display library, *Proc. Natl. Acad. Sci. USA* **96**, 5651–5656.

Kjaergaard, K., Schembri, M.A., Klemm, P. (2001) Novel Zn2+-chelating peptides selected from a fimbria-displayed random peptide library, *Appl. Environ Microbiol* **67**, 5467–5473.

Kohn, M., Wacker, R., Peters, C., Schroder, H., Soulere, L., Breinbauer, R., Niemeyer, C.M., Waldmann, H. (2003) Staudinger ligation: a new immobilization strategy for the preparation of small-molecule arrays, *Angew. Chem. Int. Ed. Engl.* **42**, 5830–5834.

Lam, K.S., Wu, J. (1994) Use of Selectide technology to identify protein kinase A substrate motif, *Methods* **6**, 401–403.

Lam, K.S., Wu, J.Z., Lou, Q. (1995) Identification and characterization of a novel synthetic peptide substrate specific for Src-family protein tyrosine kinases, *Int. J. Pept. Protein Res.* **45**, 587–592.

Lam, K.S., Hruby, V.J., Lebl, M., Knapp, R.J., Kazmierski, W.M., Hersh, E.M., Salmon, S.E. (1993) The chemical synthesis of large random peptide libraries and their use for the discovery of ligands for macromolecular acceptors, *Bioorg. Med. Chem. Lett.* **3**, 419–424.

Lam, K.S., Lou, Q., Zhao, Z.-G., Smith, J., Chem, M., Pleshko, E., Salmon, S.E. (1995) Idiotype specific peptides bind to the surface immunoglobulins of two murine B-cell lymphoma lines, inducing signal transduction, *Biomed. Pept. Proteins Nucleic Acids* **1**, 205–210.

Lau, D.H., Guo, L., Liu, R., Song, A., Shao, C., Lam, K.S. (2002) Identifying peptide ligands for cell surface receptors using cell-growth-on-bead assay and one-bead one-compound combinatorial library, *Biotechnol. Lett.* **24**, 497–500.

LeProust, E., Pellois, J.P., Yu, P., Zhang, H., Gao, X., Srivannavit, O., Gulari, E., Zhou, X. (2000) Digital light-directed synthesis. A microarray platform that permits rapid reaction optimization on a combinatorial basis, *J. Comb. Chem.* **2**, 349–354.

Lesaicherre, M.-L., Uttamchandani, M., Chen, G.Y.J., Yao, S.Q. (2002) Developing site-Specific immobilization strategies of peptides in a microarray, *Bioorg. Med. Chem. Lett.* **12**, 2079–2083.

Liu, C.F., Tam, J.P. (1994) Chemical ligation approach to form a peptide bond between unprotected peptide segments – concept and model study, *J. Am. Chem. Soc.* **116**, 4149–4153.

Liu, R., Lam, K.S. (2001) Automatic Edman microsequencing of peptides containing multiple unnatural amino acids, *Anal. Biochem.* **295**, 9–16.

Liu, R., Marik, J., Lam, K.S. (2002) A novel peptide-based encoding system for "one-bead one-compound" peptidomimetic and small molecule combinatorial libraries, *J. Am. Chem. Soc.* **124**, 7678–7680.

Lou, Q., Leftwich, M.E., Lam, K.S. (1996) Identification of GIYWHHY as a novel peptide substrate for human p60c-src protein tyrosine kinase, *Bioorg. Med. Chem. Lett.* **4**, 677–682.

MacBeath, G., Koehler, A.N., Schreiber, S.L. (1999) Printing small molecules as microarrays and detecting protein-ligand

interactions en masse, *J. Am. Chem. Soc.* **121**, 7967–7968.

Maeji, N.J., Bray, A.M., Geysen, H.M. (1990) Multi-pin peptide synthesis strategy for T-cell determinant analysis, *J. Immunol. Methods* **134**, 23–33.

Marik, J., Xu, Q., Wang, X., Peng, L., Lam, K.S. (2004) A novel encoded high-density chemical microarray platform for proteomics and drug development, in: Sawyer, T.K., Chorev M. (Eds.) *Peptides: Peptide Revolution, Genomics, Proteomics and Therapeutics*, American Peptide Society, Boston: Proceedings of 18th American Peptide Symposium. In press.

Marron, B.E., Jayawickreme, C.K. (2003) Going to the well no more: lawn format assays for ultra-high-throughput screening, *Curr. Opin. Chem. Biol.* **7**, 395–401.

McGall, G., Labadie, J., Brock, P., Wallraff, G., Nguyen, T., Hinsberg, W. (1996) Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists, *Proc. Natl. Acad. Sci. USA* **93**, 13555–13560.

Meldal, M., Svendsen, I., Breddam, K., Auzanneau, F.I. (1994) Portion-mixing peptide libraries of quenched fluorogenic substrates for complete subsite mapping of endoprotease specificity, *Proc. Natl. Acad. Sci. USA* **91**, 3314–3318.

Metzger, J.W., Kempter, C., Wiesmuller, K.H., Jung, G. (1994) Electrospray Mass Spectrometry and Tandem Mass Spectrometry of Synthetic Multicomponent Peptide Mixtures – Determination of Composition and Purity, *Anal. Biochem.* **219**, 261–277.

Moran, E.J., Sarshar, S., Cargill, J.F., Shahbaz, M.M., Lio, A., Mjalli, A.M.M., Armstrong, R.W. (1995) Radio frequency tag encoded combinatorial library method for the discovery of tripeptide-substituted cinnamic acid inhibitors of the protein tyrosine phosphatase Ptp1b, *J. Am. Chem. Soc.* **117**, 10787–10788.

O'Boyle, D.R., Pokornowski, K.A., McCann, P.J., Weinheimer, S.P. (1997) Identification of a novel peptide substrate of HSV-1 protease using substrate phage display, *Virology* **236**, 338–347.

Ohkubo, S., Miyadera, K., Sugimoto, Y., Matsuo, K., Wierzba, K., Yamada, Y. (2001) Substrate phage as a tool to identify novel substrate sequences of proteases, *Comb. Chem. High Throughput Screen.* **4**, 573–583.

Pellois, J.P., Wang, W., Gao, X.L. (2000) Peptide synthesis based on t-Boc chemistry and solution photogenerated acids, *J. Comb. Chem.* **2**, 355–360.

Pennington, M.E., Lam, K.S., Cress, A.E. (1996) The use of a combinatorial library method to isolate human tumor cell adhesion peptides, *Mol. Divers.* **2**, 19–28.

Pinilla, C., Appel, J.R., Blanc, P., Houghten, R.A. (1992) Rapid identification of high affinity peptide ligands using positional scanning synthetic peptide combinatorial libraries, *Biotechniques* **13**, 901–905.

Rogers, Y.-H., Jiang-Baucom, P., Huang, Z.-J., Bogdanov, V., Anderson, S., Boyce-jacino, M.T. (1999) Immobilization of oligonucleotides onto a glass support via disulfide bonds: a method for preparation of DNA microarrays, *Anal. Biochem.* **266**, 23–30.

Salmon, S.E., Liu-Stevens, R.H., Zhao, Y., Lebl, M., Krchnak, V., Wertman, K., Sepetov, N., Lam, K.S. (1996) High-volume cellular screening for anticancer agents with combinatorial chemical libraries – a new methodology, *Mol. Divers.* **2**, 57–63.

Sanger, F., Nicklen, S., Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.

Schatz, P.J. (1993) Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme – a 13 residue consensus peptide specifies biotinylation in Escherichia-Coli, *Biotechnology* **11**, 1138–1143.

Scott, J.K., Smith, G.P. (1990) Searching for peptide ligands with an epitope library, *Science* **249**, 386–390.

Silen, J.L., Lu, A.T., Solas, D.W., Gore, M.A., MacLean, D., Shah, N.H., Coffin, J.M., Bhinderwala, N.S., Wang, Y., Tsutsui, K.T., Look, G.C., Campbell, D.A., Hale, R.L., Navre, M., DeLuca-Flaherty, C.R. (1998) Screening for novel antimicrobials from encoded combinatorial libraries by using a two-dimensional agar format, *Antimicrob. Agents Chemother.* **42**, 1447–1453.

Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., Cerrina, F. (1999) Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array, *Nat. Biotechnol.* **17**, 974–978.

Smith, M.H., Lam, K.S., Hersh, E.M., Lebl, M., Grimes, W.J. (1994) Peptide sequences

binding to Mhc class I proteins, *Mol. Immunol.* **31**, 1431–1437.

Soellner, M.B., Dickson, K.A., Nilsson, B.L., Raines, R.T. (2003) Site-specific protein immobilization by Staudinger ligation, *J. Am. Chem. Soc.* **125**, 11790–11791.

Song, A., Zhang, J., Lebrilla, C.B., Lam, K.S. (2003) A novel and rapid encoding method based on mass spectrometry for "one-bead-one-compound" small molecule combinatorial libraries, *J. Am. Chem. Soc.* **125**, 6180–6188.

Songyang, Z., Shoelson, S.E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X.R., Barbacid, M., Sabe, H., Hanafusa, H., Yi, T., Ren, R., Baltimore, D., Ratnofsky, S., Feldman, R.A., Cantley, L.C. (1994) Specific motifs recognized by the Sh2 domains of Csk, 3bp2, Fps Fes, Grb-2, Hcp, Shc, Syk, and Vav, *Mol. Cell. Biol.* **14**, 2777–2785.

Songyang, Z., Carraway, K.L. 3rd, Eck, M.J., Harrison, S.C., Feldman, R.A., Mohammadi, M., Schlessinger, J., Hubbard, S.R., Smith, D.P., Eng, C., Lorenzo, M.J., Ponder, B.A., Mayer, B.J., Cantley, L.C. (1995) Catalytic specificity of protein-tyrosine kinases is critical for selective signalling, *Nature* **373**, 536–539.

Stevanovic, S., Jung, G. (1993) Multiple sequence analysis – pool sequencing of synthetic and natural peptide libraries, *Anal. Biochem.* **212**, 212–220.

St Hilaire, P.M., Willert, M., Juliano, M.A., Juliano, L., Meldal, M. (1999) Fluorescence-quenched solid phase combinatorial libraries in the characterization of cysteine protease substrate specificity, *J. Comb. Chem.* **1**, 509–523.

Stillman, B.A., Tonkinson, J.L. (2000) FAST (TM) slides: A novel surface for microarrays, *Biotechniques* **29**, 630–632, 634–635.

Van der Zee, R., Van Eden, W., Meloen, R.H., Noordzij, A., Van Embden, J.D.A. (1989) Efficient mapping and characterization of a T cell epitope by the simultaneous synthesis of multiple peptides, *Eur. J. Immunol.* **19**, 43–48.

Wang, Z.X., Laursen, R.A. (1992) Multiple peptide synthesis on polypropylene membranes for rapid screening of bioactive peptides, *Pept. Res.* **5**, 275–280.

Wang, X., Zhang, J., Song, A., Lebrilla, C.B., Lam, K.S. (2004) Encoding method for OBOC small molecule libraries using a biphasic approach for ladder-synthesis of coding tags, *J. Am. Chem. Soc.* **126**, 5740–5749.

Wu, J.Z., Ma, Q.Y.N., Lam, K.S. (1994) Identifying substrate motifs of protein kinases by a random library approach, *Biochemistry* **33**, 14825–14833.

Wu, J.J., Afar, D.E.H., Phan, H., Witte, O.N., Lam, K.S. (2002) Recognition of multiple substrate motifs by the c-ABL protein tyrosine kinase, *Comb. Chem. High Throughput Screen.* **5**, 83–91.

Xu, Q., Miyamoto, S., Lam, K.S. (2004) A novel approach to chemical microarray using ketone-modified macromolecular scaffolds: application in micro cell-adhesion assay, *Mol. Divers.* **8**, 301–310.

Youngquist, R.S., Fuentes, G.R., Lacey, M.P., Keough, T. (1995) Generation and screening of combinatorial peptide libraries designed for rapid sequencing by mass spectrometry, *J. Am. Chem. Soc.* **117**, 3900–3906.

Zuckermann, R.N., Kerr, J.M., Kent, S.B.H., Moos, W.H. (1992) Efficient method for the preparation of peptoids [oligo(N-substituted glycines)] by submonomer solid-phase synthesis, *J. Am. Chem. Soc.* **114**, 10646–10647.

Zuckermann, R.N., Kerr, J.M., Siani, M.A., Banville, S.C., Santi, D.V. (1992) Identification of highest-affinity ligands by affinity selection from equimolar peptide mixtures generated by robotic synthesis, *Proc. Natl. Acad. Sci. USA* **89**, 4505–4509.

Phagocytosis

Rebecca M. Henry and Joel A. Swanson
University of Michigan Medical School, Ann Arbor, MI, USA

1 Phagocytosis and Physiology 3

2 Types of Phagocytes 4
2.1 Professional Phagocytes 4
2.1.1 Macrophages 4
2.1.2 Polymorphonuclear Leukocytes 5
2.1.3 Dendritic Cells 5
2.2 Nonprofessional Phagocytes 5
2.2.1 Renewal of the Sensory Retina 5
2.2.2 Clearance of Apoptotic Cells 6
2.2.3 Phagocytosis and Pathogens 6

3 Cell Biology of Phagocytosis 6
3.1 Models of Phagocytic Signaling 7
3.2 Recognition of the Phagocytic Target 7
3.2.1 Direct Recognition 8
3.2.2 Indirect Recognition 8
3.3 Early Signaling Events in Phagocytosis 9
3.4 Internalization 10
3.5 Cytoskeleton in Phagocytosis 11
3.6 Phagosome Maturation in Macrophages 12
3.6.1 The Endocytic Pathway of Macrophages 12
3.6.2 Membrane Traffic in Phagosome Maturation 12
3.6.3 Production of Microbicidal Molecules within Phagosomes 13

4 Microbes and Phagocytosis 14
4.1 Evasion of Phagocytosis 14
4.2 Induction of Phagocytosis 14
4.3 Intracellular Survival within Phagocytes 15

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

5 Conclusions 15

Bibliography 16
Books and Reviews 16
Primary Literature 16

Keywords

Apoptosis
The process of programmed cell death; the body's normal method of disposing of damaged, unwanted, or unneeded cells.

Cytoskeleton
An internal network of filaments and tubules that give the eukaryotic cell its ability to move, to assume a characteristic shape, to divide, to arrange its organelles, and to transport them from one location to another.

Endocytosis
The process by which soluble material is internalized into a cell via invagination of cell surface membrane.

Endosome
A membrane-bound organelle that carries material newly ingested by endocytosis. It passes many of the materials to the terminal degradative compartment (lysosome) for degradation.

Lysosome
A membrane-bound organelle containing digestive enzymes with acidic pH optima. Lysosomes are found in almost nucleated all cells and are vehicles for secretion as well as digestion.

Opsonin
A substance naturally present in the blood that renders bacteria susceptible to phagocytosis through the promotion or enhancement of attachment via a cell surface receptor.

Phagocyte
An immune cell that engulfs and digests dead or dying cells, microorganisms, or other foreign bodies in the bloodstream and tissues.

Phagosome
A vacuole formed by the internalization of surface membrane along with an attached particle.

Pseudopod
Literally "false foot"; extensions of the plasma membrane used for a variety of cell activities, most notably locomotion and phagocytosis.

Phagocytosis, the process by which eukaryotic cells internalize other cells, cell fragments, protein aggregates, and foreign bodies, has an essential role in physiology. In primitive organisms, it is primarily used for acquisition of nutrients, whereas, in higher organisms, it occurs in specialized cells such as monocytes and neutrophils and is required for a wide variety of biological events. These include clearance of infectious agents and senescent cells, and tissue repair and remodeling. Because phagocytosis is essential for survival and homeostasis, complex signaling pathways have evolved both to regulate the engulfment process and to couple uptake of particulate material with appropriate cellular responses. Many pathogens manipulate the signaling pathways involved in phagocytosis to avoid phagocytic killing and to establish intracellular infections within phagocytes.

1 Phagocytosis and Physiology

Phagocytosis is the process by which eukaryotic cells internalize other cells, cell fragments, protein aggregates, and foreign bodies. While targets for engulfment may be up to many microns in diameter, the term *phagocytosis* is generally reserved for particles that are more than 1 μm in diameter, because, at this size, the mechanisms for internalization clearly differ from those used for endocytosis of soluble material. Phagocytosis was first described in 1883 by Elie Metchnikoff, who observed cells in starfish larvae capable of ingesting and destroying foreign substances. His work led to the discovery of the role of phagocytes in host defense against pathogens. This was a significant contribution to the field of immunology. Since that time, studies of phagocytosis have not only elucidated aspects of immune defense but also cell biology, developmental biology, microbiology, and ecology.

Phagocytosis is a component of many different physiological processes. For example, amoebae phagocytose bacteria and other organisms as a food source, and the nitrogen released by soil amoebae stimulates the growth of nearby plant life. In higher organisms, the most prominent roles for phagocytosis include the removal of tissue debris after injury, senescent erythrocytes, apoptotic cells, and the destruction and removal of invading microbes.

Survival of protozoan amoebae depends on their ability to internalize particulate matter. These single-celled animals feed primarily on bacteria, soluble organic matter, and sometimes fungi. Soil protozoa play an important role in mineralizing nutrients, making them available for use by plants and other soil organisms. After soil protozoa internalize and digest bacteria, they release excess nitrogen, and since they are concentrated near plant roots, the plant can benefit from this supply of nutrients. These plants, in turn, are used as a food source for many vertebrates. In

this way, phagocytosis performed by tiny, single-celled organisms is a vital part of the food chain.

Marine bivalves, such as mussels, oysters, and clams, utilize phagocytosis as a primary means of defense against microbes. Bivalves filter great volumes of water through their gills. These invertebrates trap and accumulate large numbers of microorganisms present in the water. To cope with this challenge, bivalves possess phagocytic cells called *hemocytes*, which engulf and destroy pathogenic microbes.

In mammals, phagocytosis is the primary mechanism for removal of invading microorganisms, dead and dying cells, tissue debris, protein aggregates, and foreign bodies. Phagocytosis is essential for host defense against infectious diseases, for organ remodeling in embryonic development, for tissue repair after injury, and for the removal of aging and senescent cells. Although many mammalian cells are able to perform phagocytosis, it is the primary function of specialized leukocytes, monocytes, and neutrophils. These phagocytes circulate throughout the body in search of microbes and senescent cells. Once internalized, a number of signaling pathways are initiated within phagocytes, which regulate complex host immune responses such as inflammation.

2 Types of Phagocytes

2.1 Professional Phagocytes

The host defense aspects of phagocytosis require recognition of potential pathogens. This is a function of specialized leukocytes, myeloid cells called *macrophages*, *polymorphonuclear leukocytes* (PMNs), and *dendritic cells*, collectively referred to as *professional phagocytes*. All professional phagocytes originate from precursor cells in the bone marrow. They migrate through blood and tissues, surveying for bacteria, foreign particles, and cellular debris, which they internalize and digest. Although PMNs, macrophages, and dendritic cells share the common function of phagocytosis, each cell type has distinct functions in the immune system.

2.1.1 Macrophages

Mononuclear phagocytes are released from the bone marrow as monocytes, which migrate from the bloodstream into different tissues, where they further differentiate into mature macrophages. The phenotype of these cells is highly dependent on the tissue in which they reside. Kupffer cells, resident macrophages of the liver, remove dying erythrocytes and other debris from the bloodstream. Alveolar macrophages in the lung ingest and remove inhaled particles that are small enough to reach the terminal airways. Microglia in the central nervous system ingest degenerated myelin. Macrophages are also found in the submucosa of the gastrointestinal and urogenital tracts, in the skin, and in the synovial membranes lining joint cavities. At these sites, they remove bacteria that frequently cross epithelial barriers into these tissues from the gastrointestinal tract or the lung.

Macrophages have important roles in the innate immune response. They internalize and destroy infectious agents, and release inflammatory mediators, including interleukin (IL)-8, which attracts neutrophils to sites of infection, and IL-12 or IL-4, which activates T cells. In response to IL-12, T cells produce interferon gamma, which augments the effector functions of macrophages.

2.1.2 Polymorphonuclear Leukocytes

Polymorphonuclear leukocytes (PMNs), or neutrophils, are short-lived cells (average life span is 24 to 48 h) equipped with numerous antimicrobial effector mechanisms. Unlike resident macrophages and immature dendritic cells, PMNs do not reside in peripheral tissues prior to infection. Rather, PMNs are recruited from the circulation to sites of infection by signals produced by pathogens and other immune cells. This process of oriented migration in response to chemical gradients is called *chemotaxis*. Recruited PMNs accumulate at sites of infection, phagocytose pathogens and kill them, using several microbicidal mechanisms. PMNs contain several types of granules that are specialized for the storage and secretion of antimicrobial products, including lysozyme, which can digest the peptidoglycan component of most bacterial cell walls, and myeloperoxidase, which converts hydrogen peroxide and hydrochloric acid to hydrochlorous acid. After PMNs internalize a pathogen, the membrane-bound granules fuse with the phagosome, placing the effector molecules in direct contact with the ingested material. In addition to intracellular killing of phagocytosed bacteria, PMNs are also capable of killing extracellular microbes by secreting granules into the extracellular space.

2.1.3 Dendritic Cells

Dendritic cells (DCs) are so named because of the branched morphology of their surface. DCs may develop directly from common myeloid progenitors or from monocytes and plasmacytoid cells. The different progenitors give rise to different subsets of DCs: Langerhans DC, interstitial DC, monocyte-derived DC, and plasmacytoid-derived DC. DCs are found in virtually all organs.

DCs are best known for their role in the initiation of adaptive immune responses. Their function in immunity alters with their maturity. Immature DCs in the peripheral tissues are highly phagocytic. Once they have phagocytosed pathogens, they mature, becoming less phagocytic. Mature DCs then migrate to lymphoid organs, where they transmit information to T cells about the nature of the pathogen and the infected tissues.

2.2 Nonprofessional Phagocytes

In multicellular organisms, a significant amount of phagocytosis is carried out by nonprofessional phagocytes. These are usually fibroblasts or epithelial cells whose phagocytic function facilitates general processes such as maintenance of the extracellular matrix, response to injury, development of inflammation, and regulation of immune responses. Phagocytosis by these cells is also essential to specific processes such as maintenance of vision and clearance of apoptotic cells. Nonprofessional phagocytosis is also an important vehicle through which pathogens gain access to the intracellular space.

2.2.1 Renewal of the Sensory Retina

The retina is a part of the central nervous system directly exposed to external environmental stimuli. The outer segments of photoreceptors are damaged by oxidative stress and need to undergo a renewal process. Phagocytic cells of the retinal pigmented epithelium (RPE) contribute to retinal renewal by ingesting and degrading the spent tips of photoreceptor outer segments. RPE cells are the most active phagocytes in the body, ingesting approximately 25 000 to 30 000 discarded segments per day. Phagocytosis of rod

outer segment tips by RPE cells is vitally important for maintaining structural and functional integrity of the retina. When RPE cells are unable to perform this task, irreversible photoreceptor degeneration and blindness occur. RPE phagocytic defects are common in the aging eye and may contribute to age-related macular degeneration.

2.2.2 Clearance of Apoptotic Cells

As a part of normal growth and development, certain cells undergo programmed cell death, or apoptosis. In contrast to necrosis, which is a form of cell death that results from acute tissue injury, apoptosis is carried out in an ordered process that minimizes inflammation. Although apoptotic cells may be removed by monocytes and macrophages, nonprofessional phagocytes play an important role in the clearance of apoptotic cells. This is particularly true in tissues that undergo frequent remodeling, such as epithelial barriers, and tissues that have been subjected to trauma. While professional phagocytes often need to be recruited to these areas, nonprofessional phagocytes are already present, allowing for immediate recognition and removal of dead cells. The timely engulfment of apoptotic cells prevents the release of their potentially proinflammatory and proimmunogenic contents.

After a cell has undergone apoptosis, it breaks down into apoptotic particles with intact membranes. These particles have certain molecules on their surface, including lectins and the lipid phosphatidylserine (PS), which signal their readiness to be phagocytosed. Unlike phagocytosis of pathogens, which is followed by the release of inflammatory mediators, phagocytosis of apoptotic cells is immunologically "silent" in that it does not lead to the induction of inflammatory responses.

2.2.3 Phagocytosis and Pathogens

Phagocytosis is critical to the survival of many pathogens. It provides a means by which many pathogens can avoid circulating immune factors such as antibodies and complement (see Sect. 3.2.2). The host cell cytoplasm also contains nutrients and other factors that many pathogens require. For example, the food-borne pathogen *Listeria monocytogenes* induces its uptake into intestinal epithelial cells and utilizes host cell actin to travel from cell to cell without ever leaving the intracellular environment. By using actin polymerization to propel itself, *Listeria* can distend the plasma membrane of one epithelial cell to project into the plasma membrane of another. For further discussion of this topic, see Sect. 4.

3 Cell Biology of Phagocytosis

Phagocytosis requires stimulation of receptors on the surface of the phagocyte by ligands on a phagocytic target. The receptors activate a number of signaling pathways that together orchestrate rearrangement of the actin cytoskeleton, extension of the plasma membrane, and fusion to form a vacuole termed the *phagosome*. Several downstream signaling events regulate interactions of the phagosome with the endocytic pathway and the production of microbicidal molecules. An overwhelming number of signaling molecules participate in phagocytosis. This section will highlight several molecules involved in receptor function during phagocytosis, in the process of ingestion, in the process of phagosome maturation, and in the linkage of ingestion to specific cellular responses.

3.1
Models of Phagocytic Signaling

The signaling initiated by particle binding to a phagocyte receptor causes local changes in the plasma membrane that lead to internalization. It follows that signaling events during phagocytosis must be highly localized. Two models of signaling have been proposed to describe the spatial organization of the phagocytic response: the zipper model and the trigger model.

According to the "zipper hypothesis", phagocytosis requires repeated interactions between ligands on the target particle and receptors on the phagocytic cell. In this model, membrane protrusions of the phagocytic cell move over the target like a zipper, as ligands on the particle sequentially engage phagocyte receptors. An important feature of this model is that the membrane events that lead to engulfment must be very localized. Signaling is spatially confined to the region of the cytoplasm near the receptor. The phagocytic response is limited to the particle that initiates phagocytosis.

In contrast, the trigger model entails phagocytic signaling that is not entirely confined to the region of the cytoplasm near the stimulated receptor. Signaling may radiate from the receptor site. As a result, the phagocyte may not only engulf the particle that initiated engulfment but also nearby particles. Phagocytosis in this model does not require repeated interactions between the particle and phagocyte receptors. Particle binding results in a signal that is propagated through a region of cytoplasm to nearby receptors. Once this signal is above a certain threshold, it leads to particle engulfment. The first pseudopod formed is often enough to complete ingestion. Some bacteria enter cells by stimulating this kind of all-or-none response.

Subsequent to engulfment, the newly formed phagosome participates in a series of membrane fission and fusion events, a process called *phagosome maturation*. The locally guided signaling that leads to engulfment in zippering phagocytosis does not appear to guide phagosome maturation. If maturation were strictly locally regulated, then identical phagosomes should all mature alike. Instead, identical phagosomes do not always mature identically. This indicates that signals governing maturation are not strictly locally regulated, but are integrated over the entire organelle.

Numerous different receptors can stimulate particle ingestion; the cytoskeletal elements mediating internalization differ according to the receptor stimulated and the nature of the pathogen being internalized, and the intracellular fates of the particle can differ. Despite the complexity associated with different phagocytic mechanisms, all share a number of features: Particle internalization is initiated by the interaction of specific receptors on the surface of the phagocyte with ligands on the surface of the particle. This leads to the polymerization of actin at the site of ingestion, and the internalization of the particle. The internalized particle is contained in a membrane-limited phagosome that matures via fission and fusion reactions with the endosomal system. Maturation is completed as the phagosome fuses with lysosomes and the phagosomal contents are degraded. However, many pathogens are able to avoid this fate by altering phagosome maturation.

3.2
Recognition of the Phagocytic Target

Recognition of the phagocytic target can either be direct or indirect. Direct

recognition occurs through interactions between phagocyte receptors and components of the target cell wall. Indirect recognition occurs through the process of opsonization. Opsonization is the binding of host cell proteins (opsonins) to the surface of microbes, apoptotic cells, or cellular debris. Opsonins promote or enhance the binding of the particle to phagocyte receptors.

3.2.1 Direct Recognition

Professional phagocytes express a set of surface receptors that recognize invading organisms and dead or dying cells. These target motifs are called *pathogen-associated molecular patterns* (PAMPs). The phagocyte receptors that recognize PAMPs are called *pattern recognition receptors* (PRRs). A number of these receptors recognize specific features that differentiate the target from normal cells. For example, there are phagocyte receptors for lipopolysaccharide (LPS), a component of the gram-negative bacterial cell wall, for peptidoglycan of gram-positive organisms, β-glucans in many fungal cell walls, denatured proteins, and phosphatidylserine (a surface-exposed component of apoptotic cells). These receptors directly bind their ligands on the target and initiate signaling cascades that lead to their uptake.

Toll-like Receptors Toll-like receptors (TLRs) are PRRs important in innate immunity. TLRs are a family of receptors that recognize a variety of molecules specific to pathogens, including surface components such as LPS, peptidoglycan, and lipoteichoic acid. TLRs derive their name from the Drosophila protein, Toll, which is involved in early development and host defense in the fly. TLRs are widely expressed on both professional and nonprofessional phagocytes. TLRs are recruited to phagosomes containing ingested targets. However, there is no evidence that TLRs themselves initiate phagocytosis. Signaling through certain TLRs is thought to identify a pathogen and trigger a tailored immune response. Thus, it is likely that TLRs localized to phagosomes act primarily to identify the target and to transmit signals that may activate specific microbicidal mechanisms.

3.2.2 Indirect Recognition

The host immune response has evolved soluble proteins that recognize invading microorganisms and, in turn, are recognized by specific receptors on phagocytic cells. These serum proteins are called *opsonins*, from the Greek *opson*, meaning, "to prepare to eat." Opsonins increase the range of microbes recognized as potential pathogens, or cells recognized as damaged or defective, and they engage potentially more efficient mechanisms for clearance than occurs through direct recognition. The classic opsonins are antibodies and complement, for which there are various receptors on phagocytic cells.

Fc Receptor Fc receptors (FcR) bind to the Fc portion of antibody molecules. Receptors for IgA (FcαR) and IgG (FcγR) are expressed on professional phagocytes. Phagocytosis through FcγR has been extensively studied. The Fcγ receptors on human monocytes that activate phagocytosis are FcγRI, FcγRIIA, and FcγRIII. These receptors contain immune-tyrosine activation motifs (ITAMs) in their cytoplasmic domains, which initiate signaling for particle engulfment through the activation of tyrosine kinases.

Complement Receptor The complement component C3 is the second major opsonin in serum. During complement activation, C3 is first cleaved to C3b and then further to inactivated C3bi. C3b or C3bi fragments attach nonspecifically to the carbohydrate surface of pathogens. Phagocytes express receptors for both of these C3 fragments, and these complement receptors may initiate phagocytosis. Unlike Fc receptors, complement receptors are not competent for ingestion in unactivated cells. Phagocytes ingest C3-opsonized targets efficiently only after activation. Activation stimuli can be found at sites of infection and inflammation, indicating that complement receptor-mediated phagocytosis occurs primarily at these sites.

Complement receptors 3 and 4 (CR3 and CR4) mediate ingestion of particles, while complement receptor 1 (CR1) is primarily responsible for adhesion of the opsonized target to the phagocytic cell. CR3 has a central role in PMN phagocytosis, and its expression is required for other receptors to interact appropriately with the cytoskeleton for adhesion and ingestion. The importance of this receptor in PMN function is highlighted by the phenotype of individuals with leukocyte adhesion deficiency type I, who lack the $\beta 2$ chain of CR3. These patients have severe recurrent infections due to failure of PMNs to migrate to the site of infection and failure of the few cells that do reach the site of infection to phagocytose efficiently.

CR3 and CR4 are integrin receptors that link to the cytoskeleton through their cytoplasmic tails. Their role in phagocytosis may be linked to their effects on the cytoskeletal organization of phagocytes. These receptors also activate tyrosine kinases that are essential for particle engulfment.

3.3 Early Signaling Events in Phagocytosis

When FcRs are ligated and clustered, tyrosine kinases are activated. ITAM-mediated activation of FcRs requires the tyrosine kinase Syk. Syk is required for FcR-mediated ingestion, but not CR3-mediated phagocytosis. Syk activation requires phosphorylation of the tyrosines within the ITAM, and, in many cases, members of the Src family of tyrosine kinases perform this function. However, genetic deletion of Hck, Fgr, and Lyn, three of the major Src family kinases of macrophages, delays but does not inhibit FcR-mediated phagocytosis. It is possible that Syk can phosphorylate the ITAMs itself when the receptors are sufficiently clustered.

A number of signaling events relevant to phagocytosis occur as a result of Syk activation. Syk can act not only as a kinase but also as a docking molecule for other signaling effectors. Syk activation leads to actin polymerization through the activation of phosphatidylinositol 3-kinase (PI-3 kinase), Rho family guanosine triphosphatases (GTPases), and p21-activated kinase. Syk also leads to the activation of phospholipase C (PLC), which, in turn, activates protein kinase C (PKC).

PI-3 kinase catalyzes the phosphorylation of phosphatidylinositol (PI) $(4,5)P_2$ to $PI(3,4,5)P_3$, a phospholipid important in recruiting signaling molecules such as the kinase Akt/PKB to specific regions of the membrane. Inhibition of PI 3-kinase blocks phagocytosis of a broad spectrum of particles that include IgG- and complement opsonized particles and bacteria, clearly demonstrating an important role for PI-3 kinase activation in particle internalization. Blockade of PI-3 kinase does not inhibit particle binding or actin polymerization beneath the particle, which

suggests that phagocytic signaling is intact. Instead, PI-3 kinase is required for membrane extension and fusion behind the particle, perhaps by regulating the insertion of new membrane at the site of particle internalization.

PLC cleaves PI(4,5)P_2, releasing inositol triphosphate (IP_3) and diacylglycerol (DAG). DAG activates PKC, which participates in the formation of actin filaments beneath the site of particle binding. Like PI-3 kinase, PLC is recruited to phagosomes containing IgG-opsonized particles and inhibition of its activity blocks internalization. Unlike PI-3 kinase inhibitors, PLC inhibitors completely block the formation of actin filaments beneath the site of particle contact, strongly suggesting that the main role of PLC is to activate PKC.

Phagocytosis induces activation of small GTPases of the Rho family, Rho, Rac, and Cdc42 that regulate cytoskeletal rearrangements. FcR-mediated phagocytosis depends on Rac and Cdc42, while CR3-mediated uptake of particles activates Rho. Cdc42 and Rac coordinate actin polymerization through the activation of WASP (Wiskcott–Aldrich syndrome protein) and Scar/WAVE (WASP family Verprolin-homologous) respectively and the Arp2/3 complex, which together lead to actin polymerization and pseudopod extension. Rac is also involved in the assembly and activation of the phagocyte NADPH oxidase complex, which generates microbicidal reactive oxygen molecules. Rho is required for myosin-dependent cell contractility, which is morphologically similar to the mechanism of CR3-mediated phagocytosis.

3.4 Internalization

Particle internalization requires activation of a number of signaling pathways that together orchestrate rearrangement of the actin cytoskeleton, extension of the plasma membrane, and fusion to form the phagosome. The precise mechanism of internalization is remarkably heterogeneous and depends on the type of receptor or receptors that participate in particle recognition and on the nature of the particle. Thus, although FcR-mediated phagocytosis is initiated by membrane protrusions that extend around the particle, particles internalized through CR3 appear to sink into the cell without significant membrane protrusions. These differences are likely due to different cytoskeletal elements that are assembled at the site of phagocytosis.

FcR-mediated phagocytosis occurs through a zippering mechanism. Sequential ligation of FcRs by particle-bound IgG results in pseudopod extension over the particle (Fig. 1). Alterations in the actin cytoskeleton are essential for pseudopod extension during FcR-mediated phagocytosis. Pseudopod advance is followed by contractile activities that constrict the outer margin of these extensions to enclose the particle. Myosins are utilized for the contractile activity of phagosome closure. FcR-mediated phagocytosis results in a response proportional to the size of the particle, which results in the formation of a phagosome tightly apposed to the particle surface.

Phagocytosis mediated through CR3 occurs through a variation of the zipper method; it does not involve pseudopod formation, but rather C3-opsonized particles simply sink into phagocytes. In addition to the actin cytoskeleton, microtubules are required for CR3-mediated phagocytosis.

Several bacteria are internalized through methods very different from zippering phagocytosis. *Salmonella typhimurium* triggers generalized membrane ruffling that results in the internalization of

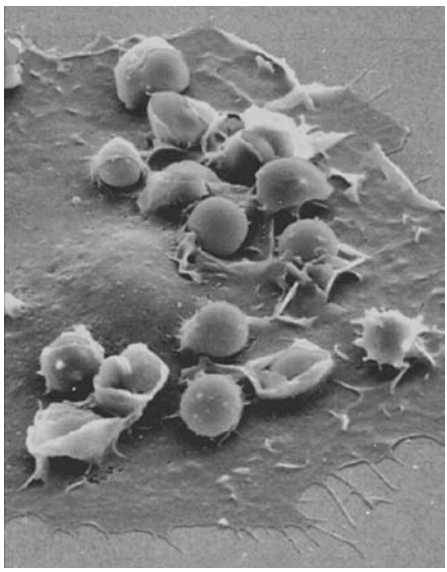

Fig. 1 Scanning electron micrograph of a macrophage internalizing IgG-opsonized erythrocytes.

the bacterium into a large compartment resembling a macropinosome (a large compartment of internalized extracellular fluid). The phagosome is large relative to the bacterium and has been called a *spacious phagosome*. *Legionella pneumophila* contains a surface protein that fixes complement to the surface of the bacterium, thereby facilitating binding to the macrophage surface through the CR. After binding, the bacterium induces the formation of an extended host cell pseudopod that spirals around the bacterium, forming a structure termed a *coiled phagosome*. Since this structure of this phagosome is very different than that induced by CR3, it is likely that additional *L. pneumophila* signals must be required.

3.5 Cytoskeleton in Phagocytosis

The cytoskeleton mediates a variety of essential biological functions in all eukaryotic cells. In addition to providing the structural framework around which cell shape and polarity are defined, its dynamic properties provide the driving force for cells to move. The changes in cell shape that occur during phagocytosis are dependent on reorganization of the actin cytoskeleton. The cytoskeleton also serves as a platform to bring surface receptors, enzymes, and their substrates into close proximity. Thus, agents that depolymerize the cytoskeleton not only inhibit shape changes but also inhibit signals initiated by receptor ligation.

All types of phagocytosis require myosins and actin polymerization at the site of ingestion. Myosins have been localized to phagosomes and play a key role in phagosome formation. Actin polymerization generates the force necessary for particle engulfment during phagocytosis. This is particularly evident in FcR-mediated phagocytosis during which a band of filamentous actin (F-actin) moves over the particle during pseudopod extension. Subsequent to FcR activation, localized actin polymerization occurs underneath the phagocytic target and promotes assembly of a continuous F-actin cup. In contrast, CR3-mediated internalization leads to the transient accumulation of F-actin and cytoskeletal elements in small foci underneath bound particles. Although phagocytosis mediated through FcR and CR3 both require actin, only phagocytosis mediated by CR3 is sensitive to agents that disrupt microtubules. This difference may account for the different mechanisms of internalization initiated by the receptors.

The cytoskeleton also plays a part in phagosome maturation. Microtubules are required for positioning of various

components of the endocytic pathway in defined domains of the cell and for organelle motility. Microtubules provide the substrate for motor proteins, dynein, and kinesin, which propel organelles in specific directions along the length of the microtubule. Microtubule-based motility contributes to phagosome maturation, as disruption of microtubules can slow fusion of phagosomes with lysosomes.

3.6 Phagosome Maturation in Macrophages

The internalization of a particle into a membrane-limited vacuole creates a new intracellular organelle, the phagosome. In order to progress to the hostile environment required for degradation, the phagosome must be completely remodeled, a process termed *phagosome maturation*. As the phagosome matures within the macrophage, both its membrane composition and contents are modified through sequential interactions with the endocytic pathway. The process occurs differently in neutrophils, which have a small endocytic compartment and specialized lysosomes called *primary granules*. Primary granules fuse directly with the base of the phagosome as it develops.

3.6.1 The Endocytic Pathway of Macrophages

The endocytic pathway is a series of intracellular vesicular organelles, distinguishable by their membranes and contents. Within the endocytic compartment, molecules mix and redistribute by organelle fusion and cytoskeleton-associated organelle motility. The distinct identities of the compartment organelles are maintained during this mixing by segregation of membrane molecules, together with vesicle fission reactions that restore or maintain compartment identities.

The endocytic pathway consists of endosomes and lysosomes. Endosomes are acidic organelles that receive internalized molecules from the plasma membrane, and sort them efficiently to other vesicular compartments, or recycle them back to the plasma membrane. There are early endosomes and late endosomes, which differ in membrane markers and luminal pH. Early and late endosomes have a characteristic pH of 6.5 and 6.0 respectively. Membrane markers of early endosomes include Rab5 and early endosome antigen 1 (EEA1), while Rab7 is a marker of late endosomes. Lysosomes are distinguished from endosomes by a pH of 5.0 to 5.5 and the presence of acid hydrolases. Lysosomes are the terminal degradation compartment of the endocytic pathway.

3.6.2 Membrane Traffic in Phagosome Maturation

The nascent phagosome membrane was originally thought to be similar in composition to the plasma membrane, containing a sample of the extracellular medium. Recent analyses of phagosomes containing latex beads identified a number of membrane proteins that typically reside the endoplasmic reticulum (ER). The presence of these proteins on phagosomal membranes indicated that the ER contributes membrane to the forming phagosome.

After internalization, phagosome maturation is temporally coordinated through the sequential activities of molecules that direct fusion with elements of the endocytic pathway. Rab GTPases and phosphoinositides regulate diverse functions associated with membrane trafficking, such as vesicle formation, vesicle docking and fusion, and organelle association with the cytoskeleton. Rab5 directs interactions of

the phagosome with early endosomes and recruits the type III PI-3 kinase hVps34, which generates PI(3)P. Rab5 and PI(3)P are present on the phagosomal membrane at early time points and may facilitate the assembly of the NADPH oxidase complex on early phagosomes.

Rab7 is acquired by phagosomes during the progression from the early to late endosomal stages. Rab7 is an important regulator of membrane transport from early to late endosomes and of lysosome biogenesis. Late-stage phagosomes fuse with lysosomes to form a hybrid organelle in which the phagosomal contents are degraded by acid hydrolases.

3.6.3 Production of Microbicidal Molecules within Phagosomes

In concert with phagosomal maturation, there is a progressive acidification of the interior of the phagosome. The drop in pH occurs as a result of the acquisition of the vacuolar proton-ATPase to the phagosome, which actively pumps protons into the phagosome. The acidic lumen can inhibit bacterial growth and viability, and enhance other microbicidal mechanisms within phagosomes. For example, low pH can activate pH-sensitive proteases and lipases that are delivered to the phagosome during fusion with lysosomes. Creating an inhospitable environment for pathogens may not be the only role of phagosomal acidification. The lumenal pH contributes to membrane traffic. Inhibition of phagosomal acidification prevents phagosome-lysosome fusion.

Microbicidal anions are also produced within phagosomes. The presence of Rac and PI(3)P on early phagosomes facilitates the assembly of the NADPH oxidase complex on early phagosomes. Activation of NADPH oxidase does not occur as a result of CR3-mediated phagocytosis, as CR3-mediated uptake of particles activates Rho, but not Rac. NADPH oxidase catalyzes the production of superoxide anion. Superoxide can interact with chelated metals to produce hydroxyl radical or dismutate to hydrogen peroxide, from which additional reactive oxygen intermediates (ROI) are generated. ROI are potent cytotoxic molecules, which damage bacterial DNA and membranes. The critical importance of ROI generation by NADPH oxidase in immunity to microbes is emphasized by the symptoms of patients with chronic granulomatous disease (CGD), a genetic deficiency in any of the NADPH oxidase subunits. The hallmarks of CGD are recurrent bacterial infections and chronic inflammation with granuloma formation.

Superoxide and hydrogen peroxide, in addition to their own antimicrobial activity, can also be used as substrates for the neutrophil enzyme, myeloperoxidase. Myeloperoxidase is stored in the granules of neutrophils and is also expressed in monocytes. Using the products of NADPH oxidase as its substrates, myeloperoxidase generates hypochlorous acid as well as other ROI, all of which have potent microbicidal activities.

Cytotoxic reactive nitrogen intermediates (RNI) are also produced inside phagosomes of mouse macrophages activated by interferon gamma and LPS. Inducible nitric oxide synthetase localizes to vesicles and produces nitric oxide (NO) from L-arginine. NO is toxic to bacteria, as it damages membranes, and inactivates DNA and proteins. Superoxide and NO can combine to produce another microbicidal molecule, peroxynitrite.

4 Microbes and Phagocytosis

Phagocytosis evolved, in part, as a defense mechanism against pathogenic microorganisms. In turn, pathogenic bacteria have evolved methods that avoid destruction by phagocytes. Some bacteria avoid phagocytosis, while others survive within phagocytes. Paradoxically, many pathogens actually stimulate cells to initiate phagocytosis, and thrive in the intracellular environment (see Fig. 2).

4.1 Evasion of Phagocytosis

Phagocytosis requires binding of the particle by phagocyte cell surface receptors. *Streptococcus pyogenes* avoid phagocytosis by producing a slippery capsule, which phagocytes cannot grasp. The capsule also prevents opsonization by antibodies or complement. Other pathogenic bacteria have evolved mechanisms to regulate Rho family GTPases as a means of avoiding immune clearance. *Yersiniae* species avoid phagocytosis by host immune cells in part by secreting a protein called YopE into the cytoplasm of host cells. YopE rapidly deactivates Cdc42, Rac, and Rho, thereby inhibiting the actin rearrangements necessary for phagocytosis.

4.2 Induction of Phagocytosis

Many pathogens can evade immune destruction by taking advantage of the phagocytic uptake of apoptotic cells. The parasite

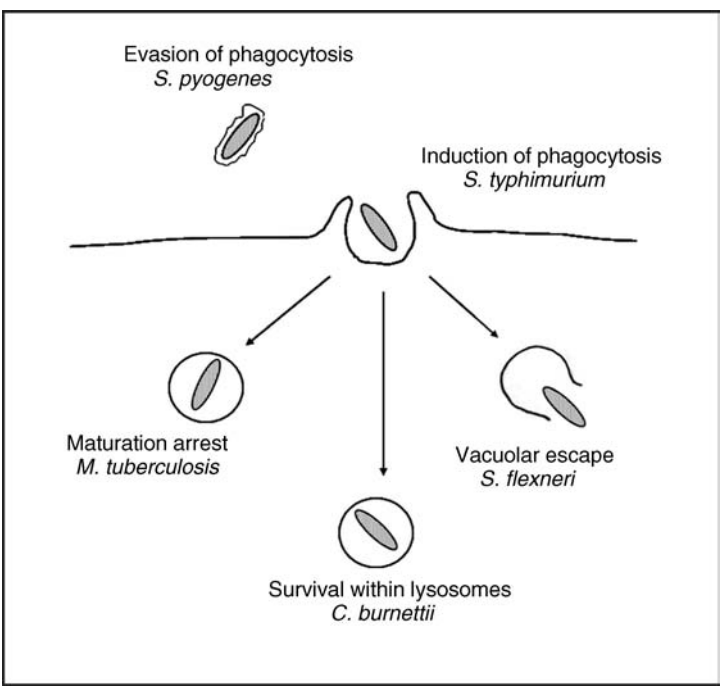

Fig. 2 Bacterial mechanisms for escaping phagocytic clearance. Selected examples of bacteria that use the indicated mechanisms are given.

Trypanosoma brucei induces apoptosis in host cells and is engulfed along with the apoptotic particles into new host cells. *Leishmania* spp are obligate intracellular parasites; that is, they must reside inside cells for survival. To reach the intracellular environment, these parasites mimic apoptotic cells by expressing PS on their outer membranes; they are phagocytosed by macrophages.

Many pathogens, such as *L. monocytogenes,* induce their uptake into host cells by directly engaging receptors on nonprofessional phagocytes. In this way, they can gain intracellular entry by receptor-mediated phagocytosis using receptors normally used for adhesion or growth control. Others, such as *Escherichia coli* and *Salmonella*, secrete factors that can selectively induce phagocytosis. Commonly, pathogens induce phagocytosis by triggering signal transduction cascades leading to cytoskeletal rearrangements.

4.3 Intracellular Survival within Phagocytes

Subsequent to phagocytosis, pathogens may survive within phagocytes by modifying the normal membrane trafficking that accompanies phagosome maturation. Some bacteria halt phagosome maturation prior to phagosome–lysosome fusion, and others avoid the endocytic pathway all together. *Mycobacteria tuberculosis* prevent fusion of phagosomes with lysosomes by inhibiting the function of EEA1, an early endosome marker required for phagosome–endosome interactions. These bacteria also prevent phagosomal acidification by excluding ATP, an energy source required for the proton ATPase pump, from its limiting membrane. *Legionella* and *Chlamydia* phagosomes have little or no interaction with the endocytic compartment but instead reside in compartments derived from the endoplasmic reticulum and Golgi apparatus, respectively. Compartments containing *Chlamydia* are not microbicidal; they do not acidify and do not interact with lysosomes. *Legionella* phagosomes resemble autophagosomes, which eventually fuse with lysosomes.

A few intracellular pathogens avoid destruction within lysosomes by producing toxins that rupture the phagosome and allow the bacteria to survive within the host cell cytosol. In the safe cytosolic environment, these bacteria are ensured a neutral pH, ample access to nutrients, and protection from many of the host cell defenses by enzymes like superoxide dismutase and catalase that eliminate ROI. *Shigella flexneri* and *L. monocytogenes* are examples of bacteria that employ phagosomal exit as a survival strategy.

Instead of avoiding the lysosomal environment, some bacteria can survive within it. *L. pneumophila* and *Coxiella burnetii* are unique among intracellular pathogens, in that they thrive within the acidic environment of lysosomes. *C. burnetii* has an absolute requirement for the acidic pH to activate its metabolism.

5 Conclusions

Phagocytosis is a fundamental cellular process that serves multiple functions in tissue homeostasis and immunity. The targets of phagocytosis, their recognition motifs, the receptors essential for their uptake, and the intracellular signals necessary for their entry into phagosomes are as diverse as the biological processes to which phagocytosis contributes. The molecular mechanisms underlying phagocytosis are

just beginning to be uncovered. The mechanisms utilized by pathogens to modify the phagocytic pathway to their own advantage have provided useful information about the development of the phagosome. Further insights into the mechanisms of phagocytosis may suggest novel strategies for modulation of the immune response and treatment of human infection.

See also Electron Microscopy in Cell Biology; Endocytosis; Membrane Traffic: Vesicle Budding and Fusion; Membrane Transport.

Bibliography

Books and Reviews

Abramson, J.S., Wheeler, J.G. (Eds.) (1993) *The Neutrophil*, Oxford University Press, Oxford, UK.

Aderem, A., Underhill, D.M. (1999) Mechanisms of phagocytosis in macrophages, *Ann. Rev. Immunol.* **17**, 593–623.

Burke, B., Lewis, C.E. (Eds.) (2002) *The Macrophage*, Oxford University Press, Oxford, UK.

Gordon, S. (Ed.) (1999) *Phagocytosis: The Host*, Vol. 5, JAI Press, Stanford, CT.

van Furth, R. (1994) *Macrophage-Pathogen Interactions*, Immunology series, Vol. 60. Marcel Dekker, Inc, New York.

Primary Literature

Adam, T. (2001) Exploitation of host factors for efficient infection by Shigella, *Int. J. Med. Microbiol.* **291**, 287–298.

Aderem, A., Ulevitch, R.J. (2000) Toll-like receptors in the induction of the innate immune response, *Nature* **406**, 782–787.

Aderem, A., Underhill, D.M. (1999) Mechanisms of phagocytosis in macrophages, *Annu. Rev. Immunol.* **17**, 593–623.

Allen, L.A., Aderem, A. (1996a) Mechanisms of phagocytosis, *Curr. Opin. Immunol.* **8**, 36–40.

Allen, L.A., Aderem, A. (1996b) Molecular definition of distinct cytoskeletal structures involved in complement- and Fc receptor-mediated phagocytosis in macrophages, *J. Exp. Med.* **184**, 627–637.

Allen, L.H., Aderem, A. (1995) A role for MARCKS, the alpha isozyme of protein kinase C and myosin I in zymosan phagocytosis by macrophages, *J. Exp. Med.* **182**, 829–840.

Allison, A.C., Davies, P., De Petris, S. (1971) Role of contractile microfilaments in macrophage movement and endocytosis, *Nat. New Biol.* **232**, 153–155.

Alpuche-Aranda, C.M., Racoosin, E.L., Swanson, J.A., Miller, S.I. (1994) Salmonella stimulate macrophage macropinocytosis and persist within spacious phagosomes, *J. Exp. Med.* **179**, 601–608.

Anderson, D.C., Schmalstieg, F.C., Arnaout, M.A., Kohl, S., Tosi, M.F., Dana, N., Buffone, G.J., Hughes, B.J., Brinkley, B.R., Dickey, W.D., Abramson, J.S., Springer, T., Boxer, L.A., Hollers, J.M., Smith, C.W. (1984) Abnormalities of PMN function associated with a heritable deficiency of high molecular weight surface glycoproteins (GP138): common relationship to diminished cell adherence, *J. Clin. Invest.* **74**, 536–551.

Araki, N., Johnson, M.T., Swanson, J.A. (1996) A role for phosphoinositide 3-kinase in the completion of macropinocytosis and phagocytosis by macrophages, *J. Cell. Biol.* **135**, 1249–1260.

Araki, N., Hatae, T., Furukawa, A., Swanson, J.A. (2003) Phosphoinositide-3- kinase-independent contractile activities associated with Fcgamma-receptor-mediated phagocytosis and macropinocytosis in macrophages, *J. Cell Sci.* **116**, 247–257.

Babior, B.M. (1999) NADPH oxidase: an update, *Blood* **93**, 1464–1476.

Baca, O.G., Li, Y.P., Kumar, H. (1994) Survival of the Q fever agent Coxiella burnetii in the phagolysosome, *Trends Microbiol.* **2**, 476–480.

Beckman, J.S., Koppenol, W.H. (1996) Nitric oxide, superoxide, and peroxynitrite: the good, the bad, and ugly, *Am. J. Physiol.* **271**, C1424–C1437.

Bellinger-Kawahara, C., Horwitz, M.A. (1990) Complement component C3 fixes selectively to the major outer membrane protein (MOMP) of Legionella pneumophila and mediates phagocytosis of liposome-MOMP complexes

by human monocytes, *J. Exp. Med.* **172**, 1201–1210.

Bok, D. (1993) The retinal pigment epithelium: a versatile partner in vision, *J. Cell Sci. Suppl.* **17**, 189–195.

Bok, D., Hall, M.O. (1971) The role of the pigment epithelium in the etiology of inherited retinal dystrophy in the rat, *J. Cell Biol.* **49**, 664–682.

Botelho, R.J., Teruel, M., Dierckman, R., Anderson, R., Wells, A., York, J.D., Meyer, T., Grinstein, S. (2000) Localized biphasic changes in phosphatidylinositol-4,5-bisphosphate at sites of phagocytosis, *J. Cell Biol.* **151**, 1353–1368.

Brown, E.J. (1991) Complement receptors and phagocytosis, *Curr. Opin. Immunol.* **3**, 76–82.

Canesi, L., Gallo, G., Gavioli, M., Pruzzo, C. (2002) Bacteria-hemocyte interactions and phagocytosis in marine bivalves, *Microsc. Res. Tech.* **57**, 469–476.

Cardelli, J. (2001) Phagocytosis and macropinocytosis in Dictyostelium: phosphoinositide-based processes, biochemically distinct, *Traffic* **2**, 311–320.

Castellano, F., Chavrier, P., Caron, E. (2001) Actin dynamics during phagocytosis, *Semin. Immunol.* **13**, 347–355.

Christoforidis, S., McBride, H.M., Burgoyne, R.D., Zerial, M. (1999) The Rab5 effector EEA1 is a core component of endosome docking, *Nature* **397**, 621–625.

Cox, D., Tseng, C.C., Bjekic, G., Greenberg, S. (1999) A requirement for phosphatidylinositol 3-kinase in pseudopod extension, *J. Biol. Chem.* **274**, 1240–1247.

de Freitas Balanco, J.M., Moreira, M.E., Bonomo, A., Bozza, P.T., Amarante-Mendes, G., Pirmez, C., Barcinski, M.A. (2001) Apoptotic mimicry by an obligate intracellular parasite downregulates macrophage microbicidal activity, *Curr. Biol.* **11**, 1870–1873.

Deretic, V., Fratti, R.A. (1999) Mycobacterium tuberculosis phagosome, *Mol. Microbiol.* **31**, 1603–1609.

Desjardins, M., Huber, L.A., Parton, R.G., Griffiths, G. (1994) Biogenesis of phagolysosomes proceeds through a sequential series of interactions with the endocytic apparatus, *J. Cell Biol.* **124**, 677–688.

Diakonova, M., Bokoch, G., Swanson, J.A. (2002) Dynamics of cytoskeletal proteins during Fcgamma receptor-mediated phagocytosis in macrophages, *Mol. Biol. Cell.* **13**, 402–411.

Diebold, B.A., Bokoch, G.M. (2001) Molecular basis for Rac2 regulation of phagocyte NADPH oxidase, *Nat. Immunol.* **2**, 211–215.

DiRita, V.J., Engleberg, N.C., Heath, A., Miller, A., Crawford, J.A., Yu, R. (2000) Virulence gene regulation inside and outside, *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* **355**, 657–665.

Duclos, S., Diez, R., Garin, J., Papadopoulou, B., Descoteaux, A., Stenmark, H., Desjardins, M. (2000) Rab5 regulates the kiss and run fusion between phagosomes and endosomes and the acquisition of phagosome leishmanicidal properties in RAW 264.7 macrophages, *J. Cell Sci.* **113**(Pt 19), 3531–3541.

Dunne, D.W., Resnick, D., Greenberg, J., Krieger, M., Joiner, K.A. (1994) The type I macrophage scavenger receptor binds to gram-positive bacteria and recognizes lipoteichoic acid, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1863–1867.

Ershov, A.V., Lukiw, W.J., Bazan, N.G. (1996) Selective transcription factor induction in retinal pigment epithelial cells during photoreceptor phagocytosis, *J. Biol. Chem.* **271**, 28458–28462.

Fadok, V.A., Chimini, G. (2001) The phagocytosis of apoptotic cells, *Semin. Immunol.* **13**, 365–372.

Feng, Y., Press, B., Chen, W., Zimmerman, J., Wandinger-Ness, A. (2001) Expression and properties of Rab7 in endosome function, *Methods Enzymol.* **329**, 175–187.

Finlay, B.B., Rosenshine, I., Donnenberg, M.S., Kaper, J.B. (1992) Cytoskeletal composition of attaching and effacing lesions associated with enteropathogenic Escherichia coli adherence to HeLa cells, *Infect. Immun.* **60**, 2541–2543.

Fitzer-Attas, C.J., Lowry, M., Crowley, M.T., Finn, A.J., Meng, F., DeFranco, A.L., Lowell, C.A. (2000) Fcgamma receptor-mediated phagocytosis in macrophages lacking the Src family tyrosine kinases Hck, Fgr, and Lyn, *J. Exp. Med.* **191**, 669–682.

Gagnon, E., Duclos, S., Rondeau, C., Chevet, E., Cameron, P.H., Steele-Mortimer, O., Paiement, J., Bergeron, J.J., Desjardins, M. (2002) Endoplasmic reticulum-mediated phagocytosis is a mechanism of entry into macrophages, *Cell* **110**, 119–131.

Gillooly, D.J., Simonsen, A., Stenmark, H. (2001) Phosphoinositides and phagocytosis, *J. Cell Biol.* **155**, 15–17.

Girard, M., Bisser, S., Courtioux, B., Vermot-Desroches, C., Bouteille, B., Wijdenes, J., Preud'homme, J.L., Jauberteau, M.O. (2003) In vitro induction of microglial and endothelial cell apoptosis by cerebrospinal fluids from patients with human African trypanosomiasis, *Int. J. Parasitol.* **33**, 713–720.

Green, S.J., Nacy, C.A., Meltzer, M.S. (1991) Cytokine-induced synthesis of nitrogen oxides in macrophages: a protective host response to Leishmania and other intracellular pathogens, *J. Leukoc. Biol.* **50**, 93–103.

Greenberg, S., el Khoury, J., di Virgilio, F., Kaplan, E.M., Silverstein, S.C. (1991) Ca(2+)-independent F-actin assembly and disassembly during Fc receptor-mediated phagocytosis in mouse macrophages, *J. Cell Biol.* **113**, 757–767.

Griffin, F.M. Jr., Silverstein, S.C. (1974) Segmental response of the macrophage plasma membrane to a phagocytic stimulus, *J. Exp. Med.* **139**, 323–336.

Gruenberg, J., Maxfield, F.R. (1995) Membrane transport in the endocytic pathway, *Curr. Opin. Cell Biol.* **7**, 552–563.

Hackstadt, T. (2000) Redirection of host vesicle trafficking pathways by intracellular parasites, *Traffic* **1**, 93–99.

Hackstadt, T., Baehr, W., Ying, Y. (1991) C. trachomatis developmentally regulated protein is homologous to eukaryotic histone H1, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3937–3941.

Hall, A. (1998) Rho GTPases and the actin cytoskeleton, *Science* **279**, 509–514.

Hampton, M.B., Kettle, A.J., Winterbourn, C.C. (1998) Inside the neutrophil phagosome: oxidants, myeloperoxidase, and bacterial killing, *Blood* **92**, 3007–3017.

Hardt, W.D., Chen, L.M., Schuebel, K.E., Bustelo, X.R., Galan, J.E. (1998) S. typhimurium encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells, *Cell* **93**, 815–826.

Henry, R.M., Hoppe, A.D., Joshi, N., Swanson, J.A. (2004) The uniformity of phagosome maturation in macrophages, *J. Cell Biol.* **164**, 185–194.

Herre, J., Gordon, S., Brown, G.D. (2004) Dectin-1 and its role in the recognition of beta-glucans by macrophages, *Mol. Immunol.* **40**, 869–876.

Horwitz, M.A. (1984) Phagocytosis of the Legionnaires' disease bacterium occurs by a novel mechanism: engulfment within a pseudopod coil, *Cell* **36**, 27–33.

Ingham, E.R. (1998) *The Soil Biology Primer, Soil Protozoa*, USDA, Natural Resources Conservation Science, Soil Quality Institute, Ames, IA.

Kaplan, G. (1977) Differences in the mode of phagocytosis with Fc and C3 receptors in macrophages, *Scand. J. Immunol.* **6**, 797–807.

Kaplan, S.S., Lancaster, J.R. Jr., Basford, R.E., Simmons, R.L. (1996) Effect of nitric oxide on staphylococcal killing and interactive effect with superoxide, *Infect. Immun.* **64**, 69–76.

Kiefer, F., Brumell, J., Al-Alawi, N., Latour, S., Cheng, A., Veillette, A., Grinstein, S., Pawson, T. (1998) The Syk protein tyrosine kinase is essential for Fcgamma receptor signaling in macrophages and neutrophils, *Mol. Cell. Biol.* **18**, 4209–4220.

Klebanoff, S.J. (1999) Oxygen Metabolites from Phagocytes, in: Gallin, J.I., Snyderman R. (Eds.) *Inflammation: Basic Principles and Clinical Correlates*, Lippincott Williams and Wilkins, Philadelphia, PA., pp. 721–721.

Kornfeld, S., Mellman, I. (1989) The biogenesis of lysosomes, *Annu. Rev. Cell Biol.* **5**, 483–525.

Larsen, E.C., DiGennaro, J.A., Saito, N., Mehta, S., Loegering, D.J., Mazurkiewicz, J.E., Lennartz, M.R. (2000) Differential requirement for classic and novel PKC isoforms in respiratory burst and phagocytosis in RAW 264.7 cells, *J. Immunol.* **165**, 2809–2817.

Lukacs, G.L., Rotstein, O.D., Grinstein, S. (1990) Phagosomal acidification is mediated by a vacuolar-type H(+)-ATPase in murine macrophages, *J. Biol. Chem.* **265**, 21099–21107.

Luzio, J.P., Rous, B.A., Bright, N.A., Pryor, P.R., Mullock, B.M., Piper, R.C. ((2000) Lysosome-endosome fusion and lysosome biogenesis, *J. Cell Sci.* **113**(Pt 9), 1515–1524.

McLauchlan, H., Newell, J., Morrice, N., Osborne, A., West, M., Smythe, E. (1998) A novel role for Rab5-GDI in ligand sequestration into clathrin-coated pits, *Curr. Biol.* **8**, 34–45.

Mellman, I. (1996) Endocytosis and molecular sorting, *Annu. Rev. Cell Dev. Biol.* **12**, 575–625.

Mullock, B.M., Bright, N.A., Fearon, C.W., Gray, S.R., Luzio, J.P. (1998) Fusion of lysosomes with late endosomes produces a hybrid organelle of intermediate density and is NSF dependent, *J. Cell Biol.* **140**, 591–601.

Nielsen, E., Severin, F., Backer, J.M., Hyman, A.A., Zerial, M. (1999) Rab5 regulates motility of early endosomes on microtubules, *Nat. Cell Biol.* **1**, 376–382.

Ozinsky, A., Underhill, D.M., Fontenot, J.D., Hajjar, A.M., Smith, K.D., Wilson, C.B., Schroeder, L., Aderem, A. (2000) The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13766–13771.

Ravetch, J.V. (1997) Fc receptors, *Curr. Opin. Immunol.* **9**, 121–125.

Savill, J. (2000) Apoptosis in resolution of inflammation, *Kidney Blood Press. Res.* **23**, 173–174.

Savill, J., Fadok, V. (2000) Corpse clearance defines the meaning of cell death, *Nature* **407**, 784–788.

Segal, A.W., Wientjes, F., Stockley, R.W., Dekker, L.V. (1999) Components and Organization of the NADPH Oxidase of Phagocytic Cells: Its Role in Microbial Killing and in the Molecular Pathology of Chronic Granulomatous Disease, in: Gordon, S. (Ed.) *Phagocytosis: The Host*, JAI Press, Stanford, CT, pp. 441–441.

Sengelov, H. (1995) Complement receptors in neutrophils, *Crit. Rev. Immunol.* **15**, 107–131.

Sheetz, M.P. (1996) Microtubule motor complexes moving membranous organelles, *Cell Struct. Funct.* **21**, 369–373.

Simonsen, A., Lippe, R., Christoforidis, S., Gaullier, J.M., Brech, A., Callaghan, J., Toh, B.H., Murphy, C., Zerial, M., Stenmark, H. (1998) EEA1 links PI(3)K function to Rab5 regulation of endosome fusion, *Nature* **394**, 494–498.

Sory, M.P., Boland, A., Lambermont, I., Cornelis, G.R. (1995) Identification of the YopE and YopH domains required for secretion and internalization into the cytosol of macrophages, using the cyaA gene fusion approach, *Proc. Natl. Acad. Sci. U. S. A.* **92**, 11998–12002.

Sturgill-Koszycki, S., Swanson, M.S. (2000) Legionella pneumophila replication vacuoles mature into acidic, endocytic organelles, *J. Exp. Med.* **192**, 1261–1272.

Sturgill-Koszycki, S., Schlesinger, P.H., Chakraborty, P., Haddix, P.L., Collins, H.L., Fok, A.K., Allen, R.D., Gluck, S.L., Heuser, J., Russell, D.G. (1994) Lack of acidification in Mycobacterium phagosomes produced by exclusion of the vesicular proton-ATPase, *Science* **263**, 678–681.

Swanson, J., Bushnell, A., Silverstein, S.C. (1987) Tubular lysosome morphology and distribution within macrophages depend on the integrity of cytoplasmic microtubules, *Proc. Natl. Acad. Sci. U. S. A.* **84**, 1921–1925.

Swanson, J.A., Johnson, M.T., Beningo, K., Post, P., Mooseker, M., Araki, N. (1999) A contractile activity that closes phagosomes in macrophages, *J. Cell Sci.* **112**(Pt 3), 307–316.

Swanson, M.S., Isberg, R.R. (1995) Association of Legionella pneumophila with the macrophage endoplasmic reticulum, *Infect. Immun.* **63**, 3609–3620.

Tauber, A.I., Chernyak, L. (1991) *Metchnikoff and the Origins of Immunology. From Metaphor to Theory*, Oxford University Press, Oxford, UK.

Underhill, D.M., Ozinsky, A. (2002) Toll-like receptors: key mediators of microbe detection, *Curr. Opin. Immunol.* **14**, 103–110.

Vazquez-Boland, J.A., Kuhn, M., Berche, P., Chakraborty, T., Dominguez-Bernal, G., Goebel, W., Gonzalez-Zorn, B., Wehland, J., Kreft, J. (2001) Listeria pathogenesis and molecular virulence determinants, *Clin. Microbiol. Rev.* **14**, 584–640.

Vodovotz, Y., Russell, D., Xie, Q.W., Bogdan, C., Nathan, C. (1995) Vesicle membrane association of nitric oxide synthase in primary mouse macrophages, *J. Immunol.* **154**, 2914–2925.

Weiss, A., Littman, D.R. (1994) Signal transduction by lymphocyte antigen receptors, *Cell* **76**, 263–274.

Zhan, Y., Virbasius, J.V., Song, X., Pomerleau, D.P., Zhou, G.W. (2002) The p40phox and p47phox PX domains of NADPH oxidase target cell membranes via direct and indirect recruitment by phosphoinositides, *J. Biol. Chem.* **277**, 4512–4518.

Pharmaceuticals Expressed in Plants: *see* Plant-based Expression of Biopharmaceuticals

Pharmacogenomics and Drug Design

Philip Dean, Paul Gane, and Edward Zanders
DeNovo Pharmaceuticals Limited, Cambridge, UK

1 Introduction to *In Silico* Drug Design to Protein Targets 198
1.1 General Background 198
1.2 The Need for Protein Structure Information 199
1.3 Protein Structure and Variation in Drug Targets – the Scale of the Problem 200

2 Mutations in Drug Targets Leading to Changes in the Ligand-binding Pocket 201
2.1 β_2-adrenergic Receptor 201
2.2 STI-571 and BCR-ABL 202
2.3 Resistance to Human Immunodeficiency Virus (HIV) 202

3 *In Silico* Design of Small Molecules 203
3.1 Automated Drug Design Methods 203
3.2 Structure-based Drug Design 204
3.3 Ligand-based Drug Design 207

4 Future Directions 208

Bibliography 209
Books and Reviews 209
Primary Literature 209

Keywords

***In silico* Drug Design**
Design of small organic molecules using computer algorithms.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

single nucleotide polymorphism (SNP)
Alteration of single base pair in DNA, leading to genetic mutations that may be associated with alterations in protein structure or expression.

Structure-based Drug Design
In silico drug design process where structure of ligand-binding site on protein is known.

Ligand-based Drug Design
In silico drug design process where activity data from a number of chemically distinct ligands are available – does not require protein structure.

The Human Genome Project is largely complete, so the problem facing scientists is how to maximize the use of the newly acquired data for improving health care. It is estimated that the number of therapeutic targets available for drug discovery will increase from the current number of 600 to 1000 to perhaps as many as 5000 to 10 000. In addition to the challenges that this number provides to the pharmaceutical industry, there is the issue of sequence variation through single nucleotide polymorphisms (SNPs), some of which may impact upon the way that the body handles drug treatment. This may be due to a direct effect on the binding site of the protein target through nonconservative alterations of the amino acid sequence, or else through indirect effects on drug metabolizing enzymes. In order to meet these challenges, the marriage of structural proteomics and computer-aided small molecule design will provide opportunities for creating new molecules *in silico*; these may be designed to bind to selected pharmacogenetic variants of a protein in order to overcome the nonresponsiveness of certain patient groups to a particular medicine. The basic aspects of these technologies, and their applicability to selected targets showing structural variation, form the basis of this chapter.

1 Introduction to *In Silico* Drug Design to Protein Targets

1.1 General Background

Much has been written concerning the expense and problems encountered by the modern pharmaceutical industry in developing novel drugs. As a further difficulty, the patient population (market) is becoming fragmented because of genetic variation in the response to medicines, resulting from alterations in the drug target or in the metabolism of the compound once ingested. The relatively new discipline of pharmacogenetics is concerned with the inheritance of these variations, measured using single nucleotide polymorphism (SNP) analysis at the level of genomic DNA. Pharmacogenomics, on the other hand, is relevant when the genetic variation leads to changes in protein expression or conformation of key ligand

binding sites. This may result in loss of drug efficacy due to lack of binding; when the loss of binding occurs in drug metabolizing enzymes, this may result in levels of compound in the blood that exceed safety thresholds.

While these aspects of human genetics are problematical for both patients and pharmaceutical companies alike (for obviously different reasons), there are ways forward; these are emerging from new drug discovery technologies, including *in silico* approaches to drug design. The relevant technologies will be described in this chapter, but first we shall discuss the background to lead discovery and pharmacogenomics by concentrating on the structural aspects of the protein targets for small molecule drugs.

1.2 The Need for Protein Structure Information

Traditional drug discovery has relied upon the chemical modification of biologically active natural products or high-throughput screening of compound libraries to obtain hits that may be converted to leads and ultimately drugs. This process is generally inefficient, since the nature of compound collections used for screening is often a reflection of the historical activity of the company in question. This means that most compounds will be limited in coverage of the variety of targets encountered in drug research, namely, enzymes, receptors, and ion channels. Even the advent of combinatorial chemistry in the mid-1990s has failed to deliver a noticeable increase in good drug leads.

The problems highlighted above have been compounded by the increase in targets afforded by the genome sequencing projects, culminating in the recent publications of the human sequence. The realization that it will be impossible to find suitable small molecule candidates for every potential drug target using high-throughput screening alone has driven the search for alternatives based on an understanding of the three-dimensional structure of the protein. The structural information on the ligand-binding site may then be used for the *in silico* design of compounds that make strong interactions with appropriate residues within these sites; alternatively, existing small molecule structures may be docked into the sites and optimized using medicinal chemistry techniques.

The availability of three-dimensional protein structures is clearly one of the rate-limiting steps in this process. Publicly available structures, derived by X-ray diffraction or NMR techniques, are deposited in the protein data bank (PDB), and currently number over 20 000 entries. There is considerable redundancy in this, however, with the number of single entries for human proteins being considerably less. Owing to the technical difficulties associated with certain classes of protein there is a strong bias toward enzymes in the database. This excludes, therefore, the G-protein-coupled receptors (GPCR) and ion channel classes that make up a large proportion of drug targets. A number of public and private structural proteomics initiatives have been established in order to increase the number and variety of structures that may be used for studies of protein folding or drug discovery. Where structures are not available, it is possible in many instances to create a homology model using the structure of a related protein as a guide. The success of drug design corresponds to

the percentage identity between the sequence and the known structure. Despite these advances (including the structural determination of the GPCR rhodopsin, it has to be accepted that some proteins will not yield to current structural techniques. This does not mean, however, that *in silico* drug design techniques cannot be used with these proteins, since ligand-based design can be employed (see Sect. 3.3).

1.3 Protein Structure and Variation in Drug Targets – the Scale of the Problem

Variation in the response of individual patients to medicines is an important issue that is currently being addressed by the pharmaceutical industry. It would be useful to have some idea of the scale of the problem by determining the number of SNPs in the human genome and their relative effects on both the level of expression and functional activity of the protein target. From the perspective of rational drug design, it is necessary to consider the alteration of the tertiary structure of the translated protein in which function is retained, but not the binding of an existing drug. This is directly analogous to the situation with the microbial targets (e.g. HIV reverse transcriptase) in which sequence variations affect the structure of the target, resulting in the problem of drug resistance.

An analysis of 1.42 million nonredundant human SNPs has been recently published by an international collaborative group. The majority of these are in repetitive regions, whereas 60 000 lie within coding and untranslated regions of exons. Bearing in mind that many more SNPs are being identified in the public and private domains, the total number that are potentially able to disrupt protein structure, while small in comparison with the total genome complement, is still likely to be significant in terms of pharmaceutical research opportunities. This study gave an overview of the total number of SNPs available within the public domain as of the first quarter of 2001.

An analysis of polymorphisms in coding regions was published by Landers group in 1999. They identified 560 SNPs (392 in coding regions) in 106 genes of relevance to cardiovascular disease, neuropsychiatry, and endocrinology. Only a minority of polymorphic changes gave rise to nonconservative amino acid substitutions, most likely due to evolutionary selection against deleterious mutations in the human genome.

Nevertheless, there is considerable activity in identifying SNPs within protein targets for current marketed drugs. The whole purpose of the study of genetic variations in drug responses is to identify patients who may benefit from a particular medicine and avoid wasteful (or dangerous) prescription to others. Examples of drug targets and detoxification systems that have been studied using SNP or other mutational analyses are listed in Table 1.

These comparatively early studies are beginning to highlight a number of points, including the fact that some SNPs are specific for particular ethnic populations. In addition, some drug targets appear not to have amino acid sequence variation as a result of SNP polymorphism, and therefore will not require a number of different ligands to accommodate this. However, since variations in tertiary structure resulting from mutations in the coding region offer opportunities for structure-based design, some relevant examples will be discussed in detail below.

Tab. 1 Examples of SNPs that influence drug metabolism or disease state.

Drug target	*Disease*	*Comments*
NMDAR1 receptor	Schizophrenia	SNPs in coding region, but no functional significance
Quinone oxidoreductase, Sulfotransferase	Drug metabolism	6 out of 22 coding region SNPs gave amino acid substitutions
75 candidate genes for blood pressure homeostasis	Hypertension	874 candidate SNPs with 387 within coding sequence; 54% of these predicted to change protein sequence
Thiopurine *S*-methyltransferase	Drug metabolism	SNPs in promoter region, introns, and 3′UTR, but not coding region
41 candidate genes	Ischemic heart disease	SNPs restricted to ethnic group (Japanese) used in study
CYP3A	Drug metabolism	SNPs cause alternative splicing and protein truncation
β_2-adrenergic receptor	Asthma	SNPs cause alteration of ligand binding
BCR-ABL tyrosine kinase	Cancer	Point mutation causes resistance to STI-571 compound

2 Mutations in Drug Targets Leading to Changes in the Ligand-binding Pocket

2.1 β_2-adrenergic Receptor

This well-characterized drug target for antiasthma medications represents one of the earliest examples of a natural mutation, leading to an alteration in ligand binding. Mutation of Thr164 to Ile in the fourth transmembrane-spanning domain of this G-protein-coupled receptor occurs at low frequency in the population tested, but gives rise to altered ligand binding through interaction with an adjacent Ser165, as illustrated in Fig. 1.

This example is one where the accurate three-dimensional structure of the protein is unknown; under these circumstances, it is necessary to create a computer model. The development of inhibitors that are designed to overcome the effects of this mutation could not be based on the

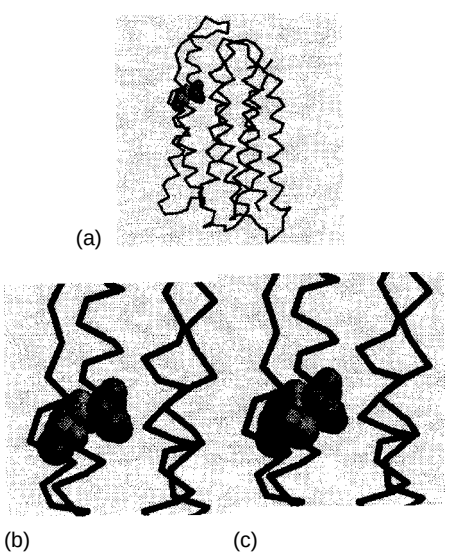

Fig. 1 Mutation of Thr164 to Ile in the fourth transmembrane region of the β_2-adrenergic receptor leads to steric interference with Ser165 within the active site of the receptor. (a) General domain structure and position of the residues, (b) wild type, (c) mutant showing interaction of Ile164 with Ser165.

accurate structure of a ligand-binding site; here, ligand-based design would be appropriate (see Sect.3.3).

2.2 STI-571 and BCR-ABL

STI-571 is a 2-phenylamino pyrimidine chemotherapeutic agent that targets the tyrosine kinase domain of the BCR-ABL oncogene present in chronic myelogenous leukemia (CML). After many years of research into compounds that are selective for specific oncogenes, this compound is showing real clinical benefit against CML. Unfortunately, due to the selective pressure on tumor cells to evolve resistance, a specific mutation in BCR-ABL inhibits the action of STI-571. Since crystal structures of ligand bound into the wild-type protein are available (PDB: 1IEP), it is possible to visualize the effects of a C-T nucleotide mutation, resulting in a substitution of Thr315 by Ile (Fig. 2). The change of amino acid results in disruption of a key hydrogen bond between oxygen on the threonine and a secondary amine nitrogen on STI-571.

In this example, structure-based design of inhibitors to the mutant enzyme may be undertaken using available X-ray structures and homology models.

2.3 Resistance to Human Immunodeficiency Virus (HIV)

The human retrovirus HIV can be controlled using chemotherapy directed at the reverse transcriptase and aspartyl protease

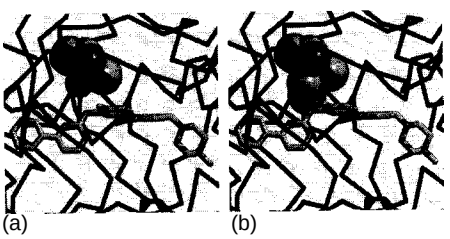

Fig. 2 Mutation of Thr315 (a) to Ile (b) in BCR-ABL interferes with the binding of STI-571.

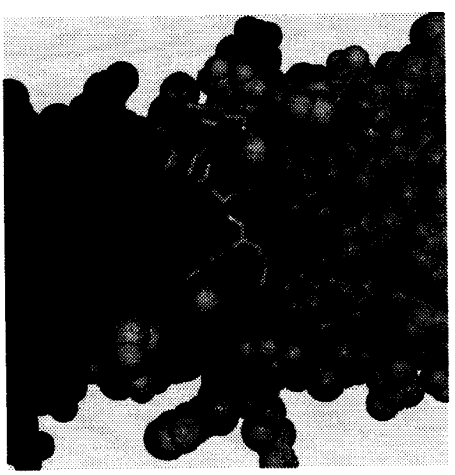

Fig. 3 Binding of the HIV protease inhibitor ritonavir. Amino acids highlighted in cyan are mutated in resistant strains of the virus and tend to occur at the extremities of the inhibitor.

encoded by the viral genome; with other microbial pathogens, however, resistance to drug therapy becomes a major problem. Figure 3 shows a crystal structure (PDB: 1HXW) of the HIV protease, where mutated amino acids (shown in cyan) lead to disrupted binding of the clinically effective inhibitor ritonavir.

3 *In Silico* Design of Small Molecules

Changes induced by SNPs (or other mutations) in the protein targets of drugs may result in the medicine being ineffective. In this case, the relevant disease may remain untreated, or else an expensive drug discovery effort will have to be undertaken to identify new molecules that are able to affect the mutant protein. Given that most current small molecule discovery strategies rely on expensive high-throughput screening and lead optimization programs, this is an unattractive prospect. Advances in computer-aided design, however, may make this process more efficient and less costly through creation of virtual small molecules that can by optimized for interaction with the mutant protein before any chemical synthesis is undertaken. The remainder of this chapter will be concerned with an overview of this exciting and rapidly evolving area of research.

3.1 Automated Drug Design Methods

The pharmaceutical industry faces a large increase in the number of therapeutic targets available for exploration. New methods must be developed for drug discovery that can be automated, are cheap, and that optimally explore the chemical space available for drug design. The number of different possible chemical structures has been estimated to be of the order 10^{200}, with perhaps 10^{60} structures with drug-like properties; these numbers can be compared with the estimated mass of the visible universe at 10^{56} g. Thus, it would not be possible to make a single gram of each potential drug molecule. In contrast, the number of different chemical structures synthesized is small by comparison, and has not yet reached 10^{8}. Clearly, *in silico* methods will have to be developed to cope with searches through the vastness of chemical space.

Design paradigms fall into two types:

- Structure-based methods that use three-dimensional structural information about the site.
- Ligand-based methods that can be used for drug discovery in the absence of structural information.

The applicability of structure-based methods is critically dependent on the availability of relevant structural data. The Structural Genomics Initiative is an international project that seeks to provide high-quality crystalographic or NMR data about new proteins. While promising for future drug discovery efforts, the rate of introduction of new structures into the public domain is still too slow. To speed this up, alternative methods for the generation of three-dimensional data using homology modeling techniques have been introduced. These are less precise, but show increasingly better performance, as elucidated by the computer-aided structure prediction (CASP) challenges (http://PredictionCenter.llnl.gov). The logical next step would be to automate protein structure prediction in order to provide extensive coverage of the human proteome. One of the main drivers for this

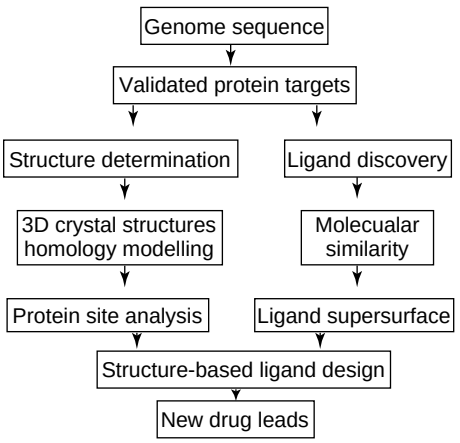

Fig. 4 Scheme for drug design utilizing genomic data. The left track illustrates the process of drug design based on structural_determination of binding sites. The right track outlines design from known ligands.

effort is the need for three-dimensional structural models of whole gene families for use in drug design.

Ligand-based design methods are built upon identified molecular similarities in known active ligands. The similarity methods have to cope with finding unspecified partial molecular similarities within the ligand data sets for freely flexible ligands. There are often many solutions to molecular similarity and choices have to be made as to which one to use. Figure 4 summarizes the two approaches to drug design.

3.2 Structure-based Drug Design

The automation of structure-based design can be divided into key areas for development:

- identification of binding sites
- automated analysis of binding sites
- prioritization of strategies for design within sites
- *de novo* design of scaffolds (active templates)
- elaboration of scaffolds into combinatorial libraries.

The identification of binding sites can be initiated from protein sequence information. Many sites show conservation of structural motifs; for example, many enzymatic catalytic sites have highly conserved local sequences within a gene family. Many of these motifs can be identified at the annotation stage of ascribing a function to a novel gene. Furthermore, the identification of cofactor binding sites can also provide additional clues to function. At the moment, the prediction of the molecular structure of the substrate has not been solved precisely, although docking algorithms could provide clues for potential substrate fitting into binding sites.

Ligands bind to their sites through specific molecular interactions, including hydrogen bonds, electrostatic interactions, steric interactions, interactions through hydration, and hydrophobic interactions. This pattern lies at the surfaces of both ligand and binding site, so that complementary interactions are optimized for efficient binding. If the target proteins have 3D coordinate data, computer algorithms can be used to automatically analyze the binding features and categorize them on a grid map of the protein surface. The features mapped onto grid

points are known as site points. Structurally and functionally related proteins may then be compared by superposing the binding sites and aligning the pattern of chemical features. This facility is important in pharmacogenomics, where we may wish either to design compounds that bind to a class of sites within related proteins, or to design compounds specific for a particular member of the protein family. In the latter case, the aim is to acquire selectivity. Figure 5 illustrates the family of caspase enzymes; caspases 1, 4, and 5 form a separate group – the ICE subfamily, the rest are in the CED-3 subfamily.

The CED-3 subfamily shows very similar three-dimensional structures. Figure 6 shows the superpositions of the Cα atoms. The structural backbones of the proteins within the family are closely superposed.

The concept of dividing sites into a collection of site points for design provides a method for directing design to specific regions. In some respects, this process has parallels to drug design from pharmacophores, where a small number of points are used to search compound collections by virtual screening. Four-point pharmacophores provide three-dimensionality to the search. However, the number of site points encountered in a site, often about 30, creates a combinatorial problem for selection. Suppose that there are n site points; if r site point interactions are

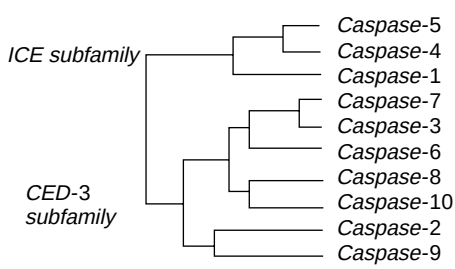

Fig. 5 Sequence similarities of the caspase family.

Fig. 6 Superposed Cα atoms from a selection of proteins in the CED-3 family of caspases.

needed for a molecule to have measurable affinity, the number of combinations of r site points, C(n,r) is given by $C(n,r) = n!/(r!(n-r)!)$ C(n,r) is a maximum when $r = n/2$. For the case where $n = 30$ and $r = 5$, there are 140 000 subsets of 5 site points. Each subset would be a strategy for design. However, some strategies would lead to preferred designs, for example, a strategy that included interactions with the catalytic residues would be expected to consistently perform well in designing inhibitors. Moreover, if there were crystalographic evidence of substrates or known inhibitors binding to certain residues, these would form a basic set for *de novo* design. Other methods for prioritizing design strategies could be built on hydrogen bonding regions, where there are overlaps in hydrogen-bonding probability. In the pharmacogenomic problem of design within a family of sites, the choice of design strategy is paramount for obtaining selectivity. The selection of site points should be such that the set chosen for the target protein should be as different as possible from nearly corresponding sets within the family of proteins.

The affinity of the designed drug molecule for its site is governed by the free energy of interaction DG. DG can be determined experimentally from the equation $\Delta G = -RT \ln K_b$ where K_b is the experimental binding constant, R is the gas constant and T is the temperature. The free energy has an enthalpic and entropic component $\Delta G = \Delta H - T\Delta S$ where ΔH is the change in enthalpy on binding and ΔS is the change in entropy.

De novo methods for drug design attempt to build novel molecules within specified regions of the site. Given the large number of molecules that could be created, it is important to assess the relative differences in predicted binding affinity for the designed ligands and the site. This is achieved using a scoring function based on the equations described above.

A composite scheme for *de novo* design is shown in Fig. 7.

Molecules are built from small molecular fragments, the latter being derived from databases of drug molecules such as the World Drug Index, or from proprietary compound collections. Fragments are labeled for attachment and substitution positions. Molecules are then built into the site by a stochastic assembly process from randomly selected fragments. This process is controlled by chemical rules that govern the addition, removal, or exchange of a fragment on the evolving skeleton. At each modification, the assembled structure is fitted in the site by positional and conformational transformations, with the scoring function being

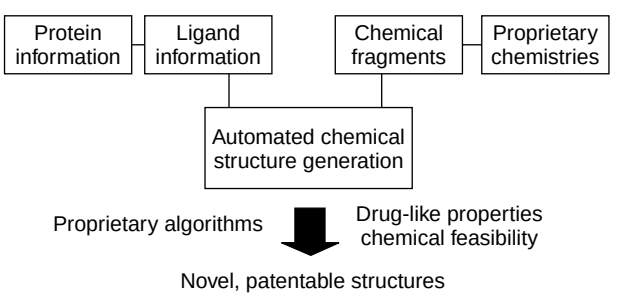

Fig. 7 Scheme for *de novo* design.

used to monitor the prediction of the free energy of the interaction between ligand and site. An optimal assembly procedure is performed using combinatorial optimization routines, such as simulated annealing or genetic algorithms.

Simulated annealing is an ergodic optimization method, meaning that, given sufficient time, the algorithm converges on the optimal solution irrespective of the initial starting position. If time constraints are placed on the algorithm run time, it behaves nonergodically and different solutions can be obtained that are acceptable, but suboptimal. In structure assembly, this is important, since a variety of molecular structures can be produced, thus giving a choice of different chemistries. These solutions can be analyzed statistically to identify different classes of structures that fit a chosen starting strategy.

Constraints can be placed on the structure assembly to limit the size of the molecule being generated so that it can be used as a template for further decoration by combinatorial chemistry. Virtual combinatorial chemical libraries can be created round the designed template to explore regions close to where the template has been designed to bind. The in-site enumeration of combinatorial libraries designed for a particular site is very important for chemical genomics, since it offers a way of designing molecules to distinguish between mutational differences in a particular site.

3.3 Ligand-based Drug Design

In the absence of useful three-dimensional information about the site, the designer can only build novel molecular structures on the basis of molecular similarity with existing active molecules. Ligand-based design requires the resolution of a number of technical issues. These are itemized below:

- A set of molecules is needed to extract spatially distributed chemical features for inclusion in drug design.
- Structurally dissimilar, active molecules provide more information about the site than a series of active close structural homologs.
- The molecules in the initial ligand set may have numerous torsion angles, making identification of the active conformation difficult.
- The choice of molecular similarity procedure needs to be made from a wide variety of available methods, preferably including one that includes a search for partial molecular similarity. This allows a more useful set of similarities to be identified.
- Molecules within the ligand set could have different binding modes, which need to be identified before selecting a subset for a design strategy.
- Molecules within the ligand set, with similar binding modes, need to be superposed such that the site points inferred from the corresponding ligand points show common spatial positions.
- The constraining supersurface of the set of molecules needs to be constructed to provide limits for automated design procedures.
- Since there is no information about the site, a scoring function based on free energy methods cannot be used.

A molecular similarity method that is applicable to this method of design has recently been published as the algorithm SLATE. This algorithm takes a set of molecules and compares them in a pairwise fashion for molecular similarity. The

molecules are allowed to flex. In the case of potential hydrogen-bonding interactions between the ligand points and possible complementary site points, site points are projected from the ligand surfaces. Hydrophobic regions are handled as local regional centroids. The molecular description is reduced to a set of points with associated properties. These sets of points can be matched using distance geometry and the search optimized by simulated annealing through conformational space. The procedure generates a similarity matrix; partial similarity is handled by null correspondences.

An application of this procedure to the design of novel compounds for the histamine H3 receptor has been published. Similarity data suggested that four hydrogen-bonding site points, together with two lipophilic regions, could be identified with a common spatial distribution. Furthermore, the constraining supersurface showed strong similarity with each of the selected ligand conformations. From these *in silico* designs, histamine H3 antagonists were synthesized and showed high affinity for the site. Thus, computational methods could be used to design potent ligands without any prior knowledge of the sequence or structure of the receptor.

4 Future Directions

New developments in pharmacogenomics will impact on drug design at three main levels:

- The interaction of the drug with its receptor binding site
- The absorption and distribution of the drug
- The elimination of the drug from the body.

If we are ever to achieve personalized drug therapies, the above issues will have to be addressed.

The bulk of this chapter has discussed SNPs in terms of mutated drug binding sites and the resulting changes that occur in response to medicines. In each of these cases, drug design strategies may need to be modified to develop drug molecules that would be selective and effective in the mutated site.

Absorption, in general, is treated as a physicochemical-transport process based on computations of logP (the octanol/water partition coefficient) and solubility governed by factors such as polar surface area on the molecule. It is conceivable that SNPs in drug transporter genes will affect the pharmacokinetic properties of compounds and, therefore, these may have to be taken into consideration in the design process.

Elimination mechanisms may include active excretion of drug molecules from cells, as with multiple drug resistance that arises in cancer chemotherapy. Patients with genotypes for particular MDR pathways could be identified and alternative treatments provided. Pharmacogenetic differences in the response to drugs can often be related to population differences in drug metabolizing enzymes. Crystal structures or homology models of these enzymes can be used to screen compounds designed *de novo* before synthesis has begun. This can be done at two levels: The virtual compounds can be input into metabolism prediction programs, such as Metabolexpert or Meteor, to identify principal pathways of expected drug metabolism. The virtual compounds can be screened against structural models of the metabolizing enzymes, including the known SNP variants. These procedures are becoming widely adopted

for the cytochrome P_{450} isozymes involved in oxidative drug metabolism.

Genomic prescreening of patients for SNP mutations in drug metabolism will improve the utility of clinical trials by focusing attention on the response of specific subpopulations to drug treatment. The consequence of this approach is that it should be possible to industrialize the creation of individual therapies, although at a significant increase in cost.

Pharmacogenomics will have a major influence on drug discovery, through drug design, as well clinical practice. The subtle changes in biomolecular structure caused by small SNP-induced amino acid alterations will profoundly affect the search for personalized medicines. In the future, greater emphasis will be placed on very precise differences in drug structures that select between mutated proteins. Along with being scientifically challenging, this approach will clearly have significant economic effects on the pharmaceutical industry.

See also Antitumor Steroids; Cancer Chemotherapy, Theoretical Foundations of; HIV Therapeutics, Biochemistry of; Medicinal Chemistry; Oncology, Molecular; Pharmacokinetics and Pharmacodynamics of Biotech Drugs.

Bibliography

Books and Reviews

Bailey, D., Brown, D. (2001) High-throughput chemistry and structure-based design: survival of the smartest, *Drug Discov. Today* **6**, 57–59.

Burley, S.K. (2000) An overview of structural genomics, *Nat. Struct. Biol.* **7**(Suppl.), 932–934.

Dean, P.M. (1995) Defining Molecular Similarity and Complementarity for Drug Design, in: Dean, P.M. (Ed.) *Molecular Similarity in Drug Design*, Blackie Academic & Professional, Oxford, UK 1–23.

Drews, J.J. (2000) Drug discovery: A historical perspective, *Science* **287**, 1960–1964.

Evans, W.E., Relling, M.V. (1999) Pharmacogenomics: translating functional genomics into rational therapeutics, *Science* **286**, 487–491.

Maggio, E.T., Ramnarayan, K. (2001) Recent developments in computational proteomics, *Trends Biotechnol.* **19**, 266–272.

Roses, A.D. (2000) Pharmacogenetics and the practice of medicine, *Nature* **405**, 857–865.

Primary Literature

Boden, D., Markowitz, M. (1998) Resistance to human immunodeficiency virus type 1 protease inhibitors, *Antimicrob. Agents Chemother.* **42**, 2775–2783.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G.Q., Lander, E.S. (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes, *Nat. Genet.* **3**, 231–238.

Darvas, F. (1988) Predicting metabolic pathways by logic programming, *J. Mol. Graph.* **6**, 80–86.

De Esch, I.J.P., Mills, J.E.J., Perkins, T.D.J., Romeo, G., Hoffmann, M., Wieland, K., Leurs, R., Menge, W.M.P.B., Nederkoorn, P.H.J., Dean, P.M., Timmerman, H. (2001) Development of a pharmacophore model for histamine H3 receptor antagonists, using the newly developed molecular modeling program SLATE, *J. Med. Chem.* **44**, 1666–1674.

Druker, B.J., Sawyers, C.L., Kantarjian, H., Resta, D.J., Reese, S.F., Ford, J.M., Capdeville, R., Talpaz, M. (2001) Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome, *N. Engl. J. Med.* **344**, 1084–1086.

Druker, B.J., Tamura, S., Buchdunger, E., Ohno, S., Segal, G.M., Fanning, S., Zimmermann, J.,

Lydon, N.B. (1996) Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells, *Nat. Med.* **2**, 561–566.

Gorre, M.E., Mohammed, M., Ellwood, K., Hsu, N., Paquette, R., Rao, P.N., Sawyers, C.L. (2001) Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification, *Science* **293**, 876–880.

Green, S.A., Cole, G., Jacinto, M., Innes, M., Ligett, S.B. (1993) A polymorphism of the human β_2 adrenergic receptor within the fourth transmembrane domain alters ligand binding and functional properties of the receptor, *J. Biol. Chem.* **268**, 23116–23121.

Greene, N., Judson, P.N., Langowski, J.J., Marchant, C.A. (1999) Knowledge-based expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR. SAR and QSAR, *Environ. Res.* **10**, 299–314.

Halushka, M.K., Fan, J.B., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R., Chakravarti, A. (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis, *Nat. Genet.* **22**, 239–247.

Iida, A., Sekine, A., Saito, S., Kitamura, Y., Kitamoto, T., Osawa, S., Mishima, C., Nakamura, Y. (2001) Catalog of 320 single nucleotide polymorphisms (SNPs) in 20 quinone oxidoreductase and sulfotransferase genes, *J. Hum. Genet.* **46**, 225–240.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Klebe, G. (2000) Virtual screening: An alternative or complement to high throughput screening, *Perspect Drug Discov.* **20**, 9 (preface).

Kuehl, P., Zhang, J., Lin, Y., Lamba, J., Assem, M., Schuetz, J., Watkins, P.B., Daly, A., Wrighton, S.A., Hall, S.D., Maurel, P., Relling, M., Brimer, C., Yasuda, K., Venkataramanan, R., Strom, S., Thummel, K., Boguski, M.S., Schuetz, E. (2001) Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression, *Nat. Genet.* **27**, 383–391.

Marth, G., Yeh, R., Minton, M., Donaldson, R., Li, Q., Duan, S., Davenport, R., Miller, R.D., Kwok, P.Y. (2001) Single-nucleotide polymorphisms in the public domain: how useful are they? *Nat. Genet.* **4**, 371–372.

Mills, J.E.J., De Esch, I.J.P., Perkins, T.D.J., Dean, P.M. (2001) SLATE: a method for the superposition of flexible ligands, *J. Comput. Aided Mol. Design* **15**, 81–96.

Ohnishi, Y., Tanaka, T., Yamada, R., Suematsu, K., Minami, M., Fujii, K., Hoki, N., Kodama, K., Nagata, S., Hayashi, T., Kinoshita, N., Sato, H., Sato, H., Kuzuya, T., Takeda, H., Hori, M., Nakamura, Y. (2000) Identification of 187 single nucleotide polymorphisms (SNPs) among 41 candidate genes for ischemic heart disease in the Japanese population, *Hum. Genet.* **106**, 288–292.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M., Miyano, M. (2000) Crystal structure of rhodopsin: a G protein-coupled receptor, *Science* **289**, 739–745.

Rice, S.R., Niu, N., Berman, D.B., Heston, L.L., Sobell, J.L. (2001) Identification of single nucleotide polymorphisms (SNPs) and other sequence changes and estimation of nucleotide diversity in coding and flanking regions of the NMDAR1 receptor gene in schizophrenic patients, *Mol. Psychiatry* **6**, 274–284.

Richman, D.E. (2001) HIV chemotherapy, *Nature* **410**, 995–1001.

Seki, T., Tanaka, T., Nakamura, Y. (2000) Genomic structure and multiple single-nucleotide polymorphisms (SNPs) of the thiopurine S-methyltransferase (TPMT) gene, *J. Hum. Genet.* **45**, 299–302.

The International SNP Map Working Group (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms, *Nature* **409**, 928–933.

Venter, J.C., Adams, M.D., Myers, E.W., et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Pharmacokinetics and Pharmacodynamics of Biotech Drugs

Bernd Meibohm[1] *and Hartmut Derendorf*[2]
[1] *The University of Tennessee Health Science Center, Memphis, TN*
[2] *University of Florida, Gainesville, FL*

1 Introduction 214

2 Bioanalytical Challenges 215

3 Pharmacokinetics of Peptides and Proteins 217
3.1 Absorption 217
3.2 Distribution 218
3.3 Protein Binding 219
3.4 Elimination 219
3.4.1 Proteolysis 220
3.4.2 Gastrointestinal Elimination 221
3.4.3 Renal Elimination 221
3.4.4 Hepatic Elimination 221
3.4.5 Receptor-mediated Elimination 223
3.5 Species Specificity and Allometry 224
3.6 Immunogenicity 224

4 Pharmacokinetics of Oligonucleotides 226

5 Pharmacokinetics of DNA 227

6 Exposure/response Correlations for Biotech Drugs 228
6.1 Direct Link PK/PD Models 229
6.2 Indirect Link PK/PD Models 229
6.3 Indirect-response PK/PD Models 230
6.4 Precursor Pool PK/PD Models 232
6.5 Complex PK/PD Models 232

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

7 Summary 235

Bibliography 235
Books and Reviews 235
Primary Literature 236

Keywords

Biotech Drug
Biotechnologically derived drug products including proteins, peptides, antibodies, and antibody fragments, as well as antisense oligonucleotides and DNA preparations for gene therapy.

Absorption
Process by which the unchanged drug proceeds from the site of administration to the systemic circulation, that is, the blood or plasma compartment.

Distribution
Reversible transfer of drug between different organs, tissues, and fluids within the body.

Elimination
Irreversible loss of drug from the systemic circulation, that is, the blood or plasma compartment. Elimination comprises excretion and metabolism.

Excretion
Irreversible loss of chemically unchanged drug from the body.

Metabolism
Sum of all chemical reactions involved in biotransformation of a drug, that is, conversion of one chemical species into another within a living organism.

Clearance (CL)
The hypothetical volume of blood or plasma from which the drug is cleared per time unit by elimination mechanisms.

Volume of Distribution (V_d)
The hypothetical volume of a biological fluid that would be required to dissolve the total amount of drug within the body to achieve the same concentration as that found in

blood or plasma. It is a proportionality constant relating the amount of drug in the body to the concentration measured in the systemic circulation, that is, the blood or plasma compartment. Volume of distribution provides some general information about a drug's extent of distribution into body tissues. Together with clearance, it determines the elimination half-life $t\frac{1}{2}$ of a drug (*via* $t\frac{1}{2} = 0.693 \times V_d/CL$).

Exposure /Response Correlation
The relationship between exposure and response to a drug. Exposure comprises drug input to the body as quantified by acute or integrated measures of drug concentrations in blood, plasma, or other biological fluid. Response includes biomarkers, potential or accepted surrogate endpoints and short-term or long-term clinical effects related to either efficacy or safety.

PK/PD Model
A mathematically defined relationship between dose, concentration, and effect of a drug in a biological system. Pharmacokinetic/pharmacodynamic (PK/PD) modeling links the dynamic change in concentration over time as assessed by pharmacokinetics to what is, in most cases, the static relationship between the concentration at the effect site and the intensity of observed response as quantified by pharmacodynamics.

In recent years, biotechnologically derived peptide- and protein-based drugs have developed into mainstream therapeutic agents. Peptide and protein drugs, as well as oligonucleotides and DNA, now constitute a substantial portion of the compounds under preclinical and clinical development in the global pharmaceutical industry. The pharmacokinetic and pharmacodynamic properties of a biotech drug determine the relationship between administered dose, resulting systemic exposure, and subsequent pharmacologic response. Pharmacokinetics, and pharmacodynamics are, therefore, frequently crucial determinants of a drug's efficacy and safety, and their evaluation is an important component during the preclinical and clinical development of a drug candidate. However, pharmacokinetic and exposure/response evaluations for peptides and protein therapeutics are frequently complicated by their similarity to endogenous peptides and proteins as well as protein nutrients. In addition, pharmacokinetic/pharmacodynamic (PK/PD) correlations for biotech drugs are frequently convoluted by their close interaction with endogenous substances and physiologic regulatory feedback mechanisms. Extensive use of pharmacokinetic and exposure/response concepts in all phases of drug development has, in the past, been identified as a crucial factor for the success of a scientifically driven, evidence-based, and hence accelerated drug development process. Thus, PK/PD concepts are likely to continue and expand their role as a fundamental factor in the successful development of biotechnologically derived drug products in the future.

1 Introduction

In the last two decades, an increasing fraction of pharmaceutical R&D has been devoted to biotechnology-derived drugs (biotech drugs) – large molecules such as soluble proteins, monoclonal antibodies, and antibody fragments, as well as smaller peptides, antisense oligonucleotides, and DNA preparations for gene therapy. Biotech and genomics companies currently perform nearly one-fifth of all pharmaceutical R&D, a figure that is set to double within the next ten years. These biotech-related drug development efforts have so far been quite successful. Biotech products accounted for more than 35% of the 37 new active substances (NASs) that were launched in 2001, and it has been predicted that half of all NASs developed in the next 10 to 15 years will result from research into antibodies alone. Numerous approved biotech drug products have revolutionized pharmacotherapy in various indications, including erythropoietin (Epogen®, Procrit®), abciximab (ReoPro®), and trastuzumab (Herceptin®) to name only a few. Since the development of biotech drugs generally rests on a fundamental understanding of the related disease, their clinical development has also proven to be more successful than for conventional small molecules (new molecular entities: NCEs). Only 8% of the NCEs that entered clinical drug development between 1996 and 1998 reached the market compared with 34% in the case of biotech drugs. On the basis of these facts, it can be predicted that biotech drugs will play a major, if not dominant, role in the drug development arena of the coming decades. Thus, biotech-based medications might serve as the key for the aspired "personalized medicine" in the health care systems of the future.

The basis for the pharmacotherapeutic use of biotech drugs is similar to that for small molecules – a defined relationship between the intensity of the therapeutic effect and the amount of drug in the body or, more specifically, the drug concentration at its site of action, that is, the exposure–response relationship. The relationship between the administered dose of a drug, the resulting concentrations in body fluids, and the intensity of produced outcome may be either simple or complex, and thus either obvious or hidden. However, if no simple relationship is obvious, it would be misleading to conclude *a priori* that no relationship exists at all, rather than that it is not readily apparent.

The dose–concentration–effect relationship is defined by the pharmacokinetic (PK) and pharmacodynamic (PD) characteristics of a drug. PKs comprises all processes that contribute to the time course of drug concentrations in various body fluids, generally blood or plasma, that is, all processes affecting drug absorption, distribution, metabolism, and excretion. In contrast, PDs characterizes the effect intensity and/or toxicity resulting from certain drug concentrations at the assumed effect site. To put it simply, PKs characterizes "*what the body does to the drug*," whereas PDs assesses "*what the drug does to the body*." The combination of both pharmacological disciplines by integrated pharmacokinetic/pharmacodynamic modeling (PK/PD modeling) allows a continuous description of the effect-time course directly resulting from the administration of a certain dose (Fig. 1).

PK and PD principles are as applicable to biotech drugs such as peptides, proteins, and oligonucleotides as they are to

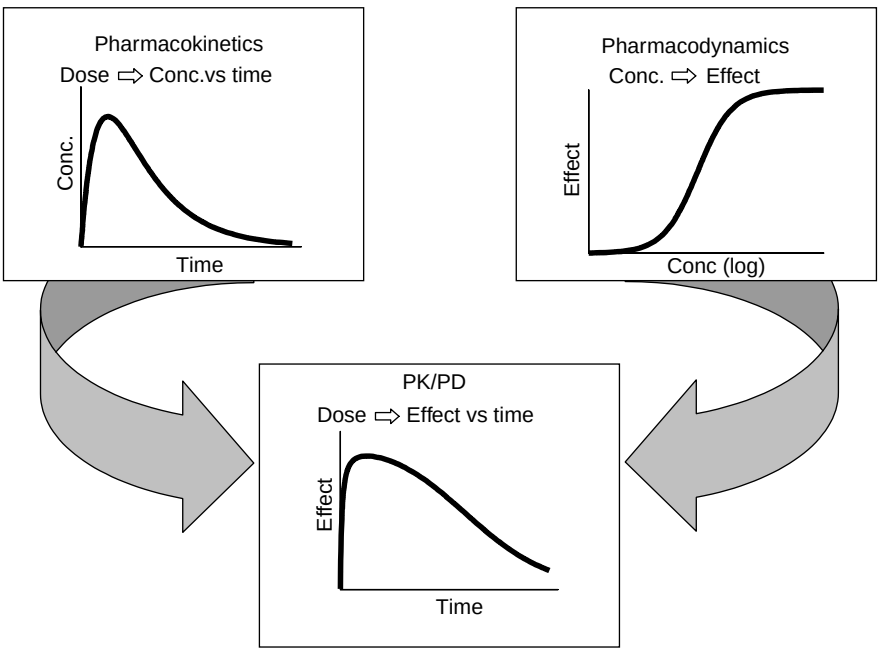

Fig. 1 Pharmacokinetic/pharmacodynamic (PK/PD) modeling as combination of the classic pharmacological disciplines, pharmacokinetics and pharmacodynamics (Reprinted with permission from Derendorf, H., Meibohm, B. (1999) Modeling of pharmacokinetic/pharmacodynamic (PK/PD) relationships: concepts and perspectives, *Pharm. Res.* **16**, 176–185).

conventional small-molecule drugs. These also include PK/PD related recommendations for drug development, such as the recently published exposure–response guidance document of the US Food and Drug Administration and the ICH E4 guideline of the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. Since biotech drugs are frequently identical or similar to endogenous substances, they oftentimes exhibit unique PK and PD properties. These may pose extra challenges and questions during their preclinical and clinical drug development that are different from small-molecule drug candidates and may require additional resources and unique expertise. Some of these problems and challenges will be discussed in the following text.

2 Bioanalytical Challenges

The availability of an accurate, precise, and specific bioanalytical technique for the quantification of active drug moieties in plasma, blood, or other biological fluids is an essential prerequisite for the evaluation of the relationship between dose, concentration, and effect of biotech drugs. In analogy to small molecules, these analytical techniques have to be validated and have to meet prespecified criteria regarding accuracy, precision, selectivity,

sensitivity, reproducibility, and stability – for example, those recommended by the US Food and Drug Administration. Additional requirements for bioanalytical method validation for macromolecules have recently been published.

A further level of complexity is often added to the bioanalytics of biotech drugs by the fact that many of these compounds are endogenous substances that are already in the body before drug administration. Thus, the analytical technique will detect a so-called baseline level prior to drug exposure. This baseline level can either be constant or undergo complex changes, for example, circadian rhythms or irregular time courses. In order to characterize the clinical pharmacology of biotech drugs naturally present in the body, baseline values have either to be accounted for in the PK and PD analysis or be suppressed before exogenous drug administration. The suppression of endogenous baseline levels is oftentimes facilitated *via* physiological regulation or feedback mechanisms. This approach was, for example, used to suppress the endogenous release of insulin and somatotropin (growth hormone, GH) *via* infusions of glucose and somatostatin respectively, prior to their exogenous administration for evaluation of their PKs.

In contrast to the bioanalytics of small-molecule drugs, immunoassays and bioassays are frequently applied to quantify peptides and proteins in biological samples. Immunoassays are considered to be the analytical method of choice for concentration determinations of peptides and proteins, as they can be performed relatively rapidly and easily. These generally comprise enzyme immunoassays as well as radioimmunoassays with poly- and monoclonal antibodies. While both techniques have a sufficient sensitivity and reproducibility, their specificity for the active drug compound to be quantified has to be carefully evaluated during the assay validation process. Immunoassays may be insensitive to relatively minor changes in the primary or secondary structure of proteins. For recombinant interferon-γ, for example, bioavailability was reported to be $>100\%$ for subcutaneous compared to intravenous administration, a finding that was produced by assay artifacts due to slightly modified degradation products. Other potential interferences with immunoassays include matrix effects, specific binding proteins, proteases, and cross-reactivity toward endogenous proteins.

Bioassays are frequently used as an alternative, or in addition to immunoassay techniques. Bioassays, in contrast to immunoassays, quantify not the pharmacologically active substance, but its biological activity, for example, in cell culture models based on cell differentiation, cell proliferation, or cytotoxicity, as well as gene expression assays or whole animal models. Frequent major problems with bioassays include a high variability in the measured parameters, lack of precision, and their time- and labor-intensive nature. Furthermore, bioassays oftentimes also lack specificity for the measured compound, as they may also detect the response to bioactive metabolites.

Because of some of the problems with bioassays and immunoassays, liquid chromatography (LC) based techniques are increasingly being applied as an alternative. While modern LC-based assays have a comparable sensitivity to immunoassays, they are oftentimes characterized by a higher selectivity. LC/mass spectrometry with matrix-assisted laser desorption ionization (MALDI) was, for example, used in *ex vivo* PK studies in combination with

enzyme-inhibition experiments to investigate the complex metabolism of dynorphin A1-13, a peptide with opioid activity, up to the fifth metabolite generation.

Biodistribution studies for peptides are frequently performed with radioactively labeled compounds. The radioactivity can be introduced either by external labeling with ^{125}I or by internal labeling of already present atoms, for example, *via* the addition of ^{3}H, ^{14}C, or ^{35}S radioactively labeled amino acids to cell cultures producing recombinant proteins. External labeling chemically modifies the protein, which may affect its activity and PKs. Internal labeling circumvents these potential problems, but has the disadvantage that radioactively labeled amino acids might be reused in the endogenous amino acid pool. Thus, when using radioactive labeling, it is generally necessary to investigate whether the physicochemical and biological properties of the proteins remain unchanged. In addition, it is crucial to determine whether the measured radioactivity represents intact protein, labeled metabolites, or the released label itself. Radioactivity that can be precipitated with trichloroacetic acid, for example, can be used to delineate active protein from released label and metabolites of small molecular weight.

3 Pharmacokinetics of Peptides and Proteins

Although traditional PK principles are also applicable for peptides and proteins, their *in vivo* disposition is, to a large degree, affected by their physiological function. Peptides, for example, which frequently show hormone activity, usually have short elimination half-lives, which is desirable for a close regulation of their endogenous levels and, thus, their function. By contrast, transport proteins like albumin or antibodies have elimination half-lives of several days, which enables and ensures the continuous maintenance of necessary concentrations in the bloodstream. The reported terminal half-life for SB209763, a humanized respiratory syncytial virus monoclonal antibody, for example, was reported as 22 to 50 days.

3.1 Absorption

Traditionally, the largest obstacle for successful pharmacotherapy with peptide and protein drugs is their delivery to the desired site of action. A clinically usable absorption of exogenously applied peptides and protein after oral application with conventional dosage forms is usually not present. This lack of systemic bioavailability is caused mainly by two factors: high-gastrointestinal enzyme activity and the function of the gastrointestinal mucosa as an absorption barrier. There is substantial peptidase and protease activity in the gastrointestinal tract, making it the most efficient body compartment for peptide and protein metabolism. Furthermore, the gastrointestinal mucosa presents a major absorption barrier for water-soluble macromolecules like peptides and protein. This is, at least for peptides, complemented by the functional system of cytochrome P450 3A and *p*-glycoprotein activity.

Because of the lack of activity of most peptides and proteins when orally administered, nonoral pathways have been utilized, in addition to parental administration. These pathways include nasal, buccal, rectal, vaginal, percutaneous, ocular, and pulmonary drug delivery. Drug delivery *via* these administration routes, however,

is also frequently accompanied by presystemic degradation processes. The bioavailability of numerous peptides and proteins is, for example, markedly reduced after subcutaneous, or intramuscular administration compared to their intravenous administration. The pharmacokinetically derived apparent absorption rate constant is thus the combination of absorption into the systemic circulation and presystemic degradation at the absorption site. The true absorption rate constant k_a can then be calculated as

$$k_a = F \cdot k_{app} \quad (1)$$

where F is the bioavailability compared to intravenous administration. A rapid apparent absorption rate constant k_{app} can thus be the result of a slow absorption and a fast presystemic degradation, that is, a low systemic bioavailability.

3.2 Distribution

Whole body distribution studies are essential for classical small-molecule drugs in order to exclude tissue accumulation of potentially toxic metabolites. This problem does not exist for protein drugs in which catabolic degradation products are amino acids recycled in the endogenous amino acid pool. Therefore, biodistribution studies for peptides and proteins are primarily performed to assess targeting to specific tissues as well as to identify the major elimination organs.

The volume of distribution of proteins is usually small and limited to the volume of the extracellular space because of their high molecular weight and the related limited mobility because of impaired passage through biomembranes. After intravenous application, peptides and proteins usually follow a biexponential plasma concentration–time profile that can best be described by a two-compartment PK model. This has, for example, been described for leuprorelin, a synthetic agonist analog of luteinizing hormone-releasing hormone (LH-RH), for clenoliximab, a macaque–human chimeric monoclonal antibody specific to the CD4 molecule on the surface of T-lymphocytes, and for AJW200, a humanized monoclonal antibody to von Willebrand factor. The central compartment in this model represents primarily the vascular space, and the interstitial space of well-perfused organs with permeable capillary walls, especially liver and kidneys, while the peripheral compartment comprises the interstitial space of poorly perfused tissues like skin and (inactive) muscle.

Thus, the volume of distribution of the central compartment in which peptides and proteins initially distribute after intravenous administration is typically 3 to 8 L, which is approximately equal to or slightly higher than the plasma volume (approximate body water volumes for a 70-kg person: interstitial 12 L, intracellular 27 L, intravascular 3 L). The total volume of distribution (V_d) is frequently 14 to 20 L, not more than twice the initial volume of distribution (V_c). This distribution pattern has, for example, been described for the somatostatin analog octreotid (V_c of 5.2–10.2 L, V_d of 18–30 L), the t-PA analog tenecteplase (V_c of 4.2–6.3 L, V_d of 6.1–9.9 L), and the glycoprotein IIb//IIIa inhibitor eptifibatide (V_c of 9.2 L). Active tissue uptake and binding to intra- and extravascular proteins, however, can substantially increase the volume of distribution of peptide and protein drugs, as, for example, observed with atrial natriuretic peptide (ANP).

There is a tendency for V_d and V_c to correlate with each other, which implies

that the volume of distribution is predominantly determined by distribution in the vascular and interstitial space as well as unspecific protein binding in these distribution spaces. The distribution rate is inversely correlated with molecular size and is similar to that of inert polysaccharides, suggesting that passive diffusion through aqueous channels is the primary distribution mechanism.

Distribution, elimination, and PDs are, in contrast to conventional drugs, frequently interrelated for peptides and proteins. The generally low volume of distribution should not necessarily be interpreted as low tissue penetration. Receptor-mediated specific uptake into the target organ, as one mechanism, can result in therapeutically effective tissue concentrations despite a relatively small volume of distribution. Nartogastrim, a recombinant derivative of the granulocyte-colony stimulating factor (G-CSF), for example, is characterized by a specific, dose-dependent, and saturable tissue uptake into the target organ bone marrow, presumably *via* receptor-mediated endocytosis.

3.3 Protein Binding

It is a general PK principle, which is also applicable to peptides and proteins, that only the free, unbound fraction of a drug substance is accessible to distribution and elimination processes as well as interactions with its target structure (e.g. receptor) at the site of action. Hence, the activity of a drug is better characterized by its free rather than total concentration if there is no constant relationship between the two.

Physiologically active endogenous peptides and proteins are frequently interacting with specific binding proteins that are involved in their transport and regulation. Furthermore, interaction with binding proteins may enable or facilitate cellular uptake processes and thus affect the drug's PDs. Specific binding proteins were, for example, described for IGF-1 (insulin-like growth factor), t-PA, interleukin-2, and somatotropin.

Six specific binding proteins were identified for IGF-1, with one binding at least 95% of IGF-1 in plasma. Since the binding affinity of IGF-1 to this binding protein is substantially higher than IGF receptors, the binding protein is assumed to have a reservoir function that protects the body from insulin-like hypoglycemia. Furthermore, the elimination half-life for bound IGF-1 is significantly longer than for free IGF-1, since only the unbound IGF-1 is accessible to elimination *via* glomerular filtration or peritubular extraction.

Somatotropin, another example, has at least two binding proteins in plasma. This protein binding substantially reduces somatotropin elimination with a tenfold smaller clearance of total compared to free somatotropin, and also decreases its activity *via* reduction of receptor interactions.

Apart from these specific bindings, peptides and proteins may also be nonspecifically bound to plasma proteins. For example, metkephamid, a met-enkephalin analog, was described to be 44 to 49% bound to albumin, and octreotid is up to 65% bound to lipoproteins.

3.4 Elimination

Peptide and protein drugs are nearly exclusively metabolized *via* the same catabolic pathways as endogenous or dietetic proteins, leading to amino acids that are reutilized in the endogenous amino acid

pool for the *de novo* biosynthesis of structural or functional body proteins.

Nonmetabolic elimination pathways such as renal or biliary excretion are negligible for most peptides and proteins. Amino acids as well as some peptides and proteins such as immunoglobulin A, however, are excreted into the bile. For octreotid, biliary excretion is, at least in rat and dog, an important elimination pathway. If biliary excretion of peptides and proteins occurs, it generally results in subsequent metabolism of these compounds in the gastrointestinal tract (see Sect. 3.4.2).

The elimination of peptides and proteins can occur unspecifically, nearly everywhere in the body, or can be limited to a specific organ or tissue. Locations of intensive peptide and protein metabolism are liver, kidneys, gastrointestinal tissue, and also blood and other body tissues. Molecular weight determines the major metabolism site as well as the predominant degradation process (Table 1).

The metabolism rate generally increases with decreasing molecular weight from large to small proteins to peptides, but is also dependent on other factors like secondary and tertiary structure as well as glycosylation. The clearance of a peptide or protein describes the irreversible removal of active substance from the extracellular space, which also includes cellular uptake besides metabolism. Because of the unspecific degradation of numerous peptides and proteins in blood, clearance can exceed cardiac output, that is, $>5\ \mathrm{L\ min^{-1}}$ for blood clearance and $>3\ \mathrm{L\ min^{-1}}$ for plasma clearance. Investigations of the detailed metabolism of peptides and proteins are relatively difficult because of the myriad molecule fragments that may be formed.

3.4.1 Proteolysis

Proteolytic enzymes such as proteases and peptidases are ubiquitously available throughout the body, but are especially localized in blood, in the vascular endothelium, and also on cell membranes and

Tab. 1 Molecular weight as major determinant of the elimination mechanisms of peptides and proteins. As indicated, mechanisms may overlap. Endocytosis may occur at any molecular weight range (Modified from McMartin, C. (1992) *Adv. Drug Res.* **22**, 39–106; Braeckman, R. (1997) Pharmacokinetics and pharmacodynamics of peptide and protein drugs, in: Crommelin, D., Sindelar, R. (Eds.) *Pharmaceutical Biotechnology*, Harwood Academic Publishers, Amsterdam, pp. 101–122).

Molecular weight	*Elimination site*	*Predominant elimination mechanisms*	*Major determinant*
<500	Blood, liver	Extracellular hydrolysis, passive lipoid diffusion	Structure, lipophilicity
500–1000	Liver	Carrier-mediated uptake Passive lipoid diffusion	Structure, lipophilicity
1000–50 000	Kidney	Glomerular filtration and subsequent degradation processes (see Fig. 2)	Molecular weight
50 000–200 000	Kidney, liver	Receptor-mediated endocytosis	Sugar, charge
200 000–400 000		Opsonization	α2-macroglobulin, IgG
>400 000		Phagocytosis	Particle aggregation

within cells. Thus, intracellular uptake is *per se* more an elimination rather than a distribution process. While peptidases and proteases in the gastrointestinal tract and in lysosomes are relatively unspecific, soluble peptidases in the interstitial space and exopeptidases on the cell surface have a higher selectivity and determine the specific metabolism pattern of an organ. The proteolytic activity of subcutaneous tissue, for example, results in a partial loss of activity of subcutaneously administrated interferon-γ compared to intravenously administrated interferon-γ.

3.4.2 Gastrointestinal Elimination

For orally administered peptides and proteins, the gastrointestinal tract is the major site of metabolism. Presystemic metabolism is the primary reason for their lack of oral bioavailability. Parenterally administered peptides and proteins, however, may also be metabolized in the intestinal mucosa following intestinal secretion. At least 20% of the degradation of endogenous albumin takes place in the gastrointestinal tract.

3.4.3 Renal Elimination

For parenterally administered and endogenous peptides and proteins, the kidneys are the major elimination organ if they are smaller than the glomerular filtration limit of ~60 kD, although the effective molecule radius is probably the limiting factor. The importance of the kidney as an elimination organ could, for example, be shown for interleukin-2, M-CSF (human macrophage colony-stimulating factor) and interferon-α.

Various renal processes contribute to the elimination of peptides and proteins (Fig. 2). For most substances, glomerular filtration is the dominant, rate-limiting step, as subsequent degradation processes are not saturable under physiologic conditions. Hence, the renal contribution to the overall elimination of peptides and proteins is reduced if the metabolic activity for these proteins is high in other body regions, and it becomes negligible in the presence of unspecific degradation throughout the body. In contrast to that, the renal contribution to total clearance approaches 100% if the metabolic activity is low in other tissues or if distribution is limited. For recombinant IL-10, for instance, elimination correlates closely with glomerular filtration rate, making dosage adjustments necessary in patients with impaired renal function.

After glomerular filtration, small linear peptides undergo intraluminal metabolism, predominantly by exopeptidases in the luminal brush border membrane of the proximal tubules. The resulting amino acids are transcellularly transported back into the systemic circulation. Larger peptides and proteins are actively reabsorbed in the proximal tubules *via* endocytosis. This cellular uptake is followed by addition of lysosomes and hydrolysis to peptide fragments and amino acids, which are returned to the systemic circulation. Therefore, only minuscule amounts of intact protein are detectable in urine. An additional renal elimination mechanism is peritubular extraction from post glomerular capillaries with subsequent intracellular metabolism, which has, for example, been described for vasopressin and calcitonin.

3.4.4 Hepatic Elimination

Apart from proteolysis and the kidneys, the liver contributes substantially to the metabolism of peptide and protein drugs. Degradation usually starts

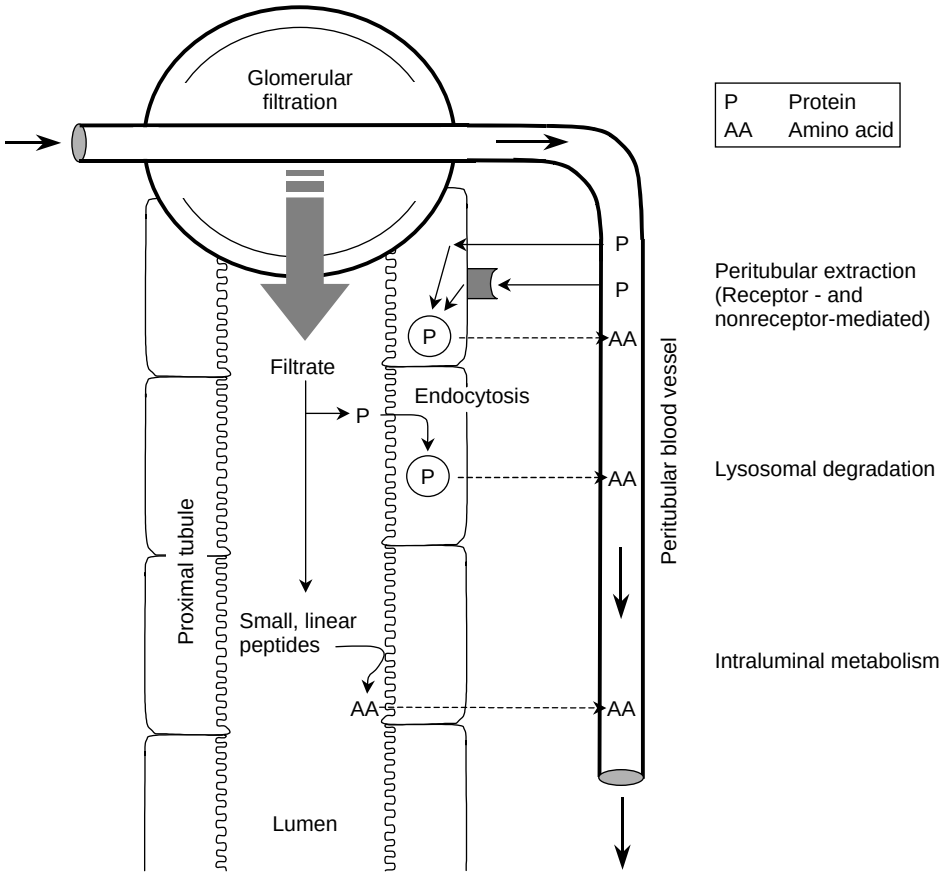

Fig. 2 Renal elimination processes of peptides and proteins: glomerular filtration, intraluminal metabolism, tubular reabsorption with intracellular lysosomal metabolism, and peritubular extraction with intracellular lysosomal metabolism.

with endopeptidases that attack in the middle part of the protein, and the resulting oligopeptides are then further degraded by exopeptidases. The ultimate metabolites of proteins, amino acids, and dipeptides are finally reutilized in the endogenous amino acid pool. The rate of hepatic metabolism is largely dependent on specific amino acid sequences in the protein.

An important first step in the hepatic metabolism of proteins and peptides is the uptake into hepatocytes. Small peptides may cross the hepatocyte membrane *via* passive diffusion if they have sufficient hydrophobicity. The uptake of larger peptides and proteins is facilitated *via* various carrier-mediated, energy-dependent transport processes. Receptor-mediated endocytosis is an additional mechanism for uptake into hepatocytes (see Sect. 3.4.5). In addition, peptides such as metkephamid can be metabolized on the surface of hepatocytes or endothelial cells.

3.4.5 Receptor-mediated Elimination

For conventional small-molecule drugs, receptor binding is usually negligible compared to the total amount of drug in the body, and rarely affects their PK profile. By contrast, a substantial fraction of a peptide or protein dose can be bound to receptors. This binding can lead to elimination through receptor-mediated uptake and subsequent intracellular metabolism. The endocytosis process is not limited to hepatocytes, but can occur in other cells as well, including the therapeutic target cells. Since the number of receptors is limited, their binding and the related drug uptake can usually be saturated within therapeutic concentrations. Thus, receptor–mediated elimination constitutes a major source for nonlinear PK behavior of numerous peptide and protein drugs, that is, a lack of dose proportionality.

M-CSF, for example, undergoes, besides linear renal elimination, a nonlinear elimination pathway that follows Michaelis–Menten kinetics and is linked to a receptor-mediated uptake into macrophages. At low concentrations, M-CSF follows linear PKs, while at high concentrations, nonrenal elimination pathways are saturated, resulting in nonlinear PK behavior (Fig. 3). Other examples for receptor-mediated elimination are insulin, t-PA, epidermal growth factor (EGF), ANP, and interleukin-10.

A mechanism-based, target-mediated drug distribution model was, for example, developed in order to accurately describe the nonlinear PKs of a recombinant human vascular endothelial growth factor ($rhVEGF_{165}$) in patients with coronary artery disease. Nonlinearity was caused by elimination of $rhVEGF_{165}$ by binding

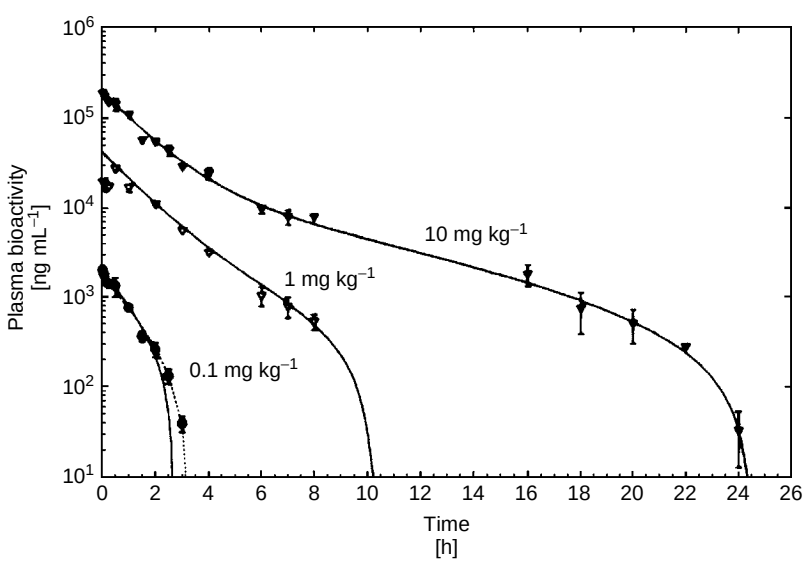

Fig. 3 Nonlinear pharmacokinetics of M-CSF, presented as measured and modeled plasma concentration–time curves (mean ± SE) after intravenous injection of 0.1 mg kg^{-1} ($n = 5$), 1.0 mg kg^{-1} ($n = 3$) and 10 mg kg^{-1} ($n = 8$) in rats (Reprinted with permission from Bauer, R.J., Gibbons, J.A., Bell, D.P., Luo, Z.P., Young, J.D. (1994) Nonlinear pharmacokinetics of recombinant human macrophage colony-stimulating factor (M-CSF) in rats, *J. Pharmacol. Exp. Ther.* **268**, 152–158).

to specific and saturable high-affinity receptors followed by internalization and degradation.

3.5 Species Specificity and Allometry

Peptides and proteins exhibit distinct species specificity with regard to structure and activity. Peptides and proteins with identical physiological function may have different amino acid sequences in different species and may have no activity or be even immunogenic if used in a different species. The extent of glycosylation is another factor of species differences, for example, for interferon-α or erythropoietin, which may alter the drug's clearance. This is of particular importance if the production of human proteins is carried out using bacterial cells.

Extrapolation of animal data to predict PK parameters by allometric scaling is an often-used tool in drug development with multiple approaches available at variable success rates. In the most frequently used approach, PK parameters between different species are related to each other *via* body weight using a power function:

$$P = a \cdot W^b \qquad (2)$$

where P is the PK parameter scaled, W is the body weight in kilograms, a is the allometric coefficient, and b is the allometric exponent. a and b are specific constants for each parameter of a compound. General tendencies for the allometric exponent are 0.75 for rate constants (i.e. clearance, elimination rate constant), 1 for volumes of distribution, and 0.25 for half-lives.

For most traditional, small-molecule drugs, allometric scaling is often imprecise, especially if hepatic metabolism is a major elimination pathway and/or if there are interspecies differences in metabolism. For peptides and proteins, however, allometric scaling has frequently proven to be much more precise and reliable, probably because of the similarity in handling peptides and proteins between different mammalian species. Clearance and volume of distribution of numerous therapeutically used proteins like somatotropin or t-PA follow a well-defined, weight-dependent physiologic relationship between lab animals and humans. This allows relatively precise quantitative predictions for toxicology and dose-ranging studies based on preclinical findings.

Figure 4 shows the allometric plots of PK parameters for the recombinant, soluble, and chimeric form of P-selectin glycoprotein ligand-1 (rPSGL-Ig), an antagonist to P-selectin for the treatment of P-selectin-mediated diseases like thrombosis, reperfusion injury, and deep vein thrombosis. Human rPSGL-Ig PK parameters could accurately be predicted on the basis of data from mouse, rat, monkey, and pig, using allometric power functions.

3.6 Immunogenicity

Because of the antigenic potential of proteins, the formation of antibodies is a frequently observed phenomenon during chronic therapy with protein drugs, especially if human proteins are used in animal studies or if animal-derived proteins are applied in human clinical studies.

Most monoclonal antibodies are murine in nature and their systemic administration can lead to the development of human antimouse immunoglobulin antibody (HAMA) response, which is in most cases directed against the constant

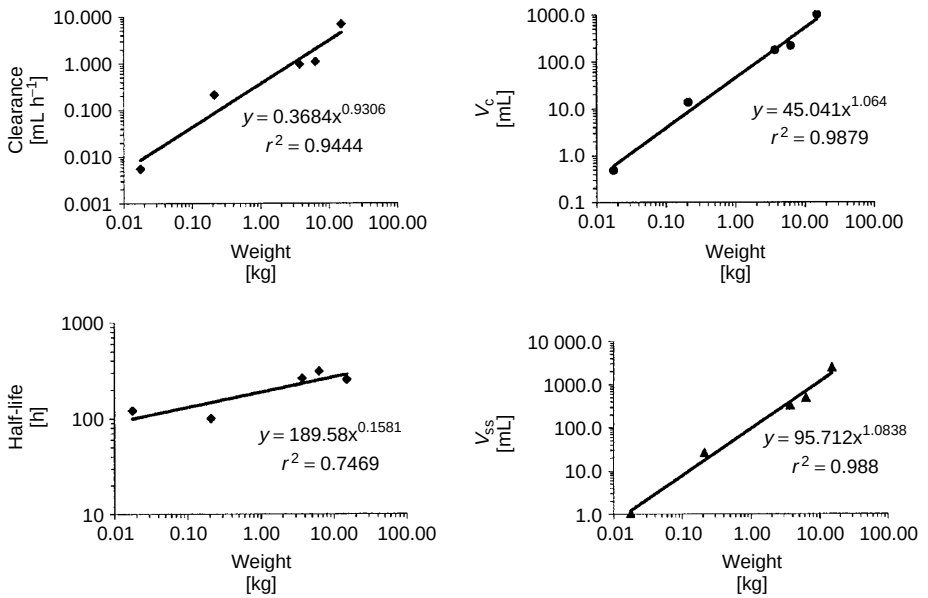

Fig. 4 Allometric plots of the pharmacokinetic parameters clearance, volume of the central compartment (V_c), volume of distribution at steady state (V_{ss}), and elimination half-life of rPSGL-Ig. Each data point within the plot represents an averaged value of the pharmacokinetic parameter with increasing weight from mouse, rat, monkey (3.74 kg), monkey (6.3 kg), and pig, respectively. The solid line is the best fit with a power function to relate pharmacokinetic parameters to body weight (Reprinted with permission from Khor, S.P., McCarthy, K., DuPont, M., Murray, K., Timony, G. (2000) Pharmacokinetics, pharmacodynamics, allometry, and dose selection of rPSGL-Ig for phase I trial, *J. Pharmacol. Exp. Ther.* **293**, 618–624).

regions of the immunoglobulin. Genetically engineered mouse–human chimeric antibodies try to minimize this immunogenicity in man by joining variable domains of the mouse to the constant regions of human immunoglobulins. The anti-EGFR (epidermal growth factor receptor) IgG monoclonal antibody cetuximab is an example of a murine–human chimeric antibody recently approved for various cancer indications.

Extravascular injection is known to stimulate antibody formation more than intravenous application, most likely due to the increased immunogenicity of protein aggregates and precipitates formed at the injection site. The presence of antibodies can obliterate the biological activity of a protein drug. In addition, protein–antibody complexation can also modify the distribution, metabolism, and excretion, that is, the PK profile, of the protein drug. Elimination can be either increased or decreased. Faster elimination of the complex occurs if the reticuloendothelial system is stimulated. Elimination is slowed down if the antibody–drug complex forms a depot for the protein drug. This effect would prolong the drug's therapeutic activity and might be beneficial if the complex formation does not decrease therapeutic activity. Furthermore, antibody binding may also interfere with bioanalytical methods like immunoassays.

4
Pharmacokinetics of Oligonucleotides

Antisense oligonucleotides hold great promise as novel therapeutic agents designed to specifically and selectively inhibit the production of disease-related products, with fomivirsen being the first approved antisense oligonucleotide drug product. So far, a significant body of preclinical and human PK data is available only for phosphothioate oligonucleotides (PONs).

Oral bioavailability is generally very low, ranging from 1 to 3%. Ongoing studies, however, indicate that oral bioavailability can be increased by the appropriate release of drug and permeability-enhancing excipients. PONs have also been administered *via* subcutaneous, intradermal, and pulmonary application routes.

After intravenous administration, PONs generally follow two-compartment characteristics and are rapidly cleared from plasma, predominantly *via* distribution processes with a half-life of 0.5 to 1 h depending on the dose. The ICAM-1 inhibitor alicaforsen, for example, has a distribution half-life of 1.0 to 1.2 h in humans. Plasma PKs are nonlinear, with a more than proportional increase in area-under-the-curve (AUC) with dose. This is most probably due to saturation of tissue uptake. Figure 5, for example, shows the disproportional increase in systemic exposure after escalating doses of the HIV-inhibitor trecovirsen. The plasma PKs of various ONs are generally independent of their sequence and chemistry as plasma clearance is primarily determined by distribution processes.

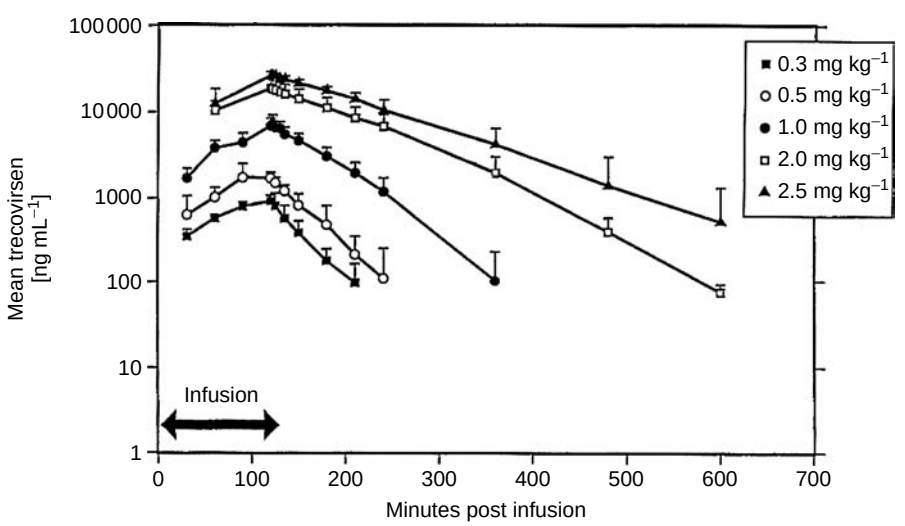

Fig. 5 Plasma concentration-time course of the antisense oligonucleotide trecovirsen in HIV-positive subjects, indicating a disproportional increase in systemic exposure after escalating doses. Trecovirsen was administered by 2-h intravenous administration to groups of six subjects (12 subjects for the 1.0 mg kg^{-1} dose group). (Reprinted with permission from Sereni, D., Tubiana, R., Lascoux, C., Katlama, C., Taulera, O., Bourque, A., Cohen, A., Dvorchik, B., Martin, R.R., Tournerie, C., Gouyette, A., Schechter, P.J. (1999) Pharmacokinetics and tolerability of intravenous trecovirsen (GEM 91), an antisense phosphorothioate oligonucleotide, in HIV-positive subjects, *J. Clin. Pharmacol.* **39**, 47–54).

After intravenous administration, PONs are detected in nearly all tissues and organs except brain and testes, suggesting significant transport barriers in these tissues. The extent of tissue uptake is dependent on the dose amount as well as dose rate. A major accumulation of PON occurs in liver and kidneys, and to a lesser extent in spleen, bone marrow, and lymph nodes, which seems to be independent of PON sequence. Chemical modification of the phosphothioate backbone structure may, however, alter protein binding and organ distribution. The mechanisms for uptake into target cells have not been fully elucidated yet, but these processes are energy-, temperature-, and time-dependent, and probably include pinocytosis and podocytosis.

PONs are cleared from tissues by nuclease-mediated metabolism, with half-lives that vary between 20 and 120 h, depending on the organ or tissue. Successive removal of bases from the 3′-end is the major metabolic pathway in plasma, while both 3′ and 5′ exonuclease excision may occur in tissues. Exonuclease metabolism in plasma and tissues is rapid, with 30 to 40% of PON having at least one nucleotide removed after 5 min in plasma. Endonuclease-mediated degradation of PONs is generally not observed.

PONs are highly bound to plasma proteins that protect them from renal filtration. Plasma protein binding of ISIS2503, for example, ranged from 95 to 97% in rats and monkeys, but was saturable at high concentrations. Urinary excretion is a major route of excretion for PONs, regardless of sequence or chemical structure, with the majority being shorter-length metabolites rather than unchanged parent drug. Urinary excretion is nonlinear, with a greater fraction excreted at higher doses. Potential mechanisms include saturation of plasma protein binding as well as tubular reuptake mechanisms. Only a minor fraction of the dose is excreted into feces, although enterohepatic recirculation has been suggested.

ISIS 104838 is a tumor necrosis factor-α (TNF-α) inhibiting second-generation ON containing five 2′-O-(2-methoxyethyl) modified (2′-MOE) nucleosides at the 3′- and 5′-terminus, respectively. The PK pattern for this second-generation PON was similar to first-generation PONs, except for a less pronounced nonlinearity in systemic exposure and a substantially prolonged terminal half-life of 27 ± 3.8 days, most probably due to the complete blockade of exonuclease digestion by the MOE modification.

5 Pharmacokinetics of DNA

In comparison to peptides, proteins, and oligonucleotides, much less is known about the PKs of recombinant plasmid DNA used like a "drug" in the novel treatment approach of gene therapy.

The *in vivo* disposition of plasmid DNA and its complexes depends largely on its physicochemical characteristics, a strong negative charge, and high molecular weight. After intravenous administration in rats, pDNA is detected in all major organs including lungs, liver, kidney, and spleen. Low-level detection in the brain is most likely an artifact from residual blood, given that pDNA is unlikely to cross the blood-brain barrier.

After intravenous administration in mice, pDNA is rapidly eliminated from the plasma due to extensive uptake into the liver as well as rapid degradation by nucleases, with hepatic uptake clearance approaching liver plasma

flow. pDNA is preferentially taken up by liver nonparenchymal cells, such as Kupffer and endothelial cells *via* receptor-mediated processes.

Analysis of the functional forms of pDNA in rats revealed that supercoiled pDNA rapidly disappears from plasma with a half-life of 0.15 min. Approximately 60% of supercoiled pDNA is degraded to open circular pDNA, which is subsequently almost completely converted to linear pDNA. The conversion of open circular to linear pDNA followed Michaelis–Menten kinetics, while linear pDNA was removed with a half-life of 2.1 min. The slower elimination of open circular and linear pDNA as compared to supercoiled pDNA was suggested to be related to a stronger interaction with plasma macromolecules that might offer some protection from plasma nuclease degradation.

6 Exposure/response Correlations for Biotech Drugs

Since biotech drugs are usually highly potent compounds with steep dose–effect curve, a careful characterization of the dose–concentration–effect relationship should receive particular emphasis during the preclinical and clinical drug development process. Integrated PK/PD modeling approaches have widely been applied for the characterization of biotech drugs. PK/PD modeling not only allows for a continuous description of the time course of effect as a function of the dosing regime and comprehensive summary of available data but also enables testing of competing hypotheses regarding processes altered by the drug, allows to make predictions of drug effects under new conditions, and facilitates the estimation of inaccessible system variables.

The application of PK/PD modeling is beneficial in all phases of preclinical and clinical drug development, with a focus on dosage optimization and identification of covariates that are causal for intra- and interindividual differences in drug response and/or toxicity. It has recently been further endorsed by the publication of the Exposure–Response Guidance document by the US Food and Drug Administration. Mechanism-based PK/PD modeling, which appreciates the physiological events involved in the elaboration of the observed effect, has been promoted as a superior modeling approach when compared to empirical modeling, especially because it not only describes observations but also offers some insight into the underlying biological processes involved and thus provides flexibility in extrapolating the model to other clinical situations. Since the molecular mechanism of action of biotech drugs is generally well understood, it is often straightforward to transform this available knowledge into a mechanism-based PK/PD modeling approach that appropriately characterizes the real physiological process leading to the drug's therapeutic effect.

In the following, the application of the three most common PK/PD modeling classes, direct-link models, indirect-link models, and indirect-response models, will be discussed in more detail. In addition, extensions of these concepts and more complex approaches will be introduced in illustrative examples. However, it should be mentioned that PK/PD models for biotech drugs are not limited only to continuous responses as shown in the following but are also used for binary or graded responses. A logistic PK/PD modeling approach was, for example, used to

link cumulative AUC of the anti-TNF-α protein etanercept with the American College of Rheumatology response criterion of 20% improvement (ARC20) in patients with rheumatoid arthritis.

6.1 Direct Link PK/PD Models

While drug concentrations are usually analytically quantified in plasma, serum, or blood, the magnitude of the observed response is determined by the concentration of the drug at its effect site, the site of action in the target tissue. The relationship between the drug concentration in plasma and at the effect site may either be constant or undergo time-dependent changes. If equilibrium between both concentrations is rapidly achieved or the site of action is within plasma, serum, or blood, there is practically a constant relationship between both concentrations with no temporal delay between plasma and effect site. In this case, measured concentrations can directly serve as input for a PD model. The most frequently used direct-link PD model is the sigmoid E_{max}-Model:

$$E = \frac{E_{max} \cdot C^n}{EC_{50}^n + C^n} \quad (3)$$

with E_{max} as maximum achievable effect, C as drug concentration at the effect site, and EC_{50} the concentration of the drug that produces half of the maximum effect. The Hill coefficient n is a shape factor that allows for an improved fit of the relationship to the observed data. Thus, a direct-link model directly connects measured concentration to the observed effect without any temporal delay.

A. Racine-Poon et al. provided an example for a direct-link model by relating the serum concentration of the antihuman immunoglobulin E (IgE) antibody CGP 51901 for the treatment of seasonal allergic rhinitis to the reduction of free IgE *via* an inhibitory E_{max}-model. Similarly, a direct-link model was applied to assess the effect of recombinant interleukin-10 on the *ex vivo* release of the proinflammatory cytokines TNF-α and interleukin-1β (IL-1β) in LPS-stimulated leukocytes.

6.2 Indirect Link PK/PD Models

The concentration–effect relationship of many biotech drugs, however, cannot be described by direct-link PK/PD models, but is characterized by a temporal dissociation between the time courses of plasma concentration and effect. In this case, concentration maxima would occur before effect maxima, effect intensity would increase despite decreasing plasma concentrations and would persist beyond the time when drug concentrations in plasma are no longer determinable. The relationship between measured concentration and observed effect follows a counterclockwise hysteresis loop. This phenomenon can be caused either by an indirect-response mechanism (see Sect. 6.3) or by a distributional delay between the concentrations in plasma and at the effect site. The latter can conceptually be described by an effect-compartment model, which attaches a hypothetical effect-compartment to a PK compartment model. The effect-compartment does not account for mass balance and only defines the changes in concentration at the effect site *via* the time course of the effect itself.

An effect-compartment approach was, for example, applied to quantify the reduction in mean arterial blood pressure by the antiadrenergic peptoid CHIR 2279. The same concept was used by Pihoker

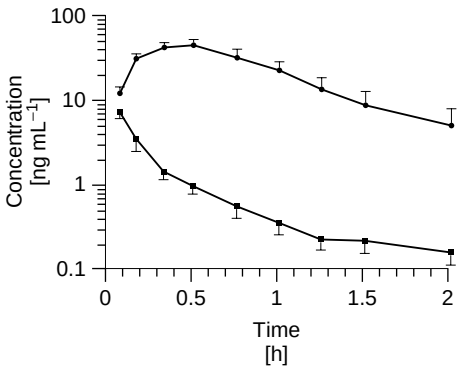

Fig. 6 Serum concentration profiles (mean ± SD) of GHRP-2 (growth hormone–releasing peptid-2; ■) and endogenous somatotropin (●) after intravenous administration of 1μ g kg^{-1} GHRP-2 in prepubertal children ($n = 10$). (Reprinted with permission from Pihoker, C., Kearns, G.L., French, D., Bowers, C.Y. (1998) Pharmacokinetics and pharmacodynamics of growth hormone-releasing peptide-2: a phase I study in children, *J. Clin. Endocrinol. Metab.* **83**, 1168–1172. Copyright 1998, The Endocrine Society).

et al. to characterize the relationship between the serum concentration of the somatotropin-releasing peptide GHRP-2 and somatotropin (Fig. 6).

6.3 Indirect-response PK/PD Models

The effect of most biotech drugs, however, is not mediated *via* a direct interaction between drug concentration and response systems, but frequently involves several transduction processes that include, at their rate-limiting step, the stimulation or inhibition of a physiologic process, for example, the synthesis or degradation of a molecular response mediator like a hormone or cytokine. In these cases, mechanism-based indirect-response models should be used that appreciate the underlying physiological process involved in mediating the observed effect. Indirect-response models generally describe the effect on a representative response parameter *via* the dynamic equilibrium between increase or synthesis and decrease or degradation of the response, with the former being a zero-order and the latter a first-order process. Each of these processes can be stimulated or inhibited in four derived basic model variants.

Two variants of the indirect-response models, stimulation of synthesis and of degradation processes were, for example, used in a recent study for modeling the effect of recombinant erythropoietin on the two response parameters, free ferritin (*Fr*) concentration and soluble transferrin receptor (*Tfr*) concentration. While erythropoietin reduces *Fr*, it increases *Tfr* (Fig. 7). The temporal change in both response variables can be described by the following equations:

$$\frac{\mathrm{d}Fr}{\mathrm{d}t} = k_{\mathrm{in},F} - k_{\mathrm{out},F} \cdot \left(1 + \frac{E_{\max} \cdot C_m^n}{EC_{50}^n + C_m^n}\right) \cdot Fr$$

$$\frac{\mathrm{d}Tfr}{\mathrm{d}t} = k_{\mathrm{in},T} \cdot \left(1 + \frac{E_{\max} \cdot C_m^n}{EC_{50}^n + C_m^n}\right) - k_{\mathrm{out},T} \cdot Tfr \qquad (4)$$

with k_{in} as endogenous formation rate of *Fr* and *Tfr*, and k_{out} as first-order degradation rate constant, respectively. C_m is the erythropoietin concentration that was additionally modulated *via* a transduction process with 50-h delay.

Similarly, a modified indirect-response model was used to relate the concentration of the humanized antifactor IX antibody

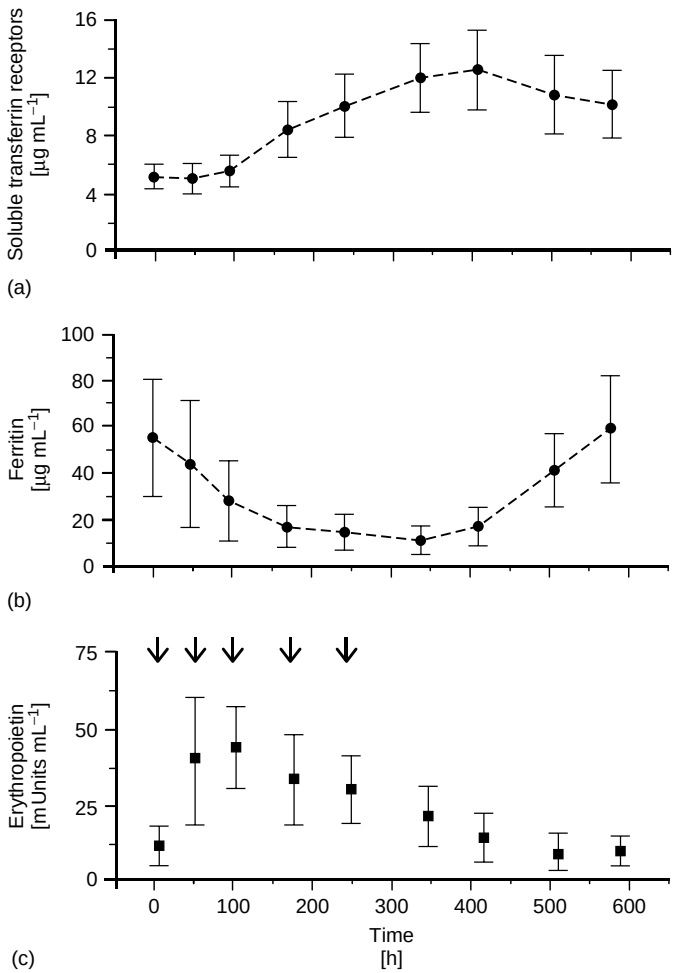

Fig. 7 Serum concentration–time course (mean ± SD) of erythropoietin (c) and its effect on the concentrations of free ferritin (b) and soluble transferrin receptor (a) after repeated subcutaneous administration of 200 U kg^{-1} recombinant erythropoietin in athletes ($n = 18$) (Reprinted with permission from Bressolle, F., Audran, M., Gareau, R., Pham, T.N., Gomeni, R. (1997) Comparison of a direct and indirect population pharmacodynamic model: application to recombinant human erythropoietin in athletes, *J. Pharmacokinet. Biopharm.* **25**, 263–275).

SB249417 to factor IX activity in Cynomolgus monkeys as well as humans. The drug effect in this model was introduced by interrupting the natural degradation of Factor IX by sequestration of factor IX by the antibody.

Indirect-response models were also used for the effect of somatotropin on endogenous IGF-1 concentration, as well as the immune suppressive activity of the monoclonal antibody mAb 5c8.

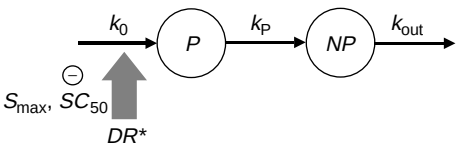

Fig. 8 Schematic representation of an indirect-response model with precursor pool used to describe the effect of interferon-β 1a (represented by its internalized drug-receptor complex DR^*) on endogenous neopterin concentrations (NP) *via* stimulation of the synthesis of its precursor neopterin triphosphate (P). See text for discussion of the details (Modified from Mager, D.E., Neuteboom, B., Efthymiopoulos, C., Munafo, A., Jusko, W.J. (2003a) Receptor-mediated pharmacokinetics and pharmacodynamics of interferon-β 1a following intravenous and subcutaneous dosing in monkeys, *J. Pharmacol. Exp. Ther.* **26**, 26.

Although physiologically related mechanism-based modeling should be preferred, an indirect-response–based temporal dissociation between time course of concentration and effect can also be modeled with the effect-compartment approach. The effect of the growth hormone–releasing peptid (GHRP) ipamorelin on somatotropin, for example, was described by a physiologically based indirect-response model, while the already mentioned GHRP-2 effect on somatotropin was characterized with an effect-compartment approach. Similarly, effect-compartment as well as indirect-response models were applied for characterizing the effect of insulin on blood glucose levels. A recent comparative study, however, suggests that a mechanism-based indirect-response model is a more appropriate approach for modeling the PK/PD of insulin.

6.4 Precursor Pool PK/PD Models

An extension of indirect-response models are precursor pool-dependent indirect-response models that include the liberation of an endogenous compound from a storage pool. These models possess the unique ability to characterize both tolerance and rebound phenomena. Such a model was, for example, used to describe the effect of interferon-β1a on neopterin, an endogenous marker for cell-mediated immunity in humans and monkeys (Fig. 8). The primary elimination mechanism of interferon-β1a was modeled as receptor-mediated endocytosis, and the PD model was driven by the amount of internalized drug-receptor complex DR^*:

$$\frac{dP}{dt} = k_0 \cdot \left(1 + \frac{S_{max} \cdot DR^*}{SC_{50} + DR^*}\right) - k_p \cdot P \tag{5}$$

where P is the concentration of neopterin precursor (neopterin triphosphate), k_0 is the zero-order production rate of precursor P, and k_p is the first-order rate constant characterizing the conversion of precursor P to neopterin. The amount of internalized DR^* stimulates precursor production *via* a stimulation function governed by the maximum effect parameter S_{max} and a sensitivity parameter SC_{50}, the concentration that results in half of S_{max}. The rate of change in neopterin concentration NP is then defined by the following expression:

$$\frac{dNP}{dt} = k_p \cdot P - k_{out} \cdot NP \tag{6}$$

with k_{out} as the first-order elimination rate constant of NP in the body.

6.5 Complex PK/PD Models

Since the effect of most biotech drugs is mediated *via* complex regulatory physiologic processes including feedback mechanisms and/or tolerance phenomena, some

PK/PD models that have been described for biotech drugs are much more sophisticated than the four classes of models previously discussed. One example of such a complex modeling approach has been developed for the therapeutic effects of the LH-RH antagonist cetrorelix.

Cetrorelix is used for the prevention of premature ovulation in women undergoing controlled ovarian stimulation in *in vitro* fertilization protocols. LH-RH antagonists suppress the LH levels and delay the occurrence of the preovulatory LH surge, and this delay is thought to be responsible for postponing ovulation. The suppression of LH was modeled in the PK/PD approach with an indirect-response model approach directly linked to cetrorelix plasma concentrations (Fig. 9). The shift in LH surge was linked to cetrorelix concentration, with a simple E_{max}-function *via* a hypothetical effect-compartment to account for a delay in response *via* complex signal transduction steps of unknown mechanism of action. The combined effect of LH suppression and delaying the LH surge was described by the following relationship:

$$\frac{dLH}{dt} = k_0 \cdot \left(1 - \frac{C}{IC_{50} + C}\right) \cdot \left(1 + \frac{SA}{\left(\frac{t - \left(T_0 + \frac{E_{max} \cdot C_e}{EC_{50} + C_e}\right)}{SW}\right)^N + 1}\right) - k_e \cdot LH \tag{7}$$

where LH is the LH concentration, k_0 and k_e are the zero-order production rate and first-order elimination rate constants for LH at baseline, C and C_e are the cetrorelix concentrations in plasma and a hypothetical effect-compartment respectively, SA is the LH surge amplitude, t is time, T_0 is the time at which the peak occurs under baseline conditions, SW is the width of the peak in time units, IC_{50} is the cetrorelix

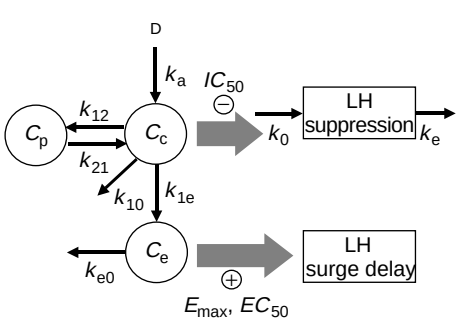

Fig. 9 PK/PD model for the LH suppression and the delay in LH surge following administration of a cetrorelix dose (D). C_C, C_P, and C_e: Concentration of drug in central, peripheral, and effect compartments, respectively; k_a, k_{12}, k_{21}, k_{10}, k_{1e}, k_{e0}: pharmacokinetic first-order rate constants. See text for discussion of the details (Modified from Nagaraja, N.V., Pechstein, B., Erb, K., Klipping, C., Hermann, R., Locher, M., Derendorf, H. (2003) Pharmacokinetic/pharmacodynamic modeling of luteinizing hormone (LH) suppression and LH surge delay by cetrorelix after single and multiple doses in healthy premenopausal women, *J. Clin. Pharmacol.* **43**, 243–251).

concentration that suppresses LH levels by 50%, E_{max} is the maximum delay in LH surge, and EC_{50} is the cetrorelix concentration that produces half of E_{max}. N describes the slope of the surge peak and is an even number. Baseline data analysis indicated that N and SW were best fixed at values of 4 and 24 h, respectively.

Figure 10 shows the application of this PK/PD model to characterize the LH

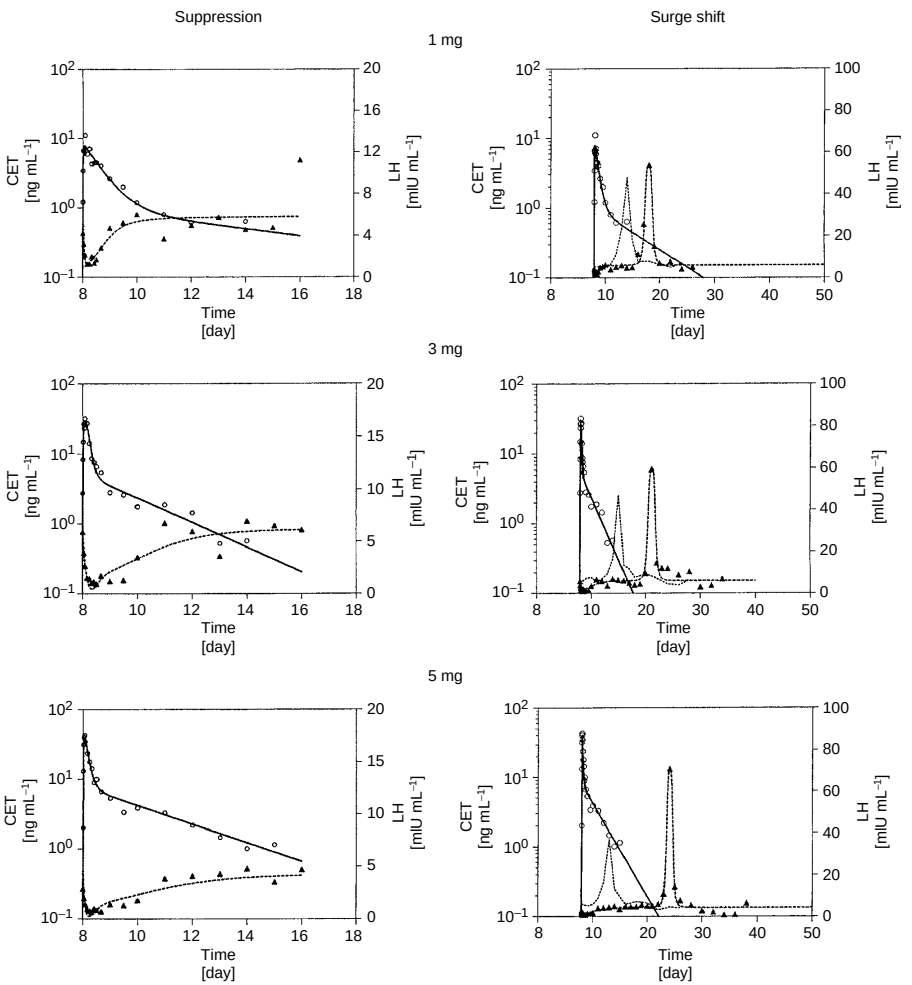

Fig. 10 PK and PD relationship between cetrorelix (O) and LH concentrations (▲) after single doses of 1, 3, and 5 mg cetrorelix in representative subjects. Left panel: LH suppression. Right panel: LH suppression and LH surge profiles. The thick solid line represents the model-fitted cetrorelix concentration; the broken line the model-fitted LH concentration; and the thin dotted line the pretreatment LH profile (not fitted) (Reprinted with permission from Nagaraja, N.V., Pechstein, B., Erb, K., Klipping, C., Hermann, R., Niebch, G., Derendorf, H. (2000) Pharmacokinetic and pharmacodynamic modeling of cetrorelix, an LH-RH antagonist, after subcutaneous administration in healthy premenopausal women, *Clin. Pharmacol. Ther.* **68**, 617–625).

suppression and LH surge delay after subcutaneous administration of cetrorelix to groups of 12 women at different dose levels. The analysis revealed a marked dose–response relationship for the LH surge, and thus predictability of drug response to cetrorelix.

An even more complex, mechanism-based modeling approach including tolerance phenomena was used for the effect of antide, an LH-RH antagonist, on the endogenous regulatory mechanisms and plasma concentrations of LH and testosterone.

7 Summary

In general, biotech drugs conform to the same PK and PD principles as traditional, small-molecule drugs. On the basis of their similarity to endogenous compounds or nutrients, however, numerous caveats and pitfalls related to bioanalytics and PKs have to be considered and addressed during the development process, and this may require additional resources. Furthermore, PDs is frequently complicated owing to close interaction with endogenous substances and specific feedback mechanisms.

Biotech drugs, including peptides, proteins and antibodies, oligonucleotides, and DNA, are projected to cover a substantial market share in the health care systems of the future. It will be crucial for their widespread application in pharmacotherapy, however, that their respective drug development programs are successfully completed in a rapid, cost-efficient, and goal-oriented manner. A more widespread application of PK and PD concepts including exposure–response correlations has repeatedly been promoted by industry, academia, and regulatory authorities for all preclinical and clinical phases of drug development and is believed to result in a scientifically driven, evidence-based, more focused and accelerated drug product development process. Thus, PK/PD concepts are likely to continue and expand their role as a decisive factor in the successful development of biotechnologically derived drug products in the future.

See also Antitumor Steroids; Cancer Chemotherapy, Theoretical Foundations of; HIV Therapeutics, Biochemistry of; Medicinal Chemistry; Oncology, Molecular; Pharmacogenomics and Drug Design.

Bibliography

Books and Reviews

Braeckman, R. (1997) Pharmacokinetics and pharmacodynamics of peptide and protein drugs, in: Crommelin, D., Sindelar, R. (Eds.) *Pharmaceutical Biotechnology*, Harwood Academic Publishers, Amsterdam, pp. 101–122.

Derendorf, H., Meibohm, B. (1999) Modeling of pharmacokinetic/pharmacodynamic (PK/PD) relationships: concepts and perspectives, *Pharm. Res.* **16**, 176–185.

Holford, N.H., Sheiner, L.B. (1982) Kinetics of pharmacologic response, *Pharmacol. Ther.* **16**, 143–166.

Kayser, O., Müller, R.H. (Eds.) (2004) *Pharmaceutical Biotechnology*, Wiley-VCH, Weinheim.

Levin, A.A. (1999) A review of the issues in the pharmacokinetics and toxicology of phosphorothioate antisense oligonucleotides, *Biochim. Biophys. Acta* **1489**, 69–84.

Meibohm, B., Derendorf, H. (1997) Basic concepts of pharmacokinetic/pharmacodynamic (PK/PD) modeling, *Int. J. Clin. Pharmacol. Ther.* **35**, 401–413.

Meibohm, B., Derendorf, H. (2002) Pharmacokinetic/pharmacodynamic studies in drug product development, *J. Pharm. Sci.* **91**, 18–31.

Nagle, T., Berg, C., Nassr, R., Pang, K. (2003) The further evolution of biotech, *Nat. Rev. Drug. Discov.* **2**, 75–79.

Rowland, M., Tozer, T. (1993) *Clinical Pharmacokinetics: Concepts and Applications*,, Williams & Wilkins, Media, PA.

Tang, L., Persky, A.M., Hochhaus, G., Meibohm, B. (2004) Pharmacokinetic aspects of biotechnology products, *J. Pharm. Sci.* **93**, 2184–2204.

Primary Literature

Alton, K.B., Kosoglou, T., Baker, S., Affrime, M.B., Cayen, M.N., Patrick, J.E. (1998) Disposition of 14C-eptifibatide after intravenous administration to healthy men, *Clin. Ther.* **20**, 307–323.

Andersen, S., Lambrecht, L., Swan, S., Cutler, D., Radwanski, E., Affrime, M., Garaud, J. (1999) Disposition of recombinant human interleukin-10 in subjects with various degrees of renal function, *J. Clin. Pharmacol.* **39**, 1015–1020.

Bartocci, A., Mastrogiannis, D.S., Migliorati, G., Stockert, R.J., Wolkoff, A.W., Stanley, E.R. (1987) Macrophages specifically regulate the concentration of their own growth factor in the circulation, *Proc. Natl. Acad. Sci. U. S. A.* **84**, 6179–6183.

Bauer, R.J., Gibbons, J.A., Bell, D.P., Luo, Z.P., Young, J.D. (1994) Nonlinear pharmacokinetics of recombinant human macrophage colony-stimulating factor (M-CSF) in rats, *J. Pharmacol. Exp. Ther.* **268**, 152–158.

Benincosa, L.J., Chow, F.S., Tobia, L.P., Kwok, D.C., Davis, C.B., Jusko, W.J. (2000) Pharmacokinetics and pharmacodynamics of a humanized monoclonal antibody to factor IX in cynomolgus monkeys, *J. Pharmacol. Exp. Ther.* **292**, 810–816.

Berger, H., Heinrich, N., Schafer, H., Baeger, I., Mehlis, B. (1988) Gonadotropin-releasing hormone (GnRH) pharmacokinetics: peptide hormone pharmacokinetics needs clarification, *Life Sci.* **42**, 985–991.

Boxenbaum, H. (1982) Interspecies scaling, allometry, physiological time, and the ground plan of pharmacokinetics, *J. Pharmacokinet. Biopharm.* **10**, 201–227.

Bressolle, F., Audran, M., Gareau, R., Pham, T.N., Gomeni, R. (1997) Comparison of a direct and indirect population pharmacodynamic model: application to recombinant human erythropoietin in athletes, *J. Pharmacokinet. Biopharm.* **25**, 263–275.

CDER/FDA (2001) Bioanalytical Method Validation: Guidance for Industry, US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Rockville.

CDER/FDA (2003) Exposure Response Relationships: Guidance for Industry, US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Rockville.

Chanson, P., Timsit, J., Harris, A.G. (1993) Clinical pharmacokinetics of octreotide. Therapeutic applications in patients with pituitary tumours, *Clin. Pharmacokinet.* **25**, 375–391.

Chow, F.S., Benincosa, L.J., Sheth, S.B., Wilson, D., Davis, C.B., Minthorn, E.A., Jusko, W.J. (2002) Pharmacokinetic and pharmacodynamic modeling of humanized anti-factor IX antibody (SB 249417) in humans, *Clin. Pharmacol. Ther.* **71**, 235–245.

Clemmons, D.R., Dehoff, M.L., Busby, W.H., Bayne, M.L., Cascieri, M.A. (1992) Competition for binding to insulin-like growth factor (IGF) binding protein-2, 3, 4, and 5 by the IGFs and IGF analogs, *Endocrinology* **131**, 890–895.

Crooke, S.T., Graham, M.J., Zuckerman, J.E., Brooks, D., Conklin, B.S., Cummins, L.L., Greig, M.J., Guinosso, C.J., Kornbrust, D., Manoharan, M., Sasmor, H.M., Schleich, T., Tivel, K.L., Griffey, R.H. (1996) Pharmacokinetic properties of several novel oligonucleotide analogs in mice, *J. Pharmacol. Exp. Ther.* **277**, 923–937.

Dayneka, N.L., Garg, V., Jusko, W.J. (1993) Comparison of four basic models of indirect pharmacodynamic responses, *J. Pharmacokinet. Biopharm.* **21**, 457–478.

Dedrick, R.L. (1973) Animal scale-up, *J. Pharmacokinet. Biopharm.* **1**, 435–461.

Dvorchik, B.H. (2000) The disposition (ADME) of antisense oligonucleotides, *Curr. Opin. Mol. Ther.* **2**, 253–257.

Eppler, S.M., Combs, D.L., Henry, T.D., Lopez, J.J., Ellis, S.G., Yi, J.H., Annex, B.H., McCluskey, E.R., Zioncheck, T.F. (2002) A target-mediated model to describe the pharmacokinetics and hemodynamic effects of recombinant human vascular endothelial growth factor in humans, *Clin. Pharmacol. Ther.* **72**, 20–32.

Fattinger, K.E., Verotta, D., Porchet, H.C., Munafo, A., Le Cotonnec, J.Y., Sheiner, L.B. (1996) Modeling a bivariate control system: LH and testosterone response to the GnRH antagonist antide, *Am. J. Physiol.* **271**, E775–E787.

Gibbons, J.A., Hancock, A.A., Vitt, C.R., Knepper, S., Buckner, S.A., Brune, M.E., Milicic, I., Kerwin, J.F., Jr., Richter, L.S., Taylor, E.W., Spear, K.L., Zuckermann, R.N., Spellmeyer, D.C., Braeckman, R.A., Moos, W.H. (1996) Pharmacologic characterization of CHIR 2279, an N-substituted glycine peptoid with high-affinity binding for alpha 1-adrenoceptors, *J. Pharmacol. Exp. Ther.* **277**, 885–899.

Gobburu, J.V., Agerso, H., Jusko, W.J., Ynddal, L. (1999) Pharmacokinetic-pharmacodynamic modeling of ipamorelin, a growth hormone releasing peptide, in human volunteers, *Pharm. Res.* **16**, 1412–1416.

Gobburu, J.V., Tenhoor, C., Rogge, M.C., Frazier, D.E., Jr., Thomas, D., Benjamin, C., Hess, D.M., Jusko, W.J. (1998) Pharmacokinetics/dynamics of 5c8, a monoclonal antibody to CD154 (CD40 ligand) suppression of an immune response in monkeys, *J. Pharmacol. Exp. Ther.* **286**, 925–930.

Herbst, R.S., Langer, C.J. (2002) Epidermal growth factor receptors as a target for cancer treatment: the emerging role of IMC-C225 in the treatment of lung and head and neck cancers, *Semin. Oncol.* **29**, 27–36.

Houk, B.E., Martin, R., Hochhaus, G., Hughes, J.A. (2001) Pharmacokinetics of plasmid DNA in the rat, *Pharm. Res.* **18**, 67–74.

International Conference on Harmonization (1994) ICH E4-Dose-response information to support drug registration, European Agency for the Evaluation of Medicinal Products, London.

Kageyama, S., Yamamoto, H., Nakazawa, H., Matsushita, J., Kouyama, T., Gonsho, A., Ikeda, Y., Yoshimoto, R. (2002) Pharmacokinetics and pharmacodynamics of AJW200, a humanized monoclonal antibody to von Willebrand factor, in monkeys, *Arterioscler. Thromb. Vasc. Biol.* **22**, 187–192.

Khor, S.P., McCarthy, K., DuPont, M., Murray, K., Timony, G. (2000) Pharmacokinetics, pharmacodynamics, allometry, and dose selection of rPSGL-Ig for phase I trial, *J. Pharmacol. Exp. Ther.* **293**, 618–624.

Kuwabara, T., Uchimura, T., Kobayashi, H., Kobayashi, S., Sugiyama, Y. (1995) Receptor-mediated clearance of G-CSF derivative nartograstim in bone marrow of rats, *Am. J. Physiol.* **269**, E1–E9.

LaBella, F.S., Geiger, J.D., Glavin, G.B. (1985) Administered peptides inhibit the degradation of endogenous peptides. The dilemma of distinguishing direct from indirect effects, *Peptides* **6**, 645–660.

Lan, L.B., Dalton, J.T., Schuetz, E.G. (2000) Mdr1 limits CYP3A metabolism in vivo, *Mol. Pharmacol.* **58**, 863–869.

Lee, H., Kimko, H.C., Rogge, M., Wang, D., Nestorov, I., Peck, C.C. (2003) Population pharmacokinetic and pharmacodynamic modeling of etanercept using logistic regression analysis, *Clin. Pharmacol. Ther.* **73**, 348–365.

Lin, S., Chien, Y.W. (2002) Pharmacokinetic-pharmacodynamic modelling of insulin: comparison of indirect pharmacodynamic response with effect-compartment link models, *J. Pharm. Pharmacol.* **54**, 791–800.

Mager, D.E., Jusko, W.J. (2002) Receptor-mediated pharmacokinetic/pharmacodynamic model of interferon-beta 1a in humans, *Pharm. Res.* **19**, 1537–1543.

Mager, D.E., Wyska, E., Jusko, W.J. (2003b) Diversity of mechanism-based pharmacodynamic models, *Drug Metab. Dispos.* **31**, 510–518.

Mager, D.E., Neuteboom, B., Efthymiopoulos, C., Munafo, A., Jusko, W.J. (2003a) Receptor-mediated pharmacokinetics and pharmacodynamics of interferon-β1a following intravenous and subcutaneous dosing in monkeys, *J. Pharmacol. Exp. Ther.* **26**, 26.

Mahmood, I. (2002) Interspecies scaling: predicting oral clearance in humans, *Am. J. Ther.* **9**, 35–42.

Mahmood, I., Balian, J.D. (1999) The pharmacokinetic principles behind scaling from preclinical results to phase I protocols, *Clin. Pharmacokinet.* **36**, 1–11.

Meissner, H.C., Groothuis, J.R., Rodriguez, W.J., Welliver, R.C., Hogg, G., Gray, P.H., Loh, R., Simoes, E.A., Sly, P., Miller, A.K., Nichols, A.I., Jorkasky, D.K., Everitt, D.E., Thompson, K.A. (1999) Safety and pharmacokinetics of an intramuscular monoclonal

antibody (SB 209763) against respiratory syncytial virus (RSV) in infants and young children at risk for severe RSV disease, *Antimicrob. Agents Chemother.* **43**, 1183–1188.

Miller, K.J., Bowsher, R.R., Celniker, A., Gibbons, J., Gupta, S., Lee, J.W., Swanson, S.J., Smith, W.C., Weiner, R.S. (2001) Workshop on bioanalytical methods validation for macromolecules: summary report, *Pharm. Res.* **18**, 1373–1383.

Mordenti, J., Chen, S.A., Moore, J.A., Ferraiolo, B.L., Green, J.D. (1991) Interspecies scaling of clearance and volume of distribution data for five therapeutic proteins, *Pharm. Res.* **8**, 1351–1359.

Mould, D.R., Davis, C.B., Minthorn, E.A., Kwok, D.C., Elliott, M.J., Luggen, M.E., Totoritis, M.C. (1999) A population pharmacokinetic-pharmacodynamic analysis of single doses of clenoliximab in patients with rheumatoid arthritis, *Clin. Pharmacol. Ther.* **66**, 246–257.

Müller, S., Hochhaus, G. (1995) Metabolism of dynorphin A 1-13 in human blood and plasma, *Pharm. Res.* **12**, 1165–1170.

Müller, S., Hutson, A., Arya, V., Hochhaus, G. (1999) Assessment of complex peptide degradation pathways via structured multicompartmental modeling approaches: the metabolism of dynorphin A1-13 and related fragments in human plasma, *J. Pharm. Sci.* **88**, 938–944.

Murakami, T., Misaki, M., Masuda, S., Higashi, Y., Fuwa, T., Yata, N. (1994) Dose-dependent plasma clearance of human epidermal growth factor in rats, *J. Pharm. Sci.* **83**, 1400–1403.

Nagaraja, N.V., Pechstein, B., Erb, K., Klipping, C., Hermann, R., Locher, M., Derendorf, H. (2003) Pharmacokinetic/pharmacodynamic modeling of luteinizing hormone (LH) suppression and LH surge delay by cetrorelix after single and multiple doses in healthy premenopausal women, *J. Clin. Pharmacol.* **43**, 243–251.

Nagaraja, N.V., Pechstein, B., Erb, K., Klipping, C., Hermann, R., Niebch, G., Derendorf, H. (2000) Pharmacokinetic and pharmacodynamic modeling of cetrorelix, an LH-RH antagonist, after subcutaneous administration in healthy premenopausal women, *Clin. Pharmacol. Ther.* **68**, 617–625.

Nakao, K., Sugawara, A., Morii, N., Sakamoto, M., Yamada, T., Itoh, H., Shiono, S., Saito, Y., Nishimura, K., Ban, T. (1986) The pharmacokinetics of alpha-human atrial natriuretic polypeptide in healthy subjects, *Eur. J. Clin. Pharmacol.* **31**, 101–103.

Nishikawa, M., Takemura, S., Takakura, Y., Hashida, M. (1998) Targeted delivery of plasmid DNA to hepatocytes in vivo: optimization of the pharmacokinetics of plasmid DNA/galactosylated poly(L-lysine) complexes by controlling their physicochemical properties, *J. Pharmacol. Exp. Ther.* **287**, 408–415.

Pechstein, B., Nagaraja, N.V., Hermann, R., Romeis, P., Locher, M., Derendorf, H. (2000) Pharmacokinetic-pharmacodynamic modeling of testosterone and luteinizing hormone suppression by cetrorelix in healthy volunteers, *J. Clin. Pharmacol.* **40**, 266–274.

Pihoker, C., Kearns, G.L., French, D., Bowers, C.Y. (1998) Pharmacokinetics and pharmacodynamics of growth hormone-releasing peptide-2: a phase I study in children, *J. Clin. Endocrinol. Metab.* **83**, 1168–1172.

Racine-Poon, A., Botta, L., Chang, T.W., Davis, F.M., Gygax, D., Liou, R.S., Rohane, P., Staehelin, T., van Steijn, A.M., Frank, W. (1997) Efficacy, pharmacodynamics, and pharmacokinetics of CGP 51901, an anti-immunoglobulin E chimeric monoclonal antibody, in patients with seasonal allergic rhinitis, *Clin. Pharmacol. Ther.* **62**, 675–690.

Radwanski, E., Chakraborty, A., Van Wart, S., Huhn, R.D., Cutler, D.L., Affrime, M.B., Jusko, W.J. (1998) Pharmacokinetics and leukocyte responses of recombinant human interleukin-10, *Pharm. Res.* **15**, 1895–1901.

Reilly, R.M., Sandhu, J., Alvarez-Diez, T.M., Gallinger, S., Kirsh, J., Stern, H. (1995) Problems of delivery of monoclonal antibodies. Pharmaceutical and pharmacokinetic solutions, *Clin. Pharmacokinet.* **28**, 126–142.

Sereni, D., Tubiana, R., Lascoux, C., Katlama, C., Taulera, O., Bourque, A., Cohen, A., Dvorchik, B., Martin, R.R., Tournerie, C., Gouyette, A., Schechter, P.J. (1999) Pharmacokinetics and tolerability of intravenous trecovirsen (GEM 91), an antisense phosphorothioate oligonucleotide, in HIV-positive subjects, *J. Clin. Pharmacol.* **39**, 47–54.

Sewell, K.L., Geary, R.S., Baker, B.F., Glover, J.M., Mant, T.G., Yu, R.Z., Tami, J.A., Dorr, F.A. (2002) Phase I trial of ISIS 104838, a 2′-methoxyethyl modified antisense oligonucleotide targeting tumor necrosis factor-alpha, *J. Pharmacol. Exp. Ther.* **303**, 1334–1343.

Shah, V., Midha, K., Dighe, S., McGilveray, I., Skelly, J., Yacobi, A., Layloff, T., Viswanathan, C., Cook, C., McDowall, R., Pittman, K., Spector, S. (1992) Analytical method validation: bioavailability, bioequivalence, and pharmacokinetic studies, *J. Pharm. Sci.* **81**, 309–312.

Sharma, A., Jusko, W. (1998) Characteristics of indirect pharmacodynamic models and applications to clinical drug responses, *Br. J. Clin. Pharmacol.* **45**, 229–239.

Sharma, A., Ebling, W.F., Jusko, W.J. (1998) Precursor-dependent indirect pharmacodynamic response model for tolerance and rebound phenomena, *J. Pharm. Sci.* **87**, 1577–1584.

Sheiner, L.B., Stanski, D.R., Vozeh, S., Miller, R.D., Ham, J. (1979) Simultaneous modeling of pharmacokinetics and pharmacodynamics: application to d-tubocurarine, *Clin. Pharmacol. Ther.* **25**, 358–371.

Sun, Y.N., Jusko, W.J. (1999) Role of baseline parameters in determining indirect pharmacodynamic responses, *J. Pharm. Sci.* **88**, 987–990.

Sun, Y.N., Lee, H.J., Almon, R.R., Jusko, W.J. (1999) A pharmacokinetic/pharmacodynamic model for recombinant human growth hormone effects on induction of insulin-like growth factor I in monkeys, *J. Pharmacol. Exp. Ther.* **289**, 1523–1532.

Takakura, Y., Nishikawa, M., Yamashita, F., Hashida, M. (2001) Development of gene drug delivery systems based on pharmacokinetic studies, *Eur. J. Pharm. Sci.* **13**, 71–76.

Taki, Y., Sakane, T., Nadai, T., Sezaki, H., Amidon, G.L., Langguth, P., Yamashita, S. (1998) First-pass metabolism of peptide drugs in rat perfused liver, *J. Pharm. Pharmacol.* **50**, 1013–1018.

Tanswell, P., Modi, N., Combs, D., Danays, T. (2002) Pharmacokinetics and pharmacodynamics of tenecteplase in fibrinolytic therapy of acute myocardial infarction, *Clin. Pharmacokinet.* **41**, 1229–1245.

Wacher, V.J., Silverman, J.A., Zhang, Y., Benet, L.Z. (1998) Role of P-glycoprotein and cytochrome P450 3A in limiting oral absorption of peptides and peptidomimetics, *J. Pharm. Sci.* **87**, 1322–1330.

Yacyshyn, B.R., Barish, C., Goff, J., Dalke, D., Gaspari, M., Yu, R., Tami, J., Dorr, F.A., Sewell, K.L. (2002) Dose ranging pharmacokinetic trial of high-dose alicaforsen (intercellular adhesion molecule-1 antisense oligodeoxynucleotide) (ISIS 2302) in active Crohn's disease, *Aliment Pharmacol. Ther.* **16**, 1761–1770.

Yacyshyn, B.R., Chey, W.Y., Goff, J., Salzberg, B., Baerg, R., Buchman, A.L., Tami, J., Yu, R., Gibiansky, E., Shanahan, W.R. (2002) Double blind, placebo controlled trial of the remission inducing and steroid sparing properties of an ICAM-1 antisense oligodeoxynucleotide, alicaforsen (ISIS 2302), in active steroid dependent Crohn's disease, *Gut* **51**, 30–36.

Yu, R.Z., Geary, R.S., Leeds, J.M., Watanabe, T., Moore, M., Fitchett, J., Matson, J., Burckin, T., Templin, M.V., Levin, A.A. (2001) Comparison of pharmacokinetics and tissue disposition of an antisense phosphorothioate oligonucleotide targeting human Ha-ras mRNA in mouse and monkey, *J. Pharm. Sci.* **90**, 182–193.

Pharmacology and Emerging Clinical Application of RNA

Sailen Barik and Vira Bitko
University of South Alabama, Mobile, AL, USA

1 Clinically Relevant Features of RNA and its Metabolism 243
1.1 RNA in Central Dogma 243
1.2 Transcription and Replication of RNA Genomes 245
1.3 Noncanonical Base Pairs in RNA 246
1.4 RNA Splicing 246
1.4.1 Alternate RNA Splicing: Regulation by Antisense RNA 246
1.4.2 Trans-cleaving Ribozymes: Destruction of Undesired RNA 248
1.4.3 Trans-splicing Ribozymes: Repair of Defective RNA 248
1.5 RNA Interference: Sirna and miRNA 251
1.6 RNA as a Protein Antagonist: SELEX, Aptamers, and Spiegelmers 252
1.7 Polyamide (''Peptide'') Nucleic Acid (PNA) 253
1.8 Summary: Pharmaceutical Advantages of RNA 254

2 Emerging Clinical Applications of RNA 255
2.1 Ribozymes Against Infectious Agents 255
2.2 Ribozymes Against Cellular Disorders 257
2.3 RNAi Against Viruses 258
2.4 RNAi as an Antiparasitic Approach 259
2.5 RNAi Against Genetic Disorders and Cancer 259
2.6 Antisense RNA in Therapy 261
2.7 RNA Aptamers in Therapy 261
2.8 Spiegelmers in Therapy 262
2.9 PNA in Therapy 262
2.10 Immunotherapy by RNA 263

3 Design, Synthesis, Delivery, and Pharmacokinetics of RNA 264
3.1 Design and Synthesis of an Effective RNA Drug 264
3.2 Delivery and Pharmacokinetics of RNA 265

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

4 An RNA Drug for Every Disease? 266

Acknowledgment 267

Bibliography 267
Books and Reviews 267
Primary Literature 268

Keywords

Aptamer
A short single-stranded nucleic acid selected from combinatorial libraries to bind specific target molecules through a protocol termed SELEX (Systematic Evolution of Ligands by EXponential enrichment).

Pharmacokinetics
The action and metabolism of pharmaceuticals (drugs) in the body, including the processes of absorption, biotransformation, distribution to tissues, and duration of action and elimination.

PNA (Peptide Nucleic Acid)
A nucleic acid analog in which the entire phosphate sugar backbone has been replaced by an uncharged polyamide backbone with the side groups, purine and pyrimidine bases, found in biological nucleic acids.

Ribozyme
Catalytic RNA, capable of cleaving target RNA of specific sequence.

RNA Drug
Different types of RNA molecules (aptamers, ribozyme, siRNA, and so on) that can be used for therapeutical and/or diagnostic purposes.

RNA interference (RNAi)
A novel mechanism induced by double-stranded RNA that leads to posttranscriptional gene silencing by sequence-specific degradation or translational inhibition of specific transcripts.

Spiegelmer
Biostable aptamers that are synthesized from L-nucleotides (the mirror image of natural nucleotides), and, therefore, cannot be degraded by naturally occurring nucleases.

In the living cell, the RNA is an important intermediate of genetic information transfer between DNA and protein. RNA is a single-stranded, flexible, and versatile biopolymer that incorporates many pharmacologically desirable traits of DNA and protein without some of their disadvantages. It is water soluble and nontoxic, produced by chemical synthesis as well as recombinant cloning, faithfully replicates the parental DNA sequence, can be converted to DNA or protein, possesses enzymatic self-cleaving activity, and makes specific RNA–RNA, RNA – DNA, and RNA–protein interactions. The biologically remarkable forms and functions of RNA, such as antisense RNA, ribozymes, RNA decoys, aptamers, and RNA interference, are rooted in this fundamentally unique combination of biochemical traits of the RNA. It is only recently that some of these properties of RNA have been exploited to interfere with or repair dysfunctional or harmful nucleic acids or proteins and to induce therapeutic gene products in a variety of pathological syndromes or infectious diseases. First and second generations of RNA drugs have indeed produced promising results, thus raising new hopes in bringing RNA therapeutics from the bench to the bedside. In this chapter, we first describe the basic features and discoveries of RNA that are relevant to its clinical usage, and then chronicle the emerging clinical applications of RNA.

1 Clinically Relevant Features of RNA and its Metabolism

1.1 RNA in Central Dogma

The core of the Central Dogma of molecular biology (Fig. 1) charts the genetic information flow in all living cells: DNA makes RNA and RNA makes protein. DNA is copied into RNA in "transcription," and the messenger RNA (mRNA) is deciphered to protein in "translation." To this is added the fact that DNA and RNA genomes are duplicated in "replication," and, in retroviruses, the viral RNA genome can be copied in the "retro" (hence the name of these viruses) or "reverse" direction to produce DNA in a process aptly called "reverse transcription." DNA replication is catalyzed by DNA polymerase and transcription by RNA polymerase (RNAP). The enzyme that copies RNA into RNA is sometimes called "replicase" or "transcriptase" (as the name "RNA polymerase" has already been taken by its DNA-dependent counterpart). In this chapter, however, we will follow the recent trend to call it RdRP (RNA-dependent RNA polymerase). Lastly, the reverse transcriptase (RT) catalyzes the conversion of RNA to DNA. Relevant aspects of the Central Dogma are detailed below.

DNA, the stored but otherwise useless genetic information in all cells, is double-stranded and, in comparison to RNA and protein, is relatively rigid and inflexible. Practically all naturally occurring DNA, except the very small ones (such as plasmids), are intimately complexed with proteins (such as histones in mammalian cells) to produce the chromosomes. The principal types of RNA transcribed from the DNA in a cell are listed in Table 1. Each class of RNA is produced by one of three

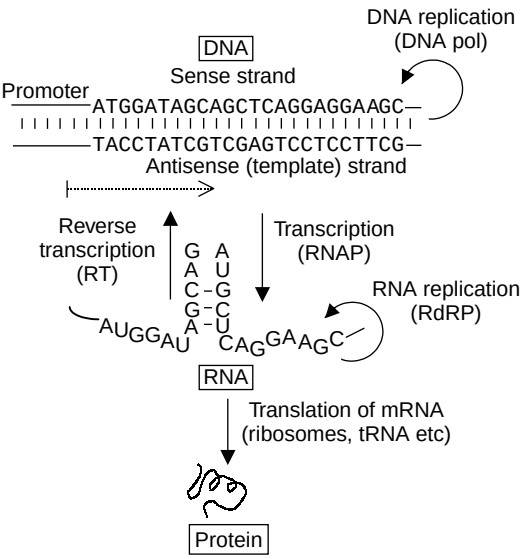

Fig. 1 The Central Dogma of molecular biology (Sect. 1.1). This Figure is not meant to serve as a comprehensive diagram of all aspect of replication, transcription, and translation, but only to provide minimal relevant information needed to understand the rest of the chapter. RNA splicing is omitted for simplicity, but is detailed in Sect. 1.4. The mRNA is shown with its 5′ cap structure and the 3′ poly(A) tail is omitted. An oversimplified RNA hairpin is shown. Nucleotides are paired by hydrogen bonds, indicated by small lines. The most important feature to note is that the RNA is located between DNA and protein, form local base pairs that resemble DNA, and at the same time folds and bends like proteins.

kinds of RNAP, often with the assistance of transcription factors, many of which are in turn regulated by physiological signals. The messenger RNA (mRNA) is the only kind of RNA that is translated into protein; hence its name as the messenger of genetic information. The majority of mammalian mRNAs are produced as pre-mRNAs containing "exons" and "introns"; RNA splicing, described in Sect. 1.4, removes the introns and joins the exons to produce the processed mRNA ready for translation. By virtue of its protein-coding nature, the mRNA has evolved to be more labile and subject to a relatively greater temporal, developmental, and tissue-specific regulation than the rRNA and tRNA. While replication and transcription of DNA genomes occur in the nucleus, translation is cytoplasmic. Replication and transcription of RNA genomes, such as those of viruses, is also a cytoplasmic event, with few exceptions.

RNAP initiates transcription at a specific sequence on the DNA, known as a *promoter*, and copies only one strand of the DNA. There are various conventions to describe the two nucleic acid strands, depending on the focus of the discussion. We prefer to call the RNA the "sense" strand because this is the first step in the Central Dogma where a molecule that "makes sense" (to the translational machinery) is produced from the information encrypted in the DNA. The complementary strand of the DNA is, therefore, "antisense" (Fig. 1). In other words, RNAP copies the antisense strand of the DNA as template to produce the RNA. To rephrase, the RNA sequence is identical to the sense strand of the DNA (except for the T-to-U change) and complementary to the antisense strand. In describing viral RNA genomes, virologists often follow another nomenclature whereby the mRNA-sense RNA is referred to as the "positive" or "plus" strand and its complementary RNA is called "negative" or "minus" strand. As most cellular functions – structural as well as enzymatic – are performed by proteins, the overwhelming majority of conventional drugs directly target proteins. In

contrast, the majority of the emerging RNA drugs target the mRNAs that encode the proteins, and thus work at a higher level in the Central Dogma.

Translation, the last step in Central Dogma, is a far more complex process requiring the participation of ribosomes, transfer RNA, and a slew of translational factors, and, thus, there is no such single entity as "RNA-dependent protein polymerase"! Protein is also the ultimate, dead-end product of Central Dogma, and, thus, cannot be reverse-translated. Lastly, proteins do not replicate. Note that RNA is the intermediate between the DNA and the protein, and shares features of both. In short, it has the nucleic acid property of DNA and the flexibility of protein. The following section elaborates this further.

1.2 Transcription and Replication of RNA Genomes

RNA genomes in nature are found nearly exclusively in RNA viruses that comprise a large group of pathogens that infect practically all species including human. Viral RNA genomes can be linear or circular, single or double-stranded, segmented or nonsegmented, and the single-stranded genomes can be positive or negative, sense or even ambisense (a mixture of the two). They can vary in size from a little over a hundred nucleotides (as in hepatitis delta virus and plant viroids) to around 30 000 nucleotides (as in coronoviruses that cause the severe acute respiratory syndrome, or SARS). Viruses, by definition, are obligatory parasites, and must infect host cells in order to multiply. Owing to their small size, viral genomes lack the capacity to code for the highly complex translation machinery (Fig. 1); therefore, all viruses must utilize the translational apparatus of the host. Animal cells, in general, and human cells, in particular, lack any enzymatic activity that can copy RNA templates, thus compelling all RNA viruses to encode such activities of their own. While the retroviral RNA genomes encode RT that copies the genomic RNA into complementary DNA (cDNA) to be integrated into the host chromosomal DNA, the nonretroviral RNA genomes encode RdRP that transcribes and replicates the genome exclusively in the cytoplasm. The functional viral RdRP is usually a multisubunit complex consisting of a large polymerase subunit and one or more accessory subunits. The RNA genome of positive-strand RNA viruses is essentially a sense-strand mRNA that is translated immediately after infection. The antisense genome of negative-strand RNA viruses, in contrast, must first be transcribed by the associated RdRP activity to produce mRNAs that are then translated to produce new viral proteins. Many RNA genomes, especially the negative-strand ones, are tightly wrapped by the nucleocapsid protein (N), producing the N-RNA nucleoprotein complex, which is the biological template for the RdRP. The mRNAs, in contrast, are naked. Thus, the genomic RNA is resistant to various onslaughts including prospective drugs, while the mRNA is a therapeutic target.

In contrast to the DNA genomes, RNA genomes are also highly mutable, mainly because the RdRP lacks a proofreading activity and thus is highly error prone. As a result, even the purest, clonally purified RNA virus preparation may contain a mixed population of genomes of various sequences, leading to the concept of "quasi-species." This has two important implications. First, it allows the RNA virus to mutate quickly and produce "best-fit" mutants in a rapidly changing environment, thus enhancing its chances

of survival. By the same token, however, it also creates major hurdles in designing vaccines and other antiviral strategies against RNA viruses. Indeed, in recurring RNA viral epidemics, such as the annual episodes of human flu (caused by influenza virus), fresh vaccines are made against the new strain as the antigenic makeup of the new virus may be sufficiently different from those of the previous years, thus making the past vaccines ineffective.

1.3 Noncanonical Base Pairs in RNA

Although RNA is described as single-stranded in comparison to DNA, in reality more than half the nucleotides in the typical RNA participate in base pairing. Several motifs with various combinations of double-stranded helix, bulge, and loop are commonly found in RNA secondary structures. In small viral RNA genomes, such as viroids, virusoids, satellite RNA, and human hepatitis delta virus RNA, essentially all the bases are paired such that the RNA is practically all double-stranded and highly folded. While the Watson-Crick A:T and G:C base pairs are universal in all DNA, RNAs may contain many modified bases in addition to the standard A, C, G, and U, and the flexibility of the RNA chain allows various noncanonical base pairs as well. The most prevalent noncanonical RNA base pairs with the standard bases are GU, GA, and AC. The tRNAs often contain the modified base inosine (I), created posttranscriptionally, in the first wobble position of the anticodon; as I can pair with A, C, or U, this allows more promiscuous codon–anticodon pairing, such that fewer different tRNAs are required. As discussed later, such modified bases and alternate pairing rules may be used in designing better RNA drugs.

1.4 RNA Splicing

Typical mammalian genes are organized on the DNA in exon–intron structure, with an average of about nine exons per gene. Transcription of the DNA produces pre-mRNA, from which the introns are removed in a process called *splicing*. The process is catalyzed by a large ribonucleoprotein complex, called *spliceosome*, containing various RNA of the snRNA family (Table 1) and multiple proteins. The spliceosome recognizes sequence features of the intron–exon boundaries (splice junctions) and performs the actual cutting and joining of the RNA segments. This section will summarize the fundamental principles of how alternate splicing can be regulated by an RNA drug, and specific splice variants or faulty RNA can be destroyed or repaired by catalytic RNA drugs.

1.4.1 Alternate RNA Splicing: Regulation by Antisense RNA

Alternative splicing occurs when the introns of a particular pre-mRNA are spliced in more than one way, generating multiple mature mRNA species – and hence multiple proteins – from a single gene (Fig. 2). More than half of all human gene transcripts are estimated to undergo alternate splicing, yielding an enormous variety of mRNAs out of a relatively small repertoire of genes. As in everything else in biology, alternate splicing is highly regulated, and its pattern can be specific to tissue, development stage, physiological condition, or disease.

A representative example of one kind of alternate splicing and its exploitation

Tab. 1 The principal types of RNA produced in cells.

Type of RNA	*Property*	*Transcribed by*
mRNA	Messenger RNA: codes for proteins.	RNAP II
rRNA	Ribosomal RNA: forms the basic structure of the ribosome and catalyzes translation.	RNAP I
tRNA	Transfer RNA: functions as adaptors between amino acids and mRNA during translation.	RNAP III
snRNA	Small nuclear RNAs: function in a variety of nuclear processes, including splicing.	RNAP III
snoRNA	Small nucleolar RNA: used to process and chemically modify rRNAs.	
SiRNA	Short interfering RNA: double-stranded RNA that eventually degrades target RNA in the RNA interference (RNAi) pathway; important in RNA therapeutics.	RNAP II
miRNA.	microRNA: related to siRNA in biogenesis. These and other small noncoding RNAs (ncRNAs) function in diverse cellular processes, including translation inhibition, X chromosome inactivation, telomere synthesis, and centromere silencing. Important in RNA therapeutics.	RNAP II

Intermediates and precursors are not described. For example, pri-miRNA is first processed by the nuclear RNase, Drosha, to produce pre-miRNA, which is further processed by Dicer in the cytoplasm to generate miRNA (Fig. 6). When first discovered, miRNA was called "small temporal RNA" (stRNA).

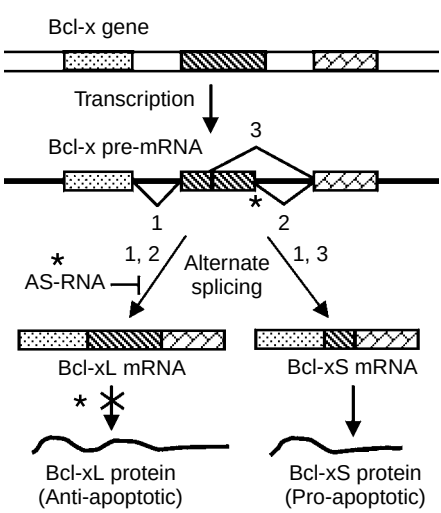

Fig. 2 Regulation of alternate splicing by antisense RNA Drug. The three possible splicing events on the Bcl-x pre-mRNA are labeled 1, 2, and 3. Combination of 1 and 2 splicing yields the short Bcl-xS mRNA and protein, while 1 and 3 splicing yields the long Bcl-xL mRNA and protein. The antisense RNA (AS-RNA, denoted by asterisk) inhibits splicing 3, thus blocking synthesis of the shorter protein and promoting death of cancer cells.

in therapeutics is shown in Fig. 2. Apoptosis or programmed cell death is regulated by differential expression of a large number of proteins by alternative RNA splicing, some of which may even have opposite apoptotic functions. For example, the Bcl-x pre-mRNA generates two alternative splice variants, namely, Bcl-x(long)

(Bcl-xL) and Bcl-x(short) (Bcl-xS). While the Bcl-xL protein is antiapoptotic, Bcl-xS is proapoptotic and sensitizes cancer cells to chemotherapeutic agents. Elevated levels of Bcl-xL generally correlate with decreased cellular sensitivity toward chemotherapy. In a recent approach, an antisense nucleic acid targeted to a complementary alternative splice site of Bcl-x pre-mRNA shifted splicing from Bcl-xL to Bcl-xS, thus lowering the apoptotic threshold of prostate cancer cells and increasing the efficacy of chemotherapeutic drugs. Thus, it is conceivable that a properly designed antisense RNA can selectively alter the splicing pattern of a pre-mRNA to produce a therapeutic effect.

In summary, an antisense RNA, by binding to a specific sequence, can alter the structure and function of the target RNA, which has obvious clinical implications. As discussed in the following sections (Sects. 1.4.2 and 1.4.3), ribozymes and siRNAs (Sect. 1.5) cannot inhibit site-specific splicing, but can be used to destroy a specific splice variant by targeting a sequence that is unique to the variant.

1.4.2 Trans-cleaving Ribozymes: Destruction of Undesired RNA

Ribozymes are RNA molecules that catalyze a chemical reaction, that is, scission of a covalent bond. Historically, the first catalytic RNA was described in a group I intron from *Tetrahymena thermophila* that could undergo self-cleavage (in cis), which was followed by the discovery of trans-cleaving RNA that could act on other RNA substrates. It was soon recognized that the latter class could be used as "catalytic antisense" tools in gene therapy, whereas the purely antisense RNA described above could only inhibit target function without causing loss of the target or alteration of the target sequence. By far, the most favored trans-cleaving ribozymes in clinical applications are the hammerhead ribozymes that share a typical secondary structure of the catalytic core composed of three helices and variable as well as invariant nucleotides at specific positions (Fig. 3a). Whereas helix II is formed intramolecularly, helices I and III are made of paired hammerhead and substrate sequences. Cleavage occurs 3′ to nucleotide H (Fig. 3a) of the target, and efficient cleavage requires a U to the 5′ side of this H to pair with a critical A residue (asterisked in Fig. 3a) in the hammerhead. The variable nucleotides (N in Fig. 3a) in these helices allow one to target virtually any RNA sequence. The repertoire of target sequences can be further expanded by taking advantage of noncanonical RNA base pairing and modified bases (Sect. 1.3). For example, replacement of the asterisked A (in Fig. 3a) with a more promiscuous base, inosine (I), resulted in a hammerhead ribozyme that could cleave at a CH consensus via I:C pairing. In summary, a synthetically engineered hammerhead ribozyme can act as "molecular scissors" destroying unwanted RNA of specific sequence such as a viral RNA, a splice variant, or a mutant RNA associated with a pathological condition such as cancer.

1.4.3 Trans-splicing Ribozymes: Repair of Defective RNA

Once defective genetic information has been transcribed into RNA, it is still possible to use RNA therapeutics to mend the defective RNA sequence without altering the DNA gene. Two major types of strategies have been developed that exemplify the use of trans-splicing ribozymes in RNA repair.

The first method (Fig. 4) employs the originally discovered self-splicing group I intron from *Tetrahymena*, which mediates

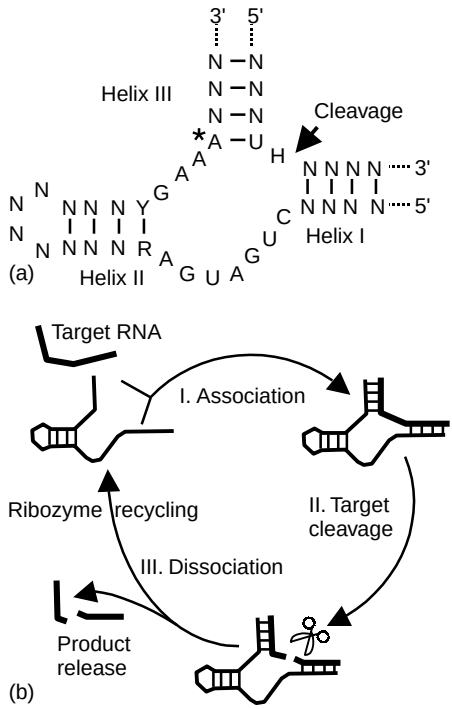

Fig. 3 (a) Secondary structure of a trans-cleaving hammerhead ribozyme with bound substrate. The ribozyme is in plain letters; substrate RNA is in bold. N = any nucleotide, Y = pyrimidine (U or C), R = purine (A or G), H = any nucleotide except G. Note that A:C and G:U base pairs are permissible in all RNA (Sect. 1.3), including ribozymes. The hammerhead ribozyme, associated with the substrate RNA, is depicted to form the three structurally and functionally important helices (I, II, III). In clinical applications, exogenously used hammerheads are generally 35 to 40 nt long, and the intramolecular helix II is formed by four base pairs; however, minimized hammerheads with faster cleavage rates have been described that contain only two base pairs in this helix. The cleavage site in the substrate RNA is as shown. The required UH dinucleotide is circled; this U can be substituted by a C if the pairing A (asterisked) is changed to I. (b) The target (thick line) cleavage cycle of a ribozyme (thin line). The dissociation step (product release) is the slowest and hence rate-limiting, but once dissociated from the cleaved target the ribozyme is recycled as a true catalyst to carry out multiple rounds of cleavage.

trans-splicing of an exon attached to its 3′ end onto a targeted 5′ exon RNA that is a separate RNA molecule. Repair of *lacZ* transcripts in *Escherichia coli* and in cultured mammalian cells using shortened versions of this ribozyme has been demonstrated. In such studies, the ribozyme, which is engineered to contain the wild-type sequence, base pairs with the mutant RNA using an internal guide sequence, cleaves off the mutant segment, and then ligates the wild-type sequence to the cleaved product (Fig. 4). In a variation of this theme, a "twin ribozyme" is created by tandem duplication of a hairpin ribozyme, such that four extra nucleotides are added to the replaced RNA segment. If this is an mRNA, then the resultant repaired mRNA would contain an extra amino acid followed by a frameshift, which may have a therapeutic value.

The second, related method is named spliceosome-mediated RNA trans-splicing (SMaRT) (Fig. 5), which does not require an exogenous ribozyme because it uses the spliceosome to catalyze a trans-splicing reaction between the target pre-mRNA and pre-trans-splicing RNA (termed PTM). Note that the PTM is not a ribozyme; its trans-splicing repairs the mutant RNA by simply competing with the natural cis-splicing.

In an interesting diagnostic application, "half-ribozyme" ligase molecules have

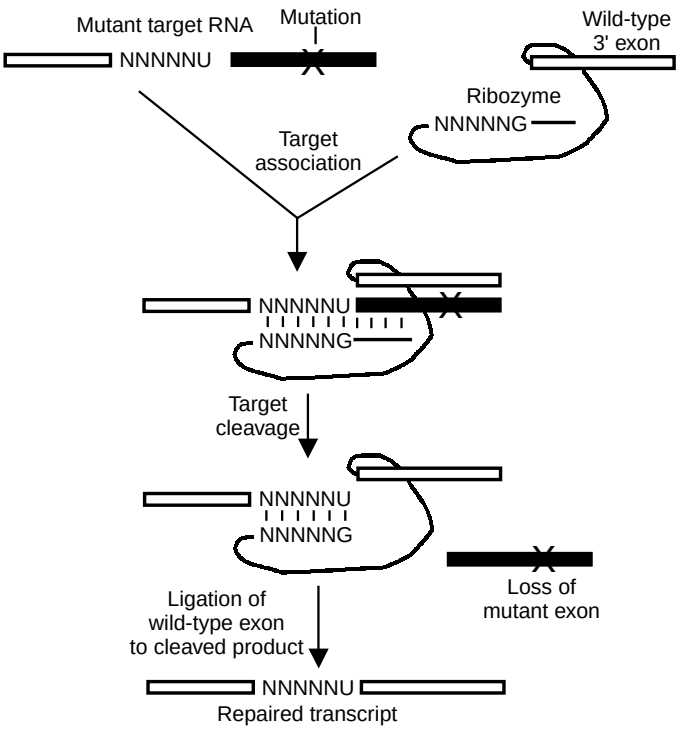

Fig. 4 Ribozyme-mediated repair of RNA. The mutant exon of the target RNA with the mutation X is shown in black. The trans-splicing ribozyme, introduced into the cell, brings in the wild-type exon (white). The N_5G stretch and some extra nucleotides acts as the internal guide sequence to direct trans-cleavage that results in the loss of the mutant exon and ligation of the wild-type exon to produce the repaired transcript (bottom).

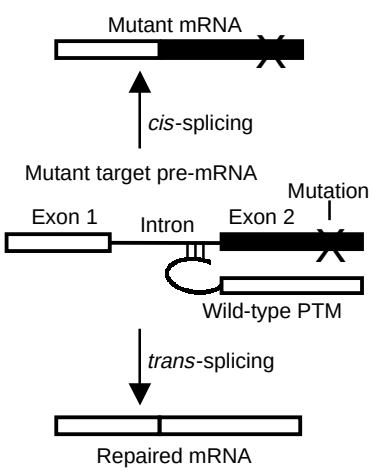

Fig. 5 Spliceosome-mediated trans-splicing repair of mutant RNA by SMART (Sect. 1.4.3). The mutant exon with the mutation X is shown in black and its wild-type counterpart in the PTM in white.

been designed that are activated upon binding to their RNA targets. The catalytic nature of these enzymes leads to multiple rounds of ligation, and hence signal amplification, allowing the detection of attomol (10^{-18}) quantities of the target, such as viral RNA in clinical samples.

1.5 RNA Interference: Sirna and miRNA

A relatively recent discovery, the phenomenon of RNA interference (RNAi) and the various aspects of its mechanism and application, has captivated the imagination of the biological community. A literature search through the PubMed site of NCBI (http://www.ncbi.nlm.nih.gov/) revealed that a total of 146 RNAi-related papers were published in just one month (May of 2004), averaging roughly five papers a day and underscoring the feverish activity in this field. The RNAi pathway is activated by double-stranded RNA (dsRNA), which, in principle, can form whenever sense RNA meets antisense. In nature, segments of many viral genomes are transcribed in both orientations, generating both sense (positive) and antisense (negative) RNA in the infected cell, which may then base pair to form dsRNA. In the laboratory, dsRNA can be synthesized chemically or by recombinant technology and then introduced into the cell, tissue, or animal. The mechanism of the RNAi pathway is briefly illustrated in Fig. 6, whereby a long dsRNA is precisely processed by Dicer, a member of the RNase III superfamily, to produce 21- to 28-nt long dsRNA fragments with 3′-overhangs, called siRNA. This step can be bypassed by chemically synthesized siRNAs that are directly introduced into the cell by transfection or electroporation. The siRNA is incorporated into a multiprotein complex called *RNA-induced silencing complex* (RISC), followed by unwinding of the strands by a helicase activity. The unwinding is asymmetric and initiates from the energetically easier terminus, that is, the one with the higher A–U base-pair

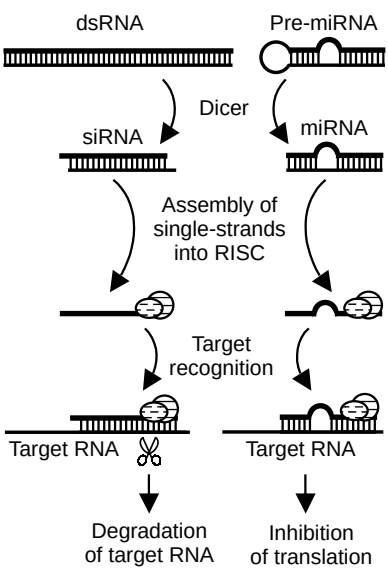

Fig. 6 RNA interference (RNAi). Only the major steps are shown. In contrast to siRNA, the miRNA starts as a hairpin and eventually inhibits the translation of an mRNA with imperfect match, although exceptions are known in that some miRNAs do degrade the target. The protein subunits of RISC are indicated by the spheres, but their exact number remains unclear. It is likely that the siRNA-RISC and miRNA-RISC assemble from common as well as unique subunits.

content. The activated RISC with the antisense strand engages its complementary target RNA. A RISC-associated RNase activity, the exact nature or composition of which remains unknown, then cleaves the target RNA 10 nucleotides from the 5′ end of the siRNA strand, resulting in knockdown or silencing of the corresponding gene function.

Recently, a related class of single-stranded cellular RNA, named microRNA (miRNA), has been discovered that is also 21- to 24-nucleotides (nt) long. They are transcribed from specialized endogenous genes as a ~70-nt-long hairpin precursor. The biogenesis and function of siRNA and miRNA are highly similar, although the full extent of overlap between the two pathways is yet to be determined. Dicer is involved in the processing of both from their respective precursors (Fig. 6), although different Dicers may be preferred by siRNA and miRNA. In general, it appears that siRNA and miRNA that are a perfect or a near-perfect match with their target degrade the target RNA, whereas those with mismatches repress translation instead, although a few exceptions are known. RISC is an exclusively cytoplasmic entity, and, therefore, siRNA and miRNA are only useful against cytoplasmic targets such as mRNA and not nuclear pre-mRNA. Viral genomic RNAs that are either highly structured or covered with protein are also resistant to RNAi.

The exact number of the small noncoding RNAs of the siRNA/miRNA family in the human cell is still unclear, but the current estimate runs in the hundreds. The accumulating evidence indicates that they regulate a variety of normal cellular pathways, including differentiation and development. Owing to their remarkable efficiency and selectivity, the siRNA and miRNA have emerged as major tools for knocking down gene expression in basic scientific as well as clinical applications.

1.6 RNA as a Protein Antagonist: SELEX, Aptamers, and Spiegelmers

The variability of nucleotide sequence and the ability to fold allow single-stranded DNA and RNA to attain a wide range of structures that can interact with high affinity and specificity with other biomolecules. In principle, starting with a large library of DNA or RNA sequences (about 10^{15}), one can select specific molecules or "aptamers" to bind to practically any target, such as a protein, by using a combinatorial and iterative Darwinian-type *in vitro* evolution process that has been termed SELEX for "systematic evolution of ligands by exponential enrichment" (Fig. 7). The aptamers are composed of a central randomized (degenerate) region flanked by fixed primer binding sites on either side for PCR-based amplification. Whereas antisense RNA and siRNA (Sect. 1.5) can target other RNAs only by Watson–Crick base pairing, the aptamers, in contrast, function by forming discrete structures that bind specific targets. Conceptually, aptamers are similar to antibodies in terms of specificity, but have several advantages over antibodies such as stability and chemical synthesis, as elaborated later (Sect. 1.8). Aptamers possess affinity in the low nanomolar to picomolar range and can discriminate between targets that are 96% identical in sequence. RNA aptamers derived from *in vitro* selections (SELEX) are being used inside cells for at least four purposes: (1) to antagonize cellular proteins; (2) as decoys to natural RNA-binding proteins; (3) as regulatory modules to govern gene expression; and (4) to antagonize nonprotein targets.

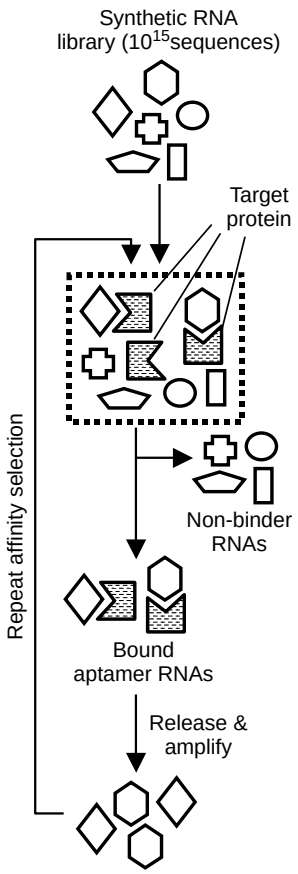

Fig. 7 RNA SELEX and aptamer. In this schematic diagram, the degenerate RNA molecules are depicted as pieces of various sizes and shapes, some of which fit the immobilized target protein (in gray). The unbound RNAs are discarded, and the bound aptamers are eluted and amplified by reverse transcription PCR. They can be further screened through additional cycles of binding and reamplification until the desired strength and specificity of binding is achieved.

Photoaptamers are defined as aptamers with photocross-linking functionality. Photocross-linking, so far, has never lowered the specificity and has, in fact, significantly increased it in some cases. Photoaptamers thus add a second dimension of specificity, since they recognize both the complex shape and charge distribution of their targets and the presence of specific amino acid residues at specific sites.

Spiegelmers (German: "Spiegel", means "mirror") are mirror-image aptamers made of L-nucleotides. The L-enantiomers act just like standard aptamers but have the added advantage of being highly biostable owing to their resistance to nucleases.

1.7 Polyamide ("Peptide") Nucleic Acid (PNA)

PNA (polyamide ("peptide") nucleic acid) is a synthetic nucleic acid analog in which the entire phosphate sugar backbone is replaced by an uncharged polyamide (peptide-like) backbone that is nuclease-resistant. The side groups are still the natural purine and pyrimidine bases and are in similar configuration as in natural single-stranded nucleic acids. Thus, PNAs possess the base-pairing capability and specificity of nucleic acids and hence can function as antisense, but are more stable and bind more strongly and specifically to their DNA or RNA targets.

A unique property of some PNA sequences is to function as antigenes. These PNAs bind to double-stranded DNA by an invasion mechanism; two PNA molecules form a triplex structure with the complementary DNA target sequence, whereas the other strand of the DNA duplex is displaced into a single-stranded loop. Although this type of binding is generally limited to pyrimidine-rich sequences, the resulting PNA–DNA complex is extremely stable. If this happens at the promoter or enhancer sequences, binding of RNAP

or specific transcription factors can be abrogated in cis, leading to silencing of specific genes.

1.8 Summary: Pharmaceutical Advantages of RNA

The relevant advantages of RNA over other forms of traditional and prospective pharmaceuticals are summarized in Table 2. Proteins contain 20 different amino acids, each with its unique chemical and physical property. RNA (or DNA), in contrast, is made of only four nucleotides, thus reducing the complexity. The costs of chemical synthesis of the three biological macromolecules of the Central Dogma are currently in the order: Protein > RNA > DNA. Recombinant expression of protein is an uncertain prospect due to a variety of factors such as mRNA and protein stability and processing, translational regulation, and codon bias. Finally, biological activity of many proteins requires posttranslation modifications such as glycosylation and phosphorylation that do not occur in bacterial expression systems. Purification of proteins in active form can be a significant hurdle due to insolubility and improper folding. A drawback of DNA is the ability of its CpG motifs to activate toll-like receptors and thus induce a general innate immune response, which, in fact, underlies the apparent therapeutic effects of several successful antisense DNA molecules.

Small molecules currently constitute the largest class of drugs and owe their origin to two major sources: natural products, often from folklore, and structure-based drug design. The former is subject to the whims of serendipity and nature's bounty, and the latter requires expensive cutting-edge technology. Synthesis of small-molecule drugs by chemical means has been the norm of the pharmaceutical industry, but the chemical reactions are often challenging and need to be revised for every new compound or new substitution. RNA can be rather easily made *in vitro* as well as *in vivo* by transcription from cloned cDNA. Small RNA molecules, such

Tab. 2 The unique advantage of RNA as pharmaceutical.

Relevant property	*DNA*	*RNA*	*Protein*	*Small molecule*
Chemical synthesis and modification	Yes	Yes	Difficult	Some
Biochemical synthesis	Yes	Yes	Yes	Some
Replication	Yes	Yes	No	No
Sequence-specific interaction with DNA or RNA	Yes	Yes	Some	No
Specific interaction with proteins	No	Yes	Some	Some
Stability without refrigeration	Yes	Yes	No	Some
Intricately folded structure	No	Yes	Yes	No
Catalytic activity	No	Yes	Yes	No
Cloning possible	Yes	Yes	Yes	No
Easy to isolate and purify	Yes	Yes	No	No
Safe for autoimmunity	Yes	Yes	Some	Yes
Nontoxic	Yes	Yes	No	Some
Tissue-specific targeting	Yes	Yes	No	No
Can repair mutant gene or disable expression	Yes	Yes	No	No
Relatively low cost	Yes	Yes	No	No

as siRNA, are better made chemically, and one can expect that the cost will diminish with innovative chemical and process engineering and increased market volume.

Ribozymes and siRNAs are likely to be the major RNA weapon of the future in our war against pathogenic RNA. There are multiple advantages to targeting RNA instead of proteins including sequence-specific binding, a greater number of sites accessible for interaction, selective inhibition, exploitation of multivalent interaction, and more facile attack on drug resistance. RNA's link to genomics and bioinformatics also allows for increased discovery and specificity testing.

2 Emerging Clinical Applications of RNA

2.1 Ribozymes Against Infectious Agents

One of the first obvious viral targets for ribozyme therapy is HIV-1, causative agent of AIDS, due in part to the extraordinary death toll of the disease and the limited efficacy and toxicity of the multidrug cocktails targeting the viral reverse transcriptase. Various sequences in the HIV genome can be targeted for ribozyme-mediated therapy. The two types of ribozymes (hairpin and hammerhead) show similar antiviral efficacy in cell culture in terms of their ability to inhibit the HIV replication and fare well in clinical trials as well. In representative studies, autologous lymphocytes were transduced with a hairpin ribozyme that cleaves the U5 region of HIV-1 RNA, and the resulting cell population consisting of transduced and nontransduced cells was infused. When tested in peripheral blood, the ribozyme-containing cells indeed exhibited preferential survival over non-ribozyme-containing cells. Similarly, hammerhead ribozymes targeted to a conserved region of the HIV *tat* gene (*Rz2*) were transduced into a population of syngeneic CD4+ T lymphocytes. Four patients, infused with syngeneic lymphocytes and subsequently tested by PCR analysis showed hammerhead expression and viral inhibition up to 4 years postinfusion (the latest time point examined). These procedures were well tolerated with no serious adverse events. Together, these results indicate that infusion of gene-altered, activated T cells in HIV-infected patients is safe, and that transduced cells persist for long intervals. Results also suggest that ribozyme-transduced cells may possess a survival advantage *in vivo*.

A different approach for ribozyme gene therapy is to transfer the ribozyme expressing constructs into hematopoietic stem cells of human immunodeficiency virus-1 (HIV-1)-infected individuals. This might be a potent therapeutic approach to provide mature cells arising from transduced stem cells resistant to the destructive events associated with HIV-1 infection. To determine the feasibility of gene therapy for AIDS patients, peripheral blood $CD34^+$ cells were isolated from HIV-1-infected individuals and transduced with retroviral vectors containing three different anti-HIV-1 genes: the RNAbinding domain of the Rev-responsive element (RRE decoy) (L-RRE-neo), a double hammerhead ribozyme vector targeted to cleave the Tat and Rev transcripts (L-TR/TAT-neo), and a trans-dominant mutant of Rev (M10) (L-M10-SN). As a control, a vector mediating only neomycin resistance (LN) was used. After 3 days of transduction on allogeneic stroma in the presence of stem cell factor, and cytokines interleukin-6 (IL-6) and IL-3, the cultures were neomycin-selected, and then challenged with HIV-1.

Compared to the control cultures, the L-RRE-neo-, L-TR/TAT-neo-, and L-M10-SN-transduced cultures displayed up to 1000-fold inhibition of HIV-1 replication after viral challenge. This study shows that anti-HIV-1 genes can be introduced into $CD34^+$ cells from individuals already infected with HIV-1, and that they strongly inhibit HIV-1 replication *in vivo*.

Herpes simplex virus 1 (HSV-1) and human cytomegalovirus (HCMV) are pervasive herpes viruses that cause severe morbidity or mortality in neonates and immunocompromised individuals. In particular, HSV-1 is the causative agent for cold sores and encephalitis in newborns. HCMV is one of the leading causes of birth defects in the United States and is accountable for more than 50% of deaths in organ transplant patients; furthermore, HCMV retinitis is the leading cause of AIDS-related blindness. In recent studies, ribozymes efficiently cleaved mRNAs of both viruses and effectively reduced viral growth in cultured cells. Targeting of the mRNA encoding the major transcriptional activator ICP4 of HSV-1, an immediate-early gene that is essential for the expression of most of the viral early and late genes, caused about 80% reduction in ICP4 and a 1000-fold decrease in viral growth. Meanwhile, other immediate-early gene (e.g. α47 and ICP27) expressions were not affected, showing that the ribozyme is highly specific in targeting the desired mRNA. Ribozyme expression was not cytotoxic, and the expressing cells were indistinguishable from the parental cells in terms of cell growth and viability for up to 2 months. Similar results were also seen when the overlapping region of HCMV immediate-early gene 1 and 2 (IE1/IE2) was targeted by ribozyme. In HCMV-infected human cells, this ribozyme reduced IE1/IE2 expression by 85% and inhibited HCMV growth 150-fold. Reduction of IE1/IE2 expression levels also resulted in the downregulation of other early and late gene products such as US2, UL44, gB, and gH, whereas transcripts that are not regulated by IE1/IE2 (e.g. 5 kb RNA and UL36 mRNA) were unaffected.

Hepatitis C virus (HCV) infection may lead to chronic hepatitis, cirrhosis, and hepatocellular carcinoma. Therapy of chronic HCV infection has greatly improved with the combined use of ribavirin and α-interferon, and, recently, PEG-interferons. However, more than half of the patients do not accrue lasting benefits from these therapies. Ribozyme strategies to target certain HCV sequences have been proposed as anti-HCV treatments. Specifically, the 5′ noncoding region of the HCV RNA is highly conserved and plays an essential role in translation initiation by functioning as an internal ribosome entry site. In several studies, endogenously expressed as well as exogenously delivered ribozymes targeted against this region inhibited HCV replication. In 2000, Ribozyme Pharmaceuticals, Inc (now Sirna Therapeutics) and Eli Lilly and Company completed a 28-day study of Heptazyme following daily subcutaneous injections that showed successful safety, tolerability, and pharmacokinetics. In 2001, phase-2 clinical trials began, designed to evaluate the safety and efficacy of Heptazyme alone and in combination with interferon. Initial results showed that 10% of Heptazyme recipients experienced biologically meaningful reductions in serum HCV RNA levels. In 2002, however, the trial was halted due to blindness that occurred in one animal 12 months after receiving high-dose Heptazyme during toxicology testing.

Urogenital human papillomavirus (HPV) infections are the most common sexually transmitted viral diseases

in women. Recently, a self-processing triple-ribozyme cassette, consisting of two cis-acting hammerhead ribozymes flanking an internal trans-acting hammerhead (ITRz), was tested against HPV-11 E6/E7 mRNA, whereupon significant reductions (80–90%) were achieved for both. A similar ribozyme, designed against hepatitis B virus (HBV), caused a >80% reduction in viral liver DNA in two weeks in a transgenic mouse model.

2.2 Ribozymes Against Cellular Disorders

Recently, ribozymes have been demonstrated as a highly effective treatment against a number of cellular disorders and especially cancer. Point mutations in the K-*ras* gene are observed at a high incidence in human pancreatic carcinomas. Anti-K-*ras* ribozyme against codon 12 of the mutant K-*ras* mRNA, when expressed via a recombinant adenoviral vector, suppressed tumor growth and promoted regression. It was also possible to accomplish efficient reversion of the malignant phenotype in human pancreatic tumors with K-*ras* gene mutation.

It is well known that tumor growth beyond a few cubic millimeters does not occur without the induction of a new vascular supply network, known as *angiogenesis*. Inhibition of angiogenesis has been heralded as a potential approach to cancer therapy. Vascular endothelial growth factor (VEGF) is the best-characterized proangiogenic factor, and effective blockade of the VEGF pathway has been demonstrated with multiple agents such as neutralizing antibody, receptor tyrosine kinase inhibitors, and ribozyme or antisense molecules. Promising preclinical data document the potential of these agents for tumor growth inhibition and even tumor regression, yet translation of novel therapeutics targeting the VEGF pathway to the clinic has proved a substantial challenge in itself. Angiozyme, developed by Sirna Therapeutics, specifically inhibits formation of the vascular endothelial growth factor receptor (VEGFR-1), a key component in the VEGF pathway. This drug is currently in clinical trials for several cancers, including breast and colon cancers. There have been no significant toxicities related to drug use in trials reported and the occurrence of side effects has been low.

Hepatocyte growth factor/scatter factor (HGF/SF) elicits a number of biological activities, including invasion and migration through activation of its tyrosine kinase receptor c-Met, overexpression of which is implicated in prostate cancer development and progression. Targeting the HGF/SF receptor by way of a hammerhead ribozyme is considered an important therapeutic approach in these cancers.

Ribozyme has been particularly useful in targeting oncogenic mRNAs resulting from random chromosomal translocation. For example, translocation between chromosome 9 and 22 fuses the genes of BCR and c-ABL to produce BCR-ABL, resulting in chronic myelogenous leukemia (CML). Ribozymes with the guide sequences that target the fused sequence of the transcript effectively destroys the BCR-ABL fusion transcript found in the cancerous cells of leukemia patients. More importantly, the inhibition is highly specific, as the ribozyme only cleaves the targeted chimeric transcript, but not the normal cellular counterpart, thus making this a promising approach in cancer therapy.

In the management of asthma, hammerhead and hairpin ribozymes targeting conserved sequences within IL-4, IL-5, ICAM, and NF-κB mRNA have been already designed. Pharmacokinetics in mice

revealed that when administered intratracheally, these ribozymes have improved therapeutic benefits compared to traditional drugs.

2.3 RNAi Against Viruses

RNAi-mediated inhibition of virus replication in animal cells was first reported by our laboratory only three years ago using respiratory syncytial virus (RSV) as a prototype model, which opened up a new direction in antiviral drugs. RSV, an RNA virus, is a major cause of respiratory infection in infants and the elderly, and claims a few millions lives throughout the world annually. To this date, there is no reliable vaccine or antiviral. Use of siRNA against essential RSV transcripts inhibited virus growth in cultured lung epithelial cells and, recently, in a mouse model. Silencing of the RdRP subunits, L and P, reduced the progeny viral titer 10^3 to 10^4 fold. No cytopathic effect has been detected in uninfected siRNA-treated cells. A number of viruses have since been targeted by siRNA *in vitro* as well as *in vivo* with highly encouraging results.

As implied earlier, HIV-1 often escapes commonly used therapies because of its high mutation rate and the complexity of the pathogenesis of AIDS. RNAi, targeting various HIV sequences that encode the structural proteins Gag and Env, the reverse transcriptase, the regulatory proteins Tat and Rev, and two accessory proteins Nef and Vif, were all effective to various degrees in inhibiting HIV growth. RNAi targeting of nontranslated RNA sequences in the long terminal repeats (LTRs) of HIV, which contains important regulatory elements required for HIV replication, was also effective. In principle, one can also inhibit virus infection by silencing the expression of essential cell factors that are critically involved in viral life cycle. Inactivation of the cellular receptor (CD4) and coreceptor (CXCR4, CCR5) by specific siRNA, in fact, reduces HIV infection of T cells.

Influenza A virus causes widespread infection in the human respiratory tract, but existing vaccines and drug therapy are of limited value. siRNAs, against influenza's nucleocapsid or a component of RdRP, abolish viral transcription and replication in cell lines and embryonated chicken eggs. These results should provide a basis for the development of siRNA as preventive treatment and therapeutic drug for influenza infection in humans.

RNAi has also been used against several other viruses, including Dengue virus, flock house virus (FHV), rhesus rotavirus (RRV), Semliki forest virus (SFV), and poliovirus. All hepatitis viruses, for example, hepatitis B (HBV), C (HCV), and delta (HDV), are major public health hazards; RNA interference appears to be an attractive option against them. siRNA against the core region of HBV, when cotransfected with the full-length HBV DNA into Huh-7 and HepG2 cells, effectively inhibited viral growth in cell culture and mice. Essentially, similar results were obtained with siRNA against the *S* gene of HBV. HCV, a *Flaviviridae* with a small RNA genome, is a uniquely difficult virus to study because of the lack of *in vitro* cell culture models. However, recently, subgenomic replicon systems for HCV have been developed using siRNAs against viral NS3 and NS5B as well as against the 5′-UTR, and they were shown to effectively inhibit viral replication. In HDV, recent studies showed that the delta antigen mRNA can be successfully targeted by siRNAs in cell culture.

In contrast to viral mRNAs, the genomic and antigenomic RNAs of many viruses are resistant to siRNA action. This is either due to their encapsidated nature (as in negative-strand RNA viruses) or possibly due to extensive secondary structure (as in HDV). Early studies had indeed confirmed that the RNAi machinery is exclusively cytoplasmic and fails to knock down unexported nuclear pre-mRNA.

2.4 RNAi as an Antiparasitic Approach

Parasites are lower eukaryotes and cause diseases of considerable medical and veterinary importance throughout Africa, Asia, and the Americas. The advantages of RNA interference have offered new hope, and siRNA drugs have indeed caused substantial inhibition of a number of parasites in the last few years. Among the parasites where RNAi has been extensively demonstrated is the protozoan parasite, *Trypanosoma brucei* (African sleeping sickness), and *Schistosoma* or flukes (liver, kidney, and intestinal damage). It should be noted, however, that the exact mechanism of siRNA action in many parasites remains to be elucidated.

The Apicomplexan parasite, *Toxoplasma*, can cause blindness, mental retardation, and miscarriage. We have recently discovered that the genome of *Toxoplasma gondii* contains orthologs with significant similarity to traditional RNAi-related genes, such as Argonaute, Dicer, RdRp, and various RDE-related genes. Expression of dsRNA, homologous to parasitic uracil phosphoribosyltransferase (UPRT) mRNA, but not of antisense RNA alone, was shown to downregulate the function of the corresponding enzyme. However, the connection of these results with the RNAi pathway needs to be established. RNAi-like phenomena were also reported in *Plasmodia* (another Apicomplexan parasite and causative agent of malaria), but the mechanism(s) remain uncertain because of the absence of any obvious orthologs of the RNAi pathway. Validation of the RNAi pathway will provide a major breakthrough in the genetic manipulation of these otherwise difficult organisms.

Vector-borne infections, including many mosquito-borne viruses and arboviruses, in general, pose a tremendous burden to the human society, particularly in the developing countries. Studies have shown that RNAi can be used to knock down gene expression both in the adult mosquito and immortalized cell lines. The RNAi mechanism could potentially be used to develop transgenic vectors that provide innate resistance to the development and growth of arthropod-borne pathogens. To explore RNA silencing in disease vectors, new studies have taken advantage of the recent completion of genome sequence of *Anopheles gambiae*, the main vector for human malaria in Africa. Homology searches revealed two Dicer-like and five Argonaute-like genes in *A. gambiae*. siRNA directed against mRNAs of these representative genes indeed caused a reduction of RNAi efficiency in *A. gambiae* Sua1B cells, confirming their functional role. Thus, RNAi may become a widely undertaken approach in mosquito management.

2.5 RNAi Against Genetic Disorders and Cancer

The fastidious sequence-specificity of siRNA allows one to specifically target the dominant and codominant mutations that underlie many noninfectious diseases. Recently, siRNAs have proven

highly effective against a number of genetic abnormalities such as autoimmune diseases, asthma, AMD, diabetes, and diabetic retinopathy, CNS disorders, obesity, ALS or Lou Gehrig's disease, and various forms of cancer.

Specific inhibition of the oncogenic K-RAS V12 expression in human tumor cells by siRNA resulted in loss of anchorage-independent growth and tumorigenicity. Since siRNA-mediated gene silencing is highly sequence-specific, siRNA can be designed to silence cancer-derived transcripts that harbor point mutations. The guardian of the genome, p53, is inactivated by point mutation in >50% of human cancers. A single base difference in siRNA discriminated between mutant and wild-type tumor suppressor p53 in cells expressing both forms, resulting in the restoration of wild-type protein function.

Unnatural fusion proteins are relatively common in cancer. Transfection of leukaemic cells with siRNAs targeting a BCR–ABL (breakpoint cluster region – Abelson protooncogene) fusion transcript induced apoptosis comparable to that triggered by the ABL tyrosine kinase inhibitor STI571 (signal transduction inhibitor 571). In Ewing's sarcoma, the transcription factor EWS is fused to a variety of other transcription factors such as FLI and Ets at various breakpoints. The NPM-ALK fusion protein, for example, is found in about 75% of pediatric anaplastic large cell lymphomas. Specific siRNA drugs designed against such chimeric mRNA may constitute a potent and specific form of anticancer therapy. Osteosarcoma is the most common highly malignant bone tumor with primary appearance during the second and third decades of life. It is associated with a high risk of relapse, often resulting from resistance developed to chemotherapy agents. Interestingly, the apurinic endonuclease 1 (APE1) is usually overexpressed in human osteosarcoma. Using siRNA against APE1, protein levels were reduced by more than 90% within 24 h, remained low for 72 h, and returned to normal levels at 96 h. There was also a clear loss of APE1 endonuclease activity following APE1-siRNA treatment. A decrease in APE1 levels in siRNA-treated human osteogenic sarcoma cells led to enhanced cell sensitization to DNA damaging and chemotherapeutic agents, thus improving prognosis.

At least eight human neurodegenerative disorders, including Huntington's disease (HD) and spinobulbar muscular atrophy (SBMA) (Kennedy's disease), are caused by expansion of trinucleotide repeats, most common among which is the repeat of the CAG codon, coding for glutamine. Transfection of cells expressing the CAG-expanded androgen receptor mRNA shows response to siRNA treatment by a reduction in mutated RNA level and, more importantly, by a rescue of the polyglutamine-induced toxicity. This example represents a proof of principle to apply siRNA technology to diseases associated with mutated transcripts arising from one allele (e.g. other neurodegenerative disorders) without affecting the other (healthy) allele.

Recently, anti-VEGF siRNAs have received much attention owing to their efficacy in preventing angiogenesis. In the mouse model, targeting VEGF by injecting siRNA into tumors resulted in growth inhibition, tumor cell killing, and sensitization of the treated cells to other therapies. Anti-VEGF siRNAs therapy of age-related macular degeneration (AMD) and diabetic retinopathy (DR) is a major area of pharmaceutical interest.

2.6 Antisense RNA in Therapy

Antisense (AS) therapy has been widely used to specifically and selectively inhibit the expression of selected genes at the mRNA level. Use of AS-RNA against cancer-associated mRNA may lead to specific protein silencing and death of the cancer cells. Indeed, the anti-apoptotic protein Bcl-2 and telomerase were both successfully silenced in cultured cells, leading to cell growth inhibition. The insulin-like growth factor 1 receptor (IGF-1R) is an important signaling molecule in cancer cells and plays an essential role in the establishment and maintenance of the transformed phenotype. Inhibition of IGF-1R signaling thus appears to be a promising strategy to interfere with the growth and survival of cancer cells. AS-RNA, designed to inhibit IGF-1R gene expression, efficiently reduced IGF-1-dependent proliferation and survival in a number of human and rodent cancer cell lines. Furthermore, a decrease in tumor size occurred when cells carrying the AS-IGF-1R were injected into syngeneic mice. In addition to blocking tumor growth, the AS-RNA treatment also inhibited metastasis. Many other genes have been selected as targets for AS-therapy, including HER-2/neu, PKA, TGF-α, EGFR, TGF-β, P12, MDM2, BRCA, Bcl-2, ER, VEGF, MDR, ferritin, transferrin receptor, IRE, C-fos, HSP27, C-myc, C-raf, and metallothionein. In many of these studies, specific inhibition of tumor cell growth was demonstrated. Combinations of AS-RNA with chemotherapeutic agents may offer important advantages in cancer treatment. Several antisense drugs, especially Oblimersen (G3139), have shown promising results in animal experiments, and have entered clinical trials.

2.7 RNA Aptamers in Therapy

Several highly specific, nucleic acid aptamers targeting select HIV proteins have been described that effectively blocked viral replication. Aptamers directed against HIV-1 Rev protein were tested in a surrogate animal model harboring human tissue. In the SCID-hu mouse *in vivo* human thymopoiesis model, differentiated thymocytes derived from reconstituted grafts expressed anti-Rev aptamers and showed significant resistance to HIV-1 infection upon challenge.

Blood clotting, which can trigger heart attacks and strokes, is a significant area of RNA aptamer application. Although anticlotting drugs are currently available, most have serious drawbacks. For example, heparin and its antidote, used during and after surgery respectively, can often cause adverse reactions; the dosage of another common blood-thinning drug, Warfarin, is difficult to regulate, and it has no antidote. One trillion RNA aptamers were recently screened for the ability to block specific protein factors crucial to the blood clotting process. Once a clot-stopping aptamer was chosen, its antidote was also designed, which is simply another length of RNA with a complementary sequence. This second RNA snippet sticks to the first, disabling it. In the test tube, different amounts of antidote either regulated or reversed the aptamer's anticlotting ability. Studies in animals are now under way.

The treatment of immunomediated glomerulonephritides is presently based upon a limited series of drugs. Several original and innovative approaches to treat inflammatory glomerular diseases have been recently reported, including RNA drugs designed to limit the

effect of proinflammatory and prosclerotic cytokines. The application of peptide aptamers specifically binding to the IGF-1R represents a novel approach to target IGF-1R signaling in cancer. Integration of peptide aptamers into targeted protein degradation vehicles and their transduction into cells allows the temporary elimination of the receptor protein.

Myasthenia gravis is a neuromuscular disorder associated with muscular weakness and fatigability. The pathogenesis of MG mainly results from antibody-mediated autoimmune response to nicotinic acetylcholine receptors (AChR) located in the postsynaptic muscle cell membrane. Recently, aptamers were successfully used to treat experimental autoimmune myasthenia gravis (EAMG) in animals. Clinical symptoms of EAMG in rats were efficiently inhibited by a truncated RNA aptamer, but not by control scrambled RNA. Moreover, the loss of AChR in the animals induced by the antibody was also significantly blocked with the modified RNA aptamer.

Aptamers can also be used against parasites. African trypanosomes cause sleeping sickness in humans and Nagana in cattle, multiply in the blood and escape the immune response of the infected host by antigenic variation, that is, periodic change of the parasite surface antigen, which consists of a variant glycoprotein known as *variant surface glycoprotein* (VSG). Aptamers that bind to VSGs with subnanomolar affinity are able to recognize different VSG variants and bind to the surface of live trypanosomes. Aptamers tethered to an antigenic side group are capable of directing antibodies to the surface of the parasite *in vitro*.

2.8 Spiegelmers in Therapy

Spiegelmer that inhibits the action of the migraine-associated target calcitonin gene-related peptide 1 (alpha-CGRP) was identified as a lead compound for *in vivo* studies. Gonadotropin-releasing hormone (GnRH) is a key peptide hormone in the regulation of mammalian reproduction. It is the trigger signal for the cascade of hormones responsible for controlling the production of the gonadotropins, luteinizing hormone (LH), and follicle-stimulating hormone. GnRH and its receptor, therefore, have been identified as therapeutic targets for sex steroid–dependent conditions such as prostate cancer, breast cancer, and endometriosis, as well as in assisted-reproduction techniques. A spiegelmer with high affinity for gonadotropin-releasing hormone (GnRH) was isolated that acted as an antagonist to GnRH in Chinese hamster ovary cells stably expressing the human GnRH receptor. In a castrated rat model, the spiegelmer further demonstrated strong GnRH antagonist activity. These studies suggest that spiegelmers could be of substantial interest in the development of new pharmaceutical approaches against GnRH and other targets.

2.9 PNA in Therapy

PNAs demonstrated significant promise against papillomavirus-induced human cancers. Approximately 100% of cervical carcinomas are caused by infections with human papillomaviruses (HPVs). The expression of the *E6* and *E7* genes from high-risk HPV16 and 18 is crucial for development, immortalization, and maintenance of the malignant phenotype

of cervical carcinoma, and constitutes important targets for anticancer therapies. Different PNAs directed against HPV18 *E6* and *E7* genes were able to regulate the growth of HeLa-S cervical cancer cells. Telomerase activity, which is below detectable level in almost all types of diploid cells, is re-activated in most immortal and cancer cells. In recent studies, PNAs directed against the human telomerase RNA (hTR) and catalytic subunit (human telomerase reverse transcriptase, hTERT), effectively arrested the growth of gastric cancer cells. In another study, upregulation of the *c-myc* oncogene in Burkitt's lymphoma cells was inhibited by PNA complementary to a specific unique E mu intronic sequence and blocked the expression of the *c-myc* oncogene under E mu control.

PNAs were also effective antivirals. The RNA genome of HCV contains a well-defined and highly conserved secondary structure that functions as an internal ribosomal entry site (IRES), necessary for translation and viral replication. PNAs can invade critical sequences within the HCV IRES and thereby inhibit translation. In cells infected with pseudotyped HIV-1 virions, PNAs exhibited dramatic reduction of HIV-1 replication.

2.10 Immunotherapy by RNA

With the realization that cancer or malignant tumor is a disease of defective genetic programming, attempts have been made to identify tumor-specific proteins and train the body's immune system against these prospective antigens. The strategy of immunotherapy of cancer is based on the knowledge that nonself (or foreign) proteins are proteolytically degraded inside bone marrow–derived dendritic cells (DCs) to produce short peptides that associate with the major histocompatibility complex (MHC) and are transported to the cell surface. Naïve cytotoxic T cells (CTLs) recognize the displayed peptide–MHC complexes and undergo an activation process to kill the targets. In the cancer patient, the capture of tumor antigens by DCs and/or the stimulation of tumor-specific CTLs is apparently inefficient. In RNA immunotherapy, the strategy is to transfect DCs with mRNA for specific tumor antigens or with total tumor-derived mRNA population, and introduce them into the patient. In animals as well as human volunteers, DCs loaded with tumor mRNAs indeed stimulated CTL response. As with other mammalian cells, cationic lipid reagents and electroporation have been used for mRNA transfection into DCs; interestingly, mRNA alone is effective, which reflects the extraordinary sensitivity of the immune system to small amounts of antigen. The success of the mRNA loading obviates the need for difficult and laborious alternatives such as cloning the mRNAs into cDNA or expression and purification of tumor-specific proteins. In fact, in direct comparison, mRNA-loaded DCs often fare better than those transfected with cDNAs or proteins. Endosomal compartments of DCs also contain specialized toll-like receptors that are activated by single-stranded RNA, resulting in the activation of IFN-γ. This apparent side effect can be actually beneficial for therapy against viruses and cancer. RNA immunotherapy does not require an understanding of the mechanism of tumorigenesis, and yet offers a natural and biological anticancer treatment that can be custom-designed and administered against a particular tumor in a matter of hours.

3
Design, Synthesis, Delivery, and Pharmacokinetics of RNA

3.1
Design and Synthesis of an Effective RNA Drug

A number of algorithms and software have been developed to design ribozymes, siRNA, and aptamers, and analyze their interaction with prospective targets *in silico*, although many of them are based on semiempirical and thermodynamic studies such that the final selections must be experimentally tested for efficacy. Many such programs and other RNA-related resources are available at various web sites.

The principles of Ribozyme design have been discussed (Sects. 1.4.2 and 1.4.3). The siRNAs are 21- to 23-nt-long, double-stranded RNA with 2-nt overhangs. Synthetic siRNAs are generally made to conform to the sequence $NA(N)_{19}TT$ where N can be any nucleotide; however, variations of the overhangs have been tried with success. The $(N)_{19}$ core should be perfectly complementary to the target RNA sequence, but some mismatch may be tolerated, especially if it occurs near the termini of the siRNA. Recently, a set of eight rules has been suggested for rational design of the $(N)_{19}$ core of the siRNA on the basis of experimental analysis and thermodynamic annealing parameters. A few important ones are as follows: 30 to 52% GC content, absence of internal repeats or hairpins, and, most importantly, three or more A/Us at positions 15 to 19 of the sense strand so that preferential unwinding occurs from this end of the siRNA duplex, resulting in the formation of a RISC with the antisense strand (Fig. 6). siRNA and degradative ribozymes should be designed away from the 5′ and 3′ termini of the target mRNA as these regions interact with, and are protected by, translational machinery or factors.

RNA can be synthesized either exogenously (*in vitro*) and then delivered into cells, or transcribed endogenously (*in vivo*) from DNA clones introduced into cells. Exogenous RNA can be made using synthetic chemistry or by transcription of DNA clones. A major issue in the therapeutic use of RNA is its stability. The only difference between DNA and RNA (besides the T to U change) is the 2′-OH group of the ribose ring in the RNA. This has important consequences, the most prominent of which is the hydrolysis of the phosphodiester bond of the RNA catalyzed by nucleophilic attack of the electrons of the extra oxygen atom – fundamentally the same reaction that is catalyzed by ribozymes. A major effort in synthetic RNA chemistry is, therefore, directed toward modifying the 2′-OH group. An ideal modification must provide improved stability and better pharmacokinetics without affecting the base-pairing characteristic or function of the RNA. Notable modifications include amino, fluoro, methyl, and allyl derivatives of the 2′-OH group. Another common modification is to replace the oxygen atoms of the phosphodiester bonds by sulphur. The resulting phosphothioester bond is unhydrolyzable, but it is now realized that the phosphorothioate RNA exhibit significantly higher cellular toxicity and nonspecific binding to proteins. The current trend is to use a limited number of phosphorothioate linkages combined with 2′-modifications. In another approach, an inverted T is added at the 3′ end to form a 3′-3′ phosphodiester linkage, which makes the RNA resistant to 3′ exonucleases. In designing a siRNA, deoxythymidines (dT) are used to substitute for the two T overhangs with the hope of

increasing the nuclease-resistance at the 3′ end. Ribozyme cores are also susceptible to hydrolysis by endonucleases that primarily attack the pyrimidine nucleotides. Thus, 2′-modification of the two U nucleotides of the CUGAUG consensus (Fig. 3a) produces a more stable hammerhead without compromising ribozyme activity. In yet another approach, deoxyribozymes have been developed that exhibit some advantages over RNA ribozymes, such as greater stability, improved catalytic efficiency, and potentially lower toxicity.

Clearly, chemical modifications are only possible *in vitro* and not *in vivo*, when the RNA is being produced through transcription inside a cell. On the other hand, *in vivo* synthesis of RNA from recombinant DNA clones obviates the need for manufacturing and delivering the RNA and uses the natural transcriptional machinery of the target cell itself. The RNA is only produced inside the cell, thus avoiding losses in blood and degradation by serum nucleases. A variety of vectors have been used for *in vivo* expression of RNA, the most common ones being either plasmids or viral vectors with strong promoters. Small RNA molecules such as ribozymes and siRNA are often transcribed from a RNAP III promoter engineered into these vectors, such as the U6, H1, or 7SK promoters of mouse or human origin. Controlled expression of RNA is often achieved through the use of inducible and tissue-specific promoters.

Like small molecules, RNA drugs are highly amenable to automated high-throughput screening (HTS) procedures, based either on direct binding to targets or function-based alteration of reporter gene expression. Combined with chemical modification of synthetic RNA, such assays can be adapted to use almost any kind of readout format, including, but not limited to, fluorescence intensity (FI), fluorescence lifetime (FLT), fluorescence polarization (FP), fluorescence resonance energy transfer (FRET), solid-state (membrane or bead) binding assay, enzyme-linked assay, and radioactivity.

3.2 Delivery and Pharmacokinetics of RNA

The issues of delivery of an exogenously made RNA drug are not much unlike those of DNA transfection and gene therapy. In cell culture, Oligofectamine (Life Technologies, Gaithersburg, MD, USA) and TransIT-TKO reagent (Mirus Corp., Madison, WI, USA) are used by many laboratories with success. Cellular permeation is also improved by conjugation with specific peptides such as helical peptides, Tat protein of HIV, and Antennapedia of *Drosophila*. In live animals, consistent delivery of sufficient quantities of RNA remains a challenge. In mice, "hydrodynamic injection" through the tail vein effectively delivers the RNA into hepatocytes. The optimum amount is 10 to 15% of the animal's body weight injected within 5 to 7 s. Injection of a large bolus is believed to result in short-term right heart failure and backflow of a large volume into the liver. However, hydrodynamic injection through the tail vein may not transport the RNA to all cells of the body, and is an impossible procedure in human subjects.

Once the RNA is delivered, the major issues are its pharmacokinetic properties, namely, stability in tissues and body fluids, metabolism and urinary excretion, and the potential toxicity of large amounts of RNA needed for the intended therapeutic effect. Although generally well tolerated, each RNA needs to be tested for these parameters because it may have unique effects on cellular gene expression. This

is particularly important in designing second-generation RNA drugs conjugated to novel non-RNA moieties, which results in unique conjugates that do not exist in nature. It is also to be remembered that each tissue or organ may have unique interactions with RNA. The uptake and distribution of RNA in tumor tissues are typically poor compared to normal tissues. While working with live animals and human patients, considerations must also be given to the possibility that viral vectors may cause systemic infections and immune reactions.

4
An RNA Drug for Every Disease?

In this article, we have detailed a relatively large number of recent and emerging clinical applications of various forms of RNA. Despite its phenomenal prospect, however, it should be borne in mind that RNA is a relatively new entrant in the pharmaceutical arena, and considerable work still needs to done for RNA-based drugs to become common items in the family medicine cabinet. The major areas where improvement is still desirable are as follows: cost, delivery, stability, and specificity. The first three issues have been discussed. When compared to most small molecular drugs, RNA is in fact highly specific, primarily because essentially all of its clinical applications are sequence-dependent. Double-stranded RNAs (dsRNAs) that are longer than about 35 base pairs tend to trigger the so-called "interferon response," in which the dsRNA binds to and activates dsRNA-activated protein kinase, PKR. Among the many cellular proteins that are substrates of PKR, a strategically important one is the translation initiation factor, eIF2α, phosphorylation of which leads to global translational shutoff and sometimes, cell death. Primarily by virtue of their shorter length, siRNAs do not activate the IFN response, which is the key to their target-specific effect. However, siRNAs sometimes do affect off-target gene expression, especially when large amounts are applied. This is, in part, due to their ability to tolerate some degree of mismatch. The mechanism or extent of such nonspecificity and its potential impact on the clinical applications of RNA remain an area of active debate. On a practical note, it should be realized that there is no chemical or medicine that is totally free of side effects, especially when subjected to exquisitely sensitive molecular biological screening such as microarray analysis, and that the real-life issue in medicine is the balance between risk and benefit.

In principle, multiple RNA-based strategies can be applied to a given target. For example, an mRNA can be silenced by antisense RNA, ribozyme, or siRNA. Currently, there is no clear *a priori* guideline to choose one over the others, primarily because there are few studies comparing them under identical conditions. The general consensus is that siRNAs may perform as well as or better than antisense and ribozyme. One can envisage that their relative effectiveness would be influenced by a variety of factors, including tissue or cell type, transfection technique, target sequence, and chemical modification.

From an entrepreneurial perspective, the annual worldwide market for RNA-based therapy is estimated to be as high as $200 billion (US). The excitement in this area is underscored by the growing number of Biotech

companies that have added RNA-based drugs or reagents in their R&D portfolio. Some of the major names are as follows: Acuity Pharmaceuticals, Ambion, Antisense Pharma GmbH, Archemix Corporation, AVI BioPharma, Avocel, Benitec, CytRx, Dharmacon Research, Devgen, Enzo Biochem, EpiGenesis Pharmaceuticals, Genta, Hybridon, Intradigm, Imgenex Corporation, Immusol, Intronn, International Therapeutics, Intradigm, Isis Pharmaceuticals, Lorus Therapeutics, Merix Bioscience, MethylGene, Mirus Corporation, NascaCell GmbH, NeoPharm, Neucleonics, NOXXON Pharma AG, Polgen, PTC Therapeutics, Qiagen NV, Ribozyme Pharmaceuticals, Sequitur, Sirna Therapeutics, SomaGenics, and SomaLogic.

In the past few years, RNA drugs of virtually every category described here have progressed through different stages of development including clinical trials, and have produced encouraging results. Indeed, the breadth and scope of the emerging clinical applications of RNA are only matched by the diversity of the biological tasks assigned to RNA by Mother Nature. It is fair to say that the full potential of RNA as a pharmaceutical entity has only begun to be appreciated, and that RNA drugs against a variety of diseases and infections will achieve blockbuster status in the foreseeable future.

Acknowledgment

We apologize to readers and colleagues for often referring to comprehensive reviews instead of original research papers due to space limitations. Work in our laboratory was generously supported by Burroughs Wellcome Foundation, American Heart Association Southeast Affiliate (AL G970031), and NIH, USA (AI045803, EY013826, F32 AI049682, AI37938). We acknowledge Titus Barik for help in sequence analysis and other computational projects.

See also Gene Targeting; Oncology, Molecular.

Bibliography

Books and Reviews

Ahlquist, P. (2002) RNA-dependent RNA polymerases, viruses, and RNA silencing, *Science* **296**, 1270–1273.

Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function, *Cell* **116**, 281–297.

Burgstaller, P., Jenne, A., Blind, M. (2002) Aptamers and aptazymes: accelerating small molecule drug discovery, *Curr. Opin. Drug Discov. Devel.* **5**, 690–700.

Demidov, V.V., Frank-Kamenetskii, M.D. (2004) Two sides of the coin: affinity and specificity of nucleic acid interactions, *Trends Biochem. Sci.* **29**, 62–71.

Dorsett, Y., Tuschl, T. (2004) siRNAs: applications in functional genomics and potential as therapeutics, *Nat. Rev. Drug. Discov.* **3**, 318–329.

Doudna, J.A., Cech, T.R. (2002) The chemical repertoire of natural ribozymes, *Nature* **418**, 222–228.

Mattick, J.S. (2003) Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms, *Bioessays* **25**, 930–939.

Nielsen, P.E. (2002) PNA technology, *Methods Mol. Biol.* **208**, 3–26.

Peracchi, A. (2004) Prospects for antiviral ribozymes and deoxyribozymes, *Rev. Med. Virol.* **14**, 47–64.

Scherer, L.J., Rossi, J.J. (2003) Approaches for the sequence-specific knockdown of mRNA, *Nat. Biotechnol.* **21**, 1457–1465.

Sullenger, B.A., Gilboa, E. (2002) Emerging clinical applications of RNA, *Nature* **418**, 252–258.

Vater, A., Klussmann, S. (2003) Toward third-generation aptamers: Spiegelmers and their therapeutic prospects, *Curr. Opin. Drug. Discov. Devel.* **6**, 253–261.

Primary Literature

Adelman, Z.N., Blair, C.D., Carlson, J.O., Beaty, B.J., Olson, K.E. (2001) Sindbis virus-induced silencing of dengue viruses in mosquitoes, *Insect Mol. Biol.* **10**, 265–273.

Adelman, Z.N., Sanchez-Vargas, I., Travanty, E.A., Carlson, J.O., Beaty, B.J., Blair, C.D., Olson, K.E. (2002) RNA silencing of dengue virus type 2 replication in transformed C6/36 mosquito cells transcribing an inverted-repeat RNA derived from the virus genome, *J. Virol.* **76**, 12925–12933.

Al-Anouti, F., Ananvoranich, S. (2002) Comparative analysis of antisense RNA, double-stranded RNA, and delta ribozyme-mediated gene regulation in *Toxoplasma gondii*, *Antisense Nucleic Acid Drug Dev.* **12**, 275–281.

Alvarez-Salas, L.M., Benitez-Hess, M.L., DiPaolo, J.A. (2003) Advances in the development of ribozymes and antisense oligodeoxynucleotides as antiviral agents for human papillomaviruses, *Antivir. Ther.* **8**, 265–278.

Anderson, J., Banerjea, A., Planelles, V., Akkina, R. (2003) Potent suppression of HIV type 1 infection by a short hairpin anti-CXCR4 siRNA, *AIDS Res. Hum. Retroviruses* **19**, 699–706.

Barik, S. (2004) Control of nonsegmented negative-strand RNA virus replication by siRNA, *Virus Res.* **102**, 27–35.

Barik, S., Rud, E.W., Luk, D., Banerjee, A.K., Kang, C.Y. (1990) Nucleotide sequence analysis of the L gene of vesicular stomatitis virus (New Jersey serotype): identification of conserved domains in L proteins of nonsegmented negative-strand RNA viruses, *Virology* **175**, 332–337.

Bartel, D.P., Szostak, J.W. (1993) Isolation of new ribozymes from a large pool of random sequences, *Science* **261**, 1411–1418.

Been, M.D., Barfod, E.T., Burke, J.M., Price, J.V., Tanner, N.K., Zaug, A.J., Cech, T.R. (1987) Structures involved in Tetrahymena rRNA self-splicing and RNA enzyme activity, *Cold Spring Harb. Symp. Quant. Biol.* **52**, 147–157.

Berget, S.M. (1995) Exon recognition in vertebrate splicing, *J. Biol. Chem.* **270**, 2411–2414.

Bitko, V., Barik, S. (2001) Phenotypic silencing of cytoplasmic genes using sequence-specific double-stranded short interfering RNA and its application in the reverse genetics of wild type negative-strand RNA viruses, *BMC Microbiol.* **1**, 34.

Bitko, V., Oldenburg, A., Garmon, N.E., Barik, S. (2003) Profilin is required for viral morphogenesis, syncytium formation, and cell-specific stress fiber induction by respiratory syncytial virus, *BMC Microbiol.* **3**, 9.

Blandin, S., Moita, L.F., Kocher, T., Wilm, M., Kafatos, F.C., Levashina, E.A., Vriend, G., Lagueux, M. (2002) Reverse genetics in the mosquito *Anopheles gambiae*: targeted disruption of the Defensin gene, *EMBO Rep.* **3**, 852–856.

Boczkowski, D., Nair, S.K., Nam, J.H., Lyerly, H.K., Gilboa, E. (2000) Induction of tumor immunity and cytotoxic T lymphocyte responses using dendritic cells transfected with messenger RNA amplified from tumor cells, *Cancer Res.* **60**, 1028–1034.

Boyle, J.P., Wu, X.J., Shoemaker, C.B., Yoshino, T.P. (2003) Using RNA interference to manipulate endogenous gene expression in Schistosoma mansoni sporocysts, *Mol. Biochem. Parasitol.* **128**, 205–215.

Braun, K., Ehemann, V., Waldeck, W., Pipkorn, R., Corban-Wilhelm, H., Jenne, J., Gissmann, L., Debus, J. (2004) HPV18 E6 and E7 genes affect cell cycle, pRB and p53 of cervical tumor cells and represent prominent candidates for intervention by use peptide nucleic acids (PNAs), *Cancer Lett.* **209**, 37–49.

Breaker, R.R., Joyce, G.F. (1994) A DNA enzyme that cleaves RNA, *Chem. Biol.* **1**, 223–229.

Bridge, A.J., Pebernard, S., Ducraux, A., Nicoulaz, A.L., Iggo, R. (2003) Induction of an interferon response by RNAi vectors in mammalian cells, *Nat. Genet.* **34**, 263–264.

Burge, C.B., Tuschl, T., Sharp, P.A. (1999) Splicing of Precursors to mRNAs by the Spliceosomes, in: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.) *The RNA World II*, Cold Spring Harbor Laboratory Press, New York, pp. 525–560.

Butticaz, C., Ciuffi, A., Munoz, M., Thomas, J., Bridge, A., Pebernard, S., Iggo, R., Meylan, P.,

Telenti, A. (2003) Protection from HIV-1 infection of primary CD4 T cells by CCR5 silencing is effective for the full spectrum of CCR5 expression, *Antivir. Ther.* **8**, 373–377.

Cantin, E.M., Podsakoff, G., Willey, D.E., Openshaw, H. (1992) Antiviral effects of herpes simplex virus specific anti-sense nucleic acids, *Adv. Exp. Med. Biol.* **312**, 139–149.

Caplen, N.J., Zheng, Z., Falgout, B., Morgan, R.A. (2002) Inhibition of viral gene expression and replication in mosquito cells by siRNA-triggered RNA interference, *Mol. Ther.* **6**, 243–251.

Caplen, N.J., Taylor, J.P., Statham, V.S., Tanaka, F., Fire, A., Morgan, R.A. (2002) Rescue of polyglutamine-mediated cytotoxicity by double-stranded RNA-mediated RNA interference, *Hum. Mol. Genet.* **11**, 175–184.

Capodici, J., Kariko, K., Weissman, D. (2002) Inhibition of HIV-1 infection by small interfering RNA mediated RNA interference, *J. Immunol.* **169**, 5196–5201.

Cech, T.R., Uhlenbeck, O.C. (1994) Ribozymes. Hammerhead nailed down, *Nature* **372**, 39–40.

Chang, J., Taylor, J.M. (2003) Susceptibility of human hepatitis delta virus RNAs to small interfering RNA action, *J. Virol.* **77**, 9728–9731.

Chen, C.J., Banerjea, A.C., Harmison, G.G., Haglund, K., Schubert, M. (1992) Multitarget-ribozyme directed to cleave at up to nine highly conserved HIV-1 env RNA regions inhibits HIV-1 replication – potential effectiveness against most presently sequenced HIV-1 isolates, *Nucleic Acids Res.* **20**, 4581–4589.

Chernicky, C.L., Yi, L., Tan, H., Gan, S.U., Ilan, J. (2000) Treatment of human breast cancer cells with antisense RNA to the type I insulin-like growth factor receptor inhibits cell growth, suppresses tumorigenesis, alters the metastatic potential and prolongs survival *in vivo*, *Cancer Gene. Ther.* **7**, 384–395.

Chi, J.T., Chang, H.Y., Wang, N.N., Chang, D.S., Dunphy, N., Brown, P.O. (2003) Genomewide view of gene silencing by small interfering RNAs, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6343–6346.

Cobaleda, C., Sanchez-Garcia, I. (2000) *In vivo* inhibition by a site-specific catalytic RNA subunit of RNase P designed against the BCR-ABL oncogenic products: a novel approach for cancer treatment, *Blood* **95**, 731–737.

Coburn, G.A., Cullen, B.R. (2002) Potent and specific inhibition of human immunodeficiency virus type 1 replication by RNA interference, *J. Virol.* **76**, 9225–9231.

Cooper, D., Penny, R., Symonds, G., Carr, A., Gerlach, W., Sun, L.Q., Ely, J. (1999) A marker study of therapeutically transduced CD4+ peripheral blood lymphocytes in HIV discordant identical twins, *Hum. Gene Ther.* **10**, 1401–1421.

Cutrona, G., Carpaneto, E.M., Ponzanelli, A., Ulivi, M., Millo, E., Scarfi, S., Roncella, S., Benatti, U., Boffa, L.C., Ferrarini, M. (2003) Inhibition of the translocated c-myc in Burkitt's lymphoma by a PNA complementary to the E mu enhancer, *Cancer Res.* **63**, 6144–6148.

Davies, G., Watkins, G., Mason, M.D., Jiang, W.G. (2004) Targeting the HGF/SF receptor c-met using a hammerhead ribozyme transgene reduces in vitro invasion and migration in prostate cancer cells, *Prostate* **60**, 317–324.

Dector, M.A., Romero, P., Lopez, S., Arias, C.F. (2002) Rotavirus gene silencing by small interfering RNAs, *EMBO Rep.* **3**, 1175–1180.

Diebold, S.S., Kaisho, T., Hemmi, H., Akira, S., Reis e Sousa, C. (2004) Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA, *Science* **303**, 1529–1531.

Domingo, E., Holland, J.J. (1997) RNA virus mutations and fitness for survival, *Annu. Rev. Microbiol.* **51**, 151–178.

Dreyfus, D.H., Matczuk, A., Fuleihan, R. (2004) An RNA external guide sequence ribozyme targeting human interleukin-4 receptor alpha mRNA, *Int. Immunopharmacol.* **4**, 1015–1027.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells, *Nature* **411**, 494–498.

Ellington, A.D., Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands, *Nature* **346**, 818–822.

Emilsson, G.M., Breaker, R.R. (2002) Deoxyribozymes: new activities and new applications, *Cell. Mol. Life. Sci.* **59**, 596–607.

Filleur, S., Courtin, A., Ait-Si-Ali, S., Guglielmi, J., Merle, C., Harel-Bellan, A., Clezardin, P., Cabon, F. (2003) SiRNA-mediated inhibition of vascular endothelial growth factor severely limits tumor resistance to antiangiogenic thrombospondin-1 and

slows tumor vascularization and growth, *Cancer Res.* **63**, 3919–3922.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*, *Nature* **391**, 806–811.

Forster, A.C., Davies, C., Hutchins, C.J., Symons, R.H. (1990) Characterization of self-cleavage of viroid and virusoid RNAs, *Methods Enzymol.* **181**, 583–607.

Ge, Q., Filip, L., Bai, A., Nguyen, T., Eisen, H.N., Chen, J. (2004) Inhibition of influenza virus production in virus-infected mice by RNA interference, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8676–8681.

Ge, Q., McManus, M.T., Nguyen, T., Shen, C.H., Sharp, P.A., Eisen, H.N., Chen, J. (2003) RNA interference of influenza virus production by directly targeting mRNA for degradation and indirectly inhibiting all viral RNA transcription, *Proc. Nat. Acad. Sci. U.S.A.* **100**, 2718–2723.

Gervaix, A., Li, X., Kraus, G., Wong-Staal, F. (1997) Multigene antiviral vectors inhibit diverse human immunodeficiency virus type 1 clades, *J. Virol.* **71**, 3048–3053.

Giladi, H., Ketzinel-Gilad, M., Rivkin, L., Felig, Y., Nussbaum, O., Galun, E. (2003) Small interfering RNA inhibits hepatitis B virus replication in mice, *Mol. Ther.* **8**, 769–776.

Gitlin, L., Karelsky, S., Andino, R. (2002) Short interfering RNA confers intracellular antiviral immunity in human cells, *Nature* **418**, 430–434.

Grunweller, A., Wyszko, E., Bieber, B., Jahnel, R., Erdmann, V.A., Kurreck, J. (2003) Comparison of different antisense strategies in mammalian cells using locked nucleic acids, 2'-O-methyl RNA, phosphorothioates and small interfering RNA, *Nucleic Acids Res.* **31**, 3185–3193.

Hafner, M., Zawatzky, R., Hirtreiter, C., Buurman, W.A., Echtenacher, B., Hehlgans, T., Mannel, D.N. (2001) Antimetastatic effect of CpG DNA mediated by type I IFN, *Cancer Res.* **61**, 5523–5528.

Hamasaki, K., Nakao, K., Matsumoto, K., Ichikawa, T., Ishikawa, H., Eguchi, K. (2003) Short interfering RNA-directed inhibition of hepatitis B virus replication, *FEBS Lett.* **543**, 51–54.

Harborth, J., Elbashir, S.M., Vandenburgh, K., Manninga, H., Scaringe, S.A., Weber, K., Tuschl, T. (2003) Sequence, chemical, and structural variation of small interfering RNAs and short hairpin RNAs and the effect on mammalian gene silencing, *Antisense Nucleic Acid Drug Dev.* **13**, 83–105.

Heil, F., Hemmi, H., Hochrein, H., Ampenberger, F., Kirschning, C., Akira, S., Lipford, G., Wagner, H., Bauer, S. (2004) Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8, *Science* **303**, 1526–1529.

Heiser, A., Maurice, M.A., Yancey, D.R., Coleman, D.M., Dahm, P., Vieweg, J. (2001) Human dendritic cells transfected with renal tumor RNA stimulate polyclonal T-cell responses against antigens expressed by primary and metastatic tumors, *Cancer Res.* **61**, 3388–3393.

Hoa, N.T., Keene, K.M., Olson, K.E., Zheng, L. (2003) Characterization of RNA interference in an *Anopheles gambiae* cell line, *Insect Biochem. Mol. Biol.* **33**, 949–957.

Hotchkiss, G., Maijgren-Steffensson, C., Ahrlund-Richter, L. (2004) Efficacy and mode of action of hammerhead and hairpin ribozymes against various HIV-1 target sites, *Mol. Ther.* **10**, 172–180.

Hu, W.Y., Myers, C.P., Kilzer, J.M., Pfaff, S.L., Bushman, F.D. (2002) Inhibition of retroviral pathogenesis by RNA interference, *Curr. Biol.* **12**, 1301–1311.

Hwang, B., Han, K., Lee, S.W. (2003) Prevention of passively transferred experimental autoimmune myasthenia gravis by an *in vitro* selected RNA aptamers, *FEBS Lett.* **31**, 85–89.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., Linsley, P.S. (2003) Expression profiling reveals off-target gene regulation by RNAi, *Nat. Biotechnol.* **21**, 635–637.

Jacque, J.M., Triques, K., Stevenson, M. (2002) Modulation of HIV-1 replication by RNA interference, *Nature* **418**, 435–438.

Jia, Q., Sun, R. (2003) Inhibition of gammaherpesvirus replication by RNA interference, *J. Virol.* **77**, 3301–3306.

Jiang, M., Milner, J. (2002) Selective silencing of viralgene expression in HPV-positive human cervical carcinoma cells treated with siRNA, a primer of RNA interference, *Oncogene* **21**, 6041–6048.

Kapadia, S.B., Brideau-Andersen, A., Chisari, F.V. (2003) Interference of hepatitis C virus

RNA replication by short interfering RNAs, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2014–2018.

Kaushik, N., Pandey, V.N. (2002) PNA targeting the PBS and A-loop sequences of HIV-1 genome destabilizes packaged tRNA3(Lys) in the virions and inhibits HIV-1 replication, *Virology* **303**, 297–308.

Kijima, H., Yamazaki, H., Nakamura, M., Scanlon, K.J., Osamura, R.Y., Ueyama, Y. (2004) Ribozyme against mutant K-ras mRNA suppresses tumor growth of pancreatic cancer, *Int. J. Oncol.* **24**, 559–564.

Klein, C., Bock, C.T., Wedemeyer, H., Wustefeld, T., Locarnini, S., Dienes, H.P., Kubicka, S., Manns, M.P., Trautwein, C. (2003) Inhibition of hepatitis B virus replication in vivo by nucleoside analogues and siRNA, *Gastroenterology* **125**, 9–18.

Knop, A.E., Arndt, A.J., Raponi, M., Boyd, M.P., Ely, J.A., Symonds, G. (1999) Artificial capillary culture: expansion and retroviral transduction of CD4+ T-lymphocytes for clinical application, *Gene Ther.* **6**, 373–384.

Konishi, M., Wu, C.H., Wu, G.Y. (2003) Inhibition of HBV replication by siRNA in a stable HBV-producing cell line, *Hepatology* **38**, 842–850.

Koppelhus, U., Nielsen, P.E. (2003) Cellular delivery of peptide nucleic acid (PNA), *Adv. Drug Deliv. Rev.* **55**, 267–280.

Kossen, K., Vaish, N.K., Jadhav, V.R., Pasko, C., Wang, H., Jenison, R., McSwiggen, J.A., Polisky, B., Seiwert, S.D. (2004) High-throughput ribozyme-based assays for detection of viral nucleic acids, *Chem. Biol.* **11**, 807–815.

Kretschmer-Kazemi Far, R., Sczakiel, G. (2003) The activity of siRNA in mammalian cells is related to structural target accessibility: a comparison with antisense oligonucleotides, *Nucleic Acids Res.* **31**, 4417–4424.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena, *Cell* **31**, 147–157.

Kuwabara, T., Amontov, S.V., Warashina, M., Ohkawa, J., Taira, K. (1996) Characterization of several kinds of dimer minizyme: simultaneous cleavage at two sites in HIV-1 tat mRNA by dimer minizymes, *Nucleic Acids Res.* **24**, 2302–2310.

Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., Tuschl, T. (2003) New microRNAs from mouse and human, *RNA* **9**, 175–179.

Lai, J.C., Benimetskaya, L., Santella, R.M., Wang, Q., Miller, P.S., Stein, C.A. (2003) G3139 (oblimersen) may inhibit prostate cancer cell growth in a partially bis-CpG-dependent non-antisense manner, *Mol. Cancer Ther.* **2**, 1031–1043.

Lee, M.T., Coburn, G.A., McClure, M.O., Cullen, B.R. (2003) Inhibition of human immunodeficiency virus type 1 replication in primary macrophages by using Tat- or CCR5-specific small interfering RNAs expressed from a lentivirus vector, *J. Virol.* **77**, 11964–11972.

Lee, N.S., Dohjima, T., Bauer, G., Li, H., Li, M.J., Ehsani, A., Salvaterra, P., Rossi, J. (2002) Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells, *Nat. Biotechnol.* **20**, 500–505.

Levashina, E.A., Moita, L.F., Blandin, S., Vriend, G., Lagueux, M., Kafatos, F.C. (2001) Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*, *Cell* **104**, 709–718.

Li, H., Li, W.X., Ding, S.W. (2002) Induction and suppression of RNA silencing by an animal virus, *Science* **296**, 1319–1321.

Liu, X., Jiang, Q., Mansfield, S.G., Puttaraju, M., Zhang, Y., Zhou, W., Cohn, J.A., Garcia-Blanco, M.A., Mitchell, L.G., Engelhardt, J.F. (2002) Partial correction of endogenous DeltaF508 CFTR in human cystic fibrosis airway epithelia by spliceosome-mediated RNA trans-splicing, *Nat. Biotechnol.* **20**, 47–52.

Long, M.B., Jones, J.P. 3rd, Sullenger, B.A., Byun, J. (2003) Ribozyme-mediated revision of RNA and DNA, *J. Clin. Invest.* **112**, 312–338.

Lorger, M., Engstler, M., Homann, M., Goringer, H.U. (2003) Targeting the variable surface of African trypanosomes with variant surface glycoprotein-specific, serum-stable RNA aptamers, *Eukaryot. Cell* **2**, 84–94.

Maggon, K., Barik, S. (2004) New drugs and treatment for respiratory syncytial virus, *Rev. Med. Virol.* **14**, 149–168.

Mansfield, S.G., Clark, R.H., Puttaraju, M., Kole, J., Cohn, J.A., Mitchell, L.G., Garcia-Blanco, M.A. (2003) 5′ exon replacement and repair by spliceosome-mediated RNA trans-splicing, *RNA* **9**, 1290–1297.

Martinez, M.A., Gutierrez, A., Armand-Ugon, M., Blanco, J., Parera, M., Gomez, J., Clotet, B., Este, J.A. (2002) Suppression of

chemokine receptor expression by RNA interference allows for inhibition of HIV-1 replication, *AIDS* **16**, 2385–2390.

Martinez, L.A., Naguibneva, I., Lehrmann, H., Vervisch, A., Tchenio, T., Lozano, G., Harel-Bellan, A. (2002) Synthetic small inhibiting RNAs: efficient tools to inactivate oncogenic mutations and restore p53 pathways, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14849–14854.

McCaffrey, A.P., Nakai, H., Pandey, K., Huang, Z., Salazar, F.H., Xu, H., Wieland, S.F., Marion, P.L., Kay, M.A. (2003) Inhibition of hepatitis B virus in mice by RNA interference, *Nat. Biotechnol.* **21**, 639–644.

McCall, M.J., Hendry, P., Mir, A.A., Conaty, J., Brown, G., Lockett, T.J. (2000) Small, efficient hammerhead ribozymes, *Mol. Biotechnol.* **14**, 5–17.

McCown, M., Diamond, M.S., Pekosz, A. (2003) The utility of siRNA transcripts produced by RNA polymerase I in down regulating viral gene expression and replication of negative- and positive-strand RNA viruses, *Virology* **313**, 514–524.

Mercatante, D.R., Mohler, J.L., Kole, R. (2002) Cellular response to an antisense-mediated shift of Bcl-x pre-mRNA splicing and antineoplastic agents, *J. Biol. Chem.* **277**, 49374–49382.

Michienzi, A., Castanotto, D., Lee, N., Li, S., Zaia, J.A., Rossi, J.J. (2003) RNA-mediated inhibition of HIV in a gene therapy setting, *Ann. N. Y. Acad. Sci.* **1002**, 63–71.

Miyagishi, M., Taira, K. (2004) RNAi expression vectors in mammalian cells, *Methods Mol. Biol.* **252**, 483–491.

Miyagishi, M., Hayashi, M., Taira, K. (2003) Comparison of the suppressive effects of antisense oligonucleotides and siRNAs directed against the same targets in mammalian cells, *Antisense Nucleic Acid Drug Dev.* **13**, 1–7.

Nagaswamy, U., Larios-Sanz, M., Hury, J., Collins, S., Zhang, Z., Zhao, Q., Fox, G.E. (2002) NCIR: a database of non-canonical interactions in known RNA structures, *Nucleic Acids Res.* **30**, 395–397.

Nair, S.K., Heiser, A., Boczkowski, D., Majumdar, A., Naoe, M., Lebkowski, J.S., Vieweg, J., Gilboa, E. (2000) Induction of cytotoxic T cell responses and tumor immunity against unrelated tumors using telomerase reverse transcriptase RNA transfected dendritic cells, *Nat. Med.* **6**, 1011–1017.

Novina, C.D., Murray, M.F., Dykxhoorn, D.M., Beresford, P.J., Riess, J., Lee, S.K., Collman, R.G., Lieberman, J., Shankar, P., Sharp, P.A. (2002) siRNA directed inhibition of HIV-1 infection, *Nat. Med.* **8**, 681–686.

Nulf, C.J., Corey, D. (2004) Intracellular inhibition of hepatitis C virus (HCV) internal ribosomal entry site (IRES)-dependent translation by peptide nucleic acids (PNAs) and locked nucleic acids (LNAs), *Nucleic Acids Res.* **32**, 3792–3798.

Pan, W.H., Xin, P., Morrey, J.D., Clawson, G.A. (2004) A self-processing ribozyme cassette: utility against human papillomavirus 11 E6/E7 mrna and hepatitis B virus, *Mol. Ther.* **9**, 596–606.

Park, W.S., Miyano-Kurosaki, N., Hayafune, M., Nakajima, E., Matsuzaki, T., Shimada, F., Takaku, H. (2002) Prevention of HIV-1 infection in humanperipheral blood mononuclear cells by specific RNA interference, *Nucleic Acids Res.* **30**, 4830–4835.

Pass, R.F. (2001) Cytomegalovirus, in: Knipe, D.M., Howley, P.M. (Eds.) *Fields Virology*, Lippincott-William & Wilkins, Philadelphia, PA, pp. 2675–2706.

Paul, C.P., Good, P.D., Winer, I., Engelke, D.R. (2002) Effective expression of small interfering RNA in human cells, *Nat. Biotechnol.* **20**, 505–508.

Persengiev, S.P., Zhu, X., Green, M.R. (2004) Nonspecific, concentration-dependent stimulation and repression of mammalian gene expression by small interfering RNAs (siRNAs), *RNA* **10**, 12–18.

Phipps, K.M., Martinez, A., Lu, J., Heinz, B.A., Zhao, G. (2004) Small interfering RNA molecules as potential anti-human rhinovirus agents: in vitro potency, specificity, and mechanism, *Antiviral Res.* **61**, 49–55.

Poch, O., Sauvaget, I., Delarue, M., Tordo, N. (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements, *EMBO J.* **8**, 3867–3874.

Popescu, F.D. (2003) New asthma drugs acting on gene expression, *J. Cell. Mol. Med.* **7**, 475–486.

Puttaraju, M., Jamison, S.F., Mansfield, S.G., Garcia-Blanco, M.A., Mitchell, L.G. (1999) Spliceosome-mediated RNA trans-splicing as a tool for gene therapy, *Nat. Biotechnol.* **17**, 246–252.

Qin, X.F., An, D.S., Chen, I.S., Baltimore, D. (2003) Inhibiting HIV-1 infection in human

T cells by lentiviral-mediated delivery of small interfering RNA against CCR5, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 183–188.

Ramezani, A., Ding, S.F., Joshi, S. (1997) Inhibition of HIV-1 replication by retroviral vectors expressing monomeric and multimeric hammerhead ribozymes, *Gene Ther.* **4**, 861–867.

Reed, J.C., Cuddy, M., Haldar, S., Croce, C., Nowell, P., Makover, D., Bradley, K. (1990) BCL2-mediated tumorigenicity of a human T-lymphoid cell line: synergy with MYC and inhibition by BCL2 antisense, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3660–3664.

Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W.S., Khvorova, A. (2004) Rational siRNA design for RNA interference, *Nat. Biotechnol.* **22**, 326–330.

Rothenfusser, S., Tuma, E., Wagner, M., Endres, S., Hartmann, G. (2003) Recent advances in immunostimulatory CpG oligonucleotides, *Curr. Opin. Mol. Ther.* **5**, 98–106.

Rubinson, D.A., Dillon, C.P., Kwiatkowski, A.V., Sievers, C., Yang, L., Kopinja, J., Rooney, D.L., Ihrig, M.M., McManus, M.T., Gertler, F.B., Scott, M.L., Van Parijs, L. (2003) A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference, *Nat. Genet.* **33**, 401–406.

Santoro, S.W., Joyce, G.F., Sakthivel, K., Gramatikova, S., Barbas, C.F. 3rd (2000) RNA cleavage by a DNA enzyme with extended chemical functionality, *J. Am. Chem. Soc.* **122**, 2433–2439.

Sarver, N., Cantin, E.M., Chang, P.S., Zaia, J.A., Ladne, P.A., Stephens, D.A., Rossi, J.J. (1990) Ribozymes as potential anti-HIV-1 therapeutic agents, *Science* **247**, 1222–1225.

Scherr, M., Battmer, K., Winkler, T., Heidenreich, O., Ganser, A., Eder, M. (2003) Specific inhibition of bcr-abl gene expression by small interfering RNA, *Blood* **101**, 1566–1569.

Schmidt, C., Welz, R., Muller, S. (2000) RNA double cleavage by a hairpin-derived twin ribozyme, *Nucleic Acids Res.* **28**, 886–894.

Schwarz, D.S., Hutvagner, G., Haley, B., Zamore, P.D. (2002) Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways, *Mol. Cell* **10**, 537–548.

Scotlandi, K., Maini, C., Manara, M.C., Benini, S., Serra, M., Cerisano, V., Strammiello, R., Baldini, N., Lollini, P.-L., Nanni, P., Nicoletti, G., Picci, P. (2002) Effectiveness of insulin-like growth factor I receptor antisense strategy against Ewing's sarcoma cells, *Cancer Gene Ther.* **9**, 296–307.

Semizarov, D., Frost, L., Sarthy, A., Kroeger, P., Halbert, D.N., Fesik, S.W. (2003) Specificity of short interfering RNA determined through gene expression signatures, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6347–6352.

Sen, A., Steele, R., Ghosh, A.K., Basu, A., Ray, R., Ray, R.B. (2003) Inhibition of hepatitis C virus protein expression by RNA interference, *Virus Res.* **96**, 27–35.

Seo, M.Y., Abrignani, S., Houghton, M., Han, J.H. (2003) Small interfering RNA-mediated inhibition of hepatitis C virus replication in the human hepatoma cell line Huh-7, *J. Virol.* **77**, 810–812.

Shlomai, A., Shaul, Y. (2003) Inhibition of hepatitis B virus expression and replication by RNA interference, *Hepatology* **37**, 764–770.

Shtivelman, E., Lifshitz, B., Gale, R.P., Canaani, E. (1985) Fused transcript of abl and bcr genes in chronic myelogenous leukaemia, *Nature* **315**, 550–554.

Sledz, C.A., Holko, M., de Veer, M.J., Silverman, R.H., Williams, B.R. (2003) Activation of the interferon system by short-interfering RNAs, *Nat. Cell Biol.* **5**, 834–839.

Strobel, S.A., Doudna, J.A. (1997) RNA seeing double: close-packing of helices in RNA tertiary structure, *Trends Biochem. Sci.* **22**, 262–266.

Sullenger, B.A., Cech, T.R. (1994) Ribozyme-mediated repair of defective mRNA by targeted, trans-splicing, *Nature* **371**, 619–622.

Sumimoto, H., Miyagishi, M., Miyoshi, H., Yamagata, S., Shimizu, A., Taira, K., Kawakami, Y. (2004) Inhibition of growth and invasive ability of melanoma by inactivation of mutated BRAF with lentivirus-mediated RNA interference, *Oncogene* **23**(36) 6031–6039.

Sun, L.Q., Wang, L., Gerlach, W.L., Symonds, G. (1995) Target sequence-specific inhibition of HIV-1 replication by ribozymes directed to tat RNA, *Nucleic Acids Res.* **23**, 2909–2913.

Surabhi, R.M., Gaynor, R.B. (2002) RNA interference directed against viral and cellular targets inhibits human immunodeficiency virus type 1 replication, *J. Virol.* **76**, 12963–12973.

Suyama, E., Wadhwa, R., Kaur, K., Miyagishi, M., Kaul, S.C., Kawasaki, H., Taira, K. (2004) Identification of metastasis-related

genes in a mouse model using a library of randomized ribozymes, *J. Biol. Chem.* **279**(37), 38083–38086.

Talerico, M., Berget, S.M. (1994) Intron definition in splicing of small *Drosophila* introns, *Mol. Cell. Biol.* **14**, 3434–3445.

Tinoco Jr., I., Bustamante, C. (1999) How RNA folds, *J. Mol. Biol.* **293**, 271–281.

Tomar, R.S., Matta, H., Chaudhary, P.M. (2003) Use of adeno-associated viral vector for delivery of small interfering RNA, *Oncogene* **22**, 5712–5715.

Tompkins, S.M., Lo, C.Y., Tumpey, T.M., Epstein, S.L. (2004) Protection against lethal influenza virus challenge by RNA interference *in vivo*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8682–8686.

Trang, P., Kilani, A., Lee, J., Hsu, A., Liou, K., Kim, J., Nassi, A., Kim, K., Liu, F. (2002) RNase P ribozymes for the studies and treatment of human cytomegalovirus infections, *J. Clin. Virol.* **25**(Suppl. 2), S63–S74.

Tuerk, C., Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, *Science* **249**, 505–510.

Uhlenbeck, O.C. (1987) A small catalytic oligoribonucleotide, *Nature* **328**, 596–600.

Ullu, E., Djikeng, A., Shi, H., Tschudi, C. (2002) RNA interference: advances and questions, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **29**, 65–70.

Usman, N., Blatt, L.M. (2000) Nuclease-resistant synthetic ribozymes: developing a new class of therapeutics, *J. Clin. Invest.* **106**, 1197–1202.

Vacek, M., Sazani, P., Kole, R. (2003) Antisense-mediated redirection of mRNA splicing, *Cell. Mol. Life Sci.* **60**, 825–833.

Vater, A., Jarosch, F., Buchner, K., Klussmann, S. (2003) Short bioactive Spiegelmers to migraine-associated calcitonin gene-related peptide rapidly identified by a novel approach: tailored-SELEX, *Nucleic Acids Res.* **31**, e130.

Vickers, T.A., Koo, S., Bennett, C.F., Crooke, S.T., Dean, N.M., Baker, B.F. (2003) Efficient reduction of target RNAs by small interfering RNA and RNase H-dependent antisense agents. A comparative analysis, *J. Biol. Chem.* **278**, 7108–7118.

Wang, K., Zhang, Q.F., Wang, X.S., Xue, Y.W., Pang, D., Fu, S.B. (2004) Peptide nucleic acids arrest the growth of gastric cancer cells SGC7901, *Chin. Med. J. (Engl)*. **117**, 566–570.

Welch, P.J., Yei, S., Barber, J.R. (1998) Ribozyme gene therapy for hepatitis C virus infection, *Clin. Diagn. Virol.* **10**, 163–171.

Welch, P.J., Tritz, R., Yei, S., Leavitt, M., Yu, M., Barber, J. (1996) A potential therapeutic application of hairpin ribozymes: in vitro and in vivo studies of gene therapy for hepatitis C virus infection, *Gene Ther.* **3**, 994–1001.

Weng, D.E., Usman, N. (2001) Angiozyme: a novel angiogenesis inhibitor, *Curr. Oncol. Rep.* **3**, 141–146.

Wilson, J.A., Jayasena, S., Khvorova, A., Sabatinos, S., Rodrigue-Gervais, I.G., Arya, S., Sarangi, F., Harris-Brandts, M., Beaulieu, S., Richardson, C.D. (2003) RNA interference blocks gene expression and RNA synthesis from hepatitis C replicons propagated in human liver cells, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2783–2788.

Wlotzka, B., Leva, S., Eschgfaller, B., Burmeister, J., Kleinjung, F., Kaduk, C., Muhn, P., Hess-Stumpp, H., Klussmann, S. (2002) *In vivo* properties of an anti-GnRH Spiegelmer: an example of an oligonucleotide-based therapeutic substance class, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 8898–9902.

Wong-Staal, F., Poeschla, E.M., Looney, D.J. (1998) A controlled, phase 1 clinical trial to evaluate the safety and effects in HIV-1 infected humans of autologous lymphocytes transduced with a ribozyme that cleaves HIV-1 RNA, *Hum. Gene Ther.* **9**, 2407–2425.

Xia, H., Mao, Q., Paulson, H.L., Davidson, B.L. (2002) siRNA-mediated gene silencing *in vitro* and *in vivo*, *Nat. Biotechnol.* **20**, 1006–1010.

Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., Nemzer, S., Pinner, E., Walach, S., Bernstein, J., Savitsky, K., Rotman, G. (2003) Widespread occurrence of antisense transcription in the human genome, *Nat. Biotechnol.* **21**, 379–386.

Ying, C., De Clercq, E., Neyts, J. (2003) Selective inhibition of hepatitis B virus replication by RNA interference, *Biochem. Biophys. Res. Commun.* **309**, 482–484.

Yokota, T., Sakamoto, N., Enomoto, N., Tanabe, Y., Miyagishi, M., Maekawa, S., Yi, L., Kurosaki, M., Taira, K., Watanabe, M., Mizusawa, H. (2003) Inhibition of intracellular hepatitis C virus replication by synthetic and vector-derived small interfering RNAs, *EMBO Rep.* **4**, 1–7.

Yoshinari, K., Miyagishi, M., Taira, K. (2004) Effects on RNAi of the tight structure, sequence and position of the targeted region, *Nucleic Acids Res.* **32**, 691–699.

Zhou, C., Bahner, I.C., Larson, G.P., Zaia, J.A., Rossi, J.J., Kohn, E.B. (1994) Inhibition of HIV-1 in human T-lymphocytes by retrovirally transduced anti-tat and rev hammerhead ribozymes, *Gene* **149**, 33–39.

Zinnen, S.P., Domenico, K., Wilson, M., Dickinson, B.A., Beaudry, A., Mokler, V., Daniher, A.T., Burgin, A., Beigelman, L. (2002) Selection, design, and characterization of a new potentially therapeutic ribozyme, *RNA* **8**, 214–228.

Zou, H., Lee, J., Kilani, A.F., Kim, K., Trang, P., Kim, J., Liu, F. (2004) Engineered RNase P ribozymes increase their cleavage activities and efficacies in inhibiting viral gene expression in cells by enhancing the rate of cleavage and binding of the target mRNA, *J. Biol. Chem* **279**, 32063–32070.

Phospholipids

Dennis E. Vance
University of Alberta, Edmonton, Alberta, Canada

1 Structure of Phospholipids 279

2 Biosynthesis of Phospholipids 280
2.1 Pathways 280
2.2 Enzymes of Phosphatidylcholine Biosynthesis 282
2.2.1 Historical Background 282
2.2.2 Choline Transport and Oxidation 283
2.2.3 Choline Kinase 283
2.2.4 CTP:Phosphocholine Cytidylytransferase (CT) 283
2.2.5 CDP-choline: 1,2-diacylglycerol Cholinephosphotransferase 285
2.2.6 Phosphatidylethanolamine *N*-methyltransferase 285
2.2.7 Phosphatidylcholine in Bacteria 287
2.3 Enzymes of Phosphatidylethanolamine Biosynthesis 287
2.3.1 Historical Background 287
2.3.2 Enzymes of the CDP-ethanolamine Pathway 287
2.4 Phosphatidylserine Synthase and Decarboxylase 288
2.4.1 Historical Developments and Biosynthesis 288
2.4.2 Phosphatidylserine Decarboxylase 289
2.5 Bis(monoacylglycerol)phosphate, Phosphatidylglycerol, and Diphosphatidylglycerol 290
2.6 Phosphatidylinositol Biosynthesis and Phosphorylation 291
2.6.1 Historical Developments 291
2.6.2 CDP-diacylglycerol Synthase 292
2.6.3 Phosphatidylinositol Synthase 292
2.7 Remodeling of the Acyl Substituents of Phospholipids 293
2.8 Regulation of Phosphatidylcholine Biosynthesis 293
2.8.1 Transcriptional and Posttranscriptional Regulation of CTα 296
2.8.2 Transgenic and Gene-disrupted Murine Models of CTα 296
2.9 Regulation of Gene Expression in Yeast 297

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

3 Phospholipid Transport 298
3.1 Phospholipid Transfer Proteins 299
3.2 Vesicular and Contiguous Transfer of Phospholipids 299
3.3 Genetic Analysis of Phospholipid Transport in Yeast 300

4 Functional Aspects of Phospholipids 300
4.1 Sources of Second Messengers 300
4.2 Anchoring of Proteins to Membranes 301

5 Perspectives 302

Bibliography 302
Books and Reviews 302
Primary Literature 303

Keywords

Arachidonic Acid
A 20 carbon fatty acid with 4 double bonds. It is the fatty acid precursor of eicosanoids.

Diacylglycerol
An intermediate in phospholipid biosynthesis and an activator of protein kinase C.

Inositol-trisphosphate
A cyclohexane ring substituted with a hydroxyl group on each carbon and a phosphate on the hydroxyl groups of carbon number 1,4, and 5.

Phosphatidylcholine
Quantitatively the most important phospholipid found in eucaryotes.

Platelet-activating Factor
A potent compound that at a concentration of 10^{-11} M will cause platelets to aggregate and has many other biological activities.

Protein Kinase
An enzyme that phosphorylates a protein on the amino acids, serine, threonine, or tyrosine.

Second Messenger
A compound that is formed within a cell in response to a hormone or other agonists binding to a specific receptor on the cell surface.

Phospholipids are biological compounds that contain phosphorus and have both hydrophobic and hydrophilic moieties. They are ubiquitous components of all biological membranes. Phospholipids provide the basic structure and the permeability barriers of cellular membranes. The absolute requirement for phospholipids in life is underscored by the apparent lack in animals of gene mutations in phospholipid biosynthesis. Such mutations would result in severely defective enzyme activities and would probably be lethal. Phospholipids also act as sources for cellular second messengers such as diacylglycerol and inositol-trisphosphate. Phospholipids are made within each cell from precursors that include fatty acids, glycerol, CTP, ATP, and compounds that contain a free hydroxyl group such as ethanolamine, choline, glycerol, inositol, or serine. The phospholipids are made by biosynthetic enzymes and degraded by phospholipases. Many of the genes for these enzymes from *Escherichia coli*, yeast, mice, and humans have been cloned and characterized. Recent progress in the targeting of murine genes encoding several phospholipids by biosynthetic enzymes has provided insight into the function of these genes.

1 Structure of Phospholipids

The basic strucutre of all phospholipids is shown in Fig. 1. In mammals, the fatty acid attached to carbon 1 is usually saturated [palmitate (16 : 0) or stearate (18 : 0)]. In the abbreviation of fatty acids, the number to the left of the colon indicates the number of carbons and the number to the right indicates the number of double bonds. Unsaturated fatty acids [oleate (18 : 1), linoleate (18 : 2), arachidonate (20 : 4), docosahexaenoate (22 : 6)] are usually found on carbon 2. However, there are exceptions where saturated fatty acids are found on both carbons, or unsaturated fatty acids are found on both carbons. Small percentages (1 to 5%) of such phospholipids are found in most tissues. However, lung surfactant (which keeps lungs from collapsing when air is expelled) contains approximately 70% of its phosphatidylcholine (PC) as dipalmitoyl-phosphatidylcholine. The retina has as a major phospholipid species in its membrane, phosphatidylcholine that has docosahexaenoate (22 : 6) on both carbons. The function of this highly unsaturated phospholipid in the retina is unknown. The phospholipids in *E. coli* do not contain polyunsaturated fatty acids. Thus, usually a saturated fatty acid is on carbon 1 and an 18-carbon monounsaturated fatty acid (*cis*-vaccenic acid) on carbon 2.

Linked to phosphate on carbon 3 is one of the five small compounds (choline, ethanolamine, serine, glycerol, inositol) that have a hydroxyl functional group. The phospholipids are named "phosphatidyl" (the diacylglycerol phosphate moiety), followed by the name of the small compound such as choline (e.g. PC). A sixth phospholipid, diphosphatidylglycerol (DPG), also known as *cardiolipin*, has phosphatidylglycerol (PG) linked to the phosphatidyl moiety. There are other minor phospholipids, some of which will be mentioned later in this chapter.

In addition to the diacyl-phospholipids, there are two other classes in which the

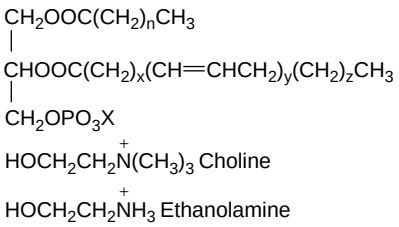

Fig. 1 General structure of a phospholipid. The *X* component can be one of five compounds that contains a hydroxyl group linked to the phosphate residue. The structures of two of these compounds is shown. Usually, there is a saturated fatty acid on carbon 1 of the glycerol moiety and an unsaturated fatty acid on carbon 2. In some cases, the substituent on carbon 1 can be a long-chain alcohol instead of a fatty acid.

acyl group on carbon 1 has been replaced by an ether linkage. If the substituent is a saturated alkyl ether, the phosphatidyl moiety is then referred to as *plasmanyl*. If the substituent is a vinyl (1,2-cis double bond) alkyl ether, the phosphatidyl moiety is referred to as *plasmenyl*. Plasmenylethanolamine is a major component of many cells and tissues such as the heart. The ether-containing phospholipids are absent from many but not all prokaryotes.

The phospholipid composition of tissues and cells is surprisingly consistent with PC and phosphatidylethanolamine (PE) being present at the highest concentrations (20 to 40% each). Phosphatidylserine (PS), phosphatidylinositol (PI), PG and DPG are each present at concentrations in the range of 10% of the total phospholipid. The subcellular membranes usually have minor differences in the phospholipid compositions with the exception of mitochondria. Phosphatidylglycerol and DPG are generally found only in this organelle along with the other phospholipids.

2 Biosynthesis of Phospholipids

2.1 Pathways

The major pathways for the biosynthesis of phospholipids in animals are shown in Fig. 2. In the first stage, fatty acids are esterified to glycerol-3-phosphate to yield phosphatidic acid (PA). The latter compound can react with cytidine triphosphate (CTP) to yield CDP-diacylglycerol (CDP-DG). In animal cells, CDP-DG reacts with inositol to form PI or with glycerol to form phosphatidylglycerol (PG). The PG then reacts with another molecule of CDP-DG to form DPG (cardiolipin). Alternatively, the phosphate of PA can be removed by the enzyme phosphatidic acid phosphohydrolase (PAP) to yield diacylglcyerol. Diacylglcyerol (DG) reacts with either CDP-choline to make PC or CDP-ethanolamine to make PE. Note that the formation of a phospholipid *de novo* always requires CTP.

Phospholipids can be modified in several ways. For example, PE (or PC) can react with serine to form PS that can be decarboxylated to PE (Fig. 2). This appears to be the only way in nature that ethanolamine is generated. The ethanolamine can then be used via the CDP-ethanolamine pathway to make PE. Hence, the CDP-ethanolamine pathway is sometimes referred to as a *salvage pathway*. In another modification, PE can be methylated to PC by three successive transfers of methyl groups from *S*-adenosylmethionine catalyzed by phosphatidylethanolamine *N*-methyltransferase (PEMT). This appears to be the major pathway by which the choline moiety of PC is made *de novo*, since

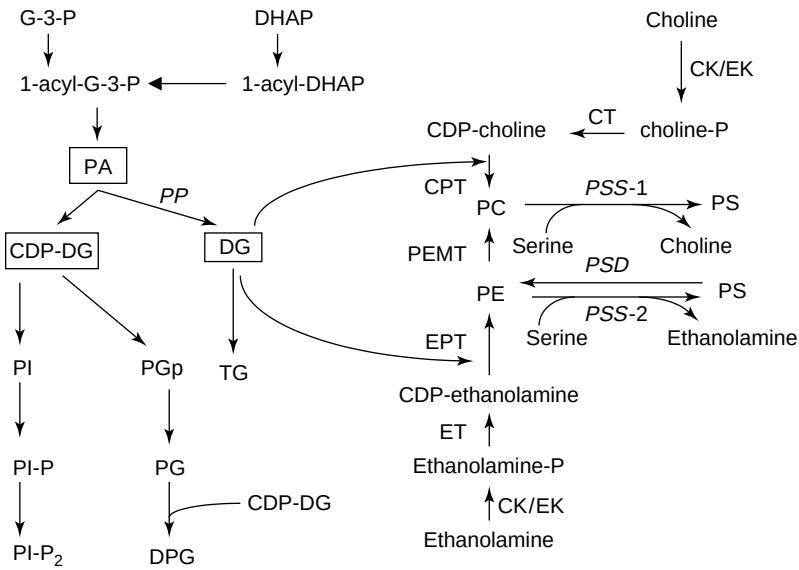

Fig. 2 Outline for the biosynthesis of the major phospholipids in human cells. The abbreviations are DHAP, dihydroxyacetone phosphate; G-3-P, glycerol-3-phosphate; PA, phosphatidic acid; DG, diacylglycerol; TG, triacylglycerol; CDP-DG, cytidine diphosphodiacylglycerol; PI, phosphatidylinositol; PG, phosphatidylglycerol; PGP, phosphatidylglycerol phosphate; DPG, diphosphatidylglycerol; PP, phosphatidic acid phosphohydrolase; PE, phosphatidylethanolamine; PC phosphatidylcholine; PEMT, phosphatidylethanolamine *N*-methyltransferase; CT, CTP: phosphocholine cytidylyltransferase; CDP-choline, cytidine diphosphocholine; choline-P, phosphocholine; CDP-ethanolamine, cytidine diphosphoethanolamine; ethanolamine P, phosphoethanolamine; PS, phosphatidylserine; CK/EK, choline kinase/ethanolamine kinase; CPT, CDP-choline:1,2-diacylglycerol cholinephosphotransferase; EPT, CDP-ethanolamine:1,2-diacylglycerol ethanolaminephosphotransferase; ET, CTP: phosphoethanolamine cytidylyltransferase; PSD, phosphatidylserine decarboxylase; PSS, phosphatidylserine synthase.

PC can be degraded by phospholipases to yield choline, which can then be reutilized to make PC by the CDP-choline pathway. In animal tissues, the methylation pathway is abundant in the liver but present only at much lower levels in other tissues. The methylation pathway is absent in most prokaryotes, but is present in yeast. Also, though not shown in Fig. 2, the fatty acid moieties of phospholipids can be removed (deacylated) and replaced by another fatty acid (reacylated). Finally, the phosphocholine moiety of PC can be transferred to ceramide (*N*-acyl-sphingosine) to form sphingomyelin (SM). This particular phospholipid is usually found enriched in the plasma membrane of cells.

The enzymes of phospholipid biosynthesis are found in the cytosol, endoplasmic reticulum (ER), mitochondria-associated membrane (MAM, a section of the ER that sediments with mitochondria after homogenization of cells/tissues), Golgi, peroxisomes, mitochondria, and the nucleus. Only a few membrane-bound phospholipid biosynthetic enzymes from

animal sources have been purified to homogeneity. This means that purification requires solubilization of the enzymes from the membranes by detergents such as deoxycholate, Triton X-100, or others. Unfortunately, to keep the enzyme soluble it is necessary to have detergents present during the purification steps and this greatly complicates the procedures.

2.2 Enzymes of Phosphatidylcholine Biosynthesis

2.2.1 Historical Background

PC was first described by Gobley in 1847 as a component of egg yolk and named "lecithin" after the Greek equivalent for egg yolk (*lekithos*). In the 1860s, Diakonow and Strecker demonstrated that lecithin contained two fatty acids linked to glycerol and that choline was attached to the third hydroxyl by a phosphodiester linkage. The first significant advance in understanding PC biosynthesis occurred in 1932 with the discovery by Charles Best that animals had a dietary requirement for choline. In the 1950s, the CDP-choline pathway for PC biosynthesis (Fig. 3) was described by Eugene Kennedy and coworkers. A key observation was that CTP, rather than ATP, was the activating nucleotide for PC biosynthesis. CTP is required not only for PC biosynthesis but also for the *de novo* synthesis of all phospholipids (prokaryotic and eukaryotic, excluding PA, which can

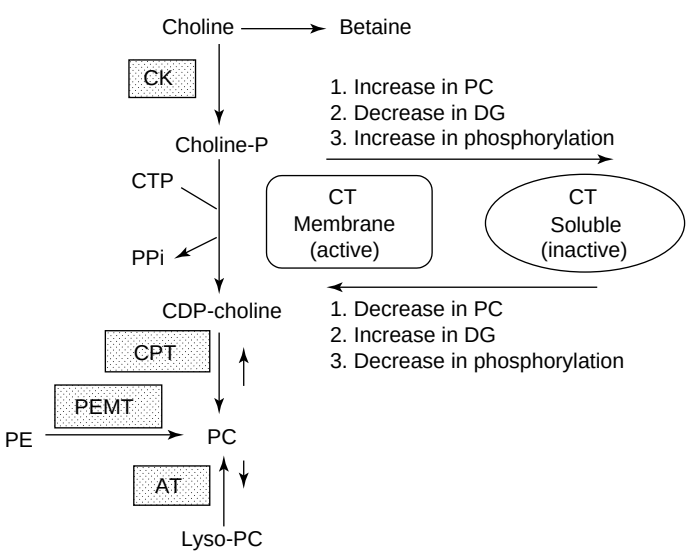

Fig. 3 Regulation of phosphatidylcholine (PC) biosynthesis via the CDP-choline pathway by modulation of the binding of CTP:phosphocholine cytidylyltransferase (CT) to membranes. Three different modes of regulation of CT activity are shown. Thus, a decrease in the concentration of PC will promote the translocation and activation of CT. An increase in PC will cause CT to be released from the membranes into a soluble, inactive form. The abbreviations are CK, choline kinase; CPT, CDP-choline: 1,2-diacylglycerol cholinephosphotransferase; PEMT, phosphatidylethanolamine *N*-methyltransferase; AT, lyso-PC acyltransferase; PE, phosphatidylethanolamine; DG, diacylglycerol.

be considered to be an intermediate in glycerolipid biosynthesis).

An alternative pathway for PC biosynthesis, of quantitative significance only in liver, is the conversion of PE to PC via PE methylation (Fig. 2). The first observation of this pathway was in 1941 when Stetten fed [^{15}N]ethanolamine to rats and isolated [^{15}N]choline. Two decades later, Bremer and Greenberg detected a microsomal enzyme that converted PE to PC via transfer of methyl groups from *S*-adenosylmethionine.

2.2.2 Choline Transport and Oxidation

Choline is not made *de novo* in animal cells except by methylation of PE to PC and subsequent hydrolysis of the choline moiety. Therefore, choline must be imported from extracellular sources. There are two distinct transport mechanisms for choline: a high affinity (K_m or $K_t < 5\ \mu M$), Na-dependent transporter and a lower affinity ($K_t > 30\ \mu M$), Na-independent transporter. Several cDNAs encoding proteins that show high affinity transport of choline have been reported. A human cDNA is predicted to have 13 transmembrane-spanning domains.

Once choline is inside the cell, its normal fate is rapid phosphorylation by choline kinase (Fig. 3). In neurons, choline is also converted to the neurotransmitter, acetylcholine. Choline is also oxidized to betaine [$^{-}$00C-CH$_2$-N^{+} (CH$_3$)$_3$] in the liver and kidney. In the liver, betaine is an important donor of methyl groups for methionine biosynthesis and the one carbon pool. Betaine is produced in mitochondria into which choline is transported by a specific transporter on the inner membrane. Next, choline is oxidized to betaine aldehyde by choline dehydrogenase on the inner leaflet of the inner mitochondrial membrane. The conversion to betaine is catalyzed by betaine-aldehyde dehydrogenase located in the mitochondrial matrix. In the renal medulla, betaine accumulates as an osmolyte (a small organic solute that accumulates in response to hypertonicity without adverse affect to the cell or organism). Hypertonicity of the renal medulla is important for the kidney's ability to concentrate urine.

2.2.3 Choline Kinase

The enzyme was first demonstrated in yeast extracts by J. Wittenberg and A. Kornberg (more famous for his contributions to DNA replication) in 1953. The enzyme was purified by K. Ishidate in 1984 from rat kidney and shown also to phosphorylate ethanolamine. This kinase is now referred to as *choline/ethanolamine kinase ß*. The cDNA for a rat liver choline/ethanolamine kinase encoded an enzyme that is now referred to as *choline/ethanolamine kinase* α1. Northern analyses indicate that the mRNA for choline/ethanolamine kinase α1 is most abundant in testis. Choline/ethanolamine kinase α2 appears to be a splice variant of choline/ethanolamine kinase α1. The choline/ethanolamine kinase α and ß genes have been characterized. The length of the gene was 40 kb for the choline/ethanolamine kinase α gene, whereas the ß gene was only 3.5 kb in length.

Choline is not only required in the diet of animals but also in the medium of animal cells in culture. Choline is essential because of the cell's requirement for PC to grow and divide.

2.2.4 CTP:Phosphocholine Cytidylytransferase (CT)

This enzyme was first described by Kennedy and Weiss in 1955. Over three

decades later, cytidylytransferase (CT) was finally purified to homogeneity. The CT gene was cloned from *Saccharomyces cerevisiae* by S. Yamashita in 1987 by complementation of a yeast mutant defective in CT activity. The cDNA of rat liver CT was subsequently cloned by R.B. Cornell and coworkers in 1990. CT is a homodimer in soluble extracts of rat liver and is also found on membranes. In most cells, CT is thought to exist in an inactive reservoir in its soluble form and to be activated when associated with membranes (Fig. 3).

Two genes encode different forms of CT: α and ß. The *CTα* gene spans approximately 26 kb. Exon 1 is untranslated, exon 2 encodes the translation start site and a nuclear localization signal, exons 4–7 encode the catalytic domain, exon 8 codes for the alpha helical membrane–binding domain, and exon 9 encodes a C-terminal phosphorylation domain.

The *CTß* gene is located on the X chromosome and encodes three isoforms, CTß1, CTß2, and CTß3 (Fig. 4), presumably derived by mRNA splicing. Presently CTß1 has only been found in humans. All isoforms differ from CTα at the amino terminus, lack the nuclear localization signal, and are found in the cytoplasm of animal cells. The primary sequences of CTß1and CTß2 are identical except at the carboxyl terminal. CTß1 lacks most of the phosphorylation domain that is present in CTß2. There are significant differences between the sequences of the phosphorylation domains of CTα and CTß.

CT has classically been considered to be a cytoplasmic enzyme since its activity is found in the cytosol and on microsomal membranes in cellular homogenates. However, Kent and coworkers demonstrated that CT was found in the nuclear matrix and associated with the nuclear membrane. There is intriguing evidence that CTα migrates into the cytoplasm during the G1 phase of the cell cycle, a time when PC biosynthesis is activated. Thus, the role the nuclear localization signal of CT plays in cellular PC biosynthesis remains an intriguing question.

The lipid-binding domain and the phosphorylated domains are involved in the regulation of CT activity. These domains

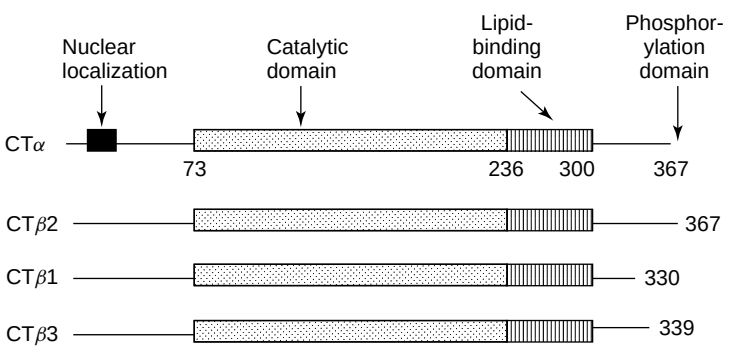

Fig. 4 Domain structures of CTP:phosphocholine cytidylyltransferase (CT) α, ß1, ß2, and ß3. CTα contains a nuclear localization signal, an N-terminal catalytic domain, an amphipathic helical (lipid-binding) domain, and a C-terminal phosphorylation domain. The CT ß forms lack the nuclear localization signal but contain catalytic and amphipathic helical domains. CT ß1 is missing the phosphorylation domain, whereas CTß2 and CTß3 have a phosphorylation domain that is different from that of CTα.

of CTα have been deleted by either proteolysis with chymotrypsin or by construction of CTα truncation mutants. CTα cDNAs that were truncated in the region of residue 314 (Fig. 4) lacked the phosphorylation segment, and CT truncated at residues 236, 231, or 228 lacked both the phosphorylation and lipid-binding domains. When the lipid-binding and phosphorylation domains were deleted, CT was a soluble, active enzyme that did not bind to membranes. Thus, the lipid-binding domain is regulatory for the binding to membranes and the activation of CT. The binding of phospholipids to CT appears to activate the enzyme by decreasing the apparent K_m value for CTP.

CT activity is modulated by phosphorylation. Experiments with CT truncation mutants have demonstrated that the phosphorylation domain is not required for lipid-binding or CT activity. Exactly what role phosphorylation of CT plays in a physiologically relevant system remains to be demonstrated.

2.2.5 CDP-choline: 1,2-diacylglycerol Cholinephosphotransferase

This enzyme was also discovered by Kennedy and coworkers and is considered to be located on the ER but is also found on the Golgi, MAM, and nuclear membranes. Even though the enzyme has been known for more than four decades, despite intense efforts in many laboratories, the cholinephosphotransferase has never been purified. The difficulty is that the enzyme is an intrinsic membrane-bound protein that requires detergents for solubilization. Moreover, the detergents complicate purification procedures commonly used, such as gel filtration, because the protein binds to micelles that are hard to separate on the basis of molecular size. The purification of membrane-bound enzymes has been described as "masochistic enzymology".

Yeast genetics and molecular biology have, however, allowed for the cholinephosphotransferase to be cloned. The products of two genes, *CPT1* and *EPT1*, each account for 50% of the cholinephosphotransferase activity in yeast extracts. By the use of null mutations in these two genes, it has been established that *CPT1* is responsible for 95% of the PC made and the *EPT1* gene product accounts for 5%. The *EPT1* gene product utilizes both CDP-choline and CDP-ethanolamine, whereas *CPT1* catalyzes only reactions with CDP-choline.

More recently, a human choline/ethanolamine-phophotransferase cDNA (hCE PT1) was cloned and expressed, and an additional human cDNA was cloned that encoded a CDP-choline-specific enzyme (hCPT1) with 60% sequence identity to hCEPT1.

Cholinephosphotransferase acts at a branch point in the metabolism of DG that can also be converted to PE, triacylglycerol (TG) or PA (Fig. 2). Most studies indicate that there is an excess of cholinephosphotransferase in cells; hence, the amount of active enzyme does not limit PC biosynthesis. However, it is clear that the *in vivo* activity of cholinephosphotransferase is regulated by substrate supply. The supply of CDP-choline is regulated by the activity of CT. The supply of DG in the liver seems to be controlled by the supply of fatty acids. The excess DG that is not utilized for PC or PE biosynthesis is stored in the liver as TG.

2.2.6 Phosphatidylethanolamine *N*-methyltransferase

All nucleated cells contain PC and the CDP-choline pathway. Thus, it was not

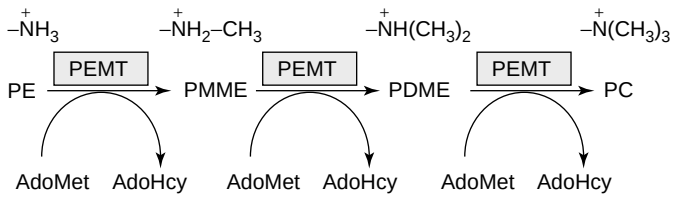

Fig. 5 Reactions catalyzed by PEMT, phosphatidylethanolamine *N*-methyltransferse; PE, phosphatidylethanolamine; AdoMet, *S*-adenosylmethionine; AdoHcy, *S*-adenosylhomocysteine; PMME, phosphatidylmonomethylethanolamine; PDME, phosphatidyldimethylethanolamine; and PC, phosphatidylcholine.

obvious why the pathway for PE methylation (Fig. 5) survived during evolution. Nor was it obvious why PEMT activity is mostly found in the liver, whereas 2% or less of the hepatic PEMT activity is found in other tissues of the body.

PEMT was purified from rat liver microsomes although it is an intrinsic membrane protein. The sequencing of the amino terminal enabled the cloning of the cDNA for PEMT. Preparation of an antibody to the deduced sequence of the carboxyl terminal peptide permitted subcellular localization of the enzyme. The major activity for PEMT is found on the ER, but the antibody only recognizes a protein that is exclusively localized to MAM. This isoenzyme of PEMT is referred to as PEMT2 and the activity on the ER is called PEMT1. Both PEMTs catalyze all three transmethylation reactions that convert PE to PC (Fig. 5).

A mouse was generated in which the *Pemt* gene was disrupted and there was no PEMT activity. The *Pemt*$^{-/-}$ mice lived and bred normally and there was a 50% increase in CT activity in their livers. Since the mice retained the CDP-choline pathway, the lack of an obvious phenotype was not surprising. However, when the mice were fed a choline-deficient diet for 3 days, which attenuates PC synthesis via the CDP-choline pathway, the *Pemt*$^{-/-}$ mice exhibited severe liver failure. *Pemt*$^{+/+}$ mice fed a choline-deficient diet were normal with no obvious liver pathology. Thus, it seems that the PEMT pathway has survived in evolution to provide PC at times when the CDP-choline pathway is less active such as might occur during starvation. Moreover, pregnant rats and suckling mothers can also have choline reserves depleted; hence, the PEMT pathway might provide an evolutionary advantage in this respect. The structurally related compound, dimethylethanolamine [$H0CH_2$-CH_2-N^+ $H(CH_3)_2$] would not substitute for choline in the *Pemt*$^{-/-}$ mice even though it was converted to phosphatidyldimethylethanolamine. Thus, it seems that the third methyl group on the phospholipid has a critical function in mice.

The human gene-encoding PEMT has been cloned and characterized. Whereas only one mRNA transcript has been identified in mice, the human liver has three separate mRNAs that differ only at the 5′ end, in a noncoding region of the transcript. Thus, the three transcripts encode the same protein. The function of separate PEMT mRNAs is going to be difficult to study in humans.

Yeast also has both the PE methylation pathway and the CDP-choline pathway. In

yeast, two enzymes are used for the conversion of PE to PC. The methylation of PE to phosphatidylmonomethylethanolamine is catalyzed by the *PEM1/CHO2* gene product, whereas the subsequent two methylations are catalyzed by the *PEM2/OPI3* gene product. Deletion of both *PEM1* and *PEM2* genes is lethal unless the yeast is supplied with choline. Yeast normally grows in the absence of choline and depends on the PEMT pathway. Thus, the CDP-choline pathway and the PE methylation pathway can compensate for each other in yeast.

2.2.7 Phosphatidylcholine in Bacteria

About 90% of bacteria do not contain PC but *Rhodobacter sphaeroides* make PC by methylation of PE. Interestingly, this enzyme (PC) is soluble and has virtually no homology to PEMT or the yeast enzymes. Also, in one bacterium, a novel choline-dependent pathway was recently discovered in *Sinorhizobium meliloti*, in which choline reacts with CDP-diacylglycerol, catalyzed by PC synthase, to form PC. A number of symbiotic bacteria and pathogenic bacteria also produce PC via the same PC synthase pathway.

2.3 Enzymes of Phosphatidylethanolamine Biosynthesis

2.3.1 Historical Background

PE was first alluded to in a book published by Thudichum in 1884. He described "kephalin" as a nitrogen- and phosphorus-containing lipid that was different from lecithin. In 1913, Renall and Baumann independently isolated ethanolamine from kephalin. In 1930, Rudy and Page isolated the first pure preparation of PE. The structure of PE was established in 1952 by Baer and colleagues.

The biosynthesis of PE in eukaryotes can occur via four pathways (Fig. 6). The route via CDP-ethanolamine constitutes *de novo* synthesis of PE. The other pathways arise as a result of the modification of a preexisting phospholipid. The CDP-ethanolamine pathway was first described by Kennedy and Weiss in 1956. The decarboxylation of PS to yield PE (Fig. 6) was shown in 1960 to occur in animal cells. PS decarboxylation is the only route for PE biosynthesis in *E. coli*. The PE generated by this pathway can react with serine to generate PS and ethanolamine (Fig. 6). This appears to be one mechanism by which ethanolamine is made in cells. The other involves degradation of sphingosine. The ethanolamine generated by either pathway can be utilized for PE biosynthesis via the CDP-ethanolamine pathway. No one has ever been able to show the decarboxylation of serine to ethanolamine in animal cells. Such a reaction was shown to occur in a plant, *Arabidopsis thaliana*. PE can also be formed by reacylation of lyso-PE or reaction of ethanolamine with PS (Fig. 6).

2.3.2 Enzymes of the CDP-ethanolamine Pathway

As mentioned in Sect. 2.2.3, the phosphorylation of ethanolamine in liver can be catalyzed by choline/ethanolamine kinase (Fig. 6). The cDNA encoding an ethanolamine kinase was cloned from *Drosophila*. These scientists had not planned on cloning this cDNA since their approach was to determine the gene responsible for the easily shocked (*eas*) phenotype in this insect. These mutant flies display transient paralysis following a brief mechanical shock. In the *eas* mutant, a 2-bp deletion caused the formation of a premature stop codon in the ethanolamine kinase gene. Analysis of the phospholipids

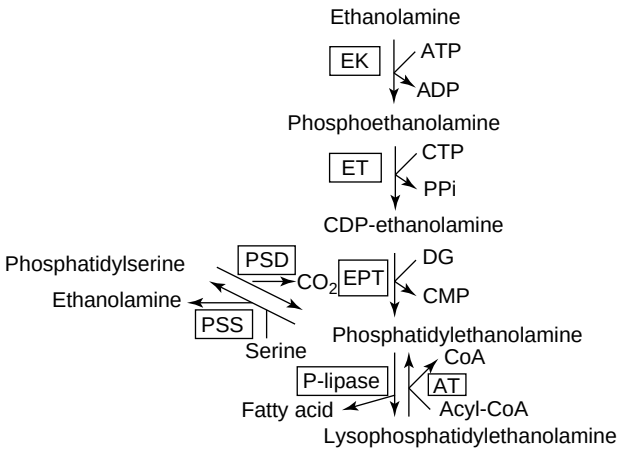

Fig. 6 Pathway for biosynthesis of phosphatidylethanolamine and phosphatidylserine in animal cells. The abbreviations are EK, ethanolamine(choline) kinase; ET, CTP: phosphoethanolamine cytidylyltransferase; EPT, CDP-ethanolamine: 1,2-diacylglycerol ethanolaminephosphotransferase; PSS, phosphatidylserine synthase; PSD, phosphatidylserine decarboxylase; P-lipase, phospholipase A_2; AT, acyl-CoA: lyso-phosphatidylethanolamine acyltransferase; DG, diacylglycerol.

showed a decrease in PE from 59% of the total phospholipid in the wild type to 56% in *eas*. Whether this minor change mediates the paralysis is not known. The difference may reflect a major change in PE content in a particular tissue or subcellular membrane. More recently, the gene that encodes a yeast ethanolamine kinase and a human cDNA for ethanolamine-specific kinase were cloned and expressed.

The second step in the CDP-ethanolamine pathway is catalyzed by CTP:phosphoethanolamine cytidylyltransferase. The enzyme is distinct from CT and is not activated by lipids. Although the phosphoethanolamine cytidylyltransferase is recovered in cytosol from cell extracts, much of the enzyme has been localized to rough ER of rat liver by immunoelectron microscopy. Unlike CTα, there is no report of the phosphoethanolamine cytidylyltransferase in the nucleus.

CDP-ethanolamine: 1,2-diacylglycerol ethanolaminephosphotransferase is an integral membrane protein found on the ER, Golgi, and MAM. The enzyme shows a distinct preference for DG species that contain 1-palmitoyl-2-docosahexaenoyl (22 : 6) fatty acids. In hepatocytes in culture, nearly 50% of PE made via the ethanolaminephosphotransferase reaction belongs to this species. The purpose of this extraordinary selectivity is unknown. The bovine hepatic enzyme was purified, and exhibited both ethanolamine- and cholinephosphotransferase activity.

2.4 Phosphatidylserine Synthase and Decarboxylase

2.4.1 Historical Developments and Biosynthesis

PS accounts for 5 to 15% of the phospholipids in eukaryotic cells. The lower

concentration of PS compared to PC and PE is probably the reason PS was not discovered as a separate component of "kephaline" (originally identified to be only PE in 1930) until 1941 by Folch. The correct structure was proposed by Folch in 1948 and confirmed by chemical synthesis in 1955 by Baer and Maurukas. PS is a required cofactor for protein kinase C and is required for initiation of the blood-clotting cascade. In the plasma membrane of cells, PS is normally located on the inner monolayer. During apoptosis, exposure of PS on the cell surface (outer monolayer) leads to recognition and removal of these cells by macrophages.

PS is made in prokaryotes, in some plants and in yeast via the CDP-diacylglycerol pathway. This route does not exist in animals. Instead, PS is made by a base-exchange reaction catalyzed by PS synthase, first described by Hübscher in 1959, in which the head group of PC or PE is exchanged for serine.

Chinese hamster ovary (CHO) mutants were generated that were auxotrophic for PS and demonstrated that these cells have two PS synthases. PS synthase 1 utilizes PC and serine as substrates, whereas PS synthase 2 utilizes only serine and PE. The two PS synthases, when coupled with PS decarboxylase, yield PS at the expense of PC and generate both choline and ethanolamine, which could be recycled into the biosynthesis of PC and PE. As a result, PS and PE can both be generated without a decline in the amount of PC.

PS synthase 1

$$\text{PC} + \text{serine} \longrightarrow \text{PS} + \text{choline}$$

PS decarboxylase

$$\text{PS} \longrightarrow \text{PE} + \text{C0}_2$$

PS synthase 2

$$\text{PE} + \text{serine} \longrightarrow \text{PS} + \text{ethanolamine}$$

The sum of the reactions is

$$\text{PC} + \text{two serines} \longrightarrow \text{PS} + \text{choline} + \text{ethanolamine} + \text{CO}_2$$

A CHO mutant defective in PS synthase 1 was used to clone by complementation of the cDNA for this enzyme. The deduced amino acid sequence for murine PS synthase 1 was >90% identical to the CHO enzyme. The cDNA for PS synthase 2 from CHO cells was cloned and shown to be 32% identical in amino acid sequence to PS synthase 1. Immunoblot analysis indicated that both of the murine PS synthases are mainly localized to MAM. The source of the substantial PS synthase activity in the rough and smooth ER remains unknown and may possibly be a third PS synthase activity. The mRNAs encoding PS synthases 1 and 2 were found in all murine tissues examined, but PS synthase 2 was enriched in testis and kidney.

The murine gene for PS synthase 1 has been cloned and characterized. This is an important step toward the generation of mice with a disrupted gene for PS synthase 1. The gene for PS synthase 2 has been disrupted and the mice are viable.

In yeast and in *E. coli*, PS is made by a different reaction than in mammals. In these organisms, CDP-diacylglycerol condenses with serine to yield cytidine mono phosphate (CMP) plus PS.

2.4.2 Phosphatidylserine Decarboxylase

PS decarboxylase is found in both prokaryotes and the mitochondria of eukaryotes. The enzyme activity was first described by Kanfer and Kennedy in 1964. The enzyme has not been purified from a

eukaryotic source but the gene was cloned and expressed from CHO cells and yeast. The yeast gene (*PSD1*) encodes a protein that is localized to mitochondria. However, when *PSD1* was disrupted in yeast, 5% PS decarboxylase activity remained and the yeast continued to grow. Subsequently, a second gene, *PSD2*, was isolated. When both *PSD1* and *PSD2* were disrupted, the yeast became ethanolamine auxotrophs. The PSD2 protein has been localized to the vacuolar and Golgi compartments. The function of *PSD2* is not known other than that it can supply enough PS decarboxylase to allow growth of yeast in the absence of *PSD1*. The rate of PS decarboxylation is determined by the rate of PS transport into mitochondria.

2.5 Bis(monoacylglycerol)phosphate, Phosphatidylglycerol, and Diphosphatidylglycerol

Diphosphatidylglycerol (DPG), commonly known as cardiolipin, was discovered in 1942 in beef heart by Pangborn. The correct structure was proposed in 1956–1957 and confirmed by chemical synthesis in 1965–1966 by de Haas and van Deenen. Phosphatidylglycerol was first isolated in 1958 from algae by Benson and Mauro. The structure was confirmed by Haverkate and van Deenen in 1964–1965. The third lipid in this class, bis(monoacylglycerol)phosphate was recovered from pig lung by Body and Gray in 1967. The stereochemistry differs from PG and DPG since bis(monoacylglycerol)phosphate contains *sn*-(monoacyl)glycerol-1-phospho-*sn*-1′-(monoacyl)-glycerol rather than a *sn*-glycerol-3-phospho linkage.

These three lipids are widely distributed in animals, plants, and microorganisms. In animals, DPG is found in highest concentration in the cardiac muscle (9–15% of phospholipid), hence the name cardiolipin, and is exclusively found in the mitochondria. PG is generally present at a concentration of less than 1% of total cellular phospholipids, except in the lung, where it comprises 2 to 5% of the phospholipid. In pulmonary surfactant and alveolar type II cells, PG is 7 to 11% of the total lipid phosphorus. Bis(monoacylglycerol)phosphate comprises less than 1% of total phospholipids in animal tissues, except in alveolar (lung) macrophages, where it is 14 to 18% of the total phospholipid.

The biosynthesis of PG was elucidated by Kennedy and coworkers in 1963 (Fig. 2). For DPG biosynthesis, PA is transferred from CDP-DG to PG to yield DPG. DPG synthesis in *E. coli* differs and involves the condensation of two molecules of PG.

Understanding the biosynthesis of bis(monoacylglycerol)phosphate has been a particular challenge because the carbon linked to the phosphate residue is the sn-1 rather than the sn-3 configuration. An intermediate in the biosynthesis of bis(monoacylglycerol)phosphate is 1-acyl-lyso-PG, also known as *lysobis-PA*. Recent studies have shown that the inner membranes of late endosomes are enriched in lysobis-PA and that these membranes play an important role in the sorting of insulin growth factor receptor 2 and the mannose-6-phosphate receptor. Moreover, lysobis-PA cross-reacts with antibodies produced in patients with an antiphospholipid syndrome. Possibly, some of the pathological defects in this disease could arise from disruption of endosomal traffic.

PG can be made in mitochondria and microsomes from various animal cells and, except for lung, appears to be primarily

converted to DPG. DPG is biosynthesized exclusively on the matrix side of the mitochondrial inner membrane and is found only in this organelle. There is evidence that the rate-limiting step in DPG biosynthesis is the conversion of PA into CDP-DG. Consistent with this idea, the levels of CTP have been shown to regulate DPG biosynthesis in cardiac myoblasts.

Using techniques developed by Raetz and coworkers, M. Nishijima and coworkers isolated a temperature-sensitive mutant in PG-P synthase of CHO cells. The mutant had 1% of wild-type CHO PG-P synthase activity at 40 °C and a temperature-sensitive defect in PG and DPG biosynthesis. This mutant was used to show that DPG is required for the NADH-ubiquinone reductase (complex I) activity of the respiratory chain.

In yeast, DPG synthesis has been genetically disrupted. The yeast grows at temperatures between 16 and 30 °C without DPG but fails to grow at 37 °C on fermentable carbon sources such as glucose, even though intact mitochondria are, therefore, not required for ATP synthesis. Thus, mitochondria must have some necessary function in yeast survival other than generating energy.

The fatty acyl content of phospholipids can also impact on mitochondrial function. Incubation of cardiomyocytes with palmitic acid increased the palmitic acid content of PA and PG, and decreased DPG levels in mitochondria with a concomitant release of cytochrome c, leading to apoptosis.

Phosphatidylglycerolphosphate synthase has been purified from *E. coli* and the cDNA cloned. The phosphate group is cleaved from phosphatidylglycerolphosphate by a phosphatase. Genetic manipulations have demonstrated that the gene-encoding phosphatidylglycerophosphate synthase is essential for *E. Coli*. This is in contrast to PS decarboxylase, which is apparently not essential in *E. coli*. That is, these organisms seem to grow without significant amounts of the PS decarboxylase product, PE.

Cardiolipin (DPG) synthase from *E. coli* has also been purified and the cDNA cloned. This enzyme is also not an essential enzyme. An alternative pathway or enzyme seems to provide the small amount of cardiolipin that is required for growth of *E. coli*.

2.6 Phosphatidylinositol Biosynthesis and Phosphorylation

2.6.1 Historical Developments

A major outcome of PA is conversion to DG that is metabolized to PC, PE, and TG (Fig. 2). Alternatively, PA can react with CTP to form CDP-DG that is utilized for the biosynthesis of the inositol phospholipids, PG and diphosphatidylglycerol (Fig. 2).

Inositol is a cyclohexane derivative in which all 6 carbons are substituted with hydroxyl groups. The most common isoform is *myo*-inositol, but other less-abundant inositols with different structures also occur. The first report of an inositol-containing lipid was in 1930 from *Mycobacteria*, which is ironic since inositol lipids are rarely found in bacteria. Brain is the richest source of these lipids, as first discovered by Folch and Wooley in 1942. In 1949, Folch described PI phosphate (PI-P), which was later found to include PI and PI bisphosphate (PI-P_2). The chemical structures of PI, PI-P, and PI-P_2 were determined by Ballou and coworkers between 1959 and 1961. PI (1.7μmol/g liver) constitutes around 10% of the phospholipids in a cell

or tissue. PI-P and PI-P_2 are present at much lower concentrations (1–3% of PI). Agranoff et al. published the first experiments in 1958 on the incorporation of [^{3}H]inositol into PI. Subsequently, Paulus and Kennedy showed that CTP was the preferred nucleotide donor.

2.6.2 CDP-diacylglycerol Synthase

Regulation of the conversion of PA to CDP-DG is not well understood. The enzyme, CDP-DG synthase, is largely microsomal, but is also found in the mitochondrial inner membrane.

A cDNA-encoding CDP-DG synthase 1 was cloned from *Drosophila*. This isoform is specifically located in photoreceptor cells of *Drosophila*. Mutations in this isoform lead to a defect in PI-P_2 biosynthesis. As a result, mutant photoreceptor cells show severe defects in their phospholipase C-mediated signal transduction that can be rescued by reintroduction of the CDP-DG synthase cDNA.

cDNAs encoding human and murine CDP-DG synthases 1 and 2 have recently been cloned. CDP-DG synthase 2 is expressed during embryogenesis in the central nervous system, whereas CDP-DG synthase 1 has a high level of expression in the adult retina.

Curiously, in *S. cerevisiae*, CDP-DG synthase activity is found in microsomes and the mitochondrial inner membrane, even though only one gene encodes this activity. Since only a single mRNA species was found, there may not be alternative splicing of the yeast gene. The yeast CDP-DG synthase gene is essential for cell viability as well as germination of spores.

2.6.3 Phosphatidylinositol Synthase

Three potential sources for cellular inositol are diet, *de novo* biosynthesis, and recycling of inositol. Biosynthesis of inositol from glucose occurs in the brain and testes, and in other tissues to a lesser extent. The rate-limiting step appears to be the synthesis of inositol-3-phosphate from glucose-6-phosphate. Inositol-3-phosphate is hydrolyzed to inositol by a phosphatase.

PI synthase was purifed from human placenta. When the cDNAs encoding either CDP-DG synthase 1 or phosphatidylinositol synthase, or both, were overexpressed in COS 7 cells, there was no change in the rate of PI biosynthesis, indicating that the level of these enzymes was not limiting for PI biosynthesis. Disruption of the PI synthase gene in yeast is lethal, indicating that PI is essential.

PI-4,5-P_2 is an important source of the second messengers, inositol-1,4,5-P_3 and DG. The biosynthesis of PI-P_2 begins with the reaction of CDP-DG and inositol to form PI. The phosphorylation of PI on the 4 position appears to be catalyzed by several different isoenzymes. One has been purified to homogeneity with a molecular weight of 55 000 and is specific for PI on the 4 position of the inositol moiety. A second kinase with a molecular mass of approximately 200 kDa has also been partially purified. The role of the different isoenzymes of PI-4 kinase remains to be elucidated.

The next enzyme in the sequence is PI-4P kinase. There appear to be three distinct forms of this enzyme in bovine brain, and one with a molecular mass of 53 kDa has been purified to homogeneity from brain and from human red cells. A different form has also been purified from erythrocytes. A major difficulty with the purification of these PI kinases from cells is that they are not abundant enzymes in extracts.

Another derivative of PI, PI-3P, appears to be involved in a signaling pathway that is distinct from the metabolism of

PI-4,5-P_2. The purifed bovine brain PI-3 kinase is composed of an 85 and a 110-kDa subunit. The 85-kDa subunit lacks PI-3 kinase activity and appears to mediate the coupling of the 110-kDa subunit with protein tyrosine kinases that phosphorylate PI-3 kinase. The 110-kDa subunit has been cloned and shows PI-3 kinase activity when associated with the 85-kDa subunit. The 110-kDa subunit has been shown to be encoded by the *VPS*34 gene in yeast and is essential for protein sorting to the lysosome-like vacuole in yeast. Exactly how PI-3P is involved in protein sorting remains to be determined.

2.7 Remodeling of the Acyl Substituents of Phospholipids

Phospholipids are made *de novo* with the fatty acid compositions present in the precursors DG and CDP-DG. Once the phospholipid is made, the fatty acid substitutents can be remodeled via deacylation–reacylation reactions (Fig. 7). Remodeling can occur on either the sn-1 or the sn-2 positions of the glycerolipid. For example, a major molecular species formed from the conversion of PE to PC is 16 : 0–22 : 6-PC. This species of PC has a half-life of less than 6 h and appears not to be significantly degraded but rather converted to other molecular species, particularly those with 18 : 0 on the sn-1 position and 20 : 4, 18 : 2, or 22 : 6 on the sn-2 position. Other studies have suggested that the main products of *de novo* PC and PE biosynthesis are 16 : 0–18 : 2, 16 : 0–18 : 1, 16 : 0–22 : 6, and 18 : 1–18 : 2. The major remodeled product is 18 : 0–20 : 4 for both PC and PE. Why 18 : 0–20 : 4-PC and -PE are made by this circuitous route, rather than directly, is not known.

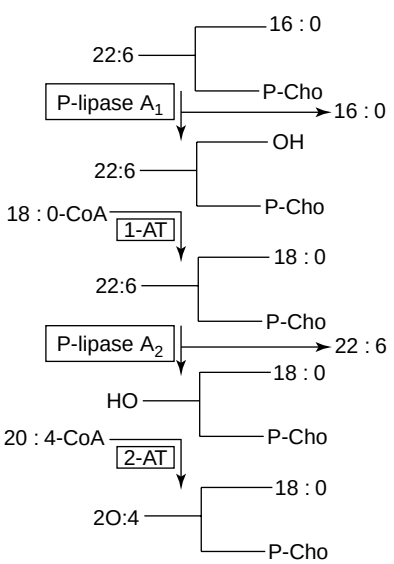

Fig. 7 One pathway by which phospholid molecular species are remodeled to change the fatty acid composition. P-lipase, phospholipase; AT, acyl-CoA: lyso-phosphatidylcholine acyltransferase; cho, choline.

2.8 Regulation of Phosphatidylcholine Biosynthesis

Of all the phospholipids, the regulation of PC biosynthesis is best understood. Generally, the activity of CTP:phosphocholine cytidylyltransferase (CT in Fig. 3) is rate-limiting and regulated. Considerable evidence has demonstrated that the rate of the CT reaction can limit PC biosynthesis. The first evidence in favor of this conclusion was the measurement of pool sizes of the aqueous precursors by Sundler and Åkesson (in rat liver choline = 0.23 mM, phosphocholine = 1.3 mM, CDP-choline = 0.03 mM). The concentration of phosphocholine is approximately 40-fold higher than

CDP-choline. The accumulation of phosphocholine is consistent with a bottleneck in the pathway at the reaction catalyzed by CT. Pulse-chase experiments with ^{3}H-choline demonstrate the bottleneck more vividly. After incubating cells with labeled choline for a short period, for example, 15 min, the cells are then "chased" (incubated with unlabeled choline for various durations) and the incorporation of radioactivity into PC and its precursors is measured. At the end of the pulse, approximately 90% of the radioactivity is in phosphocholine (Fig. 8). Over the next few hours, the radioactivity disappears from phosphocholine and appears quantitatively in PC. There is only a small amount of label in CDP-choline, consistent with the rapid formation of PC from CDP-choline.

Cytidylyltransferase (CT) exists as a dimer of identical subunits in the cytosol of cell homogenates where it is inactive. CT is activated by translocation of the enzyme to membranes (primarily the ER and nuclear membrane in liver) as shown in Fig. 3. The binding to the membranes appears to be via an amphipathic helix (the helix is composed of amino acids that results in a hydrophobic and hydrophilic side). Deletion of this amphipathic helix by construction of carboxyl truncation mutants of CT resulted in a protein that was active but would no longer bind to membrane lipids. The binding to the lipids in the ER activates the enzyme, at least in part, by lowering the apparent K_m for CTP. Binding of CT to membranes begins by electrostatic adsorption, followed by hydrophobic interactions that involve intercalation of the protein into the nonpolar core of the membrane. When insertion of CT into the membrane lipids is blocked by using viscous gel–phase lipids, CT binds electrostatically to the membrane, but is not activated. Four properties of membranes promote CT insertion: (1) interfacial packing defects as might occur when lipids with small head groups such as DG are in the membrane; (2) low lateral surface pressure (loose packing) as observed in highly curved compared to planar bilayers; (3) acyl chain disorder that can be caused by oxidation of the fatty acyl chains; and (4) curvature strain that would occur when

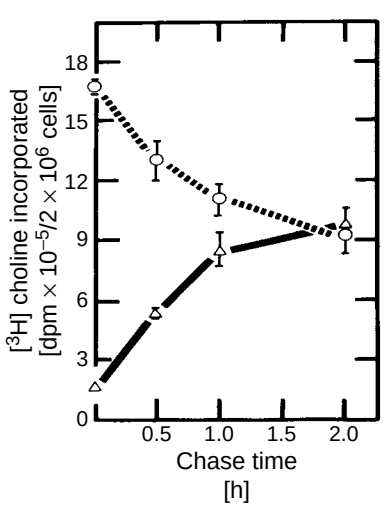

Fig. 8 Incorporation of [^{3}H-*methyl*]choline into phosphocholine and PC as a function of time. Hepatocytes from rat liver were incubated with labeled choline for 30 min. Subsequently, the cells were washed thoroughly and incubated (chased) for various times with unlabeled choline. The disappearance of radioactivity from phosphocholine (dashed line) and its appearance in PC (solid line) are shown.

membranes are enriched in hexagonal phase-preferring lipids such as PE and DG. Synthesis of PC would reverse these properties of membranes and form a more stable bilayer.

Understanding the regulation of CT activity is complicated by the observation that the enzyme is highly phosphorylated in the cytosol (potentially 16 different serine residues) and much less phosphorylated when bound to the ER. The phosphate residues are located in the carboxyl domain (residues 315 to 367) of CTα. The function of the phosphorylation of CT is being actively investigated at the present time. Experiments with mutagenesis (serine to alanine) of CT or construction of truncated forms of CT demonstrated that the phosphorylation domain of CT is not required for lipid-binding or CT activity. Recent evidence has shown that CT appears to bind to the membrane in response to a change in the lipid composition and is subsequently dephosphorylated. Thus, a change in the lipid composition of a membrane may be of primary importance compared to phosphorylation in regulating CT translocation. However, there is other evidence in studies with macrophages that dephosphorylation of CT may regulate how active CT is on the membranes.

Regulation of a metabolic pathway by product inhibition has been commonly observed. In livers or hepatocytes derived from choline-deficient rats, the rate of PC biosynthesis was inhibited by approximately 70% compared to choline-supplemented rats and there was a corresponding increased binding of CT to cellular membranes. The CT appeared to recognize a need for increased PC biosynthesis and was sitting poised on the membrane prepared for catalysis. However, because of a lack of substrate (phosphocholine), the amount of PC made was insufficient. When choline was supplied to hepatocytes under this condition, there was a nearly positive correlation between the increase in the level of PC in these cells and the release of CT into the cytosol. Similar correlations were observed when the level of PC was increased by providing methionine for enhanced conversion of PE to PC, or by providing lyso-PC, which is taken into hepatocytes and acylated to PC. This was the first study in which a mechanism for regulation of PC in cell cultures could be directly related to a physiologically relevant animal model.

An elegant feedback regulation of CT has been shown in the yeast *S. cerevisiae*. SEC14p is a phospholipid transfer protein that when assayed *in vitro* prefers PI and PC and is an essential gene product. SEC14p inhibited the CDP-choline pathway when PC was bound to SEC14p. In contrast, when PI was bound to SEC14p, there was minimal inhibition of CT. Thus, in yeast under conditions where PC is abundant, there is a feedback inhibition of CT and the CDP-choline pathway.

CTα has classically been considered to be a cytoplasmic enzyme since its activity is found in the cytosol and on microsomal membranes in cellular homogenates. However, immunofluorescence studies have demonstrated that CTα in CHO cells and many other cell lines is predominantly nuclear. In other studies, it was clear that in liver and primary hepatocytes, CT was predominantly cytoplasmic and was also present in the nucleus. Some of the CT in the cytoplasm is probably CTß, which does not have a nuclear localization signal. There is also evidence for the shuttling of CTα from the nucleus into the cytoplasm during active periods of PC biosynthesis.

CTP has also been implicated as being regulatory in animal systems and yeast. Overexpression of CTP synthetase in yeast

stimulated the biosynthesis of PC via the CDP-choline pathway.

2.8.1 Transcriptional and Posttranscriptional Regulation of CTα

Most studies on CT activity and PC biosynthesis have not indicated regulation at the level of gene expression. The ability of a cell to activate the soluble form of CT would normally satisfy the cell's requirement for PC. Nevertheless, some control over the expression of the genes encoding CT must occur. The proximal promoter of the *CTα* gene has numerous potential regulatory elements. Subsequent studies showed that Sp1, the first mammalian transcription factor purified and cloned, had an important role in regulating the expression of the *CTα* gene. The related nuclear factor, Sp3, could also activate CTα transcription. DNase protection assays indicated several elements in the proximal promoter-bound unidentified nuclear factors. The yeast one hybrid system was utilized to clone the cDNA for one of these factors, and the transcription enhancer factor 4 (TEF4) was identified as a regulator of CTα transcription. More recently, Ets has also been identified as an activating transcription factor. These initial studies were based on experiments involving transfections with various cDNA constructs. The first experiments to search for transcriptional regulation in a cell physiology–relevant system were on CTα expression in a murine fibroblast cell line as a function of the cell cycle. During the G0 to G1 phase of the cell cycle, there is an increase in PC biosynthesis, but there was no enhanced transcription of the *CTα* gene. Instead, increased transcription occurred during the S-phase of the cell cycle, possibly to increase the amount of CTα in preparation for mitosis. More recent studies have shown that the increased transcription is due to enhanced binding of Sp1 to the proximal promoter.

Sterol response element binding proteins (SREBPs) play a critical role in regulating the expression of genes involved in fatty acid and cholesterol metabolism. Thus, researchers in several laboratories explored whether SREBPs might alter the expression of the *CTα* gene. It appears that SREBP can activate CT transcription by binding to an element on the promoter, but the major activation appears to be indirect via the synthesis of fatty acids that will cause CT translocation to the membrane.

Obviously, there is much to be done to elucidate the factors and DNA elements involved in transcriptional regulation of the *CTα* gene. The transcription of the *CTß* gene appears to be activated in neuronal cells when exposed to nerve growth factor. However, the identity of the factor(s) involved remains to be determined.

The level of CTα mRNA can also be regulated by alterations in mRNA stability. When a macrophage cell line was depleted of the colony-stimulating factor and then repleted, there was a fourfold induction of mRNA for CTα. The stability of CTα mRNA increased after the addition of the colony-stimulating factor. An increase in CTα mRNA in fetal lung type II cells has also been ascribed to enhanced mRNA stability.

Finally, enhanced turnover of CTα via the ubiquitin–proteosome pathway appears to be the mechanism by which tumor necrosis factor decreases the level of CTα in alveolar type II cells.

2.8.2 Transgenic and Gene-disrupted Murine Models of CTα

To determine if enhanced PC biosynthesis would protect macrophages from

excess cholesterol-induced toxicity, genetically modified mice have been generated. A truncated version of CTα lacking the phosphorylation domain was expressed specifically in macrophages of mice under control of the scavenger receptor promoter. These cells were protected from cholesterol-induced toxicity. In another approach, CTα expression was eliminated in macrophages using the Cre-lox method for selective disruption of a gene in specific cells. The lack of the *CTα* gene and hence decreased PC biosynthesis caused enhanced sensitivity to cholesterol loading. In the absence of cholesterol loading, the macrophages without CTα appeared normal, possibly due to increased expression of CTß2. Further studies on tissue-specific knockouts of CTα should provide further information on the function of this enzyme.

2.9 Regulation of Gene Expression in Yeast

The pathways for the biosynthesis of phospholipids in yeast were largely elucidated by Lester and coworkers in the late 1960s (Fig. 9). These pathways are similar to those found in other eukaryotes except that PS in yeast is made via a pathway similar to that found in *E. coli*, where CDP-DG reacts with serine to yield PS and CMP.

Considerable interest in yeast as a model system has developed over the past two decades. Reasons for choosing *S. cerevisiae* include a large knowledge base in classical genetics, knowledge of the sequence of the genome, the ease of making mutant strains, and the ability to grow large amounts of yeast. Whereas understanding the regulation of expression of phospholipid biosynthetic enzymes in animal cells is still in its infancy, considerable progress has been made in the yeast system. When yeast cells are grown in the presence of choline and inositol, the expression of the enzymes involved in the conversion of PA and glucose-6-P to PI, PC, and PE is depressed (Fig. 9).

Both positive and negative regulatory factors are involved in the regulation of expression of phospholipid biosynthetic enzymes in yeast. The *INO2* and *INO4* genes encode transcription factors that are required for the expression of inositol-1-P synthase (*INO1*). *In vitro* transcribed and translated proteins derived from *INO2* and *INO4* form a heterodimer that binds a specific DNA fragment of the *INO1* gene referred to as UAS_{INO}. Ino4p (the protein encoded by *INO4*) and Ino2p exhibit basic helix–loop–helix domains. The Ino2p-Ino4p heterodimer binds to UAS_{INO} of the *INO1* promoter that contains two copies of a binding site (CANNTG) for basic helix-loop-helix–containing proteins.

The *OPI1* gene encodes a protein that is a negative regulatory factor for phospholipid biosynthesis. Opi1p contains a leucine zipper, a motif implicated in protein–DNA interactions, and transcriptional control. *Opi1* mutants exhibit a twofold increase in the constitutive expression of inositol-1-P synthase and other enzymes involved in PI, PC, and PE biosynthesis. The mechanism by which Opi1p mediates its negative regulatory role is unknown. Opi1p does not interact directly with UAS_{INO} or with Ino2p or Ino4p. Phosphorylation of Opi1 by protein kinase C may be involved.

Recent experiments have identified other proteins that interact with Ino4p, indicating that there is still much to learn about transcriptional regulation of phospholipid biosynthetic genes in yeast. How the regulatory genes (*INO2, INO4, OPI1*)

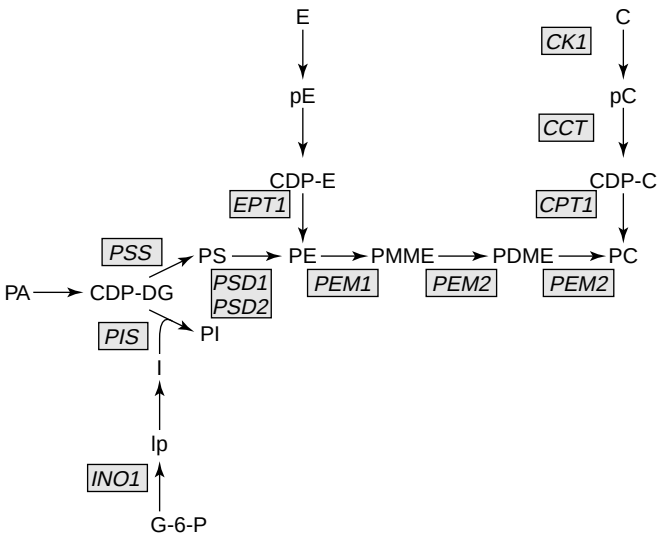

Fig. 9 Pathways for phospholipid biosynthesis in yeast and designation of the genes (italics in boxes) encoding the enzymes that catalyze the reactions. The abbreviations are E, ethanolamine; pE, phosphoethanolamine; CDP-E, CDP-ethanolamine; C, choline; pC, phosphocholine; CDP-C, CDP-choline; PE, phosphatidylethanolamine; PMME, phosphatidylmonomethylethanolamine; PDME, phosphatidyldimethylethanolamine; PC, phosphatidylcholine; PS, phosphatidylserine; PA, phosphatidic acid; CDP-DG, CDP-diacylglycerol; PI, phosphatidylinositol; I, inositol; Ip, inositol phosphate; G-6-P, glucose-6-phosphate. The genes encode the following enzymes: *INO1*, I-1-P synthase; *PIS*, PI synthase; *PSS* (also known as *CHO1*), PS synthase; *EPT1*, CDP-E:1,2-diacylglycerol ethanolaminephosphotransferase; *PEM1* (*CHO2*), PE methyltransferase; *PEM2* (*OPI3*), phospholipid methyltransferase; *CK1*, choline kinase; *CCT*, CTP:phosphocholine cytidylyltransferase (abbreviated as CT elsewhere in this chapter); *CPT1*, CDP-C: 1,2-diacylglycerol cholinephosphotransferase; *PSD1* and *PSD2*, PS decarboxylase.

are themselves regulated is just beginning to be studied.

3 Phospholipid Transport

The bulk of cellular phospholipids are made on the ER. How these lipids are directed to other cellular membranes is an important question of current interest. For example, PC is made on the ER and not on the mitochondria. However, approximately 30% of the phospholipids in mitochondria are PC. How does PC move to the mitochondria and what regulates this process?

Three modes are currently being considered for transport of phospholipids intracellularly: phospholipid transfer proteins, vesicular transport, and membrane continuity between adjacent organelles.

3.1 Phospholipid Transfer Proteins

These are low molecular weight proteins (14–18 kDa) that were first discovered in 1968 by Wirtz and Zilversmit. These proteins bind a single molecule of phospholipid and catalyze an exchange of phospholipids between donor and acceptor membranes. There are three categories of these exchange proteins: (1) those that are specific for PC; (2) those that prefer PI but also have transfer activity with PC; and (3) nonspecific transfer proteins that transfer most phospholipids and cholesterol. Several of these proteins have been purified, sequenced, and cloned. Considerable information is, therefore, available about the properties and *in vitro* activity of the transfer proteins. What has been lacking is evidence that the proteins actually function in cells as phospholipid transfer proteins.

The first success on understanding the possible function of the phospholipid-transfer proteins occurred several years ago when Dowhan and colleagues successfully inactivated the PI/PC transfer protein gene in yeast and demonstrated that the gene product was an essential protein. The authors correctly noted that this did not prove that the protein functioned in the transfer of PI or PC in the organism. Shortly thereafter, Bankaitis, Dowhan, and colleagues demonstrated that the *PI/PC* gene in yeast was identical to the previously identified *SEC14* gene. *Sec14* mutants of yeast are defective in secretion, fail to grow, and accumulate large amounts of Golgi membranes. Further research demonstrated that suppresor mutants of *sec14* (normal growth and function are recovered), which still had no PI/PC transfer protein, were defective in the biosynthesis of PC via the CDP-choline pathway. Thus, the PI/PC protein was somehow involved in regulation of the PI/PC ratio of Golgi membranes, and a decrease in PC biosynthesis achieved the same result. When PC is bound to the PI/PC transfer protein, it inhibits CT of the CDP-choline pathway. When PI is bound, there is no effect on CT activity. Thus, in the absence of the PI/PC transfer protein in *sec14* mutants, the biosynthesis of PC is unregulated so that PC accumulates in the Golgi and somehow causes Golgi dysfunction. Hence, at least in the case of *sec 14*, the transfer activity of the protein does not appear to be involved directly in lipid transfer, but rather in the binding of PC or PI to this protein, and the ligand that is bound determines the effect on CT activity. Further studies are required to elucidate the function of other phospholipid transfer proteins.

3.2 Vesicular and Contiguous Transfer of Phospholipids

A very attractive mechanism for movement of phospholipids from ER and Golgi (which also has the capacity for phospholipid biosynthesis) to other membranes is by vesicular transport. It is well known that this is how proteins are moved within cells to the plasma membrane or for secretion. The vesicles have phospholipid bilayers and fuse with target membranes. Therefore, it seems certain that some component of phospholipids in membranes such as the plasma membrane is delivered via this vesicular transport. Whether vesicular transport is the only source for plasma membrane phospholipids remains uncertain.

How mitochondria receive phospholipids and how PE, specifically made in mitochondria by the decarboxylation of

PS, is delivered to other organelles are questions of current interest. Voelker has used the organelle-specific decarboxylation of PS as a tool to study the mechanism of transport of PS, made on the ER, to the mitochondria. His studies have shown that the process does not require cytosol (source of phospholipid transfer protein) or PS biosynthesis, requires ATP, and may involve a membrane continuity between the ER and mitochondria. In agreement with the idea that there may be some type of noncovalent bridge between ER and mitochondria, J. Vance has isolated a MAM that is enriched in PS synthase compared to ER, has a protein composition similar to, but different from ER, and has low activity of a classical ER marker, nicotinamide adenine diphosphate (NADPH) cytochrome c reductase. If membrane continuity between the ER and mitochondria could be demonstrated, this might also be an explanation for much of phospholipid transport between the ER and other intracellular membranes such as the nuclear membrane and lysosomes. Then, the question of how the phospholipid composition of organelles is regulated can be addressed.

3.3
Genetic Analysis of Phospholipid Transport in Yeast

Selection of yeast mutants that require ethanolamine for growth has provided novel information about the transport of PS from the ER to mitochondria or Golgi. In the transport of PS to the Golgi, two mutant genes have been identified. One encodes a PI-4-kinase and the second gene encodes a protein that is closely related to the PI transfer protein encoded by the *SEC14* gene (Sect. 3.1). The process of PS transfer seems complex and it will be a challenge to unravel how these and other proteins are involved in the transfer of PS to the Golgi and mitochondria.

4
Functional Aspects of Phospholipids

4.1
Sources of Second Messengers

Phospholipids as a source of cellular second messengers is a relatively new concept. It has been known for several decades that phospholipids store arachidonic acid that is released for the biosynthesis of eicosanoids. These, however, are not truly second messengers, but are hormones that act on nearby cells. Approximately 40 years ago, there was a clue that derivatives of PI might be playing a signaling role in cells when Hokin and Hokin observed the rapid turnover of PI in pancreatic extracts from pigeon incubated with acetylcholine. However, significant progress occurred only in the early 1980s when it was recognized that within 20 to 30 s after the binding of an agonist to a receptor, there was a rapid catabolism of PI-4,5-P_2 to DG and inositol-1,4,5-P_3. It is now known that this process involves a coupling between the receptor on the cell surface and a phospholipase C by trimeric guanidine triphosphate (GTP) binding proteins.

Ten isoenzymes of PI-specific phospholipase C have been identified. All isoenzymes require calcium and degrade PI, PI-4-P, and PI-4,5-P_2. Which lipid is preferentially degraded depends on the conditions of the assay, but PC and PE are not substrates. Although these phospholipases have been purified from cytosol, the substrate is in the plasma membrane, and this is the probable site for hydrolysis.

The inositol-1,4,5-P_3 mediates the release of calcium from the ER. A receptor on the ER has been identified

that has a molecular weight of 313 kDa and binds inositol-P_3, besides having calcium-channel activity. Calcium released by this mechanism after the binding of vasopressin to hepatocyte receptors is known to be involved in the activation of phosphorylase.

The other product of PI-4,5-P_2 hydrolysis is DG and this appears to function as a second messenger by the activation of protein kinase C. This kinase was discovered in 1977 by Nishizuka and coworkers and has been shown to phosphorylate many different cellular proteins. Such phosphorylated proteins are involved in the regulation of cellular processes such as secretion, exocytosis, gene expression, and down-regulation of receptors. For example, the epidermal growth factor receptor is a substrate for protein kinase C. When this receptor is phosphorylated, it exhibits a decreased binding of the growth factor and its tyrosine kinase activity is attenuated.

Ten isoenzymes of protein kinase C have been identified and there may be more. All have a regulatory domain and a kinase domain. Protein kinase C exists in the cytosol as an inactive reservoir. On the generation of DG in the plasma membrane, the kinase binds to the membrane in which it is activated. Some isoenzymes of protein kinase C require calcium and PS for activity, whereas others do not.

More recently, there has been evidence that ceramide derived from SM hydrolysis might be a cellular second messenger and may activate a protein phosphate phosphatase. Similarly, DG released from PC hydrolysis has been implicated as a second messenger in cells. It is likely that we are only now seeing the "tip of the iceberg" with respect to the production of cellular second messengers derived from phospholipids.

In 1979, a potent, biologically active phospholipid was discovered that could aggregate platelets at a concentration of 10^{-11} M. The compound is now usually referred to as *platelet-activating factor* (PAF) and the structure has been identified as 1-alkyl-2-acetyl-*sn*-glycerol-3-phosphocholine. Thus, PAF is an unusual derivative of PC in which there is an ether-linked alkyl chain on carbon 1 and an acetyl group, rather than a long-chain fatty acid, on carbon 2 of the glycerol backbone. Various blood cells, healthy tissues, and tumor cells can synthesize and degrade PAF. PAF exerts many different kinds of biological responses. For example in neutrophils, PAF causes aggregation and degranulation responses, induces chemotaxis, increases cell adherence, enhances the respiratory burst, and stimulates the release of arachidonic acid for eicosanoid production. PAF has been implicated as a contributing factor in diseases such as asthma, hypertension, allergies, inflammation, and others. Specfic receptors for PAF are present on the plasma membrane of cells and thus most of PAF's effects seem to be mediated through receptors.

4.2 Anchoring of Proteins to Membranes

Most proteins on the exterior surface of cells are linked to the plasma membrane through a transmembrane domain. Over a decade ago, it was discovered that a variety of cell surface proteins are linked covalently to a glycosyl PI (GPI) in the plasma membrane. The GPI anchors bind a wide variety of proteins to the cell surface, such as hydrolytic enzymes, cell surface antigens, protozoan antigens (the most intensely studied is the variant surface glycoproteins of *Trypansoma brucei*), and proteins involved in cell–cell interactions.

The advantage to the cell of having cell surface proteins attached via the GPI anchor is not known. However, proteins linked in this manner can be released from the cell surface by hydrolysis of the PI moiety. A specific phospholipase D, which hydrolyzes the bond between the phosphate and inositol moiety, has been purified from human serum and cloned. Thus, one possible role for the GPI anchor would be to allow selective release of proteins from the cell surface. The release of glycoprotein antigens from the cell surface of *Trypanosomes* may be primarily involved in the parasite's ability to evade the immune system of infected hosts.

The core structure of GPI consists of PI linked to glucosamine, which is attached to a string of 3 mannose residues. To the terminal mannose, an ethanolamine phosphate residue is attached. The amine of the ethanolamine is linked to the carboxy terminus of the protein in an amide linkage. This core structure is found in a variety of cell types from brain to *T. brucei*. There are variations on the structure of the GPI moiety that include the addition of additional ethanolamine residues, fatty acids, and saccharides. In animal cells, the fatty acid at the 1 position of glycerol is often replaced by an ether-linked chain.

The biosynthetic pathway for the GPI anchor has now been established. Initially, there is a transfer of *N*-acetyl-glucosamine to PI followed by deacetylation. The mannose residues are then donated from dolichol-P-mannose. The ethanolamine P residue appears to be donated from PE. The proteins appear to be linked to GPI in the ER shortly after protein synthesis is completed. There is evidence that a transmembrane segment of the protein is hydrolyzed before or concomitant with addition of the GPI anchor.

5 Perspectives

The field of phospholipid research has entered a new and exciting period with the advent of molecular techniques. With the availability of various genomic sequences and the advent of proteomics, we can expect that most genes and proteins involved in phospholipid biosynthesis will be discovered in the next decade. Transcriptional regulation of phospholipid biosynthetic enzymes should emerge into a major area of research. The utilization of gene targeting will provide tremendous new insight into the function of these genes and proteins. The application of mass spectrometry to lipid analysis has emerged as lipidomics, where we can expect that most of the major lipids in a particular cell or tissue will be identified and quantitated. Variations in lipid compositions under various physiological perturbations (e.g. fasting, obesity) should provide new insight into the function of these lipids.

See also FTIR of Biomolecules; Intracellular Fatty Acid Binding Proteins and Fatty Acid Transport; Lipid and Lipoprotein Metabolism.

Bibliography

Books and Reviews

Baumann, N.A., Menon, A.K. (2002) Lipid Modifications of Proteins, in: Vance, D.E., Vance, J.E. (Eds.) *Biochemistry of Lipids, Lipoproteins and Membranes*, Elsevier Science, Amsterdam, pp. 37–54.

Dowhan, W., Bogdanov, M. (2002) Functional Roles of Lipids in Membranes, in: Vance, D.E., Vance, J.E. (Eds.) *Biochemistry of Lipids, Lipoproteins and Membranes*, Elsevier Science, Amsterdam, pp. 1–35.

Heath, R.J., Jackowski, S., Rock, C.O. (2002) Fatty Acid and Phospholipid Metabolism in Prokaryotes, in: Vance, D.E., Vance, J.E. (Eds.) *Biochemistry of Lipids, Lipoproteins and Membranes*, Elsevier Science, Amsterdam, pp. 55–92.

McPhail, L.C. (2002) Glycerolipids in Signal Transduction, in: Vance, D.E., Vance, J.E. (Eds.) *Biochemistry of Lipids, Lipoproteins and Membranes*, Elsevier Science, Amsterdam, pp. 315–340.

Schmid, K.M., Ohlrogge, J.B. (2002) Lipid Metabolism in Plants, in: Vance, D.E., Vance, J.E. (Eds.) *Biochemistry of Lipids, Lipoproteins and Membranes*, Elsevier Science, Amsterdam, pp. 93–126.

Smith, W.L., Murphy, R.C. (2002) The Eicosanoids: Cyclooxygenase, Lipoxygenase, and Epoxygenase Pathways, in: Vance, D.E., Vance, J.E. (Eds.) *Biochemistry of Lipids, Lipoproteins and Membranes*, Elsevier Science, Amsterdam, pp. 341–371.

Snyder, F., Lee, T.-C., Wykle, R.L. (2002) Ether-linked Lipids and their Bioactive Species, in: Vance, D.E., Vance, J.E. (Eds.) *Biochemistry of Lipids, Lipoproteins and Membranes*, Elsevier Science, Amsterdam, pp. 233–261.

Sohlenkamp, C., Lopez-Lara, I.M., Geiger, O. (2003) Biosynthesis of phosphatidylcholine in bacteria, *Prog. Lipid Res.* **42**, 115–162.

Tabas, I. (2002) Lipids and Atherosclerosis, in: Vance, D.E., Vance, J.E. (Eds.) *Biochemistry of Lipids, Lipoproteins and Membranes*, Elsevier Science, Amsterdam, pp. 573–600.

Vance, D.E. (2002) Phospholipid Biosynthesis in Eukaryotes, in: Vance, D.E., Vance, J.E. (Eds.) *Biochemistry of Lipids, Lipoproteins and Membranes*, Elsevier Science, Amsterdam, pp. 205–232.

Voelker, D.R. (2002) Lipid Assembly into Cell Membranes, in: Vance, D.E., Vance, J.E. (Eds.) *Biochemistry of Lipids, Lipoproteins and Membranes* Elsevier Science, Amsterdam, pp. 449–481.

Primary Literature

Aitken, J.F., van Heusden, G.P.H., Temkin, M., Dowhan, W. (1990) The gene encoding the phosphatidylinositol transfer protein is essential for cell growth, *J. Biol. Chem.* **265**, 4711–4717.

Aoyama, C., Ohtani, A., Ishidate, K. (2002) Expression and characterization of the active molecular forms of choline/ethanolamine kinase-alpha and -beta in mouse tissues, including carbon tetrachloride-induced liver, *Biochem. J.* **363**, 777–784.

Aoyama, C., Yamazaki, N., Terada, H., Ishidate, K. (2000) Structure and characterization of the genes for murine choline/ethanolamine kinase isozymes alpha and beta, *J. Lipid Res.* **41**, 452–464.

Atkinson, K., Fogel, S., Henry, S.A. (1980) Yeast mutant defective in phosphatidylserine synthesis, *J. Biol. Chem.* **255**, 6653–6661.

Audubert, F., Vance, D.E. (1983) Pitfalls and problems in studies on the methylation of phosphatidylethanolamine, *J. Biol. Chem.* **258**, 10695–10701.

Bakovic, M., Waite, K., Tang, W., Tabas, I., Vance, D.E. (1999) Transcriptional activation of the murine CTP:phosphocholine cytidylyltransferase gene (Ctpct): combined action of upstream stimulatory and inhibitory cis-acting elements, *Biochim. Biophys. Acta* **1438**, 147–165.

Bell, R.M., Coleman, R. (1980) Enzymes of glycerolipid synthesis in eukaryotes, *Ann. Rev. Biochem.* **49**, 459–487.

Bergo, M.O., Gavino, B.J., Steenbergen, R., Sturbois, B., Parlow, A.F., Sanan, D.A., Skarnes, W.C., Vance, J.E., Young, S.G. (2002) Defining the importance of phosphatidylserine synthase 2 in mice, *J. Biol. Chem.* **277**, 47701–47708.

Bladergroen, B.A., Houweling, M., Geelen, M.J.H., van Golde, L.M.G. (1999) Cloning and expression of CTP:phosphoethanolamine cytidylyltransferase cDNA from rat liver, *Biochem. J.* **343**, 107–114.

Bligh, E.G., Dyer, W.J. (1959) A rapid method of total lipid extraction and purification, *Can. J. Biochem. Physiol.* **37**, 911–917.

Blok, M., Wirtz, K.W.A., Scherphof, G.L. (1971) Exchange of phospholipids between microsomes and inner and outer mitochondrial membranes of rat liver, *Biochim. Biophys. Acta* **233**, 61–75.

Bogdanov, M., Umeda, M., Dowhan, W. (1999) Phospholipid-assisted refolding of an integral membrane protein. Minimum structural features for phosphatidylethanolamine to act as a molecular chaperone, *J. Biol. Chem.* **274**, 12339–12345.

Borkenhagen, L.F., Kennedy, E.P., Fielding, L. (1961) Enzymatic formation and decarboxylation of phosphatidylserine, *J. Biol. Chem.* **236**, 28–32.

Bremer, J., Figard, P.H., Greenberg, D.M. (1960) The biosynthesis of choline and its relation to phospholipid metabolism, *Biochim. Biophys. Acta* **43**, 477–488.

Carter, J.M., Waite, K.A., Campenot, R.B., Vance, J.E., Vance, D.E. (2003) Enhanced expression and activation of CTP:phosphocholine cytidylyltransferase beta 2 during neurite outgrowth, *J. Biol. Chem.* **278**, 44988–44994.

Cases, S., Stone, S.J., Zhou, P., Yen, E., Tow, B., Lardizabal, K.D., Voelker, T., Farese, R.V. Jr. (2001) Cloning of DGAT2, a second mammalian diacylglycerol acyltransferase, and related family members, *J. Biol. Chem.* **276**, 38870–38876.

Cornell, R.B., Northwood, I.C. (2000) Regulation of CTP:phosphocholine cytidylyltransferase by amphitropism and relocalization, *Trends Biochem. Sci.* **25**, 441–447.

Cui, Z., Houweling, M., Chen, M.H., Record, M., Chap, H., Vance, D.E., Terce, F. (1996) A genetic defect in phosphatidylcholine biosynthesis triggers apoptosis in Chinese hamster ovary cells, *J. Biol. Chem.* **271**, 14668–14671.

Cui, Z., Vance, J.E., Chen, M.H., Voelker, D.R., Vance, D.E. (1993) Cloning and expression of a novel phosphatidylethanolamine N-methyltransferase, *J. Biol. Chem.* **268**, 16655–16663.

DeLong, C.J., Shen, Y.-J., Thomas, M.J., Cui, Z. (1999) Molecular distinction of phosphatidylcholine synthesis between the CDP-choline pathway and phosphatidylethanolamine methylation pathway, *J. Biol. Chem.* **274**, 29683–29688.

Dowhan, W., Wickner, W.T., Kennedy, E.P. (1974) Purification and properties of phosphatidylserine decarboxylase from Escherichia coli, *J. Biol. Chem.* **249**, 3079–3084.

Esko, J.D., Nishijima, M., Raetz, C.R.H. (1982) Animal cells dependent on exogenous phosphatidylcholine for membrane biogenesis, *Proc. Natl. Acad. Sci. USA* **79**, 1698–1702.

Gaynor, P.M., Gill, T., Toutenhoofd, S., Summers, E.F., McGraw, P., Homann, M.J., Henry, S.A., Carman, G.M. (1991) Regulation of phosphatidylethanolamine methyltransferase and phospholipid methyltransferase by phospholipid precursors in Saccharomyces cerevisiae, *Biochim. Biophys. Acta* **1090**, 326–332.

Golfman, L.S., Bakovic, M., Vance, D.E. (2001) Transcription of the CTP:phosphocholine cytidylyltransferase alpha gene is enhanced during the S phase of the cell cycle, *J. Biol. Chem.* **276**, 43688–43692.

Hajra, A., Bishop, J.E. (1982) Glycerolipid biosynthesis in peroxisomes via the acyl dihydroxyacetone phosphate pathway, *Ann. N.Y. Acad. Sci.* 170–182.

Holub, B.J. (1978) Molecular species of phosphatidylcholine and phosphatidylethanolamine, *J. Biol. Chem.* **253**, 691–696.

Ishidate, K. (1997) Choline/ethanolamine kinase from mammalian tissues, *Biochim. Biophys. Acta* **1348**, 70–78.

Ishidate, K., Nakagomi, K., Nakazawa, Y. (1984) Complete purification of choline kinase from rat kidney and preparation of rabbit antibody against rat kidney choline kinase, *J. Biol. Chem.* **259**, 14706–14710.

Jamil, H., Yao, Z., Vance, D.E. (1990) Feedback regulation of CTP:Phosphocholine cytidylyltransferase translocation between cytosol and endoplasmic reticulum by phosphatidylcholine, *J. Biol. Chem.* **265**, 4332–4339.

Kanoh, H., Imai, S.-I., Yamada, K., Sakane, F. (1992) Purification and properties of phosphatidic acid phosphatase from porcine thymus membranes, *J. Biol. Chem.* **267**, 25309–25314.

Kennedy, E.P. (1957) Metabolism of Lipides, *Ann. Rev. Biochem.* **26**, 119–148.

Kennedy, E.P., Weiss, S.B. (1956) The function of cytidine coenzymes in the biosynthesis of phospholipides, *J. Biol. Chem.* **222**, 193–214.

Kent, C. (1997) CTP:phosphocholine cytidylyltransferase, *Biochim. Biophys. Acta* **1348**, 79–90.

Kuge, O., Nishijima, M., Akamatsu, Y. (1985) Isolation of a somatic cell mutant defective in phosphatidylserine biosynthesis, *Proc. Natl. Acad. Sci. USA* **82**, 1926–1930.

Kuge, O., Nishijima, M., Akamatsu, Y. (1991) A cloned gene encoding phosphatidylserine decarboxylase complements the phosphatidylserine biosynthetic defect of a Chinese hamster ovary cell mutant, *J. Biol. Chem.* **266**, 6370–6376.

Lagace, T.A., Miller, J.R., Ridgway, N.D. (2002) Caspase processing and nuclear export of CTP:phosphocholine cytidylyltransferase alpha during farnesol-induced apoptosis, *Mol. Cell Biol.* **22**, 4851–4862.

Lykidis, A., Baburina, I., Jackowski, S. (1999a) Distribution of CTP:phosphocholine cytidylyltransferase (CCT) isoforms: identification of a new CCTß splice variant, *J. Biol. Chem.* **274**, 26992–27001.

Lykidis, A., Baburina, I., Jackowski, S. (1999b) Distribution of CTP:phosphocholine cytidylyltransferase (CCT) isoforms, *J. Biol. Chem.* **274**, 26992–27001.

Lykidis, A., Jackowski, S. (2001) Regulation of mammalian cell membrane biosynthesis, *Prog. Nucl. Acid Res. Mol. Biol.* **65**, 361–393.

Lykidis, A., Murti, K.G., Jackowski, S. (1998) Cloning and characterization of a second human CTP:phosphocholine cytidylyltransferase, *J. Biol. Chem.* **273**, 14022–14029.

Morand, J.N., Kent, C. (1989) Localization of the membrane-associated CTP:phosphocholine cytidylyltransferase in chinese hamster ovary cells with an altered membrane composition, *J. Biol. Chem.* **264**, 13785–13792.

Nikoloff, D.M., Henry, S.A. (1991) Genetic analysis of yeast phospholipid biosynthesis, *Annu. Rev. Genet.* **25**, 559–583.

Noga, A.A., Stead, L.M., Zhao, Y., Brosnan, M.E., Brosnan, J.T., Vance, D.E. (2003) Plasma homocysteine is regulated by phospholipid methylation, *J. Biol. Chem.* **278**, 5952–5955.

Noga, A.A., Vance, D.E. (2003) A gender-specific role for phosphatidylethanolamine N-methyltransferase-derived phosphatidylcholine in the regulation of plasma high density and very low density lipoproteins in mice, *J. Biol. Chem.* **278**, 21851–21859.

Noga, A.A., Zhao, Y., Vance, D.E. (2002) An unexpected requirement for phosphatidylethanolamine *N*-methyltransferase in the secretion of very low density lipoproteins, *J. Biol. Chem.* **277**, 42358–42365.

Pagano, R.E., Sleight, R.G. (1985) Defining lipid transport pathways in animal cells, *Science* **229**, 1051–1057.

Pelech, S.L., Vance, D.E. (1984) Regulation of phosphatidylcholine biosynthesis, *Biochim. Biophys. Acta* **779**, 217–251.

Pelech, S.L., Vance, D.E. (1989) Signal transduction via phosphatidylcholine cycles, *Trends Biochem. Sci.* **14**, 28–30.

Ridgway, N.D., Vance, D.E. (1987) Purification of phosphatidylethanolamine *N*-methyltransferase from rat liver, *J. Biol. Chem.* **262**, 17231–17239.

Rothman, J.E., Kennedy, E.P. (1977) Rapid transmembrane movement of newly synthesized phospholipids during membrane assembly, *Proc. Natl. Acad. Sci. USA* **74**, 1821–1825.

Ruetz, S., Gros, P. (1994) Phosphatidylcholine translocase: a physiological role for the mdr2 gene, *Cell* **77**, 1071–1081.

Seegmiller, A.C., Dobrosotskaya, I., Goldstein, J.L., Ho, Y.K., Brown, M.S., Rawson, R.B. (2002) The SREBP pathway in drosophila: regulation by palmitate, not sterols, *Dev. Cell* **2**, 229–238.

Shiao, Y.-J., Lupo, G., Vance, J.E. (1995) Evidence that phosphatidylserine is imported into mitochondria via a mitochondria-associated membrane and that the majority of mitochondrial phosphatidylethanolamine is derived from decarboxylation of phosphatidylserine, *J. Biol. Chem.* **270**, 11190–11198.

Shields, D.J., Lehner, R., Agellon, L.B., Vance, D.E. (2003) Membrane topography of human phosphatidylethanolamine *N*-methyltransferase, *J. Biol. Chem.* **278**, 2956–2962.

Smit, J.J.M., Schinkel, A.H., Oude Elferink, R.P.J., Groen, A.K., Wagenaar, E., van Deemter, L., Mol, C.A.A.M., Ottenhoff, R., van der Lugt, N.M.T., van Roon, M.A., van der Valk, M.A., Offerhaus, G.J.A., Berns, A.J.M., Borst, P. (1993) Homozygous disruption of the murine mdr2 P-glycoprotein gene leads to a complete absence of phospholipid form bile and to liver disease, *Cell* **75**, 451–462.

Sugimoto, H., Bakovic, M., Yamashita, S., Vance, D.E. (2001) Identification of transcriptional enhancer factor-4 as a transcriptional modulator of CTP:phosphocholine cytidylyltransferase α, *J. Biol. Chem.* **276**, 12338–12344.

Sugimoto, H., Sugimoto, S., Tatei, K., Obinata, H., Bakovic, M., Izumi, T., Vance, D.E. (2003) Identification of Ets-1 as an important transcriptional activator of CTP:phosphocholine cytidylyltransferase alpha in COS-7 cells and co-activation with transcriptional enhancer factor-4, *J. Biol. Chem.* **278**, 19716–19722.

Sundler, R., Akesson, B. (1975) Regulation of phospholipid biosynthesis in isolated rat

hepatocytes. Effect of different substrates, *J. Biol. Chem.* **250**, 3359–3367.

Trotter, P.J., Pedretti, J., Voelker, D.R. (1993) Phosphatidylserine decarboxylase from Saccharomyces cerevisiae: isolation of mutants, cloning of the gene, and creation of a null allele, *J. Biol. Chem.* **268**, 21416–21424.

Trotter, P.J., Pedretti, J., Yates, R., Voelker, D.R. (1995) Phosphatidylserine decarboxylase 2 of Saccharomyces cerevisiae. Cloning and mapping of the gene, heterologous expression and creation of the null allele, *J. Biol. Chem.* **270**, 6071–6080.

van den Bosch, H. (1974) Phosphoglyceride metabolism, *Ann. Rev. Biochem.* **43**, 243–277.

Vance, D.E. (1990a) Phosphatidylcholine metabolism: masochistic enzymology, metabolic regulation, and lipoprotein assembly, *Biochem. Cell Biol.* **68**, 1151–1165.

Vance, J.E. (1990b) Phospholipid synthesis in a membrane fraction associated with mitochondria, *J. Biol. Chem.* **265**, 7248–7256.

Vance, J.E. (1998) Eukaryotic lipid-biosynthetic enzymes: the same but not the same, *Trends Biochem. Sci.* **23**, 423–428.

Vance, D.E., Choy, P.C. (1979) How is phosphatidylcholine biosynthesis regulated? *Trends Biochem. Sci.* **4**, 145–148.

Vance, D.E., Choy, P.C., Farren, S.B., Lim, P.H., Schneider, W.J. (1977) Asymmetry of phospholipid biosynthesis, *Nature* **270**, 268–269.

Vance, J.E., Pan, D., Vance, D.E., Campenot, R.B. (1991) Biosynthesis of membrane lipids in rat axons, *J. Cell Biol.* **115**, 1061–1068.

Vance, D.E., Pelech, S.L. (1984) Enzyme translocation in the regulation of phosphatidylcholine biosynthesis, *TIBS* **9**, 17–20.

Vance, J.E., Vance, D.E. (1990) Lipoprotein assembly and secretion by hepatocytes, *Ann. Rev. Nutr.* **10**, 337–356.

Voelker, D.R. (1989) Reconstitution of phosphatidylserine import into rat liver mitochondria, *J. Biol. Chem.* **264**, 8019–8025.

Walkey, C.J., Donohue, L.R., Bronson, R., Agellon, L.B., Vance, D.E. (1997) Disruption of the murine gene encoding phosphatidylethanolamine N-methyltransferase, *Proc. Natl. Acad. Sci. USA* **94**, 12880–12885.

Walkey, C.J., Yu, L., Agellon, L.B., Vance, D.E. (1998) Biochemical and evolutionary significance of phospholipid methylation, *J. Biol. Chem.* **273**, 27043–27046.

Wang, Y., MacDonald, J.I.S., Kent, C. (1995) Identification of the nuclear localization signal of rat liver CTP:phosphocholine cytidylyltransferase., *J. Biol. Chem.* **270**, 354–360.

Weinhold, P.A., Rounsifer, M.E., Feldman, D.A. (1986) The purification and characterization of CTP:phosphorylcholine cytidylyltransferase from rat liver, *J. Biol. Chem.* **261**, 5104–5110.

Wetterau, J.R., Zilversmit, D.B. (1985) Purification and characterization of microsomal triglyceride and cholesteryl ester transfer protein from bovine liver microsomes, *Chem. Phys. Lipids* **38**, 205–222.

Yamashita, S., Oshima, A. (1980) Regulation of phosphatidylethanolamine methyltransferase level by myo-inositol in Saccharomyces cerevisiae, *Eur. J. Biochem.* **104**, 611–616.

Yao, Z., Vance, D.E. (1988) The active synthesis of phosphatidylcholine is required for very low density lipoprotein secretion from rat hepatocytes, *J. Biol. Chem.* **263**, 2998–3004.

Zhou, Q., Zhao, J., Stout, J.G., Luhm, R.A., Wiedmer, T., Sims, P.J. (1997) Molecular cloning of human plasma membrane phospholipid scramblase: a protein mediating transbilayer movement of plasma membrane phospholipids, *J. Biol. Chem.* **272**, 18240–18244.

Zilversmit, D.B., Hughes, M.E. (1977) Extensive exchange of rat liver microsomal phospholipids, *Biochim. Biophys. Acta* **469**, 99–110.

Physical Labeling: *see* Labeling, Biophysical

Plant Gene Expression, Regulation of

Pravendra Nath, Vidhu A. Sane, Aniruddha P. Sane, and Prabodh K. Trivedi
National Botanical Research Institute, Lucknow, India

1 Basic Mechanism of Plant Gene Expression 312
1.1 RNA Polymerases 313
1.2 Promoters and Other Regulatory Elements 314
1.3 Assembly of the Transcription Initiation Complex 315
1.4 Transcription Factors and Their Motifs 317
1.5 RNA Processing 317
1.5.1 3′-End Formation 317
1.5.2 5′-End Capping 319
1.5.3 Splicing of Introns 320
1.5.4 Alternative Splicing 321
1.6 Translation 322
1.6.1 Initiation of Translation 322
1.6.2 Elongation of Polypeptide Chain 323
1.6.3 Termination of the Polypeptide Synthesis 324
1.7 Expression of Plant Organelle Genome 324
1.7.1 Expression of Chloroplast Genome 324
1.7.2 Expression of Mitochondrial Genome 327

2 Light Regulation of Gene Expression 329
2.1 Phytochromes 330
2.1.1 Gene Regulation by PhyA Signaling 330
2.1.2 Gene Regulation by PhyB Signaling 331
2.2 Cryptochromes 331
2.3 Phototropins 331
2.4 COP1 – A Major Regulator of Light Responses 331
2.5 Light Mediated Changes in Plastids 332
2.6 Light Regulation of the Circadian Clock 332
2.7 Light Regulation of Flowering Time 333
2.8 Interaction of the Light Signal with Hormonal Pathways 334

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

2.9 Nuclear-organelle Interactions 335
2.9.1 Signaling Between Nucleus and the Organelles 336

3 Hormonal Regulation of Gene Expression 337
3.1 Regulation by Auxins 337
3.2 Regulation by Cytokinin 339
3.3 Regulation by Gibberellins 340
3.4 Regulation by Abscisic Acid 341
3.5 Regulation by Ethylene 342
3.6 Regulation by Jasmonic Acid 344
3.7 Regulation by Brassinosteroids 345
3.8 Control of Gene Expression by Proteolytic Degradation: Role of the Proteasome 346
3.9 Cross Talk Between Phytohormones 347

4 Gene Regulation During Stress 348
4.1 Regulation during Abiotic Stress 348
4.2 Regulation during Biotic Stress and Defense 349
4.3 Regulation by Metabolites 350
4.3.1 Role of Nitric Oxide in Defense 350
4.3.2 Role of Reactive Oxygen in Stress 350

5 Developmental Regulation of Genes 351

6 Epigenetic Regulation of Plant Development 352
6.1 Role of miRNAs in Plant Development 352
6.2 Repression by DNA Methylation 353

Summary 353

Bibliography 354
Books and Reviews 354
Primary Literature 354

Keywords

Transcription
Transcription is the process by which genetic information stored in DNA is copied into RNA transcript. It is achieved by recruiting DNA dependant RNA polymerase, which binds in the promoter region of the gene along with several other factors and slides down in a 3′ to 5′ direction on the complementary strand of DNA adding ribonucleotides to give an exact copy of the sense strand.

Transcript Processing
Majority of eukaryotic RNAs are synthesized as pre-RNA and processed posttranscriptionally to give a mature RNA molecule. Processing could include 5′ capping, 3′-end nucleolytic cleavage followed by polyadenylation, intron removal, alternative splicing, and RNA editing.

Translation
Translation is the process by which a mature transcript is translated into a polypeptide chain. This requires recruitment of ribosomal complex, aminoacyl tRNA, various translational and elongational factors on the mRNA and finally termination of the chain.

Cis-Elements and Trans-Acting Regulatory Factors
The cis-elements are DNA motifs present within and outside a gene including promoter and terminator regions and that allow other factors (known as *trans-factors*), mostly proteins, to bind with these in order to initiate/regulate transcription.

Photoreceptors
These are protein molecules with a chromophore, which sense light of different wavelengths and intensities and signal to provide various responses that lead to transcription of several genes.

Nuclear Organelle Genome Interaction
Chloroplast and mitochondria, though they have their own genome, depend largely on the gene products from nucleus, which are transported to these organelles and assemble together with organelle genome products to give a functional unit. The messages are mostly carried from nucleus to organelle; however, recently a vice versa interaction has also been reported.

Phytohormones
These are low molecular weight substances produced by plants, which exert their effects in very minute quantities mostly after binding to their receptors. The binding to the receptors evokes a cascade of events including expression of several genes (factors), which ultimately regulate expression of target genes.

Proteasome
This is a complex that degrades proteins marked by polyubiquitination, and brought to it by SCF complex. Various signaling pathways are activated by the proteasome proteolysis in order to transduce responses on the target genes.

Abiotic and Biotic Stress Responses
Plants sense various abiotic (heat, cold, drought, salinity, etc.) and biotic (pathogens, insects, herbivores) stresses and respond to these through hormones, metabolites, and

other physiological means that initiate protective measures by activating stress related genes.

MADS-Box Genes

These are transcription factors that possess a highly conserved 180 nucleotide domain known as *MADS box*. They are mainly involved in flower and fruit development and in plants contain an additional K-box for dimerization.

Epigenetic Regulation of Transcription

Activation and repression of genes by means that do not involve changes in the gene sequence. It includes transcriptional gene silencing (TGS) and posttranscriptional gene silencing (PTGS), is brought about by histone modification and DNA methylation as well as RNA degradation/inhibition of translation by siRNAs and miRNAs.

Eukaryotic genomes are much more complex than prokaryotes and plant genomes present more complexities than other eukaryotes. The continuous process of development, absence of germ-line, flexible and reversible cellular differentiation, polyploidy, and so on, are characteristic features that distinguish plant genomes from other eukaryotes. The complexities observed in plant genomes can possibly be attributed to three major factors. Firstly, plants are nonmotile and hence exposed to various environmental conditions constantly. In the absence of any neuronal network or immunological system, plants have evolved various local and systemic mechanisms to adapt to the environment to cope with biotic and abiotic stresses. This is achieved mainly through various signaling systems of phytohormones and other metabolites and the scores of genes participating in it. Secondly, plants from various species and genera synthesize about 45 000 different natural products or secondary metabolites. The functions of many of these are unknown; however, a number of these are used in defense, as attractants and volatile signals. This involves complex metabolic pathways involving several genes that require specific regulatory mechanism for expression. Thirdly, and the most puzzling attribute that still remains to be answered satisfactorily, is the variability observed in its genome size. The genome size as such is quite variable throughout the biological world; however, plants show the greatest variation of any kingdom in the biological world. For example, angiospermic plants that do not differ much in their basic activities contain a genome varying from 7×10^4 kb/haploid (*Arabidopsis*) to 1×10^8 kb/haploid (Lilly). Majority of the DNA in plant genome is not transcribed and may account for up to 90% of the genome on the basis of the number of active genes. Further, DNA motifs ranging in length from single base to thousands of bases, repeated many hundreds or thousands of time, are characteristics

of all eukaryotic genomes; however, plant genome may represent 50 to 90% or more of such DNA. The biological significance of this repetitive DNA is still not resolved properly. Of sequence motifs that are highly repeated, some are conserved from one species to another. Satellite DNA, minisatellite, variable-number tandem repeats (VNTR) are some of the major features of plant genome, which make them different from other eukaryotic genomes. The taxonomical usage of these sequences have added great knowledge in establishing phylogeny as well as biodiversity; however, whether they play any major role in plant gene expression and regulation is still not clear.

It was in the early 1960s that the basic principles of gene regulation and its expression in prokaryotes were established. While in prokaryotes the gene regulatory elements were defined as promoters, operators, and positive control elements, unfortunately the mechanism of eukaryotic gene regulation had to wait till recombinant DNA technology was in practice in the early 1970s. High throughput sequencing of genome (genes) and development of various softwares for the analysis of these sequences made it possible to look into functional and regulatory aspects of plant genome. On the basis of the information obtained through *Arabidopsis* and rice genome sequencing and various ESTs (Expressed-sequence tags), plants may express 25 000 to 50 000 genes depending upon species and genera. Of these, about 30 to 35% may be expressed in all parts of the plant (constitutive), whereas others could be organ specific or event specific. The gene product may localize to a certain cell type, appear only at a specific developmental stage, or accumulate following distinct environmental stimuli. More than one inducer may turn on some genes or a single stimulus may have different effects on different genes. The existence of cross talk between different hormones, light, and environment during plant development further complicated the task of defining gene regulation of expression in plants. Fortunately, gene transfer technology and development of mutants and functional genomics have added tremendously to the knowledge of gene expression, regulation, and its function. Reverse genetics is being applied to investigate the function of candidate genes through approaches such as RNA interference (RNAi), TILLING (target induced local lesions in genome) and use of systematically generated T-DNA populations. Comparative genomic approaches, as well as functional assays using mini chromosomes, hold great promise for understanding both cis-DNA elements and the trans-protein components.

In this article, we have tried to include some of the common and recent features of plant gene expression and its regulation. Though it is difficult to describe every gene regulation, the article will provide an overview of various classes of genes that are expressed in plants in general and the mechanism of their regulation. We will describe here topics like the basic mechanism of plant gene transcription and translation; hormonal and light regulation of genes; regulation of expression during biotic and abiotic stress; gene regulation during development; mitochondrial and chloroplast gene expression and regulation; and recent advances in plant gene expression and regulation. The emphasis has been given on the various DNA elements identified within and outside gene and transcription factors, which hold key to all gene expression and regulation.

1
Basic Mechanism of Plant Gene Expression

The flow of genetic information contained in a portion of genetic material, DNA, involves the generation of RNA molecule followed by synthesis of polypeptide. Understanding how the information in DNA is converted into functional proteins became possible through the discovery of messenger RNA and transfer RNA and the decoding of the genetic code. The process by which information in DNA is finally expressed as polypeptides is known as *Central dogma*. This includes three major steps. The first is replication, the copying of parental DNA to form daughter DNA molecules having identical nucleotide sequences. The second is transcription, the process by which parts of DNA are copied precisely to form RNA. The third is translation, in which the genetic message coded in mRNA is translated on the ribosome, through tRNA molecules, into a polypeptide with a specific sequence of amino acids. We will discuss here the process of transcription and translation pertaining to transcribed and/or translated region of the plant DNA. A schematic representation depicting the general mechanism of eukaryotic gene transcription and translation has been presented in Fig. 1

The first step in gene expression is transcription, the synthesis of RNA copy of DNA template. Studies have suggested that in addition to continuous presence of mRNA from constitutive genes, each organ in the plant contains a large number of organ specific mRNAs. Various protein complexes, that bind to DNA and are again subjected to various types of regulation, carry out the control of transcription. Much of our understanding of the basic elements of the transcription is derived from early work on the bacterial system; however, transcription in plants and other eukaryotes is far more complex than in prokaryotes. In prokaryotes, transcription is coupled to translation, that is, as the transcript elongates, it binds to ribosomes and begins synthesizing proteins. In eukaryotes, however, the nuclear envelope separates the genome from the translational machinery. This may allow selective movement of mRNA from the nucleus to reach the ribosomal machinery in cytoplasm for translation purposes. The expression of genes present in organelle genome of the mitochondria and chloroplast, differ from the expression of nuclear genome in that they follow a prokaryotic mode of transcription and translation and yet depend on nucleus for expression of their genes.

RNA polymerases and other DNA binding proteins are responsible for determining which genes are to be expressed at a particular time. For each gene, the transcription process can be conveniently divided into two stages. The first stage is the initiation of transcription and involves assembly of the complex of proteins, including the RNA polymerase enzyme and its various accessory polypeptides upstream of the gene. This step determines whether the gene is actually expressed. In the second stage, the RNA polymerase moves along the gene to synthesize the primary transcript, which is a direct copy of the gene. This second stage is often accompanied by RNA processing events, which include removal of introns, processing of 5′ and 3′ ends, and so on. We will examine each step one by one in more detail in the following section predominantly for nuclear genes. The organelle genome expression and regulation will be described separately at the end of this section.

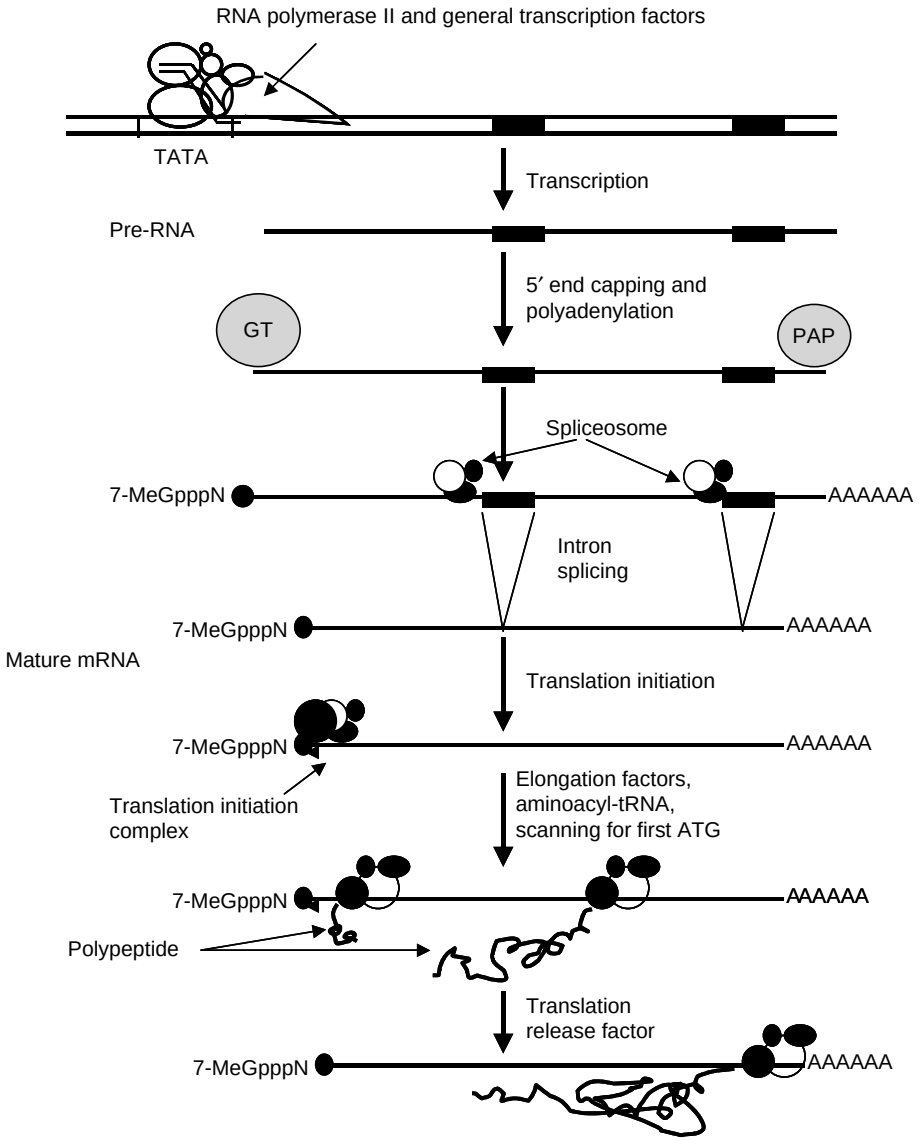

Fig. 1 Basic transcription and translation scheme during plant gene expression. GT, guanylyl transferase; PAP, poly(A) polymerase.

1.1 RNA Polymerases

The enzyme responsible for copying DNA into RNA is called *DNA dependent RNA polymerase*. Transcription of eukaryotic genes requires three different RNA polymerases, namely, RNA polymerase I, RNA polymerase II, and RNA polymerase III. These three RNA polymerases have different responsibilities and locations in the nucleus. RNA polymerase I, which

resides in the nucleolus is responsible for transcribing the genes coding for rRNA; RNA polymerase II is located in the nucleoplasm and required for synthesizing heterogeneous nuclear RNA (hnRNA), the precursor for mRNA. RNA polymerase III is a minor enzyme, located in the nucleoplasm and synthesizes tRNAs and other small RNAs. All plant and other eukaryotic RNA polymerases are large multisubunit (8–12 subunits) proteins with a molecular weight more than 500 kDa. Structurally, these polymerases are quite similar to each other. The largest subunit in RNA polymerase II has a carboxy-terminal domain (CTD), which consists of multiple repeats of the consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser. The sequence is unique to RNA polymerase II. The number of repeats is important, because deletion or removal of more than half a repeat is lethal. The phosphorylation/dephosphorylation of this domain at serine or threonine residue determines the initiation of transcription.

The RNA polymerases that transcribe organellar genes are prokaryotic in nature unlike their counterparts in the nucleus, reflecting the bacterial origins of mitochondria and chloroplast. The organelle genomes are much smaller, hence, the RNA polymerase needs to transcribe relatively few genes, and the control of transcription is likely to be simpler. The mitochondrial RNA polymerase, encoded by the nucleus, consists of a single subunit with a molecular mass of 140 kDa. This enzyme is very closely related to bacteriophage RNA polymerase. The mitochondrial polymerase works in conjunction with other accessory polypeptides. Chloroplasts have RNA polymerases encoded by chloroplast genome as well as nuclear genome. Both types of RNA polymerases are involved in transcription of different sets of genes present in chloroplast.

1.2 Promoters and Other Regulatory Elements

In plants and other eukaryotes, the term "promoter" is used to describe all the nucleotide sequences that are important in initiation and regulation of transcription of genes. For some genes, these sequences can be numerous and diverse in their functions, including not only the "core promoter" or "basal promoter," which is the site at which initiation complex is assembled but also one or more elements upstream of the core promoter. We can divide the structure of a plant promoter into two parts, the core (basal) or minimal promoter, consisting of the minimum upstream sequence required for gene expression, and the additional regulatory sequences that control the activity of the core promoter. Each of the three RNA polymerases binds to a different type of promoter. RNA polymerase I promoters consist of a core promoter spanning the transcription start site, between nucleotide −45 to +20, and an upstream control element. The minimum promoter for genes transcribed by RNA polymerase II typically extends about 100 bp upstream of the transcription initiation site and includes several sequence elements referred to as *proximal promoter sequences*. The TATA box, which plays a crucial role in the transcription has been found to be located about 25 to 30 bp upstream of the transcription start site. It has a sequence of TATAAA(A) and serves as a site for assembly of the transcription initiation complex. Approximately 85% of the plant genes sequenced so far contain TATA box. In addition to the TATA

box, the minimum promoter also contains "CAAT box" and "GC box." These two regulatory elements are the site for binding of transcription factors, proteins that enhance the rate of transcription by facilitating the assembly of the initiation complex. Because these sequences are adjacent to the transcription unit, they are considered regulatory and known as *cis-acting sequences*. The transcription factors that bind to the cis-acting sequences are called *trans-acting factors*, since the genes encoding these are located elsewhere in the genome. Various cis-acting elements and their respective trans-acting factors for the regulation of genes during plant growth and development has been shown in Fig. 2.

Numerous other cis-acting sequences are found to be located further upstream of the proximal promoter and exert either positive or negative control over plant gene expression. These sequences are termed *distal regulatory sequences* and these are usually located within 1 kbp of the transcription initiation site. The positively acting transcription factors binding to these sites are known as *activators* whereas the negative transcription factors binding to these sequences are known as *repressors*. In addition to having regulatory elements within distal regulatory sequences, plant as well as other eukaryotic genes are also regulated by other control elements located tens of thousands of base pairs away from the transcription start site. These elements are known as *enhancers* and may be located either upstream or downstream from the distal regulatory sequences. In plants, many developmentally important genes have been shown to be regulated by enhancers.

RNA polymerase III promoters are unusual in a way that they are located in the gene for which they promote the transcription. These promoters are variable and fall into three different categories. Usually, the core promoter spans approximately 50 to 100 bp. One category of RNA polymerase III is very similar to RNA polymerase II with a range of upstream promoter elements.

1.3 Assembly of the Transcription Initiation Complex

In *Escherichia coli*, a direct contact is established between the promoter and RNA polymerase, whereas in plants, as well as in other eukaryotes, transcription initiation is much more complex. For genes transcribed by RNA polymerase II, the initiation contact is made by the general transcription factor, TFIID (Transcription Factor of RNA polymerase II protein D), which is a complex made up of the TATA-binding proteins (TBP) and at least 12 TATA-binding protein-associated factors (TAFs). Most of these proteins are DNA-binding proteins and have sequence specificity. Both TBP and TAFs appear to have important function in recognition and are indispensable, at least for some genes. The TFIID recognizes the TATA component of the RNA polymerase II core promoter. After TFIID has attached itself to the core promoter, the preinitiation complex is formed by binding of additional general transcription factors in the order TFIIA, TFIIB, TFIIF/RNA polymerase II, TFIIE, and TFIIH. Some of these GTFs (general transcription factors) have DNA-binding properties and others interact purely by protein–protein interaction. The binding of TFIIF, recruits the RNA polymerase II to the complex. After that, TFIIH acts as a DNA helicase, which unwinds the DNA as well as acts as a kinase that adds phosphate

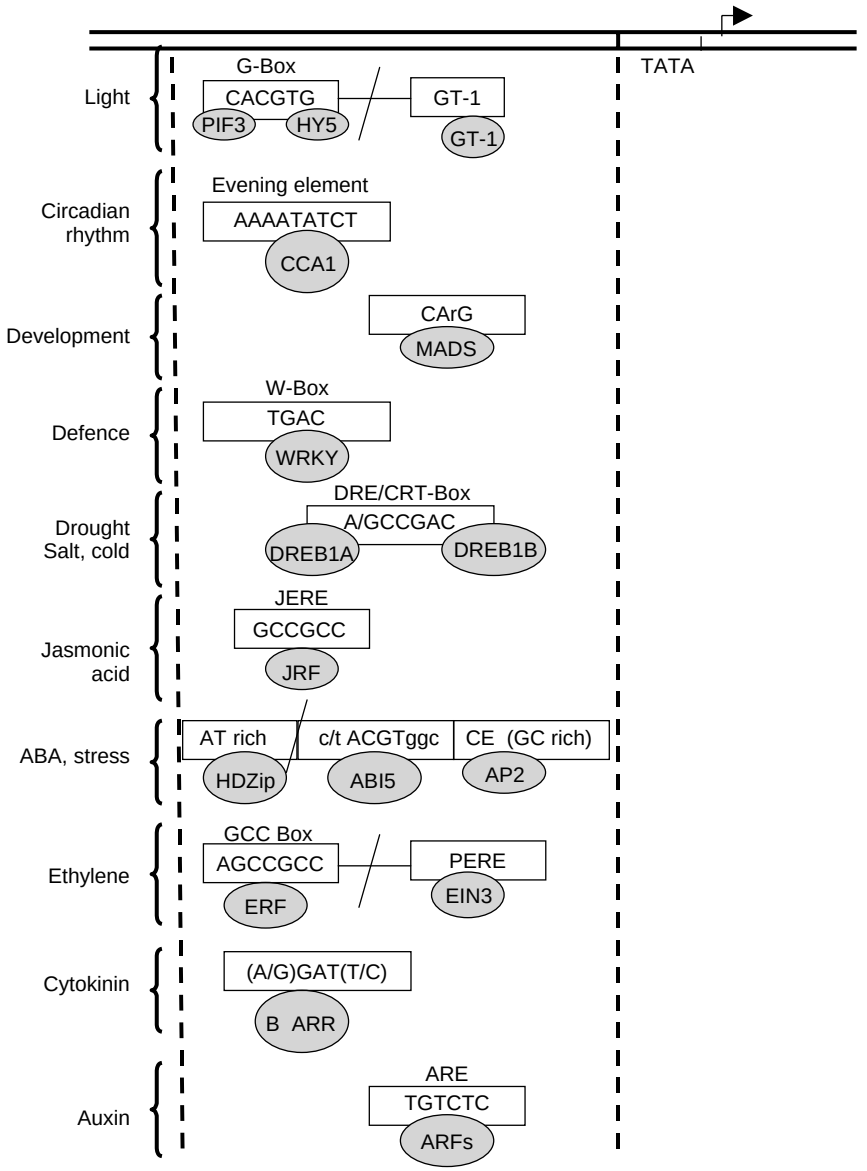

Fig. 2 Various cis-acting elements and associated trans-acting factors during plant gene expression in response to different stimuli.

group to the C-terminal domain of RNA polymerase II. RNA polymerase then separates from the initiation complex and proceeds along the antisense strand in the 3′ to 5′ direction. While some of the general transcription factors dissociate from the complex at this point, others remain bound for another RNA polymerase molecule and initiate another round of transcription.

1.4 Transcription Factors and Their Motifs

In most of the transcription factors, three structural features are invariably common. These are DNA binding domain, a transcription-activating domain, and a ligand binding domain. Some of the transcription factors are needed for the coordinated regulation of genes in plants and other eukaryotes. In plants, the effects induced by light and hormones as well as the adaptive responses caused by various types of stresses involve the coordinate regulation of the groups of the genes that share a common response element upstream of the promoter. The genes that act as major developmental switches, such as the homeotic genes, encode transcription factors that bind to common regulatory sequence present in dozens, or even in hundreds, of genes scattered throughout the genome. A detailed list of major transcription factors and genes regulated by these is given in Table 1.

1.5 RNA Processing

Many RNAs, especially in eukaryotes, are initially synthesized as precursor or pre-RNAs, which have to be processed before they can carry out their function. The most important processing events are splicing, 5′-end capping, polyadenylation, and RNA editing, which are carried out singularly or in conjunction depending upon the message type. Each of these processes is regulated and may provide a means by which expression of individual genes can be modulated. For example, the mRNA resulting from both editing and alternative splicing often displays tissue specificity thereby increasing the coding capacity of the genome without the requirement for additional copies of the relevant genes. We shall describe each processing event below in detail with respect to plants.

1.5.1 3′-End Formation

The poly (A) tail present in the 3′ end of the mature mRNA plays an important role in many aspects of RNA function and metabolism, particularly in mRNA translation and turnover. The mature 3′ end is formed by endonucleolytic cleavage followed by the addition of poly (A) tail to the 3′ end of the cleavage product. The 3′-end formation is a necessary prerequisite for the transcription termination. However, not much is known about mRNA 3′-end processing in plants. The discovery of the importance of the AAUAAA motif for polyadenylation in animal systems prompted the search for similar motifs in plant genes. The analysis suggested that the AAUAAA element is not a prerequisite for 3′-end formation in plants as more than 60% of the 3′ UTRs of the plant genes do not contain an exact match of this sequence at an appropriate position. AAUAAA is nevertheless found as a consensus of the 3′ UTR of many plant genes; however, the information regarding the elements required for the constitution of poly (A) signal is based on sequence comparisons and data from a small number of functional studies. Plant poly(A) signals appear to be much more diffuse, redundant, and complex as compared to their animal counterparts. To our knowledge, only three genes, one from dicot (pea rbcS-E9) and two from monocots (maize 27 kDa zein gene and wheat histone H3 gene) have been functionally analyzed for 3′-end formation. The modular architecture of a plant poly(A) signal can be described

Tab. 1 Major transcription factor families in plants (numbers are given for *Arabidopsis*).

Family	*Genes*	*Function*	*Examples*
Families specific to plants			
1. AP2/ERF	144		
a. AP2/RAV subfamily	20	Floral organ specification, seed development, cell identity	*ANT, APETALA2*
b. ERF subfamily	134	Ethylene response, abiotic stress, defense response	*ERF1, ERF3, ERF4, DREB1a, DREB1b,*
2. NAC domain	109	Auxin response, meristem development, organ separation	*NAC1, NAP, CUC1, CUC2*
3. WRKY	72	Defense response	*tWRKY3, tWRKY4*
4. DOF/CO-like/YABBY (Zn-C2C2 subfamily)	70	Flowering time, circadian rhythm, seed germination, carbon metabolism, abaxial cell fate	*CO, DAG1, CRC*
5. GARP	56	Cytokinin response, photosynthetic development, abaxial cell identity	*ARR1, ARR2, KAN1, PHR1*
6. GRAS	32	Gibberellin response, phytochrome response, root, shoot development	*GA1, RGA1, SCR, SHR*
7. Aux/IAA	27	Auxin response, root development	*IAA3, IAA7, AXR3,*
Families with a significantly higher representation in plants			
1. MYB family	190		
a. MYB (R1)R2R3	136	Stem cell function, trichome and root development, defense response, ABA response, phyA signaling	*GL1, AtMYB2, AtMYB4, LAF1,*
b. MYB related	54	Circadian clock, root development	*CCA1, LHY, CPC*
2. MADS Box	82	Meristem identity, flower and fruit development, abscission, flowering time	*AG, AP1, FLC, SEP1, SEP2, RIN, SHP1, GL15,*
Other major families			
1. basic HLH	139	Phytochrome responses, trichome, root development, abiotic stress response, secondary metabolism	*PIF3, HFR1, rd22BP1, GL3*
2. C2H2 (Zn)	112	Organ development, flowering time, abiotic stress, light response	*EMF2, SUP, VRN2*
3. Homeodomain	90	Development (leaf, stem, root, ovule), stem cell identity, sucrose signaling, shade avoidance	*PHV, REV, STM, WUS, KNOX, ATHB, KNAT1,*
4. bZIP	77	Photomorphogenic response, flower, leaf development, defense, ABA, GA response	*HY5, HYH, ABI5, PAN*

by three features – the near upstream element (NUE), the far upstream element (FUE) and the cleavage site itself. The NUEs characterized to date range from AAUAAA-like motifs to rather unrelated sequences. Although often present in a variant form, the FUE is present in all plant poly(A) signals characterized indicating a requirement for this element to enhance the overall processing efficiency. This is presumably achieved via interaction with protein factors belonging to the processing complex. The FUE of different plant poly(A) signals are interchangeable demonstrating their functional conservation. No conserved sequence element common to all FUE has been reported though they contain U- or UG-rich sequences. Examination of poly(A) sites reveals a consensus of YA at the cleavage site. It has been suggested that the cleavage site itself be defined as a cis-acting component of the poly(A) signal in plants. However, in most cases where the original cleavage site is removed or mutated cleavage still takes place at an appropriate position downstream of a functional NUE/FUE, although with less precision. Taken together, it suggests that although certain nucleotides are preferred as the substrate for a cleavage site, the cleavage site is determined by the position of the NUE.

The exact model for the 3′-end formation in plants is not clear. It seems that cleavage and polyadenylation specificity-like factor (CPSF-like), cleavage stimulation factor (CST-like) and cleavage factor (CF) in addition to poly(A) polymerase play important roles in 3′-end formation in plants. A protein, which has been shown to influence 3′-end processing via interaction with upstream sequences, is the U1 snRNP (U1A) in mammals. The cloning of U1A from potato and *Arabidopsis* revealed that the autoregulatory mechanism present in mammals is not conserved in plants.

Apart from NUE and FUE, the presence of Poly (A) polymerase (PAP) itself can safely be assumed in higher plants. PAP enzyme has been cloned from different sources. Owing to lack of any direct nucleotide sequence homology, it has not been possible to use either the mammalian or yeast PAP clones as molecular probes to isolate a plant homolog. The PAP activity in plants was first reported more than 25 years ago and has been described in a diverse range of plants such as cotton, maize, wheat, pea, and tobacco. A cDNA from pea seedling encodes a polypeptide, which was found active as poly(A) polymerase either in association with RNA binding factor or free polyA as RNA substrate.

1.5.2 5′-End Capping

Another modification before pre-RNA is processed is the 5′-end capping by the addition of 7-methylguanosine at the 5′-end of the transcript via a 5′-to-5′ linkage. The pre-RNA is capped almost immediately after initiation of mRNA synthesis. One of the functions of the 5′ cap is to protect the growing RNA transcript from degradation by RNases. A nuclear enzyme, guanylyl transferase, catalyzes the addition of terminal G at the 5′ end.

Poly(A) tail and 5′-end capping have several functions. They protect transcripts against RNases and therefore increase stability of the transcript. Both the processes are required for transit of the RNA from nuclear pore and increase the efficiency of the translation on the ribosome. The requirement of eukaryotic mRNAs to have both the 5′-cap and poly(A) tail ensures that only properly processed transcripts will reach the ribosome and be translated. By electroporating capped or uncapped

mRNAs and mRNAs with or without poly(A) tail into tobacco protoplast, it has been found that 5′ cap stabilizes reporter transcript two- to fourfold and the poly(A) tail stablizes reporter transcript two- to threefold. Although it is unclear how the 5′-cap and poly(A) tail protect the mRNAs from degradation, an appealing model would be that the physical interaction between the cap and poly(A) tail via their associated factors would sequester the ends of the transcript, protecting them from nucleases.

1.5.3 Splicing of Introns

The removal of introns or untranslatable sequences from pre-mRNA transcript and the concomitant ligation of exons is known as *pre-mRNA splicing*. It is a fundamental feature of gene expression in eukaryotes and a key step at which expression is regulated. Higher plant genes contain introns and the majority contains multiple introns. The gene encoding the 140-kDa subunit of RNA polymerase II in *Arabidopsis* contains 24 introns while the gene encoding acetyl-coenzyme A carboxylase contains 34 introns. Genes encoding highly conserved eukaryotic proteins such as histones, which lack introns in mammals, can contain introns in plants. However, introns are not a universal feature of plant genes, for example, maize zein genes lack introns.

Higher plants introns vary widely in length, but around two-third are shorter than 150 nt and most fall within the size range of 80 to 139 nt. A minimum functional length of 70 nt has been determined for efficient processing of synthetic introns in higher plants and consistent with this, naturally occurring introns shorter than 70 nt are rare. Very few introns in plants have been reported to be more than 2 to 3 kb, for example, maize *pericarp* gene, which has intron of 7 kb. Higher plant introns, belonging particularly to dicots, are characteristically AU rich. Introns of dicot origin average 70% AU and very few (2%) have an AU content less than 59%. Pre-mRNA splicing in monocots does not exhibit the same requirement for AU-rich sequence. Introns of monocot origin average 60% AU, but many (38%) have an AU content less than 59%, with some being as low as 30 to 35%.

The 5′ splice site dinucleotide/GU is conserved in virtually all naturally occurring plant introns. However, in rare instances, the 5′ splice site dinucleotide is /GC. A /GC 5′-splice site is conserved in seven myrosinase genes characterized from three different plant species. The wider 5′-splice site consensus in higher plants, AG/GUAAGU, matches that derived from mammals. The 3′-splice site dinucleotide, AG/, is invariant in naturally occurring higher plant introns. Mutation of the terminal AG/ by deletion, substitution, or point mutation of the terminal G results in abolition of its use. Point mutation in the 3′-dinucleotide AG to UG in maize *bronze*-2 intron abolished processing. The wider consensus derived from plant 3′-splice sites (UGYAG/GU) though it resembles, is more extensive than those for mammals (YAG/G).

Pre-mRNA splicing can, but need not necessarily, occur cotranscriptionally. It takes place within a large ribonucleoprotein complex termed as *spliceosome*. This spliceosome assembles on pre-RNA in a stepwise manner governed by multiple cis-acting elements. Interest in pre-RNA splicing in higher plants increased when it was realized that heterologous introns are not usually processed in transformed plant cells. This indicated that splicing in plants has some unique requirements that

distinguish it from mammals and yeast. The major subunits of the spliceosome are uridylate-rich small nuclear ribonucleoprotein particles (U snRNPs), U1, U2, U4/6 and U5. In addition to spliceosomal U snRNPs, a number of other non-U snRNP proteins are essential for constitutive and regulated splicing. The U snRNPs assemble on the pre-mRNA via an ordered pathway that requires the transient association of diverse non-U snRNP proteins. The defined pathway of assembly and disassembly is called *the spliceosome cycle*. Nascent pre-RNA transcripts first associate with multiple distinct heterogeneous ribonucleoprotein particles (hnRNPs) to form hnRNP complex. This complex is not a functional intermediate in spliceosome assembly, nor a complex specific to the splicing reaction but a substrate recognized by spliceosome machinery. After this, a stepwise complex is formed with spliceosome to process the introns. This stepwise assembly of the spliceosome mediates dynamic array of RNA–RNA interaction at the heart of the catalytically active complex.

Each of the key RNA sequences that form the dynamic skeleton of the spliceosome are conserved in higher plants: the 5′ sequence of U1 snRNA that interacts with 5′ splice site; the U2 sequence that interacts with the branch site and forms helix I with U6; the U5 loop I sequence that makes exons interact; and the U6 sequence that interacts with U2 in helix I and 5′ splice site, are all invariant among plant U snRNAs. Genes encoding each of the spliceosome snRNAs, U1, U2, U4, U5, and U6 have been cloned from several different plant species and analysis suggest that these are highly conserved with their mammalian counterparts in terms of size, sequence, and secondary structures. The fundamental mechanism of pre-RNA splicing in higher plants, in all likelihood, closely resembles that of yeast and mammals. The most direct evidence in support of this observation comes from the demonstration that the splicing in plants is through a branched lariate intermediate.

1.5.4 Alternative Splicing

Primary transcripts of many genes arising from a single transcription unit in animals can be differentially spliced to produce multiple mRNAs giving rise to proteins with different function. The process is needed to regulate gene expression during development and/or in a tissue-specific manner. The number of alternatively spliced genes reported so far in plants is much smaller than that of mammals, but with explosive growth of available EST and genomic sequences the number is increasing. There are five basic models for a gene to be internally and alternatively spliced. In the first model, the intron is retained with the exon. In the second and third models, the intron is alternatively spliced because of alternative donor or acceptor site, which causes part of the intron to be spliced from the pre-RNA. The fourth possibility arises due to exon skip and in that case the donor and acceptor sites of two different introns work together, which results in an mRNA lacking an exon between the two introns. In the fifth model, splicing takes place because of mutually exclusive exons. There are two additional models for alternative splicing and they depend on transcription start selection and transcription termination selection. More than one possibility can apply to the same gene, and the combinations result in a tremendous variation in gene products.

Recent data indicate that intron retention is the most common form of alternative splicing in plants and account for more than 30% of all alternative splicing. Recently, one database "Plant Alternative Splicing Database" (PASDB) has been created with the sequence information of all plant species under *Viridiplantae*. Examples of alternative splicing in plants include rubisco activase, which is a nuclear-encoded chloroplast protein required for the light activation of ribulose 1,5-bisphosphate carboxylase/oxygenase *in vivo*. In most of the plants that are known to have a spliced variant for this gene, two isoforms were found that differ only at the carboxy-terminus, resulting from alternative splicing of a common pre-RNA. In *Arabidopsis*, both isoforms can activate rubisco activity, although light modulation of rubisco activity requires the larger rubisco activase isoform. Alternative splicing has also been observed in MADS-box genes. Spliced variant of MADS-box genes have been reported in maize, barley, rose, and blue gum. Another example of alternative splicing in rice and maize is granule bound starch synthase, an enzyme that is responsible for the synthesis of amylase. Similarly, the starch branching enzyme, which defines the chain length and frequency of branch points for amylopectin, is also known to alternatively splice in maize, rice, wheat, and kidney bean. Recently, phosphoenolpyruvate carboxylase kinase gene has been shown to alternatively splice in different members of *Solanaceae*.

1.6 Translation

Translational efficiency of any system is defined as the rate of polypeptide synthesis per mRNA per unit time. Levels of protein synthesis are therefore determined not only by message abundance but more so by the ability of the transcript to be successfully translated. The translation of mRNA to polypeptide is divided into three different steps. In the first step, the mRNA is recruited to the ribosome (initiation). In the second step, polypeptide chain synthesis takes place (elongation) whereas in the third step, polypeptide chain is released from the translation machinery for further processing (termination). In plants, translation is frequently controlled at initiation and appears to be regulated by competition for limiting translational components, different requirement of specific factors, and cis-acting mRNA elements.

1.6.1 Initiation of Translation

Recruitment of mRNA for translation is a complex process that requires interactions between the mRNA, the two ribosomal subunits and at least nine translational factors, as well as the hydrolysis of ATP and GTP. Many initiation factors (eIF1 to eIF6) have been identified in plant and their role in protein synthesis has been established. In cap-dependent translation, the initiation factor 4 (eIF4) components are responsible for the selection of 5′-capped transcripts and the removal of the secondary structure to facilitate 5′-to-3′ migration of 40S preinitiation complex toward the initiator AUG. The preinitiation complex includes the mRNA, elF4 proteins, eIF2-tRNA^{met}-40S ribosomal subunit complex, eIF3, eIF1A, and eIF5. The preinitiation complex searches for the first AUG triplet. The recognition of the first AUG triplet is assisted by eIF1 and eIF1A. The binding of the 60S ribosomal subunit and hydrolysis of GTP complete the initiation. Sometimes specific trans-acting factors are needed for efficient translation initiation.

A protein that binds to the 5′-GpppN cap of the mRNA mediates the first phase of recruitment of mRNA for translation. In mammals, availability of the cap-binding protein eIF4E is regulated by its interactions with 4E-binding proteins. Plants possess at least three distinct cap-binding proteins. The plant eIf4E homolog interacts with the plant eIF4G homolog. There is also a second and distinct cap-binding protein, eIFiso4E that is apparently unique to plants. This protein interacts with eIFiso4G, an 86-kDa protein unique to plants that also interacts with eIF4A, eIF4B, and poly(A) binding protein. A third cap-binding protein, nCBP, was identified in *Arabidopsis*. It interacts with eIF4iso4G in the yeast two-hybrid systems and functionally replaces eIF4E or eIFiso4E in an *in vitro* translation system.

Factors that regulate the initiation of translation include phosphorylation/ dephosphorylation of factors and 5′- and 3′-UTR interaction. The eIF4A is present in phosphorylated as well as nonphosphorylated forms. The level of phosphorylated eIF4A increases approximately 50% in response to oxygen deprivation and heat shock. The phosphorylation status of eIF4A and eIF4B might influence protein–protein or protein–mRNA interaction, which is important for mRNA selection for translation. In plants, the phosphorylation of ribosomal protein S6 is observed in several instances. In *Arabidopsis*, heat shock results in S6 phosphorylation by heat inactivation of the S6 kinase. Presence of 5′-cap structure and a 3′-poly(A) tail promotes a synergistic enhancement of translation of mRNA in plants and yeast cell protoplast. These observations suggest that interaction between the 5′-and 3′-termini causes circularization of actively translated mRNAs. In plants, 5′ to 3′-interaction via PABP and cap-binding complex might be further promoted by elF4B because this factor stimulates binding of PABP to poly(A). The 5′ to 3′-interaction might promote the selection of mRNA for initiation, ribosome cycling, and/or to ensure that the 5′ end is correctly capped and transcripts are polyadenylated.

1.6.2 Elongation of Polypeptide Chain

Elongation of polypeptide chain is carried out by various factors, that is, elongation factors (eEF1A, eEF1B, and eEF2), different aminoacyl-tRNAs, GTP, and ribosomes. The elongation process in plant is very much similar to what is known for other eukaryotes and prokaryotes. The traditional role of eEF1 is to bring the correct aminoacyl tRNA to the A site on the ribosome in preparation for the next peptide bond to be formed during elongation. Plant eEF1, like in other eukaryotes, is composed of four nonidentical subunits (p52, p48, p36, and p34). The eEF1 has been isolated from a variety of plants. The p52, p48, p36, and p34 subunits are known as *eEF1A* (*EF-1α*), *eEF1Bγ*, *eEF1Bβ*, *and eEF1Bα* respectively. eEF1A is one of the most abundant proteins in the eukaryotic cell representing up to 5% of the soluble protein and has a very important function in making the elongation complex. The eEF1A is a member of GTPase super family and hydrolyzes GTP with the help of eEF1B. The amino acid sequence of this factor is highly conserved (>95%) between plant species. The expression of the gene responsible for encoding eEF1A is highest in plant tissues with rapid growth such as in meristem, root tips, young leaves, and developing ovules.

The eEF2 elongation factor catalyzes the GTP-dependent translocation of the peptidyl-tRNA from the A site to the P

site of the ribosome, resulting in the movement of the mRNA three nucleotide toward the 3′ end and release of deacylated tRNA from P site. eEF2 is also a member of the GTPase superfamily but in contrast to eEF1A, this does not need a recycling factor for the exchange of GDP to GTP. The most important feature of the eEF2 is that it is posttranslationally modified at His715 to yield a unique residue, diphthamide, found only in eEF2. This unique residue is the target for specific and covalent attachment of ADP-ribose to eEF2 for several bacterial toxins, which inactivates the eEF2 and leads to cell death. The phosphorylation of eEF2 by eEF2 kinase leads to inhibition of eEF2 activity.

1.6.3 Termination of the Polypeptide Synthesis

The process of release of the polypeptide chain from the ribosome begins when the ribosome encounters one of the three termination codons in the A site. A release factor binds to the ribosome, the peptidyl-tRNA is hydrolyzed, and the nascent polypeptide is released. The hydrolysis of GTP is necessary for the termination, although the exact point of hydrolysis in the process is not known. In prokaryotes, release factors RF1 and RF2 are codon specific and require the presence of RF3, which is needed for hydrolysis of GTP. The release factors have been identified but not purified in plants as yet.

1.7 Expression of Plant Organelle Genome

In plants, organelles such as chloroplast and mitochondria contain their own genome, which encodes polypeptides associated with the functioning of the respective genome. Both the genomes are semiautonomous, that is, most of the polypeptides necessary for the structure and function of the organelles are encoded by the nuclear genome. We will present an overview of the expression of chloroplast and mitochondrial genomes separately in the following section.

1.7.1 Expression of Chloroplast Genome

Chloroplast genome of higher plants is a circular molecule of 120 to 160 kbp and contains about 130 genes. About half of these gene products are involved in chloroplast genome expression (transcription/translation) and about 30 encode for the subunits of the five photosynthesis complexes that is photosystem I (PSI), photosystem II (PSII), cytochrome b6f complex, ATP synthase, and ribulose bisphosphate carboxylase/oxygenase complex. There are large intergenic regions in chloroplast genome and about 20 to 21 ORFs encode polypeptides of unknown functions. The parasitic plant, *Epifagus virginiana* has the smallest chloroplast genome (approx 71 kbp) since it does not contain genes related to photosynthesis. Unlike chloroplast genes of algal species, most chloroplast genes in higher plants are organized in clusters and are cotranscribed as polycistronic pre-RNA, which are then extensively processed into mono- and oligo-cistronic RNAs.

The transcription–translation machinery of chloroplast has a number of prokaryote-like features. For example, chloroplast ribosomes are 70S in size, chloroplast rRNAs and tRNAs are very similar in sequence to *E. coli* counterparts, and chloroplast mRNAs contain triphosphates at their 5′ ends and lack poly(A) tail. Most of the components of chloroplast transcription–translation machinery are not encoded by chloroplasts but are instead nuclear encoded. Therefore, information of the chloroplast genetic

system requires the coordinated expression of nuclear and chloroplastic genes. The upstream region of many transcription initiation sites on chloroplast DNA contains DNA sequence similar to *E. coli* "−10" and "−35" consensus sequences. These consensus (promoter) sequences are recognized by the plastid-encoded plastid (PEP) RNA polymerase and contain two conserved hexameric sequences, TTGACA and TATAAT with 17 to 19 nucleotides in between them. Analyses using genes with deleted or mutated promoter regions and *in vitro* transcription system have confirmed that "−10" and "−35"-like sequences are required for proper transcription of a number of genes. A DNA dependent RNA polymerase catalyzes the transcription of chloroplast gene, which is distinct from nuclear RNA polymerase. Multiple RNA polymerase activities have been found in chloroplasts. For example, in mustard chloroplasts, two distinct RNA polymerase activities (peak A and B) have been identified, both of which correspond to large multisubunit complexes. The genes encoding the core α, β, and β'-subunit homolog of the plastid RNA polymerase are encoded by plastid genome. On the other hand, several genes including some tRNA genes lack the usual prokaryotic promoter sequences in the region upstream of their transcription initiation sites, for example, the *rps16* gene possesses the "−10" motif but "−35" motif is missing. Other than "−10" and "−35" sequence NEP promoters (for nuclear-encoded chloroplastic RNA polymerases) are also reported. These promoters have a short, conserved sequence around the transcription initiation site. Multiple transcription initiation sites are known especially for the chloroplast genes *psbD, atpB,* and 16S rDNA. On the basis of biochemical evidence, it is now known that promoter recognition specific to PEP is conferred by three σ-like factors of 67, 52, and 29 kDa. These factors are nuclear encoded. In addition to PEP, there are nuclear-encoded RNA polymerases (NEP) in chloroplast. At least two NEPs are directed to the chloroplast in *Arabidopsis*. The NEPs resemble the phage-type single subunit enzyme related to the mitochondrial RNA polymerase. Thus, either chloroplast-encoded or nuclear-encoded RNA polymerases or both contribute to the transcription of individual genes. In general, it appears that the nuclear-encoded RNA polymerase transcribes genes for the chloroplast genetic system during the early stage of plastid development and subsequently transcription of photosynthesis-related genes is carried out by chloroplast-encoded RNA polymerase.

Though several chloroplast genes are monocistronically transcribed, most of the chloroplast genes of higher plants are co-transcribed. Polycistronic pre-RNA thus obtained consists of message for proteins or rRNA with related functions such as photosynthesis, transcription, and translation. Chloroplast polycistronic pre-RNAs are generally processed into many overlapping shorter RNA species. For example, *psbB* operon is transcribed from a bacterial type promoter as a polycistronic precursor (consisting of *psbB-psbT-psbH-petB-petD*). Two cis-splicing and multiple cleaving events take place to generate shorter RNA species from the polycistronic RNA. Many intermediate transcripts are generated during this process; at least 17 different species in spinach and 20 in maize have been observed. Processing ultimately generates monocistronic transcripts for *psbB* and *psbH* and dicistronic/monocistronic transcripts for petB-petD. It has been shown that processing often removes secondary structures that hinder translation of

genes when present in unprocessed polycistronic transcript, for example, psbA in *Arabidopsis*. A set of transcripts in chloroplast may also contain multiple 5′ ends. A different 5′ end for the same mRNA in the chloroplast provides a mechanism for translational control and differential stabilization of mRNA. In barley, an alternative transcript for rbcL accumulates, which has additional 35 nucleotides at the 5′ end. The additional sequence has high homology to the complementary sequence of the 3′ end of 16S rRNA. It is possible that formation of intermolecular complex between *rbcL* and 16S rRNA could hinder the translation of the *rbcL*.

Several chloroplast genes are interrupted by introns. In the tobacco chloroplast genome, 18 genes contain introns of 503 to 2526 bp in length, 6 for tRNAs and 12 for protein coding genes. Most introns can be folded into a characteristic, evolutionarily conserved secondary structure similar to that of the group II intron class. Chloroplast group II-type intron have conserved boundary sequences (5′-GTGYGRY···RYCNAYYYYTAY-3′), which are similar in part to those of nuclear pre-RNA introns. The intron of the gene encoding $tRNA^{Leu}$ (UAA) can be classified into the group I-type intron of fungal mitochondrial genes. In higher plant chloroplast, $tRNA^{Leu}$(UAA) is the only gene harboring a group I-type intron, but attempts to demonstrate self-splicing of its RNA *in vitro* have so far been unsuccessful. No self-splicing has been observed in group II-type introns. It is possible that RNA-protein interaction as in the case of nuclear transcripts is essential for splicing of chloroplastic transcripts.

Higher plant organelle mRNAs can be subjected to another posttranscriptional processing known as *RNA editing*. In chloroplast, it was first described with respect to the transcript of maize *rpl2* gene in which the AUG (Met) start codon is created by a C to U transition from ACG (Thr) codon. RNA editing in plastids seems to be an early step of pre-mRNA processing, which takes place independent of splicing and cleavage of polycistronic and monocistronic mRNAs. Transcripts, which are edited in one species, may contain the correct nucleotide at the DNA level in another species. For example, the *psbF* mRNA is edited in spinach plastids, changing a serine to a phenylalanine codon, whereas in tobacco a phenylalanine codon already exists at the DNA level. The simplest mechanism for RNA editing seems to be deamination of the C4 amide of cytosine to directly form a uracil residue; alternatively, a cytosine deletion and uridine insertion mechanism would accomplish the same task. The molecular mechanism that specifies the bases to be edited is still unknown, but flanking sequences are clearly important in the selection of editing sites.

The steady state level of chloroplast RNAs varies substantially for different genes, during developmental stages and in response to environmental stimuli, and fluctuates during circadian rhythms. In general, mRNAs encoding photosynthesis-related proteins accumulate at higher levels in chloroplasts from green tissues or during chloroplast development while levels are very low in plastids of nonphotosynthetic tissues, such as amyloplasts in roots and chromoplasts in fruits. The steady state level of chloroplast transcripts is regulated either by transcriptional activity or by stability of the particular transcript and is greatly influenced by light. It has been shown that often transcripts marked for degradation in the chloroplast are polyadenylated. Thus, polyadenylation of

transcripts renders them unstable. Thus, steady state levels of a particular transcript are regulated by factors governing the stability of the transcript rather than transcription rate in chloroplasts.

The stability of chloroplast RNAs varies from RNA species to species, depending on chloroplast development and environmental conditions. Most of the chloroplastic mRNAs contain invert repeat region in their 3′ UTR, which can fold to give stem-loop structure. These invert repeats were initially known to function as transcription terminators as in prokaryotes. However, it has been demonstrated that these structures probably act as processing signals rather than termination signals. Invert repeat-containing pre-RNA are processed in a 3′ to 5′ direction by 3′ exonucleases present in the transcription system, generating nearly homogeneous 3′ ends distal to the invert repeats. The processed 3′ ends form stem-loop structures, which bind to different nuclear-encoded polypeptides to provide stability to the transcript. These nuclear factors, also known as *RNA-binding proteins* (RNPs) are involved in splicing and/or processing of pre-RNA. In tobacco, five such proteins were isolated, which contain four domains necessary for binding to stem loop. These domains are a transit peptide, an acidic amino terminal domain and two conserved RNA binding domains. The RNA binding domain of approximately 80 amino acids belong to a superfamily, called *consensus sequence-type RNA binding domain* (CS-RBD) and is responsible for RNA binding. The nuclear-encoded RNA-binding proteins have been isolated from many different plants in the last decade and their role in binding to 5′ or 3′ UTR has been established. These polypeptides not only provide stability to transcripts after binding to 3′ UTR but also work as translational activators after binding to 5′ UTR of different genes present in chloroplast.

1.7.2 Expression of Mitochondrial Genome

Plant mitochondrial genomes are very large compared to human and yeast mitochondrial genome and extremely variable in size. The size ranges from 200 kbp in *Solanum tuberosum* to 2400 kbp in *Cucumis melo*. Much of the large size is the result of noncoding sequences that are present at multiple places in the genome. The mitochondrial genomes are much more complex compared to the chloroplast genome. The complexity of the plant mitochondrial genome can be represented by a single large circle designated the ''Master Circle''. The master circle contains repeated regions (in direct and/or inverted orientation) that are active in recombination. Recombination between direct repeats results in deletions. The deletions result in the production of ''subgenomic'' molecules. The abundance of each subgenomic molecule varies. Many appear to exist in very low copy number, and are called *sublimons*. Nuclear genes presumably play some role in regulating the relative abundance of various mitochondrial DNA configurations. *Zea mays*, has five different sets of repeats that are active in recombination. This results in a large number of different subgenomic molecules. If these contain active repeats, they can undergo further intra- and intermolecular recombinations. The result is a complex multipartite organization in which the same gene may be present in several different contexts. Given all this, it is not surprising that there is little or no conservation of gene order or physical maps among the mitochondrial genomes of the different plant species.

Even within a species, there is little or no conservation. Variation in the number of genes expressed in the mitochondria has been observed frequently. The *Arabidopsis thaliana* mt-genome of 366 924 nucleotides encodes 57 identified genes only. These are components of respiratory complexes, rRNA and tRNA genes, rRNA proteins, a group of proteins involved in cytochrome *c* biogenesis and 3 orfs common to most plant mitochondrial genomes. The genome of *Marchantia polymorpha* (the liverwort – a nonvascular plant) encodes more polypeptides than that of *Arabidopsis* – particularly ribosomal proteins. Only one-third of the mitochondrial genome has been shown to stably transcribe. The remaining appears to be nonessential, partly derived from chloroplast, partly from nucleus, and in part from the duplication of mitochondrial sequences. Extensive recombination has been shown to occur in mitochondria resulting in a lot of variation in the intergenic regions of the mitochondrial genome.

Most of the essential genes of mitochondria are expressed as monocistronic transcripts from individual promoters, for example, the *atp6*, *atp9*, *cytb*, *cox1*, *cox2* and *cox3* in most plants. The transcripts of most mitochondrial genes are found to be generally much larger than the actual coding regions of the genes. The change in size is because the transcripts contain extended noncoding 5′- and 3′-transcribed regions of several hundred nucleotides. Genes located closely show cotranscription invariably, such as 18S-5S-RNA genes or the *nad3-rps12* transcription unit. The transcribed spacer between 18S-5S rRNA varies among plants, ranging from just over 100 to over 500 nucleotides.

Inspection of the sequences at transcription initiation sites revealed several conserved sequences surrounding the first transcribed nucleotide. The minimal plant mitochondrial promoter (about 18 nucleotides in length) exhibits a short loose consensus sequence (CRTA) overlapping the transcription initiation site. *In vitro* experiments confirm the presence of a set of different types of promoter structure in plant mitochondria. Some of the promoters for protein coding genes differ from the consensus CRTA motifs, for example, ribosomal rRNA genes appear to be transcribed from a different type of recognition sequence in both monocots and dicots. In *Zea perennis*, such a nonconsensus promoter region was dependent on a nuclear gene product encoded by MCT gene, confirming that indeed distinct transcription cofactors act individually on these alternative promoter sequences. The mt-genome does not possess any gene for RNA polymerase; instead, a nuclear gene encoded polymerase is required. Recently, genes encoding mitochondrial-targeted, single subunit, phage T-7 type polymerases have been shown to be directed to mitochondria in *Arabidopsis* and maize.

The steady state level of mitochondrial transcripts may be regulated at the transcriptional level for the individual gene and/or by differential stability of particular transcript. While the higher transcriptional activity of the rRNA genes results in higher abundance of steady state RNA, the transcriptional activity of the protein-encoding genes may not correlate with the amount of the respective transcripts in the mitochondrial *in vivo* RNA population. RNA stability thus appears to play a major role in regulating steady state levels of RNA.

Several of the 5′-termini mapped for mRNAs as well as rRNAs in plant mitochondria cannot be labelled by *in vitro* capping experiments and may thus be derived from 5′-processing events. Some of

the *atp6* mRNA in maize mitochondria, for example, were found to initiate translation *in vitro*, although these could not be cap labelled. This suggests that inability to cap the RNA does not reflect that the processing has generated these. The 3′ termini of the plant mitochondrial mRNAs are not polyadenylated and thus require other means of sequence stabilization. Extensive 3′ processing is required because additional noncoding regions are usually included in the transcripts. Studies on many of the genes characterized in plant mitochondria suggest that 3′ UTR of mRNAs can fold into either single or double stem-loop structures. The secondary structure is most likely associated with the formation and/or stabilization of this RNA end but their role in transcriptional termination has not been ruled out.

RNA editing is a characteristic feature of plant mitochondrial genes. RNA editing modifies virtually every mitochondrial mRNA by C to U and also U to C modifications, which is required to generate correct open reading frames. RNA editing is thus an important regulatory step for the expression of genes. In wheat mitochondria, 19 mRNAs have been found edited at 364 positions. The least edited transcript to date is nad5 (5.5 editing sites per 1000 nucleotides), whereas the most edited is orf206 involved in the biogenesis of *c*-type cytochrome, (67.9 editing sites in 1000 nucleotides). Recombination studies suggest that sequences 5′ to an editing site are important for correct editing process. Analysis of other editing sites in the recombined maize *rps12* gene showed that at least 20 nucleotides 5′ to an editing site are required to specify editing. Analysis of RNA editing following the electroporation of mutated wheat mitochondrial *coxII* gene constructs into isolated mitochondria demonstrated that 16 5′ flanking nucleotides and six 3′ flanking nucleotides are required to specify an editing site. In general, at least 25 nucleotides 5′ to the editing site are critical to specify editing and some sites appear to have additional requirements for a limited number of 3′ nucleotides. While partially edited transcripts are sometimes translated in plant mitochondria, the corresponding proteins may fail to accumulate or accumulate but fail to assemble into protein complexes.

In summary, the transcription of plant genes is carried out by three different RNA polymerases whose activities are modulated by a diverse group of cis-acting regulatory sequences. RNA polymerase II is responsible for the synthesis of pre-mRNA. General transcription factors assemble into the initiation complex at the TATA box, which lies within 100 bp of the transcription initiation site of the gene. Pre-RNA transcripts must be processed by splicing, capping, and addition of poly(A) tail to produce a mature transcript, which must exit the nucleus to initiate translation in the cytosol. The chloroplast and mitochondrial gene expression is different from nuclear genes and requires completely different transcriptional and translational machinery.

2 Light Regulation of Gene Expression

Plants being sessile organisms adapt to various environmental stimuli to sustain their development. Light is probably the most important of these and forms the basis of plant and animal life. Almost every aspect of plant development is regulated by light and appears sensitive not just to light quality but also to light quantity and duration. Light drives photosynthetic reactions and these help in converting light

energy to chemical energy in the form of ATP and also provide reducing power for the synthesis of sugars, lipids, and amino acids. A byproduct of the reactions is evolution of oxygen, which sustains life. The fine-tuning of the response to light has been achieved by evolution of several light receptors that sense the different wavelengths and intensity. These include the phytochromes (for red and far red light), the cryptochromes, for blue light and UV light, and the phototropins. These photoreceptors help in mediating responses that lead to seed germination, hypocotyl elongation, growth, photosynthesis, photoperiodism, flowering, and so on. We shall be describing the role of these receptors in the regulation of gene expression in the following sections.

2.1 Phytochromes

These are homodimers of 124 kDa each containing a phytochromobilin as chromophore and a histidine kinase domain. They exist in two inter convertible forms – a Pr (inactive) form that absorbs red light and gets converted to Pfr (active) form by photoconversion and vice versa by absorption of far red light. At least five phytochromes (A–E) are known in *Arabidopsis,* which mediate responses to different wavelengths such as red, far red, the fluence of light (low and high irradiance), and the ratios of the red/far red lights. Phytochromes are present in the cytosol in their inactive (Pr) forms but are translocated to the nucleus upon photoconversion to their active (Pfr) forms. In the nucleus, they are localized to speckles where they are believed to interact with other nuclear proteins responsible for transcription. While phytochrome A (phyA) mediates far red light responses, phytochrome B (phyB) brings about responses to red light. Of all phytochromes, only phytochrome A is light labile.

2.1.1 Gene Regulation by PhyA Signaling

The phyA regulation of gene expression is most widely studied. PhyA is responsible for the high irradiance response of phytochrome action and helps in seedling de-etiolation, hypocotyl elongation, expansion of cotyledons, anthocyanin formation as well as perception of day length. Recent experiments have shown that the initial far red light responses mediated by phytochrome A probably occurs through activation of a group of master transcription factors within one hour of light sensing. These master regulators in turn activate a diverse group of transcription factors that result in extensive branching out of the response downstream leading to varied changes. Indeed, 10% of the total *Arabidopsis* genes are activated by phytochrome A. Of these about 50% are late acting and related to photosynthesis and general metabolism while about 44% are early responsive (activated within 1 h of light treatment) and are basically transcription factor genes belonging to different classes such as the zinc finger proteins, bZIP factors, basic helix–loop–helix proteins, Myb domain proteins, AP2 domain proteins, and so on. Among the main transcription factors, genes activated by phyA are the *PIF1* (*PHYTOCHROME INTERACTING FACTOR 1*), *HY5* (*HYPOCOTYL5*), *LAF1* (*LONG AFTER FAR RED LIGHT 1*), *TOC1 (TIMING OF CAB EXPRESSION 1), CCA1 (CIRCADIAN CLOCK ASSOCIATED 1), LHY (LATE ELONGATED HYPOCOTYL), CO (CONSTANS)*, and so on. Of these, PIF1, a bHLH protein, can bind both PHYA and PHYB and activate a subset of G-box (CACGTG) containing genes. The G-box has been found in

the promoters of several light responsive genes. HY5 is a bZIP type of factor that activates transcription of genes mainly in response to far red light and can bind the G-box in some light responsive genes such as *CHS* (*CHALCONE SYNTHASE*, involved in anthocyanin formation). Another transcription factor HYH, also a bZIP factor, can form heterodimers with HY5 *in vivo* but functions more in mediating blue light responses from cryptochromes. HFR1, a bHLH protein, is responsible for accumulation of *CAB* (CHLOROPHYLL AB BINDING PROTEIN) and *RBCS* (encoding the small subunit of RUBP carboxylase) transcripts. CCA1, LHY and TOC1, also activated by PhyA, are factors that regulate the circadian clock, while CO is known to be involved in flowering response.

2.1.2 Gene Regulation by PhyB Signaling

PhyB mainly senses red light and is responsible for red light mediated effects. Its main responses include the low fluence response promotion of seedling germination, inhibition of hypocotyl elongation, shade avoidance response (in which plants sense shade due to a low red:far red ratio and elongate to overcome the shade), inhibition of flowering, and so on. Signaling by phyB is mediated by PIF3, which binds to phyB in a light-dependent fashion. The other phytochromes (C, D, and E) also play a role in shade avoidance response and in flowering.

2.2 Cryptochromes

These are blue light-sensing flavoproteins with similarity to photolyases but without the DNA repair activity of photolyases. Two cryptochromes, CRY1 and CRY2 have been identified and are localized to the nucleus. The *c*-terminal portions of both the proteins are required for light signaling. The *c*-terminal can also interact with COP1 (a key negative regulator of light responses) and repress COP1. Cryptochromes have been shown to activate almost the same number of genes as red light and activate several transcription factors such as HYH. A ser/thr phosphatase has been shown to positively regulate blue light signal. While in animals cryptochromes form the central oscillator of the circadian clock, in plants they are responsible for light input to the clock. PhyA is also required for some cryptochrome responses.

2.3 Phototropins

These include PHOT1 and PHOT2 and are blue light and UVA sensing receptor kinases that control phototropism of stem and roots as well as chloroplast movement and stomatal opening. The phototropin signal is mediated possibly by NPH3 (NON-PHOTOTROPIC HYPOCOTYL 3) and RPT2 (ROOT PHOTOTROPISM2) since both contain BTB/POZ domains that are known to interact with transcription factors. Another transcription factor activated by these receptors is NPH4/ARF7 – an auxin response factor, indicating an interaction between phototropins and auxin signaling.

2.4 COP1 – A Major Regulator of Light Responses

In recent years, it has become increasingly apparent that plant responses to light and hormonal stimuli are regulated posttranscriptionally by protein degradation. The first evidence of the role of selective protein degradation came from the isolation of

the COP1 – an E3 ubiquitin ligase-like protein containing RING finger domains and WD repeats. COP1 localizes to the nucleus in the dark where it binds to transcription factors involved in light signaling such as HY5, HYH, and LAF, and promotes their ubiquitination and degradation via the proteasome. This prevents expression of light inducible genes in the dark. COP1 can also interact with CRY2, PHYB, and SPA1 and these interactions are believed to modulate its activity. In light, COP1 moves out of the nucleus into the cytosol while phytochromes move in. This allows light responsive transcription factors to transcribe genes. COP1 thus negatively regulates light responses and prevents light responses in the dark. Microarray analyzes have shown that expression of almost 30% of all *Arabidopsis* genes is influenced by COP1. COP1 interacts with the COP9 signalosome – believed to be the regulatory 19S subunit lid of the 26S proteasome. The COP9 signalosome has been shown recently to govern the degradation of regulatory factors involved in several responses such as those of auxins, JA (jasmonic acid), ABA (abscisic acid), gibberellins, flowering initiation, disease resistance, and so on.

2.5 Light Mediated Changes in Plastids

The major change that occurs upon receiving light is the conversion of the etioplasts to active chloroplasts. Phytochromes as well as cryptochromes are involved in this process of transformation. Absorption of light by the chlorophylls and carotenoids drives the photosynthetic electron transport and generates reducing power for synthesis of sugars. Since the whole process is sensitive to light, both nuclear and plastid genes have to be regulated in light. Most genes of the plastid, especially those involved in the light reactions, show a marked light induced upregulation of transcription or a light mediated increase in translation or stability of transcripts. Several light responsive elements such as the G-box, GT box, and so on, have been identified in the promoters of light responsive genes that bind accessory factors to drive light mediated transcription. Light quality and intensity also play a major role in regulating gene expression, for example, the psbD promoter has been shown to contain a light responsive element that is sensitive to red and blue light and responds to signals from both the cryptochromes and phytochrome A. Gene expression of light responsive genes is also regulated posttranscriptionally either through differential stability of transcripts or through differential translation.

2.6 Light Regulation of the Circadian Clock

The circadian clock of plants helps the plant in anticipating and preparing for the arrival of dusk and dawn and is thus closely linked to light input from photoreceptors. Light input from the photoreceptors (PhyA, PhyB, the cryptochromes, and possibly ZTL/FKF1/LKP1) entrains the clock with the environment making it sensitive to small changes in light duration. The clock itself consists of a feedback loop consisting of components that cycle between day and night. The central oscillator consists of two myb domain transcription factor genes *CCA1* and *LHY* along with *TOC1* and *ELF4* as primary components of the clock. *TOC1* is a gene that is responsible for activation of the two myb genes *CCA1* and *LHY*. These two genes in turn repress the transcription of the *TOC1* gene by binding on to an "evening"

element (AAAATATCT) in its promoter. This results in cycling of the *TOC1* mRNA in such a way that *TOC1* mRNA and expression are low during the day but begin increasing late in the day. When levels of TOC1 increase, it begins transcribing the *CCA1* and *LHY* genes causing a peak increase in their levels in the morning. This represses *TOC1* during the day, which in turn affects (reduces) CCA1 and LHY levels in the evening. This results in a strong circadian rhythm where the expression of the *CCA1* and *LHY* are out of phase with *TOC1*. In addition, an F-box containing gene *ZTL* (*ZEITLUPE*) has been shown to be responsible for the degradation of TOC1 at night, thus creating a robust rhythm. Mutations in all the clock genes result in early flowering. The circadian clock has been shown to regulate the expression of several genes related to photosynthesis and pigment biosynthesis, which peak just before dawn in anticipation of light and those of lipid modification peak just before dusk and help in altering the desaturation levels of the lipids. Regulation of many of these genes involves the "evening" element in the promoter of these genes that are regulated by CCA1. Entrainment of the clock with the environment is brought about by the photoreceptors via PIF3 (binding PhyB), and other light sensing domain containing proteins such as ZTL, FKF1. Genes such as *ELF3* (*EARLY FLOWERING 3*) and *GI* (*GIGANTIA*) help in mediating the light input to the clock. PIF3 has been shown to upregulate both *CCA1* and *LHY*.

2.7 Light Regulation of Flowering Time

Another major effect of light is in determining the flowering time in some plants. This is known as *photoperiodism* and is based on the duration of light that triggers flowering. Plants can be classified as short day plants (SDPs, where flowering is initiated only below a critical day length), long day plants (LDPs, where flowering is initiated only above a critical day length) and day neutral plants that flower independent of the duration of day length. The ability to flower in a given day length is closely linked to the plant physiology and its adaptability to environmental conditions in that season that maximize the chances for proper seed set and survival.

At the genetic level, there are several genes involved in determining flowering time and floral initiation. The expression of the floral organ determinants in LDPs and SDPs is closely associated with light input to the circadian clock and regulation of its components by the photoreceptors, light quality, temperature. The circadian clock of the plant measures the time and day length and under favorable conditions genes responsible for initiating flowering are activated.

Phytochrome A and cryptochrome 2 are believed to promote flowering by activating floral pathway integrators and repressing phyB. On the other hand, phytochrome B as well as phyD and phyE repress flowering, with these effects occurring in a temperature-dependent fashion. PhyB repression is predominant at 22 °C while that of phyE is predominant at 16 °C.

In the last couple of years, the gene *CONSTANS* has been shown to be a major determinant in initiating flowering in both short day plants and long day plants. *CONSTANS* encodes a transcription factor with two B-box zinc finger domains and can directly activate *FT* (*FLOWERING LOCUS T*). Levels of *CONSTANS* transcript undergo a circadian rhythm and begin increasing during the day and reach

a peak at the end of day. However, levels of the protein begin accumulating only late in the evening during LDs. Another gene *FKF1* (*FLAVIN BINDING, KELCH REPEAT F-BOX CONTAINING*) has been shown to be responsible at least partly for the high levels of *CONSTANS* mRNA in light. By performing some elegant experiments with the use of mutants for flowering time as well as circadian clock it has been proposed that in long day plants such as *Arabidopsis*, the expression of *CONSTANS* mRNA and protein reaches a maximum at the end of the day and the coincidence of this expression with the light period in long days allows it to activate the flowering time gene FT. On short days, the peak of CONSTANS expression occurs in the dark and it is unable to activate FT. Moreover, there is a dark induced degradation of the CONSTANS protein by the proteasome brought about by ZTL. In light, the stability of the CONSTANS has been shown to be regulated by phytochrome A and cryptochrome 2, which stabilize it at the end of the day while phytochrome B is responsible for its instability and low levels during the mornings. These effects on stability are independent of the mRNA accumulation. Interestingly, in short day plants such as rice, the *CONSTANS* homolog *HD1* (*HEADING DATE1*) also shows a circadian peak with similar expression patterns, yet it brings about opposite effects as far as flowering is concerned. Experiments suggest that in short day plants, during the light phase, *CONSTANS* expression represses the expression of a transcription factor responsible for activation of *HD3a* (*HEADING DATE 3A*- or the homolog of FT in rice). This light mediated repression by CONSTANS of the transcription factor is activated by phytochrome. At night, expression of CONSTANS is high. This, along with the fact that phytochromes no longer activate it at night results in lack of repression of *HD3a* transcription and therefore the initiation of flowering. On long days, the CONSTANS protein remains activated by phytochrome and continues to repress transcription of *HD3a* resulting in no flowering.

2.8 Interaction of the Light Signal with Hormonal Pathways

Light signaling also interacts with other hormonal pathways to bring about plant development. Auxin has been shown to be involved in hypocotyl elongation and light regulates the expression of many auxin inducible genes. In fact, *IAA1, IAA9, PsIAA4,* and *AXR3*, have been shown to be phosphorylated by Phytochrome A. Degradation of some auxin regulated products occurs by the COP9 signalosome. Recently, a suppressor of *COP1, fin219,* was also shown to encode an auxin inducible gene. PhyB is known to promote seedling germination by activating gibberellin (GA) biosynthesis by activation of GA 3 β-hydroxylase. GA synthesis is responsible for the inhibition of hypocotyl elongation. Light also interacts with ethylene in mediating seedling development. Ethylene inhibits hypocotyl elongation in the dark but promotes it in light. Cytokinins are required during the greening process and application of cytokinins to plants grown in darkness causes de-etiolation, development of chloroplasts, and the expression of many light-regulated genes.

Light (both quality and quantity) thus affects several processes such as chloroplast biogenesis, entrainment of the circadian clock, flowering by photoperiodic responses, and general plant development

as well as several other aspects (not covered in this article) such as photosynthetic carbon fixation, carbon metabolism (interconversion of starch and sugars), nitrogen assimilation and so on. The responses and gene regulation are brought about by the action of different photoreceptors and their interaction with other hormone-signaling pathways as well as indirectly by the byproducts of photosynthesis such as reducing equivalents, sugars, ATP, and so on.

2.9 Nuclear-organelle Interactions

As has been previously described, plant cells contain chloroplasts and mitochondria, each with their own genomes. They perform the important functions of photosynthesis, lipid, and amino acid biosynthesis, respiration, and so on that are absolutely essential for the life of the plant. Large membrane-bound multiprotein complexes that contain polypeptides synthesized by both nucleus and the organelles bring about these processes. Almost 80 to 90% of the proteins located in the organelles are imported from the nucleus. Many genes, especially in the chloroplast, are transcribed as multicistronic mRNAs forming subunits of more than one complex that are differentially regulated by the environment or even showing tissue-specific expression as in the case of the *psbB* operonic genes. Gene expression of the organelles thus has to be tightly regulated so as to maintain the stoichiometry of the subunits in the multisubunit complexes. This is possible through a close coordination and communication between the nucleus and the organelles. Much of the transcription and translation of the organellar genes is controlled by the nucleus. The nucleus encodes at least three organelle directed RNA polymerases. One each is targeted at the chloroplast and the mitochondria respectively while the third has been shown to be directed to both the organelles. The three RNAPs are active at different times of plant development and undergo transcriptional regulation.

Besides synthesizing subunits for the photosynthetic and respiratory complexes of the organelles, the nucleus also exerts a great deal of control on the expression of the organellar genes and thus on their biogenesis. In recent years, many nuclear genes that are specifically required for the transcription, RNA processing, and stability, and translation of the organellar genes have been identified and characterized. Many of the regulatory genes are involved in gene-specific mRNA processing (both at the 5′ and 3′ ends) and translation such as in regulation of the *psbB* operon. The nuclear gene products of *NAC2, MBB1/HCF107, MCD-1* are necessary for processing, stability, and/or translation of the *psbD, psbB/psbH* and *petD* gene transcripts of the chloroplasts respectively, by acting at the 5′-end of the transcripts. Similarly, RB60 (a protein disulfide isomerase) and RB47 (a polyA binding protein member) are proteins that have been shown to bind the 5′ UTR of psbA and influence its translation. Secondary structures at the 3′ end of the plastid transcripts may also bind nuclear factors, for example, CSP41, CSP55, and CSP29 proteins have been shown to bind the stem loop and AU-rich element at the 3′ end of the *petD* pre-mRNA transcript in spinach. CSP41 has both RNA binding and endonuclease activities. Some gene products may control more than one process, for example, CRP1 seems to be independently responsible for *petD* transcript processing and *petA* translation in maize while the *HCF109*

in *Arabidopsis*, encodes a termination release factor that affects the expression of many plastid genes. Although the nucleus governs the expression of the bulk of the organellar proteins, there has to be compatibility between the nuclear and organellar genomes for a proper functional interaction to occur. Interactions between the whole nuclear plastid genomes have been studied in the subgenus *Euoenothera* of *Oenothera* and it has been shown that of the three nuclear and five plastid genomes that exist in this subgenus, only a few combinations exist in nature. Other combinations can be very disruptive leading to drastic dysfunction of the plastids resulting in nongreen tissues.

The effect of nuclear genes on mitochondrial gene expression has been studied mainly through the use of cytoplasmic male sterility (CMS) phenotype. CMS is a defect wherein plants are unable to produce functional pollen and are consequently male sterile. It has been shown that CMS results from incompatible nuclear-mitochondrial interactions and inheritance of the trait is non-Mendelian and maternally inherited. In several plants, CMS has been shown to occur due to expression of chimeric genes in mitochondria. These chimeric genes involve the fusion of a known mitochondrial gene with an unknown reading frame and this results in a novel open reading frame. Examples include the *T-urf13* gene in the CMS-T cytoplasm of maize (in which the coding portion contains the 3′ noncoding portion of *rrn26*), the *pcf* gene in petunia (containing portions of the *atp9*, *coxII* and *urfS*), the *pet1* gene in sunflower, *orf224* in Brassica (containing coding portions of *atp8* and *rps3*), *orf107* in sorghum (containing portions of *atp9*), and so on. In some cases, the expression of these chimeric genes may affect the normal expression or processing of the genes adjacent to them such as in the case of *orf224* in radish, which affects *atp6* expression lying downstream to it. In most cases, however, the chimeric protein may disrupt mitochondrial functioning by itself as in the case of TURF13 in maize and PVS in *Phaseolus*. This probably disrupts the energy processes especially in tissues requiring high mitochondrial activity such as the tapetum of male gametophytes. Restoration to fertility occurs when plants are crossed with restorer lines. These lines possess fertility restorer genes that are able to suppress the expression of the CMS genes. In recent years, fertility restorer genes or their loci have been identified for several CMS lines. Many of these belong to or contain genes that are part of a novel protein family termed the *PPR (penta-tricopeptide repeat) family of proteins*, for example, the restorer genes of *pcf* in petunia, *Brassica*, radish, and rice. Members of the PPR family are characterized by tandem repeats of 35 amino acids. Interestingly, this large family of proteins in *Arabidopsis* and rice, consisting of almost 400 members, is unique to plants and only one or two members have been identified in animals. Even more strikingly, almost two-thirds of these proteins are predicted to target the organelles indicating a predominantly organelle-specific function for this family of proteins.

2.9.1 Signaling Between Nucleus and the Organelles

Just as the nucleus controls the expression of the plastids, there is also a reverse signaling that signals the nucleus about the status of the chloroplasts and regulates expression of the nuclear genes. Experiments performed with the *gun* (*genomes uncoupled*) mutants in which LHCB expression is not repressed completely even

when plastid development is prevented by photo oxidative damage indicate that the *gun* mutants are affected in genes that encode products involved in the synthesis of tetrapyroles. Thus, much of plastid to nucleus signaling probably involves the chlorophyll precursors, Mg^{2+} protoporphyrin IX and protoporphyrin IX and a subunit of the Mg^{2+} chetalase. When levels of these precursors increase, they negatively regulate expression of nuclear genes such as *CAB* and *RbcS*. Expression of plastid and nuclear genes is also sensitive to redox regulation. This is not surprising considering that photosynthetic electron transport in light causes a change in the membrane potential as well as pH. The redox poise of the plastoquinone pool probably regulates LHCII kinase and is responsible for affecting nuclear gene expression of ferredoxin, and some PSI genes.

Thus, a close interaction and compatibility between all the three genomes, that is, the nuclear, plastid, and mitochondrial genomes is very essential and tight controls are exerted to help the plant in regulating the different process in the organelles and therefore plant development.

3 Hormonal Regulation of Gene Expression

Phytohormones are low molecular weight substances produced by plants that affect many processes throughout the development of plant life and its adaptation to the environment. They act specifically at very low concentrations (10^{-6} to 10^{-8} M) and usually trigger the effect after binding to specific receptors. Their effects are specific and not interchangeable. There are five classical groups of phytohormones: auxins, cytokinins, abscisic acid, gibberellins, and ethylene. Besides the five classical hormones, there are four more groups that regulate plant growth and are now considered as hormones. These are jasmonic acid, brassinosteroids, systemin, and oligosaccharides. Salicylic acid and nitric oxide are also involved in gene regulation during different stresses but so far they have been considered only as metabolites and not as hormones. This section will give brief information about five classical hormones, namely, auxin, cytokinin, gibberellin, abscisic acid and ethylene, and jasmonic acid and brassinosteroids. The receptors that have been identified so far and genes that are responsive to these hormones including responsive elements will be discussed. Systemin will be covered in the subsection of jasmonic acid.

3.1 Regulation by Auxins

Auxins play a crucial role in most of the major growth responses during plant development. They regulate various cellular processes such as cell division, differentiation, and cell extension, as well as other visible whole plant processes such as tropism (phototropism, gravitropism), root initiation, and apical dominance. The genes that are regulated by auxins within minutes of its application are referred to as *early or primary auxin response genes*. These genes are characterized as three different classes: Aux/IAAs, SAURs (small auxin up RNAs), and GH3s.

The aux/IAA genes are exclusively found in plants and belong to a multigene family with at least 27 Aux/IAA genes identified in *Arabidopsis* and other plants. The proteins encoded by these genes are small in size (about 20–35 kDa) with four highly conserved domains. They are localized in the nucleus and have a very high turnover

rate. These genes can also be induced even after treatment with cycloheximide (a eukaryotic protein synthesis inhibitor), indicating that these are kept inactive by a rapidly turned over repressor, which is somehow inactivated by auxins. There are some evidences that AUX/IAA proteins act as a repressor for their own transcription.

The SAUR genes, as the name suggests, encode small RNAs that are induced within 2 to 5 min of auxin exposure. They lack introns and encode small proteins of 9 to 10 kDA. At least 70 SAUR genes are known in *Arabidopsis*. Regulation of these genes is at the transcriptional level by auxin. These genes are not induced by cyclohexamide (protein synthesis inhibitor) but the treatment results in an increased accumulation of transcript probably due to the stabilization of SAUR transcripts. The role of the SAUR genes is not clear.

Auxin-regulated GH3 genes, like Aux/IAA and SAUR genes, were also identified through differential hybridization technique. Protein synthesis inhibitor does not affect GH3 transcripts, like those of SAUR and Aux/IAA. In *Arabidopsis*, the GH3 gene family consists of 19 members, which are divided in three subfamilies depending upon sequence homology and splicing patterns. Most of these show auxin induction. Interestingly, members of GH3 family, play a role in photomorphogenesis also. Two mutants *fin219* (*far red insensitive219*) and *fin19* of GH3 gene family have a role in light signaling. Another mutant *dfl1-D* (*dwarf in light*) has defective DFL1-D/GH3–6. In this mutant, GH3–6 transcript is highly expressed and shows dwarf phenotypes. Different GH3 genes may have specific roles in auxin and light signaling.

Another important family of genes is the PIN-formed (PIN) family that encodes auxin efflux carriers. These carriers are responsible for the active and directional transport of auxin and are required for tropical growth, regulation of root and shoot branching, and developmental patterning.

Promoters for different auxin responsive genes from all the three families have been cloned and studied. On the basis of studies on the different promoters, a 6-bp sequence TGTCTC has been identified as auxin responsive and is known as the *auxin responsive element* (ARE). Naturally occurring AREs are functional as simple elements or as composite AREs in which the ARE is functional only in combination with coupling or constitutive element. A trans-acting auxin responsive factor (ARF), which binds the TGTCTC of ARE, has been identified. At least 27 ARFs that express in almost all tissues are known in *Arabidopsis*. The ARF transcript is not auxin inducible. The N-terminal domain of ARF proteins has a DNA binding domain with specificity to ARE. The C-terminal domain is related to domain III and IV of AUX/IAA proteins, which probably facilitate heterodimerization. The central region of ARF functions as an activation/repression domain. ARFs that have a proline or a serine rich central region act as repressors (ARF1 and ARF2), whereas ARFs containing a glutamine rich middle region function as activators (ARF6, 7, and 8).

Regulation of gene expression by auxins is complex and knowledge about all the factors is limiting. A hypothetical model for auxin regulation has been proposed and according to this, at low auxin concentration ARF activators remain associated with AUX/IAA repressor through the carboxy terminal–dimerization domain resulting in repression of auxin responsive genes. At elevated levels of auxin, AUX/IAA repressor dissociates from ARFs, is bound

by the SCF^{TIR1} complex (an E3 ubiquitin ligase) and degraded by the proteasome pathway. This dissociation and degradation of AUX/IAA allow ARFs to bind to AuxRE to activate/derepress auxin responsive genes.

3.2 Regulation by Cytokinin

Plant hormone cytokinin controls cell division, shoot meristem initiation, leaf and root differentiation, chloroplast biogenesis, stress tolerance, and senescence. Cytokinin also plays an important role in shoot regeneration in plant tissue culture.

Analyzes of mutants in *Arabidopsis* suggest that the cytokinin-signaling pathway represents a two-component signaling system similar to bacterial two-component system. It involves a hybrid histidine kinase (AHK) as the cytokinin receptor and signal from AHKs to nuclear response regulators (ARRs) is transmitted through histidine phosphotransfer proteins (AHPs, which are about 12 kDa in size). Three genes that encode cytokinin receptors (*AHK4, AHK2,* and *AHK3*) have been identified in *Arabidopsis* and all of them are activated by exogenous cytokinin treatment. Two more genes, *CKI1* and *CKI2,* also exhibit cytokinin responses but are not activated by cytokinin and probably act through different perception mechanisms. Green fluorescent protein-based experiments provide evidence that AHPs shuttle from cytoplasm to nucleus in a cytokinin dependent manner.

There are about 22 ARRs in *Arabidopsis* and these are divided into two classes; type A ARRs have a response regulator and a short N- and C-terminal end and are cytokinin inducible, whereas type B ARRs are not affected by cytokinin but act upstream of type A ARRs to activate them. The type B ARRs have a C-terminal domain with a GARP motif that is distantly related to the Myb repeat of transcription factors and probably bind DNA as a monomer at the sequence 5′(A/G)GAT(T/C)3′. The type B ARRs have a central role in cytokinin signaling and overexpression of the C-terminal domain that contains the GARP domain and an acidic domain rich in P/Q residues, mimics cytokinin responses. The A-type ARR (ARR4, ARR5, ARR6, and ARR7) are induced within 10 min of cytokinin treatment. Expression of most type A ARRs is maximum in the roots. About 823 upregulated and 917 downregulated genes have been identified in *Arabidopsis* following cytokinin application. While some cytokinin genes such as *AHK4* are expressed in the globular embryo and play a role in embryonic cell proliferation, vascular morphogenesis, and root meristem activity, other genes such as *ARR2* regulate genes that are involved in synthesis of components of mitochondrial complex I and play a role in pollen formation and sporogenesis. *ARR4* can interact with phyB and stabilize it and it also interacts with auxin responsive genes such as *AtDBP1* and *AtDBP2* indicating the involvement of cytokinin in light and auxin effects. Cytokinins also regulate expression of some ribosomal proteins like S14, S16, L13a, and L30 and hence these are probably involved in translational processes as well besides regulating at transcriptional level.

The hypothetical model for cytokinin signal transduction suggests that cytokinin signal is perceived by AHKs. AHKs initiate a signaling cascade through phosphorelay and shuttle AHP to nucleus. Activated AHPs interact with repressed ARRs in the nucleus and release their repression. The derepressed ARRs bind to multiple cis-elements in the promoters of target genes.

It has been recently suggested that the increase in endogenous cytokinin is perceived by a two-component system and transduced into the nucleus by MAP-kinase signaling and translated into a response by inducing WRKY transcription factors, which later activate downstream genes.

3.3 Regulation by Gibberellins

Gibberellins (GAs) influence the growth and development of the plant and are especially important in leaf expansion, stem elongation, seedling starch reserve hydrolysis, and flowering. The knowledge about GA function comes mainly from studies carried out in *Arabidopsis* and on the barley seed aleurone layer. Little is known about GA perception, downstream-signaling components and GA responsive genes. DNA microarray analyses have led to the identification of 138 upregulated and 120 downregulated GA responsive genes. GA has been shown to activate expression of the expansin gene in rice that causes internode expansion allowing rice plants submerged in floods to rise above the water. The α-amylase gene, responsible for starch breakdown in the aleurone layer, is another important GA-regulated gene. These, as well as other genes encoding hydrolytic enzymes, are rapidly activated in the cereal aleurone layer in response to GA. A Myb factor, GAMYB, has been shown to bind to the amylase promoter and activate its transcription as well as that of other hydrolytic enzyme genes. Another gene, *HRT* (*HORDEUM REPRESSOR OF TRANSCRIPTION*), encodes a zinc finger repressor that binds to the promoter of the amylase gene and represses transcription. The D1 gene, that encodes a putative α-subunit of the hetero-trimeric G-protein, has been shown to positively regulate GA responses.

Studies carried out on *Arabidopsis* mutants (dominant and recessive) indicate that GA signal transduction pathway is activated by inhibition of the repressor protein GAI (GA INSENSITIVE) and RGA (REPRESSOR OF *GA1-3*). These are homologs and function as nuclear transcriptional regulators. GAI and RGA possess a DELLA motif in their N-terminal domain (hence referred to as *DELLA proteins*) and belong to the large GRAS family in *Arabidopsis* that act as transcriptional regulators. Five GAI/RGA related genes (*GAI, RGA, RGL1, RGL2,* and *RGL3*) are present in *Arabidopsis*. Deletion of the 17 amino acid residues within the DELLA sequence in RGA (rga-Δ17) produce a constitutively active repressor that is not inactivated by GA. Apparently, this deletion makes mutant protein stable and resistant to GA induced degradation. Probably, the N-terminus of DELLA protein is a regulatory domain, which senses the GA signal, whereas the C-terminus functions as the repressor of GA signaling. DELLA proteins repress GA signaling in other plant species and GA-insensitive dwarf plants have been identified in maize (*D8*), rice (*slr-1*), barley (*sln1-d*), and grape (*Vvgai1*). In all these mutant species, the DELLA proteins are altered in the N-terminal region. The information on rice suggests that it contains a single DELLA protein gene and knocking out this gene results in a taller phenotype. This dramatic change, upon knocking out a DELLA protein gene, is not seen in *Arabidopsis* due to functional redundancy amongst its five DELLA members. Other repressors of GA signaling include SPINDLY (an O-linked *N*-acetylglucosamine transferase) and possibly SHORT INTERNODES while positive regulators include the *SLEEPY1*

gene, that encodes an F-box protein involved in degradation of the GA repressors. PICKLE, a CHD3 chromatin-remodeling factor, that suppresses embryonic differentiation in *Arabidopsis*, is also likely to be a positive regulator.

GA also plays an important role in flowering and this is achieved by direct activation of the floral meristem identity gene, *LEAFY*. An 8-bp cis-acting element, necessary for GA dependent expression, has been identified in the LEAFY promoter. However, this 8-bp element is not common for all GA responsive genes and some other unidentified elements regulate GA response. In potato, *PHOR1* (*PHOTOPERIOD RESPONSIVE 1*) has been identified as a GA upregulated gene that may function as an E3 ubiquitin ligase to degrade GA repressors. This gene is also regulated by photoperiodic responses and thus provides an evidence for interaction between GA and light signaling. It is interesting that many GA downregulated genes also have ABA responsive elements in their promoters indicating cross talk between GA and ABA.

3.4 Regulation by Abscisic Acid

Abscisic acid plays a role in various processes such as plant development, seed dormancy, germination, and stomata closure. It also has a major role in abiotic stresses such as drought and salt. At a cellular level, it modulates cell division and cellular responses to biotic and abiotic stresses.

Genome-wide gene expression profiling in *Arabidopsis* reveals that at least 1300 genes are either up- or downregulated by ABA. More than 90% of these genes showed impaired regulation in the *abi1-1* mutant (which encodes protein phosphatase2C) indicating an important role for this enzyme in ABA signaling. Besides transcription factors such as VIVIPAROUS/ABA INSENSITIVE3 and the bZIP and AP2 type of transcription factors that mediate ABA gene regulation, ABA also induces the expression of the LEA (LATE EMBRYOGENESIS ABUNDANT) class of proteins during seed formation and abiotic stress. These proteins are required for protection against desiccation. Many ABA-regulated genes contain ABA responsive elements (ABRE) consisting of ACGT-containing G-box. However, optimal ABA responsiveness requires a second GC rich coupling element (CE), which is similar to ABRE and approximately 30 bp downstream of the main ABRE. This coupling element can bind AP2 domain type of transcription factors. Other cis-elements that bind Myc and Myb domain proteins have also been reported.

Gene regulation during seed and early seedling development is brought about by the bZIP factor *ABA INSENSITIVE5* (with 9 nine members in *Arabidopsis*), while the bZIP factor AREB/ABF (ABRE BINDING FACTOR) acts later during development. Both are activated and stabilized by phosphorylation by SNF1-related protein kinases, ABA-INDUCED PROTEIN KINASE1 (PKABA1)/ABA ACTIVATED PROTEIN KINASE (AAPK). The factors activate transcription by binding the ABRE in the promoters of several genes. ABI5 can also interact with ABI3 to enhance transcription of some of these genes. When the ABA signal is to be attenuated or abolished, a protein ABF (ABI5 BINDING PROTEIN) binds ABI5 and targets the dephosphorylated form of ABI5 for degradation by the proteasome.

Another regulator of ABA responses is the homeodomain leucine zipper (HDZip) protein HOMEODOMAIN PROTEIN 6

(AtHB6). This protein binds to an AT rich cis-element and interacts with ABI1. Nuclear localization of this protein desensitizes guard cells and seeds to ABA. ABA also controls gene expression at the posttranscriptional level by downregulating expression of ribosomal proteins and upregulating genes involved in proteolysis. Posttranscriptional control is also evident in the form of dephosphorylation of the carboxy-terminal domain of RNA polymeraseII by the genes *AtCPL1* and *AtCPL3* (*CARBOXY-TERMINAL DOMAIN PHOSPHATE-LIKE1* and *3)*, which results in downregulation of ABA inducible genes. In contrast, phosphorylation of AKIP1 (AAPK INTERACTING PROTEIN1) by AAPK targets AKIP1 at mRNAs of dehydrin, allowing their translation in the guard cells. Another protein ABA-HYPERSENSITIVE1 (ABH1) encodes a subunit of the cap-binding complex and participates in several RNA processing events that may affect transcripts that limit ABA-signaling events.

ABA response regulators comprise G-proteins, protein phosphatases, and protein kinases of the Ca^{2+} dependent protein kinase (CDPK) and SUCROSE NON FERMENTING protein (SNF-1)-like groups. The RHO-like small G-protein ROP10 (as well as ROP2 and ROP6) negatively regulate ABA mediated stomatal closure, germination, and growth inhibition. This protein mimics the role of small G-protein RAS in the mitogenic responses of mammals. These proteins are also associated with increased H_2O_2 production because of their activation of NADPH oxidase. Other negative regulators include four protein phosphatases (PP2Cs) ABI1, ABI2, AtPP2CA, and AtP2CHA found in *Arabidopsis*. However, these are transcriptionally upregulated by ABA. ABI2 mediates its action by binding to CIPK15, which is required for ABA action, and negatively regulates the signal relay.

Cellular responses to ABA vary in different cell types and it is possible that there exist separate ABA-signaling pathways in different cells. No receptors have yet been identified for ABA. ABA regulates plasma membrane and ion channel activities and it is possible that ABA interacts directly with transport proteins or metabolic factors. Studies on stomatal closure suggest that cytosolic Ca^{2+} mediates stomatal closure and Ca^{2+} channel activation is ABA mediated. Phospholipase C (PLC) and phospholipase D (PLD) are activated during ABA signaling and generate IP3 and phosphatidic acid as second messengers. Studies on transgenic tobacco plants with reduced PLC expression in guard cells reveal impairment of ABA related activities in both mediation of stomatal closure and inhibition of stomatal opening.

The pH and redox status of a cell are also crucial in regulating ABA signal transduction. H_2O_2 and nitric oxide (NO) concentration increase before ABA-induced stomatal closure. The redox signals provided by varying concentrations of H_2O_2 and NO are characterized by short biological half-life. Studies on nitrate reductase–deficient double mutants (*nia1* and *nia2*) indicate strongly impaired NO production and stomatal closure in the presence of ABA. Results also infer that NO is active upstream of Ca^{2+} and NO and H_2O_2 activate inward rectifying Ca^{2+} channel. ABA signaling downstream of Ca^{2+} is negatively regulated by protein phosphatase ABI1 and ABI2.

3.5 Regulation by Ethylene

Ethylene affects all the major developmental processes in plants starting from

germination, flower and leaf senescence, fruit ripening, abscission to programmed cell death. It also plays an important role in various biotic and abiotic stresses. Ethylene causes exaggeration of the apical hook, shortening and thickening of the hypocotyls, and inhibition of root growth in dark grown seedlings of *Arabidopsis* – a phenomenon known as *triple response*.

Ethylene is perceived by a family of five receptors (ETR1, ETR2, ERS1, ERS2, and EIN4) in *Arabidopsis*, which share similarity with the bacterial two-component signaling system. These receptors show redundancy in function and act as negative regulators of ethylene responses. These contain an amino terminal ethylene binding domain (sensor domain) that binds ethylene via a Cu^{++} and a C-terminal histidine kinase domain whose activity is not necessary for receptor function. It has been suggested that this domain in ethylene receptors is probably required for localization, protein stability, and/or interaction with other functions. Besides *Arabidopsis*, a large number of studies have been carried out in tomato to understand the role of ethylene in development and in fruit ripening. In tomato, six ethylene receptors (LeETR1–6) have been identified. These are very similar to those in *Arabidopsis*. Most of the information available on tomato ethylene signaling is through *NEVER RIPE* (*Nr*) and *ripening inhibitor* mutants that are affected in tomato fruit ripening.

The receptors are localized in endoplasmic reticulum. The ethylene signal is transmitted by the ethylene receptors to CTR1 (an Raf-like MAPKKK-like protein). Ethylene signal is then transduced downstream via a signal transduction cascade that involves MAP kinases to finally activate a series of transcription factors that are specific for different genes. The signaling is unique in the sense that while the first two components of the signal cascade, that is, ethylene receptors and CTR1 are negative regulators, the downstream components (MAPKK onwards) are positive regulators of the signal. Also, it couples a two-component system with an MAP kinase cascade. In the absence of ethylene, the receptors are in an active state and interact with N-terminus of CTR1, which then represses the downstream ethylene responses. Interestingly, the MAPK cascade genes involved in ethylene signaling have also been shown to play a role in biotic stress related ethylene release and salt stress signaling.

In *Arabidopsis*, the EIN3 family has been identified as containing transcription factors, which mediate all ethylene responses. There are six members in the family but only EIN3 and EIL1 play a major role in ethylene responses. The rest of the members (EIL2–5) play roles in specific developmental stages or tissues. It is speculated that these may function in other pathways, which are not ethylene dependent. Other plant species like tomato, tobacco, and mung bean also have orthologs of EIN3. The EIN3 family of transcription factors is regulated by ethylene through a posttranscriptional mechanism that involves protein degradation mediated by the ubiquitin/proteasome pathway. Interestingly, glucose enhances EIN3 degradation through functional glucose sensing hexokinase (AtHXK1) indicating an interaction between ethylene and sugar signaling.

Genetic and biochemical studies have identified an ethylene responsive gene *ERF1* as a first target for EIN3. *ERF1* binds to 11 bp GCC-box element in promoters of many ethylene and pathogen induced genes. ERF1 belongs to a large

AP2/EREBP family (ETHYLENE- RESPONSE ELEMENT BINDING protein), which is exclusive to the plant kingdom. This family consists of 144 members in *Arabidopsis* and is characterized by the presence of a conserved AP2 domain of 58 to 59 amino acids. This large family is divided into two subgroups: The AP2-like subfamily consists of two AP2 domains and includes genes such as *APETALA2, AINTEGUMENTA* (*ANT*) and *GLOSSY15*. These genes are involved in seed development and floral organ specification. The subfamily EREBP has a single AP2-like domain and includes ERF1, ERF2, ERF3, and ERF4 from tobacco, and AtEBP, TINY, CBF1, OsEBP89, and so on. ERFs can act both as activators and repressors of gene expression. ERFs that repress gene expression contain an additional EAR (ERF-associated amphiphilic repression) motif that is also present in many zinc finger repressor proteins. These repressor-ERFs function mainly in response to biotic and abiotic stresses. Important targets of the ERFs include genes such as chitinase, β-1,3-glucanase and pathogenesis related proteins that play a role in defence. In addition, ethylene induction of β-1,3-glucanase also plays a role in seedling germination, mobilization of storage reserves in the endosperm, pollen development, and so on. Studies carried out on ethylene-regulated genes involved in fruit ripening and senescence have so far not detected GCC-box. These genes, instead, have conserved octanucleotide sequence PERE (ATTTCAAA) that confers ethylene sensitivity. This sequence is also present in E4 and E8 promoters of two tomato fruit specific genes. In a recent experiment, 624 genes out of 22 000 studied, were found to be regulated by ethylene in the seedlings of *Arabidopsis*. Other fruiting plants such as tomato, banana, peach, apple, and so on, are likely to possess many more genes regulated by ethylene due to its role in climacteric fruit ripening. Ethylene can also interact with JA in defense and wounding responses and these effects are mediated by ERF1. It was recently shown that ERF1 acts as a common downstream component for the two hormones and blocking either pathway by genetic means prevents ERF1 induction by ethylene and JA, either alone or in combination.

3.6 Regulation by Jasmonic Acid

Jasmonic acid and its derivatives affect a variety of processes in plants such as pollen maturation, root growth, tendril coiling, fruit ripening, biotic and abiotic stresses, and defense against insects. The biochemistry of JA synthesis is well known but its perception and receptors are not yet very well understood. The current information for JA signaling is mainly available through *Arabidopsis* and tomato but with several discrepancies.

JA signaling is induced by osmotic stress, wounding, drought, and various elicitors (chitins, oligosaccharides, etc.). Transgenic studies reveal that in tomato a MAP kinase, WIPK, is required for wound-induced JA biosynthesis. In *Arabidopsis*, MPK4 is responsible for mediating JA responses indicating that different MAP kinases act at different points in the JA pathway. Insect attack on tomato leaves activates systemic JA synthesis through a small polypeptide systemin, an 18-amino acids long peptide that has been identified as a signal substance for systemic activation of defense genes. It is produced from a larger precursor and is transported by the vascular system of the plant. Systemin and oligosaccharides use a similar transducer although the

oligosaccharide effect is local, whereas systemin leads to systemic effects.

Microarray analyses suggest that JA activates at least five genes for its own biosynthesis. Transcriptional activation of these five genes (*DAD1*, *LOX2*, *AOS*, *OPR3*, and *JMT*) occurs at the site of JA biosynthesis indicating that JAs are synthesized locally in response to stress. Till date, no receptor has been identified for JA signal transduction. Different studies carried out in *Arabidopsis* indicate involvement of two different pathways in JA signal transduction. One pathway, responsible for defense responses, recognizes OPDA (12-oxophytodienoic acid) or JA. The other pathway recognizes only JA for pollen maturation. Screening of various mutants for JA receptors identified only *COI1* (*CORONATINE INSENSITIVE 1*) and *JAR1* (*JASMONATE RESISTANT1*) genes. But the proteins encoded by these genes do not show homology to any plant receptor protein. So far two genes that mediate the JA signal transduction have been identified. The first one, *ERF1*, is common to both ethylene and JA responses during pathogenesis. The other *AtMYC2* differentially regulates two groups of genes; those related to defense responses are repressed by it while those related to JA mediated systemic responses to wounding are activated by it. Interestingly, *ERF1* has the opposite effect on these groups of genes. Thus, responses of ethylene and JA can be integrated during plant defenses but are antagonistic during the wound response. JA can also act synergistically with ABA in inhibiting seed germination and has a close interaction (both positive and negative) with salicylic acid.

Like other hormones, regulation of JA responsive genes occurs through proteasome-mediated degradation. Studies on *coi1-1* mutant, which is insensitive to coronatine (a structural analog of JA), reveal that this mutant fails to express JA-regulated genes such as *VSP* (VEGETATIVE STORAGE PROTEIN), *Thi2.1* (THIONIN 2.1) and *PDF1.2* (PLANT DEFENSIN1.2). The *COI1* gene encodes F-box protein, which functions as a receptor, recruiting regulatory proteins as a substrate for ubiquitin mediated destruction. Immunological studies suggest that COI-1 forms an SCF complex, which degrades other proteins by proteasome system. Only one transfactor, ORCA3, is known to be involved in JA responses. This transcription factor has a JA responsive APETALA2 (AP2) domain. Its overexpression results in induced expression of several genes for metabolite synthesis. ORCA3 binds to JA and elicitor responsive elements (JERE) in the promoters of JA responsive genes. The JA responsive element in the promoter of plant defensin1.2 (PDF1.2) also contains GCC-box, which is required for ERF1 binding.

3.7 Regulation by Brassinosteroids

Brassinosteroids (BR) are plant steroids that act as plant hormones and brassinolide (BL) is the most bioactive form of BRs. Brassinosteroids have growth-promoting activity and affect leaf bending, cell elongation, cell division, proton pumping, and membrane polarization. In addition to this, these steroid hormones also affect sensors and stress responses. BR-deficient mutants show dark green dwarf phenotypes in the absence of light suggesting potential interaction between BR and light-signaling pathways. Inhibitors of BR biosynthesis (Brassinozol) also generate phenotypes mimicking those of BR-deficient mutants.

BL is perceived by putative receptors BRASSINOSTEROID INSENSITIVE 1(BRI1), which is a plasma membrane localized leucine rich repeat (LRR) receptor kinase. BRI belongs to the specific family of receptors, which like kinases, have Cf-9, TOLL, LRR, and CLAVAT2 receptors. The extracellular portion of BRI is required for BR dependant activation of the intracellular Ser/Thr-specific kinase domain. Interestingly, a tomato receptor kinase SR160LRR identified as the receptor of peptide hormone systemin is identical to BRI1. This indicates that brassinosteroid and systemin share one receptor for signaling pathway.

BRs inhibit the genes required for their own biosynthesis. The transcription of CPD gene (encoding C-23 steroid hydroxylase CYP90A1) is downregulated by BL and other bioactive BRs. The BRI-5 mutant has impaired BR perception and shows derepressed expression of CPD. Besides BR biosynthetic genes, BL also negatively regulates the transcription of a number of genes. The BR-activated gene, *BRUI*, identified in soybean encodes xyloglucan endotransglycosylase. Brassinolides also induce xyloglucan endotransglycosylases (XETs) in several plants and this XET induction correlates with cell wall loosening during BL-induced growth responses. The other important genes that are BR inducible include cyclin gene *CycD3* (involved in cell cycle) and CDC2b cyclin dependent kinase in the absence of light. A homolog of animal translation initiation factor eIF_3 (TRIP) has been found to be upregulated by BR in *Arabidopsis* and can establish a link between BL-signaling and developmental pathways controlled by the homolog of translation initiation factor.

At present, all the components of BR signaling have not been elucidated and a hypothetical model for BR signaling has been proposed. According to this model, BRI1 is the receptor for signal perception. BR promotes autophosphorylation and activation of BRI1 triggering downstream responses by its binding to BRI1. KAPP-like protein phosphatase has been suggested to modulate BRI1. BRI1 deactivates the cytoplasmic GSK3/SHAGGY-like BIN2 kinase, which is a negative regulator of BR signaling and it phosphorylates downstream components BES1 and BZR1 and prevents their translocation to the nucleus. A MAPK cascade may come into play and activate transcription factors that control activation or repression of BR-responsive genes. Alternately, it is possible that BRI1 directly phosphorylates transcription factors and other signaling components like TRIP modulating diverse cellular responses.

3.8 Control of Gene Expression by Proteolytic Degradation:Role of the Proteasome

In recent years, proteolyses of key repressors or activators of response mediators in hormonal regulation, disease resistance, flowering, and light regulation has been shown to be a very effective means of controlling responses. It involves the polyubiquitination of proteins marked for degradation and their transfer to the proteasome by the SCF complex proteins. The F-box proteins that have specificity for the target protein as well as the SCF proteasome play an important role in this process. More than 600 F-box proteins are believed to be present in *Arabidopsis*. In hormonal responses, degradation of repressors of auxin, gibberellins and JA signaling, and activators of ethylene and ABA responses provide a rapid means to regulate responses. Repressors for auxin signaling such as the Aux/IAA proteins

are targeted for degradation by the SCF^{TIR} complex, while the DELLA motif containing repressors of GA signaling, RGA and GAI, are degraded by the proteasome by SCF^{SLY}complex. On the other hand, ethylene signaling is prevented in the absence of ethylene by the degradation of EIN3 – a transcription factor that activates ERFs, while ABA signaling in the absence of ABA is prevented by degradation of the bZIP factor ABI5. This is similar to the prevention of light responses in the dark by degradation of light responsive transcription factors by COP1.

3.9 Cross Talk Between Phytohormones

Modern research on the classical plant hormones is providing interesting information about the interaction between these regulatory components. The cross talk between these hormones influences plant growth and development.

Ethylene and abscisic acid interact with each other during root elongation and seed dormancy. Mutant analyzes of *era3* (*enhanced response to ABA3*) reveal that this gene is allelic to ethylene insensitive *ein2* mutant. Increased seed dormancy of mutant *ein2–45* is counteracted by ABA insensitive mutant *abi3–4*. The double mutant *abi3–4ein2–45* germinates faster than wild type indicating a strong interaction between ethylene and ABA signal transduction pathway. In contrast to seed dormancy, ethylene, and ABA, both inhibit root elongation. ABA inhibits root growth more effectively if ethylene concentration is low indicating that ethylene negatively regulates root responsiveness to ABA. It has been hypothesized that ABA acts through ethylene signal transduction and both of these hormones compete with each other to activate that pathway.

Auxins and GA promote stem elongation and normal levels of auxins are required to maintain the normal levels of GAs. Auxin upregulates *LE* mRNA, which encodes the enzyme involved in GA_1 biosynthesis. Different studies made on pea plants suggest that auxin is transported to elongating internodes from apical buds. There it induces the transcription of *LE* mRNA and downregulates the expression of PsGA2ox1, which encodes an enzyme that catalyzes the deactivation steps of GA_1 to GA_8. Microarray studies also indicate that GA affects ethylene biosynthesis and its responses. Auxin and ethylene affect GA-regulated root and hypocotyl growth by modifying the stability of RGA. Other examples of phytohormone cross talk have already been mentioned in the text. The SCF E3 ubiquitin ligase mediated protein degradation plays a central role in each of the hormones (auxin, ethylene, ABA, GA, and JA) signal cascade and it is possible that this system provides the molecular link to the phytohormone cross talk.

JA signal pathway interacts with ethylene signal pathway in the expression of defense response and in development. It also interacts with auxin and salicylic acid signaling. Brassinosteroids (BR) interact synergistically with auxins in hypocotyls elongation. Several studies indicate that BR-induced effects are mediated via auxin, whereas BR alters the levels of endogenous auxin or enhances the sensitivity to auxin. BR upregulates many early auxin-inducible genes suggesting overlap of auxin and BR-regulated genes.

This cross talk between different hormones is indicative of the fact that possibly fewer elements are necessary for overall hormonal regulation. Probably, due to this cross talk, a single hormone is able to cause different responses in different

organs by interacting differently with other hormones that are present.

4
Gene Regulation During Stress

Plants are sessile organisms and are constantly exposed to changes in environment conditions. They are exposed to biotic stresses like pathogen or fungal attack, insect attack, herbivores, and wounding. Plants are also challenged by abiotic stresses like salt, drought, and cold. Plants have evolved various mechanisms to overcome the damage caused by stress. The response to environmental change is usually rapid and involves regulation of various genes. Hormones like ABA and ethylene and metabolites like salicylic acid, nitric oxide, and reactive oxygen species have been shown to play an important role during these stresses evoking various gene responses.

4.1
Regulation during Abiotic Stress

Environmental stresses such as drought, cold, and salinity, induce several genes that provide stress tolerance to plants. The genes induced by abiotic stresses have been classified into two major groups. One group encodes proteins that directly protect plant cells against stresses (such as heat shock proteins, LEA proteins, ROS scavenger proteins like SOD, catalase etc.), whereas the proteins encoded by the other group regulate gene expression and signal transduction. Many stress inducible genes include those that encode enzymes involved in phospholipid metabolism, various protein kinases like MAP kinases, Ca^{2+} dependent kinases, receptor-like kinases and histidine kinases.

Plants undergoing abiotic stresses in general face dehydration and hence almost 50% of the genes activated by these stresses such as drought, salinity, or ABA treatment are common. In contrast, only 10% of the genes are common to cold stress. Many transcription factors have been found to be stress responsive. These include members of DRE binding protein (DREB), ethylene responsive factor (ERF), the MYB family, the zinc finger family, the homeodomain transcription factor family and the WRKY family. Studies on the proline dehydrogenase gene (*ProDH*), which is rehydration inducible, reveal an important cis-acting element ACTCAT, which is found in many rehydration inducible genes. The gene *RESPONSIVE TO DEHYDRATION 29 (RD29A,* also known as *COLD REGULATED 78* or *COR78 / LOW TEMPERATURE INDUCED* or *LT178*) is responsive to drought, salt, and cold. The promoter of this gene has two distinct cis-acting elements; one DRE/CRT (*d*ehydration *r*esponsive *e*lement/*C r*epea*t*), responsive to drought, salt, and cold and the other ABRE, which is ABA responsive element. DREB proteins bind to DRE/CRT elements and are classified in two groups – DREB1 and DREB2 that are induced specifically by cold and drought respectively. Transgenic studies show that drought and salt trigger both ABA-dependent and ABA-independent pathways, whereas cold signal is transmitted through ABA-independent pathway because low temperature does not induce accumulation of ABA in plants. Homologs of DREB1 and DREB2 have also been found in rice and function similar to *Arabidopsis* homologs in stress responsive–gene expression and stress tolerance. Several drought-inducible genes are not activated by cold or ABA. These genes include *ERD1* (*EARLY RESPONSIVE TO*

DEHYDRATION1), which encodes a Clp protease regulatory subunit. This *ERD1* gene is also upregulated by senescence.

Recent studies indicate that Ca^{2+} acts as a second messenger in abiotic stress and ABA response. According to present hypothesis, Ca^{2+} transmits the stress signal downstream by interacting with protein sensors. The Ca^{2+} sensors such as calmodulin, calcineurin B-like (CBLs) and calmodulin domain protein kinase (CDPK) bind Ca^{2+} and induce conformational change. CBL and calmodulin are small Ca^{2+} sensors, which function by interacting with and regulating the function of their target genes. CBL protein interacts with a specific group of protein kinases called *CBL interacting protein kinases* (CIPKs). There are evidences that *CaM* gene functions as a negative regulator of stress gene expression, whereas CBL1 acts as both a positive and a negative regulator in different signaling pathways. It acts as a positive regulator of salt and drought stress and as a negative regulator of cold response.

4.2 Regulation during Biotic Stress and Defense

Plants interact with a huge array of organisms such as viruses, bacteria, fungi, insects, nematodes, and herbivores throughout their lives. Complex and integrated defense signaling pathways that allow these to respond to individual organisms have been evolved in a variety of ways. The disease resistance in plants against pathogens invokes hypersensitive responses. There is an activation of an array of defense genes and the production of antimicrobial phytoalexins. Plants are also able to develop systemic acquired resistance in uninfected portions of plant subsequent to attack by the same or unrelated pathogen.

These responses are initiated by the recognition of pathogen. A plant resistance gene interacts with pathogen avirulence (*avr*) gene and may trigger a hypersensitive response that results in cell death. Recent studies point out that MAPK cascade is the integrative point to downstream responses after perception of different pathogens and elicitors. Fungal elicitor xylanase from *Trichoderma vivde* and bacterial elicitor harpin have been found to induce MAPK-like activities. MAPK such as SIPK (salicylic acid induced protein kinase) and WIPK (wound induced protein kinase) are activated by various biotic and abiotic responses.

Virus resistance is mediated by the synthesis of pathogenesis related proteins (PRPs) and virus-induced gene silencing, which involves the recognition of the dsRNA of the virus, its degradation by a DICER-like complex to small 21 to 25 nt fragments and the methylation of promoters that drive gene expression of these viral genes. Bacterial resistance involves a family of proteins known as the *NBS-LRR* (*nucleotide binding site-leucine rich repeat*) family of receptors. These receptor kinases have specificity for different pathogenic proteins and mediate gene for gene resistance against several pathogens. Fungal resistance may involve expression of chitinases, endoglucanases, while herbivory and insect attack is prevented by release of toxic volatiles such as nitriles, isothiocyanates, epithionitriles, and so on, by the action of myrosinases on glucosinolates. Insect attack is also resisted in some plant legumes such as soybean and cowpea by the expression of protease inhibitors that block insect gut proteases.

Many defense genes are activated by a group of transcription factors belonging to the WRKY family. These genes

recognize W-box elements containing a conserved TGAC motif present in several plant defense genes such as class I chitinases, some PRPs, a glutathione S-transferase, and so on. Defense response is also initiated by activation of genes of the phenylpropanoid pathway that results in production of phytoalexins. In addition, many members of ethylene-response factors (ERF) subfamily of AP2 transcription factors (discussed in ethylene section) are also involved in defense. Several ERF genes are regulated by plant hormones, ethylene, jasmonic acid, and salicylic acid and also by pathogen attack. Pti4, Pti5, Pti6 (Pto interacting kinase), which interact with tomato disease resistance proteins are members of ERF family. AtERF3 and AtERF4 are transcriptional repressors, which are induced by both ethylene and JA. Pti4–6 indirectly regulates SA (salicylic acid) responses through interaction with other TFs. These ERF genes are regulated posttranscriptionally by MAPK mediated phosphorylation. ERFs bind target sequence GCC-box in promoters of many pathogenesis related and hormone-induced genes. The overexpression of many of these ERFs confers increased resistance to pathogens. Oligogalacturonide products produced from wound-induced plant polygalacturonase and chitinase also elicit defense responses against attack from both herbivores and pathogens. Systemin, a wound inducible hormone potentiates oxidative bursts in response to oligogalacturonide products to activate genes near the wound site.

4.3 Regulation by Metabolites

4.3.1 Role of Nitric Oxide in Defense

Recent studies have highlighted the role of NO and its interaction with other hormones in many plant defense responses. NO signaling is mediated by three second messengers, namely, cGMP, cADPribose, and Ca^{++}. NO affects the expression of many genes such as *PHENYLALANINE AMMONIA LYASE* (*PAL*) and *PR1*. NO also plays a role in regulating plant cell death and probably facilitates the cell-to-cell spread of hypersenstive response. However, NO can also protect cells from the reactive oxygen species (ROS) mediated cell damage by inducing expression of catalase, SOD, glutathione S-transferase. It is thus a bifunctional modulator of cell death. NO also plays a role in mediating ABA-induced stomatal guard cell closure.

There is an interrelationship between NO, SA, and jasmonic acid in plant defense. NO induces levels of endogenous SA in tobacco and *Arabidopsis*. NO inhibitors attenuate SA-induced systemic acquired resistance (SAR). On the other hand, NO function requires SA and overexpression of SA-degrading enzyme salicylate hydroxylase abolishes NO-induced resistance against TMV. There is cross talk between NO and JA/wounding signaling pathway during defense. Both wounding and JA induce NO synthesis through an unidentified enzyme. Conversely, NO activates genes that encode JA biosynthetic enzymes. This NO-induced synthesis of JA is observed in SA-deficient plants indicating that SA negatively regulates NO-mediated JA synthesis.

4.3.2 Role of Reactive Oxygen in Stress

Biotic and abiotic stresses lead to production of reactive oxygen intermediates (ROS/ROI). These were first considered to be toxic byproducts of aerobic metabolism but now these are also considered as signaling molecules, which activate various processes such as abiotic stress,

defense against pathogen acclimation, and programmed cell death. These are also involved in stomatal closure. Various types of stresses like salt, chilling, heavy metals, ultraviolet radiation, ozone, mechanical stress, and pathogen attack enhance cellular ROI levels from 240 $\mu Ms^{-1}O_2^-$ to 720 $\mu Ms^{-1}O_2^-$. Since ROI, though toxic, participates in signaling events, plants require two different mechanisms to regulate intracellular ROI concentration; one enables the detoxification of excess ROI and the other enables fine modulation for low levels of ROIs for signaling purposes.

Superoxide dismutase (SOD), ascorbate peroxidase (APX) and catalase (CAT) are major ROI scavengers. SOD and APX pathways are found in almost all the cellular compartments whereas CAT pathway for ROI scavenging is restricted to only peroxisomes. APX has very high affinity for H_2O_2 and is believed to be a fine modulator of ROI for signaling, whereas CAT is thought to be responsible for the removal of excess ROI during stress. Ascorbic acid and glutathione play crucial role in plant defense against oxidative stress. Mutant and transgenic plants with suppressed/altered ascorbic acid and glutathione level are hypersensitive to stress conditions. It is important to maintain a high ratio of reduced:oxidized state of these antioxidants in the cell for proper scavenging of ROI.

Besides scavenging, reduced ROI production can also be helpful in stress conditions. Plants reduce ROI during stress through anatomical and physiological adaptation. ROI production is also decreased by alternative channeling of electrons in organelle electron transport chains by alternative oxidases (AOX). ROIs also play a role in defense of plants against biotic stress that is different from abiotic stress. During pathogen attack, enhanced enzymatic activities of plasma membrane–bound NADPH oxidases, amine oxidases, and cell wall bound peroxidases produce ROI. H_2O_2, produced during the response, activates many plant defense responses along with SA and NO. SA and NO suppress activities of APX and CAT resulting in overaccumulation of ROI and activation of programmed cell death (PCD). The suppression of ROI scavenging mechanisms together with the synthesis of NO is crucial for the activation of PCD. The opposite roles played by ROI during abiotic and biotic stress could be due to the action of SA and NO, from cross talk between different signaling pathways or from differences in the steady state level of ROIs produced during different stresses.

A simplified hypothetical model for ROI mediated–signal transduction pathway involves sensing of H_2O_2 by a sensor followed by activation of calmodulin and an MAP kinase cascade. This in turn activates or deactivates several transcription factors. These factors regulate the response of plants to oxidative stress. Recently, an *Arabidopsis* mutant *radical-induced cell death1* (*rcd1*) has been identified which, upon ozone exposure, exhibits higher ROI production, reduced sensitivity to ABA, ethylene and methyl jasmonate, and has altered expression of several hormone-regulated genes. RCD1 belongs to the WWE protein–protein interaction domain protein family and is suggested to act as an integrative node in hormonal signaling and in regulation of stress responsive genes.

5 Developmental Regulation of Genes

In plants, development of different organs is governed mainly by members of the MADS-box gene family and the

homeodomain family of transcription factors. Different genes responsible for meristem identity and development of specific organs are expressed at specific stages where they activate or repress other genes to bring about development. While the homeodomain genes specify organ identity in roots, stem, and leaves, the MADS-box genes are mainly involved in flower and fruit development. Owing to the complexity of developmental regulation, only the major transcription factor genes involved in each process will be mentioned.

Seed development is governed by the action of *ABI3* and it activates genes for LEA proteins that protect the seed from desiccation effects. Apical basal pattern of the plant is determined by *GNOM* (*GN*), *MONOPTEROS* (*MP*), *GURKE* (*GK*) and *FACKEL* (*FK*), while radial patterning is governed by ARABIDOPSIS THALIANA MERISTEM LAYER1 (*ATML1*) and *KNOLLE*. Endodermal cell identity is governed by *SCARECROW* (*SCR*) and *SHORT ROOTS* (*SHR*), while root meristem is regulated by *HOBBIT (HBT)*. Shoot meristem and shoot development are governed by *SHOOT MERISTEMLESS* (*STM*), *KNOTTED* (*KN*) and *WUSCHEL* (*WUS*), while leaf development is controlled by *AUGUSTFOLIA* (*AN*), *ROTUNDIFOLIA* (*ROT*), *PHAVULOTA* (*PHV*), *PHABULOSA* (*PHB*), and *PHANTASTICA* (*PHAN*), the last of which is a repressor of *KNOX* genes. Trichome development is regulated by *GLABRA1(GL1*) and *TRANSPARENT TESTA GLABRA*(*TTG*). The floral meristem genes *LEAFY* (*LFY*) and its repressor *FLOWERING LOCUS C* (*FLC*) control the transition from vegetative phase to flowering phase. The floral organ identity genes consist of four classes. Their interactions determine the development of the sepals, petals, stamens, and carpels. These include *APETALA1* (*AP1*) and *APETALA2* (*AP2*), which are A class genes, *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), which are B class genes, *AGAMOUS* (*AG*), which is a C class gene and *SEPALLATA 1, 2,* and *3 (SEP1, SEP2, SEP3),* which interact with all the three classes. Genes such as *RIPENING INHIBITOR* (*RIN*) play a role in fruit ripening while *SHATTERPROOF1* and *2* (*SHP1* and *SHP2*), *AGAMOUS LIKE 15* (*AGL15*), and *JOINTLESS* control dehiscence and abscission.

The control of development is brought about by a complex interaction of the different gene products. MADS-box genes are believed to function as homo- and heterodimers through interaction from their K-box, which is unique to plant MADS genes. They recognize a sequence motif $CC(A/T)_6TGG$ also known as the *CArG box*.

6 Epigenetic Regulation of Plant Development

In recent years, studies on epigenetic control of eukaryotic gene expression have added a new dimension to our understanding of gene regulation. It is apparent now that activation and repression of genes by histone modification and chromatin remodeling as well as by small RNA molecules such as miRNAs and siRNAs plays an important role in development and temporal and spatial expression of genes. In plants too, several recent studies have highlighted the role of these processes in gene expression.

6.1 Role of miRNAs in Plant Development

MicroRNAs are small 22 to 25 nt RNAs that regulate gene expression by binding

specific sequences in target genes and suppressing their expression. A study of the plant microRNAs and their putative target sequences showed that in about 70% of the cases the target genes were transcription factors and these are mainly related to plant developmental processes.

Three class-III HDZIP genes *PHABULOSA* (*PHB*), *PHAVULOTA* (*PHV*), and *REVOLUTA* (*REV*) have recently been shown to regulate miRNAs. While *PHB* and *PHV* are involved in determining adaxial and abaxial polarity in lateral organs, *REV* is required for leaf development and vascular patterning. All three genes contain targets for two miRNAs, miR165 and miR166, and regulation probably occurs by RNA degradation.

Some AP2 (*APETALA2*)-like genes have also been shown to be regulated translationally by miR172. It has been shown that AP2 (which is involved in flower development and is transcribed in all four whorls of the flower), restricts the expression of AGAMOUS, a class C gene that is responsible for carpel and stamen identity, to whorls 3 and 4. This is made possible by expression of miR172 in whorls 3 and 4 and its repression of AP2 translation in whorls 3 and 4. This results in accumulation of AGAMOUS in these whorls. miR172 is not expressed in whorls 1 and 2 and thus allows AP2 protein to accumulate in these whorls and repress *AGAMOUS* expression.

6.2 Repression by DNA Methylation

Another mechanism of regulation involves the POLYCOMB group of genes that are involved in methylation of DNA and long-term memory of repression. In plants, the Polycomb group (PcG) of genes are involved in regulation of MADS-box genes and thus control development, floral identity, response to vernalization, flowering and seed development, and so on. Examples include the repression of *AGAMOUS* (responsible for carpel and stamen development) by *CURLY LEAF* and *EMBRYONIC FLOWER* and of *FLOWERING LOCUS C* (a floral transition repressor) by *VERNALIZATION*. In both cases, the target genes are repressed by hypermethylation of an intron.

Other genes related to DNA methylation and chromatin remodeling have been identified in plants and include *DECREASED DNA METHYLATION (DDM1), DNA METHYLTRANSFERASE1* (*MET1*), *CHROMOMETHYLASE 3* (*CMT3* – a DNA methyltransferase), *LIKE HETEROCHROMATIN PROTEIN1* (*LHP1*), *DOMAINS REARRANGED METHYLASE 1* and 2 (*DRM1* and *DRM2*), *KRYPTONITE* (*KYP* – an H3K9 methyltransferase), and *DEMETER* (a DNA glycosylase).

Summary

Plant gene expression presents a complex picture of gene activity when compared to other eukaryotes. The nuclear genome and organelle genome show a lot of difference in their transcription and translation machinery. Several RNA polymerases are encoded, which have specific roles in nuclear and organelle gene transcription. In general, several factors together with RNA polymerase bind to the promoter and other cis-elements of the gene and begin transcription. The mature transcript may be generated by a variety of posttranscriptional processing, which prepares the transcript for translation on ribosomes. The signal for the start of transcription

may be constitutive, developmental or hormonal as a result of various changes sensed by plants in its environment. The signal may include ions, small molecules like hormones and other metabolites, and light and various kinds of stresses to which plants are exposed. Several of these signals are mediated through a signaling system, which includes receptors, transcription factors, and a variety of MAP kinases. Besides transcriptional regulation, genes are also regulated posttranslationally. Selective degradation of activators and repressors of signal pathway by the proteasome is turning out to be a very important means of controlling signal transduction. Changes in chromatin structure histone modification and DNA methylation have also been shown to regulate gene expression in plants.

Bibliography

Books and Reviews

Abler, M.L., Green, P.J. (1996) Control of mRNA stability in higher plants, *Plant Mol. Biol.* **32**, 63–78.

Brown, T.A. (2002) *Genomes*, BIOS Scientific Publishers, Ltd, Oxford, UK.

Futterer, J., Hohn, T. (1996) Translation in plants-rules and exceptions, *Plant Mol. Biol.* **32**, 159–189.

Kawaguchi, R., Bailey-Serres, J. (2002) Regulation of translational initiation in plants, *Curr. Opin. Plant Biol.* **5**, 460–465.

Knoop, V. (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective, *Curr. Genet.* **46**, 123–139.

Quail, P.H. (2002) Phytochrome photosensory signalling networks, *Nat. Rev. Mol. Cell Biol.* **3**, 85–93.

Sharrock, R.A., Quail, P.H. (1989) Novel phytochrome sequences in *Arabidopsis thaliana*: structure, evolution, and differential expression of a plant regulatory photoreceptor family, *Genes Dev.* **3**, 1745–1757.

Simpson, G.G., Filipowicz, W. (1996) Splicing of precursors to mRNA in higher plants: mechanism, regulation and sub-nuclear organisation of the spliceosomal machinery, *Plant Mol. Biol.* **32**, 1–41.

Srivastava, A., Nath, P. (2002) Mitochondrial Genome: Rearrangements, RNA Editing and Cytoplasmic Male Sterility in Higher Plants, in: Nath, P., Mattoo, A.K., Ranade, S.A., Weil, J.H. (Eds.) *Molecular Insight in Plant Biology*, Oxford and IBH publishing Co. Pvt. Ltd., New Delhi, India and Science Publisher Inc., Plymouth (UK) and Enfield (USA), pp. 13–32.

Ulmasov, T., Hagen, G., Guilfoyle, T.J. (1999) Activation and repression of transcription by auxin-response factors, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5844–5849.

Primary Literature

Baba, K., Schmidt, J., Espinosa-Ruiz, A., Villarejo, A., Shiina, T., Gardeström, P., Sane, A.P., Bhalerao, R.P. (2004) Organellar gene transcription and early seedling development are affected in the *rpoT;2* mutant of *Arabidopsis*, *Plant J.* **38**, 38–48.

Bentolila, S., Alfonso, A.A., Hanson, M.R. (2002) A pentatrico-peptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10887–10892.

Bock, R. (1998) Analysis of RNA editing in plastids, *Methods* **15**, 75–83.

Brady, S.M., Sarkar, S.F., Bonetta, D., McCourt, P. (2003) The ABSCISIC ACID INSENSITIVE 3 (ABI3) gene is modulated by farnesylation and is involved in auxin signalling and lateral root development in Arabidopsis, *Plant J.* **33**, 1–9.

Cameron, R.K., Paiva, N.L., Lamb, C.J., Dixon, R.A. (1999) Accumulation of salicylic acid and PR-1 gene transcripts in relation to the systemic acquired resistance (SAR) response induced by pseudomonas syringae pv tomato in Arabidopsis, *Physiol. Mol. Plant Pathol.* **55**, 121–130.

Chamovitz, D.A., Wei, N., Osterlund, M.T., von Arnim, A.G., Staub, J.M., Matsui, M., Deng, X.W. (1996) The COP9 complex, a novel multisubunit nuclear regulator involved in light control of a plant developmental switch, *Cell* **86**, 115–121.

Chen, X. (2004) A microRNA as a translational repressor of *APETALA2* in Arabidopsis flower development, *Science* **303**, 2022–2025.

Chen, C., Chen, Z. (2000) Isolation and characterization of two pathogen- and salicylic acid-induced genes encoding WRKY DNA-binding proteins from tobacco, *Plant Mol. Biol.* **42**, 387–396.

Coen, E.S., Meyerowitz, E.M. (1991) The war of the whorls: genetic interactions controlling flower development, *Nature* **353**, 31–37.

Covello, P.S., Gray, M.W. (1989) RNA editing in plant mitochondria, *Nature* **341**, 662–666.

D'Agostino, I.B., Derue're, J., Kieber, J.J. (2000) Characterization of the response of the Arabidopsis response regulator gene family to cytokinin, *Plant Physiol.* **124**, 1706–1717.

Dargeviciute, A., Roux, C., Decreux, A., Sitbon, F., Perrot-Rechenmann, C. (1998) Molecular cloning and expression of the early auxin-responive Aux/IAA gene family in *nicotiana tabacum*, *Plant Cell Physiol.* **39**, 993–1002.

Das Gupta, J., Li, Q.S., Thomson, A.B., Hunt, A.G. (1998) Characterization of a cDNA encoding a novel plant poly(A) polymerase, *Plant Mol. Biol.* **37**, 729–734.

Davies, B., Egea-Cortines, M., Andrade Silva, E., Saedler, H., Sommer, H. (1996) Multiple interactions amongst floral homeotic MADS box proteins, *EMBO J.* **15**, 4330–4343.

Delledonne, M., Zeier, J., Marocco, A., Lamb, C. (2001) Signal interactions between nitric oxide and reactive oxygen intermediates in the plant hypersensitive disease resistance response, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13454–13459.

Desikan, R., Griffiths, R., Hancock, J., Neill, S. (2002) A new role for an old enzyme: nitrate reductase-mediated nitric oxide generation is required for abscisic acid-induced stomatal closure in *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16314–16318.

Dewey, R.E., Levings, C.S., Timothy, D.H. III (1986) Novel recombinations in the maize mitochondrial genome produce a unique transcriptional unit in the Texas male-sterile cytoplasm, *Cell* **44**, 439–449.

Dill, A., Jung, H.S., Sun, T.P. (2001) The DELLA motif is essential for gibberellin-induced degradation of RGA, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14162–14167.

Dixit, R., Trivedi, P.K., Nath, P., Sane, P.V. (1999) Organization and post-translational processing of the *psbB* operon from chloroplasts of *Populus deltoides*, *Curr. Genet.* **36**, 165–172.

Felder, S., Meierhoff, K., Sane, A.P., Meurer, J., Driemel, C., Plucken, H., Klaff, P., Stein, B., Bechtold, N., Westhoff, P. (2001) The nucleus-encoded *HCF107* gene of *Arabidopsis* provides a link between intercistronic RNA processing and the accumulation of translation-competent *psbH* transcripts in chloroplasts, *Plant Cell* **13**, 2127–2141.

Finkelstein, R.R., Lynch, T.J. (2000) The *Arabidopsis* abscisic acid response gene *ABI5* encodes a basic leucine zipper transcription factor, *Plant Cell* **12**, 599–609.

Fisk, D.G., Walker, M.B., Barkan, A. (1999) Molecular cloning of the maize gene crp1 reveals similarity between regulators of mitochondrial and chloroplast gene expression, *EMBO J.* **18**, 2621–2630.

Gubler, F., Chandler, P.M., White, R.G., Llewellyn, D.J., Jacobsen, J.V. (2002) Gibberellin signaling in barley aleurone cells: control of SLN1 and GAMYB expression, *Plant Physiol.* **129**, 191–200.

Guo, H., Ecker, J.R. (2003) Plant responses to ethylene gas are mediated by $SCF^{EBF1/EBF2}$ dependent proteolysis of EIN3 transcription factor, *Cell* **115**, 667–677.

Hamilton, A., Voinnet, O., Chappell, L., Baulcombe, D. (2002) Two classes of short interfering RNA in RNA silencing, *EMBO J.* **21**, 4671–4679.

He, Z., Wang, Z.Y., Li, J., Zhu, Q., Lamb, C., Ronald, P., Chory, J. (2000) Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1, *Science* **288**, 2360–2363.

Hoch, B., Maier, R.M., Appel, K., Igloi, G.L., Kossel, H. (1991) Editing of a chloroplast mRNA by creation of an initiation codon, *Nature* **353**, 178–180.

Hua, J., Meyerowitz, E.M. (1998) Ethylene responses are negatively regulated by a receptor gene family in *Arabidopsis thaliana*, *Cell* **94**, 261–271.

Jarillo, J.A., Capel, J., Tang, R.H., Yang, H.Q., Alonso, J.M., Ecker, J.R., Cashmore, A.R. (2001) An *Arabidopsis* circadian clock component interacts with both CRY1 and phyB, *Nature* **410**, 487–490.

Krizek, B.A., Meyerowitz, E.M. (1996) The *Arabidopsis* homeotic genes *APETALA3* and *PISTILLATA* are sufficient to provide the B

class organ identity function, *Development* **122**, 11–22.

Larkin, R.M., Alonso, J.M., Ecker, J.R., Chory, J. (2003) GUN4, a regulator of chlorophyll synthesis and intracellular signaling, *Science* **299**, 902–906.

Leyser, H.M.O., Pickett, F.B., Dharmasiri, S., Estelle, M. (1996) Mutations in the *AXR3* gene of *Arabidopsis* result in altered auxin response including ectopic expression from the *SAUR-AC1* promoter, *Plant J.* **10**, 403–413.

Li, J., Nagpal, P., Vitart, V., McMorris, T.C., Chory, J. (1996) A role for brassinosteroids in light-dependent development of *Arabidopsis*, *Science* **272**, 398–401.

Luccioni, L.G., Oliverio, K.A., Yanovsky, M.J., Boccalandro, H.E., Yin, Y., Wang, Z.Y., Mora-Garcia, S., Li, J., Yoshida, S., Asami, T., Chory, J. (2002) BES1 accumulates in the nucleus in response to brassinosteroids to regulate gene expression and promote stem elongation, *Cell* **109**, 181–191.

Marsh, J.T., Sullivan, S., Hartwell, J., Nimmo, H.G. (2003) Structure and expression of phosphoenolpyruvate carboxylase kinase genes in solanaceae. A novel gene exhibits alternative splicing, *Plant Physiol.* **133**, 2021–2028.

Meinhard, M., Grill, E. (2001) Hydrogen peroxide is a regulator of ABI1, a protein phosphatase 2C from Arabidopsis, *FEBS Lett.* **508**, 443–446.

Meurer, J., Lezhneva, L., Amann, K., Godel, M., Bezhani, S., Sherameti, I., Oelmuller, R. (2002) A peptide chain release factor 2 affects the stability of UGA-containing transcripts in Arabidopsis chloroplasts, *Plant Cell* **14**, 3255–3269.

Meyer, K., Leube, M.P., Grill, E. (1994) A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*, *Science* **264**, 1452–1455.

Neff, M.M., Chory, J. (1998) Genetic interactions between phytochrome A, phytochrome B, and cryptochrome 1 during *Arabidopsis* development, *Plant Physiol.* **118**, 27–35.

Neuhaus, G., Bowler, C., Kern, R., Chua, N-H. (1993) Calcium/calmodulin-dependent and -independent phytochrome signal transduction pathways, *Cell* **73**, 937–952.

Nivison, H.T., Hanson, M.R. (1989) Identification of a mitochondrial protein associated with cytoplasmic male sterility in petunia, *Plant Cell* **1**, 1121–1130.

Osterlund, M.T., Hardtke, C.S., Wei, N., Deng, X.W. (2000) Targeted destabilization of HY5 during light-regulated development of Arabidopsis, *Nature* **405**, 462–466.

Rounsley, S., Ditta, G., Yanofsky, M. (1995) Diverse roles for MADS box genes in *Arabidopsis* development, *Plant Cell* **7**, 1259–1269.

Ruud, K.A., Kuhlow, C., Goss, D.J., Browning, K.S. (1998) Identification and characterization of a novel cap-binding protein from *Arabidopsis thaliana*, *J Biol. Chem.* **273**, 10325–10330.

Sane, V.A., Nath, P., Aminuddin, Sane, P.V. (1997) Development of insect resistant transgenic plants using plant genes: expression of cowpea trypsin inhibitor in transgenic tobacco plants, *Curr. Sci.* **72**, 741–747.

Satoh, R., Nakashima, K., Seki, M., Shinozaki, K., Yamaguchi-Shinozaki, K. (2002) ACTCAT, a novel cis-acting element for proline- and hypo-osmolarity-responsive expression of the ProDH gene encoding proline dehydrogenase in *Arabidopsis*, *Plant Physiol.* **130**, 709–719.

Solano, R., Stepanova, A., Chao, Q., Ecker, J.R. (1998) Nuclear. events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1, *Genes Dev.* **12**, 3703–3714.

Steber, C.M., McCourt, P. (2001) A role for brassinosteroids in germination in *Arabidopsis*, *Plant Physiol.* **125**, 763–769.

Suzuki, T., Miwa, K., Ishikawa, K., Yamada, H., Aiba, H., Mizumo, T. (2001) The *Arabidopsis* sensor His-kinase, AHK4, can respond to cytokinins, *Plant Cell Physiol.* **42**, 107–113.

Timmermans, M.C.P., Hudson, A., Becraft, P.W., Nelson, T. (1999) ROUGH SHEATH2: A Myb protein that represses KNOX homeobox genes in maize organ primordia, *Science* **284**, 151–153.

Tucker, M.L., Whitelaw, C.A., Nicholas, N., Lyssenko, N.N., Nath, P. (2002) Functional analysis of regulatory elements in the gene promoter for an abscission-specific cellulase from bean *(Phaseolus vulgaris*) and isolation, expression and binding affinity of three TGA-type bZIP transcription factors, *Plant Physiol.* **130**, 1487–1496.

Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A., Coupland, G. (2004)

Photoreceptor regulation of CONSTANS protein in photoperiodic flowering, *Science* **303**, 1003–1006.

Wang, Z-Y., Tobin, E.M. (1998) Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED (CCA1) gene disrupts circadian rhythms and suppresses its own expression, *Cell* **93**, 1207–1217.

Werneke, J.M., Chatfield, J.M., Ogren, W.L. (1989) Alternative mRNA splicing generates the two ribulosebisphosphate carboxylase/oxygenase activase polypeptides in spinach and *Arabidopsis*, *Plant Cell* **1**, 815–825.

Yamaguchi, S., Smith, M.W., Brown, R.G., Kamiya, Y., Sun, T. (1998) Phytochrome regulation and differential expression of gibberellin 3-hydroxylase genes in germinating Arabidopsis seeds, *Plant Cell* **10**, 2115–2126.

Yamaguchi-Shinozaki, K., Shinozaki, K. (1994) A novel *cis*-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salinity stress, *Plant Cell* **6**, 251–264.

Yang, J., Schuster, G., Stern, D. (1996) CSP41, a sequence-specific chloroplast mRNA binding protein, is an endoribonuclease, *Plant Cell* **8**, 1409–1420.

Yohn, C.B., Cohen, A., Danon, A., Mayfield, S.P. (1998) A poly(A) binding protein functions in the chloroplast as a message-specific translation factor, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2238–2243.

Zou, Z., Eibl, C., Koop, H.U. (2003) The stem-loop region of the tobacco *psbA* 5'UTR is an important determinant of mRNA stability and translation efficiency, *Mol. Genet. Genomics* **269**, 340–349.

Plant Pathology, Molecular

Matt Dickinson
The University of Nottingham, Nottingham, UK

1 Introduction 361

2 How do Pathogens Cause Disease? 361
2.1 Viruses 361
2.1.1 General Properties of Plant Viruses 361
2.1.2 Virus Replication and Assembly 362
2.1.3 Viral Movement Within Plants and Transmission between Plants 364
2.1.4 Affects of Virus Infection on Plant Processes 365
2.2 Bacteria 366
2.2.1 General Properties of Plant-Pathogenic Bacteria 366
2.2.2 Establishing Infection 366
2.2.3 Pathogenicity Factors 366
2.3 Fungi and Oomycetes 369
2.3.1 General Properties 369
2.3.2 Establishing Infection 370
2.3.3 Pathogenicity Factors of Necrotrophs 371
2.3.4 Pathogenicity of Biotrophs 372

3 Plant Defense against Disease 373
3.1 General Properties 373
3.2 Constitutive Defenses 373
3.3 Induced Mechanisms 375
3.4 Triggers for Induced Resistance 376
3.5 The *R* Genes 376
3.6 Downstream Signaling of Resistance 377
3.7 Coordination of Defense Responses 379

4 Concluding Remarks 379

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

Bibliography 380
Books and Reviews 380
Primary Literature 381

Keywords

Pathogenicity Factors
Proteins, toxins, exopolysaccharides, and other elicitors used by viruses, bacteria, fungi, and oomycetes to cause disease on plants.

Necrotrophy
Infection where the pathogen causes plant cell death and utilizes the nutrients released.

Biotrophy
Infection where the pathogen taps into living plant cells to obtain its nutrients.

Nonhost Resistance
The general resistance that makes a particular plant species resistant to a potential pathogen.

Host Resistance
The resistance that occurs within a plant species whereby some individuals are resistant while others are susceptible.

Resistance Genes
Genes present in plants that are responsible for triggering host resistance to pathogens.

Systemic Resistance
A form of resistance response that spreads from the initial trigger through the plant via the vascular system.

The interactions between plants and microbes can take many forms and involve complex molecular signaling between different organisms. In some cases, the interactions result in plant disease as the virus, bacterium, fungus, or oomycete pathogen colonizes the host tissue and accesses nutrients for its own development and proliferation. The mechanisms through which this occurs are both elegant and diverse, often involving production of specific pathogenicity factors, such as toxins or cell wall degrading enzymes, and suppression of the plant defense responses. In other cases, these plant defenses, which also include complex molecular signaling pathways within the plant and may involve structural barriers, nonhost resistance, induced resistance, and systemic resistance, are successful and combat the potential pathogen.

1 Introduction

The interactions between plants and microbes involve complex molecular signaling between different organisms and can have a range of outcomes. Many microbes exist as saprophytes living off nutrients released from plants, but some have developed specific mechanisms and strategies to manipulate plant metabolism and basal defenses for their own growth and development. It is these specialist organisms, and, in particular, those that infringe on our food supply and environment, that form the basis for studies in plant pathology.

The most significant causes of plant disease are viruses, bacteria, fungi, and oomycetes, and there are numerous historical examples where these have had a major impact on human populations and health through crop failures, starvation, and toxicity. The potato blight oomycete *Phytophthora infestans*, for example, resulted in complete failure of the potato crop in Ireland in 1846, contributing directly to 1 million deaths, while an epidemic of Southern corn leaf blight, *Bipolaris maydis*, resulted in a 50% corn yield reduction in the United States in 1970 at an estimated cost of $1 billion. Today, plant diseases are estimated to be responsible for a reduction in potential crop yields of around 10 to 15% annually worldwide. Furthermore, new diseases continue to emerge as pathogens evolve to overcome current control measures, or through the movement of plants and plant products between countries as part of global trade.

While there are a diverse range of strategies used by pathogens for infection and by plants for defense, by concentrating resources on model plant systems such as *Arabidopsis thaliana*, rice, and maize, and the pathogens that infect these plants, there has been a rapid increase in our understanding of molecular plant pathology. There are increasing numbers of genome sequences and tools for mining this information. The capacity to generate large numbers of mutagenized plants combined with elegant screens to select mutant phenotypes allows specific pathways to be targeted, while transformation systems and targeted gene silencing allow sequences to be connected to function through reverse genetics approaches. Gene expression profiling, proteomics, and metabolomics are being used extensively, and much of this information is readily available through bioinformatics resources. In this chapter, we will review the molecular mechanisms used by pathogens to cause disease on plants along with the methods used by plants to defend themselves against these pathogens.

2 How do Pathogens Cause Disease?

2.1 Viruses

2.1.1 General Properties of Plant Viruses

Viruses are common in plants and have been organized into 73 genera on the basis of size, shape, nucleic acid composition, sequence comparisons, transmission characteristics, and host range. The majority (90%) of plant viruses possess single-strand RNA genomes in the plus (+) or sense polarity. There are, however, some significant plant viruses with genomes of dsRNA (the phytoreoviruses), ssDNA (the geminiviruses), and dsDNA (the caulimoviruses). In most cases, the viral genome is a single segment of nucleic acid encapsidated in a simple protein coat, and 50% of plant viruses are rod

shaped (flexuous or rigid), while most of the rest are isometric. There are, however, some viruses that have segmented genomes and in many cases the segments are encapsidated in different particles resulting in multicomponent viruses, where all the components must be present in the plant cell simultaneously for it to replicate and spread.

Many plant virus genomes have been sequenced and they are generally very compact, with the genes separated by very short intergenic regions or overlapping. Normally, they encode only a small number of polypeptides, relying on host cell components for aspects of their translation, replication, and movement. Some of the encoded polypeptides are multifunctional, and, in general, the basic requirements for a successful virus are a replicase to promote replication of the genome, a coat protein to encapsidate this genome, a movement protein to enable the virus to move between cells, and a vector transmission protein to allow transmission between plants. The basic infection strategy for a model plant virus is shown in Fig. 1.

2.1.2 Virus Replication and Assembly

The majority of plant viruses with ssRNA genomes replicate in the cytoplasm of plant cells, often associated with membrane-bound cytoplasmic inclusions (viroplasms) containing host endoplasmic reticulum. A number of replication strategies have been developed by viruses to ensure that all the functions encoded by their genomes are translated. The first of these is the production of subgenomic RNAs (see Fig. 1). In this process, the capped viral genome is recognized by the host translation machinery as mRNA, and the 5′-most cistron up to the first stop codon is translated. In these viruses, this cistron is invariably the replicase, which has three functional components, an RNA-dependent RNA polymerase (RdRp), a helicase, and a methyltransferase. The helicase and RdRp, possibly in conjunction with host factors, act on the (+)-strand RNA template to produce a complementary (−) strand, and then use signals within this strand to produce more full-length (+) strands and subgenomic (+) strands for each of the downstream genes. These subgenomics are capped by the methyltransferase and in turn translated to produce the separate functional proteins such as the movement and coat proteins.

An alternative approach used by some (+)-strand viruses such as *Cucumber mosaic virus* is to effectively package each of the subgenomic RNAs into separate particles resulting in multicomponent viruses, while a third strategy exemplified by the *Potyviridae* is to have no stop codons in the genomic RNA and produce a single large polyprotein. This is then cleaved post-translationally into the different functional polypeptides, triggered by virally encoded proteinases within the polyprotein. For double-strand and negative-strand RNA viruses, similar translation mechanisms occur, but the initial replication process is promoted through the packaging of 10 to 20 copies of the replicase protein into the viral particles.

The ssDNA geminiviruses have a different problem to overcome for replication. Replication occurs in the nucleus of the infected plant by recruiting plant enzymes for conversion of the circular ssDNA genome into dsDNA, which is then transcribed to produce the viral replicase, and replicated by a rolling-circle mechanism. However, for this to happen, the virus has to induce a permissive state for DNA replication in already differentiated plant cells, and it appears to do this through

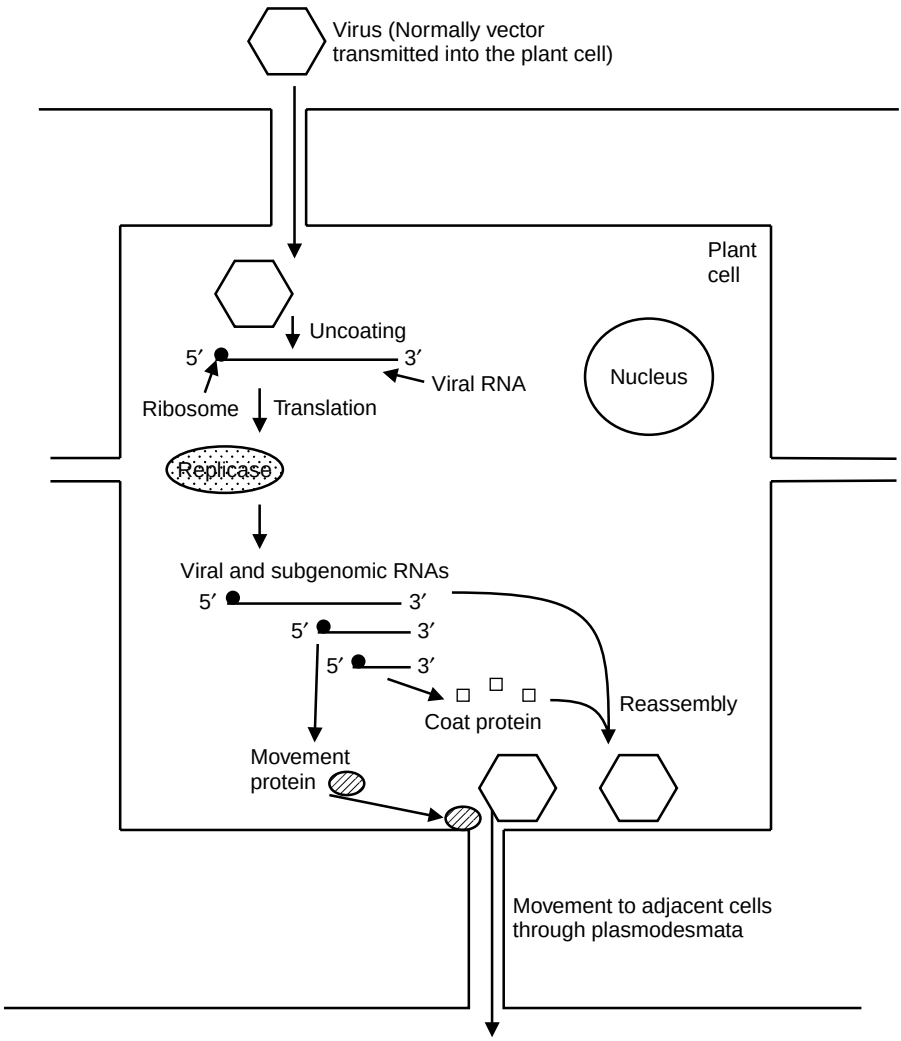

Fig. 1 General scheme for virus spread within plants. The virus enters the plant cell through wounds made mechanically or by a vector. The RNA then becomes uncoated and, at the same time, is surrounded by ribosomes for protection. Following replication and production of more coat protein, the virus reassembles or forms a ribonucleoprotein complex and is then moved between adjacent cells through the plasmodesmata, facilitated by the virally encoded movement protein.

association of the replicase with a plant retinoblastoma–related protein, which effectively regulates the cell cycle and shifts mature plant cells into the S-phase.

The dsDNA *Caulimoviridae* are plant retroviruses that replicate via reverse transcription, and use a ribosome-shunting mechanism to ensure that all the open reading frames on the polycistronic 35S transcript are translated. This 35S RNA is a full-length transcript, and is then copied by the virally encoded reverse

transcriptase into the DNA viral genome, although unlike retroviruses of vertebrates and bacteria, this does not appear to involve integration into the plant genome.

Assembly of plant viruses within cells following replication and translation is a process that has been best studied for *Tobacco mosaic virus* (TMV). This rod-shaped virus has a 6400-nucleotide ssRNA genome encapsidated in 2130 copies of a single 17.6-kDa coat protein subunit. These subunits reassemble spontaneously into discs containing 34 units in 2 rings of 17, and in the presence of the viral RNA, these rings dislocate to form a helix of 16.33 subunits/turn. The viral RNA possesses a specific protein-binding motif, which binds into a groove within the helix, and additional disks are then recruited and dislocated as the remainder of the RNA is wound up inside the extending helix until the entire genome has been encapsidated. Isometric virus particles appear to follow similar spontaneous reassembly processes, with the subunits taking up a basic 60T structure, where T is an integer.

2.1.3 Viral Movement Within Plants and Transmission between Plants

To spread within plants, viruses must move from the initial infection site systemically to other parts of the plant. The majority of viruses are initially introduced into epidermal or mesophyll cells by insect vectors and must then spread through the leaf tissue from cell to cell and then into the phloem. This short-distance movement between cells occurs through symplastic connections known as *plasmodesmata*. However, the normal size exclusion of plasmodesmata only allows solutes of less than 1.5-nm radius to pass, which is too narrow for the passage of intact virions. Viruses therefore encode movement proteins, which widen these channels to allow either intact particles or ribonucleoprotein complexes to pass through. The main features of these movement proteins are that they are nonstructural nucleic acid binding proteins that are often not virus-specific and can facilitate the movement of unrelated viruses. They appear to work by one of two mechanisms. One group, typified by the TMV 30-kDa protein, act as chaperones, binding to the viral RNA, and guiding it first toward the plasmodesmata, and then through them in a process controlled by phosphorylation mechanisms. The second group appears to form a tubule structure through the plasmodesmata and then forces the 40- to 80-nm diameter intact virions through these by a ''treadmilling'' mechanism. Once the viruses or ribonucleoproteins have passed through the plasmodesmata, the movement proteins return these channels to their normal size exclusion to prevent leakage of other solutes. Eventually, cell-to-cell movement results in the viruses or ribonucleoproteins entering the vascular system, and, once there, they move from source to sink tissues along with photoassimilates. They can also be unloaded back into nonvascular tissue, essentially by the same short-distance movement mechanisms.

To move between plants, most viruses recruit vectors and, in particular, arthropods such as aphids, whiteflies, and hoppers. The relationship between the virus and its vector is generally a specific one, and the amount of genetic variation that can be tolerated in a viral genome is often determined by the need to maintain this specific virus/vector relationship. Traditionally, viruses have been categorized as nonpersistent, where the virus associates with the mouthparts of the insect rapidly, but is lost equally rapidly when the insect

feeds again, and persistent, where the virus is taken up into the gut and hemolymph during feeding and then enters the salivary glands of the insect and so is transmitted every time the insect feeds subsequently. Persistent viruses in turn are subdivided into propagative, which can replicate in the insect as well as the plant, and circulative, which only replicate in plants.

Many viruses, whether transmitted by insects, nematodes, or other vectors encode polypeptides and/or polypeptide domains within their coat proteins that are essential for the transmission. These domains encompass specific motifs that are required for protein-protein interactions and ensure that the virus binds to specific proteins within the insect. There are also a significant number of viruses (around 20%) that are transmitted through seed rather than by vectors, either moving into the gametes prior to fertilization, or directly infecting the embryo after fertilization. This capacity for seed transmission is also controlled by genes within the virus and the host.

2.1.4 Affects of Virus Infection on Plant Processes

Viral infections cause major alterations to host gene expression and metabolic processes. These include accumulation of starch and soluble sugars, increased respiration, decreased photosynthesis, an increase in the levels of amino and organic acids in infected areas, and alterations to plant defense responses. As a viral infection front passes through a plant, it induces a permissive state in cells for its own replication, and the host cells then return to near normality once the front has passed. This permissive state involves host gene "shutoff" in which the majority of genes are temporarily downregulated, except for those that are important for the virus, levels of which remain the same or increase. For example, expression of genes for actin and beta-tubulin, important for cell-to-cell movement of the virus, may remain unchanged, while heat-shock protein gene expression increases, which may be important for ensuring that the changes in gene expression are only transient, and that the host cell does not die.

Although host gene expression effectively returns to normal once the infection front has passed, it is clear that changes do occur to some metabolic processes, as indicated by the changes to respiration, starch accumulation, and so on, and these changes result in the characteristic viral symptoms of patches of chlorosis, often in the form of mosaics. It has been postulated that plant signal molecules such as salicylic acid and soluble sugars may be responsible for these physiological changes.

A further property of plant viruses that has recently been identified is their ability to interfere with plant defense mechanisms. Plants have a specific defense mechanism, posttranscriptional gene silencing (PTGS), which is targeted against foreign RNA such as viral genomes in the cell cytoplasm. This mechanism, like RNAi in animals, involves recognition of dsRNA, (for example, viral replicative forms), and degradation of this RNA by Dicer RNase III-type enzymes. In plants, the resultant small interfering ribonucleoprotein (siRNP) complexes also enter the vascular system and appear to enter a race with the virus, such that if the virus moves faster systemic infection occurs, while if the silencing mechanism catches the virus, infection is aborted. What has now been shown is that some viruses encode within their genomes specific proteins, such as the *Potato virus Y* (PVY) HC-Pro, that are able to block both the initiation and

maintenance of PTGS in a counterdefense strategy.

2.2 Bacteria

2.2.1 General Properties of Plant-Pathogenic Bacteria

Most plant-pathogenic bacteria are rod-shaped and Gram negative and possess flagella, and the best studied at the molecular level are species of *Pseudomonas*, *Xanthomonas*, *Erwinia*, *Pantoea*, *Ralstonia*, and *Agrobacterium*. Prior to colonization, bacteria that are potential pathogens may inhabit the leaf surface or colonize the soil surrounding roots at low density. For pathogenicity to occur, specific pathogenicity factors, often encoded in pathogenicity islands (PAIs) within the bacterial nucleic acids, are required. These factors may be exopolysaccharides, toxins, cell wall degrading enzymes (CWDEs), plant hormones, or other effector proteins, which enable the bacterium to access nutrients for its own growth, while at the same time suppressing or evading the plant defense mechanisms. Furthermore, these factors are often delivered into the host cell cytosol by specific secretion mechanisms, and the whole process is initiated through complex molecular signaling within bacterial populations, and between bacteria and plants.

2.2.2 Establishing Infection

A common mechanism used by bacteria that has been shown to be important for initial plant colonization is quorum sensing. This is a communication mechanism used by bacteria that enables them to sense population density and respond through the regulation of expression of specific genes. In Gram-negative systems, the bacteria produce autoinducers, through the LuxI family of acyl-homoserine lactone (AHL) synthases, which are diffusible signal molecules that can pass in and out of bacterial membranes. At high cell density, these autoinducers reach a threshold level in the external environment that is detected by *LuxR*-encoded transcriptional activators present in the bacterial membranes, which regulate transcription of target genes.

In many cases, further regulatory complexities involving chemical signals emanating from plants have been added to the basic LuxI/LuxR system, such as in the regulation of CWDE and exopolysaccharide (EPS) production of *Erwinia* spp., *Pantoea* spp. and *Ralstonia solanacearum*, and in the conjugal transfer of the Ti (tumor inducing) plasmid in *Agrobacterium tumefaciens*. In *R. solanacearum*, for example, swimming motility mediated by flagella is regulated through PhcA, a Lys-R type global regulator, expression of which is in turn controlled by a quorum sensing mechanism. At low bacterial cell density, as occurs in the rhizosphere, PhcA is not expressed so that the motility and *CWDE* genes are switched on. As bacterial population density increases, for example, through colonization of the xylem, PhcA is expressed, the motility and *CWDE* genes are downregulated, and production of the EPS pathogenicity factors is upregulated.

2.2.3 Pathogenicity Factors

Bacteria utilize a wide range of pathogenicity factors and effector molecules to colonize plants, and different secretion mechanisms are used to deliver these into the plant cytosol. Necrotrophic bacteria, such as the soft rots *Erwinia carotovora* and *Erwinia chrysanthemi* secrete cocktails of CWDEs, including pectinases,

cellulases, and proteinases to soften and macerate the plant tissue and release nutrients. The proteinases, which contain C-terminal secretion signals, are secreted by a type I secretion mechanism driven by ATP hydrolysis, while the pectinases and cellulases are secreted by the type II Out secretion mechanism. Regulation of *CWDE* gene expression is complex and involves plant cell wall breakdown products as well as other plant factors, and environmental factors such as temperature, osmolarity, and nitrogen starvation.

Bacterial toxins, which generally induce chlorosis in plants and increase the severity of many diseases, are important pathogenicity factors, particularly of many *Pseudomonas syringae* pathovars. Coronatine, for example, which has structural similarities to methyl jasmonates (a signaling molecule in higher plants), causes cell wall thickening and changes in chloroplast structure, and its production is regulated through a two-component histidine kinase regulatory system in response to temperature changes. Phaseolotoxin, produced by halo blight of beans, is also temperature regulated and inhibits the enzyme ornithine carbomyltransferase (OCTase) in plants, resulting in arginine deficiency and ornithine accumulation. The bacterium itself copes with this toxin by producing an isozyme of OCTase that resists the toxin.

While these toxins are important for increasing the severity of many pseudomonas diseases, they are not essential and are only part of a complex of effector molecules that these and other plant pathogenic bacteria use to cause disease. Many of these effectors appear to be delivered into plants by a type III secretion mechanism, in which a structure is produced that spans the bacterial membranes and forms an external hrp-pili appendage that acts as a needle through which the effectors are translocated (Fig. 2). These secretion systems are encoded by clusters of genes referred to as *hrp* (hypersensitive response and pathogenicity) genes analogous to the virulence associated type III secretion genes of animal pathogenic bacteria such as *Yersinia* spp. Expression of the *hrp* genes is regulated by environmental and host plant factors, and it has been shown that they are not induced in rich artificial media, but are in minimal media that mimics the content of plant apoplastic fluids. In *R. solanacearum,* for example, the key regulatory protein PrhA is induced by plant cell wall factors, though, interestingly, this occurs with both host and nonhost cell wall extracts, suggesting that the host range of the bacterium is determined by the factors that are secreted and not by the ability to form a secretion mechanism. In *Erwinia* spp. and *Pseudomonas* spp., regulation of *hrp* genes is by a different mechanism involving responses to low levels of carbon and nitrogen, low pH, and low temperature.

The effector molecules secreted by type III systems are diverse in structure and function. Some of these were initially characterized as avirulence proteins because they were identified as components of gene-for-gene resistance (Sect. 3.4). However, it is now apparent that their main role may be to suppress plant defense responses to ensure survival of the colonizing bacteria, and their classification as avirulence peptides is a consequence of the ability of plant defense surveillance mechanisms to detect them. For example, AvrPtoB from *P. syringae* has been shown to act as a general inhibitor of programmed cell death, while others have nuclear localization signals and are suggested to regulate host gene expression.

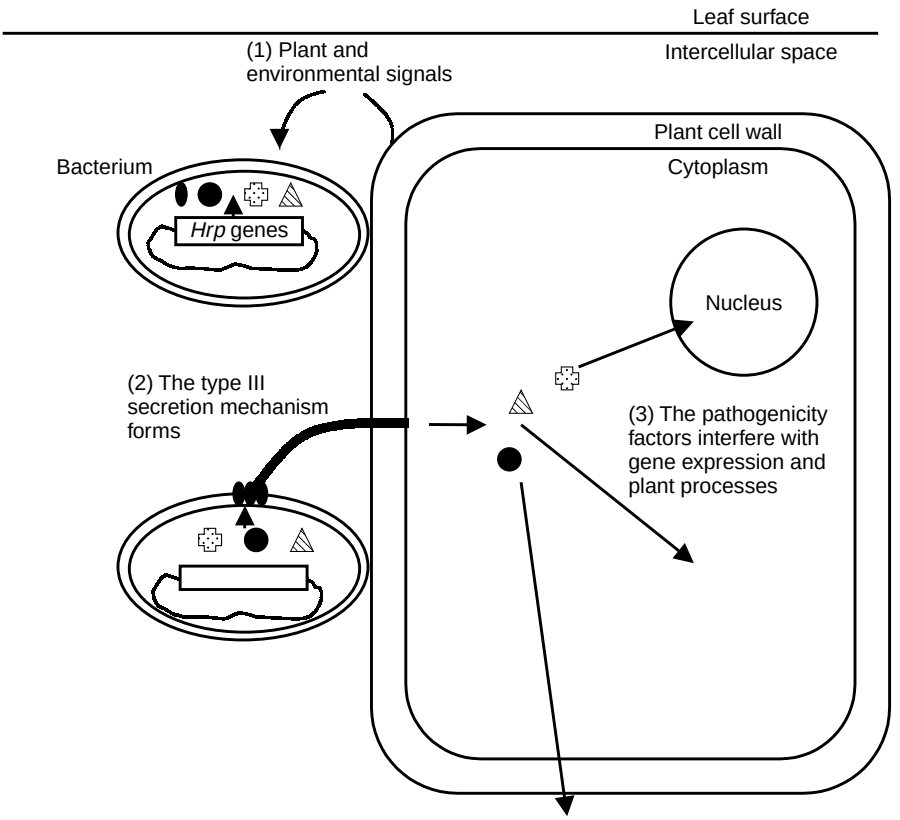

Fig. 2 Model for the infection process of a bacterial plant pathogen. Plant and environmental signals regulate the *hrp* genes (1) resulting in the formation of the type III secretion mechanism and hrp pilus (2) through which pathogenicity factors are secreted that interfere with plant processes, (3) facilitating bacterial multiplication and disease.

The role of many of the other effectors secreted by type III systems has yet to be established.

In some species of *P. syringae*, such as *P. syringae* pv. *savastanoi*, the hrp system is involved in the secretion of enzymes involved in the biosynthesis of auxins. The resultant auxin production in infected plants then proceeds in an uncontrolled way, since this biosynthetic pathway is different to the one normally used by plants for auxin production, and results in gall formation. Equivalent auxin biosynthetic genes and gall production are also a feature of the crown gall bacterium *Agrobacterium tumefaciens*. However, in this bacterium, the genes are carried on a 180-kb pathogenicity island referred to as the Ti (tumor-inducing) plasmid, and it is the genes themselves that are secreted into the plant by a type IV secretion mechanism. This type IV secretion is encoded by a set of genes on the Ti plasmid called the *vir* genes, and the expression of these is regulated by chemical signals emanating from wounded plants such as lignin and flavonoid precursors. These wound signals are detected and transmitted via

a two-component histidine kinase regulatory system (VirA and VirG) to switch on the other *vir* genes. These gene products, in turn, excise a fragment of DNA (the T-DNA) encoding the auxin biosynthetic genes and opine biosynthetic genes from the Ti plasmid, and this DNA is then enveloped in other Vir proteins and secreted out of the bacterium through the T-transporter (assembled from further Vir proteins), which is a pilus that acts as a conduit into the host cytoplasm. The T-complex then targets and enters the plant nucleus, and the T-DNA integrates into the genomic DNA facilitated by plant proteins. The auxin and opine biosynthetic genes are then expressed, resulting in the gall formation and biosynthesis of opines, which are modified amino acids that the bacterium uses as a source of nutrients as it proliferates in the gall.

A final class of virulence factors used by some bacteria such as *R. solanacearum, Xanthomonas campestris, Pantoea stewartii,* and *Erwinia amylovora* is the production of exopolysaccharides (EPS). Gene expression for the biosynthesis of these is upregulated in response to quorum sensing (see Sect. 2.2.2) and the resultant sugar polymers appear to have a double role in pathogenicity. Firstly, they produce a fluidal slime barrier around the bacterium to protect it against desiccation on leaf surfaces and so on, and, secondly, they provide a defense against toxic plant compounds and plant defenses. The regulation of these to ensure that they are only produced when the bacteria reaches a high cell density may be a mechanism to ensure that plant defenses are not activated too early in invasion, so that the bacteria have time to prepare themselves for a "mob" attack.

2.3 Fungi and Oomycetes

2.3.1 General Properties

Fungi form the most extensive group of plant pathogens, and of the more than 74 000 known species, more than 10 000 can parasitize living plants to varying degrees. In addition to the true fungi (which characteristically produce mycelia containing glucans and chitin), a number of organisms that were once classified as fungi, but do not contain chitin, also cause plant disease. These are the oomycetes, which are now classified in the Chromista kingdom, and contain such notorious pathogens as *Phytophthora infestans* (potato blight), the white rusts, the water moulds, and the downy mildews. However, both oomycetes and true fungi have similar mechanisms of pathogenicity and infection strategies in that they are dispersed as spores, which may be carried long distances by wind and air currents or may be deposited in the soil and/or splash dispersed, so they will be considered together in this section.

To colonize plants, fungi and oomycetes adopt a range of infection strategies. Some, such as certain blights, are necrotrophic and use CWDEs and toxins to break down plant cells to release nutrients for their own growth. Others, such as the rusts and mildews, are obligate biotrophs in that they can only grow on living plant cells, and tap into these to obtain their nutrients. Others, often referred to as *hemibiotrophs*, use a combination of these approaches, with an initial biotrophic phase as they establish infection becoming necrotrophic as the infection progresses. A further group of fungi, such as ergot, cause limited damage to plants, but are classified as plant pathogens because they produce

mycotoxins that can be harmful to human and animal health.

2.3.2 Establishing Infection

Once dispersed, the spores of a fungus or oomycete must eventually find a suitable host plant, germinate, and penetrate to access the nutrients. Spores of fungi that are dispersed in the air are produced in vast quantities so that by chance some of them will be deposited on suitable host plants. In the presence of moisture, a key requirement for all spore germination and penetration, adhesive pads are formed on the leaf surface containing adhesins and adhesive materials released from the leaf through esterase and cutinase action, which may also act as the chemical triggers of germination. Spores of fungi that are soilborne may remain dormant in the soil for many years until they detect specific chemical signals and/or electrical currents generated around plant roots. In the case of many oomycetes, these signals result in germination of the dormant zoospores, which produce flagella and swim through moisture in the soil toward the source of the signal.

Once the spore has adhered and germinated, it must grow into the plant. In many cases, this involves differentiation of the germ tube to form a specific swelling known as the *appressorium*, and, depending on the fungal species, this may result in direct penetration through the plant cuticle, or through natural openings such as stomata, or wounds, which the germ tube may grow toward directionally in response to chemical and/or physical cues on the leaf surface. One of the best-studied model systems for the mechanisms of direct penetration is the ascomycete *Magnaporthe grisea* that causes the rice blast disease. Here it has been shown that transmembrane integrin proteins sense the leaf surface topography and hardness and transmit this signal through the cell wall, where this and other environmental cues result in changes to the heterotrimeric G proteins inside the cell. These are important intermediates in a number of signaling pathways involving Ca^{2+}-dependent protein kinases, cAMP and protein kinase A, and mitogen-activated protein kinases (MAPK). These result in changes to gene expression such that the tip of the germ tube stops elongating and forms the appressorium, which becomes separated from the germ tube by a septum. Hydrophobins accumulate and self-assemble in this appressorium wall, which then becomes melanized, apart from a small zone in contact with the leaf surface, which forms the appressorial pore. Trehalose is broken down within the fungus along with glycogen granules and lipid bodies, resulting in accumulation of glycerol to concentrations of up to 3.2 M. Moisture surrounding the germinated spore passes through the appressorial wall as these osmotically active solutes accumulate and turgor pressure builds up, which is translated into a mechanical force that pushes the penetration peg out through the appressorial pore and through the plant cuticle. While this physical force may be enough to penetrate the cuticle for some fungi, in other cases not enough force is generated and CDWE are also produced to soften the plant tissue and aid the penetration. These include pectinases, cutinases, and cellulases, and most fungi have multigene families encoding these enzymes.

Having breached the outer defenses, the fungus may then need to overcome further plant defenses to establish infection. These will include preformed antimicrobial compounds (phytoanticipins) and

induced compounds (phytoalexins) (see Sect. 3.2). Some fungi, such as species of *Colletotrichum* and *Botrytis* enter a period of quiescence and wait until the levels of these compounds decline during fruit ripening before they become aggressive. Others such as *Gaeumannomyces graminis* (take-all of cereals) encode enzymes (avenacinase) that detoxify the phytoanticipins (avenacin) or phytoalexins that are produced, and it has also been suggested that the breakdown products of these may in some cases induce a signal transduction mechanism in the host plant that results in suppression of plant defense responses. A further strategy used by *M. grisea* and some other fungi is to use ATP-binding cassette (ABC) transporters to export the antimicrobial compounds out of the fungus, while others secrete compounds (oxalic acid from *Sclerotinium sclerotiorum* or mannitol from *Alternaria alternata*) to suppress active oxygen species-mediated defenses.

2.3.3 Pathogenicity Factors of Necrotrophs

Once established in a plant, a necrotroph aims to kill the host plant cells and utilize the nutrients released. Usually, the resultant macroscopic symptoms of this are brown necrotic areas of dead cells in and around the site of infection. The main weapons used by these fungi are CDWEs, as discussed previously, and toxins. There may also be a role for a group of proteins referred to as the *NLP family* (Nep1-like protein) that have been identified in the range of fungi such as *Fusarium oxysporum*, oomycetes such as *Phytophthora* and *Pythia* species, and some necrosis-inducing bacteria. These proteins appear to induce a programmed cell death in plants by triggering defense responses, although the advantages of this to fungal pathogenicity are uncertain.

Fungal toxins can be divided into three groups, the host-selective toxins, produced mainly by *Alternaria* and *Cochliobolus* species, which are only toxic to the hosts for these fungi, the nonselective toxins that can cause damage to a wider range of plant species, and the mycotoxins, which cause no damage to the plants, but can be fatal to the animals and humans that consume them. The host-selective toxins, such as T-toxin and Victorin of *Cochliobolus* spp. and AAL of *Alternaria alternata* have a broad range of chemical properties (T-toxin is a polyketol, Victorin a chlorinated peptide, and AAL a sphinganine-related compound), and act by affecting mitochondria or host plasma membranes to cause leakage of electrolytes and other small molecules to disrupt photophosphorylation and ATP synthesis, resulting in plant cell death. An interesting feature of these toxins is that their biosynthetic genes tend to be clustered into pathogenicity islands in isolates that produce the toxins, and are completely absent in non–toxin-producing isolates, suggestive of acquisition by horizontal gene transfer. Host nonselective toxins generally have less severe affects on plants, causing chlorotic lesions and wilt, often through the induction of reactive oxygen species in the plant that cause membrane breakdown and nutrient leakage. The mycotoxins by contrast have no obvious affects on plants, but the ergot alkaloids, aflatoxins, trichothecines, and fumonisins that fit into this definition are amongst the most potent toxins known in animals, causing gangrene and convulsions (ergotism) and liver damage and cancers (aflatoxins). Interestingly, the biosynthetic genes for production of these toxins are also generally in pathogenicity islands, although the benefit to the fungus in maintaining these gene clusters is not obvious.

2.3.4 Pathogenicity of Biotrophs

Obligate biotrophic fungi such as the rusts, powdery mildews, and downy mildews use stealth to obtain their nutrients from plants. The concentrations of sugars and amino acids in the apoplast of plants are low and not conducive to rapid fungal growth, so most of these organisms form specialized feeding structures known as *haustoria* to invaginate into nutrient-rich plant cells (Fig. 3). Interestingly, convergent evolution has meant that basidiomycete fungi (rusts), ascomycete fungi (powdery mildews), and oomycetes (downy mildews) have all developed this same strategy for acquisition of their nutrients.

Haustoria form by puncturing the plant cell wall, but not the plasma membrane, which remains intact and surrounds the haustoria to form the extrahaustorial membrane. How this occurs without the plant defense responses being triggered is unclear, but analysis of gene expression within haustoria suggests that suppressors, such as bax inhibitors of cell

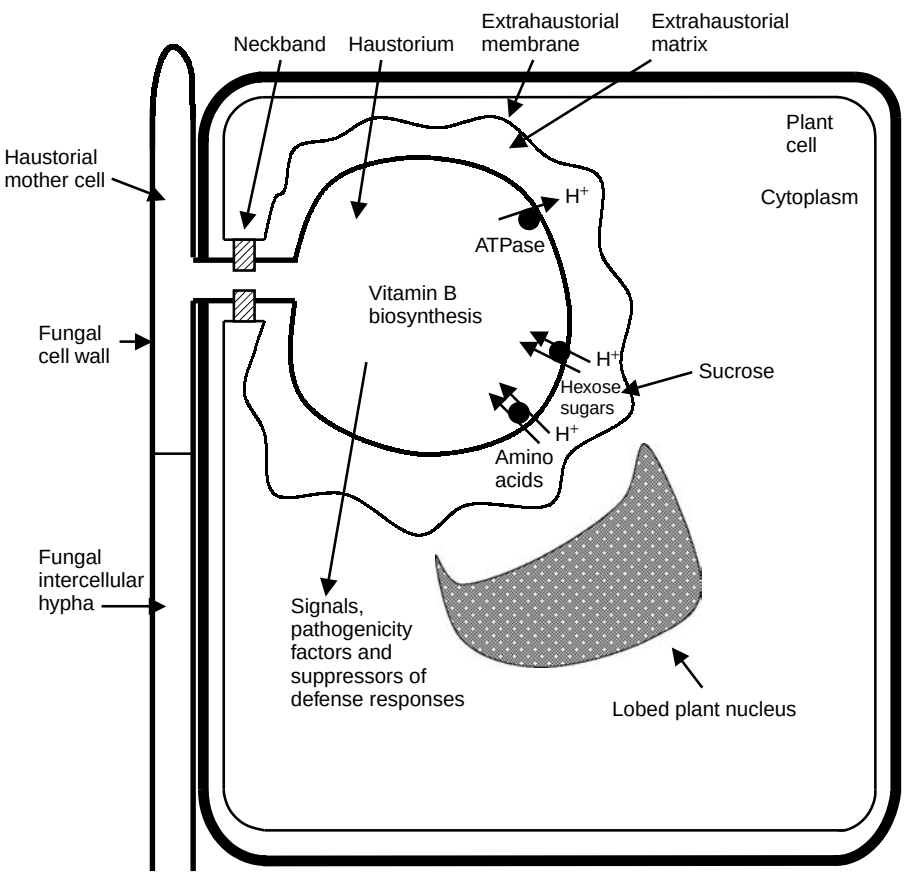

Fig. 3 General properties of a stylized haustorium. The haustorium forms as an invagination into the plant cell and absorbs nutrients (e.g. hexoses and amino acids from the extrahaustorial matrix via proton-symport transporters. Vitamin B biosynthesis also occurs, and signals, pathogenicity factors, and suppressors of defense responses are secreted into the plant cell.

death may be produced by the fungus and secreted along with other signals. Other mechanisms for evading defense responses have also been identified, such as alterations to the cell wall composition of infection structures so that they are not recognized by the plant, controlling levels of proline in plant cells to reduce stress responses, or disrupting plant cell wall/plasma membrane connections (Hechtian strands) so that cell wall associated defense responses are not activated.

For nutrient acquisition, it has been shown that plasma membrane H^+-ATPases, hexose transporters, and amino acid transporters are highly expressed in haustoria and are involved in nutrient transfer into the fungus involving a proton-symport mechanism. To access sucrose from plant cells, for example, it appears that the fungus may secrete an invertase to convert the sucrose to glucose and fructose, which are then taken up by the hexose transporters. Gene expression for thiamine (vitamin B1) biosynthesis, an essential cofactor for several enzymes in carbon metabolism, is also upregulated in haustoria.

A further role for the signals produced by haustoria appears to be to delay senescence in plant cells surrounding the haustoria, resulting in "green islands" around infection sites. It has been suggested that the levels of plant hormones and polyamines are being deliberately manipulated in these cells by the fungus to keep them alive and photosynthetically active. The roles of many other effector molecules secreted into plants by fungi and oomycetes have yet to be established, and, as with bacterial effectors (see Sect. 2.3.2), some of these are being identified indirectly as avirulence gene products. In the fungus *Cladosporium fulvum*, for example, a number of avirulence peptides have been identified and are generally small cysteine rich peptides. Of these, it has been proposed that Avr9 may have a housekeeping role in nitrogen metabolism, while Avr4 has chitin-binding properties. In *Rhynchosporium secalis*, the Nip1 avirulence protein is a host nonselective toxin, and mutants have significantly reduced virulence. Over the next few years, as the genomes of more of these biotrophic organisms are sequenced, it is likely that the roles of many more of these effectors in nutrient acquisition and suppression of defense responses will be established.

3 Plant Defense against Disease

3.1 General Properties

Having discussed the strategies used by pathogens to cause disease, we should now consider how plants defend themselves. There are multiple layers to these defenses, as outlined in Fig. 4, and they can be broadly divided into constitutive and induced mechanisms. In many plant species, a combination of these mechanisms will ensure that the particular species is not infected by a particular pathogen (nonhost resistance). For example, the potato blight oomycete is unable to infect wheat. In other cases, the mechanisms may result in particular genotypes within a plant species being resistant, while others are susceptible (host resistance).

3.2 Constitutive Defenses

The first barrier that defends plants against microbes is the wax layer present on the cuticle. This forms a water-repellent

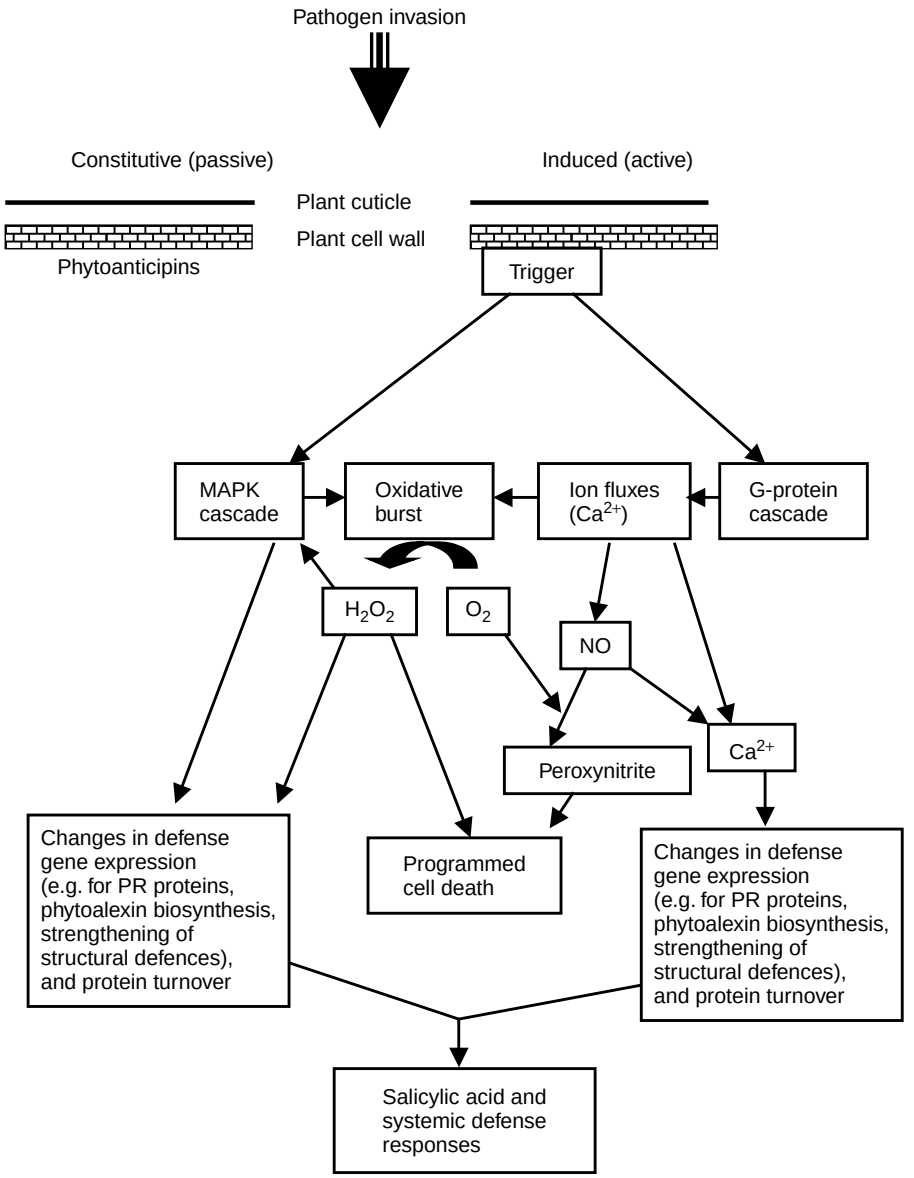

Fig. 4 The multiple layers to plant defenses and signaling. Constitutive defenses are indicated on the left-hand side of the diagram, while induced defenses are shown down the right. The scheme also indicates the interplay between different signaling pathways (see text for details).

surface that prevents the formation of water droplets necessary for bacterial ingress and fungal spore germination. In addition, the epidermal cell walls form an impenetrable barrier to all but those fungi that have the necessary mechanisms for direct penetration (see Sect. 2.3.2). This leaves wounds and natural openings, such

as stomata, as the main weaknesses in structural defenses.

Should a pathogen breach these defenses, antimicrobial chemicals within the plant form a second line of defense. Low molecular weight secondary metabolites such as terpenoid derivatives, saponins, aliphatic acid derivatives, phenolics, phenylpropanoids, alkaloids, and sulfur-containing compounds sequestered in plant cells have been shown to be antimicrobial, either as preformed substances (phytoanticipins) or when their biosynthetic pathways are induced as part of induced resistance (phytoalexins).

3.3 Induced Mechanisms

Induced resistance involves the plant recognizing that it is being invaded and mounting a defense through a triggering of changes in gene expression and production of antimicrobial compounds (Fig. 4). This response is often referred to as the *hypersensitive response* (HR) and involves kinase signaling cascades, ion fluxes, the oxidative burst, nitric oxide (NO), programmed cell death, induction of genes for the biosynthesis of phytoalexins, genes for the strengthening of structural defenses such as peroxidases and lignin biosynthetic enzymes, and production of various novel proteins known collectively as the *pathogenesis-related* (PR) *proteins*. These PR proteins have been grouped into at least 14 classes, and while the biochemical roles of some are not clear, others such as chitinases (PR-3, PR-4, PR-8, PR-11), glucanases (PR-2), proteinases (PR-7), and RNAses (PR-10), are presumed to have roles as hydrolytic enzymes against the invading pathogen, while the defensins (PR-12) and thionins (PR-13) are classified because of their similarities to these antimicrobial compounds found in insects and mammals. The induction of this same battery of defenses appears to occur whether the invading pathogen is a virus, bacterium, fungus or oomycete, and the macroscopic symptoms are small necrotic lesions on the plant. The overall strategy appears to be to combat the pathogen with a range of antimicrobial compounds, and, at the same time, invoke programmed cell death in cells surrounding the site of infection and changes to structural defenses to produce a barrier against further invasive growth.

A further consequence of local induction of the hypersensitive response is a systemic resistance activated throughout the plant, termed *systemic acquired resistance* (SAR). This has some conceptual similarities to the immune response in animals, in that it is long lasting, and can be boosted by repeat inoculations, but, unlike animal immunity, it can be effective against a range of pathogens, conferring broad spectrum resistance. Salicylic acid is a key signaling molecule in SAR and specific genes such as *PR-1, PR-2,* and *PR-5* are taken as markers for this form of resistance. Interestingly, necrosis-inducing pathogens such as *Alternaria* spp. induce a different systemic defense mechanism characterized by jasmonate and ethylene signaling and different *PR* gene induction, while nonpathogens such as the plant growth – promoting rhizobacteria induce a third type of systemic resistance, induced systemic resistance (ISR). Some components of these different systemic mechanisms overlap, while others are antagonistic, but the overall result appears to be to prime host cells in systemic tissues for a more rapid future deployment of defenses. There is also some evidence that these systemic signals can result in the production of volatile signaling molecules,

such as methyl jasmonates and methyl salicylates, that are detected by adjacent plants to result in a form of "communal" resistance.

3.4 Triggers for Induced Resistance

For induction of resistance, plants must be able to recognize and respond to the invading organism. The triggers for this appear to be a group of constitutively produced peptides capable of interacting with and responding to elicitors emanating from the pathogens. In nonhost resistance, these elicitors may be general proteins such as flagellin from bacteria, or oligomers of chitins and glucans from fungi, that the plant recognizes as "pathogen associated molecular patterns" (PAMPs), as part of a defense mechanism with similarities to innate immunity (the ability to recognize self from nonself) in animals. However, as mentioned in Sect. 2.3.4, some organisms have evolved the ability to mask these elicitors and/or suppress the defense responses. When this has happened, the second form of induced resistance, host or gene-for-gene resistance may have evolved. Here, the plant encodes specific "receptors" that detect the presence of specific elicitors from a pathogen. These elicitors may be individual viral proteins (coat protein, replicase etc.), proteins secreted from bacteria by type III secretion mechanisms, or those secreted through fungal and oomycete haustoria, that the plant has found a way of recognizing. Because they are specific elicitors, they will be encoded by specific genes (referred to as *Avr* genes), and the recognition protein in the plant will similarly be encoded by a specific gene (the *R* gene).

This is the basis for the gene-for-gene concept where resistance will occur as long as the plant has the dominant allele of the specific *R* gene that recognizes the *Avr* gene-encoded elicitor. However, if the plant has no dominant *R* allele, or the pathogen loses this elicitor (through selection and/or mutation), disease will occur, although if the avirulence elicitor is a key component of pathogenicity, loss of function may result in reduced pathogenicity. This, in turn, may result in an evolutionary "arms race" developing with complementary changes occurring in the plant and pathogen populations. In natural plant/pathogen ecosystems, this often results in evolutionary stability, but, in agriculture, where resistance is bred into commercial cultivars that are used as extensive monocultures, it can be responsible for severe epidemics if the resistance gene is overcome. There are, however, some *R* genes that have been shown to be highly effective and long lasting in crop protection, and this has led to the concept that *R* genes can be subdivided into those that are that are slow-evolving and presumably recognize elicitors that are essential pathogenicity factors, and those that are rapidly evolving and presumably detect avirulence elicitors that have little fitness penalty on the pathogen.

3.5 The *R* Genes

Because of the genetic simplicity of the gene-for-gene concept, these major resistance genes have been the focus of extensive research to understand the molecular basis of disease resistance. As a result, a significant number of *R* genes have been cloned in the last decade from a number of plant species and against a range of viral, bacterial, fungal, and oomycete pathogens. These genes are often found clustered in the plant genomes as families

residing at complex loci, and the loci may include apparently inactive resistance gene analogs (RGAs), sequences disrupted by frameshifts, mutations, and transposons, and *R* genes capable of recognizing more than one pathogen.

On the basis of their sequences, *R* genes have been grouped into a number of classes, with the largest class encoding a nucleotide binding site plus a carboxy-terminal leucine-rich repeat domain (NBS-LRR). Some of these also possess a domain with toll/interleukin-1-receptor homology (TIR) at their amino terminal (the TNLs), while others have a loosely predicted coiled-coil (CC) domain (the CNLs). A smaller class of resistance genes possess serine-threonine kinase domains either physically linked to an LRR domain (e.g. the *Xa21* gene from rice) or functionally linked (the *Pto* gene from tomato linked to *Prf*). Other smaller classes of resistance genes have more unique characteristics.

The leucine-rich repeat domains, which consist of multiple serial amino acid repeats containing leucines at regular intervals, have similarities to domains in mammalian proteins involved in protein–protein recognition, and it has therefore been postulated that they are involved in recognition of the pathogen elicitors. Evidence to support this comes from domain swaps between resistance genes such as the *L* alleles of flax, where they determine the rust resistance specificity, although in some cases regions of the NBS also appear to be important. The predicted cellular locations of the LRR domains add further support to a role in recognition. Those that detect apoplastic fungi, such as *C. fulvum*, have an extracellular location, while most others have a cytoplasmic location in which they are most likely to detect viruses, bacterial effector proteins, and secreted fungal peptides.

Attempts to confirm that R proteins interact directly with avirulence proteins have, however, been less conclusive. The rice Pi-ta protein has been shown to interact with the rice blast Avr-Pita protein in yeast two-hybrid experiments, but in most other cases such interactions have not been found. Furthermore, the bacterial effector protein AvrPto interacts with the Pto resistance protein of tomato, which has no LRR domain. A proposed model to account for this is the "Guard hypothesis" in which the LRR-type resistance proteins (in the case of AvrPto/Pto, this might be Prf) is there to guard a key plant protein and to detect the interaction between this and the pathogen elicitor. Further evidence that R protein/Avr protein recognition involves a number of other protein–protein interactions come from studies on *Pseudomonas syringae* and *Arabidopsis*. Here, a protein Rin4 has been shown to interact with AvrB and AvrRPM1, and also with the resistance gene product RPM1 that recognizes both of these avirulence peptides genetically. The Rin4 protein also interacts with a separate resistance gene protein RPS2, but in an inhibitory way such that removal of Rin4 when the complementary avirulence peptide AvrRPt2 is present results in RPS2 activation. This clearly indicates that R protein/Avr protein interactions are often more complex than a simple "receptor"/ligand model, and involve multiple protein/protein interactions.

3.6 Downstream Signaling of Resistance

The other domains present in resistance gene proteins are predicted to have functions in the signaling that results in the HR (see Sect. 3.3). Protein kinases have known roles in signaling. TIR domains

have homology to the cytoplasmic signaling domains of the *Drosophila* toll protein and interleukin-1-receptors of birds and animals, which are part of the innate immune response, while NBS domains have ATP- or GTP-binding activities that regulate protein–protein interactions and occur in apoptosis proteins such as APAF-1 (apoplastic protease activation factor) in humans and CED-4 (*Caenorhabditis elegans* death regulator protein 4) in *C. elegans*.

To identify components of *R* gene signaling pathways, genetic analysis, particularly in *Arabidopsis*, has been important. Mutants have been identified, for example, with increased susceptibility, or deficiencies in pathogen-induced phytoalexin accumulation, and these have been mapped on to signaling pathways. Figure 4 indicates a simplified scheme for how these signaling pathways proposed in plant defense responses overlap to culminate in changes in gene expression, programmed cell death, and systemic signaling.

As in all eukaryotes, MAPKs are key signaling components involved in regulating target gene expression in response to signals, and specific MAPKKK, MAPKK, MAPK, and WRKY-type transcription factors have been identified associated with some signaling cascades such as that involved in the flagellin receptor FLS2. In addition, one of the most rapid measurable changes that occurs to a plant during the hypersensitive response is a large transient ion flux across plasma membranes, in particular, K^+ efflux and Ca^{2+} influx. Resistance genes interacting with elicitors directly or indirectly are thought to modulate this through G-protein signaling cascades operating through phospholipase C, diacylglycerol (DAG), and inositol 1,4,5-triphosphate (IP_3), and this in turn activates kinase-signaling cascades to regulate target gene expression.

Along with calcium fluxes, reactive oxygen species (ROS) are rapidly induced as part of HR via NADPH/NADH oxidases or via apoplastic peroxidases. The resultant hydrogen peroxide may have direct antimicrobial affects, but is more likely to be involved in signaling, activating MAPK cascades, interacting with NO signaling, and setting off systemic defense responses. In addition, the hydrogen peroxide reacts through the Fenton cycle of reactions with intracellular transition metals to form the hydroxyl radical, which causes DNA fragmentation and cell damage, both in the invading pathogen and in plant cells to form a barrier against further pathogen ingress. Reactive oxygen species may also have a role in regulating protein turnover; proteins with CHORD (cysteine and histidine rich) domains, such as RAR1 in barley, have been identified in defense responses with proposed roles in coordinating the degradation and turnover of proteins involved in defense signaling. RAR1 interacts with the protein SGT1, which regulates a group of RING-type E3 ubiquitin ligases. These are the specificity determinants that mediate transfer of ubiquitin to target proteins, which undergo degradation at the 26S proteasome. However, it is not yet known which proteins are the targets for this process in defense responses. ROSs also work in conjunction with NO to regulate plant cell apoptosis and signaling, and NO may also interact with oxygen to form peroxynitrite, which has direct antimicrobial properties.

Downstream, low molecular weight signaling molecules such as salicylic acid and jasmonic acid are produced and are involved in both local and systemic resistance responses. Salicylic acid is synthesized from chorismate in plants by two pathways, one involving phenylalanine and the enzyme phenylalanine ammonia lyase

(PAL), which is known to be upregulated as part of HR. The other pathway involves isochorismate synthase, which has been identified in *Arabidopsis* through mutant screens for enhanced disease susceptibility. Jasmonic acid is synthesized by plasma membrane–located lipases, which are believed to be induced by calcium fluxes.

Three key components identified as part of these downstream signaling pathways are EDS1, NDR1, and NPR1. EDS1 has been mapped onto TNL-type resistance pathways and is required, both for HR development and in association with PAD4, for salicylic acid accumulation. NDR1, which has been identified as a transmembrane protein, appears to be involved in CNL-type resistance. Both these pathways converge at NPR1, which is an ankyrin-repeat protein that interacts with transcription factors of the TGA type of basic leucine zippers. These, in turn, appear to bind to the promoter elements upstream of *PR*-protein genes to regulate their expression. Induced systemic resistance also appears to require the NPR1 protein, but uses jasmonates and ethylene as the signals, while the systemic pathways from necrosis-inducing pathogens are independent of NPR1 and may be mutually antagonistic to act through plant defensins and thionins.

3.7 Coordination of Defense Responses

The coordination of these signaling pathways to result in defense responses is complex and, like in animal systems, is believed to involve the mitochondria. In animals, the mitochondria interprets and amplifies stress signals such as calcium fluxes, ion fluxes, ROS, NO, and so on, to result in release from the mitochondrion of cell death activators, inhibitors, and inhibitor derepressors such as cytochrome *c* to regulate apoptosis proteins, caspase cascades, and cell death. In plants, however, there is limited evidence for the involvement of cytochrome *c*, and the caspase-like proteins identified are not direct homologs of those in animals. It has been proposed that this lack of conservation is related to the way that defense responses manifest themselves in plants and animals. In animals, programmed cell death has to be strictly controlled to be rapid and to ensure that the dead cells are removed to prevent inflammation of adjacent cells. In plants, however, the same evolutionary constraints do not apply and a form of programmed oncosis appears to occur where spread to adjacent cells can act to the advantage of the plant by forming a barrier of dead cells through which the pathogen cannot grow.

4 Concluding Remarks

Much has been learnt about the molecular basis of disease resistance in plants and about the mechanisms of fungal, oomycete, bacterial, and viral pathogenicity. The dynamic coevolutionary processes have resulted in multiple facets to the mechanisms, with microbes acquiring pathogenicity factors so that they can colonize plants, plants evolving a complex battery of defensive mechanisms, pathogens developing ways to evade these, and plants developing further layers of resistance. Added to the general mechanisms that have been discussed, there are also examples of resistance conferred by recessive genes, such as the barley *mlo* gene for resistance to powdery mildew, and there are also many cases where resistance

is controlled quantitatively by polygenic resistance mechanisms, and our knowledge of the molecular basis of this form of resistance is in its infancy.

From a practical perspective, understanding the molecular basis of plant–pathogen interactions has a range of applications in crop protection. Molecular diagnostics for detection and identification of organisms can be developed to supplement and enhance traditional diagnostic techniques, and can be used to ensure that appropriate and targeted crop protection strategies are used. Molecular markers linked to traits such as resistance genes can be used in breeding programs to improve commercial cultivars. Molecular genetics can also help with the identification of new resistance genes. Specific target elicitors from pathogens can be screened on different accessions and wild relatives of crop plants using transient expression vectors to identify new sources of resistance for breeding into commercial cultivars. Alternatively, DNA shuffling techniques could be used to create large numbers of recombinant resistance genes from cloned resistance genes and resistance gene analogs, and these could be screened for new specificities.

There are also many technically feasible and tested approaches for controlling disease through the genetic modification of plants, using methods such as pathogen-derived resistance, overexpression of defense-related genes or cloned resistance genes. For example, viral coat protein genes have been introduced into plants to give effective pathogen-derived resistance against many viruses, and the *Xa21* gene for rice resistance to *Xanthomonas oryzae* has been transferred at the Rice Research Institute in the Philippines from the wild species *Oryzae longistaminata* into elite cultivars, a process that has been difficult to achieve through conventional breeding.

Whether molecular biology will be used in the future primarily to support conventional crop protection methods or for the genetic modification of plants will depend on public acceptance and government policies. Appropriate deployment is also crucial to ensure that such techniques do not simply perpetuate the development of new ''super-pathogens'' capable of overcoming the control methods. However, there is no doubt that the use of molecular biology has dramatically increased our understanding of plant-pathogen interactions and will continue to do so as the functions of more genes and gene products are identified in both pathogenicity and resistance mechanisms.

Bibliography

Books and Reviews

Agrios, G.N. (1997) *Plant Pathology*, 4th edition, Academic Press, San Diego, CA.

Cao, H., Baldini, R.L., Rahme, L.G. (2001) Common mechanisms for pathogens of plants and animals, *Annu. Rev. Phytopathol.* **39**, 259–284.

Culver, J.N. (2002) Tobacco mosaic virus assembly and disassembly: determinants in pathogenicity and resistance, *Annu. Rev. Phytopathol.* **40**, 287–308.

Dickinson, M.J. (2003) *Molecular Plant Pathology*, BIOS Scientific Publishers, London and New York.

Kronstad, J.W. (Ed.) (2000) *Fungal Pathology*, Kluwer Academic Press, Dordrecht.

Lucas, J.A. (1998) *Plant Pathology and Plant Pathogens*, 3rd edition, Blackwell Science, Oxford.

Mandahar, C.L. (Ed.) (1999) *Molecular Biology of Plant Viruses*, Kluwer Academic Press, Boston.

Oliver, R.P., Schweizer, M. (Eds.) (1999) *Molecular Fungal Biology*, Cambridge University Press, Cambridge.

Wolpert, T.J., Dunkle, L.D., Ciuffetti, L.M. (2002) Host-selective toxins and avirulence determinants: what's in a name? *Annu. Rev. Phytopathol.* **40**, 251–285.

Primary Literature

Abramovitch, R.B., Kim, Y.J., Chen, S., Dickman, M.B., Martin, G.B. (2003) *Pseudomonas* type III effector AvrPtoB induces plant disease susceptibility by inhibition of host programmed cell death, *EMBO J.* **22**, 60–69.

Aranda, M., Maule, A. (1998) Virus-induced host gene shutoff in animals and plants, *Virology* **243**, 261–267.

Asai, T., Tena, G., Plotnikova, J., Willman, M.R., Chiu, W-L., Gomez-Gomez, L., Boller, T., Ausubel, F.M., Sheen, J. (2002) MAP kinase signaling cascade in *Arabidopsis* innate immunity, *Nature* **415**, 977–983.

Axtell, M.J., Staskawicz, B.J. (2003) Initiation of RPS2-specified disease resistance in *Arabidopsis* is coupled to the AvrRpt2-directed elimination of RIN4, *Cell* **112**, 369–377.

Ayliffe, M.A., Roberts, J.K., Mitchell, H.J., Zhang, R., Lawrence, G.J., Ellis, J.G., Pryor, T.J. (2002) A plant gene upregulated at rust infection sites, *Plant Physiol.* **129**, 169–180.

Azevedo, C., Sadanandom, A., Kitagawa, K., Freialdenhoven, A., Shirasu, K., Schulze-Lefert, P. (2002) The RAR1 interactor SGT1, an essential component of *R* gene-triggered disease resistance, *Science* **295**, 2073–2076.

Berger, S. (2002) Jasmonate-related mutants of *Arabidopsis* as tools for studying stress signalling, *Planta* **214**, 497–504.

Bergelson, J., Kreitman, M., Stahl, E.A., Tian, D. (2001) Evolutionary dynamics of plant *R*-genes, *Science* **292**, 2281–2284.

Bolwell, G.P. (1999) Role of active oxygen species and NO in plant defence responses, *Curr. Opin. Plant Biol.* **2**, 287–294.

Bonas, U., Lahaye, T. (2002) Plant disease resistance triggered by pathogen-derived molecules: refined models of specific recognition, *Curr. Opin. Microbiol.* **5**, 44–50.

Büttner, D., Bonas, U. (2003) Common infection strategies of plant and animal pathogenic bacteria, *Curr. Opin. Plant Biol.* **6**, 312–319.

Dangl, J.L., Jones, J.D.G. (2001) Plant pathogens and integrated defence responses to infection, *Nature* **411**, 826–833.

Deising, H.B., Werner, S., Wernitz, M. (2000) The role of fungal appressoria in plant infection, *Microbes Infect.* **2**, 1631–1641.

De Lorenzo, G., Ferrari, S. (2002) Polgalacturonase-inhibiting proteins in defense against phytopathogenic fungi, *Curr. Opin. Plant Biol.* **5**, 295–299.

Del Sorbo, G., Schoonbeek, H-J., De Waard, M.A. (2000) Fungal transporters involved in efflux of natural toxic compounds and fungicides, *Fungal Genet. Biol.* **30**, 1–15.

Denny, T.P. (1999) Autoregulator-dependent control of extracellular polysaccharide production in phytopathogenic bacteria, *Eur. J. Plant Pathol.* **105**, 417–430.

Devoto, A., Muskett, P.R., Shirasu, K. (2003) Role of ubiquitination in the regulation of plant defence against pathogens, *Curr. Opin. Plant Biol.* **6**, 307–311.

Dixon, R.A. (2001) Natural products and plant disease resistance, *Nature* **411**, 843–847.

Drugeon, G., Urcuqui-Inchima, S., Milner, M., Kadaré, G., Valle, R.P.C., Voyatzakis, A., Haenni, A-L., Schirawski, J. (1999) The strategies of plant virus gene expression: models of economy, *Plant Sci.* **148**, 77–88.

Fan, W., Dong, X. (2002) *In vivo* interaction between NPR1 and transcription factor TGA2 leads to salicylic acid-mediated gene activation in *Arabidopsis*, *Plant Cell* **14**, 1377–1389.

Farmer, E.E. (2001) Surface-to-air signals, *Nature* **411**, 854–856.

Feys, B.J., Parker, J.E. (2000) Interplay of signalling pathways in plant disease resistance, *Trends Genet.* **16**, 449–455.

Francis, M.S., Wolf-Watz, H., Forsberg, A. (2002) Regulation of Type III secretion systems, *Curr. Opin. Microbiol.* **5**, 166–172.

Genin, S., Boucher, C. (2002) *Ralstonia solanacearum*: secrets of a major pathogen unveiled by analysis of its genome, *Mol. Plant Pathol.* **3**, 111–118.

Gómez-Gómez, L., Boller, T. (2002) Flagellin perception: a paradigm for innate immunity, *Trends Plant Sci.* **7**, 251–256.

Grant, J.J., Loake, G.J. (2000) Role of reactive oxygen intermediates and cognate redox signalling in disease resistance, *Plant Physiol.* **124**, 21–29.

Grant, M., Mansfield, J. (1999) Early events in host-pathogen interactions, *Curr. Opin. Plant Biol.* **2**, 312–319.

Gutierrez, C. (2000) DNA replication and cell cycle in plants: learning from geminiviruses, *EMBO J.* **19**, 792–799.

Hacker, J., Kaper, J.B. (2000) Pathogenicity islands and the evolution of microbes, *Annu. Rev. Microbiol.* **54**, 641–679.

Heiser, I., Oßwald, W., Elstner, E.F. (1998) The formation of reactive oxygen species by fungal and bacterial phytotoxins, *Plant Physiol. Biochem.* **36**, 703–713.

Hentschel, U., Steinert, M., Hacker, J. (2000) Common molecular mechanisms of symbiosis and pathogenesis, *Trends Microbiol.* **8**, 226–231.

Herbers, K., Yakahata, Y., Melzer, M., Mock, H-P., Hajirezaei, M., Sonnewald, U. (2000) Regulation of carbohydrate partitioning during the interaction of potato virus Y with tobacco, *Mol. Plant Pathol.* **1**, 51–59.

Hugouvieuxc-Cotte-Pattat, N., Condemine, G., Nasser, W., Reverchon, S. (1996) Regulation of pectinolysis in *Erwinia chrysanthemi*, *Annu. Rev. Microbiol.* **50**, 213–257.

Idnurm, A., Howlett, B.J. (2001) Pathogenicity genes of phytopathogenic fungi, *Mol. Plant Pathol.* **2**, 241–255.

Innes, R.W. (2001a) Targeting the targets of Type III effector proteins secreted by phytopathogenic bacteria, *Mol. Plant Pathol.* **2**, 109–115.

Innes, R.W. (2001b) Mapping out the roles of MAP kinases in plant defense, *Trends Plant Sci.* **6**, 392–394.

Jin, Q., He, S.Y. (2001) Role of the Hrp pilus in type III protein secretion in *Pseudomonas syringae*, *Science* **294**, 2556–2558.

Jones, A. (2000) Does the plant mitochondrion integrate cellular stress and regulate programmed cell death? *Trends Plant Sci.* **5**, 225–230.

Jorgensen, R.A., Atkinson, G.R., Forster, R.L.S., Lucas, W.J. (1998) An RNA-based information superhighway in plants, *Science* **279**, 1486–1487.

Kahmann, R., Basse, C. (2001) Fungal gene expression during pathogenesis-related development and host plant colonisation, *Curr. Opin. Microbiol.* **4**, 374–380.

Kim, Y.J., Lin, N.C., Martin, G.B. (2002) Two distinct *Pseudomonas* effector proteins interact with the Pto kinase and activate plant immunity, *Cell* **109**, 589–598.

Kjemtrup, S., Nimchuk, Z., Dangl, J.L. (2000) Effector proteins of phytopathogenic bacteria: bifunctional signals in virulence and host recognition, *Curr. Opin. Microbiol.* **3**, 73–78.

Kunkel, B.N., Brooks, D.M. (2002) Cross talk between signalling pathways in pathogen defense, *Curr. Opin. Plant Biol.* **5**, 325–331.

Lahaye, T., Bonas, U. (2001) Molecular secrets of bacterial type III effector proteins, *Trends Plant Sci.* **6**, 479–485.

Lai, E-M., Kado, C.I. (2000) The T-pilus of *Agrobacterium tumefaciens*, *Trends Microbiol.* **8**, 361–369.

Lam, E., Kato, N., Lawton, M. (2001) Programmed cell death, mitochondria and the plant hypersensitive response, *Nature* **411**, 848–853.

Laugé, R., de Wit, P.J.G.M. (1998) Fungal avirulence genes: structure and possible functions, *Fungal Genet. Biol.* **24**, 285–297.

Laugé, R., Joosten, M.H.A.J., Haanstra, J.P.W., Goodwin, P.H., Lindhout, P., de Wit, P.J.G.M. (1998) Successful search for a resistance gene in tomato targeted against a virulence factor of a fungal pathogen, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9014–9018.

Laxalt, A.M., Munnik, T. (2002) Phospholipid signalling in plant defence, *Curr. Opin. Plant Biol.* **5**, 332–338.

Leach, J.E., Vera Cruz, C.M., Bai, J., Leung, H. (2001) Pathogen fitness penalty as a predictor of durability of disease resistance genes, *Annu. Rev. Phytopathol.* **39**, 187–224.

Lee, J-Y., Lucas, W.J. (2001) Phosphorylation of viral movement proteins – regulation of cell-to-cell trafficking, *Trends Microbiol.* **9**, 5–8.

Lindbo, J.A., Fitzmaurice, W.P., della-Cioppa, G. (2001) Virus-mediated reprogramming of gene expression in plants, *Curr. Opin. Plant Biol.* **4**, 181–185.

Loh, J., Pierson, E.A., Pierson, L.S., Stacey, G., Chatterjee, A. (2002) Quorum sensing in plant-associated bacteria, *Curr. Opin. Plant Biol.* **5**, 285–290.

Luderer, R., Joosten, M.H.A.J. (2001) Avirulence proteins of plant pathogens: determinants of victory and defeat, *Mol. Plant Pathol.* **2**, 355–364.

Mackey, D., Holt, B.F., Wiig, A., Dangl, J. (2002) RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in Arabidopsis, *Cell* **108**, 743–754.

Markham, J.E., Hille, J. (2001) Host-selective toxins as agents of cell death in plant-fungus interactions, *Mol. Plant Pathol.* **2**, 229–239.

Martin, G.B. (1999) Functional analysis of plant disease resistance genes and their downstream effectors, *Curr. Opin. Plant Biol.* **2**, 273–279.

Maule, A., Leh, V., Lederer, C. (2002) The dialogue between viruses and hosts in compatible interactions, *Curr. Opin. Plant Biol.* **5**, 279–284.

McDonald, B.A., Linde, C. (2002) Pathogen population genetics, evolutionary potentials, and durable resistance, *Annu. Rev. Phytopathol.* **40**, 349–379.

Mellersh, D., Heath, M.C. (2001) Plasma membrane-cell wall adhesion is required for the expression of plant defense responses during fungal penetration, *Plant Cell* **13**, 413–424.

Mendgen, K., Struck, C., Voegele, R.T., Hahn, M. (2000) Biotrophy and rust haustoria, *Physiol. Mol. Plant. Pathol.* **56**, 141–145.

Michelmore, R.W., Meyers, B.C. (1998) Clusters of resistance genes in plant evolve by divergent selection and a birth-and-death process, *Genome Res.* **8**, 1113–1130.

Moffett, P., Farnham, G., Peart, J.R., Baulcombe, D.C. (2002) Interaction between domains of a plant NBS-LRR protein in disease resistance-related cell death, *EMBO J.* **21**, 4511–4519.

Morris, E.R., Walker, J.C. (2003) Receptor-like protein kinases: the keys to response, *Curr. Opin. Plant Biol.* **6**, 339–342.

Panstruga, R. (2003) Establishing compatibility between plants and obligate biotrophic pathogens, *Curr. Opin. Plant Biol.* **6**, 320–326.

Peart, J.R., Lu, R., Sadanandom, A., Malcuit, I., Moffett, P., Brice, D.C., Schauser, L., Jaggard, D.A., Xiao, S., Coleman, M.J., Dow, M., Jones, J. D. G., Shirasu, K., Baulcombe, D. C. (2002) Ubiquitin ligase-associated protein SGT1 is required for host and non-host disease resistance in plants, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10865–10869.

Perfect, S.E., Green, J.R. (2001) Infection structures of biotrophic and hemibiotrophic fungal plant pathogens, *Mol. Plant Pathol.* **2**, 101–108.

Power, A.G. (2000) Insect transmission of plant viruses: a constraint on virus variability, *Curr. Opin. Plant Biol.* **3**, 336–340.

Pühler, A., Arlat, M., Becker, A., Gottfert, M., Morrissey, J.P., O'Gara, F. (2004) What can bacterial genome research teach us about bacteria-plant interactions? *Curr. Opin. Plant Biol.* **7**, 137–147.

Rathjen, J.P., Moffett, P. (2003) Early signal transduction events in specific plant disease resistance, *Curr. Opin. Plant Biol.* **6**, 300–306.

Ramonell, K.M., Somerville, S. (2002) The genomics parade of defense responses: to infinity and beyond, *Curr. Opin. Plant Biol.* **5**, 291–294.

Schell, M.A. (2000) Control of virulence and pathogenicity genes of *Ralstonia solanacearum* by an elaborate sensory network, *Annu. Rev. Phytopathol.* **38**, 263–292.

Schneider, D.S. (2002) Plant immunity and film noir: what gumshoe detectives can teach us about plant-pathogen interactions, *Cell* **109**, 537–540.

Shah, J. (2003) The salicylic acid loop in plant defense, *Curr. Opin. Plant Biol.* **6**, 365–371.

Singh, K.B., Foley, R.C., Oñate-Sánchez, L. (2002) Transcription factors in plant defense and stress responses, *Curr. Opin. Plant Biol.* **5**, 430–436.

Snoeijers, S.S., Pérez-García, A., Joosten, M.H.A.J., de Wit, P.J.G.M. (2000) The effect of nitrogen on disease development and gene expression in bacterial and fungal plant pathogens, *Eur. J. Plant Pathol.* **106**, 493–506.

Soanes, D.M., Skinner, W., Keon, J., Hargreaves, J., Talbot, N.J. (2002) Genomics of Phytopathogenic fungi and the development of bioinformatic resources, *Mol. Plant Microbe Interact.* **15**, 421–427.

Staples, R.C. (2001) Nutrients for a rust fungus: the role of haustoria, *Trends Plant Sci.* **6**, 496–498.

Staskawicz, B.J., Mudgett, M.B., Dangl, J.L., Galan, J.E. (2001) Common and contrasting themes of plant and animal diseases, *Science* **292**, 2285–2289.

Stuiver, M.H., Custers, J.H.H.V. (2001) Engineering disease resistance in plants, *Nature* **411**, 865–868.

Sweeney, M.J., Dobson, A.D.W. (1999) Molecular biology of mycotoxin production, *FEMS Microbiol. Lett.* **175**, 149–163.

Tao, Y., Xie, Z.Y., Chen, W.Q., Glazebrook, J., Chang, H.S., Han, B., Zhu, T., Zou, G.Z., Katagiri, F. (2003) Quantitative nature of *Arabidopsis* responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*, *Plant Cell* **15**, 317–330.

Thordal-Christensen, H. (2003) Fresh insights into processes of nonhost resistance, *Curr. Opin. Plant Biol.* **6**, 351–357.

Tyler, B.M. (2001) Genetics and genomics of the oomycete-host interface, *Trends Genet.* **17**, 611–614.

Vance, V., Vaucheret, H. (2001) RNA silencing in plants – defense and counterdefense, *Science* **292**, 2277–2280.

Van Etten, H., Temporini, E., Wasmann, C. (2001) Phytoalexin (and phytoanticipin) tolerance as a virulence trait: why is it not required by all pathogens? *Physiol. Mol. Plant. Pathol.* **59**, 83–93.

Ward, D.V., Zupan, J.R., Zambryski, P.C. (2002) *Agrobacterium* VirE2 gets the VIP1 treatment in plant nuclear import, *Trends Plant Sci.* **7**, 1–3.

Waterhouse, P.M., Wang, M-B., Lough, T. (2001) Gene silencing as an adaptive defence against viruses, *Nature* **411**, 834–842.

Wendehenne, D., Pugin, A., Klessig, D.F., Durner, J. (2001) Nitric oxide: comparative synthesis and signalling in animal and plant cells, *Trends Plant Sci.* **6**, 177–183.

White, F.F., Yang, B., Johnson, L.B. (2000) Prospects for understanding avirulence gene function, *Curr. Opin. Plant Biol.* **3**, 291–298.

Wildermuth, M.C., Dewdney, J., Wu, G., Ausubel, F.M. (2001) Isochorismate synthase is required to synthesize salicylic acid for plant defence, *Nature* **414**, 562–565.

Wojtaszek, P. (1997) Oxidative burst: an early plant response to pathogen infection, *Biochem. J.* **322**, 681–692.

Wu, Y., Wood, M.D., Tao, Y., Katagiri, F. (2003) Direct delivery of bacterial avirulence proteins into resistant *Arabidopsis* protoplasts leads to hypersensitive cell death, *Plant J.* **33**, 131–137.

Xu, J-R. (2000) MAP Kinases in fungal pathogens, *Fungal Genet. Biol.* **31**, 137–152.

Zupan, J., Muth, T.R., Draper, O., Zambryski, P. (2000) The transfer of DNA from *Agrobacterium tumefaciens* into plants: a feast of fundamental insights, *Plant J.* **23**, 11–28.

Plant-based Expression of Biopharmaceuticals

Jörg Knäblein
Schering AG, Berlin, Germany

1 Introduction 387

2 Alternative Expression Systems 387

3 History of Plant Expression 389

4 Current Status of Plant-based Expression 390
4.1 SWOT Analysis Reveals a Ripe Market for Plant Expression Systems 390
4.2 Risk Assessment and Contingency Measures 392

5 The Way Forward: Moving Plants to Humanlike Glycosylation 396

6 Three Promising Examples: Tobacco (Rhizosecretion, Transfection) and Moss (Glycosylation) 398
6.1 Harnessing Tobacco Roots to Secrete Proteins 398
6.2 High Protein Yields Utilizing Viral Transfection 399
6.3 Simple Moss Performs Complex Glycosylation 401

7 Other Systems Used for Plant Expression 404

8 Analytical Characterization 405

9 Conclusion and Outlook 406

Acknowledgments 407

Bibliography 407
Books and Reviews 407
Primary References 407

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

Keywords

GMP
Good Manufacturing Practice (GMP) was established by WHO in 1968 to guarantee the optimum degree of quality during production and processing of pharmaceuticals (cGMP means under the current regulations of the authorities).

Transgenic
Organisms that have externally introduced foreign DNA/genes stably integrated into their genome to, for example, produce desired substances like human insulin.

Plant-based Expression
Transgenic plants can be genetically modified with a gene of interest to produce a biopharmaceutical of interest.

Glycosylation
It is the addition of polysaccharides to a certain molecule such as a protein. The majority of proteins are synthesized in the rough endoplasmic reticulum (ER) where they undergo glycosylation.

Bioreactor
It is a vessel in which a (bio)chemical process that involves organisms or biochemically active substances (e.g. enzymes) derived from such organisms is carried out.

Biopharmaceuticals are currently the mainstay products of the biotechnology market and represent the fastest growing and, in many ways, the most exciting sector within the pharmaceutical industry. The term ''biopharmaceutical'' was originated in the 1980s, when a general consensus evolved that it represented a class of therapeutics produced by means of modern biotechnologies. Already a quarter of a century ago, ''humulin'' (recombinant human insulin, produced in *E. coli* and developed by Genentech in collaboration with Eli Lilly) was approved and received marketing authorization in the United States of America in 1982. Since then the market for biopharmaceuticals has been steadily growing and currently nearly 150 biopharmaceuticals have gained approval for general human use (EU and USA). Over this period it became obvious that production capacities for biopharmaceuticals with ''conventional'' bioreactors would be a bottleneck and that worldwide fermentation capacities are limited. One exciting solution to these ''capacity crunches'' is the use of transgenic plants to produce biopharmaceuticals. This article describes different plant expression systems, their advantages and limitations, and concludes by considering some of the innovations and trends likely to influence the future of plant-based biopharmaceuticals.

1 Introduction

Biopharmaceuticals, which are large molecules produced by living cells, are currently the mainstay products of the biotechnology industry. Indeed, biologics such as Genentech's (Vacaville, CA, USA) human growth factor somatropin or Amgen's (Thousand Oaks, CA, USA) recombinant erythropoietin (EPO) have shown that biopharmaceuticals can benefit a huge number of patients and also generate big profits for these companies at the same time. The single most lucrative product is EPO and combined sales of the recombinant EPO products "Procrit" (Ortho biotech) and "Epogen" (Amgen) have reportedly surpassed the $6.5 billion mark. But it has also become obvious over the last couple of years that current fermentation capacities will not be sufficient to manufacture all biopharmaceuticals (in the market already or in development), because the market and demand for biologics is continuously and very rapidly growing; for antibodies alone (with at least 10 monoclonal antibodies approved and being marketed), the revenues are predicted to expand to US$3 billion in 2002 and US$8 billion in 2008. The 10 monoclonal antibodies on the market consume more than 75% of the industry's manufacturing capability. And there are up to 60 more that are expected to reach the market in the next six or seven years. Altogether, there are about 1200 protein-based products in the pipeline with a 20% growth rate and the market for current and late stage (Phase III) is estimated to be US$42 billion in 2005 and even US$100 billion in 2010. But, there are obvious limitations of large-scale manufacturing resources and production capacities – and pharmaceutical companies are competing (see ref Knäblein (2004), review).

To circumvent this capacity crunch, it is necessary to look into other technologies rather than the established ones, like, for example, *Escherichia coli* or CHO (Chinese hamster ovary) cell expression. One solution to avoid these limitations could be the use of transgenic plants to express recombinant proteins at low cost, in GMP (good manufacturing practice) quality greenhouses (with purification and fill finish in conventional facilities). Plants therefore provide an economically sound source of recombinant proteins, such as industrial enzymes, and biopharmaceuticals. Furthermore, using the existing infrastructure for crop cultivation, processing, and storage will reduce the amount of capital investment required for commercial production. For example, it was estimated that the production costs of recombinant proteins in plants could be between 10 and 50 times lower than those for producing the same protein in *E. coli* and Alan Dove describes a factor of thousand for cost of protein (US dollar per gram of raw material) expressed in, for example, CHO cells compared to transgenic plants. So, at the dawn of this new millennium, a solution is imminent to circumvent expression capacity crunches and to supply mankind with the medicines we need. Providing the right amounts of biopharmaceuticals can now be achieved by applying our knowledge of modern life sciences to systems that were on this planet long time before us – plants.

2 Alternative Expression Systems

Currently, CHO cells are the most widely used technology in biomanufacturing because they are capable of expressing eukaryotic proteins (processing, folding,

and posttranslational modifications) that cannot be provided by *E. coli*. A long track record exists for CHO cells, but unfortunately they bring some problems along when it comes to scaling up production. Transport of oxygen (and other gases) and nutrients is critical for the fermentation process, as well as the fact that heat must diffuse evenly to all cultured cells. According to the Michaelis–Menten equation, the growth rate depends on the oxygen/nutrient supply; therefore, good mixing and aeration are a prerequisite for the biomanufacturing process and are usually achieved by different fermentation modes (see Fig. 1). But the laws of physics set strict limits on the size of bioreactors. For example, an agitator achieves good heat flow and aeration, but with increased fermenter size, shear forces also increase and disrupt the cells – and building parallel lines of bioreactors multiplies the costs linearly. A 10 000-L bioreactor costs between US$ 250 000 to 500 000 and takes five years to build (conceptual planning, engineering, construction, validation, etc.). An error in estimating demand for, or inaccurately predicting the approval of, a new drug can be incredibly costly. To compound the problem, regulators in the United States and Europe demand that drugs have to be produced for the market in the same system used to produce them for the final round of clinical trials, in order to guarantee bioequivalence (e.g. toxicity, bioavailability, pharmacokinetics, and pharmacodynamics) of the molecule. So, companies have to choose between launching a product manufactured at a smaller development facility (and struggling to meet market demands) or building larger, dedicated facilities for a drug that might never be approved!

Therefore, alternative technologies are used for the expression of biopharmaceuticals, some of them also at lower costs involved (see Fig. 2). One such alternative is the creation of transgenic animals (''pharming''), but this suffers from the disadvantage that it requires a long time to establish such animals (approximately 2 years). In addition to that, some of the human biopharmaceuticals could be detrimental to the mammal's health, when expressed in the mammary glands. This is why ethical debates sometimes arise from the use of transgenic mammals for production of biopharmaceuticals. Although there are no ethical concerns involved with plants, there are societal ones that will be addressed later. Another expression system (see Fig. 2) utilizes transgenic chicken. The eggs, from which the proteins are harvested, are natural protein-production systems. But production of transgenic birds is still

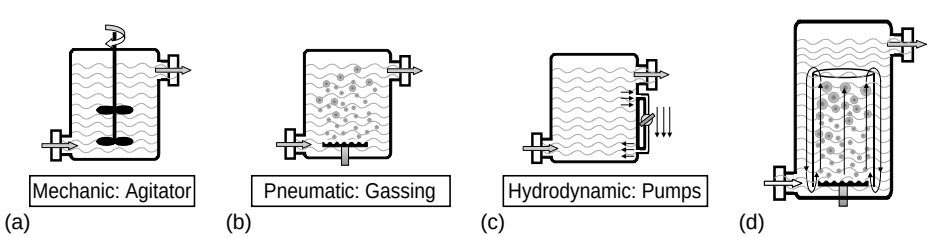

Fig. 1 Different fermentation modes for bioreactors. In order to achieve best aeration and mixing and to avoid high shear forces, different fermentation modes are applied. (a) mechanical, (b) pneumatical, (c) hydrodynamic pumps, (d) airlift reactor. Source: Knäblein J. (2002) *Transport Processes in Bioreactors and Modern Fermentation Technologies*, Lecture at University of Applied Sciences, Emden, Germany.

Major technology	Mammalian (CHO) cells	Transgenic mammal milk	Transgenic chicken eggs	Transgenic plants
companies	Amgen (Thousand Oaks, CA) Genentech (S. San Francisco, CA) other current biologics manufacturers: Crucell (Leiden, Netherlands) uses human cells	GTC Biotherapeutics (Framingham, MA) PPL Therapeutics (Edinburgh, UK) BioProtein (Paris, France)	Avigenics (Athens, GA) Origen Therapeutics (Burlingame, CA) TranXenoGen (shrewsbury, MA) Viragen (Plantation, FL) GeneWorks (Ann Arbor, MI) Vivalis (Nantes, France)	Croptech (Blacksburg, VA) Epicyte (San Diego, CA) Large Scale Biology (Owensboro, KY) Meristem Therapeutics (Clermont-Ferrand, France) Prodigene (College Station, TX)
Estimated cost (cost/g raw material)*	$150	$1–$2	$1–$2	$0.05

*Company estimates

Fig. 2 Companies and technologies in biomanufacturing. A comparison of different expression systems shows the big differences in terms of costs, ranging from 150 US$ per gram for CHO cells to 0.05 US$ per gram for transgenic plants. Source: Dove, A. (2002) Uncorking the biomanufacturing bottleneck, *Nat. Biotechnol.* **20**, 777–779.

several years behind transgenic mammal technology. Intensive animal housing constraints also make them more susceptible to disease (e.g. Asia 1997 or Europe 2003: killing of huge flocks with thousands of chicken suffering from fowl pest). In the light of development time, experience, costs, and ethical issues, plants are therefore the favored technology, since such systems usually have short gene-to-protein times (weeks), some are already well established, and as mentioned before, the involved costs are comparatively low. This low cost of goods sold (COGS) for plant-derived proteins is mainly due to low capital costs: greenhouse costs are only US$ 10 per m^2 versus US$ 1000 per m^2 for mammalian cells.

3 History of Plant Expression

Plants have been a source of medicinal products throughout human evolution. These active pharmaceutical compounds have been primarily small molecules, however. One of the most popular examples is aspirin (acetylsalicylic acid) to relieve pain and reduce fever. A French pharmacist first isolated natural salicin (a chemical relative of the compound used to make aspirin) from white willow bark in 1829. Advances in genetic engineering are now allowing for the production of therapeutic proteins (as opposed to small molecules) in plant tissues. Expression of recombinant proteins in plants has been well documented since the 1970s and has slowly gained credibility in the biotechnology industry and regulatory agencies. The first proof of concept has been the incorporation of insect and pest resistance into grains. For example, ''Bt corn'' contains genes from *Bacillus thuringensis* and is currently being grown commercially. Genetic engineering techniques are now available for the manipulation of almost all commercially valuable plants. Easy transformation and cultivation make plants suitable

for production of virtually any recombinant protein.

Plants have a number of advantages over microbial expression systems, but one of them is of outmost importance: they can produce eukaryotic proteins in their native form, as they are capable of carrying out posttranslational modifications required for the biological activity of many such proteins (see Fischer Schillberg (2004), books). These modifications can be acetylation, phosphorylation, and glycosylation, as well as others. Per se, there is no restriction to the kind of proteins that can be expressed in plants: vaccines (e.g. pertussis or tetanus toxins), serum proteins (e.g. albumin), growth factors (e.g. vascular endothelial growth factor (VEGF), erythropoietin), or enzymes (e.g. urokinase, glucose oxidase, or glucocerebrosidase). However, enzymes sometimes have very complex cofactors, which are essential for their catalytic mode of action, but cannot be supplied by most expression systems. This is why, for the expression of some enzymes, expression systems with special features and characteristics need to be developed. Another very important class of proteins is the antibodies (e.g. scFv, Fab, IgG, or IgA). More than 100 antibodies are currently used in clinical trials as therapeutics, drug delivery vehicles, in diagnostics and imaging, and in drug discovery research for both screening and validation of targets. Again, plants are considered as the system of choice for the production of antibodies ("plantibodies") in bulk amounts at low costs. Since the initial demonstration that transgenic tobacco (*Nicotiana tabacum*) is able to produce functional IgG1 from mouse, full-length antibodies, hybrid antibodies, antibody fragments (Fab), and single-chain variable fragments (scFv) have been expressed in higher plants for a number of purposes. These antibodies can serve in health care and medicinal applications, either directly by using the plant as a food ingredient or as a pharmaceutical or diagnostic reagent after purification from the plant material. In addition, antibodies may improve plant performance, for example, by controlling plant disease or by modifying regulatory and metabolic pathways.

4 Current Status of Plant-based Expression

4.1 SWOT Analysis Reveals a Ripe Market for Plant Expression Systems

When I analyzed the different expression systems regarding their strengths, weaknesses, opportunities, and threats (SWOT), the advantages of plants and their potential to circumvent the worldwide capacity limitations for protein production became quite obvious (see Fig. 3). Comparison of transgenic animals, mammalian cell culture, plant expression systems, yeast, and bacteria shows certain advantages for each of the systems. In the order in which the systems were just mentioned, we can compare them in terms of their development time (speed). Transgenic animals have the longest cycle time (18 months to develop a goat), followed by mammalian cell culture, plants, yeast, and bacteria (one day to transform *E. coli*). If one looks at operating and capital costs, safety, and scalability, the data show that plants are beneficial: therefore, in the comparison (see Fig. 3), they are shown on the right-hand side already. But even for glycosylation, multimeric assembly and folding (where plants are not shown on the right-hand side, meaning other systems are advantageous), some plant expression systems are moving in

Strengths
- Access new manufacturing facilities
- High production rates/high protein yield
- Relatively fast 'gene to protein' time
- Safety benefits;no human pathogens/no TSE
- Stable cell lines/high genetic stability
- Simple medium (water, minerals & light)
- Easy purification (ion exchange vs protA)

Weaknesses
- No approved products yet (but Phase III)
- No final guidelines yet (but drafts available)

Opportunities
- Reduce projected COGS
- Escape capacity limitations
- Achieve human-like glycosylation

Threats
- Food chain contamination
- Segregation risk

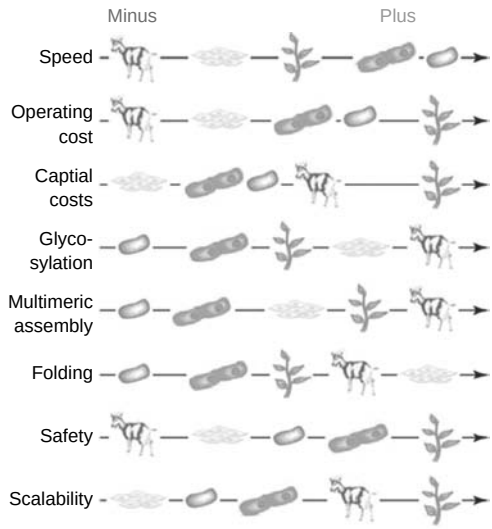

Fig. 3 SWOT analysis of plant expression systems. Plant expression systems have a lot of advantages (plus) over other systems and are therefore mostly shown on the right-hand side of the picture (Raskin, I., Fridlender, B., et al. (2002) Plants and human health in the twenty-first century, *Trends Biotechnol.* **20**, 522–531). Herein different systems (transgenic animals, mammalian cell culture, plants, yeast, and bacteria) are compared in terms of speed (how quickly they can be developed), operating and capital costs, and so on, and plants are obviously advantageous. Even for glycosylation, assembly, and folding, where plants are not shown on the right-hand side (meaning other systems are advantageous), some plant expression systems are moving in that direction (as will be shown exemplarily in the section on moss). Also the weaknesses and threats can be dealt with, using the appropriate plant expression system. Source: Knäblein J. (2003) *Biotech: A New Era In The New Millennium – From Plant Fermentation To Plant Expression Of Biopharmaceuticals*, PDA International Congress, Prague, Czech Republic.

that direction. An example of this is the moss system from the company greenovation Biotech GmbH (Freiburg, Germany), which will be discussed in detail in the example section. This system performs proper folding and assembly of even such complex proteins like the homodimeric VEGF. Even the sugar pattern could successfully be reengineered from plant to humanlike glycosylation.

In addition to the potential of performing human glycosylation, plants also enjoy the distinct advantage of not harboring any pathogens, which are known to harm animal cells (as opposed to animal cell cultures and products), nor do the products contain any microbial toxins, TSE (Transmissible Spongiform Encephalopathies), prions, or oncogenic sequences. In fact, humans are exposed to a large, constant dose of living plant viruses in the diet without any known effects/illnesses. Plant production of protein therapeutics also has advantages with regard to their scale and speed of production. Plants can be grown in ton quantities (using

existing plant/crop technology, like commercial greenhouses), be extracted with industrial-scale equipment, and produce kilogram-size yields from a single plot of cultivation. These economies of scale are expected to reduce the cost of production of pure pharmaceutical-grade therapeutics by more than 2 orders of magnitude versus current bacterial fermentation or cell culture reactor systems (plus raw material COGS are estimated to be as low as 10% of conventional cell culture expenses).

Although a growing list of heterologous proteins were successfully produced in a number of plant expression systems with their manifold advantages, there are also obvious downsides. One weakness is that no product has been approved for the market yet (but will be soon, since some are in Phase III clinical trials already, see Table 1). The other weakness is that no final regulatory guidelines exist. But as mentioned before, regulatory authorities (Food and Drug Administration (FDA), European Medicine Evaluation Agency (EMEA), and Biotechnology Regulatory Service (BRS) and the Biotechnology Industry Organization (BIO) have drafted guidelines on plant-derived biopharmaceuticals (see Table 2) and have asked the community for comments. The FDA has also issued several PTC (Points To Consider) guidelines about plant-based biologics, and review of the July 2002 PTC confirms that the FDA supports this field and highlights the benefits of plant expression systems – including the absence of any pathogens to man from plant extracts. The main concerns of using plant expression systems are societal ones about environmental impacts, segregation risk, and contamination of the food chain. But these threats can be dealt with, using nonedible plants (nonfood, nonfeed), applying advanced containment technologies (GMP greenhouses, bioreactors) and avoiding open-field production.

Owing to the obvious strengths of plant expression systems, there has been explosive growth in the number of start-up companies. Since the 1990s, a number of promising plant expression systems have been developed, and in response to this "blooming field" big pharmaceutical companies have become more interested. Now, the plant expression field is "ripe" for strategic alliances, and, in fact, the last year has seen several major biotech companies begin partnerships with such plant companies. The selection of several such partnerships shown in Table 1 clearly demonstrates that, in general, there has been sufficient experimentation with various crops to provide the overall proof of concept that transgenic plants can produce biopharmaceuticals. However, and this can be seen in the table as well, the commercial production of biopharmaceuticals in transgenic plants is still in the early stages of development and yet the most advanced products are in Phase III clinical development.

4.2 Risk Assessment and Contingency Measures

For a number of reasons, including the knowledge base developed on genetically modifying its genome, industrial processes for extracting fractionated products and the potential for large-scale production, the preferred plant expression system has been corn. However, the use of corn touches on a potential risk: some environmental activist groups and trade associations are concerned about the effect on the environment and possible contamination of the food supply. These issues

Tab. 1 Plant-derived biopharmaceuticals in clinical trials.

Company	*Partner*	*Protein/indication*	*Host*	*Stage*
Monsanto	Guy's Hospital London	Anticaries antibody	Corn	Phase III
Large Scale Biology	Own product	scFv (non-Hodgkin)	Tobacco	Phase IIIs
Meristem Therapeutics	Solvay Pharmaceuticals	Gastric lipase	Corn	Phase II
Large Scale Biology	ProdiGene, Plant Bioscience	Anti-ideotype antibody	Tobacco	Phase I
Monsanto	NeoRx	Antitumor antibody	Corn	Phase I
ProdiGene	Own product	TGEV vaccine	Corn	Phase I
Epicyte Pharmaceutical	Dow, Centocor	Anti-HSV antibody	Corn	Phase I
Crop Tech	Immunex	Enbrel (arthritis)	Tobacco	Preclinical
Crop Tech	Amgen	Therapeutic antibodies	Tobacco	Preclinical
AltaGen Bioscience Inc.	U.S. Army 3 + biotechs	Antibodies	Potato	Preclinical
Meristem Therapeutics	CNRS	Human lactoferrin	Corn	Preclinical
MPB Cologne GmbH	Aventis CropScience	Confidential	Potato	Preclinical

Tab. 2 Drafted guidelines on plant-derived biopharmaceuticals.

Agency	*Guideline*	*Status*
BRS (Biotechnology Regulatory Services	"Case study on plant-derived biologics" for Office of Science and Technology Policy/Council on Environmental Quality	Released: Mar 5, 2001
BIO (Biotechnology Industry Organization)	"Reference Document for Confinement and Development of Plant-made Pharmaceuticals in the United States"	Released: May 17, 2002
BIO (Biotechnology Industry Organization)	"BIO Position on Geographic Restrictions for Plant-made Pharmaceuticals and Industrials"	Released: Oct 22, 2002
EMEA (European Medicine Evaluation Agency)	"Concept Paper on the Development of a Committee for Proprietary Medicinal Products (CPMP) Points to Consider on the Use of Transgenic Plants in the Manufacture of Biological Medicinal Products for Human Use"	Released: Mar 01, 2001
FDA (Food and Drug Administration)	"Drugs, Biologics, and Medical Devices Derived from Bioengineered Plants for Use in Humans and Animals"	Issued: Sep 6, 2002
EMEA (European Medicine Evaluation Agency)	"Points To Consider Quality Aspects of Medicinal Products containing active substances produced by stable transgene expression in higher plants"	Issued: Mar 13, 2002

are reflected in the regulatory guidelines and have been the driving force to investigate other plants as well. While many mature and larger companies have been working in this area for many years, there are a number of newcomers that are developing expertise as well. These smaller companies are reacting to the concerns by looking at the use of nonedible plants that can be readily raised in greenhouses. All potential risks have to be assessed and contingency measures need to be established. Understanding the underlying issues is mandatory to make sophisticated decisions about the science and subsequently on the development of appropriate plant expression systems for production of biopharmaceuticals.

Ongoing public fears from the food industry and the public, particularly in Europe ("Franken Food") could have spillover effects on plant-derived pharmaceuticals. Mistakes and misunderstandings have already cost the genetically enhanced grain industry hundreds of millions of dollars. The only way to prevent plant expression systems from suffering the same dilemma is to provide the public with appropriate information on emerging discoveries and newly developed production systems for biopharmaceuticals. Real and theoretical risks involve the spread of engineered genes into wild plants, animals, and bacteria (horizontal transmission). For example, if herbicide resistance was transmitted to weeds, or antibiotic resistance was to be transmitted to bacteria, superpathogens could result. If these genetic alterations were transmitted to their progeny (vertical transmission), an explosion of the pathogens could cause extensive harm. An example of this occurred several years ago, when it was feared that pest-resistant genes had been transmitted from Bt corn to milkweed – leading to the widespread death of Monarch butterflies. Although this was eventually not found to be the case, the public outcry over the incident was a wake-up call to the possible dangers of transgenic food technology. To avoid the same bad perception for biopharmaceuticals expressed in plants, there is the need for thorough risk assessment and contingency planning. One method is the employment of all feasible safety strategies to prevent spreading of engineered DNA (genetic drift), like a basic containment in a greenhouse environment. Although no practical shelter can totally eradicate insect and rodent intrusion, this type of isolation is very effective for self-pollinators and those plants with small pollen dispersal patterns. The use of species-specific, fragile, or poorly transmissible viral vectors is another strategy. Tobacco mosaic virus (TMV), for example, usually only infects a tobacco host.

It requires an injury of the plant to gain entry and cause infection. Destruction of a field of TMV-transformed tobacco requires only plowing under or application of a herbicide. These factors prevent both horizontal and vertical transmission. In addition, there is no known incidence of plant viruses infecting animal or bacterial cells. Another approach is to avoid stable transgenic germlines and therefore most uses of transforming viruses do not involve the incorporation of genes into the plant cell nucleus. By definition, it is almost impossible for these genes to be transmitted vertically through pollen or seed. The engineered protein product is produced only by the infected generation of plants. Another effective way to reduce the risk of genetic drift is the use of plants that do not reproduce without human aid. The modern corn plant cannot reproduce without cultivation and

the purposeful planting of its seeds. If a plant may sprout from grain, it still needs to survive the wintering-over process and gain access to the proper planting depth. This extinction process is so rapid, however, that the errant loss of an ear of corn is very unlikely to grow a new plant. Another very well-known example of self-limited reproduction is the modern banana. It propagates almost exclusively through vegetative cloning (i.e. via cuttings).

Pollination is the natural way for most plants to spread their genetic information, make up new plants, and to deliver their offspring in other locations. The use of plants with limited range of pollen dispersal and limited contact with compatible wild hosts therefore is also very effective to prevent genetic drift. Corn, for example, has pollen, which survives for only 10 to 30 min and, hence, has an effective fertilizing radius of less than 500 m. In North America, it has no wild-type relatives with which it could cross-pollinate. In addition to being spatially isolated from nearby cornfields, transgenic corn can be "temporally isolated" by being planted at least 21 days earlier or 21 days later than the surrounding corn, to ensure that the fields are not producing flowers at the same time. Under recent USDA (U.S. Department of Agriculture) regulations, the field must also be planted with equipment dedicated to the genetically modified crop. For soybeans, the situation is different, since they are virtually 100% self-fertilizers and can be planted in very close proximity to other plants without fear of horizontal spread. Another option is the design of transgenic plants that have only sterile pollen or – more or less only applicable for greenhouses – completely prevent cross-pollination by covering the individual plants. One public fear regards spreading antibiotic resistance from one (transgenic donor) plant to other wild-type plants or bacteria in the environment. Although prokaryotic promoters for antibiotic resistance are sometimes used in the fabrication and selection of transgenic constructs, once a transgene has been stably incorporated into the plant genome, it is under the control of plant (eukaryotic) promoter elements. Hence, antibiotic-resistance genes are unable to pass from genetically altered plants into bacteria and remain functional. As stated earlier, another common fear is the creation of a "super bug." The chance of creating a supervirulent virus or bacterium from genetic engineering is unlikely, because the construction of expression cassettes from viral or bacterial genomes involves the removal of the majority of genes responsible for the normal function of these organisms. Even if a resultant organism is somewhat functional, it cannot compete for long in nature with normal, wild-type bacteria of the same species.

As one can see from the aforementioned safety strategies, considerable effort is put into the reduction of any potential risk from the transgenic plant for the environment. In general, the scientific risk can be kept at a minimum, if common sense is applied – in accordance with Thomas Huxley (1825–1895) that "Science is simply common sense at its best." For example, protein toxins (for vaccine production) should never be grown in food plants.

Additionally, the following can be employed as a kind of risk management to prevent the inappropriate or unsafe use of genetically engineered plants:

- An easily recognized phenotypic characteristic can be coexpressed in an

engineered product (e.g. tomatoes that contain a therapeutic protein can be selected to grow in a colorless variety of fruit).

- Protein expression can be induced only after harvesting or fruit ripening. For example, CropTech's (Blacksburg, VA, USA) inducible expression system in tobacco, MeGA-PharM, leads to very efficient induction upon leaf injury (harvest) and needs no chemical inducers. This system possesses a fast induction response and protein synthesis rate, and thus leads to high expression levels with no aged product in the field (no environmental damage accumulation).
- Potentially antigenic or immunomodulatory products can be induced to grow in, or not to grow in, a certain plant tissue (e.g. root, leaf/stem, seed, or pollen). In this way, for example, farmers can be protected from harmful airborne pollen or seed dusts.
- Although no absolute system can prevent vandalism or theft of the transgenic plants, a very effective, cheap solution has been used quietly for many years now in the United States. Plots of these modified plants are being grown with absolutely no indication that they are different from a routine crop. In the Midwest, for example, finding a transgenic corn plot among the millions of acres of concurrently growing grain is virtually impossible. The only question here is, if this approach really helps facilitating a fair and an open discussion with the public. Asking the same question for the EU is not relevant: owing to labeling requirements, this approach would not be feasible, as, in general, it is much more difficult to perform open-field studies with transgenic plants.

5 The Way Forward: Moving Plants to Humanlike Glycosylation

As discussed earlier, plant production of therapeutic proteins has many advantages over bacterial systems. One very important feature of plant cells is their capability of carrying out posttranslational modifications. Since they are eukaryotes (i.e. have a nucleus), plants produce proteins through an ER (endoplasmatic reticulum) pathway, adding sugar residues also to the protein – a process called *glycosylation*. These carbohydrates help determine the three-dimensional structures of proteins, which are inherently linked to their function and their efficacy as therapeutics. This glycosylation also affects protein bioavailability and breakdown of the biopharmaceutical; for example, proteins lacking terminal sialic acid residues on their sugar groups are often targeted by the immune system and are rapidly degraded. The glycosylation process begins by targeting the protein to the ER. During translation of mRNA (messenger RNA) into protein, the ribosome is attached to the ER, and the nascent protein fed into the lumen of the ER as translation proceeds. Here, one set of glycosylation enzymes attaches carbohydrates to specific amino acids of the protein. Other glycosylation enzymes either delete or add more sugars to the core structures. This glycosylation process continues into the Golgi apparatus, which sorts the new proteins, and distributes them to their final destinations in the cell (see Fig. 4). Bacteria lack this ability and therefore cannot be used to synthesize proteins that require glycosylation for activity. Although plants have a somewhat different system of protein glycosylation from mammalian cells, the differences usually prove not to be a problem. Some proteins, however, require

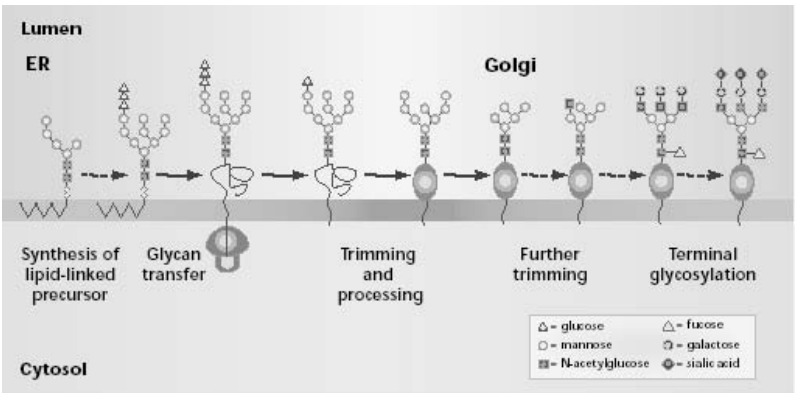

Fig. 4 The glycosylation pathway via ER and Golgi apparatus. In the cytosol, carbohydrates are attached to a lipid precursor, which is then transported into the lumen of the ER to finish core glycosylation. This glycan is now attached to the nascent, folding polypeptide chain (which is synthesized by ribosomes attached to the cytosolic side of the ER from where it translocates into the lumen) and subsequently trimmed and processed before it is folded and moved to the Golgi apparatus. Capping of the oligosaccharide branches with sialic acid and fucose is the final step on the way to a mature glycoprotein. Source: Dove, A. (2001) The bittersweet promise of glycobiology, *Nat. Biotechnol.* **19**, 913–917.

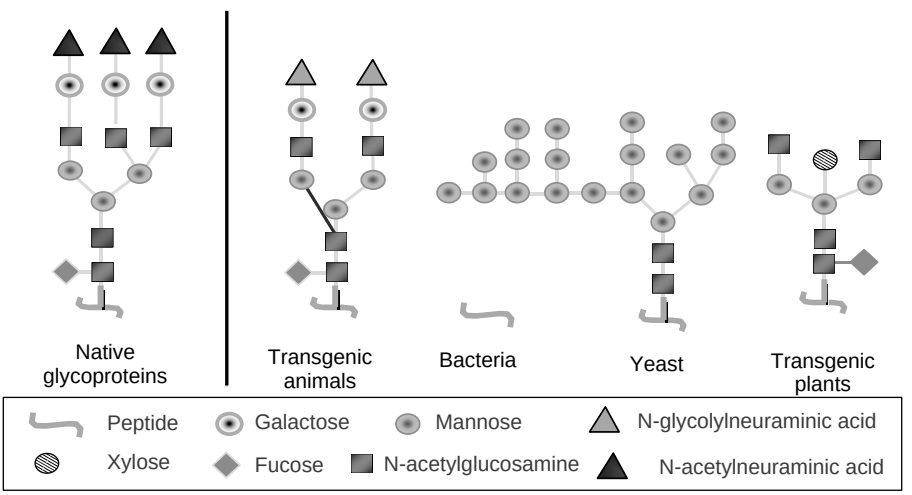

Fig. 5 Engineering plants to humanlike glycosylation. The first step to achieve humanlike glycosylation in plants is to eliminate the plant glycosylation pattern, that is, the attachment of β 1–2 linked xylosyl- and α 1–3 linked fucosyl sugars to the protein. Because these two residues have allergenic potential, the corresponding enzymes Xylosyl- and Fucosyl Transferase are knocked out. In case galactose is relevant for the final product, Galactosyl Transferase is inserted into the host genome. Galactose is available in the organism so that this single gene insertion is sufficient to ensure galactosylation. Source: Knäblein J. (2003) *Biotech: A New Era In The New Millennium – Biopharmaceutic drugs manufactured in novel expression systems*, DECHEMA-Jahrestagung der Biotechnologen, Munich, Germany, 21.

humanlike glycosylation (see Fig. 5) – they must have specific sugar structures attached to the correct sites on the molecule to be maximally effective. Therefore, some efforts are being made in modifying host plants in such a way that they provide the protein with human glycosylation patterns. One example of modifying a plant expression system in this way is the transgenic moss, which will be discussed in the next section.

6
Three Promising Examples: Tobacco (Rhizosecretion, Transfection) and Moss (Glycosylation)

To further elaborate on improving glycosylation and downstream processing, three interesting plant expression systems will be discussed. All systems share the advantage of utilizing nonedible plants (nonfood and nonfeed) and can be kept in either a greenhouse or a fermenter to avoid any segregation risk. Another obvious advantage is secretion of the protein into the medium so that no grinding or extraction is required. This is very important in light of downstream processing: protein purification is often as expensive as the biomanufacturing and should never be underestimated in the total COGS equation.

6.1
Harnessing Tobacco Roots to Secrete Proteins

Phytomedics (Dayton, NJ, USA) uses tobacco plants as an expression system for biopharmaceuticals. Besides the advantage of being well characterized and used in agriculture for some time, tobacco has a stable genetic system, provides high-density tissue (high protein production), needs only simple medium, and can be kept in a greenhouse (see Fig. 6). Optimized antibody expression can be rapidly verified using transient expression assays (short development time) in the plants before creation of transgenic suspension cells or stable plant lines (longer development time). Different vector systems, harboring targeting signals for subcellular compartments, are constructed in parallel and used for transient expression. Applying this screening approach, high expressing cell lines can rapidly be identified. For example, transgenic tobacco plants, transformed with an expression cassette containing the GFP (Green Fluorescent Protein) gene fused to an *aps* (amplification-promoting sequence), had greater levels of corresponding mRNAs and expressed proteins compared to transformants lacking *aps*. Usually, downstream processing (isolation/extraction and purification of the target protein) is limiting for such a system, for example, if the protein has to be isolated from biochemically complex plant tissues (e.g. leaves), this can be a laborious and expensive process and a major obstacle to large-scale protein manufacturing. To overcome this problem, secretion-based systems utilizing transgenic plant cells or plant organs aseptically cultivated *in vitro* would be one solution. However, *in vitro* systems can be expensive, slow growing, unstable, and relatively low yielding. This is why another interesting route was followed. Secretion of molecules is a basic function of plant cells and organs in plants, and is especially developed in plant roots. In order to take up nutrients from the soil, interact with other soil organisms, and defend themselves against numerous pathogens, plant roots have evolved sophisticated mechanisms based on the secretion of different biochemicals (including proteins

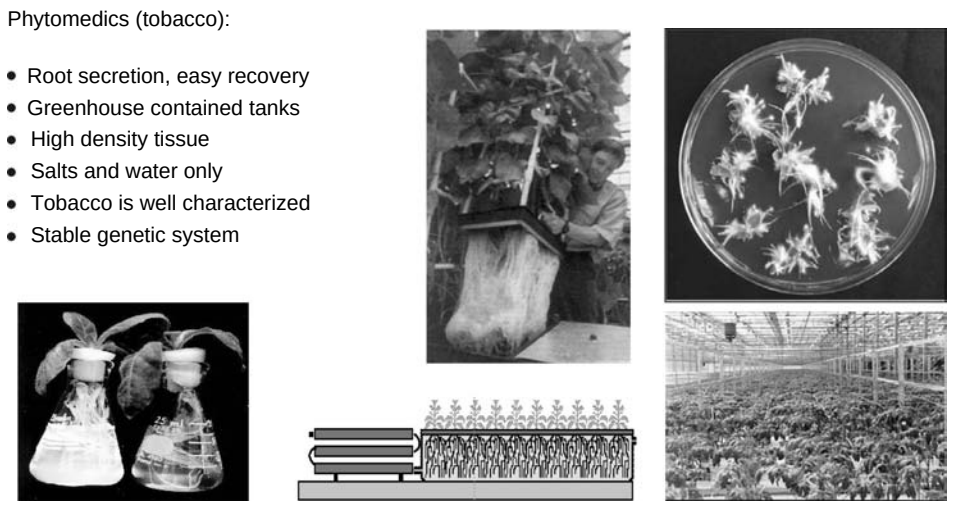

Fig. 6 Secretion of the biopharmaceuticals via tobacco roots. The tobacco plants are genetically modified in such a way that the protein is secreted via the roots into the medium ("rhizosecretion"). In this example, the tobacco plant takes up nutrients and water from the medium and releases GFP (Green Fluorescent Protein). Examination of root cultivation medium by its exposure to near ultraviolet-illumination reveals the bright green-blue fluorescence characteristics of GFP in the hydroponic medium (left flask in panel lower left edge). The picture also shows a schematic drawing of the hydroponic tank, as well as tobacco plants at different growth stages, for example, callus, fully grown, and greenhouse plantation. Source: Knäblein J. (2003) *Biotech: A New Era in the New Millennium – Biopharmaceutic Drugs Manufactured in Novel Expression Systems*, DECHEMA-Jahrestagung der Biotechnologen, Munich, Germany, 21. (See color plate. p. xxv)

like toxins) into their neighborhood (rhizosphere). In fact, Borisjuk and coworkers could demonstrate that root secretion can be successfully exploited for the continuous production of recombinant proteins in a process termed "*rhizosecretion*." Here, an endoplasmic reticulum signal peptide is fused to the recombinant protein, which is then continuously secreted from the roots into a simple hydroponic medium (based on the natural secretion from roots of the intact plants). The roots of the tobacco plant are sitting in a hydroponic tank (see Fig. 6), taking up water and nutrients and continuously releasing the biopharmaceutical. By this elegant set up, downstream processing becomes easy and cost-effective, and also offers the advantage of continuous protein production that integrates the biosynthetic potential of a plant over its lifetime and might lead to higher protein yields than single-harvest and extraction methods. Rhizosecretion is demonstrated in Fig. 6, showing a transgenic tobacco plant expressing GFP and releasing it into the medium.

6.2 High Protein Yields Utilizing Viral Transfection

ICON Genetics (Halle, Germany) has developed a protein-production system that relies on rapid multiplication of viral vectors in an infected tobacco plant (see Fig. 7). Viral transfection systems offer

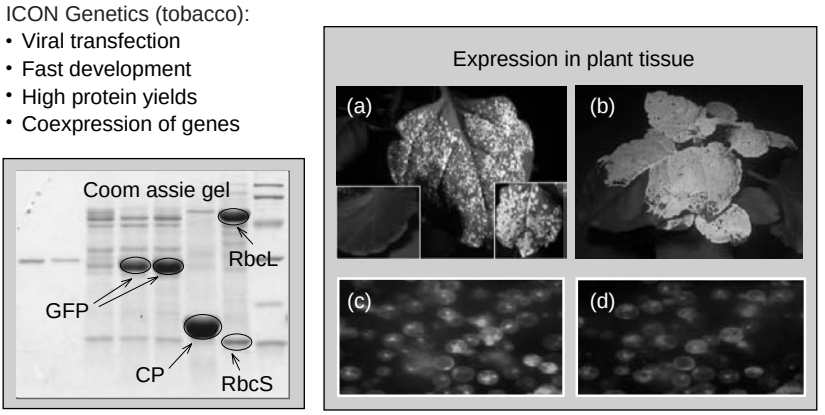

Fig. 7 Viral transfection of tobacco plants. This new generation platform for fast (1 to 2 weeks), high-yield (up to 5 g kg^{-1} fresh leaf weight) production of biopharmaceuticals is based on proviral gene amplification in a nonfood host. Antibodies, antigens, interferons, hormones, and enzymes could successfully be expressed with this system. The picture shows development of initial symptoms on a tobacco following the Agrobacterium-mediated infection with viral vector components that contain a GFP gene (a); this development eventually leads to a systemic spread of the virus, literally converting the plant into a sack full of protein of interest within two weeks (b). The system allows to coexpress two proteins in the same cell, a feature that allows expression of complex proteins such as full-length monoclonal antibodies. Panels (c) and (d) show the same microscope section with the same cells, expressing Green Fluorescent Protein (c) and Red Fluorescent Protein (d) at the same time. The yield and total protein concentration achievable are illustrated by a Coomassie gel with proteins in the system: GFP (protein of interest), CP (coat protein from wild-type virus), RbcS and RbcL (small and large subunit of ribulose-1,5-bisphosphate carboxylase). Source: Knäblein J. (2003) *Biotech: A New Era in the New Millennium – Biopharmaceutic Drugs Manufactured in Novel Expression Systems*, DECHEMA-Jahrestagung der Biotechnologen, Munich, Germany, 21. (See color plate. p. xxv)

a number of advantages, such as very rapid (1 to 2 week) expression time, possibility of generating initial milligram quantities within weeks, high expression levels, and so on. However, the existing viral vectors, such as TMV-based vectors used by, for example, Large Scale Biology Corp. (Vacaville, CA, USA) for production of single-chain antibodies for treatment of non-Hodgkin lymphoma (currently in Phase III clinical trials, see Table 1), had numerous shortcomings, such as inability to express genes larger than 1 kb, inability to coexpress two or more proteins (a prerequisite for production of monoclonal antibodies, because they consist of the light and heavy chains, which are expressed independently and are subsequently assembled), low expression level in systemically infected leaves, and so on. ICON has solved many of these problems by designing a process that starts with an assembly of one or more viral vectors inside a plant after treating the leaves with agrobacteria, which deliver the necessary viral vector components. ICON's proviral vectors provide advantages of fast and high-yield amplification processes in a plant cell, simple and inexpensive assembly of expression cassettes *in planta*, and

full control of the process. The robustness of highly standardized protocols allows the use of inherently the same safe protocols for both laboratory-scale as well as industrial production processes. In this system, the plant is modified transiently rather than genetically and reaches the speed and yield of microbial systems while enjoying posttranslational capabilities of plant cells. De- and reconstructing of the virus adds some safety features and also increases efficiency. There is no "physiology conflict," because the "growth phase" is separated from the "production phase," so that no competition occurs for nutrients and other components required for growth and also for expression of the biopharmaceutical at the same time.

This transfection-based platform allows the production of proteins in a plant host at a cost of US$1 to 10 per gram of crude protein. The platform is essentially free from limitations (gene insert size limit, inability to express more than one gene) of current viral vector-based platforms. The expression levels reach 5 g per kilogram of fresh leaf tissue (or some 50% of total cellular protein!) in 5 to 14 days after inoculation. Since the virus process (in addition to superhigh production of its own proteins, including the protein of interest) leads to the shutoff of the other cellular protein synthesis, the amount of protein of interest in the initial extract is extremely high (Fig. 7). It thus results in reduced costs of downstream processing. Milligram quantities can be produced within two weeks, gram quantities in 4 to 6 months, and the production system is inherently scalable. A number of high-value proteins have been successfully expressed, including antibodies, antigens, interferons, hormones, and enzymes (see Klimyuk, Marillonnet, Knäblein, McCaman, Gleba (2005), books).

6.3 Simple Moss Performs Complex Glycosylation

Greenovation Biotech GmbH (Freiburg, Germany) has established an innovative production system for human proteins. The system produces pharmacologically active proteins in a bioreactor, utilizing a moss (*Physcomitrella patens*) cell culture system with unique properties (see Fig. 8). It was stated before that posttranslational modifications for some proteins are crucial to gain complete pharmacological activity. Since moss is the only known plant system that shows a high frequency of homologous recombination, this is a highly attractive tool for production strain design. By establishing stable integration of foreign genes (gene knockout and new transgene insertion) into the plant genome, it can be programmed to produce proteins with modified glycosylation patterns that are identical to animal cells. The moss is photoautotrophic and therefore only requires simple media for growth, which consist essentially of water and minerals. This reduces costs and also accounts for significantly lower infectious and contamination risks, but in addition to this, the system has some more advantages:

- The transient system allows production of quantities for a feasibility study within weeks – production of a stable expression strain takes 4 to 6 months.
- On the basis of transient expression data, the yield of stable production lines is expected to reach 30 mg L^{-1} per day. This corresponds to the yield of a typical fed-batch culture over 20 days of 600 mg L^{-1}.
- Bacterial fermentation usually requires addition of antibiotics (serving as selection marker and to avoid loss of the

Greenovation (moss system):

- Simple medium (photoautotrophic plant needs only water and minerals)
- Robust expression system (good expression levels from 15 to 25°C)

- Secretion into medium via human leader sequence (broad pH range: 4-8)
- Easy purification from low salt medium via ion exchange

- Easy genetic modifications to cell lines
- Stable cell lines / high genetic stability

- Codon usage like human (no changes required)
- Inexpensive bioreactors from the shelf

- Nonfood plant (no segregation risk)
- Good progress on genetic modification of glycosylation pathways (plant to human)

Fig. 8 Greenovation use a fully contained moss bioreactor. This company has established an innovative production system for human proteins. The system produces pharmacologically active proteins in a bioreactor, utilizing a moss (*Physcomitrella patens*) cell culture system with unique properties. Source: Knäblein J. (2003) *Biotech: A New Era in the New Millennium – Biopharmaceutic drugs Manufactured in Novel Expression Systems*, DECHEMA-Jahrestagung der Biotechnologen, Munich, Germany, 21.

expression vector). For moss cultivation, no antibiotics are needed – this avoids the risk of traces of antibiotics having a significant allergenic potential in the finished product.

- Genetic stability is provided by the fact that the moss is grown in small plant fragments and not as protoplasts or tissue cultures avoiding somaclonal variation.
- As a contained system, the moss bioreactor can be standardized and validated according to GMP standards mandatory in the pharmaceutical industry.
- Excretion into the simple medium is another major feature of the moss bioreactor, which greatly facilitates downstream processing.

As discussed in detail, the first step to get humanlike glycosylation in plants is to eliminate the plant glycosylation, for example, the attachment of β-1-2-linked xylosyl and α-1-3-linked fucosyl sugars to the protein, because these two residues have allergenic potential. Greenovation was able to knockout the relevant glycosylation enzymes xylosyl transferase and fucosyl transferase, which was confirmed by RT-PCR (reverse transcriptase PCR). And indeed, xylosyl and fucosyl residues were completely removed from the glycosylation pattern of the expressed protein as confirmed by MALDI-TOF (matrix assisted laser desorption ionization time of flight) mass spectroscopy analysis (see Fig. 9).

A very challenging protein to express is VEGF because this homodimer consists of two identical monomers linked via a disulfide bond. To produce VEGF in

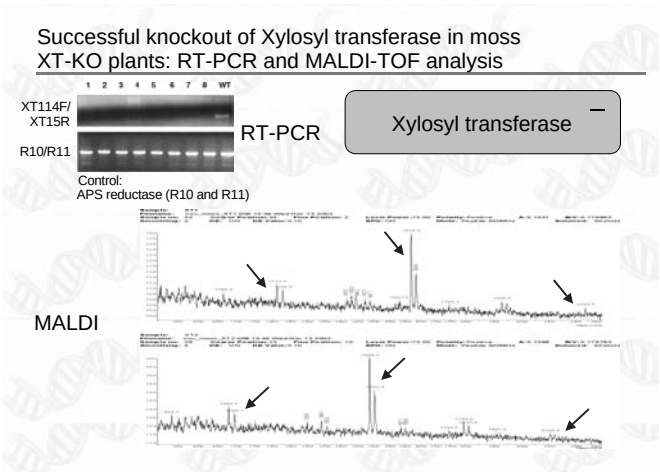

Fig. 9 Knockout of Xylosyl Transferase in moss. To avoid undesired glycosylation, greenovation knocked out the Xylosyl and Fucosyl Transferase, as confirmed by RT-PCR. MALDI-TOF results show that indeed, xylosyl- and fucosyl-residues were completely removed from the glycosylation pattern of the expressed protein (data for knockout of Fucosyl Transferase not shown). Source: Knäblein J. (2003) *Biotech: A New Era in the New Millennium – Biopharmaceutic Drugs Manufactured in Novel Expression Systems*, DECHEMA-Jahrestagung der Biotechnologen, Munich, Germany, 21.

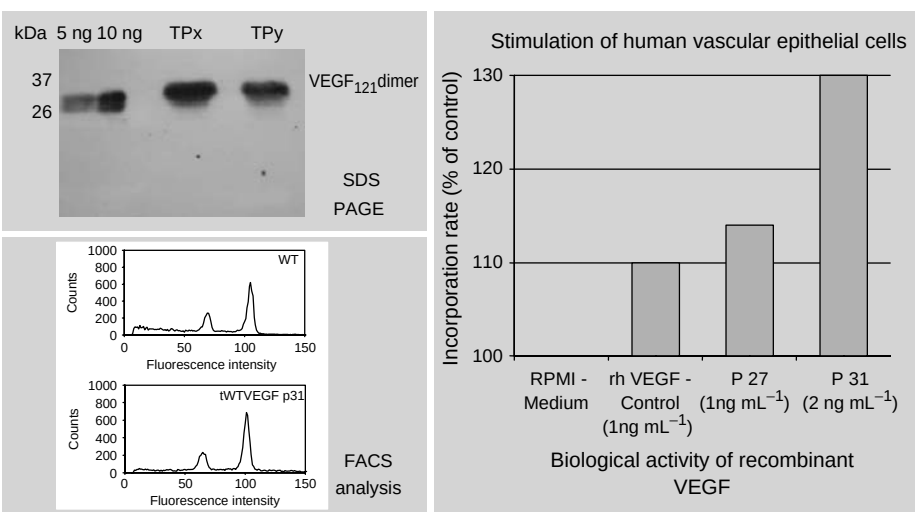

Fig. 10 Greenovation could successfully express the biopharmaceutical VEGF. This growth factor is a very complex protein consisting of two identical monomers linked via a disulfide-bond. To produce VEGF in an active form, the monomers need to be expressed to the right level, correctly folded, assembled, and linked via the disulfide-bond. The analytical assays clearly show that expression in moss yielded completely active VEGF. Source: Knäblein J. (2003) *Biotech: A New Era in the New Millennium – from Plant Fermentation to Plant Expression of Biopharmaceuticals*, PDA International Congress, Prague, Czech Republic.

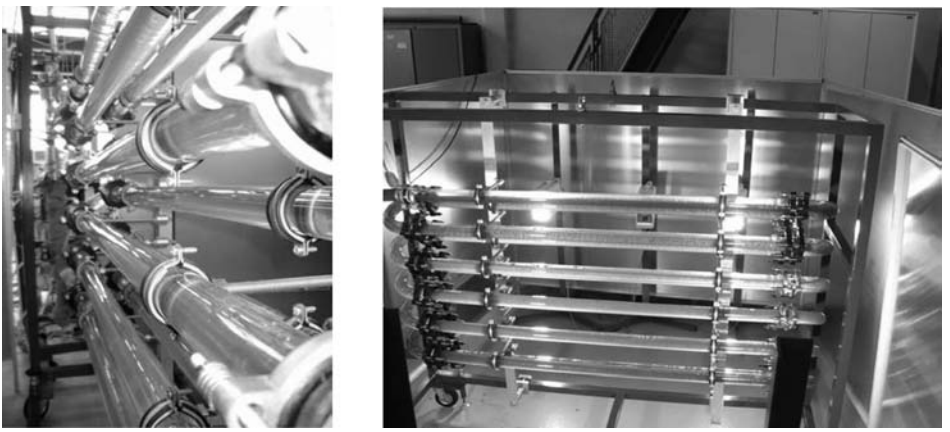

30 L pilot reactor for moss

Two weeks after incubation

Fig. 11 Scaling of photobioreactors up to several 1000 L. The moss bioreactor is based on the cultivation of *Physcomitrella patens* in a fermenter. The moss protonema is grown under photoautotrophic conditions in a medium that consists essentially of water and minerals. Light and carbon dioxide serve as the only energy and carbon sources. Cultivation in suspension allows scaling of the photobioreactors up to several 1000 L. Adaptation of existing technology for large-scale cultivation of algae is done in cooperation with the Technical University of Karlsruhe. Source: greenovation Biotech GmbH (Freiburg, Germany) and Professor C. Posten, Technical University (Karlsruhe, Germany).

an active form, the following need to be provided:

- Monomers need to be expressed to the right level.
- Monomers need to be correctly folded.
- Homodimer needs to be correctly assembled and linked via a disulfide bond.
- Complex protein needs to be secreted in its active form.

And in fact, all this could be achieved with the transgenic moss system as shown in Fig. 10. These results are very promising because they demonstrate that this system is capable of expressing even very complex proteins. In addition to that, the moss system adds no plant-specific sugars to the protein – a major step toward humanlike glycosylation. Furthermore, moss is a robust expression system leading to high yields at 15 to 25 °C and the pH can be adjusted from 4 to 8 depending on the optimum for the protein of interest. Adapting existing technology for large-scale cultivation of algae, fermentation of moss in suspension culture allows scaling of the photobioreactors up to several 1000 L (see Fig. 11). Finally, the medium is inexpensive, since only water and minerals are sufficient.

7 Other Systems Used for Plant Expression

Several different plants have been used for the expression of proteins in plants. All these systems have certain advantages regarding edibility, growth rate, scalability, gene-to-protein time, yield, downstream processing, ease of use, and so on, which I will not discuss in further detail here. A selection of different expression systems is listed:

Alfalfa	Ethiopian mustard	Potatoes
Arabidopsis	*Lemna*	Rice
Banana	Maize	Soybean
Cauliflower	Moss	Tomatoes
Corn	Oilseeds	Wheat

Some of these systems have been used for research on the basis of their ease of transformation, well-known characterization, and ease to work with. However, they are not necessarily appropriate for commercial production. Which crop is ultimately used for full-scale commercial production will depend on a number of factors including

- time to develop an appropriate system (gene-to-protein);
- section of the plant expressing the product/possible secretion;
- cost and potential waste products from extraction;
- "aged" product/ease of storage;
- long-term stability of the storage tissue;
- quantities of protein needed (scale of production).

Depending on the genetic complexity and ease of manipulation, the development time to produce an appropriate transgenic plant for milligram production of the desired protein can vary from 10 to 12 months in corn as compared to only weeks in moss. Estimates for full GMP production in corn are 30 to 36 months and approximately 12 months for moss. Expression of the protein in various tissues of the plant can result in a great variation in yield. Expression in the seed can often lead to higher yields than in the leafy portion of the plant. This is another explanation for the high interest in using corn, which has a relatively high seed-to-leaf ratio. Extraction from leaf can be costly as it contains a high percentage of water, which could result in unavoidable proteolysis during the process. Proteins stored in seeds can be desiccated and remain intact for long periods of time. The purification and extraction of the protein is likely to be done by adaptations of current processes for the extraction and/or fractionation. For these reasons, it is anticipated that large-scale commercial production of recombinant proteins will involve grain and oilseed crops such as maize, rice, wheat, and soybeans. On the basis of permits for open-air test plots issued by the USDA for pharmaceutical proteins and industrial biochemicals, corn is the crop of choice for production with 73% of the permits issued. The other major crops are soybeans (12%), tobacco (10%), and rice (5%).

In general, the use of smaller plants that can be grown in greenhouses is an effective way of producing the biopharmaceuticals and alleviating concerns from environmental activist groups that the transgenic plant might be harmful to the environment (food chain, segregation risk, genetic drift, etc.).

8 Analytical Characterization

Validated bioanalytical assays are essential and have to be developed to characterize the biopharmaceuticals during the production process (e.g. in-process control) and to release the final product for use as a drug in humans. These assays are applied to determine characteristics such as purity/impurities, identity, quantity, stability, specificity, and potency of the recombinant protein during drug development. Since the very diverse functions of different proteins heavily depend on their structure, one very valuable parameter in

protein characterization is the elucidation of their three-dimensional structure. Although over the last couple of years a lot of effort was put into a method for improving the elucidation of protein structures (during my PhD thesis, I was also working in this fascinating field together with my boss Professor Robert Huber, Nobel Prize Laureate in 1988, "for the determination of the three-dimensional structure of a photosynthetic reaction centre"), it is still very time consuming to solve the 3-D structure of larger proteins. This is why despite the high degree of information that can be obtained from the protein structure, this approach cannot be applied on a routine basis. Therefore, tremendous efforts are put into the development of other assays to guarantee that a potent biopharmaceutical drug is indeed ready for use in humans.

9
Conclusion and Outlook

The production of protein therapeutics from transgenic plants is becoming a reality. The numerous benefits offered by plants (low cost of cultivation, high biomass production, relatively fast gene-to-protein time, low capital and operating costs, excellent scalability, eukaryotic post-translational modifications, low risk of human pathogens, lack of endotoxins, as well as high protein yields) virtually guarantee that plant-derived proteins will become more and more common for therapeutic uses. Taking advantage of plant expression systems, the availability of cheap protein-based vaccines in underdeveloped countries of the world is possible in the near future. The cost of very expensive hormone therapies (erythropoietin, human growth hormone, etc.) could fall dramatically within the next decade because of the use of, for example, plant expression systems. Fears about the risks of the plant expression technology are real and well founded, but with a detailed understanding of the technology, it is possible to proactively address these safety issues and create a plant expression industry almost free of mishaps. For this purpose, the entire set up, consisting of the specific plant expression system and the protein being produced, needs to be analyzed and its potential risks assessed on a case-by-case basis. As plant-derived therapeutics begin to demonstrate widespread, tangible benefits to the population and as the plant expression industry develops a longer safety track record, public acceptance of the technology is likely to improve continuously. Plants are by far the most abundant and cost-effective renewable resource uniquely adapted to complex biochemical synthesis. The increasing cost of energy and chemical raw materials, combined with the environmental concerns associated with conventional pharmaceutical manufacturing, will make plants even more compatible in the future. With the words of Max Planck (1858–1947) "How far advanced Man's scientific knowledge may be, when confronted with Nature's immeasurable richness and capacity for constant renewal, he will be like a marveling child and must always be prepared for new surprises," we will definitely discover more fascinating features of plant expression systems. But there is no need to wait: combining the advantages of some technologies that we already have in hand could lead to the ultimate plant expression system. This is what we should focus on, because, then, at the dawn of this new millennium, this would for the first time yield large-enough amounts of biopharmaceuticals to treat everybody on our planet!

Acknowledgments

I would like to thank the companies greenovation Biotech GmbH (Freiburg, Germany), ICON Genetics (Halle, Germany), and Phytomedics (Dayton, NJ, USA) for providing some data and figures to prepare this manuscript.

See also Bioprocess Engineering; Expression Systems for DNA Processes; Plant Gene Expression, Regulation of.

Bibliography

Books and Reviews

Fischer, R., Schillberg S. (Eds.) (2004) *Molecular Farming: Plant-made Pharmaceuticals and Technical Proteins*, Wiley, ISBN: 3-527-30786-9

Fischer, R., et al. (2004) Plant-based production of biopharmaceuticals, *Curr. Opin. Plant Biol.* **7**(2), 152–158.

Horn, M.E., Woodard, S.L., Howard, J.A. (2004) Plant molecular farming: systems and products, *Plant Cell Rep.* **22**(10), 711–720.

Klimyuk, V., Marillonnet, S., Knäblein, J., McCaman, M., Gleba, Y. (2005) Production of Recombinant Proteins in Plants, in: *Modern Biopharmaceuticals – Design, Development and Optimization*, Wiley-VCH, in press.

Knäblein, J. (2004) Biopharmaceuticals Expressed in Plants – A New Era in the New Millennium, in: Müller, R., Kayser, O. (Eds.) *Applications in Pharmaceutical Biotechnology*, Wiley-VCH, ISBN 3-527-30554-8.

Ma, J.K., Drake, P.M., Christou, P. (2003) The production of recombinant pharmaceutical proteins in plants, *Nat. Rev. Genet.* **4**(10), 794–805.

Stoger, E., et al. (2004) Antibody production in transgenic plants, *Methods Mol. Biol.* **248**, 301–318.

Primary References

Arakawa, T., Chong, D.K.X., Langridge, W.H.R. (1998) Efficacy of a food plant-based oral cholera toxin B subunit vaccine, *Nat. Biotechnol.* **16**, 292–297.

Arakawa, T., Yu, J., Chong, D.K., Hough, J., Engen, P.C., Langridge, W.H. (1998) A plant-based cholera toxin B subunit-insulin fusion protein protects against the development of autoimmune diabetes, *Nat. Biotechnol.* **16**, 934–938.

Arthur D. Little, Inc. (ADL), *AgIndustries Research*, Cambridge, MA, Copyright© 2002.

Artsaenko, O., et al. (1998) Potato tubers as a biofactory for recombinant antibodies, *Mol. Breeding* **4**, 313–319.

Beachy, R.N., Fitchen, J.H., Hein, M.B. (1996) Use of Plant Viruses for Delivery of Vaccine Epitopes, in: Collins, G.B., Sheperd, R.J. (Eds.) *Engineering Plants for Commercial Products and Applications*, New York Academy of Sciences, New York, pp. 43–49.

Boothe, J.G., Parmenter, D.L., Saponja, J.A. (1997) Molecular farming in plants: oilseeds as vehicles for the production of pharmaceutical proteins, *Drug Develop. Res.* **42**, 172–181.

Borisjuk, N.V., Raskin, I., et al. (1999) Production of recombinant proteins in plant root exudates, *Nat. Biotechnol.* **17**, 466–469.

Borisjuk, N.V., Raskin, I., et al. (2000) Tobacco ribosomal DNA spacer element stimulates amplification and expression of heterologous genes, *Nat. Biotechnol.* **18**, 1303–1306.

Cabanes-Macheteau, M., et al. (1999) N-glycosylation of a mouse IgG expressed in transgenic tobacco plants, *Glycobiology* **9**, 365–372.

Chance, R.E., Frank, B.H. (1993) Research, development, production and safety of biosynthetic human insulin, *Diabetes care* **16**(3), 133–142.

Chaudhary, S., Parmenter, D.L., Moloney, M.M. (1998) Transgenic Brassica carinata as a vehicle for the production of recombinant proteins in seeds, *Plant Cell Rep.* **17**, 195–200.

Conrad, U., Fiedler, U. (1994) Expression of engineered antibodies in plant cells, *Plant Mol. Biol.* **26**, 1023–1030.

Conrad, U., Fiedler, U., Artsaenko, O., Phillips, J. (1998) High-level and stable accumulation of single-chain Fv antibodies in plant storage organs, *J. Plant Physiol.* **152**, 708–711.

Cramer, C.L., Boothe, J.G., Oishi, K.K. (1999) Transgenic plants for therapeutic proteins: linking upstream and downstream strategies, *Curr. Top. Microbiol. Immunol.* **240**, 95–118.

Cramer, C.L., et al.. (1996) Bioproduction of Human Enzymes in Transgenic Tobacco, in: Collins, G.B., Sheperd, R.J. (Eds.) *Engineering Plants for Commercial Products and Applications*, New York Academy Of Sciences, New York, 62–71.

Dalsgaard, K., et al. (1997) Plant-derived vaccine protects target animals against a viral disease, *Nat. Biotechnol.* **15**, 248–252.

Davies, L., Plieth, J. (2001) The challenge of meeting the escalating demand for proteins, *Scr Mag* **10**, 25–29.

Della-Cioppa, G., Grill, L.K. (1996) Production of Novel Compounds in Higher Plants by Transfection with RNA Viral Vectors, in: Collins, G.B., Sheperd, R.J. (Eds.) *Engineering Plants for Commercial Products and Applications*, New York Academy of Sciences, New York, pp. 57–61.

Dieryck, W., et al. (1997) Human haemoglobin from transgenic tobacco, *Nature* **386**, 29–30.

Doran, P.M. (2000) Foreign protein production in plant tissue cultures, *Curr. Opin. Biotechnol.* **11**, 199–204.

Dove, A. (2001) The bittersweet promise of glycobiology, *Nat. Biotechnol.* **19**, 913–917.

Dove, A. (2002) Unkorking the biomanufacturing bottleneck, *Nat. Biotechnol.* **20**, 777–779.

Drake, P.M., Chargelegue, D., Vine, N.D., Van Dolleweerd, C.J., Obregon, P., Ma, J.K. (2002) Transgenic plants expressing antibodies: a model for phytoremediation, *FASEB J.* **16**(14), 1855–1860.

Drug & Market Development Publications, *Antibody Engineering: Technologies, Applications and Business opportunities*, Westborough, MA, Copyright© 2003.

Evangelista, R.L., Kusnadi, A.R., Howard, J.A., Nikolov, Z.L. (1988) Process and economic evaluation of the extraction and purification of recombinant glucouronidase from transgenic corn, *Biotechnol. Prog.* **14**, 607–614.

Fischer, R., Emans, N. (2000) Molecular farming of pharmaceutical proteins, *Transgenic Res.* **9**, 279–299.

Fischer, R., Hoffmann, K., Schillberg, S., Emans, N. (2000) Antibody production by molecular farming in plants, *J. Biol. Regul. Homeost. Agents* **14**, 83–92.

Ganz, P.R., et al.. (1996) Expression of Human Blood Proteins in Transgenic Plants: The Cytokine GM-CSF as a Model Protein, in: Owen, M.R.L., Pen, J. (Eds.) *Transgenic Plants: A Production System for Industrial and Pharmaceutical Proteins*, John Wiley & Sons, London, UK, 281–297..

Garber, K. (2001) Biotech industry faces new bottleneck, *Nat. Biotechnol.* **19**, 184–185.

Giddings, G., Allison, G., Brooks, D., Carter, C. (2000) Transgenic plants as factories for biopharmaceuticals, *Nat. Biotechnol.* **18**, 1151–1155.

Goddijn, O.J.M., Pen, J. (1995) Plants as bioreactors, *Trends Biotechnol.* **13**, 379–387.

Hamamoto, H., et al. (1993) A new tobacco mosaic virus vector and its use for the systematic production of angiotensin-I-converting enzyme inhibitor in transgenic tobacco and tomato, *Biotechnology* **11**, 930–932.

Hiatt, A., Cafferkey, R., Bowdish, K. (1989) Production of antibodies on transgenic plants, *Nature* **342**, 76–78.

Hood, E.E., Jilka, J.M. (1999) Plant-based production of xenogenic proteins, *Curr. Opin. Biotechnol.* **10**, 382–386.

Johnson, E. (1996) Edible plant vaccines, *Nat. Biotechnol.* **14**, 1532–1533.

Knäblein, J. (2003) Biotech: A new era in the new millennium – fermentation and expression of biopharmaceuticals in plants, *SCREENING – Trends Drug Discov* **4**, 14–16.

Knäblein, J., McCaman, M. (2003) Modern biopharmaceuticals-recombinant protein expression in transgenic plants, *SCREENING – Trends Drug Discov* **6**, 33–35.

Knäblein, J., Huber, R., et al. (1997) [Ta6Br12]2+, a tool for phase determination of large biological assemblies by X-ray crystallography, *J. Mol. Biol.* **270**, 1–7.

Kumagai, M.H., et al. (1993) Rapid, high-level expression of biologically active alpha-trichosanthin in transfected plants by an RNA viral vector, *Proc. Natl. Acad. Sci. USA* **90**, 427–430.

Kusnadi, A., Nikolov, Z.L., Howard, J.A. (1997) Production of recombinant proteins in transgenic plants: practical considerations, *Biotechnol. Bioeng.* **56**, 473–484.

Ma, J.K.C. (2000) Genes, greens, and vaccines, *Nat. Biotechnol.* **18**, 1141–1142.

Ma, J.K.C., Hein, M.B. (1995) Plant antibodies for immunotherapy, *Plant Physiol.* **109**, 341–346.

Ma, J.K.C., Hein, M.B. (1996) Antibody Production and Engineering in Plants, in: Collins, G.B., Sheperd, R.J. (Eds.) *Engineering Plants for Commercial Products and*

Applications, New York Academy of Sciences, New York, pp. 72–81.

Ma, J.K.C., Hiatt, A. (1996) Expressing Antibodies in Plants for Immunotherapy, in: Owen, M.R.L., Pen, P. (Eds.) *Transgenic Plants: A Production System for Industrial and Pharmaceutical Proteins*, John Wiley & Sons, London, UK, pp. 229–243.

Ma, J.K.C., Vine, N.D. (1999) Plant expression systems for the production of vaccines, *Curr. Top. Microbiol. Immunol.* **236**, 275–292.

Ma, J.K.C., et al. (1998) Characterization of a recombinant plant monoclonal secretory antibody and preventive immunotherapy in humans, *Nat. Med.* **4**(5), 601–606.

Ma, S.W., et al. (1997) Transgenic plants expressing autoantigens fed to induce oral immune tolerance, *Nat. Med.* **3**, 793–517.

McCormick, A.A., et al. (1999) Rapid production of specific vaccines for lymphoma by expression of the tumor-derived single-chain Fv epitopes in tobacco plants, *Proc. Natl. Acad. Sci. USA* **96**, 703–708.

McGarvey, P.B., et al. (1995) Expression of the rabies virus glycoprotein in transgenic tomatoes, *Biotechnology* **13**, 1484–1487.

Moloney, M.M. (1995) "Molecular farming" in plants: achievements and prospects, *Biotechnol. Eng.* **9**, 3–9.

Morrow, K.J. (2002) Economics of antibody production, *Genet. Eng. News* **22**, 34–39.

Mushegian, A.R., Shepard, R.J. (1995) Genetic elements of plant viruses as tools for genetic engineering, *Microbiol. Rev.* **59**, 548–578.

Parmenter, D.L., et al. (1995) Production of biologically active hirudin in plant seeds sing oleosin partitioning, *Plant Mol. Biol.* **29**, 1167–1180.

Pen, J. (1996) Comparison of Host Systems for the Production of Recombinant Proteins, in: Owen, M.R.L., Pen, J. (Eds.) *Transgenic Plants: A Production System for Industrial and Pharmaceutical Proteins*, John Wiley & Sons, London, UK, pp. 149–167.

Ponstein, A.S., Verwoerd, T.C., Pen, J. (1996) Production of Enzymes for Industrial Use, in: Collins, G.B., Sheperd, R.J. (Eds.) *Engineering Plants for Commercial Products and Applications*, Vol. 792, New York Academy of Sciences, New York, pp. 91–98.

Raskin, I., Fridlender, B., et al. (2002) Plants and human health in the twenty-first century, *Trends Biotechnol.* **20**, 522–531.

Richter, L.J., Thanavala, Y., Arntzen, C.J., Mason, H.S. (2000) Production of hepatitis B surface antigen in transgenic plants for oral immunization, *Nat. Biotechnol.* **18**, 1167–1171.

Ruggiero, F., et al. (2000) Triple helix assembly and processing of human collagen produced in transgenic tobacco plants, *FEBS Lett.* **469**, 132–136.

Sijmons, P.C., et al. (1990) Production of correctly processed human serum albumin in transgenic plants, *Biotechnology* **8**, 217–221.

Smith, M.D. (1996) Antibody production in plants, *Biotechnol. Adv.* **14**, 267–281.

Smith, M.D., Glick, B.R. (2000) The production of antibodies in plants, *Biotechnol. Adv.* **18**, 85–89.

Stoger, E., et al. (2000) Cereal crops as viable production and storage systems for pharmaceutical scFv antibodies, *Plant Mol. Biol.* **42**, 583–590.

Tacket, C.O., Mason, H.S. (1999) A review of oral vaccination with transgenic vegetables, *Microbes Infect.* **1**, 777–783.

Tacket, C.O., et al. (1998) Immunogenicity in humans of a recombinant bacterial antigen delivered in a transgenic potato, *Nat. Med.* **4**, 607–609.

Technology Catalysts International Corporation, *Biopharmaceutical Farming*, Falls Church, VA, Copyright© 2002.

Thanavala, Y., et al. (1995) Immunogenicity of transgenic plant-derived hepatitis B surface antigen, *Proc. Natl. Acad. Sci. USA* **92**, 3358–3361.

The Context Network, Biopharmaceutical Production in Plants, *Biopharma Prospectus*, West Des Moines, IA, Copyright© 2002.

Tomsett, B., Tregova, A., Garoosi, A., Caddick, M. (2004) Ethanol-inducible gene expression: first step toward a new green revolution? *Trends Plant Sci.* **9**(4), 159–161.

Valdes, R., et al. (2003) Hepatitis B surface antigen immunopurification using a plant-derived specific antibody produced in large scale, *Biochem. Biophy. Res. Commun.* **310**, 742–747.

Vandekerckhove, J., et al. (1989) Enkephalines produced in transgenic plants using modified 2S storage proteins, *Biotechnology* **7**, 929–932.

Whitelam, G.C. (1995) The production of recombinant proteins in plants, *J. Sci. Food Agric.* **68**, 1–9.

Whitelam, G.C., Cockburn, W. (1996) Antibody expression in transgenic plants, *Trends Plant Sci.* **1**, 268–272.

Zhong, G.Y., et al. (1999) Commercial production of aprotinin in transgenic maize seeds, *Mol. Breeding* **5**, 345–356.

Plasmids

Dhruba K. Chattoraj
Center for Cancer Research, NCI, NIH, Bethesda, MD, USA

1 Basic Facts 413

2 Maintenance 414
2.1 Replication 414
2.1.1 Principle of Replication Control 414
2.1.2 Control of Primer Formation by Antisense RNA 415
2.1.3 Control of Initiator Synthesis by Antisense RNA 418
2.1.4 Control of Initiator Synthesis by Iterons 422
2.2 Partition 425
2.2.1 Regulation of *par* 425
2.2.2 Partition Dynamics 426
2.2.3 Unexpected Aspects 427
2.3 Postsegregational Killing 428
2.3.1 Classification and Regulation of Addiction Systems 429
2.3.2 Addiction Systems in Chromosomes 432
2.4 Multimer Resolution 432
2.5 Incompatibility 434
2.6 Cell cycle Control 435

3 Horizontal Spread 436
3.1 Conjugation 437
3.1.1 Mechanisms of DNA Transfer 437
3.1.2 Establishment of Transferred DNA 437
3.1.3 Regulation of Transfer 438
3.2 Broad Host Range 439
3.3 Bacterial Evolution 440
3.4 Impact on Public Health 441

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

4 Concluding Remarks 442
4.1 As Model Chromosomes 442
4.2 Contributions 442
4.3 Future Potential 443

Bibliography 443
Books and Reviews 443
Primary Literature 444

Keywords

Antisense RNA
Discovered as RNA that negatively controls replication initiation by hybridizing to primer RNA. Subsequently, it was found to control mRNAs particularly involved in initiator synthesis. These RNA–RNA interactions are the mechanistic basis of incompatibility in plasmids under antisense control.

Basic Replicon
Minimal sequences that allow plasmid replication characteristic of the native plasmid.

Horizontal Transfer
Horizontal (also called lateral) gene transfer is the gene exchange between individuals and may occur across species boundaries.

Incompatibility
Defined as the inability of two plasmids to be maintained in the same cell without selection. This happens if the plasmids in question have a common replication control system. Incompatibility can also arise if the two plasmid partition systems are closely related.

Iterons
Repeat sequences that bind plasmid-encoded initiator protein for replication initiation. The binding is also involved in controlling replication and causing incompatibility.

Partition
A process that ensures plasmid distribution from mother to both daughter cells.

Plasmids are extrachromosomal elements common in bacteria and archea but also found in virally transformed eukaryotic cells. The study of plasmid biology has contributed significantly to understanding regulatory principles in chromosome replication and segregation, gene targeting, gene amplification, maintenance of genome identity, and genome evolution. However, the most profound impact of plasmids on life sciences has been through their use as cloning and expression vectors. Plasmids can also have remarkably broad host range that can cross species and even kingdom boundaries. Thus, plasmids are powerful agents for gene transfer, and by the same token, they are a potential risk to human health as they often carry genes for antibiotic resistance and virulence factors. Plasmids also carry useful genes, for example, those that encode functions capable of degrading organic pollutants or fixing nitrogen. Given the prevalence of plasmids in natural bacterial isolates, undoubtedly, much remains to be discovered about the roles they play.

1 Basic Facts

Information needed for life's basic processes is present in a coded form in the DNA of chromosomes, whether the cells are bacterial, fungal, plant, or animal. But many species, bacteria in particular, contain additional DNA molecules much smaller than chromosomes, which they transmit faithfully from one generation to the next. These molecules are called *plasmids*. The term was introduced in 1952 by Joshua Lederberg as a generic term for genetic elements that can persist outside of chromosomes. Our knowledge of plasmids has come mostly from studies in bacteria, and in this article, the term "plasmid" will refer to extrachromosomal elements of bacterial provenance. Plasmids are not essential to their host cells but rather carry specialized information that enables the cell to cope with and survive sometimes threatening environments. Plasmids have played a crucial role in the understanding of life processes, out of proportion to their diminutive size and dispensability to the life of the cell. These features actually made plasmids attractive as research objects. Their small size made physical manipulations easy and their dispensability allowed manipulations of genes connected to vital functions such as DNA replication without worrying about consequences to the host survival. Plasmids are of interest in several disciplines. With the advent of recombinant DNA technology in the early 1970s, they have become indispensable tools for research in molecular biology. They are studied for their special features such as their being carriers of antibiotic resistance and virulence genes that have impacted profoundly the public health. The postgenomic era has kindled new interest in plasmids for their astonishing role in bacterial adaptation and diversity. Finally, plasmids are studied in their own rights to understand functions that maintain them so efficiently outside of chromosomes.

This article mainly covers the current understanding of plasmid-maintenance functions where research has been most in-depth and rewarding in terms of understanding biological regulatory principles.

2 Maintenance

DNA replication is the primary requirement for plasmid maintenance. Additionally, plasmids may have several other functional units (cassettes) that contribute to plasmid maintenance in a growing host population. Their requirement for stability becomes more critical the lower the plasmid copy number. The maintenance functions are highly controlled. Understanding the molecular basis of control has been the main goal of basic research in plasmid biology.

2.1 Replication

The basic mechanism of DNA replication is a well-conserved process and plasmids are no exception. Replication can be divided into three stages: initiation, elongation, and termination. Initiation usually occurs in conjunction with the binding of a plasmid-specified initiator protein to a locus called the origin or *ori*, resulting in the formation of a higher order structure that also involves host-encoded proteins. This leads to opening of the strands of the origin DNA duplex that allows loading of helicase for unwinding the rest of the plasmid. The synthesis of new DNA strands on the opened template strands by DNA polymerases requires a primer providing a 3′-OH end. A variety of priming mechanisms have been found from studies on bacteriophages and from the *Escherichia coli* chromosomal origin. The events subsequent to priming, elongation, and termination of replication, are remarkably similar in different replicons. Since plasmids control their replication at the stage of initiation, aspects of elongation and termination will not be discussed here.

2.1.1 Principle of Replication Control

Plasmids have defined copy numbers that can range from one to more than one hundred in a newly born cell. However, because of the stochastic nature of all biological reactions, fluctuations in copy number can arise despite stringent controls. Large fluctuations also arise in the case of plasmids like ColE1, which is believed to distribute randomly to daughter cells. If these fluctuations are not adjusted, the result is a broadening of distribution of plasmid copies per cell. This is undesirable for plasmid maintenance. Cells with too few plasmids can generate plasmid-free daughter cells. At the other extreme, cells with too many plasmids grow slowly because of the excess genetic load or because of toxicity from excessive plasmid-encoded gene products. In the absence for selection for a plasmid, these cells are likely to be outcompeted by faster growing plasmid-free cells of the population. Thus, narrowing of plasmid copy number distribution is imperative for stable plasmid maintenance. Adjusting the rate of replication does this. Instead of doubling the copy number in a cell cycle, cells with too many plasmids make fewer new copies, and the cells with too few copies overreplicate their plasmids to catch up to the desired number. This ability to make the rate of replication a decreasing function of plasmid copy number (autogenous control or negative feedback control) is the most basic requirement for plasmid maintenance. In fact, the elements that control replication have been found to be plasmid-borne without exception, signifying the importance of the function for plasmid survival. A few prototypes of

plasmid initiation control mechanisms are illustrated below.

2.1.2 Control of Primer Formation by Antisense RNA

Control of initiation is best understood for the high copy plasmid, ColE1, of *E. coli*. ColE1 copy number is about 15, meaning it is maintained at about 15 times higher than the number of *E. coli* chromosomal origin, *oriC*. (The *oriC* number increases with growth rate and so does the plasmid copy number such that the ratio of the two remains relatively constant). ColE1 has been crucial in understanding replication control for the commonly used multicopy cloning vectors such as pBR322 and pACYC184.

Replication of ColE1 type plasmids differ from many other replicons in that the ColE1 replicons do not encode any initiator protein and they depend totally on host-encoded proteins for their replication. The plasmids nevertheless code for elements that control their copy number. The mechanisms of initiation and its control have been elaborated in great detail by *in vitro* studies pioneered by J. Tomizawa at NIH. The housekeeping RNA polymerase of *E. coli* makes the preprimer RNA (RNAII) constitutively (Fig. 1). When the RNA is about 550 nucleotides (nt) long, it starts hybridizing to the template making a R-loop that could be more than 200 bp long. Normally during transcription elongation, only 10 nt of the leading edge of the RNA hybridize to the template. To form the long R-loop, RNAII must acquire specific secondary and tertiary structures, which form as RNA II is being synthesized. RNase H then digests the RNA out of the loop almost completely except for the 4 to 5 bases in the beginning of the loop (primer maturation). This residual hybrid suffices to serve as the primer for DNA polymerase I to start DNA synthesis. The switch point of RNA to DNA has been defined as the origin for this plasmid.

A shorter RNA (RNA I) about 110 nt long mediates the control of initiation. It is fully complementary to the 5′-end of RNA II and is therefore called *antisense RNA*. The hybridization of RNA I to RNA II prevents R-loop formation at the origin sequence and thereby inhibits priming of DNA synthesis. The RNA polymerase now behaves canonically and transcribes through the origin DNA as any other template without making long R-loops. It is reasonably clear that proper folding of RNAII is crucial for R-loop formation. The folding is altered upon hybridization with RNA I. In ColE1, primer maturation is rate-limiting for initiation of replication and this process is controlled by RNA I. There is an accessory regulator, Rom, that strengthens RNA I–RNA II hybridization, thereby making primer maturation less probable (Fig. 1). In the absence of Rom, the copy number increases about two/three fold.

Several general remarks on replication control can be made using the ColE1 example. Plasmids control their replication at the stage of initiation. The rate of initiation is controlled primarily by a plasmid-encoded negative regulator (RNA I). While the primer RNA acts in cis, the negative controller must be able to work in trans, and its inhibitory power must increase with plasmid copy number. (The negative element has to be able to function in trans so that the plasmid copies can communicate with each other, but the positive element can be in either cis or trans; the only requirement is that the replication frequency in the absence of inhibitor is higher than one per plasmid and cell cycle, preferably much higher.) RNA I is made

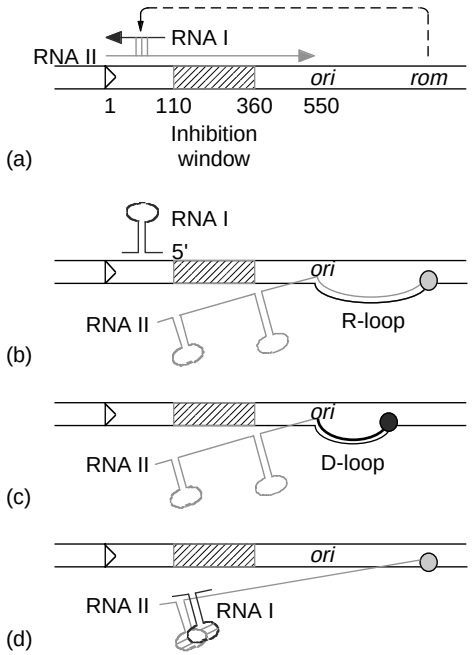

Fig. 1 Replication control in plasmid ColE1. Double-stranded DNA is shown as two parallel lines. Base pair 1 is the starting point of transcription of preprimer RNA, RNA II (gray arrow, a). For replication initiation, it is imperative that RNA II, when about 550 nt long, starts hybridizing to template strand and forms a large R-loop (b). RNase H subsequently cleaves the R-loop (not shown) and the 3′-end of the transcript is used as the primer for DNA synthesis (c). The transition point of RNA to DNA is called origin (*ori*) in this system. If RNA I (which is complementary to the 5′-terminal region of RNA II covering nt 110 to 2, leftward arrow in a) hybridizes to RNA II during the time it is being made from the inhibition window (bp 110 to 360), R-loop formation and replication initiation do not happen (d). The Rom protein helps in RNA I/RNA II hybrid formation. The schematic also shows that RNA II folding relevant to R-loop formation alters when it hybridizes to RNA I. Formation of one or more stem-loop structures is common among the regulatory RNAs in which the stem provides metabolic stability and the loop, the sequences for initial interactions with complementary RNA sequences. Other symbols: empty arrow head, promoter; gray circle, RNA polymerase; black circle, DNA polymerase.

from a constitutive promoter and is diffusible (i.e. can work in trans). As the plasmid copy number increases so does the concentration of RNA I. This in turn increases the probability of interaction with the target, the preprimer RNA. When the plasmid concentration approaches a characteristic threshold, the RNA I concentration becomes high enough to engage every preprimer RNA and prevent any further increase in plasmid copy number. Replication can resume when the RNA I concentration is reduced by dilution due to the increase in cell volume due to cell growth, and by degradation by RNases.

Efficiency of ColE1 replication control Plasmid stability improves with the rapidity with which copy number fluctuations are adjusted in a cell cycle. In an efficiently controlled system, the replication on/off switch should be tripped in response to small deviations from the steady state copy number. Ideally, replication should be negligible when plasmid concentration is high and should be frequent when plasmid concentration is lower than the steady state.

An understanding of the efficiency of control requires analysis of the dynamics of the process. It has been shown that R-loop formation can only be inhibited when the primer RNA being made is of lengths from 110 to 360 nt (Fig. 1). When the RNAII is shorter than 110 nt, it does not hybridize to RNA I. When it is longer than 360, the hybrid formation does not prevent R-loop formation, since the structure of the folded RNA relevant for R-loop formation has already been formed. This suggests that to be effective as a negative regulator, RNA I concentration has to be sufficiently high since the window of opportunity to catch the primer RNA is only the few seconds that it takes RNA polymerase to transcribe the 250-bp region. Mathematical modeling studies by Paulsson et al. have shown that the efficiency of control would be greatly improved if the 250-bp region provided multiple ($\leq$250) opportunities for hybrid formation. In the multistep mode of hybridization, the probability of replication initiation is an exponential function of RNA I concentration as opposed to a hyperbolic function in the single-step hybridization mode. The greater the number of steps, the more sensitive the replication initiation becomes to changes in RNA I concentration, and the narrower becomes the copy number distribution. For ColE1, the efficiency of control remains to be experimentally established but the available data are consistent with exponential inhibition mechanism.

In ColE1, the regulatory circuit senses copy number deviations by changes in RNA I concentration. In order to be an up to date reporter of plasmid concentration, the proportionality between RNA I and plasmid copy number needs to be maintained. If the RNA I concentration lags behind the plasmid copy number because of the delays in RNA I synthesis, the replication rate would remain high. Alternatively, elevated levels of RNA I concentration due to its finite half-life, would keep the replication rate artificially lower. Out-of-date RNA I concentrations will thus broaden plasmid copy number distribution. Thus for efficient control, constitutive synthesis of the copy number reporter (RNA I) is not enough; additionally, the regulator needs to be degraded rapidly. Counterintuitively, it is not the synthesis rate constant that determines how rapidly the RNAI concentration is increased but rather the half-life of RNAI that determines both its speed of increase and decrease.

The modeling studies also show that the rate of copy number adjustments is higher with increased transcription of RNA II. The steady state plasmid copy number (average copy number during logarithmic phase of host growth) also depends on the rate of primer RNA synthesis (the plasmid copy number increases if the primer promoter strength is increased). Thus, the three contributors to the efficiency of control are the mechanism of inhibition (hyperbolic vs exponential), high degradation rate of RNA I, and high initiation rate of RNA II. The latter (a positive function) plays as important a role in control as the negative regulator (RNA I). Plasmids are said to be under negative control because without a negative regulator, continued increase in plasmid copy number cannot be checked, which is lethal to the host (runaway replication). However, when subjected to rigorous mathematical modeling, the ColE1 example teaches us that both the steady state copy number and the rapidity with which adjustments are made to deviations from steady state copy number, are the result of both positive and negative activities. Adjusting the strength of either activity can alter the steady state and adjustment rate values.

Stable maintenance demands correction of fluctuations in either direction. In both up- and down-fluctuations, a high transcription rate of RNA II and repression by RNAI are required. In other words, although necessary, negative control is not sufficient for efficient replication control.

2.1.3 Control of Initiator Synthesis by Antisense RNA

In several plasmids, the copy number is determined by the availability of initiator proteins (i.e. initiators are rate-limiting for replication). In all these cases, the primary control element for initiator synthesis is an antisense RNA. The interaction of the antisense RNA with its target RNA inhibits initiator synthesis. Some better-studied examples are discussed below.

A straightforward mechanism is in the case of high copy streptococcal plasmid pMV158 (copy number about 20). The antisense RNA directly interacts with the translation initiation signals (ribosome binding site or RBS) of the RepB initiator mRNA and thereby downregulates its translation (Fig. 2a). The RepB–mRNA level itself is also regulated by a transcriptional repressor (CopG), a second control element for the plasmid initiator. The antisense RNA is stronger of the two control elements as it exerts stronger incompatibility (see Sect. 2.5)

In plasmid R1, a low copy enterobacterial plasmid, the initiator (RepA) translation is regulated indirectly by the antisense RNA (CopA) (Fig. 2b). RepA can be synthesized from either of the two transcripts, the longer of which also encodes the CopB protein. CopB represses the promoter for the shorter transcript (P2), making the constitutively made longer transcript practically the only source of RepA. The CopA antisense RNA binds to a homologous region of RepA mRNA. The binding inhibits translation of an adjacent small gene called *tap*. The translation of *tap* is required to break a stable RNA secondary structure that blocks the RBS of *repA*. The translation of *repA* is therefore coupled to that of *tapA*. Thus, CopA regulates *repA* translation indirectly through TapA.

When R1 copy number becomes exceptionally low, CopB level also drops releasing repression of P2, the stronger of the two promoters that can synthesize RepA. The resulting stimulation of initiator synthesis quickly increases R1 copy number until CopB repression is reestablished and the copy number returns to the steady state level. The increased P2 activity also increases RepA synthesis by interfering with *copA* expression (convergent

Fig. 2 Mechanisms for limiting replication initiator proteins for controlling plasmid copy number. The primary control element in these plasmids is also an antisense RNA. Hybridization of antisense RNA with complimentary sense RNA sequesters the translation initiation signal (RBS) of the initiator itself (a) or an activator of initiator translation (tap in b, RepB in d). The hybridization can also prevent initiator transcription by prematurely terminating the initiator mRNA (c). All these plasmids also have a protein that negatively regulates initiator transcription (a–c) or initiator translation (d). The regulators sometimes function in more than one way. For example, when CopB level is down, derepression of the P2 promoter of R1 not only increases *repA* mRNA but also helps in initiator synthesis by interfering with (convergent) antisense transcription. In pMU720, the antisense RNA (RNA I) prevents simultaneously translational activation of *repA* through RepB and pseudoknot formation. The complementary sequences for pseudoknot formation are shown in thick gray lines with an empty circle. Other symbols are as in Fig.1.

transcription) to some extent. Thus, CopB is there primarily to boost copy number if it becomes abnormally low and thereby avoid plasmid loss, but normally CopA suffices to adjust for fluctuations in copy number during steady state growth.

In IP501, a streptococcal plasmid whose genetic organization is similar to that of R1/pMV158 involving two negative regulators, the antisense RNA limits initiators by transcriptional rather than translational control. CopR represses the pII

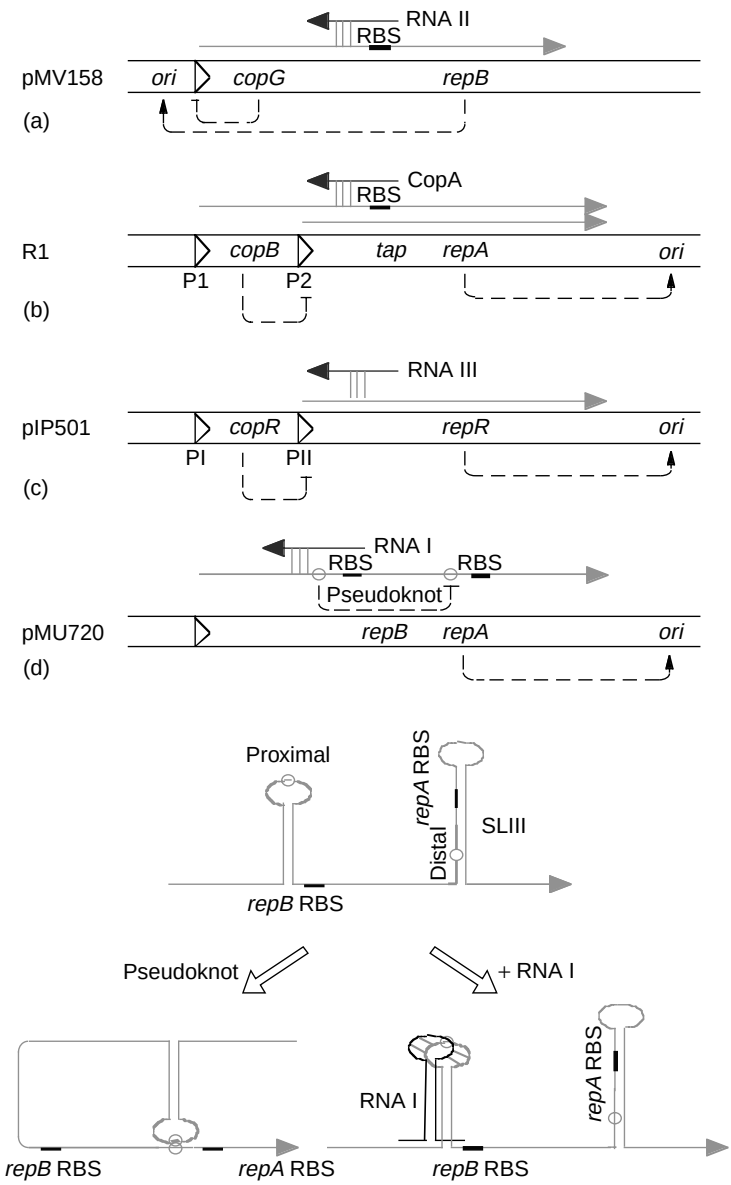

promoter, the only promoter for the initiator, RepR (Fig. 2c). The basal level of pII activity is high enough to maintain the plasmid. During transcription elongation, the *repR*–mRNA can adopt two mutually exclusive conformations depending on whether the antisense RNA is present or not. The binding of the antisense RNA leads to the formation of a transcriptional terminator (attenuator) that aborts transcription before the full-length *repR*–mRNA is made. Transcription attenuation is well known in many biosynthetic operons but the modulating translation elongation in the leader region of the message generates the transcription terminator.

In a group of plasmids (pMU270/pCol-IB-P9) unrelated to R1, the replication control region is similarly organized but the *rep* translation additionally requires activation by formation of a RNA pseudoknot (a tertiary motif in which nucleotides from the loop region of a stem-loop pair with a single-stranded distal region). In pMU270, the *repA* translation is coupled to that of a leader peptide (RepB) as in plasmid R1 (Fig. 2d). The translation of *repB* disrupts a stable stem-loop structure (SLIII) that otherwise sequesters the distal pseudoknot sequences, and the RBS and translation start codon of *repA*. Unlike R1, in these plasmids, the translation coupling is not enough: *repA* translation further requires activation by the formation of a pseudoknot immediately upstream of the *repA* RBS. A small antisense RNA, RNA I, controls replication by inhibiting translation of repB and thereby translation coupling, and by preempting pseudoknot formation by sequestering the proximal pseudoknot bases.

In pT181 family of plasmids also, the antisense RNA controls initiator (RepC) availability by transcriptional attenuation (Fig. 3). An additional but completely novel mechanism operates to limit RepC availability further. At the termination of one round of replication, the initiator is inactivated, preventing its recycling. The plasmid pT181serves as a classic plasmid example in which replication starts by the rolling circle mechanism i.e. by nicking the plus strand of the double-strand origin (called DSO, Fig. 3) by one of the subunits of RepC dimer. The inactivation is associated with a second nicking at DSO at the termination of replication. For reasons not clear yet, replication continues beyond one round by about 10 nt and then stops. This regenerates the DSO, which is cleaved by the unused subunit of the RepC dimer. This initiates a concerted strand transfer cascade, generating a single-stranded (SS) monomer corresponding to the displaced plus strand, and release of the initiator protein as an inactive heterodimer with one subunit of RepC dimer covalently linked to an oligonucleotide corresponding to about 10 nucleotides immediately 3′ to the nick site. The heterodimer RepC/C* is inactive in nicking and in initiation of replication. The replication cycle therefore involves two nickings, one at initiation and one at termination, and two closings, one for the new and one for the old leading strand, justifying why RepC needs to be a dimer. In other plasmids (pC194, pUB110), although the nicking-closing activities reside on a monomer, it has two separate active sites.

Replication kinetics in R1 and pT181 The kinetics of copy number control has been experimentally analyzed in plasmid R1. The rate-limiting step, also the case in ColE1, is the initial complex formation between the sense and antisense RNAs. The antisense RNA is unstable

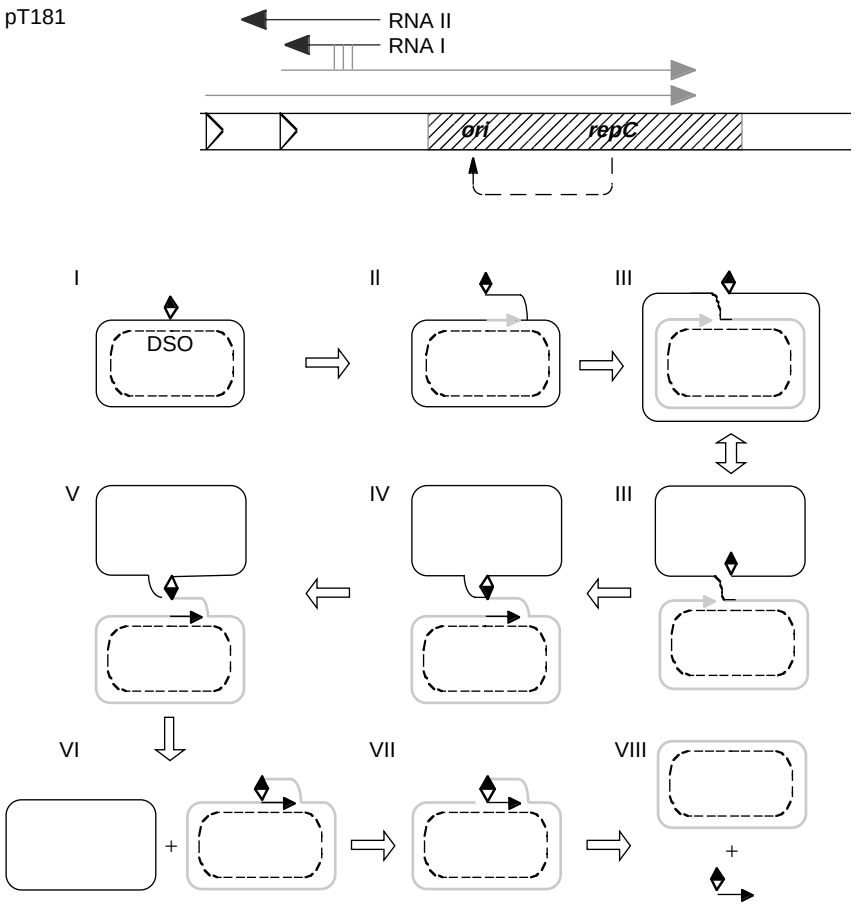

Fig. 3 Steps in pT181 replication with inactivation of initiator protein RepC at the termination of replication. The replicon is shown at the top. One subunit of a RepC dimer (black and white diamond) makes a nick at the origin (DSO) on the plus strand (I) and gets covalently attached to the 5′ end. The free 3′-OH end is then extended by new synthesis (gray line with arrowhead showing the growing end, II). At the end of one round of synthesis (III), the replication continues beyond the nick site a short distance (bold arrow, IV). The second subunit of the dimer initiates a concerted strand-transfer reaction by nicking the newly synthesized origin (V). The newly generated 3′OH end then attacks the old RepC–DNA bond releasing RepC and restoring the continuity of the old plus strand, which is then released as a single-stranded circle (VI). The free RepC subunit (white triangle of the diamond) then attacks the nick site generated on the newly synthesized strand. Finally, the 3′OH end generated by the nicking (VII) attacks the DNA–RepC bond generated in step V. The end products of this reaction are a double-stranded plasmid molecule with one newly synthesized strand, and a RepC/C* heterodimer with the covalently bound short oligonucleotide (about 12mer) to one of the subunits (VIII). The heterodimer is inactive as an initiator.

($t_{1/2} = 1–2$ min), and therefore its concentration can be taken to sense plasmid concentration. As in ColE1, the key parameter in determining the initiation rate is the concentration of antisense RNA but the inhibition kinetics is hyperbolic (e.g. replication initiation frequency decreases twofold if the copy number increases twofold). This will lead to a gradual restoration of copy number to the steady state value. The relatively slower rate of correction (compared to the exponential inhibition mechanism proposed for ColE1) apparently has not been a problem because of other maintenance functions, particularly, the CopB-mediated boosting of initiation rates when the copy number drops abnormally.

Replication kinetics studies in the case of a multicopy plasmid, pT181, have been particularly telling on how the copy number approaches the steady state value from low copy situations (termed repopulation). The lowering of copy number was achieved using a *repC*(ts) mutant plasmid at the nonpermissive temperature. Upon return to the permissive temperature, copy number increased rapidly and went beyond the steady state value before replication was inhibited. The copy number then returned to the steady state value gradually. The repopulation kinetics was similar when a plasmid copy was introduced to a plasmid-free cell. In the low copy state, when the inhibitor concentration was correspondingly low, replication rate was maximal that brought the copy number up to and beyond the steady state value rapidly. The overshoot in copy number is expected because inhibitor accumulation lags behind copy number increase. However, when sufficient inhibitor accumulates, negative control becomes the overriding force that prevents continued increase in plasmid copy number.

Although the instability of antisense RNA has been argued to be important for rapid copy number correction, an exception has been found in plasmid pIP501 (Fig. 2c). The antisense RNA in pIP501 is quite stable ($t_{1/2}$ = about 30 min as opposed to about 1–2 min in other cases) but the copy number correction is apparently not sluggish enough to cause significant plasmid instability. The Cop repressor of the plasmid is made constitutively (like R1). A reduction in copy number may therefore lead to derepression of the Rep promoter and increased initiator synthesis that would help bring the copy number back to the steady state level. The binding of Cop repressor to its operator in pIP501 is weaker compared to the situation in R1, and this may allow easier depression of initiator synthesis and prevent undershoot of copy number that is expected from a stable (''out-of-date'') reporter of plasmid copy number. Cop of pIP501 is also stable ($t_{1/2} = 40$ min, comparable to the generation time). With both the negative regulators being stable, this plasmid should have a relatively broad copy number distribution in a population of cells.

2.1.4 Control of Initiator Synthesis by Iterons

Antisense RNA is not the only negative regulator known to control plasmid copy number. There is a separate group of plasmids (prominent members include F, RK2, R6K, P1, pSC101, and pSL10) in which the negative regulators are short (20 bp or so) repeated sequences, called *iterons*. The hallmark of this group of plasmids is the presence of five or so iterons in the origin of replication. Iterons are essential for initiation and therefore they serve dual function of both effecting

initiation and controlling negatively the initiation frequency. As the iteron concentration increases with increases in plasmid copy number, the negative regulatory activity of the iterons becomes the overriding feature that apparently turns off replication altogether when the copy number becomes excessive.

The known activities of iterons are that they bind plasmid-encoded replication initiator protein Rep, and they pair via the bound Rep. The Rep protein concentration is usually controlled by transcriptional autoregulation and posttranslationally by the formation of inactive dimers. The *rep* promoter region is either found buried within the iterons in the origin or has overlapping inverted half-iterons (dyads). Rep binds to iterons as monomer and to the dyad as dimers and shuts off the promoter activity efficiently. Thus, Rep autoregulates its own synthesis. Although the monomers serve as initiators, the protein is found largely as dimer, which are inactive in iteron binding and hence cannot serve as initiators. The transcriptional autorepression and dimerization are the two known mechanisms that can limit active initiator availability. In the presence of some of the molecular chaperones (DnaJ, DnaK, GrpE, ClpA, and ClpB), the inactive Rep is converted to the active monomeric form. The chaperones thus contribute positively to the initiator pool. An alternate mechanism of activation is interaction with iteron DNA itself (DNA induces conformation change in Rep that dissociates dimers into active form). The primary mechanism to prevent plasmid overreplication is currently believed to be pairing of iterons by bound Rep (dubbed ''handcuffing''). These multitudes of controls involving the initiator apparently maintain the plasmid copy number within narrow limits. Rep proteins have domains in common with replication initiators encoded by organisms in all three kingdoms of life, making their detailed study into one of general importance.

The current view of the replication cycle based primarily on results of P1 is as follows (Fig. 4). The replication initiates when origin iterons are saturated with Rep. The binding is most likely cooperative, best shown in the case of R6K. The act of replication elongation frees the *rep* promoter of bound Rep and provides a window of opportunity for maximal *rep* transcription (''replication-induced transcription''). The newly formed Rep is inactive and forms dimer. Thus, although Rep synthesis maximizes immediately after an initiation event, this does not encourage another round of replication right away. If this were the case, runaway replication would follow. Thus, Rep inactivation by dimerization appears obligatory to prevent uncontrolled overreplication. The preexisting active Rep most likely gets distributed to daughter origins (titration) so that none can be saturated and this can aid to prevent premature initiation also. The origins can pair by the bound Rep and this is believed to cause steric hindrance to origin function. It is generally accepted that formation of a higher order nucleoprotein complex activates an origin, and it is reasonable to assume that handcuffing will be on the way to the formation of such a complex. As active Rep accumulates by the mediation of chaperones, the origins can saturate again. Our best guess is that intramolecular cooperative interactions among bound Rep molecules, Rep interactions with host factors (DnaA, DnaB, and DnaG), and the presence of architectural proteins like IHF or HU all cooperate to form a complex that reverses the handcuffing and makes initiation possible again.

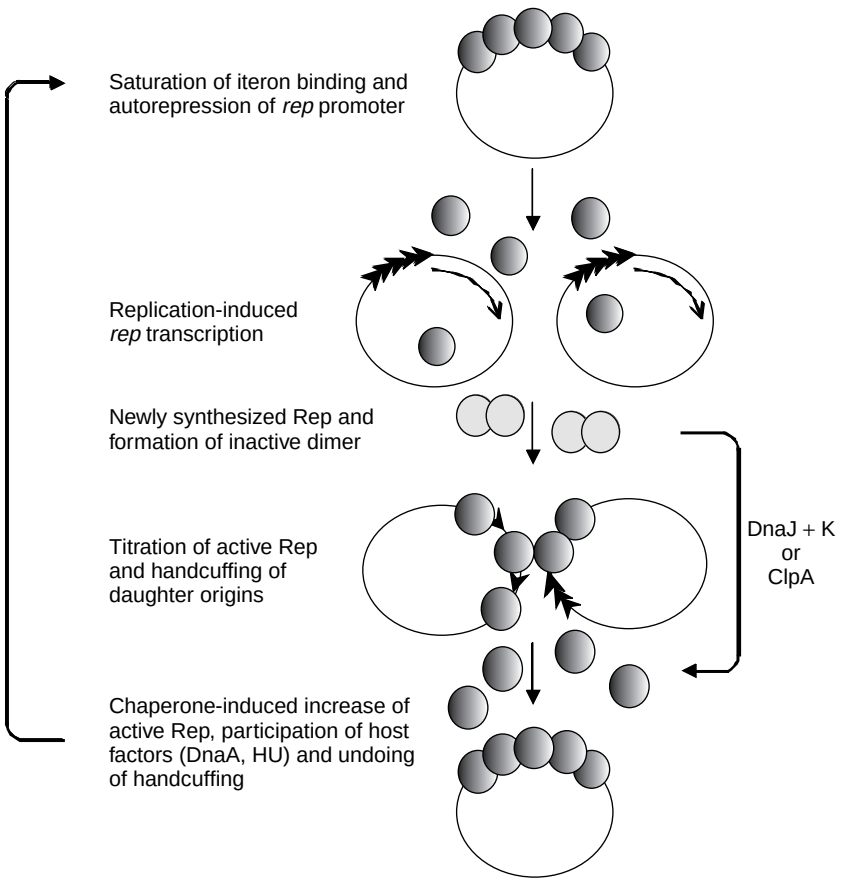

Fig. 4 A model on how Rep protein controls replication of an iteron-carrying plasmid, P1. Rep protein (gray sphere) binds to five repeated sites (iterons) in the origin that simultaneously achieves initiation as well as autorepression (the iterons also serve as operators of *rep* transcription). When the replication fork traverses the iterons (black arrow heads), the promoter is cleaned of bound Rep, promoting a burst of *rep* transcription (replication-induced transcription; bent arrows, *rep* transcription). However, the newly made Rep misfolds (light gray spheres) that encourages dimerization. Rep dimers are defective in iteron binding and require chaperones (DnaJ and DnaK or ClpA) for activity. The active Rep is titrated by the daughter origins that also allow pairing of origins by Rep–Rep interactions (handcuffing). Handcuffing is believed to be an impediment to initiation and must reverse to allow the next round of replication. Increase in Rep activity and participation of other host-encoded initiators apparently make the origin initiation-competent again and the cycle continues.

Iteron mediated control is very precise as it can maintain plasmids faithfully as low as one in a newborn cell. How the precision is achieved is not known but the presence of multiple iterons, and multiple modes of control on Rep synthesis and activity (cooperative binding, autoregulation, inactivation by dimer formation, activation by chaperones, titration and handcuffing, and interactions with multiple host

factors) are there to avoid fluctuations that otherwise would be inherent in a low frequency event.

2.2 Partition

Other than precise duplication of the genome, chromosome maintenance in all organisms involves equal distribution of DNA copies (sister chromatids) to daughter cells (mitosis). A mitosis-like partition system, called *par*, also functions for maintenance of low copy number plasmids such as F, P1, and R1. Plasmids differ from chromosomes in that plasmid copies are identical to each other whereas in mitosis only sister chromatids are identical. Unlike mitosis, there is no obligatory need for plasmids to be equipartitioned. In principle, plasmid maintenance can be assured even if each daughter receives just one plasmid copy. An efficient replication control system can then bring the copy number to the steady state value rapidly. Experimental evidence in this regard is rather meager and thus in support of the equipartition model.

Plasmids like ColE1 do not encode a *par* system. Nonetheless, the plasmids are maintained stably because of their high copy number. Since the probability that a plasmid will not enter a particular daughter is 0.5, and if there are $2N$ plasmid copies in a dividing cell (N being the copy number in a new born cell), the probability that none ends up in that daughter is $(0.5)^{2N}$, which is the expected loss frequency per cell per generation. For a plasmid like pBR322 in which $2N$ can be 40, the loss frequency is expected to be about 10^{-12}, whereas for plasmids like P1 and F, where $2N$ can be 2, the loss frequency can be as high as one out of four cells. These loss rates make it obvious why high copy number plasmids do not have *par* and vice versa.

2.2.1 Regulation of *par*

The *par* genes exist as a cassette and are remarkably uniform in terms of their genetic organization. In all cases, they encode two trans-acting proteins, usually called ParA and ParB, and one cis-acting region called *parS* (*S* for site), the centromere analog (Fig. 5). The genes for ParA and ParB are in an operon that is autoregulated. ParA is an ATPase and ParB binds *parS*. However, the ATPase can be of two types and on this criterion, plasmids have been classified as type I with Walker-type ATPase (as in P1 and F), and type II with actin-type ATPase (as in R1). The type I plasmids have been further separated into two types depending upon whether the operator/promoter region for autoregulation is a separate locus (type Ia) or coincides with the *parS* site (type Ib). The autoregulatory site of type II plasmids also coincides with *parS* as in type Ib. The regulation of the *par* operon is known in detail in several plasmids. It is particularly intriguing in type Ia plasmids in which ParA serves as the repressor primarily when in the ADP-bound form in those plasmids in which the two forms have been distinguished. ParB and *parS* serve as corepressors and significantly improves the repressive effects of ParA.

The tight regulation of *par* suggests that the levels of the components are critical for function. In fact, overproduction of any of the three components in trans destabilizes plasmids, that is, causes partition-mediated incompatibility (see Sect. 2.5). The mechanisms of destabilization are largely unknown but likely to vary depending upon the destabilizing (incompatibility exerting) component. The incompatibility due to *parS* could be either

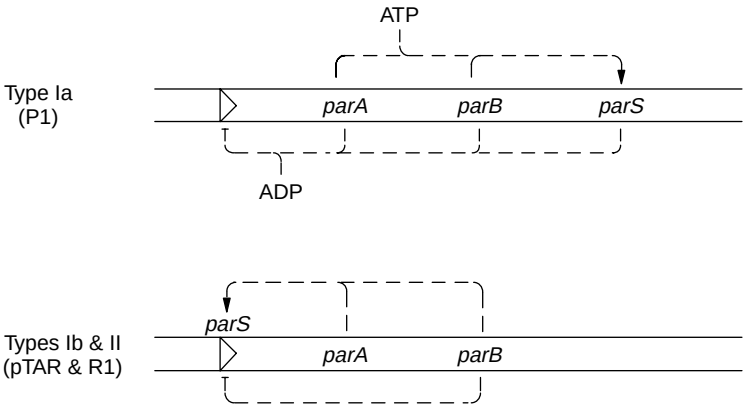

Fig. 5 Genetic organization of *par* operons. The operons consist of genes for two trans-acting proteins, ParA and ParB, and a cis-acting site, *parS*, which serves the function analogous to a centromere. The operons have been classified according to whether the ATPase domain of ParA is Walkerlike (type I) or actinlike (type II), and whether the promoter and *parS* are separate (type Ia) or overlapping (type Ib and II). The operons are autoregulated in all cases, sometimes, involving all three components as in plasmid P1. The ATP form of ParA is involved in partition whereas the ADP form is used in autorepression.

due to titration of an essential plasmid or host component for partition. A more attractive possibility is that the cloned *parS* pairs with a target (incompatibility sensitive) plasmid thereby precludes its pairing with its sister, believed to be an essential early step in partition. In P1 and F, the mechanism of destabilization by excess ParB is likely to be spreading of ParB on DNA and silencing of genes on its wake (see Sect. 2.2.3). ParA could be exerting incompatibility by breaking plasmid pairs.

2.2.2 Partition Dynamics

Our understanding of the dynamics of partition process owes greatly to recent technological developments that allow mapping of proteins and DNA sites within cells by fluorescence microscopy. Both the proteins and the sites can be directly tagged with green fluorescence protein (GFP) or its derivatives, and visualized in live cells in real time. Alternatively, mapping can also be done using immunofluorescence microscopy that does not require tagging *in vivo*, but this can only be done in dead cells. On the basis of their results on R1, Gerdes and colleagues have provided a model for plasmid partition for the type II (actin-type) *par* cassette. Following plasmid replication in the midcell, the sister plasmid copies pair via ParR (ParB analog) bound to *parS*. The pairing process also requires ParM (ParA analog) bound to ATP. The pairing apparently provides a signal to ParM to start forming a linear polymer that pushes the sister plasmids away from each other towards the cell pole. On reaching the polar location, the sisters stay anchored till the cell division is completed. The situation is analogous to mitosis in that the plasmids first pair at the centromere and are then pulled or pushed apart to opposite halves of the cell. Several recent findings, not necessarily from R1 alone, fit with the above model. The replication factory is

stationed in the midcell and plasmids are incorporated in the midcell for replication. In smaller cells, plasmids are found in midcell and in longer cells either in cell quarters (future midcells as in P1 and F) or in more polar positions (as in R1). Simultaneous mapping of several plasmids has indicated that plasmids have discreet nonoverlapping locations. The ParM (ParA analog of R1) forms filaments that can span the entire cell length. This requires an intact ATPase domain of ParA and other two *par* components, ParR and *parS*. The genetic requirements for filament formation and partition are thus identical, implying that the former is important for the latter.

The filament formation has explained how plasmids can move apart at speeds that are an order of magnitude faster than the cell elongation rate. Previously, attachment to cell membrane and movement in sync with membrane growth was the proposed mechanism of moving daughter chromosomes/plasmids apart. This mode of movement is too slow to account for the observed speed of plasmid movement. Hitchhiking on the chromosome instead of membrane was another possibility, but this is unlikely since plasmids can segregate to chromosome-less (anucleate) cells. ParA filaments may help in an alternate way by providing tracks on which plasmids can be transported, as is the situation in mitosis where molecular motors carry chromosomes along the microtubules tracks. It is not known whether the ATPases of type I plasmids also form filaments or whether they provide the energy of movement in some other way. Determination of the details of plasmid movement remains the most challenging aspect of understanding the partition problem.

It is also not known how the plasmids stay in their polar destination after they reach there, and how they are recruited for replication. To date, no host genes involved in plasmid partition have been identified. Some of the genes implicated in partition of bacterial chromosomes (e.g. *mukB, smc*), apparently, do not contribute to plasmid partition. This is not surprising considering the additional complexity of partitioning a much larger size element such as a chromosome. In fact, chromosomal *par* genes so far identified all seem to be involved in condensing the nucleoid. Compaction afforded by negative superhelicity may suffice for plasmids and additional mechanisms are called for only when the chromosome size gets unwieldy.

2.2.3 Unexpected Aspects

Binding of SopB of F and its analog ParB of P1 to their centromeric sites (*sopC/parS*) can silence surrounding gene expression over several kilobases. The silenced region has also been shown to be inaccessible to DNA gyrase and DNA adenine methylase. Such long-range effects on DNA for sequence-specific DNA-binding proteins are surprising since transcriptional repressors only control expression in an operon-specific way. ParB type proteins also serve as classical repressors as they participate in autoregulation of *par* operons. ParB homologs are also known to act as transcriptional factors sometimes without playing any role in partition (e.g. VirB protein of *Shigella* virulence plasmid). But only binding to the centromeric sites seem to mediate silencing so far.

The silencing is likely to be due to spreading of ParB bidirectionally on DNA from the *parS* site by nonspecific binding. DNA covered with ParB apparently precludes other DNA–protein interactions.

An alternative model is that ParB concentrates on subcellular compartments forming a patch that attracts *parS* and sequesters adjoining DNA by nonspecific binding. The nonspecific binding is expected to follow because of high local concentration of ParB in the patch (homing site). The relevance of silencing to partition is not obvious, but the latter model is easier to appreciate because it can bring a *parS*-containing plasmid to subcellular location and thereby help in their segregation. In eukaryotes, centromeres are associated with heterochromatin and are transcriptionally silent. In fission, yeast heterochromatin proteins serve a critical role in chromosome segregation as they are required to load components of mitosis such as cohesin. In budding yeast, unstable plasmids lacking a centromere can be stabilized efficiently by transcriptionally silent mating-type locus or telomeric repeats. These spindle-independent segregation mechanisms require proteins that are also required for silencing. One such protein, Sir4p, apparently tethers plasmid DNA to an immobile nuclear component that divides between progeny at mitosis. These links of gene silencing with chromosome segregation may indicate that silencing is important in partition of bacterial plasmids.

ParA (of plasmid pB171) has shown another novel phenotype – it oscillates between the two cell extremes with a periodicity about one fifth of cell generation time. The oscillation requires the ATPase activity and other two components of *par*, ParB, and *parS*. Two other proteins related to ParA by virtue of having Walker-type ATPase motif, Soj of *Bacillus subtilis*, and MinD of *E. coli* also can oscillate requiring other related components. MinD oscillation helps to define the midcell (cell division plane) but how the oscillation could help plasmid segregation is unclear. The requirements for ParA movement and plasmid partition being identical suggest that the former is important for the latter.

Fluorescence microscopy has revealed that plasmids stay in clusters but there is no clue yet as to how they do so. Clustering should be problematic for stability of plasmids lacking *par*, as the number of segregation units will be greatly reduced. In spite of clustering, the stability measurements indicate as if the plasmids diffuse freely, independent of each other. A conceptual solution to this paradox has been proposed. The idea is that replication can decluster plasmids and make them segregate as free copies that follow replication. The Austin group has presented a dynamic view of clustering and declustering for plasmid P1 and this could lead to a radically different view of partition. Plasmid copies are found first to coalesce at midcell. At the time of cell division, the cluster splits and plasmid copies are ejected away from each other. This way, one or more plasmid copies are segregated to each half of the dividing cell. The placement of the plasmid cluster at the cell division plane makes the completion of cell division contingent upon clearing the plane of plasmids that may otherwise cause steric hindrance. The model is attractive for its inherent checkpoint control feature.

2.3 Postsegregational Killing

Postsegregational killing (PSK) systems are widespread among plasmids and also in chromosomes of bacteria and archea indicating their importance in the prokaryotic world. The PSK systems usually

consist of a two-gene (bicistronic) operon that stabilizes plasmids by a different mechanism from that of partition. One gene codes for a stable toxin and the second a relatively unstable antidote that neutralizes the effect of the toxin. As long as a plasmid is present in a cell, the regulation of the operon is such that the antidote dominates over the toxin and there is no undesirable effect on the cell. In case the plasmid is lost, the relatively stable toxin is retained longer than the antidote and causes growth arrest or even death of the plasmid-free cell. The toxin effect is restricted to plasmid-free cells only, and their selective loss increases the fraction of plasmid-carrying cells in the population. The plasmid therefore appears more stable (PSK does not reduce the frequency of appearance of plasmid-free cells). Once a cell has acquired a plasmid with the PSK system, it essentially cannot survive without it. For this reason, the PSK system has also been called a *plasmid addiction* system or, since these operons also exist in genetic elements other than plasmids, a *genetic addiction* system. The killing of plasmid-free cells is quite analogous to programmed cell death or apoptosis (a Greek word meaning ''falling off'') in multicellular eukaryotes.

Note that the killing by the addiction system is from within. There are also plasmids like ColE1 that can increase plasmid stability by killing plasmid-free neighboring cells from without. ColE1 produces a bacteriocidal agent called *colicin*, which does not harm the plasmid-carrying cells because the plasmid also carries an immunity gene, but the secreted colicin can kill plasmid-free cells in the immediate environment, thereby conferring a competitive growth advantage to the plasmid-carrying cells.

2.3.1 Classification and Regulation of Addiction Systems

The antidote prevents toxin action either by direct protein–protein interactions as in proteic systems, or by preventing toxin translation as in antisense-regulated systems, or by modifying the target of the toxin as in restriction–modification systems. A classic example of a proteic system is the Ccd (control of cell death) system encoded in the *ccdAB* operon of plasmid F. CcdB is the toxin and its expression in the absence of the antidote protein, CcdA, causes induction of the SOS pathway, which entails cell filamentation and ultimately cell death. In plasmid-free bacterial segregants, CcdA is degraded by Lon protease but CcdB persists, converting DNA gyrase to an endonuclease. DNA gyrase is an enzyme that is essential for DNA supercoiling and cell viability. CcdB traps an intermediate of the supercoiling reaction in which the gyrase is covalently bound to cleaved strands of DNA. The double-strand break can cause cell filamentation by inducing SOS response, which is known to be a biological response to DNA damaging agents. However, it is not known whether the death is due to poisoned gyrase that only cleaves DNA or due to loss of DNA supercoiling or both. CcdB can also stabilize the cleaved DNA–gyrase complex and these complexes can serve as roadblocks for elongating RNA polymerases. This can be an additional mechanism contributing to cell death. Interestingly, the antidote CcdA can reverse the cleaved complex and restore DNA integrity, as well as remove the roadblock to transcription, by sequestering CcdB as a tight CcdA–CcdB complex that also frees gyrase from DNA. Most likely, in plasmid-carrying cells, there are not sufficient free CcdB to cause any harm to the cell but the capability to reverse damage ensures survival of

plasmid-carrying cells, should the amount of CcdB transiently rise to toxic levels.

Expression of the addiction operons is regulated efficiently. This is expected since the relative levels of the two gene products are critical for cell survival. A common mechanism in proteic systems is autoregulation in which the antidote acts as a transcriptional repressor by binding to the operator–promoter region of the operon (Fig. 6). In many cases, the toxin acts as a corepressor and greatly improves repression (e.g. in *ccd* of F and *phd–doc* of plasmid P1). Moreover, the ratio of the two proteins is important for DNA binding. In the case of *ccd,* the binding is efficient with a CcdA:CcdB ratio of one. Excess of CcdA still maintains repression but derepression occurs when CcdB is in excess of CcdA. This way, the harmful effect of CcdB can be prevented in plasmid-carrying cells but there will be enough of it to cause harm in plasmid-free cells. The autoregulation is less complex in some other systems that have a three-protein operon and the extra protein solely serves as the autoregulator of transcription (e.g. Omega in SegB region of plasmid pSM19035).

The paradigm for antisense-regulated control is the *hok/sok* system of plasmid R1. Here the toxin, Hok (host killing), is also a protein but the antidote, Sok (suppression of killing), is an unstable RNA that does not directly interact with the toxin as in the proteic systems. Instead, it interacts with the *hok* message and prevents its translation. Binding of the two RNAs allows RNase III to degrade the *hok* message and thereby prevent its killing activity. In plasmid-free cells, the unstable Sok RNA is quickly depleted that allows translation of the *hok* message. The product, Hok, kills cells by damaging their membrane.

The regulation of the *hok/sok* system is entirely posttranscriptional but more intricate than that has been discussed in connection with replication control (Fig. 2). The activity of *hok* mRNA depends upon a complicated folding pathway (Fig. 6). During transcription, the 5′ end is locked into a metastable hairpin structure that sequesters a sequence required for activation of *mok* translation. Translation of Hok is coupled to that of Mok, and therefore the metastable structure prevents Hok

Fig. 6 Regulation of plasmid addiction operons. (a) In the *ccd* operon of F plasmid the antidote–toxin pair, CcdA and CcdB, also regulate the transcription of the operon. CcdA is the primary repressor. CcdB joins CcdA and helps in the repression thereby serving as a corepressor. The operon is therefore autoregulated. (b) The SegB region contains a three-gene operon in which the first gene is a dedicated autorepressor that indirectly controls the expression of the antidote–toxin pair. (c) The *hok/sok* system of plasmid R1 has three genes *mok, sok,* and *hok.* The *sok* antidote is an antisense RNA that inhibits translation of *mok* protein, the toxin. Translation of *hok* is coupled to *mok*, and therefore, the antidote RNA controls the toxin translation indirectly through *mok.* The toxin translation is actually under more complex control. Initially, the *hok* message is not translated because of the formation of a metastable structure at the beginning of the message that includes a sequence (tac) required to activate *mok* translation. The metastable structure stabilizes ucb/*mok*RBS hairpin and thereby makes *mok*RBS unavailable for translation. After a full-length transcription, a refolding takes place destabilizing the metastable hairpin, but the ucb/*mok*RBS hairpin is still maintained as the tac sequence forms a new structure that includes a hybrid with the 3′ end of the message. The 3′ end is then slowly processed that triggers refolding of the message that allows tac to hybridize with ucb, freeing the *mok*RBS for translation initiation. The refolded structure also allows hybridization with Sok RNA.

translation also. In the full-length *hok* mRNA, the metastable structure is lost but the message is refolded into a structure that is still inactive and requires processing from the 3′ end before it can be translated. However, the processed message is also the target for Sok RNA, and therefore translation is possible in plasmid-free cells. The toxin production thus requires breakdown of RNA secondary structures, refolding and processing, then translational coupling, and finally escape from Sok hybridization and RNase III attack. Avoidance of cell killing is particularly important when an addiction module is trying to establish itself in a new host. One of the

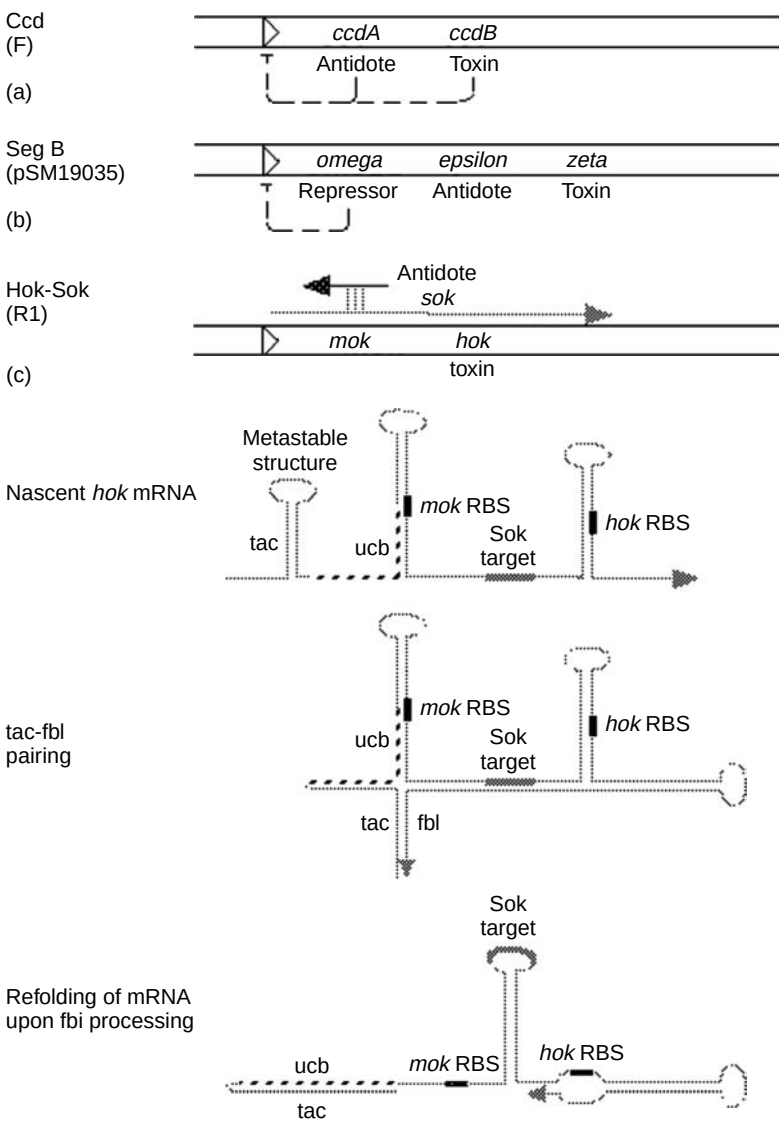

reasons for employing the delay tactics in Hok toxin synthesis may be for establishment without killing the host. These tactics will also discourage accidental toxin production in established cells.

In the restriction–modification system, the toxin is the restriction enzyme that can cleave specific sequences on DNA, and the antidote is the methylase that protects the sequences by modifying them by attaching methyl groups. The antidote thus does not directly interact with the toxin nor prevents its synthesis, but protects the target of the toxin. In the presence of the restriction–modification gene pair, nearly all, if not all, of the restriction sequences are modified. When the gene pair is lost, mere dilution of the enzymes will suffice to kill the cell and no differential stability of the toxin–antidote pair is required. This is because of the differential demand on the two enzymes. While survival would require modification of nearly all restriction sites and therefore a heavy demand on the methylase, restriction of even a single site can be potentially lethal. Although restriction–modification systems are known for almost fifty years, that they are powerful addiction systems has been realized relatively recently.

The restriction–modification genes are regulated at the level of transcription and some modules have a third gene solely for regulatory purposes as has been discussed for other types of addiction systems. When *E. coli* K12 *hsdK* genes are transferred to a new host, modification activity is observed right away but restriction activity appears about 15 min later. Such staging of expression would help establish the system in an unmodified host.

2.3.2 Addiction Systems in Chromosomes

All three types of addiction modules have been found in chromosomes, in bacteria, and, sometimes, in multiple copies in the same chromosome. Since addiction modules cannot increase the stability of the chromosomes as they do for plasmids, and since they are potentially toxic, the question arises as to why they are so prevalent in the chromosome. Their accumulation can be understood because once in, be it in the plasmid or chromosome, they would be hard to get rid of. Even then potentially harmful elements are expected to be eliminated unless they confer some benefit to the host. The restriction–modification genes can destroy invading foreign DNA thereby help maintain the genetic identity of a species. The chromosomal *mazEF* operon is induced by nutritional stress. The operon is autoregulated by the combined action of MazE and MazF antidote–toxin pair like the Ccd system (Fig. 6). Additionally, the expression of the operon is inhibited by high concentration of ppGpp, which is synthesized upon nutritional stress. Since MazE antidote is labile, its level falls rapidly that leaves MazF toxin to exert its toxicity. The dead cells not only do not compete for nutrients but they may also provide their resources to surviving siblings. Similarly in stresses like infection, precocious death of the infected population prevents the spread of the infectious agent to the uninfected population and thereby saves them. Thus, addiction modules can be viewed as stability elements for the chromosome as well. In any event, death of a subpopulation to save the majority (altruism) appears to be ubiquitous in nature.

2.4 Multimer Resolution

Formation of plasmid multimers is rather common since it is caused by homologous recombination, a highly conserved

function in bacteria. Multimerization is undesirable for plasmid maintenance as it reduces the number of segregation units. This is because plasmid replication control systems generally count origins per cell rather than number of plasmid copies. To avoid plasmid loss due to multimerization and consequent lowering of copy number, plasmids encode systems that reduce multimers to monomers by (intramolecular) site-specific recombination. The effect of multimerization must be significant since multimer resolution (MR) systems are widespread among plasmids from such as ColE1, P1, F and RK2 Gram-negative and pAM1, pIP50 and pSM19035 from Gram-positive bacteria. The presence of MR systems is easily rationalized for low copy number plasmids as further lowering of their segregation units would have a more significant effect on the production of plasmid-free cells.

Detailed study in ColE1 has revealed the requirement of the MR system even for a high copy number plasmid. Even if dimers (and higher multimers) form at a low frequency, they can become overrepresented in a fraction of the cell population. Because dimers have twice the origin number, they enjoy a higher chance of being replicated as plasmids are chosen at random for replication. Once replicated, the chance of one of the progeny dimers being selected remains high and this leads to an ever-increasing bias for dimer replication that can eventually spawn dimer-only cells (*plasmid apartheid*). Since plasmid copy number would be lowest in dimer-only cells, they would be the primary progenitor of plasmid-free cells. Plasmid apartheid can increase the probability of plasmid loss by 30 000-fold. This high loss rate due to plasmid dimerization and segregation to dimer-only cells has been called *dimer catastrophe* by Summers. The role of plasmid apartheid became evident when the proportion of dimers was correlated with the plasmid loss rate by computer modeling. In pBR322, only 4% of the plasmid DNA is found as dimers. If the dimers were assumed to be distributed randomly in the cell population, their contribution to lowering of plasmid copy number was found to be too low to account for the observed plasmid loss rate. For satisfactory accounting, about half of the dimers needed to be present in dimer-only cells and the rest as a mixture with monomers. Thus, for plasmid stability, in addition to the mean dimer copy number, it is important how the dimers are distributed in the population.

The MR system in the case of ColE1 requires two host recombinases, XerC and XerD, and a site in the plasmid, *cer*, where catalysis occurs. The system can only convert dimer to monomer and not the reverse, that is, works only intramolecularly. For this directionality of recombination, two other accessory host factors, ArgR and PepA, are required. Furthermore, the sites need to be directly repeated and the dimer needs to be supercoiled. These topological constraints make perfect sense if the MR systems have evolved to resolve dimers. A detailed molecular model has been proposed where the two *cer* sites bound by XerC, XerD, and PepA synapse via an ArgR bridge, which is too weak unless reinforced by supercoiling. This explains why the sites need to be in cis. XerC and XerD also participate in resolving chromosomal dimers that arise in about 15% of the cells in each generation. The chromosomal site is called *dif*, and instead of ArgR and PepA, the resolution of chromosomal dimers require a cell division protein FtsK. Unlike the plasmid case in which preventing cell division buys more time for recombination, the connection to cell division allows

chromosomal dimer resolution only at the time of cell division preempting the scope for any back reaction.

The MR system alone, however, is not necessarily efficient enough to eliminate the catastrophic plasmid loss due to dimerization. In ColE1, the monomerization rate is too slow compared to the rate with which dimers build up by replication. Fortunately, a regulator of cell division, *rcd*, is made only from plasmid dimers (multimers). The promoter for *rcd* maps in *cer*, and the synaptic complex formation on dimers allows proper geometry of the promoter to be active. The cell division is delayed as dimers build up. The extension of cell generation time allows more opportunities for the MR system for monomerization and this helps to minimize plasmid loss. The *rcd* in fact is a short RNA but how it interferes with cell division is not known. Interestingly, it downregulates expression of chromosomal genes but not the plasmid genes. Thus by overexpressing *rcd*, cells can be directed to make pure cloned gene products.

It remains to be seen if systems like Rcd are there to help other MR systems. In iteron-carrying plasmids, dimers are discriminated against for replication. The presence of two origins in cis allows more efficient pairing (handcuffing) of the origins that is believed to inactivate both the origins. Thus, iteron-carrying plasmids seem to have been selected to prevent rather than cure plasmid apartheid.

2.5 Incompatibility

Wild strains of bacteria often contain multiple plasmids that can be stably inherited for many generations. In *E. coli*, it has been possible to maintain eight different plasmids in the same strain at their natural copy number. In contrast, cells doubly transformed with genetically related plasmids usually fail to maintain them during growth without selection for both. These plasmids are called *incompatible* and those that can be stably maintained are called *compatible*. As pointed out by Novick and Hoppensteadt, to be incompatible the plasmids in question (1) must form a common pool on the basis of some shared determinants (due to genetic relatedness), and (2) are withdrawn at random for replication and/or partition. That plasmids are chosen at random for replication has been verified: in a cell cycle, the same plasmid can replicate more than once whereas others none at all. Plasmids with compatible replicons can also show partition-mediated incompatibility if their *par* systems are similar enough and important for stable inheritance.

Incompatibility (*inc*) due to replication can be viewed as a phenotype of plasmid negative control functions. If a plasmid pair is equally sensitive to each other's control systems, their total copy number would be characteristic of the individual plasmids but the ratio of the two plasmid kinds will vary from cell to cell because they are chosen at random for replication. Taken to the limit, this disproportion leads to loss of one or the other plasmid type, manifesting incompatibility. In compatible pairs, the control system of one does not recognize the other, and each is dedicated to maintain the copy number of the same kind. Since all plasmids control their copy number, they all exhibit incompatibility. This phenotype has been exploited in plasmid classification into separate incompatibility groups and this has proven useful in terms of understanding relatedness of control functions.

In practical terms, knowledge of incompatibility becomes necessary for designing experiments that call for maintaining multiple plasmids in the same cell. With the ease of DNA sequencing, plasmid relatedness is initially determined by direct sequence comparisons. However, point mutations in control regions can alter control specificity, making experiment-based incompatibility classification a more reliable approach.

Determinants of *inc* are conveniently identified by cloning and overexpression by their ability to destabilize the parent plasmid in trans. It should be noted that not all negative control elements exert incompatibility. Whereas all antisense RNAs are strong *inc* determinants, none of the accessory proteins like Rom of ColE1 and the Cop proteins of plasmids under antisense control are (see Sects. 2.1.2 and 2.1.3). The steady state concentration of the accessory proteins are probably saturating so that further increase of their concentration does not affect replication. It appears that, to be an *inc* determinant, the control element activity needs to be concentration sensitive near its steady state value. Gross overproduction can cause destabilization artifactually. This is another reminder that concentration sensitivity of the activity, a control element, is at the heart of measuring and correcting copy number deviations.

2.6 Cell cycle Control

Plasmids do not seem to have mechanisms to directly communicate with host cell cycle events to ensure that their replication and partition are complete before the cell division. Cross talks among the cell cycle events, chromosome replication, segregation, and cell division, are also not apparent for the host itself. For a growing culture of *E. coli*, the events occur in an orderly fashion with a periodicity of once per cell generation. However, it is clear that replication and segregation can happen independent of cell division since blockage to cell division does not prevent the other two events. It has been suggested by Boye and Nordström that the events are best viewed as independently regulated processes that run in parallel, meaning that initiation of one event does not wait for the completion of another as in biosynthetic pathways. In the latter, the product of a step serves as the substrate for the next one. If the first step is inhibited, the next one cannot start, unlike the situation in bacterial cell cycle events. How is the order of cell cycle events maintained? The suggestion is that some behind-the-scene growth-related factors regulate the cell cycle events independently and indirectly maintain their order. In the case of perturbation such as DNA damage, checkpoint controls are induced that helps to restore the order. For example, damage to DNA blocks replication fork movement that induces synthesis of an unstable cell division inhibitor. When the damage is repaired and replication elongation resumes, no new inhibitor is synthesized and cell division is no longer blocked. The communication channel in the normal course (without DNA damage) between DNA replication and cell division has not been identified. In *E. coli*, small signaling molecules cAMP and ppGpp, and in *Caulobacter crescentus* the CtrA protein, which control several events are best candidates for cell cycle regulators at the moment.

Plasmid replication does not depend on plasmid segregation. The *par* genes can

be deleted without affecting replication. This is best evident in integrative suppression experiments in which defects in host replication initiation are suppressed efficiently by integrating a plasmid replicon in the chromosome. This process remains unchanged whether or not *par* genes are present. On the contrary, it is believed that plasmid and chromosome segregation are facilitated by replication. We have discussed earlier the possibility that replication can unclump plasmids and thereby help in their segregation (see Sect. 2.2). Chromosomal replication has also been implicated in disposing daughter DNAs to opposite cell poles and thereby their segregation. In P1, cell division may trigger plasmid segregation (see Sect. 2.2.3).

Plasmid replication and segregation appear to be unlinked to their chromosomal counterparts. Plasmids have to adjust for copy number fluctuations constantly. A cell-cycle-specific unique timing of replication almost surely cannot apply for plasmids. Helmstetter and coworkers have studied replication of several plasmids, and none has shown cell-cycle-specific timing of initiation under conditions the chromosomal origin did. It should be noted that evidence for cell-cycle-specific replication, even if found, could be a side effect of replication control. For instance, if the plasmid such as P1 has a very low copy number and a replication control that senses a threshold concentration of initiators, then all cells will replicate their plasmids at the same time in the cell cycle, even if there is no explicit dependence on cell cycle events (the same idea proposed by Boye and Nordstrom). Genes such as *mukB* that help chromosome segregation by condensing the DNA do not affect plasmid segregation. Overall, the cell cycle connections remain tenuous and will remain so until specific regulators like CtrA are identified.

3 Horizontal Spread

Plasmids not only move vertically but also horizontally. The horizontal spread is easily rationalized for its potential for gene transfer across species and genera, thereby contributing to bacterial fitness and evolution. Some plasmids can have unusually broad host range (BHR) for extrachromosomal existence, and potentially even a broader range of recipient hosts in which they can be transferred by conjugation. Even if the plasmid cannot replicate in a new host, the DNA can be cleaved and recombined into the host chromosome. A BHR plasmid, RK2, has been established in forty species of Gram-negative bacteria. Plasmid F has been transferred from *E. coli* to yeast, RK2 from *E. coli* to cultured mammalian cells, and a segment (T-DNA) of Ti plasmid is transferred naturally from *Agrobacterium* to the nucleus of plant cells, demonstrating the capacity of horizontal gene transfer (HGT) between genera and kingdoms. The magnitude of HGT among microbes, as revealed by genome sequencing, has been startling. As many as 40 genes have been found to be shared exclusively by bacteria and human, that is, without any occurrence in invertebrates. Since the year 2000, four textbooks have appeared on HGT (also called LGT, L for lateral) that speaks for the current interest on this subject (Mobile DNA II, ASM Press; Horizontal Gene Transfer, Academic Press; Lateral DNA Transfer, CSHL Press; and Horizontal Gene Pool, Harwood Academic Publishers).

3.1 Conjugation

The primary mechanism for HGT is conjugation. The transfer of DNA requires more elaborate machinery than their maintenance functions discussed above and can involve even up to 50 plasmid-encoded genes, usually called *tra*. Unlike vertical movement where plasmids capitalize on host cell division machinery, HGT requires cell-to-cell contact (mating pair formation), synthesis of an extra strand of DNA specifically for transfer and transporting DNA across the donor and recipient membrane barriers through a mating pore. Finally, survival of the transferred DNA in the new host environment must have entered in the selection process during evolution of conjugative systems. These steps of conjugation are highly conserved whether the transfer is between bacteria or that between bacteria and eukaryotic cells.

3.1.1 Mechanisms of DNA Transfer

In Gram-negative bacteria, the mating pair formation requires building a special apparatus, pilus, which is a tube like extracellular structure whose role is to reach out and touch the recipient and bring about cell–cell contact by pilus retraction (by depolymerization of pilin subunits). In Gram-positive bacteria, cell–cell contact is mediated not by pili but by plasmid-encoded aggregating substances that are secreted in response to pheromones secreted by plasmid-free recipients. The cell–cell contact is believed to send a signal to the donor to form a mating pore through which the DNA is transferred. The conjugative proteins that form the pore belong to the type IV secretion systems, a widely used system to transport protein across bacterial membranes. DNA synthesis for transfer is well understood and requires an enzyme relaxase (nickase) that cleaves one strand of *oriT* (origin of transfer). The free 3′-OH of the nick serves as the primer for rolling circle replication. The replication steps are grossly similar to what has been described for pT181 (see Fig. 3; Sect. 2.1.3). The relaxase remains covalently bound to the 5′end, which is guided through the mating pore into the recipient and brings with it the attached single-strand as it is being released by replication. This process continues until the entire plasmid has been regenerated in the donor and a complete strand is transferred to the recipient. According to de la Cruz and colleagues, conjugation is a union of two basic processes: rolling circle replication and protein transport, helped by a coupling protein that guides the entry of the relaxase (pilot protein) into the protein transport machinery (mating pore/type IV secretion system).

3.1.2 Establishment of Transferred DNA

Establishment of the transferred strand in the recipient cell requires first the strand to be circularized. The relaxase covalently bound to the leading (5′) end of the transferred strand does this job. The 3′end most likely extends beyond one full length of the plasmid regenerating the *nic* site. The 5′-bound relaxase recognizes this site and by a cleavage–joining reaction regenerates a monomer circle. The next step is synthesis of the complementary strand. The synthesis would require multiple initiations as in any lagging strand synthesis and seems to be aided by plasmid-encoded primases that are transferred along with DNA to recipient. There are several safeguards to protect the transferred DNA in the recipient from attack by nucleases. The relaxase could be protecting the 5′-end from exonucleases. The lagging strand is most likely

synthesized as it enters the recipient. The conversion to double-stranded DNA could provide stability to inherently less stable single-stranded DNA and repair possibilities in case of strand breakage. Some plasmids have genes for SSB (single-strand binding) and PsiB proteins. The latter can prevent SOS response. Both the genes are transferred early and induced in the recipient during conjugation. SSB can protect single-stranded DNA and is known to facilitate many DNA metabolic processes, including replication. PsiB may interfere with RecA binding to single-stranded DNA and thereby speed up their conversion to double strand. Another early transferred gene is *ardA* whose role possibly is to protect the unmodified DNA from host restriction.

3.1.3 Regulation of Transfer

It is natural that a complex and energy-consuming process like conjugation will be highly regulated. The transfer genes are normally kept repressed (fertility inhibition), and both extra- and intracellular signals are involved in induction of the transfer process. Apparently, a balance is struck to minimize metabolic burden on the host and make the transfer an efficient process. Considerable details on regulation are known for plasmids like F, RP4/RK2, *Enterococcal* plasmid pAD1 and *Agrobacterium* plasmid Ti. It is also clear that the network of *tra* control coordinates with other plasmid processes like vegetative replication. Conjugative plasmids encode genes, which are expressed in the recipient at the completion of transfer to prevent a second bacterium to transfer the same or closely related plasmid. This requires coating the outer membrane of the recipient with plasmid encoded proteins or lipoproteins that discourage mating pair formation. Avoiding wasteful transfers also prevents cell lysis that may occur if donors outnumber the recipient and the conditions permit intense mating.

When the entire machinery for conjugation is encoded by plasmids themselves, they are called *self-transmissible*. There are also plasmids that only encode the requirements to initiate rolling circle replication (*oriT*, relaxase and a few accessory factors). These are called mobilizable plasmids, and a coresident self-transmissible plasmid is required for mating pair formation and transfer. Plasmids without any *tra* gene can also be mobilized if they can hop into a self-transmissible plasmid by transposon or other jumping genes (IS elements) or by homologous recombination. When F plasmid integrates with the *E. coli* chromosome, it can mobilize the entire 4.7 Mb of chromosomal DNA into a recipient indicating the efficiency and processivity of the conjugative process. This may require pushing DNA though the mating pore by a motor protein, as is the case in chromosome segregation at cell division by FtsK and during sporulation by SpoIIIE.

Horizontal transfer can also be considered a strategy for hereditary stability. In principle, a plasmid can be conjugated to multiple recipients and thereby outpace the rate of plasmid spread over cell growth rate. Although the transfer is tightly repressed and normally rare, plasmids in the recipient cells (exconjugants) remain derepressed transiently and this may allow further transfer to plasmid-free recipients. Thus, HGT can be considered an alternate means of plasmid maintenance. HGT should not affect plasmid maintenance by vertical transmission since transfer is tied to a new round of replication that should leave the plasmid copy number intact.

3.2 Broad Host Range

Although a plasmid can explore and establish in an unrelated bacterial species by conjugative transfer, its maintenance in the plasmid state requires, most importantly, vegetative replication. Several plasmids seem to have acquired this capability by minor adaptation of the mechanisms that maintain the narrow range plasmids. The adaptation has taken place in multiple ways that either results in changes in origin sequences, or in the plasmid-encoded initiator protein, in acquisition of multiple replicons in the same plasmid or in reduced dependence on host functions.

In one of the best-studied BHR plasmids, RK2, an iteron-type plasmid belonging to IncP group, a 35-bp insertion in the minimal origin inactivates the origin activity in *E. coli* and in *Pseudomonas putida* but not in *Pseudomonas aeruginosa*. In a different BHR plasmid, R1162, the minimal origin for replication in *E. coli* needs to be extended for replication in *P. putida*. It appears that one strategy to broaden plasmid host range is to increase the origin information content. The initiator protein of RK2, TrfA, also appears to have dispensable information for replication in some hosts that may become essential in some other hosts. The protein is made both as full length, TrfA-44, and as a smaller version, TrfA-33, from an internal in-frame translational start site. Although the smaller initiator suffices in many Gram-negative bacteria, the full-length protein is essential for replication in *P. aeruginosa*.

The understanding for the requirement for the larger TrfA-44 for replication in *P. aeruginosa* was greatly aided from studies *in vitro*. A crucial requirement for initiation of replication for any replicon is the loading of a replicative helicase. Replicons use a variety of mechanisms for this purpose. At the chromosomal origin of *E. coli*, the helicase DnaB loading requires both the host initiator DnaA and the helicase loader DnaC. DnaA interacts with origin DNA as well as DnaB but its transfer to origin strands requires DnaC. DnaA also binds to many plasmid origins including RK2 and apparently plays a similar role in loading of DnaB. However, TrfA-44 can load DnaB directly in *P. aeruginosa* without requiring DnaA and DnaC. It appears that the N-terminal part of TrfA-44 that is missing in TrfA-33 plays the role of both inviting DnaB to RK2 origin (normally the role of DnaA) and unloading of DnaB (normally the role of DnaC). As expected, DnaA-binding sites are not required for RK2 replication in *P. aeruginosa*. However, for loading of DnaB to RK2 origin in *E. coli*, both TrfA-44 and TrfA-33 still requires DnaA and DnaC. A better understanding of these results would require actual affinity measurements of various DNA–protein and protein–protein interactions. For example, TrfA-44 interaction with DnaB may be reversible enough to obviate the need for DnaC in *P. aeruginosa*.

Some BHR plasmids like RSF1010 (belonging to IncQ group) have opted to play safe by encoding both the plasmid-origin-specific initiator and the helicase so that the uncertainty of proper interfacing of the two is avoided in a new host. In *E. coli*, the DnaB helicase interacts with the DnaG primase and recruits it to the origin for the priming of DNA replication. Since RSF1010 makes its own helicase, it is in order that the plasmid also makes its own primase to ensure helicase–primase compatibility. These plasmids by not requiring DnaA, DnaB, DnaC, and DnaG can replicate in a wide variety of hosts including some Gram-positive bacteria.

The presence of more than one replicon in the same plasmid is a common occurrence (e.g. F, P1, R6K, pCU1) and in some cases (pCU1 belonging to IncN group) helps to extend the host range. Host-specific replicons have also been combined in the laboratory (called shuttle vectors) that replicate efficiently in both of their natural hosts. This provides the proof of principle that combining replicons is a feasible strategy to extend host range. A narrow host range plasmid can extend its host range by forming a cointegrate with a BHR plasmid. The alternate replicons, even if they do not increase the host range, can serve as back-up replicons in case the primary replicon fails to function and thereby contribute to plasmid stability.

An independent approach to study the basis of BHR has been to start with a narrow host range plasmid and extend the host range by mutation. Characterization of the mutants may then identify the relevant basis. It is clear that improving interactions between the plasmid-specific initiator and the host initiator improves the replication success in a new host. Plasmid pPS10 replicates efficiently in *P. aeruginosa* and in *P. putida* but not in *E. coli*. The inability to replicate in *E. coli* was overcome by mutating either the plasmid initiator RepA or the host initiator DnaA, implying that RepA–DnaA interaction was not optimal in *E. coli* for helicase loading. Since DnaA initiates the chain of events involving other replication proteins (DnaB, DnaC, DnaG etc.), once the plasmid initiator can interact with DnaA, the interactions with rest of the host proteins should follow automatically. It may seem that the adaptation in a new host may minimally require the ability to recruit DnaA to the plasmid origin. However, the two most promiscuous plasmids known to date, RK2 and RSF1010, have adopted strategies for independence from DnaA altogether for helicase loading.

Efficient maintenance would require that the replication be controlled and other maintenance functions like partition are operative. The BHR plasmids described above mostly belong to the iteron family of plasmids where genetic requirements for initiation and its control are the same. Control of replication will automatically follow with replication success. BHR plasmids also employ antisense RNA and/or a separate repressor protein for controlling copy number. Examples include several plasmids of Gram-positive origin such as pMV158, a *Streptococcal* plasmid and pC194, a *Staphylococcal* plasmid that can replicate even in a Gram-negative bacterium, *E. coli*. Extension of host range from Gram-positive to Gram-negative has also been obtained by mutating plasmids. The host range of maintenance functions other than replication is less known except that chromosomal *par* system of Gram-positive *B. subtilis* and Gram-negative *P. putida* stabilize a partition defective F plasmid in *E. coli*. The host range of other maintenance functions are likely to be broad as their requirement for BHR plasmids is expected to be as important as they are for the narrow host range plasmids.

3.3 Bacterial Evolution

A particularly striking discovery from genomic sequencing is the degree of occurrence of horizontal (lateral) gene transfer (HGT) among microorganisms. HGT is now recognized to be a major force in shaping microbial genomes and enhancing genetic diversity and adaptation in bacteria. For instance, there is more than 20%

difference in the genetic make up between two strains of *E.* coli (pathogen O157:H7 vs benign K12). Much of this difference (speciation) can be explained by gene acquisition and loss by HGT. Plasmids and other mobile genetic elements offer blocks of DNA (an extreme example being the transfer of entire bacterial chromosome by conjugation) that appear to have accelerated bacterial adaptation and evolution much more than what would have been possible with point mutation and selection. Within a brief span of seven years, diverse strains of nonsymbiotic rhizobia were found to have acquired symbiotic genes by HGT in the environment. One reason for fast plasmid-mediated adaptation is because plasmids can replicate faster than the host by spreading laterally to cells that lack them (conjugative transfer always involves an extra round of replication so that the donor plasmid remains in the donor cell). Currently, there is great interest in the studies of bacterial population genetics and evolution and on the role of mobile genetic elements in maintaining bacterial diversity and adaptation.

3.4 Impact on Public Health

Although understanding the mechanisms of BHR is mechanistically challenging, and cloning vectors based on BHR plasmids are boon to the burgeoning field of research in non–*E. coli* bacteria, the plasmids are potentially the most dangerous for spreading genes undesirable for human health. Usually genes conferring resistance to a variety of antibiotics and toxic heavy metals, and encoding virulence factors are found in mobile genetic elements such as plasmids, transposons, and bacteriophages, not in the host chromosome. Plasmid-borne virulence factors are found to be involved in a wide variety of functions such as adhesion, host cell penetration and invasiveness, high-affinity iron transport, and diseases caused by a bacterial toxins including anthrax, tetanus, diarrhea (due to labile toxin, LT) and food poisoning (due to enterotoxin D). However, by far, the biggest impact on public health has been the rapid spread of antibiotic resistance bacteria in the environment. Every class of antibiotics has developed resistance within a year to a decade of clinical use. In the United States, resistance to *Streptococcus pneumoniae* to two commonly prescribed antibiotics is predicted to reach 40% by July 2004. Antibiotics resistance is therefore not a matter of if but when. The widespread use of antibiotics is the cause of selection for this high rate of resistance. The problem is compounded by the presence of several resistance genes in the same plasmid or other mobile genetic elements, and consequently, in the selection of multidrug resistance bacteria that can resist all available antibiotics. Normally, transposable elements are the primary source of resistance genes. The much-studied conjugative plasmid R1(also called R100/NR1, see Sect. 2.1.3), isolated in late 1950s, was found to carry genes for resistance to chloramphenicol, fusidic acid, spectinomycin, streptomycin, sulfonamide, and tetracycline primarily due to transposon hops. Time has come to explore alternates to antibiotics to treat bacterial infections, and fundamental research in plasmid will obviously guide the future line of research on topics such as on how to avert conjugal transfer, which is presently thought to be the major means of transfer of antibiotic–resistance determinants in the environment.

4
Concluding Remarks

Using a few well-studied systems, the article tries to present our current thinking on how plasmids are maintained and has only touched upon other facets of plasmid research such as public health, and bacterial diversity and adaptation. The emphasis has been on mechanisms. Use of plasmid as research tools and in biotechnology have not been discussed. These topics deserve to be covered in separate articles.

4.1
As Model Chromosomes

Because of their small size and dispensable nature, plasmids appeared attractive as models for studying chromosome maintenance in the 1970s. It was certainly logical to expect that the issues involved would be similar. However, it turns out that the *E. coli* chromosome is regulated differently from its plasmids. The cell can tolerate about 30 extra copies of its origin whereas plasmid replication essentially stops with a two-fold increase of copy number. To date also no chromosomal centromere has been found in *E. coli*, suggesting that the chromosome segregates differently from plasmids. Hopefully, emerging interest in non–*E. coli* systems will permit generalization on maintenance functions in the bacterial world.

Plasmids need to regulate its replication independent from that of the host. Since plasmids need to adjust the rate of replication to correct for fluctuations in copy number, it cannot afford to follow once per cell cycle replication characteristic of the host chromosome. This requirement for independence from the host appears sufficient to justify encoding of replication control functions in plasmids without exception. The structure of iteron-mediated plasmids does give the impression that its origin has evolved from the host origin by incorporating plasmid-specific control elements, the iterons. Although plasmid *par* genes are found in the chromosome of several non–*E. coli* bacteria such as *Bacillus subtilis*, *Vibrio cholerae*, *Pseudomonas putida*, and *Caulobacter crescentus*, it remains to be seen whether the mechanisms are similar. It is already evident that because chromosome partition has to deal with a much larger piece of DNA, additional requirements are there for chromosome condensation, which can be dispensed with for plasmids. The *par* genes are essential in *Caulobacter crescentus* suggesting that the genes might serve additional functions. Genetic requirements for dimer resolution also are different for the ColE1 plasmid and the *E. coli* chromosome. The plasmid maintenance function that was least expected in the chromosome, the toxin-antidote system, however, is found in multiple copies in the chromosome (see Sect. 2.3.2). It seems that the common ground is likely to widen as our knowledge of plasmids and chromosomes progress.

4.2
Contributions

Even if plasmids do not exactly fit the bill as model bacterial chromosomes, the research in plasmid has been rewarding in revealing several new regulatory principles. The first and foremost is the seminal discovery of regulation by antisense RNA (Sect. 2.1.2). Other novel mechanisms include initiator inactivation by a covalent DNA adduct (Sect. 2.1), gene silencing by spreading of a site-specific DNA-binding to flanking DNA (Sect. 2.2), microtubule-like

polymer formation in bacteria (Sect. 2.2), pole-to-pole oscillation of a partition protein in plasmid segregation (Sect. 2.2), plasmid addiction by toxin–antitdote gene pair (Sect. 2.3), and plasmid multimerization dependent gene expression of *rcd* (Sect.2.4). One can also consider the elucidation of the steps of maturation of *hok* RNA as an outstanding achievement (Sect. 2.3.1).

4.3 Future Potential

The boundary between chromosome and plasmid has become fuzzier with the discovery of chromosome-size megaplasmids such as pSym plasmids of *Shinorhizobium meliloti*, distribution of genome into multiple genetic elements as in *Borellia*, and ever-increasing discovery of plasmid genes in chromosomes. The new realization of the impact of plasmids on bacterial evolution has made understanding of plasmid behavior particularly in the environment all the more necessary. Little is known about how cells communicate and prepare themselves for conjugative plasmid transfer. BHR plasmids are particularly powerful to study adaptation at the molecular level. Plasmids will remain attractive and convenient systems to study how genes are counted, their fluctuations are adjusted, how they are sorted, and how they migrate intra- and extracellularly. Much of these processes need to be understood quantitatively and in kinetic terms. The present day technologies are allowing single cell and even single molecule measurements. These are welcome developments as steady state measurement is an imperfect proxy to understand control processes. If past performance is any guide, a wealth of new knowledge remains to be gained by studying existing plasmids in greater depth in the laboratory and in their natural niche, and studying new plasmids from exotic habitats.

See also Bacterial Growth and Division; Bacterial Pathogenesis, Molecular Basis of; Chromosome, Microdissection and Microcloning; DNA Libraries.

Bibliography

Books and Reviews

Boye, E., Nordstrom, K. (2003) Coupling the cell cycle to cell growth, *EMBO Rep.* **4**, 757–760.

del Solar, G., Giraldo, R., Ruiz-Echevarria, M.J., Espinosa, M., Diaz-Orejas, R. (1998) Replication and control of circular bacterial plasmids, *Microbiol. Mol. Biol. Rev.* **62**, 434–464.

Funnell, B.E. (2004) *The Biology of Plasmids*, ASM, Washington, DC.

Gerdes, K., Moller-Jensen, J., Bugge Jensen, R. (2000) Plasmid and chromosome partitioning: surprises from phylogeny, *Mol. Microbiol.* **37**, 455–466.

Hayes, F. (2003) Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest, *Science* **301**, 1496–1499.

Koonin, E.V., Makarova, K.S., Aravind, L. (2001) Horizontal gene transfer in prokaryotes: quantification and classification, *Annu. Rev. Microbiol.* **55**, 709–742.

Llosa, M., Gomis-Ruth, F.X., Coll, M., de la Cruz Fd, F. (2002) Bacterial conjugation: a two-step mechanism for DNA transport, *Mol. Microbiol.* **45**, 1–8.

Summers, D.K. (1996) *The Biology of Plasmids*, Blackwell Science, Oxford.

Summers, D. (1998) Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability, *Mol. Microbiol.* **29**, 1137–1145.

Thomas, C.M. (2000) *The Horizontal Gene Pool*, Harwood Academic Publishers, Amsterdam.

Primary Literature

Afif, H., Allali, N., Couturier, M., Van Melderen, L. (2001) The ratio between CcdA and CcdB modulates the transcriptional repression of the ccd poison-antidote system, *Mol. Microbiol.* **41**, 73–82.

Ansari, A., Gartenberg, M.R. (1997) The yeast silent information regulator Sir4p anchors and partitions plasmids, *Mol. Cell Biol.* **17**, 7061–7068.

Austin, S., Nordström, K. (1990) Partition-mediated incompatibility of bacterial plasmids, *Cell* **60**, 351–354.

Belogurov, A.A., Delver, E.P., Agafonova, O.V., Belogurova, N.G., Lee, L.Y., Kado, C.I. (2000) Antirestriction protein Ard (Type C) encoded by IncW plasmid pSa has a high similarity to the "protein transport" domain of TraC1 primase of promiscuous plasmid RP4, *J. Mol. Biol.* **296**, 969–977.

Boye, E., Nordstrom, K. (2003) Coupling the cell cycle to cell growth, *EMBO Rep.* **4**, 757–760.

Chattoraj, D.K. (2000) Control of plasmid DNA replication by iterons: no longer paradoxical, *Mol. Microbiol.* **37**, 467–476.

Chattoraj, D., Cordes, K., Abeles, A. (1984) Plasmid P1 replication: negative control by repeated DNA sequences, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6456–6460.

Chatwin, H.M., Summers, D.K. (2001) Monomer–dimer control of ColE1 P_{cer} promoter, *Microbiology* **147**, 3071–3081.

Clewell, D.B., Francia, M.V., Flannagan, S.E., An, F.Y. (2002) Enterococcal plasmid transfer: sex pheromones, transfer origins, relaxases, and the Staphylococcus aureus issue, *Plasmid* **48**, 193–201.

Critchlow, S.E., O'Dea, M.H., Howells, A.J., Couturier, M., Gellert, M., Maxwell, A. (1997) The interaction of the F plasmid killer protein, CcdB, with DNA gyrase: induction of DNA cleavage and blocking of transcription, *J. Mol. Biol.* **273**, 826–839.

de la Hoz, A.B., Ayora, S., Sitkiewicz, I., Fernandez, S., Pankiewicz, R., Alonso, J.C., Ceglowski, P. (2000) Plasmid copy-number control and better-than-random segregation genes of pSM19035 share a common regulator, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 728–733.

del Solar, G., Alonso, J.C., Espinosa, M., Diaz-Orejas, R. (1996) Broad-host-range plasmid replication: an open question, *Mol. Microbiol.* **21**, 661–666.

del Solar, G., Hernandez-Arriaga, A.M., Gomis-Ruth, F.X., Coll, M., Espinosa, M. (2002) A genetically economical family of plasmid-encoded transcriptional repressors involved in control of plasmid copy number, *J. Bacteriol.* **184**, 4943–4951.

Diaz-Lopez, T., Lages-Gonzalo, M., Serrano-Lopez, A., Alfonso, C., Rivas, G., Diaz-Orejas, R., Giraldo, R. (2003) Structural changes in RepA, a plasmid replication initiator, upon binding to origin DNA, *J. Biol. Chem.* **278**, 18606–18616.

Dibbens, J.A., Muraiso, K.M., Chattoraj, D.K. (1997) Chaperone-mediated reduction of RepA dimerization is associated with RepA conformational change, *Mol. Microbiol.* **25**, 185–195.

Draper, G.C., Gober, J.W. (2002) Bacterial chromosome segregation, *Annu. Rev. Microbiol.* **56**, 567–597.

Eguchi, Y., Itoh, T., Tomizawa, J. (1991) Antisense RNA, *Annu. Rev. Biochem.* **60**, 631–652.

Fang, L., Davey, M.J., O'Donnell, M. (1999) Replisome assembly at oriC, the replication origin of E. coli, reveals an explanation for initiation sites outside an origin, *Mol. Cell.* **4**, 541–553.

Funnell, B.E., Gagnier, L. (1995) Partition of P1 plasmids in Escherichia coli mukB chromosomal partition mutants, *J. Bacteriol.* **177**, 2381–2386.

Giraldo, R. (2003) Common domains in the initiators of DNA replication in bacteria, archaea and eukarya: combined structural, functional and phylogenetic perspectives, *FEMS Microbiol. Rev.* **26**, 533–554.

Godfrin-Estevenon, A.M., Pasta, F., Lane, D. (2002) The parAB gene products of Pseudomonas putida exhibit partition activity in both P. putida and Escherichia coli, *Mol. Microbiol.* **43**, 39–49.

Gordon, G.S., Sitnikov, D., Webb, C.D., Teleman, A., Straight, A., Losick, R., Murray, A.W., Wright, A. (1997) Chromosome and low copy plasmid segregation in E. coli: visual evidence for distinct mechanisms, *Cell* **90**, 1113–1121.

Hao, J.J., Yarmolinsky, M. (2002) Effects of the P1 plasmid centromere on expression of P1 partition genes, *J. Bacteriol.* **184**, 4857–4867.

Hayes, F. (2003) Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest, *Science* **301**, 1496–1499.

Helinski, D.R., Toukdarian, A.E., Novick, R.P. (1996) Replication control and other stable maintenance mechanisms of plasmids, *Escherichia coli Salmonella* **2**, 2295–2324.

Ho, T.Q., Zhong, Z., Aung, S., Pogliano, J. (2002) Compatible bacterial plasmids are targeted to independent cellular locations in Escherichia coli, *EMBO J.* **21**, 1864–1872.

Hunding, A., Ebersbach, G., Gerdes, K. (2003) A mechanism for ParB-dependent waves of ParA, a protein related to DNA segregation during cell division in prokaryotes, *J. Mol. Biol.* **329**, 35–43.

Ingmer, H., Miller, C., Cohen, S.N. (2001) The RepA protein of plasmid pSC101 controls Escherichia coli cell division through the SOS response, *Mol. Microbiol.* **42**, 519–526.

Jacob, F. (1993) The replicon: thirty years later, *Cold Spring Harbor Symp. Quant. Biol.* **58**, 383–387.

Jiang, Y., Pacek, M., Helinski, D.R., Konieczny, I., Toukdarian, A. (2003) A multifunctional plasmid-encoded replication initiation protein both recruits and positions an active helicase at the replication origin, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8692–8697.

Konieczny, I., Helinski, D.R. (1997) The replication initiation protein of the broad-host-range plasmid RK2 is activated by the ClpX chaperone, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 14378–14382.

Kulakauskas, S., Lubys, A., Ehrlich, S.D. (1995) DNA restriction-modification systems mediate plasmid maintenance, *J. Bacteriol.* **177**, 3451–3454.

Lander, E.S. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Lederberg, J. (1998) (1952–1997) Plasmid (1952–1997), *Plasmid* **39**, 1–9.

Lemonnier, M., Bouet, J.Y., Libante, V., Lane, D. (2000) Disruption of the F plasmid partition complex in vivo by partition protein SopA, *Mol. Microbiol.* **38**, 493–505.

Leonard, A.C., Helmstetter, C.E. (1988) Replication patterns of multiple plasmids coexisting in Escherichia coli, *J. Bacteriol.* **170**, 1380–1383.

Li, Y., Austin, S. (2002) The P1 plasmid is segregated to daughter cells by a 'capture and ejection' mechanism coordinated with Escherichia coli cell division, *Mol. Microbiol.* **46**, 63–74.

Llosa, M., Gomis-Ruth, F.X., Coll, M., de la Cruz Fd, F. (2002) Bacterial conjugation: a two-step mechanism for DNA transport, *Mol. Microbiol.* **45**, 1–8.

Lobner-Olesen, A. (1999) Distribution of minichromosomes in individual Escherichia coli cells: implications for replication control, *EMBO J.* **18**, 1712–1721.

Lu, Y.B., Datta, H.J., Bastia, D. (1998) Mechanistic studies of initiator-initiator interaction and replication initiation, *EMBO J.* **17**, 5192–5200.

Lynch, A.S., Wang, J.C. (1995) SopB protein-mediated silencing of genes linked to the sopC locus of Escherichia coli F plasmid, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1896–1900.

Maestro, B., Sanz, J.M., Faelen, M., Couturier, M., Diaz-Orejas, R., Fernandez-Tresguerres, E. (2002) Modulation of pPS10 host range by DnaA, *Mol. Microbiol.* **46**, 223–234.

Moller-Jensen, J., Jensen, R.B., Lowe, J., Gerdes, K. (2002) Prokaryotic DNA segregation by an actin-like filament, *EMBO J.* **21**, 3119–3127.

Mukherjee, S., Patel, I., Bastia, D. (1985) Conformational changes in a replication origin induced by an initiator protein, *Cell* **43**, 189–197.

Naito, T., Kusano, K., Kobayashi, I. (1995) Selfish behavior of restriction-modification systems, *Science* **267**, 897–899.

Nonaka, N., Kitajima, T., Yokobayashi, S., Xiao, G., Yamamoto, M., Grewal, S.I., Watanabe, Y. (2002) Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast, *Nat. Cell. Biol.* **4**, 89–93.

Nordström, K. (1984) Equipartition and other modes of partition: on the interpretation of curing kinetics using rep(ts) plasmids, *Mol. Gen. Genet.* **198**, 185–186.

Nordström, K., Gerdes, K. (2003) Clustering versus random segregation of plasmids lacking a partitioning function: a plasmid paradox? *Plasmid* **50**, 95–101.

Novick, R.P. (2003) Mobile genetic elements and bacterial toxinoses: the superantigen-encoding pathogenicity islands of Staphylococcus aureus, *Plasmid* **49**, 93–105.

Novick, R.P., Hoppensteadt, F.C. (1978) On plasmid incompatibility, *Plasmid* **1**, 421–434.

Onogi, T., Miki, T., Hiraga, S. (2002) Behavior of sister copies of mini-F plasmid after synchronized plasmid replication in

Escherichia coli cells, *J. Bacteriol.* **184**, 3142–3145.
Paulsson, J., Nordström, K., Ehrenberg, M. (1998) Requirements for rapid plasmid ColE1 copy number adjustments – a mathematical model of inhibition modes and RNA turnover rates, *Plasmid* **39**, 215–234.
Pedersen, K., Gerdes, K. (1999) Multiple hok genes on the chromosome of Escherichia coli, *Mol. Microbiol.* **32**, 1090–1102.
Praszkier, J., Pittard, A.J. (2002) Pseudoknot-dependent translational coupling in repBA genes of the IncB plasmid pMU720 involves reinitiation, *J. Bacteriol.* **184**, 5772–5780.
Rawlings, D.E., Tietze, E. (2001) Comparative biology of IncQ and IncQ-like plasmids, *Microbiol. Mol. Biol. Rev.* **65**, 481–496.
Rine, J. (1999) On the mechanism of silencing in Escherichia coli, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8309–8311.
Rodionov, O., Lobocka, M., Yarmolinsky, M. (1999) Silencing of genes flanking the P1 plasmid centromere, *Science* **283**, 546–549.
Sat, B., Reches, M., Engelberg-Kulka, H. (2003) The Escherichia coli mazEF suicide module mediates thymineless death, *J. Bacteriol.* **185**, 1803–1807.
Sherratt, D.J. (2003) Bacterial chromosome dynamics. *Science*, **301**, 780–785.
Sullivan, J.T., Patrick, H.N., Lowther, W.L., Scott, D.B., Ronson, C.W. (1995) Nodulating strains of Rhizobium loti arise through chromosomal symbiotic gene transfer in the environment, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8985–8989.
Waters, V.L. (2001) Conjugation between bacterial and mammalian cells, *Nat. Genet.* **29**, 375–376.
Wickner, S., Hoskins, J., McKenney, K. (1991) Monomerization of RepA dimers by heat shock proteins activates binding to DNA replication origin, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7903–7907.
Woldringh, C.L. (2002) The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation, *Mol. Microbiol.* **45**, 17–29.
Yamaichi, Y., Niki, H. (2000) Active segregation by the Bacillus subtilis partitioning system in Escherichia coli, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14656–14661.
Yarmolinsky, M.B. (1995) Programmed cell death in bacterial populations [comment], *Science* **267**, 836–837.
Yates, P., Lane, D., Biek, D.P. (1999) The F plasmid centromere, sopC, is required for full repression of the sopAB operon, *J. Mol. Biol.* **290**, 627–638.
Yates, J., Aroyo, M., Sherratt, D.J., Barre, F.X. (2003) Species specificity in the activation of Xer recombination at dif by FtsK, *Mol. Microbiol.* **49**, 241–249.

Population, Specific Genetic Markers and Disease: *see* Human Genetic Variation and Disease

Porins, Mitochondrial: *see* Mitochondrial Porins, Eukaryotic

Pragmatic Computational Biology: Sequence Analysis and Biological Systems Description

Andrzej K. Konopka
BioLingua Research, Inc., Gaithersburg, MD

1 Preliminaries 451
1.1 Codes and (Biological) Function: A Logical Trap 451
1.2 Classification Codes 451
1.3 Pragmatic Inference 451
1.4 Determining Classification Code Words 451

2 Statistical Modeling and Frequency Counts 453
2.1 Nonoverlapping and Overlapping Frequency Counts 453
2.2 Ad Hoc (Simple) Frequency Counts 455
2.3 Ranked Oligonucleotide Frequencies; Yule Distribution 456

3 Measuring Deviations from Discrete Uniform Distribution (DUD) 458
3.1 Entropy and Redundancy of Nucleotide Sequences 458
3.2 Coincidence Indices and Their Use in Sequence Analysis 460

4 Noncontiguous Patterns and Distance Charts 464

5 Local Compositional Complexity: A Maximum Entropy Principle 466
5.1 Local Compositional Complexity 466
5.2 Surprisal 467
5.3 Log-odds Complexity Distribution 468
5.4 Patchy Complexity 468

6 Computer-assisted Systems Biology 469
6.1 Pragmatic and Semantic Theories of Information 471
6.2 Logic of Heuristic Reasoning 472
6.3 Question of Measurement in Biology 472
6.4 Convoluted Nature of Biological Systems 473

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

Bibliography 473
Books and Reviews 473
Primary Literature 474

Keywords

k-GRAM (k-tuple)
String (sequence) of k symbols chosen from a given elementary alphabet. In the case of nucleotide sequences, k-grams correspond to oligonucleotides of length k. The most frequently used nucleic acids elementary alphabet is {A, C, G, T (or U)}, where letters stand for adenine, cytosine, guanine, and thymine (or uracil) nucleotides respectively. Other alphabets used in nucleic acids studies include: {K, M}, where K is either guanine or thymine (uracil in RNA) and M is either adenine or cytosine; {R, Y} where R is a purine (adenine or guanine) and Y is pyrimidine (either cytosine or thymine); {S, W} where S is either cytosine or guanine and W is either adenine or thymine (or uracil in RNA).
If the elementary alphabet contains n symbols and k is fixed, we refer to a set of all n^k k-grams as *k-gram* (nonelementary) *alphabet*.

Pattern
There is no general definition of pattern. It can be any logical, geometrical or (broadly) factual connection between elements of a model that attracts our attention. ''Pattern'' is usually understood ostensively by experienced explorers of the same model.

Motif
Motif (sequence motif) is a term used in sequence analysis to refer to detectable and stable feature that can be used to characterize a conserved meaningful function or structure. There are several different specific meanings of ''motif,'' which depend on the subject matter (specific function or structure) to which they are applied. Most frequently used definitions of the word ''motif'' in biomolecular sequence analysis include the following:

1. Motif is a string or regular expression that significantly often occurs in a collection of similar sequences. Vast majority of ''motifs'' discussed in molecular biology literature are ''signature'' sequence patterns that distinguish a specific set of protein (or nucleic acid) sequences from all other sequences. The word ''motifs'' reflects here a tacit assumption of the existence of meaningful functional and structural constraints that are unique and detectable at the sequence level. Some authors also assume the existence of a true homology (as opposite to sequence or structure similarity) relating sequences in a given set of proteins (or just domains) or nucleic acids primary structures (see also sequence pattern and motif descriptor).
2. Motif is a pattern that can be considered a letter (unit of pattern) in an alphabet that is adequate for studying a specific correlation between sequences and their functional or structural role.

Motif Descriptor
A data structure used to define a sequence motif. Best- known descriptors include consensus sequences, weight matrices, and profiles.

Sequence Pattern
Given a k-gram alphabet, we select k-grams that we call *motifs* (meaningful strings) and k-grams that we call *punctuations* (semantically empty strings). (Heuristic principles for such choice are complex because they do involve knowledge of ''extra-sequential'' facts.) A *sequence pattern* is then a string of motifs and punctuations that begins and ends with a motif. *Contiguous Pattern* is a sequence pattern that does not contain punctuations whereas *noncontiguous pattern* is a sequence pattern that does contain punctuations.

Distance Chart
We select a particular k-gram to be a motif and then consider oligonucleotides separating the instances of this k-gram to be punctuations. The frequency of occurrence of punctuations plotted against their length is a *distance chart*. We refer to a distance chart in which the motif is not allowed to be a substring of any punctuation as the *shortest distance chart*.

Code
A mapping from one symbolic representation (input representation) of a system into another representation (output representation). If the input and output representations can be expressed in the form of elementary symbols chosen from finite alphabets, a code can be seen as a relation between input and output alphabets.

In a narrow sense, pragmatic computational biology is today's equivalent of theoretical molecular biology of the 1970s and 1980s combined with genome bioinformatics of 1990s and early 2000s. Similarly, to the latter it aims toward extended interpretation of laboratory and computational experiments in molecular genetics (also known as functional genomics), molecular evolution, and structural biology. The ''extended interpretation'' consists of several activities, such as the following:

1. Relating the mechanistic and the informational aspects of biological phenomena through appropriate generalization and modeling, for example, postulating the triplet nature of the genetic code before it had been determined experimentally.
2. Devising principles for analyzing inexact data (quantitative representations of qualitative descriptions included), for example, database design, annotation, and searches; logic of heuristic reasoning.
3. Relating principles and laws inferred from studies at the molecular level with those derived from studies of organisms, populations, and ecosystems, for example, molecular evolution, molecular evolutionary genetics.

4. Deriving properties of biological systems from principles of physics and chemistry, for example, biological thermodynamics and kinetics; theories of origin of life; molecular mechanics.
5. Modeling (usually mathematical) properties of laboratory instrumentation and designing experiments, for example, analysis and interpretation of molecular spectra, X-ray data, electrophoretic data, microchips, and so on.
6. Modeling properties of computational tools and designing computational experiments, for example, sequence alignment.

   These activities often draw on methods that in themselves are established research areas with their own methodological foundations. Some of the methods pertain to physics (for instance, molecular mechanics), some to chemistry (for example, reaction kinetics) and yet others are parts of applied mathematics (combinatorics, for example) or computer science (for instance, theory of algorithms). Pragmatic computational biology differs from chemistry, physics, and even biochemistry by systematically exploring symbolic aspects (besides the material ones) of the biologic role(s) of nucleic acid and protein sequences.

   In a larger sense, pragmatic computational biology is concerned with cognitive systematization of concepts, theories, models, and observations that could contribute to an explanation of the phenomenon of life itself. In this (larger) sense, the principles of modeling biological systems appear to be more important than design and interpretations of specific experiments. A nonexhaustive list of activities pertaining to the larger aspect of pragmatic computational biology includes the following:
7. Developing a general system theory and research on specific classes of general systems, for example, anticipatory systems and their use for the description of (biochemical) reaction networks.
8. Developing foundations of logic and mathematic such that the resulting mathematical objects would be appropriate for modeling biological systems, for example, theory of categories as a language for formulating models and theories.
9. Developing foundations of general information theory and information science, for example, work on semantic and pragmatic theories of information, knowledge representation, data integration, and so on.

Activities 7 to 9 draw on methods from other research programs such as systems theory, cognitive science, logic, and epistemology but do not belong to the current paradigms of these fields.

This article is primarily devoted to those aspects of computational biology that pertain to sequence analysis and genomics (activities 1 and 2 above) and thereby are of interest to molecular biologists proper. However, in order to satisfy those readers who would like to predict future research directions in the field, a brief discussion of topics belonging to the larger aspect of computer-assisted systems biology (activities 7–9 above) is given in Sect. 6.

1 Preliminaries

1.1 Codes and (Biological) Function: A Logical Trap

The focus on the relationship between biopolymer (nucleic acids and proteins) sequences and possible mechanisms of biological processes makes "code" the central, unifying metaphor in molecular biology. From the early period in molecular biology, biopolymer sequences and structures were thought of as texts that carry encoded messages. The three best-known "biological codes" are

1. Watson-Crick-Chargaff base-pairing rules,
2. translation code for protein biosynthesis (the genetic code), and
3. cleavage sites for restriction enzymes.

In the past, many biologists assumed that more *functional codes* of this kind existed. For instance, we could think of replication code, homologous and illegitimate recombination codes, mutation hot-spots code, protein or RNA folding codes, and so on. However, the adequate description of biological function at the molecular level is a formidable task. It requires knowledge of a great deal of mechanistic details that constitute the function. This is a major methodological obstacle in specifying a functional code. In most cases in which we know or surmise that a code exists, we are unable to appropriately describe the biological function concerned (the genetic code is a notable exception from this common situation).

1.2 Classification Codes

The foregoing methodological problems can be avoided when we replace the notion of functional codes by an alternative concept of *classification codes*. In principle, there are many-to-one mappings from a set of patterns found in (or resulting from) a class of sequences into descriptors (such as a name or a list of features) of this class. We specify a classification code by just listing patterns from its domain. We call each individual pattern from the list a *classification code word* or just a *code word*. Not every pattern found in a given collection of sequences is a classification code word. To qualify as a code word, the pattern must be *significant* according to predetermined criteria.

1.3 Pragmatic Inference

The criteria of significance usually involve a complex heuristics called *pragmatic inference*. (The name indicates a focus on the "rules" of usage for patterns). Pragmatic inference explores the entire body of available biological knowledge in addition to statistical analyzes. The resulting measure of significance often does not correspond to statistical significance because the rules of pragmatic inference are different from the rules of statistical inference. However, statistical modeling is a vital part of those steps of pragmatic inference in which selection of motifs and punctuations is made. Figure 1 shows a conceptual scheme of pragmatic inference for sequence research as seen by most computational biologists.

1.4 Determining Classification Code Words

Methods of determining code words draw on several assumptions concerning those patterns of characters (from an elementary alphabet) that are correct code

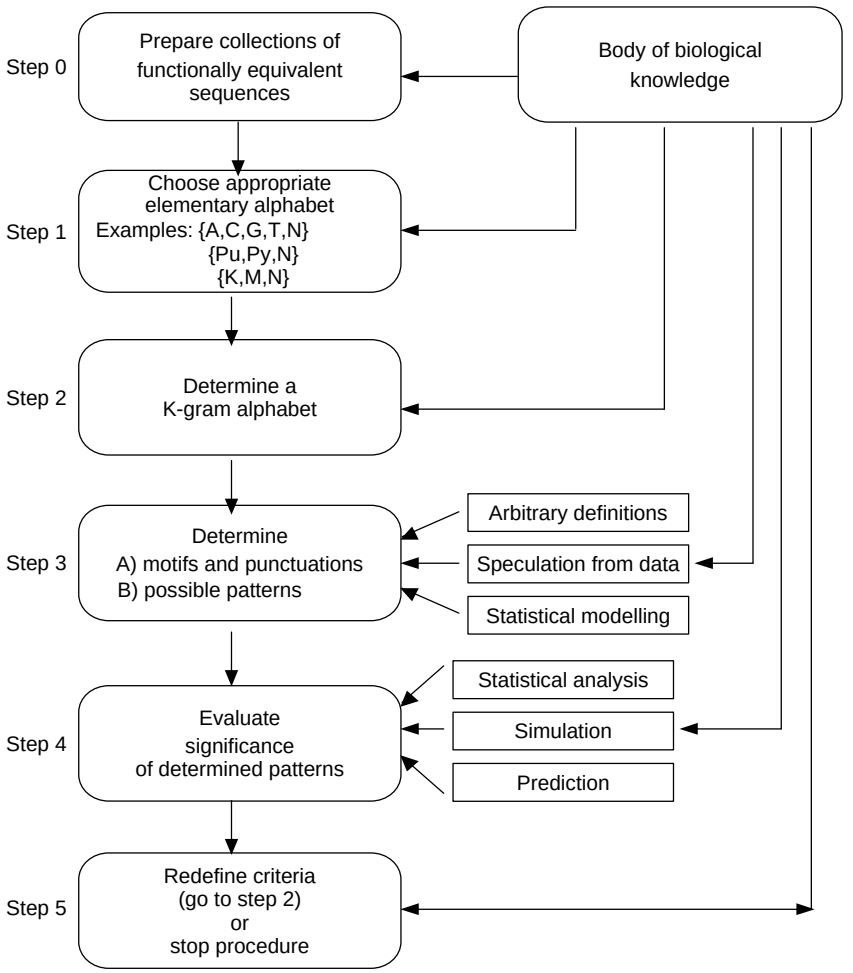

Fig. 1 General protocol of pragmatic inference pertaining to computer-assisted sequence analysis. Vaguely defined "biological knowledge" is taken into account at every step of inference.

words. A nonexhaustive list of assumptions includes:

A1. Relative frequency of a given code word should be the same (with the appropriate accuracy) in all sufficiently long sequences from a given functional class. (For sequences of infinite length, we expect that the relative frequencies converge to definite, invariant values (probabilities).)

A2. Probabilities of at least some individual code words should markedly differ from one another in all sequences from a given functional class.

A3. Probability distribution of code words should be the same (with appropriate accuracy) in all sequences from a given functional class.

A4. Selected code words should occupy distinct relative positions in every sequence from a given functional

class. (This assumption is known in linguistics as a postulate of existence of distributional structures.)

For methods to determine classification code words from corpora of sequences, explore assumptions A1 to A3. The most important are perhaps those based on coincidence indices (string length test and string repetition test) or on Markov analyses (including hidden Markov chains and nonhomogeneous Markov chains). A couple of powerful methods explore distributional structures (assumption A4) in the form of distance charts or Fourier analysis.

Knowing classification code words has practical consequences. It allows us to develop algorithms for prediction of putative functional domains in unannotated sequences. These algorithms (and the corresponding software) are of paramount importance for all genome-sequencing projects but we shall not discuss them here. The relevant discussion can be found in specialized publications on computational biology and genomics listed at the end of this article.

2 Statistical Modeling and Frequency Counts

To determine how "unusual" the occurrence of a given k-gram in a long sequence is, we have to compare the occurrence frequency in the data with the frequency expected from a model of chance.

It is in general a laborious task to find the best model of chance (i.e. prior probability distribution) for a given situation. A model often used in sequence research is a *Bernoulli text* in which all 1-grams occur independently of each other and with probabilities equal to each other. The corresponding probability distribution is known as *discrete uniform distribution* (DUD). (The reverse distribution assigns a k-gram to a given probability.)

Let determined from data frequency of ith k-gram be $F(i)$. Let the expected frequency of this k-gram be $F_0(i)$ and the variance (determined from a model of chance) be $V(i)$. The "degree of unusuality" of the occurrence of ith k-gram in the data is a function of $F(i)$, $F_0(i)$, and $V(i)$. The most common such function is a *z-score* defined as:

$$z(i) = \frac{F(i) - F_0(i)}{\sqrt{V(i)}} \qquad (1)$$

After having calculated z-score $z(i)$, we compare it with a threshold value z_0.

If $|z(i)| > z_0$, we consider the frequency of ith k-gram unusually high for $z(i) - z_0 > 0$ (overrepresentation) or unusually low for $z(i) - z_0 < 0$ (underrepresentation).

2.1 Nonoverlapping and Overlapping Frequency Counts

There are two general ways of calculating k-gram frequencies in a given long sequence (of length L): nonoverlapping count and overlapping counts. In the case of nonoverlapping count, we take into account k-grams in positions 1, $1 + k$, $1 + 2k, \ldots, 1 + mk < L$. The number $(m + 1)$ of k-grams counted in this way equals the integer part of L/k. Cases of overlapping counts of k-grams differ from each other by "counting steps" of 1 through $k - 1$ (counting step k is nonoverlapping count). In the overlapping count of step 1 we take into account k-grams in positions $1, 2, 3, \ldots, L - k + 1$ and there are exactly $L - k + 1$ k-grams in the sample of total length L. (This counting is equivalent to shifting the window of length k by

1 nucleotide at the time). If we chose a counting step of 2 and k is greater than 2, we take into account k-grams in every other position (i.e. 1, 3, 5, . . ., and so on). Counting step of 3 (if k is greater than 3) means taking into account every third position (i.e. 1, 4, 7, . . ., and so on). Most molecular biologists understand the overlapping count as a count of step 1 and we will use this convention in this article (i.e. we will not discuss cases of counting steps other than 1 and k).

When the elementary alphabet has n letters (1-grams) and when k is fixed, the k-gram alphabet contains $N = nk$ elements. In a Bernoulli text of length L the expected frequency of each k-gram is therefore $F_0 = L/nk$. For the nonoverlapping count variance is a simple function of n, k, and L given by the equation

$$V = \frac{L}{n^k} \cdot \left(1 - \frac{2k-1}{n^k}\right) \tag{2}$$

For the overlapping, count variance is no longer a simple function of n, k, and L and has to be calculated for each individual k-gram separately. The reason for this complication is the variance's dependence on the *self-overlap capacity* of individual k-grams.

The self-overlap capacity of a given k-gram S is determined from a polynomial:

$$K_S = \sum_{i=0}^{k-1} a_i \cdot x^i \tag{3}$$

Coefficients a_i equal 1 if the first and last $(k-i)$-grams in S are identical and equal to 0 otherwise. For example, the polynomial for trinucleotide GGG is $K_{\mathrm{GGG}} = 1 + x + x^2$. For trinucleotides GAG and GAT, we have $K_{\mathrm{GAG}} = 1 + x^2$ and $K_{\mathrm{GAT}} = 1$ (no self - overlap) respectively.

The variance of frequency distribution for a k-gram S in a Bernoulli text of length L over an elementary alphabet containing n letters is a function of $K_S(1/n)$ given by the formula:

$$V_S = \frac{L}{n^k} \cdot \left[2K_S\left(\frac{1}{n}\right) - 1 - \frac{2k-1}{n^k}\right] \tag{4}$$

Fig. 2 Examples of dinucleotide frequency counts in representative (large) samples of DNA sequences from seven different classes of genomes. Three different elementary alphabets were taken into account. (a) – {K, M} alphabet, where K is either guanine or thymine and M is either adenine or cytosine. Assuming that the threshold z-score values for unusually high frequency is +3 and for unusually low frequency is −3, none of dinucleotides over this alphabet occur significantly more often or significantly less often than others in all sequence samples studied. However, "complex" dinucleotides KM and MK tend to be slightly but systematically underrepresented. (b) – {R, Y} alphabet, where R stands for purine (adenine or guanine) and Y stands for pyrimidine (cytosine or thymine). "Complex" dinucleotides RY and YR appear to be unusually underrepresented in organella and animal (i.e. primate, rodent, other mammalian, other vertebrate and invertebrate) genomes. This underrepresentation might be due in part to the well-known avoidance of long alternating purine-pyrimidine tracks (sequences with a potential for B-DNA – Z-DNA tertiary structure transition). Note that in other genomes (bacteria, phages, yeast, viruses, and plants), slight underrepresentation of RY and YR takes place as well. (c) – {S, W} alphabet, where S stands for either guanine or cytosine and W stands for either adenine or thymine. It can be seen that dinucleotide SS is underrepresented in phage, organella (mitochondria and chloroplast) and yeast genomes. In contrast, dinucleotide WW is overrepresented in these genomes. Notably, dinucleotides SW and WS are significantly underrepresented in organella (z-score < -3) and tend to be underrepresented in other genomes as well.

Failure to consider self-overlap capacity in overlapping counts leads to huge errors (hundreds of percent) in variance calculation, particularly in cases of small-size elementary alphabets and large values of k. (Variance for nonoverlapping count (Eq. 2) can be derived from Eq. 4 by putting $K_S(1/n) = 1$ (no self-overlap).)

2.2 Ad Hoc (Simple) Frequency Counts

Examples of the overlapping 2-gram frequency counts in seven large collections of DNA sequences are shown in Fig. 2. Although the "folklore" of describing and interpreting k-gram frequency counts

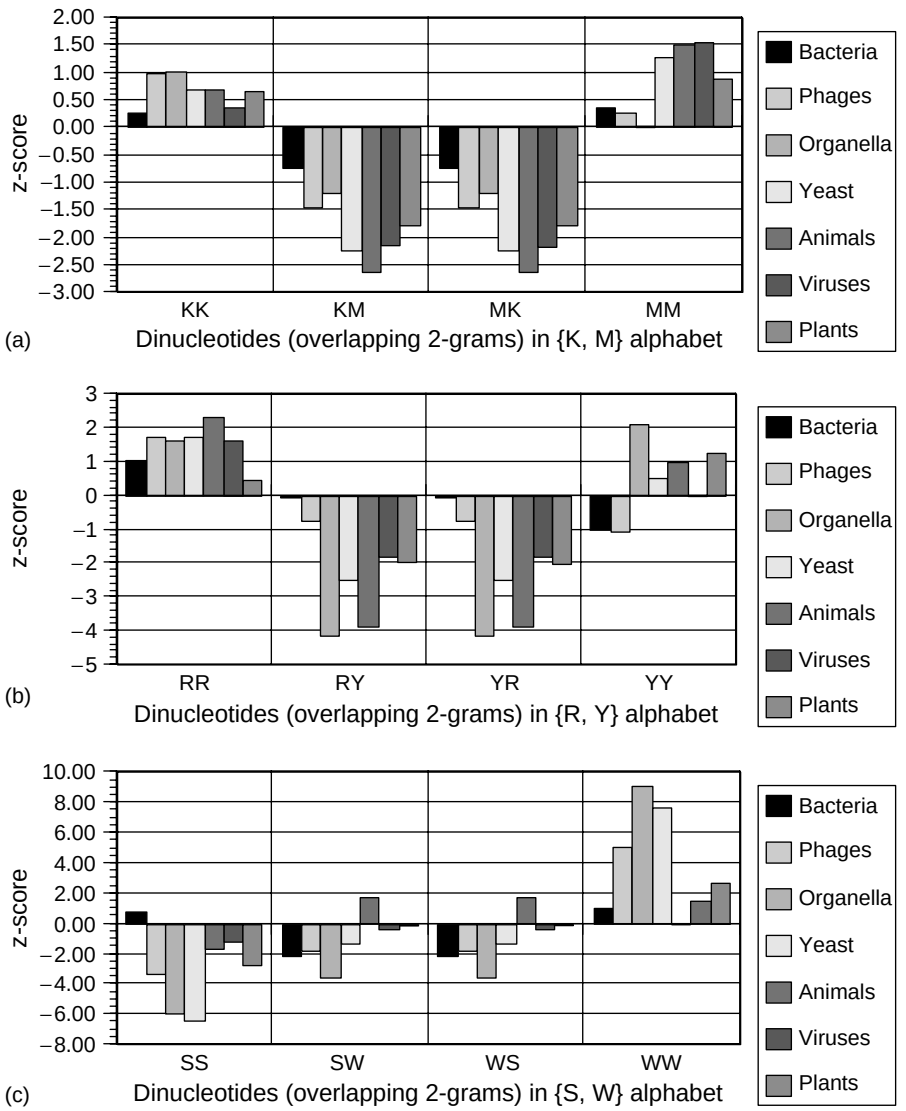

resembles descriptive biology of the nineteenth century (see caption to Fig. 2), it plays an invaluable role in selecting elementary alphabets and k-gram candidates for motifs.

2.3 Ranked Oligonucleotide Frequencies; Yule Distribution

Ranks of word frequency in printed alphabetic languages appear to follow the so-called *Zipf distribution*, which is given by the equation:

$$F = a \cdot R^k \quad \text{and } k < 0 \tag{5}$$

where F is frequency of occurrence and R is rank from most frequent to least frequent.

The same distribution also characterizes a host of nonlinguistic phenomena such as the number of species per genus, the number of papers per author in a given field of science, the distribution of income, and the population of cities in a nation.

Attempts to find a fit of ranked oligonucleotide frequencies to Zipf distribution failed soon after the first nucleotide sequences were known. It is clear by now that short oligonucleotide frequencies from all known genomes do not follow Zipf distribution but a more general 3-parameter density called the *Yule distribution*, which is given by the equation:

$$F = a \cdot R^k \cdot b^R \tag{6}$$

Tab. 1 Fit of Yule equation to ranked hexanucleotide frequencies in several large collections of intron and exon sequences. When plotted on log–log coordinates, all of the data exhibit the convex form shown in Fig. 3. The mean values of parameters a, b and k in introns and exons are the same within error limits (a-intron $= 0.0034 \pm 0.0026$; a-exon $= 0.0018 \pm 0.0005$; b-intron $= 0.9981 \pm 0.0036$; b-exon $= 0.9994 \pm 0.0001$; k-intron $= 0.25 \pm 0.11$; k-exon $= 0.15 \pm 0.03$). This implies that Yule parameters should not be used to discriminate between coding and noncoding regions in nucleotide sequences (the sampling errors in noncoding regions are too high). However, high ρ^2 values show that fit to the Yule distribution is excellent for both coding and noncoding regions regardless of the species. (Based on Martindale, C., Konopka, A.K. (1996), *Comput. Chem.* **20**(1), 35–38.)

	Base pairs	*Yule distribution*			
		a	*k*	*b*	ρ^2
Primates introns	346,953	0.0016	−0.15	0.9994	0.96
Primates exons	500,899	0.0017	−0.14	0.9993	0.97
Vertebrates introns	232,272	0.0017	−0.16	0.9995	0.96
Vertebrates exons	370,327	0.0021	−0.18	0.9994	0.97
Invertebrates introns	181,184	0.0034	−0.28	0.9995	0.99
Invertebrates exons	499,601	0.0016	−0.15	0.9994	0.95
Organella introns	105,189	0.0045	−0.33	0.9996	0.99
Organella exons	531,943	0.0020	−0.17	0.9994	0.97
Plant introns	255,535	0.0027	−0.24	0.9995	0.98
Plant exons	588,172	0.0015	−0.13	0.9994	0.95
Yeast introns	18,821	0.0089	−0.43	0.9999	0.97
Yeast exons	333,789	0.0025	−0.19	0.9993	0.99
Viral introns	75,359	0.0015	−0.15	0.9995	0.99
Viral exons	733,231	0.0011	−0.09	0.9995	0.89
Bacterial genes	876,257	0.0015	−0.12	0.9993	0.95
Phage genes	353,104	0.0022	−0.19	0.9994	0.98

where F is frequency and R is the rank; a, b, and k are parameters.

Zipf distribution (Eq. 5) is a special case of the Yule distribution (Eq. 6) in which parameter $b = 1$. Even if b is very close to 1 but not exactly equal to 1 (for instance $b = 0.9995$), the Yule distribution remarkably differs from the Zipf distribution.

Parameters of Yule distribution for hexanucleotide frequencies in large collections of intron and exon sequences are shown in Table 1. It appears that both intron and exon data fit the Yule distribution extremely well (and about equally well; see caption to Table 1). Example plots (in log-rank-log-frequency coordinates) of the Yule distribution are shown in Fig. 3.

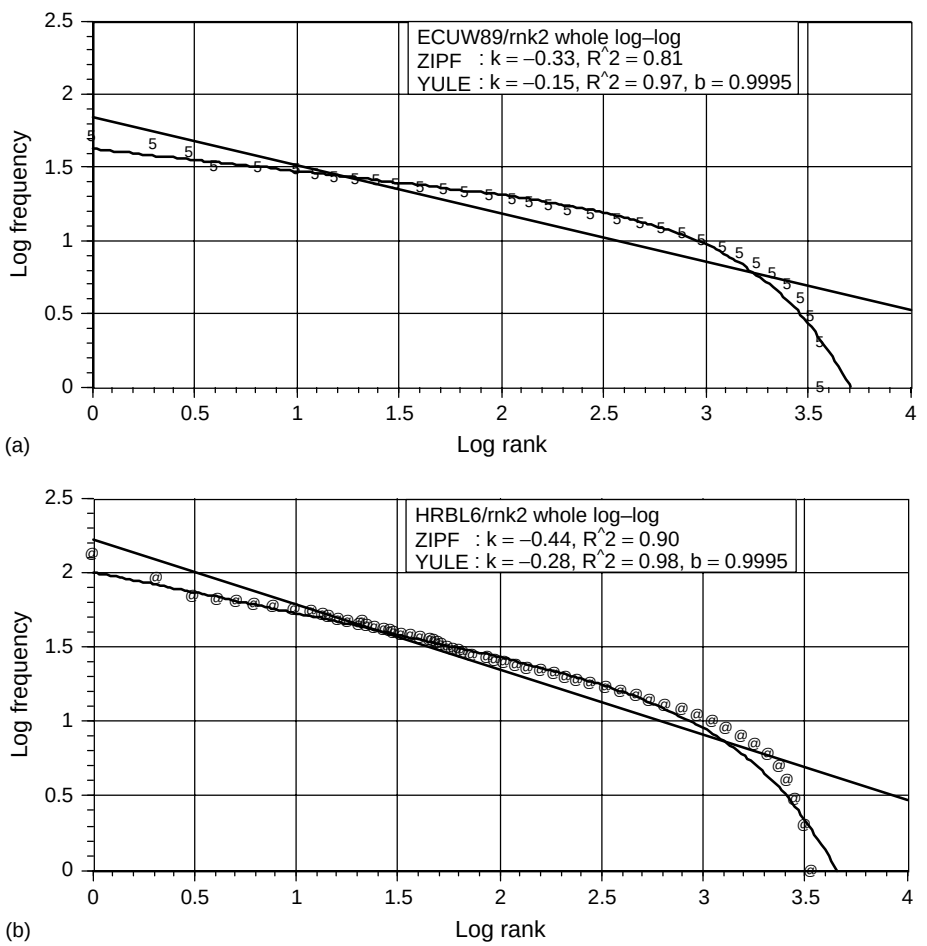

Fig. 3 Ranked hexanucleotide frequencies fit the Yule distribution (convex lines on Figures (a) and (b)) much better than the Zipf distribution (straight lines on Figures (a) and (b))
(a) ECOUW89 is a long sequence (176,195 nt) in which 82.1% of nucleotides constitute protein-coding regions. (b) HUMRETBLAS is a long sequence (180,388 nt) in which only 1.5% of nucleotides belong to exons [Based on: Martindale, C. Konopka, A.K., (1996) *Comput. Chem.*, **20**(1), 35–38].

3 Measuring Deviations from Discrete Uniform Distribution (DUD)

3.1 Entropy and Redundancy of Nucleotide Sequences

The mathematical apparatus of Shannon's Theory of Communication ("Information" Theory) has been applied to biology quite against Shannon's preferences. In 1954, he dismissed psychology and biology as fields of science that could benefit from applying his theory because he believed that semantic aspects of communication are more important in these fields than syntactic aspects. Of course, he was unaware of the important role that statistical studies of nucleotide and protein sequences would play in biology. Today, we know that indeed communication theory does not explain much of biology (nor did anyone sane ever expect it to do so) but, on the other hand, it is an extraordinary source of statistical methods for studying sequences.

Shannon's studies of k-grams of letters in printed alphabetic languages have been a spectacular application of his theory of communication ("information" theory). The central concept in these studies is an abstract device (called *source*) that generates sequences of symbols (messages) chosen from a finite alphabet by means of a random mechanism. Particularly important are ergodic sources in which the random mechanism leads to "typical" messages with high probability (close to 1) and to "atypical" sequences of symbols with negligible probability. (We can say, in other words, that sequences produced by an ergodic source are statistically homogeneous.)

It has long been known that frequency distributions of letters are roughly the same in all texts written in the same language. It has then been assumed (but not proved) that any alphabetic writing system can be modeled by an ergodic source. (In particular, it has been suggested that a writing system can be regarded as a limit of a succession of ergodic Markov sources that in turn is itself an ergodic source.)

The probability distribution of letters from an alphabet of size n in a given ensemble of texts can be characterized by a function called *entropy* (of the distribution):

$$H_1 = -\sum_{i=1}^{n} p(i) \cdot \log_r p(i) \qquad (7)$$

where $p(i)$ are probabilities of letters and the log is taken to the base $r = 2$ (unless otherwise specified) and a convention $0 * \log_r(0) = 0$ is applied.

If textual elements (letters) from a finite alphabet of size n were combined in a text by chance alone, it would have to be a Bernoulli text. Therefore, the letters should occur independently of each other and with probabilities following a DUD (i.e. equal to $1/n$). The entropy of DUD obviously equals $log_r(n)$, that is:

$$H_0 = -n \cdot \frac{1}{n} \cdot \log_r \frac{1}{n} = \log_r n \qquad (8)$$

The *entropy of strings of length k* composed of characters from an (elementary) alphabet of size n is defined as

$$H^{(k)} = -\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_k=1}^{n} p(i_1, i_2, \ldots, i_k) \cdot \log_r p(i_1, i_2, \ldots, i_k) \qquad (9)$$

The *k-gram entropy* is defined as

$$H_k = H^{(k)} - H^{(k-1)} \quad \text{for} \quad k > 1 \qquad (10)$$

with $H^{(1)} = H_1$ given by (7) and H_0 given by (8).

Note that entropy has a dimension that is determined by the basis r of logarithmic function and by the alphabet taken into account. In the general case, when the basis of log function is r, the k-gram entropy H_k is given in r-ary units per letter and the strings of length k entropy is given in r-ary units per k-gram. The most frequently used value of r is 2. In this case, $H^{(k)}$ is in bits per k-gram and H_k in bits per letter ("bit" is a name of 2-ary unit).

A *source* S has entropy $H(S)$ if

$$\lim_{k\to\infty} \frac{H^{(k)}}{k} \tag{11}$$

exists and equals $H(S)$.

The ergodic sources always have entropy. A given writing system (sometimes referred to as *a language*) L is assumed to be an ergodic source (a limit of a succession of ergodic Markov sources). Hence, the writing system's entropy $H(L)$ exists and is bounded by the following (obvious) chain of inequalities:

$$H(L) < \cdots < H_k < \cdots < H_3 < H_2 < H_1 < H_0 \tag{12}$$

The ratio of entropy of source to the maximum value it could have while still restricted to the same symbols is called *relative entropy* (of source). Since it follows from (12) that H_0 is the maximum value of entropy, we have:

$$H(L)' = \frac{H(L)}{H_0} \tag{13}$$

In a similar way, we can define *relative k-gram entropy*:

$$H_k' = \frac{H_k}{H_0} \tag{14}$$

Values of relative entropy are dimensionless real numbers form the interval [0, 1].

Writing system's (language) *redundancy* is defined as:

$$R(L) = 1 - H(L)' \tag{15}$$

Similarly, we can define k-gram redundancy as:

$$R_k = 1 - H_k' \tag{16}$$

It follows from (12) that $R(L)$ is higher than any k-gram redundancy, that is:

$$0 < R_1 < R_2 < \cdots < R_k < \cdots < R(L) \tag{17}$$

Because of high sampling errors of k-gram entropies in finite-length sequences, it is not obvious that the ensembles of functionally equivalent nucleotide sequences have entropy. Dedicated computational studies will be needed in the future to determine how well k-gram entropies converge to the source entropy values. For the time being, we assume that functionally equivalent sequences do have entropy and that "functionality" can be thought of as an ergodic source.

Table 2 shows the results of redundancy calculation for several large collections of functionally equivalent nucleotide sequences. As expected the sequences of eukaryotic illegitimate recombination regions display very high redundancy (around 20%). This is due to the frequent occurrence of mononucleotide repeats. Interestingly, the lowest redundancy can be seen for protein-coding genes and the highest for intergenic spacers and flanking regions of genes. However, because of very high error values, redundancy in exons and the corresponding introns appears to be the same even at the level

Tab. 2 Estimated redundancy in various large collections of functionally equivalent nucleotide sequences. The estimates for $k \to \infty$ were obtained from regression analysis of 1- through 4-grams' or 1- through 5-grams' redundancies (depending on the length of sequences in a given collection).The error values listed are standard errors of estimate for hyperbolic regression. (Based on: Konopka, A.K. (1994) Sequences and Codes: Fundamentals of Biomolecular Cryptology, in: Smith, D. (Ed.) *Biocomputing: Informatics and Genome Projects,* Academic Press, San Diego, CA, pp. 119–174.)

	Redundancy
Human exons (in mature mRNAs)	0.070 ± 0.009
Human introns	0.082 ± 0.017
Human 3′UTRs	0.081 ± 0.020
Human 5′UTRs	0.083 ± 0.022
Human 3′ flanks of protein-coding genes	0.160 ± 0.034
Human 5′ flanks of protein-coding genes	0.122 ± 0.025
Eukaryotic exons (in mature mRNAs)	0.076 ± 0.021
Eukaryotic introns	0.087 ± 0.022
Eukaryotic Intergenic Spacers (non-RNA, nonhistone)	0.112 ± 0.007
Eukaryotic Intergenic Spacers (histone)	0.142 ± 0.027
Eukaryotic Illegitimate Recombination Regions	0.202 ± 0.032
Bacterial protein-coding genes	0.075 ± 0.023
Bacterial spacers (RNA genes)	0.090 ± 0.023
18s-rRNA genes	0.095 ± 0.040
Small nuclear RNAs	0.127 ± 0.032
16s-rRNA genes	0.118 ± 0.038
5s-rRNAs	0.092 ± 0.028
tRNAs	0.155 ± 0.041
23s-RNAs	0.127 ± 0.041
28s-RNAs	0.169 ± 0.024

of confidence of 33%. This precludes using the "function-associated" redundancy as a criterion for determining exon and intron location in unannotated sequences. This also means that redundancy is unlikely to be a basis for a classification code.

3.2 Coincidence Indices and Their Use in Sequence Analysis

pt For a discrete probability distribution $[p(1), p(2), \ldots, p(n)]$ over an ensemble of n events the *index of coincidence* can be defined as:

$$I = n \cdot \sum_{i=1}^{n} p^2(i) \tag{18}$$

For the text of finite length L, index of coincidence can be estimated as:

$$I = \frac{n \cdot \sum_{i=1}^{n} f(i) \cdot [f(i) - 1]}{M \cdot (M - 1)} \tag{19}$$

where $f(i)$ is a frequency of ith element of the alphabet n is a size of the alphabet and M is the number of elements of the alphabet present in the text (for single letters $M = L$, for strings of 2

letters $M = L - 1$ for strings of k letters $M = L - k + 1$).

The variance of I has been found to be

$$\mathrm{Var}(I) = \frac{2I \cdot (n - I)}{M \cdot (M - 1)} \tag{20}$$

In statistical linguistics and cryptanalysis tests based on coincidence, indices are reliable tools to determine substitution block encodings. They work particularly well in the cases in which each string of the code alphabet has the same length as other strings. It also helps if the size of (encoded) "space bars" is negligible compared to the size of (also encoded) "message units".

Classification codes for different ensembles of functionally equivalent sequences can be determined by the two tests based on coincidence indices

Both, *string repetition test* and *length test*, apply to the following situation: Given a sequence of length L represented in an elementary alphabet E = {A, C, G, T}, consider m code alphabets containing $N(k) = 4^k$ strings each ($k = 1, 2, \ldots, m$ and $m \ll L$). Choose "the best" code alphabet (to represent a given sequence) from the m alphabets given.

In the string repetition test we pick the k-th alphabet and calculate coincidence index (19) along with its variance (20) in our sequence of length L (i.e. containing $M = L - k + 1$ "overlapping" or $M' = \mathrm{Int}\{L/k\}$ "nonoverlapping" strings belonging to the alphabet of $N(k)$ elements). For a Bernoulli text, the k-gram probability distribution is a DUD and therefore $I_{\text{expected}} = 1$. The z-score of I in a given alphabet is:

$$z(I) = \frac{I - 1}{\sqrt{\mathrm{Var}(I)}} \tag{21}$$

After calculating $z(I)$ for each alphabet, we consider those alphabets that maximize (21) to be "the best".

In the string length test, mean coincidence indices (per letter) in k-grams over an n-letter alphabet are compared with the coincidence index over 1-grams in *entire* sequence. We pick the kth alphabet and calculate the coincidence index for all letters at all positions $j = 1, 2, \ldots, k$ within each k-gram in a given sequence, that is:

$$I(j) = \frac{n}{M \cdot (M - 1)} \cdot \sum_{i=1}^{n} f(i, j) \cdot [f(i, j) - 1] \tag{22}$$

where $f(i, j)$ is the frequency of ith letter (out of n) at the jth position of a k-gram. Of course, for nucleotide sequences expressed in {A,C,G,T/U) alphabet we have $n = 4$, while for binary alphabets such as {R,Y} or {K,M}, n equals 2.

The above calculations provide us with a vector $\boldsymbol{J} = [I(1), I(2), \ldots, I(k)]$ from which we can determine the mean coincidence index $\langle I \rangle$, that is:

$$\langle I \rangle = \frac{1}{k} \cdot \sum_{j=1}^{k} I(j) \tag{23}$$

The expected value of $\langle I \rangle$ for nucleotide sequences in {A, C, G, T} alphabet can be estimated from (19) with $M = L$ (number of 1-letter subsequences equals the length of sequence) and $n = 4$, that is,:

$$\mathrm{E}(\langle I \rangle) = \frac{4 \cdot \sum_{i=1}^{4} f(i) \cdot [f(i) - 1]}{L \cdot (L - 1)} \tag{24}$$

The corresponding variance is obtained from (20) by again putting $M = L$ and $n = 4$ (if we use {A, C, G, T/U} alphabet; for binary alphabets such as {R, Y} we would need to put $n = 2$ and $M = L$), that is:

$$\mathrm{Var}(\langle I \rangle) = \frac{2I \cdot (4 - I)}{L \cdot (L - 1)} \tag{25}$$

The z-score for this test is then:

$$z(\langle I \rangle) = \frac{\langle I \rangle - \mathrm{E}(\langle I \rangle)}{\sqrt{\mathrm{Var}(\langle I \rangle)}} \qquad (26)$$

After calculating $z(\langle I \rangle)$ for each alphabet, we consider the alphabet that maximizes (26) to be "the best."

Both tests indicate unambiguously (see Table 3) that the trinucleotide code is the best to represent protein-coding regions in nucleotide sequences. This result is remarkable because it leads to a rediscovery of codons of the genetic code solely on the basis of the statistical structure of sequences. This means that knowledge of the protein-coding function of genes is not required to find genes among other sequences. Needless to say, methodology of this kind can be used for practical purposes of finding exons in unannotated sequences.

In fact, about 98.5% of known translated regions of genes (exons and bacterial genes) display excessively high, positive values of $z(\langle I \rangle)$ for the alphabets of 3-grams, 6-grams, 9-grams, and generally, $3m$-grams ($m = 1, 2, 3, \ldots, 20$). In addition, the large fractions of remaining 1.5% are the exons that are subject to alternative splicing (i.e. are also parts of introns).

As far as sequences that are not translated are concerned, the results of string length test are less clear. For example, in a large sample of nuclear eukaryotic introns, about 20% of the sequences test positively for the code word's length 2 [i.e. $2m$-grams ($m = 1, 2, 3, \ldots$) are unambiguously indicated by the test (26); see Table 3]. Another 3% of intron sequences display clear preference for strings of length 4. Both these results are consistent with the 2-base periodicities in large collections of intron sequences (see Sect. 5). However, the test (26) also indicates that about 7% of intron sequences display a clear preference for strings of length 5 and there exist introns (only about 1.3% in the sample collection) that "favor" heptanucleotides.

There are at least three plausible interpretations of the above findings. One possibility is that there are many different kinds of introns playing different biological roles before being spliced out of hnRNAs. Another possibility is that the "intronic" code alphabet contains strings of variable length. In such a case, the string length test would detect the mean length of a code word but not the actual lengths of individual code words. Finally, it may well be that only fragments of introns are involved in some (unknown) biological processes. In such a case, long stretches of sequence within introns would not code for any function and therefore there would be no reason to "employ" a statistically distinguishable k-grams' alphabet. This last situation would be similar to a text in which long strings of space bars separate individual words (or even letters) and many different symbols would be allowed to stand for a space bar.

As far as viral introns are concerned they display conspicuously high z-scores for string lengths multiples of 3 (examples are listed in Table 3), as do many mitochondrial introns (data not shown). This "exonic" characteristics is quite unexpected in introns but can be easily explained by the fact that viral genomes contain overlapping genes. A given viral intron can be an intron in one gene and a part of exon in another (overlapping) gene. Therefore, the "exonic" characteristics (in the form of preference for strings of lengths multiple of 3) of viral introns can be detected. Similarly, those mitochondrial introns that display "exonic" characteristics (code word lengths 3, 6, 9, ...) overlap

Tab. 3 Examples of *z*-scores calculated from the string length test [formula (26) in the text] for sequences belonging to different functional classes. Only string lengths between 2 and 9 are listed. (Calculations based on: Shulman, M.J., Steinberg, C.M., Westmoreland, N. (1981) The coding function of nucleotide sequences can be discerned by statistical analysis, *J. Theor. Biol.* **88**, 409–420.)

	Length of k-strings (i.e. k)							
	2	***3***	***4***	***5***	***6***	***7***	***8***	***9***
	Typical 2-base periodic IVS (~20% of all IVS)							
>DL;FOSMS	2.1	−0.4	3.5	−0.2	7.2	2.4	4.9	−3.4
>DL;TVMSP1	2.7	−1.6	2.9	−3.1	6.1	−1.5	0.2	−5.3
>DL;HVMS3	7.5	0.1	7.0	−1.8	7.5	−1.0	7.1	−4.3
>DL;HVMS44	3.4	1.8	4.8	−3.6	5.0	−0.7	3.3	5.7
>DL;IGCMD	3.7	1.6	2.3	0.6	4.7	−0.2	4.2	−0.2
>DL;G2MSA	4.1	0.3	5.7	−2.5	5.4	−0.6	4.3	5.8
>DL;HLMSE1	2.0	0.0	6.1	0.1	1.5	−0.2	3.7	−3.0
>DL;CELDEB	2.7	−0.7	4.0	−3.0	0.6	0.6	7.3	−0.5
>DL;DROGLD	3.9	1.8	3.6	1.1	8.1	3.9	2.7	−1.0
>DL;DROLGL	4.2	−1.0	6.3	−0.5	2.3	−0.4	5.3	−1.9
>DL;DROMHC	3.9	0.1	3.6	−0.7	4.1	−2.6	3.2	−2.3
	Typical translated regions of genes ("true" coding regions)							
>DL;HPHU2R	−0.2	29.9	−0.5	1.4	29.1	2.5	2.8	32.5
>DL;KFHUG	−1.0	26.8	−0.3	0.3	23.8	1.2	−1.7	26.3
>DL;KXHU	1.6	71.0	4.4	3.8	73.9	−3.9	5.1	71.4
>DL;UBHUB	1.2	95.5	−0.3	−2.5	98.2	−0.6	−1.6	91.1
>DL;LPHU4A	−0.9	106.6	0.8	−1.3	106.8	−3.6	0.5	106.1
>DL;FGHUGG	1.4	17.6	0.6	−1.1	20.1	2.1	−3.6	19.9
>DL;APOCHO	4.5	151.0	2.9	−2.9	176.1	−1.1	1.0	229.9
>DL;BBO58M	0.0	51.2	0.4	−0.5	52.3	−1.4	0.2	47.3
>DL;BBOMER	−0.6	43.9	−1.5	−2.5	44.4	−1.6	−0.1	43.9
>DL;DROCID	−1.0	59.8	−1.3	0.0	59.0	−0.8	−3.4	61.5
>DL;DROCOP	0.1	175.9	−0.6	−2.9	179.8	−2.7	3.1	175.7
>DL;DROCOP	−1.1	62.5	1.2	9.2	62.7	2.7	0.6	63.6
	Typical viral translated regions of genes							
>DL;ADLEIB	−0.6	22.3	−1.9	−2.0	20.6	4.6	−0.7	28.0
>DL;ADTA$_1$	−0.1	10.0	−0.8	0.4	9.2	0.8	−0.8	10.9
>DL;ADV40E	0.7	7.2	0.3	−1.6	6.8	−2.1	−1.5	3.5
>DL;ADVFIB	0.2	26.1	0.0	−0.1	24.9	2.4	0.8	25.9
>DL;ADX1AB	0.2	12.6	−1.1	−2.7	11.2	−0.5	−2.1	13.7
	Typical viral IVS							
>DL;ADLEIB	0.0	19.1	−1.9	−0.2	18.1	4.4	0.1	22.9
>DL;ADX1AB	−0.1	10.3	−1.1	−2.4	9.4	−1.7	−3.7	9.8
>DL;ALRCG_	−1.0	10.0	−1.8	−0.9	9.3	−2.3	−1.9	10.0
>DL;HIVBH1	−0.8	20.6	−0.9	−2.7	20.0	−3.0	0.6	24.0
>DL;PPH47C	0.0	12.6	−1.2	1.2	13.9	−1.5	−0.3	11.2

with genes that encode the endonuclease responsible for their splicing.

4 Noncontiguous Patterns and Distance Charts

Distributional structures in nucleotide sequences from the same functional class (as introns, exons, tRNAs, 5sRNA, illegitimate recombination sites, and so on) can be seen on k-gram distance charts. If there were no distributional structures in sequences studied, there would be no particular reason for any punctuation length to occur more frequently than other lengths. In other words, with no distributional structures, punctuation lengths should follow a discrete, uniform frequency distribution. When the distributional structures are present, some punctuation lengths should be much more frequent and some others much more rare than expected from discrete, uniform distribution.

Similar reasoning applies to the shortest distance charts. In the absence of distributional structures, the frequency of punctuation lengths in the shortest distance charts would follow geometric distribution. Again, when distributional structures are present, some punctuation lengths will occur more frequently and some others less frequently than expected from geometric distribution.

Distance charts (z-scores with regard to discrete, uniform distribution) for motif AC in large collections of nuclear introns and exons from animal genomes (primates, rodents, other mammals, vertebrates, and invertebrates) are shown in Fig. 4(a). The charts for an internally mirror-symmetric motif ACA appears in Fig. 4(b). The corresponding shortest distance charts (z-scores with regard to the appropriate geometric distribution) are shown in Figs. 4(c and d). It is clearly seen from Figs. 4(a–d) that exon and intron sequences display k-gram distributional structures and that these structures are transparently different for exons than for introns. The former display a strong overrepresentation of punctuation lengths 0, 3, 6, ... and generally $3k$ ($k = 0, 1, 2, 3, \ldots$) for trinucleotide ACA whereas the later favor distances $0, 2, 4, \ldots$ and generally $2k$ ($k = 0, 1, 2, \ldots$) for dinucleotide AC. Inspection of shortest distances (Figs. 4(c and d) shows that preferred nearest distances between ACs are 0 and 2 in introns. However, exon preferred shortest distances between ACAs (Fig. 4d) are the same as distances from Fig. 4(b) (i.e. 0, 3, 6, 9, ... and generally $3k$). These findings suggest that AC is involved in "true" 2-base periodic patterns in introns whereas ACA is involved in strongly pronounced 3-base quasi-periodic patterns in exons.

Other than AC, nonhomopolymeric dinucleotides and the vast majority of nonhomopolymeric (including mirror-symmetric) trinucleotides display distance charts almost identical to those shown in Fig. 4 (data not shown). This suggests that predominant patterns in exons and introns are 3-base quasi-periodic arrangements of trinucleotides and 2-base periodic repeats of dinucleotides respectively.

Notably, distance charts from the above example allowed us to rediscover the triplet structure of the genetic code from exon sequences alone (i.e. without taking into account our knowledge of the genetic code). In the light of this rediscovery, 2-base periodicities observed in introns must reflect meaningful, intron-specific distributional structures.

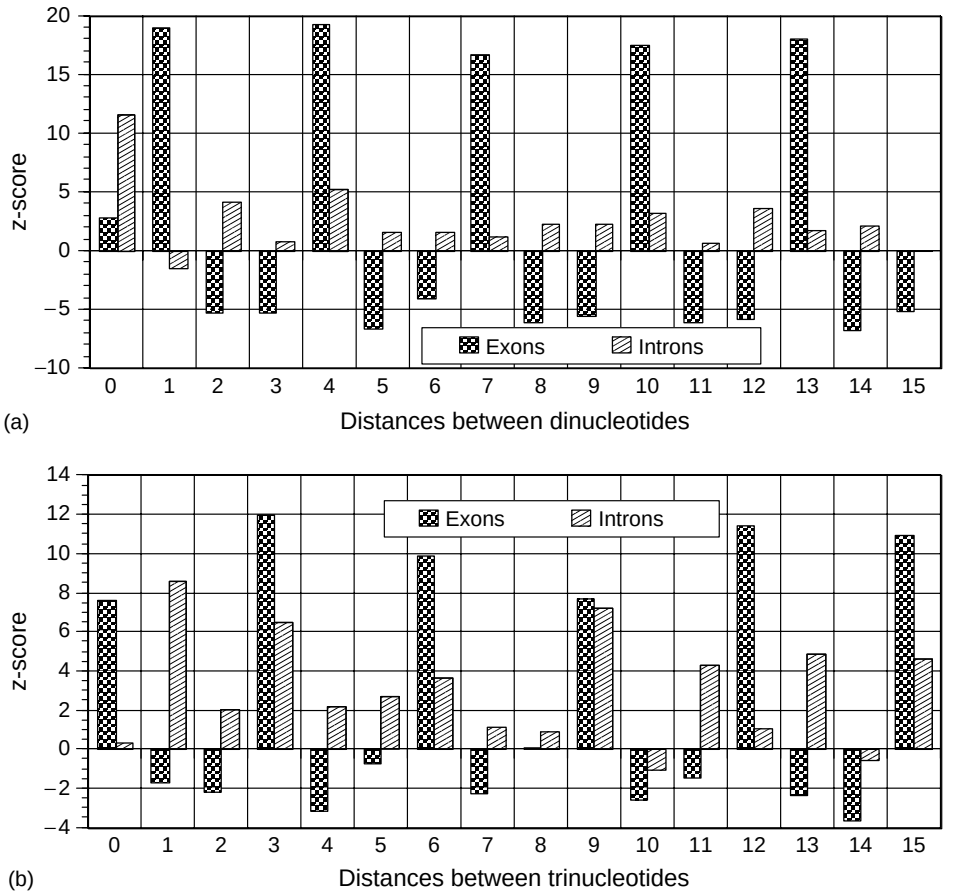

Fig. 4 Examples of distance charts for nonhomopolymeric dinucleotides and the corresponding mirror-symmetric trinucleotides in large corpora of animal, nuclear exons and introns. (Based on and modified from: Konopka, A.K., Smythers, G.W. (1987b) *Comput. Appl. Biosci.* **3,** 193–201.) (a) – Dinucleotides tend to be separated by gaps of lengths divisible by 2 in introns. In exons, we can see a tendency of those gaps lengths to follow an arithmetic progression with increment 3. This leads to predominant distances of 1, 4, 7, 10, . . . and generally $1 + 3k$ ($k = 0, 1, 2, \ldots$). (b) – Trinucleotides (including the mirror-symmetric ones) tend to be separated by gaps of lengths divisible by 3 in exons. (c) – Shortest distance chart displays the same predominant dinucleotide distances in exons as the distance chart (Fig. 2a). However, the predominant shortest distance for dinucleotides in introns is 0 nucleotides. This means that 2-base periodicity implied by distance chart is a ''true'' periodicity (see the text and Fig. 1). (d) – Shortest distance chart for trinucleotides in exons displays the same, divisible by 3, predominant distances as the distance chart from Fig. 2b. The 3-base periodicity implied by Fig. 2b is not a ''true'' periodicity but a 3-base quasi periodicity (see the text and Fig. 1).

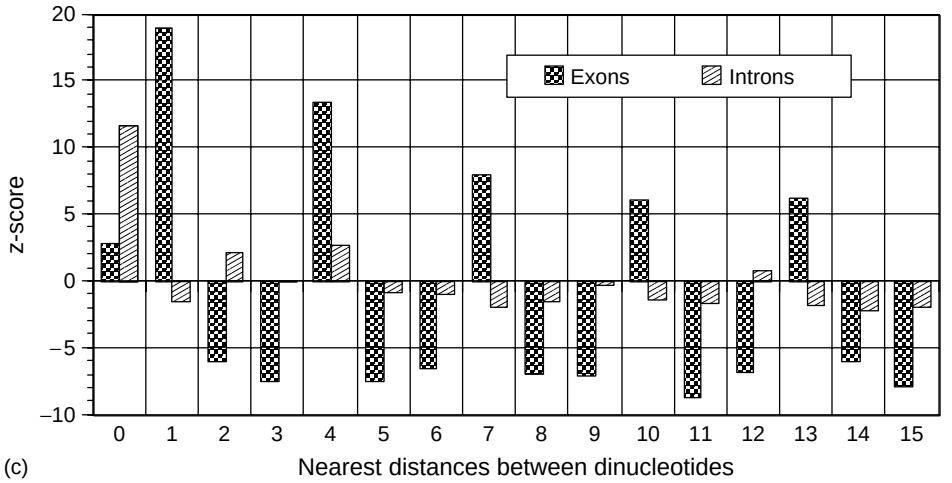

(c)

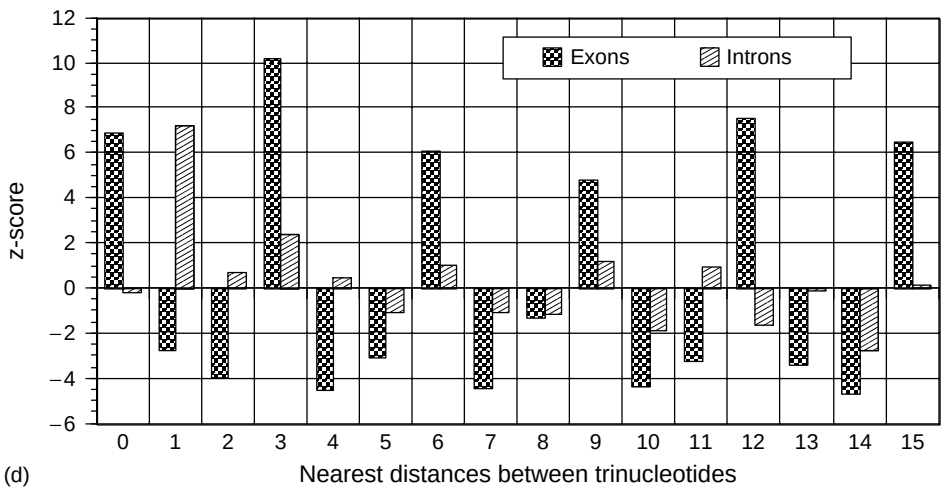

(d)

Fig. 4 (*Continued*)

5 Local Compositional Complexity: A Maximum Entropy Principle

5.1 Local Compositional Complexity

Local compositional complexity can be defined by two different measures. Both of these measures are based on a concept of *repetition vector* (sometimes called *complexity state vector*). For a string of length N, the repetition vector is:

$$\tilde{\boldsymbol{n}} = (n_1, n_2, \ldots, n_L) \tag{27}$$

where L is the number of letters in the alphabet, and n_i is the number of occurrences of letter i in the string (i.e. $n_1 + n_2 + \cdots + n_L = N$). The elements of the repetition vector are ordered according to size, so $n_1 \geq n_2 \geq n_3 \ldots \geq n_L$. For example, the string ACTGAAGT has the repetition vector (3,2,2,1). The repetition

vector is not dependent on the order of the letters, so any permutation of a given string will have the same repetition vector, for example, AAATTGGC also has the repetition vector (3,2,2,1). Similarly, any permutation of the alphabet yields a string with the same repetition vector, for example, TGACTTCA has the repetition vector (3,2,2,1).

One measure of compositional complexity is:

$$\mathcal{K}_1 = (1/N) * \log_r \left(N! \Big/ \prod_{i=1}^{L} n_i! \right) \quad (28)$$

where L, N, and n_i are defined as above and r equals to L. For example, for a DNA fragment, one alphabet would be the possible nucleotide bases in the DNA, that is, adenine, cytosine, guanine, and thymine, or A, C, G, T. In this case, L and r would be 4, and the nis would be the numbers of As, Cs, Gs, and Ts in the fragment.

Another measure of complexity is based on Shannon entropy:

$$\mathcal{K}_2 = -\sum_{i=1}^{L} [((n_i/N) * \log_r(n_i/N))] \quad (29)$$

where n_i, N, L, and r are the same as above.

Although $\mathcal{K}_1$ and $\mathcal{K}_2$ converge as N approaches infinity, for short oligonucleotides they differ significantly.

5.2 Surprisal

We can define a *complexity class* as a set of strings that have the same repetition vector, and therefore the same complexity. For a string of length $S \gg N$ we can calculate the *surprisal*, or *log-odds ratio*, for each such class.

Surprisal (log-odds ratio) is defined as:

$$\mathcal{S}(\tilde{\boldsymbol{n}}) = \log_r [P_{\text{observed}}(\tilde{\boldsymbol{n}}) / P_{\text{expected}}(\tilde{\boldsymbol{n}})] \quad (30)$$

where $\tilde{\boldsymbol{n}}$ is the repetition vector $(n_1, n_2, \ldots, n_L)$, P_{observed} is the observed probability of the repetition vector n in the string, and P_{expected} is the *a priori* probability of the repetition vector $\tilde{\boldsymbol{n}}$ in the string. The expected probability for a given repetition vector is the sum of probabilities of each of the distinct strings that share that repetition vector. Assuming that each of the letters is equally probable, then each of these strings has equal probability, and the probability for any repetition vector class is simply the number of different strings that share that repetition vector divided by the number of possible strings. For example, for a DNA fragment and the alphabet ACGT, the total number of possible strings is L^N, the number of permutations of a given string is

$$\mathrm{w}(\tilde{\boldsymbol{n}}) = N! \Big/ \prod_{i=1}^{L} n_i! \quad (31)$$

and the number of distinct permutations of the alphabet for a given repetition vector is

$$\mathrm{F}(\tilde{\boldsymbol{n}}) = L! \Big/ \prod_{i=1}^{j} h_i! \quad (32)$$

where j is the number of distinct values among the elements of the repetition vector n, and h_i is the number of elements in $\tilde{\boldsymbol{n}}$ that have the jth value. For example, for $\tilde{\boldsymbol{n}} = (3, 2, 2, 1), j = 3, h_1 = 1$ because there is one element with the value 3, $h_2 = 2$ because there are two elements with the value 2, and $h_3 = 1$ because there is one element with the value 1. Note that $L!$ is the number of permutations of the alphabet, and it is divided to eliminate redundancy.

Thus, the *a priori* probability for $\tilde{\boldsymbol{n}}$ is:

$$P_{\text{expected}}(\tilde{\boldsymbol{n}}) = [\mathrm{F}(\tilde{\boldsymbol{n}}) * \mathrm{w}(\tilde{\boldsymbol{n}})]/L^N \qquad (33)$$

The probability distribution generated from the assumption of equal probabilities for each of the letters in the alphabet we call P_0.

5.3 Log-odds Complexity Distribution

All collections of functionally equivalent nucleotide sequences studied thus far display a linear tendency in surprisal versus complexity distributions (see two examples in Fig. 5). The slope of this linear relation is negative in all cases studied. This is interpreted as a *maximum entropy relationship* constrained by mean complexity.

It is likely that this maximum entropy relation is a general law governing all natural nucleotide sequences. (Preliminary evidence exists that protein sequences "obey" this principle as well.) On the other hand, sequences belonging to different putative functional domains display systematically different slope values. Discriminative power of the slopes is much stronger than the criteria based on sequence periodicity or uneven distribution of mononucleotides between three positions in codons. This suggests that slope values can be used as reliable classification code words.

5.4 Patchy Complexity

Strings of symbols used to determine repetition vector $\tilde{\boldsymbol{n}}$ can be generated by selecting noncontiguous nucleotides from a given (long) sequence. We shall discuss complexity of strings of L symbols obtained by taking every kth nucleotide, $k = 1, 2, \ldots, 20$, from original nucleotide sequences. *Patchy oligonucleotides* defined in this way are characterized by patchiness, k and block length, L. For instance, "normal" (i.e. nonpatchy) hexanucleotides are characterized by $k = 0$ and $L = 6$ whereas hexanucleotides obtained by concatenating (six times) every other nucleotide in a given sequence are described by $k = 1$ and $L = 6$. Similarly, octanucleotides generated by taking 8 times every third nucleotide from the original sequence are characterized by $k = 2$ and $L = 8$ but "normal" octanucleotide are of patchiness $k = 0$.

We refer to complexity calculated over repetition vectors of patchy oligonucleotides as to *patchy complexity*. Of course, the "normal" (i.e. nonpatchy) local compositional complexity is a special case of patchy complexity.

All standard results obtained for local compositional complexity of nucleotide sequences hold for patchy complexities. In particular:

1. For every value of patchiness k, the linear relation between complexity of patchy L-nucleotides and surprisal is observed. The slope values are negative in majority of cases studied.
2. The larger block length L is the more negative the slope of surprisal versus complexity relationship is (for a given k).

Examples of the relationship between slope values and patchiness for the block lengths $3 \leq L \leq 13$ in {A, C, G, T/U} alphabet are shown in Figs. 6(a–d). It can be seen in Fig. 6(a) that the relationship is clearly a modulo-3 periodic function of k for exons (minimum slope values for $k = 2$, 5 and 8; a clear period of 3) but this is not the case for introns (clear minimum at $k = 1$ for $L > 4$ as well as weakly pronounced minima at $k = 1, 3, 5,$

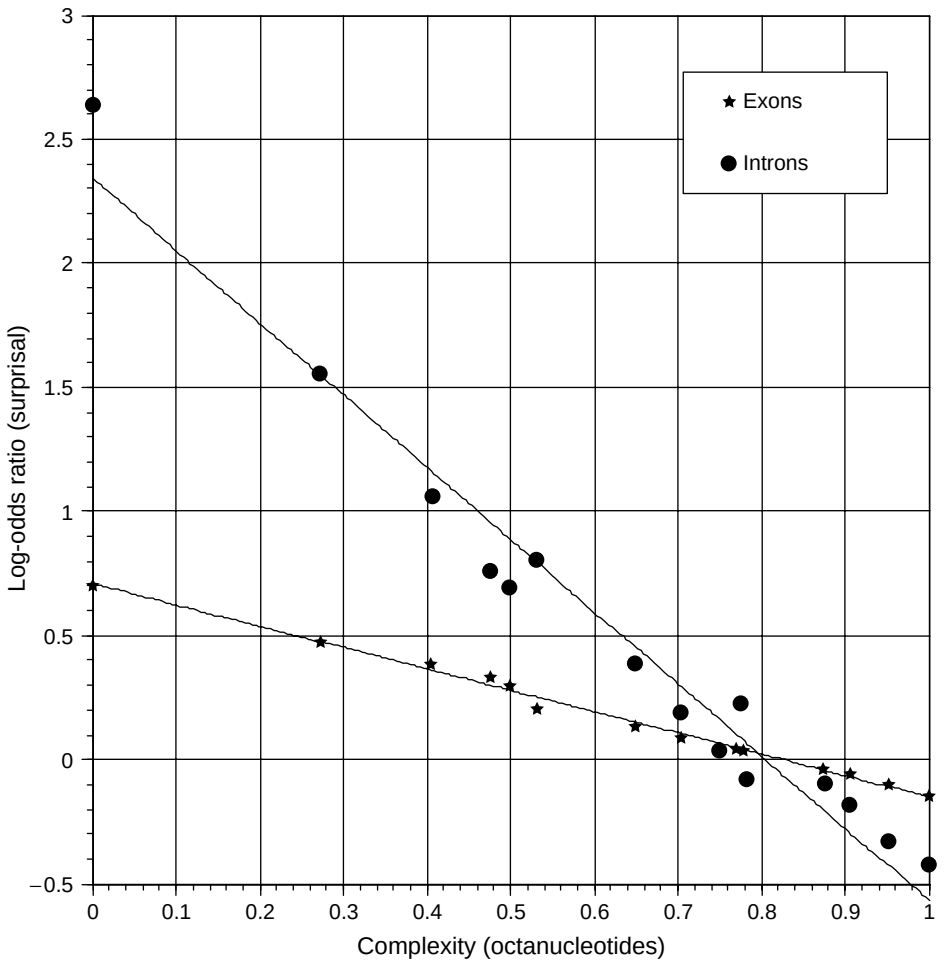

Fig. 5 Example of maximum entropy relationship between complexity and surprisal (see the text) for octanucleotides in invertebrate introns and exons. All naturally occurring nucleotide sequences studied thus far (over 30 large corpora of sequences) display a linear trend (with negative slope) for tetra- through octanucleotides. (Based on Salamon, P., Konopka, A.K. (1992) *Comput. Chem.* **16,** 117–124.)

and 7, which correspond to the well-known weak period of 2; Fig. 6b). Slopes for 3′UTR (Fig. 6c) are an increasing function of k with a very weakly pronounced minima at $k = 5$ and $k = 7$ for large L. The 5′UTRs (Fig. 6d) display a weakly pronounced period of 3 but generally lower slope values than exons. The difference between slope values for different block lengths L appears to be much more dramatic than in the case of exons.

6 Computer-assisted Systems Biology

With the passage of time, most of today's molecular biology seems to have

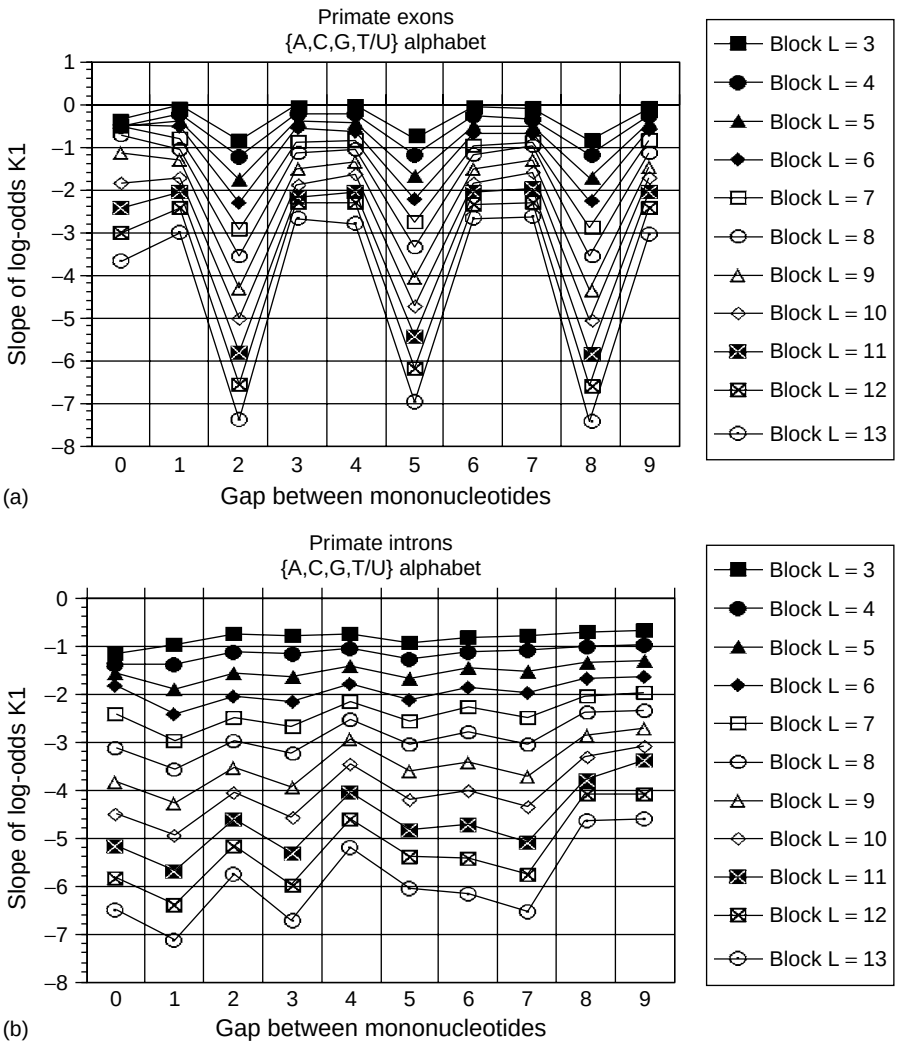

Fig. 6 Slopes of log-odds patchy complexity distributions in (a) – Primate exons (b) – Primate introns (c) – Human 3′ untranslated regions of protein-coding genes (3′-UTRs) (d) – Human 5′ untranslated regions of protein-coding genes (5′-UTRs) (Based on unpublished work by Konopka, A.K., Felts, B.E., and Hand, C.; Extended Abstracts – Open Problems of Computational Molecular Biology (4), Telluride, CO, (1994) and Notes – Gene-Finding and Gene Structure Prediction workshop, Philadelphia, PA (1995).)

been transformed into biotechnology. On the other hand, the scientific remainder of molecular biology is in a desperate need of reconciliation with the rest of biology and with other fields of science. As far as computational molecular biology is concerned, research agendas that pertain to integration of sequence analysis with biology proper involve conceptual foundations of life sciences, methods

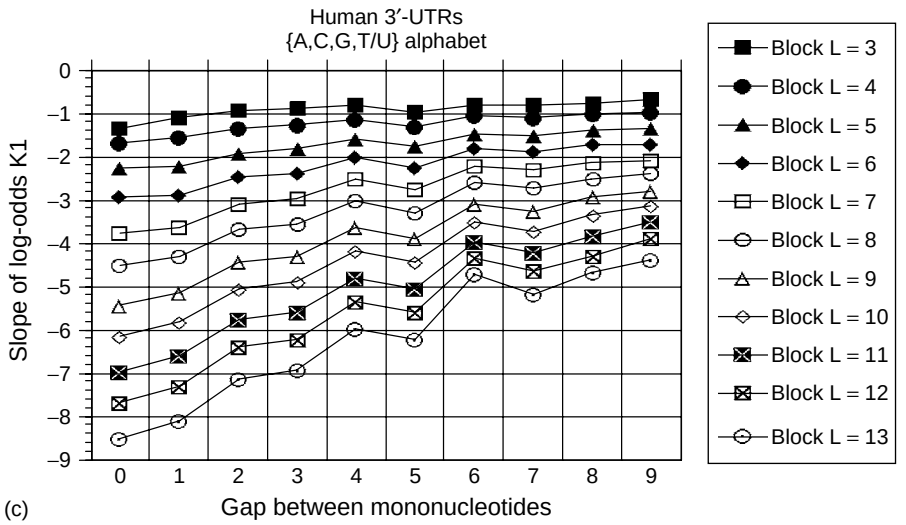

(c)

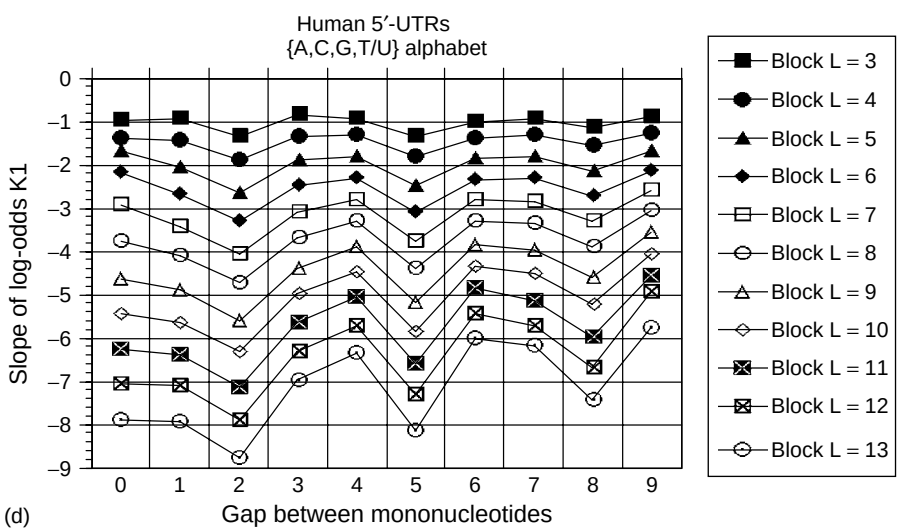

(d)

Fig. 6 (*Continued*)

of modeling biological systems, methods and concepts of data integration, as well as computer-assisted techniques of pragmatic inference. In this section, I briefly discuss selected general topics that seem to be representative of future research in computer-assisted systems biology.

6.1 Pragmatic and Semantic Theories of Information

The existing (Hartley-Shannon) theory of communication has been created for telecommunication systems and digital computers. Despite numerous proposals

stating otherwise, it does not seem to provide much new insight into molecular biology (it can be a useful source of statistical methods though). On the other hand, concepts of "information," "meaning," "function," and "organization" seem to be integral parts of biology. If we are to comprehend the phenomenon of life despite our language limitations, we need to develop a theory that could handle these concepts. The need for a theory of information pertinent to biology has been already postulated in neurobiology and in developmental biology. Although such a need has not troubled molecular biologists quite yet, it begins to emerge as evidence accumulates for the role of "epigenetic information" in controlling gene expression and regulation.

6.2 Logic of Heuristic Reasoning

As most of us know, heuristic reasoning can be conclusive. Yet biologists face the problem of not being able to specify all the rules applied to derive conclusions. Nor are they able to list all assumptions on which those rules ought to operate. It seems that these inabilities are a reflection of the complexity of biological systems themselves. Biological phenomena are often represented by models that are still too complex to be described in a communicable manner. Further and further modeling is required until our observations can be communicated in a linguistically comprehensive way. The cascade of models gives us the advantage of creating "communicable reality" but does not help us to judge the evidence pertinent to "real" (i.e. not necessarily communicable) reality. On the contrary, the more advanced a model in a cascade is, the further is its "distance" (in terms of number of modeling steps) from the modeled system.

From a logical point of view, there are two problems here. First, we have no formal system of inference to judge *formal correctness* of observation sentences that have a variable true value (credibility) and that use ill-defined terms. Second, we have no formal system to judge the *material adequacy* of sentences derived from a "distant" model to the properties of modeled phenomena.

The problem of formal correctness seems to be solvable in principle. Progress in dealing with it can be noticed already in the fields of Artificial Intelligence (AI), Pattern Recognition (especially development of fuzzy mathematical techniques) and Situation Logic. In the near future, we can expect to have formal tools to derive plausible conclusions from imprecise premises. Or, at least, we will have the option of relying on a machine (i.e. reliable AI software) that will perform plausible reasoning for us.

As far as formal judgment of material adequacy is concerned, it is unlikely that the problem is addressable in all its generality. At least within mathematics, the celebrated Gödel theorem precludes such a possibility. However, informal (more or less educated common-sense) judgments of material adequacy are not only possible but all fields of science explore them.

6.3 Question of Measurement in Biology

The notion that only measurable properties of objects and phenomena belong to reality is a persistent legacy of neopositivism. For physicists, it means that

"real" things need to be viewed as systems of states with measurable observables (i.e. selected boundary conditions) assigned to those states. However, interpretation of biological activities in terms of microphysical phenomena will almost certainly lead to observables not known to physics so far. Moreover, preliminary evidence suggests that these new observables will be irreducible to the traditional (i.e. known) ones. How biologists will handle this measurement problem is not clear yet. Some authors suggest that perhaps it will be necessary for biologists to abandon the concept of states of the system as a foundation for modeling.

6.4 Convoluted Nature of Biological Systems

Complexity of biological systems is often described as a pivotal concept in systems biology. Unfortunately, this notion is terribly vague because there are over thirty different meanings in which it could be used and none of these meanings satisfactorily reflects the nature of living things. A satisfactory meaning (for "complex") should pertain to *functionally convoluted* systems (such as an organism, immune system, or ecosystem) that can be approximately described by several complementary models but cannot be adequately represented by a single model. This postulate of nonexistence of the largest model has several consequences such as:

1. Necessity to study pairs consisting of a modeler (usually a human) and the system to be modeled instead of the latter (i.e. modeled system) in isolation.
2. Promotion of the idea that functional organization of the system as a whole can be described by concepts and principles, which are independent from the properties of components.
3. The existence of general laws (or rules) that govern systems behavior (functioning) as a whole, which do not need to be derivable from the properties of the components of the system.

The foregoing general postulates of systems biology will likely lead to future research programs that will challenge causal reductionism (particularly the principle of upward causation) of the past explanations of emergence. Computer-assisted experimentation and model-making (theorizing) will without a doubt constitute a central methodology within these future paradigms.

See also Anthology of Human Repetitive DNA; Informatics (Computational Biology); Nucleic Acid and Protein Sequence Analysis and Bioinformatics.

Bibliography

Books and Reviews

Bishop, M.J., Rawlings, C.J. (1997) DNA and Protein Sequence Analysis, IRL Press, Oxford.

Doolittle, R.F. (1990) Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences, *Methods in Enzymology*, Vol. 183, Academic Press, San Diego, CA.

Doolittle, R.F. (1996) Computer Methods for Macromolecular Sequence Analysis, *Methods in Enzymology*, Vol. 266, Academic Press, San Diego, CA.

Durbin, R., Eddy, S., Krogh, A., Mitchison, G. (1998) *Biological Sequence Analysis*, Cambridge University Press, Cambridge.

Fickett, J.W. (1996) The gene identification problem: an overview for developers, *Comput. Chem.* **20**, 103–118.

Gentleman, J.F., Mullin, R.C. (1989) The distribution of the frequency of occurrence of nucleotide subsequences, based on their overlap capability, *Biometrics* **45**, 35–52.

Gribskov, M., Veretnik, S. (1996) Identification of Sequence Patterns with Profile Analysis, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 198–212.

Hamming, R.W. (1991) *The Art of Probability for Scientists and Engineers*, Addison-Wesley, Redwood City, CA.

Higgins, D., Taylor, W. (2000) *Bioinformatics: Sequence, Structure, and Databanks*, Oxford University Press, Oxford.

Konopka, A.K. (2003a) Information Theories in Molecular Biology and Genomics, in: Cooper, D.N. (Ed.) *Nature Encyclopedia of the Human Genome*, Vol. 3, Nature Publishing Group Reference, London, pp. 464–469.

Konopka, A.K. (2003b) Biomolecular Sequence Analysis: Pattern Acquisition and Frequency Counts, in: Cooper, D.N. (Ed.) *Nature Encyclopedia of the Human Genome*, Vol. 1, Nature Publishing Group Reference, London, pp. 311–322.

Konopka, A.K., Crabbe, M.J.C. (2004) *Compact Handbook of Computational Biology*, Dekker/CRC Press, Boca Raton, FL.

Kullback, S. (1976) *Statistical Methods in Cryptanalysis*, Aegean Park Press, Laguna Hills, CA.

Li, M., Vitányi, P. (1993) An Introduction to Kolmogorov Complexity and Its Applications, Springer-Verlag, New York.

Li, W. (1997) The study of correlation structures of DNA sequences: a critical review, *Comput. Chem.* **21**, 257–272.

Mikulecky, D.C. (2001) Robert Rosen (1934–1998): a snapshot of biology's Newton, *Comput. Chem.* **25**, 317–327.

Pevzner, P.A. (2000) *Computational Molecular Biology: An Algorithmic Approach*, The MIT Press, Cambridge, MA.

Rosen, R. (2000) *Essays on Life Itself*, Columbia University Press, New York.

Sankov, D., Kruskal, J. (1983/1999) *TIME WARPS, STRING EDITS, AND MACROMOLECULES: The Theory and Practice of Sequence Comparison*, CSLI Publications, Stanford, CA.

Sinkov, A. (1968) *Elementary Cryptanalysis, a Mathematical Approach*, Random House, New York.

Waterman, M.S. (1995) *Introduction to Computational Biology: Maps, sequences and genomes*, Chapman & Hall/CRC-Press, London.

Primary Literature

Akhmanova, O.S., Mel'chuk, I.A., Frumkina, R.M., Paducheva, E.V. (1963) *Exact Methods in Linguistic Research. Rand project R-397-PR*, The Rand Corporation, Santa Monica, CA, p. 300.

Allison, L. (2000) Intelligent systems in molecular biology (ISMB98), 27 June-1 July, 1998, Montreal, *Comput. Chem.* **24**, 129–132.

Allison, L., Stern, L., Edgoose, T., Dix, T.I. (2000) Sequence complexity for biological sequence analysis, *Comput. Chem.* **24**, 43–55.

Almagor, H. (1983) A Markov chain analysis of DNA sequences, *J. Theor. Biol.* **104**, 633–645.

Altschul, S.F., Gish, W. (1996) Local Alignment Statistics, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 460–480.

Altschul, S.F., Boguski, M.S., Gish, W., Wootton, J.C. (1994) Issues in searching molecular sequence databases, *Nat. Genet.* **6**, 119–129.

Argos, P. (1992) The language of protein folding: many forked Tongues, *Comput. Chem.* **16**, 93–102.

Argos, P., Abagyan, R. (1994) The protein folding problem: finding a few minimums in a near infinite space, *Comput. Chem.* **18**, 225–232.

Arques, D.G., Michel, C.J. (1987) Periodicities in introns, *Nucleic Acids Res.* **15**, 7581–7592.

Avery, O.T., MacLeod, C.M., McCarty, M. (1944) Studies of the chemical nature of the substance inducing transformation of pneumococcal types.I. induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III, *J. Exp. Med.* **79**, 137–158.

Bacro, J.N., Comet, J.P. (2001) Sequence alignment: an approximation law for the Z-value with applications to databank scanning, *Comput. Chem.* **25**, 401–410.

Balazs, A. (2004) What does a molecule want? The myth of the self-replicating molecule (comments on the "selfish-gene" paradigm), *Biosystems* **73**, 1–11.

Bar-Hillel, Y., Carnap, R. (1953) Semantic information, *Br. J. Phil. Sci.* **4**, 147–157.

Barker, W.C., Pfeiffer, F., George, D.G. (1996) Superfamily Classification in PIR-International Protein Sequence Database, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 59–71.

Bell, G.I. (1992) Roles of repetitive sequences, *Comput. Chem.* **16**, 135–143.

Bell, G. (1996) Evolution of simple sequence repeats, *Comput. Chem.* **20**, 41–48.

Bell, G.I., Torney, D.C. (1993) Repetitive DNA sequences: some considerations for simple sequence repeats, *Comput. Chem.* **17**, 185–190.

Blaisdell, B.E. (1983a) A prevalent persistent nonrandomness that distinguishes coding and noncoding eukaryotic nuclear DNA sequences, *J. Mol. Evol.* **19**, 122–133.

Blaisdell, B.E. (1983b) Choice of base at silent codon site 3 is not selectively neutral in eukaryotic structural Genes: It maintains excess Short runs of weak and strong hydrogen bonding bases, *J. Mol. Evol.* **19**, 226–2236.

Blaisdell, B.E. (1985) Markov chain analysis finds a significant influence of neighboring bases on the occurrence of a base in eukaryotic nuclear DNA sequences both protein-coding and noncoding, *J. Mol. Evol.* **21**, 278–288.

Blake, R.D., Hess, S.T. (1992) The pattern of substitution mutation in different Nearest-neighbor environments of the human genome, *Comput. Chem.* **16**, 165–170.

Bork, P., Gibson, T.J. (1996) Applying Motif and Profile Searches, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 162–184.

Borodovsky, M.Y., Gusein-Zade, S.M. (1989) A general rule for ranged series of codon frequencies in different genomes, *J. Biomol. Struct. Dyn.* **6**, 1001–1012.

Borodovsky, M.Y., Kleffe, J. (1992) First and second moment of counts of words in random texts generated by Markov chains, *Comp. Appl. Biosci.* **8**, 433–441.

Borodovsky, M., Peresetsky, A. (1994) Deriving Nonhomogeneous Markov chain models by cluster analysis algorithm minimizing multiple alignment entropy, *Comput. Chem.* **18**, 259–268.

Bougueleret, L., Tekaia, F., Sauvaget, I., Claverie, J.-M. (1988) Objective comparison of exon and intron sequences by the mean of 2-dimensional data analysis methods, *Nucleic Acids Res.* **16**, 1729–1738.

Bowie, J.U., Zhang, K., Wilmanns, M., Eisenberg, D. (1996) Three-dimensional Profiles for Measuring Compatibility of Amino Acid Sequence with Three-dimensional Structure, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 598–616.

Breen, S., Waterman, M.S., Zhang, N. (1985) Renewal theory for several patterns, *J. Appl. Probab.* **22**, 228–234.

Brendel, V., Karlin, S. (1994) Applications of statistical criteria in protein sequence analysis: case study of yeast RNA polymerase II subunits, *Comput. Chem.* **18**, 251–253.

Brendel, V., Beckmann, J.S., Trifonov, E.N. (1986) Linguistics of nucleotide sequences: morphology and comparison of vocabularies, *J. Biomol. Struct. Dyn.* **4**, 11–21.

Bucher, P., Karplus, K., Moeri, N., Hoffman, K. (1996) A flexible motif search technique based on generalized profiles, *Comput. Chem.* **20**, 3–23.

Burton, N.G., Licklider, J.C.R. (1955) Long-range constraints in the statistical structure of printed english, *Am. J. Psychol.* **68**, 650–653.

Chaitin, G.J. (1966) On the length of programs for computing finite binary sequences, *J. ACM* **13**, 547–569.

Chapanis, A. (1954) The reconstruction of abbreviated printed messages, *J. Exp. Psychol.* **48**, 496–510.

Chichester, C., Nikitin, F., Ravarini, J.-C., Lisacek, F. (2003) Consistency checks for characterizing protein forms, *Comput. Biol. Chem.* **27**, 29–35.

Churchill, G.A. (1989) Stochastic models for heterogeneous DNA sequences, *Bull. Math. Biol.* **51**, 79–94.

Churchill, G.A. (1992) Hidden Markov chains and the analysis of genome structure, *Comput. Chem.* **16**, 107–115.

Claverie, J.-M. (1992) Sequence "Signals": artifacts or reality? *Comput. Chem.* **16**, 89–91.

Claverie, J.-M. (1994) Some useful statistical properties of position-weight matrices, *Comput. Chem.* **18**, 287–294.

Claverie, J.-M., Bougueleret, L. (1986) Heuristic information analysis of sequences, *Nucleic Acids Res.* **14**, 179–196.

Condon, E.U. (1928) Statistics of vocabulary, *Science* **67**, 300–303.

Crick, F.H.C. (1958) On protein synthesis, *Symp. Soc. Exp. Biol.* **12**, 138–163.

Delorme, M.O., Henaut, A. (1988) Merging of distance matrices and classification by dynamic clustering, *Comput. Appl. Biosci.* **4**, 453–458.

Delorme, M.O., Henaut, A. (1991) Codon usage is imposed by the gene location in the transcription unit, *Curr. Genet.* **20**, 353–358.

Dembo, A., Karlin, S. (1991) Strong limit theorems of empirical functionals for large exceedances of partial sums of i.i.d. variables, *Ann. Probab.* **19**, 1737–1755.

Dembo, A., Karlin, S., Zeitouni, O. (1994) Limit distribution of maximal nonaligned two-sequence segmental score, *Ann. Probab.* **22**, 2022–2039.

Doolittle, R.F. (1996a) Computer Methods for Macromolecular Sequence Analysis, in: Abelson, J.N., Simon, M.I. (Eds.) *Methods in Enzymology*, Vol. 266, Academic Press, San Diego, CA, p. 711.

Doolittle, R.F. (ed.) (1996b) *Computer Methods for Macromolecular Sequence Analysis*, (Methods in Enzymology Vol. 266), Academic Press, San Diego, CA.

Fickett, J.W. (1982) Recognition of protein coding regions in dna sequences, *Nucleic Acids Res.* **10**, 5303–5318.

Fickett, J.W. (1996) The gene identification problem: an overview for developers, *Comput. Chem.* **20**, 103–118.

Foulcer, D., Karlin, S. (1987) Maximal success durations for a semi-Markov process, *Stoch. Proc. Appl.* **24**, 203–224.

Francis, W., Kucera, H. (1982) *Frequency Analysis of English Usage: Lexicon and Grammar*, Houghton Mifflin, Boston, MA.

Garden, P.W. (1980) Markov analysis of viral DNA/RNA sequences, *J. Theor. Biol.* **82**, 679–684.

Garnier, J., Girbat, J.-F., Robson, B. (1996) GOR Method for Predicting Protein Secondary Structure from Amino Acid Sequence, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 540–553.

Gattiker, A., Michoud, K., Rivoire, C. et al. (2003) Automated annotation of microbial proteomes in SWISS-PROT, *Comput. Biol. Chem.* **27**, 49–58.

Gelfand, M.S., Kozhukhin, C.G., Pevzner, P.A. (1992) Extendable words in nucleotide sequences, *Comput. Appl. Biosci.* **8**, 129–135.

Gouzy, J., Corpet, F., Kahn, D. (1999) Whole genome protein domain analysis using a new method for domain clustering, *Comput. Chem.* **23**, 333–340.

Gribskov, M., Robinson, N.L. (1996) Use of receiver operating characteristic (ROC) analysis to evaluate sequence matching, *Comput. Chem.* **20**, 25–33.

Gribskov, M., Veretnik, S. (1996) Identification of Sequence Patterns with Profile Analysis, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 198–212.

Gribskov, M., Devereux, J., Burgess, R.R. (1984) The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression, *Nucleic Acids Res.* **12**, 539–549.

Guibas, L.J., Odlyzko, A.M. (1981) String overlaps, pattern matching and nontransitive games, *J. Comp. Theory (A)* **30**, 183–208.

Harris, Z.S. (1954) Distributional structure, *Word* **10**, 775–793.

Harris, Z.S. (1955) From phoneme to morpheme, *Language* **31**, 190–222.

Henaut, A., Limaiem, J., Vigier, P. (1985) The origins of the strategy of codon use, *Biochimie* **67**, 475–483.

Henikoff, J.G., Henikoff, S. (1996) Bloks Database and Its Applications, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 88–105.

Heringa, J. (1994) The evolution and recognition of protein sequence repeats, *Comput. Chem.* **18**, 233–244.

Heringa, J. (1999) Two strategies for sequence comparison: profile-preprocessed and secondary structure-induced multiple alignment, *Comput. Chem.* **23**, 341–364.

Heringa, J. (2000) *Post-genome Informatics: Minoru Kanehisa*, Oxford University Press, Oxford, p. 148; *Comput. Chem.* (2001) **25**, 425–426.

Hershey, A.D., Chase, M. (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage, *J. Gen. Physiol.* **36**, 39–56.

Higgins, D.G., Thompson, J.D., Gibson, T.J. (1996) Using CLUSTAL for Multiple Sequence Alignments, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 383–402.

Inoue, Y., Ikeda, M., Shimizu, T. (2004) Proteome-wide classification and identification of mammalian-type GPCRs by binary topology pattern, *Comput. Biol. Chem.* **28**, 39–49.

Khinchin, A.I. (1953) The entropy concept in probability theory, *Uspiekhi Matematicheskikh Nauk (Russian)* **VIII**, 3–20.

Klaerr-Blanchard, M., Chiapello, H., Coward, E. (2000) Detecting localized repeats in genomic sequences: a new strategy and its application to bacillus subtilis and Arabidopsis thaliana sequences, *Computers & Chemistry* **24**, 57–70.

Kolmogorov, A.N. (1965) Three approaches to the definition of the concept "Quantity of Information", *Probl. Peredachi Inform. (Russian)* **1**, 3–11.

Konings, D.A.M. (1992) Coexistence of multiple codes is messenger RNA molecules, *Comput. Chem.* **16**, 153–163.

Konopka, A.K. (1984) Is the information content of DNA evolutionarily significant? *J. Theor. Biol.* **107**, 697–704.

Konopka, A.K. (1985) Theory of degenerate coding and informational parameters of protein coding genes, *Biochimie* **67**, 455–468.

Konopka, A.K. (1990) Towards Mapping Functional Domains in Indiscriminantly Sequenced Nucleic Acids: A Computational Approach, in: Sarma, R.H., Sarma, M.H. (Eds.) *Human Genome Initiative and DNA Recombination*, Vol. 1, Adenine Press, Guiderland, NY, pp. 113–125.

Konopka, A.K. (1994) Sequences and Codes: Fundamentals of Biomolecular Cryptology, in: Smith, D. (Ed.) *Biocomputing: Informatics and Genome Projects*, Academic Press, San Diego, CA, pp. 119–174.

Konopka, A.K. (2003a) Sequence Complexity and Composition, in: Cooper, D.N. (Ed.) *Nature Encyclopedia of the Human Genome*, Vol. 5, Nature Publishing Group Reference, London, pp. 217–224.

Konopka, A.K. (2003b) Systems Biology: Aspects Related to Genomics, in: Cooper, D.N. (Ed.) *Nature Encyclopedia of the Human Genome*, Vol. 5, Nature Publishing Group Reference, London, pp. 459–465.

Konopka, A.K., Brendel, V. (1985) The missense errors in protein can be controlled by selective synonymous codon usage at the level of transcription, *Biochimie* **67**, 469–473.

Konopka, A.K., Smythers, G.W., Owens, J., Maizel, J.V. Jr. (1987a) Distance analysis helps to establish characteristic motifs in intron sequences, *Gene Anal. Techn.* **4**, 63–74.

Konopka, A.K., Smythers, G.W. (1987b) DISTAN-a program which detects significant distances between short oligonucleotides, *Comput. Appl. Biosci.* **3**, 193–201.

Konopka, A.K., Martindale, C. (1995) Noncoding DNA, Zipf's Law and language, *Science* **268**, 789.

Konopka, A.K., Owens, J. (1990) Complexity charts can be used to map functional domains in DNA, *Gene Anal. Techn. Appl.* **7**, 35–38.

Koonin, P.Ba.E.V. (1996) Protein sequence motifs, *Curr. Opin. Struct. Biol.* **6**, 366–376.

Koonin, E.V., Tatusov, R.L., Rudd, K.E. (1996) Protein Sequence Comparison at Genome Scale, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 295–322.

Kozhukhin, C.G., Pevzner, P.A. (1991) Genome inhomogeneity is determined mainly by WW and SS dinucleotides, *Comput. Appl. Biosci.* **7**, 39–49.

Lakoff, G. (1993) The Contemporary Theory of Metaphor, in: Ortony, A. (Ed.) *Metaphor and Thought*, 2nd edition, Cambridge University Press, Cambridge, MA, pp. 11–52.

Landes, C., Henaut, A., Risler, J.L. (1992) A comparison of several similarity indices used in the classification of protein sequences: a multivariate analysis, *Nucleic Acids Res.* **20**, 3631–3637.

Lawrence, C. (1994) Toward the unification of sequence and structural data for identification of structural and functional constraints, *Comput. Chem.* **18**, 255–258.

Lewontin, R.C. (1996) Evolution as Engineering, in: Collado-Vides, J., Smith, T., Magasanik, B., (Eds.) *Integrative Approaches to Molecular Biology*, The MIT Press, Cambridge, MA, 1–10.

Lewontin, R. (2000) *The Triple Helix: Gene Organism and Environment*, Harvard University Press, Cambridge, MA.

Lewontin, R. (2001) *It Ain't Necessarily So: The Dream of the Human Genome and Other*

Illusions, 2nd edition, New York Review Books, New York.

Li, W. (1999) Statistical properties of open reading frames in complete genome sequences, *Comput. Chem.* **23**, 283–301.

Li, W.-H., Gu, X. (1996) Estimating Evolutionary Distances between DNA Sequences, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 449–459.

Li, W., Bernaola-Galvan, P., Carpena, P., Oliver, J.L. (2003) Isochores merit the prefix 'iso', *Comput. Biol. Chem.* **27**, 5–10.

Lipman, D.J., Wilbur, W.J. (1983) Contextual constraints on synonymous codon choice, *J. Mol. Biol.* **163**, 363–376.

Llabres, M., Rossello, F. (2004) A new family of metrics for biopolymer contact structures*1, *Comput. Biol. Chem.* **28**, 21–37.

Luc, N., Risler, J.-L., Bergeron, A., Raffinot, M. (2003) Gene teams: a new formalization of gene clusters for comparative genomics, *Comput. Biol. Chem.* **27**, 59–67.

Markov, A.A. (1913) An example of statistical investigation of the poem "Eugene Onegin" illustrating the connection of events in chain, *Izvestia Acad. Nauk (in Russian)* Series VI, **3**, 153–162.

Martindale, C., Konopka, A.K. (1996) Oligonucleotide frequencies in DNA follow a Yule distribution, *Comput. Chem.* **20**, 35–38.

Mengeritzky, G., Smith, T.F. (1987) Recognition of characteristic patterns in sets of functionally equivalent DNA sequences, *Comput. Appl. Biosci.* **3**, 223–227.

Michel, C.J. (1986) New statistical approach to discriminate between protein coding and noncoding regions in DNA sequences and its evaluation, *J. Theor. Biol.* **120**, 223–236.

Mikulecky, D.C. (2001a) Robert Rosen (1934–1998): a snapshot of biology's Newton, *Comput. Chem.* **25**, 317–327.

Mikulecky, D.C. (2001b) The emergence of complexity: science coming of age or science growing old? *Comput. Chem.* **25**, 341–348.

Mikulecky, D.C. (2001c) Network thermodynamics and complexity: a transition to relational systems theory, *Comput. Chem.* **25**, 369–391.

Miller, G.A. (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information, *Psychol. Rev.* **63**, 81–97.

Miller, G.A., Friedman, E. (1957) A. The reconstruction of mutilated english texts, *Inform. Control* **1**, 38–55.

Nakata, K., Kanehisa, M., DeLisi, C. (1985) Prediction of splice junctions in mRNA sequences, *Nucleic Acids Res.* **13**, 5327–5340.

Nirenberg, M.W., Matthaei, H.J. (1961) The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides, *Proc. Natl. Acad. Sci. U.S.A.* **47**, 1588–1602.

Nyquist, H. (1924) Certain factors affecting telegraphy speed, *Bell Syst. Tech. J.* **3**, 324–326.

Pattee, H.H. (1969) How does a Molecule become a Message? in: Lang, A. (Ed.) *28th Symposium of the Society of Developmental Biology*, Academic Press, New York, pp. 1–16.

Pattee, H.H. (2001) The physics of symbols: bridging the epistemic cut, *BioSystems* **60**, 5–21.

Patthy, L. (1996) Consensus Approaches in Detection of Distant Homologies, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 184–198.

Pauling, L.C. (1958) Current opinion: molecular disease, *Pfizer Spectrum* **6**, 234–235.

Pearson, W.R. (1996) Effective Protein Sequence Comparison, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 227–258.

Pedersen, A.G., Baldi, P., Chauvin, Y., Brunak, S. (1999) The biology of eukaryotic promoter prediction-a review, *Comput. Chem.* **23**, 191–207.

Pesole, G., Attimonelli, M., Saccone, C. (1996) Linguistic Analysis of Nucleotide Sequences: Algorithms for Pattern Recognition and Analysis of Codon Strategy, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 281–294.

Pesole, G., Grillo, G., Liuni, S. (1996) Databases of mRNA untranslated regions for Metazoa, *Comput. Chem.* **20**, 141–144.

Pevzner, P.A. (1992) Nucleotide sequences versus Markov models, *Comput. Chem.* **16**, 103–106.

Pevzner, P.A. (1994) Rearrangements of DNA sequences and SBH, *Comput. Chem.* **18**, 221–223.

Pevzner, P.A., Borodovsky, M.Yu., Mironov, A.A. (1989a) Linguistics of nucleotide sequences I:

the significance of deviations from mean statistical characteristics and prediction of the frequencies of occurrence of words, *J. Biomol. Struct. Dyn.* **6**, 1013–1026.

Pevzner, P.A., Borodovsky, M.Yu., Mironov, A.A. (1989b) Linguistics of nucleotide sequences II: stationary words in genetic texts and the zonal structure of DNA, *J. Biomol. Struct. Dyn.* **6**, 1027–1038.

Phillips, G.J., Arnold, J., Ivarie, R. (1987) Mono- through hexanucleotide composition of Escherichia coli genome: a Markov chain analysis, *Nucleic Acids Res.* **15**, 2611–2626.

Poch, O., Delarue, M. (1996) Converting Sequence Block Alignments into Structural Insights, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 662–680.

Quentin, Y., Voiblet, C., Martin, F., Fichant, G. (1999) Protein-coding region discovery in organisms under-represented in databases, *Comput. Chem.* **23**, 209–217.

Régnier, M. (1998) A Unified Approach to Word Statistics, *Proceedings of RECOMB 98*, New York, pp. 207–213.

Reidys, C., Stadler, P.F. (1996) Bio-molecular shapes and algebraic structures, *Comput. Chem.* **20**, 85–94.

Ridley, M. (1996) *Evolution*, Blackwell Science, Inc., Cambridge, MA.

Risler, J.L., Delorme, M.O., Delacroix, H., Henaut, A. (1988) Amino acid substitutions in structurally related proteins. A pattern recognition approach. Determination of a new and efficient scoring matrix, *J. Mol. Biol.* **204**, 1019–1029.

Rosen, R. (1994) What is biology? *Comput. Chem.* **18**, 347–352.

Rosen, R. (1996) Biology and the measurement problem, *Comput. Chem.* **20**, 95–100.

Rost, B. PhD (1996) Predicting One-dimensional Protein Structure by Profile-Based Neural Networks, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 525–539.

Saitou, N. (1996) Reconstruction of Gene Trees from Sequence Data, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 427–449.

Salamon, P., Konopka, A.K. (1992) A maximum entropy principle for distribution of local complexity in naturally occurring nucleotide sequences, *Comput. Chem.* **16**, 117–124.

Salamon, P., Wootton, J.C., Konopka, A.K., Hansen, L.K. (1993) On the robustness of maximum entropy relationships for complexity distributions of nucleotide sequences, *Comput. Chem.* **17**, 135–148.

Sarai, A. (1989) Molecular recognition and information gain, *J. Theor. Biol.* **140**, 137–143.

Sellers, P.H. (1974) On the theory and computation of evolutionary distances, *SIAM J. Appl. Math.* **26**, 787–793.

Shannon, C.E. (1948). A mathematical theory of communication, *Bell Syst. Tech. J.* **27**, 379–423, and 623–656.

Shannon, C.E. (1949) Communication theory of secrecy systems, *Bell Syst. Tech. J.* **28**, 657–715.

Shannon, C.E. (1951) Prediction and entropy of printed english, *Bell Syst. Tech. J.* **30**, 50–64.

Shepherd, J.C.W. (1981) Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and its possible evolutionary justification, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1596–1600.

Shulman, M.J., Steinberg, C.M., Westmoreland, N. (1981) The coding function of nucleotide sequences can be discerned by statistical analysis, *J. Theor. Biol.* **88**, 409–420.

Smith, L., Yeganova, L., Wilbur, W.J. (2003) Hidden Markov models and optimized sequence alignments, *Comput. Biol. Chem.* **27**, 77–84.

Snyder, E.E., Stormo, G.D. (1993) Identification of coding regions in genomic DNA sequences: an application of dynamic programming and neural networks, *Nucleic Acids Res.* **21**, 607–613.

Sober, E. (1994) *Conceptual Issues in Evolutionary Biology*, The MIT Press, Cambridge, MA.

Solomonoff, R.J. (1964) A formal theory of inductive inference, *Inform. Control* **7**, 224–254.

Staden, R. (1984a) Measurement of the effects that coding for a protein has on a DNA sequence and their use for finding genes, *Nucleic Acids Res.* **12**, 551–567.

Staden, R. (1984b) Computer methods to locate signals in nucleic acid sequences, *Nucleic Acids Res.* **12**, 505–519.

Staden, R. (1984c) Graphic methods to determine the function of nucleic acid sequences, *Nucleic Acids Res.* **12**, 521–538.

Staden, R. (1996) Indexing and Using Sequence Databases, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 105–114.

Stormo, G.D. (1990a) Consensus Patterns in DNA, in: Doolittle, R.F. (Ed.) *Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences*, Vol. 183, Academic Press, San Diego, CA, pp. 211–220.

Stormo, G.D. (1990b) Identifying Regulatory Sites from DNA Sequence Data, in: Sarma, R.H., Sarma, M.H. (Eds.) *Structure and Methods*, Vol. 1, Adenine Press, Guiderland, NY, pp. 103–112.

Tautz, D., Trick, M., Dover, G.A. (1986) Cryptic simplicity in DNA is a major source of genetic variation, *Nature* **322**, 652–656.

Taylor, W.R. (1993) Protein structure prediction from sequence, *Comput. Chem.* **17**, 117–122.

Taylor, W.R. (1996) Multiple Protein Sequence Alignment: Algorithms and Gap Insertion, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 343–367.

Taylor, W.R., Brown, N.P. (1999) Iterated sequence databank search methods, *Comput. Chem.* **23**, 365–385.

Taylor, W.R., Saelensminde, G., Eidhammer, I. (2000) Multiple protein sequence alignment using double-dynamic programming, *Comput. Chem.* **24**, 3–12.

Taylor, W.R., Munro, R.E.J., Petersen, K., Bywater, R.P. (2003) Ab initio modeling of the N-terminal domain of the secretin receptors, *Comput. Biol. Chem.* **27**, 103–114.

Taylor, W.R., Xiao, B., Gamblin, S.J., Lin, K. (2003) A knot or not a knot? SETting the record 'straight' on proteins, *Comput. Biol. Chem.* **27**, 11–15.

Terzian, C., Laprevotte, I., Brouillet, S., Henaut, A. (1997) Genomic signatures: tracing the origin of retroelements at the nucleotide level, *Genetica* **100**, 271–279.

Trifonov, E.N. (1989) The multiple codes of nucleotide sequences, *Bull. Math. Biol.* **51**, 417–432.

Turing, A.M. (1936) On computable numbers with an application to the Entscheidungsproblem, *Proc. Lond. Math. Soc., ser. 2* **42**, 230–265.

Turing, A.M. (1950) Computing machinery and intelligence, *Mind*, **59**(236), 433–460.

Uberbacher, E.C., Mural, R.J. (1991) Locating protein-coding regions in human DNA sequences by a multiple-sensor neural network approach, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 11261–11265.

Ulanowicz, R.E. (1999) Life after Newton: an ecological metaphysics, *BioSystems* **50**, 127–142.

Ulanowicz, R.E. (2001) The organic in ecology, *Ludus Vitalis* **9**, 183–204.

Watson, J.D., Crick, F.H.C. (1953a) Molecular structure of nucleic acids: a structure for deoxyribonucleic acid, *Nature* **171**, 737–738.

Watson, J.D., Crick, F.H.C. (1953b) General implications of the structure of deoxyribonucleic acid, *Nature* **171**, 964–967.

White, O., Kerlavage, A. (1996) TDB: New Databases for Biological Discovery, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Vol. 266, Academic Press, San Diego, CA, pp. 27–40.

Wilbur, W.J., Neuwald, A.F. (2000) A theory of information with special application to search problems, *Comput. Chem.* **24**, 33–42.

Wills, C. (1996) Improving the analysis of phylogenetic data, *Comput. Chem.* **20**, 61–66.

Wittgenstein, L. (1922) *Tractatus Logico-philosophicus*, Routledge and Kegan Paul, London.

Wittgenstein, L. (1953) *Philosophical Investigations*, Blackwell, London.

Wolpert, L. (1969) Positional information and pattern, *J. Theor. Biol.* **25**, 1–49.

Wootton, J.C. (1994) NonGlobular domains in protein sequences: automated segmentation using complexity measures, *Comput. Chem.* **18**, 269–286.

Wootton, J.C., Federhen, S. (1993) Statistics of local complexity in amino acid sequences and sequence databases, *Comput. Chem.* **17**, 149–1163.

Wright, L. (1994) Functions, in: Sober, E. (Ed.) *Conceptual Issues in Evolutionary Biology*, The MIT Press, Cambridge, MA, pp. 27–47.

Yockey, H.P. (1974) An application of information theory to the central dogma and the sequence hypothesis, *J. Theor. Biol.* **46**, 369–406.

Yockey, H.P. (1979) Do overlapping genes violate molecular biology and the theory of evolution? *J. Theor. Biol.* **80**, 21–26.

Zadeh, L.A. (1999) From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions, *IEEE Trans. Circ. Syst. – I: Fundam. Theory Appl.* **45**, 105–119.

Zhurkin, V.B. (1983) Local mobility in the DNA double helix; comparison of conformational calculations with experiment, *Mol. Biol. (Russian)* **17**, 495–498.

Zipf, G.K. (1935) *The Psycho-biology of Language*, Houghton, Boston, MA.

Zipf, G.K. (1949) *Human Behavior and the Principle of Least Effort*, Addison-Wesley, Cambridge, MA.

Zuckerkandl, E., Pauling, L. (1962) *Molecules as Documents of Evolutionary History: Gates and Crellin Laboratories of Chemistry – California Institute of Technology*. Pasadena, CA Contribution No. 3041, pp. 1–19.

Preimplantation Mammalian Development, Regulation of Gene Expression in

Melvin L. DePamphilis
National Institute of Child Health and Human Development Bethesda, MD, USA

1 Preimplantation Development 485

2 Maternal to Zygotic Gene Transition 486
2.1 Timing 486
2.2 Genes 487

3 Zygotic Clock 488
3.1 Protein Translocation 489
3.2 Protein Phosphorylation 489
3.3 Translational Control 489

4 Developmental Acquisition of Transcriptional Regulation 490
4.1 Chromatin-mediated Repression 490
4.2 Enhancer Function 493
4.3 DNA Replication 495
4.4 TATA-box Function 496
4.5 DNA Methylation 496
4.6 Differential Gene Expression 498

Bibliography 499
Books and Reviews 499
Primary Literature 500

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

Keywords

Enhancer
Cis-acting regulatory DNA element involved in transcriptional activation of genes. Enhancers are located distal to mRNA start site, function in either orientation and at either end of the gene, and bind specific proteins required for enhancer activity.

Promoter
Cis-acting regulatory DNA element involved in transcriptional activation of genes. Promoters are located proximal to the mRNA start site, exhibit orientation dependence, and bind specific proteins required for transcription.

TATA-box
A DNA sequence motif with the consensus TATA(A/T)A(A/T) that is commonly found 20- to 30-bp upstream of the mRNA start site in RNA polymerase II genes. The TATA-box binds TFIID, a transcription factor complex containing the TATA recognition protein, TBP (TATA-binding protein).

Preimplantation development in the mouse involves expression of about 11 000 genes, only a few hundred of which appear during the transition from maternal to zygotic gene expression. Transcription begins in most, if not all, mammals during the late one-cell stage (phase I), but expression of most zygotic genes is delayed until the 2-cell to the 16-cell stage, depending on the mammal. In mice, a small group of genes is expressed immediately after the first mitosis, while a larger group of genes is expressed during the subsequent G2-phase. Zygotic gene activation (ZGA) is delayed by a time-dependent mechanism (''zygotic clock'') that regulates both transcription and translation. It appears to involve posttranslational modification of RNA polymerases, translational control of maternal gene expression, developmental acquisition of chromatin-mediated repression, and the ability to alleviate this repression with sequence-specific enhancers. This delay allows remodeling of chromatin into a form that globally represses gene activity so that selected genes can then be activated in a temporally and spatially specific program. Some transcription factors (e.g. Sp1, TBP (TATA-binding protein)) function from oocyte to embryo, while others (e.g. Tead2, Oct4) are selectively expressed during ZGA. Dramatic changes also occur in DNA methylation. However, while DNA methylation has been linked directly to X-chromosome inactivation, genomic imprinting, and silencing of transposable elements, a direct role for DNA methylation in regulating gene expression during animal development has yet to be demonstrated. For example, the expression pattern of closely linked genes that are expressed in different cell types in adult animals is not determined by differential DNA methylation. DNA methylation appears to serve primarily to restrict gene expression by insuring that repressed genes remain silent.

1
Preimplantation Development

Preimplantation development in mammals encompasses the events that take place between fertilization (day 1) and formation of the blastocyst (day 4 in mice). In mice, fertilization activates the first of six to seven cell cleavage cycles over a four-day period; these cycles culminate in the formation of a blastocyst containing 64 to 128 cells, comprising two cell types (Fig. 1). The outer layer of cells comprises the trophectoderm that will form the placenta, while the inner cell mass consists of about 25 totipotent embryonic stem cells that will produce the embryo. The first round of DNA replication occurs in both the maternal and paternal pronucleus of the fertilized egg (or one-cell embryo). About 4 h later the first mitosis occurs to produce a two-cell embryo in which each "zygotic nucleus" contains a complete set of parental chromosomes. This cleavage event is rapidly followed by the second S-phase and expression of a small number of zygotic genes. Changes in totipotency of individual cells are first detected at the eight-cell stage, although clearly differentiated cells first appear at the blastocyst stage. Here the blastomeres of the embryo exhibit different cellular

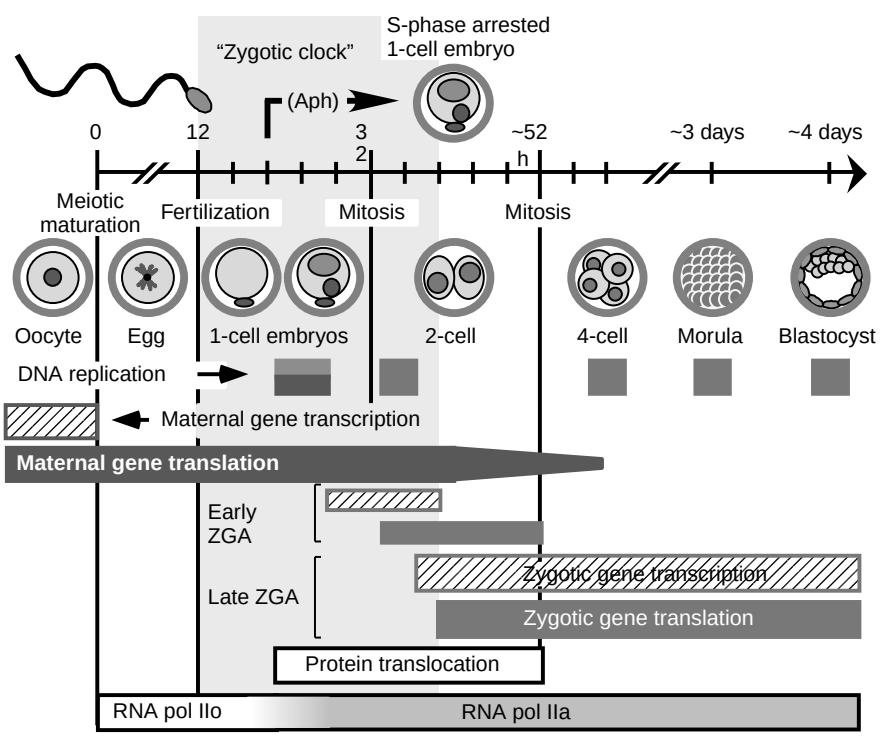

Fig. 1 Maternal to zygotic gene transition in the mouse. Maternal events are indicated in red, paternal events in blue, and zygotic events in green. Open bars apply to all three. Embryonic stem cells ("inner cell mass") are indicated in yellow and trophectodermal cells in orange. Periods of transcription are indicated by hatched bars and translation is indicated by solid bars (see color plate p. xxi).

features that presumably lead to the formation of embryonic (inner cell mass) and extraembryonic (trophectoderm) lineages.

The series of specific morphological, biochemical, and molecular events that occur postfertilization, and that allow selected genes to be expressed, have been reviewed elsewhere.

2 Maternal to Zygotic Gene Transition

All of the genes in sperm and eggs are transcriptionally silent. Therefore, the first one or two cell-cleavage events and the activation of zygotic gene expression must rely on the reservoir of mRNAs and proteins stored in the egg. This transition from maternal to zygotic gene dependence and its associated changes in chromatin organization, gene expression, DNA replication, and cell division is a phenomenon characteristic of most, if not all, of the metazoa.

2.1 Timing

While transcription of the haploid genome in oocytes from all species stops when they mature into eggs, the extent of development and the time that elapses before transcription begins again after fertilization varies considerably among species. For example, before zygotic genes are activated in flies, the fertilized egg has undergone 10 nuclear divisions in 1.5 h, those from frogs have undergone 11 cell cleavages in 6 h, while those from mice have undergone only one cleavage event in 24 h. But even among mammals, the major onset of ZGA varies from the 2-cell stage in mice to the 4-cell stage in pigs, to the 8- to 16-cell stage in cows, sheep, and rabbits, presumably reflecting differences in the amount of one or more maternally inherited proteins or mRNAs. For example, mouse oocytes specifically express the protein "zygote arrest 1" that is required for progression of one-cell to two-cell embryos, and the protein "matter" that is required for progression beyond the two-cell stage. Rad51 and profilin also have been shown to be required for preimplantation development.

In the mouse, a growing oocyte that is arrested at diplotene of its first meiotic prophase transcribes and translates many of its own genes, thereby producing a store of proteins sufficient to support development to the eight-cell stage. When an oocyte matures into an egg, it arrests in the metaphase of its second meiotic division where transcription stops and translation of mRNA is reduced (Fig. 1). Fertilization triggers completion of meiosis and formation of a one-cell embryo containing a haploid paternal pronucleus derived from the sperm and a haploid maternal pronucleus derived from the oocyte. DNA replication begins after formation of the pronuclear envelopes, and occurs in each pronucleus from 10 to 16 hours postfertilization (hpf). The first mitosis occurs from 17 to 20 hpf to produce a two-cell embryo containing two diploid "zygotic" nuclei, each with a set of paternal and a set of maternal chromosomes.

Formation of a mouse two-cell embryo marks the transition from maternal to zygotic gene dependence (Fig. 1). Degradation of maternal mRNA is triggered by meiotic maturation and is ~90% complete in two-cell embryos, although translation of maternal mRNA continues into the eight-cell stage. Injection of plasmid DNA and transplantation of nuclei from two-cell stage embryos into one-cell embryos has

shown that late one-cell embryos are transcriptionally competent. Transcription of endogenous genes has also been detected in late one-cell mouse embryos where it begins at the end of S-phase. However, transcription-dependent protein synthesis does not begin until 2 to 4 h after completion of the first mitosis and the beginning of S-phase in two-cell embryos. Eight to ten hours later, during G2-phase of two-cell embryos, expression of zygotic genes increases in both amount and complexity.

Thus, ZGA has at least three recognizable phases: transcription without translation begins in late one-cell embryos (phase I), but transcription coupled to translation does not begin until the early two-cell stage in development (phase II), and robust transcription coupled to translation does not begin until the late two-cell stage (phase III). The existence of the first and third phases can be demonstrated clearly by injection of a plasmid-encoded reporter gene. The reporter gene is transcribed as soon as the one-cell embryo becomes transcriptionally competent, but it is not translated until the third phase of ZGA begins in late two-cell embryos. The second phase of ZGA is specific for a subset of zygotic genes. An early phase of ZGA consisting of a small group of genes appears to occur shortly after fertilization in all vertebrates; it is the late, robust phase of ZGA that is delayed to different extents in different species.

2.2 Genes

The available data suggest that 38 to 282 genes are specifically expressed during ZGA in mice, but only a few of these genes have been identified. Unfertilized mouse eggs contain 258 expressed sequence tags not found in adult mouse cDNA libraries.

Tab. 1 Genes expressed at the beginning of mouse development.

Gene	*Expression begins*	*Imprinting*
Hsp70.1	Early 2-cell	
Hsp68	Early 2-cell	
U2afbp	Early 2-cell	Paternal
eIF-1A	Early 2-cell	
Histone H1	Late 2-cell	
Oct4(Oct3)	Late 2-cell	
Tead2(TEF-4)	Late 2-cell	
cyclin-A	Late 2-cell	
Xist	Late 2-cell	Paternal
Sry	Late 2-cell	
Zfy	Late 2-cell	
3 Zn^{++} finger proteins	2-cell on; 8-cell off	
SRp20	2-cell to 4-cell	
Soggy	2-cell to 4-cell	
Snrpn	4-cell	Paternal

Analysis of cDNA expression libraries has identified 9718 to 11 483 genes expressed from the oocyte to the blastocyst stage, 798 to 1585 of which were not found elsewhere in the mouse. About 2.5% of newly synthesized proteins show a transient increase at the two-cell stage, and about 2.9% of unique expressed sequence tags in preimplantation embryos sharply increased in amount at the two-cell stage. Early *ZGA* genes (Fig. 1; Table 1) include heat shock genes, the translation initiation factor eIF-1A, a gene of unknown function, and three as yet unidentified proteins known as the "transcription-requiring complex." Another candidate, *Fas*, is a gene involved in apoptosis, whose mRNA appears to be expressed specifically at the two-cell stage in rats and the four-cell stage in humans. Examples of late *ZGA* genes include linker histone H1, transcription factors Oct-4(Oct-3) and Tead2(TEF-4), splicing factor SRp20, and *Sgy*, a gene of unknown function that

is closely linked to *Tead2* (Table 1). In the rabbit, ribosomal proteins S20 and L5, splicing factor p55, α-tubulin, a membrane transport protein VDAC2, and chaperonin are transcribed at the 8- to 16-cell stage (late ZGA). Expression of all of these genes is α-amanitin sensitive, and therefore presumed to be RNA polymerase II promoters.

Although less well characterized, RNA polymerase I and III dependent gene expression follows the same time course as RNA polymerase II dependent gene expression. Pol I, II, and III promoters are all recognized in late one-cell mouse embryos. Small nuclear RNAs U1 to U5, whose synthesis is dependent on pol II, as well as U6 whose synthesis depends on pol III, begin to accumulate at the two-cell stage in mice, concurrent with late ZGA. Similarly, U2 begins to accumulate at the 8- to 16-cell stage in cows. Changes in nucleolar morphology and nucleolar protein markers that provide a sensitive indicator of pol I dependent ribosomal gene expression are localized to the middle of the two-cell stage during mouse development, and the earliest stage at which rRNA synthesis has been detected is the mid to late two-cell stage. These results are consistent with the hypothesis that the high rate of ribosome assembly that begins in two-cell embryos requires the coordinate expression of pol I dependent rRNA synthesis and pol II dependent expression of genes for ribosomal proteins. In pigs and cows, rRNA genes are also activated at the four-cell and eight-cell stages, respectively, concurrent with late ZGA. Thus, it is clear that the transition from maternal to zygotic gene expression is a highly orchestrated sequence of events, spanning the first 40 h in the life of a fertilized mouse egg, and up to several days in other animals.

3 Zygotic Clock

In organisms that undergo rapid nuclear division after fertilization (e.g. amphibians, fish, echinoderms, flies), ZGA is determined by the ratio of nuclei to cytoplasm. Thus, ZGA occurs after a preset number of nuclear divisions have occurred, regardless of whether cellularization occurs (e.g. 13-cell cleavages in *Xenopus*; 10 nuclear divisions in the syncytium of *Drosophila*). However, in mice, ZGA is a time-dependent event that is delayed for about 24 hpf, and therefore begins after formation of a two-cell embryo (Fig. 1). The fact that ZGA is independent of DNA synthesis, cytokinesis, and the ratio of cytoplasm to nuclei allows ZGA to begin in S-phase-arrested one-cell embryos after they have advanced temporally to the "two-cell stage" (Fig. 1). The mechanism that determines when ZGA begins (referred to as the "zygotic clock") is not simply the time required to convert sperm and egg chromatin into a transcribable form, because it also regulates expression of injected plasmid-encoded genes utilizing RNA polymerase I, II or III dependent promoters. Therefore, the zygotic clock must involve delayed expression or activation of one or more transacting factors required for transcription and/or translation. Nevertheless, early *ZGA* genes are unique in that they begin translation immediately following the first mitosis, whereas plasmid-encoded genes are not translated until late ZGA, even though they are transcribed when injected into late one-cell embryos. Therefore, the zygotic clock regulates both transcription and translation of mRNA. Furthermore, the fact that plasmid-encoded genes are expressed during late ZGA reveals that

early *ZGA* genes are not simply more accessible to transcription factors than late *ZGA* genes; early ZGA genes are in some way predestined for immediate expression, perhaps by association with the transcription machinery.

3.1 Protein Translocation

Following fertilization, transcription factor Sp1, TATA-box binding protein, and RNA polymerase II all undergo a time-dependent concentration in the pronuclei, and both Sp1 mRNA and transcription factor activity increases markedly from the one-cell to the four-cell stage (Fig. 1). Both Sp1 and HSF1 are present in the pronuclei of one-cell embryos and are required for Hsp70.1 gene expression during early ZGA, but translocation of transcription factors to the nucleus cannot be the only component of the zygotic clock because, in contrast to ZGA, the translocation of these proteins does not require concomitant protein synthesis.

3.2 Protein Phosphorylation

In mice, rabbits, and frogs, the carboxyl-terminal domain (CTD) of the RNA pol II catalytic subunit is hyperphosphorylated and transcriptionally inactive in mature eggs (RNA pol IIo) (Fig. 1). Fertilization triggers dephosphorylation of the CTD into a hypophosphorylated form IIa that is required for initiation of transcription, and a reduced phosphorylated form IIe characteristic of somatic cells. Form IIa appears in late one-cell embryos prior to early ZGA in both mice and rabbits but form IIe is not observed until late ZGA. Late ZGA occurs in late 2-cell mouse embryos, at the 8- to 16- cell stage in rabbits, and at the midblastula transition in frogs. Like ZGA, these changes in phosphorylation require concomitant protein synthesis, and they may also account for the observation that ZGA in the mouse requires protein kinase activity. Posttranslational modifications of other proteins required for transcription, such as Sp1 and TBP, may also contribute to the zygotic clock mechanism, but this remains more speculative.

3.3 Translational Control

Maternal mRNA degradation is triggered by meiotic maturation and is about 90% completed in two-cell embryos, although maternal protein synthesis continues into the eight-cell stage. Nevertheless, some of the newly synthesized proteins are made from maternally inherited transcripts that are polyadenylated after fertilization and then translated. For example, newly synthesized histone H1, *Tead2*, Sp1, TBP, and cyclin-A are first translated from maternal mRNA in late one-cell or early two-cell embryos and then from zygotic mRNA during the late two-cell to four-cell stage. Such proteins are presumably required for chromatin remodeling and cell proliferation. As much as one-third of the mRNAs in two-cell embryos appear to contain a cytoplasmic polyadenylation element.

Despite the fact that transcription is first detected in G2-phase one-cell embryos, fertilized mouse eggs can delay expression of zygotic genes by uncoupling translation from transcription. Injection of plasmid-encoded genes revealed that the time courses for nascent mRNA accumulation from pol I, II or III dependent promoters are biphasic, with the second phase of transcription occurring just prior to late ZGA. An RNA polymerase II dependent gene can be transcribed prior to the

first mitosis, but it is not translated until zygotic gene expression begins (up to 15 h later). At this time, translation of RNA pol II transcripts becomes tightly linked to transcription (the second phase). Thus, the small amount of RNA synthesis that occurs in late one-cell embryos may simply represent premature transcription resulting from the appearance of transcription machinery before the parental genomes are completely masked by chromatin structure. In fact, one or more factors capable of repressing DNA transcription and replication are absent from one-cell embryos; they are produced just prior to ZGA. Thus, by delaying transcription and by preventing any transcripts that are made from being translated, the zygotic clock provides a window of opportunity for remodeling parental chromosomes in the absence of gene expression.

4 Developmental Acquisition of Transcriptional Regulation

4.1 Chromatin-mediated Repression

Transient expression from plasmid-encoded reporter genes and nuclear transplantation experiments have revealed the presence of transacting factors that can repress the activities of promoters and replication origins from 20- to >500-fold as fertilized mouse eggs develop into two-cell and four-cell embryos. The same appears to be true in rabbits, despite the fact that ZGA in rabbits does not occur until the 8- to 16-cell stage. While maternal nuclei in both oocytes and one-cell embryos exhibit this repression, the paternal pronucleus exhibits repression only when one-cell embryos develop beyond S-phase (Fig. 2). Sperm chromatin, which contains protamines but not core histones H2B and H3, becomes dispersed about 1 h after fertilization, and then within the following 7 h protamines are replaced by core histones and the DNA is condensed, the whole process being completed prior to DNA replication. Paternal pronuclei generally replicate more quickly than maternal pronuclei, which inherit a full set of histones from the oocyte and exhibit a different pattern of histone acetylation. Thus, paternal pronuclei may exhibit an ''unrepressed'' chromatin state while maternal pronuclei exhibit a ''repressed'' chromatin state.

This hypothesis is supported by observations that expression of microinjected genes in paternal pronuclei is ~5-fold greater than in maternal pronuclei, and inhibitors of histone deacetylase stimulate transcription in maternal pronuclei while reducing transcription in paternal pronuclei. The factor(s) responsible for this repression is absent from the cytoplasm of early one-cell embryos; it is not simply excluded from the paternal pronucleus. It is produced in the cytoplasm sometime between S-phase in a one-cell embryo and formation of a two-cell embryo where it can operate within any nucleus, regardless of its parental origin or ploidy. Repression increases as development proceeds from the two-cell to the four-cell stage. Such changes in chromatin structure could account for the fact that Hsp70.1 expression, which is constitutive in two-cell embryos, is repressed to a basal level during the four- to eight-cell stage.

These observations correlate well with changes in the synthesis and modification of chromatin-bound histones at the beginning of mouse development (Fig. 2). Oocytes, which can repress microinjected

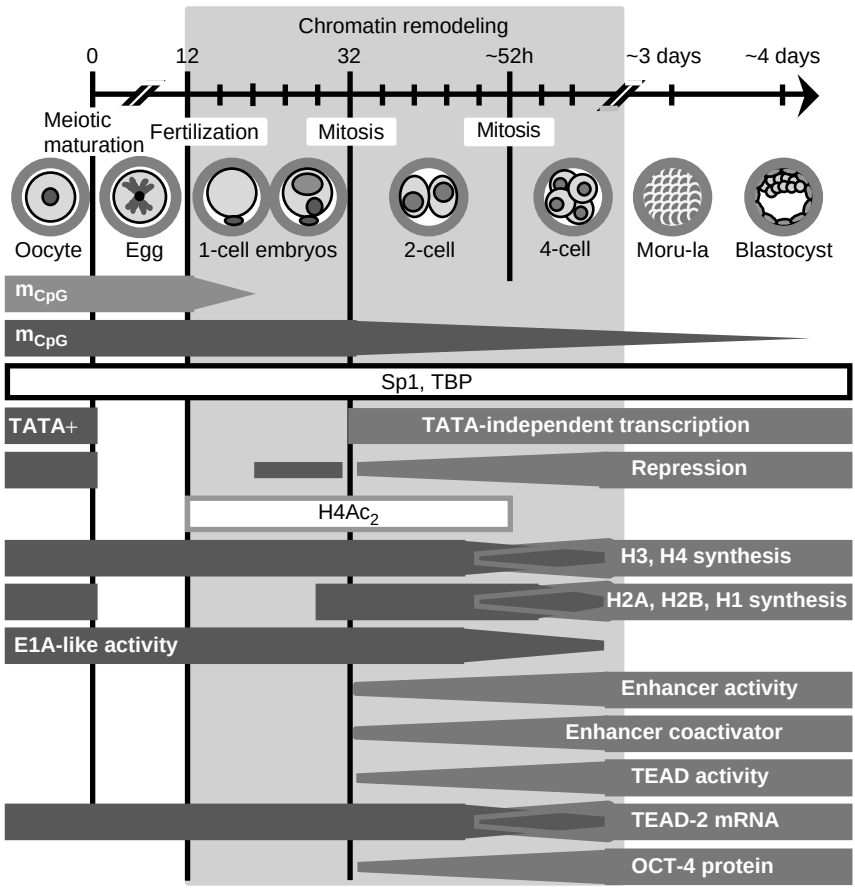

Fig. 2 Developmental changes affecting regulation of gene expression at the beginning of mouse development. These include changes in DNA methylation of the paternal and maternal genomes, changes in histone synthesis and modification that lead to chromatin-mediated repression, acquisition of chromatin-mediated repression, and acquisition of enhancer function. Some transcription factors such as Sp1 and TBP are ubiquitous. TATA-box function appears to be restricted to differentiated cells. Color coding and bar coding are the same as in Fig. 1 (see color plate p. xxii).

promoter activity, synthesize a full complement of histones; in addition, histone synthesis, which originates from mRNA inherited from the oocyte, can be observed up to the early two-cell stage. However, while histones H3 and H4 continue to be synthesized in early one-cell embryos, the synthesis of histones H2A, H2B, and H1 (proteins required for chromatin condensation) is delayed until the late one-cell stage, reaching its maximum rate in early two-cell embryos. Moreover, histone H4 in both one-cell and two-cell embryos is predominantly diacetylated (a modification that facilitates DNA transcription and chromatin assembly). Deacetylation toward the unacetylated and monoacetylated H4 population observed in differentiated

cells, such as the fibroblasts, begins at the late two-cell to four-cell stage. These changes correlate with the establishment of chromatin-mediated repression during formation of a two-cell embryo as well as the increase in repression from the two-cell to four-cell stage, where accumulation of linker histone H1 and deacetylation of core histones are observed. The arresting of development at the beginning of S-phase in one-cell embryos prevents both the appearance of chromatin-mediated repression of transcription in paternal pronuclei and synthesis of new histones. The pattern of hyperacetylated histones throughout the nuclei also changes as embryos progress from the one-cell to the four-cell stage, but the significance of these changes remains speculative.

Similar changes in histone composition occur during frog development. Histone H4 is stored in a diacetylated form in *Xenopus* eggs and then progressively deacetylated after ZGA begins at the blastula stage. Histone deacetylase inhibitors can then induce expression of specific genes. The type of linker histone changes from the maternal histone H1 variant (B4) at the midblastula transition to the somatic histone H1 variant at the end of gastrulation, resulting in specific repression of oocyte 5 S RNA gene expression. The initial activation of H1 synthesis occurs entirely by activation of maternal transcripts that then disappear by the early gastrula stage. Thus, the induction of chromatin-mediated repression may be a common feature of the maternal-to-zygotic gene transition in all metazoa.

Oocytes and early embryos of many nonmammalian species lack the somatic cell form of the linker histone H1; somatic H1 was originally believed to be absent from mouse and bovine oocytes and then first expressed at the four- to eight-cell stage of development. mRNA for H1 variants could be detected in these cells, and $H1^0$ protein could be detected in oocytes, suggesting that $H1^0$, rather than somatic H1, was present during ZGA. However, the incorporation of radiolabeled amino acids revealed that somatic H1 is indeed synthesized in mouse oocytes and in preimplantation embryos starting from the late one-cell stage. Moreover, improved methods of immunological detection confirmed that somatic H1 is present throughout this period, although it is not present on maternal metaphase II chromatin. Upon formation of pronuclear envelopes, somatic H1 is rapidly incorporated onto maternal and paternal chromatin, and the amount of somatic H1 on embryonic chromatin steadily increases to the eight-cell stage. These data suggest that the major changes in histone H1 that accompany ZGA involve changes in the amount and location of somatic H1 rather than in selective expression of H1 variants. Nevertheless, a maternally derived H1 variant (H1oo) that bears significant homology with the oocyte-specific *Xenopus* B4 histone and sea urchin cs-H1 histone is specifically expressed in mouse oocytes – one-cell and two-cell embryos – but disappears by the four- to eight-cell stages. Thus, while somatic H1 appears responsible for global changes in chromatin structure, selected genes may be regulated by oocyte-specific H1 variants.

Manipulation of the ratio of chromatin-bound proteins to DNA can alter ZGA. For example, injection of somatic cell histone H1 into mouse one-cell embryos retards the onset of RNA polymerase II transcription in late one-cell embryos (phase I of ZGA), presumably by increasing chromatin condensation, while injection of HMG-1 (a protein that binds AT-rich sequences and that can modulate

gene expression) advances it. Similarly, the depletion of HMG-14 and HMG-17 – two chromosomal proteins that generally stimulate transcription – by injecting one-cell embryos with antisense oligonucleotides delays, but does not prevent, preimplantation development. The fact that these experimental insults did not prevent development of one-cell embryos into blastocysts suggests that cells recover by degrading injected proteins and antisense oligonucleotides.

Global inhibition of ZGA by H1 does not affect all genes, because injection of histone H1 into one-cell embryos does not inhibit expression of the "transcription-requiring complex" of proteins (phase II of ZGA). This may also explain why neither the injection of H1 nor of HMG-1 had a significant effect on the development of one-cell embryos. Nevertheless, coinjection of histones H1, H2A, H2B, H3, and H4 together with a plasmid-encoded reporter gene into the paternal pronuclei of one-cell embryos can recreate the behavior observed when plasmid-encoded genes are injected into the nuclei of two-cell embryos: the injected gene is repressed and this repression can be relieved by including either a functional enhancer or histone deacetylase inhibitors. In these experiments, the effect of H1 was to reduce the amount of core histones needed to repress gene expression. Thus, chromatin-mediated repression that can be relieved by embryo-responsive enhancers (phase III of ZGA) can be made to occur in one-cell paternal pronuclei simply by assembling the DNA into somatic cell chromatin.

4.2 Enhancer Function

Direct evidence that repression of promoter activity at the beginning of mouse development is mediated through chromatin structure comes from the effects of inhibitors of histone deacetylases and enhancers on gene expression. Repression can be relieved either by treating cells with inhibitors of histone deacetylase or, in cleavage stage embryos (two-cell embryos to morula), by linking the promoter or replication origin to an embryo-responsive enhancer. Inhibitors of histone deacetylase, such as trichostatin A and sodium butyrate, increase the fraction of hyperacetylated core histones and thereby stimulate transcription from specific genes. This is consistent with the fact that transcriptionally active eukaryotic genes are generally associated with acetylated core histones. In mouse one-cell embryos, butyrate can stimulate promoter activity on genes injected into the maternal pronucleus ("repressed"), but not into the paternal pronucleus ("unrepressed"). Thus, stimulation is caused by changes in chromatin structure of the injected plasmid and not by increased synthesis of transcription factors. If the latter were the cause of stimulation, the butyrate would affect both pronuclei.

Enhancers can substitute for butyrate in stimulation of promoter activity, but only after formation of a two-cell embryo (Fig. 2). Enhancers are inactive in oocytes and one-cell embryos even when the appropriate sequence-specific activation protein required for enhancer activity is present (e.g. GAL4:VP16, Tead, Sp1), because an enhancer-specific coactivator activity is not produced until ZGA. This coactivator activity works on most, if not all, enhancers and likely represents one or more of the proteins identified in both yeast and mammalian cells that appears to mediate the interaction of upstream sequence-specific transcription factors with the RNA pol II complex.

In vitro, enhancers do not stimulate transcription unless the DNA substrate is organized into chromatin. *In vivo*, enhancers have little, if any, effect on promoters injected into two-cell embryos cultured in butyrate. Therefore, the primary role of enhancers is not simply to provide additional transcription factors to facilitate formation of an active initiation complex, but to relieve repression of weak promoters by chromatin (Fig. 3). Direct proof for this hypothesis comes from coinjection of purified histones and a plasmid-encoded reporter gene into the paternal pronuclei of one-cell embryos. Acquisition of chromatin-mediated repression of the injected promoter and subsequent relief of this repression either

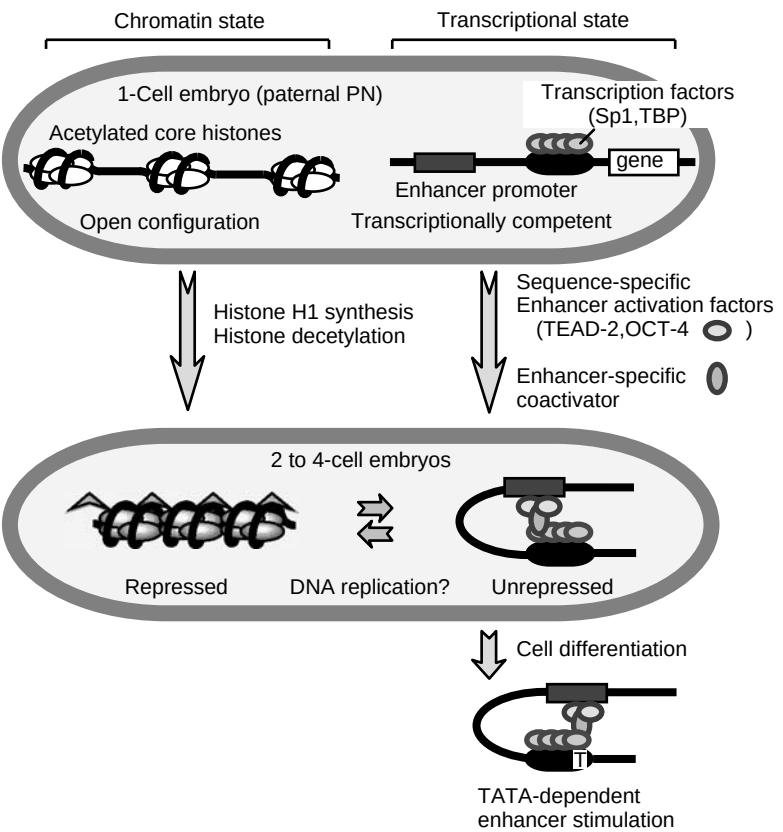

Fig. 3 Relationships between chromatin structure, promoter/enhancer activity and DNA replication during the maternal-to-zygotic gene transition in the mouse. Chromatin in the paternal pronucleus is in an open configuration that does not suppress promoter activity. Formation of a two-cell embryo is accompanied by changes in chromatin structure that repress promoter activity. This repression can be relieved by enhancer activity that requires sequence-specific-enhancer binding proteins such as Tead2 or Oct-4, and an as-yet-unidentified enhancer-specific coactivator, all of which first become available during ZGA. DNA replication may facilitate activation of some genes. Cell differentiation is accompanied by the need for a TATA-box element in the promoter to enable enhancers to function.

by functional enhancers or by histone deacetylase inhibitors occurred at a specific ratio of histones to DNA. The extent of the enhancer-mediated stimulation was found to be inversely related to the acetylation status of the histones and dependent on the developmental acquisition of the enhancer-specific coactivator activity.

The results described above reveal that the "zygotic clock" delays ZGA in mice by preventing the bulk of transcription and translation until two-cell embryos have entered G2-phase (36 to 40 hpf). By that time, chromatin-mediated repression has begun and the ability to use enhancers has been acquired. The net result is a global repression of mammalian gene expression, thus permitting specific genes to be activated at specific times and in specific tissues through the expression of specific promoter and enhancer activation proteins that activate their promoter/enhancer, as well as by specific proteins that relieve chromatin-mediated repression (e.g. histone acetyltransferase, GAGA, SWI/SNF complex).

4.3 DNA Replication

As discussed above, the ability of chromatin structure to repress DNA transcription or replication, and the ability of enhancers to relieve this repression first appear in mouse development with formation of two-cell and four-cell embryos. However, enhancers alone cannot always relieve chromatin-mediated repression. Once a repressed state is formed, it may be necessary for DNA to replicate in order to reprogram itself into a transcriptionally active state (Fig. 3). When DNA is injected into either pronucleus of one-cell embryos, and the injected embryo then undergoes mitosis to form a two-cell embryo, the injected promoter becomes "irreversibly" repressed, in that neither enhancers nor butyrate can restore its activity. This is not due to loss of plasmid DNA from the injected pronucleus during mitosis, because repression is reversible when the injected pronucleus is transplanted to a two-cell embryo that then undergoes mitosis. Therefore, something happens to DNA between completion of S-phase in a one-cell embryo and formation of a two-cell embryo, which prevents activation of injected genes while allowing embryonic genes to undergo ZGA. One explanation is that plasmid DNA does not replicate when injected into mouse embryos unless it contains a viral replication origin, whereas the genome of a one-cell embryo undergoes one round of replication prior to early ZGA and two rounds prior to late ZGA. DNA replication may be required to restore the newly remodeled zygotic genome to a transcriptionally competent state. Chromatin assembly in one-cell embryos occurs in the absence of at least one factor required for enhancer function; this factor does not appear until the two-cell stage. Therefore, if chromatin-mediated repression begins in late one-cell embryos, before enhancers are functional, DNA replication may be required to disrupt the repressed state so that appropriate transcription factors can bind. Conversely, once an enhancer has acted to prevent repression of its adjunct promoter, the resulting transcription complex may remain active until replication again allows reprogramming. Thus, the fraction of genes encoded by plasmid DNA that are "on" or "off" will depend on the relative amounts of repressor *versus* enhancer activation proteins present at the time of injection.

Recent evidence provides some support for this hypothesis. Complete inhibition of DNA synthesis in one-cell embryos with

aphidicolin (a specific inhibitor of replicative DNA polymerases) does not prevent early ZGA, but it does reduce the amount of transcription observed by nearly one-third. It also inhibits transcription of the eIF-4C gene by nearly one-half and the synthesis of the "transcription-requiring complex" (TRC) of proteins by ~90%. These results suggest that expression of some of the early *ZGA* genes is dependent on prior replication of some or all of the genome. The effect, however, is overestimated by measuring TRC synthesis, because inhibition of DNA replication in one-cell embryos is accompanied by a general decrease in total protein biosynthesis that accompanies the decrease in TRC synthesis. Nevertheless, for integrated transgenes, enhancer activity first occurs at the two- or four-cell stage, but only after completion of DNA replication., suggesting that DNA replication is required to relieve chromatin repression. Thus, the possibility that disruption of chromatin structure by DNA replication permits reprogramming of gene expression remains an attractive hypothesis.

4.4 TATA-box Function

Once ZGA begins, the requirements for transcription can differ from those observed at later stages in development. In yeast, proteins binding to upstream activator sequences can facilitate binding of TBP to the TATA-box, thus accelerating transcription. Moreover, the TATA-box element common to many RNA pol II promoters appears to be required only when a promoter is stimulated either by an upstream enhancer or by a transacting factor equivalent to herpesvirus ICP4 or adenovirus E1A proteins. However, prior to cell differentiation, enhancer stimulation is observed in the absence of TATA-box cell differentiation. This would imply that the TATA-less promoters that drive transcription of typical housekeeping genes can be stimulated selectively in cleavage stage embryos, but not in differentiated cells (Figs. 2, 3). In fact, oocytes are unique in that they express an E1A-like activity that may serve to stimulate genes in place of enhancer activity (Fig. 2).

4.5 DNA Methylation

The bulk of DNA methylation in eukaryotes consists of 5-methyl cytosines within CpG dinucleotides. In mammals, three families of DNA methyltransferases (Dnmt 1, 2, and 3) have been identified. Dnmt-1 is the enzyme responsible for maintaining DNA methylation patterns during DNA replication by converting hemimethylated CpG dinucleotides into fully methylated CpG dinucleotides. Dnmt-3 is involved in *de novo* methylation. The role of Dnmt-2 is not yet known. In addition to these enzymes, five proteins have been identified that bind to ^{m}CpGs (MeCP1 and 2, and MBD proteins 1 to 4; MBD-2 is the ^{m}CpG-binding protein in the multiprotein complex MeCP1). These proteins are associated with ^{m}CpGs and are believed to be involved in gene silencing.

Clearly, DNA methylation is required for proper embryonic development. For example, loss of Dnmt-1 results in embryonic stem (ES) cells that proliferate normally with their DNA highly demethylated but that die upon differentiation, with the consequence that embryos lacking this enzyme are delayed in development, and do not survive past midgestation. Similarly, mice lacking MBD-3 die during early embryogenesis. The precise reasons for these effects, however, and the identity of

the factors that control methylation pattern dynamics during gametogenesis and early development, are largely unknown.

Global methylation patterns are erased in primordial germ cells and then reestablished in sex-specific patterns in mature male and female germ cells. There is an additional wave of demethylation during preimplantation development, although some sequences, notably certain imprinted genes, retain gametic methylation patterns at all stages. While primordial germ cells are highly methylated when compared to somatic cells, both male and female primordial germ cells undergo global demethylation that is completed by embryonic day 13 or 14. Both imprinted genes and single-copy genes are demethylated at this stage. Whether this demethylation step involves active or passive demethylation is unknown. However, when embryonic germ cells are fused with somatic cells, somatic nuclei undergo demethylation, suggesting the existence of factor(s) in germs cells that confer demethylation in "trans." In the male germ cells, remethylation takes place at the prospermatogonia stage (E15/16 onward), which precedes the mitosis/meiosis step. In contrast, remethylation of female primordial germ cells occurs during the growth of oocytes in young animals. These demethylation/remethylation steps are thought to be required in order to reset the imprinting cues and reprogram epigenetic modifications in germ cells.

Demethylation and remethylation also take place during embryogenesis (Fig. 2). The paternal genome is actively demethylated within the first 8 h after fertilization, prior to DNA replication. Although the nature of the demethylating mechanism is not yet clear, it would appear that chromatin remodeling of sperm DNA exposes it to enzymes that can demethylate it, while the chromatin structure of oocyte DNA protects it from demethylation. In contrast to the paternal genome, the maternal genome is passively demethylated by DNA replication in the absence of Dnmt-1 activity, as cells undergo cleavage from the 2-cell to the 32-cell (morula) to blastula stages. Dnmt-1 is localized in the cytoplasm during preimplantation development. CpG methylation activity returns after implantation, and confers methylation patterns similar to what is observed in somatic cells.

DNA methylation patterns associated with imprinted genes appear to be shielded from demethylation during preimplantation development. Part of the mechanism that prevents loss of these methylation imprints involves a novel form of Dnmt-1. Mouse oocytes and preimplantation embryos lack Dnmt1, but they express a variant called Dnmt-1o. After fertilization, this extremely abundant form of Dnmt1 is relegated to the cytoplasm until the eight-cell stage when, for a single cell cycle, it enters the nucleus. Dnmt-1o is then excluded from the nucleus in subsequent cell cycles until postimplantation day eight.

Although genomic methylation patterns are established normally in Dnmt-1o-deficient oocytes, embryos derived from such oocytes show a loss of allele-specific expression and methylation at certain imprinted loci. Experiments with mice lacking functional Dnmt-1o suggest that the maintenance of imprinted methylation is critically dependent on the role of this enzyme during the eight-cell stage. Here the absence of Dnmt-1o results in failure of the newly replicated paternal strand to become methylated. Subsequently maintenance methylation occurs normally, but there is a 50% reduction of methylation on newly replicated paternal alleles. This heritable change results in some cells having

either differentially methylated parental alleles or ones that are unmethylated. Thus, Dnmt-1o is not required for the establishment of maternal genomic imprints, but instead appears to be required specifically for maintenance methylation of imprinted loci.

While CpG methylation has been directly linked to a number of biological phenomena including X-chromosome inactivation, imprinting, and silencing of repeat sequences and retroviruses, its direct role in regulating gene expression has yet to be demonstrated. Methylation of CpG dinucleotides is commonly correlated with a loss of gene expression both *in vivo* and *in vitro*. Nevertheless, the absence of a change in the DNA methylation pattern of several tissue-specific genes during early mouse development has led to the hypothesis that CpG methylation is a consequence of the absence of transcription rather than the cause, and thereby serves to insure that repressed genes remain silent.

What is the purpose of these demethylation/remethylation events in preimplantation embryos? One general consequence of global demethylation is to remove one of the major obstacles to the binding of proteins to DNA. These proteins may include repressor proteins whose binding is abrogated by methylation. With the exception of the five proteins known to bind specifically to ^{m}CpG dinucleotides, the affinity for DNA of all other DNA binding proteins appears to be reduced by methylation of their DNA binding sites. Thus, demethylation may facilitate remodeling of sperm chromatin into somatic cell chromatin in one-cell embryos, and it may allow more subtle remodeling to take place in both paternal and maternal genomes during the preimplantation period.

DNA methylation may also prevent specific genes from being activated during ZGA, because one characteristic of genes that are destined to be transcribed during ZGA appears to be the absence of DNA methylation at their regulatory regions in sperm. For example, the regulatory regions of the *Sgy* and *Tead2* genes (Table 1) are not methylated in either sperm or oocytes. A number of genes that are expressed during ZGA are expressed only from the paternal allele (Table 1), and in the case of U2afbp-rs and Xist, these alleles enter development in an undermethylated state. DNA methylation patterns associated with imprinted genes appear to be shielded from demethylation during preimplantation development. In contrast, the tissue-specific mouse skeletal α-actin promoter is fully methylated in the sperm and essentially unmethylated in the oocyte.

4.6 Differential Gene Expression

The average spacing of genes in human cells has been estimated to be 85 kb. However, the number of closely spaced, divergently expressed genes is surprisingly high. In some cases, two closely spaced genes are transcribed in the same cell (coordinately expressed), whereas in other cases, only one of the two genes is transcribed in a particular cell type (differentially expressed). A common feature of these bidirectional loci is the presence of a CpG island, suggesting that DNA methylation may regulate their expression during animal development. However, one detailed analysis of such a locus has revealed (1) that differential expression of closely spaced genes is developmentally acquired, and (2) that DNA methylation restricts, but does not dictate, differential gene expression during mouse development.

Tead2 is the only member of the *Tead* gene family of transcription factors that is expressed in mouse embryos during the first seven days of development suggesting that it plays a unique role at the beginning of mammalian development, when it is presumed to be responsible for the ability of preimplantation mouse embryos to utilize Tead-dependent promoters and enhancers. In adult animals, *Tead2* continues to be expressed to varying extents in most cell types. Soggy (*Sgy*) is a novel single-copy gene whose mRNA start site is located only 3.8-kb upstream of the *Tead2* mRNA start site. The function of *Sgy* is unknown, but it is presumed important because its expression pattern in the adult is restricted to lymphocytes and pre-spermatocytes, and it is partially homologous to the Dickkopf gene family, potential effectors of the Wnt signaling pathway. Since *Tead2* and *Sgy* are transcribed in opposite directions, their regulatory elements lie in close proximity. In fact, the same locus is found in humans on chromosome 19q13.3, except that the two mRNA start sites are separated by only 1.5 kb. Both mRNA start sites lie within CpG islands.

Both *Sgy* and *Tead2* are expressed concurrently during zygotic gene activation in mouse preimplantation embryos and in embryonic stem cells where their regulatory regions are unmethylated. However, with the onset of cell differentiation and DNA methylation, cells express one gene or the other, but not both. This differential gene expression is restricted by DNA methylation, because neither gene is expressed in cells in which its regulatory region is methylated, and *Sgy* expression in permissive cells is inversely related to the extent of methylation in the *Sgy* regulatory region. Consequently, a sharp transition between unmethylated and methylated DNA marks the upstream boundary of each active promoter. Moreover, recovery of *Sgy* expression is accompanied by demethylation of the *Sgy* promoter. Nevertheless, DNA methylation *per se* does not determine expression at the *Sgy*/*Tead2* locus, because some cells do not express one of these genes even when its regulatory sequences are unmethylated. Thus, while DNA methylation can restrict gene expression during mammalian development, it is not the primary regulatory mechanism. Therefore, the developmental acquisition of differential gene expression must depend primarily on the presence of gene-specific activators and repressors.

See also Genomic Imprinting, Molecular Genetics of; Heterochromatin and Eurochromatin – Organization, Packaging, and Gene Regulation; Homeodomain Proteins.

Bibliography

Books and Reviews

Adachi, N., Lieber, M.R. (2002) Bidirectional Gene Organization: a common architectural feature of the human genome, *Cell* **109**, 807–809.

Latham, K.E. (1999) Mechanisms and control of embryonic genome activation in mammalian embryos, *Int. Rev. Cytol.* **193**, 71–124.

Latham, K.E., Schultz, R.M. (2001) Embryonic genome activation, *Front. Biosci.* **6**, D748–D759.

Majumder, S., DePamphilis, M.L. (1995) A unique role for enhancers is revealed during early mouse development, *Bioessays* **17**, 879–889.

Nothias, J.Y., Majumder, S., Kaneko, K.J., DePamphilis, M.L. (1995) Regulation of gene expression at the beginning of mammalian development, *J. Biol. Chem.* **270**, 22077–22080.

Reik, W., Dean, W., Walter, J. (2001) Epigenetic reprogramming in mammalian development, *Science* **293**, 1089–1093.

Schultz, G.A., Heyner, S. (1992) Gene expression in preimplantation mammalian embryos, *Mutat. Res.* **296**, 17–31.

Schultz, R.M. (1993) Regulation of zygotic gene activation in the mouse, *Bioessays* **15**, 531–538.

Schultz, R.M. (2002) The molecular foundations of the maternal to zygotic transition in the preimplantation embryo, *Hum. Reprod. Update*. **8**, 323–331.

Schultz, R.M., Davis, W., Jr., Stein, P., Svoboda, P. (1999) Reprogramming of gene expression during preimplantation development, *J. Exp. Zool.* **285**, 276–282.

Primary Literature

Adachi, N., Lieber, M.R. (2002) Bidirectional Gene Organization: a common architectural feature of the human genome, *Cell* **109**, 807–809.

Adenot, P.G., Mercier, Y., Renard, J.P., Thompson, E.M. (1997) Differential H4 acetylation of paternal and maternal chromatin precedes DNA replication and differential transcriptional activity in pronuclei of 1-cell mouse embryos, *Development* **124**, 4615–4625.

Adenot, P.G., Campion, E., Legouy, E., Allis, C.D., Dimitrov, S., Renard, J., Thompson, E.M. (2000) Somatic linker histone H1 is present throughout mouse embryogenesis and is not replaced by variant H1 degrees, *J. Cell. Sci.* **113**, 2897–2907.

Almouzni, G., Khochbin, S., Dimitrov, S., Wolffe, A.P. (1994) Histone acetylation influences both gene expression and development of Xenopus laevis, *Dev. Biol.* **165**, 654–669.

Aoki, E., Schultz, R.M. (1999) DNA replication in the 1-cell mouse embryo: stimulatory effect of histone acetylation, *Zygote* **7**, 165–172.

Aoki, F., Worrad, D.M., Schultz, R.M. (1997) Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo, *Dev. Biol.* **181**, 296–307.

Baran, V., Brochard, V., Renard, J.P., Flechon, J.E. (2001) Nopp 140 involvement in nucleologenesis of mouse preimplantation embryos, *Mol. Reprod. Dev.* **59**, 277–284.

Beaujean, N., Bouniol-Baly, C., Monod, C., Kissa, K., Jullien, D., Aulner, N., Amirand, C., Debey, P., Kas, E. (2000) Induction of early transcription in one-cell mouse embryos by microinjection of the nonhistone chromosomal protein HMG-I, *Dev. Biol.* **221**, 337–354.

Bellier, S., Chastant, S., Adenot, P., Vincent, M., Renard, J.P., Bensaude, O. (1997) Nuclear translocation and carboxyl-terminal domain phosphorylation of RNA polymerase II delineate the two phases of zygotic gene activation in mammalian embryos, *EMBO J.* **16**, 6250–6262.

Bestor, T.H. (2000) The DNA methyltransferases of mammals, *Hum. Mol. Genet.* **9**, 2395–2402.

Bevilacqua, A., Fiorenza, M.T., Mangia, F. (1997) Developmental activation of an episomic hsp70 gene promoter in two-cell mouse embryos by transcription factor Sp1, *Nucleic Acids Res.* **25**, 1333–1338.

Bouniol, C., Nguyen, E., Debey, P. (1995) Endogenous transcription occurs at the 1-cell stage in the mouse embryo, *Exp. Cell Res.* **218**, 57–62.

Bouniol-Baly, C., Nguyen, E., Besombes, D., Debey, P. (1997) Dynamic organization of DNA replication in one-cell mouse embryos: relationship to transcriptional activation, *Exp. Cell Res.* **236**, 201–211.

Bouvet, P., Dimitrov, S., Wolffe, A.P. (1994) Specific regulation of Xenopus chromosomal 5S rRNA gene transcription in vivo by histone H1, *Genes Dev.* **8**, 1147–1159.

Brandeis, M., Ariel, M., Cedar, H. (1993) Dynamics of DNA methylation during development, *Bioessays* **15**, 709–713.

Brunet-Simon, A., Henrion, G., Renard, J.P., Duranthon, V. (2001) Onset of zygotic transcription and maternal transcript legacy in the rabbit embryo, *Mol. Reprod. Dev.* **58**, 127–136.

Cardoso, M.C., Leonhardt, H. (1999) DNA methyltransferase is actively retained in the cytoplasm during early development, *J. Cell Biol.* **147**, 25–32.

Carlson, L.L., Page, A.W., Bestor, T.H. (1992) Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting, *Genes Dev.* **6**, 2536–2541.

Christians, E., Rao, V.H., Renard, J.P. (1994) Sequential acquisition of transcriptional control during early embryonic development in the rabbit, *Dev. Biol.* **164**, 160–172.

Christians, E., Campion, E., Thompson, E.M., Renard, J.P. (1995) Expression of the HSP 70.1 gene, a landmark of early zygotic activity in the

mouse embryo, is restricted to the first burst of transcription, *Development* **121**, 113–122.

Christians, E., Michel, E., Adenot, P., Mezger, V., Rallu, M., Morange, M., Renard, J.P. (1997) Evidence for the involvement of mouse heat shock factor 1 in the atypical expression of the HSP70.1 heat shock gene during mouse zygotic genome activation, *Mol. Cell. Biol.* **17**, 778–788.

Clarke, H.J., Bustin, M., Oblin, C. (1997) Chromatin modifications during oogenesis in the mouse: removal of somatic subtypes of histone H1 from oocyte chromatin occurs post-natally through a post-transcriptional mechanism, *J. Cell. Sci.* **110**, 477–487.

Davis, W., Jr., Schultz, R.M. (1997) Role of the first round of DNA replication in reprogramming gene expression in the preimplantation mouse embryo, *Mol. Reprod. Dev.* **47**, 430–434.

Davis, W., Jr., Schultz, R.M. (2000) Developmental change in TATA-Box utilization during preimplantation mouse development, *Dev. Biol.* **218**, 275–283.

Dean, W.L., Seufert, A.C., Schultz, G.A., Prather, R.S., Simerly, C., Schatten, G., Pilch, D.R., Marzluff, W.F. (1989) The small nuclear RNAs for pre-mRNA splicing are coordinately regulated during oocyte maturation and early embryogenesis in the mouse, *Development* **106**, 325–334.

DePamphilis, M.L., Martinez-Salas, E., Cupo, D., Hendrickson, E.A., Fritze, C.E., Folk, W.R., Heine, U. (1988) Initiation of Polyomavirus and SV40 DNA Replication, and the Requirements for DNA Replication During Mammalian Development, in: Stillman, B., Kelly, T. (Eds.) *Eukaryotic DNA Replication*, Vol. 6, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 165–175.

Dimitrov, S., Almouzni, G., Dasso, M., Wolffe, A.P. (1993) Chromatin transitions during early Xenopus embryogenesis: changes in histone H4 acetylation and in linker histone type, *Dev. Biol.* **160**, 214–227.

Dooley, T.P., Miranda, M., Jones, N.C., DePamphilis, M.L. (1989) Transactivation of the adenovirus EIIa promoter in the absence of adenovirus E1A protein is restricted to mouse oocytes and preimplantation embryos, *Development* **107**, 945–956.

Eden, S., Cedar, H. (1994) Role of DNA methylation in the regulation of transcription, *Curr. Opin. Genet. Dev.* **4**, 255–259.

Eden, S., Constancia, M., Hashimshony, T., Dean, W., Goldstein, B., Johnson, A.C., Keshet, I., Reik, W., Cedar, H. (2001) An upstream repressor element plays a role in Igf2 imprinting, *EMBO J.* **20**, 3518–3525.

Ferreira, J., Carmo-Fonseca, M. (1997) Genome replication in early mouse embryos follows a defined temporal and spatial order, *J. Cell. Sci.* **110**, 889–897.

Flechon, J.E., Kopecny, V. (1998) The nature of the 'nucleolus precursor body' in early preimplantation embryos: a review of fine-structure cytochemical, immunocytochemical and autoradiographic data related to nucleolar function, *Zygote* **6**, 183–191.

Forlani, S., Bonnerot, C., Capgras, S., Nicolas, J.F. (1998) Relief of a repressed gene expression state in the mouse 1-cell embryo requires DNA replication, *Development* **125**, 3153–3166.

Fuchimoto, D., Mizukoshi, A., Schultz, R.M., Sakai, S., Aoki, F. (2001) Posttranscriptional regulation of cyclin A1 and cyclin A2 during mouse oocyte meiotic maturation and preimplantation development, *Biol. Reprod.* **65**, 986–993.

Hatada, I., Kitagawa, K., Yamaoka, T., Wang, X., Arai, Y., Hashido, K., Ohishi, S., Masuda, J., Ogata, J., Mukai, T. (1995) Allele-specific methylation and expression of an imprinted U2af1-rs1 (SP2) gene, *Nucleic Acids Res.* **23**, 36–41.

Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A., Bird, A. (2001) Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development, *Genes Dev.* **15**, 710–723.

Henery, C.C., Miranda, M., Wiekowski, M., Wilmut, I., DePamphilis, M.L. (1995) Repression of gene expression at the beginning of mouse development, *Dev. Biol.* **169**, 448–460.

Howell, C.Y., Bestor, T.H., Ding, F., Latham, K.E., Mertineit, C., Trasler, J.M., Chaillet, J.R. (2001) Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene, *Cell* **104**, 829–838.

Hyttel, P., Laurincik, J., Viuff, D., Fair, T., Zakhartchenko, V., Rosenkranz, C., Avery, B., Rath, D., Niemann, H., Thomsen, P.D., Schellander, K., Callesen, H., Wolf, E., Ochs, R.L., Greve, T. (2000) Activation of ribosomal RNA genes in preimplantation cattle and swine embryos, *Anim. Reprod. Sci.* **60–61**, 49–60.

Johnson, M.H., Maro, B. (1986) Time and Space in the Mouse Early Embryo: A Cell Biological Approach to Cell Diversification, in: Rossant, J., Pedersen, R.A. (Ed.) *Experimental Approaches to Mammalian Embryonic Development*, Cambridge University Press, New York, pp. 35–65.

Kandolf, H. (1994) The H1A histone variant is an in vivo repressor of oocyte-type 5S gene transcription in Xenopus laevis embryos, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7257–7261.

Kaneko, K.J., DePamphilis, M.L. (2000) Soggy, a spermatocyte-specific gene, lies 3.8 kb upstream of and antipodal to TEAD-2, a transcription factor expressed at the beginning of mouse development, *Nucleic Acids Res.* **28**, 3982–3990.

Kaneko, K.J., Cullinan, E.B., Latham, K.E., DePamphilis, M.L. (1997) Transcription factor mTEAD-2 is selectively expressed at the beginning of zygotic gene expression in the mouse, *Development* **124**, 1963–1973.

Kaneko, K.J., Rein, T., Guo, Z.S., Latham, K., DePamphilis, M.L. (2004) DNA methylation may restrict but does not determine differential gene expression at the Sgy/Tead2 locus during mouse development, *Mol. Cell. Biol.* **24**, 1968–1982.

Kawamura, K., Fukuda, J., Kodama, H., Kumagai, J., Kumagai, A., Tanaka, T. (2001) Expression of Fas and Fas ligand mRNA in rat and human preimplantation embryos, *Mol. Hum. Reprod.* **7**, 431–436.

Ko, M.S., Kitchen, J.R., Wang, X., Threat, T.A., Hasegawa, A., Sun, T., Grahovac, M.J., Kargul, G.J., Lim, M.K., Cui, Y., Sano, Y., Tanaka, T., Liang, Y., Mason, S., Paonessa, P.D., Sauls, A.D., DePalma, G.E., Sharara, R., Rowe, L.B., Eppig, J., Morrell, C., Doi, H. (2000) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, *Development* **127**, 1737–1749.

Lander, E.S., Linton, L.M., Birren, B. et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Latham, K.E. (1999) Mechanisms and control of embryonic genome activation in mammalian embryos, *Int. Rev. Cytol.* **193**, 71–124.

Latham, K.E., Rambhatla, L. (1995) Expression of X-linked genes in androgenetic, gynogenetic, and normal mouse preimplantation embryos, *Dev. Genet.* **17**, 212–222.

Latham, K.E., Schultz, R.M. (2001) Embryonic genome activation, *Front. Biosci.* **6**, D748–D759.

Latham, K.E., Solter, D., Schultz, R.M. (1992) Acquisition of a transcriptionally permissive state during the 1-cell stage of mouse embryogenesis, *Dev. Biol.* **149**, 457–462.

Latham, K.E., Garrels, J.I., Chang, C., Solter, D. (1991) Quantitative analysis of protein synthesis in mouse embryos. I. Extensive reprogramming at the one- and two-cell stages, *Development* **112**, 921–932.

Lawinger, P., Rastelli, L., Zhao, Z., Majumder, S. (1999) Lack of enhancer function in mammals is unique to oocytes and fertilized eggs, *J. Biol. Chem.* **274**, 8002–8011.

Li, E., Bestor, T.H., Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality, *Cell* **69**, 915–926.

Lin, P., Clarke, H.J. (1996) Somatic histone H1 microinjected into fertilized mouse eggs is transported into the pronuclei but does not disrupt subsequent preimplantation development, *Mol. Reprod. Dev.* **44**, 185–192.

Majumder, S., DePamphilis, M.L. (1994) TATA-dependent enhancer stimulation of promoter activity in mice is developmentally acquired, *Mol. Cell. Biol.* **14**, 4258–4268.

Majumder, S., DePamphilis, M.L. (1995) A unique role for enhancers is revealed during early mouse development, *Bioessays* **17**, 879–889.

Majumder, S., Miranda, M., DePamphilis, M.L. (1993) Analysis of gene expression in mouse preimplantation embryos demonstrates that the primary role of enhancers is to relieve repression of promoters [published erratum appears in EMBO J 1993 Oct; 12(10): 4042], *EMBO J.* **12**, 1131–1140.

Majumder, S., Zhao, Z., Kaneko, K., DePamphilis, M.L. (1997) Developmental acquisition of enhancer function requires a unique coactivator activity, *EMBO J.* **16**, 1721–1731.

Malik, S., Roeder, R.G. (2000) Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells, *Trends Biochem. Sci.* **25**, 277–283.

Martinez-Salas, E., Linney, E., Hassell, J., DePamphilis, M.L. (1989) The need for enhancers in gene expression first appears during mouse development with formation of a zygotic nucleus, *Genes Dev.* **3**, 1493–1506.

Masui, Y., Wang, P. (1998) Cell cycle transition in early embryonic development of Xenopus laevis, *Biol. Cell.* **90**, 537–548.

Matsumoto, K., Anzai, M., Nakagata, N., Takahashi, A., Takahashi, Y., Miyata, K. (1994) Onset of paternal gene activation in early mouse embryos fertilized with transgenic mouse sperm, *Mol. Reprod. Dev.* **39**, 136–140.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., Haaf, T. (2000) Demethylation of the zygotic paternal genome, *Nature* **403**, 501–502.

McLay, D.W., Clarke, H.J. (1997) The ability to organize sperm DNA into functional chromatin is acquired during meiotic maturation in murine oocytes, *Dev. Biol.* **186**, 73–84.

Melin, F., Miranda, M., Montreau, N., DePamphilis, M.L., Blangy, D. (1993) Transcription enhancer factor-1 (TEF-1) DNA binding sites can specifically enhance gene expression at the beginning of mouse development, *EMBO J.* **12**, 4657–4666.

Memili, E., First, N.L. (1999) Control of gene expression at the onset of bovine embryonic development, *Biol. Reprod.* **61**, 1198–1207.

Mohamed, O.A., Bustin, M., Clarke, H.J. (2001) High-mobility group proteins 14 and 17 maintain the timing of early embryonic development in the mouse, *Dev. Biol.* **229**, 237–249.

Nothias, J.Y., Miranda, M., DePamphilis, M.L. (1996) Uncoupling of transcription and translation during zygotic gene activation in the mouse, *EMBO J.* **15**, 5715–5725.

Nothias, J.Y., Majumder, S., Kaneko, K.J., DePamphilis, M.L. (1995) Regulation of gene expression at the beginning of mammalian development, *J. Biol. Chem.* **270**, 22077–22080.

Oh, B., Hwang, S., McLaughlin, J., Solter, D., Knowles, B.B. (2000) Timely translation during the mouse oocyte-to-embryo transition, *Development* **127**, 3795–3803.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., Walter, J. (2000) Active demethylation of the paternal genome in the mouse zygote, *Curr. Biol.* **10**, 475–478.

Palancade, B., Bellier, S., Almouzni, G., Bensaude, O. (2001) Incomplete RNA polymerase II phosphorylation in Xenopus laevis early embryos, *J. Cell. Sci.* **114**, 2483–2489.

Paranjape, S.M., Kamakaka, R.T., Kadonaga, J.T. (1994) Role of chromatin structure in the regulation of transcription by RNA polymerase II, *Annu. Rev. Biochem.* **63**, 265–297.

Pikaart, M.J., Recillas-Targa, F., Felsenfeld, G. (1998) Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators, *Genes Dev.* **12**, 2852–2862.

Rajkovic, A., Yan, M.S.C., Klysik, M., Matzuk, M. (2001) Discovery of germ cell-specific transcripts by expressed sequence tag database analysis, *Fertil. Steril.* **76**, 550–554.

Rastelli, L., Robinson, K., Xu, Y., Majumder, S. (2001) Reconstitution of enhancer function in paternal pronuclei of one-cell mouse embryos, *Mol. Cell. Biol.* **21**, 5531–5540.

Reik, W., Dean, W., Walter, J. (2001) Epigenetic reprogramming in mammalian development, *Science* **293**, 1089–1093.

Rougier, N., Bourc'his, D., Gomes, D.M., Niveleau, A., Plachot, M., Paldi, A., Viegas-Pequignot, E. (1998) Chromosome methylation patterns during mammalian preimplantation development, *Genes Dev.* **12**, 2108–2113.

Schultz, G.A., Heyner, S. (1992) Gene expression in pre-implantation mammalian embryos, *Mutat. Res.* **296**, 17–31.

Schultz, R.M. (1993) Regulation of zygotic gene activation in the mouse, *Bioessays* **15**, 531–538.

Schultz, R.M. (2002) The molecular foundations of the maternal to zygotic transition in the preimplantation embryo, *Hum. Reprod. Update* **8**, 323–331.

Schultz, R.M., Davis, W., Jr., Stein, P., Svoboda, P. (1999) Reprogramming of gene expression during preimplantation development, *J. Exp. Zool.* **285**, 276–282.

Siegfried, Z., Eden, S., Mendelsohn, M., Feng, X., Tsuberi, B.Z., Cedar, H. (1999) DNA methylation represses transcription in vivo, *Nat. Genet.* **22**, 203–206.

Smith, L.C., Meirelles, F.V., Bustin, M., Clarke, H.J. (1995) Assembly of somatic histone H1 onto chromatin during bovine early embryogenesis, *J. Exp. Zool.* **273**, 317–326.

Smith, S.S. (2000) Gilbert's conjecture: the search for DNA (cytosine-5) demethylases and the emergence of new functions for eukaryotic DNA (cytosine-5) methyltransferases, *J. Mol. Biol.* **302**, 1–7.

Stanton, J.L., Green, D.P. (2001) Meta-analysis of gene expression in mouse preimplantation embryo development, *Mol. Hum. Reprod.* **7**, 545–552.

Stein, P., Schultz, R.M. (2000) Initiation of a chromatin-based transcriptionally repressive state in the preimplantation mouse embryo:

lack of a primary role for expression of somatic histone H1, *Mol. Reprod. Dev.* **55**, 241–248.

Stein, P., Worrad, D.M., Belyaev, N.D., Turner, B.M., Schultz, R.M. (1997) Stage-dependent redistributions of acetylated histones in nuclei of the early preimplantation mouse embryo, *Mol. Reprod. Dev.* **47**, 421–429.

Tada, M., Tada, T., Lefebvre, L., Barton, S.C., Surani, M.A. (1997) Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells, *EMBO J.* **16**, 6510–6520.

Tanaka, M., Hennebold, J.D., Macfarlane, J., Adashi, E.Y. (2001) A mammalian oocyte-specific linker histone gene H1oo: homology with the genes for the oocyte-specific cleavage stage histone (cs-H1) of sea urchin and the B4/H1M histone of the frog, *Development* **128**, 655–664.

Taylor, K.D., Piko, L. (1992) Expression of ribosomal protein genes in mouse oocytes and early embryos, *Mol. Reprod. Dev.* **31**, 182–188.

Temeles, G.L., Ram, P.T., Rothstein, J.L., Schultz, R.M. (1994) Expression patterns of novel genes during mouse preimplantation embryogenesis, *Mol. Reprod. Dev.* **37**, 121–129.

Thompson, E.M., Legouy, E., Christians, E., Renard, J.P. (1995) Progressive maturation of chromatin structure regulates HSP70.1 gene expression in the preimplantation mouse embryo, *Development* **121**, 3425–3437.

Tong, Z.B., Gold, L., Pfeifer, K.E., Dorward, H., Lee, E., Bondy, C.A., Dean, J., Nelson, L.M. (2000) Mater, a maternal effect gene required for early embryonic development in mice, *Nat. Genet.* **26**, 267–268.

Tremblay, K.D., Duran, K.L., Bartolomei, M.S. (1997) A 5′ 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development, *Mol. Cell. Biol.* **17**, 4322–4329.

Tsuzuki, T., Fujii, Y., Sakumi, K., Tominaga, Y., Nakao, K., Sekiguchi, M., Matsushiro, A., Yoshimura, Y., Morita, T. (1996) Targeted disruption of the Rad51 gene leads to lethality in embryonic mice, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6236–6240.

Venter, J.C., Adams, M.D., Myers, E.W. et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Vogel, J.L., Kristie, T.M. (2000) The novel coactivator C1 (HCF) coordinates multiprotein enhancer formation and mediates transcription activation by GABP, *EMBO J.* **19**, 683–690.

Wade, P.A., Pruss, D., Wolffe, A.P. (1997) Histone acetylation: chromatin in action, *Trends Biochem. Sci.* **22**, 128–132.

Walsh, C.P., Bestor, T.H. (1999) Cytosine methylation and mammalian development, *Genes Dev.* **13**, 26–34.

Wang, Q., Latham, K.E. (1997) Requirement for protein synthesis during embryonic genome activation in mice, *Mol. Reprod. Dev.* **47**, 265–270.

Wang, Q., Latham, K.E. (2000) Translation of maternal messenger ribonucleic acids encoding transcription factors during genome activation in early mouse embryos, *Biol. Reprod.* **62**, 969–978.

Warnecke, P.M., Clark, S.J. (1999) DNA methylation profile of the mouse skeletal alpha-actin promoter during development and differentiation, *Mol. Cell. Biol.* **19**, 164–172.

Warnecke, P.M., Mann, J.R., Frommer, M., Clark, S.J. (1998) Bisulfite sequencing in preimplantation embryos: DNA methylation profile of the upstream region of the mouse imprinted H19 gene, *Genomics* **51**, 182–190.

Wassarman, P.M., Kinloch, R.A. (1992) Gene expression during oogenesis in mice, *Mutat. Res.* **296**, 3–15.

Watson, A.J., Wiemer, K.E., Arcellana-Panlilio, M., Schultz, G.A. (1992) U2 small nuclear RNA localization and expression during bovine preimplantation development, *Mol. Reprod. Dev.* **31**, 231–240.

Wiekowski, M., Miranda, M., DePamphilis, M.L. (1991) Regulation of gene expression in preimplantation mouse embryos: effects of the zygotic clock and the first mitosis on promoter and enhancer activities, *Dev. Biol.* **147**, 403–414.

Wiekowski, M., Miranda, M., DePamphilis, M.L. (1993) Requirements for promoter activity in mouse oocytes and embryos distinguish paternal pronuclei from maternal and zygotic nuclei, *Dev. Biol.* **159**, 366–378.

Wiekowski, M., Miranda, M., Nothias, J.Y., DePamphilis, M.L. (1997) Changes in histone synthesis and modification at the beginning of mouse development correlate with the establishment of chromatin mediated repression of transcription, *J. Cell. Sci.* **110**, 1147–1158.

Witke, W., Sutherland, J.D., Sharpe, A., Arai, M., Kwiatkowski, D.J. (2001) Profilin I is essential for cell survival and cell division in early mouse development, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3832–3836.

Wolffe, A.P. (1996) Chromatin Structure and DNA Replication: Implications for Transcriptional Activity, in: DePamphilis, M.L. (Ed.) *DNA Replication In Eukaryotic Cells*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 271–293.

Wolffe, A.P., Jones, P.L., Wade, P.A. (1999) DNA demethylation, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5894–5896.

Woodland, H.R., Flynn, J.M., Wyllie, A.J. (1979) Utilization of stored mRNA in Xenopus embryos and its replacement by newly synthesized transcripts: histone H1 synthesis using interspecies hybrids, *Cell* **18**, 165–171.

Worrad, D.M., Schultz, R.M. (1997) Regulation of gene expression in the preimplantation mouse embryo: temporal and spatial patterns of expression of the transcription factor Sp1, *Mol. Reprod. Dev.* **46**, 268–277.

Worrad, D.M., Ram, P.T., Schultz, R.M. (1994) Regulation of gene expression in the mouse oocyte and early preimplantation embryo: developmental changes in Sp1 and TATA box-binding protein, TBP, *Development* **120**, 2347–2357.

Worrad, D.M., Turner, B.M., Schultz, R.M. (1995) Temporally restricted spatial localization of acetylated isoforms of histone H4 and RNA polymerase II in the 2-cell mouse embryo, *Development* **121**, 2949–2959.

Wu, X., Viveiros, M.M., Eppig, J.J., Bai, Y., Fitzpatrick, S.L., Matzuk, M.M. (2003) Zygote arrest 1 (Zar1) is a novel maternal-effect gene critical for the oocyte-to-embryo transition, *Nat. Genet.* **33**, 187–191.

Xiao, H., Friesen, J.D., Lis, J.T. (1995) Recruiting TATA-binding protein to a promoter: transcriptional activation without an upstream activator, *Mol. Cell. Biol.* **15**, 5757–5761.

Yasuda, G.K., Schubiger, G. (1992) Temporal regulation in the early embryo: is MBT too good to be true? *Trends Genet.* **8**, 124–127.

Preparedness for Bioterrorism

Nancy Khardori
University School of Medicine, Springfield, IL, USA

1 Biowarfare versus Bioterrorism 509

2 Chronology of Biodefense Actions 511

3 The Concept of "Bioterror" 512

4 Bioterrorism Preparedness – The Rationale 513

5 Bioterrorism Preparedness – The Avenues 514
5.1 National Security and Law 514
5.2 Medical and Scientific Preparedness 516
5.2.1 Assessing the Threat 516
5.2.2 Training and Recognition 517
5.2.3 Disease Reporting – A Tool for Determining Bioterrorism Preparedness 518
5.2.4 Planning and Readiness – Public Health Preparedness 519
5.2.5 Decontamination and Infection-control Procedures 520
5.2.6 Pharmaceutical Readiness 522
5.3 Political Preparedness 523
5.4 Global Preparedness Against the Threat of Bioterrorism 523

Bibliography 525
Books and Reviews 525
Primary Literature 526

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

Keywords

Biological Weapons
Microbial agents or their toxins with the potential of causing mass casualties.

Biological Weapons Convention (BWC)
Convention on the prohibition of development, production, and stockpiling of biological and toxin weapons and their destruction.

Bioterrorism
The intentional use of a pathogen or biological product to cause harm to humans, animals, plants or other living organisms to influence the conduct of government or coerce a civilian population.

Bioterrorism Preparedness
Preparedness of the public health infrastructure, the political/administrative system, and the health care system against bioterrorism threats and events.

Biowarfare
The intentional use of microbial agents or their toxins in warfare.

During the Great Depression, Hans Zinser, a bacteriologist and historian wrote that "Infectious Disease is one of the few genuine adventures left in the world." Emphasizing that this has remained so, Madeline Drexler in 2002 pointed out that "Modern adventurers like to up the *ante*, but even the most extreme sports wouldn't produce the adrenaline of a race against pandemic influenza or a cloud of anthrax at the Super Bowl. In the field of infectious disease, reality is stranger than anything a writer could dream up. The most menacing bioterrorist is Mother Nature herself." Accepting nature as the constant reservoir and source of infectious agents puts bioterrorism preparedness in the more appropriate context of preparedness against the infectious agents. By their very nature, infectious agents have the capability to spread fast and cross national borders. The course of human history has frequently been altered by epidemics that swept unchecked across continents, claiming more lives and creating more social devastation than wars. With the ease and speed of global travel now, the spread of infectious agents happens faster than ever before. An example of this is the ongoing (at the time of this writing) outbreak of Severe Acute Respiratory Syndrome (SARS). The past and the present efforts of the World Health Organization and many national agencies to strengthen global preparedness for defense against infectious disease threats is part and parcel of bioterrorism preparedness. Infectious agents have been, and in the foreseeable future will remain, potential tools of mass casualties. Historically, outbreaks (wars) of microbial species against the human species have killed far more people than war itself. Examples include: (1) death of 95% of Pre-Columbian Native American populations by diseases like smallpox, measles, plague, typhoid, and influenza; (2)

death of 25 million Europeans (a quarter of the population) caused by Bubonic Plague in the fourteenth century; and (3) 21 million deaths due to the influenza pandemic of 1918–1919. Worldwide, naturally occurring infectious diseases remain the major causes of death. In the United States, the impact of a number of very virulent biological agents and/or their toxins has been drastically reduced because of a very accessible health care system and excellent public health infrastructure. Still, a substantial number of people (approximately 170 000) die each year from infectious diseases. The travel and trade necessary for economic globalization, continued potential for transmission of infectious agents from animals to humans, and large populations living in proximity in major urban areas of the world, make disease outbreaks a major threat. The resistance of common pathogens to the available antimicrobial agents adds significantly to the threat. Advances in public health, diagnostic, and pharmacological interventions are needed to protect the population from emerging and reemerging infectious diseases, including those related to potential bioterrorism events.

1
Biowarfare versus Bioterrorism

The intentional use of microbial agents in warfare dates back to days when neither the pathogens nor the host pathogen interactions were known. It is important to recognize that biological agents have been used in warfare from antiquity to the present. The intentional use of living organisms or infected materials derived from them has occurred over centuries during war and "peace" time by armies, states, groups, and individuals. One of the first recorded uses of a biological agent in the war was in 184 B.C. The Carthaginian soldiers led by Hannibal used snakes in the battle against King Eumenes of Perganium and achieved a victory. A strong association was made between foul odors and disease, well before the development of the germ theory of disease, and many early civilizations employed crude forms of biological warfare. As early as 300 B.C., the Greeks polluted the wells and drinking water supplies of their enemies with animal corpses. The same tactics were used later by the Romans and Persians. The bodies of dead soldiers as well as animals were used to pollute wells during a battle in Italy in 1155. Pollution (poisoning) of potable water was used as an effective and calculated method of gaining advantage in warfare throughout the Classical, Medieval, and Renaissance periods. In the more modern times, General Johnson used the bodies of sheep and pigs to pollute drinking water at Vicksburg during the US Civil War.

The use of catapults and siege machines during medieval times introduced new technology to biological warfare. In 1346, the attacking Tartar force experienced an epidemic of plague during the siege of Kaffa (now Feodosia, Ukraine). They catapulted the cadavers of their deceased into the city to initiate a plague epidemic. An outbreak of plague did occur followed by the retreat of defending forces and the conquest of Kaffa. It is thought that this contributed to the second plague pandemic as the ships carrying

plague-infected refugees and rats sailed to Constantinople, Genoa, Venice, and other Mediterranean ports.

The inhabitants of Central and South America were decimated by smallpox and measles that accompanied the Spanish Conquistadores. In 1763, the introduction of a specific disease, smallpox, as a weapon in the North American Indian Wars became a significant milestone in the history of biological warfare. This adaptation of Trojan horse use was employed by Captain Ecuyes, on June 24, 1763, by handing blankets and a handkerchief from the smallpox hospital to the Native Americans. Sir Jeffery Amherst, Commander of British forces in North America, had suggested the deliberate use of smallpox to "reduce" Native American tribes hostile to the British.

The development of the science of bacteriology in the nineteenth century expanded the scope of potential biological warfare agents. *Bacillus anthracis* was the first specific agent attributed to human disease when Robert Koch fulfilled his own "postulates" concerning this organism in 1877. During World War I, Germany was accused of using cholera in Italy and plague in St. Petersburg in 1915. There is evidence that Germany used glanders and anthrax organisms to infect horses, mules, and cattle (1916–1918). These covert operations were conducted on neutral trading partners of the allies to infect livestock and to contaminate animal feed to be exported to allied forces.

The modern era of biological weapons development began immediately before and during World War II. The Japanese released fleas infected with plague in Chinese cities in the 1930s and 1940s. Between 1940 and 1942, Japanese unit 731 dropped bombs containing up to 15 million plague-infected fleas on two Chinese cities – Quxian and Ning-hsien, resulting in thousands of deaths. Water supplies and food items were contaminated with *B. anthracis*, Cholera, Shigella, Salmonella, and *Yersinia pestis*.

The origin of the biological weapons program of the former Soviet Union dates back to statements by Lenin. Although experimental work was started in the late 1920s, the modern era was ushered in only with the postwar military building programs. The Allied Biological Weapons program had shifted from the British research on anthrax (and the development of the World War II anthrax cattle cake retaliation weapon) to a large US-based research, development, and production capability. The US military had accepted seven types of classified agents and could produce 650 tons of an agent per month at plants such as the one at Pine Bluff in Arkansas. This offensive program was unilaterally abandoned in 1969, giving impetus to the creation of the Biological and Toxin Weapons Convention. The Soviet Union signed the Convention at its inception in 1972. Unfortunately, the number of countries engaged in biological weapons experimentation grew from 4 in the 1960s to 11 in the 1990s. It is estimated that at least 10 nations and possibly 17 possess biological warfare agents. Of the seven countries listed by the United States Department of State as sponsoring international terrorism, at least five are suspected to have biological warfare programs. Nations and dissident groups have the access to the skills needed to selectively cultivate some of the most dangerous pathogens and to deploy them as agents of biological terrorism and war.

Iraq's biological weapons program dates back to at least 1974 and has been carried out secretly after the Biological and Toxin Weapons Convention had been signed. In

1995, Iraq confirmed that it had produced and deployed bombs, rockets, and aircraft spray tanks containing *B. anthracis* and botulinum toxin. In 1973, the Soviet Politburo formed the organization known most recently as *Biopreparat*, designed to carry out offensive biological weapons programs concealed behind civil biotechnology research. In January 1991, the first ever visit to Biopreparat facilities was undertaken by a Joint United Kingdom/United States technical team. By the mid-1990s, substantial changes took place within Biopreparat, and a concerted effort is underway to help the Russians civilianize these former biological weapons research and development establishments. The current capability of the old Russian Ministry of Defense sites remains largely unknown. The status of one of Russia's largest and most sophisticated former bioweapons facilities called *Vector* in Koltsovo, Novosibirsk, is of concern. The facility housed the smallpox virus as well as work on Ebola, Marburg, and the hemorrhagic fever viruses (e.g. Machupo and Crimean-Congo). A visit in 1997 found a half-empty facility protected by a handful of guards. No one is clear where the scientists have gone. The confidence that this is the only storage site for smallpox outside the Centers for Diseases Control and Prevention is lacking.

2
Chronology of Biodefense Actions

1910–1920s–The first use of chemical and biological weapons in combat led to efforts to ban their use.

1925–The use of biological and chemical weapons in war was prohibited by The Geneva Protocol. The United States signed but failed to ratify the treaty. The treaty contained no provision for verifications and inspection.

1950s–1970s–The Soviet Union and United States built arsenals of biological and chemical weapons. International pressures mounted to draw up new treaties to curb such weapons.

November 25, 1969–President Richard M. Nixon unilaterally renounced the use of biological weapons in war by the United States and restricted research to immunization and safety efforts. Three months later, he extended the ban to include toxins. The United States Army Medical Research Institute for Infectious Diseases (USAMRIID) at Fort Detrick, Maryland, was established to conduct unclassified research on safety against potential agents of bioterrorism.

1972–Convention on the prohibition of the development, production, and stockpiling of bacteriological (biological) and toxin weapons and their destruction opened for signature at Washington, London, and Moscow on April 10, 1972.

1975–The United States Congress ratified the Biological and Toxin Weapons Convention (BWC) as well as the 1925 Geneva Protocol on January 22, 1975. The Biological and Toxin weapons Convention entered into force on March 26, 1975.

There are now 143 state parties to the convention and an additional 18 signatories. Article VI of the Convention has proved to be an inadequate mechanism for actions against noncompliance.

1980s–Arms control initiatives failed to curb biological and chemical weapons proliferation.

1984–President Reagan's administration presented a draft treaty to ban the production and storage of chemical

weapons to the Conference on Disarmament in Geneva.

1990s – Concerns over exposure to chemical and biological weapons during the Persian Gulf War increased support for international treaties.

May 13, 1991 – Shortly after the Allied victory against Iraq, President George Bush announced that the United States will renounce the use of chemical weapons for any reason, and an international treaty banning them took effect.

April 1992 – Russian President Boris N. Yeltsin declared that Russia's biological weapons program is being discontinued.

January 1993 – President George Bush signed the Chemical Weapons treaty at the convention, banning the production and use of chemical weapons.

January 7, 1997 – The Presidential Advisory Committee on Gulf War Veterans' illnesses found no conclusive evidence linking Gulf War Syndrome to chemical or biological weapons.

April 15, 1997 – New regulations aimed at limiting access to chemicals and pathogens that could be made into weapons went into effect under the 1996 Antiterrorism and Effective Death Penalty Act.

April 29, 1997 – The Chemical Weapons Convention (CWC) went into effect. It has more than 160 signatories and 65 ratifications.

July 25, 2001 – The United States rejected a protocol to strengthen the Biological Weapons Convention as well as the whole approach to it. As in the Chemical Weapons Conventions (CWC), a strong bioweapons protocol could add to the deterrence of bioweapons, which are a much greater threat than chemical weapons.

November 19, 2001 – Fifth BWC Review Conference was held.

3
The Concept of "Bioterror"

As the ease of cultivating microorganisms increased and technology became cheap, dissident groups and well-financed organizations used these agents to terrorize civilians in order to accomplish political goals. Examples of these attempts include the following:

- Weather Underground (1970), a US revolutionary group, intended to obtain agents from Fort Detrick by blackmail and to temporarily incapacitate US cities to demonstrate the "impotence of the federal government." Report originated with a US Customs informant.
- R.I.S.E. (1972). College students influenced by ecoterrorist ideology and 1960s drug culture planned on using agents of typhoid fever, diphtheria, dysentery, and meningitis to target the entire world population initially and later narrowed the plan to five cities near Chicago. The attack was aborted when cultures were discarded.
- Bulgarian defector Georgi Markov was assassinated in Lauda (1978) using ricin-filled pellets and a spring-loaded device disguised in an umbrella. Similar device used against a second defector in the same area was unsuccessful.
- Sverdlovsk, Russia (1979). Accidental release of anthrax from Soviet bioweapons facility caused an epidemic of inhalational anthrax with at least 77 cases and 60 deaths.
- Red Army Faction (1980). Members of a Marxist revolutionary ideology group allegedly cultivated botulinum toxin in

a Paris safe house and planned attacks against at least nine German officials and civilian leaders. This probably was an erroneous report, later repudiated by the German government.

- Rajneeshee Cult (1984). Indian religious cult headed by Rajneeshee plotted to contaminate restaurant salad bars with *Salmonella typhimurium* in Dalles, Oregon. The motive was to incapacitate voters, win local elections, and seize political control of the county. The incident resulted in a large community outbreak of salmonellosis, involving 751 patients and at least 45 hospitalizations. The plot was revealed when the cult collapsed and members turned informants.
- Minnesota Patriots Council (1991). Right-wing "Patriot" movement obtained ricin extracted from castor beans by mail order. They planned to deliver ricin through skin with dimethyl sulfoxide (DMSO) and aloe vera or as dry aerosol against IRS officials, US Deputy Marshals, and local law enforcement officials. The group was penetrated by Federal Bureau of Investigation (FBI) informants.
- Aum Shinrikyo (1995). New Age Doomsday cult, seeking to establish a theocratic state in Japan, attempted at least 10 times to use anthrax, botulinum toxin, Q fever agent, and Ebola virus in aerosol form. All attempts with biological weapons failed. Multiple chemical weapon attacks with Sarin, Vx, and hydrogen cyanide in Matsumato, Tokyo, and assassination campaigns were conducted. Nerve gas Sarin killed 12 and injured 5500 in Tokyo subway.
- Texas (1997). Intentional contamination of muffins and donuts with laboratory cultures of *Shigella dysenteriae* caused gastroenteritis in 45 laboratory workers and 4 were hospitalized.
- Unknown individual group (2001). Intentional dissemination of anthrax spores through the US Postal System led to the death of five people, infection of 22 others, and contamination of several government buildings. Investigation into the attacks so far has not led to any conclusions.

Although events like these do not reach the level of national or international security, they can have significant public health consequences and, therefore, require preparedness at the local level. Active surveillance and rapid response at the local level will fulfill the definition for Bioterrorism preparedness – "think locally, act globally."

The use of microbial agents by smaller groups or individuals who may have very limited targets is highly likely but the public health consequences are low.

- Larry Wayne Harris (1998) allegedly threatened to release "military grade anthrax" in Las Vegas, Nevada. He obtained plague and anthrax (vaccine strains), and repeatedly isolated several other bacteria. He then made vague threats against US federal officials on behalf of right-wing "patriot" groups and was arrested when he talked openly about biological weapons terrorism.

4
Bioterrorism Preparedness – The Rationale

Bioterrorism has been defined as the intentional use of a pathogen or biological product to cause harm to humans, animals, plants, or other living organisms to influence the conduct of government or to intimidate or coerce a civilian population. This definition put bioterrorism

under National Security and has prompted the recent upgrading of the Homeland Security Department in the United States. Biological agents are easy to develop as weapons, inexpensive, and difficult to detect. Biological weapons are very attractive to the terrorist because of several characteristics. Aerosols of biological agents are invisible, silent, odorless, tasteless, and are relatively easily dispersed. They are 600 – 2000 times cheaper than other weapons of mass destruction. It is estimated that the cost would be about 0.05% of the cost of conventional weapons to produce similar numbers of mass casualties per square kilometer. The production is relatively easy, using the common technology available for the production of antibiotics, vaccines, foods, and beverages. The delivery systems such as spray devices from an airplane, boat, or car are commonly available. The natural lead time provided by the organism's incubation period (3 to 7 days for most potential organisms) would allow for the terrorists' escape before any investigation starts. In addition, the use of an endemic infectious agent may cause confusion because of the inability to differentiate a biological warfare attack from a natural epidemic. For some agents, potential exists for secondary or tertiary cases by person-to-person transmission or by natural vectors. Depending on the agent, a lethal or incapacitating outcome can occur. In the military context, incapacitating agents may be more effective because the unit will not be able to perform their mission and casualties will consume scarce medical and evacuation assets.

The consequences of biological weapons use are many. They can rapidly produce mass effect that overwhelms services and the health care system of the communities. Most of the civilian population is susceptible to infections caused by these agents. They are associated with high morbidity and mortality rates. The resulting illness is usually difficult to diagnose and treat early, particularly in areas where the disease is rarely seen. One kilogram of anthrax powder has the capability to kill up to 100 000 people, depending on the mechanism of delivery. The economic impact of a biological attack has been estimated to be from 478 million/100 000 persons exposed (brucellosis scenario) to 26.2 billion/100 000 persons exposed (anthrax scenario).

5 Bioterrorism Preparedness – The Avenues

5.1 National Security and Law

Two simulated biological attacks, Dark Winter (smallpox) and TOPOFF (plague), in the United States demonstrated serious weaknesses in the public health system that could prevent an effective response to bioterrorism or severe outbreaks of naturally occurring infectious diseases. The Hanuman Redux 2001 exercise run in Louisville, Kentucky, under the auspices of the CDC, a few weeks prior to September 11, 2001, used tularemia as the biological agent. The exercise demonstrated that local, state and federal health, law enforcement, and other agencies can work together effectively. It also showed that regions with adequate bioterrorism response plans can respond to even a sizeable attack within the time frame needed to prevent mass casualties. The intentional dispersal of anthrax through the US Postal Service that followed the terrorist attacks of October 2001 brought these issues into a clear focus. The US government began

a process to strengthen the public health infrastructure.

The need for law reform was recognized as law has long been considered an important tool for the safety of the public's health. The power to act to preserve the public's health is constitutionally reserved primarily to the states as an exercise of their police powers. Some states like Colorado and Rhode Island had developed legislation or administrative public health plans for a bioterrorism event prior to September 11, 2001. The Model Act was designed to update and modernize the state public health statutes and to avoid problems of inconsistency, inadequacy, and obsolescence. The Model State Emergency Health Powers Act (MSEHPA or Model Act) was drafted by the Center for Law and the Public's Health at Georgetown and Johns Hopkins Universities at the request of the Centers for Disease Control and Prevention (CDC) and in collaboration with members of national organizations representing governors, legislators, attorneys, general and health commissioners. This act provides state actors with the powers to detect and contain bioterrorism or a naturally occurring disease outbreak. The Model Act is structured to facilitate five basic public health functions: (1) Preparedness, comprehensive planning for a public health emergency; (2) Surveillance, measures to detect and track public health emergencies; (3) Management of Property, ensuring adequate availability of vaccines, pharmaceuticals, and hospitals as well as providing power to abate hazards to the public's health; (4) Protection of Persons, powers to compel vaccination, testing, treatment, isolation, and quarantine when clearly necessary; and (5) Communication, providing clear and authoritative information to the public. The act also contains a modernized, extensive set of principles and requirements to safeguard personal rights. On the basis of MSEHPA, legislative bills have been introduced in 34 states and the District of Columbia. As of July 26, 2002, 16 states and the District of Columbia had enacted a version of the act and the states enacting or expected shortly to enact legislation, influenced by the Model Act, were Arizona, Florida, Georgia, Hawaii, Maine, Maryland, Minnesota, Missouri, New Hampshire, New Mexico, Oklahoma, South Carolina, South Dakota, Tennessee, Utah, and Virginia. A critique of the Model State Emergency Health Powers Act was published by Anna G, J., which raised concerns about the impact of the Model Act on civil liberties.

At the federal level, Public Health Security and Bioterrorism Preparedness and Response Act of 2002, HR 3448, was passed by the US Congress on May 23, 2002, and signed into law (Public Law 107–188) June 12, 2002. The bill is intended to improve the health system's ability to respond to bioterrorism, protect the nation's food supply and drinking water from bioterrorist attacks, speed up the development and production of new drug treatments and vaccines, address shortages of specific types of health professions, improve coordination of federal antibioterrorism activities, increase investment in federal, state, and local preparedness, and expand controls over the most dangerous biological agents and toxins. The American Society for Microbiology (ASM) has testified before congress on issues surrounding biosecurity and has worked closely with congress in the drafting of Title II of this law to balance public health concern over safety and security with the need to protect legitimate scientific research and diagnostic testing. Important new provisions for

the possession, use, and transfer of select agents (42 biological agents and toxins listed in Appendix A of 42 CFR part 72) are included in Title II of HR 3448, leading to enhanced Controls on Dangerous Biological Agents and Toxins. On July 12, 2002, the CDC announced preliminary guidance for notification of possession of select agents as mandated in Section 202(a) of Public Law 107–188, the Public Health Security and Bioterrorism Preparedness and Response Act of 2002 (Appendix B). The notice stated that each facility should designate a responsible facility official (RFO) to complete the notification of possession form by September 10, 2002. The RFO will need to inventory the facility and consult with others (e.g. Principal Investigators) to obtain the required information.

5.2 Medical and Scientific Preparedness

5.2.1 Assessing the Threat

A US congressional initiative to upgrade national public health capabilities for response to acts of bioterrorism was begun in 1999. The Centers for Disease Control and Prevention (CDC) was designated by the Department of Health and Human Services as the lead agency for overall public health planning. A Bioterrorism Preparedness and Response Office was formed to target several areas for initial preparedness activities including planning, improved surveillance and epidemiological capabilities, rapid diagnostic testing, enhanced communications, and medical therapeutics stockpiling. The biological agents toward which the efforts should be targeted needed to be formally identified and placed in a priority order. The strategic usefulness of multiple agents on the battlefield has formally been assessed by the military. The Working Group on Civilian Biodefense has identified several biological agents with the potential of high impact against civilian populations.

In June 1999, national public health experts, academic infectious disease experts, Department of Health and Human Services representatives, civilian and military intelligence experts, and law enforcement officials met to review the impact of various agents on civilian population. The general criteria used to assess the threat were: (1) public health impact based on illness and death, (2) delivery potential to large populations based on mass production, stability, virulence, and potential for person-to-person transmission, (3) public perception as related to public fear and potential civil disruption, and (4) special public health preparedness needs based on enhanced surveillance, diagnostic needs, and stockpile requirements. A risk-matrix analysis process was used to evaluate potential biological threat agents. On the basis of the overall ratings received in each of the four areas, the agents received final category assignments of A, B, or C (Table 1).

Category A Category A includes the highest priority agents that pose a risk to national security because of the following features:

1. They can be easily disseminated or transmitted person to person, causing secondary and tertiary cases.
2. They cause high mortality with potential for major public health impact, including the impact on health care facilities.
3. They may cause public panic and social disruption.
4. They require special action for public health preparedness.

Tab. 1 Potential agents of bioterrorism.

Category A	***Category B***	***Category C***
Variola major (smallpox) *Bacillus anthracis* (anthrax) *Yersinia pestis* (plague) *Clostridium botulinum* toxin (botulism) *Francisella tularensis* (tularemia) Filoviruses, Ebola hemorrhagic fever Marburg hemorrhagic fever Arenaviruses, Lassa (Lassa fever) Junin (Argentine hemorrhagic fever) and related viruses	*Coxiella burnetii* (Q fever) Brucella species (brucellosis) *Burkholderia mallei* (glanders) *Burkholderia pseudomallei* (Melioidosis) Alphaviruses, Venezuelan encephalomyelitis Eastern and Western equine\Encephalomyelitis Ricin toxin from *Ricinus communis* (castor beans) Epsilon toxin of *Clostridium perfringens* Staphylococcal enterotoxin B T2 – Mycotoxins Food or waterborne pathogens include but not limited to Salmonella species *Shigella dysenteriae* *Escherichia coli* 0157.H7 *Vibrio cholerae*, and *Cryptosporidium parvum Rickettsia prowazekii* (*Typhus* fever) *Chlamydia psittaci* (Psittacosis)	Nipah virus Hantaviruses Tick-borne hemorrhagic fever viruses Tick-borne encephalitis viruses Yellow fever Multidrug-resistant tuberculosis

Category B This category contains the second highest priority agents because they

1. are moderately easy to disseminate;
2. cause moderate morbidity and low mortality;
3. require specific enhancement of CDC's diagnostic capacity and enhanced disease surveillance.

Category C The agents in this group with the third highest priority include emerging pathogens that could be engineered for mass dissemination. The characteristics that render them amenable to bioterrorism are:

1. Availability
2. Ease of production and dissemination
3. Potential for high morbidity and major health impact.

5.2.2 Training and Recognition

To diagnose the illnesses caused by potential bioterrorism agents, the health care professionals must be aware of the symptoms and epidemiology patterns that may indicate a covert or unannounced attack.

A terrorist attack may occur in two scenarios – overt and covert. In the past, emergency responses were prepared on the basis of overt attacks like bombings and chemical agents that cause immediate and obvious effects. However, attacks with biological agents are more likely to be covert. They pose different challenges and require emergency planning, with the involvement of the public health

infrastructure. The attack by a biological agent will not have an immediate impact because of the delay between exposure and onset of illness (i.e. the incubation period). Therefore, the first victims of a bioterrorism action will need to be identified by physicians or other primary health care providers. On the basis of the first wave of victims, public health officials will need to determine that an attack has occurred, identify the organism, and prevent more casualties through prevention strategies (e.g. mass vaccination, prophylactic treatment) and infection-control procedures. The clues to a potential bioterrorist attack include an outbreak of a rare or new disease, an outbreak of diseases in a nonendemic area, a seasonal disease during an off-season time, a known pathogen with unusual resistance or unusual epidemiologic features, an unusual clinical presentation or age distribution, a genetically identical pathogen emerging rapidly in different geographical areas, reports of sick or dead animals or plants, and intelligence information.

It is now recognized that emergency departments (EDs) are critical partners in the early recognition of emerging public health threats. Early recognition depends on epidemiologic warning networks from the local and national public health agencies and the clinical knowledge and increased index of suspicion by the health care providers working in emergency departments and urgent care centers. The ED Surveillance System, known as *EMERGEncyIDNET*, designed to detect emerging infectious diseases, started as a collaboration between 11 academic EDs and the CDC. Food Net, the Foodborne Disease Active Surveillance Network is an analogous program also coordinated by the CDC and is designed to monitor and track foodborne diseases.

Although very useful, such large databases may not be able to provide timely response without vigilance of local health care providers. Surveillance systems based in ambulatory care settings provide worthwhile additional information in a more timely fashion. The Department of Defense Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) is one of the best known such systems. Minnesota operates another such system. A system of automated ambulatory care encounter records for the detection of acute illness clusters has been developed between the Centers for Disease Control and Prevention and the Massachusetts Department of Public Health, a large group practice, a health plan, and an academic department. The system is designed to produce next-day information about illness clusters, based on ambulatory care visits and telephone calls. Such measures to strengthen the surveillance systems and improved training of health care providers will be critical to effective planning and readiness. A task force of health care professionals, central to detecting and responding to a bioterrorism incident, was convened by the American College of Emergency Physicians. The goal is to examine educational strategies, curricula, teaching methods, and certification processes with the outcome of better-prepared "frontline" health care professionals.

5.2.3 Disease Reporting – A Tool for Determining Bioterrorism Preparedness

The timely and accurate reporting of the diseases caused by biological agents would likely be the triggering event for investigating and responding to a bioterrorism

event. The quality and distribution of laws mandating the reporting of disease caused by certain biological agents is one measurement of the public health system's level of bioterrorism preparedness. The CDC's Bioterrorism Preparedness and Response Program commissioned a study of state and local laws requiring the reporting of diseases caused by specific biological agents. These agents are regarded as "critical biological agents" because of their potential to harm the public health, especially if used in a terrorist act. As of March 31, 2001, this commission reported a significant variability in disease-reporting requirements pertaining to critical biological agents in the 50 states, the District of Columbia, Chicago, New York, and Los Angeles County. Some jurisdictions may have revised these requirements since September 2001. The commission's findings suggest the draft Model State Emergency Health Powers Act as a useful template for states considering revisions to reporting laws. The Model State Act addresses which diseases should be reported, who should be legally obligated to report, the manner and time frame in which a disease should be reported, and the enforcement of disease-reporting laws. Such laws may also serve as educational tools and provide incentives for health care providers to obtain the necessary training and skills for diagnosing and managing potential agents of bioterrorism.

5.2.4 Planning and Readiness – Public Health Preparedness

The National Association of Counties recently conducted a survey of the Directors of County Public Health Departments, which represented 300 counties in 36 states. A significant number of responding counties reported less than optimal levels of preparedness for biological threats and for policies and procedures to enforce quarantine. Among the reasons cited for lack of preparedness were insufficient funding, work force, and communication networks. In most cities in the United States, large health care institutions have disaster plans and various types of task forces with "experts" in different areas in place. However, they need to be updated and modified to include new information on potential agents of bioterrorism. The CDC was designated by the Department of Health and Human Services to prepare the US Public Health System to respond to bioterrorism events. A strategic plan was developed by the CDC in collaboration with the National Center for Infectious Diseases, the National Center for Environmental Health, the Public Health Practice Program Office, the Epidemiology Program Office, the National Institute for Occupational Safety and Health, the Office of Health and Safety, the National Immunization Program, and the National Center for Injury Prevention and Control. In order to enhance the preparedness at local and state levels, the CDC funded cooperative agreements with states and several large cities. During the first three years of the program (1999–2001), five areas were emphasized: (1) Preparedness, planning, and readiness assessment; (2) surveillance and epidemiology capacity; (3) biological laboratory capacity; (4) chemical laboratory capacity and; (5) health alert network and training.

Recognizing the pivotal role of rapid laboratory identification of bioterrorism agents and the limited laboratory experience with Category A agents in the United States, the Laboratory Response Network (LRN) for Bioterrorism was established.

The CDC and the Association of Public Health Laboratories developed LRN for bioterrorism in collaboration with the Federal Bureau of Investigation (FBI), United States Army Medical Research Institute of Infectious Diseases, the Naval Medical Research Center, and Lawrence Livermore National Laboratory. The LRN became operational in August 1999, and is a multilevel system that links state and some local public health laboratories with military, veterinary, agricultural, water, and food-testing laboratories. The network of laboratories is designated Levels A through D with progressively stringent levels of safety, containment, and technical proficiency needed to perform the essential rule-out, rule-in, and referral functions required for identification of agents. Level A laboratories are the hospital and other clinical laboratories that participate in the LRN by ruling out or referring the critical agents to the nearest Level B or C laboratory. Levels B, C, and D laboratories have access to the biodetection assays and specialized agents that are used in protocols for the confirmation of critical agents. Level B laboratories are state and local public health laboratories with Biosafety Level 2, where Biosafety Level 3 practices are observed. Level C laboratories are Public Health laboratories with Biosafety Level 3 facilities or certified animal facilities for performing the mouse toxicity assay for botulinum toxin. They perform all Level B tests and additional tests requiring Biosafety Level 3 containment, including the handling of powders suspected of containing anthrax spores. Level D laboratories are the federal laboratories at the CDC and the United States Army Medical Research Institute of Infectious diseases. They have Biosafety Level 4 capacity to handle agents like smallpox and Ebola.

5.2.5 Decontamination and Infection-control Procedures

The information in this section refers to procedures used by health care institutions if and when patients with communicable diseases including those caused by the agents of bioterrorism are managed.

The Association of Professionals in Infection Control and Epidemiology (APIC) in cooperation with the CDC has prepared a Bioterrorism Readiness plan to serve as a reference document and a template to facilitate preparation of bioterrorism readiness plans for individual institutions.

Decontamination methods have an important role in the control of infectious diseases. However, the procedures have to be safe for people and materials they are used on. Examples of mechanical decontamination are water filtration, air filtration, and washing with water. Skin surfaces can be easily decontaminated by careful washing with soap and water. Chemical decontamination (disinfectants) used in the form of liquid, gas, or aerosol render biological agents harmless. Some of these are harmful to humans, animals, and the environment. Grossly contaminated skin surfaces should be washed with 0.5% sodium hypochlorite solution with a contact time of 10–15 min. Physical agents (head, radiation) are employed for decontamination of objects. The potential agents of bioterrorism that can be transmitted person to person require patient-isolation precautions. In addition to standard precautions employed in the care of all patients, airborne precautions, droplet precautions, and contact precautions are recommended for patients with smallpox, pneumonic plague, and viral hemorrhagic fevers respectively.

The personal protective equipment currently in use in the health care facilities in the United States is adequate for dealing with infectious agents with various modes of transmission. The USAMRIID maintains an aeromedical isolation team to minimize the risk of transmission from the troops to the aircrew, caregivers, and civilians.

Patient-isolation precautions – Appendix B: USAMRIID's medical management of biological casualties handbook

Standard Precautions

- Wash hands after patient contact.
- Wear gloves when touching blood, body fluids, secretions, excretions, and contaminated items.
- Wear a mask and eye protection, or a face shield during procedures likely to generate splashes or sprays of blood, body fluids, secretions, or excretions.
- Handle used patient-care equipment and linen in a manner that prevents the transfer of microorganisms to people or equipment.

Use care when handling sharps and use a mouthpiece or other ventilation device as an alternative to mouth-to-mouth resuscitation when practical.

Standard precautions are employed in the care of ALL patients.

Airborne precautions

Standard Precautions Plus

- Place the patients in a private room that has monitored negative air pressure, a minimum of six air changes/hour, and appropriate filtration of air before it is discharged from the room.
- Wear respiratory protection when entering the room.
- Limit movement and transport of the patients. Place a mask on the patients if they need to be moved.

Conventional diseases requiring airborne precautions: measles, varicella, pulmonary tuberculosis.

Biothreat diseases requiring airborne precautions: smallpox

Droplet precautions

Standard Precautions Plus

- Place the patients in a private room or cohort them with someone with the same infections. If not feasible, maintain at least 3 ft between patients.
- Wear a mask when working within 3 ft of the patients.
- Limit movement and transport of the patients. Place a mask on the patients if they need to be moved.

Conventional diseases requiring droplet precautions: Invasive *Haemophilus influenzae* and meningococcal disease, drug-resistant Pneumococcal disease, diphtheria, pertussis, mycoplasma, GABHS, influenza, mumps, rubella, parvovirus.

Biothreat diseases requiring droplet precautions: pneumonic plague.

Contact precautions

Standard Precautions Plus

- Place the patients in a private room or cohort them with someone with the same infection if possible.
- Wear gloves when entering the room. Change gloves after contact with infective material.
- Wear a gown when entering the room if contact with patient is anticipated or if the patient has diarrhea, a colostomy, or wound drainage not covered by a dressing.

- Limit the movement or transport of the patient from the room.
- Ensure that patient-care items, bedside equipment, and frequently touched surfaces receive daily cleaning.
- Dedicate use of noncritical patient-care equipment (such as stethoscopes) to a single patient, or cohort of patients with the same pathogen. If not feasible, adequate disinfection between patients is necessary.

Conventional diseases requiring contact precautions: MRSA, VRE, *Clostridium difficile*, RSV, parainfluenza, enteroviruses, enteric infections in the incontinent host, skin infections (SSSS, HSV, impetigo, lice, scabies), hemorrhagic conjunctivitis.

Biothreat diseases requiring contact precautions: viral hemorrhagic fevers.

5.2.6 Pharmaceutical Readiness

The United States Food and Drug Administration (FDA) is participating in an interagency group for response in a civilian emergency. The group represents the Department of Defense, the Veteran's Administration, CDC, NIH, and the Office of Emergency Preparedness. The regulation of medical products by the FDA is based on science, law, and public health considerations. Research conducted at the FDA's Center for Biologics Evaluation and Research contributes to the defense against bioterrorism agents. These efforts include design of new vaccines, pathogenesis and mechanism of replication of bioterrorism agents, new methods and standards to expedite the review of new vaccines and immunoglobulins. The FDA also proposes standards for the use of animal efficacy data in approving new products and ensures appropriate product surveillance and compliance.

Stockpiling of supplies, antimicrobial agents, and vaccines is considered the key mechanism for rapid response to a bioterrorism threat and/or event. Stockpiles have been created by the Department of Health and Human Services that can reach any community within 12 h. Special teams will accompany the stockpiles to aid in unpacking and preparing for distribution. The use of stockpiled antibiotics for mass protection has the disadvantages of side effects, intolerances, and lack of efficacy against genetically engineered bioterrorism agents. In addition to the use of antimicrobial agents for postexposure prophylaxis, administration of passive antibody has been proposed as a biological defense initiative. Specific antibodies are active against the major agents of bioterrorism, have low toxicity, and provide immediate protection. Various models and estimates of the economic impact of bioterrorism attacks have been published. Rapid implementation of a postattack prophylaxis program is the single most important means of reducing the huge economic impact. The model proposed by Kaufamnn et al. provides economic justification for preparedness measures. Unlike the military, there are not enough emergency providers who have received sufficient immunizations to care for patients with infections due to critical biological agents. Efforts to immunize health care providers against smallpox and create a "smallpox response team" in most areas of the country have been started in the United States. After immunizing about 40 000 personnel, the process is currently suspended. However, smallpox vaccination can be protective even when given 3–4 days after infection. Thus, the personnel needed to manage patients will be able to be protected from the time smallpox is diagnosed. The currently available anthrax

vaccine is being given to all deployed military personnel. More efficacious anthrax vaccines with fewer injections are under study.

5.3 Political Preparedness

The single most important issue in any preparedness plan is to know who will be "in charge." This question must be adequately answered before an emergency arises. All local hospitals and governments should have an action plan, with responsible personnel and their alternates in place prior to the incident.

Following the Oklahoma City bombing, President Clinton signed Presidential Decision Directive 39 (PDD-39), which places the management of such attacks under the Federal Bureau of Investigation (FBI) and the Federal Emergency Management Agency (FEMA) for domestic incidents. The preparations include site surveys, assessment of local resources to treat victims, and status of local stocks of antidotes. The US Congress followed this with Public Law 104–201, the Defense against Weapons of Mass Destruction Act.

The Office of Homeland Security, created in the aftermath of the September 11, 2001, attacks, was upgraded to a full Department of the Federal Government in June 2002 by President George W. Bush. The Department is likely to have a budget of $32 billion a year, including several hundred million for research. The biodefense research will yield results applicable to natural infectious diseases also. However, the advantages of disseminating the results to researchers can be offset by the information also reaching the terrorists.

The Department of Defense has the greatest capability in Biodefense for the military forces. In the civilian context, the responsibility for dealing with the effects of biological agents falls into multiple venues, including many federal, state, and local governments, and ultimately the civilian medical community. The Department of Health and Human Services has the lead role and the bulk of the resources.

5.4 Global Preparedness Against the Threat of Bioterrorism

As in the case of outbreaks of natural infectious diseases, diseases caused by potential agents of bioterrorism have quick and devastating global implications. Most countries do not have the resources to prepare against bioterrorism attacks or the spread of bioterrorism agents. The World Health Organization (WHO) with over 50 years of experience has the task of a rapid, multifaceted response needed to contain outbreaks quickly and prevent their international spread. The staff, consultants, and expert advisors of WHO have privileged access to all countries. WHO's permanently positioned geographical resources include the Geneva headquarters, six regional offices, and 141 country offices located in close proximity to ministries of health. The country offices are concentrated in the areas where epidemics are most frequent and new diseases are most likely to emerge. All offices are linked electronically to WHO and its global network of institutional resources and collaborators. The network of formally designated WHO collaborating centers consists of 250 laboratories and institutions. In addition, many laboratories and public health institutions collaborate with WHO on a regular basis. WHO has a very close

working relationship with CDC, including direct secondment of staff. It also uses technical support from laboratories included in the US Department of Defense Global Emerging Infections Surveillance and Response System (DOD – GEIS) and similar laboratories in other WHO member states. Other US agencies that provide technical and financial assistance for WHO sponsored programs are USAMRIID, NASA-Goddard Space Flight Center, the National Centers for Environmental prediction, the National Institutes of Health, and the US State Department's Bureau of Population, Refugees, and Migration.

WHO is politically neutral and greatly needed in the developing world. Ministries of Health in these countries rely upon WHO as the single most important source of authoritative advice and technical assistance pertaining to the control of infectious diseases. The epidemiological techniques needed to investigate natural and deliberate outbreaks are the same. WHO coordinates a large number of electronic "detective" systems and databases to keep watch over known risks as well as unexpected or unusual disease events. FluNet, established more than 50 years ago, is the oldest of these and has served as the prototype for the design and implementation of many other systems. The recent establishment of innovative mechanisms for detection and response to unknown diseases or unusual or unexpected disease events has added to the global preparedness against bioterrorism. WHO scans the world through a semi-automatic electronic system developed by Health Canada that continuously and systemically crawls websites, news wires, public health e-mail services, electronic discussion groups (e.g. US-based ProMED), and local on-line newspapers for definitive and rumor-based information. The World Health Assembly, the supreme governing body of WHO, adopted a consensus resolution on global health security in May 2001. This allows the WHO to investigate and verify rumored outbreaks even prior to receipt of an official notification from the government of the country concerned. To combine heightened vigilance with a rapid response, WHO has formed the Global Outbreak Alert and Response Network and issued guidelines for the behavior of foreign nationals during and after field operations in the host country.

Strengthening and building national epidemic detection and response is far more cost effective than mobilizing an international response. In order to improve preparedness in different parts of the world, WHO has formed the Training Programs in Epidemiology and Public health Interventions Network and established a working group on long-term preparedness for outbreak response. Through these programs, Early Warning and Response Networks have been started in partnership with nongovernmental organizations present in the field. Efforts at systematic capacity building among the local communities help sustain the effects of international assistance. Sustained improvements in civil administration through the establishment of a community registry of births and deaths is the most significant step in the path to detection and global preparedness against infectious diseases threats including those related to bioterrorism and will allow the medical community to "think locally and act globally." WHO has issued an updated edition of its standard guide to health aspects of chemical and biological weapons.

Like the national efforts, international efforts of protecting global health are governed by laws. International Health Regulations (IHR) adopted in 1951 provided the only legally binding instrument implemented by WHO, governing the reporting of epidemic-prone diseases and the application of measures to prevent their spread. The IHR's contribution to international governance on infectious diseases was questioned in the late nineteenth century (called *International Sanitary Conventions*, then) and its problems analyzed in the 1960s and 1970s. WHO launched efforts to revise the IHR for new globalization challenges. Currently, the global Network operates without the revised IHR in place. The strengthening of global surveillance under the revised IHR centers on requiring member states to report all "public health emergencies of international concern" to WHO. The definition of public health emergency would be based on WHO-developed criteria, including whether or not the event is serious, unexpected, and likely to involve the international spread and to trigger trade and travel restrictions. WHO has also proposed that the member states make confidential, provisional notifications and have national surveillance systems that meet minimum requirements.

The second objective of IHR is to provide maximum protection against international disease spread with minimum interference with world traffic. In the recent WHO proposals, changes are mentioned as "recommendations." WHO already possesses the power to issue recommendations (WHO Constitution, Article 23). Before any rule becomes binding, World Health Assembly approval is required (WHO Constitution, Article 21). Owing to the guarding of sovereignty by member states and the required approval by the World Health Assembly, the revised IHR would allow WHO to only make recommendations. The revised IHR does not address what happens when member states ignore WHO recommendations. The role of international law in the control of infectious disease is more important today than it has ever been but remains uncertain.

See also Anthrax (*Bacillus anthracis*), Molecular Biology of; Epidemiology, Molecular; Immunologic Memory; Infectious Disease Testing by LCR; Innate Immunity.

Bibliography

Books and Reviews

Davis, C.J. (1999) Nuclear blindness: An overview of the biological weapons programs of the former Soviet Union and Iraq, *Emerg. Infect. Dis.* **5**, 509–512.

Eitzen, E.M., Takafuji, E.T. (1997) Historical Overview of Biological Warfare, in: Sidell, F.R., Takafuji, E.F., Franz, D.R. (Eds.) *Medical Aspects of Chemical and Biological Warfare*, Borden Institute, Washington, DC, pp. 415–423.

Parkman, F. (1901) *The Conspiracy of Pontiac and the Indian War After the Conquest of Canada*, Little Brown & Co Inc., Boston, MA, pp. 2.

Rorberts, B. (1993) New Challenges and New Policy Priorities for the 1990's, in: *Biologic Weapons; Weapons of the Future*, Center for Strategic and International Studies, Washington, DC.

U.S. Army (1977) *Activity in the U.S. Biological Warfare Programs*, U.S. Department of the Army, Washington, DC, Publication Number B193427L.

US Army (2001) *USAMRIID's Medical Management of Biological Casualties Handbook*, US Army Medical Research Institute of

Infectious Diseases, Fort Detrick, Fredrick, MD.

World Health Assembly (1995) Revision and updating of the International health Regulations, World Health Assembly Resolution 48.7, May 12.

World Health Organization (1983) *International Health Regulation*, 3rd edition, The Organization, Geneva, Switzerland.

World Health Organization (2002) *Global crises – Global solutions: Managing public health emergencies through the revised International health Regulations Doc.* WHO/CDS/CSR/GAR/2002.4, The Organization, Geneva.

Zinsser, H. (1963) *Rats, Lice and History*, The Atlantic Monthly Press by Little, Brown and Company, Boston, MA.

Primary Literature

Anna, S.G.J. (2002) Bioterrorism, public health and civil liberties, *N. Engl. J. Med.* **345**, 1337–1342.

Arnon, S.A., Schecter, R., Inglesby, T.V., Henderson, D.A., Bartlett, J.G., Ascher, M.S. et al (2001) For the Working Group on Civilian Biodefense. Botulinum toxin as a biological weapon, *JAMA* **285**, 1059–1070.

Atlas, R.M. (2002) Bioterrorism: from threat to reality, *Annu. Rev. Microbiol.* **56**, 167–185.

Barbera, J., Macintyre, A., Gostin, L. et al. (2001) Large scale quarantine following biological terrorism in the United States, *JAMA* **286**, 2711–2717.

Bartlett, J.G. (1999) Thoughts on bioterrorism, *Ann. Intern. Med.* **131**, 273–280.

Beeching, N.G., Dance, D., Alastair, R.O. et al (2002) Biological warfare and bioterrorism, *BMJ* **321**, 336–339.

Biological Warfare (2001) Preparing for the unthinkable emergency, *ED Manag.* **13**(11), 121–123.

Brumfiel, G. (2002) Mission Impossible? *Nature* **419**, 10–11.

Carus, W.S. (1999) *Bioterrorism and Biocrimes: The Illicit Use of Biological Agents in the 20th Century*, Center for Counterproliferation Research, National Defense University, Washington, DC.

Casadeva, A. (2002) Passive antibody administration (immediate immunity) as a specific defense against biological weapons, *Emerg. Infect. Dis.* **8**, 833–841.

CDC (2002) Biological and chemical terrorism: strategic plan for preparedness and response: Recommendations of the CDC Strategic Planning Workgroup, *MMWR Recomm. Rep.* **49**(RR-4), 1–26.

Christopher, G.W., Eitzen, E.M. (1999) Air evacuation under high-level biosafety containment: The Aeromedical Isolation Team, *Emerg. Infect. Dis.* **5**, 241–246.

Christopher, G.W., Cieslak, T.J., Pavlin, J.A. et al (1997) Biological warfare. A historical perspective, *JAMA* **278**, 412–417.

Danzig, R., Berkowsky, P.B. (1997) Why should we be concerned about biological weapons, *JAMA* **278**, 431–432.

Darling, R.G., Catlett, C.L., Huebner, K.D. et al (2002) Threats in bioterrorism. I:CDC category A agents, *Emerg. Med. Clin. North Am.* **20**, 273–409.

Delon, P.J. (1975) *The International Health Regulations: A Practical Guide*, World Health Organization, Geneva, Switzerland.

Dennis, D.T., Inglesby, T.V., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E. et al (2001) For the Working Group on Civilian Biodefense. Tuleraremia as a biological weapon, *JAMA* **285**, 2763–2773.

Diamond, J. (1999) *Guns, Germs and Steel*, www/Norton and Co, New York.

Dorelle, P. (1969) Old plagues in the jet age: international aspects of present and future control of communicable diseases, *WHO Chron.* **23**, 103–111.

Dorey, E. (2001) US rejects stronger bioweapons treaty, *Nat. Biotechnol.* **19**, 793.

Drexler, M. (2002) *Secret Agents: The Menace of Emerging Infections*, National Academic Press, Washington, DC.

English, J.F., et al. (1999) Bioterrorism readiness Plan: A template for Healthcare Facilities. A document prepared by APIC Bioterrorism Task Force and CDC Hospital Infections Program Bioterrorism Working Group, April 13.

Fidler, D. (2003) Emerging trends in International law concerning global infectious disease control, *Emerg. Infect. Dis.* **9**, 285–290. Vol.

Gostin, L. (2002) *Public Health Law and Ethics: A Reader*, University of California Press and Milbank Memorial Fund, Berkeley and New York.

Gostin, L.O., Sapsin, J.W., Teret, S.P. et al (2002) The model state emergency health powers act, *JAMA* **288**, 622–628.

Hawley, R.J., Eitzen, E.M. Jr. (2001) *Biological* weapons – a primer for microbiologists, *Annu. Rev. Microbiol.* **55**, 235–253.

Henderson, D.A. (1998) Bioterrorism as a public health threat, *Emerg. Infect. Dis.* **4**, 488–492.

Henderson, D.A., Inglesby, T.V., Bartlett, J.G., Ascher, M.S., Eitzen, E., Jahrling, P.B. et al (1999) For the Working Groups on Civilian Biodefense. Small pox as a biological Weapon: Medical and Public Health Management, *JAMA* **281**, 2127–2137.

Henderson, D.A. (2001) Testimony Before the Foreign Relations Committee: Hearing on the Threat of Bioterrorism and the Spread of Infectious Disease, *107th Congress, 1st Session*, University of Pittsburgh Medical Center.

Heyman, D. (2000) *The Global Infectious Disease Threat and its Implication for the United States*, National Intelligence Council, Washington, DC, Publication NIE 99–170.

Heymann, D.L. (2001) Strengthening Global Preparedness for Defense Against Infectious Disease Threats. Committee on Foreign Relations, United States Senate. Hearing on the Threat of Bioterrorism and the Spread of Infectious Diseases.

Hopkins, D.R. (1983) *Princes and Peasants Small pox in History*, University of Chicago Press, Chicago, IL.

Horton, H.H., Misrahi, J.J., Matthews, G.W., Kocher, P.L. (2002) Critical biological agents: Disease reporting as a tool for determining bioterrorism preparedness, *J. Law Med. Ethics* **30**, 262–266.

Howard-Jones, N. (1975) *The Scientific Background of the International Sanitary Conferences 1851–1938*, World Health Organization, Geneva, Switzerland.

Hoffman, R.E., Norton, J.E. (2000) Lessons learned from a full-scale bioterrorism exercise, *Emerg. Infect. Dis.* **6**, 652–653.

Humbaugh, K. (2001) Discussion of the lessons learned from the recent CDC bioterrorism exercise held in Louisville. http://www.louisville.edu/medschool/cme/bioterrorismwebcast.html.

Inglesby, T.V., Dennis, D.T., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E. et al (2000) For the Working Group on Civilian Biodefense. Plague as biological weapon: Medical and public health management, *JAMA* **283**, 2281–2290.

Inglesby, T.V., Grossman, R., O'Toole, T. et al (2001) A plague on your city: observations from TOPOFF, *Clin. Infect. Dis.* **32**, 436–445.

Inglesby, R.V., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Friedlander, A.M. et al (1999) For the working group on civilian biodefense. Anthrax as a biological weapon: medical and public health management, *JAMA* **281**, 2127–22137.

Kaufman, A.F., Meltzer, M., Schmid, G.P. (1997) The Economic Impact of a bioterrorist attack: Are prevention and post attack intervention programs justifiable, *Emerg. Infect. Dis.* **3**, 83–94.

Khan, A.S., Moise, S.A., Lillibridge, S. (2000) Public health preparedness for biological terrorism in the USA, *The Lancet* **356**, 1179–1182.

Lazarus, R., Kleinman, K., Dashevsky, I. et al (2002) Use of automated ambulatory care encounter records for detection of acute illness clusters, including potential bioterrorism events, *Emerg. Infect. Dis.* **8**, 753–760.

Mair, J.S., Sapsin, J., Teret, S. (2002) The model state emergency health powers act and beyond, *Biodefense Q.* **3**, 1–12.

Martinez, B. (2001) Questions of security: Healthpartners use reach, speedy data to hold watch for bioterrorism attacks, *Wall Street J.* **Sect A**, 10.

Meyer, R.F., Moise, S.A. (2002) Bioterrorism preparedness for the public health and medical communities, *Mayo Clin. Proc.* **77**, 619–621.

Miller, J.M. (2001) Agents of bioterrorism; preparing for bioterrorism at the community health care level, *Infect. Dis. Clin. North Am.* **15**, 1127–1156.

National Association of Counties. County Public Health preparedness http://www.naco.org/pubs/surveys/pubhealth/index.cfm.

O'Toole, T., Mair, M., Inglesby, T.V. (2002) Shining light on dark winter, *Clin. Infect. Dis.* **34**, 972–983.

Poupard, J., Miller, L. (1992) History of biological warfare: catapults to capsomeres, *Ann. N. Y. Acad. Sci.* **666**, 9–20.

Relman, D.A., Olson, J.E. (2001) Bioterrorism preparedness: What practitioners need to know, *Infect. Med.* **18**, 497–514.

Richards, C.F., Burstein, J.L., Waeckerle, J.F. et al (1999) Emergency physicians and biological terrorism, *Ann. Emerg. Med.* **34**, 183–190.

Roelsgaard, E. (1974) Health regulations and international travel, *WHO Chron.* **28**, 265–268.

Rotz, L.D., Khan, A.S., Lillibridge, S.R., Ostroff, S.M., Hughes M.H. (2002) Public Health assessment of potential biological terrorism agents, *Emerg. Infect. Dis.* **8**, 225–230.

Tucker, J.B. (1999) Historical trends related to bioterrorism: An empirical analysis, *Emerg. Infect. Dis.* **5**, 498–504.

U.S. Department of Defense (1999) *Annual report, fiscal year 1999*, Walter Reed Army Institute of Research, Silver Spring,(MD).

Velimirovic, B. (1976) Do we still need international health regulations? *J. Infect. Dis.* **133**, 478–482.

Wheelis, M. (2000) Investigating disease outbreaks under a protocol to the biological and toxin weapons convention, *Emerg. Infect. Dis.* **6**, 595–600.

Zalinskas, R.A. (1997) Iraq's biological weapons: The past or future? *JAMA* **278**, 418–424.

Zoon, K.C. (1999) Vaccines, Pharmaceutical products, and bioterrorism: Challenges for the US Food and Drug Administration, *Emerg. Infect. Dis.* **5**, 241–246.

Principles and Applications of Embryogenomics

Vincent Vanburen and Minoru S. H. Ko
National Institute on Aging, National Institutes of Health, Baltimore, MD, USA

1 Introduction 532

2 Approaches 533
2.1 Complementary DNA (cDNA) Projects 533
2.2 Large-scale Analysis of Gene Expression at the Transcriptome Level 536
2.2.1 EST Frequency 536
2.2.2 SAGE 537
2.2.3 Subtractive Cloning 537
2.2.4 Differential Display 538
2.2.5 DNA Microarrays 538
2.2.6 Experimental Design 538
2.2.7 Reproducibility 539
2.2.8 Gene Coverage and Analysis 539
2.2.9 Sensitivity and Specificity 540
2.2.10 Spatial Resolution 540
2.2.11 Large-scale ISH 540
2.3 Large-scale Analysis of Gene Expression at the Proteome Level 541
2.4 Development and Evolution: Comparative Genomics 542
2.5 Functional Genomics/Large-scale Manipulation of Expression 543
2.5.1 Mutagenesis 543
2.5.2 Gene Trapping 543
2.5.3 RNAi/siRNA 544

3 Model Organisms for Embryogenomics 544
3.1 Non-mammalian Animals 545
3.1.1 Drosophila Melanogaster 545
3.1.2 Anopheles Gambiae 545
3.1.3 Caenorhabditis Elegans 545
3.1.4 Xenopus Laevis 547

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

3.1.5 Brachydanio Rerio 547
3.1.6 Strongylocentrotus Purpurratus 547
3.2 Mammalian 548
3.2.1 Homo Sapiens 548
3.2.2 Mus Musculus 548
3.2.3 Rattus norvegicus and others 549
3.3 Plants 549
3.3.1 Arabidopsis Thaliana 550
3.3.2 *Zea Maize* and Others 550
3.4 Suitability of Approaches for Particular Model Organisms Applied to the Study of Development 550

4 Summary Remarks 550

Bibliography 551
Books and Reviews 551
Primary Literature 551

Keywords

BAC
Bacterial artificial chromosome.

cDNA
Complementary DNA, DNA that corresponds to all or part of a gene transcript.

DNA Microarray
Sometimes also called a *"gene chip,"* spots of DNA are placed in a two-dimensional array on a surface, typically either a glass surface or a nylon surface, so that a nucleotide sequence labeled for detection may be hybridized to the DNA on the surface and quantified.

Embryo
After fertilization of an ovum, the resultant zygote and subsequent stages until birth are collectively considered "embryonic." In human and other larger animals, the later stage embryo is called a fetus.

Genome
The entire nucleotide sequence of an organism's chromosomes.

Genomics
Field of study that addresses biological questions at the scale of the genome or transcriptome.

Hybridization
The process of annealing complementary nucleotide sequences, as performed in a variety of techniques, including DNA Microarrays, SAGE, ISH, and RT-PCR.

ISH
In situ hybridization, a technique used to characterize the localization of mRNA in cells and tissues by visualizing (using microscopy) labeled probes that hybridize with RNA in the cell.

PCR
Polymerase Chain Reaction, a method using known probes flanking a nucleotide sequence of interest and cycles of polymerase action to amplify, or make many copies of, that sequence of interest.

RT-PCR
Real Time PCR (RT-PCR), a variation of PCR that scales the amplified nucleotide sequence of interest to the original frequency of that sequence, thereby allowing an accurate measure of the original abundance.

SAGE
Serial Analysis of Gene Expression, a technique that estimates the amount of mRNA expression by ligating small pieces, or tags, of the entire set of expressed transcripts, determining the sequence of this ligated sequence, and counting the occurrence of each tag.

Transcriptome
The collection of sequences and abundances of all transcripts, or RNAs, uniquely collected from a cell type, tissue, organ, or whole organism.

Embryogenomics, or developmental genomics, is defined as a systematic analysis of embryo development with large-scale approaches. This systematic analysis permits consideration of the system as a whole by keeping whole genome sequences and a complete gene catalog in perspective. By merging genomics and developmental biology, embryogenomics contributes to both disciplines: large-scale methodologies and genomewide concepts are contributed to developmental biology and the materials and knowledge of genes specifically expressed in embryos are contributed to genomics. Here, we review the history and recent strides in this effort, including discussions of the genome projects, cDNA library construction, and large-scale experiments and analysis for model organisms that have the greatest impact on the field of embryology. We will emphasize progress in mouse embryogenomics as it provides the basis of discussion for what is arguably the most accessible model organism with a developmental plan that has high similarity to that of humans. Finally, the status of current technologies as they relate to embryogenomics, the obstacles to overcome, and the future aims for this new field are discussed.

1 Introduction

One of the fundamental questions in developmental biology is how the expression of individual genes is regulated in specific cell types and how the differential expression/action of genes drive the development of complex organisms from a simple form (e.g., fertilized eggs in animals). In the 1960s and early 1970s, hybridization kinetic analyses of early embryos and many different cell types was used to examine global transcriptomes and this established that different groups of genes are indeed actively transcribed in each cell type. The majority of these RNA molecules, however, could not be identified at that time. The advent of molecular cloning in mid-1970s made it possible to isolate and characterize individual genes that play crucial roles in the development of organisms. This one-gene-at-a-time approach has become the standard of molecular developmental biology and forms the foundation of our current understanding of development. Using such an approach, however, it is difficult to provide a global view of genome-wide gene actions. Simultaneous expression analyses of hundreds of genes during development began in the early 1980s with hybridization measurements from cDNA clones spotted on membranes. The extension of such analyses to entire genomes began to be formulated in the early 1990s. The earlier approaches can satisfy both global aspects of gene expression as in 1960s to 1970s and individual gene aspects of gene expression since the 1970s, but owing to the much larger scale of whole transcriptome/genome studies, the development of new technologies was inevitable. The Human Genome Project that began in late 1980s revolutionized molecular biology by developing and incorporating large-scale technologies into biology. Massive handling and analysis of DNA clones, aided by robotics and informatics, are for the first time brought into biology. As such, embryogenomics emerges at the interface between genomics and developmental biology.

Genomics offers three main conceptual frameworks to developmental biology. First, large-scale approaches offer investigators the ability to screen a large number of candidate molecules simultaneously to determine which candidates are of interest for further study of the phenomenon of interest. Large-scale efforts in molecular biology offer a cost-effective method of providing comprehensive resources and data sets. The development of permanent, reusable resources (e.g. cDNA clone archives), once and for all establish a strong infrastructure, thus increasing the efficiency of development research. These reusable resources for large-scale approaches offer a huge advantage and increase in the pace of research over previous techniques that allowed only a very limited subset of molecules to be screened in a single experiment. This advantage also brings the paradoxical problem of having more data available from one experiment than one would previously obtain from hundreds or even thousands of experiments prior to the advent of technologies such as DNA microarrays. With this tidal wave of data comes the necessity of improving analytical and data-visualization techniques so that the results can be fully analyzed and understood.

Second, while the cost-effectiveness of comprehensive coverage provides a central motivation for large-scale efforts, an ongoing pursuit in the use of large-scale approaches in molecular biology is motivated in part by the desire to move

from using large-scale endeavors as merely screening tools (data) to synthesizing the results with the results of other techniques to build sophisticated models (knowledge). Embryogenomics poses questions at a large scale compared to how questions are framed with traditional molecular biology techniques. Integrative approaches to large-scale questions have the best hope of providing cohesive answers. While questions that aim at a larger scale have the merits of efficiency and broadened perspective, the real advantage of embryogenomics is the ability to begin considering the combinatorial actions of molecules in the specification of developmental processes. This interface of embryogenomics with systems biology approaches allows a reevaluation of the one-gene-at-a-time paradigm to development research.

Third, the evolutionary aspects of developmental biology will be explored through comparative genomics. The completion or draft completion of many projects to sequence entire genomes has ushered in the so-called postgenomic era for molecular biology. The National Center for Biotechnology Information (NCBI) currently reports some genomic sequence entries for 59 metazoan species (Fig. 1). Comparisons of the genome sequences of different species provides opportunities to identify orthologous genes and conserved regulatory sequences. The emerging view is that molecular mechanisms executed by orthologous genes operate similarly in different species.

Genomics approaches also benefit from the interface with embryogenomics because of the experimental knowledge and materials offered by embryogenomics approaches. While available genome sequences play a central role in large-scale approaches (Fig. 2), much of our understanding of transcript expression and gene structure within the genome comes from cDNA collections, where transcripts or partial transcripts may be assembled into consensuses and mapped to the genome. Construction of these cDNA libraries is normally performed with adult tissues because these tissues are easier to obtain. Consequently, cDNAs uniquely expressed in embryonic tissues will not be represented in these libraries, and inference of gene structure from transcripts could not be achieved for these embryo-specific sequences. Efforts to build libraries from embryonic tissues have contributed to the ability of genomics approaches to identify gene structures, both for the genes actually represented by these libraries and for others through computational predictions using these unique transcripts as part of their training set.

Our definition of embryogenomics is somewhat broad, and so we will focus our discussion primarily on organisms and large-scale approaches that are most aptly suited for investigation of development at the level of the transcriptome.

2 Approaches

2.1 Complementary DNA (cDNA) Projects

Expressed sequence tags (ESTs) are short sequence reads from either the 5′ or 3′ end of complementary DNA (cDNA). Construction of cDNA libraries and large-scale analysis of these libraries forms part of the core of embryogenomics. Building a gene catalog from the fundamental gene identification and gene structure information provided by EST projects is a prerequisite for most of the expression profiling techniques we discuss below (Sects 2.2 and

- *Metazoa*
  - *Protostomia*
    - *Mollusca*
      - *Albinaria coerulea*
      - *Cepaea nemoralis*
      - *Katharina tunicata*
    - *Arthropoda*
      - *Anopheles gambiae*
      - *Drosophila melanogaster*
      - *Drosophila yakuba*
      - *Apis mellifera ligustica*
      - *Locusta migratoria*
      - *Artemia franciscana*
    - *Annelida*
      - *Lumbricus terrestris*
    - *Chordata*
      - *Vertebrata*
        - *Sauropsida*
          - *Alligator mississippiensis*
          - *Gallus gallus*
          - *Rhea americana*
          - *Struthio camelus*
          - *Dinodon semicarinatus*
          - *Pelomedusa subrufa*
        - *Mammalia*
          - *Balaenoptera musculus*
          - *Balaenoptera physalus*
          - *Bos taurus*
          - *Ovis aries*
          - *Ceratotherium simum*
          - *Rhinoceros unicornis*
          - *Equus asinus*
          - *Equus caballus*
          - *Dasypus novemcinctus*
          - *Felis catus*
          - *Halichoerus grypus*
          - *Phoca vitulina*
          - *Gorilla gorilla*
          - *Homo sapiens*
          - *Pan paniscus*
          - *Pan troglodytes*
          - *Pongo pygmaeus*
          - *Mus musculus*
          - *Rattus norvegicus*
          - *Myoxus glis*
          - *Oryctolagus cuniculus*
          - *Didelphis virginiana*
          - *Macropus robustus*

- *Metazoa*
  - *Protostomia*
    - *Chordata*
      - *Vertebrata (continued)*
        - *Amphibia*
          - *Xenopus laevis*
        - *Coelacanthiformes*
          - *Latimeria chalumnae*
        - *Dipnoi*
          - *Protopterus dolloi*
        - *Actinopterygii*
          - *Crossostoma lacustre*
          - *Cyprinus carpio*
          - *Oncorhynchus mykiss*
          - *Salmo salar*
          - *Polypterus ornatipinnis*
        - *Chondrichthyes*
          - *Scyliorhinus canicula*
        - *Agnatha*
          - *Petromyzon marinus*
      - *Cephalochordata*
        - *Branchiostoma lanceolatum*
    - *Echinodermata*
      - *Arbacia lixula*
      - *Paracentrotus lividus*
      - *Strongylocentrotus purpuratus*
      - *Asterina pectinifera*
      - *Florometra serratissima*
      - *Balanoglossus carnosus*
    - *Pseudocoelomata*
      - *Secernentea*
        - *Rhabditia*
          - *Ascaridida*
            - *Ascaris suum*
          - *Rhabditida*
            - *Caenorhabditis elegans*
        - *Spiruria*
          - *Onchocerca volvulus*

Fig. 1 Taxonomy of the 59 metazoan species with genome entries in NCBI's database. These species are listed as those for which NCBI holds submitted genomic sequences. Adapted from the NCBI Web site: http://www.ncbi.nlm.nih.gov/PMGifs/genomes/new_euk.html.

2.3). ESTs give gene signatures, which help define gene structure by alignment with the genome, and they allow for an evaluation of the frequency of randomly selected cDNA clones within a library. The cDNA clones from which ESTs are derived are useful for functional studies of a particular gene or may be used as part of a collection in large-scale efforts such as spotting on cDNA microarrays or in large-scale *in situ* hybridization.

Early efforts at cDNA library construction and EST sequencing typically resulted in clones with relatively short inserts

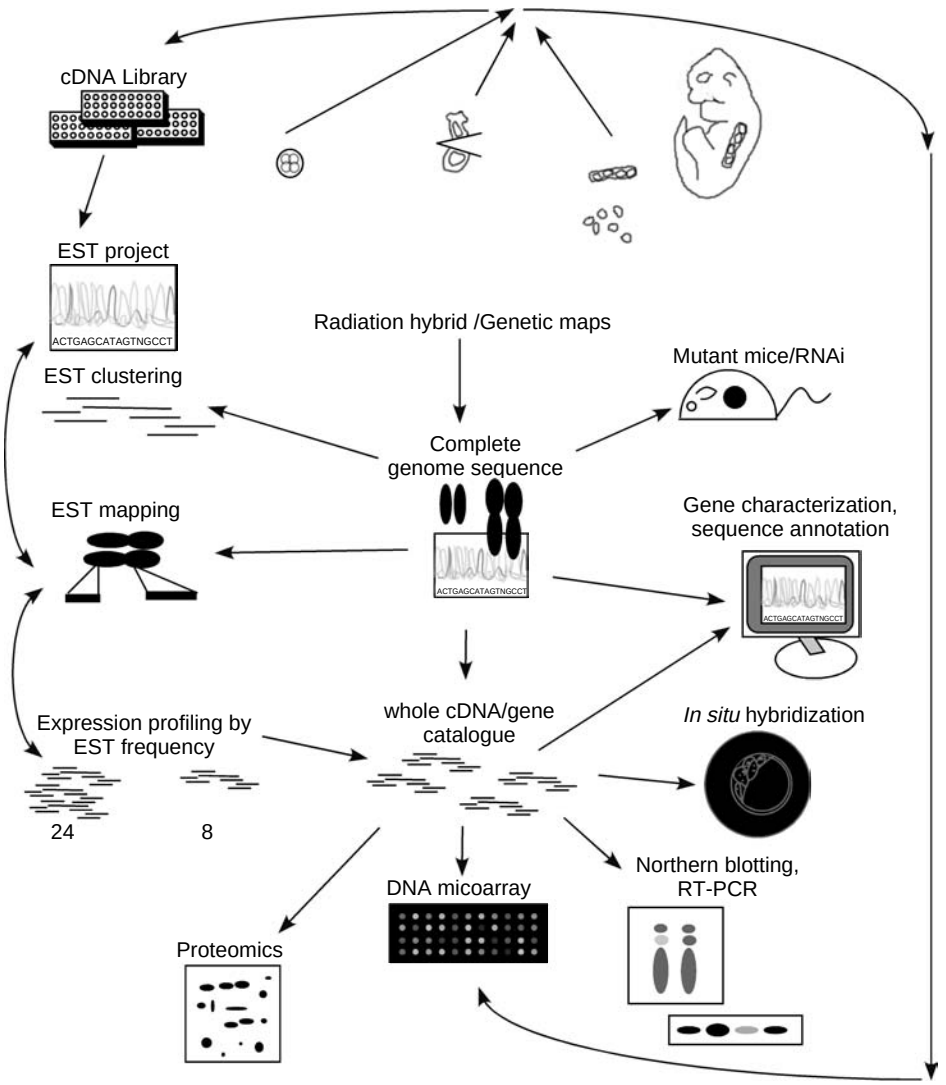

Fig. 2 Available genome sequences play a central role in embryogenomics and other large-scale biology.

(~1.5 kb average insert size) with relatively short EST sequence reads (100–500 bp). These truncated sequences are useful as unique tags for functional studies and for defining gene boundaries. Demand for more complete sequences and gene definitions have driven development of technologies to increase clone insert length and sequence read length, and there is high interest in collecting cDNA clones with full-length inserts.

While several cDNA projects began in the early 1990s, most of these early efforts constructed libraries from adult tissues, resulting in the assembly of gene indices mainly representing adult organisms. As

expression of some genes is likely to be specific to certain developmental stages or tissues, this practice of constructing cDNA libraries from adult tissues results in the exclusion of uniquely expressed developmental genes from gene indices.

2.2 Large-scale Analysis of Gene Expression at the Transcriptome Level

Profiling the expression of genes at a large scale gives valuable insights into which genes are important to, or have unique expression patterns in, a particular cell type, or which genes have enriched or depleted expression in response to a given cue.

Several issues, all concerning the complexity of biochemical and gene regulatory networks *in vivo* and *in vitro*, confound our efforts to analyze expression data and reconstruct gene regulatory networks from that analysis: (1) transcript expression and protein activity sometimes do not correlate; (2) although we readily gather information about how a single gene varies expression between samples, finding absolute expression values so the expression levels of different genes may be compared is not yet a standard practice; (3) comparisons of different cell types may lead to erroneous conclusions because of heterogeneity of some sample collections and volumetric differences between cell types; and (4) even when the experimental conditions are well described (see MIAME compliance, below), it is not yet clear how the results from different experimental platforms and methods may be compared.

Genomics approaches, such as the use of microarrays, force questions or hypotheses to be framed at a different scale or purpose. While genomics approaches may not always give consistent answers or provide biological "truth," the same is true for traditional biological methods, which have certainly provided some mistaken conclusions. Just as integrative approaches are used to independently verify traditional experiments, the same is necessary for microarrays and other high-throughput approaches, and this should not be of any surprise. Conversely, these high-throughput approaches are also not a panacea for biological discovery but instead a useful tool to ask very broad questions that cannot be otherwise asked.

We will first discuss some of the primary techniques used to determine expression profiles without spatial resolution, and then end this section with a discussion of recent efforts at large-scale *in situ* hybridization. In situ hybridization (ISH) offers spatial resolution of transcripts, but most approaches have poor quantitative resolution. ISH thus complements quantitative techniques of expression profiling.

2.2.1 EST Frequency

Standard (not normalized or subtracted) EST libraries are thought to be random samplings of the transcriptome from the cell type or tissue it was derived from. As such, it has been proposed by Meyerowitz that EST frequencies correlate with the relative abundance of each transcript in the sample. EST sequences are obtained from cDNA libraries, which are transcript-encoding sequences or sequence fragments that have been cloned into plasmids and then cultured in a biological host (Fig. 3(a)). The primary consideration for the construction of cDNA libraries is to identify and analyze gene structures, to make these cDNA clone sequences available for spotting on cDNA microarrays, or to make these clones available for downstream functional studies. Availability of EST sequences for the analysis of

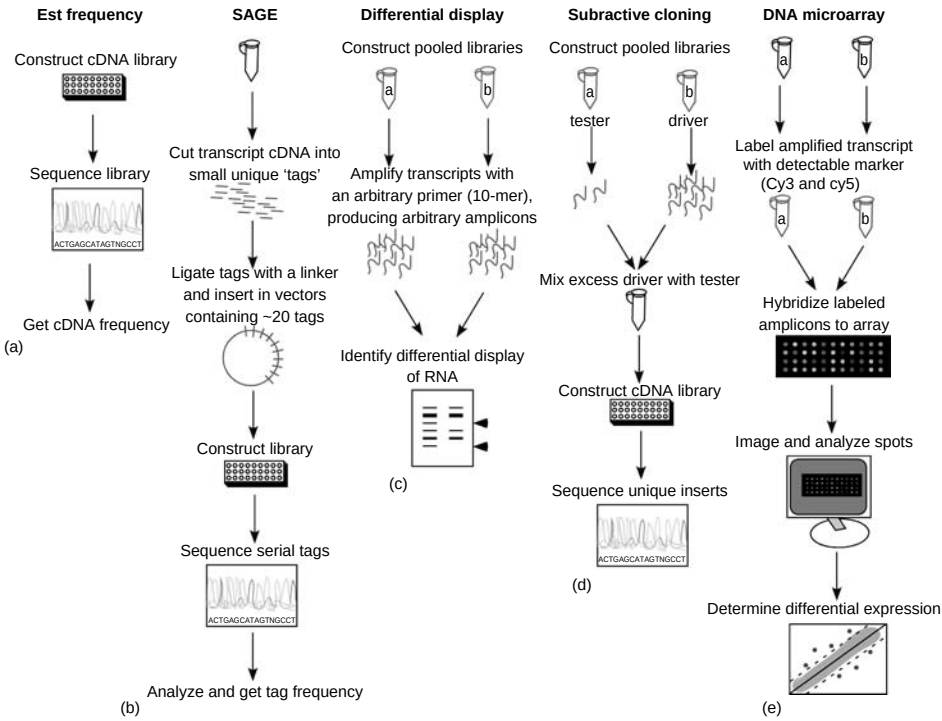

Fig. 3 Techniques for measuring differential expression.

EST frequencies is an additional benefit of cDNA projects, which have the primary goal of providing sequenced cDNA clones as reagents for molecular biology investigations.

2.2.2 SAGE

Serial Analysis of Gene Expression (SAGE) offers an alternate strategy for quantifying gene expression. SAGE is essentially an accelerated version of determining EST frequency from an EST sequencing project. Expressed sequences are cut with restriction enzymes to produce small unique "tags" representing each expressed sequence. These tags are ligated together with linker sequences into vectors containing approximately 20 tags separated by the linker sequences. Sequencing of these vectors allows frequency of the expressed sequences to be determined in a high-throughput capacity, when compared to normal EST sequencing efforts (Fig. 3b). Many plasmid clones must be sequenced for SAGE, but these costly sequencing efforts produce only the expression profiles without producing the reagents to be used for other purposes as in EST projects.

2.2.3 Subtractive Cloning

One strategy for cloning transcripts that are uniquely expressed in one of a pair of transcriptomes is to use subtractive cloning, which removes transcripts common to the pair. An investigator chooses the transcriptome in which they are looking for novel transcripts, called the *tester*, and chooses which transcriptome will be

used to subtract nonunique clones, called the *driver*. Using modern PCR methods, double-stranded cDNA from the cells or tissues of interest is created, linkers are ligated to the ends of the cDNA, and the cDNA pools are amplified using PCR. After amplification, the linkers are removed and unique linkers are ligated to the tester pool. The tester pool is then hybridized to a large excess of the driver. Tester–tester hybrids (cDNA unique to the tester pool) are then amplified using PCR (Fig. 3c). The so-called representational difference analysis (RDA) of cDNA makes use of two rounds of hybridization, because tester–tester hybrids in the first phase of hybridization may arise because of random annealing events and thus may not necessarily be unique to the tester pool.

2.2.4 Differential Display

Another PCR technique used to find unique transcripts is differential display, or RNA fingerprinting, which amplifies the transcriptome by first using reverse transcriptase and oligo-dT primer (or another arbitrary primer) to make cDNA from the transcriptome; then, arbitrary transcript fragments are amplified using poly-dT primer and another arbitrary primer (usually a 10-mer). Differences in transcriptomes are evaluated by electrophoresis of the two amplified pools (Fig. 3d).

2.2.5 DNA Microarrays

DNA microarray technology has been leading the way in offering applications and knowledge-seeking strategies in large-scale biology. Microarrays give a snapshot of gene expression for thousands of transcripts simultaneously.

Three types of microarrays are currently used: (1) cDNAs spotted on either glass slide or nylon membranes; (2) 25-mer oligonucleotides synthesized by the lithographic technique (Affymetrix); (3) 50- to 80-mer oligonucleotides synthesized *in situ* (e.g. Agilent) or spotted on either glass slide or nylon membrane. Transcripts are labeled, typically with Cy3 and Cy5 flourophores for two-channel glass arrays, and with the radioisotope phosphorous-32 for one-channel nylon membranes. When transcripts are collected from rare sample tissues such as preimplantation embryos, they are often linear-amplified by *in vitro* RNA polymerase reaction and then hybridized to the microarray. The intensity of the spot is used as a means of quantifying the transcript abundance relative to the abundance of the same transcript from a different sample, which may be measured in the second channel on the same two-channel array, or measured from a different array when using one-channel arrays (Fig. 3e).

While DNA microarrays are the best large-scale expression profiling method at this point and offer many insights previously inaccessible by traditional experiments, there are still problems facing those who would use microarrays in their investigations, including questions of how to design experiments, how to analyze the data, and the biggest problem: how to achieve reproducibility between different experiments and investigators.

2.2.6 Experimental Design

For the present, the most important aspect of array experiments is proper experimental design. The analysis of microarray data is a subject of discussion and controversy, but several currently popular techniques of analysis will be effective if experiments are properly designed. Some difficulties may arise in choosing a design strategy because no one design is suitable for all types of experiments or for all

practical considerations, such as the availability of resources. Some of the design problems one is faced with include: the biological questions being asked (or which comparisons are the most important to consider), whether to use a universal reference sample, the number of hybridizations that can be performed, sample availability, technical variability, biological variability, and the desired compatibility with future experiments.

2.2.7 Reproducibility

Although many laboratories have produced quality microarray experiments with acceptable reproducibility, the reproducibility of experiments and conclusions between different labs remains a strong challenge to the scientific community. Even when researchers experiment to answer the same questions with microarrays of the same construction, the results and conclusions may be very different. The formulation of guidelines for data sharing in the Minimal Information About a Microarray Experiment (MIAME) standard has been an important step toward realizing the goal of reproducibility. Here "Minimal" is something of a misnomer, as the MIAME standard requires reporting all aspects of the microarray experiment, including detailed information about the platform, sample collection, and all hybridizations.

Standardization of microarray data analysis will also be of benefit to the scientific community by promoting a common understanding of techniques used and by allowing for greater ease of comparing results from different groups. This will come with time as the best techniques emerge from the numerous available ideas. The use of the False Discovery Rate (FDR) is one example of recent progress toward an improved standard of microarray analysis. This method controls the rate of false positives when multiple hypotheses are tested simultaneously. As such, FDR provides a moderate stringency that provides much more reliable results than simple T-statistics, but unlike the stringent Bonferroni correction, FDR will typically uncover some statistically significant differential expressions for further study.

2.2.8 Gene Coverage and Analysis

A more complete coverage of genes by microarrays can be expected to give a better foundation for global assessments and reconstruction of biochemical pathways. One obstacle here is that transcripts may be of low abundance, and thus be missed when looking for genes that are significantly differentially expressed and measurable above background intensities. Without addressing the problem of measuring transcripts with low abundance, complete assessment of the transcriptome cannot be achieved.

In the spirit of attaining a more complete coverage of nucleotide target sequences in an efficient, reproducible experimental system, all probe sequences should be present on the same array, so a single sample can be applied to the probe sequence cohort simultaneously. Future arrays should include, in order of increasing array density, the following:

1. Complete gene coverage (every gene represented at least once)
2. Complete transcriptome coverage (with all alternatively spliced forms)
3. Complete exon coverage (all possible exons)
4. Complete genome coverage (with representative probes evenly spaced throughout genome).

2.2.9 Sensitivity and Specificity

Amplification of target sequences continues to be used in instances where target samples are rare, such as with mammalian oocytes. There has previously been some concern that amplification, however, may skew the apparent abundance of transcripts, especially when multiple rounds of amplification are used. Multiple groups have addressed this problem, and with only a few rounds of linear amplification, reproducibility was found to be good.

Sensitivity of microarray measurements remains one of the largest technical concerns for acquiring comprehensive data. A better probe design is one way to approach the problem of low-abundance transcripts with measured intensities not significantly different from background levels, providing a way to increase measured signal intensity without corrupting the data. Knowledge of how probe qualities affect measurements also gives us the ability to grapple with the problem of deriving absolute abundance estimates from these measures. Current practice of hybridization on two-channel microarrays yields only information about *relative* changes in expression for the same gene between two different samples. To reconstruct genetic/biochemical networks from transcript data, however, knowledge of the *absolute* abundance of transcripts will be invaluable. Some attempts at making a conversion from measured fluorescence intensity have produced useful results, but these early efforts have not been widely applied to the diverse commercial and custom platforms used throughout the scientific community.

2.2.10 Spatial Resolution

The application of DNA microarrays can also address the question of *where* transcription occurs to the extent that tissues or cells of interest may be microdissected or otherwise separated from other tissues and cells before examination with a microarray. Combining laser capture microscopy and microarray analysis seems to offer a solution for this problem, but to have enough spatial resolution, a prohibitively large-number of arrays must be used. This requires the use of ISH techniques described below.

2.2.11 Large-scale ISH

One of the key strategies of investigating developmental processes is to examine the differential distribution of molecules as development progresses. To accomplish this with regard to transcripts, ISH is the most direct way to visualize tissue- or even subcellular-localization of transcripts. Some high-throughput ISH efforts have produced image databases of tremendous value to the scientific community, and efforts such as this are likely to accelerate in the near future with the recent development of robots and other technologies that streamline this laborious technique.

ISH uses hybridization of detectable probes to the transcripts of fixed tissues for the localization of gene expression and offers a characterization of gene expression that is complementary to the quantification of transcript abundance provided by DNA microarrays. Many tissues used in microarray studies are heterogeneous, or made up of several cell types, and so cell-type-specific enriched expression may be missed by microarray studies because of dilution within the heterogeneous tissue. Some tissues are difficult or presently impossible to separate by microdissection for applications in array experiments, so the ability to localize expression patterns to particular cell types should provide many insights inaccessible

to array experiments. Furthermore, cellular compartmentalization of transcripts is also generally not accessible to microarray studies, so ISH offers the ability to explore expression patterns at a higher resolution, which often provides insights that are invaluable for elucidating gene function. Probes with detectable tags attached are created either by synthesis of an oligo from a known sequence or by using PCR-amplified cDNA sequences of the target sequences.

Two recent reports described new gene expression atlases of human chromosome 21 orthologs in mouse at various developmental stages. In the first of these papers, researchers used an ISH robot they developed and an automated microscope to analyze about 6500 sections to generate their atlas covering E9.5, E10.5, E14.5, and adult tissues. Of the 178 confirmed human genes on chromosome 21, the researchers recorded expression for 98% of the 161 mouse orthologs. The second paper, by the HSA21 expression map initiative, describes efforts that examine mapping for 158 mouse homologs of human chromosome 21 confirmed genes (258 human chromosome 21 genes, 168 identified mouse cognates). Here, most of this group's emphasis was on mapping whole mount E9.5 embryos, although a subset of clones was mapped at other stages.

Each of these ISH mapping efforts has produced a freely available, high quality online database for retrieval of the ISH images and the supporting data and methods. Many new gene patterns of expression were discovered in each project. This work has provided an important resource for future work on any of the genes with transcripts mapped by these works.

2.3 Large-scale Analysis of Gene Expression at the Proteome Level

Proteomics, or the large-scale study of proteins, began in 1975, when the introduction of two-dimensional (2-D) electrophoresis allowed for the separation of and quantification of proteins by pI in the first dimension and by size in the second dimension. 2-D gels have been used in the creation of extensive databases cataloging spots and for measuring the differential expression of proteins. While work with 2-D gels is the classical definition of proteomics, large-scale yeast two-hybrid experiments and protein identification and detection of posttranslational modification by mass spectrometry (MS) techniques have joined this field. There are several computational techniques emerging to decipher metabolic networks, including the combination of yeast two-hybrid data with microarray studies to help define network topology. Additionally, efforts to create a "protein-chip" analogous to a DNA microarray are aimed to supplant the utility of 2-D gels.

One promising new technique is the use of matrix-assisted laser desorption ionization mass spectrometry (MALDI MS) of tissues that are laser capture microdissected (LCM) or contact blotted to characterize the spatial distribution of a large number of proteins in thin (~12 μm) tissue slices. MALDI MS techniques such as these join the ranks of immunohistochemistry and fluorescence microscopy for protein detection in tissues or cells, but MALDI MS presents an opportunity to achieve a much larger scale of analysis in a single experiment (hundreds of simultaneous protein measurements compared to one or a few simultaneous measures in the former techniques). MALDI MS

with contact-blotted tissues offers a spatial resolution of ~50 μm and has been used with striking results in sections of mouse brain and human glioblastomas.

2.4 Development and Evolution: Comparative Genomics

Recently it has become clear that many organisms share a similar mechanism of development; in many cases, genes that are conserved among species convey a similar function. For example, some transcription factors from one species can function in cells of other species after transfection.

One approach to finding genes and regulatory sequences essential for execution of a developmental plan is to compare the genomes of closely related species, looking for highly conserved sequences. The converse, looking for genes uniquely expressed in a developmental profile, may yield insights into how that profile generates the unique features of its developmental plan. The two forerunners for this kind of analysis are comparisons between *Drosophila* and *Anopheles*, and between *Mus musculus* and *Homo sapiens*. At this time, comparative genomics approaches have mainly been performed at the global level, without a published account detailing a comparison of genes with important developmental roles between genomes. Analysis of the current draft genomes with a focus on embryology has some beginnings in the supplemental materials accompanying the published comparative genomics of the *Drosophila* and *Anopheles* genomes, but we find little else that approaches this problem in the literature. This scarcity of "comparative embryogenomics" approaches is likely a consequence of our presently poor understanding of large regulatory networks for the specification of development. It is difficult to make comprehensive comparisons of the molecular details of development in different organisms without the corresponding network topologies to use as a conceptual framework of comparison.

It has been proposed by Meyerowitz that comparison of plant and animal developmental processes offers the only true comparative studies of development, as studies between animal species compare only developmental plans that are similar by descent. For example, segmental identity is established in *Drosophila* by spatially specific transcriptional activation of the HOX master regulatory genes. Vertebrates exhibit similar processes in the developing brain. *Arabidopsis*, however, has master regulatory genes, mainly in the MADS family (named for the founding family members Mcm1, Agamous, Deficiens, and serum response factor (SRF)), but these sequences have no similarity to the *HOX* genes. A second pattern specification in animals, the dorsal-ventral specification, has a logical analog in plant, but most of the animal genes involved in this specification are absent in plants. Gurken, the receptor tyrosine kinases, Ras protein, and Pipe, which are all implemented in the dorsal-ventral specification in *Drosophila*, have no homologs in plants. Adaxial–abaxial axes in the leaves of Arabidopsis, the final example of pattern formation given by Meyerowitz, involves activation of the YABBY (named for the australian fresh water crayfish ("yabby") as a link to Crabs claw, the founding member of the gene family) family of transcription factors by the GARP (glycoprotein A repetitions predominant) family of transcription factors. No animal genomes are known to encode members of the YABBY or GARP protein families, again demonstrating the lack of molecular similarity between patterning in

these two kingdoms. Plants and animals have a very different molecular specification for pattern formation, but the logic of that patterning has striking similarity. This relationship presents an opportunity for "true" comparative developmental studies where the focus of study will be how similar logical patterning is achieved with a different molecular toolset.

Now that draft genomes for human and mouse are available, comparative genomics approaches should yield valuable insights into development. The mouse and human genomes have 217 syntenic blocks, from 303 kb to 64.9 Mb in size. 90.2% of the human genome and 93.3% of the mouse genome compose these syntenic regions. While much comparative analysis was completed with the publication of the mouse genome, there was no special attention given to developmental processes in this analysis.

2.5 Functional Genomics/Large-scale Manipulation of Expression

2.5.1 Mutagenesis

The availability of mutant model organisms underpins much of our understanding about development and disease. For the mouse, less than 5% of genes have been mutated, although there are several large-scale mutagenesis projects underway. Most of these mutations were created as targeted mutagenesis by homologous recombination in ES cells. While targeted mutations offer precise selection of the gene to mutate, the standard strategy causes the creation of a null allele. It is often preferable to generate point mutations that may have a range of effects, as this provides more knowledge about protein function. Chemical mutagenesis strategies have the advantages that they allow for phenotype screening, which may be preferable to targeting a specific gene, and they can produce point mutations that have a spectrum of effects on protein function.

Mutagenesis by X-ray, chlorambucil (CHL), and ethylnitrosurea (ENU) all produce random genomic mutations. X-rays produce large mutations and has a low efficiency. CHL also creates large deletions, but creates mutations more efficiently. ENU has the advantage that it creates point mutations. ENU causes ethylation of either O or N, with the most frequent base pair changes being A/T→T/A (44%), A/T→G/C (38%), and G/C→A/T (8%), most of which are missense mutations (64%). ENU is used with the idea of affecting the early spermatogonial cells of male mice, so these males will produce mutated sperm for the remainder of their lives (after a period of sterility).

2.5.2 Gene Trapping

Gene trapping is a technique that permits reverse genetics and provides a mechanism that does not require additional steps to isolate the gene responsible for the phenotype of interest. The principle method of gene trapping is to randomly insert a promoterless construct with reporter, selectable marker, and splice site(s) into the genome. Proper integration and splicing of this construct will result in the creation of a fusion protein with one or more exons of the trapped locus and, ideally, the native transcript is truncated, so it will be nonfunctional.

Constructs with a splice acceptor (SA) or both a SA and a splice donor (SD) are the most commonly used gene trap vectors. The murine engrailed-2 and the adeno virus late major transcript SA are both widely used, and the human *BCL-2* gene intron 2/exon 3 SA has been introduced more recently. One potential problem is

that typical reporters are only detectable if the insertion is in-frame, but with gene-trapping the insertion of the trapping vector is random. Reporter activity is enhanced by the insertion of an internal ribosomal entry site (IRES) between the SA and the reporter in the gene trap vector. The IRES allows independent translation of the reporter, and thus reporter detection does not require in-frame gene trapping when the gene trap vector contains an IRES.

2.5.3 RNAi/siRNA

RNA interference (RNAi) is perhaps the most exciting technology introduced to molecular biology in recent years. RNAi takes advantage of cellular mechanisms that carry out the destruction of specified transcripts and may provide a loss-of-function phenotype, thus providing insights and motivation for further analysis. Before the application of RNAi, gene knockouts and knockdowns could take many months or even years to achieve useful results. Now, large-scale transient inactivation of individual genes (in *Caenorhabditis elegans*, for example) can be achieved for thousands of genes (separately) in a few months (see Sect. 3.1.3).

Briefly, the mechanism of RNAi begins in the cell with the introduction of double-stranded RNA (dsRNA), which mediates interference with the target of interest. Short interfering RNAs (siRNAs), 21- to 23-nt, are produced by the cutting action of the enzyme Dicer on dsRNA. The siRNAs are incorporated into the RNAi-induced silencing complex (RISC), which is a multisubunit protein complex. The RISC may unwind the siRNA, thus facilitating interaction with the target transcript. Those siRNAs that interact with targets lead to target cleavage by DICER and liberate new siRNA as a result. The remaining mRNA target is degraded by exonucleases. It also seems likely that siRNA can act as a primer for an RNA-dependent polymerase (RdRP), which would allow for the amplification of siRNA sequences. The positive feedback loop provided by an RdRP could explain how a relatively small amount of siRNA can cause the degradation of a larger amount of target and how it is that RNAi is inheritable in worms.

RNAi may be applied to many eukaryotic systems, but of the metazoans, *C. elegans* is particularly well suited for this approach because these worms may be fed bacteria transfected with the dsRNA used as an RNAi stimulus.

3 Model Organisms for Embryogenomics

Model organisms are chosen in order to optimize suitability for a particular type of investigation and the utility of the knowledge gained in the study. By "suitability," we mean the accessibility of the organism's physiology to investigation, including considerations such as gestation period, physical properties of the body tissues of interest, and overall effort required in maintenance of the organism. Utility of the knowledge gained refers to the closeness of the model organism's genetics and physiology to the organism (or more generally, organisms) we are modeling (for example, Homo sapiens). Here we discuss the various model organisms studied in embryology using genomics approaches. We will consider the advantages and disadvantages of the various model organisms, and review the current status of genomics approaches to developmental questions. Each organism discussed here has a rich history as a model for developmental studies and so our discussions will necessarily be cursory and highlight

the salient features of each organism and their place in the study of development in general and in embryogenomics.

3.1 Non-mammalian Animals

3.1.1 Drosophila Melanogaster

The developmental stages of metamorphosis in the fruit fly *Drosophila melanogaster*, hereafter *Drosophila*, are the first developmental plans to be studied with microarray analysis. *Drosophila* has a small genome, an extremely short life cycle, and is relatively inexpensive to maintain, which makes this organism appealing for use in scientific investigations. *Drosophila* has been the most powerful system for genetic analysis with the ease of mutant screening and the large archives of such mutants. Completion of the *Drosophila* genome has increased the value of this model species further. There is a large body of experimental knowledge about development in *Drosophila* and over 125 000 cDNA clones from *Drosophila* have been derived from embryonic stages (Table 1). Systematic large-scale ISH of 2179 genes has also been reported recently.

3.1.2 Anopheles Gambiae

The genome sequencing of the malaria mosquito, *Anopheles gambiae*, hereafter *Anopheles*, was published in 2002, thus providing the opportunity to compare the genomes of the two dipteran species *Drosophila* and *Anopheles*. Prior to the availability of this genome sequence, interest in *Anopheles* was primarily focused on the prevention and treatment of malaria. Although there has been interest in the life cycle of *Anopheles*, this has mainly been with the goal of understanding the corresponding life cycle of *Plasmodium falciparum* (the malaria-causing organism that invades erythrocytes) better, and there are no cDNA libraries with sequences in dbEST that report derivation from *Anopheles* embryonic stages.

Comparative genomics analysis of *Anopheles* and *Drosophila* appeared in a companion paper to the *Anopheles* draft genome paper. Although most of this analysis emphasized a global comparison, the supplementary materials for this work included a simple analysis of some developmentally important genes. Flybase (*Drosophila*) was queried with keywords related to embryos and development to identify 314 *Drosophila* genes (after manual corrections), and the orthologs in *Anopheles* were identified. This information was placed in a table according to *Drosophila* gene name, and the number of *Anopheles* orthologs was listed. This procedure demonstrated that most of the characterized *Drosophila* genes listed that are important in development had orthologs in *Anopheles* (88%), whereas ~12% of these developmentally important genes appear to be absent in *Anopheles*. This analysis offers some clues about genes that may drive important differences in the developmental plan of *Drosophila*. The absence of the reciprocal analysis underscores that there is much less gene information available for *Anopheles*.

3.1.3 Caenorhabditis Elegans

Caenorhabditis elegans, hereafter *C. elegans*, has been of profound importance for developmental studies, primarily because it has been possible to follow the developmental lineage and migration of every cell in the organism from one cell through adulthood.

In addition to the usefulness of *C. elegans* in small-scale functional studies of development, this organism has proved useful beyond its anticipated advantages as a model organism because of the application

Tab. 1 EST sequences (S) and unique clones (C) reported for cDNA libraries by NCBI unigene (as of July 18, 2003).

	Preimplantation (S)	Preimplantation (C)	Embryo/fetus (S)	Embryo/fetus (C)	Neonate/newly enclosed/seedling (S)	Neonate/newly enclosed/seedling (C)	Beyond neonate/adult (S)	Beyond neonate/adult (C)	Umbilical cord/placenta (S)	Umbilical cord/placenta (C)	Others and/or unclassified (S)	Others and/or unclassified (C)	Total (S)	Total (C)
Homo sapiens	0	0	336 833	225 975	76 338	48 248	744 771	621 783	58 042	38 745	2 864 345	2 611 912	4 080 329	3 546 663
Mus musculus	123 444	115 579	811 031	724 873	358 450	315 208	1 208 998	1 043 887	15 472	12 584	1 123 943	937 396	3 641 338	3 149 527
Rattus norvegicus	0	0	39 549	34 416	244	242	212 488	186 515	12 304	12 159	248 933	244 231	513 518	477 563
Brachydanio rerio	N/A	N/A	42 967	34 962	0	0	123 890	105 638	N/A	N/A	96 963	78 711	263 820	219 311
Xenopus laevis	N/A	N/A	138 154	95 538	0	0	35 811	25 099	N/A	N/A	117 104	87 857	291 069	208 494
Anopheles gambiae	N/A	N/A	0	0	0	0	82 983	82 971	N/A	N/A	15 857	13 161	98 840	96 132
Drosophila melanogaster	N/A	N/A	132 196	126 470	11 454	10 767	90 530	85 900	N/A	N/A	21 164	20 083	255 344	243 220
Caenorhabditis elegans	N/A	N/A	134 346	76 165	0	0	0	0	N/A	N/A	52 886	31 814	187 232	107 979
Arabidopsis thaliana	N/A	N/A	39 239	31 736	1639	1082	41 166	36 036	N/A	N/A	84 548	78 734	166 592	147 588
													9 498 082	8 196 477

Note: Most "others and/or unclassified" may be regarded as from adult or from mixed sources.

of RNAi. *C. elegans* may be fed bacteria carrying RNAi vectors, and these vectors can cause the destruction of targeted native transcripts in the worm. Kamath, et al. have recently performed a comprehensive functional analysis of *C. elegans* using RNAi. By their count, there are 19 427 predicted genes in *C. elegans*, and they inhibited transcript function for 16 757 (about 86%) of this total. 1528 single-gene mutant phenotypes were identified, where a mutant phenotype was defined as reproducible sterility, embryonic or larval lethality, slow postembryonic growth, or any of a number of specifically defined postembryonic defects. This work also established that there is high reproducibility of RNAi techniques, regardless of methods, and showed there is a false positive rate of <1%. Kamath, et al. claim to have identified more than 70% of all embryonic lethal phenotypes in *C. elegans*. Their efforts have produced both an intellectual contribution and a tremendous resource for the scientific community.

3.1.4 Xenopus Laevis

Xenopus laevis has been of value to embryology in the areas of early fate decisions and morphogenesis (with detailed fate maps constructed for blastula and from gastrula through organogenesis), patterning of the vertebrate body plan, and early organogenesis. Moreover, several functional genomics studies have focused on *X. laevis* development. When females are injected with human chorionic gonadotropin, they ovulate and lay hundreds of oocytes that may be manipulated or subjected to *in vitro* fertilization. The early embryos are quite large and are thus suitable for microinjections and dissections. The embryos are also transparent, thus making the effects of experimental conditions easy to visualize. Recently, *X. tropicalis* has been identified as a model system that retains the advantages of *X. laevis*, but has the additional benefits of having a diploid genome suitable for genetics study and a shorter generation time, making this species more accessible to multigenerational approaches.

3.1.5 Brachydanio Rerio

The zebrafish *Brachydanio rerio*, hereafter *Danio*, has been a cornerstone model organism for developmental processes because *Danios* are extremely hardy and easy to raise, a large number of *Danio* mutants have been created by mutagenesis, and most importantly, *Danio* embryos are transparent, thereby facilitating their study by noninvasive observation. Large-scale ISH has been done to identify genes specific to embryonic stages in *Danios*. There are more than 34 000 cDNA clones derived from *Danio* embryos available (Table 1). Recently, the first cDNA microarray constructed from *Danio* cDNA libraries, including a library derived from three-day-old embryonic heart, has been used to study development and hypoxia.

3.1.6 Strongylocentrotus Purpurratus

The long history of the sea urchin as a developmental model has its beginnings with the discovery by Theodore Boveri that chromosomes contain information for controlling the development of embryos. Of the developmental plans in the metazoans, only the sea urchin endomesoderm specification has been reconstructed from existing experimental knowledge. This tremendous effort combined many years of acquired experimental knowledge with computer simulation to construct what Davidson, et al., describe as a provisional model for the specification of the endomesoderm in *Strongylocentrotus purpurratus*.

3.2 Mammalian

Studies of mammalian systems are expected to offer the most relevant insights into human developmental processes. We begin our discussion of mammalian model organisms with a look at human, mouse, and rat as systems for inquiry into developmental questions. Moral and practical considerations severely restrict studies with human tissues, so the emphasis of our discussion will be on the nonhuman mammals, *Mus musculus* and *Rattus norvegicus*, with most of our attention on *Mus musculus*, the preeminent model organism that most closely matches human developmental patterns.

3.2.1 Homo Sapiens

Mammalian species are of greatest interest to embryonic investigations because the traditional view is that these species will provide the most relevant model for the human. Ethical concerns limit the availability of human embryonic tissues and essentially forbid any *in vivo* embryological experimentation. Despite these limitations, several large-scale experimental studies have been performed using human embryo tissues, and there is great interest in studies with stem cells derived from human. The human draft genome sequence was made available in 2001 and the completed human genome was announced in March 2003, although new analysis of the present sequence has not been published as of this writing. The recent availability of the mouse genome will greatly contribute to our understanding of regulatory elements and conserved genes important to development by comparison with the human sequence.

3.2.2 Mus Musculus

The house mouse *Mus musculus* is the forerunner of model organisms for use in studies of development and pluripotency. Similarity to human development, short generation time, and the relatively low cost of maintenance all contribute to the notion that *Mus musculus* is the optimal model organism for relevant and rapid large-scale developmental investigations. Of mouse, rat, and human, only mouse has cDNA libraries derived from preimplantation stage embryos. Transcript expression specific to these early embryonic stages may be overlooked in rat and human because of the difficulty in working with and acquiring these rare tissues. Another advantage of the mouse as a model is that the largest number of full-length cDNA clones have been assembled and sequenced for mouse.

In response to the underrepresentation of embryonic tissues in cDNA collections, many new libraries have been created from these rare tissues in mouse. The availability of over 115 000 cDNA clones derived from preimplantation tissues in mouse offer a compelling reason to choose a mouse model for study of early development. Sequencing and subsequent analysis has identified many novel or rare transcripts from early embryonic stages. Given multiple EST sequences from these libraries, an effort to assemble collections of representative transcripts culminated in the construction of the NIA Mouse 15K and NIA Mouse 7.4K cDNA collections. Each of these collections has been distributed worldwide to more than 100 cDNA microarray facilities and these clones are used in both spotting cDNA microarrays and in downstream functional studies.

The human and mouse genomes together represent valuable tools for comparative studies, revealing important *cis*

regulatory elements, definition of gene structures, and key genes for shared developmental processes. The draft sequence and analysis of the mouse genome was published recently. 1859 Mb of the sequence is draft sequence (74%) and 477 Mb is finished sequence (19%) for a total mouse genome size of 2.5 Gb (Human, 2.9 Gb). Over 96% of the assembled nucleotides have only about 1 mismatch per 10 000 bases. When finished bacterial artificial chromosome (BAC) sequences were compared to the assembly, the discrepancy rate was 0.005%, corresponding to about 1 mismatch per 20 000 bases. Comparisons of the high quality mouse and human genomes underscore the usefulness of comparative studies. For example, the Sox family of transcription factors is important to many developmental processes in both mouse and human and includes the *Sry* gene on the Y chromosome, which specifies testis determination. A comparative analysis of the human and mouse genomes has determined the precise number and ortholog identity of the members of Sox family of transcription factors, thus resolving the nomenclature for this important family of genes.

The utility of dsRNA for RNAi in mouse oocytes and preimplantation embryos has been demonstrated. It was shown that dsRNA could be used to prevent gene expression of a GFP transgene (MmGFP) in heterozygotes where the transgene was parentally derived. Two-cell expression of the GFP transgene was prevented by injection of dsRNA MmGFP in the one-cell zygote. This work showed that strong, specific depletion of target transcripts could be achieved from the four- to six-cell stage through the blastocyst stage. It was later demonstrated that hairpin RNA could achieve similar results. For these promising techniques, the main difficulty is that injection of the RNAi vector is required in each instance. This difficult procedure currently limits the extent of the application of RNAi in these early embryonic tissues, and poses a challenge to any present notion of making this procedure large scale.

3.2.3 Rattus norvegicus and others

The first genome assembly of *Rattus norvegicus*, hereafter rat, was recently announced by the Rat Genome Sequencing Consortium (announcement June 19, 2003). The rat genome is slightly larger than the mouse genome (2.57 Gb compared to 2.5 Gb). Details of the assembly, as well as comparative genomics of rat to another species, awaits future publication as of this writing. Rat is primarily used as a model organism for cardiovascular or physiological research, but the availability of genome sequence may spark interest in comparative genomics approaches toward a better understanding of the common features of genetic programs of development among mammals. Similarly, genome projects for other mammals or vertebrates (Fig. 1) may also spark interest in large-scale comparative studies of development.

3.3 Plants

Plant embryology enters our discussion because although plant and animal development and physiology are quite distinct at the molecular level, many of the experimental procedures and analysis share common techniques and strategies for elucidation of developmental pathways. There is also the additional motivation that comparative studies between plant and animal development may yield evolutionary insights and provide clues about

convergence of some aspects of their developmental plans (see Sect. 2.4).

3.3.1 Arabidopsis Thaliana

Completion of genome sequencing for the thale cress *Arabidopsis thaliana* has strengthened the utility of this plant model for comparative approaches to development. Arabidopsis has long been of interest for genomics and developmental studies because of its small genome and short germination and development cycle. Of plants, *Arabidopsis* is the canonical species for large-scale studies.

3.3.2 *Zea Maize* and Others

Crop plants like *Zea maize* (corn) and *Orzo sativa* (rice) are heavily studied because of their importance as a food source for much of the world. Genome projects for each of these organisms are progressing and will contribute to the comparative genomics approaches between plants and the more general comparisons of development between kingdoms.

3.4 Suitability of Approaches for Particular Model Organisms Applied to the Study of Development

The traditional view has been that mammalian systems are closest to the human in genome organization and developmental plan, so these systems are of greatest interest as model species. The biggest trade-off here is that viviparity makes mammalian systems much more difficult to work with than egg-laying organisms. The increased cost and effort involved in working with mammalian systems provides incentive for working with other organisms that are more suitable for some large-scale efforts. For example, the nematode *C. elegans* has been an apt choice for developmental studies including large-scale mutagenesis for many years. More recently, *C. elegans* is the first metazoan to be extensively studied using RNAi to selectively render transcripts nonfunctional.

The primary benefits of the mouse as a developmental model are the closeness of developmental plan to that of the human and the availability of a large number of cDNA clones throughout the preimplantation stages and later embryo development. *C. elegans* offers a rich base of experimental knowledge, relatively easy application of RNAi, a large number of available mutants, and easy maintenance. *Xenopus* offers a replenishable source of oocytes and a large oocyte (~1 mm) that may be collected and manipulated relatively easily. *Danios* offer a rich base of experimental knowledge, a large number of available mutants, a translucent embryo for easy visualization of development, and are easily maintained.

4 Summary Remarks

Embryogenomics presents an exciting new area that continues to grow and draw great interest from the scientific community. Although there are some difficulties with current hardware and analysis technologies, these issues are being aggressively addressed in an effort to approach the attractive goal of effective systems biology approaches that yield accurate predictions that guide our understanding of development and provide fertile ground for future investigations.

See also Developmental Cell Biology.

Bibliography

Books and Reviews

Anderson, K.V., Ingham, P.W. (2003) The transformation of the model organism: a decade of developmental genetics, *Nat. Genet.* **33**(Suppl.), 285–293.

Chaurand, P., Schwartz, S.A., Caprioli, R.M. (2002) Imaging mass spectrometry: a new tool to investigate the spatial organization of peptides and proteins in mammalian tissue sections, *Curr. Opin. Chem. Biol.* **6**, 676–681.

Chen, Y., Xu., D. (2003) Computational analyses of high-throughput protein-protein interaction data, *Curr. Protein Pept. Sci.* **4**, 159–181.

Davidson, E.H. (1986) *Gene Activity in Early Development*, Academic Press, Orlando, FL.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z., Schilstra, M.J., Clarke, P.J., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H. (2002a) A genomic regulatory network for development, *Science* **295**, 1669–1678.

Fishman, M.C. (2001) Genomics. Zebrafish–the canonical vertebrate, *Science* **294**, 1290–1291.

Gerhart, J., Kirschner, M. (1997) *Cells, Embryos, and Evolution*, Blackwell Science, Cambridge, MA.

Hirsch, N., Zimmerman, L.B., Grainger, R.M. (2002) Xenopus, the next generation: X. tropicalis genetics and genomics, *Dev. Dyn.* **225**, 422–433.

Ko, M.S. (2001) Embryogenomics: developmental biology meets genomics, *Trends Biotechnol.* **19**, 511–518.

Kohane, I.S., Kho, A.T., Butte, A.J. (2003) *Microarrays for an Integrative Genomics*, MIT Press, Cambridge, MA, p. 306.

Lennon, G.G., Lehrach, H. (1991) Hybridization analyses of arrayed cDNA libraries, *Trends Genet.* **7**, 314–317.

Lohr, J.L., Yost, H.J. (2000) Vertebrate model systems in the study of early heart development: Xenopus and zebrafish, *Am. J. Med. Genet.* **97**, 248–257.

Meyerowitz, E.M. (2002) Plants compared to animals: the broadest comparative study of development, *Science* **295**, 1482–1485.

Pandey, A., Mann, M. (2000) Proteomics to study genes and genomes, *Nature* **405**, 837–846.

Perkins, A.S. (2002) Functional genomics in the mouse, *Funct. Integr. Genomics* **2**, 81–91.

Raff, R.A. (1996) *The Shape of Life*, University of Chicago Press, Chicago and London, IL.

Reinke, V., White, K.P. (2002) Developmental genomic approaches in model organisms, *Annu. Rev. Genomics Hum. Genet.* **3**, 153–178.

Schlessinger, D., Ko, M.S. (1998) Developmental genomics and its relation to aging, *Genomics* **52**, 113–118.

Schuler, G.D. (1997) Pieces of the puzzle: expressed sequence tags and the catalog of human genes, *J. Mol. Med.* **75**, 694–698.

Somerville, C., Somerville, S. (1999) Plant functional genomics, *Science* **285**, 380–383.

Primary Literature

Adams, M.D., Soares, M.B., Kerlavage, A.R., Fields, C., Venter, J.C. (1993) Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library, *Nat. Genet.* **4**, 373–380.

Adams, M.D., Celniker, S.E., Holt, R.A., et al. (2000) The genome sequence of Drosophila melanogaster, *Science* **287**, 2185–2195.

Arbeitman, M.N., Furlong, E.E., Imam, F., Johnson, E., Null, B.H., Baker, B.S., Krasnow, M.A., Scott, M.P., Davis, R.W., White, K.P. (2002) Gene expression during the life cycle of Drosophila melanogaster, *Science* **297**, 2270–2275.

Aronow, B.J., Richardson, B.D., Handwerger, S. (2001) Microarray analysis of trophoblast differentiation: gene expression reprogramming in key gene function categories, *Physiol. Genomics* **6**, 105–116.

Benjamini, Y., Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B.* **57**, 289–300.

Bogi, C., Levy, G., Lutz, I., Kloas, W. (2002) Functional genomics and sexual differentiation in amphibians, *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **133**, 559–570.

Boyes, D.C., Zayed, A.M., Ascenzi, R., McCaskill, A.J., Hoffman, N.E., Davis, K.R., Gorlach, J. (2001) Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants, *Plant Cell.* **13**, 1499–1510.

Carter, M.G., Hamatani, T., Sharov, A.A., Carmack, C.E., Qian, Y., Aiba, K., Ko, N.T., Dudekula, D.B., Brzoska, P.M., Hwang, S.S., Ko, M.S. (2003) In situ-synthesized novel microarray optimized for mouse stem cell and early developmental expression profiling, *Genome Res.* **13**, 1011–1021.

Carulli, J.P., Artinger, M., Swain, P.M., Root, C.D., Chee, L., Tulig, C., Guerin, J., Osborne, M., Stein, G., Lian, J., Lomedico, P.T. (1998) High throughput analysis of differential gene expression, *J. Cell. Biochem. Suppl.* **30–31**, 286–296.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Schilstra, M.J., Clarke, P.J., Rust, A.G., Pan, Z., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H. (2002b) A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo, *Dev. Biol.* **246**, 162–190.

Drees, B.L., Sundin, B., Brazeau, E., Caviston, J.P., Chen, G.C., Guo, W., Kozminski, K.G., Lau, M.W., Moskow, J.J., Tong, A., Schenkman, L.R., McKenzie 3rd, A., Brennwald, P., Longtine, M., Bi, E., Chan, C., Novick, P., Boone, C., Pringle, J.R., Davis, T.N., Fields, S., Drubin, D.G. (2001) A protein interaction map for cell polarity development, *J. Cell Biol.* **154**, 549–571.

Fiehn, O., Weckwerth, W. (2003) Deciphering metabolic networks, *Eur. J. Biochem.* **270**, 579–588.

Finkelstein, D., Ewing, R., Gollub, J., Sterky, F., Cherry, J.M., Somerville, S. (2002) Microarray data quality analysis: lessons from the AFGC project. Arabidopsis functional genomics consortium, *Plant Mol. Biol.* **48**, 119–131.

Floss, T., Wurst, W. (2002) Functional genomics by gene-trapping in embryonic stem cells, *Methods Mol. Biol.* **185**, 347–379.

Gepstein, L. (2002) Derivation and potential applications of human embryonic stem cells, *Circ. Res.* **91**, 866–876.

Gitton, Y., Dahmane, N., Baik, S., Ruiz i Altaba, A., Neidhardt, L., Scholze, M., Herrmann, B.G., Kahlem, P., Benkahla, A., Schrinner, S., Yildirimman, R., Herwig, R., Lehrach, H., Yaspo, M.L. (2002) A gene expression map of human chromosome 21 orthologues in the mouse, *Nature* **420**, 586–590.

Grammer, T.C., Liu, K.J., Mariani, F.V., Harland, R.M. (2000) Use of large-scale expression cloning screens in the Xenopus laevis tadpole to identify gene function, *Dev. Biol.* **228**, 197–210.

Haab, B.B., Dunham, M.J., Brown, P.O. (2001) Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions, *Genome Biol.* **2**, RESEARCH0004.

Harper, J.C., Wells, D. (1999) Recent advances and future developments in PGD, *Prenat. Diagn.* **19**, 1193–1199.

Holt, R.A., Subramanian, G.M., Halpern, A., et al. (2002) The genome sequence of the malaria mosquito Anopheles gambiae, *Science* **298**, 129–149.

Initiative, T.A.G. (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana, *Nature* **408**, 796–815.

Ivanova, N.B., Dimos, J.T., Schaniel, C., Hackney, J.A., Moore, K.A., Lemischka, I.R. (2002) A stem cell molecular signature, *Science* **298**, 601–604.

Jacob, H.J. (1999) Functional genomics and rat models, *Genome Res.* **9**, 1013–1016.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D.P., Zipperlen, P., Ahringer, J. (2003) Systematic functional analysis of the Caenorhabditis elegans genome using RNAi, *Nature* **421**, 231–237.

Ko, M.S. (1990) An 'equalized cDNA library' by the reassociation of short double-stranded cDNAs, *Nucleic Acids Res.* **18**, 5705–5711.

Ko, M.S., Kitchen, J.R., Wang, X., Threat, T.A., Hasegawa, A., Sun, T., Grahovac, M.J., Kargul, G.J., Lim, M.K., Cui, Y., Sano, Y., Tanaka, T., Liang, Y., Mason, S., Paonessa, P.D., Sauls, A.D., DePalma, G.E., Sharara, R., Rowe, L.B., Eppig, J., Morrell, C., Doi, H. (2000) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, *Development* **127**, 1737–1749.

Komiya, T., Tanigawa, Y., Hirohashi, S. (1997) A large-scale in situ hybridization system using an equalized cDNA library, *Anal. Biochem.* **254**, 23–30.

Kudoh, T., Tsang, M., Hukriede, N.A., Chen, X., Dedekian, M., Clarke, C.J., Kiang,

A., Schultz, S., Epstein, J.A., Toyama, R., Dawid, I.B. (2001) A gene expression screen in zebrafish embryogenesis, *Genome Res.* **11**, 1979–1987.

Lander, E.S., Linton, L.M., Birren, B., et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Maslov, S., Sneppen, K. (2002) Specificity and stability in topology of protein networks, *Science* **296**, 910–913.

Menssen, A., Hermeking, H. (2002) Characterization of the c-MYC-regulated transcriptome by SAGE: identification and analysis of c-MYC target genes, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6274–6279.

Neidhardt, L., Gasca, S., Wertz, K., Obermayr, F., Worpenberg, S., Lehrach, H., Herrmann, B.G. (2000) Large-scale screen for genes controlling mammalian embryogenesis, using high-throughput gene expression analysis in mouse embryos, *Mech. Dev.* **98**, 77–94.

O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins, *J. Biol. Chem.* **250**, 4007–4021.

Okazaki, Y., Furuno, M., Kasukawa, T., et al. (2002) Analysis of the mouse transcriptome based on functional annotation of 60 770 full-length cDNAs, *Nature* **420**, 563–573.

Okubo, K., Matsubara, K. (1997) Complementary DNA sequence (EST) collections and the expression information of the human genome, *FEBS Lett.* **403**, 225–229.

Pollet, N., Schmidt, H.A., Gawantka, V., Niehrs, C., Vingron, M. (2000) In silico analysis of gene expression patterns during early development of Xenopus laevis, *Pac. Symp. Biocomput.* **5**, 440–451.

Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R.C., Melton, D.A. (2002) "Stemness": transcriptional profiling of embryonic and adult stem cells, *Science* **298**, 597–600.

Reymond, A., Marigo, V., Yaylaoglu, M.B., Leoni, A., Ucla, C., Scamuffa, N., Cacciopoli, C., Dermitzakis, E.T., Lyle, R., Banfi, S., Eichele, G., Antonarakis, S.E., Ballabio, A. (2002) Human chromosome 21 gene expression atlas in the mouse, *Nature* **420**, 582–586.

Rubin, G.M., Yandell, M.D., Wortman, J.R., Gabor Miklos, G.L., Nelson, C.R., Hariharan, I.K., Fortini, M.E., Li, P.W., Apweiler, R., Fleischmann, W., Cherry, J.M., Henikoff, S., Skupski, M.P., Misra, S., Ashburner, M., Birney, E., Boguski, M.S., Brody, T., Brokstein, P., Celniker, S.E., Chervitz, S.A., Coates, D., Cravchik, A., Gabrielian, A., Galle, R.F., Gelbart, W.M., George, R.A., Goldstein, L.S., Gong, F., Guan, P., Harris, N.L., Hay, B.A., Hoskins, R.A., Li, J., Li, Z., Hynes, R.O., Jones, S.J., Kuehl, P.M., Lemaitre, B., Littleton, J.T., Morrison, D.K., Mungall, C., O'Farrell, P.H., Pickeral, O.K., Shue, C., Vosshall, L.B., Zhang, J., Zhao, Q., Zheng, X.H., Lewis, S. (2000) Comparative genomics of the eukaryotes, *Science* **287**, 2204–2215.

Sargent, T.D., Dawid, I.B. (1983) Differential gene expression in the gastrula of Xenopus laevis, *Science* **222**, 135–139.

Sasaki, N., Nagaoka, S., Itoh, M., Izawa, M., Konno, H., Carninci, P., Yoshiki, A., Kusakabe, M., Moriuchi, T., Muramatsu, M., Okazaki, Y., Hayashizaki, Y. (1998) Characterization of gene expression in mouse blastocyst using single-pass sequencing of 3995 clones, *Genomics* **49**, 167–179.

Schepers, G.E., Teasdale, R.D., Koopman, P. (2002) Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families, *Dev. Cell* **3**, 167–170.

Schmitt, A.O., Specht, T., Beckmann, G., Dahl, E., Pilarsky, C.P., Hinzmann, B., Rosenthal, A. (1999) Exhaustive mining of EST libraries for genes differentially expressed in normal and tumour tissues, *Nucleic Acids Res.* **27**, 4251–4260.

Solter, D., de Vries, W.N., Evsikov, A.V., Peaston, A.E., Chen, F.H., Knowles, B.B. (2002) Fertilization and Activation of the Embryonic Genome, in: Rossant, J., Tam, P.P.L. (Eds.) *Mouse Development: Patterning, Morphogenesis, and Organogenesis*, Academic Press, San Diego, CA, pp. 5–19.

Stancheva, I., El-Maarri, O., Walter, J., Niveleau, A., Meehan, R.R. (2002) DNA methylation at promoter regions regulates the timing of gene activation in Xenopus laevis embryos, *Dev. Biol.* **243**, 155–165.

Stapleton, M., Liao, G., Brokstein, P., Hong, L., Carninci, P., Shiraki, T., Hayashizaki, Y., Champe, M., Pacleb, J., Wan, K., Yu, C., Carlson, J., George, R., Celniker, S., Rubin, G.M. (2002) The Drosophila gene collection: identification of putative full-length cDNAs for 70%

of D. melanogaster genes, *Genome Res.* **12**, 1294–1300.

Steffen, M., Petti, A., Aach, J., D'Haeseleer, P., Church, G. (2002) Automated modelling of signal transduction networks, *BMC Bioinform.* **3**, 34.

Stoeckli, M., Chaurand, P., Hallahan, D.E., Caprioli, R.M. (2001) Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues, *Nat. Med.* **7**, 493–496.

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Ketteman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E., Schnerch, A., Schein, J.E., Jones, S.J., Marra, M.A. (2002) Generation and initial analysis of more than 15 000 full-length human and mouse cDNA sequences, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16899–16903.

Svoboda, P., Stein, P., Schultz, R.M. (2001) RNAi in mouse oocytes and preimplantation embryos: effectiveness of hairpin dsRNA, *Biochem. Biophys. Res. Commun.* **287**, 1099–1104.

Tanaka, T.S., Kunath, T., Kimber, W.L., Jaradat, S.A., Stagg, C.A., Usuda, M., Yokota, T., Niwa, H., Rossant, J., Ko, M.S. (2002) Gene expression profiling of embryo-derived stem cells reveals candidate genes associated with pluripotency and lineage specificity, *Genome Res.* **12**, 1921–1928.

Tanaka, T.S., Jaradat, S.A., Lim, M.K., Kargul, G.J., Wang, X., Grahovac, M.J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R., Doi, H., Wood 3rd, W.H., Becker, K.G., Ko, M.S. (2000) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15 000 mouse developmental cDNA microarray, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9127–9132.

Tao, Q., Chang, Y.L., Wang, J., Chen, H., Islam-Faridi, M.N., Scheuring, C., Wang, B., Stelly, D.M., Zhang, H.B. (2001) Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis, *Genetics* **158**, 1711–1724.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S.E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S.E., Rubin, G.M. (2002) Systematic determination of patterns of gene expression during Drosophila embryogenesis, *Genome Biol.* **3**, RESEARCH0088.

Ton, C., Stamatiou, D., Dzau, V.J., Liew, C.C. (2002) Construction of a zebrafish cDNA microarray: gene expression profiling of the zebrafish during development, *Biochem. Biophys. Res. Commun.* **296**, 1134–1142.

Tong, A.H., Drees, B., Nardelli, G., Bader, G.D., Brannetti, B., Castagnoli, L., Evangelista, M., Ferracuti, S., Nelson, B., Paoluzi, S., Quondam, M., Zucconi, A., Hogue, C.W., Fields, S., Boone, C., Cesareni, G. (2002) A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules, *Science* **295**, 321–324.

VanBuren, V., Piao, Y., Dudekula, D.B., Qian, Y., Carter, M.G., Martin, P.R., Stagg, C.A., Bassey, U.C., Aiba, K., Hamatani, T., Kargul, G.J., Luo, A.G., Kelso, J., Hide, W., Ko, M.S. (2002) Assembly, verification, and initial annotation of the NIA mouse 7.4K cDNA clone set, *Genome Res.* **12**, 1999–2003.

Venter, J.C., Adams, M.D., Myers, E.W., et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Waterston, R.H., Lindblad-Toh, K., Birney, E., et al. (2002) Initial sequencing and comparative analysis of the mouse genome, *Nature* **420**, 520–562.

White, K.P., Rifkin, S.A., Hurban, P., Hogness, D.S. (1999) Microarray analysis of Drosophila development during metamorphosis, *Science* **286**, 2179–2184.

Wianny, F., Zernicka-Goetz, M. (2000) Specific interference with gene function by double-stranded RNA in early mouse development, *Nat. Cell Biol.* **2**, 70–75.

Woo, Y.M., Hu, D.W., Larkins, B.A., Jung, R. (2001) Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression, *Plant Cell.* **13**, 2297–2317.

Xiang, C.C., Chen, M., Kozhich, O.A., Phan, Q.N., Inman, J.M., Chen, Y., Brownstein, M.J. (2003) Probe generation directly from small numbers of cells for DNA microarray studies, *Biotechniques* **34**, 386–388, **390**, 392–393.

Xiao, J., Xu, M., Li, J., Chang Chan, H., Lin, M., Zhu, H., Zhang, W., Zhou, Z., Zhao, B., Sha, J. (2002) NYD-SP6, a novel gene potentially involved in regulating testicular development/spermatogenesis, *Biochem. Biophys. Res. Commun.* **291**, 101–110.

Xu, B.J., Caprioli, R.M., Sanders, M.E., Jensen, R.A. (2002) Direct analysis of laser capture microdissected cells by MALDI mass spectrometry, *J. Am. Soc. Mass Spectrom.* **13**, 1292–1297.

Zdobnov, E.M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R.R., Christophides, G.K., Thomasova, D., Holt, R.A., Subramanian, G.M., Mueller, H.M., Dimopoulos, G., Law, J.H., Wells, M.A., Birney, E., Charlab, R., Halpern, A.L., Kokoza, E., Kraft, C.L., Lai, Z., Lewis, S., Louis, C., Barillas-Mury, C., Nusskern, D., Rubin, G.M., Salzberg, S.L., Sutton, G.G., Topalis, P., Wides, R., Wincker, P., Yandell, M., Collins, F.H., Ribeiro, J., Gelbart, W.M., Kafatos, F.C., Bork, P. (2002) Comparative genome and proteome analysis of Anopheles gambiae and Drosophila melanogaster, *Science* **298**, 149–159.

Prions

Stanley B. Prusiner
University of California, San Francisco, CA

1 Prions 560

2 Prions are Distinct from Viruses 561

3 Disease Paradigms 562

4 Nomenclature 563

5 Discovery of the Prion Protein 566

6 Prion Protein Isoforms 567
6.1 Cell Biology of PrP^{Sc} Formation 567

7 Rodent Models of Prion Disease 568

8 PrP Gene Structure and Organization 568
8.1 Expression of the PrP Gene 569
8.2 Overexpression of Wild-type PrP Transgenes 569
8.3 PrP Gene Dosage Controls the Incubation Time 570
8.4 PrP-deficient Mice 570
8.5 Species Variations in the PrP Sequence 571
8.6 N-terminal Sequence Repeats 571
8.7 Conserved Ala-Gly Region 574

9 Structures of PrP Isoforms 574
9.1 NMR Structures of recPrP 575
9.2 Electron Crystallography of PrP^{Sc} 577

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

10 Prion Replication 579
10.1 Mechanism of Prion Propagation 580
10.2 Template-assisted Prion Formation 580
10.3 Evidence for Protein X 581
10.4 Dominant-negative Inhibition 582

11 Sporadic Human Prion Diseases 583

12 Heritable Human Prion Diseases 583
12.1 GSS and Genetic Linkage 584
12.2 fCJD Caused by Octarepeat Inserts 584
12.3 fCJD in Libyan Jews 584
12.4 Penetrance of fCJD 584
12.5 Fatal Familial Insomnia 585
12.6 Human PrP Gene Polymorphisms 585

13 Infectious Human Prion Diseases 586
13.1 Human Growth Hormone 586
13.2 Variant Creutzfeldt–Jakob Disease (vCJD) 586
13.3 Transmission of vCJD Prions by Blood Transfusion 587

14 Strains of Prions 587
14.1 Isolation of New Strains 588
14.2 Interplay between the Species and Strains of Prions 589

15 Prion Diseases of Animals 589
15.1 PrP Polymorphisms in Sheep, Cattle, and Elk 589
15.2 Bovine Spongiform Encephalopathy 590
15.3 Monitoring Cattle for BSE Prions 591
15.4 Compelling Evidence of Transmission of Bovine Prions to Humans 591
15.5 Chronic Wasting Disease 592

16 Fungal Prions 593
16.1 Yeast Prion Domains 593
16.2 Dependence of Yeast Prions on Molecular Chaperones 593
16.3 Differences between Yeast and Mammalian Prions 594

17 Prion Diseases are Disorders of Protein Conformation 594

18 Prevention and Therapeutics for Prion Diseases 595
18.1 Prion Therapeutics 595
18.1.1 Quinacrine and other Acridine Derivatives 595
18.1.2 Anti-PrP Antibodies 596
18.2 Dominant-negative Sheep 596

19 Some Principles of Prion Biology 597
19.1 Implications for Common Neurodegenerative Diseases 597

Bibliography 598
Books and Reviews 598
Primary Literature 600

Keywords

Prion
Proteinaceous infectious particle.

PrP^{C}
Normal isoform of the prion protein.

PrP^{Sc}
Disease-causing isoform of the prion protein.

Prions are infectious proteins that have been described in both mammals and fungi. That prions are composed solely of proteins makes them unprecedented infectious pathogens. Prions are proteins that can adopt at least two different conformations; they multiply by forcing the precursor protein to acquire a second conformation. Different conformations of proteins in the prion state encipher distinct strains and are prone to aggregation. In mammals, prions accumulate to high levels in the nervous system where they cause dysfunction and fatal degeneration.

Both mammalian and fungal prions have been produced in cell-free systems. Synthetic prion protein (PrP) peptides and recombinant PrP fragments have been used to form mammalian prions while N-terminal regions, called *prion domains* that are rich in glutamine and asparagine, have been used to form fungal prions.

Prions cause a group of invariably fatal, neurodegenerative diseases. Prion diseases may present as genetic, infectious, or sporadic disorders, all of which involve modification of PrP. The tertiary structure of PrP is profoundly altered as prions are formed, and as such, prion diseases represent disorders of protein conformation. Creutzfeldt-Jakob disease (CJD) generally presents as a progressive dementia in humans while scrapie of sheep and bovine spongiform encephalopathy (BSE) usually manifest as ataxic illnesses.

1
Prions

In mammals, prions reproduce by recruiting the normal, cellular isoform of the prion protein (PrP^C) and stimulating its conversion into the disease-causing isoform (PrP^{Sc}). PrP^C has a high α-helical content and little β-sheet structure, whereas PrP^{Sc} has less α-helical structure and a high β-sheet content. Comparisons of secondary structures of PrP^C and PrP^{Sc} were performed on proteins purified from Syrian hamster (SHa) brains. Limited proteolysis of PrP^{Sc} produces PrP 27–30, which retains prion infectivity; under these conditions, PrP^C is completely hydrolyzed (Fig. 1).

Solution structures of recombinant SHa and mouse (Mo) PrPs produced in bacteria showed three α-helices denoted A, B, and C as well as two short β-strands using nuclear magnetic resonance (NMR) imaging. The atomic structure of PrP^{Sc} has not been determined because of the insolubility of the protein. Two-dimensional crystals of PrP 27–30 have been used to constrain computational models of PrP^{Sc} suggesting it contains α-helix C and a portion of α-helix B while α-helix A, the two β-strands, and the surrounding segments form a β-helix. From these studies as well as from ionizing radiation inactivation analyses, it is likely that the infectious monomer is a trimer of PrP^{Sc} molecules. Recombinant

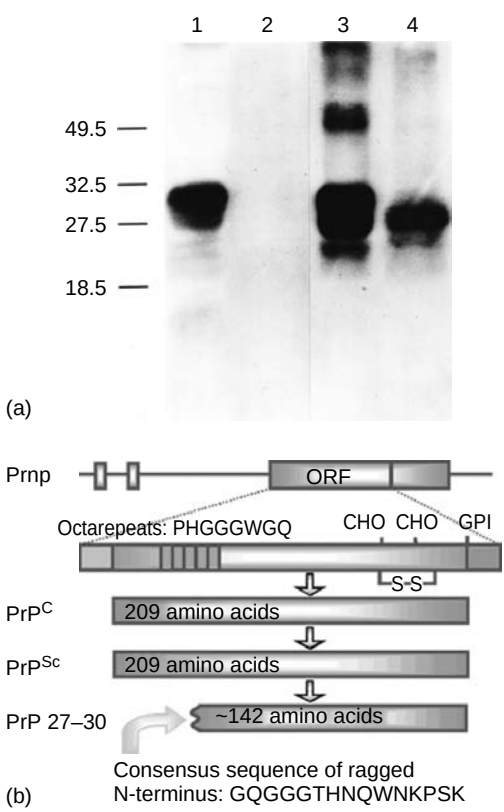

Fig. 1 Prion protein isoforms. Western immunoblot of brain homogenates from uninfected (lanes 1 and 2) and prion-infected (lanes 3 and 4) Syrian hamsters. Samples in lanes 2 and 4 were digested with 50 µg mL^{-1} of proteinase K for 30 min at 37 °C. PrP^C in lanes 2 and 4 was completely hydrolyzed under these conditions, whereas approximately 67 amino acids were digested from the N-terminus of PrP^{Sc} to generate PrP 27–30. After polyacrylamide gel electrophoresis (PAGE) and electrotransfer, the blot was developed with anti-SHaPrP R073 polyclonal rabbit antiserum. Molecular weight markers are depicted in kilodaltons (kDa). (b) Bar diagram of the SHaPrP gene that encodes a protein of 254 amino acids. After processing of the N- and C-termini, both PrP^C (green) and PrP^{Sc} (red) consist of 209 residues. After limited proteolysis, the N-terminus of PrP^{Sc} is truncated to form PrP 27–30, which is composed of approximately 142 amino acids, the N-terminal sequence of which was determined by Edman degradation. (See color plate p. xxvii).

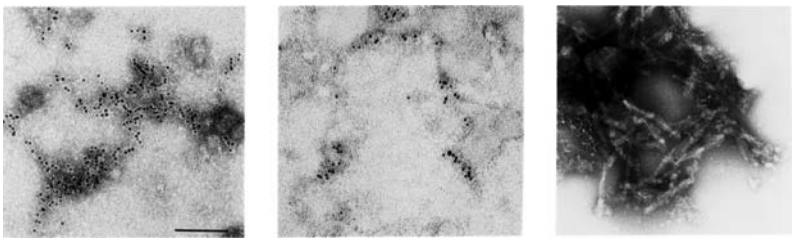

(a) (b) (c)

Fig. 2 Electron micrographs of negatively stained and immunogold-labeled prion proteins. (a) PrP^{C} and (b) PrP^{Sc}. Neither PrP^{C} nor PrP^{Sc} forms recognizable, ordered polymers. (c) Prion rods composed of PrP 27–30 were negatively stained. The prion rods are indistinguishable from many purified amyloids. Bar = 100 nm. Reprinted with permission, from (Pan, K-M., Baldwin, M., Nguyen, J., Gasset, M., Serban, A., Groth, D., Mehlhorn, I., Huang, Z., Fletterick, R.J., Cohen, F.E., Prusiner, S.B. (1993) Conversion of α-helices into β-sheets features in the formation of the scrapie prion proteins, *Proc. Natl. Acad. Sci. U. S. A.* **90**, 10962–10966), copyright 1993 National Academy of Sciences, USA.

antibody fragments have been used to map the surfaces of PrP^{C} and PrP^{Sc} and those results are consistent with the foregoing structural data. The structural transition from an α-helical protein into a β sheet–rich molecule seems to be the fundamental event underlying the formation of nascent prions.

Limited proteolysis truncates the N-terminus of PrP^{Sc} to produce PrP 27–30 consisting of the C-terminal ~142 amino acids. PrP 27–30 polymerizes into amyloid fibrils that are indistinguishable from fibrils found in amyloid plaques of the brains of mammals with prion disease. When full-length PrP^{Sc} was purified from SHa brain, only amorphous aggregates were seen by electron microscopy. While limited digestion of purified PrP^{Sc} with proteinase K produced PrP 27–30, the ultrastructural appearance of the samples remained unchanged (Fig. 2). Addition of a detergent, such as Sarkosyl, either before or after limited proteolysis provoked the assembly of PrP 27–30 into rod-shaped structures with the ultrastructural and tinctorial properties of amyloid.

2
Prions are Distinct from Viruses

Major features that distinguish prions from viruses are: (1) the ability to create prion infectivity by modifying the conformation of a polypeptide devoid of nucleic acid and (2) PrP is encoded by a chromosomal gene, designated *PRNP* in humans and *Prnp* in mice. The PrP gene is located on the short arm of chromosome 20 in humans and the syntenic region of chromosome 2 in mice.

Prions differ from viruses and viroids in that they lack a nucleic acid genome that directs the synthesis of their progeny. Many investigators argued for a nucleic acid genome within the infectious prion particle while others contended for a small, noncoding polynucleotide of either foreign or cellular origin. No nucleic acid has been found despite intensive searches using a wide variety of techniques and approaches. On the basis of a wealth of evidence, it is reasonable to assert that such a nucleic acid has not been found because it does not exist. Prions are composed of an

alternative isoform of a cellular protein, whereas most viral proteins are encoded by a viral genome and viroids are devoid of protein.

In contrast to viruses, prions are non-immunogenic. Prions do not elicit an immune response because the host is tolerant to PrP^{C}, which prevents the host from mounting an immune response to PrP^{Sc}. In contrast, foreign proteins of viruses that are encoded by the viral genome often elicit a profound immune response. Thus, it seems unlikely that vaccination, which has been so effective in preventing many viral illnesses, will be a useful strategy for preventing or treating prion diseases.

When prions and viruses are passaged from one host species to another, the consequences are very different. The passage of prions from one host to another is accompanied by the acquisition of a new PrP amino acid sequence encoded by the genome of the new host. The crossing of prions from one species to another is restricted by what has been called the ''species barrier''. The closer the evolutionary relationship is between the host in which the prions last replicated and the newly infected animal, the more likely replication will occur in the new host. However, particular strains of prions can change this rule. For example, variant CJD prions from humans replicate much more readily in transgenic (Tg) mice expressing bovine (Bo) PrP than in Tg mice expressing human or chimeric mouse-human PrP.

The issue of prion strains posed a profound conundrum for many years. How could an infectious pathogen composed only of protein encipher biological information? This riddle was solved when prion strains with different physical properties were isolated. Subsequently, new strains of prions have been isolated either by passage through mice expressing artificial PrP transgenes or by forming synthetic prions from PrP produced in *Escherichia coli*.

3 Disease Paradigms

Despite some similarities between prion and viral illnesses, these disorders are very different. Prion diseases are uniformly fatal. No human or animal has ever recovered from a prion disease once neurologic dysfunction is manifest. No host defenses are mounted in response to prion infection: no humoral immunity, no cellular immunity, and no interferons are elicited to the replicating prion.

While prions can be spread from one host to another, the most common form of prion disease is spontaneous or sporadic. In these illnesses, prions arise endogenously and replicate. Recent experiments with synthetic prions demonstrate that only PrP^{C} is required to generate prion infectivity. These findings contrast with viruses in which exogenous infection is required except in the case of latent retroviral genomes. For example, after infection with exogenous HIV, the virus may disappear but its RNA genome may be reverse-transcribed into DNA and the DNA copies may remain dormant for years.

The dramatically different principles that govern prion biology have often been poorly understood. This lack of understanding has led to some regrettable decisions of great economic, political, and possibly public health importance. For example, scrapie and BSE have different names, yet they are the same disease in two different species. Scrapie and BSE differ in only two respects: first, the PrP sequence in sheep differs from that of cattle at seven

or eight positions of 270 amino acids, which results in different PrP^{Sc} molecules. Second, some aspects of each disease are determined by the particular prion strain that infects the respective host.

Understanding prion strains and the "species barrier" is of paramount importance with respect to the BSE epidemic in Britain, in which more than 180 000 cattle have died over the past decade. Brain extracts from eight cattle with BSE resulted in similar incubation times and patterns of vacuolation in the neuropil when inoculated into a variety of inbred mice. Incubation times and profiles of neuronal vacuolation have been used for three decades to study prion strains. Brain extracts prepared from three domestic cats, one nyala, and one kudu, all of which died of a neurologic illness, produced incubation times and lesion profiles indistinguishable from those found in the BSE cattle. Cats and exotic ungulates (such as the kudu) presumably developed prion disease from eating food containing bovine prions.

4 Nomenclature

Although the prions that cause transmissible mink encephalopathy (TME) and BSE are referred to as *TME prions* and *BSE prions*, this may be unjustified because both are thought to originate from the oral consumption of scrapie prions in sheep-derived foodstuffs and because many lines of evidence argue that the only difference among various prions is the sequence of PrP, which is dictated by the host and not the prion itself. Human (Hu) prions present a similar semantic conundrum. Transmission of Hu prions to laboratory animals produces prions carrying PrP molecules with sequences dictated by the PrP gene of the last host, not that of the inoculum.

To simplify the terminology, the generic term PrP^{Sc} was suggested in place of such terms as PrP^{CJD}, PrP^{BSE}, and PrP^{res}. To distinguish PrP^{Sc} found in humans and cattle from that found in other animals, "$HuPrP^{Sc}$" and "$BoPrP^{Sc}$" are suggested instead of PrP^{CJD} and PrP^{BSE}, respectively. Once human prions, or $HuPrP^{Sc}$ molecules, have been passaged into animals, the prions and PrP^{Sc} are no longer of the human species unless they were formed in an animal expressing a HuPrP transgene.

The "Sc" superscript of PrP^{Sc} was initially derived from the term "scrapie" because scrapie was the prototypic prion disease. Because all of the known prion diseases (Table 1) of mammals involve aberrant metabolism of PrP similar to that observed in scrapie, the "Sc" superscript was suggested for all abnormal, pathogenic PrP isoforms. In this context, the "Sc" superscript is used to designate the disease-causing isoform of PrP (Table 2). The development of the conformation-dependent immunoassay (CDI) for PrP^{Sc} led to the discovery of a protease-sensitive form of PrP^{Sc}, designated $sPrP^{Sc}$. The CDI is based on measuring antibody binding to an epitope that is exposed in PrP^{C} but buried in PrP^{Sc}. Under conditions of limited proteolysis in which the protease-resistant form of PrP^{Sc}, designated $rPrP^{Sc}$, is converted into PrP 27–30, $sPrP^{Sc}$ is completely hydrolyzed. Whether $sPrP^{Sc}$ is an intermediate in the formation of infectious prions remains to be established.

In the case of mutant PrP, mutations and polymorphisms can be denoted in parentheses following the PrP isoform.

Tab. 1 The prion diseases.

Disease	*Host*	*Mechanism of pathogenesis*
Kuru	Fore people	Infection through ritualistic cannibalism
Iatrogenic CJD	Humans	Infection from prion-contaminated HGH, dura mater grafts, etc.
Variant CJD	Humans	Infection from bovine prions?
Familial CJD	Humans	Germline mutations in PrP gene
GSS	Humans	Germline mutations in PrP gene
FFI	Humans	Germline mutations in PrP gene (D178N,M129)
Sporadic CJD	Humans	Somatic mutation or spontaneous conversion of PrP^{C} into PrP^{Sc}?
sFI		Somatic mutation or spontaneous conversion of PrP^{C} into PrP^{Sc}?
Scrapie	Sheep	Infection in genetically susceptible sheep
Bovine spongiform encephalopathy	Cattle	Infection with prion-contaminated MBM
Transmissible mink encephalopathy	Mink	Infection with prions from sheep or cattle
Chronic wasting disease	Mule deer, elk	Unknown
Feline spongiform encephalopathy	Cats	Infection with prion-contaminated bovine tissues or MBM
Exotic ungulate encephalopathy	Greater kudu, nyala, oryx	Infection with prion-contaminated MBM

Notes: CJD: Creutzfeldt–Jakob disease; FFI: fatal familial insomnia; sFI: sporadic fatal insomnia; GSS: Gerstmann–Sträussler–Scheinker disease; HGH: human growth hormone; MBM: meat and bone meal.

For fatal familial insomnia (FFI), in which it might be important to identify the mutation, the prions would be designated $HuPrP^{Sc}$(D178N,M129) (Table 3).

Parentheses following PrP^{Sc} can also be used to notate a particular prion strain. For example, prions from Syrian hamsters inoculated with Sc237 or 139H prion strains can be designated $SHaPrP^{Sc}$(Sc237) or $SHaPrP^{Sc}$(139H), respectively. Similarly, sheep inoculated with scrapie or BSE prions produce $OvPrP^{Sc}$(Sc) or $OvPrP^{Sc}$(BSE), respectively.

The terms "PrP^{res}" and "PrP-res" were derived to describe the protease resistance of PrP^{Sc} and have sometimes been used interchangeably with PrP^{Sc}. The use of "PrP^{res}" and "PrP-res" became particularly problematic with the discovery of $sPrP^{Sc}$. Protease resistance, insolubility, and high β-sheet content should be considered only as surrogate markers of PrP^{Sc} infectivity because not all characteristics may be present. For example, $MoPrP^{Sc}$(P101L) from Tg mice that express high levels of MoPrP(P101L) is transmissible to Tg(MoPrP,P101L)196/$Prnp^{0/0}$ mice, but these prions are sensitive to proteolytic digestion at 37 °C. When digestions were performed at 4 °C or the CDI was used to detect PrP^{Sc}, then $MoPrP^{Sc}$(P101L) was detected. In contrast, when Tg196 mice were inoculated with mouse RML prions, the resulting $MoPrP^{Sc}$(P101L) was resistant to proteolytic digestion at 37 °C. These

Tab. 2 Glossary of prion terminology.

Term	*Description*
Prion	A proteinaceous infectious particle that lacks nucleic acid. Prions are composed largely, if not entirely, of PrP^{Sc} molecules.
PrP^{Sc}	Abnormal, pathogenic isoform of the prion protein that causes illness. This protein is the only identifiable macromolecule in purified preparations of prions.
PrP^{C}	Cellular isoform of the prion protein.
PrP 27–30	N-terminally truncated PrP^{Sc}, generated by digestion with proteinase K.
PRNP	Human PrP gene located on chromosome 20.
Prnp	Mouse PrP gene located on syntenic chromosome 2. *Prnp* controls the length of the prion incubation time and is congruent with the incubation-time genes *Sinc* and *Prn-i*. PrP-deficient ($Prnp^{0/0}$) mice are resistant to prion infection.
PrP amyloid	Fibril of PrP fragments derived from PrP^{Sc} by proteolysis. Plaques containing PrP amyloid are found in the brains of some mammals with prion disease.
Prion rod	An amyloid polymer composed of PrP 27–30 molecules. Created by detergent extraction and limited proteolysis of PrP^{Sc}.
Protein X	A hypothetical macromolecule that is thought to act as a molecular chaperone in facilitating the conversion of PrP^{C} into PrP^{Sc}.

Tab. 3 Examples of human PrP gene mutations found in the inherited prion diseases.

Inherited prion disease	*PrP gene mutation*
Gerstmann–Sträussler–Scheinker disease	P102L[a]
Gerstmann–Sträussler–Scheinker disease	A117V
Familial Creutzfeldt–Jakob disease	D178N, V129
Fatal familial insomnia	D178N, M129[a]
Gerstmann-Sträussler-Scheinker disease	F198S[a]
Familial Creutzfeldt–Jakob disease	E200K[a]
Gerstmann–Sträussler–Scheinker disease	Q217R
Familial Creutzfeldt–Jakob disease	octarepeat insert[a]

[a]Signifies genetic linkage between the mutation and the inherited prion disease.

findings emphasize the ambiguities that may arise from simply assessing resistance to proteolytic digestion. Whether PrP^{res} is useful in denoting PrP molecules that have been subjected to procedures modifying resistance to proteolysis but that neither convey infectivity nor cause disease remains questionable. Perhaps PrP^{res} is best reserved for PrP molecules that exhibit resistance to limited proteolysis after binding to PrP^{Sc}.

The term "PrP*" has been used in two different ways. First, it has been used to identify a fraction of PrP^{Sc} molecules that are infectious. Such a designation is thought to be useful because

there are $\sim 10^5$ PrP^{Sc} molecules per infectious (ID_{50}) unit. Second, PrP* has been used to designate a metastable intermediate of PrP^C that is bound to a putative conversion cofactor, provisionally designated protein X. It is noteworthy that neither a subset of biologically active PrP^{Sc} molecules nor a metastable intermediate of PrP^C has been identified, to date.

In mice, *Prnp* is now known to be identical to two genes, *Sinc* and *Prn-i,* that are known to control the length of the incubation time in mice inoculated with prions. A gene, designated *Pid-1,* on mouse chromosome 17 also appears to influence experimental CJD and scrapie incubation times but information on this locus is limited.

Distinguishing among CJD, FFI, and Gerstmann–Sträussler–Scheinker syndrome (GSS) has grown increasingly difficult with the recognition that familial (f) CJD, GSS, and FFI are autosomal-dominant diseases caused by mutations in *PRNP* (Table 3). Initially, it was thought that each *PRNP* mutation was associated with a particular cliniconeuropathologic phenotype, but more exceptions are being recognized. Multiple examples of variations in the cliniconeuropathologic phenotype have been recorded within a single family in which all affected members carry the same *PRNP* mutation. Most patients with a *PRNP* mutation at codon 102 present with ataxia and have PrP amyloid plaques; such patients are generally given the diagnosis of GSS, but some individuals present with dementia, a clinical characteristic that is usually associated with CJD. For most inherited prion disease, the disease is specified by the respective mutation, such as fCJD(E200K) and GSS(P101L). In the case of FFI, describing the D178N mutation and M129 polymorphism seems unnecessary because this is the only known mutation–polymorphism combination that results in the FFI phenotype. The sporadic form of fatal insomnia is denoted sFI.

5 Discovery of the Prion Protein

The discovery of PrP 27–30 in SHa brain fractions progressively enriched for scrapie infectivity transformed research on scrapie and related diseases. PrP 27–30 was so named because it has an apparent molecular weight (M_r) of 27 kDa to 30 kDa (Fig. 1). PrP 27–30 is derived from the larger PrP^{Sc} by N-terminal truncation; both PrP 27–30 and PrP^{Sc} are infectious. PrP 27–30 not only provided a molecular marker that is specific for prion disorders but it was later shown to be the sole component of the prion particle.

The molecular biology and genetics of prions began with the purification of PrP 27–30 that allowed determination of its N-terminal, amino acid sequence. Multiple signals in each cycle of the Edman degradation suggested that either multiple proteins were present in these "purified fractions" or a single protein with a ragged N-terminus was present. When the signals in each cycle were grouped according to their strong, intermediate, and weak intensities, it became clear that a single protein with a ragged N-terminus was being sequenced (Fig. 3). Determination of a single, unique sequence for the N-terminus of PrP 27–30 permitted the synthesis of isocoding mixtures of oligonucleotides that were subsequently used to identify incomplete PrP cDNA clones from hamster and mouse. cDNA clones encoding the entire open reading

Relative amount	Amino acid sequence*
1	G-Q-G-G-G-T-H-N-Q-W-N-K-P-S-K
0.4	X-X-X-T-H-N-X-W-X-K-P
0.2	X-X-P-W-X-Q-X-X-X-T-H-X-Q-W

*Single-letter amino acid code. X = amino acid not determined at that cycle.

Fig. 3 Interpreted sequence, shown by single-letter amino acid codes, of the N-terminus of PrP 27–30. The "ragged ends" of PrP 27–30 are shown. X = amino acid was not detected at that cycle of the Edman degradation.

frames (ORFs) of SHaPrP and MoPrP were eventually recovered.

6 Prion Protein Isoforms

In Syrian hamsters, PrP^C and PrP^{Sc} are 209-residue proteins that are anchored to the cell surface by a glycosylphosphatidyl inositol (GPI) moiety and have the same covalent structure (Fig. 1). The N-terminal sequencing, the deduced amino acid sequences from PrP cDNA, and immunoblotting studies argue that PrP 27–30 is a truncated protein of approximately 142 residues, which is derived from PrP^{Sc} by limited proteolysis of the N-terminus.

In general, $\sim 10^5$ PrP^{Sc} molecules correspond to one ID_{50} unit (U) using the most sensitive bioassay. PrP^{Sc} is probably best defined as the alternative or abnormal isoform of PrP, which stimulates conversion of PrP^C into nascent PrP^{Sc}, accumulates, and causes disease. Although resistance to limited proteolysis has proved to be a convenient tool for detecting PrP^{Sc}, not all PrP^{Sc} molecules possess protease resistance, as discussed above. Some investigators equate protease resistance with PrP^{Sc} and this erroneous view has been compounded by the use of the term "PrP-res".

Although insolubility and protease resistance were used in initial studies to differentiate PrP^{Sc} from PrP^C, subsequent investigations showed that these properties are only surrogate markers of infectivity, as are high β-sheet content and polymerization into amyloid. When these surrogate markers are present, they are useful, but their absence does not establish a lack of prion infectivity. PrP^{Sc} is usually not detected by Western immunoblotting if fewer than 10^5 ID_{50} U mL^{-1} are present in a sample. Furthermore, PrP^{Sc} from different species or prion strains may exhibit different degrees of protease resistance.

6.1 Cell Biology of PrP^{Sc} Formation

In scrapie-infected cells, PrP^C molecules destined to become PrP^{Sc} exit to the cell surface prior to conversion into PrP^{Sc}. Like other GPI-anchored proteins, PrP^C appears to reenter the cell through a subcellular compartment bounded by cholesterol-rich, detergent-insoluble membranes, which might be caveolae or early endosomes. Within this cholesterol-rich, nonacidic compartment, GPI-anchored PrP^C can be either converted into PrP^{Sc} or partially degraded. Subsequently, PrP^{Sc} is trimmed at the N-terminus in an acidic compartment in scrapie-infected cultured cells to form PrP 27–30. In contrast, N-terminal truncation of PrP^{Sc} is minimal in the brain, where little PrP 27–30 is found.

7
Rodent Models of Prion Disease

Mice and hamsters are commonly used in experimental studies of prion disease. The shortest incubation times are achieved with intracerebral inoculation of prions with a sequence identical to that of the host animal; under these conditions, all animals develop prion disease within a narrow interval for a particular dose. When the PrP sequence of the donor prion differs from that of the recipient host, the incubation time is prolonged, and can be quite variable; often, many of the inoculated animals do not develop disease. This phenomenon is generally called the "species barrier".

8
PrP Gene Structure and Organization

Prnp is a member of the *Prn* gene family. The second member of this family to be identified is the *Prnd* gene that lies approximately 19 kb downstream from the PrP locus and encodes the doppel (Dpl) protein. The respective genes that encode PrP and Dpl appear to represent ancient gene duplication that occurred an prior to the speciation of mammals. The sequences are approximately 25% identical but the structures of the two proteins are highly conserved (Fig. 4). In contrast to PrP, which is expressed in many different tissues, Dpl expression is confined to the testis. Both Dpl and PrP are found on the surface of sperm but their functions are unknown. In contrast to PrP-deficient ($Prnp^{0/0}$) mice, Dpl-deficient ($Prnd^{0/0}$) mice are sterile. The knockout of both PrP and Dpl genes resulted in a sterile phenotype.

The entire ORF of all known mammalian and avian PrP genes resides within a single exon, which eliminates the possibility that PrP^{Sc} arises from alternative RNA splicing. The two exons of the SHaPrP gene are separated by a 10-kb intron; exon 1 encodes a portion of the 5′ untranslated leader sequence. The PrP

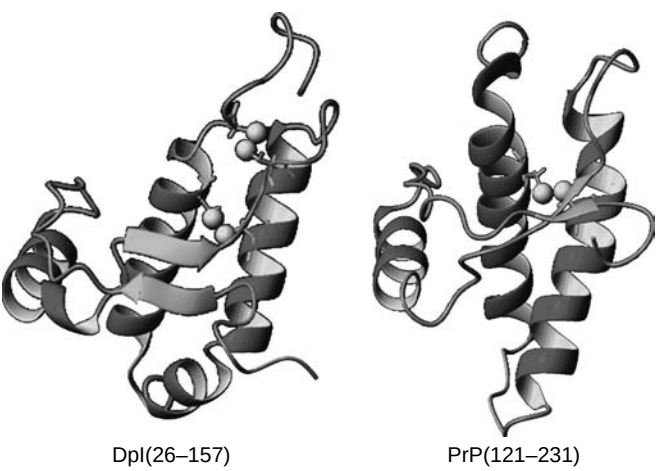

Fig. 4 Comparison of the NMR structures of mouse Dpl and PrP. Backbone topology of mouse Dpl(26–157) and mouse PrP(121–231). Figure prepared by Jane Dyson and Peter Wright.

genes of Syrian hamster, mouse, sheep, and rat contain three exons, with exon 3 encoding the ORF and 3′ untranslated region. The promoters of both the SHaPrP and MoPrP genes contain multiple copies of G-C–rich repeats and are devoid of TATA boxes. These G-C nonamers represent a motif that may function as a canonical binding site for the transcription factor Sp1.

Like the PrP gene, the entire ORF of the Dpl gene is encoded within a single exon. In some lines of *Prnp*$^{0/0}$ mice, high levels of Dpl expression were found in the brain. The expression of Dpl in the CNS was due to intergenic splicing of the nontranslated exons of the PrP gene with the translated exon of Dpl. That Dpl expression is neurotoxic was subsequently demonstrated by construction of Tg mice expressing Dpl in the brain.

8.1 Expression of the PrP Gene

Although PrP mRNA is constitutively expressed in the brains of adult animals, it is highly regulated during development. In the septum, levels of PrP mRNA and choline acetyltransferase were found to increase in parallel during development. In other brain regions, PrP gene expression occurs at an earlier age. *In situ* hybridization studies show that the highest levels of PrP mRNA are found in neurons.

Because no antibodies are currently available that clearly distinguish between PrP^{C} and PrP^{Sc}, PrP^{C} is generally measured in tissues from uninfected control animals, in which no PrP^{Sc} is found. PrP^{Sc} staining was minimal in the regions that intensely stained for PrP^{C}. A similar relationship between PrP^{C} and PrP^{Sc} was found in the amygdala. In contrast, PrP^{Sc} accumulated in the medial habenular nucleus, the medial septal nuclei, and the diagonal band of Broca; these areas were virtually devoid of PrP^{C}. In the white matter, bundles of myelinated axons contained PrP^{Sc} but were devoid of PrP^{C}. These findings suggest that prions are transported along axons, which is consistent with earlier findings in which scrapie infectivity migrated in a pattern consistent with retrograde transport. While the rate of PrP^{Sc} synthesis appears to be a function of the level of PrP^{C} expression in Tg mice, the level of PrP^{Sc} accumulation appears to be independent of PrP^{C} concentration.

8.2 Overexpression of Wild-type PrP Transgenes

Mice expressing different levels of the wild-type (wt) SHaPrP transgene were constructed. Inoculation of these Tg(SHaPrP) mice with SHa prions demonstrated abrogation of the species barrier, resulting in abbreviated incubation times due to a nonstochastic process. The length of the incubation time after inoculation with SHa prions was inversely proportional to the level of $SHaPrP^{C}$ in the brains of Tg(SHaPrP) mice. Bioassays of brain extracts from clinically ill Tg(SHaPrP) mice inoculated with Mo prions revealed that Mo prions, but no SHa prions, were produced. Conversely, inoculation of Tg(SHaPrP) mice with SHa prions led to the synthesis of only SHa prions.

During transgenetic studies, uninoculated older mice harboring high copy numbers of wt PrP transgenes derived from Syrian hamsters, sheep, and *Prnp*^{b} mice spontaneously developed truncal ataxia, hind-limb paralysis, and tremors. These Tg mice exhibited profound necrotizing

myopathy involving skeletal muscle, demyelinating polyneuropathy, and focal vacuolation of the CNS. Development of disease was dependent on transgene dosage. For example, Tg(SHaPrP$^{+/+}$)7 mice homozygous for the SHaPrP transgene array regularly developed disease between 450 and 600 days of age, while hemizygous Tg(SHaPrP$^{+/0}$)7 mice developed disease after >650 days. Attempts to demonstrate spontaneous generation of SHa prions in Tg(SHaPrP$^{+/+}$)7 mice have been unsuccessful.

8.3 PrP Gene Dosage Controls the Incubation Time

Incubation times have been used to isolate prion strains inoculated into sheep, goats, mice, and hamsters. The *Sinc* gene is a major determinant of incubation periods in mice. Once molecular clones of *Prnp* became available, a study was performed to determine if control of the length of the incubation time is genetically linked to the PrP gene. Because the availability of VM mice with prolonged incubation times that were used to define *Sinc* was restricted, I/LnJ mice were used in the crosses. Indeed, the incubation-time locus, designated *Prn-i*, was found to be either congruent with or closely linked to *Prnp*.

Although the amino acid substitutions in PrP that distinguish *Prnp*^{a} from *Prnp*^{b} mice argued for the congruency of *Prnp* and *Prn-i*, experiments with *Prnp*^{a} mice expressing *Prnp*^{b} transgenes demonstrated a "paradoxical" shortening of incubation times. These Tg mice were predicted to exhibit a prolongation of the incubation time after inoculation with RML prions based on (*Prnp*^{a} × *Prnp*^{b}) F1 mice, which exhibit long incubation times. Those findings were described as a "paradoxical" shortening because we and others had believed for many years that long incubation times are dominant traits. From studies of congenic and transgenic mice expressing different numbers of the *a* and *b* alleles of *Prnp*, these findings were discovered not to be paradoxical; indeed, they resulted from increased PrP gene dosage. When the RML isolate was inoculated into congenic and transgenic mice, increasing the number of copies of the *a* allele was found to be the major determinant in reducing the incubation time; however, increasing the number of copies of the *b* allele also reduced the incubation time, but not to the same extent as that seen with the *a* allele. Gene-targeting studies established that the *Prnp* gene controls the incubation time and as such, is congruent with both *Prn-i* and *Sinc*.

8.4 PrP-deficient Mice

Ablation of the PrP gene in mice did not affect development of these animals. *Prnp*$^{0/0}$ mice remain healthy for more than two years. Acute suppression of PrP expression by addition of doxycycline to the drinking water of bigenic mice with an inducible PrP transgene under the control of the tetracycline promoter did not result in any untoward effects in adult mice. Similarly, bigenic mice, in which neuronal PrP expression was terminated in adulthood, showed no discernable deficits.

In two *Prnp*$^{0/0}$ lines, Purkinje cell loss was accompanied by ataxia beginning at approximately 70 weeks of age. Crossing one of these *Prnp*$^{0/0}$ lines with Tg mice overexpressing MoPrP rescued the ataxic phenotype. With the discovery of Dpl, it became clear, as described above, that Dpl

expression in the brains of these *Prnp*$^{0/0}$ mice provoked cerebellar degeneration.

Prnp$^{0/0}$ mice inoculated with prions are resistant to infection. *Prnp*$^{0/0}$ mice were sacrificed 5 days, 60 days, 120 days, and 315 days after inoculation with RML prions and brain extracts were bioassayed in Swiss CD-1 mice. Except for residual infectivity from the inoculum detected at 5 days after inoculation, no infectivity was found in the brain extracts, as the Swiss CD-1 mice did not develop disease.

Prnp$^{0/0}$ mice crossed with Tg(SHaPrP) mice were rendered susceptible to SHa prions but remained resistant to Mo prions. Because the absence of PrP^{C} expression does not provoke disease, it is likely that scrapie and other prion diseases are a consequence of PrP^{Sc} accumulation rather than inhibition of PrP^{C} function. Such an interpretation is consistent with the dominant inheritance of familial prion diseases.

Mice heterozygous (*Prnp*$^{0/+}$) for ablation of the PrP gene had prolonged incubation times when inoculated with Mo prions and developed signs of neurologic dysfunction at 400 to 460 days after inoculation. These findings are in accord with studies on Tg(SHaPrP) mice in which increased SHaPrP expression was accompanied with shortened incubation times.

Because *Prnp*$^{0/0}$ mice do not express PrP^{C}, we reasoned that they might more readily produce α-PrP antibodies. *Prnp*$^{0/0}$ mice immunized with Mo or SHa prion rods produced α-PrP antisera that bound MoPrP, SHaPrP, and HuPrP. These findings contrast with earlier studies in which α-MoPrP antibodies could not be produced in mice presumably because the mice had been rendered tolerant by the presence of MoPrP^{C}. That *Prnp*$^{0/0}$ mice readily produce α-PrP antibodies is consistent with the hypothesis that the lack of an immune response in prion diseases is due to the fact that PrP^{C} and PrP^{Sc} share many epitopes.

8.5 Species Variations in the PrP Sequence

PrP is posttranslationally processed to remove a 22-amino acid, N-terminal signal peptide. The C-terminal 120 amino acids contain two conserved disulfide-bonded cysteines and a sequence that beckons addition of a GPI anchor. Twenty-three residues are removed during the addition of this GPI moiety, which anchors the protein to the cell membrane. Contributing to the mass of the protein are two Asn side chains linked to large oligosaccharides with multiple structures that have been shown to be complex and diverse. Although many species variants of PrP have now been sequenced, only the chicken sequence has been found to differ greatly from the human sequence. The alignment of the translated sequences from more than 40 PrP genes shows a striking degree of conservation between the mammalian sequences and is suggestive of the retention of some important function through evolution (Fig. 5). Cross-species conservation of PrP sequences makes it difficult to draw conclusions about the functional importance of many of the individual residues in the protein.

8.6 N-terminal Sequence Repeats

The N-terminal domain of mammalian PrP contains five copies of a P(H/Q)GGG (G)WGQ octarepeat sequence, occasionally more, as in the case of one sequenced bovine allele, which has six copies. These repeats are remarkably conserved between species, which implies a functionally

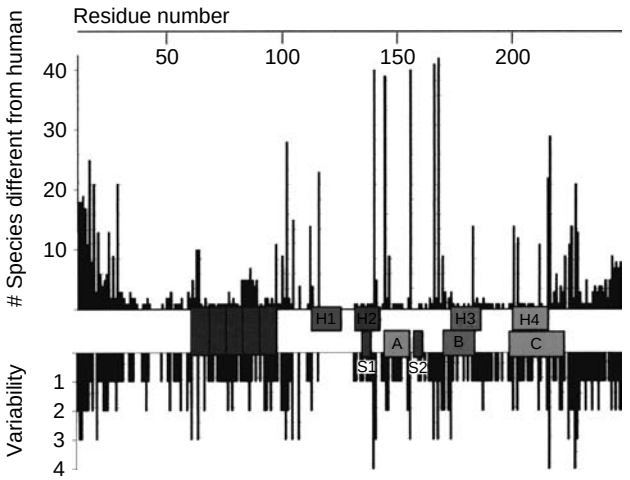

Fig. 5 Species variations and mutations of the prion protein gene. The *x*-axis represents the human PrP sequence, with the five octarepeats and H1–H4 regions of putative secondary structure shown as well as the three α-helices A, B, and C and the two β-strands S1 and S2 as determined by NMR. Vertical bars above the axis indicate the number of species that differ from the human sequence at each position. Below the axis, the length of the bars indicates the number of alternative amino acids at each position in the alignment. Data were compiled by Paul Bamborough and Fred E. Cohen (see color plate p. xxvi).

important role. The chicken sequence contains a different repeat, PGYP(H/Q)N. Although insertions of extra repeats have been found in patients with familial prion disease, naturally occurring deletions of single octarepeats do not appear to cause disease, and deletion of all these repeats does not prevent PrP^C from undergoing a conformational transition into PrP^{Sc}.

It was suggested that the histidine residues in the octarepeats might bind metal ions and that immobilized metal ion affinity chromatography (IMAC) might facilitate the purification of PrP; indeed, IMAC did prove to be useful in the purification of PrP^C. The availability of purified PrP^C allowed studies comparing the secondary structures of PrP^C and PrP^{Sc}. The metal ion-peptide complexes were found to be much more soluble than the metal ions bound to full-length PrP. Using full-length recombinant (rec) PrP, Cu^{2+} was found to bind with a much higher avidity than any other metal ion, but the concentration for half-maximal binding for Cu^{2+} was 14 μM at pH 6.0, indicating a rather low affinity of PrP for Cu^{2+}. At neutral pH, Cu^{2+}–PrP complexes tended to form large aggregates and precipitate, so, synthetic peptides containing the octarepeats were used to study more extensively the interaction of Cu^{2+} with PrP. The concentration for half-maximal binding for Cu^{2+} to a peptide (51–75) containing two octarepeats was 6 μM at pH 7.5. When a peptide (58–91) containing four octarepeats was studied, the binding of Cu^{2+} was found to be cooperative and highly pH-dependent (Fig. 6). The midpoint of the pH-dependence transition was pH

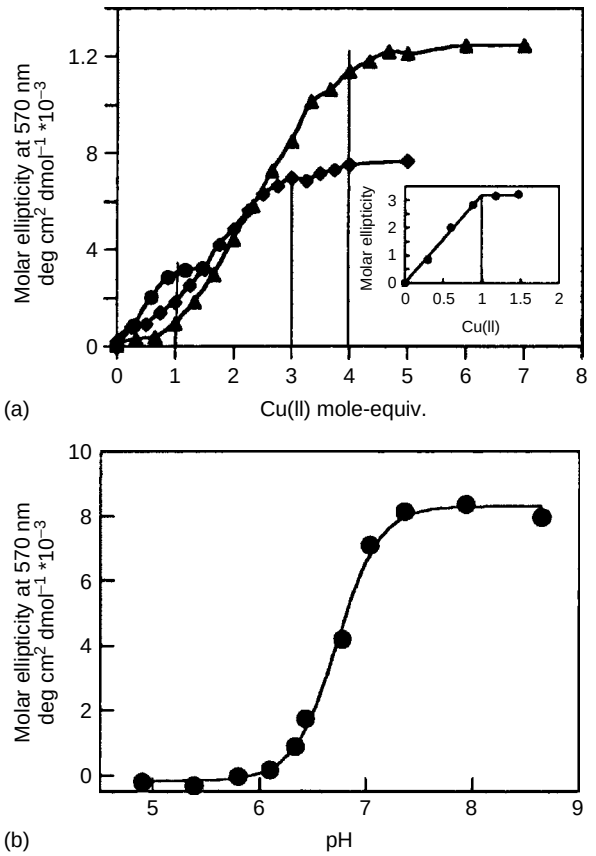

Fig. 6 Binding of copper ions to synthetic peptides containing the octarepeats of PrP. (a) Cu(II) binding curves: molar ellipticity at 570 nm with increasing amounts of Cu(II), pH 7.5. Filled circles, 2-His peptide, PrP(51–75) (0.34 mM). Filled diamonds, 3-His peptide, PrP(66–91) (0.021 mM). Filled triangles, 4-His peptide, PrP(58–91) (0.033 mM). (b) pH dependence of the ellipticity at 570 nm for PrP(58–91). The pH-dependence curve has been fitted to the following equation: $\Delta\varepsilon_{obs} = [\Delta\varepsilon_{acid}[H^+]^n + \Delta\varepsilon_{base}[H^+]K_a]/[H^+]^n + K_a]$, in which n = Hill coefficient and K_a = acid dissolution constant for the transition. The midpoint of the transition is pH 6.7. Reprinted with permission, from (Viles, J.H., Cohen, F.E., Prusiner, S.B., Goodin, D.B., Wright, P.E., Dyson, H.J. (1999) Copper binding to the prion protein: structural implications of four identical cooperative binding sites, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 2042–2047), copyright 1999 National Academy of Sciences, USA.

6.7, suggesting that the binding of Cu^{2+} occurred through the imidazole nitrogens of histidine residues.

In studies of full-length recPrP, Cu^{2+} was found to catalyze the oxidation of the histidine residues within the octarepeats. These findings argue that PrP may function as Cu^{2+}-binding protein. Consistent with studies on recPrP, membranes isolated from the brains of *Prnp*$^{0/0}$ mice were reported to have substantially lower levels of Cu^{2+} than wt mice. However, attempts to confirm this relationship between brain Cu^{2+} and PrP levels were unsuccessful. Interestingly, the copper-chelating reagent cuprizone administered to rodents causes spongiform degeneration resembling that induced by prions.

8.7 Conserved Ala-Gly Region

In addition to the octarepeat, the other region of notable conservation is in the sequence at the C-terminal end of the last octarepeat. Here, an unusual glycine- and alanine-rich region from A113 to Y128 is found (Fig. 5). Although no differences have been found in this part of the sequence between species, a single point mutation A117V is linked to GSS. The conservation of structure suggests an important role in the function of PrP^{C}; in addition, this region is likely to be important in the conversion of PrP^{C} into PrP^{Sc}.

9 Structures of PrP Isoforms

Mass spectrometry and gas phase sequencing were used to search for posttranslational chemical modifications that might explain the differences in the properties of PrP^{C} and PrP^{Sc}. No modifications differentiating PrP^{C} from PrP^{Sc} were found. These observations forced the consideration of the possibility that a conformational change distinguishes the two PrP isoforms.

When the secondary structures of the PrP isoforms were compared by optical spectroscopy, they were found to be markedly different. Fourier transform infrared (FTIR) and circular dichroism (CD) spectroscopy studies showed that PrP^{C} contains approximately 40% α-helix and little β-sheet while PrP^{Sc} is composed of approximately 30% α-helix and 45% β-sheet. That the two PrP isoforms have the same amino acid sequence runs counter to the widely accepted view that the amino acid sequence specifies only one biologically active conformation of a protein. Like PrP^{Sc}, PrP 27–30 has a high β-sheet content, which is consistent with the earlier finding that PrP 27–30 polymerizes into amyloid fibrils. Denaturation of PrP 27–30 under conditions that reduced infectivity resulted in a concomitant diminution of β-sheet content.

Prior to comparative studies on the structures of PrP^{C} and PrP^{Sc}, metabolic labeling studies showed that the acquisition of protease resistance in PrP^{Sc} is a posttranslational process. In a search for chemical differences that would distinguish PrP^{Sc} from PrP^{C}, ethanolamine was identified in hydrolysates of PrP 27–30, which signaled the possibility that PrP might contain a GPI anchor. Both PrP isoforms were found to carry GPI anchors. PrP^{C} was found on the surface of cells, where it could be released by cleavage of the anchor. Subsequent studies showed that PrP^{Sc} formation occurs after PrP^{C} reaches the cell surface and localizes to caveolae-like domains (CLDs).

9.1 NMR Structures of recPrP

The NMR structures of more than 10 PrPs from different species have been determined to be quite similar (Table 4). The recPrP molecules were produced in *E. coli* and isotopically labeled. When refolded into proteins with a high α-helical content, these recPrPs appear to be representative of PrP^C. It is noteworthy that recPrPs produced in *E. coli* have neither N-linked sugar chains nor a GPI anchor, both of which are attached to PrP^C. However, synthetic prions produced from recMoPrP(89–230) demonstrated that neither N-linked sugar chains nor a GPI anchor is required for PrP^{Sc} formation. The structure of MoDpl

Tab. 4 Structures of PrP, Dpl, and Ure2p[a].

Species	*Protein*	*Determined by*	*PDB accession codes*
Mouse	recPrP(121–231)	NMR	1AG2
Mouse	recPrP(23–231)	NMR	data not released?
Syrian hamster	recPrP(90–231)	NMR	1B10, updated 1999
Syrian hamster	recPrP(29–231)	NMR	data not released?
Syrian hamster	synPrP(104–113) in complex with 3F4	NMR	1CU4 in complex with 3F4
Syrian hamster	recPrP(90–231)	NMR	used to update 1B10
Human	recPrP(23–230)	NMR	1QLX and 1QLZ
Human	recPrP(90–230)	NMR	1QM0 and 1QM1
Human	recPrP(121–230)	NMR	1QM2 and 1QM3
Cattle	recPrP(121–230)	NMR	1DWY and 1DWZ
Cattle	recPrP(23–230)	NMR	1DX0 and 1DX1
Human	recPrP(121–230), M166V	NMR	1E1G and 1E1J
Human	recPrP(121–230), S170N	NMR	1E1P and 1E1S
Human	recPrP(121–230), R220K	NMR	1E1U and 1E1W
Human	recPrP(90–231), E200K	NMR	1QM0, 1FKC, and 1FO7
Human	recPrP(90–231)	X ray	1I4M
Human	recPrP(121–230), M166C, E221C	NMR	1H0L
Sheep	synPrP(145–169)	NMR	1G04
Sheep	synPrP(145–169)	NMR	1M25
Mouse	recDpl(26–157)	NMR	1I17
Human	recDpl(24–152)	NMR	1LG4
S. cerevisiae	recUre2p(95–354)	X ray	1G6W and 1G6Y
S. cerevisiae	recUre2p(95–354) Glutathion complex	X ray	1JZR, 1K0B, and 1K0D
S. cerevisiae	recUre2p(95–354) S-P-Nitrobenzylglutathione complex	X ray	1K0C
S. cerevisiae	recUre2p(95–354) S-Hexylglutathione complex	X ray	1K0A
S. cerevisiae	recUre2p(97–354)	X ray	1HQO

[a]Data compiled by Holger Wille.

is similar to that of MoPrP even though the two proteins share only 25% sequence similarity (Fig. 4).

The NMR structure of recSHaPrP (90–231) was determined after the protein was purified and refolded (Fig. 7a). Residues 90–112 are not shown because marked conformational heterogeneity was found in this region while residues 113–126 constitute the conserved hydrophobic region that also displays some structural plasticity. Although some features of the structure of recSHaPrP(90–231) are similar to those reported earlier for the smaller recMoPrP(121–231) fragment, substantial differences were found. For example, the loop at the N-terminus of helix B is defined in recSHaPrP(90–231) but is disordered in recMoPrP(121–231); in addition, helix C is composed of residues 200–227 in recSHaPrP(90–231) but extends only from residues 200–217 in recMoPrP(121–231). The loop and the C-terminal portion of

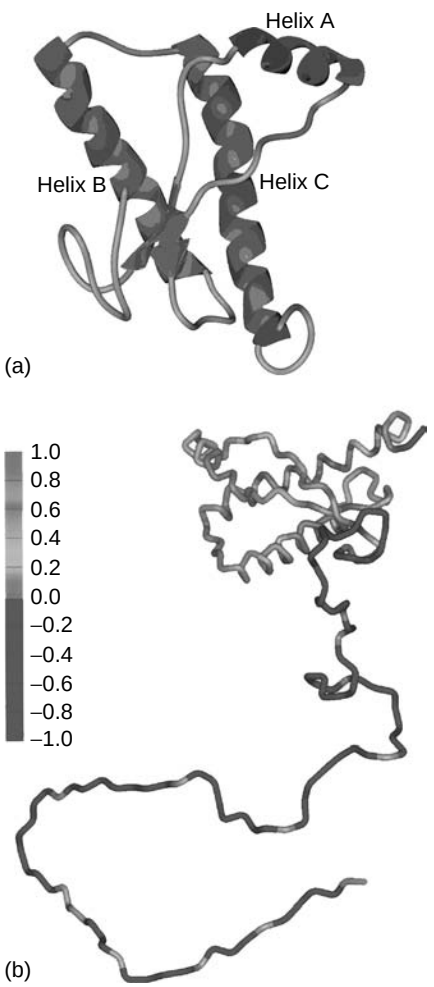

Fig. 7 Structures of PrP^C. (a) NMR structure of Syrian hamster (SHa) recombinant (rec) PrP(90–231). Presumably, the structure of the α-helical form of recPrP(90–231) resembles that of PrP^C. recPrP(90–231) is viewed from the interface where PrP^{Sc} is thought to bind to PrP^C. The color scheme is: α-helices A (residues 144–157), B (172–193), and C (200–227) in blue; loops in yellow; residues 129–134 encompassing strand S1 and residues 159–165 encompassing strand S2 in green. (b) Schematic diagram showing the flexibility of the polypeptide chain for PrP(29–231). The structure of the portion of the protein representing residues 90–231 was taken from the coordinates of PrP(90–231). The remainder of the sequence was hand-built for illustration purposes only. The color scale corresponds to the heteronuclear $[^1H]$-^{15}N NOE data: red for the lowest (most negative) values, where the polypeptide is most flexible, to blue for the highest (most positive) values in the most structured and rigid regions of the protein. (See color plate. p. xxviii)

helix C are particularly important, as described below. Whether the differences between the two recPrP fragments are due to (1) their different lengths, (2) species-specific differences in sequences, or (3) the conditions used for solving the structures remains to be determined.

NMR studies of recMoPrP(23–231) and recSHaPrP(29–231) showed that like the shorter fragments described above, full-length PrP is also a three helix–bundle protein with two short antiparallel β-strands. While the three helices form a globular C-terminal domain, the N-terminal domain is highly flexible and lacks identifiable secondary structure under the experimental conditions employed (Fig. 7b). Studies of recSHaPrP(29–231) indicate transient interactions between the C-terminal end of helix B and the highly flexible, N-terminal random-coil containing the octarepeats (residues 29–125).

9.2 Electron Crystallography of PrP^{Sc}

Because the insolubility of PrP^{Sc} has frustrated structural studies by X-ray crystallography or NMR spectroscopy, electron crystallography was used to characterize the structure of two infectious variants of PrP. Isomorphous, two-dimensional crystals of PrP 27–30 and a miniprion ($PrP^{Sc}106$) were identified by negative-stain electron microscopy. Image processing allowed the extraction of limited structural information to 7-Å resolution. Comparing projection maps of PrP 27–30 and $PrP^{Sc}106$, the 36-residue internal deletion of the miniprion was visualized and the N-linked sugars were localized (Fig. 8). The dimensions of the monomer and the locations of the deleted segment and sugars were used as constraints in the construction of models for PrP^{Sc}. Only models featuring parallel β-helices as the key element could satisfy the constraints.

A single antiparallel β-sheet as previously suggested was not consistent with the observed densities in the projection maps obtained from the 2D crystals. Specifically, the sheets were far too wide to fit into the observed hexameric arrangement. Efforts to adjust the sheet morphology to fit the density required the use of shorter strands. The amount of β-structure in these altered β-sheets was no longer compatible with the amounts of β-sheet observed by FTIR spectroscopy. Furthermore, antiparallel β-sheets typically have a twist of $\sim 20^\circ$ per strand.

From a study of 119 all-β folds observed in globular proteins, it was determined that if PrP^{Sc} follows a known protein fold, it adopts either a β-sandwich or parallel β-helical architecture. With increasing evidence arguing for a parallel β-sheet organization in amyloids, it was argued that the sequence of PrP is compatible with a parallel left-handed β-helical fold. Left-handed β-helices readily form trimers, providing a natural template for a trimeric model of PrP^{Sc} (Fig. 9a). This trimeric model accommodates the PrP sequence from residues 89–175 in a β-helical conformation with the C-terminus (residues 176–227) retaining the disulfide-linked α-helical conformation observed in PrP^{C}. In addition, the proposed model matches the structural constraints of the crystals, positioning residues 141–176 and the N-linked sugars appropriately. The parallel left-handed β-helical model provides a coherent framework that is consistent with the stacking of trimeric units of PrP 27–30 to form PrP amyloid fibrils (Fig. 9b).

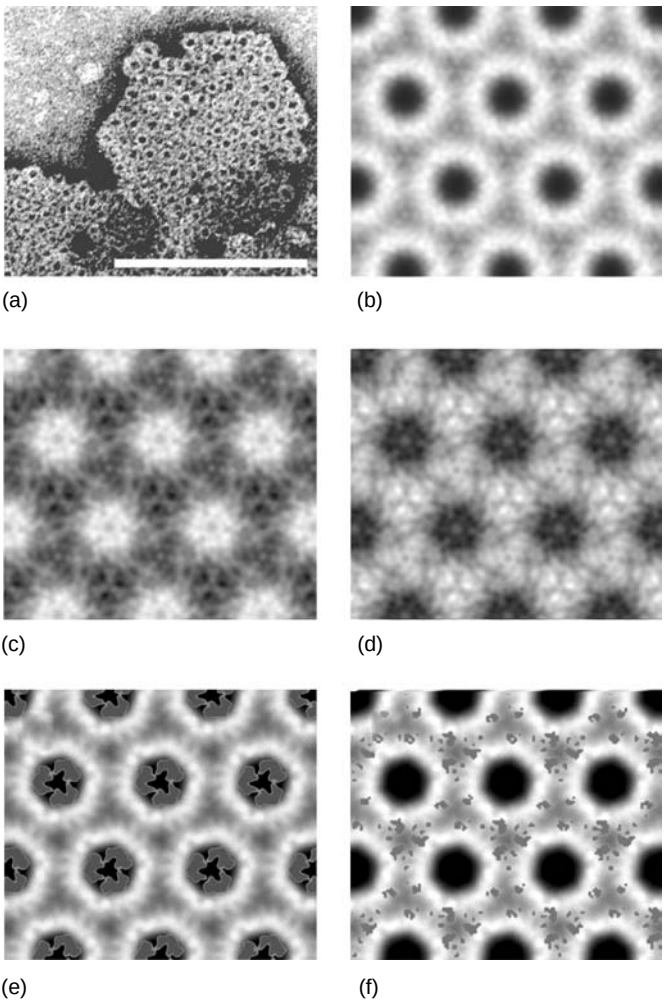

Fig. 8 Two-dimensional crystals of PrP^{Sc}. (a) A 2D crystal of $PrP^{Sc}106$ stained with uranyl acetate. Bar = 100 nm. (b) Image processing result after correlation-mapping and averaging followed by crystallographic averaging. (c and d) Subtraction maps between the averages of PrP 27–30 and $PrP^{Sc}106$ (panel B). (c) $PrP^{Sc}106$ minus PrP 27–30 and (d) PrP 27–30 minus $PrP^{Sc}106$, showing major differences in lighter shades. (e and f) The statistically significant differences between PrP 27–30 and $PrP^{Sc}106$ calculated from (c) and (d) in red and blue, respectively, overlaid onto the crystallographic average of PrP 27–30. Reprinted with permission, from Wille, H., Michelitsch, M.D., Guénebaut, V., Supattapone, S., Serban, A., Cohen, F.E., Agard, D.A. Prusiner, S.B. (2002) Structural studies of the scrapie prion protein by electron crystallography, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 3563–3568 Wille et al., (2002), copyright 2002 National Academy of Sciences, USA. (See color plate. p. xxix)

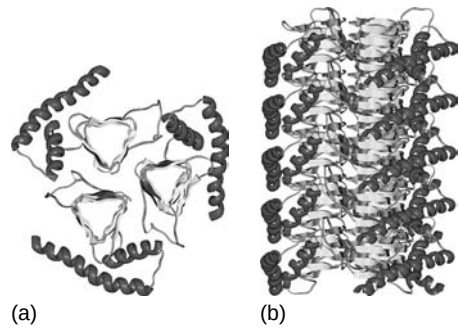

Fig. 9 Left-handed β-helical models of PrP 27–30. (a) Trimeric model of PrP 27–30 built by superimposing three monomeric models onto the coordinates of the C'_αs of the structure of trimeric carbonic anhydrase from *Methanosarcina thermophila* (PDB ID 1THJ). (b) A model of the PrP 27–30 fiber, constructed by stacking five trimeric discs. Figure prepared by Cédric Govaerts.

10 Prion Replication

The mechanism of prion replication is not well understood. The formation of nascent prions requires that PrP^C be converted into PrP^{Sc}. Some investigators have argued that this process involves a nucleation–polymerization (NP) reaction while others have contended that a more likely mechanism employs a template-assisted process.

In an uninfected cell, PrP^C with the wt sequence is likely to exist in equilibrium in its monomeric α-helical, protease-sensitive state or bound to some other protein, such as protein X (Fig. 10). The conformation of PrP^C that is bound to protein X is denoted PrP*; this conformation is likely to be different from that determined under aqueous conditions for monomeric recPrP. The PrP*–protein X complex will bind PrP^{Sc}, thereby creating a replication-competent assembly. Order-of-addition experiments demonstrate that protein X binding to PrP^C precedes productive PrP^{Sc} interactions. A conformational change takes place wherein PrP*, in a shape competent for binding to protein X and PrP^{Sc}, represents the initial phase in the formation of infectious PrP^{Sc}.

Although many lines of evidence reviewed below argue for a template-assisted process in the formation of nascent PrP^{Sc}

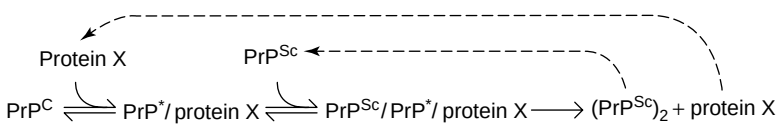

Fig. 10 Hypothetical scheme for template-assisted PrP^{Sc} formation. In the initial step, PrP^C binds to protein X to form the PrP*–protein X complex. Next, PrP^{Sc} binds to PrP* that has already formed a complex with protein X. When PrP* is transformed into a nascent molecule of PrP^{Sc}, protein X is released and a dimer of PrP^{Sc} remains. The inactivation target size of an infectious prion suggests that it is composed of a dimer of PrP^{Sc}. In the model depicted here, a fraction of infectious PrP^{Sc} dimers dissociates into uninfectious monomers as the replication cycle proceeds, while a majority of the dimers accumulates in accord with the increase in prion titer that occurs during the incubation period. Another fraction of PrP^{Sc} is cleared presumably by cellular proteases. The precise stoichiometry of the replication process remains uncertain. Reprinted with permission, from (Prusiner, S.B., Scott, M.R., DeArmond, S.J., Cohen, F.E. (1998) Prion protein biology, *Cell* **93**, 337–348), copyright 1998 Cell Press.

in mammalian cells, the formation of synthetic prions from PrP(89–143,P101L) as well as from recPrP(89–230) establishes that infectious prions can be generated *de novo* without the aid of auxiliary proteins.

10.1 Mechanism of Prion Propagation

In attempting to probe the mechanism of prion formation, it is reasonable to determine the rate-limiting step in prion formation. First, the impact of the concentration of PrP^{Sc} in the inoculum, which is inversely proportional to the length of the incubation time, must be considered. Second, the sequence of PrP^{Sc} that forms an interface with PrP^{C} must be taken into account. When the sequences of the two isoforms are identical, the shortest incubation times are observed. Third, the strain-specific conformation of PrP^{Sc} must be considered. Some prion strains exhibit longer incubation times than others; the mechanism underlying this phenomenon is not understood. From these considerations, a set of conditions exist under which initial PrP^{Sc} concentrations can be rate limiting. These effects presumably relate to the stability of the PrP^{Sc}, its targeting to the correct cells and subcellular compartments, and its ability to be cleared. Once infection in a cell is initiated and endogenous PrP^{Sc} production is operative, then the following discussion of PrP^{Sc} formation seems most applicable. If the assembly of PrP^{Sc} into a specific dimeric or multimeric arrangement is difficult, then an NP formalism would be relevant. In NP processes, nucleation is the rate-limiting step and elongation, or polymerization, is facile. These conditions are frequently observed in peptide models of aggregation phenomena; however, studies in ScN2a cells and with Tg mice expressing foreign PrP genes suggest that a different process occurs.

10.2 Template-assisted Prion Formation

From investigations with mice expressing both the SHaPrP transgene and the endogenous MoPrP gene, designated Tg(SHaPrP) mice, it is clear that PrP^{Sc} provides a template for directing prion replication; a template is defined as a catalyst that leaves its imprint on the product of the reaction. Inoculation of these mice with $SHaPrP^{Sc}$ leads to the production of nascent $SHaPrP^{Sc}$ and not $MoPrP^{Sc}$. Conversely, inoculation of the Tg(SHaPrP) mice with $MoPrP^{Sc}$ results in $MoPrP^{Sc}$ and not $SHaPrP^{Sc}$ formation.

Even stronger evidence for templating has emerged from studies of prion strains passaged in Tg(MHu2M)$Prnp^{0/0}$ mice, which express a chimeric mouse–human gene, as described in more detail below. Even though the conformational templates were initially generated with PrP^{Sc} molecules having different sequences from patients with inherited prion diseases, these templates are sufficient to direct replication of distinct PrP^{Sc} molecules when the amino acid sequences of the substrate PrPs are identical.

Another line of evidence for template-assisted prion replication comes from studies of FI. In FFI, the protease-resistant fragment of PrP^{Sc} after deglycosylation has an M_r of 19 kDa when measured by immunoblotting of brain extracts from humans with FFI or from Tg(MHu2M)$Prnp^{0/0}$ mice that were inoculated with FFI prions. In both humans and inoculated Tg(MHu2M)$Prnp^{0/0}$ mice on both first and second passage, PrP^{Sc} is confined largely to the thalamus. These findings argue that the conformation of

PrP^{Sc} that yields a 19-kDa polypeptide after deglycosylation is propagated both in humans and in mice expressing an artificial chimeric transgene; in addition, prion accumulation is confined to the thalamus. Additional evidence supporting these assertions comes from a patient who died after developing a clinical disease similar to FFI. Because both PrP alleles encoded the wt PrP sequence and methionine at position 129, we labeled this sFI. At autopsy, the spongiform degeneration, reactive astrocytic gliosis, and PrP^{Sc} deposition were confined to the thalamus. Moreover, the PrP^{Sc} both in the patient's brain and in the brains of inoculated Tg(MHu2M)*Prnp*$^{0/0}$ mice yielded a 19-kDa fragment after deglycosylation. These findings argue that the clinicopathologic phenotype is determined by the conformation and not the amino acid sequence of PrP^{Sc}. Additionally, FI prions can be replicated with mutant human PrP(D178N), wt human PrP, or chimeric human–mouse PrP. With these different PrPs as substrates, the conformation of PrP^{Sc} must be faithfully copied; template-assisted prion replication provides for such a mechanism.

10.3 Evidence for Protein X

Protein X was postulated to explain the results of the transmission of Hu prions to Tg mice (Table 5). Mice expressing both $MoPrP^{C}$ and $HuPrP^{C}$ were resistant while those expressing only $HuPrP^{C}$ were susceptible to Hu prions. These results argue that $MoPrP^{C}$ inhibited transmission of Hu prions, that is, the formation of nascent $HuPrP^{Sc}$. In contrast, mice expressing both $MoPrP^{C}$ and chimeric $MHu2MPrP^{C}$ were susceptible to Hu prions and mice expressing $MHu2MPrP^{C}$ alone were only slightly more susceptible. These findings contend that $MoPrP^{C}$ has only a minimal effect on the formation of chimeric $MHu2MPrP^{Sc}$.

When the data on Hu prion transmission to Tg mice were considered together, they suggested that $MoPrP^{C}$ prevented the conversion of $HuPrP^{C}$ into PrP^{Sc} by binding to another Mo protein (protein X) with a higher affinity than $HuPrP^{C}$ does. It was postulated that $MoPrP^{C}$ has little effect on the formation of PrP^{Sc} from $MHu2MPrP^{C}$ (Table 5) because $MoPrP^{C}$ and $MHu2MPrP^{C}$ share the same amino

Tab. 5 Influence of prion species and strain on transmission to Syrian hamsters and mice[a].

Inoculum	*Recipient*	*Incubation time [days ± SEM] (n/n_0)*	
		Sc237 prions	*139H prions*
SHa>SHa	SHa	77 ± 1 (48/48)	167 ± 1 (94/94)
SHa>SHa	non-Tg mice	>700 (0/9)	499 ± 15 (11/11)
SHa>SHa	Tg(SHaPrP)81 mice	75 ± 2 (22/22)	110 ± 2 (19/19)

[a]The species of prion is encoded by the primary structure of PrP^{Sc} and the strain of prion is enciphered by the tertiary structure of PrP^{Sc}. We recognize that the primary structure as well as posttranslational chemical modifications determine the tertiary structure of PrP^{C}, but we argue that the conformation of PrP^{C} is modified by PrP^{Sc} as it is refolded into a nascent PrP^{Sc} molecule.

acid sequence at the C-terminus. This also suggested that MoPrP^{C} only weakly inhibited transmission of SHa prions to Tg(SHaPrP) mice because SHaPrP is more closely related to MoPrP than to HuPrP.

The search for protein X has been frustrating because many proteins are known to bind to PrP^{C} but none have been shown to participate in PrP^{Sc} formation. Whether protein X is one protein or a complex of proteins remains to be established. It is reasonable to expect that mice deficient in protein X would not replicate prions or do so extremely slowly, resulting in prolonged incubation times.

10.4 Dominant-negative Inhibition

To extend the foregoing findings in Tg mice, an expression vector with an insert encoding a chimeric mouse–hamster PrP, designated MHM2PrP, was transfected into ScN2a cells. MHM2PrP carries two Syrian hamster residues, which create an epitope that is recognized by the anti-SHaPrP 3F4 mAb that does not react with mouse PrP. Substitution of a Hu residue at position 214 or 218 in MHM2PrP prevented formation of MHM2PrP^{Sc} in ScN2a cells. The side chains of these residues protrude from the same surface of the C-terminal α-helix, forming a discontinuous epitope with residues 167 and 171 in an adjacent loop. Like MHM2PrP(Q218K), substitution of a basic residue at position 167 or 171 prevented PrP^{Sc} formation. When MHM2PrP and MHM2PrP(Q218K) were coexpressed, the conversion of MHM2PrP^{C} into PrP^{Sc} was inhibited, arguing that MHM2PrP(Q218K) was acting as a dominant negative. Similar results were obtained when studies were performed with cells expressing MHM2PrP(Q167R).

One interpretation of dominant-negative inhibition of prion propagation is that mutant PrP binds more avidly to protein X than does wt PrP. In this scenario, mutant PrP binds to protein X and prevents binding of wt PrP. This explanation is also consistent with the protective effects of basic polymorphic residues in PrP of humans and sheep. The E219K substitution seems to render humans resistant to sCJD and Q171R renders sheep resistant to scrapie. In MHM2, the Q218K mutation corresponds to E219K in humans and the Q167R mutation corresponds to Q171R in sheep (Fig. 11).

To determine whether dominant-negative inhibition of prion formation occurs *in vivo*, Tg mice expressing PrP with either the Q167R or Q218K substitution alone or in combination with wt PrP were produced. Tg(MoPrP,Q167R)*Prnp*$^{0/0}$ mice expressing mutant PrP at levels equal to non-Tg FVB mice were inoculated with prions and remained healthy for more than 550 days, indicating that inoculation did not initiate a process sufficient to cause disease. Immunoblots of brain homogenates and histologic analysis revealed no abnormalities. Tg(MoPrP,Q167R)*Prnp*$^{+/+}$ mice expressing both mutant and wt PrP exhibited neurologic dysfunction at ~450 days after inoculation; the brains of three of these mice that were sacrificed at 300 days revealed low levels of PrP^{Sc} as well as numerous vacuoles and severe astrocytic gliosis. Both Tg(MoPrP,Q218K)*Prnp*$^{0/0}$ and Tg(MoPrP,Q218K)*Prnp*$^{+/+}$ mice expressing the transgene product at 16× remained healthy for more than ~450 days after inoculation. Neither PrP^{Sc} deposition nor neuropathology was found. Tg mice expressing MoPrP(Q218K) at 32× developed spontaneous neurodegeneration at ~450 days of age. These studies

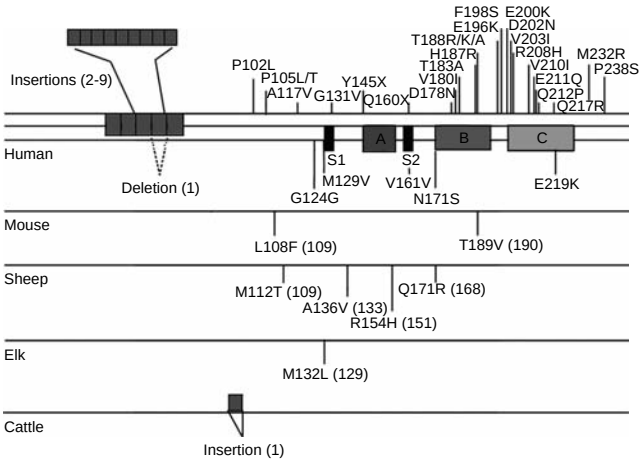

Fig. 11 Mutations and polymorphisms of the prion protein gene. PrP mutations causing inherited human prion disease and PrP polymorphisms found in humans, mice, sheep, elk, and cattle. Above the line of the human sequence are mutations that cause inherited prion disease. Below the lines are polymorphisms, some but not all of which are known to influence the onset as well as the phenotype of disease. Residue numbers in parentheses correspond to the human codons. Much of the data was compiled by J.-L. Laplanche.

demonstrate that although dominant-negative inhibition of wt PrP^{Sc} formation occurs, expression of dominant-negative PrP at the same level as wt PrP does not completely prevent prion formation. However, expression of dominant-negative PrP alone had no deleterious effects on the mice and did not support prion propagation.

11 Sporadic Human Prion Diseases

In most patients with CJD, there is neither an infectious nor heritable etiology. How prions arise in patients with sCJD is unknown; hypotheses include: (1) horizontal transmission from humans or animals; (2) somatic mutation of the *PRNP* ORF; (3) spontaneous conversion of PrP^{C} into PrP^{Sc}; and (4) the accumulation of PrP^{Sc}, which is normally present at very low levels. Numerous attempts to establish an infectious link between sCJD and a preexisting prion disease in animals or humans have been unrewarding. Studies demonstrating the formation of synthetic prions from PrP(89–143,P101L) as well as from recPrP(89–230) show that only PrP^{C} is needed for an animal for producing infectious prions. Thus, the spontaneous conversion of PrP^{C} to PrP^{Sc} seems reasonable and any mammal expressing PrP^{C} is capable of forming prions.

12 Heritable Human Prion Diseases

The recognition that 10 to 15% of CJD cases are familial led to the suspicion that genetics plays a role in this disease. As with scrapie, the relative contributions of genetic and infectious etiologies in the human prion diseases remain puzzling.

More than 30 different mutations of the PrP gene have been shown to segregate with the heritable human prion diseases. Five of these mutations have been genetically linked to the inherited human prion diseases (Table 3). Virtually all cases of GSS and FFI appear to be caused by germline mutations in the PrP gene. The brains of humans dying of inherited prion disease contain infectious prions that have been transmitted to experimental animals.

12.1 GSS and Genetic Linkage

The discovery that GSS, which is a familial disease, could be transmitted to apes and monkeys was first reported when many still thought that scrapie, CJD, and related disorders were caused by viruses. With the discovery that the P102L mutation of the human PrP gene was genetically linked to GSS, prion diseases were concluded to have both genetic and infectious etiologies. In that study, the P102L mutation was linked to development of GSS with a logarithm of the odds (LOD) score exceeding 3, demonstrating a tight association between the altered genotype and disease phenotype (Fig. 3B). This mutation may be caused by the deamination of a methylated CpG in a germline PrP gene, which results in the substitution of a leucine for proline. This mutation has been found in many families in numerous countries, including the first family identified to have GSS.

12.2 fCJD Caused by Octarepeat Inserts

An insert of 144 bp containing six octarepeats at codon 53, in addition to the five that are normally present, was described in patients with CJD from four families residing in southern England. Genealogic investigations showed that all four families are related, arguing for a single founder born more than two centuries ago. The LOD score for this extended pedigree exceeds 11. Studies from several laboratories have demonstrated that inserts of two, four, five, six, seven, eight, or nine octarepeats in addition to the normal five are found in individuals with inherited CJD (Fig. 11).

12.3 fCJD in Libyan Jews

The unusually high incidence of CJD among Israeli Jews of Libyan origin was thought to be due to the consumption of lightly cooked sheep brain or eyeballs. Molecular genetic investigations revealed that Libyan and Tunisian Jews with fCJD have a PrP gene point mutation at codon 200, resulting in a Glu-to-Lys substitution (Fig. 11). The E200K mutation has been genetically linked to fCJD, with an LOD score exceeding 3; the same mutation has also been found in patients from Orava in North Central Slovakia, in a cluster of familial cases in Chile, and in a large German family living in the United States.

Most patients are heterozygous for the mutation and thus, express both mutant and wt PrP^{C}. In the brains of patients who die of fCJD(E200K), mutant PrP^{Sc} is both insoluble and protease-resistant, whereas much of wt PrP differs from both PrP^{C} and PrP^{Sc} in that it is insoluble but readily digested by proteases. Whether this form of PrP is an intermediate in the conversion of PrP^{C} into PrP^{Sc} remains to be established.

12.4 Penetrance of fCJD

Life-table analyses of carriers harboring the codon 200 mutation exhibit complete

penetrance. Therefore, if the carriers live long enough, they will all eventually develop prion disease. Some investigators have argued that the inherited prion diseases are not fully penetrant and thus an environmental factor, such as the ubiquitous "scrapie virus," is required for illness to be manifest, but no viral pathogen has been found to date.

12.5 Fatal Familial Insomnia

The D178N mutation has been linked to the development of FFI, with an LOD score exceeding 5. More than 30 families worldwide with FFI have been recorded. Studies of inherited human prion diseases demonstrate that the amino acid at polymorphic residue 129 with the D178N pathogenic mutation alters the clinical and neuropathologic phenotype. The D178N mutation combined with M129 results in FFI. In this disease, adults generally over age 50 years present with a progressive sleep disorder and usually die within one year. In their brains, deposition of PrP^{Sc} is confined largely within the anteroventral and the dorsal medial nuclei of the thalamus. In contrast, the same D178N mutation with V129 produces fCJD, in which the patients present with dementia and widespread deposition of PrP^{Sc} is found postmortem. The first family to be recognized with CJD was found to carry the D178N mutation.

12.6 Human PrP Gene Polymorphisms

Several polymorphisms in the PrP gene may affect the development and expression of prion disease. A Met/Val polymorphism at PrP codon 129 has been identified (Fig. 11). This polymorphism appears to influence expression of the inherited as well as the iatrogenic and sporadic forms of prion disease. A second polymorphism resulting in an amino acid substitution at codon 219 (Glu/Lys) has been reported to occur with a frequency of approximately 12% in the Japanese population but not in Caucasians. In people heterozygous for the lysine substitution, this expression seems to act as a dominant negative and protects against CJD by the avid binding of mutant PrP^{C} to protein X. By sequestering protein X, PrP^{C}(K219) prevents PrP^{C}(E219) from being converted into PrP^{Sc}. In people homozygous for the lysine substitution, it seems likely that PrP^{C}(K219) is not converted into PrP^{Sc}. A third polymorphism results in an amino acid substitution at codon 171 (Asn/Ser), which lies adjacent to the protein X binding site. This polymorphism has been found in Caucasians but it has not been studied extensively and it is not known to influence the binding of PrP^{C} to protein X. A fourth polymorphism is the deletion of a single octarepeat (24 bp), which has been found in 2.5% of Caucasians. In another study of over 700 individuals, this single octarepeat deletion was found in 1.0% of the population.

Studies of Caucasian patients with sCJD have shown that most are homozygous for Met or Val at codon 129. This contrasts with the general Caucasian population, in which frequencies for the codon 129 polymorphism are 12% Val/Val, 37% Met/Met, and 51% Met/Val. In contrast, the frequency of the valine allele in the Japanese population is much lower. Heterozygosity at codon 129 is more frequent in Japanese CJD patients (18%) than in the general Japanese population, in which the polymorphism frequencies are 0% Val/Val, 92% Met/Met, and 8% Met/Val.

While no specific mutations have been identified in the PrP gene of patients with sCJD, homozygosity at codon 129 is consistent with the results of Tg mouse studies. The finding that homozygosity at codon 129 predisposes to sCJD supports a model of prion production that favors PrP interactions between homologous proteins, as appears to occur in Tg(SHaPrP) mice inoculated with either hamster prions or mouse prions, as well as in Tg(MHM2) mice inoculated with "artificial" prions.

13 Infectious Human Prion Diseases

Prions from different sources have infected humans. Human prions have been transmitted to others both by ritualistic cannibalism and iatrogenic means. Kuru in the highlands of New Guinea was transmitted by ritualistic cannibalism, as people in the region attempted to immortalize their dead relatives by eating their brains. Iatrogenic transmissions include prion-tainted human growth hormone (HGH) and gonadotropin, dura mater grafts, and corneal transplants from people who died of CJD. In addition, CJD cases have been recorded after neurosurgical procedures in which ineffectively sterilized depth electrodes or instruments were used.

13.1 Human Growth Hormone

More than 165 young adults have been diagnosed with iatrogenic (i) CJD between 4 and 30 years after receiving HGH or gonadotropin from cadaveric pituitaries. The longest incubation periods (20 to 30 years) are similar to those associated with more recent cases of kuru. Since 1985, recombinant HGH produced in *E. coli* has been used in place of cadaveric HGH. With recombinant HGH, no cases of iCJD have been identified.

13.2 Variant Creutzfeldt–Jakob Disease (vCJD)

In 1994, the first cases of CJD in teenagers and young adults that were eventually labeled variant (v) CJD occurred in Great Britain. More than 150 teenagers and young adults have died of vCJD in Britain, Ireland, Italy, and France. While the average age of vCJD patients is 26 years of age, the youngest patient was 12 years old and the oldest was 74 years of age. The median duration of illness was 13 months, with the range from 6 to 69 months.

In addition to the young age of these patients, vCJD is characterized by numerous PrP amyloid plaques surrounded by a halo of intense spongiform degeneration in the brain. These unusual neuropathologic changes have not been seen in CJD cases in the United States, Australia, or Japan. Both macaque monkeys and marmosets developed neurologic disease several years after inoculation with bovine prions, but only the macaques exhibited numerous PrP plaques similar to those found in vCJD.

The majority of vCJD patients presented with psychiatric symptoms, including dysphoria, withdrawal, anxiety, insomnia, and loss of interest. Generally, neurologic deficits did not appear until at least four months later; these neurologic changes consisted of memory loss, paresthesias, sensory deficits, gait disturbances, and dysarthria. To date, all vCJD cases have been reported from Britain, except for six cases in France, one in Italy, one in Ireland, and one in the United States. The US case is a 23-year-old woman, who is thought to have been exposed to bovine prions while living in Britain during the first 12 years of

her life. From both epidemiologic and experimental studies, evidence is now quite compelling that vCJD is the result of prions being transmitted from cattle with BSE to humans through consumption of contaminated beef products.

Studies of the prion diseases have taken on new significance with the identification of vCJD. The number of cases of vCJD caused by bovine prions that will occur in the years ahead is unknown. Until more time passes, it will be impossible to assess the magnitude of this problem. These tragic cases have generated a continuing discourse concerning mad cows, prions, and the safety of the human and animal food supplies throughout the world. Untangling politics and economics from the science of prions seems to have been difficult in disputes between Great Britain and other European countries over the safety of beef and lamb products.

13.3 Transmission of vCJD Prions by Blood Transfusion

vCJD has been identified in two patients who received blood transfusions from donors that later died of vCJD. In one case, the recipient was a 69-year-old male who was transfused 6.5 years before the onset of neurologic dysfunction. His PrP gene encoded Met/Met at position 129. Many details of the second case are not published but the patient is known to have died of a non-neurologic disease. While vCJD prions were found in the spleen and cervical lymph nodes of this patient, none were found in the brain. The patient's PrP gene encoded Met/Val at position 129 and as such, is the first heterozygous codon 129 human identified with vCJD prions.

A glimpse of future vCJD cases caused by prion-tainted transfused blood may come from a survey of tissues collected during appendectomies and tonsillectomies. This UK survey reports that of the 12 674 appendectomy specimens examined, three were positive for PrP^{Sc} by immunohistochemistry. This finding argues that as many as 3800 people in the United Kingdom may be replicating vCJD prions in their lymphoid tissues. Considering that immunohistochemistry is less sensitive than Western immunoblotting, the number of people with vCJD prions in their lymphoid tissues may be substantially greater than the above estimate.

14 Strains of Prions

That goats with scrapie can manifest two different syndromes, one in which the goats became hyperactive and the other in which they became drowsy, raised the possibility that strains of prions might exist. Subsequent studies with mice documented the existence of multiple strains through careful measurements of incubation times and the distribution of vacuoles in the CNS.

The search for an explanation of how biological information encrypting the disease phenotype could be enciphered within the prion posed a conundrum. Many investigators kept arguing for a small nucleic acid but none could be found. Others argued that this information must be enciphered within the structure of PrP^{Sc}. The first evidence supporting the hypothesis that strain-specific information is enciphered in PrP^{Sc} came from studies on prions causing TME, which were passaged into Syrian hamsters. On serial passaging, two strains

emerged: one strain (HY) produced hyperactivity in Syrian hamsters and the other (DY) was manifest as a drowsy syndrome, as was seen with scrapie strains in goats. The HY strain was similar to the Sc237 prion strain with respect to protease resistance and sedimentation properties, while the DY strain showed minimal protease resistance and was less readily sedimented.

Because the HY and DY strains emerged after passaging TME prions in Syrian hamsters, concern over the origin of these strains made the conclusions somewhat questionable. Because of such concerns, a more convincing investigation came from the transmission of FFI and fCJD to Tg mice. In this study, two different strains of prions were generated *de novo* in patients with PrP gene mutations and propagated in mice expressing the chimeric MHu2MPrP transgene. Brain homogenates of FFI and fCJD(E200K) patients transmitted disease to Tg(MHu2M) mice approximately 200 days after inoculation. The FFI inoculum induced formation of a 19-kDa PrP^{Sc} fragment, as measured by SDS-PAGE after limited proteolysis and removal of Asn-linked carbohydrates; in contrast, fCJD(E200K) produced a 21-kDa PrP^{Sc} fragment. On second passage, Tg(MHu2M) mice inoculated with FFI prions showed an incubation time of ~130 days and a 19-kDa PrP^{Sc} fragment while those inoculated with fCJD(E200K) prions exhibited an incubation time of ~170 days and a 21-kDa PrP^{Sc} fragment. The experimental data demonstrated that $MHu2MPrP^{Sc}$ can exist in two different conformations, based on the sizes of the protease-resistant fragments, within an invariant amino acid sequence. The results of these studies argue that PrP^{Sc} acts as a template for the conversion of PrP^{C} into nascent PrP^{Sc}. Imparting the size of the protease-resistant fragment of PrP^{Sc} through conformational templating provides a mechanism for both the generation and propagation of prion strains.

In another set of studies, prions from cattle with BSE and humans with vCJD were used to demonstrate distinct prion strains in humans expressing wt PrP, in contrast to the studies described above with prions from patients harboring the E200K mutation. Both BSE and vCJD prions transmitted to Tg mice expressing BoPrP but not to mice expressing either HuPrP or MHu2MPrP. Moreover, sCJD, iCJD, and fCJD(E200K) prions failed to transmit disease to Tg(BoPrP) mice after more than 500 days. These findings demonstrate the prions that cause BSE and vCJD are similar, but distinct from those responsible for sCJD, iCJD, and fCJD(E200K).

14.1 Isolation of New Strains

Additional evidence implicating PrP in the phenomenon of prion strains comes from studies on the interspecies transmission of strains. Such studies were especially revealing when mice expressing chimeric hamster–mouse transgenes were used. Evidence was obtained showing that new strains pathogenic to Syrian hamsters could be obtained from prion strains that had been previously replicated by limiting dilution in mice. Both the generation and propagation of prion strains result from interactions between PrP^{Sc} and PrP^{C}. Additionally, strains that were isolated from different breeds of sheep with scrapie and thought to be distinct were shown to have identical properties. Such findings argue for the convergence of some strains and raise the issue of the limits of prion diversity.

Studies with different strains of prions propagated in Syrian hamsters and subsequently in Tg mice expressing MH2MPrP are most illustrative with respect to the isolation of new prion strains. These Tg mice were inoculated with the two prion strains, Sc237 and DY that had been serially passaged in Syrian hamsters. On first passage in Tg(MHM2)*Prnp*$^{0/0}$ mice, the SHa(Sc237) prions exhibited prolonged incubation times, typical of a species barrier. On subsequent passage in Tg(MHM2)*Prnp*$^{0/0}$ mice, the MH2M(Sc237) strain showed a profound shortening of incubation time. Moreover, PrP^{Sc} of the MH2M(Sc237) strain possesses different structural properties from PrP^{Sc} of the SHa(Sc237) strain, as demonstrated by relative conformational stability measurements. Conversely, transmission of SHa(DY) prions to Tg(MHM2)*Prnp*$^{0/0}$ mice did not encounter a species barrier and the MH2M(DY) strain retained the conformational and phenotypic properties of SHa(DY). These results extend the findings described above for FFI and fCJD(E200K) prions and contend that a change in PrP^{Sc} conformation is intimately associated with the emergence of a new prion strain.

14.2 Interplay between the Species and Strains of Prions

The recent advances in our understanding of the role of the primary and tertiary structures of PrP in the transmission of disease have given new insights into the pathogenesis of the prion diseases. The amino acid sequence of PrP encodes the species of the prion (Table 5), and the prion derives its PrP^{Sc} sequence from the last mammal in which it was passaged. While the primary structure of PrP is likely to be the most important or even sole determinant of the tertiary structure of PrP^{C}, existing PrP^{Sc} seems to function as a template in determining the tertiary structure of nascent PrP^{Sc} molecules as they are formed from PrP^{C}. In turn, prion diversity appears to be enciphered in the conformation of PrP^{Sc} and prion strains represent different conformers of PrP^{Sc}. The total number of prion strains, that is, different conformations of PrP^{Sc}, in existence remains to be established.

15 Prion Diseases of Animals

The prion diseases of animals include scrapie of sheep and goats, BSE, TME, chronic wasting disease (CWD) of mule deer and elk, feline spongiform encephalopathy, and exotic ungulate encephalopathy (Table 1).

15.1 PrP Polymorphisms in Sheep, Cattle, and Elk

In 1962, Parry argued that host genes were responsible for the development of scrapie in sheep. He was convinced that natural scrapie is a genetic disease that could be eradicated by proper breeding protocols. He considered its transmission by inoculation of importance primarily for laboratory studies and communicable infection of little consequence in nature. Other investigators viewed natural scrapie as an infectious disease and argued that host genetics only modulates susceptibility to an endemic infectious agent.

In sheep, polymorphisms at codons 136, 154, and 171 of the PrP gene that produce amino acid substitutions have been studied with respect to the

incidence of scrapie (Fig. 11). Studies of natural scrapie in the United States have shown that 85% of afflicted sheep are of the Suffolk breed. Only those Suffolk sheep homozygous for Gln (Q) at codon 171 developed scrapie although healthy controls with Gln/Gln, Gln/Arg, and Arg/Arg genotypes were also found. These results argue that susceptibility in Suffolk sheep is governed by the PrP codon 171 polymorphism. As with the Suffolk breed, the PrP codon 171 polymorphism in Cheviot sheep has a profound influence on susceptibility to scrapie, and codon 136 also modulates susceptibility but less so than codon 171.

In contrast to sheep, different breeds of cattle have no specific PrP polymorphisms. The only polymorphism recorded in cattle is a variation in the number of octarepeats: most cattle, like humans, have five octarepeats but some have six. However, the presence of six octarepeats does not seem to be overrepresented in BSE.

In studies of CWD, the susceptibility of elk, but not deer, seems to be modulated by codon 132, which corresponds to codon 129 in humans (Fig. 11). Elk with CWD consistently express Met/Met at position 132; no elk with CWD expressing leucine at this residue have been found.

15.2 Bovine Spongiform Encephalopathy

Prion strains and the species barrier are of paramount importance in understanding the BSE epidemic in Britain, in which it is estimated that between one and two million cattle were infected with prions. The mean incubation time for BSE is approximately 5 years. Therefore, most cattle did not manifest disease because they were slaughtered between 2 and 3 years of age. Nevertheless, more than 180 000 cattle, primarily dairy cows, have died of BSE over the past decade. BSE is a massive common-source epidemic caused by meat and bone meal (MBM) fed primarily to dairy cows. MBM was prepared from the offal of sheep, cattle, pigs, and chickens as a high-protein nutritional supplement. In the late 1970s, the hydrocarbon-solvent extraction method used in the rendering of offal began to be abandoned, resulting in MBM with a much higher fat content. It is now thought that this change allowed scrapie prions from sheep to survive the rendering process and to be passed into cattle. Alternatively, bovine prions may have been present at low levels prior to modification of the rendering process and with the processing change, survived in sufficient numbers to initiate the BSE epidemic when reintroduced into cattle through ingestion of MBM. Perhaps, a particular conformation of $BoPrP^{Sc}$ selected for heat-resistance during the rendering process and then reselected multiple times as cattle infected by ingesting prion-contaminated MBM were slaughtered and their offal rendered into more MBM. Against the latter hypothesis is the widespread geographical distribution of the initial 17 cases of BSE throughout England, which occurred almost simultaneously.

In July 1988, the practice of feeding MBM to sheep and cattle was banned. Although statistical analyses demonstrate that the epidemic is disappearing as a result of this ruminant feed ban, reminiscent of the disappearance of kuru in the Fore people of New Guinea, it is unclear how many current cases of BSE are due to infection and how many arise spontaneously. In 2003, there were 1390 cases of BSE; how many were due to tainted feed is unknown. Although evidence of a preexisting prion disease of

cattle, either in Great Britain or elsewhere, is scant, an outbreak of TME in Wisconsin has been cited as evidence for sporadic BSE. Moreover, the formation of synthetic prions from wt PrP argues that any mammal expressing PrP^{C} can produce prions spontaneously.

The origin of BSE prions cannot be determined by examining the amino acid sequence of PrP^{Sc} from cattle with BSE because it has the bovine sequence regardless of whether the initial prions originated from sheep or cattle. The bovine PrP sequence differs from that of sheep at seven or eight positions.

Brain extracts from cattle with BSE cause disease in cattle, sheep, mice, pigs, and mink after intracerebral inoculation, but prions in brain extracts from sheep with scrapie fed to cattle produced an illness substantially different from BSE. However, no exhaustive effort has been made to test different strains of sheep prions or to examine the disease following bovine-to-bovine passage.

15.3 Monitoring Cattle for BSE Prions

Although no reliable, specific test for prion disease in live animals is available, immunoassays for PrP^{Sc} in the brainstems of cattle currently provide the best available approach for preventing BSE prions from entering the human food supply. Determining how early in the incubation period PrP^{Sc} can be detected by immunological methods is now possible because a reliable bioassay is now available with the creation of Tg(BoPrP)$Prnp^{0/0}$ mice. Prior to the construction of Tg(BoPrP)$Prnp^{0/0}$ mice, cattle were used for bioassays of bovine prions. In a very limited study using cattle bioassays, bovine prions were undetectable in the obex of the brainstem until 26 months after oral inoculation. In these studies, prion infectivity was detected much earlier in the lymphoid tissue of the distal ileum.

Prior to development of Tg(BoPrP) $Prnp^{0/0}$ mice, non-Tg mice inoculated intracerebrally with BSE brain extracts required more than 300 days to develop disease. Generally, many of the inoculated mice fail to develop prion disease. It is noteworthy that a highly stable, mouse prion strain denoted 301V was derived by passage of BSE inocula into non-Tg $Prnp^{b/b}$ mice.

Both Western blots and ELISA tests are currently used in the postmortem evaluation of bovine brainstems. As configured, these tests are quite insensitive. The CDI, whose sensitivity is superior to conventional Western blots and ELISAs, offers a new approach to measuring PrP^{Sc} in tissues. The sensitivity of the CDI for BSE prions was found to be similar to that of bioassays using Tg(BoPrP)$Prnp^{0/0}$ mice. These studies suggest that the CDI may be able to detect BSE prions in asymptomatic cattle even when titers are quite low.

15.4 Compelling Evidence of Transmission of Bovine Prions to Humans

The restricted geographical occurrence and chronology of vCJD raise the possibility that BSE prions have been transmitted to humans. That only ~150 vCJD cases have been recorded and the incidence has remained relatively constant made establishing the origin of vCJD difficult. No set of dietary habits distinguishes vCJD patients from apparently healthy people. Moreover, there is no explanation for the predilection of vCJD for teenagers and young adults. Why have older individuals

not developed vCJD-based neuropathologic criteria? It is noteworthy that epidemiologic studies over the past three decades have failed to find evidence of transmission of sheep prions to humans. Attempts to predict the future number of cases of vCJD, assuming exposure to bovine prions prior to the offal ban, have been uninformative because so few cases of vCJD have occurred. Are we at the beginning of a prion disease epidemic in Britain, as seen for BSE and kuru, or will the number of vCJD cases remain small, as seen with iCJD caused by cadaveric HGH?

Recent studies of PrP^{Sc} from brains of patients who died of vCJD show a PrP glycoform pattern different from those found for sCJD and iCJD. The utility of measuring PrP glycoforms is questionable in trying to relate BSE to vCJD because PrP^{Sc} is formed after the protein is glycosylated and enzymatic deglycosylation of PrP^{Sc} requires denaturation. Moreover, synthetic prions have been formed from unglycosylated recMoPrP(89–230), demonstrating that glycosylation is unnecessary for prion formation.

Prior to the investigations recorded below laboratory studies were unconvincing in establishing a relationship between the conformations of PrP^{Sc} from cattle with BSE and those from humans with vCJD. Perhaps slightly more persuasive is the relationship demonstrated between vCJD and BSE based on similar incubation times in non-Tg RIII mice, of ~310 days after inoculation with human vCJD or bovine BSE prions. But such studies suffer from transmission of both BSE and vCJD prions to a heterologous host, that is, non-Tg mice expressing $MoPrP^{C}$. Using Tg mice, as was done for strains generated in the brains of patients with FFI or fCJD, compelling evidence for the transmission of bovine prions to humans was found. BSE prions transmitted to all inoculated Tg(BoPrP) mice after ~240 days but not to Tg mice expressing either HuPrP or MHu2MPrP. On second passage to Tg(BoPrP)$Prnp^{0/0}$ mice, the incubation time was unaltered, demonstrating the complete absence of a species barrier. Similar to BSE prions, vCJD prions transmitted readily to Tg(BoPrP)$Prnp^{0/0}$ mice, with a slightly longer incubation time of ~270 days, but poorly to Tg(HuPrP) and Tg(MHu2M) mice. Moreover, sCJD, iCJD, and fCJD(E200K) prions failed to transmit disease to Tg(BoPrP) mice after more than 500 days. On second passage of vCJD prions to Tg(BoPrP)$Prnp^{0/0}$ mice, the incubation time was reduced to ~225 days, demonstrating a small but expected species barrier. These findings argue that the strain-specific PrP^{Sc} conformations from BSE and vCJD prions are quite similar despite substantial differences in the amino acid sequences of BoPrP and HuPrP. Clearly, the conformation of $HuPrP^{Sc}$(vCJD) makes these prions much more readily transmittable to Tg(BoPrP)$Prnp^{0/0}$ mice than to either Tg(HuPrP)$Prnp^{0/0}$ or Tg(MHu2M)$Prnp^{0/0}$ mice.

15.5 Chronic Wasting Disease

Mule deer, white-tailed deer, and elk have been reported to develop CWD, which is unique among the prion diseases because it seems to be far more communicable than scrapie, BSE, or TME; moreover, it is the only prion disease known in free-ranging animals. CWD was first described in 1967 and reported to be a spongiform encephalopathy in 1978 based on histopathology in the brain. CWD has been found in the United States, Canada, and South Korea. In the United States,

CWD has been reported in Colorado, Wyoming, South Dakota, Nebraska, Oklahoma, Montana, New Mexico, Minnesota, and Wisconsin. In captive cervid herds, up to 90% of mule deer have been reported to be positive for prions and up to 60% of elk in Colorado and Wyoming develop CWD. Moreover, the incidence of CWD in cervids living in the wild has been estimated to be as high as 15%. The mode of transmission of CWD prion among mule deer, white-tailed deer, and elk is unresolved but contamination of grass with prions excreted in fecal matter seems to be a likely source. The high content of PrP^{Sc} in the intestinal lymphoid tissue of cervids with CWD supports such a scenario.

Brain homogenates from mule deer with CWD have transmitted disease to 4 of 13 cattle after intracerebral inoculation. These findings are particularly important because there is great concern that CWD prions might be transmitted to cattle grazing in contaminated pastures. In addition, CWD has been transmitted to ferrets, mink, squirrel monkeys, goats, and mice after intracerebral inoculation; however, only mule deer demonstrate efficient transmission of CWD prions by intracerebral inoculation. To date, endpoint titrations of CWD prions have not been performed in mule deer and Tg mice susceptible to CWD prions have not been developed, but the CDI has been adapted to measure CWD prions rapidly.

16 Fungal Prions

Although prions were originally defined in the context of an infectious pathogen, it is now becoming widely accepted that prions are elements that impart and propagate variability through multiple conformers of a normal, cellular protein. It is likely that such a mechanism will not be restricted to a single class of transmissible pathogens. Indeed, it is probable that this original definition will need to be extended to encompass other situations in which a similar mechanism of information transfer occurs. Two notable prion-like determinants, [*URE3*] and [*PSI*], have been described in yeast and another prion-like determinant has been reported in other fungi.

16.1 Yeast Prion Domains

In both [*PSI*] and [*URE3*], the "functional" determinants have been mapped to the C-terminal region of the protein, distinct from the "prion" domain, which comprises the N-terminal residues 65 and 114 of Ure2p and Sup35p, respectively. Although neither of the prion domains displays sequence identity either to each other or to PrP, the N-terminal regions of both Sup35p and mammalian PrP contain short repeated sequence elements: PQGGYQQYN in Sup35p and PHGGGWGQ in PrP. Interestingly, when the prion domains of both Sup35 and Ure2p are expressed in *E. coli* and purified, they polymerize spontaneously into amyloid-like fibrils. Polymerization of the prion domains of both Sup35 and Ure2p are able to create [*PSI*] and [*URE3*] states in yeast, respectively.

16.2 Dependence of Yeast Prions on Molecular Chaperones

The intrinsic power of the yeast genetic system has provided striking evidence of

the involvement of chaperones in the propagation of yeast [*PSI*] "prions." A genetic screen for factors that suppress the [*PSI*] phenotype resulted in the isolation of a single suppressor plasmid, which was found to contain the chaperone Hsp104. Furthermore, propagation of [*PSI*] was eliminated by either overproduction or absence of Hsp104, and treatment of cells with guanidine or UV light led to induction of Hsp104. The significance of Hsp104 is unclear because no published data exist to indicate that [*URE3*] utilizes Hsp104; furthermore, overexpression of Sup35 at high levels can induce [*PSI*] in the absence of Hsp104.

16.3 Differences between Yeast and Mammalian Prions

A wealth of studies has established the concept of fungal prions. While fungal prions are produced in the cytoplasm, mammalian prions are produced within cholesterol-rich microdomains on the surface of cells called *rafts* or *CLDs*. Whether the process of fungal and mammalian prion replication are fundamentally different or quite similar remains to be established. It seems likely that many more similarities than differences will emerge because distinct strains of both fungal and mammalian prions seem to represent different conformations of a particular protein. This finding necessitates a mechanism by which a particular conformation can be templated and reproduced with a high degree of fidelity.

Interestingly, the prion state in yeast is proposed to be functionally inert in the case of both [*PSI*] and [*URE3*] and produces the same phenotype as inactivation of the maintenance gene. In contrast, prion diseases in mammals cannot be explained simply by the loss of function of PrP because ablation of the PrP gene does not as yet have a detectable deleterious effect.

17 Prion Diseases are Disorders of Protein Conformation

The study of prions has followed several unexpected directions over the past three decades. The discovery that prion diseases in humans are uniquely genetic and infectious has greatly strengthened and extended the prion concept. To date, more than 30 different mutations in the human PrP gene, all resulting in nonconservative substitutions have been found either to be linked genetically to or to segregate with the inherited prion diseases (Fig. 11).

Understanding how PrP^{C} unfolds and refolds into PrP^{Sc} will be of paramount importance in transferring advances in the prion diseases to studies of other degenerative illnesses. The mechanism by which PrP^{Sc} is formed must involve a templating process in which existing PrP^{Sc} directs the refolding of PrP^{C} into nascent PrP^{Sc} with the same conformation. Undoubtedly, molecular chaperones of some type participate in a process that appears to be associated with CLDs on the cell surface.

Studies of prions in fungi have been extremely helpful in establishing the prion concept. In the case of yeast, Sup35 and Ure2p fold into alternative conformations that create the new metabolic states [*PSI*] and [*URE3*], respectively. Whether PrP^{Sc} in mammals represents a misfolded protein or an alternatively folded protein, as with yeast prions, remains to be established.

18 Prevention and Therapeutics for Prion Diseases

Because people at risk for inherited prion diseases can now be identified decades before neurologic dysfunction is evident, the development of an effective therapy for these fully penetrant disorders is imperative. Although it is difficult to predict the number of individuals who may develop neurologic dysfunction from bovine prions in the future, seeking an effective therapy now seems most prudent. Interfering with the conversion of PrP^{C} into PrP^{Sc} would seem to be the most attractive therapeutic target.

Defining the pathogenesis of prion disease is an important issue with respect to developing an effective therapy. The issue of whether large aggregates of misprocessed proteins or misfolded monomers (or oligomers) cause CNS degeneration has been addressed in several studies of prion diseases in humans as well as in Tg mice. In humans, the frequency of PrP amyloid plaques varies from 100% in GSS and vCJD to ~70% in kuru and ~10% in sCJD, arguing that these plaques are a nonobligatory feature of the disease. In Tg mice expressing both MoPrP and SHaPrP, animals inoculated with hamster prions produced hamster prions and developed amyloid plaques composed of $SHaPrP^{Sc}$. In contrast, Tg mice inoculated with mouse prions did not develop plaques even though they produced mouse prions and died of prion disease.

18.1 Prion Therapeutics

Various compounds have been proposed as potential therapeutics for treatment of prion diseases; these include polysulfated anions, dextrans, Congo red dye, oligonucleotides, and cyclic tetrapyrroles, all of which have been shown to increase survival time when given prior to prion infection in rodents, but not when administered a month or more after infection has been established.

Besides studies in rodents, ScN2a cells chronically infected with scrapie prions have been used to identify several candidate antiprion drugs but none have been shown to be effective in halting prion diseases in either animals or humans. Structure-based drug design based on dominant-negative inhibition of prion formation has produced several lead compounds. Prion replication depends on protein–protein interactions and a subset of these interactions gives rise to dominant-negative phenotypes produced by single residue substitutions. A particularly interesting set of drugs is the branched polyamines, or dendrimers, which enhance the clearance of PrP^{Sc} from cells. While these compounds cure cultured cells of prion infection, they have not been successfully deployed in mice because of difficulties in delivering such highly charged compounds to the CNS.

18.1.1 Quinacrine and other Acridine Derivatives

Tricyclic derivatives of acridine exhibit half-maximal inhibition of PrP^{Sc} formation at effective concentrations (IC_{50}) between 0.3 μM and 3 μM in cultured ScN2a cells. The IC_{50} for chlorpromazine was 3 μM while quinacrine was 10 times more potent. A variety of 9-substituted, acridine-based analogs of quinacrine were synthesized, which demonstrated variable potencies similar to chlorpromazine and emphasized the importance of the side chain in mediating the inhibition of PrP^{Sc}

formation. These studies showed that tricyclic compounds with an aliphatic side chain at the middle ring moiety constitute a new class of antiprion agents. Because quinacrine and chlorpromazine have been used for many years in humans as antimalarial and antipsychotic drugs, respectively, and are known to penetrate the blood-brain barrier (BBB), these compounds became immediate candidates for the treatment of CJD and other human prion diseases.

An asymmetric carbon in the side chain of quinacrine creates two stereoisomers. (*S*)-Quinacrine is two to three times more potent than the (*R*)-quinacrine enantiomer in cultured cells. Whether the use of (*S*)-quinacrine in place of the racemic mixture will allow significantly more drug to be given remains to be established. Studies performed almost six decades ago in humans demonstrated that (*S*)-quinacrine is selectively metabolized and that (*R*)-quinacrine remains intact. Whether (*R*)-quinacrine or the metabolites of (*S*)-quinacrine are responsible for the side effects recorded in patients receiving the racemic mixture is unknown. Side effects of quinacrine include liver toxicity, cardiomyopathy, and toxic psychoses.

18.1.2 Anti-PrP Antibodies

A panel of recombinant antibody fragments (recFabs), recognizing different epitopes on PrP, was studied with respect to inhibition of prion propagation in cultured ScN2a cells. Recombinant Fabs binding to PrP^C on the cell surface inhibited PrP^{Sc} formation in a dose-dependent manner. In ScN2a cells treated with the most potent recFab D18, prion replication was completely abolished and preexisting PrP^{Sc} rapidly cleared, suggesting that this antibody may cure established infection. The activity of recFab D18 is associated with its ability to recognize more completely the total population of PrP^C molecules on the cell surface than other recFabs and with the location of its epitope on PrP^C. In other studies, a monoclonal antibody, pH4, which is thought to bind to the same region of PrP as recFab D18, was found to inhibit prion accumulation in ScN2a cells.

Whether antibodies or Fabs can be effectively administered to humans for the prevention and treatment of prion diseases is unclear. Neither antibodies nor Fabs cross the BBB in high concentration so that delivery of these proteins to the CNS remains a critical issue.

It is notable that anti-PrP antibodies delivered by transgenetics or by injection were successful in blocking prion replication in mice inoculated intraperitoneally. The same anti-PrP antibodies were ineffective in treating prion infection initiated by intracerebral inoculation.

18.2 Dominant-negative Sheep

The production of domestic animals that do not replicate prions may also prove to be a practical way to prevent prion disease. Sheep encoding Arg/Arg at polymorphic position 171 seem resistant to scrapie; presumably, this was the genetic basis of Parry's scrapie eradication program in Great Britain 30 years ago. A more effective approach applying dominant-negative inhibition for producing prion-resistant domestic animals, including sheep and cattle, is probably the expression of PrP transgenes encoding K219 and/or R171 (Fig. 11). Such an approach has been evaluated in Tg mice and shown to be effective. In fact, the replacement of all sheep with Arg/Arg at codon 171 is underway in the Netherlands. Such an approach can

be instituted by artificial insemination using sperm from males homozygous for Arg/Arg at residue 171.

In comparison, a less practical approach is the production of PrP-deficient cattle and sheep. Although such animals would not be susceptible to prion disease, they might suffer some deleterious effects from the ablation of the PrP gene.

19 Some Principles of Prion Biology

Many principles of prion replication are clearly unprecedented in biology. As such, it is not surprising that some of these principles have been slow to be embraced. Although prion replication resembles viral replication superficially, the underlying principles are quite different. For example, in prion replication, the substrate is a host-encoded protein, PrP^{C}, which undergoes modification to form PrP^{Sc}, the only known component of the infectious prion particle. In contrast, viruses carry a DNA or RNA genome that is copied and directs the synthesis of most, if not all, of the viral proteins. The mature virion consists of a nucleic acid genome surrounded by a protein coat, whereas a prion appears to be composed of a dimer of PrP^{Sc}. When viruses pass from one species to another, they often replicate without any structural modification, whereas prions undergo a profound change. The prion adopts a new PrP sequence, which is encoded by the PrP gene of the current host. That change in amino acid sequence can result in a restriction of transmission for some species while making the new prion permissive to others. In viruses, different properties exhibited by distinct strains are encoded in the viral genome, whereas in prions, strain-specific properties are enciphered in the conformation of PrP^{Sc}.

19.1 Implications for Common Neurodegenerative Diseases

Understanding how PrP^{C} unfolds and refolds into PrP^{Sc} may also open new approaches to deciphering the causes of and to developing effective therapies for some common neurodegenerative diseases, including Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis (ALS). Whether or not therapies designed to prevent the conversion of PrP^{C} into PrP^{Sc} will be effective in these more common neurodegenerative diseases is unknown. Alternatively, developing a therapy for the prion diseases might provide a blueprint for designing somewhat different drugs for these common disorders. Like the inherited prion diseases, subsets of Alzheimer's disease and ALS are caused by mutations that result in nonconservative amino acid substitutions in proteins expressed in the CNS.

As knowledge about prions continues to expand, our understanding of how prions replicate and cause disease will undoubtedly evolve. It is important to add that many of the basic principles of prion biology are becoming increasingly well understood.

See also Alzheimer's Disease; Motor Neuron Diseases: Cellular and Animal Models; Motor Neuron Diseases: Molecular Mechanism, Pathophysiology, and Treatments.

Bibliography

Books and Reviews

Aguzzi, A., Heppner, F.L., Heikenwalder, M., Prinz, M., Mertz, K., Seeger, H., Glatzel, M. (2003) Immune system and peripheral nerves in propagation of prions to CNS, *Br. Med. Bull.* **66**, 141–159.

Alpers, M.P. (1968) Kuru: Implications of its Transmissibility for the Interpretation of its Changing Epidemiological Pattern, in: Bailey, O.T., Smith, D.E. (Eds.) *The Central Nervous System: Some Experimental Models of Neurological Diseases*, Williams and Wilkins Company, Baltimore, MD, pp. 234–251.

Behrens, A. (2003) Physiological and pathological functions of the prion protein homologue Dpl., *Br. Med. Bull.* **66**, 35–42.

Brandner, S. (2003) CNS pathogenesis of prion diseases, *Br. Med. Bull.* **66**, 131–139.

Bruce, M., Chree, A., McConnell, I., Foster, J., Pearson, G., Fraser, H. (1994) Transmission of bovine spongiform encephalopathy and scrapie to mice: strain variation and the species barrier, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **343**, 405–411.

Bruce, M.E. (1996) Strain Typing Studies of Scrapie and BSE, in: Baker, H.F., Ridley, R.M. (Eds.) *Methods in Molecular Medicine: Prion Diseases*, Humana Press, Totowa, NJ, pp. 223–236.

Bruce, M.E. (2003) TSE strain variation: An investigation into prion disease diversity, *Br. Med. Bull.* **66**, 99–108.

Budka, H. (2003) Neuropathology of prion diseases, *Br. Med. Bull.* **66**, 121–130.

Caughey, B. (2003) Prion protein conversions: insight into mechanisms, TSE transmission barriers and strains, *Br. Med. Bull.* **66**, 109–120.

Cervenáková, L., Brown, P., Piccardo, P., Cummings, J.L., Nagle, J., Vinters, H.V., Kaur, P., Ghetti, B., Chapman, J., Gajdusek, D.C., Goldfarb, L.G. (1996) 24-Nucleotide Deletion in the *PRNP* Gene: Analysis of Associated Phenotypes, in: Court, L., Dodet, B. (Eds.) *Transmissible Subacute Spongiform Encephalopathies: Prion Diseases*, Elsevier, Paris, France, pp. 433–444.

Chesebro, B. (2003) Introduction to the transmissible spongiform encephalopathies or prion diseases, *Br. Med. Bull.* **66**, 1–20.

Chien, P., Weissman, J.S., DePace, A.H. (2004) Emerging principles of conformation-based prion inheritance, *Annu. Rev. Biochem.* **73**, 617–656.

Collinge, J., Palmer, M.S. (1997) Human Prion Diseases, in: Collinge, J., Palmer, M.S. (Eds.) *Prion Diseases*, Oxford University Press, Oxford, UK, pp. 18–56.

DeArmond, S.J., Prusiner, S.B. (1997) Prion Diseases, in: Lantos, P., Graham, D. (Eds.) *Greenfield's Neuropathology*, 6th edition, Edward Arnold, London, UK, pp. 235–280.

Dickinson, A.G., Outram, G.W. (1988) Genetic Aspects of Unconventional Virus Infections: the Basis of the Virino Hypothesis, in: Bock, G., Marsh, J. (Eds.) *Novel Infectious Agents and the Central Nervous System. Ciba Foundation Symposium 135*, John Wiley and Sons, Chichester, UK, pp. 63–83.

Diringer, H. (1991) Transmissible spongiform encephalopathies (TSE) virus-induced amyloidoses of the central nervous system (CNS), *Eur. J. Epidemiol.* **7**, 562–566.

Dormont, D. (2003) Approaches to prophylaxis and therapy: An investigation into prion disease diversity, *Br. Med. Bull.* **66**, 281–292.

Erdtmann, R., Sivitz, L.B. (Eds.) (2004) *Advancing Prion Science: Guidance for the National Prion Research Program*, National Academies Press, Washington, DC

Gambetti, P., Kong, Q., Zou, W., Parchi, P., Chen, S.G. (2003) Sporadic and familial CJD: classification and characterisation, *Br. Med. Bull.* **66**, 213–239.

Gambetti, P., Parchi, P., Petersen, R.B., Chen, S.G., Lugaresi, E. (1995) Fatal familial insomnia and familial Creutzfeldt-Jakob disease: clinical, pathological and molecular features, *Brain Pathol.* **5**, 43–51.

Goodman, L.S., Gilman, A. (Eds.) (1970) *The Pharmacological Basis of Therapeutics; A Textbook of Pharmacology, Toxicology, and Therapeutics for Physicians and Medical Students*, Macmillan, New York.

Harris, D.A. (2003) Trafficking, turnover and membrane topology of PrP: protein function in prion disease, *Br. Med. Bull.* **66**, 71–85.

Harris, D.A. (Ed.) (2004) *Mad Cow Disease and Related Spongiform Encephalopathies*, Current Topics in Microbiology and Immunology, Springer-Verlag, Berlin, Germany.

Hill, A.F., Collinge, J. (2003) Subclinical prion infection in humans and animals, *Br. Med. Bull.* **66**, 161–170.

Hunter, N. (2003) Scrapie and experimental BSE in sheep, *Br. Med. Bull.* **66**, 171–183.

Kimberlin, R.H. (1996) Speculations on the Origin of BSE and the Epidemiology of CJD, in: Gibbs Jr., C.J. (Ed.) *Bovine Spongiform Encephalopathy: The BSE Dilemma*, Springer Verlag, New York, pp. 155–175.

Kübler, E., Oesch, B., Raeber, A.J. (2003) Diagnosis of prion diseases, *Br. Med. Bull.* **66**, 267–279.

Lasmézas, C.I. (2003) Putative functions of PrP^C, *Br. Med. Bull.* **66**, 61–70.

Malmgren, R., Kurland, L., Mokri, B., Kurtzke, J. (1979) The Epidemiology of Creutzfeldt-Jakob Disease, in: Prusiner, S.B. Hadlow, W.J. (Eds.) *Slow Transmissible Diseases of the Nervous System*, Vol. 1, Academic Press, New York, pp. 93–112.

Parry, H.B. (1983) *Scrapie Disease in Sheep*, Academic Press, New York.

Pattison, I.H. (1965) Experiments with Scrapie with Special Reference to the Nature of the Agent and the Pathology of the Disease, in: Gajdusek, D.C. Gibbs Jr., C.J., Alpers, M.P. (Eds.) *Slow, Latent and Temperate Virus Infections, NINDB Monograph 2*, U.S. Government Printing, Washington, DC, pp. 249–257.

Prusiner, S.B. (1989) Scrapie prions, *Annu. Rev. Microbiol.* **43**, 345–374.

Prusiner, S.B. (1991) Molecular biology of prion diseases, *Science* **252**, 1515–1522.

Prusiner, S.B. (1997) Prion diseases and the BSE crisis, *Science* **278**, 245–251.

Prusiner, S.B. (Ed.) (2004) *Prion Biology and Diseases*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Prusiner, S.B., Scott, M.R., DeArmond, S.J., Cohen, F.E. (1998) Prion protein biology, *Cell* **93**, 337–348.

Prusiner, S.B., Baron, H., Carlson, G., Cohen, F.E., DeArmond, S.J., Gabizon, R., Gambetti, P., Hope, J., Kitamoto, T., Kretzschmar, H.A., Laplanche, J.-L., Tateishi, J., Telling, G., Will, R. (2000) Prions, in: van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R., Wickner, R.B. et al., *Virus Taxonomy – Classification and Nomenclature of Viruses*, Academic Press, San Diego, CA, 1032–1039.

Riesner, D. (2003) Biochemistry and structure of PrP^C and PrP^{Sc}, *Br. Med. Bull.* **66**, 21–33.

Sigurdson, C.J., Miller, M.W. (2003) Other animal prion diseases, *Br. Med. Bull.* **66**, 199–212.

Smith, P.G., Bradley, R. (2003) Bovine spongiform encephalopathy (BSE) and its epidemiology, *Br. Med. Bull.* **66**, 185–198.

Solassol, J., Crozet, C., Lehmann, S. (2003) Prion propagation in cultured cells, *Br. Med. Bull.* **66**, 87–97.

Tabrizi, S.J., Elliott, C.L., Weissmann, C. (2003) Ethical issues in human prion diseases, *Br. Med. Bull.* **66**, 305–316.

Tateishi, J., Kitamoto, T. (1993) Developments in diagnosis for prion diseases, *Br. Med. Bull.* **49**, 971–979.

Taylor, D.M. (2003) Preventing accidental transmission of human transmissible spongiform encephalopathies, *Br. Med. Bull.* **66**, 293–303.

Tuite, M.F., Lindquist, S.L. (1996) Maintenance and inheritance of yeast prions, *Trends Genet.* **12**, 467–471.

Wadsworth, J.D., Hill, A.F., Beck, J.A., Collinge, J. (2003) Molecular and clinical classification of human prion disease, *Br. Med. Bull.* **66**, 241–254.

Weissmann, C. (1991) A "unified theory" of prion propagation, *Nature* **352**, 679–683.

Weissmann, C., Flechsig, E. (2003) PrP knockout and PrP transgenic mice in prion research, *Br. Med. Bull.* **66**, 43–60.

Wells, G.A.H., Wilesmith, J.W. (1995) The neuropathology and epidemiology of bovine spongiform encephalopathy, *Brain Pathol.* **5**, 91–103.

Wickner, R.B., Liebman, S.W., Saupe, S.J. (2004) Prions of Yeast and Filamentous Fungi: [URE3], [PSI^+], [PIN^+], and [Het-s], in: Prusiner, S.B. (Ed.) *Prion Biology and Diseases*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 305–372.

Will, R.G. (2003) Acquired prion disease: iatrogenic CJD, variant CJD, kuru, *Br. Med. Bull.* **66**, 255–265.

Will, R.G., Alpers, M.P., Dormont, D., Schonberger, L.B. (2004) Infectious and Sporadic Prion Diseases, in: Prusiner, S.B. (Ed.) *Prion Biology and Diseases*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 629–671.

Williams, E.S., Miller, M.W. (2002) Chronic wasting disease in deer and elk in North America, *Rev. Sci. Tech.* **21**, 305–316.

Primary Literature

Anderson, R.M., Donnelly, C.A., Ferguson, N.M., Woolhouse, M.E.J., Watt, C.J., Udy, H.J., MaWhinney, S., Dunstan, S.P., Southwood, T.R.E., Wilesmith, J.W., Ryan, J.B.M., Hoinville, L.J., Hillerton, J.E., Austin, A.R., Wells, G.A.H. (1996) Transmission dynamics and epidemiology of BSE in British cattle, *Nature* **382**, 779–788.

Anfinsen, C.B. (1973) Principles that govern the folding of protein chains, *Science* **181**, 223–230.

Asante, E.A., Linehan, J.M., Desbruslais, M., Joiner, S., Gowland, I., Wood, A.L., Welch, J., Hill, A.F., Lloyd, S.E., Wadsworth, J.D., Collinge, J. (2002) BSE prions propagate as either variant CJD-like or sporadic CJD-like prion strains in transgenic mice expressing human prion protein, *EMBO J.* **21**, 6358–6366.

Baker, H.F., Ridley, R.M., Wells, G.A.H. (1993) Experimental transmission of BSE and scrapie to the common marmoset, *Vet. Rec.* **132**, 403–406.

Barry, R.A., Prusiner, S.B. (1986) Monoclonal antibodies to the cellular and scrapie prion proteins, *J. Infect. Dis.* **154**, 518–521.

Basler, K., Oesch, B., Scott, M., Westaway, D., Wälchli, M., Groth, D.F., McKinley, M.P., Prusiner, S.B., Weissmann, C. (1986) Scrapie and cellular PrP isoforms are encoded by the same chromosomal gene, *Cell* **46**, 417–428.

Bastian, F.O., Foster, J.W. (2001) Spiroplasma sp. 16S rDNA in Creutzfeldt-Jakob disease and scrapie as shown by PCR and DNA sequence analysis, *J. Neuropathol. Exp. Neurol.* **60**, 613–620.

Bateman, D., Hilton, D., Love, S., Zeidler, M., Beck, J., Collinge, J. (1995) Sporadic Creutzfeldt-Jakob disease in a 18-year-old in the UK (Lett.), *Lancet* **346**, 1155–1156.

Behrens, A., Genoud, N., Naumann, H., Rülicke, T., Janett, F., Heppner, F.L., Ledermann, B., Aguzzi, A. (2002) Absence of the prion protein homologue Doppel causes male sterility, *EMBO J.* **21**, 3652–3658.

Bellinger-Kawahara, C.G., Kempner, E., Groth, D.F., Gabizon, R., Prusiner, S.B. (1988) Scrapie prion liposomes and rods exhibit target sizes of 55 000 Da, *Virology* **164**, 537–541.

Belt, P.B., Muileman, I.H., Schreuder, B.E.C., Ruijter, J.B., Gielkens, A.L.J., Smits, M.A. (1995) Identification of five allelic variants of the sheep PrP gene and their association with natural scrapie, *J. Gen. Virol.* **76**, 509–517.

Bertoni, J.M., Brown, P., Goldfarb, L., Gajdusek, D., Omaha, N.E. (1992) Familial Creutzfeldt-Jakob disease with the PRNP codon 200^{lys} mutation and supranuclear palsy but without myoclonus or periodic EEG complexes, *Neurology* **42**(4), Suppl. 3), 350 [Abstr].

Bessen, R.A., Marsh, R.F. (1994) Distinct PrP properties suggest the molecular basis of strain variation in transmissible mink encephalopathy, *J. Virol.* **68**, 7859–7868.

Billette de Villemeur, T., Deslys, J.-P., Pradel, A., Soubrié, C., Alpérovitch, A., Tardieu, M., Chaussain, J.-L., Hauw, J.-J., Dormont, D., Ruberg, M., Agid, Y. (1996) Creutzfeldt-Jakob disease from contaminated growth hormone extracts in France, *Neurology* **47**, 690–695.

Borchelt, D.R., Scott, M., Taraboulos, A., Stahl, N., Prusiner, S.B. (1990) Scrapie and cellular prion proteins differ in their kinetics of synthesis and topology in cultured cells, *J. Cell Biol.* **110**, 743–752.

Britton, T.C., Al-Sarraj, S., Shaw, C., Campbell, T., Collinge, J. (1995) Sporadic Creutzfeldt-Jakob disease in a 16-year-old in the UK (Lett.), *Lancet* **346**, 1155.

Brown, D.R., Qin, K., Herms, J.W., Madlung, A., Manson, J., Strome, R., Fraser, P.E., Kruck, T., von Bohlen, A., Schulz-Schaeffer, W., Giese, A., Westaway, D., Kretzschmar, H. (1997) The cellular prion protein binds copper *in vivo*, *Nature* **390**, 684–687.

Brown, P., Cathala, F., Raubertas, R.F., Gajdusek, D.C., Castaigne, P. (1987) The epidemiology of Creutzfeldt-Jakob disease: conclusion of a 15-year investigation in France and review of the world literature, *Neurology* **37**, 895–904.

Bruce, M., Chree, A., McConnell, I., Foster, J., Fraser, H. (1993) Transmissions of BSE, scrapie and related diseases to mice (Abstr.), *Proceedings of the IXth International Congress of Virology*, Glasgow, Scotland, Aug. 9–12, 93.

Bruce, M.E., Dickinson, A.G. (1987) Biological evidence that the scrapie agent has an independent genome, *J. Gen. Virol.* **68**, 79–89.

Bruce, M.E., McConnell, I., Fraser, H., Dickinson, A.G. (1991) The disease characteristics of different strains of scrapie in *Sinc* congenic mouse lines: implications for the nature of the agent and host control of pathogenesis, *J. Gen. Virol.* **72**, 595–603.

Bruce, M.E., Will, R.G., Ironside, J.W., McConnell, I., Drummond, D., Suttie, A., McCardle, L., Chree, A., Hope, J., Birkett, C., Cousens, S., Fraser, H., Bostock, C.J. (1997) Transmissions to mice indicate that 'new variant' CJD is caused by the BSE agent, *Nature* **389**, 498–501.

Büeler, H., Raeber, A., Sailer, A., Fischer, M., Aguzzi, A., Weissmann, C. (1994) High prion and PrP^{Sc} levels but delayed onset of disease in scrapie-inoculated mice heterozygous for a disrupted PrP gene, *Mol. Med.* **1**, 19–30.

Büeler, H., Aguzzi, A., Sailer, A., Greiner, R.-A., Autenried, P., Aguet, M., Weissmann, C. (1993) Mice devoid of PrP are resistant to scrapie, *Cell* **73**, 1339–1347.

Büeler, H., Fisher, M., Lang, Y., Bluethmann, H., Lipp, H.-P., DeArmond, S.J., Prusiner, S.B., Aguet, M., Weissmann, C. (1992) Normal development and behaviour of mice lacking the neuronal cell-surface PrP protein, *Nature* **356**, 577–582.

Burns, C.S., Aronoff-Spencer, E., Legname, G., Prusiner, S.B., Antholine, W.E., Gerfen, G.J., Peisach, J., Millhauser, G.L. (2003) Copper coordination in the full-length, recombinant prion protein, *Biochemistry* **42**, 6794–6803.

Buschmann, A., Pfaff, E., Reifenberg, K., Müller, H.M., Groschup, M.H. (2000) Detection of cattle-derived BSE prions using transgenic mice overexpressing bovine PrP^{C}, *Arch. Virol. [Suppl]*. **16**, 75–86.

Carlson, G.A., Westaway, D., DeArmond, S.J., Peterson-Torchia, M., Prusiner, S.B. (1989) Primary structure of prion protein may modify scrapie isolate properties, *Proc. Natl. Acad. Sci. U. S. A.* **86**, 7475–7479.

Carlson, G.A., Kingsbury, D.T., Goodman, P.A., Coleman, S., Marshall, S.T., DeArmond, S., Westaway, D., Prusiner, S.B. (1986) Linkage of prion protein and scrapie incubation-time genes, *Cell* **46**, 503–511.

Carlson, G.A., Ebeling, C., Yang, S.-L., Telling, G., Torchia, M., Groth, D., Westaway, D., DeArmond, S.J., Prusiner, S.B. (1994) Prion isolate specified allotypic interactions between the cellular and scrapie prion proteins in congenic and transgenic mice, *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5690–5694.

Caughey, B., Race, R.E. (1992) Potent inhibition of scrapie-associated PrP accumulation by Congo red, *J. Neurochem.* **59**, 768–771.

Caughey, B., Raymond, G.J. (1991) The scrapie-associated form of PrP is made from a cell-surface precursor that is both protease- and phospholipase-sensitive, *J. Biol. Chem.* **266**, 18217–18223.

Caughey, B., Kocisko, D.A., Raymond, G.J., Lansbury Jr., P.T. (1995) Aggregates of scrapie-associated prion protein induce the cell-free conversion of protease-sensitive prion protein to the protease-resistant state, *Chem. Biol.* **2**, 807–817.

Caughey, B., Raymond, G.J., Ernst, D., Race, R.E. (1991) N-terminal truncation of the scrapie-associated form of PrP by lysosomal protease(s): implications regarding the site of conversion of PrP to the protease-resistant state, *J. Virol.* **65**, 6597–6603.

Caughey, B., Neary, K., Butler, R., Ernst, D., Perry, L., Chesebro, B., Race, R.E. (1990) Normal and scrapie-associated forms of prion protein differ in their sensitivities to phospholipase and proteases in intact neuroblastoma cells, *J. Virol.* **64**, 1093–1101.

Caughey, B.W., Dong, A., Bhat, K.S., Ernst, D., Hayes, S.F., Caughey, W.S. (1991) Secondary structure analysis of the scrapie-associated protein PrP 27–30 in water by infrared spectroscopy, *Biochemistry* **30**, 7672–7680.

Centers for Disease Control (1996) Surveillance for Creutzfeldt-Jakob disease – United States, *MMWR Morb. Mortal. Wkly. Rep.* **45**, 665–668.

Chapman, J., Ben-Israel, J., Goldhammer, Y., Korczyn, A.D. (1994) The risk of developing Creutzfeldt-Jakob disease in subjects with the *PRNP* gene codon 200 point mutation, *Neurology* **44**, 1683–1686.

Chernoff, Y.O., Lindquist, S.L., Ono, B., Inge-Vechtomov, S.G., Liebman, S.W. (1995) Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi^{+}], *Science* **268**, 880–884.

Chesebro, B. (1998) Prion diseases: BSE and prions: uncertainties about the agent, *Science* **279**, 42–43.

Chesebro, B., Race, R., Wehrly, K., Nishio, J., Bloom, M., Lechner, D., Bergstrom, S., Robbins, K., Mayer, L., Keith, J.M., Garon, C., Haase, A. (1985) Identification of scrapie prion protein-specific mRNA in scrapie-infected and uninfected brain, *Nature* **315**, 331–333.

Clousard, C., Beaudry, P., Elsen, J.M., Milan, D., Dussaucy, M., Bounneau, C., Schelcher, F., Chatelain, J., Launay, J.-M., Laplanche, J.-L. (1995) Different allelic effects of the codons 136 and 171 of the prion protein gene in

sheep with natural scrapie, *J. Gen. Virol.* **76**, 2097–2101.

Cohen, F.E., Pan, K.-M., Huang, Z., Baldwin, M., Fletterick, R.J., Prusiner, S.B. (1994) Structural clues to prion replication, *Science* **264**, 530–531.

Collinge, J., Palmer, M.S., Dryden, A.J. (1991) Genetic predisposition to iatrogenic Creutzfeldt-Jakob disease, *Lancet* **337**, 1441–1442.

Collinge, J., Sidle, K.C.L., Meads, J., Ironside, J., Hill, A.F. (1996) Molecular analysis of prion strain variation and the aetiology of "new variant" CJD, *Nature* **383**, 685–690.

Collinge, J., Whittington, M.A., Sidle, K.C., Smith, C.J., Palmer, M.S., Clarke, A.R., Jefferys, J.G.R. (1994) Prion protein is necessary for normal synaptic function, *Nature* **370**, 295–297.

Collinge, J., Palmer, M.S., Sidle, K.C., Hill, A.F., Gowland, I., Meads, J., Asante, E., Bradley, R., Doey, L.J., Lantos, P.L. (1995) Unaltered susceptibility to BSE in transgenic mice expressing human prion protein, *Nature* **378**, 779–783.

Cousens, S.N., Vynnycky, E., Zeidler, M., Will, R.G., Smith, P.G. (1997) Predicting the CJD epidemic in humans, *Nature* **385**, 197–198.

Cousens, S.N., Harries-Jones, R., Knight, R., Will, R.G., Smith, P.G., Matthews, W.B. (1990) Geographical distribution of cases of Creutzfeldt-Jakob disease in England and Wales 1970–84, *J. Neurol. Neurosurg. Psychiatry* **53**, 459–465.

Dawson, M., Wells, G.A.H., Parker, B.N.J. (1990) Preliminary evidence of the experimental transmissibility of bovine spongiform encephalopathy to cattle, *Vet. Rec.* **126**, 112–113.

Dawson, M., Wells, G.A.H., Parker, B.N.J., Scott, A.C. (1990) Primary parenteral transmission of bovine spongiform encephalopathy to the pig, *Vet. Rec.* **127**, 338.

DeArmond, S.J., McKinley, M.P., Barry, R.A., Braunfeld, M.B., McColloch, J.R., Prusiner, S.B. (1985) Identification of prion amyloid filaments in scrapie-infected brain, *Cell* **41**, 221–235.

Derkatch, I.L., Chernoff, Y.O., Kushnirov, V.V., Inge-Vechtomov, S.G., Liebman, S.W. (1996) Genesis and variability of [*PSI*] prion factors in *Saccharomyces cerevisiae*, *Genetics* **144**, 1375–1386.

Dickinson, A.G., Fraser, H., Outram, G.W. (1975) Scrapie incubation time can exceed natural lifespan, *Nature* **256**, 732–733.

Dickinson, A.G., Meikle, V.M.H., Fraser, H. (1968) Identification of a gene which controls the incubation period of some strains of scrapie agent in mice, *J. Comp. Pathol.* **78**, 293–299.

Dickinson, A.G., Meikle, V.M., Fraser, H. (1969) Genetical control of the concentration of ME7 scrapie agent in the brain of mice, *J. Comp. Pathol.* **79**, 15–22.

Dickinson, A.G., Young, G.B., Stamp, J.T., Renwick, C.C. (1965) An analysis of natural scrapie in Suffolk sheep, *Heredity* **20**, 485–503.

Doel, S.M., McCready, S.J., Nierras, C.R., Cox, B.S. (1994) The dominant *PNM2*$^-$ mutation which eliminates the ψ factor of *Saccharomyces cerevisiae* is the result of a missense mutation in the *SUP35* gene, *Genetics* **137**, 659–670.

Doh-ura, K., Iwaki, T., Caughey, B. (2000) Lysosomotropic agents and cysteine protease inhibitors inhibit scrapie-associated prion protein accumulation, *J. Virol.* **74**, 4894–4897.

Doh-ura, K., Kitamoto, T., Sakaki, Y., Tateishi, J. (1991) CJD discrepancy, *Nature* **353**, 801–802.

Doh-ura, K., Tateishi, J., Sasaki, H., Kitamoto, T., Sakaki, Y. (1989) Pro→Leu change at position 102 of prion protein is the most common but not the sole mutation related to Gerstmann-Sträussler syndrome, *Biochem. Biophys. Res. Commun.* **163**, 974–979.

Donne, D.G., Viles, J.H., Groth, D., Mehlhorn, I., James, T.L., Cohen, F.E., Prusiner, S.B., Wright, P.E., Dyson, H.J. (1997) Structure of the recombinant full-length hamster prion protein PrP(29–231): the N terminus is highly flexible, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 13452–13457.

Edenhofer, F., Rieger, R., Famulok, M., Wendler, W., Weiss, S., Winnacker, E.-L. (1996) Prion protein PrP^C interacts with molecular chaperones of the Hsp60 family, *J. Virol.* **70**, 4724–4728.

Ehlers, B., Diringer, H. (1984) Dextran sulphate 500 delays and prevents mouse scrapie by impairment of agent replication in spleen, *J. Gen. Virol.* **65**, 1325–1330.

Enari, M., Flechsig, E., Weissmann, C. (2001) Scrapie prion protein accumulation by scrapie-infected neuroblastoma cells abrogated by exposure to a prion protein antibody, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9295–9299.

Endo, T., Groth, D., Prusiner, S.B., Kobata, A. (1989) Diversity of oligosaccharide structures linked to asparagines of the scrapie prion protein, *Biochemistry* **28**, 8380–8388.

Fink, J.K., Peacock, M.L., Warren, J.T., Roses, A.D., Prusiner, S.B. (1994) Detecting prion protein gene mutations by denaturing gradient gel electrophoresis, *Hum. Mutat.* **4**, 42–50.

Fischer, M., Rülicke, T., Raeber, A., Sailer, A., Moser, M., Oesch, B., Brandner, S., Aguzzi, A., Weissmann, C. (1996) Prion protein (PrP) with amino-proximal deletions restoring susceptibility of PrP knockout mice to scrapie, *EMBO J.* **15**, 1255–1264.

Foster, J., Goldmann, W., Parnham, D., Chong, A., Hunter, N. (2001) Partial dissociation of PrP^{Sc} deposition and vacuolation in the brains of scrapie and BSE experimentally affected goats, *J. Gen. Virol.* **82**, 267–273.

Fraser, H., Dickinson, A.G. (1973) Scrapie in mice. Agent-strain differences in the distribution and intensity of grey matter vacuolation, *J. Comp. Pathol.* **83**, 29–40.

Fraser, H., McConnell, I., Wells, G.A.H., Dawson, M. (1988) Transmission of bovine spongiform encephalopathy to mice, *Vet. Rec.* **123**, 472.

Fraser, H., Bruce, M.E., Chree, A., McConnell, I., Wells, G.A.H. (1992) Transmission of bovine spongiform encephalopathy and scrapie to mice, *J. Gen. Virol.* **73**, 1891–1897.

Furukawa, H., Kitamoto, T., Tanaka, Y., Tateishi, J. (1995) New variant prion protein in a Japanese family with Gerstmann-Sträussler syndrome, *Mol. Brain Res.* **30**, 385–388.

Gabizon, R., Telling, G., Meiner, Z., Halimi, M., Kahana, I., Prusiner, S.B. (1996) Insoluble wild-type and protease-resistant mutant prion protein in brains of patients with inherited prion disease, *Nat. Med.* **2**, 59–64.

Gabizon, R., Rosenmann, H., Meiner, Z., Kahana, I., Kahana, E., Shugart, Y., Ott, J., Prusiner, S.B. (1993) Mutation and polymorphism of the prion protein gene in Libyan Jews with Creutzfeldt-Jakob disease (CJD), *Am. J. Hum. Genet.* **53**, 828–835.

Gabriel, J.-M., Oesch, B., Kretzschmar, H., Scott, M., Prusiner, S.B. (1992) Molecular cloning of a candidate chicken prion protein, *Proc. Natl. Acad. Sci. U. S. A.* **89**, 9097–9101.

Gajdusek, D.C. (1977) Unconventional viruses and the origin and disappearance of kuru, *Science* **197**, 943–960.

Gajdusek, D.C., Gibbs Jr., C.J., Asher, D.M., Brown, P., Diwan, A., Hoffman, P., Nemo, G., Rohwer, R., White, L. (1977) Precautions in medical care of, and in handling materials from, patients with transmissible virus dementia (Creutzfeldt-Jakob disease), *N. Engl. J. Med.* **297**, 1253–1258.

Gambetti, P., Parchi, P. (1999) Insomnia in prion diseases: sporadic and familial, *N. Engl. J. Med.* **340**(21), 1675–1677.

Gasset, M., Baldwin, M.A., Fletterick, R.J., Prusiner, S.B. (1993) Perturbation of the secondary structure of the scrapie prion protein under conditions that alter infectivity, *Proc. Natl. Acad. Sci. U. S. A.* **90**, 1–5.

Gauczynski, S., Peyrin, J.M., Haik, S., Leucht, C., Hundt, C., Rieger, R., Krasemann, S., Deslys, J.P., Dormont, D., Lasmézas, C.I., Weiss, S. (2001) The 37-kDa/67-kDa laminin receptor acts as the cell-surface receptor for the cellular prion protein, *EMBO J.* **20**, 5863–5875.

Ghani, A.C., Ferguson, N.M., Donnelly, C.A., Anderson, R.M. (2000) Predicted vCJD mortality in Great Britain, *Nature* **406**, 583–584.

Glasse, R.M. (1967) Cannibalism in the kuru region of New Guinea, *Trans. NY Acad. Sci. [Ser. 2]* **29**, 748–754.

Glover, J.R., Kowal, A.S., Schirmer, E.C., Patino, M.M., Liu, J.-J., Lindquist, S. (1997) Self-seeded fibers formed by Sup35, the protein determinant of [PSI^+], a heritable prion-like factor of S. cerevisiae, *Cell* **89**, 811–819.

Goldfarb, L., Brown, P., Goldgaber, D., Garruto, R., Yanaghiara, R., Asher, D., Gajdusek, D.C. (1990) Identical mutation in unrelated patients with Creutzfeldt-Jakob disease, *Lancet* **336**, 174–175.

Goldfarb, L.G., Mitrova, E., Brown, P., Toh, B.H., Gajdusek, D.C. (1990) Mutation in codon 200 of scrapie amyloid protein gene in two clusters of Creutzfeldt-Jakob disease in Slovakia, *Lancet* **336**, 514–515.

Goldfarb, L.G., Haltia, M., Brown, P., Nieto, A., Kovanen, J., McCombie, W.R., Trapp, S., Gajdusek, D.C. (1991) New mutation in scrapie amyloid precursor gene (at codon 178) in Finnish Creutzfeldt-Jakob kindred, *Lancet* **337**, 425.

Goldfarb, L.G., Brown, P., Goldgaber, D., Asher, D.M., Rubenstein, R., Brown, W.T., Piccardo, P., Kascsak, R.J., Boellaard, J.W., Gajdusek, D.C. (1990) Creutzfeldt-Jakob disease and kuru patients lack a mutation consistently

found in the Gerstmann-Sträussler-Scheinker syndrome, *Exp. Neurol.* **108**, 247–250.

Goldfarb, L.G., Brown, P., McCombie, W.R., Goldgaber, D., Swergold, G.D., Wills, P.R., Cervenakova, L., Baron, H., Gibbs, C.J.J., Gajdusek, D.C. (1991) Transmissible familial Creutzfeldt-Jakob disease associated with five, seven, and eight extra octapeptide coding repeats in the *PRNP* gene, *Proc. Natl. Acad. Sci. U. S. A.* **88**, 10926–10930.

Goldfarb, L.G., Petersen, R.B., Tabaton, M., Brown, P., LeBlanc, A.C., Montagna, P., Cortelli, P., Julien, J., Vital, C., Pendelbury, W.W. (1992) Fatal familial insomnia and familial Creutzfeldt-Jakob disease: disease phenotype determined by a DNA polymorphism, *Science* **258**(5083), 806–808.

Goldfarb, L.G., Brown, P., Mitrova, E., Cervenakova, L., Goldin, L., Korczyn, A.D., Chapman, J., Galvez, S., Cartier, L., Rubenstein, R., Gajdusek, D.C. (1991) Creutzfeldt-Jacob disease associated with the *PRNP* codon 200^{Lys} mutation: an analysis of 45 families, *Eur. J. Epidemiol.* **7**, 477–486.

Goldgaber, D., Goldfarb, L.G., Brown, P., Asher, D.M., Brown, W.T., Lin, S., Teener, J.W., Feinstone, S.M., Rubenstein, R., Kascsak, R.J., Boellaard, J.W., Gajdusek, D.C. (1989) Mutations in familial Creutzfeldt-Jakob disease and Gerstmann-Sträussler-Scheinker's syndrome, *Exp. Neurol.* **106**, 204–206.

Goldmann, W., Hunter, N., Manson, J. Hope, J. (1990) The PrP gene of the sheep, a natural host of scrapie, *Proceedings of the VIIIth International Congress of Virology*, Berlin, Germany, Aug. 26–31, 284.

Goldmann, W., Hunter, N., Benson, G., Foster, J.D., Hope, J. (1991) Different scrapie-associated fibril proteins (PrP) are encoded by lines of sheep selected for different alleles of the *Sip* gene, *J. Gen. Virol.* **72**, 2411–2417.

Goldmann, W., Hunter, N., Martin, T., Dawson, M., Hope, J. (1991) Different forms of the bovine PrP gene have five or six copies of a short, G-C-rich element within the protein-coding exon, *J. Gen. Virol.* **72**, 201–204.

Goldmann, W., Hunter, N., Smith, G., Foster, J., Hope, J. (1994) PrP genotype and agent effects in scrapie: change in allelic interaction with different isolates of agent in sheep, a natural host of scrapie, *J. Gen. Virol.* **75**, 989–995.

Goldmann, W., Hunter, N., Foster, J.D., Salbaum, J.M., Beyreuther, K., Hope, J. (1990) Two alleles of a neural protein gene linked to scrapie in sheep, *Proc. Natl. Acad. Sci. U. S. A.* **87**, 2476–2480.

Gorodinsky, A., Harris, D.A. (1995) Glycolipid-anchored proteins in neuroblastoma cells form detergent-resistant complexes without caveolin, *J. Cell Biol.* **129**, 619–627.

Govaerts, C., Wille, H., Prusiner, S.B., Cohen, F.E. (2004) Evidence for assembly of prions with left-handed β-helices into trimers, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8342–8347.

Graner, E., Mercadante, A.F., Zanata, S.M., Forlenza, O.V., Cabral, A.L.B., Veiga, S.S., Juliano, M.A., Roesler, R., Walz, R., Mineti, A., Izquierdo, I., Martins, V.R., Brentani, R.R. (2000) Cellular prion protein binds laminin and mediates neuritogenesis, *Mol. Brain Res.* **76**, 85–92.

Grassi, J., Comoy, E., Simon, S., Creminon, C., Frobert, Y., Trapmann, S., Schimmel, H., Hawkins, S.A., Moynagh, J., Deslys, J.P., Wells, G.A. (2001) Rapid test for the preclinical postmortem diagnosis of BSE in central nervous system tissue, *Vet. Rec.* **149**, 577–582.

Hamir, A.N., Cutlip, R.C., Miller, J.M., Williams, E.S., Stack, M.J., Miller, M.W., O'Rourke, K.I., Chaplin, M.J. (2001) Preliminary findings on the experimental transmission of chronic wasting disease agent of mule deer to cattle, *J. Vet. Diagn. Invest.* **13**, 91–96.

Hammick, D.L., Chambers, W.E. (1945) Optical activity of excreted mepacrine, *Nature* **155**, 141.

Haraguchi, T., Fisher, S., Olofsson, S., Endo, T., Groth, D., Tarantino, A., Borchelt, D.R., Teplow, D., Hood, L., Burlingame, A., Lycke, E., Kobata, A., Prusiner, S.B. (1989) Asparagine-linked glycosylation of the scrapie and cellular prion proteins, *Arch. Biochem. Biophys.* **274**, 1–13.

Harries-Jones, R., Knight, R., Will, R.G., Cousens, S., Smith, P.G., Matthews, W.B. (1988) Creutzfeldt-Jakob disease in England and Wales, 1980–1984: a case-control study of potential risk factors, *J. Neurol. Neurosurg. Psychiatry* **51**, 1113–1119.

Harris, D.A., Falls, D.L., Walsh, W., Fischbach, G.D. (1989) Molecular cloning of an acetylcholine receptor-inducing protein, *Soc. Neurosci.* **15**, 70. 7.

Heppner, F.L., Musahl, C., Arrighi, I., Klein, M.A., Rülicke, T., Oesch, B., Zinkernagel, R.M., Kalinke, U., Aguzzi, A. (2001) Prevention of scrapie pathogenesis by transgenic expression of antiprion protein antibodies, *Science* **294**, 178–182.

Hill, A.F., Desbruslais, M., Joiner, S., Sidle, K.C.L., Gowland, I., Collinge, J., Doey, L.J., Lantos, P. (1997) The same prion strain causes vCJD and BSE, *Nature* **389**, 448–450.

Hilton, D.A., Ghani, A.C., Conyers, L., Edwards, P., McCardle, L., Ritchie, D., Penney, M., Hegazy, D., Ironside, J.W. (2004) Prevalence of lymphoreticular prion protein accumulation in UK tissue samples, *J. Pathol.* **203**, 733–739.

Hornshaw, M.P., McDermott, J.R., Candy, J.M. (1995) Copper binding to the N-terminal tandem repeat regions of mammalian and avian prion protein, *Biochem. Biophys. Res. Commun.* **207**, 621–629.

Hornshaw, M.P., McDermott, J.R., Candy, J.M., Lakey, J.H. (1995) Copper binding to the N-terminal tandem repeat region of mammalian and avian prion protein: structural studies using synthetic peptides, *Biochem. Biophys. Res. Commun.* **214**, 993–999.

Hsiao, K., Baker, H.F., Crow, T.J., Poulter, M., Owen, F., Terwilliger, J.D., Westaway, D., Ott, J., Prusiner, S.B. (1989) Linkage of a prion protein missense variant to Gerstmann-Sträussler syndrome, *Nature* **338**, 342–345.

Hsiao, K., Meiner, Z., Kahana, E., Cass, C., Kahana, I., Avrahami, D., Scarlato, G., Abramsky, O., Prusiner, S.B., Gabizon, R. (1991) Mutation of the prion protein in Libyan Jews with Creutzfeldt-Jakob disease, *N. Engl. J. Med.* **324**, 1091–1097.

Hsiao, K.K., Cass, C., Schellenberg, G.D., Bird, T., Devine-Gage, E., Wisniewski, H., Prusiner, S.B. (1991) A prion protein variant in a family with the telencephalic form of Gerstmann-Sträussler-Scheinker syndrome, *Neurology* **41**, 681–684.

Hsiao, K.K., Groth, D., Scott, M., Yang, S.-L., Serban, H., Rapp, D., Foster, D., Torchia, M., DeArmond, S.J., Prusiner, S.B. (1994) Serial transmission in rodents of neurodegeneration from transgenic mice expressing mutant prion protein, *Proc. Natl. Acad. Sci. U. S. A.* **91**, 9126–9130.

Huang, Z., Prusiner, S.B., Cohen, F.E. (1995) Scrapie prions: a three-dimensional model of an infectious fragment, *Fold. Des.* **1**, 13–19.

Hundt, C., Peyrin, J.M., Haik, S., Gauczynski, S., Leucht, C., Rieger, R., Riley, M.L., Deslys, J.P., Dormont, D., Lasmézas, C.I., Weiss, S. (2001) Identification of interaction domains of the prion protein with its 37-kDa/67-kDa laminin receptor, *EMBO J.* **20**, 5876–5886.

Hunter, N., Foster, J.D., Benson, G., Hope, J. (1991) Restriction fragment length polymorphisms of the scrapie-associated fibril protein (PrP) gene and their association with susceptiblity to natural scrapie in British sheep, *J. Gen. Virol.* **72**, 1287–1292.

Hunter, N., Goldmann, W., Smith, G., Hope, J. (1994) Frequencies of PrP gene variants in healthy cattle and cattle with BSE in Scotland, *Vet. Rec.* **135**, 400–403.

Hunter, N., Goldmann, W., Benson, G., Foster, J.D., Hope, J. (1993) Swaledale sheep affected by natural scrapie differ significantly in PrP genotype frequencies from healthy sheep and those selected for reduced incidence of scrapie, *J. Gen. Virol.* **74**, 1025–1031.

Hunter, N., Moore, L., Hosie, B.D., Dingwall, W.S., Greig, A. (1997) Association between natural scrapie and PrP genotype in a flock of Suffolk sheep in Scotland, *Vet. Rec.* **140**, 59–63.

Hunter, N., Cairns, D., Foster, J.D., Smith, G., Goldmann, W., Donnelly, K. (1997) Is scrapie solely a genetic disease? *Nature* **386**, 137.

Ikeda, T., Horiuchi, M., Ishiguro, N., Muramatsu, Y., Kai-Uwe, G.D., Shinagawa, M. (1995) Amino acid polymorphisms of PrP with reference to onset of scrapie in Suffolk and Corriedale sheep in Japan, *J. Gen. Virol.* **76**, 2577–2581.

Ingrosso, L., Ladogana, A., Pocchiari, M. (1995) Congo red prolongs the incubation period in scrapie-infected hamsters, *J. Virol.* **69**, 506–508.

Ironside, J.W. (1997) The new variant form of Creutzfeldt-Jakob disease: a novel prion protein amyloid disorder (Editorial), *Amyloid: Int. J. Exp. Clin. Invest.* **4**, 66–69.

James, T.L., Liu, H., Ulyanov, N.B., Farr-Jones, S., Zhang, H., Donne, D.G., Kaneko, K., Groth, D., Mehlhorn, I., Prusiner, S.B., Cohen, F.E. (1997) Solution structure of a 142-residue recombinant prion protein corresponding to the infectious fragment of the scrapie isoform, *Proc. Natl. Acad. Sci. U. S. A.* **94**(19), 10086–10091.

Jeffrey, M., Wells, G.A.H. (1988) Spongiform encephalopathy in a nyala *(Tragelaphus angasi)*, *Vet. Pathol.* **25**, 398–399.

Jendroska, K., Heinzel, F.P., Torchia, M., Stowring, L., Kretzschmar, H.A., Kon, A.,

Stern, A., Prusiner, S.B., DeArmond, S.J. (1991) Proteinase-resistant prion protein accumulation in Syrian hamster brain correlates with regional pathology and scrapie infectivity, *Neurology* **41**, 1482–1490.

Kahana, E., Milton, A., Braham, J., Sofer, D. (1974) Creutzfeldt-Jakob disease: focus among Libyan Jews in Israel, *Science* **183**, 90–91.

Kaneko, K., Zulianello, L., Scott, M., Cooper, C.M., Wallace, A.C., James, T.L., Cohen, F.E., Prusiner, S.B. (1997) Evidence for protein X binding to a discontinuous epitope on the cellular prion protein during scrapie prion propagation, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 10069–10074.

Kaneko, K., Ball, H.L., Wille, H., Zhang, H., Groth, D., Torchia, M., Tremblay, P., Safar, J., Prusiner, S.B., DeArmond, S.J., Baldwin, M.A., Cohen, F.E. (2000) A synthetic peptide initiates Gerstmann-Sträussler-Scheinker (GSS) disease in transgenic mice, *J. Mol. Biol.* **295**(4), 997–1007.

Kascsak, R.J., Rubenstein, R., Merz, P.A., Tonna-DeMasi, M., Fersko, R., Carp, R.I., Wisniewski, H.M., Diringer, H. (1987) Mouse polyclonal and monoclonal antibody to scrapie-associated fibril proteins, *J. Virol.* **61**, 3688–3693.

Kellings, K., Prusiner, S.B., Riesner, D. (1994) Nucleic acids in prion preparations: unspecific background or essential component? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **343**, 425–430.

Kellings, K., Meyer, N., Mirenda, C., Prusiner, S.B., Riesner, D. (1992) Further analysis of nucleic acids in purified scrapie prion preparations by improved return refocussing gel electrophoresis (RRGE), *J. Gen. Virol.* **73**, 1025–1029.

Keshet, G.I., Bar-Peled, O., Yaffe, D., Nudel, U., Gabizon, R. (2000) The cellular prion protein colocalizes with the dystroglycan complex in the brain, *J. Neurochem.* **75**, 1889–1897.

Kimberlin, R.H. (1982) Reflections on the nature of the scrapie agent, *Trends Biochem. Sci.* **7**, 392–394.

Kimberlin, R.H. (1990) Scrapie and possible relationships with viroids, *Semin. Virol.* **1**, 153–162.

Kimberlin, R.H., Walker, C.A. (1983) The antiviral compound HPA-23 can prevent scrapie when administered at the time of infection, *Arch. Virol.* **78**, 9–18.

Kimberlin, R.H., Collis, S.C., Walker, C.A. (1976) Profiles of brain glycosidase activity in cuprizone-fed Syrian hamsters and in scrapie-affected mice, rats, Chinese hamsters and Syrian hamsters, *J. Comp. Pathol.* **86**, 135–142.

Kimberlin, R.H., Field, H.J., Walker, C.A. (1983) Pathogenesis of mouse scrapie: evidence for spread of infection from central to peripheral nervous system, *J. Gen. Virol.* **64**, 713–716.

Kimberlin, R.H., Walker, C.A., Fraser, H. (1989) The genomic identity of different strains of mouse scrapie is expressed in hamsters and preserved on reisolation in mice, *J. Gen. Virol.* **70**, 2017–2025.

King, C.-Y., Tittman, P., Gross, H., Gebert, R., Aebi, M., Wüthrich, K. (1997) Prion-inducing domain 2–114 of yeast Sup35 protein transforms *in vitro* into amyloid-like filaments, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 6618–6622.

Kingsbury, D.T., Kasper, K.C., Stites, D.P., Watson, J.D., Hogan, R.N., Prusiner, S.B. (1983) Genetic control of scrapie and Creutzfeldt-Jakob disease in mice, *J. Immunol.* **131**, 491–496.

Kirkwood, J.K., Cunningham, A.A., Wells, G.A.H., Wilesmith, J.W., Barnett, J.E.F. (1993) Spongiform encephalopathy in a herd of greater kudu (*Tragelaphus strepsiceros*): epidemiological observations, *Vet. Rec.* **133**, 360–364.

Kitamoto, T., Tateishi, J. (1994) Human prion diseases with variant prion protein, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **343**, 391–398.

Klatzo, I., Gajdusek, D.C., Zigas, V. (1959) Pathology of kuru, *Lab. Invest.* **8**, 799–847.

Klitzman, R.L., Alpers, M.P., Gajdusek, D.C. (1984) The natural incubation period of kuru and the episodes of transmission in three clusters of patients, *Neuroepidemiology* **3**, 3–20.

Koch, T.K., Berg, B.O., DeArmond, S.J., Gravina, R.F. (1985) Creutzfeldt-Jakob disease in a young adult with idiopathic hypopituitarism. Possible relation to the administration of cadaveric human growth hormone, *N. Engl. J. Med.* **313**, 731–733.

Kocisko, D.A., Come, J.H., Priola, S.A., Chesebro, B., Raymond, G.J., Lansbury Jr., P.T., Caughey, B. (1994) Cell-free formation of protease-resistant prion protein, *Nature* **370**, 471–474.

Korth, C., May, B.C.H., Cohen, F.E., Prusiner, S.B. (2001) Acridine and phenothiazine derivatives as pharmacotherapeutics for prion disease, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9836–9841.

Korth, C., Kaneko, K., Groth, D., Heye, N., Telling, G., Mastrianni, J., Parchi, P., Gambetti, P., Will, R., Ironside, J., Heinrich, C., Tremblay, P., DeArmond, S.J., Prusiner, S.B. (2003) Abbreviated incubation times for human prions in mice expressing a chimeric mouse – human prion protein transgene, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 4784–4789.

Korth, C., Stierli, B., Streit, P., Moser, M., Schaller, O., Fischer, R., Schulz-Schaeffer, W., Kretzschmar, H., Raeber, A., Braun, U., Ehrensperger, F., Hornemann, S., Glockshuber, R., Riek, R., Billeter, M., Wuthrick, K., Oesch, B. (1997) Prion (PrP^{Sc})-specific epitope defined by a monoclonal antibody, *Nature* **389**, 74–77.

Kretzschmar, H.A., Neumann, M., Stavrou, D. (1995) Codon 178 mutation of the human prion protein gene in a German family (Backer family): sequencing data from 72 year-old celloidin-embedded brain tissue, *Acta Neuropathol.* **89**, 96–98.

Kretzschmar, H.A., Prusiner, S.B., Stowring, L.E., DeArmond, S.J. (1986) Scrapie prion proteins are synthesized in neurons, *Am. J. Pathol.* **122**, 1–5.

Kretzschmar, H.A., Honold, G., Seitelberger, F., Feucht, M., Wessely, P., Mehraein, P., Budka, H. (1991) Prion protein mutation in family first reported by Gerstmann, Sträussler, and Scheinker, *Lancet* **337**, 1160.

Kurschner, C., Morgan, J.I. (1995) The cellular prion protein (PrP) selectively binds to Bcl-2 in the yeast two-hybrid system, *Mol. Brain Res.* **30**, 165–168.

Laplanche, J.-L., Chatelain, J., Launay, J.-M., Gazengel, C., Vidaud, M. (1990) Deletion in prion protein gene in a Moroccan family, *Nucleic Acids. Res.* **18**, 6745.

Laplanche, J.-L., Chatelain, J., Beaudry, P., Dussaucy, M., Bounneau, C., Launay, J.-M. (1993) French autochthonous scrapied sheep without the 136Val PrP polymorphism, *Mamm. Genome* **4**, 463–464.

Lasmézas, C.I., Deslys, J.-P., Demaimay, R., Adjou, K.T., Hauw, J.-J., Dormont, D. (1996) Strain specific and common pathogenic events in murine models of scrapie and bovine spongiform encephalopathy, *J. Gen. Virol.* **77**, 1601–1609.

Lasmézas, C.I., Deslys, J.-P., Demaimay, R., Adjou, K.T., Lamoury, F., Dormont, D., Robain, O., Ironside, J., Hauw, J.-J. (1996) BSE transmission to macaques, *Nature* **381**, 743–744.

Lasmézas, C.I., Deslys, J.-P., Robain, O., Jaegly, A., Beringue, V., Peyrin, J.-M., Fournier, J.-G., Hauw, J.-J., Rossier, J., Dormont, D. (1997) Transmission of the BSE agent to mice in the absence of detectable abnormal prion protein, *Science* **275**(17), 402–405.

Legname, G., Baskakov, I.V., Nguyen, H.-O.B., Riesner, D., Cohen, F.E., DeArmond, S.J., Prusiner, S.B. (2004) Synthetic mammalian prions, *Science* **305**, 673–676.

Li, G., Bolton, D.C. (1997) A novel hamster prion protein mRNA contains an extra exon: increased expression in scrapie, *Brain Res.* **751**, 265–274.

Liu, H., Farr-Jones, S., Ulyanov, N.B., Llinas, M., Marqusee, S., Groth, D., Cohen, F.E., Prusiner, S.B., James, T.L. (1999) Solution structure of Syrian hamster prion protein rPrP(90–231), *Biochemistry* **38**(17), 5362–5377.

Lledo, P.-M., Tremblay, P., DeArmond, S.J., Prusiner, S.B., Nicoll, R.A. (1996) Mice deficient for prion protein exhibit normal neuronal excitability and synaptic transmission in the hippocampus, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 2403–2407.

Llewelyn, C.A., Hewitt, P.E., Knight, R.S., Amar, K., Cousens, S., Mackenzie, J., Will, R.G. (2004) Possible transmission of variant Creutzfeldt-Jakob disease by blood transfusion, *Lancet* **363**, 417–421.

Locht, C., Chesebro, B., Race, R., Keith, J.M. (1986) Molecular cloning and complete sequence of prion protein cDNA from mouse brain infected with the scrapie agent, *Proc. Natl. Acad. Sci. U. S. A.* **83**, 6372–6376.

Lugaresi, E., Medori, R., Montagna, P., Baruzzi, A., Cortelli, P., Lugaresi, A., Tinuper, P., Zucconi, M., Gambetti, P. (1986) Fatal familial insomnia and dysautonomia with selective degeneration of thalamic nuclei, *N. Engl. J. Med.* **315**, 997–1003.

Maddelein, M.L., Dos Reis, S., Duvezin-Caubet, S., Coulary-Salin, B., Saupe, S.J. (2002) Amyloid aggregates of the HET-s prion protein are infectious, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7402–7407.

Mallucci, G., Dickinson, A., Linehan, J., Klohn, P.C., Brandner, S., Collinge, J. (2003) Depleting neuronal PrP in prion infection prevents disease and reverses spongiosis, *Science* **302**, 871–874.

Manson, J.C., Clarke, A.R., Hooper, M.L., Aitchison, L., McConnell, I., Hope, J. (1994) 129/Ola

mice carrying a null mutation in PrP that abolishes mRNA production are developmentally normal, *Mol. Neurobiol.* **8**, 121–127.

Manson, J.C., Jameison, E., Baybutt, H., Tuzi, N.L., Barron, R., McConnell, I., Somerville, R., Ironside, J., Will, R., Sy, M.-S., Melton, D.W., Hope, J., Bostock, C. (1999) A single amino acid alteration (101L) introduced into murine PrP dramatically alters incubation time of transmissible spongiform encephalopathy, *EMBO J.* **18**, 6855–6864.

Marsh, R.F., Bessen, R.A., Lehmann, S., Hartsough, G.R. (1991) Epidemiological and experimental studies on a new incident of transmissible mink encephalopathy, *J. Gen. Virol.* **72**, 589–594.

Martins, V.R., Graner, E., Garcia-Abreu, J., de Souza, S.J., Mercadante, A.F., Veiga, S.S., Zanata, S.M., Neto, V.M., Brentani, R.R. (1997) Complementary hydropathy identifies a cellular prion protein receptor, *Nat. Med.* **3**, 1376–1382.

Masison, D.C., Wickner, R.B. (1995) Prion-inducing domain of yeast Ure2p and protease resistance of Ure2p in prion-containing cells, *Science* **270**, 93–95.

Masters, C.L., Gajdusek, D.C., Gibbs Jr., C.J. (1981) Creutzfeldt-Jakob disease virus isolations from the Gerstmann-Sträussler syndrome, *Brain* **104**, 559–588.

Masters, C.L., Harris, J.O., Gajdusek, D.C., Gibbs Jr., C.J., Bernouilli, C., Asher, D.M. (1978) Creutzfeldt-Jakob disease: patterns of worldwide occurrence and the significance of familial and sporadic clustering, *Ann. Neurol.* **5**, 177–188.

Mastrianni, J., Nixon, F., Layzer, R., DeArmond, S.J., Prusiner, S.B. (1997) Fatal sporadic insomnia: fatal familial insomnia phenotype without a mutation of the prion protein gene, *Neurology* **48** (Suppl.), A296.

Mastrianni, J.A., Nixon, R., Layzer, R., Telling, G.C., Han, D., DeArmond, S.J., Prusiner, S.B. (1999) Prion protein conformation in a patient with sporadic fatal insomnia, *N. Engl. J. Med.* **340**(21), 1630–1638.

McKinley, M.P., Meyer, R.K., Kenaga, L., Rahbar, F., Cotter, R., Serban, A., Prusiner, S.B. (1991) Scrapie prion rod formation *in vitro* requires both detergent extraction and limited proteolysis, *J. Virol.* **65**, 1340–1351.

McKnight, S., Tjian, R. (1986) Transcriptional selectivity of viral genes in mammalian cells, *Cell* **46**, 795–805.

Medori, R., Montagna, P., Tritschler, H.J., LeBlanc, A., Cortelli, P., Tinuper, P., Lugaresi, E., Gambetti, P. (1992) Fatal familial insomnia: a second kindred with mutation of prion protein gene at codon 178, *Neurology* **42**, 669–670.

Meggendorfer, F. (1930) Klinische und genealogische Beobachtungen bei einem Fall von spastischer Pseudosklerose Jakobs, *Z. Gesamte Neurol. Psychiatr.* **128**, 337–341.

Meyer, R.K., McKinley, M.P., Bowman, K.A., Braunfeld, M.B., Barry, R.A., Prusiner, S.B. (1986) Separation and properties of cellular and scrapie prion proteins, *Proc. Natl. Acad. Sci. U. S. A.* **83**, 2310–2314.

Miller, M.W., Wild, M.A., Williams, E.S. (1998) Epidemiology of chronic wasting disease in captive Rocky Mountain elk, *J. Wildl. Dis.* **34**, 532–538.

Miller, M.W., Williams, E.S., McCarty, C.W., Spraker, T.R., Kreeger, T.J., Larsen, C.T., Thorne, E.T. (2000) Epizootiology of chronic wasting disease in free-ranging cervids in Colorado and Wyoming, *J. Wildl. Dis.* **36**, 676–690.

Mills, J.L., Schonberger, L.B., Wysowski, D.K., Brown, P., Durako, S.J., Cox, C., Kong, F., Fradkin, J.E. (2004) Long-term mortality in the United States cohort of pituitary-derived growth hormone recipients, *J. Pediatr.* **144**, 430–436.

Minor, P., Newham, J., Jones, N., Bergeron, C., Gregori, L., Asher, D., Van Engelenburg, F., Stroebel, T., Vey, M., Barnard, G., Head, M. (2004) Standards for the assay of Creutzfeldt-Jakob disease specimens, *J. Gen. Virol.* **85**, 1777–1784.

Miyazono, M., Kitamoto, T., Doh-ura, K., Iwaki, T., Tateishi, J. (1992) Creutzfeldt-Jakob disease with codon 129 polymorphism (Valine): a comparative study of patients with codon 102 point mutation or without mutations, *Acta Neuropathol.* **84**, 349–354.

Mo, H., Moore, R.C., Cohen, F.E., Westaway, D., Prusiner, S.B., Wright, P.E., Dyson, H.J. (2001) Two different neurodegenerative diseases caused by proteins with similar structures, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 2352–2357.

Mobley, W.C., Neve, R.L., Prusiner, S.B., McKinley, M.P. (1988) Nerve growth factor increases mRNA levels for the prion protein and

the β-amyloid protein precursor in developing hamster brain, *Proc. Natl. Acad. Sci. U. S. A.* **85**, 9811–9815.

Moore, R.C., Hope, J., McBride, P.A., McConnell, I., Selfridge, J., Melton, D.W., Manson, J.C. (1998) Mice with gene targeted prion protein alterations show that *Prnp, Sinc* and *Prni* are congruent, *Nat. Genet.* **18**, 118–125.

Moore, R.C., Mastrangelo, P., Bouzamondo, E., Heinrich, C., Legname, G., Prusiner, S.B., Hood, L., Westaway, D., DeArmond, S.J., Tremblay, P. (2001) Doppel-induced cerebellar degeneration in transgenic mice, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 15288–15293.

Moore, R.C., Lee, I.Y., Silverman, G.L., Harrison, P.M., Strome, R., Heinrich, C., Karunaratne, A., Pasternak, S.H., Chishti, M.A., Liang, Y., Mastrangelo, P., Wang, K., Smit, A.F.A., Katamine, S., Carlson, G.A., Cohen, F.E., Prusiner, S.B., Melton, D.W., Tremblay, P., Hood, L.E., Westaway, D. (1999) Ataxia in prion protein (PrP)-deficient mice is associated with upregulation of the novel PrP-like protein doppel, *J. Mol. Biol.* **292**(4), 797–817.

Mouillet-Richard, S., Ermonval, M., Chebassier, C., Laplanche, J.-L., Lehmann, S., Launay, J.M., Kellermann, O. (2000) Signal transduction through prion protein, *Science* **289**, 1925–1928.

Nathanson, N., Wilesmith, J., Griot, C. (1997) Bovine spongiform encephalopathy (BSE): cause and consequences of a common-source epidemic, *Am. J. Epidemiol.* **145**, 959–969.

Nishida, N., Tremblay, P., Sugimoto, T., Shigematsu, K., Shirabe, S., Petromilli, C., Erpel, S.P., Nakaoke, R., Atarashi, R., Houtani, T., Torchia, M., Sakaguchi, S., DeArmond, S.J., Prusiner, S.B., Katamine, S. (1999) A mouse prion protein transgene rescues mice deficient for the prion protein gene from Purkinje cell degeneration and demyelination, *Lab. Invest.* **79**(6), 689–697.

Oesch, B., Teplow, D.B., Stahl, N., Serban, D., Hood, L.E., Prusiner, S.B. (1990) Identification of cellular proteins binding to the scrapie prion protein, *Biochemistry* **29**, 5848–5855.

Oesch, B., Westaway, D., Wälchli, M., McKinley, M.P., Kent, S.B.H., Aebersold, R., Barry, R.A., Tempst, P., Teplow, D.B., Hood, L.E., Prusiner, S.B., Weissmann, C. (1985) A cellular gene encodes scrapie PrP 27–30 protein, *Cell* **40**, 735–746.

O'Rourke, K.I., Holyoak, G.R., Clark, W.W., Mickelson, J.R., Wang, S., Melco, R.P., Besser, T.E., Foote, W.C. (1997) PrP genotypes and experimental scrapie in orally inoculated Suffolk sheep in the United States, *J. Gen. Virol.* **78**, 975–978.

O'Rourke, K.I., Besser, T.E., Miller, M.W., Cline, T.F., Spraker, T.R., Jenny, A.L., Wild, M.A., Zebarth, G.L., Williams, E.S. (1999) PrP genotypes of captive and free-ranging Rocky Mountain elk (Cervus elaphus nelsoni) with chronic wasting disease, *J. Gen. Virol.* **80**, 2765–2769.

Owen, F., Poulter, M., Collinge, J., Crow, T.J. (1990) Codon 129 changes in the prion protein gene in Caucasians, *Am. J. Hum. Genet.* **46**, 1215–1216.

Owen, F., Poulter, M., Lofthouse, R., Collinge, J., Crow, T.J., Risby, D., Baker, H.F., Ridley, R.M., Hsiao, K., Prusiner, S.B. (1989) Insertion in prion protein gene in familial Creutzfeldt-Jakob disease, *Lancet* **1**, 51–52.

Paisley, D., Banks, S., Selfridge, J., McLennan, N.F., Ritchie, A.M., McEwan, C., Irvine, D.S., Saunders, P.T., Manson, J.C., Melton, D.W. (2004) Male infertility and DNA damage in Doppel knockout and prion protein/Doppel double-knockout mice, *Am. J. Pathol.* **164**, 2279–2288.

Palmer, M.S., Dryden, A.J., Hughes, J.T., Collinge, J. (1991) Homozygous prion protein genotype predisposes to sporadic Creutzfeldt-Jakob disease, *Nature* **352**, 340–342.

Palmer, M.S., Mahal, S.P., Campbell, T.A., Hill, A.F., Sidle, K.C.L., Laplanche, J.-L., Collinge, J. (1993) Deletions in the prion protein gene are not associated with CJD, *Hum. Mol. Genet.* **2**, 541–544.

Pan, K.-M., Stahl, N., Prusiner, S.B. (1992) Purification and properties of the cellular prion protein from Syrian hamster brain, *Protein Sci.* **1**, 1343–1352.

Pan, K.-M., Baldwin, M., Nguyen, J., Gasset, M., Serban, A., Groth, D., Mehlhorn, I., Huang, Z., Fletterick, R.J., Cohen, F.E., Prusiner, S.B. (1993) Conversion of α-helices into β-sheets features in the formation of the scrapie prion proteins, *Proc. Natl. Acad. Sci. U. S. A.* **90**, 10962–10966.

Paramithiotis, E., Pinard, M., Lawton, T., LaBoissiere, S., Leathers, V.L., Zou, W.Q., Estey, L.A., Lamontagne, J., Lehto, M.T., Kondejewski, L.H., Francoeur, G.P., Papadopoulos, M., Haghighat, A., Spatz, S.J.,

Head, M., Will, R., Ironside, J., O'Rourke, K., Tonelli, Q., Ledebur, H.C., Chakrabartty, A., Cashman, N.R. (2003) A prion protein epitope selective for the pathologically misfolded conformation, *Nat. Med.* **9**, 893–899.

Parchi, P., Capellari, S., Chin, S., Schwarz, H.B., Schecter, N.P., Butts, J.D., Hudkins, P., Burns, D.K., Powers, J.M., Gambetti, P. (1999) A subtype of sporadic prion disease mimicking fatal familial insomnia, *Neurology* **52**, 1757–1763.

Parchi, P., Capellari, S., Chen, S.G., Petersen, R.B., Gambetti, P., Kopp, P., Brown, P., Kitamoto, T., Tateishi, J., Giese, A., Kretzschmar, H. (1997) Typing prion isoforms (Lett.), *Nature* **386**, 232–233.

Parry, H.B. (1962) Scrapie: a transmissible and hereditary disease of sheep, *Heredity* **17**, 75–105.

Patino, M.M., Liu, J.-J., Glover, J.R., Lindquist, S. (1996) Support for the prion hypothesis for inheritance of a phenotypic trait in yeast, *Science* **273**, 622–626.

Pattison, I.H., Millson, G.C. (1961) Scrapie produced experimentally in goats with special reference to the clinical syndrome, *J. Comp. Pathol.* **71**, 101–108.

Pattison, I.H., Jebbett, J.N. (1971) Clinical and histological observations on cuprizone toxicity and scrapie in mice, *Res. Vet. Sci.* **12**, 378–380.

Peden, A.H., Head, M.W., Ritchie, D.L., Bell, J.E., Ironside, J.W. (2004) Preclinical vCJD after blood transfusion in a *PRNP* codon 129 heterozygous patient, *Lancet* **364**, 527–529.

Peretz, D., Scott, M., Groth, D., Williamson, A., Burton, D., Cohen, F.E., Prusiner, S.B. (2001) Strain-specified relative conformational stability of the scrapie prion protein, *Protein Sci.* **10**, 854–863.

Peretz, D., Williamson, R.A., Legname, G., Matsunaga, Y., Vergara, J., Burton, D., DeArmond, S.J., Prusiner, S.B., Scott, M.R. (2002) A change in the conformation of prions accompanies the emergence of a new prion strain, *Neuron* **34**, 921–932.

Peretz, D., Williamson, R.A., Matsunaga, Y., Serban, H., Pinilla, C., Bastidas, R.B., Rozenshteyn, R., James, T.L., Houghten, R.A., Cohen, F.E., Prusiner, S.B., Burton, D.R. (1997) A conformational transition at the N-terminus of the prion protein features in formation of the scrapie isoform, *J. Mol. Biol.* **273**, 614–622.

Peretz, D., Williamson, R.A., Kaneko, K., Vergara, J., Leclerc, E., Schmitt-Ulms, G., Mehlhorn, I.R., Legname, G., Wormald, M.R., Rudd, P.M., Dwek, R.A., Burton, D.R., Prusiner, S.B. (2001) Antibodies inhibit prion propagation and clear cell cultures of prion infectivity, *Nature* **412**, 739–743.

Perrier, V., Wallace, A.C., Kaneko, K., Safar, J., Prusiner, S.B., Cohen, F.E. (2000) Mimicking dominant-negative inhibition of prion replication through structure-based drug design, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6073–6078.

Perrier, V., Kaneko, K., Safar, J., Vergara, J., Tremblay, P., DeArmond, S.J., Cohen, F.E., Prusiner, S.B., Wallace, A.C. (2002) Dominant-negative inhibition of prion replication in transgenic mice, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 13079–13084.

Peters, J., Miller, J.M., Jenny, A.L., Peterson, T.L., Carmichael, K.P. (2000) Immunohistochemical diagnosis of chronic wasting disease in preclinically affected elk from a captive herd, *J. Vet. Diagn. Invest.* **12**, 579–582.

Petersen, R.B., Tabaton, M., Berg, L., Schrank, B., Torack, R.M., Leal, S., Julien, J., Vital, C., Deleplanque, B., Pendlebury, W.W., Drachman, D., Smith, T.W., Martin, J.J., Oda, M., Montagna, P., Ott, J., Autilio-Gambetti, L., Lugaresi, E., Gambetti, P. (1992) Analysis of the prion protein gene in thalamic dementia, *Neurology* **42**, 1859–1863.

Phillips, N.A., Bridgeman, J. Ferguson-Smith, M. (2000) Findings & Conclusions. *The BSE Inquiry*, Vol. 1, Stationery Office, London, UK.

PHS (1997) Report on Human Growth Hormone and Creutzfeldt-Jakob Disease, Public Health Service Interagency Coordinating Committee. **14**: 1–11.

Poulter, M., Baker, H.F., Frith, C.D., Leach, M., Lofthouse, R., Ridley, R.M., Shah, T., Owen, F., Collinge, J., Brown, G., Hardy, J., Mullan, M.J., Harding, A.E., Bennett, C., Doshi, R., Crow, T.J. (1992) Inherited prion disease with 144 base pair gene insertion. 1. Genealogical and molecular studies, *Brain* **115**, 675–685.

Priola, S.A., Raines, A., Caughey, W.S. (2000) Porphyrin and phthalocyanine antiscrapie compounds, *Science* **287**, 1503–1506.

Proske, D., Gilch, S., Wopfner, F., Schätzl, H.M., Winnacker, E.L., Famulok, M. (2002) Prion-protein-specific aptamer reduces PrP^{Sc} formation, *Chembiochem* **3**, 717–725.

Prusiner, S.B. (1982) Novel proteinaceous infectious particles cause scrapie, *Science* **216**, 136–144.

Prusiner, S.B. (2004) Detecting mad cow disease, *Sci. Am.* **291**, 86–93.

Prusiner, S.B., Groth, D.F., Bolton, D.C., Kent, S.B., Hood, L.E. (1984) Purification and structural studies of a major scrapie prion protein, *Cell* **38**, 127–134.

Prusiner, S.B., Bolton, D.C., Groth, D.F., Bowman, K.A., Cochran, S.P., McKinley, M.P. (1982) Further purification and characterization of scrapie prions, *Biochemistry* **21**, 6942–6950.

Prusiner, S.B., McKinley, M.P., Bowman, K.A., Bolton, D.C., Bendheim, P.E., Groth, D.F., Glenner, G.G. (1983) Scrapie prions aggregate to form amyloid-like birefringent rods, *Cell* **35**, 349–358.

Prusiner, S.B., Groth, D., Serban, A., Koehler, R., Foster, D., Torchia, M., Burton, D., Yang, S.-L., DeArmond, S.J. (1993) Ablation of the prion protein (PrP) gene in mice prevents scrapie and facilitates production of anti-PrP antibodies, *Proc. Natl. Acad. Sci. U. S. A.* **90**, 10608–10612.

Prusiner, S.B., Fuzi, M., Scott, M., Serban, D., Serban, H., Taraboulos, A., Gabriel, J.-M., Wells, G., Wilesmith, J., Bradley, R., DeArmond, S.J., Kristensson, K. (1993) Immunologic and molecular biological studies of prion proteins in bovine spongiform encephalopathy, *J. Infect. Dis.* **167**, 602–613.

Prusiner, S.B., Scott, M., Foster, D., Pan, K.-M., Groth, D., Mirenda, C., Torchia, M., Yang, S.-L., Serban, D., Carlson, G.A., Hoppe, P.C., Westaway, D., DeArmond, S.J. (1990) Transgenetic studies implicate interactions between homologous PrP isoforms in scrapie prion replication, *Cell* **63**, 673–686.

Raymond, G.J., Hope, J., Kocisko, D.A., Priola, S.A., Raymond, L.D., Bossers, A., Ironside, J., Will, R.G., Chen, S.G., Petersen, R.B., Gambetti, P., Rubenstein, R., Smits, M.A., Lansbury Jr., P.T., Caughey, B. (1997) Molecular assessment of the potential transmissibilities of BSE and scrapie to humans, *Nature* **388**, 285–288.

Requena, J.R., Groth, D., Legname, G., Stadtman, E.R., Prusiner, S.B., Levine, R.L. (2001) Copper-catalyzed oxidation of the recombinant SHa(29–231) prion protein, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 7170–7175.

Rieger, R., Edenhofer, F., Lasmézas, C.I., Weiss, S. (1997) The human 37-kDa laminin receptor precursor interacts with the prion protein in eukaryotic cells, *Nat. Med.* **3**, 1383–1388.

Riek, R., Hornemann, S., Wider, G., Billeter, M., Glockshuber, R., Wüthrich, K. (1996) NMR structure of the mouse prion protein domain PrP(121–231), *Nature* **382**, 180–182.

Robinson, M.M., Hadlow, W.J., Knowles, D.P., Huff, T.P., Lacy, P.A., Marsh, R.F., Gorham, J.R. (1995) Experimental infection of cattle with the agents of transmissible mink encephalopathy and scrapie, *J. Comp. Path.* **113**, 241–251.

Rogers, M., Yehiely, F., Scott, M., Prusiner, S.B. (1993) Conversion of truncated and elongated prion proteins into the scrapie isoform in cultured cells, *Proc. Natl. Acad. Sci. U. S. A.* **90**, 3182–3186.

Rogers, M., Serban, D., Gyuris, T., Scott, M., Torchia, T., Prusiner, S.B. (1991) Epitope mapping of the Syrian hamster prion protein utilizing chimeric and mutant genes in a vaccinia virus expression system, *J. Immunol.* **147**, 3568–3574.

Rossi, D., Cozzio, A., Flechsig, E., Klein, M.A., Rülicke, T., Aguzzi, A., Weissmann, C. (2001) Onset of ataxia and Purkinje cell loss in PrP null mice inversely correlated with Dpl level in brain, *EMBO J.* **20**, 694–702.

Ryou, C., Legname, G., Peretz, D., Craig, J.C., Baldwin, M.A., Prusiner, S.B. (2003) Differential inhibition of prion propagation by enantiomers of quinacrine, *Lab. Invest.* **83**, 837–843.

Saeki, K., Matsumoto, Y., Hirota, Y., Matsumoto, Y., Onodera, T. (1996) Three-exon structure of the gene encoding the rat prion protein and its expression in tissues, *Virus Genes* **12**, 15–20.

Safar, J., Roller, P.P., Gajdusek, D.C., Gibbs, C.J.J. (1993) Thermal-stability and conformational transitions of scrapie amyloid (prion) protein correlate with infectivity, *Protein Sci.* **2**, 2206–2216.

Safar, J., Wille, H., Itri, V., Groth, D., Serban, H., Torchia, M., Cohen, F.E., Prusiner, S.B. (1998) Eight prion strains have PrP^{Sc} molecules with different conformations, *Nat. Med.* **4**(10), 1157–1165.

Safar, J.G., Scott, M., Monaghan, J., Deering, C., Didorenko, S., Vergara, J., Ball, H., Legname, G., Leclerc, E., Solforosi, L., Serban, H., Groth, D., Burton, D.R., Prusiner, S.B., Williamson, R.A. (2002) Measuring prions causing bovine spongiform encephalopathy

or chronic wasting disease by immunoassays and transgenic mice, *Nat. Biotechnol.* **20**, 1147–1150.

Sakaguchi, S., Katamine, S., Nishida, N., Moriuchi, R., Shigematsu, K., Sugimoto, T., Nakatani, A., Kataoka, Y., Houtani, T., Shirabe, S., Okada, H., Hasegawa, S., Miyamoto, T., Noda, T. (1996) Loss of cerebellar Purkinje cells in aged mice homozygous for a disrupted PrP gene, *Nature* **380**, 528–531.

Schätzl, H.M., Da Costa, M., Taylor, L., Cohen, F.E., Prusiner, S.B. (1995) Prion protein gene variation among primates, *J. Mol. Biol.* **245**, 362–374.

Schmitt-Ulms, G., Legname, G., Baldwin, M.A., Ball, H.L., Bradon, N., Bosque, P.J., Crossin, K.L., Edelman, G.M., DeArmond, S.J., Cohen, F.E., Prusiner, S.B. (2001) Binding of neural cell adhesion molecules (N-CAMs) to the cellular prion protein, *J. Mol. Biol.* **314**, 1209–1225.

Scott, M., Foster, D., Mirenda, C., Serban, D., Coufal, F., Wälchli, M., Torchia, M., Groth, D., Carlson, G., DeArmond, S.J., Westaway, D., Prusiner, S.B. (1989) Transgenic mice expressing hamster prion protein produce species-specific scrapie infectivity and amyloid plaques, *Cell* **59**, 847–857.

Scott, M., Groth, D., Foster, D., Torchia, M., Yang, S.-L., DeArmond, S.J., Prusiner, S.B. (1993) Propagation of prions with artificial properties in transgenic mice expressing chimeric PrP genes, *Cell* **73**, 979–988.

Scott, M.R., Köhler, R., Foster, D., Prusiner, S.B. (1992) Chimeric prion protein expression in cultured cells and transgenic mice, *Protein Sci.* **1**, 986–997.

Scott, M.R., Groth, D., Tatzelt, J., Torchia, M., Tremblay, P., DeArmond, S.J., Prusiner, S.B. (1997) Propagation of prion strains through specific conformers of the prion protein, *J. Virol.* **71**, 9032–9044.

Scott, M.R., Will, R., Ironside, J., Nguyen, H.-O.B., Tremblay, P., DeArmond, S.J., Prusiner, S.B. (1999) Compelling transgenetic evidence for transmission of bovine spongiform encephalopathy prions to humans, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 15137–15142.

Scott, M.R., Safar, J., Telling, G., Nguyen, O., Groth, D., Torchia, M., Koehler, R., Tremblay, P., Walther, D., Cohen, F.E., DeArmond, S.J., Prusiner, S.B. (1997) Identification of a prion protein epitope modulating transmission of bovine spongiform encephalopathy prions to transgenic mice, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 14279–14284.

Serban, D., Taraboulos, A., DeArmond, S.J., Prusiner, S.B. (1990) Rapid detection of Creutzfeldt-Jakob disease and scrapie prion proteins, *Neurology* **40**, 110–117.

Sethi, S., Lipford, G., Wagner, H., Kretzschmar, H. (2002) Postexposure prophylaxis against prion disease with a stimulator of innate immunity, *Lancet* **360**, 229–230.

Shibuya, S., Higuchi, J., Shin, R.-W., Tateishi, J., Kitamoto, T. (1998) Protective prion protein polymorphisms against sporadic Creutzfeldt-Jakob disease, *Lancet* **351**, 419.

Sigurdson, C.J., Williams, E.S., Miller, M.W., Spraker, T.R., O'Rourke, K.I., Hoover, E.A. (1999) Oral transmission and early lymphoid tropism of chronic wasting disease PrP^{res} in mule deer fawns (*Odocoileus hemionus*), *J. Gen. Virol.* **80**, 2757–2764.

Somerville, R.A., Chong, A., Mulqueen, O.U., Birkett, C.R., Wood, S.C., Hope, J. (1997) Biochemical typing of scrapie strains, *Nature* **386**, 564.

Sparkes, R.S., Simon, M., Cohn, V.H., Fournier, R.E.K., Lem, J., Klisak, I., Heinzmann, C., Blatt, C., Lucero, M., Mohandas, T., DeArmond, S.J., Westaway, D., Prusiner, S.B., Weiner, L.P. (1986) Assignment of the human and mouse prion protein genes to homologous chromosomes, *Proc. Natl. Acad. Sci. U. S. A.* **83**, 7358–7362.

Spencer, M.D., Knight, R.S., Will, R.G. (2002) First hundred cases of variant Creutzfeldt-Jakob disease: retrospective case note review of early psychiatric and neurological features, *Br. Med. J.* **324**, 1479–1482.

Spudich, S., Mastrianni, J.A., Wrensch, M., Gabizon, R., Meiner, Z., Kahana, I., Rosenmann, H., Kahana, E., Prusiner, S.B. (1995) Complete penetrance of Creutzfeldt-Jakob disease in Libyan Jews carrying the E200K mutation in the prion protein gene, *Mol. Med.* **1**, 607–613.

Stahl, N., Baldwin, M.A., Burlingame, A.L., Prusiner, S.B. (1990) Identification of glycoinositol phospholipid linked and

truncated forms of the scrapie prion protein, *Biochemistry* **29**, 8879–8884.

Stahl, N., Borchelt, D.R., Hsiao, K., Prusiner, S.B. (1987) Scrapie prion protein contains a phosphatidylinositol glycolipid, *Cell* **51**, 229–240.

Stahl, N., Baldwin, M.A., Teplow, D.B., Hood, L., Gibson, B.W., Burlingame, A.L., Prusiner, S.B. (1993) Structural analysis of the scrapie prion protein using mass spectrometry and amino acid sequencing, *Biochemistry* **32**, 1991–2002.

Stekel, D.J., Nowak, M.A., Southwood, T.R.E. (1996) Prediction of future BSE spread, *Nature* **381**, 119.

Stöckel, J., Safar, J., Wallace, A.C., Cohen, F.E., Prusiner, S.B. (1998) Prion protein selectively binds copper (II) ions, *Biochemistry* **37**(20), 7185–7193.

Sulkowski, E. (1985) Purification of proteins by IMAC, *Trends Biotechnol.* **3**, 1–7.

Supattapone, S., Nguyen, H.-O.B., Cohen, F.E., Prusiner, S.B., Scott, M.R. (1999) Elimination of prions by branched polyamines and implications for therapeutics, *Proc. Natl. Acad. Sci. U. S. A.* **96**(25), 14529–14534.

Supattapone, S., Wille, H., Uyechi, L., Safar, J., Tremblay, P., Szoka, F.C., Cohen, F.E., Prusiner, S.B., Scott, M.R. (2001) Branched polyamines cure prion-infected neuroblastoma cells, *J. Virol.* **75**, 3453–3461.

Supattapone, S., Bosque, P., Muramoto, T., Wille, H., Aagaard, C., Peretz, D., Nguyen, H.-O.B., Heinrich, C., Torchia, M., Safar, J., Cohen, F.E., DeArmond, S.J., Prusiner, S.B., Scott, M. (1999) Prion protein of 106 residues creates an artificial transmission barrier for prion replication in transgenic mice, *Cell* **96**, 869–878.

Taraboulos, A., Scott, M., Semenov, A., Avrahami, D., Laszlo, L., Prusiner, S.B. (1995) Cholesterol depletion and modification of COOH-terminal targeting sequence of the prion protein inhibits formation of the scrapie isoform, *J. Cell Biol.* **129**, 121–132.

Tatzelt, J., Maeda, N., Pekny, M., Yang, S.-L., Betsholtz, C., Eliasson, C., Cayetano, J., Camerino, A.P., DeArmond, S.J., Prusiner, S.B. (1996) Scrapie in mice deficient in apolipoprotein E or glial fibrillary acidic protein, *Neurology* **47**, 449–453.

Taylor, D.M., Fernie, K., Steele, P.J., McConnell, I., Somerville, R.A. (2002) Thermostability of mouse-passaged BSE and scrapie is independent of host PrP genotype: implications for the nature of the causal agents, *J. Gen. Virol.* **83**, 3199–3204.

Telling, G.C., Haga, T., Torchia, M., Tremblay, P., DeArmond, S.J., Prusiner, S.B. (1996) Interactions between wild-type and mutant prion proteins modulate neurodegeneration in transgenic mice, *Genes Dev.* **10**, 1736–1750.

Telling, G.C., Scott, M., Mastrianni, J., Gabizon, R., Torchia, M., Cohen, F.E., DeArmond, S.J., Prusiner, S.B. (1995) Prion propagation in mice expressing human and chimeric PrP transgenes implicates the interaction of cellular PrP with another protein, *Cell* **83**, 79–90.

Telling, G.C., Parchi, P., DeArmond, S.J., Cortelli, P., Montagna, P., Gabizon, R., Mastrianni, J., Lugaresi, E., Gambetti, P., Prusiner, S.B. (1996) Evidence for the conformation of the pathologic isoform of the prion protein enciphering and propagating prion diversity, *Science* **274**, 2079–2082.

Telling, G.C., Scott, M., Hsiao, K.K., Foster, D., Yang, S.-L., Torchia, M., Sidle, K.C.L., Collinge, J., DeArmond, S.J., Prusiner, S.B. (1994) Transmission of Creutzfeldt-Jakob disease from humans to transgenic mice expressing chimeric human-mouse prion protein, *Proc. Natl. Acad. Sci. U. S. A.* **91**, 9936–9940.

Ter-Avanesyan, M.D., Dagkesamanskaya, A.R., Kushnirov, V.V., Smirnov, V.N. (1994) The *SUP35* omnipotent suppressor gene is involved in the maintenance of the non-Mendelian determinant [*psi*$^+$] in the yeast *Saccharomyces cerevisiae*, *Genetics* **137**, 671–676.

Tobler, I., Gaus, S.E., Deboer, T., Achermann, P., Fischer, M., Rülicke, T., Moser, M., Oesch, B., McBride, P.A., Manson, J.C. (1996) Altered circadian activity rhythms and sleep in mice devoid of prion protein, *Nature* **380**, 639–642.

Tremblay, P., Ball, H.L., Kaneko, K., Groth, D., Hegde, R.S., Cohen, F.E., DeArmond, S.J., Prusiner, S.B., Safar, J.G. (2004) Mutant PrP^{Sc} conformers induced by a synthetic peptide and several prion strains, *J. Virol.* **78**, 2088–2099.

Tremblay, P., Meiner, Z., Galou, M., Heinrich, C., Petromilli, C., Lisse, T., Cayetano, J., Torchia, M., Mobley, W., Bujard, H., DeArmond, S.J., Prusiner, S.B. (1998) Doxycycline control of prion protein transgene expression modulates prion disease in mice, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 12580–12585.

Vey, M., Pilkuhn, S., Wille, H., Nixon, R., DeArmond, S.J., Smart, E.J., Anderson, R.G., Taraboulos, A., Prusiner, S.B. (1996) Subcellular colocalization of the cellular and scrapie prion proteins in caveolae-like membranous domains, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 14945–14949.

Viles, J.H., Cohen, F.E., Prusiner, S.B., Goodin, D.B., Wright, P.E., Dyson, H.J. (1999) Copper binding to the prion protein: structural implications of four identical cooperative binding sites, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 2042–2047.

Vnencak-Jones, C.L., Phillips, J.A. (1992) Identification of heterogeneous PrP gene deletions in controls by detection of allele-specific heteroduplexes (DASH), *Am. J. Hum. Genet.* **50**, 871–872.

Waggoner, D.J., Drisaldi, B., Bartnikas, T.B., Casareno, R.L., Prohaska, J.R., Gitlin, J.D. Harris, D.A. (2000) Brain copper content and cuproenzyme activity do not vary with prion protein expression level, *J. Biol. Chem.* **275**(11), 7455–7458.

Weissmann, C. (1991) Spongiform encephalopathies – the prion's progress, *Nature* **349**, 569–571.

Wells, G.A., Hawkins, S.A., Austin, A.R., Ryder, S.J., Done, S.H., Green, R.B., Dexter, I., Dawson, M., Kimberlin, R.H. (2003) Studies of the transmissibility of the agent of bovine spongiform encephalopathy to pigs, *J. Gen. Virol.* **84**, 1021–1031.

Wells, G.A.H. (2002) Report on TSE infectivity distribution in ruminant tissues (State of Knowledge, December 2001), European Commission: 10–37.

Westaway, D., Cooper, C., Turner, S., Da Costa, M., Carlson, G.A., Prusiner, S.B. (1994) Structure and polymorphism of the mouse prion protein gene, *Proc. Natl. Acad. Sci. U. S. A.* **91**, 6418–6422.

Westaway, D., Goodman, P.A., Mirenda, C.A., McKinley, M.P., Carlson, G.A. Prusiner, S.B. (1987) Distinct prion proteins in short and long scrapie incubation period mice, *Cell* **51**, 651–662.

Westaway, D., Zuliani, V., Cooper, C.M., Da Costa, M., Neuman, S., Jenny, A.L., Detwiler, L., Prusiner, S.B. (1994) Homozygosity for prion protein alleles encoding glutamine-171 renders sheep susceptible to natural scrapie, *Genes Dev.* **8**, 959–969.

Westaway, D., Mirenda, C.A., Foster, D., Zebarjadian, Y., Scott, M., Torchia, M., Yang, S.-L., Serban, H., DeArmond, S.J., Ebeling, C., Prusiner, S.B., Carlson, G.A. (1991) Paradoxical shortening of scrapie incubation times by expression of prion protein transgenes derived from long incubation period mice, *Neuron* **7**, 59–68.

White, A.R., Enever, P., Tayebi, M., Mushens, R., Linehan, J., Brandner, S., Anstee, D., Collinge, J., Hawke, S. (2003) Monoclonal antibodies inhibit prion replication and delay the development of prion disease, *Nature* **422**, 80–83.

Wickner, R.B. (1994) [URE3] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*, *Science* **264**, 566–569.

Wilesmith, J.W. (1991) The epidemiology of bovine spongiform encephalopathy, *Semin. Virol.* **2**, 239–245.

Wilesmith, J.W., Ryan, J.B.M. Atkinson, M.J. (1991) Bovine spongiform encephalopathy – Epidemiologic studies on the origin, *Vet. Rec.* **128**, 199–203.

Will, R.G., Ironside, J.W., Zeidler, M., Cousens, S.N., Estibeiro, K., Alperovitch, A., Poser, S., Pocchiari, M., Hofman, A., Smith, P.G. (1996) A new variant of Creutzfeldt-Jakob disease in the UK, *Lancet* **347**, 921–925.

Wille, H., Zhang, G.-F., Baldwin, M.A., Cohen, F.E. Prusiner, S.B. (1996) Separation of scrapie prion infectivity from PrP amyloid polymers, *J. Mol. Biol.* **259**, 608–621.

Wille, H., Michelitsch, M.D., Guénebaut, V., Supattapone, S., Serban, A., Cohen, F.E., Agard, D.A. Prusiner, S.B. (2002) Structural studies of the scrapie prion protein by electron crystallography, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 3563–3568.

Williams, E.S., Young, S. (1980) Chronic wasting disease of captive mule deer: a spongiform encephalopathy, *J. Wildl. Dis.* **16**, 89–98.

Williamson, R.A., Peretz, D., Smorodinsky, N., Bastidas, R., Serban, H., Mehlhorn, I., DeArmond, S.J., Prusiner, S.B., Burton, D.R. (1996) Circumventing tolerance to generate autologous monoclonal antibodies to the prion protein, *Proc. Natl. Acad. Sci. U. S. A.* **93**: 7279–7282.

Wyatt, J.M., Pearson, G.R., Smerdon, T.N., Gruffydd-Jones, T.J., Wells, G.A.H., Wilesmith, J.W. (1991) Naturally occurring scrapie-like spongiform encephalopathy in five domestic cats, *Vet. Rec.* **129**, 233–236.

Yehiely, F., Bamborough, P., Costa, M.D., Perry, B.J., Thinakaran, G., Cohen, F.E., Carlson, G.A., Prusiner, S.B. (1997) Identification of candidate proteins binding to prion protein, *Neurobiol. Dis.* **3**, 339–355.

Zulianello, L., Kaneko, K., Scott, M., Erpel, S., Han, D., Cohen, F.E., Prusiner, S.B. (2000) Dominant-negative inhibition of prion formation diminished by deletion mutagenesis of the prion protein, *J. Virol.* **74**(9), 4351–4360.

Programmed Cell Death

David L. Vaux and Andreas Strasser
The Walter and Eliza Hall Institute of Medical Research, Royal Parade, Parkville, Victoria, Australia

1 Introduction 619

2 Mechanisms of Apoptosis 619
2.1 Effector Mechanism 619
2.1.1 Caspases 620
2.1.2 Caspase Substrates 621
2.1.3 Caspase Adaptor Proteins 622
2.2 Control of Apoptosis 622
2.2.1 IAPs 622
2.2.2 IAP Antagonists 623
2.2.3 Bcl-2 Protein Family 623
2.3 Cell-death Signaling Pathways 626
2.3.1 Death-receptor Pathway 626
2.3.2 Bcl-2 Inhibitable Pathway 627
2.4 Fate of Apoptotic Cells 627

3 Apoptosis in Normal Physiology 627
3.1.1 Development 627
3.1.2 Homeostasis 628
3.1.3 Immune System 628

4 Apoptosis in Disease 629
4.1.1 Cancer 629
4.1.2 Autoimmune Disease 630
4.1.4 Sepsis 630

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

4.1.5 Degenerative Disease 630
4.1.6 Infectious Disease 631
4.2 Therapeutics 631

Bibliography 631
Books and Reviews 631
Primary Literature 632

Keywords

Apoptosis
A physiological cell-suicide process whose mechanism is shared among the metazoans.

Bcl-2
An apoptosis inhibitory protein whose gene is often translocated in follicular lymphoma.

Caspase
A family of cysteine proteases that cleave their substrates after aspartic-acid residues.

Death Receptor
Members of the tumor necrosis factor (TNF) receptor family that send proapoptotic signals when ligated.

IAP
Member of the inhibitor of apoptosis protein family.

p53
A tumor-suppressor gene that is commonly mutated in human cancers.

Programmed Cell Death
Cell death carried out by a physiological process whose role is to cause the death of the cell.

Apoptosis is a physiological process of cell suicide that is implemented by activation of latent proteases that demolish vital cellular constituents. Apoptosis is used to sculpt the body during development, to maintain constant cell number by balancing cell production by mitosis, and to remove cells that are damaged or infected. Failure of cells to undergo apoptosis can lead to cancer and autoimmune disease, and inappropriate apoptosis may contribute to or cause organ damage in degenerative disease and ischemic vascular injury.

1 Introduction

All cells are mortal, and they can die in a number of ways. Cells may be killed if exposed to a physical insult, drug, toxin, or abnormal gene product that interferes with a process required for ongoing survival, or if they fail to receive the nutrients needed to maintain their metabolism. In vertebrates, this kind of cell death is relatively uncommon under physiological conditions. By far, the most common way by which cells die is by activating a physiological, genetically programmed, suicide mechanism (Fig. 1).

Physiological cell death – in which a cell dies by activating an endogenous process whose role is to kill the cell – has been observed in most multicellular organisms and some single-celled organisms. In single-celled organisms, it is used as a contingent strategy to protect a group of individuals by limiting spread of infection or by reducing the consumption of limited resources. In multicellular organisms, cell death is programmed to occur in particular cells in particular circumstances.

Animal cells undergoing physiological cell death typically exhibit an appearance that has been termed *apoptosis*. Apoptotic features include shrinkage of the cell, appearance of blebs in the plasma membrane, and condensation of the chromatin against the nuclear membrane. In culture, apoptotic cells subsequently lose the integrity of their plasma membranes, whereas within a tissue of an animal, apoptotic cells are rapidly engulfed by other cells and degraded in their lysosomes.

The molecular mechanism of apoptosis is shared amongst the metazoans. Although defensive (against pathogens) and developmental cell death occur commonly in plants, and some of the signaling pathways that regulate these processes resemble those found in apoptosis signaling in animals, the effector mechanism of physiological cell death in plants and animals appears to be different.

The key event in apoptosis is activation of a family of latent proteases, termed *caspases*. Numerous control and signaling pathways exist so that cells undergo apoptosis in response to developmental cues, at the appropriate time.

As well as being important for development and homeostasis of cell number, apoptosis is also involved in many disease processes, such as removal of infected or damaged cells. An increase or a decrease in the amount of apoptosis has been observed in many diseases. In some cases, defects in apoptosis signaling are the primary cause of disease, whereas, in other cases, the appearance of apoptotic cells is a consequence of the disease process.

A detailed understanding of apoptosis is important because it may allow development of new therapies to treat diseases in which cells die inappropriately or in which they fail to die when they should.

2 Mechanisms of Apoptosis

2.1 Effector Mechanism

The key event in apoptosis is the cleavage of intracellular proteins by a family of aspartate-specific cysteine proteases, termed *caspases*. Caspases exist within healthy cells as zymogens. Once activated, they cleave other proteins within the cell, after particular aspartic-acid residues. If enough of a protein required for a cell to maintain viability is cleaved, the cell will die. Within the tissues of animals, cells

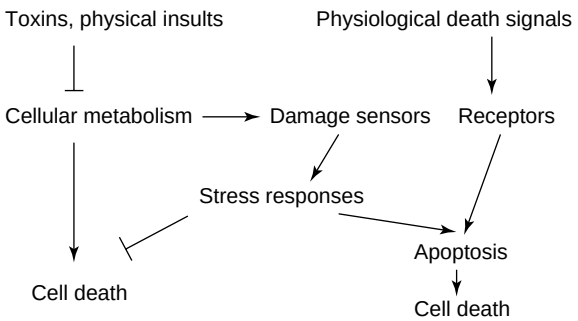

Fig. 1 All cells are mortal and can be killed if a vital metabolic process is blocked. If this process is only partially inhibited, a cell may detect the damage and mount a number of stress responses. Some of these responses protect the cell and repair the damage; however, cells may also respond to stress by committing suicide. Animal cells can activate their apoptotic cell-death mechanism either as a stress response or under normal physiological conditions to aid in development or to maintain homeostasis of cell number. Cells dying following noxious stimuli can simultaneously exhibit apoptotic and necrotic changes.

undergoing apoptosis are rapidly engulfed by neighboring cells or by specialist migratory phagocytic cells.

2.1.1 **Caspases**

Caspases are one of about 20 different classes of cysteine proteases, all of which cause cleavage of the peptide bond using catalytic cysteine and histidine residues. Caspases are intracellular endopeptidases that cleave their substrates, after aspartic-acid residues.

Mammals have genes for about 12 caspases. Some, such as caspases 1, 4, and 5 in humans and 1, 11, and 12 in mice, have roles in innate immunity by cleaving, and thereby activating prointerleukin 1β, but it is possible that these enzymes also function in apoptosis. Caspase-14 is thought to play a role in keratinocyte differentiation, but all other caspases (3, 6, 7, 8, 9, and human 10) are thought to be critical only for apoptosis.

As with all cysteine proteases, during proteolysis, the sulfur of the catalytic cysteine in caspases acts as a nucleophile, forming a thiol ester with the carbon of the substrate peptide bond, and then, the catalytic histidine provides a hydrogen to the nitrogen that follows, allowing separation of the peptide bond.

What makes caspases different from other cysteine proteases is the scaffold they use to align their substrate and to promote catalysis. These differences also account for the differences in substrate preference between individual caspases. While they all cleave after aspartic-acid residues, and they all prefer small residues such as glycine, serine or alanine to follow the aspartic acid, they show specific preferences for the P4–P2 residues. Caspase 3, for example, prefers to cleave substrates with an acidic P4 residue (e.g. ...DEVD*G...), whereas caspase 9 prefers an aliphatic P4 (e.g...IETD*G...).

Caspases are synthesized as single-chain proenzymes (zymogens) that contain small (p10) and large (p20) subunits. The large subunit bears the catalytic cysteine and histidine residues, and residues in the small subunit help define substrate specificity.

Mammalian caspases can be grouped into two different classes on the basis of their substrate preferences, structure, and mode of activation.

The "downstream" or "effector" caspases 3, 6, and 7 account for most of the proteolytic activity in apoptotic cells. They have short prodomains at their amino-termini and are activated following cleavage of their zymogen forms by the "upstream" or "initiator" caspases (caspases 1, 2, 4, 5, 8, 9, 10, 11, and 12). In nonapoptotic cells, the effector caspases exist as dimers with very low zymogen activity. The two polypeptide chains that make up this dimer form catalytic sites in cis position once the flexible linkers that connect the small and large subunits of the monomers is cut by another caspase. The active form of effector caspases is a heterotetramer with two small (p10) subunits and two large (p20) subunits and two active sites. Thus, for effector caspases, their processing correlates well with their activity, and they are activated by proteolytic processing by an "upstream" or "initiator" caspase.

The "upstream" or "initiator" caspases (1, 2, 4, 5, 8, 9, 10, 11, and 12) exist as monomers in nonapoptotic cells. They have long amino-terminal prodomains bearing protein-interaction motifs, either CARDs (caspase recruitment domains) or DEDs (death effector domains), that allow them to bind to similar domains on caspase-activating adaptor molecules. Adaptor proteins promote activation of initiator caspases by inducing their dimerization or multimerization. Unlike the effector caspases, the initiator caspases do not need to be proteolytically processed in order to have significant activity. Nevertheless, once dimerized and activated, they can autoprocess, and this is thought to result in enhanced initiator-caspase activation. In addition, initiator caspases can be activated when cleaved by effector caspases, allowing a feed-forward cycle of caspase activation (Fig. 2).

2.1.2 Caspase Substrates

Several hundred potential caspase substrates have been identified by adding active caspases to cell extracts *in vitro* and by proteomic analysis of extracts of cells undergoing apoptosis. Currently, it is unclear which of these proteins must be cleaved for a cell to die, but cleavage of some cells has been shown to be important for some of the morphological or biochemical changes associated with

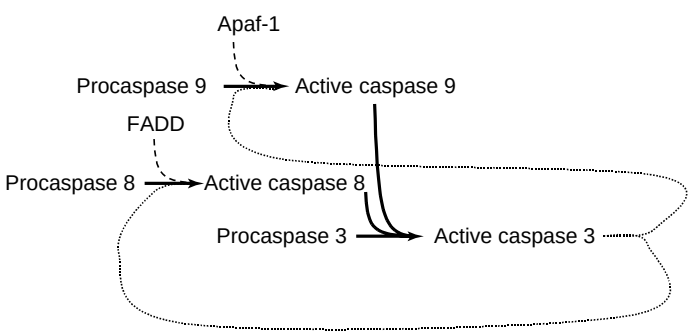

Fig. 2 Upstream caspases such as caspase 8 and caspase 9 are activated by dimerization induced by the adaptors FADD and Apaf-1 (dashed lines). These caspases can then cleave and activate downstream caspases such as caspase 3. Active, processed caspase 3 can feed back to cleave and activate further caspase 8 and caspase 9 (dotted lines).

apoptosis, such as membrane blebbing or DNA degradation.

Inhibitor of caspase activated DNAse (ICAD), for example, functions as a chaperone and inhibitor of the latent endonuclease CAD (caspase activated DNAse). After ICAD is cleaved by caspases, active CAD is liberated, which digests the cell's DNA. Cleavage of the Rho effector protein ROCK I by caspases generates a truncated active form that has been reported to be responsible for membrane blebbing seen in apoptosis.

2.1.3 Caspase Adaptor Proteins

Apaf-1 is the adaptor protein responsible for activating caspase 9. 'From the close similarity of mice deficient in caspase 9 with those deficient in Apaf-1, it is apparent that Apaf-1 is the main, if not the only, activator of caspase 9, and that Apaf-1 does not activate other caspases'? Apaf-1 bears a CARD domain that enables it to bind to the CARD of caspase 9. It also bears a nucleotide-binding P loop and a number of WD40 repeats. Apaf-1 is activated by binding to cytochrome *c* and dATP, and then, together with caspase 9, it aggregates into a large molecular-weight complex known as the "*apoptosome*." The apoptosome then recruits and activates caspase 3.

FADD is the adaptor responsible for activating caspase 8. Recruitment of FADD to aggregated members of the tumor necrosis factor (TNF) receptor family, which have an intracellular death domain (DD) (e.g. TNFR1, Fas and TRAIL receptors), occurs via DDsin FADD and results in formation of a complex known as the DISC (death inducing signaling complex). Within the DISC, FADD aggregates caspase 8 (and in humans, also caspase 10) through an interaction involving the DED of FADD and the two DEDs of procaspase 8 (or 10). The caspase then dimerizes, activates, and autoprocesses.

2.2 Control of Apoptosis

2.2.1 IAPs

Cell death can be controlled at multiple levels. For example, caspases can be controlled even after they are activated by the inhibitor of apoptosis (IAP) proteins. Four members of the IAP family, XIAP, cIAP1, cIAP2 and ML-IAP, have been shown to bind to and inhibit caspases, although XIAP has been the most extensively studied.

All the IAPs bear one to three copies of a domain, termed a BIR (baculoviral IAP repeat), that is a zinc-binding fold. Each of the antiapoptotic IAPs also carries a RING finger, another zinc-binding domain that allows IAPs to act as ubiquitin E3 ligases. Using their RING domains, IAPs can ubiquitinate themselves and other interacting proteins, targeting them for degradation by the proteasome.

XIAP, which has 3 BIRs, can inhibit apoptosis by binding to the catalytic sites of active caspases 3 and 7, via a region just amino-terminal to BIR2. XIAP can also bind via the core of BIR3 to the processed active caspase 9, thereby inhibiting it. Although XIAP binds to and inhibits caspases with high affinity, and can potently inhibit apoptosis of cultured cells exposed to certain stress stimuli, it has not yet been shown to confer long-term survival. Moreover, no major phenotype has been detected in those mice in which XIAP is deleted. It is, however, possible that IAPs have essential but overlapping functions in mammals, and interesting phenotypes may only appear in mice lacking two or more of these proteins.

2.2.2 IAP Antagonists

It has been clearly demonstrated that regulation of apoptosis by IAPs is very important in the fly, *Drosophila melanogaster*. Deletion of DIAP1 results in early embryonic lethality with massive apoptosis. In *Drosophila*, IAPs are themselves controlled by the proapoptotic IAP inhibitors, HID, Reaper, Grim, and Sickle, each of which bind to the core of BIR3 of DIAP1 to liberate active caspases. Two mammalian proteins, Smac/Diablo and HtrA2/Omi, have been identified that can bind to mammalian IAPs in a similar way and displace active caspases *in vitro*. Unlike the insect IAP antagonists, which are located in the cytoplasm, Smac/Diablo and HtrA2/Omi normally reside in the mitochondria such that they are only able to bind to IAPs and release caspases once the mitochondria are disrupted (Fig. 3). This may indicate that these mammalian IAP-binding proteins may regulate a caspase amplification mechanism rather than the initiation of the caspase cascade. Furthermore, unlike the insect IAP antagonists, Smac/Diablo is dispensable for normal development as Smac/Diablo-deficient mice are phenotypically normal and the primary role of HtrA2/Omi is to degrade or refold misfolded mitochondrial proteins, rather than to regulate apoptosis. As mentioned above, this may indicate that multiple IAP inhibitors have essential but overlapping functions in mammalian development.

2.2.3 Bcl-2 Protein Family

In addition to being controlled after their activation by IAPs, caspases are subject to regulation of their activation. Bcl-2, the product of a gene frequently translocated in human follicular center lymphoma, was initially identified as an inhibitor of cell death because it could prevent apoptosis of growth-factor-dependent cells deprived of cytokine. Three groups of Bcl-2-like proteins, all of which share one or more Bcl-2 homology (BH) domains, have now been identified. Some members, including mammalian Bcl-2, Bcl-xL, Bcl-w, A1/Bfl1, Mcl-1, Boo/Diva, *Caenorhabditis elegans* CED-9, and certain viral homologs such as adenovirus E1B19kD and Epstein Barr virus BHRF1, inhibit apoptosis. Bax, Bak, and Bok/Mtd share extensive sequence (2 or 3 BH domains) and structural homology with their prosurvival relatives, but they promote apoptosis and are essential for programmed apoptosis in a wide variety of circumstances. The third group, whose similarity is confined to a single BH3 domain (hence their name "BH3-only proteins"), are also apoptosis promoters and are required for programmed cell death in response to many stimuli. This group includes proteins such as mammalian Bid,

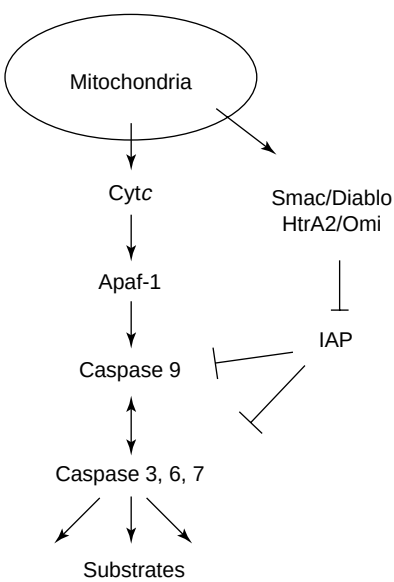

Fig. 3 IAPs can inhibit activated caspases, but they can be antagonized by proteins released from the mitochondria, such as Smac/Diablo and HtrA2/Omi.

Bik/Blk/Nbk, Bad, Bim/Bod, Hrk/DP5, Puma/Bbc3, Bcl-G_S, Noxa, and Bmf as well as *C. elegans* EGL-1.

Although Bcl-2 was the first component of the cell-death mechanism to be recognized, exactly how this component and its family members function remains controversial. There are two main models (Fig. 4). The first suggests that Bcl-2 functions like the *C. elegans* cell-death inhibitor CED-9, which acts by directly binding to the Apaf-1-like adaptor protein CED-4, thereby preventing it from activating the caspase CED-3. According to this model, Bcl-2 and other antiapoptotic family members prevent cell death by binding to CED-4-like adaptors to prevent them from activating caspases. In this model, apoptosis is triggered when proapoptotic Bax or Bak or BH3-only proteins bind to the antiapoptotic proteins, such as Bcl-2, and displace the adaptor proteins, thereby allowing them to activate initiator caspases. Genetic and biochemical studies have shown that BH3-only proteins and Bax/Bak-like proteins are both essential for apoptosis, but it is not clear how they interact biochemically. It is possible that the Bax/Bak-like proteins participate directly in the stimulation of adaptor proteins, but their function in caspase activation might be indirect.

The second model suggests that Bcl-2 prevents apoptosis by stopping release of proteins from the mitochondria. This idea achieved prominence when it was found

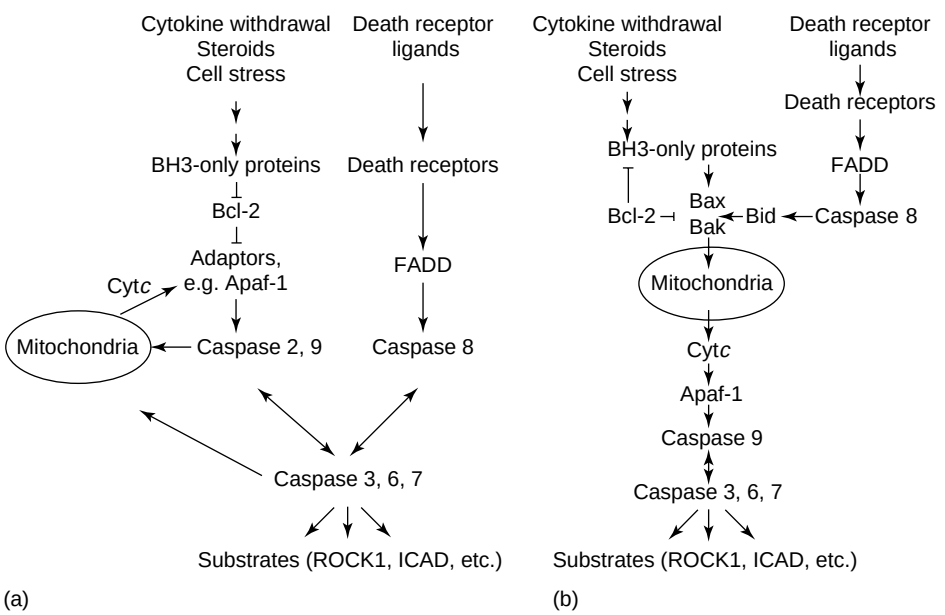

Fig. 4 Alternative models of the cell-death mechanism. In both models, independent pathways converge on a common effector mechanism. In the model (a), caspases are essential for apoptosis and the mitochondria play a nonessential role that serves to accelerate cell death. Bcl-2 promotes cell survival by preventing activation of caspases. In the model (b), the mitochondria play an obligatory role in cell death and caspases play a nonessential role as cells will die from loss of mitochondrial function, whether caspases are active or not. Bcl-2 promotes cell survival by preserving mitochondrial integrity.

that when released from the mitochondria into the cytosol, cytochrome *c* could bind to and activate Apaf-1, allowing it, in turn, to activate caspase 9. Since that time, it has been found that other proapoptotic proteins, such as Smac/Diablo and HtrA2/Omi, are also released from the mitochondria during apoptosis. According to this model, the proapoptotic Bcl-2 family members Bax and Bak form channels or pores in the mitochondrial outer membrane that directly or indirectly lead to release of apoptogenic proteins. In this model, the BH3-only proteins (directly or indirectly) cause changes in Bax and Bak that allow them to form the pores, and the antiapoptotic family members, such as Bcl-2, inhibit cell death by sequestering the BH3-only proteins.

The key difference between these models is that in the first, caspases are essential for apoptosis, and therefore, a pharmaceutical agent that can inhibit the relevant caspases would allow long-term survival of a cell. In this model, the critical events in mammalian apoptotic cell death would be similar to those in invertebrates, where experiments in the worm *C. elegans* have shown that the caspase CED-3 is required for all programmed cell deaths and experiments in *Drosophila* have shown that transgenic expression of the caspase inhibitor p35 prevents both programmed cell death and degenerative cell death in a retinitis pigmentosa model. In the second model, loss of mitochondrial function and release of mitochondrial proteins is the key event, so that even if caspase activity were blocked, cells would still die either because of loss of energy production by the mitochondria or, in a caspase independent manner, because of release of reactive oxygen species.

In both models, the proapoptotic signaling input comes from the BH3-only proteins. These proteins can integrate many different cell-death stimuli onto the apoptotic effector mechanism, and they themselves are regulated by a number of transcriptional and posttranslational mechanisms. Noxa and Puma/Bbc3 are transcriptionally upregulated by the tumor suppressor p53 and are therefore thought to be critical for DNA damage-induced apoptosis. Hrk/DP5 and Bim are induced by AP1 transcription factors, and Bim is also induced by the forkhead transcription-factor family member FKHRL1. Bim, Bmf, Bad, Bik, and Bid can be regulated at the posttranslational level. In healthy cells, Bim associates with microtubular dynein motor complex, and Bmf associates with the actin cytoskeleton-based myosin V motor complex, in such a way that they can be released following changes to the intracellular scaffolding. Bid is activated by proteolytic cleavage, for example, by caspase 8 or other caspases and may therefore form part of an amplification mechanism in the caspase cascade. Phosphorylation induced by cytokine-receptor stimulation allows Bad to be kept in an inactive state by sequestration to 14-3-3 scaffold proteins, and the proapoptotic activity of Bik/Blk/Nbk may instead be induced in response to phosphorylation.

In order to signal apoptosis, the BH3-only proteins bind to antiapoptotic members of the Bcl-2 family. The BH3 alpha helix of the BH3-only protein binds to a shallow groove on the surface of the antiapoptotic family member. In nonapoptotic cells, this groove may be occupied by the hydrophobic C-terminal tail of the antiapoptotic Bcl-2 family member. It is possible that binding of the BH3-only protein displaces the hydrophobic tail, which would then be free to interact with cellular membranes.

2.3 Cell-death Signaling Pathways

In mammalian cells, there are two distinct caspase-activation pathways. One is activated by members of the TNF receptor family with an intracellular DD (also known as *death receptors*, e.g. TNF-R1, Fas, TRAIL-R1, TRAIL-R2) and signals via the adaptor protein FADD to activate caspase 8. This pathway is known as the *death-receptor pathway*. It has also been called the *extrinsic pathway*, but this term is somewhat misleading because some of the stimuli that activate Bcl-2-regulated apoptosis signaling (see Sect. 3.1.3) are also extrinsic (e.g. addition of glucocorticoids or TGFβ). The nondeath receptor apoptosis-signaling pathway can be activated by stress conditions, such as removal of cytokines, addition of glucocorticoids, or DNA damage. In these cases, Bcl-2 is effective at blocking cell death, and release of cytochrome *c* from the mitochondria is an early feature. This pathway is known as the *Bcl-2-inhibitable pathway, intrinsic pathway*, or the *mitochondrial pathway*. We believe that the term "intrinsic pathway" is not entirely appropriate because, as mentioned above, this pathway can be activated by certain extrinsic stimuli, such as treatment with glucocorticoids or TGFβ. The term "mitochondrial pathway" is also problematic because, although in this pathway, mitochondria are damaged and apoptogenic molecules are released, there is now evidence that these changes are not essential for cell death but are a consequence of activation of caspases acting upstream.

2.3.1 Death-receptor Pathway

Ligation of members of the TNF receptor family with an intracellular DD (e.g. TNFR1, Fas, TRAIL-R1, and TRAIL-R2) can lead to apoptosis in a pathway that requires the adaptor molecule FADD and caspase 8 (with an involvement of caspase 10 in human cells). It should be noted that activation of caspase 8 and apoptosis is not the only event stimulated by these receptors, as ligand binding can also activate transcription factors NF-kB and Jun. Moreover, experiments with genetically modified mice have shown that in T lymphocytes, FADD is not only required for "death-receptor"-induced apoptosis but is also essential for mitogen-induced proliferation. It is believed that regulatory proteins, such as FLIP (which is induced by NF-kB), may prevent activation of caspase 8 by FADD and thereby allow other signaling functions of FADD to come to the fore.

Once caspase 8 is activated, it can directly cleave some protein substrates within the cell, but it also cleaves and activates effector caspases 3 and 7. Mitochondria lose their membrane potential and proteins such as cytochrome *c* and Smac/Diablo leave the mitochondria and enter the cytosol in death receptor–induced, caspase 8–dependent apoptosis. These events are dependent on caspase 8 activity because the mitochondrial events do not occur if caspase 8 is inhibited, for example, by the Cowpox-derived protein CrmA. In most cells, Bcl-2 family members do not play critical roles in death receptor–induced apoptosis as lymphocytes or myeloid cells in which Bcl-2 is overexpressed or Bax and Bak are deleted remain sensitive to Fas ligand.

Some believe that a subset of cells, known as *Type II cells*, activate less caspase 8 following death-receptor ligation such that death receptor–induced apoptosis in these cells can be partly inhibited by Bcl-2 family members. It is proposed that in these Type II cells, caspase 8 cleaves the

BH3-only protein Bid, which then activates the Bcl-2-inhibitable death pathway.

2.3.2 Bcl-2 Inhibitable Pathway

When apoptosis occurs after cytokine withdrawal, addition of glucocorticoids, or following cell stress caused by chemotherapeutic drugs, toxins, hypoxia, γ- or UV-radiation, or high temperature, it can almost always be inhibited by Bcl-2 (and its prosurvival homologs). It is very important to note, however, that inhibition of apoptosis by Bcl-2 will not necessarily save the cell; it will still die if a process essential for its continued survival is blocked but it will not exhibit the classical apoptotic morphology. The fact that cells commonly respond to sublethal damage by undergoing apoptosis means that many cells treated with toxins or chemotherapeutic drugs exhibit both apoptotic as well as necrotic changes.

2.4 Fate of Apoptotic Cells

In vivo, apoptotic cells are rapidly engulfed by either neighboring cells or professional phagocytes before they lyse. A key event in the recognition of apoptotic cells by the engulfing cells is the exposure of phosphatidylserine (PS) on the outer surface of the plasma membrane. In healthy cells, PS is maintained on the inner surface through an active process that is thought to be mediated by a translocase. While activation of the apoptotic effector machinery leads to rapid surface exposure of PS, it is not currently known how this happens, although it is presumed to be a caspase-dependent event that causes either inactivation of a PS translocase or activation of a PS scramblase.

Phosphatidylserine on the apoptotic cell is recognized by a specific receptor, termed the PSR, but other receptors, such as CD36, CD-14, and ABC1 and the integrins $\alpha_V\beta_5$ and $\alpha_V\beta_3$ as well as macrophage-secreted MFG-E8, are also implicated in the recognition of apoptotic cells by phagocytes. Binding of the apoptotic cell activates a signaling pathway within the engulfing cell that involves phosphorylation of p130 cas, recruitment of Crk-II and Dock-180, and activation of Rac-1, which ultimately causes formation of pseudopodia around the dying cell. Engulfed cells are then fully degraded within the lysosomal compartment.

3 Apoptosis in Normal Physiology

3.1.1 Development

Development in mammals, as in insects, worms, and possibly all multicellular organisms, is characterized by generation of excess cells, followed by removal of those that are not needed by apoptosis. For example, during development in *C. elegans*, 131 of the 1090 somatic cells formed are fated to undergo programmed cell death. When cell death is blocked, for example, by mutating the genes for the caspase CED-3, its activator CED-4, or the BH3-only protein EGL-1, most of the 131 cells that fail to die adopt the fate of their sister cells.

In vertebrates, apoptosis is used to remove interdigital tissue, allowing the fingers and toes to separate. It is also used to create hollow tubes and to remove epithelial cells when tissues fuse at the midline. It is estimated that over 80% of neurons formed during intrauterine development are destined to die by apoptosis in the perinatal period. Consistent with this, deletion of genes for proteins that promote apoptosis, such as Apaf-1, caspase 9, and caspase

3, lead to abnormal accumulation of neuronal cells, and animals often die perinatally with incomplete closure of the neural tube. However, when Apaf-1, caspase 9, or caspase 3 mutations are crossed onto the C57BL/6 background, the phenotype is much milder, and some apparently normal animals live to adulthood and reproduce, demonstrating that there is no absolute requirement for any of these proteins. Furthermore, clonogenic experiments have shown that in hemopoietic cells at least, neither Apaf-1, caspase 9, nor caspase 2 are required for cell death following cytokine withdrawal or treatment with cytotoxins, but are only required for rapid display of the classic apoptotic morphology.

3.1.2 Homeostasis

Apoptosis is also used to maintain constant cell numbers by balancing mitotic cell production. For example, if a lobe of the liver is removed, the hepatocytes in the remaining lobes of the liver grow and proliferate. Typically, there is an overshoot, but apoptosis occurs to bring the liver back to the appropriate size.

3.1.3 Immune System

The immune system uses apoptosis very extensively, both for selection of its repertoire of antigen-receptor specificities amongst lymphocytes and as a tool for combating pathogens. Developing B lymphocytes and thymocytes that fail to express antigen receptors cannot transmit signals to sustain their survival and undergo an apoptotic "death by neglect." In early B cells, this death is mediated by the Bcl-2 regulated pathway, and in T cell progenitors, it requires the caspase adaptor FADD, indicating an involvement of death receptors. Developing B and T cells with receptors that have a high affinity for a self antigen also undergo apoptotic "negative selection" that is important for the maintenance of tolerance to self, thus imposing a barrier against autoimmune disease. Experiments in genetically modified mice have shown that the BH3-only protein Bim is critical for negative selection and that Bcl-2 can inhibit this process. During an immune response, large numbers of effector B and T lymphocytes are produced, and at the end of the response, most of the cells undergo apoptosis, but a few "memory cells" remain. It appears that both Bcl-2-regulated apoptosis signaling that involves Bim as well as death-receptor signaling are involved in the culling of effector lymphocytes.

Cytotoxic T lymphocytes (CTL) kill virally infected cells by inducing them to undergo apoptosis. CTL can kill by two different mechanisms. The granule-exocytosis mechanism depends on perforin and calcium and involves introduction of components of the granules of CTL into the target cell. Among the granule components are several serine proteases termed *granzymes*. Unlike other serine proteases, but like the caspases, granzyme B has preference for cleaving after aspartic-acid residues. Granzyme B can cleave and activate caspases *in vitro* and can also cleave caspase substrates such as vital structural components of the cell and the BH3-only protein Bid.

Because they often express membrane-bound Fas ligand, CTL can also kill certain target cells that bear Fas. Comparison of mice lacking genes for perforin with those bearing mutations in the genes for Fas (*lpr*) or Fas ligand (*gld*) indicate that the perforin-dependent granule-exocytosis mechanism is much more important for CTL activity than Fas Ligand/Fas-dependent killing.

4 Apoptosis in Disease

4.1.1 Cancer

For a tumor to expand, the rate of cell division must exceed the rate of cell death. Although many oncogenes act by promoting cell proliferation, the association of translocations involving Bcl-2 with follicular lymphoma and the production of transgenic mice that overexpress Bcl-2 proved that inhibition of cell death could also be oncogenic. Furthermore, because mice bearing a transgene that stimulates cell division (c-*myc*) as well as a transgene that inhibits cell death (*bcl*-2) develop malignancies much more rapidly than mice bearing either of the transgenes alone, it is clear that promotion of cell division and inhibition of apoptosis are potently synergistic.

The gene most commonly mutated in human cancers encodes the tumor suppressor p53. In response to DNA damage, p53 becomes stabilized and functionally activated. p53 acts in two different ways to prevent the formation of tumors: it can promote production of the cell-cycle inhibitor p21, leading to arrest in G1 that allows time for the DNA to be repaired and it can promote apoptosis via the Bcl-2-inhibitable pathway. Precisely how p53 does this is not yet clear, but the favored model is that p53 induces transcription of the BH3-only proteins PUMA and Noxa.

Transgenically expressed Bcl-2 can prevent apoptosis of irradiated thymocytes, pre-B cells, and mature B and T lymphocytes as does deletion of p53, but p53−/−mice develop thymomas much more readily than Bcl-2-transgenic mice, so not all of the tumor-suppressor activity of p53 is due to its proapoptotic activity. On the other hand, because mutant mice lacking the cell-cycle inhibitor p21 also have much lower incidence of tumors than p53−/−mice, it appears that the proapoptotic as well as the cell-cycle regulatory functions of p53 are critical for its tumor-suppressive activity.

Mutations of proapoptotic genes and overexpression of antiapoptotic genes have been searched for in many different cancer types, but apart from translocations involving *bcl-2* and mutation of *p53*, these alterations are not common. Mice lacking Fas are prone to plasmacytoma development, and loss of FADD function promotes T lymphomagenesis in rag-deficient mice. Moreover, loss of Fas and Bcl-2 overexpression synergize potently in the development of myeloid leukemia in mice, demonstrating again that "death receptors" and Bcl-2 regulate distinct apoptosis-signaling pathways. Translocations involving cIAP2 are found in mucosa associated lymphoid tissue (MALT) lymphomas, but it seems unlikely that an antiapoptotic activity of cIAP2 is involved. Bax mutations have been found in some gastrointestinal tumors. Although there have been numerous reports of correlations between increased expression of cell-death inhibitors in cancers, or decreased expression of proapoptotic genes, a causal role has rarely been established, and an increased incidence of spontaneous tumors has not been reported to date in mice lacking genes for caspases, Apaf-1, Bax, Bak, or BH3-only proteins, or in mice transgenic for XIAP or CrmA.

In theory, increased expression of apoptosis inhibitors, or selection of cancer cells with higher levels of apoptosis inhibitors, may contribute to resistance to chemotherapeutic drugs. Because apoptosis, mediated via the Bcl-2 inhibitable pathway, is a common response to cell stress, including that caused by chemotherapeutic agents or γ-radiation, these drugs can cause cell

death in two very different ways. They can kill cells directly by interfering with a vital cellular process, such as protein synthesis or DNA replication, or they can kill cells indirectly by provoking an apoptotic stress response. At present, it is unclear which mechanism of cell death is more important for their therapeutic effect.

4.1.2 Autoimmune Disease

Transgenic mice that overexpress Bcl-2 and those lacking the proapoptotic genes for Bim, Fas (*lpr*), and Fas ligand (*gld*) develop autoimmune syndromes that resemble systemic lupus erythematosus. While these experiments and observations have proven that a properly regulated apoptotic mechanism is required to avoid autoimmune disease, the common human autoimmune diseases have rarely been found to be due to defects in cell-death genes.

Acute vascular disease Because hypoxia is one of the many cell stresses that can provoke an apoptotic response, it is possible that many of the neurons that die after a stroke, and the cardiomyocytes that die after a heart attack, are not dying purely from lack of oxygen but may be doing so because of activation of their apoptotic mechanism. Indeed, apoptotic cells have been found in hypoxic tissues of experimental animals, and experiments with first generation caspase-inhibitory drugs suggest that the total amount of cell death can be reduced by inhibiting apoptosis.

4.1.4 Sepsis

So far, the most promising results with antiapoptotic drugs have been obtained in preclinical studies of sepsis, where caspase inhibitors or Bcl-2 transgene expression in lymphocytes could prevent the death of mice that had undergone cecal ligation and puncture (CLP). Since Bcl-2 expression in lymphocytes only was effective, it appears that death or survival of effector lymphocytes determines the outcome of sepsis.

4.1.5 Degenerative Disease

Neurodegenerative diseases such as motor neurone disease, Parkinson's disease, and Alzheimer's disease are characterized by gradual loss of neurons. In some cases, signs have been found suggesting that this death is due to apoptosis, but because apoptosis is a rapid event, whereas progress of these diseases is chronic, it is very difficult to catch cells in the act of dying.

Because the processes that initiate these neurodegenerative diseases is still obscure, it may be that even if the cells ultimately die by apoptosis, they may have already been rendered nonfunctional by the primary abnormality, perhaps indicating that caspase-inhibitory drugs would not be able to halt or reverse tissue damage.

Nevertheless, the possibility that inhibiting apoptosis might retard progression of these diseases may justify trials of apoptosis inhibitors. The chronic nature of these diseases would necessitate long-term treatment, so it is likely that trials of cell-death inhibitors for neurodegenerative diseases will only occur after the inhibitors have been tested in more acute syndromes, such as septic shock (see Sect. 4.1.1).

Mice deficient for Bcl-2 die at an early age (3–6 weeks) from degenerative polycystic kidney disease and also have lymphopenia and abnormal melanocyte death, causing premature graying (Veis et al., 1993). All these defects could be prevented by concomitant loss of the BH3-only protein Bim; remarkably, loss of even

one allele of *bim* completely prevented polycystic kidney disease. This indicates that proapoptotic Bcl-2 family members may play critical roles in cell killing in degenerative diseases.

4.1.6 Infectious Disease

Cell death is an effective strategy to combat viral infections because viruses need to subvert the machinery of an intact cell in order to replicate. A cell that can detect the presence of a viral particle and respond by activating its apoptotic mechanism can prevent the virus from reproducing and infecting further cells.

This apoptotic antiviral strategy has put selective pressure on viruses to carry genes for apoptosis inhibitors. For example, certain baculoviruses (which infect insect cells) encode the caspase inhibitors p35 and IAP, cowpox virus carries the caspase-1 and -8 inhibitor CrmA, and many herpes viruses as well as adenoviruses, Epstein Barr virus and African Swine Fever Virus, encode Bcl-2 homologs.

4.2 Therapeutics

Hopefully, this understanding of the mechanisms of apoptosis will allow development of drugs to treat diseases in which there is either failure of or inappropriate apoptosis. With the benefit of hindsight, it is now clear that several therapies that were found to be useful empirically, such as γ-radiation, glucocorticoids, TNF antagonists, and some monoclonal antibodies, work to a significant extent by promoting or inhibiting apoptosis.

New caspase inhibitors, death-receptor agonists or antagonists, as well as inhibitors of anti- or proapoptotic Bcl-2 family members are likely to be tested in the clinic in the near future.

See also Aging and Sex, DNA Repair in; Molecular Mediators: Cytokines; Nucleic Acids (DNA) Damage and Repair.

Bibliography

Books and Reviews

Adams, J.M., Cory, S. (1998) The Bcl-2 protein family – arbiters of cell survival, *Science* **281**, 1322–1326.

Ashkenazi, A., Dixit, V.M. (1999) Apoptosis control by death and decoy receptors, *Curr. Opin. Cell Biol.* **11**, 255–260.

Barry, M., McFadden, G. (1998) Apoptosis regulators from DNA viruses, *Curr. Opin. Immunol.* **10**, 422–430.

Earnshaw, W.C., Martins, L.M., Kaufmann, S.H. (1999) Mammalian caspases: structure, activation, substrates, and functions during apoptosis, *Annu. Rev. Biochem.* **68**, 383–424.

Ellis, R.E., Yuan, J.Y., Horvitz, H.R. (1991) Mechanisms and functions of cell death, *Annu. Rev. Cell. Biol.* **7**, 663–698.

Fadok, V.A., Chimini, G. (2001) The phagocytosis of apoptotic cells, *Semin. Immunol.* **13**, 365–372.

Green, D.R., Reed, J.C. (1998) Mitochondria and apoptosis, *Science* **281**, 1309–1312.

Gross, A., McDonnell, J.M., Korsmeyer, S.J. (1999) BCL-2 family members and the mitochondria in apoptosis, *Genes. Dev.* **13**, 1899–1911.

Huang, D.C.S., Strasser, A. (2000) BH3-Only proteins – essential initiators of apoptotic cell death, *Cell* **103**, 839–842.

Kerr, J.F., Wyllie, A.H., Currie, A.R. (1972) Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics, *Br. J. Cancer* **26**, 239–257.

Kumar, S., Vaux, D.L. (2002) A Cinderella caspase takes center stage, *Science* **297**, 1290–1291.

Lane, D.P. (1992) Cancer. p53, guardian of the genome, *Nature* **358**, 15–16.

Makin, G., Dive, C. (2001) Apoptosis and cancer chemotherapy, *Trends Cell Biol.* **11**, S22–S26.

Nagata, S. (1997) Apoptosis by death factor, *Cell* **88**, 355–365.

Newton, K., Strasser, A. (2001) Cell death control in lymphocytes, *Adv. Immunol.* **76**, 179–226.

Nicholson, D.W. (1999) Caspase structure, proteolytic substrates, and function during apoptotic cell death, *Cell Death Differ.* **6**, 1028–1042.

Puthalakath, H., Strasser, A. (2002) Keeping killers on a tight leash: transcriptional and posttranslational control of the pro-apoptotic activity of BH3-only proteins, *Cell Death Differ.* **9**, 505–512.

Richardson, H., Kumar, S. (2002) Death to flies: *Drosophila* as a model system to study programmed cell death, *J. Immunol. Methods* **265**, 21–38.

Salvesen, G.S., Duckett, C.S. (2002) IAP proteins: blocking the road to death's door, *Nat. Rev. Mol. Cell Biol.* **3**, 401–410.

Stennicke, H.R., Salvesen, G.S. (2000) Caspases – controlling intracellular signals by protease zymogen activation, *Biochim. Biophys. Acta – Protein Struct. Mol. Enzymol.* **1477**, 299–306.

Strasser, A., O'Connor, L., Dixit, V.M. (2000) Apoptosis signaling, *Annu. Rev. Biochem.* **69**, 217–245.

Thornberry, N.A., Lazebnik, Y. (1998) Caspases – enemies within, *Science* **281**, 1312–1316.

Vaux, D.L., Korsmeyer, S.J. (1999) Cell death in development, *Cell* **96**, 245–254.

Vaux, D.L., Flavell, R.A. (2000) Apoptosis genes and autoimmunity, *Curr. Opin. Immunol.* **12**, 719–724.

Vaux, D.L., Haecker, G., Strasser, A. (1994) An evolutionary perspective on apoptosis, *Cell* **76**, 777–779.

Zheng, T.S., Hunot, S., Kuida, K., Flavell, R.A. (1999) Caspase knockouts: matters of life and death, *Cell Death Differ.* **6**, 1043–1053.

Primary Literature

Acehan, D., Jiang, X.J., Morgan, D.G., Heuser, J.E., Wang, X.D., Akey, C.W. (2002) Three-dimensional structure of the apoptosome: Implications for assembly, procaspase-9 binding, and activation, *Mol. Cell* **9**, 423–432.

Alnemri, E.S., Livingston, D.J., Nicholson, D.W., Salvesen, G., Thornberry, N.A., Wong, W.W., Yuan, J. (1996) Human ICE/CED-3 protease nomenclature, *Cell* **87**, 171.

Aravind, L., Dixit, V.M., Koonin, E.V. (2001) Apoptotic molecular machinery: vastly increased complexity in vertebrates revealed by genome comparisons, *Science* **291**, 1279–1282.

Atkinson, E.A., Barry, M., Darmon, A.J., Shostak, I., Turner, P.C., Moyer, R.W., Bleackley, R.C. (1998) Cytotoxic t lymphocyte-assisted suicide – caspase 3 activation is primarily the result of the direct action of granzyme b, *J. Biol. Chem.* **273**, 21261–21266.

Barry, M., Heibein, J.A., Pinkoski, M.J., Lee, S.F., Moyer, R.W., Green, D.R., Bleackley, R.C. (2000) Granzyme B short-circuits the need for caspase 8 activity during granule-mediated cytotoxic T-lymphocyte killing by directly cleaving bid, *Mol. Cell. Biol.* **20**, 3781–3794.

Birnbaum, M.J., Clem, R.J., Miller, L.K. (1994) An apoptosis inhibiting gene from a nuclear polyhedrosis virus encoding a polypeptide with cys/his sequence motif, *J. Virol.* **68**, 2521–2528.

Boldin, M.P., Goncharov, T.M., Goltsev, Y.V., Wallach, D. (1996) Involvement of mach, a novel mort1/fadd-interacting protease, in fas/apo-1- and tnf receptor-induced cell death, *Cell* **85**, 803–815.

Bouillet, P., Cory, S., Zhang, L., Strasser, A., Adams, J. (2001) Degenerative disorders caused by Bcl-2 deficiency prevented by loss of its BH3-only antagonist Bim, *Dev. Cell* **1**, 645–653.

Bouillet, P., Metcalf, D., Huang, D.C.S., Tarlinton, D.M., Kay, T.W.H., Kontgen, F., Adams, J.M., Strasser, A. (1999) Proapoptotic Bcl-2 relative bim required for certain apoptotic responses, leukocyte homeostasis, and to preclude autoimmunity, *Science* **286**, 1735–1738.

Bouillet, P., Purton, J.F., Godfrey, D.I., Zhang, L.C., Coultas, L., Puthalakath, H., Pellegrini, M., Cory, S., Adams, J.M., Strasser, A. (2002) BH3-only Bcl-2 family member Bim is required for apoptosis of autoreactive thymocytes, *Nature* **415**, 922–926.

Cain, K., Bratton, S.B., Langlais, C., Walker, G., Brown, D.G., Sun, X.M., Cohen, G.M. (2000) Apaf-1 oligomerizes into biologically active similar to 700-kDa and inactive similar to 1.4-MDa apoptosome complexes, *J. Biol. Chem.* **275**, 6067–6070.

Cecconi, F., Alvarezbolado, G., Meyer, B.I., Roth, K.A., Gruss, P. (1998) Apaf1 (ced-4 homolog) regulates programmed cell death in mammalian development, *Cell* **94**, 727–737.

Cheng, E., Wei, M.C., Weiler, S., Flavell, R.A., Mak, T.W., Lindsten, T., Korsmeyer, S.J. (2001) BCL-2, BCL-X-L sequester BH3 domain-only molecules preventing BAX- and BAK-mediated mitochondrial apoptosis, *Mol. Cell* **8**, 705–711.

Chinnaiyan, A.M., Tepper, C.G., Seldin, M.F., O'Rourke, K., Kischkel, F.C., Hellbardt, S., Krammer, P.H., Peter, M.E., Dixit, V.M. (1996) Fadd/mort1 is a common mediator of CD95 (Fas/APO-1) and tumor necrosis factor receptor-induced apoptosis, *J. Biol. Chem.* **271**, 4961–4965.

Chiou, S.K., Rao, L., White, E. (1994) bcl 2 blocks p53 dependent apoptosis, *Mol. Cell. Biol.* **14**, 2556–2563.

Christich, A., Kauppila, S., Chen, P., Sogame, N., Ho, S.I., Abrams, J.M. (2002) The damage-responsive *Drosophila* gene sickle encodes a novel IAP binding protein similar to but distinct from reaper, grim, and hid, *Curr. Biol.* **12**, 137–140.

Coleman, M.L., Sahai, E.A., Yeo, M., Bosch, M., Dewar, A., Olson, M.F. (2001) Membrane blebbing during apoptosis results from caspase-mediated activation of ROCK I, *Nat. Cell. Biol.* **3**, 339–345.

Conradt, B., Horvitz, H.R. (1998) The *C. elegans* protein EGL-1 is required for programmed cell death and interacts with the Bcl-2-like protein CED-9, *Cell* **93**, 519–529.

Davidson, F.F., Steller, H. (1998) Blocking apoptosis prevents blindness *Drosophila* retinal degeneration mutants, *Nature* **391**, 587–591.

Davidson, W.F., Giese, T., Fredrickson, T.N. (1998) Spontaneous development of plasmacytoid tumors in mice with defective fas-fas ligand interactions, *J. Exp. Med.* **187**, 1825–1838.

Day, C.L., Dupont, C., Lackmann, M., Vaux, D.L., Hinds, M.G. (1999) Solution structure and mutagenesis of the caspase recruitment domain (CARD) from Apaf-1, *Cell Death Differ.* **6**, 1125–1132.

Deng, C.X., Zhang, P.M., Harper, J.W., Elledge, S.J., Leder, P. (1995) Mice lacking p21(cip1/waf1) undergo normal development, but are defective in G1 checkpoint control, *Cell* **82**, 675–684.

Dierlamm, J., Baens, M., Wlodarska, I., Stefanova-Ouzounova, M., Hernandez, J.M., Hossfeld, D.K., De Wolf-Peeters, C., Hagemeijer, A., Van den Berghe, H., Marynen, P. (1999) The apoptosis inhibitor gene API2 and a novel 18q gene, MLT, are recurrently rearranged in the t(11;18)(q21;q21) associated with mucosa-associated lymphoid tissue lymphomas, *Blood* **93**, 3601–3609.

Dijkers, P.F., Birkenkamp, K.U., Lam, E.W.F., Thomas, N.S.B., Lammers, J.W.J., Koenderman, L., Coffer, P.J. (2002) FKHR-L1 can act as a critical effector of cell death induced by cytokine withdrawal: protein kinase B-enhanced cell survival through maintenance of mitochondrial integrity, *J. Cell Biol.* **156**, 531–542.

Du, C.Y., Fang, M., Li, Y.C., Li, L., Wang, X.D. (2000) Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation by eliminating IAP inhibition, *Cell* **102**, 33–42.

Ekert, P.G., Read, S.H., Silke, J., Marsden, V.S., Kaufmann, H., Hawkins, C.J., Gerl, R., Kumar, S., Vaux, D.L. (2004) Apaf-1 and caspase-9 accelerate apoptosis, but do not determine whether factor-deprived or drug-treated cells die, *J. Cell Biol.* **165**, 835–842.

Ellis, H.M., Horvitz, H.R. (1986) Genetic control of programmed cell death in the nematode *C. elegans*, *Cell* **44**, 817–829.

Enari, M., Hug, H., Nagata, S. (1995) Involvement of an ice-like protease in fas-mediated apoptosis, *Nature* **375**, 78–81.

Enari, M., Sakahira, H., Yokoyama, H., Okawa, K., Iwamatsu, A., Nagata, S. (1998) A caspase-activated dnase that degrades dna during apoptosis, and its inhibitor icad, *Nature* **391**, 43–50.

Hakem, R., Hakem, A., Duncan, G.S., Henderson, J.T., Woo, M., Soengas, M.S., Elia, A., Delapompa, J.L., Kagi, D., Khoo, W., Potter, J., Yoshida, R., Kaufman, S.A., Lowe, S.W., Penninger, J.M., Mak, T.W. (1998) Differential requirement for caspase 9 in apoptotic pathways in vivo, *Cell* **94**, 339–352.

Han, B.H., Xu, D.G., Choi, J.J., Han, Y.X., Xanthoudakis, S., Roy, S., Tam, J., Vaillancourt, J., Colucci, J., Siman, R., Giroux, A., Robertson, G.S., Zamboni, R., Nicholson, D.W., Holtzman, D.M. (2002) Selective, reversible caspase-3 inhibitor is neuroprotective and reveals distinct pathways of cell death after neonatal hypoxic-ischemic brain injury, *J. Biol. Chem.* **277**, 30128–30136.

Hanayama, R., Tanaka, M., Miwa, K., Shinohara, A., Iwamatsu, A., Nagata, S. (2002) Identification of a factor that links apoptotic cells to phagocytes, *Nature* **417**, 182–187.

Harlin, H., Reffey, S.B., Duckett, C.S., Lindsten, T., Thompson, C.B. (2001) Characterization of XIAP-deficient mice, *Mol. Cell. Biol.* **21**, 3604–3608.

Harris, C.A., Johnson, E.M. (2001) BH3-only Bcl-2 family members are coordinately regulated by the JNK pathway and require Bax to induce apoptosis in neurons, *J. Biol. Chem.* **276**, 37754–37760.

Hartley, S.B., Cooke, M.P., Fulcher, D.A., Harris, A.W., Cory, S., Basten, A., Goodnow, C.C. (1993) Elimination of self-reactive B lymphocytes proceeds in two stages: arrested development and cell death, *Cell* **72**, 325–335.

Hegde, R., Srinivasula, S.M., Zhang, Z.J., Wassell, R., Mukattash, R., Cilenti, L., DuBois, G., Lazebnik, Y., Zervos, A.S., Fernandes-Alnemri, T., Alnemri, E.S. (2002) Identification of Omi/HtrA-2 as a mitochondrial apoptotic serine protease that disrupts inhibitor of apoptosis protein-caspase interaction, *J. Biol. Chem.* **277**, 432–438.

Hildeman, D.A., Zhu, Y.N., Mitchell, T.C., Bouillet, P., Strasser, A., Kappler, J., Marrack, P. (2002) Activated T cell death in vivo mediated by proapoptotic Bcl-2 family member Bim, *Immunity* **16**, 759–767.

Hinds, M.G., Norton, R.S., Vaux, D.L., Day, C.L. (1999) Solution structure of a baculoviral inhibitor of apoptosis (IAP) repeat, *Nat. Struct. Biol.* **6**, 648–651.

Hotchkiss, R.S., Swanson, P.E., Knudson, C.M., Chang, K.C., Cobb, J.P., Osborne, D.F., Zollner, K.M., Buchman, T.G., Korsmeyer, S.J., Karl, I.E. (1999) Overexpression of Bcl-2 in transgenic mice decreases apoptosis and improves survival in sepsis, *J. Immunol.* **162**, 4148–4156.

Hotchkiss, R.S., Chang, K.C., Swanson, P.E., Tinsley, K.W., Hui, J.J., Klender, P., Xanthoudakis, S., Roy, S., Black, C., Grimm, E., Aspiotis, R., Han, Y., Nicholson, D.W., Karl, I.E. (2000) Caspase inhibitors improve survival in sepsis: a critical role of the lymphocyte, *Nat. Immunol.* **1**, 496–501.

Hu, S.M., Snipas, S.J., Vincenz, C., Salvesen, G., Dixit, V.M. (1998) Caspase-14 is a novel developmentally regulated protease, *J. Biol. Chem.* **273**, 29648–29653.

Huang, D.C.S., Hahne, M., Schroeter, M., Frei, K., Fontana, A., Villunger, A., Newton, K., Tschopp, J., Strasser, A. (1999) Activation of Fas by FasL induces apoptosis by a mechanism that cannot be blocked by Bcl-2 or Bcl-x(L), *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14871–14876.

Jones, J.M., Datta, P., Srinivasula, S.M., Ji, W., Gupta, S., Zhang, Z., Davies, E., Hajnoczky, G., Saunders, T.L., Van Keuren, M.L., Fernandes-Alnemri, T., Meisler, M.H., Alnemri, E.S. (2003) Loss of Omi protease activity causes the neuromuscular disorder of *mnd2* mutant mice, *Nature* **425**, 721–727.

Joza, N., Susin, S.A., Daugas, E., Stanford, W.L., Cho, S.K., Li, C.Y.J., Sasaki, T., Elia, A.J., Cheng, H.Y.M., Ravagnan, L., Ferri, K.F., Zamzami, N., Wakeham, A., Hakem, R., Yoshida, H., Kong, Y.Y., Mak, T.W., Zuniga-Pflucker, J.C., Kroemer, G., Penninger, J.M. (2001) Essential role of the mitochondrial apoptosis-inducing factor in programmed cell death, *Nature* **410**, 549–554.

Kagi, D., Vignaux, F., Ledermann, B., Burki, K., Depraetere, V., Nagata, S., Hengartner, H., Golstein, P. (1994) Fas and perforin pathways as major mechanisms of T cell-mediated cytotoxicity, *Science* **265**, 528–530.

Kischkel, F.C., Hellbardt, S., Behrmann, I., Germer, M., Pawlita, M., Krammer, P.H., Peter, M.E. (1995) Cytotoxicity-dependent APO-1 (Fas/CD95)-associated proteins form a death-inducing signaling complex (DISC) with the receptor, *EMBO J.* **14**, 5579–5588.

Kountouras, J., Boura, P., Lygidakis, N.J. (2001) Liver regeneration after hepatectomy, *HepatoGastroenterology* **48**, 556–562.

Kovacs, P., Bak, I., Szendrei, L., Vecsernyes, M., Varga, E., Blasig, I.E., Tosaki, A. (2001) Non-specific caspase inhibition reduces infarct size and improves post-ischaemic recovery in isolated ischaemic/reperfused rat hearts, *Naunyn Schmiedebergs Arch. Pharmacol.* **364**, 501–507.

Kuida, K., Haydar, T.F., Kuan, C.Y., Gu, Y., Taya, C., Karasuyama, H., Su, M., Rakic, P., Flavell, R.A. (1998) Reduced apoptosis and cytochrome c-mediated caspase activation in mice lacking caspase 9, *Cell* **94**, 325–337.

Lassus, P., Opitz-Araya, X., Lazebnik, Y. (2002) Requirement for caspase-2 in stress-induced apoptosis before mitochondrial permeabilization, *Science* **297**, 1352–1354.

Lens, S.M.A., Kataoka, T., Fortner, K.A., Tinel, A., Ferrero, I., MacDonald, R.H.,

Hahne, M., Beermann, F., Attinger, A., Orbea, H.A., Budd, R.C., Tschopp, J. (2002) The caspase 8 inhibitor c-FLIPL modulates T-cell receptor-induced proliferation but not activation-induced cell death of lymphocytes, *Mol. Cell. Biol.* **22**, 5419–5433.

Leonard, J.R., Klocke, B.J., D'Sa, C., Flavell, R.A., Roth, K.A. (2002) Strain-dependent neurodevelopmental abnormalities in caspase-3-deficient mice, *J. Neuropathol. Exp. Neurol.* **61**, 673–677.

Li, H.L., Zhu, H., Xu, C.J., Yuan, J.Y. (1998) Cleavage of bid by caspase 8 mediates the mitochondrial damage in the fas pathway of apoptosis, *Cell* **94**, 491–501.

Li, P., Nijhawan, D., Budihardjo, I., Srinivasula, S.M., Ahmad, M., Alnemri, E.S., Wang, X.D. (1997) Cytochrome c and dATP-dependent formation of apaf-1/caspase-9 complex initiates an apoptotic protease cascade, *Cell* **91**, 479–489.

Lindsten, T., Ross, A.J., King, A., Zong, W.X., Rathmell, J.C., Shiels, H.A., Ulrich, E., Waymire, K.G., Mahar, P., Frauwirth, K., Chen, Y.F., Wei, M., Eng, V.M., Adelman, D.M., Simon, M.C., Ma, A., Golden, J.A., Evan, G., Korsmeyer, S.J., MacGregor, G.R., Thompson, C.B. (2000) The combined functions of proapoptotic Bcl-2 family members Bak and Bax are essential for normal development of multiple tissues, *Mol. Cell* **6**, 1389–1399.

Liu, X.S., Zou, H., Slaughter, C., Wang, X.D. (1997) DFF, a heterodimeric protein that functions downstream of caspase-3 to trigger DNA fragmentation during apoptosis, *Cell* **89**, 175–184.

Lowin, B., Hahne, M., Mattmann, C., Tschopp, J. (1994) Cytolytic T-cell cytotoxicity is mediated through perforin and Fas lytic pathways, *Nature* **370**, 650–652.

Luo, X., Budihardjo, I., Zou, H., Slaughter, C., Wang, X.D. (1998) Bid, a Bcl2 interacting protein, mediates cytochrome c release from mitochondria in response to activation of cell surface death receptors, *Cell* **94**, 481–490.

Marsden, V.S., O'Connor, L., O'Reilly, L.A., Silke, J., Metcalf, D., Ekert, P.G., Huang, D.C.S., Cecconi, F., Kuida, K., Tomaselli, K.J., Roy, S., Nicholson, D.W., Vaux, D.L., Bouillet, P., Adams, J.M., Strasser, A. (2002) Apoptosis initiated by Bcl-2-regulated caspase activation independently of the cytochrome c/Apaf-1/caspase-9 apoptosome, *Nature* **419**, 634–637.

McDonnell, T.J., Korsmeyer, S.J. (1991) Progression from lymphoid hyperplasia to high-grade malignant lymphoma in mice transgenic for the t(14; 18), *Nature* **349**, 254–256.

Medema, J.P., Scaffidi, C., Kischkel, F.C., Shevchenko, A., Mann, M., Krammer, P.H., Peter, M.E. (1997) FLICE is activated by association with the CD95 death-inducing signalling complex (DISC), *EMBO J.* **16**, 2794–2804.

Muzio, M., Stockwell, B.R., Stennicke, H.R., Salvesen, G.S., Dixit, V.M. (1998) An induced proximity model for caspase-8 activation, *J. Biol. Chem.* **273**, 2926–2930.

Newton, K., Harris, A.W., Strasser, A. (2000) FADD/MORT1 regulates the pre-TCR checkpoint and can function as a tumour suppressor, *EMBO J.* **19**, 931–941.

Newton, K., Harris, A.W., Bath, M.L., Smith, K., Strasser, A. (1998) A dominant interfering mutant of FADD/Mort1 enhances deletion of autoreactive thymocytes and inhibits proliferation of mature t lymphocytes, *EMBO J.* **17**, 706–718.

Nuttall, M.E., Lee, D., McLaughlin, B., Erhardt, J.A. (2001) Selective inhibitors of apoptotic caspases: implications for novel therapeutic strategies, *Drug Discov. Today* **6**, 85–91.

Okada, H., Suh, W.K., Jin, J., Woo, M., Du, C., Elia, A., Duncan, G.S., Wakeham, A., Itie, A., Lowe, S.W., Wang, X., Mak, T.W. (2002) Generation and characterization of Smac/DIABLO-deficient mice, *Mol. Cell. Biol.* **22**, 3509–3517.

Pistritto, G., Jost, M., Srinivasula, S.M., Baffa, R., Poyet, J.L., Kari, C., Lazebnik, Y., Rodeck, U., Alnemri, E. (2002) Expression and transcriptional regulation of caspase-14 in simple and complex epithelia, *Cell Death Differ.* **9**, 995–1006.

Poe, M., Blake, J.T., Boulton, D.A., Gammon, M., Sigal, N.H., Wu, J.K., Zweerink, H.J. (1991) Human cytotoxic lymphocyte granzyme B. Its purification from granules and the characterization of substrate and inhibitor specificity, *J. Biol. Chem.* **266**, 98–103.

Puthalakath, H., Huang, D.C.S., O'Reilly, L.A., King, S.M., Strasser, A. (1999) The proapoptotic activity of the Bcl-2 family member Bim is regulated by interaction with the dynein motor complex, *Mol. Cell* **3**, 287–296.

Puthalakath, H., Villunger, A., O'Reilly, L.A., Beaumont, J.G., Coultas, L., Cheney, R.E., Huang, D.C.S., Strasser, A. (2001) Bmf: a proapoptotic BH3-only protein regulated by interaction with the myosin V actin motor complex, activated by anoikis, *Science* **293**, 1829–1832.

Qin, H.X., Srinivasula, S.M., Wu, G., Fernandes-Alnemri, T., Alnemri, E.S., Shi, Y.G. (1999) Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1, *Nature* **399**, 549–557.

Rampino, N., Yamamoto, H., Ionov, Y., Li, Y., Sawai, H., Reed, J.C., Perucho, M. (1997) Somatic frameshift mutations in the bax gene in colon cancers of the microsatellite mutator phenotype, *Science* **275**, 967–969.

Renatus, M., Stennicke, H.R., Scott, F.L., Liddington, R.C., Salvesen, G.S. (2001) Dimer formation drives the activation of the cell death protease caspase 9, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14250–14255.

Robertson, G.S., Crocker, S.J., Nicholson, D.W., Schulz, J.B. (2000) Neuroprotection by the inhibition of apoptosis, *Brain Pathol.* **10**, 283–292.

Sattler, M., Liang, H., Nettesheim, D., Meadows, R.P., Harlan, J.E., Eberstadt, M., Yoon, H.S., Shuker, S.B., Chang, B.S., Minn, A.J., Thompson, C.B., Fesik, S.W. (1997) Structure of Bcl-x(l)-Bak peptide complex – recognition between regulators of apoptosis, *Science* **275**, 983–986.

Scaffidi, C., Fulda, S., Srinivasan, A., Friesen, C., Li, F., Tomaselli, K.J., Debatin, K.M., Krammer, P.H., Peter, M.E. (1998) Two cd95 (apo-1/fas) signaling pathways, *EMBO J.* **17**, 1675–1687.

Shi, L., Kraut, R.P., Aebersold, R., Greenberg, A.H. (1992) A natural killer cell granule protein that induces DNA fragmentation and apoptosis, *J. Exp. Med.* **175**, 553–566.

Shresta, S., Russell, J.H., Ley, T.J. (1997) Mechanisms responsible for granzyme B-independent cytotoxicity, *Blood* **89**, 4085–4091.

Silke, J., Ekert, P.G., Day, C.L., Hawkins, C.J., Baca, M., Chew, J., Pakusch, M., Verhagen, A.M., Vaux, D.L. (2001) Direct inhibition of caspase 3 is dispensable for the anti-apoptotic activity of XIAP, *EMBO J.* **20**, 3114–3123.

Smith, K.G.C., Strasser, A., Vaux, D.L. (1996) CrmA expression in T lymphocytes of transgenic mice inhibits CD95 (Fas/APO-1)-transduced apoptosis, but does not cause lymphadenopathy or autoimmune disease, *EMBO J.* **15**, 5167–5176.

Srinivasula, S.M., Hegde, R., Saleh, A., Datta, P., Shiozaki, E., Chai, J.J., Lee, R.A., Robbins, P.D., Fernandes-Alnemri, T., Shi, Y.G., Alnemri, E.S. (2001) A conserved XIAP-interaction motif in caspase-9 and Smac/DIABLO regulates caspase activity and apoptosis, *Nature* **410**, 112–116.

Srinivasula, S.M., Datta, P., Kobayashi, M., Wu, J.W., Fujioka, M., Hegde, R., Zhang, Z.J., Mukattash, R., Fernandes-Alnemri, T., Shi, Y.G., Jaynes, J.B., Alnemri, E.S. (2002) Sickle, a novel *Drosophila* death gene in the reaper/hid/grim region, encodes an IAP-inhibitory protein, *Curr. Biol.* **12**, 125–130.

Strasser, A., Harris, A.W., Cory, S. (1991a) bcl-2 transgene inhibits T cell death and perturbs thymic self-censorship, *Cell* **67**, 889–999.

Strasser, A., Harris, A.W., Cory, S. (1993) E mu-bcl-2 transgene facilitates spontaneous transformation of early pre-B and immunoglobulin-secreting cells but not T cells, *Oncogene* **8**, 1–9.

Strasser, A., Harris, A.W., Bath, M.L., Cory, S. (1990) Novel primitive lymphoid tumours induced in transgenic mice by cooperation between myc and bcl-2, *Nature* **348**, 331–333.

Strasser, A., Harris, A.W., Corcoran, L.M., Cory, S. (1994a) Bcl-2 expression promotes B- but not T-lymphoid development in SCID mice, *Nature* **368**, 457–460.

Strasser, A., Harris, A.W., Jacks, T., Cory, S. (1994b) DNA damage can induce apoptosis in proliferating lymphoid cells via p53-independent mechanisms inhibitable by Bcl-2, *Cell* **79**, 329–339.

Strasser, A., Harris, A.W., Huang, D.C.S., Krammer, P.H., Cory, S. (1995) Bcl-2 and fas/apo-1 regulate distinct pathways to lymphocyte apoptosis, *EMBO J.* **14**, 6136–6147.

Strasser, A., Whittingham, S., Vaux, D.L., Bath, M.L., Adams, J.M., Cory, S., Harris, A.W. (1991b) Enforced BCL2 expression in B-lymphoid cells prolongs antibody responses and elicits autoimmune disease, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8661–8665.

Stennicke, H.R., Salvesen, G.S. (2000) Caspases – controlling intracellular signals by protease zymogen activation, *Biochim. Biophys. Acta – Protein Struct. Mol. Enzymol.* **1477**, 299–306.

Stennicke, H.R., Deveraux, Q.L., Humke, E.W., Reed, J.C., Dixit, V.M., Salvesen, G.S. (1999) Caspase-9 can be activated without proteolytic processing, *J. Biol. Chem.* **274**, 8359–8362.

Sun, X.M., MacFarlane, M., Zhuang, J.G., Wolf, B.B., Green, D.R., Cohen, G.M. (1999b) Distinct caspase cascades are initiated in receptor-mediated and chemical-induced apoptosis, *J. Biol. Chem.* **274**, 5053–5060.

Sun, C., Cai, M., Gunasekera, A.H., Meadows, R.P., Wang, H., Chen, J., Zhang, H., Wu, W., Xu, N., Ng, S.C., Fesik, S.W. (1999a) NMR structure and mutagenesis of the inhibitor-of-apoptosis protein XIAP, *Nature* **401**, 818–821.

Sutton, V.R., Davis, J.E., Cancilla, M., Johnstone, R.W., Ruefli, A.A., Sedelies, K., Browne, K.A., Trapani, J.A. (2000) Initiation of apoptosis by granzyme B requires direct cleavage of Bid, but not direct granzyme B-mediated caspase activation, *J. Exp. Med.* **192**, 1403–1413.

Suzuki, M., Youle, R.J., Tjandra, N. (2000) Structure of Bax: coregulation of dimer formation and intracellular localization, *Cell* **103**, 645–654.

Thornberry, N.A., Ranon, T.A., Pieterson, E.P., Rasper, D.M., Timkey, T., Garciacalvo, M., Houtzager, V.M., Nordstrom, P.A., Roy, S., Vaillancourt, J.P., Chapman, K.T., Nicholson, D.W. (1997) A combinatorial approach defines specificities of members of the caspase family and granzyme B – functional, relationships established for key mediators of apoptosis, *J. Biol. Chem.* **272**, 17907–17911.

Traver, D., Akashi, K., Weissman, I.L., Lagasse, E. (1998) Mice defective in two apoptosis pathways in the myeloid lineage develop acute myeloblastic leukemia, *Immunity* **9**, 47–57.

Troy, C.M., Rabacchi, S.A., Hohl, J.B., Angelastro, J.M., Greene, L.A., Shelanski, M.L. (2001) Death in the balance: alternative participation of the caspase-2 and -9 pathways in neuronal death induced by nerve growth factor deprivation, *J. Neurosci.* **21**, 5007–5016.

Vaux, D.L., Hacker, G. (1995) Hypothesis – apoptosis caused by cytotoxins represents a defensive response that evolved to combat intracellular pathogens, *Clin. Exp. Pharmacol. Physiol.* **22**, 861–863.

Vaux, D.L., Cory, S., Adams, J.M. (1988) Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells, *Nature* **335**, 440–442.

Veis, D.J., Sorenson, C.M., Shutter, J.R., Korsmeyer, S.J. (1993) Bcl-2-deficient mice demonstrate fulminant lymphoid apoptosis, polycystic kidneys, and hypopigmented hair, *Cell* **75**, 229–240.

Verhagen, A.M., Ekert, P.G., Pakusch, M., Silke, J., Connolly, L.M., Reid, G.E., Moritz, R.L., Simpson, R.J., Vaux, D.L. (2000) Identification of DIABLO, a mammalian protein that promotes apoptosis by binding to and antagonizing IAP proteins, *Cell* **102**, 43–53.

Verhagen, A.M., Silke, J., Ekert, P.G., Pakusch, M., Kaufmann, H., Connolly, L.M., Day, C.L., Tikoo, A., Burke, R., Wrobel, C., Moritz, R.L., Simpson, R.J., Vaux, D.L. (2002) HtrA2 promotes cell death through its serine protease activity and its ability to antagonize inhibitor of apoptosis proteins, *J. Biol. Chem.* **277**, 445–454.

Verhoven, B., Krahling, S., Schlegel, R.A., Williamson, P. (1999) Regulation of phosphatidylserine exposure and phagocytosis of apoptotic T lymphocytes, *Cell Death Differ.* **6**, 262–270.

Verma, S., Zhao, L.J., Chinnadurai, G. (2001) Phosphorylation of the pro-apoptotic protein BIK – mapping of phosphorylation sites and effect on apoptosis, *J. Biol. Chem.* **276**, 4671–4676.

Villunger, A., O'Reilly, L.A., Holler, N., Adams, J., Strasser, A. (2000) Fas ligand, Bcl-2, granulocyte colony-stimulating factor, and p38 mitogen-activated protein kinase: regulators of distinct cell death and survival pathways in granulocytes, *J. Exp. Med.* **192**, 647–657.

Wang, X.D. (2001) The expanding role of mitochondria in apoptosis, *Genes. Dev.* **15**, 2922–2933.

Wei, M.C., Zong, W.X., Cheng, E.H.Y., Lindsten, T., Panoutsakopoulou, V., Ross, A.J., Roth, K.A., MacGregor, G.R., Thompson, C.B., Korsmeyer, S.J. (2001) Proapoptotic BAX and BAK: a requisite gateway to mitochondrial dysfunction and death, *Science* **292**, 727–730.

Wing, J.P., Karres, J.S., Ogdahl, J.L., Zhou, L., Schwartz, L.M., Nambu, J.R. (2002) *Drosophila* sickle is a novel grim-reaper cell death activator, *Curr. Biol.* **12**, 131–135.

Yang, Y., Fang, S.Y., Jensen, J.P., Weissman, A.M., Ashwell, J.D. (2000) Ubiquitin protein ligase activity of IAPs and their degradation in proteasomes in response to apoptotic stimuli, *Science* **288**, 874–877.

Yonish, R.E., Resnitzky, D., Lotem, J., Sachs, L., Kimchi, A., Oren, M. (1991) Wild-type p53 induces apoptosis of myeloid leukaemic cells that is inhibited by interleukin-6, *Nature* **353**, 345–347.

Yoshida, H., Kong, Y.Y., Yoshida, R., Elia, A.J., Hakem, A., Hakem, R., Penninger, J.M., Mak, T.W. (1998) Apaf1 is required for mitochondrial pathways of apoptosis and brain development, *Cell* **94**, 739–750.

Zha, J.P., Harada, H., Yang, E., Jockel, J., Korsmeyer, S.J. (1996) Serine phosphorylation of death agonist Bad in response to survival factor results in binding to 14-3-3 not Bcl-x(l), *Cell* **87**, 619–628.

Zhang, J.K., Cado, D., Chen, A., Kabra, N.H., Winoto, A. (1998) Fas-mediated apoptosis and activation-induced T-cell proliferation are defective in mice lacking FADD/Mort1, *Nature* **392**, 296–300.

Zong, W.X., Lindsten, T., Ross, A.J., MacGregor, G.R., Thompson, C.B. (2001) BH3-only proteins that bind pro-survival Bcl-2 family members fail to induce apoptosis in the absence of Bax and Bak, *Genes. Dev.* **15**, 1481–1486.

Zou, H., Henzel, W.J., Liu, X.S., Lutschg, A., Wang, X.D. (1997) Apaf-1, a human protein homologous to c-elegans ced-4, participates in cytochrome c-dependent activation of caspase-3, *Cell* **90**, 405–413.

Prokaryotic and Eukaryotic Cells in Biotech Production

Irmgard Merfort[1], *Stefan Pelzer*[2], *Dirk Hoffmeister*[1], *and Andreas Bechthold*[1]
[1]*Albert-Ludwigs-Universität Freiburg, Freiburg, Germany*
[2]*Combinature Biopharm AG, Berlin, Germany*

1 Actinomycetes in Biotech Production 641
1.1 Actinomycetes: Producer of Commercially Important Drugs 641
1.2 Actinomycetes Genetics: The Basis for Understanding Antibiotic Biosynthesis 644
1.3 Urgent Need for the Development of New Antimicrobial Drugs 646
1.4 Strategies for the Identification and Development of New Antimicrobial Drugs 646
1.4.1 Approaches to Explore Nature's Chemical Diversity 646
1.4.2 Exploiting the Enormous Genotypic Potential of Actinomycetes by "Genome Mining" 647
1.4.3 Generation of Novel Antibiotics by Targeted Manipulation of the Biosynthesis (Combinatorial Biosynthesis) 647
1.4.4 Novel Natural Compounds by Glycorandomization 649
1.4.5 Novel Natural Compounds by Mutasynthesis 653

2 Saccharomyces Cerevisiae and Other Fungi in Biotech Production 654
2.1 Generation of Engineered Strains of *Saccharomyces Cerevisiae* for the Production of Alcoholic Beverages 654
2.2 Generation of Engineered Strains of *Saccharomyces Cerevisiae* for Lactic Acid, Xylitol, and Strictosidine Production 655
2.3 The Use of Fungi in the Production of Statins, Cyclosporin, and ß-Lactam Antibiotics 655

3 Plants in Biotech Production 657
3.1 Transgenic Plants as Functional Foods or Neutraceuticals 657
3.2 Transgenic Plants and Plant Cell Culture as Bioreactors of Secondary Metabolites 658
3.3 Transgenic Plants as Bioreactors of Recombinant Proteins 660

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 10
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30552-1

Bibliography 661
Books and Reviews 661
Primary Literature 661

Keywords

Combinatorial Biosynthesis
Genetic method used to generate novel natural compounds.

Functional Food
Food that is selected and advertised for a high content of therapeutically active molecules.

Genetic Engineering
Process of manipulating genes in an organism, mostly outside of its reproductive process.

Genome Mining
Genetic method to detect strains that are able to produce a given compound.

Glycorandomization
Method to generate novel glycosylated products.

Metabolic Engineering
Modification of cellular metabolism and properties by introducing, deleting, and modifying metabolic pathways.

Molecular Farming
Use of plants as biopharming systems for production of pharmaceutical proteins.

Mutasynthesis
Process to produce novel compounds by feeding an alternative precursor to a mutated strain.

Polyketides
Natural compounds derived from acetate.

Streptomyces
Bacteria, used for the production of antibiotics.

Transgenic Plant
Plant transfected with plasmids as vectors carrying the desired gene.

The production of compounds used in the food and pharmaceutical industries by biotech processes is both an old and a very young business. Over the past 70 years, fermentation of microorganisms or the use of yeast and plants in the production of important pharmaceuticals has been well established. The promises of genomics in drug discovery and drug production have now been fulfilled in many areas.

Bacterial strains, especially Actinomycetes have been used in biotech production and drug discovery for years. Genetic methods now open the field of combinatorial biosynthesis that has improved impressingly in the past couple of years. Also, the productivity of yeast and other fungi in a variety of different processes has improved significantly since genetic methods have been introduced. In addition, a number of recent works considerably widens the potential of plant biotechnology. This review covers examples describing the use of prokaryotic cells and plant cells in biotech production.

1 Actinomycetes in Biotech Production

Soil bacteria of the order Actinomycetes are the most important producers of pharmaceutically relevant bioactive metabolites including antibiotics, antitumor agents, immunosuppressants, antiparasitic agents, herbicides, and enzyme-inhibiting agents. The success story of these bacteria began about 60 years ago with the groundbreaking work of Waksman, who discovered and described streptomycin as the first antibiotic synthesized by an Actinomycete. Ever since, systematic large-scale screens performed by the pharmaceutical industry have revealed numerous therapeutically relevant drugs. More than two-thirds of all naturally derived antibiotics currently used are produced by Actinomycetes strains, underlining their importance to medicine.

1.1 Actinomycetes: Producer of Commercially Important Drugs

Natural products (''secondary metabolites'') have been the largest contributors to drugs in the history of medicine. Before antibiotics were introduced in the 1940s and 1950s (see above), patients with bacteremia faced low survival chances, and the mortality from tuberculosis was 50%. It has been stated that the doubling of our life span in the twentieth century is mainly due to the use of plant and microbial secondary metabolites. Of the 520 new drugs approved between 1983 and 1994, 39% were natural products or those derived from natural products and 60 to 80% of antibacterial and anticancer drugs were derived from natural products. Almost half of the best-selling pharmaceuticals are natural or related to them. In 2001, over 100 natural product–derived compounds were in clinical development. Natural products and their derivatives account for annual revenues of about US$30 billion in the anti-infectives market, US$20 billion in the anticancer market, and US$14 billion in the lipid-lowering market. Actinomycetes and, particularly, Streptomycetes (Fig. 1) are the largest antibiotic-producing genus in the microbial world discovered so far. Of the 12 000 or so antibiotics known in 1995, 55% were produced by Streptomycetes and an additional 11% by other

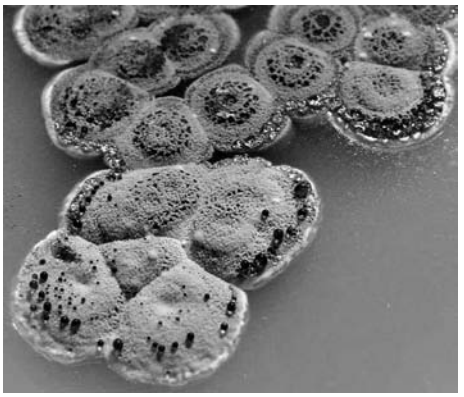

Fig. 1 Photography of a sporulated streptomyces strain growing on solid medium. The blue drops indicate the production of an antibiotic (aromatic polyketide). (See color plate. p. xxii)

Actinomycetes. A compilation of numerous bioactive and commercially important metabolites, which are all synthesized by Actinomycetes strains, is shown in Table 1. This list includes not only very important drugs such as the macrolide erythromycin A synthesized by *Saccharopolyspora (Sac.) erythraea* (in 2000, the annual sales of semisynthetic derivatives reached US$2.6 billion), the glycopeptide vancomycin synthesized by *Amycolatopsis (A.) orientalis* (in 2000, the annual sales of glycopeptides reached US$424 million,) and tetracycline synthesized by *Streptomyces (S.) aureofaciens* (in 2000, the annual sales reached US$217 million) but also anticancer agents like doxorubicin synthesized by *Streptomyces peucetius* (Fig. 2). Many compounds produced by Actinomycetes belong to the large family of polyketides. Polyketides are structurally diverse (Fig. 2) and exhibit a wide scope of bioactivities. More than 500 aromatic polyketides have been characterized from

Tab. 1 Origin, target, and application of commercially important secondary metabolites originating from actinomycetes.

Antibiotic	*Producer*	*Molecular target*	*Application*
A40 926	*Nonomurea* sp.	Cell wall synthesis	Antibacterial
Amphotericin	*S. nodosus*	Membrane (pore-formation)	Antifungal
Ascomycin (FK520)	*S. hygroscopicus*	FKBP12	Immunosuppressive
Avermectin	*S. avermitilis*	Membrane (ion-channel)	Antiparasitic
Avilamycin	*S. viridochromogenes*	Ribosome	Antibacterial
Avoparcin	*A. coloradensis*	Cell wall synthesis	Antibacterial growth promotant
Bleomycin	*S. verticillus*	DNA binding	Antitumor
Bialaphos	*S. hygroscopicus*	Glutamine synthetase	Herbicide
Candicidin	*S. griseus*	Membrane (pore-formation)	Antifungal

Tab. 1 (*Continued*)

Antibiotic	***Producer***	***Molecular target***	***Application***
Clavulanic acid	*S. clavuligerus*	Beta-lactamases	Combined with β-lactam antibacterial
Chloramphenicol	*S. venezuelae*	Ribosome	Antibacterial
Chlortetracycline	*S. aureofaciens*	Ribosome	Antibacterial
Cyclohexamide	*S. griseus*	Ribosome	Antibiotic
Dactinomycin	*S. parvulus*	DNA intercalation	Antitumor
Daptomycin	*S. roseosporus*	Cell wall synthesis	Antibacterial
Daunorubicin	*S. peucetius*	DNA intercalation	Antitumor
Doxorubicin	*S. peucetius var caesius*	DNA intercalation	Antitumor
Erythromycin A	*Saccharopolyspora erythraea*	Ribosome	Antibacterial
Gentamicin	*Micromonospora purpurea*	Ribosome	Antibacterial
Geldanamycin	*S. hygroscopicus*	Hsp90	Antitumor
Kanamycin	*S. kanamyceticus*	Ribosome	Antibacterial
Lincomycin	*S. lincolnensis*	Ribosome	Antibacterial
Milbemycin	*S. hygroscopicus*	Membrane (ion-channel)	Antiparasitic
Mithramycin	*S. argillaceus*	DNA alkylation	Antitumor
Mitomycin C	*S. lavendulae*	DNA alkylation	Antitumor
Moenomycin	*S. ghanaensis*	Cell wall synthesis	Antibacterial, growth promotant
Monensin	*S. cinnamonensis*	Membrane (ionophore)	Anticoccidial, growth promotant
Natamycin	*S. nataensis*	Membrane (pore-formation)	Antifungal
Neomycin	*S. fradiae*	Ribosome	Antibacterial
Nikkomycin	*S. tendae*	Chitin synthase	Antifungal
Novobiocin	*S. niveus*	DNA gyrase	Antibacterial
Nystatin	*S. noursei*	Membrane (pore-formation)	Antifungal
Oxytetracycline	*S. rimosus*	Ribosome	Antibacterial
Pristinamycin	*S. pristinaespiralis*	Ribosome	Antibacterial
Ramoplanin	*Actinoplanes spec.*	Cell wall synthesis	Antibacterial
Rapamycin	*S. hygroscopicus*	FKBP	Immunosuppressive
Rifamycin	*A. mediterranei*	RNA polymerase	Antibacterial
Salinomycin	*S. albus*	Membrane (ionophore)	Anticoccidial, growth promotant
Spinosyn	*Sac. spinosa*	unknown	Insecticidal
Spiramycin	*S. ambofaciens*	Ribosome	Antibacterial
Staurosporin	*S. staurosporeus*	Protein kinase C	Antibacterial
Streptomycin	*S. griseus*	Ribosome	Antibacterial
Tacrolimus (FK506)	*Streptomyces spec.*	FKBP	Immunosuppressive
Teicoplanin	*A. teicomyceticus*	Cell wall synthesis	Antibacterial
Tetracycline	*S. aureofaciens*	Ribosome	Antibacterial
Thienamycin	*S. cattleya*	Cell wall synthesis	Antibacterial
Tylosin	*S. fradiae*	Ribosome	Growth promotant
Vancomycin	*A. orientalis*	Cell wall synthesis	Antibacterial
Virginiamycin	*S. virginiae*	Ribosome	Growth promotant

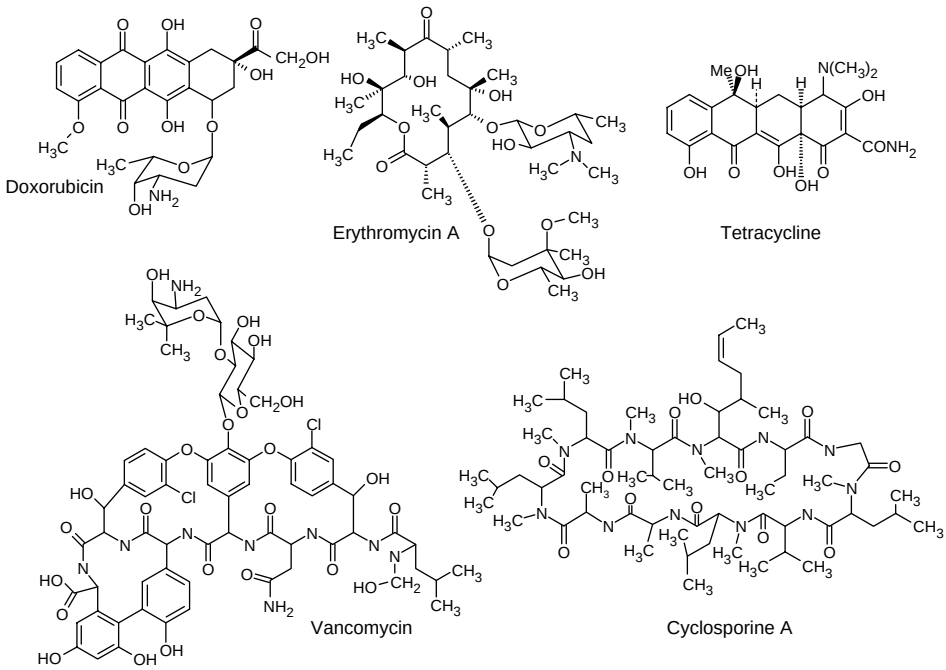

Fig. 2 Structures of natural compounds used as important drugs in pharmacy and medicine.

Actinomycetes. Polyketides are particularly important for drug discovery, since statistics show that 1 out of 100 polyketides will make its way to commercialization. With an average of as low as 1 out of 5000 compounds, other substances are far less likely to hit the market. Sales of drugs based on polyketides exceed US$15 billion a year. In general, for industrial production, overproducing strains have to be developed. Today, modern processes allow the production of compounds at concentrations even higher than 10 g L^{-1}.

1.2 Actinomycetes Genetics: The Basis for Understanding Antibiotic Biosynthesis

Streptomyces coelicolor A3 (2) is the genetically best-characterized strain among the filamentous Actinomycetes. Actinomycetes genetics has been the subject of research since 1958 when Prof. Dr. Sir D. Hopwood published the first linkage map of *S. coelicolor* performing the first genetic recombination experiments considering six marker genes. Afterwards, Actinomycetes genetics developed continuously, by the identification of mutants interrupted in the biosynthesis of actinorhodin. After identification and isolation of easily selectable antibiotic resistance genes, the first gene cloning in *Streptomyces* was described in 1980. In 1984, Malpartida and Hopwood demonstrated for the first time that antibiotic biosynthesis genes are usually organized as a gene cluster of structural, regulatory, export, and self-resistant genes. Hence, once a single gene within a cluster has been located, the others may be identified quickly by chromosomal

walking. In the course of the last two decades, many molecular tools including vector systems (phage and plasmid-based) have been developed along with DNA transfer and gene inactivation techniques, which were all necessary for targeted manipulation of Actinomycetes. The excellent manual published by the John Innes Institute summarizes all the necessary information related to handling and molecular genetic tools essential for working with Actinomycetes. Several hundred biosynthetic gene clusters have so far been identified and genes encoding about 80 pathways for secondary metabolites have been cloned, at least partially sequenced and made available to the public. The average size of an antibiotic biosynthetic gene cluster ranges from about 20 kb for a simple aromatic polyketide like actinorhodin to 120 kb for complex polyketides like the antifungal antibiotic nystatin.

A highlight in Actinomycetes genetics was the completion of the first genome sequence of the model Actinomycete *S. coelicolor*A3(2) in July 2001. Recently, the genome sequence of a second *Streptomyces* strain, the avermectin producer *Streptomyces avermitilis*, has been published, opening up new perspectives for comparative genomics with Actinomycetes. A characteristic feature of Actinomycete chromosomes is their linear structure. The genome of *S. coelicolor* comprises 8 667 507 bp (G/C content of 72.1%), whereas the *S. avermitilis* genome contains 9 025 608 bp (G/C content of 70.7%). Both genomes are densely packed and harbor 7574 ORFs in *S. avermitilis* and 7825 ORFs in *S. coelicolor*, respectively. Comparative analysis of the *S. coelicolor* and *S. avermitilis* chromosomes revealed that both the genomes had an unusual biphasic structure with a core region of 5 and 6.5 Mb, respectively. The most interesting feature in both the completed *Streptomyces* genomes that will impact biotechnology is the abundance of secondary metabolite gene clusters. Before the genome of *S. coelicolor* was sequenced, three antibiotics and a spore pigment were known to be synthesized from this strain. The genome sequence revealed that 23 gene clusters (about 5% of the total genome) are directly dedicated to secondary metabolism including clusters for further putative antibiotics, pigments, complex lipids, signaling molecules, and iron-scavenging siderophores. In *S. avermitilis*, 30 gene clusters related to secondary metabolites were identified, corresponding to 6.6% of the genome. From these clusters, 5 out of 30 are putatively involved in pigments and siderophores, 5 in terpenes, 8 in nonribosomal peptides, and 12 in polyketide biosynthesis. With avermectin, oligomycin, and a polyene antibiotic, only three complex polyketide clusters have been characterized from this strain before. The completed genome-sequence data can also be used to study the regulatory network of primary metabolism pathways and the cross talk between primary and secondary metabolism (i.e. the carbon flux). The knowledge gained by these analyses will be useful for the construction of improved strains produced in a rational approach by deleting undesired pathways or adding advantageous pathways, generating precursors and essential cofactors. Moreover, targeted modifications will improve cell growth and fermentation properties (metabolic engineering). Successful metabolic engineering of a strain producing doramectin, a commercial antiparasitic avermectin analog, is an excellent example of the importance of this technology.

1.3 Urgent Need for the Development of New Antimicrobial Drugs

Stimulated by the discovery of numerous novel antibacterial agents, which reached a peak in the 1970s, US Surgeon General William Stewart declared in 1969 in the US Congress that it was "time to close the book on infection diseases." Today, unfortunately, we know that antibiotics have not won the fight against infectious microorganisms and therefore there is a permanent need for new antibiotics.

One main reason for this development is the problem of emerging resistant forms of pathogens. As an example, according to the WHO more than 95% of *Staphylococcus aureus* strains worldwide are resistant to penicillin G, and up to 60% are resistant to its derivative methicillin. Several reasons (e.g. use of antibiotics as growth promoters, changes in the spectrum of pathogens) are responsible for this development.

The past decades witnessed a major decrease in the number of newly discovered compounds. In an almost 40-year period (1962–2000), no new class of antibiotic was introduced to the market (nalidixic acid in 1962, the oxazolidinone antibiotic linezolid in 2000). The major reason for the decrease in the number of newly discovered compounds might be a decline in screening efforts. Ironically, some of the leading pharmaceutical companies are currently cutting back their antiinfective programs, especially for natural products. They would rather focus their activities on the semisynthetic modification of existing antibiotics to produce second- and third-generation antibiotics with improved properties.

Nevertheless, there is no need to resign. According to biomathematical modeling, only 3% of all antibacterial agents synthesized in *Streptomyces* have been reported so far. Additionally, less than 10% of the world's biodiversity has been tested for biological activity, and many more useful natural lead compounds are yet to be discovered.

1.4 Strategies for the Identification and Development of New Antimicrobial Drugs

1.4.1 Approaches to Explore Nature's Chemical Diversity

Vicuron Pharmaceuticals Inc., formerly Biosearch Italia, a company screening for new antibiotics, focuses its activities on a proprietary strain collection of 50 000 microorganisms, including unusual filamentous Actinomycetes and filamentous fungi or strains that are difficult to isolate. The rationale behind this campaign is that these organisms have not been intensively screened in the past and that they may be producers of novel compound classes.

Another strategy to reveal the chemical diversity of a single strain is the One strain – many Compounds (OSMAC) approach described by Bode et al. By systematic alteration of cultivation parameters, the number of secondary metabolites increased tremendously in a single strain. When this method was applied, up to 20 different metabolites with, in some cases, high production titers were detected. Since recent estimates suggest that only 0.1 to 1% of the microbial flora in the environment can be kept in culture, the "metagenome" of the unculturable microorganisms should also have a potential to generate novel secondary metabolites. Indeed, several reports demonstrated that it is possible to construct DNA libraries from "soil-DNA" and to use them for the production of novel metabolites in a heterologous *Streptomyces* host.

1.4.2 Exploiting the Enormous Genotypic Potential of Actinomycetes by "Genome Mining"

The completion of the sequence of the two *Streptomyces* genomes demonstrated that between 5 and 6.6% of the whole genome are directly involved in the biosynthesis of predominantly unknown secondary metabolites (see above). Prior to genome sequencing, a number of reports were published in which cryptic or silent secondary metabolite pathways were identified during the search for gene clusters for known metabolites. Hence, the occurrence of multiple "orphan" gene clusters has been reported for various compound classes like nonribosomal peptides, PKSI and PKSII. Combinature Biopharm AG is a Berlin-based company using modern high-throughput genomics for the systematic genetic screening of several Actinomycetes genomes to identify known and "orphan" clusters. Recently, Zazopoulos et al. described how a genomics-guided approach can be rewarding for the discovery and expression of cryptic metabolic pathways (genome mining). The genetic information of these biosynthesis clusters is used for the targeted generation and modification of novel compounds in an approach termed "*combinatorial biosynthesis.*"

1.4.3 Generation of Novel Antibiotics by Targeted Manipulation of the Biosynthesis (Combinatorial Biosynthesis)

Researchers have started using biosynthetic genes to alter the structure of natural compounds by genetic engineering or to combine genes from different biosynthetic pathways. This new technology named *combinatorial biosynthesis* results in the formation of novel natural products.

New drugs by targeted gene disruption Inactivation of specific selected genes is a very common methodology for the generation of novel structural variations of known natural products. Erythromycin is a macrolide antibiotic that is clinically useful in the treatment of infections by gram-positive bacteria. A hydroxyl group at C6 of the erythronolide macrolactone is responsible for acidic inactivation in the stomach by conversion into anhydroerythromycin. Erythromycin derivatives lacking this hydroxyl group are therefore interesting from the therapeutic and pharmacological point of view. The gene *eryF* that encodes a cytochrome P450 monooxygenase responsible for the introduction of this hydroxyl group into the macrolactone was inactivated and the mutant produced 6- deoxyerythromycin A. This is a much more acid-stable antibiotic and as efficient as erythromycin because of its higher stability.

The orthosomycins are a prominent class of antibiotics produced by various Actinomycetes. Members of this class are active against a broad range of gram-positive pathogenic bacteria. Prominent examples of orthosomycins are the avilamycins and the everninomicins produced by *Streptomyces viridochromogenes* Tü57 and *Micromonospora carbonacea*, respectively. Avilamycins and everninomicins are poorly soluble in water, which poses a major obstacle for their use as therapeutics. The avilamycin biosynthetic gene cluster has been cloned and sequenced. Several putative methyltransferase genes have been found in the cluster. Double and triple mutants have been generated by deleting two or three methyltransferase genes in the chromosome of the producer strain (Fig. 3). All mutants produced novel avilamycin derivatives with improved water solubility.

Improved yield by expression of genes Pristinamycin, produced by *Streptomyces*

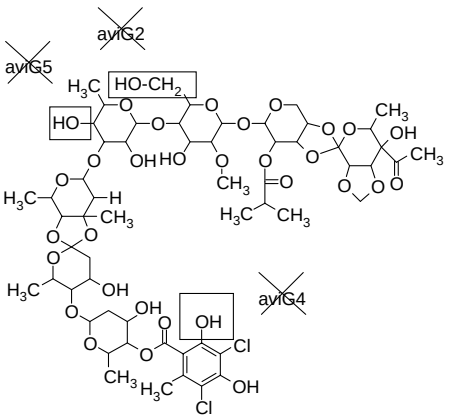

Fig. 3 Production of gavibamycin H1 by a mutant with deletions in three methyltransferase genes

pristinaespiralis, is a mixture of two types of macro-cyclic lactone peptolides, pristinamycins I (PI), a branched cyclic hexadepsipeptide of the streptogramin B group, and pristinamycins II (PII), a polyunsaturated cyclic peptolide of the streptogramin A group. Both the compounds inhibit the growth of bacteria. In combination, they display a synergistic bactericidal activity. The PII component of pristinamycin is produced mainly in two forms, called *PIIA* (80%) *and PIIB* (20%). A water-soluble derivative of pristinamycin, now being marketed under the trade name Synercid, was obtained by the chemical modification of PIIA. To generate a PIIA-specific producer strain, two genes, *snaA* and *snaB*, were isolated from the biosynthetic gene cluster. The enzymes encoded by *snaA* and *snaB* catalyze the conversion of PIIB to PIIA. Both genes were placed under the transcription control of a strong promoter and were cloned into an integrative vector. The integration of this vector into the chromosome of the producer strain resulted in the production of 100% PIIA and this was achieved in high concentrations.

New drugs by expression of genes Mithramycin is an aromatic polyketide, which is clinically used as an anticancer agent. It possesses a tricyclic chromophore and is glycosylated at two different positions. Urdamycin A is an angucycline polyketide produced by *Streptomyces fradiae* Tü2717, which also shows antitumor activity. It consists of the aglycon aquayamycin, which contains a C-glycosidically linked D-olivose, and three additional O-glycosidically linked deoxyhexoses. The UrdGT2 glycosyltransferase catalyzes the C-glycosyl transfer of activated D-olivose as the first glycosylation step during the urdamycin biosynthesis. Landomycins are produced by *S Streptomyces cyanogenus* S136 and contain an unusual hexasaccharide consisting of four D-olivose and two L-rhodinose units. These polyketides also show antitumor activities, in particular, against prostate cancer cell lines. To generate novel compounds, genes out of the urdamycin and landomycin clusters were expressed in mutants of *Streptomyces argillaceus*: coexpression of *urdGT2* (urdamycin biosynthesis) together with *lanGT1*, (landomycin biosynthesis) in a mutant of the mithramycin producer led to the hybrid molecule 9-C-diolivosyl-premithramycinone. This example was

listed as a highlight in the field of combinatorial biosynthesis as genes from three different organisms yielded a rationally designed product (Fig. 4).

Recently, a plasmid-based strategy has been described that allows the use of deoxysugar biosynthetic genes to produce a variety of deoxysugars in a cell, which can then be attached to an aglycon by the use of different glycosyltransferases. As an example, a plasmid was generated harboring all the genes necessary for the biosynthesis dTDP-D-olivose. This plasmid was coexpressed with the highly substrate-flexible glycosyltransferase gene *elmG* in *S. albus*. When 8-demethyl-tetracenomycin C was fed to this strain, D-olivosyltetracenomycin was produced. In a similar way, L-rhamnosyl-tetracenomycin C, L-olivosyl-tetracenomycin C, and L-rhodinosyl-tetracenomycin C were generated depending on the deoxysugar biosynthetic genes used in each case.

Polyketides are synthesized by the action of polyketide synthases (PKSs), which have been classified into two types, type I (modular PKSs) and type II (iterative PKSs).

Modular PKSs are large multifunctional enzymes. Active sites (domains) within these enzymes ketosynthases (KS), acyltransferases (AT), dehydratases (DH), enoyl reductases (ER), ketoreductases (KR), acyl carrier proteins (ACP) and thioesterases (TE) are organized into modules such that each module catalyzes the stereospecific addition of a new monomer onto a growing polyketide chain and also sets the reduction level of the carbon atoms of the resulting intermediate. In 1994, the heterologous expression of the complete erythromycin polyketide synthase was accomplished. The recombinant strain produced 6-deoxyerythronolide B. This polyketide synthase was then used for a variety of experiments in which modules/domains of the polyketide synthase were exchanged. As an example, compounds produced by the substitution of KR domains are shown in Fig. 5. In these examples, KR domains from the erythromycin PKS have been replaced by domains from the rapamycin producer.

New drugs by expression of "Artificial" genes UrdGT1b and UrdGT1c, involved in urdamycin biosynthesis, share 91% identical amino acids. However, the two GTs show different specificities for both nucleotide sugar and acceptor substrate. Targeted amino acid exchanges reduced the number of amino acids, potentially dictating the substrate specificity to 10 in either enzyme. Subsequently, a gene library was created such that only codons of these 10 amino acids from both parental genes were independently combined. Screening of almost 600 library members *in vivo* revealed 40 active members, acting either like the parental enzymes UrdGT1b and UrdGT1c or displaying a novel specificity. The novel enzymatic specificity is responsible for the biosynthesis of urdamycin P carrying a branched saccharide side chain, hitherto unknown for urdamycins (Fig. 6).

1.4.4 Novel Natural Compounds by Glycorandomization

As discussed above, combinatorial biosynthesis represents an *in vivo* methodology to diversify natural products, and, among them, numerous glycosides. It is superior to traditional synthetic chemistry in that it makes use of specific enzymes for efficient modifications of complex scaffolds and avoids solubility problems. The reactions are fed out of the host's metabolism, and its cytoplasm is used as an aqueous phase instead of toxic organic solvents that need to be carefully removed from the product

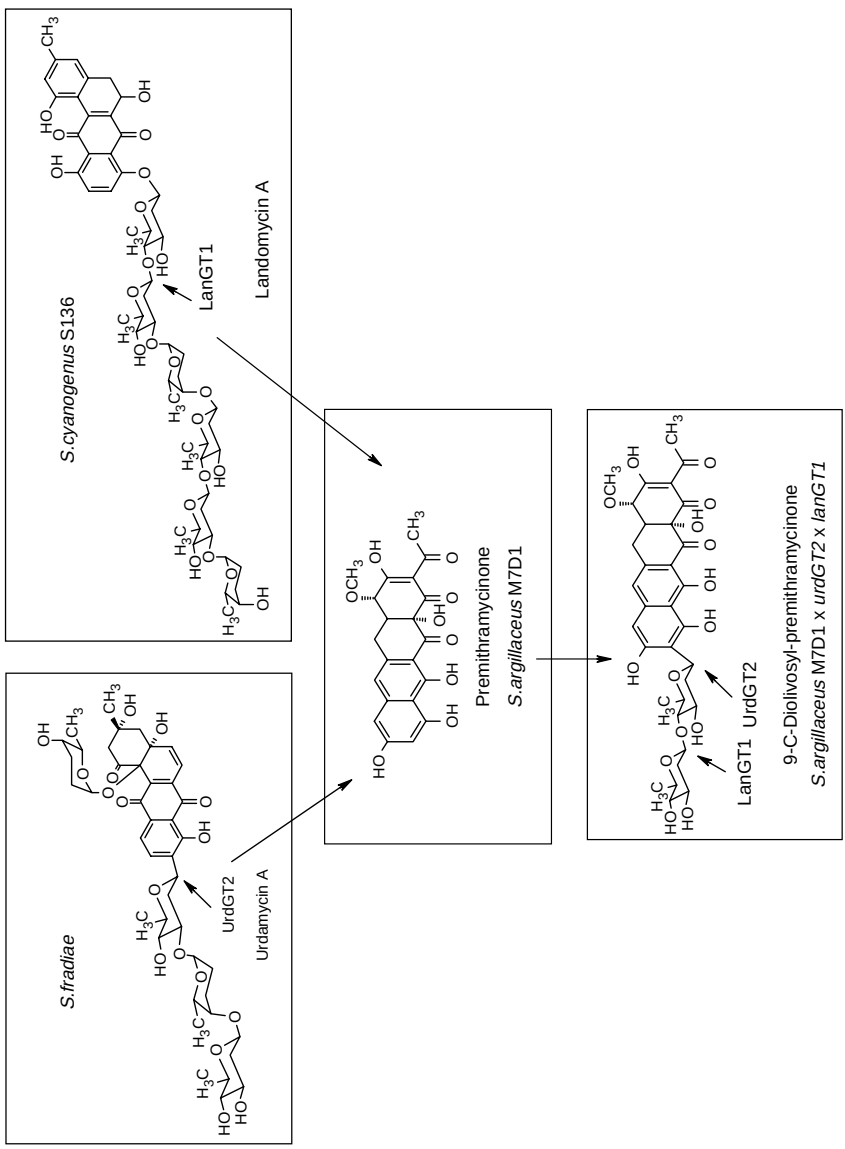

Fig. 4 Generation of a rationally designed hybrid product by coexpression of genes from three different organisms.

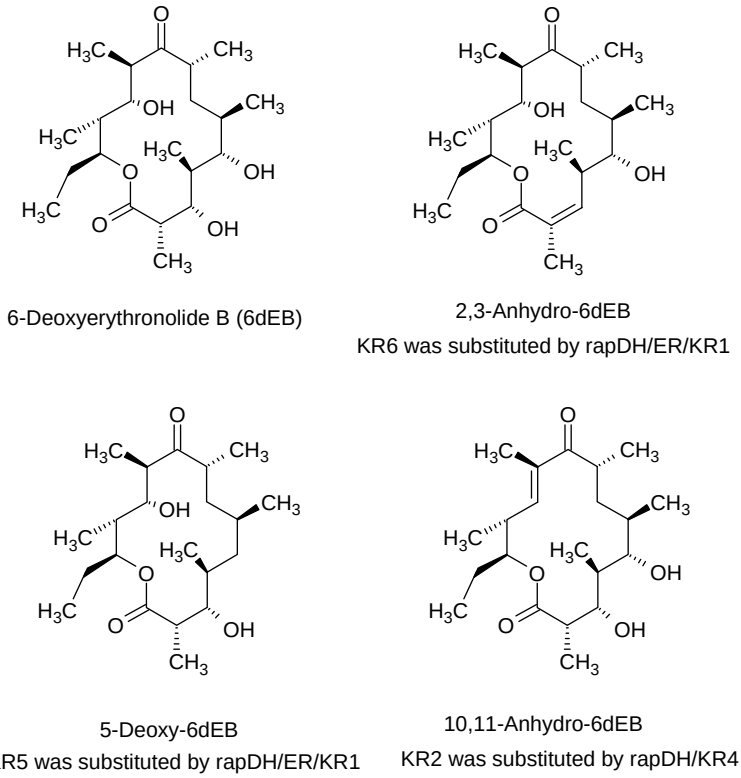

Fig. 5 Production of novel natural compounds by exchanging of modules of polyketide synthases.

and, in most cases, are expensive to dispose of. However, a still unsolved issue is how a novel active compound is dealt with by the host strain. Highly antimicrobial or cytotoxic agents that are not immediately and effectively detoxified by the host's intrinsic mechanisms will kill the host strain long before the compound is detected in a screen. Therefore, especially in case of antimicrobials, scientists run the risk of selecting for structures innocuous to pathogens, and truly combinatorial approaches suffer from limited sugar diversity as a second drawback since only those sugars are available for drug-lead diversification whose biosynthetic routes are understood and the genes involved cloned.

To make use of the possibilities microbial enzymes, particularly natural product glycosyltransferases, offer to pharmaceutical biotechnologists, the glycorandomization paradigm has been proposed and developed. Its two key reactions are the bioconversion of free sugars (natural or unnatural ones) via sugar phosphates into nucleotide-bound sugars as substrates for flexible glycosyltransferases.

The glycorandomization process starts with a library of chemically synthesized sugars, is therefore only limited by the available chemistry and includes, for example, halogenated, thio-, or azidosugars. Thus, the range of available sugars is extended far beyond what is possible

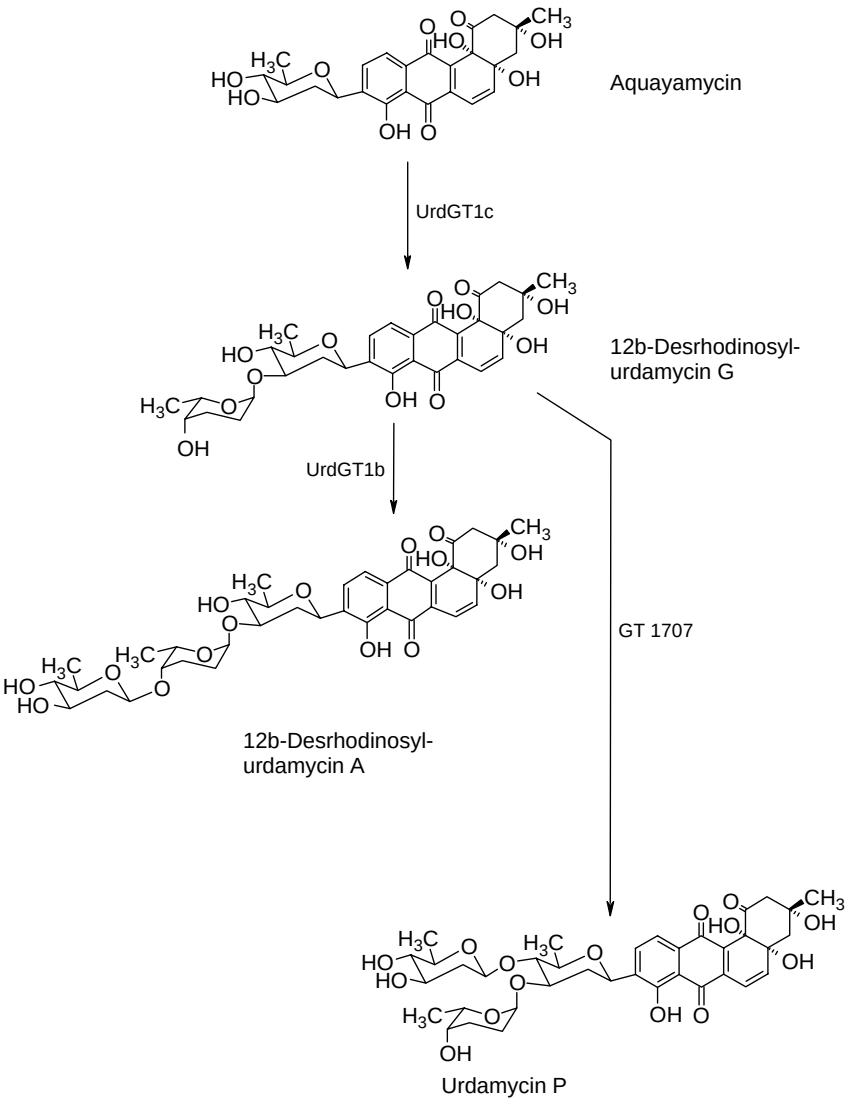

Fig. 6 Generation of urdamycin P by the expression of an "artificial" glycosyltransferase gene.

by biosynthetic pathways. These sugars are enzymatically converted to their anomeric phosphates during the kinase reaction, catalyzed by engineered *Escherichia coli* galactokinases and subsequently to dTDP- or UDP-derivatives by RmlA, the α-D-glucopyranosyl phosphate thymidylyltransferase from *Salmonella enterica* LT2. These enzymes display remarkable substrate flexibility after directed evolution or structure-based engineering. The RmlA-generated sugar nucleotides eventually

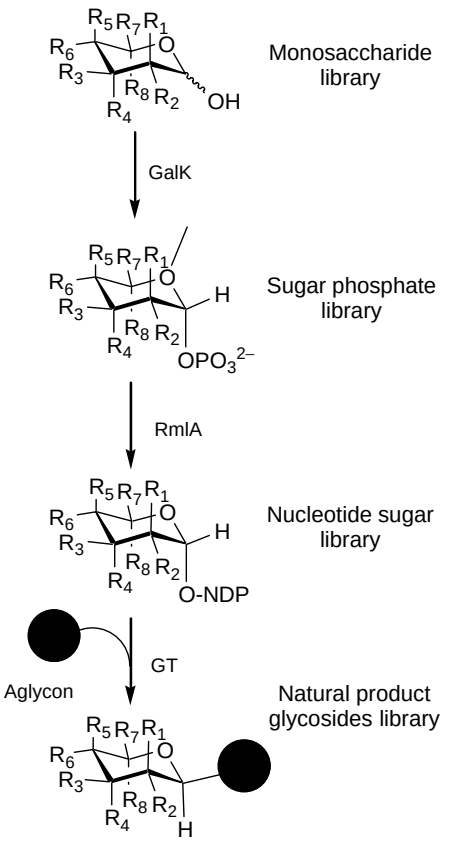

Fig. 7 Schematic of the glycorandomization strategy. Three enzymatic steps convert a library of free sugars into a library of natural product glycosides. GalK: galactokinase, RmlA: thymidylyltransferase, GT: glycosyltransferase, NDP: nucleosid diphosphate.

represent the sugar donor substrates for enzymatic glycosyltransfer.

Thus, glycorandomization involves only three enzymes (galactokinase, thymidylyltransferase, and a glycosyltransferase) to diversify natural product glycosides, thereby eliminating the need for large sets of biosynthetic genes (Fig. 7). If the sugar carries a reactive group, such as an azido-substituent, the glycorandomized molecules can be subjected to further chemical modifications under mild conditions thus creating an even more diverse pool of natural product derivatives.

Although being a very recent technique, glycorandomization has already demonstrated its versatility in its first application toward diversifying vancomycin and teicoplanin; nonribosomally generated sugar-decorated heptapeptides, which are in clinical use as antimicrobial drugs of last resort.

Combinatorial biosynthesis, and glycorandomization alike, are promising approaches to make use of the metabolic potential of prokaryotic cells and should promote drug development in the future. Potentially, glycorandomization can also be executed *in vivo* in a streptomycete host. In this case, the diverse intrinsic metabolites of Streptomycetes could be tapped as a rich source for drug backbone structures, the process could be scaled up easily, and expensive cosubstrates, for example, ATP, would be provided by the host.

1.4.5 Novel Natural Compounds by Mutasynthesis

The substrate flexibility of enzymes is also the basis for the "mutasynthesis" approach. During "mutasynthesis," a microorganism containing a defined mutation in an important precursor biosynthesis gene of an interesting metabolite can be fed with alternative or even synthetic precursors. Consequently, derivatives of complex natural products are generated, which may not have been obtained by synthetic methods. This technology was successfully applied to generate the first fluorinated vancomycin-type antibiotics.

2
Saccharomyces Cerevisiae and Other Fungi in Biotech Production

Saccharomyces (Sa.) cerevisiae might be viewed to be one of the most important fungal organisms used in biotechnology. It has been used in the "old biotechnology" for baking and brewing since prehistory. Yeast genetics, yeast biochemistry, and, finally, yeast molecular biology have substantially contributed to the importance of *Sa. cerevisiae* also in the "new biotechnology" area.

Yeast is a unicellular organism, which, unlike more complex eukaryotes, is amenable to mass production. It can be grown on defined media, giving the investigator a complete control over environmental parameters. The availability of the complete genome sequence of *Sa. cerevisiae* opened the age of "new biotechnology." In this chapter, we first review how genetic engineering of *Sa. cerevisiae* resulted in improved productivity and yield of important biotech products. Later, three examples of fungal natural products (or their derivatives) are described that have found their way into clinical use.

2.1
Generation of Engineered Strains of *Saccharomyces Cerevisiae* for the Production of Alcoholic Beverages

Alcohol fermentation is one of the most important processes of biotechnology. Generally, it is initiated by adding yeast to a carbon source and discontinues at a given alcohol concentration. The extension of the substrate range of *Sa. cerevisiae* is of major importance for the large-scale production of several metabolites. *Sa. cerevisiae* is not able to degrade starch and dextrin, since it does not produce starch-decomposing enzymes. Therefore, it is necessary to add starch-decomposing enzymes before fermentation. Attempts have been undertaken to use recombinant strains that contain the decomposing enzymes-encoding genes in order to avoid the preincubation process. The complete assimilation of starch (>98%) was accomplished by coexpression of the *sta2* gene of *Sa. diastaticus* encoding a glucoamylase, the *amy1* gene of *Bacillus amyloliquefaciens* encoding an α-amylase, and the *pulA* gene of *Klebsiella pneumoniae* encoding a pullulanase.

Further, genetically engineered strains have been developed, which are able to utilize lactose, melobiose, xylose, and other materials. A thermostable β-galactosidase encoded by *lacA* from *Aspergillus niger* was expressed in *Sa. cerevisiae*, enabling the strain to use lactose as carbon source. A melibiase-producing yeast was constructed by overexpressing the *melI* gene from another *Saccharomyces* strain. Moreover, after the coexpression of *xyl1* and *xyl2,* encoding a xylose reductase and xylitol dehydrogenase from *Pichia stipitis* along with the overexpressed xylulolkinase XKS1 from *Sa. cerevisiae*, xylose was converted to ethanol.

Especially in the large-scale production of beer, of highest significance is not ethanol production but a balanced flavor to obtain the desired taste. One unpleasant off-flavor compound is diacetyl, which is a nonenzymatically degraded product of α-acetolactate. Diacetyl is then enzymatically converted to acetoin and subsequently to 2,3-butanediol. The nonenzymatic-degradation step is very slow and requires long lager periods.

One way to avoid the off-flavor is to introduce an alternative route of degradation of α-acetolactate directly to acetoin.

α-Acetolactate decarboxylases from different organisms were successfully overexpressed in the beer-producer strains, accelerating the brewing process by diminishing the time of lagering by weeks.

Another interesting example is the expression of a β-glucanase of *Bacillus subtilis* in yeast. β-Glucans, the highly viscous side products during fermentation, impede beer filtration, which is still an important separation technique in the brewing industry. The presence of β-glucanase during fermentation did not affect the beer quality and taste but improved the filtration process.

2.2 Generation of Engineered Strains of *Saccharomyces Cerevisiae* for Lactic Acid, Xylitol, and Strictosidine Production

Several lactate dehydrogenases (LDHs) were expressed in *Sa. cerevisiae* in order to produce lactic acid. Most successful was the expression of a fungal (*Rhizopus oryzae*) lactate dehydrogenase (LDH). A recombinant strain accumulated approximately 40% more lactic acid with a final concentration of 38 g L^{-1} lactic acid and a yield of 0.44 g of lactic acid per gram of glucose. Xylitol is an attractive sweetener used in the food industry. Xylitol production in yeast was performed by the expression of *xyl1* of *P. stipitis*, encoding axylitosereductase.

A transgenic *Sa. cerevisiae* was constructed harboring the cDNAs that encodesstrictosidine synthase (STR) and strictosidine β-glucosidase (SGD) from the medicinal plant *Catharanthus roseus*. Both enzymes are involved in the biosynthesis of terpenoid indole alkaloids. The yeast culture was found to express high levels of both enzymes. Upon feeding of tryptamine and secologanin, this transgenic yeast culture produced high levels of strictosidine.

2.3 The Use of Fungi in the Production of Statins, Cyclosporin, and ß-Lactam Antibiotics

Statins are the secondary metabolites of a number of different filamentous fungi. Their medical importance and commercial value stem from their ability to inhibit the enzyme (3*S*)-hydroxy-3-methylglutaryl-CoA (HMG-CoA) reductase. Since this enzyme catalyzes a key step in the endogenous cholesterol biosynthetic pathway, statins have become the widely used antihypercholesterolemic drugs. Along with some synthetic statins, the most prominent examples are lovastatin, mainly from *Aspergillus terreus*, and mevastatin produced by *Penicillium citrinum*, which was the first statin to be discovered.

Chemical modification, for example, hydroxylation, turned out to be rather unproductive during derivatization efforts. Therefore, biotransformation and biotechnological approaches have become the strategy of choice relatively early. For example, a two-step fermentation/biotransformation process has been established for the clinically important pravastatin: Its direct precursor mevastatin is obtained out of a culture in the first step and is then subjected to biotransformation, for example, by *Streptomyces carbophilus* to complete pravastatin biosynthesis by introducing a hydroxyl group at C-6. Later, improved *Aspergillus* and *Monascus* strains for direct pravastatin production were described. Pioneering work on the genetics and enzymology underlying lovastatin biosynthesis was published in 1999, paving the way for

the generation of novel derivatives. Recently, an approach termed *association analysis* was developed to further improve statin producers. The interrelationship between secondary metabolite production levels and genome-wide gene expression was profiled for a minilibrary of *Aspergillus terreus* strains engineered to express either wild type or engineered genes that are part of the lovastatin cluster itself or implicated in secondary metabolite regulation. The authors found that multiple genes/proteins attributed to cellular processes as diverse as primary metabolism, secondary metabolism, carbohydrate utilization, sulfur assimilation, transport, proteolysis, and many more correlate with increased (or decreased) lovastatin production levels. This approach revealed multiple points from which to start engineering and may help manipulate statin-producing filamentous fungi for industrial purposes.

Cyclosporin A (INN: ciclosporin) is a cyclic, nonribosomally-synthesized undecapeptide from *Tolypocladium inflatum* (Fig. 2). Apart from its antifungal properties, it represents a potent immunosuppressive drug as it interferes with lymphokine production. Cyclosporin A has been introduced into clinical use to prevent allograft rejection after organ transplants.

The biosynthesis of cyclosporin A has been extensively investigated. A huge 45.8-kb open reading frame was identified as a putative gene coding for the cyclosporine synthetase, a multifunctional peptide synthetase. With a molecular mass of 1689 kDa, it represents one of nature's largest enzymes. Fungal peptide synthetases are somewhat different from their prokaryotic counterparts in that they usually consist of one huge single enzyme, as is the case for cyclosporin, whereas bacteria generally use multi-unit synthetases. Further, fungal peptides often include D-configured amino acids. However, these peptide synthetases do not harbor an epimerization domain. In the case of cyclosporin, an external alanin racemase responsible for supplying the synthetase with D-alanine has been purified and characterized. Even before the cyclosporin synthetase was characterized, it had become obvious that the biosynthetic pathway tolerates a number of substrate analogs. For example, D-alanine can be replaced by D-serine as was demonstrated by precursor-directed biosynthesis thereby leading to novel cyclosporin derivatives.

For enhanced scaled-up cyclosporin production, a modified sporulation/immobilization method has been proposed, increasing the cyclosporin yield up to tenfold, compared to suspended culture techniques. The immobilization of cells on celite carrier beads decreases the culture's viscosity and therefore increases the mass transfer. It is possible to run the fermentation continuously since the fungal spores can be trapped in the fermenter vessel to populate and germinate on freshly added beads.

Ever since penicillin was discovered and further developed into a drug for use in humans, there has hardly been a natural product that parallels its impact on medicine and pharmacy. Yet, penicillin is only one example of the β-lactam antibiotics class, along with other fungal (and streptomycete) secondary metabolites, for example, cephalosporin C from *Acremonium chrysogenum*. The commercial importance of penicillins and cephalosporins is evident from the worldwide annual sales for these compounds (including their semisynthetic derivatives), which is estimated to reach US$ 15 billion, and of that, US$ 4.8 billion derived from sales

in the United States. β-Lactam biosynthesis has been thoroughly investigated. In brief, a nonribosomal peptide synthetase assembles the building blocks, L-α-aminoadipic acid, L-valine (which is epimerized to D-valin), and L-cysteine, into a linear tripeptide, which is then cyclized to isopenicillin N (IPN) by the enzyme IPN synthase. The latter enzyme has been crystallized and, by elegant X-ray diffraction investigations, been used to support the notion that the bicyclic enzymatic product IPN is synthesized in a two-step process, with β-lactam formation preceding the closure of the thiazolidine ring. IPN represents the last common intermediate along the pathways toward penicillin and cephalosporins. While penicillin biosynthesis requires only side-chain modifications, the five-membered penicillin thiazolidine ring is expanded to a dihydrothiazine system on the route to cephalosporins. Gene clusters coding for the enzymatic machinery of β-lactam biosynthesis have been cloned from *Penicillium chrysogenum, A. nidulans* (Penicillins), and *Acremonium chrysogenum* (Cephalosporin C). Strain development, fermentation, recovery, and purification conditions for β-lactam producers have been subjected to optimization ever since commercial production started. For example, modern fed-batch fermentations yield penicillin titers exceeding 40 g L^{-1} (compared to less than 1 g L^{-1} in 1950). Penicillin G and V are produced in highly automated fermentation vessels with a capacity in the 100 to 400 000 L range. Although costs for energy and labor have increased, the advances in penicillin production techniques have led to a drastic decrease in bulk prices from ~US$300 per kg in the early 1950s to ~US$15 to 20 per kg today.

3 Plants in Biotech Production

From the very roots of humanity, plants have made a crucial contribution to our well being. Plant products have been used as food and as medicine. Even today, plants are an important source for the discovery of novel pharmacologically active compounds, though the recent competition from combinatorial chemistry and computational drug design has declined the interest. In recent years, gene technology has opened up exciting perspectives and offers tools not only to improve the existing properties of plants, such as the amount of bioactive compounds, but also to create transgenic plants with new properties. Facile transformation and cultivation not only make plants suitable for the production of secondary metabolites but also for recombinant proteins. Plants are capable of carrying out acetylation, phosphorylation, and glycosylation as well as other posttranslational protein modifications required for the biological activity of many eukaryotic proteins. This chapter gives some examples describing the use of plant cells in biotech production by which pharmaceuticals as well as functional and medicinal food are obtained. In this article, we restrict ourselves to a secondary metabolite plant in the low molecular-weight range.

3.1 Transgenic Plants as Functional Foods or Neutraceuticals

A few years ago, industry started the age of engineered functional food. Numerous examples such as the generation of golden rice, the production of healthy plant oils, and engineered plants with

Tab. 2 Examples of new properties in transgenic plants used as neutraceuticals.

Property	*Plant*
Resveratrol	Peanut
Increased amount of iron and its bioavailability by the reduction of phytic acid	Maize
Lactoferrin	Rice
Enriched ferritin leading to the binding of iron and consequently to its accumulation	Lettuce
Removal of bitter-tasting compounds (glycoalkaloids)	Potatoes

increased levels of essential vitamins and nutrients have been published (Table 2). Golden rice was engineered with two plant genes from *Narcissus pseudonarcissus* encoding a phytoene synthase and a lycopene ß-cyclase along with one bacterial gene from *Erwinia uredovora* encoding a phytoene desaturase to synthesize ß-carotene, a precursor of vitamin A. This was possible because the transformation of rice was well established and all carotinoid biosynthetic pathway genes had been identified. Despite these promising results, golden rice is not yet on the market. Further work aims to increase the provitamin A amount and to unify high-iron rice lines with provitamin A lines, as it is known that provitamin A potentially increases iron bioavailability.

Wheat is a further target in the area of engineered functional food. A gene encoding a stilbene synthase was expressed in rice that enabled wheat to produce resveratrol. This natural antioxidant possesses positive effects against the development of thrombosis and arteriosclerosis. Moreover, in its glycosidic form it enhances resistance to fungal pathogenes, which also occurs in transgenic plants.

The current health-related objective of plant seed engineering is to increase the content of "healthy fatty acids" and reduce "unhealthy fatty acids" in oilseed crops, such as soybean, oil palm, rape-seed, and sunflower. Genetic engineering was successful in reducing the levels of trans-unsaturated fatty acids and in reducing the ratio between omega-6 and omega-3 unsaturated fatty acids in some vegetable oils. Metabolic engineering also succeeded in increasing the vitamin E, vitamin C, and the lycopene content in plants, as well as the content of bioflavonoids, known for their antioxidant, anticancer, and estrogenic properties. In addition, human milk proteins like lactoferrin can now be expressed in plants. These proteins are believed to have a multitude of biological activities that benefit the newborn infant. Functional food selected and advertised for its high content of therapeutically active molecules is already offered in the shelves of supermarkets, leaving the determination of their true medical benefit to the consumer.

A further research area is the elimination of natural compounds from a plant to avoid severe side effects. As an example, peanut causes allergies due to several proteins. Researchers are now working on the generation of plants with reduced levels of these proteins.

3.2 Transgenic Plants and Plant Cell Culture as Bioreactors of Secondary Metabolites

Biotech methods are also used to increase the amount of pharmaceutically interesting compounds in plants. Leaves of *Atropa belladonna* contain high amounts of L-hyoscyamin, but negligible

contents of L-scopolamin due to the low activity of hyoscyamin-6ß-hydroxylase (H6H) in roots. H6H-cDNA was isolated from *Hyoscyamus niger*, cloned, and introduced in *A. belladonna* using *Agrobacterium tumefaciens*. Transgenic plants as well as the sexual descendents contain the transgene and accumulate up to 1% of L-scopolamin, but only traces of L-hyoscyamin (Table 3).

Another approach to produce biologically active secondary metabolites is the use of plant cell cultures. Plant cell cultures are advantageous in that they are not limited by environmental, ecological, or climatic conditions. Further, cells can proliferate at higher growth rates than whole plants in cultivation. As shown in Table 4, some metabolites in plant cell cultures have been reported to accumulate with a higher titer compared to those in the parent plants. Some industrial processes harness this potential, for example, for shikonin, ginseng, and paclitaxel production. Especially *Taxus* cell cultures are an interesting alternative to the isolation of paclitaxel from plantation- grown plants as the slow growth of *Taxus* species, the significant variation in taxoid content, and the costly purification of 10-deacetylbaccatin III from co-occurring taxoids are significantly limiting parameters. Using cell cultures, taxol production rates up to 23.4 mg L^{-1} per day with paclitaxel comprising 13 to 20% of the total taxoid fraction can now be achieved.

However, the use of plant cell cultures for the production of interesting

Tab. 3 Transgenic plants with selected improved production of secondary metabolites.

Compound	*Target protein*	*Gene donor*	*Gene recipient*
Cadaverin	Lysin-decarboxylase	*Hafnia alvei*	*Nicotiana tabacum*
Sterols	HMG-CoA-reductase	*Hevea brasiliensis*	*Nicotiana tabacum*
Nicotin	Ornithin-decarboxylase	*Saccharomyces cerevisiae*	*Nicotiana tabacum*
Resveratrol	Stilbene synthase	*Vitis vinifera*	*Nicotiana tabacum*
Scopolamin	Hyoscyamin-6ß-hydroxylase	*Hyoscyamus niger*	*Atropa belladonna*

Tab. 4 Comparison of product yield of secondary metabolites in cell culture and parent plants.

Product	*Plant*	*Yield [% Dw[a]]*	
		Culture	*Plant*
Anthocyanin	*Vitis* sp.	16	10
	Euphorbia milli	4	0.3
	Perilla frutescens	24	1.5
Anthraquinone	*Morinda citrifolia*	18	2.2
Berberine	*Coptis japonica*	13	4
	Thalictrum minor	10	0.01
Rosmarinic acid	*Coleus blumei*	27	3.0
Shikonin	*Lithospermum erythrorhizon*	14	1.5

[a] Dry weight.

molecules has not gained acceptance in industry as yet. Usually, low productivity and high costs are the most important negative parameters. Nevertheless, research is still going on and is well described in the next article. Further application of plant-cell-suspension cultures are aimed at the production of ajmalicine, vinblastine, vincristine, podophyllotoxins, and camptothecin.

3.3 Transgenic Plants as Bioreactors of Recombinant Proteins

Nowadays, plants such as tobacco, potato, tomato, banana, legumes, and cereals as well as alfalfa, are used in molecular farming and have emerged as promising biopharming systems for production of pharmaceutical proteins, such as antibodies, vaccines, regulatory proteins, and enzymes (Table 5). The advantages offered by plants include low cost of cultivation, high biomass production, relatively fast protein synthesis, low operating costs, excellent scalability, eukaryotic posttranslational modifications, low risk of pathogenicity toward humans and endotoxins, and a relatively high protein expression level. Using transgenic plants as a host is highly attractive in that proteins can be administered in fruits and vegetables as a source of antigens for oral vaccination. Thus, potatoes expressing a synthetic *LT-B* gene, a labile toxin from *E. coli* were successfully used in a clinical study to examine an edible plant vaccine. Further, interesting vaccines that have been tested clinically are directed against viral diarrhea and hepatitis B. No plant-derived protein has still been developed to be used as a drug, but molecular farming has gained attention as plants can be turned into molecular medicine factories.

Recently, plant cells have also been considered to be an alternative host for the production of recombinant proteins since they are able to glycosylate proteins. Of the various systems used for

Tab. 5 Examples of therapeutic antibodies produced in plants.

Antibody format	*Antigen*	*Cellular location*	*Transgenic plant*	*Max. expression level*
dAb	Substance P	Apoplast	*Nicotiana benthamiana*	1% TSP leaves
IgG1, Fab	Human creatine kinase	Apoplast	*N. tabacum*	0.044% TSP leaves
			Arabidopsis thaliana	1.3% TSP leaves
IgG1	Streptococcal surface antigen (I/II)	Plasma membrane	*N. tabacum*	1.1% TSP leaves
IgG1	Human IgG	Apoplast	*Medicago sativa*	1% TSP
IgG1	Herplex simplex virus 2	Apoplast	*Glycine max*	Not reported
SigA	Streptococcal surface antigen (I/II)	Apoplast	*N. tabacum*	0.5 mg g^{-1} FW leaves
ScFv	Carcinoembryonic antigen	ER	*Pisum sativum*	0.009 mg g^{-1} seed

Notes: dAb: single domain antibody; FW: fresh weight; TSP: total soluble protein.

cultivation, such as hairy roots, immobilized cells, and free cell suspensions, the latter is regarded to be most suitable for large-scale applications. Full-size antibodies, antibody fragments, and fusion proteins can be expressed in transgenic plant–cell systems, such as *Nicotiana tabacum*, pea, wheat, and rice using shake-flask or fermentation cultures. Yet, these systems are still of low commercial importance as their productivity is not advantageous.

See also Bacterial Cell Culture Methods; Bacterial Growth and Division; Bioprocess Engineering; Expression Systems for DNA Processes; Mammalian Cell Culture Methods; Plant-based Expression of Biopharmaceuticals.

Bibliography

Books and Reviews

August, P.R., Yu, T-W., Floss, H.G. (2000) in: Grabley, S., Thiericke, R. (Eds.) *Drug Discovery from Nature*, Springer, Berlin-Heidelberg, Germany.

Berdy, J. (1995) *Proceedings of the 9th International Symposium on the Biology of Actinomycetes*, Part 1, Allerton Press, New York, pp. 3–23.

Kreis, W., Baron, D., Stoll, G. (Eds.) (2001) *Biotechnologie der Arzneistoffe*, Dtsch. Apoth. Verlag, Stuttgart, pp. 250–260.

Muth, G., Brolle, F., Wohlleben, D.F. (1999) in: Demain, A.L., Davis, J.E., Atlas, R.M. (Eds.) *Manual of Industrial Microbiology and Biotechnology*, ASM Press, Washington, DC, pp. 353–367.

Skarbek, J., Speedie, M. (1981) in: Aszalos, A. (Ed.) *Antitumor Compounds of Natural Origin*, Vol. 1, CRC Press, Boca Raton, FL, pp. 191–235.

Strohl, W.R. (1997) in: Strohl, W.R. (Eds.) *Biotechnology of Antibiotics*, 2nd edition, Marcel Dekker, New York, pp. 1–47.

Strohl, W.R. (1999) Secondary Metabolites; Antibiotics, in: Flickinger, M.C., Drew, S.W. (Eds.) *Encyclopedia of Bioprocess Technology: Fermentation, Biocatalysis and Bioseparations*, Wiley, New York, pp. 2348–2365.

Zähner, H., Fiedler, H-P. (1995) in: Hunter, P.A., Darby, G.K., Russel, N.J. (Eds.) *Past Perspectives and Future Friends*, Cambridge University Press, Cambridge, MA, pp. 67–84.

Primary Literature

Askenazi, M., Driggers, E.M., Holtzman, D.A., Norman, T.C., Iverson, S., Zimmer, D.P., Boers, M.E., Blomquist, P.R., Martinez, E.J., Monreal, A.W., Feibelman, T.P., Mayorga, M.E., Maxon, M.E., Sykes, K., Tobin, J.V., Cordero, E., Salama, S.R., Trueheart, J., Royer, J.C., Madden, K.T. (2003) Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains, *Nat. Biotechnol.* **21**, 150–156.

Baltz, R.H. (1998) Genetic manipulation of antibiotic-producing Streptomyces, *Trends Microbiol.* **6**, 76–83.

Bax, R.P., Anderson, R., Crew, J., Fletcher, P., Johnson, T., Kaplan, E., Knaus, B., Kristinsson, K., Malek, M., Strandberg, L. (1998) Antibiotic resistance–what can we do? *Nat. Med.* **4**, 545–546.

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J., Hopwood, D.A. (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2), *Nature* **417**, 141–147.

Bibb, M., Schottel, J.L., Cohen, S.N. (1980) A DNA cloning system for interspecies gene transfer in antibiotic-producing Streptomyces, *Nature* **284**, 526–531.

Bode, H.B., Bethe, B., Höfs, R., Zeeck, A. (2002) Big effects from small changes: possible ways to explore nature's chemical diversity, *Chem. Biochem.* **3**, 619–627.

Brautaset, T., Sekurova, O.N., Sletta, H., Ellingsen, T.E., StrLm, A.R., Valla, S., Zotchev, S.B. (2000) Biosynthesis of the polyene antifungal antibiotic nystatin in Streptomyces noursei ATCC 11455: analysis of the gene cluster and deduction of the biosynthetic pathway, *Chem. Biol.* **7**, 395–403.

Breithaupt, H. (1999) The new antibioptics, *Nat. Biotechnol.* **17**, 1165–1169.

Burzlaff, N.I., Rutledge, P.J., Clifton, I.J., Hensgens, C.M.H., Pickford, M., Adlington, R.M., Roach, P.L., Baldwin, J.E. (1999) The reaction cycle of isopenicillin N synthase observed by X-ray diffraction, *Nature* **401**, 721–724.

Choi, J.W., Cho, G.H., Byun, S.Y., Kim, D.I. (2001) Integrated bioprocessing for plant cell cultures, *Adv. Biochem. Eng. Biotechnol.* **72**, 63–102.

Coates, A., Hu, Y., Bax, R., Page, C. (2002) The future challenges facing the development of new antimicrobial drugs, *Nat. Rev. Drug Discov.* **11**, 895–910.

Cragg, G.M., Newman, D.J., Snader, K.M. (1997) Natural products in drug discovery and development, *J. Nat. Prod.* **60**, 52–60.

Cropp, T.A., Wilson, D.J., Reynolds, K.A. (2000) Identification of a cyclohexylcarbonyl CoA biosynthetic gene cluster and application in the production of doramectin, *Nat. Biotechnol.* **18**, 980–983.

Cusido, R.M., Palazon, J., Bonfill, M., Navia-Osorio, A., Morales, C., Pinol, M.T. (2002) Improved Paclitaxel and Baccatin III production in suspension cultures of Taxus media, *Biotechnol. Prog.* **18**, 418–423.

Demain, A.L. (1999) Pharmaceutically active secondary metabolites of microorganisms, *Appl. Microbiol. Biotechnol.* **52**, 455–463.

Demain, A.L. (2002) Prescription for an ailing pharmaceutical industry, *Nat. Biotechnol.* **20**, 331.

Donadio, S., Sosio, M., Lancini, G. (2002) Impact of the first Streptomyces genome sequence on the discovery and production of bioactive substances, *Appl. Microbiol. Biotechnol.* **60**, 377–380.

Donadio, S., Monciardini, P., Alduina, R., Mazza, P., Chiocchino, C., Cavaletti, L., Sosio, M., Puglia, A.M. (2002) Microbial technologies for the discovery of novel bioactive metabolites, *J. Biotechnol.* **99**, 187–198.

Drautz, H., Zähner, H., Rohr, J., Zeeck, A. (1986) Metabolic products of microorganisms. Urdamycins, new angucycline antibiotics from Streptomyces fradiae. I. Isolation, characterization and biological properties, *J. Antibiot.* **39**, 1657–1669.

Elander, R.P. (2003) Industrial production of β-lactam antibiotics, *Appl. Microbiol. Biotechnol.* **61**, 385–392.

Fischer, R., Drossard, J., Commandeur, U., Schillberg, S., Emans, N. (1999) Towards molecular farming in the future: moving from diagnostic protein and antibody production in microbes to plants, *Biotechnol. Appl. Biochem.* **30**, 101–108.

Fu, X., Albermann, C., Jiang, J., Liao, J., Zhang, C., Thorson, J.S. (2003) Antibiotic optimization via in vitro glycorandomization, *Nat. Biotechnol.* **21**, 1467–1469.

Geerlings, A., Redondo, F.J., Contin, A., Memelink, J., van der Heijden, R., Verpoorte, R. (2001) Biotransformation of tryptamine and secologanin into plant terpenoid indole alkaloids by transgenic yeast, *Appl. Microbiol. Biotechnol.* **56**, 420–424.

Govinden, R., Pillay, B., van Zyl, W.H., Pillay, D. (2001) Xylitol production by recombinant Saccharomyces cerevisiae expressing the Pichia stipitis and Candida shehatae XYL1 genes, *Appl. Microbiol. Biotechnol.* **55**, 76–80.

Ho, N.W., Chen, Z., Brainard, A.P., Sedlak, M. (1999) Successful design and development of genetically engineered Saccharomyces yeasts for effective cofermentation of glucose and xylose from cellulosic biomass to fuel ethanol, *Adv. Biochem. Eng. Biotechnol.* **65**, 163–192.

Hoffmeister, D., Ichinose, K., Bechthold, A. (2001) Two sequence elements of glycosyltransferases involved in urdamycin biosynthesis are responsible for substrate specificity and enzymatic activity, *Chem. Biol.* **8**, 557–567.

Hoffmeister, D., Wilkinson, B., Foster, G., Sidebottom, P.J., Ichinose, K., Bechthold, A. (2002) Engineered urdamycin glycosyltransferases are broadened and altered in substrate specificity, *Chem. Biol.* **9**, 287–295.

Hopwood, D.A. (1997) Genetic Contributions to Understanding Polyketide Synthases, *Chem. Rev.* **10**, 2465–2498.

Hopwood, D.A. (1999) Forty years of genetics with Streptomyces: from in vivo through in vitro to in silico, *Microbiology* **145**, 2183–2202.

Hopwood, D.A. (2003) The Streptomyces genome-be prepared, *Nat. Biotechnol.* **21**, 505–506.

Ikeda, H., Ishikawa, J., Hanamoto, A., Spinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., Omura, S. (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*, *Nat. Biotechnol.* **21**, 526–531.

Janse, B.J., Pretorius, I.S. (1995) One-step enzymatic hydrolysis of starch using a recombinant strain of Saccharomyces cerevisiae producing alpha-amylase, glucoamylase and pullulanase, *Appl. Microbiol. Biotechnol.* **42**, 873–883.

Kennedy, J., Auclair, K., Kendrew, S.G., Park, C., Vederas, J.C., Hutchinson, C.R. (1999) Modulation of Polyketide Synthase Activity by Accessory Proteins During Lovastatin Biosynthesis, *Science* **284**, 1368–1372.

Kumar, V., Ramakrishnan, S., Teeri, T.T., Knowles, J.K., Hartley, B.S. (1992) Saccharomyces cerevisiae cells secreting an Aspergillus niger beta-galactosidase grow on whey permeate, *Biotechnology* **10**, 82–85.

Lee, T.H., Chun, G.T., Chang, Y.K. (1997) Development of sporulation/immobilization method and its application for the continuous production of cyclosporin A by *Tolypocladium inflatum*, *Biotechnol. Prog.* **13**, 546–550.

Lin, Y.S., Kieser, H.M., Hopwood, D.A., Chen, C.W. (1993) The chromosomal DNA of Streptomyces lividans 66 is linear, *Mol. Microbiol.* **10**, 923–933.

Liu, H., Reynolds, K.A. (2001) Precursor supply for polyketide biosynthesis: the role of crotonyl-CoA reductase, *Metab. Eng.* **3**, 40–48.

Losey, H.C., Jiang, J., Biggins, J.B., Oberthür, M., Ye, X-Y., Dong, S.D., Kahne, D., Thorson, J.S., Walsh, C.T. (2002) Incorporation of Glucose Analogs by GtfE and GtfD from the Vancomycin Biosynthetic Pathway to Generate Variant Glycopeptides, *Chem. Biol.* **9**, 1305–1314.

MacNeil, I.A., Tiong, C.L., Minor, C., August, P.R., Grossman, T.H., Loiacono, K.A., Lynch, B.A., Phillips, T., Narula, S., Sundaramoorthi, R., Tyler, A., Aldredge, T., Long, H., Gilman, M., Holt, D., Osburne, M.S. (2001) Expression and isolation of antimicrobial small molecules from soil DNA libraries, *J. Mol. Microbiol. Biotechnol.* **3**, 301–308.

Malpartida, F., Hopwood, D.A. (1984) Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host, *Nature* **309**, 462–464.

Martin, J.F. (1998) New aspects of genes and enzymes for β-lactam antibiotic biosynthesis, *Appl. Microbiol. Biotechnol.* **50**, 1–15.

Mason, H.S., Warzecha, H., Mor, T., Arntzen, C.J. (2002) Edible plant vaccines: applications for prophylactic and therapeutic molecular medicine, *Trends Mol. Med.* **8**, 324–329.

McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., Ashley, G. (1999) Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1846–1851.

Mendez, C., Künzel, E., Lipata, F., Lombo, F., Cotham, W., Walla, M., Bearden, D.W., Brana, A.F., Salas, J.A., Rohr, J. (2002) Oviedomycin, an unusual angucyclinone encoded by genes of the oleandomycin-producer Streptomyces antibioticus ATCC11891, *J. Nat. Prod.* **65**, 779–782.

Metsä-Ketelä, M., Palmu, K., Kunnari, T., Ylihonko, K., Mentsala, P. (2003) Engineering anthracycline biosynthesis toward angucyclines, *Antimicrob. Agents Chemother.* **47**, 1291–1296.

Miele, L. (1997) Plants as bioreactors for biopharmaceuticals: regulatory considerations, *Trends Biotechnol.* **15**, 45–50.

Moore, B.S., Piel, J. (2000) Engineering biodiversity with type II polyketide synthase genes, *Antonie Van Leeuwenhoek* **78**, 391–398.

Moraes, R.M., Bedir, E., Barrett, H., Burandt, C., Canel, C., Khan, I.A. (2002) Evaluation of Podophyllum peltatum accessions for podophyllotoxin production, *Planta Med.* **68**, 341–344.

Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. (2001) Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12215–12220.

Ostergaard, S., Olsson, L., Nielsen, J. (2000) Metabolic engineering of Saccharomyces cerevisiae, *Microbiol. Mol. Biol. Rev.* **64**, 34–50.

Paradkar, A., Trefzer, A., Chakraburtty, R., Stassi, D. (2003) Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host, *Crit. Rev. Biotechnol.* **23**, 1–27.

Pelzer, S., Reichert, W., Huppert, M., Heckmann, D., Wohlleben, W. (1997) Cloning and analysis of a peptide synthetase gene of the balhimycin producer Amycolatopsis mediterranei DSM5908 and development of a gene disruption/replacement system, *J. Biotechnol.* **56**, 115–128.

Randak, C., Brabletz, T., Hergenrother, M., Sobotta, I., Serfling, E. (1990) Cyclosporin A suppresses the expression of the interleukin 2 gene by inhibiting the binding of lymphocyte-specific factors to the IL-2 enhancer, *EMBO J.* **9**, 2529–2536.

Raskin, I., Ribnicky, D.M., Komarnytsky, S., Ilic, N., Poulev, A., Borisjuk, N., Brinker, A., Moreno, D.A., Ripoll, C., Yakoby, N., O'Neal, J.M., Cornwell, T., Pastor, I., Fridlender, B. (2002) Plants and human health in the twenty-first century, *Trends Biotechnol.* **20**, 522–531.

Redenbach, M., Scheel, J., Schmidt, U. (2000) Chromosome topology and genome size of selected actinomycetes species, *Antonie Van Leeuwenhoek* **78**, 227–235.

Rix, U., Fischer, C., Remsing, L., Rohr, J. (2002) Modification of post-PKS tailoring steps through combinatorial biosynthesis, *Nat. Prod. Rep.* **19**, 542–580.

Rodriguez, L., Aguirrezabalaga, I., Allende, N., Brana, A.F., Mendez, C., Salas, J.A. (2002) Engineering deoxysugar biosynthetic pathways from antibiotic-producing microorganisms. A tool to produce novel glycosylated bioactive compounds, *Chem. Biol.* **9**, 721–729.

Rosamond, J., Allsop, A. (2000) Harnessing the power of the genome in the search for new antibiotics, *Science* **287**, 1973–1976.

Ruan, X., Stassi, D., Lax, S.A., Katz, L. (1997) A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC 29253, a rapamycin-producing strain, *Gene* **203**, 1–9.

Rudd, B.A.M., Hopwood, D.A. (1979) Genetics of actinorhodin biosynthesis by Streptomyces coelicolor A3(2), *J. Gen. Microbiol.* **114**, 35–43.

Schillberg, S., Fischer, R., Emans, N. (2003) 'Molecular farming' of antibodies in plants, *Naturwissenschaften* **90**, 115–155.

Sezonov, G., Blanc, V., Bamas-Jacques, N., Friedmann, A., Pernodet, J.L., Guerineau, M. (1997) Complete conversion of antibiotic precursor to pristinamycin IIA by overexpression of Streptomyces pristinaespiralis biosynthetic genes, *Nat. Biotechnol.* **15**, 349–353.

Skory, C.D. (2003) Lactic acid production by Saccharomyces cerevisiae expressing a Rhizopus oryzae lactate dehydrogenase gene, *J. Ind. Microbiol. Biotechnol.* **30**, 22–27.

Sosio, M., Bossi, E., Bianchi, A., Donadio, S. (2000) Multiple peptide synthetase gene clusters in Actinomycetes, *Mol. Gen. Genet.* **264**, 213–221.

Strohl, W.R. (2001) Biochemical engineering of natural product biosynthesis pathways, *Metab. Eng.* **3**, 4–14.

Thompson, C.J., Ward, J.M., Hopwood, D.A. (1980) DNA cloning in Streptomyces: resistance genes from antibiotic-producing species, *Nature* **286**, 525–527.

Tobert, J.A. (2003) Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors, *Nat. Rev. Drug Discov.* **2**, 517–526.

Trefzer, A., Blanco, G., Remsing, L., Künzel, E., Rix, U., Lipata, F., Brana, A.F., Mendez, C., Rohr, J., Bechthold, A., Salas, J.A. (2002) Rationally designed glycosylated premithramycins: hybrid aromatic polyketides using genes from three different biosynthetic pathways, *J. Am. Chem. Soc.* **124**, 6056–6062.

Trefzer, A., Hoffmeister, D., Künzel, E., Stockert, S., Weitnauer, G., Westrich, L., Rix, U., Fuchser, J., Bindseil, K.U., Rohr, J., Bechthold, A. (2000) Function of glycosyltransferase genes involved in urdamycin A biosynthesis, Function of glycosyltransferase genes involved in urdamycin A biosynthesis, *Chem. Biol.* **7**, 133–142.

Verdine, G.L. (1996) The combinatorial chemistry of nature, *Nature* **384**, 11–13.

Vincent, S.F., Bell, P.J., Bissinger, P., Nevalainen, K.M. (1999) Comparison of melibiose utilizing baker's yeast strains produced by genetic engineering and classical breeding, *Lett. Appl. Microbiol.* **28**, 148–152.

Wang, G.Y., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J., Meurer, G., Saxena, G., Andersen, R.J., Davies, J. (2000) Novel natural products from soil DNA libraries in a streptomycete host, *Org. Lett.* **2**, 2401–2404.

Watve, M.G., Tickoo, R., Jog, M.M., Bhole, B.D. (2001) How many antibiotics are produced by the genus Streptomyces? *Arch. Microbiol.* **176**, 386–390.

Weber, G., Schörgendorfer, K., Schneider-Scherzer, E., Leitner, E. (1994) The peptide synthetase catalyzing cyclosporine production in Tolypocladium niveum is encoded by a giant 45.8-kilobase open reading frame, *Curr. Genet.* **26**, 120–125.

Weber, J.M., Leung, J.O., Swanson, S.J., Idler, K.B., McAlpine, J.B. (1991) An erythromycin derivative produced by targeted gene disruption in Saccharopolyspora erythraea, *Science* **252**, 114–117.

Weber, T., Welzel, K., Pelzer, S., Vente, A., Wohlleben, W. (2003) Exploiting the genetic potential of polyketide producing streptomycetes, *J. Biotechnol.* **106**, 221–232.

Weist, S., Bister, B., Puk, O., Bischoff, D., Pelzer, S., Nicholson, G.J., Wohlleben, W., Jung, G., Süssmuth, R.D. (2002) Fluorobalhimycin–a new chapter in glycopeptide antibiotic research, *Angew. Chem., Int. Ed. Engl.* **41**, 3383–3385.

Weitnauer, G., Mühlenweg, A., Trefzer, A., Hoffmeister, D., Süssmuth, R.D., Jung, G., Welzel, K., Vente, A., Girreser, U., Bechthold, A. (2001) Biosynthesis of the orthosomycin antibiotic avilamycin A: deductions from the molecular analysis of the avi biosynthetic gene cluster of Streptomyces viridochromogenes Tu57 and production of new antibiotics, *Chem. Biol.* **8**, 569–581.

Wohlleben, W., Pelzer, S. (2002) New compounds by combining ''modern'' genomics and ''old-fashioned'' mutasynthesis, *Chem. Biol.* **9**, 1163–1164.

Yamano, S., Kondo, K., Tanaka, J., Inoue, T. (1994) Construction of a brewer's yeast having alpha-acetolactate decarboxylase gene from Acetobacter aceti ssp. xylinum integrated in the menome, *J. Biotechnol.* **14**, 173–178.

Yang, J., Hoffmeister, D., Liu, L., Thorson, J.S. (2004) Natural product glycorandomization, *Bioorg. Med. Chem.* **12**, 1577–1584.

Zazopoulos, E., Huang, K., Staffa, A., Liu, W., Bachmann, B.O., Nonaka, K., Ahlert, J., Thorson, J.S., Shen, B., Farnet, C.M. (2003) A genomics-guided approach for discovering and expressing cryptic metabolic pathways, *Nat. Biotechnol.* **21**, 187–190.

Zhong, J.J. (2001) Biochemical engineering of the production of plant-specific secondary metabolites by cell suspension cultures, *Adv. Biochem. Eng. Biotechnol.* **72**, 1–26.

Proteasomes

Martin Rechsteiner
Department of Biochemistry University of Utah, Salt Lake, UT, USA

1 Proteasomes – A Quick Summary 5

2 The 20S Proteasome 5
2.1 Structure 5
2.2 Enzyme Mechanism and Proteasome Inhibitors 6
2.3 Immunoproteasomes 6

3 The 26S Proteasome 7
3.1 The Ubiquitin–proteasome System 7
3.2 Ultrastructure of the 26S Proteasome and Regulatory Complex 8
3.3 The 19S Regulatory Complex (RC) 9
3.4 ATPases of the RC 9
3.5 The non-ATPase Subunits 11
3.6 Biochemical Properties of the Regulatory Complex 12
3.6.1 Nucleotide Hydrolysis 12
3.6.2 Chaperone-like Activity 12
3.6.3 Proteasome Activation 12
3.6.4 Ubiquitin Isopeptide Hydrolysis 12
3.6.5 Substrate Recognition 13

4 Proteasome Biogenesis 13
4.1 Subunit Synthesis 13
4.2 Biogenesis of the 20S Proteasome 13
4.3 Biogenesis of the RC 13
4.4 Posttranslational Modification of Proteasome Subunits 13
4.5 Biogenesis of the 26S Proteasome 14

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

5 Substrate Recognition by Proteasomes 14
5.1 Degradation Signals (Degrons) 14
5.2 Ubiquitin-dependent Recognition of Substrates 14
5.3 Substrate Selection Independent of Ubiquitin 15

6 Proteolysis by the 26S Proteasome – Mechanism 16
6.1 Contribution of Chaperones to Proteasome-mediated Degradation 16
6.2 Presumed Mechanism 16
6.3 Processing by the 26S Proteasome 18

7 Proteasome Activators 18
7.1 REGs or PA28s 18
7.2 PA200 19
7.3 Hybrid Proteasomes 19
7.4 ECM29 19

8 Protein Inhibitors of the Proteasome 19

9 Physiological Aspects of Proteasomes 20
9.1 Tissue and Subcellular Distribution of Proteasomes 20
9.2 Physiological Importance 20
9.3 The Ubiquitin–proteasome System and Human Disease 20

Bibliography 21
Books and Reviews 21
Primary Literature 22

Keywords

AAA ATPases
A large subfamily of ATPases involved in separating and unfolding proteins. The six ATPases of the 19S regulatory complex are members of the AAA subfamily.

Aggresome
"Abnormal proteins" – pericentrosomal accumulations of abnormal proteins that recruit chaperones and proteasomes.

Base
A nine-protein subcomponent of the 19S regulatory complex, containing six ATPases and three proteins that recognize ubiquitin or ubiquitin-like proteins.

Chaperone
Chaperones are proteins that assist in the folding of nascent proteins or the refolding of denatured proteins. The heat shock proteins such as hsap90 or hsp70 are prime examples.

COP9 Signalosome (Csn)
A protein complex containing a metalloisopeptidase that removes the ubiquitin-like protein NEDD8 from ubiquitin ligases. Each of the eight Csn subunits is evolutionarily related to the eight lid subunits in the 19S regulatory complex.

Degron
Short peptide motifs (or even single N-terminal amino acids) that confer rapid proteolysis on the polypeptides bearing them.

Ecm29
A proteasome-associated protein proposed to function either as a clamp holding the 19S RC to the 20S proteasome or as an adaptor that localizes the 26S proteasome to specific membrane compartments within cells.

Epoxomicin
A bacterial compound that is a specific inhibitor of the proteasome.

Hybrid Proteasome
A 20S proteasome with a 19S regulatory complex at one end and either PA28 or PA200 at the other.

Immunoproteasome
A 20S proteasome where each of the three constitutive catalytic subunits is replaced by an active subunit with altered substrate specificity. Immunoproteasomes are induced by immune cytokines and play a role in specifying epitopes presented by the Class-I immune pathway.

Lactacystin
A fungal metabolite that is a reasonably specific inhibitor of the proteasome's active sites.

Lid
An eight-protein subcomponent of the 19S regulatory complex, containing a metalloisopeptidase that removes ubiquitin chains from the substrate protein.

Metabolic Regulation
Metabolic regulation refers to the control of biological processes by changes in the concentration of metabolites or proteins or by changes in the activities of enzymes. Phosphorylation and ubiquitilation are widespread mechanisms for controlling enzyme activity.

PA28s (aka 11S REGs)
Donut shaped, heptameric protein complexes that bind the ends of the 20S proteasome and promote peptide entry to or efflux from its central proteolytic chamber.

PA200
A large nuclear protein that binds the ends of the 20S proteasome and is thought to play a role in DNA repair.

Proteasome Activators
Single polypeptide chains or small protein complexes that bind 20S proteasomes and stimulate peptide hydrolysis. To date, three have been discovered.

20S Proteasome
A cylindrical proteolytic particle composed of 28 subunits arranged as a stack of four heptameric rings. The enzyme's active sites face an internal chamber.

26S Proteasome
The only ATP-dependent protease discovered so far in the nuclear and cytosolic compartments of eukaryotic cells. The enzyme consists of one or two 19S regulatory complexes attached to the ends of the 20S proteasome.

19S Regulatory Complex (aka PA700)
A multisubunit particle containing six ATPases and eleven additional proteins that functions to bind, unfold, and translocate substrate proteins into the central proteolytic chamber of the 20S proteasome.

Ubiquitin
An exceptionally conserved 76-amino acid protein that is covalently attached to a wide variety of other eukaryotic proteins. Chains of ubiquitin attached to substrate proteins target them for destruction by the 26S proteasome.

VCP
A large hexameric ATPase that transfers some polyubiquitylated substrates to the 26S proteasome.

Velcade
A peptide boronic acid inhibitor of the proteasome used clinically to treat multiple myeloma.

The 20S proteasome was discovered in 1980 and the 26S proteasome six years later. Research over the past two decades has made it abundantly clear that the Ub-proteasome system is of central importance in eukaryotic cell physiology and medicine. At the cellular and biochemical levels, there are a number of unresolved problems. We need a crystal structure of the 19S RC, or better yet, the 26S proteasome, for they would surely provide an insight into the mechanism by which the 26S proteasome degrades its substrates. How the 26S proteasome itself is regulated and the extent to which proteasomal components vary among

tissues in higher eukaryotes are other important unresolved problems. As for the medical perspective, we need to know how many diseases will be found to arise because of defects in the ubiquitin–proteasome system. Hopefully, these problems will generate research on the UPS by some readers of this article.

1 Proteasomes – A Quick Summary

Proteasomes are multisubunit, cylindrical proteases found in eukaryotes, eubacteria, and archaebacteria. The proteasome's active sites face a central chamber buried within the cylindrical particle. Thus, the proteasome is an ideal intracellular protease because cellular proteins can only be degraded if they are actively transferred to the enzyme's central chamber. Eukaryotic proteasomes come in two sizes, the 20S proteasome and the considerably larger ATP-dependent 26S proteasome. The latter is formed when the 20S proteasome binds one or two multisubunit ATPase-containing particles known as *19S regulatory complexes*. The 26S proteasome is responsible for degrading ubiquitylated proteins and is therefore essential for a vast array of cellular processes including cell-cycle traverse, control of transcription, regulation of enzyme levels, and apoptosis. Being the key protease of the ubiquitin system, the 26S proteasome also impacts a number of human diseases, especially cancer, cachexia, and neurodegenerative diseases. Both 20S and 26S proteasomes can associate with other protein complexes. As their name implies, proteasome activators stimulate peptide hydrolysis and may serve to localize 20S and 26S proteasomes within cells. Hybrid proteasomes consist of the 20S proteasome with a 19S regulatory complex bound at one end and a proteasome activator at the other. Immunoproteasomes are formed when the catalytic subunits found in constitutive proteasomes are replaced by interferon-inducible subunits with different substrate specificities. Immunoproteasomes are found mainly in immune tissues where they play a role in Class-I antigen presentation.

2 The 20S Proteasome

2.1 Structure

We know the molecular anatomy of archaebacterial, yeast, and bovine proteasomes in great detail since high-resolution crystal structures have been determined for all three enzymes. The archaebacterial proteasome is composed of two kinds of subunits, called α and β. Each subunit forms heptameric rings that assemble into the 20S proteasome by stacking four deep on top of one another to form a "hollow" cylinder. Catalytically inactive α-rings form the ends of the cylinder, while proteolytic β-subunits occupy the two central rings. The quaternary structure of the 20S proteasome can therefore be described as $\alpha 7 \beta 7 \beta 7 \alpha 7$. The active sites of the β-subunits face a large central chamber about the size of serum albumin (see Fig. 1). The α-rings seal off the central proteolytic chamber and two smaller antechambers from the external solvent. Archaebacterial proteasomes, with

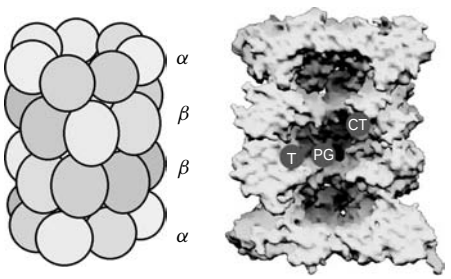

Fig. 1 20S Proteasomes. At the left is a schematic representation of a eukaryotic 20S proteasome. Different colors serve to emphasize the diversity of α and β subunits. At the right is a cutaway view of the yeast proteasome, showing the three internal cavities and the three protease activities: T (trypsin-like), PG (post glutamyl cleaving) and CT (chymotrypsin-like).

their fourteen identical β-subunits, preferentially hydrolyze the peptide bonds following hydrophobic amino acids and are therefore said to have chymotrypsin-like activity. Eukaryotic proteasomes maintain the overall structure of the archaebacterial enzyme, but they exhibit a more complicated subunit composition. There are seven different α-subunits and at least seven distinct β-subunits arranged in a precise order within their respective rings (see Fig. 1). Although current evidence indicates that only three of its seven β-subunits are catalytically active, the eukaryotic proteasome cleaves a wider range of peptide bonds containing, as it does, two copies each of trypsin-like, chymotrypsin-like, and post-glutamyl-hydrolyzing subunits. For this reason, it is capable of cleaving almost any peptide bond, having difficulty only with proline-X, glycine-X, and, to a lesser extent, with glutamine-X bonds.

2.2 Enzyme Mechanism and Proteasome Inhibitors

Whereas standard proteases use serine, cysteine, aspartate, or metals to cleave peptide bonds, the proteasome employs an unusual catalytic mechanism. N-terminal threonine residues are generated by self-removal of short peptide extensions from the active β-subunits, and they act as nucleophiles during peptide-bond hydrolysis. Given its unusual catalytic mechanism, it is not surprising that there are highly specific inhibitors of the proteasome. The fungal metabolite lactacystin and the bacterial product epoxomicin covalently modify the active-site threonines and inhibit the enzyme. Both compounds are commercially available; other inhibitors include vinylsulfones and various peptide aldehydes that are generally less specific. Recently, Velcade, a peptide boronate inhibitor of the proteasome, has been approved for the treatment of multiple myeloma.

2.3 Immunoproteasomes

Interferonγ is an immune cytokine that increases expression of a number of cellular components involved in Class-I antigen presentation. Among the IFNγ inducible components are three catalytically active β-subunits of the proteasome, called LMP2, LMP7, and MECL1. Each replaces its corresponding constitutive subunit, resulting in altered peptide-bond cleavage preferences of 20S immunoproteasomes. For example, immunoproteasomes exhibit much reduced cleavage after acidic residues and enhanced hydrolysis of peptide bonds following branch-chain amino acids such as isoleucine or valine. Class-I molecules preferentially bind peptides with hydrophobic or positive C-termini,

and proteasomes generate the vast majority of Class-I peptides. Hence, the observed β-subunit exchanges are well suited for producing peptides able to bind Class-I molecules.

3 The 26S Proteasome

3.1 The Ubiquitin–proteasome System

Bacteria can express as many as five ATP-dependent proteases. By contrast, the 26S proteasome is the only ATP-dependent protease discovered so far in the nuclear and cytosolic compartments of eukaryotic cells. Because the 20S proteasome's internal cavities are inaccessible to intact proteins, openings must be generated in the enzyme's outer surface for proteolysis to occur. A number of protein complexes have been found to bind the proteasome and stimulate peptide hydrolysis. (see Fig. 2). The most important of the proteasome-associated components is the 19S regulatory complex (RC), for it is a major part of the 26S ATP-dependent enzyme that degrades ubiquitin-tagged proteins in eukaryotic cells. Ubiquitin (Ub) is treated in a separate chapter of this Encyclopedia. Still, this important protein must be briefly covered since it plays a central role in substrate recognition by the 26S proteasome.

Ubiquitin is a small, evolutionarily conserved eukaryotic protein that can be attached to a wide variety of intracellular proteins, including itself. Although Ub serves nonproteolytic roles, such as histone modification or viral budding, the protein's major function is targeting proteins for destruction. To do so,

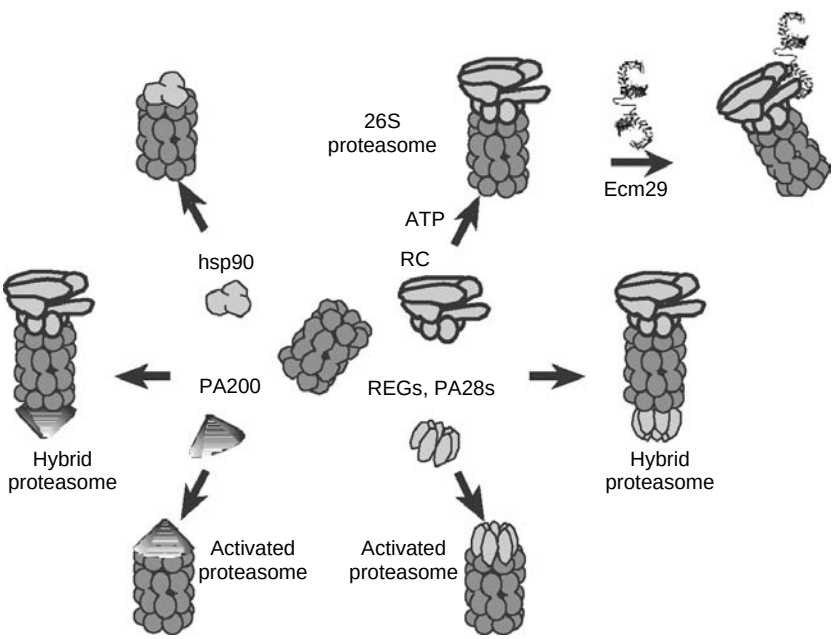

Fig. 2 Schematic representation of the 20S proteasome assembling with various activator proteins (RC, REGs, PA200) or with the chaperone hsp90, a protein that inhibits the enzyme.

the carboxyl terminus of ubiquitin is activated by an ATP-consuming enzyme (E1) and is transferred to one of several small carrier proteins (E2s) in the form of a reactive thiolester. The carboxyl terminus of an activated Ub then forms an isopeptide bond with lysine amino groups on proteolytic substrates (S) that have been selected by members of several large families of Ub ligases or E3s. Chains of Ub are formed, and the Ub-conjugated substrate is recognized by the 26S proteasome and degraded; Ub is recycled for use in further rounds of proteolysis (see Fig 3). It is important to note that ubiquitin contains seven lysine residues and polyUb chains formed via Lys6, Lys27, Lys29, Lys48, and Lys63 exist in nature. Lys29 and Lys48 chains form directly on proteolytic substrates and target them for destruction. Lys27 chains have been found on the cochaperone BAG1, and they target degradation of misfolded proteins bound by the Hsp70 chaperone to the 26S proteasome. Ub monomers linked to each other through Lys63 are involved in endocytosis and DNA repair, but not, apparently, in targeting proteins to the 26S proteasome.

3.2 Ultrastructure of the 26S Proteasome and Regulatory Complex

Electron micrographs (EMs) of purified 26S proteasomes reveal a dumbbell-shaped particle, approximately 40 nm in

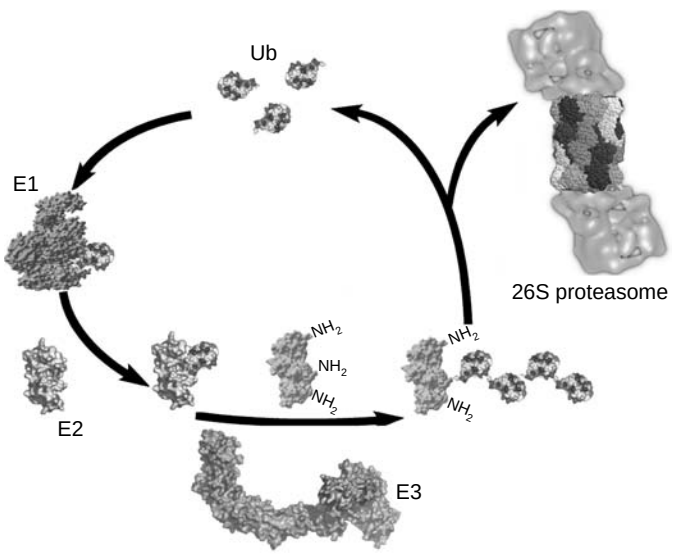

Fig. 3 Schematic representation of the ubiquitin-proteasome pathway. Ubiquitin molecules are activated by an E1 enzyme (shown in green at one-third scale) in an ATP-dependent reaction, transferred to a cysteine residue (yellow) on an E2 or Ub carrier protein and subsequently attached to amino groups (NH_2) on a substrate protein (lysozyme shown in purple) by an E3 or ubiquitin ligase, (the multicolored SCF complex). Note that chains of Ub are generated on the substrate, and these are recognized by the 26S proteasome depicted in the upper right at one-twentieth scale. (See color plate p. xxiv).

length, in which the central 20S proteasome cylinder is capped at one or both ends by asymmetric regulatory complexes looking much like Chinese dragonheads (see Fig 4). In doubly capped 26S proteasomes, the regulatory complexes face in opposite directions, indicating that contact between the proteasome's α-rings and the RC is highly specific. However, the contacts may not be especially tight since image analysis of Drosophila 26S proteasomes suggests movement of the RCs relative to the 20S proteasome. EM images of the 26S proteasome appear the same in all organisms, indicating that the overall architecture of the enzyme has been conserved in evolution. This conclusion is also supported by sequence conservation among RC subunits (see Sect. 3.4). A yeast mutant lacking the RC subunit Rpn10 contains a salt-labile RC that dissociates into two subcomplexes called the *lid* and the *base* (see Fig 4). The base contains 9 RC subunits, which include six ATPases described below in Sect. 3.4, the two largest RC subunits S1 and S2, as well as S5a; the lid contains the remaining RC subunits. Thus, the RC is composed of two subcomplexes separated on one side by a cavity, that is, the dragon's mouth. Ultrastructural studies have also been performed on the lid and on a related protein complex called the *COP9 signalosome*. Both particles lack obvious symmetry. Some particles exhibit a negative-stain-filled central groove; other classes of particles exhibit seven or eight lobes in a disc-like arrangement. Since both particles are composed of eight subunits, the lobes may represent individual subunits.

3.3 The 19S Regulatory Complex (RC)

The regulatory complex is also called the 19S cap, PA700, and the μ-particle. As its most common name suggests, the 19S regulatory complex is roughly the same size as the 20S proteasome. In fact, it is a more complicated protein assembly containing 17 or 18 different subunits, ranging in size from 25 to about 110 kDa. In animal cells, the subunits are designated S1 through S15. Homologs for each of these subunits are present in budding yeast where an alternate nomenclature has been adopted (see Table 1). Sequences for the 18 RC subunits permit their classification into a group of 6 ATPases and another group containing the 12 nonATPases.

3.4 ATPases of the RC

The six ATPases belong to the rather large family of AAA ATPases (for ATPases

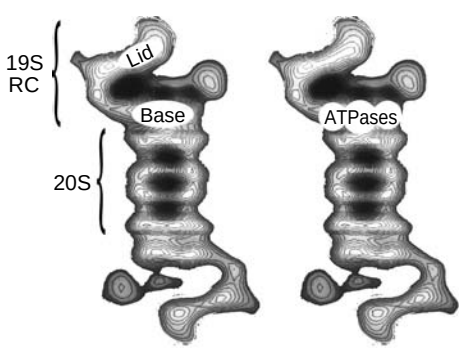

Fig. 4 Electron microscopic reconstructions of the 26S proteasome. Two images of negatively stained, doubly capped 26S proteasomes are presented to illustrate the positions of the 19S RC lid and base subcomplexes and to identify the most probable location of the RC ATPases.

Tab. 1 Subunits of the 19S regulatory complex.

Mammalian nomenclature	*Yeast nomenclature*	*Function*	*Motifs*
S1	Rpn2	Ub/Ubl binding	Leucine-rich repeats, KEKE
S2	Rpn1	Ub/Ubl binding	Leucine-rich repeats, KEKE
S3	Rpn3	?	PCI
p55	Rpn5	?	PCI
S4	Rpt2	ATPase	AAA nucleotidase
S5a	Rpn10	polyUb binding	UIM, KEKE
S5b	none	?	
S6	Rpt3	ATPase	AAA nucleotidase
S6'	Rpt5	ATPase	AAA nucleotidase
S7	Rpt1	ATPase	AAA nucleotidase
S8	Rpt6	ATPase	AAA nucleotidase
S9	Rpn6	?	PCI
S10a	Rpn7	?	PCI, KEKE
S10b	Rpt4	ATPase	AAA nucleotidase
S11	Rpn9	?	PCI
S12	Rpn8	?	MPN, KEKE
S13	Rpn11	Isopeptidase	MPN
S14	Rpn12	?	PCI

Associated with a variety of cellular Activities) whose members include the motor protein dynein, the membrane fusion factor NSF, and the chaperone VCP/Cdc48. The six ATPases, denoted S4, S6, S6′, S7, S8, and S10b in mammals, are about 400 amino acids in length and homologous to one another. On the basis of their sequences, one can distinguish three major regions: (1) a central nucleotide binding domain of about 200 amino acids, which is roughly 60% identical among members of the RC subfamily; (2) the C-terminal region, approximately 100 amino acids in length and with a lesser, though significant, degree of conservation (~40%); and (3) a highly divergent N-terminal region (<20% identity) around 120 amino acids in length; this region contains heptad repeats characteristic of coiled-coil proteins. Despite sequence differences among RC ATPases within an organism, each ATPase has been conserved during evolution with specific subunits being almost 75% identical between yeast and humans. The high degree of conservation encompasses the entire sequence, making it likely that even the divergent N-terminal regions play an important role in RC function. Conceivably, they are used to select substrates for degradation by the 26S proteasome. The "helix-shuffle" hypothesis proposes that the N-terminal coiled-coils of S4 subfamily ATPases bind unassembled substrate proteins such as fos or jun through the latter's unpaired leucine zippers. Alternatively, the variable N-terminal regions in the RC ATPases may be involved in the assembly of the RC by promoting the specific placement of the ATPase subunits within the complex. In this regard, the six ATPases associate with one another in highly specific pairs: S4 binds S7, S6 binds S8, and S6′ binds to S10b. Moreover, the N-terminal regions of RC ATPases are required for

partner-specific binding. Staining patterns of two-dimensional gels show the six RC ATPases to be present at comparable levels and the affinity capture of yeast 26S proteasomes indicate the presence one copy of each in the regulatory complex. Mutational analysis in yeast demonstrates that the ATPases are not functionally redundant since mutation of yeast S4 has a particularly profound effect on peptide hydrolysis. It is probable that the ATPases form a hexameric ring like other members of the AAA family of ATPases such as NSF or VCP/Cdc48. But this assumption has not been experimentally verified. Finally, it is quite likely that the ATPases directly bind the α-ring of the 20S proteasome. Evidence favoring this arrangement comes from chemical cross-linking experiments and the presence of the ATPases in the base subcomplex of the yeast 26S proteasome.

3.5 The non-ATPase Subunits

Whereas the six RC ATPases are homologous and relatively uniform in size, the nonATPases are heterogeneous in size and sequence. Nonetheless, they can be grouped on the basis of their location, for example, lid versus base, on the presence or absence of certain sequence motifs, for example, PCI and MPN domains, and on their affinity for Ub chains or Ub-like proteins. Eight RC subunits are found in the lid subcomplex. One of these subunits (S13) is a metalloisopeptidase that removes Ub chains from the tagged substrate prior to its translocation into the proteasome for degradation. Each of the eight lid subunits is homologous to a subunit in a separate protein complex called the COP9 signalosome. Six of the eight lid subunits contain PCI domains, stretches of about 200 residues so named from their occurrence in Proteasome, COP9 signalosome, and the eukaryotic Initiation factor 3 subunits. The PCI domains are thought to mediate subunit–subunit interactions. Two lid subunits contain 140 amino acid long MPN domains (Mpr1p and Pad1p N-terminal regions), with one of these subunits being the S13 isopeptidase. Although several models have been proposed from 2-hybrid screens of lid and COP9-signalosome subunits, the arrangement of the eight subunits within the lid subcomplex is not known with certainty. S13 stands out because it exhibits isopeptidase activity. Functions of the remaining seven lid subunits have not been discovered, although two lid subunits, S3 and S9, are critical for the degradation of specific substrates. The presence of the S13 isopeptidase in the lid explains why the lid is necessary for degradation of ubiquitylated proteins, even though the RC–base complex supports the ATP-dependent degradation of some small nonubiquitylated proteins. Interestingly, the COP9-signalosome also exhibits isopeptidase activity that removes the ubiquitin-like protein, NEDD8, from certain ubiquitin ligases.

In addition to the six ATPases, the base subcomplex contains the two largest RC subunits (S1, S2) and a smaller subunit called S5a. Besides their common location, these three subunits share the property of binding polyUb chains or Ub-like domains. S5a binds polyubiquitin chains even after it has been transferred from SDS-PAGE gels, and it displays many features that match polyubiquitin recognition by the 26S proteasome. However, S5a cannot be the only ubiquitin recognition component in the 26S proteasome because deletion of the gene-encoding yeast S5a has only a modest impact on proteolysis. This strongly suggests that there

are other ubiquitin recognition components in the RC, with S1 and S2 being prime candidates. S1 and S2 display significant homology to each other, and both can be modeled as α-helical toroids. They have been shown to bind the ubiquitin-like domains of RAD23 and Dsk2, adapter proteins that target ubiquitylated proteins to the 26S proteasome. It has also been found that the S6′ ATPase can be cross-linked to Ub. Currently, then it appears that the RC contains three, or possibly four, subunits able to recognize Ub or Ub-like proteins. As discussed below, there are other ways in which the RC can select proteins for destruction.

3.6 Biochemical Properties of the Regulatory Complex

3.6.1 Nucleotide Hydrolysis

Both the 26S proteasome and the regulatory complex hydrolyze all four nucleotide triphosphates, with ATP and CTP preferred over GTP and UTP. K_ms for hydrolysis by the 26S proteasome are two- to fivefold lower for each nucleotide and virtually identical to the Kms required for Ub-conjugate degradation. Although ATP hydrolysis is required for conjugate degradation, the two processes are not strictly coupled. Complete inhibition of the peptidase activity of the 26S proteasome by calpain inhibitor I has little effect on the ATPase activity of the enzyme. The nucleotidase activities of the RC and the 26S proteasome closely resemble those of *Escherichia coli* Lon protease, which is composed of identical subunits that possess both proteolytic and nucleotidase activities in the same polypeptide chain. Like the regulatory complex and 26S proteasome, Lon hydrolyzes all four ribonucleotide triphosphates, but not ADP or AMP.

3.6.2 Chaperone-like Activity

AAA nucleotidases share the common property of altering the conformation or association state of proteins. So, it is not surprising that the RC has been shown to prevent aggregation of several denatured proteins including citrate synthase and ribonuclease A. The chaperone activity of the RC may explain why the RC plays a role in DNA repair even in the absence of an attached 20S proteasome.

3.6.3 Proteasome Activation

The 20S proteasome is a latent protease because of the barrier imposed by the α-subunit rings on peptide entry. Consequently, a readily measured activity of the RC is activation of fluorogenic peptide hydrolysis by the 20S proteasome. The extent of activation is generally found to be in the range 3- to 20-fold. Activation is relatively uniform for all three proteasome catalytic subunits and presumably reflects opening by the attached RC of a channel, leading to the proteasome's central chamber.

3.6.4 Ubiquitin Isopeptide Hydrolysis

The channel through the proteasome α-ring into the central chamber measures 1.3 nm in diameter, a size too small to permit passage of a folded protein even as small as ubiquitin. This consideration coupled with the fact that Ub is recycled intact upon substrate degradation requires an enzyme to remove the polyUb chain prior to or concomitant with proteolysis. Several isopeptidases that remove Ub from substrates has been found associated with the 26S proteasome. Of these, S13 is an integral component of the enzyme. S13 is a metalloisopeptidase stimulated by nucleotides and is active only in the fully assembled 26S proteasome. Thus, it is not

strictly correct to list isopeptidase activity as a property of the RC.

3.6.5 Substrate Recognition

It is clear that the RC plays a predominant role in selecting proteins for degradation. This important topic is covered in Sect. 5 in the context of substrate recognition by both 20S and 26S proteasomes.

4 Proteasome Biogenesis

4.1 Subunit Synthesis

The synthesis of proteasome subunits is markedly affected by proteasome function. For example, inhibition of proteasome activity by lactacystin induces coordinate expression of both RC and 20S proteasome subunits. Similarly, impaired synthesis of a given RC subunit results in overexpression of all RC subunits. Proteasome subunit synthesis in yeast is controlled by Rpn4p, a short-lived positive transcription factor that binds PACE elements upstream of proteasome genes. Rpn4p is a substrate of the 26S proteasome, suggesting that the transcription factor functions in a feedback loop in which proteasome activity limits its concentration, thereby regulating proteasome levels. To date, an Rpn4-like factor has not been identified in higher eukaryotes; however, the presence of such a factor could explain why proteasome inhibition results in higher expression of proteasome subunits in mammalian and *Drosophila* cells.

4.2 Biogenesis of the 20S Proteasome

Proteasome β-subunits are synthesized with N-terminal extensions, and they are inactive because a free N-terminal threonine is required for peptide-bond hydrolysis. The precursor β-subunits assemble with α-subunits to form half proteasomes composed of one α-and one β-ring. These two ring intermediates dimerize to form the 20S particle, and the N-terminal extensions are removed, thereby generating a new unblocked N-terminal threonine in the catalytically active β-subunits. A small accessory protein called Ump1 in yeast and POMP or proteassemblin in mammalian cells assists in the final assembly of the 20S proteasome. Interestingly, Ump1/POMP is apparently trapped in the proteasome's central chamber and degraded upon maturation of the enzyme.

4.3 Biogenesis of the RC

Assembly pathways for the RC are virtually unknown. As mentioned above, the ATPases interact with one another and complexes containing all six S4 subfamily members have been observed following *in vitro* synthesis. Impaired synthesis of a lid subunit can result in the absence of the entire lid, so presumably, lid and base subcomplexes assemble independently and associate in the final stages of RC formation cells. *In vivo*, 26S proteasomes assemble from preformed regulatory complexes and 20S proteasomes.

4.4 Posttranslational Modification of Proteasome Subunits

Proteasome and RC subunits are subjected to a variety of posttranslational modifications including phosphorylation, acetylation, and even myristoylation in the case of the RC ATPase S4. In yeast, all seven α-subunits as well as two

β-subunits are acetylated. Since acetylation of the N-terminal threonine in an active β-subunit would poison catalysis, it has been suggested that the propeptide extensions function to prevent acetylation. Three members of the S4 ATPase subfamily (S4, S6, and S10b) and two 20S α-subunits (C8 and C9) are known to be phosphorylated. Phosphorylation appears to be particularly important for 26S proteasome assembly since the kinase inhibitor staurosporine reduces 26S proteasome levels in mouse lymphoma cells.

4.5 Biogenesis of the 26S Proteasome

The RC and 20S proteasome associate to form the 26S proteasome in the presence of ATP. Comparison of the cross-linking patterns of RC and assembled 26S proteasomes indicates that this association is accompanied by subunit rearrangement. In yeast, two proteins play a special role in 26S proteasome assembly or stability. Nob1p is a nuclear protein required for biogenesis of the 26S proteasome. It is degraded following assembly of the 26S enzyme, suffering the same fate as Ump1 does following 20S maturation. The molecular chaperone Hsp90 also plays a role in the assembly and maintenance of yeast 26S proteasomes since functional loss of Hsp90 results in 26S proteasome dissociation.

5 Substrate Recognition by Proteasomes

5.1 Degradation Signals (Degrons)

One of the major insights of twentieth century cell biology was the recognition that proteins possess built-in signals, targeting them to specific locations within cells. Selective proteolysis can be considered to be targeting out of existence, and a number of short peptide motifs have been discovered to confer rapid proteolysis on their bearers. These include PEST sequences, destruction boxes, KEN boxes, and even the N-terminal amino acid. These motifs are recognized by one or more ubiquitin ligases that mark the substrate protein by addition of a polyUb chain. However, some proteins are degraded by the 26S proteasome without prior marking by Ub. Denatured proteins are also selectively degraded by both 26S and 20S proteasomes. It is not clear what features of denatured proteins are recognized by proteasomes or by components of the Ub proteolytic system.

5.2 Ubiquitin-dependent Recognition of Substrates

Most well-characterized substrates of the 26S proteasome are ubiquitylated proteins, so our discussion starts with them. Efficient proteolysis of ubiquitylated proteins by the 26S proteasome requires a chain containing at least four Ub monomers. This matches well with the Ub-binding characteristics of the RC subunit S5a. It also selectively binds ubiquitin polymers composed of four or more ubiquitin moieties and exhibits increased affinity for longer chains. In addition, binding to S5a is impaired by mutations in ubiquitin that allow chain formation but reduce the targeting competence of the chains. These properties and the fact that denatured S5a can readily regain affinity for polyUb suggest that multiple, short sequences within S5a form "loops" that are able to bind "grooves" in the polyubiquitin

chain. S5a molecules from a number of higher eukaryotes contain two repeated motifs that are independent polyubiquitin binding sites. Each motif is approximately 30-residues long and is characterized by five hydrophobic residues that consist of alternating large and small hydrophobic residues, for example, leu-ala-leu-ala-leu. Similar motifs have been found in other proteins of the ubiquitin system and are now called UIMs, for Ubiquitin Interacting Motifs. Because a hydrophobic patch on the surface of ubiquitin is critical for substrate targeting, current models envision direct hydrophobic interaction between the UIMs in S5a and the hydrophobic patches on ubiquitin molecules in polyUb chains.

Whereas S5a provides for direct recognition of polyubiquitylated substrates, a second mechanism involves adaptor proteins possessing both a UbL (a Ubiquitin-like domain) and one or more UbA domains (Ubiquitin-associated domains are polyUb binding domains found in several proteins of the Ub system). The adaptor proteins include RAD23 and Dsk2 in yeast and recruit substrates to the 26S proteasome through UbL binding to 26S proteasome subunits, while the UbA domain binds substrate-tethered Ub chains. In yeast, the RC subunits S1 and S2 serve as UbL binding components; in mammals, S5a also serves this purpose.

The cochaperone BAG1 illustrates a third way in which polyUb can target substrate proteins to the 26S proteasome. In this case, the substrate is not polyubiquitylated; rather, it is bound to the chaperone Hsp70. A polyUb chain, linked through Lys27, is attached to the Hsp70-associated cochaperone BAG1. Apparently, the Lys27 chain promotes association of the chaperone–substrate complex with the 26S proteasome, after which the substrate is degraded while BAG1, Hsp70, and ubiquitin are recycled. Direct interaction between E3 ubiquitin ligases and RC subunits can also deliver ubiquitylated substrates to the protease. The yeast E3 UFD4 binds RC ATPases, and UFD4-mediated delivery of substrates bypasses the requirement for S5a. In what appears to be a similar delivery system, the mammalian E3 Parkin uses a UbL to bind the 26S proteasome. Mutational analyses in yeast have shown that whereas deletion of either S5a or Rad23 has mild impact on proteolysis, loss of both proteins produces a severe phenotype. Yeast lacking S5a, RAD23, and Dsk2 are not viable, indicating that direct delivery by E3s cannot compensate for the absence of all three proteins.

5.3 Substrate Selection Independent of Ubiquitin

The 26S proteasome also degrades nonubiquitylated proteins. The short-lived enzymeornithine decarboxylase (ODC) and the cell-cycle regulator p21Cip provide well-documented examples of Ub-independent proteolysis by the 26S enzyme. ODC degradation is stimulated by a protein called *antizyme* that binds to both ODC and the 26S proteasome. Antizyme functions as an adapter much like RAD23 and Dsk2 except that polyUb chains are not involved, although interestingly, free Ub chains compete with antizyme–ODC for degradation. p21Cip is also degraded in a nonUb-dependent reaction. This was clearly demonstrated by substitution of arginines for all the lysine residues in p21Cip, thereby preventing ubiquitylation of p21Cip. The lysine-less protein was still degraded in human fibroblasts by the proteasome. The C-terminal region of p21Cip binds to the proteasome α-subunit C8,

and in vitro p21Cip is degraded by the 20S proteasome alone. Direct binding of p21Cip to the 20S proteasome may open a channel through the α-ring, allowing the loosely structured protein to enter the central chamber. c-Jun, "aged" calmodulin, troponin C, and p53 are other proteins that can be degraded by the 26S proteasome absent marking by Ub. The destruction of p53 and p21Cip can proceed by both Ub-dependent and Ub-independent pathways. Other 20S proteasome substrates, *in vitro* at least, include oxidized proteins, small denatured proteins, and loosely folded proteins such as casein. Whether the 20S proteasome degrades proteins within cells is an unresolved problem.

6 Proteolysis by the 26S Proteasome – Mechanism

6.1 Contribution of Chaperones to Proteasome-mediated Degradation

Chaperones are connected to proteasomes in at least four ways. First, chaperones can deliver substrates to the proteasome as described above for the cochaperone BAG1. In a similar fashion, the chaperone VCP/Cdc48 is required for the degradation of several Ub pathway substrates. VCP is a member of the AAA family of ATPases. The large hexameric ATPase appears to function as a protein separase able to remove ubiquitylated monomers from multisubunit complexes. In some cases, the liberated proteins are degraded by the 26S proteasome; in other cases, the separated proteins may change their intracellular location. The proteasome also degrades endoplasmic-reticulum membrane proteins. If these ER membrane proteins possess a large cytoplasmic domain, their proteasomal degradation can require Hsps 40, 70, and 90 as well as VCP. Hsp90 is required to assemble and stabilize the yeast 26S proteasome, providing a third connection between chaperones and proteasomes. Hsp90 is also able to bind and suppress peptide hydrolysis by the 20S proteasome. Finally, both chaperones and proteasomes are induced by the accumulation of denatured proteins within eukaryotic cells.

6.2 Presumed Mechanism

Proteolysis of ubiquitylated proteins by the 26S proteasome can be thought to be consisting of seven steps: (1) chaperone-mediated substrate presentation; (2) substrate association with RC subunits; (3) substrate unfolding; (4) detachment of polyUb from the substrate; (5) translocation of the substrate into the 20S proteasome central chamber, (6) peptide-bond cleavage; and (7) release of peptide products as well as polyUb (see Fig 5). Step 1 is optional, depending on the substrate, and in principle, steps 3 and 4 could occur in either order. Step 4 is unnecessary for substrates like ODC that are not ubiquitylated. The other steps almost have to occur as presented. Although it is easy to conceptualize the reaction sequence, few experimental findings bear directly on any of the proposed subreactions, and virtually nothing is known about molecular movements within the 26S proteasome. It has been shown that steps 6 and 7 are not required for sequestration of ODC by the 26S proteasome. But which RC subunits actually recognize ODC–antizyme complexes has not been discovered. For ubiquitylated substrates, it is very likely

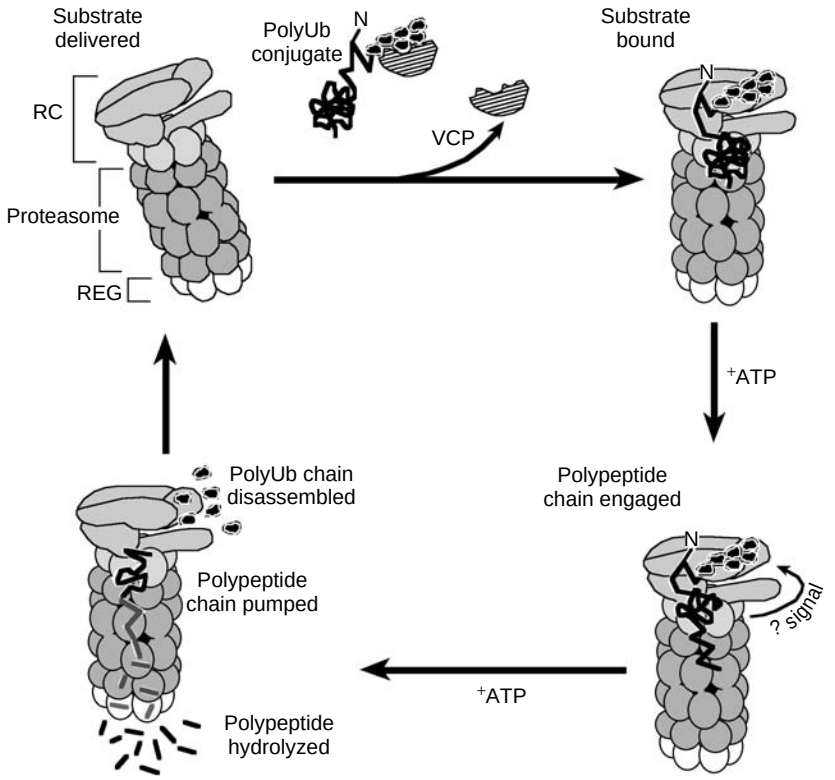

Fig. 5 Hypothetical reaction cycle for the 26S proteasome. A polyubiquitylated substrate is delivered to the 26S proteasome possibly by chaperones such as VCP (step 1). Substrate is bound by polyubiquitin-recognition components of the RC until the polypeptide chain is engaged by the ATPases (step 2). As the polypeptide chain is unfolded and pumped down the central pore of the proteasome, a signal is conveyed to the metalloisopeptidase to remove the polyUb chain (step 3). The unfolded polypeptide is eventually degraded within the inner chamber of the proteasome (step 4) and peptide fragments exit the enzyme.

that S1, S2, and S5a are recognition components, but this is not certain. Even if they are, we do not know whether they recognize just polyUb or polyUb and portions of the substrate. Translocation is thought to proceed by the six ATPases threading the polypeptide through a channel in the 20S proteasome's α-ring, but the possibility that convulsive movements transfer substrate has not been ruled out. It is also thought that the RC ATPases processively unravel substrates from degradation signals within the polypeptide chain and are able to "pump" the polypeptide chain in either N-terminal to C-terminal or the opposite direction. The ATPases may even be capable of transferring loops into the 20S enzyme. One thing that is certain is that the peptide fragments generated in the central chamber are generally 5 to 10 residues in length, but fragments as long as 35 amino acids can be present. How these fragments exit the central chamber is not known. Clearly, there is much to be discovered about the internal workings of the 26S proteasome.

6.3
Processing by the 26S Proteasome

In some cases, the 26S proteasome partially degrades the substrate protein releasing processed functional domains. The best-studied example of processing involves the transcriptional activator NFκB. The C-terminal half of a 105-kDa precursor is degraded by the 26S proteasome to yield a 50-kDa N-terminal domain that is the active transcription component. A glycine-rich stretch of amino acids at the C-terminal boundary of p50 is an important factor in limiting proteolysis. It is possible that polypeptide translocation by the RC starts at the glycine-rich region and proceeds in only one direction because of the presence of the tightly folded N-terminal domain. Or the RC may start translocation at the C-terminus and stop when the ATPases encounter the glycine-rich region. Another example of partial processing involves SPT23, a yeast protein embedded in the endoplasmic-reticulum membrane. SPT23 controls unsaturated fatty acid levels and membrane fluidity regulates 26S proteasomal generation of a freely diffusible transcription factor from the SPT23 precursor. Partial processing may be a more widespread regulatory mechanism than is currently thought.

7
Proteasome Activators

In addition to the RC, there are two protein complexes, REG$\alpha\beta$ and REGγ, and a single polypeptide chain, PA200, that bind the 20S proteasome and stimulate peptide, but not protein, degradation (see Fig. 2). Like the RC, proteasome activators bind the ends of the 20S proteasome, and importantly, they can form mixed or hybrid 26S proteasomes in which one end of the 20S proteasome is associated with a 19S RC and the other is bound to a proteasome activator. This latter property raises the possibility that proteasome activators serve to localize the 26S proteasome within eukaryotic cells.

7.1
REGs or PA28s

There are three distinct REG subunits called $\alpha\beta\gamma$. REG$\alpha\beta$ form donut-shaped hetero-heptamers found principally in the cytoplasm, whereas, REGγ forms a homo-heptamer located in the nucleus. REG$\alpha\beta$ are abundantly expressed in immune tissues while REGγ is highest in brain. The REGs also differ in their activation properties. REG$\alpha\beta$ activate all three proteasome active sites; REGγ only activates the trypsin-like subunit. There is reasonably solid evidence that REG$\alpha\beta$ play a role in Class-I antigen presentation, but we have no idea what REGγ does, especially since REGγ knockout mice have almost no phenotype. The crystal structure of REGα reveals that the seven subunits form a donut-shaped structure with a central aqueous channel, and the structure of a REG–proteasome complex provides important insight into the mechanism by which REGα activates the proteasome. The carboxyl tail on each REG subunit fits into a corresponding cavity on the α-ring of the proteasome and loops on the REG subunits cause N-terminal strands on several proteasome α-subunits to reorient upward into the aqueous channel of the REG heptamer. These movements open a continuous channel from the exterior solvent to the proteasome central chamber.

7.2
PA200

A new proteasome activator, called PA200, was recently purified from bovine testis. Human PA200 is a nuclear protein of 1843 amino acids that activates all three catalytic subunits, with preference for the PGPH active site. Homologs of PA200 are present in budding yeast, worms, and plants. A single chain of PA200 can bind each end of the proteasome, and when bound, PA200 molecules look like volcanoes in negatively stained EM images. Evidence from both yeast and mammals suggests that PA200 is involved in DNA repair.

7.3
Hybrid Proteasomes

As the α-rings at each end of the 20S proteasome are equivalent, the 20S proteasome is capable of binding two RCs, two PA28s, two PA200s, or combinations of these components. In fact, 20S proteasomes simultaneously bound to RC and PA28 or PA200 have been observed, and they are called *hybrid proteasomes*. In HeLa cells, hybrid proteasomes containing PA28 at one end and an RC at the other have been estimated to be twice as abundant as 26S proteasomes capped at both ends by 19S RCs. Hybrid 26S proteasomes containing PA200 appear to be much less abundant in HeLa cells.

7.4
ECM29

Another proteasome-associated protein, called ecm29, has been identified in several recent proteomic screens. Ecm29p is reported to stabilize the yeast 26S proteasome by clamping the RC to the 20S cylinder. However, in human culture cells, ecm29p is found predominately associated with the endoplasmic-reticulum Golgi intermediate compartment, a location suggesting a role in secretion rather than stability of 26S proteasome. Ecm29p clearly associates with 26S proteasomes; whether it activates proteasomal peptide hydrolysis is unknown.

8
Protein Inhibitors of the Proteasome

A number of proteins have been found to suppress proteolysis by the proteasome. One of these is PI31, another is the abundant cytosolic chaperone, Hsp90, and a third is a proline/arginine-rich 39-residue peptide called PR39. Both PI31 and Hsp90 may affect how the proteasome functions in Class-I antigen presentation. PI31 is a 30-kDa proline-rich protein that inhibits peptide hydrolysis by the 20S proteasome and can block activation by both RC and REG$\alpha\beta$. Although surveys of various cell lines show PI31 to be considerably less abundant than RC or REG$\alpha\beta$, when overexpressed, PI31 is reported to inhibit Class-I antigen presentation by interfering with the assembly of immunoproteasomes. A number of studies have shown that Hsp90 can bind the 20S proteasome and inhibit its chymotrypsin-like and PGPH activities. Interestingly, inhibition by Hsp90 is observed with constitutive but not with immunoproteasomes, a finding consistent with proposals that Hsp90 shuttles immunoproteasome-generated peptides to the endoplasmic reticulum for Class-I presentation. PR39 was originally isolated from bone marrow as a factor able to induce angiogenesis and inhibit inflammation. Two hybrid screens showed that PR39 binds the 20S

proteasome. Apparently, PR39 affects angiogenesis and inflammation by inhibiting respectively the degradation of HIF1 or IkBα, the latter being an inhibitor of NFkB. Finally, HIV's Tat protein inhibits the 20S proteasome's peptidase activity. Tat also competes with REG$\alpha\beta$ for proteasome binding, and by doing so, Tat can inhibit Class-I presentation of certain epitopes.

9 Physiological Aspects of Proteasomes

9.1 Tissue and Subcellular Distribution of Proteasomes

Proteasomes are found in all organs of higher eukaryotes, but the degree to which the composition of proteasomes and its activators varies among tissues is largely unexplored territory. Proteasomes are very abundant in testis since the organ contains almost fivefold more 20S subunits than skeletal muscle. At the cellular level, there are about 800 000 proteasomes in a HeLa cell and roughly 20 000 proteasomes in a yeast cell. At the subcellular level, 26S proteasomes are present in cytosol and nucleus where they appear to be freely diffusible. They are not found in the nucleolus or within membrane-bound organelles other than the nucleus. When large amounts of misfolded proteins are synthesized by a cell, the aberrant polypeptides often accumulate around the centrosome in what are called "aggresomes." Under these conditions, 26S proteasomes, chaperones, and proteasome activators also redistribute to the aggresomes presumably to refold and/or degrade the misfolded polypeptides.

9.2 Physiological Importance

Deletion of yeast genes encoding 20S proteasome and 19S RC subunits is usually lethal, indicating that the 26S proteasome is required for eukaryotic cell viability. Known substrates of the 26S proteasome include transcription factors, cell-cycle regulators, protein kinases, and so on – essentially, most of the cell's important regulatory proteins. Surprisingly, even proteins secreted into the endoplasmic reticulum are returned to the cytosol for degradation by the 26S proteasome. Given the scope of its substrates, it is hardly surprising that the ubiquitin–proteasome system contributes to the regulation of a vast array of physiological processes, ranging in higher eukaryotes. Discussion of these fascinating regulatory mechanisms is beyond the scope of this article. A brief review of proteasomes and disease does, however, seem appropriate.

9.3 The Ubiquitin–proteasome System and Human Disease

As the central protease in the ubiquitin system, the 26S proteasome impacts a number of diseases, especially cancer, neurological diseases, and muscle wasting. Ubiquitin-mediated destruction of proteins driving cell division imparts directionality to the cell cycle. So it is not surprising that several cancers arise because of an impaired or overactive ubiquitin–proteasome system. VonHippel-Lindau (VHL) disease is a hereditary cancer syndrome, characterized by a wide range of malignant tumors. The protein defective in VHL is the substrate-recognition component of a Ub–ligase complex that targets

the transcription-component hypoxia-inducible factor 1 (HIF1) for destruction. HIF1 induces synthesis of VEGF, a growth factor important for angiogenesis, and many of the tumors in VHL involve the vasculature. The human papilloma virus oncoprotein E6 provides an example of an overactive Ub ligase causing cancer. E6 is a positive regulator of the Ub ligase that targets the tumor-suppressor protein p53 for degradation. Downregulation of p53 prevents apoptosis of HPV-transformed cells, thereby promoting malignancy. Angelman and Liddle's syndromes and juvenile-onset Parkinson's disease are three other human diseases involving E3 Ub ligases.

Impairment of the proteasome occurs in a number of neurodegenerative diseases, especially those characterized by formation of protein inclusions, such as Parkinson's and Huntington's diseases. In these diseases, proteasomes and ubiquitylated proteins accumulate at the inclusions. It is thought that sequestration and/or inhibition of proteasomes may impair normal protein turnover, eventually resulting in neuronal cell death.

The ubiquitin–proteasome system also plays a key role in cellular immunity. There is considerable evidence that proteasomes generate the majority of peptides presented on Class-I molecules to cytotoxic lymphocytes. Class-I antigen presentation is a major defense against viral infection, and viruses have devised a number of ways to prevent Class-I molecules from reaching the cell surface. In several cases, viruses actually hijack the Ub-proteasome system to degrade newly synthesized MHC Class-I molecules, thereby preventing surface presentation of viral peptides. Given its pervasive role in metabolic regulation, there is no doubt that many more human diseases will be shown to result from malfunction of the Ub-proteasome system.

See also Electron Microscopy in Cell Biology; Electron Microscopy of Biomolecules; Membrane Traffic: Vesicle Budding and Fusion; Motor Proteins; Ubiquitin-Proteasome System for Controlling Cellular Protein Levels.

Bibliography

Books and Reviews

Attaix, D., Briand, Y. (2003) The proteasome in the regulation of cell function, *Int. J. Biochem.* **35**, 545–755.

Bashir, T., Pagano, M. (2003) Aberrant ubiquitin-mediated proteolysis of cell cycle regulatory proteins and oncogenesis, *Adv. Cancer Res.* **88**, 101–144.

Blobel, G. (2000) Protein targeting (Nobel lecture), *Chembiochem* **1**, 86–102.

Cope, G.A., Deshaies, R.J. (2003) COP9 signalosome: a multifunctional regulator of SCF and other cullin-based ubiquitin ligases, *Cell* **114**, 663–671.

Ferrell, K., Wilkinson, C.R., Dubiel, W., Gordon, C. (2000) Regulatory subunit interactions of the 26S proteasome, a complex problem, *Trends Biochem. Sci.* **25**, 83–88.

Goldberg, A.L., Cascio, P., Saric, T., Rock, K.L. (2002) The importance of the proteasome and subsequent proteolytic steps in the generation of antigenic peptides, *Mol. Immunol.* **39**, 147–164.

Hampton, R.Y. (2002) ER-associated degradation in protein quality control and cellular regulation, *Curr. Opin. Cell Biol.* **14**, 476–482.

Hartmann-Petersen, R., Seeger, M., Gordon, C. (2003) Transferring substrates to the 26S proteasome, *Trends Biochem. Sci.* **28**, 26–31.

Hegde, A.N., DiAntonio, A. (2002) Ubiquitin and the synapse, *Nat. Rev. Neurosci.* **3**, 854–861.

Hershko, A., Ciechanover, A., Varshavsky, A. (2000) Basic Medical Research Award. The ubiquitin system, *Nat. Med.* **6**, 1073–1081.

Hill, C.P., Masters, E.I., Whitby, F.G. (2002) The 11S regulators of 20S proteasome activity, *Curr. Top. Microbiol. Immunol.* **268**, 73–89.

Hilt, W., Wolf, D.H. (2004) The ubiquitin-proteasome system: past, present and future, *Cell Mol. Life Sci.* **61**, 1545–1632.

Kopito, R.R. (2000) Aggresomes, inclusion bodies and protein aggregation, *Trends Cell Biol.* **10**, 524–530.

Lipford, J.R., Deshaies, R.J. (2003) Diverse roles for ubiquitin-dependent proteolysis in transcriptional activation, *Nat. Cell Biol.* **5**, 845–850.

McCracken, A.A., Brodsky, J.L. (2003) Evolving questions and paradigm shifts in endoplasmic-reticulum-associated degradation (ERAD), *BioEssays* **25**, 868–877.

Muratani, M., Tansey, W.P. (2003) How the ubiquitin-proteasome system controls transcription, *Nat. Rev. Mol. Cell Biol.* **4**, 192–201.

Orlowski, M., Wilk, S. (2003) Ubiquitin-independent proteolytic functions of the proteasome, *Arch. Biochem. Biophys.* **415**, 1–5.

Peters, J.M., Harris, J.R. Finley, D. (1998) *Ubiquitin and the Biology of the Cell*, Plenum Press, New York, pp. 472.

Rechsteiner, M., Realini, C., Ustrell, V. (2000) The proteasome activator 11S REG (PA28) and class I antigen presentation, *Biochem. J.* **345**, 1–15.

Rechsteiner, M., Rogers, S.W. (1996) PEST sequences and regulation by proteolysis, *Trends Biochem. Sci.* **21**, 267–271.

Reed, S.I. (2003) Ratchets and clocks: the cell cycle, ubiquitylation and protein turnover, *Nat. Rev. Mol. Cell Biol.* **4**, 855–864.

Varshavsky, A. (1996) The N-end rule: Functions, mysteries, uses, *Proc. Natl. Acad. Sci. USA* **93**, 12142–12149.

Vierstra, R.D. (2003) The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins, *Trends Plant Sci.* **8**, 135–142.

Yewdell, J.W., Hill, A.B. (2002) Viral interference with antigen presentation, *Nat. Immunol.* **3**, 1019–1025.

Zwickl, P., Baumeister, W. (2002) The Proteasome-Ubiquitin Protein Degradation Pathway, *Curr. Top. Microbiol. Immunol.* **268**, 1–213.

Primary Literature

Bogyo, M., McMaster, J.S., Gaczynska, M., Tortorella, D., Goldberg, A.L., Ploegh, H.L. (1997) Covalent modification of the active site threonine of proteasomal β subunits and the *Escherichia coli* homolog HsIV by a new class of inhibitors, *Proc. Natl. Acad. Sci. USA* **94**, 6629–6634.

Braun, B.C., Glickman, M., Kraft, R., Dahlmann, B., Kloetzel, P.M., Finley, D., Schmidt, M. (1999) The base of the proteasome regulatory particle exhibits chaperone-like activity, *Nat. Cell Biol.* **1**, 221–226.

Brooks, P., Fuertes, G., Bose, S., Murray, R.Z., Knecht, E., Rechsteiner, M.C., Hendil, K.B. (2000) Subcellular localization of proteasomes and their regulatory complexes in mammalian cells, *Biochem. J.* **346**, 155–161.

Chau, V., Tobias, J.W., Bachmair, A., Marriott, D., Ecker, D.J., Gonda, D.K., Varshavsky, A. (1989) A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein, *Science* **243**, 1576–1583.

Cope, G.A., Suh, G.S., Aravind, L., Schwarz, S.E., Zipursky, S.L., Koonin, E.V., Deshaies, R.J. (2002) Role of predicted metalloprotease motif of Jab1/Csn5 in cleavage of Nedd8 from Cul1, *Science* **298**, 608–611.

Deveraux, Q., Ustrell, V., Pickart, C., Rechsteiner, M. (1994) A 26 S protease subunit that binds ubiquitin conjugates, *J. Biol. Chem.* **269**, 7059–7061.

Elsasser, S., Gali, R.R., Schwickart, M., Larsen, C.N., Leggett, D.S., Muller, B., Feng, M.T., Tubing, F., Dittmar, G.A., Finley, D. (2002) Proteasome subunit Rpn1 binds ubiquitin-like protein domains, *Nat. Cell Biol.* **4**, 725–730.

Fenteany, G., Standaert, R.F., Lane, W.S., Choi, S., Corey, E.J., Schreiber, S.L. (1995) Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin, *Science* **268**, 726–731.

Flynn, J.M., Neher, S.B., Kim, Y.I., Sauer, R.T., Baker, T.A. (2003) Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals, *Mol. Cell* **11**, 671–683.

Gaczynska, M., Osmulski, P.A., Gao, Y., Post, M.J., Simons, M. (2003) Proline- and arginine-rich peptides constitute a novel class of

allosteric inhibitors of proteasome activity, *Biochemistry* **42**, 8663–8670.

Glickman, M.H., Rubin, D.M., Coux, O., Wefes, I., Pfeifer, G., Cjeka, Z., Baumeister, W., Fried, V.A., Finley, D. (1998) A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and elF3, *Cell* **94**, 615–623.

Groll, M., Ditzel, L., Lowe, J., Stock, D., Bochtler, M., Bartunik, H.D., Huber, R. (1997) Structure of 20S proteasome from yeast at 2.4 A resolution, *Nature* **386**, 463–471.

Guterman, A., Glickman, M.H. (2004) Complementary roles for rpn11 and ubp6 in deubiquitination and proteolysis by the proteasome, *J. Biol. Chem.* **279**, 1729–1738.

Harris, J.L., Alper, P.B., Li, J., Rechsteiner, M., Backes, B.J. (2001) Substrate specificity of the human proteasome, *Chem. Biol.* **8**, 1131–1141.

Hendil, K.B., Khan, S., Tanaka, K. (1998) Simultaneous binding of PA28 and PA700 activators to 20 S proteasomes, *Biochem. J.* **332**, 749–754.

Hofmann, K., Falquet, L. (2001) A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems, *Trends Biochem. Sci.* **26**, 347–350.

Hoppe, T., Matuschewski, K., Rape, M., Schlenker, S., Ulrich, H.D., Jentsch, S. (2000) Activation of a membrane-bound transcription factor by regulated ubiquitin/proteasome-dependent processing, *Cell* **102**, 577–586.

Johnson, E.S., Ma, P.C., Ota, I.M., Varshavsky, A. (1995) A proteolytic pathway that recognizes ubiquitin as a degradation signal, *J. Biol. Chem.* **270**, 17442–17456.

Kane, R.C., Bross, P.F., Farrell, A.T., Pazdur, R. (2003) Velcade: U.S. FDA approval for the treatment of multiple myeloma progressing on prior therapy, *Oncologist* **8**, 508–513.

Kenniston, J.A., Baker, T.A., Fernandez, J.M., Sauer, R.T. (2003) Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine, *Cell* **114**, 511–520.

Kimura, Y., Saeki, Y., Yokosawa, H., Polevoda, B., Sherman, F., Hirano, H. (2003) N-Terminal modifications of the 19S regulatory particle subunits of the yeast proteasome, *Arch. Biochem. Biophys.* **409**, 341–348.

Kleijnen, M.F., Alarcon, R.M., Howley, P.M. (2003) The ubiquitin-associated domain of hPLIC-2 interacts with the proteasome, *Mol. Biol. Cell* **14**, 3868–3875.

Lam, Y.A., Lawson, T.G., Velayutham, M., Zweier, J.L., Pickart, C.M. (2002) A proteasomal ATPase subunit recognizes the polyubiquitin degradation signal, *Nature* **416**, 763–767.

Lee, C., Schwartz, M.P., Prakash, S., Iwakura, M., Matouschek, A. (2001) ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal, *Mol. Cell* **7**, 627–637.

Leggett, D.S., Hanna, J., Borodovsky, A., Crosas, B., Schmidt, M., Baker, R.T., Walz, T., Ploegh, H., Finley, D. (2002) Multiple associated proteins regulate proteasome structure and function, *Mol. Cell* **10**, 495–507.

Liu, C.W., Millen, L., Roman, T.B., Xiong, H., Gilbert, H.F., Noiva, R., DeMartino, G.N., Thomas, P.J. (2002) Conformational remodeling of proteasomal substrates by PA700, the 19 S regulatory complex of the 26 S proteasome, *J. Biol. Chem.* **277**, 26815–26820.

Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., Huber, R. (1995) Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution, *Science* **268**, 533–539.

Meng, L., Mohan, R., Kwok, B.H., Elofsson, M., Sin, N., Crews, C.M. (1999) Epoxomicin, a potent and selective proteasome inhibitor, exhibits *in vivo* anti-inflammatory activity, *Proc. Natl. Acad. Sci. USA* **96**, 10403–10408.

Mueller, T.D., Feigon, J. (2003) Structural determinants for the binding of ubiquitin-like domains to the proteasome, *EMBO. J.* **22**, 4634–4645.

Murakami, Y., Matsufuji, S., Kameji, T., Hayashi, S., Igarashi, K., Tamura, T., Tanaka, K., Ichihara, A. (1992) Ornithine decarboxylase is degraded by the 26S proteasome without ubiquitination, *Nature* **360**, 597–599.

Neher, S.B., Sauer, R.T., Baker, T.A. (2003) Distinct peptide signals in the UmuD and UmuD' subunits of UmuD/D' mediate tethering and substrate processing by the ClpXP protease, *Proc. Natl. Acad. Sci. U S A* **100**, 13219–13224.

Peng, J., Schwartz, D., Elias, J.E., Thoreen, C.C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D., Gygi, S.P. (2003) A proteomics approach to understanding protein ubiquitination, *Nat. Biotechnol.* **21**, 921–926.

Petroski, M.D., Deshaies, R.J. (2003) Context of multiubiquitin chain attachment influences the rate of Sic1 degradation, *Mol. Cell* **11**, 1435–1444.

Reits, E.A., Benham, A.M., Plougaste, B., Neefjes, J., Trowsdale, J. (1997) Dynamics of proteasome distribution in living cells, *EMBO J.* **16**, 6087–6094.

Rubin, D.M., Glickman, M.H., Larsen, C.N., Dhruvakumar, S., Finley, D. (1998) Active site mutants in the six regulatory particle ATPases reveal multiple roles for ATP in the proteasome, *EMBO J.* **17**, 4909–4919.

Sakata, E., Yamaguchi, Y., Kurimoto, E., Kikuchi, J., Yokoyama, S., Yamada, S., Kawahara, H., Yokosawa, H., Hattori, N., Mizuno, Y., Tanaka, K., Kato, K. (2003) Parkin binds the Rpn10 subunit of 26S proteasomes through its ubiquitin-like domain, *EMBO Rep.* **4**, 301–306.

Seemuller, E., Lupas, A., Stock, D., Lowe, J., Huber, R., Baumeister, W. (1995) Proteasome from Thermoplasma acidophilum: a threonine protease, *Science* **268**, 579–582.

Sheaff, R.J., Singer, J.D., Swanger, J., Smitherman, M., Roberts, J.M., Clurman, B.E. (2000) Proteasome turnover of $p21^{Cip1}$ does not require $p21^{Cip1}$ ubiquitination, *Mol. Cell* **5**, 403–410.

Studemann, A., Noirclerc-Savoye, M., Klauck, E., Becker, G., Schneider, D., Hengge, R. (2003) Sequential recognition of two distinct sites in sigma(S) by the proteolytic targeting factor RssB and ClpX, *EMBO J.* **22**, 4111–4120.

Taxis, C., Hitt, R., Park, S.H., Deak, P.M., Kostova, Z., Wolf, D.H. (2003) Use of modular substrates demonstrates mechanistic diversity and reveals differences in chaperone requirement of ERAD, *J. Biol. Chem.* **278**, 35903–35913.

Thrower, J.S., Hoffman, L., Rechsteiner, M., Pickart, C.M. (2000) Recognition of the polyubiquitin proteolytic signal, *EMBO J.* **19**, 94–102.

Unno, M., Mizushima, T., Morimoto, Y., Tomisugi, Y., Tanaka, K., Yasuoka, N., Tsukihara, T. (2002) The structure of the mammalian 20S proteasome at 2.75 A resolution, *Structure (Camb.)* **10**, 609–618.

Ustrell, V., Hoffman, L., Pratt, G., Rechsteiner, M. (2002) PA200, a nuclear proteasome activator involved in DNA repair, *EMBO J.* **21**, 3403–3412.

van Nocker, S., Sadis, S., Rubin, D.M., Glickman, M., Fu, H., Coux, O., Wefes, I., Finley, D., Vierstra, R.D. (1996) The Multiubiquitin-Chain-Binding Protein Mcb1 is a Component of the 26S Proteasome in *Saccharomyces cerevisiae* and Plays a Nonessential, Substrate-Specific Role in Protein Turnover, *Mol. Cell Biol.* **16**, 6020–6028.

Verma, R., Chen, S., Feldman, R., Schieltz, D., Yates, J., Dohmen, J., Deshaies, R.J. (2000) Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes, *Mol. Biol. Cell* **11**, 3425–3439.

Walz, J., Erdmann, A., Kania, M., Typke, D., Koster, A.J., Baumeister, W. (1998) 26S proteasome structure revealed by three-dimensional electron microscopy, *J. Struct. Biol.* **121**, 19–29.

Whitby, F.G., Masters, E.I., Kramer, L., Knowlton, J.R., Yao, Y., Wang, C.C., Hill, C.P. (2000) Structural basis for the activation of 20 S proteasomes by 11 S regulators, *Nature* **408**, 115–120.

Wojcik, C., Yano, M., DeMartino, G.N. (2004) RNA interference of valosin-containing protein (VCP/p97) reveals multiple cellular roles linked to ubiquitin/proteasome-dependent proteolysis, *J. Cell Sci.* **117**, 281–292.

Wu-Baer, F., Lagrazon, K., Yuan, W., Baer, R. (2003) The BRCA1/BARD1 heterodimer assembles polyubiquitin chains through an unconventional linkage involving lysine residue K6 of ubiquitin, *J. Biol. Chem.* **278**, 34743–34746.

Xie, Y., Varshavsky, A. (2001) RPN4 is a ligand, substrate, and transcriptional regulator of the 26S proteasome: a negative feedback circuit, *Proc. Natl. Acad. Sci. U S A* **98**, 3056–3061.

Xie, Y., Varshavsky, A. (2002) UFD4 lacking the proteasome-binding region catalyses ubiquitination but is impaired in proteolysis, *Nat. Cell Biol.* **4**, 1003–1007.

Yamano, T., Murata, S., Shimbara, N., Tanaka, N., Chiba, T., Tanaka, K., Yui, K., Udono, H. (2002) Two distinct pathways

mediated by PA28 and hsp90 in major histocompatibility complex class I antigen processing, *J. Exp. Med.* **196**, 185–196.

Young, P., Deveraux, Q., Beal, R., Pickart, C., Rechsteiner, M. (1998) Characterization of two polyubiquitin binding sites in the 26S protease subunit 5a, *J. Biol. Chem.* **273**, 5461–5467.

Zaiss, D.M., Standera, S., Kloetzel, P.M., Sijts, A.J. (2002) PI31 is a modulator of proteasome formation and antigen processing, *Proc. Natl. Acad. Sci. U S A* **99**, 14344–14349.

Zhang, M., Pickart, C.M., Coffino, P. (2003) Determinants of proteasome recognition of ornithine decarboxylase, a ubiquitin-independent substrate, *EMBO J.* **22**, 1488–1496.

Protein Aggregation: *see* Aggregation, Protein

Protein and Nucleic Acid Enzymes

Daniel L. Purich
University of Florida College of Medicine, Gainesville, FL, USA

1 **Enzymes as Life's Actuators 29**

2 **Chemical Nature of Enzymes 30**

3 **A Brief History of Enzyme Science 30**

4 **Systematic Classification of Enzyme-catalyzed Reactions 32**

5 **Mechanochemical Enzyme Reactions 33**

6 **The Catalytic Reaction Cycle 35**

7 **Basic Enzyme Kinetics 41**

8 **Enzyme Regulation 41**

9 **Inborn Errors in Enzyme Function 43**

Bibliography 43
Books and Reviews 43

Keywords

Catalytic Antibody
An antibody that catalyzes chemical reactions.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

Catalytic Proficiency
A measure of the rate enhancement achieved in the presence of an enzyme; the dimensionless parameter v_{enz}/v_{non}, where v_{enz} and v_{non} are the reaction velocities in the presence and absence of an enzyme respectively.

Energase
A mechanochemical enzyme that catalyzes energy-dependent translocation, energy-dependent transport, and/or energy-dependent force production.

Enzyme
A biological catalyst that speeds up the making and breaking of chemical bonds without altering the reaction's equilibrium.

Protein Enzyme
A biological catalyst composed of one or more polypeptide chains.

RNA Enzyme
A biological catalyst composed of ribonucleic acid.

As is true of any catalyst, the defining quality of every enzyme is the ability to enhance chemical reactivity without altering the thermodynamic equilibrium of the reaction catalyzed. An enzyme's catalytic proficiency is measured by its rate enhancement factor, a parameter that equals v_{enz}/v_{non}, where v_{enz} and v_{non} are the reaction velocities in the presence and absence of an enzyme respectively. Although the rates of uncatalyzed reactions are often too slow to be determined under physiologic conditions, enhancement factors of enzymes are thought to range typically from 10^9 to 10^{17}, with a few approaching 10^{22} and others falling well below 100. Each catalytic cycle shares common features: (1) the enzyme must combine with its substrate, which is then converted into the transition state; (2) and this least stable reactant configuration then undergoes equipartition (i.e. half of the time, reforming EnzymeSubstrate complex, and just as often proceeding to form EnzymeProduct complex, the latter followed by product release). After each catalytic reaction cycle, an enzyme returns to its original form, such that an enzyme's action is without any effect on the reaction's equilibrium constant. Because product release is so often the slowest step in catalysis, discrete bond-making and bond-breaking steps, and even the formation of covalent intermediates, rarely hinder catalysis. The rates of highly perfected enzymes are limited only by the diffusion-limited availability of substrate; other enzymes are so sluggish that each complete catalytic cycle operates on the seconds to minutes timescale. In terms of natural selection, however, an organism accrues no competitive advantage should one of its enzymes evolve catalytic proficiency beyond that needed for efficient metabolism.

1
Enzymes as Life's Actuators

Because normal physiologic conditions (e.g. neutral pH, aqueous environment, moderate ionic strength, pressures near one atmosphere, and a narrow 0–50 °C temperature range) conspire to deactivate most chemical reactions, their uncatalyzed rates are frequently extraordinarily low. Consider, for example, the hexokinase reaction: $Mg{\cdot}ATP^{2-}$ + D-Glucose = D-Glucose 6-Phosphate + $Mg{\cdot}ADP^{1-}$. At cellular glucose and ATP concentrations, detectable glucose-6-P would require hundreds or thousands of years in the absence of hexokinase, but occurs in seconds at cellular enzyme concentrations. Even when nonenzymatic reactions are fairly fast, as is true for the reversible hydration of carbon dioxide to form bicarbonate anion, the presence of an enzyme (carbonic anhydrase, in this case) assures the reaction rate is always sufficient for effective metabolism.

The speed and specificity of enzyme-catalyzed reactions provides cells with the capacity to organize a series of these reactions into pathways responsible for the synthesis (*anabolism*) and breakdown (*catabolism*) of essential life-sustaining intermediates. With hexokinase as an example once again, exclusive phosphorylation at position-6 of the β-D-glucopyranose form would be impossible in the absence of enzyme. Even when nonenzymatic reactions are reasonably rapid, as in the case of reversible hydration of carbon dioxide to form bicarbonate anion, an enzyme is required to assure the reaction rate is sufficiently fast for efficient metabolism. In this respect, enzymes play roles analogous to on–off switches and amplifiers within electronic circuits. Substrate specificity also reduces the formation of toxic substances that would otherwise accumulate in side reactions. By coupling energy-requiring and energy-yielding reactions, enzymes also assist in managing cellular bioenergetics, thus permitting cells to store energy in many forms (e.g. as low molecular–weight metabolites, as macromolecules, as ion and solute concentration gradients, etc.). Enzymes catalyzing the first committed step in a metabolic pathway often possess regulatory sites that bind activators and inhibitors, allowing them to respond to regulatory cues as substances are produced or consumed within their own and other metabolic pathways. Moreover, because some enzymes are oligomeric (i.e. they are composed of multiple subunits), regulatory information can be transferred across their subunit–subunit interfaces in a manner that alters their catalytic efficiency in response to small fluctuations in the concentration of regulatory (or effector) molecules. This phenomenon, known as *cooperativity*, is attended by enzyme shape changes (i.e. allosteric transitions) that can greatly alter catalytic activity.

Certain highly specialized mechanochemical enzymes interact with their binding partners to adopt low- and high-affinity interaction states, the interconversion of which is driven by the Gibb free energy liberated upon the hydrolysis of certain metabolites, most often ATP or GTP. These affinity-modulated systems can also derive free energy during the dissipation of transmembrane solute concentration gradients. Such energy-transducing enzymes can also generate the forces needed for cell division and motility, solute transport, as well as intracellular trafficking, and sorting of molecules and/or organelles. By generating the

forces and mechanical work needed to power sensory processes, transport and translocation, and the ability to divide, these mechanochemical enzymes bestow upon organisms the essentially animating qualities of irritability, motility, and proliferation.

2 Chemical Nature of Enzymes

From earliest times, enzymes were invariably found to be associated with proteins. On the basis of his Nobel prize–winning work establishing the structures of heme and chlorophyll molecules, the German chemist Willstätter incorrectly asserted what he believed to be a universal principle, namely, that the catalytic entity of every enzyme would prove to be some small molecule that is bound to a protein that serves only passively as a stabilizing colloidal support. Sumner's crystallization of urease proved that the active enzyme is itself a protein, and this accomplishment earned him a Nobel Prize. However, while most known enzymes are proteins, others are partly or entirely composed of nucleic acid. Nobel laureates Cech and Altman demonstrated that certain RNA molecules are efficient catalysts of RNA self-splicing, phosphotransfer, and even peptide bond-forming reactions. These catalytic RNA molecules, also known as *ribozymes*, often achieve rate enhancements approaching 10^{11}. The hammerhead-shaped ribozyme was the first RNA motif observed to catalyze sequence- specific self-cleavage. With only around 30 nucleotides in their catalytic cores, these enzymes are among the smallest known catalytic RNAs. They display Michaelis–Menten kinetics, with Michaelis constants values ranging from 20–200 nM and turnover numbers in the range of $0.03\ s^{-1}$. Product release is generally fast, suggesting that the rate-determining step is phosphodiester bond-scission. Ribozyme-catalyzed phosphoryl transfer appears to involve destabilization of the substrate's ground state. This transesterification reaction is analogous to the catalytic mechanisms used by the mRNA spliceosome as well as DNA topoisomerases. Some evolution experts have argued that the likely early appearance of nucleic acids necessitated reliance on their catalytic properties in what might have been a "preprotein" biotic world.

One cannot dismiss the possibility that other biological substances (e.g. polysaccharides, complex lipids, etc.) may prove to be biological catalysts. The phenomenon of micellar catalysis, for example, is already firmly rooted in modern organic chemistry. In fact, some micellar catalysts even exhibit chiral recognition (i.e. the capacity to combine with and transform substrate molecules in a stereoselective manner). Whether the same holds for cellular membranes remains to be determined.

3 A Brief History of Enzyme Science

On the basis of his studies of diastase, a crude preparation of the enzyme α-amylase, the Swedish chemist Berzelius advanced the germinal concept that, to hasten product formation, a catalyst must first combine with its reactant. In keeping with his idea that a catalyst must weaken the bonds within a reactant, he coined the term *catalysis* by combining the Greek words *kata* ("break") and *lyein* ("release") to describe its ability to break apart chemical reactants. The birth of

the field of enzymology can be traced to 1810, when the French chemist and physicist Gay-Lussac demonstrated that anaerobic growth of yeast on sugars produces mainly ethanol and carbon dioxide. In 1878, Kühne coined the name enzyme, taken from the Greek words *en* ("within") and *zyme* ("yeast"), to designate the fermentative catalyst within yeast. In 1896, Büchner successfully opened yeast cells in a mortar and pestle through the abrasive action of fine grains of sand on the microorganism's cell wall. Upon addition of sugars to this cell-free extract, he noted the evolution of CO_2 bubbles, a property that he took as a telltale sign of active metabolism. Büchner's sensational conclusion that biochemical reactions can occur outside a living cell gave birth to investigative enzyme chemistry and earned him the Nobel Prize. The discipline of enzyme kinetics was firmly established in 1913, when Michaelis and Menten published the famous rate law that now bears their names. A decade later, Briggs and Haldane introduced the concept of a steady state to enzyme kinetics and metabolism. By improving rapid mixing kinetic methods, Chance detected EnzymeSubstrate complexes spectrally and characterized their subsequent conversion to product. About the same time, Racker also discovered an enzyme (glyceraldehydes 3-phosphate dehydrogenase) that formed covalent enzyme–substrate intermediates during catalysis, thus dispelling notions that formation of covalently bound intermediates would be an unproductive detour (or *umweg*, as the German Nobelist Warburg put it), serving only to retard catalysis. Surging interest in shock waves and sonar during World War II, along with Einstein's extension of Maxwell's theory of chemical relaxation to include pressure-induced perturbations, provided an impetus for investigating the individual steps (or elementary reactions) within multistep mechanisms. For their pioneering work in developing and exploiting novel techniques that detect and quantify changes in rapid chemical and biological processes, Eigen, Norrish, and Porter shared the Nobel Prize in chemistry.

Enzymology is now a highly sophisticated and innovative discipline that continues to provide powerful insights into enzyme catalysis and metabolic regulation. Among its most notable contributors are: Abeles – for mechanism-based (or suicide) inhibitor design; Albery and Knowles – for enzyme energetics and catalytic perfection; Altman and Cech – for the discovery of catalytic RNA; Benkovic and Lerner – for catalytic antibodies; Boyer – for isotope tracer kinetics, equilibrium exchange theory, and ATP synthase binding change mechanism; Chance – for stopped-flow rapid mixing techniques, heme–protein kinetics; Cleland – for nomenclature, exchange-inert metal-nucleotides, exchange kinetics, and kinetic isotope effects; Cohn – for oxygen-18 probes of phosphotransfer reactions and NMR distance measurements; Eigen, De Maeyer, and Hammes – for temperature-jump reaction techniques to enzyme systems; Fersht – for site-directed mutagenesis as mechanistic probes as well as kinetic proofreading mechanism of protein folding; Frieden – for enzyme hysteresis; Fromm – for reversible inhibitors as tools for discriminating multisubstrate systems; Hartley – for burst kinetic methods for detecting reaction intermediates; Jencks – for catalytic strategies in chemistry and enzymology, as well as the conceptual basis for catalytic antibodies; Koshland – for the induced-fit hypothesis and cooperative ligand-binding

mechanisms; Mitchell – for the chemiosmotic principle; Monod, Wyman and Changeux – for cooperativity and allosteric models; Rose – for dynamic stereochemical probes and isotope trapping methods; Westheimer – for enzyme stereochemistry, stereochemistry of phosphoryl transfer, and photoaffinity labeling; and Wolfenden – for transition-state inhibitor theory and the conceptual basis of catalytic proficiency.

4 Systematic Classification of Enzyme-catalyzed Reactions

The number of classified protein enzymes now approaches 7000; fewer than several hundred reactions are known to require RNA enzymes. Given the diversity of living organisms, there must be thousands more that catalyze reactions remaining to be defined. The Enzyme Commission and the Joint Commission on Biochemical Nomenclature have distinguished enzymes into six chemically distinct classes, based entirely on the ways that enzymes break, rearrange, and form covalent bonds (See Table 1).

For more than a half century, enzymes have been named and classified only on the basis of the making and breaking of covalent bonds. There are, however, numerous examples of biological catalysis involving the interconversion of noncovalent substrate-like and product-like states. Consider the hundreds of mechanochemical enzyme reactions (see following text) that catalyze reactions that accomplish mechanical work or that generate force. Another prominent example is the entire group of so-called nucleotide exchange factors. These proteins play important physiologic roles, as illustrated by the facilitated exchange of solution-phase GTP with tightly bound GDP on GTP-regulatory proteins (or G-Proteins), resulting in the formation of G-ProteinGTP complex and solution-phase GDP. Another example is the adenine nucleotide exchange factor known as profilin (Reaction: $\text{Actin}\cdot\text{ADP} + \mathbf{ATP}_{\text{solution-phase}} = \text{Actin}\cdot\mathbf{ATP} + \text{ATP}_{\text{solution-phase}}$, where boldface type is used to indicate that exchange is the purely physical release of bound ADP, followed by binding of **ATP**). This uncatalyzed exchange reaction is too slow to sustain the high filament elongation rates (400–500 monomer per filament per second) during actin-based motility. To overcome this kinetic barrier, the protein profilin accelerates the ADP ↔ ATP exchange reaction by a factor of 150, without any effect on the equilibrium constant for the exchange reaction. These features indicate that G-protein exchange factors and profilin must be regarded as biological catalysts, and hence as enzymes.

In the face of the growing number of instances where biological catalysis is not attended only by changes in covalent bonding, Purich offered a definition in 2001: "An enzyme as a biological catalyst that facilitates the making and/or breaking of chemical bonds." While appearing to be no more encompassing than existing definitions of enzyme catalysis, the crucial difference lies in the use of *chemical* in place of *covalent* to describe the bonding changes. In *The Nature of the Chemical Bond*, Linus Pauling offered the following definition of chemical bonds: "We shall say that there is a chemical bond between two atoms or groups of atoms in case that the forces acting between them are such as to lead to the formation of an aggregate with sufficient stability to make it convenient

Tab. 1 Systematic classification of enzyme-catalyzed reactions

Class 1: Oxidoreductases catalyze redox reactions.	
Dehydrogenase:	Ethanol + NAD^+ = Acetaldehyde + NADH + H^+
Class 2: Transferases catalyze the transfer of functional groups.	
Sugar kinase:	Glucose + $MgATP^{2-}$ = Glucose = 6-P^{1-} + $MgADP^{1-}$
Aminotransferase:	Oxaloacetate + Glutamate = Aspartate + 2-Oxoglutarate
Class 3: Hydrolases catalyze hydrolysis reactions.	
Esterase:	Ester + H_2O = Acid + Alcohol
Amidase:	Glutamiine + H_2O = Glutamic Acid + Ammonia
Phosphatase:	Glucose-6-P^{2-} + H_2O = Glucose + $HPO_4{}^{2-}$
Class 4: Lyases catalyze addition/elimination reactions.	
Fumarase:	Malate = Fumarate + H_2O.
Aspartase:	Aspartate = Oxaloacetate + NH_3
Class 5: Isomerases catalyze intramolecular rearrangements.	
Isomerase:	Glucose-6-P^{2-} = Fructose-6-P^{2-}
Class 6: Ligases catalyze joining reactions.	
Synthetase:	Glutamate + NH_3 + $MgATP^{2-}$ = Glutamine + $MgADP^{1-}$
Polymerase:	DNA_n + dNTP = DNA_{n+1} + PP_i

for the chemist to consider it as an independent molecular species.'' Note that Pauling made no comment about covalent bonding. In this respect, because many protein conformational states and numerous ProteinLigand complexes are sufficiently long lived to exhibit chemically definable properties, their formation and/or transformation must be considered as chemical reactions. The same logic applies to the persistent, chemically definable position of a solute relative to the faces of a membrane, representing substrate-like and product-like states. For this reason, transporters should be regarded as enzymes. Finally, Purich in 2001 suggested that each mechanochemical ATPase can be termed an *energase*, a name he offered for those specialized enzymes catalyzing energy-dependent translocation, energy-dependent transport, energy-dependent force production, and so on. He also suggested that these mechanochemical enzymes should be considered as an entirely separate class, as these enzymes cannot be rationally accommodated in the current Enzyme Commission classification.

5 Mechanochemical Enzyme Reactions

There are literally hundreds of catalyzed force-generating and/or work-producing reactions that culminate in a change of some molecular condition (i.e. transport against a concentration gradient, locomotion against an opposing load, or stabilization of an otherwise unstable conformation/configuration). These reactions are typically driven by the Gibbs free energy liberated upon the hydrolysis of such molecules as ATP, GTP, P-enolpyruvate, and acetyl-phosphate. With the input of sufficient Gibbs free energy, these mechanochemical enzymes facilitate their reactions by interconverting between two or more noncovalent substrate- and product-like interactions states:

$$\text{Condition}_{\mathbf{A}} + \cdot ENZ \cdot \text{A}\mathbf{T}\text{P} = \text{Condition}_{\mathbf{A}} \cdot ENZ \cdot \mathbf{AT}\text{P} \quad (1)$$

$$\text{Condition}_{\mathbf{A}} \cdot ENZ \cdot \text{ATP} = \text{Condition}_{\mathbf{A}} \cdot ENZ \cdot \text{A}\mathbf{D}\text{P} \cdot \text{Pi} \quad (2)$$

$$\text{Condition}_{\mathbf{A}} \cdot ENZ \cdot \text{A}\mathbf{D}\text{P} \cdot \text{Pi} = \text{Condition}_{\mathbf{B}} \cdot ENZ \cdot \text{A}\mathbf{D}\text{P} + \text{Pi} \quad (3)$$

$$\text{Condition}_{\mathbf{B}} \cdot ENZ \cdot \text{ADP} + \text{A}\mathbf{T}\text{P} = \text{Condition}_{\mathbf{B}} + \cdot ENZ \cdot \text{A}\mathbf{T}\text{P} \quad (4)$$

$$\text{Condition}_{\mathbf{A}} + \text{ATP} = \text{Condition}_{\mathbf{B}} + \text{ADP} + \text{Pi} \quad \sum(1\text{–}4)$$

Note that the above reaction forms a catalytic reaction cycle, culminating in recycling of the enzyme to its initial form. In some cases (e.g. microtubule assembly/disassembly), the entire process may also require a considerable period of time to complete its cycle.

These reactions utilize Gibbs free energy liberated upon the hydrolysis of molecules such as ATP, GTP, P-enolpyruvate, and acetyl-phosphate. While certain bonds within these substances are thermodynamically highly unstable, they are highly stable kinetically. Mechanochemical enzymes share the capacity to catalyze the hydrolysis of these P—O—P or C—O—P anhydride bonds, thereby triggering a conformational transition from one interaction state to another.

Active solute transporters promote the transport of ionic and neutral solutes by forming a high-affinity TransporterSolute complex on the side of the membrane where the solute concentration is low. Hydrolysis of transporter-bound ATP subsequently drives the conversion of the high-affinity TransporterSolute complex into its low-affinity TransporterSolute complex, such that solute release is facilitated on the membrane's high-solute concentration side (see Table 2). In a similar manner, molecular motors power contractile processes, intracellular organelle trafficking, and cell crawling. Likewise, the building up and tearing down of supramolecular structures depends on ATP/GTP-driven affinity-modulated interactions (e.g. regulatory G-proteins, chaperonins, and proteasomes, and various other latching and switching molecules). The field of cell biology can be regarded as structural metabolism, where the supramolecular components are continually formed, remodeled, and degraded. Endocytosis and organelle traffic, cell crawling, signal transduction, and mitosis/meiosis are processes that are taking on the appearance of the pathways of intermediary metabolism. Even long-term potentiation, a neuronal process lying at the root of our memory and consciousness, is now known to depend on actin polymerization motors that remodel and maintain structures known as dendritic spines into functional neuronal synapses.

The essence of these energase-type enzymes is a molecular discrimination (or affinity-modulating) mechanism comprised of a succession of conformational states that transit between high- and low-affinity ligand interaction states. In a similar fashion, electron transport systems convert the free energy of NADH oxidation into the proton-motive force (pmf) driving ATP synthesis in oxidative phosphorylation. In its well-choreographed succession of electron transfer steps, a

Tab. 2 Mechanochemical enzymes: Affinity-modulated clamps, latches and switches

Reaction	*Energy source*	*High-affinity interaction partner*
Transporters	ATP	Ligand on low-concentration side of membrane
ATP synthase	ATP	Itself ($\alpha\beta-\gamma$ subunit interface)
Myosin	ATP	Actin filament
Electron transport chain	emf	Proton on high pH side of gradient
Dynein and kinesin	GTP	Microtubule
GTP-regulatory proteins	GTP	Target protein
DNA topoisomerase (type II)	ATP	Itself (at its dimer interface)
ATP sulfurylase	GTP	Active site
Actoclamplin	ATP	Clampin (Ena/VASP, N-WASP and formin protein families)

membrane-bound electron transport protein complex undergoes conformational change(s) altering the acidity of one or more acid-base groups, thus permitting the transmembrane protein to extrude one or more protons across a membrane bilayer. After NADH is formed upon the oxidation of foodstuffs, it undergoes oxidation (a net loss of electrons) and initiates the chemical reduction (a net gain of electrons) of electron transport proteins (ETP) situated in the inner mitochondrial membrane. Subsequent chemical reduction of molecular oxygen is mediated by various ETP complexes, resulting in the accumulation of protons within a submitochondrial compartment. Precisely how the electromotive force (emf) of electrochemical gradient generates the proton-motive force, remains an area of intense investigation. The current view is that electron transport drives a conformational change that lowers the affinity of carboxyl groups within the ETP complex for their protons. When viewed in this manner, the action of these multiprotein ETP complexes represents a highly specialized example of molecular discrimination.

6 The Catalytic Reaction Cycle

The long-standing problem in biological catalysis concerns the origin of enzyme rate enhancement and the functional coupling of enzyme and reactant in transformations of EnzymeSubstrate, EnzymeIntermediate, and EnzymeProduct complexes. The following table (Table 3) lists some of the ways in which an enzyme is thought to increase chemical reactivity. There is presently no single enzyme-catalyzed reaction for which the precise contribution of each factor has been unambiguously determined. The alert reader will also recognize that many of these factors are surely acting simultaneously and synergistically, and in various degrees, at defined stages in the reaction cycle.

Of all the explanations advanced to account for the catalytic proficiency of enzymes, the most appealing is that enzymes promote catalysis by stabilizing reaction transition states. This concept was assumed implicitly in writings of Haldane in 1930 and was explicitly stated by Pauling in 1940: "In order to catalyze a reaction an enzyme must recognize the transition state in a selective way (i.e. it must

Tab. 3 Physicochemical factors thought to increase catalytic proficiency

Approximation	Enzymes increase the local effective reactant concentration by bringing reactants (as well as essential coenzymes and metal ion cofactors) into the vicinity of catalytic groups within the enzyme's active site.
Orientation	Enzymes adjust the mutual orientation of reactive groups on the substrate(s) and enzyme.
Push–pull catalysis	Enzymes have suitably positioned acid and base groups that can provide ("push") or withdraw ("pull") protons to and from the substrate, product or reaction intermediate at appropriate times in the reaction cycle.
Stereoselection	Enzymes possess the capacity to serve as three-dimensional templates that discriminate between various reactants on the basis of the stereochemical arrangement of their atoms.
Chelation	Enzymes and substrates often employ transition metal ions to form "bridge" complexes (e.g. enzyme–metal–ion substrate; enzyme–substrate–metal ion; or cyclic complexes as well).
Lewis acid effects	Enzymes use metal ions as electron-pair acceptors to polarize otherwise unreactive groups on the substrate or enzyme and/or to alter their acid/base properties.
Redox effects	Enzymes alter the redox properties of certain metal ions to transiently achieve oxidation states that would be otherwise unavailable in aqueous solutions.
Desolvation	Enzymes remove water molecules from the active site in order to increase the nucleophilic reactivity of functional groups on the substrate or enzyme.
Hydrophobic effects	Enzymes can alter the polar and apolar microenvironment within the active site to increase/decrease reactivity.
Proton shuttling	Enzymes create extended chains of hydrogen bonded water molecules that serve as "proton wires" for providing or removing protons from the active site on an as-needed basis.
Transition-state stabilization	Enzymes have active sites that are structurally complementary to one or more activated complexes formed along the reaction trajectory.
Ground-state stabilization	Enzymes can promote catalysis by destabilizing an otherwise unreactive ground-state enzyme–substrate complex.
Entropic effects	Enzymes can temporarily alter the thermodynamic properties of bound reactants in a manner that favors attainment of a highly reactive configuration.
Covalent catalysis	Enzymes can transiently form covalent intermediates that facilitate catalysis by managing the group transfer potential of reactants or by masking/unmasking functional groups to promote catalysis.
Entatic effects	Enzymes can speed up reactions by exerting substantial push–plull forces on reacting species, especially over short distances.

stabilize it better than the substrate).'' The following diagram (Fig. 1) illustrates this fundamental concept by comparing an uncatalyzed reaction proceeding through a highly unstable transition-state ($X^{\ddagger}$) and a hypothetical enzyme operating by a three-stage mechanism entailing the formation of an Enzyme·Substrate complex (E·S), an Enzyme·Transition State ($E{\cdot}X^{\ddagger}$), and an Enzyme·Product complex (E·P). Note that the stabilization free energy may be written as: $\Delta\Delta G^{\ddagger} = \Delta G^{\ddagger}_{\text{non}} - \Delta G^{\ddagger}_{\text{enz}}$, where $\Delta G^{\ddagger}_{\text{non}}$ is the energy change for forming the nonenzymatic transition state, and $\Delta G^{\ddagger}_{\text{enz}}$ is the energy change for forming the enzymatic transition state. On the basis of absolute rate theory, the $\nu_{\text{non}}/\nu_{\text{enz}}$ may be estimated from the relationship: $\nu_{\text{non}} = \nu_{\text{enz}} e^{-\Delta\Delta G^{\ddagger}/RT}$, where R and T are the universal gas constant and the absolute temperature.

Hammes in 2002 offered the view that ''when a substrate binds to an enzyme, it becomes an integral part of the macromolecule. The subsequent dynamics of the macromolecular conformational changes are then the catalytic process itself. This view of catalysis means that the making and breaking of noncovalent bonds within the structure are part of the catalytic process, and that these events can occur close to and far from the active site. The advantage of having hundreds of intramolecular interactions dynamically involved in catalysis is that the energetics of the reaction can be easily manipulated to produce catalysis, and extremely fine tuning is provided by hundreds of intramolecular interactions. This mechanism could be viewed as a ''gear shift'' mechanism: the conformational transitions are analogous to shifting gears, and the interactions between the enzyme and substrate correspond to the gear coupling mechanism. Asking what ''drives'' the reaction is not terribly meaningful, as the essence of cooperative processes is that many events are occurring essentially simultaneously.''

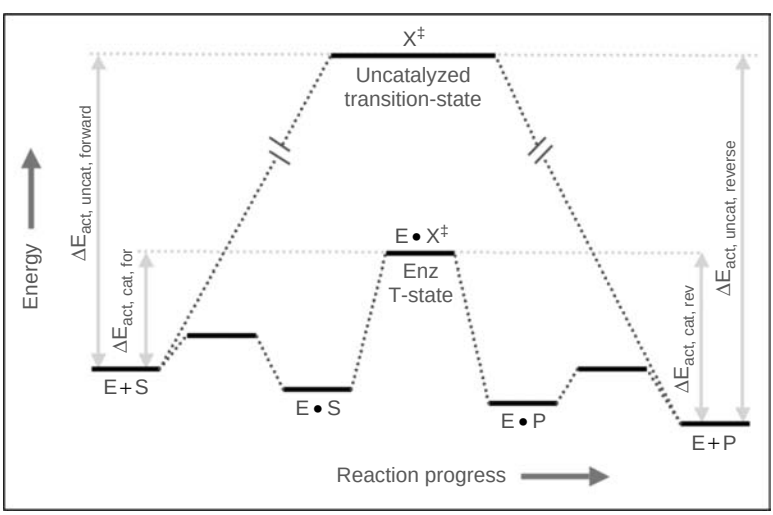

Fig. 1 Reaction coordinate diagram illustrating the steps in an uncatalyzed reaction (upper pathway) and those in an enzyme-catalyzed process (lower pathway). Note that the energy of the initial and final states are the same for the enzymatic and nonenzymatic processes.

These fundamentally important ideas are tantamount to saying that an enzyme creates a catalytic environment, one that dynamically promotes the reaction trajectory from substrate to product. Although $E{\cdot}X^{\ddagger}$ has been widely employed to represent the transition state, the latter should be represented as $(E{\cdot}X)^{\ddagger}$ to indicate that enzyme and substrate are mutually altered as the catalytic reaction cycle proceeds. That enzyme and substrate jointly reach the transition state also complies with the principle of thermodynamic reciprocity, which requires that substance a cannot affect substance without the latter likewise affecting the former. Traditional use of $E{\cdot}X^{\ddagger}$ leaves a mistaken impression that the enzyme is an immutable template, such that only the substrate reaches the activated transition-state configuration.

The chief ambition of enzyme chemists is to understand the catalytic process used by enzymes by obtaining a complete description of all of the physical and chemical events occurring during an enzyme-catalyzed reaction. Such a daunting undertaking inevitably leads to the formulation of a reaction mechanism, a chemical model or framework consisting of all of the elementary reactions comprising the overall catalytic process. The mechanism must be consistent with the reaction's overall stoichiometry and stereochemistry, its kinetic and thermodynamic properties, the structures of detected intermediates and inferred transition state(s), the energetics of intermediates involved in rate-determining step(s), as well as effects of temperature, ph, ionic strength, and solvent. modern enzymology has benefited enormously from the atomic level molecular structures, as provided by X-ray crystallography and high-resolution, multidimensional nmr spectroscopy. even so, while structural biologists have glimpsed various stages of catalysis, there is no such thing as a "tell-all motion picture" of even the simplest catalytic process. Enzyme chemists must also rely on stereochemistry to learn whether certain covalent intermediates are formed during catalysis. The optimal approach is to construct hypotheses offering testable predictions that can be recursively tested in kinetic experiments designed to discriminate among rival explanations or at least to suggest a more definitive experiment.

Intensive investigation of proteolytic enzymes over the past half century has culminated in a self-consistent mechanism for chymotrypsin catalysis (Fig. 2). This enzyme cleaves peptide bonds where the carboxyl group–donating residue contains a hydrophobic side chain. The reaction is facilitated by push–pull acid/base interactions involving three residues common to hundreds of other mechanistically related "serine" proteases. An acyl-serine intermediate permits one product (designated by the R-group) to dissociate such that water can replace the departing amino group in a manner that leads to hydrolysis of the peptidyl-acyl-enzyme and subsequent release of the second peptide fragment (designated by R′).

Enzyme chemists are reasonably confident of the general outline of this reaction mechanism, especially in the light of the wealth of structural, chemical, and kinetic information gleaned from persistent and systematic investigation. That said, there remain obvious voids in our understanding of chymotrypsin catalysis. A major limitation stems from an almost exclusive reliance on data obtained from the use of synthetic chromogenic substrates (i.e. molecules that generate spectral change

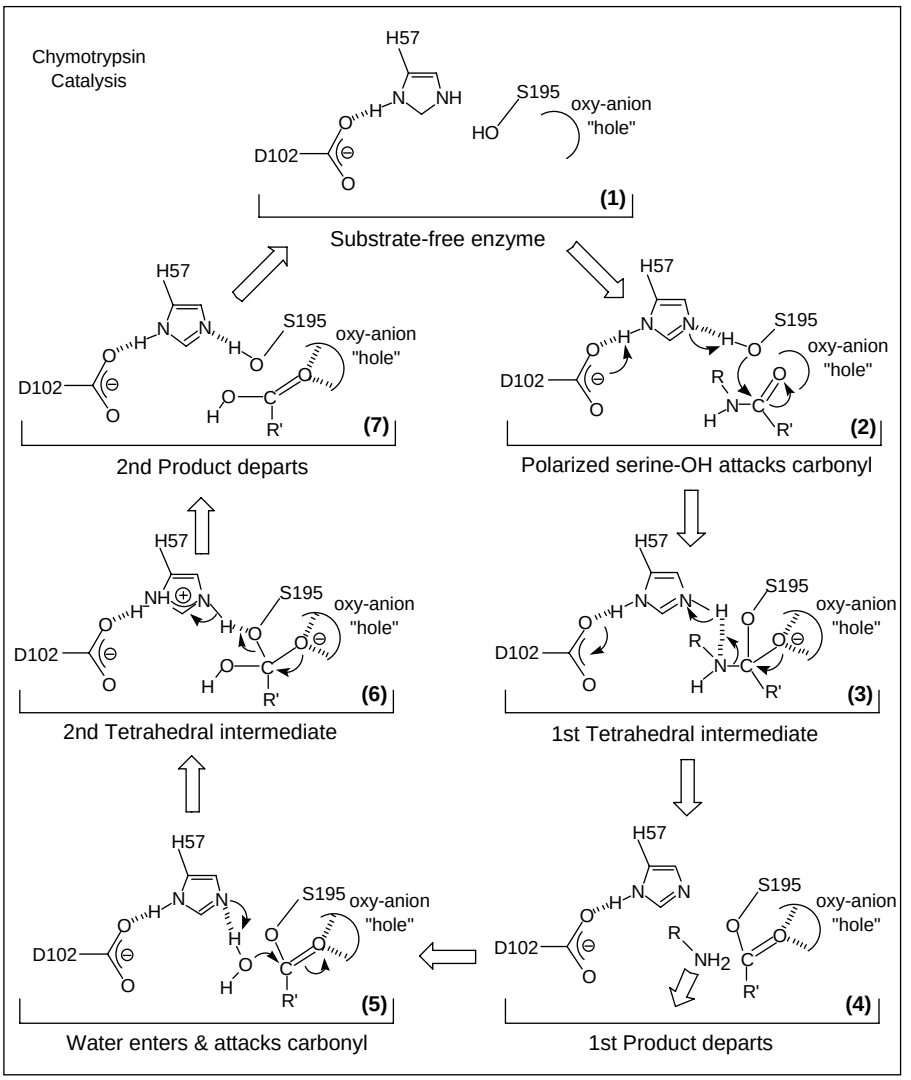

Fig. 2 Catalytic mechanism of the proteolytic enzyme chymotrypsin. The substrate-free enzyme (**1**) with its triad of aspartate, histidine, and serine residues, as well as a diagrammatic representation of the positively charged residues making up the ''oxy-anion hole''. The triad is thought to facilitate nucleophilic attack of the serine hydroxyl group, as shown in (**2**), thereby forming a negatively charged tetrahedral intermediate (**3**) that rearrange to the *O*-acyl-enzyme (**4**). The reaction cycle proceeds after the amine-containing product departs and is replaced by a water molecule (**5**). Nucleophilic attack by water then forms the second tetrahedral intermediate (**6**), which rearranges to form the second product-bound (**7**). The same mechanism applies to ''serine'' protease/esterase superfamily.

in the visible and/or ultraviolet spectrum upon peptide bond cleavage). Virtually nothing is known about chymotrypsin's action on natural protein substrates.

Whenever it becomes feasible to explain the nature of biological catalysis, biochemists can also begin to understand the structural, dynamic, and catalytic features responsible for effective regulation. Activators and inhibitors of enzymes have the effect of respectively lowering and raising reaction barrier(s), as do the activating and inhibitory effects of posttranslational covalent modifications of enzymes. Another benefit of developing detailed, well-reasoned models for enzyme catalysis is the improved prospect for designing highly effective inhibitors that may be employed as therapeutic agents. Not surprisingly, pharmaceutical companies continually make huge investments of time and money in their investigative efforts to gain such valuable information about key enzyme reactions. An excellent example is angiotensinogen-converting enzyme (ACE), for which well over a million compounds have been synthesized and evaluated in terms of their potential to become reliable blood pressure–regulating drugs.

Finally, a long-standing aspiration in chemistry is the creation of synthetic molecules that attain the stereochemical selectiveness, if not the catalytic proficiency, of naturally occurring enzymes. The rationale for this approach is to circumvent the limited stability of proteins, especially in nonaqueous solutions. One approach for obtaining novel catalysts for specific organochemical reactions involves the creation of catalytic antibodies (or "abzymes"). The idea that enzymes promote catalysis by stabilizing reaction transition states led Jencks in 1969 to offer the following prediction that appropriately selected antibodies would prove to be effective catalysts. "If complementarity between the active site and the transition state contributes significantly to enzymatic catalysis, it should be possible to synthesize an enzyme by constructing such an active site. One way to do this is to prepare an antibody to a antigenic group which resembles the transition state of a given reaction. The combining sites of such antibodies should be complementary to the transition state and should cause an acceleration by forcing bound substrates to resemble the transition state." Because transition states are intrinsically unstable, catalytic antibodies are selected by using chemically stable transition-state analogs as immunigens. For example, antibodies generated against a bent porphyrin ring were found to catalyze the metal ion insertion into heme groups, presumably by straining the planar substrate toward a bent transition-state conformation. Extensive investigations, chiefly in the laboratories of Lerner and Benkovic, demonstrated that the observed rate enhancements correlated with an antibody's affinity for analog versus that for the reactant. While many details of antibody catalysis remain to be elucidated, it is safe to assert that antibodies are unlikely to approach the efficiency of enzymes, at least not in the foreseeable future. Beyond stabilizing one or more transition states in a catalytic reaction cycle, synthetic enzymes must also undergo an intricate series of energetically and temporally controlled conformational changes, which permit naturally occurring enzymes the capacity to control their acid/base, nucleophilic, electrophilic, and/or redox chemistry. Achieving the skill and insight to "teach" synthetic catalysts the finely honed art of catalysis remains a stubbornly elusive goal in modern enzyme science.

7
Basic Enzyme Kinetics

The Michaelis–Menten treatment of enzyme-catalyzed reactions is based on the following key assumptions: (1) the enzyme combines rapidly and reversibly with the substrate to form a single E·X complex, such that the Michaelis constant K_S equals $\{([E_{total}] - [E{\cdot}X]) \times ([S_{total}] - [E{\cdot}X])\}/[E{\cdot}X]$; (2) the total enzyme concentration is negligible relative to the total substrate concentration (i.e. $[E_{total}] \ll [S_{total}]$), thereby simplifying the earlier expression to: $K_S = \{([E_{total}] - [E{\cdot}X]) \times [S_{total}]\}/[E{\cdot}X]$; (3) the reaction rate (i.e. $v = k_{cat}\ [E{\cdot}X]$) is constant during the initial reaction phase, when substrate consumption and product accumulation is minimal; and (4) the concentration of active enzyme remains unchanged. The resulting Michaelis–Menten rate equation is:

$$v = V_{max}[S]/\{K_S + [S]\}$$

where V_{max} is the maximal rate (units = Molarity per second). [Note: $V_{max}/[E_{total}]$ equals the catalytic constant k_{cat} (units = second^{-1}), a measure of the frequency of catalytic cycling.] In more rigorous steady state treatments, particularly those accounting for isomerizations among many enzyme-bound species (e.g. $E + S = E{\cdot}S_1 = E{\cdot}S_2 = \cdots E{\cdot}S_n \cdots = (E{\cdot}X)^{\ddagger} = E{\cdot}P_n \cdots = E{\cdot}P_2 = E{\cdot}P_1 = E{+}P$), the Michaelis constant K_M (units = Molarity) is more complicated than the thermodynamic parameter K_S. For this reason, the Michaelis constant should not be regarded as an accurate measure of the dissociation constant for the E·S complex.

Most enzyme-catalyzed reactions actually utilize more than one substrate, especially when ATP, NAD^+, Coenzyme A, and similar cosubstrates are required. The kinetic behavior of multisubstrate enzyme systems depends on the order in which substrates bind and products are released. In hydrolase-type reactions, water is the second substrate; however, its concentration (55.5 M) is so high that, for kinetic purposes, water cannot be treated as a variable.

8
Enzyme Regulation

Homeostasis arises from the ability of enzymes, membrane-bound and mobile receptors, as well as key regulatory proteins to sense changes in the concentration of substrates, coenzymes, activators, as well as inhibitors, to integrate these input signals and to make appropriate responses that modify metabolic activity. Of particular significance are the enzymes catalyzing the first committed reactions within a pathway, and often play a pacemaking role in controlling metabolic flux (or throughput). Other enzymes within a pathway may simply adjust their individual catalytic properties in response to changes in the availability of pathway intermediates. Because most pathways are interconnected at branch points (or nodes), enzymes catalyzing first steps of pathways must be regulated in a hierarchical manner. Fast signaling occurs on the millisecond-to-second timescale and requires rapid binding and release of regulatory molecules to and from specific binding sites on target enzymes. On the second-to-minute timescale, cellular regulation is achieved by hormone and/or effector controlled enzymes catalyzing posttranslational modifications (e.g. phosphorylation, methylation, nucleotidylation, ADPRibosylation, acetylation, proteolysis, etc.) of

target enzymes, thereby altering the latter's catalytic performance. Subsequent removal of the transferred group by certain phosphatases, phosphodiesterases, glycosylases, and hydrolases restores the target enzyme's original activity state. Regulatory processes taking place on the minutes-to-days timescale often reflect changes in the rates of biosynthesis or degradation of regulated enzymes. Some enzymes have short biological half-lives (e.g. 12 min for ornithine decarboxylase), whereas others, like lactate dehydrogenase, are highly stable and persist for weeks or longer.

Feedback inhibition allows enzymes catalyzing committed steps to be inhibited as pathway end products eventually accumulate. Feedback inhibitors most often bind at discrete regulatory sites that are topologically remote from the active site, and they possess the ability to modify catalytic activity by altering the properties of active-site residues. Inhibitors have a net effect of raising the height of energy barriers limiting the rate of substrate binding and/or catalysis; activators have the opposite effect. It should be stressed that feedback inhibition at pathway branch points presents the opportunity for highly coordinated metabolic regulation of highly branched metabolic networks.

The most common mode of enzyme regulation results from spontaneous or ligand-induced conformational changes. In the case of spontaneous changes, a protein may exist in two states X and Y that are in thermodynamic equilibrium with each other. The equilibrium concentrations of X and Y are determined by the relative stabilities of each conformation. In other cases, a ligand molecule selectively alters the stability of one conformational state of the enzyme. In either case, structural alterations may arise from short-range effects (i.e. at distances much less than a nanometer) or they can occur from long-range changes at the quaternary structural level (i.e. through rearrangements in subunit–subunit interactions). The term *allostery* is derived from the Greek prefix *allo* (meaning "another") and the suffix "steros" (meaning "shape" or "form"). Allosteric interactions are attended by changes in substrate binding affinity and/or catalytic efficiency, by raising or lowering thermodynamic barriers for substrate binding and/or catalysis. Allosteric activators have the net effect of increasing an enzyme's affinity for its substrate(s) and/or its catalytic performance; allosteric inhibitors have the opposite effect.

Unlike those enzymes lacking regulatory mechanism, almost every allosteric enzyme possesses multiple, mutually interactive subunits, each with its own active site and one or more sites for activators and/or inhibitors. Cooperativity is the phenomenon wherein effector molecule binding on one subunit of an oligomeric protein results in local structural changes that are communicated to and from sites on other subunits. In the Monod–Wyman–Changeux concerted transition model, an oligomeric protein undergoes an all-or-none transition in response to changes in the concentration of effector molecules; alternatively, in the Adair–Koshland model, the conformational changes occur sequentially, again in response to changes in regulatory molecule concentration. Such binding interactions alter catalytic activity, often over narrow ranges of substrate and/or effector concentrations. The need for such responsiveness becomes clear where one realizes that most enzymes are not particularly responsive to changes in the

concentration of input signals. Enzymes obeying the Michaelis–Menten kinetics (see preceding text) require large changes in substrate concentration, whereas the activity of highly cooperative enzymes responds over a more narrow concentration range.

9
Inborn Errors in Enzyme Function

Genetic mechanisms minimize the occurrence of mutant enzymes that improperly catalyze a reaction or inappropriately respond to regulatory cues. Such renegade enzymes can wreak havoc on cells, tissues, and organs. Biomedical research has uncovered countless instances wherein improper catalysis/control of even a single enzyme can greatly distress a living organism. In fact, many animal and plant diseases arise from point mutations altering enzyme efficiency through an incorrect substitution of a single amino acid. Other injurious mutations result in over/underproduction of enzymes, defective regulatory interactions, impaired stability, incorrect posttranslational modification, improper subcellular targeting, and so on. Diseases are linked to the failure of specialized enzymes designed to remediate any DNA damage arising from photolysis, oxidation, hydrolysis, racemization, and other molecular rearrangements. A notable example of an enzyme-associated pathophysiology is amyotrophic lateral sclerosis (ALS) (also known as *Lou Gehrig's disease*), a terrible neurodegenerative disorder linked to impaired action of superoxide dismutase. This enzyme plays a powerful role in mitigating oxidative stress by catalyzing the conversion of two superoxide ($O_2^{\bullet -}$) ions to produce dioxygen and the less toxic substance, hydrogen peroxide. Excess superoxide destroys essential cellular components, especially in neurons where oxidative stress has profound pathological sequelae. Another example, taken from research on Alzheimer's disease, is the very finding that enzyme-catalyzed prolyl isomerization can regulate folding of the cytoskeletal protein Tau in axons, such that enzyme knockouts in mice bring about progressive age-dependent neuropathy characterized by motor and behavioral deficits, tau hyperphosphorylation, paired helical filament formation, and neurodegeneration. Finally, the devastating effects observed in primary oxalemia have been traced to the mistargeting of an oxalate-degrading enzyme from peroxisomes to mitochondria.

See also Heme Enzymes; Metalloenzymes; Organic Cofactors as Coenzymes; Ribozymes; Telomerase.

Bibliography

Books and Reviews

Abeles, R.H., Frey, P.A., Jencks, W.P. (1992) *Biochemistry*, Jones & Bartlett, Boston, MA, p. 838.

Fersht, A.R. (1998) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W. H. Freeman, New York, p. 650.

Haldane, J.B.S. (1930) *Enzymes*, Longmans-Green, London.

Hammes, G.G. (2002) Multiple conformational changes in enzyme catalysis, *Biochemistry* **41**, 8221.

Jencks, W.P. (1969) *Catalysis in Chemistry and Enzymology*, McGraw-Hill, San Francisco, CA.

Lerner, R.A., Benkovic, S.J. (1988) Principles of antibody catalysis, *Bioessays* **9**, 107–112.

Purich, D.L. (2001) Enzyme catalysis: a new definition accounting for noncovalent substrate- and product-like states, *Trends Biochem. Sci.* **26**, 417.

Scriver, C.R., Sly, W.S., Childs, B., Beaudet, A.L., Valle, D., Kinzler, K.W., Vogelstein, B. (2000) *The Metabolic & Molecular Basis of Inherited Disease*, 7th edition, McGraw-Hill, New York.

Voet, D., Voet, J.G. (2003) *Biochemistry Volume 1 Biomolecules, Mechanisms of Enzyme Action, and Metabolism* 3rd edition, J. Wiley, New York, p. 1178.

Wolfenden, R. (2001) The depth of chemical time and the power of enzymes as catalysts, *Accts Chem. Res.* **34**, 938–945.

Protein Circular Dichroism: *see* Circular Dichroism in Protein Analysis

Protein Complexes, Supramolecular: *see* Supermolecular Protein Complexes

Protein-DNA Interactions: *see* DNA–Protein Interactions

Protein Expression by Expansion of the Genetic Code

Jason W. Chin[1] *and Thomas J. Magliery*[2]
[1] *MRC Laboratory of Molecular Biology, Cambridge, UK*
[2] *Yale University, New Haven, CT, USA*

1 Expanding the Genetic Code 47
1.1 The Chemical and Biosynthetic Incorporation of Unnatural Amino Acids into Proteins 47
1.2 A General Strategy for Expanding the Genetic Code of Organisms 48
1.3 Unique Codons 49

2 Expanding the Genetic Code of Bacteria 50
2.1 Orthogonal Aminoacyl-tRNA-Synthetase/tRNA Pairs for Bacteria 50
2.2 Methods for Evolving Synthetases with Unnatural Amino Acid Specificity in *E. coli* 52

3 Expanding the Genetic Code of Eukaryotic Cells 58
3.1 Stoichiometric Aminoacylation Methods 58
3.2 Catalytic Aminoacylation Methods 58
3.3 Orthogonal Aminoacyl-tRNA Synthetase/tRNA Pairs in Eukaryotic Cells 58
3.4 Methods for Evolving Synthetases with Unnatural Amino Acid Specificity in Eukaryotic Cells 60

4 Using Expanded Genetic Codes 62
4.1 The Utility of Expanded Genetic Codes 62
4.2 Biophysical Probes 63
4.3 Unique Reactive Handles 64
4.4 Posttranslational Modifications 64
4.5 Organism Evolution 64

5 Summary 65

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

Acknowledgment 65

Bibliography 66
Books and Reviews 66
Primary Literature 66

Keywords

Amber
The name given to the UAG stop codon, which is the least used of the three stop codons in *E. coli* and yeast.

Amber-suppressor tRNA
A tRNA that is capable of decoding the *amber* stop codon, UAG, is said to "suppress" *amber* nonsense mutations.

Aminoacyl-tRNA Synthetase (aaRS)
An enzyme capable of esterifying an amino acid onto a tRNA. These enzymes activate free amino acids to adenylates using ATP, and some actively "edit" misacylated tRNAs by hydrolysis of the errant amino acid.

Codon
The unit of mRNA that codes for an amino acid. Natural codons are three consecutive ribonucleotides (G, A, U and C); there are therefore 64 natural codons of which three (UAG, UGA, and UAA) code for termination of protein synthesis (i.e. "stop"). Codons are read by complementarity to anticodons on tRNAs at the ribosome.

Identity Element
The portions of tRNAs responsible for conferring specificity for a cognate aminoacyl-tRNA synthetase.

Library
A collection of related molecules that vary in a specified way, typically synthesized in a highly parallel fashion. For example, a library might consist of a large collection of mutants of a specific protein, such as an aminoacyl-tRNA synthetase.

Orthogonal Pairs
Aminoacyl-tRNA synthetases and their tRNA substrates that do not interact with the tRNAs and synthetases of the host cell in which they are expressed.

Selection
A procedure for sorting cells on the basis of their ability to survive under specified conditions. Selections can be used to sort libraries of enzymes expressed in cells,

if the survival of the organism requires the activity of the enzyme. In contrast, a screen is a procedure for sorting cells on the basis of phenotype, such as fluorescence.

tRNA
Small RNAs, typically about 75 nt ending with the sequence CCA-3′, which deliver amino acids for protein synthesis at the ribosome in response to the mRNA codons to which they correspond. tRNAs are acylated with amino acid by aminoacyl-tRNA synthetases.

Unnatural Amino Acids
Protein monomer building blocks not inserted by the biosynthetic machinery of living organisms. There are 20 natural amino acids normally incorporated into proteins.

The Genetic Code consists of 64 three-base codons that code for 20 amino acids and three stop signals. Recent engineering of the cell's protein synthesis machinery has allowed unnatural amino acids to be encoded in the genetic code of prokaryotic and eukaryotic organisms. Importantly, the approaches discussed add unnatural amino acids to the genetic code, without detracting from the cell's ability to incorporate the common 20 amino acids into proteins. Here we discuss the intellectual and practical strategies that have been employed in the challenging, but ultimately successful expansion of the genetic code. We also discuss the emerging applications of expanded genetic codes in solving basic biological and biophysical problems and in creating a new synthetic biology.

1 Expanding the Genetic Code

1.1 The Chemical and Biosynthetic Incorporation of Unnatural Amino Acids into Proteins

One of the remarkable chemical mysteries of life is that molecules composed of just 20 amino acid monomers carry out the entire repertoire of protein function – which spans structural roles, metabolism, signal transduction, immunity, replication and degradation, to name just a few. The chemical repertoire of proteins is naturally enlarged by the occasional insertion of selenocysteine or pyrrolysine, and by the common addition of posttranslational modifications (including phosphorylation, sulfation, lipidation, biotinylation, glycosylation, hydroxylation, methylation and acetylation), particularly in higher organisms. Those taking a chemical approach to biology have long recognized that expanding and precisely controlling the functionalities that can be incorporated into proteins would allow the creation of powerful tools for enzymatic, biophysical, and cellular studies. Despite its obvious utility, achieving the facile incorporation of "unnatural" amino acids has been and continues to be a difficult problem.

A number of methods have been developed for the insertion of unnatural amino acids into proteins, from synthetic to biosynthetic. Solid-phase peptide synthesis combined with enzymic or chemoselective ligation now permits the *in vitro* synthesis of small proteins with total control over the amino acid composition. Larger proteins are accessible *in vitro* using semisynthetic methods such as expressed protein ligation, wherein a synthetic peptide is chemoselectively ligated to a recombinant protein fragment. Biosynthetic approaches give access to essentially any size protein, but have their own limitations. For example, addition of near-homologs of natural amino acids (such as selenomethionine) to the growth medium sometimes results in widespread replacement of a natural amino acid, but, while useful, this does not actually expand the genetic code. Site-specific biosynthetic expansion of the genetic code has been achieved by chemical esterification of a tRNA with an unnatural amino acid, both in cell-free translation systems and with microinjection or electroporation of cells. Here, the rarest stop codon is used as a "blank" in the genetic code, read by the esterified tRNA. While this is a powerful approach, its use has been limited by the heroic synthetic effort required. Recent ribozyme-based approaches for acylation of tRNAs with unnatural amino acids may expand the scope of this method.

A quantum leap forward has occurred in recent years in this field – the first organisms have been engineered that site specifically utilize a twenty-first amino acid monomer. One advantage to this approach is that the researcher need only transform an expression vector for the protein of interest into a strain with the appropriate biosynthetic machinery, and addition of unnatural amino acid to the growth medium results in protein containing unnatural amino acid. The ease of use finally makes available proteins with unique reactive handles, homogeneous posttranslational modifications, and intrinsic biophysical probes, to the experimentalists such as enzymologists and cell biologists who can make the best use of them. Here we describe the engineering of prokaryotic and, more recently, eukaryotic organisms with expanded genetic codes, highlighting uses, challenges and future directions.

1.2 A General Strategy for Expanding the Genetic Code of Organisms

Expanding the genetic code of an organism involves the addition of an unnatural amino acid to the organism's genetic repertoire without loss of the ability to insert any of the 20 natural amino acids into proteins. An organism with an expanded genetic code will incorporate unnatural amino acids at specific sites in a protein in response to a unique codon, through a unique enzyme (an aminoacyl-tRNA synthetase) capable of acylating a tRNA for delivery of the unnatural amino acid in response to that codon (see Fig. 1). Ideally, the amino acid will be incorporated with a fidelity rivaling that for natural amino acid–incorporation. To achieve an organism with an expanded genetic code, the following are required:

1. A host organism: Owing to the extensive engineering required to expand the genetic code, well-understood systems like *Escherichia coli* and yeast have proven good starting points.
2. An insertion signal (codon): The *amber* stop codon is a good first choice, since it is rarely used in *E. coli* and yeast, and

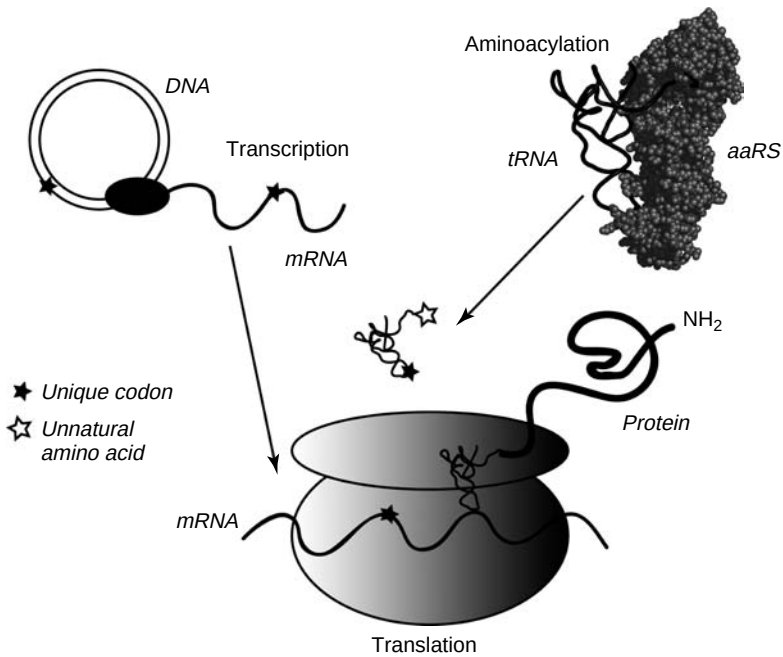

Fig. 1 Scheme of protein biosynthesis: Expansion of the genetic code first requires a unique codon (e.g. *amber* stop codon or 4-base codon), which is specified in the DNA and transcribed into the mRNA. A unique aminoacyl-tRNA synthetase (aaRS) must acylate a unique tRNA with the unnatural amino acid. This tRNA then delivers the unnatural amino acid to the ribosome in response to the unique codon, incorporating it into protein site specifically.

since robust "suppressors" of *amber* mutations are known. Expansion of the genetic code with multiple distinct unnatural amino acids is possible using extended insertion signals such as four-base codons.

3. An "orthogonal" tRNA corresponding to the insertion signal that is neither acylated nor deacylated ("edited") by any endogenous aminoacyl-tRNA synthetase (aaRS).
4. An orthogonal aaRS, that acylates only the orthogonal tRNA, and only with an unnatural amino acid. For simplicity, an aaRS without hydrolytic editing is a good choice.
5. An unnatural amino acid that is neither toxic nor incorporated into proteins by endogenous aaRSs, but that is readily taken up by the cell (or produced there).

1.3 Unique Codons

The genetic code consists of 64 three-base codons that code for 20 amino acids and three stop signals. Since only one stop codon is needed for translational termination, the other two can in principle be used to encode unnatural amino acids. The *amber* stop codon (UAG) has been extensively used in the *in vitro* and *in vivo* biosynthesis of unnatural amino acid–containing proteins. This codon is also the least used stop codon in both *E. coli* and in yeast, and several laboratory *E. coli* strains

contain suppressors that insert a natural amino acid in response to UAG without significantly affecting growth rates. Suppressors have also been discovered for *amber* codons in mammalian cells and in higher organisms, such as yeast, worms and mice. It should therefore be possible to incorporate unnatural amino acids in response to an *amber* codon in a range of host organisms from prokaryotes to eukaryotes without significant perturbation of the host. In fact, orthogonal pairs that decode *amber* codons and insert unnatural amino acids have been developed for *E. coli*, yeast, and mammalian cells.

An alternative and potentially complementary set of codons to the *amber* codon are four-base and other extended codons. Naturally occurring frameshift suppressors in yeast and *Salmonella* are known, typically functioning by decoding a 4-base codon using a tRNA with an extended anticodon loop (8 nt instead of 7 nt). Suppression of 4-base and 5-base codons has been demonstrated in *E. coli* and in *E. coli* cell-free translation systems. Variants of $tRNA^{Ser}$ with randomized, extended anticodon loops (8 or 9 nt instead of 7 nt) have been selected for the ability to suppress randomized four-base codons in the gene for β-lactamase. The selected variants included suppressors for the codons AGGA, UAGA, CCCU, and CUAG, with efficiencies comparable to *amber* suppression. A similar approach afforded fairly efficient 5-base codon suppressors of CUAGU, CUACU, and AGGAU.

In practice, it has been demonstrated that a four-base codon AGGA and the *amber* codon UAG can be used to direct the incorporation of two distinct unnatural amino acids into a single site in protein in *E. coli* with high fidelity and specificity. Moreover, simultaneous insertion of two unnatural amino acids has been accomplished in cell-free transcription/translation systems with two 4-base codons. The number of 4-base codons with reasonable suppression efficiency probably places a small, finite upper limit (perhaps 4) on the number of amino acids that can be inserted this way. An alternative approach is to develop "unnatural codons" that utilize a nonnatural base pair, and this is actively being pursued by several groups. The ability to generate a stable base pair that can faithfully be enzymatically incorporated into both DNA and RNA in living cells (which must make or uptake the monomers) is a difficult problem that is still far from a practical solution.

2 Expanding the Genetic Code of Bacteria

2.1 Orthogonal Aminoacyl-tRNA-Synthetase/tRNA Pairs for Bacteria

The first stage in expanding the genetic code of an organism is to generate an orthogonal tRNA/aaRS pair, and second, to change the specificity of the aaRS with respect to amino acid. (Hereafter, the notation $Oo\text{tRNA}^{Aaa}$(NNN) for tRNAs and *Oo*AaaRS for aminoacyl-tRNA synthetases shall be used, where *Oo* denotes the organism, Aaa denotes the amino acid, and NNN denotes the anticodon.) All initial work was carried out in *E. coli* using *amber*-suppressing tRNAs. The Schultz laboratory first employed an engineering approach to generate an orthogonal tRNA/synthetase pair based on the exceedingly well-characterized *E. coli* glutaminyl-tRNA synthetase/$tRNA^{Gln}$(CUA) pair. (GlnRS naturally acylates an efficient *amber*

suppressor in *E. coli*.) Despite a remarkable change in specificity toward the orthogonal tRNA, the engineered mutant GlnRS was still capable of efficiently acylating the wild-type $tRNA^{Gln}$ as well. If the amino acid specificity of this enzyme had been subsequently altered, the unnatural amino acid would have been inserted at Gln codons throughout the proteome, in addition to *amber* codons.

Selection against aminoacylation of endogenous wild-type tRNAs might lead to aaRSs with specificity for the orthogonal tRNA. RajBhandary and coworkers pursued such a strategy with a mutant *E. coli* initiator tRNA that is not aminoacylated by endogenous synthetases and the tyrosyl-tRNA synthetase from *Saccharomyces cerevisiae*. The wild-type *Sc*TyrRS also aminoacylates *E. coli* $tRNA^{Pro}$, and is therefore toxic in *E. coli*. However, when a library of *Sc*TyrRS mutants was screened simultaneously for a lack of toxicity and the ability to suppress *amber* mutations in the gene for chloramphenicol acetyltransferase (which confers resistance to the antibiotic chloramphenicol), nontoxic variants of TyrRS were isolated, which discriminated 15-fold against *E. coli* $tRNA^{Pro}$.

A second strategy for generating new orthogonal pairs involves importing a heterologous *amber*-suppressor aaRS/tRNA pair from another organism. This approach was, in part, driven by the observation that the identity elements in tRNAs for eukaryotic aaRSs often differ significantly from their prokaryotic counterparts. Many eukaryotic aminoacyl-tRNA synthetases therefore do not charge *E. coli* tRNAs and should be orthogonal in *E. coli*. In particular, several yeast synthetases (notably glutamine, aspartic acid, arginine, and tyrosine) do not aminoacylate total *E. coli* tRNA, and might be useful for the construction of orthogonal tRNA-synthetase pairs.

On the basis of the observation that *S. cerevisiae* $tRNA^{Gln}$ (*Sc*$tRNA^{Gln}$) is not acylated by *E. coli* GlnRS (*Ec*GlnRS), Liu and Schultz demonstrated that a modified *Sc*$tRNA^{Gln}$(CUA) and *Sc*GlnRS constitute a functional orthogonal pair in *E. coli*. Despite nearly 10 man-years of work engineering this pair, it was not possible to alter the amino acid specificity of *Sc*GlnRS, even though much of the selection and library construction technology was worked out on this system.

Archaeal tRNA synthetases are more similar to their eukaryotic than prokaryotic counterparts in terms of homology and tRNA identity elements. However, unlike synthetases from eukaryotic cells, synthetases from archaea can be expressed in good yields in *E. coli* and can be readily purified in active form. Numerous archaeal genome sequences are currently available, which, together with the absence of introns in archeal genomes, facilitate the cloning of the archaeal synthetase genes. Moreover, early work on the halophile *Halobacterium cutirebrum* indicated that almost all its tRNAs (notably those for leucine, arginine, tyrosine, serine, histidine, and proline) cannot be charged by *E. coli* aminoacyl-tRNA synthetases, suggesting that these tRNAs are potentially orthogonal in *E. coli*. Consequently, attention has recently focused on the archaea, rather than on eukaryotes, as a source of orthogonal tRNA-synthetase pairs for *E. coli*.

It was shown that a $tRNA^{Tyr}$(CUA) and TyrRS from the archea *Methanococcus jannaschii* constitute an orthogonal pair in *E. coli*. However, while the *Mj*TyrRS is considerably more active than *Sc*GlnRS, the *Mj*$tRNA^{Tyr}$(CUA) was not "as orthogonal" as *Sc*$tRNA^{Gln}$(CUA); that is: the

*Mj*tRNA^{Tyr}(CUA) is detectably acylated by *E. coli* synthetases. A semirational design strategy was used to increase the orthogonality of this tRNA using a variant of the "double-sieve" selection for unnatural amino acids (see the following section). The resulting orthogonal tRNA/synthetase pair has been the most useful one so far in *E. coli*, due to the high activity of the synthetase and good orthogonality of the tRNA. The relatively large, hydrophobic amino acid binding pocket of TyrRS has allowed the incorporation of many useful amino acids. Recently, solution of the *Mj*TyrRS crystal structure permitted design of a mutant synthetase with better recognition of the *Mj*tRNA^{Tyr}(CUA), which may be useful in further work altering the amino acid specificity of the synthetase.

A number of other pairs have also been introduced into *E. coli* including yeast tRNA^{Asp}(CUA) and a point-mutant of yeast AspRS; derivatives of *Halobacterium* NRC-1 tRNA^{Leu} and *Methanobacterium thermoautotrophicum* LeuRS that result in suppression of *amber, opal* (UGA), and 4-base (AGGA) codons; and a consensus-derived archaeal tRNA^{Glu}(CUA) and *Pyrococcus horikoshii* GluRS. However, there has been no success in altering the amino acid specificity of any of these synthetases. Recently a mutant *Pyrococcus horikoshii* LysRS derivative has been selected for the incorporation of homoglutamine in *E. coli* in response to the four-base codon AGGA using a consensus-derived archaeal tRNA^{Lys}(UCCU). This is only the second orthogonal pair shown to be useful for unnatural amino acid incorporation in *E. coli*, and the fact that it allows four-base suppression means that it is now possible to simultaneously, site specifically introduce two unnatural amino acids into proteins.

2.2 Methods for Evolving Synthetases with Unnatural Amino Acid Specificity in *E. coli*

Once an orthogonal tRNA-synthetase pair has been developed for use in *E. coli*, the next step is the modification of the synthetase specificity such that it aminoacylates the orthogonal tRNA with a desired unnatural amino acid and no other endogenous amino acid.

This problem is complicated considerably by at least five factors. (1) The small amino acid substrates are contacted by a large number of residues concentrated in three-dimensional space but often spread over the primary structure of the synthetase. This makes it technically difficult to construct reasonable mutagenic libraries. (2) In spite of this, mutations far from the active site of aaRSs are known to affect aminoacylation kinetics. (3) Mutations to the synthetase can also affect recognition of tRNA, thereby altering the orthogonality of the enzyme. (4) It is extremely difficult to alter the specificity of an aaRS without reducing its aminoacylation activity, but a certain level of activity is required to support high-level protein synthesis. Thus, the engineering of synthetases with weak aminoacylation activity toward native substrates has proven to be exceedingly difficult. (5) Selecting for unnatural amino acid specificity requires that the survival of the organism be tied to the insertion of an unnatural amino acid, and no direct way to do this has been developed.

A small number of attempts have been made to alter the specificity of aaRSs by inspection or semiempirically, with varying results. For example, from genetic screens, it was known that the *Ec*PheRS Ala294 → Ser mutant resisted incorporation of *p*-F-Phe. Ibba and Hennecke showed that

the Ala294 → Gly mutant of this enzyme, with an expanded binding pocket, is capable of inserting *p*-Cl-Phe and *p*-Br-Phe, and the Tirrell group has shown that *para*-iodo, azido-, cyano-, and ethynyl-Phe are also accepted. Furter used the associated *p*-F-Phe resistant *E. coli* strain to engineer the first bacterium able to site-selectively insert an unnatural amino acid by introducing a *Sc*PheRS/*Sc*tRNA^{Phe}(CUA) pair that accepts *p*-F-Phe. However, the *Ec*PheRS(A294S) accepts *p*-F-Phe weakly,

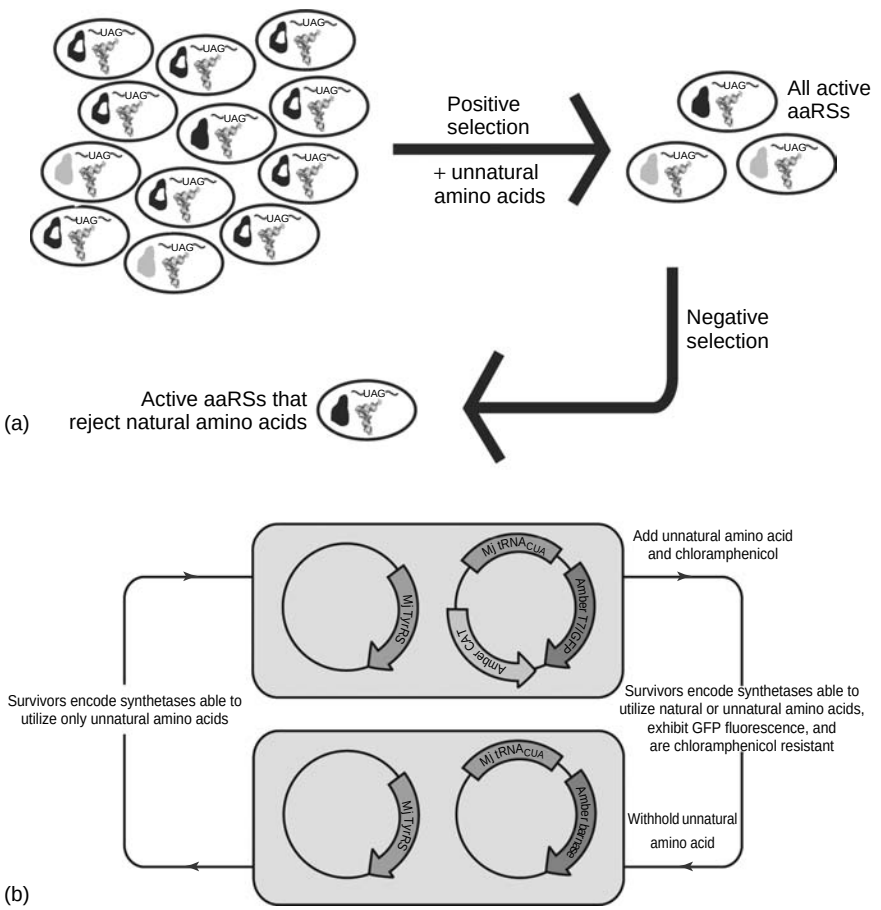

Fig. 2 ''Double-sieve'' selections for unnatural amino acid activity: (a) Schematic of a generic double-sieve selection. A library of orthogonal aminoacyl-tRNA synthetases is transformed into cells with an orthogonal tRNA. First, positive selection is done in the presence of unnatural amino acids, such that the selective phenotype requires acylation of the orthogonal tRNA and suppression of the unique codon. Cell with inactive synthetases (white) die, but those with synthetases active toward natural (grey) or unnatural (black) amino acids survive. Second, a negative selection is carried out in the absence of unnatural amino acid, where the selective phenotype requires that the tRNA *not* be acylated. Only cells with synthetases that lack activity toward natural amino acids survive. Thus, survival of both selections implies activity only toward unnatural amino acids. (b): A double-sieve selection in *E. coli*. (c): A double-sieve selection in yeast. (See color plate p. xxvi.)

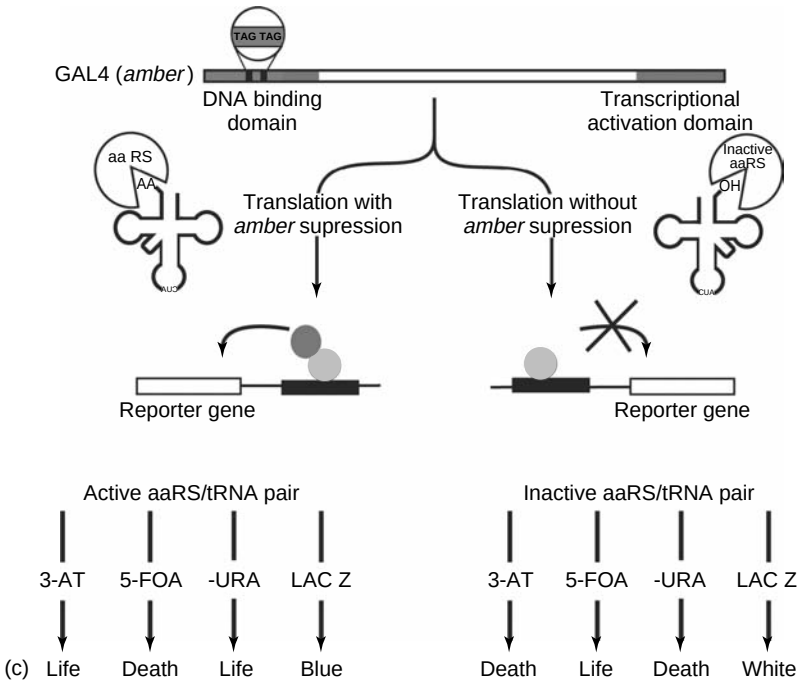

Fig. 2 (*Continued*)

and the *Sc*tRNA^{Phe}(CUA) is also acylated with Phe and Lys (presumably by *Ec*LyrsRS, in the latter case). Kiga et al. converted *Ec*TyrRS into an enzyme capable of accepting 3-iodotyrosine (3-I-Tyr) as a substrate by selecting three sites for mutagenesis based on structure, and examining the aminoacylation properties of 50 mutants *in vitro*. Interestingly, while >95% of *amber*-encoded sites were occupied by 3-I-Tyr in a wheat-germ extract translation reaction, Tyr-containing protein was produced if 3-I-Tyr was left out of the reaction mixture. This is potentially a problem when moving this technology into cells, since the concentration of unnatural amino acid cannot be controlled arbitrarily. A number of attempts have been made to change GlnRS into GluRS, but the resulting enzymes still prefer Gln (although by a considerably smaller amount than wild-type GlnRS). Additionally, some attempts to computationally predict mutations to alter amino acid specificity have been reported.

By far the most successful and general method for altering the specificity of aminoacyl-tRNA synthetases has been directed combinatorial mutagenesis of the amino acid binding pocket of synthetase enzymes coupled to a general "double-sieve" selection (see Fig. 2). Several general methods have been developed to alter the specificity of orthogonal aminoacyl-tRNA synthetase/tRNA(CUA) pairs that function in *E. coli* so that they use only a desired unnatural amino acid.

The selection of a library of aaRS variants was first carried out by a two-step procedure. In the first step, the library of synthetases is selected in bacteria bearing tRNA(CUA) and an *amber* mutant

of β-lactamase, with ampicillin and unnatural amino acid supplementation of the medium. Use of a permissive site in β-lactamase ensures that survivors of the selection contain aaRSs that are capable of acylating the tRNA(CUA) – thus producing β-lactamase and conferring resistance to ampicillin – but which amino acid was esterified onto the tRNA is unknown. The synthetases from the survivors of this positive selection are then expressed in bacteria bearing the tRNA(CUA) and a barnase gene containing multiple amber codons (wild-type barnase is a lethal ribonuclease in *E. coli*). No unnatural amino acids are added to the medium, so survival at this stage ensures that no cellular amino acids (proteinogenic or otherwise) are a substrate for the mutant synthetases. Clones that survive *both* steps of the selection should therefore be active *only* toward the unnatural amino acid. The initial implementation of this strategy (with the *Sc*GlnRS/*Sc*tRNA^{Gln}(CUA) orthogonal pair, a large library of mostly commercially available amino acids, and synthetase libraries created by random mutation) resulted in no useful synthetases.

Two critical technical improvements to the methodology were (1) the replacement of the β-lactamase positive selection with a chloramphenicol acetyltransferase-based selection, and (2) use of a directed, semirational library-construction method. Chloramphenicol is bacteriostatic and acts cytosolically, allowing a broad range of chloramphenicol concentration to be suitable for selection. Random mutagenesis of wild-type GlnRS had two negative consequences. First, most of the library members were very active toward Gln, requiring exceptional performance from the negative selection. Second, the probability of mutation in the proximity of the amino acid substrate was low. Instead, guided by the ternary X-ray crystal structure of *Ec*GlnRS, tRNA^{Gln}, and a Gln-AMP analog, 5–10 residues proximal to the substrate were first mutated to Ala and then randomized to all 20 amino acids. The resulting libraries contained few synthetases with activity toward Gln, allowing one to forego the negative selection in early rounds, and were almost certain to have an altered amino acid binding pocket.

Unfortunately, libraries designed for use with carboxamide N-alkylated Gln analogs, analogs elaborated from the γ-carbon, and the α-hydroxy acid analog of Gln were not found to contain any synthetases capable of activating the unnatural amino acids at useful levels. However, using the very same library construction and selection methodology, it has been possible to alter the specificity of *Mj*TyrRS and *Ph*KRS, allowing the delivery of several useful unnatural amino acids (see section 4 and Fig. 4). Santoro et al. extended the selection technology by developing a fluorescence-based screen for *amber* suppression. An *amber* mutant of T7 RNA polymerase was used to drive the transcription of green fluorescent protein (GFP). In combination with fluorescence-activated cell sorting (FACS), this system can be used as both a positive screen for *amber* suppression (where one sorts for fluorescent cells in the presence of unnatural amino acids) and a negative screen against *amber* suppression (sorting for dim cells in the absence of unnatural amino acids). A multivalent system, where chloramphenicol-based selection is possible in addition to screening, has been especially effective.

The fidelity of incorporation of each amino acid was established by several methods. First the clones were characterized by their ability to confer chloramphenicol resistance on cells transformed

with the CAT(*amber*) reporter in the presence and absence of a particular unnatural amino acid. In all successful cases, the IC_{50} values with respect to chloramphenicol were large (100 μg mL^{-1}) in the presence of unnatural amino acid, and were generally less than 10 μg mL^{-1} in the absence of the unnatural amino acid. Similarly, cellular fluorescence can be measured with the T7/GFP system in the presence and absence of unnatural amino acid.

Expression of *amber* mutants of DHFR, myoglobin, and the Z-domain of protein A was also directly assessed. In the presence of unnatural amino acid, synthetase, and tRNA, milligrams of each protein could be purified from a liter of liquid culture. Yields were typically 25 to 30% of wild-type protein. In the absence of any one component synthetase, tRNA, or amino acid, no protein could be observed by Western blot, silver stain, or Coomassie stain. Proteins were subjected to mass analysis of intact protein, and tryptic digests confirmed that the unnatural amino acid and only this amino acid is incorporated in response to the UAG codon and at no other site in the protein.

It is likely that the high activity of the *Mj*TyrRS and the excellent orthogonality of the synthetase and tRNA account for the malleability of this enzyme's substrate specificity. Though all of the libraries of *Mj*TyrRS were designed using the *Bacillus stearothermophilus* structure, the structure of the *M. jannaschii* protein has recently been solved. There are a number of differences in the tyrosine binding pocket that will improve further library design with the enzyme, and may even suggest ways to improve some of the existing mutant synthetases (see Fig. 3).

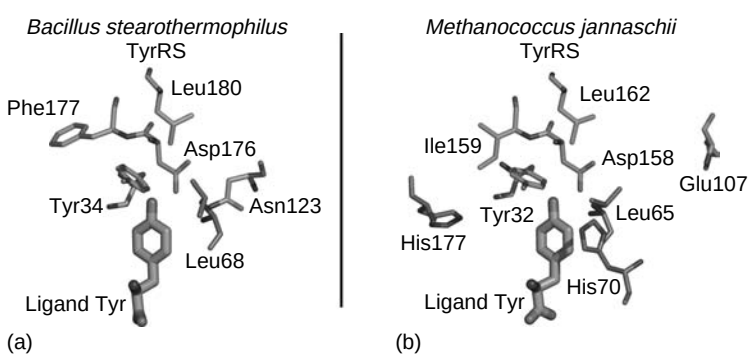

Fig. 3 The active site of TyrRS: Amino acids near the substrate tyrosine of (a) *Bacillus stearothermophilus* and (b) *Manococcus jannaschii* TyrRS. The Bacillus structure is a good model for the *Ec*TyrRS used to expand the genetic code of yeast. All the *Ec*TyrRS variants for the insertion of unnatural amino acids were derived from variants of Tyr37 (Tyr34), Asn126 (Asn123), Asp182 (Asp176), Phe183 (Phe177) and Leu186 (Leu180), where the corresponding amino acids in *Bs*TyrRS are noted parenthetically. Nearly all of the *Mj*TyrRS derivatives used for insertion of unnatural amino acids in *Escherichia coli* were generated by randomization of the Tyr32 (Tyr34), Glu107 (Asn123), Asp158 (Asp176), Ile159 (Phe177), and Leu162 (Leu180). These amino acids were selected from the *Bs*TyrRS structure, which was available at the time, but the recent *Mj*TyrRS structure suggests some possible improvements. Rendered using PyMOL (Warren Delano, 2002, http://www.pymol.org) from PDB entries 4TS1 and 1J1U.

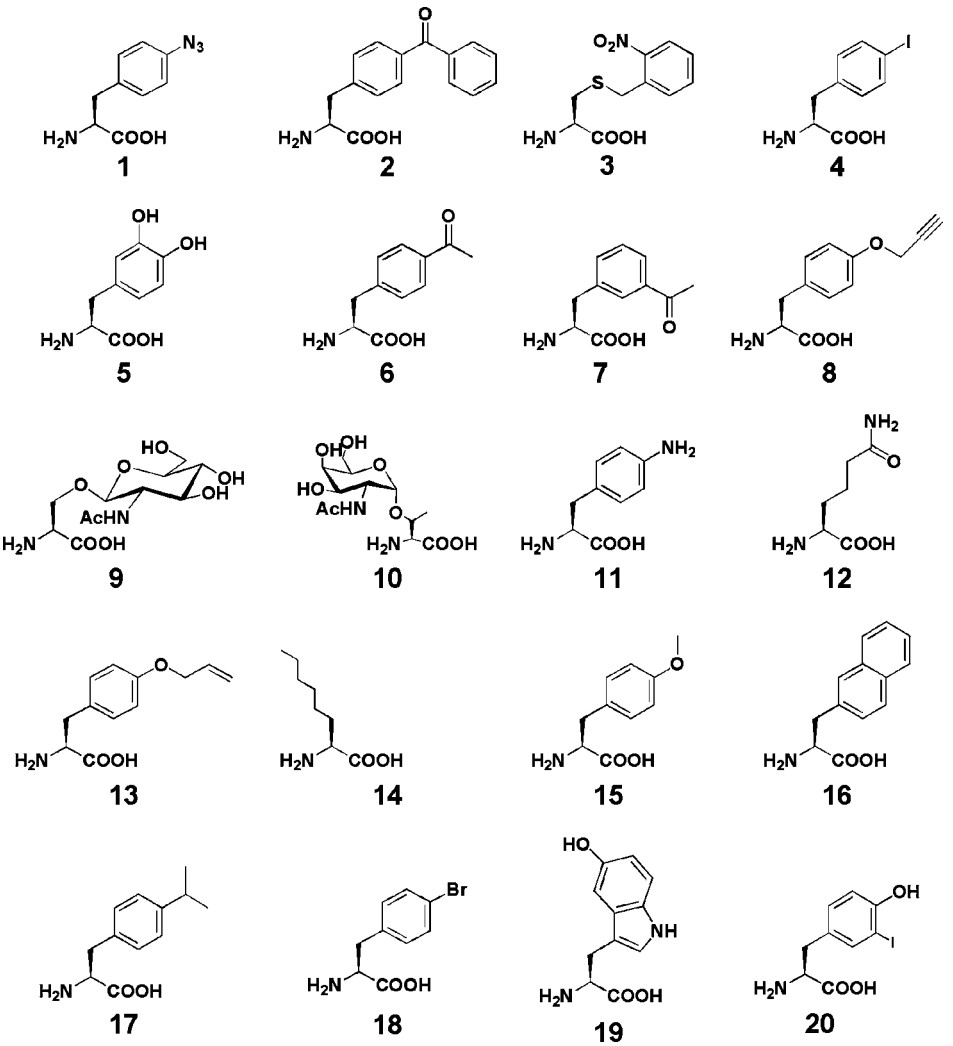

Fig. 4 Unnatural amino acids incorporated site specifically in living cells with modified aminoacyl-tRNA synthetases:
(1) *p*-azido-L-phenyalanine,
(2) *p*-benzoyl-L-phenylalanine,
(3) *o*-nitrobenzyl-L-cysteine,
(4) *p*-iodo-L-phenylalanine,
(5) *m*-hydroxy-L-tyrosine,
(6) *p*-acetyl-L-phenylalanine,
(7) *m*-acetyl-L-phenylalanine,
(8) *p*-propargyloxy-L-phenylalanine,
(9) β-GlcNAc-L-serine,
(10) β-GlcNAc-L-threonine,
(11) *p*-amino-L-phenylalanine,
(12) homoglutamine, (13) *O*-allyl-L-tyrosine,
(14) α-aminocaprylic acid,
(15) *O*-methyl-L-tyrosine,
(16) 3-(2-naphthyl)-L-alanine,
(17) *p*-isopropyl-L-phenylalanine,
(18) *p*-bromo-L-phenylalanine,
(19) 5-hydroxy-L-tryptophan,
(20) *m*-iodo-L-tyrosine.

3 Expanding the Genetic Code of Eukaryotic Cells

3.1 Stoichiometric Aminoacylation Methods

The first expansion of a cell's genetic code was performed in *Xenopus* oocytes by microinjection of a chemically misacylated *Tetrahymena thermophila* tRNA(CUA), and a mutant mRNA containing the *amber* nonsense codon. An advantage of this approach, shared with the *in vitro* incorporation methods from which it is derived, is that the chemical groups that can be incorporated are constrained only by the translational machinery of the cell (which is highly permissive to a variety of amino acids with altered side chains) and by the chemistry of aminoacylation. This methodology allowed the detailed biophysical study of the nicotinic acetylcholine receptor in oocytes by the introduction of amino acids containing side chains with unique physical or chemical properties and the study their effect using patch clamp methods. A disadvantage of this approach for studying other cellular proteins is that it is not applicable to producing mutant versions of many proteins at their normal *in vivo* levels or for producing enough protein to assay using standard cell biological approaches.

Recent extensions of this approach have extended the type of eukaryotic cell that it can be used to study, by altering the method of delivery of chemically acylated tRNAs and mRNA. It has been shown that chemically aminoacylated tRNAs can be electroporated or transfected into mammalian cells (including neurons) in culture. Again, since the tRNA cannot be re-acylated *in vivo*, the yield of protein produced is very low and this potentially limits the range of cellular processes that these methods can be used to study.

3.2 Catalytic Aminoacylation Methods

There has been considerable interest in extending the principles for expanding the genetic code of bacteria to eukaryotic cells. This would have significant technical and practical advantages over stoichiometric incorporation methods, since tRNAs would be acylated by their cognate synthetases, leading to large amounts of mutant protein. Moreover, genetically encoded aminoacyl-tRNA synthetases and tRNAs are, in principle, heritable, allowing the unnatural amino acid to be incorporated into proteins through many cell divisions without exponential dilution.

The steps necessary to add new amino acids to the genetic code of *E. coli* have been described; similar principles are likely to be useful for expanding the genetic code of eukaryotes. However, because the cellular tRNAs and aminoacyl-tRNA synthetases are not well conserved between prokaryotes and eukaryotes, the translational components from *M. janaschii* added to *E. coli* are not orthogonal in eukaryotic cells. The challenges in extending this approach are therefore to discover new orthogonal pairs in which the tRNA is produced in a functional form *in vivo*, to devise selection methods for altering the specificity of the orthogonal synthetase, and to create large libraries of mutant synthetases in a eukaryotic cell.

3.3 Orthogonal Aminoacyl-tRNA Synthetase/tRNA Pairs in Eukaryotic Cells

The orthogonal pairs reported to date for eukaryotes are derived from bacterial

pairs. This reflects the fact that the divergence of identity elements of tRNAs from prokaryotes to eukaryotes is greater than that from eukaryotes and archea and between one eukaryote and another. However, it is clear that in some instances bacterial tRNA cannot simply be cut and pasted into eukaryotic cells. In eukaryotes, the synthesis of tRNAs is dependent on RNA Polymerase III and its associated factors, which recognize A and B box sequences that are internal to the tRNA structural gene ("internal promoters"). The A and B box consensus sequences place limitations on the tRNA sequences that can be used in eukaryotic cells, and some bacterial tRNAs that diverge from the eukaryotic A and B box sequence are simply not transcribed in eukaryotic cells. Moreover, many eukaryotic tRNAs have the 3′ CCA sequence added enzymatically by CCA transferase, while all *E. coli* tRNAs encode this sequence in the DNA sequence, and it may be necessary to delete the 3′ CCA in the *tRNA* gene imported from *E. coli*. The processing or transcription of a tRNA appears, in some cases, to require organism-specific flanking sequences 5′ and 3′ to the tRNA gene. Posttranscriptional nucleotide modifications, that may be absent in the host eukaryote, may also be necessary to produce a functional tRNA molecule. In addition, since the tRNA is transcribed in the nucleus, it must be exported to the cytoplasm to function in translation via an exportin-dependent process. Nonetheless, several potential orthogonal pairs have been described in eukaryotes.

E. coli tyrosyl-tRNA(CUA) can be expressed and correctly processed, as judged by Northern blot, in *S. cerevisiae* when the flanking sequences 5′ and 3′ to the structural gene are included and the 3′ flanking sequence is altered to a poly T stretch to allow Pol III transcriptional termination. To demonstrate the functionality and orthogonality of *E. coli* tyrosyl-tRNA(CUA) in yeast, a reporter strain was constructed that contained three auxotrophic *amber*-suppressible alleles, allowing qualitative comparison of *amber* suppression on media lacking methionine, tryptophan, or histidine. When this strain was transformed with a plasmid containing the *E. coli* tRNA(CUA) at a variety of copy numbers, all transformants were MET^- HIS^- TRP^-, indicating that any aminoacylation of this tRNA by yeast cytoplasmic synthetases was at a low level. When the same yeast strain was cotransformed with plasmids encoding the *E. coli* tRNA(CUA) and its cognate synthetase, suppression of all three alleles was observed, and shown to be correlated with the expression level of the tRNA. These experiments demonstrated that *E. coli* tyrosyl-tRNA is orthogonal in *S. cerevisiae*, and can be aminoacylated by its cognate synthetase. Numerous *in vitro* experiments have demonstrated that *E. coli* tyrosyl-tRNA synthetase does not aminoacylate total *S. cerevisiae* tRNA. Thus, the *E. coli* tyrosyl-tRNA synthetase/tRNA pair is a candidate orthogonal pair. Two other orthogonal pairs have subsequently been reported for *S. cerevisiae*. One is derived from *E. coli* glutamine-tRNA synthetase and an amber suppressor derived from human initiator tRNA while the other is derived from an *E. coli* leucine pair. Thus far the tyrosine and leucine orthogonal pairs form *E. coli* have been used to expand the genetic code of *S. cerevisiae*.

There have also been efforts to create orthogonal pairs in higher eukaryotic cells. A derivative of the *E. coli* glutamine pair is orthogonal and functional in mammalian cells. The tRNA of this pair, a variant of *E. coli* $tRNA^{Gln}$(CUA), was created by replacing C9 in the bacterial tRNA with the

mammalian A box consensus base A9, and the native flanking sequences were replaced with those of the human initiator tRNA. These changes were shown to drastically increase the transcription of the tRNA in a HeLa cell-free system. When COS-1 cells were transformed with a CAT reporter containing an amber codon at a permissive site, a $tRNA^{Gln}$(CUA) gene driven to high copy on an SV40-based vector and an *Ec*GlnRS gene, *amber* suppression was observed in a CAT assay. CAT activity was dependent on the presence of all three components. Since *Ec*GlnRS does not charge total mammalian tRNA, the *Ec*GlnRS/tRNA(CUA) constitute an orthogonal pair in mammalian cells. However, the amount of suppression observed with this pair was much lower than observed for a human serine amber suppressor, suggesting that this synthetase/tRNA pair is far from optimal. Yokoyama and coworkers have subsequently investigated bacterial tyrosyl-tRNA synthetase tRNA(CUA) pairs for their potential as orthogonal pairs in mammalian cells. These experiments showed that *E. coli* tyrosyl synthetase/*Bst*RNA(CUA) form a potential orthogonal pair in mammalian cells. The efficiency of this pair appears to be good, with Ras yields from an amber mutant of the *ras* gene being about 25% of those from the wild-type *ras* gene in the same system. The synthetase of this pair has subsequently been altered to allow the incorporation of numerous unnatural amino acids in mammalian cells (see the following). In addition, it has subsequently been shown that the tryptophanyl synthetase and cognate opal suppressor derived from *B. subtilis* are orthogonal in mammalian cells. The efficiency of the pair appears to be comparable to the *E. coli* tyrosine pair and it was used to incorporate an unnatural amino acid in mammalian cells.

3.4 Methods for Evolving Synthetases with Unnatural Amino Acid Specificity in Eukaryotic Cells

With an orthogonal aminoacyl-tRNA synthetase that aminoacylates the orthogonal suppressor tRNA in hand, a method is required to alter the specificity of the aminoacyl-tRNA synthetase so that it aminoacylates its cognate tRNA, and only its cognate tRNA with desired unnatural amino acid and no endogenous amino acid present in a eukaryotic cell.

The first approach to this problem was to screen a small number of active site variants. Using this method, Yokoyama and coworkers evaluated a small collection of *Ec*TyrRS mutants in an *in vitro* wheat-germ translation system and discovered an *Ec*TyrRS variant that utilizes 3-iodotyrosine more effectively than tyrosine. However, this enzyme still incorporates tyrosine in the absence of the unnatural amino acid. More recently, Yokoyama and coworkers have also demonstrated that this *Ec*TyrRS mutant functions with a tRNA(CUA) from *B. stearothermophilus* (see preceding text for a discussion of this orthogonal pair) to suppress *amber* codons in mammalian cells. The fidelity of incorporation was shown by LC-MS to be greater than 95%. A similar approach was subsequently used to discover mutants of the orthogonal *B. subtilis* TrpRS/tRNA(UCA) that incorporate 5-hydroxy-L-tryptophan (5-HTPP). On the basis of the crystal structure of the homologous enzyme from *B. stearothermus*, Val144 that points toward C5 of bound tryptophan adenylate was mutated to all other 19 amino acids. One mutant Val144 → Pro was able to suppress a TGA codon placed in the *foldon* gene in the presence of 5-HTPP. Moreover, in the absence of the unnatural

amino acid, no full-length Foldon protein was synthesized, suggesting the mutant synthetase is specific for 5-HTPP. Mass spectrometry confirmed the incorporation with a fidelity >97%.

To add numerous unnatural amino acids to the genetic code of eukaryotic cells, Chin et al. developed a general, *in vivo* selection method for the discovery of variants of *Ec*TyrRS/tRNA(CUA) or other orthogonal pairs that function in yeast to incorporate unnatural amino acids, but no endogenous amino acids, in response to the *amber* codon UAG. A major advantage of a selection is that enzymes that incorporate unnatural amino acids can be rapidly selected and enriched from libraries of 10^8 *Ec*TyrRS active site variants, 6 to 7 orders of magnitude more diverse than can be easily screened *in vitro*. This increase in diversity vastly increases the likelihood of isolating *Ec*TyrRS variants for the incorporation of a diverse range of useful functionality with very high fidelity.

A key to the selection was a strain of *S. cerevisiae*, MaV203:pGADGAL4 (2TAG) (see Fig. 2), that contains a plasmid encoded gene for the transcriptional activator protein GAL4 in which the codons for Thr44 and Arg110 (for which substitution with numerous other amino acids has minimal effects on GAL4 activity) were converted to amber nonsense codons. Suppression of these amber codons leads to the production of full-length GAL4, which in turn leads to transcriptional activation of genomic GAL4 responsive *HIS3*, *URA3*, and *LACZ* reporter genes. Expression of *HIS3* and *URA3* complement the histidine and uracil auxotrophy in this strain, allowing clones expressing active aaRS/tRNA pairs to be selected. Upon addition of 5-fluoroorotic acid (5-FOA), which is converted to a toxic product by the URA3 protein, clones expressing inactive aaRS/tRNA pairs can be selected. The *lacZ* reporter allows active and inactive synthetase/tRNA pairs to be discriminated colorimetrically. Thus, a single strain of *S. cerevisiae* can be used for positive and negative selections, minimizing the number of manipulations necessary to identify aaRS/$tRNA_{CUA}$ pairs that use unnatural amino acids.

To use this selection method to expand the eukaryotic genetic code, a library of EcTyrRS mutants was generated. On the basis of the crystal structure of the homologous *Bs*TyRS in complex with tyrosine, five residues within 7 Å of the *para* position of the bound tyrosine were randomly mutated to all amino acids, and the library was transformed into the selection strain yielding 10^8 transformants. To select for cells containing mutant synthetases that selectively aminoacylate the orthogonal *E. coli* tRNA(CUA), a two-step selection was carried out. In the first step, a positive selection for cells able to grow on media lacking uracil or containing 20 mM 3-aminotriazole (3-AT, a competitive inhibitor of the HIS3 protein) in the absence of histidine was performed. If a mutant TyrRS charges tRNA(CUA) with an amino acid, then the cell biosynthesizes histidine and uracil and survives. In a second step aimed at removing clones that incorporate endogenous amino acids in response to an amber codon, a negative selection was applied. Cells were grown on media containing 0.1% 5-FOA, but lacking the unnatural amino acid. Those cells expressing URA3, as a result of suppression of the GAL4 amber mutations with natural amino acids, convert 5-FOA to a toxic product and cause cell death. Surviving clones were amplified in the presence of unnatural amino acid and reapplied to the positive selection. Synthetases for the incorporation of numerous unnatural

amino acids have been developed through this method.

The specificity and activity of each mutant synthetase/tRNA pair was tested in several ways. First, individual clones were tested for the ability to confer survival on MaV203:pGADGAL4 on – Ura or – His +20 mM 3-AT in the presence and absence of an unnatural amino acid, and the blue/white phenotype was tested on X-Gal in the presence and absence of the amino acid. Clones that grow on selective media and turn blue on X-Gal only in the presence of the unnatural amino acid were selected for further characterization. Second, an amber mutant of hexahistidine tagged human superoxide dismutase (SOD) was overexpressed in the presence of each synthetase/tRNA pair with and without cognate unnatural amino acid (see Fig. 5). In the presence of the unnatural amino acid, SOD-His_6 was produced, but in the absence of added amino acid no SOD could be purified. Moreover, the SOD purified from yeast cell growth in the presence of each unnatural amino acid was shown, by mass spectrometry, to contain the desired unnatural amino acid at the desired position with fidelity in excess of 99.9%. The total yield of purified SOD was about 20% of that produced from the analogous wild-type *SOD* gene.

Synthetases evolved in yeast can be readily transplanted to higher eukaryotes when paired with a functional cognate orthogonal tRNA for the higher organism. Using this strategy, the unnatural amino acid dependent synthetases evolved in yeast have been transplanted into mammalian cells, along with the cognate amber-suppressor tRNA derived from *B. stearothermophilus*, and incorporation of each unnatural amino acid demonstrated.

4 Using Expanded Genetic Codes

4.1 The Utility of Expanded Genetic Codes

Several applications of expanded genetic codes have been demonstrated. Numerous future applications can be predicted from applications that have been achieved with other methods including the *in vitro* incorporation of amino acids, multisite amino acid–incorporation, and expressed protein ligation. Perhaps even more exciting are the entirely novel applications that are beginning to emerge, in which *in vivo* incorporation of unnatural amino

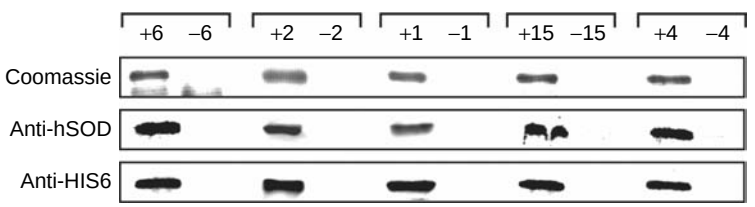

Fig. 5 Unnatural amino acid dependent protein expression in yeast: An *hSOD* gene containing an amber codon at position 33 was used to produce protein in yeast strains with aminoacyl-tRNA synthetase/tRNA pairs for specific amino acids (see Fig. 4). HSOD was purified from yeast in the presence (+) and absence (−) of the unnatural amino acid indicated and SOD expression was visualized by Coomassie stain or Western blot using antibodies against hSOD and the C-terminal hexahistidine tag.

acids are used to interrogate processes inside living cells with molecular precision. Here we focus on the utility of the expanded genetic code in *E. coli* using pairs derived from *Mj*TyrRS/tRNA(CUA) or in yeast using the *Ec*TyrRS/tRNA(CUA) pair or EcLeuRS/tRNA(CUA).

4.2 Biophysical Probes

Aryl iodides have been site specifically incorporated into proteins in both *E. coli* and yeast, and have subsequently been used for X-ray structure determination via single wavelength anomalous diffraction on a home X-ray source. The method appears to require significantly less data than selenomethionine incorporation. Both *p*-benzoyl-phenylalanine and *p*-azido-phenylalanine have been incorporated into proteins in bacteria and yeast (see Fig. 6). These amino acids allow photoaffinity labeling of interacting proteins upon irradiation at 360 nm. The covalently linked proteins can then be isolated or visualized electrophoretically after *in vitro* or *in vivo* cross-linking. This utility of this method was demonstrated in a model system, and has subsequently been used to determine a binding site of a client protein on the chaperone ClpB. Isotopic labeling (e.g. ^{19}F) of unnatural amino acids should aid NMR structure determination. Other biophysical probes such as spin labels for EPR studies are likely future targets.

The genetic code of yeast has been expanded to include a nitrobenzyl cysteine. This photocaged cysteine can be deprotected, allowing the temporal control of protein function. Several unnatural amino acids have been useful for engineering chemical scission sites into proteins, though these have not yet been incorporated into proteins *in vivo*. Hydroxyacids inserted through chemical acylation of tRNA result in an ester linkage that can be cleaved in base. Similarly, specific cleavage at aminooxyacetic acid can be achieved with zinc and acetic acid, and cleavage of the protein backbone at allylglycine

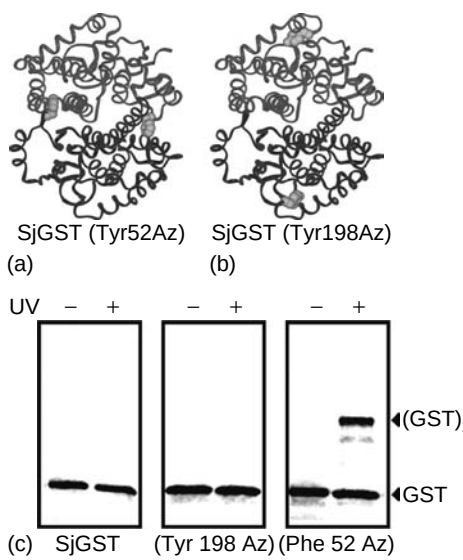

Fig. 6 An example of the utility of expanded genetic codes: *p*-azido-L-phenylalanine (1) is inserted into glutathionine-*S*-transferase at position 52 (a) or position 198 (b). On irradiation with UV light (c) the azido group photocross-links protein monomers in the dimeric GST complex. This cross-linking is dependent on the azido group being at the protein interface (position 52). Similar experiments have been used to map protein–protein interactions both *in vitro* and in cells. (See color plate p. xxv.)

is possible upon treatment with iodine. Photochemical proteolysis occurs upon irradiation at 2-nitrophenylglycines.

4.3 Unique Reactive Handles

Expanding the reactivity of proteins, particularly in a site-specific way, is a potentially useful alteration to proteins. Chemical modification of natural proteins is basically limited to use of the nucleophilic groups: cysteine's thiol, amines on lysine, and at the N terminus, histidine's imidazole, and the alcohols, depending upon pH. However, it is often difficult to prevent cross-reaction with other nucleophilic groups in the protein, and extensive mutation (for example, to yield a single-Cys version of a protein) is often required. It is possible to introduce an electrophile into a protein by mild oxidation of N-terminal serine or threonine, but more general approaches to unique reactive handles have many potential applications.

The genetic codes of *E. coli* and yeast have been expanded to include a ketone that allows the chemoselective formation of hydrazones and oximes using hydrazides and hydroxylamines respectively. Hydrazone formation is possible in living cells to label ketone-containing protein with fluorophore hydrazides. Both cytosolic and outer-membrane proteins have been labeled in this fashion. Ketones also allow for the possibility of protein stabilization by formation of an imine with lysine. The addition of an azides and alkynes to the genetic code of both *E. coli* and yeast allows site specific–protein derivatization via a Cu(I) catalyzed $3 + 2$ cycloaddition. This reaction has been used to specifically PEGylate proteins or to add biotin or fluoroscein groups. An alternative derivatization strategy for azides would involve a Staudinger-like ligation, which has been used to selectively modify sugar moieties on cell surfaces, in mice, and for the labeling of azidohomoalanine introduced into proteins using *in vivo* amino acid replacement.

4.4 Posttranslational Modifications

Understanding the roles of proteins in cellular function necessarily means understanding the extent and function of posttranslational modification, particularly in eukaryotes. However, it is difficult to obtain specifically, homogeneously modified protein. The genetic code of *E. coli* has been expanded for the incorporation of β-GlcNAc-serine, where proteins are not normally glycosylated, permitting homogeneous preparation of large amounts of glycoprotein. Expressed protein ligation has also been used to study posttranslational modifications via the synthesis of phosphorylated or nonhydrolyzable phosphorylation mimic amino acids, or via geranylgeranyl amino acids. It is likely that the expansion of the genetic code will allow the direct incorporation of these and other posttranslational modifications into proteins *in vivo*.

4.5 Organism Evolution

In all the above studies, the amino acid must be added to the growth medium and taken up by the cell in order to be incorporated into a protein in response to the *amber* codon. An alternative strategy is to engineer a pathway for the biosynthesis of the unnatural amino acid in the host organism. In this way, a completely autonomous bacterium has been generated that contains genes for the biosynthesis of *p*-amino-L-phenylalanine

(pAF) from simple carbon sources, an aminoacyl-tRNA synthetase that uses this new amino acid and no other endogenous amino acids, and a tRNA that delivers into proteins in response to the *amber* codon.

To biosynthesize pAF, the *papA*, *papB*, *papC* genes from *S. venezuelae* were imported into *E. coli*. The PapA protein, 4-amino-4-deoxychorismate, was expected to convert chorismate to 4-amino-4-deoxychorismic acid using ammonia from glutamine in a simple addition-elimination reaction. PapB should convert 4-amino-4-deoxychorismic acid to 4-amino-4-deoxyprephenic acid, and PapC should then convert 4-amino-4-deoxyprephenic acid to *p*-aminophenylpyruvic acid. The endogenous aminotransferases (*tyrB*, *asps*, *ilvE*) were expected to convert this to pAF. *E. coli* containing these genes were found to produce *p*-amino-L-phenylalanine at levels comparable to those of other aromatic amino acids and had normal growth rates. Cells producing *p*-amino-L-phenylalanine were able to incorporate it into myoglobin in response to an *amber* codon with high fidelity, as judged by mass spectrometry of the mutant protein and in high yield of 2 mg L^{-1} of culture. These yields are comparable to those observed when the unnatural amino acid is added to the growth media directly. In addition to *p*-amino-L-phenylalanine, it should be possible to biosynthesize and genetically encode other amino acids *in vivo* as well including methylated, acetylated, and glycosylated amino acids.

The creation of a 21 or 22 amino acid code removes a billion year constraint on the evolution of life, allowing the experimental assessment of the fitness of organisms with expanded genetic codes. By placing *amber*, *opal*, or four-base codons throughout the genome, we can compare the survival of various 20, 21, and 22 amino acid organisms in response to selective pressures. Some unnatural amino acids, such as ketones, expand the reactivity of proteins, permitting new mechanisms of stabilization (imines with lysine, for example) and new chemistries (ketones and aldehydes are found in cofactors but not among the natural amino acids). It will be interesting to see how nature can make use of these new reactivities in response to stresses like extremes of temperature of pH.

5 Summary

Expanding the natural repertoire of amino acids is creating a powerful set of tools for biochemical, biophysical, and cellular studies far beyond the possibilities afforded by site-directed mutagenesis. Engineering the biosynthetic machinery also tests our understanding of how organisms achieve fidelity in protein synthesis, and has resulted in some of the most profound changes in enzyme function attained by any method of protein engineering. However, it is still not possible to easily generate a synthetase with substrate specificity for any desired amino acid, and there is a strict limit on code expansion imposed by the small number of blanks in the genetic code from suppressible stop codons and 4-base codons. The development of further highly active and orthogonal tRNA/synthetase pairs and the development of unnatural codons will be major areas of effort in the upcoming decade.

Acknowledgment

Thomas J. Magliery is an N.I.H. Postdoctoral Fellow (GM065750), and he

thanks Lynne Regan (Yale University) for additional support while writing this manuscript.

See also Bacterial Cell Culture Methods; Synthetic Peptides: Chemistry, Biology, and Drug Design.

Bibliography

Books and Reviews

Cornish, V.W., Mendel, D., Schultz, P.G. (1995) Probing protein-structure and function with an expanded genetic-code, *Angew. Chem. Int. Ed. Engl.* **34**, 621–633.

Cropp, T.A., Schultz, P.G. (2004) An expanding genetic code, *Trends Genet.* **20**, 625–630.

Dawson, P.E., Kent, S.B. (2000) Synthesis of native proteins by chemical ligation, *Annu. Rev. Biochem.* **69**, 923–960.

Hohsaka, T., Sisido, M. (2002) Incorporation of non-natural amino acids into proteins, *Curr. Opin. Chem. Biol.* **6**, 809–815.

Magliery, T.J. (2005) Unnatural protein engineering: producing proteins with unnatural amino acids, *Med. Chem. Rev. Online* (expected August, 2005).

Magliery, T.J., Pasternak, M., Anderson, J.C., Santoro, S.W., Herberich, B., Meggers, E., Wang, L., Schultz, P.G. (2003) In vitro Tools and in vivo Engineering: Incorporation of Unnatural Amino Acids into Proteins, in: Lapointe, J., Brakier-Gingras, L., (Eds.) *Translation Mechanisms*, Landes Bioscience/Eurekah.com, Georgetown, TX and Kluwer Academic/Plenum, New York, pp. 95–114.

Muir, T.W. (2003) Semisynthesis of proteins by expressed protein ligation, *Annu. Rev. Biochem.* **72**, 249–289.

Soll, D., RajBhandary, U.L. (Eds.) (1995) *tRNA: Structure, Biosynthesis and Function*, ASM Press, Washington, DC.

Primary Literature

Anderson, J.C., Schultz, P.G. (2003) Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression, *Biochemistry* **42**, 9598–9608.

Anderson, J.C., Magliery, T.J., Schultz, P.G. (2002) Exploring the limits of codon and anticodon size, *Chem. Biol.* **9**, 237–244.

Anderson, J.C., Wu, N., Santoro, S.W., Lakshman, V., King, D.S., Schultz, P.G. (2004) An expanded genetic code with a functional quadruplet codon, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7566–7571.

Chin, J.W., Schultz, P.G. (2002) In vivo photocrosslinking with unnatural amino acid mutagenesis, *ChemBioChem* **11**, 1135–1137.

Chin, J.W., Martin, A.B., King, D.S., Wang, L., Schultz, P.G. (2002) Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11020–11024.

Chin, J.W., Cropp, T.A., Chu, S., Meggers, E., Schultz, P.G. (2003) Progress toward an expanded eukaryotic genetic code, *Chem. Biol.* **10**, 511–519.

Chin, J.W., Santoro, S.W., Martin, A.B., King, D.S., Wang, L., Schultz, P.G. (2002) Addition of *p*-Azido-L-phenylalanine to the genetic code of *Escherichia coli*, *J. Am. Chem. Soc.* **124**, 9026–9027.

Chin, J.W., Cropp, T.A., Anderson, J.C., Mukherji, M., Zhang, Z., Schultz, P.G. (2003) An expanded eukaryotic genetic code, *Science* **301**, 964–967.

Deiters, A., Cropp, T.A., Summerer, D., Mukherji, M., Schultz, P.G. (2004) Site-specific PEGylation of proteins containing unnatural amino acids, *Bioorg. Med. Chem. Lett.* **14**, 5743–5745.

Deiters, A., Cropp, T.A., Mukherji, M., Chin, J.W., Anderson, J.C., Schultz, P.G. (2003) Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*, *J. Am. Chem. Soc.* **125**, 11782–11783.

Drabkin, H.J., Park, H.J., RajBhandary, U.L. (1996) Amber suppression in mammalian cells dependent upon expression of an *Escherichia coli* aminoacyl-tRNA synthetase gene, *Mol. Cell. Biol.* **16**, 907–913.

Edwards, H., Schimmel, P. (1990) A bacterial amber suppressor in *Saccharomyces cerevisiae* is selectively recognized by a bacterial aminoacyl-tRNA synthetase, *Mol. Cell. Biol.* **10**, 1633–1641.

Furter, R. (1998) Expansion of the genetic code: site-directed *p*-fluoro-phenylalanine

incorporation in *Escherichia coli*, *Protein Sci.* **7**, 419–426.

Henry, A.A., Romesberg, F.E. (2003) Beyond A, C, G and T: augmenting nature's alphabet, *Curr. Opin. Chem. Biol.* **7**, 727–733.

Hohsaka, T., Ashizuka, Y., Sisido, M. (1999) Incorporation of two nonnatural amino acids into proteins through extension of the genetic code, *Nucleic Acids Symp. Ser.* **42**, 79–80.

Ibba, M., Hennecke, H. (1995) Relaxing the substrate specificity of an aminoacyl-tRNA synthetase allows in vitro and in vivo synthesis of proteins containing unnatural amino acids, *FEBS Lett.* **364**, 272–275.

Kiga, D., Sakamoto, K., Kodama, K., Kigawa, T., Matsuda, T., Yabuki, T., Shirouzu, M., Harada, Y., Nakayama, H., Takio, K., Hasegawa, Y., Endo, Y., Hirao, I., Yokoyama, S. (2002) An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9715–9720.

Kobayashi, T., Nureki, O., Ishitani, R., Yaremchuk, A., Tukalo, M., Cusack, S., Sakamoto, K., Yokoyama, S. (2003) Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion, *Nat. Struct. Biol.* **10**, 425–432.

Kohrer, C., Xie, L., Kellerer, S., Varshney, U., RajBhandary, U.L. (2001) Import of amber and ochre suppressor tRNAs into mammalian cells: a general approach to site-specific insertion of amino acid analogues into proteins, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14310–14315.

Kowal, A.K., Kohrer, C., RajBhandary, U.L. (2001) Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2268–2273.

Link, A.J., Mock, M.L., Tirrell, D.A. (2003) Non-canonical amino acids in protein engineering, *Curr. Opin. Biotechnol.* **14**, 603–609.

Liu, D.R., Schultz, P.G. (1999) Progress toward the evolution of an organism with an expanded genetic code, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4780–4785.

Liu, D.R., Magliery, T.J., Pasternak, M., Schultz, P.G. (1997) Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10092–10097.

Magliery, T.J., Anderson, J.C., Schultz, P.G. (2001) Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in *Escherichia coli*, *J. Mol. Biol.* **307**, 755–769.

Mehl, R.A., Anderson, J.C., Santoro, S.W., Wang, L., Martin, A.B., King, D.S., Horn, D.M., Schultz, P.G. (2002) Generation of a bacterium with a 21 amino acid genetic code, *J. Am. Chem. Soc.* **125**, 935–939.

Monahan, S.L., Lester, H.A., Dougherty, D.A. (2003) Site-specific incorporation of unnatural amino acids into receptors expressed in Mammalian cells, *Chem. Biol.* **10**, 573–580.

Noren, C.J., Anthonycahill, S.J., Griffith, M.C., Schultz, P.G. (1989) A general-method for site-specific incorporation of unnatural amino-acids into proteins, *Science* **244**, 182–188.

Nowak, M.W., Kearney, P.C., Sampson, J.R., Saks, M.E., Labarca, C.G., Silverman, S.K., Zhong, W., Thorson, J., Abelson, J.N., Davidson, N., Schultz, P.G., Dougherty, D.A., Lester, H.A. (1995) Nicotinic receptor-binding site probed with unnatural amino-acid-incorporation in intact-cells, *Science* **268**, 439–442.

Pasternak, M., Magliery, T.J., Schultz, P.G. (2000) A new orthogonal suppressor tRNA/aminoacyl-tRNA synthetase pair for evolving an organism with an expanded genetic code, *Helv Chim Acta* **83**, 2277.

Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., Kobayashi, T., Kitabatake, M., Takio, K., Saito, K., Shirouzu, M., Hirao, I., Yokoyama, S. (2002) Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells, *Nucleic Acids Res.* **30**, 4692–4699.

Santoro, S.W., Wang, L., Herberich, B., King, D.S., Schultz, P.G. (2002) An efficient system for the evolution of aminoacyl-tRNA synthetase specificity, *Nat. Biotechnol.* **20**, 1044–1048.

Schlieker, C., Weibezahn, J., Patzelt, H., Tessarz, P., Strub, C., Zeth, K., Erbse, A., Schneider-Mergener, J., Chin, J.W., Schultz, P.G., Bukau, B., Mogk, A. (2004) Substrate recognition by the AAA+ chaperone ClpB, *Nat. Struct. Mol. Biol.* **11**, 607–615.

Trezeguet, V., Edwards, H., Schimmel, P. (1991) A single base pair dominates over the novel identity of an *Escherichia coli* tyrosine tRNA in *Saccharomyces cerevisiae*, *Mol. Cell. Biol.* **11**, 2744–2751.

Wang, L., Schultz, P.G. (2001) A general approach for the generation of orthogonal tRNAs, *Chem. Biol.* **8**, 883–890.

Wang, L., Magliery, T.J., Liu, D.R., Schultz, P.G. (2000) A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins, *J. Am. Chem. Soc.* **122**, 5010–5011.

Wang, L., Brock, A., Herberich, B., Schultz, P.G. (2001) Expanding the genetic code of *Escherichia coli*, *Science* **292**, 498–500.

Wolin, S.L., Matera, A.G. (1999) The trials and travels of tRNA, *Genes Dev.* **13**, 1–10.

Wu, N., Deiters, A., Cropp, T.A., King, D., Schultz, P.G. (2004) A genetically encoded photocaged amino acid, *J. Am. Chem. Soc.* **126**, 14306–14307.

Xie, J., Wang, L., Wu, N., Brock, A., Spraggon, G., Schultz, P.G. (2004) The site-specific incorporation of *p*-iodo-L-phenylalanine into proteins for structure determination, *Nat. Biotechnol.* **22**, 1297–1301.

Zhang, Z., Alfonta, L., Tian, F., Bursulaya, B., Uryu, S., King, D.S., Schultz, P.G. (2004) Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8882–8887.

Zhang, Z., Gildersleeve, J., Yang, Y.Y., Xu, R., Loo, J.A., Uryu, S., Wong, C.H., Schultz, P.G. (2004) A new strategy for the synthesis of glycoproteins, *Science* **303**, 371–373.

Protein Mediated Membrane Fusion

Reinhard Jahn
Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany

1 Introduction and Overview 71

2 Physical Principles of Bilayer Fusion 72

3 Transition States of Protein-mediated Fusion and Fusion Pores 74

4 Viral Fusion Proteins 75
4.1 Structure of Class I Fusion Proteins 75
4.2 Fusion Mechanism of Class I Proteins 76
4.3 Class II Viral Fusion Proteins 78

5 Mitochondrial Fusion 79

6 Fusion Reactions in the Secretory Pathway 79
6.1 Overview 79
6.2 Tethering and Docking – the Role of Rab Proteins and Their Effectors 81
6.3 Fusion – SNARE Proteins as Fusion Catalysts 83
6.3.1 The SNARE Assembly– Disassembly Cycle 84
6.3.2 Mechanism of SNARE-mediated Membrane Fusion 86
6.3.3 Conformational Intermediates of SNAREs 87
6.3.4 Topology and Specificity of SNAREs in Intracellular Fusion Reactions 89
6.4 Regulators of SNAREs 91
6.4.1 SM Proteins 91
6.4.2 Synaptotagmin and Complexin – Regulators of Ca^{2+}-dependent Exocytosis 92
6.4.3 "Pseudo" SNAREs 94

7 Conclusion and Outlook 94

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

Bibliography 95
Books and Reviews 95
Primary Literature 95

Keywords

Exocytosis
Fusion of an intracellular organelle (usually a trafficking vesicle or a specialized secretory vesicle) with the plasma membrane. Exocytosis can be further subdivided into regulated and constitutive exocytosis. Examples for the former are neurotransmitter release by exocytosis of synaptic vesicles or hormone and enzyme release by secretion from exocrine and endocrine glands. Examples for the latter include the continuous incorporation of trafficking vesicles as required for receptor recycling and cell growth. For many years, neuronal exocytosis has served as a paradigm for studying the mechanisms of membrane fusion.

Fusion Peptides
Stretches of usually 20–60 amino acids that are part of viral fusion proteins and are thought to be inserted into the target membrane upon activation. Fusion peptides are amphiphilic, that is, when forming an α-helix one face contains hydrophobic/large-side chains whereas the rest of the surface is hydrophilic. Isolated fusion peptides insert into membranes, and several of them were shown to be potent fusogens, that is, they are capable of fusing membranes even without being attached to a protein.

Fusion Pore
The opening of an aqueous connection between the distal sides of the two fusing membranes. In a protein-mediated fusion, fusion pores are tight so that no leakage occurs between the distal and proximal sides of the fusing membranes. While fusion pores can be characterized with electrophysiological methods, the molecular structure of these intermediates in the fusion pathway is unknown. Both proteinaceous and lipidic intermediates are discussed.

Hemifusion
A hypothetical intermediate in bilayer fusion in which there is already a free exchange of membrane lipids, although an aqueous connection (fusion pore) is not yet formed.

SNAREs
A superfamily of small proteins (molecular weight between 10 and 35 kDa) that catalyze membrane fusion in the secretory pathway of eukaryotic cells. Each fusion step involves a specific set of SNAREs that spontaneously assemble into ternary or quaternary complexes, bridging the membrane and initiating fusion. After fusion, SNARE complexes are disassembled by the ATPase NSF (NEM-sensitive factor) in conjunction with cofactors termed SNAPs (soluble NSF attachment proteins).

Viral Fusion Proteins
Membrane proteins on the surface of enveloped viruses, that is, viruses that are surrounded by a phospholipid bilayer. Viral fusion proteins are metastable. Upon activation during cell contact or after endocytosis they undergo a spontaneous conformational change that mediates the fusion of the viral with the cell membrane, resulting in the injection of the nucleocapsid into the host cell cytoplasm. Viral fusion proteins are divided into class I and class II fusion proteins, which differ with respect to overall structure and fusion kinetics.

The fusion of biological membranes is one of the most fundamental reactions of life. It involves the initial recognition of the membranes, the close apposition of the membranes followed by the generation of nonbilayer intermediates, and finally the opening of an aqueous fusion pore. While the intermediates in various fusion reactions share common features, diverse classes of proteins evolved independently to carry out the task. Single proteins may be responsible for the entire reaction sequence as is the case for fusion proteins of enveloped viruses. In contrast, intracellular fusion reactions are mediated by cascades of reactions carried out by multiprotein complexes allowing for tight spatial and temporal regulation. Although the precise mechanism of fusion proteins is not yet fully understood, it is becoming clear that fusion proteins are nanomachines that exert mechanical force upon the membranes destined to fuse. Intriguingly, all characterized fusion proteins undergo major refolding during the reaction. Refolding proceeds downhill an energy gradient and yields highly stable oligomeric complexes with a stiff rodlike structure suggesting similarities in the fusion pathway.

1 Introduction and Overview

Membranes are fundamental to all forms of life. They form barriers that cannot be penetrated by hydrophilic molecules and that allow the enclosing of compartments containing biological macromolecules. Every cell is surrounded by a membrane that usually is represented by a bilayer of amphiphiles (phospholipids and related compounds). Furthermore, eukaryotic cells contain membrane-surrounded organelles.

Since the barrier function of membranes must be maintained at all times, special mechanisms that enable membranous compartments to merge or to split have been evolved. These mechanisms are referred to as *membrane fusion* and *membrane fission*. Both fusion and fission of biological membranes are characterized by specific features:

- no leakage occurs between the inside and the outside of the membrane-delineated spaces
- the asymmetry of the membrane is maintained, that is, amphiphiles do not randomize between the two monolayers
- they are mediated by specific proteins.

Biological membrane fusion events can be classified in

- extracellular fusions between cells, for example, sperm–egg fusion and fusions yielding polynucleated cells such as muscle cells or osteoclasts
- fusions of enveloped viruses with the plasma membrane of host cells
- intracellular fusions of and between organelles with the surrounding plasma membrane.

In recent years, significant progress has been made in identifying the proteins responsible for fusing biological membranes. Although there is still considerable debate about the precise role of many of these proteins, it has become clear that proteins capable of fusing membranes originated several times independently during evolution. Best understood are the fusion proteins of enveloped viruses. They fall into at least two major classes that are structurally unrelated to each other, and the diversity in either one of the classes is so great that even for these a common evolutionary origin is doubtful. Eukaryotic cells contain at least three classes of fusion proteins: those involved in extracellular fusions whose identities are largely unknown, those involved in the homotypic fusion of mitochondria, and finally those involved in fusion events of the so-called secretory pathway.

2 Physical Principles of Bilayer Fusion

Membranes are typically organized in bilayers composed of two leaflets of amphiphilic molecules such as phospholipids. These molecules have an elongated shape, with a hydrophilic head group and a hydrophobic tail that is often composed of hydrocarbon chains. The hydrophilic groups are exposed to the aqueous environment, whereas the hydrophobic tails face the hydrophobic interior of the bilayer.

Membrane fusion first requires that the two membranes destined to fuse be brought into close contact, requiring that electrostatic and hydration repulsion is overcome. Next, transition states that locally disrupt the bilayer structure must be generated, involving both hydrophilic and hydrophobic contacts between the membrane constituents. These transition states are governed by the hydrophobic effect, that is, by forces that minimize the exposure of hydrophobic surfaces to water. Finally, the bilayer structure is reinstated, which is associated with the local opening of an aqueous fusion pore.

Fusion sites are "dirty nanostructures" in that they have a variable composition involving hundreds of amphiphiles but with a size too small for resolving them with a light microscope. Furthermore, they dynamically change during the reaction, and it is not easy to arrest the fusion reaction at a defined intermediate state. Thus, the structure of the nonbilayer transition states cannot be directly determined. However, both theoretical and experimental approaches have led to well-founded models that describe the steps involved in membrane fusion in accordance with physical principles. The most widely accepted model is referred to as *stalk hypothesis* (Fig. 1). In its original version, the stalk hypothesis is based on macroscopic physics (continuum models) and describes the energy landscape of the fusion pathway on the basis of parameters such as bending energy, intrinsic curvature of the lipids, tilt of hydrocarbon chains, and it takes into account energy penalties due to the presence of hydrophobic voids in the

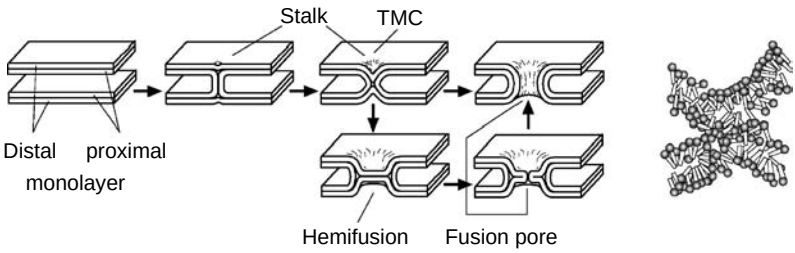

Fig. 1 Transition states in membrane fusion according to the stalk hypothesis. The monolayers are shown as smooth and bendable sheets. TMC, transmembrane contact. On the right, a snapshot of a nonbilayer intermediate is shown that is derived from a simulation of membrane fusion using simplified (coarse-grain) models for membrane lipids. Reprinted from Jahn, R., Lang, T., Südhof, T.C. (2003) Membrane fusion, *Cell* **112**, 519–533.

nonbilayer transition states. According to this hypothesis, the first steps involve the merger of proximal monolayers, a process that may be facilitated by thermal motions of individual amphiphiles perpendicular to the membrane plane. As a result, a fusion stalk is formed, followed by a symmetrical structure in which the first contact between the distal monolayers is established, which is also referred to as transmembrane contact (TMC). From this intermediate, a fusion pore may directly form or else, a hemifusion diaphragm is formed in which the contact zone of the distal monolayers expands, with a fusion pore forming at a later state (Fig. 1).

A large body of experimental evidence supports the overall description by the stalk hypothesis of the fusion pathway. For instance, structures similar to the intermediate involving the first TMC (Fig. 1) have been observed by X-ray diffraction in stacks of dehydrated bilayers. Similarly, such structures are part of the so-called rhombohedral/hexagonal phase of certain membrane lipids. Furthermore, addition of amphiphiles, which promotes either positive or negative curvature promotes or inhibits progression through the transition states as predicted by the model, and in many fusion events lipid mixing is observable before an aqueous connection is formed. However, the structure and precise orientation of the membrane lipids during the transition states is not known, and it has not been possible to show directly that stalks are intermediates of biological fusion reactions. Recent efforts to simulate membrane fusion using coarse-grain models also resulted in stalk-like intermediates, but the transition states appeared to be more disordered than originally assumed by the stalk hypothesis. While these models presently suffer from an arbitrary choice of fitting parameters, it is hoped that full atomistic simulations using refined force fields will yield more realistic and less biased structural models of fusion intermediates.

Unless strained or otherwise destabilized, membranes do not spontaneously fuse with each other under normal conditions (ambient temperature and pressure). However, amphiphilic molecules such as polyethylene glycol or certain amphipatic peptides can mediate spontaneous fusion of membranes. Similarly, certain amphipatic proteins (e.g. annexins) induce fusion of artificial membranes although there is no evidence that they act as fusion

proteins in their native environment. Apparently, the activation energy barrier for membrane fusion is not very high, requiring less specificity for the catalyst than, for example, an enzyme-catalyzed metabolite conversion.

3 Transition States of Protein-mediated Fusion and Fusion Pores

In the 1980s, two models were proposed explaining how proteins may mediate fusion. They describe two possible extreme situations that are instructive and that are still widely used as reference (Fig. 2). The first model describes the initial steps of fusion analogous to the formation and activation of gap junction channels. Each membrane contains oligomeric fusion proteins that assemble into a closed, ringlike oligomeric structure. Upon membrane contact, these rings connect and form an aqueous channel that represents the first opening of the fusion pore. Capacitance patch-clamping experiments revealed that fusion pores with an initial diameter of about 2 to 5 nm may reversibly open and close at a millisecond timescale ("flickering"), supporting the view that fusion pores, at least in their initial state, resemble channel proteins. Subsequently, both the oligomeric assembly and the TMC need to be broken up in order for the fusion pore to expand and for fusion to be completed. According to this model, the initial states of fusion including the transition states are primarily governed by protein–protein interactions and subsequently by protein–lipid interactions.

The second model assumes that the role of proteins in fusion is confined to providing a scaffold that lowers the activation energy barrier, while the transition states and the initial fusion pore is exclusively lipidic and thus must be in accordance with lipid biophysics. Accordingly, there is no need for the proteins to span the membrane. Rather, the role of fusion proteins may be

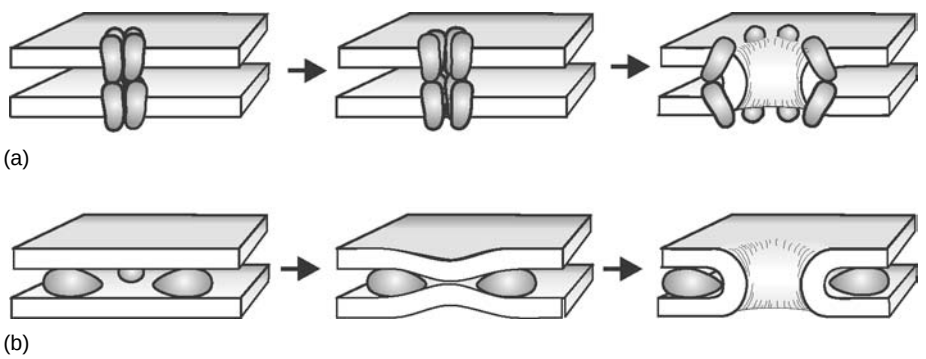

Fig. 2 Classical models for protein-mediated membrane fusion. (a) Channel-model. Here, the fusion proteins form a closed channel that connects in "trans," resulting in a tight connection between the two membranes that is reminiscent of gap junctions. Formation of a fusion pore is due to the opening of a proteinaceous channel. Enlargement is mediated by lipid invasion between the dissociating subunits. (b) Scaffold model. Here, the role of the proteins is confined to bending of the membrane (in the original version it was assumed that a dimple is created in one of the membranes), with all transition states governed by lipid biophysics.

confined to forming a tight connection between the membranes, or else, they may destabilize the hydrophilic–hydrophobic boundary of the membrane to facilitate nonbilayer transition states. Fusion pore flickering, originally considered to be a hallmark of protein-mediated fusion, was later also observed in fusion events of protein-free artificial membranes, thus depriving the "channel" model of one of its strongest supportive arguments.

As discussed in more detail below, it is still unknown which of the two models is more accurate. Furthermore, the hypothesis that different classes of fusion proteins operate by different mechanisms cannot be excluded. While most investigators agree that lipids play a major role in the early transition states as suggested by the scaffold model, modifications of fusion proteins were reported to directly affect fusion pore characteristics indicating that fusion proteins do more than just clamping membranes together.

4 Viral Fusion Proteins

Enveloped viruses are enclosed by a host cell–derived bilayer that protects the nucleocapsid containing the viral genome. In order to infect cells, the viral membrane needs to fuse with the host cell membrane, resulting in the delivery of the nucleocapsid into the cytoplasm of the host cell. For this task, enveloped viruses contain specific membrane proteins whose purpose is to execute the fusion of the viral with the host cell membrane. Viral fusion proteins can be divided into class I and class II proteins, with class I proteins being better understood.

4.1 Structure of Class I Fusion Proteins

Class I fusion proteins typically are homooligomeric glycoproteins with a single transmembrane domain. They are oriented perpendicular to the viral membrane and often form characteristic spikes. At or near the N-terminal end, these proteins contain an amphiphilic region of 15 to 35 amino acids, referred to as *fusion peptide*, which is essential for function. All class I fusion proteins have one remarkable feature in common: before the viruses bud from the host cell, the otherwise stably folded fusion proteins are posttranslationally processed by a single proteolytic cut yielding a stable globular N-terminal and a metastable C-terminal fragment containing the transmembrane domain. The metastable C-terminal fragments are the fusion catalysts. When these proteins are triggered by binding to a surface receptor or by low pH, as encountered after endocytosis of the virus particle, they spontaneously undergo major conformational changes. As a result, elongated helical bundles (usually consisting of trimeric coiled coils) are formed and the previously buried fusion peptides are exposed at the tip of the bundle. After completion of the conformational transition, the resulting protein is very stable and inactive. Thus, viral fusion proteins are "single-shot" devices resembling a loaded spring: they can be triggered only once. Biologically, this makes sense since they are needed only once in an infectious fusion reaction.

The fusion peptides play an essential role in virus-mediated fusion. These domains insert into the membrane of the host cell upon activation with the large hydrophobic side chains buried in the interior of the membrane. Perturbation of the amphiphilic nature of the fusion peptide,

for example by site-directed mutagenesis, either cripples or completely inactivates fusion activity. Interestingly, many isolated fusion peptides derived from viral fusion proteins are efficient fusogens: when added to a suspension of liposomes, spontaneous fusion occurs. However, it is likely that their prime role is to provide a firm anchor in the target membrane.

Much of our understanding of viral membrane proteins is derived from the influenza hemagglutinin (HA) fusion protein that for many years served as a paradigm for class I fusion proteins. It is the only fusion protein for which fragments were crystallized both for the metastable prefusion state and the relaxed postfusion state. In the prefusion state, the HA trimer contains a central trimeric coiled coil that is connected to the fusion peptide by a long loop that folds back to the base of the molecule. Triggering of the protein by low pH results in two major rearrangements. First, the connecting loop is converted into a helix that now extends the coiled-coil region, effectively propelling the fusion peptide from the base to the top of the elongated molecule. Second, a reverse turn is generated in the middle of the original helix, rotating the C-terminal part by 180°, with these helical parts and some nonhelical extensions now fitting into the grooves of the central trimer (Fig. 3). While the purpose of the first molecular motion is obvious (exposure of the fusion peptide and its "shooting" into the target membrane), the significance of the second motion was only fully appreciated when structures of other viral fusion proteins (e.g. gp41 of the human immunodeficiency virus) were solved in which such back folding (resulting in an inner coiled coil and an outer ring of inverted helices) is more obvious than in HA. It is thought that the second motion repositions the transmembrane domain from the base of the molecule to its tip, that is, adjacent to the fusion peptide. It should be borne in mind, however, that the region adjacent to the transmembrane domain was not part of the structure, and thus the precise position of the transmembrane domain in the relaxed form remains to be determined. In addition to HA, the crystal structures of several additional class I proteins have been determined, but for these only the relaxed structures are available. They all contain a central trimeric coiled coil and three surrounding outer helices although there is no sequence similarity between most of these proteins.

4.2 Fusion Mechanism of Class I Proteins

Despite a wealth of structural information and numerous studies on fusions mediated by wild-type and mutated class I fusion proteins, there is still no consensus about the mechanism by which these proteins fuse membranes. It is undisputed that any interference with the conformational change (or part of it) either inhibits fusion completely or leads to an arrested intermediate state, suggesting that these fusion proteins are nanomachines that exert mechanical force on the membranes destined to fuse. The most popular model ("jackknife model") envisions fusion as a two-step process progressing via defined intermediates (Fig. 3). In the first step, the fusion peptide is propelled at the tip of the protein, resulting in its insertion and thus anchoring in the host cell membrane. In this state, the fusion protein is in an extended conformation, being connected to the viral membrane by its transmembrane domain and to the host cell membrane by its fusion peptide. In the second step, the

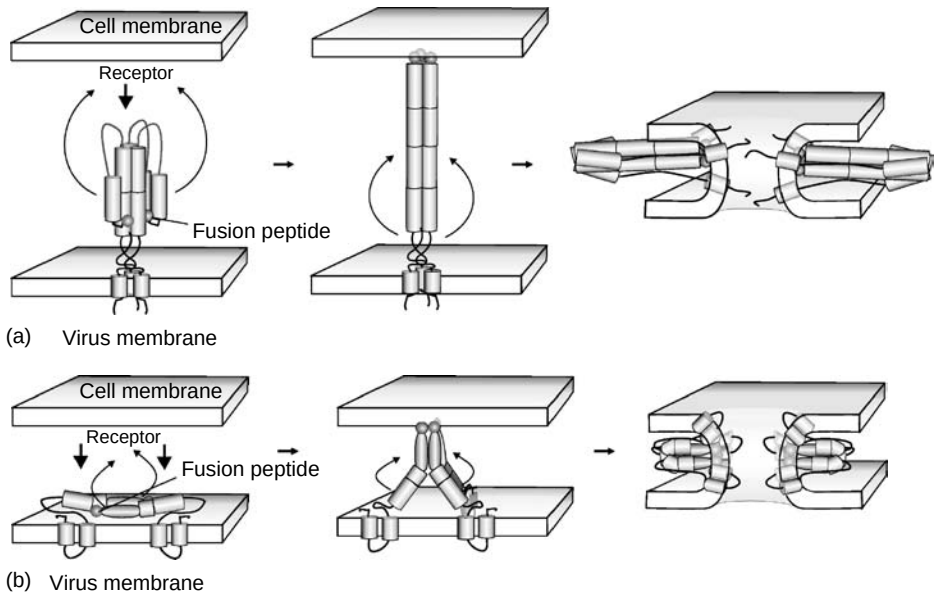

Fig. 3 Models for membrane fusion mediated by class I (a) and class II (b) viral fusion proteins involving a two-step mechanism. In the first step, the fusion peptides are exposed and are inserted in the target membrane. In the second step, the C-terminal part of the proteins undergoes another conformational change, resulting in bending/pulling that results in fusion. Red balls depict the fusion peptides, green cylinders the transmembrane domains. Note that the cellular receptors and the viral receptor–binding proteins are not shown. The colors indicate different domain of the proteins. Note that in class I proteins the colored boxes indicate α-helices, whereas in class II proteins they define folded domains that are not helical. Arrows indicate the presumed movements (see color plate p. xxiii).

outer helices revert and fold back on the rigid coiled coil, thus dragging the transmembrane domains from the base to the top of the molecule adjacent to the fusion peptide. As a result, the two membranes are forced into close apposition, with the fusion peptides probably destabilizing the monolayer surface by displacing water at the contact site, facilitating the formation of nonbilayer transition states.

Although widely accepted, the jackknife model suffers from some shortcomings. First, the rotational symmetry of the trimeric fusion proteins (see cartoon in Fig. 3) precludes bending to one side (as mandatory for the jackknife model) unless one assumes that the second folding step occurs asynchronically, that is, that back folding in only one of the subunits allows for lateral bending of the molecule. Second, it has been known for some time that activation of HA in the absence of a target membrane results in insertion of the fusion peptide into the viral membrane. Again, it is difficult to explain how such ''reverse'' insertion is prevented upon contact with a host membrane. To overcome these problems, alternative mechanisms were proposed that differ from the jackknife model in that the fusion peptide is either exclusively or partially inserted into the viral membrane. According to the first scenario, the formation of the coiled coil

would then exert a strong bending force on the viral membrane, while the protein remains attached to the cell membrane by binding to the receptor. Fusion proteins grouped in a ring around the prospective fusion site would create a membrane protrusion that extends in the direction of the target membrane, resulting in its destabilization and initiation of fusion. A second alternative is that the fusion peptides of a given trimer are inserted randomly into the viral and the target membrane. In this scenario, the extension of the central coiled coil rather than the backfolding would exert the primary pulling force that brings the membranes in close apposition. While both of the latter models account for the "symmetry problem," experimental support for either of them presently is lacking.

4.3 Class II Viral Fusion Proteins

Class II fusion proteins including the E-glycoproteins of flaviviruses and alphaviruses are also activated by low pH to undergo conformational changes, but it appears that the mechanisms are very different from those of class I proteins. Recently, the structure of whole virus particles containing class II fusion proteins has been solved by a combination of cryoelectron microscopy and crystallography. These fusion proteins form an interconnected lattice on the viral surface. They are oriented horizontally rather than perpendicular to the viral membrane. Proteolytic activation usually occurs in an accessory protein rather than in the fusion protein itself. α-helices and coiled coils are not part of the protein structure, rather, they are primarily composed of β-strands. The E-glycoproteins have three domains: an N-terminal domain, a middle domain that carries the fusion peptide (rather than being positioned near the N-terminus), and a C-terminal domain containing the transmembrane domain. The fusion peptides form short loops, which, in the nonactivated form are located close to the viral membrane and are covered either by an accessory protein involved in receptor binding or by a dimerization interface. Upon activation, the fusion peptides are exposed by what appears to be a coordinated conformational change of neighboring molecules. In the case of the Semliki forest virus, activation reorients the E1-fusion protein from a dimer with its accessory protein E2 to a homotrimer, associated with a movement away from the viral membrane surface (Fig. 3). This motion probably causes the insertion of at least part of the fusion peptides into the target membrane, probably in a β-barrel conformation. *In vitro* fusion of viruses carrying class II proteins with artificial membranes is considerably faster than that of viruses with class I proteins.

Despite the differences in structure, recently solved crystal structures of class II fusion proteins have revealed striking similarities between class I and class II fusion proteins. In both cases, activation leads to conformational rearrangements that result in a juxtapositioning of the transmembrane domain with the fusion peptide, with the final conformation being very stable. In the relaxed low pH trimeric form, domains 1 and 2 form a thick, rodlike structure with the fusion peptides at their tips, again reminiscent of the class I protein structure. Thus, it is conceivable that despite a very different structural basis the fusion mechanisms of class I and class II proteins are similar (Fig. 3).

5 Mitochondrial Fusion

Mitochondria continuously undergo dynamic shape changes including frequent fusion and fission events. Since mitochondria are surrounded by an inner and an outer membrane with very different protein compositions, homotypic fusion of mitochondria must involve the consecutive fusion of the outer and inner membrane, respectively.

Until recently, the protein machinery involved in mitochondrial fusion has remained enigmatic although it is clearly distinct from the protein complexes mediating fusions in the secretory pathway. A few years ago, a first putative protein participating in mitochondrial fusion has been discovered. Originally identified as a gene involved in the formation of a special mitochondrial aggregate in the sperm of Drosophila, a conserved integral membrane protein, termed fzo (for "fuzzy onion"), was discovered to be essential for fusion. Fzo represents a small protein family of evolutionary highly conserved proteins (termed *mitofusins* in mammals) that possess two transmembrane domains and an N-terminal GTPase domain, with both the C- and N-terminus facing the cytoplasm. Adjacent to the membrane-anchor domain are regions that are predicted to have a high propensity for the formation of coiled coils, thus bearing superficial similarities with bona fide fusion proteins such as the class I viral fusion proteins and the SNAREs. While yeast possesses a single family member (Fzo1p), higher organisms and mammals contain two mitofusins.

Recently it has been discovered that mitofusins can form dimers in which the C-terminal regions of mitofusins form antiparallel coiled coils, thus tethering the outer membranes of two adjacent mitochondria. Furthermore, an active GTPase domain is required for mitochondria to fuse. However, it is not clear at present whether the role of mitofusins is confined to tethering or whether they are also involved in later states of the fusion reaction. Additional proteins are likely to be involved including, for example, Ugo1p, an outer membrane protein with several transmembrane domains, and Mgm1p, a dynamin-like GTPase on the outer surface of the inner membrane. Furthermore, recent screens in yeast have uncovered more genes that appear to be involved in mitochondrial fusion and whose molecular roles are yet to be elucidated.

6 Fusion Reactions in the Secretory Pathway

6.1 Overview

The term *secretory pathway* refers to all organelles that originate from the endoplasmic reticulum and that are connected by defined vesicular trafficking routes. They include both the biosynthetic route, leading from the endoplasmic reticulum to the Golgi complex, and from there lead either to the plasma membrane or to endosomal intermediates, and the endocytotic/degradative route leading from the plasma membrane to the lysosomes via early and late endosomal intermediates. Additional connections and shortcuts are provided by further specialized trafficking routes, some of which may be tissue and cell-type specific such as regulated exocytosis (Fig. 4).

Each trafficking step in the secretory pathway can be broken down into three

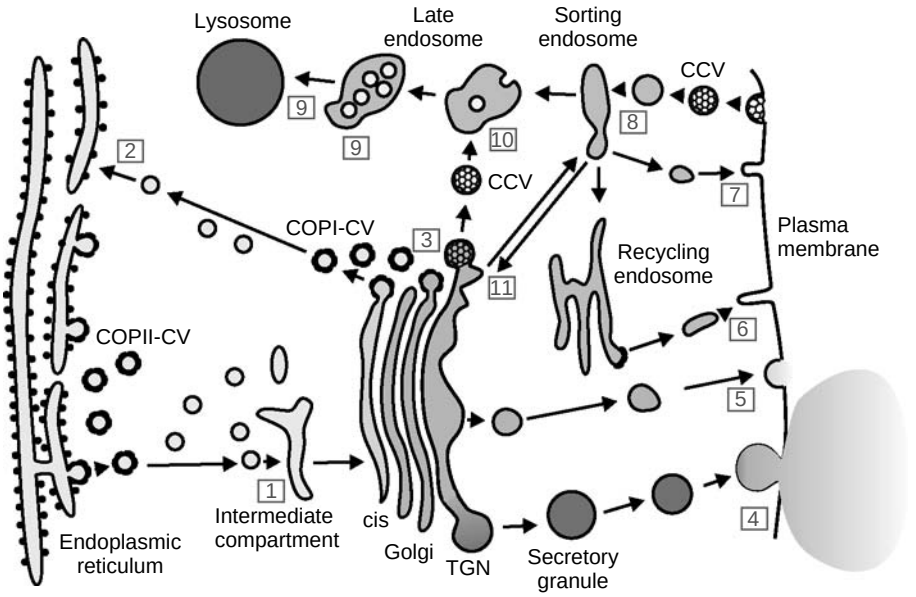

Fig. 4 Overview of the secretory pathway in a prototype eukaryotic cell. The boxed numbers indicate distinct fusion steps. CCV, clathrin-coated vesicle, COPI/II-CV, COPI/II-coated vesicle, TGN, *trans*-Golgi network. The biosynthetic route is depicted in green, the endocytotic pathway in pink (see color plate p. xxv).

elementary steps: the generation of a transport vesicle by budding from a precursor membrane, the movement of the vesicle to its target membrane mediated by the cytoskeletal machinery, and the consumption of the transport vesicle by fusion with its target compartment. In addition, there is ''homotypic'' fusion between organelles of the same class such as endosomes and cisternae of the endoplasmic reticulum. Intracellular trafficking pathways are highly organized, and they exhibit a high (but not absolute) degree of specificity. Thus, a vesicle derived from the endoplasmic reticulum will normally not fuse with the plasma membrane. Furthermore, vesicles originating from common sorting platforms such as the *trans*-Golgi network or the early endosome may have different destinations. This specificity is a feature of the organelles and trafficking vesicles, and as discussed in more detail below it is encoded in specific proteins that provide layers of regulation.

Unlike viral fusion proteins that single handedly carry out the jobs of docking and fusion, complex supramolecular assemblies mediate fusion in the secretory pathway, which act in concert with each other and carry out cascades of sequential reactions. These supramolecular machines are not represented by stable entities with fixed stoichiometries. Rather, they are assembled on demand and dissociate from the fusion site when the task is completed.

Despite the diversity of intracellular fusion reactions with respect to structure and kinetics, it appears that the fusion machines of the secretory pathway operate by principally similar mechanisms. This is suggested by the fact that several of the

proteins required for docking and fusion represent evolutionarily conserved protein families that have diversified to adjust to the specific demands of individual fusion reactions. Such conserved proteins include the small GTPases of the Ypt/Rab family, the SM proteins, and the SNARE proteins.

Before intracellular trafficking organelles fuse, they need to be transported into close proximity. Furthermore, signaling between the membranes is required in order for them to "know" that another potential fusion partner is nearby. This also involves "proofreading," that is, mechanisms that ensure that the fusion partner is appropriate. These steps may or may not be associated with the establishment of physical contact that proceeds via first loose ("tethering") and then firm ("docking") contact. Next, the fusion reaction is executed which may require prior activation of the fusion machinery. After fusion, the fusion machinery is disassembled and regenerated to be available for another round of fusion.

Intriguingly, it appears that with exception of the presumed fusion catalysts (the SNARE proteins) and some regulatory proteins involved in specialized fusion reactions (such as the synaptotagmins), all other proteins are recruited on demand from the cytoplasm and thus are not permanently associated with the membrane. The advantage of such a mechanism is probably to minimize the "efforts" a cell needs to spend on protein recycling. After fusion of a trafficking vesicle, membrane-resident fusion proteins need to be returned by membrane trafficking to the precursor compartment for the proteins to be reused. This necessity also explains why membrane-bound fusion proteins (SNAREs) are not only present in the organelle for whose fusion they are responsible but also in all other organelles belonging to the recycling pathway required for returning these proteins to their original compartment.

6.2 Tethering and Docking – the Role of Rab Proteins and Their Effectors

What is the first step in the reaction cascade leading to fusion of two proximal organelles? Very little is known about the initial signaling events. However, it is well established that Rab GTPases play a central role in defining a site for fusion and in orchestrating the recruitment of protein complexes.

Rab proteins belong to the ras-superfamily of small monomeric GTPases. They operate as molecular switches that exist in an active GTP-bound and an inactive GDP-bound conformation. The switch between these conformations is made by regulatory proteins, the GTPase-activating proteins (GAPs) that stimulate the intrinsic GTPase activity of the rabs, and the guanine nucleotide exchange factors (GEFs) that replace GDP with GTP. Superimposed on the GTP-GDP cycle is a membrane association–dissociation cycle. Rab proteins carry posttranslationally attached polyprenyl chains that anchor the proteins in the membrane. GDP-Rab proteins are removed from the membrane by the action of a universal protein termed GDP-dissociation inhibitor (GDI) that forms soluble complexes with all GDP-Rabs by hiding the prenyl anchor in a hydrophobic pocket. Rebinding is associated with GTP-reloading by GEFs and may require assistance of additional proteins such as the members of the Yip/Pra1 family. These proteins displace GDI and thus are termed GDFs (for GDI displacement factors).

Rab proteins display the most organelle-specific localization of all known trafficking proteins, and they have therefore been assigned a major role in defining the identity of organelles. More than 60 Rab proteins are known in mammalian species. Individual organelles may contain multiple Rabs that probably define functionally distinct subdomains on the organellar surface. However, it is still not understood by which mechanisms this exquisite specificity is achieved.

The main job of active Rab proteins is to recruit specific proteins (termed Rab effectors) to defined sites at the membrane. While many effectors are exquisitely specific for a given Rab protein, others are shared between multiple rabs. Examples for the latter are rabaptin, the TRAPP complex, and mss4. The number of known Rab effectors is increasing steadily, with individual Rab proteins known to bind to dozens of different proteins. Presently, there are no common denominators between the many prospective Rab effectors. However, in some well worked out cases such as the homotypic fusion of early endosomes it is becoming clear that Rab effectors orchestrate the initial steps in docking and fusion. Effectors of the early endosomal Rab5 include large proteins such as EEA1, which is capable of forming homooligomeric structures, tethering prospective fusion partners. Other effectors include a phosphatidylinositol 3-kinase and proteins containing FYVE and PX domains that in turn bind to specific polyphosphoinositides. Phosphoinositides are known to be centrally involved in intracellular membrane traffic, and it is possible that together with an active Rab they provide a local binding site for further proteins that is highly specific for a given fusion step. For instance, EEA1 is both an effector for Rab5 and it contains a FYVE domain that binds to phosphatidylinositol-3-phosphate, suggesting that only the combination of both – active Rab5 and a sufficiently high concentration of phosphatidylinositol-3-phosphate – is effective in recruiting EEA1 to the membrane.

While it is generally agreed upon that active Rabs recruit docking and tethering complexes, the underlying molecular mechanisms are not known. Many Rab effectors are multiprotein complexes including the exocyst (presumably involved in constitutive exocytosis), the GARP, TRAPPI, TRAPPII, and COG complexes that are involved in trafficking steps at the Golgi apparatus, and the HOPS/VpsC-complex that functions in the homotypic fusion of yeast vacuoles. For the homotypic fusion of yeast vacuoles, sequentially occurring assembly and disassembly reactions were worked out involving components of the VpsC-complex and additional proteins such as LMA1. An example from the other end of the spectrum is represented by the RIM proteins, effectors of the Rab3 that function in neurotransmitter release at the synapse. RIM is part of the so-called active zone, an electron-dense structure at the presynaptic plasma membrane to which synaptic vesicles dock. Active zones consist of an insoluble supramolecular complex containing a set of large and insoluble proteins including Piccolo/aczonin, bassoon, ERCs (acronym for ELKS/Rab3-interacting molecule/CAST), RIM-binding proteins 1 and 2, and α-liprins. Taken together, there is little similarity between the components of these multiprotein complexes, and it is thus conceivable that the diversity of Rab effectors is responsible for imprinting specific features to each fusion reaction in the secretory pathway and to specialized fusions in certain cell types.

6.3 Fusion – SNARE Proteins as Fusion Catalysts

SNARE proteins (acronym for soluble NSF acceptor protein receptors) represent a superfamily of small and mostly membrane-bound proteins that are conserved throughout the eukaryotic kingdom, with 24 members in yeast, at least 35 members in mammals, and more than 50 in plants. SNARE proteins typically contain a C-terminal transmembrane domain, an adjacent stretch of 60–70 amino acids arranged in heptad repeats that is characteristic for the entire SNARE superfamily and that is referred to as *SNARE motif*, and an N-terminal region that either consists of only few amino acids or else, is composed of a separately folded domain (Fig. 5). There is considerable heterogeneity among SNARE family members both with respect to primary structure and to domain structure. In addition to the prototypical SNARE described above, some family members contain two instead of one SNARE motifs. Examples include the neuronal SNARE SNAP-25, its relatives SNAP-29 and SNAP-23, and the yeast SNAREs Sec9p and Spo20. In these proteins, the SNARE motifs are connected with a linker region. These SNAREs lack a transmembrane domain; instead, some of them carry hydrophobic palmitoyl side chains. Other exceptions include SNAREs with single SNARE motifs that carry either a posttranslationally added polyprenyl anchor (such as Ykt6) and the yeast SNARE Vam7 that lacks a membrane anchor altogether.

The importance of SNARE proteins was originally recognized because of three independent lines of evidence: first, the SNAREs involved in neuronal exocytosis (including the proteins synaptobrevin/VAMP, SNAP-25, and syntaxin 1) were identified as the targets of botulinum and tetanus toxins, second, the proteins were recognized in genetic screens in yeast as being essential for intracellular fusion reactions, and third, the SNAREs were identified as the membrane receptor

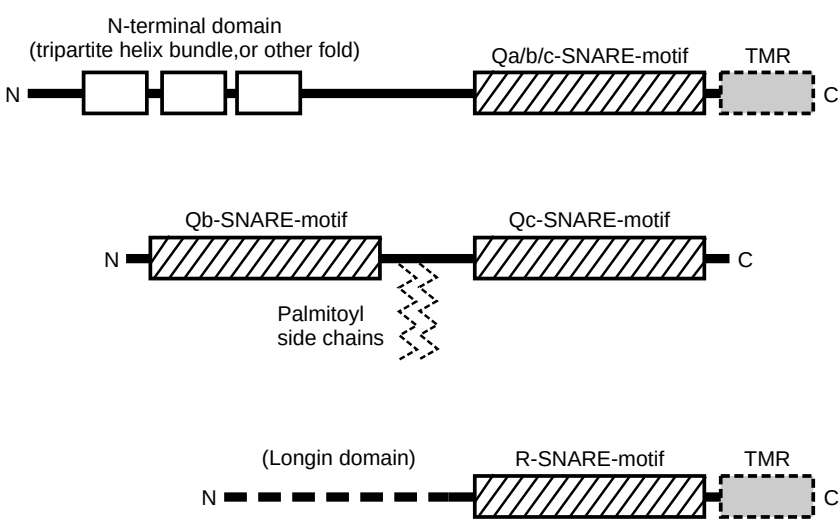

Fig. 5 Domain structures of SNARE proteins. Dotted lines delineate domains that may be absent in some of the family members. TMR, transmembrane region.

of the ATPase NSF (hence the name SNARE) that was previously recognized as an important component of intracellular fusion reactions.

6.3.1 The SNARE Assembly–Disassembly Cycle

SNARE proteins undergo an assembly–disassembly cycle that involves the SNARE motifs and that is essential for membrane fusion. Monomeric SNARE motifs are unstructured in solution. However, when appropriate sets of SNAREs are combined, they spontaneously form complexes in which four different SNARE motifs form a stable helical bundle. The crystal structures of two only distantly related SNARE complexes revealed an extraordinary degree of structural conservation despite limited sequence homology. The structures are represented by four elongated α-helices that form a coiled coil containing 16 layers of interacting amino acid side chains stacked on top of each other. These amino acids are mostly hydrophobic with the exception of an unusual layer in the middle of the complex, termed "0" layer, to which three SNARE motifs each contribute a glutamine, and one an arginine (Fig. 6). On the basis of this highly conserved layer, SNAREs are now classified into subfamilies termed

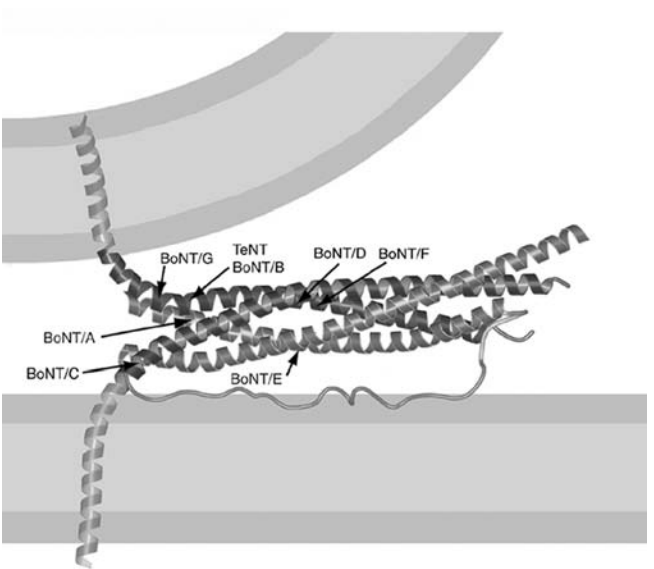

(a)

Fig. 6 (a) Crystal structure of the neuronal SNARE complex containing syntaxin 1 (red), SNAP-25 (green), and synaptobrevin/VAMP (blue). The structure is modeled between two membranes, the transmembrane domains and immediately adjacent residues are not part of the structure. (a) Structure of the ionic "0" layer that is conserved among all SNAREs and that is the basis for the classification of SNAREs into Q- and R-SNAREs, respectively. Modified from Sutton, B., Fasshauer, D., Jahn, R., Brünger, A.T. (1998) Crystal structure of a core synaptic fusion complex at 2.4 Å resolution, *Nature* **395**, 347–353 (see color plate p. xxviii).

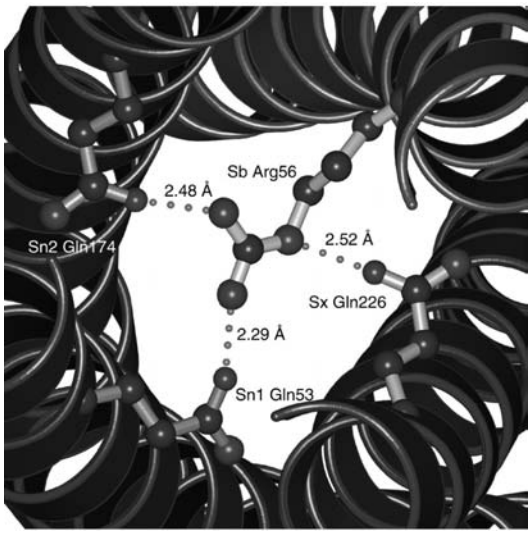

(b)

Fig. 6 (*Continued*)

Q-SNAREs and R-SNAREs, with the Q-SNAREs being further subdivided into Qa-SNAREs (representing the "true" syntaxins), Qb-SNAREs, Qc-SNAREs (homologs of the two SNARE motifs of SNAP-25, respectively), and R-SNAREs (relatives of VAMP/synaptobrevin). The division of SNAREs into these four subfamilies is strongly supported by additional similarities that were derived from sequence comparisons. As a general rule, SNARE complexes can only form if one SNARE motif of each subfamily is contributed. It replaces the previous classification of SNAREs to v-SNAREs (vesicular SNAREs) and t-SNAREs (SNAREs residing in target membranes) since the intracellular topology that was used for this classification does not agree with the subclasses as defined by activity, crystal structures, and sequence alignment. In fact, *in vitro* SNARE motifs of the appropriate subclasses can – within certain limits – substitute for each other.

Although due to an unusual hysteresis, it is not possible to determine the free energy of SNARE assembly, assembly proceeds downhill an energy gradient. The stability of assembled SNARE complexes is remarkable, requiring temperatures above 80 °C or very strong denaturants such as 5 M guanidinium chloride for dissociation. Both the elongated structure of the helical bundle and the stability toward heat, denaturants, and proteases is strikingly parallel to the relaxed conformation of class I viral fusion proteins discussed above. However, unlike the viral fusion proteins, SNAREs are reactivated by disassembly, involving the ATPase NSF (acronym for NEM-sensitive factor). On its own, however, NSF cannot interact with SNAREs; it requires cofactors termed SNAPs (acronym for soluble NSF attachment proteins). Both in yeast and in mammals only one NSF gene has been found. In contrast, mammalian species contain three variants of SNAP termed α-, β-, and γ-SNAP respectively,

with α-SNAP being the ubiquitous isoform. Three SNAPs bind to one SNARE complex, resulting in the recruitment of NSF. Remarkably, the SNAP-NSF system interacts with all known SNARE complexes in this manner despite the fact that the surface amino acid composition is highly variable with respect to charge and size. The crystal structure of the yeast SNAP-homolog Sec17p revealed a rigidly packed protein with an N-terminal region in which nine α-helices are packed against each other in an antiparallel fashion, resulting in a bent sheet and a C-terminal globular region. The N-terminal region is responsible for SNARE binding, whereas the C-terminal region recruits NSF. Several SNAP mutants are known, which are defective in binding and/or activation of NSF, they are strong inhibitors of all fusion reactions of the secretory pathway.

NSF is a hexameric ATPase that belongs to the AAA superfamily (ATPases associated with other activities). AAAs frequently operate as unfoldases, that is, they dissociate tightly packed or aggregated proteins. NSF monomers consist of three domains: an N-terminal region that binds to the substrate and that undergoes large conformational changes upon ATP hydrolysis, and two homologous ATP binding regions termed D1 and D2, with only D1 being catalytically active and D2 being responsible for hexamer formation. Single-particle and cryoelectron microscopy as well as crystallographic studies on NSF domains and on VCP (valosin-containing protein), a relative of NSF, have revealed that the protein is represented by a double barrel structure with a conspicuous hole in the middle. Presently it is unknown how AAAs manage to disentangle proteins as tightly packed and stable as SNARE complexes. It has been proposed that unfolding involves threading of the substrate through the central channel (resembling some chaperones). Alternatively, it is possible that the barrel opens sideways (''clamp loader'') thus getting a ''grip'' on the substrate and exposing it to hydrophobic side chains in the interior of the oligomeric molecule. The ATP-requirement has not yet been precisely determined, and thus it is not known whether a full catalytic cycle of all six subunits is required for disassembly. Mutations of NSF are known both in yeast and in Drosophila, which loose activity at mildly elevated temperature, resulting in a general impairment of membrane traffic. In Drosophila, the dominant effect is in synapses where exocytotic release of neurotransmitters is affected first, resulting in paralysis (''comatose'') and an accumulation of SNARE complexes.

6.3.2 Mechanism of SNARE-mediated Membrane Fusion

The most widely accepted model describing how SNAREs mediate membrane fusion implies that it is the spontaneous assembly of SNAREs into SNARE complexes that drives the membrane merger. According to this model, each of the membranes destined to fuse contains complementary sets of SNAREs that upon close contact spontaneously interact with each other in ''trans'' and form a SNARE complex by zippering up from the N- toward the C-terminal ends of the respective SNARE motifs. Such zippering would force the transmembrane domains in close apposition. The energy being released during assembly is used to overcome the repulsive energy barrier separating the membranes, thus exerting strain on the membranes. During fusion, the SNAREs change from a strained ''trans'' into a relaxed ''cis'' conformation, in a remarkable parallelism to the class I viral fusion proteins. After fusion is complete, the NSF-SNAP

disassembly system then dissociates the complex and makes available the SNAREs for another round of fusion.

This model, although not undisputed, is supported by a large body of evidence such as

- experiments using yeast vacuolar fusion *in vitro* have shown that NSF is not required for the fusion reaction but rather is needed for a preceding activation step;
- many SNARE mutants from yeast, flies, and worms causing functional impairments map to amino acids of the central interacting layers, supporting the view that the formation of the helical bundle is critical for fusion;
- cells can be enticed to fuse when SNAREs are "flipped," that is, expressed by facing the exterior rather than the cytoplasm;
- any interference with SNAREs in intact cells or in cell-free assays using antibodies, clostridial neurotoxins, or recombinant SNARE motifs as competitors invariably inhibits fusion;
- proteoliposomes reconstituted with appropriate sets of SNAREs spontaneously fuse with each other, and this fusion requires the formation of SNARE complexes.

6.3.3 Conformational Intermediates of SNAREs

Recent studies have shown that the zippering model described above, while providing a useful overall description of the SNARE mechanism, is an oversimplification that needs to be refined. For instance, while SNAREs spontaneously form complexes and fuse liposomes *in vitro*, these reactions are many orders of magnitude slower than their biological counterparts, although high protein concentrations were used in the assays, and artificial docking of the liposomes also does not accelerate the reaction. Similar to viral fusion proteins, the attention has therefore focused on conformational intermediates of the SNARE cycle.

A model outlining potential intermediate steps is shown in Fig. 7. This model is primarily based on studies of the SNARE proteins involved in neuronal exocytosis that served as paradigms. It is possible that certain depicted states do not exist in other SNARE complexes. The following features are noteworthy:

- *Hotspots of concentrated clusters of SNAREs.* In plasma membranes, SNAREs are not uniformly distributed across the membrane but rather concentrated in microdomains. These microdomains are dependent on the presence of cholesterol in the membrane but are not resistant to detergent, that is, they are different from rafts. The local concentration of SNAREs in these hotspots appears to be very high. This is particularly noteworthy in neurons where the two Q-SNAREs syntaxin 1 and SNAP-25 each contribute at least 1% of total brain protein, making them by far the most abundant membrane proteins in the nervous system. In PC12 cells, secretory granules exclusively undergo exocytosis on such hotspots suggesting that for fusion to be efficient many SNARE complexes need to cooperate. In addition, the SNAREs in these hotspots readily form complexes with both exogenous and endogenous SNAREs suggesting that activation of SNAREs prior to fusion is not rate limiting.
- *SNARE acceptor complexes.* Considering that the SNARE motifs are unstructured in solution, it is not surprising that SNARE complex formation *in vitro*

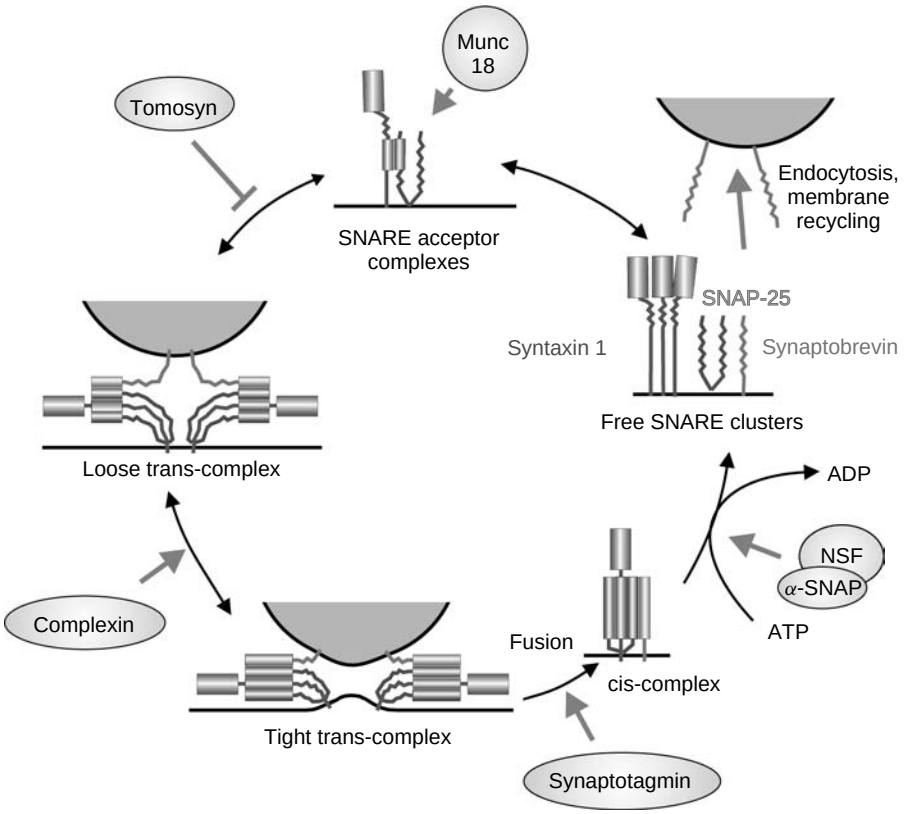

Fig. 7 Model of the cycle of the neuronal SNAREs and its regulation by accessory proteins. See text for details.

is exceedingly slow as it requires the simultaneous association of four different molecules. *In vitro*, the neuronal Q-SNAREs syntaxin 1 and SNAP-25 form a binary complex consisting of two syntaxin and one SNAP-25 molecule. This complex resembles the SNARE core complex but is less stable. For synaptobrevin to bind, one of the syntaxins needs to be displaced which appears to be rate limiting. Recent work revealed that an unstable intermediate is formed between SNAP-25 and a single syntaxin molecule. Binding of synaptobrevin to this complex appears to be considerably faster. This intermediate may represent the "true" acceptor complex that can either recruit a second syntaxin, resulting in a nonconstructive "dead-end" reaction, or bind synaptobrevin in a pathway leading to fusion. Intriguingly, independent lines of evidence suggest that such acceptor complexes are stabilized by SM proteins (see below).

- *SNARE trans-complexes*. Regulated exocytosis implies that the pathway leading to fusion is arrested at some step, requiring activation by a trigger that is usually an elevation of intracellular Ca^{2+}. In neuronal exocytosis, the delay time between Ca^{2+} triggering and membrane fusion is extremely short (<1 ms). The

question arises whether SNARE assembly can be arrested at an intermediate state, with the N-terminal parts of the SNARE motifs being already assembled, whereas the C-terminal parts are still apart. Evidence for such intermediates primarily stems from studies of exocytosis of chromaffin cells, which can be monitored with high time resolution using electrophysiological techniques. Following a jump in intracellular Ca^{2+}, distinct pools of granules are observed that undergo exocytosis with a characteristic time constant: a cytoplasmic depot pool, a reserve pool representing membrane-attached vesicles that still need to be activated ("primed"), and two pools of release-ready vesicles, referred to as slowly releasable pool (SRP) and rapidly releasable pool (RRP). These pools are in dynamic equilibria with each other. Recent studies suggested that in the RRP, SNAREs are associated with each other in a trans-complex prior to fusion, and similar conclusions have also been reached from studies at the crayfish neuromuscular junction. While all these studies are indirect (i.e. conclusions about molecular states are drawn from kinetic data), they suggest that metastable folding intermediates may exist in the SNARE assembly pathway, again paralleling similar observations on class I viral fusion proteins.

- *Contribution of SNARE transmembrane domains.* For the final step in membrane fusion, the model predicts that a pulling force is exerted on the membranes, suggesting that the linkers between the transmembrane domains and the SNARE motifs are stiff. Both theoretical calculations and experimental evidence suggest that this may be the case. Furthermore, it is possible that the transmembrane domains do more than just providing a membrane anchor. For instance, peptides corresponding to SNARE transmembrane domains have recently been shown to be capable of fusing liposomes, suggesting that they may also be involved in destabilizing the bilayer. Furthermore, recent studies indicate that the transmembrane domains of syntaxin may contribute to the fusion pore structure. Also, transmembrane domains of syntaxin and synaptobrevin have the capacity to form homooligomers and heterooligomers, again supporting the view that in addition to providing membrane attachment they may participate in transition states in a specific manner.

6.3.4 Topology and Specificity of SNAREs in Intracellular Fusion Reactions

The model outlined above for SNARE function implies that of the fusing membranes each one carries at least one of a complementary set of SNAREs that contains a transmembrane domain. In neuronal exocytosis this requirement is fulfilled, with synaptobrevin (R-SNARE) being localized primarily to the synaptic vesicle and syntaxin 1 (Qa-SNARE) to the plasma membrane. Intracellular fusion reactions, however, often involve SNARE complexes in which three or all four SNARE motifs are contributed by SNAREs possessing a transmembrane domain, allowing, at least in theory, variable distributions between the membranes. However, it is conceivable that the need for preformation of acceptor complexes limits the choices, for example, if, as in the case of neuronal exocytosis, an acceptor complex consists of Qa/Qb/Qc-SNARE motifs. Indeed, liposome fusion experiments suggested that only one type of topology of a given SNARE complex results in fusion. However, in yeast there are examples of SNAREs that

may function in different complexes with different topology. For instance, the yeast R-SNARE Sec22p functions both in anterograde as well as in retrograde traffic between the endoplasmic reticulum and the Golgi apparatus. In anterograde traffic, Sec22p appears to be colocalized with the Qb- and Qc-SNAREs Bos1p and Bet1p respectively, on the transport vesicle, with the Qa-SNARE Sed5p residing in the acceptor membrane. In contrast, all three Q-SNAREs for retrograde traffic including Qa (Ufe1p), Qb (Sec20p), and Qc (Use1p) are localized to the endoplasmic reticulum, with Sec22p being the sole functional SNARE on the retrograde transport vesicle.

The example of yeast anterograde and retrograde traffic between the endoplasmic reticulum and the Golgi apparatus also raises the question to which extent SNARE pairing is specific. Although this issue continues to be controversially discussed, the following points can be made:

- *In vitro*, SNAREs can be assembled into SNARE complexes rather promiscuously as long as the QabcR-SNARE rule for SNARE complexes is obeyed. However, while in some experiments no differences were observed between different R- or Q-SNAREs, respectively, it is likely that promiscuity is not absolute, that is, that not every SNARE of a given subfamily is capable of substituting for a particular family member in a given complex.
- Certain SNAREs are known to be involved in different trafficking steps involving different SNARE partners. In addition to Sec22p discussed above, other yeast examples include Sed5p (a Qa-SNARE operating in ER to Golgi, and intra Golgi trafficking steps), Vti1p (a Qb-SNARE operating in traffic to the *cis*-Golgi, and from the Golgi to the prevacuole (late endosome) and the vacuole compartment), and Ykt6p (an R-SNARE operating in retrograde traffic to the Golgic, in traffic from the Golgi to the prevacuole (late endosome) and the vacuole compartment, and in anterograde traffic from the ER to the Golgi).
- Conversely, in a given SNARE complex certain SNAREs can functionally substitute for each other, albeit the efficiency appears to be lower than with the "cognate" SNARE. For instance, in yeast Ykt6p can substitute for Sec22p in anterograde but not retrograde ER to Golgi traffic, and SNAP-23 can substitute for SNAP-25 in exocytosis of chromaffin granules. Such functional substitutions of one SNARE for another one may explain why genetic deletion of individual SNAREs in some cases is known to yield surprisingly mild phenotypes.

The question then arises to which extent SNARE complexes are specific for a given fusion step. Although a definitive answer cannot be given at this time, it appears that the SNAREs active in a specific intracellular fusion event are precisely regulated, that is, a given SNARE complex functions only in one defined step even if some flexibility is allowed in the identity of some of its participating SNARE proteins. It should be noted, however, that presently it cannot be excluded that more than one SNARE complex operates in parallel in certain fusion events.

It is not known how nonspecific pairing (as observed *in vitro*) is prevented. Furthermore, it is not known by which mechanism the correct SNAREs are selected for a given fusion step. As integral membrane proteins, SNAREs need to recycle by membrane traffic to their resident

membrane after fusion, explaining the widespread intracellular distribution of many SNAREs. Probably every trafficking vesicle carries "passenger" SNAREs in addition to the ones needed for the forthcoming fusion step. For instance, in neurons and neuroendocrine cells the recycling pathway of synaptic vesicles includes at least two fusion steps – exocytosis and fusion with an early endosomal intermediate that is reached after endocytosis via clathrin-coated vesicles (CCV). The SNAREs involved in these two fusion steps are distinct (exocytosis: syntaxin 1 (Qa), SNAP-25 (Qbc), synaptobrevin (Qc); endosome fusion: syntaxin 13 or 16 (Qa), Vti1a (Qb), syntaxin 6 (Qc), and VAMP4 (R)). Functionally, there appears to be no cross talk at least between the R-SNAREs as only exocytosis but not endosome fusion is inhibited by clostridial neurotoxins cleaving the R-SNARE synaptobrevin.

6.4 Regulators of SNAREs

An almost overwhelming panoply of proteins is reported to interact with SNAREs in a functionally relevant manner, either regulating fusion or else, being regulated by the SNAREs. In particular, the neuronal SNARE proteins were the target of intense investigations, with syntaxin 1 alone being thought to bind to more than 40 different proteins including many ion channels and receptors. However, in most of these cases the evidence is limited to qualitative binding studies, complemented with perturbation studies in intact cells, making it difficult to assess the specificity of the reported interactions. As discussed above, SNAREs can assume different conformations, and for proteins thought to specifically interact with SNAREs it is essential to determine where exactly these proteins act in the conformational cycle of SNAREs. While the discussion will be restricted here to those proteins for which experimentally supported molecular models are available, it needs to be borne in mind that this is an actively ongoing field of research, with additional proteins likely to be involved.

6.4.1 SM Proteins

SM proteins comprise a small group of conserved cytoplasmic proteins that, as far known, are as essential as SNAREs for all fusion steps of the secretory pathway. In fact, impairment or knock out of SM proteins leads to dramatic phenotypes. For instance, genetic deletion of Munc-18, the SM protein involved in neuronal exocytosis, leads to mice whose nervous system by and large develops normally but in which synapses are completely silent, that is, no exocytotic release of neurotransmitter occurs. There are fewer SM proteins than SNAREs, with four members in yeast and seven in mammals. Crystallographic studies on several SM proteins revealed a high degree of structural conservation, being represented by globular, arch-shaped molecules with a conspicuous cleft.

Numerous genetic and biochemical studies suggest that SM proteins regulate SNAREs. However, major confusion has arisen from the fact that despite a high degree of structural conservation among both SNAREs and SM proteins, the interaction between these proteins, at least *in vitro*, occurs by different mechanisms:

- Munc-18, the SM protein involved in neuronal exocytosis binds to the N-terminal region of the Qa-SNARE syntaxin 1 with high affinity. Like many other SNAREs, syntaxin 1 possesses an antiparallel three-helix bundle at its N-terminus that is connected to the

SNARE motif by a linker region. In a reversible intramolecular transition between an "open" and a "closed" conformation, this N-terminal domain can fold back onto the SNARE motif, preventing the latter from entering SNARE complexes. *In vitro*, Munc-18 exclusively binds to the closed conformation, effectively preventing it from interacting with its SNARE partners. Indeed, the crystal structure of the Munc-18/syntaxin complex reveals that both the N-terminal helix bundle and part of the SNARE motif are bound to the large cleft in the middle of Munc-18, with numerous crystal contacts between Munc-18 and both regions of syntaxin.

- The yeast SM proteins Sly1p and Vps45 (and probably also their mammalian counterparts) bind to the respective Qa-SNAREs Sed5p and Tlg2p in a completely different manner. Here, a short sequence at the N-terminal end of the Qa-SNAREs (i.e. beyond the folded part) binds to the outside of the SM protein. This interaction does not prevent formation of SNARE complexes; rather, it appears to facilitate them.
- In trafficking steps of the endocytotic pathway, the SM proteins do not appear to interact directly with SNAREs. Rather they form complexes with other proteins. Interestingly, this includes not only the SM protein Vps33p that is part of the VpsC or HOPS complex involved in the fusion of vacuoles but also Vps45p that forms a complex with Vac8p. Thus, one and the same SM protein may either bind directly to a SNARE, or else, operate via multiprotein complexes without displaying direct SNARE binding.

Recent evidence suggests that, at least in the case of Sly1p, direct binding to the SNARE is not required for function. Thus, at least some of the bindings described above may be recruiting mechanisms to ensure a sufficiently high concentration of SM proteins at the site of action rather than reflecting part of the mechanism these proteins are involved in, explaining the surprising structural diversity of the SM-SNARE complexes. How then are SM proteins regulating SNAREs? Several lines of evidence suggest that some SM proteins facilitate SNARE assembly, raising the possibility that the main job of SM proteins is to form and maintain SNARE acceptor complexes. Such acceptor complexes are probably unstable and thus hard to measure – the more stable they are, the less energy is left for the ensuing fusion reaction. Furthermore, it has recently been suggested that SM proteins may be part of a proofreading system that ensures pairing of appropriate SNAREs.

6.4.2 Synaptotagmin and Complexin – Regulators of Ca^{2+}-dependent Exocytosis

Exocytosis of synaptic vesicles represents a membrane fusion event that is tightly regulated by the intracellular Ca^{2+} concentration. Delay times between increases in intracellular Ca^{2+} and fusion are extremely short (below 1 ms), suggesting that the system is arrested at a very late step in the pathway of protein–protein interactions leading to fusion. The Ca^{2+} receptor responsible for triggering exocytosis was identified as synaptotagmin, an integral membrane protein of synaptic vesicles. Genetic deletion of synaptotagmin 1 leads to a massive loss of Ca^{2+}-stimulated exocytosis, whereas spontaneous transmitter release persists. Intriguingly, stimulation of exocytosis by means of bypassing the Ca^{2+} trigger (e.g. using the active ingredient of black widow spider venom) persisted but was sensitive to clostridial neurotoxins, documenting that SNARE function

in fusion was not affected in the deletion mutant.

Synaptotagmin is represented by a protein family of hitherto 15 members in mammals. Synaptotagmin 1 and 2 that function in neuronal exocytosis are type I transmembrane glycoproteins, that is, with the C-terminus facing the cytoplasm and the glycosylated N-terminus being exposed to the vesicle lumen. Characteristic for synaptotagmins are two tandem C2-domains (C2A and C2B) that represent Ca^{2+} binding modules binding 3 and 2 Ca^{2+} ions, respectively. Whereas the intrinsic Ca^{2+} affinity of both C2-domains is very low (close to 1 mM), it is increased by almost three orders of magnitude in the presence of acidic phospholipids. Structural studies revealed that the C2-domains provide only a partial coordinations sphere for Ca^{2+} ions that requires acidic phospholipids for completion. C2-domains are present in many additional proteins, and in some instances (e.g. protein kinase C) they are known to convey Ca^{2+}-dependent translocation of the protein to the membrane. Furthermore, Ca^{2+} binding is very rapid and it is not associated with a conformational change in the C2-domain.

While it is undisputed that Ca^{2+} binding to the C2-domains of synaptotagmin triggers exocytosis, the mechanism by which this occurs is still not clear. One possibility is that Ca^{2+}-dependent binding to phospholipids at a site where SNAREs are already forming trans-complexes suffices to further destabilize the membrane surface, leading to the formation of a fusion pore. Indeed, both mutant analysis and ion dependence suggest that phospholipid binding is critical for synaptotagmin function. On the other hand, both Ca^{2+}-dependent and Ca^{2+}-independent binding of synaptotagmin to SNAREs has been observed, raising the possibility that synaptotagmin directly acts upon the SNAREs, for example, by driving the assembly of a partially zippered trans-complex to completion. Intriguingly, Sr^{2+} ions, known to substitute for Ca^{2+} in stimulating exocytosis, promote only phospholipid but not SNARE binding of synaptotagmin. While this finding shows that Ca^{2+}-dependent phospholipid interaction is essential for function, it is conceivable that binding to SNAREs increases efficiency, for example, by positioning the C2-domains precisely before the arrival of the Ca^{2+} signal.

Intriguingly, a phenotype similar to synaptotagmin deletion albeit weaker was observed upon the deletion of complexins, two small and highly homologous isoforms that bind to the surface of assembled neuronal SNARE complexes. Structural studies revealed that part of the complexin forms a helix positioned in the groove formed by syntaxin I and synaptobrevin, which bends away from the SNARE complex toward the C-terminal transmembrane domains. Thus, it is possible that complexins bind and stabilize trans-SNARE complexes, maintaining them in a metastable state and allowing for synaptotagmin to act efficiently in promoting the reaction. Lack of complexins may thus result in less stable SNARE complexes (i.e. complexes that reversibly loosen and tighten), requiring more Ca^{2+} binding to synaptotagmin for efficient fusion.

It remains to be established whether similar regulatory mechanisms also apply to other SNARE-mediated fusion reactions. Many intracellular fusion reactions are known to require Ca^{2+} albeit at a concentration well below 1 μM, that is, the normal cytoplasmic concentration suffices for fusion to proceed. It is conceivable that other ubiquitously expressed synaptotagmin isoforms participate in

such intracellular fusion reactions. However, some synaptotagmins do not bind Ca^{2+} at all, raising the possibility that synaptotagmins may have other functions. Conversely, several additional synaptic proteins possess tandem C2-domains including the active zone components RIM (an effector of Rab3) and Piccolo/Aczonin.

6.4.3 "Pseudo" SNAREs

Two soluble proteins, termed *tomosyn* and *amisyn* respectively are known, which possess C-terminal R-SNARE motifs but otherwise have no similarities to SNAREs. Tomosyn is a large protein of approximately 110 kDa (SNAREs rarely exceed 35 kDa) whose large N-terminal region bears similarity to the lethal giant larvae proteins, whereas amisyn is represented by a smaller protein. Both proteins, when introduced into secretory cells, inhibit exocytosis. Detailed studies on the R-SNARE motif of tomosyn revealed that it directly competes with synaptobrevin in the formation of SNARE complexes, thus reducing the number of active acceptor complexes that are available for fusion. Presently, it cannot be decided whether competition by soluble SNARE motifs (resulting in the formation of nonconstructive cis-complexes) is a general physiological means to downregulate SNARE complexes or whether these proteins are specialized for neuronal exocytosis.

7 Conclusion and Outlook

Despite the diversity of fusion proteins, the examples discussed above show that there appear to be some common denominators between viral fusion proteins and SNAREs. In both cases, fusion is associated with a spontaneous conformational change that involves movement of membrane-anchor domains (transmembrane domains or fusion peptides) over large distances. At the end of the conformational change, an elongated and stiff rodlike structure is generated in which all membrane-anchor domains are localized at one end. Furthermore, both viral fusion proteins and SNAREs are oligomers that exhibit at least some rotational symmetry (trimers in the case of viral fusion proteins, trimers and tetramers in the case of SNAREs). These intriguing features suggest that the basic mechanism of operation is principally similar in all cases. However, there are still many unresolved issues. For instance, it is unknown whether all fusion reactions transit through hemifusion intermediates. Furthermore, it remains to be established whether the fusion proteins are either part of the initial fusion pores or at least directly influence their diameter and enlargement kinetics.

There is still only scant knowledge about the proteins involved in cell–cell fusion. Several candidate proteins were identified during the last years but in most cases it is not clear whether these proteins are primarily functioning in attachment rather than fusion. These include the tetraspanins CD 9 and CD 81, the single-pass membrane proteins of the ADAM family that may be involved in sperm-egg fusion, and the *Caneorhabditis elegans* protein EEF-1. Of these, EEF-1 is presently the strongest candidate for a bona fide fusion protein. It is both necessary and sufficient for cell–cell fusion, and it is structurally related to class II viral fusion proteins, containing an internal putative fusion peptide. Indeed, the search for cellular fusion proteins was driven for many years by the expectation that cellular fusion proteins must be similar

to viral fusion proteins. After all, viral fusion proteins must have originated from primordial eukaryotes.

See also Membrane Traffic: Vesicle Budding and Fusion; Membrane Transport; Protein Translocation Across Membranes.

Bibliography

Books and Reviews

Bonifacino, J.S., Glick, B.S. (2004) The mechanisms of vesicle budding and fusion, *Cell* **116**, 153–166.

Chernomordik, L.V., Kozlov, M.M. (2003) Protein-lipid interplay in fusion and fission of biological membranes, *Annu. Rev. Biochem.* **72**, 175–207.

Earp, L.J., Delos, S.E., Park, H.E., White, J.M. (2005) The many mechanisms of viral membrane fusion proteins, *Curr. Top. Microbiol. Immunol.* **285**, 25–66.

Eckert, D.M., Kim, P.S. (2001) Mechanisms of viral membrane fusion and its inhibition, *Annu. Rev. Biochem.* **70**, 777–810.

Fasshauer, D. (2003) Structural insights into the SNARE mechanism, *Biochim. Biophys. Acta* **1641**, 87–97.

Jahn, R., Lang, T., Südhof, T.C. (2003) Membrane fusion, *Cell* **112**, 519–533.

Maxfield, F.R., McGraw, T.E. (2004) Endocytic recycling, *Nat. Rev. Mol. Cell Biol.* **5**, 121–132.

Smith, A.E., Helenius, A. (2004) How viruses enter animal cells, *Science* **304**, 237–242.

Südhof, T.C. (2004) The synaptic vesicle cycle, *Annu. Rev. Neurosci.* **27**, 509–547.

Tamm, L.K., Crane, J., Kiessling, V. (2003) Membrane fusion: a structural perspective on the interplay of lipids and proteins, *Curr. Opin. Struct. Biol.* **13**, 453–466.

Primary Literature

Almers, W., Tse, F.W. (1990) Transmitter release from synapses: does a preassembled fusion pore initiate exocytosis? *Neuron* **4**, 813–818.

Antonin, W., Fasshauer, D., Becker, S., Jahn, R., Schneider, T.R. (2002) Crystal structure of the endosomal SNARE complex reveals common structural principles of all SNAREs, *Nat. Struct. Biol.* **9**, 107–111.

Atlashkin, V., Kreykenbohm, V., Eskelinen, E.L., Wenzel, D., Fayyazi, A., Fischer von Mollard, G. (2003) Deletion of the SNARE vti1b in mice results in the loss of a single SNARE partner, syntaxin 8, *Mol. Cell. Biol.* **23**, 5198–5207.

Bai, J., Chapman, E.R. (2004) The C2 domains of synaptotagmin – partners in exocytosis, *Trends Biochem. Sci.* **29**, 143–151.

Bock, J.B., Matern, H.T., Peden, A.A., Scheller, R.H. (2001) A genomic perspective on membrane compartment organization, *Nature* **409**, 839–841.

Bracher, A., Weissenhorn, W. (2002) Structural basis for the Golgi membrane recruitment of Sly1p by Sed5p, *EMBO J.* **21**, 6114–6124.

Breckenridge, L.J., Almers, W. (1987) Currents through the fusion pore that forms during exocytosis of a secretory vesicle, *Nature* **328**, 814–817.

Bressanelli, S., Stiasny, K., Allison, S.L., Stura, E.A., Duquerroy, S., Lescar, J., Heinz, F.X., Rey, F.A. (2004) Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation, *EMBO J.* **23**, 728–738.

Brose, N., Petrenko, A.G., Südhof, T.C., Jahn, R. (1992) Synaptotagmin: a Ca^{2+} sensor on the synaptic vesicle surface, *Science* **256**, 1021–1025.

Brunger, A.T. (2001) Structural insights into the molecular mechanism of calcium-dependent vesicle-membrane fusion, *Curr. Opin. Struct. Biol.* **11**, 163–173.

Brunger, A.T., DeLaBarre, B. (2003) NSF and p97/VCP: similar at first, different at last, *FEBS Lett.* **555**, 126–133.

Bullough, P.A., Hughson, F.M., Skehel, J.J., Wiley, D.C. (1994) Structure of influenza hemagglutinin at the pH of membrane fusion, *Nature* **371**, 37–43.

Chanturiya, A., Chernomordik, L.V., Zimmerberg, J. (1997) Flickering fusion pores comparable with initial exocytotic pores occur in protein-free phospholipid bilayers, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 14423–14428.

Chen, X., Tomchick, D.R., Kovrigin, E., Arac, D., Machius, M., Südhof, T.C., Rizo, J. (2002) Three-dimensional structure of the complexin/SNARE complex, *Neuron* **33**, 397–409.

Christoforidis, S., McBride, H.M., Burgoyne, R.D., Zerial, M. (1999) The Rab5 effector EEA1 is a core component of endosome docking, *Nature* **397**, 621–625.

Dilcher, M., Köhler, B., von Mollard, G.F. (2001) Genetic interactions with the yeast Q-SNARE VTI1 reveal novel functions for the R-SNARE YKT6, *J. Biol. Chem.* **276**, 34537–34544.

Dimmer, K.S., Fritz, S., Fuchs, F., Messerschmitt, M., Weinbach, N., Neupert, W., Westermann, B. (2002) Genetic basis of mitochondrial function and morphology in Saccharomyces cerevisiae, *Mol. Biol. Cell* **13**, 847–853.

Fasshauer, D., Margittai, M. (2004) A transient N-terminal interaction of SNAP-25 and syntaxin nucleates SNARE assembly, *J. Biol. Chem.* **279**, 7613–7621.

Fasshauer, D., Sutton, B., Brünger, A.T., Jahn, R. (1998) Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15781–15786.

Fasshauer, D., Antonin, W., Subramaniam, V., Jahn, R. (2002) SNARE assembly and disassembly exhibit a pronounced hysteresis, *Nat. Struct. Biol.* **9**, 144–151.

Fasshauer, D., Otto, H., Eliason, W.K., Jahn, R., Brunger, A.T. (1997) Structural changes are associated with soluble N-ethylmaleimide-sensitive fusion protein attachment protein receptor complex formation, *J. Biol. Chem.* **272**, 28036–28041.

Fasshauer, D., Antonin, W., Margittai, M., Pabst, S., Jahn, R. (1999) Mixed and non-cognate SNARE complexes. Characterization of assembly and biophysical properties, *J. Biol. Chem.* **274**, 15440–15446.

Fernandez, I., Ubach, J., Dulubova, I., Zhang, X., Südhof, T.C., Rizo, J. (1998) Three-dimensional structure of an evolutionarily conserved N-terminal domain of syntaxin 1A, *Cell* **94**, 841–849.

Fujita, Y., Shirataki, H., Sakisaka, T., Asakura, T., Ohya, T., Kotani, H., Yokoyama, S., Nishioka, H., Matsuura, Y., Mizoguchi, A., Scheller, R.H., Takai, Y. (1998) Tomosyn: a syntaxin-1-binding protein that forms a novel complex in the neurotransmitter release process, *Neuron* **20**, 905–915.

Gallwitz, D., Jahn, R. (2003) The riddle of the Sec1/Munc-18 proteins – new twists added to their interactions with SNAREs, *Trends Biochem. Sci.* **28**, 113–116.

Garner, C.C., Kindler, S., Gundelfinger, E.D. (2000) Molecular determinants of presynaptic active zones, *Curr. Opin. Neurobiol.* **10**, 321–327.

Geppert, M., Goda, Y., Hammer, R.E., Li, C., Rosahl, T.W., Stevens, C.F., Südhof, T.C. (1994) Synaptotagmin I: a major Ca^{2+} sensor for transmitter release at a central synapse, *Cell* **79**, 717–727.

Gibbons, D.L., Erk, I., Reilly, B., Navaza, J., Kielian, M., Rey, F.A., Lepault, J. (2003) Visualization of the target-membrane-inserted fusion protein of Semliki forest virus by combined electron microscopy and crystallography, *Cell* **114**, 573–583.

Gibbons, D.L., Vaney, M.C., Roussel, A., Vigouroux, A., Reilly, B., Lepault, J., Kielian, M., Rey, F.A. (2004) Conformational change and protein-protein interactions of the fusion protein of Semliki Forest virus, *Nature* **427**, 320–325.

Han, X., Wang, C.T., Bai, J., Chapman, E.R., Jackson, M.B. (2004) Transmembrane segments of syntaxin line the fusion pore of Ca^{2+}-triggered exocytosis, *Science* **304**, 289–292.

Hanson, P.I., Roth, R., Morisaki, H., Jahn, R., Heuser, J.E. (1997) Structure and conformational changes in NSF and its membrane receptor complexes visualized by quick-freeze/deep-etch electron microscopy, *Cell* **90**, 523–535.

Hardwick, K.G., Pelham, H.R.B. (1992) SED5 encodes a 39-kD integral membrane protein required for vesicular transport between the ER and the Golgi complex, *J.Cell Biol.* **119**, 513–521.

Hatsuzawa, K., Lang, T., Fasshauer, D., Bruns, D., Jahn, R. (2003) The R-SNARE motif of tomosyn forms SNARE core complexes with syntaxin 1 and SNAP-25 and down-regulates exocytosis, *J. Biol. Chem.* **278**, 31159–31166.

Hsu, S.C., TerBush, D., Abraham, M., Guo, W. (2004) The exocyst complex in polarized exocytosis, *Int. Rev. Cytol.* **233**, 243–265.

Hu, C., Ahmed, M., Melia, T.J., Söllner, T.H., Mayer, T., Rothman, J.E. (2003) Fusion of cells by flipped SNAREs, *Science* **300**, 1745–1749.

Jahn, R., Grubmuller, H. (2002) Membrane fusion, *Curr. Opin. Cell Biol.* **14**, 488–495.

Jardetzky, T.S., Lamb, R.A. (2004) Virology: a class act, *Nature* **427**, 307–308.

Koshiba, T., Detmer, S.A., Kaiser, J.T., Chen, H., McCaffery, J.M., Chan, D.C. (2004) Structural

basis of mitochondrial tethering by mitofusin complexes, *Science* **305**, 858–862.

Kozlov, M.M., Chernomordik, L.V. (1998) A mechanism of protein-mediated fusion – coupling between refolding of the influenza hemagglutinin and lipid rearrangements, *Biophys. J.* **75**, 1384–1396.

Laage, R., Rohde, J., Brosig, B., Langosch, D. (2000) A conserved membrane-spanning amino acid motif drives homomeric and supports heteromeric assembly of presynaptic SNARE proteins, *J. Biol. Chem.* **275**, 17481–17487.

Lang, T., Margittai, M., Holzler, H., Jahn, R. (2002) SNAREs in native plasma membranes are active and readily form core complexes with endogenous and exogenous SNAREs, *J. Cell Biol.* **158**, 751–760.

Lang, T., Bruns, D., Wenzel, D., Riedel, D., Holroyd, P., Thiele, C., Jahn, R. (2001) SNAREs are concentrated in cholesterol-dependent clusters that define docking and fusion sites for exocytosis, *EMBO J.* **20**, 2202–2213.

Langosch, D., Crane, J.M., Brosig, B., Hellwig, A., Tamm, L.K., Reed, J. (2001) Peptide mimics of SNARE transmembrane segments drive membrane fusion depending on their conformational plasticity, *J. Mol. Biol.* **311**, 709–721.

Lehman, K., Rossi, G., Adamo, J.E., Brennwald, P. (1999) Yeast homologues of tomosyn and lethal giant larvae function in exocytosis and are associated with the plasma membrane SNARE, Sec9, *J. Cell Biol.* **146**, 125–140.

Lin, R.C., Scheller, R.H. (1997) Structural organization of the synaptic exocytosis core complex, *Neuron* **19**, 1087–1094.

Littleton, J.T., Barnard, R.J., Titus, S.A., Slind, J., Chapman, E.R., Ganetzky, B. (2001) SNARE-complex disassembly by NSF follows synaptic-vesicle fusion, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12233–12238.

Liu, Y., Barlowe, C. (2002) Analysis of Sec22p in endoplasmic reticulum/Golgi transport reveals cellular redundancy in SNARE protein function, *Mol. Biol. Cell* **13**, 3314–3324.

Markin, V.S., Albanesi, J.P. (2002) Membrane fusion: stalk model revisited, *Biophys. J.* **82**, 693–712.

Masuda, E.S., Huang, B.C.B., Fisher, J.M., Luo, Y., Scheller, R.H. (1998) Tomosyn binds t-SNARE proteins via a VAMP-like coiled coil, *Neuron* **21**, 479–480.

McNew, J.A., Parlati, F., Fukuda, R., Johnston, R.J., Paz, K., Paumet, F., Söllner, T.H., Rothman, J.E. (2000) Compartmental specificity of cellular membrane fusion encoded in SNARE proteins, *Nature* **407**, 153–159.

Misura, K.M., Scheller, R.H., Weis, W.I. (2000) Three-dimensional structure of the neuronal-Sec1-syntaxin 1a complex, *Nature* **404**, 355–362.

Modis, Y., Ogata, S., Clements, D., Harrison, S.C. (2004) Structure of the dengue virus envelope protein after membrane fusion, *Nature* **427**, 313–319.

Monck, J.R.F. (1995) The exocytotic fusion pore and neurotransmitter release, *Neuron* **12**, 707–716.

Novick, P., Guo, W. (2002) Ras family therapy: Rab, Rho and Ral talk to the exocyst, *Trends Cell Biol.* **12**, 247–249.

Pelham, H.R. (2001) SNAREs and the specificity of membrane fusion, *Trends Cell Biol.* **11**, 99–101.

Peng, R., Gallwitz, D. (2004) Multiple SNARE interactions of an SM protein: Sed5p/Sly1p binding is dispensable for transport, *EMBO J.* **23**, 3939–3949.

Pfeffer, S. (2003) Membrane domains in the secretory and endocytic pathways, *Cell* **112**, 507–517.

Pobbati, A.V., Razeto, A., Boddener, M., Becker, S., Fasshauer, D. (2004) Structural basis for the inhibitory role of tomosyn in exocytosis, *J.Biol.Chem.* **279**, 47192–47200.

Reim, K., Mansour, M., Varoqueaux, F., McMahon, H.T., Südhof, T.C., Brose, N., Rosenmund, C. (2001) Complexins regulate a late step in Ca^{2+}-dependent neurotransmitter release, *Cell* **104**, 71–81.

Rizo, J. (2003) SNARE function revisited, *Nat. Struct. Biol.* **10**, 417–419.

Rizo, J., Südhof, T.C. (2002) Snares and Munc18 in synaptic vesicle fusion, *Nat. Rev. Neurosci.* **3**, 641–653.

Roy, R., Laage, R., Langosch, D. (2004) Synaptobrevin transmembrane domain dimerization-revisited, *Biochemistry* **43**, 4964–4970.

Scales, S.J., Hesser, B.A., Masuda, E.S., Scheller, R.H. (2002) Amisyn, a novel syntaxin-binding protein that may regulate SNARE complex assembly, *J. Biol. Chem.* **277**, 28271–28279.

Schiavo, G., Matteoli, M., Montecucco, C. (2000) Neurotoxins affecting neuroexocytosis, *Physiol. Rev.* **80**, 717–766.

Schoch, S., Deak, F., Königstorfer, A., Mozhayeva, M., Sara, Y., Südhof, T.C., Kavalali, E.T. (2001) SNARE function analyzed in synaptobrevin/VAMP knockout mice, *Science* **294**, 1117–1122.

Seabra, M.C., Wasmeier, C. (2004) Controlling the location and activation of Rab GTPases, *Curr. Opin. Cell Biol.* **16**, 451–457.

Sesaki, H., Jensen, R.E. (2004) Ugo1p links the Fzo1p and Mgm1p GTPases for mitochondrial fusion, *J. Biol. Chem.* **279**, 28298–28303.

Skehel, J.J., Wiley, D.C. (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin, *Annu. Rev. Biochem.* **69**, 531–569.

Söllner, T., Bennet, M.K., Whiteheart, S.W., Scheller, R.H., Rothman, J.E. (1993) A protein assembly-disassembly pathway in vitro that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion, *Cell* **75**, 409–418.

Söllner, T., Whiteheart, S.W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P., Rothman, J.E. (1993) SNAP receptors implicated in vesicle targeting and fusion, *Nature* **362**, 318–324.

Sorensen, J.B. (2004) Formation, stabilization and fusion of the readily releasable pool of secretory vesicles, *Pflugers Arch.* **448**, 347–362.

Sørensen, J.B., Nagy, G., Varoqueaux, F., Nehring, R.B., Brose, N., Wilson, M.C., Neher, E. (2003) Differential control of the releasable vesicle pools by SNAP-25 splice variants and SNAP-23, *Cell* **114**, 75–86.

Stein, K.K., Primakoff, P., Myles, D. (2004) Sperm-egg fusion: events at the plasma membrane, *J. Cell Sci.* **117**, 6269–6274.

Südhof, T.C. (2002) Synaptotagmins: why so many? *J. Biol. Chem.* **277**, 7629–7632.

Sutton, B., Fasshauer, D., Jahn, R., Brünger, A.T. (1998) Crystal structure of a core synaptic fusion complex at 2.4 Å resolution, *Nature* **395**, 347–353.

Toonen, R.F., Verhage, M. (2003) Vesicle trafficking: pleasure and pain from SM genes, *Trends Cell Biol.* **13**, 177–186.

Verhage, M., Maia, A.S., Plomp, J.J., Brussaard, A.B., Heeroma, J.H., Vermeer, H., Toonen, R.F., Hammer, R.E., van den Berg, T.K., Missler, M., Geuze, H.J., Südhof, T.C. (2000) Synaptic assembly of the brain in the absence of neurotransmitter secretion, *Science* **287**, 864–869.

Wang, Y., Okamoto, M., Schmitz, F., Hofmann, K., Südhof, T.C. (1997) Rim is a putative Rab3 effector in regulating synaptic-vesicle fusion, *Nature* **388**, 593–598.

Washbourne, P., Thompson, P.M., Carta, M., Costa, E.T., Mathews, J.R., Lopez-Bendito, G., Molnar, Z., Becher, M.W., Valenzuela, C.F., Partridge, L.D., Wilson, M.C. (2002) Genetic ablation of the t-SNARE SNAP-25 distinguishes mechanisms of neuroexocytosis, *Nat. Neurosci.* **5**, 19–26.

Weber, T., Paesold, G., Galli, C., Mischler, R., Semenza, G., Brunner, J. (1994) Evidence for H^+-induced insertion of influenza hemagglutinin HA2 N-terminal segment into viral membrane, *J. Biol. Chem.* **269**, 18353–18358.

Weber, T., Zemelman, B.V., McNew, J.A., Westermann, B., Gmachi, M., Parlati, F., Söllner, T., Rothman, J.E. (1998) SNAREpins: minimal machinery for membrane fusion, *Cell* **92**, 759–772.

Weisman, L.S. (2003) Yeast vacuole inheritance and dynamics, *Annu. Rev. Genet.* **37**, 435–460.

Weissenhorn, W., Dessen, A., Harrison, S.C., Skehel, J.J., Wiley, D.C. (1997) Atomic structure of the ectodomain from HIV-1 gp41, *Nature* **387**, 426–430.

Westermann, B. (2003) Mitochondrial membrane fusion, *Biochim. Biophys. Acta* **1641**, 195–202.

Whiteheart, S.W., Matveeva, E.A. (2004) Multiple binding proteins suggest diverse functions for the N-ethylmaleimide sensitive factor, *J. Struct. Biol.* **146**, 32–43.

Whyte, J.R., Munro, S. (2002) Vesicle tethering complexes in membrane traffic, *J. Cell Sci.* **115**, 2627–2637..

Wickner, W. (2002) Yeast vacuoles and membrane fusion pathways, *EMBO J.* **21**, 1241–1247.

Wickner, W., Haas, A. (2000) Yeast homotypic vacuole fusion: a window on organelle trafficking mechanisms, *Annu. Rev. Biochem.* **69**, 247–275.

Xu, T., Rammner, B., Margittai, M., Artalejo, A.R., Neher, E., Jahn, R. (1999) Inhibition of SNARE complex assembly differentially affects kinetic components of exocytosis, *Cell* **99**, 713–722.

Yang, B., Gonzalez, L. Jr., Prekeris, R., Steegmaier, M., Advani, R.J., Scheller, R.H. (1999) SNARE interactions are not selective. Implications for membrane fusion specificity, *J. Biol. Chem.* **274**, 5649–5653.

Zerial, M., McBride, H. (2001) Rab proteins as membrane organizers, *Nat. Rev. Mol. Cell Biol.* **2**, 107–117.

Protein Microarrays

Jens R. Sydor[1], *David S. Wilson*[2], *and Steffen Nock*[2]
[1] *Infinity Pharmaceuticals Inc., Cambridge, MA*
[2] *Absalus Inc., Mountain View, CA*

1 Introduction 103

2 Protein Microarrays for Protein Expression Profiling 104
2.1 Principle 104
2.2 Content for Protein Expression Profiling Chips 104
2.2.1 Source of Antigen 105
2.2.2 Types of Protein-binding Agents 105
2.3 Protein-immobilization Methods 107
2.4 Detection Techniques 109
2.5 Applications 110
2.5.1 Antibody Arrays 110
2.5.2 Antigen Arrays 111
2.6 Alternative Technologies 112

3 Protein Microarrays for Enzyme Activity Profiling 113
3.1 Principle 113
3.2 Immobilization of Enzymes or their Substrates 114
3.3 Detection of Enzymatic Activity on Microarrays 117
3.4 Applications 118
3.5 Alternative Technologies 119

4 Protein Microarrays for Protein Interaction Profiling 120
4.1 Principle 120
4.2 Applications 120
4.3 Alternative Technologies 124

5 Concluding Remarks 124

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

Bibliography 125
Books and Reviews 125
Primary Literature 126

Keywords

Capture Agents
An antibody or antibody mimetic that is immobilized on a surface for the purpose of binding an analyte of interest from a test sample.

Enzyme Activity Profiling
Determination of the activities of multiple enzymes in parallel, or of a single enzyme with respect to multiple substrates and/or potential inhibitors.

Multiplexed Protein Assay
An assay in which a single sample aliquot is simultaneously tested with respect to the abundance or activity of multiple proteins.

Protein Expression Profiling
Determination of the expression levels of multiple proteins in parallel and changes of expression levels.

Protein Interaction Profiling
Identification of the interactions between proteins and other molecules such as DNA, RNA, metabolites, or other proteins.

Protein Microarray
A two-dimensional surface with different antibodies or proteins attached at different defined locations, used for the study of the abundance or activity of proteins.

Protein microarrays allow for the parallel analysis of a large number of proteins with respect to their abundance or activity and, therefore, are playing an emerging role in proteome research.

Protein expression profiling, detection of enzymatic activity and protein interaction mapping are the most important applications of protein microarrays. The biological content, as well as the detection strategies, vary significantly between these applications, making the protein microarray technologies much more varied than those used with DNA microarrays.

In most cases, the biological content on the microarray, as well as the analytes, are proteins, which, due to their delicate nature and the heterogeneity of their properties,

complicate the manufacturing and assay design. The potential advantages of protein microarrays over nonparallel, macroscopic technologies cannot, however, be overestimated.

1 Introduction

Microarrays are miniaturized surface-based assay devices that allow for the multiplexed analysis of biological samples. In the case of DNA microarrays, which were developed in the 1990s, sequences of DNA are immobilized on a glass surface in a two-dimensional (2D) array format. These DNA microarrays are used for a number of applications in genome research.

In gene expression profiling, for example, differences in the expression levels of genes can be measured. DNA probes, each of which is composed of a sequence that is complementary to an mRNA sequence of interest, are immobilized on a glass surface in an array format. In order to detect different expression levels, mRNA from different biological samples is then extracted, amplified into cDNA and labeled with different fluorophores representing the different biological samples. Differences in gene expression can then be detected by an increase or decrease in fluorescence.

Today, DNA microarrays are commercially available and have been proven to be a very important tool in genomic research for the multiplexed comparative analysis of gene expression and gene polymorphism. Many important insights into gene expression patterns associated with disease states have been derived from such technology.

It is the proteins, however, that are the functional product of almost all genes, and mRNA levels correlate only weakly with protein levels. The measurement of mRNA also does not provide information about the activity or posttranslational modification of proteins. To gain a realistic understanding of biological systems, it is necessary to understand the expression and activity of the full complement of proteins in different biological systems.

Inspired by the power of DNA arrays to measure mRNA expression, several academic and industrial groups have created protein microarrays, with the aim of multiplexing and miniaturizing protein analysis. Protein microarrays allow researchers to quantify the abundance or activity of many proteins in parallel while expending minimal amounts of precious sample.

The main types of protein arrays are (1) expression profiling arrays, consisting of immobilized protein-capture molecules such as antibodies, which provide information about protein abundance; (2) functional arrays, consisting of arrays of correctly folded proteins, peptides or other molecules, which can be used to characterize enzymatic activities; and (3) protein interaction arrays, which can be used to characterize protein–protein, protein–DNA, protein–carbohydrate and protein–small-molecule interactions. Protein array technology has been developed to such an extent that it now has the requisite robustness, breadth, reproducibility, and flexibility to play a leading role in proteomics research.

2 Protein Microarrays for Protein Expression Profiling

2.1 Principle

The rationale for using protein arrays for expression profiling has been explained in the previous section, and can be justified by several studies showing that more information can be culled from the analysis of multiple molecular markers than could be obtained from analyzing individual ones.

The principle behind the construction of arrays for measuring multiple proteins in parallel is similar to that used in the construction of DNA arrays for multiplexed mRNA analysis. To construct DNA arrays, each feature is derivatized with a probe that hybridizes to a single mRNA species. The production of these DNA-based "capture agents" is relatively straightforward: one simply needs to design and synthesize a DNA sequence that is complementary to the mRNA that is to be measured. For protein arrays, however, there is no comparable method for creating capture agents using bioinformatics tools alone. Protein-capture agents – usually antibodies – must be developed experimentally, usually by immunizing an animal with the protein target and then isolating the resulting antibodies for immobilization.

The other major difference between arrays for measuring mRNA versus those for measuring proteins relates to how the captured analytes are detected. Captured nucleic acids can be labeled with fluorescent probes and directly detected using a confocal fluorescence scanner. Proteins in a sample can also be directly labeled, but the process is much less reproducible than is the labeling of nucleic acids. Alternatively, proteins captured on an array can be indirectly detected by the use of labeled antibodies that recognize a different epitope on the captured target proteins. These and other issues regarding the development of the biological content for expression profiling applications are discussed in the following section.

2.2 Content for Protein Expression Profiling Chips

Most protein-profiling platforms rely on "sandwich pair" reagents – two different antibodies that can simultaneously bind to a target protein. For each target protein, one of these antibodies is immobilized onto the array, and the other one, which contains a label, is used to detect the target protein that has been captured on the array (Fig. 1). The principle behind the sandwich assay, and its advantages, are discussed in Sect. 2.4.

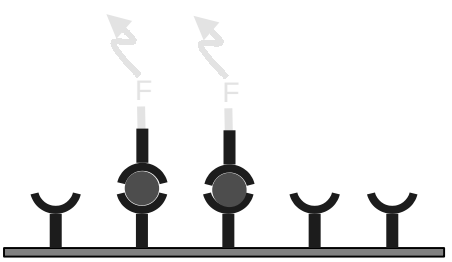

Fig. 1 Principle of a sandwich assay. A capture antibody is immobilized on a solid support. The captured analyte is detected using a second, fluorescently labeled antibody.

2.2.1 Source of Antigen

The first step in developing such sandwich pairs is usually to express and purify significant amounts of the target protein in a form as close as possible to its native state. This task can require as much, if not more, work than creating the capture agents themselves. Most proteins of interest need to be expressed in systems such as mammalian or insect cells, which require the investment of considerable time and cost. Most mammalian proteins are misfolded and insoluble when expressed in bacteria, and refolding protocols need to be individually tailored to each protein. The possibility of forgoing protein production altogether by using peptide antigens is rarely successful since most of the resulting antibodies have insufficient affinity against native proteins.

There are several sources of cDNAs for cloning into expression vectors, such as the NIH Mammalian Gene Collection, the Harvard Institute of Proteomics, the IMAGE Consortium, and the American Tissue Culture Collection (ATCC), and private companies such as Open Biosystems (Huntsville, AL) and Invitrogen (Carlsbad, CA). Modern cloning methods that avoid the use of restriction enzymes and contain various characteristics that favor correct insert orientation can make the cloning as high throughput and generic as possible. Multihost vectors and recombination-based transfer systems allow for the production of target proteins, including purification and detection tags for expression in multiple hosts. The first step is to transfer the gene into a "transfer" or "donor" vector and confirm its sequence; in the second step, the insert is shuttled into vectors for expressing the proteins in mammalian or insect cells. Both Invitrogen (Carlsbad, CA) and BD Biosciences Clontech (Palo Alto, CA) have commercialized such systems. The inclusion of N- or C-terminal purification tags in the encoded constructs allow for "generic" purification systems. The use of two affinity tags, such as the His tag plus the Glutathione-S-transferase (GST), or FLAG tag, can result in >95% pure protein using generic operating procedures.

For proteins that are large or contain important posttranslational modifications, it is often necessary to use expression systems based on higher eukaryotes. Recent improvements in technologies for introducing DNA constructs into cells, either virally or through transient transfection, have allowed for yields of up to 50 mg L^{-1} in mammalian cells or even higher in baculovirus-infected insect cells.

2.2.2 Types of Protein-binding Agents

Monoclonal antibodies (mAbs) are the most commonly used protein-binding agents for protein expression profiling platforms since they are monospecific and can be produced in unlimited quantities at low cost once a hybridoma cell line is developed. In some cases, proteolytic fragments of antibodies (Fabs) are used, since these can give improved performance due to the fact that an oriented presentation of the binding agent is possible. For sandwich-based approaches, the capture agent is almost always a mAb but the detection agents are occasionally polyclonal antibodies.

Monoclonal antibodies used in protein biochips should have dissociation constants (K_D) in the single-digit nanomolar range. Lower affinity antibodies generally have rapid ($>10^{-3}$ s^{-1}) dissociation rates (k_{off}), which result in the loss of the vast majority of captured antigen during the assay steps of washing and (in the

case of sandwich assays) incubation with detection antibodies. Capture antibodies should preferably have $k_{off} < 10^{-4}\ s^{-1}$ and $K_D < 10^{-10}$ M. It usually requires repeated attempts at immunization to obtain such reagents.

Polyclonal antibodies are generally poorly suited as capture agents for protein biochips because only 0.1 to 10% of the antibodies in a preparation are specific to the immunogen used to create them. The only way to improve on this is to affinity-purify the antigen-specific antibodies away from the bulk immunoglobulin, but this requires large amounts of antigen, and such affinity-purified polyclonal antibodies cannot compare to mAbs in terms of homogeneity and lot-to-lot consistency. For these reasons, polyclonal antibodies instead find a more suitable role as the detection agents in multiplexed platforms using arrays of mAbs. In the detection step, the captured antigen will be selectively targeted by the highest affinity antibodies in the polyclonal preparation, and the fact that >95% of the antibodies in this preparation have limited or no affinity to the antigen is relatively unimportant.

The high expense and slow nature of mAb development has inspired the invention of alternative methods for making antibodies and antibody mimetics. Display technologies in which recombinant antibody fragments are linked to the genetic information encoding them are increasingly used. Such systems include phage display, ribosome display, mRNA display, yeast display, and bacterial display. The displayed libraries are usually in a Fab or single-chain variable region (scFv) configuration, the latter of which is less suitable due to its instability and conformational heterogeneity, which leads to lower affinities toward antigens. Artificial libraries of "antibody mimetics" have been created in which certain surface loops of nonantibody proteins are replaced by random sequences to make artificial complementarity-determining regions (CDRs). Although there have been several demonstrations of the creation of high-affinity protein-binding agents from naïve libraries using display methods, typical dissociation constants are in the 10^{-6} to 10^{-8} M range and therefore not suitable for multiplexed protein analysis platforms. Another drawback of these methods is that antibody fragments are significantly more expensive to manufacture than traditional mAbs.

Perhaps the most radical approach for creating the protein-binding agents is to use nucleic acid–based molecules. Libraries of random nucleic acid sequences can be subjected to systematic evolution of ligands by exponential enrichment (SELEX) to identify "aptamers" – rare single-stranded molecules that fold into a structure capable of binding to protein targets. Arrays of aptamers can be used to create protein expression profiling arrays. "Photoaptamers" are bromodeoxyuridine-substituted DNA molecules that can photo-crosslink to their target proteins. The photo-crosslinking event provides a second dimension of specificity since it requires close juxtaposition of reactive groups. The covalent crosslink allows for stringent washing of the arrays to remove nonspecifically bound proteins. Nonspecific protein-reactive dyes can be used to detect the captured proteins on the arrays. Photoaptamers can crosslink low picomolar concentrations of target proteins with very high specificity, even in the presence of high concentrations of noncognate proteins from serum. A fluorescent image of a photoaptamer array is shown in Fig. 2.

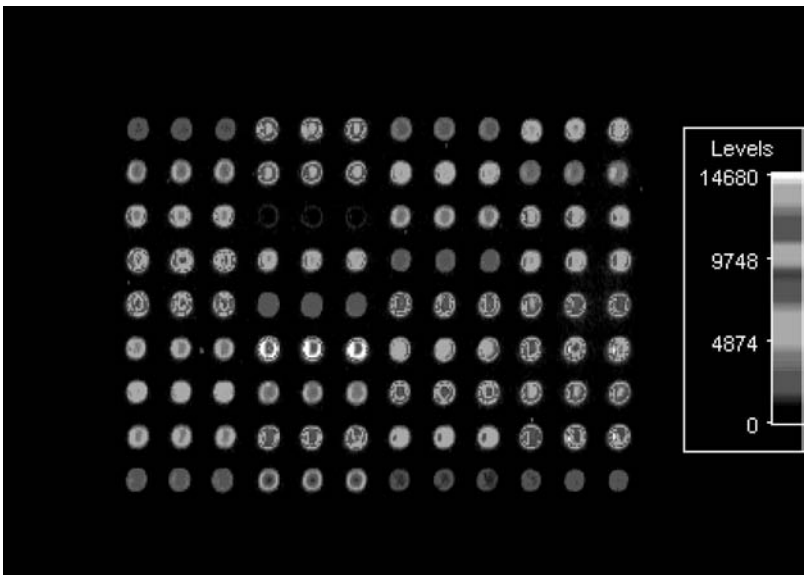

Fig. 2 Example of a photoaptamer array. Universal protein stain (UPS) photoaptamer array for the simultaneous assay of 26 analytes. The array was exposed to 10% human serum. After photo-crosslinking, the array was washed and stained with Alexa 647 dye and read on a TECAN LS300 scanner. (Figure courtesy SomaLogic Inc).

2.3 Protein-immobilization Methods

Proteins are very fragile molecules that are easily denatured at liquid–solid and liquid–air interfaces, and therefore the nature of the surfaces to which proteins are attached, as well as the dispensing method, can have pronounced effects on the surface protein density and the fraction of immobilized protein that is active. This section discusses dispensing methods and surface chemistries used for two-dimensional protein arrays.

Three approaches to protein immobilization are in common use.

Physical adsorption. The easiest way to immobilize proteins is by physical adsorption to surfaces with inherent nonspecific protein-binding activity. The surface-binding energy is hydrophobic in nature when using materials such as plastic, or electrostatic in nature when using charged surfaces such as polylysine-coated glass or nitrocellulose. A protein film is spontaneously formed on such surfaces when brought into contact with a protein solution, making this the simplest procedure for creating arrays. In such films, each protein is attached via multiple weak noncovalent interactions, but since the overall interaction surface on each protein is large, the sum of these weak interactions results in binding energies that are orders of magnitude stronger than typical specific protein–protein interactions. The main disadvantage of these approaches is that the interaction energies are so great that they frequently denature or otherwise inactivate the immobilized proteins. Generally, 75 to 90% of proteins attached to surfaces in this way are inactivated.

Random chemical attachment. The second most facile protein-immobilization method is based on random covalent attachment of proteins to substrates. Most methods take advantage of the fact that proteins typically contain several surface-exposed lysine residues. The amino group on the lysine side chains can react with surfaces containing, for example, *N*-hydroxysuccinimide moieties. The random chemical coupling results in heterogeneity with respect to the number of crosslinking sites per protein, the position of the crosslinks, and, as a result, the orientation of the proteins with respect to the surface. Many of the proteins are therefore inactivated by modification, by inaccessibility of the binding site, or by denaturation arising from strain caused by multiple attachment sites. A variation on this theme relies on random chemical biotinylation of lysine residues followed by exposure to streptavidin-coated surfaces.

Site-specific attachment. Optimal protein activity-retention is achieved by using single point oriented surface attachment of the binding agents. The most facile way of achieving this is to conjugate a binding agent through a unique cysteine residue that is either naturally present or has been introduced by genetic engineering. Surfaces can be derivatized by a number of thiol-reactive functionalities such as maleimide. Coupling can be made to occur almost exclusively to the thiol group on the unique cysteine, giving homogeneously oriented binding agents. Several studies have demonstrated the superior performance of protein arrays made using this method. Alternatively, the glycosylated functionalities on the Fc region of antibodies, which is distal from the antigen-binding site, can be used as an immobilization handle. Recombinant proteins can also be attached to surfaces in an oriented fashion by incorporating affinity tags such as the hexahistidine tag or the biotin-accepting peptide, which can be enzymatically biotinylated. These tags allow for the oriented immobilization on surfaces coated with Ni-complexes or streptavidin.

The choice of methods for immobilization depends highly on the sensitivity required in the assay and the number of binding agents that will be immobilized. For example, the amount of work involved in creating protein arrays with site-specifically attached proteins may not be justified when the number of proteins is high. Alternatively, arrays of 10 to 20 proteins in which a 2 to 5-fold increased sensitivity would be highly valued would justify the added effort required to create arrays with specifically oriented binding agents.

Flat surfaces have an inherent limitation on the number of protein molecules that can be immobilized per unit area. To extend beyond this limit, one must build a brush or gel structure onto the surface and attach the proteins to it. One theoretical drawback to this approach may be poor penetration into and out of the structure, leading to slow equilibration times and high nonspecific binding. This approach led to the commercialization of a "hydrogel" slide based on polyacrylamide, now sold by Perkin-Elmer Life Sciences (Foster City, CA). Slides such as these have been shown capable of immobilizing active proteins at a density at least fivefold higher than is possible using optimized flat surfaces. Another advantage is that the hydrogel can protect proteins from dehydration and therefore provide for a higher specific activity.

2.4 Detection Techniques

Nearly all reports of protein arrays rely on detection by fluorescence or chemiluminescence. Charge-coupled device-based cameras or laser scanners with confocal detection optics are commercially available. The proteins in a biological sample can be directly labeled with fluorescent dyes or with biotin, captured by the binding agents on the features of the array, and then detected using such fluorescent scanners (Fig. 3). To compare expression patterns between two different samples, each of the two samples is labeled with a different fluorescent dye. After labeling, the two samples are mixed together and then exposed to the array. A confocal fluorescence scanner then quantifies the signal from the two dyes (typically Cy3 and Cy5). However, such methods generally have very poor sensitivity for the following reasons: (1) fewer labels can be conjugated per protein without causing loss of solubility than for nucleic acids; (2) many proteins in samples are prone to nonspecific binding to surfaces, resulting in high background; and (3) different proteins are labeled with different efficiencies, usually relating to the number of accessible lysine residues. Owing to these reasons, chemical labeling of proteins has very limited applications for protein arrays.

Orders of magnitude higher sensitivity and reproducibility can be obtained with the use of ''sandwich pair'' antibodies (Fig. 1). In this case, two different antibodies, reacting to different regions (epitopes) of the target protein, are exploited. One antibody – the ''capture agent'' – is immobilized onto the array feature and captures the target protein; the other one – the ''detection agent'' – carries a label and binds to the captured target protein, forming a ''sandwich'' around it. The use of two antibodies greatly increases both the absolute amount of signal per captured target protein (since the detection antibody can be derivatized with very bright labels such as phycoerythrin), as well as the specificity of the signal (since two antibody-target protein-binding events must occur to produce signal). Such an assay format has the additional advantage of lending itself to high sample throughput since the biological sample need not be chemically labeled. The disadvantage of the use of sandwich reagent pairs is that two antibodies, both of which must recognize the target protein in its native form, must be developed and validated.

The sensitivity of sandwich assays can be pushed to new limits by amplifying the signal from the detection antibody using rolling circle amplification (RCA). To accomplish this, the detection antibody is labeled with a DNA primer annealed to a circular DNA template. Upon localization of the detection antibody on the array, the DNA primer on the antibody can be extended by a DNA polymerase on the circular template. Since the template has no end, the polymerase creates very

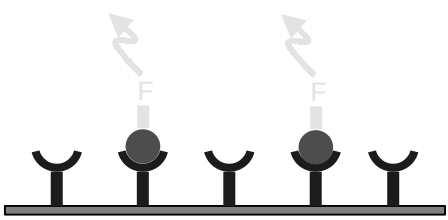

Fig. 3 Fluorescently labeled proteins are captured on an antibody array.

long DNA products in a "rolling circle amplification" reaction. Hybridization of fluorescently labeled oligonucleotides to this DNA product leads to a high density of fluorophores on the feature. Using this approach, it is possible to detect dozens of cytokines in parallel in the low picomolar range.

Another strategy for the fluorescent detection of proteins on array surfaces is to use planar waveguide technology in conjunction with fluorescently labeled detection antibodies. In this case, only the surface-bound fluorescently labeled molecules are detected, so no washing step is required. This allows one to take measurements under equilibrium conditions, thus avoiding the time-consuming array-washing steps.

In contrast with fluorescence and other optical readouts, mass spectrometry (MS) provides information as to the molecular weights of the captured proteins. The most widely used system relies on matrix-assisted laser desorption–ionization (MALDI) or surface-enhanced laser desorption–ionization (SELDI). The target proteins are captured by surface immobilized antibodies. One limitation of this system is the relatively low sensitivity in comparison to sandwich assay–based methods and the need to treat the surface-captured proteins with proteases prior to MALDI analysis. It is also difficult to obtain quantitative data on the basis of a purely MS-based approach.

2.5 Applications

2.5.1 Antibody Arrays

Antibody arrays have been used to monitor relatively small numbers of proteins in parallel, mainly due to the lack of developed antibody content for these platforms. Most arrays investigating >100 specificities have relied upon direct sample labeling since there are few proteins for which high quality sandwich pair antibodies are available. Generally, the two-dye labeling method mentioned above is used to compare two different samples. Several groups have analyzed changes in protein expression among cells as they become cancerous, or progress through different cytological stages, or are exposed to radiation.

Protein expression changes have been identified in squamous cell carcinoma, for example, with the aid of laser capture microdissection to separate out the different cell types prior to sample labeling. Samples were applied to an array of 368 antibodies against known signaling proteins. Eleven different proteins were identified that changed expression levels, some of which were in the cancer cells and others were in the surrounding stroma. A similar approach of combining laser capture microdissection with antibody arrays to find cancer biomarkers was used to study hepatocellular carcinoma. Antibodies to 83 proteins likely to be involved in proliferation made up the array, and differences in expression were observed for 32 of them. The data was confirmed by the use of western blots and tissue microarrays.

Most published reports of protein expression profiling arrays focus on a smaller number (10–75) of specificities, generally all within a restricted class of proteins, and use sandwich-based antibody pairs for detection. One high-density array that allowed for the measurement of expression levels of 75 cytokines was constructed. Similar products have been reported, but with a smaller number of specificities. Such cytokine-detection arrays have made it possible to uncover the complex changes in cytokine abundance that result from exposure to proinflammatory factors such

as lipopolysaccharide or tumor necrosis factor-α. Similar arrays have also delineated differences in cytokine production between TH1 and TH2 helper cells, as well as uncovered an unexpected relationship between vitamin E levels and the cytokine monocyte chemoattractant protein-1.

In another study, the relationship between 78 inflammatory mediators and cerebral palsy was studied. Cord blood from 19 cerebral palsy children and 19 gestation matched controls were analyzed and significant differences were observed in the expression of 11 of the 78 proteins studied. In addition, it was found that cord blood from preterm infants with cerebral palsy showed even more elevated levels of certain inflammatory mediators than the gestation matched controls.

2.5.2 Antigen Arrays

Antigen arrays, sometimes referred to as *reverse arrays*, are used to measure the levels of antibodies with defined specificities in serum. Proteins of interest, allergens, or proteins known to elicit an autoimmune response are arrayed on a substrate. Serum samples of patients are applied, allowing the antibodies from the patient to bind to the arrayed proteins, if antibodies with such binding specificity are present. Then, a fluorescently labeled antihuman antibody–antibody is added to reveal which types of antibodies are present in the serum (Fig. 4).

Antigen arrays have been successfully applied to the fields of allergy, autoimmune disease, and inflammation. An array of 430 distinct proteins and overlapping peptides that represent the simian-human immunodeficiency virus (SHIV) proteome, for example, was used to analyze serum from vaccinated and nonvaccinated primates after challenge with SHIV. It was observed that the vaccinated versus nonvaccinated animals could be distinguished and survival predicted. A

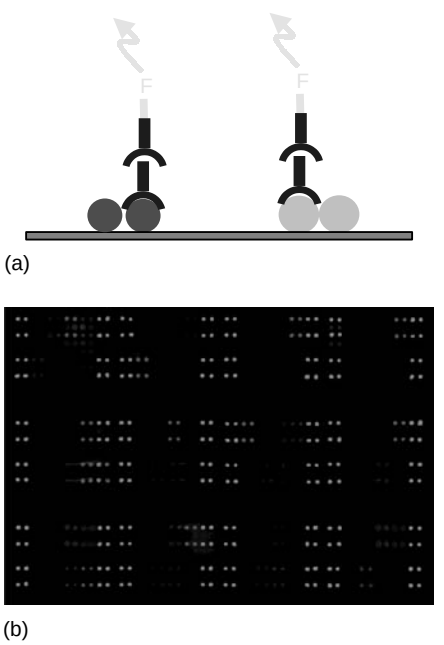

Fig. 4 Scheme and example of an antigen array. (a) Antigens are arrayed on a two-dimensional surface. Autoantibodies in serum are captured and detected with a fluorescently labeled antihuman antibody-antibody. (b) "Arthritis Array" characterization of autoantibodies in systemic lupus erythematosus. Arthritis arrays were produced by printing common lupus antigens including DNA, histone proteins, and additional nuclear proteins on microscope slides. Arrays were incubated with patient serum, and binding of autoimmune antibodies was detected with green fluorescent markers (green spots). The yellow spots are used as "marker features" to orient the arrays. (Figure courtesy Dr. William Robinson, Stanford University) (see color plate p. xxix).

second study used an array of proteins and peptide fragments representing all the known proteins in myelin to monitor autoantibody development in experimental autoimmune encephalomyelitis, a model for multiple sclerosis.

Antigen arrays were used to profile antibodies from patients diagnosed with various rheumatoid diseases, including rheumatoid arthritis, systemic lupis, systemic sclerosis, and Sjogrens syndrome. A characteristic autoantibody pattern for each of the diseases was observed, giving hope that such an array could be used as a diagnostic and monitoring tool.

It is well known that cancer patients often develop antibodies against proteins expressed in their tumors. In an effort to use such immunological information in the diagnosis of cancer, Qiu and colleagues fractionated protein lysates from a human lung cancer cell line and then arrayed them onto nitrocellulose-coated slides. The serum from individuals with or without lung cancer were then applied to the arrays to visualize the autoreactivity to cancer antigens. The signals from duplicate spots (within-slide) and duplicate slides were highly reproducible, exhibiting correlation values >0.9. Of the 1840 arrayed fractions, 63 demonstrated increased reactivity in cancer patients relative to normal, as measured by a rank-based statistic ($p < 0.008$). This type of approach, or one based on arraying purified cancer antigens, may be useful in the future in the identification and classification of patients.

2.6 Alternative Technologies

To increase the sample throughput and still maintain at least a modest degree of multiplexing, several groups have constructed antibody arrays within the wells of microtiter plates, generally containing 96 wells/plate. In such cases, the readout is typically based on a sandwich assay in which the detection antibodies are labeled with an enzyme that catalyzes a reaction to form locally precipitating fluorescent or chemiluminescent products. A commercial system for cytokine analysis has been commercialized by Pierce (Rockford, IL).

A more radical departure from the array configurations considered in the previous chapters is the so-called ''liquid array'' – suspended, encoded particles, each coated with a single antibody specificity. The most common way to encode beads relies on the incorporation of different ratios of two or more fluorophores with distinguishable spectral properties. A flow cytometer is exploited to read particles one at a time, thus decoding the identity of each particle (and thereby the identity of the immobilized antibody), while simultaneously reading a third fluorescent signal that represents binding of the target protein. This third signal is generally produced by the binding of a detection antibody labeled with the strongly fluorescent protein phycoerythrin, in a sandwich setup. The most popular suspension array platform is marketed by Luminex (Austin, TX). Several companies sell kits for multiplexed detection of cytokines and other proteins. Instruments and reagents are also available commercially to allow researchers to construct their own liquid arrays. Some of the key similarities of the liquid versus the more standard two-dimensional array platforms are shown in Table 1. The key advantages of the liquid array system are as follows: (1) manufacturing is easier since antibodies can be incubated in batch mode with the beads, thus overcoming difficulties in reproducible spotting due to denaturation, evaporation, and so on; (2) commercial systems are already

Tab. 1 Comparison of multiplexed protein analysis platforms.

	Two-dimensional arrays	*Suspended particle arrays*
Assay type	Direct capture, sandwich assay, functional assays	Direct capture, sandwich assay
Array construction method	Robotic spotting	Batch incubation of encoded particles with antibodies followed by particle mixing
Detection	Fluorescence, radioactivity, SPR, mass spectrometry	Fluorescence
Multiplexing	High	Medium
Sample throughput	Low	High
Storage	Dry	In buffer

available for fully automating the assay procedure; (3) ability to "mix and match" beads with a variety of specificities, as needed; and (4) compatibility with microtiter plates, since samples are loaded from 96-well plates and instruments can analyze an entire plate of samples in less than one day. Because of these advantages, the liquid array systems are becoming more popular and useful in life sciences than the two-dimensional arrays and will continue to be so for multiplexing applications in which the number of specificities is <100 and when sample throughput is important.

There are numerous publications using liquid arrays for multiplexed protein expression profiling, mostly involving the measurement of a limited number of cytokines or other secreted signaling molecules. One example is the measurement of cytokines in the serum of children infected with rotavirus. Three cytokines showed significantly higher expression levels in infected children. In another study, Soldan and colleagues profiled seven cytokines from stimulated peripheral blood mononuclear cells from females with relapsing-remitting multiple sclerosis (RRMS), and secondary progressive multiple sclerosis (SPMS). Two cytokines were significantly different between the RRMS and SPMS patients, which may indicate therapeutic directions for treating these two ailments.

3 Protein Microarrays for Enzyme Activity Profiling

3.1 Principle

Although it is of great value to determine the abundance of proteins in complex biological systems, it is also important to measure the enzymatic activity of proteins. Kinases, for example, play an essential role in almost all signal transduction processes in cells. The multiplexing capability of microarrays is especially suited for these screening efforts.

Two microarray formats are most commonly used for enzyme activity screening. In one case, the substrates of enzymes such as kinases or proteases are immobilized (Fig. 5a). The reaction can be monitored by a variety of different detection strategies that depend on the modification. If used for the determination of enzyme abundance, this format can also lead to

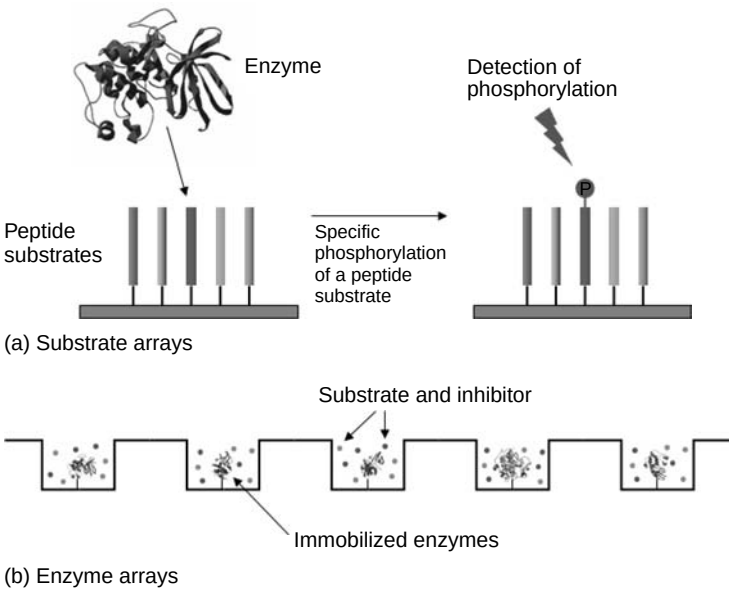

Fig. 5 The two most common formats for enzyme activity screening on microarrays. (a) Substrates, in this case of kinases, are arrayed on a surface, and their modification by enzymes detected. (b) Enzymes are immobilized, typically in nanowells, and their activity is measured by incubation with substrates in the absence or presence of inhibitors.

improved sensitivity, as the enzyme abundance is measured by its activity, which only requires the presence of a specific substrate and not through its binding to an immobilized capture agent, in which case the sensitivity depends on the affinity.

In the other case, an array of enzymes is prepared and their activity detected at their location, usually in a nanowell format (Fig. 5b). This format can also be used for inhibitor screening.

3.2 Immobilization of Enzymes or their Substrates

As in the case of protein-profiling microarrays, the immobilization strategy is also a key feature for the sensitivity and reproducibility of enzyme assays in microarray formats. The advantage in the manufacturing of peptide or small-molecule arrays, however, is the lack of complex and environmentally dependent three-dimensional structures of these molecules as compared with proteins. To ensure reproducibility and allow for a quantitative comparison of results across microarrays, however, it is still crucial to immobilize peptide substrates in an oriented and uniform fashion.

Peptide substrates for enzyme activity screening can be readily produced in a high-throughput fashion by solid-phase synthesis, which also allows for the incorporation of a variety of functional groups for covalent surface attachment. Several elegant approaches for covalent immobilization have been developed using different reaction schemes (Table 2).

Tab. 2 Selected reactions for the covalent immobilization of peptides on surfaces.

Functional group in substrate	*Functional group on surface*	*Reaction product for immobilization of substrate on surface*
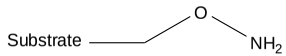 Alkoxyamino group	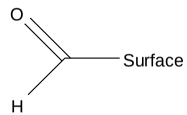 Aldehyde group	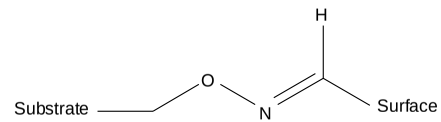 Oxime group
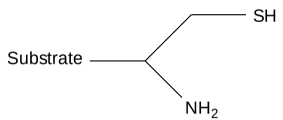 1,2-Aminothiol group	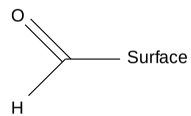 Aldehyde group	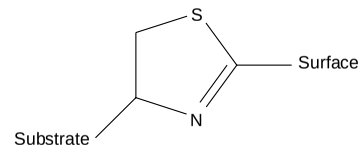 Thiazolidine group
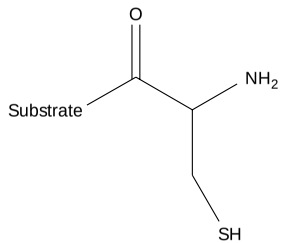 Cysteine group	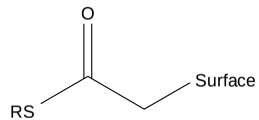 Thioester group	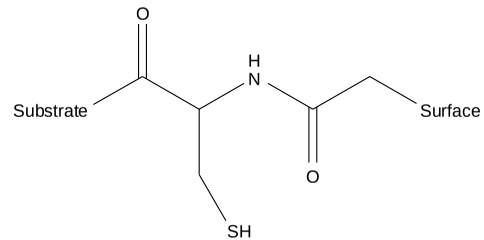 Amide group

(*continued overleaf*)

Tab. 2 (*Continued*)

Functional group in substrate	*Functional group on surface*	*Reaction product for immobilization of substrate on surface*
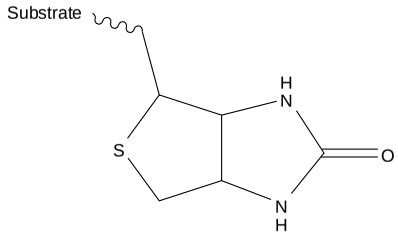 Biotin group	Avidin – surface	High-affinity interaction between avidin on the surface and biotin in the substrate
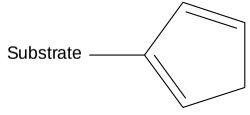 Cyclopentadiene group	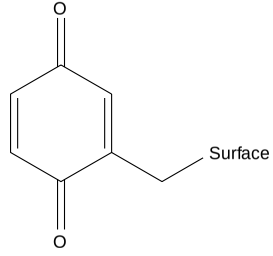 Benzoquinone group	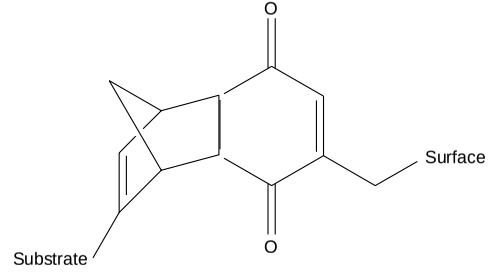 Bicyclic group
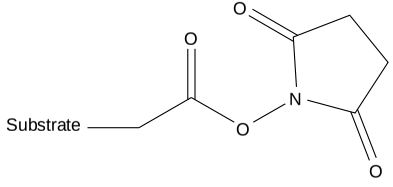 NHS-activated ester group	H_2N —— Surface Amino group	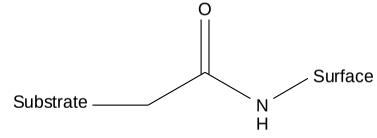 Amide group

An advantage of these approaches is the oriented immobilization, thereby decreasing the fraction of peptides that are immobilized in an orientation that inhibits their participation in the enzymatic reaction.

Another approach towards manufacturing peptide arrays is the direct synthesis of the peptides on surfaces. In the so-called SPOT synthesis, reaction reagents, together with the amino acid monomers, are spotted onto a supporting membrane, such as cellulose, allowing for the construction of arrays with about 100 sequences/cm^2. An additional on-chip peptide synthesis method has been described by Gao and coworkers, which uses t-Boc-based chemistry to generate acids on the surface by photolithography for spatial deprotection of the amino acids during the synthesis process.

Nitrocellulose has been used to create arrays, not only of proteins, but also of peptides. This surface, like many others, has a very high nonspecific protein-binding tendency, which can however, interfere with assays. Bovine serum albumin (BSA) is a very common blocking agent in protein microarray assays to minimize nonspecific protein binding to microarrays surfaces. In peptide microarrays, however, it can mask the much smaller peptides and prevent enzymes from gaining access to them. The use of oligo(ethylene glycol) groups on the surface has proven to be very useful in the prevention of nonspecific adsorption. These groups can also act as a linker between the surface and the immobilized peptides. Another approach exploits the nonspecific adsorption of BSA onto surfaces such as glass; the film of BSA can then be modified by attaching *N*-hydroxysuccinimide groups that can react with amino groups on the peptide substrates.

In cases where the enzyme, rather than the substrate, is arrayed, it is often necessary to create a diffusion boundary between the different features of the array; as a result, any modified substrate remains associated with the feature containing the arrayed enzyme. Fabricated chips with microwells containing covalently attached kinases or proteases have been used to characterize the specificity of enzymes and to identify inhibitors. In this case, the assay strategy is very similar to plate-based enzyme assays, in which the enzyme is incubated with its substrate in the presence of different inhibitors and the extent of substrate modification is an indication of the inhibitory potential of the inhibitors.

3.3 Detection of Enzymatic Activity on Microarrays

The strategy for detecting enzymatic activity on a microarray depends on the nature of the catalyzed reaction. To profile proteolytic or otherwise hydrolytic reactions, one can take advantage of the reaction releasing an immobilized fluorescent dye from the substrate. Coumarin-based derivatives are the most commonly exploited dyes for such studies and have been used to monitor proteolysis of peptide substrates (Fig. 6).

Kinase activity results in a phosphorylation of a specific target sequence in a peptide or protein substrate on the microarray. Such a reaction requires the presence of a cosubstrate such as ATP, from which the phosphate group is transferred to the substrate on the surface. A very sensitive approach for the detection of phosphorylation events is to use radioactive [γ-^{32}P]-ATP and quantify the amount of radioactive phosphate transferred to different arrayed

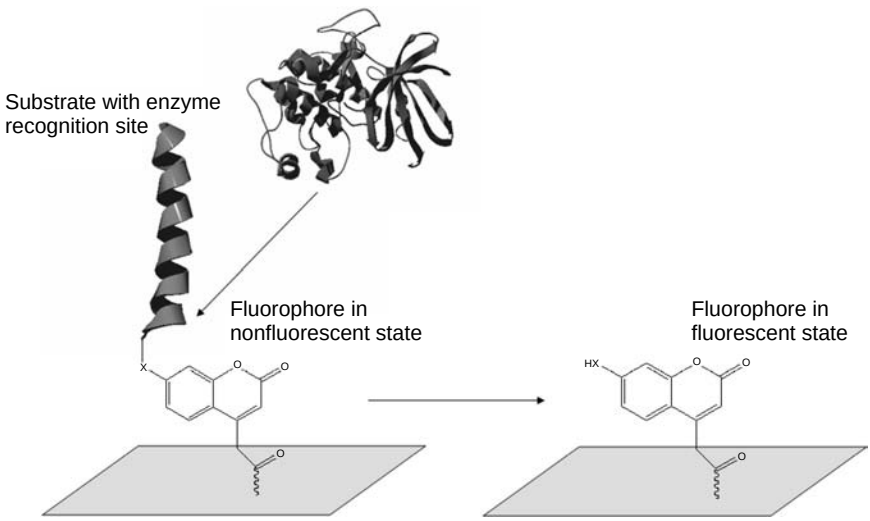

Fig. 6 Detection of protease activity on microarray surfaces with fluorescently labeled peptide substrates. The peptide substrates are immobilized on the array surface with fluorophores adjacent to the enzyme recognition site. Cleavage releases the peptide substrate and the immobilized fluorophores become fluorescent.

substrates using phosphorimaging analysis or autoradiography. Because these detection methods suffer from low spatial resolution, an alternative is to apply a photographic emulsion directly to the arrays (after the enzymatic phosphotransfer reactions were complete), and visualize the results by light microscopy.

Owing to the inconvenience and potential health hazard of working with radioactivity, nonradioactive alternatives to the detection of phosphotransfer reactions on arrays can also be applied. Phosphorylated peptide substrates can be detected using fluorescently labeled antiphosphopeptide antibodies. The universal application of this technique compared with radioactive detection, however, is problematic, as the affinity of antibodies varies depending on the phosphoamino acid sequence context. Another possible detection strategy for kinase microarray assays exploits the use of phosphospecific fluorescent dyes; such dyes are used in gel electrophoresis for the detection of phosphorylation.

Figure 7 summarizes the different detection technologies for kinase activity arrays.

3.4 Applications

Microarrays of substrates or enzymes are becoming increasingly important for the characterization, abundance determination and inhibitor screening of enzymes. Perhaps because it is far easier to array the substrates than the enzymes themselves, most published applications to date describe peptide microarrays used to detect protease or kinase activity. Mrksich and coworkers, for example, developed a peptide chip used to characterize inhibitors of the nonreceptor tyrosine kinase c-Src. The chip was prepared by immobilizing the kinase substrate onto a self-assembled

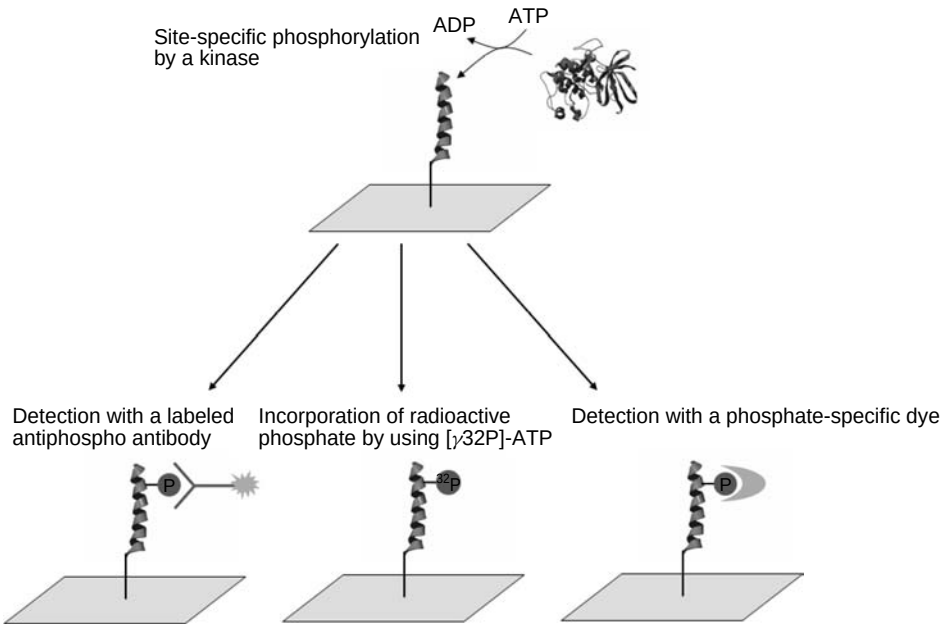

Fig. 7 Overview of the different detection technologies for kinase activity arrays.

monolayer of alkanethiolates on gold. Phosphorylation of the immobilized peptides was characterized by surface plasmon resonance, fluorescence, and phosphorimaging. Three inhibitors of the enzyme were quantitatively evaluated in an array format on a single, homogeneous substrate.

Most applications that have been described so far, however, are proof-of-concept studies evaluating either the immobilization or the detection strategies that have been described above. Examples of these applications to characterize kinases or proteases with substrate microarrays can be found in the list of the primary articles listed at the end of this chapter.

Novel microwell-based protein chip technology for high-throughput analysis of biochemical activities has also been described for the analysis of nearly all protein kinases from the yeast *Saccharomyces cerevisiae*. Of the 122 known and predicted yeast protein kinases, 119 were overexpressed and analyzed using 17 different substrates. Studies of this nature have shown that many novel activities can be identified and that a large number of protein kinases, not known to be capable of phosphorylating tyrosine peptide sequences, do possess this biochemical activity.

3.5 Alternative Technologies

The advent of microarrays for enzyme activity profiling has made tremendous progress in the last few years, as reflected by the number of recent publications, especially concerning peptide microarrays for kinase activity profiling and protease fingerprinting. The use of this technology in industrial research, however, is still limited

owing to the fact that most laboratories today perform assays in 384- or 1536-well plates, and existing automation technology facilitates high-throughput screening. Such plate-based assays do not require as much technology and assay development as surface-based setups.

One interesting alternative to microarrays for multiplexing the analysis of enzymatic activity was reported by Kozarich, who applied the principles of enzyme chemistry, mechanism of action and inhibitor design to create a set of activity-based probes. The use of suicide inhibitors allows for interrogation of enzyme family members, both known and unknown, in cells and protein extract fractions without the need for individual assay development and isolation. The serine hydrolases and cysteine proteases have provided the proofs of concept for this type of activity-based proteomics platform, and other studies are rapidly following.

4 Protein Microarrays for Protein Interaction Profiling

4.1 Principle

The interactions of proteins with each other and with other molecules constitute an essential component of all biological processes. In addition to the profiling of enzymatic activity, functional proteomics is therefore also focused on the elucidation of protein interaction networks, also called protein interaction maps. Different research groups, using different methods, however, often report conflicting results, emphasizing the importance of using a variety of methods in order to verify results.

Because of the complexity of these networks and the large number of interactions, high-throughput methods are essential for generating comprehensive data sets on any of the critical pathways controlling biological events. Microarrays are ideally suited for the parallel analysis of thousands of interactions of proteins with each other and with other key metabolic and regulatory molecules such as DNA, RNA, lipids, carbohydrates, metabolites, and allosteric effectors. Most of the reports on the use of protein microarrays for interaction profiling have, to date, focused on the interaction of proteins with non-protein molecules. The main reasons for this are (1) high-throughput methods already exist for studying protein–protein interactions, most notably the yeast two-hybrid system and (2) the difficulty in creating and purifying large collections of active proteins (as discussed in Sect. 2.2.1) in comparison to the relative ease of producing small molecules, peptides, RNA, or DNA.

The on-chip detection of substances, whether proteins or other molecules, that bind to the proteins on the arrays poses an additional challenge. If the analyte contains a complex mixture of molecules, it is not trivial to identify which ones are binding to the arrayed proteins. In the simplest case, for example, the analyte contains only one substance that is labeled by a fluorescent group–in which case the readout is straightforward (Fig. 8).

The used label is usually very similar to the ones that were described in Chapter 2 for protein expression profiling microarrays.

4.2 Applications

To date, at least one whole proteome microarray has been developed, consisting of

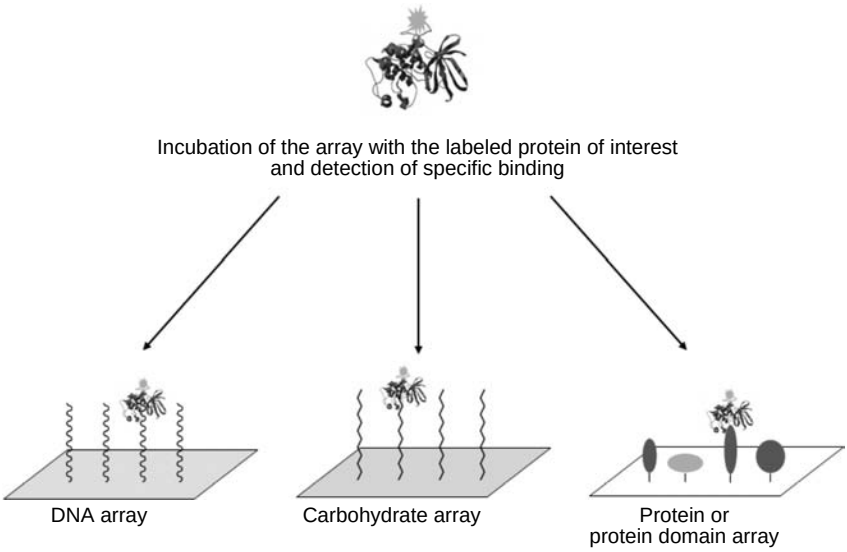

Fig. 8 Principle of a protein interaction microarray.

almost the entire yeast proteome on a chip. High-throughput protein expression and purification methods are critical for the success of such endeavors. To establish such a proteome microarray, open reading frames encoding the proteins must be cloned into an expression vector that allows for the production of fusion proteins with affinity tags that can be used in high-throughput purification (e.g. GST tag for purification on glutathione columns) and immobilization (e.g. His tag for immobilization on Ni^{2+}-nitrilotriacetic acid (Ni-NTA), coated glass slides). These proteome microarrays can then be probed, for example, with biotin-labeled proteins or other molecules and the interactions detected by adding fluorescently labeled streptavidin.

More commonly, a subset of proteins from an organism, defined by a protein or disease class are immobilized and then probed for interactions.

An elegant approach for alleviating the difficulties with fabrication of protein microarrays, as opposed to the relatively straightforward methods for arraying DNA, is the following: DNA sequences encoding proteins of interest, genetically fused to an immobilization tag, are incubated in a cell-free expression lysate on a tag-binding surface. Proteins are synthesized *in situ* and directly immobilized by the interaction of the tag with the surface (Fig. 9).

Despite these efforts, the high-throughput production of correctly folded proteins still poses an obstacle in the fabrication of protein microarrays (Sect. 2.2.1). One way to alleviate this is to synthesize and immobilize smaller, functional domains of proteins that are responsible for a specific activity or interaction. The so-called SPOT synthesis, for example (see Sect. 3.2), has been used for the preparation of microarrays of protein domains that can then be probed for interaction with target molecules.

Protein–DNA interactions are crucial for the regulation of gene transcription and DNA replication. The fact that the methods for the construction of DNA microarrays are well established makes

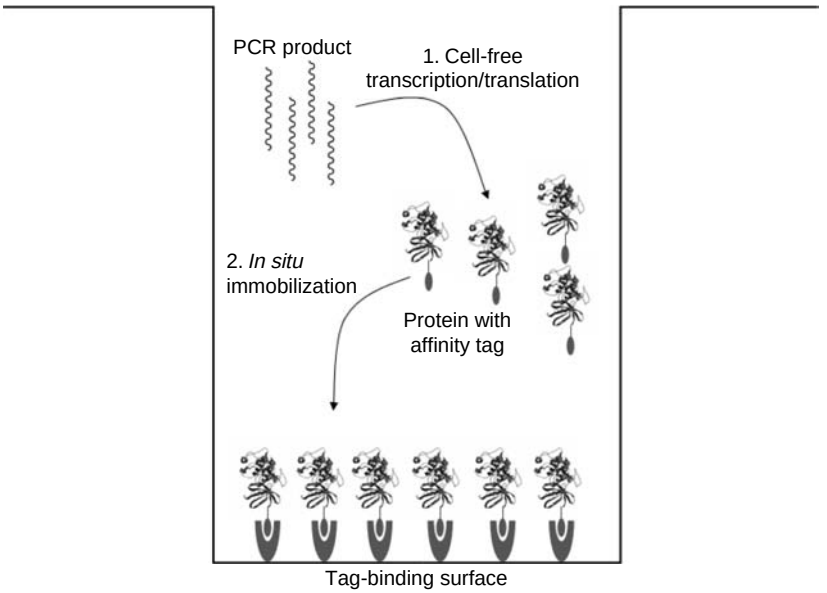

Fig. 9 Example of *in situ* immobilization of proteins on surfaces by the DiscernArray™ technology described by He and colleagues.

them an ideal tool for the screening of DNA–protein interactions. This approach can, for example, be used to screen the interaction of transcription factors or DNA-modifying enzymes.

Carbohydrates constitute another important class of molecules that interact with proteins. Advancements in the synthesis of carbohydrates, as well as carbohydrate-compatible immobilization strategies, led to the first manufacturing of arrays in 2002 and established carbohydrate microarrays as an important tool in "glycomics." Carbohydrates can be immobilized on microarray surfaces and remain accessible to interacting with proteins. The application of carbohydrate arrays to answer important biological questions has been demonstrated by Seeberger and coworkers, who used them to characterize the binding specificity of different gp120-binding proteins, important for HIV entry into host cells.

Another potentially important application in drug discovery will be the use of small-molecule microarrays for the screening of protein inhibitors. Small molecules can be routinely synthesized in large numbers by parallel synthesis technologies and modified with functional groups for specific immobilization onto surfaces. Small-molecule microarrays are therefore of growing interest to the research community. Most of the groundbreaking work in this field has been performed by Schreiber and coworkers, who developed a variety of immobilization schemes for small molecules and demonstrated the use of small-molecule microarrays in elucidating pathways in glucose signaling and in discovering transcription factor inhibitors.

Additional recent innovations with respect to the immobilization of small molecules have been reported. The so-called Staudinger ligation has been

successfully used selectively to immobilize azide-functionalized molecules on phosphane-derivatized glass slides by forming an amide bond. Another strategy, which allows for the immobilization of non-functionalized molecules, including natural products, can also be applied for the production of small-molecule microarrays by using photoaffinity labeling. Although the small molecules are not immobilized in an oriented fashion, this technology is widely applicable and the random orientation enables one to display the molecules such that most, if not all, of the functional groups will be available on some proportion of the immobilized compounds.

A different but elegant approach has been to use a DNA array to derive information about protein–small-molecule interactions. To accomplish this, each of the small molecules is coupled to a unique PNA (peptide nucleic acid hybrid molecule) sequence and a fluorescent dye. A library of fluorescent small molecule-PNA hybrids are then mixed together and incubated with a target enzyme to allow binding to occur. Next, size-exclusion chromatography is used to separate the bound from the unbound library members. The bound members are then hybridized to an array of DNA sequences that hybridize to the different PNA tags, and a fluorescence scan of the array reveals which molecules have affinity for the target protein (Fig. 10).

The possibility of screening large libraries of small molecules with a variety of target proteins, under conditions of minimal sample consumption, could potentially have an important impact on the efficiency of the early steps in the drug discovery process.

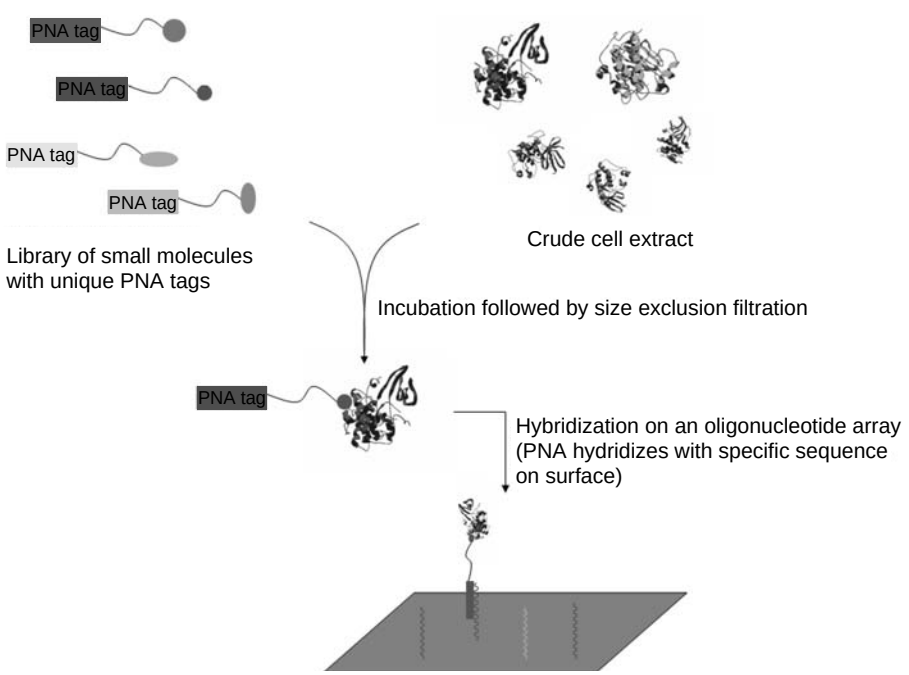

Fig. 10 Screening of protein interactions with PNA-encoded small molecule libraries as described by Winssinger and colleagues.

4.3 Alternative Technologies

The most important technology to date for the elucidation of protein–protein interactions are the two-hybrid systems. These are genetic-based *in vivo* screening systems that have been used extensively for the discovery of new interactions between proteins. The most widely used such system is in yeast and is based on the functional reconstitution of the transcriptional activator Gal4p. The interaction of two proteins is detected by fusing one protein to the DNA-binding domain of Gal4p and the other to the transcriptional activation domain. The interaction of the two proteins reconstitutes the activity of Gal4p, allowing for the transcription of reporter genes. Using the yeast two-hybrid system in large-scale screens, a detailed analysis of the whole "interactome" of yeast *S. cerevisiae* has already been performed. A disadvantage of the yeast-based system, however, is the fact that all interactions take place in the yeast nucleus, which is not an optimal environment for the proper folding of certain proteins, especially with respect to post-translational modifications. To solve this problem, other two-hybrid systems have been developed that allow for the interaction of two proteins in the cytoplasm of yeast or in mammalian cells.

Another *in vivo* screening system is based on reconstituting an enzyme from two polypeptide fragments by genetically fusing one protein of interest to one fragment and another protein to the other fragment. If the two proteins of interest interact, they bring together and reconstitute the genetically fragmented enzyme, making it active. The activity of the enzyme can then be followed by using a substrate that is converted into a fluorescent product, for example.

Many protein–protein interactions are not binary in nature but rather require the formation of multiprotein complexes. Different approaches have been developed to tackle this problem in which a large number of proteins are cloned into expression vectors and expressed with fusion tags, such that they can form protein complexes under physiological conditions. The tag can then be used to extract the protein complexes from cell lysates and then identify the proteins in these complexes by gel electrophoresis and mass spectrometry.

Each of these techniques has its advantages and disadvantages. The two-hybrid system, for example, can only elucidate binary interactions but is very efficient at detecting transient or weak interactions, in contrast to methods requiring the extraction of protein complexes. Only a combination of techniques can result in a more complete picture of protein interaction maps for complex biological systems. Protein microarrays are especially useful for the determination of protein interactions with other molecules such as DNA, RNA or carbohydrates. Even for the detection of protein–protein interactions, protein arrays can be *in vitro* alternatives to the cell-based two-hybrid system and may be useful where the latter is deficient, such as for identifying interactions involving secreted proteins.

5 Concluding Remarks

The emergence of proteomics has generated the need for new technologies that introduce parallel measurements and miniaturization into protein analysis. Protein expression profiling is particularly well suited to miniaturization, especially

for the analysis of soluble proteins such as growth factors and cytokines. Dozens of platforms that rely on different types of configurations and detection methods have been established. Both planar and suspension-based antibody arrays, using fluorescent sandwich pair antibodies for detection, have demonstrated that dozens of proteins can be analyzed quantitatively at picomolar concentrations in parallel. Further broadening the applications of this powerful technology will require the development of hundreds to thousands of new high quality antibodies. Multiplexed analysis of enzyme activity and protein interaction profiling is also technically well established and already has enabled important advances. For these technologies also, the rate-limiting step in developing further applications is the development of biological content, that is, collections of functional recombinant proteins. As such content becomes available, there is little doubt that protein microarrays will become a mainstream tool for biological research.

See also Microarray-Based Technology: Basic Principles, Advantages and Limitations; Peptide and Non-Peptide Combinatorial Libraries; Total Analysis Systems, Micro.

Bibliography

Books and Reviews

Amaratunga, D., Cabrera, J. (2003) *Exploration and Analysis of DNA Microarray and Protein Array Data*, Wiley, New York.

Cahill, D.J., Nordhoff, E. (2002) Protein arrays and their role in proteomics, *Adv. Biochem. Eng. Biotechnol.* **83**, 177–187.

Drickamer, K., Taylor, M.E. (2002) Glycan arrays for functional glycomics, *Genome Biol.* **3**, pp. 1034.1–1034.4.

Feizi, T., Fazio, F., Chai, W., Wong, C.H. (2003) Carbohydrate microarrays – a new set of technologies at the frontiers of glycomics, *Curr. Opin. Struct. Biol.* **13**, 637–645.

Fodor, S.P., Rava, R.P., Huang, X.C., Pease, A.C., Holmes, C.P., Adams, C.L. (1993) Multiplexed biochemical assays with biological chips, *Nature* **364**, 555–556.

Fung, E. (2004) *Protein Arrays: Methods and Protocols (Methods in Molecular Biology)*, Humana Press, Totowa, NJ.

Goddard, J.-P., Reymond, J.-L. (2004) Enzyme assays for high-throughput screening, *Curr. Opin. Biotechnol.* **15**, 314–322.

Haab, B.B. (2003) Methods and applications of antibody microarrays in cancer research, *Proteomics* **3**, 2116–2122.

Houseman, B.T., Mrksich, M. (2002) Towards quantitative assays with peptide chips: a surface engineering approach, *Trends Biotechnol.* **20**, 279–281.

Huang, R.P. (2003) Cytokine antibody arrays: a promising tool to identify molecular targets for drug discovery, *Comb. Chem. High Throughput Screen.* **6**, 769–775.

Jain, K.K. (2004) Applications of biochips: from diagnostics to personalized medicine, *Curr. Opin. Drug Discov. Dev.* **7**, 285–289.

Jenkins, R.E., Pennington, S.R. (2001) Arrays for protein expression profiling: towards a viable alternative to two-dimensional gel electrophoresis? *Electrophoresis* **1**, 13–29.

Jessani, N., Cravatt, B.F. (2004) The development and application of methods for activity-based protein profiling, *Curr. Opin. Chem. Biol.* **8**, 54–59.

Jona, G., Snyder, M. (2003) Recent developments in analytical and functional protein microarrays, *Curr. Opin. Mol. Ther.* **5**, 271–277.

Kambhampati, D. (2004) *Protein Microarray Technology*, Wiley, New York.

Koch, J., Mahler, M. (2002) *Peptide Arrays on Membrane Supports: Synthesis and Applications*, Springer-Verlag, Heidelberg, Germany.

Kozarich, J.W. (2003) Activity-based proteomics: enzyme chemistry redux., *Curr. Opin. Chem. Biol.* **7**, 78–83.

Kumble, K.D. (2003) Protein microarrays: new tools for pharmaceutical development, *Anal. Bioanal. Chem.* **377**, 812–819.

Mantripragada, K.K., Buckley, P.G., de Stahl, T.D., Dumanski, J.P. (2004) Genomic microarrays in the spotlight, *Trends Genet.* **20**, 87–94.

Michnik, S.W. (2003) Protein fragment complementation strategies for biochemical network mapping, *Curr. Opin. Biotechnol.* **14**, 610–617.

Miller, J., Stagljar, I. (2004) Using the yeast two-hybrid system to identify interacting proteins, *Methods Mol. Biol.* **261**, 247–262.

Mirzabekov, A., Kolchinsky, A. (2002) Emerging array-based technologies in proteomics, *Curr. Opin. Chem.* **6**, 70–75.

Ng, J.H., Ilag, L.L. (2003) Biochips beyond DNA: technologies and applications, *Biotechnol. Annu. Rev.* **9**, 1–149.

Nielsen, U.B., Geierstanger, B.H. (2004) Multiplexed sandwich assays in microarray format, *J. Immunol. Methods* **90**, 107–120.

Panicker, R.C., Huang, X., Yao, S.Q. (2004) Recent advances in peptide-based microarray technologies, *Comb. Chem. High Throughput Screen.* **7**, 547–556.

Petach, H., Gold, L. (2002) Dimensionality is the issue: use of photoaptamers in protein microarrays, *Curr. Opin. Biotechnol.* **13**, 309–314.

Phelan, M.L., Nock, S. (2003) Generation of bioreagents for protein chips, *Proteomics* **3**, 2123–2134.

Predki, P.F. (2004) Functional protein microarrays: ripe for discovery, *Curr. Opin. Chem. Biol.* **8**, 8–13.

Robinson, W.H., Steinman, L., Utz, P.J. (2003) Protein arrays for autoantibody profiling and fine-specificity mapping, *Proteomics* **3**, 2077–2084.

Schena, M. (2002) *Microarray Analysis*, Wiley, New York.

Schena, M. (2004) *Protein Microarrays*, Jones & Bartlett Publishers, Boston, MA.

Skerra, A. (2000) Engineered protein scaffolds for molecular recognition, *J. Mol. Recognit.* **13**, 167–187.

Sydor, J.R., Nock, S. (2003) Protein expression profiling arrays: tools for the multiplexed high-throughput analysis of proteins, *Proteome Sci.* **1**, 3.

Templin, M.F., Stoll, D., Bachmann, J., Joss, T.O. (2004) Protein microarrays and multiplexed sandwich immunoassays: what beads the bead? *Comb. Chem. High Throughput Screen.* **7**, 223–229.

Templin, M.F., Stoll, D., Schwenk, J.M., Potz, O., Kramer, S., Joos, T.O. (2003) Protein microarrays: promising tools for proteomic research, *Proteomics* **3**, 2155–2166.

Tucker, C.L., Gera, J.F., Uetz, P. (2001) Towards an understanding of complex protein networks, *Trends Cell Biol.* **11**, 102–106.

Uttamchandani, M., Chen, G.Y.J., Lesaicherre, M.-L., Yao, S.Q. (2004) Site-specific peptide immobilization strategies for the rapid detection of kinase activity on microarrays, *Methods Mol. Biol.* **264**, 191–204.

Walhout, A.J.M., Vidal, M. (2001) High-throughput yeast two-hybrid assays for large-scale protein interaction mapping, *Methods* **24**, 297–306.

Walsh, D.P., Chang, Y.T. (2004) Recent advances in small molecule microarrays: applications and technology, *Comb. Chem. High Throughput Screen.* **7**, 557–564.

Wiesner, A. (2004) Detection of tumor markers with ProteinChip technology, *Curr. Pharm. Biotechnol.* **5**, 45–67.

Wilson, D.S., Nock, S. (2003) Recent developments in protein microarray technology, *Angew. Chem., Int. Ed.* **42**, 494–500.

Winter, G., Griffiths, A.D., Hawkins, R.E., Hoogenboom, H.R. (1994) Making antibodies by phage display technology, *Annu. Rev. Immunol.* **12**, 433–455.

Yeo, D.S., Panicker, R.C., Tan, L.P., Yao, S.Q. (2004) Strategies for immobilization of biomolecules in a microarray, *Comb. Chem. High Throughput Screen.* **7**, 213–221.

Zhou, F.X., Bonin, J., Predki, P.F. (2004) Development of functional protein microarrays for drug discovery: progress and challenges, *Comb. Chem. High Throughput Screen.* **7**, 539–546.

Zhu, H., Snyder, M. (2003) Protein chip technology, *Curr. Opin. Chem. Biol.* **7**, 55–63.

Primary Literature

Adams, E.W., Ratner, D.M., Bokesch, H.R., McMahon, J.B., O'Keefe, B.R., Seeberger, P.H. (2004) Oligosaccharide and glycoprotein microarrays as tools in HIV glycobiology: glycan-dependent gp120/protein interactions, *Chem. Biol.* **11**, 875–881.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I., Rosenwald, S.A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I.,

Yang, L., Marti, G.E., Moore, T. Jr., Hudson, J., Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Staudt, L.M. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling, *Nature* **403**, 503–511.

Anderson, L., Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver, *Electrophoresis* **18**, 533–537.

Barnes-Seeman, D., Park, S.B., Koehler, A.N., Schreiber, S.L. (2003) Expanding the functional group compatibility of small-molecule microarrays: discovery of novel calmodulin ligands, *Angew. Chem., Int. Ed.* **42**, 2376–2379.

Bernard, A., Fitzli, D., Sonderegger, P., Delamarche, E., Michel, B., Bosshard, H.R., Biebuyck, H. (2001) Affinity capture of proteins from solution and their dissociation by contact printing, *Nat. Biotechnol.* **19**, 866–869.

Bertozzi, C.R., Kiessling, L.L. (2001) Chemical glycobiology, *Science* **291**, 2357–2364.

Bhaduri, A., Sowdhamini, R. (2003) A genome-wide survey of human tyrosine phosphatases, *Protein Eng.* **16**, 881–888.

Boisguerin, P., Leben, R., Ay, B., Radziwill, G., Moelling, K., Dong, L., Volkmer-Engert, R. (2004) An improved method for the synthesis of cellulose membrane-bound peptides with free C termini is useful for PDZ domain binding studies, *Chem. Biol.* **11**, 449–459.

Borrebaeck, C.A., Ekstrom, S., Hager, A.C., Nilsson, J., Laurell, T., Marko-Varga, G. (2001) Protein chips based on recombinant antibody fragments: a highly sensitive approach as detected by mass spectrometry, *BioTechniques* **30**, 1126–1130, 1132.

Boutell, J.M., Hart, D.J., Godber, B.L.J., Kozlowski, R.Z., Blackburn, J.M. (2004) Functional protein microarrays for parallel characterization of p53 mutants, *Proteomics* **4**, 1950–1958.

Bryan, M.C., Lee, L.V., Wong, C.-H. (2004) High-throughput identification of fucosyltransferase inhibitors using carbohydrate microarrays, *Bioorg. Med. Chem. Lett.* **14**, 3185–3188.

Bulyk, M.L., Gentalen, E., Lockhart, D.J., Church, G.M. (1999) Quantifying DNA-protein interactions by double-stranded DNA arrays, *Nat. Biotechnol.* **17**, 573–577.

Bulyk, M.L., Huang, X., Choo, Y., Church, G.M. (2001) Exploring the DNA-binding specificities of zinc fingers with DNA microarrays, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7158–7163.

Butler, J.E., Ni, L., Brown, W.R., Joshi, K.S., Chang, J., Rosenberg, B., Voss, E.W. (1993) The immunochemistry of sandwich ELISAs-VI. Greater than 90% of monoclonal and 75% of polyclonal anti-fluorescyl capture antibodies (CAbs) are denatured by passive adsorption, *Mol. Immunol.* **30**, 1165–1175.

Carson, R.T., Vignali, D.A. (1999) Simultaneous quantitation of 15 cytokines using a multiplexed flow cytometric assay, *J. Immunol. Methods* **227**, 41–52.

Daugherty, P.S., Olsen, M.J., Iverson, B.L., Georgiou, G. (1999) Development of an optimized expression system for the screening of antibody libraries displayed on the Escherichia coli surface, *Protein Eng.* **12**, 613–621.

de Haard, H.J., van Neer, N., Reurs, A., Hufton, S.E., Roovers, R.C., Henderikx, P., de Bruine, A.P., Arends, J.W., Hoogenboom, H.R. (1999) A large non-immunized human Fab fragment phage library that permits rapid isolation and kinetic analysis of high affinity antibodies, *J. Biol. Chem.* **274**, 18218–18230.

Dhanasekaran, S.M., Barrette T.R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K.J., Rubin, M.A., Chinnaiyan, A.M. (2001) Delineation of prognostic biomarkers in prostate cancer, *Nature* **412**, 822–826.

Durocher, Y., Perret, S., Kamen, A. High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells, *Nucleic Acids Res.* **30**, E9.

Duveneck, G.L., Bopp, M.A., Ehrat, M., Balet, L.P., Haiml, M., Keller, U., Marowsky, G., Soria, S. (2003) Two-photon fluorescence excitation of macroscopic areas on planar waveguides, *Biosens. Bioelectron.* **18**, 503–510.

Eppinger, J., Funeriu, D.P., Miyake, M., Denizot, L., Miyake, J. (2004) Enzyme microarrays: On-chip determination of inhibition constants based on affinity-label detection of enzymatic activity, *Angew. Chem., Int. Ed.* **43**, 3806–3810.

Falsey, J.R., Renil, M., Park, S., Li, S., Lam, K.S. (2001) Peptide and small molecule microarray for high-throughput cell adhesion and functional assays, *Bioconjug. Chem.* **12**, 346–353.

Feldhaus, M.J., Siegel, R.W., Opresko, L.K., Coleman, J.R., Feldhaus, J.M., Yeung, Y.A., Cochran, J.R., Heinzelman, P., Colby, D., Swers, J., Graff, C., Wiley, H.S., Wittrup, K.D.

(2003) Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library, *Nat. Biotechnol.* **21**, 163–170.

Feng, Y. Ke, X., Ma, R., Chen, Y., Hu, G., Liu, F. (2004) Parallel detection of autoantibodies with microarrays in rheumatoid diseases, *Clin. Chem.* **50**, 416–422.

Fields, S., Song, O. (1989) A novel genetic system to detect protein-protein interactions, *Nature* **340**, 245–246.

Figeys, D. (2004) Combining different 'omics' technologies to map and validate protein-protein interactions in humans, *Brief. Funct. Genomic. Proteomic.* **2**, 357–365.

Fotin-Mleczek, M., Rottmann, M., Rehg, G., Rupp, S., Johannes, F.-J. (2000) Detection of protein-protein interactions using a green fluorescent protein-based mammalian two-hybrid system, *BioTechniques* **29**, 22–26.

Frank, R. (1992) Spot synthesis an easy technique for positionally addressable, parallel chemical synthesis on a membrane, *Tetrahedron* **48**, 9217–9232.

Frank, R. (2002) The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports-principles and applications, *J. Immunol. Methods* **267**, 13–26.

Fukui, S., Feizi, T., Galustian, C., Lawson, A.M., Chai, W. (2002) Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate-protein interactions, *Nat. Biotechnol.* **20**, 1011–1017.

Gao, X., Zhou, X., Gulari, E. (2003) Light directed massively parallel on-chip synthesis of peptide arrays with t-Boc chemistry, *Proteomics* **3**, 2135–2141.

Gavin, A.C., Bosche, M., Krause, R., et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes, *Nature* **415**, 141–147.

Golden, M.C., Collins, B.D., Willis, M.C., Koch, T.H. (2000) Diagnostic potential of PhotoSELEX-evolved ssDNA aptamers, *J. Biotechnol.* **81**, 167–178.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science* **286**, 531–537.

Gygi, S.P., Rochon, Y., Franza, B.R., Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast, *Mol. Cell. Biol.* **19**, 1720–1730.

He, M. (2004) Generation of protein in situ arrays by DiscernArray technology, *Methods Mol. Biol.* **264**, 25–31.

Hergenrother, P.J., Depew, K.M., Schreiber, S.L. (2003) Small-molecule microarrays: Covalent attachment and screening of alcohol-containing small molecules on glass slides, *J. Am. Chem. Soc.* **122**, 7849–7850.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., et al. (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry, *Nature* **415**, 180–183.

Houseman, B.T., Mrksich, M. (2002) Carbohydrate arrays for the evaluation of protein binding and enzymatic modification, *Chem. Biol.* **9**, 443–454.

Houseman, B.T., Huh, J.H., Kron, S.J., Mrksich, M. (2002) Peptide chips for the quantitative evaluation of protein kinase activity, *Nat. Biotechnol.* **20**, 270–274.

Huang, R. (2001) Detection of multiple proteins in an antibody-based protein microarray system, *J. Immunol. Methods* **255**, 1–13.

Huang, R.P., Huang, R., Fan, Y., Lin, Y. (2001) Simultaneous detection of multiple cytokines from conditioned media and patient's sera by an antibody-based protein array system, *Anal. Biochem.* **294**, 55–62.

Hubsman, M., Yudkovsky, G., Aronheim, A. (2001) A novel approach for the identification of protein-protein interaction with integral membrane proteins, *Nucleic Acids Res.* **29**, E18.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4569–4574.

Jiang, B., Snipes-Magaldi, L., Dennehy, P., Keyserling, H., Holman, R.C., Bresee, J., Gentsch, J., Glass, R.I. (2003) Cytokines as mediators for or effectors against rotavirus disease in children, *Clin. Diagn. Lab Immunol.* **10**, 995–1001.

Kanoh, N., Kumashiro, S., Simizu, S., Kondoh, Y., Hatakeyama, S., Tashiro, H., Osada, H. (2003) Immobilization of natural products on glass slides by using a photoaffinity

reaction and the detection of protein-small-molecule interactions, *Angew. Chem., Int. Ed.* **42**, 5584–5587.

Kaukola, T., Satyaraj, E., Patel, D.D., Tchernev, V.T., Grimwade, B.G., Kingsmore, S.F., Koskela, P., Tammela, O., Vainionpaa, L., Pihko, H., Aarimaa, T., Hallman, M. (2004) Cerebral palsy is characterized by protein mediators in cord serum, *Ann. Neurol.* **55**, 186–194.

Kettman, J.R., Davies, T., Chandler, D., Oliver, K.G., Fulton, R.J. (1998) Classification and properties of 64 multiplexed microsphere sets, *Cytometry* **33**, 234–243.

Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wolle, J., Pluckthun, A., Virnekas, B. (2000) Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides, *J. Mol. Biol.* **296**, 57–86.

Knezevic, V., Leethanakul, C., Bichsel, V.E., Worth, J.M., Prabhu, V.V., Gutkind, J.S., Liotta, L.A., Munson, P.J. III, Petricoin, E.F., Krizman, D.B. (2001) Proteomic profiling of the cancer microenvironment by antibody arrays, *Proteomics* **1**, 1271–1278.

Koehler, A.N., Shamji, A.F., Schreiber, S.L. (2003) Discovery of an inhibitor of a transcription factor using small molecule microarrays and diversity-oriented synthesis, *J. Am. Chem. Soc.* **125**, 8420–8421.

Koehn, M., Wacker, R., Peters, C., Schroeder, H., Soulere, L., Breinbauer, R., Niemeyer, C.M., Waldmann, H. (2003) Staudinger ligation: A new immobilization strategy for the preparation of small-molecule arrays, *Angew. Chem., Int. Ed.* **42**, 5830–5834.

Kohl, A., Binz, H.K., Forrer, P., Stumpp, M.T., Pluckthun, A., Grutter, M.G. (2003) Designed to be stable: crystal structure of a consensus ankyrin repeat protein, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1700–5.

Koide, A., Bailey, C.W., Huang, X., Koide, S. (1998) The fibronectin type III domain as a scaffold for novel binding proteins, *J. Mol. Biol.* **284**, 1141–1151.

Kramer, A., Reineke, U., Dong, L., Hoffmann, B., Hoffmuller, U., Winkler, D., Volkmer-Engert, R., Schneider-Mergener, J. (1999) Spot synthesis: observations and optimizations, *J. Pept. Res.* **54**, 319–327.

Kuruvilla, F.G., Shamji, A.F., Sternson, S.M., Hergenrother, P.J., Schreiber, S.L. (2002) Dissecting glucose signalling with diversity-oriented synthesis and small-molecule microarrays, *Nature* **416**, 653–657.

Landgraf, C., Panni, S., Montecchi-Palazzi, L., Castagnoli, L., Schneider-Mergener, J., Volkmer-Engert, R., Cesareni, G. (2004) Protein interaction networks by proteome peptide scanning, *PLoS Biol.* **2**, 94–103.

Lesaicherre, M.-L., Uttamchandani, M., Chen, G.Y.J., Yao, S.Q. (2002a) Developing site-specific immobilization strategies of peptides in a microarray, *Bioorg. Med. Chem. Lett.* **12**, 2079–2083.

Lesaicherre, M.-L., Uttamchandani, M., Chen, G.Y.J., Yao, S.Q. (2002b) Antibody-based fluorescence detection of kinase activity on a peptide array, *Bioorg. Med. Chem. Lett.* **12**, 2085–2088.

Lin, Y., Huang, R., Santanam, N., Liu, Y.G., Parthasarathy, S., Huang, R.P. (2002) Profiling of human cytokines in healthy individuals with vitamin E supplementation by antibody array, *Cancer Lett.* **187**, 17–24.

MacBeath, G., Koehler, A.N., Schreiber, S.L. (1999) Printing small molecules as microarrays and detecting protein-ligand interactions en masse, *J. Am. Chem. Soc.* **121**, 7967–7968.

MacBeath, G., Schreiber, S.L. (2000) Printing proteins as microarrays for high-throughput function determination, *Science* **289**, 1760–1763.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S. (2002) The protein kinase complement of the human genome, *Science* **298**, 1912–1934.

Martin, K., Steinberg, T.H., Cooley, L.A., Gee, K.R., Beechem, J.M., Patton, W.F. (2003) Quantitative analysis of protein phosphorylation status and protein kinase activity on microarrays using a novel fluorescent phosphorylation sensor dye, *Proteomics* **3**, 1244–1255.

Mendoza, L.G., McQuary, P., Mongan, A., Gangadharan, R., Brignac, S., Eggers, M. (1999) High-throughput microarray-based enzyme-linked immunosorbent assay (ELISA), *BioTechniques* **27**, 778–788.

Miller, J.C., Zhou, H., Kwekel, J., Cavallo, R., Burke, J., Butler, E.B., The, B.S., Haab, B.B. (2003) Antibody microarray profiling of human prostate cancer sera: Antibody screening and identification of potential biomarkers, *Proteomics* **3**, 56–63.

Moody, M.D., van Ardel, S.W., Orencole, S.F., Burns, C. (2001) Array-based ELISAs for high-throughput analysis of human cytokines, *BioTechniques* **31**, 186–194.

Neuman de Vegvar, H.E., Amara, R.R., Steinman, L., Utz, P.J., Robinson, H.L., Robinson, W.H. (2003) Microarray profiling of responses against simian-human immunodeficiency virus: postchallenge convergence of reactivities independent of host histocompatibility type and vaccine regime, *J. Virol.* **77**, 11125–11138.

Nielsen, T.O., West, R.B., Linn, S.C., Alter, O., Knowling, M.A., O'Connell, J.X., Zhu, S., Fero, M., Sherlock, G., Pollack, J.R., Brown, P.O., Botstein, D., van de Rijn, M. (2002) Molecular characterisation of soft tissue tumours: a gene expression study, *Lancet* **359**, 1301–1307.

Nord, K., Gunneriusson, E., Ringdahl, J., Stahl, S., Uhlen, M., Nygren, P.A. (1997) Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain, *Nat. Biotechnol.* **15**, 772–777.

Pawlak, M., Schick, E., Bopp, M.A., Schneider, M.J., Oroszlan, P., Ehrat, M. (2002) Zeptosens' protein microarrays: a novel high performance microarray platform for low abundance protein analysis, *Proteomics* **2**, 383–393.

Pellois, J.P., Zhou, X., Srivannavit, O., Zhou, T., Gulari, E., Gao, X. (2002) Individually addressable parallel peptide synthesis on microchips, *Nat. Biotechnol.* **20**, 922–926.

Peluso, P., Wilson, D.S., Do, D., Tran, H., Venkatasubbaiah, M., Quincy, D., Heidecker, B., Poindexter, K., Tolani, N., Phelan, M., Witte, K., Jung, L.S., Wagner, P., Nock, S. (2003) Optimizing antibody immobilization strategies for the construction of protein microarrays, *Anal. Biochem.* **312**, 113–124.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S.X., Lonning, P.E., Borresen-Dale, A.L., Brown, P.O., Botstein, D. (2000) Molecular portraits of human breast tumours, *Nature* **406**, 747–752.

Petricoin, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone, C., Fishman, D.A., Kohn, E.C., Liotta, L.A. (2002) Use of proteomic patterns in serum to identify ovarian cancer, *Lancet* **359**, 572–577.

Pomeroy, S.L., Tamayo, P., Gaasenbeek, M., Sturla, L.M., Angelo, M., McLaughlin, M.E., Kim, J.Y., Goumnerova, L.C., Black, P.M., Lau, C., Allen, J.C., Zagzag, D., Olson, J.M., Curran, T., Wetmore, C., Biegel, J.A., Poggio, T., Mukherjee, S., Rifkin, R., Califano, A., Stolovitzky, G., Louis, D.N., Mesirov, J.P., Lander, E.S., Golub, T.R. (2002) Prediction of central nervous system embryonal tumour outcome based on gene expression, *Nature* **415**, 436–442.

Qiu, J., Madoz-Gurpide, J., Misek, D.E., Kuick, R., Brenner, D.E., Michailidis, G., Haab, B.B., Omenn, G.S., Hanash, S. (2004). Development of natural protein microarrays for diagnosing cancer based on an antibody response to tumor antigens, *J. Proteome Res.* **3**, 261–267.

Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S., Rosen, B., Lau, A.Y., Walter, J.C., LaBaer, J. (2004) Self-assembling protein microarrays, *Science* **305**, 86–90.

Robertson, M.P., Ellington, A.D. (2001) In vitro selection of nucleoprotein enzymes, *Nat. Biotechnol.* **19**, 650–655.

Rowe, C.A., Scruggs, S.B., Feldstein, M.J., Golden, J.P., Ligler, F.S. (1999) An array immunosensor for simultaneous detection of clinical analytes, *Anal. Chem.* **71**, 433–439.

Rowe, C.A., Tender, L.M., Feldstein, M.J., Golden, J.P., Scruggs, S.B., MacCraith, B.D., Cras, J.J., Ligler, F.S. (1999) Array biosensor for simultaneous identification of bacterial, viral, and protein analytes, *Anal. Chem.* **71**, 3846–3852.

Salisbury, C.M., Maly, D.J., Ellman, J.A. (2002) Peptide microarrays for the determination of protease substrate specificity, *J. Am. Chem. Soc.* **124**, 14868–14870.

Schaffitzel, C., Hanes, J., Jermutus, L., Pluckthun, A. (1999) Ribosome display: an in vitro method for selection and evolution of antibodies from libraries, *J. Immunol. Methods* **231**, 119–135.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science* **270**, 467–470.

Schweitzer, B., Wiltshire, S., Lambert, J., O'Malley, S., Kukanskis, K., Zhu, Z., Kingsmore, S.F., Lizardi, P.M., Ward, D.C. (2000) Immunoassays with rolling circle DNA amplification: a versatile platform for

ultrasensitive antigen detection, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10113–10119.

Schweitzer, B., Roberts, S., Grimwade, B., Shao, W., Wang, M., Fu, Q., Shu, Q., Laroche, I., Zhou, Z., Tchernev, V.T., Christiansen, J., Velleca, M., Kingsmore, S.F. (2002) Multiplexed protein profiling on microarrays by rolling-circle amplification, *Nat. Biotechnol.* **20**, 359–365.

Sheets, M.D., Amersdorfer, P., Finnern, R., Sargent, P., Lindquist, E., Schier, R., Hemingsen, G., Wong, C., Gerhart, J.C., Marks, J.D., Lindqvist, E. (1998) Efficient construction of a large nonimmune phage antibody library: the production of high-affinity human single-chain antibodies to protein antigens, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6157–6162.

Soldan, S.S., Alvarez-Retuerto, A.I., Sicotte, N.L., Voskuhl, R.R. (2004) Dysregulation of IL-10 and IL-12p40 in secondary progressive multiple sclerosis, *J. Neuroimmunol.* **146**, 209–215.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Thorsen, T., Quist, H., Matese, J.C., Brown, P.O., Botstein, D., Eystein Lonning, P., Borresen-Dale, A.L. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10869–10874.

Sreekumar, A., Nyati, M.K., Varambally, S., Barrette, T.R., Ghosh, D., Lawrence, T.S., Chinnaiyan, A.M. (2001) Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins, *Cancer Res.* **61**, 7585–7593.

Stillman, B.A., Tonkinson, J.L. (2000) FAST slides: a novel surface for microarrays, *BioTechniques* **29**, 630–635.

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., et al. (2002) Generation and initial analysis of more than 15 000 full-length human and mouse cDNA sequences, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16899–16903.

Sydor, J.R., Mariano, M., Sideris, S., Nock, S. (2002) Establishment of intein-mediated protein ligation under denaturing conditions: C-terminal labeling of a single-chain antibody for biochip screening, *Bioconjug. Chem.* **13**, 707–712.

Sydor, J.R., Scalf, M., Sideris, S., Mao, G.D., Pandey, Y., Tan, M., Mariano, M., Moran, M.F., Nock, S., Wagner, P. (2003) Chip-based analysis of protein-protein interactions by fluorescence detection and on-chip immunoprecipitation combined with microLC-MS/MS analysis, *Anal. Chem.* **75**, 6163–6170.

Tam, S.W., Wiese, R., Lee, S., Gilmore, J., Kumble, K.D. (2002) Simultaneous analysis of eight human Th1/Th2 cytokines using microarrays, *J. Immunol. Methods* **261**, 157–165.

Tannapfel, A., Anhalt, K., Hausermann, P., Sommerer, F., Benicke, M., Uhlmann, D., Witzigmann, H., Hauss, J., Wittekind, C. (2003) Identification of novel proteins associated with hepatocellular carcinomas using protein microarrays, *J. Pathol.* **201**, 238–249.

Tegge, W.J., Frank, R. (1998) Analysis of protein kinase substrate specificity by the use of peptide libraries on cellulose paper (SPOT-method), *Methods Mol. Biol.* **87**, 99–106.

Toepert, F., Knaute, T., Guffler, S., Pires, J.R., Matzdorf, T., Oschkinat, H., Schneider-Mergener, J. (2003) Combining SPOT synthesis and native peptide ligation to create large arrays of WW protein domains, *Angew. Chem., Int. Ed.* **42**, 1136–1140.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., Rothberg, J.M. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae, *Nature* **403**, 623–627.

Uttamchandani, M., Chan, E.W.S., Chen, G.Y.J., Yao, S.Q. (2003) Combinatorial peptide microarrays for the rapid determination of kinase specificity, *Bioorg. Med. Chem. Lett.* **13**, 2997–3000.

Vaughan, T.J., Williams, A.J., Pritchard, K., Osbourn, J.K., Pope, A.R., Earnshaw, J.C., McCafferty, J., Hodits, R.A., Wilton, J., Johnson, K.S. (1996) Human antibodies with subnanomolar affinities isolated from a large non-immunized phage display library, *Nat. Biotechnol.* **14**, 309–314.

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., Bork, P. (2002) Comparative assessment of large-scale data

sets of protein-protein interactions, *Nature* **417**, 399–403.

Wahler, D., Badalassi, F., Crotti, P., Reymond, J.-L. (2001) Enzyme fingerprints by fluorogenic and chromogenic substrate arrays, *Angew. Chem. Int. Ed.* **40**, 4457–4460.

Wang, D., Liu, S., Trummer, B.J., Deng, C., Wang, A. (2002) Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells, *Nat. Biotechnol.* **20**, 275–281.

Wiese, R., Belosludtsev, Y., Powdrill, T., Thompson, P., Hogan, M. (2001) Simultaneous multianalyte ELISA performed on a microarray platform, *Clin. Chem.* **47**, 1451–1457.

Winssinger, N., Ficarro, S., Schultz, P.G., Harris, J.L. (2002) Profiling protein function with small molecule microarrays, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11139–11144.

Xu, L., Aha, P., Gu, K., Kuimelis, R., Kurz, M., Lam, T., Lim, A., Liu, H., Lohse, P., Sun, L., Weng, S., Wagner, R., Lipovsek, D. (2002) Directed evolution of high-affinity antibody mimics using mRNA display, *Chem. Biol.* **9**, 933–942.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R.A., Gerstein, M., Snyder, M. (2001) Global analysis of protein activities using proteome chips, *Science* **293**, 2101–2105.

Zhu, Q., Uttamchandani, M., Li, D., Lesaicherre, M.L., Yao, S.Q. (2003) Enzymatic profiling in a small-molecule microarray, *Org. Lett.* **5**, 1257–1260.

Zhu, H., Klemic, J.F., Chang, S., Bertone, P., Casamayor, A., Klemic, K.G., Smith, D., Gerstein, M., Reed, M.A., Snyder, M. (2000) Analysis of yeast protein kinases using protein chips, *Nat. Genet.* **26**, 283–289.

Protein Modeling

Marian R. Zlomislic and D. Peter Tieleman
University of Calgary, Albeta, Canada

1 Introduction 136
1.1 Primary, Secondary, Tertiary, and Quaternary Structure of Proteins 136
1.2 Relationship between Structure and Function 136
1.3 Experimental Structure Determination Methods 138
1.4 Role of Modeling 139

2 Structure Prediction Methods 139
2.1 The Protein-folding Problem 139
2.2 *De Novo/Ab Initio* Methods 140
2.3 Fold Recognition Methods 141
2.4 Homology Modeling 142

3 Structure-based Modeling 143
3.1 Molecular Graphics-based Methods 143
3.2 Poisson–Boltzmann/Electrostatics Calculations 144
3.3 Ligand Docking 145
3.4 Protein–Protein Interactions 146

4 Simulations of Protein Dynamics 146
4.1 Molecular Dynamics Simulation 146
4.2 Free-energy Calculations 148
4.3 Brownian Dynamics Calculations 148

5 Example Applications 149
5.1 Acetylcholinesterase 149
5.2 Water Transport Through Aquaporins 150
5.3 ABC Transporters and Multidrug Resistance Proteins 152

6 Perspectives 154

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

Acknowledgment 155

Bibliography 156
Books and Reviews 156
Primary Literature 156

Keywords

***Ab Initio* Modeling**
The goal of *ab initio* modeling is to predict the structure of a protein from only its amino acid sequence.

Brownian Dynamics Simulation
Brownian dynamics simulations are related to molecular dynamics, but do not include the same amount of detail. Solvent in particular is normally represented as a continuous medium, which significantly decreases the complexity of the models and allows much longer simulations.

CASP, CAFASP, CAPRI
Critical Assessment of Structure Prediction (CASP), Critical Assessment of Fully Automated Structure Prediction (CAFASP), and Critical Assessment of PRedicted Interactions (CAPRI) are large-scale assessments in which hundreds of research groups submit their best structure predictions on the same set of targets, allowing a critical comparison between methods and an assessment of strengths and weaknesses of different methods and the field as a whole.

Docking
The term docking incorporates a group of methods for studying the interactions of small molecules and proteins. The process typically involves extensive searching of possible orientations of the small molecule in a binding site of the protein, and a scoring function that ranks different orientations by their likelihood of occurrence.

Fold Recognition
The goal of fold recognition is to predict which fold, out of ca. 1000 observed protein folds, a given sequence will adopt.

Homology Modeling
Predicting the three-dimensional structure of a protein on the basis of known structure of a related protein.

Molecular Dynamics Simulation
A computer simulation method in which the motions of all atoms in a molecular model, including water and ions are calculated over a period of up to a microsecond.

This computationally expensive method is one of the most detailed ways of studying the dynamics of macromolecules.

Molecular Graphics
Modern computer graphics are used to visualize molecular structures and models. The ability to highlight different features or properties in a three-dimensional model is a powerful way to explore molecules. Stereo hardware and immersive environments further enhance the usefulness of molecular graphics.

Poisson–Boltzmann Equation
The Poisson–Boltzmann equation describes electrostatic interactions between charges in a continuous medium (e.g. water) in the presence of salt. It can be solved numerically by several popular software packages to give insight into electrostatic interactions in proteins, between proteins and their substrates, between proteins and membranes, and in other systems. A common application is the calculation of pK_a-shifts for acidic and basic residues in proteins due to their environment.

Primary, Secondary, Tertiary, and Quaternary Protein Structure
The amino acid sequence of a protein is its primary structure. The α-helices, β-strands, and random coils are typical secondary structures that form between consecutive residues. They are defined by the hydrogen-bonding pattern and dihedral angles observed in the protein backbone. The tertiary structure of a protein describes the relative position of all the protein atoms in three-dimensional space. Elements of secondary structure that were far apart sequentially may now form higher order structures such as parallel or antiparallel β-sheets, or helix bundles. The quaternary structure describes how multiple proteins are arranged together in larger complexes.

QM/MM Simulation
A computer simulation method in which an important part of a system, such as the active site of an enzyme, is described by quantum mechanics, with the rest of the system described by molecular dynamics methods. The part of the system treated by QM includes explicit treatment of electrons.

Structural Genomics
A concerted effort to determine experimentally the high-resolution structure of all proteins in a genome, often with a focus on proteins that are predicted to be structurally different from all structures that are already in the database. This greatly increases the likelihood that a given sequence can be modeled using homology modeling.

Protein modeling consists of a broad range of computational techniques to understand the properties of proteins and has become an integral part of structural biology and drug design. Modeling can be used to predict the secondary structure or fold of a protein on the basis of its sequence alone, to predict the three-dimensional structure of a protein on the basis of knowledge of the structure of a related protein,

to design new proteins, and to predict properties that depend on the experimentally determined three-dimensional structure of a protein. Examples of such properties include drug binding, protein–protein interactions, and interactions with elements in a protein's environment, including ions, lipids, carbohydrates, and nucleic acids. Conformational changes in proteins can be investigated by molecular dynamics simulations to provide detailed insight into the dynamics of proteins, a crucial aspect of protein function. In recent developments, quantum mechanical calculations are used more and more frequently to study reactions in proteins. With the ever-increasing power of computers, increasingly detailed aspects of protein function can now be investigated by modeling methods, at a scale and level of detail that is often very difficult or impossible to achieve by experimental methods.

In this chapter, the main principles and techniques involved in protein modeling are introduced. A few examples from the literature will highlight how protein modeling can be used in complement with other methods.

1 Introduction

1.1 Primary, Secondary, Tertiary, and Quaternary Structure of Proteins

Proteins are structured at different levels. The primary structure of a protein is its amino acid sequence, encoded by DNA. All proteins are constructed from about 20 common amino acids as well as some amino acids that are formed through chemical modification. The secondary structure of a protein is a sequence of common structural, three-dimensional elements or building blocks. These secondary structure elements include α-helices and β-strands as best-known elements, but there are many other building blocks. The tertiary structure of a protein is its three-dimensional structure in space, which can be thought of as placing the secondary structure elements together in space. The quaternary structure of a protein describes how multiple proteins interact through noncovalent forces to form larger protein complexes. An example of the different levels of structure is shown in Fig. 1, based on the protein barstar. It is shown in a protein complex with barnase to illustrate quaternary structure. The minimal requirement for modeling proteins is knowledge of the primary structure. Since the primary sequence is one of the main results of genome sequencing, this is a very modest requirement. As more experimental structural information becomes available, modeling typically becomes more accurate by using that experimental information to identify plausible models. Any experimental information that puts limitations on possible models is useful.

1.2 Relationship between Structure and Function

A basic assumption in protein science is that the function of a protein follows from its structure. Yet, proteins with very similar three-dimensional structure can have very different functions, so

MKKAVINGEQIRSISDLHQTLKKELALPEYYGENLAALWDCLTGWVEYPLVLEWRQ
FEQSKQLTENGAESVLQVFREAKAEGCDITIILS
(a)

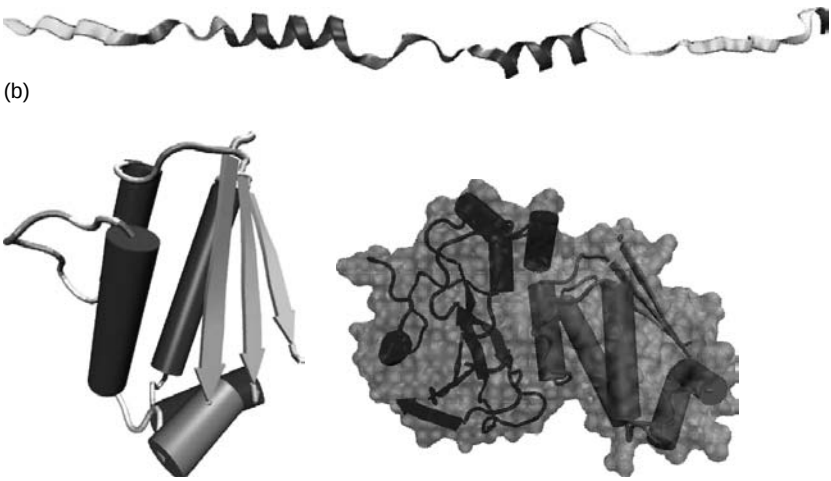

(b)

(c) (d)

Fig. 1 (a) The primary sequence of the protein barstar in FASTA format. (b) The secondary structure of the first 60 residues of barstar, colored purple for helices, and yellow for β-strands. All are shown in ribbon format. (c) The complete tertiary structure of barstar in cartoon format. The helices are now bundled together, and the β-strands have formed a three-stranded parallel β-sheet. (d) Example of quaternary structure: the barnase/barstar complex. Barstar (red) and barnase (blue) area shown in ribbon and spacefilling formats. Images created with VMD using the PDB entry 1B2U. (See color plate p. xxx.)

precise details of the three-dimensional (tertiary/quaternary) structure are very important. Nonetheless, there is a large amount of information contained in the primary and secondary structure of a protein. Indeed, nature in most cases only seems to need the primary sequence, but computational methods are not yet able to reliably predict the three-dimensional structure of proteins based on primary sequence alone. At the current state of the art, a primary sequence is useful to compare proteins of known function with a similar sequence and sometimes allows assignment of a tentative function. Knowing the secondary structure of a protein significantly narrows down the range of possible three-dimensional structures. It allows a comparison with databases of the secondary sequence of known proteins, which often makes it possible to recognize the three-dimensional fold of a protein on the basis of the secondary structure alone. Interestingly, although in principle the number of possible three-dimensional structures is practically infinite, only about 1000 different "folds" seem to occur in nature. This is particularly useful for whole-genome studies: if we can use the primary sequence and predictions or knowledge of the secondary structure, chances are reasonable that the fold of the protein can be recognized. This is a

constructive step toward attaining a three-dimensional structure.

The three-dimensional structure of a protein is the most useful level for molecular modeling. At this level, the interactions between small molecules (drugs, substrate) and an enzyme, permeation properties of ion channels and transporters, activation of receptors by ligand binding, and chemical reactions in enzymes and other processes that require a very detailed knowledge of atomic structure to understand them can be studied. This level is also required to study interactions between protein subunits or for protein–protein or protein–peptide interactions. In practice, the best way to obtain knowledge of the three-dimensional structure is from experimental structure determination, but prediction methods primarily based on homology modeling (see Sect. 2) are increasingly becoming accurate enough to provide useful starting points for further computational studies.

1.3 Experimental Structure Determination Methods

There are several experimental methods to determine the structure of proteins. The primary structure is determined by sequencing, either of the protein itself or of the DNA that encodes the protein. The secondary structure can be measured by spectroscopic methods, which usually determine a percentage of various secondary structure elements. Common techniques to do this are CD (circular dichroism) spectroscopy, IR (infrared) spectroscopy, and NMR (nuclear magnetic resonance) spectroscopy.

As a basis for detailed modeling problems, high-resolution 3-dimensional structures are typically required. The two main methods to experimentally determine 3D structures are X-ray crystallography and nuclear magnetic resonance. Electron microscopy is a third method that has been used to solve the structure of several membrane proteins. Crystallography is a very powerful method that can be used on proteins and protein complexes of any size. The main limitation of the method is that a protein must form regular crystals. Not all proteins do this (especially membrane proteins), and sometimes crystallization forces proteins in structures that are probably not physiologically relevant. NMR experiments can determine the structure of a protein in solution, which is usually a more realistic environment. The main limitations of solution NMR are that it is difficult to apply the method in practice to proteins that are larger than ca. 40 kDa, and that it is usually considerably more labor intensive to determine a structure by NMR compared to crystallography. These limitations have spurred the development of new methods that are catered to work with larger proteins, and to make the process of structure calculation from a measured spectrum more automated. Solid-state protein NMR is a relatively new field that can be used to determine the structure of proteins in ordered systems, such as membranes. Electron microscopy as a method to determine 3-D structures has been mainly used for membrane proteins, which are especially difficult to crystallize for X-ray crystallography. The method is very labor intensive. It does have a significant strength for very large complexes, where the overall arrangement of large subunits can be seen at low resolution. The details of the structure can then be added from crystal structures of individual proteins that make up the complex.

1.4 Role of Modeling

Modeling has many potential uses, at different levels of protein structure. Protein modeling is such a broad field it may be beneficial to distinguish two separate general goals.

One set of goals consists of predicting aspects of the structure of a protein based on less detailed information. If only the primary structure (the sequence) of a protein is known, we can try to predict the secondary structure with reasonable accuracy, often in the 75 to 85% range. On the basis of the degree of hydrophobicity of predicted helices, it is usually also possible to predict transmembrane segments, identifying a protein as a membrane protein, with reasonable accuracy. The secondary structure can be useful to predict the general fold of a protein, which may make it possible to assign a tentative function to the protein. Predicting the tertiary structure directly from the sequence is a very difficult problem with a low success rate. If a high-resolution structure of a related protein is known, then homology modeling can be used to provide a good model of the target protein. In principle, such modeling efforts make it possible to circumvent experimental structure determination. One major goal of protein structure research at the moment is to improve all steps in the process that leads to an accurate three-dimensional model, both by improving *ab initio* prediction methods and by experimentally determining enough protein structure that homology modeling becomes feasible for most proteins in the genome – ideally, one could then obtain structural models for all proteins in an organism directly from its genome sequence. As the number of structures solved experimentally increases, the algorithms to detect homology between remotely related proteins and built homology models will improve, making this goal increasingly feasible.

A second set of goals starts from an experimentally determined high-resolution structure (or a very high-quality homology model) and uses physics-based methods to model aspects of a protein that cannot be easily determined experimentally. This type of approach could include electrostatics calculations to understand how proteins interact with substrates or other proteins, docking calculations to understand differences in binding constants and design new inhibitors for enzymes, or detailed molecular dynamics simulations to investigate the dynamics of proteins. This type of goal usually involves much more biochemical knowledge of the protein and is often centered on very specific questions: why does mutating residue x to alanine change the binding affinity of a particular drug by a factor of 1000? Why is this ion channel selective for potassium over sodium, even though the only difference between potassium and sodium superficially is a small difference in radius?

In the following sections we consider each set of goals in more detail and give specific examples.

2 Structure Prediction Methods

2.1 The Protein-folding Problem

One of the most challenging problems in biophysics is to understand how it is possible that proteins fold rapidly (microseconds to seconds) into a well-defined structure when based on their primary sequence alone, a practically infinite number

of structures is possible. Levinthal showed that folding on a realistic timescale could therefore not occur through a systematic search of all possible conformations (Levinthal's paradox). Major research efforts are devoted to understanding protein folding, with a significant focus on the physics of simplified models. Clearly, if it is possible to fold a protein *in vivo* from the sequence only, then with a proper understanding of the laws governing protein folding, we should be able to predict any protein structure computationally. This is still a lofty goal, but in the past years major progress has been made in improving prediction methods at all levels of structure. This progress has been documented strikingly in the proceedings of the semiannual CASP competition (Critical Assessment of Structure Prediction).

CASP is an interesting and exciting venture involving a large scientific community. Its main goal is to obtain an in-depth and objective assessment of our current abilities and inabilities in the area of protein structure prediction (http://predictioncenter.llnl.gov). It involves a neutral organizing committee that collects unpublished, but already solved protein structures from experimental structure determination groups. Specific information, such as the protein's primary sequence, is published on a Web site, and modeling groups can attempt to predict the structures. A panel of assessors judges the predictions against the real structures, and at the CASP scientific meeting the results are discussed, areas of progress and areas of problems are identified, and the progress and directions of the field as a whole are examined. This procedure makes CASP a very fair way of establishing the merits of particular methods and of identifying where future efforts can be most productively focused. CASP6 was held in December 2004 in Italy. In the last few years, CASP has begun to introduce automated prediction software as a separate exercise (CAFASP – Critical Assessment of Fully Automated Structure Prediction). The process has also been adapted to predicting protein–protein interactions (CAPRI – Critical Assessment of PRediction of Interactions), which saw its first edition in 2001. More details about CAPRI can be found at http://capri.ebi.ac.uk. These competitions give a novice in the field of structure prediction an excellent starting point to compare the many protein modeling programs and help a user to decide which program would work best with their requirements. Despite differences in the performance of these programs, many of them draw from the same fundamental approaches to protein structure prediction. The three main approaches are discussed in the following.

2.2 *De Novo/Ab Initio* Methods

De novo methods are a general designation for methods that predict the structure of a protein on the basis of the primary sequence alone. The ideal method to do this would be based on simple physical laws only. Although this would be a true *de novo* method, it is exceedingly difficult and has only been modestly successful for peptides and small proteins. The difficulty is the extremely large number of possible conformations for a polypeptide chain; predicting the structure of a protein from the sequence only would indeed be equivalent to solving the protein-folding problem. In practice, *de novo* methods incorporate a variety of information derived from a database

of existing structures. This can be very direct information, such as taking the conformation of a loop with the same sequence directly from the database, or less direct, such as through heuristical rules derived from statistical analysis of all known protein structures.

2.3 Fold Recognition Methods

Fold recognition methods attempt to identify structures within a new protein (the query) by comparing its sequence to proteins with known structure (the template). This can be called the *inverse folding problem*, since the goal is to find probable folds that might fit the sequence instead of trying to determine how the sequence will fold.

There are two broad approaches to fold recognition methods. The first approach is sequence-based. It relies on finding homologous sequences and then assumes that strong sequence similarity equates to strong structure and function similarity. While this may seem straightforward, current algorithms may not be sensitive enough to recognize distant homologs, and therefore the "correct" template might be missed; as well, mutations between these sequences, including gaps and insertions, sometimes make it difficult to identify the correct sequence alignment. It can be misleading to rely on the sequence alignment alone for fold recognition. It is possible to identify proteins with some sequence identity that do not have any structural or functional similarity to the query, and so, blind sequence alignments may not yield a successful prediction. Classifying protein sequences into families, and then identifying the patterns within the family such as strongly conserved residues (which might be important to function) and the pattern of mutations (For example, is a hydrophobic residue always replaced by another hydrophobic residue? That is an acceptable mutation.) can improve the confidence of your alignment. PSI-BLAST is one of the most popular homology recognition algorithms, although there are a number of others that also perform well.

The second approach to fold recognition is structure-based. "Threading" is a popular term for this approach, which evokes the image of the thread-like primary sequence being "threaded" through the three-dimensional structure of template proteins. Finding the optimal fit of the sequence with each framework is a complex process, particularly if there are gaps and deletions among the sequences of the two proteins. It is very common to use sequence-based methods in conjunction with threading to identify the best matches. Each iteration of the threading process is scored and this score is used to determine the best matches. Following Anfinsen's hypothesis that the native state of the protein is the lowest energy structure, the evaluated energy of the query's fit to the template is a strong predictor of which models are "good" within an experiment. There are a number of threading algorithms available, varying in their choice of protein model (backbone atoms vs just α-carbon atoms, side chains vs interaction sites), their method of alignment, as well as their method of evaluating the energy of the proposed structure, just to name a few differences.

When embarking on a fold recognition study, the state of the art of current fold recognition methods can be judged from results of the previously discussed CASP competition, and also of the Live Bench Project. Unlike CASP, the targets in this exercise are recently deposited structures

from the Protein Data Bank. All of the methods assessed are fully automated. An expanding number of fold recognition servers (as well as metapredictors and sequence comparison servers) participate in this weekly assessment exercise, and the most current results can be found on the Web at http://bioinfo.pl/livebench/. It is important to recognize that fold recognition methods are mostly knowledge-based. They are making predictions based only on what is already known. This means that it is impossible to identify any novel folds. Novel fold prediction would fall into the category of an *ab initio* method, where the predictions are based primarily on physical principles.

2.4 Homology Modeling

Homology modeling is the only generally useful method to predict the total three-dimensional structure of a protein. It is based on the observation that the structure of a protein during evolution is much more conserved than the sequence, so that sequences that differ somewhat are still likely to have the same structure. "Somewhat" can be quantified on the basis of the database of known structures: if two proteins with more than ca. 50 residues have a sequence identity of ca. 20 to 30%, they will likely have the same structure. Thus, homology modeling can be used if there is a protein with a known structure (the template) that is sufficiently homologous to the protein whose structure we want to model (the target). In practice, homology modeling follows a series of steps, each of which involve choices by the modeler, or an automated computer server. Following Krieger et al., homology modeling can be thought of as involving seven steps:

1. Template recognition and initial alignment
2. Alignment correction
3. Backbone generation
4. Loop modeling
5. Side-chain modeling
6. Model optimization
7. Model validation

The overall strategy is to begin with aligning the sequence of the target with that of the template, to build the overall chain structure of the target, and then to consider "details." The accuracy of this alignment is the most important determinant of the accuracy of the model. Step 2 is a manual correction of the alignment if there is a compelling reason to do so. For example, in potassium channels there is a nearly universally conserved motif TVGYG that is likely to be aligned even if multiple-sequence alignment algorithms offer alternative alignments. The backbone of the target is typically initially copied from the template, followed by algorithms that allow some flexible adjustments where necessary. The major problems in homology modeling typically involve insertions that are not present in the template. In potassium channels, the loops between transmembrane helices can be very different between channels, even if the transmembrane segments are very similar. Similarly, the loops connecting the transmembrane helices in G-protein coupled receptors convey much of the specificity of different receptors, but this most interesting part generally cannot be modeled by homology modeling. Thus, being able to model loops for the targets that are missing in the template is an important step, and a significant research focus. Once the backbone has been optimized as much as possible, side chains can be modeled and the model tested against experimental

information. In many cases, steps have to be repeated.

It is now usually fairly easy to generate homology models, provided a suitable template is available. Commonly used programs include Modeller and WhatIf, as well as automated Web servers such as SwissModel. The key measure of success for homology models is their experimental validation, but less than perfect models can be very useful in interpreting experiments and guiding the design of new experiments.

3 Structure-based Modeling

3.1 Molecular Graphics-based Methods

Over the past 5 to 10 years, the personal computer has begun to replace the large supercomputers and dedicated workstations once needed to use molecular visualization tools. Today, a number of molecular graphics programs are available for use on personal computers, often as freeware. Any program with basic rendering capabilities can be a valuable tool to study a particular protein. A variety of representations, illustrated in Fig. 2 for the protein barnase, can help the modeler understand the link between the protein's structure and its function. By manipulating colors and representations, one can interrogate the structure to see where conserved residues/sequences are in the protein's structure, look at the overall surface of the protein as a guide to potential protein–protein interactions, and look for cavities that might be ligand-docking sites, just to name a few examples. In this section, we aim to point out a few commonly used programs for protein

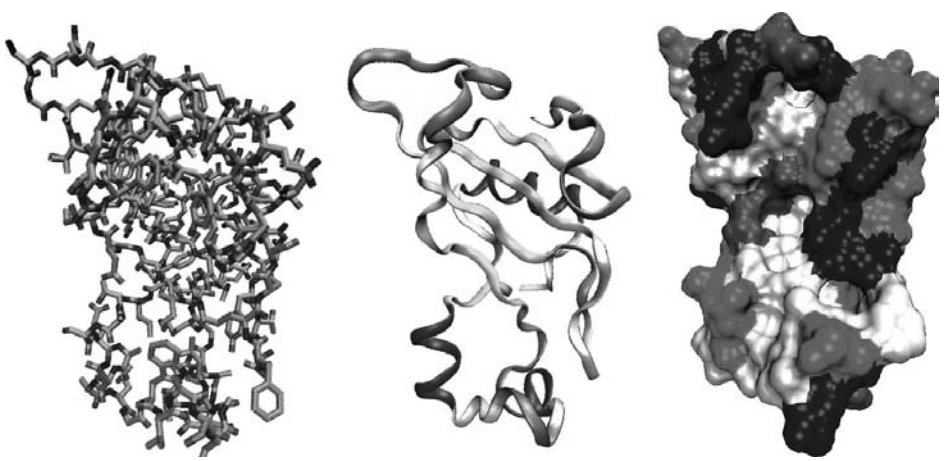

(a) (b) (c)

Fig. 2 Barnase rendered in three formats: (a) colored by atom-type (carbon – blue, nitrogen – navy blue, oxygen – red); (b) colored by secondary structure features in ribbon format, where β-sheets are colored yellow, helices are colored purple, and turns and random coil features are colored blue and white; (c) rendered in surface format, colored by residue type, where charged residues are colored red or blue, polar residues are green, and nonpolar residues are white. All figures are rendered with the same view of the protein, looking into the barstar binding pocket. All figures were rendered with VMD. (See color plate p. xxxii.)

modeling. Besides those mentioned here, a few more are described by Tate, and even more are listed on the PDB Web site (http://www.rcsb.org/pdb/software-list.html).

Rasmol is one of the most widely used programs for molecular visualization. Besides the basic options in which the display of the protein can be altered from sticks to spacefilling to ribbon format, and so on, there are a variety of tools that can be used through the command line interface. Examples include selection of specific residues of the protein, measuring distances, and highlighting parts of the protein within a certain radius. Rasmol can be used to probe all of the information contained in a ".pdb file," but it has few tools to alter that information. With Swiss-PDB Viewer, protein structures can be manipulated in a number of ways. Residues can be "mutated," and dihedral angles can be modified. This program provides a graphical interface within which several protein structures can be analyzed and manipulated at the same time. Moreover, Swiss-PdbViewer is tightly linked to SwissModel, an automated homology-modeling server. Working with these two programs greatly reduces the amount of work necessary to generate models, as it is possible to thread a protein primary sequence onto a three-dimensional template and get an immediate feedback of how well the threaded protein will be accepted by the reference structure before submitting a request to build missing loops and refine side-chain packing.

VMD is another very powerful visualization program, which can also be used to view trajectories from molecular dynamics simulations. It has a broad range of representations for proteins, including molecular surfaces, electrostatic potential maps, and crystal information. Other useful programs are PyMol and MolMol.

3.2 Poisson–Boltzmann/Electrostatics Calculations

Poisson–Boltzmann theory (PB) has become an important tool for studying biomolecular systems. This type of calculation provides a view of a protein in terms of regions of positive or negative potential. PB calculations are useful for identifying interaction sites on proteins with charged ligands or other proteins. They can also be used for calculating amino acid pK_as in different environments, such as in enzyme active sites where catalytic residues often have large pK_a shifts due to their local environment. PB calculations also provide a basis for Brownian dynamics simulations (see Sect. 4).

PB is based on the Poisson equation, which describes electrostatic interactions in general, but includes implicitly the presence of ions in solution. It is assumed that at equilibrium, the distribution of mobile ions in the system can be approximated by a continuous charge density $\rho_{eq}(\mathbf{r})$, determined by the Boltzmann factor:

$$\rho_{eq}(\mathbf{r}) = \sum_i z_i e n_{0i} \exp\left[-\frac{z_i e \phi(\mathbf{r})}{kT}\right]$$

where i is an ionic species, z_i the charge of ion species i, n_{0i} a reference number density for species i, e is the unit charge, and $\rho(\mathbf{r})$ and $\phi(\mathbf{r})$ are the local equilibrium charge density and the average electrostatic potential, respectively. The electrostatic potential $\phi(\mathbf{r})$ is the solution of Poisson's equation:

$$\varepsilon_0 \nabla.[\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = -\rho_{eq}(\mathbf{r}) - \rho_{ex}(\mathbf{r})$$

Assuming a 1 : 1 electrolyte, these two equations can be combined to give:

$$\varepsilon_0 \nabla.[\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = 2en_0 \times \sinh\left[\frac{e\phi(\mathbf{r})}{kT}\right] - \rho_{ex}(\mathbf{r})$$

Here the mobile charges represented by ρ_{eq} are in equilibrium, and the fixed charges on the protein or membrane are represented by ρ_{ex}. This is the full nonlinear Poisson–Boltzmann equation, which can be solved numerically. It can be linearized by expanding sinh and retaining only the leading term. When there are no fixed charges and assuming a spatially homogeneous dielectric, this simplifies to:

$$\nabla^2\phi = \kappa^2\phi$$

with κ^{-1} is Debye screening length, given by:

$$\kappa^{-1} = \sqrt{\frac{\varepsilon_0 \varepsilon kT}{2e^2 n_0}}$$

The Debye screening length $\kappa^{-1} = 8$ Å for 150 mM salt at 298 K in water with a dielectric constant of 80. The screening of a central charge in bulk solution is about 80% at $r = 3\kappa^{-1}$, or ca. 25 Å. This distance decreases to about 9 Å for 1 M salt solution.

The Poisson–Boltzmann equation, either in its linear or nonlinear form, can be solved for macromolecules by a number of commonly used programs, including DelPhi, UHBD, and APBS.

3.3 Ligand Docking

The binding of small molecules to proteins is a key determinant of many biological processes. The ligand-docking problem is the challenge to predict where, in what orientation, and with what affinity a small molecule binds to a binding site in a protein. In the ultimate state of the art, one would like to predict all three with high accuracy, for a high computational cost if necessary. In addition, we also need methods that allow rapid (computationally inexpensive) screening of a library of millions of compounds to search for drug leads that may interact favorably with potential drug targets. Broadly speaking, ligand docking can be approached from two angles. In the first approach, compounds of known structure are docked into the active site of a protein of interest, and some form of scoring function is a measure of how likely the compound will bind tightly to the protein. In the second approach, starting from the known structure of the protein, a molecule can be designed by fitting functional groups in the active site to optimize the interactions with the protein target. The resulting "computational" molecule can then be synthesized and tested. Several classes of scoring functions exist. One class is based on a combination of interaction energies, with terms that are similar to the potential function used in molecular dynamics simulations. Simpler functions based on, for example, the number of hydrogen bonds or penalties for steric clashes are also useful and can be computationally faster. A second class tries to directly calculate the free energy of binding using a parameterized function or an approximate description of the free energy of binding derived from a potential function. A third class uses statistical information about similar molecules and general features of the binding site to rank ligand–protein complexes. At the moment a combination of methods seems to give the best results, but to our knowledge there is no consensus on which method is the best in any given case.

3.4 Protein–Protein Interactions

Protein–protein interactions are emerging as one of the most important themes in biochemistry. Some interactions are sufficiently stable that a complete complex can be crystallized, but this is relatively rare. More often, proteins interact transiently with other proteins at some point in their lives, often in critical processes such as signaling. The protein database contains mainly single proteins, because to crystallize them they are normally purified to a high degree, thus disrupting all but the strongest noncovalent interactions.

The general goal of protein–protein docking could be described as predicting correctly the structure of a protein–protein complex, given only the structures of the two independent proteins. This assumes that these structures are already known. Whether the proteins are allowed to change their structure somewhat upon binding, which seems realistic, is an additional complexity. There are many approaches in the literature, with two common themes. First, there is a remarkable degree of surface complementarity in most stable complexes. This suggests that it should be possible to draw a molecular surface on both proteins (e.g. see Fig. 1d) and search for all possible relative orientations of the two proteins for the most favorable orientation. This type of search is expensive, and the challenge is to identify the most favorable orientation among many possible orientations. Usually, some form of energy function is used, which gives more favorable values for orientations that optimize interactions such as hydrogen bonding, burying of hydrophobic exposed area, and matching up complementary electrostatic potentials on the surfaces of the proteins that form the complex. A second approach combines experimental data with energy functions to guide the two proteins to a solution that is compatible with experimental data. A recent successful example of this approach is HADDOCK, which combines a molecular mechanics energy function, a simplified form of dynamics to enable flexibility of parts of the interface, and incorporation of "ambiguous interaction restraints." These are "soft" restraints that, by themselves, are insufficient to guide docking but combine to yield more reliable solutions than energy functions alone, provided there is experimental information that can be used as restraints. As mentioned above, progress in the area of protein–protein docking is monitored in the CAPRI challenge (http://capri.ebi.ac.uk).

4 Simulations of Protein Dynamics

Three common simulation methods used to study protein dynamics are explained here.

4.1 Molecular Dynamics Simulation

The most common atomistic simulation technique is molecular dynamics (MD). In MD simulations, the interactions between all atoms in the system are described by empirical potentials. An example of a frequently used potential function is:

$$V(\mathbf{r}^N) = \sum_{\text{bonds}} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{\text{angles}} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{\text{torsions}} \frac{V_n}{2}(1 + \cos(n\omega - \gamma))$$

$$+\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]+\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right)$$

This potential function contains harmonic terms for bonds and angles, a cosine expansion for torsion angles, and Lennard–Jones and Coulomb interactions for nonbonded interactions. The constants k_i are harmonic force constants, l_i, is the current bond length, $l_{i,0}$ the reference bond length, θ_i the current angle, $\theta_{i,0}$ the reference angle, V_n, n and γ are the barrier height, multiplicity, and off-set from the origin for the cosine function used to describe dihedral angles (rotations around a central bond), ε and σ are Lennard–Jones parameters (a different pair for each possible combination of two different atom types), q_i and q_j are (partial) atomic charges, and r_{ij} is the distance between two atoms, and ε_0 (different than ε in the Lennard–Jones potential) is the dielectric constant of the medium. Using this potential function, the forces (the derivative of the potential with respect to position) on all atoms in the system of interest are calculated and used to solve classical equations of motions to generate a trajectory of all atoms in time. An example of a system studied with MD is shown in Fig. 3: a simulation snapshot of the ABC transporter BtuCD in a realistic environment consisting of lipids, water, and ions.

The primary result of the simulation is a trajectory of all atoms in time, from which specific details of the system can be analyzed. This is an exciting idea, because atoms can be followed as they move in real time on a timescale of up to ca. 100 ns, although longer simulations have also been reported. In principle, any properties that depend on coordinates, velocities, or forces can be calculated, given sufficient simulation time. No assumptions are required about the nature of the solvent, there is no need to choose dielectric boundaries because all atoms are explicitly present, and in principle all interactions (water–ions,

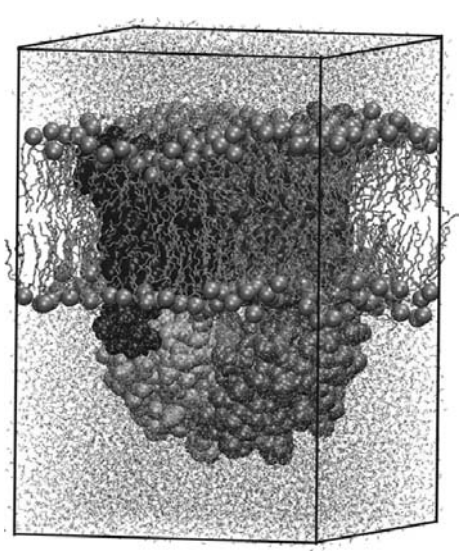

Fig. 3 An orthographic view of the periodic box for the ATP-bound BtuCD simulation, showing water (red and white), lipid with phosphorus atoms enlarged, and the protein. The transporter consists of two transmembrane domains (blue and purple) and two nucleotide binding domains (orange and ochre). The two docked MgATP molecules are partially visible (green and red). Figure courtesy of E. Oloo. See Oloo, E.O., Tieleman, D.P. (2004) Conformational transitions induced by the binding of MgATP to the vitamin B_{12} ABC-transporter BtuCD, *J. Biol. Chem.* **279**, 45013–45019 for more details. Rendered with VMD. (See color plate p. xxx.)

water–protein, water–lipid, lipid–protein etc.) are incorporated.

The main limitations of MD are its computational cost, the limited time and length scale that can be treated in a simulation, and technical limitations such as the accuracy of current empirical force fields.

4.2 Free-energy Calculations

Several useful extensions of the basic simulation method make it possible to calculate free-energy differences from simulations. If a reaction coordinate can be identified, processes that are orders of magnitude slower can be studied than would be possible by direct simulation. Such a reaction coordinate could be a concerted conformational change, or a pathway for ion permeation in ion channels. In umbrella sampling, a biasing potential is used to restrict a simulated system to sample phase space within a specified region (called a *window*). By placing windows along the reaction coordinate, one can generate a free-energy profile (also called *potential of mean force*), which will quantitatively describe why one region of space is more favorable than another.

Relative free energies of different side-chain mutations in a protein can be studied using free-energy perturbation by slowly changing one side chain into another within the computer (also known as computational alchemy). This makes it possible to investigate in the computer the effect of mutations on the stability of a protein or on, for example, the binding affinity of a ligand. The theory behind this kind of calculations is well developed and the algorithms have been implemented in many molecular dynamics software packages.

4.3 Brownian Dynamics Calculations

In Brownian dynamics (BD) simulations, the trajectories of individual particles (ions, molecules) are calculated using the Langevin equation:

$$m_i \frac{d\mathbf{v}_i}{dt} = -\gamma_i \mathbf{v}_i + \mathbf{F}_R + \mathbf{F}_i$$

where m_i, *and* $\mathbf{v}_i$ are the mass and velocity of atom i. Water molecules are not included explicitly, but are present implicitly in the form of a friction coefficient $\gamma = kT/D_i m_i$ (where k is the Boltzmann constant, T is temperature, and D is the diffusion coefficient) and a stochastic force $\mathbf{F}_R$ arising from random collisions of water molecules with ions, obeying the fluctuation-dissipation theorem. $\mathbf{F}_i$ is the force due to other particles in the system as well as external sources, such as an applied electric field. When the friction is large and the motions are overdamped, the inertial term $m_i d\mathbf{v}/dt$ may be neglected, and the simplified form

$$\mathbf{v}_i = \frac{D_i}{kT}\mathbf{F}_i + \mathbf{F}_R$$

may be used. This is the approximation made in Brownian dynamics.

BD simulations require only a few input parameters: in its simplest form, only the diffusion coefficients of the different species of ions and the charge on the ions are needed. However, the model can be refined. To study a protein such as an ion channel, the residues of the protein can be modeled as a set of partial charges, and some form of interaction potential between the mobile ions and the protein must be specified (as seen earlier).

The result of BD simulations is a large set of trajectories for diffusion particles such as ions, proteins, or ligands. By

averaging over these trajectories macroscopic properties such as the conductance of ion channels or association rate of protein–protein or protein–ligand encounters can be calculated. In addition, the simulations yield molecular details of the paths through space for the diffusing particles.

Although Brownian dynamics simulations are conceptually simple, replacing solvent by a continuum description and ignoring internal conformational changes in proteins are significant assumptions that require careful consideration. When BD simulations are valid, they are a very powerful method to study biological processes on a timescale that is much longer than can be reached by MD.

5 Example Applications

The literature is now very extensive, and hundreds of papers are published that are based on each of the methods described above. For illustration purposes, we give a few examples of biochemical problems that have been addressed by modeling and simulation methods.

5.1 Acetylcholinesterase

Acetylcholinesterase (AChE) is an extremely fast enzyme that hydrolyzes the neurotransmitter acetylcholine to terminate signaling in cholinergic synapses. It has been studied in great detail by a number of computational methods, including molecular dynamics simulations, continuum electrostatics calculations and Brownian dynamics simulations.

One of the first puzzles posed by the crystal structure of AChE is the question of how the substrate acetylcholine gains access to the active site, since in the crystal structure there is no unobstructed pathway to the active site. This original observation for AChE from the fish *Torpedo californica* has subsequently been reiterated in mouse and human AChE. Simulations have shown how breathing motions in the enzyme facilitate the displacement of substrate from the surface of the enzyme to the buried active site. These motions appear quite complex and spatially extensive, which suggests possible modes of regulation of the activity of the enzyme. Such a mechanism has been observed in other proteins, including hemoglobin and myoglobin, but the fast reaction rates of AChE are hard to reconcile with major structural rearrangements. MD simulations suggest that the primary point of access opens and closes on a timescale that is fast enough not to slow down the entrance of the substrate substantially. Interestingly, the protein appears to have several secondary channels that allow water to enter and leave the active site as the substrate enters.

The reaction rate of AChE is limited by the diffusion of the neurotransmitter acetylcholine. Brownian dynamics are a suitable method to study encounters of acetylcholine with AChE. The acetylcholine diffuses under the influence of intermolecular interactions between acetylcholine and the protein and a random force. The net effect is that the protein "guides" the acetylcholine to its active site through an electrostatic mechanism that enhances the rate of finding acetylcholine beyond the pure diffusional limit. A similar mechanism has since been identified for several fast-reacting proteins.

AChE is the target of the neurotoxin fasciculin, a peptide. Brownian dynamics of the peptide and AChE give insight into

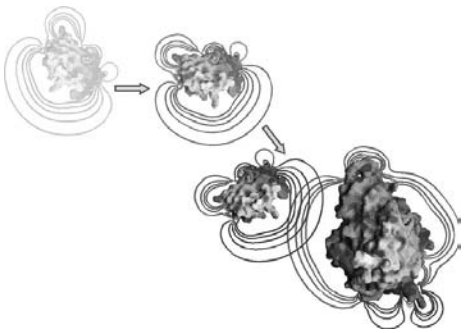

Fig. 4 Schematic illustration of the association of fasciculin (on the left) with acetylcholinesterase. Blue and red contour lines indicate regions of positive and negative electrostatic potential, respectively. Figure courtesy of D. Sept and A. Elcock. See Elcock, A.H., Gabdoulline, R.R., Wade, R.C., McCammon, J.A. (1999) Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin, *J. Mol. Biol.* **291**, 149–162 for more details. (See color plate p. xxiv.)

the binding kinetics of fasciculin and the structure of the resulting complex. In a series of Brownian dynamics simulations, Elcock et al. investigated the effect of mutations in the protein on association rate constants. The electrostatic interaction between AChE and fasciculin that promotes association is illustrated in Fig. 4. In its simplest form, the Brownian dynamics simulations reproduced the correct order of rate constants for different mutants, although the absolute values were too large by about a factor of 30. In a more accurate treatment of the continuum electrostatics problem, this discrepancy can be reduced greatly. Interestingly, these calculations make it possible to distinguish kinetic and thermodynamic effects, and suggest that for the case of fasciculin-AChE-binding mutations can separately affect the rate of association and the binding constant.

5.2 Water Transport Through Aquaporins

Aquaporins are ubiquitous membrane proteins that aid in the maintenance of crucial osmotic balance in cells, on whose discovery the Nobel Prize in chemistry was awarded in 2003. The road to high-resolution structures of these proteins, and to understanding with atomic detail how they function, is a nice example of the link between experiment, modeling, and simulations.

In early 2000, De Groot and coworkers used 4.5-Å resolution data, insufficient to identify the fold of the protein directly, to predict the fold of aquaporin based on the constraints of helix packing, atomic force microscopy data, and primary sequence. From 1440 possible folds, they were able to identify a maximum of 8 possible folds. Later that year, a 3.8-Å resolution structure of aquaporin from cryoelectron microscopy became available, as well as a 2.2-Å X-ray crystal structure for a homolog of aquaporin, GlpF, the glycerol transport facilitator. The fold of aquaporin is illustrated in Fig. 5(a). It has six transmembrane helices in an "hourglass" arrangement. The protein pore is shaped by two reentrant loops on either side of the membrane that have a conserved coil–NPA–helix motif, in which the Asn-Pro-Ala (NPA) residues of these loops meet in the middle of the channel. The partial helices of this motif are labeled HB and HE. The aromatic rich constriction region (ar/R), labeled in Fig. 5(a), is suspected of contributing to aquaporin's highly specific behavior.

Direct simulation of the AQP structure obtained from cryoelectron microscopy

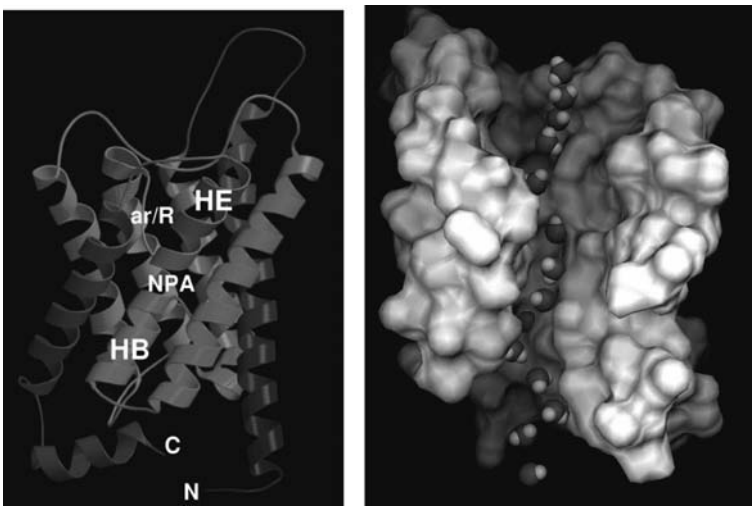

(a) (b)

Fig. 5 (a) Ribbon diagram of the aquaporin monomer and (b) spacefilling representation illustrating a water file through the channel pore. (a) Starting at the N-terminus of the monomer, there are two transmembrane helices, followed by the coil–NPA–helix motif. A third transmembrane helix completes the first half of the protein. The second half of the monomer has two transmembrane helices, followed by the coil–NPA–helix motif, followed by another transmembrane helix. (b) The path of the water channel is through the core of the protein, following the path of the coil motifs, which meet at the NPA signature. This is more clearly illustrated in 5(b). The protein is rendered as a molecular surface in white, with the front surface of the protein cut away so that we can clearly see the pore in the middle of the protein. Surfaces colored yellow are those that interact most strongly with passing water molecules. The waterfile displayed is an overlay of a number of snapshots from the 10 ns simulation. The dipole inversion of water at the NPA motif is clearly illustrated here. Figure courtesy of B. de Groot. See De Groot, B.L., Grubmuller, H. (2001) Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF, *Science* **294**, 2353–2357. De Groot, B.L., Frigato, T., Helms, V., Grubmuller, H. (2003) The mechanism of proton exclusion in the aquaporin-1 water channel, *J. Mol. Biol.* **333**, 279–293 for more details. (See color plate p. xxxi.)

showed a discontinuous water file through the channel and obvious defects in the structure, particularly at the NPA motifs, because of instability observed through the MD simulations. Further homology modeling between the available structures yielded improved stability in the core of the channel. Simulations of the high-resolution GlpF structure, and the improved homology model of AQP1 both observed continuous water files diffusing through the water channel, illustrated in Fig. 5(b), at rates comparable to those observed experimentally.

In late 2001, a 2.2-Å resolution X-ray crystal structure was obtained of human aquaporin-1. With confidence in the three-dimensional structure, the

simulation studies that followed were able to focus on measuring physical properties of the channel using classical simulation methods, and also tried to pinpoint the origin of aquaporin's ability to transport water at high diffusion rates, while restricting proton transport. The basis of proton transport is a topic better addressed by quantum mechanical simulations, and so combinations of classical dynamics methods with semiempirical methods have been employed by a number of groups to determine the basis for proton exclusion by these channels.

What is particularly exciting about aquaporin as a case study is that the biological function of this protein, to act as a water channel across the cell membrane, occurs at a rate which is accessible to study by the current state of the art in molecular dynamics simulations. Through simulations, researchers were able to point out deficiencies in the low-resolution crystal structures on the basis of structural instability, and because the observed behavior *in silico* did not agree with that measured experimentally. This observation helped to point out the significance of certain structural motifs in the protein (i.e. the reentrant loops with the NPA motif). The evolution of our understanding of the aquaporin structure through experiments, homology modeling, and simulations has lead to similar methods being applied to other aquaporins with yet unsolved structures.

5.3 ABC Transporters and Multidrug Resistance Proteins

Many human proteins of major medical interest are difficult to obtain in large enough quantities for structural studies. The only exceptions have been proteins that occur in large amounts, such as bovine lens protein rhodopsin or Aqp1 from blood (as seen earlier). An emerging theme in membrane protein structural biology over the past years has been the use of bacterial homologs of human proteins, with, as striking examples, several potassium channel structures and several structures of ABC transporters. ATP-binding cassette (ABC) transporters are modular mechanical machines that couple the hydrolysis of ATP with the transport of molecules across membranes. They consist of two nucleotide binding domains (NBDs), two transmembrane domains, and optional additional domains, organized in a varying number of polypeptide chains. Mutations in the genes encoding many of the 48 ABC transporters of human cells are associated with several diseases, including cystic fibrosis. Increased expression of certain ABC transporters is a major cause of resistance to peptide antibiotics, antifungals, herbicides, anticancer drugs, and other cytotoxic agents. Interestingly, ABC transporters probably form the largest group of homologous proteins and exist in all species. In an exciting recent development, three crystal structures of ABC transporters have been determined: MsbA from *Escherichia coli* at 4.5-Å resolution, MsbA from *Vibrio cholera* at 3.8-Å resolution, and BtuCD from *E. coli* at 3.2-Å resolution. In addition, about 10 structures of nucleotide binding domains (NBDs) have been solved as well as ATP binding domains from other proteins. Despite this impressive progress in structural studies, several of the key questions about the basic mechanism of action of ABC transporters are unresolved. Simulations and homology modeling can be used to make some progress toward understanding the dynamics of ABC transporters and to translating the bacterial structures to human homologs like P-glycoprotein.

Molecular dynamics simulation studies have been used to investigate the "real-time" dynamics of the vitamin B_{12} importer BtuCD from *E. coli*, based on the single snapshot captured in the crystal structure. An example of a simulation model that incorporates BtuCD is shown in Fig. 3. In this model, the crystal structure is incorporated in a realistic environment of phospholipids, water, and ions, and simulated for 20 ns, both in the absence and presence of MgATP in the two binding sites. The results demonstrate that the docking of ATP to the catalytic pockets progressively draws the two cytoplasmic nucleotide-binding cassettes toward each other. Movement of the cassettes into closer opposition in turn induces conformational rearrangement of α-helices in the transmembrane domain. The shape of the translocation pathway consequently changes in a manner that could aid the vectorial movement of vitamin B_{12}. These results suggest that ATP binding may indeed represent the power stroke in the catalytic mechanism. Moreover, occlusion of ATP at one catalytic site is mechanically coupled to opening of the nucleotide-binding pocket at the second site. This may indicate that asymmetric behavior at the two catalytic pockets forms the structural basis by which the transporter is able to alternate ATP hydrolysis from one site to the other. While this remains to be tested, this study is an example of the use of dynamics simulations that build on a single crystal structure to obtain more information about the full process the protein is involved in.

The multidrug resistance P-glycoprotein mediates the extrusion of chemotherapeutic drugs from cancer cells. Characterization of the drug binding and ATPase activities of the protein have made it the paradigm ATP binding cassette (ABC) transporter. Although no high-resolution structure is known, P-glycoprotein has been imaged at low resolution by electron cryomicroscopy and extensively analyzed by disulfide cross-linking. Stenham et al. used an interesting combination of approaches to create a molecular model of P-glycoprotein (shown in Fig. 6) that fits with most experimental data from cross-linking and imaging, helps interpret these experiments, and provides insight into possible mechanisms of drug transport. As described earlier, the only ABC transporter structures whose high-resolution structures are known are MsbA and BtuCD. There is no homology between the transmembrane domains of BtuCD and P-glycoprotein, but the homology between MsbA and P-glycoprotein is 23% sequence identity for the transmembrane domains and 51% sequence identity for the nucleotide binding domains. MsbA is the closest known bacterial relative of P-glycoprotein. Unfortunately, the nucleotide binding domains in MsbA are poorly resolved and the structure has a low resolution. In contrast, the nucleotide binding domains in BtuCD are well resolved. The situation is further complicated by the tertiary and quaternary organization of the domains of MsbA, which is not consistent with either the BtuCD crystal or the extensive cross-linking data already published for P-gp. Nevertheless, Stenham et al. have shown that it is possible to generate atomic scale models of P-gp by combining experimental and theoretical methods. Their homology models are based on an MsbA template but with a tertiary organization that reflects the increasingly accepted consensus NBD dimer interface. This model will be useful in designing and interpreting experimental work on P-gp. In turn, results

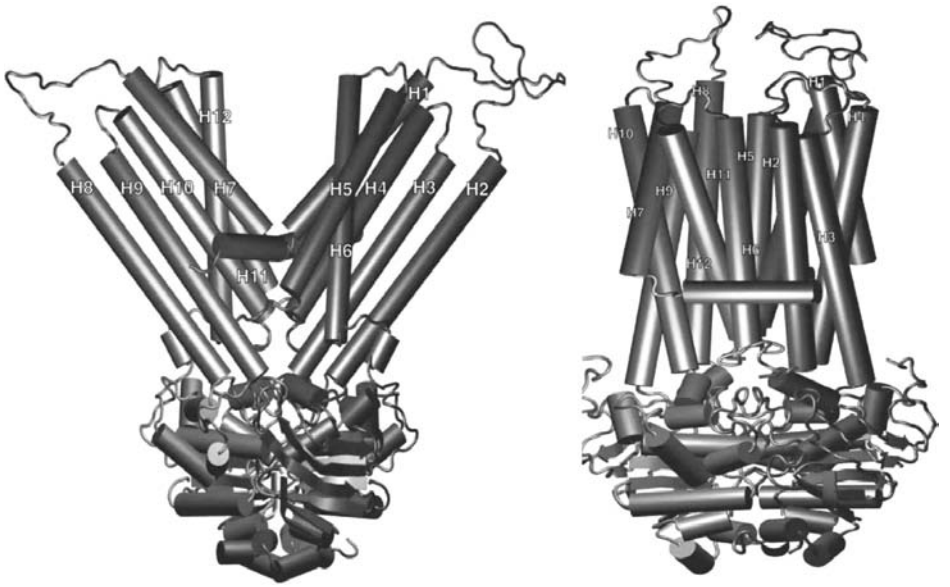

Fig. 6 (a) Construction of a molecular model for P-glycoprotein (P-gp). Each half of P-gp was modeled by homology to the crystal structure of MsbA (PDB code 1JSQ), which had been extended to a full atom representation. The two halves of P-gp were assembled such that the NBDs adopt the ATP-dependent orientation observed in MJ0796 (PDB code 1L2T). The constituent domains are individually colored. (b) Reconciliation of cross-linking data with a P-gp model. P-gp-Model-B was generated by rotation of each NBD with respect to its cognate TMD. The final model contains a parallel TMD:TMD interface (blue and gold subunits) and a consensus NBD:NBD interface (green and purple subunits). Adapted from Stenham, D.R., Campbell, J.D., Sansom, M.S.P., Higgins, C.F., Kerr, I.D., Linton, K.J. (2003) An atomic detail model for the human ATP binding cassette transporter P-glycoprotein derived from disulphide cross-linking and homology modeling, *FASEB J.* **17**, 2287–2289. (See color plate p. xxxii.)

from experimental work can be used directly to improve the model, driving an iterative process that may advance our understanding of medically important ABC transporters.

6 Perspectives

Molecular modeling methods have become standard techniques to address biochemical problems at the level of biomolecular structure. A number of factors have contributed to this success. Computers themselves have become vastly more powerful in the past decades. Initially, simulations were restricted to very simple systems on timescales and levels of detail that were typically not adequate to answer specific questions about proteins. With a 10-fold increase in speed every 5 years a wide range of problems is now accessible directly by computer simulation, such as water transport in aquaporins. The fastest folding proteins can now be studied directly

by simulations on the same timescale as the fastest experiments, with unprecedented progress in our understanding of protein folding. Computer graphics hardware has enjoyed a similar increase in speed, enabling complex surface representations of large biomolecular complexes and interactive 3D graphics. A second factor is the improvement in simulation and modeling software, which has led to the development of software that is vastly more user friendly than even a few years ago and runs on commodity hardware instead of exotic (expensive) workstations. A third factor is the improvement in algorithms and parameters that describe interactions between atoms or predict the structure of proteins. With increasing accuracy or more efficient algorithms, molecular modeling is able to answer increasingly more detailed and complicated questions. Finally, the massive advances in experimental studies at the level of biomolecular structure feed back into computer modeling. The expanding database of high-resolution protein structures is a key resource for further development of modeling methods. Advanced single-molecule experiments provide a direct link with simulations that was very rare in the past, and an increasing number of experimental techniques can now be linked to simulation results directly. An example of this is data from nuclear magnetic resonance spectroscopy, which provides both structural and dynamical data on timescales that are comparable to those of simulations.

Because molecular modeling is such a broad area, it is difficult to indicate exactly where most progress can be expected. Technical advances in hardware and software engineering, algorithm and parameter development will continue to increase the number of problems that can be addressed by molecular modeling. These advances will also facilitate incorporation of computationally intense calculations such as QM and QM/MM methods into accessible programs that are user-friendlier to the molecular modeler. The use of Grid technology, distributed computing, and massively parallel commodity-based Beowulf clusters will bring CPU-intensive calculations within reach of an increasing number of researchers, allowing us to study bigger and more complex systems. We expect that important areas of growth will be a stronger link between modeling and experiments, the routine incorporation of modeling and simulation in otherwise mostly experimental studies, and improved methods to study protein–protein interactions and chemical reactions.

Acknowledgment

Tieleman is a Scholar of the Alberta Heritage Foundation for Medical Research and a Sloan Foundation Fellow. Work in his laboratory is supported by the National Sciences and Engineering Research Council and the Canadian Institutes of Health Research.

See also Mass Spectrometry of Proteins (Proteomics); Protein NMR Spectroscopy; Protein Structure Analysis: High-throughput Approaches; Proteomics; RNA Three-Dimensional Structures, Computer Modeling of.

Bibliography

Books and Reviews

Bourne P.E., Weissig, H. (Eds.) (2003) *Structural Bioinformatics*, John Wiley & Sons Inc., New Jersey, MI.

Day, R., Daggett, V. (2003) All-atom Simulations of Protein Folding and Unfolding, in: Eisenberg, D., Kim, P. (Eds.) *Advances in Protein Chemistry*, Vol. 66, Elsevier Academic Press, New York, pp. 373–403.

Frenkel, D. Smit, B. (2001) *Understanding Molecular Simulation. From Algorithms to Applications*, 2nd edition, Academic Press.

Honig, B., Nicholls, A. (1995) Classical electrostatics in biology and chemistry, *Science* **268**, 1144–1149.

Jorgensen, W.L. (2004) The many roles of computation in drug discovery, *Science* **303**, 1813–1818.

Karplus, M., McCammon, J.A. (2002) Molecular dynamics simulations of biomolecules, *Nat. Struct. Biol.* **9**(9), 646–652.

Leach, A.R. (2001) *Molecular Modeling, Principles and Applications*, 2nd edition, Pearson Education Limited, England.

Petsko, G.A., Ringe, D. (2004) *Primers in Biology: Protein Structure and Function*, New Science Press Ltd.

Schlick, T. (2002) *Molecular Modeling and Simulation – An Interdisciplinary Guide*, Springer, New York.

Primary Literature

Agre, P. (2004). Aquaporin Water Channels (Nobel Lecture). *Angew. Chem. Int. Ed.* **43**, 4278–4290.

Altschul, S.F., Madden, T.L., Schaeffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acid Res.* **25**, 3389–3402.

Ash, W.L., Zlomislic, M.R., Oloo, E.O., Tieleman, D.P. (2004) Computer simulations of membrane proteins, *Biochim. Biophys. Acta-Biomembranes* **1666**, 158–189.

Baker, N.A., McCammon, J.A. (2003) Electrostatic Interactions, in: Bourner, P.E., Weissig, H. (Eds.) *Structural Bioinformatics*, Wiley & Sons, Inc., New Jersey, MI.

Baker, D., Sali, A. (2001) Protein structure prediction and structural genomics, *Science* **294**, 93–96.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037–10041.

Berendsen, H.J.C. (2001) Bioinformatics – Reality simulation – Observe while it happens, *Science* **294**, 2304–2305.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. (2000) The protein data bank, *Nucleic Acids Res.* **28**, 235–242.

Bradley, P., Chivian, D., Meiler, J., Misura, K.M.S., Rohl, C.A., Schief, W.R., Wedemeyer, W.J., Schueler-Furman, O., Murphy, P., Schonbrun, J., Strauss, C.E.M., Baker, D. (2003) Rosetta predictions in CASP5: Successes, failures, and prospects for complete automation, *Proteins: Struct., Funct., Genet.* **53**, 457–468.

Burykin, A., Warshel, A. (2003) What really prevents proton transport through aquaporin? Charge self-energy versus proton wire proposals, *Biophys. J.* **85**, 3696–3706.

Campbell, J.D., Biggin, P.C., Baaden, M., Sansom, M.S.P. (2003) Extending the structure of an ABC transporter to atomic resolution: modeling and simulation studies of MsbA, *Biochemistry* **42**, 3666–3673.

Chakrabarti, N., Tajkhorshid, E., Roux, B., Pomes, R. (2004) Molecular basis of proton blockage in aquaporins, *Structure* **12**, 65–74.

Chance, M.R., Fiser, A., Sali, A., Pieper, U., Eswar, N., Xu, G.P., Fajardo, J.E., Radhakannan, T., Marinkovic, N. (2004) High-throughput computational and experimental techniques in structural genomics, *Genome Res.* **14**, 2145–2154.

Chang, G. (2003) Structure of MsbA from Vibrio cholera: a multidrug resistance ABC transporter homolog in a closed conformation, *J. Mol. Biol.* **330**, 419–430.

Davis, M.E., McCammon, J.A. (1990) Electrostatics in Biomolecular Structure and Dynamics, *Chem. Rev.* **90**, 509–521.

De Groot, B.L., Grubmuller, H. (2001) Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF, *Science* **294**, 2353–2357.

De Groot, B.L., Engel, A., Grubmuller, H. (2003) The structure of the aquaporin-1 water channel: a comparison between cryo-electron microscopy and x-ray crystallography, *J. Mol. Biol.* **325**, 485–493.

De Groot, B.L., Frigato, T., Helms, V., Grubmuller, H. (2003) The mechanism of proton exclusion in the aquaporin-1 water channel, *J. Mol. Biol.* **333**, 279–293.

De Groot, B.L., Heymann, J.B., Engel, A., Mitsuoka, K., Fujiyoshi, Y., Grubmuller, H. (2000) The fold of human aquaporin-1, *J. Mol. Biol.* **300**, 987–994.

Dill, K.A., Chan, H.S. (1997) From Levinthal to pathways to funnels, *Nat. Struct. Biol.* **4**, 10–19.

Dominguez, C., Boelens, R., Bonvin, A.M. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information, *J. Am. Chem. Soc.* **125**, 1731–1737.

Elcock, A.H. (2004) Molecular simulations of diffusion and association in multimacromolecular systems, *Numerical Comput. Methods, Pt D* **383**, 166–198.

Elcock, A.H., Sept, D., McCammon, J.A. (2001) Computer simulation of protein-protein interactions, *J. Phys. Chem. B* **105**, 1504–1518.

Elcock, A.H., Gabdoulline, R.R., Wade, R.C., McCammon, J.A. (1999) Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin, *J. Mol. Biol.* **291**, 149–162.

Ermak, D.L., McCammon, J.A. (1978) Brownian dynamics with hydrodynamic interactions, *J. Chem. Phys.* **69**, 1352–1360.

Fiser, A.S., Sali, A. (2003) MODELLER: generation and refinement of homology-based protein structure models, *Macromol. Crystallogr., Pt D* **374**, 461–46.

Gabdoulline, R.R., Wade, R.C. (2001) Protein-protein association: Investigation of factors influencing association rates by brownian dynamics simulations, *J. Mol. Biol.* **306**, 1139–1155.

Godzik, A. (2003) Fold Recognition Methods, in: Bourner, P.E., Weissig, H. (Eds.) *Structural Bioinformatics*, John Wiley & Sons, Inc., New Jersey, MI, pp. 525–546.

Gues, N., Peitsch, M.C. (1997) Swiss-model and the Swiss-PDB viewer: an environment for comparative protein modeling, *Electrophoresis* **18**, 2714–2723.

Hartley, R.W. (1989) Barnase and barstar – 2 small proteins to fold and fit together, *Trends Biochem. Sci.* **14**, 450–454.

Holland, I.B., Cole, S.P.C., Kuchler, K., (Eds.) (2002). *ABC Proteins: From Bacteria to Man.* Academic Press.

Honig, B., Nicholls, A. (1995) Classical electrostatics in biology and chemistry, *Science* **268**, 1144–1149.

Humphrey, W., Dalke, A., Schulten, K. (1996) VMD – visual molecular dynamics, *J. Mol. Graph.* **14**, 33–38.

Ilan, B., Tajkhorshid, E., Schulten, K., Voth, G.A. (2004) The mechanism of proton exclusion in aquaporin channels, *Proteins* **55**, 223–228.

Janin, J., Henrick, K., Moult, J., Eyck, L.T., Sternberg, M.J., Vajda, S., Vakser, I., Wodak, S.J. (2003) CAPRI: a critical assessment of predicted interactions, *Proteins* **52**, 2–9.

Jensen, M.O., Tajkhorshid, E., Schulten, K. (2003) Electrostatic tuning of permeation and selectivity in aquaporin water channels, *Biophys. J.* **85**, 2884–2899.

Kollman, P. (1993) Free-energy calculations – applications to chemical and biochemical phenomena, *Chem. Rev.* **93**, 2395–2417.

Krieger, E., Nabuurs, S.B., Vriend, G. (2003) Homology Modeling, in: Bourner, P.E., Weissig, H. (Eds.) *Structural Bioinformatics*, Wiley & Sons, Inc, New Jersey, MI.

Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L., Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy, *Science* **302**, 1364–1368.

Lee, L.P., Tidor, B. (2001) Barstar is electrostatically optimized for tight binding to barnase, *Nat. Struct. Biol.* **8**, 73–76.

Levinthal, C. (1969). How to Fold Graciously. *Mossbauer Spectroscopy in Biological Systems*, University of Illinois Press, Urbana, Illinois, MN.

Locher, K.P., Lee, A.T., Rees, D.C. (2002) The E-coli BtuCD structure: a framework for ABC transporter architecture and mechanism, *Science* **296**, 1091–1098.

MacKinnon, R. (2004) Potassium channels and the atomic basis of selective ion conductance (Nobel lecture), *Angew. Chem.-Int. Ed.* **43**, 4265–4277.

Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., Sali, A. (2000) Comparative protein structure modeling of genes and genomes, *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291–325.

McCammon, J.A., Gelin, B.R., Karplus, M. (1977) Dynamics of folded proteins, *Nature* **267**, 585–590.

McCammon, J.A., Gelin, B.R., Karplus, M., Wolynes, P.G. (1976) Hinge-bending mode in lysozyme, *Nature* **262**, 325–326.

Oloo, E.O., Tieleman, D.P. (2004) Conformational transitions induced by the binding of MgATP to the vitamin B12 ATP-binding cassette (ABC) transporter BtuCD, *J. Biol. Chem.* **279**, 45013–45019.

Rost, B. (2003) Prediction in 1D: Secondary Structure, Membrane Helices, and Accessibility, in: Bourner, P.E., Weissig, H. (Eds.) *Structural Bioinformatics*, John Wiley & Sons, Inc, New Jersey, MI.

Rychelewski, L., Fischer, D., Elofsson, A. (2003) LiveBench-6: large-scale automated evaluation of protein structure prediction servers, *Proteins: Struct., Funct., Genet.* **53**, 542–547.

Sali, A., Blundell, T.L. (1993) Comparative protein modeling by satisfaction of spatial restraints, *J. Mol. Biol.* **234**, 779–815.

Sayle, R., Milner-White, E.J. (1995) RasMol: biomolecular graphics for all, *Trends Biochem. Sci.* **20**, 374.

Schreiber, G., Fersht, A.R. (1995) Energetics of protein-protein interactions – analysis of the barnase-barstar interface by single mutations and double mutant cycles, *J. Mol. Biol.* **248**, 478–486.

Sharp, K.A., Honig, B. (1990) Electrostatic interactions in macromolecules – theory and applications, *Annu. Rev. Biophys. Biophys. Chem.* **19**, 301–332.

Shen, T.Y., Tai, K.H., Henchman, R.H., McCammon, J.A. (2002) Molecular dynamics of acetylcholinesterase, *Acc. Chem. Res.* **35**, 332–340.

Smith, G.R., Sternberg, M.J. (2002) Prediction of protein–protein interactions by docking method, *Curr. Opin. Struct. Biol.* **12**, 28–35.

Smith, G.R., Sternberg, M.J. (2003) Evaluation of the 3D-dock protein docking suite in rounds 1 and 2 of the CAPRI blind trial, *Proteins* **52**, 74–79.

Stenham, D.R., Campbell, J.D., Sansom, M.S.P., Higgins, C.F., Kerr, I.D., Linton, K.J. (2003) An atomic detail model for the human ATP binding cassette transporter P-glycoprotein derived from disulphide cross-linking and homology modeling, *FASEB J.* **17**, 2287–2289.

Tajkhorshid, E., Nollert, P., Jensen, M.O., Miercke, L.J.W., O'Connell, J., Stroud, R.M., Schulten, K. (2002) Control of the selectivity of the aquaporin water channel family by global orientational tuning, *Science* **296**, 525–530.

Tajkhorshid, E., Aksimentiev, A., Balabin, I., Gao, M., Israelwitz, B., Phillips, J.C., Zhu, F., Schulten, K. (2003) Large Scale Simulation of Protein Mechanics and Function, in: Eisenberg, D., Kim, P. (Eds.) *Advances in Protein Chemistry*, Vol. 66, Elsevier Academic Press, New York, pp. 195–247.

Tate, J. (2003) Molecular Visualization, in: Bourner, P.E., Weissig, H. (Eds.) *Structural Bioinformatics*, Wiley & Sons, Inc, New Jersey, MI.

Taylor, R.D., Jewsbury, P.J., Essex, J.W. (2002) A review of protein-small molecule docking methods, *J. Comput. Aided Mol. Des.* **16**, 151–166.

Tieleman, D.P., Marrink, S.J., Berendsen, H.J.C. (1997) A computer perspective of membranes: molecular dynamics studies of lipid bilayer systems, *Biochim. Biophys. Acta-Rev. Biomembranes* **1331**, 235–270.

Van Gunsteren, W.F., Berendsen, H.J.C. (1990) Computer-simulation of molecular-dynamics - methodology, applications, and perspectives in chemistry, *Angew. Chem.-Int. Ed. Engl.* **29**, 992–1023.

Vaughan, C.K., Buckle, A.M., Fersht, A.R. (1999) Structural response to mutation at a protein-protein interface, *J. Mol. Biol.* **286**, 1487–1506.

Vriend, G. (1990) What if – a molecular modeling and drug design program, *J. Mol. Graph.* **8**, 52–55.

Warshel, A., Aqvist, J. (1991) Microscopic Simulations of Chemical Processes in Proteins and the Role of Electrostatic Free Energy, in: Beveridge, D.L., Lavery, R. (Eds.) *Theoretical Biochemistry and Molecular Biophysics*, Vol. 2, Adenine Press, New York, pp. 257.

Protein NMR Spectroscopy

Thomas Szyperski
State University of New York, Buffalo, NY

1 **Aspects of Multidimensional Protein NMR Spectroscopy 161**

2 **NMR Instrumentation 165**

3 **Experimental Protein NMR Parameters 165**

4 **Recently Developed Techniques 167**

5 **Structure Determination 172**

6 **Dynamics 175**

7 **Hydration 176**

8 **Folding 177**

9 **Structural Genomics 177**

Bibliography 178
Books and Reviews 178
Primary Literature 178

Keywords

Correlation Spectroscopy
"COSY": NMR experiment devised to correlate chemical shifts via through-bond scalar nuclear spin–spin couplings.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

Fourier Transformation
Mathematical transformation that provides an analysis of the angular frequencies encoded in a given function.

Nuclear Magnetic Moment
Since atomic nuclei are positively charged, the nuclear spin is proportional to a nuclear magnetic moment.

Nuclear Magnetic Resonance
Owing to their magnetic moment, atomic nuclei interact with external magnetic fields. This enables one to pursue nuclear magnetic resonance (NMR) spectroscopy.

Nuclear Overhauser Enhancement Spectroscopy
"NOESY": NMR experiment to measure through-space dipolar interactions between the magnetic moments of protons. This allows one to estimate distances between protons in proteins.

Nuclear Spin
The nuclear spin is a nonclassical angular momentum associated with atomic nuclei possessing a spin quantum number larger than zero.

Triple Resonance NMR Experiment
Experiment devised to correlate the chemical shifts of three types of nuclei, that is, ^{1}H, ^{13}C, and ^{15}N.

Nuclear magnetic resonance (NMR) spectroscopy provides unique information about protein structure, dynamics, hydration, and folding in aqueous solution, and has become a pivotal biophysical technique to investigate biological macromolecules. Although the first application of NMR spectroscopy to study proteins date back to the 1960s, its widespread use for proteins was fostered only in the late 1970s and early 1980s by introduction of two-dimensional NMR spectroscopy conducted on spectrometers equipped with superconducting high-field magnets. Advances made in the subsequent 25 years until today have established NMR as an indispensable tool to study even large proteins with molecular masses above 100 kDa. These advances were due to new approaches for efficient production of stable isotope labeled proteins, novel spin relaxation, optimized and rapid sampling of NMR data collection strategies, and the advent of highly sensitive cryogenic probes used for signal detection at field strengths corresponding to 500–900 MHz proton resonance frequency. Recently, the outstanding value of NMR was demonstrated for high-throughput structure determination in the newly emerging field of structural genomics, which makes protein NMR spectroscopy also a key tool for systems biology.

1
Aspects of Multidimensional Protein NMR Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is based on the existence of a nonclassical angular momentum, the *nuclear spin*. All atomic nuclei with a nonzero nuclear spin quantum number possess such a spin. Since nuclei are positively charged, the spin generates a colinearly oriented magnetic moment, which interacts with magnetic fields as compass needles interact with the earth's magnetic field. "Spin-$\frac{1}{2}$ nuclei," with a spin quantum number of $\frac{1}{2}$, exhibit neither electric dipole nor electric quadrupole moments. Hence, spin-$\frac{1}{2}$ nuclei are quite weakly coupled to the environment (often called the *lattice*) via their magnetic dipole moments only: these interact with random fluctuations of magnetic fields that arise from thermal motions. In proteins, the weak coupling enables the observation of coherent spatial reorientations of spins (briefly named *coherences*) about the axis of a strong external magnetic field, B_0, for time periods up to several hundred milliseconds. For a typical study, this allows accurate measurement of hundreds or even thousands of NMR parameters. This wealth of information is unrivaled among solution spectroscopic techniques and is a central reason for the success of modern protein NMR spectroscopy.

Protein NMR relies on the three spin-$\frac{1}{2}$ nuclei ^{1}H, ^{13}C, and ^{15}N. Since these nuclei possess the same spin quantum number of $\frac{1}{2}$, the absolute value of their spin angular momenta is equal. However, owing to varying charges (and charge distributions within the nuclei), the resulting magnetic moments differ greatly. ^{1}H possesses the largest magnetic moment, while the magnetic moments of ^{13}C and ^{15}N are fourfold and tenfold smaller, respectively. Since resonance frequencies scale accordingly, this has far-reaching consequences for the design of protein NMR experiments: highest sensitivity is achieved for experiments starting with proton polarization and ending with proton signal detection. Furthermore, only ^{1}H is 100% abundant in nature, while ^{13}C (natural abundance: ~1%; mainly naturally occurring is ^{12}C, a spin-0 nucleus which is "invisible" in NMR) and ^{15}N (~0.5%; mainly naturally occurring is ^{14}N, a spin-1 nucleus with an electric quadrupole moment that results in broad NMR lines) need to be enriched if a protein has to be studied with more demanding heteronuclear NMR spectroscopy based on ^{13}C and/or ^{15}N.

Radio frequency (rf) pulses created with an rf coil can induce the coherent precession of spin magnetic moments about the axis of an external magnetic field. This precession and its decay can be detected with an rf coil, and the resulting signal ("free induction decay") yields, after Fourier Transformation (FT), a one-dimensional (1D) NMR spectrum. For proteins, the resonance lines provide chemical shifts and, in favorable cases, estimates for scalar spin–spin couplings from resonance fine structures, and nuclear spin relaxation times from line widths. However, owing to the large number of nuclei, 1D NMR spectra of proteins and other biological macromolecules are very crowded. Hence, protein NMR primarily relies on two-, three-, and four-dimensional (2D, 3D, and 4D) spectral information. The high dimensional NMR spectra afford increased spectral resolution and correlate several chemical shifts in a single data set.

A 2D data set is obtained if the signals of a 1D spectrum are modulated by the evolution of an NMR parameter (usually a chemical shift) in an "indirect

dimension'' prior to acquisition (Fig. 1). Fourier transformation along the indirect dimension yields the 2D frequency domain spectrum. In the same fashion, 3D (4D) NMR data sets are obtained by recording many 2D (3D) data sets, which are modulated by an additional NMR parameter. As a result, we have the minimal measurement time, T_m, increasing steeply with dimensionality: acquiring 32 points in each indirect dimension (with one scan per 1D spectrum each second) yields T_m(3D) ~0.5 h, T_m(4D) ~9 h, T_m(5D) ~12 days and T_m(6D) ~1 year. This obstacle for acquiring the highest dimensional NMR spectral information has been named the *NMR sampling problem*.

Figure 2(a) shows a 2D [^{15}N,^{1}H]-correlation spectrum (''COSY''), which was recorded for a ^{15}N-labeled 21-kDa globular protein on a spectrometer operating at 900 MHz ^{1}H resonance

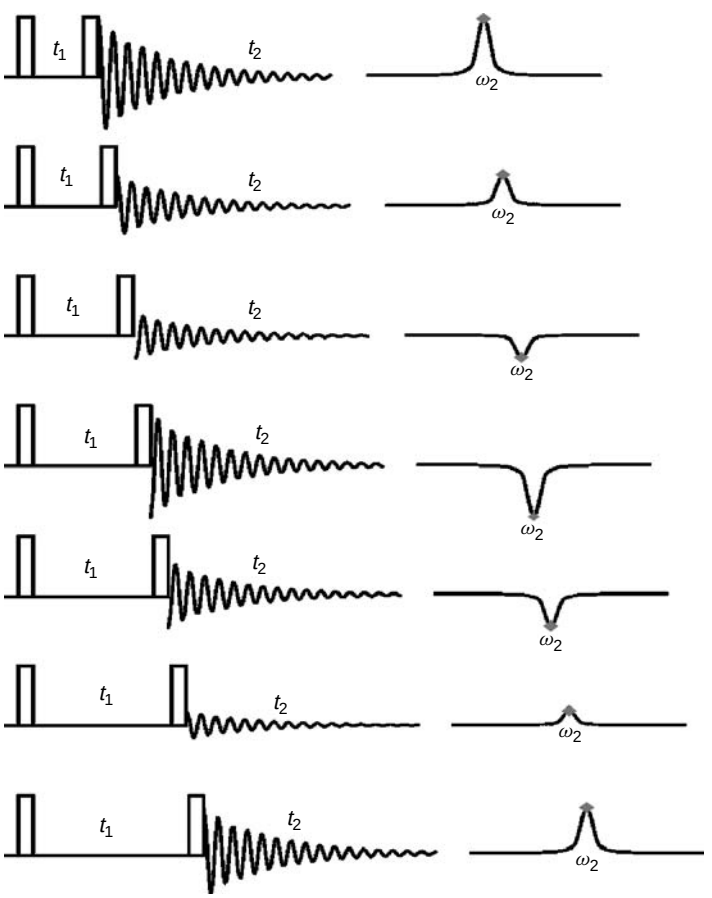

Fig. 1 Principle of 2D NMR. 1D spectra (on the right), obtained after Fourier transformation (FT) of the free induction decays along t_2, are modulated by the evolution of an NMR parameter (e.g. a chemical shift) in an indirect dimension (''t_1''). The frequency domain 2D spectrum is obtained after a second FT along t_1. Courtesy of Jeff Mills.

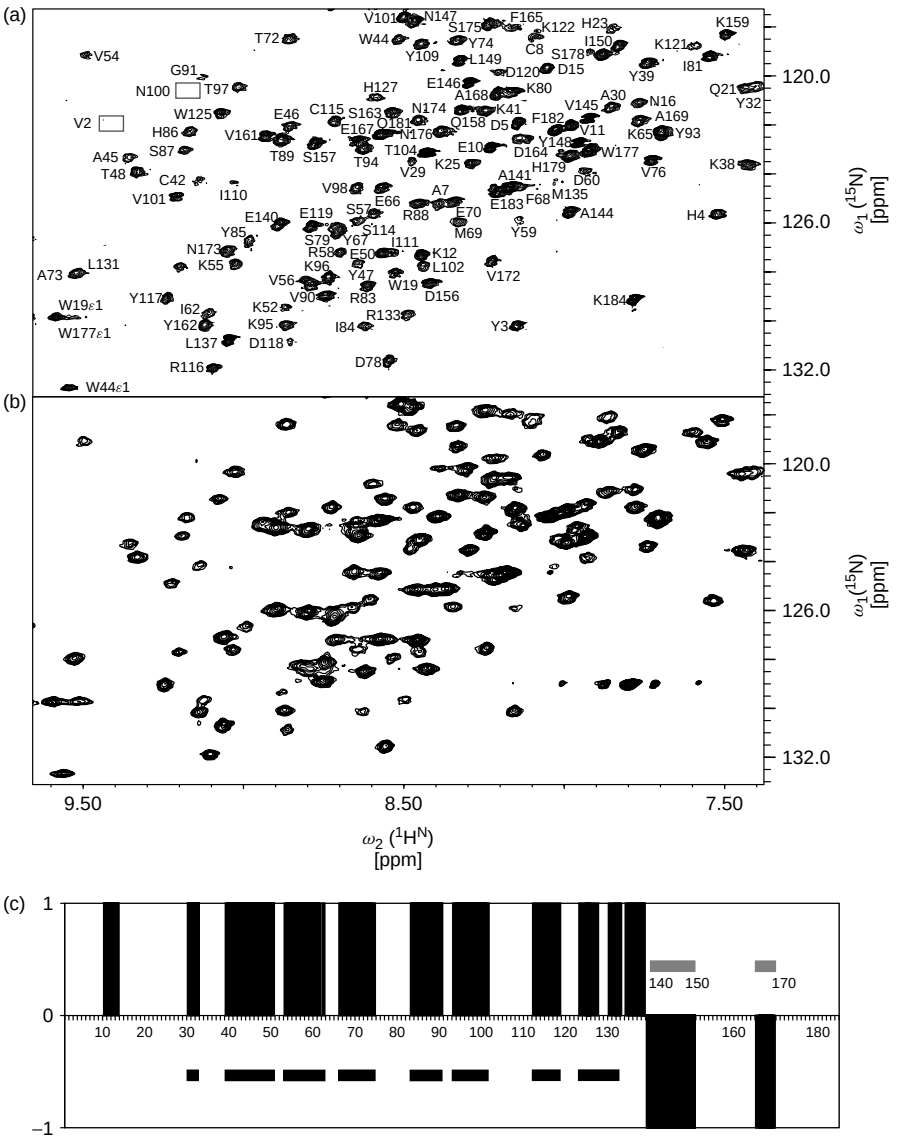

Fig. 2 (a) 2D [^{15}N,^{1}H]-COSY spectrum recorded in the transverse relaxation optimized (''TROSY'') acquisition mode (see Sect. 4) on a 900-MHz Varian INOVA spectrometer (25 °C) for the uniformly ^{15}N-labeled 21-kDa protein FluA(R95K). The peaks are labeled using the one-letter code for amino acids. (b) For comparison, a 2D [^{15}N,^{1}H]-COSY spectrum recorded without ''TROSY'' at 600 MHz is shown. (c) Chemical shift index (CSI) consensus plot for identification of regular secondary structure elements. The eight β-strands forming a β-barrel (black bars) as well as the α-helices (gray bars) are indicated. (a and b are reproduced with permission from Liu, G., Mills, J.L., Hess, T.A., Kim, S., Skalicky, J.J., Sukumaran, D.K., Kupce, E., Skerra, A., Szyperski, T. (2003) Resonance assignments for the 21 kDa engineered fluorescein-binding lipocalin FluA, *J. Biomol. NMR* **27**, 187–188).

frequency. To generate this spectrum, 1D ^{1}H NMR spectra of amide protons are modulated with the chemical shift of the covalently attached ^{15}N nucleus, and one signal is detected for each amino acid residue (except for prolinyl residues and the N-terminal residue). Owing to high intrinsic sensitivity, one can rapidly obtain site-specific information along the entire polypeptide backbone. This feature makes 2D [^{15}N,^{1}H]-COSY (correlation spectroscopy) one of the most important experiments in modern NMR-based structural biology. First, the signal dispersion allows one to assess the degree to which the protein is folded: unfolded polypeptide segments tend to exhibit signals in a comparably narrow range attributed to "random coils." Secondly, the comparison of expected and detected number of signals enables one to investigate if parts of the protein, for example, loops connecting regular secondary structure elements, are affected by slow motional modes in the microsecond to millisecond time range. Such motions quite often lead to broadening of signals beyond detection.

Prior to using NMR spectroscopy for deriving site-specific information on structure or dynamics of proteins, the nuclear magnetic resonances must be "assigned," that is, one has to identify the chemical shifts (see Sect. 3) of spins. Since proteins contain each type of amino acid several times at various positions along the polypeptide chain, it is not sufficient to identify the type of proton (e.g. $^{1}H^{\alpha}$ of glutamate) belonging to a given resonance. Instead, one has to obtain "sequence-specific" resonance assignments (e.g. $^{1}H^{\alpha}$ of the glutamate *in position 79*). Obtaining (nearly complete) sequence-specific resonance assignments is generally considered a prerequisite for determining protein structures (see Sect. 5). This so-called "assignment problem" is comparable to the crystallographic "phase problem" in the sense that a structural interpretation of experimental data depends on solving these problems. Nowadays, sequential resonance assignment of $^{15}N/^{13}C$-labeled proteins relies largely on recording $^{15}N/^{13}C/^{1}H$ triple resonance NMR experiments (Fig. 3). In these experiments, several proton, carbon, and nitrogen chemical shifts are correlated. These originate either from the same, or from two sequentially neighboring residues. Combining the information from the two experiments then allows one to effectively "walk" along the polypeptide chain to obtain sequence-specific resonance assignments for backbone (and often $^{13}C^{\beta}$ shifts). Importantly, these chemical shifts can provide the location of regular secondary structure elements (Fig. 2c) without reference to nuclear Overhauser effects (NOEs). The

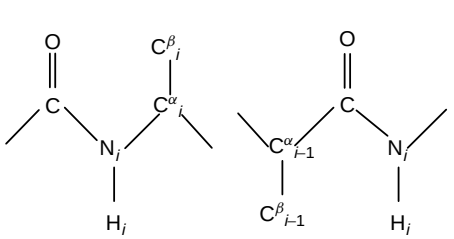

Fig. 3 Example of chemical shifts of nuclei (in color) correlated in triple resonance NMR experiments for obtaining sequence-specific backbone resonance assignments of proteins. A first experiment correlates the shifts of amide proton, nitrogen, α-carbon, and β-carbon within a given residue i (on the left). A second experiment correlates the shifts of amide proton and nitrogen of residue i with those of the α- and β-carbons of the preceding residue $i-1$. Combining the two experiments allows one to obtain sequence-specific resonance assignments for the backbone (and $^{13}C^{\beta}$ shifts) of the entire polypeptide. (See color plate p. xxix).

backbone NMR experiments are then combined with other experiments to assign resonances of side chains, thus yielding a *complete* protein resonance assignment.

2 NMR Instrumentation

Transitions between nuclear spin states are associated with radio frequency energy quanta that are, at ambient temperature, several orders of magnitude smaller than kT (k and T represent Boltzman constant and temperature). Hence, NMR transitions are *lowest energy* transitions. This results in both, the greatest strength and the greatest weakness of NMR spectroscopy. The strength is due to the fact that the low energy required for observing the system (using the nuclear spins as "spies") hardly affects even the subtlest conformational equilibria in their structural and dynamic manifestation. The weakness associated with the low interaction energy is due to the low sensitivity of NMR spectroscopy when compared with, for example, optical spectroscopy. This problem is further aggravated in protein NMR, since slowly tumbling macromolecules exhibit short transverse spin relaxation times and thus exhibit broad resonance lines. Hence, to alleviate this drawback, modern protein NMR depends on using large superconducting magnets (Fig. 4a). In 2005, the largest commercially available magnets have a price tag of about four million US dollars and can induce a ^{1}H resonance frequency of 900 MHz. Apart from increased sensitivity, which scales with $B_0{}^{3/2}$, high magnetic fields yield increased signal dispersion scaling with $B_0{}^{N}$, where N is the dimensionality of the NMR experiment. A second major breakthrough in hardware development was the construction of cryogenic NMR probes (Fig. 4b). In such probes, the rf coil for signal detection is cooled to about 25 K, which leads to a large reduction in thermal noise in the rf circuitry. In spite of the fact that the insulation between rf coil and NMR sample (which needs to be kept at ambient temperature) reduces the filling factor of the coil, 2 to 3 fold gains in sensitivity can be routinely achieved in protein NMR. This corresponds to about 5 to 10 fold reduced NMR data collection times, so that rapid NMR data collection techniques (see Sect. 4) are best suited to take advantage of the high sensitivity of such probes.

3 Experimental Protein NMR Parameters

Protein NMR spectroscopy aims at measurement of parameters, which provide the desired structural and/or dynamic information of a system under consideration. *Chemical shifts* are due to the site-specific shielding of the external magnetic field at the location of the nucleus. Hence, chemical shifts depend on the covalent and conformational environment in which a spin is embedded and are affected by long-range electrostatic and ring current effects. The dispersion of chemical shifts arising from the folding of a polypeptide into a tertiary structure is pivotal for using NMR: without such conformation-dependent dispersion, one could not resolve the site-specific information. Chemical shifts are routinely measured in multidimensional NMR spectra (see Sect. 1), and the resonance assignment is quite generally considered a prerequisite for the site-specific interpretation of other NMR parameters described below. Importantly, backbone and $^{13}C^{\beta}$ chemical shifts provide the location of the regular secondary structure elements

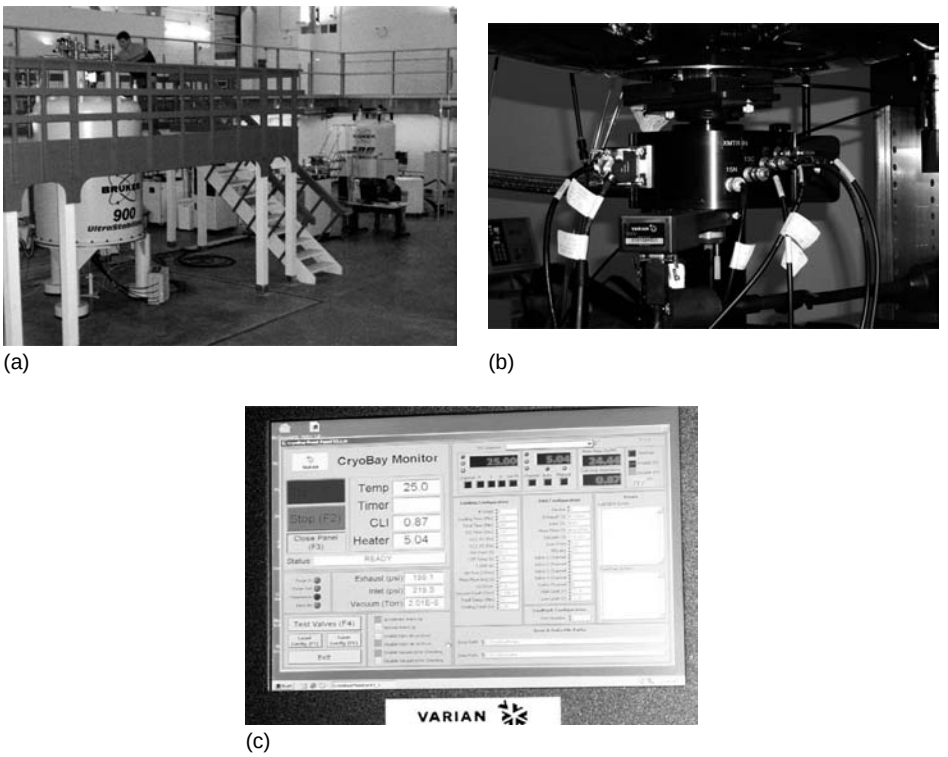

(a) (b) (c)

Fig. 4 Modern high-field NMR equipment used for protein NMR spectroscopy. (a) Spectrometer at the New York Structural Biology Center operating at 900 MHz ^{1}H resonance frequency (Courtesy of David Cowburn). (b) Cryogenic probe installed at the University at Buffalo high-field NMR facility (Courtesy of Dinesh Sukumaran). The probe is inserted from below in the magnet, and its rf coil is cooled to 25 K, as can be read off the monitor shown in (c).

(Fig. 2c). Moreover, chemical shifts are not isotropic, that is, the shielding of the external magnetic field depends on the orientation of the molecule relative to the field. Chemical shifts are thus accurately described by using a "chemical shift tensor."

Scalar nuclear spin–spin couplings, J, arise from the through-bond coupling of the nuclear magnetic moments. The *J*-coupling is the most important interaction used to devise multidimensional experiments for measurement and correlation of chemical shifts (Fig. 2a). Moreover, three-bond *J*-couplings are related through the "Karplus-relations" to dihedral angles and thus encode valuable structural information.

^{1}H$-^1$H NOEs arise from through-space coupling of the magnetic dipole moments of protons. Owing to the r^{-6} dependence of the dipolar coupling (with r being the distance between two protons), the measurement of NOEs in [^{1}H,^{1}H]-NOESY (nuclear Overhauser enhancement spectroscopy) allows one to estimate distances between protons: NOEs effectively constitute a "molecular ruler." Thus, having a large number of such ^{1}H$-^1$H distance constraints, one can calculate

three-dimensional molecular structures (see Sect. 5).

Residual dipolar couplings (RDCs) arise if a protein is partially aligned in a dilute liquid crystalline medium. Notably, for paramagnetic metalloproteins, such alignment takes place at high magnetic fields without the presence of a liquid crystalline medium. Owing to the partial alignment, the dipolar through-space coupling between dipole moments is not completely averaged out, as in the case of molecules tumbling in an isotropic medium. RDCs encode orientational constraints, that is, they provide information regarding the orientation of the internuclear axis of the dipolarly coupled nuclei relative to the principle axes of the alignment tensor. RDC-derived orientational constraints are thus quite distinct from – and complementary to – NOE-derived distance constraints for NMR structure calculation and validation. Since RDCs depend on motional averaging, they also provide valuable information about internal motional modes.

Nuclear spin relaxation parameters, such as longitudinal/transverse spin relaxation times and heteronuclear Overhauser effects reflect dynamic features of the protein under investigation. Hence, the majority of our insights into protein dynamics, be it overall rotational reorientation or internal motions, have thus far been obtained from measurement of these parameters.

4 Recently Developed Techniques

The introduction of (1) multidimensional triple resonance NMR and heteronuclear resolved [^{1}H,^{1}H]-NOESY, and (2) efficient ^{13}C/^{15}N labeling and protein deuteration protocols until the mid 1990s, made NMR a key tool to study structure and dynamics of proteins up to about 15 to 20 kDa. During the last decade, new NMR techniques were introduced to study much larger systems and to dramatically increase NMR data collection speed.

Transverse relaxation optimized NMR spectroscopy (TROSY) takes advantage of the mutual cancellation of nuclear spin relaxation pathways in slowly tumbling macromolecules studied at high magnetic fields (Fig. 5). 2D [^{15}N,^{1}H]-TROSY (Fig. 2a), and triple resonance NMR variants thereof, are most prominent and shall be discussed. Both, the ^{15}N and the ^{1}H nucleus of a polypeptide backbone H−N moiety exhibit substantial chemical shift anisotropy (CSA). As a result, two major pathways for transverse ^{15}N and ^{1}H spin relaxation are encountered: one owing to the dipolar ^{1}H−^{15}N interaction and the other owing to the CSA. While spin relaxation due to random fluctuation of the shielding is independent of spin states, the sign of the fluctuating local B-field arising from the spatial modulation of the dipolar interaction does depend on spin states. Hence, the transverse relaxation rate of a given ^{15}N or ^{1}H spin is accelerated when both the CSA and dipolar interaction random B-field fluctuations add up, whereas the transverse rate is reduced in cases where the two random fluctuations (partially) cancel. Since short transverse relaxation times lead to broad NMR lines, this phenomenon can be directly monitored in a 2D [^{15}N,^{1}H]-COSY experiment in which the one-bond ^{1}H−^{15}N J-coupling (~95 Hz) places the various transitions into separate spectral regions (no "^{15}N−^{1}H" decoupling; Fig. 5). Out of the four transitions registered for a given ^{15}N−^{1}H moiety, one is broadened in both dimensions, two are broadened in one dimension but narrow in the other,

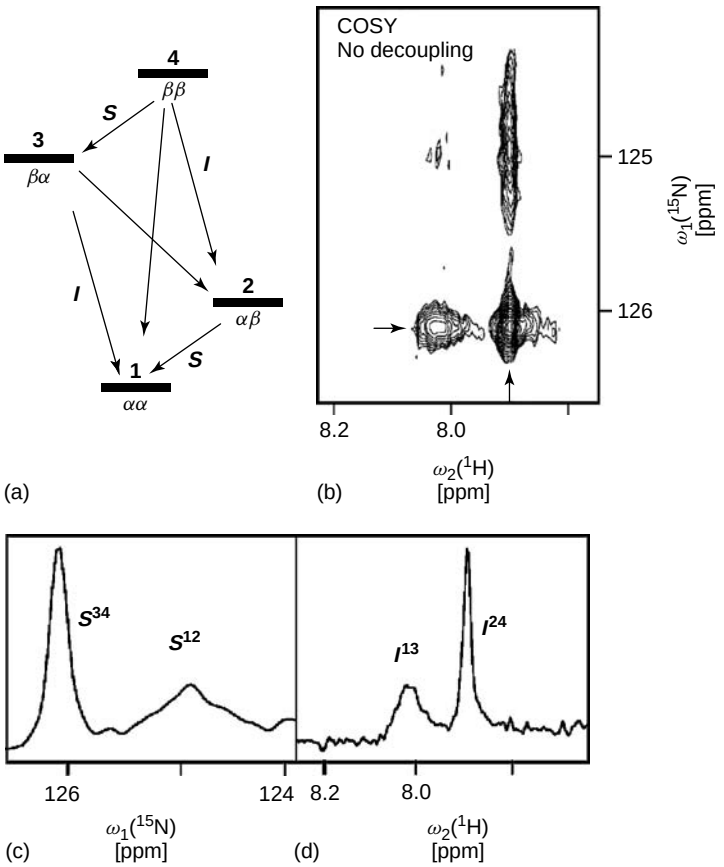

Fig. 5 (a) Energy levels of a scalarly coupled two spin-$\frac{1}{2}$ system IS (for example, $I = {}^1H$, $S = {}^{15}N$). The nuclear spin states are indicated as "α" and "β" depending on the orientation of the magnetic moments relative to the external magnetic field. (b) Cross-peak registered for a backbone N−H group in a [^{15}N,^1H]-COSY spectrum recorded (without decoupling of the $^{15}N-^1H$ scalar coupling interaction) for the 110-kDa protein aldolase at 20 °C. In this protein preparation, aliphatic and aromatic ^1H were replaced by ^2H ("deuterated") in order to eliminate dipolar interactions between backbone amide and carbon-bound protons. (c) and (d) 1D cross sections taken along ω_1 and ω_2 at the positions indicated in (b) by arrows. Comparison of line widths of the four components reflects the different transverse relaxation rates between the transitions shown in (a). A [^{15}N,^1H]-TROSY experiment relies on selecting the two transitions S^{34} and I^{24}. Reproduced with permission from Pervushin, K. (2000) Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology, *Q. Rev. Biophys.* **33**, 161–197.

and one is narrow in both dimensions. TROSY relies on selecting the latter narrow component, while discarding the others.

Since CSA-based transverse relaxation scales with $B_0{}^2$, TROSY is best performed at a magnetic field strength where the two relaxation pathways cancel (nearly) entirely. Fortunately, ^{1}H and ^{15}N CSA are of comparable magnitude (in Hz), so that about the same optimal TROSY field strength is predicted for both ^{1}H and ^{15}N transitions (corresponding to 900–1100 MHz ^{1}H resonance frequency). This makes triple resonance [^{15}N,^{1}H]-TROSY feasible. In conjunction with protein deuteration, this allows one to study systems with molecular masses well above 100 kDa, that is, up to about an order of magnitude larger than what could be approached without TROSY.

Methodology for rapid acquisition of NMR data focuses on resolving the "NMR sampling problem" (see Sect. 1), that is, on making high dimensional spectral information available at short measurement times. Currently, two broader classes of rapid data sampling approaches can be distinguished.

First, *G-matrix Fourier Transform* (GFT) NMR is a projection technique, which can also serve to reconstruct the higher-dimensionality parent NMR spectra one would obtain if no projection is applied. GFT NMR is based on the joint sampling of indirect chemical shift evolution periods that leads to signals with a multiplet fine structure, encoding several chemical shifts. A "G-matrix transformation" is applied in order to sort the components of the shift multiplets into subspectra, so that the peaks in each subspectrum encode a distinct linear combination of the jointly measured chemical shifts (Fig. 6). Hence, monitoring of chemical shifts in FT NMR

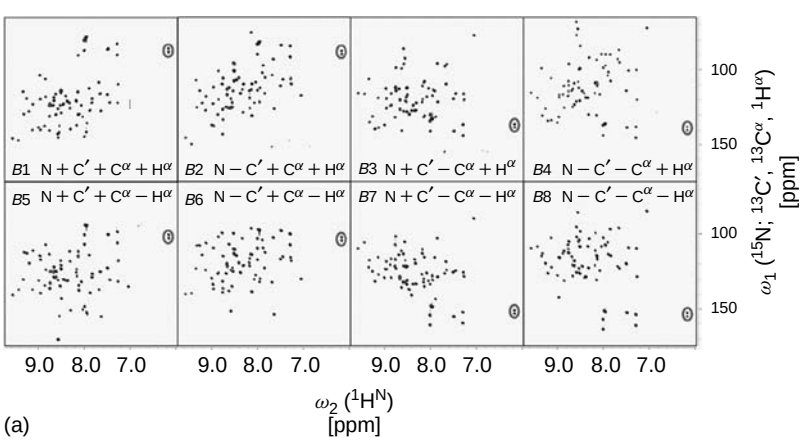

Fig. 6 15 2D planes constituting a (5,2)D HACACONHN GFT NMR experiment, which can provide the information of a 5D FT NMR experiment within less than an hour of measurement time. The type of linear combination of chemical shifts detected in a given plane is indicated. (a) The basic spectra. (b), (c) and (d) First, second, and third order central peak spectra in which linear combinations with fewer shifts are monitored to resolve assignment ambiguities. (e) Cross sections showing that signals do not broaden with an increasing number of shifts being jointly sampled. Reproduced with permission from Kim, S., Szyperski, T. (2003) GFT NMR, a new approach to rapidly obtain precise high dimensional NMR spectral information, *J. Am. Chem. Soc.* **125**, 1385–1393.

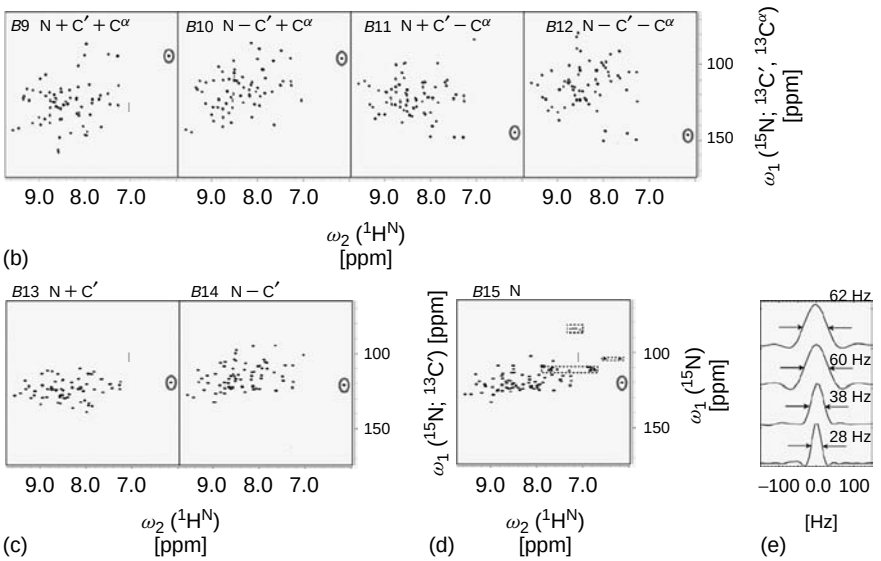

Fig. 6 (*Continued*)

spectroscopy is replaced in GFT NMR spectroscopy by measuring permutations of linear combinations of chemical shifts. This leads to approximately an order of magnitude reduction of measurement times *per* indirect dimension that is included in the joint sampling scheme.

Second, *single-scan acquisition* of 2D NMR spectra, named *ultrafast NMR*, was introduced. This approach constitutes an interface between high-resolution NMR and magnetic resonance imaging. The indirect chemical shift evolution is spatially encoded by use of pulse-field gradients applied along the axis of the external magnetic field B_0, and then "read out" together with the chemical shift evolution in the direct dimension in a single scan. In principle, this single-scan acquisition scheme can be extended to an arbitrary number of dimensions, that is, to the implementation of 3D and 4D single-scan acquisition. For proteins, however, reduced sensitivity currently limits this concept to the use of 2D single-scan acquisition. Importantly, ultrafast NMR is the only technique providing multidimensional NMR spectral information without sampling of indirect chemical shift evolution periods (Fig. 7). This allows one to acquire a multidimensional spectrum within the fraction of a second, and enables high-throughput acquisition of such spectra at unprecedented speed.

5 Structure Determination

NMR solution structures are usually solved in several major steps as outlined in the following. At the outset, a suitable sample (usually about 500 μL of a 1-mM protein solution) is prepared. Nowadays, both prokaryotic and eukaryotic high-yield overexpression systems serve for that purpose. If the molecular weight of a protein exceeds about 10 kDa, labeling with the NMR active spin-$\frac{1}{2}$ nuclei ^{13}C and ^{15}N is required to resolve spectral

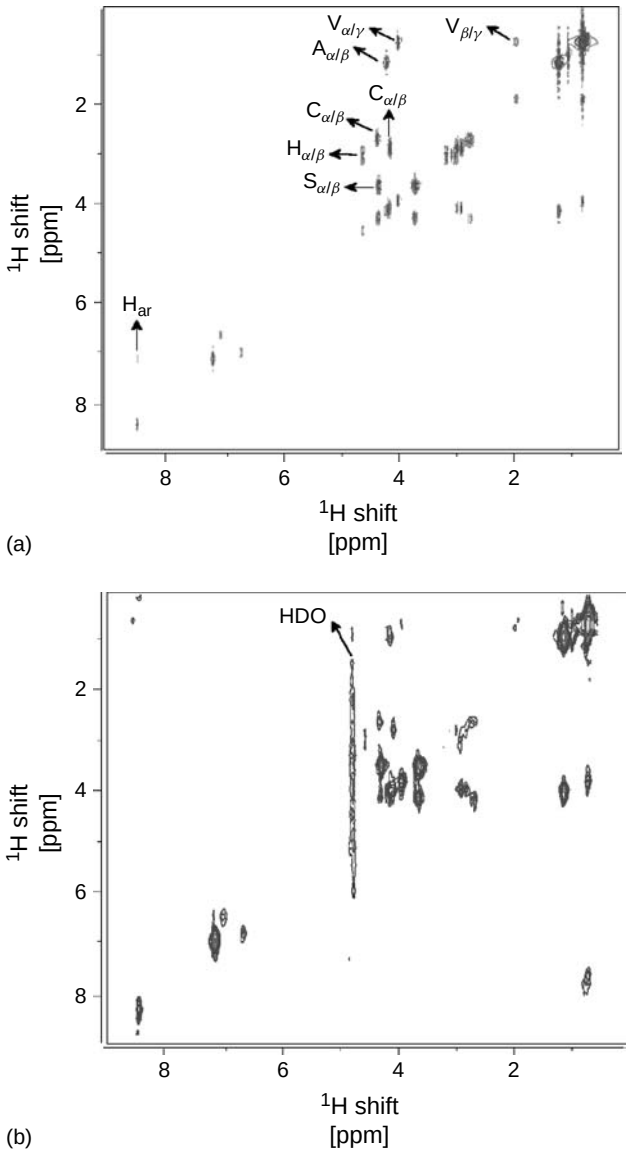

Fig. 7 Comparison of a conventionally acquired 2D [^{1}H,^{1}H] total correlation spectrum (a) acquired in 90 min for the hexapeptide CSHAVC (in the one-letter amino acid code) with the corresponding ultrafast congener recorded in 150 ms (b). In the upper panel, resonance assignments are indicated. Courtesy of Lucio Frydman.

overlap and to measure heteronuclear chemical shift correlations. For systems above about 25 kDa, one would usually consider deuterating the protein.

The sample thus obtained is used to record a set of multidimensional NMR experiments at temperatures around 30 °C. These allow (nearly) complete sequential NMR assignments to be obtained, and the conformation-dependent dispersion of the shifts enables one to derive experimental constraints for the NMR structure calculation. In most cases, structures are calculated on the basis of $^1H-^1H$ upper distance limit constraints derived from [^1H,^1H]-NOESY (Fig. 8).

The assignment of NOESY cross peaks and the calculation of the NMR structure is quite generally pursued in an iterative fashion. An initial set of distance constraints, which can be unambiguously obtained from the chemical shift data alone, is used to determine a low-resolution description of the protein. In turn, this initial structural model allows one to resolve the remaining NOE assignment ambiguities. This improves the precision of the structural model and allows the identification of yet additional constraints in the next round. Iterations involving structure calculations and identification of new constraints are usually pursued until (nearly) all experimentally derived constraints are in agreement with a bundle of protein conformations representing the NMR solution structure, in which conformational

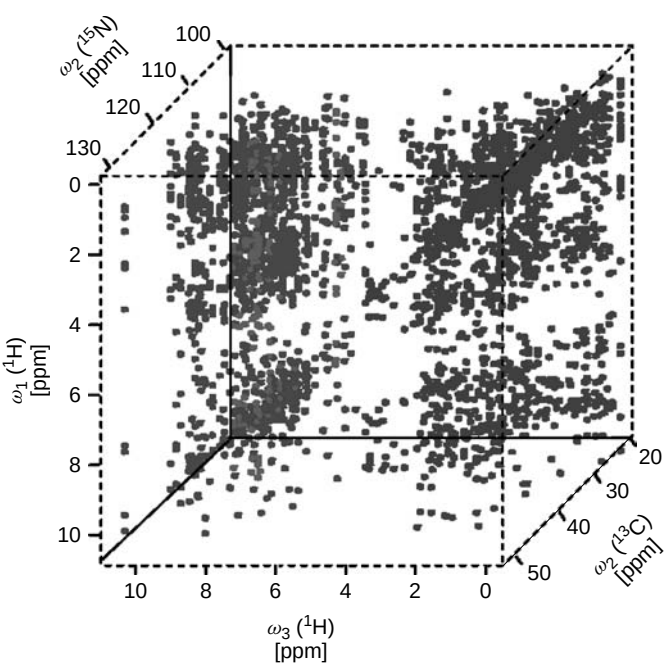

Fig. 8 Cube with dots representing NOEs detected in 3D heteronuclear resolved [^{1}H,^{1}H]-NOESY, which were used for calculating the NMR solution structure of YqfB shown in Fig. 9. Blue: signals detected along ω_3 on backbone amide protons; Red: detected on aromatic protons; Green: detected on aliphatic protons. Courtesy of Hanudatta Atreya. (See color plate p. xliv).

variations reflect the precision of the NMR structure determination. Finally, the NMR structure can be refined using molecular force fields, which in essence reflect our knowledge about conformational preferences in proteins.

In order to sample the "conformational space" allowed by experimental constraints, NMR structures are usually represented by an ensemble of conformers (typically ~10–30). Various representations (Fig. 9) are used to (1) display the structural information and (2) indicate the local and global precision of the structure determination, which is reflected by rmsd values calculated for sets of atom

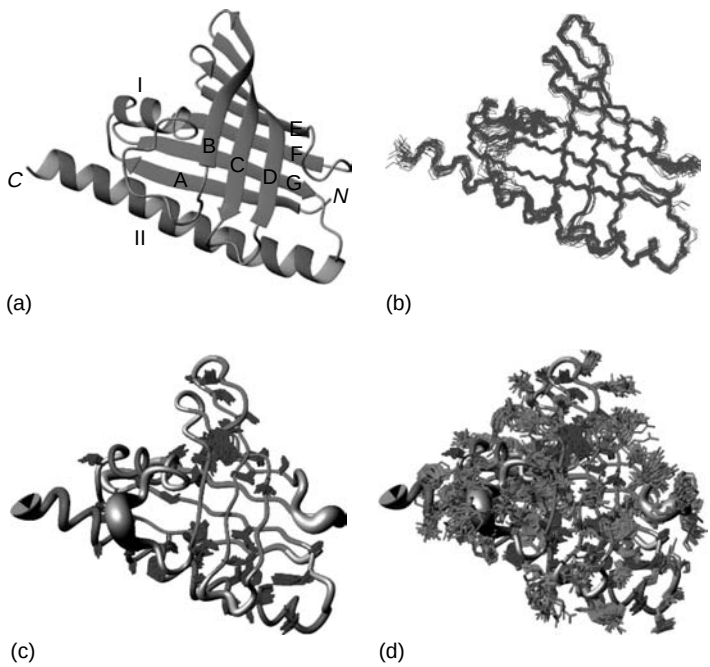

Fig. 9 Commonly used representations of NMR solution structures, exemplified for the 18-kDa Northeast Structural Genomics consortium target protein CC1736. (a) Ribbon drawing of the conformer exhibiting the smallest residual constraints violations. α-Helices and β-strands are depicted, respectively, in red/yellow, and cyan. (b) Superposition of the polypeptide backbone of the 20 conformers chosen to represent the NMR solution structure. The higher precision of the atomic coordinates of the regular secondary structure elements is apparent when compared with the loops. (c) The polypeptide backbone is represented by a spline function through the mean C^{α} coordinates, where the thickness represents the rmsd values for the C^{α}-coordinates after superposition of the regular secondary structure elements for minimal rmsd. The superpositions of the best defined side chains of the molecular core are shown in green to indicate the precision of their structural description. (d) Same as in (c), also showing the superposition of the more flexibly disordered side chains on the protein surface. The protein sample was provided by Drs. Acton and Montelione, Rutgers University. Courtesy of Yang Shen. (See color plate p. xlv).

positions for the ensemble of conformers. Additional criteria, such as the number and size of residual constraint violations or the completeness of stereo-specific resonance assignments, can be used to assess the quality of an NMR structure.

6 Dynamics

Proteins are not rigid but quite dynamic biological macromolecules. In fact, the internal mobility of proteins is of key importance for their function. Nuclear spin relaxation experiments provide unique insights into protein dynamics on the subnanosecond to millisecond timescale. This is because the KHz to MHz random magnetic field fluctuations associated with the motions can be monitored by measuring relaxation parameters. In particular, for ^{15}N, ^{13}C, and ^{2}H-labeled proteins, the measurement of corresponding nonproton relaxation parameters can afford a rather complete coverage of the entire protein.

In a typical protein dynamics study, longitudinal (T_1) and transverse relaxation times (T_2) and heteronuclear Overhauser effects are measured. These measurements can be complemented by use of multidimensional exchange spectroscopy, which provides information regarding very slow motional modes on the ~100 ms timescale. The first key challenge that is encountered for proper interpretation of relaxation parameters is the dissection of the impact arising from the overall rotational reorientation and the internal motional modes.

While the overall tumbling of the protein is important for "finding" an interaction partner, the internal motional modes are likely the key to linking function with dynamics. Proteins are usually not of spherical shape, so that their overall tumbling can be highly anisotropic. Any failure to accurately consider this anisotropy leads to a biased picture of the internal motions. Once the overall tumbling properties of a protein are known, the proper selection of an internal motional model remains as the second key challenge. In the so-called "model-free" approach, the internal motional mode is described by a minimal set of two parameters: an order parameter reflecting spatial motional restriction and a correlation time. Importantly, the model-free approach is limited to situations in which the internal motions occur on a faster timescale than the overall tumbling. In contrast, detailed assumptions on the nature of the internal motions are made within the framework of analytical models. These are inferred from either intuition or molecular dynamics simulations.

Comparable to reaction mechanisms in chemical kinetics, motional models can never be truly "proven" by measurement of spin relaxation parameters. Instead, one can only attest consistency between model and experimental parameters. This is because the complexity of possible spatial motional modes far exceeds the complexity that can be encoded in a rather small number of scalar relaxation parameters. In the same spirit, a more extensive body of data, for example, relaxation parameters measured for various types of nuclei at different magnetic field strengths, allows one to develop more refined multiparameter motional models. Very recently, the measurement of RDCs (see Sect. 3) has been established as a valuable complement to characterize protein dynamics.

Among the currently better characterized motional modes are low-amplitude

segmental motions of loops, large-amplitude flipping of aromatic rings and disulfide bonds, as well as the relative motion of entire domains in multidomain proteins. Moreover, important progress has been made in recent years to link the investigation of protein motion with functional aspects.

7 Hydration

Proteins have evolved in aqueous solution, and their surface interacts with water molecules. Hence, protein hydration plays a key role for understanding structure, dynamics, and function of proteins. In particular, macromolecular recognition involves protein surfaces: the hydration water molecules need to be "stripped off" prior to the formation of direct contacts between biological molecules. Hence, thermodynamics and kinetics of protein hydration deserve proper consideration in structural biology. Two NMR techniques provide complementary insights into protein hydration: measurement of $^{1}H-^{1}H$ NOEs between water molecules and protein, and measurement of ^{17}O and ^{2}H nuclear spin relaxation times at different magnetic field strengths.

For a few selected systems, such studies have shown that more than 95% of the water molecules being in contact with the protein surface are *less than* twofold motionally retarded compared to bulk water molecules. Hence, it appears that stripping off the hydration layer can hardly constitute a rate-limiting step for biomolecular interactions. This finding has evidently far-reaching consequences for understanding the kinetic properties of biological systems. In contrast to the very mobile hydration water molecules with "lifetimes" on the picosecond timescale, water molecules with longer lifetimes on the millisecond timescale have been identified and characterized. These may well play an integral structural role for proteins.

8 Folding

The rules governing the folding of a protein from an ensemble of "random coil" state into its native, functional conformation remain a major enigma of structural biology. NMR spectroscopy has been extensively employed to investigate thermodynamics and kinetics of protein folding. Measurement of chemical shifts (see Sect. 3), spin relaxation parameters (see Sect. 7), exchange spectroscopy (see Sect. 6), as well as determination of backbone amide proton–deuteron exchange rates can provide valuable insights. The majority of NMR studies focused on a few small model systems. These studies fostered the development of intriguing theoretical concepts shaping our view of how protein folding may take place, but have as yet not lead to algorithms enabling one to predict protein structure from sequence. In fact, even the existence of a concise algorithm to predict protein structure, for example, from first physical principles, could be called in question. John von Neumann noted in the framework of his automata theory, that systems may exhibit a degree of complexity, which is so high that their (empiric) description represents the easiest approach to assess them. Such complex systems have subsequently been named *von Neumann systems*, and one may ask the question whether proteins are possibly such systems. If so, it appears that the solution of the protein folding problem through structural

genomics (see Sect. 9) might be the most viable way to proceed. Notably, all methods currently available for "*ab initio*" prediction of protein structures, that is, without direct reference to an experimental template, rely on our empiric knowledge of protein structures. Recently, NMR studies have also proven the common existence of intrinsically unstructured but functional proteins, and showed that folding of proteins may take place upon binding to their physiological target molecules.

9 Structural Genomics

Since the genetic code is known, genomic DNA sequences can provide the amino acid sequence of all proteins encoded in a genome. However, since the protein folding problem (see Sect. 8) is not solved, the three-dimensional structure of the proteins cannot be inferred from sequence alone. One central aim of the new discipline of "structural genomics" is to make available at least one experimentally determined protein structure for each naturally occurring family of sequence homologs. This shall allow one to (1) homologically model the remaining members of the family using an experimental X-ray or NMR structure as a template, and (2) support the functional annotation of the gene encoding the protein. One may thus argue that structural genomics corresponds to seeking a "semi-empirical" solution of the protein folding problem.

Structural genomics relies on high-throughput structure determination

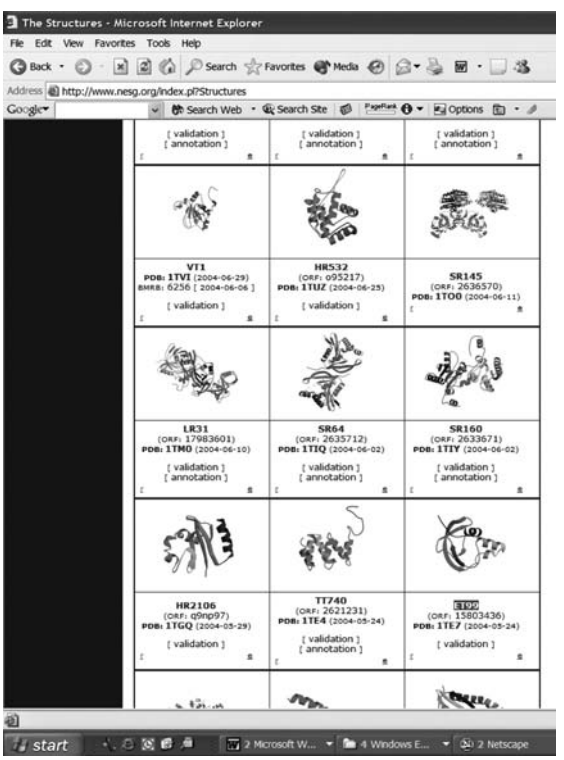

Fig. 10 "Structure gallery" accessible at the Web site of the Northeast Structural Genomics consortium in the United State (http://www.nesg.org), documenting successful implementation of a high-throughput NMR structure production pipeline.

(Fig. 10). Protein NMR spectroscopy has only very recently been shown to be a valuable technique for structural genomics (and thus also for systems biology). The ability to achieve protein structure determination (see Sect. 5) in high throughput depends primarily on: (1) sensitive high-field NMR spectrometers equipped with cryogenic probes (see Sect. 2); (2) methodology for rapid NMR data acquisition (see Sect. 4); (3) software for efficient analysis of NMR spectra and fast structure calculation; and (4) methods for automated structure quality assessment, validation, and data bank deposition. The setup of an efficient NMR-based structural genomics pipeline thus requires development of methodology, which strengthens the scientific infrastructure available for NMR-based structural biology in general.

See also Protein Modeling; Protein Structure Analysis: High-throughput Approaches; Proteomics; Structure-based Drug Design and NMR-based Screening.

Bibliography

Books and Reviews

Atreya, H., Szyperski, T. (2005) Rapid NMR data collection, *Methods Enzymol.* **394**, 78–108.

Bruschweiler, R. (2003) New approaches to the dynamic interpretation of NMR relaxation data from proteins, *Curr. Opin. Struct. Biol.* **13**, 175–183.

Cavanagh, J., Fairbrother, W.J., Palmer, A.G., Skelton, N.J. (1996) *Protein NMR Spectroscopy*, Academic Press, San Diego, CA.

Dyson, H.J., Wright, P.E. (2004) Unfolded proteins and protein folding studied by NMR, *Chem. Rev.* **104**, 3607–3622.

Ernst, R.R., Bodenhausen, G., Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford.

Montelione, G.T., Zheng, D., Huang, Y., Gunsalus, K.C., Szyperski, T. (2000) Protein NMR spectroscopy for structural genomics, *Nat. Struct. Biol.* **7**, 982–984.

Otting, G. (1997) NMR studies of water bound to biological molecules, *Prog. Nucl. Magn. Reson. Spectrom.* **31**, 259–285.

Pervushin, K. (2000) Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology, *Q. Rev. Biophys.* **33**, 161–197.

Sandström, J. (1982) *Dynamic NMR Spectroscopy*, Academic Press, London, UK.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

Primary Literature

Akke, M., Palmer, A. (1996) Monitoring macromolecular motions on microsecond to millisecond scales by R(1)rho-R(1) constant relaxation time NMR spectroscopy, *J. Am. Chem. Soc.* **118**, 911–912.

Arora, A., Abildgaard, F., Bushweller, J.H., Tamm, L.K. (2001) Structure of outer membrane protein A transmembrane domain by NMR spectroscopy, *Nat. Struct. Biol.* **8**, 334–338.

Balbach, J., Forge, V., Lau, W.S., van Nuland, N.A., Brew, K., Dobson, C.M. (1996) Protein folding monitored at individual residues during a two-dimensional NMR experiment, *Science* **274**, 1161–1163.

Bax, A. (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics, *Protein Sci.* **12**, 1–16.

Bruschweiler, R., Liao, X., Wright, P.E. (1995) Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling, *Science* **268**, 886–889.

Cano, K.E., Thrippleton, M.J., Keeler, J., Shaka, A.J. (2004) Cascaded z-filters for efficient single-scan suppression of zero-quantum coherence, *J. Magn. Reson.* **167**, 291–297.

Castellani, F., van Rossum, B., Diehl, A., Schubert, M., Rehbein, K., Oschkinat, H. (2002) Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy, *Nature* **420**, 98–102.

Chou, J.J., Li, H., Salvesen, G.S., Yuan, J., Wagner, G. (1999) Solution structure of BID, an intracellular amplifier of apoptotic signaling, *Cell* **96**, 615–624.

Fiaux, J., Bertelsen, E.B., Horwich, A.L., Wuthrich, K. (2002) NMR analysis of a 900K GroEL GroES complex, *Nature* **418**, 207–211.

Frydman, L., Scherf, T., Lupulescu, A. (2002) The acquisition of multidimensional NMR spectra within a single scan, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15858–15862.

Gardner, K.H., Kay, L.E. (1998) The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins, *Annu. Rev. Biophys. Biomol. Struct.* **27**, 357–406.

Gutmanas, A., Billeter, M. (2004) Specific DNA recognition by the Antp homeodomain: MD simulations of specific and nonspecific complexes, *Proteins* **57**, 772–782.

Hare, B.J., Wyss, D.F., Osburne, M.S., Kern, P.S., Reinherz, E.L., Wagner, G. (1999) Structure, specificity and CDR mobility of a class II restricted single-chain T-cell receptor, *Nat. Struct. Biol.* **6**, 574–581.

Kalodimos, C.G., Biris, N., Bonvin, A.M., Levandoski, M.M., Guennuegues, M., Boelens, R., Kaptein, R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes, *Science* **305**, 386–389.

Kim, S., Szyperski, T. (2003) GFT NMR, a new approach to rapidly obtain precise high dimensional NMR spectral information, *J. Am. Chem. Soc.* **125**, 1385–1393.

Kitahara, R., Yokoyama, S., Akasaka, K. (2005) NMR snapshots of a fluctuating protein structure: ubiquitin at 30bar-3kbar, *J. Mol. Biol.* **347**, 277–285.

Korzhev, D.M., Salvatella, X., Vendruscolo, M., DiNardo, A.A., Davidson, A.R., Dobson, C.M., Kay, L.E. (2004) Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR, *Nature* **29**, 586–590.

KupŠe, E., Freeman, R. (2003) Projection-reconstruction of three-dimensional NMR spectra, *J. Am. Chem. Soc.* **125**, 13958–13959.

Lipari, G., Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, *J. Am. Chem. Soc.* **104**, 4546–4559.

Liu, G., Mills, J.L., Hess, T.A., Kim, S., Skalicky, J.J., Sukumaran, D.K., Kupce, E., Skerra, A., Szyperski, T. (2003) Resonance assignments for the 21 kDa engineered fluorescein-binding lipocalin FluA, *J. Biomol. NMR* **27**, 187–188.

Luginbühl, P., Pervushin, K.V., Iwai, H., Wuthrich, K. (1997) Anisotropic molecular rotational diffusion in ^{15}N spin relaxation studies of protein mobility, *Biochemistry* **36**, 7305–7312.

Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E., Wüthrich, K. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids, *J. Mol. Biol.* **280**, 933–952.

Modig, K., Liepinsh, E., Otting, G., Halle, B. (2003) Dynamics of protein and peptide hydration, *J. Am. Chem. Soc.* **126**, 102–114.

Monleon, D., Colson, K., Moseley, H.N.B., Anklin, C., Oswald, R., Szyperski, T., Montelione, G.T. (2002) Rapid data collection and analysis of protein resonance assignments using AutoProc, AutoPeak, and AutoAssign Software, *J. Struct. Funct. Genom.* **2**, 93–101.

Montelione, G.T. (2001) Structural genomics: an approach to the protein folding problem, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13488–13489.

Mueller, T.D., Feigon, J. (2003) Structural determinants for the binding of ubiquitin-like domains to the proteasome, *EMBO J.* **22**, 4634–4645.

Orekhov, V., Pervushin, K.V., Arseniev, A.S. (1994) Backbone dynamics of (1–71)bacterioopsin studied by two-dimensional 1H-15N NMR spectroscopy, *Eur. J. Biochem.* **219**, 887–896.

Otting, G., Wüthrich, K. (1991) Protein hydration in aqueous solution, *Science* **254**, 974–980.

Perez Canadillas, J.M., Varani, G. (2003) Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein, *EMBO J.* **22**, 2821–2830.

Pervushin, K., Riek, R., Wider, G., Wüthrich, K. (1997) Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12366–12371.

Peti, W., Meiler, J., Bruschweiler, R., Griesinger, C. (2002) Model-free analysis of protein backbone motion from residual dipolarcouplings, *J. Am. Chem. Soc.* **124**, 5822–5833.

Powers, R., Garrett, D.S., March, C.J., Frieden, E.A., Gronenborn, A.M., Clore, G.M. (1992) Three-dimensional solution structure of human interleukin-4 by multidimensional heteronuclear magnetic resonance spectroscopy, *Science* **256**, 1673–1677.

Reif, B., Hennig, M., Griesinger, C. (1997) Direct measurement of angles between bond vectors in high-resolution NMR, *Science* **276**, 1230–1233.

Serber, Z., Dötsch, V. (2001) In-cell NMR spectroscopy, *Biochemistry* **40**, 14317–14323.

Skalicky, J.J., Sukumaran, D.K., Mills, J.L., Szyperski, T. (2000) Toward structural biology in supercooled water, *J. Am. Chem. Soc.* **122**, 3230–3231.

Szyperski, T., Luginbühl, P., Otting, G., Güntert, P., Wüthrich, K. (1993) Protein dynamics studied by rotating frame ^{15}N spin relaxation times, *J. Biomol. NMR* **3**, 151–164.

Tjandra, N., Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, *Science* **278**, 1111–1114.

Tolman, J.R., Flanagan, J.M., Kennedy, M.A., Prestegard, J.H. (1995) Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9279–9283.

Volkman, B.F., Lipson, D., Wemmer, D.E., Kern, D. (2001) Two-state allosteric behavior in a single-domain signalling protein, *Science* **291**, 2429–2433.

Von Neumann, J. (1966) *Theory of Self-Reproducing Automata*, University of Illinois Press, Urbana, IL.

Wüthrich, K. (2003) NMR studies of structure and function of biological macromolecules (Nobel lecture), *Angew. Chem., Int. Ed. Engl.* **42**, 3340–3363.

Yabuki, T., Kigawa, T., Dohmae, N., Takio, K., Terada, T., Ito, Y., Laue, E.D., Cooper, J.A., Kainosho, M., Yokoyama, S. (1998) Dual amino acid-selective and site-directed stable-isotope labeling of the human c-Ha-Ras protein by cell-free synthesis, *J. Biomol. NMR* **11**, 295–306.

Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le, B., Ramelot, T., Lee, G.M., Bhattacharyya, S., Gutierrez, P., Denisov, A., Lee, C.H., Cort, J.R., Kozlov, G., Liao, J., Finak, G., Chen, L., Wishart, D., Lee, W., McIntosh, L.P., Gehring, K., Kennedy, M.A., Edwards, A.M., Arrowsmith, C.H. (2002) An NMR approach to structural proteomics, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1825–1830.

Zheng, D., Aramini, J.M., Montelione, G.T. (2004) Validation of helical tilt angles in the solution NMR structure of the Z domain of Staphylococcal protein A by combined analysis of residual dipolar coupling and NOE data, *Protein Sci.* **13**, 549–554.

Protein Phosphorylation: *see* Biological Regulation by Protein Phosphorylation

Protein Purification

Richard R. Burgess
University of Wisconsin, Madison, WI, USA

1 Introduction 183

2 Types of Molecular Interactions and Variables that Affect Them 183
2.1 Hydrogen Bonds 184
2.2 Hydrophobic Interactions 184
2.3 Ionic Interactions 184
2.4 Variables that Affect Molecular Forces 185

3 Protein Properties that Can Be Used as Handles for Purification 185
3.1 Size 186
3.2 Shape 186
3.3 Charge 186
3.4 Isoelectric Point 187
3.5 Charge Distribution 187
3.6 Hydrophobicity 187
3.7 Solubility 188
3.8 Density 188
3.9 Ligand Binding 189
3.10 Metal Binding 189
3.11 Reversible Association 189
3.12 Posttranslational Modifications 189
3.13 Specific Sequence or Structure 189
3.14 Unusual Properties 189
3.15 Genetically Engineered Purification Handles 190
3.16 What Can Be Learnt from the Amino Acid Sequence of a Protein that is Useful in Purification? 190

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

4 Types of Separation Methods 191

5 Protein Inactivation and How to Prevent It 191

6 Protein Purification Strategy 194

7 Overproducing Recombinant Proteins 194

8 Refolding Proteins Solubilized from Inclusion Bodies 196
8.1 Increasing Production of Soluble Protein 196
8.2 Refolding Inclusion Bodies 196
8.3 A General Procedure for Refolding Proteins from Inclusion Bodies 197

Bibliography 197
Books and Reviews 197
Primary Literature 198

Keywords

Fractionation
The process of separating different proteins based on differences in their properties, such as size, charge, shape, solubility, and binding properties.

Inclusion Body
An insoluble form of a recombinant protein often observed when the protein is overproduced in a bacterial expression host.

Protein
A macromolecule formed by the covalent joining of amino acids to make a polypeptide. The properties of a protein are determined by the number and sequence of its amino acids, and how the polypeptide chain folds into a native structure. Proteins can function as enzymes, as structural components, or as regulators.

Protein Refolding
Protein in an inclusion body is solubilized with a denaturant and then the denaturant is removed to allow the protein to reform its native structure, determined by its primary sequence, and its biological activity.

Protein purification is the process of separating a given protein from all the other proteins, nucleic acids, polysaccharides, lipids, metabolites, and other small molecules in a cell extract. This purification is done ideally with reasonable speed, yield, final purity, and economy, while retaining the biological activity and chemical

and structural integrity. A purification scheme involves a series of fractionation steps that exploit the many differences in the properties of the protein of interest and the other components in the mixture. Design of such fractionation steps must take into account the molecular forces that determine molecular interactions between proteins and various materials used in purification. Care must be taken to prevent inactivation, degradation or absorptive losses during the purification. In recent years, most researchers purify recombinant proteins whose genes have been cloned into an expression vector and expressed in a suitable expression host organism, often the bacterium *Escherichia coli*. The use of purification tags to aid in affinity purification is common. Sometimes, the overproduced protein is found as an insoluble inclusion body that must be isolated, solubilized by denaturants, and then refolded.

1 Introduction

The magnitude of the challenge of protein purification becomes clearer when one considers the mixture of macromolecules present in a cell extract. In addition to the protein of interest, several thousand other proteins with different properties are present in the extract, along with nucleic acids (DNA and RNA), polysaccharides, lipids, and small molecules. The proteins present in the bacterium *Escherichia coli* may be dramatically visualized after resolution by two-dimensional gel electrophoresis as shown in Fig. 1. A given protein may be present at more than 10% or at less than 0.001% of the total protein in the cell. Enzymes are found in different states and locations: soluble, insoluble, membrane bound, DNA bound, in organelles, cytoplasmic, periplasmic, and nuclear. The challenge, therefore, is to separate the protein of interest from all of the other components in the cell, especially the unwanted contaminating proteins, with reasonable efficiency, speed, yield, and purity, while retaining the biological activity and chemical integrity of the polypeptide.

Sections 1 to 6 provide background on classical protein fractionation and purification; that is, the isolation of a protein from its natural source. Sections 7 and 8 give a brief introduction to overproduction and purification of recombinant proteins cloned and overexpressed in a bacterial host expression system.

2 Types of Molecular Interactions and Variables that Affect Them

With regard to protein structure and stability and the interaction between an individual protein and other proteins, DNA, or materials used in protein purification, one must understand the molecular forces involved and how the strength of these forces varies as one varies conditions such as temperature, pH, and ionic strength of a solution. The atomic interactions that seem to be the most important with regard to protein interactions are hydrogen bonds, hydrophobic interactions, and ionic interactions. These are described briefly below.

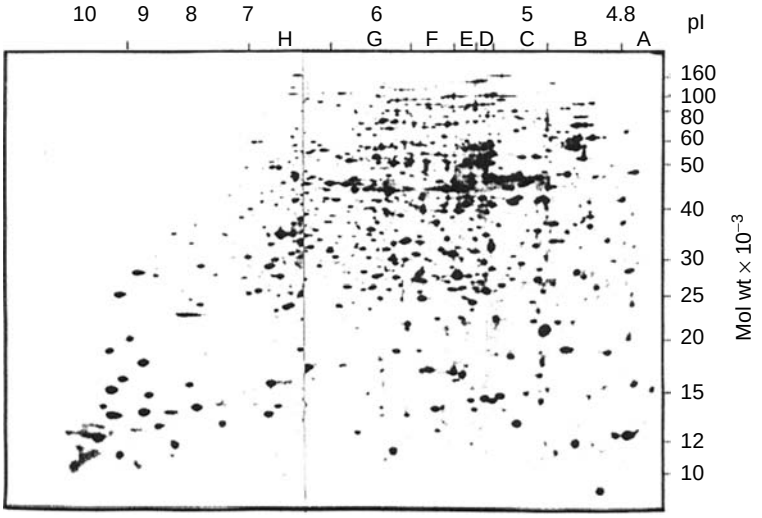

Fig. 1 *E. coli* proteins resolved on a two-dimensional gel. The approximate isoelectric point and molecular weight scales are indicated. *E. coli* K12 strain W3110 was labeled with $^{35}SO_4$ during growth in glucose minimal medium at 37 °C. A composite autoradiogram was made from nonequilibrium (*left side*) and pH 5–7 (*right side*) isoelectric focusing gels. (Adapted, with permission, from Neidhardt, F.C., Phillips, T.A. (1985) The Protein Catalog of *E. coli*, in: Celis, J.E., Bravo, R. (Eds.) *Two-Dimensional Gel Electrophoresis of Proteins*, Academic Press, New York, pp. 417–444.)

2.1 Hydrogen Bonds

Hydrogen bonds (Fig. 2) occur when a proton is shared between a proton donor (−NH and −OH) and a proton acceptor (O=C− and :N−). Optimal hydrogen bonds have a linear geometry and a distance between the donor and acceptor atoms between 2.6 and 3.1 Å. Hydrogen bonds are stronger at low temperature and are weakened as the temperature is raised.

2.2 Hydrophobic Interactions

Nonpolar residues (isoleucine, leucine, valine, phenylalanine, and tryptophan) cannot make favorable hydrogen bonds with water. In order to avoid water, they tend to come together in a so-called hydrophobic interaction (see Fig. 2) usually resulting in their being buried in the interior of a protein. Hydrophobic interactions are strengthened at high salt and high temperature.

2.3 Ionic Interactions

Ionic interactions (see Fig. 2) occur between charged molecules, with like charges repelling and opposite charges attracting. The force of the electrostatic interaction is given by an approximation of Coulomb's law, $E = Z_A Z_B e^2 / D r_{AB}$, where r_{AB} is the distance between two charges, A and B, Z_A, and Z_B are their respective number of unit charges, e is one

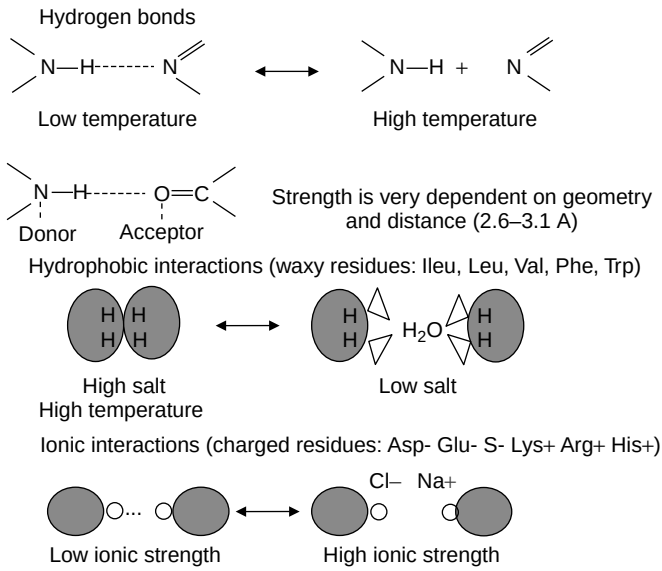

Fig. 2 Hydrogen bonds, hydrophobic interactions, ionic interactions.

unit of electronic charge, and D is the dielectric constant of the solvent. The strength of ionic interactions is therefore inversely proportional to the distance between the charges and the dielectric constant of the solvent, which varies from 2 in nonpolar solvents like hexane to 80 in highly polar solvents such as water. Ionic interactions are weakened as the ionic strength of the solvent increases and the charge is shielded by counterions. Ionic interactions are affected by the pH of the solution, since pH determines the number of charged residues.

2.4 Variables that Affect Molecular Forces

One can vary conditions to affect the relative strength of the above molecular forces. One can easily vary temperature, ionic strength, ion type, dielectric constant, and pH. In a few cases, researchers have also varied pressure.

3 Protein Properties that Can Be Used as Handles for Purification

The reason one is able to purify one protein from a mixture of thousands of proteins is that proteins vary tremendously in a number of their physical and chemical properties. These properties are the result of proteins having different numbers and sequences of amino acids. The amino acid residues attached to the polypeptide backbone may be positively or negatively charged, neutral and polar, or neutral and hydrophobic. In addition, the polypeptide is folded in a very definite secondary structure (α-helices, β-sheets, and various turns) and tertiary structure to create a unique size, shape, and distribution of residues on the surface of the protein. By exploiting the differences in properties between the protein of interest and other proteins in the mixture, one can design a rational series of fractionation steps. These properties include the following.

3.1 Size

Proteins may vary in size from peptides of a few amino acids (with molecular weights of a few hundred) to very large proteins containing over 10 000 amino acids (with molecular weights of over 1 000 000). Most proteins have molecular weights in the range 10 000 to 150 000 (see Fig. 1). Proteins that are part of multisubunit complexes may reach much larger sizes. Proteins are often fractionated on the basis of size (really on the basis of effective radius or Stokes' radius) by passing down a gel-filtration column (or size-exclusion column (SEC). The column is filled with porous beads with characteristic pore sizes. The largest proteins cannot penetrate into the bead and are excluded and elute first in what is called the void or excluded volume. Very small proteins and salts easily pass in and out of the beads and see the entire volume of the column (the column volume). Other intermediate-sized proteins elute between the void and the column volume based on how much time they spend outside and inside the beads.

3.2 Shape

Protein shapes range from approximately spherical (globular) to quite asymmetric. The shape of a protein influences its movement through a solution during centrifugation, through small pores in membranes, into beads during gel filtration, or through gels during electrophoresis. For example, consider two monomeric proteins of the same mass where one is spherical and the other is cigar shaped. During centrifugation through a glycerol gradient, the spherical protein will have a smaller Stokes' radius and, thus, will encounter less friction as it sediments through the solution. It will sediment faster, and thus, appear to be larger than the cigar-shaped protein. On the other hand, during size-exclusion chromatography, the same spherical protein with its smaller Stokes' radius will more readily diffuse into the pores of a gel-filtration bead and will elute later, thus appearing smaller than the cigar-shaped protein.

3.3 Charge

The net charge of a protein is determined by the sum of the positively and negatively charged amino acid residues. If a protein has a preponderance of aspartic and glutamic acid residues, it has a net negative charge at pH 7 and is termed an acidic protein. If it has a preponderance of lysine and arginine residues, it is considered to be a basic protein. The equilibrium between charged and uncharged groups and hence the charge of a protein is determined by the pH of the solution. The charge of the ionizable groups found on unmodified proteins as a function of pH is shown in Table 1. Generally, one used a positively charged resin (an anion-exchange column) to bind a negatively charged protein and a negatively charged resin (a cation-exchange column) to bind a positively charged protein. Bind the protein to the column at low salt (e.g. 0.1 M NaCl) and elute with an increasing salt gradient. At some stage, the ionic attraction of the protein to the column resin will become weak enough to cause the protein to dissociate from the column and elute.

Tab. 1 The charge of the ionizable groups found on unmodified proteins as a function of pH.

Ionizable group	pK_a[a]	*pH 2*	*pH 7*	*pH 12*
C-terminal (COOH)	4.0	0 0 0 0 0	- -	
Aspartate (COOH)	4.5	0 0 0 0 0 0	- -	
Glutamate (COOH)	4.6	0 0 0 0 0 0	- -	
Histidine (imidazole)	6.2	+ + + + + + + + + +	0 0 0 0 0 0 0 0 0 0 0	
N-terminal (amino)	7.3	+ + + + + + + + + + +	0 0 0 0 0 0 0 0 0	
Cysteine (SH)	9.3	0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 - - - - - - - -	
Tyrosine (phenol)	10.1	0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 - - - - - - -	
Lysine (amino)	10.4	+ + + + + + + + + + +	+ + + + + + + + + + +	0 0 0 0
Arginine (guanido)	12.0	+ + + + + + + + + + +	+ + + + + + + + + + +	+ 0

[a] pK_a is the pH at which the ionizable group is half ionized. The precise pK_a value for a given ionizable group can be influenced by the immediate local environment.

3.4 Isoelectric Point

The isoelectric point (pI) is the pH at which the charge on a protein is zero and is determined by the number and titration curves of the positively and negatively charged amino acid residues on the protein. Protein pI values generally range from 4 to 10 (Fig. 1). An example of a theoretical titration curve and pI determination of *E. coli* RNA polymerase transcription factor, sigma32 (σ^{32}), is shown in Fig. 3.

3.5 Charge Distribution

The charged amino acid residues may be distributed uniformly on the surface of the protein or they may be clustered such that one region is highly positive while another region is highly negative. Such nonrandom charge distribution can be used to discriminate among proteins. An example is the *E. coli* σ^{32} protein (Fig. 3). At pH 7.9, σ^{32} has a negative charge of -46 and a positive charge of $+40$, giving a net charge of -6. It is able to bind reasonably tightly to both anion- and cation-exchange columns, apparently because its charged residues are not evenly distributed on the surface. This property can be used to purify this protein because most proteins will not bind to both types of ion-exchange columns under a single solvent condition.

3.6 Hydrophobicity

Most hydrophobic amino acid residues are buried on the inside of a protein, but some are found on the surface. The number and spatial distribution of hydrophobic amino acid residues present on the surface of the protein determine the ability of the protein to bind to hydrophobic column materials (as in hydrophobic interaction chromatography or HIC) and, therefore, can be exploited in fractionation. Generally, load a protein mixture onto a HIC column at high salt (e.g. 1 M ammonium sulfate), where hydrophobic interactions are strongest, and then elute with a decreasing salt gradient to successively elute more and more tightly bound proteins.

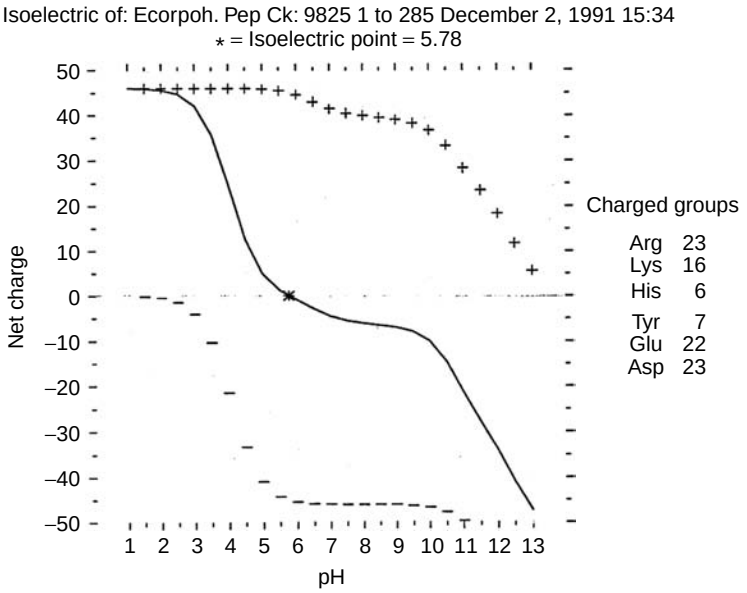

Fig. 3 Titration curve and isoelectric point (pI) of *E. coli* σ^{32}. This graph shows theoretical plots of the number of positively charged and negatively charged groups and the net charge as a function of pH for the *E. coli* RNA polymerase transcription factor σ^{32}, based on its amino acid sequence. The pI is indicated by the asterisk and is 5.78. The charged groups are Arg (23), Lys (16), His (6), Tyr (7), Glu (22), and Asp (23). This plot was generated using the Genetics Computer Group Sequence Analysis Software package.

3.7 Solubility

Proteins vary dramatically in their solubility in different solvents, all the way from being essentially insoluble ($<10\ \mu g\ mL^{-1}$) to being very soluble ($>300\ mg\ mL^{-1}$). Key variables that affect the solubility of a protein include pH, ionic strength, the nature of the ions, temperature, and the polarity of the solvent. Proteins are generally less soluble at their isoelectric point where there is less charge repulsion. Proteins are commonly fractionated by adding higher and higher concentrations of the mild salt, ammonium sulfate. Generally, the solubility of a given protein will decrease about 10-fold as the ammonium sulfate increases about 6% in saturation. (It takes about 760 gm of ammonium sulfate added to a liter of water to give a 100% saturated solution, which is about 4.1 M, at 20 °C). Since ammonium sulfate is mild and stabilizing to proteins, is relatively inexpensive and pure, and is highly soluble, it is the most common material used to fractionate proteins on the basis of solubility.

3.8 Density

The density of most proteins is between 1.3 and $1.4\ g\ cm^{-3}$ and this is not generally a useful property for fractionating proteins. However, proteins containing large amounts of phosphate (e.g. phosvitin, density = 1.8) or lipid

moieties (e.g. β-lipoprotein, density = 1.03) are substantially different in density compared with the average protein and may be separated from the bulk of proteins using density methods.

3.9
Ligand Binding

Many enzymes bind substrates, effector molecules, cofactors, or DNA sequences quite tightly. This binding affinity can be used to bind an enzyme to a column to which the appropriate ligand or DNA sequence has been immobilized. For example, the transcription factor AP-1 is purified by binding to a specific DNA affinity column.

3.10
Metal Binding

Many enzymes bind certain chelated metal ions (e.g. Cu^{++}, Zn^{++}, Ca^{++}, Co^{++}, and Ni^{++}) quite tightly, usually through interactions with cysteine or histidine residues. This binding can be used to bind an enzyme to a column to which the appropriate chelated metal ion has been immobilized. See Sect. 3.15 on the use of metal-chelate column for purification of protein tagged by the addition of 6 to 10 terminal histidines.

3.11
Reversible Association

Under certain solution conditions, some enzymes aggregate to form dimers, tetramers, and so on. For example, the ability of *E. coli* RNA polymerase to be a dimer under one condition (0.05 M NaCl) and a monomer under another condition (0.3 M NaCl) can be used if two fractionations based on size are carried out sequentially under those two different conditions.

3.12
Posttranslational Modifications

After protein synthesis, many proteins are modified by the addition of carbohydrates, acyl groups, phosphate groups, or a variety of other moieties. In many cases, these modifications provide handles that can be used in fractionation. For example, proteins containing carbohydrates on their surface can often be bound to columns containing plant lectins, which are molecules capable of binding tightly to certain carbohydrate moieties on glycoproteins. Phosphoproteins will in some cases bind to a chelated Fe^{++} column.

3.13
Specific Sequence or Structure

The precise geometric presentation of amino acid residues on the surface of a protein can be used as the basis of a separation procedure. For example, an antibody that recognizes only a particular site (epitope) on a protein can usually be obtained. An immunoaffinity column can be prepared by attaching a monospecific antibody (which binds only to the protein of interest) to a resin. Immunoaffinity chromatography (IAC) can result in highly selective separation and provides a very effective purification step. One can also immobilize a protein of interest and use it to specifically bind another protein out of a complex protein extract. This process is called *protein affinity chromatography*.

3.14
Unusual Properties

In addition to the types of properties mentioned above, certain proteins have unusual properties that can be exploited during their purification – an example is

unusual thermostability. Most proteins unfold and coagulate or precipitate when heated to 95 °C. A protein that remains soluble and active after such heat treatment can be separated easily from the bulk of the other cellular proteins. Another such property is unusual resistance to proteases. These two properties often go hand in hand. An interesting example of a purification involving these properties is that of *E. coli* alkaline phosphatase. The cellular extract is heated and the insoluble coagulated proteins are removed by centrifugation. The supernatant that contains the phosphatase is then treated with a protease, which digests the remaining contaminating proteins, leaving an essentially pure preparation of alkaline phosphatase.

3.15 Genetically Engineered Purification Handles

With the advent of genetic engineering, it has become relatively easy to clone the cDNA encoding a given protein. It is then possible to construct an overproducing strain of *E. coli* that can be induced to produce large amounts of a desired gene product. Recently, it has become common to alter the cDNA in such a way as to add a few extra amino acids on the amino terminus or the carboxyl terminus of the protein being expressed. This added "tag" can be used as an effective purification handle. One of the most popular tags is addition of 6 to 10 histidines onto the amino terminus of a protein. One then purifies the protein by its ability to bind tightly to a column containing chelated Ni^{++} or Co^{++} in which it can be washed and then eluted with free imidazole or by lowering the pH to 5.9, where histidine becomes fully protonated and no longer binds to chelated metal.

3.16 What Can Be Learnt from the Amino Acid Sequence of a Protein that is Useful in Purification?

These days it is very common to purify a protein in which the gene has been sequenced. Thus, one can easily deduce the amino acid sequence of the corresponding protein. Can this knowledge help in designing a purification scheme? The answer is that it can be somewhat, but not very, helpful. It is easy to determine the precise molecular weight of the polypeptide chain, but not to predict whether it forms dimers or tetramer or is part of a multisubunit complex. Its charge versus pH and its isoelectric point can be determined as shown in Fig. 3, which gives some idea as to which type of ion-exchange column to use, but again this will only be useful if it is not associated with other proteins. It is possible to calculate its extinction coefficient on the basis of its tryptophan and tyrosine content, which is very useful when it is pure. It is possible to determine if it has membrane-spanning regions or it has potential modification sites. One may be able to deduce that it is a member of a larger family of proteins by sequence alignment or by the presence of conserved sequence motifs that suggest cofactor affinity. However, its shape is not known, since the three-dimensional structure from sequence cannot yet be reliably predicted. Its multisubunit features cannot be predicted. Its ammonium sulfate precipitation properties cannot be predicted. Surface features such as hydrophobic patches, charge distribution, or antigenic sites cannot be predicted. Therefore, one must conclude that protein purification is still an empirical science.

4 Types of Separation Methods

There are a large number of separation processes that can be utilized to fractionate proteins on the basis of the properties listed above. These are summarized in Table 2. The sequential use of several of these separation processes will allow the progressive purification of almost any protein. If the processes are chosen carefully, and if proper attention is paid to separation conditions, and to maintaining the stability of the protein, the purification will result in reasonable efficiency, speed, yield, and purity, while retaining the biological activity and chemical integrity of the polypeptide.

An example of a hypothetical protein fractionation scheme is shown in Fig. 4. This scheme relies on three of the most common fractionation methods, ammonium sulfate precipitation, ion-exchange chromatography, and gel-filtration or size-exclusion chromatography. The purification summary in Table 3 is a typical way of summarizing the yield and specific activity of each of the major steps in a purification scheme.

5 Protein Inactivation and How to Prevent It

A protein purification scheme will generally not be considered successful if the result is a protein that is pure, but inactive. Therefore, one of the key considerations in working with a protein is to prevent it from becoming inactivated.

Tab. 2 Separation processes that can be utilized to fractionate proteins.

Separation process	***Basis of separation***
Precipitation	
Ammonium sulfate	Solubility
Acetone	Solubility
Polyethyleneimine	Charge, size
Isoelectric	Solubility, pI
Phase partitioning (e.g. with polyethylene glycol)	Solubility
Chromatography	
Ion exchange (IEX)	Charge, charge distribution
Hydrophobic interaction (HIC)	Hydrophobicity
Reverse-phase HPLC	Hydrophobicity, size
Affinity	Ligand-binding site
DNA affinity	DNA-binding site
Lectin affinity	Carbohydrate content and type
Immobilized metal affinity (IMAC)	Metal binding
Immunoaffinity (IAC)	Specific antigenic site
Chromatofocusing (CF)	pI
Gel filtration/size exclusion (SEC)	Size, shape
Electrophoresis	
Gel electrophoresis (PAGE)	Charge, size, shape
Isoelectric focusing (IEF)	pI
Centrifugation	Size, shape, density
Ultrafiltration	Size, shape

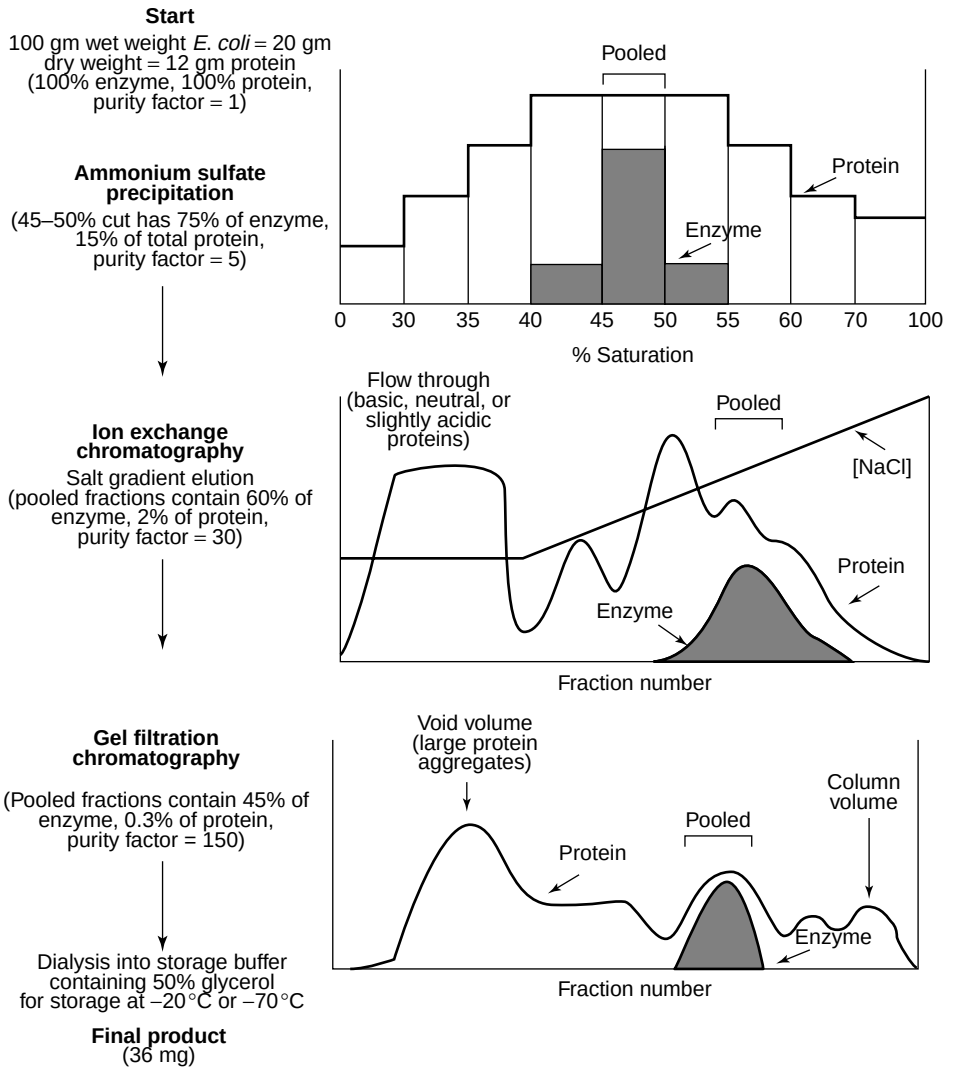

Fig. 4 A hypothetical purification scheme.

Table 4 summarizes some of the main reasons why a protein might become inactivated and what can be done to prevent this inactivation. In general, one works quickly and at low temperature (in a cold room or ice bucket) to avoid proteolytic degradation. One avoids foaming or undue exposure to oxygen and adds a reducing agent to prevent oxidation. One uses a buffer to maintain pH and a chelating agent like EDTA to protect against heavy-metal ions. Addition of 5% glycerol seems to stabilize most proteins and reduce adsorption to the walls of the container. A low salt concentration (e.g. 100 mM) helps increase solubility and prevent ionic adsorption to surfaces. Table 5 gives a good all-purpose buffer that in most

Tab. 3 Summary of hypothetical purification in Fig. 4.

Fraction	*Total Protein [mg]*	*Total Activity [%]*	*Specific Activity*	*Step Yield [%]*	*Overall Yield [%]*
Extract	12 000	100	=1		=100
				75	
AS pptn (45–50%)	1800	75	5		75
				80	
IEC (pooled peak)	240	60	30		60
				75	
Gel filtration (peak)	36	45	150 fold purification		45 final yield
Pure standard			150		

Tab. 4 Protein inactivation and ways of preventing it.

Reasons for inactivation	*How to prevent it*
Oxidation, foaming	Add DTT or TCEP, store under argon
Protease degradation	Add protease inhibitors, cooler, purer
Adsorption to container	Use polypropylene tubes, BSA carrier, glycerol nonionic detergent, protein more concentrated
Aggregation and precipitation	Store less concentrated, add salt, pH away from pI
Heavy metals	Add EDTA, cleaner tube, reagents
Temperature inactivation	Store cooler, add ligand or glycerol to stabilize
Bacterial growth	Use Tris, EDTA, azide, avoid PO_4, OAc^-
Enzymatic reaction (phosphatase)	Cooler, purer, add specific inhibitor
Dissociation of subunits/cofactors	Store more concentrated
pH changed	Avoid CO_2 in room, Tris changes pH with temperature
Inactive/misfolded conformation	Incubate at 37 °C to anneal the structure

Tab. 5 A good all-purpose buffer for keeping proteins happy.

TGED + 0.1 M NaCl	
Buffer	50 mM Tris-HCl, pH 7.9 at 20 °C
Stabilizer	5% glycerol
Chelator	0.1 mM EDTA
Reducing agent	0.1 mM DTT (dithiothreitol)
Salt	0.1 M NaCl
Storage buffer – similar to above but has 50% glycerol. Will not freeze at −20 °C. Best stored at −70 °C.	

cases will keep a protein active and happy.

6
Protein Purification Strategy

How one designs a purification scheme will, in large part, determine how successful one will be in achieving the goal of a protein purification: *that of getting a high yield of highly pure and active protein in the minimal number of steps*. Achieving a high final yield requires a high recovery at each step. Four steps at 80% step yield will be $(0.8)^4 = 0.41 = 41\%$ final yield, while 4 steps at 60% step yield will give a 13% final yield. Final purity will be guided by the intended use of the protein, but it should be free of major contaminants and any traces of enzymes that interfere with the intended use. High activity will depend on maintaining the stability of the protein as discussed in Sect. 5. By choosing high-resolution fractionation steps, one can achieve a given fold purification in fewer steps. For example, if the target protein is 0.01% of the total protein in the extract, it will require a 10^4-fold purification. This can be achieved in 4 steps, each giving a 10-fold purification, or 3 steps with a 22-fold purification, or 2 steps capable of 100-fold purification. The fewer the steps the faster the preparation, the lower the protein losses the lower the cost of the purification procedure. Some of the key considerations in designing a purification procedure are (1) to have a convenient assay to follow purification; (2) choose a starting material rich in protein; (3) take precautions to minimize damage, inactivation or loss; (4) use the minimal number of steps; (5) remove the bulk of material quickly; (6) avoid unnecessary duplication, dialysis, and delay; (7) generally use fractionation steps in the order: precipitation, ion exchange, affinity, sizing; and (8) use high-resolution steps where possible.

7
Overproducing Recombinant Proteins

The advent of genetic engineering has given us the ability to routinely clone genes and overproduce their gene products. This has changed the way we think about protein purification. For most protein purifications, one no longer starts with large amounts of naturally occurring material. Instead, clone the gene of interest, insert it into a suitable expression vector, transform the vector into a suitable expression host (e.g. *E. coli*), grow the cells, induce the transcription of the gene of interest, harvest the cells, break open the cells and purify the overproduced recombinant protein. By using an expression vector where the target gene has a strong, inducible promoter, it is possible to express the target to levels as high as 20 to 40% of the total cellular protein.

The most commonly used bacterial expression system was developed by Bill Studier and colleagues at Brookhaven National Laboratory, in the late 1980s. The host strain BL21(DE3) is a derivative of *E. coli* B that is deficient in several proteases to help prevent proteolysis of the recombinant protein and that has an inducible copy of the T7 phage RNA polymerase integrated as a phage lambda lysogen in the bacterial chromosome. The T7 RNA polymerase is kept repressed by the lactose operon repressor due to the placement of a lac operator near the promoter. The gene of interest is inserted into a multicopy expression vector under the control of a T7 RNA polymerase

promoter, also with a lac operator to keep it off. The vector also contains an extra copy of the lac repressor gene to enhance repression. The repressor and the presence of an additional plasmid (pLysS, that encodes a T7 lysozyme to inhibit low levels of T7 RNA polymerase activity) help keep uninduced transcription of the target gene low. When the cells have been grown to the desired cell density, the lac operon inducer, IPTG, is added to about 1 mM, causing the repressor to dissociate from the lac operators and allowing expression of the T7 RNA polymerase. The T7 RNA polymerase in turn actively transcribes the many copies of the plasmid-encoded recombinant gene and the resulting mRNA is efficiently translated into the protein of interest. The result can be the production of very large amounts of the recombinant protein, to as much as 20 to 40% of the total cell protein within about 4 h after induction. This means that it is possible to purify as much as 30 to 50 mg of recombinant protein from 1 gm of wet weight bacterial cell paste.

Several refinements have been introduced in the last few years to improve the chances that overproduction will be successful, especially if one is trying to overproduce a mammalian, plant, or archaeal protein. One problem is that human proteins often use codons that are rarely used in *E. coli*. These codons correspond to *E. coli* tRNAs that are very low in abundance in the cell. When one tries to overexpress a gene containing many of these rare codons, the result is that very little, if any, of the protein is produced. This problem was originally solved by changing the DNA sequence of the recombinant gene so that it did not contain rare codons, but it contained the preferred *E. coli* codons. A much easier and more elegant solution has now been developed and is marketed by several biotech research products companies. This involves creation of an improved host bacterial strain that has had 3 to 5 of its rare tRNAs augmented. Many poorly expressed proteins can now be expressed at very high levels. Table 6 lists this and several other problems that have been encountered in protein overexpression in *E. coli* along with how these problems have been solved or alleviated. There are a wide variety of elegantly engineered expression vectors and bacterial expression hosts available from many different biotechnology research

Tab. 6 Problems of poor recombinant protein expression and their solutions.

Problem	***Solution***
1. Target gene contains rare *E. coli* codons	Supplement host *E. coli* with rare tRNAs
2. Target mRNA is degraded	Use *E. coli* strain deficient in RNase E
3. Target protein is toxic to host cells	Use tighter repression, lower copy plasmid
4. Target protein is a membrane protein	Use a strain that has extra internal membranes
5. Target protein needs to form disulfide bonds	Use strain that is deficient in several key reductases, Gor, TrxB
6. Product normally forms stable heterodimers	Simultaneous coexpression of two different proteins in one *E. coli* strain

products companies worldwide. In addition to many *E. coli*-based expression hosts, there are also expression hosts such as: *Bacillus subtilis*, *Pichia pastoris*, *Aspergillus*, baculovirus/insect cells, mammalian cells, plants, and animals.

8 Refolding Proteins Solubilized from Inclusion Bodies

One of the most common problems in overexpressing a recombinant protein in *E. coli* is the fact that while large amounts of the protein are produced, most of it is not soluble, but is found as an insoluble inclusion body. Apparently, the newly synthesized protein, when partially folded into its native structure, exposes some hydrophobic regions and is quite sticky and prone to interaction with other partially folded proteins, leading to aggregation and inclusion body formation. There are two main approaches to dealing with inclusion bodies: (1) try to increase the proportion of the overproduced protein that is soluble; and (2) purify the inclusion body, solubilize it by dissolving it in a protein denaturant, and then refold it into its native structure.

8.1 Increasing Production of Soluble Protein

To purify the soluble material, one will want to increase as much as possible the proportion of the overproduced protein that is soluble. The most common approach is to induce the overproduction in cell growing at 20 to 25 °C. Apparently, the slower growth rate and lower temperatures results in more refolded protein and less aggregation and inclusion body formation. People have tried coexpressing cloned chaperone proteins to facilitate proper folding, but this is not common as it is only effective in some cases. An elegant recent approach, has been to grow the cells at 37 °C, shift the temperature briefly to 42 °C to induce the heat shock response, and then shift the cells to 20 °C for induction. Often if two proteins that normally form stable heterodimers are individually overexpressed they are insoluble, but if coexpressed in the same cell they form soluble, native heterodimers. Finally, many proteins remain soluble when overexpressed as genetic fusions with known proteins that readily fold to form stable native structures (protein fusion partners like NusA, GST, and TrxA).

8.2 Refolding Inclusion Bodies

If one chooses to refold protein solubilized from inclusion bodies, it is common to first wash the inclusion bodies with a nonionic detergent like TritonX-100 to solubilize membranes and break any unbroken cells. The washed inclusion body is then almost pure, and is ready to be solubilized. The real challenge is not the purification, but the refolding. The key to refolding without reaggregation and precipitation is to refold under low protein concentration. Under these conditions, the concentration of sticky, partially refolded material is lower, decreasing the opportunity for interaction and aggregation. However, for larger preparations, the large volumes become a major problem. Usually, one can find refolding conditions that give efficient refolding yields at reasonable protein concentrations.

8.3 A General Procedure for Refolding Proteins from Inclusion Bodies

A general method that the author has found to be quite effective for many proteins is given below.

a. Grow cells and induce overexpression of target protein.
b. Harvest cells, weigh cell pellet, store frozen at -80 °C
c. Break cells by sonication (pLysS cells are easy to break because of the presence of some T7 lysozyme). Otherwise adding lysozyme helps.
d. Centrifuge cell lysate, wash the IB pellet with 1% TritonX-100 to solubilize membranes and membrane proteins, then wash with buffer to remove TritonX-100.
e. Solubilize IB with a denaturant such as 6 M guanidium hydrochloride or 0.3% Sarkosyl to about 1 mg protein mL^{-1}. Difficult-to-refold proteins may need to be diluted to 0.1 mg mL^{-1} in denaturant; 8 M urea can also be used, but there is a risk of carbamylation of the protein.
f. Centrifuge out any undissolved material and slowly drip dilute the solubilized protein into 15 to 60 volumes of suitable refolding buffer. Additives or various buffer variables are often used to improve folding efficiency of a particular protein. These include 0.5 M arginine, 25% glycerol, varying the pH, temperature, presence of divalent ions, and presence of redox buffers.
g. Pass dilute refolded protein over a suitable high-resolution ion-exchange column, wash the column, and then elute with an increasing salt gradient. This step accomplishes several important things: (1) it concentrates the dilute protein; (2) it removes the denaturant; (3) it removes impurities that do not bind to the column or bind weaker or stronger than the target protein; and finally (4) it often separates refolded monomer from soluble multimers that tend to bind tighter and elute later in the gradient.
h. The resulting protein is usually fully active, homogeneous, and capable of forming crystals suitable for three-dimensional structure determination.

See also Proteomics.

Bibliography

Books and Reviews

Burgess, R.R. (1987) Protein Purification, in: Oxender, D., Fox, C.F. (Eds.) *Protein Engineering*, A.R. Liss, New York, pp. 71–82.

Coligan, J.E., Dunn, B.M., Ploegh, H.L., Speicher, D.W., Wingfield, P. (1997) *Current Protocols in Protein Science*, John Wiley & Sons, New York.

Creighton, T.E. (1993) *Proteins: Structures and Molecular Properties*, 2nd edition, W.H. Freeman, San Francisco, CA.

Ford, C.F., Suominen, I., Glatz, C.E. (1991) Fusion tails for the recovery and purification of recombinant proteins, *Protein Expr. Purif.* **2**, 95–107.

Marschak, D., Kadonaga, J., Burgess, R., Knuth, M., Brennan, W., Lin, S.-H. (1996) *Strategies for Protein Purification and Characterization: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor.

Neidhardt, F.C., Phillips, T.A. (1985) The Protein Catalog of *E. coli*, in: Celis, J.E., Bravo, R. (Eds.) *Two-Dimensional Gel Electrophoresis of Proteins*, Academic Press, New York, pp. 417–444.

Schein, C.H. (1989) Production of soluble recombinant proteins in bacteria, *Biotechnology* **7**, 1141–1149.

Scopes, R. (1994) *Protein Purification: Principles and Practice*, 3rd Edition, Springer-Verlag, NY.

Simpson, R.J. (Ed.) (2004) *Purifying Proteins for Proteomics: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor.

Primary Literature

Bessette, P.H., Aslund, F., Beckwith, J., Georgiou, G. (1999) Efficient folding of proteins with multiple disulfide bonds in the *E. coli* cytoplasm, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13703–13708.

Burgess, R.R. (1969) A new method for the large-scale purification of *E. coli* DNA-dependent RNA polymerase, *J. Biol. Chem.* **244**, 6160–6167.

Gill, S., von Hippel, P. (1989) Calculation of protein extinction coefficients from amino acid sequence data, *Biochemistry* **182**, 319–326.

Held, D., Yaeger, K., Novy, R. (2003) Coexpression vectors, *Innovations* **18**, 4–6.

Hochuli, E., Bannworth, W., Dobeli, R., Gentz, R., Studber, D. (1988) Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent, *Biotechnology* **6**, 1321–1325.

Lopez, P.J., Marchand, I., Joyce, S.A., Dreyfus, M., RNase, E. (1999) The C-terminal half of which organizes the *E. coli* degradosome, participates in mRNA degradation but not rRNA processing in vivo, *Mol. Microbiol.* **33**, 188–199.

Miroux, B., Walker, J. (1996) Over-production of proteins in *E. coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels, *J. Mol. Biol.* **260**, 289–298.

Novy, R., Drott, D., Yaeger, K., Mierendorf, R. (2001) Overcoming codon bias of *E. coli* for enhanced protein expression, *Innovations* **12**, 1–3.

Porath, J. (1992) Immobilized metal ion affinity chromatography, *Protein Expr. Purif.* **3**, 206–281.

Studier, W., Rosenberg, A., Dunn, J., Dubendorff, J. (1990) Use of T7 RNA polymerase to direct expression of cloned genes, *Methods Enzymol.* **185**, 60–89.

Protein Repertoire, Evolution of

Christine Vogel[1], *Rajkumar Sasidharan*[1], *and Emma E. Hill*[2]
[1] *MRC Laboratory of Molecular Biology, Cambridge*
[2] *University of California, Berkeley, California*

1 General Characteristics of the Protein Repertoire 203
1.1 Sequence, Structure, and Function 203
1.2 Related Proteins Form Families 205
1.3 Most Proteins Fold into Structures 205
1.4 Domains are the Units of Protein Evolution 206
1.4.1 Evolutionary Hierarchical Domain Classification (SCOP) 207
1.4.2 Alternative Domain Classifications 209
1.4.3 Implications and Complications of Domain Definitions 210
1.4.4 Domain Prediction 211
1.5 How Many Different Families, Superfamilies, and Folds Exist in Nature? 212
1.6 The Distribution of Protein and Domain Families Across Different Genomes 212
1.7 Proteins Interact 214
1.8 Proteins are Part of Pathways 215
1.9 Protein and Domain Function 216
1.9.1 Indirect Function Characteristics 216
1.9.2 Functional Annotation of Domains 217

2 Mechanisms in the Evolution of Single Proteins and the Protein Repertoire 217
2.1 Orthology and Paralogy 217
2.2 Duplication – Mechanisms to Create "More" 218
2.2.1 Horizontal Gene Transfer and Transposons 219
2.2.2 Duplication Leads to Protein and Domain Family Expansions 219
2.2.3 Reconstructing the Evolutionary History of Proteins 220
2.3 Divergence after Duplication 221
2.3.1 Changes in Sequence 221

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

2.3.2 Changes in Function 222
2.3.3 Structural Tolerance to Amino Acid Changes – a Case Study on Globins 223
2.3.4 Changes in Structure 224
2.4 Recombination – Mechanisms to Create ''Novel Combinations'' 225
2.4.1 Most Proteins Have More than One Domain 226
2.4.2 The Repertoire of Domain Combinations is Limited and Highly Biased 226
2.4.3 The Sequential Order of Domains and Their Geometry are Conserved 226
2.5 Selective and Random Processes 227

3 Summary and Conclusions – Evolutionary Principles 228

Bibliography 229
Books and Reviews 229
Primary Literature 230

Keywords

Alternative splicing
Creation of several different protein variants from one gene sequence using different exons.

Convergence
Independent invention of similar features.

Domain
A distinct (evolutionary) unit within a protein – can be defined from a structural, sequence-based, functional, or evolutionary perspective. Here, we concentrate on the structural and evolutionary domains.

Domain accretion
Tendency of eukaryote proteins to have more domains than their prokaryote homologs.

Domain architecture
The total number and type of domains within a protein and their relative orientations to one another.

Domain combination
Two or more structural domains that have combined at some time within a protein generally in tandem, though, domain insertions also exist.

Domain or protein family
Closely related sequences of domains or whole proteins.

Essential gene; essentiality
Gene with a lethal knockout phenotype.

Gene silencing
Upon a gene duplication event, the redundant gene copy is turned nonfunctional.

Homolog; homologous
Inferred evolutionary relationship.

Inserted domains
Protein domain A inserted into another domain B, so that B is interrupted in sequence.

Linker region; linker sequence
Amino acid sequence joining two domains.

Neo-functionalization
Emergence of novel protein function, usually after relaxation of selection upon a gene duplication event.

Nonsynonymous substitutions
Nucleotide substitutions that change the type of resulting amino acid.

Orphan (sequence)
Sequence without detectable homologs. Also called *singleton*.

Orthologue; orthologous
Homologs; homologous sequences related by descent, that is, speciation.

Paralogue; paralogous
Homologs; homologous sequences in one organism, related by gene duplication.

Protein Repertoire
The full complement of proteins encoded within an organism's genome.

Proteome
The fraction of the protein repertoire that is expressed in a particular organism or cell at a particular time.

Pseudogenes
Presumably untranscribed and untranslated, formerly active but now nonfunctional genes.

Sequence–structure–function Relationship
Relationship between the sequence, structure, and function of a protein. Protein evolution is marked by changes in these three characteristics at different rates.

Structural class
Top level of the hierarchical classification of protein domains. Groups domains according to their composition of alpha-helices and beta-strands.

Subfunctionalisation
Modification and divergence protein function, usually after relaxation of selection upon a gene duplication event.

Superfamily
One or more families of related domains or whole proteins.

Superfold
Notion of folds that are populated by many superfamilies.

Synonymous substitutions
Nucleotide substitutions that do not affect the type of resulting amino acid.

Transcriptome
The complete collection of transcribed elements from a genome.

Unifold
Notion of folds that comprise only one or very few superfamilies.

Abbreviations

aa: amino acid (residues)
BDIM: birth-death-innovation models
CATH: class(C), architecture(A), topology(T) and homologous superfamily (H)
DNA: deoxyribonucleic acid
3D: 3-dimensional
EC: enzyme commission
GO: gene ontology
HGT: horizontal gene transfer
HMM: hidden Markov model
NCBI: National Center for Biotechnology Information
NRDB: non-redundant database
NTP: nucleotide triphosphate
PDB: protein data bank
RMSD: root mean square deviation
RNA: ribonucleic acid
SCOP: structural classification of proteins
TIM: triose phosphate isomerase
WHD: winged helix domain

Proteins are the key players in our cells. Their entirety is what we call an organism's protein repertoire, encoded in its genome. Proteins are made up of smaller units called *domains*. Although single-domain proteins exist, the majority of proteins consist of at least two domains; the individual domains and nature of their interactions determine the function of the protein. Today, we have information on the sequence and structure of many proteins and domains. These data enable us to identify, describe, and compare the protein repertoires of a wide range of organisms and how they are determined by the tight interplay of sequence, structure, and function. The main evolutionary processes that form the protein repertoire are duplication, recombination, and divergence of the encoding sequences. Here, we describe these processes and how they affect protein structure and function. This allows us to draw general conclusions on the evolution of the protein repertoire on the organismal scale.

1 General Characteristics of the Protein Repertoire

Millions of different species inhabit our planet. One of the most fundamental problems in biology is to describe, organize, and understand the evolution of organisms that have very diverse characteristics. Although natural selection does not necessarily lead to an increase in complexity as organisms evolve, it is apparent that the complexity of certain lineages, such as our own, has increased during evolution.

The anatomy and physiology of an organism and its changes during evolution are determined primarily by its protein repertoire. The modularity of proteins, in turn, reveals the versatility of living organisms and their amazing ability to constantly evolve. One can imagine an increase in complexity by employment of three main mechanisms: (1) the invention of novel *modules*, for example, proteins or cells; (2) an increase in abundance of existing modules and their divergence; or (3) novel interactions between existing modules. The first of these mechanisms, *de novo* protein construction is relatively rare.

Here, we examine the protein repertoire, how it was formed via the duplication and recombination of existing modules, and how this contributed to the multitude of organisms that exist today. This provides, in contrast to single protein analyses, a much broader evolutionary insight into phylogenetic relationships. Understanding how the protein repertoire has evolved to produce functionally diverse and distinct proteins permits an appreciation of how speciation and evolution have occurred at a higher level.

1.1 Sequence, Structure, and Function

Proteins function as enzymes, regulators, messengers, or receptors; they form cellular structures, transporters, and storage devices. Proteins enable cells to grow, divide, and communicate with other cells or to respond to environmental input. They are found in cellular compartments or even outside the cell, attached to

the membrane or freely moving around. Proteins bind other molecules, such as DNA, RNA, lipids, metal ions, or carbohydrates, but they also interact with each other, forming large structures or functional complexes.

Proteins are characterized by the intimate relationship between their sequence, structure, and function (Fig. 1). Proteins of similar structure can have an enormous variety of functions, and an example of these are the TIM barrels. In contrast, the same function, for example binding to DNA, can be fulfilled by proteins of most different structures.

Similarly, the sequence and structure of a protein are intrinsically linked, and the chemistry of certain residues along with their surface location is responsible for maintaining the function of binding to the receptor protein. Even though apparently simple, understanding the relationships between these characteristics is crucial for an understanding of protein evolution. The amino acid sequence determines the structure a protein forms. Homologous proteins are of common evolutionary descent, and they often have similar sequences. Such sequence similarity is easy to detect and quantify with tools available today. However, protein structure is more conserved than sequence, and, frequently, the structure of related proteins will be similar despite the sequences having greatly diverged from one another.

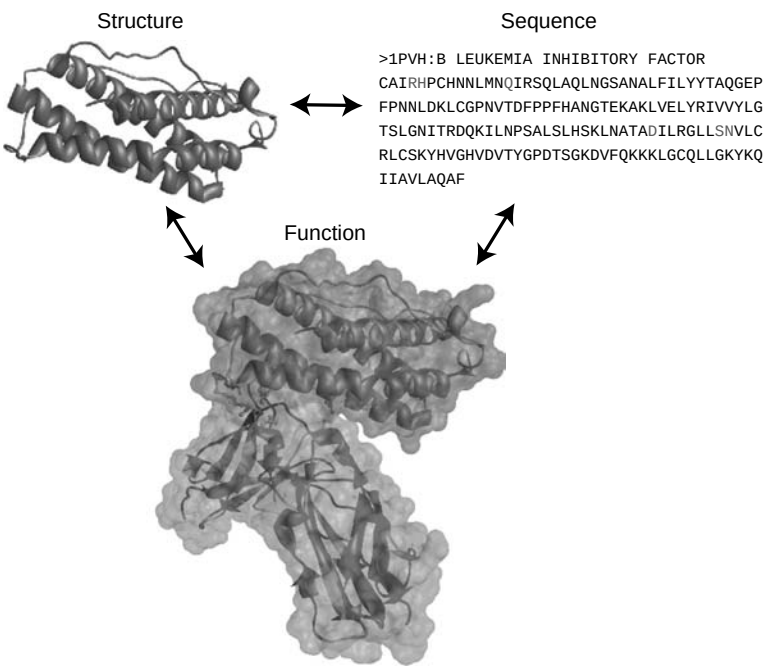

Fig. 1 Proteins are characterized by their sequence, structure, and function. The sequence, individual domain structure (colored purple), and interacting structure (with Gp130 colored blue) of leukemia inhibitory factor are shown (Protein Data Bank (PDB) ID: 1pvh). Within the sequence and interacting domains residues that are colored red form hydrogen bonds with the receptor. (See color plate p. xxxiii.)

In the triangular relationship between sequence, structure, and function (Fig. 1), small changes at the level of the sequence can produce either large or small changes in the structure, which, depending on location, may impact the protein's function. A single nucleotide substitution can produce an amino acid change, which might alter specificity or binding affinity. A classic example of this is the discovery of sickle cell anemia by Vernon Ingram in 1956, the first disease that was documented on a molecular level. Ingram showed that the replacement of glutamic acid by valine at position 6 in hemoglobin β-chain was responsible for the abnormality. This substitution produces alterations in the structure such that haem binding and thus iron transport is compromised, resulting in sickle cell–shaped red blood cells and serious anemia.

In the following sections, we describe some aspects of the sequence, structure, and function of proteins and how they relate to each other.

1.2 Related Proteins Form Families

In order to understand the protein repertoire and how it evolves, we need to know the evolutionary or phylogenetic relationships between the proteins. Margaret Dayhoff was amongst the first to introduce the notion that proteins can be grouped into *families* and *superfamilies* of related molecules and recognized the role of these relationships in protein evolution. These concepts of families and superfamilies have played heavily into construction of our present-day classifications of proteins.

A family was defined as a group of related sequences based on sequence similarity. The evolutionary relationship is obvious from the amino acid sequence alone: if two sequences are highly similar, the proteins are likely to be related. Many methods exist to detect these sequence similarities today, for example, BLAST. As we commented on earlier, the structure of two related proteins can be remarkably similar despite their sequence having diverged beyond recognizable similarity. In superfamilies, which can comprise one or more related families, sequence similarity is often undetectable, and the evolutionary relationship is demonstrated by structural and functional information.

A wonderful example of related proteins with minimal or no sequence similarity is the four-helical cytokine superfamily. These proteins are known to bind to receptors and trigger signaling cascades functioning in immune and circulatory systems. Distinctive structural features reveal, however, that four-helical cytokines can be subdivided into three different families: the interferons that have a fifth helix, the short chain cytokines that have a small two-stranded beta-sheet and subtle folding differences, and finally the long-chain cytokines, which have longer chains. The structure-based family classification was confirmed by analysis of the gene structure, which showed remarkable conservation of intron/exon structure with respect to the secondary structure elements.

1.3 Most Proteins Fold into Structures

Most proteins fold into very distinct three-dimensional globular structures. Protein structures solved by X-ray crystallography or NMR are deposited into the protein data bank (PDB), a central repository for all proteins of known structure. At a basic level, we distinguish alpha-helices, beta-strands that fold into sheets, turns,

and loops, as structural elements. These elements are connected to form simple motifs. These structural motifs fold into domains. Polypeptide chains or proteins can consist of several such domains.

A small set of proteins are fibrous proteins that often play structural roles in, for example, keratin in hair. In addition, a significant fraction of proteins do not fold into distinct structures and form so-called *unstructured* or *disordered* proteins. Often, this group of disordered proteins is overlooked when considering the protein repertoire. Machine learning neural network predictors trained to detect regions of sequence that are likely to be disordered predicted that 35 to 51% of eukaryote proteins have at least one disordered region of >50 residues, and that fully disordered proteins correspond to 11% of proteins in Swiss-Prot and 6 to 17% of proteins in various genomes. At this time, most of our protein evolution knowledge is derived from globular proteins – and it is unclear how similar the evolution of unstructured proteins is.

1.4 Domains are the Units of Protein Evolution

We now discuss one of the key concepts for studying the protein repertoire – that of protein *domains*, that is, the structural and functional units that make proteins

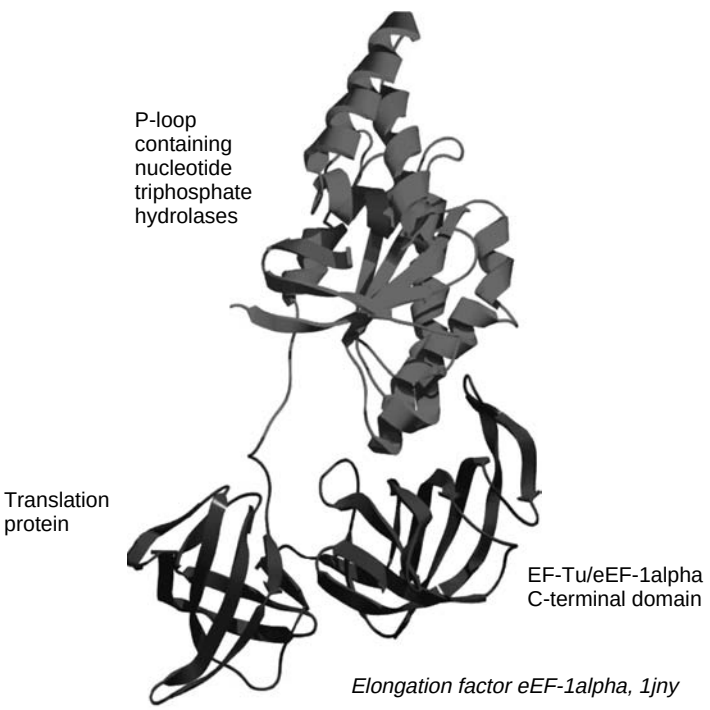

Fig. 2 Domains are the units that make proteins. The eukaryote elongation factor eEF 1-alpha is a beautiful example of a protein that is structured into domains: the P-loop hydrolase domain hydrolases GTP and thus triggers conformational change, the Translation protein domain interacts with the ribosome (not shown) in order to prolong translation.

(Fig. 2). We also discuss the definitions, classification schemes, and properties of domains, and explain how domains allow us to understand protein properties and the evolution of proteins.

The concept of protein domains arose when the first X-ray structures of proteins were determined. These structures, for example, globin or insulin, appeared to be single compact entities. Other protein structures seemed to contain several of these compact entities, and this suggested that proteins are divided into distinct structural units that fold independently, analogous to beads on a string. Typically, domains are regions of compact protein structure with a hydrophobic core. Levitt and Chothia were the first to classify the arrangements of secondary structure elements in a set of proteins available at that time. These representations led to the identification of four distinct protein classes: all-alpha, all-beta, alpha + beta and alpha/beta – these remain the main classes used today.

In 1974, Rossmann et al. realized that domains represented *evolutionary* rather than only structural units. They observed that several nucleotide-binding domains displayed structural similarity, and they speculated that these domains had descended from a domain present in an ancestral protein. Importantly, they also recognized the value of structural information for these analyses: only "a combination of structure and sequence permits a measurement of more distant evolutionary relationships". Later, Murzin et al. constructed the structural classification of proteins database (SCOP), which uses such an evolutionary domain definition (Fig. 3). In SCOP, a structural unit within a protein only qualifies as a domain if it is observed in isolation as a single-domain protein and/or in different combinations with other structural units. Evidence for the evolutionary relationship between domains is taken from detailed examination of protein structure, and, additionally, sequence and function.

Alternative views on domains include the idea that fold space consists of repeated small structural motifs, such as helix-turn-helix motifs, smaller than the compact entities considered domains in other structural approaches. Certainly, folds such as TIM-barrels and other barrel type structures consist of the same small structural motif repeated several times. It is easy to imagine how duplication and assembly of the smaller motifs might lead to the evolution of these larger protein domains. Indeed, some protein fragment matching and assembly-based approaches have been very successful recently when applied to the *ab initio* protein fold prediction problem.

Protein domains are of great diversity, and today there are many databases using their own domain definitions and different protein classifications. The definition of a protein domain remains complicated and can be approached from a structural, sequential, or functional perspective. For answering evolutionary questions, however, it is crucial to define a domain in an evolutionary context – and this is why we use the term *domain* always with reference to the evolutionary unit defined in SCOP.

1.4.1 Evolutionary Hierarchical Domain Classification (SCOP)

SCOP is a completely manual classification for proteins of known structures and is considered by many as the gold-standard classification. The hierarchical classification starts at the level of individual domains or small proteins (Fig. 3). A SCOP domain typically has 100 to 250 amino acid residues, though smaller and

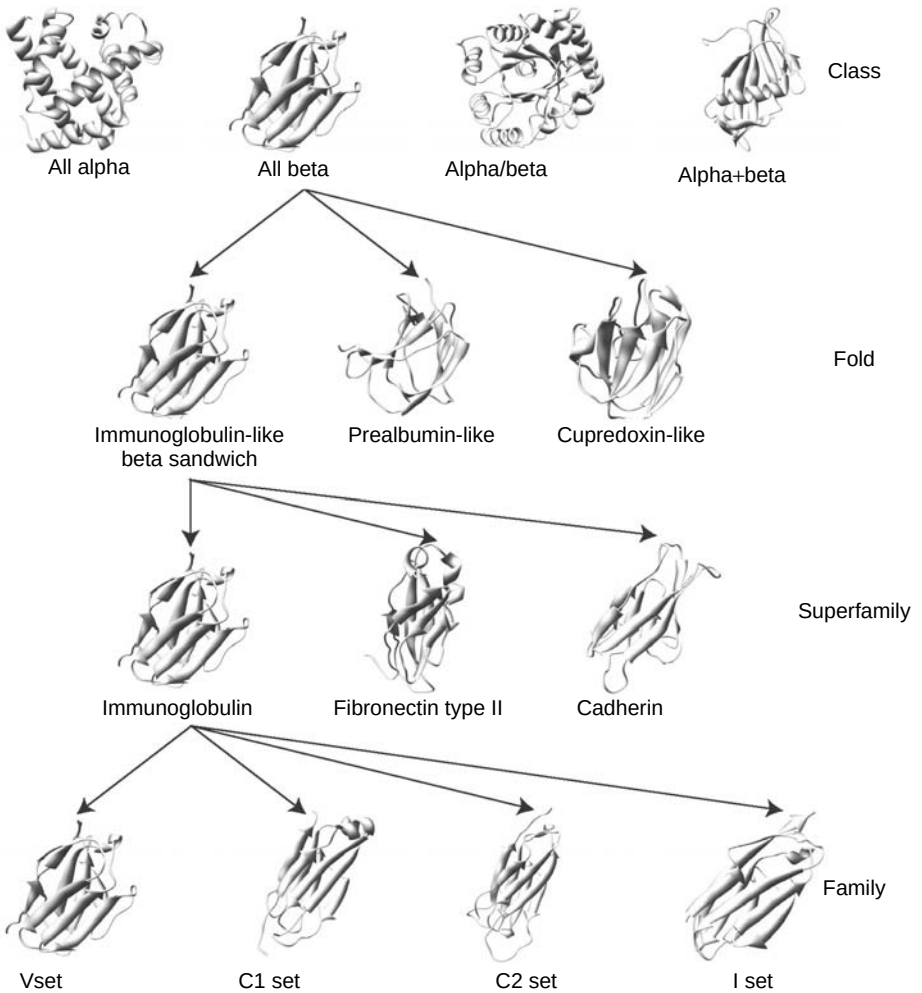

Fig. 3 The SCOP classification. The hierarchical classification is illustrated by the example of the Immunoglobulin domains. One representative structure is given at every level – note that not all folds are shown for this class, and not all superfamilies for this specific fold.

larger domains do occur. Closely related domains are then grouped into families on the basis of having a common evolutionary ancestor evidenced by structural similarity and more than 30% sequence identity. Superfamilies at the next level are made up of families for which the structure and function suggest a common evolutionary relationship, but which have sequence identities below 30%.

The subsequent level of classification is the fold. Superfamilies and families of domains of the same fold have equivalent major secondary structure elements in a similar arrangement with the same topological connections. Superfamilies within the same fold have an unresolved relationship and may be convergently or divergently related – but for which we are currently lacking information either way.

The final level of SCOP is class, which comprises the original all-alpha, all-beta, alpha/beta (a/b), alpha + beta (a + b), and seven additional, smaller classes.

Note that in the literature, the expressions *fold*, *family*, and *superfamily* are often used with different and overlapping meanings and the hierarchical levels are then blurred. For example, fold is sometimes used to describe superfamilies or even single structures; or protein families are not grouped into superfamilies. There are many potential additional relationships between proteins for which we lack evidence as yet, for example, between superfamilies of the same fold. As more protein structures and sequences become available and our methods advance, some of these relationships will be clarified and classifications updated.

1.4.2 Alternative Domain Classifications

In addition to the evolutionary domain definition used in SCOP, there are many other systems of domain classification and they use a variety of definitions and methods. Some databases center their classification completely on structure, others on sequence similarity; the classification can be done manually, completely automatically, or using a combination of both approaches. Some so-called *metasites*, for example, Interpro, combine the output of several different databases.

An alternative and also commonly used protein structural database is the classification architecture topology and homology (CATH) database. CATH uses structural criteria and a domain definition similar to the one used in SCOP. It classifies domains into homologous superfamilies, topologies, architectures, and classes using a combination of manual and automated methods. The main automated approach to domain definition is to identify segments of protein structure with more internal contacts than contacts to other regions of the structure. Additionally, there are many other structural domain classifications, for example, the families of structurally similar proteins (FSSP) or the molecular modeling database (MMDB), both of which are fully automated.

In both SCOP and CATH, the top levels of the hierarchy are defined by the three-dimensional structure, whereas lower levels are identified by sequence similarity. There is an extensive agreement of the classification of the proteins in both databases, although some exact criteria on topological similarity have yet to be determined. There are, however, some striking differences between the domain definitions of these two resources. One such example is that of protein kinases, which are classified as one single domain in SCOP, but is split into two domains in CATH (Fig. 4). In favor of the CATH definition, there are two distinct lobes within the kinase domain. These two lobes, however, are only ever found together and bind ATP at their interface, and this strongly supports the SCOP single-domain definition of kinases as an individual evolutionary domain. Such classifications are to some extent subjective – it is important to remember that alternatives exist and should be used according to the precise context and what question is being looked into.

Some databases rely on sequence rather than structural similarity for clustering domains into families. Examples of these databases are Pfam and SMART, both of which use HMMs as a basis for the alignments they provide. Other databases rely completely on the conservation of sequence stretches in multiple alignments, such as the ones used for TRIBE-MCL or ProtoMap.

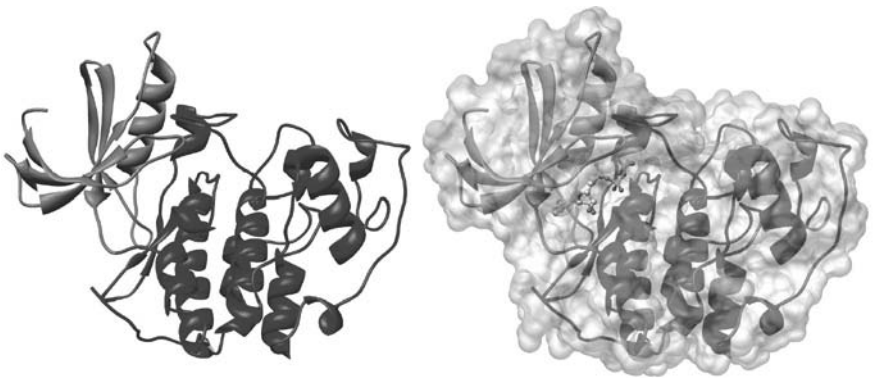

Fig. 4 Protein kinases, as defined in SCOP and CATH. On the left is the kinase domain (PDB ID: 1hck) colored according to the CATH definition of two distinct structural domains (first CATH domain residues 1–84 red, the second CATH domain residues 85–297 blue). On the right, the same coloring is implemented; however, we also show the surface of the entire kinase domain and the location of the bound ATP. This highlights the fact that in terms of surface the entire domain appears as one globular structure, and the location of the ATP binds at the interface of the two CATH defined domains. Functionally, both domains are necessary for ATP to bind and therefore the single-domain SCOP definition seems more accurate in terms of functional, evolutionary domain definition. (See color plate p. xxxiv.)

Sequence-based domain definitions have the advantage that domains of as yet unknown structure and unstructured regions can be included. Homology from sequence similarity, when present, is easier to detect than from structural similarity. We are positive that when significant similarity exists between two sequences, domains are homologous. However, in order to detect all remote relationships, sequence similarity alone is often insufficient and thus it is only of limited use for answering evolutionary questions. In such cases, other conservation of other structural features should be included for consideration. Structure and its finer details, for example, alpha-helix caps, beta-bulges, conserved turns and so on, is a potentially better tool for understanding evolution of the protein repertoire. Currently, such approaches are limited to mostly manual analyses, and we need to develop more automated tools to render their use possible on a genome-wide scale.

1.4.3 Implications and Complications of Domain Definitions

The evolutionary domain definition in SCOP and the knowledge inherent within this classification on the relationships between proteins have important implications. We discuss some of these below.

Firstly, information about the evolutionary relationships between domains is most useful to identify homologous proteins in different genomes. When we find domains of the same superfamily in two different proteins, we know that these proteins, or at least parts of them, are related by descent. The proteins must have arisen by duplication and recombination of the ancestral sequence.

Secondly, the exact definition of a domain in terms of its start and end location within the protein sequence is sometimes complicated. The domains can be joined by long *linker* regions, which sometimes fold into secondary structure elements. Many domains can also have

long insertions in the form of loops, alpha helices, or other elements. In the extreme case, whole domains, called *inserted domains*, can be introduced into the middle of another domain, interrupting the sequence contiguity. Thus, despite all efforts to consistently classify protein structures, many exceptions or special cases have to be considered.

Finally, it remains a matter of debate whether individual folds (structural topologies) arose only once during evolution or whether some of them arose by *convergence* from unrelated ancestors. The basic physicochemical interactions and associated structural and sequence motifs are often conserved across different superfamilies within one fold. In contrast, the same activity and/or function can be performed by two or more completely unrelated structures. Anecdotal cases of *convergent* evolution have been reported in the literature where there is currently no evidence of the proteins being derived from a common ancestor. A well-known example are serine proteases: chymotrypsin and subtilisin are members of two different superfamilies, but the three-dimensional arrangement of the catalytic residues and the reaction mechanism are the same. Another example is MutS and topoisomerise II, which share a structural motif that in its dimeric form is suitable for sensing DNA topology and binding recombination intermediates. Such examples might prove wrong once our ability to detect remote homology improves. When we do not have sufficient evidence of different origins of two protein folds, the relationship is really unresolved in terms of evolution rather than a definite case of convergent evolution. This general lack of proof for convergent evolution led to the suggested use of the term *parallel* rather than convergent evolution.

There is, however, very little doubt that domains with clear structure, sequence, and/or functional evidence of homology have one common ancestor, even in the absence of detectable sequence identity. While classification systems such as CATH contain some information on domain homology, to the best of our knowledge, SCOP is the only one, which defines domains from a more hierarchical evolutionary perspective and incorporates several homology levels. Thus, the evolutionary domain definition in SCOP provides the basis for many exciting analyses into the evolution of the protein repertoire. We can examine the changes in structure across homologous proteins, and how this is linked to changes in sequence and function. We can also compare orthologous proteins in different organisms and try and understand how they may contribute to organism-specific features. Finally, knowledge of the relationships between proteins can help method development, such as to establish evaluation procedures for homology recognition algorithms.

1.4.4 Domain Prediction

In order to study the protein repertoire, for example, the number, distribution, or sizes of domain families in different organisms, we need as comprehensive as possible knowledge on the presence or absence of particular domains in those organisms. While we currently know the structure of more than 25 000 proteins, this covers only a small fraction of all proteins in a particular organism. We can increase this coverage by predicting domains and their structures. This is done by inference, that is, exploiting homology with proteins of known structure, or *ab initio*, that is, prediction of the structure of proteins with new folds.

Many classification systems mentioned above include algorithms to predict domains based on sequence and/or structural homology. Today, we have the complete genome sequence for more than 130 organisms, and, for example, with the use of the SUPERFAMILY database, it is possible to make domain predictions for 40% of eukaryote sequences by inferring homology from sequence similarity – and these are the data that we discuss here. Despite some apparent biases in the proteins of known structure, for example, more globular and soluble proteins, and overrepresentation of some proteins of high biological interest, for example, immunoglobulins and globins, we assume that the conclusions drawn from this partial information are still of general relevance and depict a more refined picture of the protein repertoires than that which was previously known – and this picture will continue to be polished as more data becomes available.

1.5 How Many Different Families, Superfamilies, and Folds Exist in Nature?

The first important question that may arise, after consideration of appropriate domain definition and classification, is on the size of the protein universe, that is, how many different families actually exist in nature and how many novel families are yet to be discovered.

There are more than 1200 domain superfamilies in the current version of the SCOP database, and it is likely that these cover the most common superfamilies. Estimates on the total number of folds or superfamilies, however, vary – partly due to the blurred definitions of fold and superfamily. Early speculation about the total number of protein families and superfamilies relied largely on the rate of discovery of new families, and they predicted around 1000 common superfamilies. When other assumptions or methods were used, the estimates of the total number of existing families, superfamilies, or folds vary from a few hundred to several thousand.

In some respect, the exact number of domain superfamilies might be of little importance, since the distribution of the number of proteins per superfamily or the number of superfamilies per fold is extremely uneven: a few folds are very popular and have many superfamilies. These superfamilies, in turn, also occur in many different proteins (Fig. 5). These folds are also called *superfolds*. Superfolds often have very generic functions and contain structures such as the P-loop NTP hydrolases, Rossmann folds or TIM barrels. In contrast, the majority of novel folds that have been discovered recently tend to only have a small number of superfamily members. They are also known as*unifolds*; it remains unclear how many of the unifolds encompass the still unknown proportion of the protein repertoire.

1.6 The Distribution of Protein and Domain Families Across Different Genomes

Once we have an idea of the number of different domain families that exist in nature, we might like to know how they are distributed across different species. We obtain a first and simple measure of similarity between organisms, how they are related and evolved, and how the organisms' complexity is determined by, for example, the number of different domain superfamilies that occur, their kind, and size.

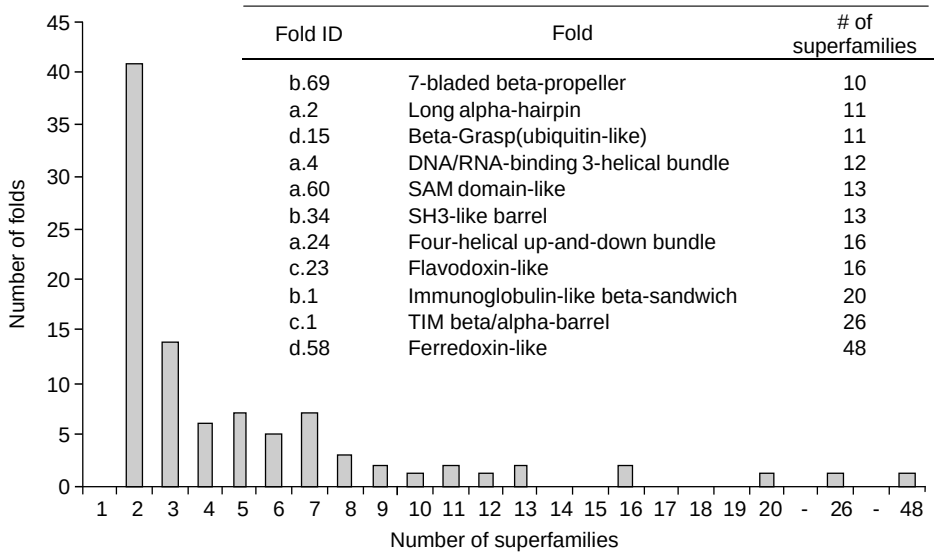

Fold ID	Fold	# of superfamilies
b.69	7-bladed beta-propeller	10
a.2	Long alpha-hairpin	11
d.15	Beta-Grasp(ubiquitin-like)	11
a.4	DNA/RNA-binding 3-helical bundle	12
a.60	SAM domain-like	13
b.34	SH3-like barrel	13
a.24	Four-helical up-and-down bundle	16
c.23	Flavodoxin-like	16
b.1	Immunoglobulin-like beta-sandwich	20
c.1	TIM beta/alpha-barrel	26
d.58	Ferredoxin-like	48

Fig. 5 Distribution of the number of superfamilies within folds in SCOP 1.65. The inset table provides details of the super- or hyper-folds with 10+ superfamilies. There are 602 folds, which have only one superfamily – this is not shown as a bar on the histogram.

Early studies provided a census of genomes in terms of structures, highlighting differences in preferences for secondary structures and folds. Some studies suggest that globular domains occurred early in evolution, with the alpha/beta class being possibly the first.

Later, these analyses were tailored to the distribution of domain families and superfamilies; and again, any conclusions we draw are from those 40% of the sequences to which one or more domains are assigned. For example, in most metazoans, we find about 800 different domain superfamilies out of a total of about 1200 known superfamilies, whereas the unicellular baker's yeast and bacteria have about 650 or even fewer (Fig. 6). Some superfamilies are conserved across different genomes; and the more closely related are two genomes, the larger the fraction of conserved superfamilies. For example, around 400 superfamilies or 35% occur in 14 eukaryote organisms including yeast, invertebrates, and vertebrates. The members of these superfamilies form up to 90% of the domains in these organisms, which makes the protein repertoires between these organisms very similar in terms of domain composition. In contrast, only about 20% of the domain superfamilies or even less occur in all three kingdoms of life. These 20% of all superfamilies comprise, however, around 80% of all domains.

This means that most of the superfamilies evolved long before the split of eukaryotic clades. Sequences encoding domains of these superfamilies have subsequently been duplicated many times and are modified and reused in many different contexts; and we discuss these mechanisms – duplication, modification, and reuse – later in more detail.

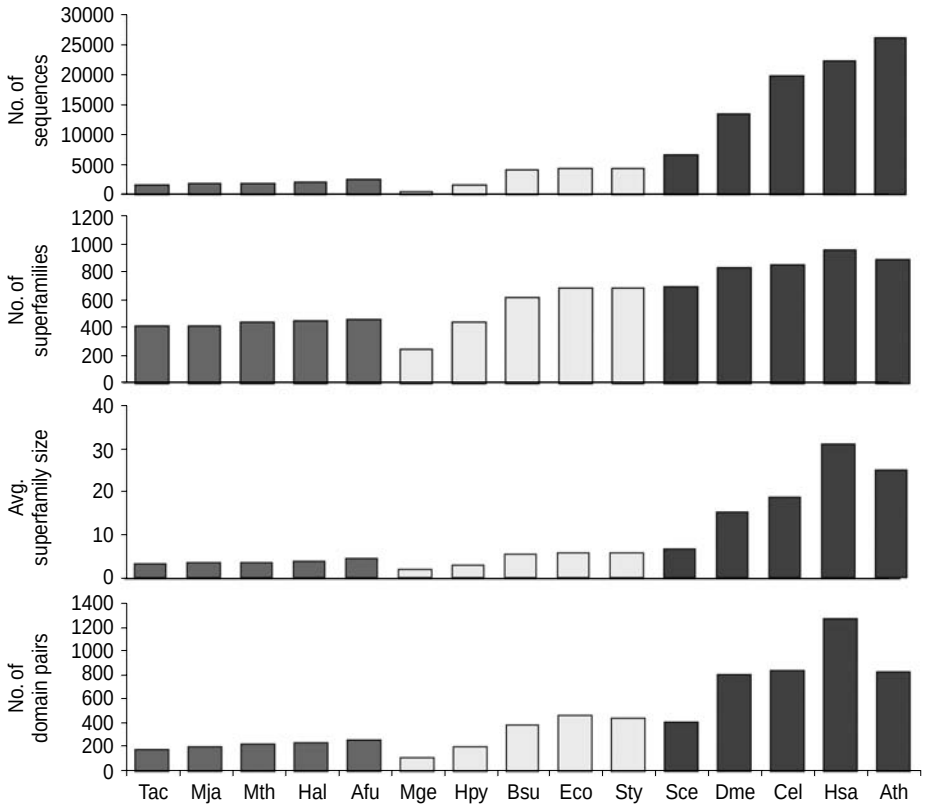

Fig. 6 Different features of the genome and protein repertoire. From top: The total number of sequences (genes) predicted for each organism. The total number of different SCOP domain superfamilies predicted with SUPERFAMILY version 1.65 (http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/). The average number of proteins per domain superfamily that is, number of duplicates per domain superfamily. Finally, the number of different two-domain combinations. Red – Archaea: Tac, *Thermoplasma acidophilum*; Mja, *Methanocaldococcus jannaschii*; Mth, *Methanothermobacter thermautotrophicus* Delta H; Hal, *Halobacterium* sp. NRC-1; Afu, *Archaeoglobus fulgidus* DSM 4304. Yellow – Bacteria: Mge, *Mycoplasma genitalium*; Hpy, *Helicobacter pylori* 26695; Bsu, *Bacillus subtilis* ssp. subtilis 168; Eco, *Escherichia coli* K12; Sty, *Salmonella typhimurium* LT2. Blue – Eukarya: Sce, *Saccharomyces cerevisiae*; Dme, *Drosophila melanogaster* 3.2; Cel, *Caenorhabditis elegans* WS123; Hsa, *Homo sapiens* 22.34d; Ath, *Arabidopsis thaliana* 5. (See color plate p. xxxv.)

1.7 Proteins Interact

Proteins interact – they are part of complexes, cellular structures, and pathways. Many *stable* protein complexes are crucial for the cell to work: the transcription factor assembly, the splicing machinery. *Transient* interactions of proteins or protein domains have important roles, for example, in signaling cascades or cell cycle regulation. These interactions are part of protein function and hence constrain the evolution of sequence. Indeed, proteins

in stable complexes are more conserved than proteins in transient complexes, and, generally, the more interaction partners a protein has, the slower it evolves. Proteins with many interaction partners tend to have well-conserved sequences and frequently have essential functions. Domains within proteins also interact, and we discuss the details of their spatial arrangement in a later section.

Recent years have produced an enormous growth of genome-scale protein interaction data and revealed that, similar to other biological phenomena, protein interaction networks are of a scale-free nature. The hubs in these networks, i.e., proteins with many interaction partners, are either part of stable complexes or are involved in crucial transient interactions mentioned above. With a map of protein interactions, we can better understand the evolution of the protein repertoire with reference to the constraints placed by interactions and the resulting functions.

1.8 Proteins are Part of Pathways

Many proteins, for example enzymes or signaling molecules, are part of pathways, such as in metabolism or signal transduction. In the *Escherichia coli* small-molecular metabolism, about 600 proteins act in about 100 pathways, and one-quarter of the enzymes are involved in multiple pathways. Even this basic set of ancient enzymes seems to have evolved by means of extensive duplication and recombination.

Whether the function of a single protein, or of the entire pathway, is essential decides whether the organism will survive in case of a mutation and thus influences the rate of evolution of the protein. Again, we see many biases: most domain superfamilies occur in only a few pathways, whereas a few domain superfamilies of very generic function occur in many pathways.

The "Horowitz" model assumes that duplicated enzymes tend to be employed in the same pathway so that enzymes in consecutive reaction steps are often homologous. However, despite extensive domain and protein duplication across metabolic enzymes, this is only partially true. For most of the homologous proteins, the catalytic mechanism or cofactor-binding properties are conserved or only slightly modified while the substrate specificity has changed. This suggests that it is much easier to evolve new binding sites than new catalytic mechanisms. Thus, the recruitment of homologous domains in novel pathways primarily occurred on the basis of catalytic mechanism or cofactor binding, and this speaks against the Horowitz model.

The network of biological pathways is amazingly robust. More than half of all genes in yeast have, when knocked out, no, or hardly any, apparent effect on the fitness of the organism, that is, they are nonessential. Many of these genes might, however, be of important function under nonhomeostatic conditions such as when subject to stress. The apparent robustness is not only due to the functional redundancy created by the many duplicate genes but also by alternative routes to producing the same molecules.

The same pathway in different organisms can contain species-specific sets of isozymes. Enzymes responsible for the same reaction in the same pathway but in different organisms can also be completely unrelated. This phenomenon is called *nonorthologous* replacement, and one example is lysyl-tRNA-synthetases, which occurs in two unrelated protein families across bacteria and archaea.

The recruitment of single enzymes from different pathways could be the driving force for pathway evolution, and some enzyme families are capable of changing both reaction and substrate chemistry. Such recruitment seems more prevalent than, for example, whole pathway duplication, enzyme specialization, *de novo* invention of pathways, or retroevolution of pathways. Together, these variations produce "mosaic" network of pathways in different organisms with a widespread plasticity.

1.9 Protein and Domain Function

The function of a protein is usually determined by the interplay of its sequence and three-dimensional structure. However, two proteins of very similar functions may have arisen from different ancestor molecules and/or can have different structures. A well-documented example of functional convergence is the Ser/His/Asp catalytic triad of the serine proteases found in several different protein folds. The converse is also true: proteins that are structurally or evolutionarily related do not necessarily have identical functions such as is seen for the large TIM-barrel fold.

Similar to structure classification, the identification and useful classification of protein function is nontrivial. Several classification schemes for protein function have been developed in recent years, and it is important to choose the scheme according to the question that is to be addressed. Each scheme has its own advantages and disadvantages. One of the oldest systems is that of EC-numbers, classifying chemical reactions of enzymes.

A more recent classification, that comprises both enzymes and nonenzyme proteins, is gene ontology (GO) which classifies protein subcellular localization, molecular function, or the biological process of proteins in three separate ontologies. The distinction between the molecular function of a protein and the biological process in which it acts is very important, and, to the best of our knowledge, GO is the only classification system thus far that makes it. For example, a kinase's molecular function is to transfer phosphate groups. Kinases can, however, also act in processes as different as signal transduction, glycolysis, or muscle movement. A disadvantage or difficulty in GO is its nonhierarchical character and that the depth of description is inconsistent at various levels of the graph.

1.9.1 Indirect Function Characteristics

The function of a protein is tightly coupled to its sequence and structure, but also to many other characteristics, such as its interaction partners, its localization in the cell, position in the genome, and so on. These characteristics co-evolve and, thus, they can be used to predict protein function and to study function evolution.

Firstly, function prediction often relies on sequence similarity, assuming that homologous proteins have related functions. The predictions are reliable if the sequence identity to a homolog of known function is greater than about 40%. If the sequence identity is low, the three-dimensional structure of the protein might still give clues as to the function, but the situation is much less clear.

Secondly, proteins of similar function tend to co-localize on the genome. In prokaryotes, genes that participate in the same process tend to be organized in an operon; but a nonrandom gene order is also observed in eukaryotes. Housekeeping genes, for example, tend to form clusters on the chromosome.

Thirdly, proteins of the same pathway tend to interact. Fourth, a functional link between two proteins is assumed to encourage the physical fusion of the two genes, and this approach has been termed *Rosetta Stone principle*. In the fifth place, two functionally linked proteins also tend to show the same phylogenetic profile. That means two genes that show the same pattern of presence or absence in a number of organisms tend to be functionally linked. Penultimately, functionally related proteins are often co-regulated in the same subcellular compartment or have similar effects on the knockout phenotype.

Last but not least, the existence of the same domains in two proteins, for example, two enzymes from different organisms, and their interactions with other domains can also tell us whether these enzymes carry out a similar function. Thus, the domain composition can be used to predict protein function or localization.

1.9.2 Functional Annotation of Domains

The function annotation of protein domains is even more difficult. GeneOntology annotation is available for some Pfam domains, but only a few hundred of them have equivalents to SCOP domain superfamilies. It is, however, difficult to use a whole-gene-based annotation scheme, such as GO, for domains. Most proteins consist of several domains, which can have different functions. For example, a kinase domain can be found with a lipid-binding and/or protein interaction domain. Such considerations made it necessary to develop a novel domain-based annotation scheme, and Bashton and Chothia have done so. Their scheme is domain-centric and emphasizes domain function in the context of domain neighbors in multidomain proteins, providing functional annotations for a subset of SCOP domains. The annotation is based on detailed examination of the protein structures. In this domain-centric functional classification scheme, domains are classified into seven categories that encompass catalytic activity, cofactor-binding, responsibility for subcellular localization, protein–protein interaction, and so forth.

2 Mechanisms in the Evolution of Single Proteins and the Protein Repertoire

In the previous sections, we described the characteristics of single proteins and of the protein repertoire as it is encoded in an organism's genome. We will now focus on the processes that form the protein repertoire during evolution: firstly, whole sequences or parts thereof are duplicated and therefore the number of copies of a protein or domain is increased in a genome. This can amplify the function of the protein, but also provides material for a second mechanism: sequences, especially of duplicated proteins or domains, diverge by mutations, deletions, and insertions that produce modified structures and functions. Thirdly, whole genes or single exons recombine so that novel protein domain arrangements emerge.

2.1 Orthology and Paralogy

Before discussing the details of these three mechanisms, it is important to introduce an important notion on homology: the distinction between paralogy and orthology.

Paralogs are homologous proteins that arose via gene duplication within one organism (Fig. 7); they are also called *duplicate* genes. The redundancy created by gene duplication relaxes selection

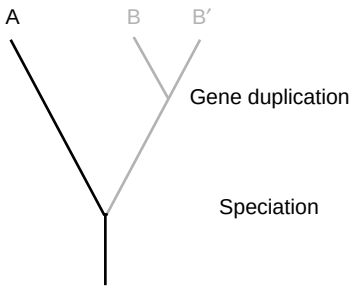

Fig. 7 Orthology and Paralogy. The tree represents the orthologous and paralogous relationships of a gene, protein, or domain. Speciation results in one copy of the gene in each species: A and B are orthologs (gray and black branch, respectively). A gene duplication event in one species (gray branch) results in B′, which is a paralog to B.

pressure and encourages the divergence of protein function. In many cases, however, paralogs still have identical or similar biochemical activities.

Orthologs are protein homologs that arose during a speciation and not a gene duplication event. Thus, orthologous proteins are still under selective pressure and will often have identical functions in different organisms. While we generally observe a core proteome of 1 : 1 orthologs across eukaryote organisms, orthologous genes can, of course, also be duplicated. If the duplication occurs *after* the speciation event, the duplicates are called *in-paralogs* or *co-orthologs*, since the orthology is ''shared'' between the duplicates. If the duplication occurs *before* the speciation event, the duplicates are called *out-paralogs*. A very useful database collects clusters of orthologous groups of proteins (COGs). Even though these groups of proteins merge paralogs and orthologs, proteins within these clusters often have similar functions.

Finally, ''horizontal'' gene transfer, as described below, introduces *xenologs* into an organism, for example, proteins that are homologous to proteins in another organism without these two organisms being directly related by descent. In addition, proteins that substitute for function in alternative pathways or in nonorthologous displacement, discussed above, could be considered *analogs*, that is, proteins of functional or structural similarity without homology.

2.2 Duplication – Mechanisms to Create "More"

Duplication of existing building blocks drives evolution and creates paralogs as mentioned above. Haldane (1932) and Muller (1935) were the first to note the importance of duplication, later followed by more exact descriptions and estimates. The genomes of organisms of all three kingdoms of life seem to be formed by many gene duplication events, accompanied by extensive loss of genes. This is also clear at the level of domains: at least 58% of the domains in *Mycoplasma* and 98% of the domains in humans are duplicates. Nice examples of evolution by gene duplication are the HisF and HisA beta/alpha barrels, which have arisen through duplication and fusion of an ancient half-barrel.

Several different mechanisms can be responsible for duplication of existing available coding sequences. First, whole genome duplication takes place as a consequence of nondisjunction during meiosis within species or hybrids. Whole genome duplication has presumably occurred several times at least during the evolution

of eukaryotes: once in yeast, later another one in vertebrates. The model plant *Arabidopsis thaliana* shows evidence of at least two rounds.

Second, whole chromosomes or parts of chromosomes can be duplicated. There are well-known human genetic disorders, which are the result of such chromosomal duplications. Perhaps the best known of these is trisomy 21 or Down syndrome in which affected people have a third copy of chromosome 21. These types of chromosomal aberrations are called *aneuploidy*. The majority of chromosome duplications or deletions are deleterious since the duplication of a large number genes causes severe dosage imbalance in the levels of the gene products.

Third, single genes can be duplicated via tandem duplication, retroposition, or transposition as the result of ectopic recombination. Single gene duplication is assumed to be a major mechanism for the creation of recent duplicates. Tandem duplications are the result of unequal crossing-overs during meiosis. Unequal crossing-over between misaligned sequences generates a chromosome with a duplicated region and a chromosome with a deleted region.

Finally, part of a gene or one exon can be duplicated in unequal crossing-overs or replication slippage during DNA replication. Repetitive sequences can increase the chances of unequal crossing-over. Replication slippage has been found to happen for domains, resulting in long stretches of repeated domains in some proteins.

2.2.1 Horizontal Gene Transfer and Transposons

In addition to gene duplication, horizontal gene transfer (HGT) plays an important role in the evolution of prokaryote genomes. HGT is the acquisition of new genes via the movement of genetic material between species, other than by descent. Many eukaryote genes are found in bacteria, interestingly, however, not so in archaea. In eukaryotes, the rates of HGT seem much lower than in prokaryotes. The presence of many genes that arrived through horizontal transfer can seriously distort our picture of the distribution of protein and domain families and impair the reconstruction of the correct evolutionary pathway. In fact, the extensive HGT in prokaryotes led to the depiction of their phylogeny being a "mesh" or network, rather than a tree without a single common bacterial ancestor.

Finally, eukaryote genomes appear to be frequently shaped by jumping genes or *transposons* that move around in the genome, copying and pasting themselves. Estimates suggest that at least half of the human DNA consists of transposons, making a total of around 4 million all together. By their influence on protein structure, for example, via insertion of additional domains, they are of some influence on the shape of the protein repertoire.

2.2.2 Duplication Leads to Protein and Domain Family Expansions

While the earliest evolution of the protein repertoire must have involved the *ab initio* invention of proteins and domains, later on, this may have been replaced by the easier and more efficient reuse of existing building blocks. Thus, it is unsurprising that even the simplest bacterial genomes are the product of extensive gene duplication and recombination.

Lineage-specific expansions of protein families result from relatively recent duplications and are prominent in genomes. Across bacteria, protein, and domain families that function as transcription factors

or in signal transduction increase exponentially in number with the size of the bacterial genome. In eukaryotes, we observe the expansion of protein or domain families with very specific functions, and this can give clues on why the organisms have different physiological features. For example, the largest superfamilies in humans are mostly involved in or part of cell surface receptors, protein–protein, or cell–cell interaction, signaling, or cytoskeleton structure, and this explains why we can have so many different cell types forming intricate tissues and organs.

Proteins that function as regulators or cell-adhesion molecules are associated with the advent of multicellularity, and thus the expansion of protein or domain families that have these functions give clues as to how the multicellular organisms evolved. For example, Immunoglobulin and Cadherin domains expanded in the fruit fly as compared to the nematode worm. The functions of these domains in cell-adhesion molecules in the embryonic development of the nervous system can explain why the fly has a much more intricate nervous system than the worm.

The observed extent of domain duplication varies – across different superfamilies and across genomes. In vertebrates, such as human and mouse, each domain superfamily occurs on average in 30 different proteins, while in invertebrates, they occur in only 20 proteins. The numbers are even smaller for bacteria. Generally the duplication results in a distribution of family sizes in genomes that follow a power law, with remarkably constant parameters across organisms. This means that there are only a few highly abundant superfamilies, for example, the P-loop NTP hydrolases, NAD(P)-binding Rossmann domains and certain kinase families. This relationship was first recognized for gene family size, later for domain families or even pseudogene families, and its general implications are discussed in a later section.

2.2.3 Reconstructing the Evolutionary History of Proteins

A classic way to study evolution of a protein repertoire is to reconstruct phylogenetic trees from sequence alignments using one of several methods. One of the most reliable but also the most computationally expensive are maximum likelihood approaches, followed by parsimony, and then distance methods. From the reconstructed trees, it is possible to follow the evolutionary pathway of a protein. This permits an understanding of one protein family – more recent approaches have involved constructing trees for all or most of an organism's proteins in order to obtain an overview of the evolutionary pathway of as many components of the organism as possible. These kinds of analyses can be performed on individual domains, entire sequences, or on concatenated alignments of multiple proteins or domains. Unfortunately, xenologs, that is, homologs in organisms that are not directly related, create difficulties. Therefore, it is advisable to consult several phylogenetic trees since their consensus may reflect the organisms' true evolutionary relationships best.

By reconciling a tree with a species tree, it is then possible to estimate where the duplication and speciation events took place and this differentiates orthologs from paralogs (Fig. 7). This method is vital in phylogenomic-type approaches, which attempt to elucidate function from a proteins phylogenetic location within a tree with respect to homologous proteins of known function.

2.3 Divergence after Duplication

In the absence of duplication, any mutation that occurs is under selective pressure to maintain the protein's structure and function. Once a domain or protein has duplicated, this pressure is relaxed and one copy of the domain can evolve a new or modified function. This can occur by sequence divergence or by combining with other domains to form a multidomain protein with a new series of domains. The positive, negative, or neutral effect on the function determines if the duplicate is retained. Apart from changes in sequence, structure, and function, which we discuss below, duplication also relaxes selection pressure on other characteristics, for example, gene regulation or expression patterns.

Many eukaryotes are diploid, which means they have two copies of each chromosome. Thus, there naturally exists a duplicate of each gene. The diploid nature of many organisms delivers potential for divergence and modification of protein function even without any additional gene duplication. This is exemplified in many well-known cases of dominant and recessive mutations that are carried through populations. One example is the recessive mutation that is the cause of sickle cell anemia, which we discussed earlier.

2.3.1 Changes in Sequence

The first level of changes that occur are at the nucleotide level. Mutations can affect a single nucleotide (point mutations) or several adjacent nucleotides (segmental mutations). Depending on the type of change caused by the mutational event, there can be substitutions, recombinations, deletions, insertions, or inversions. Substitution mutations can be transitions, where one purine is replaced by a different purine or transversions, where a purine is replaced by a pyrimidine or *vice versa*. If these changes happen in protein-coding regions, they can affect the amino acid sequence in two different ways:

Many single nucleotide changes are called *synonymous* substitutions, as they do not alter the resulting amino acid due to the redundancy of the genetic code. Synonymous substitutions are not as much affected by selection, as the protein sequence and structure remains the same, and they display similar rates across different protein families. They can therefore serve as a measure of the overall substitution rate and the absolute evolutionary distance between two sequences.

In contrast, *nonsynonymous substitutions* change the resulting amino acid and are likely to affect the protein and fall under selection. As a consequence of different selection pressures on different protein families, the nonsynonymous substitution rates vary enormously amongst protein families.

We usually distinguish three different effects of sequence changes: deleterious (or negative), neutral, and advantageous (or positive) mutations. The neutral effect of, mostly, synonymous substitutions has a long history in research. One common way to analyze for advantageous mutations is to compare the frequency of nonsynonymous and synonymous substitutions. Examples of advantageous mutations seem relatively rare, and are observed in resistance or pathogen defense genes, for example, in MHC proteins. It is very difficult to estimate the absolute or relative mutation rates, as many different factors affect them: the level of expression of the gene, the function of the resulting protein, the

essentiality of the gene, or the relative amino acid cost.

2.3.2 Changes in Function

Once a gene is duplicated within a genome, one of the two copies is free to evolve, while the first can maintain the original function. This evolution may produce modified or new molecular functions that act within the same biological process, or it could create functionality within an unrelated biological process. We distinguish three major effects on protein function after a duplication event. First, the redundant sequence can have a highly negative effect on the fitness of the organism and is consequently silenced. The whole gene is lost or turned into a "dead," untranscribed *pseudogene*. This is a very common process. Second, the redundant protein can evolve a new function that is different from the original function, for example, by sequence divergence or combination with a novel domain partner, and this process is called *neo-functionalization*. Last, if the original gene has multiple functions, they may become split between the two copies, a process called *subfunctionalization*. In eukaryotes, neofunctionalization is much more common than subfunctionalization.

The evolution of domain function is even more complicated, since a domain can either have an independent function, and/or contribute to the function of a multidomain protein in cooperation with other domains. In order to assess domain function evolution, knowledge of the three-dimensional structure is vital. Todd et al. found that for most domains within a superfamily the reaction chemistry is fully or at least largely conserved. The substrate specificity, rather than reaction chemistry, is modulated; this often occurs via changes in the kind of domains adjacent to the catalytic domains, embellishments in the active site, or by variations in the oligomeric state of relatives. Catalytic residues can "migrate" with respect to their position in the amino acid sequence, but their relative position in the three-dimensional structure is mostly preserved. In sum, domain function evolution occurs in two major ways: first, a domain can perform the same function, but in different protein contexts, that is, with different partner domains. Second, a domain may diverge and acquire a novel or modified function.

The way in which an individual protein or protein domain will evolve is very much dependent on the interactions it maintains with other molecules. There is a high level of evolutionary constraint on enzyme binding sites that have chemically specific ligands and restricted binding orientation so that mechanistically the function can still be maintained. Similarly, regulator proteins that bind to specific DNA or RNA sequences also need to maintain a very specific binding site. Conversely, a protein that binds a receptor protein is freer to evolve as the receptor may co-evolve at the same time. Therefore, proteins involved in such interactions can evolve more quickly than proteins that must bind and localize a specific small molecule.

This is exemplified by the fast-evolving four-helical cytokines, specifically human and mouse Interleukin-5, the sequences of which are 70% identical, but they will no longer cross-bind to one another's receptors. Such fast-evolving proteins often seem to have functions with the immune system – a facet of living organisms, which needs to evolve quickly in order to adapt to the many new and changing invasive viruses and bacteria present. For immunoglobulins, crucial for antigen recognition in higher eukaryotes, the ability to evolve quickly is indispensable.

Indeed, they have evolved the ability to switch only one of their domains (the variable domain or V-set) by alternative splicing in order to create more combinations, thus increasing the likelihood of antigen recognition.

In sum, the nature of a protein domain's specific interactions, and the number of interactions it is physically a part of, is likely to impact greatly the rate of evolution for that particular protein or at least the domains within it. The greater the number of interactions and the more constrained structurally the binding partners, the slower the evolution.

2.3.3 Structural Tolerance to Amino Acid Changes – a Case Study on Globins

Implicitly, alterations within the amino acid sequence of a protein will produce changes in the resulting structure. However, the effect on the structure is dependent on where in the protein the change occurs. Traditionally, there have been two approaches to understand tolerance to amino acid changes – theoretical comparison of related sequences and structures from present-day organisms and mutagenesis experiments. In mutagenesis experiments, the roles of specific residues in the folding and stability of a protein are probed by replacing them with alternative residues.

These two approaches convey a similar message on acceptability of mutations in proteins. The hydrophilic exterior of a protein is tolerant to a wide variety of mutations. However, there is a restriction on the choice of mutations in the protein interior; in general, mutations conserve the hydrophobic nature of the core. Thus, the nature and accessibility of the side chains play an important role in the folding and stability of proteins.

Globins represent wonderful examples of how evolutionarily related proteins can be of similar three-dimensional structure and function, but of hardly recognizable sequence similarity. The first two protein structures to be determined, that of sperm whale myoglobin and horse hemoglobin, have only 28% sequence identity. Lesk and Chothia examined the relationship between sequence, structure, and packing in globins in order to understand the effect of sequence divergence on structure and to understand how proteins adapt to mutations.

Although there are approximately 150 residues in globins, the principal determinants of the three-dimensional structure are the 59 residues that are involved in the packing of helices and in the interactions between the helices and the haem group. Only half of these residues are buried. Any mutations of the buried residues retain nonpolar sidechains, although the amino acid shape and size vary. The interface structure between the alpha helices within the globin domains is extensively conserved, except for some shifts and rotations between pairs of helices as a consequence of dense helix packing. If a mutation at the interface changes the volume of the amino acid side chain, the helices move to adjust the packing.

Rare or disallowed changes in globins illustrate the global effect of mutations on the structure and how they are accommodated. If the interface residues switched instead to a completely new set of contacts, the stability of the contact would be lost as the complementarity of the helix surface is compromised beyond allowed limits of variability. For example, a single insertion in the helix interior would turn at least part of the interface by 100°, destroying surface complementarity. Thus, insertions and

deletions are never found within a globin helix. Furthermore, mutations in the individual residues in contact with the haem are difficult to accommodate as the haem has comparatively little flexibility. In order to maintain the geometry of haem packing, changes in helix packings are coupled and do not occur independently.

2.3.4 Changes in Structure

How far can sequences diverge given the requirements of structural stability? In order to understand the relationship between divergence of sequence and structure, Chothia and Lesk compared 25 structures from eight different protein families. They determined the common core of a protein family by individually superposing the main-chain atoms of major secondary structure. The core regions undergo structural change as their sequences diverge and accumulate mutations. The further the sequences diverge and the lower the sequence identify, the small is the core common to the proteins. For those pairs of proteins, whose sequence identity is 50% or more, 90% of the residues of the individual structures are part of their common core. In contrast, for pairs whose residue identity drops to about 20%, their common cores contain commonly half of the residues in the individual structures. The structural differences in the common cores of members of individual protein families consist mainly of changes in the relative position and orientation of packed secondary structures and local changes in case of beta-sheets. A more recent study of the four-helical cytokines found around 5% of sites within the proteins that had any kind of minimal hydrophobicity or volume conservation within the hydrophobic cores of the proteins.

This demonstrates how rapid evolution can take place at the sequence level while still producing the same or a similar protein structure. There is, however, also great scope for alterations to protein structure.

First, secondary structure can be changed and additional regions of secondary structure (decorations) can be added around the core. Frequently, N- and C-terminal extensions are added to structural domains. Occasionally, we observe wonderful snapshots of such structural evolution in progress. For example, recent work by Newlove shows how one part of a protein domain in closely related Cro proteins can switch its secondary structure depending on its oligomeric state (Fig. 8). Other fold changes include divergent relationships between proteins with distinct topological isomers such as the catalytic domain of carboxypeptidase G_2

Fig. 8 Cro protein evolution. The dimeric Cro lambda repressor protein is shown on the left-hand side (PDB ID: 1cop with the different domains shaded differently) and on the right is the evolutionarily related Cro 434 protein (PDB ID: 2cro). Secondary structure switching occurs at the dimer interface of Cro lambda, which forms beta-strands, compared to the C-terminal helices in the equivalent position in Cro 434.

and the aminopeptidase from *Aeromonas proteolytica*, and interchangeable parts, that is, topological differences, but similar secondary structures, for example, triabin and the lipocalins.

Second, oligomeric state also contributes to the evolution of the protein repertoire. Paralogs of proteins can have different oligomeric dependencies for stability and functionality. Alteration of oligomeric state can be a mechanism for enormous functional, as well as structural, divergence as seen in the cupin family. Often, within multimeric protein structures we observe*domain-swapping,* that is, individual monomers become intertwined and even more dependent on one another. Indeed domain-swapping has been suggested as a mechanism of protein evolution toward multimerization. Thus, in multimeric proteins, we see again how the modularity of proteins, that is, by combining different numbers of domains, provides a fantastic resource for a variety of evolutionary processes.

Overall, the variety of evolutionary mechanisms, such as insertion, deletion and substitution of sequences, circular permutations, strand invasions and withdrawals, hairpin flips and swaps, can eventually lead to changes in the protein fold, and Grishin et al. detail many wonderful examples. Circular permutations are alterations of protein connectivity evident through ligation of the termini and cleavage at another site. Strand invasion refers to the insertion of a strand into a beta-sheet such that hydrogen bonds have to be broken in the existing sheet, and reformed with the inserted strand. Strand withdrawal means the removal of a strand. In hairpin flips and swaps, beta-strands are internally flipped, more commonly, this is observed for strands adjacent to one another – this creates or removes crossing loops.

This illustrates how the evolutionary relationship between two proteins, that is, homology, and similarity between their structures are different features and how one does not necessarily determine the other. Rapid evolution can take place at the sequence level and still produce a similar structure and *vice versa*.

2.4 Recombination – Mechanisms to Create "Novel Combinations"

Another way for Nature to diversify the repertoire of proteins is to shuffle their units, the domains, around and create novel combinations. Already 30 years ago, examination of protein structures showed that many proteins are formed of two or more domains; and domains from the same or from different superfamilies combine with each other. Several molecular mechanisms underlie this reshuffling. The major molecular mechanism that leads to multidomain proteins and novel combinations is nonhomologous recombination of domains, sometimes referred to as *domain shuffling*.

There is evidence that in eukaryotes there is a tendency for exon boundaries to coincide with domain boundaries, which suggests that the proteins are formed by intronic recombination.

There is a lot of debate as to whether introns were inserted early on in evolution or late. A detailed study on the conservation of the introns and exons between proteins of known structure revealed that probably some introns were inserted earlier and others later – and, in some cases, their positioning was more conserved within superfamilies than

in others. Most of the intron positions showed little or no conservation supporting the view that intron insertion appeared only after the duplication or speciation. In contrast, some extreme examples, for example, the four-helical cytokines, showed an amazing conservation of intron positions with respect to secondary structure elements, and this implied that they were inserted already in the proteins' common ancestor.

Another important recombinatorial mechanism, especially in eukaryotes, is the fusion of genes leading to so-called *Rosetta-Stone sequences*. Gene fusion seems more common than splitting or fission of genes.

2.4.1 Most Proteins Have More than One Domain

Proteins form complex networks of interactions. Thus, even a modest increase in the number of domains in a protein could result in numerous new interactions. Small proteins generally contain just one domain. Large proteins are more often formed by combinations of domains. The total number of domains found in one protein and their relative orientation determines what we call the protein's *domain architecture*. The majority, or at least two-thirds, of proteins consist of two or more domains in prokaryotes and an even larger fraction in eukaryotes. Higher eukaryote proteins also have, on average, more different types of domains than lower organisms. Last, eukaryote proteins acquire domains additional to those seen in their prokaryote orthologs, and this tendency has been termed *domain accretion*. The increasing complexity of domain architectures in higher eukaryotes (Fig. 6) has been correlated with their increased physiological complexity, since more domains in a protein allow a greater variety of interactions and functions.

2.4.2 The Repertoire of Domain Combinations is Limited and Highly Biased

Even though most proteins consist of several domains, the repertoire of combinations of different domains seems very limited. The proteins from more than one hundred different organisms contain only several thousand different combinations of two-domain superfamilies. This is far fewer than would be possible given the total number of about 1200 known superfamilies or the number of multidomain proteins per proteome. This fraction is likely to decrease even more if membrane proteins are included.

There are a few domain superfamilies that are highly versatile and have neighboring domains from many superfamilies, while most superfamilies have only a few different types of adjacent domains. The distribution of the number of partner superfamilies per superfamily follows a power law, like the distribution of superfamily sizes mentioned above. A most versatile superfamily is, for example, the P-loop hydrolase with more than 80 different combination partners in the human genome. We also observe that particular combinations of domains occur in many different proteins, such as the combination of SH3 and SH2 domains is part of many signal transduction proteins. These recurring domain combinations are *supra-domains*, for example, combinations that are duplicated as one unit and reused in different proteins. Thus, supra-domains represent evolutionary units larger than single domains – and they appear to be a major part of the protein repertoire.

2.4.3 The Sequential Order of Domains and Their Geometry are Conserved

Since most proteins consist of multiple domains, and domains determine the

function and evolutionary relationships of proteins, it is important to understand the principles of both domain combinations and interactions. If the same domain combination is observed in two different proteins, one possibility is that they have assembled independently by different recombinatorial routes. We assume, however, that most instances of the same two-domain combination or domain architecture have evolved from the same ancestor, and there are several lines of evidence that support this.

First, three-dimensional structural analyses of individual protein families such as the Rossmann domains have shown that proteins with the same domain architecture are related by descent, that is, evolved from one common ancestor. This may even be true for most two-domain protein families of known structures in the current databases. Second, with only a small fraction of exceptions, two domains occur in only one N- to C-terminal order in structural assignments to genome sequences. This conservation of domain order is likely to be historical instead of functional, as a very similar interface and functional sites could be formed by two domains in either order, for instance, given a long linker. Finally, proteins sharing a series of domains also tend to have the same function, which is rarely the case if domain order is switched.

In addition to their conserved sequential order in a protein, the spatial arrangement, or domain *geometry*, is remarkably conserved if the two domains are homologous. This has first been shown for the example of Rossmann domains and their catalytic partner domains, but is possibly true for most of the domain combinations found in SCOP. Studies of small numbers of families of protein complexes, where the interdomain geometry occurs across different polypeptide chains, have also revealed extensive conservation of geometry; and there is a tendency for geometry of interaction of protein domains to be the more conserved, the more similar the domain sequences are.

These results suggested that for proteins of unknown structure, the quarternary structure and protein complex geometry can usually be modeled on the basis of homologous polypeptide(s) of known structure; and impressive examples of such protein models of complexes are the yeast ribosome and exosome.

2.5 Selective and Random Processes

As illustrated above, the sequence, structure and function of proteins are tightly linked and place strong constraints on evolution. In contrast, many evolutionary changes seem to occur in a random fashion. Selection then modifies the outcome of these processes: it favors advantageous changes, rejects disadvantageous changes, and tolerates changes of neutral effect. Some changes may be of neutral effect. This is also true at the level of proteins and domains.

On the one hand, we have seen that many domain or protein families have properties useful in a wide range of contexts; it seems obvious that certain specialized biochemical functions are selected more often than others. Examples are the C2H2 and C2HC zinc-finger domains, which are found in a wide variety of proteins associated with DNA-binding, such as transcription factors. From a structural point of view, there is a tight interplay between the folding kinetics, the stability, and the function of a protein, and the so-called *designability* of a protein, domain, or domain combination could also be an

important selection criterion. According to this view, certain folds have topological properties that make them especially *designable* that is stable, modular, and functional. One example is the TIM-barrel fold with a uniquely symmetrical construction.

On the other hand, one of the most striking observations in recent years was that mathematical laws seem to govern many biologically relationships, and it is difficult to imagine how specific selective procedures may have produced them. In particular, power laws describe the distribution of gene family sizes pseudogene families, the lengths of proteins in terms of domains, or the domain superfamilies among folds. Power laws also describe properties of scale-free networks that depict the relationships or interactions between proteins in metabolic pathways and domains within proteins.

Stochastic birth, death, and innovation models (BDIM) deliver one explanation of the power laws. They describe a *preferential attachment* mechanism in which the "fit get fitter" : large protein families tend to get larger; highly versatile domains tend to get even more different combination partners. Innovation comprises both the divergence of sequences and processes that introduce novel sequences into a genome, such as horizontal gene transfer. BDIMs approximate the observed power law distributions remarkably well and support a random mode in evolution. However, they basically ignore the individual properties of proteins and domains – a most crucial abstraction and simplification.

3 Summary and Conclusions – Evolutionary Principles

If life was like a book, written in an unknown language, then the words of this language may be the proteins. These proteins form sentences of pathways. Several sentences in a paragraph might correspond to an organelle; organelles are parts of cells, and cells of organisms. Here, we considered the protein vocabulary, and how the protein words are made of syllables called *domains*. Since the first determination of protein structures some 40 years ago, researchers have been studying these basic units, the domains. Now that the complete genome sequence of many organisms and their domain assignments are available, we can start to understand the rules of how domains form proteins and how our molecular vocabulary evolved.

We have given a description of the repertoire of proteins and their domains, how it has evolved thus far, and the mechanisms involved in bringing about this evolution. Our description revolved around the relationship between sequence, structure, and function of these domains, and how these characteristics determine the rate of protein evolution. We showed how the duplication, recombination, and divergence of protein domains had an enormous impact on the protein repertoire as we see it today.

In a very succinct review, Chothia and colleagues gave two possible reasons why these mechanisms of domain duplication, divergence, and recombination take place so commonly rather than via *ab initio* protein construction. The first sets of proteins created must have arisen via *ab initio* invention. Once an initial set of domains and proteins had evolved that was varied enough to support a basic form of life, it was much faster and more efficient to use duplication and recombination to create new protein combinations with altered functions. In addition, once repair mechanisms present in DNA replication

and protein synthesis had been created, they made the *ab initio* process for domain invention too costly and difficult to be used very often.

In addition, many other processes, such as regulation of gene expression, alternative splicing, RNA editing or protein modification, determine how many and which kinds of proteins are active in a particular cell at a particular point in time. They represent mechanisms by which the complexity of the protein repertoire that is active in a cell is increased and delivers one explanation of why higher eukaryotes can do with relatively few genes. They are important mechanisms that deserve separate, more detailed description, and go beyond the scope of this discussion.

For an even more complete understanding of evolution, an integration of most different fields will be necessary: ''genomics'' analyses will have to join molecular evolution approaches; structural biologists, population geneticists, and experimentalists will have to interact more closely. Today, we already see some fruitful outcomes of such integrative approaches. One example is the recent work by Lynch and Conery. The authors analyze the relationship between the genome complexity and population size: species of large populations, such as bacteria, have small genomes, while for species of small populations, such as humans, we observe the opposite. Another beautiful example examines the link between molecular sequence variation and molecular and macroscopic phenotype; and this is for the chaperone Hsp90, which acts as a buffer and regulator in response to environmental changes. The chaperone assists the maturation, that is, the correct folding, of many key regulatory proteins. Thus, genetic variation can accumulate in genomes and can remain phenotypically silent due to the buffering by the chaperone. Upon environmental changes, however, for example, under stress, the chaperone's function is challenged and it releases novel phenotypes of folded protein structures.

While the groundwork for an understanding of protein structures and their evolution has been done, the exact interaction between selection and random processes is not yet clear. Only future comprehensive, integrative analyses of domain superfamilies and their combinations will reveal how the selection, in combination with random processes, can have produced a most complex protein universe. Thus, eventually, once we fully understand the vocabulary of proteins and grammar of protein evolution, we will be able to read the whole story of the book of life.

See also Genetic Analysis of Populations; Immunoassays; Molecular Systematics and Evolution; Origins of Life, Molecular Basis of; Paleontology, Molecular; Protein Expression by Expansion of the Genetic Code.

Bibliography

Books and Reviews

Branden, C., Tooze, J. (1998) *Introduction to Protein Structure*, Garland Science, New York.

Chothia, C., Gough, J., Vogel, C., Teichmann, S.A. (2003) Evolution of the protein repertoire, *Science* **300**(5626), 1701–1703.

Grishin, N.V. (2001) Fold change in evolution of protein structures, *J. Struct. Biol.* **134**, 167–185.

Koonin, E.V., Galperin, M.Y. (2002) *Sequence – Evolution – Function: Computational Approaches in Comparative Genomics*, Kluwer Academic Publishers, Boston, MA.

Koonin, E.V., Wolf, Y.I., Karev, G.P. (2002) The structure of the protein universe and genome evolution, *Nature* **420**(6912), 218–223.

Li, W.-H., Graur, D. (1999) *Fundamentals of Molecular Evolution*, Sinauer Associates Incorporated, Sunderland, MA.

Patthy, L. (1999) *Protein Evolution*, Blackwell Science, Oxford.

Petsko, G., Ringe, D. (2003) *Protein Structure and Function*, Blackwell Publishing, Oxford.

Ponting, C.P., Russell, R.R. (2002) The natural history of protein domains, *Annu. Rev. Biophys. Biomol. Struct.* **31**, 45–71. Epub 2001 Oct 25.

Rubin, G.M., Yandell, M.D., et al. (2000) Comparative genomics of the eukaryotes, *Science* **287**(5461), 2204–2215.

Todd, A.E., Orengo, C.A., Thornton, J.M. (1999) Evolution of protein function, from a structural perspective, *Curr. Opin. Chem. Biol.* **3**(5), 548–556.

Primary Literature

Alberts, B. (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists, *Cell* **92**, 291–294.

Aloy, P., Russell, R.B. (2002) Interrogating protein interaction networks through structural biology, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5896–5901.

Aloy, P., Ceulemans, H., Stark, A., Russell, R.B. (2003) The relationship between sequence and interaction divergence in proteins, *J. Mol. Biol.* **332**, 989–998.

Aloy, P., Ciccarelli, F.D., Leutwein, C., Gavin, A.C., Superti-Furga, G., Bork, P., Bottcher, B., Russell, R.B. (2002) A complex prediction: three-dimensional model of the yeast exosome, *EMBO Rep.* **3**, 628–635.

Alves, R., Chaleil, R.A.G., Sternberg, M.J.E. (2002) Evolution of enzymes in metabolism: a network perspective, *J. Mol. Biol.* **320**, 751–770.

Andrade, M.A., Perez-Iratxeta, C., Ponting, C.P. (2001) Protein repeats: structures, functions, and evolution, *J. Struct. Biol.* **134**, 117–131.

Andreeva, A., Howorth, D., Brenner, S.E., Hubbard, T.J., Chothia, C., Murzin, A.G. (2004) SCOP database in 2004: refinements integrate structure and sequence family data, *Nucleic Acids Res.* **32**, D226–D229.

Anfinsen, C.B., Haber, E., Sela, M., White, F.H. Jr. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain, *Proc. Natl. Acad. Sci. U.S.A.* **47**, 1309–1314.

Apic, G., Gough, J., Teichmann, S.A. (2001) Domain combinations in archaeal, eubacterial and eukaryotic proteomes, *J. Mol. Biol.* **310**, 311–325.

Apic, G., Huber, W., Teichmann, S.A. (2003) Multidomain protein families and domain pairs: comparison with known structures and a random model of domain recombination, *J. Struct. Funct. Genomics* **4**, 67–78.

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D.R., et al. (2000) InterPro – an integrated documentation resource for protein families, domains and functional sites, *Bioinformatics* **16**, 1145–1150.

Aravind, L., Mazumder, R., Vasudevan, S., Koonin, E.V. (2002) Trends in protein evolution inferred from sequence and structure analysis, *Curr. Opin. Struct. Biol.* **12**, 392–399.

Aravind, L., Tatusov, R.L., Wolf, Y.I., Walker, D.R., Koonin, E.V. (1998) Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles, *Trends Genet.* **14**, 442–444.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nat. Genet.* **25**, 25–29.

Bairoch, A. (2000) The ENZYME database in 2000, *Nucleic Acids Res.* **28**, 304–305.

Bajaj, M., Blundell, T. (1984) Evolution and the tertiary structure of proteins, *Annu. Rev. Biophys. Bioeng.* **13**, 453–492.

Barabasi, A.L., Albert, R. (1999) Emergence of scaling in random networks, *Science* **286**, 509–512.

Barabasi, A.L., Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization, *Nat. Rev. Genet.* **5**, 101–113.

Bartlett, G.J., Borkakoti, N., Thornton, J.M. (2003) Catalysing new reactions during evolution: economy of residues and mechanism, *J. Mol. Biol.* **331**, 829–860.

Bashton, M., Chothia, C. (2002) The geometry of domain combination in proteins, *J. Mol. Biol.* **315**, 927–939.

Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., et al. (2004) The Pfam protein families database, *Nucleic Acids Res.* **32**, D138–D141.

Beckmann, R., Spahn, C.M., Eswar, N., Helmers, J., Penczek, P.A., Sali, A., Frank, J., Blobel, G. (2001) Architecture of the protein-conducting channel associated with the translating 80S ribosome, *Cell* **107**, 361–372.

Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., et al. (2002) The protein data bank, *Acta Crystallogr. D Biol. Crystallogr.* **58**, 899–907.

Betts, M.J., Guigo, R., Agarwal, P., Russell, R.B. (2001) Exon structure conservation despite low sequence similarity: a relic of dramatic events in evolution? *EMBO J.* **20**, 5354–5360.

Bilke, S., Peterson, C. (2001) Topological properties of citation and metabolic networks, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **64**, 036106.

Bork, P., Jensen, L.J., von Mering, C., Ramani, A.K., Lee, I., Marcotte, E.M. (2004) Protein interaction networks from yeast to human, *Curr. Opin. Struct. Biol.* **14**, 292–299.

Branden, C.I., Tooze, J. (1999) *Introduction to Protein Structure*, Garland Publishing, New York.

Brenner, S.E., Chothia, C., Hubbard, T.J. (1998) Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6073–6078.

Brenner, S.E., Chothia, C., Hubbard, T.J.P., Murzin, A.G. (1996) Understanding Protein Structure: Using SCOP for Fold Interpretation, in: Doolittle, R.F. (Ed.) *Computer Methods for Macromolecular Sequence Analysis*, Chap. 37, Methods in Enzymology Vol. 266, Academic Press, Orlando, FL, pp. 635–643.

Caetano-Anolles, G., Caetano-Anolles, D. (2003) An evolutionarily structured universe of protein architecture, *Genome Res.* **13**, 1563–1571.

Castillo-Davis, C.I., Hartl, D.L., Achaz, G. (2004) cis-regulatory and protein evolution in orthologous and duplicate genes, *Genome Res.* **14**, 1530–1536.

Cavalier-Smith, T. (1991) Intron phylogeny: a new hypothesis, *Trends Genet.* **7**, 145–148.

Charlesworth, B., Sniegowski, P., Stephan, W. (1994) The evolutionary dynamics of repetitive DNA in eukaryotes, *Nature* **371**, 215–220.

Chen, J., Anderson, J.B., DeWeese-Scott, C., Fedorova, N.D., Geer, L.Y., He, S., Hurwitz, D.I., Jackson, J.D., Jacobs, A.R., Lanczycki, C.J., et al. (2003) MMDB: Entrez's 3D-structure database, *Nucleic Acids Res.* **31**, 474–477.

Chervitz, S.A., Aravind, L., Sherlock, G., Ball, C.A., Koonin, E.V., Dwight, S.S., Harris, M.A., Dolinski, K., Mohr, S., Smith, T., et al. (1998) Comparison of the complete protein sets of worm and yeast: Orthology and divergence, *Science* **282**, 2022–2028.

Chirgadze, D.Y., Demydchuk, M., Becker, M., Moran, S., Paoli, M. (2004) Snapshot of protein structure evolution reveals conservation of functional dimerization through intertwined folding, *Structure (Camb)* **12**, 1489–1494.

Chothia, C. (1992) Proteins – 1000 families for the molecular biologist, *Nature* **357**, 543–544.

Chothia, C., Lesk, A.M. (1982) Evolution of proteins formed by Beta-sheets. 1. Plastocyanin and Azurin, *J. Mol. Biol.* **160**, 309–323.

Chothia, C., Lesk, A.M. (1986) The relation between the divergence of sequence and structure in proteins, *EMBO J.* **5**, 823–826.

Chothia, C., Gough, J., Vogel, C., Teichmann, S.A. (2003) Evolution of the protein repertoire, *Science* **300**, 1701–1703.

Copley, R.R., Bork, P. (2000) Homology among (beta alpha)(8) barrels: Implications for the evolution of metabolic pathways, *J. Mol. Biol.* **303**, 627–640.

Copley, R.R., Schultz, J., Ponting, C.P., Bork, P. (1999) Protein families in multicellular organisms, *Curr. Opin. Struct. Biol.* **9**, 408–415.

Coulson, A.F., Moult, J. (2002) A unifold, mesofold, and superfold model of protein fold use, *Proteins: Struct., Funct., Genet.* **46**, 61–71.

Dandekar, T., Snel, B., Huynen, M., Bork, P. (1998) Conservation of gene order: a fingerprint of proteins that physically interact, *Trends Biochem. Sci.* **23**, 324–328.

Dandekar, T., Schuster, S., Snel, B., Huynen, M., Bork, P. (1999) Pathway alignment: application to the comparative analysis of glycolytic enzymes, *Biochem. J.* **343**(Pt 1), 115–124.

Darwin, C. (1859) *The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, John Murray, London.

Dayhoff, M.O. (1976) The origin and evolution of protein superfamilies, *Fed. Proc.* **35**, 2132–2138.

Dodson, G., Wlodawer, A. (1998) Catalytic triads and their relatives, *Trends Biochem. Sci.* **23**, 347–352.

Doolittle, W.E. (1998) You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes, *Trends Genet.* **14**, 307–311.

Doolittle, W.F. (1987) What introns have to tell us – hierarchy in genome evolution, *Cold. Spring Harb. Symp. Quant. Biol.* **52**, 907–913.

Doolittle, W.F. (1999a) Lateral genomics, *Trends Biochem. Sci.* **24**, M5–M8.

Doolittle, W.F. (1999b) Phylogenetic classification and the universal tree, *Science* **284**, 2124–2128.

Doolittle, W.F. (2000) The nature of the universal ancestor and the evolution of the proteome, *Curr. Opin. Struct. Biol.* **10**, 355–358.

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuveglise, C., Talla, E., et al. (2004) Genome evolution in yeasts, *Nature* **430**, 35–44.

Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., Obradovic, Z. (2002) Intrinsic disorder and protein function, *Biochemistry* **41**, 6573–6582.

Dunker, A.K., Obradovic, Z., Romero, P., Garner, E.C., Brown, C.J. (2000) Intrinsic protein disorder in complete genomes, *Genome Inform. Ser. Workshop Genome Inform.* **11**, 161–171.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., et al. (2001) Intrinsically disordered protein, *J. Mol. Graph. Model.* **19**, 26–59.

Dunwell, J.M., Purvis, A., Khuri, S. (2004) Cupins: the most functionally diverse protein superfamily? *Phytochemistry* **65**, 7–17.

Eisen, J.A. (1998) A phylogenomic study of the MutS family of proteins, *Nucleic Acids Res.* **26**, 4291–4300.

Enright, A., Van Dongen, S., Ouzounis, C. (2002) An efficient algorithm for large-scale detection of protein families, *Nucleic Acids Res.* **30**, 1575–1584.

Enright, A.J., Iliopoulos, I., Kyrpides, N.C., Ouzounis, C.A. (1999) Protein interaction maps for complete genomes based on gene fusion events, *Nature* **402**, 86–90.

Fitch, W.M. (1970) Distinguishing homologous from analogous proteins, *Syst. Zool.* **19**, 99–113.

Fitch, W.M. (2000) Homology a personal view on some of the problems, *Trends Genet.* **16**, 227–231.

Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C., Feldman, M.W. (2002) Evolutionary rate in the protein interaction network, *Science* **296**, 750–752.

Galperin, M.Y., Walker, D.R., Koonin, E.V. (1998) Analogous enzymes: Independent inventions in enzyme evolution, *Genome Res.* **8**, 779–790.

Gerstein, M. (1997) A structural census of genomes: comparing bacterial, eukaryotic, and archaeal genomes in terms of protein structure, *J. Mol. Biol.* **274**, 562–576.

Gilbert, W. (1978) Why genes in pieces? *Nature* **271**, 501.

Gilbert, W. (1979) *Introns and Exons: Playgrounds of Evolution*, Academic Press, New York.

Gough, J. (2004) Convergent evolution of domain architectures (is rare), *Bioinformatics*.

Gough, J., Chothia, C. (2002) SUPERFAMILY: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments, *Nucleic Acids Res.* **30**, 268–272.

Gough, J., Karplus, K., Hughey, R., Chothia, C. (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure, *J. Mol. Biol.* **313**, 903–919.

Govindarajan, S., Recabarren, R., Goldstein, R.A. (1999) Estimating the total number of protein folds, *Proteins: Struct., Funct., Genet.* **35**, 408–414.

Grishin, N.V. (2001) Fold change in evolution of protein structures, *J. Struct. Biol.* **134**, 167–185.

Gu, Z., Nicolae, D., Lu, H.H., Li, W.H. (2002) Rapid divergence in expression between duplicate genes inferred from microarray data, *Trends Genet.* **18**, 609–613.

Hadley, C., Jones, D.T. (1999) A systematic comparison of protein structure classifications: SCOP, CATH and FSSP, *Structure Fold. Des.* **7**, 1099–1112.

Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P., et al. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network, *Nature* **430**, 88–93.

Harrison, S.C. (2003) Variation on an Src-like theme, *Cell* **112**, 737–740.

Harrison, P.M., Gerstein, M. (2002) Studying genomes through the aeons: Protein families, pseudogenes and proteome evolution, *J. Mol. Biol.* **318**, 1155–1174.

Heger, A., Holm, L. (2000) Towards a covering set of protein family profiles, *Prog. Biophys. Mol. Biol.* **73**, 321–337.

Hegyi, H., Gerstein, M. (1999) The relationship between protein structure and function: a comprehensive survey with application to the yeast genome, *J. Mol. Biol.* **288**, 147–164.

Hegyi, H., Gerstein, M. (2001) Annotation transfer for genomics: measuring functional divergence in multidomain proteins, *Genome Res.* **11**, 1632–1640.

Hill, E., Morea, V., Chothia, C. (2002) Sequence conservation in families whose members have little or no sequence similarity: the four-helical cytokines and cytochromes, *J. Mol. Biol.* **322**, 205–233.

Hill, E., Broadbent, I.D., Chothia, C., Pettitt, J. (2001) Cadherin superfamily proteins in Caenorhabditis elegans and Drosophila melanogaster, *J. Mol. Biol.* **305**, 1011–1024.

Holm, L., Sander, C. (1996) The FSSP database: fold classification based on structure structure alignment of proteins, *Nucleic Acids Res.* **24**, 206–209.

Hughes, A.L., Nei, M. (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection, *Nature* **335**, 167–170.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., et al. (2000) Functional discovery via a compendium of expression profiles, *Cell* **102**, 109–126.

Hurst, L.D., Smith, N.G. (1999) Do essential genes evolve slowly? *Curr. Biol.* **9**, 747–750.

Hurst, G.D., Werren, J.H. (2001) The role of selfish genetic elements in eukaryotic evolution, *Nat. Rev. Genet.* **2**, 597–606.

Hurst, L.D., Pal, C., Lercher, M.J. (2004) The evolutionary dynamics of eukaryotic gene order, *Nat. Rev. Genet.* **5**, 299–310.

Huynen, M.A., van Nimwegen, E. (1998) The frequency distribution of gene family sizes in complete genomes, *Mol. Biol. Evol.* **15**, 583–589.

Ingram, V.M. (1956) A specific chemical difference between the globins of normal human and sickle cell anaemia haemoglobin, *Nature* **178**, 792–794.

Ingram, V.M. (2004) Sickle cell anemia hemoglobin: the molecular biology of the first "molecular disease"–the crucial importance of serendipity, *Genetics* **167**, 1–7.

Irving, J.A., Whisstock, J.C., Lesk, A.M. (2001) Protein structural alignments and functional genomics, *Proteins: Struct., Funct., Genet.* **42**, 378–382.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4569–4574.

Janin, J., Chothia, C. (1985) Domains in proteins – definitions, location, and structural principles, *Methods Enzymol.* **115**, 420–430.

Jardine, O., Gough, J., Chothia, C., Teichmann, S.A. (2002) Comparison of the small molecule metabolic enzymes of Escherichia coli and Saccharomyces cerevisiae, *Genome Res.* **12**, 916–929.

Jeong, H., Mason, S.P., Barabasi, A.L., Oltvai, Z.N. (2001) Lethality and centrality in protein networks, *Nature* **411**, 41–42.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabasi, A.L. (2000) The large-scale organization of metabolic networks, *Nature* **407**, 651–654.

Jordan, I.K., Makarova, K.S., Spouge, J.L., Wolf, Y.I., Koonin, E.V. (2001) Lineage-specific gene expansions in bacterial and archaeal genomes, *Genome Res.* **11**, 555–565.

Kaessmann, H., Zollner, S., Nekrutenko, A., Li, W.H. (2002) Signatures of domain shuffling in the human genome, *Genome Res.* **12**, 1642–1650.

Karev, G.P., Wolf, Y.I., Koonin, E.V. (2003) Simple stochastic birth and death models of genome evolution: was there enough time for us to evolve? *Bioinformatics* **19**, 1889–1900.

Karev, G.P., Wolf, Y.I., Rzhetsky, A.Y., Berezovskaya, F.S., Koonin, E.V. (2002) Birth and death of protein domains: a simple model of evolution explains power law behavior, *BMC Evol. Biol.* **2**, 18.

Kendrew, J.C. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis, *Nature* **181**, 662–666.

Khaitovich, P., Weiss, G., Lachmann, M., Hellmann, I., Enard, W., Muetzel, B., Wirkner, U., Ansorge, W., Paabo, S. (2004) A neutral model of transcriptome evolution, *PLoS Biol.* **2**, E132.

Kidwell, M.G. (2002) Transposable elements and the evolution of genome size in eukaryotes, *Genetica* **115**, 49–63.

Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge, MA.

Kinch, L.N., Grishin, N.V. (2002) Evolution of protein structures and functions, *Curr. Opin. Struct. Biol.* **12**, 400–408.

Koonin, E.V., Aravind, L., Kondrashov, A.S. (2000) The impact of comparative genomics on our understanding of evolution, *Cell* **101**, 573–576.

Koonin, E.V., Makarova, K.S., Aravind, L. (2001) Horizontal gene transfer in prokaryotes: quantification and classification, *Annu. Rev. Microbiol.* **55**, 709–742.

Koonin, E.V., Mushegian, A.R., Bork, P. (1996) Non-orthologous gene displacement, *Trends Genet.* **12**, 334–336.

Koonin, E.V., Wolf, Y.I., Karev, G.P. (2002) The structure of the protein universe and genome evolution, *Nature* **420**, 218–223.

Kuznetsov, V.A. (2002). *Computational and Statistical Approaches to Genomics*, Kluwer, Boston, MA, pp. 125–171.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Lang, D., Thoma, R., Henn-Sax, M., Sterner, R., Wilmanns, M. (2000) Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion, *Science* **289**, 1546–1550.

Lee, D., Grant, A., Buchan, D., Orengo, C. (2003) A structural perspective on genome evolution, *Curr. Opin. Struct. Biol.* **13**, 359–369.

Lercher, M.J., Urrutia, A.O., Hurst, L.D. (2002) Clustering of housekeeping genes provides a unified model of gene order in the human genome, *Nat. Genet.* **31**, 180–183.

Lesk, A.M., Chothia, C. (1980) How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins, *J. Mol. Biol.* **136**, 225–270.

Lespinet, O., Wolf, Y.I., Koonin, E.V., Aravind, L. (2002) The role of lineage-specific gene family expansion in the evolution of eukaryotes, *Genome Res.* **12**, 1048–1059.

Letunic, I., Copley, R.R., Schmidt, S., Ciccarelli, F.D., Doerks, T., Schultz, J., Ponting, C.P., Bork, P. (2004) SMART 4.0: towards genomic data integration, *Nucleic Acids Res.* **32**, D142–D144.

Levitt, M., Chothia, C. (1976) Structural patterns in globular proteins, *Nature* **261**, 552–558.

Liu, J., Rost, B. (2003) Domains, motifs and clusters in the protein universe, *Curr. Opin. Chem. Biol.* **7**, 5–11.

Liu, Y., Gerstein, M., Engelman, D.M. (2004) Evolutionary use of domain recombination: a distinction between membrane and soluble proteins, *Proc. Natl. Acad. Sci. U.S.A.* **101**(10), 3495–3497.

Lupas, A.N., Ponting, C.P., Russell, R.B. (2001) On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world? *J. Struct. Biol.* **134**, 191–203.

Luscombe, N., Qian, J., Zhang, Z., Johnson, T., Gerstein, M. (2002) The dominance of the population by a selected few: power law behavior applies to a wide variety of genomic properties, *Genome Biol.* **3**, RESEARCH0040.

Lynch, M., Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes, *Science* **290**, 1151–1155.

Lynch, M., Conery, J. (2003) The origins of genome complexity, *Science* **302**, 1401–1404.

Madera, M., Gough, J. (2002) A comparison of profile hidden Markov model procedures for remote homology detection, *Nucleic Acids Res.* **19**, 30.

Madera, M., Vogel, C., Kummerfeld, S.K., Chothia, C., Gough, J. (2004) The SUPERFAMILY database in 2004: additions and improvements, *Nucleic Acids Res.* **32**, D235–D239.

Makalowski, W., Boguski, M.S. (1998) Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9407–9412.

Makarova, K.S., Aravind, L., Daly, M.J., Koonin, E.V. (2000) Specific expansion of protein families in the radioresistant bacterium

Deinococcus radiodurans, *Genetica* **108**, 25–34.

Makarova, K.S., Aravind, L., Galperin, M.Y., Grishin, N.V., Tatusov, R.L., Wolf, Y.I., Koonin, E.V. (1999) Comparative genomics of the archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell, *Genome Res.* **9**, 608–628.

Makarova, K.S., Aravind, L., Wolf, Y.I., Tatusov, R.L., Minton, K.W., Koonin, E.V., Daly, M.J. (2001) Genome of the extremely radiation-resistant bacterium Deinococcus radiodurans viewed from the perspective of comparative genomics, *Microbiol. Mol. Biol. Rev.* **65**, 44–79.

Marchler-Bauer, A., Addess, K.J., Chappey, C., Geer, L., Madej, T., Matsuo, Y., Wang, Y., Bryant, S.H. (1999) MMDB: Entrez's 3D structure database, *Nucleic Acids Res.* **27**, 240–243.

Marcotte, E.M., Pellegrini, M., Ng, H.L., Rice, D.W., Yeates, T.O., Eisenberg, D. (1999a) Detecting protein function and protein-protein interactions from genome sequences, *Science* **285**, 751–753.

Marcotte, E.M., Pellegrini, M., Thompson, M.J., Yeates, T.O., Eisenberg, D. (1999b) A combined algorithm for genome-wide prediction of protein function, *Nature* **402**, 83–86.

Martin, A.C., Orengo, C.A., Hutchinson, E.G., Jones, S., Karmirantzou, M., Laskowski, R.A., Mitchell, J.B., Taroni, C., Thornton, J.M. (1998) Protein folds and functions, *Structure Fold. Des.* **6**, 875–884.

Mazet, F., Shimeld, S.M. (2002) Gene duplication and divergence in the early evolution of vertebrates, *Curr. Opin. Genet. Dev.* **12**, 393–396.

McKenzie, A.N., Barry, S.C., Strath, M., Sanderson, C.J. (1991) Structure-function analysis of interleukin-5 utilizing mouse/human chimeric molecules, *EMBO J.* **10**, 1193–1199.

Miles, E.W., Davies, D.R. (2000) Protein evolution. On the ancestry of barrels, *Science* **289**, 1490.

Miyata, T., Yasunaga, T. (1980) Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application, *J. Mol. Evol.* **16**, 23–36.

Miyata, T., Miyazawa, S., Yasunaga, T. (1979) Two types of amino acid substitutions in protein evolution, *J. Mol. Evol.* **12**, 219–236.

Mott, R., Schultz, J., Bork, P., Ponting, C.P. (2002) Predicting protein cellular localization using a domain projection method, *Genome Res.* **12**, 1168–1174.

Muller, A., MacCallum, R.M., Sternberg, M.J. (2002) Structural characterization of the human proteome, *Genome Res.* **12**, 1625–1641.

Murzin, A.G. (1998) How far divergent evolution goes in proteins, *Curr. Opin. Struct. Biol.* **8**, 380–387.

Murzin, A.G., Brenner, S.E., Hubbard, T., Chothia, C. (1995) Scop – a structural classification of proteins database for the investigation of sequences and structures, *J. Mol. Biol.* **247**, 536–540.

Nagano, N., Orengo, C.A., Thornton, J.M. (2002) One fold with many functions: the evolutionary relationships between TIM-barrel families based on their sequences, structures and functions, *J. Mol. Biol.* **321**, 741–765.

Nagarajan, N., Yona, G. (2004) Automatic prediction of protein domains from sequence information using a hybrid learning system, *Bioinformatics* **20**, 1335–1360.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., et al. (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima, *Nature* **399**, 323–329.

Newlove, T., Konieczka, J.H., Cordes, M.H. (2004) Secondary structure switching in Cro protein evolution, *Structure (Camb)* **12**, 569–581.

Nicola, N.A., Hilton, D.J. (1998) General classes and functions of four-helix bundle cytokines, *Adv. Protein Chem.* **52**, 1–65.

Nobbs, C.L., Watson, H.C., Kendrew, J.C. (1966) Structure of deoxymyoglobin: a crystallographic study, *Nature* **209**, 339–341.

Ogata, H., Audic, S., Barbe, V., Artiguenave, F., Fournier, P.E., Raoult, D., Claverie, J.M. (2000) Selfish DNA in protein-coding genes of Rickettsia, *Science* **290**, 347–350.

Ohno, S. (1970) *Evolution by Gene Duplication*, Springer-Verlag, New York.

Orengo, C.A., Jones, D.T., Thornton, J.M. (1994) Protein superfamilies and domain superfolds, *Nature* **372**, 631–634.

Orengo, C.A., Pearl, F.M., Thornton, J.M. (2003) The CATH domain structure database, *Methods Biochem. Anal.* **44**, 249–271.

Orengo, C.A., Sillitoe, I., Reeves, G., Pearl, F.M.G. (2001) Review: what can structural classifications reveal about protein evolution? *J. Struct. Biol.* **134**, 145–165.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., Thornton, J.M. (1997) CATH – a hierarchic classification of protein domain structures, *Structure* **5**, 1093–1108.

Ouzounis, C.A., Coulson, R.M., Enright, A.J., Kunin, V., Pereira-Leal, J.B. (2003) Classification schemes for protein structure and function, *Nat. Rev. Genet.* **4**, 508–519.

Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G.D., Maltsev, N. (1999) The use of gene clusters to infer functional coupling, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2896–2901.

Papp, B., Pál, C., Hurst, L. (2003) Dosage sensitivity and the evolution of gene families in yeast, *Nature* **424**, 194–197.

Papp, B., Pal, C., Hurst, L.D. (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast, *Nature* **429**, 661–664.

Park, J., Lappe, M., Teichmann, S.A. (2001) Mapping protein family interactions: Intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast, *J. Mol. Biol.* **307**, 929–938.

Park, J., Karplus, K., Barrett, C., Hughey, R., Haussler, D., Hubbard, T., Chothia, C. (1998) Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods, *J. Mol. Biol.* **284**, 1201–1210.

Patthy, L. (1994) Exons and introns, *Curr. Opin. Struct. Biol.* **4**, 383–392.

Patthy, L. (1999) Genome evolution and the evolution of exon-shuffling–a review, *Gene* **238**, 103–114.

Patthy, L. (2003) Modular assembly of genes and the evolution of new functions, *Genetica* **118**, 217–231.

Pawson, T., Nash, P. (2003) Assembly of cell regulatory systems through protein interaction domains, *Science* **300**, 445–452.

Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D., Yeates, T.O. (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4285–4288.

Perutz, M.F., Kendrew, J.C., Watson, H.C. (1965) Structure and function of haemoglobin: II. Some relations between polypeptide chain configuration and amino acid sequence, *J. Mol. Biol.* **13**, 669–678.

Perutz, M., Muirhead, H., Cox, J.M., Goaman, L.C. (1968) Three-dimensional Fourier synthesis of horse oxyhemoglobin at 2.8 A resolution: the atomic model, *Nature* **219**, 131–139.

Ponting, C.P., Russell, R.R. (2002) The natural history of protein domains, *Annu. Rev. Biophys. Biomol. Struct.* **31**, 45–71.

Prabu, M.M., Suguna, K., Vijayan, M. (1999) Variability in quaternary association of proteins with the same tertiary fold: a case study and rationalization involving legume lectins, *Proteins: Struct., Funct., Genet.* **35**, 58–69.

Qian, J., Luscombe, N.M., Gerstein, M. (2001) Protein family and fold occurrence in genomes: Power-law behaviour and evolutionary model, *J. Mol. Biol.* **313**, 673–681.

Ragan, M.A. (2001) Detection of lateral gene transfer among microbial genomes, *Curr. Opin. Genet. Dev.* **11**, 620–626.

Ranea, J.A., Buchan, D.W., Thornton, J.M., Orengo, C.A. (2004) Evolution of protein superfamilies and bacterial genome size, *J. Mol. Biol.* **336**, 871–887.

Reidhaar-Olson, J.F., Sauer, R.T. (1988) Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences, *Science* **241**, 53–57.

Remm, M., Storm, C.E., Sonnhammer, E.L. (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons, *J. Mol. Biol.* **314**, 1041–1052.

Riley, M., Labedan, B. (1996) E.coli Gene Products: Physiological Functions and Common Ancestries, in: Neidhart, F.C. (Ed.) *Escherichia Coli and Salmonella, Cellular and Molecular Biology*, ASM Press, Washington, DC, pp. 2118–2202.

Riley, M., Serres, M.H. (2000) Interim report on genomics of Escherichia coli, *Annu. Rev. Microbiol.* **54**, 341–411.

Rison, S.C., Teichmann, S.A., Thornton, J.M. (2002) Homology, pathway distance and chromosomal localization of the small molecule metabolism enzymes in Escherichia coli, *J. Mol. Biol.* **318**, 911–932.

Rokas, A., Williams, B., King, N., Carroll, S. (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies, *Nature* **425**, 798–804.

Rossmann, M.G., Moras, D., Olsen, K.W. (1974) Chemical and biological evolution of nucleotide-binding protein, *Nature* **250**, 194–199.

Rubin, G.M., Yandell, M.D., Wortman, J.R., Gabor Miklos, G.L., Nelson, C.R., Hariharan, I.K., Fortini, M.E., Li, P.W., Apweiler, R., Fleischmann, W., et al. (2000) Comparative genomics of the eukaryotes, *Science* **287**, 2204–2215.

Russell, R.B. (1994) Domain insertion, *Protein Eng.* **7**, 1407–1410.

Russell, R.B. (1998) Detection of protein three-dimensional side-chain patterns: new examples of convergent evolution, *J. Mol. Biol.* **279**, 1211–1227.

Russell, R.B., Alber, F., Aloy, P., Davis, F.P., Korkin, D., Pichaud, M., Topf, M., Sali, A. (2004) A structural perspective on protein-protein interactions, *Curr. Opin. Struct. Biol.* **14**, 313–324.

Rzhetsky, A., Gomez, S.M. (2001) Birth of scale-free molecular networks and the number of distinct DNA and protein domains per genome, *Bioinformatics* **17**, 988–996.

Sangster, T.A., Lindquist, S., Queitsch, C. (2004) Under cover: causes, effects and implications of Hsp90-mediated genetic capacitance, *Bioessays* **26**, 348–362.

Schmidt, S., Sunyaev, S., Bork, P., Dandekar, T. (2003) Metabolites: a helping hand for pathway evolution?, *Trends Biochem. Sci.* **28**, 336–341.

Schultz, J., Milpetz, F., Bork, P., Ponting, C.P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5857–5864.

Semple, C., Wolfe, K.H. (1999) Gene duplication and gene conversion in the Caenorhabditis elegans genome, *J. Mol. Evol.* **48**, 555–564.

Serres, M.H., Goswami, S., Riley, M. (2004) GenProtEC: an updated and improved analysis of functions of Escherichia coli K-12 proteins, *Nucleic Acids Res.* **32**, D300–D302.

Sidow, A. (1996) Genome duplications in the evolution of early vertebrates, *Curr. Opin. Genet. Dev.* **6**, 715–722.

Simillion, C., Vandepoele, K., Saeys, Y., Van de Peer, Y. (2004) Building genomic profiles for uncovering segmental homology in the twilight zone, *Genome Res.* **14**, 1095–1106.

Snel, B., Bork, P., Huynen, M. (2000) Genome evolution. Gene fusion versus gene fission, *Trends Genet.* **16**, 9–11.

Snel, B., Bork, P., Huynen, M. (2002) Genomes in flux: the evolution of archaeal and proteobacterial gene content, *Genome Res.* **12**, 17–25.

Sonnhammer, E.L., Koonin, E.V. (2002) Orthology, paralogy and proposed classification for paralog subtypes, *Trends Genet.* **18**, 619–620.

Spahn, C.M., Beckmann, R., Eswar, N., Penczek, P.A., Sali, A., Blobel, G., Frank, J. (2001) Structure of the 80S ribosome from Saccharomyces cerevisiae–tRNA-ribosome and subunit-subunit interactions, *Cell* **107**, 373–386.

Swindells, M.B., Orengo, C.A., Jones, D.T., Hutchinson, E.G., Thornton, J.M. (1998) Contemporary approaches to protein structure classification, *Bioessays* **20**, 884–891.

Tatusov, R.L., Koonin, E.V., Lipman, D.J. (1997) A genomic perspective on protein families, *Science* **278**, 631–637.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution, *Nucleic Acids Res.* **28**, 33–36.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., et al. (2003) The COG database: an updated version includes eukaryotes, *BMC Bioinform.* **4**, 41.

Teichmann, S.A. (2002) The constraints protein-protein interactions place on sequence divergence, *J. Mol. Biol.* **324**, 399–407.

Teichmann, S.A., Park, J., Chothia, C. (1998) Structural assignments to the Mycoplasma genitalium proteins show extensive gene duplications and domain rearrangements, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14658–14663.

Teichmann, S.A., Rison, S.C., Thornton, J.M., Riley, M., Gough, J., Chothia, C. (2001) Small-molecule metabolism: an enzyme mosaic, *Trends Biotechnol.* **19**, 482–486.

Timsit, Y. (2001) Convergent evolution of MutS and topoisomerase II for clamping DNA crossovers and stacked Holliday junctions, *J. Biomol. Struct. Dyn.* **19**, 215–218.

Todd, A.E., Orengo, C.A., Thornton, J.M. (1999) Evolution of protein function, from a structural perspective, *Curr. Opin. Chem. Biology.* **3**, 548–556.

Todd, A.E., Orengo, C.A., Thornton, J.M. (2001) Evolution of function in protein superfamilies, from a structural perspective, *J. Mol. Biol.* **307**, 1113–1143.

Todd, A.E., Orengo, C.A., Thornton, J.M. (2002) Plasticity of enzyme active sites, *Trends Biochem. Sci.* **27**, 419–426.

Tompa, P. (2002) Intrinsically unstructured proteins, *Trends Biochem. Sci.* **27**, 527–533.

Tsoka, S., Ouzounis, C.A. (2001) Functional versatility and molecular diversity of the metabolic map of Escherichia coli, *Genome Res.* **11**, 1503–1510.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae, *Nature* **403**, 623–627.

van Nimwegen, E. (2003) Scaling laws in the functional content of genomes, *Trends Genet.* **19**, 479–484.

Vitagliano, L., Masullo, M., Sica, F., Zagari, A., Bocchini, V. (2001) Crystal structure of Sulfolobus solfataricus Elongation factor 1 alpha in complex with GDP, *EMBO J.* **20**, 5305.

Vogel, C., Teichmann, S.A., Chothia, C. (2003) The immunoglobulin superfamily in Drosophila melanogaster and Caenorhabditis elegans and the evolution of complexity, *Development* **130**, 6317–6328.

Vogel, C., Berzuini, C., Bashton, M., Gough, J., Teichmann, S.A. (2004) Supra-domains–evolutionary units larger than single protein domains, *J. Mol. Biol.* **336**, 809–823.

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., Bork, P. (2002) Comparative assessment of large-scale data sets of protein-protein interactions, *Nature* **417**, 399–403.

Wagner, A. (2000a) Decoupled evolution of coding region and mRNA expression patterns after gene duplication: implications for the neutralist-selectionist debate, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6579–6584.

Wagner, A. (2000b) Robustness against mutations in genetic networks of yeast, *Nat. Genet.* **24**, 355–361.

Wang, Z.X. (1998) A reestimation for the total numbers of protein folds and superfamilies, *Protein Eng.* **11**, 621–626.

Webber, C., Ponting, C.P. (2004) Genes and homology, *Curr. Biol.* **14**, R332–R333.

Wetlaufer, D.B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 697–701.

Wilson, C.A., Kreychman, J., Gerstein, M. (2000) Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure and function through traditional and probabilistic scores, *J. Mol. Biol.* **297**, 233–249.

Winzeler, E.A., Liang, H., Shoemaker, D.D., Davis, R.W. (2000) Functional analysis of the yeast genome by precise deletion and parallel phenotypic characterization, *Novartis Found. Symp.* **229**, 105–109; discussion 109–111.

Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., et al. (1999) Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis, *Science* **285**, 901–906.

Wolf, Y.I., Aravind, L., Koonin, E.V. (1999a) Rickettsiae and Chlamydiae – evidence of horizontal gene transfer and gene exchange, *Trends Genet.* **15**, 173–175.

Wolf, Y.I., Grishin, N.V., Koonin, E.V. (2000) Estimating the number of protein folds and families from complete genome data, *J. Mol. Biol.* **299**, 897–905.

Wolf, Y.I., Karev, G., Koonin, E.V. (2002) Scale-free networks in biology: new insights into the fundamentals of evolution? *Bioessays* **24**, 105–109.

Wolf, Y.I., Rogozin, I.B., Koonin, E.V. (2004) Coelomata and not Ecdysozoa: evidence from genome-wide phylogenetic analysis, *Genome Res.* **14**, 29–36.

Wolf, Y.I., Brenner, S.E., Bash, P.A., Koonin, E.V. (1999b) Distribution of protein folds in the three superkingdoms of life, *Genome Res.* **9**, 17–26.

Wolfe, K.H. (2001) Yesterday's polyploids and the mystery of diploidization, *Nat. Rev. Genet.* **2**, 333–341.

Wolfe, K.H., Li, W.H. (2003) Molecular evolution meets the genomics revolution, *Nat. Genet.* **33**, 255–265.

Wolfe, K.H., Shields, D.C. (1997) Molecular evidence for an ancient duplication of the entire yeast genome, *Nature* **387**, 708–713.

Wright, C.S., Alden, R.A., Kraut, J. (1969) Structure of subtilisin BPN' at 2.5 angstrom resolution, *Nature* **221**, 235–242.

Wuchty, S. (2001) Scale-free behavior in protein domain networks, *Mol. Biol. Evol.* **18**, 1694–1702.

Wuchty, S. (2002) Interaction and domain networks of yeast, *Proteomics* **2**, 1715–1723.

Yanai, I., Wolf, Y.I., Koonin, E.V. (2002) Evolution of gene fusions: horizontal transfer versus independent events, *Genome Biol.* **3**, research0024.

Yang, Z., Bielawski, J.P. (2000) Statistical methods for detecting molecular adaptation, *Trends Ecol. Evol.* **15**, 496–503.

Yona, G., Linial, N., Tishby, N., Linial, M. (1998) A map of the protein space–an automatic hierarchical classification of all protein sequences, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 212–221.

Zhang, C., DeLisi, C. (1998) Estimating the number of protein folds, *J. Mol. Biol.* **284**, 1301–1305.

Zuckerkandl, E. (1975) The appearance of new structures and functions in proteins during evolution, *J. Mol. Evol.* **7**, 1–57.

Protein Splicing

Kenneth V. Mills
College of the Holy Cross, Worcester, MA

1 Discovery of Protein Splicing 243

2 Chemical Mechanism of Protein Splicing 243
2.1 N-S (or N-O) Acyl Rearrangement 243
2.2 Transesterification 245
2.3 Asparagine Cyclization Coupled to Peptide Bond Cleavage 246
2.4 Finishing Reactions 247

3 Intein Structures 247
3.1 Domain Organization 247
3.2 Three-dimensional Structures 248

4 Protein Splicing in Trans 248

5 Kinetic Studies of Protein Splicing 249

6 Relationship of Inteins with Introns and Homing Endonucleases 250

7 Relationship of Protein Splicing with Other Forms of Protein Autoprocessing 250
7.1 Bacterial Intein-like Domains 250
7.2 Hedgehog Proteins 251
7.3 Pyruvoyl Enzyme Autoprocessing 252
7.4 N-terminal Nucleophile Hydrolases 254
7.5 Nucleoporins 254

8 Inteins in Biotechnology 255
8.1 Self-cleavable Affinity Tags 255
8.2 Protein Ligation 255

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

8.3 Protein Cyclization 258
8.4 Screening Systems Utilizing Inteins 259

9 Perspective 259

Acknowledgments 260

Bibliography 260
Books and Reviews 260
Primary Literature 260

Keywords

Extein
The exteins are the flanking polypeptides interrupted by an intein.

HINT Domain
The HINT domain is the common structural folding pattern shared by **h**edgehog proteins and **int**eins.

Homing Endonuclease
Homing endonucleases recognize and cleave a target allele, facilitating the lateral transfer of the gene encoding an intein (or intron) to a homologous allele lacking the intervening sequence.

Intein
Inteins are intervening protein sequences that facilitate their own posttranslational excision from, as well as the ligation of, the flanking polypeptides.

Protein Splicing
Protein splicing is a posttranslational autoprocessing event by which an intervening protein facilitates its own excision from flanking polypeptides and the specific ligation of those polypeptides.

Trans-splicing
Protein splicing facilitated by a split intein. Intein reassociation must precede protein splicing in trans.

Protein splicing is a posttranslational autoprocessing event by which an intervening polypeptide sequence, the intein, facilitates both its own excision from flanking polypeptides, or exteins, as well as the ligation of the exteins. Since the discovery of protein splicing in 1990, significant progress has been made in elucidating the

chemical mechanism of protein splicing by both biochemical and structural studies. The more recent discoveries of inteins that can facilitate protein trans-splicing, inteins that splice via noncanonical mechanisms, and bacterial intein-like domains have provided further insight into the mechanism and evolution of protein splicing. The study of protein splicing has also deepened our understanding of related forms of protein autoprocessing, particularly the processing of hedgehog proteins. Inteins have also found wide use in biotechnology applications, including systems for protein purification, ligation, and cyclization.

1 Discovery of Protein Splicing

Protein splicing was discovered in 1990 in *Saccharomyces cerevisiae* (*Sce*). The *Sce VMA1* gene is interrupted by an intervening sequence, which is absent from homologous genes in related organisms. The split gene is transcribed into the corresponding full-length mRNA, with no spliced mRNA product detected. Equal amounts of ATPase and intein are detected using Western blot, indicating that the intervening gene is translated. Protein-splicing elements were also discovered in the *recA* gene of *Mycobacterium tuberculosis (Mtu)* and the DNA polymerase gene in *Thermococcus litoralis*.

The experiments that followed defined the essential elements for a functional intein. The fact that splicing occurs if the intein is inserted into a foreign context indicates that the intein itself contains all of the genetic information necessary for protein splicing. Inteins have homology to homing endonuclease domains similar to those found in mobile introns, although it was demonstrated by deletion analysis that the endonuclease activity is not essential for protein splicing. Essential residues were identified by analysis of the modulation of protein splicing by mutation. The nucleophilic cysteine or serine flanking the intein, as well as the conserved asparagine at the C-terminus of the intein, play a role in efficient *in vivo* protein splicing.

2 Chemical Mechanism of Protein Splicing

2.1 N-S (or N-O) Acyl Rearrangement

The first step of protein splicing involves the formation of a linear thioester or ester intermediate at the N-terminal splice junction. (See Fig. 1.) This ester intermediate is the result of an N-S or N-O acyl shift of the peptide bond involving the amino group of the N-terminal cysteine or serine. A linear thioester intermediate was first observed in a *Pyrococcus* sp. GBD Pol intein in which the N-terminal serine was replaced by cysteine. The thioester was detected by reaction with neutral hydroxylamine or ethylenediamine, which results in cleavage at the N-terminal splice junction. The splice junction was identified by analysis of the C-terminal hydroxamate residue of the N-terminal extein after hydroxylamine-induced cleavage. A thioester intermediate was also observed in the *Sce* VMA intein, which naturally contains an N-terminal cysteine.

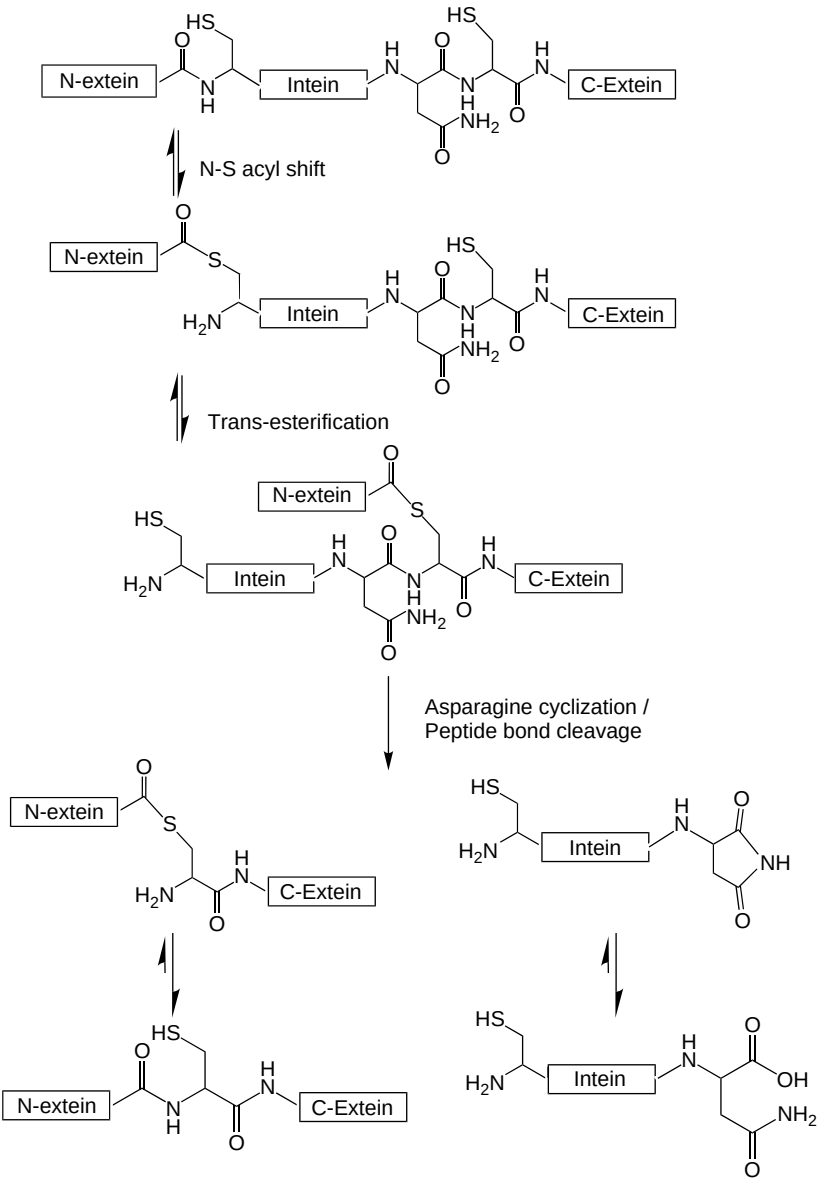

Fig. 1 Chemical mechanism of protein splicing. The nucleophilic residues are shown as cysteine in this example, although the upstream can be cysteine or serine and the downstream nucleophile can be cysteine, serine, or threonine.

Not all inteins initiate protein splicing by an N-S or N-O acyl shift of the scissile peptide bond, as inteins lacking an N-terminal nucleophile have been identified, three of which have been shown to splice. This suggests that these inteins have evolved to splice by bypassing the initial N-S or N-O shift through a direct attack by the

downstream nucleophile on the upstream scissile peptide bond. In the *Methanococcus jannaschii* KlbA intein, replacement of the N-terminal alanine with cysteine or serine yields significant amounts of spliced product, and a double mutant of the intein N-terminal alanine to cysteine and the N-terminal cysteine of the C-exteins (The term C-extein will refer to the C-terminal extein segment, and N-extein will refer to the N-terminal extein segment. Amino acid numbering begins with the first residue of the intein, such that the "N-1 residue" is the C-terminal residue of the N-extein and the "C+1" residue is the N-terminal residue of the C-extein.) to alanine can promote N-terminal cleavage by neutral hydroxylamine. The fact that this modified intein is able to facilitate an N-S acyl shift suggests that the replacement of the N-terminal nucleophilic amino acid of the intein by alanine may have occurred recently.

Uncatalyzed N-O or N-S acyl shifts have not been detected, but are known to occur in peptide bonds involving the α-nitrogen of serine or threonine in the presence of strong acid. It has been suggested that catalytic strain may serve in both accelerating the rate of this step of protein splicing as well as in shifting the equilibrium more favorably toward the ester. For instance, in the crystal structure of the *Mycobacterium xenopi* (*Mxe*) GyrA intein, the scissile bond is in a cis conformation. In addition, distortion of the scissile bond in a catalytically active *Mxe* GyrA intein is suggested by NMR data obtained using a segmental isotopically labeled intein. These data include an abnormally low one-bond dipolar coupling constant for the scissile bond and an unusually large upfield proton shift for the amide proton. Experimental evidence suggests that this bond strain is mechanistically relevant, as the identity of the N-1 residue of the *Mxe* GyrA intein has an effect on both splicing yield *in vivo* and DTT-induced N-terminal cleavage *in vitro*. In the crystal structure of the *Sce* VMA intein, the backbone τ angle of the N-1 residue is distorted by 10°. For this intein as well, a subset of mutations of the N-1 residue prevents protein splicing and DTT-induced N-terminal cleavage, and another set of mutations shifts the equilibrium of the N-S acyl shift to favor the thioester On the other hand, no distortion of the scissile bond is observed in crystal structures of nonsplicing mutants of the *Sce* VMA intein or the *Synechocystis* sp. PCC6803 (*Ssp*) DnaB intein.

2.2 Transesterification

The second step of protein splicing involves the transesterification of the linear ester by the downstream nucleophile (See Fig. 1). The result is a branched ester intermediate, first discovered in the *Psp*-GBD Pol intein. The branched intermediate was identified by reduced mobility in SDS-PAGE and by the results of sequential Edman degradation, in which two amino acids were released at each sequencing cycle, consistent with the N-terminal sequences of the N-extein and intein. Raising the pH shifts the equilibrium to the amide, and the lowering of the pH restores the branched ester, indicating full reversibility. An ester bond between the side chain of the downstream serine and the N-extein was discovered by trapping the branched ester by denaturation and cleaving it under mildly alkaline conditions. Branched thioester intermediates are also observed with inteins with cysteine at the C+1 position.

The upstream and downstream nucleophilic side chains are less than 4 Å apart in a crystal structure of a *Sce* VMA intein, suggesting facile transesterification. However, in the crystal structures of the *Sce* VMA and *Ssp* DnaB inteins, the nucleophilic side chains are separated by 8.5 and 9 Å, which suggests that either these inteins are not in an active form or require a significant conformational change to allow splicing. The downstream cysteine residue in the *Mtu* RecA intein has an unusually low pKa of 5.8, which may indicate that the transesterification can be facilitated by the unique dielectric environment of the intein. On the other hand, general base catalysis may facilitate transesterification in the *Sce* VMA intein, as the amino group of the intein N-terminal cysteine produced by the N-S shift may improve the nucleophilicity of the N-terminal thiol of the C-extein.

2.3 Asparagine Cyclization Coupled to Peptide Bond Cleavage

The third step of protein splicing is the cyclization of the C-terminal asparagine of the intein, which is coupled to cleavage of the peptide bond joining the intein and C-extein. (See Fig. 1.) Spontaneous asparagine cyclization in proteins typically results in side-chain deamidation and involves nucleophilic attack of the main-chain amide nitrogen on the side-chain carbonyl carbon to form an amidosuccinimide, followed by hydrolysis to either aspartic acid or iso-aspartic acid. An alternative mode of cyclization involves attack of the side-chain amide nitrogen on the main-chain carbonyl and leads to aminosuccinimide formation coupled to peptide bond cleavage. This may be preceded by deprotonation of the side-chain amide nitrogen, as indicated by studies of model peptides in which the rate of cleavage increases with pH. Spontaneous peptide bond cleavage via succinimide intermediates is observed in some long-lived proteins, such as α-crystallin.

The third step of protein splicing was elucidated by the isolation and analysis of C-terminal aminosuccinimide residues formed during protein splicing mediated by the *Psp*-GBD Pol intein and the *Sce* VMA intein. Branched intermediate can accumulate if the conserved asparagine at the C-terminus of the *Psp*-GBD Pol intein is mutated to aspartic acid or glutamine.

It has been proposed that the conserved penultimate histidine residue plays a role in asparagine cyclization. For instance, protein splicing is arrested and branched intermediate accumulates if the penultimate histidine in the *Psp*-GBD Pol intein is mutated to alanine. Protein splicing can be enhanced in inteins lacking a penultimate histidine by introducing a histidine residue by mutation. Structural evidence points to a role for the penultimate histidine as well. For instance, in the *Mxe* GyrA crystal structure, His197 is in a position to serve as a proton donor to the backbone amide nitrogen of the scissile bond. In the *Ssp* DnaB structure, the penultimate histidine could make a hydrogen bond with the main-chain carbonyl oxygen of the adjacent asparagine. However, at least 27 inteins lacking a penultimate histidine are known. Some of these inteins splice in the absence of the histidine residues, and in some cases splicing is impaired if the penultimate residue is replaced with histidine by mutation. It is possible that this histidine may play a role in coordinating the steps of protein splicing rather

than facilitating the asparagine cyclization directly.

Two inteins with C-terminal glutamine can facilitate splicing, as can an intein with a C-terminal aspartic acid. Although it seems likely that these inteins splice via cyclization of the C-terminal residue to a glutarimide or succinanhydride, respectively, excised inteins with cyclized C-terminal residues have not been detected, even if the C-terminal glutamine is changed to asparagine. This is surprising because other inteins with C-terminal aminosuccinimide residues have been isolated, and suggest that the noncanonical inteins may splice via an alternative mechanism.

2.4 Finishing Reactions

Asparagine cyclization leads to the excision of the intein and leaves the exteins linked by an ester. The completion of protein splicing thus requires the hydrolysis of the C-terminal aminosuccinimide residue of the intein and the conversion of the thioester or ester bond ligating the extein segments to a native peptide bond (See Fig. 1). The N-O or N-S acyl rearrangement is probably uncatalyzed because the rearrangement in synthetic depsipeptides occurs in less than a minute, much faster than the measured rates of *in vitro* splicing reactions. However, a crystal structure of a spliced *Sce* VMA intein shows that the N-terminal amino group of the intein could stabilize the oxythiazolidine intermediate in the S-N rearrangement. Enhancement of the rate of this rearrangement could prevent hydrolysis of the nascent ester-linked extein segments. The hydrolysis of the C-terminal aminosuccinimide is probably uncatalyzed as well, as the timescale of the isolation of inteins containing C-terminal aminosuccinimide is consistent with the 2.3-h half-life of the L-imide in model peptides.

3 Intein Structures

3.1 Domain Organization

Inteins can vary in size from 134 to 608 residues, and many are themselves interrupted by a homing endonuclease domain. In general, inteins contain six conserved motifs related to protein splicing, which are labeled in Fig. 2 according to the revised nomenclature of Pietrokovski. Conserved elements include an N-terminal nucleophilic residue (either Cys or Ser) in block N1, a conserved aspartic acid or glutamic acid in blocks N2 and N4, a TXXH motif in block N3, and a His-Asn sequence in block C1. Inteins are flanked at the C-terminus with a conserved cysteine, serine, or threonine residue. Although inteins can promote protein splicing in a foreign context, the efficiency of protein splicing can

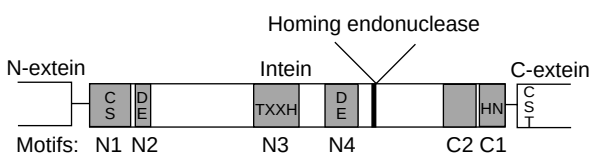

Fig. 2 Conserved elements in a typical intein. The motif nomenclature is that of Pietrokovski. Highly conserved residues are indicated in bold.

be influenced by the extein amino acid sequence. Random mutagenesis and genetic deletion studies have shown that the homing endonuclease domain is not required for protein-splicing function, and about 10% of inteins lack this central endonuclease domain, yet such mini-inteins can still promote splicing. This discovery led to the development of inteins split within the endonuclease domain that could splice in trans both *in vivo* and *in vitro*, and the discovery of a naturally split intein in *Synechocystis* sp. PCC6803.

3.2 Three-dimensional Structures

The crystal structure of the VMA intein from *S. cerevisiae* consists of two separate domains: the protein-splicing domain, which contains both termini of the protein, and the endonuclease domain. The protein-splicing domain is an elongated structure of seven β-sheets in two substructures. A crystal structure of a zinc-inactivated *Sce* VMA intein with the endonuclease domain deleted shows that the N-terminal scissile bond is in a strained trans conformation, although a later crystal structure with short N- and C-terminal exteins has a normal trans-peptide bond at both splice junctions. The *M. xenopi* GyrA intein forms a flattened "horseshoe" shape, the edges of which are composed of two long, curved antiparallel β-stands. Two three-stranded β-sheets, related by pseudo twofold symmetry, form a cleft in which the N- and C-terminal residues are separated by less than 3 Å. The N-terminal scissile bond is in a cis conformation. The crystal structure of the *Ssp* DnaB mini-intein has a similar overall folding pattern, but exhibits strain in the C-terminal scissile bond and two separate catalytic sites that may be separately responsible for the reactions at the intein N- and C-termini. The crystal structure of the Hedgehog autoprocessing domain reveals that Hedgehog domains and inteins share a common fold, as is expected from their sequence homology. This common fold is called the *HINT module* (for Hedgehog-INTein).

4 Protein Splicing in Trans

The discovery of minimal inteins provided the impetus for the search for further intein modifications that would allow protein splicing in trans (See Fig. 3). Trans-splicing was first demonstrated *in vivo* in *Escherichia coli* using a plasmid encoding the *Mtu* RecA intein in which the DNA encoding the central homing endonuclease domain was replaced by a linker containing overlapping translation stop and start codons. Protein trans-splicing *in vitro* was demonstrated using separately expressed fragments, one consisting of *E. coli* maltose-binding protein fused to the 105 N-terminal residues of the RecA intein and the other consisting of the 107 C-terminal residues of the intein fused to a domain containing a His-tag. These fragments could be reassociated by denaturation followed by joint renaturation and efficiently mediated *in vitro* protein splicing. This trans-splicing system was improved by the generation of a semisynthetic split intein by replacement of the C-terminal expressed segment of the trans-splicing intein with a synthetic polypeptide. A similar trans-splicing system was derived from the *Psp*-GBD Pol intein, although the homing endonuclease domain could not be deleted without loss of function.

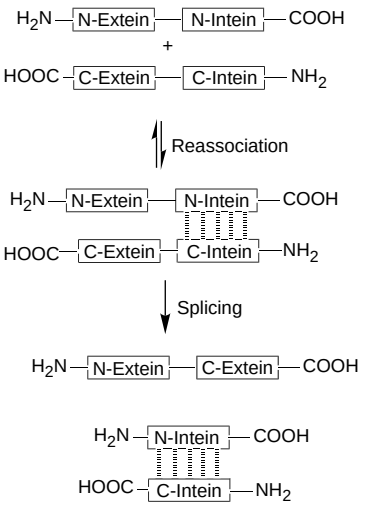

Fig. 3 Protein splicing in trans. The vertical lines are used to suggest interactions between the N- and C-intein segments.

A naturally split intein was discovered to interrupt the DnaE protein in *Synechocystis* species PCC6803. The N-terminal segment of the DnaE protein is fused to the N-intein, and the C-intein is fused to the C-terminal segment of the DnaE protein. The genes encoding these two proteins are separated by 745 kb of DNA and reside on opposite strands. The split intein can splice *in vivo* if overexpressed in *E. coli*. As a result of this discovery, an *in vitro* trans-splicing system using the DnaE intein was developed that could mediate splicing without the need for joint denaturation and renaturation, which has been used to discover zinc as an inhibitor for protein splicing, allowed for extensive kinetic analysis of protein splicing, and encouraged the development of biotechnological applications and protein-splicing screening systems that will be described below.

5 Kinetic Studies of Protein Splicing

Because inteins facilitate an autoprocessing reaction, they serve as both enzyme and substrate and are not true "catalysts." To study the kinetics of the protein-splicing reaction, one must be able to activate an intein from an inactive precursor. This was first accomplished in detail with the split *Ssp* DnaE intein. The rate constant of trans-splicing at 23 °C was 6.6×10^{-5} s^{-1}. The first step of splicing, the N-S acyl shift, was fast in comparison, with a pseudo-first-order rate constant of DTT-induced N-terminal cleavage of 1.0×10^{-3} s^{-1}. Preincubation experiments suggest that intein association is even faster. On the other hand, the rate constant of C-terminal cleavage, that is, asparagine cyclization coupled to peptide bond cleavage, was 1.9×10^{-4} s^{-1}. These data suggest that, for the *Ssp* DnaE intein, the N-S acyl rearrangement proceeds more rapidly than asparagine cyclization, which could serve to ensure that the steps of splicing proceed in order. Further detailed kinetic analysis of inteins and modified inteins could help determine how the intein catalyzes and coordinates the individual chemical steps of splicing.

6
Relationship of Inteins with Introns and Homing Endonucleases

Inteins and introns both interrupt the coding sequence of genes. As with inteins, group I and group II introns can catalyze their own excision from and the religation of flanking sequences. However, although inteins are excised posttranslationally, introns are excised at the RNA level. Like protein splicing, the splicing of introns from precursor RNA consists of a series of transesterification reactions. However, the self-splicing of introns *in vivo* is often inefficient in the absence of protein cofactors. Over 80% of inteins are interrupted by homing endonuclease domains, and many introns contain open reading frames outside of their core domains that encode homing endonucleases.

Homing endonuclease domains interrupt the majority of known inteins and are usually members of the dodecapeptide (LAGLI-DAG) family. The endonuclease domain of the *Sce* VMA intein can mediate intein homing. It cleaves at the intein insertion locus of an intein-less allele of the *Sce VMA* gene during meiosis and converts the allele lacking the intein into one containing an intein. However, the endonuclease is not required for protein-splicing activity. The crystal structure of this intein shows that it is a two-domain protein, with one domain responsible for protein-splicing activity and the other for endonuclease activity. The wide genetic distribution of mini-inteins that lack homing endonucleases suggests that the homing function may have evolved early and that modern mini-inteins lost their endonuclease domains.

The homing ability of inteins due to the endonuclease domain may explain some interesting aspects of intein distribution. For instance, few inteins are found in eubacteria aside from those in the mycobacteria and cyanobacteria. This could be explained by the fact that homing events have not been found to occur in *E. coli*. Another strange case involves the relationship of the DnaB inteins found in *Synechocystis* sp. PCC6803 and in *Rhodothermus marinus*. The DnaB inteins have 54% sequence identity and are inserted at the same site in the host protein, but the codon usage in the *R. marinus* intein is very different from the codon usage in the extein. If this is the result of homing by the same intein, it is curious how it could occur between such unrelated organisms. A third example is that the host proteins can share greater sequence similarity than the inteins. The VMA inteins in *Candida tropicalis* and *S. cerevisiae* are only 37% identical, whereas the host proteins are 88% identical. It would seem unlikely that unrelated inteins would invade the same gene if there were not some role for protein splicing that exerts genetic pressure to maintain the intein.

It is unclear if inteins or introns are more than merely parasitic DNA elements. Inteins usually interrupt conserved domains and introns are usually inserted at domain boundaries, suggesting that they might serve a regulatory function. The most likely case for possible gene regulation by an intein would seem to be the naturally split DnaE inteins, which could play a role in regulating DNA replication.

7
Relationship of Protein Splicing with Other Forms of Protein Autoprocessing

7.1
Bacterial Intein-like Domains

Bacterial intein-like domains (BILs) have been discovered in a wide range of bacterial

species. The BILs have more conserved sequence motifs than inteins, and are inserted into variable regions of their host proteins. There are two major categories of BILs: the A-type BILs contain the conserved intein sequence blocks, but mostly lack a C-terminal nucleophile, and the B-type BILs lack intein motifs C1 and C2, but usually have a nucleophilic residue at their penultimate positions.

A-type BILs have been shown to facilitate protein splicing as well as C-terminal cleavage reactions. An interesting case is an A-type BIL from *Clostridium thermocellum*, which facilitates proteins splicing, although it has a C-terminal alanine. This suggests that the BIL might splice via an alternative mechanism in which an N-S acyl rearrangement at the N-terminal splice junction and asparagine cyclization coupled to peptide bond cleavage at the C-terminal splice junction occur independently, and the free amino group of the C-extein's N-terminus then attacks the linear ester, resulting in ligation of the exteins (See Fig. 4). To date, B-type BILs have not been shown to facilitate protein splicing, but some can undergo N- and C-terminal cleavage reactions.

7.2 Hedgehog Proteins

Inteins and hedgehog proteins have many similarities, including amino acid sequence, three-dimensional structure, and catalytic function. Hedgehog proteins are involved in cell patterning during metazoan development, and disruption of the hedgehog-signaling pathway can lead to developmental disorders and tumorigenesis. They are found in almost all metazoans, whereas inteins are found only in unicellular organisms.

Hedgehog proteins, like inteins, undergo posttranslational autoprocessing. The *Drosophila* hedgehog protein was first shown to generate two protein domains from one larger protein precursor. The mechanism of this self-cleavage is similar to that of intein-mediated protein splicing (See Fig. 5). First, an N-S acyl rearrangement of the peptide bond between a conserved Gly-Cys sequence results in a thioester intermediate. Next, a transesterification between the 3β-hydroxyl of cholesterol and the newly formed thioester results in peptide bond cleavage and conjugation of the upstream fragment with cholesterol. In analogy with the role of exteins in protein splicing, the upstream signaling fragment is not required for autoprocessing, as hedgehog self-cleavage can occur in a foreign protein context.

HINT domains also occur in *Caenorhabditis elegans*, and are called *Groundhog* and *Warthog*. Although one of these HINT domains undergoes cleavage near the expected site if expressed in cultured *Drosophila* cells, the polypeptides flanking the HINT domain are dissimilar from those in the hedgehog proteins, which suggests that they may function differently in autoprocessing and cell signaling.

It is likely that the hedgehog proteins and inteins share a common ancestor. There is significant amino acid sequence homology between inteins and the autoprocessing domain of hedgehog proteins. In addition, the *Drosophila* hedgehog protein shares structural homology with intein domains. It has been suggested that inteins evolved from this common ancestor and propagated as a result of the invasion of a homing endonuclease. The hedgehog proteins may have lost the ability to splice

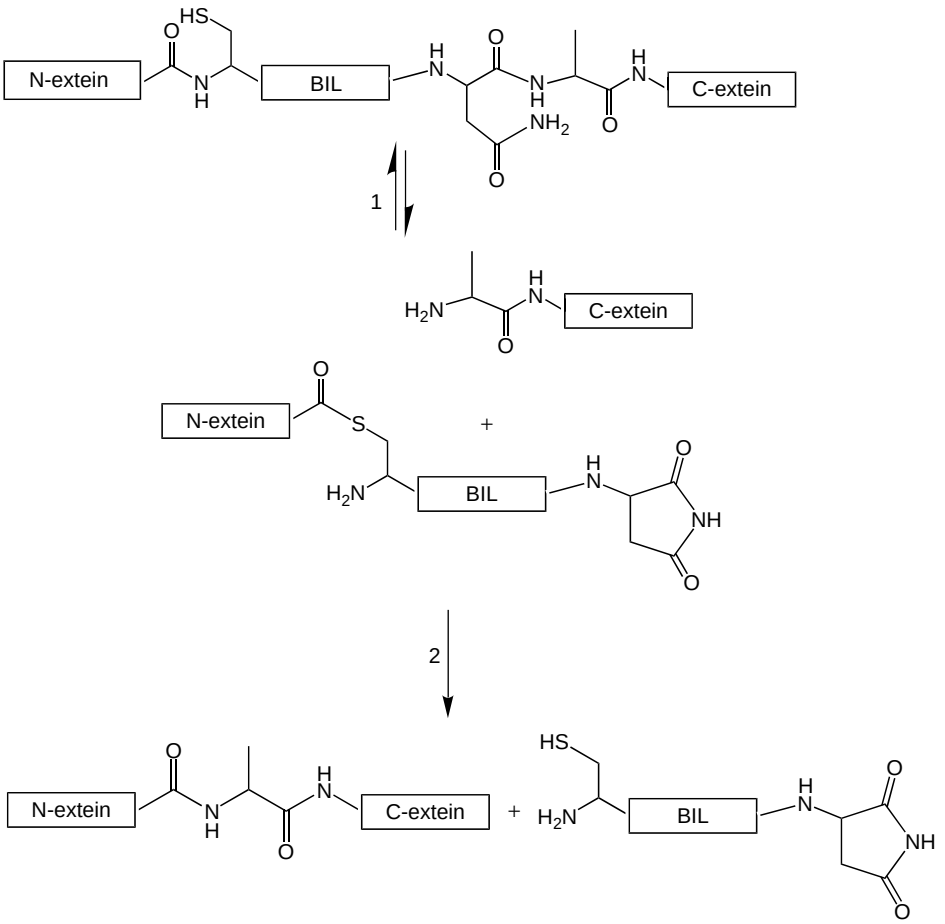

Fig. 4 Proposed alternate mechanism for splicing of a BIL without a C-terminal nucleophile.

while retaining the ability to facilitate an N-S acyl shift.

7.3 Pyruvoyl Enzyme Autoprocessing

A class of decarboxylases and reductases use a covalently linked pyruvoyl moiety as a prosthetic group instead of pyridoxal phosphate cofactors. In both histidine decarboxylase and aspartate decarboxylase, the pyruvoyl moiety is the result of self-catalyzed processing. Although there is no sequence or structural homology to inteins, the reaction is initiated by an N-O acyl rearrangement (See Fig. 5). A vestige of this reaction is seen in the crystal structure of *E. coli* aspartate decarboxylase. The precursor protein, π, is a tetramer, but only three subunits undergo processing. Autoprocessing is initiated by an N-O acyl rearrangement of the peptide bond between Gly24 and Ser25. The presence of an ester intermediate is suggested by the electron density of the unprocessed subunit.

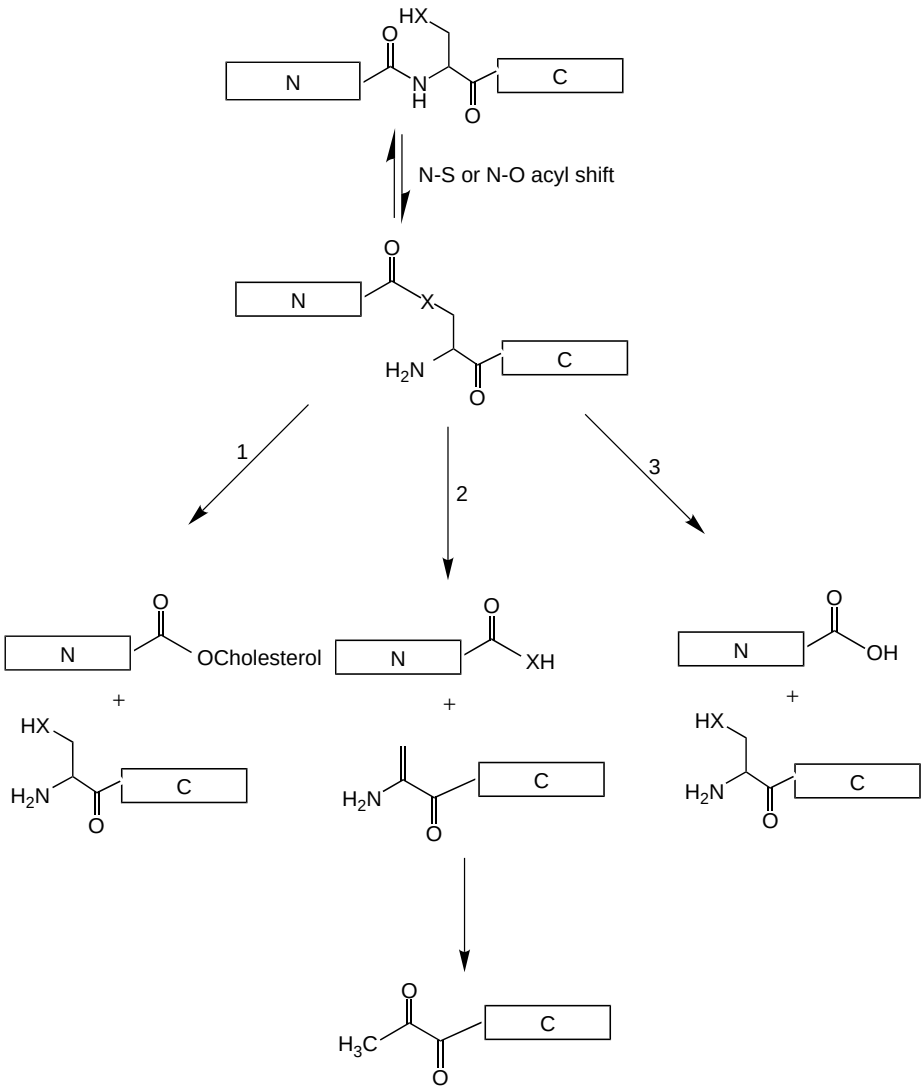

Fig. 5 Related forms of autoprocessing. Following an N-O or N-S acyl shift, the ester can be resolved by (1) transesterification with cholesterol as in hedgehog autoproteolysis; (2) β-elimination following by hydrolysis, as in pyruvoyl enzyme formation; or (3) hydrolysis as in N-terminal nucleophile or nucleoporin autocleavage. "N" and "C" in the boxes indicate the N- and C-terminal domains respectively.

Instead of the second transesterification reaction of inteins or hedgehog proteins, the breakdown of the ester is facilitated by β-elimination, followed by hydration to a carbinolamine and elimination of ammonia to form an N-terminal pyruvoyl group. As with some inteins, the position of the N-O equilibrium may be affected by catalytic strain, in this case by the presence of a tight

turn immediately preceding the scissile bond.

7.4 N-terminal Nucleophile Hydrolases

N-terminal nucleophile hydrolases, such as penicillin acylase, glutamine PRPP amidotransferase, and the proteosome are activated by cleavage of the peptide bond upstream of the catalytic residue to generate a catalytic subunit from a larger, inactive precursor. Direct evidence for intramolecular processing in this class of structurally related enzymes exists for glycosylasparaginase from *Flavobacterium meningosepticum* (*Fme*). Active glycosylasparaginase hydrolyzes the β-*N*-glycosidic bond between asparagine and *N*-acetylglucosamine in asparagine-linked glycans. Activation of the enzyme initiates with an N-O acyl rearrangement. In place of the transesterification reaction found with inteins, the ester is hydrolyzed to form two protein segments. The discovery that free glycine inhibits autoprocessing allowed the detection of a thioester intermediate by cleavage with neutral hydroxylamine in a Thr1Cys glycosylasparaginase mutant. Although there is no structural or evolutionary relationship between inteins and glycosylasparaginase, it is possible that glycosylasparaginase has also adopted conformational strain as a means to promote an amide to ester rearrangement. The crystal structure of the *Fme* glycosylasparaginase reveals a tight turn involving the scissile bond, resulting in a deviation from planarity of the peptide bond immediately downstream of the scissile bond by more than 20°, as well as deviation of the backbone τ angles by more than 9°. The distortion is stabilized by interaction of the side chain of the N-1 aspartic acid residue (Asp151) with Thr203 and the catalytic Thr152. Perhaps the release of this steric strain favors formation of the ester intermediate.

7.5 Nucleoporins

Another example of protein autoprocessing involves the self-catalyzed cleavage of the yeast nucleoporin Nup145. The nuclear pore complex is responsible for many aspects of cellular function, including the exchange of protein and RNA between the nucleoplasm and the cytosol. Nup145p is a 145-kDa protein that is posttranslationally processed into a 65-kDa N-terminal domain and an 80-kDa C-terminal domain. The C-terminal domain functions as part of a nucleoporin complex that controls mRNA export and nuclear architecture. The function of the N-terminal fragment is not clear, but it is improperly localized if separately expressed from the C-terminal fragment. Thus, Nup145p autoprocessing plays a regulatory role in the cellular localization of essential proteins. The autoprocessing is initiated by an N-O acyl rearrangement that results in hydrolysis of the peptide bond between phenylalanine and serine in a conserved His-Phe-Ser motif (See Fig. 5). Deletion of the first 712 residues of Nup145p does not eliminate autoprocessing, which suggests that the information required for processing is located mostly in the C-terminal domain. In the crystal structure of human nucleoporin Nup98, hydrogen bonding is evident between the histidine and serine of the conserved triad. Although there is no evidence of catalytic strain, there is tight anchoring of the side chain of the conserved phenylalanine of the conserved triad in a hydrophobic pocket as well as β-strand interactions downstream of the

scissile bond, which may suggest the use of constrained geometry in the promotion of the amide–ester rearrangement. Although nucleoporin autoprocessing and protein splicing share the same first step, there is no structural or sequence homology between these autoprocessing domains.

8
Inteins in Biotechnology

8.1
Self-cleavable Affinity Tags

One of the first applications developed that takes advantage of protein-splicing elements was the use of inteins modulated by mutation for use in protein purification. Protein purification is often facilitated by the use of affinity tags, but it is sometimes necessary to remove these domains from the protein of interest. Traditionally, this is done using a protease that must then be separated from the target protein. However, if the affinity tag is fused to an intein, it is possible to cleave the intein-affinity tag fusion from the target protein, using either N-terminal or C-terminal cleavage uncoupled from the overall protein-splicing reaction (See Fig. 6).

The first affinity tag purification scheme developed uses a modified form of the *Sce* VMA intein that can be induced to cleave at the intein N-terminus. Precursor protein accumulates if the C-terminal intein asparagine residue is mutated to alanine, preventing the final step of splicing. If the target protein is used as the N-extein and an affinity tag as the C-extein, the precursor can be purified using an affinity resin. The protein target can then be eluted by initiation of N-terminal cleavage by added thiol. If cysteine is used, it can be incorporated into the target protein at its C-terminus by essentially irreversible S-N acyl rearrangement after cleavage. Protocols using mini-inteins can allow purification of larger target proteins. N-terminal cleavage of the *Ssp* DnaB mini-intein can be induced by raising the pH to 8.0 or the temperature to 25 °C, eliminating the need for added thiols.

Alternatively, the target protein can be fused to the intein C-terminus, and the N-extein can be used as the affinity tag. Purification is accomplished by inducing C-terminal cleavage via asparagine cyclization. One would expect higher levels of expression of soluble fusion proteins in such a system, in which a soluble *E. coli* protein such as maltose-binding protein can be used as the N-terminal domain. The first such system described uses a *Sce* VMA intein modified by mutation such that C-terminal cleavage only occurs after N-terminal cleavage is induced by the addition of exogenous thiol. However, in this case the N-extein peptide coelutes with the C-extein target and added thiol. Later systems use modified inteins for which C-terminal cleavage does not require prior cleavage at the intein N-terminus and can be induced by a change in pH or temperature.

8.2
Protein Ligation

Protein semisynthesis in aqueous solution in the absence of enzymes has been a long-standing goal in synthetic chemistry. Two methods of protein semisynthesis are native chemical ligation and expressed protein ligation.

The native chemical ligation scheme developed by Tam and coworkers involves the ligation of a peptide with a C-terminal thioester to a peptide with an N-terminal

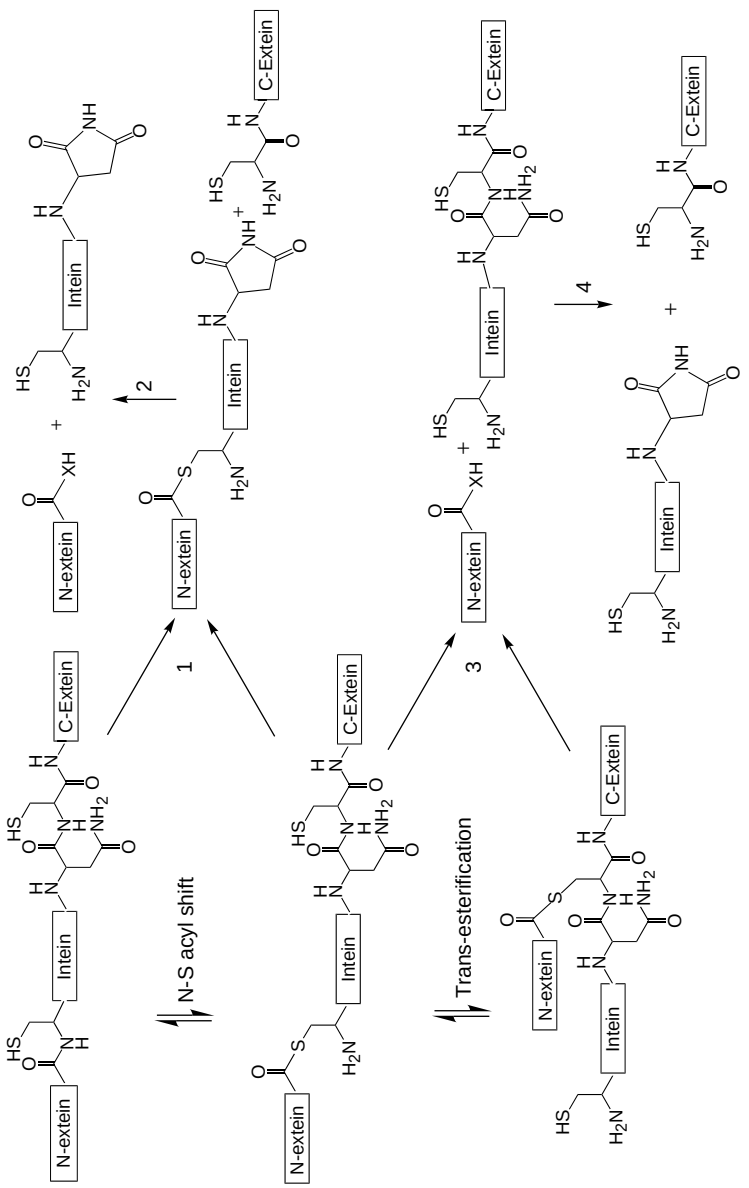

Fig. 6 Side reactions that are manipulated in intein-based protein purification and ligation systems. (1) Premature asparagine cyclization of the precursor or linear ester can lead to C-terminal cleavage, possibly followed by (2) hydrolysis or thiolysis at the N-terminal splice junction. (3) Hydrolysis or thiolysis of the linear or branched esters can lead to N-terminal cleavage, possibly followed by (4) cyclization of asparagine at the C-terminal splice junction coupled to peptide bond cleavage.

cysteine (mediated by a trialkylphosphine and an alkylthiol), or the ligation of a peptide with a C-terminal thiocarboxylic acid to a peptide with an N-terminal β-bromo amino acid at low pH. Another method, developed by Kent and coworkers, ligates a peptide with a C-terminal thioester and a peptide with an N-terminal cysteine and proceeds in the presence of Ellman's reagent. However, the bimolecular nature of both methods results in a relatively slow rate of ligation. The methods also require a cysteine residue at the junction of the ligated peptides as well as the synthesis of the N-terminal segment, which limits the size of the protein target. However, other methods avoid the requirement for the cysteine residue. For instance, in "conformationally assisted protein ligation," the ligation of ribonuclease S with the S-peptide between alanine and serine is facilitated by the prior refolding of the two fragments. Native chemical ligation between a thioester and a 2-mercaptobenzylamine-linked peptide results in a peptide bond without the need for cysteine, albeit with *o*-benzylmercaptoamine in the product. Another method uses a phosphinobenzenethiol to facilitate native chemical ligation between a peptide with a C-terminal thioester and a peptide with an N-terminal azide, yielding a native peptide with no requirement for cysteine or affinity between the peptide targets. However, this method remains limited to the ligation of small peptides because of the need to synthesize the thioester-containing peptide.

The semisynthesis of proteins with building blocks generated using inteins presents a great advantage over the synthetic peptides used in native chemical ligation procedures. This "expressed protein ligation" was first described using the *Sce* VMA intein with the target protein as the N-extein. The target protein is purified with a C-terminal thioester via N-terminal cleavage of the intein induced by thiophenol. It is then ligated to a peptide with an amino-terminal cysteine. A more efficient procedure was developed using the *Mxe* GyrA mini-intein, using 2-mercaptoethanesulfonic acid to induce N-terminal cleavage. Expressed protein ligation methods of this kind also employ the *Mth* RIR1 mini-intein and the *Ssp* DnaB mini-intein. However, these methods require the use of a synthetic peptide containing an N-terminal cysteine. Two procedures were developed to avoid the need for a synthetic peptide. In one, factor Xa is used to cleave a target peptide upstream of a cysteine residue. The second uses controlled intein C-terminal cleavage to generate a cleaved C-extein with an N-terminal cysteine. Recent applications include intein-mediated biotinylation of proteins, conjugation of an oligonucleotide to a recombinant protein, and semisynthesis of a protein with β-amino acids.

Protein splicing in trans is an essentially unimolecular method of protein ligation. It was originally used to ligate maltose-binding protein to paramyosin or a tetrapeptide using the *Psp*-GBD-Pol intein or to a His-tag-containing domain using the *Mtu* RecA intein. Using a synthetic peptide as the C-terminal fragment of the *Mtu* RecA intein, peptides containing synthetic probes or nonnatural amino acids can be ligated to target proteins. However, these methods are limited by the need for joint denaturation and renaturation. The discovery of a naturally split intein, the *Ssp* DnaE intein, opened the possibility of protein ligation *in vitro* by protein trans-splicing without prior denaturation, and can be applied to proteins that cannot be reversibly denatured. The discovery of "conditional" protein splicing

in trans, by which association and therefore subsequent splicing of the split *Sce* VMA intein is induced by the fusion of complementary components of a protein dimerization domain to the intein segments, could allow regulation of protein ligation *in vitro* and *in vivo*.

The discovery of intein-mediated protein ligation in trans has led to a number of interesting applications. For instance, each fragment of a split *Sce* VMA intein was fused to a fragment of green fluorescent protein (GFP). Although the intein fragments interact only weakly, if fused to proteins that interact strongly, they associate properly and splice to yield functional GFP. This system was used *in vivo* to measure protein–protein interactions. In another example, split *Ssp* DnaE fragments, each fused to fragments of 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), yields a reconstituted, although not spliced, enzyme. The reconstituted enzyme rescues the growth of *E. coli* cells with a depletion of the *EPSPS* gene if cultured in the presence of glyphosate. Herbicide-resistant acetolactate synthase was also split and fused to the two segments of the *Ssp* DnaE intein, which mediates splicing of the acetolactate synthase as well as resistance to valine growth inhibition in *E. coli*. Trans-splicing of EPSPS mediated by the split *Ssp* DnaE intein is also observed in the chloroplast of *Nicotiana tabacum* when a gene encoding for N-terminal intein-EPSPS fusion was integrated into the nuclear DNA and a gene encoding for the C-terminal intein-EPSPS fusion was integrated into the chloroplast genome. These transgenic plants also show an increased resistance to glyphosate. Trans-splicing of split β-glucuronidase in *Arabidopsis* using the *Ssp* DnaE intein has also been observed.

8.3 Protein Cyclization

An interesting use of inteins involves the cyclization of proteins by native chemical ligation, expressed protein ligation, and protein trans-splicing. Cyclized proteins are resistant to exoproteases and useful in the study of protein folding dynamics.

Two methods for protein cyclization have been developed on the basis of native chemical ligation. In one, an unprotected peptide is synthesized with an N-terminal cysteine and a C-terminal thioester on a Boc-Gly-aminomethyl-polystyrene resin with a thioester "handle." The thioester is derivatized with Ellman's reagent and the cyclic peptide is formed in the presence of thiophenol. Another method involves the cyclization of the 47-residue WW domain from human Yes kinase-associated protein. The protein was synthesized with a C-terminal thioester and an N-terminal cysteine, and allowed to fold. The cyclization proceeds whether the protein is folded or not, but the cyclization is an order of magnitude faster when the protein is folded because the N- and C-termini of the protein are in closer proximity.

The size limitations of protein cyclization are largely eliminated by the development of expressed protein ligation protocols. One group fused the N-terminus of a target protein to a modulated *Ssp* DnaB intein and the C-terminus of the target protein to the *Mxe* GyrA intein. Cleavage at the N-terminus of the target protein is initiated by pH-induced asparagine cyclization of the DnaB intein. This is followed by cleavage at the C-terminus of the target protein by addition of 2-mercaptoethanesulfonic acid, resulting in a target protein with an N-terminal cysteine and a C-terminal thioester, which

can either cyclize or polymerize. In another method, an SH2 domain is cyclized by fusion to the N-terminus of the *Sce* VMA intein. An N-terminal cysteine is revealed by factor Xa cleavage and reacts with the thioester at the N-terminal splice junction to liberate cyclized protein. A similar method was used to cyclize β-lactamase that is more stable to heat denaturation than the linear form.

Protein cyclization based on trans-splicing was first developed using the naturally split *Ssp* DnaE intein *in vivo*. The C-terminal segment of the DnaE intein is fused to the target protein at its N-terminus, and the N-terminal segment of the intein is fused to the target at its C-terminus, such that protein splicing yields a cyclized product. This method is called *SICLOPPS*, or split intein-mediated circular ligation of peptides and proteins. A similar system was used to circularize maltose-binding protein *in vivo* as well as *in vitro*. A split version of the *Pfu* Pol1 intein can cyclize GFP *in vivo*, which denatures at half of the rate of uncyclized GFP. Intein-mediated cyclization can be used to generate cyclic peptide libraries in cells to screen for bioactive compounds.

8.4 Screening Systems Utilizing Inteins

Inteins have been used in the design of genetic screening systems. In one such system, a plasmid-carried gene encoding for aminoglycoside phosphotransferase (aph) was interrupted with the *Mtu* RecA intein, such that cell growth in the presence of kanamycin is dependent on efficient protein splicing. This aph-intein fusion was used to screen for open reading frame insertions into the DNA sequence coding for the intein. Belfort and coworkers created a screening system based on the splicing of thymidylate synthase (TS) interrupted by the *Mtu* RecA intein. In *thyA E. coli*, one can select for protein splicing by growth in the absence of thymine, which would require protein splicing to generate a complementary TS, and select against protein splicing in the presence of thymine and a DHFR inhibitor, which would be lethal in the presence of spliced, complementary TS. This system was used to define the minimal size of a functional intein and to select for an intein that can facilitate pH-dependent C-terminal cleavage.

Temperature-sensitive intein mutants were selected using a system based on the *Mxe* GyrA intein. Splicing of the split GyrA protein results in cell death in the presence of quinoline-related gyrase inhibitors, so this system could be used to select for inhibitors of protein splicing by coupling cell survival to efficient splicing. Protein-splicing inhibitors could also be detected using a system based on the cytotoxicity of the ccdB protein. N-terminal fusion of lacZα to the cell death domain ccdB does not emeliorate the cytotoxicity of ccdB, but the insertion of the *Mtu* RecA intein between lacZα and ccdB does ablate ccdB function. Therefore, inhibition of protein splicing can be coupled to cell survival. Finally, the fluorescence of GFP interrupted by the *Mtu* RecA intein is dependent on removal of the intein by protein splicing, and this could be used as a colorimetric assay for protein-splicing inhibition.

9 Perspective

Protein splicing was first discovered in 1990. Since that time, the chemical steps

of this process have been defined, and a variety of clever and useful biotechnology applications have been developed. However, many questions and interesting avenues of inquiry remain. For instance, the means by which inteins facilitate and coordinate each step of splicing remain largely unanswered. It is also unclear if inteins are merely molecular parasites, or if they play a role in the regulation of gene expression. If so, inhibitors of protein splicing could be designed and used to dissect this biological role. Inteins could also serve as drug targets, as one could imagine that the inactivation of such interrupted exteins as the RecA or DnaB proteins of *Mtu* could have antimicrobial effects. The majority of the studies of protein splicing have been done with foreign proteins as the exteins and *in vitro* or in *E. coli*, which does not have inteins, so it will be interesting to study the splicing of inteins in their natural hosts and extein contexts. Finally, it is interesting that inteins and hedgehog domains have considerable sequence and structural homology, but that inteins are found only in unicellular organisms and do not have an identified function, whereas hedgehog proteins are found in most metazoans and play an essential role in embryonic development. Perhaps the recently discovered bacterial intein-like domains will serve as an evolutionary link between these autoprocessing domains.

Acknowledgments

I gratefully acknowledge Professor Henry Paulus of the Boston Biomedical Research Institute for his critical reading of the manuscript.

See also Protein and Nucleic Acid Enzymes; Mass Spectrometry-based Methods of Proteome Analysis; Proteomics.

Bibliography

Books and Reviews

Belfort, M., Reaban, M.E., Coetzee, T., Dalgaard, J.Z. (1995) Prokaryotic introns and inteins: a panoply of form and function, *J. Bacteriol.* **177**, 3897–3903.

Blaschke, U.K., Silberstein, J., Muir, T.W. (2000) Protein engineering by expressed protein ligation, *Methods Enzymol.* **328**, 478–496..

Cotton, G.J., Muir, T.W. (1999) Peptide ligation and its application to protein engineering, *Chem. Biol.* **6**, R247–R256.

Derbyshire, V., Belfort, M. (1998) Lightning strikes twice: intron-intein coincidence, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 1356–1357.

Evans, T.C., Xu, M.Q. Jr. (1999) Intein-mediated protein ligation: harnessing nature's escape artists, *Biopolymers* **51**, 333–342.

Giriat, I., Muir, T.W., Perler, F.B. (2001) Protein splicing and its applications, *Genet. Eng. (N. Y)*. **23**, 171–199.

Lambowitz, A.M., Belfort, M. (1993) Introns as mobile genetic elements, *Annu. Rev. Biochem.* **62**, 587–622.

Liu, X.Q. (2000) Protein-splicing intein: Genetic mobility, origin, and evolution, *Annu. Rev. Genet.* **34**, 61–76.

Mann, R.K., Beachy, P.A. (2000) Cholesterol modification of proteins, *Biochim. Biophys. Acta* **1529**, 188–202.

Paulus, H. (2000) Protein splicing and related forms of protein autoprocessing, *Annu. Rev. Biochem.* **69**, 447–496.

Paulus, H. (2003) Inteins as targets for potential antimycobacterial drugs, *Front. Biosci.* **8**, s1157–s1165.

Primary Literature

Adam, E., Perler, F.B. (2002) Development of a positive genetic selection system for inhibition of protein splicing using mycobacterial inteins

in Escherichia coli DNA gyrase subunit A, *J. Mol. Microbiol. Biotechnol.* **4**, 479–487.

Albert, A., Dhanaraj, V., Genschel, U., Khan, G., Ramjee, M.K., Pulido, R., Sibanda, B.L., von Delft, F., Witty, M., Blundell, T.L., Smith, A.G., Abell, C. (1998) Crystal structure of aspartate decarboxylase at 2.2 A resolution provides evidence for an ester in protein self-processing, *Nat. Struct. Biol.* **5**, 289–293.

Amitai, G., Dassa, B., Pietrokovski, S. (2004) Protein splicing of inteins with atypical glutamine and aspartate C-terminal residues, *J. Biol. Chem.* **279**, 3121–3131.

Amitai, G., Belenkiy, O., Dassa, B., Shainskaya, A., Pietrokovski, S. (2003) Distribution and function of new bacterial intein-like protein domains, *Mol. Microbiol.* **47**, 61–73.

Arnold, U., Hinderaker, M.P., Nilsson, B.L., Huck, B.R., Gellman, S.H., Raines, R.T. (2002) Protein prosthesis: a semisynthetic enzyme with a beta-peptide reverse turn, *J. Am. Chem. Soc.* **124**, 8522–8523.

Aspock, G., Kagoshima, H., Niklaus, G., Burglin, T.R. (1999) Caenorhabditis elegans has scores of hedgehog-related genes: sequence and expression analysis, *Genome Res.* **9**, 909–923.

Beligere, G.S., Dawson, P.E. (1999) Conformationally Assisted Protein Ligation Using C-Terminal Thioester Peptides, *J. Am. Chem. Soc.* **121**, 6332–6333.

Camarero, J.A., Muir, T.W. (1997) Chemoselective backbone cyclization of unprotected peptides, *Chem. Commun.* **15**, 1369–1370.

Camarero, J.A., Muir, T.W. (1999) Biosynthesis of a head-to-tail cyclized protein with improved biological activity, *J. Am. Chem. Soc.* **121**, 5597–5598.

Camarero, J.A., Pavel, J., Muir, T.W. (1998) Chemical synthesis of a circular protein domain: Evidence for Folding-assisted cyclization, *Angew. Chem. Int. Ed. Engl.* **37**, 5597–5598.

Capasso, S., Mazzarella, L., Sorrentino, G., Balboni, G., Kirby, A.J. (1996) Kinetics and mechanism of the cleavage of the peptide bond next to asparagine, *Peptides* **17**, 1075–1077.

Caspi, J., Amitai, G., Belenkiy, O., Pietrokovski, S. (2003) Distribution of split DnaE inteins in cyanobacteria, *Mol. Microbiol.* **50**, 1569–1577.

Chen, L., Benner, J., Perler, F.B. (2000) Protein splicing in the absence of an intein penultimate histidine, *J. Biol. Chem.* **275**, 20431–20435.

Chen, L., Pradhan, S., Evans, T.C. Jr. (2001) Herbicide resistance from a divided EPSPS protein: the split Synechocystis DnaE intein as an in vivo affinity domain, *Gene* **263**, 39–48.

Chin, H.G., Kim, G.D., Marin, I., Mersha, F., Evans, T.C., Chen, L., Xu, M.Q., Pradhan, S. Jr. (2003) Protein trans-splicing in transgenic plant chloroplast: reconstruction of herbicide resistance from split genes, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 4510–4515.

Chong, S., Xu, M.Q. (1997) Protein splicing of the Saccharomyces cerevisiae VMA intein without the endonuclease motifs, *J. Biol. Chem.* **272**, 15587–15590.

Chong, S., Williams, K.S., Wotkowicz, C., Xu, M.Q. (1998) Modulation of protein splicing of the Saccharomyces cerevisiae vacuolar membrane ATPase intein, *J. Biol. Chem.* **273**, 10567–10577.

Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F.B., Xu, M.Q. (1996) Protein splicing involving the Saccharomyces cerevisiae VMA intein. The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system, *J. Biol. Chem.* **271**, 22159–22168.

Chong, S., Montello, G.E., Zhang, A., Cantor, E.J., Liao, W., Xu, M.Q., Benner, J. (1998) Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step, *Nucleic Acids Res.* **26**, 5109–5115.

Chong, S., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner, J., Kucera, R.B., Hirvonen, C.A., Pelletier, J.J., Paulus, H., Xu, M.Q. (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element, *Gene* **192**, 271–281.

Clarke, S. (1987) Propensity for spontaneous succinimide formation from aspartyl and asparaginyl residues in cellular proteins, *Int. J. Pept. Protein Res.* **30**, 808–821.

Cooper, A.A., Chen, Y.J., Lindorfer, M.A., Stevens, T.H. (1993) Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision, *EMBO J.* **12**, 2575–2583.

Dassa, B., Haviv, H., Amitai, G., Pietrokovski, S. (2004) Protein splicing and auto-cleavage of bacterial intein-like domains lacking a C'-flanking nucleophilic residue, *J. Biol. Chem.* **279**, 32001–32007.

Daugelat, S., Jacobs, W.R. Jr. (1999) The Mycobacterium tuberculosis recA intein can be used in an ORFTRAP to select for open reading frames, *Protein Sci.* **8**, 644–653.

Davis, E.O., Sedgwick, S.G., Colston, M.J. (1991) Novel structure of the recA locus of Mycobacterium tuberculosis implies processing of the gene product, *J. Bacteriol.* **173**, 5653–5662.

Davis, E.O., Jenner, P.J., Brooks, P.C., Colston, M.J., Sedgwick, S.G. (1992) Protein splicing in the maturation of M. tuberculosis recA protein: a mechanism for tolerating a novel class of intervening sequence, *Cell* **71**, 201–210.

Dawson, P.E., Muir, T.W., Clark-Lewis, I., Kent, S.B. (1994) Synthesis of proteins by native chemical ligation, *Science* **266**, 776–779.

Derbyshire, V., Wood, D.W., Wu, W., Dansereau, J.T., Dalgaard, J.Z., Belfort, M. (1997) Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 11466–11471.

Ding, Y., Xu, M.Q., Ghosh, I., Chen, X., Ferrandon, S., Lesage, G., Rao, Z. (2003) Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing, *J. Biol. Chem.* **278**, 39133–39142.

Duan, X., Gimble, F.S., Quiocho, F.A. (1997) Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity, *Cell* **89**, 555–564.

Evans, T.C., Benner, J., Xu, M.Q. Jr. (1998) Semisynthesis of cytotoxic proteins using a modified protein splicing element, *Protein Sci.* **7**, 2256–2264.

Evans, T.C., Benner, J., Xu, M.Q. Jr. (1999) The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins, *J. Biol. Chem.* **274**, 18359–18363.

Evans, T.C., Benner, J., Xu, M.Q. Jr. (1999) The in vitro ligation of bacterially expressed proteins using an intein from Methanobacterium thermoautotrophicum, *J. Biol. Chem.* **274**, 3923–3926.

Evans, T.C., Xu, M.Q. Jr. (1999) Intein-mediated protein ligation: harnessing nature's escape artists, *Biopolymers* **51**, 333–342.

Evans, T.C., Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, I., Chen, L., Benner, J., Liu, X.Q., Xu, M.Q. Jr. (2000) Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of Synechocystis species PCC6803, *J. Biol. Chem.* **275**, 9091–9094.

Gangopadhyay, J.P., Jiang, S.Q., Paulus, H. (2003) An in vitro screening system for protein splicing inhibitors based on green fluorescent protein as an indicator, *Anal. Chem.* **75**, 2456–2462.

Geiger, T., Clarke, S. (1987) Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation, *J. Biol. Chem.* **262**, 785–794.

Ghosh, I., Sun, L., Xu, M.Q. (2001) Zinc inhibition of protein trans-splicing and identification of regions essential for splicing and association of a split intein*, *J. Biol. Chem.* **276**, 24051–24058.

Gimble, F.S., Thorner, J. (1992) Homing of a DNA endonuclease gene by meiotic gene conversion in Saccharomyces cerevisiae, *Nature* **357**, 301–306.

Gorbalenya, A.E. (1998) Noncanonical inteins, *Nucleic Acids Res.* **26**, 1741–1748.

Gu, H.H., Xu, J., Gallagher, M., Dean, G.E. (1993) Peptide splicing in the vacuolar ATPase subunit A from Candida tropicalis, *J. Biol. Chem.* **268**, 7372–7381.

Guan, C., Cui, T., Rao, V., Liao, W., Benner, J., Lin, C.L., Comb, D. (1996) Activation of glycosylasparaginase. Formation of active N-terminal threonine by intramolecular autoproteolysis, *J. Biol. Chem.* **271**, 1732–1737.

Guan, C., Liu, Y., Shao, Y., Cui, T., Liao, W., Ewel, A., Whitaker, R., Paulus, H. (1998) Characterization and functional analysis of the cis-autoproteolysis active center of glycosylasparaginase, *J. Biol. Chem.* **273**, 9695–9702.

Hall, T.M., Porter, J.A., Young, K.E., Koonin, E.V., Beachy, P.A., Leahy, D.J. (1997) Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins, *Cell* **91**, 85–97.

Hirata, R., Anraku, Y. (1992) Mutations at the putative junction sites of the yeast VMA1 protein, the catalytic subunit of the vacuolar membrane H(+)-ATPase, inhibit its processing by protein splicing, *Biochem. Biophys. Res. Commun.* **188**, 40–47.

Hirata, R., Ohsumk, Y., Nakano, A., Kawasaki, H., Suzuki, K., Anraku, Y. (1990) Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of Saccharomyces cerevisiae, *J. Biol. Chem.* **265**, 6726–6733.

Hodel, A.E., Hodel, M.R., Griffis, E.R., Hennig, K.A., Ratner, G.A., Xu, S., Powers, M.A. (2002) The three-dimensional structure of the autoproteolytic, nuclear pore-targeting domain of the human nucleoporin Nup98, *Mol. Cell.* **10**, 347–358.

Hodges, R.A., Perler, F.B., Noren, C.J., Jack, W.E. (1992) Protein splicing removes intervening sequences in an archaea DNA polymerase, *Nucleic Acids Res.* **20**, 6153–6157.

Iwai, H., Pluckthun, A. (1999) Circular beta-lactamase: stability enhancement by cyclizing the backbone, *FEBS Lett.* **459**, 166–172.

Iwai, H., Lingel, A., Pluckthun, A. (2001) Cyclic green fluorescent protein produced in vivo using an artificially split PI-PfuI intein from Pyrococcus furiosus, *J. Biol. Chem.* **276**, 16548–16554.

Iwai, K., Ando, T. (1967) N-O Acyl Shifts, *Methods Enzymol.* **11**, 263–282.

Kane, P.M., Yamashiro, C.T., Wolczyk, D.F., Neff, N., Goebl, M., Stevens, T.H. (1990) Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase, *Science* **250**, 651–657.

Kinsella, T.M., Ohashi, C.T., Harder, A.G., Yam, G.C., Li, W., Peelle, B., Pali, E.S., Bennett, M.K., Molineaux, S.M., Anderson, D.A., Masuda, E.S., Payan, D.G. (2002) Retrovirally delivered random cyclic Peptide libraries yield inhibitors of interleukin-4 signaling in human B cells, *J. Biol. Chem.* **277**, 37512–37518.

Klabunde, T., Sharma, S., Telenti, A., Jacobs, W.R., Sacchettini, J.C. Jr. (1998) Crystal structure of GyrA intein from Mycobacterium xenopi reveals structural basis of protein splicing, *Nat. Struct. Biol.* **5**, 31–36.

Lee, J.J., Ekker, S.C., von Kessler, D.P., Porter, J.A., Sun, B.I., Beachy, P.A. (1994) Autoproteolysis in hedgehog protein biogenesis, *Science* **266**, 1528–1537.

Lesaicherre, M.L., Lue, R.Y., Chen, G.Y., Zhu, Q., Yao, S.Q. (2002) Intein-mediated biotinylation of proteins and its application in a protein microarray, *J. Am. Chem. Soc.* **124**, 8768–8769.

Lew, B.M., Paulus, H. (2002) An in vivo screening system against protein splicing useful for the isolation of nonsplicing mutants or inhibitors of the RecA intein of Mycobacterium tuberculosis, *Gene* **282**, 169–177.

Lew, B.M., Mills, K.V., Paulus, H. (1998) Protein splicing in vitro with a semisynthetic two-component minimal intein, *J. Biol. Chem.* **273**, 15887–15890.

Lew, B.M., Mills, K.V., Paulus, H. (1999) Characteristics of protein splicing in trans mediated by a semisynthetic split intein, *Biopolymers* **51**, 355–362.

Liu, X.Q., Hu, Z. (1997) A DnaB intein in Rhodothermus marinus: indication of recent intein homing across remotely related organisms, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 7851–7856.

Liu, X.Q., Yang, J. (2003) Split dnaE genes encoding multiple novel inteins in Trichodesmium erythraeum, *J. Biol. Chem.* **278**, 26315–26318.

Lovrinovic, M., Seidel, R., Wacker, R., Schroeder, H., Seitz, O., Engelhard, M., Goody, R.S., Niemeyer, C.M. (2003) Synthesis of protein-nucleic acid conjugates by expressed protein ligation, *Chem. Commun. (Camb).* **7**, 822–823.

Lum, L., Beachy, P.A. (2004) The Hedgehog response network: sensors, switches, and routers, *Science* **304**, 1755–1759.

Martin, D.D., Xu, M.Q., Evans, T.C. Jr. (2001) Characterization of a naturally occurring trans-splicing intein from Synechocystis sp. PCC6803, *Biochemistry* **40**, 1393–1402.

Mathys, S., Evans, T.C., Chute, I.C., Wu, H., Chong, S., Benner, J., Liu, X.Q., Xu, M.Q. (1999) Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation, *Gene* **231**, 1–13.

Mills, K.V., Paulus, H. (2001) Reversible inhibition of protein splicing by zinc ion, *J. Biol. Chem.* **276**, 10832–10838.

Mills, K.V., Lew, B.M., Jiang, S., Paulus, H. (1998) Protein splicing in trans by purified N- and C-terminal fragments of the Mycobacterium tuberculosis RecA intein, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 3543–3548.

Mills, K.V., Manning, J.S., Garcia, A.M., Wuerdeman, L.A. (2004) Protein splicing of a

Pyrococcus abyssi intein with a C-terminal glutamine, *J. Biol. Chem.* **279**, 20685–20691.

Ming, J.E., Muenke, M. (1998) Holoprosencephaly: from Homer to Hedgehog, *Clin. Genet.* **53**, 155–163.

Mizutani, R., Anraku, Y., Satow, Y. (2004) Protein splicing of yeast VMA1-derived endonuclease via thiazolidine intermediates, *J. Synchrotron. Radiat.* **11**, 109–112.

Mizutani, R., Nogami, S., Kawasaki, M., Ohya, Y., Anraku, Y., Satow, Y. (2002) Protein-splicing reaction via a thiazolidine intermediate: crystal structure of the VMA1-derived endonuclease bearing the N and C-terminal propeptides, *J. Mol. Biol.* **316**, 919–929.

Mootz, H.D., Blum, E.S., Tyszkiewicz, A.B., Muir, T.W. (2003) Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo, *J. Am. Chem. Soc.* **125**, 10561–10569.

Muir, T.W., Sondhi, D., Cole, P.A. (1998) Expressed protein ligation: a general method for protein engineering, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6705–6710.

Nichols, N.M., Benner, J.S., Martin, D.D., Evans, T.C. Jr. (2003) Zinc ion effects on individual Ssp DnaE intein splicing steps: regulating pathway progression, *Biochemistry* **42**, 5301–5311.

Nilsson, B.L., Kiessling, L.L., Raines, R.T. (2000) Staudinger ligation: a peptide from a thioester and azide, *Org. Lett.* **2**, 1939–1941.

Offer, J., Dawson, P.E. (2000) N^{α}-2-Mercaptobenzylamine-Assisted Chemical Ligation, *Org. Lett.* **2**, 23–26.

Ozawa, T., Kaihara, A., Sato, M., Tachihara, K., Umezawa, Y. (2001) Split luciferase as an optical probe for detecting protein-protein interactions in mammalian cells based on protein splicing, *Anal. Chem.* **73**, 2516–2521.

Ozawa, T., Takeuchi, T.M., Kaihara, A., Sato, M., Umezawa, Y. (2001) Protein splicing-based reconstitution of split green fluorescent protein for monitoring protein-protein interactions in bacteria: improved sensitivity and reduced screening time, *Anal. Chem.* **73**, 5866–5874.

Perler, F.B. (2002) InBase: the Intein Database, *Nucleic Acids Res.* **30**, 383–384.

Perler, F.B., Comb, D.G., Jack, W.E., Moran, L.S., Qiang, B., Kucera, R.B., Benner, J., Slatko, B.E., Nwankwo, D.O., Hempstead, S.K., Carlow, C.K.S., Jannasch, H. (1992) Intervening sequences in an Archaea DNA polymerase gene, *Proc. Natl. Acad. Sci. U. S. A.* **89**, 5577–5581.

Pietrokovski, S. (1994) Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins, *Protein Sci.* **3**, 2340–2350.

Pietrokovski, S. (1998) Modular organization of inteins and C-terminal autocatalytic domains, *Protein Sci.* **7**, 64–71.

Poland, B.W., Xu, M.Q., Quiocho, F.A. (2000) Structural insights into the protein splicing mechanism of PI-SceI, *J. Biol. Chem.* **275**, 16408–16413.

Porter, J.A., Young, K.E., Beachy, P.A. (1996) Cholesterol modification of hedgehog signaling proteins in animal development, *Science* **274**, 255–259.

Porter, J.A., Ekker, S.C., Park, W.J., von Kessler, D.P., Young, K.E., Chen, C.H., Ma, Y., Woods, A.S., Cotter, R.J., Koonin, E.V., Beachy, P.A. (1996) Hedgehog patterning activity: role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain, *Cell* **86**, 21–34.

Romanelli, A., Shekhtman, A., Cowburn, D., Muir, T.W. (2004) Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein-intein junction, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6397–6402.

Rosenblum, J.S., Blobel, G. (1999) Autoproteolysis in nucleoporin biogenesis, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 11370–11375.

Scott, C.P., Abel-Santos, E., Wall, M., Wahnon, D.C., Benkovic, S.J. (1999) Production of cyclic peptides and proteins in vivo, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13638–13643.

Shao, Y., Paulus, H. (1997) Protein splicing: estimation of the rate of O-N and S-N acyl rearrangements, the last step of the splicing process, *J. Pept. Res.* **50**, 193–198.

Shao, Y., Xu, M.Q., Paulus, H. (1995) Protein splicing: characterization of the aminosuccinimide residue at the carboxyl terminus of the excised intervening sequence, *Biochemistry* **34**, 10844–10850.

Shao, Y., Xu, M.Q., Paulus, H. (1996) Protein splicing: evidence for an N-O acyl rearrangement as the initial step in the splicing process, *Biochemistry* **35**, 3810–3815.

Shingledecker, K., Jiang, S., Paulus, H. (2000) Reactivity of the cysteine residues in the protein splicing active center of the

Mycobacterium tuberculosis RecA intein, *Arch. Biochem. Biophys.* **375**, 138–144.

Shingledecker, K., Jiang, S.Q., Paulus, H. (1998) Molecular dissection of the Mycobacterium tuberculosis RecA intein: design of a minimal intein and of a trans-splicing system involving two intein fragments, *Gene* **207**, 187–195.

Southworth, M.W., Perler, F.B. (2002) Protein splicing of the Deinococcus radiodurans strain R1 Snf2 intein, *J. Bacteriol.* **184**, 6387–6388.

Southworth, M.W., Benner, J., Perler, F.B. (2000) An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile, *EMBO J.* **19**, 5019–5026.

Southworth, M.W., Amaya, K., Evans, T.C., Xu, M.Q., Perler, F.B. (1999) Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein, *Biotechniques* **27**, 110–114, 116, 118–120.

Southworth, M.W., Adam, E., Panne, D., Byer, R., Kautz, R., Perler, F.B. (1998) Control of protein splicing by intein fragment reassembly, *EMBO J.* **17**, 918–926.

Sun, L., Ghosh, I., Paulus, H., Xu, M.Q. (2001) Protein trans-splicing to produce herbicide-resistant acetolactate synthase, *Appl. Environ. Microbiol.* **67**, 1025–1029.

Takeda, S., Tsukiji, S., Nagamune, T. (2004) Site-specific conjugation of oligonucleotides to the C-terminus of recombinant protein by expressed protein ligation, *Bioorg. Med. Chem. Lett.* **14**, 2407–2410.

Tam, J.P., Lu, Y.A., Liu, C.F., Shao, J. (1995) Peptide synthesis using unprotected peptides through orthogonal coupling methods, *Proc. Natl. Acad. Sci. U. S. A.* **92**, 12485–12489.

Teixeira, M.T., Fabre, E., Dujon, B. (1999) Self-catalyzed cleavage of the yeast nucleoporin Nup145p precursor, *J. Biol. Chem.* **274**, 32439–32444.

Telenti, A., Southworth, M., Alcaide, F., Daugelat, S., Jacobs, W.R., Perler, F.B. Jr. (1997) The Mycobacterium xenopi GyrA protein splicing element: characterization of a minimal intein, *J. Bacteriol.* **179**, 6378–6382.

Vanderslice, P., Copeland, W.C., Robertus, J.D. (1988) Site-directed alteration of serine 82 causes nonproductive chain cleavage in prohistidine decarboxylase, *J. Biol. Chem.* **263**, 10583–10586.

Voorter, C.E., de Haard-Hoekman, W.A., van den Oetelaar, P.J., Bloemendal, H., de Jong, W.W. (1988) Spontaneous peptide bond cleavage in aging alpha-crystallin through a succinimide intermediate, *J. Biol. Chem.* **263**, 19020–19023.

Wang, S., Liu, X.Q. (1997) Identification of an unusual intein in chloroplast ClpP protease of Chlamydomonas eugametos, *J. Biol. Chem.* **272**, 11869–11873.

Wood, D.W., Wu, W., Belfort, G., Derbyshire, V., Belfort, M. (1999) A genetic system yields self-cleaving inteins for bioseparations, *Nat. Biotechnol.* **17**, 889–892.

Wu, H., Hu, Z., Liu, X.Q. (1998) Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9226–9231.

Xu, M.Q., Perler, F.B. (1996) The mechanism of protein splicing and its modulation by mutation, *EMBO J.* **15**, 5146–5153.

Xu, Q., Buckley, D., Guan, C., Guo, H.C. (1999) Structural insights into the mechanism of intramolecular proteolysis, *Cell* **98**, 651–661.

Xu, M.Q., Southworth, M.W., Mersha, F.B., Hornstra, L.J., Perler, F.B. (1993) In vitro protein splicing of purified precursor and the identification of a branched intermediate, *Cell* **75**, 1371–1377.

Xu, M.Q., Comb, D.G., Paulus, H., Noren, C.J., Shao, Y., Perler, F.B. (1994) Protein splicing: an analysis of the branched intermediate and its resolution by succinimide formation, *EMBO J.* **13**, 5517–5522.

Xu, R., Ayers, B., Cowburn, D., Muir, T.W. (1999) Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 388–393.

Xuan, J., Tarentino, A.L., Grimwood, B.G., Plummer, T.H., Cui, T., Guan, C., Van Roey, P. Jr. (1998) Crystal structure of glycosylasparaginase from Flavobacterium meningosepticum, *Protein Sci.* **7**, 774–781.

Yamamoto, K., Low, B., Rutherford, S.A., Rajagopalan, M., Madiraju, M.V. (2001) The Mycobacterium avium-intracellulare complex dnaB locus and protein intein splicing, *Biochem. Biophys. Res. Commun.* **280**, 898–903.

Yang, J., Fox, G.C., Henry-Smith, T.V. Jr. (2003) Intein-mediated assembly of a functional beta-glucuronidase in transgenic plants, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3513–3518.

Protein Structure Analysis: High-throughput Approaches

Andrew P. Turnbull and Udo Heinemann
Max Delbrück Center for Molecular Medicine, Berlin, Germany

1 Bioinformatics 269
1.1 Structure-to-function Approaches 269
1.2 Identification of Disordered Regions in a Protein 272
1.3 Protein-ligand Complexes 272

2 Protein Production 273
2.1 Yeast 274
2.2 Baculovirus–insect Cell 274
2.3 *Leishmania tarentolae* (Trypanosomatidae) 274
2.4 Cell-free Expression Systems 275

3 Purification 275

4 Structure Determination by X-ray Crystallography 276
4.1 Crystallization 276
4.2 Data Collection 278
4.3 Phasing 278
4.4 Automated Structure Determination 279
4.5 Molecular Replacement 280

5 Structure-based Drug Design 280

6 Structure Determination by NMR 280

Bibliography 282
Books and Reviews 282
Primary Literature 282

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

Keywords

Active Site
Part of an enzyme molecule made up of amino acid residues, and involved in substrate binding and catalysis.

Anomalous Diffraction
Diffraction of X-rays near the absorption edge of a scattering atom, where a phase shift occurs and the atomic scattering factor becomes a complex quantity; This phenomenon is exploited in phasing X-ray diffraction data.

Isotope Labeling
Introduction of ^{13}C, ^{15}N and other nonradioactive isotopes into a protein for structure determination by nuclear magnetic resonance (NMR) spectroscopy.

Phasing
Reconstitution of the phase relations of diffracted X-rays as part of a crystal structure analysis, using anomalous diffraction and other techniques.

Posttranslational Modification
Covalent modification of a protein molecule after translation by phosphorylation, glycosylation, acetylation and so on. This is rare in prokaryotic proteins, and common in proteins from eukaryotes.

Protein Disorder
Lack of defined three-dimensional structure in segments of proteins or complete proteins in the absence of stabilizing binding partners.

Protein Domain
Compact folding unit of a protein, recognizable on the sequence and/or three-dimensional structure level.

Protein Fold
Recurrent pattern of three-dimensional structure in a protein or protein domain. There are far fewer protein folds than sequences or sequence families.

Selenomethionine
Nonnatural amino acid that can be incorporated into proteins by gene technology in place of methionine. Its selenium atom is used as an anomalous scatterer for phasing the X-ray diffraction pattern.

Structural Genomics
Large-scale project to determine the shapes of all proteins and other important biomolecules encoded by the genomes of key organisms.

Structural Proteomics
Large-scale project to determine the shapes and functions of all proteins encoded by the genomes of key organisms.

Synchrotron Radiation
Electromagnetic radiation emitted by subatomic particles (electrons or positrons) traveling at high velocity in storage rings. X-rays produced at synchrotrons are used in crystal structure determination.

Developments in the high-throughput analysis of protein structure have been primarily driven by worldwide structural genomics initiatives that are aimed at determining the three-dimensional structures of all proteins and other important biomolecules encoded by the genomes of key organisms. Structural genomics requires a large number of procedural steps in order to convert sequence information into a three-dimensional structure; this has led to new high-throughput methods for protein production, characterization, and structure determination. Over the past decade, the most notable technological advances have been in the fields of X-ray crystallography and NMR – the principal tools of structural genomics – which have facilitated high-throughput, rapid, and cost-effective structure determinations. In recent years, these developments have resulted in an exponential increase in the number of structures being deposited in the Protein Data Bank (PDB; http://www.rcsb.org/pdb/), in which the total number currently exceeds 30 000. Major developments in the fields of bioinformatics, protein production, and structure determination, and the impact of these on high-throughput protein structure analysis will be discussed in what follows.

1 Bioinformatics

Primary sequence analysis, such as similarity searches against protein sequence databases, protein domain architecture determination, identification of specialized local structural motifs, and prediction of protein structure are possible with a variety of homology-based modeling methods (Fig. 1). Sequence database searches are particularly useful in selecting targets for structural genomics initiatives and, where possible, generating homologous probes for determining structures by the molecular replacement method. For example, the Web-based program 3D-PSSM (http://www.sbg.bio.ic.ac.uk/~3dpssm/) is a fast method for predicting the protein fold from the primary amino acid sequence, and SWISS-MODEL (http://swissmodel.expasy.org/) is a fully automated protein structure homology-modeling server.

1.1 Structure-to-function Approaches

There are a number of bioinformatics resources available that are aimed

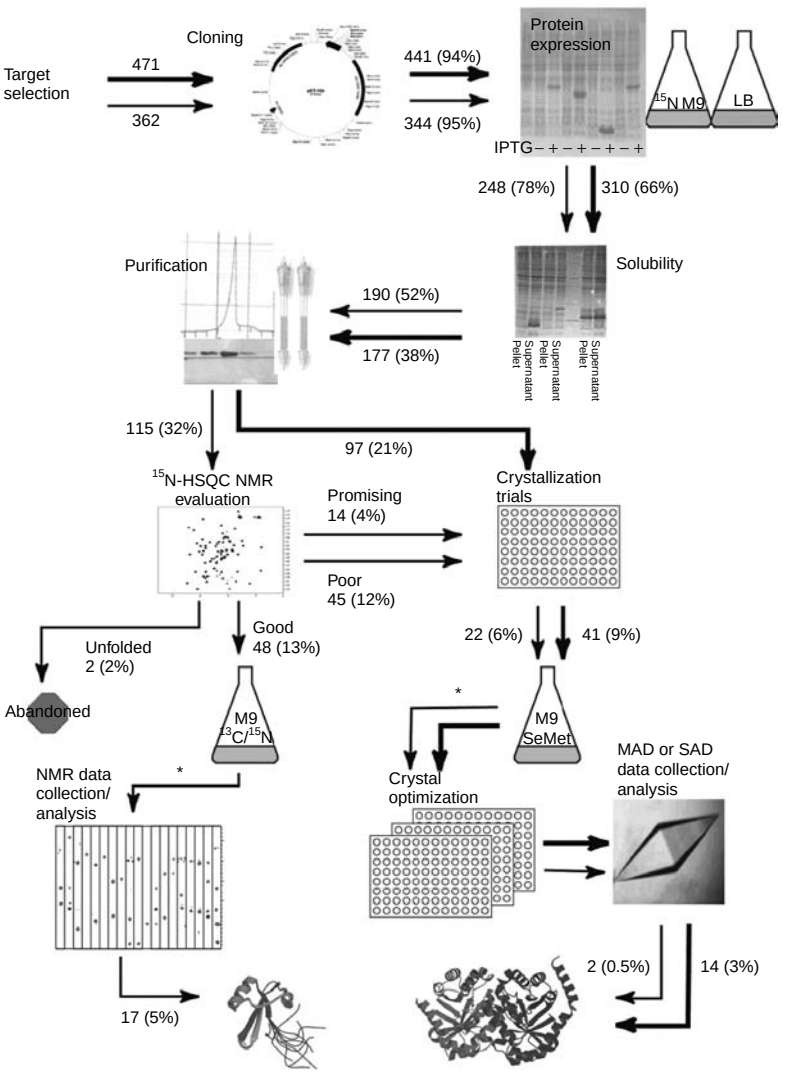

Fig. 1 Schematic flow diagram of the strategies employed in structural genomics initiatives, using *Methanobacterium thermoautotrophicum* as an example. The number of protein targets after each step and the percentage relative to the number of starting targets are indicated in brackets. Thin arrows and italicized numbers are for smaller molecular-weight proteins, and wide arrows and bold numbers are for larger molecular-weight proteins. Diagram taken from Yee, A., Pardee, K., Christendat, D., Savchenko, A., Edwards, A.M., Arrowsmith, C.H. (2003) Structural proteomics: toward high-throughput structural biology as a tool in functional genomics, *Acc. Chem. Res.* **36**, 183–189. Picture with permission from Prof. Cheryl Arrowsmith.

at identifying a protein's biochemical function from its three-dimensional structure (Fig. 2). For example, Dali (http://www2.ebi.ac.uk/dali) and VAST (Vector Alignment Search Tool; http://www.ncbi.nlm.nih.gov:80/Structure/VAST/vastsearch.html) offer Web-based servers for automatically comparing the fold of a newly determined structure against known folds, as represented by the protein structures in the PDB. Such comparisons can often reveal striking similarities between proteins that are not evident from sequence analysis alone, and that can provide important insights into biological function even in the absence of any other biochemical or functional data. However, computer-based approaches fail to assign functions to proteins that adopt novel folds. Enzymes are a notable exception to this rule, because the groupings of residues constituting their active sites tend to be highly conserved in their spatial disposition even in cases where there is no overall similarity in sequence or fold. For example, the relative positioning of the Ser-His-Asp catalytic triad of the serine proteases is highly conserved even when found in protein structures adopting different folds. Hence, screening a new protein structure against a database of enzyme active-site templates such as PROCAT (http://www.biochem.ucl.ac.uk/bsm/PROCAT/PROCAT.html) can be used for detecting key functional residues. The spatial patterns of residues can be automatically generated using various techniques including graph theory (ASSAM) and "fuzzy pattern matching" (RIGOR). Alternative approaches to identifying enzymes

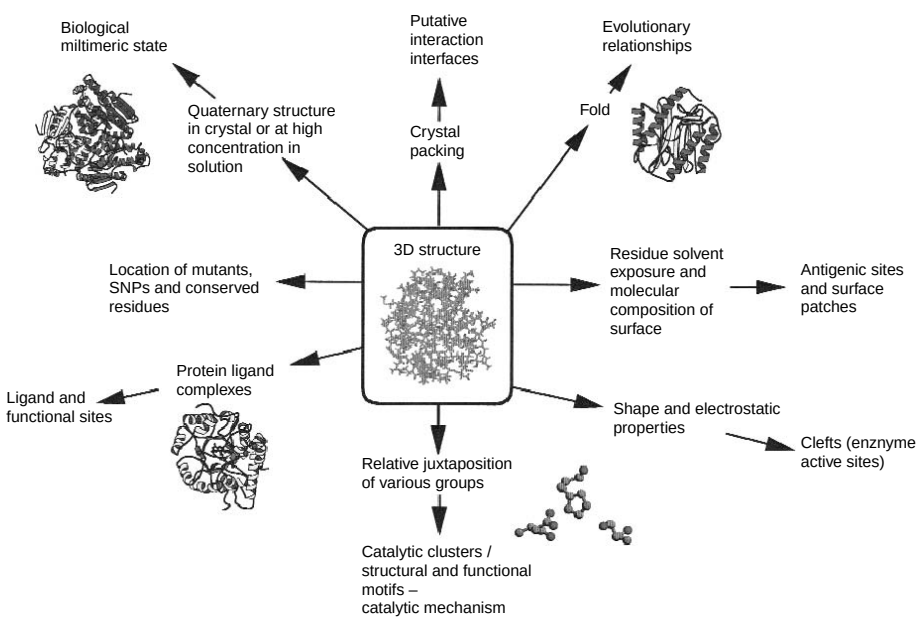

Fig. 2 Summary of the information deriving from the three-dimensional structure of a protein, relating to its biological function. Taken from Thornton, J.M., Todd, A.E., Milburn, D., Borkakoti, N., Orengo, C.A. (2000) From structure to function: approaches and limitations, *Nat. Struct. Biol.* **7**(Suppl.), 991–994.

on the basis of their three-dimensional structure and predicting their functions, have recently been reported. Here, a vector machine-learning algorithm is used, based on the secondary structure of proteins the propensities of amino acids, and surface properties, in order to discriminate enzymes from nonenzymes. Another approach analyses protein surface charges to identify conserved residues that can serve as catalytic sites. Once a general class of biochemical function of a protein has been proposed, experimental screening of enzymatic activity can be used to derive the precise biochemical function. For example, after the structure determination of BioH (an enzyme involved in biotin biosynthesis in *E. coli*) the protein structure was screened against a library of enzyme active sites, a Ser/His/Asp catalytic triad was identified, and subsequent hydrolase assays showed BioH to be a carboxylesterase.

1.2
Identification of Disordered Regions in a Protein

The occurrence of regions in proteins that lack any fixed tertiary structure is increasingly being observed in structural studies. These disordered regions or "random coils" are inherently flexible and are involved in a variety of functions, including the modulation of the specificity/affinity of protein-binding interactions, activation by cleavage, and DNA recognition. During the target selection process, it is important to consider any intrinsic protein disorder, because it can often lead to problems with the expression of protein-coding genes, protein stability, purification, and crystallization. PONDR, DisEMBL, and GlobPlot are useful tools for predicting potential disordered regions within a protein sequence that can be used to help design constructs corresponding to globular proteins or domains. PONDR (Predictor of Naturally Disordered Regions; http://www.pondr.com) and DisEMBL use methods based on artificial neural networks, whereas GlobPlot (http://globplot.embl.de) relies on a novel, propensity-based disorder-prediction algorithm. These methods can also be used to predict inherently flexible regions in protein sequences. For example, PONDR predicted that the linker between the DNA-operator-binding central domain of the transcriptional regulator KorB (KorB-O) and the KorB dimerization domain (KorB-C) is flexible, which was indeed observed in crystal structures and is thought to facilitate complex formation on circular plasmids (Fig. 3).

1.3
Protein-ligand Complexes

Protein-ligand complexes are the most useful in terms of providing functional information, because they reveal the nature of the ligand, the site at which it is bound to the protein, the location of the active site, and of the catalytic machinery (if the protein is an enzyme). There are several examples of structural analyses in which an unexpected protein-bound ligand or cofactor derived from the cloning organism was discovered. For example, the structure of the trimeric human protein p14.5 was found to have picked up benzoate molecules from the crystallization buffer at its inter-subunit tunnels, which most likely mark a hydrolytic active site (Fig. 4). When such data are available at high resolution, proposing a biological function for the protein can be relatively straightforward, because these data identify the nature of the ligand, the

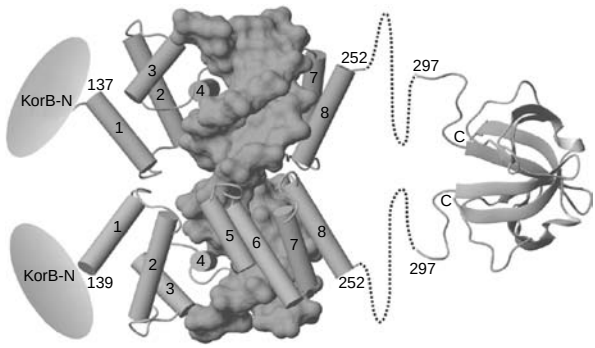

Fig. 3 Natively disordered regions in the bacterial transcriptional regulator and partitioning protein KorB. The KorB DNA-binding domains (KorB-O, center) are connected by flexible linkers to N-terminal domains of unknown structure and function (KorB-N, left), and the KorB dimerization domains (KorB-C, right). Picture taken from Khare et al., 2004.

Fig. 4 Crystal structure of the trimeric human protein, hp14.5. Benzoate molecules picked up from the crystallization buffer bind in the inter-subunit tunnels and mark putative hydrolytic active sites (Manjasetty et al., 2004). Picture with permission from Dr. B.A. Manjasetty.

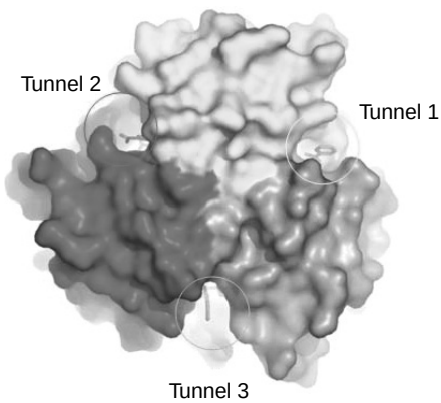

ligand-binding site, and the arrangement of catalytic residues from which a catalytic mechanism can be postulated.

2 Protein Production

Protein expression and purification play a central role in high-throughput protein structure analysis. Cloning using restriction enzymes is impractical for high-throughput approaches, because of the complications of selecting compatible and appropriate restriction enzymes for each cloning procedure, and the multiple steps of experimental refinement and treatment that must be performed. Therefore, high-throughput cloning requires procedures based on the polymerase chain reaction (PCR). High-throughput cloning and expression methods are now being developed in many laboratories, enabling the generation and testing of up to hundreds of DNA constructs for high-level expression in parallel, using rapid and

generic protocols. To generate expression vector clones, generic cloning systems can be used; the Gibco/Life Technologies GATEWAY™ system is one such example, streamlining the expression and cloning process by alleviating recloning steps and avoiding the use of restriction enzymes in the cloning and subcloning processes.

High-throughput approaches may rely on both prokaryotic and eukaryotic hosts. *Escherichia coli* expression systems are advantageous for many reasons, most notably because the overproduced protein is usually obtained without any posttranslational modification heterogeneity, and protein expression is cheaper and faster than with eukaryotic systems. However, it is not always possible to obtain soluble protein for many eukaryotic proteins, in particular, human proteins, by using the heterologous expression of eukaryotic genes, in which the codon usage for a cDNA could be suboptimal in *E. coli*. Furthermore, eukaryotic systems are necessary for the expression of proteins that require posttranslational modifications for their correct folding and activity. In situations in which a protein cannot be synthesized in *E. coli*, the eukaryotic yeasts (*Saccharomyces cerevisiae* and *Pichia pastoris*), baculovirus-infected insect cells, *Leishmania tarentolae,* and the cell-free wheat germ systems that have been developed for high-throughput approaches can be used. Furthermore, in these and other eukaryotic expression systems, the codon usage is closer to that in humans. The various merits of each system are discussed below.

2.1
Yeast

The yeasts *S. cerevisiae* and *P. pastoris* can be used for the routine production of recombinant proteins. More recently, *P. pastoris* has emerged as the preferred yeast, because of its strong, highly inducible promoter system resulting in higher yields of recombinant protein, stable genomic integration and posttranslational modifications, such as phosphorylation, which may be important for both the structure and function of some human proteins. In comparison with *S. cerevisiae,* the distribution and chain length of *N*-linked oligosaccharides are significantly shorter, and therefore this system represents a suitable alternative for the extracellular expression of human proteins. An additional advantage of the methylotrophic *P. pastoris* expression system is that it makes it possible to introduce ^{13}C label into recombinant proteins (by feeding ^{13}C-labeled methanol) for nuclear magnetic resonance (NMR) structural analyses.

2.2
Baculovirus–insect Cell

The recombinant baculovirus–insect cell expression system accomplishes most posttranslational modifications, including phosphorylation, *N*- and *O*-linked glycosylation, acylation, disulfide cross-linking, oligomeric assembly, and subcellular targeting, which may all be critical for the accurate production and function of human proteins. In contrast to bacterial expression systems, the recombinant baculovirus expression system usually produces soluble proteins without the need for induction or specific temperature conditions.

2.3
Leishmania tarentolae (Trypanosomatidae)

The use of the parasitic Trypanosomatidae species, *L. tarentolae,* as the host for

in vitro protein production has recently been reported, and can achieve high levels of protein expression. The Trypanosomatidae species of parasites naturally produce large amounts of glycoproteins, which is an advantage in the production of sialylated heterologous glycosylated proteins. Furthermore, given the natural auxotrophy of *L. tarentolae* for methionine, it has been suggested that such a system could prove to be useful for the production of selenomethionine-labeled proteins, for high-throughput X-ray crystallographic structure determination using single- or multiple-wavelength anomalous diffraction (SAD and MAD) phasing techniques.

2.4 Cell-free Expression Systems

Cell-free systems facilitate the parallel expression of many protein-coding genes, and are therefore suitable for high-throughput protein production. Obtaining the protein directly in a form suitable for X-ray crystallography or NMR spectroscopy is another prerequisite for high-throughput structure analysis. For the purpose of X-ray analysis, anomalous diffraction techniques (SAD or MAD) necessitate the substitution of methionine residues with selenomethionine. Additionally, NMR structure analysis often requires the labeling of proteins with ^{13}C and/or ^{15}N, which can be introduced through cell growth on media containing these isotopes in the form of ^{13}C-glucose and $^{15}NH_4Cl$. In this respect, the *E. coli* cell-free protein synthesis system of Kigawa et al. (1999) permits the straightforward incorporation of isotopes for NMR analysis. However, this system is not always suitable for the expression of some eukaryotic proteins and can result in aggregation, formation of insoluble inclusion bodies, and degradation of the expression product. Furthermore, in the case of multidomain proteins, which are found more often in eukaryotes, correct folding occurs more frequently in eukaryotic than in prokaryotic translation systems. Another limitation with the use of *E. coli* systems for high-throughput cell-free expression is that PCR-generated fragments are not transcribed and translated efficiently in these systems: contamination by the mRNA- and DNA-degradation enzymes originating from the cell decreases the stability of the templates and reduces yields. By contrast, in eukaryotic cell-free systems, the added mRNAs are stable for long periods of time, and therefore these systems overcome many of the limitations associated with *E. coli* cell-free systems. Furthermore, cell-free systems can produce high yields of correctly folded proteins and, unlike *in vivo* systems, facilitate the expression of proteins that would otherwise interfere with the host cell physiology. Recently, the synthesis and screening of gene products based on the cell-free system prepared from eukaryotic wheat embryos has been reported; this method bypasses many of the time-consuming cloning steps involved in conventional expression systems and lends itself to the high-throughput expression of proteins, using automated robotic systems.

3 Purification

Protein purification has also seen significant improvements because of the use of affinity tags fused to the protein of interest, so that it can be separated from the host cell proteins rapidly by using standardized purification schemes. N-terminal tags range from large tags, such as

glutathione-S-transferase (GST), maltose-binding protein (MBP), thioreductase, and the chitin-binding protein, to fairly small tags such as His_6, and various epitope tags. In some systems, the use of large tags can lead to the production of fusion products that are too large for high-level expression, but the use of such tags can improve the correct folding and stability of the overexpressed protein. His_6 is the most commonly used purification tag, because it is easier to incorporate into expression constructs and it allows a generic one-step purification using automated methods based on nickel-nitrilotriacetic acid (Ni-NTA) or other immobilized metal-affinity chromatography resins. The use of an additional C-terminal StrepII tag enables dual-affinity column purification, which ensures that only full-length gene products are separated. For crystallization to occur, it is generally agreed that the fusion tags must be removed, because they can introduce flexible regions into the protein, which can interfere with crystallization or lead to various forms of microheterogeneity. In a recent study, the compatibility of small peptide affinity tags with protein crystallization was assessed, in which the N-terminus of the chicken spectrin SH3 domain was labeled with a His_6 tag and a StrepII tag, fused to the N- and C-termini, respectively. The resulting protein, His_6-SH3-StrepII, comprised 83 amino acid residues, 23 of which originated from the tags (accounting for 23% of the total fusion protein mass). In contrast to the general consensus that the presence of affinity tags is detrimental in structural studies, this study demonstrated that the fused affinity tags did not interfere with crystallization or structure analysis, and did not change the protein structure. This suggests that, in some cases, protein constructs utilizing both N- and C-terminal peptide tags may lend themselves to structural investigations in high-throughput regimes.

4 Structure Determination by X-ray Crystallography

4.1 Crystallization

Crystallization is regarded as a major bottleneck in structure determination by X-ray crystallography and can be divided into two stages: coarse screening for initial crystallization conditions, followed by optimization of the conditions in order to produce diffraction-quality single crystals. The field of crystallization has recently been revolutionized by significant developments in automation, miniaturization, and process integration. These developments have led to the availability of robotic liquid-handling systems that are capable of rapidly and efficiently screening thousands of crystallization conditions, in which different parameters such as ionic strength, precipitant concentration, additives, pH, and temperature are altered. High-throughput screening begins with the automated preparation and/or reformatting of precipitant solutions into crystallization microplates. The latest nanoliter robotic liquid-dispensing systems are capable of dispensing very small drops (typically containing between 25 and 100 nL of protein), which reduces the amount of protein required for screening conditions and the overhead on protein production. Furthermore, smaller drops equilibrate faster than larger drops, leading to a more rapid appearance of crystals. The crystallization experiments are then regularly monitored using an imaging robot, and the collected images can either be analyzed visually

or by using automatic crystal recognition systems. HomeBase™ (The Automation Partnership, UK) represents an integrated system combining high-density plate storage and imaging.

A recent development in protein crystallization has been the use of a microfluidic system for crystallizing proteins, using the free-interface diffusion method at the nanoliter scale. This system is capable of screening hundreds of crystallization conditions in which droplets, each containing solutions of protein, precipitants, and additives in various ratios, are formed in the flow of immiscible fluids inside a polydimethylsiloxane (PDMS)/glass capillary composite microfluidic device (Fig. 5). The system is capable of performing multidimensional screening (mixing 5–10 solutions) and therefore explores more of the crystallization space than the conventional method of vapor diffusion, increasing the likelihood of obtaining crystals. Furthermore, the capillary containing protein crystals can be directly exposed to the synchrotron X-ray beam, eliminating the need to manually manipulate the crystal. These features also mean that it has the potential to serve as the basis for future high-throughput automated crystallization systems.

Membrane proteins represent the most persistent bottleneck for all preparatory procedures and analytical methods, because they are water soluble only in the presence of detergents, and are difficult to overproduce in the quantities required for structural studies. Membrane proteins, which constitute up to 30% of the protein repertoire of an organism, represent the targets for more than 50% of the drugs that are being currently used and tested. An adapted robotic system has recently been reported that enables high-throughput crystallization of membrane proteins using lipidic mesophases.

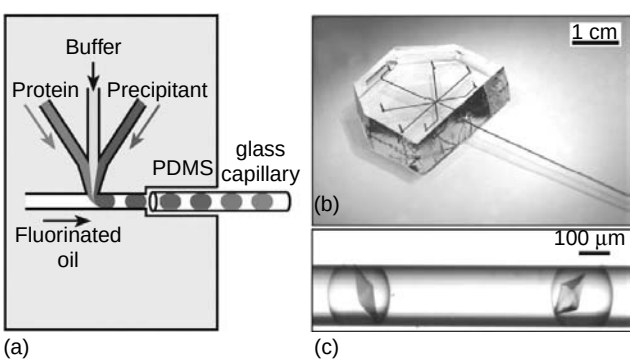

Fig. 5 (a) Schematic illustration of the droplet-based microfluidic system for protein crystallization. (b) Photograph of a polydimethylsiloxane (PDMS)/glass capillary composite microfluidic device. (c) A micrograph of thaumatin crystals grown in droplets, produced by the method outlined in panel (a), within a capillary. Taken from Zheng, B., Tice, J.D., Roach, L.S., Ismagilov, R.F. (2004) A droplet-based, composite PDMS/glass capillary microfluidic system for evaluating protein crystallization conditions by microbatch and vapor-diffusion methods with on-chip X-ray diffraction, *Angew. Chem., Int. Ed. Engl.* **43**, 2508–2511. Picture with permission from Dr. Rustem Ismagilov.

4.2 Data Collection

X-ray data collection has been revolutionized over the past decade, with the development of improved X-ray sources and detectors and the universal adoption of flash-freezing techniques that greatly reduce crystal radiation damage. New third-generation synchrotrons are now available across the world, which provide more intense and stable X-ray beams and, combined with new, faster, larger, and more sensitive X-ray detectors, allow higher-quality data to be collected much more rapidly, leading to a dramatic increase in the success rate of structure determination. Furthermore, specialized, highly collimated microfocus beamlines, such as ID13 at the ESRF (Grenoble, France; http://www.esrf.fr/exp_facilities/ID13/index.html) and the protein crystallography beamline at the Swiss Light Source (SLS), are specifically tailored to study biological crystals with small physical dimensions (microcrystals from 5 to 50 μm in size) or very large unit cell dimensions. Future developments in detector design, such as the MarResearch GmbH (Norderstedt, Germany) flat-panel detector, promise larger continuous active areas, higher spatial resolution, very low noise, and fast readout times (~1 s). High-throughput X-ray data collection has also led to the development of automated robotic sample changers that store and mount crystals sequentially while maintaining the samples at liquid-nitrogen temperatures (100 K). A novel system has recently been described, using a proprietary crystallization plate that facilitates the preliminary investigation of the diffraction properties of crystals *in situ* in the drop, by direct exposure to the X-ray beam. The BIOXHIT project (Biocrystallography (X) on a Highly Integrated Technology Platform for European Structural Genomics; http://www.bioxhit.org), comprising more than 20 research groups from all over Europe, aims to develop an integrated technology platform for synchrotron beamlines by promoting new approaches to crystallization, and fully automating diffraction data collection and structure determination.

Recent developments in technology have not been limited exclusively to synchrotron sources. The latest generation of high-intensity X-ray generators has revolutionized X-ray sources to such an extent that the latest "in-house" systems from Rigaku/MSC (e.g. the FR-E SuperBright; http://www.rigakumsc.com/) and Bruker AXS (http://www.bruker-axs.com) are comparable in intensity to first-generation synchrotron sources and, coupled with automated sample changers, make high-throughput crystallography possible in the laboratory. Furthermore, the development of the Compact Light Source (Lyncean Technologies, Inc., Palo Alto, CA, USA) promises to have a huge impact on protein structure determination, offering the possibility of a "synchrotron beamline" for home laboratory applications. This tunable, tabletop X-ray source combines an electron beam with a laser beam to generate an intense X-ray beam and, as a next-generation X-ray source, directly addresses the increasing demand for high-throughput protein crystallography.

4.3 Phasing

The central problem in X-ray crystallography is the determination of the protein phases. X-ray data collected from a crystal

consist of structure factor amplitudes, but there is no way of directly measuring the phase associated with each amplitude. Recent advances in macromolecular phasing have simplified and further automated this crucial stage of X-ray structure determination to such an extent that the eventual success of a project is usually assured, if well-diffracting crystals are available. The techniques of isomorphous replacement, anomalous scattering, molecular replacement, and single (SAD) and multiple (MAD) anomalous dispersion are commonly used to solve the phase problem. However, phase determination has been dramatically facilitated by the widespread adoption of SAD and MAD phasing techniques by the crystallographic community, primarily as a consequence of the availability of stable and tunable synchrotron sources. These allow the optimal exploitation of the anomalous effect as a source of phase information by delivering X-ray energies corresponding to absorption maxima of the anomalous scatterers, very often selenium introduced as selenomethionine through substitution of methionine. Additionally, heavy-atom labels (mercury, platinum, and others) may be bound to crystalline proteins by soaking to yield phase information by anomalous diffraction techniques.

Modern phasing techniques, such as fast halide soaks and sulfur-SAD, hold promise of simpler and faster protein structure determination than traditional methods. Fast halide soaks using bromide and iodide that diffuse rapidly into the crystal and display significant anomalous scattering signals can be used to quickly derivatize protein crystals. Heavy-atom reagents can also be incorporated into the crystal in a relatively short time if a concentration greater than 10 mM is used. These derivatives then display better isomorphism and diffraction qualities than those obtained after a standard, prolonged soak. The availability of stable synchrotron beamlines and improvements in data processing programs make it possible to collect extremely accurate diffraction data, and to determine structures using the very weak anomalous signals from atoms such as sulfur and phosphorus that are inherently present in macromolecules or nucleic acids; several novel structures have been determined by using sulfur-SAD phasing.

4.4 Automated Structure Determination

High-throughput crystallographic structure determination requires software that is automated and designed for minimum user intervention. There has been considerable development in the direct-method programs, SHELXD and SnB, which can automatically determine heavy-atom substructures from a very small signal. HKL2MAP connects several programs from the SHELX suite to guide the user from analyzing scaled diffraction data (SHELXC), through substructure solution (SHELXD) and phasing (SHELXE), to displaying an electron density map (Xfit). There are a number of other automated software systems, such as ACrS (automated crystallographic system) and PHENIX (Python-based hierarchical environment for integrated Xtallography) that are currently being developed to meet the requirements of high-throughput structure determination by combining multiple structure-determination software packages into one intuitive interface.

Finally, new algorithms for interpreting electron density maps and for automated model building, such as, SOLVE/RESOLVE, AUTOSHARP/

SHARP, and ARP/wARP, enable rapid construction of protein models without the need for significant manual intervention. At present, the success rates for these programs are dependent on the resolution of the diffraction data (typically, 2.5-Å resolution or higher is necessary for automatic chain fitting).

4.5 Molecular Replacement

When an approximate structural model of a protein under investigation is available, either from NMR, a homologous X-ray structure, or from homology modeling, initial phases can be obtained using molecular replacement where the homologous probe structure is fitted to the experimental data using three rotational and three translational parameters. The programs AMoRe and MolRep that are integrated into the CCP4i GUI simplify the problem of positioning a molecule in the asymmetric unit by running sequential rotational and translational searches. Advances in molecular replacement include the implementation of the maximum likelihood–based algorithms in BEAST, and the six-dimensional evolutionary search algorithm in EPMR. As the number of protein structures increases, it is anticipated that molecular replacement will become the standard method for structure determination, using a generalized search of all unique protein domains present in the PDB.

5 Structure-based Drug Design

Knowledge of the three-dimensional structure of proteins can play a key role in the development of small-molecule drugs, because being able to verify how lead compounds bind to their targets accelerates drug development and is more cost-effective. Notable drugs that have been successfully designed using protein three-dimensional structure information include the HIV protease inhibitors, Viracept™ (Agouron, USA and Eli Lilly, USA), and Agenerase™ (Vertex, USA; Kissei, Japan; Glaxo Wellcome, UK). However, it is usually necessary to screen a large number of protein–ligand complex structures in the iterative process of rational structure-based drug design. Hence, high-throughput protein X-ray crystallography offers an unprecedented opportunity for facilitating drug discovery. Recent advances in rapid binding-site analysis of *de novo* targets using virtual ligand (*in silico*) screening and small-molecule cocrystallization methodologies, in combination with the miniaturization and automation of structural biology, enable more rapid lead compound identification and faster optimization, providing a framework for direct integration into the drug discovery process. In the high-throughput structure determination of protein–ligand complexes, it is desirable to use tools that can locate, build, and refine the structure of the bound ligand with minimal human intervention. One such tool is X-LIGAND, part of the QUANTA software package (Accelrys, San Diego, CA, USA) that automatically searches for unoccupied regions of electron density in the structure of the protein–ligand complex in which it tries to fit the ligand.

6 Structure Determination by NMR

Solution-state NMR spectroscopy can serve as a technique complementary to X-ray crystallography in protein structure

analysis, particularly in the context of structural genomics initiatives, where many protein targets either do not crystallize or do not form crystals suitable for crystallographic studies (owing to small crystal size or poor diffraction quality). NMR measurements are performed in aqueous solution, obviating the need for growing crystals. This technique is applicable primarily to small proteins (<30 kDa) that are highly soluble (millimolar concentrations), and is particularly useful in the study of proteins that are partially unfolded in the absence of their appropriate binding partners. The additional technique of solid-state NMR is useful in providing structural information for some integral membrane proteins that may not be accessible using crystallographic methods. Furthermore, chemical shift perturbation studies can be used to validate proposed biochemical functions, to map ligand-binding epitopes, and to screen for small-molecule ligands in drug development.

High-throughput, NMR-based structure determination requires rapid and automated data acquisition and analysis methods. The major challenges in realizing this have been those of increasing instrumentation sensitivity (signal-to-noise ratio) and reducing the time required for data collection. These technical issues have been addressed by constructing new high-field magnets and by the recent introduction of cryogenic probes that operate at low temperatures (~25 K), permitting the investigation of proteins that have either low solubility or low yields from purification. Additionally, the application of TROSY (transverse relaxed optimized spectroscopy), a novel spectroscopic concept based on the selection of slowly relaxing NMR transitions, has provided significant sensitivity enhancements for large proteins.

The most time-consuming aspects of structure determination by NMR are the long data collection times necessary for independently sampling three or more indirect dimensions along with the time taken to interpret the correspondingly large number of spectra from $^{13}C/^{15}N$-isotope-labeled samples. Rapid resonance assignment is a prerequisite for high-throughput NMR structure determination; techniques such as reduced-dimensionality $^{13}C,^{15}N,^{1}H$-triple resonance NMR also avoid the sampling limited regime through the simultaneous frequency-labeling of two spin types in a single indirect dimension. Heteronuclear multidimensional data reduce complications arising from interspectral variations by maximizing the dimensionality of the spectra and decrease signal overlap of data sets sufficiently for the data to be analyzed automatically. Recent approaches to automated structure elucidation from NMR spectra include NOESY-Jigsaw, in which sparse and unassigned NMR data can be used to reasonably and accurately assess secondary structure and align it. The information thus retrieved is useful for quick structural assays for assessing folds before full structural determination and can therefore assist in fold prediction. Additionally, the program ATNOS (automated NOESY peak picking) enables automated peak picking and NOE signal identification in homonuclear 2D and heteronuclear 3D [^{1}H, ^{1}H]-NOESY spectra during *de novo* protein structure determination.

See also Protein NMR Spectroscopy; Proteomics.

Bibliography

Books and Reviews

Blundell, T.L., Mizuguchi, K. (2000) Structural genomics: an overview, *Prog. Biophys. Mol. Biol.* **73**, 289–295.

Dauter, Z. (2002) New approaches to high-throughput phasing, *Curr. Opin. Struct. Biol.* **12**, 674–678.

Guntert, P. (2004) Automated NMR structure calculation with CYANA, *Methods Mol. Biol.* **278**, 353–378.

Laskowski, R.A., Watson, J.D., Thornton, J.M. (2003) From protein structure to biochemical function? *J. Struct. Funct. Genomics* **4**, 167–177.

Norin, M., Sundstrom, M. (2002) Structural proteomics: developments in structure-to-function predictions, *Trends Biotechnol.* **20**, 79–84.

Stevens, R.C. (2000) Design of high-throughput methods of protein production for structural biology, *Struct. Fold. Des.* **8**, R177–R185.

Stewart, L., Clark, R., Behnke, C. (2002) High-throughput crystallization and structure determination in drug discovery, *Drug Discov. Today* **7**, 187–196.

Tickle, I., Sharff, A., Vinkovic, M., Yon, J., Jhoti, H. (2004) High-throughput protein crystallography and drug discovery, *Chem. Soc. Rev.* **33**, 558–565.

Primary Literature

Abola, E., Kuhn, P., Earnest, T., Stevens, R.C. (2000) Automation of X-ray crystallography, *Nat. Struct. Biol.* **7**, 973–977.

Adams, P., Gopal, k., Grossekunstleve, R.W., Hung, L.-W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Pai, R.K., Read, R.J., Romo, T.D., Sacchettini, J.C., Sauter, N.K., Storoni, L.C., Terwilliger, T.C. (2004) Recent developments in the PHENIX software for automated crystallographic structure determination, *J. Synchrotron Radiat.* **11**, 53–55.

Albala, J.S., Franke, K., McConnell, I.R., Pak, K.L., Folta, P.A., Rubinfeld, B., Davies, A.H., Lennon, G.G., Clark, R. (2000) From genes to proteins: high-throughput expression and purification of the human proteome, *J. Cell. Biochem.* **80**, 187–191.

Bailey-Kellogg, C., Widge, A., Kelley, J.J., Berardi, M.J., Bushweller, J.H., Donald, B.R. (2000) The NOESY jigsaw: automated protein secondary structure and main-chain assignment from sparse, unassigned NMR data, *J. Comput. Biol.* **7**, 537–558.

Bate, P., Warwicker, J. (2004) Enzyme/non-enzyme discrimination and prediction of enzyme active site location using charge-based methods, *J. Mol. Biol.* **340**, 263–276.

Breitling, R., Klingner, S., Callewaert, N., Pietrucha, R., Geyer, A., Ehrlich, G., Hartung, R., Muller, A., Contreras, R., Beverley, S.M., Alexandrov, K. (2002) Non-pathogenic trypanosomatid protozoa as a platform for protein research and production, *Protein Expr. Purif.* **25**, 209–218.

Brenner, S.E. (2000) Target selection for structural genomics, *Nat. Struct. Biol.* **7**(Suppl.), 967–969.

Brown, J., Walter, T.S., Carter, L., Abrescia, G.A., Aricescu, A.R., Batuwangala, T.D., Bird, L.E., Brown, N., Chamberlain, P.P., Davis, S.J., Dubinina, E., Endicott, J., Fennelly, J.A., Gilbert, R.J.C., Harkiolaki, M., Hon, W.-C., Kimberley, F., Love, C.A., Mancini, E.J., Manso-Sancho, R., Nichols, C.E., Robinson, R.A., Sutton, G.C., Schueller, N., Sleeman, M.C., Stewart-Jones, G.B., Vuong, M., Welburn, J., Zhang, Z., Stammers, D.K., Owens, R.J., Jones, E.Y., Harlos, K., Stuart, D.I. (2003) A procedure for setting up high-throughput nanoliter crystallization experiments. II. Crystallization results, *J. Appl. Crystallogr.* **36**, 315–318.

Bruel, C., Cha, K., Reeves, P.J., Getmanova, E., Khorana, H.G. (2000) Rhodopsin kinase: expression in mammalian cells and a two-step purification, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3004–3009.

Brunzelle, J.S., Shafaee, P., Yang, X., Weigand, S., Ren, Z., Anderson, W.F. (2003) Automated crystallographic system for high-throughput protein structure determination, *Acta Crystallogr.* **D59**, 1138–1144.

Cereghino, J.L., Cregg, J.M. (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*, *FEMS Microbiol. Rev.* **24**, 45–66.

Cherezov, V., Peddi, A., Muthusubramaniam, L., Zheng, Y.F., Caffrey, M. (2004) A robotic system for crystallizing membrane and soluble proteins in lipidic mesophases, *Acta Crystallogr.* **D60**, 1795–1807.

Collaborative Computational Project, N. (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallogr.* **D50**, 760–763.

Dauter, Z., Dauter, M., Rajashankar, K.R. (2000) Novel approach to phasing proteins: derivatization by short cryo-soaking with halides, *Acta Crystallogr.* **D56**, 232–237.

del Val, C., Mehrle, A., Falkenhahn, M., Seiler, M., Glatting, K.H., Poustka, A., Suhai, S., Wiemann, S. (2004) High-throughput protein analysis integrating bioinformatics and experimental assays, *Nucleic Acids Res.* **32**, 742–748.

Delbruck, H., Ziegelin, G., Lanka, E., Heinemann, U. (2002) An Src homology 3-like domain is responsible for dimerization of the repressor protein KorB encoded by the promiscuous IncP plasmid RP4, *J. Biol. Chem.* **277**, 4191–4198.

Dobson, P.D., Doig, A.J. (2003) Distinguishing enzyme structures from non-enzymes without alignments, *J. Mol. Biol.* **330**, 771–783.

Endo, Y., Sawasaki, T. (2004) High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system, *J. Struct. Funct. Genomics* **5**, 45–57.

Fortelle, Edl., Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods, *Methods enzymol.* **276**, 472–494.

Garman, E. (1999) Cool data: quantity AND quality, *Acta Crystallogr.* **D55**, 1641–1653.

Garner, E., Cannon, P., Romero, P., Obradovic, Z., Dunker, A.K. (1998) Predicting disordered regions from amino acid sequence: common themes despite differing structural characterization, *Genome Inf. Ser. Worksh. Genome Inf.* **9**, 201–213.

Garner, E., Romero, P., Dunker, A.K., Brown, C., Obradovic, Z. (1999) Predicting binding regions within disordered proteins, *Genome Inf. Ser. Worksh. Genome Inf.* **10**, 41–50.

Goodwill, K.E., Tennant, M.G., Stevens, R.C. (2001) High-throughput x-ray crystallography for structure-based drug design, *Drug Discov. Today* **15**(Suppl.), 113–118.

Grinna, L.S., Tschopp, J.F. (1989) Size distribution and general structural features of N-linked oligosaccharides from the methylotrophic yeast, *Pichia pastoris*, *Yeast* **5**, 107–115.

Heinemann, U., Frevert, J., Hofmann, K., Illing, G., Maurer, C., Oschkinat, H., Saenger, W. (2000) An integrated approach to structural genomics, *Prog. Biophys. Mol. Biol.* **73**, 347–362.

Herrmann, T., Guntert, P., Wuthrich, K. (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS, *J. Biomol. NMR* **24**, 171–189.

Holm, L., Sander, C. (1993) Protein structure comparison by alignment of distance matrices, *J. Mol. Biol.* **233**, 123–138.

Jhoti, H. (2001) High-throughput structural proteomics using X-rays, *Trends Biotechnol.* **19**, S67–S71.

Kaldor, S.W., Kalish, V.J., Davies, J.F. II, Shetty, B.V., Fritz, J.E., Appelt, K., Burgess, J.A., Campanale, K.M., Chirgadze, N.Y., Clawson, D.K., Dressman, B.A., Hatch, S.D., Khalil, D.A., Kosa, M.B., Lubbehusen, P.P., Muesing, M.A., Patick, A.K., Reich, S.H., Su, K.S., Tatlock, J.H. (1997) Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease, *J. Med. Chem.* **40**, 3979–3985.

Kelley, L.A., MacCallum, R.M., Sternberg, M.J. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM, *J. Mol. Biol.* **299**, 499–520.

Khare, D., Ziegelin, G., Lanka, E., Heinemann, U. (2004) Sequence-specific DNA binding determined by contacts outside the helix-turn-helix motif of the ParB homolog KorB, *Nat. Struct. Mol. Biol.* **11**, 656–663.

Kim, E.E., Baker, C.T., Dwyer, M.D., Murcko, M.A., Rao, B.G., Tung, R.D., Navia, M.A. (1995) Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme, *J. Am. Chem. Soc.* **117**, 1181–1182.

Kissinger, C.R., Gehlhaar, D.K., Fogel, D.B. (1999) Rapid automated molecular replacement by evolutionary search, *Acta Crystallogr.* **D55**, 484–491.

Kleywegt, G.J. (1999) Recognition of spatial motifs in protein structures, *J. Mol. Biol.* **285**, 1887–1897.

Kolb, V.A., Makeyev, E.V., Spirin, A.S. (2000) Co-translational folding of an eukaryotic multidomain protein in a prokaryotic translation system, *J. Biol. Chem.* **275**, 16597–16601.

Kuhn, P., Wilson, K., Patch, M.G., Stevens, R.C. (2002) The genesis of high-throughput structure-based drug discovery using protein

crystallography, *Curr. Opin. Chem. Biol.* **6**, 704–710.

Lamzin, V.S., Perrakis, A. (2000) Current state of automated crystallographic data analysis, *Nat. Struct. Biol.* **7**(Suppl.), 978–981.

Li, X., Obradovic, Z., Brown, C.J., Garner, E.C., Dunker, A.K. (2000) Comparing predictors of disordered protein, *Genome Inf. Ser. Worksh. Genome Inf.* **11**, 172–184.

Li, X., Romero, P., Rani, M., Dunker, A.K., Obradovic, Z. (1999) Predicting protein disorder for N-, C-, and internal regions, *Genome Inf. Ser. Worksh. Genome Inf.* **10**, 30–40.

Linding, R., Russell, R.B., Neduva, V., Gibson, T.J. (2003) GlobPlot: exploring protein sequences for globularity and disorder, *Nucleic Acids Res.* **31**, 3701–3708.

Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., Russell, R.B. (2003) Protein disorder prediction: implications for structural proteomics, *Structure* **11**, 1453–1459.

Linial, M., Yona, G. (2000) Methodologies for target selection in structural genomics, *Prog. Biophys. Mol. Biol.* **73**, 297–320.

Madej, T., Gibrat, J.F., Bryant, S.H. (1995) Threading a database of protein cores, *Proteins* **23**, 356–369.

Manjasetty, B.A., Delbruck, H., Pham, D.T., Mueller, U., Fieber-Erdmann, M., Scheich, C., Sievert, V., Bussow, K., Niesen, F.H., Weihofen, W., Loll, B., Saenger, W., Heinemann, U., Neisen, F.H. (2004) Crystal structure of Homo sapiens protein hp14.5, *Proteins* **54**, 797–800.

Montelione, G.T., Zheng, D., Huang, Y.J., Gunsalus, K.C., Szyperski, T. (2000) Protein NMR spectroscopy in structural genomics, *Nat. Struct. Biol.* **7**(Suppl.), 982–985.

Mueller, U., Bussow, K., Diehl, A., Bartl, F.J., Niesen, F.H., Nyarsik, L., Heinemann, U. (2003) Rapid purification and crystal structure analysis of a small protein carrying two terminal affinity tags, *J. Struct. Funct. Genomics* **4**, 217–225.

Navaza, J. (2001) Implementation of molecular replacement in AMoRe, *Acta Crystallogr.* **D57**, 1367–1372.

Netzer, W.J., Hartl, F.U. (1997) Recombination of protein domains facilitated by co-translational folding in eukaryotes, *Nature* **388**, 343–349.

Novotny, M., Madsen, D., Kleywegt, G.J. (2004) Evaluation of protein fold comparison servers, *Proteins* **54**, 260–270.

Oldfield, T.J. (2001) X-LIGAND: an application for the automated addition of flexible ligands into electron density, *Acta Crystallogr.* **D57**, 696–705.

Pandey, N., Ganapathi, M., Kumar, K., Dasgupta, D., Das Sutar, S.K., Dash, D. (2004) Comparative analysis of protein unfoldedness in human housekeeping and non-housekeeping proteins, *Bioinformatics* **20**, 2904–2910.

Pape, T., Schneider, T.R. (2004) HKL2MAP: a graphical user interface for phasing with SHELX programs, *J. Appl. Crystallogr.* **37**, 843–844.

Perrakis, A., Morris, R., Lamzin, V.S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol.* **6**, 458–463.

Potterton, E., Briggs, P., Turkenburg, M., Dodson, E. (2003) A graphical user interface to the CCP4 program suite, *Acta Crystallogr.* **D59**, 1131–1137.

Prinz, B., Schultchen, J., Rydzewski, R., Holz, C., Boettner, M., Stahl, U., Lang, C. (2004) Establishing a versatile fermentation and purification procedure for human proteins expressed in the yeasts Saccharomyces cerevisiae and Pichia pastoris for structural genomics, *J. Struct. Funct. Genomics* **5**, 29–44.

Read, R.J. (2001) Pushing the boundaries of molecular replacement with maximum likelihood, *Acta Crystallogr.* **D57**, 1373–1382.

Sali, A., Glaeser, R., Earnest, T., Baumeister, W. (2003) From words to literature in structural proteomics, *Nature* **422**, 216–225.

Sanchez, R., Pieper, U., Melo, F., Eswar, N., Marti-Renom, M.A., Madhusudhan, M.S., Mirkovic, N., Sali, A. (2000) Protein structure modeling for structural genomics, *Nat. Struct. Biol.* **7**(Suppl.), 986–990.

Sanishvili, R., Yakunin, A.F., Laskowski, R.A., Skarina, T., Evdokimova, E., Doherty-Kirby, A., Lajoie, G.A., Thornton, J.M., Arrowsmith, C.H., Savchenko, A., Joachimiak, A., Edwards, A.M. (2003) Integrating structure, bioinformatics, and enzymology to discover function: BioH, a new carboxylesterase from *Escherichia coli*, *J. Biol. Chem.* **278**, 26039–26045.

Sawasaki, T., Ogasawara, T., Morishita, R., Endo, Y. (2002) A cell-free protein synthesis

system for high-throughput proteomics, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14652–14657.

Schneider, T.R., Sheldrick, G.M. (2002) Substructure solution with SHELXD, *Acta Crystallogr.* **D58**, 1772–1779.

Schwede, T., Kopp, J., Guex, N., Peitsch, M.C. (2003) SWISS-MODEL: an automated protein homology-modeling server, *Nucleic Acids Res.* **31**, 3381–3385.

Spriggs, R.V., Artymiuk, P.J., Willett, P. (2003) Searching for patterns of amino acids in 3D protein structures, *J. Chem. Inf. Comput. Sci.* **43**, 412–421.

Stevens, R.C., Yokoyama, S., Wilson, I.A. (2001) Global efforts in structural genomics, *Science* **294**, 89–92.

Sun, P.D., Radaev, S. (2002) Generating isomorphous heavy-atom derivatives by a quick-soak method. Part II: phasing of new structures, *Acta Crystallogr.* **D58**, 1099–1103.

Sun, P.D., Radaev, S., Kattah, M. (2002) Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases, *Acta Crystallogr.* **D58**, 1092–1098.

Szyperski, T., Yeh, D.C., Sukumaran, D.K., Moseley, H.N., Montelione, G.T. (2002) Reduced-dimensionality NMR spectroscopy for high-throughput protein resonance assignment, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8009–8014.

Terwilliger, T.C. (2000) Maximum-likelihood density modification, *Acta Crystallogr.* **D56**, 965–972.

Terwilliger, T.C., Berendzen, J. (1999) Automated MAD and MIR structure solution, *Acta Crystallogr.* **D55**, 849–861.

Thornton, J.M., Todd, A.E., Milburn, D., Borkakoti, N., Orengo, C.A. (2000) From structure to function: approaches and limitations, *Nat. Struct. Biol.* **7**(Suppl.), 991–994.

Vagin, A., Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement, *J. Appl. Crystallogr.* **30**, 1022–1025.

Voss, S., Skerra, A. (1997) Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification, *Protein Eng.* **10**, 975–982.

Walter, T.S., Diprose, J., Brown, J., Pickford, M., Owens, R.J., Stuart, D.I., Harlos, K. (2003) A procedure for setting up high-throughput nanoliter crystallization experiments. I. Protocol design and validation, *J. Appl. Crystallogr.* **36**, 308–314.

Watanabe, N., Murai, H., Tanaka, I. (2002) Semi-automatic protein crystallization system that allows in situ observation of X-ray diffraction from crystals in the drop, *Acta Crystallogr.* **D58**, 1527–1530.

Watson, J.D., Todd, A.E., Bray, J., Laskowski, R.A., Edwards, A., Joachimiak, A., Orengo, C.A., Thornton, J.M. (2003) Target selection and determination of function in structural genomics, *IUBMB Life* **55**, 249–255.

Xu, H., Hauptman, H., Weeks, C.M. (2002) Sine-enhanced shake-and-bake: the theoretical basis for applications to Se-atom substructures, *Acta Crystallogr.* **D58**, 90–96.

Yee, A., Pardee, K., Christendat, D., Savchenko, A., Edwards, A.M., Arrowsmith, C.H. (2003) Structural proteomics: toward high-throughput structural biology as a tool in functional genomics, *Acc. Chem. Res.* **36**, 183–189.

Zheng, B., Roach, L.S., Ismagilov, R.F. (2003) Screening of protein crystallization conditions on a microfluidic chip using nanoliter-size droplets, *J. Am. Chem. Soc.* **125**, 11170–11171.

Zheng, B., Tice, J.D., Roach, L.S., Ismagilov, R.F. (2004) A droplet-based, composite PDMS/glass capillary microfluidic system for evaluating protein crystallization conditions by microbatch and vapor-diffusion methods with on-chip X-ray diffraction, *Angew. Chem., Int. Ed. Engl.* **43**, 2508–2511.

Protein Translocation Across Membranes

Carla M. Koehler and David K. Hwang
University of California Los Angeles, Los Angeles, CA, USA

1 Overview of Compartmentalization 288

2 Common Principles of Translocation 289

3 Properties of Targeting Sequences 290

4 Signal-gated Translocation Systems 292
4.1 The Sec System 292
4.1.1 The Prokaryotic Sec System 292
4.1.2 The Mammalian Sec System 294
4.2 YidC is Involved in the Assembly of Membrane Proteins 294
4.3 Mitochondrial Protein Import Pathways 295
4.4 Chloroplast Protein Import Pathways 297

5 Signal-assembled Translocation Systems 298
5.1 Peroxisomal Import Pathways 298
5.1.1 Protein Import into the Matrix 299
5.1.2 Protein Import into the Membrane 301
5.2 The Tat/ΔpH Pathway of Bacteria and Thylakoids 301

6 Conclusions 303

Acknowledgment 303

Bibliography 303
Books and Reviews 303
Primary Literature 304

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

Keywords

Molecular Chaperones
Proteins that assist with the correct folding, assembly, or disassembly of other proteins *in vivo*, but are not considered components of the final functional structure.

Receptor
The receptor, usually in association with the translocation channel, recognizes the targeting sequence and directs the substrate to the translocation channel.

Signal or Targeting Sequence
A sequence or motif in a substrate for directing it to a particular location within the cell. The targeting sequence often is a conserved or consensus sequence at the N or C terminus of the protein.

Translocase or Translocon
A system catalyzing the transfer of a substrate across or into a (membrane) barrier.

Translocation Channel
A pore formed by proteins across the membrane that mediates the transport of a protein across or into a membrane.

From prokaryotic to eukaryotic cells, the movement of proteins across membranes as well as the integration into membranes is a fundamental process with shared mechanisms. Protein translocation pathways can be viewed as modular systems with basic components. Transported substrates contain targeting information that is decoded by receptors at docking sites. The substrate is then translocated across the membrane generally through a channel. A source of energy is often required and a translocation motor usually mediates translocation through the channel. In this chapter, we will summarize translocation systems in eukaryotic and prokaryotic systems, emphasizing common themes. Our understanding of protein translocation is continuing to grow with the addition of new components to existing systems and the identification of new translocation machines. The reader is encouraged to use this chapter as a starting point for more detailed excursions into scientific literature.

1 Overview of Compartmentalization

Membranes form important boundaries in cells to regulate selected exchange of molecules and ions between aqueous environments. The plasma membrane is a universal feature of all cell types and contains a wide range of transporters to mediate the exchange between the extracellular milieu and the cytoplasm. Eukaryotic cells also have developed compartmentalization

through membrane-bound organelles that specialize in various intracellular biochemical processes.

Experimental analysis using genetics and *in vitro* assays for protein targeting to eukaryotic organelles and the bacterial periplasmic membrane have revealed molecular details of the majority of protein translocation systems. The unifying principles of these systems were first proposed in the signal hypothesis two decades ago, although the components of the systems vary for each pathway. Prokaryotes first developed translocation systems for the translocation of proteins into the periplasmic space and for integration into membranes. The secretion (Sec) system mediates the translocation of proteins to the periplasmic space and has been evolutionarily conserved in the endoplasmic reticulum and chloroplast thylakoid. When the mitochondrion and chloroplast developed in the primitive eukaryotic cell, the progenitor's DNA became integrated into the host nucleus. As genes moved into the nucleus, novel translocation systems developed within the chloroplast and mitochondrion. The peroxisome also developed a translocation system seemingly independent of other organellar systems, which can translocate folded proteins.

2 Common Principles of Translocation

All translocons possess several essential features. A substrate possesses specific targeting information that directs it to the correct location. The targeting information is referred to as a *signal sequence* and is often found at the N terminus of a substrate. Molecular chaperones maintain the precursor in an import-competent state. Receptors are able to decode the information and direct the precursor to the oligomeric membrane complex, termed the *translocon*. Receptors are often bound to the membrane and may transiently associate with the translocation channel. The channel is an aqueous pore in the membrane that is gated, and mediates passage of the substrate across or into the membrane. The precursor generally passes through the channel in an unfolded state. Translocation requires an energetic input such as a membrane potential and/or nucleotide hydrolysis. A translocation motor is required to pull or push the precursor from the cis side (site of synthesis) to the trans side (destination site) of the membrane. The substrate is often processed and assembles into a multi-subunit complex.

Translocation systems can be defined as export or import systems depending upon the difference between the compartments on either side of the membrane. The export system is located in the bacterial plasma membrane, the endoplasmic reticulum, the chloroplast thylakoid membrane, and mitochondrial inner membrane. Proteins are exported from the cytosol into an extracytoplasmic compartment (i.e. the bacterial periplasm, the lumen of the endoplasmic reticulum and thylakoid, and the mitochondrial inner membrane). The import systems are located in the mitochondrial and chloroplast outer and inner membranes, and the peroxisome; this reference is based on the fact that the protein is translocated into a compartment that is functionally equivalent to the cytosol.

The translocons are remarkably flexible, able to accommodate hundreds of distinct proteins substrates while maintaining the permeability barrier of the membrane. The translocation channel must be gated to prevent solutes and ions from passing between compartments. The channel

must also recognize "stop-transfer" signals within integral membrane proteins and then gate laterally to allow diffusion of the transmembrane domains into the lipid bilayer. The channels therefore play an active role in translocation.

Two classes of translocons can be defined on the basis of the mechanism by which the substrate is translocated. The most common class, termed *signal-gated translocon* functions like the gated ion channel. The newly synthesized substrate moves through the channel in an unfolded conformation with the assistance of molecular chaperones. Because the precursor is maintained in an unfolded state, the single translocon can accommodate a wide range of substrates. The translocation reaction therefore is similar to that of ion or metabolite transport. In contrast, the second class is able to transport fully folded and/or oligomeric proteins of large dimensions while maintaining a membrane permeability barrier. This class is referred to as *signal-assembled* translocons and is present in peroxisomes, chloroplast thylakoids, and bacteria. Stable translocation channels have not been detected in these systems; thus, it has been proposed that the translocons are assembled in variable dimensions to accommodate the size of the translocated substrate.

3 Properties of Targeting Sequences

Proteins destined for transport into or across a membrane have a signal sequence, usually at the N terminus that is proteolytically cleaved on the trans side of the membrane. Depending upon the system, it is referred to as a *signal-*, *leader-*, *targeting-*, *transit-*, or *presequence*. Generally, export signal sequences are hydrophobic whereas import signal sequences are more hydrophilic. For the prokaryotic and endoplasmic reticulum export systems, the export sequences can generally be interchanged. However, import signal sequences cannot be interchanged between different organellar sequences; swapping the targeting sequence of a resident chloroplast protein to a mitochondrial targeting presequence alters trafficking from the chloroplast to the mitochondrion. Hence, import signals are important for directing a precursor to the correct organelle in the cytosol that contains several organelles including peroxisomes, mitochondria, and chloroplasts.

Signal sequences increase the specific interaction of a precursor with the appropriate transport machinery. The signal sequence may act as a folding inhibitor to maintain the transport competency of the precursor. As a precursor exits from the ribosome in *Escherichia coli*, the signal sequence inhibits folding, thus allowing the molecular chaperone SecB to bind to the precursor. SecB subsequently binds to SecA to deliver the precursor to the translocation channel. The signal sequence then interacts with components in the translocation channel. If the signal sequence is removed, export is severely inhibited, but export can be increased to normal rates by mutations in the translocation channel. Thus, SecB–SecA interactions are important for targeting, rather than the signal sequence. However, in most cases, the signal sequence acts as a true targeting signal because it is specifically recognized by cytosolic chaperones, targeting factors, or receptors.

For the bacterial and endoplasmic reticulum export systems, the targeting sequence is generally a short (~20 amino acids) sequence consisting of a hydrophobic core and short polar flanking regions

located at the N terminus. The targeting sequence is usually cleaved at the trans side of the membrane by a processing protease. If an export signal is not cleaved and possesses a hydrophobic core of 20 amino acids, the sequence can serve as a signal-anchor sequence to anchor the protein in the membrane, generating a transmembrane protein.

A protein with several membrane-spanning domains also can be generated by the sequential coordination of a signal sequence and a downstream hydrophobic stop-transfer sequence. The signal sequence serves as the initiator for translocation and the hydrophobic stop-transfer sequence arrests translocation, probably triggering a lateral opening of the translocation channel. The protein thus escapes from the translocation channel and becomes an integral membrane protein.

Mitochondria and chloroplasts contain an N-terminal bipartite targeting sequence for targeting and subsequent sorting within the organelle. Mitochondrial precursors are sorted to the matrix, inner membrane, and intermembrane space, whereas chloroplast precursors are directed to the thylakoid membrane and lumen, the stroma, and the inner envelope. The mitochondrial targeting sequence consists of an N-terminal amphipathic helix and is approximately 20 to 50 amino acids in length. One face is positively charged while the opposite face is rich in hydrophobic residues. Proteins destined for the intermembrane space contain a bipartite targeting sequence. Specifically, a hydrophobic "stop-transfer" domain to arrest translocation within the translocon is located after the targeting sequence. Import is coordinated by proteolytic cleavage to remove the N-terminal targeting domain in the matrix, followed by precursor arrest in the translocon, and subsequent proteolytic cleavage in the intermembrane space before release. In contrast, mitochondrial precursors targeted for the inner and outer membranes lack an N-terminal targeting presequence and instead, contain targeting information within the mature protein.

Chloroplast precursors contain an N-terminal transit sequence varying in length from 20 to 150 residues. Precursors destined for the thylakoids also contain a bipartite sequence in which the first N-terminal tract designates chloroplast targeting, followed by a thylakoid transfer domain. Transit sequences carry a net positive charge and are rich in hydroxylated amino acids serine and threonine. Interestingly, several transit sequences can be phosphorylated in the cytosol, which enhances import.

Peroxisomal proteins contain three different types of peroxisomal targeting sequences (PTS). For the matrix protein cohort, PTS1 and PTS2 have been characterized: PTS1 is a C-terminal three amino acid sequence (S/A/C)(K/R/H)(L/M) that has been easily identified as genome sequences have become available, whereas PTS2 is a nonapeptide sequence (R/K)(L/V/I)X_5(H/Q)(L/A) near the N terminus. For membrane proteins, a consensus sequence has not been identified, but sequences encompassing the transmembrane domains have been shown to be required for targeting so that both interact with a peroxisome chaperone and mediate insertion into the membrane.

Substrates of the twin-arginine translocation (Tat) system contain an N-terminal cleavable presequence with a twin-arginine motif containing a consensus <u>SRR</u>*X*FLK. The Tat-type signals are highly conserved throughout plant and bacterial kingdoms. In prokaryotes, the Tat and Sec systems

have to discriminate between the two types of precursors. Whereas Tat- and Sec-type targeting signals generally are similar, Tat-type signals are less hydrophobic than Sec-type signals and thus evade the Sec system.

In general, signal sequences for a given membrane transport system lack a strict consensus sequence. 25% of randomly generated peptides can function as signal sequences for various organelles, and roughly 5% of all sequences in the *E. coli* genome can direct a protein to the mitochondrion. Therefore, the secondary structure seems to be important for specific interactions with receptors.

4 Signal-gated Translocation Systems

4.1 The Sec System

The Sec system of bacteria (Fig. 1) and the endoplasmic reticulum (Fig. 2) share common properties. In addition, the chloroplast thylakoid also contains a Sec-related translocon. These translocons contain an oligomeric membrane–protein complex and associate with a translocation motor. The protein-conducting channel of this pathway is evolutionarily conserved in all kingdoms of life.

The bacterial system contains the SecYEG membrane complex that associates with the cytoplasmic ATPase, SecA. SecA functions as both a component of the cytosolic signal receptor complex and as the protein translocation motor. The SecYEG complex associates with the ribosome for cotranslational-membrane-protein insertion and with YidC, for assembly of membrane proteins. YidC has homologs in the chloroplast (Albino3) and mitochondria (Oxa1p).

4.1.1 The Prokaryotic Sec System

The bacterial Sec system utilizes cytosolic chaperones, a translocation channel, and an ATP-dependent translocation motor to push substrates into the periplasm (Fig. 1). The cytosolic SecB tetramer binds to the targeting sequence for precursors destined for export, thus functioning as a molecule chaperone. SecA is a multifunctional protein that acts as a chaperone and ATP-dependent translocation motor. In the cell, SecA is both membrane-associated and soluble. The soluble SecA also functions as a molecule chaperone; in the homodimeric state, SecA binds to the SecB–preprotein complex. When the

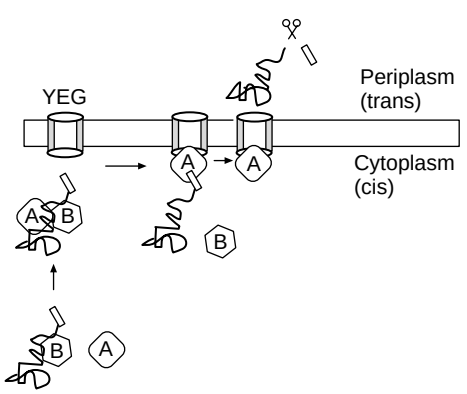

Fig. 1 The Sec translocation pathway for the export of bacterial proteins. Bacterial proteins are synthesized as precursors in the cytosol. SecB binds to the precursor and recruits SecA. The complex moves to the SecYEG translocon. SecB is released and SecA functions as a translocation motor to push the precursor through the SecYEG channel. SecA insertion is coupled with SecG inversion. Note: the channel is in a sealed conformation when not engaged or occupied by SecA.

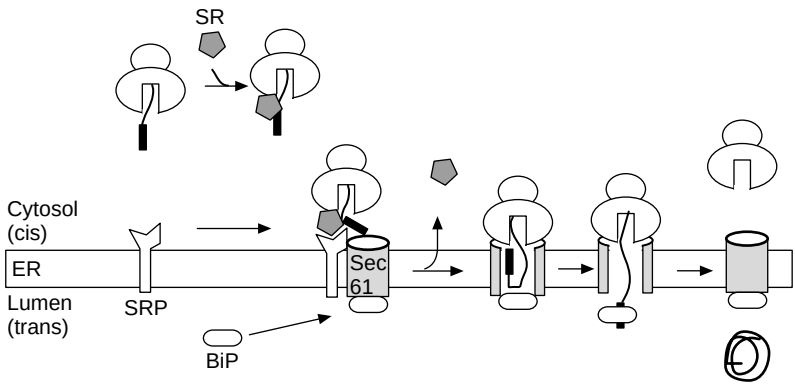

Fig. 2 The Sec cotranslational translocation pathway for the export of mammalian proteins via the endoplasmic reticulum. The soluble signal protein (SR) binds to the targeting signal as it emerges from the ribosome. The ribosome nascent chain complex is then escorted to the signal protein receptor (SRP) at the membrane. SR is released and the ribosome engages the Sec61 translocon, which causes opening of the translocon. The Hsp70 protein BiP gates the channel from the lumenal side and provides a seal to maintain membrane permeability. Translocation proceeds cotranslationally until protein synthesis terminates. The ribosome is released from the translocon and the translocon reverts to its closed state.

SecA–SecB–preprotein complex reaches the membrane, SecA associates with low affinity to negatively charged phospholipids and with high affinity to the translocation channel consisting of SecYEG. The preprotein is then transferred from SecB to SecA, and SecB is released by SecA upon the ATP-dependent initiation of translocation. The preprotein is then translocated through the SecYEG channel by multiple cycles of ATP binding and hydrolysis by SecA. Approximately 2 to 3 kDa segments are translocated per each ATP-dependent insertion reaction. The membrane potential is also required for translocation. SecA "pushing" from the cytosolic side compensates for the lack of a periplasmic energy source that could fuel molecular chaperones in the periplasmic space.

SecYEG forms the core of the protein-conducting channel and along with SecA is termed the *preprotein translocase*. SecY is the largest member of the protein-conducting channel and essential for viability and protein translocation. From experimental approaches and computer modeling, SecY contains 10 membrane-spanning domains with both N and C termini oriented to the cytosol. The fifth cytosolic loop is important for interactions with SecA. In contrast, SecE is a 14-kDa protein with three membrane-spanning domains in *E. coli*; the N terminus localizes to the cytosol while the C terminus faces the periplasm. Interestingly, other bacteria contain a SecE version that is homologous to only the third membrane-spanning domain, thus suggesting that this is an important region in SecE. SecG associates loosely with the protein-conducting channel and contains two membrane-spanning domains. SecG is an unusual protein because it undergoes an inversion in topology upon the membrane insertion of SecA. Although SecG is not essential for viability, it is required for efficient export at low temperatures.

What is the oligomeric state of the protein-conducting channel? The crystal structure from the archaeon *Methanococcus jannaschii* revealed a monomeric complex. However, the diameter of the channel is too narrow to accommodate SecA. Electron microscopy studies with purified SecYEG complex suggested a larger-sized complex of ~85 by ~65 Å with a pore of ~20 Å; the size and mass of these particles may reflect a dimeric or tetrameric form. From experimental studies including analytical ultracentrifugation and blue native gel analysis, it was revealed that the complex forms a dimer. The channel thus seems to have several different oligomeric assemblies. Further structural studies with the channel engaging SecA are required to provide a picture of the translocase in action.

4.1.2 The Mammalian Sec System

The mammalian system contains the SecYEG homologs Sec61α, Sec61β, and Sec61γ subunits, in addition to a unique subunit TRAM that is required for membrane protein integration (Fig. 2). The mammalian translocon forms an aqueous pore that spans the ER (endoplasmic reticulum) membrane, and Sec61α membrane-spanning domains line the channel similar to SecY. Ribosomes synthesizing membrane or secretory proteins are earmarked by a signal sequence at the N terminus of the nascent chain that binds to the signal recognition particle (SRP). SRP directs the ribosome nascent chain complex (RNC) to the signal receptor (SR) at the ER membrane. GTP-dependent interactions then elicit binding of the RNC directly to the translocon, followed by transfer of the nascent chain into the channel.

How is the protein translocated into the channel? The lumenal side of the pore is closed by the Hsp70 chaperone BiP. After RNC docking and elongation of the translocating chain, BiP is released from the pore, allowing passage of the chain into the ER lumen. The RNC then maintains an ion-tight seal on the cytoplasmic side. If a protein contains several membrane-spanning domains, BiP at the lumenal side and the RNC at the cytosolic side alternately gate the pore through a series of orchestrated changes, thus maintaining an ion-tight barrier between the lumen and cytoplasm during translocation of the protein. Specifically, as the nascent chain increases in length, BiP is released from the channel, while the ribosome maintains an ion-tight seal. Upon the detection of a membrane-spanning domain, the luminal end of the pore is closed by BiP, allowing the ribosome to then lift from the translocation channel. Translocated membrane proteins most likely are released laterally into the membrane via parting of the membrane-spanning domains in the Sec channel. The inner diameter of the pore changes dynamically from 9–15 Å in the ribosome-free closed state to 40–60 Å upon ribosome docking. The flexibility of the channel may be required to accommodate multiple membrane-spanning domains, which may leave the translocon in groups.

4.2 YidC is Involved in the Assembly of Membrane Proteins

YidC cooperates with the bacterial Sec system and operates independently to mediate translocation of integral membrane proteins. YidC is an integral membrane protein with six membrane-spanning domains and a large periplasmic domain. YidC interacts with the SecYEG translocon to facilitate integration of membrane proteins during translocation. YidC also acts

independently to mediate the insertion of subunits in the energy-transducing complexes such as the F_0 domain of the F_1F_0 ATPase. The chloroplast homolog albino3 is required for protein insertion into the thylakoid membrane and the mitochondrial homolog Oxa1p is required for protein insertion into the inner membrane.

4.3 Mitochondrial Protein Import Pathways

The mitochondrion contains two membranes that separate the mitochondrial matrix from the intermembrane space (Fig. 3). The inner membrane is essentially impermeable to ions because a membrane potential must be maintained for respiration. Most proteins are synthesized on cytosolic ribosomes and imported cotranslationally or posttranslationally into the mitochondrion. Proteins coded on the mitochondrial genome, however, are exported.

The mitochondrion contains a small genome and these proteins are exported to the inner membrane via Oxa1p. Oxa1p is homologous to YidC, but contains five membrane-spanning domains and a large C-terminal domain in the matrix. The ribosome docks on the C-terminal domain to facilitate translocation of newly synthesized hydrophobic membrane-spanning

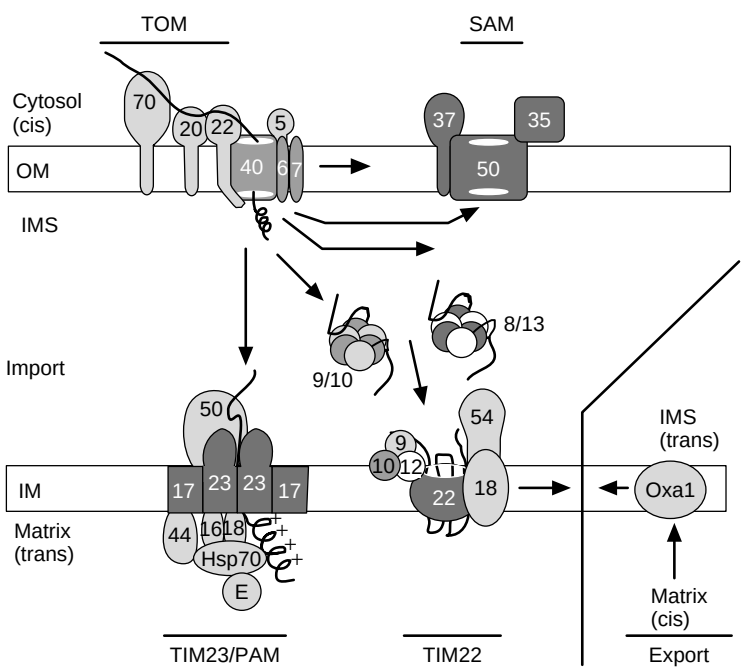

Fig. 3 Translocation pathways of the mitochondrion. Most proteins are imported from the cytosol and pass through the TOM complex. Single pass membranes are released from the TOM channel, whereas β-barrel proteins are sorted to the SAM complex for assembly. Proteins with a typical N-terminal targeting sequence are imported via the TIM23 translocon, whereas inner membrane proteins lacking an N-terminal targeting domain are imported via the TIM22 translocation system. Proteins coded by the mitochondrial genome are exported via Oxa1p.

domains into the membrane. Translocation depends upon the membrane potential. Structural studies suggest that Oxa1p functions as a tetramer, but whether it forms a translocation channel has not been established.

Most proteins are imported from the cytosol. The outer membrane contains the translocase of the outer membrane (TOM) complex through which all precursors are translocated. Cytosolic chaperones such as mitochondrial import stimulation factor and presequence binding factor have been identified in mammals, and the cytosolic chaperones Hsp70 and Hsp90 bind to the Tom70p receptor. Cotranslational import also has been implicated because mRNAs of prokaryotic origin have a signal in the 3′ untranslated region that targets them to the mitochondrion.

The TOM complex contains both, receptors and the translocation pore. The receptor Tom20p binds directly to the N-terminal presequence. In contrast, Tom70p mediates import of precursors that contain targeting information within the mature part of the protein. The precursors are then directed to the translocation pore formed by Tom40p. Small Tom proteins, Tom6p and Tom7p, are thought to gate the pore and allow proteins with one or two membrane-spanning domains to escape laterally into the lipid bilayer.

Proteins such as Tom40p and porin are β-barrel proteins and recently have been shown to be assembled by the sorting and assembly machinery (SAM) complex for outer membrane β-barrel proteins. The SAM complex consists of Sam50p, Sam35p, and Sam37p. Whereas Sam50p is also a β-barrel protein, Sam37p contains one membrane-spanning domain, and Sam35p is peripherally associated with Sam50p on the cytosolic side. The molecular mechanisms of this pathway remain to be elucidated, but preliminary studies suggest that the precursor is passed from the TOM complex via the intermembrane space to the SAM complex. Bacteria contain a homologous protein Omp85, which most likely assembles β-barrel proteins in the outer membrane and may be required for lipid biogenesis. Thus, the SAM pathway seems to be evolutionarily conserved.

For proteins with a typical N-terminal targeting presequence, they are imported from the TOM complex to the TIM23 complex. The TIM23 complex forms the membrane translocon and associates with the protein associated motor (PAM). The TIM23 complex contains the receptors Tim50p and the N-terminal domain of Tim23p, whereas Tim17p and the C-terminal domain of Tim23p form the translocation channel. A membrane potential is required to initiate translocation. The PAM then completes translocation in an ATP-dependent manner. The driving force is generated by the ATPase mHsp70, which is anchored to the membrane by Tim44p. The nucleotide exchange factor Mge1p promotes the reaction cycle of mHsp70, thereby allowing nucleotide release. Pam16p and Pam18p are associated J-proteins that stimulate the ATPase cycle. Upon translocation into the matrix, the targeting sequence may be proteolytically cleaved, and a battery of chaperones facilitates folding of proteins and assembly into functional protein complexes.

Proteins destined for the inner membrane utilize the TIM22 pathway. Precursors of this pathway include the mitochondrial carrier family such as the ADP/ATP carrier, and the phosphate carrier, as well as Tim17p, Tim22p, and Tim23p. In contrast to precursors with an N-terminal targeting presequence, these precursors contain targeting information within the mature part of the protein and cross the

TOM complex as a loop. The estimated size of the TOM translocon is ~20–26 Å, which can accommodate two unfolded polypeptide chains. As the hydrophobic precursors cross the TOM complex and enter the aqueous intermembrane space, the small Tim proteins escort the complexes to the TIM22 insertion complex at the inner membrane. The small Tim proteins are a family of five proteins in yeast. Tim8p partners with Tim13p and Tim9p partners with Tim10p to form 70-kDa complexes in the intermembrane space. The small Tim proteins bind to hydrophobic sequences in the inner membrane substrates and escort them to the insertion complex. The insertion complex contains the membrane proteins Tim22p, Tim54p, and Tim18p and associated Tim12p with a fraction of the Tim9p and Tim10p. Tim22p forms a twin-pore translocase that mediates import into the inner membrane. The mechanism of insertion into the membrane has not been elucidated, but a membrane potential is required. The auxiliary proteins Tim18p and Tim54p most likely play a role in complex assembly and stability.

4.4 Chloroplast Protein Import Pathways

The chloroplast organization is the most complex of all organelles. Proteins can be imported across three distinct membranes, the chloroplast outer and inner envelope membranes and the thylakoid membrane, and into three soluble subcompartments – the space between envelope membranes, the stroma, and the thylakoid lumen (Fig. 4). The translocase of the outer envelope (TOC) and the translocase of the inner envelope (TIC) coordinate the import of chloroplast proteins. Proteins are imported to the thylakoids by one of three different systems, the Sec pathway, the SRP pathway, and the Tat pathway. Additionally, proteins may insert spontaneously into the thylakoid membrane.

Most precursors are imported posttranslationally from the cytosol, except for proteins that are coded on the chloroplast genome. As the precursor emerges from the ribosome, molecular chaperones of the Hsp70 family prevent aggregation and maintain an unfolded import-competent state. The precursor may be phosphorylated by a cytosolic ATP-dependent protein kinase to "fast-track" import. Toc159 and Toc34 are GTPases and form the receptor system for directing the precursor to the membrane. Toc159 cycles between the cytosol and the TOC complex while Toc34 remains associated with the β-barrel channel Toc75. The GTP requirement has not been elucidated, but recent studies suggest a coordinated cycle in which GTP-dependent substrate recognition is coupled to GTP-driven translocation across the outer envelope.

The TOC complex has a molecular mass of approximately 550 kDa consisting of a stoichiometry of 4 Toc34 : 4 Toc75 : 1 Toc159. Toc75 is the most abundant outer envelope protein with a predicted topology of 16 transmembrane β-sheets. Toc75 forms a cation-selective ion channel when reconstituted into planar lipid bilayers. On the basis of electrophysiological measurements, the channel is an estimated 25 Å at the entrance and 15 to 17 Å in the middle, which could accommodate a single polypeptide chain. An accessory protein Toc64 might serve as a docking site for hydrophobic precursors bound to chaperone Hsp70.

The TIC complex is suggested to associate simultaneously with the TOC complex during translocation. In contrast

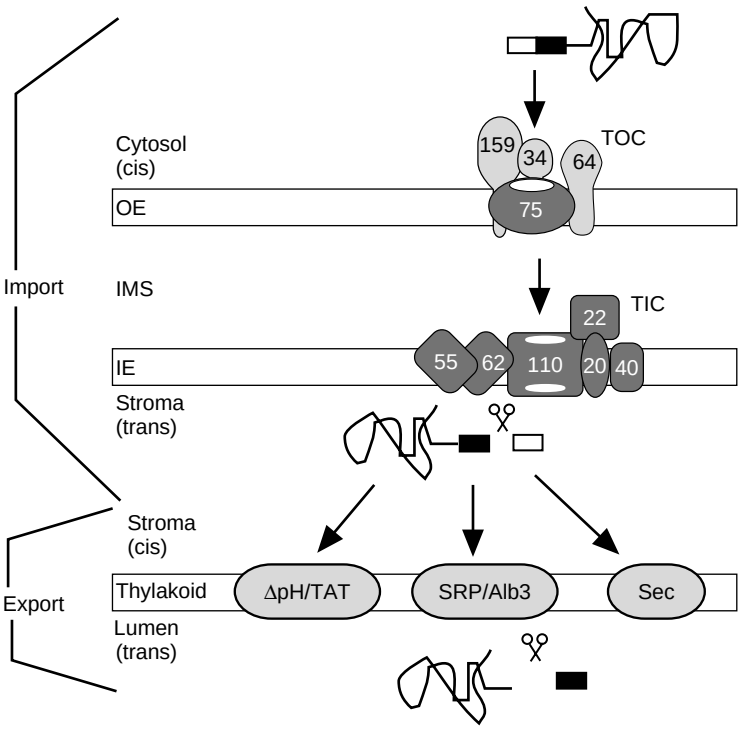

Fig. 4 Translocation pathways of the chloroplast. Precursors are imported from the cytosol and pass through the TOC complex and the TIC complex. In the stroma, the processing peptidase removes the transit sequence; for precursors destined for export to the thylakoid, cleavage of the transit sequences reveals the thylakoid targeting sequence for export via one of three systems.

to the GTP requirement at the TOC complex, ATP hydrolysis is required for translocation across the inner membrane, probably for stromal molecular chaperones that function as a translocation motor. The TIC complex contains several members including Tic110, Tic62, Tic55, Tic40, Tic22, and Tic20 and has a molecular weight of 250 kDa. Of these components, Tic110 is the best characterized and seems to bind to the TOC complex and stromal chaperones Hsp93 and chaperonin 60. Tic62 and Tic55 are redox components that may link the metabolic status of the chloroplast (based on the NAD(P)H:NAD(P) ratio) to the import capacity of the chloroplast. In *Arabidopsis thaliana*, many of the TIC and TOC components belong to multigene families, allowing complex regulation and reinforcing the importance of import in chloroplast biogenesis.

5 Signal-assembled Translocation Systems

5.1 Peroxisomal Import Pathways

Peroxisomes are single membrane-bound organelles that specialize in β-oxidation of fatty acids, hydrogen peroxide-based

respiration, and are a defense against oxidative stress. They are found in virtually every eukaryotic cell, with a typical size of 0.1–1 μ m in diameter. In plants, peroxisomes are specialized as glyoxysomes for the glyoxylate cycle, leaf peroxisomes involved in photorespiration pathways, and unspecialized peroxisomes. In mammals, peroxisomes are integral for fatty acid oxidation, and defects in its assembly lead to a class of inherited diseases referred to as *Zellweger's syndrome* or *peroxisome biogenesis disorders* (PBDs). Genetic and biochemical studies in yeast and complementation studies in cells derived from Zellweger's syndrome patients have proved instrumental in identifying many of the proteins involved in peroxisomal biogenesis, termed *peroxins* and coded by PEX genes. Approximately 25 to 30 peroxins have been identified and include both soluble and membrane proteins.

As with other organelles, events in biogenesis include import into the matrix and import of membrane proteins. The matrix protein pathway has been best characterized, but the current mechanistic understanding is not as refined as that of other organelles, because *in vitro* manipulations with peroxisomes are difficult: the organelles are difficult to purify and a robust import assay has not been developed. Biogenesis studies, however, have been advanced by groups working in several experimental systems. The striking differences of peroxisomal biogenesis in comparison with other organelles are that a typical translocation motor and channel have not been identified, and folded proteins can be imported. Additionally, peroxisomes do not possess a membrane potential.

5.1.1 Protein Import into the Matrix

Peroxisomal matrix proteins are synthesized on free ribosomes in the cytosol and imported posttranslationally into peroxisomes, emphasized with PTS1 precursors that have the targeting information at the C terminus. From experimental systems, the simple PTS1 can target a nonresident protein such as green fluorescent protein to the peroxisomal matrix, verifying that the PTS1 sequence is necessary and sufficient for targeting to peroxisomes. The import of matrix proteins depends on cytosolic receptors Pex5p and Pex7p, for PTS1 and PTS2 precursors, respectively (Fig. 5). Pex5p consists of two domains; the C-terminal half contains a high-affinity PTS1-binding site consisting of seven tetratricopeptide repeats (TPR), whereas the N-terminal half mediates peroxisomal targeting and protein–protein interactions at the organelle. Pex7p contains WD-40 motifs and works in conjunction with accessory proteins to guide the precursor to the peroxisome. The cytosolic chaperones thus bind to their cargo and the complexes dock independently at the peroxisome surface.

The docking platform consists of Pex13p, Pex14p, and Pex17p, of which Pex13p and Pex14p are integral peroxisomal membrane proteins (PMP). Both Pex5p and Pex7p converge to this dock, implicating a single translocon for both types of cargo. Peripherally associated Pex17p forms a tight core with PMP Pex14p. Pex13p exposes the N and C termini to the cytosol and recognizes both Pex5p and Pex7p. Pex7p binds to the N terminus, and Pex5p and Pex14p binds to the C terminus, which contains an Src-homology-3 (SH3) domain.

A second group of membrane proteins, Pex2p, Pex10p, and Pex12p (the

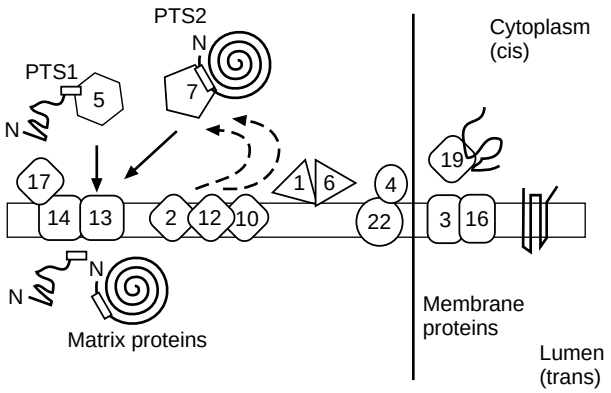

Fig. 5 Translocation pathways of the peroxisome. Pex5p chaperones PTS1 precursors and Pex7p chaperones PTS2 precursors to the docking complex consisting of Pex13p, Pex14p, and Pex17p. The precursor is translocated into the thylakoid lumen and Pex5p and Pex7p are recycled to the cytosol. Peroxisomal membrane proteins are chaperoned to the membrane by Pex19p and associate with Pex3p and Pex16p for insertion into the membrane. Peroxisome translocation mechanisms are not well understood but folded, proteins can be translocated.

RING-finger peroxins), function downstream from the receptor complexes. All three components contain cytoplasmic zinc RING domains, and Pex10p and Pex12p bind to Pex5p. How does Pex5p deliver cargo to the peroxisomal matrix? The mechanism by which Pex5p releases cargo has been under debate because Pex5p has been localized to the cytoplasm, peroxisomal membrane, and peroxisomal matrix. It has been suggested that Pex5p shuttles between the cytoplasm and peroxisomal matrix. In the "extended-shuttle" mechanism, Pex5p undergoes multiple cycles of translocation in and out of the matrix, whereas in the "simple shuttle" mechanism, Pex5p stays at the surface of the membrane and releases the cargo. The consensus is that Pex5p traverses the mammalian peroxisomal membrane through multiple rounds of entry into the matrix and export to the cytosol. Further evidence that Pex5p at least traverses the membrane is confirmed by an interaction between Pex5p and Pex8p, the only known peroxin localized to the trans side of the membrane in *Saccharomycescerevisiae*.

Additional peroxins, Pex4p, Pex22p, Pex1p, and Pex6p, are required to unload the cargo and return the receptor to the cytoplasm. Pex22p is a PMP that serves as a membrane anchor for cytosolic Pex4p, an E2-type ubiquitin conjugating enzyme. The RING-finger domains of Pex10p may interact with Pex4p to link the docking complex. ATP hydrolysis is required for import of matrix proteins, providing the potential for a nucleotide-hydrolysis–driven import cycle. Pex1p and Pex6p are two membrane-associated ATPases of the AAA family (ATPases Associated with diverse cellular Activities). Genetic studies place the requirement for Pex1p and Pex6p between receptor docking/translocation and Pex4p/Pex22p,

suggesting that Pex1p and Pex6p are required for unloading the cargo and recycling the receptor.

5.1.2 Protein Import into the Membrane

Matrix protein and PMP import seem to be independent. In contrast to matrix protein import, PMP import does not require ATP hydrolysis, and mutations in genes that abrogate matrix import usually do not affect membrane import. Furthermore, a translocation intermediate that blocked the binding of matrix proteins to isolated glyoxysomes did not affect PMP import.

Thus, of most *pex* mutants in which matrix import is abolished, PMPs are still inserted into peroxisome remnants or "ghosts" because they lack matrix proteins.

However, in cells that are deficient in the PMPs Pex3p, Pex16p, and Pex19p, peroxisome ghosts are not detectable, suggesting that this cohort defines the essential components for PMP import and sorting. Supporting this, mutant cells are seen to lack *pex3* or *pex19* and have unstable PMPs. And these three peroxins interact *in vitro* and *in vivo*.

Like peroxisomal matrix protein import, PMP import seems to share common features (Fig. 5). Pex19p is predominantly cytosolic, but can be found associated with the peroxisomal membrane, possibly via farnesylation at the C terminus. Pex19p thus acts as a chaperone to prevent PMP degradation and to guide PMPs in the cytosol to the peroxisome. As expected, Pex19p binds to several PMPs. In addition to its chaperone role, Pex19p subsequently acts as an import receptor, binding to the docking protein Pex3p. Pex3p thus acts as a docking protein, and as expected, Pex3p inserts independently of Pex19p. A specific role for Pex16p in PMP targeting is not known, but Pex16p may function at an early stage of peroxisomal membrane assembly upstream of Pex3p and Pex19p. Thus, peroxisomal import pathways share the common feature of a cytosolic chaperone, which targets both membrane and matrix proteins to docking sites on the membrane.

5.2 The Tat/ΔpH Pathway of Bacteria and Thylakoids

The Tat system (also referred to as the ΔpH pathway) is conserved in bacteria and chloroplast thylakoids (Fig. 6). The pathway was first characterized in plant chloroplasts in the early 1990s when biochemical studies showed that the pH gradient, rather than a nucleotide hydrolysis,

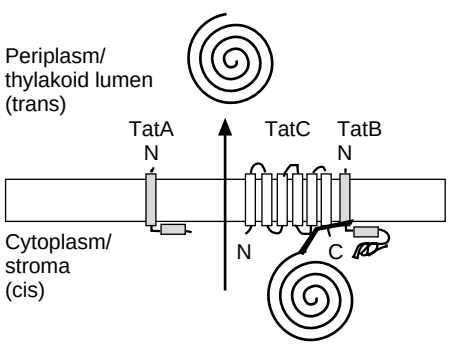

Fig. 6 The Tat/ΔpH translocation pathway for folded proteins. The Tat pathway of the thylakoid lumen and bacteria share common features. TatB and TatC assemble into a complex that binds to the Tat signal sequence. TatA is then recruited and the folded protein is translocated across the membrane in a membrane potential/ΔpH-dependent manner. A nucleotide requirement and chaperones on the cis and trans sides of the membrane have not been identified.

was required for the import of a subset of thylakoid lumenal proteins. The observation of the pathway in prokaryotes led to rapid identification of its important components. The Tat pathway mediates the translocation of large, folded proteins, such as proteins with a range of redox cofactors, including FeS, NiFe, and molybdopterin centers, across the tightly sealed membrane. In some cases, the oligomeric substrates employ a ''hitchhiker'' mechanism in which only one or a subset of subunits contain targeting signals. In most prokaryotes, the Sec system is the main route for export, but in the halophilic archaeon Halibacterium sp NRC-1, the majority of secretory proteins including traditional Sec-dependent substrates, use the Tat system. The rationale for this adaptation is that the haloarchea has a very high cytoplasmic salt concentration (4–5 M) and newly synthesized proteins must fold quickly to prevent aggregation; hence, the Tat system is required for export of these folded cytoplasmic proteins.

How does the Tat system translocate a folded protein across an energy-transducing membrane? With chloroplast thylakoid import studies, a precursor was constructed with a Tat signal fused to bovine pancreatic trypsin inhibitor (BPTI), which was internally cross-linked to prevent unfolding. This construct imported into isolated chloroplasts and targeted to the thylakoid lumen by the Tat pathway. In a second study, a larger translocation substrate was generated using the reporter dihydrofolate reductase (DHFR), which when folded binds to the folate analog methotrexate (MTX) with very high affinity. This translocation intermediate normally arrests in the mitochondrial TOM or bacterial Sec system because it cannot be unfolded, thus confirming that most translocons must completely unfold proteins prior to translocation. However, the folded DHFR-MTX construct was imported efficiently into the thylakoid lumen, and folding was confirmed because the DHFR–MTX complex was protease resistant. In bacterial systems, GFP (green fluorescent protein) fusions also were used to confirm that folded proteins were exported via the Tat system. Specifically, GFP that is engineered for export via the Sec system does not fold properly in the periplasm and does not fluoresce. In contrast, Tat-targeted GFP exported to the periplasm was folded and detected as a fluorescent halo on the cell periphery, confirming efficient translocation of a folded protein.

Because the Tat pathway is organized in one operon in *E. coli*, characterization of the translocation components was facilitated by identification of the proteins within the operon. The *E. coli* system has three Tat proteins (the tatABC operon) and plants have three (Tha4, Hcf106, and cpTatC for tatA, B, and C respectively) homologs. All Tat proteins are located in the membrane. TatA and TatB both contain one transmembrane domain (N terminus facing the periplasm), whereas TatC contains six membrane-spanning domains (N and C termini exposed to the cytoplasm). The TatABC complex is approximately 500 to 600 kDa in size and consists of approximately a 1 : 11 ratio for the individual components, although TatA is present in higher amounts or in a separate complex. From single particle electron microscope images, the TatABC complex has dimensions of 10×13 nm. TatB/Hcf106 and TatC form a stable membrane complex that binds the Tat signal. This then triggers the recruitment of TatA to form the functional translocation system, which requires a membrane potential. For the actual

translocation of the substrate, additional factors on the cis and trans side of the membrane seemingly are not required. The current available evidence suggests that the force may be applied within the membrane domain or by structural changes within the Tat translocon itself. After translocation, dissociation of the Tat translocon has been observed, which is consistent with the reversible assembly of the translocon.

Whereas the specific mechanism of transport remains to be elucidated, the transport of large folded proteins is very efficient, but costly. Careful measurements of the prevailing ΔpH during translocation of thylakoid proteins has shown that translocation requires 30 000 protons per protein transported. Moreover, two differently sized substrates required differing levels of ΔpH for translocation, suggesting that the Tat system adapts energetically to the size and/or shape of the substrate. Indeed, future mechanistic studies should provide new ideas about how large folded proteins can be exported across an energy-transducing membrane.

6 Conclusions

Significant progress has been made toward understanding the mechanism of protein translocation across membranes. Yet, there are major questions to be addressed. The events that initiate the translocation reaction have not been deciphered in all translocation systems. The mechanism by which the channel can accommodate precursors of varying sizes, particularly for the signal-assembled translocon, remains a mystery. Moreover, how defects in the translocation systems result in the diseased state in plants and animals are beginning to be elucidated. Eventually, a greater understanding of the mechanisms of translocation will provide information about the broader question of how membrane components assemble.

Acknowledgment

David Hwang was supported by the United States Public Health Service National Research Service Award (GM07185). Carla M. Koehler was supported by grants from the Beckman Foundation, National Institutes of Health, and the Muscular Dystrophy Association.

See also Membrane Traffic: Vesicle Budding and Fusion; Protein Mediated Membrane Fusion.

Bibliography

Books and Reviews

Bauer, J., Hiltbrunner, A., Kessler, F. (2001) Molecular biology of chloroplast biogenesis: gene expression, protein import and intraorganellar sorting, *Cell Mol. Life Sci.* **58**, 420–433.

Blobel, G. (1980) Intracellular protein topogenesis, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1496–1500.

Eckert, J.H., Erdmann, R. (2003) Peroxisome biogenesis, *Rev. Physiol. Biochem. Pharmacol.* **147**, 75–121.

Johnson, A.E., van Waes, M.A. (1999) The translocon: a dynamic gateway at the ER membrane, *Annu. Rev. Cell Dev. Biol.* **15**, 799–842.

Koehler, C.M. (2004) New developments in mitochondrial assembly, *Annu. Rev. Cell Dev. Biol.* **20**, 309–335.

Robinson, C., Bolhuis, A. (2004) Tat-dependent protein targeting in prokaryotes

and chloroplasts, *Biochim. Biophys. Acta* **1694**, 135–147.

Schatz, G., Dobberstein, B. (1996) Common principles of protein translocation across membranes, *Science* **271**, 1519–1526.

Schnell, D.J., Hebert, D.N. (2003) Protein translocons: multifunctional mediators of protein translocation across membranes, *Cell* **112**, 491–505.

Soll, J., Schlieff, E. (2004) Protein import into chloroplasts, *Natl. Rev.* **5**, 198–208.

Walter, P., Johnson, A.E. (1994) Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane, *Annu. Rev. Cell Biol.* **10**, 87–119.

Primary Literature

Abe, Y., Shodai, T., Muto, T., Mihara, K., Nishikawa, S.-I., Endo, T., Kohda, D. (2000) Structural basis of presequence recognition by the mitochondrial presequence receptor Tom20, *Cell* **100**, 551–560.

Akita, M., Shinkai, A., Matsuyama, S., Mizushima, S. (1991) SecA, an essential component of the secretory machinery of *Escherichia coli*, exists as homodimer, *Biochem. Biophys. Res. Commun.* **174**, 211–216.

Alder, N.N., Theg, S.M. (2003) Energetics of protein transport across biological membranes. A study of the thylakoid DeltapH-dependent/cpTat pathway, *Cell* **112**, 231–242.

Attardi, G., Schatz, G. (1988) Biogenesis of mitochondria, *Annu. Rev. Cell Biol.* **4**, 289–333.

Baker, A., Schatz, G. (1987) Sequences from a prokaryotic genome or the mouse dihydrofolate reductase gene can restore the import of a truncated precursor protein into yeast mitochondria, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3117–3121.

Becker, T., Jelic, M., Vojta, A., Radunz, A., Soll, J., Schlieff, E. (2004) Preprotein recognition by the Toc complex, *EMBO J.* **23**, 520–530.

Berks, B.C., Palmer, T., Sargent, F. (2003) The Tat protein translocation pathway and its role in microbial physiology, *Adv. Microb. Physiol.* **47**, 187–254.

Bessonneau, P., Besson, V., Collinson, I., Duong, F. (2002) The SecYEG preprotein translocation channel is a conformationally dynamic and dimeric structure, *EMBO J.* **21**, 995–1003.

Blobel, G., Dobberstein, B. (1976) Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on the membrane-bound ribosomes of murine myeloma, *J. Cell Biol.* **67**, 835–851.

Bolhuis, A. (2002) Protein transport in the halophilic archaeon Halobacterium sp. NRC-1: a major role for the twin-arginine translocation pathway? *Microbiology* **142**, 3335–3346.

Brink, S., Bogsch, E.G., Edwards, W.R., Hynds, P.J., Robinson, C. (1998) Targeting of thylakoid proteins by the ΔpH-driven twin-arginine pathway requires a specific signal in the hydrophobic domain in conjunction with the twin-arginine motif, *FEBS Lett.* **434**, 425–430.

Brix, J., Ziegler, G.A., Dietmeier, K., Schneider-Mergener, J., Schulz, G.E., Pfanner, N. (2000) The mitochondrial import receptor Tom70: identification of a 25 kDa core domain with a specific binding site for preproteins, *J. Mol. Biol.* **303**, 479–488.

Caliebe, A., Grimm, R., Kaiser, G., Lubeck, J., Soll, J., Heins, L. (2003) The chloroplastic protein import machinery contains a Rieske-type iron-sulfur cluster and a mononuclear iron-binding protein, *EMBO J.* **278**, 38617–38627.

Chen, X., Schnell, D.J. (1999) Protein import into chloroplasts, *Trends Cell Biol.* **9**, 222–227.

Clark, S.A., Theg, S.M. (1997) A folded protein can be transported across the chloroplast envelope and thylakoid membranes, *Mol. Biol. Cell* **8**, 923–934.

Cline, K., Mori, H. (2001) Thylakoid delta pH-dependent precursor proteins bind to a cpTatC-Hcf106 complex before Tha4-dependent transport, *J. Cell Biol.* **154**, 719–729.

Collins, C.S., Kalish, J.E., Morrell, J.C., McCaffery, J.M., Gould, S.J. (2000) The peroxisome biogenesis factors pex4p, pex22p, pex1p, and pex6p act in the terminal steps of peroxisomal matrix protein import, *Mol. Cell Biol.* **20**, 7516–7526.

Cristobal, S., de Gier, J.W., Nielsen, H., von Heijne, G. (1999) Competition between Sec- and TAT-dependent protein translocation in *Escherichia coli*, *EMBO J.* **18**, 2982–2990.

Crowley, K.S., Liao, S., Worrell, V.E., Reinhart, G.D., Johnson, A.E. (1994) Secretory proteins move through the endoplasmic reticulum membrane via an aqueous, gated pore, *Cell* **78**, 461–471.

Dammai, V., Subramani, S. (2001) The human peroxisomal targeting signal receptor, Pex5p, is translocated into the peroxisomal matrix and recycles to the cytosol, *Cell* **105**, 187–196.

Diestelkotter, P., Just, W.W. (1993) *In vitro* insertion of the 22 kDa peroxisomal membrane protein into isolated rat liver peroxisomes, *J. Biol. Chem.* **123**, 1717–1725.

Dodt, G., Braverman, N., Valle, D., Gould, S.J. (1996) From expressed sequence tags to peroxisome biogenesis disorder genes, *Ann. N. Y. Acad. Sci.* **804**, 516–523.

Driessen, A.J. (1993) SecA, the peripheral subunit of the *Escherichia coli* precursor protein translocase, is functional as a dimer, *Biochemistry* **32**, 13190–13197.

D'Silva, P.D., Schilke, B., Walter, W., Andrew, A., Craig, E.A. (2003) J protein cochaperone of the mitochondrial inner membrane required for protein import into the mitochondrial matrix, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13839–13844.

Eichler, J., Duong, F. (2004) Break on through to the other side – the Sec translocon, *Trends Biochem. Sci.* **29**, 221–223.

Endo, T., Yamamoto, H., Esaki, M. (2003) Functional cooperation and separation of translocators in protein import into mitochondria, the double-membrane bounded organelles, *J. Cell. Sci.* **116**, 3259–3267.

Feilmeier, B.J., Iseminger, G., Schroeder, D., Webber, J., Phillips, G.J. (2000) Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*, *J. Bacteriol.* **182**, 4068–4076.

Fekkes, P., de Wit, J.G., van der Wolk, J.P., Kimsey, H.H., Kumamoto, C.A., Driessen, A.J.M. (1998) Preprotein transfer to the *Escherichia coli* translocase requires the cooperative binding of SecB and the signal sequence to SecA, *Mol. Microbiol.* **29**, 1179–1190.

Flugge, U.I., Hinz, G. (1986) Energy dependence of protein translocation into chloroplasts, *Eur. J. Biochem.* **160**, 563–567.

Fransen, M., Wylin, T., Brees, C., Mannaerts, G.P., Van Veldhoven, P.P. (2001) Human Pex19p binds peroxisomal integral membrane proteins at regions distinct from their sorting sequences, *Mol. Cell. Biol.* **21**, 4413–4424.

Frazier, A.E., Dudek, J., Guiard, B., Voos, W., Li, Y., Lind, M., Meisinger, C., Geissler, A., Sickmann, A., Meyer, H.E., Bilanchone, V., Cumsky, M.G., Truscott, K.N., Pfanner, N., Rehling, P. (2004) Pam16 has an essential role in the mitochondrial protein import motor, *Nat. Struct. Mol. Biol.* **11**, 226–233.

Gatto, G.J. Jr., Geisbrecht, B.V., Gould, S.J., Berg, J.M. (2000) Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5, *Nat. Struct. Biol.* **7**, 1091–1095.

Ghys, K., Fransen, M., Mannaerts, G.P., Van Veldhoven, P.P. (2002) Functional studies on human Pex7p: subcellular localization and interaction with proteins containing a peroxisome-targeting signal type 2 and other peroxins, *Biochem. J.* **365**, 41–50.

Gould, S.J., Valle, D. (2000) Peroxisome biogenesis disorders: genetics and cell biology, *Trends Genet.* **16**, 340–345.

Gould, S.J., Collins, C.S. (2002) Opinion: peroxisomal protein import: Is it really that complex? *Nat. Rev. Mol. Cell Biol.* **3**, 382–389.

Gray, M.W., Burger, G., Lang, B.F. (1999) Mitochondrial Evolution, *Science* **283**, 1476–1481.

Haigh, N.G., Johnson, A.E. (2002) A new role for BiP: closing the aqueous translocon pore during protein integration into the ER membrane, *J. Cell Biol.* **156**, 261–270.

Hartl, F.U., Lecker, S., Schiebel, E., Hendrick, J.P., Wickner, W. (1990) The binding cascade of SecB to SecA to SecY/E mediates preprotein targeting to the *E. coli* plasma membrane, *Cell* **63**, 269–279.

Hell, K., Neupert, W., Stuart, R.A. (2001) Oxa1p acts as a general membrane insertion machinery for proteins encoded by mitochondrial DNA, *EMBO J.* **20**, 1281–1288.

Hettema, E.H., Girzalsky, W., van Den Berg, M., Erdmann, R., Distel, B. (2000) *Saccharomyces cerevisiae* pex3p and pex19p are required for proper localization and stability of peroxisomal membrane proteins, *EMBO J.* **19**, 223–233.

Hinnah, S.C., Hill, K., Wagner, R., Schlicher, T., Soll, J. (1997) Reconstitution of a chloroplast protein import channel, *EMBO J.* **16**, 7351–7360.

Holroyd, C., Erdmann, R. (2001) Protein translocation machineries of peroxisomes, *FEBS Lett.* **501**, 6–10.

Honsho, M., Hiroshige, T., Fujiki, Y. (2002) The membrane biogenesis of Pex16p – topogenesis and functional roles in peroxisomal membrane assembly, *J. Biol. Chem.* **277**, 44513–44524.

Huhse, B., Rehling, P., Albertini, M., Blank, L., Meller, K., Kunau, W.H. (1998) Pex17p of *Saccharomyces cerevisiae* is a novel peroxin and component of the peroxisomal protein

translocation machinery, *J. Cell Biol.* **140**, 49–60.

Imanaka, T., Shinina, Y., Takano, Y., Hashimoto, T., Osumi, T. (1996) Insertion of the 70 kDa peroxisomal membrane protein into peroxisomal membrane *in vivo* and *in vitro*, *J. Biol. Chem.* **271**, 3706–3713.

Jia, L., Dienhart, M., Schramp, M., McCauley, M., Hell, K., Stuart, R.A. (2003) Yeast Oxa1 interacts with mitochondrial ribosomes: the importance of the C-terminal region of Oxa1, *EMBO J.* **22**, 6438–6447.

Jones, J.M., Morrell, J.C., Gould, S.J. (2001) Multiple distinct targeting signals in integral peroxisomal membrane proteins, *J. Cell Biol.* **153**, 1141–1150.

Keegstra, K., Froehlich, J.E. (1999) Protein import into chloroplasts, *Curr. Opin. Plant Biol.* **2**, 471–476.

Kerscher, O., Holder, J., Srinivasan, M., Leung, R.S., Jensen, R.E. (1997) The Tim54p-Tim22p complex mediates insertion of proteins into the mitochondrial inner membrane, *J. Cell Biol.* **139**, 1663–1675.

Kessler, F., Schnell, D. (2004) Chloroplast protein import: solve the GTPase riddle for entry, *Trends Cell Biol.* **14**, 334–338.

Koehler, C.M., Merchant, S., Schatz, G. (1999) How membrane proteins travel across the mitochondrial intermembrane space, *Trends Biochem. Sci.* **24**, 428–432.

Kozjak, V., Wiedemann, N., Milenkovic, D., Lohaus, C., Meyer, H.E., Guiard, B., Meisinger, C., Pfanner, N. (2003) An essential role of Sam50 in the protein sorting and assembly machinery of the mitochondrial outer membrane, *J. Biol. Chem.* **278**, 48520–48523.

Kuhn, A., Stuart, R., Henry, R., Dalbey, R.E. (2003) The Alb3/Oxa1/YidC protein family: membrane-localized chaperones facilitating membrane protein insertion? *Trends Cell Biol.* **13**, 510–516.

Kunau, W.H. (2001) Peroxisomes: the extended shuttle to the peroxisome matrix, *Curr. Biol.* **11**, R659–R662.

Lazarow, P.B., Fujiki, Y. (1985) Biogenesis of peroxisomes, *Annu. Rev. Cell Biol.* **1**, 489–530.

Legakis, J.E., Terlecky, S.R. (2001) PTS2 protein import into mammalian peroxisomes, *Traffic* **2**, 252–260.

Liao, S., Lin, J., Do, H., Johnson, A.E. (1997) Both lumenal and cytosolic gating of the aqueous ER translocon pore are regulated from inside the ribosome during membrane protein integration, *Cell* **90**, 31–41.

Lill, R., Dowhan, W., Wickner, W. (1990) The ATPase activity of SecA is regulated by acidic phospholipids, SecY, and the leader and mature domains of precursor proteins, *Cell* **60**, 271–280.

Martsuzono, Y., Kinoshita, N., Tamura, S., Shimozawa, N., Hamasaki, M., Ghaedi, K., Wanders, R.J., Suzuki, Y., Kondo, N., Fujiki, Y. (1999) Human PEX19: cDNA cloning by functional complementation, mutation analysis in a patient with Zellweger syndrome, and potential role in peroxisomal membrane assembly, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2116–2121.

Marzioch, M., Erdmann, R., Veenhuis, M., Kunau, W.H. (1994) PAS7 encodes a novel yeast member of the WD-40 protein family essential for import of 3-oxoacyl-CoA thiolase, a PTS2-containing protein, into peroxisomes, *EMBO J.* **13**, 4908–4918.

Mihara, K., Omura, T. (1996) Cytosolic factors in mitochondrial protein import, *Experientia* **52**, 1063–1068.

Milenkovic, D., Kozjak, V., Wiedemann, N., Lohaus, C., Meyer, H.E., Guiard, B., Pfanner, N., Meisinger, C. (2004) Sam35 of the mitochondrial protein sorting and assembly machinery is a peripheral outer membrane protein essential for cell viability, *J. Biol. Chem.* **279**, 22781–22785.

Mori, H., Cline, K. (2002) A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid delta pH/Tat translocase, *J. Cell Biol.* **157**, 205–210.

Murphy, C.K., Beckwith, J. (1994) Residues essential for the function of SecE, a membrane component of the *Escherichia coli* secretion apparatus, are located in a conserved cytoplasmic region, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2557–2561.

Neupert, W., Brunner, M. (2002) The protein import motor of mitochondria, *Nat. Rev. Mol. Cell Biol.* **3**, 555–565.

Nishiyama, K., Mizushima, S., Tokuda, H. (1992) The carboxyl-terminal region of SecE interacts with SecY and is functional in the reconstitution of protein translocation activity in *Escherichia coli*, *J. Biol. Chem.* **267**, 7170–7176.

Pfanner, N., Neupert, W. (1987) Distinct steps in the import of ADP/ATP carrier into mitochondria, *J Biol. Chem.* **262**, 7528–7536.

Pfanner, N., Wiedemann, N., Meisinger, C., Lithgow, T. (2004) Assembling the mitochondrial outer membrane, *Nat. Struct. Mol. Biol.* **11**, 1044–1048.

Pohlschroder, M., Prinz, W.A., Hartmann, E., Beckwith, J. (1997) Protein translocation in three domains of life: variations on a theme, *Cell* **81**, 563–566.

Porcelli, I., de Leeuw, E., Wallis, R., van den Brink-van der Laan, E., de Kruijff, B., Wallace, B.A., Palmer, T., Berks, B.C. (2002) Characterization and membrane assembly of the TatA component of the *Escherichia coli* twin-arginine protein transport system, *Biochemistry* **41**, 13690–13697.

Purdue, P.E., Lazarow, P.B. (2001) Peroxisome biogenesis, *Annu. Rev. Cell Dev. Biol.* **17**, 701–752.

Rachubinski, R.A., Subramani, S. (1995) How protein penetrate peroxisomes, *Cell* **83**, 525–528.

Rehling, P., Model, K., Brandner, K., Kovermann, P., Sickmann, A., Meyer, H.E., Kuhlbrandt, W., Wagner, R., Truscott, K.N., Pfanner, N. (2003) Protein insertion into the mitochondrial inner membrane by a twin-pore translocase, *Science* **299**, 1747–1751.

Robinson, C., Bolhuis, A. (2001) Protein targeting by the twin-arginine translocation pathway, *Nat. Rev. Mol. Cell Biol.* **2**, 350–356.

Rodrigue, A., Chanal, A., Beck, K., Muller, M., Wu, L.F. (1999) Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial tat pathway, *J. Biol. Chem.* **274**, 13223–13228.

Sacksteder, K.A., Gould, S.J. (2000) The genetics of peroxisome biogenesis, *Annu. Rev. Genet.* **34**, 623–652.

Samuelson, J.C., Chen, M., Jiang, F., Moller, I., Wiedmann, M., Kuhn, A., Phillips, G.J., Dalbey, R.E. (2000a) YidC mediates membrane protein insertion in bacteria, *Nature* **406**, 637–641.

Santini, C.L., Bernadac, A., Zhang, M., Chanal, A., Ize, B., Blanco, C., Wu, L.F. (2001) Translocation of jellyfish green fluorescent protein via the Tat system of *Escherichia coli* and change of its periplasmic localization in response to osmotic up-shock, *J. Biol. Chem.* **276**, 8159–8164.

Sargent, F., Bogsch, E.G., Stanley, N.R., Wexler, M., Robinson, C., Berks, B., Palmer, T. (1998) Overlapping functions of components of a bacterial Sec-independent protein export pathway, *EMBO J.* **17**, 3640–3650.

Schatz, P.J., Riggs, P.D., Jacq, A., Fath, M.J., Beckwith, J. (1989) The secE gene encodes an integral membrane protein required for protein export in *Escherichia coli*, *Genes Dev.* **3**, 1035–1044.

Schwartz, M.P., Matouschek, A. (1999) The dimensions of the protein import channel in the outer and inner mitochondrial membranes, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13086–13090.

Settles, M.A., Yonetani, A., Baron, A., Bush, D.R., Cline, K., Martienssen, R. (1997) Sec-independent protein translocation by the maize Hcf106 protein, *Science* **278**, 1467–1470.

Shimozawa, N., Suzuki, Y., Zhang, Z., Imamura, A., Chaedi, K., Fujiki, Y., Kondo, N. (2000) Identification of PEX3 as the gene mutated in a Zellweger syndrome patient lacking peroxisomal remnant structures, *Hum. Mol. Genet.* **9**, 1995–1999.

Smith, M.D., Schnell, D. (2001) Peroxisomal protein import: the paradigm shifts, *Cell* **105**, 293–296.

Snyder, W.B., Koller, A., Choy, A.J., Subramani, S. (2000) The peroxin Pex19p interacts with multiple, integral membrane proteins at the peroxisomal membrane, *J. Cell Biol.* **149**, 1171–1178.

Sohrt, K., Soll, J. (2000) Toc64, a new component of the protein translocon of chloroplasts, *J. Cell Biol.* **148**, 1213–1221.

Spence, E., Sarcina, M., Ray, N., Moller, S.G., Mullineaux, C.W., Robinson, C. (2003) Membrane-specific targeting of green fluorescent protein by the Tat pathway in the cyanobacterium synechocystis PCC6803, *Mol. Microbiol.* **48**, 1481–1489.

Sylvestre, J., Margeot, A., Jacq, C., Dujardin, G., Corral-Debrinski, M. (2003) The role of the 3′ untranslated region in mRNA sorting to the vicinity of mitochondria is conserved from yeast to human cells, *Mol. Biol. Cell.* **14**, 3848–3856.

Taylor, R.D., Pfanner, N. (2004) The protein import and assembly machinery of the mitochondrial outer membrane, *Biochim. Biophys. Acta* **1658**, 37–43.

Truscott, K.N., Voos, W., Frazier, A.E., Lind, M., Li, Y., Geissler, A., Dudek, J., Muller, H., Sickmann, A., Meyer, H.E., Meisinger, C., Guiard, B., Rehling, P., Pfanner, N. (2003) A J-protein is an essential subunit of the

presequence translocase-associated protein import motor of mitochondria, *J. Cell Biol.* **163**, 707–713.

Van den Berg, B., Clemons, W.M. Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., Rapoport, T.A. (2004) X-ray structure of a protein-conducting channel, *Nature* **427**, 36–44.

van der Laan, M., Urbanus, M.L., Ten Hagen-Jongman, C.M., Nouwen, N., Oudega, B., Harms, N., Driessen, A.J., Luirink, J. (2003) A conserved function of YidC in the biogenesis of respiratory chain complexes, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5801–5806.

Waegemann, K., Soll, J. (1996) Phosphorylation of the transit sequence of chloroplast precursor proteins, *J. Biol. Chem.* **271**, 6545–6554.

Waegemann, K., Paulsen, H., Soll, J. (1990) Translocation of proteins into isolated chloroplasts requires cytosolic factors to obtain import competence, *FEBS Lett.* **261**, 89–92.

Walker, M.B., Roy, L.M., Coleman, E., Voelker, R., Barkan, A. (1999) The maize *tha4* gene functions in Sec-independent protein transport chloroplasts and is related to *hcf106, tatA,* and *tatB, J. Cell Biol.* **147**, 267–276.

Young, J.C., Hoogenraad, N.J., Hartl, F.U. (2003) Molecular chaperones Hsp90 and Hsp70 deliver preproteins to the mitochondrial import receptor Tom70, *Cell* **112**, 41–50.

Zheng, N., Gierasch, L.M. (1996) Signal sequences: the same yet different, *Cell* **86**, 849–852.

Zito, C.R., Oliver, D. (2003) Two-stage binding of SecA to the bacterial translocon regulates ribosome-translocon interaction, *J. Biol. Chem.* **278**, 40640–40646.

Proteomics

Paul Cutler[1], *Israel S. Gloger*[2], *and Christine Debouck*[3]
GlaxoSmithKline Pharmaceutical Research & Development,
[1] *Stevenage, Hertfordshire, UK*
[2] *Harlow, Essex, UK*
[3] *Collegeville, PA*

1 Definitions 313
1.1 Proteomics 313
1.2 Expression Proteomics 313
1.3 Interaction Proteomics 313
1.4 Clinical Proteomics and Toxicoproteomics 313

2 Technologies 313
2.1 2D Gel Electrophoresis (2-DE) 313
2.2 Mass Spectrometry 315
2.3 Isotope Coding 318
2.4 Yeast Two-hybrid Systems 319
2.5 Protein Arrays 321
2.6 Bioinformatics 322

3 Applications 323
3.1 Target Identification and Validation 323
3.2 Protein–protein Interactions 325
3.3 Protein Biomarkers in Clinical and Preclinical Research 325

4 Future Perspectives 327
4.1 Toward a Full Proteome Map 327
4.2 Biological Validation of Protein Functions 328
4.3 Clinical and Diagnostic Applications 328

Acknowledgments 329

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

Bibliography 329
Books and Reviews 329
Primary Literature 329

Keywords

Electrospray Ionization (ESI)
Electrospray is the process of ionization by which fractionated peptides are sprayed in solution through a charged needle toward the mass analyzer.

Isotope-coded Affinity Tag (ICAT)
Isotope-coded affinity tags are reagents used for quantifying the differential expression of peptides by mass spectrometry. ICAT tags contain three key components: a reactive group for labeling peptides (e.g. iodo-group for labeling cysteines), a group for affinity separation (e.g. biotin), and a spacer. The spacer can be derivatized with different isotopes (e.g. hydrogen or deuterium) to discriminate protein expression of peptides from different samples. The ratios of the heavy and light isotopes measured by the mass spectrometer may then be correlated with the relative abundance of the peptide in the two samples.

Liquid Chromatography–Mass Spectrometry (LC-MS)
Liquid chromatography–mass spectrometry is the separation of complex mixtures of peptides or protein samples by liquid chromatography followed by mass spectrometric analysis. Reverse phase liquid chromatography is usually performed on peptides as this utilizes volatile solvents compatible with mass spectrometry.

Mass Spectrometry
Mass spectrometry is a biophysical methodology used to measure the mass-to-charge ratio of proteins or peptides. The method involves the generation of ions in a vacuum from a solid peptide or a peptide in solution, followed by their detection and mass analysis. Mass spectrometry instruments therefore contain a source of ionization to convert the analyte into ions within the gas phase and a mass analyzer for the detection of the ions. Various methods exist for the generation of ions, including electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI). Methods for detecting and analyzing the ions include time of flight (TOF) analyzers, quadrupole electric fields and ion traps.

Matrix-assisted Laser Desorption Ionization (MALDI)
In matrix-assisted laser desorption ionization, ions are formed via the laser excitation of samples embedded in a chromophore matrix. Laser energy excites the chromophore and the energy is transferred to the analyte, which ionizes and enters the gas phase. The mass-to-charge ratios of these ions are usually measured by the ‘‘time of flight’’

(TOF) from their release until detection. The time of flight can thus be converted into accurate molecular weight values.

Posttranslational Modifications
This is the process by which proteins are structurally modified following translation by the ribosomes, which may lead to altered function or activity. These include deamidation, proteolytic cleavage such as C-terminal truncation, phosphorylation, and glycosylation.

Protein Arrays
These consist in the microarray-based analysis of proteins either in a planar (chip) or suspension format, to qualitatively or quantitatively characterize protein expression or function.

Proteome
The proteome is the complete collection of proteins expressed in a given cell or tissue under a given set of conditions (normal, disease, treated or untreated with stimulants or compounds).

RNA Interference (RNAi)
RNAi (RNA interference) refers to the introduction of homologous double-stranded RNA (dsRNA) to specifically target and thereby silence the expression of one or more genes in a variety of organisms and cell types (e.g. worms, fruit flies, mammalian cells), resulting in null or hypomorphic phenotypes.

Surface-enhanced Laser Desorption Ionization (SELDI™)
In surface-enhanced laser desorption ionization, the analyte(s) are applied in solution to a protein-binding surface or ''chip'' followed by treatment with a MALDI matrix. The essential element of SELDI is the chemically modified nature of the surface, which facilitates selective adsorption prior to analysis. Laser desorption will release the bound proteins as ions (analogous to MALDI). The ion spectrum can be analyzed and specific patterns compared across different sample groups.

Tandem MS (MS/MS)
Ions detected by mass spectrometry can be selected for further fragmentation into ion products. The fragmented ion spectra contain more detailed information, which can be used to derive *de novo* sequence information about the peptide.

Transcriptomics
Transcriptomics is the study of the transcriptome, the collection of RNAs expressed in a given cell or tissue under given conditions (normal, disease, treated or untreated with stimulants or compounds). The commonly used method to study the transcriptome is microarray, also called *gene chip*, technology that allows the analysis of gene expression to be determined for all genes in a genome.

Two-dimensional Gel Electrophoresis (2-DE)
Two-dimensional gel electrophoresis (2-DE) is the separation of proteins in a polyacrylamide gel matrix, based on two properties. Separation in the first dimension is determined by the isoelectric point of the proteins being analyzed, while the second dimension separates them according to their apparent molecular weight.

Yeast Two-hybrid (Y2H)
This method was developed for the identification of protein–protein interactions using yeast as a host organism. A gene sequence encoding either a protein or part of a protein of interest is fused to the DNA-binding domain of a known, site-specific transcription factor ("bait"), while a single target gene sequence or library of such gene sequences (which may be prepared from a range of tissues etc.) for a potentially interacting protein(s) is fused to the transactivation domain of the same transcription factor ("prey"). Association between these two fusion proteins in yeast facilitates the reconstitution of the DNA binding and transactivation domains as a functional transcription factor, resulting in activation of a selected reporter gene placed under control of a promoter regulated by the transcription factor. Expression of the reporter can be monitored by a variety of methods.

The past century has witnessed many impressive advances in molecular biology and medical research, leading to both increased understanding and improvements in the treatment of many diseases. A notable example of this is the impact that the discovery of antibiotics had in treating bacterial infections. One of the most remarkable achievements was the discovery of the structure of DNA by Crick & Watson, an event that subsequently led some 50 years later to the completion of the sequencing of the entire human genome. This milestone has created a wealth of information and novel avenues for a more focused and comprehensive search and understanding of the biochemical, regulatory and structural function of the proteins encoded by the genome – a phase of biological exploration now referred to as the *postgenome era*.

In the postgenome era, there is significant work ongoing not only on the genetic aspects of biology but also the molecular processes guiding gene expression and function at all levels. These include RNA and protein synthesis and their regulation, protein modifications, protein–protein interactions (pathways), and metabolite formation. This review will focus on Proteomics, the study of global gene expression at the protein level. We will describe both currently used and emerging technologies for exploring universal protein expression, including protein fractionation, analysis, and quantification of complex protein solutions. In particular, this review will focus on characterizing the roles of proteins in pharmaceutical research and development, including exploring applications in determining the functional role of proteins and their relevance in human disease, as surrogate biomarkers for predicting/diagnosing disease, disease progression, and for determining compound efficacy and toxicity. Finally, we will assess what are the future prospects of utilizing protein research to reach a better knowledge of gene/protein function in normal and disease biology as well as clinical applications.

1 Definitions

1.1 Proteomics

The rapidly evolving nature of protein analysis has resulted in an ever-expanding definition of the term "proteomics." While it is accepted that proteomics concerns the study and characterization of all proteins, including their synthesis, modifications, function and interactions, the evolution of separation and identification techniques has generated more specific definitions of proteomic activities such as expression proteomics and interaction proteomics. In addition, specific applications have emerged including clinical proteomics and toxicoproteomics.

1.2 Expression Proteomics

Expression proteomics is defined as the global separation, quantification and identification of expressed proteins in organisms, tissues, or cells under normal or other conditions such as disease or following compound treatment. Initially, this area was centered on 2-DE, but the recent development of nongel-based methodologies such as LC-MS and protein arrays have also started to show promise in this area.

1.3 Interaction Proteomics

Interaction proteomics is the study of specific protein–protein or protein–compound interactions and, more generally, the study of the role of protein complexes within biological pathways. Different methodologies such as the yeast two-hybrid system, phage display, protein arrays, and affinity purification of complexes followed by mass spectrometric analysis have been successfully applied to interaction proteomics studies.

1.4 Clinical Proteomics and Toxicoproteomics

Two particularly exciting emerging fields of expression proteomic research relating to clinical applications are clinical proteomics and toxicoproteomics. Clinical proteomics consists in the identification of biomarkers for disease detection or progression, or even regression after medical therapeutic treatment, whereas toxicoproteomics is the elucidation of potentially toxic or pathological effects of chemical agents or environmental factors.

2 Technologies

2.1 2D Gel Electrophoresis (2-DE)

Since the inception of proteomics as a discipline, 2-DE has been the technique most widely used for protein separation and analysis. This is still the case today, with relatively little changes from its original process developed about 30 years ago by O'Farrell, although a number of complementary approaches are being developed. In 2-DE, the separated proteins are visualized by either pre- or postelectrophoretic stains, class-specific stains (e.g. phosphorylation), or by autoradiography for radiolabeled proteins. Typically, between 2000 and 3000 proteins can be resolved in a standard 2-DE gel (Fig. 1). The technique is robust, quantitative, and relatively simple to implement with reasonably high throughput.

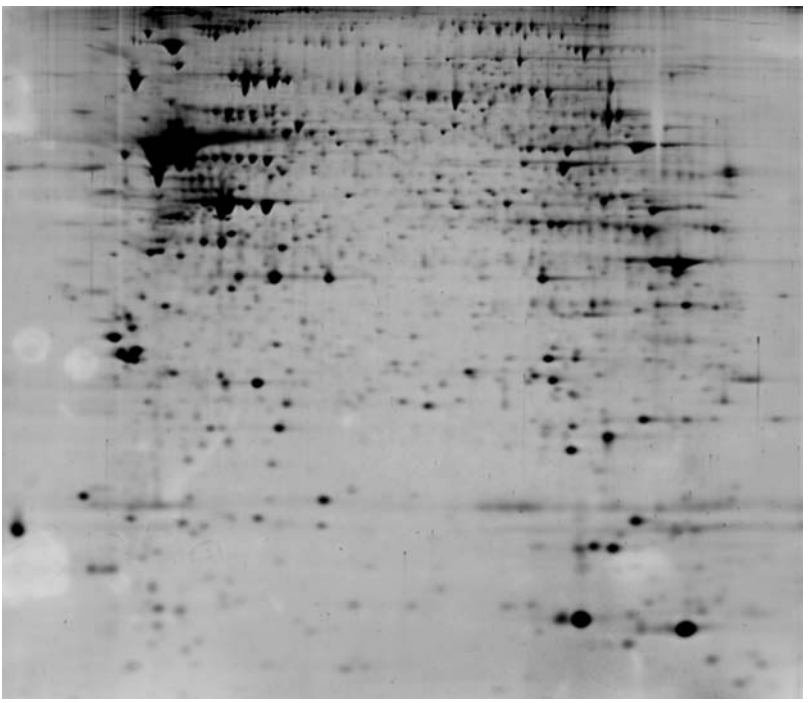

Fig. 1 Two-dimensional polyacrylamide gel separation of proteins from a complex mixture, illustrating separation by isoelectric point in the first (horizontal) dimension and separation by size in the second (vertical) dimension.

Major improvements have enabled greatly increased coverage, reproducibility, and accuracy. These include advances in sample preparation, electrophoresis methodology, as well as sensitivity and quantification of staining. Furthermore, scanning hardware has dramatically improved as have the available algorithms for image analysis and data mining.

Although unparalleled in its resolving power and flexibility, the performance of 2-DE is ultimately finite. Proteins are commonly detected in biological tissues over a dynamic range of $>10^6$. In plasma, the dynamic range has been reported to be on the order of 10^{12}. Because of this huge expression range, high abundance proteins may ultimately mask the visibility and detection of low abundance ones.

There are no "amplification" techniques for proteins analogous to the polymerase chain reaction (PCR) used for transcript identification and quantification. As proteomics is a substrate-limited technique, it is crucial that sufficient sample is available as starting material for analysis. Prefractionation is increasingly becoming a prerequisite for broad proteome coverage. In addition, the technology is biased against proteins with extremes of size, charge, and hydrophobicity. In order to address these "problematic" proteins, a number of novel approaches are being developed. These include multicompartment electrophoresis for enrichment of proteins on the basis of native charge and subcellular fractionation, in particular, isolation of organelle components such as mitochondria and

nuclei. As a consequence, initiatives to define the subcellular complement of the nuclear and mitochondrial proteomes are being actively pursued. This type of more focused proteomic analysis has increased sensitivity and has enabled analysis of the nuclear envelope and even the nuclear pore complex. Techniques have also been developed to preenrich proteins by functional class (e.g. proteases) and to specifically remove high abundance proteins. Such approaches include removal of albumin and other highly abundant proteins from plasma and serum prior to clinical proteomic studies.

The original carrier ampholyte-generated pH gradients of the 1970s have now been replaced with immobilized pH gradients (IPG) for isoelectric focusing (IEF). These matrices provide stable and reproducible gradient separations and are available commercially in broad pH range gels (e.g. pH 3–11) and extremely narrow pH range gels (e.g. pH 5–6), enabling extremely high resolution to be attained.

The performance of the second dimension has also been enhanced with the introduction of better instrumentation. In particular, reproducibility and throughput has been improved, while the introduction of the stable gradient gels has increased the resolution and flexibility of the technique. In conjunction with other advances, 2-DE is now capable of producing quantitative and reproducible data at a throughput compatible with biologically significant sample numbers.

The introduction of highly sensitive fluorescent dyes is greatly improving protein detection in gels. Fluorescent stains such as SYPRO Ruby have a better dynamic range, are more quantitative and are more compatible with mass spectrometry than previous stains (e.g. silver stains). A direct consequence of improvements to the performance of the technology is the separation of isoforms of proteins differing by single deamidation or phosphorylation modifications. Subsequently, stains specific for phosphorylation and glycosylation were developed and used successfully for proteins separated by 2-DE. An additional advance is the use of differential gel electrophoresis, incorporating prelabeling protein samples with fluorescent dyes and the ability to run two or even three samples on the same gel for comparative analysis.

Overall, 2-DE remains one of the most reliable and widely used techniques for proteome analysis. This is particularly true for sample profiling and for the analysis of differential protein expression in differing disease states or after treatment of cells or tissues with biologically or medically relevant agents or compounds. The continuous development of new techniques and instrumentation means that the limitations of the technique are constantly being reappraised and 2-DE will remain a very valuable approach for the foreseeable future.

2.2 Mass Spectrometry

One of the fundamental challenges of proteomics work is the requirement for sequence identification of proteins from a given sample following separation. Until a few years ago, the classical protein identification procedure was performed through N-terminal amino acid sequencing, known as *Edman degradation*. The development of mass spectrometry has essentially replaced Edman's method by virtue of providing more sensitive detection and higher sample throughput. These features, coupled with the rapid expansion of genomic databases, have provided

a revolutionary approach to protein detection and identification. The key steps describing the use of mass spectrometry to identify proteins from polyacrylamide gels are shown in Fig. 2.

In general, mass spectrometry is performed on tryptic peptides rather than intact proteins as their extraction from the microreticular polyacrylamide matrix of 2-DE is more facile. In addition, peptides are more readily amenable to ionization and analysis. This is not trivial as it necessitates a method for efficiently digesting proteins within the gel and recovering the resulting peptides. The proteases routinely employed, such as trypsin, are presented to the protein substrates, which are present within the gel at concentrations significantly below the Michaelis constant of the enzyme. Nevertheless, methods have been optimized, which enable efficient fragmentation of the proteins *in situ*.

The main principle of mass spectrometry is the accurate measurement of the masses of the peptide ions and derived peptide fragment ions from the analyses. The first requirement for analyte characterization is the generation of ions in the gas phase. In particular, the so-called "soft ionization" methods of matrix-assisted laser desorption ionization (MALDI) and

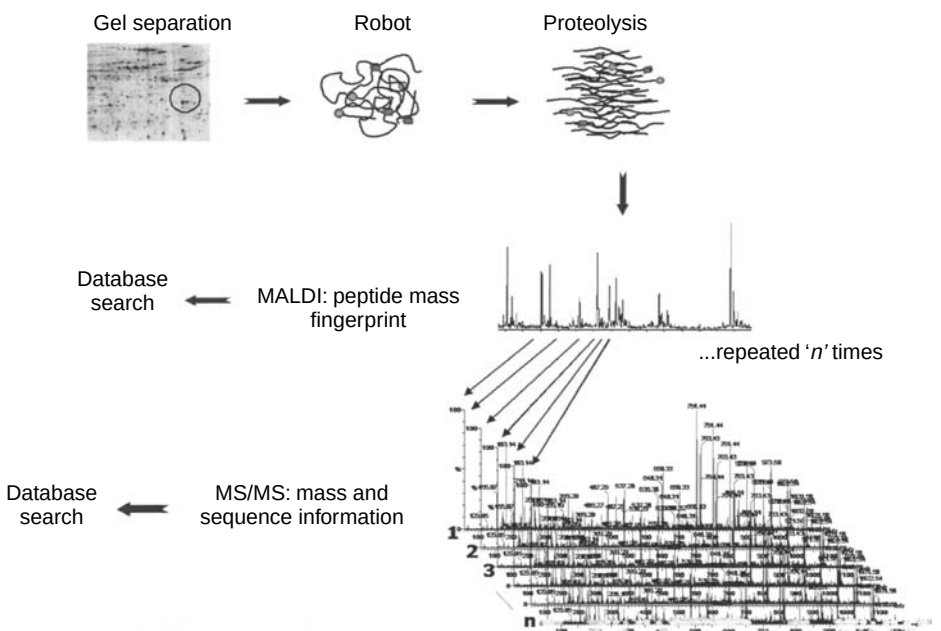

Fig. 2 Qualitative identification of proteins from gels by mass spectrometry. A spot of interest is excised from the gel and digested to peptides by trypsin. Two main methods exist for identifying a protein from the peptides generated. Firstly, the mass of the peptides may be determined by mass fingerprinting and compared to genomic data present in bioinformatics databases. Alternatively, individual peptide ions can be sequentially selected for fragmentation, which generates *de novo* sequence information. A combination of the sequence information and the peptide mass can be used to accurately assign a protein identity by comparison to the genomic database.

electrospray ionization (ESI) have made a dramatic impact on proteomics.

In MALDI, a laser pulse is applied to the sample that is embedded in a matrix (chromophore), which absorbs light at a specific wavelength. The energy in the chromophore is passed to the analyte peptide, which ionizes. The ions created are guided by a strong electrical field into the mass analyzer. MALDI is a particularly "soft" ionization technique resulting in singly charged peptide ions where the mass-to-charge (m/z) ratio reflects the mass of the peptide. This technique is most often used to measure the masses of peptides derived from enzymatic digestion of a protein. This is often referred to as the "*fingerprint*" of the protein and may be sufficient to identify the protein against a genomic database.

In ESI, ions are formed by electrospraying a peptide solution into the mass spectrometer via a small needle in the presence of a strong electrical field, creating small ionized droplets. The solvent in the droplets is eliminated by applying a counter current flow of gas or heat. Once the ions have entered the mass spectrometer, the mass analyzer separates and measures the ions on the basis of their mass-to-charge ratio (m/z) using electric or magnetic fields with subsequent registration by the detector. In ESI, a series of ions of different charges ($z = 1$, $z = 2$, $z = 3$, etc.) may be formed and these have to be deconvoluted to determine the mass of the parent peptide. This technique provides higher energy ions suitable for fragmentation and *de novo* sequencing. Relative abundance of the peptide will be reflected by the ion current in the analyzer. The ion current intensity will also be a reflection of the propensity of a particular peptide to ionize, which is in turn predicated on the amino acid composition. These confounding influences make absolute quantification without the use of internal standards complex.

In its simplest mode, the mass spectrometer defines the mass of the parent peptide ion. Peptides can then be selected manually or automatically within the mass spectrometer and further fragmented by collision with an inert gas. Peptides undergoing this process (tandem MS or MS/MS) tend to fragment at the amide bond, leading to a mass spectral ladder, whose steps reflect the mass differences between the parent ions sequentially depleted of the terminal amino acid residue (Fig. 2). It is therefore possible to derive *de novo* sequence information and, by reference to genomic databases, unequivocally identify a protein from a single peptide.

In recent years, different variants of mass spectrometry have emerged, encompassing novel approaches for trapping and selecting ions (ion traps, quadrupoles) as well as new mass analyzers such as time of flight (TOF) and Fourier transform ion cyclotron resonance (FT or FTICR-MS). As with the more classical approaches, performance is dependent upon the key parameters of mass accuracy, resolution, and sensitivity.

Biological samples are often extremely complex, and gel-derived digests often contain reagents incompatible with mass spectrometers. To circumvent these issues, separation technologies such as chromatography have been coupled to mass spectrometry. As with 2-DE, methods of fractionation have been developed to increase the proteome coverage or to analyze specific subsets of the proteome. These include methods of fractionation and specific tagging procedures for labeling phosphoproteomes.

Owing to its use of volatile solvents, reverse phase high performance liquid

chromatography (RP-HPLC) is particularly compatible with mass spectrometry and is readily interfaced via a standard ESI source (LC-MS). The mass spectrometer is able to sample and analyze the peptides as they elute sequentially from the chromatographic column. While this approach has disadvantages of being slower than MALDI-based approaches, it has the advantage of generating higher quality data.

For extremely high resolution of very complex mixtures, use of two-dimensional chromatography techniques coupled to tandem mass spectrometry (LC/LC-MS/MS), sometimes referred to as *multidimensional protein identification technology* (MudPIT), has been established. The first dimension is often ion exchange chromatography followed by RP-HPLC for the second dimension. Progress in the development of LC/MS-based proteome analysis has raised the realistic prospect of nongel-based global protein expression profiling. Intact proteins are not readily amenable to RP-HPLC in complex mixtures and so it is necessary to predigest the proteins to peptides for analysis. This creates an added dimension of complexity and often precludes adequate analysis of posttranslationally modified proteins. While this technology is still evolving, it offers potential advantages in overcoming the limitations discussed above that are associated with 2-DE approaches. Recent advances have shown that the peptide retention time by LC, the m/z value and its intensity may provide sufficient information to both quantitatively and qualitatively identify a protein in a complex mixture.

2.3 Isotope Coding

Mass spectrometry has utilized stable isotopes to quantify relative amounts of individual analytes for many years. Attempts to perform experiments on a global proteomic scale have led to a number of exciting developments. These include the use of stable isotope labeling of amino acids in culture (SILAC) and other metabolic labeling techniques. However, for *in vivo* and larger studies, labeling at the point of analysis is needed. Digestion of proteins with trypsin produces neotermini on the peptides, which provide the opportunity for differential labeling with the isotopes ^{18}O and ^{16}O derived from heavy and light water respectively. The use of isotope-coded affinity tags (ICAT) has gained some acceptance in the last few years, with one notable early study showing differential protein expression in yeast grown under different metabolic conditions.

ICAT probes are a series of labeling reagents containing an affinity tag (biotin), a spacer labeled with a light or heavy isotope, and a thiol-specific reactive group (-iodo) for labeling of cysteine-containing peptides. Labeling is optimized to permit quantitative analysis of samples (Fig. 3). Early ICAT reagents possessed spacers containing either eight deuteriums (D8, the heavy isotope) or eight hydrogens (D0, the light isotope). Two samples are labeled, one with the light reagent and the other with the heavy reagent and the samples are then combined. Once digested, separated, and analyzed by mass spectrometry, the relative abundance of the isotope tags may be used to estimate the relative levels of the individual protein species.

In the earlier incarnation of ICAT, the necessity for high-resolution LC systems to resolve the plethora of peptides resulted in the undesired resolution of the light D0 and heavy D8 form. This is problematic as the mass spectrometric analysis

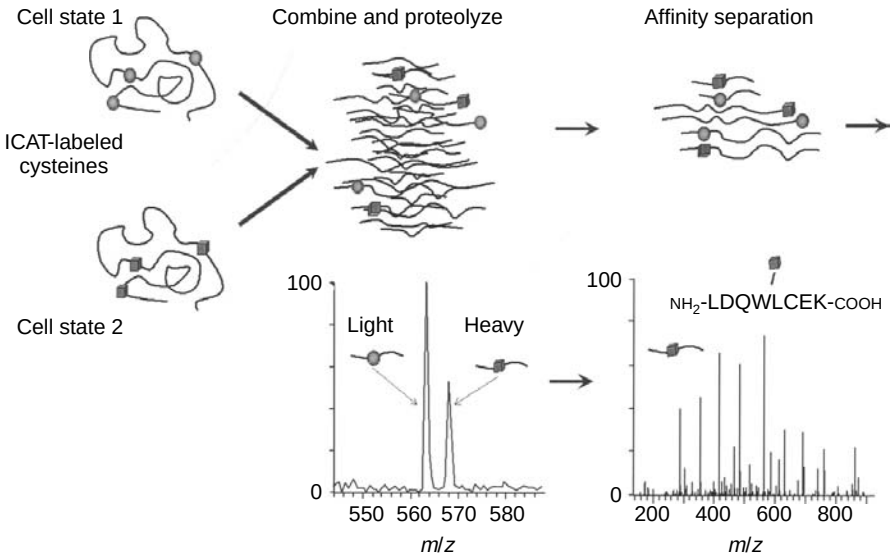

Fig. 3 Principle of isotope-coded affinity tagging. The proteins of two samples are labeled with a tag containing one of two isotopic labels such as eight deuteriums (squares) or eight protons (circles). The samples are combined, affinity separated, and analyzed by mass spectrometry. The relative abundance of the ions with the appropriate mass difference (e.g. 8Da) reflects the relative abundance of the proteins.

is predicated on the two ions entering the mass spectrometer at the same time. The affinity tag specificity for the cysteine residue also limited its applicability only to cysteine-containing proteins. In addition, the reagent itself did not perform ideally in the mass spectrometer, generating heterologous fragmentation, which confounded interpretation. Many new isotopes are being developed today, and it is hoped that future iterations will address these caveats.

2.4 Yeast Two-hybrid Systems

The yeast two-hybrid system identifies putative binary protein interactions, which, if proven, can provide powerful insights into biological pathways. The gene coding region for a protein of interest or part of a protein of interest (the "bait") is fused to the coding region for the DNA-binding domain of a yeast transcription factor (commonly GAL4 or LexA). This is engineered to ensure that the resulting fusion protein is targeted to the nucleus with a selection marker. The coding region for a single protein (the "prey") or a library of proteins from different tissues ("prey" library) is fused to the transcriptional activation domain of the same transcription factor, and also targeted to the nucleus. Screening is performed with reporter genes under the control of a promoter that respond to the reconstituted DNA binding/transcription factor activity, reflecting bait–prey interaction (Fig. 4). Common reporter genes include enzymes, which support cellular growth *(LEU2, HIS3)* and β-galactosidase, for which a simple colorimetric detection is possible.

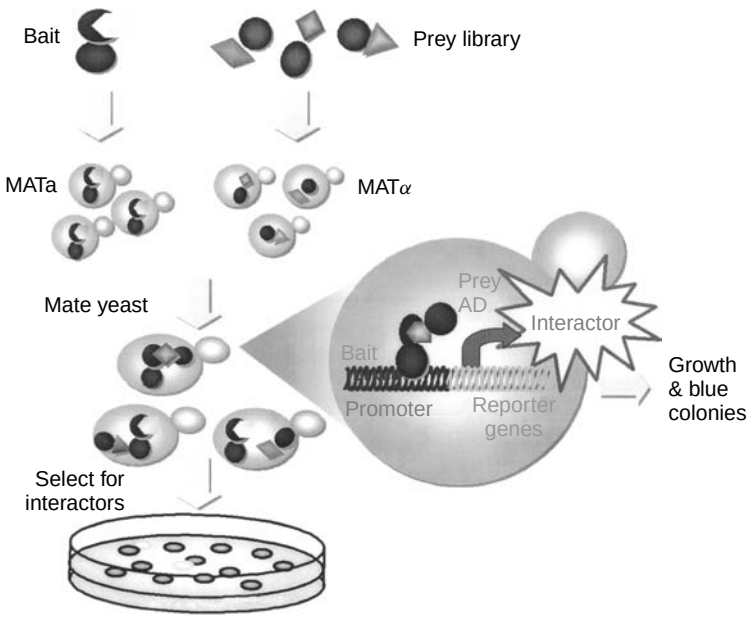

Fig. 4 Principle of the yeast two hybrid system for defining protein–protein interactions. Two chimeras, one containing a DNA-binding domain and the other an activation domain are cotransfected into yeast. If the fusion partners interact, the binding domain and activation domain are brought into proximity and activate transcription of a reporter gene.

In addition to screening prey libraries for interactors, the system can also be used for screening deletion mutants or polymorphisms to characterize the interacting domains. Variants of the yeast two-hybrid technique include yeast one-hybrid systems for protein-DNA interactions and yeast three-hybrid systems for looking at three way protein interactions or protein–drug interactions.

Although there is an association between the strength of the interaction and the level of reporter activity in the yeast two-hybrid system, the interaction is only weakly quantitative and the true strength of the technique is in its qualitative analysis. Although false-positive results due to nonspecific interactions or spontaneous transactivation activity are often likely to confound interpretation, careful experimental design and interpretation as well as follow-up with directed confirmation assays using alternate methods (e.g. coimmunoprecipitation) can help with the identification and validation of true interactors. Often critical protein–protein interactions are specific to a disease state and so the yeast model may generate false-negative data. Other limitations of the technology include aberrant interactions owing to improper folding, difficulty with finding membrane protein interactors, lack of necessary posttranslational modifications, or inappropriate cellular compartmentalization. However, it has the tremendous advantage of being cheap and simple to run, is an *in vivo* system, and can be run on a high-throughput scale. The positive aspects are reflected in its wide utility and role in many important discoveries including

the GABA-B receptor heterodimerization and producing genome scale interaction maps for *Saccharomyces cerevisiae* (yeast), *Caenorhabditis elegans* (worm), and *Drosophila* (fruit fly).

2.5 Protein Arrays

Protein arrays constitute an emerging technology, which permits the analysis of protein expression, protein interactions, or protein function. The microarray-like format permits the deposition of various biomolecules onto a solid surface in an ordered array (Fig. 5). Antibody chips allow the detection of protein expression in an analogous method to enzyme-linked immunoassays (ELISAs). Arraying of functional biological molecules will also permit protein interactions as well as protein function to be investigated.

The inherent promise of protein arrays is the capacity to perform routine assays in a multiplexed format. This should greatly improve the comparative nature of a study by reducing assay-to-assay variability. Another benefit in practical terms is the reduction in reagent usage, which has clear relevance for the analysis of scarce clinical samples. Despite these advantages, the performance of protein expression arrays appears to be comparable in terms of sensitivity to ELISAs rather than enhanced. Certain techniques are being developed to try and increase sensitivity, including the use of novel detection systems (e.g. resonance light scattering) and signal amplification systems, but the results to date remain equivocal. Future developments using more advanced detection systems such as mass spectrometry may offer increased sensitivity, although for many applications such as expression analysis, this must be achieved while retaining

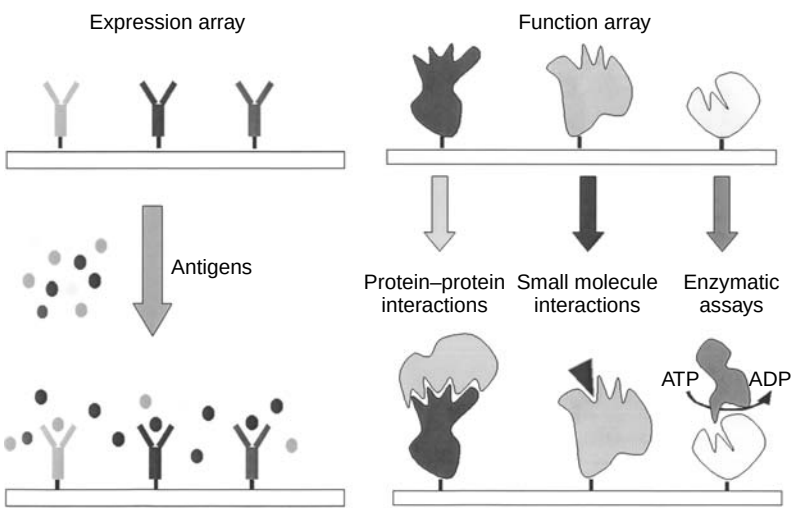

Fig. 5 Protein arrays contain a defined set of proteins arrayed at high density, typically onto a solid surface. Protein arrays can be used for protein profiling by immobilization of antibodies to specific analytes or as functional arrays for detecting protein–protein or protein–drug interactions. Proteins have also been arrayed to detect and characterize the functionality and specificity of both enzymes and substrates.

quantification. Also, as with other "omic-based" technologies, such advances will place demands upon the bioinformatics tools needed to manage and interpret the data.

Currently, the multiplexed protein array assay cannot, in general, compete with individual assays in terms of performance. Although the technology is advancing rapidly with respect to surface chemistry, detection, and data analysis, the limitations of the technology are predominantly in reagent supply. Commercial antibodies with the required sensitivity and selectivity are not available for the vast majority of proteins, and, even where they do exist, only 5% or less are suitable for use in protein arrays, having been produced for other purposes such as immunohistochemistry and western blotting. The production of antibodies is both labor intensive and time consuming. For example, production of monoclonal antibodies requires immunization, generation of hybridoma cell lines, and antibody isolation. The end result is often an antibody that can be thoroughly characterized in terms of sensitivity and selectivity, and a renewable source for producing this antibody over and over again. Although the use of phage libraries and other more novel capture reagents exist, such as aptamers and proteins based on novel scaffolds, the evidence is that maturation is needed to achieve ligands of acceptable performance and assay screening of potential reagents is often rate-limiting. The same challenge is true for functional proteins where the promise of whole genome or near whole genome cloning and expression is still to be realized. Future developments in capture reagents and protein production are likely to control the rate at which omic-scale protein arrays impact the biomedical sciences. Despite the current limitations, protein arrays for specific applications remain an exciting prospect and are beginning to yield useful and meaningful results.

2.6 Bioinformatics

A robust and expert bioinformatics platform is a key requirement for the capture, organization, analysis, and storage of proteomics data. The aim of both transcriptomics and proteomics studies is to generate high quality quantitative expression data. This needs to be managed in such a way as to enable information to be derived about complex patterns of expression and pathways consistent with pathology, physiology, drug efficacy, toxicology, and so on. By necessity, bioinformatics is a cross-disciplinary activity embracing computer science, software engineering, mathematics, and biology. The massive leap in scale accompanying "omic" technologies has necessitated ever more sophisticated bioinformatic tools to support target selection and validation, and disease understanding. The ability to relate a small piece of information on protein sequence to enable gene identification is critical and fundamental to proteomics, but increasingly important is the ability to place expression changes and interactions into a biological context of pathway and function. A concerted effort has begun to emerge to store, analyze and disseminate proteomic information, although it is still a discipline in its infancy where additional progress is eagerly awaited.

Many searchable databases available on the World Wide Web are being constantly updated to incorporate the increasing volume of data being generated from diverse genomic and proteomic platforms. Placing proteins in their correct functional

and biological context is likely to be a complex process with lessons to be learnt from transcriptomics and other disciplines.

3 Applications

3.1 Target Identification and Validation

Since proteins are the ultimate structural and functional components of life, understanding biological systems at the protein level is critical to delineating biological pathways, including those involved in disease as well as determining how drugs exert their beneficial and sometimes toxic effects. These explorations can lead to the identification of new targets (also proteins to be avoided) as well as contributing to validating the involvement of the particular protein involved in the biology of the disease. Proteomics has therefore begun to provide insights into several areas of medical research.

In the field of infectious diseases, proteomics – in conjunction with other disciplines such as transcriptomics and RNA interference (RNAi) – has been used to unravel the mechanism of certain drug activities and to define host–pathogen interactions. This holds the promise of providing new methods for diagnosing and treating infectious diseases. Initial studies focused on the effect of bacterial infection within host cells, including analysis of those proteins that are selectively expressed by the pathogen during infection. These studies also led to increasing our understanding of the inflammatory signaling networks in the process of sepsis. More recent work has focused on the mechanisms of pathogenesis. In addition, work with organisms such as *Plasmodium falciparum*, the malaria parasite, has provided critical insights into major diseases such as malaria, via delineation of the protein expression patterns at distinct stages in the life cycle of the parasite.

Protein changes correlating with disease processes, such as inflammation, have indicated dysregulation of proteins, which may subsequently be modulated by non-steroidal anti-inflammatory drugs. These included a time- and gender-dependent upregulation of acute phase proteins in a rat model. Respiratory diseases, such as chronic obstructive pulmonary disease and asthma, are known to have a strong inflammatory element to their pathology. Genetics has greatly advanced our understanding of the pathobiology underlying complex genetic disorders such as asthma, with numerous genetic loci and individual genes showing a linkage or association. Proteomics has the ability to complement these genetic results by generating biomarkers that may prove invaluable as prognostic tools for predicting asthma susceptibility and disease severity as well as predictors of drug response. As part of the drive to deliver such biomarkers, proteomics has been applied to a number of biofluids, such as bronchoalveolar lavage, together with plasma and serum.

Proteomics has also been deployed to widen the understanding of the therapeutic value of vaccines. For example, in a relatively sophisticated approach to define novel vaccine candidates, 2-DE has been incorporated into studies to define immunogenic proteins, which elicit a host response. This has been reported for therapeutic areas as diverse as infectious diseases and oncology. The latter raises the prospect of developing future cancer vaccines, making provision for a major unmet therapeutic need. Similar approaches have

been applied to advancing the diagnosis and treatment of transplant rejection via identification of early events in the process of rejection.

Research in cardiovascular biology in the pregenomic era focused on classical pathophysiological descriptions, with the later incorporation of receptor biology and signal transduction–driven approaches. Cardiovascular research is a discipline for which there are many *in vitro* and *in vivo* model systems available, which provide a vehicle for a more detailed understanding of the molecular basis of disease. In the postgenomic era, proteomics will provide an important role, bridging the gap between genetic and physiological research. Atherosclerosis is a disease of the cardiovascular system, which is now generally accepted to have a large inflammatory component. Several risk factors have been identified including cell adhesion molecules and inflammatory markers such as C reactive protein. Proteomics has recently been instrumental in discovering several putative protein markers associated with atherosclerosis, many of which have a recognized role in inflammation.

Applications of proteomics in the field of oncology are quite diverse, focusing on advancing the understanding of mechanisms, diagnosis, and prognosis, as well as defining possible therapeutic regimes in a number of cancers. One important role has been to advance the definition of mechanisms underlying cellular regulation, including those implicated in the transcriptional regulation of metastasis. Protein expression in hepatocellular carcinoma cell lines, which has proven to be a useful model of persistent viral hepatitis–induced hepatoma, has revealed interesting new proteins and confirmed others previously implicated in the disease. Proteomics has also been used to gain a better understanding of the multifactorial nature of chemoresistance demonstrated by certain cancer cells. The oncology area has been extensively studied using DNA chips (microarrays) for transcript analysis, hence there is a wonderful opportunity for synergy between transcriptomics and proteomics by comparing and contrasting the data for validation of new targets or new disease diagnostic or disease/drug prognostic markers.

In the fields of neurology and psychiatry, the understanding of neurologically relevant systems has been enhanced by significant progress in mapping the human brain proteome. This has been further advanced by the investigation of model organisms such as the mouse and *in vitro* cell systems such as synaptosomes. Alzheimer's disease is a complex genetic disorder in which dysfunction is associated with mutational loci (amyloid precursor protein, presenilin 1 and presenilin 2) as well as susceptibility loci (apolipoprotein E). Progress has been made in this area by detection of oxidatively modified proteins consistent with previously identified peroxynitrite-mediated damage and oxidatively modified RNA. Analysis of human postmortem brains from both Alzheimer's and Down syndrome patients indicates decreased levels of proteins such as synaptosomal-associated protein 25 (snap 25), possibly reflecting impaired synaptogenesis or neuronal loss. Postmortem studies and antemortem analysis of cerebrospinal fluid have also revealed several additional proteins found to be altered in expression in Alzheimer's disease, including apolipoprotein E.

In psychiatric research, the application of genomics and proteomics to disorders as diverse as depression, schizophrenia, and autism is being considered as a major

potential contributor in terms of disease diagnosis, understanding, and therapy. New avenues into disease understanding have been illuminated by genetic studies, for example, the discovery of neuregulin, DAOO, and G72 (the latter two associated with glutamate pathways) as schizophrenia susceptibility genes. To date, analysis of cerebrospinal fluid from schizophrenics has produced limited identifiable changes in protein expression and in general, validated markers of psychiatric disorders remain elusive. Despite this difficulty, it is likely that given clinical material of sufficient quality and with detailed disease phenotypic data, proteomics can make a major contribution to both mechanistic understanding and diagnosis.

3.2 Protein–protein Interactions

A key aspect of understanding biological systems is to gain an insight into protein function. Characterizing protein function in a normal state and defining changes in function in the dysregulated state associated with disease can be extremely challenging and while expression proteomics is certainly capable of providing important clues into protein function, investigation of how proteins interact with one another is of additional utility. Association of a protein of unknown function with proteins of known function or linkage to a defined pathway can be invaluable in delineating cellular processes and disease mechanisms.

Various methods exist for defining protein interactions including the yeast two-hybrid system and affinity pull downs or chemoproteomic approaches. In a relatively recent advance in chemoproteomics, a tandem affinity purification (TAP) tag is used in which a fusion of the binding domains of both protein A and calmodulin are employed to enhance selectivity. The two affinity regions are separated by a cleavable tobacco etch virus protease cleavage site. This approach has been applied successfully to create a functionally organized database for yeast–protein interactions, and provide the first insight of how a complex network of proteins may operate in a eukaryotic proteome.

Yeast two-hybrid systems have also been effective in defining new protein targets for pharmaceutical intervention including defining novel nuclear hormone receptors and their cofactors and in characterizing the specific residues or domains within proteins associated with protein function. For instance, the system has also been employed to find peptides, which bind and inhibit key enzymes such as kinases.

As described above, protein arrays are now being developed to define protein function and studies in yeast have already defined and characterized novel binding motifs for proteins such as calmodulin. Perhaps the most immediate benefit of protein interaction studies will be the assembly of large databases of hitherto undefined pathways, which will assist in associating proteins with their functions or, more broadly, with specific pathways and cellular processes.

3.3 Protein Biomarkers in Clinical and Preclinical Research

The discovery and successful application of therapeutic treatments to diseases is a very complicated and lengthy process. Through the development cycle of a new drug candidate, a vast number of hurdles need to be overcome. They mainly involve a clear establishment that the proposed molecule

is safe, tolerable, and therapeutically efficacious. During the process, the compound has to be carefully and thoroughly tested in preclinical animal models, and only if those studies are successful in demonstrating the above attributes will the drug progress to tightly controlled human trials. The complexity, scientific rigor, and vast amount of human and financial resources needed during the drug discovery and development process mean that finding ways to predict, diagnose, and understand better the manifestations of disease as well as the outcome of the proposed drug treatment are critical factors to be addressed.

Alongside transcriptomics and metabolomics, proteomics has the potential to dramatically improve the speed of clinical research, discriminating between diseases and disease subtypes. Proteomics applied to biofluids such as plasma offers a relatively noninvasive method of providing bedside/near bedside clinical support. In particular, proteomics has major potential for finding novel biomarkers that can act as predictive signatures of disease onset and progression, act as surrogate markers of therapeutic efficacy, or reveal potential toxicology. The search for novel protein biological biomarkers that can predict at an early stage the efficacy and safety profiles of a drug has benefited from a concerted and vigorous increase in effort in the last decade.

Proteomics technologies are playing a key role in the search for biomarkers, often complementing other approaches such as transcriptomics and genetics. While acknowledging recent advances, initial progress in plasma/serum proteomics was slow owing to the complexity and wide dynamic range of the protein content. For example, while the concentration of serum albumin in normal plasma is in the range of $35-50\ mg \cdot mL^{-1}$, the concentration of a cytokine such as interleukin IL6 is 10^9 times lower at $0-5\ pg \cdot mL^{-1}$. Methods for removal of albumin and other abundant proteins from plasma have now dramatically improved the scope for clinical proteomics in biofluids. In both preclinical as well as clinical arena, the use of 2-DE gel electrophoresis techniques in conjunction with mass spectrometry as well as the use of immunobased approaches is now providing important and useful data.

Adverse drug reactions are a significant issue both in terms of molecules failing both in preclinical development and in the clinic. Since the estimated cost of bringing a drug to the market is $800 million, any improvement in attrition rates, however small, will accrue significant benefits both to pharmaceutical companies and also to the patients themselves. Initially, the focus of genomic technologies in the area of toxicology was centered on transcriptomics, with the analysis of RNA expression in tissues obtained from animals treated with test compounds. However, the parallel analysis of genes and proteins has brought new insight to both *in vitro* and *in vivo* studies with the goal to predict the toxic effects of potential of new potential drugs much earlier in the development cycle. The aim of predictive toxicology via proteomics is to improve lead and candidate selection and enhance drug efficacy and safety monitoring via profiling of defined surrogate markers from easily obtainable biofluids. Proteomics via 2-DE enables drug-related alterations in expression profiles to be compared and potential biomarkers for further scrutiny to be identified. Several studies have already begun to define protein expression changes, which correlate with drug-mediated toxicity, whereas specific studies have indicated possible mechanisms of toxicity.

Recently, the surface-enhanced laser desorption ionization (SELDI™) time of flight (TOF) technology has been developed. Proteins in a mixture are selectively bound to a surface of defined chemistry and their masses profiled by mass spectrometry. Although the identity of the proteins cannot usually be ascertained from this type of data, a specific mass spectrum can be obtained and used to characterize a specific analyte. This has been applied to define potential biomarkers of ovarian, breast, and prostate cancer.

Application of "omic"-based technologies can offer additional insights over and above the growing area of pharmacogenetics. Pharmacoproteomics represents a more phenotypic approach to the analysis of patient-to-patient variability than is provided by genotyping as it will assess the effects of an individual's genetic makeup as well as the influences of the environment and lifestyle. This makes proteomics a valuable tool in disease diagnosis and the push toward defining subgroups of patients who are more suited to the individual therapies.

4 Future Perspectives

4.1 Toward a Full Proteome Map

The genomes of higher eukaryotes give rise to proteomes of high complexity owing to the predominance of posttranslational modifications and splice variants. Indeed, the level of complexity appears to correlate with that of the organism. It has been estimated that the number of proteins per gene is around 1.1 to 1.3 for bacteria, 3 for yeast and around 10 for humans, although these are estimates and may only hint at the actual levels of complexity.

In addition, the true relevance of a protein function cannot be derived from information about a single protein in isolation. The activity of a given protein will be strongly influenced by its tissue distribution and intracellular localization, as well as its proximity to and interaction with other proteins and metabolites. These factors illustrate the increasing need to move the technology toward total proteome coverage.

Although a laudable aim, currently no method is in sight that will allow the analysis of the entire proteome. Progress toward significantly higher proteome coverage is being made via prefractionation techniques to analyze subproteomes. Several studies have now analyzed proteomes at the organellar level and even at the suborganellar level. This contributes more information about the location, expression, and function of proteins, and these methods for analysis are becoming increasingly sensitive, robust and quantitative. Historically, certain types of proteins have been difficult to study, including highly hydrophobic membrane-bound proteins. Although membrane proteins have been typically underrepresented during analysis by 2-DE, progress in extraction methods for membrane proteins and the introduction of improved solubilization techniques, both in sample preparation and separation, have markedly improved membrane protein detection. In addition, the ability to select individual protein classes such as glycosylated proteins or phosphoproteins via antibodies or specific tags has increased the scope and range of proteome subset analyses. Advances in the technology and, in particular, the recent advances in nongel-based proteomics may

herald a major advance toward increased proteome coverage.

4.2 Biological Validation of Protein Functions

As mentioned a number of times in this review, proteomics offers tremendous opportunities to rapidly progress biological and medical research. The ability to analyze and potentially quantify many proteins at once in a multiplexed assay without any preselection makes proteomics a truly open multiplexed discovery platform. However, experience to date has shown that the wealth of information created can be extremely difficult to process and presents a number of novel issues. The multiplexed nature of the assay increases the chances of discovering many proteins associated with a pathology or biological phenotype. It also increases the probability of identifying changes, which occur purely by chance, or as an indirect result of the up- or downregulation of relevant proteins. Thus, the necessity for rigorous validation of protein changes associated with altered phenotypes is of paramount importance and places a large burden on what is currently a potentially resource and rate-limiting step.

While many proteins such as enzymes can be assayed functionally when defined substrates are available, individual immunoassays have long been one of the definitive approaches to measuring individual protein expression. However, as discussed above, the limited availability of high-quality antibody reagents has led to significant effort being expended on provision of antibody reagents for assays and protein arrays. Further, the creation of international consortia is being proposed to undertake and complete the monumental task of producing antibodies against every human protein. Additional validation can also be gleaned from specific pathway expansion, gene silencing and gene knockout experiments.

4.3 Clinical and Diagnostic Applications

A number of issues need to be addressed before proteomics can achieve general and routine applicability in clinical studies. In addition to the caveats of working with plasma and serum, as discussed above, there are issues surrounding the heterogeneous nature of human samples. Biological systems are complex and study populations from clinical cohorts tend to be highly variable. Many clinical studies are derived from populations representing diversity in genetic and environmental and lifestyle factors, as well as in age, disease severity and comorbidity. All of these factors conspire to confound interpretation of clinical proteomic data. The capacity to run proteomic analysis at a scale consistent with large populations is needed to facilitate extraction of biologically and statistically significant information from the confounding elements. The industrialization of 2-DE is not as advanced as is needed to address very large clinical cohorts, although it is now capable of processing thousands of samples. The immediate promise of the emerging techniques such as LC/LC-MS/MS and protein arrays will be to augment information content attainable with 2-DE rather than to provide dramatically increased throughput.

Progress to date in the area of clinical proteomics has been rapid and it will surely continue to improve at an equally impressive rate given the huge demand and benefits. The potential for defining

biomarkers for therapeutic or diagnostic use, as well as defining biological mechanisms is immense.

Acknowledgments

The authors wish to thank Jingwen Chen, Bob Hollingsworth, Arthur Moseley, Ian White, and Julia White for their help in preparing the figures, and Mike Trower and Mark Skehel for helpful comments on the manuscript.

See also Genetics, Molecular Basis of; Oncology, Molecular; Protein Modeling; Protein NMR Spectroscopy; Protein Splicing; Serial Analysis of Gene Expression.

Bibliography

Books and Reviews

Aebersold, R., Mann, M. (2003) Mass spectrometry-based proteomics, *Nature* **422**, 198–207.

Anderson, N.L., Anderson, N.G. (2002) The human plasma proteome: history, character and diagnostic prospects, *Mol. Cell. Proteomics* **1**, 845–867.

Cutler, P. (2003) Protein arrays: current state of the art, *Proteomics* **3**, 3–18.

Patterson, S.D., Aebersold, R.H. (2003) Proteomics: the first decade and beyond, *Nat. Genet.* **33**(suppl.), 311–323.

Simpson, R.J. (2002) *Proteins and Proteomics: A Laboratory Manual*, Cold Spring Harbor Press, New York.

Walgren, J.L., Thompson, D.C. (2004) Application of proteomic technologies in the drug development process, *Toxicol. Lett.* **149**, 377–385.

Zerkowski, H.-R., Grussenmeyer, T., Matt, P., Grapow, M., Engelhardt, S., Lefkovits, I. (2004) Proteomics strategies in cardiovascular research, *J. Proteome Res.* **3**, 200–208.

Zhu, H., Snyder, M. (2003) Protein Chip Technology, *Curr. Opin. Chem. Biol.* **7**, 55–63.

Zolg, J.W., Langen, H. (2004) How industry is approaching the search for new diagnostic markers and biomarkers, *Mol. Cell. Proteomics* **3**, 345–354.

Primary Literature

Arrell, D., Neverova, I., van Eyk, J. (2001) Cardiovascular proteomics: evolution and potential, *Circ. Res.* **88**, 763–773.

Baak, J.P.A., Path, F.R.C., Hermsen, M.A.J.A., Meijer, G., Schmidt, J., Janssen, E.A.M. (2003) Genomics and proteomics in cancer, *Eur. J. Cancer* **39**, 1199–1215.

Bailey, W.J., Ulrich, R. (2004) Molecular profiling approaches for identifying novel biomarkers, *Expert Opin. Drug Safety* **3**, 137–151.

Bandara, L.R., Kennedy, S. (2002) Toxicoproteomics – a new preclinical tool, *Drug Discov. Today* **7**, 411–418.

Beranova-Giorgianni, S. (2003) Proteome analysis by two-dimensional gel electrophoresis and mass spectrometry: strengths and limitations, *Trends Anal. Chem.* **22**, 273–281.

Beroza, P., Villar, H.O., Wick, M.M., Martin, G.R. (2002) Chemoproteomics as a basis for postgenomic drug discovery, *Drug Discov. Today* **7**, 807–814.

Blobel, G., Wozniak, R. (2000) Proteomics for the pore, *Nature* **403**, 835–836.

Bodovitz, S. (2003) The protein biochip content problem, *Drug Discov. World* **4**, 50–60.

Cacabelos, R. (2002) Pharmacogenomics for the treatment of dementia, *Ann. Med.* **34**, 357–379.

Chakravati, D., Fiske, M., Fletcher, L., Zagursky, R. (2001) Application of genomics and proteomics for identification of bacterial gene products as potential vaccine candidates, *Vaccine* **19**, 601–612.

Coates, P.J., Hall, P.A. (2004) The yeast two-hybrid system for identifying protein-protein interactions, *J. Pathol.* **199**, 4–7.

Cordwell, S.J., Nouwens, A.S., Walsh, B.J. (2001) Comparative proteomics of bacterial pathogens, *Proteomics* **1**, 461–472.

Davidson, P., Westman-Brinkmalm, A., Nilsson, C., Lindbjer, M., Paulson, L., Andreasen, N., Sjorgen, M., Blennow, K. (2002)

Proteome analysis of cerebrospinal fluid proteins in Alzheimer patients, *Clin. Neurosci. Neuropathol.* **13**, 611–615.

Edman, P. (1970) Sequence determination, *Mol. Biol. Biochem. Biophys.* **8**, 211–255.

Evangelista, C., Lockshon, D., Fields, S. (1996) The yeast two-hybrid system: prospects for protein linkage maps, *Trends Cell Biol.* **6**, 196–199.

Flory, M., Griffin, T., Martin, D., Aebersold, R. (2002) Advances in quantitative proteomics using stable isotope tags, *Trends Biotechnol.* **20**, S23–S29.

Furness, L.M. (2002) Analysis of gene and protein expression for drug mode of toxicity, *Curr. Opin. Drug. Discov. Devel.* **5**, 98–103.

Garfin, D.E. (2003) Two-dimensional gel electrophoresis: an overview, *Trends Anal. Chem.* **22**, 263–272.

Grandi, G. (2001) Antibacterial vaccine design using genomics and proteomics, *Trends Biotechnol.* **19**, 181–188.

Grunenfelder, B., Rummel, G., Vohradsky, J., Roder, D., Langen, H., Jenal, U. (2001) Proteomic analysis of the bacterial cell cycle, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4681–4686.

Halapi, E., Hakonarson, H. (2004) Recent development in genomic and proteomic research for asthma, *Curr. Opin. Pulm. Med.* **10**, 22–30.

Hegde, P.S., White, I.R., Debouck, C. (2003) Interplay of transcriptomics and proteomics, *Curr. Opin. Biotechnol.* **14**, 647–651.

Herbert, B.R., Harry, J.L., Packer, N.H., Gooley, A.A., Pedersen, S.K., Williams, K.L. (2001) What place for polyacrylamide in proteomics? *Trends Biotechnol.* **19**(Suppl. 10), S3–S9.

Hoving, S., Voshol, H., van Oostrum, J. (2000) Towards high performance two-dimensional gel electrophoresis using ultrazoom gels, *Electrophoresis* **21**, 2617–2621.

Hutter, G., Sinha, P. (2001) Proteomics for studying cancer cells and the development of chemoresistance, *Proteomics* **1**, 1233–1248.

Ideker, T., Thorsson, V., Ranish, J., Christmas, R., Buhler, J., Eng, J., Bumgarner, R., Goodlett, D., Aebersold, R., Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network, *Science* **292**, 929–934.

Jiang, L., Lindpaintner, K., Li, H.-F., Gu, N.-F., Langen, H., Fountoulakis, M. (2003) Proteomic analysis of the cerebrospinal fluid of patients with schizophrenia, *Amino Acids* **25**, 49–57.

Jain, K.K. (2004) Role of pharmacogenomics in the development of personalised medicine, *Pharmacogenomics* **5**, 331–336.

Johnson, J.R., Florens, L., Carucci, D.J., Yates, J.R. (2004) Proteomics in malaria, *J. Proteome Res.* **3**, 296–306.

Jungblut, P., Thiede, B. (1997) Protein identification from 2-DE gels by MALDI mass spectrometry, *Mass Spec. Rev.* **16**, 145–162.

Lander, E.S. et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Le Naour, F. (2001) Contribution of proteomics to tumor immunology, *Proteomics* **1**, 1295–1302.

Link, A. (2002) Multidimensional peptide separations in proteomics, *Trends Biotechnol.* **20**, S8–S13.

Lubec, G., Krapfenbauer, K., Fountoulakis, M. (2003) Proteomics in brain research: potentials and limitations, *Prog. Neurobiol.* **69**, 193–211.

Luo, J., Isaacs, W.B., Trent, J.M., Duggan, D.J. (2003) Looking beyond morphology: cancer gene expression profiling using DNA microarrays, *Cancer Invest.* **21**, 937–949.

Martin, D., Nelson, P. (2001) From genomics to proteomics: techniques and applications in cancer research, *Trends Cell Biol.* **11**, S60–S65.

O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins, *J. Biol. Chem.* **250**, 4007–4021.

Ong, S.-E., Foster, J., Mann, M. (2003) Mass spectrometric-based approaches in quantitative proteomics, *Methods* **29**, 124–130.

Petricoin, E.F., Liotta, L.A. (2002) Proteomic analysis at the bedside: early detection of cancer, *Trends Biotechnol.* **20**, S30–S34.

Petricoin, E.F., Zoon, K.C., Kohn, E.C., Barrett, J.C., Liotta, L.A. (2002) Clinical Proteomics: Translating benchside promise into bedside reality, *Nat. Rev. Drug Discov.* **1**, 683–695.

Pieper, R., Gatlin, C.L., Makusky, A.J., Russo, P.S., Schatz, C.R., Miller, S.S., Su, Q., McGrath, A.M., Estock, M.A., Parmer, P.P., Zhao, M., Huang, S.-T., Zhou, J., Wang, F., Esquer-Biasco, R., Anderson, N.L., Taylor, J., Steiner, S. (2003) The human serum

proteome: display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis and identification of 325 distinct proteins, *Proteomics* **3**, 1345–1364.

Rosenblatt, K.P., Bryant-Greewood, P., Killian, J.K., Mehta, A., Geho, D., Espina, V., Perticoin, E.F., Liotta, L.A. (2004) Serum Proteomics in Cancer diagnosis and management, *Annu. Rev. Med.* **55**, 97–112.

Ryan, T., Patterson, S. (2002) Proteomics: drug target discovery on an industrial scale, *Trends Biotechnol.* **20**, S45–S51.

Santoni, V., Molloy, M., Rabilloud, T. (2000) Membrane proteins and proteomics. Un amour impossible? *Electrophoresis* **21**, 1054–1070.

Smith, R., Shen, Y., Tang, K. (2004) Ultrasensitive and quantitative analyses from combined separations – mass spectrometry for the characterisation of proteomics, *Acc. Chem. Res.* **37**, 269–278.

Smolka, M., Zhou, H., Purkayasstha, S., Aebersold, R. (2001) Optimization of the isotope-coded affinity tag-labelling procedure for quantitative proteome analysis, *Anal. Biochem.* **297**, 25–31.

Srinivas, P.T., Verma, M., Zhao, Y., Srivastava, S. (2002) Proteomics for cancer biomarker discovery, *Clin. Chem.* **48**, 1160–1169.

Steiner, S., Anderson, N.L. (2000) Expression profiling in toxicology – potentials and limitations, *Toxicol. Lett.* **112–113**, 467–471.

Tao, W., Aebersold, R. (2003) Advances in quantitative proteomics via stable isotope tagging and mass spectrometry, *Curr. Opin. Biotechnol.* **14**, 110–118.

Taylor, S., Fahy, E., Ghosh, S. (2003) Global organellar proteomics, *Trends Biotechnol.* **21**, 82–88.

Walduck, A., Rudel, R., Meyer, T.F. (2004) Proteomic and gene profiling approaches to study host responses to bacterial infection, *Curr. Opin. Microbiol.* **7**, 33–38.

Washburn, M., Wolters, D., Yates, J. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology, *Nat. Biotechnol.* **19**, 242–247.

Wilson, K.E., Ryan, M.M., Prime, J.E., Pashby, D.P., Orange, P.R., O'Beirne, G., Whateley, J.G., Bahn, S., Morris, C.M. (2004) Functional genomics and proteomics: application in neurosciences, *J. Neurol. Neurosurg. Psych.* **75**, 529–538.

Zhu, H., Klemic, J., Chang, S., Bertone, P., Casamayor, A., Klemic, K., Smith, D., Gerstein, M., Reed, M., Snyder, M. (2000) Analysis of yeast protein kinases using protein chips, *Nat. Genet.* **26**, 283–289.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R., Gerstein, M., Snyder, M. (2001) Global analysis of protein activities using proteome chips, *Science* **293**, 2101–2105.

Proton Translocating ATPases

Masamitsu Futai[1], *Ge-Hong Sun-Wada*[2], *and Yoh Wada*[2]
[1] *Microbial Chemistry Research Foundation and CREST, Japan Science and Technology Agency, Kamiosaki, Shinagawa, Tokyo, Japan*
[2] *ISIR, Osaka University, Ibaraki, Osaka, Japan*

1 Introduction 335

2 Catalytic Mechanism of Proton-translocating F-ATPase 337

3 Roles of the γ-Subunit: Energy Coupling by Mechanical Rotation 338
3.1 Roles of the γ-Subunit in Energy Coupling 338
3.2 Subunit γ-Rotation 340
3.3 Mutational Analysis of the γ-Subunit Rotation 342

4 Rotational Catalysis of the F-ATPase Holoenzyme 344
4.1 Structure of F_0Sector and Proton Transport Pathway 344
4.2 Rotational Catalysis of the F-ATPase Holoenzyme 345
4.3 Rotational Catalysis of F-ATPase in Membranes 347

5 Rotational Catalysis of V-ATPASE 350
5.1 Catalytic Site and Proton Pathway 350
5.2 Subunit Rotation of V-ATPase During Catalysis 351

6 Epilogue 353

Bibliography 353
Books and Reviews 353
Primary Literature 354

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

Keywords

F_0F_1
ATP synthase or proton-translocating ATPase formed from membrane extrinsic (F_1) and intrinsic (F_0) sectors. Nomenclature from factors of oxidative phosphorylation.

F-ATPase
One of the three classes of ion transport ATPases, mostly proton-translocating ATPase. Nomenclature from F_0F_1 or factors of oxidative phosphorylation.

P-type ATPase
Ion transport ATPase forming acylphosphoenzyme intermediates during catalysis. This group contains proton-translocating ATPases such as H^+/K^+ ATPase and H^+ATPase. They are different from F-ATPase, or V-ATPase, but similar to Na^+/K^+ ATPase.

Rotational Catalysis
Continuous rotation of $\gamma\varepsilon$ C_{10} subunit complex during ATP synthesis or hydrolysis by F-ATPase. Similar mechanism found in V-ATPase.

γ-Subunit
F-ATPase subunit located at the center of the catalytic hexamer ($\alpha_3\beta_3$) formed from the α- and β-subunits, and rotating during catalysis.

V-ATPase
Vacuolar-type proton-translocating ATPase. Nomenclature from a proton pump identified initially in fungal or plant vacuoles, but now found in a wide variety of endomembrane organelles such as lysosomes and endosomes and plasma membrane of specialized cells such as osteoclast or kidney intercalated cell.

Abbreviations

DCCD: dicyclohexylcarbodiimide
Pi: inorganic phosphate

Proton-translocating ATPases synthesize or hydrolyze ATP coupling with electrochemical proton gradient, and have important roles in animal and plant physiology. The ion transporting ATPases are classified as F-ATPases (F-type ATPases), V-ATPases (V-type ATPases), and P-ATPases (P-type ATPase). The F-ATPase, also called *ATP synthase*, is found in chloroplasts, mitochondria, or bacterial membranes, and synthesizes most of the ATP, the biological energy currency. The enzyme is formed basically from eight subunits, *a*-, *b*-, and *c*-subunits forming membrane intrinsic F_0 sector with ab_2c_{10} stoichiometry, and α-, β-, γ-, δ-, and

ε-subunits forming peripheral F_1 sector with $\alpha_3\beta_3\gamma\delta\varepsilon$ stoichiometry. Three catalytic β-subunits forming a hexamer with α-subunit ($\alpha_3\beta_3$) have catalytic cooperativity. The $\gamma\varepsilon c_{10}$ subunits complex, formed from peripheral ($\gamma\varepsilon$) and intrinsic membrane (c) subunits is a cential stalk, rotating during catalysis. The V-ATPase forms inside acidic pH in endomembrane organelles such as lysosomes, endosomes, synaptic vesicles, and so on, and is similar to F-ATPase in subunit organization and structure. The same enzyme is found in plasma membrane of special cells such as osteoclast and kidney intercalated cell. Gastric proton pump H^+/K^+ATPase or plant plasma membrane H^+ATPase belongs to P-ATPases forming acyl phosphate intermediate, and is not discussed in this article.

1 Introduction

The mechanism of ATP synthesis has been a focus of biochemists for more than four decades. ATP synthase was first identified in the mitochondrial inner membrane, and then found successively in chloroplasts and bacterial membranes. This enzyme synthesizes ATP from ADP and phosphate (Pi) coupled with an electrochemical proton gradient, and is also called *proton-translocating ATPase* or F-ATPase (F-type ATPase) because of the reversible proton pumping upon ATP hydrolysis. The name F-ATPase originated from the coupling factor of oxidative phosphorylation sensitive to oligomycin. *Escherichia coli* F-ATPase is composed of a membrane extrinsic F_1 sector and a transmembrane F_0, formed from five ($\alpha_3\beta_3\gamma\delta\varepsilon$) and three ($ab_2c_{10-14}$) subunit assemblies with different stoichiometries, respectively (Fig. 1). It can also be divided into a catalytic $\alpha_3\beta_3$ hexamer, stalks ($\gamma\varepsilon ab_2$), and membrane (ab_2c_{10-14}) domains. Two stalks (central and peripheral stalk) have been observed by electron microscopy. The mitochondrial enzyme has additional subunits, possibly with regulatory functions.

The higher-ordered X-ray structure of the bovine $\alpha_3\beta_3\gamma$ complex had been solved in 1994 by John Walker and coworkers. Following this breakthrough, the structures of crystals of bovine F_1, inhibited by efrapeptin, aurovertin, NBD-Cl (7-chloro-4-nitrobenzo-2-oxa-1,3-diazole), and DCCD (dicyclohexylcarbodiimide) were solved by the same group. Bovine F_1 containing 1 mol MgADP trifluoroaluminate and 2 mol MgADP trifluoroaluminate has also been crystallized, and the structures have been determined. These studies, together with biochemical analysis, have contributed greatly to an understanding of the catalytic site and mechanism. The structures of F_1 sectors of other origins have also been reported.

The catalytic site is mainly located in the β-subunit of the $\alpha_3\beta_3$ hexamer, and the three sites (one in each β) show strong cooperativity. The amino- and carboxyl-terminal helices of the γ-subunit are located in the center of the $\alpha_3\beta_3$ hexamer, and form a central stalk with the ε-subunit. The peripheral or second stalk is formed from the a- and b-subunits, and the δ-subunit is located near the top of the α-subunit. The proton pathway is located at the interface between

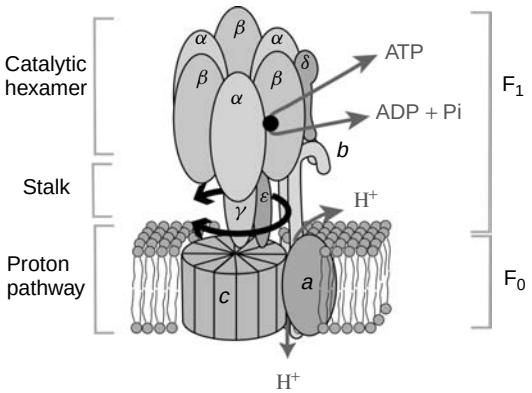

Fig. 1 Schematic model of proton-translocating F-ATPase, showing the subunit structures of the catalytic hexamer ($\alpha_3\beta_3$), stalk and membrane domain. The membrane extrinsic F_1 and transmembrane F_0 sector are shown together with energy coupling between the ATP synthesis–hydrolysis and subunit rotation.

the *a*- and *c*-subunits, and the hairpin structure of the purified *c*-subunit has been solved by Fillingame and coworkers using NMR. A ring structure formed from multiple *c*-subunits was suggested by early studies involving electron and atomic force microscopy, and was extensively analyzed recently through NMR structure and genetic approaches. The X-ray structure of yeast F_1 with a *c*-subunit ring has also been reported by Stock et al.

The mechanism of coupling of proton transport and ATP synthesis or hydrolysis has been a major question for this complicated enzyme. The binding change mechanism proposes rotation of the γ-subunit relative to the $\alpha_3\beta_3$ hexamer coupled with the chemistry at the catalytic sites. Rotation of the $\gamma\varepsilon c_{10-14}$ complex relative to $\alpha_3\beta_3\delta ab_2$ upon ATP addition was shown recently, indicating that continuous rotation of the assembly of F_1 and F_0 subunits is involved in the coupling between ATP hydrolysis and proton transport.

Vacuolar-type ATPase (V-ATPase) with significant similarity to F-ATPase was introduced later to a family of proton-translocating ATPases. V-ATPase is apparently different from F-ATPase in its physiological roles, and forms an acidic luminal pH in endomembrane organelles including lysosomes and endosomes, and in extracellular compartments such as resorption lacuna formed between osteoclasts and the bone surface. It has extrinsic membrane V_1 and V_0 transmembrane domains formed from the *A*, *B*, *C*, *D*, *E*-, *F*-, *G*-, and *H*-subunits, and *a*, *c*, *c*′, *c*″, and *d*, respectively. F- and V-ATPase share significant homology, especially in its catalytic subunits and proton pathways, and also share some unique differences, including its membrane sector and stalk region subunit compositions. A series of isoforms have been found for V-ATPase subunits. Thus, comparative studies of the two proton–translocating ATPases are pertinent for understanding the molecular mechanisms of both.

In this chapter, we discuss recent progress made in the understanding of F-ATPase, focusing mainly on the energy coupling between the chemistry and proton transport through subunit rotation, and also discuss V-ATPase from a similar aspect. It should be interesting for readers to follow a series of biochemical studies to show rotational catalysis of the F-ATPase holoenzyme leading to that of the V-ATPase.

2
Catalytic Mechanism of Proton-translocating F-ATPase

As expected from its complicated structure, F-ATPase is not a simple Michaelis–Menten type enzyme. Furthermore, the overall mechanism includes catalysis (chemistry), subunit rotation and proton translocation. X-ray structure and kinetic studies, especially substrate binding analysis with an intrinsic tryptophan probe, revealed that the three catalytic sites are asymmetric. Senior and coworkers have recently discussed and summarized the molecular catalytic mechanism of F-ATPase.

ATP synthesis and hydrolysis could be carried out kinetically at a single site (unisite catalysis), two sites operating together (bisite catalysis), or all three sites working together (trisite catalysis). Unisite catalysis has only been demonstrated for ATP hydrolysis, and can be measured experimentally with an ATP:F_1 ratio of less than 1 : 3. This rate is 10^5–10^6-fold lower than that of steady state (multisite) catalysis. The enzyme cross-linked chemically and thus could not rotate, but could still carry out unisite catalysis. This catalysis is not a part of the steady state, which includes subunit rotation.

Catalytic residues have been identified by analyzing unisite catalysis of the purified *E. coli* mutant F_1 sector (Fig. 2). Briefly, βLys155 of the β-subunit is required for binding of the γ-phosphate moiety, as shown by studies involving affinity labeling and mutant enzymes such as βLys155Ala (βLys155 $\rightarrow$ Ala) or βLys155Ser. The βArg182 residue is also involved in the binding. Enzymes with substitutions of βThr156 showed similar properties to those of βLys155. The hydroxyl moiety of βThr156 is essential, possibly for Mg^{2+} binding, since it can only be replaced by a serine residue. βGlu181 is a critical catalytic residue. Its side chain forms a hydrogen bond with a water molecule located near the γ-phosphate of ATP. However, it was shown later that this residue was not involved in nucleotide binding or Mg^{2+} coordination. βTyr331, which is stacked close to the adenine ring, is required for the binding of ADP or ATP. Results of analysis of the tryptophan fluorescence of the βTyr331Trp mutant F_1 are consistent

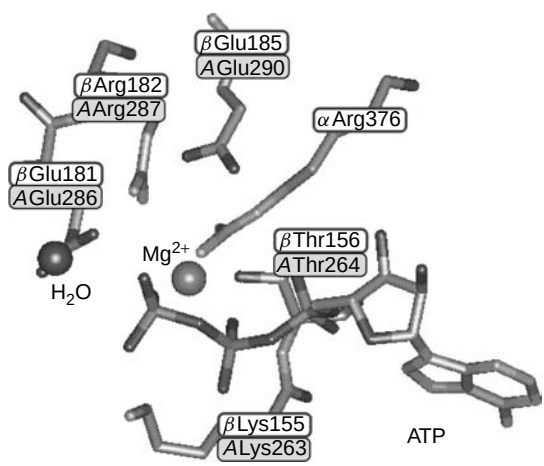

Fig. 2 Catalytic sites of F-ATPase and V-ATPase. The catalytic residues of *E. coli* F-ATPase are shown together with bound ATP. Their positions are cited according to the bovine crystal structure. Corresponding residues of yeast V-ATPase are also shown.

with the role of the βTyr331 residue. The bovine residues corresponding to those discussed above are located close to the phosphate moiety or the adenine ring of bound ATP or ADP as shown in the X-ray structure (Fig. 2).

Recent results, including the F_1 structure of all three sites filled with nucleotides, support trisite catalysis, in which the three sites are working together during the steady state, and the notion that bisite catalysis does not occur. We leave convincing discussions of these points to the article of Senior and coworkers.

It can be assumed that mutant enzymes defective in steady state catalysis should show impaired unisite catalysis. In this regard, αArg376 mutant enzymes are of interest. The αArg376 residue of the α-subunit is located close to the β- or γ-phosphate of ATP or ADP and Mg at the catalytic site (Fig. 2). However, αAg376 does not directly participate in the chemistry of ATP hydrolysis or synthesis, as shown by the unisite catalysis of mutant F_1 sectors. The αArg376Lys or αArg376Ala mutant enzyme showed 2×10^3-fold lower steady state ATP hydrolysis than the wild type. However, the mutant enzymes showed essentially the same kinetics for unisite catalysis as the wild type, suggesting that they can pass through the transition state. These results indicate that αArg376 is essential for promotion of catalysis to the steady state turnover. This notion is different from the previously suggested roles of αArg376 that was deduced from the structural model and fluoroaluminate binding, an indicator of the formation of the pentacovalent transition state. The βGlu185 residue, located close to the γ-phosphate and Mg at the catalytic site, may have a similar role to αArg376 because the mutant enzymes maintain unisite catalysis, but are defective in multisite catalysis. However, detailed analysis could not be carried out because the mutant F_1 was unstable after solubilization from membranes.

The binding change mechanism of Boyer proposes that the three sites are involved sequentially in ATP synthesis or hydrolysis: at a specific point of time during the steady state, the chemical reaction occurs reversibly at one site and ATP release and/or binding of ATP + Pi occurs at the two other sites with the expenditure of energy. Evidence supporting catalytic cooperativity and the binding change mechanism has accumulated, as reviewed by Boyer in 1997. It includes early kinetic evidence of unisite and steady state catalysis showing three K_m values and three interacting nucleotide binding sites, ^{18}O isotope exchange reactions, inhibitor studies, and so on. These studies suggested that the chemistry, "ADP + Pi $\leftrightarrow$ ATP + H_2O," at the catalytic site is reversible, and provides essentially no energy change. The mechanism is strongly supported by the X-ray structure showing asymmetric catalytic sites and predicting continuous γ-subunit rotation. However, it cannot be concluded that the binding change mechanism was proven at the molecular level, although the mechanism is conceptually accepted and has been extremely useful for understanding the enzyme.

3
Roles of the γ-Subunit: Energy Coupling by Mechanical Rotation

3.1
Roles of the γ-Subunit in Energy Coupling

The essential role of the γ-subunit in catalysis and assembly was shown by early reconstitution experiments: a catalytic core

complex exhibiting ATPase activity could be reconstituted from the purified *E. coli* α-, β-, and γ-subunits, but not without the γ-subunit. Consistently, assembly of the F_1 sector is strongly affected by γ-subunit mutations, especially those of residues interacting with the β-subunit. The isolated $\alpha_3\beta_3\gamma$ complex could functionally bind to the F_0 sector only after its assembly with the δ- and ε-subunits. These results indicate that the two minor subunits are required for the functional binding of $\alpha_3\beta_3\gamma$.

Early genetic approaches focused on the γ-subunit carboxyl-terminal region. The amino acid sequences of the γ-subunits are weakly conserved among different species, though their X-ray structures are similar. When the known γ-sequences are aligned, only 28 of the 286 residues of the *E. coli* subunit are conserved, and mostly in the carboxyl- and amino-terminal helices located at the center of the $\alpha_3\beta_3$ hexamer. Seventeen residues are conserved between residues 242 and 286 of the γ-subunit. An enzyme lacking 10 residues at the carboxyl terminus is still capable of *in vivo* ATP synthesis, indicating that the three conserved residues in this region are not required. Structural flexibility of the carboxyl terminus was also demonstrated by a frameshift mutation: the enzyme was still active with the γ-subunit having seven additional residues at its carboxyl terminus together with nine altered residues downstream of γThr277.

The enzyme with the nonsense mutation (γGln269End) was inactive, and substitution of a conserved residue (γGln269, γThr273, or γGlu275) between γGln269 and γLeu276 gave enzymes with reduced ATPase activity and energy coupling. We noticed that mutations resulted in different ratios of ATPase catalysis and proton transport; three mutants (βThr277End, βGln269Leu, and βGlu275Lys) exhibited about 15% of the wild-type ATPase activity, but showed various degrees of ATP-dependent H^+ transport and *in vivo* ATP synthesis. These results suggest active role(s) of the γ-subunit in ATPase activity and energy coupling.

In addition to the carboxyl terminus, the only other conserved region is near the amino terminus. The importance of this region was first suggested by the mutant lacking residues between γLys21 and γAla27, which resulted in failure of assembly of the F_1 complex. We introduced amino acid substitutions systematically into the amino-terminal region of the γ-subunit. Most of the changes between γIle19 and γLys33, γAsp83 and γCys87, or at γAsp65 had no effect, even with a drastic replacement such as hydrophobic to hydrophilic or acidic to basic. Interesting exceptions were the γMet23Arg and γMet23Lys substitutions. These mutants grew only slowly on succinate through oxidative phosphorylation, indicating that they were impaired in ATP synthesis. However, the membranes prepared from the γArg23 and γLys23 strains showed 100 and 65% of the wild-type ATPase activity, but formed only 32 and 17% of the electrochemical proton gradient respectively, indicating that these mutants are defective in energy coupling between ATP hydrolysis and proton transport. In the X-ray structure, the γMet23 residue is located close to the DELSEED (βAsp380–βAsp386) loop of the β-subunit. Thermodynamic and kinetic analyses of the purified γMet23Lys enzyme suggested that the introduced γLys23 residue forms an ionized hydrogen bond with βGlu381 in the loop. Consistent with this interpretation, the phenotype of γMet23Lys was restored by the second mutation, βGlu381Gln. The βGlu381Lys mutation also caused deficient energy

coupling. These results suggest that the interaction between the regions around γMet23 and βGlu381, and thus the γ-subunit residue and β-subunit DELSEED loop, are involved in energy coupling.

The γMet23Lys mutation was suppressed by substitution of carboxyl-terminal residues, including γArg242, γGlu269, γAla270, γIle272, γThr273, γGlu278, γIle279k, and γVal280. From these results, Nakamoto et al. suggested that γMet23, γArg242, and the region between γGlu269 and γVal280 are three interacting domains that are required for efficient energy coupling. The X-ray structure shows that γMet23 located in the amino-terminal helix is near γArg242, but γGly269 and γVal280 in the carboxyl-terminal helix are near the top of the $\alpha_3\beta_3$ hexamer. Furthermore, second-site mutations mapped to the amino (residues 18, 34, and 35) and carboxyl (residues 236, 238, 242, and 262) termini suppressed γGln269Glu and γThr273Val mutations. The higher-ordered structure clearly shows that γGlu269 or γThr273 does not interact directly with the residues of the second mutations. The occurrence of suppression at a distance may suggest that the two α-helices of the γ-subunit located at the center of the $\alpha_3\beta_3$ complex undergo long-range conformational changes during catalysis. As expected, the relative orientations of the γ-subunit to the three β-subunits (β_E, β_{DP}, and β_{TP}) are different in the crystal structure, strongly supporting γ-subunit rotation during ATP hydrolysis or synthesis.

3.2
Subunit γ-Rotation

The higher-ordered structure of $\alpha_3\beta_3\gamma$ indicated that α and β are arranged alternately around the amino- and carboxyl-terminal α-helices of the γ-subunit. The binding change mechanism predicts that the catalytic sites in the three β-subunits participate sequentially in ATP synthesis or hydrolysis via conformation transmission through the γ-subunit rotation. This rotation was suggested by experiments on chemical cross-linking between γCys87 and βCys380 (originally βAsp380) in the DELSEED loop, and analysis of polarized absorption recovery after photobleaching of a probe linked to the carboxyl terminus of the chloroplast γ-subunit.

The continuous unidirectional γ-subunit rotation in F_1 was recorded directly by Noji, Yoshida, and their colleagues (Fig. 3). They immobilized the Bacillus $\alpha_3\beta_3\gamma$ complex on a glass surface through histidine residues introduced into the β-subunit. The fluorescent actin filament connected to the γ-subunit rotated continuously in an anticlockwise direction during ATP hydrolysis. The rotation became slower with an increase in the filament length, and generated a frictional torque of ~40 pN nm. The ε-subunit rotation was also shown using the same approach, consistent with the tight association of γ and ε. A 120° step rotation was shown in the presence of a dilute ATP concentration, indicating that the γ-subunit rotates, interacting with the three β-subunits successively. Furthermore, a refined measurement system involving gold beads revealed that the 120° step could be divided into 90 and 30° steps. These two steps were proposed to correspond to ATP binding and ADP release, respectively.

The rotation of an actin filament connected to the γ-subunit of *E. coli* F_1 was observed using a similar system to that described for the Bacillus $\alpha_3\beta_3\gamma$ complex. The rotation was anticlockwise, inhibited by azide (F-ATPase inhibitor),

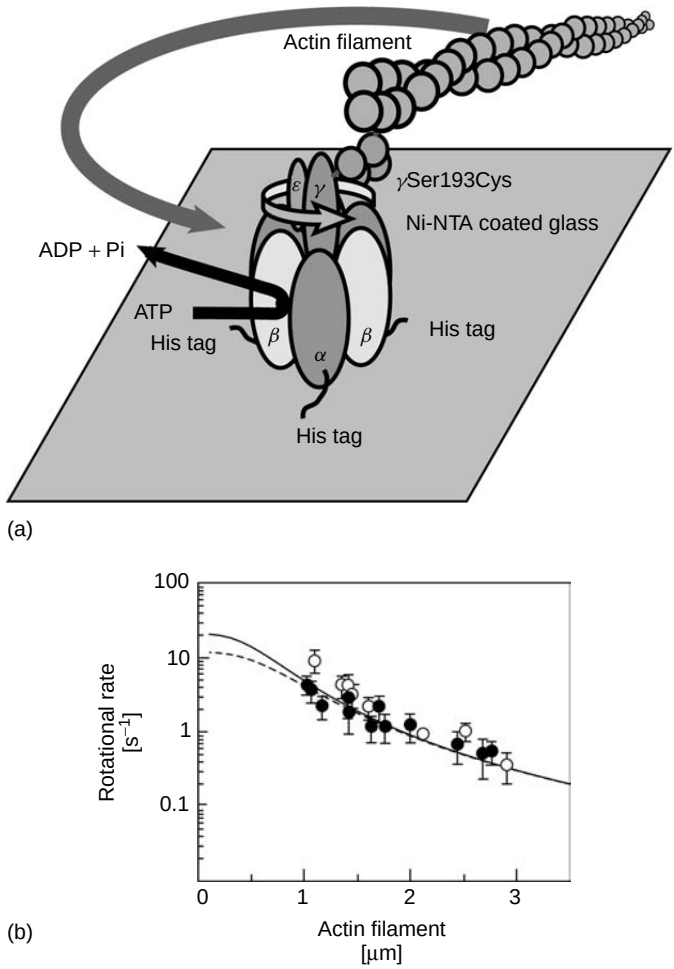

Fig. 3 Rotation of the γ-subunit of F-ATPase. F_1 was immobilized through a histidine tag introduced to the α- or β-subunit, and an actin filament was connected to the carboxyl terminus of the γ-subunit. ATP-dependent rotation of the wild-type (open circles) and γMet23Lys mutant (filled circles) F_1 sector are shown. Taken from Omote, H., Sambonmatsu, N., Saito, K., Sambongi, Y., Iwamoto-Kihara, A., Yanagida, T., Wada, Y., Futai, M. (1999) The γ-subunit rotation and torque generation in F_1-ATPase from wild-type or uncoupled mutant Esche, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7780–7784. Experimental procedure developed by Noji et al. (1997).

and generated a frictional torque of ~40 pN·nm, as reported for the Bacillus complex. The F_1 sector, proven to be a chemically driven motor, should be connected functionally to the F_0 sector to complete ATP-driven proton transport. Conversely, proton transport through F_0 should be coupled to the γ-subunit

rotation and chemistry at the catalytic sites. Questions on the energy coupling among chemistry, rotation, and proton transport will be answered with the *E. coli* enzyme by taking advantage of the accumulated genetic and biochemical information.

3.3 Mutational Analysis of the γ-Subunit Rotation

We were interested in characterizing the γ-rotation of a series of mutant enzymes defective in energy coupling and catalytic cooperativity. As described above, the γMet23Lys mutant is defective in energy coupling between catalysis and proton transport. We thought that the γ-rotation in the mutant F_1 may be defective since the γ-residue was replaced. Similar to the original mutant, the γMe23Lys enzyme engineered for rotation observation exhibited essentially the same ATPase activity as the engineered wild type. However, the enzyme could not form an electrochemical proton gradient in membrane vesicles, or carry out *in vivo* oxidative phosphorylation. Unexpectedly, an actin filament connected to the γ-subunit of γMet23Lys rotated, and generated essentially the same torque as that of the wild type (Fig. 3). These results suggest that the γMet23Lys mutant F-ATPase could couple between chemistry and rotation, but was defective in transforming mechanical work into proton translocation or vice versa. In this regard, Al-Shawi et al. showed that the γMet23Lys mutant is defective in communication between F_1 and F_0. These results also suggested that analysis of F_1 sector rotation is not enough to understand F-ATPase catalysis. It became imperative to determine which subunit complex is rotating in the F-ATPase holo enzyme purified or embedded in the membrane.

Mutant F_1 sectors with substitution of the βSer174 residue and their suppressors have been useful for understanding the rotation mechanism (Fig. 4). Replacing βSer174 with other residues altered the ATPase activity to between 150 and 10% of the wild-type level, and the larger the side chain of the residue introduced, the lower the ATPase activity observed. Both the βSer174Leu and βSer174Phe enzymes retained about 10% of the wild-type ATPase activity. However, the two mutants showed a difference in energy coupling: the βSer174Leu mutant could still grow on succinate by oxidative phosphorylation and transport protons into the isolated membrane vesicles, whereas the βSer174Phe mutant could not grow and showed no proton transport. Consistent with these observations, their F_1 sectors differed in γ-subunit rotation. The F_1 sector with βSer174Phe showed apparently slower rotation than the wild type, and generated significantly lower frictional torque ($\sim$17 pN nm), whereas the F_1 with βSer174Leu was similar to the wild type. These results suggest that the rotation or torque generation is closely related to the energy coupling with proton transport.

Biochemical defects of the βSer174Phe mutation were suppressed by a second-site mutation, βGly149 to Ser, Cys, or Ala. Double mutants such as βSer174Phe/βGly149 Ser showed essentially the same ATPase activity and proton transport as the wild type. As expected from these results, the double mutant F_1 generated the wild-type torque ($\sim$40 pN nm). The high-resolution structure of the bovine F_1 predicts that βSer174 is located on the β-subunit surface within the loop between an α-helix (helix B) and a β-sheet (β-sheet 4) (Fig. 4). βGly149, the first residue of the phosphate-binding P-loop connected to helix B, is

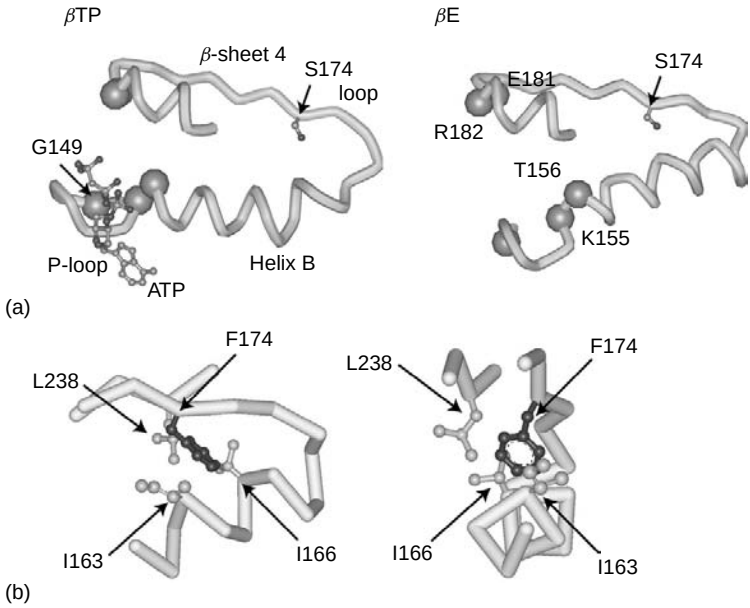

Fig. 4 Models of the *E. coli* β-subunit domain including βGly149 and βSer174. (a) Domain structure for the ATP-bound (β_{TP}) and empty (β_E)β-subunit. The bovine structure was used to model *E. coli* domains. Positions of ATP and residues discussed in the text are indicated. Nomenclatures for the α-helix and β-sheet are those of Abrahams et al. (1994). (b) Models of the βSer174Phe mutant domain structure. The positions of the residues discussed in the text are shown. Taken from Iko, Y., Sambongi, Y., Tanabe, M., Iwamoto-Kihara, A., Saito, K., Ueda, I., Wada, Y., Futai, M. (2001) ATP synthase F_1 sector rotation. Defective torque generation in the β subunit Ser-174 to Phe mutant and its suppression by second mutations, *J. Biol. Chem.* **276**, 47508–47511.

close to the catalytic site. The structure of the βGly149–βSer174 region is significantly different between the nucleotide-bound and empty β-subunits. Thus, it can be assumed that the conformational change of the catalytic site, followed by that of the βGly149–βSer174 domain, leads to the γ-subunit rotation for the energy coupling to proton transport. The βSer174Phe mutation was also suppressed by αArg296Cys of the α-subunit. It is of interest to analyze the rotation of the double mutant and related strains.

Energy minimization with a simple potential function of the modeled βSer174Phe mutant F_1 predicts that the side chain of βPhe174 interacts with that of βIle163 or βIle166 of the β-subunit with no nucleotide bound. We substituted βIle163 or βIle166 with a less bulky Ala residue in the βSer174Phe mutant. As expected, the F_1 with the βIle163Ala/βSer174Phe or βIle166Ala/βSer174Phe double mutant could rotate and generate almost similar torque to the wild type. These results suggest that the βGly149–βSer174 domain plays important roles in the rotation and torque generation essential for energy coupling.

The role of the DELSEED loop has been focused on because its conformation is significantly different among the three

β-subunits (β_E and β_T or β_D), consistent with the roles of γMet23 and βGlu381 (the second residue of the DELSEED loop) in energy coupling. However, mutagenesis studies involving replacement of residues in the DELSEED loop of thermophilic Bacillus $\alpha_3\beta_3\gamma$ did not reveal significant effects on the torque generation. The negative results may suggest that the loop is not related to the conformation change driving the β-subunit rotation. However, this is difficult to conclude from the directed mutations without structural analysis. Furthermore, the experimental system with actin filaments gives rotation rates and torque values with high deviations. Extensive substitution of related residues and their second mutations should be analyzed for the final conclusion, as discussed earlier, for a region around βSer174. Analysis of rotation in the F-ATPase holoenzyme may be necessary, as pointed out, for the γMet23Lys mutation.

4 Rotational Catalysis of the F-ATPase Holoenzyme

4.1 Structure of F_0Sector and Proton Transport Pathway

As discussed earlier, the F_0 membrane sector is composed of *a*-, *b*-, and *c*-subunits, whose stoichiometry ($1:2:10 \pm 1$, for a:b:c) was first determined from the stained bands on polyacrylamide gel electrophoresis. On the basis of structural prediction, Cox et al. proposed a model of F_0 in which the *a*- and *b*-subunit helices are surrounded by a ring of *c*-subunits. However, this model was not consistent with the electron and atomic force microscopic images. In the model derived from the images, the *a*- and *b*-subunits are attached to one side of the symmetric ring structure formed by the *c*-subunits. An atomic force microscope image indicated <12 *c*-subunits in the ring. The ab_2 complex was purified recently after solubilization of F_0 from membranes utilizing a histidine tag introduced at the *a*-subunit amino terminus. The proton pathway (F_0) was reconstituted from the ab_2 complex and *c*-subunit, confirming that ab_2 and the *c*-ring are two functional subcomplexes of the F_0 sector.

Subunit *a* spans the membrane with five helices, and its fourth helix could be cross-linked to the carboxyl-terminal helix of a *c*-subunit when Cys residues were introduced at appropriate positions, *a*Arg210 in the fourth helix is involved in proton transport, and all the substitutions, even *a*Arg210 to Lys, gave an F_0 sector with no ability to transport protons. The structure of the amino terminus (residues 1–34) of the *b*-subunit, including the transmembrane helix, was solved by NMR in a chloroform–methanol–water mixture. Extramembrane helices interact with the δ- and α-subunits at the top of the F_1 sector, which is consistent with the model proposing that the *b*-subunit closely interacts with an $\alpha_3\beta_3$ hexamer. The *b*-subunit dimer interaction with the F_1 sector is necessary for a second stalk. The carboxyl-terminal helix of a *c*-subunit can be cross-linked to the membrane helix of the *b*-subunit. Fillingame and coworkers solved the structure of the *E. coli c*-subunit by NMR: monomeric *c* (in a mixture of chloroform, methanol, and water, pH 5) gave a hairpin-like structure formed from the two helices connected by a polar region that interacts with the γ- and ε-subunits. Structural models of F_0

and the *c*-ring were reviewed recently by Fillingame and coworkers.

The front face of one *c*-subunit packs with the back face of a second *c*-subunit to form a dimer, consistent with the results of mutant studies. Fillingame and coworkers proposed that dimers form a ring of 12 monomers, with the amino and carboxyl α-helices of each *c*-subunit in the interior and at the periphery, respectively, and the ring was modeled from the results of molecular dynamic calculations. The model was supported by cross-linking analysis. However, a ring containing more than 10 monomers was not found in purified F_0F_1, and recent recalculation gave essentially the same, but a slightly smaller ring, formed from 10 copies of the *c*-subunit with the two helices in similar orientations. In an alternative model, the carboxyl- and amino-terminal α-helices form inner and outer rings, respectively. However, this model is not supported by cross-linking experiments. The X-ray structure indicated an yeast F_1 tightly bound with a *c*-ring formed from 10 monomers.

These results established that the *c*-subunits form a ring of 10 monomers. However, the number of monomers forming the ring is still controversial: atomic force microscopy of chloroplast and bacterial F_0 demonstrated rings of 14 and 11 *c*-subunits, respectively. The difference in the copy number may be due to the difference in species, loss of a part of the monomers, or reorganization of the ring during purification. Structural studies on F-ATPase or the F_0 sector will eventually explain the discrepancy.

*c*Asp61, in the middle of the second transmembrane α-helix, is responsible for proton transport, and close to *c*Ala24 and *c*Ile28, of which substitutions by other residues reduced the DCCD reactivity of *c*Asp61. The stoichiometry of the *a*- and *c*-subunits (1 : 10) indicates that one *a*Arg210 and multiple *c*Asp61 are required for proton translocation in F-ATPase. As the pK_a of the *c*Asp61 carboxyl moiety is 7.1, the NMR *c*-subunit structure at pH 5 is at the fully protonated stage. The interaction between *c*Asp61 and *a*Arg210 may lower the pK_a to facilitate proton release into the proton pathway. Rastogi and Girvin solved the *c*-subunit structure at pH 8, with *c*Asp61 being in an almost completely deprotonated form. The difference between the *c* structures at pH 5 and 8 is that the carboxyl-terminal α-helix was rotated by 140° with respect to the amino-terminal helix. From the two structures and the results of cross-linking experiments, Fillingame suggested that the carboxyl-terminal α-helix rotates during proton transport, interacts with *a*Arg210, and deprotonates *c*Asp61. This *c*-subunit structural change possibly drives stepwise rotation of the *c*-ring.

4.2 Rotational Catalysis of the F-ATPase Holoenzyme

To complete ATP hydrolysis-dependent proton transport, the γ-rotation should be transmitted to the F_0 membrane sector. Conversely, for ATP synthesis, proton transport through F_0 should generate torque to drive the γ-subunit rotation coupled to chemistry. The γ-rotation should be coupled to protonation–deprotonation of *c*Asp61 in either direction. Depending on the modeling of F_0 (*a*- and *b*-subunit inside the *c*-ring), rotation of *a*, *b*, γ, δ, and ε relative to the complex of α, β, and the *c*-ring once has been suggested. This speculation led to experimental tests, although an alternative model in which the *c*-subunit ring attached by the *a*- and

b-subunit helices is supported experimentally, as discussed above.

Different mechanisms could be proposed for the coupling between γ-rotation within the $\alpha_3\beta_3$ hexamer and proton transport through the F_0 sector: (1) the γ-subunit rotates on the surface of the *c*-ring; and (2) the γ-subunit and the *c*-ring rotate as a single complex. Models should be consistent with the proton transport continuously utilizing one *a*Arg210 and 10–14 *c*Asp61 residues of the *a*-subunit and the *c*-ring, respectively. Consistent with mechanism (2), energy transduction of F-ATPase with a counterrotating rotor ($\gamma\varepsilon c$ ring) and stator has been proposed. Convincing evidence supporting this mechanism includes the results of chemical cross-linking. Cross-linking between the γ- and *c*-subunits did not affect F-ATPase activity, indicating that sequential interaction of the rotating γ-subunits with multiple *c*-subunits is not necessary during ATP synthesis–hydrolysis. Similarly, cross-linking between the γ- and ε-subunits did not affect ATP hydrolysis, supporting the rotation of an actin filament connected to either subunit. However, $\alpha-\gamma$, $\alpha-\varepsilon$, $\beta-\gamma$, and $\beta-\varepsilon$ cross-linking resulted in loss of the activity, confirming that $\varepsilon\gamma$ rotates against $\alpha_3\beta_3$. These results suggest that a complex formed from γ-, ε-, and *c*-subunits is a mechanical unit for relative rotation to the $\alpha_3\beta_3$ hexamer.

We obtained direct evidence of continuous rotation of the *E. coli* $\varepsilon\gamma c_{10-14}$ complex during ATP hydrolysis (Fig. 5). F-ATPase engineered for rotation was solubilized from membranes, purified, and immobilized through histidine tags introduced into the α-subunit (Fig. 5a). Upon the addition of ATP, the filament connected to the *c*-ring rotated anticlockwise continuously, and generated similar torque to that observed for the γ-rotation in the F_1 sector. The rotation was inhibited by venturicidin, a specific inhibitor for F-ATPase but not for ATPase activity of the F_1 sector. The rotation and ATPase activity became less sensitive to the antibiotic when the *c*Ile38Thr mutation was introduced into the *c*-subunit, indicating that antibiotic binding to the *c*-ring inhibited rotation.

Pänke et al. also observed *c*-subunit rotation during ATP hydrolysis. They immobilized F_0F_1 through a histidine tag introduced into the β-subunit, and an actin filament was connected to the *c*-subunit through a streptag and streptavidin. The characteristics of the rotation observed were the same as those with the above systems. These results are consistent with the recent experiments showing that complete cross-linking of the γ-, ε-, and *c*-subunit had no effect on ATP hydrolysis, proton translocation, or ATP synthesis.

The important conclusion drawn from these experiments is that the *c*-subunit ring rotates when F-ATPase is immobilized through the α- or β-subunit. As discussed by Pänke et al. and Wada et al., it is not easy to prove that all the subunits were integrated into F-ATPase rotating under the microscope. However, indirect evidence supports the intactness of the enzyme used for rotation: reconstitution studies indicated that the *a*-, *b*-, and *c*-subunit are required to form F_0 capable of F_1 binding, and all F_1 subunits are required for binding to the F_0 sector. Early genetic and biochemical studies led to the conclusion that all three F_0 subunits are required for F_1 binding. The intactness of F-ATPase in these observations was also supported by the results for the membrane-bound enzyme, which rotated in essentially a similar manner to the purified F-ATPase.

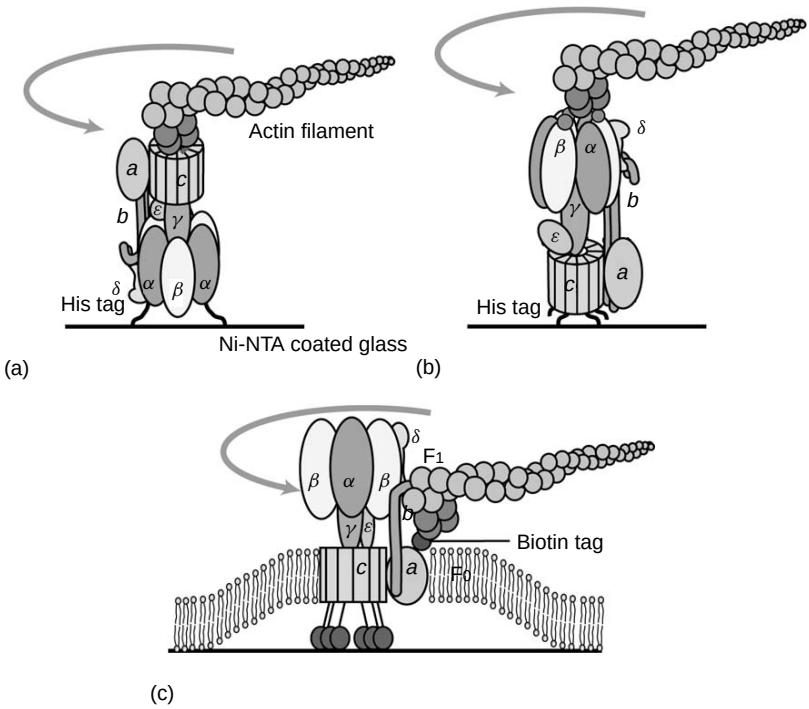

Fig. 5 Rotational catalysis of F-ATPase. F_0F_1 was immobilized through a histidine tag attached to the α-subunit (a) or *c*-ring (b, c), and a fluorescent actin filament was connected as a probe to the *c* (a), α (b), or *a* (c) subunit to observe rotational catalysis. Continuous rotation of the probe connected to the purified F_0F_1 (a or b) or membrane-embedded F_0F_1 (c) was observed upon addition of ATP.

We further addressed the basic question of whether the rotor and stator are interchangeable in F-ATPase (Fig. 5b). Thus, F-ATPase was immobilized on a glass surface through a histidine tag introduced into the *c*-subunit, and an actin filament was connected to the β-subunit through the biotin-binding domain of transcarboxylase (biotin tag). ATP-dependent filament rotation generated the same frictional torque as observed above, similar to the case of F-ATPase immobilized through the α- or β-subunit and with an actin filament connected to the *c*-subunit. Thus, either of the two subcomplexes ($\varepsilon\gamma c_{10-14}$ or $\alpha_3\beta_3\delta ab_2$) could be a rotor or a stator. Their actual roles *in vivo* will depend on the viscous drag due to the cytosol and membranes.

4.3 Rotational Catalysis of F-ATPase in Membranes

Although rotation of the $\gamma\varepsilon C_{10-14}$ complex has been shown using purified F-ATPase, a more favorable experimental system is the isolated membrane because the original integrity of the enzyme is maintained. The membrane could be used to answer one of the obvious questions of whether the *c*-ring rotates relative to the *a*-subunit. Such rotation would support the model of proton transport through the interface of the two subunits (Fig. 5c). It may be difficult to

test the rotation of a probe connected to the *c*-ring embedded in membranes by immobilizing F-ATPase through the α- or β-subunit. As described earlier, purified F-ATPase can be immobilized through a histidine tag connected to the *c*-ring, and an actin filament can be connected to the α- or β-subunit through a biotin tag. This experimental system was modified to test the rotation of F-ATPase in the membrane. As histidine and biotin tags face the periplasm and cytoplasm of the intact *E. coli* cell, respectively, membrane preparation for rotation should be carefully considered. Right-side out or everted membrane vesicles cannot be used for testing rotation, because only one of the two tags in these vesicles is accessible from the medium: the histidine tag faces the medium for right-side out vesicles, but the biotin tag is inside the same vesicles. Similarly, the histidine tag is inside everted vesicles, and thus cannot be used for immobilizing F-ATPase in these vesicles. Therefore, F-ATPase in planar membranes should be used to test for the rotation.

We prepared membrane fragments by passing *E. coli* cells through a French press by a slight modification of the procedure used to prepare everted vesicles. A labeling experiment with streptavidine-conjugated gold particles indicated that the preparation contained a significant number of planar membranes. Using this preparation, we observed ATP-dependent rotation of an actin filament connected to the β-subunit. As expected from the previous studies, the rotation was counterclockwise and sensitive to DCCD.

To complete a model of proton translocation through the rotation in F-ATPase, it is essential to show rotation of the *c*-ring relative to the *a*-subunit (Fig. 5c). The *c*-subunit (*c*Met65Cys) cross-linked with the *a*-subunit (αAsn214Cys) was highly protected from labeling with ^{14}C-DCCD. However, when F-ATPase was labeled with ^{14}C-DCCD and subjected to the ATPase reaction before cross-linking, there was significantly increased labeling of the *a*–*c* dimer. This experiment supports the rotational catalysis in F_0, although no

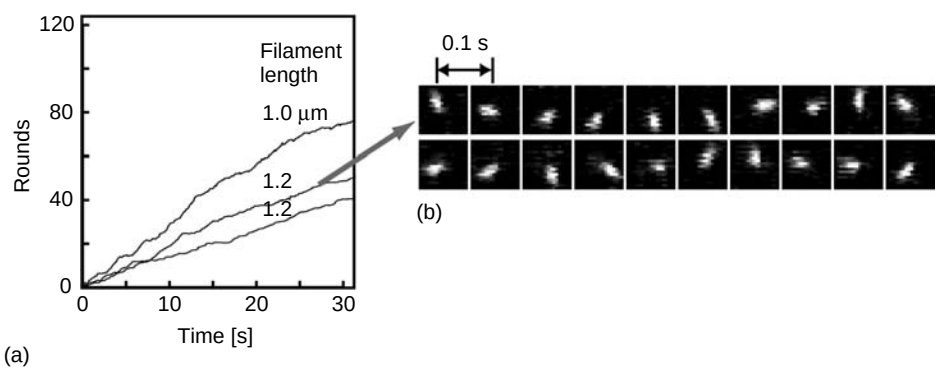

Fig. 6 Relative rotation of the *a*-subunit in membrane-embedded F-ATPase. Time courses of rotating filaments connected to the *a*-subunit of membrane F-ATPase immobilized through the *c*-subunit are shown, along with time courses of rotating filaments of varying length (a) and typical sequential video images (video interval, 100 ms) (b). Taken from Nishio, K., Iwamoto-Kihara, A., Yamamoto, A., Wada, Y., Futai, M. (2002) Subunit rotation of ATP synthase embedded in membranes: a or βsubunit rotation relative to the *c* subunit ring, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13448–13452.

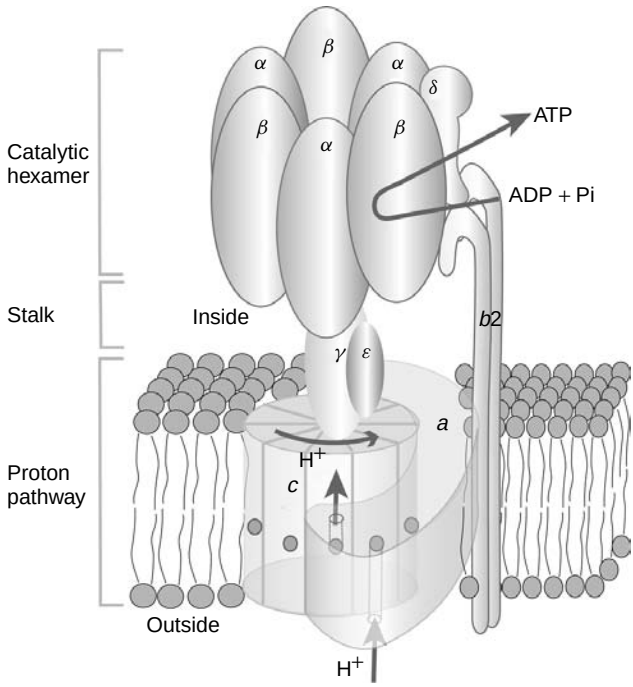

Fig. 7 Schematic mechanism of rotational catalysis by F-ATPase. For ATP synthesis, electrochemical proton transport changes the *c*-subunit conformation that drives the rotational movement of the *c*-ring together with the γ- and ε-subunits. Conversely, ATP hydrolysis drives the γ- and ε-subunit rotation together with the *c*—ring to transport protons. Modified from Wang, H., Oster, G. (1998) Energy transduction in the F_1 motor of ATP synthase, *Nature* **396**, 279–282; Fillingame, R.H., Angevine, C.M., Dmitriev. O.Y. (2002) Coupling proton movement to *c*-ring rotation in F_0F_1 ATP synthase: aqueous access channel and helix rotations at a–c interface, *Biochim. Biophys. Acta* **1555**, 29–36; Junge, W., Lill, H., Engelbrecht, S. (1997) ATP synthase: an electrochemical transducer with rotatory mechanics, *Trends Biochem. Sci.* **22**, 420–423.

information on the mechanism could be obtained. We could show rotation of the actin filament connected to the *a*-subunit relative to the *c*-ring using membrane fragments (Fig. 6).

These results and those obtained with purified F_0F_1 or the F_1 sector indicate that $\gamma\varepsilon c_{10-14}$ and $\alpha_3\beta_3\delta ab_2$ are mechanical units, that is, an interchangeable rotor and stator, respectively. Notably, the results obtained with membrane-embedded F-ATPase and those with the purified one are similar. Furthermore, the experimental system will be extremely valuable for studying the mechanism of coupling of rotation and proton transport, although further modification(s) may be necessary.

These studies established rotational catalysis by F-ATPase (Fig. 7). For ATP hydrolysis, the F-ATPase is a chemically driven motor rotating $\gamma\varepsilon c_{10-14}$ to drive proton transport. Studies on the

mechanism of the rotation in the F_0 sector were started only recently. The rotation is inhibited by venturicidin or DCCD, suggesting that the tight or covalent binding of these bulky chemicals to the *c*-subunit inhibits mechanical rotation. Junge and coworkers showed that the *c*-subunit ring with the *c*Asp61Asn mutation can still rotate during ATP hydrolysis, indicating that the proton transport is not obligatory for the chemically driven rotation (ATP hydrolysis-dependent *c*-ring rotation). For ATP synthesis, the same enzyme is a potential-driven motor rotating $\gamma\varepsilon c_{10-14}$ through an electrochemical proton gradient. This membrane system will be useful for studying the rotation in this direction.

5 Rotational Catalysis of V-ATPASE

5.1 Catalytic Site and Proton Pathway

V-ATPase (vacuolar-type ATPase) acidifies the lumens of endomembrane organelles such as lysosomes, endosomes, and synaptic vesicles. The same enzyme in the plasma membranes of specialized cells, including osteoclasts and renal intercalated cells, pumps protons into extracellular compartments such as resorption lacunae and collecting ducts, respectively. Despite significant physiological differences, V-ATPase exhibits similarities with F-ATPase. The catalytic *A*-subunit of V-ATPase is homologous to F-ATPase β. The homology is striking in the phosphate-binding P-loop and other sequences, including the F-ATPase catalytic residues discussed above. *E. coli* F-ATPase residues at the catalytic site, βLys155, βThr156, βGlu181, βArg182, and βGlu185, correspond to *A*Lys263, *A*Thr264, *A*Glu266, *A*Arg267, and *A*Glu290 of the yeast V-ATPase *A*-subunit, respectively (Fig. 2). Mutational studies of yeast V-ATPase supported the notion that they are catalytic residues. The kinetics indicate that V-ATPase has three catalytic sites exhibiting cooperativity. The cysteine residue in the V-ATPase P-loop may be involved in regulation. F-ATPase with Cys introduced at the corresponding position became sensitive to sulfhydryl reagents, similar to V-ATPase.

The membrane V_0 sector of V-ATPase has a more complicated subunit composition: yeast V_0 is formed from *c*-, *c*′-, *c*″-, *a*-, and *d*-subunits. Of the three proteolipid subunits, the *c*- and *c*′- having four transmembrane helices are a duplicated form of the F-ATPase *c*-subunit, and show 56% identity in amino acid residues with each other. The carboxyl moieties (*c*Glu137 and *c*′Glu145) for proton transport are located on the fourth helix of the *c*- and *c*′-subunits, respectively. The *c*″-subunit, exhibiting some homology with *c* and *c*′, is also required in proton transport: *c*″Glu188 in the middle of the third α-helix is also implicated for proton transport. The *c*″-subunit is conserved in mammalians, whereas the *c*′-subunit is only found in yeast. The stoichiometry of the $c : c' : c''$ subunits in yeast is n:1 : 1. The three proteolipids may form a ring structure similar to the F_0c-ring. Assuming that 4, 1, and 1 molecules of the *c*, *c*′ and *c*″ subunits, respectively, form a ring, V_0 has ~6 proton–translocating residues altogether. This is in contrast with 10–14 residues for F_0. Mutation of V-ATPase *a*Arg735 abolished proton translocation similar to that of F-ATPase *a*Arg210, although the subunit *a* of V_0 and F_0 exhibit little homology. Thus, V_0 has multiple carboxyl moieties and one arginine essential for proton translocation.

The similarities between the two ATPases raised the interesting question of whether V-ATPase can synthesize ATP. Unlike F-ATPase localized with the respiratory chain in the mitochondrial membrane, mammalian or yeast membranes containing V-ATPase do not have a system for generating an electrochemical proton gradient. Does V-ATPase synthesize ATP when a membrane potential or proton gradient is generated? To answer this question, we expressed a plant proton–translocating pyrophosphatase in yeast vacuoles. Upon hydrolysis of pyrophosphate, the vacuolar membrane vesicles could form an electrochemical proton gradient, which was lowered by the addition of ADP + Pi. As expected from this result, we observed ATP synthesis sensitive to the V-ATPase specific inhibitor bafilomycin. The results further support the similarities between the two ATPases. Plant vacuoles having both V-ATPase and pyrophosphatase may synthesize ATP similar to the yeast chimeric system.

5.2 Subunit Rotation of V-ATPase During Catalysis

Despite the similarities of V- and F-ATPase discussed earlier, they are also significantly different. Typically, F-ATPase of *E. coli* has eight subunits, whereas yeast V-ATPase has 13. Most of the subunits unique to V-ATPase are located in the stalk region. Furthermore, the presence of isoforms of subunits *B*, *E*, *G*, *d*, and *a* has been found for V-ATPase, whereas the information on isoforms is limited for F-ATPase. As pointed out above, V_0, possibly its ring structure, may be different from F_0. These differences and similarities led us to examine the rotational catalysis in V-ATPase.

As described above, rotation of an actin filament connected to the *a*-, α-, or β-subunit was observed when F-ATPase was immobilized through the *c*-subunit ring. The correspondence of minor subunits in the stalk region between the two ATPases was difficult to determine (Fig. 8).

```
Yeast G    1 -------------------------------------------------------MSQKN     5
E.coli b   1 VNLNATILGQAIAFVLFVLFCMKYVWPPLMAAIEKRQKEIADGLASAERAHKDLDLAKAS    60

           6 GIATLLQAEKEAHEIVSKARKYRQDKLKQAKTDAAKEIDSYKIQKDKELKEFEQKNAGGV    65
          61 ATDQLKKAKAEAQVIIEQANKRRSQILDEAKAEAEQERTKIVAQAQAEIEA-ERKRARE-   118

          66 GELEKKAEAG-VQGELAEIKKIAEKKKD-DVVKILIETVIKPSAEVHINAL            114
         119 -ELRKQVAILAVAGAEKIIERSVDEAANSDIVDKLVAEL------------            156

Yeast G    1 -------------------------------------MSQKNGIATLLQAEKEAHEIVSK    23
E.coli δ   1 MSEFITVARPYAKAAFDFAVEHQSVERWQDMLAFAAEVTKNEQMAELLSGALAPETLAES    60

          24 ARKYRQDKLKQAKTDAAKEIDSYK-IQKDKELK-EFEQKNAGGVG--ELEKKAEAGVQGE    79
          61 FIAVCGEQLDENGQNLIRVMAENGRLNALPDVLEQFIHLRAVSEATAEVDVISAAALSEQ   120

          80 -LAEIKKIAEKKKDDVVKIL--IETVIKPSAEVHINAL-------------------      114
         121 QLAKISAAMEKRLSRKVKLNCKIDKSVMAGVIIRAGDMVIDGSVRGRLERLADVLQS      177
```

Fig. 8 Similarities of the V-ATPase *G*-subunit with the F-ATPase *b*- or δ-subunit. The subunits are aligned to obtain maximal homology. Identical amino acid residues are boxed.

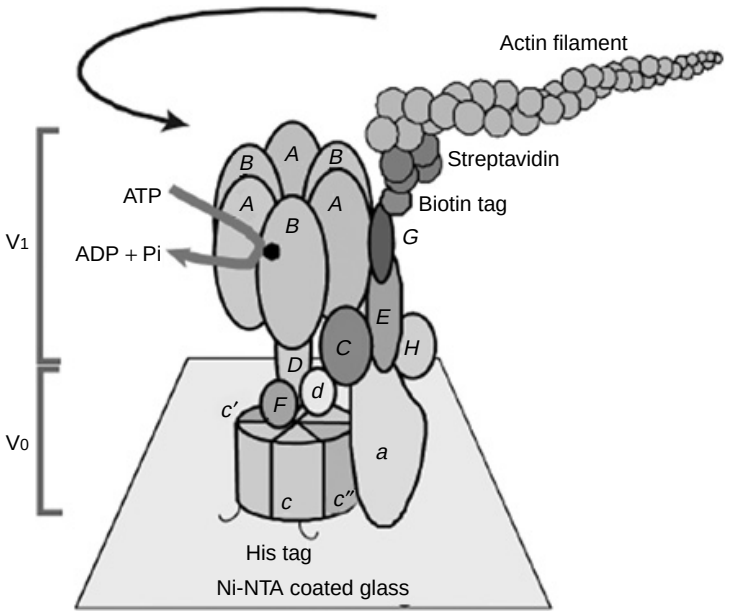

Fig. 9 Rotation of an actin filament connected to the *G*-subunit of V-ATPase. Solubilized V-ATPase was immobilized on a glass surface through the *c*-subunit, and an actin filament was connected to the *G*-subunit. Rotation was observed upon the addition of ATP, and inhibited by nitrate and concanamycin. Taken from Hirata, T., Iwamoto-Kihara, A., Sun-Wada, G.-H., Okajima, T., Wada, Y., Futai, M. (2003) Subunit rotation of vacuolar-type proton pumping ATPase: relative rotation of the G and C subunits, *J. Biol. Chem.* **278**, 23741–23719.

The *G*-subunit was suggested to exist in the stalk domain of a recent model, and to be accessible from the cytosol. The *G*-subunit and F-ATPase *b* exhibit ~24% homology, although the *G*-subunit lacks a transmembrane region. However, recent cross-linking studies suggested that the *G*-subunit is located near the top of V_1, similar to the F-ATPase δ-subunit. Thus, the *G*-subunit may correspond to the *b* or δ-subunit of F_1, and could be a candidate for connecting a probe to observe rotation, although the homology is quite limited.

On the basis of these considerations, we engineered the yeast chromosome (*VMA3* and *VMA10* for the *c*- and *G*-subunit, respectively), solubilized V-ATPase from vacuolar membranes, and tested rotational catalysis. An actin filament connected to the *G*-subunit rotated upon ATP hydrolysis (Fig. 9). The rotation was inhibited by nitrate and concanamycin, similar to their effects on ATPase activity. Concanamycin inhibition was striking: the rotation terminated within a few seconds after the addition, which is consistent with its tight affinity. This antibiotic is similar to bafilomycin, which binds to the V_0 sector, possibly to its *c*-subunit, indicating that concanamycin, possibly bound to the rotor–stator interface, terminated the rotation. Torque generated by the rotation was similar to that with F-ATPase.

These results suggest that the A_3B_3 hexamer rotates together with the *G*-subunit when the *c*-subunit ring is immobilized. An interesting question is "Which part of the V-ATPase rotates *in vivo*?" In this regard, Holiday and coworkers reported that the *B*-subunit interacts with the cytoskeleton, possibly actin. Thus, the *c*-subunit ring together with the subunit corresponding to the γ- (possibly D) subunit may be rotating *in vivo* if the V-ATPase holo enzyme is immobilized with the cytoskeleton. However, it is also possible that the rotor–stator is determined only by the slight difference in viscous drag applied to each subcomplex.

The rotational catalysis of the peripheral membrane sector of *Thermus thermophilus* ATPase was shown recently. This Archaean ATPase is more similar to F-ATPase than V-ATPase, as pointed out by the authors. In this regard, the ATPase of Archae (Archaean bacterial) plasma membranes has been classified as an A-type ATPase. Thus, three related proton–translocating ATPases carry out common rotational catalysis.

6 Epilogue

Proton-translocating F-ATPase (ATP synthase) has been a fascinating membrane enzyme for generations of biochemists. As Paul Boyer called it *A splendid molecular machine* it is more than an ordinary enzyme. F-ATPase synthesizes or hydrolyzes ATP coupling between proton translocation and chemistry, and has become a real molecular machine in this sense, since mechanical rotation of its subunit complex was included recently in its mechanism. For the engineering aspect of the mechanism, the F_1 sector immobilized on a metal surface can rotate a metal plate.

As described above, experimental systems have been established for further studies. Details of the rotation mechanism will be revealed by studies involving *E. coli* F-ATPase together with the progress regarding physical methods and the higher-ordered X-ray structure determinations of the F_0 sector.

It was interesting to learn that rotational catalysis has been expanded to V-ATPase and also distantly related Archaean ATPase. The basic mechanism of V-ATPase rotation may be similar to that of F-ATPase, as expected from the structural similarities between the two enzymes. Thus, the rotational mechanism of *E. coli* F-ATPase should be studied extensively, using its genetic and biochemical advantages. V-ATPase may have a more fascinating regulatory mechanism that is supported by a unique V_0 structure and a series of isoforms in the stalk region. The regulatory role of stalk subunit(s) including subunit *E* in energy coupling has already been discussed.

See also Biological Regulation by Protein Phosphorylation; Metabolic Basis of Cellular Energy; Protein and Nucleic Acid Enzymes; Protein Modeling.

Bibliography

Books and Reviews

Boyer, P.D. (1997) The ATP synthase – a splendid molecular machine, *Annu. Rev. Biochem.* **66**, 717–749.

Fillingame, R.H. (1996) Membrane sectors of F- and V-type H^+-transporting ATPases, *Curr. Opin. Struct. Biol.* **6**, 491–498.

Fillingame, R.H., Dmitriev, O.Y. (2002) Structural model of the transmembrane Fo

rotary sector of H^+-transporting ATP synthase derived by solution NMR and intersubunit cross-linking in situ, *Biochim. Biophys. Acta* **1565**, 232–245.

Fillingame, R.H., Angevine, C.M., Dmitriev, O.Y. (2002) Coupling proton movements to c-ring rotation in F_1F_0ATP synthase: aqueous access channels and helix rotations at the a-c interface, *Biochim. Biophys. Acta* **1555**, 29–36.

Futai, M., Noumi, T., Maeda, M. (1989) ATP synthase (H^+-ATPase): results by combined biochemical and molecular biological approaches, *Annu. Rev. Biochem.* **58**, 111–136.

Futai, M., Wada, Y., Kaplan, J. (2004) *Handbook of ATPases*, Wiley-VCH Verlag Gmbh & Co., KGaA, Weinheim.

Nelson, N., Harvey, W.R. (1999) Vacuolar and plasma membrane proton-adenosinetriphosphatases, *Physiol. Rev.* **79**, 361–385.

Nishi, T., Forgac, M. (2002) The vacuolar (H^+)-ATPases – nature's most versatile proton pumps, *Nat. Rev. Mol. Cell Biol.* **3**, 94–103.

Panefsky, H.S., Cross, R.L. (1991) Structure and mechanism of F_0F_1-type ATP synthases and ATPases, *Adv. Enzymol. Relat. Areas Mol. Biol.* **64**, 173–214.

Racker, E. (1976) *A New Look at Mechanisms in Bioenergetics*, Academic Press, New York.

Senior, A.E., Nadanaciva, S., Weber, J. (2002) The molecular mechanism of ATP synthesis by F_1F_0-ATP synthase, *Biochim. Biophys. Acta* **1553**, 188–211.

Stock, D., Gibbons, C., Arechaga, I., Leslie, A.G.W., Walker, J.E. (2000) The rotary mechanism of ATP synthase, *Curr. Opin. Struct. Biol.* **10**, 672–679.

Weber, J., Senior, A.E. (1997) Catalytic mechanism of F_1-ATPase, *Biochim. Biophys. Acta* **1319**, 19–58.

Primary Literature

Abrahams, J.P., Leslie, A.G.W., Lutter, R., Walker, J.E. (1994) Structure at 2.8 Å resolution of F_1-ATPase from bovine heart mitochondria, *Nature* **370**, 621–628.

Abrahams, J.P., Buchanan, S.K., Raaji Van, M.J., Fearnley, I.M., Leslie, A.G.W., Walker, J.E. (1996) The structure of bovine F_1-ATPase complexed with the peptide antibiotic efrapeptin, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9420–9424.

Aggeler, R., Haughton, M.A., Capaldi, R.A. (1995) Disulfide bond formation between the COOH-terminal domain of the β subunits and the γ and ε subunits of the *Escherichia coli* F_1-ATPase. Structural implications and functional consequences. *J. Biol. Chem.* **270**, 9185–9191.

Al-Shawi, M.K., Ketchum, C.J., Nakamoto, R.K. (1997) Energy coupling, turnover, and stability of the F_0F_1 ATP synthase are dependent on the energy of interaction between γ and β subunits, *J. Biol. Chem.* **272**, 2300–2306.

Arata, Y., Baleja, J.D., Forgac, M. (2002a) Localization of subunits D, E, and G in the yeast V-ATPase complex using cysteine-mediated cross-linking to subunit B, *Biochemistry* **41**, 11301–11307.

Arata, Y., Baleja, J.D., Forgac, M. (2002b) Cysteine-directed cross-linking to subunit B suggests that subunit E forms part of the peripheral stalk of the vacuolar H^+-ATPase, *J. Biol. Chem.* **277**, 3357–3363.

Bianchet, M.A., Hullihen, J., Pederson, P.L., Amzel, L.M. (1998) The 2.8-Å structure of rat liver F_1-ATPase: configuration of a critical intermediate in ATP synthesis/hydrolysis, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11065–11070.

Birkenhäger, R., Hoppert, M., Deckers-Hebestreit, G., Mayer, F., Altendorf, K. (1995) The F_0 complex of the *Escherichia coli* ATP synthase. Investigation by electron spectroscopic imaging and immunoelectron microscopy, *Eur. J. Biochem.* **230**, 58–67.

Braig, K., Menz, R.I., Montgomery, M.G., Leslie, A.G.W., Walker, J.E. (2000) Structure of bovine mitochondrial F_1-ATPase inhibited by Mg^{2+} ADP and aluminium fluoride, *Structure* **8**, 567–573.

Cox, G.B., Fimmel, A.L., Gibson, F., Hatch, L. (1986) The mechanism of ATP synthase: a reassessment of the functions of the b and a subunits, *Biochim. Biophys. Acta* **849**, 62–69.

Dancan, T.M., Bulygin, V.V., Zou, Y., Hutcheon, M.L., Cross, R.L. (1995) Rotation of subunits during catalysis by *Escherichia coli* F_1-ATPase, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10964–10968.

Dmitriev, O.Y., Jones, P.C., Fillingame, R.H. (1999) Structure of the subunit c oligomer in the F_1F_0 ATP synthase: model derived from

solution structure of the monomer and cross-linking in the native enzyme, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7785–7790.

Dmitriev, O.Y., Jones, P.C., Jiang, W.P., Fillingame, R.H. (1999) Structure of the membrane domain of subunit b of the *Escherichia coli* F_0F_1 ATP synthase, *J. Biol. Chem.* **274**, 15598–15604.

Dunn, S.D. (1982) The isolated γ subunits of *Escherichia coli* F_1 ATPase binds the ε subunit, *J. Biol. Chem.* **257**, 7354–7359.

Dunn, S.D., Futai, M. (1980) Reconstitution of a functional coupling factor from the isolated subunits of *Escherichia coli* F_1 ATPase, *J. Biol. Chem.* **255**, 113–118.

Elston, T., Wang, H., Oster, G. (1998) Energy transduction in ATP synthase, *Nature* **391**, 510–513.

Eya, S., Maeda, M., Futai, M. (1991) Role of the carboxyl terminal region of H^+-ATPase F_0F_1 a subunit from *Escherichia coli*, *Arch. Biochem. Biophys.* **284**, 71–77.

Fillingame, R.H., Jiang, W.P., Dmitriev, O.Y. (2000) The oligomeric subunit crotor in the F_0 sector of ATP synthase: unresolved questions in our understanding of function, *J. Bioenerg. Biomembr.* **32**, 433–439.

Foster, D.L., Fillingame, R.H. (1982) Stoichiometry of subunits in the H^+-ATPase complex of *Escherichia coli*, *J. Biol. Chem.* **257**, 2009–2015.

Futai, M. (1977) Reconstitution of ATPase activity from the isolated α, β, and γ subunits of the coupling factor, F_1, of *Escherichia coli*, *Biochem. Biophys. Res. Commun.* **79**, 1231–1237.

Futai, M., Omote, H., Sambongi, Y., Wada, Y. (2000) Synthase (H^+ ATPase): coupling between catalysis, mechanical work, and proton translocation, *Biochim. Biophys. Acta* **1458**, 276–288.

Futai, M., Oka, T., Sun-Wada, G.-H., Moriyama, Y., Kanazawa, H., Wada, Y. (2000) Luminal acidification of diverse organelles by V-ATPase in animal cells, *J. Exp. Biol.* **203**, 107–116.

Garcia, J.J., Capaldi, R.A. (1998) Unisite catalysis without rotation of the γ-ε domain in *Escherichia coli* F_1-ATPase, *J. Biol. Chem.* **273**, 15940–15945.

Gibbons, C., Montgomery, M.G., Leslie, A.G.W., Walker, J.E. (2000) The structure of the central stalk in bovine F_1-ATPase at 2.4 Å resolution, *Nat. Struct. Biol.* **7**, 1055–1061.

Girvin, M.E., Rastogi, V.K., Abildgaard, F., Markley, J.L., Fillingame, R.H. (1998) Solution structure of the transmembrane H^+-transporting subunit c of the F_1F_0 ATP synthase, *Biochemistry* **37**, 8817–8824.

Groth, G., Walker, J.E. (1997) Model of the c-subunit oligomer in the membrane domain of F-ATPases, *FEBS Lett.* **410**, 117–123.

Groth, G., Pohl, E. (2001) The structure of the chloroplast F_1-ATPase at 3.2 Å resolution, *J. Biol. Chem.* **276**, 1345–1352.

Gumbiowski, K., Pänke, O., Junge, W., Engelbrecht, S. (2002) Rotation of the c subunit oligomer in EF_0EF_1 mutant cD61N, *J. Biol. Chem.* **277**, 31287–31290.

Hanada, H., Moriyama, Y., Maeda, M., Futai, M. (1990) Kinetic studies of chromaffin granule H^+-ATPase and effects of bafilomycin A1, *Biochem. Biophys. Res. Commun.* **170**, 873–878.

Hara, K.Y., Noji, H., Bald, D., Yasuda, R., Kinoshita, K. Jr., Yoshida, M. (2000) The role of the DELSEED motif of the β subunit in rotation of F_1-ATPase, *J. Biol. Chem.* **275**, 14260–14263.

Hausrath, A.C., Gruber, G., Matthews, B.W., Capaldi, R.A. (1999) Structural features of the γ subunit of the *Escherichia coli* F_1 ATPase revealed by a 4.4-Å resolution map obtained by X-ray crystallography, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13697–13702.

Hirata, R., Graham, L.A., Takatsuki, A., Stevens, T.H., Anraku, Y. (1997) VMA11 and VMA16 encode second and third proteolipid subunits of the *Saccharomyces cerevisiae* vacuolar membrane H^+-ATPase, *J. Biol. Chem.* **272**, 4975–4803.

Hirata, H., Nakamura, N., Omote, H., Wada, Y., Futai, M. (2000) Regulation and reversibility of vacuolar H^+-ATPase, *J. Biol. Chem.* **275**, 386–389.

Hirata, T., Iwamoto-Kihara, A., Sun-Wada, G.-H., Okajima, T., Wada, Y., Futai, M. (2003) Subunit rotation of vacuolar-type proton pumping ATPase: relative rotation of the G and *c* subunits, *J. Biol. Chem.* **278**, 23741–23719.

Holiday, L.S., Lu, M., Lee, B.S., Nelson, R.D., Solivan, S., Zhang, L., Gluck, S.L. (2000) The amino-terminal domain of the B subunit of vacuolar H^+-ATPase contains a filamentous actin binding site, *J. Biol. Chem.* **275**, 32331–32337.

Huss, M., Ingenhorst, G., Konig, S., Gassel, M., Dorse, S., Zeeck, A., Altendorf, K., Wieczorek, H. (2002) Concanamycin A, the specific inhibitor of V-ATPases, binds to the V_0 subunit c, *J. Biol. Chem.* **277**, 40544–40548.

Hutcheon, M.L., Duncan, T.M., Ngai, H., Cross, R.L. (2001) Energy-driven subunit rotation at the interface between subunit a and the c oligomer in the F_0 sector of *Escherichia coli* ATP synthase, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8519–8524.

Ida, K., Noumi, T., Maeda, M., Fukui, T., Futai, M. (1991) Catalytic site of F_1-ATPase of *Escherichia coli* Lys-155 and Lys-201 of the β subunit are located near the γ-phosphate group of ATP in the presence of Mg^{2+}, *J. Biol. Chem.* **266**, 5424–5429.

Ihara, K., Abe, T., Sugimura, K.I., Mukohata, Y. (1992) Halobacterial A-ATP synthase in relation to V-ATPase, *J. Exp. Biol.* **172**, 475–485.

Iko, Y., Sambongi, Y., Tanabe, M., Iwamoto-Kihara, A., Saito, K., Ueda, I., Wada, Y., Futai, M. (2001) ATP synthase F_1 sector rotation. Defective torque generation in the β subunit Ser-174 to Phe mutant and its suppression by second mutations, *J. Biol. Chem.* **276**, 47508–47511.

Imamura, H., Nakano, M., Noji, H., Mureyuki, E., Ohkuma, S., Yoshida, M., Yokoyama, K. (2003) Evidence for rotation of V_1-ATPase, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2312–2315.

Iwamoto, A., Miki, J., Maeda, M., Futai, M. (1990) H^+-ATPase γ subunit of *Escherichia coli*. Role of the conserved carboxyl-terminal region, *J. Biol. Chem.* **265**, 5043–5048.

Iwamoto, A., Omote, H., Hanada, H., Tomioka, N., Itai, A., Maeda, M., Futai, M. (1991) Mutations in Ser174 and the glycine-rich sequence (Gly149, Gly150, and Thr156) in the β subunit of *Escherichia coli* H^+-ATPase. *J. Biol. Chem.* **266**, 16350–16355.

Jeanteur-De Beukelaer, C., Omote, H., Iwamoto-Kihara, A., Maeda, M., Futai, M. (1995) β-γ subunit interaction is required for catalysis by H^+-ATPase (ATP synthase). β subunit amino acid replacements suppress a γ subunit mutation having a long unrelated carboxyl terminus, *J. Biol. Chem.* **270**, 22850–22854.

Jiang, W.P., Hermolin, J., Fillingame, R.H. (2001) The preferred stoichiometry of c subunits in the rotary motor sector of *Escherichia coli* ATP synthase is 10, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4966–4971.

Jones, P.C., Fillingame, R.H. (1998) Genetic fusions of subunit c in the F_0 sector of H^+-transporting ATP synthase. Functional dimers and trimers and determination of stoichiometry by cross-linking analysis, *J. Biol. Chem.* **273**, 29701–29705.

Jones, P.C., Hermolin, J., Jiang, W.P., Fillingame, R.H. (2000) Insights into the rotary catalytic mechanism of F_0F_1 ATP synthase from the cross-linking of subunits b and c in the *Escherichia coli* enzyme, *J. Biol. Chem.* **275**, 31340–31346.

Junge, W., Lill, H., Engelbrecht, S. (1997) ATP synthase: an electrochemical transducer with rotatory mechanics, *Trends Biochem. Sci.* **22**, 420–423.

Kato-Yamada, K., Noji, H., Yasuda, R., Kinoshita, K. Jr., Yoshida, M. (1998) Direct observation of the rotation of epsilon subunit in F_1-ATPase, *J. Biol. Chem.* **273**, 19375–19377.

Kawasaki-Nishi, S., Nishi, T., Forgac, M. (2001) Arg-735 of the 100-kDa subunit a of the yeast V-ATPase is essential for proton translocation, *Proc. Natl. Acad Sci. U.S.A.* **98**, 12397–12402.

Ketchum, C.J., Al-Shawi, M.K., Nakamoto, K.K. (1998) Intergenic suppression of the γ M23K uncoupling mutation in F_0F_1 ATP synthase by β Glu-381 substitutions: the role of the β 380DELSEED386 segment in energy coupling. *Biochem. J.* **330**, 707–712.

Le, N.P., Omote, H., Wada, Y., Al-Shawi, M.K., Nakamoto, R.K., Futai, M. (2000) *Escherichia coli* ATP synthase alpha subunit Arg-376: the catalytic site arginine does not participate in the hydrolysis/synthesis reaction but is required for promotion to the steady state, *Biochemistry* **39**, 2778–2783.

Liu, Q., Kane, P.M., Newman, P.R., Forgac, M. (1996) Site-directed mutagenesis of the yeast V-ATPase B subunit (Vma2p), *J. Biol. Chem.* **271**, 2018–2022.

Liu, Q., Leng, X.H., Newman, P.R., Vasilyeva, E., Kane, P.M., Forgac, M. (1997) Site-directed mutagenesis of the yeast V-ATPase A subunit, *J. Biol. Chem.* **272**, 11750–11756.

Löbau, S., Weber, J., Wilke-Mounts, S., Senior, A.E. (1997) F_1-ATPase, roles of three catalytic site residues, *J. Biol. Chem.* **272**, 3648–3456.

Long, J.C., Wang, S., Vik, S.B. (1998) Membrane topology of subunit a of the F_1F_0 ATP synthase as determined by labeling of unique cysteine residues, *J. Biol. Chem.* **273**, 16235–16240.

MacLeod, K.J., Vasilyeva, E., Baleja, J.D., Forgac, M. (1998) Mutational analysis of the nucleotide binding sites of the yeast vacuolar proton-translocating ATPase, *J. Biol. Chem.* **273**, 150–156.

Mc Lachlin, D.T., Dunn, S.D. (2000) Disulfide linkage of the b and Δ subunits does not affect the function of the *Escherichia coli* ATP synthase *Biochemistry* **39**, 3486–3490.

Menz, R.I., Walker, J.E., Leslie, A.G.W. (2001) Structure of bovine mitochondrial F_1-ATPase with nucleotide bound to all three catalytic sites: implications for the mechanism of rotary catalysis, *Cell* **106**, 331–341.

Nadanaciva, S., Webe, J., Senior, A.E. (1999) The role of β-Arg-182, an essential catalytic site residue in *Escherichia coli* F_1-ATPase, *Biochemistry* **38**, 7670–7677.

Nadanacivam, S., Weber, J., Wilke-Mounts, S., Senior, A.E. (1999) Importance of F_1-ATPase residue αArg-376 for catalytic transition state stabilization, *Biochemistry* **38**, 15493–15499.

Nakamoto, R.K., Maeda, M., Futai, M. (1993) The γ subunit of the *Escherichia coli* ATP synthase. Mutations in the carboxyl-terminal region restore energy couplin γ to the amino-terminal mutant γ Met-23–> Lys, *J. Biol. Chem.* **268**, 867–872.

Nakamoto, R.K., Al-Shawi, M.K., Futai, M. (1995) The ATP synthase γ subunit. Suppressor mutagenesis reveals three helical regions involved in energy coupling, *J. Biol. Chem.* **270**, 14042–14046.

Nakamoto, R.K., Shin, K., Iwamoto, A., Omote, H., Maeda, M., Futai, M. (1992) *Escherichia coli* F_0F_1-ATPase. Residues involved in catalysis and coupling, *Ann. N. Y. Acad. Sci.* **671**, 335–344.

Nishio, K., Iwamoto-Kihara, A., Yamamoto, A., Wada, Y., Futai, M. (2002) Subunit rotation of ATP synthase embedded in membranes: a or βsubunit rotation relative to the c subunit ring, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13448–13452.

Noji, H., Yasuda, R., Yoshida, H., Kinoshita, K. Jr. (1997) Direct observation of the rotation of F_1-ATPase, *Nature* **386**, 299–302.

Omote, H., Maeda, M., Futai, M. (1992) Effects of mutations of conserved Lys-155 and Thr-156 residues in the phosphate-binding glycine-rich sequence of the F_1-ATPase β subunit of *Escherichia coli*, *J. Biol. Chem.* **267**, 20571–20576.

Omote, H., Le, N.P., Park, M.-Y., Maeda, M., Futai, M. (1995) β subunit Glu-185 of *Escherichia coli* H^+-ATPase (ATP synthase) is an essential residue for cooperative catalysis, *J. Biol. Chem.* **270**, 25656–25660.

Omote, H., Tainaka, K., Iwamoto-Kihara, A., Wada, Y., Futai, M. (1998) Stability of the *Escherichia coli* ATP synthase F_0F_1 complex is dependent on interactions between γGln-269 and the β subunit loop β Asp-301-β Asp-305, *Arch. Biochem. Biophys.* **308**, 277–282.

Omote, H., Sambonmatsu, N., Saito, K., Sambongi, Y., Iwamoto-Kihara, A., Yanagida, T., Wada, Y., Futai, M. (1999) The γ-subunit rotation and torque generation in F_1-ATPase from wild-type or uncoupled mutant *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7780–7784.

Orriss, G.L., Leslie, A.G.W., Braig, K., Walker, J.E. (1998) Bovine F_1-ATPase covalently inhibited with 4-chloro-7-nitrobenzofurazan: the structure provides further support for a rotary catalytic mechanism, *Structure* **6**, 831–837.

Pänke, O., Gumbiowski, K., Junge, W., Engelbrecht, S. (2000) F-ATPase: specific observation of the rotating c subunit oligomer of EF_0EF_1, *FEBS Lett.* **472**, 34–38.

Park, M.Y., Omote, H., Maeda, M., Futai, M. (1994) Conserved Glu-181 and Arg-182 residues of *Escherichia coli* H^{+-} ATPase (ATP synthase) β subunit are essential for catalysis: properties of 33 mutants between β Glu-161 and β Lys-201 residues, *J. Biochem.* **116**, 1139–1145.

Powell, B., Graham, L.A., Stevens, T.H. (2000) Molecular characterization of the yeast vacuolar H^+-ATPase proton pore, *J. Biol. Chem.* **275**, 26354–23660.

Rastogi, V.K., Girvin, M.E. (1999) Structural changes linked to proton translocation by subunit c of the ATP synthase, *Nature* **402**, 263–268.

Rodgers, A.J.W., Capaldi, R.A. (1998) The second stalk composed of the β- and Δ-subunits connects F_0 to F_1 via an α-subunit in the *Escherichia coli* ATP synthase, *J. Biol. Chem.* **273**, 29406–29410.

Sabbert, D., Engelbrecht, S., Junge, W. (1996) Functional and idling rotatory motion within F_1-ATPase, *Nature* **381**, 623–625.

Sambongi, Y., Iko, Y., Tanabe, M., Omote, H., Iwamoto-Kihara, A., Ueda, I., Yanagida, T., Wada, Y., Futai, M. (1999) Mechanical rotation of the c subunit oligomer in ATP synthase (F_0F_1): direct observation, *Science* **286**, 1722–1724.

Schneider, E., Altendorf, K. (1985) All three subunits are required for the reconstitution of an active proton channel (F_0) of *Escherichia coli* ATP synthase (F_1F_0), *EMBO J.* **4**, 515–518.

Seelert, H., Poetsch, A., Dencher, N.A., Engel, A., Stahlberg, H., Müller, D.J. (2000) Structural biology. Proton-powered turbine of a plant motor, *Nature* **405**, 418–419.

Senior, A.E., Al-Shawi, M.K. (1992) Further examination of seventeen mutations in *Escherichia coli* F_1-ATPase β-subunit, *J. Biol. Chem.* **267**, 21471–21478.

Senior, A.E., Nadanaciva, S., Weber, J. (2000) Rate acceleration of ATP hydrolysis by F_1F_0-ATP synthase, *J. Exp. Biol.* **203**, 35–40.

Shin, K., Nakamoto, R.K., Maeda, M., Futai, M. (1992) F_0F_1-ATPase γ subunit mutations perturb the coupling between catalysis and transport, *J. Biol. Chem.* **267**, 20835–20839.

Shirakihara, Y., Leslie, A.G.W., Abrahams, J.P., Walker, J.E., Ueda, T., Sekimoto, Y., Kambara, M., Saika, K., Kagawa, Y., Yoshida, M. (1997) The crystal structure of the nucleotide-free $\alpha 3\beta 3$ subcomplex of F_1-ATPase from the thermophilic *Bacillus* PS3 is a symmetric trimer, *Structure* **5**, 825–836.

Singh, S., Turina, P., Bustamante, C.J., Keller, D.J., Capaldi, R.A. (1996) Topographical structure of membrane-bound *Escherichia coli* F_1F_0 ATP synthase in aqueous buffer, *FEBS Lett.* **397**, 30–34.

Soong, R.K., Bachand, G.D., Neves, H.P., Olkhovets, A.G., Craighead, H.G., Montemagno, C.D. (2000) Powering an inorganic nanodevice with a biomolecular motor, *Science* **290**, 1555–1558.

Stahlberg, H., Müller, D.J., Suda, K., Fotiadis, D., Engel, A., Meier, T., Matthey, U., Dimroth, P. (2001) Bacterial Na^+-ATP synthase has an undecameric rotor, *EMBO Rep.* **2**, 229–233.

Stalz, W.-D., Greie, J.-C., Deckers-Hebestreit, G., Altendorf, K. (2003) Direct interaction of subunits a and b of the F_0 complex of *Escherichia coli* ATP synthase by forming an ab_2 subcomplex, *J. Biol. Chem.* **278**, 27068–27071.

Stock, D., Leslie, A., Walker, J.E. (1999) Molecular architecture of the rotary motor in ATP synthase, *Science* **286**, 1700–1705.

Sun-Wada, G.-H., Wada, Y., Futai, M. (2003) Vacuolar H^+ pumping ATPases in luminal acidic organelles and extracellular compartments: common rotational mechanism and diverse physiological roles, *J. Bioenerg. Biomembr.* **35**, 347–358.

Sun-Wada, G.-H., Murakami, H., Nakai, H., Wada, Y., Futai, M. (2001) Mouse Atp6f, the gene encoding the 23-kDa proteolipid of vacuolar proton translocating ATPase, *Gene* **274**, 93–99.

Sun-Wada, G.-H., Imai-Senga, Y., Yamamoto, A., Murata, Y., Hirata, T., Wada, Y., Futai, M. (2002) A proton pump ATPase with testis-specific E1-subunit isoform required for acrosome acidification, *J. Biol. Chem.* **277**, 18098–18105. Y.

Takeyasu, K., Omote, H., Nettikadan, S., Tokumasu, F., Iwamoto-Kihara, A., Futai, M. (1996) Molecular imaging of *Escherichia coli* F_0F_1-ATPase in reconstituted membranes using atomic force microscopy, *FEBS Lett.* **392**, 110–113.

Tanabe, M., Nishio, K., Iko, Y., Sambongi, Y., Iwamoto-Kihara, A., Wada, Y., Futai, M. (2001) Rotation of a complex of the γ subunit and c ring of *Escherichia coli* ATP synthase. The rotor and stator are interchangeable, *J. Biol. Chem.* **276**, 15269–15274.

Tsunoda, S.P., Aggeler, R., Yoshida, M., Capaldi, R.A. (2001) Rotation of the c subunit oligomer in fully functional F_1F_0 ATP synthase, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 898–902.

Valiyaveetil, F.I., Fillingame, R.H. (1998) Transmembrane topography of subunit a in the *Escherichia coli* F_1F_0 ATP synthase, *J. Biol. Chem.* **273**, 16241–16247.

Van Raaji, M.J., Abrahams, J.P., Leslie, A.G.W., Walker, J.E. (1996) The structure of bovine F_1-ATPase complexed with the antibiotic inhibitor aurovertin B, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6913–6917.

Wada, Y., Sambongi, Y., Futai, M. (2000) Biological nano motor, ATP synthase F_0F_1: from catalysis to $\gamma\varepsilon c_{10-12}$ subunit assembly rotation, *Biochim. Biophys. Acta* **1459**, 499–505.

Watts, S.D., Capaldi, R.A. (1997) Interactions between the F_1 and F_0 parts in the *Escherichia coli* ATP synthase. Associations involving the loop region of *c* subunits, *J. Biol. Chem.* **272**, 15065–15068.

Weber, J., Senior, A.E. (2000) ATP synthase: what we know about ATP hydrolysis and what we do not know about ATP synthesis, *Biochim. Biophys. Acta* **1458**, 300–309.

Weber, J., Senior, A.E. (2001) Bi-site catalysis in F_1-ATPase: does it exist? *J. Biol. Chem.* **276**, 35422–35428.

Weber, J., Hammond, S.T., Wike-Mounts, S., Senior, A.E. (1998) Mg^{2+} coordination in catalytic sites of F_1-ATPase, *Biochemistry* **37**, 608–614.

Weber, J., Lee, R.S.V., Grell, E., Wise, J.G., Senior, A.E. (1992) On the location and function of tyrosine β331 in the catalytic site of *Escherichia coli* F_1-ATPase, *J. Biol. Chem.* **267**, 1712–1718.

Wikens, S., Capaldi, R.A. (1998) ATP synthase's second stalk comes into focus, *Nature* **393**, 29.

Yasuda, R., Noji, H., Kinoshita, K. Jr., Yoshida, M. (1998) F_1-ATPase is a highly efficient molecular motor that rotates with discrete 120 degree steps, *Cell* **93**, 1117–1124.

Yasuda, R., Noji, H., Yoshida, H., Kinoshita, K. Jr. (2001) Resolution of distinct rotational substeps by submillisecond kinetic analysis of F_1-ATPase, *Nature* **410**, 898–904.

Protusions: Cilia and Flagella: *see* Mobile Structures: Cilia and Flagella

Pufferfish Genomes: *Takifugu* and *Tetraodon*

Melody S. Clark[1] *and Hugues Roest Crollius*[2]
[1] *Biological Sciences Division, British Antarctic Survey Cambridge, UK*
[2] *Laboratoire Dynamique et Organisation des Genomes (LDOG), Department of Biology, Ecole Normale Superiere, Paris, France*

1 Introduction 363

2 Classification, Habitat, and Morphology of Pufferfish 364
2.1 Classification and Geographical Distribution of Pufferfish 364
2.2 Pufferfish Morphology 365

3 Pufferfish Genomes 365
3.1 Genome Size 365
3.2 Repetitive Elements 366
3.3 Gene Number 367
3.4 Pufferfish Chromosomes 368
3.5 Pufferfish Genetics 369

4 The Genome Projects 370
4.1 Sequencing Projects 370
4.2 Physical maps 371
4.3 Resources 371

5 Pufferfish as a Genome Tool 371
5.1 Comparative Genome Analysis in Vertebrates 371
5.2 Functional Analyses Including Detection of Regulatory Elements 375

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

6 Prospects 377

Bibliography 377
Books and Reviews 377
Primary Literature 378

Keywords

Conserved Order
Demonstration that three or more homologous genes lie on one chromosome in the same order in two separate species.

Conserved Segment
The syntenic association of two or more homologous genes in two separate species that are contiguous (not interrupted by different chromosome segments) in both species.

Conserved Synteny
The linked association of two or more homologous genes in two separate species regardless of gene order or interspersion of noncontiguous segments between two markers.

Karyotype
The chromosome complement of an organism.

Synteny
This term simply means linkage, that is, if two genes are syntenic, they are both present on the same chromosome.

Takifugu
While this is a species of pufferfish, this is generally the abbreviated name used in this text for *Takifugu rubripes*. This species is also known as *Fugu*, *Fugu rubripes*, and *Torafugu*. Its common name is the Japanese pufferfish.

Tetraodon
While this is also a species of pufferfish, this is generally the abbreviated name used in this text for *Tetraodon nigroviridis*. In the past, this species had been misnamed as *Tetraodon fluviatilis*, which is a different species of the *Tetraodon* family. *T. nigroviridis* is currently designated as *Chelonodon nigroviridis* in Fishbase. Its common name is the spotted green pufferfish.

The pufferfish have some of the smallest vertebrate genomes known (350–500 Mb). This lead to two species, the Japanese pufferfish and the spotted green pufferfish (*Takifugu rubripes* and *Tetraodon nigroviridis* respectively), being adopted as model vertebrates for in-depth genome-sequencing programs. These data have provided a wealth of information for comparative genomics studies in both fish and other vertebrate species, particularly human. This article describes the generalized features of the pufferfish genomes, the genome projects and resources available, in addition to the development of the pufferfish as genome tools.

1 Introduction

The raison d'être behind the pufferfish genome projects has to be viewed in a historical context. When the initial pufferfish genome project (using *Takifugu*) was proposed by Sydney Brenner in the early 1980s, sequencing technologies were still in their infancy. The sequencing programs of *Escherichia coli* and *Caenorhabditis elegans* had been started and it was these that provided the impetus behind the development of the high-throughput technologies, which are taken for granted today. Having said that, it did not seem conceivable at that time that they could be improved to such an extent as to allow the sequencing of the human genome. Therefore, Sydney Brenner proposed the need for an economical vertebrate model genome to bridge the gap between the data arising from the bacterial and invertebrate projects and human (Table 1). It had been known since 1968, from the work of Hinegardner, that several species of pufferfish had small genomes of approximately 0.4 pg per haploid nucleus (400 Mb). This equates to approximately one-eighth the size of the human genome. After preliminary sequencing scanning experiments showed a similar gene content to human, *T. rubripes* became accepted as a universal model genome. However, *Takifugu* was in some ways rather limited; it was only available from Japan and was therefore not universally accessible to researchers; it also grows rapidly to 1 kg within the first year and so cannot be kept in small laboratory fish tanks. Therefore, an alternative pufferfish species was proposed by Hans Lehrach – the spotted green pufferfish or *T. nigroviridis*. This is a

Tab. 1 Genome sizes of selected organisms. When the pufferfish was proposed as a model genome, sequencing technologies were relatively limited, with work started only on the bacteria and *Saccharomyces cerevisiae* and *C. elegans*.

Organism	*Common name*	*Genome size [Mb]*
Escherichia coli	Bacteria	4.64
Saccharomyces cerevisiae	Yeast	14
Caenorhabditis elegans	nematode	97
Takifugu rubripes	Japanese pufferfish	365
Tetraodon nigroviridis	spotted green pufferfish	350
Danio rerio	zebrafish	1900
Xenopus laevis	African clawed toad	3100
Gallus gallus	chicken	1200
Bos taurus	cattle	3651
Rattus norvegicus	rat	3000
Mus musculus	mouse	2500
Homo sapien	human	3000

small freshwater pufferfish that is available from many aquarium shops worldwide and so can be easily kept in most laboratories, providing a readily available source of DNA and tissue-specific RNA. Paradoxically, with the massive improvement in sequencing technologies, the human genome has been sequenced in advance of the pufferfish and the excess capacity of the large sequencing factories has since enabled the sequencing of both pufferfish genomes. The data from these is proving extremely valuable in comparative genomics and in the functional characterization of genes, as will be discussed Sect. 5.

2 Classification, Habitat, and Morphology of Pufferfish

2.1 Classification and Geographical Distribution of Pufferfish

Both pufferfish species (*T. rubripes* and *T. nigroviridis*) are members of the series Percomorpha, order Tetraodontiformes, family Tetraodontidae. The Tetraodontidae family comprises 19 genera with approximately 121 species. They are largely marine species, with several of them entering and occurring in brackish and freshwater. They live in tropical and subtropical regions, in the Atlantic, Indian, and Pacific oceans. Approximately 12 species of *Cariotetraodon, Chonerhinos*, and *Tetraodon* occur only in freshwater, primarily in the Congo river and southern Asia. The geographical locations of the two pufferfish species subject to genome sequencing are detailed further:

- *T. rubripes* is found in the Northwest Pacific: western part of the Sea of Japan and the East China and Yellow Seas northward to Muroran, Hokkaido, Japan. It is a temperate marine species, which occasionally enters brackish waters. It grows rapidly to a large size (70 cm is thought to be the maximum size it attains) and is not thought to breed until three-years old. It is a very important commercial species in the Japanese fisheries/aquaculture industry.
- *T. nigroviridis* is found in Asia: Sri Lanka to Indonesia, Malaysia, and north to China. It is a tropical (24–28 °C) freshwater species. It is a relatively small pufferfish species (maximum size 14 cm) and is sold in the aquarium trade worldwide.

The family Tetraodontidae is further split into two subfamilies: the Canthigastrinae (sharpnose puffers), which only comprises one genus and 26 species, while the rest are consigned to the Tetraodontinae. Paleontological evidence indicates that the families of the Tetraodontiform fishes diverged from each other between 90 and 50 million years ago, the fossil records for the Tetraodontiformes going back at least to the early Eocene (58 million years ago), indicating a long evolutionary history. Phylogenetic trees built from mitochondrial and nuclear genes showed that *Tetraodon* and *Takifugu* shared a common ancestor between 18 and 30 million years ago.

With two pufferfish genomes available for data mining, biologists are increasingly interested in related species as a means of exploiting these as biological models. While the order Tetraodontiformes is well defined (Fig. 1), there is relatively little information on the relationships between the different family members.

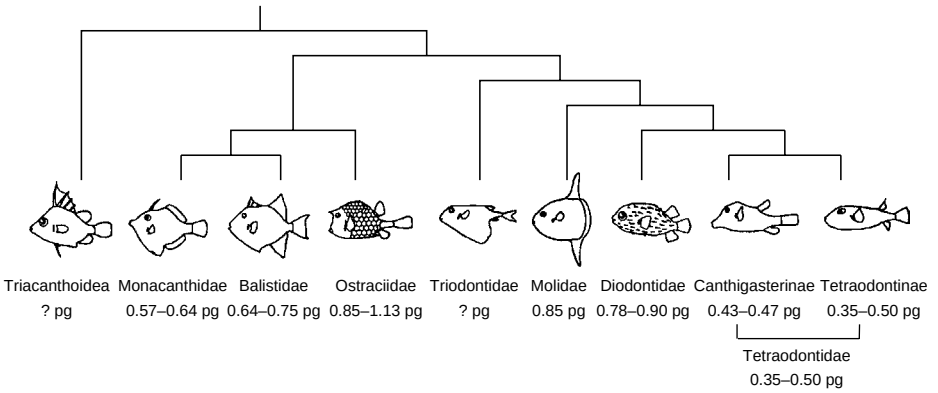

Fig. 1 Phylogeny and haploid genome sizes for sampled members of seven Tetraodontiform families. The phylogeny was based on morphological characters by Winterbottom, 1974. Figure reproduced from Brainerd, E.L., Slutz, S.S., Hall, E.K., Phillis, R.W. (2001) Patterns of genome size evolution in tetraodontiform fishes, *Evolution Int. J. Org. Evolution* **55**, 2363–2368, with permission.

2.2 Pufferfish Morphology

Pufferfish are generally recognized as the most advanced order of the teleosts, having a highly derived body form. The name Tetraodontiformes literally means four-toothed, which refers to the common pattern of four teeth in the outer jaws, although it has to be said that, on first glance, these often resemble more of a beak structure. Pufferfish are characterized by a high degree of fusion or loss of numerous bones in both the head and body. They usually have no lower ribs and have a reduced number of vertebrae when compared to most other fish species. The scales are usually modified as spines, shields, or bony plates over a thick leathery skin. The pufferfish have modified stomachs, which allows them to inflate to an enormous size by gulping down water, or in some cases air, when annoyed or threatened.

Some species, including *T. rubripes*, contain the alkaloid poison tetraodotoxin in certain tissues of their body. This is one of the most potent poisons known and acts by selectively blocking the voltage-sensitive sodium channels, which are essential components in cellular signaling pathways. The pufferfish have a point mutation in the protein sequence of the sodium-channel pump and so are resistant to the poison. In *Takifugu*, tetraodotoxin is present in high concentrations in the liver and ovaries and at lower concentrations in the intestines. The flesh, skin, and testes are not poisonous. However, despite the presence of a potentially lethal poison in certain tissues, *Takifugu* is highly prized item on the sushi menu. Hence the need for specially trained chefs to prepare *Takifugu*, ensuring that none of the toxic tissues is included in the sushi.

3 Pufferfish Genomes

3.1 Genome Size

The two pufferfish have similar genome characteristics, which will be described

in general terms here. Measurements of the nuclear DNA content showed that both contain approximately 0.35 to 0.4 pg per haploid genome, equivalent to a genome size between 350 and 400 megabases (Mb). The alignment of syntenic regions between *Tetraodon* and *Takifugu* indicates that the latter is larger by approximately 10%, although it will be necessary to wait until both genome sequences are complete for a final estimate. Both genomes are equivalent to approximately one-eighth the size of the human genome.

These compact genomes are characterized by smaller intergenic regions and introns and fewer repeat sequences (less than 10% of the genome). Data from the *Takifugu* genome shows that 75% of the introns are less than 425 bp, with a modal value of 79 bp. Only 500 introns are greater than 10 kb. The comparable figures for human are a modal value of 87 bp, which is not significantly different, but the 75% threshold is raised to 2609 bp, and over 12 000 introns are greater than 10 kb. The pufferfish genomes are prized by molecular biologists for their small genes, and, in general, pufferfish genes are scaled down compared to their human orthologs; however, a note of warning: a significant number are 1.3× larger.

Questions always arise as to why the genomes are so small and about the mechanisms involved. There is great diversity in fish genome sizes and this is also true of the Tetraodontiformes. Members of the Diodontidae (the spiny pufferfish), the sister family to the Tetraodontidae (smooth pufferfish), have genomes twice as large (Fig. 1). Studies indicate that the Diodontidae genomes have reached equilibrium, that is, are stable, but that this state of equilibrium was disturbed in the smooth pufferfish, following their divergence from the spiny puffers. The smaller genomes are thought to be the result of a decline in the rate of large-scale insertions, probably as a result of decreased transposable element activity. These are often found in the genome as middle repeat sequences (approximately 20% of the *Diodon hystrix* genome), but these are rare in *Takifugu*, a conclusion that presumably extends to the other smooth pufferfish with their compact genomes.

3.2 Repetitive Elements

Aside from a reduced number of middle repeat transposable elements, other types of such elements are also present in much lower quantities (approximately 1% of the total genome) compared to the other larger vertebrate genomes. It is the number rather than the types of transposable elements that are reduced as pufferfish contain sequences homologous to all known families of transposable elements including Ty3-gypsy, Tyl-copia, LINES, DNA transposons, and retroviruses. In general, these elements are dispersed throughout the genome but not in an even manner. *In situ* localization studies in *Tetraodon* show that they are compartmentalized in specific regions that correspond to the short arms of 10 pairs of small subtelocentric chromosomes (Fig. 2). These chromosome arms are heterochromatic and also contain a large number of pseudogenes that have arisen through duplication and retrotransposition. Characterization of such genes (iSET and Trapeze) has been carried out in *Tetraodon*, but it is not known whether they are present in *Takifugu* as the heterochromatic regions have been left out of the genome assembly to facilitate contiguation.

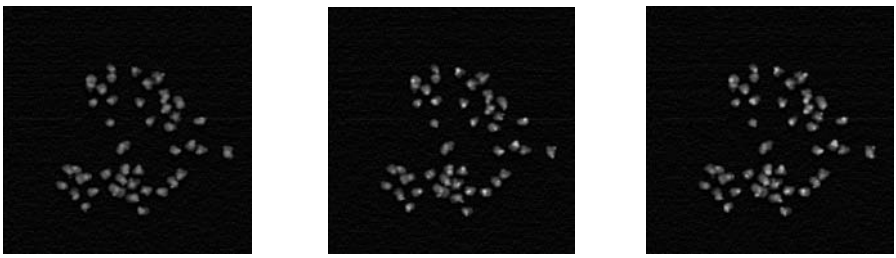

(a) (b) (c)

Fig. 2 *In situ* hybridization of two different non-LTR retrotransposons to *Tetraodon* chromosomes. (a) Green signals of Zebulon retrotransposon; (b)red signals of Barbar retrotransposon; (c) combined signal (yellow) of both retrotransposons. Both of these elements share the same sites in the genome, at the tip of some subtelocentric chromosomes. Babar is more abundant (in the shotgun data) and has more signals here, as seen by the additional red hybridizations in the combined data. (See color plate p. xxvii).

In addition to these transposable elements, the pufferfish genomes also contain two satellite sequences. The first is a 118-bp repeat, which maps to the centromeres. Interestingly, and despite their exact same size in both species, the two centromeric repeats show no noticeable sequence similarity. The second satellite is a subtelocentric repeat (10-mer in the case of *Tetraodon* (GGCGTCTGAG) and 20-mer in the case of *Takifugu* (GGCATCTGATCCTGGTAGCT)). Minisatellites are present to the extent of approximately 41% in *Tetraodon* (unknown in *Takifugu*) and microsatellites comprise between 2 to 3.5% of the pufferfish genomes. In total, repeat sequences comprise less than 10% of the genome.

The G + C content is higher in the pufferfish (46.2% in *Tetraodon* and 47.7% in *Takifugu*) compared to human (40.3%), which is thought to be due to the relative increased coding content. In *Takifugu*, the G + C heterogeneity is also less marked compared to mammals but is similar in *Tetraodon,* albeit shifted to higher G + C percentage. The reasons why *Tetraodon* contains more regions higher in G + C content than the closely related *Takifugu* are still unknown. In fact, a more homogeneous distribution in *Tetraodon* would be expected since it is a general condition associated with poikilothermic animals. The heterogeneous distribution of G + C content in mammals is associated with gene-rich and gene-poor regions, and the pufferfish also have gene-dense and gene-sparse regions. Overall, the average gene density in the pufferfish is one gene for every 10 kb, compared to 40 kb in human.

3.3 Gene Number

As regards the actual number of genes present in the pufferfish, there are only current estimates based on gene prediction programs. The number of predicted genes is greater than that of human (35 180 compared to 24 847), but it is suspected that fish have more genes than mammals owing- to additional duplication events. This is supported by results from whole-genome comparisons between *Takifugu* and *Tetraodon* on the one hand, and human and mouse on the other hand, which

show that the two fish contain a slight excess of regions conserved during evolution between tetrapods and teleosts. However, the number of transcripts does not seem to vary much between the two species (38 510 and 37 347). It should be noted that no estimate has been made of the rate of alternative transcription in the pufferfish and this could dramatically alter the number of transcripts. In human, 60% of genes are thought to be alternatively transcribed.

When the *Takifugu* draft genome was compared to the human sequence, three-quarters of human predicted genes showed a strong match on BLAST sequence similarity searching. One-quarter of human predicted proteins did not show a match and, therefore, are either highly diverged or not present. Six thousand *Takifugu* predicted genes showed no match to human predicted proteins. Although some of these genes may be fish-specific, it is unlikely that this is the case for the majority. Of the nonmatching human proteins, there were many cell surface receptor–ligand system proteins of the immune system, hemopoietic system, and energy/metabolism of homotherms. Immune cytokines, in general, were either not detected in *Takifugu* or showed very distant similarity. However, a focused effort to identify the class I and II cytokine receptors and their ligands in *Tetraodon* showed that the complete repertoire of these genes was present. This work required a careful examination of gene-structure conservation and a systematic cloning of transcripts for these genes, and highlights the fact that automatic annotation of genomes may leave out complete functional families, especially those submitted to accelerated sequence divergence.

3.4 Pufferfish Chromosomes

In depth, karyotypic analysis of most pufferfish chromosomes is difficult because of their small size (2 μm, on average). Therefore, most analysis is in the form of a basic giesma-stained karyotype. Cytogenetic information is available for 53 out of the 339 marine Tetraodontiformes. The Triacanthidae are considered the most basal clade of the Tetraodontiformes (based on morphological and osteologic criteria) and have a diploid chromosome complement of $2n = 2x = 48$. This mirrors the generalized Percomorpha complement from which they were derived and also represents the karyotypic base for all Tetraodontiformes (Fig. 3). Karyotypes of the more specialized species range from $2n = 2x = 28$ to 52, which are thought to represent derivations of this karyotype. This inference is supported by the occurrence of bi-armed chromosomes in this group, which is considered a derived condition in fish karyotypes.

Takifugu rubripes. $2n = 2x = 44$. Analysis of six different *Takifugu* species karyotypes shows that they are virtually identical, and the occurrence of natural hybrid species suggests that their morphology is indeed very similar. There is a large pair of submetacentric chromosomes within the complement, which is the characteristic marker chromosome pair of the group.

Tetraodon nigroviridis. $2n = 2x = 42$. In addition to the larger submetacentric chromosome pair, characteristic of the *Takifugu* species, *Tetraodon* also contains a large

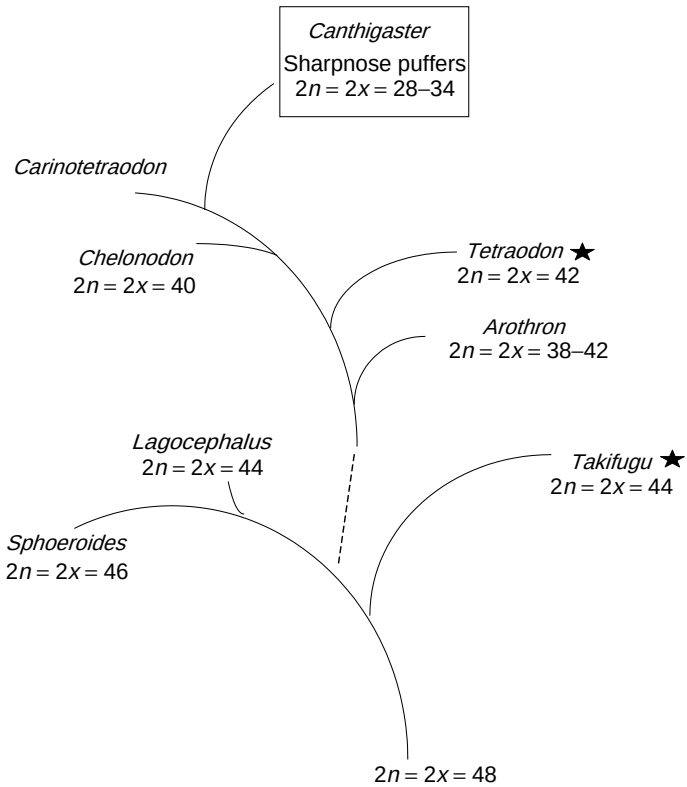

Fig. 3 Chromosome numbers of sampled Tetraodontinae overlaid on a restricted phylogeny on the basis of anatomical diversity, derived from J.C. (1980) *Osteology, phylogeny and higher classification of the fishes of the order Plectognathi (Tetraodontiformes)*. NOAA Technical Report NMFS Circular 434. US Department of Commerce, Washington, USA. Not all species of Tetraodontidae are shown. *Sphoeroides* is regarded as the most basal, having the most generalized osteology. $2n = 2x = 48$ is regarded as the ancestral karyotype. The stars indicate the species to which the sequenced pufferfish genomes belong. Virtually nothing is known about the evolutionary timescales between species in this family, although it has been estimated that *Takifugu* and *Tetraodon* diverged approximately 18 to 30 million years ago.

metacentric chromosome, presumably the result of a centric fusion between two of the smaller acrocentric chromosome pairs. This species is by far the best studied, in cytogenetic and physical mapping terms. Silver staining shows that there are two pairs of nucleolar organizer (NOR) chromosomes that are transcriptionally active.

3.5 Pufferfish Genetics

Owing to the problems in breeding these species, there are no linkage maps for pufferfish. In fact, no morphological characters enable the distinction between males and females in *Tetraodon*. Evidence

suggests that the inability to breed pufferfish may be due to some form of captive-induced stress. Histological studies of gonads have been performed in *Tetraodon* and show that both females and males are sexually mature when imported from Southeast Asia to Europe, but a high level of occlusions in female in vitellogenesis suggest that the last stages of sexual maturation are blocked because of some external factors that may be due to the stress of transportation. Breeding problems also arise with *Takifugu*, even when attempted in their native Japan, without being subjected to lengthy air-freight transportation.

4 The Genome Projects

4.1 Sequencing Projects

Both pufferfish genome projects have been generated using a Whole-Genome Shotgun (WGS) approach, where single read sequences are generated from the ends of cloned genomic DNA fragments. Both are currently at the stage of a "working draft," which means that, several times, genome coverage of clones have been assembled into a series of contigs (termed *scaffolds*). These vary dramatically in size from several 100 kb to 1 to 2 kb and are gradually being assembled into increasingly larger contigs. If both pufferfish were to be completely sequenced, then there would eventually be 21 contigs for *Tetraodon* ($2n = 2x = 42$) and 22 contigs for *Takifugu* ($2n = 2x = 44$), with each contig representing one of each chromosome pair. However, it is not known whether this will be the case; some regions, such as the telomeres and centromeres, are particularly difficult to assemble, and it is more likely that each genome will remain as a draft. Because of the nature of these projects, data are continually changing and, therefore, specific details are not given here and the reader is referred to the relevant web pages (below).

Tetraodon. This sequencing project was started at Genoscope in 1997 and they were joined by the Whitehead Institute for Genomic Research (WIGR) in June 2001. The approach taken was to end sequence a number of libraries containing different insert sizes: BAC (120–160 kb), cosmid (40 kb), and plasmid (4, 2.5, and 2 kb) libraries. The *Tetraodon* draft-sequence data is available on two sites, with an option on the French language version at Genoscope.

Genoscope: www.genoscope.cns.fr/externe/English/Projets/Projet_C/C.html

WIGR: www-genome.wi.mit.edu/annotation/Tetraodon

Takifugu. October 2000 saw the announcement by the US department of Energy that a consortium led by theJoint Genome Institute (JGI) was to completely sequence *Takifugu*. The consortium comprised groups from within the JGI, Molecular Sciences Institute, and Institute for Systems Biology in the US, along with MRC HGMP Resource Centre in the UK and the Institute of Molecular and Cell Biology in Singapore. This project mainly used end sequencing from a 2-kb plasmid library, but with additional sequence from a 5.5-kb plasmid library and BACs and cosmids. The *Takifugu* sequence information is available on four sites at the moment, all with different sequence viewers, although

the final official site is expected to be Ensembl.

Ensembl: www.ensembl.org/Fugu_rubripes

HGMP: http://www.fugu.hgmp.mrc.ac.uk

Institute of Molecular and Cellular Biology: http://www.fugu-sg.org

JGI: http://genome.jgi-psf.org/fugu6/fugu6.home.html

4.2 Physical maps

The requirement for physical mapping has, to a large extent, been superseded by the genome information in which large areas of the genome are represented by sequence contigs. These contigs are available in several databases (detailed in Sect. 4.1) and are generally well annotated with gene homology information. Genome maps are however useful for confirming the integrity of whole-genome sequence assemblies and providing sequencing material in regions where there are gaps. Because pufferfish cannot be bred to produce genetic linkage maps, the only map prior to the sequence data was generated by physically linking genomic clones across chromosomes. In *Tetraodon*, this was performed in several complementary ways. Firstly, single copy probes derived from the ends of BAC clones were hybridized onto high-density colony membranes carrying 10-fold redundant coverage of the genome in BAC clones. This generated nearly 3000 clone contigs spanning about 60% of the *Tetraodon* genome. Then, the same BAC clones were digested with restriction enzymes and their electrophoretic profile compared (the ''fingerprinting'' technique) to examine the extent of their overlap. This approach produced approximately 2000 contigs independently from the hybridization method, and both sets were then merged to provide a structure onto which the assembled genome sequence could be layered and confirmed. This mapping information, together with added paired clone end sequences, was used to generate ''ultracontigs'' by linking together sequence scaffolds. Several hundred double-color *in situ* hybridizations on metaphase chromosomes were finally necessary to anchor these ultracontigs (generally 1 to 10 megabases in size) to the *Tetraodon* genome. A similar mapping program is currently ongoing for the *Takifugu* genome, using the clone fingerprinting technique.

4.3 Resources

There are quite a few genomic resources available for both pufferfish species, with additional cDNA libraries available for both *Takifugu* and *Tetraodon*. There is also a P1 artificial chromosome library available for a related pufferfish species, *Sphoeroides nephelus* (the Southern pufferfish). This comprises 30 000 clones, which is equivalent to 7 to 8 genome equivalents. These are all detailed below in Table 2.

5 Pufferfish as a Genome Tool

5.1 Comparative Genome Analysis in Vertebrates

With the current abundance of whole-genome sequence in both the prokaryote and eukaryote kingdoms, ''comparative genomics'' is a term increasingly used to

Tab. 2 Resources (genomic and cDNA) for pufferfish.

Species	*Library*	*Type*	*Average insert size [kb]*	*Available*
Takifugu	genomic	cosmid	40	http://www.hgmp.mrc.ac.uk/geneservice/
	genomic	BAC	120	http://www.hgmp.mrc.ac.uk/geneservice/
	cDNA	plasmid		http://www.hgmp.mrc.ac.uk/geneservice/info@image.llnl.gov
Tetraodon	genomic (A)	BAC	126	www.genoscope.cns.fr/
	genomic (B)	BAC	153	www.genoscope.cns.fr/
	genomic (G)	plasmid	4	www.genoscope.cns.fr/
	cDNA	plasmid		www.genoscope.cns.fr/
Sphoeroides	genomic	P1	125–150	replicas: camemiya@vmrc.org pools: jpostle@oregon.uoregon.edu replicas and filters: http://www.rzpd.de

describe any approach where the sequence of two genomes (whole or partial) are compared to extract some knowledge about functional elements and the evolutionary processes that shape their sequences. Such studies were slow to take off in vertebrates because of the large genome size of the few species that were used as models or that were of interest to genome biologists (human, mouse, frog, zebrafish, chicken, etc.). The advent of sequencing projects in pufferfish changed this situation by providing a relatively economical and fast access to large regions of vertebrate genomes to be used in such comparative analyses.

The position of *Takifugu* and *Tetraodon* relative to mammals in the evolutionary tree is crucial to their importance as genome tools. The long divergence time since the split with tetrapods (approximately 400 million years) implies that functionally inert sequences that are submitted to a neutral rate of mutation have had time to diverge sufficiently to provide a good contrast with those that are functionally important and are thus submitted to purifying selection. This was empirically evidenced by the many small-scale analyses of regions sampled from the *Takifugu* genome and compared to their orthologous regions in human and sometimes mouse.

Initially, one of the reasons behind proposing the pufferfish as a model was the hypothesis that gene finding in human would be facilitated by a comparative positional cloning approach in a compact genome. For example, if the candidate disease region was 4 Mb in human, then by taking flanking markers to this region and mapping them to the pufferfish, this would equate to only 500 kb in the pufferfish. This left a much smaller region of DNA, in which to search, identify, and sequence the putative candidate disease gene. This relied on there being large-scale conservation of synteny between the pufferfish and human (the main reference organism). At the beginning of the 1980s, very little was known about gene-order conservation between vertebrates, so early pufferfish studies concentrated on these small-scale mapping studies. In retrospect, not surprisingly, there were many examples of conservation of synteny

between the pufferfish and human; however, many of these studies only encompassed two or three genes and often did not consider conservation of gene order. As larger data sets became available, it was clear that conservation of gene order over larger regions was rare (Fig. 4). Analysis of the *Takifugu* genome data showed that only 3.8% of the genome is in segments with no intervening sequences (perfect conservation of gene order between the pufferfish and human), while 5% is in segments with 1 intervening gene (mapping from a completely different human chromosome), 7% has up to 5 intervening genes, and 9.3% is in segments with up to 15 intervening genes. This is substantiated by mapping data in the zebrafish, which demonstrates large-scale conservation of synteny between zebrafish and human, at the whole chromosome level, but massive rearrangement of gene order between the two species. So the chance of finding a human candidate disease gene in the pufferfish by positional cloning is rare.

Map position		Gene	*Tetraodon* chromosome
9p	24.3	DMRT1	8
	24.1	AK3	1
	24.1	GLDC	9
	23	TYRP1	11
	13.3	CNTFR	8
Centromere			
9q	21.13	ANXA1	10
	21.13	GCNT1	8
	21.32	HNRPK	12
	22.32	HSD17B3	7
	22.33	TMOD	13
	33	EPB72	14
	33.2	PTGS1	8
	33.3	PBX3	15
	34.11	AK1	16
	34.11	ENG	17
	34.11	ASS	10
	34.12	ABL1	10

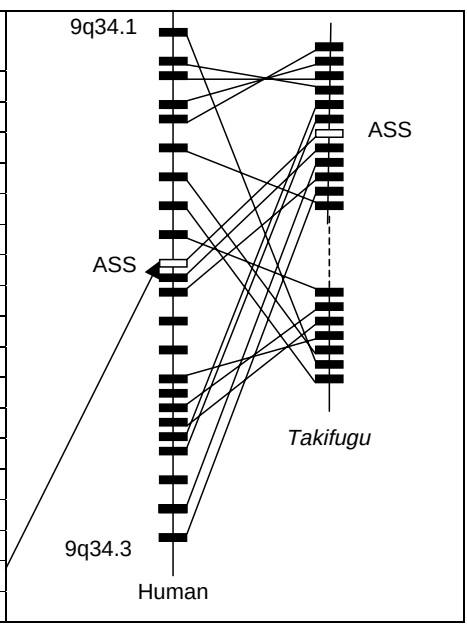

Fig. 4 Comparative mapping data from human chromosome 9, compared with *in situ* hybridization mapping data onto *Tetraodon* chromosomes (data taken from Gruetzner, F., Roest Crollius, H., Lutjens, G., Jaillon, O., Weissenbach, J., Ropers, H.H., Haaf, T. (2002) Four-hundred million years of conserved synteny of human Xp and Xq genes on three *Tetraodon* chromosomes, *Genome Res.* **12**, 1316–1322, with human map positions updated with Ensembl v.16.33.1) and also *Takifugu* BAC mapping data taken from Bouchireb, N., Gruetzner, F., Haaf, T., Stephens, R.J., Elgar, G., Green, A.J., Clark, M.S. (2001) Comparative mapping of the human 9q34 region in *Fugu rubripes, Cytogenet. Cell Genet.* **94**, 173–179. The data clearly shows that genes from human chromosome 9 are scattered across 12 different *Tetraodon* chromosomes. ASS, which maps to human 9q34.11, was mapped to a single *Takifugu* chromosome along with 2 BAC contigs containing 19 other human 9q34 genes. The diagram shows the rearrangements in gene order between the two species, which, in human, covers a region of 10 Mb.

There are exceptions to this general rule. At the gene level, there are the Hox complexes, and this is probably because of their coordinated expression in development processes. At the chromosomal level, there is some conservation of synteny of genes found on the human X. However, the X was almost certainly subjected to a special selection process during vertebrate development when compared with the rest of the chromosome complement. So, although one of the original reasons for studying the pufferfish is no longer robust, the pufferfish genomes have contributed significantly to our understanding of gene and chromosome evolution. Comparative analyses of human and pufferfish are still very useful for gene finding and also for promoter identification, as will be described in the following sections. As regards conservation of gene order between the two pufferfish, preliminary data indicates that they are nearly identical, which is not surprising when considering their close evolutionary relatedness.

Another way to exploit pufferfish genome sequence is to identify genes in other species. This became possible on a genome scale when the first samples of the *Tetraodon* WGS sequence became available. First, it was necessary to optimize sequence comparison algorithms so that the computing time required to compare a large fraction of vertebrate genome could be counted in days rather than months or years. Such parameters were thus adjusted using the BLAST sequence comparison algorithm so that about 40% of the human genome and 30% of the *Tetraodon* genome (the respective fraction of the two genomes available at the time) could be compared in about 5 days on a 100 CPU Unix cluster. The alignments produced were then filtered, on the basis of their length and their percent identity, so that only those that were thought to align over protein coding regions were retained. The criteria used in this selection were calibrated on known pufferfish and human genes. When several regions or sequence reads from *Tetraodon* overlapped on a given human region, this stretch of human DNA was dubbed an evolutionary conserved region (ecore). This first large-scale comparison between significant fractions of two vertebrate genomes produced a surprising result: the number of ecores estimated in the entire genome (once extrapolated from the 40% used in the comparison) was much lower than what would be expected if humans were to have the 80 000 to 100 000 genes given by previous estimates. Results from Exofish, the tool developed around the Human–*Tetraodon* comparison, suggested that the human genome contained between 28 000 and 34 000 genes, which is similar to the plant *Arabidopsis thaliana* and only five times more than the unicellular yeast. At the same time, this number was supported by a careful reanalysis of human EST sequence clusters, which suggested that the human genome contained no more than 35 000 genes. Today, with a finished sequence at hand, automatic annotation programs such as the Ensembl project have identified approximately 24 000 human genes and it is thought that few genes remain to be discovered. In fact, a comparison between the *Takifugu* draft genome sequence and an earlier version of the human genome suggested that less than 1000 human genes conserved between fish and mammals remained to be annotated.

Pufferfish genomes still have much to contribute to genome research. For instance, the human and mouse genomes are notorious for their abundance of transposable elements (approximately 45% of the genome). Conversely, probably the

main reason for the small pufferfish genome size is their paucity in repeats. This difference could be exploited to address a key question: how much DNA is under evolutionary selection in a vertebrate genome? Because this fraction is likely to be small, it cannot be easily measured against the background of neutrally evolving DNA represented by the repeats. In fact, a rough estimate indicated that approximately 5% of the mammalian genome could be under selection, when comparing the mouse draft sequence to the human. But aligning the two pufferfish genomes together may provide a much better answer, and, moreover, may indicate where these regions could be.

5.2 Functional Analyses Including Detection of Regulatory Elements

The "classic" gene functional analysis tools of transgenics, knock-ins, and knockouts are not possible in the pufferfish because of problems with breeding and size (in the case of *Takifugu*). Additionally, there has been only one pufferfish cell line made and so cell culture expression studies in "same species" is also not generally available. However, this will not hinder the exploitation of the pufferfish genomes. While characterization of pufferfish gene structure is valuable for gene identification (see Sect. 5.1), identifying conserved functional domains, understanding gene evolution processes, and to fish biologists working on other species, the main exploitation will be for the identification and characterization of conserved noncoding regulatory elements. Understanding gene function is the next stage for all genome projects and activity is obviously particularly strong in the mammalian field. Comparative analyses of mammalian genomes with those of the pufferfish can greatly advance this area of research.

The true extent of a gene, or rather functional unit, comprising coding regions, 5 and 3′ UTRs (untranslated regions), and regulatory elements can span many 100 kb in mammals. An example of this is the human *Sox9* gene, which is surrounded by 2 Mb 5′ and 500 kb 3′ of noncoding DNA. In *Takifugu*, these regions are reduced to 68 kb and 97 kb respectively, representing a considerable compaction of sequence. What exists in these noncoding regions between genes is only just being investigated; although, homology between species is generally very low when considering UTRs and these can be determined relatively easily by sequencing of full-length cDNAs or by RACE experiments. Regulatory regions are much more difficult to define, they are very small, often less than 10 bp in length, and may be at a considerable distance from the coding region of the gene they control. This is the real use of the pufferfish; because of their small genome size, the essential regulatory elements may also be squeezed much closer to the genes they regulate. Despite the evolutionary distance between the pufferfish and mammals, some of the regulatory sequences remain similar enough to those of mammals to be recognized by sequence comparisons and also to be functional in transgenic organisms. To identify and characterize a regulatory element, there are two requirements: that the regulatory elements are present in the DNA and that there is an assay system.

Recognition of regulatory regions can be determined using a variety of methods, such as BLAST sequence similarity searching, Clustal multiple sequence alignments, or purpose-built programs: Pipmaker (http://bio.cse.psu.edu/pipmaker), Vista (http://www-gsd.lbl.gov/vista), or

Dotter. Having determined that there are conserved regions between the pufferfish and mammals in a gene of interest, these then require experimental testing.

So far, use of pufferfish DNA in transgenics has had a mixed history, particularly with regard to the use of whole pufferfish genes. Success was achieved in rat for the *Takifugu* isotocin and oxytocin genes, but other examples, such as the HD and Wt1 genes, have resulted in aberrant splicing and nonfunctional products. The sequencing of the zebrafish genome and the increasing number of zebrafish facilities has opened up a new and potentially very powerful avenue for pufferfish research. While pufferfish genes may not generally splice correctly in mammals or mammalian cell lines, there appear to be no such problems in zebrafish, indicating that the aberrant splicing of pufferfish genes that occurs in mammals is related to evolutionary differences in splicing efficiency between fish and mammals. Having said that, it is the regulatory regions that are the real source of exploitation, rather than the actual genes themselves.

Pufferfish promoter regions have been inserted upstream of reporter constructs and expressed in a range of cell lines (mammalian and piscine). They have enabled the narrowing of regions responsible for the control of gene expression in genes as varied as the *Mx* protein (involved in the innate immune system), *Ick* (a lymphocyte-specific protein tyrosine kinase), *Pax 6* (involved in eye development), *tyrosinase* (pigmentation gene), and *SCL* (stem cell leukemia gene involved in hemopoiesis and vasculogenesis). This number of examples, currently rather limited, will dramatically increase over the next few years, as more groups access and mine the pufferfish data.

Pufferfish promoters have also been assayed, to a lesser extent, in transgenic animals. The *Takifugu* SCL locus directed appropriate expression to hemopoietic and neural tissue in transgenic zebrafish embryos and the Wt1 gene from the same species was faithfully expressed in the developing zebrafish kidney. Ick and Pax 6 *Takifugu* noncoding regions were successfully tested in transgenic mice. In the human *Pax 6* gene, there is a 3′ cis-acting regulatory element termed C1170Box123. This contains three modules in human and mouse, but only one in *Takifugu*. However, this one conserved module was assayed in mice, using both *Takifugu* and human 3′ constructs and both produced similar reporter gene-expression patterns. This latter example clearly demonstrates the point that not all regulatory regions will be conserved between fish and mammals. Indeed, this cannot be expected as there are considerable differences in morphology, physiology, and biochemistry between the two sets of species. Also, just because no homology is detected between mammalian and pufferfish promoter sequences, this does not mean that pufferfish 5′ regions cannot direct expression of mammalian genes. Although the *Takifugu* CRABP-1 promoter showed no homology to that of mouse, expression of the mouse gene was still obtained, but in a much more tissue-restricted manner.

Analysis of promoter elements using comparisons between species does rely on similarity of expression of the gene under study. Clearly there will be many examples between mammals and fish where this is not so, but fish also have an additional use, as regards dissection of promoter elements. It is well documented that fish have extra genes. It is thought that these duplicate pairs survive via the processes

of nonfunctionalization, neofunctionalization, and subfunctionalization. This duplication, which in many quarters is seen as a drawback of fish models, is actually very much an advantage. Comparison of duplicated genes between fish species and also with other vertebrates allows subfunction partitioning between the two genes, a situation that is not possible in mammalian models. Analysis of promoters between duplicated genes in fish can assign tissue- or developmental-specific functions to either or both of the pair of genes and can break up pleiotropic gene function and regulation. Few examples are available so far, but analysis of zebrafish, the two pufferfish, and *Xiphophorus*-duplicated *mitf* genes demonstrated subfunctionalization as a result of degeneration of alternative exons. This again is an area of investigation set to expand, particularly with regard to genes, which present in single copies in mammals are lethal when mutated. Mutations of such genes in fish are often still viable (as there are two copies) if only one of each duplicated pair is treated at any one time.

6 Prospects

The completion of the *Takifugu* and *Tetraodon* draft genome sequences mark a new step in the field of whole-genome sequencing in eukaryotes. Indeed, the first eukaryotes to be sequenced other than human (yeast, worm, fly, mustard weed, and mouse) were natural first choices: decades of laboratory experiments had accumulated a wealth of genetic and functional information, and their genome sequence was thought of as a conceptual framework in which to unify this diverse and heterogeneous data. In contrast, very little was known on the biology and life cycle of the two pufferfish described in this article. The reason for sequencing them lay exclusively in their position in the phylogenetic tree with respect to human and in their unusually small genome size. Their genome sequence has been a platform for gene discovery and genome comparisons.

But the *Takifugu* and *Tetraodon* genomes have not said their last word. Today, with the upcoming availability of the zebrafish genome sequence, it is expected that much will be gained by adequately combining functional tools developed in this species and comparative genomics tools represented by the pufferfish genomes.

See also Genetics, Molecular Basis of; Genomic Sequencing (Core Article); Immunoassays; Shotgun Sequencing (SGS); Whole Genome Human Chromosome Physical Mapping.

Bibliography

Books and Reviews

Ebert, E. (2001) *The Puffers of Fresh and Brackish Waters*, Aqualog verlag GmbH, Rodgau.

Helfman, G.S., Collette, B.B., Facey, D.E. (1997) *The Diversity of Fishes*, Blackwell Science, Malden, Massachusetts, USA.

Muller, F., Blader, P., Strahle, U. (2002) Search for enhancers: teleost models in comparative genomic and transgenic analysis of *cis* regulatory elements, *Bioessays* **24**, 564–572.

Nelson, J.S. (1994) *Fishes of the World*, J. Wiley & Sons, New York, USA.

Rothenberg, E.V. (2001) Mapping of complex regulatory elements by pufferfish/zebrafish transgenesis, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6540–6542.

Thomas, J.W., Touchman, J.W. (2002) Vertebrate genome sequencing: building a backbone for comparative genomics, *Trends Genet.* **18**, 104–108.

Tyler, J.C. (1980) *Osteology, phylogeny and higher classification of the fishes of the order Plectognathi (Tetraodontiformes)*. NOAA Technical Report NMFS Circular 434. US Department of Commerce, Washington, USA.

Winterbottom, R. (1974) *The Familial Phylogeny of the Tetraodontiformes (Acanthopterygii: Pisces) as Evidenced by their Comparative Mycology*, Smithsonian Contributions to Zoology, Number 155. Smithsonian Institution Press, Washington, DC.

Primary Literature

Altschmied, J., Delfgaauw, J., Wilde, B., Duschl, J., Bouneau, L., Volff, J.N., Schartl, M. (2002) Subfunctionalization of duplicate mitf genes associated with differential degeneration of alternative exons in fish, *Genetics* **161**, 259–267.

Aparicio, S., Hawker, K., Cottage, A., Mikawa, Y., Zuo, L., Venkatesh, B., Chen, E., Krumlauf, R., Brenner, S. (1997) Organization of the *Fugu rubripes* Hox clusters: evidence for continuing evolution of vertebrate Hox complexes, *Nat. Genet.* **16**, 79–83.

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., et al. (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*, *Science* **297**, 1301–1310.

Bagheri-Fam, S., Ferraz, C., Demaille, J., Scherer, G., Pfeifer, D. (2001) Comparative genomics of the SOX9 region in human and *Fugu rubripes*: conservation of short regulatory sequence elements within large intergenic regions, *Genomics* **78**, 73–82.

Barton, L.M., Gottgens, B., Gering, M., Gilbert, J.G., Grafham, D., Rogers, J., Bentley, D., Patient, R., Green, A.R. (2001) Regulation of the stem cell leukemia (SCL) gene: a tale of two fishes, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6747–6752.

Baxendale, S., Abdulla, S., Elgar, G., Buck, D., Berks, M., Micklem, G., Durbin, R., Bates, G., Brenner, S., Beck, S. (1995) Comparative sequence analysis of the human and pufferfish Huntington's disease genes, *Nat. Genet.* **10**, 67–76.

Bouchireb, N., Gruetzner, F., Haaf, T., Stephens, R.J., Elgar, G., Green, A.J., Clark, M.S. (2001) Comparative mapping of the human 9q34 region in *Fugu rubripes*, *Cytogenet. Cell Genet.* **94**, 173–179.

Brainerd, E.L., Slutz, S.S., Hall, E.K., Phillis, R.W. (2001) Patterns of genome size evolution in tetraodontiform fishes, *Evolution Int. J. Org. Evolution* **55**, 2363–2368.

Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B., Aparicio, S. (1993) Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome, *Nature* **366**, 265–268.

Brenner, S., Venkatesh, B., Yap, W.H., Chou, C.F., Tay, A., Ponniah, S., Wang, Y., Tan, Y.H. (2002) Conserved regulation of the lymphocyte-specific expression of lck in the *Fugu* and mammals, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2936–2941.

Brum, M.J.I., Galetti, Jr, P.M. (1997) Teleostei ground plan karyotype, *J. Comp. Biol.* **2**, 91–102.

Camacho-Hubner, A., Richard, C., Beermann, F. (2002) Genomic structure and evolutionary conservation of the tyrosinase gene family from *Fugu*, *Gene* **285**, 59–68.

Camacho-Hubner, A., Rossier, A., Beermann, F. (2000) The *Fugu* rubripes tyrosinase gene promoter targets transgene expression to pigment cells in the mouse, *Genesis* **28**, 99–105.

Crnogorac-Jurcevic, T., Brown, J.R., Lehrach, H., Schalkwyk, L.C. (1997) *Tetraodon fluviatilis*, a new pufferfish model for genome studies, *Genomics* **41**, 177–184.

Dalle Nogare, D.E., Clark, M.S., Elgar, G., Frame, I.G., Poulter, R.T. (2002) Xena, a full-length basal retroelement from tetraodontid fish, *Mol. Biol. Evol.* **19**, 247–255.

Dasilva, C., Hadji, H., Ozouf-Costaz, C., Nicaud, S., Jaillon, O., Weissenbach, J., Roest Crollius, H. (2002) Remarkable compartmentalization of transposable elements and pseudogenes in the heterochromatin of the *Tetraodon nigroviridis* genome, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13636–13641.

Fischer, C., Ozouf-Costaz, C., Roest Crollius, H., Dasilva, C., Jaillon, O., Bouneau, L., Bonillo, C., Weissenbach, J., Bernot, A. (2000) Karyotype and chromosomal localization of characteristic tandem repeats in the pufferfish

Tetraodon nigroviridis, *Cytogenet. Cell genet.* **88**, 50–55.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J. (1999) Preservation of duplicate genes by complementary, degenerative mutations, *Genetics* **151**, 1531–1545.

Gilligan, P., Brenner, S., Venkatesh, B. (2002) *Fugu* and human sequence comparison identifies novel human genes and conserved non-coding sequences, *Gene* **294**, 35–44.

Griffin, C., Kleinjan, D.A., Doe, B., van Heyningen, V. (2002) New 3′ elements control Pax6 expression in the developing pretectum, neural retina and olfactory region, *Mech. Dev.* **112**, 89–100.

Gruetzner, F., Lutjens, G., Rovira, C., Barnes, D.W., Ropers, H.H., Haaf, T. (1999) Classical and molecular cytogenetics of the pufferfish *Tetraodon nigroviridis*, *Chromosome Res.* **7**, 655–662.

Gruetzner, F., Roest Crollius, H., Lutjens, G., Jaillon, O., Weissenbach, J., Ropers, H.H., Haaf, T. (2002) Four-hundred million years of conserved synteny of human Xp and Xq genes on three *Tetraodon* chromosomes, *Genome Res.* **12**, 1316–1322.

Hinegardner, R. (1972) Cellular DNA content and the evolution of teleostean fishes, *Am. Nat.* **106**, 621–644.

Jaillon, O., et al. (2003) The *Tetraodon nigroviridis* genome sequence (in preparation).

Kleinjan, D.A., Dekker, S., Guy, J.A., Grosveld, F.G. (1998) Cloning and sequencing of the CRABP-I locus from chicken and pufferfish: analysis of the promoter regions in transgenic mice, *Transgenic Res.* **7**, 85–94.

Lamatsch, D.K., Steinlein, C., Schmid, M., Schartl, M. (2000) Noninvasive determination of genome size and ploidy level in fishes by flow cytometry: detection of triploid *Poecilia formosa*, *Cytometry* **39**, 91–95.

Miles, C.G., Rankin, L., Smith, S.I., Niksic, M., Elgar, G., Hastie, N.D. (2003) Faithful expression of a tagged *Fugu* WT1 protein from a genomic transgene in zebrafish: efficient splicing of pufferfish genes in zebrafish but not mice, *Nucleic Acids Res.* **31**, 2795–2802.

Miyaki, K., Tabeta, O., Kayano, H. (1995) Karyotypes in six species of pufferfishes genus *Takifugu* (Tetraodontidae, Tetraodontiformes), *Fish. Sci.* **61**, 594–598.

Neafsey, D.E., Palumbi, S.R. (2003) Genome size evolution in pufferfish: a comparative analysis of diodontid and tetraodontid pufferfish genomes, *Genome Res.* **13**, 821–830.

Postlethwait, J.H., Woods, I.G., Ngo-Hazelett, P., Yan, Y.L., Kelly, P.D., Chu, F., Huang, H., Hill-Force, A., Talbot, W.S. (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes, *Genome Res.* **10**, 1890–1902.

Poulter, R., Butler, M. (1998) A retrotransposon family from the pufferfish (*fugu*) *Fugu rubripes*, *Gene* **215**, 241–249.

Poulter, R., Butler, M., Ormandy, J. (1999) A LINE element from the pufferfish (*fugu*) *Fugu rubripes* which shows similarity to the CR1 family of non-LTR retrotransposons, *Gene* **227**, 169–179.

Roest Crollius, H., Jaillon, O., Bernot, A., Dasilva, C., Bouneau, L., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W., Weissenbach, J. (2000) Human gene number estimate provided by genome wide analysis using *Tetraodon nigroviridis* genomic DNA, *Nat. Genet.* **25**, 235–238.

Roest Crollius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C., Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A., Weissenbach, J. (2000) Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*, *Genome Res.* **10**, 939–949.

Sathasivam, K., Baxendale, S., Mangiarini, L., Bertaux, F., Hetherington, C., Kanazawa, I., Lehrach, H., Bates, G.P. (1997) Aberrant processing of the *Fugu* HD (FrHD) mRNA in mouse cells and in transgenic mice, *Hum. Mol. Genet.* **6**, 2141–2149.

Schuddekopf, K., Schorpp, M., Boehm, T. (1996) The *whn* transcription factor encoded by the nude locus contains an evolutionarily conserved and functionally indispensable activation domain, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9661–9664.

Thomas, J.W., Touchman, J.W., Blakesley, R.W., Bouffard, G.G., Beckstrom-Sternberg, S.M., Margulies, E.H., Blanchette, M., Siepel, A.C., Thomas, P.J., McDowell, J.C., et al. (2003) Comparative analyses of multi-species sequences from targeted genomic regions, *Nature* **424**, 788–793.

Venkatesh, B., Si-Hoe, S.L., Murphy, D., Brenner, S. (1997) Transgenic rats reveal functional conservation of regulatory controls between the *Fugu* isotocin and rat oxytocin genes, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12462–12466.

Yap, W.H., Tay, A., Brenner, S., Venkatesh, B. (2003) Molecular cloning of the pufferfish (*Takifugu rubripes*) Mx gene and functional characterization of its promoter, *Immunogenetics* **54**, 705–713.

Quantitative Analysis of Biochemical Data

Albert Jeltsch[1], *Jim Hoggett*[2], *and Claus Urbanke*[3]
[1] *International University Bremen, Bremen, Germany*
[2] *University of York, York, UK*
[3] *Medizinische Hochschule Hannover, Zentrale Einrichtung Biophysikalisch-Biochemische Verfahren, Hannover, Germany*

1 Quantitative Data 383

2 Numeric Modeling of Experiments 383
2.1 Formal Definition of the Mathematical Problem 383
2.2 Fitting 386
2.3 Experimental Systems 387
2.4 Measurement and Signals 388
2.5 Models 388
2.6 Selection of Appropriate Models 389
2.7 Parameters 389
2.8 Essential Steps in the Analysis 390
2.9 Fitting Data by the Method of Least Squares 390
2.10 Global Fitting of Multiple Data Sets 392
2.11 Introduction to Error Estimation 393
2.12 Introduction to Numerical Integration 394

3 Applications 396
3.1 Linear Regression 396
3.2 Michaelis–Menten Kinetics 396
3.3 Analysis of Circular Dichroism Experiments to Determine the Secondary Structure Composition of Proteins 397
3.4 Dissociation Kinetics 397
3.5 Binding Data 398
3.6 Independent Identical Binding Sites 398
3.6.1 Analysis of Simple Binding Data 399
3.6.2 Independent Nonidentical Binding Sites 399

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

3.6.3 Cooperative Binding 400
3.7 pH Dependence of Enzyme Catalyzed Reactions 401
3.8 Pre-steady State Kinetics 402
3.9 Association Kinetics Analyzed by Numerical Integration 403
3.10 Surface Binding Reactions 404
3.11 Analysis of Competition Experiments 404
3.12 Regularization – Application to Analytical Ultracentrifugation 406

Bibliography 408
Books and Reviews 408
Primary Literature 408

Keywords

Algorithm
A defined rule for calculating output numbers from input numbers, usually carried out by a computer.

Gaussian (or Normal) Distribution
A distribution of values with the probability of the distribution being proportional to the distance squared from the mean.

Global Fitting
Fitting experimental data from various data sets to a single model function.

Least-squares Fit
Method of fitting experimental data to a function by minimization of the squared deviations of the experimental and theoretical points.

Model
An abstraction (usually a simplification) of the processes that are happening in a real system that allows the processes to be described in mathematical language.

Numerical Integration
Obtaining the value of an integral or the solution of a differential equation in a stepwise procedure using only numerical values.

Occam's Razor
Also known as the principle of parsimony, this means that other things being equal, the model or hypothesis with the smaller number of assumptions or parameters is to be preferred.

Parameters
Constant elements of a function.

Biochemistry is an experimental science characterized by an analytical and quantitative approach. Methods for the analysis of quantitative data are essential tools for handling the results of many biochemical studies. Some approaches are relatively straightforward and familiar to the biochemical community; others are more specialized, and their use is consequently less common. This chapter provides an account of two principal approaches to quantitative data analysis: least-squares fitting and regularization methods. A short summary of the theoretical background of these two approaches is followed by extensive coverage of applications taken from diverse areas including: enzyme kinetics, ligand binding, analytical ultracentrifugation, circular dichroism (CD), and the pH dependence of biochemical reactions. The discussion covers the nature of mathematical modeling of systems, and emphasizes the importance of selection of an appropriate physical model to describe the data.

1 Quantitative Data

The outcome of some biochemical experiments can be expressed as simple descriptive statements, such as "the desired band or peak was observed." In others, such as studies of the rates of processes or the affinity of a ligand for its target, the results need to be given in quantitative form as numerical values of one or more parameters. Such quantitative results are derived by mathematical analysis of the raw experimental data. As an example, consider the time course of an enzymatic reaction. The raw data are the dependence of product concentration (the dependent variable, conventionally shown on the y-axis) on the time (the independent variable, conventionally on the x-axis). The mathematical methods used to handle these and similar quantitative data can be of greater or lesser complexity depending on the processes involved and on the detail or depth of analysis required.

2 Numeric Modeling of Experiments

2.1 Formal Definition of the Mathematical Problem

This paragraph gives a formal introduction to the mathematical approaches most relevant to quantitative biochemical data analysis. For details on practical application, please refer to Sects. 2.2 and 3. In quantitative biochemical experiments, the outcome R can be described by a set of values usually expressed as real numbers. The elements of this set usually have all the mathematical properties of a function f, assigning the observable quantities (concentrations, spectroscopic signals etc.) to the independent variables $(x_i^1 \cdots x_i^n)$

(place, time, concentration of reactants ...) and constant parameters ($p_1 \cdots p_m$) (temperature, pH, buffer composition ...):

$$R = \{f_1(x_1^1 \ldots x_1^n; p_1 \ldots p_m), f_2(x_2^1 \ldots x_2^n; p_1 \ldots p_m), \ldots, f_k(x_k^1 \ldots x_k^n; p_1 \ldots p_m)\} \quad (1)$$

This general treatment can be simplified if one considers only a single observable. Then the result can be represented as an n-dimensional surface in $(n+1)$-dimensional space, created by the n-dimensions of the independent variables and one additional dimension for the experimental result. For further simplification, if there is only a single independent variable, this hypersurface is reduced to a situation that is much more familiar to a biochemist, that is, a set of points in a two-dimensional plot. One should, however, keep in mind that simply connecting these points with a continuous line introduces an assumption about the order of the functions and implies that one expects "in-between" functions that have not been measured. This procedure is only valid if there are good reasons to support this assumption, but fortunately this is more often than not true.

To illustrate this general function further, let us consider an enzyme kinetic measurement where the concentration of the product formed is measured as a function of time, while the enzyme concentration, initial substrate concentration, and temperature, as well as buffer composition are kept constant. In formal mathematical language this is expressed as follows:

$$c_{\text{Product}}(t) = f(t; c_{\text{Enzyme}}, c_{\text{Substrate}}^{\text{initial}}, pH, T, c_{\text{Buffer}})$$

In the experiment that we have just envisaged, time is the only independent variable, but repeating the experiment for different enzyme concentrations or initial substrate concentrations may introduce additional independent variables such that the complete experiment R consists of a set of time series of product concentration at different enzyme and initial substrate concentrations. Depending on the scientific question that is being posed, one also might vary the pH, temperature, or buffer composition. While in most experiments it is assumed that the independent variables and the constant parameters are known *a priori*, the values of these functions, which are the experimental results, are of course subject to experimental uncertainties, in other words – they are noisy. In obtaining such an experimental value, one is effectively taking a sample from a potentially infinitely large set of possible readings whose distribution reflects the experimental uncertainty. It follows, therefore, that statistical factors must be taken into consideration in any quantitative interpretation of results, a statement that will be all too familiar to anyone who has actually been involved in carrying out experiments and analyzing data.

Evaluation of such quantitative data is thus concerned with finding a suitable algebraic description for the earlier-mentioned functions to interpret it properly. The algebraic description that was used historically was a linear function, because then the only way of carrying out a regression analysis of data was fitting to a straight line. Data that were not themselves linear were transformed in such a way that they satisfied a linear relationship. Well-known examples of such transformations are the Scatchard analysis of binding data or Lineweaver–Burk analysis of enzyme kinetics. Unfortunately, all such linearization procedures inevitably

distort the experimental data by nonlinear transformation of the noise, which is one of the major sources of error in data interpretation.

Since computing power is now not a limiting factor and linearization of data is no longer necessary, or indeed desirable, algorithms have been developed that allow the direct analysis of the raw experimental data without the necessity of transformation. Basically, an algebraic model is devised that can simulate the unknown function that is being measured experimentally. Such a model should be based on physical considerations about the experiment, but it may also be partly or completely empirical. In addition to the independent variables $(x_i^1 \cdots x_i^n)$ and constant parameters $(c_1 \cdots c_m)$ of the experiment, such a model will also include some model parameters $(p_1 \cdots p_o)$. These model parameters include physical constants, like rate constants or equilibrium binding constants whose determination usually is the aim of the experiment. The result of such modeling then is a set R_{Model} of numerical functions, ideally similar to those functions measured experimentally:

$$R_{\text{Model}} = \{f_1^{\text{Model}}(x_1^1 \cdots x_1^n; c_1 \cdots c_m; p_1 \cdots p_o), \cdots, f_k^{\text{Model}}(x_k^1 \cdots x_k^n; c_1 \cdots c_m; p_1 \cdots p_o)\} \quad (2)$$

There are essentially two strategies that can be used to find an optimal description of an experimental result:

In *fitting algorithms*, R_{Model} is used to describe the measurement directly and optimization procedures are used to find the set of model parameters $(p_1 \cdots p_o)$ that best describes the experiment. The methods to find this optimal set will be discussed in Sect. 2.2.

Regularization methods can be exploited in circumstances where the fitting procedure is used to generate a best-fit function rather than values of individual parameters. In general, the fitting is carried out to a so-called describing function, which is of the following form $\int a(r) \cdot R_{\text{Model}}(r)\mathrm{d}r$.

$R_{\text{Model}}(r)$ is a function that depends on the nature of the experiment being undertaken. For example, for kinetic data it may be appropriate to take an exponential function of the form $R = a \cdot \mathrm{e}^{-t/\tau}$ where a (the amplitude of the function) and τ (its characteristic time dependence) are parameters used in the fitting algorithm. For ultracentrifuge data, the $R_{\text{Model}}(r)$ function is generally based on Lamm's differential equation expressing the concentration c as a function of distance (x) and time (t), the independent variables, and s and D are the sedimentation and diffusion coefficients, which are time and position independent parameters.

The other element of the describing function $a(r)$ is an arbitrary function whose form is part of the fitting procedure. Different criteria can be used to decide on the most appropriate form of the function $a(r)$. One that is commonly used is the maximum entropy principle, which selects the solution with the highest informational entropy, or alternatively expresses, the minimal information expressed as $\int (a \ln a)\mathrm{d}r$. The rationale for this criterion is that, according to Occam's razor, the solution with the highest parsimony is to be preferred, a view that can be justified by Bayesian statistical arguments. In operational terms, it is important that the quality of the fit to the data should be balanced against any constraints such as maximum entropy or parsimony requirements.

As an example, one might envisage an exponential function $R = a \cdot \mathrm{e}^{-t/\tau}$. In this

function, both a and τ are the parameters that would be used in a fitting algorithm. However, a describes the height of the function (or its intensity) and τ describes the shape (or its steepness). The describing function is defined as $\int a(\tau) \cdot \mathrm{e}^{-t/\tau} \mathrm{d}\tau$. Since τ will now take all possible values (in practice the range will be limited) only $a(\tau)$ is left as an arbitrary function to be calculated to describe the experiment best. It is interesting to note that this procedure represents a projection of the experimental data onto $a(\tau)$. For example, a sum of three exponential functions $R = \sum_{i=1}^{3} a_i \mathrm{e}^{-t/\tau_i}$ would result in an $a(\tau)$, which has the value of a_i at τ_i and zero everywhere else. Provencher in 1982 had given a full description of this method, and the same is also illustrated later in this chapter (Sect. 3.12) in an analysis of the sedimentation behavior of protein mixtures in the analytical ultracentrifuge that was first described by Schuck in 2000.

2.2 Fitting

Regardless of the strategy used, data evaluation by modeling involves searching for a numerical model that best describes the experimental data according to some suitable criterion. The questions that arise when data are being analyzed quantitatively are essentially the following:

- How well does the model under consideration, which is usually proposed on the basis of previous experience, perform in explaining the experimental data, bearing in mind the accuracy of that data? Is the model satisfactory, or is it necessary to consider alternatives?
- What values of the parameters characterizing the system (rate constants, equilibrium binding constants etc.) are most consistent with the experimental data?
- How accurate are these parameters, and what are the limits of error?

There are several important criteria that all procedures for data analysis should satisfy, chiefly that:

- the experimenter should be able to see the results of the analysis graphically to check whether they are reasonable, and get a feel for the accuracy;
- however the results are manipulated, the original raw data should not be lost;
- there should be no hidden error propagation in the operations, for example, by using transformations involving $1/x$, y^x and similar functions.

The most widely used criterion for judging the quality of description or fit of the data by a model is the least-squares criterion where the sum of squared deviations (SSD) or the root mean square deviation (RMSD) is a minimum for a given parameter set.

When having o parameters and k data points we obtain:

$$SSD(p_1 \cdots p_o) = \sum_{i=1}^{k} \left[f_i^{\text{Model}} - f_i^{\text{measured}} \right]^2$$

$$RMSD(p_1 \cdots p_o) = \sqrt{\frac{1}{k} \sum_{i=1}^{k} \left[f_i^{\text{Model}} - f_i^{\text{measured}} \right]^2} \qquad (3)$$

If the original experimental measurements are normally (Gaussian) distributed around the "true" value, then such a minimization will lead to a parameter set of maximum likelihood. This set of parameters can be interpreted as the result of the experiment.

However, it should be borne in mind that the least-squares criterion is only one

possible criterion of fit, and other criteria exist, which can be expressed by using the following more general definition of model deviation D:

$$D(p_1 \ldots p_o) = \sum_{i=1}^{k} |(f_i^{\text{Model}} - f_i^{\text{measured}})|^n \tag{4}$$

where n can be any real number. In case of $n = 1$ the fitting procedure is called *robust fitting*. Robust fitting can be of advantage whenever the sampled measurements are not distributed in a Gaussian manner, but contain some "outliers" (Johnson, 2000).

The basic concepts underlying the methods of data analysis discussed here are illustrated in Fig. 1. The results of an experiment are data. A model is a description of the processes taking place in the experimental system being observed, which defines a mathematical relationship between the independent variables and the results. The model also defines physical parameters as variables to be fitted or used in regularization. With plausible initial values of the parameters, the mathematical relationships are used to obtain simulated data, which are compared with the experimental data. The values of the parameters are then varied until an optimal fit is obtained of the simulated and experimental results.

In the following sections, the basic concepts of quantitative data analysis are discussed, together with the terms used in Fig. 1.

2.3 Experimental Systems

The system is made up of various components and species. "Components" are molecules that differ in their covalent structures, for example, enzyme, substrate, and product; components can interact to form complexes, for example, an enzyme–substrate complex. "Species" are all the entities present in the solution that differ in either their covalent or

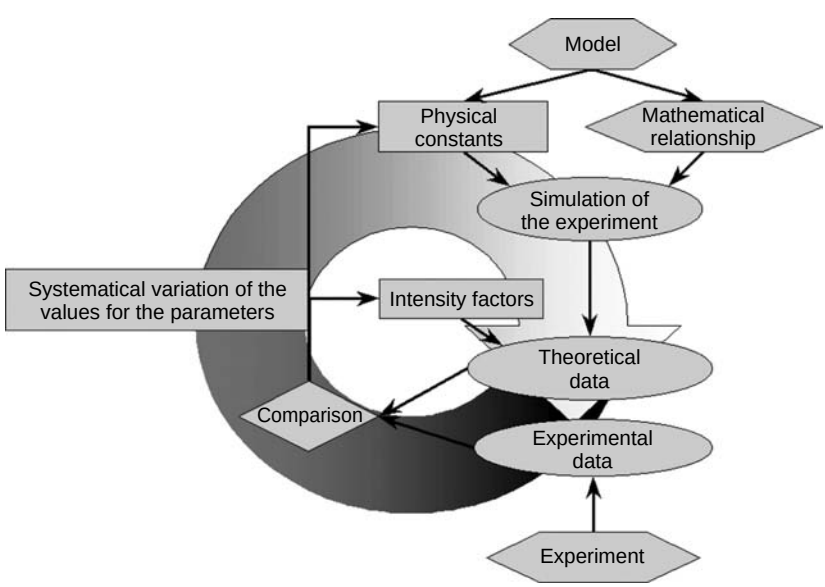

Fig. 1 Schematic illustration of fitting procedures used for quantitative data analysis.

noncovalent structures; this will include components, complexes, and where relevant, different conformations of these.

2.4 Measurement and Signals

In our analysis, we consider a general relationship between the measurement or "signal" and the composition of the solution; the signal is the experimental quantity being measured, which gives information about the processes taking place. It is assumed that the total measured signal is additive in terms of the contributions of all of the species (i) in the solution, and furthermore, that the signal from each species is proportional to its concentration (c_i). Different species contribute in differing extents to the total signal, and the proportionality constant (f_i) is termed *the intensity factor* of the species i. This intensity factor defines the relationship between the concentration of the species and the measured signal (e.g. CPM, absorbance, fluorescence intensity, etc.). It is usual to have to take account of a nonspecific but constant background signal, which we here define as the *baseline* (BL). Thus, the observed signal is given by:

$$S = BL + \sum_i f_i c_i \tag{5}$$

In many cases, this expression simplifies by using the mass conservation laws. So, for example, if we have the species A, B, that form AB in solution, then the signal is given by the expression

$$\begin{aligned} S &= BL + f_A c_A + f_B c_B + f_{AB} c_{AB} \\ &= BL + f_A c_A^0 + f_B c_B^0 + (f_{AB} - f_A - f_B) c_{AB} \\ &= C + f c_{AB} \end{aligned} \tag{6}$$

with only a single intensity factor and a constant term. To simplify the analysis, and to make the numerical analyses more stable, it is important that realistic assumptions are made about which species contribute to the signal, for example, in the case of radioactive detection, only those species that are labeled.

2.5 Models

As described in the mathematical definition section, a model represents an abstraction of the processes that are happening, or could be happening, during the experiment. From this model, we can derive a mathematical relationship between the experimental results and the independent variables. Consider the simple case of the two species A and B forming a complex AB in a time-dependent process:

$$c_{AB} = f(t, c_A, c_B)$$

The experimental results would be the concentrations c_{AB}, and the independent variables would be c_A, c_B, and the time t.

The model provides specific parameters for the fitting process, enabling theoretical data to be evaluated. These can be calculated either analytically or numerically. If the mathematical relationship between the signal (observation) and the parameters is sufficiently simple, it may be possible to obtain analytical solutions and calculate the theoretical signal directly, that is, in the present example to obtain values of c_{AB} knowing the initial concentrations of the concentrations c_A^0, and c_B^0, and the time t.

However, in many cases, the mathematical relationships are not that simple, and analytical solutions may either not be possible in principle, or be too difficult and cumbersome in practice. In such cases,

the theoretical data can be simulated by numerical methods.

A model is, of course, only a working hypothesis, whose validity is judged by its success in accounting for the data. If its performance is not satisfactory, then alternative models should be sought or devised. However, if a model is to be replaced by a more complicated one, then it is important to check that the data really warrant this. More complicated models generally have more parameters, and more parameters will always lead to better fitting of the data. One should be guided here by the Principle of Parsimony, that other things being equal, the preferred model is the simplest one with the fewest parameters.

2.6 Selection of Appropriate Models

The choice of the right model to be used to describe experimental results is one of the trickiest and most interesting tasks in scientific work, and this is a subject that can only be touched on here. As discussed earlier, we are guided by the Principle of Parsimony that in science one should seek the simplest explanation for phenomena. In the present context, it means that we should define models with as few parameters as possible, consistent with obtaining a satisfactory description of the data. This is a sensible approach, because if a simple model fits the data adequately, then so necessarily must more complicated versions of that model. It follows that experimental observations can only serve to rule out models often, but not always, because they are oversimplified; the data can never prove that a model is correct. The question naturally arises at this stage about how one can establish whether a model is successful in accounting for the data or not. There are several criteria for assessing the quality of a model.

- The absolute magnitude of the deviations between the theoretical and experimental data. Does the theoretical curve lie in the region of experimental uncertainty of the data points (taking particular care not to overestimate the accuracy of the data)?
- The direction of the deviations between the theoretical and experimental data. Are the deviations randomly distributed, sometimes above and sometimes below the curve, or are they clustered, above the curve in one region and below in another? If the deviations are not randomly distributed, this indicates that the theoretical curve is not a satisfactory fit to the experimental data. One reason for this is that the model is wrong and is not an adequate description of the situation; another is that systematic errors have been made in carrying out the experiment.
- Whether alternative models are available, which can account for the data more satisfactorily.

A good model should also have predictive power and suggest additional experiments that can be carried out to test the model further.

2.7 Parameters

Depending on the model under consideration, one obtains a set of parameters that establish the relationship between the experimental data and the assumptions underlying the model. It is important to distinguish two kinds of parameters: global and local. This distinction is important when several data sets are being

considered jointly in the analysis; the values of the global parameters must be the same in all cases, whereas, those of the local parameters may vary from one data set to another.

- Global parameters: the values of the global parameters are the same for all of the data sets that are being considered in the analysis. We are dealing here with physical quantities such as binding or rate constants whose values we wish to determine.
- Local parameters: the values of the intensity factors discussed earlier can differ from experiment to experiment. Examples of intensity factors are: radioactivity ($CPM = f_i \cdot c_i$), fluorescence intensity ($signal = f_i \cdot c_i$), absorbance spectroscopy ($OD = f_i \cdot c_i$, where f_i is the extinction coefficient of species i), ELISA ($signal = f_i \cdot c_i$), and so on. Although the precise values of these factors, which are local parameters, often are not particularly interesting in understanding the system, they are needed for the analysis.

2.8 Essential Steps in the Analysis

There are three basic steps in every data analysis (cf. Fig. 1):

- arbitrary initial values of the parameters are introduced into the model to calculate theoretical concentrations for all of the species of interest in the system.
- these theoretical concentrations are combined with initial values for the intensity factors to obtain theoretical values for the measurement or signal.
- the values of the parameters and intensity factors are varied to obtain the best fit of the theoretical values of the signal to the experimental values; the combination of parameters that best fits the data is the result of the analysis.

2.9 Fitting Data by the Method of Least Squares

The classical method for fitting data to theoretical curves is linear regression. This procedure allows the equation of the best straight line fitting a set of experimental data to be calculated directly:

$$y = a + bx$$

$$\text{slope } b = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{n\sum x^2 - \left(\sum x\right)^2} \quad (7)$$

$$y \text{ intercept } a = \frac{\left(\sum x^2\right)\left(\sum y\right) - \left(\sum x\right)\left(\sum xy\right)}{n\sum x^2 - \left(\sum x\right)^2}$$

Until relatively recently, this was the only method that could be used conveniently to fit data by regression. This is the reason why so many classical approaches for evaluating biochemical data depended on linearizing data, sometimes by quite complex transformations. The best-known examples are the use of the Lineweaver–Burk transformation of the Michaelis–Menten model to derive enzyme kinetic data, and of the Scatchard plot to analyze ligand binding equilibria. These linearization procedures are generally no longer recommended, or necessary.

In contrast to the explicit analytical solution of "least-squares fit" used in linear regression, our present treatment of data analysis relies on an iterative optimization, which is a completely different approach: as a result of the operations discussed in the previous section theoretical data are calculated, dependent on the model and

choice of parameters, which can be compared with the experimental results. The deviation between theoretical and experimental data is often expressed as the sum of the errors squared for all the data points, alternatively called the *sum of squared deviations* (SSD):

$$SSD = \sum_i (S_{i,\text{exp}} - S_{i,\text{theo}})^2 \qquad (8)$$

This deviation is now minimized by variation of the parameters. The combination of parameter values that best fit the experimental data – using this deviation as the criterion of best fit is the desired solution of the analysis. This process of finding a solution is termed *iteration* because the solution is located by trying out many possible combinations of parameters; since the equations being fitted are in general nonlinear, the process is more specifically one of the iterative nonlinear least-squares fitting.

The process is essentially as follows. All possible combinations of the parameters (physical constants and intensity factors), of number N, define an $(N+1)$-dimensional error space. Every point in that space has a characteristic value of the sum of the squared deviations or SSD, which thus generates an error surface in $(N+1)$-dimensional space. If, for simplicity, we consider a model with only two parameters, these can be represented on the x- and y-axes, and the value of SSD on the z-axis. The error surface is now simply a surface in conventional three-dimensional space. An even simpler example with one parameter is illustrated in Fig. 2, where the parameter is shown on the x-axis and the SSD is on the y-axis. The task in the fitting procedure is to locate the minimum value in the SSD curve (region A in Fig. 2). It is impracticable to try out all possible values of the combined set of parameters, particularly when there are many of them. The procedure adopted in most computer programs is, starting from initial values of the parameters (provided by the user) calculations are made of the slope (or derivative) of the error surface in $(N+1)$-dimensional space. This is done by making a small variation in each of the parameters in turn, and calculating the SSD. The program then locates the region where the slope is steepest (downwards) and it alters the parameters by a small step in that direction to generate a new set of parameters, which fit the data better. From this new set of parameters, the program repeats the operation in a second iterative cycle to locate the direction of steepest descent, and hence a new set of parameters.

This procedure depends on certain features that merit comment.

- The step length in the iteration is critical: if it is too short, then the process of

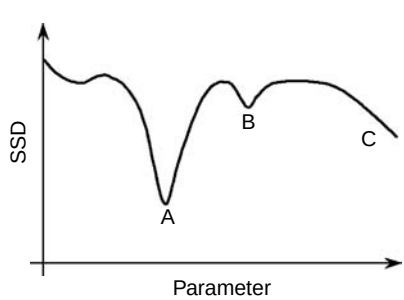

Fig. 2 Two-dimensional representation of an error surface. Region A is the location of the global minimum, region B is a local minimum, and region C represents an area where the model is no longer valid and the slope of the error surface is directed away from the minimum.

locating the minimum takes too long, whereas, if the step length is too long the algorithm used in the program can miss the target area, and thus never locate the minimum. The SOLVER algorithm used in Excel selects the step length automatically depending on the slope of the error surface and the result of the previous round of iteration.

- The "result" located by the program can be a local minimum (e.g. region B in Fig. 2). Locating the global minimum is often not straightforward, particularly when the error surface is complex, and the program can find itself trapped in a local minimum. The best means of avoiding this, or at least detecting when it is happening, is to begin the iteration process from different initial parameter estimates, and check whether the same solution is found in every case. If this does not happen, the solution with the lowest SSD corresponds to the best solution, although it should be noted that in some cases alternative solutions might be equally good in terms of their *SSD* values, bearing in mind the accuracy of the experimental data.
- To avoid local minima, most algorithms also test randomly selected points in the error surface. The extent to which a program carries out these tests determines the speed of locating the minimum and the tendency of the algorithm to become trapped in local minima.

All models have limits to their region of validity; for example, negative values of rate or binding constants do not correspond to physically meaningful situations. In such regions, mathematical errors will arise, such as attempting to find the square root of a negative number, even though all of the equations have been correctly programmed.

The slope of the error surface can lead the iterations into regions that are remote from the minimum. This situation can readily lead to failure to locate the minimum when the initial parameter estimates are not very good. To remedy this, a fresh set of initial estimates should be selected, which fit the data better. In Fig. 2, for example, it would be difficult to locate the minimum if the program started in region C since the slope in the error surface is pointing in the wrong direction.

The usual criterion of "best fit" is the sum of errors squared (the *SSD* discussed earlier) rather than the absolute magnitude of the errors. This procedure is mathematically justified when the errors in the data follow the Gaussian (or normal) distribution. Under these conditions, the error distribution function is given by the following equation in which x is the measurement, μ the mean, and σ the standard deviation:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{9}$$

When the data are distributed according to this function, the frequency of occurrence of data falls according to the square of the deviation. In practice, the sum of error square (*SSD*) criterion is also used in cases where it has not been explicitly established that the errors are normally distributed, and it appears to function quite well.

2.10 Global Fitting of Multiple Data Sets

If different sets of experiments have been carried out under circumstances in which the observations depend on a common set of parameters, then it is sensible to attempt a global fitting of the data sets to

obtain best estimates of the parameters. It is important here to distinguish clearly between the global and local parameters discussed earlier (Sect. 2.7). The global parameters are valid for all of the data sets and are fitted to all of the data, whereas the local parameters may assume different values for the various data sets. For example, if one were investigating a thermodynamic equilibrium, and monitoring the process using radioactive detection, the value of the equilibrium constant must be the same under the same conditions, whereas the specific activities of the reaction participants could well be different. In global data fitting, it is particularly important to keep the number of parameters as small as possible. There are two reasons for this. First, the general consideration that following the Principle of Parsimony one should seek to account for experimental data using the smallest number of variables. Second, that iterative fitting of the data becomes much more difficult (in fact, exponentially so) as the number of variables increases; the process becomes much slower, and there is an increasing risk that local minima will interfere with the fitting. To keep the number of parameters as small as possible, it is important to check, in particular, whether all of the local parameters are needed. For example, in the general case, it is assumed that all of the reaction participants contribute to the experimental signal or measurement, but if this is not in fact true, then it is better to set the intensity factors of as many species as possible to 0, and only allow the minimum necessary number of species to contribute to the signal. For example, in studies based on fluorescence detection, only species containing a fluorophore need to be assigned intensity factors.

One difficulty that can arise in global data analysis is that the signal intensities of different data sets can be very different. If the data are treated equally, this can lead to the situation that data sets or curves with high intensities completely dominate those with lower intensities, simply because their error squared parameters (SSD) are so much larger. The most effective way of dealing with this situation is to weight the SSDs of the different data sets or curves by a suitable factor, for example, by the mean value of the data set, or by the relevant intensity factor. It should be emphasized that weighting factors must never be treated as variables in the fitting process.

2.11 Introduction to Error Estimation

One of the most difficult tasks in day-to-day scientific activity is making reliable estimates of the errors and uncertainties in the data. How reliable are my data? How accurate are the parameters calculated from them? Can I, or should I, exclude particular models for explaining my data? These are examples of the sort of questions that need to be asked. We have already discussed the question of judging how well models perform in accounting for data; we now turn to the question of assessing accuracy, on the basis that the model used is an appropriate one. It should be noted that we are dealing here with statistical errors and the treatment of outliers; systematic errors cannot be detected by these approaches.

Statistical analysis of repeated measurements. If a very large number of data are available, then it is sensible to consider carrying out a rigorous statistical analysis. The simplest procedure is to do many replicates

of the same experiment (or series of experiments, if more than one data set is needed for the analysis) and then analyze them independently. This is a very good way of assessing the error range (determined as standard deviations, maximal range etc.) of the individual parameters; the problem is the amount of work involved.

The accuracy of individual measurements. We are concerned here with the problem of assessing accuracy when the number of available data is limited. One means of gauging error is to remove individual data points from the fitting process to get a feel for the "robustness" of the data. In effect, what this process does is to analyze the data on the assumption that the single experiment removed had not been carried out. It is possible in this way to assess how reliable the data are, and specifically to determine whether the outcome was highly dependent on the single result, implying that one would need to be very sure about it. This form of analysis is straightforward and revealing, and it ought to be a part of every data evaluation.

Analysis of error surfaces. To conclude this section, we consider a more quantitative approach to error estimation. The first step is to estimate the accuracy of the individual data points; this can either be done by analysis of the variability of replicate measurements, or from the variation of the fitted result. From that, one can assess the shape of the error surface in the region of the minimum. The procedure is straightforward: the square root of the error, defined as the SSD, is taken as a measure of the quality of the fit. A maximum allowed error is defined, which depends on the reliability of the individual points, for example, 30% more than with the best fit, if the points are scattered by about 30%. Then each variable (not the SSD as before) is minimized and also maximized. A further condition is imposed so that the sum of errors squared (SSD) should not increase by more than the fraction defined earlier. This method allows good estimates to be made of the different accuracy of the component variables, and also enables accuracy to be estimated reliably even in complex analyses. Finally, it reveals whether parameters are correlated. This is an important matter since it happens often, and in some extreme cases where parameters are tightly correlated, it leads to situations where individual constants are effectively not defined at all, merely their products or quotients. Correlations can also occur between global and local parameters. This issue is illustrated in Fig. 3, where a set of initial slopes of an enzyme reaction are analyzed using the Michaelis–Menten model (see Sect. 3.2, Eq. 12). Since K_m is defined using v_{max}, depending on the shape of the curve any change of v_{max} will cause a certain change of K_m as well.

2.12 Introduction to Numerical Integration

Kinetic processes can be described by differential equations; for example,

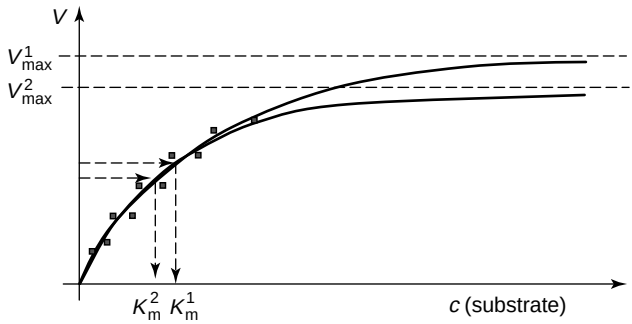

Fig. 3 Illustration of the origin of correlation of parameters with the example of a Michaelis–Menten analysis of steady state kinetics data. In this example, there is a strong correlation of k_{cat} and K_m.

for a reversible bimolecular association reaction:

$$A + B \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} AB$$

is described by

$$\frac{\partial c_{AB}}{\partial t} = k_1 c_A c_B - k_{-1} c_{AB} \qquad (10)$$

This equation defines directly the change in concentration of the species AB with given concentrations of the reactants A and B, and the product AB. This is a differential equation whose solution is an expression of the form $c_{AB} = f(t, c^0_A, c^0_B)$. The solution involves a process of integration, which is often difficult, and sometimes impossible, at least analytically. In such cases, numerical integration can be used to simulate the time-dependent variation of c_{AB} in an experiment, enabling theoretical data to be obtained even for complex systems.

The procedure for numerical integration is as follows. Initial conditions are first selected: c^0_A, c^0_B, c^0_{AB} and from this initial state the concentrations of the three component species are altered stepwise using "fluxes" defined from the differential equation given earlier, with a finite time increment Δt.

Two different fluxes exist:

F_1:"association," A + B

$\longrightarrow$ AB for which $F_1 = k_1 c_A c_B \Delta t$

F_{-1}:"dissociation," AB

$\longrightarrow$ A+B for which $F_{-1} = k_{-1} c_{AB} \Delta t$

(11)

The concentration changes are defined in terms of these fluxes as follows:

$$\Delta c_A = -F_1 + F_{-1}$$

$$\Delta c_B = -F_1 + F_{-1}$$

$$\Delta c_{AB} = -F_{-1} + F_1$$

from which new concentrations are obtained using the following general expression, where $c_{i,\text{old}}$ is the "old" concentration of the species (i) before the incremental change Δc_i:

$$c_i = c_{i,\text{old}} + \Delta c_i$$

The formulae given in Eq. (11) are prototypes for bimolecular (F_1) and monomolecular (F_{-1}) elementary reactions respectively. By combining these

prototype equations, kinetic schemes of any desired complexity can be described and analyzed. The time interval is a critical parameter in the integration process, if it is chosen too short, the process takes too long, if it is too long, the process is inaccurate.

3 Applications

3.1 Linear Regression

Situations arise very often where data need to be fitted to linear equations. Linear regression is one of the classical procedures in general regression analysis, and before the advent of accessible nonlinear fitting methods it was the only one that could be readily used. For n data pairs in the form (x, y) where y is a function of x, the linear equation of the form $y = a + bx$ that minimizes the sum of errors squared (SSD) is given by Eq. (7).

3.2 Michaelis–Menten Kinetics

The Michaelis–Menten model shown below is the simplest mechanism for describing the kinetics of enzyme catalyzed reactions:

$$\mathrm{S} + \mathrm{E} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \longrightarrow \mathrm{E} + \mathrm{P}$$

According to this mechanism, the rate constant for the reaction k (dimension: time^{-1}) depends on two parameters: K_m and k_cat. In the simple mechanism shown earlier and with the assumption that ES is in a steady state, $K_\mathrm{m} = (k_2 + k_{-1})/k_1$ and $k_\mathrm{cat} = k_2$. The dimensions of K_m and k_cat are concentration and (time)$^{-1}$ respectively. The rate of the reaction v (dimension: concentration/time) is given by the expression $v = kc_{\mathrm{E,total}}$, and v_max is equal to $k_\mathrm{cat}c_{\mathrm{E,total}}$. The dependence of the reaction rate constant (k) on substrate concentration is given by Eq. (12), from which it can be seen that the K_m value is the concentration of substrate that gives half of the maximum rate k_cat.

$$k(c_\mathrm{S}) = k_\mathrm{cat}\frac{c_\mathrm{S}}{c_\mathrm{S} + K_\mathrm{m}} \tag{12}$$

To evaluate K_m and k_cat, the rate of reaction is measured as a function of substrate concentration and the two kinetic parameters are determined using Eq. (12). The classical method of doing this is by fitting the data to a linearized form of Eq. (12) such as the Lineweaver–Burk plot shown in Eq. (13):

$$\frac{1}{k(c_\mathrm{s})} = \frac{1}{k_\mathrm{cat}} + \frac{K_\mathrm{m}}{k_\mathrm{cat}}\frac{1}{c_\mathrm{s}} \tag{13}$$

From this it follows that a plot of $1/k$ against $1/c_S$ should give a straight line with an x-intercept of -$1/K_\mathrm{m}$ and a y-intercept of $1/k_\mathrm{cat}$. The Lineweaver–Burk analysis illustrates very clearly the sort of problem that can arise when dealing with linearized data. An assumption that underlies simple linear regression following the procedure discussed in the previous section is that all of the data points have the same error, or specifically, standard deviation. This assumption is no longer valid when the data are transformed as is shown in the following diagrams.

Figure 4(a) illustrates a series of measurements where all the data have the same error; in Fig. 4(b), the same data are shown after transformation for Lineweaver–Burk analysis. It can be seen that the data points at low concentration (i.e. at high values of $1/k$ and $1/c_\mathrm{s}$) have a much higher error than the other points, and the situation

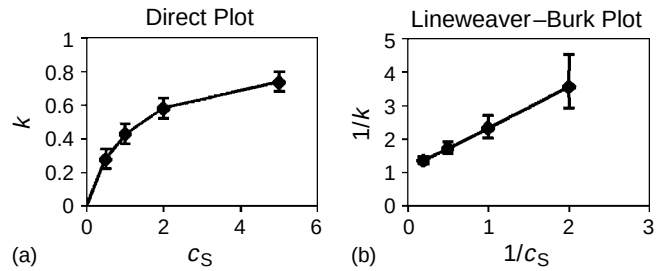

Fig. 4 Error propagation in the direct analysis and Lineweaver–Burk analysis of Michaelis–Menten kinetics.

is made worse because these inaccurate points are also the ones that exert the most leverage on the linear regression, and hence on the derived kinetic parameters. A very comprehensive analysis of the kinetics of enzyme catalyzed reactions has been presented by Bisswanger in 2000.

3.3 Analysis of Circular Dichroism Experiments to Determine the Secondary Structure Composition of Proteins

Circular dichroism (CD) is the spectroscopic method of choice to investigate the secondary structure of proteins. The CD spectrum of a protein is given by the sum of the contributions of its α-helical, β-sheet, β-turn, and disordered ("random") structural elements. The results of CD experiments can be analyzed very simply to provide information about the fraction of α-helical, β-sheet, β-turn, and random regions in a protein. The analysis depends on comparison of the CD spectrum with reference spectra for the respective secondary structure elements, which are derived from the CD spectra of a large set of proteins of known structure. Then, the CD signal is given by:

$$\theta(\lambda) = c_P(f_\alpha \varepsilon_\alpha^{ref} + f_\beta \varepsilon_\beta^{ref} + f_t \varepsilon_t^{ref} + f_r \varepsilon_r^{ref}) \quad (14)$$

where c_P is the concentration of the protein, f_α, f_β, f_t, and f_r denote the fractions of the protein in α-helical, β-sheet, turn and random conformation, and ε_i^{ref} is the extinction coefficients of the respective secondary structure element taken from the reference spectra. For analysis, $\theta(\lambda)^{theo}$ is fitted to $\theta(\lambda)^{exp}$ by variation of f_α, f_β, f_t, and f_r.

3.4 Dissociation Kinetics

Dissociation reactions of the general form $AB \rightarrow A + B$ are monomolecular processes, where the rate of decay of the complex is proportional to its concentration. The concentration dependence of c_{AB} is given by the following differential equation:

$$\frac{dc_{AB}}{dt} = k_{-1} c_{AB}$$

which on integration yields Eq. (15), where $c_{AB}(t)$ is the concentration of complex at any time t, and c_{AB}^0 is the initial concentration at time $t = 0$:

$$c_{AB}(t) = c_{AB}^0 e^{-k_{-1}t} \quad (15)$$

Analysis of dissociation processes yields values for the rate constant k_{-1}, whose dimensions are $(time)^{-1}$. This rate constant is related to the lifetime (τ) of the complex AB by the expression $\tau = (k_{-1})^{-1}$, and

to the half-life ($t_{1/2}$) by the expression $t_{1/2} = (\ln 2/k_{-1})$.

3.5 Binding Data

The equilibrium constant for a simple bimolecular association process

$$A + B \overset{K_{Ass}}{\rightleftharpoons} AB$$

is defined by the expression:

$$K_{Ass} = \frac{c_{AB}}{c_A c_B} \tag{16}$$

This equilibrium constant is expressed as the association constant, which has dimensions (*concentration*)$^{-1}$, in molar terms M^{-1}. The dissociation constant K_{Diss} is the reciprocal of K_{Ass} and has dimensions of concentration (M). The objective of the following derivation is to obtain an equation of the form $c_{AB} = f(c_{A,tot}, c_{B,tot}, K_{Ass})$, where $c_{i,tot}$ is the total or stoichiometric concentration of the component i (which is known), in contrast to the quantity c_i in Eq. (16), which is the free concentration of the species in solution, which is not known. An equation of this form will enable us to calculate the theoretical data.

Using the conservation conditions: $c_{A,tot} = c_A + c_{AB}$ and $c_{B,tot} = c_B + c_{AB}$ Eq. (16) can be written in the form:

$$K_{Ass} = \frac{c_{AB}}{[(c_{A,tot} - c_{AB})(c_{B,tot} - c_{AB})]} \tag{17}$$

The only unknown in this equation is the term c_{AB}. Expanding and rearranging Eq. (17) yields the following quadratic equation:

$$c_{AB}^2 - \left(c_{A,tot} + c_{B,tot} + \frac{1}{K_{Ass}}\right) c_{AB} + c_{A,tot} c_{B,tot} = 0$$

The solutions of a quadratic equation of the general form $x^2 + px + q = 0$ are given by the two roots x_1 and x_2:

$$x_{1,2} = -\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q}$$

In the present case, only the negative square root term is physically meaningful, so the concentration of AB is given by the following equation:

$$c_{AB} = \frac{c_{A,tot} + c_{B,tot} + 1/K_{Ass}}{2} - \sqrt{\left(\frac{c_{A,tot} + c_{B,tot} + 1/K_{Ass}}{2}\right)^2 - c_{A,tot} c_{B,tot}} \tag{18}$$

To determine values of K_{Ass} binding data are needed, where the total concentrations of either A or B are comparable in magnitude to $1/K_{Ass}$.

3.6 Independent Identical Binding Sites

The earlier-mentioned model and equations have to be modified if one of the species (say A) has several binding sites for the other species B. If the binding sites are independent and do not interact, then binding to each site on A can be described by Eq. (19). Taking all of the binding sites into account yields a hyperbolic binding curve whose binding equation only differs from Eq. (18), where for every molecule of A, n binding sites exit such that the total concentration of binding sites is $nc_{A,tot}$:

$$c_{B,bound} = \frac{nc_{A,tot} + c_{B,tot} + 1/K_{Ass}}{2} - \sqrt{\left(\frac{nc_{A,tot} + c_{B,tot} + 1/K_{Ass}}{2}\right)^2 - nc_{A,tot} c_{B,tot}} \tag{19}$$

To obtain accurate estimates of the number of binding sites (n), binding experiments (usually titrations) need to be performed under conditions where the total concentration of A is relatively high, specifically that $c_{A,tot} \gg 1/K_{Ass}$; these conditions define a "stoichiometric titration" where, effectively all of the B added is bound until the sites on A are saturated (see Fig. 5). Titrations under these conditions are insensitive to the value of the association constant, so to obtain reliable estimates of K_{Ass}, data are needed from titrations at much lower concentrations, where $c_{A,tot} \leq 1/K_{Ass}$. It should be clear from this discussion that it is not easy to evaluate both n and K_{Ass} accurately, and it is usually necessary to do a global analysis of several data sets, obtained under different concentration conditions.

3.6.1 Analysis of Simple Binding Data

The equation for n identical, noninteracting binding sites (Eq. 19) is in principle soluble, although the solution is not straightforward. When binding is more complex and the sites are of different affinity and interacting, then analytical solutions cannot be obtained. However, analysis of the binding can be simplified by carrying out experiments under conditions where one of the interacting partners (say A) is present at a much lower concentration than the other. The concentration of the partner in excess (B) is varied, and the proportion of available binding sites on A, which are occupied ($c_{AB}/c_{A,tot}$) is measured. The simplification in the analysis arises because the free concentration of B can be taken to be the same as the total concentration (since $c_{AB} \ll c_B$). Equation (16) can be simplified considerably yielding, after inserting the conservation condition for A and rearrangement:

$$\frac{c_{AB}}{c_{A,tot}} = \frac{c_B K_{Ass}}{1 + c_B K_{Ass}} \tag{20}$$

3.6.2 Independent Nonidentical Binding Sites

Consider a macromolecule A that can bind several molecules of B. In the simplest case, where A possesses two binding sites for B, there are four possible species, A, AB, BA, and BAB, whose concentrations depend on four binding constants:

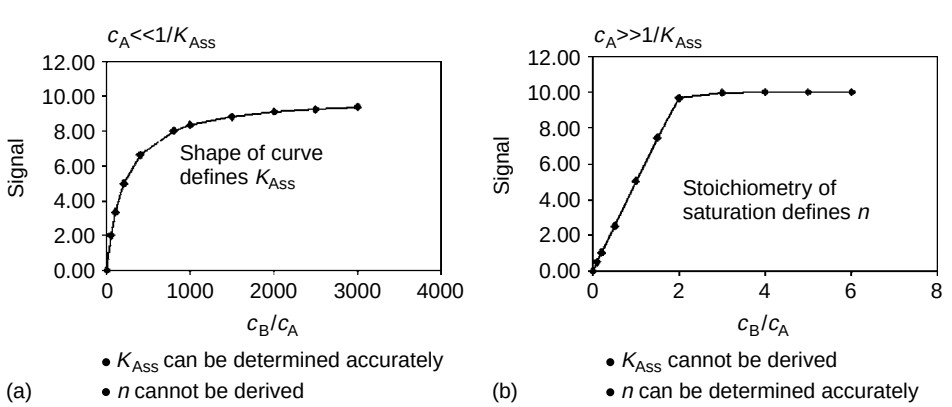

Fig. 5 Typical results of a normal and stoichiometric titration binding analysis.

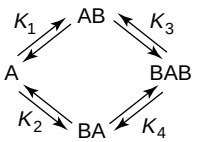

(21)

Under conditions where $c_{A,tot} \ll c_{B,tot}$, terms involving the total concentration of A do not occur in the analysis (as mentioned earlier), and it is therefore not possible to use Eq. (20) to analyze the stoichiometry of the binding equilibrium. However, even under these experimental conditions, it is possible to obtain information about the number of binding sites, provided the binding constants of the two processes are sufficiently different in magnitude.

Information about the minimum number of binding sites for B on the macromolecular species A can also be obtained if a signal can be measured, which specifically monitors the concentration of A fully saturated with B (BAB in our scheme). For example, the enzyme DNA polymerase has two binding sites for metal ions, and both need to be occupied for the enzyme to be active. If it is assumed that the two sites are independent, and hence $K_1 = K_4$ and $K_2 = K_3$ in Eq. (21), the following expression can be derived for the occupancy of the two sites (designated 1 & 2):

$$\theta_1 = \frac{c_{AB}}{c_{A,tot}} = \frac{c_B K_1}{1 + c_B K_1}$$

$$\theta_2 = \frac{c_{BA}}{c_{A,tot}} = \frac{c_B K_2}{1 + c_B K_2} \qquad (22)$$

and the proportion of A where both sites are occupied is given by:

$$\theta_{1,2} = \frac{c_{BAB}}{c_{A,tot}} = \theta_1 \theta_2$$

In the special case of identical binding sites ($K_{Ass,1} = K_{Ass,2}$), the dependence of $\theta_{1,2}$ on the total concentration of A ($c_{A,tot}$) is weakly sigmoidal at low concentrations of B, and not hyperbolic; this is a direct indication that A can bind more than one B. The total concentration of bound ligand ($= \theta_1 + \theta_2$) follows a hyperbolic dependence, as expected since the sites are independent.

3.6.3 Cooperative Binding

In the previous section, we discussed the case where the various binding sites were noninteracting; in this section we consider the other limiting case where A is either free, or fully occupied by B as the species AB_n, and the intermediate states $AB, AB_2, \ldots, AB_{n-2}$, and AB_{n-1} are not populated. This behavior arises because of positive interactions between the sites resulting in cooperative binding; according to Eq. (21), cooperative binding occurs when $K_3 \gg K_1$ and $K_4 \gg K_2$. The model considered here represents "all or none" behavior, which is not just a theoretical model but also one that does actually occur with biopolymers.

In cooperative binding following the "all or none" model

$$A + nB \rightleftharpoons AB_n$$

the association constant is defined by the expression:

$$K_{Ass}^{app} = \frac{c_{AB_n}}{c_A (c_B)^n}$$

Introducing the conservation condition for A with the further assumption that $c_{AB} \ll c_B$ yields the following equation:

$$\frac{c_{AB_n}}{c_{A,tot}} = \frac{(c_{B,tot})^n K_{Ass}^{app}}{1 + (c_{B,tot})^n K_{Ass}^{app}} \qquad (23)$$

This equation describes a sigmoidal binding curve, where the degree of sigmoidal behavior depends on the magnitude of n. The intrinsic binding constant

of B for A (K_{Ass}) can be determined from the apparent binding constant (K_{Ass}^{app}) using the following relationship:

$$K_{Ass}^{app} = (K_{Ass})^n$$

3.7 pH Dependence of Enzyme Catalyzed Reactions

The rate of an enzyme catalyzed reaction is not only dependent on the concentrations of enzyme and substrate, but also on the conditions of the reaction. An important parameter affecting rate is the pH, defined as $-\log_{10} c(H^+)$, and it is very common that enzymes have a pH optimum. The pH can have several effects: (1) protons may participate in the catalytic reaction itself; (2) the protonation state of substrates and cosubstrates may alter, with consequent effects on rate; (3) the protonation state of the enzyme itself may alter. In our example here, we deal with the last case. Proteins contain many groups that can undergo protonation–deprotonation reactions, including the N-terminal amino and C-terminal carboxyl groups, and the side chains of the following amino acids: Asp, Glu, His, Cys, Tyr, Lys, and Arg. Protonation/deprotonation reactions are examples of ligand binding equilibria, which are simplified by the experimental approach of using a buffered solution such that the concentration of H^+ in the solution remains constant throughout the experiment. The state of protonation of a group is conveniently represented by its pK_a value, which is the negative decadic logarithm of the dissociation constant for the protonation reaction:

$$K_a = \frac{c(H^+)c(A^-)}{c(HA)}$$

$$pK_a = -\log_{10} K_a$$

The proportions of a group A in the protonated $\theta(H)$ and deprotonated $\theta(-)$ states can be evaluated as follows:

$$\theta(-) = \frac{A^-}{A_{tot}} = \frac{K_a}{c(H^+) + K_a} = \frac{10^{-pK_a}}{10^{-pH} + 10^{-pK_a}}$$

$$\theta(H) = 1 - \theta(-) \qquad (24)$$

The following questions are important for analyzing the pH dependence of enzyme catalyzed reactions:

- how many protonation reactions participate?
- which protonation state must the pH-sensitive groups on the enzyme be in?
- what are the pK_a values of these groups?

We consider a general model to analyze the protonation equilibria. If the enzyme possesses n groups that can participate in protonation–deprotonation equilibria, then in principle 2^n different species can be formed. For example, if $n = 3$, all three groups can be protonated (HHH), two (HH–, H–H, and –HH), one (H–, –H–, and –H) or none (–). The probability (P) of occurrence of these species, and hence their relative concentrations, depends on the product of the probabilities that each individual group is in a particular state:

$$P(HHH) = P(\text{group 1 is protonated}) \times P(\text{group 2 is protonated}) \times P(\text{group 3 is protonated})$$

$$P(HH-) = P(\text{group 1 is protonated}) \times P(\text{group 2 is protonated}) \times P(\text{group 3 is unprotonated}) \text{ and so on.}$$

The probability of a group being in a particular protonation state is given by

Eq. (24), and combination of these probabilities multiplied by the total concentration of enzyme yields the concentrations of the different species.

The turnover rate is used as an "effect" or signal to monitor the protonation, and thus the observed rates can be used to analyze the thermodynamic protonation equilibria. In the general case, every species would be assigned an intensity factor, and the signal (observed rate of the reaction) would be the sum of all of these factors. For many analyses, one makes the simplifying assumption that only one species is catalytically active.

3.8 Pre-steady State Kinetics

Enzyme reactions proceed, in general, via several intermediate states. A simple model incorporating multiple states is shown below: enzyme and substrate associate to form an enzyme–substrate complex, which undergoes a conformational change to ES# before breaking down into enzyme and product.

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \underset{k_{-2}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES\#} \longrightarrow \mathrm{E} + \mathrm{P}$$

Since the concentrations of all the intermediate states are constant under steady state conditions, all of these states can, at least formally, be incorporated into a single kinetic intermediate state. It follows that under steady state conditions, kinetic data can provide no information about the existence and kinetic properties of intermediate enzyme–substrate complexes. An understanding of the mechanism of an enzyme catalyzed reaction needs information about these intermediate states, which is therefore usually obtained from kinetic studies before steady state has been established, usually by rapid reaction methods. Comprehensive coverage of the techniques and methods of analysis of pre-steady state kinetics is beyond the scope of this chapter, but we discuss here methods for analyzing simple exponential processes. Two approaches are used. In the first, the observed signal $S(t)$ is fitted to an exponential function of the following form:

$$S(t) = A\mathrm{e}^{-\mathrm{t}}/\tau, \text{ for decreasing signals.}$$
$$S(t) = A[1 - \mathrm{e}^{-\mathrm{t}/\tau}] \text{ for increasing signals.} \tag{25}$$

A is the amplitude of the reaction and τ the time constant, with the dimension of (time). If the kinetic mechanism of the observed process is known, then rate constants can be derived from the time constant. For example, for a simple dissociation process, the rate constant (k_{-1}) is given by $1/\tau$. In this case, the value of τ is independent of reactant concentration.

If both forward and back reactions can take place, then $1/\tau$ depends on both k_1 and k_{-1}. In the special case when the concentration of S is much greater than that of E, then the association rate constant is given by the equation $1/\tau = k_1 c_\mathrm{S} + k_{-1}$. Values of the two rate constants can be determined from the dependence of τ on the substrate concentration c_S; a linear regression of $1/\tau$ versus c_S yields k_1 as the slope of the plot and k_{-1} as the y-intercept. For this analysis to be valid, it is important to be sure that the observed reaction represents a single exponential process. If the reaction involves more than one exponential phase, then more complex models need to be considered, since the minimal number of reaction steps is given by the number of exponential processes.

This method of analysis has several disadvantages, one of which is that intermediate parameters (τ) are evaluated from the data, which then form the basis for global fitting of the data; consequently, the global fitting is not carried out on the raw data directly. A second drawback is that the predictive power of this analysis as regards mechanism is rather limited.

A more powerful method for the analysis of such systems is to use direct integration of the differential equations that describe the mechanism of the reaction. An advantage of this procedure is that the fitting is carried out directly to the raw data. Since in most cases, particularly those of any kinetic complexity, the resulting systems of differential equations cannot be integrated analytically, numerical integration has to be used, which is not a problem in the currently given computer performance.

3.9 Association Kinetics Analyzed by Numerical Integration

The rate of a bimolecular association process $A + B \rightarrow AB$ is given by Eq. (26):

$$\frac{dc_{AB}}{dt} = k_1 c_A c_B \tag{26}$$

The rate constants for bimolecular association reactions have dimensions $(\text{concentration})^{-1}\ (\text{time})^{-1}$. Although this differential equation has a very simple form, it does not have a very straightforward analytical solution. For this reason we use numerical integration methods to simulate theoretical data. This is a general approach that can be used to obtain solutions of complex kinetic processes. Although it is always easy to formulate differential equations like Eq. (26), which express the time dependence of the various concentrations, solving the equations is another matter; it is often impossible to obtain explicit analytical solutions of the form $c = f(t)$, from which concentrations of the reaction participants can be directly determined. What can, however, be evaluated is the concentration change (or "flux") for a species in a given time interval Δt under given conditions:

$$F_1 = k_1 c_A(t) c_B(t) \Delta t$$

The solution can be obtained by proceeding stepwise (using small values of Δt) and calculating $c_{AB}(t)$ using the expression:

$$c_{AB}(t + \Delta t) = c_{AB}(t) + F_1$$

This procedure is called *numerical integration*.

If we consider the following equilibrium:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES$$

There are two different fluxes:

$$E + S \longrightarrow ES, \text{ where } F_1 = k_1 c_E c_S \Delta t$$

$$ES \longrightarrow E + S, \text{ where } F_{-1} = k_{-1} c_{ES} \Delta t$$

The concentration changes are defined as follows:

$$\Delta c_E = -F_1 + F_{-1}$$

$$\Delta c_S = -F_1 + F_{-1}$$

$$\Delta c_{ES} = -F_{-1} + F_1$$

from which new concentrations can be derived using the following expression where, as before, $c_{i,\text{old}}$ is the old concentration of the species i before the new increment Δc_i:

$$c_i = c_{i,\text{old}} + \Delta c_i$$

3.10 Surface Binding Reactions

With the development and widespread use of the surface plasmon resonance (SPR) approach to study biomolecular interactions, analysis of binding reactions at surfaces is becoming increasingly important. Assuming that only one class of binding sites is present, one only has to consider that the surface has a certain binding capacity C. Since in a typical SPR experiment, the analyte is flushed along the surface in a continuous buffer flow, the concentration of analyte (c_A) available for binding is constant. Therefore, the flux (F_1) of the analyte to the surface is given by the following expression, where k_1 is the binding rate constant and Δt is the time increment used for numerical integration:

$$F_1 = k_1 c_A (C - c_{A,\text{bound}}) \Delta t \qquad (27)$$

Release of bound analyte is given by a second flux, just as described in Eq. (27):

$$F_{-1} = k_{-1} c_{A,\text{bound}} \Delta t$$

3.11 Analysis of Competition Experiments

Competition experiments are widely used in the biosciences, particularly in studies of binding interactions. A simple example is shown in the following:

$$\text{AB} + \text{C} \overset{K_1}{\rightleftharpoons} \text{A} + \text{B} + \text{C} \overset{K_2}{\rightleftharpoons} \text{AC} + \text{B} \qquad (28)$$

In this example, the equilibrium between A, B, and AB is affected by the addition of C. The popularity of the competition technique is due to the fact that it can be used to investigate interactions (in this case the binding of A + C = AC) without having to detect the participating species (free C and the complex AC). The method relies on using one interaction (here A + B = AB) as a reporter to monitor the other. This assumes, of course, that a suitable signal is available to follow the formation of AB. Figure 6 illustrates the formation of AB, and the effect of adding C to a system containing A, B, and AB: on addition of C the species AC is formed at the expense of AB whose concentration falls, with a concomitant decrease in the observed signal.

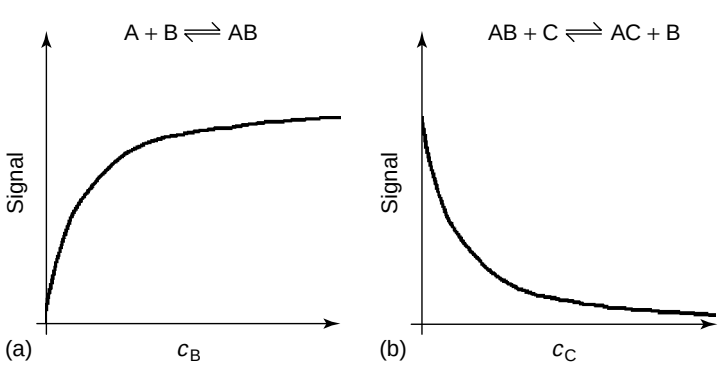

Fig. 6 Indirect analysis of molecular interaction of A and C by competition of the interaction of A with B.

It is also a desirable feature of competition experiments that they allow more precise comparison of the binding of different species (in this case B and C) to a common target (A) than is possible in separate binding experiments. It is also possible to use this approach with a single experimental set-up to test the binding of many different ligands to A, on the basis that all these ligands compete with B for the same binding site.

The analysis of coupled equilibria is the most complex problem that is considered in this chapter. It may seem surprising that such apparently straightforward systems like those shown in Eqs. (21) and (28) should present such great difficulties in analysis, the more so because it is a trivial matter to calculate the equilibrium constants, if the concentrations of the various species are known. However, that situation arises very rarely for several reasons:

- in most investigations, only some of the species can be detected;
- it is usually the case that only one "signal" is measured, whose dependence on the concentration of reaction participants may be complex, and must be derived from the model;
- experimental error.

Proceeding as we have done before with complex systems, we calculate theoretical data to deal with these systems. Since straightforward analytical solutions are not available, even for such simple cases as Eq. (28), numerical methods are used to simulate solutions of the equilibria. From an arbitrary initial state (e.g. only free A, B, and C present) we calculate for both isolated equilibria the final equilibrium concentrations, and the concentration changes needed to establish equilibrium. We then use these fluxes. Since, in every step, there are several fluxes, whose coupling is not taken into account, this approximate procedure takes us towards the equilibrium, but not directly to it. So the equilibrium has to be located iteratively.

We may write:

$$K_1 = \frac{c_{\mathrm{AB,Equil.}}}{c_{\mathrm{A,Equil.}} c_{\mathrm{B,Equil.}}}$$

or:

$$c_{\mathrm{AB,Equil.}} = K_1 c_{\mathrm{A,Equil.}} c_{\mathrm{B,Equil.}}$$

where, $c_{\mathrm{i,Equil.}}$ is the equilibrium concentration of i.

We are looking for the flux F, necessary to take us from the present concentrations to the equilibrium concentrations:

$$c_{\mathrm{AB,Equil.}} = c_{\mathrm{AB}} + F$$
$$c_{\mathrm{A,Equil.}} = c_{\mathrm{A}} - F$$
$$c_{\mathrm{B,Equil.}} = c_{\mathrm{B}} - F$$

Thus, we have:

$$(c_{\mathrm{AB}} + F) = K_1 (c_{\mathrm{A}} - F)(c_{\mathrm{B}} - F)$$

and

$$F^2 - \left(c_{\mathrm{A}} + c_{\mathrm{B}} + \frac{1}{K_1}\right) F + \frac{c_{\mathrm{A}} c_{\mathrm{B}} - c_{\mathrm{AB}}}{K_1} = 0$$

We calculate these fluxes for the two equilibria, and use these. We reach a new concentration state, which is not the true equilibrium state, since we have not taken the coupling of the equilibria into account. So the operation is repeated with newly calculated fluxes, and after ca. 50 iterations, we reach a stable equilibrium. The concentrations correspond to the joint equilibrium specified by the two thermodynamic association constants K_1 and K_2.

For practical applications, it is important that the calculated fluxes are not too large so that calculation does not overshoot the target. So, for example, if K_1 and K_2 are very large, and $c_A > c_B$ and $c_C > c_B$, our initial analysis of the first equilibrium will show that almost all of the B must flow to AB. At the same time, analysis of the second equilibrium will show that all of the B should flow to BC. When we take account of the two fluxes, we obtain negative concentrations in the first iteration. Since in this example, there are at most two fluxes for a single species (free B), this complication can be avoided by reducing every flux by 50%.

3.12 Regularization – Application to Analytical Ultracentrifugation

As outlined in the mathematical introduction (Sect 2.1), any model that describes an experiment can be used either as a discretely calculated function to describe an experiment or in a regularizing fashion to calculate the distribution of model parameters that best describe the experiment. In biochemistry, the latter approach was first applied to the analysis of relaxation kinetics or circular dichroism data. Here, we show how the approach can be used to analyze sedimentation velocity experiments in the analytical ultracentrifuge.

In an ultracentrifuge, dissolved macromolecules (e.g. proteins) are subjected to large gravitational fields, which causes them to sediment with a velocity that is described by the sedimentation coefficient s:

$$v = \frac{\mathrm{d}x}{\mathrm{d}t} = s\omega^2 x$$

where v is the velocity of the molecule, ω is the angular speed of the rotor, and x is the distance from the center of rotation. The sedimentation coefficient has the dimension of time and it is usually given in units of S (for Svedberg) with $1\ \mathrm{S} = 10^{-13}$ s. Larger molecules sediment faster and consequently have larger values of sedimentation coefficient than smaller ones. These differences in sedimentation can be exploited to detect interactions of macromolecules. In an analytical ultracentrifuge, the local concentration of particles in the centrifuge cell can be observed by optical absorption or refractive index measurements.

To illustrate such an experiment Fig. 7(a) shows the sedimentation of a mixture of two interacting proteins, N-acetyl-glutamate-kinase (NAGK) and the signal transducing protein PII from *Synechococcus* sp. PCC7942. NAGK is a hexamer with a molecular mass of 192 kg mol^{-1} and a sedimentation constant of 8 S, while PII is a tetramer with a molecular mass of 42 kg and sedimentation coefficient of 3 S. Two clearly separated boundaries can be observed, the faster sedimenting boundary (10 S) is formed by the complex of these two proteins while the slower one represents PII protein whose concentration in this experiment was chosen to be much larger than that of NAGK (Maheswaran et al. 2004). It is clear that the two boundaries are spread due to diffusion and that a straightforward analysis of the amounts of free and complex bound protein is difficult to estimate from the data directly. However, the movement of such boundaries in the analytical ultracentrifuge can be described using the Lamm equation derived by Lamm in 1929:

$$\frac{\partial c}{\partial t} = \frac{1}{x}\frac{\partial}{\partial x}\left[\frac{1}{x}\left(D\frac{\partial c}{\partial x} - s\omega^2 x^2\right)\right]$$

This equation can be solved by numerical methods and represents a model

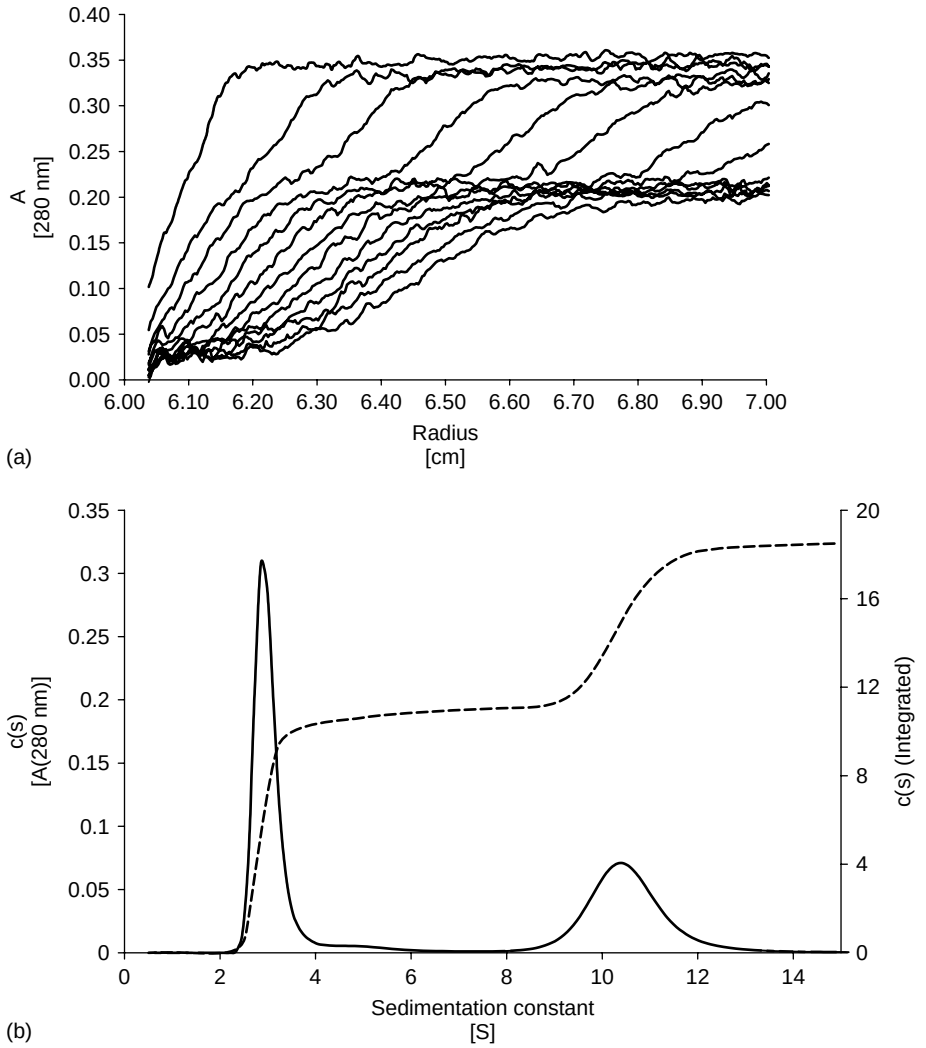

Fig. 7 Sedimentation of a mixture of 11.2 μM NAGK and 2.24 μM PII proteins (for details cf. text.) (a) Sedimentation profiles in the analytical ultracentrifuge taken every 19 min at a speed of 38 000 rpm. (b) Sedimentation coefficient distribution calculated with SEDFIT. The dotted line is the integral over the distribution and can be used to calculate the absorptions of free and bound proteins.

function of the type $c = f(x, t; s, D)$ where x and t are the independent variables and, s and D, the diffusion coefficient, are parameters that are constant throughout an experiment. Regularization algorithm given by Schuck in 2000 is similar to the following description using Lamm's differential equation. Thus, s and D are related to molecular parameters by the equations

$$s = \frac{M(1 - \bar{v}\sigma)}{N_A f} \text{ and } D = \frac{RT}{N_A f}$$

where M represents the molar mass (kg mol^{-1}), $\bar{v}$ the partial specific volume (m^3 kg^{-1}) and the f frictional coefficient (kg s^{-1}) of the particle; N_A and R are Avogadro's number and the gas constant, respectively. For a sphere of radius R_0, the frictional coefficient is given by Stokes law $f = 6 \cdot \pi \cdot \eta \cdot R_0$. Using $\bar{v}$ and the frictional ratio f/f_0, which is the ratio of the actual frictional coefficient of the particle to the frictional coefficient of a perfect sphere with the same volume, the diffusion coefficient can be related to the sedimentation coefficient by

$$D = \frac{RT}{6\pi\eta N_A}\sqrt[3]{\frac{2(1-\bar{v}\rho)}{9\bar{v}s\eta(f/f_0)^4}}$$

To analyze the sedimentation data using the regularizing procedure, we write the describing function in the form $\int c(s)f(x, t; s)\mathrm{d}s$ from which the sedimentation coefficient distribution $c(s)$ can be calculated from the original sedimentation data. The program to do this (SEDFIT) is available from Peter Schuck (http://www.analyticalultracentrifugation.com/default.htm).

Figure 7(b) shows the sedimentation profiles represented as the sedimentation coefficient distribution and from the integral of the two peaks the absorption and thus relative amounts of free and bound protein can be evaluated; from this it was calculated that the proteins form a 1 : 1 (NAGK hexamer: PII trimer) complex.

Bibliography

The present text is a modified and expanded version of Chapter 9 published in Pingoud, A., Urbanke, C., Hoggett, J. & Jeltsch, A. (2002) Biochemical Methods, Wiley-VCH. In that book, detailed step-by-step instructions can be found on how to implement many of the procedures described here in Excel worksheets.

Books and Reviews

Bisswanger, H. (2002) *Enzyme Kinetics Principles and Methods*, Wiley-VCH, Weinheim, Germany.

Brand, L. (Ed.) (1992) *Methods in Enzymology, Vol. 210: Numerical Computer Methods*, Elsevier, Amsterdam, Netherlands.

Brand, L. (Ed.) (2004) *Methods in Enzymology, Vol. 383: Numerical Computer Methods*, Part D, Elsevier, Amsterdam, Netherlands.

Johnson, M. (Ed.) (1994) *Methods in Enzymology, Vol. 240: Numerical Computer Methods*, Part B, Elsevier, Amsterdam, Netherlands.

Johnson, M. (Ed.) (2000) *Methods in Enzymology, Vol. 321: Numerical Computer Methods*, Part C, Elsevier, Amsterdam, Netherlands.

Johnson, M. (Ed.) (2004) *Methods in Enzymology, Vol. 384: Numerical Computer Methods*, Part E, Elsevier, Amsterdam, Netherlands.

Purich, D. (Ed.) (1995) *Methods in Enzymology, Vol. 249: Enzyme Kinetics and Mechanism*, Part D, Elsevier, Amsterdam, Netherlands.

Primary Literature

Johnson, M.L. (2000) Outliers and robust parameter estimation, *Methods Enzymol.* **321**, 417–424.

Lamm, O. (1929) Die differentialgleichung der ultrazentrifugierung, *Ark. Matematik, Astron. Ffys.* **21B**(2), 1–4.

Maheswaran, M., Urbanke, C., Forchhammer, K. (2004) Complex formation and catalytic activation by the PII signaling protein of N-Acetyl-L-glutamate kinase from *Synechococcus elongatus* strain PCC 7942, *J. Biol. Chem.* **279**(53), 55202–55210, Epub 2004 Oct 22.

Provencher, S.W. (1982a) A constrained regularization method for inverting data represented by linear algebraic or integral functions, *Comput. Phys. Commun.* **27**, 213–227.

Provencher, S.W. (1982b) CONTIN: a general purpose constrained regularization program for inverting noisy linear algebraic and

integral equations, *Comput. Phys. Commun.* **27**, 229–242.

Rusling, J.F., Kumosinski, T.F. (1996) *Nonlinear Computer Modeling of Chemical and Biochemical Data*, Academic Press, New York.

Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling, *Biophys. J.* **78**(3), 1606–1619.

Radioisotopes in Molecular Biology

Robert James Slater
University of Hertfordshire, Hatfield, UK

1 Principles 413
1.1 Radioactive Decay 413
1.2 Kinetics of Decay 413
1.3 Units 415

2 Radiation Protection 416
2.1 General Principles 416
2.2 Unpacking and Dispensing Radioactive Solutions 419

3 The Choice of Radionuclide 420

4 Nucleic Acid Labeling 422

5 Protein Labeling 424

6 Detection 427
6.1 Ionization Monitors 427
6.2 Scintillation Counters 427
6.3 Cerenkov Counting 429
6.4 Autoradiography 429
6.5 Personal Dosimeters 431

Bibliography 431
Books and Reviews 431
Primary Literature 432

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

Keywords

Autoradiography
The detection of radioactivity using photographic emulsions.

Fluorography
A sensitive form of autoradiography in which radioactivity is detected by light emanating from a scintillator (or "fluor") in close contact with the sample.

Half-life
The time taken for the activity of a radionuclide to lose half its value by decay.

Quenching
Any process that reduces the efficiency of detection of radioactivity.

Radioactivity
The emission of ionizing radiations by matter.

Radioisotopes
Atoms with the same atomic number (protons) but differing mass numbers (protons plus neutrons), and having unstable nuclei that decay to a stable state by the emission of ionizing radiations; for example, ^{14}C is a radioisotope of ^{12}C.

Radionuclide
An atomic species that is radioactive, for example, ^{3}H, ^{14}C and so on.

Specific Activity
The rate of decay per unit mass, for example dpm g^{-1}, Ci $mole^{-1}$, Bq $mmol^{-1}$.

Radioisotopes are used extensively in molecular biology. They can be incorporated into DNA, RNA, and protein molecules, both *in vivo* and *in vitro*. As a consequence, the presence or metabolism of macromolecules can be investigated or "traced." The incorporation of radioisotopes allows the detection of minute quantities, thereby facilitating experimental techniques that require high sensitivity. Labeling *in vitro* is much more efficient than labeling *in vivo*; specific activities regularly reach 10^8–10^9 dpm μg^{-1} for *in vitro* ^{32}P–labeling of DNA, for example. High-specific activity leads to greater sensitivity. Specific activity is inversely related to the half-life of the radioisotope. Low-energy emitters, such as ^{3}H and ^{35}S provide high resolution in autoradiography. High-energy, high-specific activity emitters such as ^{32}P for nucleic acid labeling, provide great sensitivity but low resolution in autoradiography. The radionuclides ^{3}H, ^{14}C, ^{35}S, and ^{125}I are most commonly used for protein labeling; ^{32}P and ^{33}P are most commonly used for nucleic acid labeling.

1 Principles

1.1 Radioactive Decay

Atomic nuclei contain a number of protons (referred to as the atomic number, Z) and, except in the case of hydrogen, neutrons (N) also. The total number of these nuclear atomic particles (or nucleons) is the mass number (A), that is $A = Z + N$. The atomic number determines the element (e.g. when $Z = 6$ the element is carbon). Isotopes are elements that have the same atomic number but differing mass numbers. For example, ^{12}C and ^{14}C are different isotopes of carbon, both with atomic number 6, but differing mass numbers, 12 and 14 respectively, thereby indicating that the isotopes have either 6 or 8 neutrons. Some atoms are unstable: they have the incorrect number of neutrons for nuclear stability and are referred to as radionuclides or, more commonly, radioisotopes. Radioactivity is the means by which unstable atoms reach a stable state. The process can occur in one step, as is the case with most radioisotopes used in biological experiments, or in a number of steps, such as in the case of uranium, which decays through several radionuclides to a stable isotope of lead. There are many ways in which this decay can occur.

Alpha decay is loss of a helium nucleus, $^{4}{}_{2}He$, for example:

$$^{226}{}_{88}Ra \longrightarrow {}^{222}{}_{86}Rn + {}^{4}{}_{2}He$$

Alpha particles are heavy and slow moving, have poor penetration (3–9 cm), and a net double positive charge. Alpha emitters are found mostly amongst the heavier elements and are rarely used by biologists.

Beta decay results from the conversion of a neutron to a proton (negatron emission) or a proton to a neutron (positron emission). Of most relevance to biologists is the negatron emission form of β decay, involving the loss of an electron, which forms the ionizing radiation. Many commonly used radionuclides decay by this mechanism (e.g. ^{3}H, ^{14}C, ^{35}S, ^{33}P, ^{32}P). For example, ^{14}C decays as follows:

$$^{14}{}_{6}C \longrightarrow {}^{14}{}_{7}N + \beta^{-}$$

Beta particles are less charged, lighter, and faster moving than alpha particles; their energy and therefore the range of the radiation produced, varies with the source. This has significant implications for their use and safe handling, and this is discussed later (Sect. 2).

Gamma decay is the emission of electromagnetic radiation, with properties identical to X rays. It often accompanies β-decay. For example:

$$^{131}{}_{63}I \longrightarrow {}^{131}{}_{64}Xe + \beta^{-} + \gamma$$

Electron capture is a process by which a proton is combined with an electron from the innermost K shell, forming a neutron; following a complex series of events, an X ray is emitted. An example of this kind of decay is shown by ^{125}I:

$$^{125}{}_{53}I \longrightarrow {}^{125}{}_{52}Te + \text{X ray}$$

Decay of ^{125}I is also accompanied by emission of Auger electrons, radiation of low energy, providing high-resolution detection in autoradiography (see Sect. 3).

1.2 Kinetics of Decay

It is impossible to predict when an individual unstable nucleus will decay,

but when observing the decay of a large number of unstable atoms there is a statistical probability that a certain number will decay within a given time. The number that decay per unit time is proportional to the number of unstable atoms present. Mathematically this is represented as:

$$\frac{-\mathrm{d}N}{\mathrm{d}t} \alpha N \quad \text{or} \quad \frac{-\mathrm{d}N}{\mathrm{d}t} = \lambda N \qquad (1)$$

where $N =$ number of unstable nuclei, $t =$ time and $\lambda =$ the decay constant sec^{-1}. Therefore the amount of radioactivity (A, the count rate) is:

$$A = \frac{\mathrm{d}N}{\mathrm{d}t} = \lambda N \qquad (2)$$

As a consequence, radioisotopes exhibit exponential decay as is illustrated by the graphs in Fig. 1, the rate at which a radioisotope decays being determined by its decay constant, λ. This constant is a fundamental property of a radioisotope, and is not influenced by temperature, pressure, and so on. Integration of Eq. (1) with respect to time (letting $N = N_0$ at $t = 0$) gives the following expression:

$$N = N_0 \mathrm{e}^{-\lambda t} \qquad (3)$$

and

$$1\mathrm{n}\frac{Nt}{N_0} = -\lambda t \qquad (4)$$

(where $N = Nt$ at time t)

or:

$$1\mathrm{n}\ \text{count rate(time, } t_2) = 1\mathrm{n}\ \text{count rate(time, } t_l) - \lambda t \qquad (5)$$

Thus, if the decay constant is known, the count rate at any time can be calculated from a known count rate and time. In practice, decay constants are rarely quoted;

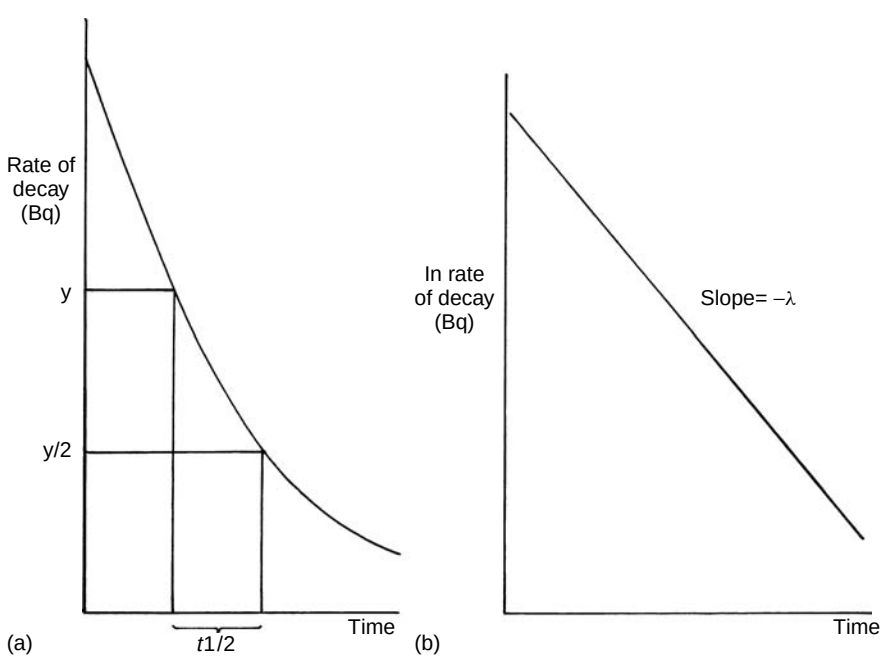

Fig. 1 The rate of decay with time on a linear (a) and semilogarithmic (b) scale.

half-life ($t_{1/2}$) is in many ways a more convenient way to express decay kinetics. The half-life is the time taken for the count rate to drop to half of a given value, and is depicted in Fig. 1.

From Eq. (4), letting $N_t = (1/2)N_0$:

$$1n\tfrac{1}{2} = -\lambda t_{1/2}$$

or

$$1n\,2 = \lambda t_{1/2} \tag{6}$$

or

$$\lambda = \frac{0.693}{t_{1/2}} \tag{7}$$

So, if the half-life is known, a combination of Eqs. (7 and 5) can be used to determine future or past levels of radioactivity. This is particularly valuable when working with short half-life radioisotopes such as ^{32}P. Decay tables based on this type of mathematics are available in the reference literature and suppliers' catalogues.

1.3 Units

The International System of Units (SI System) uses the becquerel (Bq) as the unit of radioactivity, lBq being one disintegration per second (1d.p.s.). However, a frequently used unit is the curie (Ci). This unit originates from the number of disintegrations from a gram of pure radium: 1 Ci, therefore, being 3.7×10^{10} d.p.s. (or 37 GBq). This is a very large amount of radioactivity, and most molecular biologists use quantities measured in microcuries (μCi); 1 μCi is 2.22×10^6 d.p.m. (disintegrations per minute) or 37 kBq. The becquerel is the internationally recognized unit of radioactivity and has been adopted by the International Commission for Radiological Protection (ICRP). As a consequence, most countries have incorporated the unit into their legislation regarding use and disposal of radioactivity; suppliers of radioisotopes, however, frequently still use the Ci for historical reasons.

The units Bq, Ci, d.p.s., and d.p.m. all refer to the actual number of nuclear disintegrations. The rate of decay detected by an instrument is measured in counts per minute (c.p.m.). This is usually less than the true rate, d.p.m., because counting is rarely 100% efficient. Consequently, it is normal practice to calibrate the detector and calculate d.p.m. as c.p.m. × 100/efficiency, as a percentage value. Surprisingly, it is sometimes possible for the recorded count rate (c.p.m.) to exceed the actual rate of decay. This can occur in liquid scintillation counters and is discussed in Sect. 6.2.

The term *specific activity* is important in relation to experiments with radioactivity. The term refers to the amount of radioactivity per unit mass; thus d.p.m. g^{-1}, c.p.m. g^{-1}, Ci $mole^{-1}$, Bq $mole^{-1}$ are all units of specific activity. Frequently, molecular biologists require the highest specific activity possible; for example, DNA labeling with high-specific activity ^{32}P-labeled nucleotides results in probes with the maximum possible sensitivity. The highest specific activities are obtained from radioisotopes with the shortest half-lives (see Sect. 3). Sometimes it is necessary or desirable to change the specific activity by addition of nonradioactive material (a "cold" carrier). This can be achieved by applying the following formula:

$$W = Ma\left(\frac{1}{A'} - \frac{1}{A}\right) \tag{8}$$

where W = mass of cold carrier required, expressed in mg; a = amount of radioactivity present, expressed in MBq; M = molecular weight of the compound; A = the original specific activity MBq $mmol^{-1}$;

Tab. 1 A summary of units and their definitions.

Unit	*Abbreviation*	*Definition*
Counts per minute or second	c.p.m. c.p.s.	The recorded rate of decay
Disintegration per minute or second	d.p.m. d.p.s.	The actual rate of decay
Curie	Ci	The number of d.p.s. equivalent to 1 g of radium (3.7×10^{10} d.p.s.)
Millicurie	mCi	Ci $\times 10^{-3}$ or 2.22×10^{9} d.p.m.
Microcurie	mCi	Ci $\times 10^{-6}$ or 2.22×10^{6} d.p.m.
Becquerel (SI Unit)	Bq	1 d.p.s.
Gigabecquerel (SI Unit)	GBq	10^{9} Bq or 27.027 mCi
Megabecquerel (SI Unit)	MBq	10^{6} Bq or 27.027 μCi
Electron volt	eV	The energy attained by an electron accelerated through a potential difference of 1 V. Equivalent to 1.6×10^{-19} joules.
Roentgen	R	The amount of radiation, which produces 1.61×10^{15} ion pairs per kg of air (2.58×10^{-4} coulombs kg^{-1})
Rad	rad	That dose which gives an energy absorption of 0.01 joule kg^{-1} (J kg^{-1})
Gray (SI Unit)	Gy	That dose which gives an energy absorption of 1 joule kg^{-1}. Thus, 1 Gy = 100 rad.
Rem	rem	That amount of radiation which gives a dose in human equivalent to 1 rad of X rays.
Sievert	Sv	That amount of radiation, which gives a dose in human equivalent to 1 gray of X rays. Thus, 1 Sv = 100 rem.

A' = the required specific activity MBq $mmol^{-1}$.

Units and definitions are summarized in Table 1.

2 Radiation Protection

2.1 General Principles

Toxicity is the greatest practical disadvantage of using radioisotopes. They produce ionizing radiations, which, when absorbed by cells, cause ionization and the production of free radicals. These radicals, in turn, cause mutation of DNA, hydrolysis of proteins, and ultimately cell death. The toxicity of radiation is dependent not simply on the amount present but on the amount absorbed, the energy of the radiation and its nature, and therefore the biological effect. The international unit to describe absorbed dose, taking into account its biological effectiveness, is the sievert (Sv), known as the equivalent dose. This is calculated from the absorbed dose (gray, Gy), which is a measure of the energy absorbed per unit mass (joule kg^{-1}), multiplied by a radiation-weighting factor (W_R). For X rays, γ-rays and ß-radiations this

weighting factor is one, that is the equivalent dose = absorbed dose = joule kg^{-1}. Other forms of radiation such as neutrons and α-particles are more toxic, with a W_R of up to 20.

Radiation is emitted from a source in all directions, so the level of irradiation, and therefore the dose, is inversely related to the square of the distance from the source. A commonly used parameter is dose rate (Sv per minute or hour, etc.). If this is known, either from quoted figures or measurements with a dose rate meter, total doses can be calculated. For example, if a source is delivering lmSv per hour and you work with the source for 15 min the dose will be 250 μSv. To calculate dose rates at any distance, use the formula:

$$\text{Dose rate}_1 \times \text{distance}_1^2 = \text{dose rate}_2 \times \text{distance}_2^2 \quad (9)$$

There are official dose limits for workers using radiation: 20 mSv $year^{-1}$ for the whole body. To put this in perspective, it represents about 10 times the background radiation. There are also annual dose limits for specific organs, for example, 150 mSv for the lens of the eye, and 500 mSv for the hand; however, a fundamental principle behind radiation protection is that doses must be as low as reasonably practical (the ALARP principle). In practice this means that workers should not work up to a dose limit, but rather always strive to reduce their exposure. For example, radioisotopes should only be used if there is a net benefit (see the end of Sect. 6.4); and when radioisotopes are used, all practical steps must be taken to reduce dose to workers, (for example, by choosing a low-energy emitter).

So far the discussion has been primarily concerned with *external radiation*, that is, radioisotopes outside the body. However, a potential hazard in working with radioisotopes in a biology laboratory is via *internal radiation*: radiation that has entered the body by, for example, ingestion or inhalation. The annual limit on intake (ALI) provides a guide to relevant dosimetry. ALIs for radioisotopes used in molecular biology are included in Table 2, but it should be noted that the figures are for the elemental isotopes, and not particular formulations. The ALI for ^{3}H-thymidine is lower, for example, than that for ^3H_2O, because it is more toxic due to its longer clearance time from the body. Exhaustive lists of ALI can be found in the detailed radiation protection literature.

The various radioisotopes used in molecular biology fall into two broad categories of hazard: those that present an external radiation hazard (^{32}P and ^{125}I) and those that present only a negligible (^{14}C, ^{35}S, ^{33}P) or zero (^{3}H) external radiation risk. Clearly, therefore, it is incumbent on the laboratory worker to avoid ^{32}P or ^{125}I wherever possible. All radioisotopes present an internal radiation risk, particularly when handled as liquids or gases, but again, the penetrating, higher energy radiations such as from ^{32}P and ^{125}I present the greatest risk. The figures for annual limit on intake shown in Table 2 give an indication of the relative risk from ingestion. Shielding, other than that routinely used for storage, is not usually required for ^{14}C, ^{35}S, ^{33}P or ^{3}H when used at the levels appropriate for most molecular biology experiments. Here the potential risk is through ingestion, which can be avoided by adherence to good laboratory protocols that are generally applicable to work with all radionuclides: wearing of laboratory coat, gloves, and safety spectacles, working in spill trays lined with absorbent paper, double containment of stocks for storage

Tab. 2 Properties and radiation protection data of radioisotopes commonly used in the molecular biological sciences. Radioisotopes vary considerably with respect to their properties and potential hazards. The table below summarizes some generally useful information about isotopes used in biological experiments.

	Isotope						
Property/radiation protection criteria	^{3}H	^{14}C	^{35}S	^{33}P	^{32}P	^{125}I	^{131}I
$t_{1/2}$	12.3 years	5730 years	87.4 days	25.4 days	14.3 days	59.6 days	8.04 days
Mode of decay	β^-	β^-	β^-	β^-	β^-	X (EC)	γ and β
Max β energy (MeV)	0.019	0.156	0.167	0.249	1.709	Auger electrons	0.806
Monitor	Swabs counted by liquid scintillation	β-counter	β-counter	β-counter	β-counter	γ-probe	β-counter
Biological monitoring	Urine	Urine, breath	Urine	Urine	Urine	Thyroid	Thyroid
ALI[a]	480[b]	34	15	14	6.3	1.3[c]	0.9[c]
Critical organ	Whole body	Whole body/fat	Whole body/testis	Bone	Bone	Thyroid	Thyroid
Maximum range in air	6 mm	24 cm	26 cm	49 cm	790 cm	>10 m	>10 m
Shielding required	None	Perspex 1 cm	Perspex 1 cm	Perspex 1 cm	Perspex 1 cm	Lead 0.25 mm or lead	Lead 13 mm
γ dose rate (μSv h^{-1} from 1 GBq at 1 m	–	–	–	–	(β-dose rate 210 mSv at surface of 1 MBq ml^{-1}	34	51
Special considerations	Monitoring difficulties lead to potential internal hazard. DNA precursors more toxic than tritiated water.	Avoid generation of CO_2	Avoid generation of SO_2		Potential high source of external radiation. Lead shielding and finger dosimeters for quantities greater than 300 and 30 MBq respectively.	Iodine sublimes, work in fume hood. Spills should be treated with sodium thiosulphate solution prior to decontamination. Iodine compounds may penetrate rubber gloves – wear two pairs.	

[a]Based on a dose limit of 20 mSv using the most restrictive dose coefficients for inhalation or ingestion; some figures differ in Germany.
[b]Organically bound H-3.
[c]Based on dose equivalent limit of 500 mSv to thyroid.
Source: Reproduced with permission from Oxford University Press.

and transport, regular and routine monitoring, no eating and drinking (or sucking pencils etc.), in the laboratory, no mouth pipetting, clear delineation of areas used for radioisotope work and washing hands when leaving the laboratory.

Dose rates from λ sources are quoted as the specific dose rate constant. For ^{125}I, this figure is 3.4×10^{-2} μSv h^{-1} MBq^{-1} at 1 m.

For ^{32}P the following formula can be used:

$$\text{dose rate}(\mu\text{Sv h}^{-1})\text{at 30 cm} = 54 \times \text{activity (MBq)} \quad (10)$$

Using Eqs. (9 and 10), a 1 MBq source of ^{32}P (approx 27 μCi) has a dose rate of 0.9 μSv min^{-1} at 30 cm and 0.81 Sv min^{-1} at 1 cm.

Clearly, shielding is required for work with ^{32}P and ^{125}I; the most practical is 1 cm acrylic for ^{32}P and 1 cm lead-impregnated acrylic for ^{125}I. High-energy β-emitters generate secondary X rays from absorbers, known as bremsstrahlung radiation. The generation of this radiation increases with the atomic weight of the absorber. Hence lead-impregnated acrylic can generate bremsstrahlung radiation, and is therefore not ideally suited for use with ^{32}P.

Vials or test tubes containing ^{32}P or ^{125}I must not be handled directly even when using a body shield, as doses can potentially be very high. They must be manipulated with forceps or some other appropriate device, or kept in acrylic holders such as those specifically marketed by radioisotope and radiation protection equipment suppliers. Automatic pipettes should carry small acrylic shields. Workers who frequently use ^{32}P may be advised to wear fingertip and body dosemeters. In all cases, the member of the staff who is responsible for radiation protection (e.g. Radiation Protection Supervisor or Radiation Protection Advisor) should be consulted. He/she will advise on the laboratory requirements and the type of laboratory required. In the United Kingdom, this is most likely to be a "Supervised Laboratory" (i.e. one in which it is theoretically possible to exceed 1/10 but not more than 3/10 of the recommended dose limit). When dealing with penetrating radiations, remember the basic rules of radiation protection:

maximize distance
minimize time of exposure
use shielding

Do dose estimates on your experiments using the formula described above (Eqs. 9 and 10) and the data in Table 2. Monitor regularly. A sensible precaution when using radioactivity in a protocol for the first time is to do a dummy run with ink in place of the radioisotope solution. A further simple precaution is to always add radioisotopes to mixtures last where possible; this reduces the time of exposure and risk of contamination.

2.2 Unpacking and Dispensing Radioactive Solutions

The following advice applies to unpacking and dispensing solutions emitting penetrating radiations such as from ^{32}P or ^{125}I.

1. Study the data sheet supplied with the radioisotope, do all necessary calculations (including dose assessment on the unpacking and dispensing procedure) and prepare a receptacle containing nonradioactive carrier as necessary, securely attach a label to this receptacle

that states your name, date, radionuclide and radioactive concentration. Manipulation of the stock radioactive solution should be carried out only when everything else is ready.

2. Wear laboratory coat, personal dosimeters (body and fingertip, see Sect. 6.5) safety spectacles, one or two pairs of disposable gloves, lay out all your equipment behind a body screen in a lined spill tray ensuring that no receptacles for liquids can be knocked over. Switch on a monitor and check the battery.
3. Open the sealed container, wipe test the interior with a piece of tissue paper held by forceps, and check the paper against the monitor. Regularly repeat this wipe test throughout the procedure to check for contamination outside the sample vial.
4. Remove lid of inner container surrounding the sample vial. Wipe test the top of the vial. Dispense sample without removing vial if possible; alternatively, if necessary, transfer the inner vial, with forceps, to an acrylic holder.
5. If dispensing with a syringe through a Teflon seal, insert a second needle in the seal as an air bleed. Surround the needle on the syringe with tissue to absorb any tiny drops of contamination. Handle the syringe for as short a time as possible but without reducing dexterity.
6. If dispensing with an automatic pipette, remove lid or seal with care and use an acrylic pipette shield.
7. Place contaminated tissues in a shielded container, rinse syringe/pipette tips in detergent solution for aqueous disposal (this minimizes the radioactivity in solid waste), and place disposable syringe or pipette tip in an appropriate shielded waste disposal container.
8. Repack the stock container, and wipe test all surfaces.
9. Monitor gloves, cuffs, all working surfaces, forceps and so on. Return stock to storage area. Upon completion of work remove laboratory coat, wash and monitor hands, and complete all necessary records in the radioisotope log.

If you are dispensing low-energy radioisotopes, it may be unrealistic to apply all aspects of the above procedure: ^{3}H for example, is not detected by bench monitors and the risk from external radiation is negligible, and therefore shielding is not required. However, the same **principles** apply, in that the risk of contamination should be reduced wherever possible, by, for example, working in a spill tray and preparing everything prior to dispensing the radioisotope.

3 The Choice of Radionuclide

The relative order of increasing radiation hazard is: ^{3}H, ^{14}C, ^{35}S, ^{33}P, ^{32}P, ^{131}I, and ^{125}I. Choose the radioisotope with lowest toxicity where possible. Remember that the shorter the half-life the greater the potential specific activity. The high-energy emitters give high counting efficiencies and strong signals in autoradiography. The lower-energy emitters give lower detection efficiencies but high resolution in autoradiography. The relative merits of radionuclides used in molecular biology are provided in Table 3. As a consequence, ^{32}P provides the highest sensitivity for labeling of probes and filter hybridization for genomic southern blots and northern analysis; ^{35}S and ^{33}P are preferred for DNA sequencing, and ^{35}S and ^{3}H are preferred for *in situ* hybridization and detection of unscheduled DNA synthesis. Phosphorus-33 is a good compromise for most DNA experiments as it combines the features of

Tab. 3 The relative merits of commonly used radionuclides.

	Advantages	*Disadvantages*
^{3}H	Safety High specific activity Wide choice of labeling position in organic compounds High resolution in autoradiography	Low efficiency of detection ^{3}H is large compared with ^{1}H and may cause an isotope effect Isotope exchange with the environment
^{14}C	Safety Wide choice of labeling position in organic compounds High resolution in autoradiography	Low specific activity significantly reduces value in molecular biology
^{35}S	High specific activity Good resolution in autoradiography	Relatively long biological half-life
^{33}P	High specific activity Safety relative to ^{32}P Longer half-life than ^{32}P Good resolution in autoradiography	Short half-life relative to ^{35}S Cost
^{32}P	Sensitivity of detection (10 fg or 50 dpm cm^{-2} in filter hybridization) Short half-life simplified disposal High specific activity Cerenkov counting	Short half-life effects cost and experimental design External radiation hazard Poor resolution in autoradiography
^{125}I	Ease of detection High specific activity Efficient *in vitro* labeling of proteins	Safety: radiation is penetrating and ^{125}I accumulates in thyroid, ALI is very low

relatively high-specific activity, and lower toxicity than ^{32}P.

4
Nucleic Acid Labeling

Radiolabels are most commonly introduced into nucleic acids by addition of single- or multiple-labeled nucleotides using enzyme-catalyzed reactions (Table 4 and Figs. 2 and 3). Most DNA and RNA polymerases are specific for particular pre-existing nucleic acid templates (e.g. DNA-dependent RNA polymerase) and synthesize a polynucleotide complementary to that template. There are two polymerization reactions that do not create polynucleotides complementary to a template: those catalyzed by terminal deoxynucleotidyl transferase from calf intestine, which adds deoxynucleotides to a 3′ end of DNA or RNA; and poly (A) polymerase, which adds adenosine monophosphates to 3′ ends of RNA molecules. All DNA polymerases require a nucleotide with a free 3′-hydroxyl hydrogen bonded to the template (i.e. a primer). RNA polymerases require a promoter, a specific DNA sequence on a double stranded-DNA molecule, in order to initiate transcription (RNA synthesis).

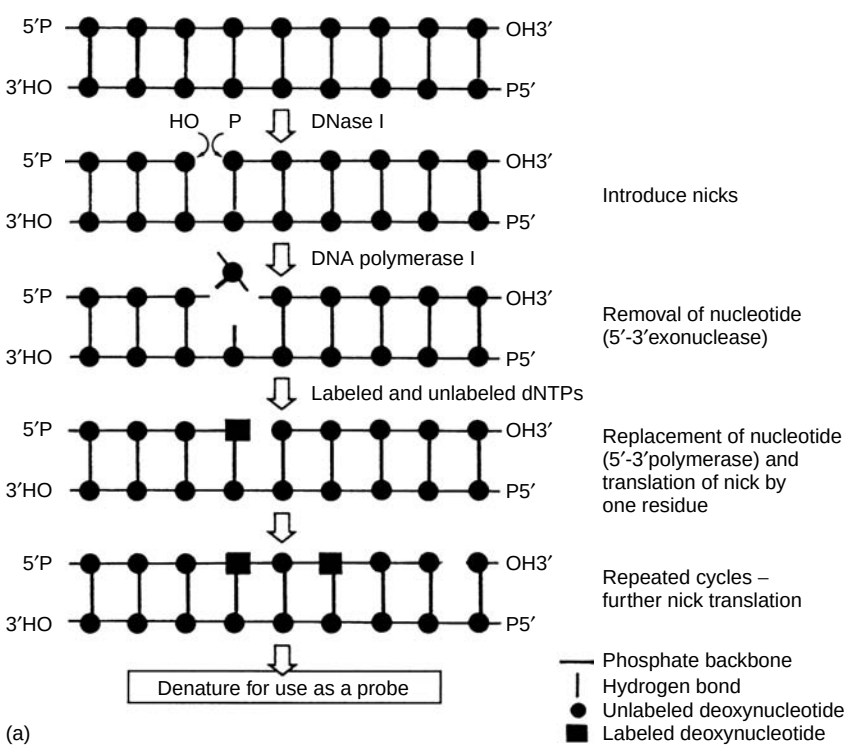

Fig. 2 Some methods for *in vitro* labeling of DNA are as follows: (a) the nick-translation reaction; (b) the random primer labeling reaction; (c) the 5′-end-labeling reaction using T4 polynucleotide kinase; (d) the 3′-end-labeling reaction using terminal deoxynucleotidyl transferase; (e) in-filling reaction using Klenow polymerase. (Reprinted with permission from Oxford University Press.)

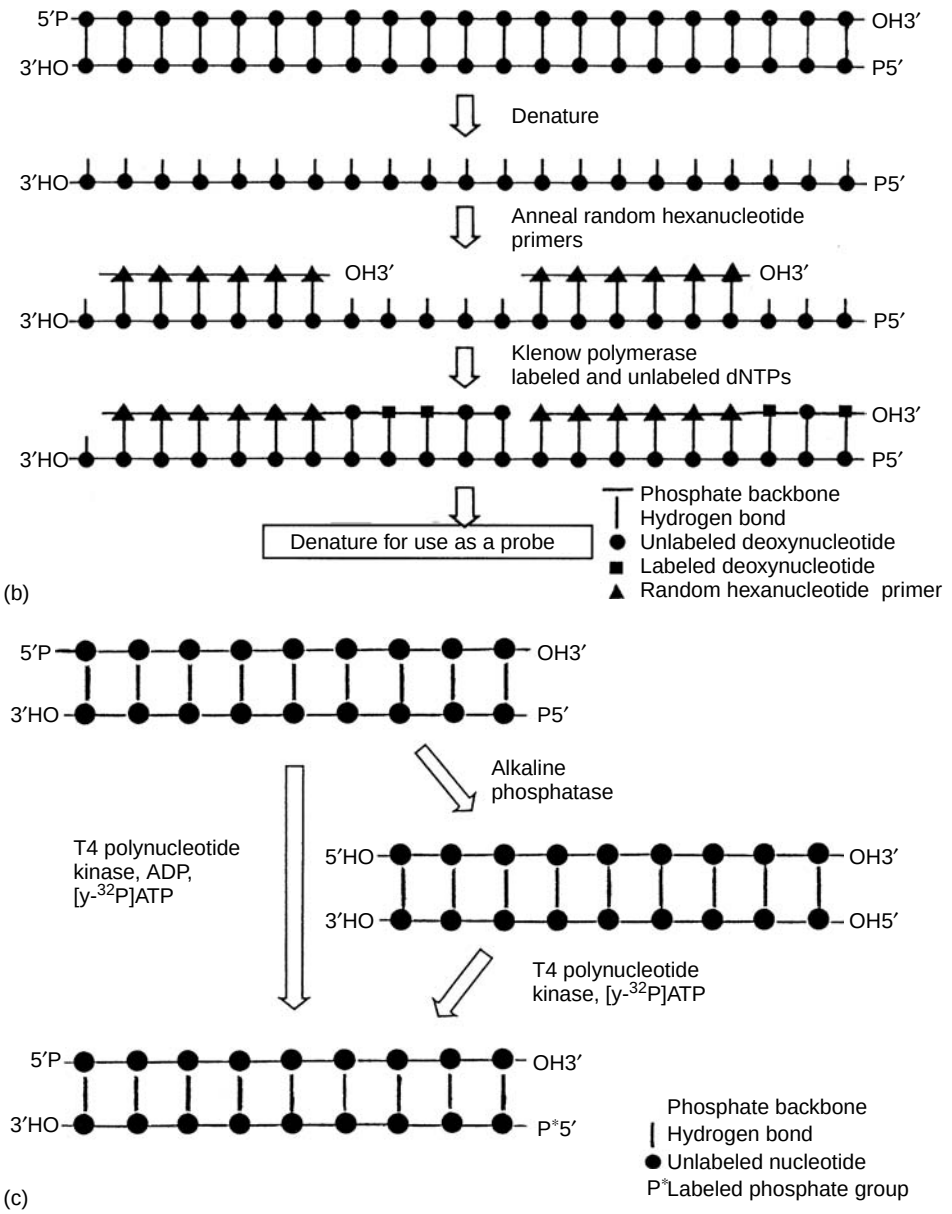

Fig. 2 (*Continued*)

Modification reactions can be used to incorporate labeling, for example, the addition of a ^{32}P phosphate to the 5′-end of a nucleic acid molecule by the enzyme T4 polynucleotide kinase using [^{32}P-γ] ATP as substrate.

When a labeling reaction is complete it is necessary to estimate the extent of

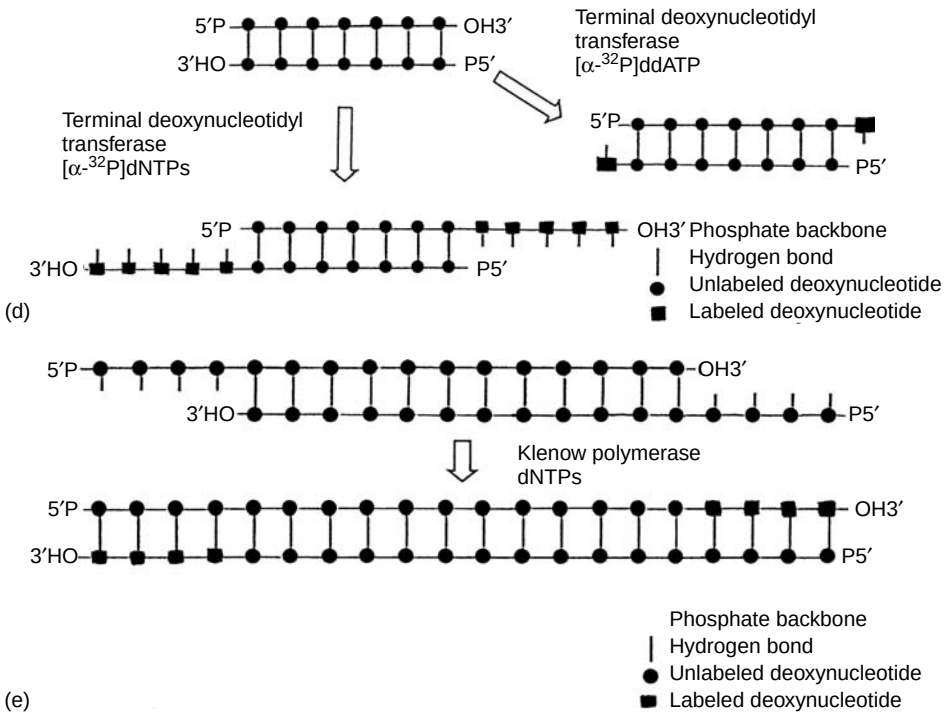

Fig. 2 (*Continued*)

incorporation. To do this, a small aliquot (2 μl) is removed from the reaction and diluted to a known volume (say 20 μl) with water. Two aliquots of this are counted in a scintillation counter. Two further aliquots are transferred to small tubes containing a few μg of carrier DNA or RNA in solution. The tubes are then filled with 10% trichloroacetic acid solution and left on ice for 30 min. The nucleic acids precipitate and can be collected on glass fiber filters. These are also counted in a scintillation counter. Comparison of these counts with the unprecipitated samples gives an indication of the percentage of incorporation, which should be at least 50% for most labeling protocols.

It is not always necessary to remove free, unincorporated labeled nucleotides from labeled probes; but if it is thought to be necessary, free label can be removed by ethanol precipitation of the probe in the presence of 0.1 M ammonium acetate or by the use of a 'spin column': Sephadex G25 swollen in buffer in a disposable syringe or pipette tip plugged with siliconized glass wool. The sample is applied to the column and, following centrifugation, the labeled probe is recovered from the centrifuge tube. Free label remains in the column. Alternatively, commercially made kits for purification of nucleic acids can be used, but these are usually more expensive.

5 Protein Labeling

Proteins are most frequently radiolabeled with ^{3}H, ^{14}C, ^{35}S, or ^{125}I. Tritium and

Tab. 4 Properties of nucleic acid–labeling reactions.

Method	*Labeling density*	*Template*	*Amount of template (typical)*	*Reaction time*	*Incorporation efficiency*	*Nature of probe*	*Amount of probe*	*Specific activity of probe (dpm μg^{-1})*
Random prime	Uniform	ssDNA (and denatured dsDNA)	25 ng	5 min–3 h	~75%	DNA	40–50 ng	5×10^9
Nick translation	Uniform	dsDNA	0.5–1 μg	~2 h	~60%	DNA	0.5–1 μg	5×10^8
Transcription labeling	Uniform	dsDNA	0.5–1 μg	1 h	~75%	RNA	~250 ng	5×10^9
3′-end labeling	End	Oligo or DNA fragment	~10 pmol ends	30–60 min	Variable	Oligo or DNA	10 pmols	5×10^6
End repair	End	Oligo or dsDNA fragment	~100 pmols ends	15 min	Variable	Oligo or DNA	100 pmols	5×10^5
5′-end labeling	End	Oligo or DNA fragment (dephosphorylated for forward reaction)	~10 pmol ends	~1 h	Variable	Oligo or DNA	10 pmols	5×10^6

Source: Printed with permission from Oxford University Press.

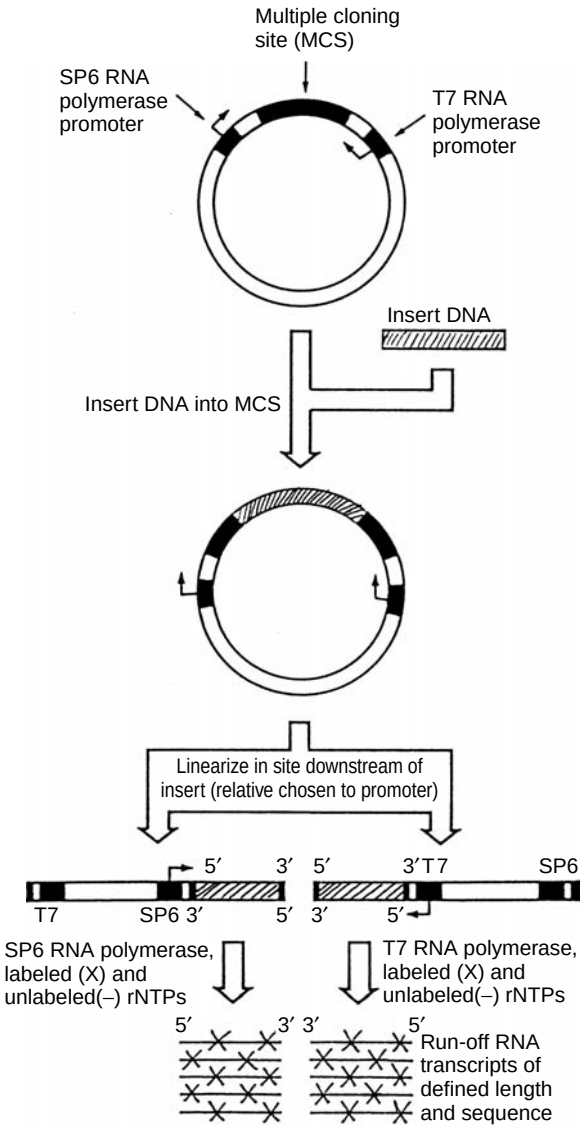

Fig. 3 Preparation of RNA probes using a plasmid containing SP6 and T7 RNA polymerase promoters. (Reprinted with permission from Oxford University Press.)

^{14}C are used for reasons of safety and practicality (because of their long half-life) whenever it is not essential to have proteins labeled to high-specific activity. Proteins can be labeled *in vivo* by incubating cells with labeled amino-acids. Alternatively they can be labeled *in vitro* with N-succinimidyl [2,3–^{3}H] propionate or by reductive methylation with [^{14}C] formaldehyde. Both techniques normally result in

proteins that retain their biological activity. Methionine labeled with ^{35}S is used for *in vivo* labeling and *in vitro* protein synthesis as it provides a higher specific activity and good resolution in autoradiography of polyacrylamide gels used for analysis of products.

Radioiodination is used when proteins with a high-specific activity are required, for example, in radioimmunoassay and related techniques. Activities of up to 10^8 dpm μg^{-1} can be achieved. There are many methods that rely on either direct oxidative iodination or indirect conjugation iodination. Probably the most commonly used technique is the chloramine T procedure, as it is relatively simple and inexpensive (for protocol see Slater, Ed., 2002).

Most experiments that involve iodinated proteins require a preparation that is free of unincorporated radioactive iodide. This is most commonly removed by either gel filtration or high performance liquid chromatography (HPLC).

Iodination of proteins is potentially one of the most hazardous techniques in molecular biology. The levels of ^{125}I employed are often in the region of 18.5 MBq (500 μCi). Rigid application of the principles of radiation protection is required, labeling should be carried out only by highly trained personnel with the permission of the radiation protection advisor, and a controlled facility is usually required (for example, if the level of ^{125}I used exceeds 10 MBq).

6 Detection

6.1 Ionization Monitors

An ionization monitor consists of a small gas chamber containing two electrodes, a voltage supply, and a scaler. Ionizing radiation enters through a thin end window, the gas is ionized and a pulse of current is recorded on the scaler, usually displayed as counts per second, counts per minute or Bq cm^{-2}. Gas ionization monitors are suitable for detecting ^{32}P, and they will pick up ^{14}C, ^{35}S and ^{33}P if a sensitive instrument with a very thin end window is used. They do not detect ^{3}H, because the β-particles cannot penetrate the end window. The instruments detect ^{125}I, but not particularly effectively, since electromagnetic radiation passes through the gas and results in relatively poor ionization.

The monitors must be regularly checked and officially calibrated annually. If you use ^{32}P, it is also worthwhile to do a simple calibration on your instrument for your own information. Place a known volume of ^{32}P in a plastic tube with the lid off and record the decay rate at various distances above the tube. Monitoring 0.1 MBq ^{32}P at 10 cm (equivalent to a dose rate of approximately 0.5 μSv min^{-1}) should give a reading in excess of 50 c.p.s. All personnel working with radioisotopes (except ^{3}H) should have constant access to a monitor while at the bench.

6.2 Scintillation Counters

These instruments work on the principle that certain substances (termed *scintillators* or *fluors*) emit light upon absorption of ionizing radiations. The scintillators can be solids (e.g. sodium iodide) or liquids (e.g. toluene). The light is detected by one or more photomultiplier tubes that give an electrical signal to a scaler. The number of electrical signals relates to the number of disintegrations in the sample, and the

strength of the signal relates to the energy of the absorbed radiation.

Scintillation probes are small handheld instruments useful for monitoring electromagnetic radiation (X or γ), say from ^{125}I. They either display c.p.s. values or, for dosimetry purposes, μSv h^{-1}. One of these instruments should be available whenever ^{125}I is used. As with ionization monitors, they should be calibrated officially once per year. You can calibrate them approximately yourself as an aid to monitoring (see Sect. 6.1).

Bench scintillation counters are designed for the analysis of large numbers of experimental samples. They either contain solid scintillators for detection of X or γ radiation or are designed for liquid scintillation counting. The latter instruments are found in most biochemical laboratories and are designed particularly for detecting β-emitters. The sample, usually in liquid form, is pipetted into a small container (or vial) to which a scintillation fluid is added.

There are many recipes for scintillation fluids (often referred to as cocktails). They are based on an organic solvent and one or more solutes termed *primary fluors* (when one solute is used) or *primary and secondary fluors* (when two solutes are used). The purpose of these fluors is to enhance the light output and shift the wavelength to the optimum for photomultiplier tubes. Most laboratories buy their cocktails ready prepared, and the manufacturer's catalogue should be carefully consulted. Essentially there are three forms of cocktail: a relatively straightforward formulation for counting organic samples, detergent-containing fluids for counting aqueous samples, and fluids that form gels for counting dispersed powders. The most commonly used ingredients are toluene or xylene as solvent, PPO (2,5-diphenyloxazole) or butyl-PBD (2-(4-t-butylphenyl)-5-(4-biphenyl)-l,3,4-oxadiazole) as primary solute, POPOP (l,4-bis(5-phenyloxazol-2-yl)benzene) as secondary solute and Triton-X100 as detergent.

An important concept in liquid scintillation counting is *quench*. This is anything that reduces counting efficiency, for example, colored compounds or chemicals such as chloroform that interfere with the scintillation process. Because a range of experimental samples may be subject to differing degrees of quench, it is necessary to calculate the counting efficiency of every sample individually. This can be done by re-counting samples following the addition of known amounts of radioactivity (an internal standard). However, this is tedious where there are a large number of samples, and contributes to the problem of waste disposal. Consequently, instruments have in-built means for automatically determining the counting efficiency. This works by counting a solid source of radioactivity within the machine that irradiates the sample vial. When quenching occurs, the energy spectrum of the radiation, as detected by the photomultipliers, appears to shift to lower-energy values. Analysis of this shift results in an assessment of counting efficiency within the sample vial. The instrument then corrects for the efficiency and prints out data as d.p.m. This system works well in most cases. However, it is essential to realize that it can be relied upon only when the sample is in solution in the vial, and is not appropriate for heterologous systems such as counting filter papers placed in scintillation fluid. This is because the filter paper will absorb some radiation, while the automatic quench correction mechanism will be recording only the properties of the solution. To accurately assess counting efficiency on filter

papers, the paper has to be removed after counting, a known aliquot of radioisotope applied, the filter re-dried, and counted again. The increase in counts divided by the known added radioactivity provides the counting efficiency.

The ability of scintillation counters to analyze the energy of radiations is of great value. Radioisotopes have characteristic energy spectra. If these are sufficiently different, a scintillation counter can detect radioisotopes separately, even when they are present in the same vial. Beta spectra are all similar in shape but different radioisotopes have differing maximum energies (see Table 2); thus ^{3}H, ^{14}C, and ^{32}P for example, can be distinguished by their energy spectra. This means that dual labeling experiments with these radionuclides is feasible.

The only way to monitor for ^{3}H is to perform wipe tests with damp tissues, and count in a liquid scintillation counter, but if other forms of contamination were present, they would be detected by the counter on the basis of spectral analysis. The fact that different isotopes have differing energies means that scintillation counters can be used to some extent to determine types of contamination (e.g. to determine whether the contamination is ^{3}H or ^{14}C).

Liquid scintillation counting sometimes presents two problems, *photoluminescence* and *chemiluminescence*, that result in light output from scintillation fluid that is not the result of radioactivity. That is, the c.p.m. will exceed the true d.p.m. It is caused by storage of scintillation fluid in bright sunlight, or by certain chemicals. Most modern counters are able to detect it and make a note to that effect on the printout. If in doubt, samples can be stored in the dark for a few hours, and then recounted.

6.3 Cerenkov Counting

High-energy β-particles travel through water faster than light; they cause polarization of molecules along their path, and these emit photons of light as they return to a ground state. The energy of radiation from ^{32}P is high enough to cause the Cerenkov effect. Consequently water can be used in place of scintillation fluid for this radioisotope. The counting efficiency is relatively low (about 30% compared with >90% in a scintillation cocktail) and the counter should be specifically calibrated for it, although an acceptable shortcut is to count the samples as if they were ^{3}H. The big advantages of Cerenkov counting are: aqueous samples can be counted and recovered for experimental use, and there is no accumulation of organic radioactive waste. It is also very inexpensive.

6.4 Autoradiography

Many experiments in molecular biology rely on this technique, in particular, analysis of nucleic acid samples separated on gels (southern and northern blotting) or detection of particular clones or plaques by hybridization screening. The basic principle underlying autoradiography is that light, a β-particle, X ray, or γ-ray can produce silver atoms in the crystals of silver halide that are present in photographic emulsions. This process is reversible if only a single atom of silver produced from one photon, forms in a crystal. However, silver atoms catalyze the conversion of silver halide to silver by the chemicals in the developer; this stabilizes the image.

Tab. 5 Use and sensitivity of autoradiography and fluorography.

Isotope	*Method*	*Example applications*	*Preflashed film*	*Inensifying screen*	*Temp. of exposure*	*Approximate detection limit after 24 h exposure [dpm/cm^{-2}]*
^{3}H	Direct	Whenever very high resolution is required, e.g. *in situ* hybridization	No	No	Room temp.	8×10^{6}
	Fluorography	Acrylamide gels	Yes	No	−70 °C	8×10^{3}
^{14}C, ^{35}S or ^{33}P	Direct	DNA sequencing	No	No	Room temp.	6×10^{3}
	Indirect	SDS-polyacrylamide gels	Yes	Yes	−70 °C	4×10^{2}
^{32}P	Direct	DNA sequencing	No	No	Room temp.	5×10^{2}
	Indirect	Southern analysis	Yes	Yes	−70 °C	5×10^{1}
^{125}I	Direct	Subcellular localization	No	No	Room temp.	2×10^{3}
	Indirect	Western blotting	Yes	Yes	−70 °C	1×10^{2}

If the radiation is of very low energy (e.g. from ^{3}H) it cannot penetrate very far and would, for example, be absorbed by the matrix of an agarose or polyacrylamide gel. Conversely, if radiation is very penetrating (e.g. from ^{32}P or ^{125}I) it passes through film without giving up much of its energy and is less efficiently detected than might be expected. To combat these difficulties there are techniques called *fluorography*, where samples such as gels are impregnated with fluors such as PPO (see Sect. 6.2), and *indirect autoradiography*, where intensifying screens partly absorb penetrating radiation and emit light. These adaptations increase sensitivity but reduce resolution. In both of these cases, exposure at low temperature is required to stabilize the latent image. In addition, a brief (<1ms) flash of light is required to preexpose the film and make it more responsive to low signals. Specific films are sold for direct or indirect autoradiography. A summary of some of the key features of autoradiography and fluorography (including use of intensifying screens) is provided in Table 5.

As discussed in Sect. 2, it is a requirement that if alternatives to radioisotopes are available, they should be used wherever possible. Nonradioactive labeling methods for DNA and proteins are improving all the time. Light-based detection techniques seem particularly successful in western blotting, for example. In the case of ^{32}P, a general rule at present is that if acceptable results are obtained within a 24-hour exposure to autoradiography, then a nonradioactive alternative should be sought. If longer exposures are required, particularly with intensifying screens, then it is likely that ^{32}P is still required for meeting the necessary sensitivity parameters.

6.5 Personal Dosimeters

These are either film badges or thermoluminescent detectors (TLDs). The latter contain a phosphor such as LiF that becomes and remains excited once exposed to radiation. Heat treatment results in light emission, the intensity of which relates to the dose received. Small TLDs can be worn on the fingers, under gloves, for hand dosimetry.

Personal dosimeters are not appropriate for workers using low-energy emitters such as ^{14}C, ^{35}S or ^{3}H but they are of value to users of ^{32}P or ^{125}I. Their issue should beon the advice of senior staff with appropriate training and responsibility for radiation protection, after dosimeter calculations and estimates of exposure levels to individuals.

See also Labeling, Biophysical; Mass Spectrometry-based Methods of Proteome Analysis; Nucleic Acid Hybrids, Formation and Structure of.

Bibliography

Books and Reviews

Ballance, P.E., Day, R.L., Morgan, J. (1992) *Phosphorus – 32: Practical Radiation Protection H and H Scientific Consultants*, Leeds, England.

Mundy, C.R., Cunningham, M.W., Read, C.A. (1991) Nucleic Acid Labelling and Detection, in: Brown, T.A. (Ed.) *Essential Molecular Biology a Practical Approach*, Oxford University Press, Oxford and New York.

Slater, R.J. (Ed.) (2002) *Radioisotopes in Biology*, 2nd edition, Oxford University Press, Oxford and New York.

Primary Literature

International Commission on Radiological Protection. (1991) *Recommendations of the International Commission on Radiological Protection*, ICRP Publication 60, Annals of the ICRP 21 Nos 1–3, Pergamon Press, Oxford.

International Commission on Radiological Protection. (1994) *Dose Coefficients for Intakes of Radionuclides by Workers*, ICRP Publication 68, Annals of the ICRP 24 No 4, Pergamon Press, Oxford.

Rat Genome (*Rattus norvegicus*)

Kim C. Worley and Preethi Gunaratne
Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA

1 The Rat and the Rat Genome 436

2 The Assembly Strategy and Results 437

3 Features of the Rat Genome 441
3.1 Genome Size 441
3.2 Telomeres and Centromeres 441
3.3 Orthologous Chromosomal Segments and Large-scale Rearrangements 443
3.4 Segmental Duplications 443
3.5 Gains and Losses of DNA 445
3.6 Substitution Rates 448
3.7 G + C Content and CpG Islands 449
3.8 Shift in Substitution Spectra between Mouse and Rat 449
3.9 Evolutionary Hotspots 450
3.10 Covariation of Evolutionary and Genomic Features 452

4 Evolution of Genes 452
4.1 Construction of Gene Set and Determination of Orthology 452
4.2 Properties of Orthologous Genes 453
4.3 Indels and Repeats in Protein-coding Sequences 454
4.4 Transcription-associated Substitution and Asymmetry 455
4.5 Conservation of Intronic Splice Signals 456
4.6 Gene Duplications 456
4.7 Conservation of Gene-regulatory Regions 459
4.8 Pseudogenes and Gene Loss 459
4.9 *In Situ* Loss of Rat Genes 460
4.10 Noncoding RNA Genes 462

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

5 Evolution of Transposable Elements 462
5.1 LINE-1 Activity in the Rat Lineage 462
5.2 Different Activity of SINEs in the Rat and Mouse Lineage 464
5.3 Colocalization of SINEs in Rat and Mouse 465
5.4 Endogenous Retroviruses and Derivatives 466
5.5 Simple Repeats 466
5.6 Prevalent, Medium-length Duplications in Rodents 467

6 Rat-specific Biology 467
6.1 Chemosensation 467
6.2 α_{2u} Globulin Pheromones 468
6.3 Detoxification 469
6.4 Proteolysis 469

7 Human Disease Gene Orthologs in the Rat Genome 469

8 Summary 474

Acknowledgment 475

Bibliography 475
Books and Reviews 475
Primary Literature 475

Keywords

BAC
A bacterial artificial chromosome, a large-insert (~200 kb) cloning vector for genomic sequences.

Contig
A contiguous set of overlapping segments of DNA.

Finished sequence
Complete, contiguous sequence generated with an accuracy of 1 error per 10 000 bp.

Genome
The entire complement of nuclear DNA in an individual or the representative sequence composite from several individuals.

Nonsynonymous
Nucleotide changes in the coding region of a gene that change the amino acid sequence of the translated protein. Some amino acid changes are more deleterious to

the protein function than others. The nonsynonymous substitution rate is the number of nonsynonymous substitutions per nonsynonymous site K_A.

ORF
Open reading frame – a sequence that translates without internal stop codons into a protein sequence.

Orthologous
Sequence regions in different organisms that originated from the same sequence in the last common ancestor of the organisms.

Scaffold
A set of contigs with sequence gaps between them that are linked by mate-pair information or marker information that may or may not give an estimated gap size.

Synonymous
Nucleotide changes in the coding region of a gene that do not change the amino acid sequence of the translated protein. The synonymous substitution rate is the number of synonymous substitutions per synonymous site K_S.

Whole-genome Shotgun (WGS)
Sequence generated randomly from a genome. Usually sheared genomic DNA is subcloned into a library in plasmid vectors, the library is then sampled and clones sequenced from both ends of the inserts.

The Brown Norway rat was the third mammalian genome to be sequenced. The three-way comparison of the human, rat, and mouse sequences resolves details of mammalian evolution and allows divergence events to be placed on different branches of the evolutionary tree. The comparison of the human to invertebrate and rodent genomes highlights the consequences of evolution over 1000 million years and 75 million years, while the comparison of the two rodent genomes describes changes that occurred in the 12 to 24 million years since the common ancestor of the rat and mouse.

A number of insights came from these comparisons:

The rat genome is 2.75 gigabases (Gb), smaller than human (2.9 Gb), and slightly larger than mouse (2.6 Gb). The three genomes encode similar numbers of genes. The majority of the genes have persisted without deletion or duplication and with well-conserved intron–exon structures. The exceptions are members of gene families that have expanded through gene duplication. Genes found in rat, but not mouse included genes producing pheromones, or involved in immunity, chemosensation, detoxification, or proteolysis.

Human genes known to be associated with disease have orthologs in the rat, but their rates of synonymous substitutions are significantly different from other genes.

Three percent of the rat genome is in large segmental duplication, primarily located near the centromeres. Expansions of major gene families are due to these genomic duplications.

About 40% of the rat genome aligns orthologously to mouse and human; this eutherian core of the genome contains the vast majority of exons and known regulatory elements (which comprise 1–2% of the genome). Only a portion of this core (5–6%) appears to be under selective constraint in rodents and primates, while the remainder appears to be evolving neutrally. Outside the eutherian core, the majority of the 30% of the rat genome that aligns only with mouse is rodent-specific repeats. More than half of the nonaligning sequence is rat-specific repeats.

More genomic changes have occurred in the rodent lineages than in the primate. Large rearrangements include approximately 250 rearrangements between a murid ancestor and human (the majority between the eutherian ancestor and the murid ancestor, and approximately 50 each from the murid ancestor to the rat and mouse). There is a threefold higher base substitution rate in neutral DNA along the rodent lineage than along the human lineage, with the rate on the rat branch 5 to 10% higher than along the mouse branch. Microdeletions are more frequent than microinsertions in both the rat and mouse branches. Most interestingly, there is a strong correlation between local rates of microinsertions and microdeletions, nucleotide substitutions, and transposable element insertions in the rat and mouse lineages, although the events occurred independently since the divergence of the two branches.

1
The Rat and the Rat Genome

The rat, although revered as the first sign in the Chinese zodiac and bearer of the Hindu god Ganesh, is a known carrier of over 70 diseases. Rats are involved in the transmission of several infectious diseases to man, including cholera, bubonic plague, typhus, leptospirosis, cowpox, and hantavirus infections. A major agricultural pest, rats and other rodents consume approximately one-fifth of the annual food harvest.

The laboratory rat (*Rattus norvegicus*) has contributed to human health by testing new drugs, and improving the understanding of essential nutrients and the pathobiology of human disease. It was the first mammal domesticated for scientific research (1828). The rat has been the model of choice for physiologists and nutritionists, and there are over 234 inbred strains developed to study genetic diseases.

The rat genome-sequencing project produced a draft sequence that, unlike mouse and human, would not ultimately be finished to remove all sequence gaps and produce a high base accuracy. For this reason, the quality of the draft was important. Although gaps remained, the overall sequence quality supported detailed analyses.

The sequence was generated from two inbred females, (BN/SsNHsd/Mcwi) by a network of centers led by the Human Genome Sequencing Center, Baylor College of Medicine. An international team representing over 20 groups contributed to the analysis reported and summarized here.

2 The Assembly Strategy and Results

The rat genome project used a combined WGS (whole-genome shotgun) with BAC clone strategy (Fig. 1). The project benefited from the logistically simpler WGS sequence generation and the local sequence assembly to resolve duplications afforded by BAC clones. The Atlas assembly suite, designed to combine these data sets, provided a BAC fisher to present localized combined BAC + WGS data to the public prior to the availability of the complete assembly.

Over 44 million DNA sequence reads were generated (Table 1). After removal of low-quality reads and vector contaminants, 36 million reads were used in the assembly where 34 million reads were retained. This was a sevenfold sequence coverage, with 60% of the reads from the WGS. Coverage estimates range from 7.3x (when estimated from the entire ''trimmed'' length of the sequence data) to 6.9x (when estimated from the sequence with quality of Phred20 or higher).

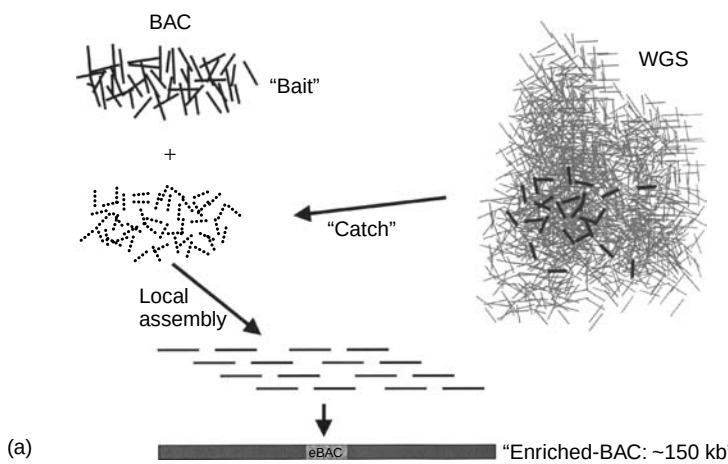

Fig. 1 The new ''combined'' sequence strategy and Atlas software. (a) Formation of ''eBACs.'' The strategy combined the advantages of both BAC and WGS sequence data. Modest sequence coverage (~1.8-fold) from a BAC is used as ''bait'' to ''catch'' WGS reads from the same region of the genome. These reads, and their mate pairs, are assembled using Phrap to form an enriched BAC or ''eBAC.'' This stringent local assembly retains 95% of the ''catch.'' (b) Creation of higher-order structures. Multiple eBACs are assembled into bactigs on the basis of sequence overlaps. The bactigs are joined into superbactigs by large clone mate-pair information (at least 2 links), extended into ultrabactigs using additional information (single links, FPC contigs, synteny, markers), and ultimately aligned to genome mapping data (RH and physical maps) to form the complete assembly. Used with permission from *Nature* **428**, 493–521 (2004).

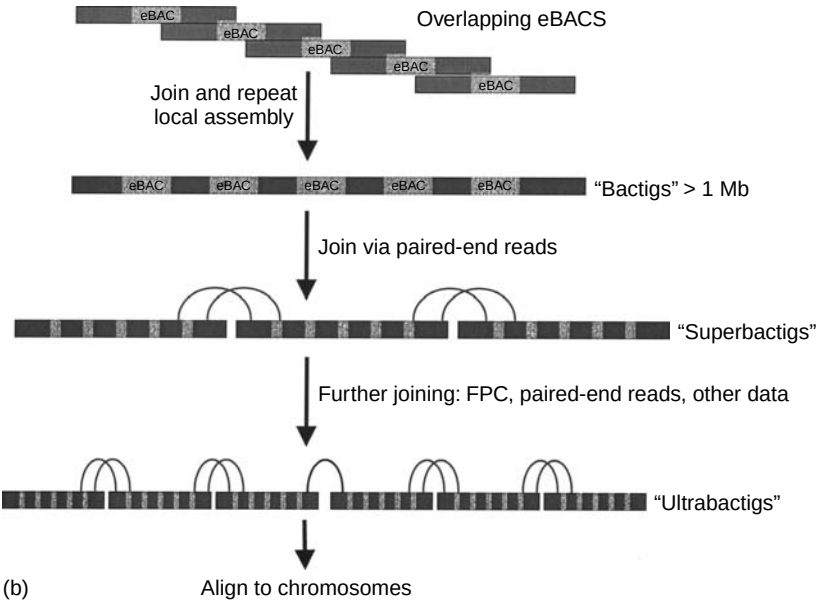

Fig. 1 (*Continued*)

Simultaneous to the sequencing, a "fingerprint contig" (FPC) map was developed, which was used in combination with the ongoing sequencing to identify BACs for sequence skimming. The parallel development of mapping and sequencing resources permitted the data-gathering phase of the project to be completed in less than two years.

The statistics of the rat draft assembly (v. 3.1) are given in Table 2. The current assembly (v. 3.4) splices in a number of finished BAC sequences to the v. 3.1 assembly. Much of the gene and protein feature analysis was developed using earlier assemblies (v. 2.0 and 2.1), while the genome description is based on v. 3.1.

The majority of the genome is in contigs larger than the expected mammalian gene (N50 = 38 kb). These contigs are linked into 783 larger scaffolds; those anchored to the radiation hybrid map had an N50 of 5.4 Mb, while the smaller scaffolds that could not be anchored had an N50 of 1.2 Mb.

The quality of the assembled sequence was assessed using sequence from finished BACs, the comparison showed that the bases within contigs were of high quality (1.32 mismatches per 10 kb), similar to the finished sequence. The majority of mismatches occurred at the ends of contigs in regions that average 750 bp and total <0.9% of the genome. Only six mismatches (insertions or deletions) were found within contigs when compared to 13 Mb of finished sequence or one per 2.2 Mb.

The assembly accuracy was judged in comparison to linkage and radiation hybrid maps. The majority of the genetic markers (13/3824) and sequence-tagged sites (96.9%) had consistent chromosome placement (Fig. 2). The maps are congruent with the assembly except for possible mismapped markers.

Tab. 1 Clones and reads used in the RGSP.

Insert size [kb][a]	*Source or vector*	*Reads [millions]*				*Bases [billions]*		*Sequence coverage*[b]		*Clone coverage*[c]
		All[d]	*Used*	*Paired*	*Assembled*	*Trimmed*	*≥Phred20*	*Trimmed*	*≥Phred20*	
2–4	Plasmid	9.6	8.6	7.4	7.9	4.8	4.5	1.8	1.6	3.70
4.5–7.5	Plasmid	4.5	4.3	3.6	3.6	2.4	2.3	0.87	0.82	2.96
10	Plasmid	8.4	7.2	6.4	6.4	4.1	3.8	1.5	1.4	11.63
50	Plasmid	1.7	1.3	1.0	1.1	0.69	0.65	0.25	0.24	9.47
150–250	BAC	0.32	0.31	0.26	0.26	0.18	0.16	0.07	0.06	9.26
Total WGS		24.5	21.7	18.7	19.2	12.1	11.3	4.4	4.1	37.0
2–5	BAC skims	19.6	14.6	13.2	14.5	8.0	7.7	2.9	2.8	4.8[e]
Total		44.1	36.3	31.9	33.7	20.2	19.0	7.3	6.9	41.8

[a]Grouped in ranges of sizes for individual libraries tracked to specific multiples of 0.5 kb.
[b]Total bases in used reads divided by sampled genome size including all cloned and sequenced euchromatic or heterochromatic regions.
[c]Estimated as sum of insert sizes divided by sampled genome size.
[d]WGS reads available on the NCBI Trace Archive as of March 21, 2003; BAC skim reads attempted at BCM–HGSC as of May 12, 2003; BAC end reads obtained directly from TIGR.
[e]Refers to coverage from 2–5 kb subclones from BACs. The BACs that were skimmed amounted to 1.58x clone coverage.

Tab. 2 Statistics of the RGSP draft sequence assembly.

Features[a]	*Number*	*N50 Length [kb]*	*Bases [Gb]*	*Bases plus gaps [Gb]*[b]	*Percentage of genome*[c]			
					Sampled [2.78 Gb]		*Assembled [2.75 Gb]*	
					Bases	*Bases + Gaps*	*Bases*	*Bases + Gaps*
Anchored contigs	127 810	38	2.476	2.481	89.1	89.2	90.0	90.2
Anchored superbactig scaffolds	783	5 402	2.476	2.509	89.1	90.3	90.0	91.2
Anchored ultrabactigs	291	18 985	2.476	2.687	89.1	96.6	90.0	97.7
Unanchored superbactigs, main scaffolds	134	1210	0.056	0.062	2.0	2.2	2.0	2.3
Unanchored ultrabactigs	128	1529	0.056	0.069	2.0	2.5	2.0	2.5
All superbactigs, main scaffolds	917	5301	2.533	2.571	91.1	92.5	92.1	93.5
Minor scaffolds	4345	8	0.033	0.038	1.2	1.4	1.2	1.4

[a]Anchored sequences are those that can be placed on chromosomes because they contain known markers. The main scaffold for each superbactig is the largest set of contigs (in terms of total contig sequence), which can be ordered and oriented using mate-pair links and ordering of BACs. Scaffolds, which cannot be ordered and oriented with respect to the main scaffold, are termed *minor scaffolds*.
[b]Ambiguous bases (Ns) are counted in the gap sizes, and excluded in the base counts.
[c]Computed as bases plus gaps divided by estimated genome size. Sampled genome size is based on oligonucleotide frequency statistics of unassembled WGS reads. Assembled genome size is based on cumulative contig sequence following assembly.

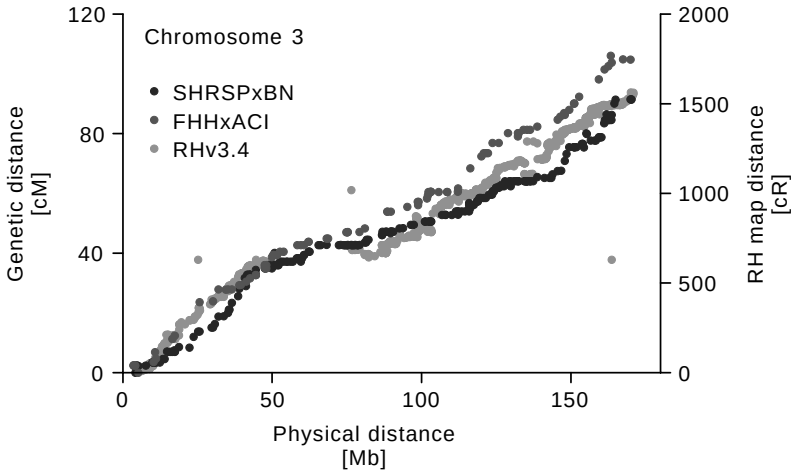

Fig. 2 Map correspondence. Correspondence between positions of markers on two genetic maps of the rat (SHRSPxBN intercross and FHHxACI intercross), on the rat radiation hybrid map, and their position on the rat genome assembly (Rnor3.1). Used with permission from *Nature* **428**, 493–521 (2004). (See color plate p. xxxix.)

3 Features of the Rat Genome

3.1 Genome Size

Genome assemblies are usually smaller than the actual genome size owing to underrepresentation of sequences due to cloning bias and sequencing and assembly difficulties. However, equating assembled genome size with the euchromatic, clonable portion of the genome (CpG) does not take into account the heterochromatic sequence included in the assembly. The rat genome size was estimated by two methods: scaling the assembled genome size by the fraction of features found in the assembly, and measuring the clonable or sampled genome size on the basis of the distribution of short oligomers in the WGS reads before assembly and in the assembly. Both estimates gave a relatively consistent measure of estimated genome size of 2.75 Gb. This conservative estimate was still considerably higher than the size estimated for the mouse draft genome sequence, which has different repeat content and appears to have underrepresented segmental duplications because of technical reasons.

3.2 Telomeres and Centromeres

The rat has both metacentric and telocentric chromosomes, unlike the wholly telocentric mouse chromosomes. As expected, the draft sequence does not contain complete telomere and centromere sequences. The approximate physical location of the centromeres relative to the genomic sequence is shown in Fig. 3. Several of the putative centromere positions coincide with segmental duplication blocks and classical satellite repeat clusters, consistent with enrichment of these sequence features in rat pericentromeric DNA. Human subtelomere regions are characterized by an abundance of segmentally duplicated DNA and an enrichment of internal (TTAGGG)n-like sequence islands.

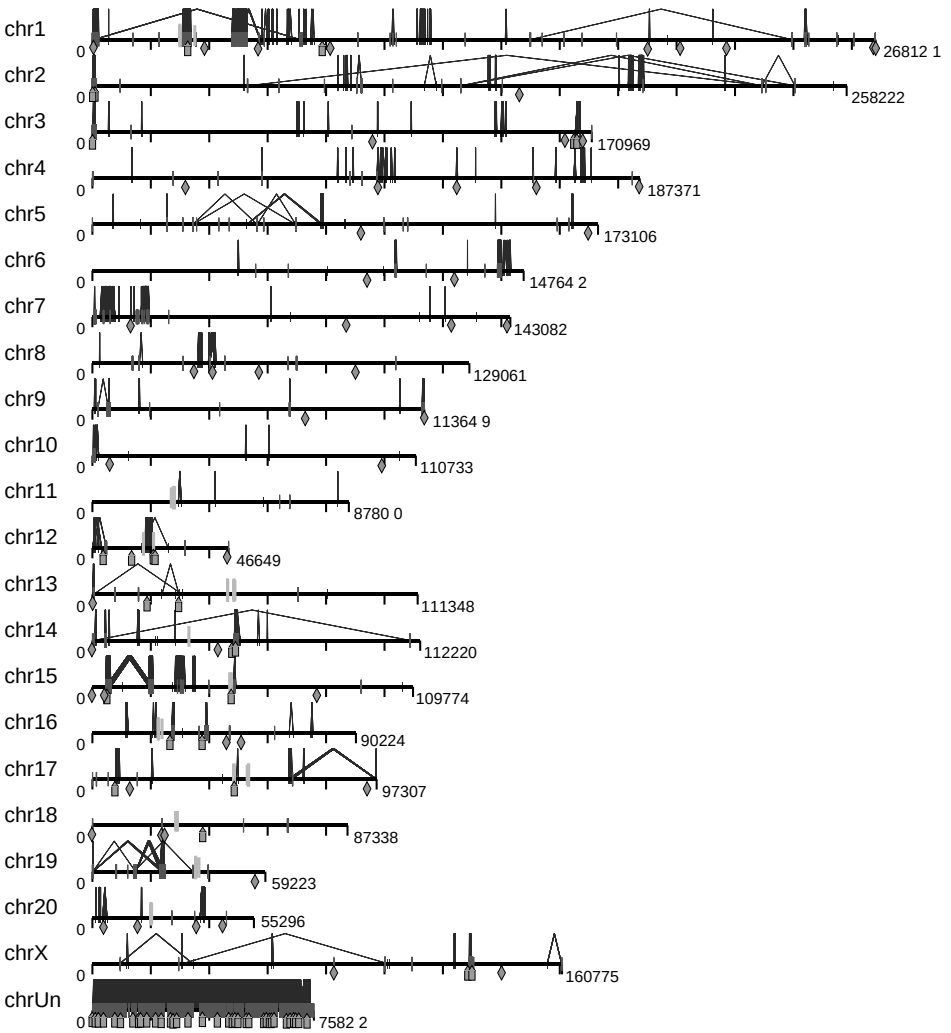

Fig. 3 Distribution of segmental duplications in the rat genome. Interchromosomal duplications (red) and intrachromosomal duplications (blue) are depicted for all duplications with ≥90% sequence identity and ≥20 kb length. The intrachromosomal duplications are drawn with connecting blue line segments; those with no apparent connectors are local duplications very closely spaced on the chromosome (below the resolution limit for the figure). P arms on the left and the q arms on the right. Chromosomes 2, 4–10, and X are telocentric; the assemblies begin with pericentric sequences of the q arms, and no centromeres are indicated. For the remaining chromosomes, the approximate centromere positions were estimated from the most proximal STS/gene marker to the p and q arm as determined by fluorescent *in situ* hybridization (FISH) (cyan vertical lines; no chromosome 3 data). The chrUn sequence is contigs not incorporated into any chromosomes. Green arrows indicate 1-Mb intervals with more than tenfold enrichment of classic rat satellite repeats within the assembly. Orange diamonds indicate 1-Mb intervals with more than tenfold enrichment of internal (TTAGGG)*n*-like sequences. For more details, see http://ratparalogy.cwru.edu. Used with permission from *Nature* **428**, 493–521 (2004). (See color plate p. xlvi.)

Approximately one-third of the euchromatic rat subtelomeric regions contain similar features, suggesting that Rnor3.1 might extend very close to the chromosome ends.

3.3 Orthologous Chromosomal Segments and Large-scale Rearrangements

Multimegabase regions of chromosomes have been passed from the primate-rodent ancestor to human and murid rodent descendants with minimal rearrangements of gene order. These regions, bounded by the breaks that occurred during ancient large-scale chromosomal rearrangements are referred to as *orthologous chromosomal segments*. The sequence of these rearrangements was tentatively reconstructed using the human genome and other outgroup data (see Fig. 4). Inspection shows events that preceded and follow the rat–mouse divergence are interleaved. At 1-Mb resolution, multiple methods report virtually indistinguishable sets of orthologous chromosomes segments: 278 between human and rat, 280 between human and mouse, and 105 between rat and mouse. The larger numbers of breaks in orthologous segments between the human and the rodents is expected because of the greater evolutionary distance.

Understanding the number and timing of rearrangement events that have occurred in each of the three individual lineages (see tree in Fig. 5a) since the common primate-rodent ancestor required a more detailed analysis. The X chromosome is presented here as an example; its history is easier to trace completely since rearrangements between the X and the autosomes are rare. There are 16 human–mouse–rat orthologous segments of at least 300 kb in size (Fig. 6a). The most parsimonious scenario requires 15 inversions in the descent from the primate-rodent ancestor. Outgroup data from cat, cow, and dog resolve the timing of these rearrangements more precisely. Most of these events occurred in the rodent lineage: five (or four) before the divergence of rat and mouse, five in the rat lineage and five in the mouse lineage. At most one rearrangement occurred in the human lineage since divergence from the common ancestor with rodents. The analysis of the whole genome showed similar results. The assignment of the considerable rearrangement activity to the rodent branch following primate-rodent divergence is consistent with previous, lower resolution studies.

3.4 Segmental Duplications

Segmental duplications are defined here as genomic regions of at least 5 kb in length that are repeated with >90% identity remaining between the copies. The rat has approximately 2.9% of its bases in these duplicated regions, whereas the human has 5 to 6%, and the mouse has 1.0 to 2.0%. These duplicated structures are particularly challenging to assemble, so some of the mouse–rat difference is attributable to the BAC-based approach used in the rat assembly compared to the WGS-mouse approach. Most of these regions have less than 99.5% identity and are, therefore, not simply overlapping sequences that were not joined by the assembly program. Nearly 44% of these blocks of segmental duplications are mapped to the "unplaced" chromosome in Rnor3.1 indicating the difficulty of anchoring these elements to the genome.

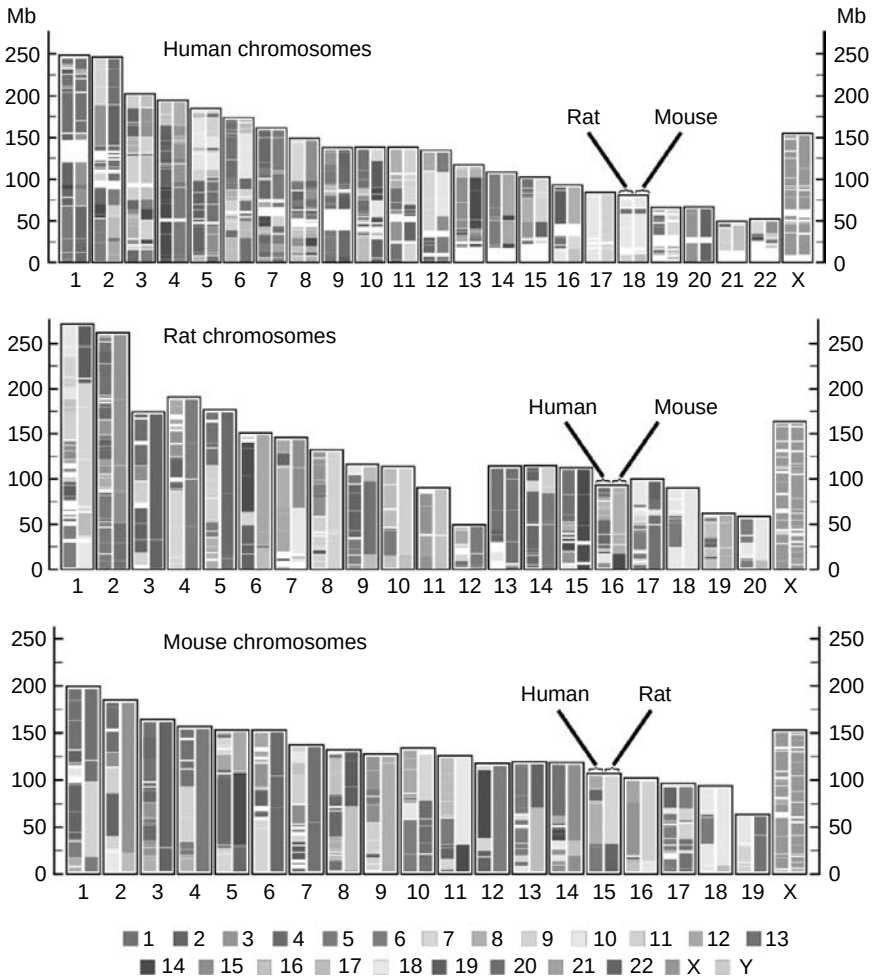

Fig. 4 Map of conserved synteny between the human, mouse, and rat genomes: For each species, each chromosome (*x*-axis) is a two column boxed pane (p-arm at the bottom) colored according to conserved synteny to chromosomes of the other two species. The same chromosome color code is used for all species (indicated below). For example, the first 30 Mb of mouse chromosome 15 is shown to be similar to part of human chromosome 5 (by the red in left column) and part of rat chromosome 2 (by the olive in right column). An interactive version is accessible (http://www.genboree.org). Used with permission from *Nature* **428**, 493–521 (2004). (See color plate p. xlvii.)

Intrachromosomal duplications are three times more common than interchromosomal duplications and are significantly enriched near the telomeric and centromeric regions (Fig. 3). The pericentromeric accumulation of segmental duplications in the rat seems to be a general property of mammalian chromosome architecture.

There is considerable clustering of segmental duplications. For many of the largest clusters, the underlying sequence

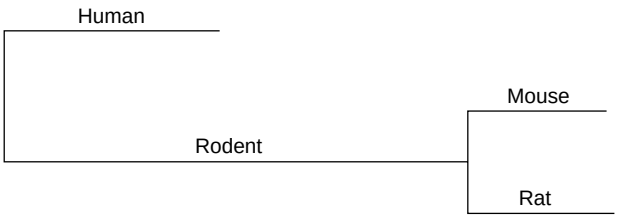

(a)

	Substitutions, insertions, and deletions			
	Human	Rodent	Mouse	Rat
Substitutions per site	0.11 ±0.0012	0.24 ±0.0012	0.073 ±0.0014	0.077 ±0.006
in neutral sites only	0.13 ±0.011	0.28 ±0.033	0.083 ±0.013	0.091 ±0.011
Insertion event per kb	2.7 ±0.94	4.74 ±1.0	1.54 ±0.84	1.43 ±0.73
Deletion events per kb	5.3 ±0.55	12 ±1.2	3.8 ±0.21	4.5 ±0.13
Inserted bases per kb	6.4 ±2.9	9.4 ±1.6	3.6 ±1.5	3.2 ±1.3
Deleted bases per kb	18 ±2.0	40 ±4.9	11 ±0.55	13 ±0.05

(b)

Fig. 5 Substitutions and microindels (1–10 bp) in the evolution of the human, mouse, and rat genomes: (a) The lengths of the labeled branches in the tree (top panel) are proportional to the number of substitutions per site inferred from all sites with aligned bases in all three genomes. (b) The table shows the midpoint and variation in these branch length estimates when estimated from different sequence alignment programs and different neutral sites, including sites from ancestral repeats, fourfold degenerate sites in codons, and rodent-specific sites (''in neutral sites only'' row). Other rows give midpoints and variation for microindels on each branch of the tree. Used with permission from *Nature* **428**, 493–521 (2004).

alignments show a wide range of sequence identity, suggesting that duplication events have occurred continuously over millions of years. In contrast, an analysis of all duplicated regions showed a bimodal distribution consistent with bursts of segmental duplication (particularly intrachromosomal duplication) that occurred approximately 5 and 8 Myr ago.

The segmental duplications of the rat genome were of considerable interest because they represent an important mechanism for the generation of new genes. Sixty-three (of 4532 total) NCBI reference sequence genes were completely or partially located within duplicated regions. As discussed in the following, many of these genes present in multiple copies, belong to recently duplicated gene families, and contribute to distinctive elements of rat biology.

3.5 Gains and Losses of DNA

In addition to large rearrangements and segmental duplications, genome architecture is strongly influenced by insertion and deletion events that add and remove DNA over evolutionary time. To characterize the origins and losses of sequence elements in the human, mouse, and rat genomes, the nucleotide bases were categorized using alignment data and annotations of the insertions of repetitive elements (Fig. 7).

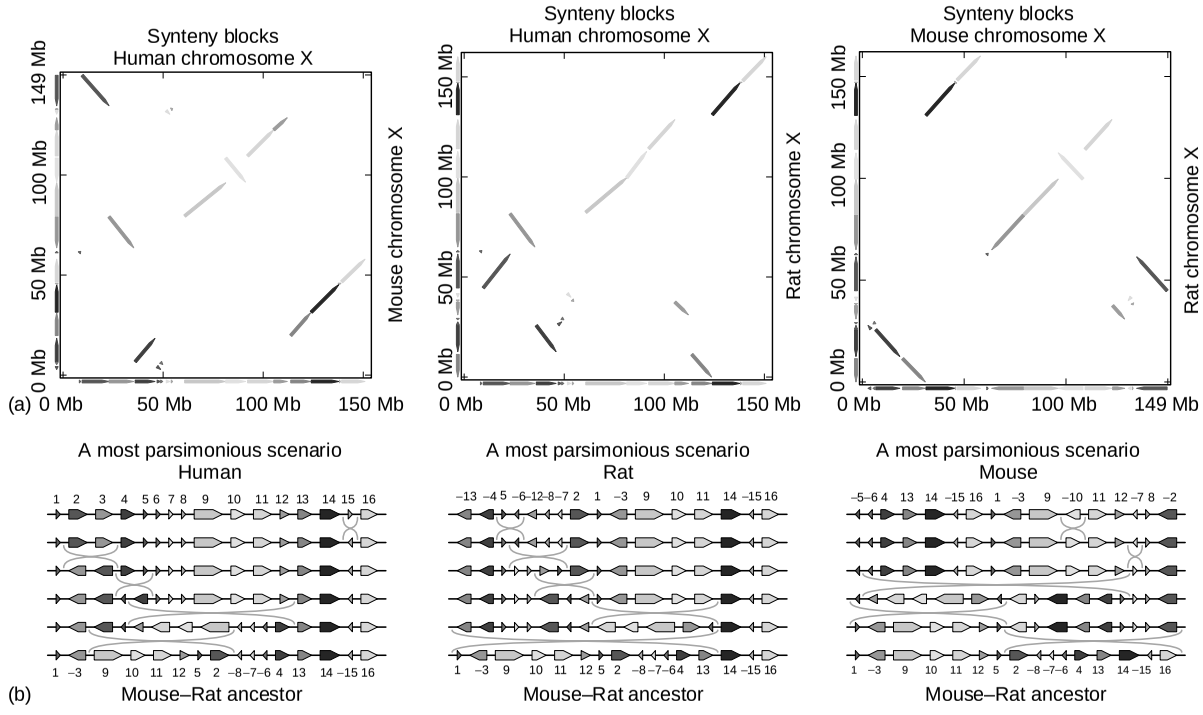

Fig. 6 X chromosome in each pair of species: (a) GRIMM–Synteny computes 16 three-way orthologous segments (≥300 kb) on the X chromosome of human, mouse, and rat, shown for each pair of species, using consistent colors. (b) The arrangement (order and orientation) of the 16 blocks implies that at least 15 rearrangement events occurred during X chromosome evolution of these species. The program MGR determined that evolutionary scenarios with 15 events are achievable and all have the same median ancestor (located at the last common mouse–rat ancestor). Shown is a possible (not unique) most parsimonious inversion scenario from each species to that ancestor. Note the last common ancestor of human, mouse, and rat should be on the evolutionary path between this median ancestor and human. Used with permission from *Nature* **428**, 493–521 (2004). (See color plate p. lv.)

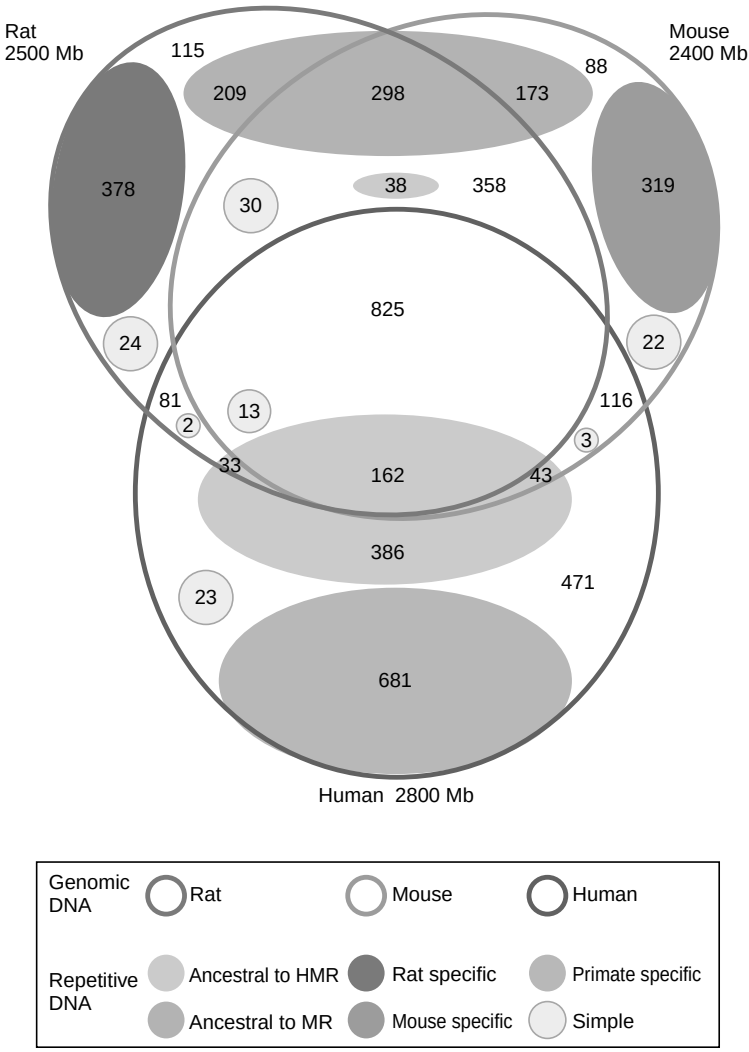

Fig. 7 Aligning portions and origins of sequences in rat, mouse, and human genomes. Each outlined ellipse is a genome, and the overlapping areas indicate the amount of sequence that aligns in all three species (rat, mouse, and human) or in only two species. Nonoverlapping regions represent sequences that do not align. Types of repeats classified by ancestry: those that predate the human–rodent divergence (gray), those that arose on the rodent lineage before the rat–mouse divergence (lavender), species-specific (orange for rat, green for mouse, blue for human), and simple (yellow), placed to illustrate the approximate amount of each type in each alignment category. Uncolored areas are nonrepetitive DNA – the bulk is assumed to be ancestral to the human–rodent divergence. Numbers of nucleotides (in Mb) are given for each sector (type of sequence and alignment category). Used with permission from *Nature* **428**, 493–521 (2004). (See color plate p. xlviii.)

The estimates of the amount of repeats represent lower bounds because some repeats, especially the older ones, are not recognized even though the rodent repeat database was greatly expanded by analyzing the rat and mouse genomes.

About a billion nucleotides (39% of the euchromatic rat genome) align in all three species, constituting an ''ancestral core,'' which is retained in these genomes. The ancestral core contains 94 to 95% of the known coding exons and regulatory regions. The levels of three-way conservation of ''mammalian ancestral repeats'' (transposon relics retained in all three species) confirm estimates that 5 to 6% of the human genome is accumulating substitutions more slowly than the neutral rate in the three lineages, and hence may be under purifying selection. In this constrained fraction, noncoding regions outnumber coding regions regardless of the constrain, an observation that supports comparative analyses on limited subsets of the genome. The preponderance of noncoding elements in the most constrained fraction of the genome underscores the likelihood that they play critical roles in mammalian biology.

About 700 Mb (28%) of the rat euchromatic genome aligns only with the mouse. At least 40% of this is rodent-specific repeats that are inserted on the branch from the primate-rodent ancestor to the murid ancestor. Some of the remainder is ancestral mammalian repeats or single copy DNA deleted in the human lineage but retained in rodents (Fig. 7). Although this 700 Mb of rodent-specific DNA is primarily neutral, it may also contain some functional elements lost in the human lineage in addition to sequences representing gains of rodent-specific functions, including some coding exons.

3.6 Substitution Rates

The alignment data allow relatively precise estimates of the rates of neutral substitutions and microindel events (<=10 bp). As in previous studies comparing human and mouse, both synonymous fourfold degenerate (''4D'') sites in protein-coding regions and sites in mammalian ancestral repeats were used in this analysis. In addition, the primarily neutral rodent-specific sites that did not align to human discussed earlier were used.

Estimates for the neutral substitution level between the two rodents range from 0.15 to 0.20 substitutions per site, while estimates for the entire tree of human, mouse, and rat range from 0.52 to 0.65 substitutions per site (Fig. 5). This difference was predictable because of the evolutionary closeness of the two rodents. Less predictable was the finding that for all classes of neutral sites analyzed, the branch connecting the rat to the common rodent ancestor is 5 to 10% longer than the mouse branch. Thus, for as yet unknown reasons, the rat lineage has accumulated substantially more point substitutions than the mouse lineage since their last common ancestor.

Four-way alignments with sequence from orthologous ancestral repeats in the three genomes and the repeat consensus sequences (to approximate the progenitor repeat sequence) were used to distinguish substitutions on the branches from the primate-rodent ancestor to either the human or rodent ancestor. This revealed a three-to-one speed-up in rodent substitution rates relative to human, larger than estimated in the mouse analysis but consistent with more recent studies, which also use multiple sequence alignments.

Estimates for rates of microdeletion events are, for all branches, approximately twofold higher than rates of microinsertion (Fig. 5b), suggesting a fundamental difference in the mechanisms that generate these mutations. Furthermore, there are substantial lineage specific rate differences. The rat lineage has accumulated microdeletions more rapidly than the mouse, while the opposite holds true for microinsertions. As with substitutions, both microinsertions and microdeletion rates are substantially slower in the human lineage. The size distributions of microindels (1–10 bp) were heavily weighted toward the smallest indels and similar for insertions, deletions, and both the mouse and rat branches.

3.7
G + C Content and CpG Islands

The G + C content of the rat varies significantly across the genome (Fig. 8a) in a distribution that more closely resembles that of mouse than human. The variation of G + C content is coupled with differences in the distribution of CpG islands – short regions that are associated with the 5′ ends of genes and gene regulation and that escape the depletion of CpG dinucleotides owing to deamination of methylated cytosine. The 2.6-Gb rat genome assembly contains 15 975 CpG islands in nonrepetitive sequences of the genome. This is similar to the 15 500 CpG islands reported in the 2.5-Gb mouse genome, but far fewer than the 27 000 reported in the human genome. The distribution of CpG islands by chromosome is shown in Fig. 8b.

The changes in CpG island content predate the rat-mouse split and are consistent with the accelerated loss of CpG dinucleotides in rodents compared with humans. It remains possible, however, that the greater number of human regions with extremely high G + C content are due to distributational changes occurring in the primate, rather than the rodent lineage.

3.8
Shift in Substitution Spectra between Mouse and Rat

The nonrepetitive fraction of the rat genome is enriched for G + C content relative to the mouse genome by ~0.35% over 1.3 billion nucleotides. This is a subtle but substantial difference that may be explained in part by differences in the spectra of mutation events that have accumulated in the mouse and rat lineages. The small minorities of substitutions where events can be assigned to either the mouse or the rat lineage by virtue of a nucleotide match in the alignment column between human and only one rodent were studied. Of the ~117 million alignment columns meeting this criteria, ~60 million involve a change in the rat lineage versus ~57 million in the mouse, reflecting the increase in the rates of point substitution in the rat lineage (Fig. 5b). Fifty percent of these events in rat involve a substitution from an A/T to a G/C compared to only 47% of all mouse changes. The complementary change, G/C to A/T, is more common in the mouse lineage (38% vs. 35% in rat). There is not a similar substantial difference between changes that do not alter G + C content. In addition, this bias is not confined to particular transition or transversion events, nor can it be explained simply as a result of divergent substitution rates of CpG dinucleotides. Thus, this shift appears to be a general

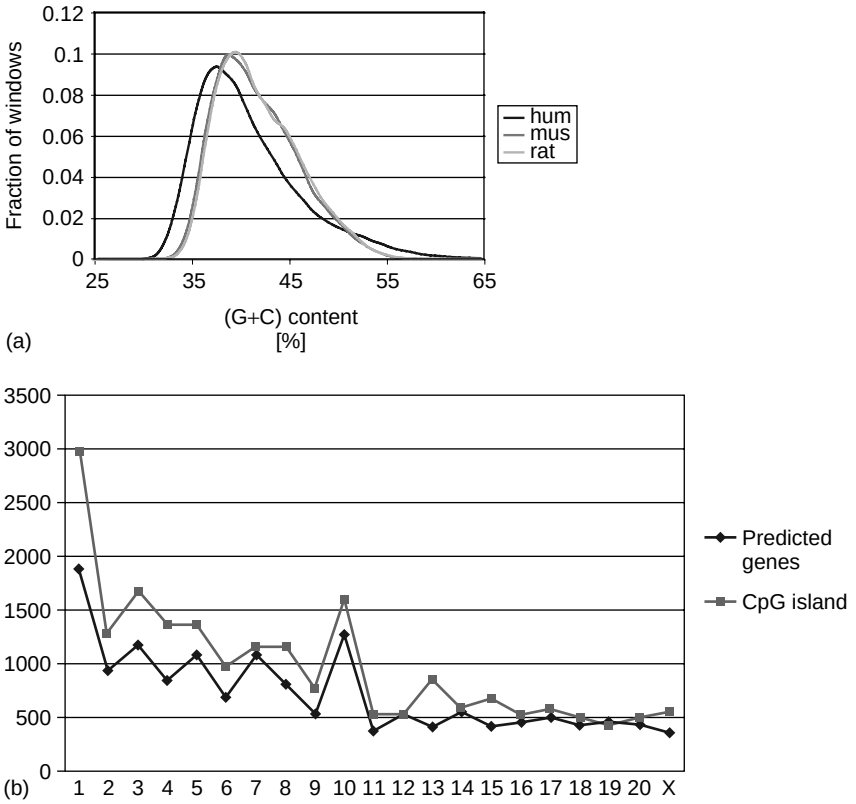

Fig. 8 Base composition distribution analysis. (a) The fraction of 20 kb nonoverlapping windows with a given G + C content is shown for human, mouse, and rat. (b) The number of Ensembl predicted genes per chromosome and the number of CpG islands per chromosome. The density of CpG islands averages 5.9 islands per Mb across chromosomes and 5.7 islands per Mb across the genome. Chromosome 1 has more CpG islands than other chromosomes, yet neither the island density nor ratio to predicted genes exceeds the normal distribution. The number of CpG islands per chromosome and the number of predicted genes are correlated ($R^2 = 0.96$). Used with permission from *Nature* **428**, 493–521 (2004). (See color plate p. xlix.)

change that results in an increase in G + C content in the rat genome. Biochemical changes in the repair or replication enzymes might be responsible, and the observation that recombination rates are slightly higher in rat than in mouse may suggest a role for G + C biased mismatch repair. However, population genetic factors, such as selection, cannot be ruled out.

3.9 Evolutionary Hotspots

Comparison of the two rodent genomes, using human as outgroup, reveals regions that are conserved yet under different levels of constraint in mouse and rat. These regions may have distinct functional roles and contribute to species-specific differences. In 5055 regions >=100 bp,

with at least a tenfold difference in the estimated number of substitutions per site on the mouse and rat branches and <0.25 substitutions per site on the human branch (to avoid fast-evolving regions), analysis found them enriched twofold in transcribed regions. Thirty-nine percent of mouse hot spots were found in 18% of the

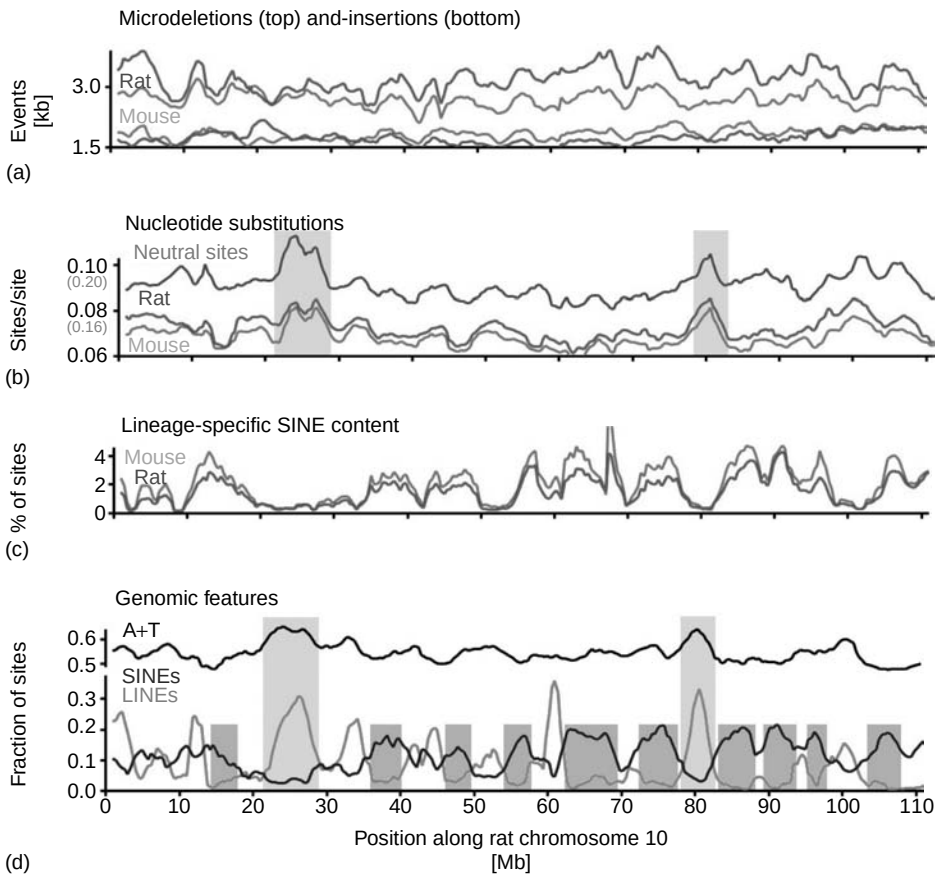

Fig. 9 Variability of several evolutionary and genomic features along rat chromosome 10. (a) Rates of microdeletion and microinsertion events (less than 11 bp) in the mouse and rat lineages since their last common ancestor, revealing regional correlations. (b) Rates of point substitution in the mouse and rat lineages. Red and green lines represent rates of substitution within each lineage estimated from sites common to human, mouse, and rat. Blue represents the neutral distance separating the rodents, as estimated from rodent-specific sites. Note the regional correlation among all three plots, despite being estimated in different lineages (mouse and rat) and from different sites (mammalian vs rodent-specific). (c) Density of SINEs inserted independently into the rat or mouse genomes after their last common ancestor. (d) A + T content of the rat, and density in the rat genome of LINEs and SINEs that originated since the last common ancestor of human, mouse, and rat. Pink boxes highlight regions of the chromosome in which substitution rates, AT content, and LINE density are correlated. Blue boxes highlight regions in which SINE density is high but LINE density is low. Used with permission from *Nature* **428**, 493–521 (2004). (See color plate p. l.)

mouse genome covered by RefSeq genes; and 17% of rat hotspots were found in 8% of the rat genome covered by RefSeq genes. Similar numbers were also observed with coding exons and EST regions. Half of all hotspots in the mouse genome lie totally in noncoding regions.

3.10 Covariation of Evolutionary and Genomic Features

A high-resolution analysis of the genomic and evolutionary landscape of rat chromosome 10 uncovered strong correlations between certain microevolutionary features. Strongly correlated, in particular, are the local rates of microdeletion ($R2 = 0.71$; Fig. 9a), microinsertion ($R2 = 0.56$; Fig. 9a), and point substitution ($R2 = 0.86$; Fig. 9b) between the two independent lineages of mouse and rat. In addition, microinsertion rates are correlated with microdeletion rates ($R2 = 0.55$; Fig. 9a). These strong correlations are also observed in an independent genome-wide analysis, both on the original data and after factoring out the effects of G + C content. Perhaps surprisingly, substantially less correlation is seen between microindel and point substitution rates (compare Figs. 9a and b).

The local point substitution rate in sites common to human, mouse, and rat strongly correlates with that in rodent-specific sites ($R2 = 0.57$; Fig. 9b, blue line versus red/green). These two classes of sites, while interdigitated at the level of tens to thousands of bases, constitute sites that are otherwise evolutionarily independent. This result confirms that local rate variation is not solely determined by stochastic effects and extends, at high resolution, the previously documented regional correlation in rate between 4D sites and ancestral repeat sites.

4 Evolution of Genes

A substantial motivation for sequencing the rat genome was to study protein-coding genes. Besides being the first step in accurately defining the rat proteome, this fundamental data set yields insights into differences between the rat and other mammalian species with a complete genome sequence. Estimation of the rat gene content is possible because of relatively mature gene-prediction programs and rodent-transcript data. Mouse and human genome sequences also allow characterization of mutational events in proteins such as amino acid repeats and codon insertions and deletions. The quality of the rat sequence also allows us to distinguish between the functional genes and pseudogenes.

An estimated 90% of rat genes possess strict orthologs in both mouse and human genomes. Studies also identified genes arising from recent duplication events occurring only in rat, and not in mouse or human. These genes contribute characteristic features to rat-specific biology, including aspects of reproduction, immunity, and toxin metabolism. In contrast, almost all human "disease genes" have rat orthologs. This underscores the importance of the rat as a model organism in experimental science.

4.1 Construction of Gene Set and Determination of Orthology

The Ensembl gene-prediction pipeline has predicted 20 973 genes with 28 516 transcripts, and 205 623 exons in rat. These genes contain an average of 9.7 exons, with a median exon number of 6.0. At least 20% of the genes are alternatively spliced, with an average of 1.3 transcripts predicted per

gene. Of the 17% single exon transcripts, 1355 contain frameshifts relative to the predicted protein and 1176 are likely processed pseudogenes. The majority of transcripts are supported by rodent (61%) or vertebrate (72%) transcript evidence, and have at least one untranslated region predicted (60%). The coding densities ranged from 1.2 to 2.2% and the coding fraction of RefSeq genes covered by these predictions ranged from 82 to 98%. The number of coding exons per gene and average exon length were similar in the three species. Differences were observed in intron length, with an average of 5338 bp in human, 4212 bp in mouse, and 5002 bp in rat.

4.2 Properties of Orthologous Genes

Orthology relationships were conservatively predicted and 12 440 rat genes showed clear, unambiguous 1 : 1 correspondence with a gene in the mouse genome. Accounting for potential errors 86 to 94% of the rat genes have 1 : 1 mouse orthologs. The remaining genes were associated with lineage-specific gene family expansions or contractions. Surprisingly, a similar proportion (89–90%) of rat genes possessed a single ortholog in the human genome, perhaps because of less resolution in the draft genomes. The majority of nucleotide changes within protein-coding regions that reflected synonymous or nonsynonymous substitutions yielded K_A/K_S ratios of less than 0.25 indicating purifying selection. Values of 1 suggest neutral evolution, and values greater than 1 indicate positive selection. Examination of ortholog pairs in orthologous genomic segments (Table 3) showed a slight increase in median K_S values, indicating that the rat lineage has more neutral substitutions in gene coding regions than the mouse lineage.

Rat genes shared with mouse, but with no counterparts in human are expected to reflect either a rapidly evolving gene set, or genes that may have arisen from noncoding DNA, or been converted to pseudogenes in the human lineage. Thirty-one ENSEMBL rat genes were collected that have no nonrodent homolog in current databases. These are twofold overrepresented among genes in paralogous gene clusters, and threefold overrepresented among genes whose proteins are likely to be secreted. This is consistent with observations that clusters of paralogous genes, and secreted proteins, evolve relatively rapidly.

Tab. 3 1 : 1 orthologous genes in human, mouse, and rat genomes[b].

	Human/Mouse	*Human/Rat*	*Mouse/Rat*
1 : 1 ortholog relationships	11 084	10 066	11 503
Median K_S values[a]	0.56 (0.39–0.80)	0.57 (0.40–0.82)	0.19 (0.13–0.26)
Median K_A/K_S values[a]	0.10 (0.03–0.24)	0.09 (0.03–0.21)	0.11 (0.03–0.28)
Median % amino acid identity[a]	88.0% (74.4–96.3%)	88.3% (75.9–96.4%)	95.0% (88.0–98.7%)
Median % nucleotide identity[a]	85.1% (77.4–90.0%)	85.1% (77.8–89.9%)	93.4% (89.2––95.7%)

[a] Numbers in parentheses represent the 16th and 83rd percentiles.
[b] Data obtained from Ensembl, *Homo sapiens* version 11.31 (24 841 genes), *Mus musculus* version 10.3 (22 345 genes), *Rattus norvegicus* version 11.2 (21 022 genes).

The paucity of rodent-specific genes indicates that *de novo* invention of complete genes in rodents is rare. This is not unexpected, since the majority of eukaryotic protein-coding genes are modular structures containing coding and noncoding exons, splicing signals, and regulatory sequences; and the chances of independent evolution and successful assembly of these elements into a functional gene are small, given the relatively short evolutionary time available since the mouse–rat split. However, individual rodent-specific exons may arise more frequently, particularly if the exon is alternatively spliced. Of the 2302 potential novel rodent-specific exons, with transcript support, none matched human transcripts but approximately half (1116) appear to be present in alternative splice forms found in rodents. These exons are speculated to contain the few successful lineage-specific survivors of the constant process of gene evolution, by birth and death of individual exons.

4.3 Indels and Repeats in Protein-coding Sequences

In contrast to small indels occurring in the bulk of the genome, indels within protein-coding regions are likely lethal, or deleterious, and rapidly removed from the population by purifying selection. Indel rates within rat coding sequences were 50-fold lower than in bulk genomic DNA. The whole genome excess of deletions compared to insertions (Fig. 5b) was also evident in coding sequences. Deletions are ~16% more likely than insertions to be removed from coding sequences by selection.

Owing to the triplet nature of the genetic code, indels of multiples of 3 nucleotides in length (3n indels) are less likely to be deleterious. Direct comparison of 3n indel rates between bulk DNA (0.77 indel/kb for mouse, 0.83 indel/kb for rat) and coding sequence (0.087 indel/kb for mouse and 0.084 indel/kb for rat) showed that 3n indels were ninefold underrepresented in coding sequences. At least 44% of indels were duplicative insertion or deletion of a tandemly duplicated sequence, collectively termed *sequence slippage*. Sequence slippage contributed approximately equally to observed insertions and deletions, the overall excess of deletions could be attributed specifically to an excess of nonslippage deletion over nonslippage insertion in both mouse and rat lineages. Of slippage indels, 13% were trinucleotide repeats known to be particularly prone to sequence slippage and encode homopolymeric amino acid tracts.

Other characteristics of amino acid repeat variations were searched to gain better understanding of dynamic changes in length of homopolymeric amino acid tracts on gene evolution and disease susceptibility. Most species-specific amino acid repeats (80–90%) were found in indel regions and regions encoding species-specific repeats were more likely to contain tandem trinucleotide repeats than those encoding conserved repeats. This was consistent with involvement of slippage in the generation of novel repeats in proteins and extended previous observations for glutamine repeats in a more limited human–mouse dataset.

The percentage of proteins containing amino acid repeats was 13.7% in rat, 14.9% in mouse, and 17.6% in human. The most frequently occurring tandem amino acid repeats were glutamic acid, proline, alanine, leucine, serine, glycine, glutamine, and lysine. Tandem trinucleotide repeats were significantly more abundant in human than in rodent coding sequences, in

striking contrast to the frequencies observed in bulk genomic sequences (29 trinucleotide repeats/Mb in rat, 32 repeats/Mb in mouse and 13 repeats/Mb in human, Sect. 5.5). The conservation of human repeats was higher in mouse (52%) than in rat (46.5%), suggesting a higher rate of repeat loss in the rat lineage than in the mouse lineage.

Functional consequences of these in-frame changes in rat, mouse, and human were investigated through clustering of proteins based on annotation of function and cellular localization, and mapping indels onto protein structural and sequence features. The rate that indels accumulated in secreted (3.9×10^{-4} indel/aa) and nuclear (4.0×10^{-4}) proteins is approximately twice that of cytoplasmic (2.4×10^{-4}) and mitochondrial (1.4×10^{-4}) proteins. Likewise, ligand-binding proteins acquire indels (3.1×10^{-4}) at a higher rate than enzymes (2.1×10^{-4}). These trends exactly mirror those observed for amino acid substitution rates, suggesting tight coupling of selective constraints between indels and substitutions. Transcription regulators showed the highest rate of indels (4.3×10^{-4}), a finding that may relate to the overrepresentation of homopolymorphic amino acid tracts in these proteins.

Known protein domains exhibited 3.3-fold fewer indels than expected by chance, again paralleling nucleotide substitution rate differences between domains and nondomain sequences. Transmembrane regions were refractory to accumulating indels, exhibiting a sixfold reduction compared to that expected by chance. Low-complexity regions were 3.1-fold enriched, reflecting their relatively unstructured nature and enrichment for indel prone trinucleotide repeats. Mapping of indels onto groups of known structures revealed indels are 21% more likely to be tolerated in loop regions than the structural core of the protein.

An interesting observation was indel frequency and amino acid repeat occurrence both correlated positively with G + C coding sequence content of the local sequence environment. This may in part be explained by the correlation of polymerase slippage-prone trinucleotide repeat sequences and G + C content. There is also a positive correlation between CpG dinucleotide frequency and coding sequence insertions, but not deletions. This effect diminishes rapidly with increasing distance from the site of the insertion.

4.4 Transcription-associated Substitution and Asymmetry

A significant strand asymmetry for neutral substitutions in transcribed regions has been reported. Within an intron the higher rate of A → G substitutions over that of T → C substitutions, together with a smaller excess of G → A over C → T substitutions, leads to an excess of G + T over C + A on the coding strand. The asymmetries are hypothesized to be a byproduct of transcription-coupled repair in germline cells. Examining the 3-way alignments of rat, mouse, and human verified that the strand asymmetries for neutral substitutions exist in introns across the rat genome (Table 4). These asymmetries are also seen if the study is limited to ancestral repeat sites, excludes ancestral repeat sites, excludes CpG dinucleotides, is limited to positions flanked by sites that are identical in the aligned sequences (in the case of observations 2 and 3 in Table 4), or considers introns of RefSeq genes for human or mouse. Thus, it appears that strand asymmetry of substitution events

Tab. 4 Strand asymmetry of substitutions in introns of rat genes.

1	Base frequencies on coding strand	Rat genome	
	$(G + T)/(C + A)$	1.060	
2	Ratio of purine transitions to pyrimidine transitions	Rat–Mouse	Rat–Human
	$\text{rate}(A \leftrightarrow G)/\text{rate}(C \leftrightarrow T)$	1.036	1.036
3	Rate of transitions	Rat	Mouse
	$\text{rate}(A \rightarrow G)/\text{rate}(T \rightarrow C)$	1.058	1.091
	$\text{rate}(G \rightarrow A)/\text{rate}(C \rightarrow T)$	1.017	1.00*

Data in (1) were computed from the rat genome, those in (2) were computed from pairwise alignments, and data in (3) were computed from 3-way alignments. All values except * were highly significant (p-values $<10^{-4}$).

within transcribed regions of the genome is a robust genome-wide phenomenon.

4.5 Conservation of Intronic Splice Signals

The dynamics of evolution of consensus splice signals in mammalian genes was examined using 6352 human–mouse–rat orthologous introns from 976 genes. Intron class is extremely well conserved: no conversion between U2 and U12 introns, nor switching within U12 introns between the major AT–AC and GT–AG subtypes was observed, although U12 switching has been documented at larger evolutionary distances. In contrast, conversions between canonical GT–AG and noncanonical GC–AG subtypes of U2 introns are not uncommon. Only ~70% of GC–AG introns are conserved between human and mouse/rat, and only 90% are conserved between mouse and rat. Using human as outgroup, we detected 9 GT to GC conversions after divergence of mouse and rat (from 6282 introns likely to be GT–AG prior to human and rodents split), and 2 GC to GT conversions (from 34 GC–AG introns likely to predate human and rodent split). Given the higher rate of conversion from GT to GC than the reverse, these results give some indication of the degree to which mutation from T to C is tolerated in donor sites. This substitution appears to be better tolerated in introns with very strong donor sites, since in these introns the proportion of GC donor sites is ~11%, which is much higher than the 0.7% overall frequency of GC donor sites in U2 introns. Very few other noncanonical configurations in U2 introns are conserved, suggesting that most correspond to transient, evolutionarily unstable states, pseudogenes, or misannotations.

4.6 Gene Duplications

Duplication of genomic segments represents a frequent and robust mechanism for generating new genes. Since there were no compelling data showing rat-specific genes arising directly from noncoding sequences, gene duplications were examined to measure their potential contribution to rat-specific biology. A previous study showed such gene clusters in mouse without counterparts in human are subject to rapid, adaptive evolution. Using methods that directly identified paralogous clusters found 784 rat paralog clusters containing 3089 genes. This was lower than in mouse (910 clusters per 3784 genes), but the difference probably reflects the

larger number of gene predictions from the mouse assembly. Using methods that analyzed genomic segmental duplications, it appears that the timing of expansion of these individual families, is reflected in local gene duplication and retention within clusters. Neutral substitution rate varies among orthologs by approximately twofold (Fig. 10). Rates of change among ancestral gene duplications (those that predate the mouse/rat split) were relatively constant. Mouse-specific and rat-specific duplications occurred at similar rates, except for those with $K_S < 0.04$ that are reduced in mouse-specific duplications (Fig. 10), though this may be accounted for by different assembly methods.

The rat paralog pairs that probably arose after the rat/mouse split (12–24 Myr ago) have a K_S value of ≥ 0.2 (Table 3). Six hundred and forty nine $K_S < 0.2$ gene duplication events were found in rat, a lower number than found in mouse (755). For both rodents, this represents a likelihood of a gene duplicating between 1.3×10^{-3} and 2.6×10^{-3} every Myr. This is consistent with a previous estimate for Drosophila genes, and an order of magnitude lower than an estimate for *Caenorhabditis elegans* genes.

Immunoglobulin, T-cell receptor α-chain, and α_{2u} globulin genes appear to be duplicating at the fastest rates in the rat genome (Table 5). Since divergence with mouse, these rat clusters have increased gene content severalfold. This recapitulates previous observations that rapidly evolving and duplicating genes are overrepresented in olfaction and odorant-detection, antigen

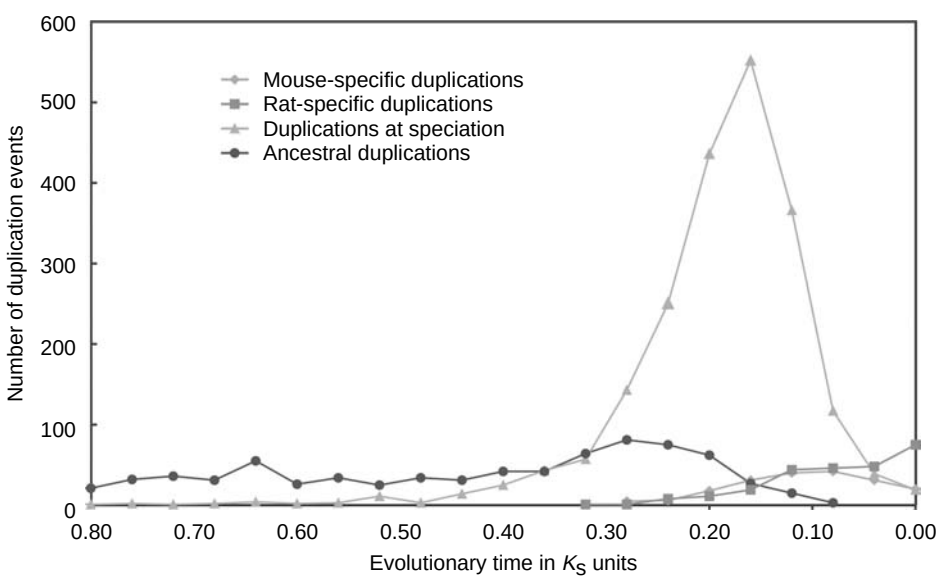

Fig. 10 Variation in the frequency of gene duplications during the evolutionary histories of the rat and mouse. The sequence of gene duplication events was inferred from phylogenetic trees determined from pairwise estimates of genetic divergence under neutral selection (K_S). The median K_S value for mouse:rat 1 : 1 orthologs is 0.19; this value corresponds to the divergence time of the mouse and rat lineages. Used with permission from *Nature* **428**, 493–521 (2004).

Tab. 5 Recent gene duplications ($K_S < 0.2$) in the rat lineage. Duplications involving retroviral genes, fragmented genes with internal repeats, and likely pseudogene clusters were removed from this list. Only gene clusters exhibiting at least 3 duplications are shown.

Cluster ID/ Chrom.	*Recent duplication events*	*Numbers of genes involved*	*Extant/ Ancestral cluster size*	*Annotation*	*Process*
249/4	38	53	60/22	Immunoglobulinκ-chain V	Immunity
640/15	38	47	53/15	TCR α-chain V	Immunity
346/6	25	35	44/15	Immunoglobulin heavy chain V	Immunity
190/3	22	42	168/146	Olfactory receptor	Chemosensation
578/13	16	28	59/43	Olfactory receptor	Chemosensation
400/8	15	26	82/67	Olfactory receptor	Chemosensation
743/20	15	21	37/22	Olfactory receptor	Chemosensation
72/1	12	22	102/90	Olfactory receptor	Chemosensation
500/10	12	18	32/20	Olfactory receptor	Chemosensation
51/1	6	7	16/10	Glandular kallikrein	Reproduction?
256/4	6	8	10/4	Vomeronasal receptor V1R	Chemosensation
488/10	6	10	11/5	Olfactory receptor	Chemosensation
644/15	6	10	14/8	Granzyme serine protease	Immunity
4/1	5	6	9/4	Trace amine receptor, GPCR	Neuropeptide receptors?
248/4	5	9	15/10	Vomeronasal receptor V1R	Chemosensation
393/8	5	10	31/26	Olfactory receptor	Chemosensation
522/10	5	8	19/14	Keratin-associated protein	Epithelial cell function
550/11	5	8	17/12	Olfactory receptor	Chemosensation
635/15	5	9	20/15	Olfactory receptor	Chemosensation
79/1	4	8	38/34	Olfactory receptor	Chemosensation
88/1	4	6	11/7	Olfactory receptor	Chemosensation
109/1	4	7	43/39	Olfactory receptor	Chemosensation
294/5	4	5	5/1	α_{2u} globulin	Chemosensation
310/5	4	5	11/7	Olfactory receptor	Chemosensation
353/7	4	7	13/9	Olfactory receptor	Chemosensation
399/8	4	5	6/2	Ly6-like urinary protein	Chemosensation?
638/15	4	6	6/2	Ribonuclease A	Immunity
690/17	4	6	21/17	Prolactin paralog	Reproduction
239/4	3	6	6/3	Prolactin-induced protein	Reproduction
253/4	3	4	5/2	Camello-like N-acetyltransferase	Developmental regulator
274/4	3	6	20/17	Ly-49 lectin natural killer cell protein	Immunity
297/5	3	4	5/2	Interferon-α	Immunity
523/10	3	4	6/3	Keratin-associated protein	Epithelial cell function
746/20	3	5	6/3	MHC class 1b (M10)	Chemosensation

recognition, and reproduction. An examination of duplicated genomic segments showed this enrichment for most of the same genes and also elements involved in foreign compound detoxification (cytochrome P450 and carboxylesterase genes). Together these are exciting findings because each of these categories can easily be associated with a familiar feature of rat-specific biology, and further investigation could explain some differences between rats and their evolutionary neighbors.

4.7 Conservation of Gene-regulatory Regions

As the third mammal to be fully sequenced, the rat can add significantly to the utility of nucleotide alignments for identifying conserved noncoding sequences. This power increases roughly as a function of total amount of neutral substitution represented in the alignment, and rat adds about 15% to the human-mouse comparison (Fig. 5). Many conserved mammalian noncoding sequences are expected to have regulatory function, and can be predicted using further analyses based upon these alignments.

Typical genome-wide human–mouse–rat alignments show strong conservation for a coding exon, as well as for several noncoding regions (Fig. 11). For example, the intronic region, in Fig. 11 contains 504 bp that are highly conserved in human, mouse, and rat. The last 100 bp of this alignment block are identical in all three species. Peaks in regulatory potential score are correlated with conservation score, and in the highly conserved intronic segment, they are higher for the three-way regulatory potential score than for the two-way scores using human and just one rodent. These data are illustrative, but form the foundation of ongoing efforts to identify genome sequences involved in gene regulation.

Requiring conservation among mammalian genomes greatly increases specificity of predictions of transcription factor binding sites. Transcription factor databases such as TRANSFAC contain known transcription factor binding sites and some knowledge of their distribution, but simply searching a sequence with these motifs provides little discriminatory power. Using a set of 164 weight matrices for 109 transcription factors extracted from TRANSFAC finds 186 792 933 matches in the April 2003 reference human genome sequence, but this was reduced to only 4 188 229 by demanding conservation in the human–mouse–rat three-way alignments. This is a 44-fold increase in specificity.

4.8 Pseudogenes and Gene Loss

Pseudogenes found in rat were classified according to whether they arose from retrotransposition, in which case they integrated into the genome randomly, or whether they arose from tandem duplication and neutral sequence substitution. Using human–rat synteny, 80% of pseudogenes exhibited no significant similarity to the corresponding human orthologous region, and therefore were considered to be retrotransposed, processed pseudogenes. The total pseudogene count and processed pseudogene proportion are consistent with those found for human. Pseudogenes are normally not subjected to selective constraint and therefore accumulate sequence modifications neutrally. Indeed, nearly all of our identified pseudogenes ($97 \pm 3\%$) evolved under neutrality according to a

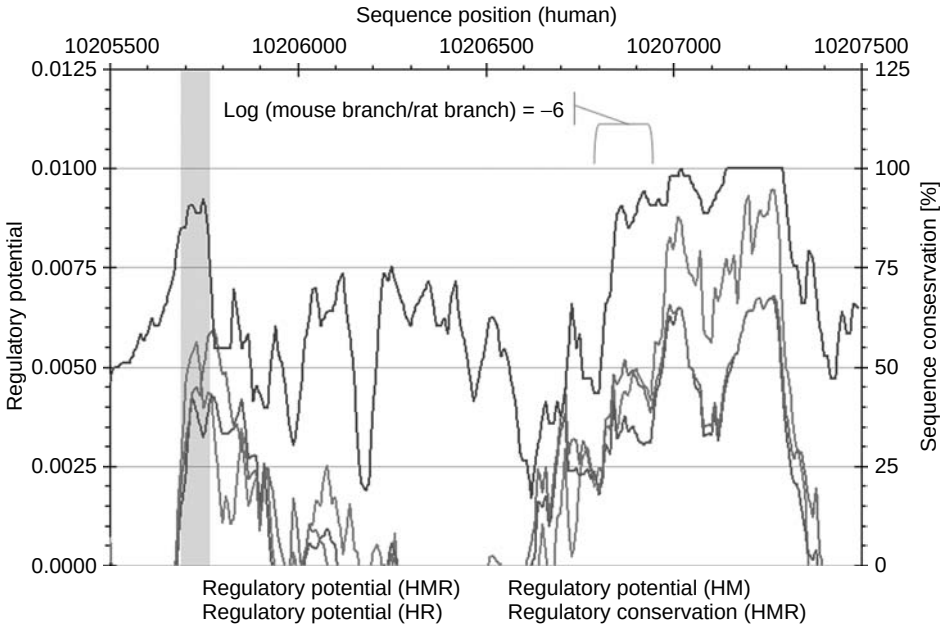

Fig. 11 Close-up of PEX14 (peroxisomal membrane protein) locus on human chromosome 1 (with homologous mouse chromosome 4 and rat chromosome 5). Conservation score computed on 3-way human–mouse–rat alignments presents a clear coding exon peak (gray bar) and very high values in a 504 bp noncoding, intronic segment (right; last 100 bp of alignment are identical in all three organisms). The latter segment showed a striking difference between the inferred mouse and rat branch lengths: the gray bracket corresponds to a phylogenetic tree where the logarithm of mouse to rat branch length ratio is −6. Regulatory potential (RP) scores that discriminate between conserved regulatory elements and neutrally evolving DNA are calculated from 3-way (human–mouse–rat) and 2-way (human-rodent) alignments. Here the 3-way regulatory potential scores are enhanced over the 2-way scores. Used with permission from *Nature* **428**, 493–521 (2004). (See color plate p. li.)

K_A/K_S test, and therefore, are consistent with being pseudogenic.

As with the human genome, the largest group of rat pseudogenes (totaling 2188) consists of ribosomal protein genes. Other large rat pseudogene families arose from olfactory receptors (552, see Sect. 6.1), glyceraldehyde 3-phosphate dehydrogenase (251), protein kinases (177), and RNA binding RNP-1 proteins (174). Pseudogenes homologous to a meiotic spindle-associated protein spindlin are particularly numerous in rat (at least 53 copies), compared to mouse (approximately 3 copies). This suggests that spindlin pseudogenes may have distributed rapidly by a recently active transposable element.

4.9 *In Situ* Loss of Rat Genes

As an organism evolves, its need for certain genes may be reduced, or lost, owing to changes in its ecological niche. Loss of selective constraints leads to accumulation of nonsense and/or frameshift mutations without retrotransposition or duplication. These nonprocessed pseudogenes are

Tab. 6 Candidate rat pseudogenes, orthologous to mouse and human functional genes.

Mouse gene	*Human gene*	*Strand*	*Rat genome Coordinates*[a]	*Frameshifts stops*[b]	*Annotation*
ENSMUSG00000013611	ENSG00000174226	+	7:92752590–92807556	1/0	Sorting nexin
ENSMUSG00000024364	ENSG00000158402	+	18:62742414–62770427	2/0	Dual specificity phosphatase CDC25c
ENSMUSG00000026293	ENSG00000077044	+	9:95634847–95692601	1/0	Diacylglycerol kinase δ
ENSMUSG00000026785	ENSG00000160447	+	3:9210762–9229984	5/0	Protein kinase PKNβ
ENSMUSG00000026829	ENSG00000148288	+	3:7662414–7664521	2/2	Forssman glycolipid synthetase
ENSMUSG00000027426	ENSG00000125846	+	3:125918806–125924149	1/1	Zinc finger protein 133
ENSMUSG00000028000	ENSG00000138799	–	2:221272797–221304350	1/0	Complement factor I
ENSMUSG00000029203	ENSG00000078140	–	14:44385206–44441888	1/0	Ubiquitin-protein ligase E2 (HIP2)
ENSMUSG00000030270	ENSG00000144550	–	20:8332585–8362331	3/0	Copine (membrane trafficking)
ENSMUSG00000035449	ENSG00000167646	+	1:67374986–67381472	1/0	Cardiac troponin I
ENSMUSG00000037029	ENSG00000105261	–	1:82728049–82730272	1/0	Zinc finger protein 146
ENSMUSG00000037432	ENSG00000158142	+	9:42465695–42498651	1/1	Dysferlin-like protein
ENSMUSG00000039660	ENSG00000167137	–	3:9320401–9326997	4/0	Similar to yeast YMR310c RNA binding protein
ENSMUSG00000042653	ENSG00000137634	+	8:49938446–49939091	1/0	Brush border 61.9 kDa-like protein

[a]Coordinates from rat v2.0.
[b]Mouse genes were used as templates for predicting rat pseudogenes.

interesting since they link environmental changes to genomic mutation events. However, predicted pseudogenes with disrupted reading frames might also be indicative of errors in genome sequence or assembly. By constraining the search to orthologous genomic regions, we identified 14 rat putative nonprocessed pseudogenes (Table 6) with apparently functional, single human, and mouse orthologs.

4.10 Noncoding RNA Genes

The abundance and distribution of noncoding RNAs was investigated in rat. Cytoplasmic tRNA gene identification in rodents is complicated by tRNA derived ID SINE elements (B2 and ID). TRNAscan–SE predicted 175 943 tRNAs (genes and pseudogenes); however, the majority (175 285) were SINEs identified by RepeatMasker. This is far greater than found in mouse (24 402/25 078) or human (25/636). Of the remaining 666 predictions, 163 were annotated as tRNA pseudogenes and 4 were annotated as undetermined by tRNAscan–SE. An additional 68 predictions were removed because their best database match in either human, mouse, or rat tRNA databases matched tRNAs with either a different amino acid or anticodon (violating the wobble rules that specify the distinct anticodons expected). The total of 431 tRNAs (including a single selenocysteine tRNA) identified in the rat genome is comparable to that for mouse (435) and human (492). These three species share a core set of approximately 300 tRNAs using a cutoff of ≥95% sequence identity and ≥95% sequence length.

A total of 454 noncodingRNAs (other than tRNAs) were identified by sequence comparison to known noncoding RNAs. These include 113 micro (mi) RNAs, 5 ribosomal RNAs, 287 small nucleolar (sno) RNAs, and small nuclear (sn) RNAs; 49 various other ncRNAs such as signal recognition particle (SRP) RNA, 7SK RNA, telomerase RNA, Rnase P RNA, brain-specific repetitive (Bsr) RNA, noncoding transcript abundantly expressed in brain (Ntab) RNA, small cytoplasmic (sc) RNA, and 626 pseudogenes. Complete 18S and 28S rRNA genes and more rRNAs were not identified presumably owing to assembly issues.

5 Evolution of Transposable Elements

Most interspersed repeats are immobilized copies of transposable elements that have accrued substitutions in proportion to their time spent fixed in the genome (for introduction see references in Nature paper). About 40% of the rat genome draft is identified as interspersed repetitive DNA derived from transposable elements, similar to mouse (Table 7) and lower than for the human (almost 50%). The latter difference is mainly because of the lower substitution rate in the human lineage, which allows us to recognize much older (Mesozoic) sequences as interspersed repeats. Almost all repeats are derived from retroposons, elements that procreate via reverse transcription of their transcripts. As in mouse, there is no evidence for activity of DNA transposons since the rat–mouse split. Many aspects of the rat and the mouse genomes' repeat structure are shared; here we focus on the differences.

5.1 LINE-1 Activity in the Rat Lineage

The long interspersed nucleotide element (LINE-1 or L1) is an autonomous

Tab. 7 Composition of interspersed repeats in the rat genome.

	Rat				*Mouse*	
	Copies [×10³]	*Total length [Mb]*	*Fraction of genome [%]*	*Lineage specific [%]*	*Fraction of genome [%]*	*Lineage specific [%]*
LINEs:	657	594.0	23.11	11.70	20.10	9.74
LINE-1	597	584.2	22.73	11.70	19.65	9.74
LINE-2	48	8.4	0.33	–	0.38	–
L3/CR1	11	1.4	0.06	–	0.06	–
SINEs:	1360	181.3	7.05	1.52	7.78	1.80
B1(Alu)	384	42.3	1.65	0.16	2.53	0.92
B4(ID_B1)	359	55.4	2.15	0.00	2.25	0.00
ID	225	19.6	0.76	0.54	0.20	0.00
B2	328	55.2	2.15	0.68	2.29	0.74
MIR	109	13.0	0.51	–	0.56	–
LTR elements:	556	232.4	9.04	1.84	10.28	2.85
ERV_classI	40	24.9	0.97	0.56	0.79	0.36
ERV_classII	141	83.4	3.24	1.02	4.13	1.73
ERV-L (III)	74	21.6	0.84	0.04	1.08	0.23
MaLRs	302	102.5	3.99	0.22	4.27	0.53
DNA elements:	108	20.9	0.81	–	0.86	–
Charlie(hAT)	80	14.8	0.58	–	0.60	–
Tigger(Tc1)	18	4.0	0.16	–	0.17	–
Unclassified:	14	7.3	0.28	–	0.37	–
Total	2690	1036	40.31	14.90	39.45	14.26
Small RNAs:	8	0.6	0.03	0.01	0.03	0.01
Satellites:	14	6.4	0.25	?	0.31	?
Simple repeats:	897	61.1	2.38	?	2.41	?

Data for Rnor3.1 and October 2003 mouse (MM4), excluding Y chromosome, using the December 17 2003 version of RepeatMasker. To highlight the differences between rat and mouse repeat content, column 5 and 6 show the fraction of the genomes comprised of lineage specific repeats. The LINE-1 numbers include all HAL1 copies, while all BC1 scRNA and >10% diverged tRNA–Ala matches, far more common than other small RNA pseudogenes and closely related to ID, have been counted as ID matches.

retroelement, with an internal RNA polymerase II promoter and two open reading frames (ORFs). ORF1 is an RNA binding protein with chaperone-like activity suggesting a role in mediating nucleic acid strand transfer steps during L1 reverse transcription. ORF2 encodes a protein with both reverse transcriptase and DNA endonuclease activity. LINEs are usually 5′ truncated so that only a small subset extends to include the promoter region and can function as a source for more copies.

Many classes of LINE-like elements exist, but only L1 has been active in rodents. Over half a million copies, in variable stages of decay, comprise 22% of the rat genome. Although over 10% of the human genome is L1 copies introduced before the rodent–primate split, only 2% of the rat genome could be recognized as such because of the fast substitution rate in the rodent lineage. Thus, probably well over one quarter of all rat DNA is derived directly from the L1 gene.

Following the mouse/rat split, L1 activity appears to have increased in rat. The 6 rat-specific L1 subfamilies, represented by 150 000 copies, cover 12% of the rat genome. Mouse-specific L1 copies accumulated over the same period cover only 10% of the genome (Table 7). This greater accumulation of L1 copies could explain some of the size difference of the rat and mouse genome.

In addition to the traditional L1 elements, there are 7500 copies (10 Mb) of a nonautonomous element that is derived from L1 by deletion of most of its ORF2. On the basis of their low divergence, the presently identified HAL1-like (HAL1 for Half-a-LINE) elements operated only a few million years ago in the mouse lineage (MusHAL1) and still propagate in the rat genome (RNHAL1). RNHAL1 only contains an ORF1, while MusHAL1 encoded an endonuclease as well, though no reverse transcriptase. The repeated origin and high copy number of HAL1s suggests that the ORF1 product, which binds strongly to its mRNA, may render this transcript a superior target for L1 mediated reverse transcription. In this way, HAL1 resembles the nonautonomous, endogenous retrovirus-derived MaLR elements (in the following), which, for over 100 million years, retained only the retroviral gag ORF that encodes an RNA binding protein. A potential advantage of HAL1 over L1 is its shorter length, which, considering the usual 5′ truncation of copies, increases the chance that a copy includes the internal promoter elements and can become a source gene.

5.2 Different Activity of SINEs in the Rat and Mouse Lineage

The most successful usurpers of the L1 retrotransposition machinery are SINEs. These are small RNA-derived sequences with an internal RNA polymerase III promoter. The human Alu SINE was shown to be transposed by L1. L1 lacks sequence specificity and rodent and primate SINE sequences are unrelated to L1. Though any transcript can be retroposed, as evidenced by the numerous mammalian processed pseudogenes, L1-dependent SINEs probably have features that make them especially efficient targets of the L1 reverse transcriptase.

While before the radiation of most mammalian orders L1 was at least as active as L2, the L2 dependent MIR was the only known (and very abundant) SINE of that time. All of the currently active SINEs in different mammalian orders appear to have arisen after the demise of L2 (and consequently MIR), as if an opportunity (or necessity) arose for the creation and expansion of other SINEs.

Four different SINEs are distinguished in rat and mouse. The B1 element and the primate Alu originated from a 7SLRNA gene, probably just before the rodent–primate split and after the speciation from most other eutherians, where Alu/B1 elements are not known. The other SINEs (B4, B2, ID) are rodent-specific and have tRNA-like internal promoter regions.

The fortunes of these SINEs during mouse and rat evolution have been different (Fig. 12). B4 probably became extinct before the mouse–rat speciation, while B2 has remained productive in both lineages, scattering >100 000 copies in each genome after this time. Interestingly, the fate of the B1 and ID SINEs has been opposite in rat and mouse. While B1 is still active in mouse, having left over 200 000 mouse-specific copies in its trail, the youngest of the 40 000 rat-specific B1 copies are 6 to 7% diverged from their source, indicating a relatively early extinction in the rat lineage.

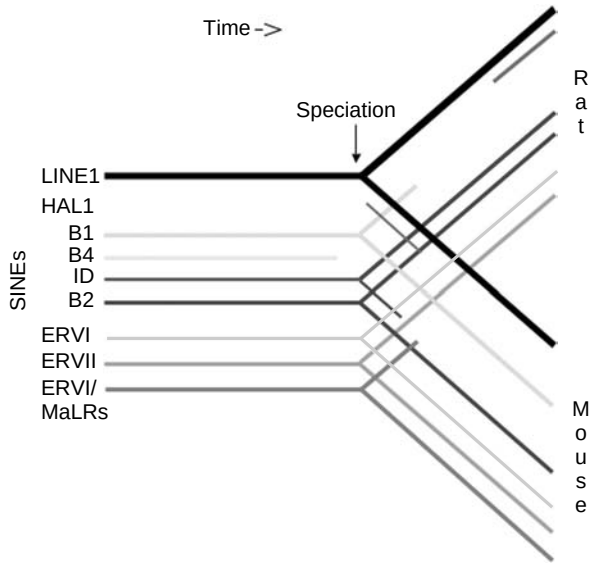

Fig. 12 Historical view of rodent repeated sequences. Relationships of the major families of interspersed repeats (Table 7) are shown for the rat and mouse genomes, indicating losses and gains of repeat families after speciation. The lines indicate activity as a function of time. Note that HAL1-like elements appear to have arisen both in the mouse and rat lineage. Used with permission from *Nature* **428**, 493–521 (2004). (See color plate p. lii.)

On the other hand, after the mouse–rat split only a few hundred ID copies may have inserted in mouse, while this heretofore minor SINE (~60 000 copies predate the speciation) picked up activity in rat to produce 160 000 ID copies.

5.3 Colocalization of SINEs in Rat and Mouse

Although the fates of SINE families differ, the number of SINEs inserted after speciation in each lineage is remarkably similar, ~300 000 copies. As MIR was replaced by L1 driven SINEs, it appears that the demise of B1 in rat allowed the expansion of IDs. Moreover, these independently inserted and unrelated SINEs (ID and B1 only share a mechanism of retroposition) accumulated at orthologous sites: the density of rat-specific SINEs in 14 243 ~100 kb windows in the rat genome is highly correlated ($R^2 = 0.83$) with the density of mouse-specific SINEs in orthologous regions in mouse (Fig. 5). These data corroborate and refine the observation of a strong correlation between the location of primate and rodent-specific SINEs in 1-Mb windows. At 100 kb, no correlation is seen for interspersed repeats other than SINEs.

Insertions of SINEs at the same location have been reported, and the correlation could reflect the existence of conserved hotspots for SINE insertions. However, primate data does not support this. Likewise, gene conversions do not significantly contribute to the observed correlation.

Figure 9(c) displays the lineage-specific SINE densities on rat chromosome 10 and in the mouse orthologous blocks, showing a stronger correlation than any other feature. The cause of the unusual distribution patterns of SINEs, accumulating in gene-rich regions where other interspersed repeats are scarce, apparently is a conserved feature independent of the primary sequence of the SINE and effective over smaller than isochore-sized regions.

In the human genome, the most recent (unfixed) Alus are distributed similar to L1, whereas older copies gradually take on the opposite distribution of SINEs. However, this temporal shift in SINE distribution pattern is not observed in mouse or rat.

Some regions of high LINE content coincide with regions that exhibit both higher AT content and an elevated rate of point substitution (Fig. 9, pink rectangles). In a genome-wide analysis, LINE content correlates strongly with substitution rates, and about 80% of this correlation is explained by higher rates in AT-rich regions. SINE density shows the opposite correlation both on chromosome 10 (Fig. 9) and genome-wide.

These phenomena, in conjunction with an overall trend in substitution rates toward AT-richness, suggest a model by which quickly evolving regions accumulate a higher-than-average AT content, which attracts LINE elements. While distinct cause–effect relationships such as this remain largely speculative, these results reinforce the idea that local genomic context strongly shapes local genomic features and rates of evolution.

5.4 Endogenous Retroviruses and Derivatives

Retrovirus-like elements are the other major contributors to interspersed repeats in the rodent genome. These have several 100-bp long terminal repeats (LTRs) with transcriptional regulatory sequences that flank an internal sequence that, in autonomous elements, encodes all proteins necessary for retrotransposition. All mammalian LTR elements are endogenous retroviruses (ERVs; classes I, II, or III) or their nonautonomous derivatives.

The most productive retrovirus in mammals has been the class III element ERV-L, primarily through its ancient nonautonomous derivatives called "MaLRs" with 350 000 copies occupying ~5% of the rat genome (Table 7). Human ERV-L and MaLR copies are >6% diverged from their reconstructed source genes and must have died out around the time of our speciation from new world monkeys. In mouse, several thousand almost identical MaLR and ERV-L copies suggest sustained activity. In contrast, rat ERV-L activity was silenced a few million years ago given that the least diverged MaLR and ERV-L copies differ by >4% from each other. Other class III ERVs were active earlier in rodent evolution, before the mouse–rat speciation. Class I and class II elements still thrive in rat, where 2 of 4 class I and 4 of 9 class II autonomous ERVs appear to be active.

5.5 Simple Repeats

Interspersed simple sequence repeats (SSRs), regions of tandemly repeated (1–6 bp) units that probably arise from slippage during DNA replication and can expand and compress by unequal crossing over. Remarkable differences were noted between the human and mouse genomes' SSR content with threefold to fivefold more base pairs contained in near (>90%) perfect SSRs in mouse. SSRs are

both more frequent and on average longer in mouse. Polypurine (or polypyrimidine) repeats are especially (10-fold) overrepresented in the mouse genome. As discussed in Sect. 4.3, this contrasts sharply with the greater frequency of triplet repeats coding for amino acids in human relative to the rodents.

Rat and mouse SSR content are more similar, with genome coverage of ~1.4% (for >90% perfect elements compared to 0.45% in human) and are of similar average length. For example, the most common mammalian SSR, the (CA)n repeat, averages 42 bp long in mouse and 44 bp in rat. Some potentially significant differences are that polypurine SSRs are of similar average length but 1.2-fold more common in mouse, while the rare SSRs containing CG dimers are 1.5-fold more frequently observed in rat.

5.6 Prevalent, Medium-length Duplications in Rodents

In addition to the transpositionally derived and simple repeats, the rat and mouse genomes contain a substantial amount of medium-length unclassified duplications (typically 100–5000 bp). These are readily seen in self-comparisons and in intrarodent comparisons after masking the known repeats, but they are much less prevalent in comparisons with the human genome. A substantial fraction of the rodent genomes consists of currently unexplained repeats, which may include (1) novel families of low-copy rodent interspersed repeats; (2) extensions of known but not fully characterized rodent repeats; and (3) duplications generated by a mechanism different from transposition.

6 Rat-specific Biology

The rat genome sequence and predicted gene set reveals genetic differences between rats and mice that might specify their differences in physiology and behavior. In particular, recently duplicated genes are enriched in elements involved in chemosensation and functional aspects of reproduction (Table 5). The differences in the gene complements of rat and mouse are illustrated by in-depth analyses of olfactory receptors (ORs), pheromones, cytochromes P450, proteases, and protease inhibitors.

6.1 Chemosensation

The ability to emit and sense specific smells is a key feature of survival for most animals in the wild. Rat and mouse pheromones, vomeronasal receptors, and ORs genes were duplicated frequently during the time since the common ancestor of rats and mice (Table 5). The rapid evolution of these genes is attributed to conspecific competition, in particular sexual selection.

The rat genome has 1866 ORs in 113 locations: 69 multigene clusters, and 49 single genes. Extrapolating for the 9.8% of the genome not in the assembly, there are ~2070 OR genes and pseudogenes. The rat therefore has ~37% more OR genes and pseudogenes than the ~1510 ORs of the mouse, assuming similar representation of recently duplicated sequences in the two genome assemblies used. Of the 1774 OR sequences that are not interrupted by assembly gaps, 1216 (69%) encode intact proteins, while the remaining 547 (31%) sequences are likely pseudogenes with in-frame stop codons, frameshifts, and/or interspersed repeat

elements. Fewer mouse OR homologs are pseudogenes (~20%), but the larger family size in rat still leaves it with substantially more intact ORs than the mouse (~1430 vs ~1210). Striking rat-specific expansions of two ancestral clusters account for much of the difference in OR family size and pseudogene content between rat and mouse, although many other clusters exhibit more subtle changes. Significant differences between human and mouse in OR families have also been reported, but the functional implications of OR repertoire size on the ability of different species to detect and discriminate odorants are not yet known.

6.2 α_{2u} Globulin Pheromones

The α_{2u} globulin genes are odorant-binding proteins that also contribute to essential survival functions in animals. α_{2u} globulin homologs are likely to be highly heterogeneous among murid species. Distinct homologs and genomic arrangement differences have been observed among mouse strains. The evolution of α_{2u} globulin genes on rat chromosome 5 has remodeled this genomic region (Fig. 13). The orthologous human genomic region contains a single homolog, suggesting that the common ancestor of rodents and human possessed one gene. The genome of C57BL/6J mice contains 4 homologous genes, and 7 pseudogenes, while the rat genome contains 10α_{2u} globulin genes and 12 pseudogenes in a single region.

Phylogenetic trees constructed using amino acid, and noncoding DNA, sequences show that, surprisingly, the rat α_{2u} globulin gene clusters appear to have arisen recently via a rapid burst of gene duplication since the rat/mouse split (Table 5). This is consistent with the Rfp37-like zinc finger-like pseudogene having accompanied virtually all of the rat-specific α_{2u} globulin gene duplications (Fig. 13). The sequences of these genes are

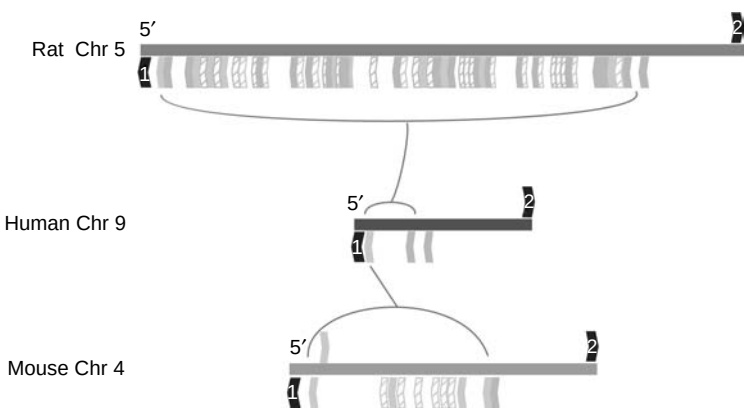

Fig. 13 Adaptive remodeling of genomes and genes. Orthologous regions of rat, human and mouse genomesencoding pheromone-carrier proteins of the lipocalin family (α_{2u} globulins in rat and major urinary proteins in mouse) shown in brown. Zfp37-like zinc finger genes are shown in blue. Filled arrows represent likely genes, whereas striped arrows represent likely pseudogenes. Gene expansions are bracketed. Arrow head orientation represents transcriptional direction. Flanking genes 1 and 2 are TSCOT and CTR1 respectively. (See color plate p. xliv.)

also evolving rapidly with median K_A/K_S values of 0.77 and 1.06, for rat and mouse genes respectively. Amino acid sites that appear to have been subject to adaptive evolution are situated both within the ligand-binding cavity, and on the solvent-exposed periphery of the α_{2u} globulin structure. This demonstrates how genome analysis can reveal the imprint of adaptive evolution from megabase to single base levels.

The rapid evolution of these genes, and the remodeling of their genomic regions, can be attributed to the known roles of rat α_{2u} globulins and mouse homologs in conspecific competition and sexual selection. These proteins are pheromones and pheromone carriers that are present in large quantities in rodent urine, and act as scent markers indicating dominance and subspecies identity.

6.3 Detoxification

Cytochrome P450 is a well-recognized participant in metabolic detoxification and rapid evolution is observed within this family. These enzymes are particularly relevant to clinical and pharmacological studies in humans as they metabolize many toxic and endogenous compounds. Rodents are important model organisms for understanding human drug metabolism, so it is important to identify 1 : 1 orthologs and species-specific expansions and losses. Compared to human genes, there are clear expansions of several rodent P450 subfamilies, but there are also significant differences between rat and mouse subfamilies (Fig. 14a). The fastest evolving subfamily seems to be CYP2J containing a single gene in human, but at least 4 in rat and 8 in mouse (Fig. 14b,c). CYP2J enzymes catalyze the NADPH-dependent oxidation of arachidonic acid to various eicosanoids, which in turn possess numerous biological activities including modulation of ion transport, control of bronchial and vascular smooth muscle tone, and stimulation of peptide hormone secretion. The genomic ordering of genes and their phylogenetic tree indicate an ongoing expansion in the rodents (Fig. 14b,c). This suggests that adaptive evolution has been involved in diversifying their functions.

6.4 Proteolysis

Protease genes also represent an example of rapid evolution in the rat genome. The rat contains 626 protease genes (~1.7% of the genes), similar to mouse (641) and more than human (561). One hundred and two rat protease genes are not found in human, and 42 are absent from mouse. Several rat gene families have expanded; most are involved in reproductive or immunological functions, and have evolved independently in the rat and mouse lineages.

These gene family expansions dramatically illustrate how large-scale genomic changes have accompanied species-specific innovation. Positive selection of duplicated genes has afforded the rat an enhanced repertoire of precisely those genes that allow reproductive success despite severe competition from both within its own, and with other species. This serves as a general illustration of the importance of chemosensation, detoxification, and proteolysis in innovation and adaptation.

7 Human Disease Gene Orthologs in the Rat Genome

The rat is recognized as the premier model for studying the physiological aspects of

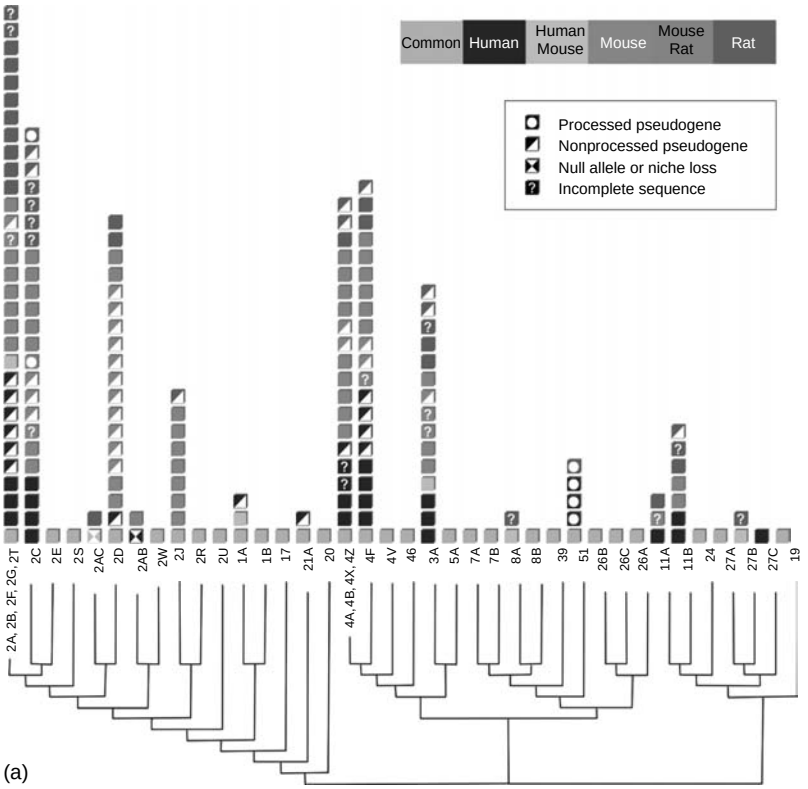

Fig. 14 Evolution of cytochrome P450 (CYP) protein families in rat, mouse, and human. (a) Dendrogram of topology from 234 full-length sequences. The 279 sequences of 300 amino acids; subfamily names and chromosome numbers are shown. Black branches have >70% bootstrap support. Incomplete sequences (they contain Ns) are included in counts of functional genes (84 rat, 87 mouse, and 57 human) and pseudogenes (including fragments not shown; 77 rat, 121 mouse, and 52 human). Thus, 64 rat genes and 12 pseudogenes were in predicted gene sets. Human CYP4F is a null allele due to an in-frame STOP codon in the genome, although a full-length translation exists (SwissProt P98187). Rat CYP27B, missing in the genome, is ''incomplete'' since there is a RefSeq entry (NP_446215). Grouped subfamilies CYP2A, 2B, 2F, 2G, 2T, and CYP4A, 4B, 4X, 4Z, occur in gene clusters; thus nine loci contain multiple functional genes in a species. One (CYP1A) has fewer rat genes than human, seven have more rodent than human, and all nine have different copy numbers. CYP2AC is a rat-specific subfamily (orthologs are pseudogenes). CYP27C has no rodent counterpart. Rodent-specific expansion, rat CYP2J is illustrated below. (b) The neighbor-joining tree, with the single human gene, contains clear mouse (Mm) and rat (Rn) orthologous pairs (bootstrap values >700/1000 trials shown). Bar indicates 0.1 substitutions per site. (c) All rat genes have a single mouse counterpart except for CYP2J 3, which has further expanded in mouse (mouse CYP2J 3a, 3b, and 3c) by two consecutive single duplications. The genes flanking the CYP2J orthologous regions (rat chromosome 5, 126.9–127.3 Mb; mouse chromosome 4, 94.0–94.6 Mb; human chromosome 1, 54.7–54.8 Mb) are hook1 (HOOK1; pink) and nuclear factor I/A (NFIA; cyan). Genes (solid) and gene fragments (dashed boxes) are shown above (forward strand) and below (reverse strand) the horizontal line. No orthology relation could be concluded for most of these cases. Used with permission from *Nature* **428**, 493–521 (2004). (See color plate p. liii.)

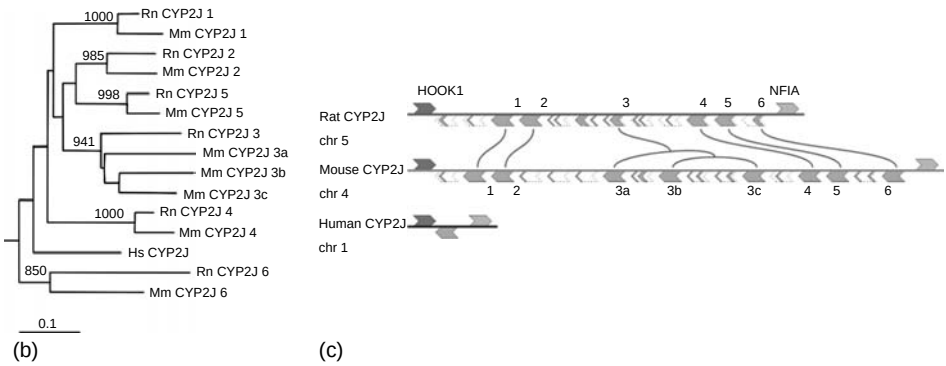

Fig. 14 (*Continued*)

many human diseases, but it has not had as prominent a role in the study of simple genetic disease traits. The rat genome provides an opportunity to use the more than 1000 human Mendelian disorders that have associated loci and alleles to link the rat to human disease examples. The precise identification of the rat orthologs of human genes that are mutated in disease creates further opportunities to discover and develop rat models for biomedical research.

Predicted rat genes were compared with 1112 well-characterized human disease genes that were verified and classified on the basis of pathophysiology. So, 844 (76%) have 1 : 1 orthologs in the rat as predicted by Ensembl. These predictions are likely to be of high quality since 97.4% of the 11 422 rat:human 1 : 1 orthologs predicted by Ensembl were found in orthologous genomic regions.

The proportion of human disease genes with single orthologs in the rat (76%) was higher than the proportion for all human genes (46%). Careful analysis of the remaining 268 human genes not predicted to show 1 : 1 orthology indicated that only six of the human disease genes lack likely rat orthologs among genome, cDNA, EST, and protein sequences. Thus, it appears that, in general, genes involved in human disease are unlikely to have diverged, or to have become duplicated, deleted, or lost as pseudogenes, between rat and human divergence.

Comparisons of K_S, K_A, and the K_A/K_S ratio values of human disease orthologs with those of all remaining ortholog pairs found only the K_S distributions differed significantly indicating that coding regions of human disease genes and their rat counterparts have mutated more rapidly than the nondisease genes. Factors influencing the specific loci could cause this result, or the disease genes may characteristically reside in genomic regions that exhibit higher mutation rates.

The disease gene set was next grouped into 16 disease-system categories and analyzed (Fig. 15). Only five disease systems exhibited significant K_A/K_S differences with respect to the remaining samples. Neurological and malformation-syndrome disease categories manifested K_A/K_S ratios consistent with purifying selection acting on these gene sets. In contrast, the pulmonary, hematological, and immune categories manifested the highest median K_A/K_S ratios, consistent with a role for more positive selection, or reduced selective constraints, among these genes.

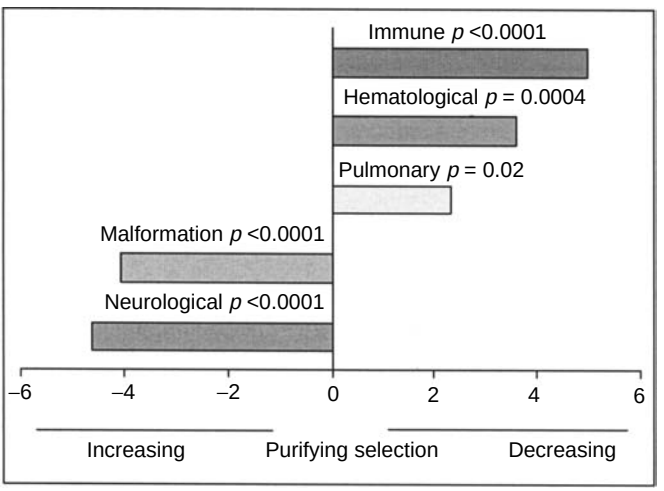

Fig. 15 Selective constraints differ for human disease systems in the rat genome. Human disease-system categories showing significant differences ($p < 0.05$) in a nonparametric test [Mann–Whitney–Wilcoxon] comparing K_A/K_S (human:rat) ratios. *P*-values from two-level tests between genes from one disease system and the remaining genes. (Mean–Mean0)/Std0 values from multilevel tests from sixteen categorized disease systems. Negative values for neurological (−4.63) and malformation-syndrome (−4.04) categories were observed to be consistent with K_A/K_S ranges where purifying selection predominates. Immune, hematological, and pulmonary categories show positive values of 4.98, 3.59, and 2.34, respectively. Used with permission from *Nature* **428**, 493–521 (2004).

Orthologs of more diverse phyla demonstrated consistent results.

These results demonstrate that various disease systems exhibit significantly different average evolutionary rates. The higher rates noted for the immune system disease genes are consistent with rapid diversification of the functions of the immune systems of rodents and humans. This is expected for genes involved in controlling species-restricted infectious agents if strong adaptive pressure acts during host–pathogen coevolution. Thus, results of studies of these rodent genes may be less directly relevant to our understanding of human immune system diseases than results obtained for other pathophysiology disease systems where conservation is greater and purifying selection is stronger.

A number of genes were examined that harbor triplet nucleotide repeats, and are involved in human neurological disorders, such as Huntington disease, a condition caused by CAG triplet repeat expansion producing abnormally long polyglutamine tracts in an otherwise normal protein. Analysis of the rat:human orthologs of these disease genes indicated that repeat length is substantially shorter in all cases in the rat than in the normal human gene (Fig. 16). To date, there are no naturally occurring rat strains described that exhibit neurological disease associated with repeat expansion mechanisms. The rat may lack repeat

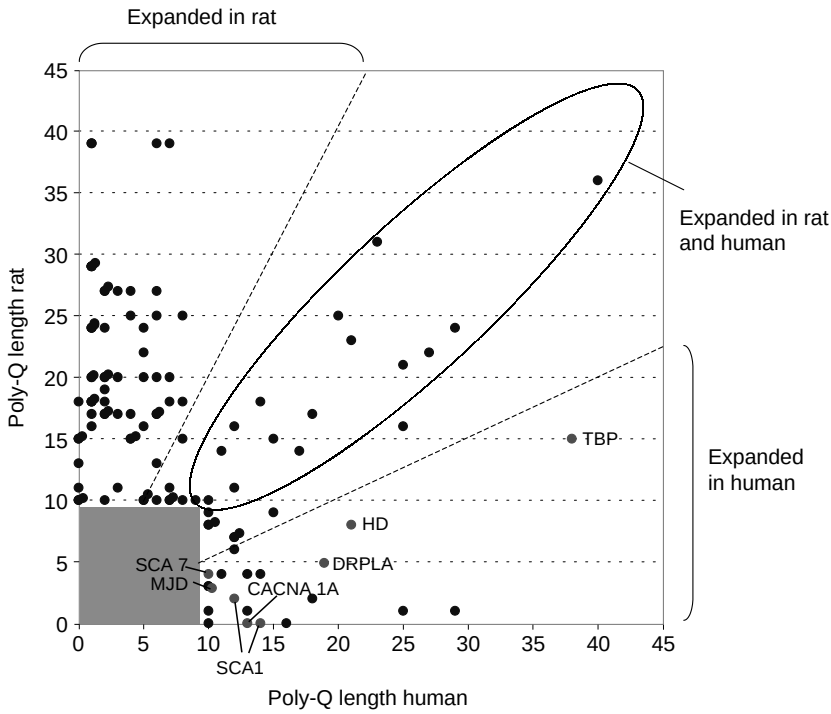

Fig. 16 Polyglutamine repeat length comparison between human and rat. Points represent protein poly-Q length for rat and human. Red points correspond to repeats in genes associated with human disease: SCA1, spinocerebellar ataxia 1 protein, or ataxin1; SCA7, spinocerebellar ataxia 7 protein; MJD, Machado–Joseph disease protein; CACNA1A, spinocerebellar ataxia 6 protein, or calcium channel α-1A subunit isoform 1; DRPLA, dentatorubro–pallidoluysian atrophy protein; HD, Huntington's disease protein, or huntingtin; TBP, TATA binding protein or spinocerebellar ataxia 17 protein. Repeat lengths over 10 were examined; green shading delineates the range not included in the analysis. Also noted are a set that are expanded in rat and human (black circle) and a set where repeats are expanded in the rat. Used with permission from *Nature* **428**, 493–521 (2004). (See color plate p. liv.)

expansion mutational mechanisms or these orthologs may fail to reach a "critical repeat length" susceptible to such mutational mechanisms. Other human genes without known disease association also contain glutamine repeats that are much shorter in the rat orthologs, and thus may be susceptible to mutations that arise through repeat expansion mechanisms. In Fig. 16, it may also be observed that a relatively high proportion of repeats are significantly longer in the rat than in their corresponding human ortholog.

In addition to enabling the direct comparison of rat/human disease orthologs, the rat genome sequence itself is an invaluable aid for discovery of other rat genes to study as disease models. First, genes underlying disease phenotypes with simple inheritance that have been mapped to chromosomal regions can be more easily pursued in both species. Indeed, the

rearrangements of conserved segments between the two species have significant value since they tighten the boundaries of the mapped disease regions and reduce the number of genes potentially associated with the disease phenotype. Second, the identification of multiple alleles contributing to quantitative and complex trait differences that are involved in disease processes can be pursued with more accuracy, both in the initial association phases, and in subsequent efforts to detect causative alleles.

8 Summary

As the third mammalian genome to be sequenced, the rat genome has provided both predictable and surprising information about mammalian species. Although clear at the outset that ongoing rat research would benefit from the resource of a genome sequence, there was uncertainty about how many new insights would be found, especially considering the superficial similarities between the rat and the already sequenced mouse. Instead, the results of the sequencing and analysis have generated some deep insights into the evolutionary processes that have given rise to these different species. In addition, the project was invaluable for further developing the methods for the generation and analysis of large genome sequence data sets.

The generation of the rat draft tested the new "combined approach" for large genome sequencing. The high quality of the overall attests of this overall strategy and the supporting software, provide a suitable approach to this problem. With a BAC "skimming" component in the underlying data set, the assembly recovered a fraction of the genome that was expected, by analogy to the mouse project to be difficult to assemble from pure WGS data. The BAC skimming component also allowed progressive generation of high quality local assemblies that were of use to the rat community as the project developed.

The rat genome data have improved the utility of the rat model enormously. The rat gene content provides a parts list that can be explored with the high degree of confidence and precision that is appropriate for biomedical research. The resources for physical and genetic mapping have also been improved, since the relative position of individual markers is now known with confidence and there are computational resources to bridge the process of genetic association with gene modeling and experimental investigation.

The expected benefit of a third mammalian sequence providing an outgroup by which to discriminate the timing of events that had already been noted between mouse and human was fully realized. Using the three sequences and other partial data sets from additional organisms, it was possible to measure some of the overall faster rate of evolutionary change in the rodent lineage shared by mice and rats, as well as the peculiar acceleration of some aspects of rat-specific evolution. The observation of specific expanded gene families in the rat should provide material for targeted studies for some time.

The sequence of the rat genome represents the beginning of the full analysis of the mammalian genome and its complex evolutionary history. Much of the additional data required to complete this story will be from other genomes, distantly related to rat. Nevertheless, the prospect of a finished rat genome, polymorphism data

from analysis of other rat strains, and targeted cDNA clone collections will provide rat-specific reagents for routine use in research. Together with the ongoing efforts to fully develop methods to genetically manipulate whole rats and provide effective "gene knockouts" the current and future rat genome resources will ensure a place for this organism in genomic and biomedical research for some time.

Acknowledgment

This chapter is based largely on the publication of the rat genome sequence by the Rat Genome Sequencing Project Consortium (*Nature* **428**, 493–521 (2004)). We thank E. Eichler, K. Fetchel, R. Hardison, P. Havlak, K. Kalafus, M. Krzywinski, A. Milosavljevic, L. Pachter, P. Pevzner, C. Ponting, K. Roskin, A. Sidow, A. Smit, G. Tessler, and D. Torrents, for the figures. The rat genome-sequencing project was funded primarily by the NHGRI and NHLBI of the NIH.

See also Genetics, Molecular Basis of; Genomic Sequencing (Core Article).

Bibliography

Books and Reviews

Chiaromonte, F. et al. (2003) The share of human genomic DNA under selection estimated from human–mouse genomic alignments, *Cold Spring Harb. Symp. Quant. Biol.* **68**, 245–254.

Danielson, P.B. (2002) The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans, *Curr. Drug Metab.* **3**, 561–597.

Hedrich, H.J. (2000) in: Krinke, G.J. (Ed.) *History, Strains, and Models in the Laboratory Rat*, Academic Press, San Diego, CA, pp. 3–16.

Lopez-Otin, C., Overall, C.M. (2002) Protease degradomics: a new challenge for proteomics, *Nat. Rev. Mol. Cell Biol.* **3**, 509–519.

Lynch, M., Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes, *Science* **290**, 1151–1155.

Montoya-Burgos, J.I., Boursot, P., Galtier, N. (2003) Recombination explains isochores in mammalian genomes, *Trends Genet.* **19**, 128–130.

Ohno, S. (1970) *Evolution by Gene Duplication*, Springer, Berlin, Germany.

Prak, E.T., Kazazian, H.H. Jr. (2000) Mobile elements and the human genome, *Nat. Rev. Genet.* **1**, 134–144.

Reddy, P.S., Housman, D.E. (1997) The complex pathology of trinucleotide repeats, *Curr. Opin. Cell Biol.* **9**, 364–372.

Scarborough, P.E., Ma, J., Qu, W., Zeldin, D.C. (1999) P450 subfamily CYP2J and their role in the bioactivation of arachidonic acid in extrahepatic tissues, *Drug Metab. Rev.* **31**, 205–234.

Smit, A.F. (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes, *Curr. Opin. Genet. Dev.* **9**, 657–663.

Weiner, A.M. (2002) SINEs and LINEs: the art of biting the hand that feeds you, *Curr. Opin. Cell Biol.* **14**, 343–350.

Willson, T.M., Kliewer, S.A. (2002) PXR, CAR and drug metabolism, *Nat. Rev. Drug Discov.* **1**, 259–266.

Primary Literature

Adams, M.D. et al. (2000) The genome sequence of *Drosophila melanogaster*, *Science* **287**, 2185–2195.

Adkins, R.M., Gelke, E.L., Rowe, D., Honeycutt, R.L. (2001) Molecular phylogeny and divergence time estimates for major rodent groups: evidence from multiple genes, *Mol. Biol. Evol.* **18**, 777–791.

Alba, M.M., Guigo, R. (2004) Comparative analysis of amino acid repeats in rodents and humans, *Genome Res.* **14**, 549–554.

Alba, M.M., Santibanez-Koref, M.F., Hancock, J.M. (1999) Conservation of polyglutamine tract size between mice and humans depends

on codon interruption, *Mol. Biol. Evol.* **16**, 1641–1644.

Altschul, S.F., Lipman, D.J. (1990) Protein database searches for multiple alignments, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5509–5513.

Antequera, F., Bird, A. (1993) Number of CpG islands and genes in human and mouse, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11995–11999.

Aparicio, S., et al. (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*, *Science* **297**, 1301–1310.

Bailey, J.A., et al. (2002) Recent segmental duplications in the human genome, *Science* **297**, 1003–1007.

Batzoglou, S., et al. (2002) ARACHNE: a whole-genome shotgun assembler, *Genome Res.* **12**, 177–189.

Benit, L., et al. (1997) Cloning of a new murine endogenous retrovirus, MuERV-L, with strong similarity to the human HERV-L element and with a gag coding sequence closely related to the Fv1 restriction gene, *J. Virol.* **71**, 5652–5657.

Bird, A.P. (1980) DNA methylation and the frequency of CpG in animal DNA, *Nucleic Acids Res.* **8**, 1499–1504.

Birdsell, J.A. (2002) Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution, *Mol. Biol. Evol.* **19**, 1181–1197.

Boffelli, D., et al. (2003) Phylogenetic shadowing of primate sequences to find functional regions of the human genome, *Science* **299**, 1391–1394.

Bourque, G., Pevzner, P.A. (2002) Genome-scale evolution: reconstructing gene orders in the ancestral species, *Genome Res.* **12**, 26–36.

Bourque, G., Pevzner, P.A., Tesler, G. (2004) Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes, *Genome Res.* **14**, 507–516.

Bray, N., Pachter, L. (2004) MAVID Constrained ancestral alignment of multiple sequence, *Genome Res.* **14**, 693–699.

Burge, C.B., Padgett, R.A., Sharp, P.A. (1998) Evolutionary fates and origins of U12-type introns, *Mol. Cell* **2**, 773–785.

Cai, L., et al. (1997) Construction and characterization of a 10-genome equivalent yeast artificial chromosome library for the laboratory rat, *Rattus norvegicus*, *Genomics* **39**, 385–392.

Cantrell, M.A., et al. (2001) An ancient retrovirus-like element contains hot spots for SINE insertion, *Genetics* **158**, 769–777.

Cavaggioni, A., Mucignat-Caretta, C. (2000) Major urinary proteins, 2U-globulins and aphrodisin, *Biochim. Biophys. Acta* **1482**, 218–228.

Chakrabarti, K., Pachter, L. (2004) Visualization of multiple genome annotations and alignments with the K-BROWSER, *Genome Res.* **14**, 716–720.

Chen, R., Sodergren, E., Gibbs, R., Weinstock, G.M. (2004) Dynamic building of a BAC clone tiling path for genome sequencing project, *Genome Res.* **14**, 679–684.

Cheung, J., et al. (2003) Recent segmental and gene duplications in the mouse genome, *Genome Biol.* **4**, R47 [online].

Clark, A.J., Hickman, J., Bishop, J. (1984) A 45-kb DNA domain with two divergently orientated genes is the unit of organisation of the murine major urinary protein genes, *EMBO J.* **3**, 2055–2064.

Cooper, G.M., et al. (2004) Characterization of evolutionary rates and constraints in three mammalian genomes, *Genome Res.* **14**, 539–548.

Cooper, G.M., Brudno, M., Green, E.D., Batzoglou, S., Sidow, A. (2003) Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes, *Genome Res.* **13**, 813–820.

Costas, J. (2003) Molecular characterization of the recent intragenomic spread of the murine endogenous retrovirus MuERV-L, *J. Mol. Evol.* **56**, 181–186.

Dehal, P., et al. (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins, *Science* **298**, 2157–2167.

Dermitzakis, E.T., et al. (2002) Numerous potentially functional but non-genic conserved sequences on human chromosome 21, *Nature* **420**, 578–582.

Dermitzakis, E.T., et al. (2003) Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs), *Science* **302**, 1033–1035.

Dewannieux, M., Esnault, C., Heidmann, T. (2003) LINE-mediated retrotransposition of marked Alu sequences, *Nat. Genet.* **35**, 41–48.

Duret, L., Mouchiroud, D. (2000) Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity

but not mutation rate, *Mol. Biol. Evol.* **17**, 68–74.

Eichler, E.E. (1998) Masquerading repeats: paralogous pitfalls of the human genome, *Genome Res.* **8**, 758–762.

Eichler, E.E. (2001) Segmental duplications: what's missing, misassigned, and misassembled – and should we care? *Genome Res.* **11**, 653–656.

Elnitski, L., et al. (2003) Distinguishing regulatory DNA from neutral sites, *Genome Res.* **13**, 64–72.

Emes, R.D., Goodstadt, L., Winter, E.E., Ponting, C.P. (2003) Comparison of the genomes of human and mouse lays the foundation of genome zoology, *Hum. Mol. Genet.* **12**, 701–709.

Emes, R.D., Beatson, S.A., Ponting, C.P., Goodstadt, L. (2004) Evolution and comparative genomics of odorant- and pheromone-associated genes in rodents, *Genome Res.* **14**, 591–602.

Felsenfeld, A., Peterson, J., Schloss, J., Guyer, M. (1999) Assessing the quality of the DNA sequence from the human genome project, *Genome Res.* **9**, 1–4.

Goff, S.A., et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*), *Science* **296**, 92–100.

Graves, J.A., Gecz, J., Hameister, H. (2002) Evolution of the human X – a smart and sexy chromosome that controls speciation and development, *Cytogenet. Genome Res.* **99**, 141–145.

Green, P. (2002) Whole-genome disassembly, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4143–4144.

Green, H., Wang, N. (1994) Codon reiteration and the evolution of proteins, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4298–4302.

Gumucio, D.L., et al. (1992) Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human and globin genes, *Mol. Cell. Biol.* **12**, 4919–4929.

Gurates, B., et al. (2002) WT1 and DAX-1 inhibit aromatase P450 expression in human endometrial and endometriotic stromal cells, *J. Clin. Endocrinol. Metab.* **87**, 4369–4377.

Guy, J., et al. (2003) Genomic sequence and transcriptional profile of the boundary between pericentromeric satellites and genes on human chromosome arm 10p, *Genome Res.* **13**, 159–172.

Haldi, M.L., et al. (1997) Construction of a large-insert yeast artificial chromosome library of the rat genome, *Mamm. Genome* **8**, 284.

Hardison, R., et al. (1993) Comparative analysis of the locus control region of the rabbit-like gene cluster: HS3 increases transient expression of an embryonic-globin gene, *Nucleic Acids Res.* **21**, 1265–1272.

Hardison, R.C., et al. (2003) Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution, *Genome Res.* **13**, 13–26.

Havlak, P., et al. (2004) The *Atlas* genome assembly system, *Genome Res.* **14**, 721–732.

Hayward, B.E., Zavanelli, M., Furano, A.V. (1997) Recombination creates novel L1 (LINE-1) elements in *Rattus norvegicus*, *Genetics* **146**, 641–654.

Hillier, L.W., et al. (2003) The DNA sequence of human chromosome 7, *Nature* **424**, 157–164.

Horvath, J.E., et al. (2003) Using a pericentromeric interspersed repeat to recapitulate the phylogeny and expansion of human centromeric segmental duplications, *Mol. Biol. Evol.* **20**, 1463–1479.

Huang, H., et al. Evolutionary conservation of human disease gene orthologs in the rat and mouse genomes. *Genome Biol.* (submitted).

Hubbard, T., et al. (2002) The Ensembl genome database project, *Nucleic Acids Res.* **30**, 38–41.

Hurst, L.D. (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution, *Trends Genet.* **18**, 486–487.

Hurst, J.L., et al. (2001) Individual recognition in mice mediated by major urinary proteins, *Nature* **414**, 631–634.

International Human Genome Sequencing Consortium. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

Jaffe, D.B., et al. (2003) Whole-genome sequence assembly for mammalian genomes: arachne 2, *Genome Res.* **13**, 91–96.

Jensen-Seaman, M.I., et al. (2004) Comparative recombination rates in the rat, mouse, and human genomes, *Genome Res.* **14**, 528–538.

Kalafus, K.J., Jackson, A.R., Milosavljevic, A. (2004) Pash: efficient genome-scale sequence anchoring by positional hashing, *Genome Res.* **14**, 672–678.

Kent, W.J., Baertsch, R., Hinrichs, A., Miller, W., Haussler, D. (2003) Evolution's cauldron: duplication, deletion, and rearrangement in

the mouse and human genomes, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11484–11489.

Kirkness, E.F., et al. (2003) The dog genome: survey sequencing and comparative analysis, *Science* **301**, 1898–1903.

Kolbe, D., et al. (2004) Regulatory potential scores from genome-wide 3-way alignments of human, mouse and rat, *Genome Res.* **14**, 700–707.

Krzywinski, M., et al. (2004) Integrated and sequence-ordered BAC and YAC-based physical maps for the rat genome, *Genome Res.* **14**, 766–779.

Kwitek, A.E., et al. (2004) High density rat radiation hybrid maps containing over 24,000 SSLPs, genes, and ESTs provide a direct link to the rat genome sequence, *Genome Res.* **14**, 750–757.

Levinson, G., Gutman, G.A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution, *Mol. Biol. Evol.* **4**, 203–221.

Li, X., Waterman, M.S. (2003) Estimating the repeat structure and length of DNA sequences using L-tuples, *Genome Res.* **13**, 1916–1922.

Loots, G.G., Ovcharenko, I., Pachter, L., Dubchak, I., Rubin, E.M. (2002) rVista for comparative sequence-based discovery of functional transcription factor binding sites, *Genome Res.* **12**, 832–839.

Ma, B., Tromp, J., Li, M. (2002) PatternHunter: faster and more sensitive homology search, *Bioinformatics* **18**, 440–445.

Makalowski, W., Boguski, M.S. (1998) Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9407–9412.

Margulies, E.H., Blanchette, M., Haussler, D., Green, E. (2003) Identification and characterization of multi-species conserved sequences, *Genome Res.* **13**, 2507–2518.

Marra, M.A., et al. (1997) High throughput fingerprint analysis of large-insert clones, *Genome Res.* **7**, 1072–1084.

Martin, S.L., Bushman, F.D. (2001) Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon, *Mol. Cell. Biol.* **21**, 467–475.

Misra, S., et al. (2002) Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biol.* **3**(12), RESEARCH0083. Epub 2002 Dec. 31.

Modrek, B., Lee, C.J. (2003) Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss, *Nat. Genet.* **34**, 177–180.

Mouse Genome Sequencing Consortium. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562.

Mural, R.J., et al. (2002) A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome, *Science* **296**, 1661–1671.

Murphy, W.J., Fronicke, L., O'Brien, S.J., Stanyon, R. (2003) The origin of human chromosome 1 and its homologs in placental mammals, *Genome Res.* **13**, 1880–1888.

Murphy, W.J., Sun, S., Chen, Z.Q., Pecon-Slattery, J., O'Brien, S.J. (1999) Extensive conservation of sex chromosome organization between cat and human revealed by parallel radiation hybrid mapping, *Genome Res.* **9**, 1223–1230.

Murphy, W.J., Bourque, G., Tesler, G., Pevzner, P., O'Brien, S.J. (2003) Reconstructing the genomic architecture of mammalian ancestors using multispecies comparative maps, *Hum. Genomics* **1**, 30–40.

Myers, E.W., et al. (2000) A whole-genome assembly of *Drosophila*, *Science* **287**, 2196–2204.

Myers, E.W., Sutton, G.G., Smith, H.O., Adams, M.D., Venter, J.C. (2002) On the sequencing and assembly of the human genome, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4145–4146.

Nadeau, J.H., Taylor, B.A. (1984) Lengths of chromosomal segments conserved since divergence of man and mouse, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 814–818.

Nekrutenko, A. (2004) Identification of novel exons from rat-mouse comparisons, *J. Mol. Evol.* **59**(5), 703–708.

Nekrutenko, A., Makova, K.D., Li, W.H. (2002) The KA/KS ratio test for assessing the protein-coding potential of genomic regions: an empirical and simulation study, *Genome Res.* **12**, 198–202.

Nekrutenko, A., Chung, W.Y., Li, W.H. (2003) An evolutionary approach reveals a high protein-coding capacity of the human genome, *Trends Genet.* **19**, 306–310.

Nelson, D.R. (1999) Cytochrome P450 and the individuality of species, *Arch. Biochem. Biophys.* **369**, 1–10.

Osoegawa, K., et al. (2004) BAC Resources for the rat genome project, *Genome Res.* **14**, 780–785.

Ostertag, E.M., Kazazian, H.H. Jr. (2001) Biology of mammalian L1 retrotransposons, *Annu. Rev. Genet.* **35**, 501–538.

Pennacchio, L.A., Rubin, E.M. (2001) Genomic strategies to identify mammalian regulatory sequences, *Nat. Rev. Genet.* **2**, 100–109.

Pevzner, P., Tesler, G. (2003a) Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7672–7677.

Pevzner, P., Tesler, G. (2003b) Genome rearrangements in mammalian evolution: lessons from human and mouse genomes, *Genome Res.* **13**, 37–45.

Pruitt, K.D., Maglott, D.R. (2001) RefSeq and LocusLink: NCBI gene-centered resources, *Nucleic Acids Res.* **29**, 137–140.

Puente, X.S., Lopez-Otin, C.A. (2004) A genomic analysis of rat proteases and protease inhibitors, *Genome Res.* **14**, 609–622.

Puente, X.S., Sanchez, L.M., Overall, C.M., Lopez-Otin, C. (2003) Human and mouse proteases: a comparative genomic approach, *Nat. Rev. Genet.* **4**, 544–558.

Quentin, Y. (1994) A master sequence related to a free left Alu monomer (FLAM) at the origin of the B1 family in rodent genomes, *Nucleic Acids Res.* **22**, 2222–2227.

Rat Genome Sequencing Project Consortium. (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution, *Nature* **428**, 493–521.

Riethman, H., et al. (2004) Mapping and initial analysis of human subtelomeric sequence assemblies, *Genome Res.* **14**, 18–28.

Roskin, K.M., Diekhans, M., Haussler, D. (2003) in: Vingron, M., Istrail, S., Pevzner, P., Waterman, M. (Eds.) *Proceedings of the 7th Annual International Conference on Research in Computational Biology (RECOMB 2003)*, ACM Press, New York, pp. 257–266, doi:10.1145/640075.640109.

Rothenburg, S., Eiben, M., Koch-Nolte, F., Haag, F. (2002) Independent integration of rodent identifier (ID) elements into orthologous sites of some RT6 alleles of *Rattus norvegicus* and *Rattus rattus*, *J. Mol. Evol.* **55**, 251–259.

Rouquier, S., Blancher, A., Giorgi, D. (2000) The olfactory receptor gene repertoire in primates and mouse: evidence for reduction of the functional fraction in primates, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2870–2874.

Roy-Engel, A.M., et al. (2002) Non-traditional Alu evolution and primate genomic diversity, *J. Mol. Biol.* **316**, 1033–1040.

Saitou, N., Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Mol. Biol. Evol.* **4**, 406–425.

Salem, A.H., et al. (2003) Alu elements and hominid phylogenetics, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12787–127891.

Salem, A.H., Kilroy, G.E., Watkins, W.S., Jorde, L.B., Batzer, M.A. (2003) Recently integrated Alu elements and human genomic diversity, *Mol. Biol. Evol.* **20**, 1349–1361.

Schwartz, S., et al. (2003) Human–mouse alignments with BLASTZ, *Genome Res.* **13**, 103–107.

Smit, A.F. (1993) Identification of a new, abundant superfamily of mammalian LTR-transposons, *Nucleic Acids Res.* **21**, 1863–1872.

Springer, M.S., Murphy, W.J., Eizirik, E., O'Brien, S.J. (2003) Placental mammal diversification and the Cretaceous-Tertiary boundary, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1056–1061.

Stanyon, R., Stone, G., Garcia, M., Froenicke, L. (2003) Reciprocal chromosome painting shows that squirrels, unlike murid rodents, have a highly conserved genome organization, *Genomics* **82**, 245–249.

Steen, R.G., et al. (1999) A high-density integrated genetic linkage and radiation hybrid map of the laboratory rat, *Genome Res.* **9**, (insert), AP1–AP8.

Stenson, P.D., et al. (2003) Human Gene Mutation Database (HGMD): 2003 update, *Hum. Mutat.* **21**, 577–581.

Tagle, D.A., et al. (1988) Embryonic epsilon and globin genes of a prosimian primate (*Galago crassicaudatus*). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints, *J. Mol. Biol.* **203**, 439–455.

Taylor, M.S., Ponting, C.P., Copley, R.R. (2004) Occurrence and consequences of coding sequence insertions and deletions in mammalian genomes, *Genome Res.* **14**, 555–566.

The International Human Genome Mapping Consortium. (2001) A physical map of the human genome, *Nature* **409**, 934–941.

Thomas, J.W., et al. (2003a) Pericentromeric duplications in the laboratory mouse, *Genome Res.* **13**, 55–63.

Thomas, J.W., et al. (2003b) Comparative analyses of multi-species sequences from targeted genomic regions, *Nature* **424**, 788–793.

Torrents, D., Suyama, M., Bork, P. (2003) A genome-wide survey of human pseudogenes, *Genome Res.* **13**, 2559–2567.

Trinklein, N.D., Aldred, S.J., Saldanha, A.J., Myers, R.M. (2003) Identification and functional analysis of human transcriptional promoters, *Genome Res.* **13**, 308–312.

Tuzun, E., Bailey, J.A., Eichler, E.E. (2004) Recent segmental duplications in the working draft assembly of the brown Norway rat, *Genome Res.* **14**, 493–506.

Venter, J.C., et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Ventura, M., Archidiacono, N., Rocchi, M. (2001) Centromere emergence in evolution, *Genome Res.* **11**, 595–599.

Vitt, U., et al. (2004) Identification of candidate disease genes by EST alignments, synteny and expression and verification of Ensembl genes on rat chromosome 1q43-54, *Genome Res.* **14**, 640–650.

Waterston, R.H., Lander, E.S., Sulston, J.E. (2002) On the sequencing of the human genome, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3712–3716.

Waterston, R.H., Lander, E.S., Sulston, J.E. (2003) More on the sequencing of the human genome. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3022–3024; author reply (2003) **100**, 3025–3026.

Wingender, E., et al. (2001) The TRANSFAC system on gene expression regulation, *Nucleic Acids Res.* **29**, 281–283.

Wolfe, K.H., Sharp, P.M. (1993) Mammalian gene evolution: nucleotide sequence divergence between mouse and rat, *J. Mol. Evol.* **37**, 441–456.

Yang, S., et al. (2004) Patterns of insertions and their covariation with substitutions in the rat, mouse and human genomes, *Genome Res.* **14**, 517–527.

Yang, Z., Goldman, N., Friday, A. (1994) Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation, *Mol. Biol. Evol.* **11**, 316–324.

Yap, V.B., Pachter, L. (2004) Identification of evolutionary hotspots in the rodent genomes, *Genome Res.* **14**, 574–579.

Young, J.M., et al. (2002) Different evolutionary processes shaped the mouse and human olfactory receptor gene families, *Hum. Mol. Genet.* **11**, 535–546.

Yu, J., et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*), *Science* **296**, 79–92.

Yunis, J.J., Prakash, O. (1982) The origin of man: a chromosomal pictorial legacy, *Science* **215**, 1525–1530.

Zhang, X., Firestein, S. (2002) The olfactory receptor gene superfamily of the mouse, *Nat. Neurosci.* **5**, 124–133.

Zhang, Z., Harrison, P., Gerstein, M. (2002) Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome, *Genome Res.* **12**, 1466–1482.

Real-Time Quantitative PCR: Theory and Practice

Gregory L. Shipley
The University of Texas Health Science Center, Houston, TX

1 Real-time PCR Basics 483
1.1 Principles of Real-time PCR 483
1.2 The Reverse Transcription Reaction 485
1.3 The Polymerase Chain Reaction 487

2 Fluorescent Dyes, Assay Chemistries, and Real-time Instruments 488
2.1 Fluorescence in Real-time PCR 488
2.2 Real-time Instruments 490
2.3 Selected Assay Chemistries 492

3 RNA and DNA Isolation for Real-time PCR 499
3.1 Isolation of DNA 499
3.2 Isolation of RNA 499

4 Standards Used for Real-time Quantitative PCR 500
4.1 RNA Standards 500
4.2 DNA Standards 501

5 Real-time PCR Assay Design 503
5.1 Available Software 503
5.2 Basic Rules for Real-time Assay Development 504
5.3 Key Factors in Assay Development 506

6 Running a Real-time PCR Experiment 509
6.1 Practical Considerations 509
6.2 Reagents 510
6.3 Controls 512

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

7 Data Analysis 513
7.1 Baseline Boundaries and Threshold Levels 514
7.2 Fluorescent Signal and Sample Normalization 514
7.3 Quantification Using a Standard Curve 516
7.4 Quantification Using the ddCt Method 518

8 Future Directions 520

Bibliography 521
Books and Reviews 521
Primary Literature 521

Keywords

Amplicon
The region of a target sequence that is amplified during the PCR, encompassing the sequence bounded by and including the two primer sequences.

Baseline
The range of early PCR cycles that contain low levels of increasing fluorescent signal that is undetectable over the baseline fluorescence of the reporter dye.

Fluorescence Resonance Energy Transfer
Light energy transferred from an excited dye to a neighboring dye in close proximity, resulting in the quenching of the visible emission of light from the primary dye.

Polymerase Chain Reaction
Primer-directed geometric increase in a DNA template using a thermostable DNA polymerase, necessary reagents, and a succession of heating and cooling steps performed in a thermocycling device.

Real-time PCR
This is the continuous collection of fluorescent signals from one or more polymerase chain reactions. Quantitative real-time PCR is the conversion of the fluorescent signal from each reaction to a numerical value for each sample. This is accomplished by calculation from a known reference sample (ddCt) or interpolation from a quantified standard curve. Real-time PCR is the most sensitive and accurate method for the detection and accurate quantification of nucleic acids from any biological source. The term real-time PCR in this article refers to both PCR and RT-PCR.

Reverse Transcription
The primer-directed synthesis of a DNA copy (cDNA) using RNA as the template and reverse transcriptase.

Threshold
A level of fluorescent signal that is significantly above the baseline fluorescence level. The threshold is set by the software or manually by the user within the geometric phase of the amplification curve and is used to define the cycle threshold (Ct) or crossing point (Cp).

Transcript quantification methodology has evolved from Northern blots to RNase protection to competitive RT-PCR. Each of these methods made strides toward greater sensitivity and accuracy in quantification. Since the introduction of the 7700 Sequence Detection System and Taqman® chemistry by Applied Biosystems in 1996, quantitative real-time PCR has become widely accepted as the most sensitive and accurate method for the quantification of both RNA and DNA from a wide variety of sources. Today, there are a number of real-time instruments from a variety of vendors and several detection methodologies and kit chemistries available. At the same time, prices for real-time platforms are falling dramatically. This trend means that the technology is moving from one found primarily in core laboratories to the bench tops of individual investigators.

There are several ways to interrogate a cell for changes induced by artificial or natural agents during a biological process. One way is to look for changes in cellular transcript levels that may indicate downstream changes in the corresponding protein. In another instance, the focus may be on the presence or absence of a viral or bacterial pathogen. In this case, detecting not only the presence but also the degree of infection provides valuable information. Alternatively, looking for the presence or the level of expression from a transgene or the inhibition of expression of an endogenous gene by an RNAi agent may be the question of interest. In all cases, quantitative real-time PCR technology can be utilized to provide the required results. However, successful implementation of the technology requires the user to have a basic background in the theoretical principles of real-time PCR as well as their practical application. The goal of this review is to provide this information.

It will not be possible to cover all aspects of real-time PCR in this review. For a more comprehensive overview of real-time PCR, "A-Z of Quantitative PCR," edited by Stephen Bustin, is highly recommended.

1 Real-time PCR Basics

Before any meaningful discussion of real-time PCR can be initiated, it will be useful to discuss the theoretical principles governing and basic chemistries utilized in real-time PCR.

1.1 Principles of Real-time PCR

Prior to real-time PCR, the effect of the primers and other components of the reagent chemistry on the amplification process during the polymerase chain reaction was something of a black box.

Tubes bearing assay chemistry, primers, and template were exposed to 30 or more cycles of heating and cooling on a thermocycler. The reaction product(s) were then separated on a gel. The presence of a single band of the correct size with any magnitude was deemed a successful experiment. With the advent of real-time PCR, the ability to observe the status of the reaction products during every cycle has brought new insight into the kinetics of template amplification. Most real-time PCR experiments are run for 40 cycles to capture the complete amplification profile from every tube or well. In an ideal PCR, the increase in template occurs geometrically throughout the entire 40 cycles. However, we now know that this is not the case. The amplification of any template can be defined by four phases: (1) the baseline; (2) the true geometric phase; (3) a linear phase as the efficiency of the PCR starts to fall in each successive cycle; and (4) a plateau as amplification arrests (Fig. 1). All three phases can be seen in the amplification plots from any real-time PCR reaction. An example of amplification curves from a 5-log dilution series of an oligonucleotide standard can be see in Fig. 2. Why the PCR slows down and then stops is not clear, but may be due to the build up of products that are inhibitory to the reaction. One thing that is clear is that the concentration of all reactants is in vast excess and none of them should be depleted during 40 cycles of PCR.

Samples with very high amounts of template have been observed, using real-time PCR, to stop amplification sooner and with fewer end products than those with lower initial template concentrations. The result is that the one with higher template will appear to have less band intensity on a gel compared to the one with

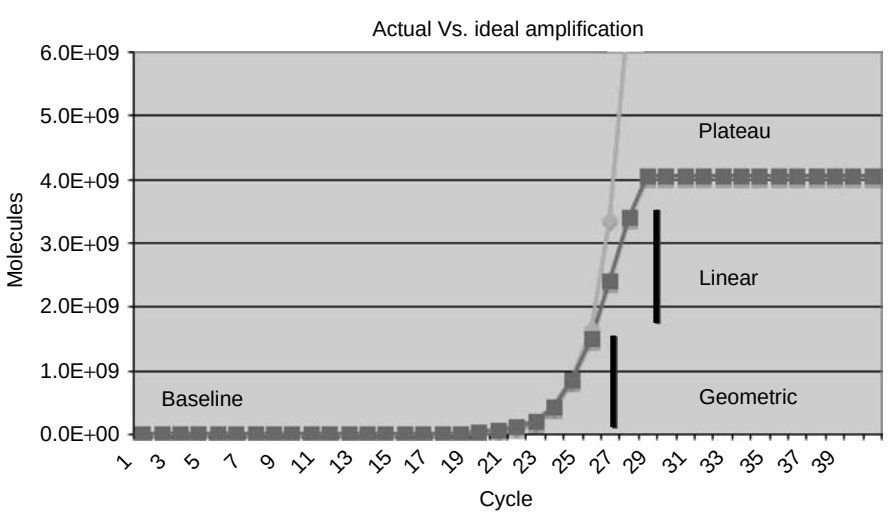

Fig. 1 Graphical representation of the fluorescent signals from an ideal versus an actual reaction over 40 cycles of real-time PCR. In an ideal PCR, there are two phases, a baseline where the signal is below the level of detection followed by a geometric increase in fluorescence that continues over the remaining cycles. However, in an actual reaction, there are four phases. As in the ideal reaction, there is a baseline followed by a geometric phase. However, the amplification becomes less than ideal, leading to a linear phase and finally a plateau where no further increase in signal occurs over the remaining signals.

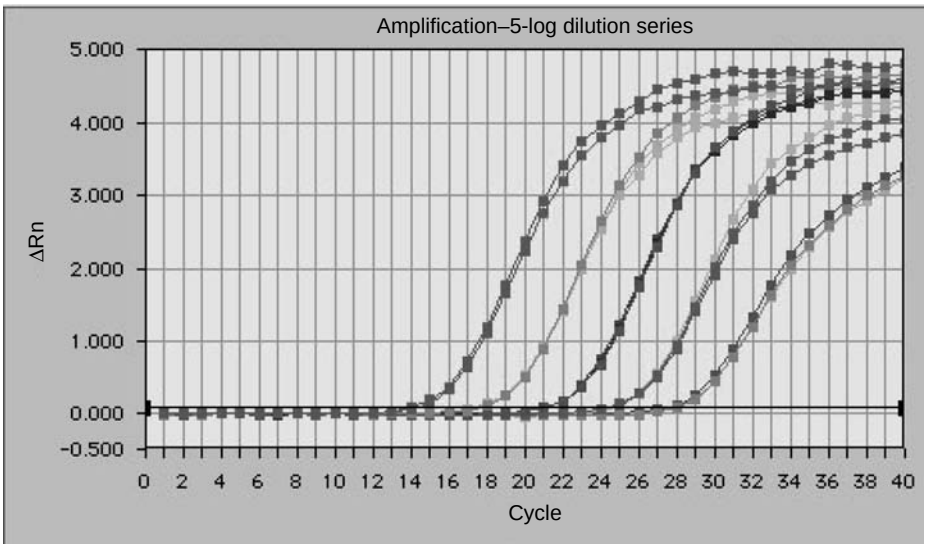

Fig. 2 Dilution curves from a real-time PCR experiment illustrating the four phases of a polymerase chain reaction. Amplification curves from a 5-log standard curve illustrating the four phases of signal generation in real-time PCR: baseline, signal being made, but not detectable by the instrument; geometric, detectable signal with maximal PCR efficiency; linear, postgeometric with slowly declining PCR efficiency; and plateau, no or very little new product made.

significantly lower template concentration after 30 or more cycles of PCR. Therefore, the only way to use the polymerase chain reaction for quantitative measurements is to compare samples in the early, geometric part of their amplification curves where the true differences in the samples can be detected. The ability to do so is what allows real-time PCR to be a *bona fide* quantitative technique.

1.2 The Reverse Transcription Reaction

Quantification of RNA begins with the reverse transcription reaction, which converts the RNA into a complementary DNA sequence or cDNA (also known as "copy DNA"). Two viral reverse transcriptase (RT) enzymes, used to make cDNA, are commercially available: AMV-RT from avian myeloblastosis virus and MMLV-RT from Moloney murine leukemia virus. AMV is a dimeric protein, while MMLV is a monomer. Both enzymes have RNase H activity, the ability to cleave RNA in a RNA/DNA hybrid. However, the activity in AMV-RT is much higher than that in MMLV-RT. This activity can be separated from the reverse transcriptase activity by site-directed mutagenesis. AMV-RT is a more processive enzyme (incorporates more nucleotides per unit time on the template) and is active at higher temperatures than MMLV-RT (42 °C vs. 37 °C for wild-type enzymes). Cloned variants of these enzymes have pushed the temperature limits to 58 °C for AMV-RT and 55 °C for MMLV-RT. Although both have been used for real-time PCR, MMLV-RT works best in practice, which is not consistent with the above characteristics. Both

enzymes are inhibitory to the PCR. The inhibition is enzyme concentration dependent. For this reason, it is important to keep the amount of reverse transcriptase as low as possible and still produce cDNA with 100% efficiency. The exact mechanism for the inhibition is not known. It has been proposed that although the high melting temperature of the PCR kills the reverse transcriptase activity, the DNA binding capacity of the enzyme may persist, forming a physical barrier that affects the activity of *Taq* polymerase. The higher temperature stability, dimeric structure, and greater RNase H activity of AMV-RT, may contribute to greater PCR inhibition than the MMLV-RT enzyme.

One major issue for RT-PCR is what primer to use in the reverse transcriptase reaction? The classic primer for cDNA library and standard RT-PCR has been oligo-dT, an oligonucleotide with 18 to 20 dT residues. Since most (roughly 95%) messenger RNA found in higher eukaryotes have poly-A tails at their 3′ ends, the oligo-dT will initiate cDNA synthesis utilizing mRNA molecules, but not rRNA or tRNAs. There are, however, two flaws in this approach for real-time quantitative PCR. The first is that long poly-A sequence stretches exist within mRNAs as well within some of the other RNA classes and these sites will initiate cDNA synthesis as well. Most important, however, is the fact that not all real-time PCR assays are targeted at sequences near the 3′ end of the transcript. Some mRNAs are over 15 Kb in length with 3′ untranslated regions (UTRs) that are 4 Kb or more. In a situation where you need to make enough cDNA to make a recombinant bacterial clone, this is not a problem. However, real-time PCR is a quantitative process where every molecule present should be detected for proper comparisons to be made. Priming with oligo-dT alone will not ensure true quantification for every transcript for every assay unless this caveat has been taken into consideration at the time of assay design. Another approach is the use of random primers that can be found as hexamers up to decamers. Random primers are short oligonucleotides that, as a population, contain every possible oligo sequence. Thus, there will be a subpopulation of the correct primer sequence for every 6-mer up to a 10-mer in every target sequence. The frequency with which a random primer will find a complementary sequence in any template is given by the formula 4^x, where x is the length of the primer sequence. For a 6-mer, a perfect match will occur every 4096 bases, while a 10-mer would be approximately every 1.05×10^6 bases. Using random primers, more than one cDNA may be synthesized for each transcript molecule resulting in the whole transcript being made into multiple cDNAs. The rationale is that every transcript in the total RNA population will be converted to cDNA with equal efficiency. However, there is no data to confirm that premise. Further, especially with random hexamers, there is the possibility that a primer will bind within the amplicon sequence resulting in only a partial cDNA, thus making that molecule undetectable in that assay. Keeping these caveats in mind, in practice, random primers can work well and are the method of choice if multiplexed assays are being performed where more than one transcript or gene will be amplified and detected simultaneously within the same reaction vessel. Some investigators hedge their bets and use a combination of both oligo-dT and random primers. For single-assay reactions, however, there is a third choice. That is to use an assay-specific primer. That is, the same primer that will

be used as the reverse primer in the following real-time PCR. This method will make, in theory, cDNA from every possible RNA template for that assay in precisely the correct location required for the following PCR. Unlike when oligo-dT or random primers have been used, the complexity of the resulting cDNA population is significantly lower when using an assay-specific primer. This means that the number of inappropriate priming events that could potentially occur during the PCR will be greatly reduced. Unlike when oligo-dT is used as a primer restricting the assayable region to the 3′-end of the message, use of assay-specific primers allows the amplified region of the assay to lie in any part of the transcript. As an aside, it should be mentioned that RNA can form complex structures resulting in self-priming and extra-assay cDNA synthesis even in the absence of any exogenous primer.

1.3 The Polymerase Chain Reaction

It is hard to think of another laboratory technique that has come along in our time that has had a more dramatic impact on scientific knowledge over such a broad range of subjects than the PCR. Real-time PCR has extended the sensitivity and specificity of nucleic acid detection greatly beyond what was available previously. The component that makes the polymerase chain reaction possible was the discovery of thermostable DNA polymerases that could withstand the multiple cycles of heating required to melt long DNA duplexes and cooling to temperature optimal for the succeeding polymerization steps. The first enzyme, and still the most commonly used, is *Taq* DNA polymerase from *Thermus aquaticus*. *Taq* is still one of the most processive thermostable enzymes commercially available. Most members of *Thermus* sp. also have a 5′-exonuclease activity that is required for cleavage of certain dual-labeled probes required for signaling.

There are two functional types of *Taq* commercially available, wild type and hot start. Hot-start enzymes are similar to their non-hot-start counterparts, but they have been reversibly inactivated by, (1) a heat activated mutation; (2) a small molecule; or (3) one or more antibodies directed at the active site of the enzyme. As the name implies, it is necessary to heat hot-start enzymes for their activation. The time varies greatly for enzyme activation with a minute at 94 to 95 °C being sufficient for the antibody inactivated enzymes while, 10 or more minutes at high temperature are necessary for mutated or chemically inactivated enzymes. All of these transient inhibition mechanisms are effective in repressing enzymatic activity until the initial melting step.

To fully appreciate the difference between using a hot-start *Taq* and non-hot-start *Taq*, you have to understand what happens to the targeted DNA sequence during the PCR. The most important part of the PCR is the first two cycles. During the first cycle, single-stranded cDNA or viral DNA is made into a double-stranded DNA molecule facilitated by new strand synthesis initiated from the forward primer in the reaction mix. Therefore, true amplification does not begin until the second PCR cycle for single-stranded templates. For a double-stranded DNA template, amplification begins during the first PCR cycle because both primers have a template. Nonspecific bands due to inappropriate priming events (false priming) occur during the first cycle of the PCR. This is true for single- or double-stranded

templates. This can occur because the reaction starts at a low temperature prior to reaching the first melting step of the PCR cycle, usually 94 to 95 °C. It is during this initial warming of the reaction mixture that false priming and extension by the DNA polymerase can occur. Thus, the use of a hot-start *Taq* that requires 1 to 10 min at 94 to 95 °C for activation will not be active in the lower temperature range during the first cycle of the PCR. For the remaining cycles of the PCR, the lowest temperature is the annealing temperature, usually 50 to 60 °C, depending on the T_m (melting temperature) of the primers. At this elevated temperature, false priming is much less likely to occur.

2 Fluorescent Dyes, Assay Chemistries, and Real-time Instruments

Real-time PCR has evolved from a single instrument that utilized one detection assay and two dyes for signaling to a large number of instruments, with a variety of assay chemistries and fluorescent dyes available. To fully comprehend how real-time PCR works, it is necessary to have a basic understanding of fluorescence and how different assay chemistries and real-time instruments work.

2.1 Fluorescence in Real-time PCR

All real-time instruments are based on the real-time detection of an increasing fluorescent signal over the length of an experiment. The increase in fluorescent signal detected is directly proportional to the corresponding increase in the number of template molecules made during amplification. It is this principle that makes fluorescence useful for quantitative real-time PCR as a detection mechanism.

Fluorescent compounds absorb light as photons within a narrow, optimal wavelength. Light at the optimal wavelength is called the excitation wavelength. When a fluorescent molecule absorbs a quantum of light, the molecule is boosted to a higher energy state. This higher energy state is transient and short lived. This higher energy state decays, falls back to the ground state, and light is given off. This form of emission is called fluorescence. The emitted light is necessarily at a longer wavelength than the initial excitation wavelength. This shift in wavelength is called a Stoke's shift. For every fluorescent dye, there is an optimal absorbance or excitation wavelength and an optimal, shifted emission wavelength used for detection.

Having a steady state fluorescent signal throughout the PCR would not be useful. What is required is a very low initial signal that increases in proportion to the products made during the PCR to a high level that is easily detectable. One way to achieve this kind of signal production is to take advantage of the principle of fluorescence resonance energy transfer or FRET. During FRET, one dye acts as a donor and the other as the acceptor. The excited donor dye transfers its emission energy to the second dye, and that energy is used to excite the electrons in the second dye. The excited acceptor dye then emits light at its emission wavelength which can then be detected. Under conditions of FRET, the light emitted from the donor dye is quenched. The efficiency of energy transfer between the two dyes falls as the 6th power base 10 of the distance between the dye molecules. Therefore, the donor and acceptor molecules have to be in close proximity (10–100 Å). Table 1 shows a list

Tab. 1 Common dyes used in real-time PCR.

Dye name	*Max. Ab [nM]*	*Max. Em [nM]*
DABCYL	453	None
Fluorescein™	492	520
6-FAM™	494	518
JOE™	520	548
TET™	521	536
Yakima Yellow™	526	448
BHQ1™	534	None
HEX™	535	556
VIC™	538	554
Cy3™	552	570
TAMRA™	565	580
ROX™	585	605
LightCycler640™	625	640
Cy5™	643	667
LightCycler705™	685	705

of the most common dyes used in real-time PCR.

There are three classes of fluorescent dyes used in real-time PCR based on their function: (1) the reporter, which is the dye monitored during the experiment; (2) the quencher, the dye responsible for quenching the reporter signal in FRET-based assays; and (3) the reference, a dye common to all reaction vessels used to normalize the signal.

- *Reporter dyes:* the most used reporter dye is 6-FAM, a fluorescein derivative with a green emission. This dye can be efficiently excited by the argon laser found in the initial real-time machines (see the following) and has remained as the reporter dye of choice of many chemistries in use today. The first multiplexing experiments, running two or more assays simultaneously in the same reaction, were done with laser-based instruments. For this reason, other fluorescein derivatives with progressively longer excitation wavelengths were made available. The instruments available today can utilize a much wider variety of reporter dyes, from the green through the red spectrum. Another reporter dye in wide use is SYBR Green I that has very low fluorescence as a free dye, but increases dramatically when bound to double-stranded DNA. The number of dyes now on the market that are suitable for use as reporters is continuing to increase.
- *Quencher dyes:* when FRET is used as the signaling mechanism, the fluorescence of the reporter dye is quenched by a neighboring dye. There is at least one exception to this rule and that will be discussed below. There are two classes of quencher dyes; those that emit light and those that do not. The initial FRET pair was 6-FAM and TAMRA, a rhodamine derivative that emits light in the red spectrum. The FAM/TAMRA combination works very well, but TAMRA is sensitive to the alkaline hydrolysis step used to remove oligonucleotides from the solid support used for their synthesis. The result is a comparatively low yield of functional FAM/TAMRA molecules at the same or higher cost compared to a FAM/dark dye oligonucleotide. The dark dyes represent the second class of quenchers. As the name implies, no light is emitted from these dyes although they work very well as quenchers and are available for a large spectrum of reporter dyes. The initial dark dye was DABSYL and is used today when a high degree of quenching is not desired when the reporter and quencher are transiently separated. An example is a molecular beacon (see the following). A second generation of dark dyes is the black hole quenchers or BHQs. As the name implies, these dyes are very efficient quenchers and emit no light. There are three different variants of these dyes available, BHQ1, 2, and 3.

The higher the number, the more red-shifted the reporter dye used in a FRET pair. BHQ dyes are particularly effective when maximal quenching is desired. An example would be Taqman® probes (see the following).

- *Reference dyes:* All real-time PCR instruments have the capability of monitoring multiple reactions during the course of an experiment. It is important that the signal from each reaction vessel be the same so that the results can be compared. One way to achieve this is to include a reference dye in the reaction mix. The most commonly used reference dye is ROX, a red dye that is well shifted from 6-FAM. In practice, any dye can be used as a reference dye as long as it doesn't significantly spectrally overlap the reporter dye(s) in the reaction. Not every instrument requires the presence of a reference dye in the reaction. This will be discussed in the following text. However, signal normalization is a part of the data analysis of all real-time PCR software packages.

2.2 Real-time Instruments

There are three major components in every real-time PCR instrument: (1) the light source; (2) the light gathering and detection system; and (3) the thermocycler. Although each real-time platform is designed to measure a fluorescent signal from the same fluorescent dyes, the way each instrument excites and detects those dyes can be quite distinct. The mechanism driving the heating and cooling cycles can also vary. These differences are driven partially by patent issues and also by an on-going desire to build a more compact, less expensive instrument that delivers comparable or better performance when compared to earlier models. A basic understanding of how real-time PCR instruments work is important for fully understanding how the data is produced.

Light Sources: All the assays currently in use utilize fluorescent dyes as a read out. Therefore, the light source has to illuminate the fluorescent dye in use near its optimum excitation wavelength. There are three different types of light sources in use today: lasers, quartz halogen-tungsten lamps, and light emitting diodes (LEDs).

Lasers emit light that is coherent, monochromatic, and directional. The only two real-time platforms to use a laser are the Applied Biosystems 7700, still widely used but no longer manufactured, and the 7900. Both use an argon laser that emits light at 488 nm. The advantage of a laser is the light is monochromatic and intense. The disadvantage is the reporter dyes used have to have an excitation maximum near 488 nM. Argon lasers work well for fluorescein dyes attached to oligonucleotides such as 6-FAM or the free dye SYBR Green I. However, as the excitation wavelength for a reporter dye moves toward the red spectrum argon lasers lose their ability to excite the dye.

Quartz halogen-tungsten lamps are capable of emitting a steady beam of light from 350 nm to over 2000 nm and are sometimes referred to as "white light" sources. For this reason, a filter must be employed that only allows a narrow band of light at or close to the optimal excitation wavelength for a particular fluorescent dye. Most machines with quartz halogen-tungsten light sources have 4 or 5 filter sets to include a broad range of dyes with excitation maxima distributed over the green–red spectrum.

Light emitting diodes or LEDs are semiconductors that emit light within a narrow bandwidth, usually 30 to 40 nm, when stimulated by an electric current. Rather than using filters, as mentioned earlier, one or more LEDs are used that emit light at increasingly longer wavelengths to match the desired excitation maxima of the reporter dye. Thus, for every dye that has a significantly different excitation wavelength, a matching LED for the new wavelength must be present. The light from most LEDs does not have the light intensity of a laser or even a quartz-halogen bulb, but there is very little heat generated during their use and they require less power to operate.

Real-time Instruments: – The following is a description of the light path in three different real-time instruments. Each one is an example of a machine that utilizes one of the three light sources described above. Although the majority of the new instruments on the market use quartz halogen-tungsten lamps, there are still those on the market that use LEDs or a laser as a light source.

The Applied Biosystems 7900HT – This is the only machine available in the market today that uses a laser light source. However, it is also the only machine that can use either 96- or 384-well plates. It may well be that the high intensity of the laser light source is required for the excitation of reporters in 384 wells. The laser emits light at 488 nm that is directed at a beam splitter that redirects the light over a scan head with 16 lenses, sitting just above the plate in the heated cover. The lenses focus the light into either 8 or 16 wells at a time, depending on which plate is in the instrument. The emitted light comes back up through the lenses and passes through a dichroic mirror that deflects the individual beams of emitted light onto a CCD camera for detection. The scan head goes down each column of the plate, using a stepper motor, and collects the emitted light from the entire plate. The scan head returns to the first column and a new scan ensues until all the cycles have been completed. Multiple scans are made for each column during the melting and annealing/extension phases of each cycle. The plate rests in a metal block that is heated and cooled by a Peltier thermocycler. Thermocyclers, based on the Peltier principle, utilize electron flow to heat and cool the metal block rather than a conventional heater unit opposed by a cooling device.

The Roche Lightcycler and Lightcycler2 – These instruments use a single blue LED light source in an optical range similar to the argon laser. The Lightcyclers have a 32-sample capacity. Each reaction chemistry is contained within a thin capillary tube, which results in a high surface to volume ratio. Cycling times in the Lightcycler units are much reduced because heated air is used instead of a metal block for heating and cooling. This is due to the excellent heat transfer possible using capillary tubes. The samples spin past the light source like tubes in a centrifuge using a stepper motor. The excitation light goes through each tube in sequence and the emitted light is focused on the capillary tip, goes through a dichroic mirror, and is then read by either 3 (Lightcycler) or 6 (Lightcycler 2) photo detection diodes, each specific for an increasing and narrow wavelength range.

The Stratagene Mx3000p – This instrument utilizes a quartz tungsten-halogen (white light) source. The excitation range is from 350 to 750 nm with a similar detection range. Excitation and emission wavelengths are selected using filter pairs

specific for each reporter dye to be detected. Light passes through an excitation filter into a fiberoptic cable and into the well of a 96-well plate. The emitted light passes back up the fiberoptic cable through an emission band pass filter and onto a photomultiplier tube (PMT) for detection. The plate is scanned once for each fluorescent dye with the filter wheels moving in tandem to bring the new filter pair into the optical path. The thermocycler is a combination of Peltier, convection, and resistive technologies unique to the Stratagene machines that results in uniform temperatures across the block. The Stratagene Mx3000p is just one example of many instruments on the market that use a quartz halogen-tungsten light source.

2.3 Selected Assay Chemistries

Just as understanding the mechanical workings of the real-time detector being used are important, knowing the principles underlying the fluorescent signaling system utilized in an assay are equally important. There are three mechanisms used to generate a fluorescent signal utilized in the large variety of assay chemistries available for real-time PCR. First is the binding of a free fluorescent dye to the newly amplified double-stranded DNA, second is a fluorescent dye that is associated with one of the PCR primers that generates a signal when incorporated into the amplified product, and third are fluorescently labeled oligonucleotides, independent of the primers called probes, that bear the fluorescent signal. There are three probe types and the method of signaling is different for each of them. In all of these assay types, the signaling molecule has to have a very low initial fluorescence that increases in direct proportion to the amount of DNA made during the PCR. Except for the intercalating dyes, all the other chemistries available depend on FRET.

SYBR Green I – The simplest assay mechanism involves the incorporation of a free dye into double-stranded DNA as it is synthesized (Fig. 3). SYBR Green I

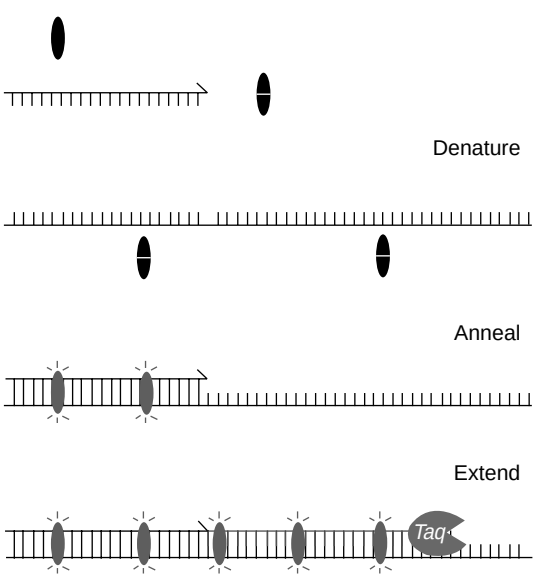

Fig. 3 Graphical representation of SYBR Green I in fluorescent signal generation during a PCR. Free dye has very low fluorescence and will not bind to single-stranded, denatured DNA. During primer annealing, a double-stranded structure is formed and SYBR green dye begins to bind and emit a fluorescent signal. During primer extension by *Taq* DNA polymerase, the number of SYBR green dye molecules bound per double-stranded molecule and the fluorescent signal increases proportionally. The process is repeated in each cycle with increasing total fluorescence (see color plate p. xxxiv).

has a very low fluorescent signal when in solution as a free dye and a low affinity for single-stranded nucleic acids at the concentration used for real-time PCR. The latter characteristic is important, as a high affinity for single-stranded DNA would be inhibitory to the progression of the PCR. Following the synthesis of double-stranded DNA, the dye binds to the minor groove and a dramatic increase in fluorescent signal can be detected. Initially, dyes such as ethidium bromide were used for this purpose. However, ethidium bromide has a relatively high signal as free dye, has a higher affinity for single-stranded DNA, and the increase in fluorescence over background following binding is fivefold lower than SYBR Green I. The positive aspects of using a dye-based real-time assay are as follows: (1) SYBR Green I is a relatively low cost signaling molecule; (2) only one dye is required for all assays; and (3) assay design is limited to two simple primers. The negative aspects of a dye-based assay are as follows: (1) lower specificity as all double-stranded DNA molecules made during the PCR will make a signal; (2) setting up a new assay requires running the initial products on a real-time instrument as well as a DNA acrylamide gel to check for nonspecific amplification products and primer dimers that can add a false signal, in the case of the former, and increased assay background noise from the latter; and (3) the use of a more expensive hot-start *Taq* is a must to reduce inappropriate priming events that can occur at sites with similar complementary sequences during the initial PCR cycle.

Primer-based signaling assays – As the title implies, the signal is generated by a reporter dye on one of the two primers required for the PCR. There are many imaginative reporter–dye primer combinations available that can be used in real-time PCR. Only two will be presented here. These two assay types represent the extremes in the level of complexity found in primer-based signally systems.

1. LUX primers (light upon extension) have a 4 to 6 base extension on the 5′ end of the primer complementary to the internal sequence of the primer that is complementary to the target sequence. A single reporter dye is covalently linked to an internal base near the 3′ end of the primer. The 5′ extended sequence forms a hairpin loop and stem structure within the primer sequence (Fig. 4). When the reporter dye is in close proximity to the double-stranded stem, particularly to guanidines, the signal is quenched without the need for another dye. Guaninidine residues are able to act as quenchers, similar to but not as efficient as an appropriate FRET dye pair. When the primer becomes incorporated into a PCR product, the stem structure is broken and this relieves the quenching of the fluorescent signal. The maximum increase in signal is about 10-fold over background. The advantages of this system are as follows: (1) relatively low cost, only a single dye is required on one of the primers; (2) relatively simple assay design; and (3) can use multiple, assay-specific reporter dyes. The disadvantages are as follows: (1) an acrylamide gel is necessary to determine assay specificity; (2) a hot-start *Taq* is necessary for optimal primer assay fidelity; and (3) it has the lowest signal to background fluorescent signal of any assay.
2. Scorpion primers are named after their mode of binding during the signaling phase and represent the most complex primer-based signaling system. Their advantage is that they have the highest

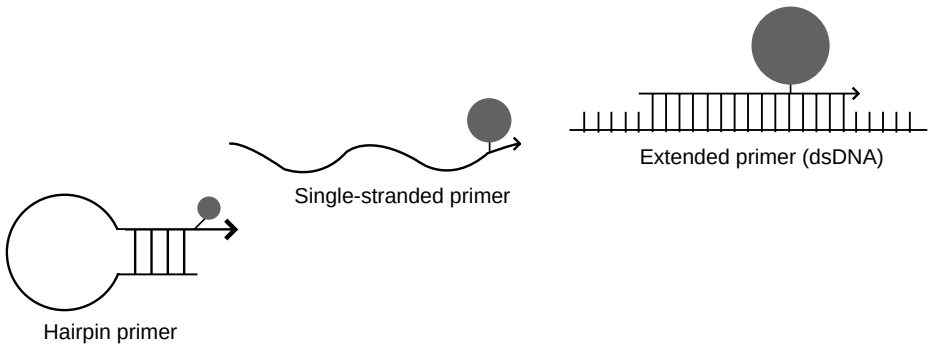

Fig. 4 Structure and mode of action of a LUX primer (light upon extension). LUX primers are inherently quenched because of a designed, short self-complementary sequence added to the 5′ end of the primer. The fluorescent moiety on the 3′ end of the primer is quenched in this confirmation. Following the melting and subsequent annealing phase, the primer binds to complementary sequence on the template and becomes incorporated into the new strand, leading to a significant increase in fluorescence.

template specificity of any primer-based assay. Two variations of the scorpion primer assay have been described, one based on a stem and loop and a newer one based on a linear duplex. Both result in efficient quenching of the reporter signal. The loop version is shown in Fig. 5 and will be described here, but they both have identical fluorescent reporter mechanisms. The basic components of scorpion primers are (1) the actual primer sequence; (2) a PCR stopper moiety to prevent PCR read through; (3) a probe-specific sequence; and (4) a fluorescence detection system comprised of reporter and quencher dyes. The primer structure begins with a standard complementary sequence for the transcript of interest on the 3′ end. Upstream from the primer sequence is the PCR stopper that prevents the upstream sequence from becoming double stranded followed by an internal DABSYL quencher dye. Next comes a stretch of oligonucleotides that can anneal to a complementary internal sequence in the PCR product, followed by 5 to 6 bases of complementary sequence to an internal region just upstream of the DABSYL dye that can form a stem structure bringing the reporter dye at the 5′ end into close proximity to the DABSYL quencher (Fig. 5). The reporter signal is quenched by FRET between the two dyes at the annealing temperature. Detection begins when the 3′ end of the scorpion primer binds to a complementary region of the single-stranded template and is extended during the extension phase. In the next PCR cycle, the resulting DNA duplex and internal stem of the scorpion primer are melted, resulting in a single-stranded molecule with the scorpion reporter incorporated. At the annealing temperature, the internal loop portion of the scorpion primer complementary to a stretch of sequence within the newly synthesized strand can now anneal to the newly made template sequence. The reporter dye is no longer quenched by FRET because the reporter dye is physically distant from the DABSYL and reporter signal

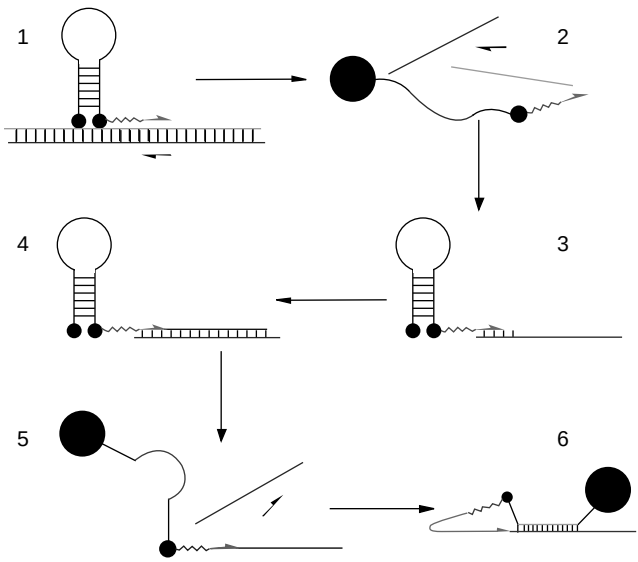

Fig. 5 The complex structure and mechanics of fluorescent signal generation by a scorpion primer. (1) dsDNA template and complete scorpion primer with a FRET quenched reporter dye; (2) melting of template and primer structure transiently releasing FRET quench of reporter dye; (3) template-complementary region of the 3′ end of the primer anneals to ssDNA template; (4) primer extension by *Taq* DNA polymerase; (5) second melting step making the newly made DNA single stranded; (6) internal complementary sequence of the scorpion primer anneals to an internal region of the newly made template releasing the reporter dye to generate a signal.

is detected. The scorpion primer is the only primer-based assay system that has a comparable template specificity to a probe-based assay. It is also the most complex and costly to make.

Probe-based signaling assays – Probe-based assays introduce one or two new oligonucleotides whose sole purpose is to provide the fluorescent signal for the assay. These extra, fluorescently labeled oligonucleotides anneal to a region in the amplicon sequence that lies between, and does not overlap, the two primers. The primers are still the driving force for amplification and they determine the efficiency and sensitivity of the assay. However, with the introduction of the intervening oligonucleotide probe, the primers are no longer the sole determinants of assay specificity. As in primer-based assays, the amplified sequence is determined by the primers alone. However, unlike primer-based assays, fluorescent signal will be realized solely from those amplicons that also have a complementary sequence for the fluorescently labeled probe.

There are three different probe designs in common use and each relies on FRET to either quench or induce the reporter signal. All three of these assay formats have increased specificity over SYBR Green I or simple primer-based systems. This is because both primers and the probe have to anneal to the correct target sequence

and work together as a unit for signaling to occur.

1. Taqman® assays are the original chemistry for real-time PCR first commercialized by Applied Biosystems in 1996 (Fig. 6). The probe consists of a dual-labeled oligonucleotide with a reporter dye on the 5′ end and a strong quencher dye such as a BHQ, on the 3′ end. The probe is designed with a higher T_m than the two primers, usually 8 to 10 °C, to ensure it will bind to the template before the primers as the temperature is lowered from the melting to the annealing setting. When free in solution during the annealing step, the probe is quenched. However, when the probe is able to anneal to a single-stranded complementary region that lies just 3′ to one of the primer binding sites, it is partially dequenched because the reporter and quencher dyes are separated and FRET is much reduced. It is critical that the probe bind to every available template prior to the initiation of primer extension to ensure that they all are detected. As *Taq* DNA polymerase synthesizes a new strand, it displaces and cleaves the probe molecule utilizing an inherent 5′-exonuclease activity within the enzyme that is separate from the polymerase activity. For the most part, only DNA polymerases from *Thermus* sp. (such as *Taq* from *T. aquaticus*) have this 5′

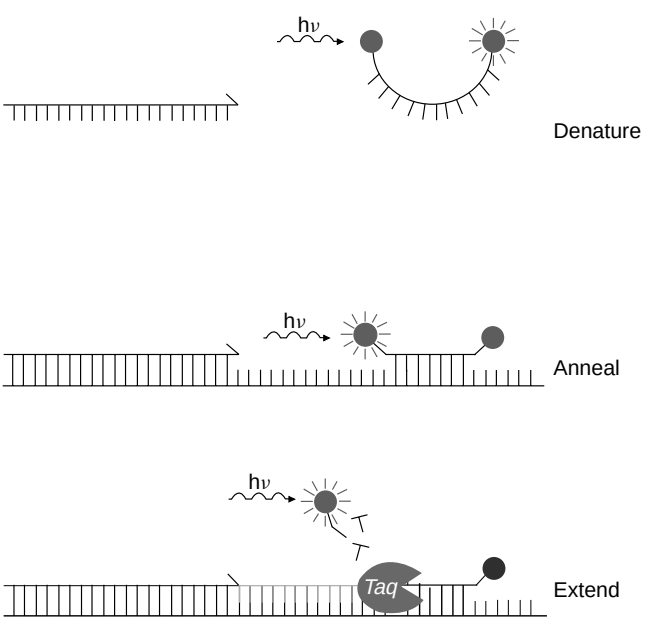

Fig. 6 Graphic showing fluorescent signal generation by a Taqman® probe. When free in solution, the reporter fluorescent dye is quenched by FRET. Owing to a higher T_m, the probe anneals to a complementary template sequence before the primer on the same strand. During primer extension by *Taq* DNA polymerase, the probe is displaced and degraded by a nuclease present in *Taq* DNA polymerase. Release of the reporter dye from the quencher dye allows the full signal to be realized.

nucleotidase activity. Following probe cleavage, the reporter dye is now free in solution and permanently dequenched after being physically removed from the quencher dye. The Taqman name stems from this cleavage step, from the computer game Pac Man.

2. Molecular beacons are similar to Taqman® probes in that they are dual-labeled oligonucleotides having a reporter dye and a DABSYL quencher. Once again, the T_m value of the probe is higher than the two primers to ensure probe binding prior to primer extension. The difference in probe structure, compared to Taqman® probes, is the addition of 4 to 6 extra self-complementary bases on both ends of the probe sequence that do not anneal to the target sequence (Fig. 7). At the annealing temperature of the PCR cycle, this forces the two dyes to come into close proximity with one another when a short stem is formed between the two short complementary sequences, maximizing the quenching of the reporter by the quencher dye via FRET when the probe is free in solution. A fluorescent reporter dye signal is generated following the annealing of the probe to a complementary sequence on the template strand. The annealing of the molecular beacon to an intervening sequence between the two primers physically separates the reporter dye from the quencher dye and FRET is no longer possible. Unlike a Taqman® probe, however, molecular beacons do not depend on cleavage by the DNA polymerase to generate a signal. The probe is simply displaced during new strand synthesis. Displacement is facilitated by keeping the T_m differences between the probe and the two primers at a lower temperature, such as 5 °C instead of 10 °C for Taqman® probes. Overall,

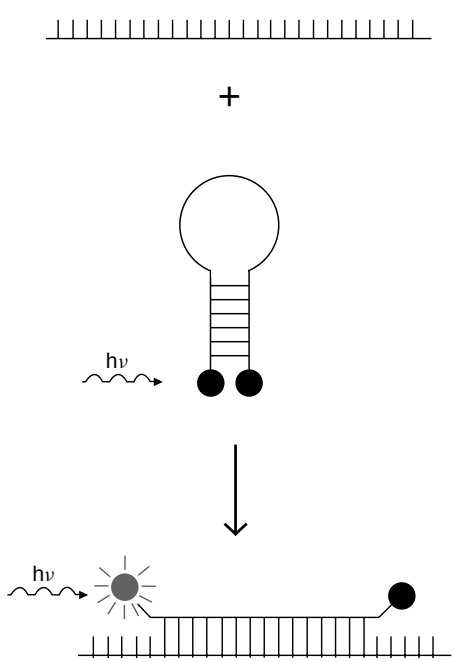

Fig. 7 Mechanism of signal generation for a molecular beacon probe. Designed complementary sequences at the 5′ and 3′ ends bring the reporter and quencher dyes in close proximity leading to much reduced reporter signal via FRET. Following template and beacon melting and annealing, full reporter signal is realized owing to loss of FRET. Unlike Taqman® probes, molecular beacons are not degraded and produce a signal only when annealed to the template.

when compared with Taqman® probes, molecular beacons have a higher signal to noise ratio due to optimal FRET between the reporter and quencher dyes. They have also been reported to work better on high G/C templates and show greater sensitivity to single-base mismatches, which makes them ideal for SNP (single nucleotide polymorphism) detection studies. However, they can be more challenging to design than Taqman® probes.

3. Hybridization probes are used primarily in conjunction with the Roche Lightcycler real-time detector, although they could be used on any real-time machine that can detect the reporter signal. For this assay system, two oligonucleotides, both with equal and higher T_m values than the primers, are made that again anneal to complementary sequences between the two PCR primers (Fig. 8). One is labeled with a donor dye on the 3′ end and the other has a signaling acceptor dye on the 5′ end. Unlike the Taqman® and molecular beacon probes that utilize FRET to quench the reporter signal, hybridization probes use the fluorescent signal induced by FRET as the reporter signal. The fluorescent signal is made when the two probes anneal to a complementary single-stranded DNA template. The excitation light stimulates the acceptor dye (green) and that emission energy is transferred to the acceptor dye (red) by FRET. The light emitted by the acceptor dye is monitored and recorded by the real-time

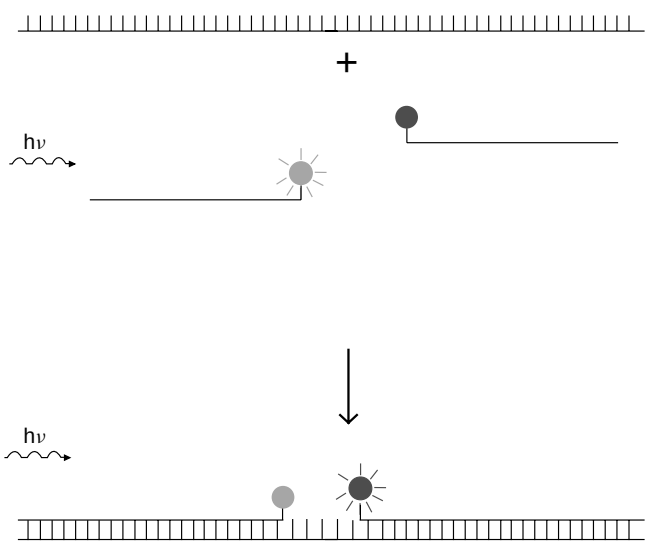

Fig. 8 Graphical representation showing the mechanism of action for a hybridization probe. Two oligo probes bearing a single dye each, one with a fluoroscein dye at the 3′ end and the other with a rhodamine dye at the 5′ end. When the two oligos anneal to a complementary template, the fluoroscein dye is excited by the light source in the instrument and transfers its energy to the rhodamine dye via FRET. FRET can only occur when the two dyes are in close proximity. The instrument is set to detect the rhodamine signal. The signal from the fluorescein dye is not monitored.

instrument as a positive signal. While free in solution, no signal is generated by the two probe molecules at the correct wavelength as the acceptor dye emission wavelength is not monitored and the reporter dye will not absorb the wavelength of the light source. As with molecular beacons, the probe molecules do not have to be cleaved. The advantage of this system is a low background. Similar to the two probe-based assays described above, this probe-based system is highly template specific. However, designing two tandem oligonucleotide probes that anneal between a pair of primers that will work with high efficiency can be even more challenging than the other two probe systems.

3 RNA and DNA Isolation for Real-time PCR

The key factor for isolating good DNA or RNA is how the sample is treated prior to nucleic acid isolation. If the sample is not treated properly and the macromolecules are degraded *in situ*, no amount of care during the nucleic acid isolation phase can rescue the sample. For animal, plant, or microbial sources, it is imperative that they are properly treated following harvest by either initiating isolation of nucleic acids immediately or processing the sample for storage that will preserve the RNA and/or DNA *in situ*. Quick-freezing of the sample in LN_2 followed by storage at −80 °C works well to stabilize the nucleic acids in most samples. A relatively new option for sample storage from Ambion, RNALater, can be used to store samples for short times at room temperature or −20 °C for extended periods of time.

Following sample collection, the next most important thing in the isolation process is to make sure the tissues or cells are disrupted rapidly in the isolation medium. For this purpose, a rotor-stator homogenizer or a bead-beater works very well for most samples.

3.1 Isolation of DNA

Procedures for the isolation of DNA from animal tissue or cells involve incubating the homogenate in buffers containing EDTA, SDS, and proteinase K at elevated temperatures (50°–60 °C). This will destroy any remaining cellular structure and contaminating proteins while preserving the DNA. Ribonucleases A and T1 are added to degrade the RNA. The digest is then extracted with phenol, phenol-chloroform, and chloroform followed by ethanol precipitation of the DNA. Isolation of DNA from some eukaryotic microbes and most plants is more involved as they have carbohydrates that copurify with the DNA. For these samples, an extra purification step utilizing CTAB (cetyltrimethylammonium bromide) will free the DNA from the contaminants. It is not necessary to worry about shearing the DNA during isolation, as the amplified region for real-time PCR is quite small, in the range of 75 to 250 bases. A final purification over a silicon-based column will result in DNA of high purity suitable for PCR.

3.2 Isolation of RNA

The isolation of RNA can be more problematic than DNA primarily because RNA is much more susceptible to degradation prior to and during isolation. The primary

reason for measuring transcripts by quantitative real-time RT-PCR is to compare mRNA levels between control and experimental samples. To ensure transcript levels that reflect the original status of the source, it is important to stop any further changes in the RNA population as rapidly as possible after harvesting the sample. The best method for RNA isolation can depend on the origin of the sample. In general, a combination of thorough homogenization in a guanidinium/phenol-based buffer followed by purification on any of the commercially available silicon-based columns results in excellent yields and the highest quality RNA for real-time PCR. This approach will work for the vast majority of samples from microbes to mammalian tissue. As with DNA isolation, those sources that present problems for DNA isolation will present the same problems for RNA isolation. Methods are available in the literature for those particular tissues or organisms. One exception is isolating RNA from fixed and embedded tissue. We have found that a detergent-based isolation medium followed by proteinase K digestion of deparafinized tissue have proven to work best with limited amounts of starting material where extractions by organic solvents are not practical for maximal yields.

Two excellent sources describing multiple methods for the extraction of nucleic acids are "Molecular Cloning" and "Current Protocols in Molecular Biology."

4 Standards Used for Real-time Quantitative PCR

The term "standard," when used in the context of real-time PCR, means any nucleic acid template that bears the amplicon sequence of the gene or transcript being interrogated. Standards are used to quantify the gene or transcript levels in unknown samples. This is usually accomplished by making a dilution series, 10-fold decrements are common, over a 5 to 6 log range of standard concentration. By assaying the dilution series of a standard along with samples of unknown target levels, it is possible to determine the level of the target sequence in the unknown samples relative to the standards.

The term "absolute" has been used in conjunction with real-time quantitative PCR. This is a misnomer as no set of universally used standard assays with independently measured and verified standards are currently available. The ideal situation would be a single, quantified standard from a central agency such as NIST (National Institute of Standards and Technology, USA) for every nucleic acid target used in a real-time PCR assay. However, it is currently not possible to measure the number of molecules in any standard preparation that will be functional in the PCR, other than by real-time PCR itself. In the absence of this ideal situation, the best that can be achieved is a measurement of target molecules in any sample relative to the nucleic acid standard being used. Standards currently in use are quantified on the basis of their total nucleic acid concentration using absorbance at A_{260} or fluorescent dye–binding assays such as picogreen, ribogreen, or oligreen from Molecular Probes/Invitrogen.

4.1 RNA Standards

Unquestionably, the best standard to use for quantitative real-time RT-PCR would be an RNA molecule bearing the target

region of the transcript being measured. This is because both the standard and the sample RNA target have to be converted into cDNA by reverse transcriptase as well as amplified by *Taq* DNA polymerase. Thus, the efficiency of both synthetic processes will be similar in both reactions for the RNA standard and RNA target sequences in the unknowns.

The simplest way to make an RNA standard is to dilute a concentrated, total RNA preparation. The problem with using total RNA to make a standard dilution curve is the highly variable amount of different transcripts in any RNA population. Thus, the number of dilution points detected will be variable for any given assay. Although it would be impractical to quantify unknown RNA samples using dilutions of a total RNA preparation, it does allow a measure of PCR efficiency (from the slope) and a determination of the relative low-end sensitivity of the assay.

The *in vitro* production of RNA of the size of an entire RNA molecule (some mRNAs are over 15 Kb in length) is not practical. However, it is possible to synthesize full-length RNA molecules *in vitro* that are <500 bases in length, in large quantities. To do so requires a double-stranded DNA template bearing the sequence that will be amplified during the PCR. The DNA template can then be cloned such that it is physically placed 3′ of a bacterial viral promoter sequence, usually T7 RNA polymerase. Alternatively, the T7 promoter can be added to the DNA target using a forward PCR primer with the promoter site added to the 5′ end of the primer. In either case, large quantities of RNA can be made per engineered DNA template. This is because RNA polymerases can use the same DNA template multiple times during a synthesis reaction. The full-length RNA is then separated and isolated from any shorter molecules in the total synthesis product utilizing an acrylamide gel. RNA standards are very stable over many years if stored at −80 °C and treated with care to avoid ribonuclease contamination. RNA standards can be quantified using an A_{260} measurement or, more specifically, by using the ribogreen fluorescent assay from Molecular Probes/Invitrogen.

The easiest RNA standard would be one that was synthesized *de novo* in the laboratory or by a company. The cost of RNA oligonucleotide synthesis has dropped dramatically in the last few years, mostly due to the increasing use of RNAi in research clinical applications. However, synthetic RNAs, the length of even the shortest assay amplicons (60–70 bases), are still very expensive compared to their DNA counterparts.

4.2 DNA Standards

From a practical point of view, DNA standards are much easier to use (no reverse transcriptase step required) and can be made in multiple ways inexpensively. The simplest DNA standard is the purified PCR product from one or more RT-PCR or PCR reactions. Once it has been determined that there is a single amplification product by polyacrylamide gel electrophoresis and excess primers and other reaction components have been removed, the PCR product can be quantified by A_{260} or using the picogreen fluorescent assay (Molecular Probes/Invitrogen). This method is easy, inexpensive, and a quick way to make a standard curve over a broad template concentration to test the efficiency and low-end sensitivity of a new real-time PCR assay.

Another approach to making a DNA standard is to clone an amplified double-stranded DNA product made from cDNA or a DNA template into a plasmid. Once a recombinant bacterial clone is established, a permanent resource for that DNA template is available. Making the DNA standard is as simple as a plasmid preparation and quantification of the purified DNA by A_{260} or picogreen assay.

As a standard for DNA samples, both of these methods are excellent. As standards for RT-PCR, however, both DNA standard types have two drawbacks: (1) DNA standards are not linked to the reverse transcription reaction and (2) cDNA is single stranded and DNA is double stranded, which means that when a single-stranded cDNA is being made into a double-stranded molecule during PCR cycle 1, a double-stranded standard is already being amplified. This must be kept in mind when using dsDNA standards to quantify RNA samples. A more detailed discussion of how standard curves are used to quantify samples is given in Sect. 6.2.

The easiest way to make a DNA standard that is also single stranded is to simply synthesize it yourself or have it made by a company on a DNA synthesizer. For this strategy to work, however, the assay amplicon has to be no longer than 100 bases, preferably less than 80. The ability of modern DNA synthesizers to make long oligonucleotides with high fidelity is quite good. However, there can be a subset of molecules in the oligonucleotide synthesis that are not completely deprotected (still have a blocking group attached), that have a

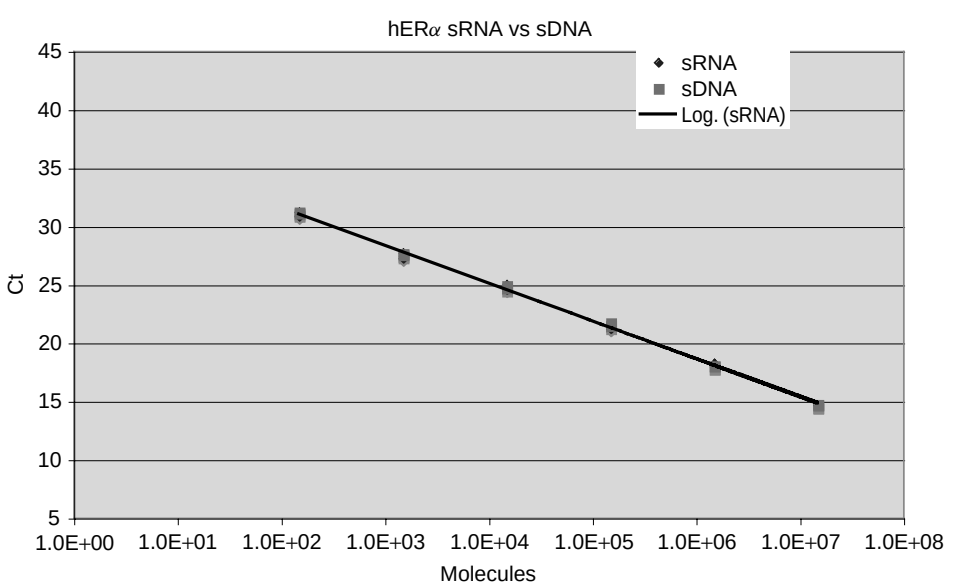

Fig. 9 A comparison made using an *in vitro* transcribed RNA standard to a synthetically manufactured oligo DNA standard for the human estrogen receptor α-(ERα) assay. Both standards were run on the same 96-well plate with identical two-step chemistry. RNA and DNA standards can work identically under the same assay conditions.

missing or inappropriate base in one or more positions and thus cannot serve as a fully functional template in the polymerase chain reaction. Yields of full-length molecules, following isolation on a polyacrylamide gel, are such that one synthesis will easily last the lifetime of the investigator running the assay.

The real question is, do DNA oligo standards yield equivalent results compared to RNA standards? Our experience has been that they do make comparable standards. One example is the comparison of standard curves generated from a single-stranded DNA oligonucleotide versus an *in vitro* transcribed RNA for the human estrogen receptor α-transcript (Fig. 9). The slopes and values for the two standards are nearly identical, showing that the RNA has been converted to cDNA and both templates amplified equally over a 6-log range of template concentrations. One caveat is that assays with amplicons over 100 bases have been difficult to test in the same way because of problems in synthesizing longer oligonucleotides. Nonetheless, in multiple assays, single-stranded DNA oligonucleotides under 100 bases have been shown to work just as well as their RNA counterparts. This means that for the relatively short amplicons used in real-time PCR assays, the RT efficiency has to be near 100% when assay-specific primers are used.

5 Real-time PCR Assay Design

The hallmarks of real-time PCR are the specificity, sensitivity, and efficiency of the assays. An assay is specific if it can be demonstrated that the detected target sequence is the only one being measured in the real-time PCR. Sensitivity is the lowest detectable amount of template that still maintains a linear relationship in a dilution series of standard template. A detectable signal for an assay that no longer falls on the standard curve is not usable for quantification and is below the sensitivity of the assay. The efficiency of an assay is 100% when every molecule in the reaction is exactly doubled. The primary factor that determines if an assay will have high levels of these three characteristics is determined by how well it is designed.

5.1 Available Software

Many multifaceted molecular biology software programs available today, such as MacVector, Vector NTI, Oligo, and others, have subroutines for designing compatible primer pairs for primer-based real-time PCR. Designing assays with an intervening oligonucleotide probe, however, requires a different software algorithm. In all software programs, primers are selected on the basis of criteria set by the investigator, the target specified, and with the goal that the oligonucleotide components not interact with one another during the PCR. Such interactions can lead to the formation of primer dimers during the PCR that increase the baseline fluorescent signal, called *noise*, of the assay. When designing a probe-based assay, the intervening probe sequence has to be taken into account as well. Only a few programs on the market are designed to find primer pairs and a compatible probe together as an assay set.

Table 2 lists the most prevalently used of the available commercial software and those that can be accessed and used over the Web.

Tab. 2 Real-time PCR assay design software.

Primer-based assays[a]		
Commercial software	*Company*	*Web site*
MacVector or DS Gene	Accelrys	http://www.accelrys.com/
Vector NTI	InforMax/Invitrogen	http://www.invitrogen.com/
Oligo	Mol. Biol. Insights	http://www.oligo.net/
DNAStar	DNASTAR, Inc.	www.dnastar.com
Probe-based Assays		
Commercial Software	*Company*	*Web site*
Primer Express	ABI	http://www.appliedbiosystems.com/
Beacon Designer	Premier Biosoft	http://www.premierbiosoft.com/
Lightcycler Software	Roche	http://www.roche-applied-science.com/
Webware	*Company*	*Web site*
Primer3	MIT	
LUX Designer	Invitrogen	
PrimeQuest	IDT	http://scitools.idtdna.com/Primerquest/

[a] Not intended to be an all inclusive list.

5.2 Basic Rules for Real-time Assay Development

Guidelines for designing primer- or probe-based assays are outlined in Table 3. By following these guidelines, we have designed assays that have worked well experimentally for many hundreds of transcripts, including viral, prokaryotic, and eukaryotic gene sequences. What has been proven empirically is that the efficiency of an assay is determined solely by how well the primers work as a pair. The probe, if present, is a signaling molecule and adds greatly to assay specificity.

SYBR Green I dye-based and simple fluorescent-labeled primer assays follow the same rules of assay design as for non-fluorescent PCR reactions. Primer concentrations should be kept relatively low, 50 to 100 nM, to discourage nonspecific priming. In contrast, when using probe-based assays and Taqman® assays, in particular, higher primer concentrations are required for optimal performance, 300 to 400 nM.

It is best to keep the T_m (melting temperature) of both primers the same. This simplifies assay quality control (QC) as changes in annealing temperature will affect the annealing of both primers equally. One way to avoid nonspecific binding is by raising the annealing temperature slightly during the PCR. Conversely, when little to no PCR product is formed, lowering the temperature may lead to increased primer binding to the template. Keeping the percent G/C content similar for both primers has also been found to be best for optimal assay performance. Pairing a short primer with a high G/C content and a long primer with low G/C content will not result in as efficient an assay as when the %G/C content is more balanced. Primers work best if they have 2 to 3 G- or C-residues in the last five bases of the primer sequence at the 3′ end. The binding specificity of PCR primers resides in the last 5 to 6 bases at their 3′ end; the rest of the primer sequence determines the annealing temperature. Many investigators insist on putting a G or C residue on

Tab. 3 Rules for the design of real-time PCR assays.

Rules for primer design		
Topic	*Rule*	*Comment*
Primer concentrations	Set primer concentrations to 50–100 nM for SYBR green and simple fluorescent primer–based assays; 300–400 nM for Taqman probe–based assays	For SYBR green assays, a low primer concentration will discourage primer dimers and false priming; fluorescently labeled primers may require higher concentrations depending on signal strength
Primer T_m	The primer T_m values should be equal	Good for assay optimization; raising or lowering the annealing temperature will have an equal effect on both primers
Primer % G/C	%G/C of a primer pair should be within 25% of one another	Primers with very disparate %G/C values work satisfactorily but keeping the %G/C close, and thus primer lengths, works best empirically
Primer structure	Primers work best with 2–3 G/Cs in the last 5 bases	Too many Gs or Cs in the last 5 bases will cause false priming and low efficiencies; too few will not prime well
3′ Clamp	G/C clamp on 3′ end works well but is not necessary	Primers with 3′ G- or C-residues work well but primers that end in an A or T can work equally well; 2 or more 3′-Ts do not work well
G residues	No more than 3 Gs in a row for primers or probes; causes a bend in the oligo structure	Multiple G-residues cause the DNA structure to bend which is not conducive to good binding for a short oligo, not such an issue for a long DNA or RNA template
Amplicon T_m	Increase T_m of the amplicon from 85 to 95 °C for high G/C regions or organisms	Often, software will set the optimal amplicon T_m at 85 °C but higher G/C sequences will require a higher setting
Rules for probe design		
Topic	*Rule*	*Comment*
5′-G residue	Reporter dye should not be linked to a G-residue	Guanine will act as a FRET pair with many fluorescein dyes and quench the signal (e.g. FAM, TET, VIC, etc.)
Probe T_m	Probes should have an 8–10 °C higher T_m than the primers	Probe T_m should be 8–10 °C above the primer T_m to ensure they can bind before the primers and generate a signal by probe cleavage or unfolding
Probe length	Taqman probes should not be over 30 bases or will not quench properly	Empirical experiments have shown that cleavage probes over 30 bases in length do not quench well and thus give a poor signal to noise ratio. High A/T rich sequences lead to long probes, lowering the T_m to 55 or 50 °C allows reasonable probe length with standard reagents, assay will still work at 55° or 60 °C
More Cs than Gs	Probes should have more Cs than Gs, particularly if there is a large imbalance between the two	Multiple G-residues cause the DNA structure to bend which is not conducive to binding in a short oligonucleotide probes that have many more G- than C-residues should be made from the complementary sequence

the 3′ end of the primer. This is called "clamping" the primer. Although primers work well when "clamped," it is not an absolute requirement for maximal PCR efficiency. Long homogeneous runs of any base should be avoided, but, in particular, G-residues. Tandem runs of G-residues over three bases in length tend to cause a bend in the oligonucleotide structure and this deviation from linear structure will impair the ability of the primer (or probe) to anneal properly.

There are rules for the design of fluorescently labeled probes as well. Probes labeled on the 5′ end with a fluorescein dye (e.g. 6-FAM, TET, VIC, etc.) should not begin with a G-residue. Guanine has the ability to effectively diminish the fluorescein signal, as mentioned previously for LUX primers. It is important that probes have a 5 to 10 °C higher T_m than the primers to ensure they bind to the template prior to the primers. This is particularly important for Taqman® probes that require cleavage of the probe for signal production. Similarly, Taqman® probes should be no longer than 30 bases. Probes longer than 30 bases will not quench well prior to cleavage because of poor FRET between the reporter and quencher dyes and generate little increase in fluorescent signal following cleavage compared to shorter, more efficiently quenched probes. When making assays for organisms with a low G/C content (lower than 40%), keeping the probe shorter than 30 bases can be problematic. An easy solution to this problem is simply to lower the optimal T_m setting during assay design. Taqman® assays work very well utilizing 50° or 55 °C annealing temperatures instead of the more common 60 °C. Probe length is not an issue for molecular beacons because of the forced close proximity of the reporter and quencher dyes in the probe design, or dual probe assays designed for the Lightcycler, where again the fluorescent signal comes from the close proximity of the dyes following binding. In general, probes work better if they have more C- than G-residues, particularly if there is a large imbalance between the two bases. This is a general rule applicable to all oligonucleotides used in real-time PCR.

5.3 Key Factors in Assay Development

The most important part of assay design is selecting the correct sequence before you begin. It is critical to have confidence in the accuracy and specificity of the sequence being investigated. There are two key issues, first to know that the sequence in question is accurate and second that the assay will be specific for the intended target. Mismatches between the actual sequence and those in the primers or probe can have a negative effect on assay performance. Genes or transcripts with regions of highly conserved sequence among family members can be problematic. Splice variants, and pseudogenes are some of the main issues to keep in mind when selecting a particular region within the sequence for assay design. For sequenced genomes with information available in public databases, accuracy usually can be determined *in silico*. However, for other organisms, it is important that the target sequence be validated prior to assay design. In this case, it may be necessary to sequence the final PCR product during assay development to ensure the intended target is being amplified.

When possible, it can be very useful to download sequences from other species for the same transcript or gene. When a human assay is the goal, for example, retrieving the human, mouse, and rat

or other mammalian sequences can yield helpful and sometimes unexpected information. An alignment of all species will sometimes illuminate conserved regions that may be useful in making a multispecies assay or point out a region to be avoided if a transgenic mouse with a human gene is the target. Multiple species alignments also may point out splice variants present in the rat or mouse sequence in this example, not known in the human transcript. In these instances, avoid exons involved in any known splice variants. It may be the same splice variant that exists in the human as well, but has not yet been observed and/or reported.

Another issue is making a transcript-specific assay for one family member of a multigene family when all members share a large amount of sequence homology. Again, alignments will allow you to find unique places to place an assay. As mentioned above, making mouse- and human-specific assays for the same transcript in a transgenic environment can be a challenge. In this case, it is critical there be no cross-detection of the transcripts from the two species. Should the sequence homology be highly conserved throughout the coding region, an assay can be placed in the 5′ or 3′ untranslated region (UTR). However, these regions of the transcript can contain a sequence that is not well balanced among the four bases and, thus, not very PCR friendly.

There are some philosophical issues to be decided prior to designing a real-time PCR assay. The first is whether to design assays across intron/exon junction regions of a transcript. The primary reason for spanning a junction region is to avoid the detection of contaminating DNA in the RNA preparation. The sequence surrounding junctions is not always compatible with efficient real-time PCR assay design. This is not a problem for primer-based assay with larger amplicons (i.e., 250 bases), but can be a problem for probe-based assays with shorter amplicons (i.e., 65–80 bases). Regardless of assay type, one thing to keep in mind is that there are a lot of pseudogenes in the mammalian genome. Pseudogenes are spliced versions of transcripts that have been introduced back into the DNA sequence, most likely by retroviruses. One estimate is the human genome contains 30 000 genes and 20 000 pseudogenes. Some genes have multiple pseudogenes and others have none. The problem is that it is impossible to predict which have them and which do not. Therefore, crossing a junction does not guarantee you will not detect contaminating DNA in your RNA sample. However, in practice, it does reduce background from DNA contamination for many transcripts. Philosophically, it is more important to have an efficient assay than to insist on crossing a junction. However, for assays that do not cross an intron/exon junction, this means that all the RNA samples have to be DNase I treated to ensure that any contaminating DNA does not contribute a significant signal. DNase I treatment is very straightforward, has little deleterious effect on the RNA, and takes only a short time. Three key factors to keep in mind when DNase I treating RNA samples: (1) keep the DNase I concentration low, a 1/10 dilution of the stock enzyme using a volume of diluted enzyme dependent on the RNA concentration; (2) DNase I is a divalent cation-requiring enzyme, 1-mM $MgCl_2$ is sufficient for the reaction; and (3) divalent cations and high heat will degrade RNA, keep the DNase I heat killing temperature and time low and brief; 75 °C and 10 min are sufficient.

Before an assay is put to use, it is important that it be tested for three important criteria: specificity, efficiency, and sensitivity. We call the empirical determination of these three assay characteristics, quality control (QC). Quality control procedures are slightly different for SYBR Green I and primer-based assays, compared to probe-based and scorpion primer assays. Determining assay specificity for primer-based assays involves running the first real-time PCR products on a DNA acrylamide gel to look closely for extra-assay bands and primer dimers. As mentioned earlier, primer dimers will increase the fluorescent background of the assay and reduce the signal to noise ratio. The latter can have a deleterious effect on assay sensitivity. A melt-curve should be run at the end of these assays, as well, to make sure the product produced makes a single, sharp melt peak. This is also true for every run done with primer-based assays. However, a word of caution with regard to the DNA profiles generated during melt-curve analysis. This method is a quick analysis of the homogeneity of the PCR products generated in every well on the plate. However, it is not a very sensitive assay. PCR products that are not that far apart in size will merely make the melt peak slightly wider, you will not get two distinct peaks unless the bands are quite disparate in size. For this assay type, the amplicon should be 200 to 250 bases in length so that primer dimers, less than 30 mers, can be readily seen as a separately resolved band on the resulting graph. The determination of assay sensitivity and efficiency is the same as for probe-based assays and will be discussed in the following text.

Probe-based assays and scorpion primers have a higher level of template specificity. For this reason, all QC analyses can be done on the real-time machine without the need for an acrylamide gel. The reason is that extra bands or primer dimers, should they be made, will not be detected because the intervening probe oligonucleotide(s) or the tail of the scorpion primer will bind solely to the central region of the correct amplicon and that binding is essential for the generation of a fluorescent signal. Therefore, for these assay types, all components must work in concert to obtain a fluorescent signal. Further, unlike SYBR Green I or fluorescent primer–based assays, the shortest amplicon possible for probe-based assays is the most desirable. The shorter the amplicon, the easier it is to achieve maximal efficiency of the PCR.

Once assay specificity has been established, the overriding theme that should be observed during assay QC is the maximization of assay efficiency and sensitivity. A proven strategy for achieving this goal utilized in designing hundreds of Taqman® probe–based assays is to order multiple primers (two forward and two reverse) around a single probe. This same design principle of multiple primers should be utilized for primer-based assays as well to minimize possible extra-assay bands and primer dimers. Testing the primers in all possible combinations will empirically determine the best primer pair for that assay. The current design software is not sophisticated enough, at present, to ensure an ideal assay every time. The primer pair that results in the lowest number of cycles necessary for a detectable signal and the highest probe dequench at a fixed template concentration is the best primer pair. In some cases, the four primer combinations may not differ greatly, but one pair will fit the optimal criteria better than the other three. In other cases, a larger 3 by 4 primer matrix (12 unique combinations) resulted

in only two acceptable primer combinations and only one of those resulted in an optimal assay. There are other examples where moving a primer one base up or down the sequence had a dramatic effect on assay performance. In all of these examples, all the primers fit the design rules outlined previously. In our experience, only by empirical experimentation can the best primer pair for an assay be determined.

Once an optimal primer pair with high template specificity has been selected, the sensitivity and efficiency of the assay must be empirically determined. Both can be found from running a dilution series of an assay standard. A standard curve should be run during assay QC, even if one is not going to be used with unknown samples, to determine assay efficiency and sensitivity. From the slope of the assay standard dilution series, assay efficiency can be determined. Acceptable slopes are −3.2 to −3.5 (see Table 7). In the real world, assays having this range of slopes are seen often and will yield excellent results. The lowest point on the standard curve that is no longer linear with the higher dilutions (falls off the standard curve) is no longer quantifiable in that assay. This is just past the low-end sensitivity for the assay. It is critical that this point be determined during assay QC so that incorrect values are not reported for the assay. This is particularly important for assays that will not include a standard dilution series on every plate.

Multiplexing is having the components for more than one assay in a single reaction. What are the design criteria for a good multiplexed assay? An acceptable multiplexed set of assays should have the same efficiency and sensitivity for their respective target sequences together that they have when run in isolation. This can only be achieved when the primers or primers and probe from one assay are able to amplify their target sequence without interfering with the other assay components in the reaction and visa versa. This means that two assays that work perfectly in isolation may very well not work with the same efficiency when combined. Thus, multiplexed assays have to be designed together as a super set to ensure operational autonomy. However, currently, there is no software available to the end user to assist in this process. Thus, beyond checking for pair-wise interactions among the reactants using available software, the optimization of a multiplex assay is an empirical one.

6 Running a Real-time PCR Experiment

It would not be practical in this review to discuss every nuance of performing a real-time PCR experiment. However, there are some important points that should be taken into consideration before any experiment is initiated.

6.1 Practical Considerations

When working with RNA, standard laboratory procedures should be followed: gloves, virgin plasticware whenever possible (RNase-free), and use of reagents and enzymes that are RNase-free. The most important component is the water. It is worth the expense to purchase commercially available nuclease-free water. DEPC-treated water made in the laboratory is not good for real-time PCR use because residual DEPC (diethyl pyrocarbonate) is a PCR inhibitor. This residue is not present in commercially available nuclease-free water.

Any lab that practices PCR will experience contamination problems, primarily from amplification products from earlier amplification reactions. Problems can be minimized by following a few, easy precautions. First, what is amplified in the tubes or plate wells should stay in the tubes or plate wells. If you have to use PCR products for a gel, for example, open the tube or plate and run the gel in a different room than the one you use to set up your RT and/or PCR reactions. Secondly, use a different set of pipettes in that room. It is best to make small aliquots of water in disposable labware for daily use and dispose of these at the end of the day. Use a different aliquot of water in each room. In time, everything that is exposed to concentrated DNA template for an assay will become contaminated by DNA-containing aerosols that are generated during pipetting. For amplicons of 70 to 250 bases, real-time PCR has the sensitivity to accurately detect template molecules in the high attogram (10^{-18} grams) range. Wipe the barrels of your pipetters down with 70% ethanol regularly to help eliminate this problem.

In an attempt to circumvent the contamination problem, many commercial kits substitute dUTP for dTTP and come with the enzyme uracil N-glycosylase (from the *Escherichia coli ung* gene). This enzyme will remove dU from any DNA molecule that contains it, degrading the template. Use of dUTP and uracil *N*-glycosylase prior to beginning the PCR is advertised by many companies as a way of eliminating potential contamination by amplified DNA. There are two problems with this approach. First, dUTP has to be used at twice the concentration of dTTP because it is not incorporated as efficiently by *Taq*. Secondly, other potentially contaminating templates in the laboratory (e.g. cDNAs made without dU, DNA standards and genomic DNA) will not contain dU and those will cause contamination anyway. In the end, using this strategy costs a lot more, increases the time of every instrument run, and adds nothing to the amplification of the final PCR product.

In practice, amplification of a target starting from an RNA substrate is more difficult than from DNA. Most failed RT-PCRs are due to a failed RT reaction. There are several reasons why the reverse transcriptase step does not succeed: (1) a very low abundance of the target transcript; (2) nonspecific binding of the primers and thus inefficient production of the correct cDNA template; (3) the inability to make enough complete templates of a long cDNA for the ensuing PCR; (4) a badly degraded RNA template or; (5) the presence of reverse transcriptase inhibitors in the RNA preparation. In contrast, the PCR is a comparatively robust reaction primarily because the DNA template is already present. Most failed PCRs stem from a bad component such as degraded dNTPs, improper or degraded primers, not enough $MgCl_2$, PCR inhibitors in the sample, and most often degraded or incorrect template.

6.2 Reagents

Most manufacturers make two kinds of reagent kits for real-time RT-PCR, a one-step and a two-step kit. In a two-step procedure, cDNA is made in one tube under ideal reaction conditions utilizing a thermocycler or water bath at the optimal temperature followed by the use of some or all of that cDNA for the succeeding PCR, also done under ideal reaction conditions, in a thermocycler or a real-time PCR machine. In a one-step protocol, RNA

template is added to a combined reagent mix containing both reverse transcriptase and a hot-start *Taq*. The RT and PCR reactions are performed sequentially in a real-time instrument. Hot-start *Taq* has made one-step protocols possible as the RT reaction is run at a lower temperature than that required for activation of the hot-start *Taq*. This method saves time and requires fewer manipulations. However, it has been shown empirically that one-step procedures are not as sensitive to low RNA template concentrations as two-step protocols. This is most likely because the components of the reaction mixture are a compromise to accommodate both enzymatic reactions. In those instances where transcript levels could be quite low and/or highly variable from assay to assay, use of a two-step protocol gives the best template sensitivity. However, if the same assays are being run repeatedly utilizing similar samples that have a known template range, a time-saving one-step procedure is more efficient.

Purchasing assay reagents from a company saves the time required to put the individual reagents together and provides a consistent reagent base for laboratory operations. When an investigator is new to real-time PCR, it is best to purchase a reagent kit to eliminate that variable should other problems arise. However, there is a financial cost to be paid for this convenience. Over time, the cost savings of constructing your own reagent mixes can be substantial. Further, knowing exactly what components are in each reaction mixture and their laboratory history makes tracking down reagent problems much easier.

The most important concept in making your own assay buffers is that of the master mix. A master mix is a large volume of all the reaction components common to each reaction, that is, for PCR that would be everything but the template, for a single assay. This will ensure that every reaction is as close to the same as possible, except for the one variable being measured.

An example of a master mix for a reverse transcriptase reaction using an assay-specific reverse primer and the reaction conditions are shown in Table 4. The RT master mix is for a single assay and is designed to easily continue on to the following real-time PCR. For wells where a no amplification control (no reverse transcriptase) reaction is required, some of this master mix can be added to a second tube just prior to the addition of the reverse transcriptase.

For SYBR Green I, or a fluorescent primer–based assay, it is important to use a hot-start *Taq* to minimize the number of nonspecific bands and primer dimers formed. An example of a master mix for the PCR we have used successively utilizing SYBR Green I can be seen in Table 5. For best results, we have found that a three-step PCR cycle will eliminate most primer dimmers during data collection at 72 °C.

Tab. 4 Example reverse transcriptase reaction.

Stock	*Component*	*Final*	*Volume [μL]*
*****	DEPC-H2O	*****	2.75
10X	RT Bfr	1X	1.00
2.5 mM	dNTP Mix[a]	500 μM	2.00
20 μM	Reverse Primer	100 nM	0.05
50 U/μL	Rtase	10 U/10 μL	0.20
			6.00

6 μL of RTMaster mix + 4 μL Sample
Thermocycler: 50 °C – 30 min; 72 °C – 5 min; 20 °C – soak

[a] Equal volumes 10 mM individual dNTPs = 2.5 mM each.

Tab. 5 Example SYBR green PCR reaction mix.

Stock	*Component*	*Final*	*Volume [µL]*
*****	PCR-H_2O[a]	*****	17.35
10X	PCR Buffer	1X	4.00
50X	ROX[b]	1X	1.00
40%	Glycerol[c]	8%	10.00
10%	Tween 20[c]	0.10%	0.50
50 mM	$MgCl_2$	5 mM	4.00
2.5 mM	dNTP Mix	200 µM	2.00
20 µM	Forward Primer (+)	100 nM	0.25
20 µM	Reverse Primer (−)	100 nM	0.20
100X	SYBR Green I[d]	1/40 000	0.50
5 U/µL	Hot-start *Taq*[e]	1.25 U	0.20
			40.00

PCR cycle:
95 °C – 1–10 min; 40 cycles: 95 °C – 15 s, 60 °C – 2 s, 72 °C – 30 s

[a]Any nuclease-free water.
[b]Invitrogen, Carlsbad, CA.
[c]High quality, nuclease-free.
[d]Molecular Probes (Invitrogen), Eugene, OR.
[e]Hot-start *Taq* required.

This buffer and cycling protocol would most likely work well for fluorescently labeled primers as well. For probe-based assays, a simpler PCR master mix is shown in Table 6. Note that with probe-based assays, a faster two-step PCR cycle can be used. In both cases, the master mix is designed to be added directly to the 10-µL post-RT cDNA reaction and inserted directly into a real-time PCR instrument for analysis.

6.3 Controls

One thing that is often not stressed enough is the importance of running the proper controls in a real-time PCR experiment. There are two essential controls to be considered when running a real-time PCR experiment. The first is a no-template control, sometimes called the "NTC". It is critical to know if there is low-level contamination present in the reaction mixture, especially if the ddCt method is being used (Sect. 7.4). One might argue that using commercial master mixes would make this control superfluous. However, the most likely source of contamination comes from the reagents added by the investigator. A second important control is for the reverse transcriptase reaction to check for signal from DNA contamination. This is done by running a reaction without reverse transcriptase, also known as a "no amplification control" or "NAC." As stated above, designing assays across intron/exon junctions is no guarantee that there will be no signal from DNA contamination. Running this control on junction-spanning assays once during QC using 20 to 50 ng of genomic DNA, much higher than any

Tab. 6 Example probe-based PCR reaction mix.

Stock	*Component*	*Final*	*Volume [µL]*
*****	PCR-H2O[a]	*****	23.00
10X	PCR Buffer	1X	5.00
100X	ROX[b]	1X	0.50
50 mM	MgCl2	5 mM	5.00
2.5 mM	dNTP Mix	200 µM	4.00
20 µM	Forward Primer(+)	400 nM	1.00
20 µM	Reverse Primer(−)	400 nM	1.00
20 µM	Probe(+)FAM	100 nM	0.25
5 U/µL	*Taq*[c]	1.25 U	0.25
			40.00

PCR cycle:
95 °C – 1–10 min; 40 cycles: 95 °C – 12 s, 60 °C – 30 s

[a]Any nuclease-free water.
[b]Invitrogen, Carlsbad, CA.
[c]Hot-start *Taq* unless assay-specific primers used for cDNA synthesis.

contamination level, will determine the sensitivity of the assay to extraneous DNA. For assays that do not cross a junction, we run one NAC for every sample on each 96-well assay plate. Even though each sample has been treated with DNase I, some assays are much more sensitive to DNA contamination than others, and uniform elimination of DNA contamination is not guaranteed. This control is particularly important for low-level transcripts where it will take 25 to 30 cycles for template detection. Control information for each assay should be included in all manuscripts that report data from real-time PCR experiments.

7 Data Analysis

In reality, it is nearly impossible for the person analyzing real-time PCR data to change the relationship of the samples with one another. That is because of the way the fluorescent signals are detected and the data recorded in the software. Having said that, data analysis can be performed incorrectly and have an impact on the final values obtained for the samples.

As with all quantitative techniques, it is important to compensate for differences in nucleic acid concentration from sample to sample. This process is called *sample normalization*. The most common method of normalizing sample values is to divide those values by values from an invariant transcript or gene, depending on whether the samples are RNA or DNA. If both the assay of interest and the one used to normalize the data have been incorrectly analyzed, the resulting ratio can be far from the true value, which can lead to an incorrect interpretation of the final data. Assay normalization will be discussed in Sect. 7.2.

7.1 Baseline Boundaries and Threshold Levels

The two most important settings for data analysis are the range of cycles that will define the baseline and the threshold. The baseline range represents a number of early cycles where only background signal can be detected. Thus, the signal is steady within this set of cycles. The rule of thumb for setting the upper end of this range is two cycles below the cycle where the highest signal on the plate or in a set of tubes is detected. Some software programs calculate the values for each individual sample on the basis of the baseline setting of the whole sample set. Others base their calculations on each individual well or tube. The low-end setting of the background defaults to cycle 3 in most software programs. This is because the first two cycles can have slightly higher overall signals as the PCR begins. This phenomenon is strictly related to the quenching efficiency of the fluorescent signaling mechanism for all, but SYBR Green I assays.

The cycle threshold (Ct) or crossing point (Cp) is the number of cycles it takes the fluorescent signal intensity from the sample to reach an arbitrary set value. In practical terms, the threshold or crossing point is represented by a horizontal line set across the rising amplification curves. The threshold setting is at a fluorescence intensity level significantly above the background signal. The minimum threshold is usually set at 10 times the standard deviation of the mean of the fluorescent signal(s) within the background range. The threshold is either preset in the software to a fixed value or determined by an algorithm in the software. In either case, it can be changed by the investigator. However, it is important that the threshold stay within the geometric part of the amplification curves of all samples being analyzed. The Ct or Cp is calculated from the fluorescence intensity values measured by the real-time instrument for all the samples in the experiment. Figure 10 shows a magnified view of the low end of the fluorescence scale from a real-time experiment to emphasize the baseline and threshold settings in relation to the amplification curves over a 6-log standard dilution series. In this case, the baseline is set from 3 to10 as the highest signal on the plate has a Ct of just over 12 cycles. The threshold is set at a fluorescent intensity (deltaRn) of 0.1, which is well within the geometric phase of this assay. This data was collected on an ABI 7700. Different software will have different relative scales for the *y*-axis. Thus, these absolute values will not be applicable to all real-time machines. However, the principles of how to set the baseline and threshold will be the same for all real-time PCR instruments.

7.2 Fluorescent Signal and Sample Normalization

The fluorescent signal values from a real-time PCR run can be reported in up to three different ways. The first is the raw fluorescence, which will have a scale commensurate with the sensitivity of the detector in the machine and the readout of the software. This is a measure of the raw signal coming from each well or tube and is often labeled R. The second value is called the deltaR (ΔR), which is calculated by subtracting the background signal in the well or tube from the raw fluorescence R and equals the net positive signal in the well or tube. This normalizes the signal for each individual

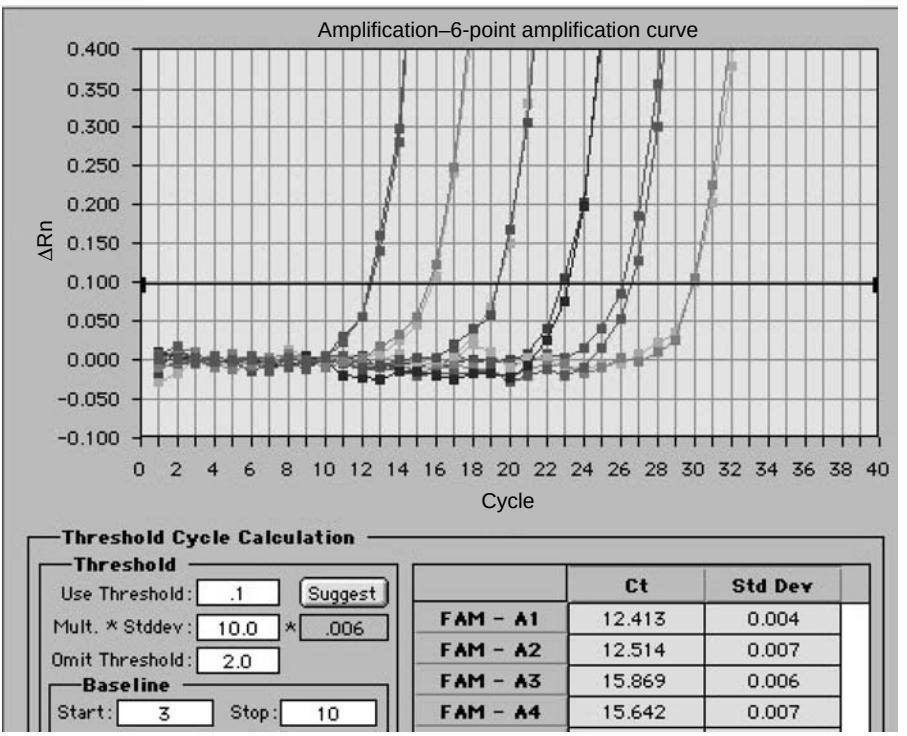

Fig. 10 Magnified view of the fluorescent signal in the baseline phase in relation to the threshold setting. A demonstration of setting the threshold high enough to be above the fluorescent signals (noise) in the baseline, but low enough to be within the geometric phase for each amplification curve. The threshold is the black horizontal line at 0.1ΔRn (see color plate p. xxxvi).

well or tube for background differences. In machines that use ROX as an internal reference dye within the master mix, the intensity can be reported in a third way, as the deltaRn (ΔRn), which equals the ΔR/ROX signal. The ROX signal detected by the real-time instrument in each well or tube should be equivalent in theory, but not always in practice. This signal will vary with the mechanical light path and sensitivity of the instrument. By dividing by the ROX signal, the net fluorescence from each sample is normalized for any differences in signal intensity from well to well. The deltaRn calculation is the first normalization process in data analysis.

As mentioned earlier, the way to level the playing field for all samples in an experiment is to divide the primary data by some value that is a measure of some invariant component within the sample. The ratio between the invariant gene or transcript and the gene or transcript of interest corrects for differences in the amount of DNA or RNA added into the assay. For DNA experiments, the genes selected for normalization are usually present in a single copy per genome such as ApoB100 for mammalian samples. For transcript normalization, however, this can be more problematic. The critical criterion for using a transcript for data normalization is that it be at the same level

in all the samples (e.g. transcripts/cell), regardless of the experimental procedure. For many experiments, transcripts such as β-actin remain constant over the course of an experiment. However, in some experiments, the treatment protocol for the experimental samples, such as the addition of drugs or changes in the physical conditions compared to controls, can cause a change in the relative level of housekeeping genes. In this case, housekeeping genes have been observed to change in tandem with the transcripts of interest. Dividing sample values by those from housekeeping genes that mirror the changes seen in the target transcripts will result in a ratio showing no change with treatment. It is critical, therefore, to make sure the transcript being used is biologically invariant throughout the experiment for all samples.

The question of what transcript to use for data normalization is one with no stock answer. Every experiment has to be monitored for changes in the transcript chosen to normalize the data. A paper looking at different concentrations of housekeeping genes commonly used to normalize real-time PCR data from many tissue types found that the amount of each housekeeping transcript was variable from tissue to tissue and that no single transcript was optimal for all tissues investigated. They concluded that all transcripts can change slightly with treatment so the best way to combat this effect was to use a minimum of three housekeeping genes for data normalization. Although this was a very a good study, the economics and time constraints of this approach have kept it from being widely adopted.

A way to skirt the problem is to use 18SrRNA or 28SrRNA. These are the two most abundant RNA species in any eukaryotic cell and are usually stable to most experimental perturbations. Some would argue that using RNA polymerase III transcripts (ribosomal RNA) to normalize RNA polymerase II transcripts (messenger RNA) is not appropriate because they are at much different levels in the cell and under different regulatory pressures within the cell. For most experimental protocols, however, the two ribosomal RNAs work well for data normalization. However, in certain experiments, it has been shown that even the ribosomal transcripts can be affected by the experimental conditions.

There are other instances where any transcript would not be appropriate for data normalization. They include studies involving apoptosis, measuring DNA or RNA in the serum, actinomycin D studies of mRNA half-life, and instances where the experimental sample has more cell types than the control, for example, sites of inflammation. In these cases, the best alternative is to measure total RNA or DNA using a fluorescent assay such as the ribogreen or picogreen assays respectively from Molecular Probes. Measuring the total nucleic acid present in a sample in this manner is far superior to estimating their concentration using absorbance at 260 nM because absorbance will detect free nucleotides, very short oligonucleotides, contaminating DNA or RNA, and impurities that absorb light at this wavelength. None of these would be detected in a real-time PCR assay, but would have an inconsistent effect on the value for each sample and, thus, on their final normalized values.

7.3 Quantification Using a Standard Curve

There are experiments where it is not necessary to quantify the results. In these

cases, a determination of the presence or absence of the target sequence is sufficient. Examples are the presence or absence of a translocation in a tumor population or a pathogen in a tissue. However, the majority of studies require an accurate measure of the number of target molecules detected.

One method for accomplishing this goal is to run a dilution series of a quantified assay template alongside the unknown samples. The kinds of standards that can be used for this purpose were discussed in Sect. 4. In practice, a 10-fold dilution series spanning a 5-log concentration range of standard template is commonly used for quantification of unknown samples. However, during QC, a 6- to 7-log curve is used to determine the efficiency and lowest endpoint (sensitivity) of the assay. An example of a 6-log standard curve is shown in Fig. 11. In this example, the lowest dilution falls on the linear line and thus defines the low end of this assay, 200 molecules, because the last dilution, 7 logs, did not fall on the standard curve (not shown). The lowest detectable level for any assay, and thus its sensitivity, is the last dilution that is still on the standard curve. Unknown samples with Ct values below this level cannot be accurately quantified by the assay. It is extremely important to determine the low-end value for each assay in use. Otherwise, values may be reported that are not within the linear range of the assay and will not represent the true measure of the target. This is because the software is written with the assumption that the Ct or Cp value for the sample falls on the standard curve. If this assumption is not true and the Ct is off the curve, the value assigned to that sample by the software will be incorrect because the Ct value will be forced to fit the curve and an overestimate will result.

The standard curve has a negative slope because the more molecules there are, the fewer PCR cycles it takes for the signal in the reaction to reach the threshold (Ct). In Fig. 11, there are three values associated with the quality of the standard curve, which is a refection of the quality of the assay. The first is the slope from which you can calculate the PCR efficiency of the reaction. The slope (rise over run) is a measure of how many PCR cycles it takes to amplify the template 10-fold. The efficiency is derived from the amplification calculation. The relationship of assay slope to PCR efficiency and template amplification per cycle can be seen in Table 7. From Table 7, a slope of -3.33 equals 100% efficiency and an ideal amplification of template per cycle of 1 to 2. Therefore, the difference between each point in the standard curve is 3.33 cycles (y-axis) for every 10-fold dilution of standard template (Fig. 11). The second value is the y-intercept, which is the theoretical number of cycles required to generate a measurable signal (Ct) from a single template molecule. It is theoretical because most real-time quantitative PCR assays are not sensitive enough to maintain linearity down to one molecule. For RT-PCR, linearity to the low hundreds of molecules is very good due to the inhibition of *Taq* DNA polymerase by reverse transcriptase. The practical utility of this number is it tells you when your standard has been calibrated properly. Values from 36 to 38 cycles are acceptable due to variations in the slope from assay to assay. The third number is r^2 or correlation coefficient, which tells you how well the values making up the standard curve fall on a straight line. For real-time PCR, this value should not be below 0.990 for a good assay and is generally higher than 0.995.

Tab. 7 PCR efficiency versus assay slope.

Assay slope	*PCR efficiency*	*Product amplification*
−3.60	89.6%	1.90
−3.55	91.3%	1.91
−0.60	93.1%	1.93
−3.45	94.9%	1.95
−3.40	96.8%	1.97
−3.35	98.8%	1.99
−3.33	100.0%	2.00
−3.30	101.0%	2.01
−0.12	103.1%	2.03
−3.20	105.4%	2.05
−3.15	107.7%	2.08
−3.10	110.2%	2.10

Exponential amplification = 10E(−1/slope).
Efficiency = [10E(−1/slope)]−1.

Molecule numbers are determined for each unknown sample by interpolation from the standard curve. Each unknown is measured in duplicate or triplicate and the mean of the measurements made by the real-time instrument, the Ct values, are averaged in software. Using the mean Ct, the numbers of molecules are determined for each unknown. Following the red lines in the example seen in Fig. 11, a Ct of 29.5 would equate to 200 molecules of target sequence in the sample, relative to this standard preparation. The software determines a formula for the standard curve using a perturbation of $y = mx + b$, where $m =$ slope and b is the y-intercept. The investigator defines how many molecules are assigned to each dilution of standard on the basis of the best measurement possible. The number of molecules, relative to the standard being used, can be calculated for a defined mass using the length of the standard and taking into account whether it is single or double-stranded. The software then uses the formula for the standard curve to calculate the number of molecules for every unknown sample based on their mean Ct values.

The advantages of using a standard curve are as follows: (1) you have a direct measure of the assay quality for each plate or group of samples and (2) data can be compared from plate to plate with confidence. Disadvantages are as follows: (1) care must be taken to ensure the standard curves for each assay are reproducibly made from run to run and (2) standard curves take up wells on the plate that could be used for unknown samples.

There are many examples where using a standard curve to determine a value for unknown samples is preferred. One would be in screening patients for a pathogen. Knowing the pathogen is present early in an infection would be very useful. Equally useful would be to know how many pathogens are present. This added knowledge would be very helpful in determining how aggressively the patient should be treated. Another example would be a large study with many animals or tissue culture plates involving controls, multiple test agents, and multiple transcript assays. Having a quantitative value for each transcript provides great flexibility in how data will be analyzed utilizing a variety of statistical methods.

7.4 Quantification Using the ddCt Method

Not all studies require that a specific value, interpolated from a standard curve, be determined for each sample. Rather, the goal of the experiment is to simply calculate a fold difference between one sample against the remaining samples on the plate or a group of similar control samples compared to one or more

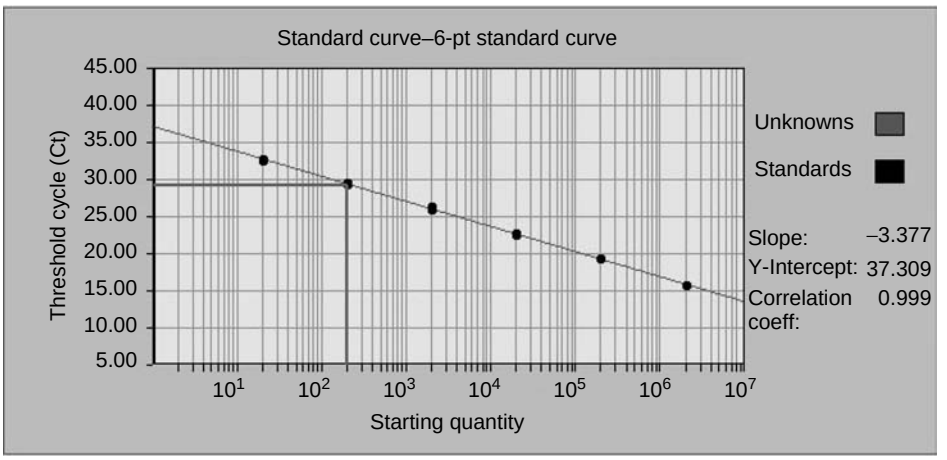

Fig. 11 A 6-log standard curve with a graphic demonstration of how the software determines molecules from the Ct values measured by real-time instruments. A Ct (threshold cycle) value of approximately 29.5 (*y*-axis) would result in an interpolated value of 200 molecules (*x*-axis) using this standard curve. A change in the slope of the curve or the measured Ct values of the standards would change the number of molecules obtained from the same Ct reading (see color plate p. xxxvi).

experimental groups on the plate. For these studies, there is a way to calculate a fold difference directly from the Ct values obtained from the real-time PCR data. A second assay, using an invariant housekeeping transcript or one of the ribosomal transcripts, must be performed to normalize the data for differences in sample loading. Armed with the Ct values from the transcript of interest and one used for normalization, a method called the delta delta Ct determination or ddCt can be employed.

As the name implies, the deltas involve subtraction steps. The first is to subtract the Ct of the assay used for data normalization from the Ct of the assay of interest. This now gives a normalized Ct value for each sample on the plate as follows:

Housekeeping gene assay
Sample 1 – Ct = 21.5
Sample 2 – Ct = 20.9
Sample 3 – Ct = 21.2

Gene of interest
Sample 1 – Ct = 25.2
Sample 2 – Ct = 22.7
Sample 3 – Ct = 21.4

Therefore, the deltaCt, normalized for loading for each sample is as follows:

Sample 1 = 3.7
Sample 2 = 1.8
Sample 3 = 0.2

Let us assume that, in this experiment, we want to know the fold difference of Samples 2 and 3 compared to Sample 1. We then have to subtract again, for the second delta:

$$\text{deltaCt S2} - \text{deltaCt S1} = 1.8 - 3.7 = -1.9$$

$$\text{deltaCt S3} - \text{deltaCt S1} = 0.2 - 3.7 = -3.7$$

Relative fold difference can be determined by the formula, $2^{-\mathrm{ddCt}}$

$$\text{Sample 2 relative to Sample 1} = 2^{-(-1.9)} = 3.73 \text{ fold}$$

$$\text{Sample 3 relative to Sample 1} = 2^{-(-3.7)} = 13.00 \text{ fold}$$

Advantages of using the ddCt method are as follows: (1) you do not have to run a standard curve on each plate and (2) the calculation is easy to perform. Disadvantages are as follows: (1) there is no direct evidence each plate has run as expected (slope); (2) comparing data from plate to plate must be controlled by a common sample(s); and (3) final data can only be expressed as a fold difference.

The ddCt method works very well for large studies where there are many samples being compared to a common control. One example of this kind of assay could be the screening of a small molecule library where assaying a maximal number of compounds in each assay set is critical for throughput. The purpose of such a screen could be to find the most active compounds affecting a target transcript from a biological pathway. Comparing this data with another screen utilizing another target transcript from a second pathway might aid in the selection of a compound that has selective activity; high in one assay and low in the second. Those that fit the criteria of the screen would then be moved on to other assays to confirm and elaborate the findings of the real-time PCR screen.

8 Future Directions

Since the inception of real-time PCR by ABI in 1996 with the 7700, the footprint of newer instruments has been steadily getting smaller, the cost has fallen from about $100 000 per machine to about one-quarter of that price, and, in some ways, the newer machines are better than the original. Not only can you get high quality data for less money but the software is also more sophisticated in assisting the user in analyzing and presenting the data. The result of the smaller sizes and lower cost points is that these instruments are migrating from the domain of the core laboratory into individual research laboratories at an increasing rate. The only problem with this trend is that the companies selling real-time PCR instrumentation are not keeping up with training and/or the training is inadequate. The rapid increase in instruments in individual laboratories will put even more pressure on companies selling real-time PCR instruments. Hopefully, this review has provided enough of an introduction to the technique to at least make the reader aware of the major issues involved in performing real-time PCR correctly.

One thing the newer real-time instruments excel at is multiplexing. The hardware has evolved to the point that selective collection of fluorescent light over the entire green–red spectrum is now possible. There are multiple fluorescent dyes available within this broad range capable of being used as reporters covalently attached to fluorescently labeled primers or probes. However, assay design software currently on the market is only capable of designing assays that work in isolation. One way of keeping down costs and saving time would be to obtain information on the assay of interest as well as information for a second transcript useful for data normalization, design two compatible assays, and then run them together in the same, multiplexed reaction. The ability of the

investigator to design good multiplexed assays in the same way that single assays can be made, presently, has not kept up with the hardware. Assay design software and the commercial sector will have to catch up with the currently available instrumentation before multiplexing becomes a true advantage to the consumer. Fortunately, there are promising signs of new software for multiplexing on the horizon.

The real value of real-time PCR instruments is their ability to collect data over time. It is this temporal component that adds an invaluable dimension to the data set. It would seem logical then that these instruments could be used for more than studies utilizing nucleic acids or, at the least, more than just PCR. There may be a number of assays that are currently read at end point, that would be amenable to the detection capabilities of a real-time instrument, and could benefit from the addition of a temporal component. One challenge yet unmet by manufacturers of real-time instruments is to expand the use of these instruments to other facets of investigation beyond PCR.

See also Immuno-PCR; Oligonucleotides; RNA Methodologies.

Bibliography

Books and Reviews

Ausubel, F.M. (2001) *Current Protocols in Molecular Biology*, Wiley InterScience, New York.

Bustin, S.A. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays, *J. Mol. Endocrinol.* **25**, 169–193.

Bustin, S.A. (2004) *A-Z of Quantitative PCR*, International University Line, La Jolla, CA.

Chomczynski, P. (1993) A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples, *Bio Techniques* **15**, 532–534, 536–537.

Gibson, U.E., Heid, C.A., Williams, P.M. (1996) A novel method for real time quantitative RT-PCR, *Genome Res.* **6**, 995–1001.

Heid, C.A., Stevens, J., Livak, K.J., Williams, P.M. (1996) Real time quantitative PCR, *Genome Res.* **6**, 986–994.

Herdewijn, P. (2004) *Oligonucleotide Synthesis: Methods and Applications*, Humana Press, Totowa, NJ

Sambrook, J., Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Selvin, P.R. (1995) Fluorescence resonance energy transfer, *Methods Enzymol.* **246**, 300–334.

Tricarico, C., Pinzani, P., Bianchi, S., Paglierani, M., Distante, V., Pazzagli, M., Bustin, S.A., Orlando, C. (2002) Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies, *Anal. Biochem.* **309**, 293–300.

Primary Literature

Al-Robaiy, S., Rupf, S., Eschrich, K. (2001) Rapid competitive PCR using melting curve analysis for DNA quantification, *Bio Techniques* **31**, 1382–1386, 1388.

Barragan-Gonzalez, E., Lopez-Guerrero, J.A., Bolufer-Gilabert, P., Sanz-Alonso, M., De la Rubia-Comos, J., Sempere-Talens, A. (1997) The type of reverse transcriptase affects the sensitivity of some reverse transcription PCR methods, *Clin. Chim. Acta* **260**, 73–83.

Berney, C., Danuser, G. (2003) FRET or no FRET: a quantitative comparison, *Biophys. J.* **84**, 3992–4010.

Bustamante, L.Y., Crooke, A., Martinez, J., Diez, A., Bautista, J.M. (2004) Dual-function stem molecular beacons to assess mRNA expression in AT-rich transcripts of Plasmodium falciparum, *Bio Techniques* **36**, 488–492, 494.

Bustin, S.A. (2002) Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems, *J. Mol. Endocrinol.* **29**, 23–39.

Bustin, S.A., Nolan, T. (2004) Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction, *J Biomol. Tech.* **15**, 155–166.

Calogero, A., Hospers, G.A., Timmer-Bosscha, H., Koops, H.S., Mulder, N.H. (2000) Effect of specific or random c-DNA priming on sensitivity of tyrosinase nested RT-PCR: potential clinical relevance, *Anticancer Res.* **20**, 3545–3548.

Casabianca, A., Orlandi, C., Fraternale, A., Magnani, M. (2004) Development of a real-time PCR assay using SYBR Green I for provirus load quantification in a murine model of AIDS, *J Clin. Microbiol.* **42**, 4361–4364.

Chen, R., Huang, W., Lin, Z., Zhou, Z., Yu, H., Zhu, D. (2004) Development of a novel real-time RT-PCR assay with LUX primer for the detection of swine transmissible gastroenteritis virus, *J. Virol. Methods* **122**, 57–61.

Chomczynski, P., Mackey, K., Drews, R., Wilfinger, W. (1997) DNAzol: a reagent for the rapid isolation of genomic DNA, *Bio Techniques* **22**, 550–553.

Eleaume, H., Jabbouri, S. (2004) Comparison of two standardisation methods in real-time quantitative RT-PCR to follow Staphylococcus aureus genes expression during in vitro growth, *J. Microbiol. Methods* **59**, 363–370.

Fehlmann, C., Krapf, R., Solioz, M. (1993) Reverse transcriptase can block polymerase chain reaction, *Clin. Chem.* **39**, 368–369.

Gerard, G.F., Fox, D.K., Nathan, M., D'Alessio, J.M. (1997) Reverse transcriptase. The use of cloned Moloney murine leukemia virus reverse transcriptase to synthesize DNA from RNA, *Mol. Biotechnol.* **8**, 61–77.

Gilliland, G., Perrin, S., Blanchard, K., Bunn, H.F. (1990) Analysis of cytokine mRNA and DNA: detection and quantitation by competitive polymerase chain reaction, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2725–2729.

Gjerdrum, L.M., Sorensen, B.S., Kjeldsen, E., Sorensen, F.B., Nexo, E., Hamilton-Dutoit, S. (2004) Real-time quantitative PCR of microdissected paraffin-embedded breast carcinoma: an alternative method for HER-2/neu analysis, *J. Mol. Diagn.* **6**, 42–51.

Huang, Y., Kong, D., Yang, Y., Niu, R., Shen, H., Mi, H. (2004) Real-time quantitative assay of telomerase activity using the duplex scorpion primer, *Biotechnol. Lett.* **26**, 891–895.

Kellogg, D.E., Sninsky, J.J., Kwok, S. (1990) Quantitation of HIV-1 proviral DNA relative to cellular DNA by the polymerase chain reaction, *Anal. Biochem.* **189**, 202–208.

Maaroufi, Y., Ahariz, N., Husson, M., Crokaert, F. (2004) Comparison of different methods of isolation of DNA of commonly encountered Candida species and its quantitation by using a real-time PCR-based assay, *J. Clin. Microbiol.* **42**, 3159–3163.

Mallet, F., Oriol, G., Mary, C., Verrier, B., Mandrand, B. (1995) Continuous RT-PCR using AMV-RT and *Taq* DNA polymerase: characterization and comparison to uncoupled procedures, *Bio Techniques* **18**, 678–687.

Margraf, R.L., Page, S., Erali, M., Wittwer, C.T. (2004) Single-tube method for nucleic acid extraction, amplification, purification, and sequencing, *Clin. Chem.* **50**, 1755–1761.

Martins, T.B., Hillyard, D.R., Litwin, C.M., Taggart, E.W., Jaskowski, T.D., Hill, H.R. (2000) Evaluation of a PCR probe capture assay for the detection of Toxoplasma gondii. Incorporation of uracil N-glycosylase for contamination control, *Am. J. Clin. Pathol.* **113**, 714–721.

Mhlanga, M.M., Malmberg, L. (2001) Using molecular beacons to detect single nucleotide polymorphisms with real-time PCR, *Methods* **25**, 463–471.

Myers, T.W., Gelfand, D.H. (1991) Reverse transcription and DNA amplification by a Thermus thermophilus DNA polymerase, *Biochemistry* **30**, 7661–7666.

Parks, S.B., Popovich, B.W., Press, R.D. (2001) Real-time polymerase chain reaction with fluorescent hybridization probes for the detection of prevalent mutations causing common thrombophilic and iron overload phenotypes, *Am. J. Clin. Pathol.* **115**, 439–447.

Pawlowski, V., Revillion, F., Hornez, L., Peyrat, J.P. (2000) A real-time one-step reverse transcriptase-polymerase chain reaction method to quantify c-erbB-2 expression in human breast cancer, *Cancer. Detect. Prev.* **24**, 212–223.

Peters, I.R., Helps, C.R., Hall, E.J., Day, M.J. (2004) Real-time RT-PCR: considerations for efficient and sensitive assay design, *J. Immunol. Methods* **286**, 203–217.

Robert, C., McGraw, S., Massicotte, L., Pravetoni, M., Gandolfi, F., Sirard, M.A. (2002) Quantification of housekeeping transcript

levels during the development of bovine preimplantation embryos, *Biol. Reprod.* **67**, 1465–1472.

Rutledge, R.G., Cote, C. (2003) Mathematics of quantitative kinetic PCR and the application of standard curves, *Nucleic Acids Res.* **31**, e93.

Sears, J.F., Khan, A.S. (2003) Single-tube fluorescent product-enhanced reverse transcriptase assay with Ampliwax (STF-PERT) for retrovirus quantitation, *J. Virol. Methods* **108**, 139–142.

Sellner, L.N., Coelen, R.J., Mackenzie, J.S. (1992) Reverse transcriptase inhibits *Taq* polymerase activity, *Nucleic Acids Res.* **20**, 1487–1490.

Szabo, A., Perou, C.M., Karaca, M., Perreard, L., Quackenbush, J.F., Bernard, P.S. (2004) Statistical modeling for selecting housekeeper genes, *Genome Biol.* **5**, R59.

Telesnitsky, A., Goff, S.P. (1993) RNase H domain mutations affect the interaction between Moloney murine leukemia virus reverse transcriptase and its primer-template, *Proc. Natl. Acad. Sci. U. S. A.* **90**, 1276–1280.

Tosh, C., Hemadri, D., Sanyal, A., Pattnaik, B., Venkataramanan, R. (1997) One-tube and one-buffer system of RT-PCR amplification of 1D gene of foot-and-mouth disease virus field isolates, *Acta Virol.* **41**, 153–155.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, *Genome Biol.* **3**, RESEARCH0034.1–0034.11.

Vu, H.L., Troubetzkoy, S., Nguyen, H.H., Russell, M.W., Mestecky, J. (2000) A method for quantification of absolute amounts of nucleic acids by (RT)-PCR and a new mathematical model for data analysis, *Nucleic Acids Res.* **28**, E18.

Wang, C., Gao, D., Vaglenov, A., Kaltenboeck, B. (2004) One-step real-time duplex reverse transcription PCRs simultaneously quantify analyte and housekeeping gene mRNAs, *Bio Techniques* **36**, 508–516, 518–509.

Whelan, J.A., Russell, N.B., Whelan, M.A. (2003) A method for the absolute quantification of cDNA using real-time PCR, *J Immunol. Methods* **278**, 261–269.

Wilhelm, J., Pingoud, A., Hahn, M. (2003) Validation of an algorithm for automatic quantification of nucleic acid copy numbers by real-time polymerase chain reaction, *Anal. Biochem.* **317**, 218–225.

Wilkening, S., Bader, A. (2004) Quantitative real-time polymerase chain reaction: methodical analysis and mathematical model, *J. Biomol. Tech.* **15**, 107–111.

Yin, J.L., Shackel, N.A., Zekry, A., McGuinness, P.H., Richards, C., Putten, K.V., McCaughan, G.W., Eris, J.M., Bishop, G.A. (2001) Real-time reverse transcriptase-polymerase chain reaction (RT-PCR) for measurement of cytokine and growth factor mRNA expression with fluorogenic probes or SYBR Green I, *Immunol. Cell. Biol.* **79**, 213–221.

RecA Superfamily Proteins

Dharia A. McGrew and Kendall L. Knight
University of Massachusetts Medical School, Worcester, MA

1 RecA: Function in Homologous Recombination 528
1.1 RecA: Structure and History 528

2 ATP Binding and Hydrolysis 531
2.1 The Walker A motif 531
2.2 The Walker B motif 531
2.3 The ATP Cap 532
2.4 ATP Hydrolysis 533

3 DNA Binding Domains 534
3.1 Loop1 and Loop2 534
3.2 C-terminal Gateway 534
3.3 Eukaryotic and Archaeal Uncertainty 534
3.3.1 Helix-hairpin-helix 536

4 Oligomeric Interface Domains 536
4.1 N-terminal Polymerization Motif 536
4.2 Other Cross-subunit Contacts 537

5 Allosteric Mechanism 538
5.1 Allosteric Switch: Gln194 538
5.2 ATP Sensing: Phe217 538

6 Proteins that interact with RecA 540
6.1 RecBCD 540
6.2 RecF, O, and R 542
6.3 RecX and DinI 542

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

6.4 Coprotease Substrates 543
6.5 PolV 543
6.6 Proteins that Interact with Rad51 543

7 Future Directions 544

Acknowledgments 545

Bibliography 545
Books and Reviews 545
Primary Literature 545

Keywords

Allosteric Switch
A glutamine residue conserved in all RecA superfamily proteins that interacts with the γ-phosphate of bound ATP and triggers conformational changes that contribute to formation of the active form of the protein filament.

ATP Cap
A loop formed by 6 to 7 residues that extends into the ATP binding site of the neighboring subunit to make contacts that are likely involved in stabilizing the structure of the filament and perhaps transmitting ATP-induced allosteric information across the subunit interface.

C-terminal Gateway
A domain formed by the C-terminal region of RecA proteins involved in regulating access of double-stranded DNA into the core of active RecA/ssDNA/ATP filaments.

Helix-hairpin-helix motif
A domain typically involved in non-sequence-specific DNA binding found in the N-terminal region of RadA and Rad51 proteins that may regulate binding of double-stranded DNA in a manner similar to the C-terminal gateway in RecA proteins.

Holliday Junction
The X-shaped crossover structure resulting from the action of RecA-like proteins during the ''homologous pairing and strand exchange'' step in the HR pathway. This type of structure had been proposed as an intermediate in the HR pathway by Robin Holliday in 1964.

Homologous Recombination (HR)
The transfer of genetic information between two double-stranded DNA molecules of similar sequence that results from a physical exchange of DNA strands.

N-terminal Polymerization Motif
A short β-strand that forms a section of the oligomeric interface by interacting with another β-strand in the core region of the neighboring subunit. This motif is conserved in all RecA superfamily structures solved to date, as well as in the human BRCA2 protein where it likely serves to regulate assembly of Rad51 filaments at the site of a DNA break.

Rad51-mediator Activity
First described for the yeast Rad52 protein regarding its ability to assist Rad51 filament formation, this function now appears to have been conserved in all organisms in which RecA/Rad51-mediated recombination occurs. Examples include RecFOR/RecA and BRCA2/human Rad51 interactions.

Translesion Bypass
DNA replication catalyzed by an error-prone polymerase across a damaged region of template DNA. Specific protein–protein interactions between RecA and the polymerase at the site of damage are important to optimal polymerase function.

The bacterial RecA protein is the central component in both the catalysis of homologous recombinational DNA repair and the regulation of expression of a family of DNA repair genes. It functions in a number of mechanistically distinct processes, all of which contribute to the maintenance of genomic integrity, and all of which require formation of an active RecA/ATP/DNA complex. As cellular, biochemical, and structural studies of the eukaryotic and archaeal RecA homologs progress (Rad51 and RadA, respectively), it is clear that despite fundamental similarities in the mechanism of DNA strand exchange catalyzed by both proteins, there are distinct differences in their catalytic mechanisms as well as dramatic differences in the cellular pathways in which each protein plays important roles. For example, both RecA and the human Rad51 protein require ATP binding and hydrolysis for the catalysis of DNA strand exchange, but the use of ATP is distinctly different for each protein. Additionally, whereas RecA plays a specific role in transcriptional regulation of DNA repair genes, no such activity is seen for Rad51. In this review, we offer a comparative view of the biochemical functions of RecA and Rad51, with a particular focus on descriptions of our current understanding of the molecular design, catalytic organization, and allosteric regulation of enzyme function.

1
RecA: Function in Homologous Recombination

The bacterial RecA protein participates in a remarkably diverse set of biological functions, all of which contribute to the maintenance of genomic integrity during the lifetime of the organism as well as to ensure that genetic information is transmitted to subsequent generations with high fidelity. Bacterial RecA is directly responsible for (1) regulation of expression of a set of DNA repair genes referred to as the SOS regulon; (2) catalysis of DNA strand exchange; and (3) promoting bypass of DNA lesions by a specialized DNA polymerase during a cell's recovery from DNA damage. There are RecA-like proteins in all free living organisms, and, in all cases, their recombination function plays an important role in the repair of damaged DNA, whether the damage results from inherent errors in the machinery responsible for replicating DNA, or from exposure to exogenous mutagens. In fact, in their search for mutants that showed defects in homologous recombination (HR), Clark and Margulies discovered in 1965 that defects in *recA* greatly increased the sensitivity of bacteria to the DNA-damaging effects of UV light, thus establishing the link between homologous recombination and DNA repair. The primary enzymatic function of RecA-like proteins is the catalysis of HR. This involves exchange of genetic information between DNA molecules of similar sequence, a process resulting from the physical transfer of DNA strands. The HR pathway is traditionally broken down into four steps as seen in Fig. 1, and RecA-like proteins specifically catalyze the "homologous pairing and strand exchange" step, thus creating crossover structures known as *Holliday junctions* (named after Robin Holliday). In the scheme depicted in Fig. 1, the "initiation" step involves processing double-strand DNA breaks by exonucleases. In bacteria, this is accomplished by the RecBCD enzyme, but the mammalian counterpart(s) have not yet been identified. This leaves an ssDNA tail that serves as substrate for RecA proteins. "Pairing and strand exchange" results when the RecA/ssDNA complex invades a homologous duplex DNA, for example, a sister chromatid. The RecFOR complex is indicated at this step, but current work indicates that this complex regulates assembly and disassembly of the RecA filament specifically during repair of DNA gaps, as described below. The RecA protein itself can perform "branch migration," but, *in vivo*, this process is carried by specialized motor proteins, for example, the bacterial RuvAB helicase complex. Again, although there are a number of candidate helicases in mammalian cells, the one(s) responsible for branch migration of Holliday crossovers have yet to be identified. "Resolution" is achieved by the action of structure-specific endonucleases ("resolvases"), and, in bacteria, this function is performed by the RuvC protein. In mammalian cells, the Rad51C/Xrcc3 protein complex has been shown to be associated with Holliday junction resolvase activity, but it is currently unclear if one of these proteins or an associated nuclease is responsible for junction cleavage.

1.1
RecA: Structure and History

The first X-ray crystal structure of the *Escherichia coli* RecA protein was solved in 1992 by Story et al. In this work, the structures solved were that of protein alone and the protein–ADP complex. In both cases, RecA formed a helical filament

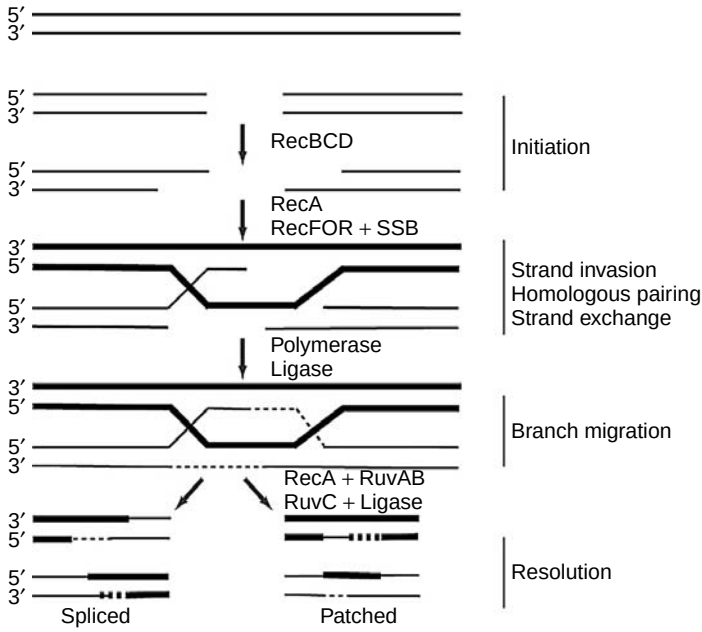

Fig. 1 Steps in the homologous recombination pathway. A basic model for homologous recombinational repair of DNA double-strand breaks (DSBs) in *E. coli* that can be separated into four major steps and are described in the text. The function of RecA in "homologous pairing and strand exchange" is the primary focus of this review.

with six subunits per turn and a distance between turns (pitch) of approximately 83 Å. Previous electron microscopic investigations had clearly established that the RecA filament could exist in two distinct oligomeric states, an "inactive" and "active" filament. The inactive filament is formed by RecA in the absence of nucleotide or by a RecA/DNA/ADP complex and has 6.2 subunits per turn with a pitch of approximately 68 Å. The active RecA filament forms in the presence of DNA and ATP, also has 6.2 subunits per turn but has an extended pitch between 95 and 100 Å. Because the RecA crystal structure contained ADP and has a pitch intermediate between the inactive and active filament, it was not clear what aspects reflected the structure of an active RecA filament. Since that time, many groups have attempted to solve the structure of a catalytically active RecA/DNA/ATP complex, but with no success. Therefore, the original RecA structure has served for more than a decade as a model for the design of experiments aimed at characterizing the molecular mechanism of HR. During this time, and now with the availability of new structures of archaeal RadA and eukaryotic Rad51 proteins, a more complete picture is emerging regarding the structural dynamics of RecA-like filaments, and the relationship of filament structure to its catalytic function.

This review focuses on our current understanding of the structural design of the bacterial RecA protein, the functional coordination of its various domains, and

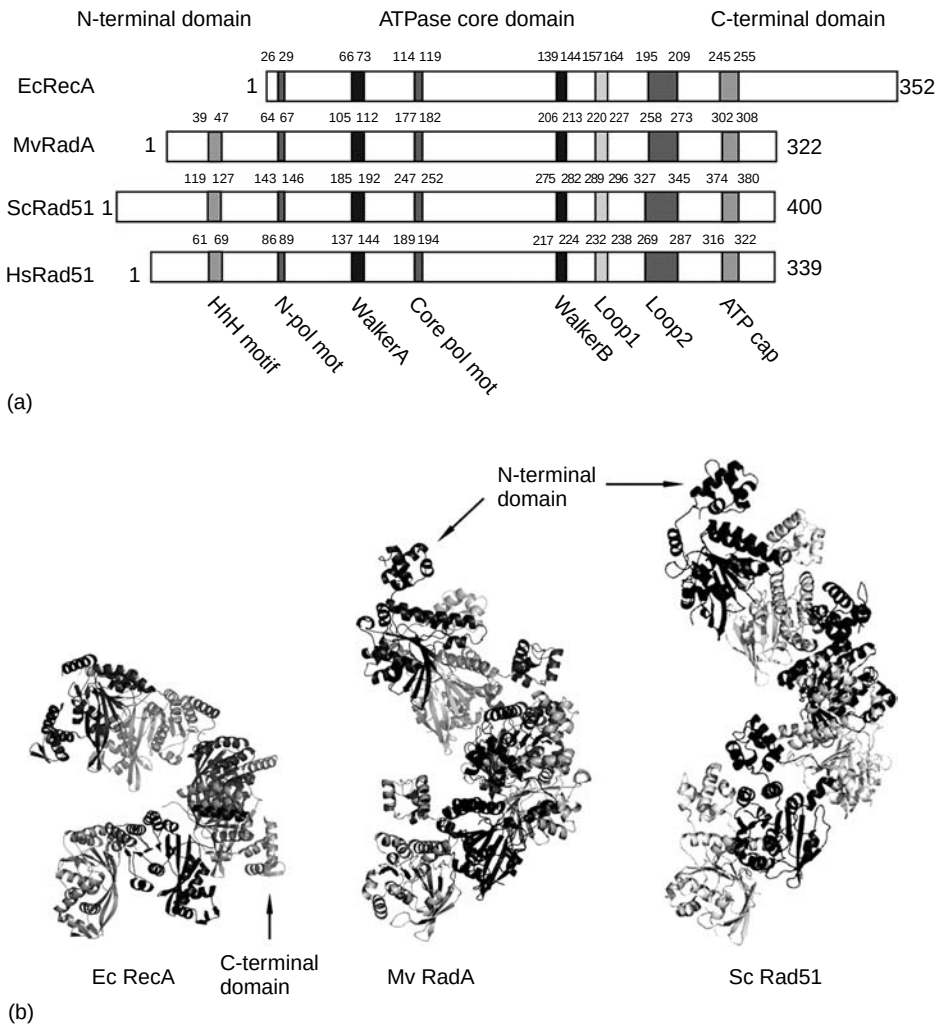

Fig. 2 Domain maps and filament structures of RecA superfamily proteins. (a) Domain map alignment of EcRecA, MvRad51, ScRad51, and HsRad51. HhH, helix-hairpin-helix; N-pol mot, N-terminal polymerization motif (see text). (b) One helical turn (6 subunits) of EcRecA, MvRadA, and ScRad51 protein filament with alternating subunits shown in dark and light colors. These filaments are arranged so that the N-terminal domain (MvRadA and ScRad51) is shown on top, and C-terminal domain (RecA) is positioned at the bottom, as indicated by arrows. Images are made from PDB files: 1REA, 1T4G, and 1SZP.

how new structures of RecA-like proteins now provide important insights into the catalytic organization of these recombinase enzymes. The alignment in Fig. 2(a) indicates some of the conserved domains within RecA, RadA, and Rad51 proteins that will be discussed below. Each member of the RecA superfamily contains the consensus Walker A and B ATP binding site motifs, and recent structural analysis of

RadA and Rad51 proteins reveal a new motif referred to as the *ATP cap* that appears to play important roles in ATP binding and cross-subunit communication. The L1 and L2 domains, as well as the extended C-terminus of RecA and N-terminus of RadA and Rad51, are discussed in terms of their roles in DNA binding. Given the fact that all of these proteins form head-to-tail polymers, large areas of the protein surface are involved in formation of the subunit interface. However, one region, in particular, the N-terminal polymerization motif, deserves mention because of its structural conservation across all members of the superfamily, and especially because of its potential role in specifically regulating Rad51 assembly in human cells.

2 ATP Binding and Hydrolysis

2.1 The Walker A motif

RecA is a DNA-dependant ATPase and contains the traditional Walker A and Walker B sequences that are found in most NTP-requiring enzymes. The Walker A motif, whose consensus sequence is defined as GXXXXGK(T/S), is also known as the P-loop because it contains residues that interact with the phosphate groups of bound NTP. In *E. coli* RecA, this motif extends from residues 66 to 73, and the highly conserved GPESSGKT sequence is identical in 61 of 64 bacterial RecA proteins. Despite this high degree of sequence identity, only four residues have been shown to be absolutely critical to RecA function: Gly66, Gly71, Lys72, and Thr73. In all currently available X-ray crystal structures of RecA proteins, the P-loop lies on the inside surface of the protein filament and near, but not directly within, the subunit–subunit interface. Although numerous mutant RecA proteins have been created with substitutions in the P-loop, work with one mutant, in particular, (Lys72Arg) has provided significant insight into important aspects of the catalytic mechanism of RecA. Although this substitution inhibits ATP hydrolysis, but not binding, and blocks recombination activities *in vivo*, the purified protein still catalyzes strand exchange up to 1.5 kb pairs. This established the idea that ATP binding, rather than ATP hydrolysis, induced a conformationally active form of RecA. Later studies using this same mutant protein showed that ATP turnover is important for coupling the exchange of subunits between free and bound states with further progression of strand exchange.

2.2 The Walker B motif

The Walker B motif in *E. coli* RecA extends from residues 139 to 144 with Asp139 and Asp144 flanking four hydrophobic residues. Although the function of Walker B motifs are less well defined than Walker A and may differ depending on the specific protein, they are most frequently thought to assist in coordinating Mg^{2+} in the bound Mg^{2+} –NTP complex via an Asp residue, play roles in transmitting ATP-induced allosteric information between proteins domains, or, in some cases, may be directly involved in ATP-mediated DNA binding. In the RecA crystal structure, Asp144 binds the Mg^{2+} in the ATP site and is identical in 64 of 64 RecA sequences. Asp139 lies within the subunit interface and is seen in the RecA structure to make a cross-subunit ionic contact with Lys6. However, the relevance of this

interaction to the active form of RecA has been questioned because substitution of various polar and nonpolar amino acids for Asp139 has no deleterious effects on RecA function. A five-residue sequence following the Walker B motif, residues 145 to 149, is identical in 63 of 64 RecA sequences, and analysis of *Mycobacterium tuberculosis* RecA in complex with various NTPs suggests that this "connector" segment serves to link the L2 and L1 regions, both physically and functionally upon binding of NTP, thus playing a role in coordinating the primary and secondary DNA binding sites of RecA.

2.3 The ATP Cap

Recently solved structures of archaeal RadA proteins and yeast Rad51 now provide new and interesting insights into the arrangement of residues at the subunit interface that may be important for stabilizing the binding of ATP and regulating ATP-induced allosteric effects within the protein filaments. The archaeal *Methanococcus voltae* RadA structures were solved in the presence of the nonhydrolyzable ATP analog (AMP-PNP) and form a filament with six subunits per turn and a pitch of approximately 104 to 107 Å. This pitch approximates what has been determined to be the pitch of the active form of RecA. In these structures, the subunit interface near the ATP binding site shows significant differences relative to the arrangement of residues in currently published RecA structures. These MvRadA structures reveal a previously uncharacterized structural region, referred to as the "ATP cap," in which residues from one subunit make specific contacts with the ATP in the binding site of the neighboring subunit. The ATP cap in MvRadA spans residues 302 to 308 and one particularly notable feature is Pro307, which stacks directly against the purine base of AMP-PNP (Fig. 3). Remarkably, this Pro is conserved in all RecA superfamily proteins, including all bacterial RecAs currently sequenced. Other residues in the ATP cap are in close contact with the neighbor's AMP-PNP. For example, Asp302 forms a hydrogen bond with the hydroxyl proton at the 3'position of ribose as well as several water-mediated hydrogen bonds with phosphate oxygen at the α and β positions of AMP-PNP. Interestingly, the equivalent residue in *E. coli* RecA is Lys248, which could make similar bonding interactions with the neighboring ATP. A Lys248Ala inhibits RecA function by disrupting the stability of the protein filament. The new yeast Rad51 structure forms a filament with a pitch of 130 Å, which is within the limits seen by electron microscopy for active RecA/ATPγS/dsDNA filaments. Although crystallization solutions contained ATPγS and a 96-base poly-dT oligonucleotide, only a sulfate ion was observed in the ATP binding site in a position that would be occupied by the β-phosphate of bound ATPγS. Importantly, a region equivalent to the ATP cap in MvRadA is seen to extend into the ATP binding site of the neighboring subunit, and Pro379 is in virtually the same position as the equivalent Pro307 in the MvRadA structure (Fig. 3).

When considering how these new RadA and Rad51 structures may relate to RecA, it is important to note the model proposed by VanLoock et al. In this study, the authors performed a rigid body rotation of a subunit in the RecA crystal structure and manually docked the subunit volume into a reconstruction of the active form of

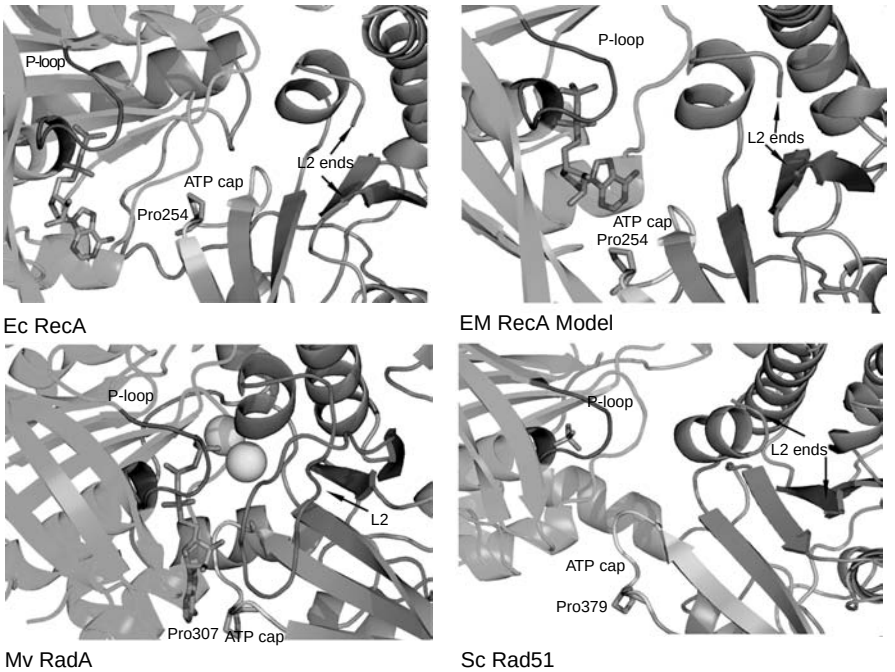

Fig. 3 ATP binding region and the "ATP cap." This view of subunit interface is seen from the inside of the filament. The ribbon diagram of the α-carbon backbone is blue for one subunit and green for the neighbor. In the green subunit, the Walker A motif, or P-loop, is in red, and the following ligands are seen bound to this motif: ADP in EcRecA and EcRecA model; AMP-PNP in MvRadA; a sulfate ion in ScRad51. In the blue subunit, the L2 region (MvRadA) or residues that flank L2 (other 3 panels) are in pink, the "ATP cap" is in yellow, and its universally conserved Proline is in orange. The yellow spheres in the MvRadA structure correspond to two K^+ atoms that are bound to the AMP-PNP. Images are made from PDB files: 1REA, 1N03, 1XU4, and 1SZP (see color plate p. xxxvii).

the RecA filament derived from analyses of electron micrographs. The resulting structure showed an arrangement of subunit–subunit interactions, very similar to those seen in the new RadA and Rad51 structures, and the region that would be the RecA ATP cap is shown in Fig. 3.

2.4 ATP Hydrolysis

While the binding of ATP induces a conformationally active form of RecA, the mechanistic requirement for ATP hydrolysis is still the subject of ongoing studies. Two important features of RecA-mediated strand exchange that require ATP hydrolysis are the ability to progress in a directional fashion and to bypass a variety of DNA lesions or certain lengths of heterology. A recent review of the motor activity of RecA describes models for the coupling of ATP hydrolysis and strand exchange. In the structures of RecA, Glu96 is positioned within the ATP binding site such that it can promote hydrolysis of ATP to ADP, and mutational analysis supports this role. Although both the yeast

and human Rad51 proteins are DNA-dependent ATPases, the mechanistic steps required for ATP binding and hydrolysis are not yet well understood (see Sect. 5: Allosteric Mechanism).

3 DNA Binding Domains

3.1 Loop1 and Loop2

To carry out the process of DNA strand exchange, RecA and Rad51 filaments must bind both, an ssDNA substrate and the homologous duplex DNA. Despite the availability of several new crystal structures for RecA, RadA, and Rad51 proteins, none has included a DNA substrate. Therefore, detail regarding the specific interactions between protein and DNA is still lacking. However, a number of mutagenesis and biochemical efforts have provided a good understanding of domains within the RecA protein that bind DNA and their positions within the filament structure. On the basis of the original RecA structure, Story et al. proposed that two disordered regions were likely to be the primary (or ssDNA) and secondary (or dsDNA) binding sites. Loop 2 (L2) extends from residues 195 to 207 and is considered to be the primary DNA binding site, while loop 1 (residues 157–164) is considered to be part of the secondary DNA binding site (Fig. 4). L2 lies within the inner surface of helical filament structure, while L1 is seen on the top surface of the filament (Fig. 4). Some studies suggest that the L2 and L1 DNA binding domains may interact both physically and functionally, and a "connector" segment extending from residues 145 to 149 may be specifically involved in transmitting NTP-induced structural changes in L2 to L1. The extensive number of biochemical, mutational, and structural studies giving rise to the current model of DNA binding by the RecA protein have been reviewed recently. Interestingly, the MvRadA structures reveal a much closer physical approach of the L2 and L1 regions than is seen in the RecA structures (Fig. 4). Therefore, if the MvRadA structures do reflect the active form of the protein filament, it would support the very reasonable idea that the primary and secondary DNA binding sites are likely to interact both physically and functionally during the catalysis of DNA strand exchange.

3.2 C-terminal Gateway

In the currently available RecA structures, the C-terminal 25 residues are unresolved, but biochemical studies suggest that this region of the protein serves as a "gateway" that regulates the accessibility of dsDNA to its DNA binding site. Further work suggests that a network of salt bridges in this region precludes access of dsDNA to the filament interior, and increasing Mg^{2+} disrupts these interactions thereby opening the "gateway." How this might be accomplished *in vivo* remains to be determined.

3.3 Eukaryotic and Archaeal Uncertainty

While the identity and position of DNA binding domains within various RecA proteins is now fairly well-modeled, the DNA binding domains within the eukaryotic RadA and Rad51 proteins are currently not well defined. The L2 region of HsRad51 includes residues 270 to 287 and lies near the filament axis. However, mutagenesis studies suggest that L2 is not involved in

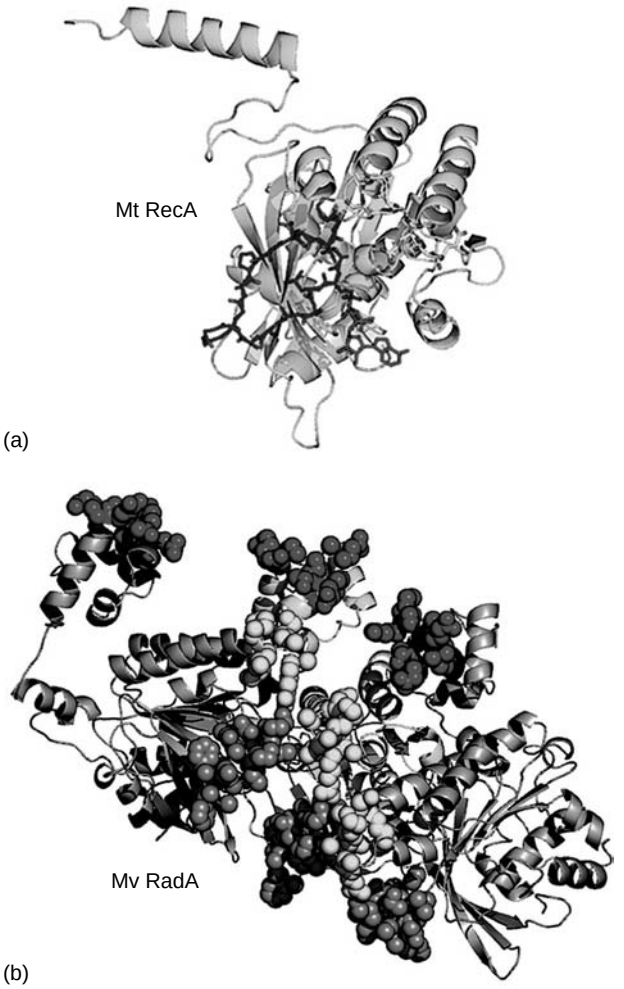

Fig. 4 DNA binding domains of RecA and RadA. (a) Monomer of MtRecA, with the L1 and L2 DNA binding regions shown in yellow and blue respectively. (b) A trimer of MvRadA as seen from the inside of the filament, showing amino acid side chains of L1 (yellow), L2 (blue) and the N-terminal Helix-hairpin-Helix domain (red) as space-filling spheres. Images are made from PDB file: 1G18/1G19 and 1XU4. The green MtRecA subunit (a) is in approximately the same orientation as the green MvRadA subunit (b) (see color plate p. xxxviii).

binding to ssDNA. We have recently created mutant HsRad51 proteins carrying either multiple or single Ala substitutions throughout the L2 region and find that all mutant proteins bind ssDNA with an affinity similar to wild-type HsRad51. Therefore, it seems that the L2 region in HsRad51 does not serve the same function

as L2 in RecA. In contrast, several mutant HsRad51 proteins with Ala substitutions in the L1 region (residues 231–239) show significant decreases in binding to ssDNA.

3.3.1 Helix-hairpin-helix

Relative to the bacterial RecA proteins the eukaryotic and archaeal Rad51 and RadA proteins have an extended N-terminus that contains a helix-hairpin-helix motif ("HhH motif" in Fig. 2a; HsRad51 residues 61–70). NMR chemical shift perturbation mapping of this region suggests that it has a role in directly binding dsDNA in HsRad51. Crystal structures of Rad51 and RadA filaments show that this region is positioned along the upper surface of the filament and could serve as the dsDNA binding domain or be involved in regulating access of dsDNA to the filament interior similar to the C-terminal region of RecA (Fig. 2b, Fig. 4).

4 Oligomeric Interface Domains

The oligomeric form of RecA-like proteins required for the catalysis of DNA strand exchange is an extended helical polymer that wraps around ss- and dsDNA, making contact with the nucleic acid polyphosphate backbone, but not the bases. Formation of the nucleoprotein filament extends the helical pitch of dsDNA from an axial rise per base pair of 3.4 Å (standard B-form dsDNA) to 5.2 Å. These DNA binding properties of RecA led to the idea that one of the primary functions of RecA-like proteins was to reveal the DNA bases in a way that optimizes strand switching and formation of new Watson–Crick bases pairs during the catalysis of strand exchange.

4.1 N-terminal Polymerization Motif

The first crystal structure of *E. coli* RecA revealed an extensive surface on each side of the subunits that formed the oligomeric interface. Earlier genetic and biochemical studies had predicted that a domain in the extreme N-terminus of RecA played an important role in oligomer formation. Comparison of all RecA, RadA, and Rad51 structures now reveals a small conserved structural domain recently termed the *polymerization motif* by Shin et al. Not only is this motif important for the oligomeric stability of RecA-like proteins, in the case of human Rad51, it appears to play an important role in the BRCA2-mediated regulation of Rad51 nucleoprotein filament assembly at the site of DNA damage. The motif forms a short β-strand ("N-pol mot" in Fig. 2a) that interacts in antiparallel fashion with a β-strand in the ATPase core ("Core pol mot" in Fig. 2a) of the neighboring subunit (Fig. 5). Alignment of this conserved region in *E. coli* RecA and several eukaryotic RadA and Rad51 proteins is shown in Table 1. This region is highly conserved among 64 bacterial RecA sequences, and, although not well conserved with the equivalent sequences in eukaryotic RadA and Rad51 proteins, the structure is conserved among all RecA-like proteins. In the Pf Rad51 structure Phe97 and Ala100 interact with the neighboring subunit through contacts with

Tab. 1 Alignment of N-terminal Polymerization Motif across RecA Superfamily Proteins.

MvRadA	63-	G		F	K	S	G	I	-68
HsRad51	85-	G		F	T	T	A	T	-90
PfRad51	95-	G	T	F	M	R	A	D	-101
ScRad51	143-	G		F	V	T	A	A	-148
HsBRC4	1523-	G		F	H	T	A	S	-1528
EcRecA	24	G	S	I	M	R	L	G	-30

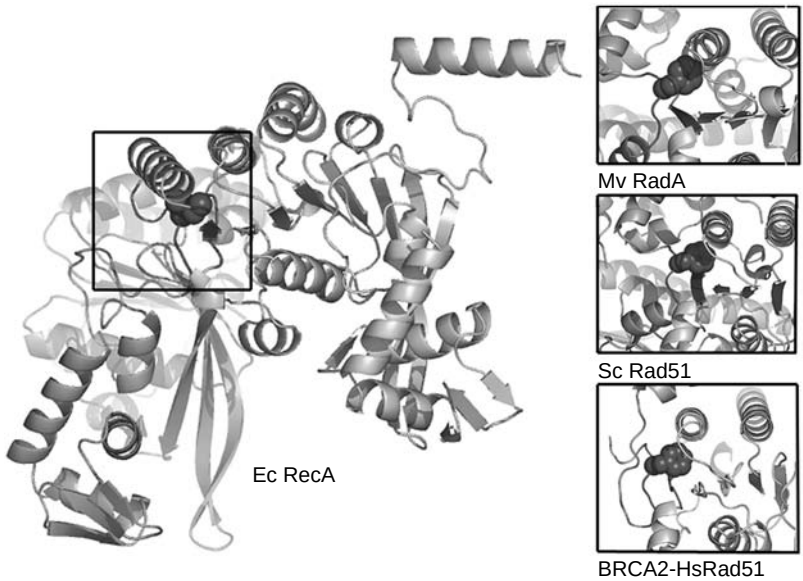

Fig. 5 N-terminal Polymerization motif. The left-hand image is two subunits of the EcRecA filaments (viewed from the outside surface of the filament), showing the N-terminal polymerization motif (green) and the Core Polymerization Motif (red). Isoleucine 26 is shown in space-filling spheres. The boxes on the right show the equivalent interface region in MvRadA (Phe64), ScRad51 (Phe144), and the BRC4-HsRad51 fusion protein (Phe1524) with indicated phenylalanine residues interacting with the neighboring subunit. Images were made from PDB files: 1REA, 1XU4, 1SZP and 1NOW (see color plate p. xxxix).

hydrophobic regions, and similar cross-subunit interactions are observed for the conserved residues in the ScRad51 and MvRadA. Mutational analysis of the equivalent residues in HsRad51, Phe86 and Ala89, supports the idea that both are important for oligomer stability. Further support for the idea that this region is an important part of the oligomeric interface comes from fluorescence studies of Xenopus Rad51. The structure of an HsRad51-BRC4 fusion protein gave rise to the idea that the BRC repeats in BRCA2 interact with Rad51 by mimicking the N-terminal polymerization motif surface. The alignment in Table 1 shows that Phe1524 and Ala1527 in the BRC4 repeat are conserved in this motif. Examination of the *E. coli* RecA, MvRadA, ScRad51, and HsRad51-BRC4 protein structures shows a striking similarity in cross-subunit interactions in this region of the interface, highlighted by insertion of a conserved Phe or Ile into a hydrophobic area of the neighboring subunit (Fig. 5). Further details of the mechanism by which the BRCA2 protein regulates assembly of HsRad51 nucleoprotein filaments at the site of DNA damage is provided in a recent review by Pellegrini and Venkitaraman.

4.2 Other Cross-subunit Contacts

Residues involved in making cross-subunit contacts account for a high

percentage of the conserved residues seen in 64 bacterial RecA sequences. Roca and Cox noted a highly conserved region corresponding to residues 214 to 222 in *E. coli* RecA that they termed a *RecA signature sequence*. In this stretch, 5 of the 9 residues are identical in 64 bacterial RecA sequences (Ala214, Leu215, Lys216, Phe217, Arg222) and the other positions show a high degree of chemical conservation. Modeling studies of *E. coli* RecA and the crystal structures of the RadA and Rad51 proteins support the idea that this region of the interface is more extensively involved with direct contacts with the ATP binding site in the neighboring subunit than the original RecA structure revealed. This region will be discussed in greater detail in the Sect. 5: Allosteric Mechanism.

5 Allosteric Mechanism

5.1 Allosteric Switch: Gln194

Early biochemical studies of RecA showed that ATP serves as an allosteric effector of enzyme function, that is, its binding causes RecA to become catalytically active. For example, although RecA binds with low affinity to ssDNA in the absence of ATP, its DNA binding affinity is increased dramatically in the presence of ATP or a nonhydrolyzable analog such as ATPγS. The RecA crystal structure shows that Gln194 is positioned within the ATP site near to where the γ-phosphate of bound ATP would be, and, therefore, may serve as the sensor of ATP binding. Mutational and biochemical studies confirmed that Gln194 is indeed an "allosteric switch" that triggers conformational changes within the protein in response to ATP binding, an action that is critical for RecA function (see below). Subsequent studies suggested that Gln194 also plays a more direct role in ATP hydrolysis in a manner similar to G-proteins.

5.2 ATP Sensing: Phe217

Large-scale mutagenesis efforts identified Phe217 as a critical subunit interface residue involved in transmitting ATP-mediated allosteric information throughout the oligomeric filament structure. Of 15 different substitutions at Phe217, only Tyr maintained wild-type-like activity *in vivo*, while all others were significantly inhibited. Biochemical studies showed that the Phe217Tyr substitution resulted in a hypercooperativity mutant that increased the kinetics of ATP-induced filament assembly by more than 250-fold. Together with other mutagenesis studies, these data suggested that the following molecular events occur within the RecA structure upon ATP binding. The interaction of Gln194 with the γ-phosphate of ATP triggers a conformational change in the L2 region (residues 195–209). Glycine residues at positions 211 and 212 are nonmutable, and, undoubtedly, are critical to propagating the conformational change into a small helix immediately downstream of L2 (residues 213–218). This results in a specific movement of Phe217 toward the neighboring subunit. However, in the original RecA crystal structure, it was not apparent what residues in the neighbor might interact with Phe217. Subsequent modeling studies using image analyses of electron micrographs suggested that subunits in the active RecA filament are reoriented such that Phe217 and other nearby side chains, for example, Lys216

and Arg222, lie very close to the ATP binding site of the neighboring subunit. In fact, the structures of MvRadA and ScRad51 show that this is likely the case (Fig. 6). The MvRadA/AMP-PNP structure, solved in potassium-rich media, shows that the NTP sits directly within the interface between two subunits. His280, which aligns with *E. coli* RecA Phe217, makes a hydrogen bond with the γ-phosphate of AMP-PNP in the neighboring subunit (Fig. 6). A similar arrangement is seen in the ScRad51 structure in which His352 (equivalent to RecA Phe217 and MvRadA His280) lies close to a sulfate ion in the ATP binding site of the neighboring subunit (sulfate results from slow hydrolysis of ATPγS; Fig. 6). In this study, a His352Ala mutant was shown to be inhibited for ssDNA binding, thereby supporting the model in which His352 is important for formation of a filament that is competent for Rad51 function. Interestingly, when MvRadA was crystallized in low potassium media, the details of side chain conformations at the interface differ, in that access to the neighboring ATP binding site by His280 is blocked by Phe107 (Fig. 6). Similarly,

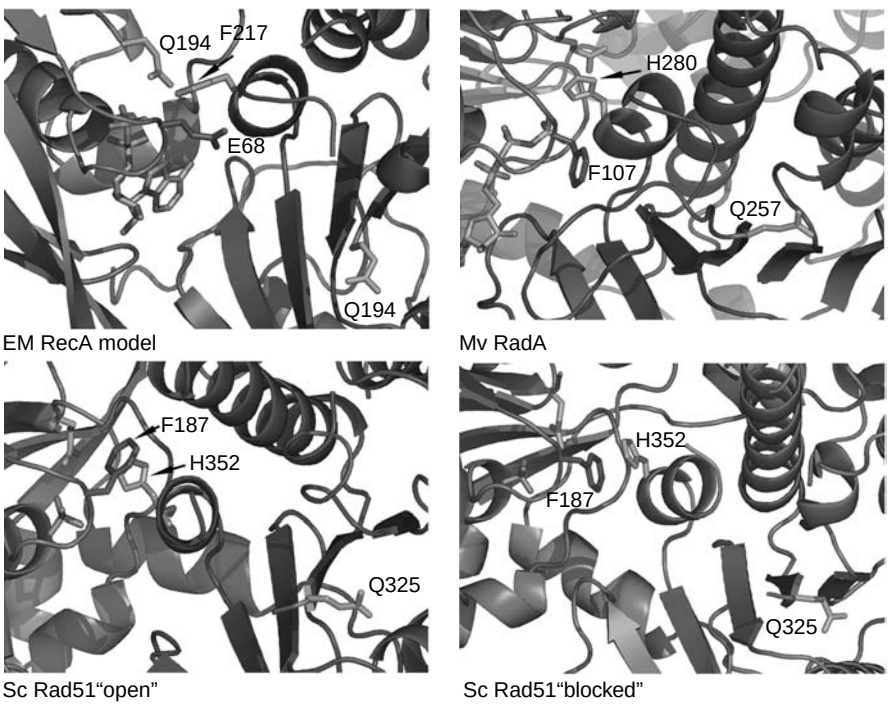

Fig. 6 Allosteric mechanism. Four panels showing different conformations of residues important in sensing ATP binding and conformational change of the filament structure. Glutamine 194 (EcRecA) and its homologous residues in MvRadA (Gln257) and ScRad51 (Gln325) are shown in yellow, while Phe217 (EcRecA) and the homologous histidine residues (MvRadA, His280; ScRad51, His352) are shown in purple. Glutamate 68 (EcRecA) and equivalent phenylalanine residues are in red. The region between Q194 and Phe217 (loop2) is unstructured in all but MvRadA. Images are made from PDB files: 1N03, 1XU4, and 1SZP (see color plate p. xl).

alternating interfaces in the ScRad51 structure show that His352 is blocked from interacting with the neighbor's ATP site by Phe187 (equivalent to MvRadA Phe107). These results may be indicative of the inherent flexibility within a RecA or Rad51 filament as it progresses through the cycle of ATP binding and hydrolysis during the catalysis of DNA strand exchange (see below). Therefore, given these new archaeal RadA and yeast Rad51 structures it seems likely that the positioning of bound ATP within the subunit interface contributes directly to subunit communication and to the cooperative nature of many aspects of recombinase function. A mechanistic similarity to other oligomeric ATPases, for example, DNA helicases, regarding the positioning of ATP binding sites has recently been described by Egelman. Although the contribution made by His280 in MvRadA and His352 in ScRad51 seems to involve direct hydrogen bonding with ATP in the neighboring subunit, it is currently not clear what kind of cross-subunit interactions are made by Phe217 in RecA. It is possible that Phe217 stacks with the purine ring of bound ATP, but further studies will be required to understand the specific molecular interactions of Phe217 in the active form of the RecA filament.

Despite a very similar arrangement of equivalent side chains in the ATP binding site and subunit interface, the human Rad51 protein shows a distinct difference regarding its use of ATP as an allosteric effector compared to RecA and yeast Rad51. Whereas RecA and ScRad51 require ATP to achieve a high-affinity ssDNA binding conformation, HsRad51 does not. It binds ssDNA cooperatively and with high affinity in the absence of ATP. This can be seen in Fig. 7, which shows biosensor DNA binding traces for *E. coli* RecA, ScRad51, and HsRad51. In this assay, protein is added directly to a cuvette containing immobilized ssDNA and binding of protein is measured in real time. Additionally, whereas a Gln194Asn or Gln194Ala mutation in RecA prevents high-affinity binding to ssDNA, equivalent mutations in HsRad51, for example, Gln268Asn, show no impairment of ssDNA binding. Although catalysis of DNA strand exchange by HsRad51 *in vitro* appears to be dependent on ATP hydrolysis, conflicting reports have appeared in the literature regarding the requirement for ATP hydrolysis for HsRad51 function *in vivo*. There are clear differences in the mechanisms by which various recombinase enzymes use NTP cofactors to drive catalysis of DNA strand exchange and further work will be required to fully understand these distinct requirements.

6 Proteins that interact with RecA

Although RecA shows proficient strand exchange activity *in vitro*, its activities *in vivo* depend on specific interactions with a variety of other proteins that are either directly involved in the catalysis of DNA repair or in the regulation of expression of DNA repair genes. Recent efforts have improved our understanding of both specific regions within the RecA protein structure responsible for these interactions as well as the functional consequences of these interactions.

6.1 RecBCD

The RecBCD enzyme is a highly processive helicase and nuclease whose activities undergo significant modification upon encountering a recombination hotspot

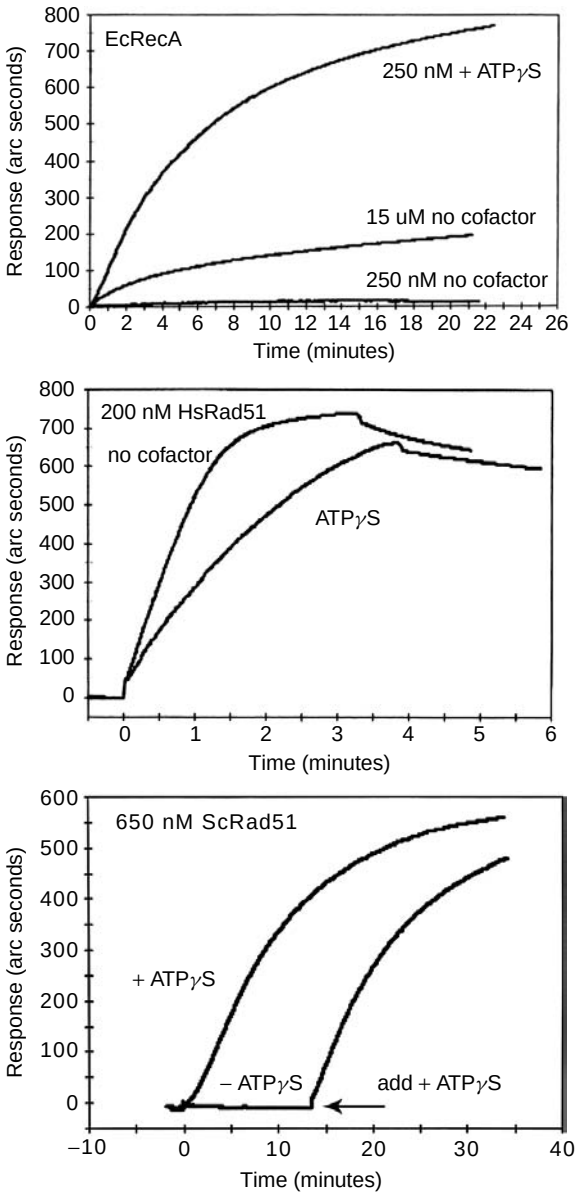

Fig. 7 Biosensor DNA binding data. These curves represent real-time binding of the indicated protein to ssDNA immobilized in an IAsys biosensor cuvette. The concentration of protein, as well as the presence or absence of nucleotide cofactor (ATPγS), is indicated in each panel.

known as *chi*. RecBCD generates 3'single-stranded ends that serve as a substrate for RecA binding and strand exchange. RecBCD also promotes loading of RecA onto ssDNA by virtue of a direct protein–protein interaction mediated by the C-terminal domain of RecB and some region of the core domain of RecA. The recently published X-ray structure of RecBCD now provides a clearer view of how the many activities of this enzyme are coordinated, as well as how the RecA loading site in RecB is likely revealed following RecBCDs encounter with *chi*.

6.2 RecF, O, and R

While early genetic studies suggested that the RecF, RecO, and RecR proteins play subtle, but important, roles in RecA-mediated recombination, recent studies show that they are specifically involved in RecA-mediated repair of gapped DNA lesions that frequently result when a replication complex stalls at a DNA lesion. It appears that these proteins can function both as subcomplexes as well as a heterotrimeric complex. RecFR and RecOR regulate the loading of RecA and the degree of RecA filament extension, respectively, within a gapped region. The RecFOR complex has been reported to serve as a structure-specific mediator of RecA by targeting recombinational repair to the ssDNA–dsDNA junctions in gapped DNA. Other studies support a role for RecFOR in clearing RecA from recombination intermediates to allow replication restart. Although no specific protein–protein interactions have been demonstrated between RecA and RecF, RecO, or RecR, studies by Bork et al. suggest that RecR specifically contacts RecA.

6.3 RecX and DinI

The *recX* and *dinI* genes are part of the SOS regulon and, through direct interaction with the RecA filament, serve to regulate RecA activity. In many bacterial species, the RecX protein is a negative modulator of RecA activity, inhibiting RecA ATPase, strand exchange, and coprotease activities, and serves to prevent harmful overexpression of RecA protein caused by DNA damage. However, in *Neisseria gonorrhoeae* RecX appears to be necessary for normal RecA function. Three-dimensional reconstructions of electron micrographs show that RecX binds to RecA nucleoprotein filaments by spanning the subunit–subunit interface, contacting the C-terminal domain of one subunit and the core domain of the neighboring subunit. However, recent studies show that the maximal inhibitory effect of RecX occurs at low, substoichiometric amounts of RecX relative to RecA. Drees et al. also show that RecX blocks extension of RecA nucleoprotein filaments through a proposed end-capping mechanism, resulting in a net filament disassembly. The interaction between RecX and RecA is modulated by the negatively charged C-terminal region of RecA, and is also enhanced by increasing Mg^{+} concentration.

Until very recently, DinI had been thought to be a negative regulator of RecA function, serving to attenuate the SOS response following DNA repair. However, new studies support the idea that the action of DinI opposes that of RecX and serves to stabilize active RecA nucleoprotein filaments. The inhibitory effects of DinI on RecA function occur only at elevated concentrations of DinI, whereas lower concentrations of DinI promote RecA activities. These studies have shown the

opposing effects of RecX and DinI to occur both *in vitro* and *in vivo*. Interestingly, at concentrations of DinI that stabilize RecA nucleoprotein filaments, RecA-mediated cleavage of UmuD is inhibited, whereas cleavage of LexA is not. DinI binds to the core domain of RecA and may specifically compete for UmuD binding, but not LexA. Additionally, the C-terminal domain of RecA modulates DinI function and provides regulated access to the core domain of RecA, perhaps in the same way that the C-terminal domain regulates access of dsDNA to the interior of the RecA filament.

6.4 Coprotease Substrates

The active complex of RecA/ATP/ssDNA also catalyzes the autoproteolytic inactivation of several repressor proteins, for example, LexA, phage λ cI and the phage ϕ80 repressors as well as the autoproteolytic activation of the UmuD protein. This activity has no mechanistic resemblance to the strand exchange activity of RecA, nor does it require extended nucleoprotein filaments. The original assumption was that RecA itself had a protease activity, but with the discovery by Little in 1984 that LexA and the phage λ cI repressors undergo a specific autodigestion reaction, it became clear that binding to RecA induced a conformation in these substrate proteins that was competent for autocleavage. Although binding sites for coprotease substrates within the RecA structure have yet to be specifically defined, a large number of mutagenesis, biochemical, and biophysical studies have provided strong evidence that the coprotease binding site includes specific positions within the groove of the RecA filament as well as positions on the upper surface of the filament that overlap with the L1 DNA binding domain. This subject and all of the studies leading up to this model have been reviewed recently by McGrew and Knight.

6.5 PolV

In addition to its activities as a strand exchange and coprotease enzyme, RecA participates directly in an important process referred to "translesion bypass DNA synthesis." The $UmuD'_2C$ complex (pol V) is an error-prone DNA polymerases in *E. coli*, and, as part of the recovery from DNA damage, pol V is activated to ensure that some of the unrepaired DNA lesions will not block DNA synthesis to an extent that kills the cell. Owing to its low fidelity, pol V gives rise to an increased level of mutations, and this process in bacteria has been termed *SOS mutagenesis* or *UV-induced mutagenesis*. RecA has several distinct roles in this process: (1) it serves to induce all SOS genes, including the *umuC* and *umuD* genes, by catalyzing autodigestion of the LexA repressor; (2) it catalyzes the autoproteolytic activation of UmuD to the active UmuD' form; (3) and it interacts directly with the $UmuD'_2C$ complex at the site of lesion bypass. The region of the RecA structure that interacts with $UmuD'_2C$ is located on the surface of the subunit interface that would be at the 3′ end of bound DNA, the end of the filament that will be facing the lesion needing to be bypassed by polV.

6.6 Proteins that Interact with Rad51

A number of proteins have been identified that interact with either yeast or human Rad51 using various screens such

as the "yeast 2-hybrid assay" or co-immunoprecipitations. This work, and descriptions of the likely biological functions of many of these proteins, has been reviewed recently by Thompson and Schild. A number of ongoing studies are aimed at characterizing the functional consequences of these interactions and providing some molecular detail regarding specific protein domains and residues involved in these interactions. Briefly, one of the best-characterized interactions is between yeast Rad51 and Rad52 proteins. This work led to the identification of Rad52 as a "mediator" of Rad51 function, that is, Rad52 assists with formation of Rad51 filaments at the site of a DNA double-strand break, and also helps Rad51 compete with other DNA binding proteins that may block its function. This particular aspect of Rad52 function is particularly interesting because it appears to have been conserved in all organisms that perform RecA/Rad51-mediated HR. For example, RecFOR assists RecA in a way similar to the Rad52/Rad51 interaction, and, in human cells, the BRCA2 protein displays a "mediator" function regarding its role in regulating assembly of Rad51 nucleoprotein filaments.

Important questions remain regarding how the mechanistic differences seen for bacterial RecA and eukaryotic RadA/Rad51 proteins relate to the various mediators and other proteins that assist each recombinase, as well as varied cellular environments in which these proteins catalyze recombination events.

7
Future Directions

Homologous recombination is one of several major DNA repair pathways dedicated to preserving the integrity of genetic information and is found in all free living organisms. This pathway operates to repair errors that occur frequently during normal metabolic processes such as DNA replication, as well as more catastrophic DNA lesions resulting from exposure to environmental mutagens. In eukaryotic organisms, recombination also plays an important role in proper alignment and segregation of chromosomes during meiotic division. Recent studies show that defects in homologous recombination correlate with the occurrence of several cancer syndromes in humans, thus emphasizing the need for a more detailed mechanistic understanding of recombination proteins. Studies over the past several years have significantly improved our understanding of the molecular design of the bacterial RecA protein and how its many structural domains coordinate their functions to carry out a wide variety of activities that are mechanistically unrelated, such as regulation of gene expression, catalysis of DNA strand exchange, and promotion of DNA synthesis by error-prone polymerases. RecA is a classical allosterically regulated enzyme, and ATP binding triggers a number of conformational changes that activate the protein for its diverse functions. Recent insights from new modeling procedures as well as new structures of archaeal and eukaryotic RecA-like proteins (RadA and Rad51, respectively) have significantly improved our understanding of the catalytic organization of RecA. RecA-like proteins in higher organisms must catalyze the same DNA strand exchange reaction as in bacteria, but the greater complexity of the cellular environment as well as the DNA substrate itself (chromatin) demands a more sophisticated interplay between Rad51 and various other proteins

involved directly or indirectly in catalyzing homologous recombination. A major challenge that is currently driving many studies in this area is to fully understand the complexities of homologous recombination in higher organisms, not only with regard to the function of Rad51 and its associated proteins but also to further our understanding of how the cell coordinates various pathways involved in the detection of DNA damage, slowing of cell growth to facilitate DNA repair, and the activation of DNA repair pathways.

Acknowledgments

The authors wish to acknowledge Dr. Martin Marinus, Dr. Nick Rhind, and Dr. Phoebe Rice for their commentary on the manuscript. Thanks to Dr. Phoebe Rice and Dr. Bill Royer for aid in making structural images. This work was supported by NIH grants GM44772 and GM65851 to K.L.K.

See also Aging and Sex, DNA Repair in; DNA Repair in Yeast; Nucleic Acids (DNA) Damage and Repair; Recombination and Genome Rearrangements; Regulome; Repair and Mutagenesis of DNA.

Bibliography

Books and Reviews

Conway, A.B., Lynch, T.W., Zhang, Y., et al. (2004) Crystal structure of a Rad51 filament, *Nat. Struct. Mol. Biol.* **11**, 791–796.

Cox, M.M. (2003) The bacterial RecA protein as a motor protein, *Annu. Rev. Microbiol.* **57**, 551–577.

Datta, S., Ganesh, N., Chandra, N.R., et al. (2003) Structural studies on MtRecA-nucleotide complexes: insights into DNA and nucleotide binding and the structural signature of NTP recognition, *Proteins* **50**, 474–485.

Kelley, J.A., Knight, K.L. (1997) Allosteric regulation of RecA protein function is mediated by Gln194, *J. Biol. Chem.* **272**, 25778–25782.

Kowalczykowski, S.C. (2000) Initiation of genetic recombination and recombination-dependent replication, *Trends Biochem. Sci.* **25**, 156–165.

McGrew, D.A., Knight, K.L. (2003) Molecular design and functional organization of the RecA protein, *Crit. Rev. Biochem. Mol. Biol.* **38**, 385–432.

Pellegrini, L., Yu, D.S., Lo, T., et al. (2002) Insights into DNA recombination from the structure of a RAD51-BRCA2 complex, *Nature* **420**, 287–293.

Shin, D.S., Pellegrini, L., Daniels, D.S., et al. (2003) Full-length archaeal Rad51 structure and mutants: mechanisms for RAD51 assembly and control by BRCA2, *EMBO J.* **22**, 4566–4576.

Story, R.M., Weber, I.T., Steitz, T.A. (1992) The structure of the E. coli recA protein monomer and polymer, *Nature* **355**, 318–325.

VanLoock, M.S., Yu, X., Yang, S., et al. (2003) ATP-Mediated conformational changes in the RecA filament, *Structure (Camb.)* **11**, 187–196.

Wu, Y., Qian, X., He, Y., et al. (2004) Crystal structure of an ATPase-active form of Rad51 homolog from methanococcus voltae: insights into potassium-dependence, *J. Biol. Chem.* **280**, 722–728.

Primary Literature

Aihara, H., Ito, Y., Kurumizaka, H., et al. (1999) The N-terminal domain of the human Rad51 protein binds DNA: structure and a DNA binding surface as revealed by NMR, *J. Mol. Biol.* **290**, 495–504.

Baumann, P., Benson, F.E., West, S.C. (1996) Human Rad51 protein promotes ATP-dependent homologous pairing and strand transfer reactions in vitro, *Cell* **87**, 757–766.

Bork, J.M., Cox, M.M., Inman, R.B. (2001) The RecOR proteins modulate RecA protein function at 5′ ends of single-stranded DNA, *EMBO J.* **20**, 7313–7322.

Campbell, M.J., Davis, R.W. (1999) On the in vivo function of the RecA ATPase, *J. Mol. Biol.* **286**, 437–445.

Churchill, J.J., Kowalczykowski, S.C. (2000) Identification of the RecA protein-loading domain of RecBCD enzyme, *J. Mol. Biol.* **297**, 537–542.

Clark, A.J., Margulies, A.D. (1965) Isolation and characterization of recombination-deficient mutants of Escherichia coli K12, *Proc. Natl. Acad. Sci. U.S.A.* **53**, 451–459.

De Zutter, J.K., Knight, K.L. (1999) The hRad51 and RecA proteins show significant differences in cooperative binding to single-stranded DNA, *J. Mol. Biol.* **293**, 769–780.

Drees, J.C., Lusetti, S.L., Cox, M.M. (2004) Inhibition of RecA protein by the Escherichia coli RecX protein: modulation by the RecA C-terminus and filament functional state, *J. Biol. Chem.* **279**, 52991–52997.

Drees, J.C., Lusetti, S.L., Chitteni-Pattu, S., et al. (2004) A RecA filament capping mechanism for RecX protein, *Mol. Cells* **15**, 789–798.

Egelman, E.H. (1993) What do X-ray crystallographic and electron microscopic structural studies of the RecA protein tell us about recombination?, *Curr. Opin. Struct. Biol.* **3**, 189–197.

Egelman, E.H. (2003) A tale of two polymers: new insights into helical filaments, *Nat. Rev. Mol. Cell Biol.* **4**, 621–630.

Egelman, E.H., Stasiak, A. (1993) Electron microscopy of RecA-DNA complexes: two different states, their functional significance and relation to the solved crystal structure, *Micron* **24**, 309–324.

Eldin, S., Forget, A.L., Lindenmuth, D.M., et al. (2000) Mutations in the N-terminal region of RecA that disrupt the stability of free protein oligomers but not RecA-DNA complexes, *J. Mol. Biol.* **299**, 91–101.

Goodman, M.F. (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes, *Annu. Rev. Biochem.* **71**, 17–50.

Hortnagel, K., Voloshin, O.N., Kinal, H.H., et al. (1999) Saturation mutagenesis of the E. coli RecA loop L2 homologous DNA pairing region reveals residues essential for recombination and recombinational repair, *J. Mol. Biol.* **286**, 1097–1106.

Karlin, S., Brocchieri, L. (1996) Evolutionary conservation of RecA genes in relation to protein structure and function, *J. Bacteriol.* **178**, 1881–1894.

Kelley De Zutter, J., Forget, A.L., Logan, K.M., et al. (2001) Phe217 regulates the transfer of allosteric information across the subunit interface of the RecA protein filament, *Structure (Camb.)* **9**, 47–55.

Kiselev, V.I., Glukhov, A.I., Tarasova, I.M., et al. (1988) [Accumulation of N-terminal fragment of recA protein in the htpR- mutant impairs the SOS-function of Escherichia coli cells], *Mol. Biol. (Mosk.)* **22**, 1198–1203.

Koller, T., Dicapua, E., Stasiak, A. (1983) Complexes of RecA with single stranded DNA, in: Cozzarelli N. (Ed.) Alan R. Liss, *Mechanisms of DNA Replication and Recombination*, NY, 723–729.

Konola, J.T., Logan, K.M., Knight, K.L. (1994) Functional characterization of residues in the P-loop motif of the RecA protein ATP binding site, *J. Mol. Biol.* **237**, 20–34.

Konola, J.T., Nastri, H.G., Logan, K.M., et al. (1995) Mutations at Pro67 in the RecA protein P-loop motif differentially modify coprotease function and separate coprotease from recombination activities, *J. Biol. Chem.* **270**, 8411–8419.

Krejci, L., Song, B., Bussen, W., et al. (2002) Interaction with Rad51 is indispensable for recombination mediator function of Rad52, *J. Biol. Chem.* **277**, 40132–40141.

Kurumizaka, H., Aihara, H., Ikawa, S., et al. (1996) A possible role of the C-terminal domain of the RecA protein. A gateway model for double-stranded DNA binding, *J. Biol. Chem.* **271**, 33515–33524.

Kurumizaka, H., Aihara, H., Kagawa, W., et al. (1999) Human Rad51 amino acid residues required for Rad52 binding, *J. Mol. Biol.* **291**, 537–548.

Little, J.W. (1984) Autodigestion of lexA and phage lambda repressors, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1375–1379.

Little, J.W. (1993) LexA cleavage and other self-processing reactions, *J. Bacteriol.* **175**, 4943–4950.

Little, J.W., Mount, D.W. (1982) The SOS regulatory system of Escherichia coli, *Cell* **29**, 11–22.

Liu, Y., West, S.C. (2004) Happy hollidays: 40th anniversary of the holliday junction, *Nat. Rev. Mol. Cell Biol.* **5**, 937–944.

Liu, Y., Stasiak, A.Z., Masson, J.Y., et al. (2004) Conformational changes modulate the activity of human RAD51 protein, *J. Mol. Biol.* **337**, 817–827.

Logan, K.M., Knight, K.L. (1993) Mutagenesis of the P-loop motif in the ATP binding site of the RecA protein from Escherichia coli, *J. Mol. Biol.* **232**, 1048–1059.

Luo, Y., Pfuetzner, R.A., Mosimann, S., et al. (2001) Crystal structure of LexA: a conformational switch for regulation of self-cleavage, *Cell* **106**, 585–594.

Lusetti, S.L., Shaw, J.J., Cox, M.M. (2003) Magnesium ion-dependent activation of the RecA protein involves the C terminus, *J. Biol. Chem.* **278**, 16381–16388.

Lusetti, S.L., Drees, J.C., Stohl, E.A., et al. (2004) The DinI and RecX proteins are competing modulators of RecA function, *J. Biol. Chem.* **279**, 55073–55079.

Menetski, J.P., Kowalczykowski, S.C. (1985) Interaction of recA protein with single-stranded DNA. Quantitative aspects of binding affinity modulation by nucleotide cofactors, *J. Mol. Biol.* **181**, 281–295.

Menetski, J.P., Kowalczykowski, S.C. (1990) Biochemical properties of the Escherichia coli recA430 protein. Analysis of a mutation that affects the interaction of the ATP-recA protein complex with single-stranded DNA, *J. Mol. Biol.* **211**, 845–855.

Morimatsu, K., Kowalczykowski, S.C. (2003) RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange. A universal step of recombinational repair, *Mol. Cells* **11**, 1337–1347.

Nastri, H.G., Guzzo, A., Lange, C.S., et al. (1997) Mutational analysis of the RecA protein L1 region identifies this area as a probable part of the co-protease substrate binding site, *Mol. Microbiol.* **25**, 967–978.

Nguyen, T.T., Muench, K.A., Bryant, F.R. (1993) Inactivation of the recA protein by mutation of histidine 97 or lysine 248 at the subunit interface, *J. Biol. Chem.* **268**, 3107–3113.

Ogawa, T., Wabiko, H., Tsurimoto, T., et al. (1979) Characteristics of purified recA protein and the regulation of its synthesis in vivo, *Cold Spring Harb Symp. Quant. Biol.* **43**(Pt 2), 909–915.

Papavinasasundaram, K.G., Colston, M.J., Davis, E.O. (1998) Construction and complementation of a recA deletion mutant of Mycobacterium smegmatis reveals that the intein in Mycobacterium tuberculosis recA does not affect RecA function, *Mol. Microbiol.* **30**, 525–534.

Pellegrini, L., Venkitaraman, A. (2004) Emerging functions of BRCA2 in DNA recombination, *Trends Biochem. Sci.* **29**, 310–316.

Rehrauer, W.M., Kowalczykowski, S.C. (1993) Alteration of the nucleoside triphosphate (NTP) catalytic domain within Escherichia coli recA protein attenuates NTP hydrolysis but not joint molecule formation, *J. Biol. Chem.* **268**, 1292–1297.

Roca, A.I., Cox, M.M. (1997) RecA protein: structure, function, and role in recombinational DNA repair, *Prog. Nucleic Acid Res. Mol. Biol.* **56**, 129–223.

Sano, Y. (1993) Role of the recA-related gene adjacent to the recA gene in Pseudomonas aeruginosa, *J. Bacteriol.* **175**, 2451–2454.

Saraste, M., Sibbald, P.R., Wittinghofer, A. (1990) The P-loop–a common motif in ATP- and GTP-binding proteins, *Trends Biochem. Sci.* **15**, 430–434.

Sedgwick, S.G., Yarranton, G.T. (1982) Cloned truncated recA genes in E. coli. I. Effect on radiosensitivity and recA+ dependent processes, *Mol. Gen. Genet.* **185**, 93–98.

Selmane, T., Camadro, J.M., Conilleau, S., et al. (2004) Identification of the subunit-subunit interface of Xenopus Rad51.1 protein: similarity to RecA, *J. Mol. Biol.* **335**, 895–904.

Silver, M.S., Fersht, A.R. (1982) Direct observation of complexes formed between recA protein and a fluorescent single-stranded deoxyribonucleic acid derivative, *Biochemistry* **21**, 6066–6072.

Singleton, M.R., Dillingham, M.S., Gaudier, M., et al. (2004) Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks, *Nature* **432**, 187–193.

Skiba, M.C., Knight, K.L. (1994) Functionally important residues at a subunit interface site in the RecA protein from Escherichia coli, *J. Biol. Chem.* **269**, 3823–3828.

Sonoda, E., Takata, M., Yamashita, Y.M., et al. (2001) Homologous DNA recombination in vertebrate cells, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8388–8394.

Sprang, S.R. (1997a) G protein mechanisms: insights from structural analysis, *Annu. Rev. Biochem.* **66**, 639–678.

Sprang, S.R. (1997b) G proteins, effectors and GAPs: structure and mechanism, *Curr. Opin. Struct. Biol.* **7**, 849–856.

Stark, J.M., Hu, P., Pierce, A.J., et al. (2002) ATP hydrolysis by mammalian RAD51 has a key role during homology-directed DNA repair, *J. Biol. Chem.* **277**, 20185–20194.

Stasiak, A., Di Capua, E., Koller, T. (1981) Elongation of duplex DNA by recA protein, *J. Mol. Biol.* **151**, 557–564.

Stohl, E.A., Seifert, H.S. (2001) The recX gene potentiates homologous recombination in Neisseria gonorrhoeae, *Mol. Microbiol.* **40**, 1301–1310.

Stohl, E.A., Brockman, J.P., Burkle, K.L., et al. (2003) Escherichia coli RecX inhibits RecA recombinase and coprotease activities in vitro and in vivo, *J. Biol. Chem.* **278**, 2278–2285.

Story, R.M., Steitz, T.A. (1992) Structure of the recA protein-ADP complex, *Nature* **355**, 374–376.

Sugiyama, T., Kowalczykowski, S.C. (2002) Rad52 protein associates with replication protein A (RPA)-single-stranded DNA to accelerate Rad51-mediated displacement of RPA and presynaptic complex formation, *J. Biol. Chem.* **277**, 31663–31672.

Sukchawalit, R., Vattanaviboon, P., Utamapongchai, S., et al. (2001) Characterization of Xanthomonas oryzae pv. oryzae recX, a gene that is required for high-level expression of recA, *FEMS Microbiol. Lett.* **205**, 83–89.

Sutton, M.D., Smith, B.T., Godoy, V.G., et al. (2000) The SOS response: recent insights into umuDC-dependent mutagenesis and DNA damage tolerance, *Annu. Rev. Genet.* **34**, 479–497.

Thompson, L.H., Schild, D. (2001) Homologous recombinational repair of DNA ensures mammalian chromosome stability, *Mutat. Res.* **477**, 131–153.

Thompson, L.H., Schild, D. (2002) Recombinational DNA repair and human disease, *Mutat. Res.* **509**, 49–78.

VanLoock, M.S., Yu, X., Yang, S., et al. (2003a) Complexes of RecA with LexA and RecX differentiate between active and inactive RecA nucleoprotein filaments, *J. Mol. Biol.* **333**, 345–354.

Venkatesh, R., Ganesh, N., Guhan, N., et al. (2002) RecX protein abrogates ATP hydrolysis and strand exchange promoted by RecA: insights into negative regulation of homologous recombination, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12091–12096.

Vierling, S., Weber, T., Wohlleben, W., et al. (2000) Transcriptional and mutational analyses of the Streptomyces lividans recX gene and its interference with RecA activity, *J. Bacteriol.* **182**, 4005–4011.

Voloshin, O.N., Ramirez, B.E., Bax, A., et al. (2001) A model for the abrogation of the SOS response by an SOS protein: a negatively charged helix in DinI mimics DNA in its interaction with RecA, *Genes. Dev.* **15**, 415–427.

Voloshin, O.N., Wang, L., Camerini-Otero, R.D. (1996) Homologous DNA pairing promoted by a 20-amino acid peptide derived from RecA, *Science* **272**, 868–872.

Voloshin, O.N., Wang, L., Camerini-Otero, R.D. (2000) The homologous pairing domain of RecA also mediates the allosteric regulation of DNA binding and ATP hydrolysis: a remarkable concentration of functional residues, *J. Mol. Biol.* **303**, 709–720.

Walker, G.C. (1984) Mutagenesis of inducible responses to deoxyribonucleic acid damage in Escherischia coli, *Microbiol. Rev.* **48**, 60–93.

Walker, J.E., Saraste, M., Runswick, M.J., et al. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945–951.

Webb, B.L., Cox, M.M., Inman, R.B. (1999) ATP hydrolysis and DNA binding by the Escherichia coli RecF protein, *J. Biol. Chem.* **274**, 15367–15374.

Witkin, E.M. (1976) Ultraviolet mutagenesis and inducible DNA repair in Escherichia coli, *Bacteriol. Rev.* **40**, 869–907.

Wu, Y., He, Y., Moya, I.A., et al. (2004) Crystal structure of archaeal recombinase RADA: a snapshot of its extended conformation, *Mol. Cells* **15**, 423–435.

Xu, L., Marians, K.J. (2002) A dynamic RecA filament permits DNA polymerase-catalyzed extension of the invading strand in recombination intermediates, *J. Biol. Chem.* **277**, 14321–14328.

Xu, L., Marians, K.J. (2003) PriA mediates DNA replication pathway choice at recombination intermediates, *Mol. Cells* **11**, 817–826.

Yasuda, T., Morimatsu, K., Horii, T., et al. (1998) Inhibition of Escherichia coli RecA coprotease activities by DinI, *EMBO J.* **17**, 3207–3216.

Yoon-Robarts, M., Blouin, A.G., Bleker, S., et al. (2004) Residues within the B' motif are critical for DNA binding by the superfamily 3 helicase Rep40 of adeno-associated virus type 2, *J. Biol. Chem.* **279**, 50472–50481.

Yoshimasu, M., Aihara, H., Ito, Y., et al. (2003) An NMR study on the interaction of Escherichia coli DinI with RecA-ssDNA complexes, *Nucleic Acids Res.* **31**, 1735–1743.

Receptor Biochemistry

Tatsuya Haga[1] *and Kimihiko Kameyama*[2]
[1] *Gakushuin University, Tokyo, Japan*
[2] *National Institute of Advanced Industrial Science and Technology, Tsukuta, Japan*

1 Classification of Receptors 554

2 Ion Channel-Coupled Receptors (ICCR) 555
2.1 Nicotinic Acetylcholine Receptors 557
2.2 Glutamate Receptors 559
2.3 $GABA_A$ Receptors 564
2.4 Other Ion Channel-Coupled Receptors 564

3 G Protein-Coupled Receptors (GPCR) 565
3.1 Structure, Function, and Species of GPCRs 567
3.2 GTP-Binding Regulatory Protein (G-Protein) 572
3.3 Classification of GPCRs and Specificity of GPCR-G Protein Interaction 574
3.4 Regulator of G-Protein Signaling (RGS) and G Protein-coupled Receptor Kinase (GRK) 575
3.5 Neurotransmitter Receptors: Roles of ICCRs and GPCRs 577

4 Protein Kinase-Coupled Receptors (PKCR) 578
4.1 Growth Factor Receptors with Tyrosine Kinase Activity 579
4.1.1 Growth Factor Receptors 579
4.1.2 Neurotrophin Receptors 582
4.1.3 Activation of the Ras-MAPK Cascade 583
4.2 Cytokine Receptors Linked with Tyrosine Kinases 585
4.2.1 Cytokine Receptors 585
4.2.2 Activation of the JAK-STAT Pathway 585
4.3 TGF-β Receptors with Serine/Threonine Kinase Activity 588
4.4 Intracellular Signaling: Common Pathways Triggered by GPCRs and PKCRs 589

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

5 Other Receptors 589

Acknowledgments 590

Bibliography 590
Books and Reviews 590
Primary Literature 591

Keywords

Agonists
Compounds that bind to and activate receptors and include endogenous ligAnds such as hormones and neurotransmitters, chemically synthesized compounds, and natural products like alkaloids.

Antagonists
Compounds that bind to but do not activate receptors, and hence inhibit the action of agonists.

Cytokines and Growth Factors
Proteins that influence proliferation, differentiation, maturation, and survival of cells. Cytokines include interleukins (IL), which are produced by leukocytes and control leukocyte function, and interferons (IFL) which are produced in response to virus infection. Growth factors are produced constitutively and act on nonhematopoietic cells, whereas cytokines are produced transiently and are mostly involved in hematopoiesis and immune or inflammatory responses. In a broader sense, cytokines include growth factors.

GRKs (G Protein–Coupled Receptor Kinases)
Protein serine/threonine kinases that phosphorylate activated forms of GPCRs (e.g. agonist-bound GPCRs and light-activated rhodopsin) specifically.

GTP-binding Regulatory Proteins (G-Proteins or Heterotrimeric G-Proteins)
Proteins that are activated by hormone- or neurotransmitter-bound receptors and activates effectors such as adenylate cyclase, phospholipase C, and ion channels. G-proteins are composed of an α-subunit with GTP-binding activity and $\beta\gamma$ subunits.

Hormones
Chemical substances that are released from one cell and modulate the function of other cells, which are usually located in remote organs.

MAP Kinase (MAPK)
Mitogen-activated protein kinase, a serine/threonine kinase in the signal transduction pathway that is triggered by activation of growth-factor receptors. Extracellular signal-regulated kinase (ERK) is another name. MAP kinase kinase (MAPKK or MEK (MAPK/ERK kinase)) phosphorylates both tyrosine and threonine residues in MAP kinase and activate MAP kinase. MAPKK kinase (MAP kinase kinase kinase, MAPKKK, or MEK kinase) phosphorylates serine residues in MAPKK and activates MAPKK.

Neurotransmitters or Transmitters
Chemical substances that are released from presynaptic terminals of an excited neuron and regulate the excitability of postsynaptic membranes of the other neuron.

Protein Serine/Threonine Kinases
Enzymes that phosphorylate serine or threonine residues in proteins, and include second messenger-activated protein kinases, such as cAMP-dependent protein kinase, protein kinase C, or Ca^{2+}/calmodulin-dependent protein kinase, and protein kinases in the signal transduction pathways such as MAP kinase.

Protein Tyrosine Kinases
Enzymes that phosphorylate tyrosine residues in proteins. Some tyrosine kinases are receptors for growth factors and are activated by binding of growth factors, whereas others are not receptors and are activated by cytokine receptors or other factors.

RGS Proteins (Regulator of G Protein–Signaling)
Proteins that accelerate GTPase activity of G-protein α-subunits and then terminate G-protein signaling.

Second Messengers
Compounds that are formed in the cells in response to the stimulation by hormones or neurotransmitters on the surface of cells (e.g. cyclic AMP, cyclic GMP, diacylglycerol, and inositol triphosphate).

SH2 Domains
The src homology domain 2 that is present in many proteins in the signal transduction pathways and interacts with domains containing phosphorylated tyrosine residues in growth-factor receptors or other proteins. A product of the src gene (sarcoma gene), src protein, contains an SH2, an SH3 (another src homology domain present in many other proteins), and a tyrosine kinase catalytic site.

Communications between cells are performed by cell–cell contact or by chemical substances like neurotransmitters, hormones, cytokines, or growth factors. These chemical substances are secreted by a cell and recognized by receptors in target cells. The binding of these ligands to their receptors initiates a series of chemical reactions. Some hormones such as steroids and thyroxin pass through cell membranes and interact with their receptors inside the cell. Most other hormones and all known neurotransmitters, cytokines, and growth factors cannot penetrate cell membranes and bind to their receptors, which are transmembrane glycoproteins, on the surface of cells. Membrane-bound receptors are classified into three major groups, ion channel-coupled receptors (ICCR), G protein–coupled receptors (GPCR), and protein kinase-coupled receptors (PKCR). In this section, we describe the molecular properties of membrane-bound receptors and intracellular signals triggered by activation of these receptors.

1 Classification of Receptors

Receptors have been named and classified by their endogenous ligands and subclassified by specific agonists and antagonists since the introduction of the receptor concept in the early 1900s by Langley. Molecular entities of receptors have been identified in the last quarter-century, which enables us to classify receptors on the basis of their structure and function. For example, receptors for acetylcholine have been classified into two types, nicotinic acetylcholine receptors and muscarinic acetylcholine receptors, on the basis of their interactions with specific agonists, nicotine and muscarine, respectively. Nicotinic and muscarinic acetylcholine receptors are further characterized by their interaction with specific antagonists, tubocurarine and atropine, respectively. It has been shown by means of purification of nicotinic and muscarinic receptors and their functional reconstitution into artificial membranes that these two receptors have completely distinct structural and functional characteristics, and that nicotinic and muscarinic receptors are, respectively, acetylcholine-gated ion channels and activators of G-proteins.

Cloning of complementary DNAs (cDNAs) encoding receptor proteins and deduction of their amino acid sequences have revealed that a number of receptors have structural and functional characteristics similar to the nicotinic receptor, constituting a superfamily of ion channel-coupled receptors (ICCRs), and a greater number of receptors have structural and functional similarity to the muscarinic receptor, constituting a superfamily of G protein–coupled receptors (GPCRs). Endogenous ligands for ICCRs are restricted to a small number of neurotransmitters. In contrast, endogenous ligands for GPCRs include almost all kinds of neurotransmitters, various hormones, autacoids, chemotactants, pheromones, and external stimulants (e.g. odorants, taste substances, and light). The structural characteristics common to all members of the two superfamilies are several hydrophobic regions, which are assumed to be transmembrane segments: GPCRs have seven such regions, and each subunit of ICCRs

has two to four. Thus, GPCRs are also referred to as seven transmembrane receptors or serpentine receptors.

Several groups of receptors for cytokines and growth factors, on the other hand, are characterized by a single transmembrane segment per subunit. Some receptors with a single transmembrane segment are tyrosine kinases with the catalytic activity to phosphorylate tyrosine residues in the receptors themselves and/or related proteins, and others do not have enzyme activity but function as activators of related tyrosine kinases. Members of another group of receptors with a single transmembrane segment have serine/threonine kinase activity. Thus, these receptors with the single transmembrane segment are somehow coupled with protein kinases and are referred to here as protein kinase–coupled receptors (PKCRs).

Table 1 summarizes the characteristics of the three major groups of membrane receptors, ICCR, GPCR, and PKCR. ICCR, GPCR, and PKCR, respectively, can be discriminated from each other by endogenous ligands (neurotransmitters, hormones/neurotransmitters, cytokines/growth factors), coupled function (ion channels, G-protein activators, protein kinases/protein kinase activators), and by the order of response rates (milliseconds, seconds, hours).

Cloning of cDNAs encoding receptor proteins and determination of the human genome sequence has revealed that multiple kinds of receptors or receptor subtypes exist for a given endogenous ligand including most neurotransmitters, hormones, and cytokines. Furthermore, a number of putative receptors are called *orphan receptors* because their cDNAs or genes have been identified on the basis of their sequence homology to those of known receptors but their endogenous ligands remain unidentified. An extreme case is the odorant receptors: approximately five hundred human receptors are believed to exist but their ligands remain to be identified except for a few receptors. One of the future challenges will be to develop specific ligands for a given receptor with a known amino acid sequence.

2 Ion Channel-Coupled Receptors (ICCR)

Endogenous ligands for ICCRs are limited to several neurotransmitters, which are acetylcholine, glutamate (and possibly aspartate), γ-aminobutyric acid (GABA), glycine, serotonin (or 5-hydroxytryptamine, 5HT), and ATP (adenosine triphosphate). ICCRs are not known for catecholamines like adrenaline and dopamine, or for peptides.

ICCRs are oligomers composed of heterogeneous subunits, although some might be homooligomers. They incorporate ion-channel function in the oligomeric structure with a transmitter-gated pore in the center of the oligomers. Receptors for acetylcholine, glutamate, serotonin, and ATP are cation channels, which are permeable to both Na^+ and K^+ (and Ca^{2+} in some cases), and receptors for GABA and glycine are anion (chloride ion) channels. Neurotransmitters are released from the presynaptic terminal and bind to ICCRs in the postsynaptic membranes. The binding leads to the opening of ion channels, followed by depolarization and excitation of the postsynaptic membranes in the case of cation channels and by hyperpolarization and inhibition of excitation in the case of anion channels. In other words, the excitatory postsynaptic potential (EPSP) and inhibitory postsynaptic potential (IPSP)

Tab. 1 Characteristics of three major groups of membrane receptors.

	Ion channel-coupled receptors (ICCR)	*G protein-coupled receptors (GPCR)*	*Protein kinase-coupled receptors (PKCR)*
Endogenous ligands	Some neurotransmitters	Neurotransmitters, hormones, autacoids, and external stimulants (odorants, taste substances, and light)	Cytokines, including growth factors, neurotrophins, and TGF-β
Structure and number of transmembrane (TM) segments	Oligomers with a pore in the center Two to four TM segments per subunit	Monomers and some dimers Seven TM segments	Dimers or oligomers Single TM segment per subunit
Immediate function of receptors induced by agonist binding	Permeation of cations or anions through pores causing depolarization or hyperpolarization of membranes	Activation of G-proteins by facilitating GTP–GDP exchange; activation of G protein-coupled receptor kinases (GRKs)	Phosphorylation of tyrosine or serine/threonine residues in receptors or activation of tyrosine kinases
Cellular responses induced by activation of receptors	Excitation of neurons and other excitable cells or inhibition of excitation	Changes in excitability of neurons or other cells: regulation of protein function and gene expression	Regulation of gene expression, cell proliferation, and differentiation
Timescale of responses	Milliseconds	Seconds	Hours

are caused by opening of transmitter-gated cation and anion channels, respectively. The time course of these responses is rapid (within the order of milliseconds). Thus, ICCRs are specified for the purpose of rapid communication in the nervous system. This does not necessarily mean, however, that the role of ICCRs is restricted to the production of EPSP and IPSP. ICCRs may also take part in the stimulation or inhibition of neurotransmitter release through presynaptic receptors, in the long-term facilitation or depression of EPSP through a subtype of glutamate receptor (NMDA receptor), or in the action on glial cells.

2.1 Nicotinic Acetylcholine Receptors

The nicotinic acetylcholine receptor is the most extensively studied and best characterized among all the receptors. Abundant nicotinic receptors are present in the electric organs of the electric eel or *torpedo*, which are homologous to skeletal muscles. Nicotinic receptors were purified from the electric organs of *torpedo* by taking advantage of their abundance and their specific interaction with snake venoms such as bungarotoxin. Purified nicotinic receptors were found to be pentameric glycoproteins composed of four distinct subunits ($2 \times \alpha$, β, γ, δ) and to show acetylcholine-gated cation channel activity when reconstituted in lipid membranes. In 1982, cDNA encoding the α-subunit of nicotinic receptors in the electric organ of *torpedo* was cloned by using its partial amino acid sequence. Thereafter, cDNAs encoding each subunit of nicotinic receptors in the *torpedo* electric organ and in the calf muscle were cloned by using the sequence homology among subunits. Amino acid sequences of each subunit were deduced from corresponding nucleotide sequences of cDNAs. The finding of homology in amino acid sequences among the four subunits indicates derivation from a common ancestor.

Each subunit is composed of 437–501 amino acid residues and has four hydrophobic domains (M1–M4) with approximately 20 amino acid residues each. Extensive studies of nicotinic receptors have been carried out by using biochemical methods such as crosslinking of receptors with acetylcholine analogs or channel blockers, gene technology (e.g. mutagenesis), and physical methods (e.g. electron microscopy and electron diffraction of receptor microcrystals). Results are summarized in the following list, and an atomic model is shown in Fig. 1(A).

1. The amino-terminal portions of each subunit are located extracellularly and are bound with sugars. Two molecules of acetylcholine bind to the amino-terminal portions of the α-subunits.
2. The four hydrophobic domains (M1–M4) of each subunit form α-helices and transverse the plasma membrane forming a pore (ion channel) in the middle of 20 transmembrane segments.
3. The five M2 segments line the pore. Leucine and valine residues in the middle of the five M2 segments are conserved among all four subunits and thought to constitute the gate of the pore.
4. The binding of acetylcholine causes a small rotation of the extracellular parts of the five subunits, which triggers a conformational change in the M2 segments and moves the hydrophobic side chains away from the ion path. Thus, these hydrophobic residues function as a gate (Fig. 1(B)).

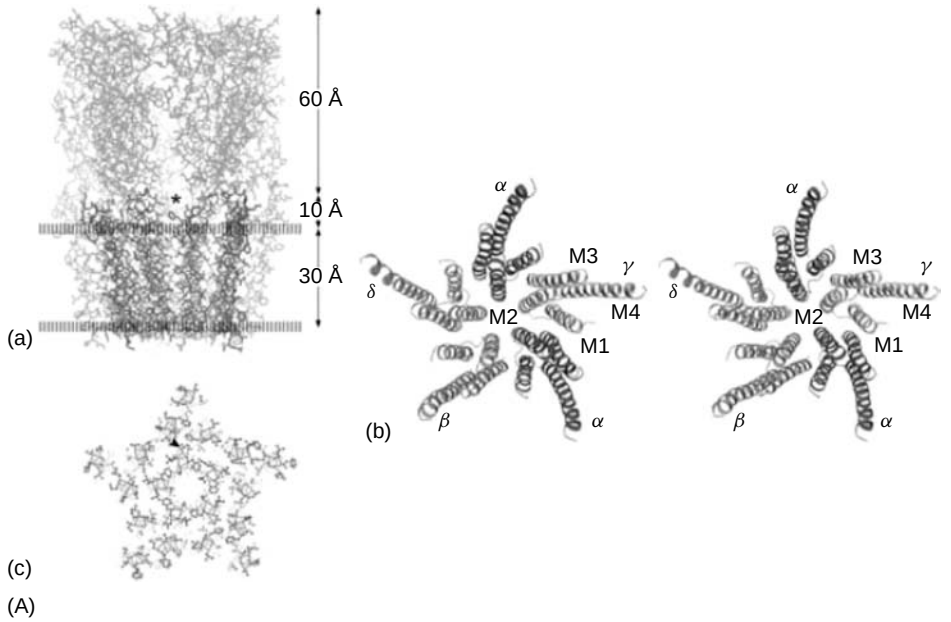

Fig. 1 Structural model of the nicotinic acetylcholine receptor (nAChR: an ion channel-coupled receptor, ICCR). (A) An atomic model with the closed pore was determined from electron images by electron cryo-microscopy of crystalline postsynaptic membranes and their reconstruction by optical diffraction and computation. The side (a) and cross-sectional (c) views show an extracellular ligand-binding domain (green), an inner, and the transmembrane segments composed of pore-facing M2 helices (blue) and lipid-facing M1, M3, M4 helices (red). (b) Stereo view of the pore, as seen from the extracellular face. (c) Cross-sectional view through the pentamer at the middle of the membrane. The pore is shaped by an inner ring of five α helices (M2), which curve radially to create a tapering path for ions, and an outer ring of 15 α-helices (M1, M3, M4), which coil around each other and shield the inner ring from the lipid layer. (B) A model for the gating mechanism. With binding of acetylcholine (ACh), two M2 α-helices of α-subunits rotate to opposite directions. This movement triggers the concerted movement of M2 α-helices of the other subunits resulting in the opening of the gate that consists of hydrophobic residues in the central part of five M2 segments. (From Miyazawa, A. et al. (2003) Structure and gating mechanism of the acetylcholine receptor pore, *Nature* **424**, 949–955) (see color plate p. xli).

5. The rate of channel opening is fast with a time constant of approximately 20 µs. The conductance of a single channel is 30 ps ($= 30 \times 10^{-12}$ A V^{-1}). This means that 2×10^4 cation molecules pass through the channel in an open time of 1 ms when the cation is derived by a potential of 100 mV: 30×10^{-12} (A V^{-1}) $\times$ 0.1 (V) $\times$ 1×10^{-3} (s) $\times$ 1/96 500 (mol C^{-1}) $\times$ 6.2×10^{23} (molecules mol^{-1}) $= 1.9 \times 10^4$ molecules.
6. The channel is permeable to both Na^+ and K^+ but not to Ca^{2+} or to anions. Anionic charges on both sides of the M2 segments contribute to the permeation of cations but not of anions.
7. In the continued presence of agonists, the channel closes with time constant of 50 to 100 ms at 20 °C (desensitization).

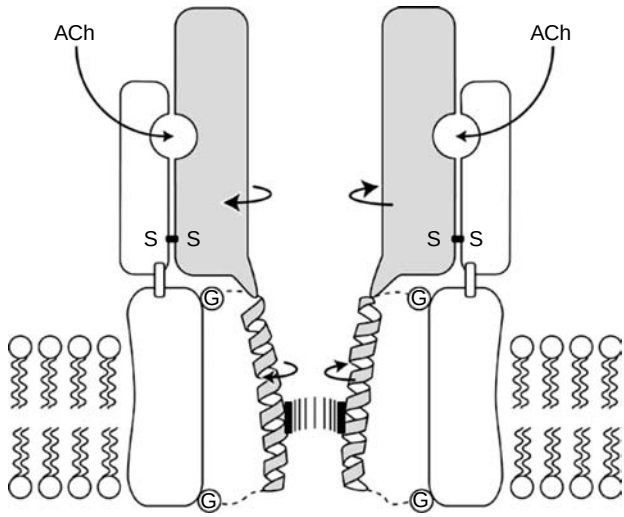

(B)

Fig. 1 (*Continued*)

Phosphorylation of nicotinic receptors by cAMP-dependent protein kinase, protein kinase C, or tyrosine kinase facilitates desensitization.

There are two kinds of nicotinic receptors in the skeletal muscle: the embryonic type ($\alpha_2\beta\gamma\delta$) and the adult type ($\alpha_2\beta\varepsilon\delta$). The denervation of skeletal muscle induces the expression of the γ-subunit and decreases the expression of the ε-subunit, which results in the conversion from the adult type to the embryonic type. The embryonic-type receptor is distributed widely in the whole area of skeletal muscle membranes, whereas the adult-type receptor is localized in the postsynaptic membranes. This specific distribution is regulated by distinct expression of *nAChR* genes in the nucleus adjacent to the synapse.

Different kinds of nicotinic receptor subtypes are present in the autonomous ganglia and the brain. Seven kinds of α-subunits (α_2–α_9, no α_8 in human), which are homologous to the skeletal α-subunit (α_1), and three kinds of β-subunits (β_2–β_4) have been identified (see Table 2). Functional receptors are thought to be pentamers like $(\alpha_4)_2(\beta_3)_3$. The effect of bungarotoxin on brain nicotinic receptors is weak or absent, except that α_7 and α_9 bind the toxin. Some brain nicotinic receptors are present in presynaptic membranes and regulate the release of glutamate. Ca^{2+} is reported to permeate through the α_7 homooligomer, while Ca^{2+} does not permeate through other nicotinic receptors including the skeletal muscle type.

2.2 Glutamate Receptors

Glutamate receptors are classified into ionotropic receptors (ion channel-coupled receptors, ICCRs) and metabotropic receptors (G protein–coupled receptors, GPCRs). Ionotropic receptors are divided by the use of a specific agonist, N-methyl D-aspartic acid (NMDA), into NMDA and non-NMDA types. Non-NMDA types are

Tab. 2 Subunits, genes, and gene loci of human ion channel-coupled receptors (ICCRs).

Subunit	*Gene*	*Gene locus*	*Subunit*	*Gene*	*Gene locus*	*Subunit*	*Gene*	*Gene locus*
Nicotinic acetylcholine receptor			**$GABA_A$ receptor**			**Glutamate receptor (AMPA Type)**		
nAChRα1	*CHRNA1*	2q24-q32	$GABA_A$Rα1	*GABRA1*	5q34–q35	GluR1	*GRIA1*	5q31.1
nAChRα2	*CHRNA2*	8q21	$GABA_A$Rα2	*GABRA2*	4p12	GluR2	*GRIA2*	4q32-q33
nAChRα3	*CHRNA3*	15q24	$GABA_A$Rα3	*GABRA3*	Xq28	GluR3	*GRIA3*	Xq25-q26
nAChRα4	*CHRNA4*	20q13.2-q13.3	$GABA_A$Rα4	*GABRA4*	4p12	GluR4	*GRIA4*	11q22
nAChRα5	*CHRNA5*	15q24	$GABA_A$Rα5	*GABRA5*	15q11.2-q12	**(Kainate type)**		
nAChRα6	*CHRNA6*	8p11.21	$GABA_A$Rα6	*GABRA6*	5q34	GluR5	*GRIK1*	21q22.11
nAChRα7	*CHRNA7*	15q14	$GABA_A$Rβ1	*GABRB1*	4p12	GluR6	*GRIK2*	6q16.3-q21
nAChRα9	*CHRNA9*	4p14	$GABA_A$Rβ2	*GABRB2*	5q34	GluR7	*GRIK3*	1p34-p33
nAChRα10	*CHRNA10*	11p15.5	$GABA_A$Rβ3	*GABRB3*	15q11.2-q12	KA1	*GRIK4*	11q22.3
nAChRβ1	*CHRNB1*	17p13.1	$GABA_A$Rδ	*GABRD*	1p36.3	KA2	*GRIK5*	19q13.2
nAChRβ2	*CHRNB2*	1q21.3	$GABA_A$Rε	*GABRE*	Xq28	**(NMDA type)**		
nAChRβ3	*CHRNB3*	8p11.2	$GABA_A$Rγ1	*GABRG1*	4p12	NR1	*GRIN1*	9q34.3
nAChRβ4	*CHRNB4*	15q24	$GABA_A$Rγ2	*GABRG2*	5q31.1-q33.1	NR2A	*GRIN2A*	16p13.2
nAChRδ	*CHRND*	2q33-q34	$GABA_A$Rγ3	*GABRG3*	15q11-q13	NR2B	*GRIN2B*	12p12
nAChRε	*CHRNE*	17p13-p12	$GABA_A$Rπ	*GABRP*	5q33-q34	NR2C	*GRIN2C*	17q25

nAChRγ	*CHRNG*	2q33–34	$GABA_A R\theta$	*GABRQ*	Xq28	NR2D	*GRIN2D*	19q13.1-qter
			$GABA_A R\rho 1$	*GABRR1*	6q13-q16.3	NR3A	*GRIN3A*	9q31.1
			$GABA_A R\rho 2$	*GABRR2*	6q13-q16.3	NR3B	*GRIN3B*	19p13.3
Glutamate receptor (δ type)			**Glycine receptor**			**Purinergic receptor (P2X type)**		
GluRδ1	*GRID1*	10q22	GlyRα1	*GLRA1*	5q32	P2X1	*P2RX1*	17p13.3
GluRδ2	*GRID2*	4q22	GlyRα2	*GLRA2*	Xp22.1-p21.3	P2X2	*P2RX2*	12
			GlyRα3	*GLRA3*	4q33-q34	P2X3	*P2RX3*	4q33-q34
Serotonin receptor ($5\text{-}HT_3$ type)			GlyRβ	*GLRB*	4q31.3	P2X4	*P2RX4*	4q31.3
						P2X5	*P2RX5*	17p13.3
5HTR3A	*HTR3A*	11q23.1-q23.2				P2X7	*P2RX7*	12q24
5HTR3B	*HTR3B*	11q23.1				P2X1like	*P2RXL1*	22q11.21

Receptors are classified by their neurotransmitters and types, which are written in **bold** face. Most commonly used names are listed for receptors and receptor subunits. Genes for receptor subunits are expressed by names approved by the gene nomenclature committee in HUGO (Human Genome Organization). Gene loci are mapped on chromosomes of the human genome.

further subdivided into AMPA (α-amino-3-hydroxy-5-methyl-4-isoxalone propionic acid) and kainic acid (KA) types, where AMPA and KA are specific agonists for the respective subtypes. A number of different genes have been identified for subunits of each glutamate receptor (seven genes for NMDA receptor subunits, four for AMPA receptors, five for KA receptors)(see Table 2). There are also two genes encoding δ-type glutamate receptors, which have significant homology to other ionotropic glutamate receptors but whose functions remain to be identified. Some of the genes encoding glutamate receptor subunits are reported to have splicing variants: for example, eight kinds of polypeptides derived from the NMDA receptor 1 (*NMDAR1*) gene are expressed in the brain. Each glutamate receptor is believed to be a heterotetramer composed of the same type of receptor subunits.

The molecular properties of glutamate receptors are not as well understood as those of the nicotinic receptors, although the fundamental properties are assumed to be similar. Each subunit of AMPA, KA, and NMDA receptors has four hydrophobic domains (M1–M4), like the nicotinic receptors and other ICCRs, and there are some similarities in these domains between glutamate receptors and other ICCRs. Each glutamate receptor subunit, however, is thought to have three, rather than four, transmembrane domains, where M1, M3, and M4 but not M2 are the transmembrane segments. In addition, glutamate receptors are thought to be composed of four, but not five, subunits in contrast with nicotinic receptors. Thus, the extracellular part of glutamate receptors is thought to be composed of four amino-terminal parts and four loops between the M3 and M4 transmembrane segments, and to constitute the glutamate-binding site. Actually the protein complex corresponding to the extracellular domains of an AMPA receptor has been expressed, purified, crystallized, and used for X-ray analysis. The X-ray analysis showed an atomic structure composed of two domains with a clam-shaped structure and with a glutamate-binding site in the cleft of the clam shell. Four clams with eight halves of shells are assembled, and the binding of multiple glutamate molecules is supposed to cause a conformational change so as to close the shells, which should be somehow transmitted to the transmembrane domains. Recently, an atomic structure of the extracellular domain of an NMDA receptor subunit, NR1, was determined and shown to be clam shaped, as seen with the AMPA receptor. NMDA receptors have to bind both glutamate and glycine (or D-serine) to be activated, where the NR1 subunit binds to glycine (or D-serine) and the NR2 subunit to glutamate. The cleft of the clamshells in NR1 was shown to be open in the presence of antagonist, and closed in the presence of glycine or D-serine.

The M2 segment is thought to be inserted into the membrane from the intracellular side forming a loop. A certain amino acid residue in the M2 segment is critical for the selectivity of permeable ions. The genetic code for the amino acid is asparagine for NMDA receptors and glutamine for AMPA receptors. This site, however, is converted to arginine by RNA editing for the GluR2 subunit of AMPA receptors. NMDA receptors are permeable to Na^+, K^+, and Ca^{2+} ions. AMPA receptors comprised of the subunits with glutamine in this site are permeable to Na^+, K^+, and also Ca^{2+} ions with inwardly rectified voltage dependency in the current. On the other hand, AMPA receptors including the edited GluR2 subunit with the arginine residue are not permeable to Ca^{2+}

ions, and are permeable only to Na^+ and K^+ ions with linear voltage dependency in the current. The ratio between edited and nonedited GluR2 mRNA increases after birth, and almost 100% of GluR2 mRNA is edited in the adult brain.

Non-NMDA glutamate receptors play a major role in the excitatory synaptic transmission in the central nervous system. NMDA receptors are different from non-NMDA receptors in that NMDA receptors are usually not active, as the ion permeability of NMDA receptors is inhibited by Mg^{2+} ions. The inhibitory effect of Mg^{2+} is suppressed by depolarization. The NMDA receptor channel opening is also regulated by glycine or D-serine. Thus, the NMDA receptor opens only when both glutamate and glycine (or D-serine) are present and the cell membrane containing NMDA receptors has been depolarized.

Phenomena referred to as long-term potentiation (LTP) and long-term depression (LTD) have been studied in detail, particularly in the hippocampus, and are considered to represent cellular models for memory formation. LTP is defined as a sustained increase in the efficiency of synaptic transmission, and is caused by brief trains of high-frequency stimulation (e.g. 30–100 stimuli/s with duration of 5–10 s, that is, consecutive 150–1000 stimuli). On the other hand, LTD is defined as a sustained decrease in the efficiency of synaptic transmission, and is caused by low-frequency stimulation (1–3 stimuli/s with duration of 5–15 min, that is, consecutive 300–1000 stimuli). NMDA receptors are involved in both LTP and LTD in the hippocampal CA1 area. High-frequency stimuli induce depolarization of postsynaptic neurons through the AMPA receptor, which is followed by opening of NMDA receptors, Ca^{2+} entry, and activation of Ca^{2+}-calmodulin-dependent protein kinase II (CaMKII). Autophosphorylation of CaMKII makes it constitutively active, and the persistent active form of CaMKII binds to the NMDA receptor NR2B subunit in postsynaptic membranes. High-frequency stimulation causes the translocation of the AMPA receptor containing the GluR1 subunit from intracellular organelles into synaptic membranes, which results in increases of the excitatory postsynaptic current. CaMKII activation and AMPA receptor translocation are thought to be somehow related to the formation of LTP. On the other hand, low-frequency stimuli induce mild activation of NMDA receptors followed by low-level elevation of Ca^{2+} concentrations, which causes the activation of protein phosphatases. Phosphatase activity induces the internalization of AMPA receptors into intracellular vesicle pools, which decreases the excitatory postsynaptic current. Thus, LTP and LTD, which are measured as the increase and decrease in the excitatory postsynaptic current, respectively, are linked with recruiting and internalization of the AMPA receptor on the synaptic surface and also with the activation of CaMKII and phosphatases, but the exact molecular mechanisms are yet to be identified.

Glutamate receptors are highly enriched in postsynaptic membranes and are anchored by so-called scaffolding proteins. Excitatory synapses in the central nervous system are characterized by the electron dense regions in postsynaptic sites, which are known as postsynaptic density (PSD). One major component in postsynaptic densities is PSD-95, which binds to NMDA receptor NR2 subunits and some voltage-gated potassium channels. The carboxy termini of the latter proteins have the consensus sequence of Thr/Ser-X-Val (X is any amino acid residue), to which the

so-called PDZ domain in PSD-95 binds. As there are three PDZ domains in a single PSD-95 molecule, coexpression of PSD-95 with NMDA receptors in cultured cells induces their clustering. Proteins referred to as GRIP (glutamate receptor interacting protein) also have PDZ domains and bind to AMPA receptors. These molecules function as scaffolding proteins contributing to assembling receptors, ion channels, and enzymes in postsynaptic sites. These molecules bind directly or indirectly to motor proteins of the kinesin family, which are supposed to transport receptor molecules from the cell body to the dendrite.

2.3 $GABA_A$ Receptors

There are two kinds of GABA receptors, $GABA_A$ and $GABA_B$, which are ion channel-coupled receptors (ICCR) and G protein–coupled receptors (GPCR), respectively. Eighteen kinds of $GABA_A$ receptor subunits have been identified in the human genome (six α, three β, three γ, one ε, one δ, one π, one θ, two ρ). Various oligomers with different combinations of these subunits are thought to exist *in situ*. The fundamental structure of $GABA_A$ receptors is thought to be similar to that of nicotinic receptors, that is, a pentamer of subunits with four transmembrane segments. $GABA_A$ receptors, however, are anion channels, in contrast with nicotinic receptors, and there are more positive amino acid residues on both sides of the M2 segments of $GABA_A$ receptor subunits compared with the M2 segments of nicotinic receptors. These positive charges are considered to help attract chloride ions before their permeation through the pores. Binding of GABA to $GABA_A$ receptors and the subsequent opening of chloride channels causes the hyperpolarization of postsynaptic membranes and thereby the suppression of excitation.

$GABA_A$ receptors take a major role in the inhibitory synaptic transmission in the brain. Different kinds of drugs are known to interact with $GABA_A$ receptors. Benzodiazepines and related compounds are tranquilizers and facilitate the action of GABA by increasing the affinity of the receptor for GABA. Barbiturates, which induce sleep, also increase the action of GABA. Picrotoxin, on the other hand, blocks the channel and then acts as an inducer of seizures. The effect of drugs appears to be subtype-specific. $GABA_A$ receptors that do not contain γ-subunits are not affected by benzodiazepines. This suggests the possibility that various drugs with different subtype specificity might be developed in the future and be used for different kinds of neurological or psychiatric diseases.

2.4 Other Ion Channel-Coupled Receptors

Glycine receptors are chloride channels and play a role in the inhibitory synaptic transmission mostly in the spinal cord. There are four subunit genes in human (three α, one β). Glycine receptors are thought to have a structure similar to nicotinic receptors and $GABA_A$ receptors.

A number of receptors for serotonin (or 5-hydroxytryptamine, 5HT) are GPCRs, and only serotonin $5HT_3$ receptors are cation channels. The $5HT_3$ receptors are expressed in forebrain, hippocampus, brainstem, and spinal cord, although the expression is restricted to some populations of neurons. At first only one gene corresponding to $5HT_{3A}$ was identified, and the channel properties of the $5HT_{3A}$ homopentamer were found to be different from those of native

brain $5HT_3$ receptors. After homology search of the human genome sequence, another gene corresponding to $5HT_{3B}$ was found, which had significant homology to the *5HT*$_{3A}$ gene. When both $5HT_{3A}$ and $5HT_{3B}$ cDNAs were coexpressed, the expressed receptor showed channel properties similar to those of native brain receptors.

Purinergic receptors are divided into two groups, P2X receptors and P2Y receptors, which are ion channel-coupled receptors (ICCR) and G protein–coupled receptors (GPCR), respectively. There are at least seven different subunit genes for P2X receptors in the human genome. These subunits are clearly distinct from subunits of either nicotinic or glutamate receptors in that P2X subunits have only two transmembrane domains. The loop between the two transmembrane domains is located extracellularly and is assumed to be the ATP-binding domain because it has an ATP-binding motif and ATP acts on the channel from the extracellular phase. P2X receptors appear to form either homo- or heterooligomers, and some receptors are suggested to be trimers although this remains to be proven in a more direct manner. P2X receptors are ATP-gated cation channels with equal permeability to Na^+ and K^+ and a significant permeability to Ca^{2+}. P2X receptors are expressed in post- and presynaptic membranes in the central nervous system, and may contribute to ATP-driven excitatory postsynaptic current (EPSC) although the ATP-driven EPSC is relatively small and less abundant as compared with that driven by glutamate receptors. P2X3 is expressed in some sensory neurons, and is thought to serve as a pain sensor. Recently, P2X4 in microglia is reported to play an important role in evoking and maintaining a kind of neuropathic pain.

3
G Protein-Coupled Receptors (GPCR)

Endogenous ligands of GPCRs include all neurotransmitters except glycine, most hormones and autacoids, several chemotactic factors, and external stimulants such as pheromones, odorants, and taste substances. Chemically, these ligands are diverse and include amino acids, peptides, proteins, amines, nucleosides, nucleotides, lipids such as prostaglandins and leukotrienes, Ca^{2+} ion, H^+ ion, and various organic compounds acting as odorants. Rhodopsin and optic pigments, which are receptors for light, are also GPCRs (Tables 3 and 4). The number of GPCR gene species in the human genome is estimated to be 800 to 900, approximately half of which are odorant receptors. Three to four hundred GPCRs are known to have endogenous ligands, but more than one hundred GPCRs are considered to be orphan receptors for which the endogenous ligands remain to be identified.

Each GPCR was thought to be a homogeneous protein and to exist as a monomer. However, recently $GABA_B$ and metabotropic glutamate receptors (mGluRs) have been shown to exist as dimers. Some GPCRs in the rhodopsin family, which is the major group of GPCRs, are also reported to form homo- or heterooligomers with intact functions. However, it is not clear yet whether oligomerization is a prerequisite for the function of most GPCRs or whether it is required for only a small number of GPCRs.

The function of GPCRs is to activate heterotrimeric G-proteins ($\alpha_{GDP}\beta\gamma$), that is,

to stimulate the dissociation of GDP from G-proteins and thereby facilitate the binding of GTP to G-proteins followed by dissociation into α_{GTP} and $\beta\gamma$ subunits (see Fig. 2). Activated G-proteins, that is α_{GTP} and $\beta\gamma$ subunits, stimulate or inhibit enzymes that synthesize or break down second messengers such as cAMP, diacylglycerol and inositol 1,4,5 triphosphate (IP_3). Thus, the information of the presence of hormones, neurotransmitters, odorants or other molecules outside of cells is converted to changes in concentrations of second messengers, which affect activities of protein serine/threonine kinases; for example, cAMP and diacylglycerol activate cAMP-dependent protein kinase (A kinase) and protein kinase C (C kinase), respectively. In addition, IP_3 interacts with and opens the IP_3-gated Ca^{2+}-channel in intracellular vesicles causing an increase in intracellular Ca^{2+} concentrations, which may activate Ca^{2+}-calmodulin-dependent protein kinases. These protein kinases affect activities of different kinds of proteins including enzymes, ion channels, and so on. In addition, expression of some genes is regulated through these protein kinases; for example, CREB (cAMP responsive element (CRE) binding protein) is phosphorylated by cAMP-dependent protein kinase. Phosphorylated CREB, together with CREB binding protein, stimulates the transcription of genes that reside downstream of CRE with a sequence of TGACGTCA.

As for the relation between GPCR and ion channels, there are different types depending on the species of ion channels; some ion channels like the inward-rectified K^+ channel in the heart are directly regulated by G-protein $\beta\gamma$ subunits, cation channels in visual or olfactory cells are regulated by cGMP or cAMP, and many other ion channels are thought to be directly or indirectly regulated by phosphorylation by second messenger-dependent protein kinases. In any case, the time courses of their opening or closing after ligand binding to the receptor are much

Fig. 2 Structures of rhodopsin and a heterotrimeric G-protein. (A) An atomic model for rhodopsin was obtained by X-ray analysis of the rhodopsin crystal. (a) A stereoview in parallel to the plane of the membrane. The upper and lower parts of this model correspond to the intra- and extracellular domains respectively. There are seven transmembrane α-helixes (I-VII), another α-helix in the C-terminal region (VIII), 11-cis-retinal (yellow) in the transmembrane region, and two β-sheets (1–2 and 3–4) and sugars in the N-terminal region. View of the receptor from the cytoplasmic (b) and intradiscal (extracellular) side (c) of the membrane. (From Palczewski, K. et al. (2000) Crystal structure of rhodopsin: A G protein-coupled receptor, *Science* **289**, 739–745). (B) Receptor-G protein interaction. (a) A structural model for the cytoplasmic surface of the M_1 muscarinic receptor. Residues important for interaction with G protein are shown in red, magenta (i3 loop), and yellow. The arrows indicates the outward movement of TMb and increased mobility of TM7 and helix 8, which are thought to accompany receptor activation. (b) A structural model for the interaction between G protein-coupled receptor (GPCR) and G-protein $\alpha\beta\gamma$ trimer. Structures are taken from crystal structures of rhodopsin and a chimera of transducin (Gt) and Gi. The N- and C-termini of the transducin α-subunit and the α4-β6 loop sequences, which are thought to be involved in the interaction with GPCR, are shown in red. The switch-II helix in the G-protein α-subunit is bound to the β-subunit when the α-subunit is bound to GDP, and is subject to conformational change and is released from the β-subunit when the α-subunit is bound with GTP. (From Lu, Z.L. et al. (2002) Seven-transmembrane receptors: crystals clarify, *Trends Pharmacol. Sci.* **23**, 140–146) (see color plate p. xlii).

slower (10 ms–1 min) as compared to IC-CRs (ms).

3.1
Structure, Function, and Species of GPCRs

The amino-terminal domain of GPCRs is considered to be located extracellularly, and the carboxy-terminal tail in the cytoplasm. This assumption is consistent with the facts that the consensus sequence for attachment of sugars (Asn-X-Ser/Thr) is present in the amino-terminal tail for almost all members of this family and that phosphorylation sites are in the carboxy-terminal tail for some GPCRs. The amino and carboxy-terminal tails have a variable length ranging from several tens to several hundreds of amino acid residues. Seven hydrophobic regions (20–25 amino acid residues each) are assumed to transverse the plasma membrane forming α-helices. The three intracellular and extracellular loops are generally short (10–30 amino acid residues) except that some receptors have a long third intracellular loop. Recently, the crystal structure of rhodopsin was determined at atomic resolution (see Fig. 2). This structure confirmed the above assumption

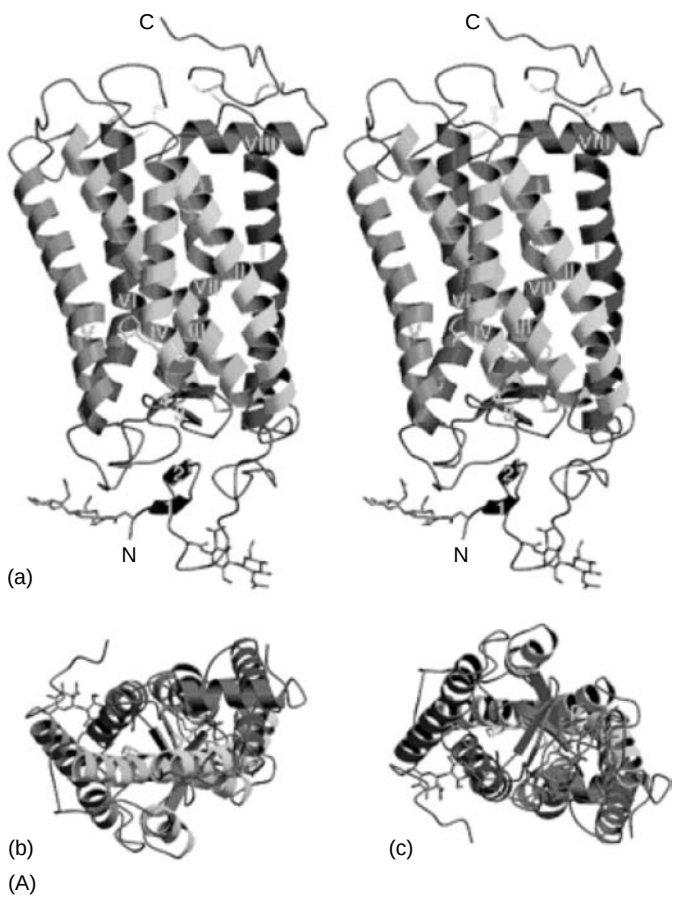

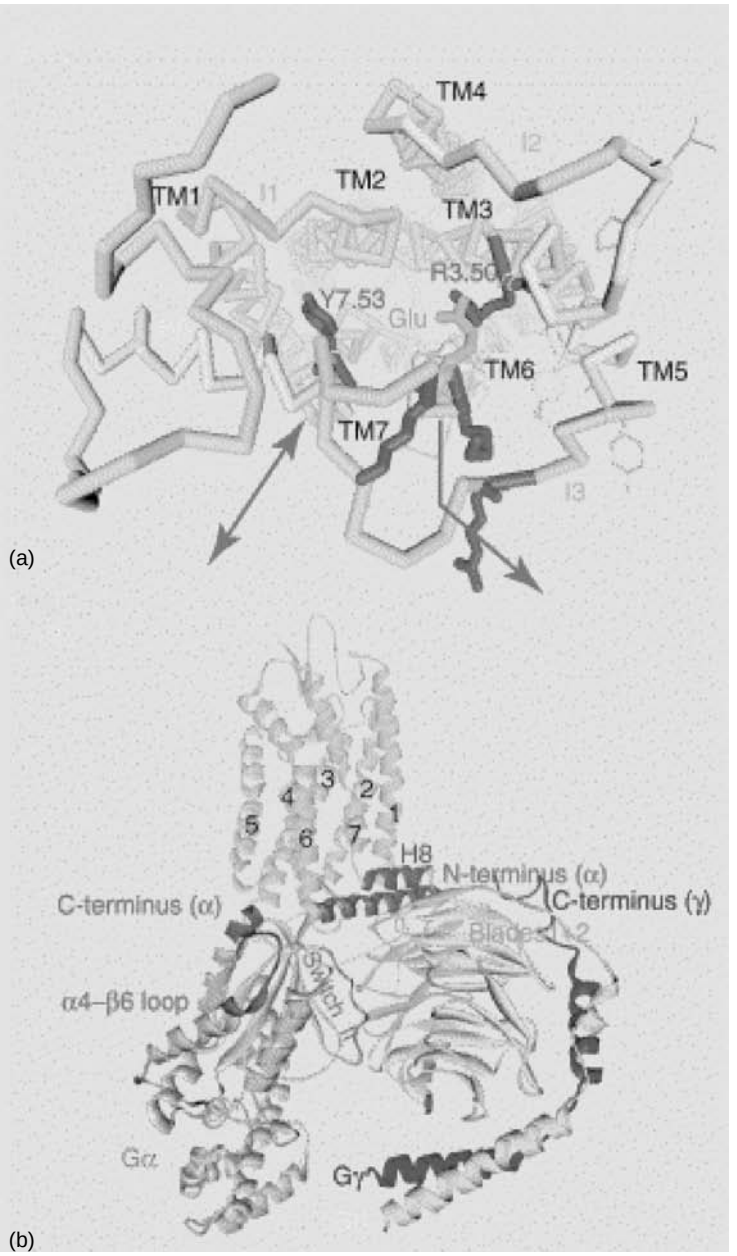

(B)

Fig. 2 (*Continued*)

Tab. 3 Neurotransmitters and receptors.

Neurotransmitters	***Ion channel-coupled receptors (ICCR) (Species of subunits)***	***G protein–coupled receptors (GPCR) (Species of coupled G proteins)***
Acetylcholine	Nicotinic Muscle type($\alpha 1_2 \beta 1 \gamma \delta$ or $\alpha 1_2 \beta 1 \varepsilon \delta$) Neuronal type (pentamer of α_{2-9} and β_{2-4})	Muscarinic $M_{1,3,5}$ (G_q) $M_{2,4}$ (G_i/G_o)
Glutamate	Ionotropic AMPA (GluR1-4) Kainate (GluR5-7, KA1,2) NMDA(NMDAR1, 2A-2D, 3A-3B)	Metabotropic mGluR1,5 (G_q) mGluR2-4, mGluR6-8 (G_i/G_o)
Serotonin (5HT)	$5HT_3$ ($5HT_{3A,3B}$)	$5HT_{2A-2C}$ (G_q) $5HT_{1A,1B,1D-1F}$ (G_i/G_o) $5HT_{4,6,7}$ (G_s)
ATP	$P2X_{1-5,7}$	$P2Y_1$ (G_i/G_o) $P2Y_2$ (G_s)
GABA	$GABA_A$ (six α, four β, three γ, one δ, one ε, one π, one θ, two ρ)	$GABA_B$ ($GABA_BR_1$, $GABA_BR_2$) (G_i/G_o)
Glycine	Glycine(three α, one β)	
Noradrenaline (Adrenaline)		$\alpha_{1A,1B,1D}$ (G_q) $\alpha_{2A,2B,2C}$ (G_i/G_o) $\beta_{1,2,3}$ (G_s)
Dopamine		D_{2-4} (G_i/G_o) D1, D_5 (G_s)
Histamine		H_1 (G_q) H_2 (G_s) H_3 (G_i/G_o)
Opioid		μ, δ, κ, (G_i/G_o)
Tachykinin		NK1–3 (G_q)
Adenosine		A_{1-3} (G_i/G_o) $A_{2A,2B}$ (G_s)

and demonstrated that rhodopsin actually has seven transmembrane α-helices with amino- and carboxy-terminal tails in the outside and inside of the cells respectively. In addition, it was revealed that rhodopsin has an eighth α-helix between the seventh transmembrane α-helix and the palmitoylated cysteine residues in the carboxy terminus, and a disulfide bond between cysteine residues in the third transmembrane segment and in the loop between the fourth and fifth transmembrane segments.

On the basis of sequence similarities, mammalian GPCRs are divided into three major groups. Group A (rhodopsin family) is the largest one and includes rhodopsin and odorant receptors, receptors for neurotransmitters like acetylcholine and other amines, and receptors for peptides and other ligands such as nucleosides, nucleotides, and lipids. Approximately 20 amino acid residues in the transmembrane segments are highly conserved among members of this group, and most members have cysteine residue(s) at conserved

Tab. 4 G protein-coupled receptors (GPCRs) for hormones, exogenous stimulants, and other ligands except typical neurotransmitters.[a]

Ligands	*Receptor subtypes*	*G-proteins*	*Effectors*	*Second messengers and their downstreams*
Group 1[b]				
CGRP (calcitonin gene-related peptide)	CGRP	G_s	Adenylate cyclase activation	cAMP increase Activation of cAMP-dependent protein kinase
Vasopressin/oxytocin	V2			
Prostanoid	DP, IP, EP2			
Glucagon				
Thyrotropin				
Lutropin-gonadotropin				
Odorants	300–500 species	G_{olf}		
Group 2				
Cannabinoid	CB1	G_{i1}, G_{i2}, G_{i3}	Adenylate cyclase inhibition	cAMP decrease
Somatostatin	SSTR1,2	G_o		
VIP (Vasoactive intestinal peptide)	VIP1,2,3		K^+ channel open Ca^{2+} channel close	Hyperpolarization Ca^{2+} decrease
Chemokines	MIPIa, MCP1, IL8A,B			
Neuropeptide Y	Y1			
Light	Rhodopsin cone pigments (blue, red, green)	G_{t1} G_{t2}	cGMP – phosphodiesterase activation	cGMP decrease Hyperpolarization
Taste	20–30 species	G_g	cAMP – phosphodiesterase activation	cAMP Decrease Depolarization

Group 3				
PAF (platelet-activating factor)		G_q G_{11}	Phosphatidylinositol-specific phospholipase C β types	Increase in diacylglycerol (DG) and IP_3
Thromboxane		G_{14}		
Prostanoid	FP, TP, EP1,3	G_{15}		Activation of protein kinase C
Bombesin	B1,2			
Endothelin	ET(A), (B)			Increase of Ca^{2+}, and activation of Ca^{2+}-calmodulin kinase
Angiotensin	AT1			
Vasopressin/oxytocin	V1A,1B, OT			
Cholecystokinin/gastrin	CCK(A),(B)			
Bradykinin	BB1,2			
Group 4				
Thrombin		G_{12}	ρ GEF activation	Activation of ρ and then ρ-kinase
Thromboxan		G_{13}		Reorganization of cytoskeleton and change in cell motility

[a]GPCRs for typical neurotransmitters are shown in Table 3. It is difficult in some cases to distinguish between neurotransmitters and nonneurotransmitters (hormones, autacoids, or neuromodulators), and the distribution of some compounds between Table 3 and Table 4 is more or less arbitrary. Noradrenaline and adrenaline may be taken as both neurotransmitters and hormones, and serotonin; histamine, and adenosine are both neurotransmitters and autacoids. Somatostatin, VIP, and other peptides are known to be released from neurons and to act on neurons although they are listed in Table 4.

[b]Receptors are grouped by their G-protein coupling properties.

positions in the carboxy termini, homologous to the ones that are palmitoylated in rhodopsin. Group B (secretin receptor family) consists of receptors for rather large peptides like secretin, glucagon, and so on, and these receptors have a large amino-terminal domain with conserved disulfide bonds. Group C (glutamate receptor family) consists of metabotropic glutamate receptors (mGluR), $GABA_B$ receptors, and putative taste receptors. These receptors have large amino- and carboxy-terminal domains and short and highly conserved third intracellular loops. Cysteine residues in the extracellular loops two and three, which are known to form a disulfide bond in rhodopsin, are conserved in most GPCRs. The ligand-binding pocket is formed by transmembrane domains in the case of Group A receptors. On the other hand, Group B and C receptors contain the ligand-binding domain in the large amino-terminal segment. The glutamate-binding domain of mGluRs is homologous to the amino acid binding domain of bacterial periplasmic binding proteins.

GPCRs were believed to exist as monomers, but various lines of evidence have indicated that class C receptors form dimers and function as dimers. An atomic structure of the ligand-binding domain of the mGluR1 indicated that the ligand-binding domain consists of a homodimer, each subunit of which has two globular domains with a cleft for the ligand. In the case of the $GABA_B$ receptor, two kinds of $GABA_B$ receptors, $GABA_BR1$ and $GABA_BR2$, were shown to have a higher affinity for GABA when they were coexpressed in cultured cells than when each receptor was expressed separately, indicating the formation of a heterooligomer. Native $GABA_B$ receptors are thought to be heterooligomers because they have high affinity for GABA, and $GABA_BR1$ and $GABA_BR2$ are coimmunoprecipitated from extracts solubilized from brain membranes. Taste receptors with a long amino terminus were also shown to have a high affinity for amino acids or nucleosides when heterooligomers were formed. A number of examples were reported indicating that class A receptors also may form homo- or heterooligomers and that the formation of oligomers may affect the affinity and specificity for ligands and receptor trafficking. Oligomerization with components other than GPCRs was also reported: for example, calcitonin receptor-like receptor interacts with a protein with a single transmembrane segment called *RAMP* (receptor activity–modifying protein), and the dopamine D5 receptor interacts with the $GABA_A$ receptor. It is clear that some class A receptors may form and function as oligomers, but it remains to be clarified whether their oligomerization is a general phenomenon under physiological conditions.

The binding sites in GPCRs for G-proteins are thought to be the second and third intracellular loops and the carboxy-terminal tail. Following ligand binding, GPCRs are supposed to undergo conformational changes in the ligand-binding site at first and then, secondarily, in the G protein–binding sites. The molecular details involved in these conformational changes, however, remain to be clarified although extensive modeling studies have been carried out with the rhodopsin system.

3.2 GTP-Binding Regulatory Protein (G-Protein)

G-proteins are composed of three subunits, $\alpha\beta\gamma$, the molecular weights of which are approximately 40, 35, and 5–10 kDa,

respectively. The crystal structure of the $\alpha\beta\gamma$ trimer was determined (Fig. 2). G-proteins are not transmembrane proteins but are bound to cell membranes through prenyl groups and fatty acids attached to the carboxy terminus of the γ-subunit and the amino terminus of the α-subunit, respectively. The α-subunit has a binding site for GTP or GDP and harbors GTPase activity. The β-subunit has seven repeated domains forming a tertiary structure like a seven-bladed propeller, and the γ-subunit is tightly bound with the amino-terminal part of the β-subunit. G-proteins bound with GTP dissociate into α_{GTP} and $\beta\gamma$ subunits, and the α_{GTP} and $\beta\gamma$ subunits activate or inhibit effectors. The activity of the α_{GTP} and $\beta\gamma$ subunits is terminated by hydrolysis of GTP and the subsequent formation of $\alpha_{GDP}\beta\gamma$ trimer. The $\alpha_{GDP}\beta\gamma$ trimer is converted into the α_{GTP} and $\beta\gamma$ subunits by action of the agonist-bound receptor. The agonist-bound receptor (aR) facilitates release of GDP from G-proteins and forms a complex with guanine nucleotide-free G-proteins (aRG complex), as evidenced by reconstituting purified receptors and G-proteins in artificial membranes. The aRG complex is thought to be formed as a transient intermediate *in vivo*, forming the transition state. The activation energy of the reaction path $G_{GDP}-G-G_{GTP}$ is considered to be reduced by taking the route of $G_{GDP}-aRG-G_{GTP}$, as is the case for the formation of enzyme–substrate complexes. Actually, aR activates multiple G-proteins like a catalyst.

Seventeen different α-subunit genes have been identified, and they are classified into four groups according to the homology of their amino acid sequences: G_s, G_i/G_o, G_q, and G_{12} (see Table 3). Multiple species of β (5) and γ (11) subunits have been identified, although their functional differences remain to be clarified. Several kinds of $\beta\gamma$ subunits are shown to interact equally well with both G_s-α and G_i-α, and several combinations of β- and γ-subunits have been expressed in cultured cells and shown to form a complex with each other and with G_i-α subunits. The specificity of interaction of the three subunits, however, has not been fully elucidated.

G_s-α (or α_s) that is bound with GTP activates adenylate cyclase. G_{olf}-α is an olfactory bulb-specific α-subunit and activates adenylate cyclase as G_s-α does. At least nine different adenylate cyclases have been cloned and all of them are activated by G_s-α. The addition of $\beta\gamma$ subunits in the presence of G_s-α further stimulates adenylate cyclase II but inhibits adenylate cyclase I. The $\beta\gamma$ subunit alone neither stimulates nor inhibits adenylate cyclase I or II. Ca^{2+}-calmodulin further stimulates adenylate cyclase I in the presence of G_s-α, whereas Ca^{2+} inhibits adenylate cyclase V and VI. These synergistic mechanisms of stimulation or inhibition may be utilized for detection of coincident stimulation of a single neuron by one receptor linked to G_s and the other receptor linked to elevation of $\beta\gamma$ subunits or Ca^{2+}.

Most α-subunits in the second group of G-proteins (G_{i1}-α, G_{i2}-α, G_{i3}-α, G_o-α, G_{t1}-α, G_{t2}-α, and G_g-α) are ADP-ribosylated by pertussis toxin. G-proteins in the other groups and G_z in this group are not ADP-ribosylated by pertussis toxin. The ADP-ribosylated G-proteins lose the ability to interact with receptors, and hence the inhibition by pertussis toxin of receptor-mediated reactions indicates the involvement of these G-proteins. G_i-α and G_z-α inhibit adenylate cyclase I, V, and VI, and G_o-α inhibits adenylate cyclase I. The amounts of G-proteins present in the mammalian brain show the following order: $G_o > G_{i1} > G_{i2}, G_s >$

G_q. This indicates that the major source of G-protein $\beta\gamma$ subunits in the brain is from members of this second group. Thus, activation of G-proteins in this group causes the stimulation of adenylate cyclase II through $\beta\gamma$ subunits and inhibition of adenylate cyclase I through both α- and $\beta\gamma$ subunits. In addition, the $\beta\gamma$ subunit directly interacts with and opens an inward-rectified K^+ channel (GIRK or Kir3). The K^+ channel in the heart is regulated by acetylcholine released from the vagus nerve through muscarinic M_2 receptors and G_i, probably G_{i2}. Similar inward-rectified K^+ channels are known to be present in the brain. A certain kind of voltage-gated Ca^{2+} channel is known to be closed by G-proteins belonging to this group through direct interaction between channels and G-protein α-subunits.

G_t (or transducin) is a retina-specific G-protein, and G_{t1}-α and G_{t2}-α are present in rod and cone cells, respectively. G_{t1} and G_{t2} are activated by light-absorbed rhodopsin and cone pigments, respectively. G_t-α bound with GTP activates cGMP-phosphodiesterase resulting in a decrease of cGMP concentrations. As cGMP interacts with and opens cation channels and depolarizes visual cells, light stimulation causes hyperpolarization of visual cells. G_g (or gustducin) is a taste bud–specific and transducin-like G-protein, and is activated by receptors for taste substances and activates cAMP phosphodiesterase.

G_q-α and related α-subunits (α_{11}, α_{14}, α_{15} (or α_{16})) activate phosphatidylinositol-specific phospholipase C β types (PLCβ), which catalyze the formation of inositol 1,4,5 triphosphate (IP_3) and diacylglycerol (DG) from phosphatidyl inositol 4,5-bisphosphate. The GTPase activity of G_q-α, G_{11}-α, or G_{14}-α is stimulated by PLCβ, which shortens the lifetime of the active α_{GTP} forms. There are several species of PLCβ; PLCβ2 and PLCβ3 are stimulated by G-protein $\beta\gamma$ subunits, whereas PLCβ1 is either inhibited or stimulated by $\beta\gamma$ subunits. Stimulatory effects of $\beta\gamma$ subunits are observed even in the absence of G_q-α, in contrast with the fact that the stimulation by $\beta\gamma$ subunits of adenylate cyclase II requires the presence of G_s-α. Stimulation of phosphatidyl inositol turnover by G_i- or G_o-coupled receptors may be due to activation of PLCβ2 or PLCβ3 by G-protein $\beta\gamma$ subunits.

G_{12} and G_{13} constitute a different group and are present ubiquitously. Receptors for thrombin and thromboxan A2 are coupled with $G_{12/13}$, and the GTP-bound form of $G_{12/13}$-α activates ρ GEF (a guanine nucleotide exchange factor of the small GTP-binding protein ρ). The activated GTP-bound form of ρ activates ρ-kinase resulting in the rearrangement of the actin cytoskeleton. Thus, G_{12} appears to be involved in cell motility.

3.3 Classification of GPCRs and Specificity of GPCR-G Protein Interaction

GPCRs may be classified according to the homology of amino acid sequences, the kind of endogenous ligands, or the species of G-protein α-subunits with which they interact. Neurotransmitter receptors are summarized in Table 3, where GPCRs are classified by their ligands. Other GPCRs are classified by G-protein interactions and are shown in Table 4. The best-characterized receptors, which interact with G_s-, G_i/G_o-, and G_q-type G-proteins, are β-adrenergic receptors, M_2 muscarinic and α_2 adrenergic receptors, and M_1 muscarinic and α_1 adrenergic receptors, respectively. There are at least three

subtypes for each of the β, α_2, and α_1 adrenergic receptors (Table 3). Thus, adrenaline or noradrenaline may activate at least three different G-proteins through nine different receptors. The function of adrenaline (noradrenaline) on a certain cell will depend on species of receptors expressed in the cell. In the case of acetylcholine, there are five different muscarinic acetylcholine receptors; M_2 and M_4 muscarinic receptors are coupled with G_i/G_o family G-proteins, and M_1, M_3, and M_5 receptors with G_q family G-proteins.

The specificity of interaction between receptor subtypes and distinct G-proteins is different depending on receptor and G-protein species. In reconstituted systems of purified receptors and G-proteins, the M_2 muscarinic receptor interacts equally well with G_{i1}, G_{i2}, and G_o and the M_1 muscarinic receptor interacts equally well with G_q, G_{11}, G_{14}, but the M_2 receptor does not interact with G_q family G-proteins nor does the M_1 receptor couple to G_i/G_o family G-proteins. Some receptors show a stricter specificity for members of the G_i/G_o family than the M_2 receptor; the α_{2A} adrenergic and A_1 adenosine receptors show higher affinity for G_{i3} than for G_{i1}, G_{i2}, and G_o, and the D_{2L} dopamine receptor shows a higher affinity for G_{i2} than for other G_i/G_o proteins. On the other hand, some receptors including the turkey β-adrenergic and the substance P (NK1 tachykinin) receptors are known to interact with both G_s and G_q family G-proteins. Among 146 GPCRs examined, 36 (25%), 76 (52%), and 51 (35%) interact with G_s, G_i/G_o, and G_q respectively, and 14 GPCRs interact with more than two kinds of G-proteins: 1, 3, 7, and 3 interact with G_s/G_i, G_s/G_q, G_i/G_q, $G_s/G_i/G_q$, respectively (cf. The IUPHAR compendium of receptor characterization and classification, IUPHAR Media, 1998). In the interactions of muscarinic receptors and G-proteins, amino acid sequences (approximately 20 amino acids) in the N- and C-terminal portions of the third intracellular loop in muscarinic receptors are critical for determining the specificity of their interaction (M_1, M_3, M_5-G_q and M_2, M_4-G_i/G_o). On the other hand, alternatively spliced forms of prostaglandin EP_3 receptors, which have different carboxy-terminal tails with distinct amino acid sequences, interact with different G-proteins including G_i/G_o, G_s, and G_q. As for G-proteins, the last 10 amino acids of the α-subunits are reported to be most critical for determining the specificity for receptors. $G_{q/i}$-α, which is a chimera of G_q-α and the carboxy-terminal tail of G_i-α, interacts with G_i-coupled receptors as G_i-α does, and then activates PLCβ1 as G_q-α does. G_{16} belongs to the G_q family but interacts promiscuously with G_i-, G_s-, and G_q-coupled receptors. The specificity or promiscuity of $G_{q/i}$ or G_{16} is not general, since each of them does not interact with some G_i-coupled receptors.

3.4 Regulator of G-Protein Signaling (RGS) and G Protein-coupled Receptor Kinase (GRK)

Signaling from GPCRs to G-proteins is regulated by regulators of G-protein signaling (RGS) and G protein-coupled receptor kinases (GRK). RGS proteins are GTPase-activators of G-protein α-subunits and accelerate the inactivation of G-protein. GRKs are serine/threonine kinases, which phosphorylate activated forms of GPCRs specifically, and initiate the attenuation of GPCR-mediated responses, which is called *desensitization*.

At least 19 different RGS members have been identified, which are subdivided into several groups such as R4 (RGS1, 2, 4, 5, 8, 13, 16), R3 (RGS3), RGSZ (RGS17, 19, 20), R7 (RGS6, 7, 9, 11), and R12 (RGS10, 12, 14). All members have RGS domains, which are responsible for the activation of GTPase of G-protein α-subunits. The specificity for G-proteins depends on the species: for example, R4 and R3 RGS act on both G_i-α and G_q-α: RGS20 on G_z-α and RGS17, 19 on G_i-α; RGS12, 14 on G_i-α. R7 RGS proteins have a G-protein γ-subunit-like domain, besides an RGS domain, and bind to the G-protein β_5 subunit. The RGS–Gβ_5 complex acts on $G_o\alpha$. RGS9 is expressed in the retina and the RGS9–Gβ_5 complex acts on the transducin α-subunit (G_t-α). The specificity for G_t-α appears to be due to the binding of Gβ_5, because the RGS domain of RGS9 can bind to any α_{GTP} subunits of G_t, G_i, or G_q and activate their GTPase activity. RGS9 is anchored to photoreceptor membranes by the transmembrane protein R9AP. Patients with recessive mutations in the RGS9 or R9AP genes are reported to have difficulty in adapting to sudden changes in luminance levels. RGS proteins for G_s were missing until recently, but RGS-PX1 was found to stimulate the GTPase activity of α-subunit of G_s but not other G-proteins. Some RhoGEFs have an RGS-like domain, which functions as GTPase activator of α-subunits of G_{12} or G_{13}. In the case of p115-Rho-GEF, not only RGS-like domains but also other regions are required for full GTPase activation, although most other RGS proteins have essentially the same GTPase-activating activity as the RGS domain alone.

There are other subfamilies called *RGS-like proteins* containing RGS-like domains, which share significant similarities with RGS domains, although some of the RGS-like domains do not have GTPase-activating activity. RGS and RGS-like proteins have other domain(s) that interact with other proteins including GPCRs. Some RGS proteins of the R4 family directly interact with specific receptors via their amino-terminal domain. RGS1, 4, and 16 attenuate the IP_3 response mediated by M_3 muscarinic receptors in pancreatic acinar cells more effectively than that mediated by cholecystokinin receptor, whereas RGS2 attenuates both signals equally. RGS3 has several splicing variants, and the difference in their amino-terminal sequences specifies their target proteins. RGSL interacts with $\beta\gamma$ subunits suppressing their stimulating effect on PLCβ. RGS3T binds directly to adenylate cyclase inhibiting its activity, and has a nuclear localization signal, which enables it to translocate into the nucleus and induce apoptosis of the cell. PDZ-RGS3 has PDZ domains, which bind to the ephrin-B receptor and attenuate signaling through GTPase activation. One RGS12 splicing variant also has PDZ domain, which binds to the interleukin 8 (IL8) receptor and terminates the receptor signal specifically.

GPCRs such as rhodopsin, β-adrenergic, α_2 adrenergic, M_2 muscarinic, and other receptors are known to be phosphorylated by G protein–coupled receptor kinases (GRKs) in a light- or agonist-dependent manner. There are six kinds of GRKs. GRK1 is rhodopsin kinase and phosphorylates light-absorbed rhodopsin. Phosphorylated rhodopsin interacts with a protein called *arrestin*, and this interaction prevents rhodopsin from interacting with and activating G-proteins. The phosphorylation of rhodopsin by rhodopsin kinase is inhibited by a calmodulin-like protein called *s-modulin* or *recoverin* in a Ca^{2+}-dependent manner. The phosphorylation

of rhodopsin by rhodopsin kinase and its inhibition by s-modulin are thought to be linked to desensitization and light adaptation, respectively.

GRK2 and GRK3 are originally known as β-adrenergic receptor kinases or βARK1 and βARK2. There are $\beta\gamma$-binding sites in the carboxy terminus of GRK2 and GRK3 but not in other GRKs. The binding of $\beta\gamma$ subunits to GRK2 or 3 facilitates their translocation from the cytoplasm to the plasma membrane. GRK2 is synergistically stimulated by G-protein $\beta\gamma$ subunits and agonist-bound receptors, providing an explanation why GRK2 phosphorylates agonist-bound receptors specifically. The phosphorylation of β-adrenergic receptors by GRK2 causes uncoupling of β-adrenergic receptors from G-proteins through action of β-arrestin, which is a protein similar to arrestin and binds to phosphorylated receptors. The agonist-dependent phosphorylation of muscarinic receptors and other GPCRs by GRK2 and other GRKs facilitates their internalization into intracellular vesicles. GRK2 phosphorylation sites and sites responsible for receptor internalization reside in the carboxy terminus for β-adrenergic receptors and in the central part of the third intracellular loop for muscarinic receptors. Agonist-dependent internalization may occur through either dynamin-dependent or -independent processes depending on the species of receptors or receptor subtypes involved.

GRK2 and 3 have an RGS-like domain in their amino termini, which binds to G_q-α and suppresses its activity. The RGS-like domain has GTPase-activating activity at G_q-α although the activity is much less compared with other RGS proteins. Thus, signaling by G_q-coupled receptors may be attenuated by GRK2 in two parallel ways, that is, the phosphorylation of the receptor followed by uncoupling from G-protein and receptor internalization, and the suppression of G_q-α activity by GRK2 binding. The activity of some G_q-coupled receptors is attenuated by GRK2 although they are not substrates of GRK2.

3.5 Neurotransmitter Receptors: Roles of ICCRs and GPCRs

Neurotransmitters may be classified based on the type of receptors they interact with (cf. Table 3). Glutamate, acetylcholine, serotonin, and ATP constitute the first family, and there are both ICCRs with cation channels and GPCRs for these neurotransmitters. The fast excitatory synaptic transmission is mostly mediated by nicotinic acetylcholine receptors in the peripheral nervous system and by AMPA-type glutamate receptors in the central nervous system. GABA and glycine form the second group of neurotransmitters and participate in fast inhibitory synaptic transmission through their ICCRs with anion channels. The GPCRs for GABA are pharmacologically identified as $GABA_B$ receptors. $GABA_B$ receptors have characteristics similar to metabotropic glutamate receptors. Glycine is a unique neurotransmitter that does not seem to act on GPCRs. Amines like dopamine, noradrenaline, and histamine, peptides like opioids and tachykinins, and adenosine constitute the third group and only act on GPCRs. Some peptides like VIP and neuropeptide Y, which are listed in Table 4, are also released from neurons and act on GPCRs on the post- or presynaptic membranes. In that sense, these substances could be considered neurotransmitters, although they are usually not classified as typical neurotransmitters.

ICCRs are specified for rapid information transfer in the nervous system. GPCRs, on the other hand, are present ubiquitously in all tissues and in different species from yeasts to human beings, and the rate of signal transduction is not so fast as observed with ICCRs (Table 1). Thus, neurotransmitters that only act on GPCRs should participate only in slow synaptic transmission. The question is raised what are the functions of GPCRs for neurotransmitters that act on both ICCRs and GPCRs. Neurotransmitters released at time zero may depolarize or hyperpolarize postsynaptic membranes through ICCRs in less than a millisecond and may also activate various events through GPCRs more than 10 to 100 ms later. The fast signal transmission mediated through activation of ICCRs by neurotransmitters cannot be affected by activation of GPCRs by the same neurotransmitters released at the same time. GPCRs can be considered monitors of synaptic activities at time zero and affect synaptic activities later; for example, modification of subsequent generation of action potentials through slow excitatory or inhibitory postsynaptic potentials in a timescale of around 100 ms, modification of subsequent release of neurotransmitters through presynaptic receptors, or modification of excitability in a longer timescale by phosphorylation of ion channels or other proteins or by modification of gene expression. It will be reasonable to assume that GPCR and subsequent signal transduction are involved in LTP and LTD as well as in learning and memory formation. Actually, various lines of experimental evidence support the idea that GPCR-mediated processes are involved in the long-term regulation of brain function.

All receptors for dopamine, noradrenaline, and orexin are GPCRs, and those for acetylcholine and serotonin are also mostly GPCRs in the brain. Neurons containing these neurotransmitters are localized in specific regions of the brainstem or basal forebrain, and send axons to many regions of the brain. These neurons are thought to regulate various general activities that do not need fast information transfer, such as emotion, sleep/wakefulness, depression/excitation, or attention.

GPCRs are also involved in the recognition of external stimuli such as light, odor, or taste, although it will take at least 10 ms or may be even 100 ms or 1 s. Speed of information transfer may be critical in some cases such as a batter who is going to hit a ball thrown at a speed of 150 km h^{-1} ($= 40$ m $s^{-1} = 4$ cm m s^{-1}) or an antelope who smells a panther dashing up at a speed of 150 km h^{-1}. The delay of 10 or 100 ms corresponds to 40 cm to 4 m of movement of a ball or a panther. It is interesting to note that GPCRs have survived as the cell sensor in spite of their slow speed and ICCR-type of sensors have not appeared in the process of evolution.

4
Protein Kinase-Coupled Receptors (PKCR)

The proteins that control the proliferation, survival, maturation, differentiation, and function of cells in specific tissues or organs are grouped together as cytokines and growth factors. Cytokines include interleukin (IL) and interferons (IFN), and in the broader definition, growth factors are also considered cytokines. Different tissues produce distinct growth factors such as epidermal growth factor (EGF), platelet-derived growth factor (PDGF), hepatocyte growth factor (HGF), nerve growth factor (NGF), and so on. Both

cytokines/growth factors and hormones are defined as molecular signals that are released from one cell and regulate the function of other cells. However, they are distinct in several respects:

1. Hormones generally act on cells in tissues remote from their tissue of origin, whereas cytokines/growth factors exert major actions within the tissues that produce them.
2. Hormones are amines, lipids, peptides, and also proteins, whereas cytokines/growth factors are proteins.
3. Receptors for most hormones, which do not penetrate cell membranes, are GPCRs, whereas receptors for cytokines/growth factors are directly or indirectly linked to activation of tyrosine kinases.
4. In general, activation of GPCRs causes metabolic changes in target cells through activation or inhibition of enzymes by protein serine/threonine phosphorylation, whereas activation of tyrosine kinases through cytokine/growth-factor receptors leads to expression of specific genes and proliferation or differentiation of target cells. However, there are cases in which gene expression and metabolic functions are regulated by hormones and cytokine/growth factors, respectively.

Insulin is a hormone in the sense of (1) but a growth factor in the sense of (3). Receptors for insulin and insulin-like growth factor (IGF-1) are tyrosine kinases. Insulin itself has an activity to stimulate proliferation of cells, and insulin-like growth factor (IGF-1) has an even stronger activity. On the other hand, the receptor for interleukin 8 is a GPCR despite "rule" (3).

Receptors for growth factors/cytokines share the following characteristics:

1. With some exceptions (e.g. the α-subunit of the insulin receptor), they have a single transmembrane segment per subunit (Fig. 3).
2. The binding of agonist (protein ligand) to the extracellular domain of receptors induces the dimerization or oligomerization of receptors.
3. Dimerization or oligomerization results in receptor activation.
4. Receptor activation is directly or indirectly linked to activation of tyrosine kinases.
5. The activation of tyrosine kinases finally leads to the activation or inhibition of cell-specific transcription factors and the expression of specific genes.

Receptors for growth factors/cytokines are divided into three subgroups according to their function (Table 5). Group 1 receptors, which include receptors for growth factors like insulin, EGF, PDGF, FGF, and neurotrophins have a catalytic domain of tyrosine kinase in their intracellular domain. Group 2 receptors, which include receptors for most cytokines except growth factors and transforming growth-factor β (TGFβ), do not have catalytic domains but interact with and activate tyrosine kinases. Group 3 receptors, which include receptors for TGFβ, have a catalytic domain of serine/threonine kinase in their intracellular domain.

4.1 Growth Factor Receptors with Tyrosine Kinase Activity

4.1.1 Growth Factor Receptors

Receptors for insulin, IGF-1 and HGF are composed of two subunits, α and β. After removal of the precursor signal peptide, the insulin and HGF receptors (c-met protooncogene) are posttranslationally cleaved into two chains (α and β) that

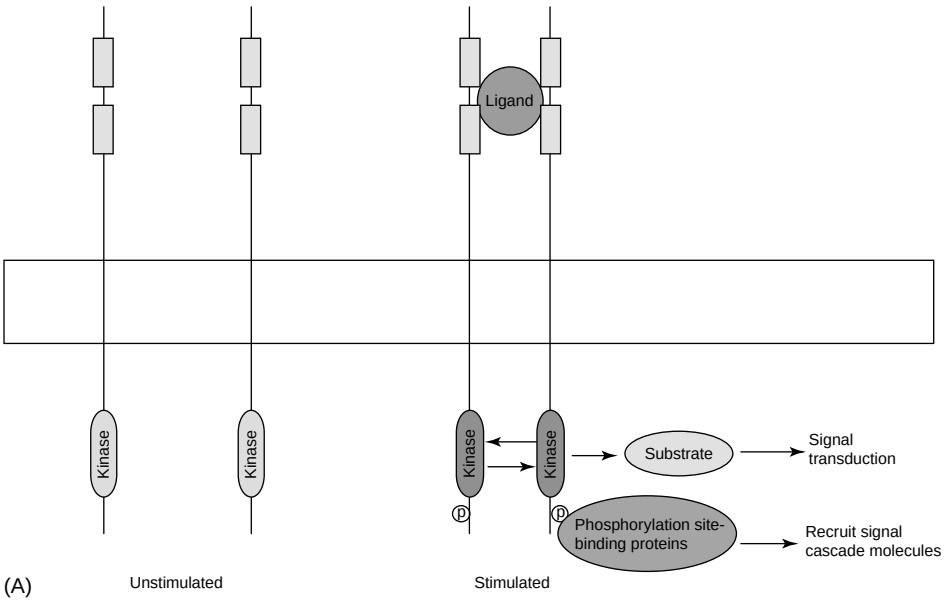

Fig. 3 Interaction of growth factors and growth-factor receptors. (A) A general model for the growth-factor receptor and other PKCRs with a single transmembrane segment. Following ligand binding, most receptors of this family form dimers (or trimers or tetramers in some cases). In many cases, ligands are dimers and then ligand binding automatically induces the oligomerization of receptors, as shown here. Dimerization of receptors triggers the signal transduction cascade, starting with autophosphorylation of a receptor subtype by the other and *vice versa*, followed by binding to phosphorylated tyrosine residues of signaling molecules, which are bound by other molecules. (B) Structure of the complex of epidermal growth factor (EGF) with the extracellular domain of EGF receptor. Binding of two EGF molecules induces the direct interaction of two EGF receptor molecules without direct interaction of EGF molecules. Two EGF molecules are observed on opposite sites. This is in contrast with the model shown in A regarding the mode of dimerization, although the outcome (receptor dimerization) is the same in both cases. (a) the ribbon diagram, (b) the top view of (a), (c) a surface model corresponding to (a). (From Ogiso, H. et al. (2002) Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains, *Cell* **110**, 775–787).

are covalently linked. The α-subunit is in the extracellular space and has a ligand-binding domain, whereas the β-subunit has a single transmembrane domain and a tyrosine kinase catalytic domain in the intracellular space. Receptors for insulin and IGF-1 are composed of two α- and two β-subunits, which are linked to each other by S–S bonds between the α- and β-subunits and between the two β-subunits. Thus, the receptors for insulin and IGF-1 are tetramers even in the absence of ligands. Hence, ligand binding does not induce the dimerization of receptors, in contrast with other growth-factor receptors. The binding of insulin or IGF-1 is thought to cause conformational changes in their receptors, which lead to their activation. A major substrate for tyrosine kinase of activated receptors is insulin receptor substrate (IRS). There are four kinds of IRSs in the human genome. The expression levels of these molecules differ

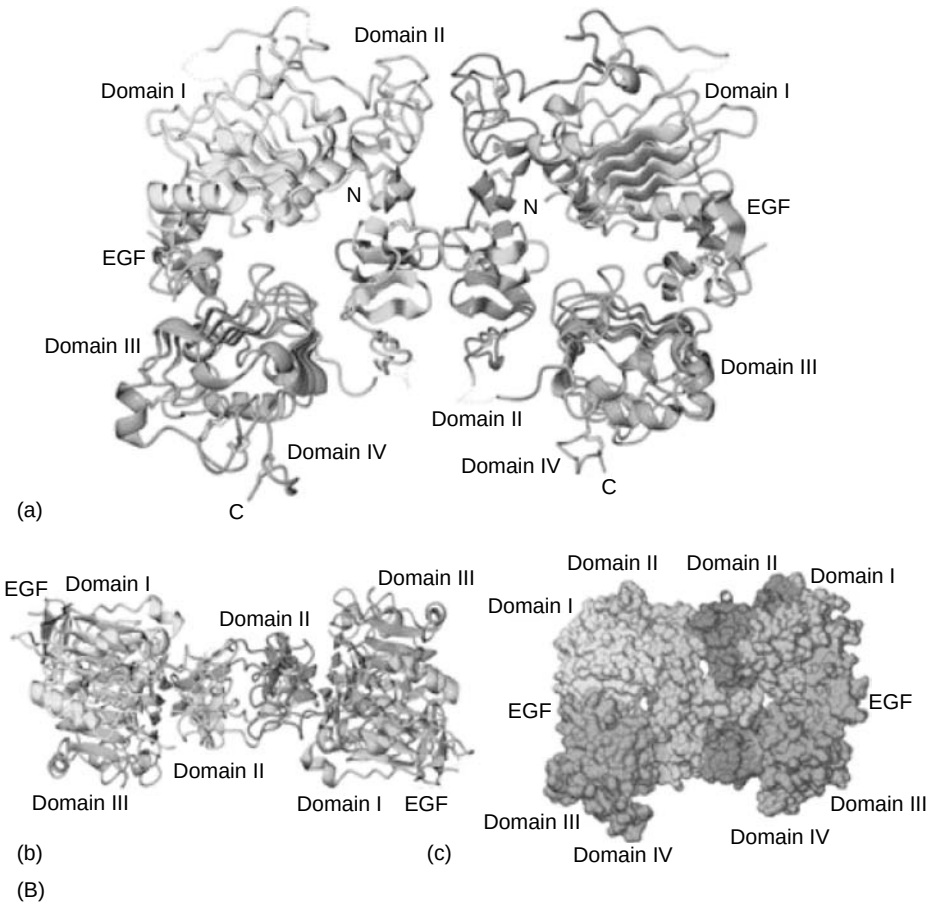

Fig. 3 (*Continued*)

from one tissue to the other, explaining the different effects of insulin in different tissues. IRS proteins have PH and PTB domains, and function as adaptor molecules. Following phosphorylation of specific tyrosine residues, IRS proteins recruit SH2 domain-containing proteins to the membrane. One of the molecules that binds to IRSs is phosphatidyl inositol 3 kinase, which is one of the key mediators in intracellular signal cascades.

Each receptor for PDGF, EGF, and FGF is a membrane protein with an extracellular ligand-binding domain, a transmembrane domain, and an intracellular tyrosine kinase domain. The molecular weights of the EGF and PDGF receptors are 120 to 130 kDa for the protein portions and 170 to 180 kDa when sugar is added. There are several characteristic domains in the extracellular part, such as cysteine-rich and immunoglobulin (Ig)-like domains. Receptors for EGF, PDGF, FGF have two cysteine-rich domains, five Ig-like domains, and two to three Ig-like domains, respectively. The cysteine-rich domains in EGF receptors and the Ig-like domains in PDGF and FGH receptors are

known to be involved in the binding of their respective growth factors.

PDGF is a dimer composed of A- and B-subunits and there are three kinds of dimers, AA, AB, and BB. There are two kinds of PDGF receptors, α- and β-receptors. The A-subunit binds only with the α-receptor, while the B-subunit binds with either α- or β-receptor. Thus, PDGF-AA induces $\alpha 2$ receptor homodimerization, PDGF-AB leads to $\alpha 2$ and $\alpha\beta$ homo- or heterodimerization, and PDGF-BB induces the dimerization of three different forms $\alpha 2$, $\beta 2$ or $\alpha\beta$. Receptor dimerization results in cross phosphorylation of tyrosine residues in one receptor subunit by the other receptor subunit and *vice versa*. Three different dimers may allow the binding of different kinds of proteins to their phosphorylated regions, which may lead to different kinds of cellular responses. Recently, other kinds of PDGF named PDGF C and D have been reported. They induce homodimers, and PDGF-CC interacts with $\alpha 2$ receptors, and PDGF-DD with $\beta 2$ (and possibly $\alpha\beta$) receptors.

EGF is a monomer, and two molecules of EGF bind to two molecules of EGF receptors (EGFR) causing formation of an EGF–EGFR–EGFR–EGF complex, where the direct interaction between two EGFRs is induced by binding of EGF but there is no direct interaction between the two EGF molecules. This is in contrast with the dimerization of NGF receptors, which is based on the direct interaction of two NGF molecules within the dimer. The EGF receptor also forms heterodimers with ErbB family receptors.

There are 23 members of the FGF family in the human genome and some of them have alternative splicing variants. There are four distinct genes for FGF receptors, *FGFR1–4*, and alternative splicing variants. These receptors form homo- and heterodimers when bound to FGFs, and intermolecular transphosphorylation occurs. FGF mediates either proliferation and/or differentiation; for example, FGF causes proliferation of NIH3T3 cells, neurite outgrowth of pheochromocytoma-derived cells (PC12 cells), and mesoderm induction in *Xenopus* oocytes. These responses appear to be mediated by a common pathway involving activation of a small GTP-binding protein, ras, but it remains unknown how these responses are differentially induced in different cells.

4.1.2 Neurotrophin Receptors

Neurotrophins are survival and differentiation factors of neurons. NGF was characterized as a survival factor for sympathetic and sensory neurons. Effects of NGF in the central nervous system are restricted to cholinergic neurons (neurons containing acetylcholine as a transmitter) in the striatum and basal forebrain. Four other neurotrophins, BDNF (brain-derived neurotrophic factor), NT-3, and NT-4/5, have been identified. All these neurotrophins are proteins with approximately 120 amino acids and form homodimers with a common three-dimensional structure.

There are two kinds of neurotrophin receptors, $p75^{NTR}$ (p75) and $p140^{Trk}$ (Trk; three subtypes TrkA, TrkB, and TrkC). Each Trk is a membrane protein with an extracellular ligand-binding domain, a transmembrane domain, and an intracellular tyrosine kinase domain. The binding of neurotrophins is thought to induce the dimerization of Trks and the autophosphorylation of dimerized receptors. p75 is also a transmembrane protein with a ligand-binding domain but does not have a tyrosine kinase domain (see Fig. 3). p75 receptors also undergo ligand-induced

dimerization. Neurotrophin-receptor complexes are internalized and transported from the axon to the soma in a retrograde fashion. The complex of phosphorylated Trk and neurotrophin activates Erk5 (a kinase of the MAP kinase family), which is followed by phosphorylation of CREB (Cyclic AMP-responsive element binding protein) and expression of specific genes for cell survival.

TrkA and TrkC specifically interact with NGF and NT-3, respectively, whereas TrkB interacts with either BDNF or NT-4/5. Expression of TrkA is localized to only a few neuronal types in the central and peripheral nervous systems, whereas both TrkB and TrkC are widely distributed throughout the brain and exist in the majority of neurons in the central nervous system, but not in nonneuronal cells such as glia. On the other hand, p75 is broadly expressed in neuronal and nonneuronal tissues. p75 is able to interact with each neurotrophin with equal affinity, but is reported to have a higher affinity for neurotrophin precursors rather than for matured neurotrophins cleaved by proteases. p75 has also been shown to bind to aggregated forms of β-amyloid and prion proteins, and to have a death domain in its intracellular region, which is somehow related to apoptotic signals.

4.1.3 Activation of the Ras-MAPK Cascade

Ligand binding to receptors results in autophosphorylation of tyrosine residues in receptors (or in IRSs in the case of insulin receptors). It is not clear and may be different among receptors if the phosphorylation of tyrosine residues further activates the catalytic activity or not. Phosphorylated tyrosine residues become targets of several proteins, which contain specific domains called SH2 (src homology domain 2). SH2 domains consist of approximately 100 amino acid residues, and recognize phosphorylated tyrosine (Y*)-containing motifs such as Y*DNV, Y*ENP, and Y*MXM. SH2 domains are included in different proteins such as nonreceptor type tyrosine kinases like src, phospholipase C γ (PLCγ), phosphatidyl inositol 3 kinase (PI3 kinase), phosphotyrosine phosphatase (PTP1C), GTPase-activating proteins (GAP), and adaptor proteins like Grb2 (growth-factor receptor bound protein 2). Amino acid sequences of various SH domains are similar but distinct from each other, and the amino acid sequences around the phosphorylated tyrosine-residue are also distinct. Thus, the interaction between SH2 domains and phosphorylated tyrosine-containing motifs is specific. For example, nine different tyrosine residues in the PDGF-β receptor are phosphorylated when activated by PDGF, and motifs containing residues 716, 771, 1009 interact with Grb2, GAP, and PTP-1D, respectively, motifs containing residues 579 and 581, 740 and 751, and 1009 and 1021 with src, PI3-kinase, and PLC-γ, respectively. Thus, a single kind of receptor may interact with and activate distinct proteins that have different kinds of SH2 domains.

A major downstream effector pathway of tyrosine kinase receptors is the MAPK (mitogen-activated protein kinase) cascade. Tyrosine kinase receptors recruit the adaptor protein Grb2. Adaptor proteins bind two different proteins and convert the binding of one protein into activation of the other protein; Grb2 has one SH2 and two SH3 domains, which are another kind of src homology domain, and binds with tyrosine kinase receptors through the SH2 domain and activates a protein called *son of sevenless* (sos) by interacting with it through SH3 domains. Sos is

a GDP–GTP exchange protein, and facilitates the release of GDP from ras, a small GTP-binding protein, and thereby facilitates the binding of GTP to ras. GTP-bound ras is the activated form of ras and interacts with and activates a serine/threonine kinase, raf kinase (or MAPKKK, MAP kinase kinase kinase). Raf kinase phosphorylates and activates MAPKK (or MEK (MAPK/ERK kinase)), which phosphorylates both tyrosine and threonine residues in MAPK (or ERK, extracellular signal-regulated kinase). Phosphorylated- and activated-MAPK is translocated into the nucleus and phosphorylates several proteins such as CREB. Finally, specific transcription factors are activated and specific genes are expressed. MAPK in PC12 chromaffine cells is activated through activation of either NGF or EGF receptors, but the final outcomes of activation of NGF and EGF receptors are opposite; activation of NGF receptors results in differentiation and arrest of proliferation, whereas activation of EGF receptors results in proliferation of PC12 cells. It seems that the duration of activation of MAPK is a critical factor and that differentiation is caused by a longer duration of MAPK activation. Activation of MAPK is also mediated through activation of other tyrosine kinase receptors including those for PDGF, insulin, and so on. There appear to be several routes from receptors to activation of ras, and also several routes for ras-mediated and ras-independent activation of MAPK.

PLC-γ binds to PDGF or FGF receptors and is translocated to membranes by interaction between phosphorylated tyrosine-containing motifs and SH2 domains, and is phosphorylated by receptors. PLC-γ is activated through translocation to membranes. Phosphorylation of PLC-γ by receptors does not affect its activity *in vitro* but is necessary to activate it in intact cells. PLC-γ catalyzes the formation of diacylglycerol and IP_3 from phosphatidylinositol 4,5-bisphosphate (PIP_2), as does PLC-β activated by the G-protein G_q. PIP_2 is known to bind to actin, α-actinin and other cytoskeletal proteins. PIP_2 bound to these proteins is not a substrate of PLC-γ but becomes a substrate of PLC-γ phosphorylated by tyrosine kinase receptors. There is a possibility that the activation of PLC-γ affects the organization of the cytoskeleton by allowing the interaction between actin and actin binding proteins through degradation of PIP_2.

PI_3 kinase catalyzes the phosphorylation of the 3′-OH of phosphatidyl inositol (PI), PI 4-phosphate, and $PI(4,5)P_2$, which results in the production of PI3-phosphate (PI3P), PI3,4-bis phosphate ($PI(3,4)P_2$), and PI3,4,5-triphosphate ($PI(3,4,5)P_3$). $PI(3,4)P_2$ and $PI(3,4,5)P_3$ bind to the Pleckstrin homology (PH) domain of intracellular signaling proteins and anchor them to cell membranes. PI3 kinase is composed of two subunits, p110 and p85, representing the catalytic subunit and an SH2-containing protein, respectively. The catalytic subunit is thought to be activated through interaction of p85 with IRSs phosphorylated by insulin receptors or with other phosphorylated proteins including receptor tyrosine kinases. Activation of PI3 kinase is necessary for insulin-mediated increase of glucose uptake, which is due to the translocation of the glucose transporter GLUT4 into cell membranes, in adipocyte, liver, and skeletal muscles. PI3 kinase increases the concentration of 3-phosphoinositide, which activates 3-phophoinositide-dependent protein kinase (PDK-1). PDK-1 phosphorylates the atypical protein kinase C and protein kinase B (PKB): PDK-1 and PKB have PH domains.

These kinases phosphorylate critical serine residues of GLUT4 resulting in its translocation to the plasma membrane. It has been suggested that PI3 kinase is involved in the regulation of cytoskeletal proteins and cell motility through activation of a subfamily of small GTP-binding proteins, rac.

4.2 Cytokine Receptors Linked with Tyrosine Kinases

4.2.1 Cytokine Receptors

Receptors for cytokines including interleukins (ILs) and interferons (IFNs) do not have tyrosine kinase activity, in contrast with receptors for growth factors. There are 30 IL genes and 25 IFN genes in the human genome. ILs are largely synthesized in lymphocytes and affect proliferation, differentiation, and functions of lymphocytes, which are involved in immune reactions. IFNs are secreted in response to virus infection, and suppress the proliferation of virus and also affect cell growth or immune responses. Receptors for ciliary neurotrophic factor (CNTF) and leukemia inhibitory factor (LIF) also lack tyrosine kinase activity, in contrast with NGF or other neurotrophin receptors. CNTF exerts potent survival effects on many peripheral neurons, in common with neurotrophins and FGF. LIF was identified also as a cholinergic differentiation factor, and has the ability to convert sympathetic adrenergic neurons into cholinergic neurons in *in vitro* experiments.

Most cytokine receptors are composed of an extracellular part, a single transmembrane domain, and an intracellular part. The extracellular parts of cytokine receptors contain cysteine-rich, Ig-like, or fibronectin type III-like domains, depending on the receptor species. The intracellular parts of the receptors do not have any known catalytic domains but have conserved motifs called *box 1 and box 2*, which are located adjacent to the transmembrane segments.

Table 5 shows some examples of cytokines and their receptors. Each of the receptors for growth hormones, erythropoietin, and prolactin is homogeneous and forms a homodimer when activated by the respective ligand. On the other hand, most cytokine receptors consist of more than two subunits, and one subunit is often shared by other cytokines. LIF, CNTF, and interleukin 6, 11 share a common receptor called gp130. LIF binds to a heteromeric complex of LIF receptor and gp130. CNTF binds to the CNTF receptor, which lacks a cytoplasmic domain, and the ligand–receptor complex then induces a heteromeric complex with the LIF receptor and gp130. Similarly, the complex of IL6 receptor with IL6 induces a complex containing a homodimer of gp130. The gp130 homodimer or gp130-LIF receptor heterodimer are thought to trigger signal transduction, while the CNTF and IL6 receptors facilitate the formation of dimers containing gp130. Receptors for IL3, GM-CSF (granulocyte macrophage-colony stimulating factor) are composed of a ligand-specific α-subunit and a common β_c-subunit, where the subscript c stands for "common." Receptors for IL2, 4, 5, 7, 9, 13, 15, and 21 are composed of a specific β-subunit and a common γ-subunit. An α-subunit contributes to an increase in the affinity of ligands to receptors but its significance in signal transduction is not known.

4.2.2 Activation of the JAK-STAT Pathway

Cytokine receptors do not have protein kinase activity but have the ability to activate nonreceptor tyrosine kinases such

Tab. 5 Protein kinase-coupled receptors (PKCRs) for growth factors, cytokines, and TGF-β, and TNF receptors.

Endogenous ligands	*Receptors*	*Events induced by ligand binding*
Group 1: Growth factor receptors with tyrosine kinase activity		
(a) Insulin, insulin-like growth factor 1(IGF-1), hepatocyte growth factor (HGF)	$\alpha_2\beta_2$ tetramer; α-subunit has ligand binding activity and β-subunit has a tyrosine kinase catalytic domain	(1) Homo- or heterodimerization of receptors (2) Transphosphorylation of tyrosine residues (3) Binding of phosphorylated tyrosine-containing motifs and SH2 domains in Grb2, PLCγ and so on. (4) Activation of SH2-containing proteins and a signal transduction cascade, Grb2-sos-ras-raf1-MAPKK-MAPK, and other cascades (5) Translocation of MAPK into nucleus and expression of specific genes
(b) Epidermal growth factor (EGF), Transforming growth factor α (TGF-α)	Homo- or heterodimer Two cysteine-rich domains	
(c) Platelet-derived growth factor (PDGF)(AA, AB, and BB dimers)	$\alpha\alpha$, $\alpha\beta$, $\beta\beta$ dimers Five Ig-like domains	
(d) Fibroblast growth factor, 23 members including acidic FGF (aFGF) and basic FGF (bFGF)	Four receptors FGFR1–4 Three Ig-like domains	
(e) Neurotrophin (NGF, brain-derived neurotrophin (BDNF), NT-3, NT-4/5)	trkA, trkB, or trkC with tyrosine kinase activity and p75 without tyrosine kinase activity	
Group 2: Cytokine receptors linked with tyrosine kinases		
(a) Growth hormone, erythropoietin (EPO), prolactin, granulocyte colony stimulating factor (G-CSF),	Homodimers	(1) Homo- or heterodimerization or oligomerization of receptors (2) Binding of Janus kinases (JAK) to receptors and activation of JAK (3) Phosphorylation by JAK of STAT (Signal transducer and activator of transcription) (4) Dimerization and translocation into nucleus of phosphorylated STATs (5) Stimulation of transcription of specific genes by dimerized STATs with or without other proteins
(b) Leukemia inhibitory factor (LIF), ciliary neurotrophic factor (CNTF), interleukin 6 (IL-6) IL-11,	Common subunit gp130	
(c) IL-3, IL-5, IL-21, granulocyte macrophage-colony stimulating factor (GM-CSF)	Common β_c subunit and specific α-subunits Common γ subunit and specific β subunits	
(d) IL-2, 4, 7, 9, 15 IL-10, 22, 28, 29 IL-12, 23	Common IL-10β-subunit and specific α-subunit Common IL—12β subunit and specific α subunit	

(e) Type 1 interferon(IFN) IFNα1, 2, 4, 5, 6, 7, 8, 10, 14, 16, 17, 21 IFN-β1, 3, IFN-τ, IFN-ω, FN-κ	Heterodimer of IFNαR1 and IFNαR2	
(f) Type 2 interferon IFN-γ	Tetramer of two IFNγR1 and two IFNγR2	
Group 3: TGFβ receptor with serine/threonine kinase activity		
Transforming growth factor β1 (TGF-β1), TGF-β2, TGF-β3, activin BMPs (bone morphogenetic proteins) (2–7)	Tetramers composed of two type I and two type II receptors Type II receptors exist as dimers and are constitutively active even in the absence of ligands	(1) Binding of TGFβ to type II receptor dimers (2) Binding of type I receptors to the TGFβ- type II receptor complex (3) Phosphorylation of type I receptor and its activation (4) Phosphorylation of SMAD (5) Heterodimerization of phosphorylated SMAD with SMAD 4 (6) Translocation of SMAD heterodimer into nucleus and stimulation of specific gene expression
Group 4: TNF receptor family		
TNF, FasL NGF	TNFR Homo-trimers FAS homotrimer NGFRp75 homodimer	(1) Binding of TNF trimer to TNFR-1, release of the inhibitory protein silencer of death domains from TNF-R1 intracellular domain (ICD) (2) Binding of the adaptor protein TRDD (TNF receptor-associated death domain) to aggregated ICD (3) TRDD recruits various adaptor proteins including caspase 8, IκB kinase, and so on. (4) Caspase 8 *triggers* a protease cascade, IκBK phosphorylates IkB that results in its dissociation from NFκB, and JNK is activated leading to phosphorylation of c-Jun (5) Initiation of apoptosis, and stimulation of specific gene transcription by NFκB and c-Jun

as JAK. JAK was named after the Roman god with two faces, Janus, because they have two kinase-like domains, although only one of them turned out to be active. Agonist binding to cytokine receptors induces the dimerization or oligomerization of receptors, which leads to activation of JAKs. JAKs bind to box 1 and box 2 of cytokine receptors. Four members of the JAK family are known; JAK1, 2, 3, and TYK2. The common receptors, gp 130, β_c, and γ, appear to interact with and activate JAK1/JAK2/TYK2, JAK2, and JAK1/JAK3, respectively. Two molecules of JAK kinases are thought to interact with ligand-bound receptor dimers and cross-phosphorylate and activate each other. Major targets of activated JAKs are STATs (signal transducer and activator of transcription). Seven members of the STAT family have been identified in the human genome (STAT1, 2, 3, 4, 5A, 5B, and 6), and each cytokine is thought to activate a specific set of STATs. STATs have SH2 domain, and the interaction between SH2 domain and phosphorylated tyrosine-containing motifs may contribute to the formation of homo- or heterodimers of phosphorylated STATs. Phosphorylated and dimerized STATs are translocated from the cytoplasm into the nucleus. STAT dimers alone or with another protein (e.g. p48) interact with DNA and stimulate transcription of specific genes.

4.3 TGF-β Receptors with Serine/Threonine Kinase Activity

Transforming growth factor-β (TGF-β) is a prototype for a large family of factors that regulate cell growth and differentiation, including activin/inhibin, bone morphogenic protein (BMP), Müllerian inhibition substances (MIS), and glial cell line-derived neurotrophic factor (GDNF). These factors have homology in amino acid sequence and tertiary structure, and all factors except MIS are dimers connected with an S–S bond. TGF-β was originally identified as a factor that causes transformation of fibroblast, and is now known to suppress proliferation of cells. Activin and inhibin were isolated as factors that stimulate or inhibit the release of follicle-stimulating hormone from cells in the pituitary, respectively. Activin is now known to function as an inducer of mesoderm. There are two kinds of inhibin dimers $\alpha\beta_A$ and $\alpha\beta_B$, and three kinds of activin homo- or heterodimers, $\beta_A\beta_A$, $\beta_A\beta_B$, $\beta_B\beta_B$. BMPs are known to induce the formation of bone and cartilage. GDNF was found as a protein that functions as a trophic factor for dopaminergic cells in the midbrain and was recently shown to support survival of motor neurons.

Receptors for TGF-β contain serine/threonine kinase activity in the intracellular part. Ligand binding to the extracellular surface of TGF receptors induces the oligomerization of type I and type II subtypes, most likely resulting in the tetramerization of two type I and two type II receptors. Type II receptors exist as dimers and are constitutively active even in the absence of TGF-β. The binding of TGF-β to type II receptor dimers causes the binding of type I receptor to the TGF-β-type II receptor complex and the phosphorylation of type I receptors. The phosphorylation of type I receptors activates its own kinase and induces the phosphorylation of downstream signal molecules named SMAD. There are eight kinds of SMAD genes in the human genome. SMAD1, 2, 3, 5, and 8 are directly phosphorylated by the type I receptor on conserved serine residues in the

carboxy terminus. Phosphorylated SMAD forms a heterodimer with SMAD4, and the heterodimer is translocated to the nucleus where it binds to transcription regulatory proteins. On the other hand, SMAD 6 and 7 inhibit the signals induced by phosphorylated SMAD. Disruption of the activity of TGF-β and SMAD family members may cause a variety of human diseases including cancer.

4.4 Intracellular Signaling: Common Pathways Triggered by GPCRs and PKCRs

Major downstream effector cascades of growth-factor receptors and cytokine receptors are the ras-MAPK and JAK-STAT pathways, respectively. There are also crossing pathways. Most cytokines also activate the ras-MAPK cascade. The src family tyrosine kinase, the adaptor protein SHC, and the carboxy-terminal domain of cytokine receptors are suggested to be involved in this response. At the same time, tyrosine kinase receptors activate STATs; for example, EGF and PDGF induce the phosphorylation of STAT1a. In addition, a pathway similar to the MAPK cascade is suggested to be present downstream of TGF-β receptors. Crosstalk between pathways triggered by GPCRs and PKCRs is also known. MAPK may be activated through GPCRs. One pathway is likely to be mediated through activation of G_q followed by activation of protein kinase C and/or increase in intracellular Ca^{2+} concentrations and another through G-protein $\beta\gamma$ subunits. In addition, activation of JAKs and STATs is reported to be caused by activation of the angiotensin II receptor (a GPCR).

Activation of PLC is mediated through different receptors. Receptors linked to G_q, G_i/G_o and growth-factor receptors may activate PLC-β1, PLC-β2/β3, and PLC-γ, respectively, and therefore any agonists interacting with these receptors may cause the increase of intracellular Ca^{2+}, and activation of Ca^{2+}-calmodulin-dependent protein kinases and protein kinase C. In addition, it has been reported that there are two kinds of PI3 kinases, which are activated by growth-factor receptors and G-protein $\beta\gamma$ subunits, respectively. PLC-γ and PI3 kinase are also known to be activated by some cytokines.

These findings indicate that the relation between receptors and intracellular signaling pathways is not straightforward and that the same outcome may be induced by activation of either GPCRs or PKCRs. Major consequences of simulation of PKCRs and GPCRs are expression of specific genes and regulation of specific protein functions, respectively. It is also possible, however, that PKCRs and GPCRs may regulate protein functions and gene expression, respectively.

5 Other Receptors

Receptors for natriuretic peptides are composed of three segments; an extracellular ligand-binding domain, a single transmembrane domain, and a cytoplasmic domain with guanylate cyclase activity and a protein kinase-like domain. There are three kinds of natriuretic peptides, ANP (atrial type), BNP (brain type) and CNP (C type), and two kinds of receptors, GC-A and GC-B. GC-A interacts with both ANP and BNP, and GC-B with CNP. Another kind of receptor for all three natriuretic peptides lacks catalytic domains, serves to internalize these peptides, and is called a *clearance receptor*. The other

type of guanylate cyclase is a soluble protein, lacks the transmembrane segment, and is activated by nitric oxide through its interaction with a heme group in the enzyme.

There are approximately 20 members in the superfamily of tumor necrosis factor receptors (TNFRs), including TNFR1, TNFR2, FAS, and CD40. Tumor necrosis factor (TNF) is a major mediator of apoptosis, inflammation, and immunity, and is thought to be involved in various diseases. TNF is a homotrimer of 157 amino acid subunits produced by activated macrophages. Most other ligands also act as a trimer. They share the homologous sequence in the domains responsible for trimerization, whereas the other regions that recognize their receptors are divergent. The receptor also forms a trimer after ligand binding. These receptors have several cysteine-rich domains in the extracellular domain, a membrane-spanning domain, and a cytoplasmic tail where some members have a death domain and a TRAF (TNF receptor-associated factor) binding site. Upon ligand binding, inhibitory proteins are released from the aggregated form of intracellular domains in the receptor and various proteins are bound there. There are six TRAFs in the human genome: TRAF1 and TRAF2 form homo- or heterodimers, and recruit IκB kinase (IKKα and IKKβ) and upstream kinases for c-jun NH_2-terminal kinase (JNK). IκB (inhibitor of nuclear factor κB (NF-κB) retains NF-κB in the cytoplasm in an inactive form, and phosphorylation of IκB results in its dissociation from NF-κB and translocation of NF-κB into the nucleus. NF-κB and c-jun stimulate transcription of specific genes in the nucleus. FAS-associated death domain (FADD) proteins are bound to the death domain, and caspase 8 is recruited to FADD. Caspase 8 initiates a proteolysis cascade leading to apoptosis. The molecular events involved in these pathways are complex and many details remain to be elucidated.

Acknowledgments

We are deeply indebted to Dr. J. Wess in N.I.H. for detailed comments and corrections throughout the whole manuscript. We also thank Dr. K. Inoue in National Institute of Health Sciences, Dr. Hiroyasu Furukawa in Columbia University, Dr. U. Langel in Stockholm University, Dr. G. Berstein in Wyeth Pharmaceuticals for reading a part or all of this manuscript and giving us valuable comments.

See also Endocrinology, Molecular; Membrane Traffic: Vesicle Budding and Fusion; Receptor Targets in Drug Discovery; Receptor, Transporter and Ion Channel Diseases; Signal Transduction Mediated by Heptahelical Receptors and Heterotrimeric G Proteins; Vitamin Receptors.

Bibliography

Books and Reviews

Aaronson, D.S., Horvath, C.M. (2002) A road map for those who do not know JAK-STAT, *Science* **296**, 1653–1655.

Attisano, L., Wrana, J.L. (2002) Signal transduction by the TGF-beta superfamily, *Science* **296**, 1646–1647.

Chen, G., Goeddel, D.V. (2002) TNF-R1 signaling: a beautiful pathway, *Science* **296**, 1634–1635.

Dijke, P.T., Hill, C.S. (2004) New insights into TGF-β-Smad signalling, *Trends Biochem. Sci.* **28**, 265–273.

Grotzinger, J. (2002) Molecular mechanisms of cytokine receptor activation, *Biochim. Biophys. Acta* **1592**(3), 215–223.

Ihle, J.N. (2001) The Stat family in cytokine signaling, *Curr. Opin. Cell Biol.* **13**, 211–217.

Khakh, B.S. (2001) Molecular physiology of P2X receptors and ATP signaling at synapses, *Nat. Rev. Neurosci.* **2**, 165–174.

Locksley, R.M., Killeen, N., Lenardo, M.J. (2001) The TNF and TNF receptor superfamilies: integrating mammalian biology, *Cell* **104**, 487–501.

Lu, Z.L., Saldanha, J.W., Hulme, E. C. (2002) Seven-transmembrane receptors: crystals clarify, *Trends Pharmacol. Sci.* **23**, 140–146.

Malenka, R.C., Nicoll, R.A. (1999) Long-term potentiation–a decade of progress?, *Science* **285**, 1870–1874.

Malinow, R., Malenka, R.C. (2002) AMPA receptor trafficking and synaptic plasticity, *Annu. Rev. Neurosci.* **25**, 103–126.

Okada, T., et al. (2001) Activation of rhodopsin: new insights from structural and biochemical studies, *Trends Biochem. Sci.* **26**, 318–324.

Pestka, S. (2000) The human interferon alpha species and receptors, *Biopolymers* **55**, 254–287.

Pierce, K.L., Premont, R.T., Lefkowitz, R.J. (2002) Seven-transmembrane receptors, *Nat. Rev. Mol. Cell Biol.* **3**, 639–650.

Rios, C.D., Gomes, J.I., Devi, L.A. (2001) G-protein-coupled receptor dimerization: modulation of receptor function, *Pharmacol. Ther.* **92**, 71–87.

Sierra, D.A., Popov, S., Wilkie, T.M. (2000) Regulators of G-protein signaling in receptor complexes, *Trends Cardiovasc. Med.* **10**, 263–268.

Song, I., Huganir, R.L. (2002) Regulation of AMPA receptors during synaptic plasticity, *Trends Neurosci.* **25**, 578–588.

Wieland, T., Mittmann, C. (2003) Regulators of G-protein signaling: multifunctional proteins with impact on signaling in the cardiovascular system, *Pharmacol. Ther.* **97**, 95–115.

Primary Literature

Furukawa, H., Gouaux, E. (2003) Mechanisms of activation, inhibition and specificity: crystal structure of the NMDA receptor NR1 ligand-binding core, *EMBO J.* **22**: 2873–2885.

Kunishima, N., Shimada, Y., Tsuji, Y., Sato, T., Yamamoto, M., Kumasaka, T., Nakanishi, S., Jingami, H., Morikawa, K. (2000) Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor, *Nature* **407**, 971–977.

Miyazawa, A., Fujiyoshi, Y., Unwin, N. (2003) Structure and gating mechanism of the acetylcholine receptor pore, *Nature* **424**, 949–955.

Ogiso, H., et al. (2002) Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains, *Cell* **110**, 775–787.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C., Motoshima, H., Fox, B.A., Trong, I.Le, Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M., Miyano, M. (2000) Crystal structure of rhodopsin: A G protein-coupled receptor, *Science* **289**, 739–745.

Schlessinger, J. (2002) Ligand-induced, receptor-mediated dimerization and activation of EGF receptor, *Cell* **110**, 669–672.

Unwin, N., Miyazawa, A., Li, J., Fujiyoshi, Y. (2002) Activation of the nicotinic acetylcholine receptor involves a switch in conformation of the alpha subunits, *J. Mol. Biol.* **319**, 1165–1176.

Receptor Targets in Drug Discovery

Michael Williams[1], *Christopher Mehlin*[2], *Rita Raddatz*[1], *and David J. Triggle*[3]
[1] *Cephalon Incorporated, West Chester, PA*
[2] *University Of Washington, Seattle, WA*
[3] *State University of New York, Buffalo, NY*

1 Drug Discovery 596

2 Cellular Communication 597

3 Receptor Concepts 600
3.1 Occupancy Theory 600
3.2 Rate Theory 604

4 Receptor Complexes and Allosteric Modulators 604

5 Ternary Complex Models 606

6 Constitutive Receptor Activity 606

7 Efficacy Considerations 607

8 Receptor Dynamics 608

9 Receptor Nomenclature 609

10 Receptor Classes 609
10.1 G-protein–Coupled Receptors 610
10.1.1 G-protein–Associated or – Modulating Proteins 611
10.2 Ligand-gated Ion Channels 611
10.2.1 $GABA_A$/Benzodiazepine (BZ) Receptor 611
10.2.2 The NMDA Receptor 612
10.2.3 The Nicotinic Cholinergic Receptor, nAChR 612
10.2.4 The P2X Receptor 612

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

10.3 Steroid Receptor Superfamily 613
10.4 Intracellular Receptors 613
10.5 Non-GPCR-linked Cytokine Receptors 613
10.6 Orphan Receptors 613
10.7 Neurotransmitter Transporters 614
10.8 Neurotransmitter Binding Proteins 614
10.9 Drug Receptors 614

11 Molecular Biology of Receptors 615

12 Functional Genomics: Target Validation/Confidence Building 615

13 Compound Properties 618
13.1 Structure-activity Relationships 618
13.2 Defining the Receptor–Ligand Interaction 619
13.2.1 Receptor-binding Assays 620
13.2.2 Functional Assays 622
13.3 Receptor Sources 623
13.4 ADME 623
13.5 Compound Databases 624

14 Lead Compound Discovery 625
14.1 High-throughput Screening 625
14.2 Compound Sources 626
14.2.1 Natural Product Sources 626
14.2.2 Pharmacophore-based Ligand Libraries 628
14.2.3 Diversity-based Ligand Libraries 631
14.3 Biologicals and Antisense 632

15 Future Directions 633

Acknowledgment 633

Bibliography 633
Books and Reviews 633
Primary Literature 635

Keywords

Agonist
A ligand that interacts with a receptor to activate and produce a defined response.

Antagonist
A ligand that interacts with a receptor to block the effects of an agonist.

Drug
A compound that, via its interaction with a defined receptor target, reverses a disease phenotype. Drugs are characterized by their selectivity, safety, and pharmacokinetic properties, which are consistent with their reaching their target site of action within the body.

G-protein–Coupled Receptors
The best-known receptor family through which the largest number of known drugs act to produce their pharmacological effects, for example, cimetidine (histamine H_2 receptor), and propranolol (β-adrenoceptor). These receptors are defined by a common structure of seven transmembrane domains and their interactions with members of the heterotrimeric G-protein family.

Heterotrimeric G-protein
A family of oligomeric proteins consisting of three subunits (α, β, and γ). These protein complexes transduce signals from activated GPCRs to a variety of intracellular signaling cascades.

High-throughput Screening
Testing of large sets of compounds (typically 100 000–2 million) for interaction with a receptor target.

Ion Channels
Typically composed of protein multimers that form channels to conduct anions and/or cations through the membrane (e.g. the neuronal nicotinic receptor). Activation of ion channels can occur by ligand activation (e.g. nicotine, glutamate), by changes in membrane voltage, pH, mechanical distension, and temperature.

Ligand
A defined chemical entity that is natural (e.g. histamine) or synthetic and (e.g. propranolol) that binds to a receptor.

New Chemical Entity (NCE)
A compound with potential druglike properties.

Orphan Receptor
A receptor, identified on the basis of its structural motif for which the endogenous ligand and physiological function have not yet been identified.

Receptor

1. A membrane associated protein that selectively responds to extracellular messenger molecules to alter cellular function to maintain homeostasis and normal cell/tissue function. Receptors comprise two major classes, G-protein–coupled receptors (GPCRs) and ion channels.

2. A drug target – a concept that expands the definition of a receptor to intracellular sites that include enzymes, transporters, and signal transduction elements.

SAR
Dose/concentration dependent differences in the efficacy (agonist or antagonist) of compounds in a ligand series, which can be associated with changes in chemical structure.

Therapeutic Index
The ratio between the dose at which a drug or NCE produces efficacy in a disease state or disease model, and the dose at which it produces side effects.

Receptors, located on both the cell surface and within the cell are the defined molecular targets through which drugs produce their beneficial effects in various disease states. Initially conceptualized over a century ago by Ehrlich and Langley, receptor concepts and receptor theory have undergone continuous modification as their behavior in normal and disease states/tissues have been more clearly characterized. Since the isolation of the nicotinic acetylcholine receptor (nAChR) from the *Torpedo*, some 50 years ago, new techniques of molecular biology have made it relatively routine to isolate receptors, including orphan receptors. Once these have been validated, they can be used in conjunction with high- throughput screening approaches to identify "hits," molecules that bind with relatively high affinity to these targets. Such hits can then be optimized to druglike entities using combinatorial/parallel synthesis technology platforms to yield clinical candidates that are potent, efficacious, and bioavailable entities with appropriate safety profiles.

1
Drug Discovery

Drug discovery, the process of identifying and developing novel chemical entities (NCEs) to treat human disease states, has entered the new century with a wealth of sophisticated technologies and information generation platforms, which have the potential to allow more rapid development of new drugs with improved selectivity and safety profiles.

Successful drugs from the pharmaceutical industry that include small-molecule drugs (NCEs), biologicals, vaccines, and antibodies, have resulted in an increase in human life expectancy, with the concomitant expectation that new drugs will be available for the aging population, which will improve the quality of life while adding to its span. The draft map of the human genome has led to the possibility of understanding diseases at a level of precision that was never before possible. Thus, in the last decade, there has been an exponential increase in understanding the potential pathophysiologies of disorders like pain, diabetes, some forms of

cancer, and neurodegenerative processes (Alzheimer's and Parkinson's diseases).

The challenge in effectively using the copious information now available for disease-related gene identification and the molecular targets involved in the cause(s) of various diseases is how to handle and productively focus the bewildering flow of data, much of it archival, to cost-effectively discover and develop new drugs – a challenge that has given rise to key disciplines like bio- and chemo-informatics.

Irrespective of compound or drug target sources, more productive high-throughput screening (HTS) processes, and more detailed structural information on putative drug targets, it is clear that continued success in the drug discovery process will depend on the iteration and integration of lead compound identification and optimization through the hierarchical complexity of *in vitro* and *in vivo* assays that measure efficacy, selectivity, side effect liability, absorption, distribution, metabolism and excretion (ADME), and potential toxicity. To do this in an effective manner, it is critical to understand how existing knowledge of drug targets has evolved and what the realistic potential is for identifying, validating, and prioritizing new ones.

2 Cellular Communication

The transfer of information between cells and the subsequent integration of multiple inputs that are necessary to maintain cellular homeostasis and tissue viability – under both normal and adverse disease-related conditions – involves a variety of different external signaling modalities. These include: temperature, membrane potential, mechanical distension and stress, alterations in ion (H^+/K^+) concentrations, pheromones, oderants, as well as more traditional neurotransmitters, neuromodulators, and hormones. These physical stimuli and endogenous chemicals elicit their effects through interactions with cell surface targets, usually proteins, which are termed *receptors* (Table 1). Once receptors are activated by an *agonist* ligand, a compound capable of producing a cellular response, they transduce or couple the energy associated with the binding event to a cellular effect via a signal transduction process(es) involving protein–protein interactions or second messenger signaling systems, for example, G-proteins, protein kinase or phosphatase modulation, alterations in lipid metabolism, calcium translocation, and so on, to produce acute or more long-term effects on cell and tissue function. Thus, neurotransmitters, neuromodulators, and hormones can produce transient increases in second messengers like cyclic AMP or inositol triphosphate (IP_3) or more long-term changes that involve changes in gene expression through activation of transcription factors.

Alterations in receptor function occur by: (1) a functional overstimulation of receptors with their consequent desensitization, a phenomenon resulting from excess ligand availability or an enhanced coupling of the ligand-activated receptor to second messenger systems, or (2) a reduction in stimulation resulting from decreased ligand availability or dysfunctional receptor coupling processes.

Drugs that effectively treat human disease states by restoring disease-associated defects in signal transduction can act either by replacing endogenous transmitters, for example, L-dopa treatment to replace the dopamine (DA) lost in Parkinson's disease, or blockade of excess agonist stimulation,

Tab. 1 Drug recognition sites classified as receptors.

Type	*Class*	*Examples*
Receptor	GPCR	Dopaminergic, adrenergic, $GABA_B$, mGluR
	Ion channel	
	LGIC	NMDA, nicotinic, $GABA_A$
	VGIC	Na_v, Ca_v
	Others	ASIC, TRPV
	Enzyme Assoc. R	
	Tyrosine kinase	PDGF, EGF, Trk
	Others	CNTF, cytokine
	Nuclear hormone R	
	Steroid R	GR, PR
	Others	PPAR, RAR
Neurotransmitter Transporter	Plasma membrane neurotransmitter transporters	SERT, NET, GAT, DAT
Enzyme	Cell surface/ extracellular	ACE, NOS, HMG CoA reductase, acetylcholinesterase
	Intracellular	Caspase, CDK, ROCK DNA polymerase

Note: GABA: γ-aminobutyric acid; mGluR: metabotropic glutamate receptor; LGIC: ligand-gated ion channel; VGIC: voltage-gated ion channel; NMDA: *N*-methyl-D-aspartic acid; ASIC: acid sensing ion channel; TRPV: transient receptor potential vanilloid; PDGF: platelet-derived growth factor; EGF: epidermal growth factor; CNTF: ciliary neurotrophic factor; GR: glucocorticoid receptor; PR: progesterone receptor; PPAR: peroxisome proliferator-activated receptor; RAR: retinoic acid receptor; SERT: serotonin transporter; NET: norepinephrine transporter; GAT: GABA transporter; DAT: dopamine transporter; ACE: angiotensin-converting enzyme; NOS: nitric oxide synthase; CDK: cyclin-dependent kinase; ROCK: rho-dependent kinase.

for example, the histamine H_2 receptor antagonist, cimetidine, which blocks histamine-induced gastric acid secretion and thus reduces ulcer formation. Of the approximately 450 targets through which known drugs act, 71% are receptors in the G-protein–coupled receptor (GPCR), ion channels, or neurotransmitter transporter (SCDNT) families. Of these drugs, approximately 90% produce their effects by antagonizing the actions of endogenous agonists

Enzymes, by producing products that regulate second messenger availability, for example, adenylyl cyclase-catalyzed production of cyclic AMP or phosphodiesterase-mediated degradation of cAMP, or by modifying protein targets, for example, by adding (protein kinases) or removing (protein phosphatases) phosphates on serine or threonine residues, serve a similar role in cellular homeostasis, and as such, also represent key drug targets that can be conceptually included in the category of receptors.

The seminal concept that therapeutic agents produce their effects by acting as "magic bullets" at discrete molecular targets on tissues within the body is attributed to Ehrlich and Langley, who at the beginning of the twentieth century independently generated experimental data that led to the evolution of the "lock and key" hypothesis. A ligand (L) thus acted as

the unique "key" to selectively modulate receptor (R) activity, the latter functioning as the "lock" for the "entry" of external signals into the cell. In this model, an agonist ligand forms an RL complex and has the ability to "turn" the lock (Equation 1), whereas a receptor antagonist would occupy the lock and prevent agonist access.

$$\text{Receptor} + \text{ligand} \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} \text{RL} \longrightarrow \text{Signaling event} \qquad (1)$$

Despite quantal advances in the technology used to study receptor interactions that have occurred over the past century, the receptor–ligand (RL) concept and the similar enzyme–substrate (ES) complex have remained the conceptual foundations for understanding receptor and enzyme function, disease pathophysiology, and medicinal chemistry-driven approaches to drug discovery. Over the past decade, however, with the explosion in the number and diversity of receptors driven by receptor cloning approaches and the mapping of the human genome, there has been an increased appreciation of the inherent complexity of receptor function.

At one time, it was thought that a receptor-mediated response was a predictable, linear process that involved ligand-induced activation of a protein monomer and its signal transduction pathway independently, or with very little influence, from other membrane proteins. It has become increasingly evident, however, that receptors can physically interact both with one another and with other membrane proteins. Numerous examples exist of receptor coexpression and interactions, for example, $GABA_BR1$ (γ-aminobutyric acid) and $GABA_BR2$, dopamine with somatostatin, and $GABA_A$ receptors and opioid receptors with α_2-adrenoceptors. These interactions are often necessary to obtain functional cell surface receptors and allow interactions with one protein partner to modulate the function of the entire signaling complex. Effects of receptors on protein subunits can be mediated through orthosteric sites, for example, the site that the endogenous ligand(s) interact with, or through allosteric sites (e.g. the benzodiazepine (BZ) site in the $GABA_A$ receptor, and the glycine receptor on the *N*-methyl-D-aspartate (NMDA) receptor complex). More recently, allosteric interactions with GPCRs (muscarinic adenosine A_1) and neurotransmitter transporters (NET) have been reported. Integration of the functional effects of signal transduction pathways also affects receptor functions in a cell-specific manner (receptor cross talk). There also exist several discrete classes of receptor-associated proteins including receptor-activity-modifying proteins (RAMPs) and trafficking chaperones, which currently number in the many hundreds and play key roles in cellular events like receptor trafficking from the endoplasmic reticulum to the cell surface, and modulation of cell surface responses. Examples of the complexity of receptor signaling at the postsynaptic level include the NMDA receptor, where proteomic analysis demonstrated more than 70 proteins other than the receptor potentially involved in the function of the receptor complex; and the ATP-sensitive $P2X_7$ receptor, where a signaling complex comprised of 11 proteins including laminin-3, integrin 2, actin, supervillin, MAGuK, three heat shock proteins, and phosphatidylinositol 4-kinase; and the receptor protein tyrosine phosphatase (RPTP) was identified, which appears to modulate $P2X_7$ receptor function through control of its phosphorylation state.

3 Receptor Concepts

Receptor theory is based on the classical Law of Mass Action as developed by Michaelis and Menten for the study of enzyme catalysis. The extrapolation of classical enzyme theory to receptors is, however, an approximation. In an enzyme–substrate (ES) interaction, the substrate S undergoes an enzyme-catalyzed conversion to a product or products. Because of the equilibrium established, product accumulation has the ability to reverse the reaction process. Alternatively, the latter can be used in other cellular pathways and is thus removed from the equilibrium situation or can act as a feedback modulator to alter the ES reaction either positively or negatively (Equation 2).

$$\text{Enzyme} + \text{Substrate} \rightleftharpoons \text{ES} \rightleftharpoons \text{E} + \text{Product} \qquad (2)$$

For the receptor–ligand interaction, binding of the ligand to the receptor to form the RL complex results in a response driven by the thermodynamics of the binding reaction that leads to functional changes in the target cell (Equation 1). Whereas conformational changes occur in the ligand, the receptor, or both, there is no chemical change in the ligand resulting from the RL interaction such that there is no chemical product derived from the ligand that results from the RL interaction. Despite events such as receptor internalization, receptor phosphorylation, second messenger system activation, and so on, the bound ligand is chemically unchanged by the binding event, and thus, there is no equilibrium established between the RL complex and the consequences of receptor activation.

After the formation of an RL complex, a functional response to receptor activation can be related to the concentration of the ligand that is present (Fig. 1a). *Occupancy theory*, developed by Clark in 1926, is based on a dose/concentration-response relationship. This theory has undergone continuous refinement on the basis of experimental data to aid in further delineating the increasingly complex concept of ligand efficacy and to accommodate allosteric site modulation of receptor function, ternary complex models (TCM), and their extension to constitutively active receptor systems – those that are functional in the absence of an identified ligand.

3.1 Occupancy Theory

The basic premise of Clark's occupancy theory, based on Michaelis–Menten theory, was that the effect produced by an agonist was dependent on the number of receptors occupied by that agonist, a reflection of the agonist concentration present. This theory was developed from Clark's observations that acetylcholine (ACh) receptor antagonists such as atropine caused a rightward shift in the ACh dose-response (DR) curve in muscle preparations when plotted logarithmically. The basic tenets of occupancy theory were as follows: (1) the RL complex was assumed to be reversible; (2) the association of the receptor with the ligand to form the RL complex was defined as a bimolecular process with dissociation being a monomolecular process; (3) all receptors in a given system were assumed to be equivalent to one another and able to bind ligand independently of one another; (4) formation of the RL complex did not alter the free (F) concentration of the ligand or the affinity of the receptor for the ligand; (5) the

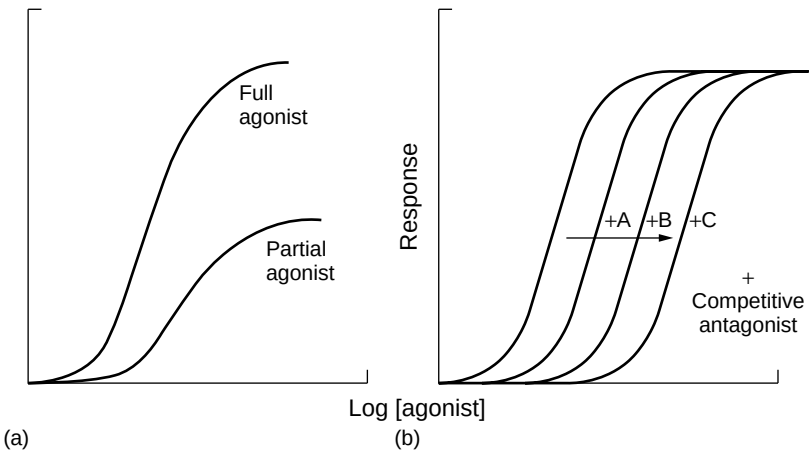

Fig. 1 Dose-response curve. The addition of increasing concentrations of an agonist ligand causes an increase in a biological response. Plotted on a logarithmic scale, a sigmoidal curve is obtained. (a) A full agonist produced a maximal response, whereas a partial agonist reaches a plateau that is only part of the response seen with a full agonist. (b) In the presence of antagonist concentrations A, B, C, the dose-response curve is progressively moved to the right. Increasing agonist concentrations overcome the effects of the antagonist.

response elicited by receptor occupancy was directly proportional to the number of receptors occupied; and (6) the biological response was dependent on attaining an equilibrium between R and L according to Equation 1.

Although it is not always possible to determine the concentration of free ligand (F) or that of the RL complex, rearranging the latter, the equilibrium dissociation constant, K_d, equaling $k - 1/k + 1$, can be derived from the equation:

$$K_d = \left(\frac{[R][L]}{[RL]}\right) \qquad (3)$$

and is equal to the concentration of L that occupies 50% of the available receptors.

Antagonist interactions with the receptor were defined by Gaddum as being the result of receptor occupancy with the antagonist ligand being unable to elicit a functional response. Antagonists thus block agonist actions. Agonists overcome the effects of a competitive (e.g. reversible/surmountable) antagonist when their concentration is progressively increased (Fig. 1b) such that in the presence of increasing fixed concentrations of a competitive antagonist, a series of parallel agonist DR curves can be generated that shift progressively to the right (Fig. 1b). A Schild regression relationship, a plot of log (DR-1) versus the log antagonist concentration (Fig. 2), can be used to derive the pA_2 value for an antagonist from the intercept of the abscissa. The pA_2 value is the negative logarithm of the affinity of an antagonist for a given receptor in a defined biological system and is equal to $-\log_{10} K_B$ where K_B is the dissociation constant for a competitive antagonist with a slope of near unity. Not all antagonists are competitive. Non- or uncompetitive antagonists that act at allosteric sites or that bind irreversibly to the agonist site have slopes that are

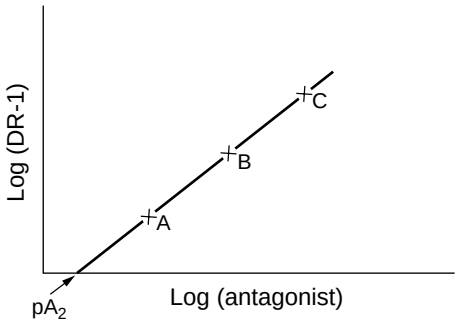

Fig. 2 Schild plot regression. Data from Fig. 1(b) for antagonist concentrations A, B, and C can be plotted by the method of Schild (29) to yield a pA_2 value, a measure of antagonist activity. A slope of unity indicates a competitive antagonist.

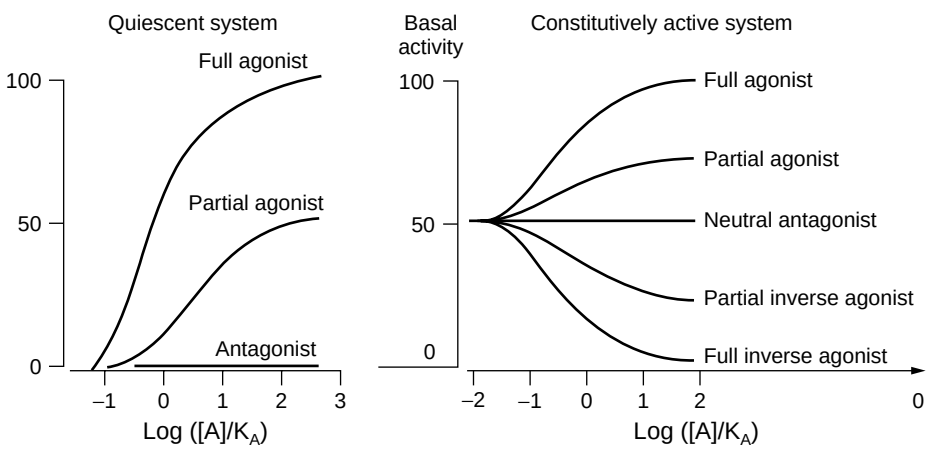

Fig. 3 Efficacy in quiescent and constitutively active systems. In a quiescent system, three types of ligands can be defined, full agonist, partial agonist, and antagonist, depending on the response elicited. In a constitutively active system, an antagonist from a quiescent system is defined as a neutral antagonist with ligands that inhibit the activity of the constitutively active system and are defined as full and partial inverse agonists depending on the degree of inhibition.

significantly less than unity. The Schild plot can thus be used to determine the mechanism by which an antagonist produces its effects.

Occupancy theory was modified by Ariens on the basis of data showing that not all cholinergic agonists were able to elicit a maximal response in a skeletal muscle preparation, even when administered at supramaximal concentrations. This led to the introduction of the concept of the *intrinsic activity* of a ligand. A full agonist was defined as having a value of 1.0 with the value for an antagonist being zero. However, many compounds were subsequently identified that bound to the receptor but were able to only produce a portion of the response seen with a full agonist. These were defined as partial agonists (Fig. 1a). By definition, these compounds were also partial antagonists.

Other agonists have also been identified that produce a response greater than the previously defined full agonist. These have been termed *super agonists*. One example

is the muscarinic cholinergic receptor agonist, L 670,207 (Fig. 3), an arecoline bioisostere that was 70% more active than arecoline and thus had an intrinsic activity of 1.7. From the activity seen in response to L 670,207, the system that is used to characterize these muscarinic agonists was obviously capable of a greater response than that seen with arecoline, making the latter compound a partial agonist.

The partial agonist concept was additionally refined by Stephenson in introducing the concept of *efficacy*, ε, which differed from intrinsic activity in that the latter was defined as a proportion of the maximal response [effect = (RL)]. This concept was extended to situations in which a maximal response to an agonist could occur when only a small proportion of the total number of receptors on a tissue were occupied (a condition termed *receptor reserve*), as in the situations when receptors were inactivated by alkylating agents. This resulted in a nonlinear occupancy relationship with the response then being defined as the stimulus, S, a product of the fraction of receptors occupied and the ligand efficacy. A nonlinear functional response clearly complicates data interpretation, especially when spare receptor or receptor reserve concepts are introduced to rationalize individual data sets. An additional issue in defining efficacy was the degree to which the receptor activation event and its blockade by antagonists was measured through events that were spatially and temporally removed from the receptor activation event, and also the degree to which the response could be amplified through cofactor and signal transduction cascades.

Kenakin described ligand-mediated responses in a given tissue in terms of four parameters: (1) receptor density; (2) the efficiency of the transductional process; (3) the equilibrium dissociation constant of the RL complex; and (4) the intrinsic efficacy of the ligand at the receptor. *In vivo*, receptor occupancy for exogenous ligands is primarily dependent on pharmacokinetic parameters, whereas that for native endogenous ligands is most probably under intrinsic homeostatic controls including rates of production, release, metabolism, and reuptake of ligands.

Classical receptor theory assumed that affinity and efficacy were independent parameters, suggesting that no consistent relationship existed between the affinity of a ligand and its ability to elicit a full response. A ligand with relatively low affinity ($<10^{-6}$ M) could still be a full agonist when a sufficient concentration interacted with the receptor. With the discovery of constitutive receptor activity, it seems that this lack of a consistent relationship is more reflective of an inability to measure receptor-mediated activity than a potency disconnect. Thus, one may conclude that all ligands have efficacy if the appropriate system is used to measure this parameter.

The relationship between the receptor and ligand in the classical lock and key model with the RL interaction resulting in no change in the receptor conformation essentially described a static situation. In 1958, however, Koshland described the induced fit model of the ES complex, where substrate binding to the enzyme caused a change in the three-dimensional structure of the protein, leading to a change in activity. Together with the pioneering work of Hill on hemoglobin allosterism, the concept of allosteric modulation of receptor function was proposed.

3.2 Rate Theory

On the basis of experimental data showing the persistence of antagonist-mediated responses and agonist "fade" where maximal responses occur transiently, to be followed by lesser responses of longer duration and agonist-mediated blockade of agonist effects, Paton modified the concept of occupancy to include a chemically based rate term. According to rate theory, it was not only the number of receptors occupied by a ligand that determined the tissue response but also the *rate* of RL formation. The resultant effect, E, was considered equal to a proportionality factor, ϕ, that included an efficacy component and the velocity of the RL interaction, V. Thus,

$$E = \omega Veq \tag{4}$$

The rate of RL formation, like that of neurotransmitter release, was measured in quantal terms with discrete "all or none" changes in receptor-mediated events. Pharmacokinetic considerations also play a major role determining the rate of RL formation.

The primary factor delineating occupancy and rate theory seemed to be the dissociation rate constant. Thus, if this factor was large, the ligand was an agonist; if the factor was small, reducing the quantal response to receptor occupancy, the ligand functioned as an antagonist. The kinetic aspects of rate theory did not, however, take into account the efficacy of transductional coupling and the potential for amplification after the initial binding event leading to its description as "a provocative conceptualization ... with limited applicability," a phrase that can be applied to many of the newer aspects of receptor theory.

4 Receptor Complexes and Allosteric Modulators

The studies of A.V. Hill on the binding of oxygen by hemoglobin demonstrated that identical binding sites on protein oligomers could influence one another such that the binding of the first ligand (in the case of hemoglobin in oxygen) facilitated the binding of a second, identical ligand, and so on for sequentially bound ligands such that the saturation curve describing the interaction of the ligand with its recognition site is steeper than that which would be predicted from classical Michaelis–Menten kinetics. The process of one ligand, homologous or heterologous, interacting with the binding of another is thought to occur by a cooperative, conformational change in the binding protein for the second ligand from a site adjacent to the ligand recognition site. Koshland's induced fit model extended these findings, leading to the development of two models of cooperativity or allosterism: the *sequential or induced fit* model described by Koshland; and the *concerted model* of Monod, Wyman, and Changeux, the key elements of which are outlined in Table 2.

Both models assume the existence of oligomeric protein units existing in two states that are in equilibrium with one another in the absence of ligand. Ligand binding induces a conformational change in the protein(s), moving the equilibrium of the two states to favor that with the higher affinity for the ligand. This in turn alters the kinetic and functional properties of the oligomeric complex. This model has been further refined in terms of *ligand-stabilizing conformational ensembles.*

The site on a receptor that defines its pharmacology and membership of

Tab. 2 Allosteric receptor models.

Monod, Wyman, Changeux Concerted Model
Receptor complex is a multicomponent oligomer comprised of a finite number of identical binding sites.
Subunits are symmetrically arranged each having a single ligand binding site.
Receptor complex exists in two conformational states, one of which has a preference for ligand binding.
Conformational transition state involves a simultaneous shift in the state of all subunits.
No hybrid states exist implying cooperativity.
Koshland Nemethy Filmer Sequential Model
Receptor complex is a multicomponent oligomer with symmetrically arranged protomers each with a single ligand binding site.
Protomers exist in two conformational states with transition induced by ligand binding.
Receptor symmetry is lost on ligand binding.
Hybrid states of the receptor complex can be stabilized by protomers.
Stabilization is equivalent to negative cooperativity.

a particular receptor superfamily, for example, 5HT, nicotine, and so on, is termed the *orthosteric site*. Ligands that bind to this receptor have a spatial overlap for the binding site such that their binding is mutually exclusive (unless an antagonist covalently binds to the orthosteric site). In contrast, the *allosteric site* (of which there may be more than one associated with a single orthosteric site and which can affect that site) is distinct from the latter in that ligands that bind to the allosteric site(s) can produce effects on ligand binding and efficacy to the orthosteric site through an indirect, conformational modulation of this site, which probably involves alterations in either the association or dissociation rates of the orthosteric ligand. While much of the early work on allosterism derived from studies on enzymes and ligand-gated ion channels (LGICs), it is now clear that GPCRs that were once considered as monomeric proteins with only an orthosteric site are now known to contain allosteric sites and can form oligomeric complexes. The concept of allosterism becomes more complex when considering multiple ligand sites on a receptor that have different pharmacological profiles, for example, when the ligand recognition sites are totally heterogeneous.

The identification of allosteric ligands that can have both positive and negative effects on the function of the orthosteric site occurred in a largely serendipitous manner. The first drug identified as an allosteric modulator was the BZ, diazepam, which has anxiolytic, hypnotic, and muscle relaxant activities and produces its effects by facilitating the actions of the $GABA_A$ receptor. Unlike the majority of directly acting $GABA_A$ receptor agonists, diazepam has a relatively safe side effect profile.

Allosteric modulators have three apparent advantages over drugs acting via orthosteric sites: their effects are saturable such that there is a ceiling effect to their activity, which results in a good margin of safety in human use; their effects are selective, and as mentioned earlier, frequently "use-dependent." Thus, the actions of an allosteric modulator occur only when the endogenous orthosteric ligand is present. In the absence of the latter, an allosteric

modulator is theoretically quiescent and may thus represent an ideal prophylactic treatment for disease states associated with sporadic or chronotropic occurrence. Finally, allosteric modulators are considered to be more specific in their effects, partly because of the nature of their binding sites that are distinct from the orthosteric site and partly because of the extent to which their effects depend on the degree of cooperativity between the allosteric and orthosteric sites.

The "cys-loop" family of LGICs including the $GABA_A$, glycine, $5HT_3$, and nicotinic cholinergic receptors are the best characterized of the allosterically regulated receptors. GPCRs are also subject to allosteric modulation. The families identified to date demonstrating this property include adenosine (P1) and α-adrenergic, dopamine, chemokine, $GABA_B$, endothelin, metabotropic glutamate, neurokinin-1, P2Y, and muscarinic cholinergic, as well as some members of the 5HT superfamily. The cone snail conotoxin, ρ-TIA is an allosteric modulator of the α_{1B}-adrenoceptor. Changes in GPCR function resulting from alterations in the ionic milieu also reflect the potential for allosteric modulation of receptor function.

5
Ternary Complex Models

The ternary complex model or TCM describes allosteric interactions between orthosteric and allosteric sites present on a single protein monomer. It can also be extended to reflect the other two-state interactions involving sites on adjacent proteins and the effect of signaling proteins, for example, G-proteins on the function of the receptor. Christopoulos notes allosteric interactions as being reciprocal such that the effects of a ligand A on the binding properties of ligand B also imply an effect on the binding of ligand B on the properties of ligand A. Similarly, because GPCRs alter the conformation of G-proteins to elicit transductional events and an alteration in cell function, changes in G-protein conformation and interactions alter receptor function, and this may be reflected in desensitization.

6
Constitutive Receptor Activity

A basic principle of receptor theory is that when a receptor is activated by a ligand, the effect produced by the ligand is proportional to the concentration of the ligand; for example, it follows the Law of Mass Action. It is now becoming apparent that receptors spontaneously form active complexes as a result of interactions with other proteins. This is especially true when receptor cDNA is *overexpressed* in cell systems such that the relative abundance of a receptor is in excess of that normally occurring in the native state, or associates with proteins that reflect the host cell milieu in which the receptor is expressed, rather than an intrinsic property of the receptor in its natural environment. This is a major issue in the characterization of ligand efficacy. A spontaneous interaction between receptor and effector can occur more frequently in a system where the proteins are in excess and where the normally present factors that control such interactions are absent. This is shown graphically in Fig. 3, where constitutive activity is shown in the range of 0–50 and where the theoretical effects of inverse agonists, full and partial, are shown. A quiescent system that is more reflective of classical receptor theory shows a full agonist, partial agonist, and what

is now defined as a neutral antagonist. Constitutive receptor activation has been described in terms of *protein ensemble* theory and in terms of *allosteric transition*, where changes in receptor conformation can occur through random thermal events.

7 Efficacy Considerations

Historical receptor theory describes a ligand efficacy continuum, with full agonism at one end and full antagonism at the other. Between the two ends of this continuum lie partial agonists that, as already noted, imply that ligands can also be partial antagonists. Antagonism *per se* implied that a ligand could bind to a receptor without producing any effect and limiting access to the native agonist – block receptor activation. Such a compound is now called a *neutral antagonist*.

With the ability to measure constitutive receptor activity, some compounds, for example, the β2-adrenoceptor antagonist, ICI 118551, inhibited constitutive activity, thus functioning as an *inverse agonist* or *negative antagonist*.

The actions of a competitive antagonism can be surmounted by the addition of increasing concentrations of the agonist ligand, resulting in a functional DR curve that undergoes a rightward shift with approximately the same shape and maximal effect (Fig. 1). Noncompetitive or uncompetitive antagonists interact at sites distinct from the agonist recognition site, and can modulate agonist binding either by proximal interactions with this site from a site adjacent to the recognition site or by allosteric modulation. The effects of noncompetitive antagonists are usually not reversible by the addition of excess agonist. This type of antagonism, whether competitive or noncompetitive, occurring at a distinct molecular target is known as *pharmacological antagonism* and involves the interactions between ligands and the receptor site (Fig. 4). In contrast, *functional antagonism* refers to a situation in which an antagonist that does not interact with a given receptor can still block the actions of an agonist of that receptor, and is typically measured in intact tissue preparations or whole animal models.

In the cartoon outlined in Fig. 4, neurotransmitter A released from neuron A interacts with A-type receptors on neuron B. Antagonist α can block the effects of A on cell B by interacting with A receptors. Antagonist α is thus a pharmacological

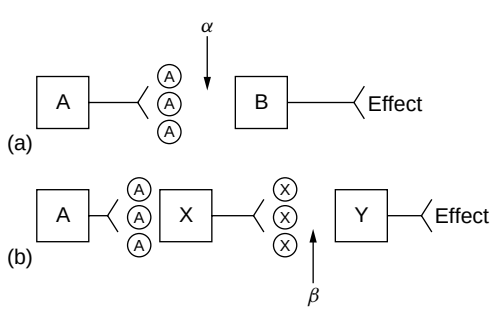

Fig. 4 Pharmacological versus functional antagonism. In panel (a), neurotransmitter A is released from neuron A, directly interacting with neuron B to produce a functional response. Antagonist blocks the effects of A, a direct pharmacological antagonism of the effects of A. In panel (b), neurotransmitter A is released from neuron A, directly interacting with neuron X, which in turn releases neurotransmitter X, which acts on cell Y to produce a functional response. Antagonist blocks the effects of X on cell Y, but in the absence of other data on the actions of antagonist, seems to block the functional effects of A because of the circuitry involved. Antagonist thus acts as a functional antagonist.

antagonist of A receptors. In the second example in Fig. 4, neuron A releases neurotransmitter A, which interacts with A-type receptors located on neuron X. In turn, neuron X releases neurotransmitter X that interacts with X-type receptors on neuron Y. Antagonist β is a competitive antagonist that interacts with X receptors to block the effect of neurotransmitter X, and in doing so, indirectly blocks the actions of neurotransmitter A. Antagonist β is thus a pharmacological antagonist of receptors for the neurotransmitter X, but a functional antagonist for neurotransmitter A.

In interpreting functional data in complex systems, it is always important to consider the possibility that a ligand has more than one effect mediated through a single class of receptor. For this reason, in advancing new ligands from *in vitro* evaluation to more complex tissue systems or animal models, it is extremely helpful to have a ligand-binding profile, for example, the activity of a compound at a battery of 90 or more receptors and enzymes (a Cerep profile), to fully understand any new findings. For instance, when a ligand for a new receptor is advanced to animal models and found to elicit changes in blood pressure, it is extremely helpful to know whether in addition to its defined activity at the new receptor, it has some other properties that relate to the blood pressure effects, rather than assuming that some unknown mechanism related to activation of the new receptor has cardiovascular-related liabilities.

8 Receptor Dynamics

Receptors, present in their active form at the cell surface are not static entities. Ligand binding, alterations in gene function and tissue dysfunction due to disease, trauma or aging, can alter the number, half-life, and responsiveness of receptors, channels, and transporters. Receptors turnover as a normal consequence of cell growth, with half-lives that vary between hours and days. Ligand binding can result in receptor internalization through phosphorylation-dependent events often initiated as a result of the ligand-binding process, exposing serine and threonine residues in the receptor protein. Histamine H_2 antagonists acting as inverse agonists in constitutively active systems can upregulate their cognate receptors, potentially increasing cell sensitivity. It is reasonable to assume that the target cells for endogenous effector agents, neurotransmitters, neuromodulators, and neurohormones operate under tonic control, one example being chronotropic, varying with the circadian rhythm of the organism. In contrast, the effects of exogenously administered ligands, for example, drugs, are rarely under normal homeostatic control, and as a result, their effects frequently become blunted on repeated administration or when they are administered in controlled release forms. Therefore, it should not be a surprise that the majority of effective therapeutic agents are antagonists of receptor function.

In Parkinson's disease, the presynaptic nerve cells in the *substantia nigra* that normally produce dopamine die as the result of an as yet unknown disease etiology. This defect results in a decrease in dopamine levels and a consequent hypersensitivity of postsynaptic responses as the cellular homeostatic events attempt to compensate for a lack of endogenous ligand. Agonists that rapidly desensitize a receptor, for example, ATP at the $P2X_3$ receptor and nicotine at the $\alpha 4\beta 2$ receptor,

appear to be antagonists because their net effect is to attenuate the normal receptor response.

9
Receptor Nomenclature

Even before the publication of the draft maps of the human genome, the techniques of molecular biology had already resulted in an explosion in the number of putative receptor families and subtypes within families, as well as classes of receptors known as *orphan receptors*. These were structurally related members of receptor classes for which the endogenous ligand and associated function was unknown.

New receptors have been (and continue to be) identified in different laboratories often simultaneously, leading to different names causing considerable confusion in the literature. The *International Union of Pharmacology* (IUPHAR) has undertaken the development of a systematic nomenclature system, based in part on naming families of receptors for their cognate endogenous ligands. The deliberations of the various committees enlisted to devise a systematic nomenclature are published periodically in *Pharmacological Reviews*. Compendiums are also regularly published, the most current and comprehensive being *The Sigma-RBI Handbook*.

10
Receptor Classes

Receptors can be divided into four major classes (Table 1): heptahelical, 7-transmembrane (7-TM) GPCRs, ion channels, transcription factor receptors and enzyme-associated receptors. Of these, the 7-TM receptors have historically represented the most fertile class for drug discovery, as these have been the most studied. This receptor superfamily includes receptors for the bioamine neurotransmitters (5-HT, dopamine, etc), peptide hormones, and lipid signaling molecules. Ion channels are further subdivided into LGIC, voltage-sensitive calcium channel (VSCC) and potassium (Kir), and ion/pH-modulated (ASIC) subtypes, all of which have similar but distinct, multimeric structural motifs. TRPV (transient receptor potential vanilloid) ion channels can be modulated by temperature, for example, vanilloid receptors. The largest group within the transcription factor receptor superfamily is the nuclear hormone receptors that include receptors for steroids (glucocorticoid (GR), progesterone (PR), mineralocorticoid (MR), androgen (AR)), and nonsteroids (thyroid hormones (TRs), PPARs, the retinoic acid receptors (RXR, RAR), and vitamin D_3 (VDR)). In addition, many orphan receptors have been identified with structural similarity to the known transcription factor receptors. The enzyme-associated receptor superfamily is a family of single- or multi-subunit proteins, which contain a subunit with a single transmembrane domain. The largest groups within this superfamily are the single-subunit–tyrosine kinase receptors, such as the PDGF (platelet-derived growth factor) and EGF (epidermal growth factor) receptors, and the multimeric complexes that utilize kinases such as the JAK-type kinases for their signal transduction (e.g. cytokine and growth hormone receptor families). Also included are receptors with serine/threonine kinase or guanylyl cyclase activity.

In addition to this multitude of receptor classes, further complexity in conceptualizing receptors as distinct, classifiable

entities is exemplified by recent findings related to GPCRs. By implicit definition, these receptors produce their physiological effects by coupling through the G-protein family. However, there are instances in which ion channels produce their effects by coupling through G-proteins. Receptor classification thus has few absolutes.

10.1 G-protein–Coupled Receptors

The 7-TM motif of GPCRs based on the X-ray structure of rhodopsin is a single protein comprised of approximately 300–500 amino acids with discrete amino acid motifs in the transmembrane (TM) regions and on the C-terminal extracellular loop, determining the ligand specificity of the receptor and those on the amino terminal and intracellular loops designating G-protein interactions. For example, RL interactions for many of the bioamine neurotransmitter receptors are thought to occur within a pocket in the 7-TM motif that is generically designated to lie between TMs III, IV, and VI. Posttranscriptional alternative splicing can alter the GPCR to create isoforms that may be species, tissue, and disease-state dependent.

The number of GPCRs present in the human genome, including orphan receptors, has been estimated to be 1000–2000 with over 1000 of these coding for odorant and pheromone receptors. Currently of greatest interest in drug discovery efforts are the 367 nonchemosensory GPCRs, over 100 of which remain "orphan receptors."

GPCRs are organized into four main families. *Family 1* includes the majority of GPCRs for peptide hormones, neurotransmitters, odorants, and a large group of orphans. Family 1 GPCRs can be divided by structural similarity into subfamilies: 1a that includes rhodopsin, adrenoceptors, thrombin, and the adenosine A_{2A} receptor, with a binding site localized within the 7-TM motif; 1b that includes receptors for peptides with the ligand-binding site in the extracellular loops, the N terminal, and the superior regions of the TM motifs; and 1c that comprises receptors for glycoproteins. Ligand binding to this receptor class is mostly extracellular. *Family 2* is morphologically related, but not sequence related to the 1c family and consists of four GPCRs activated by the hormones glucagon, secretin, and VIP-PACAP. *Family 3* contains four metabotropic glutamate receptors (mGluR) and three $GABA_B$ receptors. *Family 4* is a group of GPCRs that includes "frizzled" and "smoothened," both involved in embryonic development. Additional GPCRs have been described in invertebrate species that may define additional receptor families.

Interactions with heterotrimeric G-proteins and other associated signaling molecules have the potential for considerable complexity. Heterotrimeric G-proteins exist as complexes formed from α, β, and γ-type subunits, for each of which there are multiple subtypes, thereby offering a considerable variety of potential functional G-proteins. The α subunits are best characterized and often determine the signaling pathway activated There are four major G-protein α-subunit families that interact with GPCRs: (1) G_s that activates adenylyl cyclase; (2) $G_{i/o}$ that inhibits adenylyl cyclase and can also regulate ion channels and activation of cGMP PDE; (3) G_q that activates phospholipase C; and (4) G_{12} that regulates Na^+/H^+ exchange. The β and γ subunits are less well studied, but are also involved in signaling events such as interactions with ion channels and certain isoforms of phospholipase C, as well as in membrane trafficking and receptor interaction events.

10.1.1 G-protein–Associated or – Modulating Proteins

The cyclic nucleotide phosphodiesterases (PDEs) that are responsible for hydrolytic degradation of the cyclic nucleotides, cAMP and cGMP, exist in more than 15 isoforms. While the protein kinases, PKA and PKC, are responsible for protein phosphorylation, the GPCR kinases, GRK 1–6, are responsible for GPCR phosphorylation, and the protein phosphatases are responsible for dephosphorylation, the latter potentially numbering in excess of 300, significantly increasing the complexity of GPCR-associated signaling processes.

Superimposed on these signaling proteins are the calmodulins that mediate calcium modulation of receptor function, the β-arrestins, involved in inactivation of phosphorylated receptors (6 members), a group of proteins termed RGS (regulators of G-protein signaling), RAMPs, and a protein known as Sst2p that is involved in receptor desensitization.

Given the number of GPCRs present in the human genome, including orphan receptors, multiple G-protein subtypes, phosphatases, and the various GPCR-associated signaling proteins described earlier, there is obviously considerable scope for complexity in cell signaling associated with the GPCR family. GPCRs can also form homo- and heteromeric forms (e.g. $GABA_B$, adenosine A_1 and A_{2A}, angiotensin, bradykinin, chemokine, dopamine, metabotropic glutamate, muscarinic, opioid, serotonin, and somatostatin), further increasing the potential complexity of ligand-driven GPCR signaling processes and offering an opportunity to explore new targets in medicinal chemistry.

Applying an evolutionary trace method (ETM) to assess potential protein–protein interactions, Dean et al., identified functionally important residue clusters on TM helices 5 and 6 in over 700 aligned GPCR sequences. Similar clusters were found on TMs 2 and 3. TM 5 and 6 clusters were consistent with 5,6-contact and 5,6-domain swapped dimer formation. Additional application of ETM to 113 aligned G-protein sequences identified two functional sites: one associated with adenylyl cyclase, and regulator of G-protein signaling (RGS) binding, and the other extending from the ras-like to helical domain that seems to be associated with GPCR dimer binding. From such findings, it was concluded that GPCR dimerization and heterodimerization occur in all members of the GPCR superfamily and its subfamilies.

From these findings, potential new approaches to ligand design and by extrapolation, drug discovery, include: (1) antagonists that can act by inhibiting dimer formation, for example, transmembrane peptide mimics; (2) bivalent compounds/binary conjugates; and (3) compounds targeting the GPCR–G-protein interface.

10.2 Ligand-gated Ion Channels

Ion channels consist of homo- or heteromeric complexes numbering between three (P2X) and eight (Kirs) subunits. Examples of these are the $GABA_A$/benzodiazepine and NMDA/glycine receptor, neuronal nicotinic receptors (nAChR), and P2X receptors.

10.2.1 $GABA_A$/Benzodiazepine (BZ) Receptor

$GABA_A$/benzodiazepine receptor is an LGIC that is the target site for numerous clinically effective anxiolytic, anticonvulsant, muscle relaxant, and hypnotic drugs,

which produce their therapeutic effects by enhancing the actions of the inhibitory neurotransmitter, GABA. It is a pentameric LGIC, the constituent subunits of which are formed from a family of six, four, one, and two subunits, leading to the potential existence of several thousand different pentamers. The functional receptor complex contains a $GABA_A$ receptor, a BZ recognition site, and by virtue of its pentameric structure, a central chloride channel. Allosteric recognition sites on this complex include those for ethanol, avermectin, barbiturates, picrotoxin, and neurosteroids like allopregnanolone. The pharmacology and function of the allosteric sites depends on the actual subunit composition. The 1 subunit, which is present in nearly 60% of $GABA_A$ receptors, in mouse brain mediates the sedative, anticonvulsant, and amnestic effects of diazepam; the 2 subunit, the antianxiety effects of diazepam; and the 5 subunit, associative temporal memory. With this knowledge, rather than screening new ligands for the $GABA_A$ receptor in various animal models to derive a profile for a nonsedating anxiolytic, by understanding the structural determinants of BZ interactions with 1 and 2 subunits and designing compounds that preferentially interact with the latter, the process of designing novel ligands can be considerably enhanced.

10.2.2 The NMDA Receptor

The NMDA receptor is a member of the glutamate receptor superfamily that mediates the effects of the major excitatory transmitter, glutamate. It is composed of an NMDA receptor, a central ion channel that binds magnesium, the dissociative anesthetics, ketamine, and phencyclidine (PCP), the noncompetitive NMDA antagonist, dizocilpine (MK 801), glycine and polyamine binding sites, activation of which can markedly alter NMDA receptor function, and some 70 other ancillary proteins, the physiological function of which remains to be determined. The activation state of the receptor can define the effects of the allosteric modulators. Thus, some are termed *use-dependent*, reflecting modulatory actions only when the channel is opened by glutamate.

10.2.3 The Nicotinic Cholinergic Receptor, nAChR

The nicotinic cholinergic receptor is another pentameric LGIC composed of distinct functional regions and subunits. The subunit composition of the receptor varies, imparting different functionality when the channel is activated by nicotine, or by the endogenous ligand, ACh. Allosteric modulators of neuronal nAChRs include dizocilpine, avermectin, steroids, barbiturates, and ancillary proteins.

10.2.4 The P2X Receptor

The P2X receptor is an LGIC responsive to ATP that functions as a trimer formed from a family of seven subunits that can form both homo- and heteromers leading to the existence of at least seven distinct receptor subtypes.

Little is known regarding the structural elements involved in the ligand pharmacology and function of LGICs, such that signaling transduction pathways have not been extensively characterized. For nAChRs, it is known that the recognition site for ACh is formed between two subunits. Thus multiple orthosteric sites for ACh are possible, depending on the types of subunit forming the receptor. Many new classes of LGIC are still being discovered.

10.3 Steroid Receptor Superfamily

The steroid receptor superfamily includes the glucocorticoid (GR), progesterone (PR), mineralocorticoid (MR), androgen (AR), thyroid hormone (TR), and vitamin D_3 (VDR) receptors. These receptors bind steroid hormones and are then translocated to the nucleus, where they bind to hormone responsive elements on DNA promotor regions to alter gene expression. While steroids are very effective anti-inflammatory agents, they have a multiplicity of serious side effects that limit their full use.

The antiestrogen, tamoxifen, is the most commonly used hormonal therapy for breast cancer and has demonstrated positive effects on the cardiovascular and skeletal systems of postmenopausal women but is associated with an increased risk of uterine cancer. Tamoxifen is described as a selective estrogen receptor modulator (SERM) with a tissue selective profile dependent on the different distribution of the α- and β-subtypes of the estrogen receptor (ER-α and ER-β) that activate and inhibit transcription respectively. These selective effects have been ascribed to differential interactions with gene promotor elements and coregulatory proteins depending on whether the ER interacts directly, or in a tethered manner with DNA.

10.4 Intracellular Receptors

Members of the intracellular receptor family include the cytochrome P450 (CYP) family, the SMAD family of tumor suppressors, intracellular kinases and phosphatases, nitric oxide synthases (NOS), caspases, the RXR, RAR superfamilies, receptor activated transcription factors (RAFTs), and signal transducers and activators of transcription (STATs) such as AP-1, NFκB, NF-AT, STAT-1, PPARs, various hormone responsive elements on DNA and RNA promoters, and ribozymes. These represent a bewildering number of potential drug targets especially in the metabolic disease and cancer areas. Given their roles in normal cell function, it will be a challenge to ensure specificity in ligands interacting with these targets.

10.5 Non-GPCR-linked Cytokine Receptors

Cytokines are polypeptide mediators and are involved in the inflammatory/immune response. There are three cytokine families: *hematopoietin*, which includes IL-2–IL7, L-9–13; IL-15–IL-17; IL-19, 1L-21, IL-22, GMCSF, GCSF, EPO, LIF, OSM, and CNTF, with primary signal transduction through the Jak/STAT pathway; *tumor necrosis factor*, comprising the receptors, TNFRSF1–18 that signal through NFB, TRAF, and caspases; and the *interleukin 1/TIR* family that includes IL-1RI and IL-1RII, IL-1Rrp2, and IL1RAPL. TIGGR-1, ST2, IL-18, and Toll 1–9 also signal through NFB and TRAF.

10.6 Orphan Receptors

Orphan receptors have been generally defined as proteins with a receptor motif that lack both a ligand and function. Approximately 160–300 orphan GPCRs are thought to be in the draft sequence of the human genome and intense efforts are currently ongoing to identify the ligands for these and their function as novel intellectual property for the drug discovery process. Orphan receptor validation can be done using expression profiling to

identify tissues rich in the expression of the receptor of interest, and a technique known as *reverse pharmacology* that can be used to identify a ligand for the orphan receptor. In the latter, the orphan receptor is used as "bait" to bind selective ligands. These can then be used to further characterize receptor function. Nearly 30 orphan GPCRs have been validated in this manner. While most of the current interest on orphan receptors is focused on GPCRs because of the considerable body of existing knowledge about this receptor class, it is anticipated that orphan receptors will also be discovered for other receptor classes, for example, orphan nuclear receptors.

The orphan receptor approach to drug discovery is exemplified by the orphanin/FQ receptor, ORL1, a structural homolog of the opioid receptor. Identified in 1995 using a homology-based screening strategy, ORL1 had low affinity for known opioid ligands. A novel heptadecapeptide ligand for the ORL1 receptor, orphanin/FQ, was subsequently isolated from brain regions rich in ORL1 providing the key tool to validate the target and identify a functional role for the receptor in stress-related situations in animal models. From an intensive screening program, an antagonist of this receptor was identified, Ro 64–6198, which represented a novel anxiolytic/antidepressant drug candidate.

10.7 Neurotransmitter Transporters

Neurotransmitter transporters (NTs) are responsible for terminating the effects of neurotransmitters by removing them from the extracellular space. NTs are integral membrane proteins present on both the plasma membrane and on intracellular vesicles where they effect vesicular packaging of neurotransmitters. Their activity is dependent on the Na+ intracellular/extracellular gradient. Most NTs (e.g. dopamine transporters (DAT), norepinephrine transporters (NET), serotonin transporters (SERT) also require Cl^- and are termed *Na^+/Cl^- dependent NTs*, (SCDNTs) Transporters for glutamate also require K^+ (SKDGTs). These proteins bind their cognate neurotransmitters with high affinity and transport them across the plasma membrane into the cell. GABA transporters comprise the GAT family.

NT inhibitors represent a major class of drugs producing their effects by blocking neurotransmitter uptake, thus potentiating the actions of endogenous neurotransmitters. As such, NTs can be considered as indirect receptor agonists. Inhibitors of SERT (also known as *selective serotonin reuptake inhibitors*, SSRIs) are major antidepressants, for example, fluoxetine. Mixed SERT/NET inhibitors (SNRIs) are also antidepressants and analgesics while DAT inhibitors are well-known stimulants, e.g. cocaine, amphetamine and so on. Inhibitors of GABA transporter (GAT) are anticonvulsants (e.g. gabatril).

10.8 Neurotransmitter Binding Proteins

Binding proteins for corticotrophin releasing factor (CRF) and acetylcholine (AChBP) have been identified and exploited as potential drug discovery targets. These proteins can act as reservoirs or "decoy receptors" for their respective neurotransmitters, playing a general role in buffering synaptic message transfer.

10.9 Drug Receptors

Most drugs interact with receptors (or enzymes) for which the natural ligand

(or substrate) is known. There are, however, a number of receptors, distinct from the evolving class of orphan receptors, for which the synthesized drug is the only known ligand. The best example of this is the central benzodiazepine receptor present on the $GABA_A$ ion channel complex that was originally identified using radiolabeled diazepam. Because this is the site of action of the widely used BZ anxiolytic drug class, this is a *bona fide* receptor with clinical relevance. However, despite considerable efforts and a number of interesting candidate compounds, no endogenous ligand has yet been unambiguously identified which would represent an endogenous modulator of anxiety acting through this site. Other examples of drug receptors for which synthetic ligands were identified before the endogenous agonists, are the cannabinoid receptors, CB_1 and CB_2, through which the following compounds act: 9-Δ-tetrahydrocannibol, the active ingredient of the psychoactive recreational drug/analgesic/antiemetic drug, marijuana; the vanilloid receptor, VR-1, the known ligand for which is capsaicin, the ingredient in red pepper that evokes heat sensation; and the opioid receptor family, the site of action of morphine and other derivatives of the poppy. For each of these drug receptors, endogenous mammalian ligands, anandamide, the endovanilloid, *N*-arachidonyldopamine (NADA), enkephalins, and orphanin/FQ have been identified.

11
Molecular Biology of Receptors

Recombinant DNA methods have been extensively used to isolate and analyze the sequence of numerous receptors from various mammalian complementary DNA (cDNA) libraries using the polymerase chain reaction (PCR) technique to clone receptors that can then be expressed in various prokaryotic and eukaryotic cell lines and *Xenopus* oocyte. Conserved regions in receptors that are involved in ligand binding, coupling to transductional systems and ion channel formation, have thus been identified. Receptor cloning, sequencing, and expression are now an intimate part of the drug discovery process.

From the receptor/ligand modeling standpoint, the ability to specifically alter the structure of cloned receptors through the modification of a small number of nucleotides, the process known as *site-directed mutagenesis*, the removal of specific regions of the receptor gene "deletion mutagenesis," and the construction of hybrid (chimeric) receptor proteins can provide additional clues on the relative importance of individual amino acids in the process of ligand recognition and receptor function. These techniques are also used to manipulate receptors to produce recombinant expression systems with better cell surface expression, more convenient signal transduction coupling, and other qualities important for efficient drug discovery.

12
Functional Genomics: Target Validation/Confidence Building

The identification of novel receptor targets from the human genome and their subsequent use in the drug discovery process requires that the target be validated. Their validation requires that the ligand and function of the receptor be known, and this topic has been addressed in relation to orphan receptors mentioned earlier.

Target validation, especially that related to novel, genomically derived targets, is

a highly complex and resource intense process, adding significantly to the cost of drug discovery. In addition to orphan receptors, identified on the basis of their homology to members of known receptor classes, novel genomic targets can be identified by drug-related differential gene expression analysis, population genetic approaches, or a combination of both. Once a gene is associated with a disease, its protein product needs to be characterized and a biochemical function, for example, receptor, enzyme, or transporter, ascribed to it. The probable function of the protein can then be assessed and its cognate ligand or substrate identified. Not all proteins identified in this manner are obviously involved in mammalian cell function. The protein product of a novel lithium-related (NLR) gene identified in mice exposed to lithium has sequence homology to a bacterial nitrogen permease. The potential role of this protein in the etiology of bipolar affective disorder remains unclear. Patients can also be genotyped for susceptibility related to a drug target or drug side effects, to identify "responsive" patients for more efficient and safer drug use.

To identify what are termed *druggable* gene products, the still somewhat mystical process of functional genomics can be used to construct protein–protein interaction maps, to identify other proteins in a pathway/network that can represent more facile entry points to the protein cluster associated with the genetically identified target and with the disease state. Using techniques like yeast two-hybrid and *gal*-pull down, the function of an uncharacterized protein can be assigned on the basis of the known function of its interacting partners, involving the extensive use of bioinformatics tools.

Targeted gene disruption, antisense, and ribozyme inhibition (RNAi) are other techniques for assessing the phenotype of a given gene. Antisense to the rat $P2X_3$ receptor had marked hyperalgesic activity showing an unambiguous role for this receptor in chronic inflammatory and neuropathic pain states. Transgenic animals, in which the function of a gene is knocked out by genetic techniques, are less useful in understanding the function of a protein. In addition to being restricted to the mouse, this approach is high in cost and time (taking 1–2 years to generate sufficient animals for evaluation), and in many instances, the absence of a given protein has no overt effect on the phenotype of an animal because of compensatory changes during the developmental phase. Alternatively, the gene knockout may lead to an animal that has a limited, if any, life span or reproductive capability.

Another target validation approach involves the use of ligands to define the function of a new target. Already discussed in regard to orphan receptors, other ligand-directed target validation models fall under the rubric of *chemical genomics*. A key to identifying useful ligands that selectively interact with proteins of unknown function to define their function is a sufficient level of pharmacophore diversity to ensure success.

Assuming that a novel protein target is amenable to HTS approaches to identify its cognate ligand, that the ligand is identified, and that other approaches are able to define a putative function for the target, the next step in target validation is to identify a "druggable" molecule that can be advanced to the clinic. For ORL1, Ro 64–6198 is an excellent example of this approach.

This raises the key test of *target validation* or *target confidence building*. Does a compound that has the appropriate potency, selectivity, and ADME properties, that is active in "predictive" animal models, and is free of other systems toxicology work in the targeted human disease state? This is the ultimate, and only, validation of a target-based drug discovery program. All events leading up to this point are more accurately defined as target confidence building. With the considerable compound attrition rates in moving from animals to diseased humans, this has not proven to be a predictable transition. A case in point is that of the NK1 receptor activated by the peptide, substance P.

Over the past decade, data from both animals and humans has implicated this receptor in a variety of human pain states including migraine and neuropathic pain. As the result of several successful HTS campaigns run in parallel at Pfizer, Lilly, Merck, and Sanofi Aventis, a number of highly efficacious and selective antagonists of the NK1 receptor were identified which were then chemically optimized for use as drug candidates. These compounds, with varying degrees of potency, were active in animal models of pain and were free of overt side effects in phase I human clinical trials. However, they uniformly failed in phase II studies as novel analgesic agents in patients with various pain conditions.

The reasons for this remain unknown and have focused on the relevance of animal pain models to the human condition, for example, a lack of understanding of the true human disease condition and various nuances of substance P signaling pathways. However, before the results with NK1 antagonists, drug discovery in the analgesia area was considered one of the most robust. All known analgesics, for example, aspirin and morphine, were active in one or more of the animal pain models; new receptors could be mapped in pain pathways and their function assessed using knockout and antisense procedures, and the occupancy of receptors in human brain and pain-sensing pathways could be noninvasively imaged. The only limitation was whether the side effect profile of a putative analgesic agent acting on a novel target or suboptimal ADME characteristics would limit human exposure.

On the basis of the available data, the NK1 receptor has not been validated as a target for pain treatment despite an overwhelming body of robust preclinical data. This example thus serves to underline the many significant challenges of validating targets in the drug discovery process at the present time.

A second example, more directly related to the human genome, is the search for genes that are associated, and by inference may be causative, in producing the psychiatric disorder schizophrenia. This disease affects nearly 1% of the population and has a strong genetic association. From studies comparing schizophrenic patients with individuals lacking symptoms of the disorder, LOD scores of 2.4–6.5 on markers between or close to markers D1S1653 and D1S1679 on chromosome 1q have been reported. An LOD score of above a numerical value of 3 is considered to be indicative of a relevant association akin to a significant P value ($P < 0.05$). With this information, a search for the gene on chromosome 1q and the delineation of its function in humans would be a logical approach to finding novel genomic targets that could lead to new drugs that would more effectively and safely treat schizophrenia. However, a subsequent study using eight individually collected schizophrenia populations and a sample set that was "100% powered to detect a large genetic effect under the

reported recessive model,'' showed no evidence for a linkage between chromosome 1q and schizophrenia; a finding that could not be ascribed to ethnicity, statistical approach, or population size. Another group using two prefrontal cortex tissues from two separate schizophrenic populations showed an upregulation of apolipoprotein L1 gene expression that was not seen in tissue from patients with bipolar disorder or depression. The genes related to apolipoprotein L1 gene expression are clustered on the chromosome locus 22q12, providing another target for the functional genomic approach to target discovery and validation. Another locus at chromosome 3q has been reported.

To the researcher embarking on a well-funded, science-driven approach to new targets for schizophrenia in 1997, these new findings in 2002, as well as others showing gene association with schizophrenia on chromosomes 6 and 13, would give pause to wonder how best to proceed. Do all five locations represent *bona fide* targets for functional genomics? Should one continue work on the original chromosome 1 findings in light of the failure to replicate? Is the only real validation, in light of conflicting findings, to proceed to the identification of multiple compounds that can be tested in schizophrenic populations – to validate the genome-based approach? Given that the cost of initiating a project, finding leads, optimizing these, and running a single clinical candidate to phase IIa clinical proof of principle is in the range of $24–28 million, how many organizations can afford the $120 million+ cost to undertake scientifically logical yet financially prohibitive strategies? These are difficult questions that can be applied to many other psychiatric and neurological diseases, as well as to any other disease state with a potential genetic causality. One answer to these questions obviously points to multifactorial genetic causes in the genesis of many diseases and thus negates the overly simplistic ''one gene, one disease'' mantra that heralded the age of drug discovery based on the human genome map. Strategies resulting from the answers to these questions also make the process of genome-based target and drug discovery a much more costly endeavor, with good evidence for two to three or more potential targets for each disease that need to be examined in parallel. Hopefully, the quality of the target finally validated using compounds in the human disease state will be a magnitude of order superior to currently existing drugs and will thus justify the cost of this approach.

13 Compound Properties

Once a target has been identified, the challenge then becomes the identification of compounds that selectively interact with the target and can be used as the basis of a lead optimization program to identify potential drug candidates.

An ideal compound should have the appropriately unique recognition characteristics to impart affinity and selectivity for its target, have the necessary efficacy to alter the assumed deficit in cell function associated with the targeted disease state, be bioavailable, metabolically, and chemically stable (drug like, be chirally pure, and be easy and cost-effective to synthesize).

13.1 Structure-activity Relationships

The structure-activity relationship (SAR) of a compound series is a means to relate changes in chemical diversity to the

biological activity of the compound *in vitro* and *in vivo*, as well as the pharmacokinetic (gut/blood brain barrier transit, liver metabolic stability, plasma protein binding, etc.) and toxicological properties of the molecule. These SARs, when known, are frequently distinct such that changes that improve the bioavailability of a compound often decrease its activity and/or selectivity. Compound optimization is thus a highly iterative and dynamic process.

For the purpose of the present review, SAR is used to describe classical compound efficacy unless otherwise stated. Quantitative SAR (QSAR) involves a more mathematical approach involving neural networks and computer-assisted design.

Following the characterization of the SAR and the documentation of the effects of different pharmacophores and various substituents on biological activity, it is possible to theoretically model the way in which the ligand interacts with its target and thus derive a two- or three-dimensional approximation of the active site of the receptor or enzyme. Computer-assisted molecular design (CAMD) techniques can then be used to predict – sometimes successfully – the key structural requirements for ligand binding, thus defining those regions of the receptor target that are necessary for ligand recognition and/or functional coupling to second messenger system to permit the design of new pharmacophores.

The SAR can be further delineated in terms of the type of activity measured as the readout. *In vitro*, this can be displacement of a radiolabeled ligand from a receptor, receptor activation as measured in a functional assay, blockade of receptor function by an antagonist ligand, and so on. An additional ligand property is that of selectivity, the degree to which a ligand interacts with the target of choice compared with related structural targets. The degree of selectivity typically determines the side effect profile of a new compound, given that the targeted mechanism itself does not produce untoward effects when stimulated beyond the therapeutic range. As noted, the development of a ligand binding profile for a ligand active at a new target is very useful in assessing its effects in more complex tissue systems.

13.2 Defining the Receptor–Ligand Interaction

The complex physiochemical interactions that describe the interaction of a small molecule with a protein, despite the many sophisticated technologies used to study this interaction, are still highly empirical, being implicitly defined by the SAR for a series of active and less active compounds. In an increasing number of instances, however, the ability to clone, express, and readily derive crystals of a receptor or enzyme and analyze the interaction of a ligand or substrate with these using X-ray crystallographic, NMR, and amino acid point mutation approaches, has provided information on the actual topography of the selected drug target that can then be used in *de novo* ligand design.

There are a variety of approaches to derive information on the RL interaction for use in understanding the key features required in ligand necessary to dictate a potent and selective interaction with its target. Analysis of the interactions of a series of structurally distinct pharmacophores, agonists, and antagonists with the target can be combined with point mutation changes in the target to elucidate key amino acids involved in compound recognition. This can be complemented by X-ray crystallographic and

NMR-derived data to design novel pharmacophores. However, many proteins of interest, especially membrane-bound receptors and ion channels, are not available in the soluble form amenable to the use of these techniques such that the design of new compounds is based on the conceptualization of the target, for example, virtual receptors. There are also limitations to structure-based design approaches, for example, approximations of the hydration state of the isolated protein, the impact of removal from its native environment, and three-dimensional structural issues with recombinantly derived proteins. Nonetheless, these technologies have been useful in compound design. More recently, automated, high-throughput approaches have been applied to X-ray crystallography, providing a more rapid means to generate information on multiple ligand interactions with a given crystal.

13.2.1 Receptor-binding Assays

Until the 1960s, newly synthesized compounds and compounds isolated from natural sources were assessed by a mixture of *in vivo*, whole animal screens, and classical tissue assays. While many useful therapeutic agents were identified by this approach, the cost, in terms of compound quantity as well as time and animal use was considerable. *In vivo* test paradigms also suffered from the possible elimination of interesting compounds on the basis of unknown pharmacokinetic properties as test paradigms were usually routine in terms of timing, and as a result, many potentially interesting compounds that exhibited short plasma half-lives were considered "inactive" because data on their actions was sought after their peak plasma concentration. This type of screening approach also provided little useful chemical information about discrete interaction of the drug with its target, limiting the design of analogs. The specificity of the response, ignoring caveats related to pharmacokinetics, was not ideal because the mechanism inducing the overt response and potential points of intervention to block the response were unknown.

This empirical approach would predict, in the absence of any data related to specific interactions with a molecular target, that because β-adrenoceptor antagonists lower blood pressure, then any compound that lowers blood pressure is by definition a β-adrenoceptor antagonist. The 1970s saw the development of a number of *in vitro* biochemical screens that moved the measurement of the RL interaction a little closer to the molecular level. Nonetheless, the major challenge was to develop assays that measured the RL interaction independently of "downstream" events such as enzyme activation and second and third messenger systems. By such means, the ability of a compound to bind to a receptor could be determined on the basis of the SAR and thus provide the chemist with a more direct means to model the RL interaction.

Snyder, building on pioneering work by Roth, Cuatrecasas, and Rang developed radioligand binding as a valuable tool in the drug discovery process in 1973. This led to an explosion in the identification and characterization of new receptors and their subtypes, which was enhanced by the application of tools of recombinant DNA technology to the process.

While the technique of radioligand binding has been largely supplanted by activity-based assays in many high-throughput settings, it remains the gold standard for compound characterization and SAR development. It is done by measuring the RL interaction (Equation 1) *in vitro* using a radioactive ligand, R*, to bind with high

affinity and selectivity to receptor sites. The interaction of unlabeled ("cold") ligands with the receptor competes with the radioligand, decreasing its binding. This simple technique revolutionized compound evaluation in the 1970s, allowing SARs to be determined with milligram amounts of compound in a highly cost- and resource-effective manner. At steady state, the RL* complex can either be separated out from free radioligand using filtration or assayed using a scintillation proximity assay.

The parameters measured in a binding assay are the dissociation constant, K_d, the reciprocal of the affinity constant, K_a. The K_d is a measure of the affinity of a radioligand for the target site: the B_{max}, usually measured in moles per milligram protein, a measure of the concentration of binding sites in a given tissue source and the IC_{50} value. The K_d and B_{max} values can be determined using a saturation curve where the concentration of radioligand is increased until all the ligand recognition sites are occupied, or by measuring radioligand association and dissociation kinetics; K_d is the ratio of the dissociation and association rate constants.

The IC_{50} value is the concentration of unlabeled ligand required to inhibit 50% of the specific binding of the radioligand. This value is determined by running a competition curve (Fig. 5) with a fixed concentration of radioligand and tissue, and varying concentrations of the unlabeled ligand. To accurately determine the IC_{50}, it is essential that sufficient data points be included. As shown in Fig. 5, if the data used to derive the IC_{50} value are clustered over a range that reflects 40 to 60% of the competition curve, much useful information is lost. Ideally, the competition curve should encompass the range of 10 to 90% of the competition curve and include a minimum of 20 data points. On the basis of Michaelis–Menten kinetics, when binding is the result of the interaction of the displacer with one recognition site, 10 to 90% of the radioligand is inhibited over an 82-fold concentration range of the displacer. The slope of a competition curve can then be analyzed to assess the potential cooperation of the RL interactions. When binding is complex resulting in the interaction of the displacer with more than one recognition site, a greater than 82-fold concentration of displacer is required to inhibit the same 10 to 90% of specific radioligand binding.

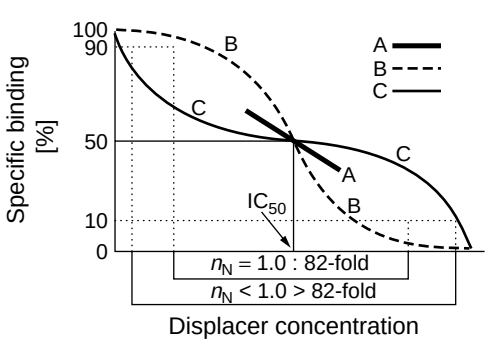

Fig. 5 Measuring ligand interactions in a receptor-binding assay. Binding of the radioligand is 100%. In curve A, close to the 50% point, an IC_{50} value can be obtained but ignores complexities of the displacement curve. More complete displacement, as in curves B and C, provides more information on the ligand. Displacement curve B has a Hill coefficient of unity requiring an 82-fold difference in displacer concentration to displace 10 to 90% of binding. Displacement curve C has a Hill coefficient of less than one, requiring a greater than 82-fold difference in displacer concentration to displace 10 to 90% of binding, and indicating the presence of more than one site. By extending the range of concentrations used, curves B and C can assess the possibility of multiple sites being present.

The IC_{50} value for a given compound is dependent on the assay conditions: the concentration of the radioligand used, the receptor density, and the affinity of the receptor, the K_d, for the radioligand. To compare the activity of a "cold" ligand across different radioligand binding assays, the Cheng–Prusoff equation is used to compensate for differences in K_d and the radioligand concentration to obtain a K_i value derived by the relationship:

$$K_i = \frac{IC_{50}}{1} + \frac{[L]}{K_d} \quad (5)$$

where [L] is the concentration of radioligand used and K_d is the dissociation constant for the radioligand at the receptor. This relationship thus corrects for inherent differences in assay conditions.

Binding assays can be rapidly used to assess compound recognition characteristics but are generally limited in their ability to delineate agonists from antagonists, especially in a high-throughput setting.

13.2.2 Functional Assays

Biochemical assays involving the measurement of cAMP production or phosphatidylinositol turnover have given way in HTS scenarios to reporter systems in which receptor activation or inhibition can be measured using a fluorescence-based readout. This depends on the use of calcium sensing dyes coupled with real-time measurement using charge coupled device (CCD) cameras and data capture. While many types of plate reader are available, a widely used system is the fluorescence imaging plate reader (FLIPR). Using a 96- or 384-well microtiter plate format, the throughput on an FLIPR is such that compound libraries of 0.5 million distinct compounds can be assayed in weeks.

Other approaches to functional assay include various reporter gene constructs in which formation of the RL complex leads to the expression of a gene that produces a response that can be read immediately. These include various luciferase reporter gene or aequorin-based assays that produce light that can be measured by multiple photodiode arrays, and α-galactosidase reporter gene that leads to a colorimetric readout. Frog melanocytes transfected with GPCR cDNA represent yet another approach to determining whether a ligand is an agonist or antagonist that is independent of a reporter gene construct. Addition of melatonin to a transfected oocyte reduces intracellular cAMP concentrations resulting in the aggregation of the pigment in the cells. Agonist stimulation of the transfected cell increases cAMP resulting in pigment dispersal, a reaction that can not only be immediately determined visually on a plus/minus basis but can also be quantified in terms of light transmission.

The ability to measure colocalization of two cellular components in living cells is an increasingly powerful technique in understanding ligand–receptor and protein–protein interactions. Two such approaches are fluorescence resonance energy transfer (FRET) that measures the proximity of two proteins through the use of a luciferase tag on one partner and green fluorescent protein (GFP) on the other, and fluorescence polarization (FP) that relies on a polarized excitation light source to illuminate a binding reaction mixture. The smaller partner (e.g. the ligand) must be fluorescently labeled, and if this is unbound then it will, by tumbling in solution, emit depolarized fluorescence. This is thus a means of looking at the amount of unbound ligand in the binding mixture and has the

benefit of not requiring any washing steps. Current systems do not work well with turbid solutions or fluorescent compounds, although the latter has largely been dealt with through the use of red fluorescent dyes.

13.3 Receptor Sources

The choice of a tissue or cell line as a receptor source has a significant impact on the data generated. The natural receptor concentration in most tissues is in the femtomole to picomole range; brain tissue has a much higher density of receptors because of more extensive nerve innervation. Expression of the drug target in a cell line can, however, lead to differences in the number of receptors expressed per clone, a factor dependent on the relative proportion of transient to stable expressed cells and the passage number of the transfectants. Thus, the number of receptors can vary, affecting the apparent activity of unknown ligands that compete for binding with the radioligand.

As the drug targets of greatest interest are those in the human, the use of a human receptor or enzyme would seem ideal in defining the SAR of a potential drug series. The drawback, however, is that nearly all the toxicology and safety studies done in preparing a compound for clinical trials are conducted in rodents, dogs, and nonhuman primates. If there are no species differences between rat and human, this testing becomes a moot point. If on the other hand, the human target is substantially different from that in rat or dog or monkey, and there are many examples of this, the safety and toxicology studies may be conducted on a compound that has limited interactions with the drug target in species other than human. One approach is to incorporate human receptor orthologs into mice.

The use of transfected cell lines in compound evaluation can lead to a number of potential artifacts. Activation of transfected receptor may result in an increase in cAMP, a second messenger effect that may already be known to be a consequence of ligand activation of the receptor in its natural state. It is also possible that the transfected receptor may activate a cell-signaling pathway in the transfected cell that is not linked to the receptor in its normal tissue environment. In this instance, the second messenger readout actually functions as a "reporter," a G-protein–linked phenomenon that results from the introduction of the cDNA for a GPCR and the generic or promiscuous interaction of the receptor with the G-protein systems. The introduction of the cDNA for any GPCR may then act to elicit a similar response. It is then advisable to use caution in extrapolating events occurring in the transfected cell to the physiological milieu of the intact tissue, and there is at least one case in which a compound identified as an antagonist in a cell system overexpressing the receptor of interest was subsequently found to be a partial agonist after it exacerbated disease symptomatology in phase II clinical trials.

13.4 ADME

In evaluating new compounds, it is only in the past 15 years that the ability of a compound to reach its putative site of action has been a priority in the discovery phase of compound identification. For many years, it was naively assumed that there was a generic approach that could be used on compounds with poor bioavailability that would turn them into

drug candidates. With attrition rates of lead compounds in the clinical development process of 50 to 60% range, with one estimate of greater than 90%, this was clearly not the case. Hodgson has noted, "a chemical cannot be a drug, no matter how active nor how specific its action, unless it is also taken appropriately into the body (absorption), distributed to the right parts of the body, metabolized in a way that does not instantly remove its activity and eliminated in a suitable manner – a drug must get in, move about, hang around and then get out." The factors involved in defining *in vivo* activity form the basis of Lipinski's "Rule of 5." In this widely cited, retrospective study, a number of compounds have been assessed and used to design new molecules. The physical properties that were determined as limiting bioavailability were as follows: molecular weight greater than 500; more than 5 hydrogen bond donors; more than 10 hydrogen bond acceptors; and a $C\log P$ value less than 5.

A retrospective evaluation of the oral bioavailability of 1100 novel drug candidates with an average molecular weight of 480 from SmithKline Beecham's drug discovery efforts in rats by Veber established that reduced molecular flexibility, measured by the number of rotatable bonds and low polar surface area or total hydrogen bond count (sum of acceptors and donors), were important predictors of good (>20–40%) oral bioavailability, independent of molecular weight. A molecular weight cutoff of 500 did not significantly separate compounds with acceptable oral bioavailability from those with poor oral bioavailability, the predictive value of molecular weight was more correlated with molecular flexibility than molecular weight *per se*. From this retrospective analysis, Veber et al. suggested that compounds with 10 or fewer rotatable bonds and a polar surface area equal to or less than 140 A^2 (representing 12 or fewer H-bond acceptors and donors) have a high probability of having good oral bioavailability in rats. Reduced polar surface area was a better predictor of artificial membrane permeation than lipophilicity ($C\log P$), with increased rotatable bond count having a negative effect on permeation and having no correlation with *in vivo* clearance.

ADME is now a routine part of the efforts of a drug discovery team with the use of a number of *in vitro* approaches, for example, Caco 2 intestinal cell lines, human liver slices, or homogenates to assess potential metabolic pathways, in addition to classical rat, dog, and nonhuman primate *in vivo* studies, none of which has yet shown reliable predictability for the human situation. More recently, proteomic approaches have been used to derive potential toxicological profiles.

13.5 Compound Databases

With the exponential increase in information flow resulting from the ability to make many more compounds and test these in multiple assays, the capture, analysis, and management of data is a critical success factor in drug discovery.

Many databases used in drug discovery are ISIS/Oracle-based using MDL structural software and a variety of data entry and analysis systems, some are PC/Mac-based, and some are on a server. The ability to capture data and then reassess its value through *in silico* approaches has the potential to be a vast improvement over the "individual memory" systems that many drug companies used for the better part of the last century. Thus, with the retirement of a key scientist, the whole history

of a project or even a department disappeared, and whatever folklore existed regarding unexplained findings with compounds 10 or 20 years before was lost with the individual.

14 Lead Compound Discovery

As evidenced by the compound code numbers used by pharmaceutical companies, many thousands of new chemical entities have been made since the industry began in the late nineteenth century. Until the advent of combinatorial chemistry, the chemical libraries at most of the major pharmaceutical companies numbered from 50 000 to 800 000 compounds and comprised of newly synthesized compounds as well as those from fermentation and natural product sources. Approximately 2 to 5 million compounds were identified in the search for new drugs to treat human disease states over the past century. This number has obviously leaped to the billions with combinatorial approaches. Given decomposition and/or depletion of compounds, the 2 to 5 million compounds could be rounded down to 1 million.

The 2001 edition of the *Merck Index*, the compendium of drugs and research tools, lists a total of 10 250 compounds. Thus, from a hypothetical million compounds, only 1% has proven to be of sustained interest as either therapeutic agents or research tools.

In the early 1980s, with the promise that compounds could be created and tested on a computer screen (a promise as yet unrealized), the screening of large numbers of compounds against selected targets in biological systems was viewed as irrational. Indeed, the head of research at one of the top 20 pharmaceutical companies told the medicinal chemists at that company in the early 1980s that the demand for their skills was becoming less and would cease by the 1990s, a viewpoint somewhat akin to the apocryphal story of the head of the US Patent Office in the early 1900s who recommended its closure because "everything that could be discovered had been discovered."

The breakthrough for screening came in 1984 with the identification by Chang and colleagues at Merck of the CCK antagonist, asperlicin, which led to the clinical candidate, MK 329. Considerable effort has been expended in the pharmaceutical industry worldwide to capitalize on this approach, with significant successes, which have enhanced the search for novel chemical entities as well as providing research tools to better understand receptor and enzyme function.

14.1 High-throughput Screening

To identify compounds, it is imperative that a rapid, economical, and information-rich evaluation of biological activities be available. The term high-throughput screening describes a set of techniques designed to permit rapid and automated (robotic) analysis of a library of compounds in a battery of assays that generate specific receptor- or enzyme-based signals. These signals may be membrane-based (radioligand binding, enzyme catalysis) or cell-based (flux, fluorescence). The primary purpose of HTS is not to identify candidate drugs, but rather to identify lead structures, preferably containing novel chemical features, which may serve as a guide for more tailored iterative optimization. HTS should generate as few false-positive leads as possible because exploitation of

leads is an expensive component of the drug discovery process. HTS is designed to give information principally about potency, and a combination of follow-up screens may provide information about selectivity and specificity. The increasing use of cell-based assays provides additional information, including agonist/antagonist characteristics, and a biological "readout" under physiological or nearly physiological conditions. Additionally, cell-based assays can provide information about the cytotoxicity and bioavailability of molecules. The increased use of "designer cells" with visual and fluorescent signal readouts will continue to facilitate the screening process, and in the future, will doubtlessly include measures of metabolism, toxicity, bioavailability, and other important pharmacokinetic parameters.

Using FLIPR-like instrumentation, 100 000 or more compounds can be screened through 10 to 20 targeted assays in less than a week, and complete libraries numbering in the millions can be assessed at single concentration points in about a month. The issue is the success rate, typically in the 0.1% range, and the ability to capture and store the data for posterity. Clearly, the quality and diversity of the compounds and the robustness of the assays play a key role in a successful screening program.

14.2 Compound Sources

New chemical entities (NCEs) are discovered or developed/optimized from the following: (1) natural products and biodiversity screening; (2) exploitation of known pharmacophores; (3) rationally planned approaches, for example, computer-assisted molecular design; (4) combinatorial or focused library chemistry approaches; and (5) evolutionary chemistry.

14.2.1 Natural Product Sources

Approximately 70% of the drugs currently in human use originate from natural sources including morphine, pilocarpine, physostigmine, theophylline, cocaine, digoxin, salicylic acid, reserpine, and many antibiotics. Medicinal and herbal extracts form the basis for the health care of approximately 80% of the world's population; some 21 000 plant species are used worldwide. Screening of natural products led to the discovery of the immunosuppressants, cyclosporin, rapamycin, and FK 506, and there is a continued search for new compounds, even in relatively well-explored areas such as China.

The continued destruction of habitat with the accompanying loss of animal and plant species may impede further natural product-based drug discovery. Many interesting drugs have vanished. For example, a plant called *silphion* by the Greeks and sylphium by the Romans grew around Cyrene in North Africa. It may have been an extremely effective antifertility drug in the ancient world but was harvested to extinction.

While 100% of the world's mammals are known, fewer than 1 to 5% of other species, bacteria, viruses, fungi, and most invertebrates, are well characterized. It has been estimated that only 0.00002 to 0.003% of the world's estimated 3 to 500×10^6 species are used as a source of modern drugs. Exploration of environments previously assumed to be hostile to life has revealed bacterial species living at extreme depths, at extraordinary temperatures, and in the presence of high concentrations of heavy metals.

The sea covers almost three-quarters of the earth's surface and contains a broader

genetic variation among species relative to the terrestrial environment. Although a number of important molecules have been derived from marine sources, including arabinosyl nucleotides, didemnin B, and bryostatin 1, there has been an inadequate focus on this potentially chemically productive biosphere. Sea snails, termed *nature's combinatorial chemistry factories*, produce a bewildering array of novel conotoxins active as mammalian drug targets. These toxins are typically 10–30 amino acid residues in length, contain several disulfide bridges, and are rigid in structure. There are several hundred varieties of cone snails, and each may secrete more than 100 toxins. Therefore, there are likely to be several tens of thousands of these toxins, representing a library of substantial structural and functional diversity. These peptides are first synthesized as larger precursors from which the mature peptide is cleaved. In the mature peptide, there is a constant N terminus region and a hypervariable C terminus region from which the biological diversity is derived. Conus toxins have proven to be invaluable as molecular probes for a variety of ion channels and neuronal receptors and as templates for drug design.

Poison frogs of the *Dendrobatidae* family contain a wide variety of skin-localized poisonous alkaloids that are presumably secreted for defensive purposes. Among the chemical structures present are the batrachotoxins, pumiliotoxins, histrionicotoxins, gephyrotoxins, and decahydroquinolines (Fig. 6) that target both voltage- and ligand-gated ion channels. The alkaloid epibatidine (Fig. 6), present in trace amounts in *Epipedobates tricolor*, is of particular interest because it has powerful analgesic activities, being 200 times more potent than morphine. The alkaloid is a potent neuronal nicotinic channel agonist selective for the $\alpha 4\beta 2$ subtype. Isolated by Daly and Myers in 1974, the structure of epibatidine was not determined until 1992. The discovery of epibatidine as a novel and potent analgesic led to the identification of ABT-594, which had equivalent analgesic efficacy to epibatidine but with reduced side effect liabilities. Despite continued successes in isolating new compounds with pharmaceutical potential from natural sources, the pharmaceutical

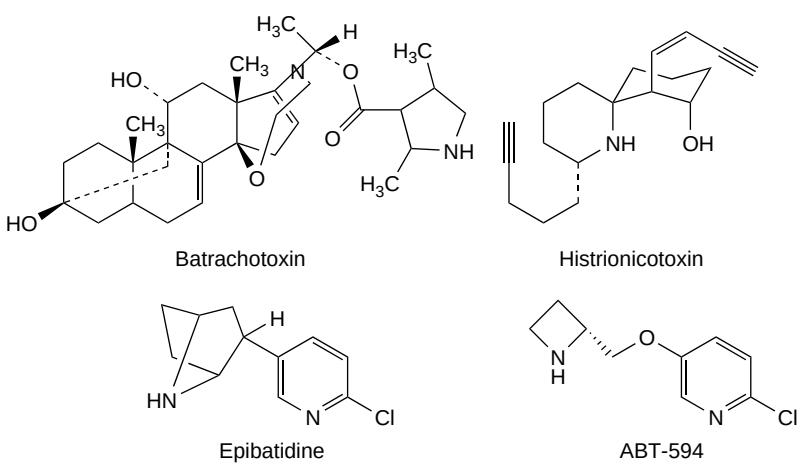

Fig. 6 Alkaloids and related structures.

industry has tended to loose interest in this approach limiting an important aspect of chemical diversity.

14.2.2 Pharmacophore-based Ligand Libraries

The majority of pharmaceutical companies have relatively large chemical libraries representing the cumulative synthetic efforts of the medicinal chemists within the company. Typically, the chemical diversity in these libraries is limited as the synthetic approach to drug design revolves around defined pharmacophores in lead series and the rational and systematic development of the SAR. Companies will thus have a large number of similar compounds based on the approaches and successes attendant to a therapeutic area.

The systematic modification of existing structures, both natural and synthetic, is an approach to compound optimization with improvements in potency, selectivity, efficacy, and pharmacokinetics being linked to discrete changes in molecular constituents on the basic pharmacophore.

Angiotensin-converting enzyme (ACE) inhibitors are important cardiovascular drugs that block the conversion of angiotensin I, formed by the action of renin on substrate angiotensinogen, to angiotensin II (a powerful pressor and growth factor agent). Until 1973, peptide inhibitors from the venom of the Brazilian viper were the only known inhibitors of this enzyme. The nonapeptide, teprotide, was an orally active and competitive inhibitor of ACE. Benzylsuccinic acid, a potent inhibitor of carboxypeptidase A, an enzyme with structural and mechanistic similarities to ACE, led Ondetti to select *N*-succinyl-l-proline as a lead ($IC_{50} = 330$ p*M*). This was subsequently optimized on the basis of the presence of a Zn^{2+} in the active site of both ACE and carboxypeptidase and the likely presence of hydrophobic pockets. An SH group to coordinate Zn^{2+} was incorporated into the stereoselectively active methyl analog, which eventually led to captopril and the analogs, enalapril, cilazapril, and lisinopril (Fig. 7).

14.2.2.1 **Molecular modeling.** Rationally planned approaches to structure design are an increasingly important part of the drug discovery process, whether planned *ab initio*, derived from structural knowledge of a putative ligand-binding site on a biological target in the absence of ligand information, or derived by rational exploitation of an existing chemical lead.

14.2.2.2 **Privileged pharmacophores.** It is increasingly apparent that a small number of common structures – "basic pharmacophores," "templates," or "scaffolds" – are associated with a multiplicity of diverse biological activities. These structures represent facile starting points for combinatorial chemical approaches to ligand diversity.

Benzodiazepines The benzodiazepines are well established as anxiolytics, hypnotics, and muscle relaxants as represented by diazepam, clonazepam, midazolam, and triazolam (Fig. 8). The BZ nucleus also occurs in natural products. Asperlicin is a naturally occurring ligand that is a weak, albeit selective, antagonist at cholecystokinin receptors and contains a benzodiazepine nucleus. From this lead was derived a series of potent and selective benzodiazepine ligands, active at CCK_1 and CCK_2 receptors, for example, L-364 718. Other BZs (Fig. 8) with activity at receptors distinct from the BZ receptor are as follows: tifluadom (opioid receptor);

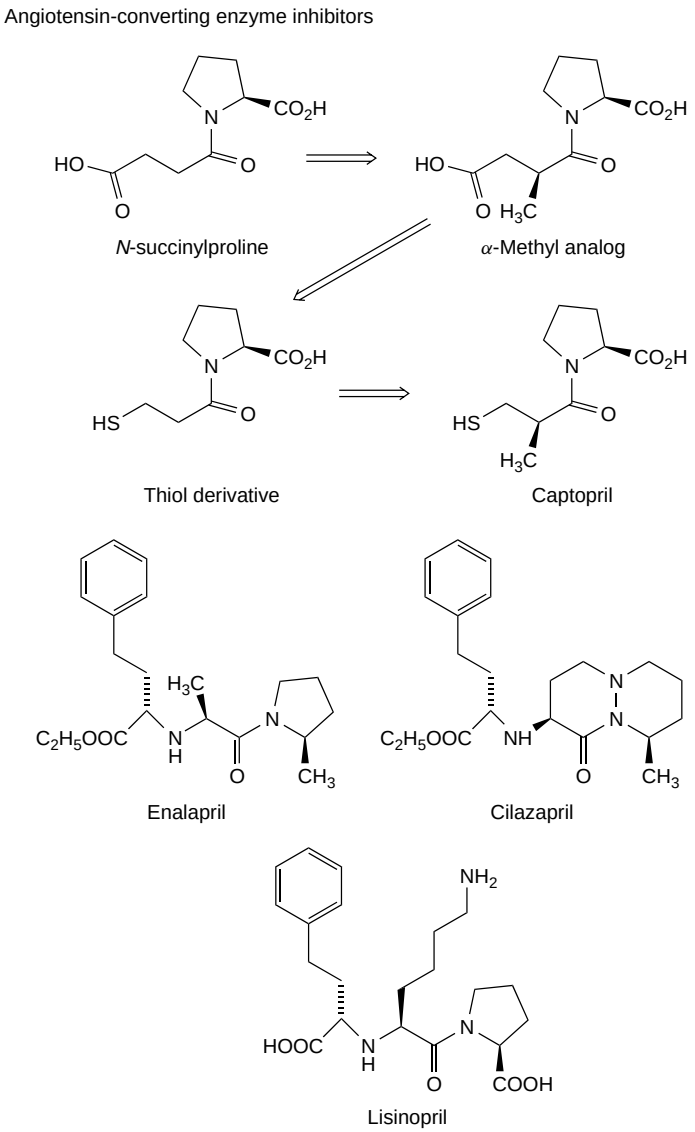

Fig. 7 ACE Inhibitors.

somatostatin; GYKI 52466 (glutamate receptor); and inward-rectifying potassium channels.

1,4-Dihydropyridines(DHP) The DHP pharmacophore is a well-established chemical entity. Nifedipine (Fig. 9) and several related DHPs, including amlodipine, felodipine, nicardipine, nimodipine, and nisoldipine, are well-established antihypertensive and vasodilating agents that act through voltage-gated L-type Ca^{2+} channels in the vasculature. However, DHPs also interact with lower activity at

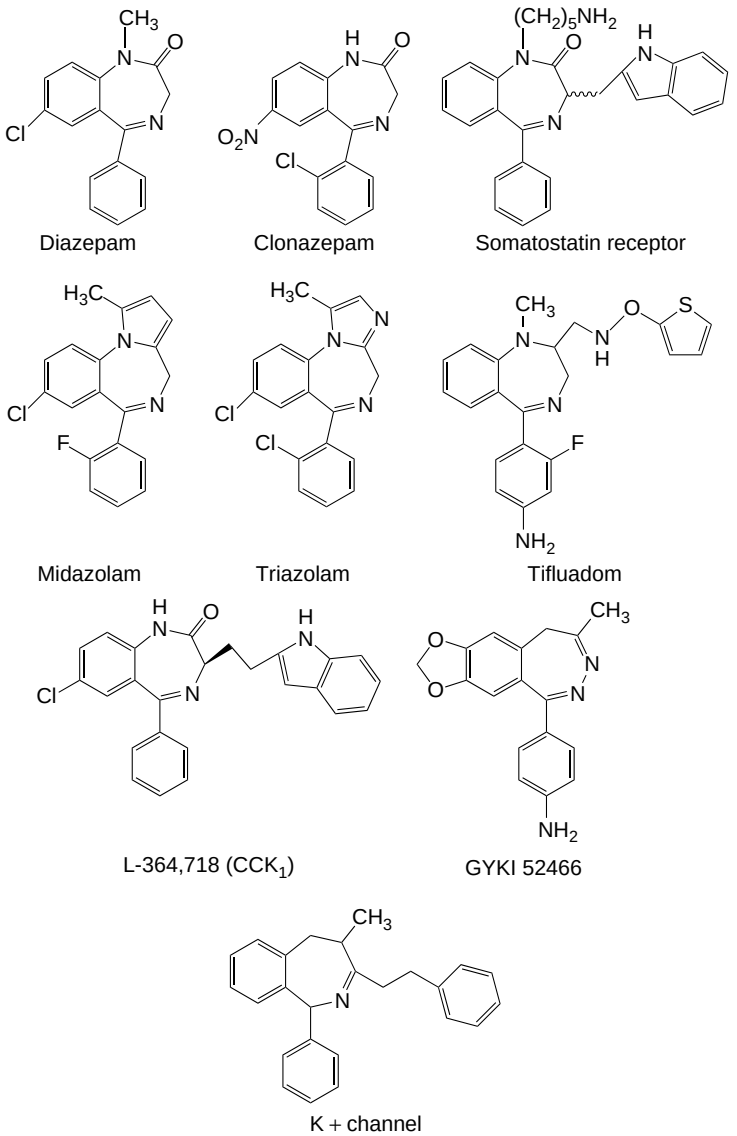

Fig. 8 Benzodiazepines: anxiolytics, hypnotics, and compounds with activity at other targets.

other classes of Ca^{2+} channels, including N- and T-type channels, and "leak" channels. DHPs can block delayed rectifier K^+ channels and cardiac Na^+ channels. Other DHP analogs (Fig. 9) are active at PAF receptors (UK 74,505), adenosine A_3 receptors, K^+ ATP channels (ZM 24 405 and A-278 367), capacitative SOC channels (MRS 1845), and α_{1A}-adrenoceptors (SNAP 5089).

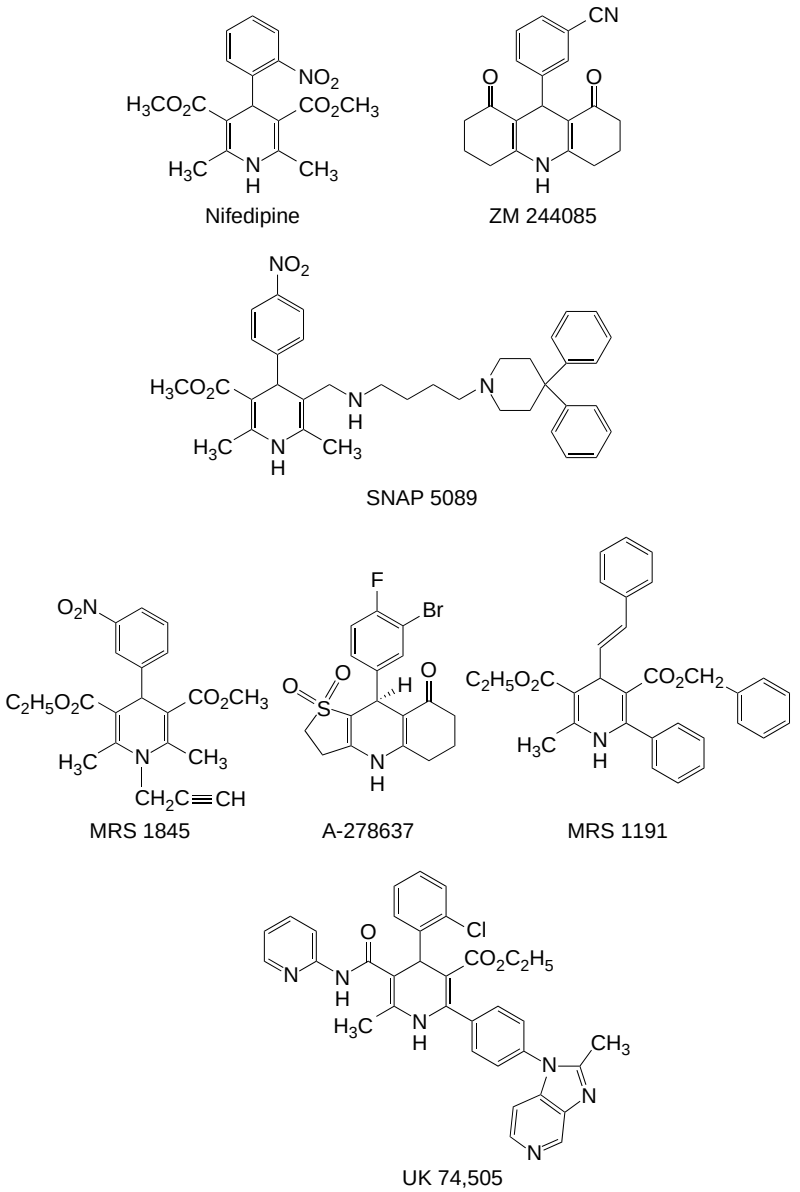

Fig. 9 Dihydropyridines; calcium channel, and other receptor antagonists.

14.2.3 Diversity-based Ligand Libraries

The issue of diversity reflects the need to enhance the scope of the library beyond those available within a company. This can be done by compound exchange with other companies, by acquiring compounds from university departments and commercial sources (Sigma, Asinex, Bader, Mayhew, Cookson, etc.). Additional sources of novel synthetic compounds include

the libraries of major present or former chemical/agricultural companies and their successors like Eastman Kodak, Stauffer, FMC, and Shell, and chemicals made in the former Eastern Block countries that are now being brokered to pharmaceutical companies. In conjunction with the use of computerized cluster programs, the selection of compounds based on diverse structures can be considerably enhanced to provide maximum coverage of molecular space. This can be done by generating libraries of approximately 2000 compounds that is used to rapidly identify potential leads for a new drug target using the SAR generated in an HTS assay. Such libraries are assembled from compounds that are available in relatively large supply and may not necessarily be proprietary to the company. Their value is in rapidly eliminating unlikely structures in a systematic manner for each new target.

Combinatorial chemistry has provided the means to synthesize literally billions of molecules, a major step forward in the exploration of "molecular space." This random or "brute-force" approach to the search for new leads has been a major disappointment in its contribution to product pipelines and the sacrifice of quality for quantity. However, combinatorial exploration around lead compounds or pharmacophores to generate dedicated libraries, when coupled with HTS, clearly provides an economical and efficient way to rapidly generate lead compounds.

Combinatorial diversity can be exploited in nature in the repertoire of antibodies with their remarkable combination of high affinity and selectivity and is also reflected in the conotoxins.

The polyketides are a family of natural products containing many important pharmaceutical agents that are synthesized through the multienzyme complex, polyketide synthase, which can display substantial molecular diversity with respect to chain length, monomer incorporation, reduction of keto groups, and stereochemistry at chiral centers. This variability, together with the existence of several discrete forms of polyketide synthase, allows the generation of diverse structures like erythromycin, avermectin, and rapamycin. This biochemical diversity has been considerably expanded by the introduction of new substrate species that were used by the enzymes to produce new or unnatural polyketides.

These methods are likely to be more extensively used in the future to provide a highly focused combinatorial approach to generating molecular diversity. The multienzyme pathway responsible for converting simple linear unsaturated allylic alcohols to sterols, carotenoids, and terpenes is incompletely characterized, but offers excellent potential for genetic manipulation to provide directed biocombinatorial chemistry.

14.3 Biologicals and Antisense

The use of the naturally occurring hormones as drugs is not new as evidenced by the use of insulin and epinephrine. The techniques of molecular biology allow the production of an increasing number of native and modified hormones and their soluble forms. Erythropoietin (EPO) is a classic example of a successful drug taken from the human body. Soluble receptors like Enbrel and humanized antibodies, for example, herceptin and D2E7 (Humira), are additional examples of how cloning techniques have altered the concept of rational drug design and what can be considered as a drug. Anticytokine therapies are particularly amenable to this approach.

Antisense oligonucleotides and RNA interference (RNAi) approaches may similarly provide novel, specific to disease treatment by the suppression/modification of a gene product. The fundamental concept is that an oligonucleotide, complementary to a disease-causing gene will anneal to that gene and prevent its transcription. In a similar manner, RNAi approaches use the fact that double-stranded RNA is rapidly degraded, and double-stranded RNA oligomers can trigger degradation of the genes encoded by them. While delivery issues remain a concern, one topically active antisense drug, Vitravene (ISIS) is approved for the treatment of CMV-induced retinitis.

15 Future Directions

The gradual evolution of the receptor concept as the basis for drug discovery over the past century has led to major advances in the understanding of biological systems and human disease states.

While the development of ever more sophisticated structure-based technologies and ultra, micro-HTS that encompass activity, ADME, toxicity, and structural data generation has resulted in an ever-increasing body of knowledge, it does not seem to have added significantly to success as measured by increased numbers of quality INDs (investigative new drug applications), but rather, together with genomics and proteomics, significantly added to the cost of the search for new medicines. The medicinal chemist and the pharmacologist play key roles in integrating and interpreting the data flow that constitutes much of the day-to-day workings of the drug discovery environment. Their challenge is to provide the intellectual framework to use this data to find drugs rather than play a technology-driven numbers game.

Acknowledgment

The present text is a modified and updated version of the chapter "Receptor-Targets in Drug Discovery and Development" published in "Burger's Medicinal Chemistry and Drug Discovery" (ISBN 0-471-37032-0) edited by D.J. Abraham in 2003.

See also Biotransformations of Drugs and Chemicals; Medicinal Chemistry; Pharmacogenomics and Drug Design; Pharmacokinetics and Pharmacodynamics of Biotech Drugs; Receptor Biochemistry; Structure-based Drug Design and NMR-based Screening; Synthetic Peptides: Chemistry, Biology, and Drug Design; Targeting and Intracellular Delivery of Drugs.

Bibliography

Books and Reviews

Ariens, E.J. (1954) Affinity and intrinsic activity in the theory of competitive inhibition. I. Problems and theory, *Arch. Int. Pharmacodyn.* **99**, 32–49.

Blundell, T.L., Jhoti, H., Abell, V. (2002) High-throughput crystallography for lead discovery in drug design, *Nat. Rev. Drug Discov.* **1**, 45–54.

Bockaert, J., Pin, J.P. (1999) Molecular tinkering of G protein-coupled receptors: an evolutionary success, *EMBO J.* **18**, 1723–1729.

Bredt, D.S., Snyder, S.H. (1994) Nitric oxide: a physiologic messenger molecule, *Annu. Rev. Biochem.* **63**, 175–195.

Christopoulos, A. (2002) Allosteric binding sites on cell-surface receptors: novel targets for drug discovery, *Nat. Rev. Drug Discov.* **1**, 198–210.

Copeland, R.E. (2005) *Evaluation of Enzyme Inhibitors in Drug Discovery*, Wiley Interscience, Hoboken, NJ.

Daly, J.W., Garraffo, H.M., Spander, T.F., Decker, M.W., Sullivan, J.P., Williams, M. (2000) Alkaloids from frog skin: the discovery of epibatidine and the potential for developing novel non-opioid analgesics, *Nat. Prod. Rep.* **17**, 131–135.

De Souza, E., Grigoriadis, D.E. (2002) in: Davis, K.L., Charney, D., Coyle, J.T., Nemeroff, C. (Eds.) *Psychopharmacology: Fifth Generation of Progress*, Lippincott Williams and Wilkins, Philadelphia, PA, pp. 91–107. Eds.),

Dower, S.K. (2000) Cytokines, virokines and the evolution of immunity, *Nat. Immunol.* **1**, 367–368.

Drews, J. (2000) Drug discovery: a historical perspective, *Science* **287**, 1960–1964.

Edwards, S.W., Tan, C.M., Limbird, L.L. (2000) Localization of G-protein-coupled receptors in health and disease, *Trends Pharmacol. Sci.* **21**, 304–308.

Ellis, C. (2004) The state of GPCR research in 2004, *Nat. Rev. Drug Discov.* **3**, 577–626.

Estibeiro, P., Godfray, J. (2001) Antisense as a neuroscience tool and therapeutic agent, *Trends Neurosci.* **24**, S56–S62.

Flower, R. (2002) Drug receptors: a long engagement, *Nature* **415**, 587.

Furchgott, R.F. (1964) Receptor mechanisms, *Annu. Rev. Pharmacol.* **4**, 21–50.

Hill, R. (2000) NK1 (substance P) receptor antagonists–why are they not analgesic in humans? *Trends Pharmacol. Sci.* **21**, 244–246.

Hill, S.J., Baker, J.G., Rees, S. (2001) Reporter-gene systems for the study of G-protein-coupled receptors, *Curr. Opin. Pharmacol.* **1**, 526–532.

Hodgson, J. (2001) ADMET? turning chemicals into drugs, *Nat. Biotechnol.* **19**, 722–726.

Horrobin, D. (2001) Realism in drug discovery-could Cassandra be right? *Nat. Biotechnol.* **19**, 1099–1100.

Jacobson, K.A., Jarvis, M.F., Williams, M. (2002) Purine and pyrimidine (P2) receptors as drug targets, *J. Med. Chem.* **45**, 4057–4093.

Kenakin, T.P. (2001) Inverse, protean, and ligand-selective agonism: matters of receptor conformation, *FASEB J.* **15**, 598–611.

Kenakin, T. (2002) Efficacy at G-protein-coupled receptors, *Nat. Rev. Drug Discov.* **1**, 103–110.

Kenakin, T. (2003) Predicting therapeutic value in the lead optimization phase of drug discovery, *Nat. Rev. Drug Discov.* **2**, 429–438.

Kenakin, T.P., Bond, R.A., Bonner, T.I. (1992) Definition of pharmacological receptors, *Pharmacol. Rev.* **44**, 351–362.

Khosla, J.R., Zawada, R.J.X. (1996) Generation of polyketide libraries via combinatorial biosynthesis, *Trends Biotechnol.* **14**, 137–142.

Kopec, K., Bozyczko-Coyne, D.B., Williams, M. (2005) Commentary: target identification and validation in drug discovery: the role of proteomics, *Biochem. Pharmacol.* **69**, 1133–1139.

Legrain, P., Wojcik, J., Gauthier, J.-M. (2001) Protein–protein interaction maps: a lead towards cellular functions, *Trends Genet.* **17**, 346–352.

Le Novere, N., Changeux, J.P. (1995) Molecular evolution of the nicotinic acetylcholine receptor: an example of multigene family in excitable cells, *J. Mol. Evol.* **40**, 155–172.

Liebmann, C. (2004) G protein-coupled receptors and their signaling pathways: classical therapeutical targets susceptible to novel therapeutic concepts, *Curr. Pharm. Des.* **10**, 1937–1958.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug. Deliv. Rev.* **46**, 3–26.

Lloyd, G.K., Williams, M. (2000) Neuronal nicotinic acetylcholine receptors as novel drug targets, *J. Pharmacol. Exp. Ther.* **292**, 461–467.

Lutz, M., Kenakin, T.P. (1999) *Quantitative Molecular Pharmacology and Informatics In Drug Discovery*, Wiley, Chichester, England.

Mangelsdorf, D.M., Evans, R.M. (1995) The RXR heterodimers and orphan receptors, *Cell* **83**, 841–850.

Masson, J., Sagné, C., Hamon, M., El Mestikawy, S. (1999) Neurotransmitter transporters in the central nervous system, *Pharmacol. Rev.* **51**, 439–464.

Milligan, G. (2001) Oligomerisation of G-protein-coupled receptors, *J. Cell. Sci.* **114**, 1265–1271.

Milligan, G. (2003) Constitutive activity and inverse agonists of G protein coupled receptors: a current perspective, *Mol. Pharm.* **64**, 1271–1276.

Monod, J., Wyman, J., Changeux, J.-P. (1965) On the nature of allosteric transitions: a plausible model, *J. Mol. Biol.* **12**, 88–118.

Neubig, R.P. (2002) Regulators of G-protein signaling as new central nervous system drug targets, *Nat. Rev. Drug Discov.* **1**, 187–197.

Neubig, R.R., Spedding, M., Kenakin, T., Christopoulos, A. (2003) International Union of Pharmacology Committee on Receptor Nomenclature and Drug Classification. XXXVIII, update on terms and symbols in quantitative pharmacology, *Pharmacol. Rev.* **55**, 597–606.

Onaran, H.O., Scheer, A., Cotecchia, S., Costa, T. (2000) *Handbook of Experimental Pharmacology*, Vol. 148, Springer, New York, 217–280.

Ondetti, M.A. (1994) From peptides to peptidases: a chronicle of drug discovery, *Annu. Rev. Pharmacol. Toxicol.* **34**, 1–16.

O'Neil, M.J., Smith, A., Heckelman, P.E., Obenchain, J.R., Gallipeau, J.A.R., D'Arecca, M.A. (2001) *The Merck Index*, 13th edition, Merck, Rahway, NJ.

Schreiber, S.L. (2000) Target-oriented and diversity-oriented organic synthesis in drug discovery, *Science* **287**, 1964–1969.

Triggle, D.J. (1992) Calcium-channel antagonists: mechanisms of action, vascular selectivities, and clinical relevance, *Cleve. Clin. J. Med.* **59**, 617–627.

Walters, W.P., Stahl, M.T., Murko, M.A. (1998) Virtual screening – an overview, *Drug Discov. Today* **3**, 160–178.

Watling, K.J. (2001) *The RBI Handbook*, 4th edition, Sigma-RBI, Natick, MA.

Williams, M., Coyle, J.T., Shaikh, S., Decker, M.W. (2001) Same brain, new century. Challenges in CNS drug discovery in the.postgenomic, proteomic era, *Annu. Rep. Med. Chem.* **36**, 1–10.

Williams, M., Giordano, T., Elder, R.A., Reiser, H.J., Neil, G.L. (1993) Biotechnology in the drug discovery process: strategic and management issues, *Med. Res. Rev.* **13**, 399–448.

Wise, A., Gearing, K., Rees, S. (2002) Target validation of G-protein coupled receptors, *Drug Discov. Today* **7**, 235–246.

Primary Literature

Chang, R.S., Lotti, V.J., Monaghan, R.L., Birnbaum, J., Stapley, E.O., Goetz, M.A., Albers-Schonberg, G., Patchett, A.A., Liesch, J.M., Hensens, O.D., Springer, J.P. (1985) A potent nonpeptide cholecystokinin antagonist selective for peripheral tissues isolated from Aspergillus alliaceus, *Science* **230**, 177–179.

Davis, R.J. (2000) Signal transduction by the JNK Group of MAP kinases, *Cell* **103**, 239–252.

Dean, M.K., Higgs, C., Smith, R.E., Bywater, R.P., Snell, C.R., Scott, P.D., Upton, G.J., Howe, T.J., Reynolds, C.A. (2001) Dimerization of G-protein-coupled receptors, *J. Med. Chem.* **44**, 4595–4614.

Enmark, E., Pelto-Huikko, M., Grandien, K., Lagercrantz, S., Lagercrantz, J., Fried, G., Nordenskjold, M., Gustafsson, J.A. (1997) Human estrogen receptor beta-gene structure, chromosomal localization, and expression pattern, *J. Clin. Endocrinol. Metab.* **82**, 4258–4265.

Honore, P., Kage, K., Mikusa, J., Watt, A.T., Johnston, J.F., Wyatt, J.R., Faltynek, C.R., Jarvis, M.F., Lynch, K. (2002) Analgesic profile of intrathecal $P2X_3$ antisense oligonucleotide treatment in chronic inflammatory and neuropathic pain states in rats, *Pain* **99**, 19–27.

Husi, H., Ward, M.A., Choudhary, J.S., Blackstock, W.P., Grant, S.G. (2000) Proteomic analysis of NMDA receptor-adhesion protein signaling complexes, *Nat. Neurosci.* **3**, 661–669.

Jacobsen, J.R., Hutchinson, C.R., Cane, D.E., Khosla, C. (1997) Precursor-directed biosynthesis of erythromycin analogs by an engineered polyketide synthase, *Science* **277**, 367–369.

Jenck, F., Wichmann, J., Dautzenberg, F.M., Moreau, J.L., Ouagazzal, A.M., Martin, J.R., Lundstrom, K., Cesura, A.M., Poli, S.M., Roever, S., Kolczewski, S., Adam, G., Kilpatrick, G. (2000) A synthetic agonist at the orphanin FQ/nociceptin receptor ORL1: anxiolytic profile in the rat, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4938–4943.

Jones, K.A., Borowsky, B., Tamm, J.A., Craig, D.A., Durkin, M.M., Dai, M., Yao, W.J., Johnson, M., Gunwaldsen, C., Huang, L.Y., Tang, C., Shen, Q., Salon, J.A., Morse, K., Laz, T., Smith, K.E., Nagarathnam, D., Noble, S.A., Branchek, T.A., Gerald, C. (1998) GABA(B) receptors function as a heteromeric assembly

of the subunits GABA(B)R1 and GABA(B)R2, *Nature* **396**, 674–679.

Kim, M., Jiang, L.H., Wilson, H.L., North, R.A., Surprenant, A. (2001) Proteomic and functional evidence for a $P2X_7$ receptor signalling complex, *EMBO J.* **20**, 6347–6358.

Koshland, D.E., Nemethy, G., Filmer, D. (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits, *Biochemistry* **5**, 365–385.

Morisset, S., Rouleau, A., Ligneau, X., Gbahou, F., Tardivel-Lacombe, J., Stark, H., Schunack, W., Ganellin, C.R., Schwartz, J.C., Arrang, J.M. (2000) High constitutive activity of native H_3 receptors regulates histamine neurons in brain, *Nature* **408**, 860–864.

Schoonbroodt, S., Piette, J. (2000) Oxidative stress interference with the nuclear factor-kappa B activation pathways, *Biochem. Pharmacol.* **60**, 1075–1083.

Shang, Y., Brown, M. (2002) Molecular determinants for the tissue specificity of SERMs, *Science* **295**, 2465–2468.

Sharpe, I.A., Gehrmann, J., Loughnan, M.L., Thomas, L., Adams, D.A., Atkins, A., Palant, E., Craik, D.J., Adams, D.J., Alewood, P.F., Lewis, R.J. (2001) Two new classes of conopeptides inhibit the alpha1-adrenoceptor and noradrenaline transporter, *Nat. Neurosci.* **4**, 902–907.

Smit, M.J., Leurs, R., Alewijnse, A.E., Blauw, J., Van Nieuw Amerongen, G.P., Van De Vrede, Y., Roovers, E., Timmerman, H. (1996) Inverse agonism of histamine H2 antagonist accounts for upregulation of spontaneously active histamine H2 receptors, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6802–6807.

Smit, A.B., Syed, N.I., Schaap, D., van Minnen, J., Klumperman, J., Kits, K.S., Lodder, H., van der Schors, R.C., van Elk, R., Sorgedrager, B., Brejc, K., Sixma, T.K., Geraerts, W.P. (2001) A glia-derived acetylcholine-binding protein that modulates synaptic transmission, *Nature* **411**, 261–268.

Stephenson, R.P. (1956) A modification of receptor theory, *Br. J. Pharmacol.* **11**, 379–392.

Thoma, J.A., Koshland, D.E. (1960) Competitive inhibition by substrate during enzyme action – evidence for the inducedfit theory, *J. Am. Chem. Soc.* **82**, 3329–3333.

Vassilatis, D.K., Hohmann, J.G., Zeng, H., Li, F., Ranchalis, J.E., Mortud, M.T., Brown, A., Rodriguez, S.S., Weller, J.R., Wright, A.C., Bergamnn, J.E., Gaitanaris, G.A. (2003) The G-protein-coupled receptor repertoires of human and mouse, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4903–4908.

Veber, D.F., Johnson, S.R., Cheng, H.Y., Smith, B.R., Ward, K.W., Kopple, K.D. (2002) Molecular properties that influence the oral bioavailability of drug candidates, *J. Med. Chem.* **45**, 2615–2623.

Wolf, B., Green, D.R. (1999) Suicidal tendencies: apoptotic cell death by caspase family proteinases, *J. Biol. Chem.* **274**, 20049–20052.

Receptor, Transporter and Ion Channel Diseases

J. Jay Gargus
University of California, Irvine, CA, USA

1 Introduction 645

2 Physiological Function of Ion Channels 646

3 Biochemical Structure of Ion Channels 650

4 Ion Channels as a Mechanism of Disease 656
4.1 Contribution of Genes and Environment to Disease 656
4.2 Monogenic Ion Channel Disease 656
4.3 Polygenic Ion Channel Disease 658
4.4 Pharmacogenetic Ion Channel Disease 658
4.5 Ion Channel Toxins 659
4.6 Mechanism of Ion Channel Disease Pathogenesis 659

5 Ion Channel Diseases of the Heart 660
5.1 Function of Ion Channels in the Heart 660
5.2 Ion Channel Dysfunction in Arrhythmia 661
5.3 Long QT Syndrome 662
5.4 Nature of LQT Genetic Loci and Alleles 663
5.5 The Dominant LQT Alleles 663
5.6 The Recessive LQT Alleles 665
5.7 Acquired LQT Channelopathies 666
5.8 Pharmacogenetic and Polygenic Arrhythmia Syndromes 666

6 Ion Channel Diseases of Muscle 667
6.1 Function of Ion Channels in Muscle 667
6.2 Ion Channel Dysfunction in Paralysis and Myotonia 670
6.3 Periodic Paralysis Syndromes 672
6.4 Malignant Hyperthermia 673

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 11
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30648-X

6.5 Other Monogenic Muscle Ion Channel Syndromes 674
6.6 Autoimmune Channelopathies in Muscle 674

7 Ion Channel Diseases of Nerve 675
7.1 Function of Ion Channels in Neurons 675
7.2 Ion Channel Dysfunction in Seizures 676
7.3 Hereditary Epilepsy Syndromes 677
7.4 LQT-like Loci and Alleles in Neuronal Syndromes 678
7.5 Other Monogenic Ion Channel Epilepsy Syndromes 678
7.6 Monogenic Ataxia and Migraine Syndromes 681
7.7 Other Monogenic Neuronal Ion Channel Syndromes 682
7.8 Acquired CNS Ion Channel Syndromes 683
7.9 Polygenic Syndromes of Altered CNS Excitability 684

8 Ion Channel Diseases of Epithelia 684
8.1 Structure and Function of Epithelial Tissues 684
8.2 Function of Ion Channels in Epithelia 685
8.3 Ion Channel Dysfunction and Diseases in Epithelia 686
8.4 Cystic Fibrosis 686
8.5 Renal Channelopathies 688
8.6 Pseudohypoaldosteronism and Liddle Syndrome 690
8.7 Bartter and Gitelman Syndrome 691
8.8 Hypomagnesemia 694
8.9 Polycystic Kidney Disease 695
8.10 Glucosuria 695
8.11 Nephrogenic Diabetes Insipidus 696
8.12 Hereditary Kidney Stone 697

9 Ion Channel Disease of Other Nonexcitable Tissue 698

10 Ion Channels as Targets for Drug Therapy in Common Diseases 699

Bibliography 700
Books and Reviews 700
Primary Literature 700

Keywords

Action potential
An all-or-none transient electrical depolarization propagated in excitable tissues.

ADH
Antidiuretic hormone or vasopressin; peptide hormone from the pituitary that controls plasma osmolarity by regulating aquaporins in the apical membrane of renal Principle cells.

Aldosterone
Adrenal steroid hormone that controls plasma volume and salt balance by regulating EnaC channels in the apical membrane of renal Principle cells.

Allele
The specific sequence of a given gene. Each possible variation of that sequence, functional or not, reflects a different allele.

Apical membrane
The specialized face of an epithelial cell that generally faces into the lumen of the gland and often is folded into a brush border with microvilli.

Aquaporin
A transmembrane protein that allows the transport of water free of solvated ions.

Ataxia
Inability to coordinate voluntary muscle activity, generally due to defect in cerebellum or spinal cord.

Channel
Transmembrane protein that allows the conductive transmembrane transport of small molecules, most commonly ions and water.

Channelopathy
A disease originating from dysfunction of a channel, or channel-like membrane protein.

Depolarization
A change in membrane potential tending to make the cell interior more positive.

Diuretic
A drug that serves to enhance urine flow, generally by acting to inhibit a renal sodium transport mechanism.

Dominant Allele
One copy of a mutation is sufficient to produce the mutant phenotype, such that the heterozygote expresses the phenotype.

Dominant Negative Mutation
A dominant mutation that results in less activity of the gene product than would be produced in a heterozygote with one wild-type and one null allele. Most commonly occurs because gene product functions as an inhibitory subunit in protein multimer.

EKG
Electrocardiogram; standardized clinical recording of extracellular electrical activity generated by the heart.

EMG
Electromyogram; standardized clinical recording of electrical activity generated by muscle contraction.

Environmental Disease
A disease caused by an agent external to the host.

Epilepsy
A chronic condition of seizure susceptibility, characterized by paroxysmal brain dysfunction and synchronized neuronal depolarizations.

Epithelium
A coupled sheet of polarized cells.

Excitation-contraction Coupling
The process by which electrical depolarization of a muscle membrane leads to activation of contractile fibers and cell contraction.

Gain-of-function Mutation
A mutation that increases the activity of the gene product to any degree.
Syn = hypermorphic.

Gate
Portions of channel protein that close access to permeability pore and is operated in response to ligand binding, voltage, or other stimulus.

Gene Family
An assembly of genes related by sequence homology, likely reflecting decent from a common ancestor followed by divergence.

G-protein Coupled Receptor
A large family of membrane proteins having seven transmembrane spans, the prototype being rhodopsin, which serve in sensory transduction by coupling to a signaling cascade through trimeric GTP binding proteins.

Haploinsufficiency
A mechanism of dominant inheritance whereby a heterozygote with one wild-type and one null mutant allele manifests the mutant phenotype. This is an atypical situation since rarely is 50% of wild-type activity insufficient for physiological function.

Heteromultimer
A combination of protein product subunits encoded by distinct genes uniting to produce a functional protein assembly.

Heterozygote
An individual having two different alleles of a gene. The default use implies that one copy is wild-type and the other mutant; however, this can also be an individual having two different mutant alleles, in which case they are specified.

Homozygote
An individual having identical copies for both alleles of a gene. These can both be wild-type or mutant.

Hypermorphic
A mutation that increases the activity of the gene product to any degree. Syn = gain of function.

Hyperpolarization
A change in membrane potential tending to make the cell interior more negative.

Hypomorphic Mutation
A mutation that inactivates the gene product to any degree. Syn = loss of function.

Inactivation
A nonconducting conformation of an ion channel from which it is incapable of carrying out an activating transition into a conducting conformation.

Ion channel
A transmembrane protein that allows the selective transport of ions freed from solvating water.

Ligand-gated Ion Channel
Binding a chemical compound to a site on the channel protein operates the gating mechanism to open or close substrate access to the permeability pore.

Loss-of-function Mutation
A mutation that inactivates the gene product to any degree. Syn = hypomorphic.

Loss of Heterozygosity
Regional loss of the chromosome carrying a normal copy of a gene, exposing clinical manifestations of a recessive mutation that is otherwise carried silently on the other copy of the chromosome.

Mendelian Disease
The same as a monogenic disease. Its inheritance follows Mendel's laws and can be either dominant or recessive.

Migraine
A condition characterized by paroxysmal onset of severe headache, generally unilateral, often associated with vertigo, nausea and photophobia and proceeded by a sensory aura.

Miniature Endplate Potentials (MEPPs)
Spontaneous depolarizations recorded at the motor endplate, which represent postsynaptic neuroreceptor response to spontaneous quantal release of neurotransmitter vesicles into the synapse.

Monogenic Disease
A disease caused by a major-effect gene mutation, such that the presence of this one mutation alone is highly predictive of the disease phenotype.

Motor Endplate
Postsynaptic membrane of specialized synapse of the motor neuron onto the skeletal muscle sarcolemma.

Myotonia
Delayed relaxation of skeletal muscle.

Neomorphic Mutation
A mutation that conveys to the gene product a new activity not present in wild-type.

Nernst Equilibrium Potential
The electrical potential difference at which no net ionic flux is observed across a permeable membrane despite the presence of a concentration difference between the sides. $E_K = [RT/nF]\ln[K_o/K_i]$ gives the Nernst potential for potassium, E_K, where R is the gas constant, T the absolute temperature Kelvin, n the valance, F the Faraday constant, and K_o and K_i the extracellular and intracellular potassium concentration, respectively.

Neuromuscular Junction
Specialized synapse of the motor neuron onto the skeletal muscle.

Null Mutation
A total loss of gene function, the extreme limit of a hypomorphic mutation, can include deletion or disruption (as in a knockout allele) of the gene itself.

Pacemaker
Endogenous source of rhythmic activity. In the heart, this is the sinoatrial node and the rhythm arises via spontaneous depolarization of the membrane.

Paracellular
Pathways across an epithelial sheet that pass between cells.

Paralysis
Inability to contract muscles.

Patch Electrode
Fire-polished glass electrode with 1-μm tip diameter that forms a molecularly-tight high resistance seal when applied to cell membrane, allowing recording from single channel molecules.

Pathophysiological Mechanism
Alteration in biological function leading to a disease process. This can be resolved from the molecular to the organismal and ecological levels.

Permeability Pore
Pathway though the channel protein taken by the conducted substrate.

Pharmacogenetic Disease
A disease caused by the interaction of an inciting ''triggering'' drug and an individual with a susceptible ''at risk'' genotype.

Plateau Phase
Segment of the cardiac myocyte action potential that coincides with the QT interval as observed on the EKG.

Polygenic Disease
A disease caused by an assembly of minor-effect gene mutations, such that the presence of any one of the mutations contributes only a small additional susceptibility to the disease.

Primary Active Transport
A transmembrane transport process that is capable of moving a substrate against its electrochemical potential gradient by directly utilizing the chemical energy of ATP hydrolysis.

QT Interval
The phase of cardiac ventricular repolarization as observed on the EKG. This coincides with the plateau phase of the cardiac myocytes action potential.

Recessive Allele
Carrying one copy of the mutation, the heterozygote appears like wild-type, not mutant. Only when both gene copies are mutant is the mutant phenotype observed.

Resorptive Epithelium
An epithelial sheet designed to create a flow of salt and water out of the lumen and into the blood.

Sarcoplasmic Reticulum
The specialized endoplasmic reticulum of skeletal muscle cells.

Secondary Active Transport
A transmembrane transport process that is capable of moving a substrate against its electrochemical potential gradient by coupling to the electrochemical potential of another chemical species, often sodium.

Secretory Epithelium
An epithelial sheet designed to create a flow of salt and water into the lumen and out of the blood.

Segregation
The disjunction of homologous chromosomes at meiosis such that offspring receive only one copy of each parental chromosome. This process underlies the characteristic Mendelian patterns of inheritance.

Seizure
Paroxysmal brain dysfunction associated with synchronized neuronal depolarizations.

Selective Advantage
An attribute of an allele that evolution favors, reflected by the survival of an excess of its descendants in subsequent generations.

Selectivity Filter
Portion of channel protein that restricts conductance to only specific substrates.

Splice Variant
Gene products that result from the use of alternative exons, often by exon skipping.

Subcellular Compartment
Membrane enclosed regions within eukaryotic cells. Examples include the endoplasmic reticulum, Golgi, mitochondria, and lysosomes.

Toxin
A natural biomolecule with potent biological activity. Ion channel toxins typically act to block or open ion channels at picomolar concentrations.

Transcellular
Pathways across an epithelial sheet that go through the cytosol.

T-tubules
Specialized invaginations of sarcolemma that play specialized role in excitation-contraction coupling of muscle fiber contraction.

Voltage Clamp
A feedback amplifier that adjusts externally applied current across a membrane such that the membrane potential is kept at a constant value even though its permeability character is varying.

Voltage-gated Ion Channel
Ion channel that is gated in response to changes in the membrane potential.

Wild-type
The allele of a gene most abundantly found in nature, for simplicity this can be considered the normal, nonmutant version of the gene.

Ion channels reflect an important but still very new mechanism of disease, first formalized in 1989 with the discovery of mutations in the *CFTR* ion channel gene causing cystic fibrosis, thereby carving out a new category of genetic inborn error, the channelopathy. Now, however, ion channel disorders are recognized to cover the gamut of medical disciplines and to cause significant pathology in virtually every organ system. The opening of this era was driven largely by the success of positional cloning strategies in isolating the genes underlying the rare monogenic disorders of ion channels and the explosive growth in basic ion channel physiology and biophysics that rendered these pathogenic mutant alleles interpretable, but in addition crystal structures have recently become available. Ion channels are a large family of over 400 related proteins representing over 1% of our genetic endowment, and many of the features shared among the rare monogenic ion channel diseases provide a solid foundation to begin to target these mechanisms for the development of novel therapeutics. They also render members of this gene family important "functional candidates" in the much more common complex polygenic diseases.

1 Introduction

Ion channel diseases reflect a relatively new category of genetic disease, or inborn error. Monogenic Mendelian ion channel diseases, or *channelopathies*, were first recognized in 1989 with the isolation of pathogenic alleles causing cystic fibrosis. The field explosively advanced as positional cloning revealed new pathogenic loci and alleles, and patch electrode

electrophysiology allowed detailed characterization of the functional effects of mutations. This work most powerfully converged in the dissection of components of the action potential of excitable tissues – nerve, muscle and heart and while it remains true that diseases of excitable tissue still most clearly illustrate this family of diseases, ion channel disorders now cover the gamut of medical disciplines, causing significant pathology in virtually every organ system, producing a surprising range of often unanticipated symptoms, and providing valuable targets for pharmacological intervention.

2 Physiological Function of Ion Channels

Ion channels are a large family of related transmembrane proteins that provide ions, predominantly K^+, Na^+, Ca^{++}, and Cl^-, a passive pathway through which they can rapidly diffuse down their electrochemical gradient across the hydrophobic barrier of the plasma membrane. The standing electrochemical gradients that drive ion movements though channels are established by ion pumps – such as the Na/K-ATPase, and ion carriers – such as the Na-K-Cl cotransporter and the Na-H antiporter. While classical kinetic studies portray channels, pumps, and carriers as radically different transport mechanisms, with channels conducting ions 4 orders of magnitude faster and seemingly quite differently than the one-at-a-time transmembrane turnover of the pumps and carriers, we now recognize much similarity in their molecular mechanisms. Most cells maintain a high intracellular K^+, and a low intracellular Na^+ and Ca^{++}, compared with the surroundings, while Cl^- is often passively distributed at its *Nernst equilibrium*, which in a cell with a typical −60 mV membrane potential (inside negative) would be 10-fold lower on the inside (i.e. $E_{Cl^-} = (RT/nF) \log [Cl^-]_i/[Cl^-]_o \sim -59$ mV). The K^+ gradient, created at the cost of ATP hydrolysis by the ion pumps, is typically 10-fold in mammalian cells, and since the K^+ permeability, P_{K^+}, of most membranes overwhelms that of other ionic species, the K^+ diffusion potential predominates, setting the cell membrane potential, and giving rise to the typical −60 mV resting membrane potential near E_{K^+}. *Note that the physiological ions both create and respond to electrical phenomena across the cell membrane.*

In many ways, channels act like highly selective water filled pores that can be opened and closed in a controlled fashion to allow a specific ion species to flow, a flow that in turn can be perceived as both a miniscule chemical flux, with the use of isotopic tracer, or, more commonly, as an electrical current that changes the membrane potential toward the Nernst potential of the conducted ion, an interior positive *depolarization* created by opening Na^+ and Ca^{++} channels, an increasingly negative *hyperpolarization* created by opening K^+ channels, and a stabilization of the membrane potential by opening Cl^- channels. Note that unlike enzymes, channels carry out no biochemical transformations; the product and substrate ions differ only in regard to the side of the membrane on which they are found. Ion channels are further distinguished from enzymes in that rarely is the transported ion itself of any physiological consequences, since the flux through channels minimally changes the ion's concentration across the cell membrane; it is predominantly the electrical consequences of the ion current flow that underlies the physiology of the ion channels. Calcium is often an important

exception to this rule, since it plays a critical role in coupling electrical activity to biochemical pathways. A second exception is that the chemical consequences of ion channel flux often predominates their function in epithelial cells.

We now recognize that ion channel proteins have structures that impose the three cardinal functions of a channel mentioned above: a *permeability pore* that provides ions but not water a passageway across the bilayer, a *selectivity filter* that allows some ions access while excluding others, and one or more *gates* that opens and closes the passageway. Three eras of Nobel Prize-winning technological and conceptual breakthroughs have led to our current level of understanding channel function (Fig. 1). All three key features of channel function were first defined, as abstract entities modeled with mathematical exactitude, in the era of the pioneers in membrane biophysics, an era that was opened up by *voltage clamp* technology fifty years ago and crowned by the Nobel Prize-winning classic squid axon studies of Hodgkin and Huxley (Fig. 1a). However, recent Nobel Prize-winning crystallographic work from the lab of Rod MacKinnon has now defined all three aspects in exquisite molecular detail for at least a few prototype channels (Fig. 1c). An essential era linking the mathematical and structural eras was that brought about by the breakthrough of *patch electrode* electrophysiology and crowned by the Nobel Prize-winning work of Neher and Sakmann. This era was critical in defining channels as biochemical mechanisms, and worked hand-in-glove with the molecular genetics of gene isolation and *in vitro* expression studies of cloned wild type and mutant channels, a breakthrough enabling paradigm pioneered in the lab of Shosaku Numa. Since the patch electrode would seal onto the plasma membrane so tightly that ions, and therefore ionic currents, could not leak between the two structures, magically tight gigaohm seals could be achieved that allowed electronic amplifiers the sensitivity to monitor in real time a single ion channel molecule in its millisecond stochastic dance between conformations: "closed," where it is nonconducting, and "open," where it passes on the order of 10^7 ions s^{-1}, producing a current of 10^{-12} amps, a picoamp, with stochastic transitions between the two states giving rise to the beautiful square wave traces of single-channel current records that are emblematic of the era (Fig. 1b).

One can separate the greater family of ion channels into various large classes based upon the cardinal channel features of selectivity and gating. One major distinction is based upon the nature of the selectivity filter and the predominant permeant ion species: hence, since the era of Hodgkin and Huxley, there have been K^+ channels, Na^+ channels and so on (Fig. 1a). Another major classification is based upon the gating mechanism. One large class of channels gate in response to their direct binding of ligand – the *ligand-gated ion channels*. Many of these ligands are classical neurotransmitters: hence, there are acetylcholine-gated ion channels, dopamine-gated ion channels and so on. But a wide range of other types of extracellular and intracellular ligands are also able to directly gate ion channels through binding. In addition, a large family of ion channels is indirectly gated by ligands, many by the same neurotransmitters mentioned above, but in this case neurotransmitter binding occurs to a heptahelical *G-protein coupled receptor* (GPCR) and the channel is activated by a second messenger ligand or a covalent modification, such as a protein

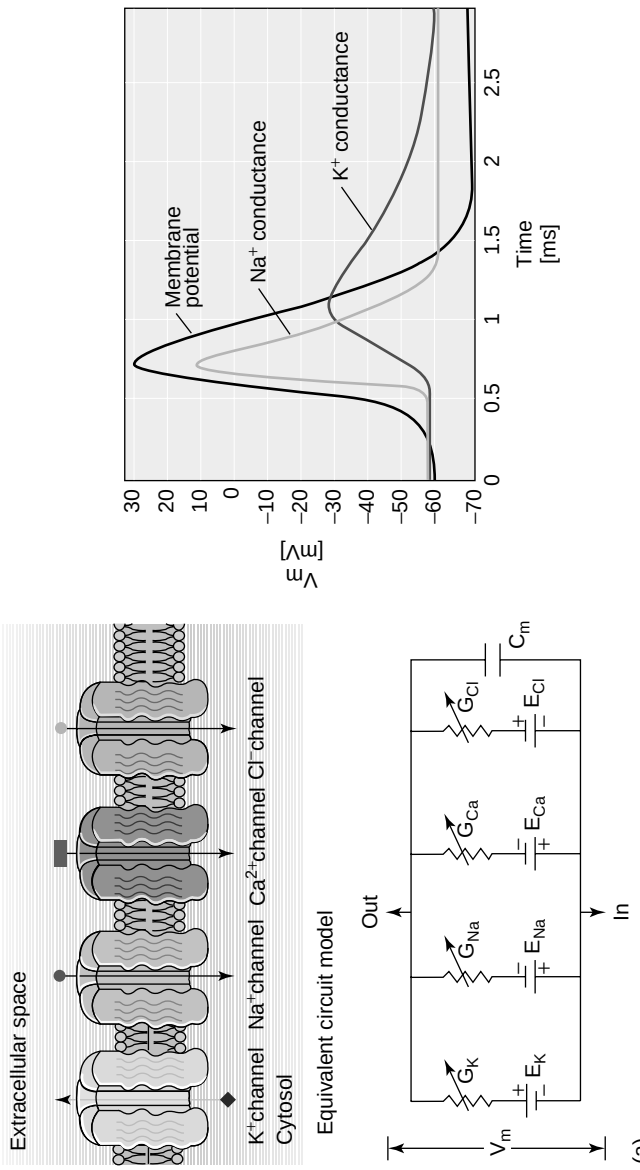

Fig. 1 3 ERAS of the ion channel. (a) Hodgkin–Huxley era, (b) Patch-clamp era, (c) Crystallography era.

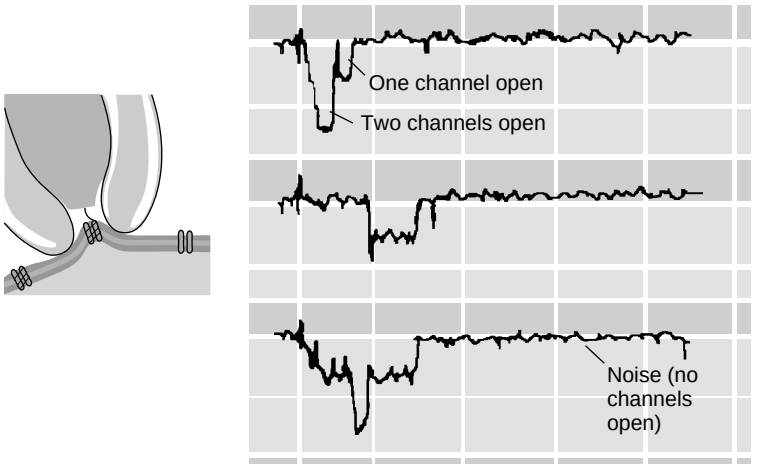

(b)

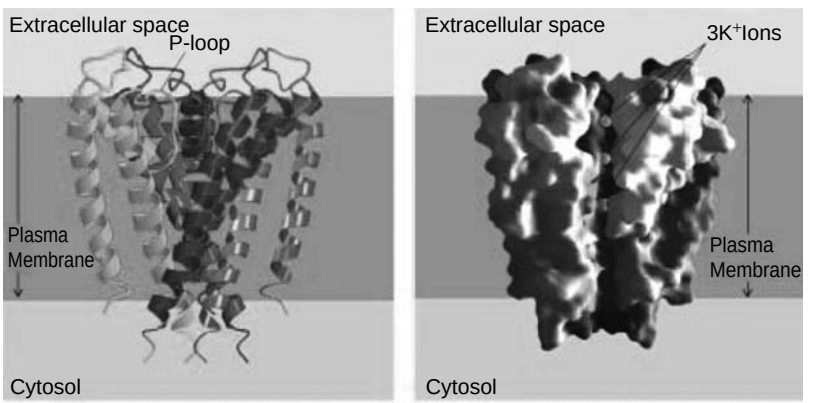

(c)

Fig. 1 (*Continued*)

phosphorylation or dephosphorylation reaction. Finally, a very large class of ion channels gate in response to changes in the electrical potential across the membrane, the *voltage-gated ion channels*. You can imagine that these are particularly tricky to study since the channel's own ionic current that is being controlled by the membrane potential *itself* alters the membrane potential, a vicious cycle that could only be broken by the voltage clamp amplifier. This experimentally vexing behavior of voltage-gated ion channels is, however, critical to their function in the perpetuation of a *propagating action potential* (AP). As a patch of membrane begins to depolarize, often because of the activation of a ligand-gated channel, voltage-gated Na^+ and Ca^{++} channels begin to respond to the voltage change by undergoing a conformational change from the "closed" into the "open" conformation (Fig. 2), increasing the membrane's permeability to sodium and calcium (P_{Na^+} and $P_{Ca^{++}}$) and hence

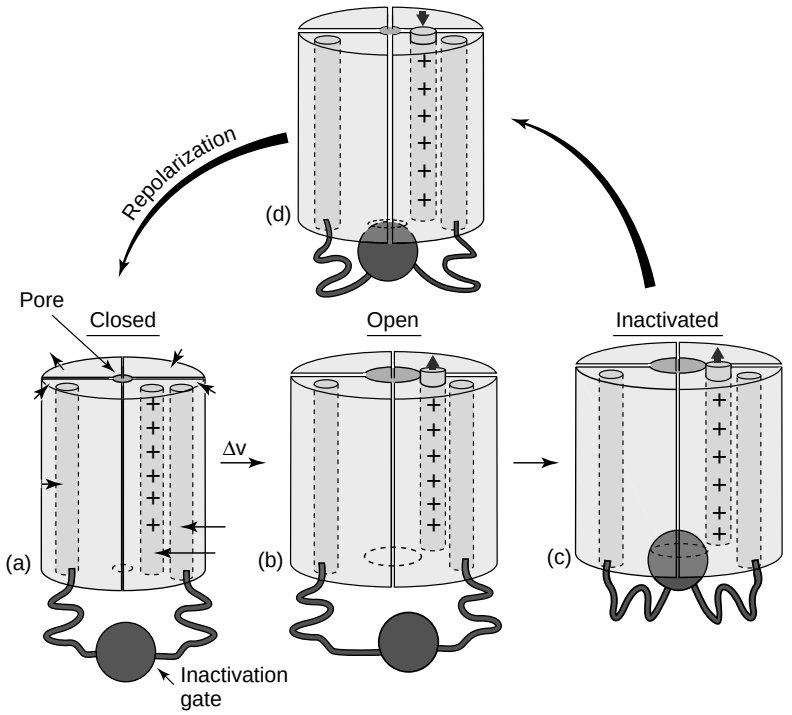

Fig. 2 Three states of the channel.

driving the membrane potential toward *their* inside-positive Nernst potential, a depolarization that serves to further activate more channels in surrounding patches of membrane in an explosive self-reinforcing cycle, recognized as a propagating AP, a depolarizing wave that rapidly spreads through an excitable tissue until all of its channels have been activated. Those channels next spontaneously enter an *inactive*, nonconducting conformation that is distinctly different from the "closed" state, since while "inactive," a channel cannot be opened and the tissue is rendered nonexcitable. Only by restoring the resting membrane potential does the channel's conformation become reset to the "closed" state that has the potential to open, and excitability is restored to the tissue. This resetting and restoring job is carried out by the more slowly activating voltage-activated K^+ channels. Since cells have but one membrane potential, it is always integrating inputs from all channel mechanisms in the membrane, and for this reason it is the great integrator in signaling pathways. Also, unlike biochemical pathways whose overall flux is predominated by a single rate-limiting step, membrane potential is intrinsically a continuous variable fully reflecting subtle changes in the entire host of contributing channels.

3 Biochemical Structure of Ion Channels

Ion channel proteins share many structural features in common that today allow

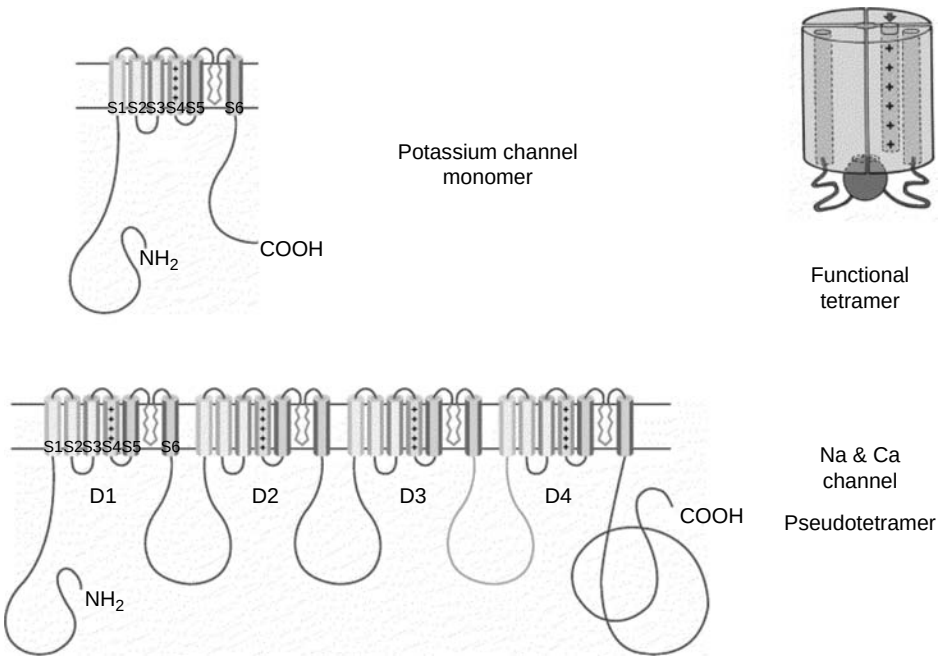

Fig. 3 Voltage-gated channels.

us to model their consensus structures and functions based upon the few prototype channel structures that have emerged from crystallography (Fig. 1c). For example, the large superfamily of voltage-gated potassium channels share a monomer structure of 6 transmembrane helices, each helix characteristically 20 to 25 amino acids in length, and both the N- and C-termini facing the cytosol (Fig. 3). The N-terminus forms a tethered ''ball and chain'' motif that gates the channel under the influence of electrostatics and the altered membrane potential its activation has brought about. It does this by swinging into the inner mouth of the pore and thereby effects the ''opened'' to ''inactive'' transition of the channel. The forth helix characteristically contains a cluster of 4 to 8 positively charged residues, converting that helix into a dipole that allows it to move under the influence of the membrane potential. This allows it to serve as the voltage sensor that gates the channel from the ''closed'' to the ''opened'' conformation. Between the fifth and sixth helix is found the heart of channel function, a signature pore domain characterized by a hairpin motif that dips partially through the membrane to form the conductive pathway used by the ion (Fig. 4). Some families of channels, such as the inward rectifiers, contain only this pore and the two flanking helices (Fig. 5). This pore domain contains the K^+ selectivity filter, a nonhelical motif conveyed by the amino acid sequence, TVGYG. It is, however, the strongly dipolar carbonyl groups of the α-carbons in the peptide backbone of this amino acid sequence, not the specialized side chains, that line the pore and create, through electrostatics and geometry, the selectivity filter that allows potassium ions

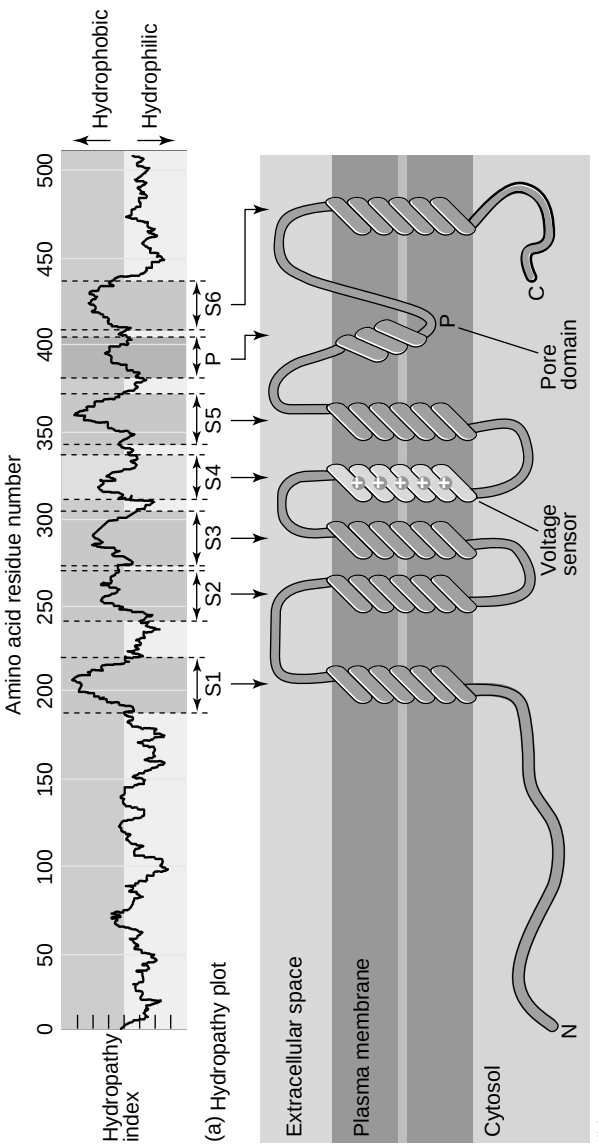

Fig. 4 Hydropathy plot and secondary structure prediction of amino acid sequence of potassium channel monomer.

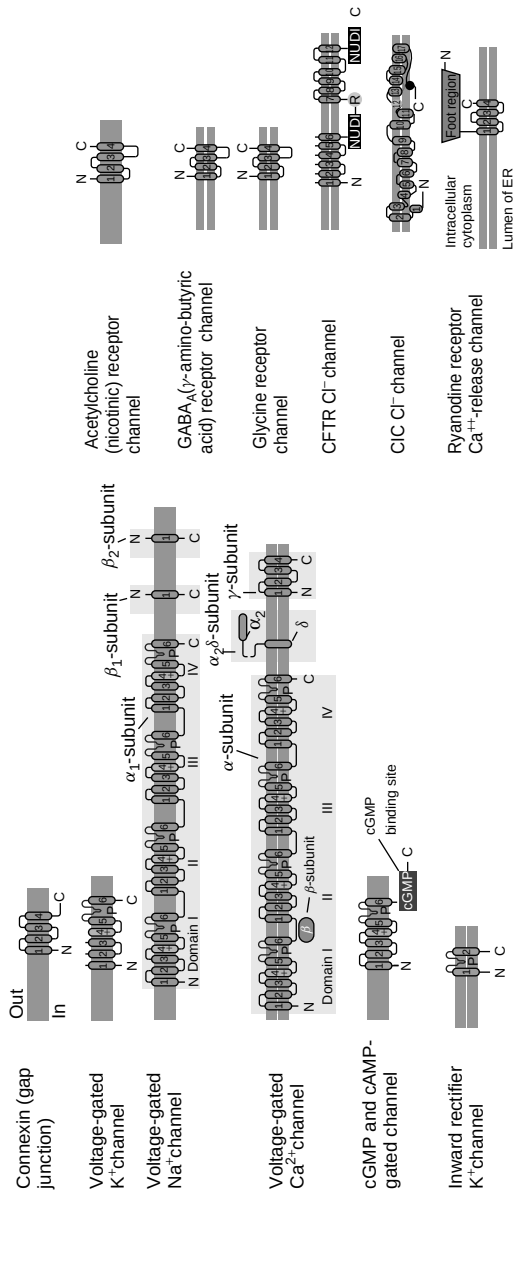

Fig. 5 Secondary structure of channel families.

entrance, but excludes sodium and other ions. The functional potassium channel that creates the transmembrane pore is a tetramer and the overall structure of the tetrameric assembly is that of an inverted teepee created by the tertiary structure of the fifth through sixth helices of the four interacting subunits, with the large vestibule of the teepee facing outward and lined by the selectivity filter residues from each monomer (Fig. 1c). Helices 1 to 4 of the monomers surround this core structure and present a surface that interacts with the lipid bilayer. Upon channel opening, the potassium ions move in a single file fashion along the pore, rapidly hopping between interactions with each of the critical residues of the filter motif, thousands of ions flowing during each channel's opening.

Conserved features that families of ion channels share in common have also proven useful in the identification of their genes in the genome. The human genome contains over 400 channel genes representing between 1 to 2% of our genetic endowment, although the substrate transported by most remains to be identified. While there are already known several large families of ion channels that apparently carry out extremely similar biochemical functions (Figs. 5 and 6), for instance, over 100 voltage-gated channels, the large number of channel genes almost certainly does *not* reflect redundancy. Since their job is not primarily contingent on their ability to transport ions from one side of the membrane to the other, but rather on the shape of the electrical signal they produce during conductive ion flow, it is apparently this that is under evolutionary selective pressure and it suggests that physiology requires the diverse vocabulary provided by multiple members of a gene family, which carry out nearly indistinguishable transport reactions with subtly differing kinetics. Furthermore, it is becoming increasingly clear that gene-family diversity is additionally greatly expanded through alternative exon splicing, allowing one channel gene locus to produce multiple splice isoforms and hence multiple functionally distinct protein products. Finally, for at least a handful of channels, and best characterized in the glutamate receptor channels, the exotic complex process of RNA editing is carried out to change single amino acid residues in the permeability pathway that critically tune ion selectivity and other aspects of channel function. Clearly, very subtle differences in channel function are important since evolution seems to have gone to very great lengths to preserve and enhance channel gene diversity.

For the potassium channels, diversity is further tremendously amplified through combinatorial monomer assembly into the functional tetramer (Fig. 3). This predominantly arises because channel protein subunits encoded by all members of a channel gene *family*, not just the allelic splice-variant protein products, coassemble to create a functional channel that differs in biophysical behavior as a function of the stoichiometry of the constituent subunits in the assembly. The panoply of combinatorial heteromultimers formed as monomers from family members coassemble, creates tremendous functional diversity in the potassium currents. This finely tunes the behavior of the population of expressed functional channels. This diversity conveys the essence of potassium channel function, allowing them to be tuned to modulate signal processing as they interact and jointly shape the membrane potential, thereby "deciding" the cell's aggregate signaling output.

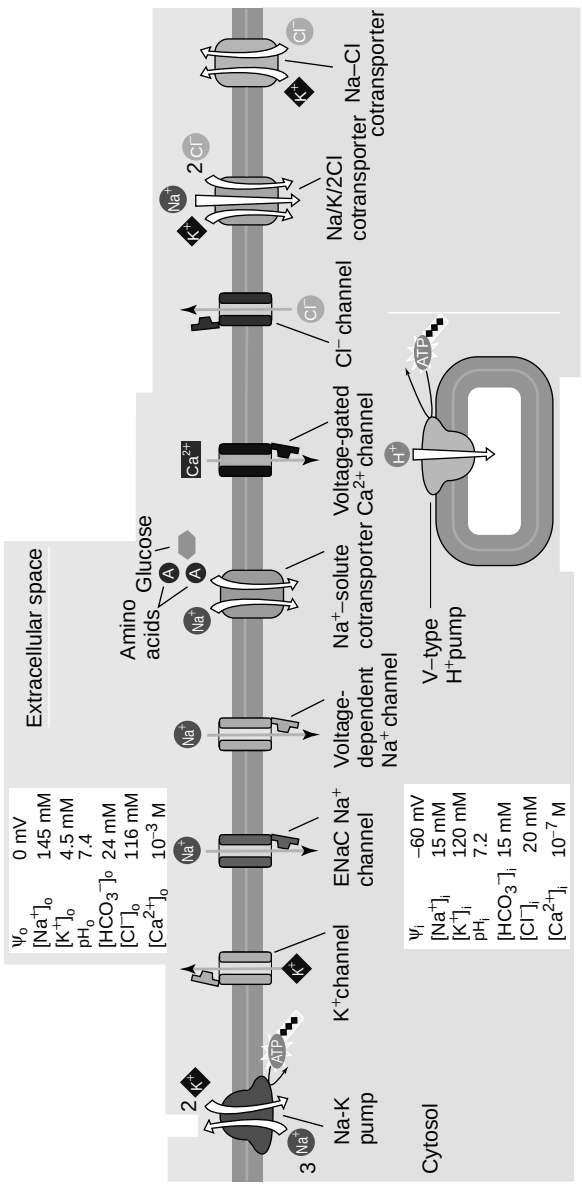

Fig. 6 Ion pumps, carriers, and channels.

In contrast to the broad heterogeneity of composition and function of the "contemplative" potassium channels, the voltage-gated sodium and calcium channels that are the "do-ers," able to execute the result of signal processing, are extremely homogeneous in their structure and function in a given tissue. While they can be perceived to share the structure of the potassium channel monomer, in their case, where a crisp uniformity of channel function is required within a tissue, a fixed pseudo-tetramer structure is hardwired into the large endoduplicated genes that encode a pseudotetrameric functional monomer (Figs. 3 and 5). Tissue-specific diversity in the function of these channels is conveyed by tissue-specific isoforms of the gene family of voltage-gated sodium and calcium channels.

In addition to the major, α, channel subunits that directly contribute to the conducting pore of ion channels, discussed above, there are also a host of auxiliary subunits, each represented by gene families often labeled $\beta, \gamma, \delta, \varepsilon$, that modify channel function (Fig. 5). This obviously adds another enormous combinatorial explosion of channel diversity. Much of this diversity allows channel function to be tuned for its specific use in a specific tissue, since many channel genes are expressed in a remarkably tissue-selective fashion. On the other hand other channels may be found to be broadly expressed, but to predominate in the physiology of only a few tissues. As will become clear below, both of these features allow channels to produce tissue-selective disease. This tissue selectivity also points out that ion channels should offer highly focused targets for therapeutic intervention, perhaps devoid of the cross-tissue "side effects" that complicate many current receptor-based drug targets.

4 Ion Channels as a Mechanism of Disease

4.1 Contribution of Genes and Environment to Disease

While classically diseases are recognized to arise from a host of extrinsic environmental causes, such as infectious agents, or from intrinsic causes, most simply a reflection of our genes, we now recognize that genes and the environment are inextricably connected in all disease, and in most we see various contributions from a small number of "major-effect" genes, smaller contributions from a large number of "minor-effect" genes, and variable contributions from a host of factors in the environment. While recognizing this interconnectedness of genes and environment, we still tend to refer to diseases as being "caused" by the most salient of these three factors, the *monogenic* diseases, the *polygenic* diseases, and the environmental diseases. *Most parsimoniously, disease reflects a mismatch between genotype and environment.*

4.2 Monogenic Ion Channel Disease

Rare mutations arise *de novo* in the population each generation, breaking the molecular mechanism encoded by every gene in the genome in every possible way, most mutations being either neutral or deleterious to the gene's function; should the effect disrupt a critical physiological pathway, they are "disease-causing." *Dominant* mutations are expressed in a phenotype even in the presence of one normal allele, so the phenotype occurs in the individual in whom they first arise; those producing a phenotype incompatible with reproduction being immediately

eliminated from the population, but being constantly replenished though new mutations. Thus, dominant mutations that are maintained and transmitted in the population tend to produce a later onset or milder course of disease. Recessive mutations that arise *de novo* are carried silently in a *heterozygote*, only rarely producing a disease on those rare occasions when two copies of the mutant gene are contributed to a *homozygous* offspring of two rare heterozygotes, this likelihood tremendously enhanced in a consanguineous mating. As most recessive alleles in the population are carried in heterozygotes, even lethal homozygote phenotypes have little effect on the disease-allele's frequency. On the other hand some "disease-causing" alleles are now very abundant in a given population, exceeding 1% of the population's alleles, frequencies typical of *polymorphic variants* that are commonly considered to be of "neutral" functional significance. An example of such a common polymorphism, carried by approximately 1 in 20 Caucasians, is the mutation that causes the most common genetic disease in Caucasians and the first recognized genetic ion channel disease, cystic fibrosis (CF). To understand the apparent anomaly of a lethal mutation rising to such abundance, a historical perspective is required. Often such high "deleterious" allele frequencies arise because of some *selective advantage* conveyed to the heterozygote carrier. In other instances, it simply appears that by chance heterozygotes for a rare mutation were among the founders who established an isolated population; with inbreeding, this allele became abundant and the disease frequency in this isolate became appreciable. Many common disease-causing mutations seem to have served in our evolutionary past as primitive "inoculations" protecting heterozygous carriers, who possess most of the "disease-alleles" in the population, against specific disease agents. The best-known example is the assent of the globin sickle cell and thalassemia hemolytic anemia alleles in the malaria belt, but CF alleles appear to be particularly protective against the infectious secretory diarrheas that imposed a high infant mortality in our recent past. While carrying two copies of the common CF allele is sufficient to produce a disease phenotype, independent of the contribution of other genes or environmental agents, and hence is recognized as a monogenic disease, the rare, non-polymorphic mutant alleles of genes more commonly cause the rare monogenic diseases, such as the Long QT syndrome ion channel disorder, or channelopathy, to be discussed below.

Diseases such as those discussed above, in which there is a major single gene contribution to the disease phenotype are classically recognized as monogenic or *Mendelian* genetic diseases. The classical Mendelian pattern of recessive disease arises most commonly from *loss of function*, *null*, or *hypomorphic*, mutations, predominantly in a situation where sufficient gene product is available from the sole remaining copy of an essential gene for homeostasis. The dominant pattern of inheritance commonly requires a *gain of function*, such that a phenotype is created even in the presence of a normal copy of the gene; this can be an enhanced normal *hypermorphic* activity, a *neomorphic* activity not present in the native protein, or an inhibitory activity, *negative-dominance*, generally arising from an inactivating multimerization with the native protein. An alternative mechanism of dominance is *haploinsufficiency*, the less common situation where 50% of the physiological level of the gene product is *not* enough for homeostasis and in such cases

a loss-of-function allele can be dominant. Even for these "monogenic" diseases we recognize that other genes in the genome and the environment still contribute to the disease phenotype, and these effects are acknowledged under the classical terms of *penetrance* and *expressivity*; two individuals carrying exactly the same two alleles of a major-effect gene may have vastly different phenotypes with one showing the classical disease and the other ranging from having such minor manifestations that no aspect of the phenotype reveals the presence of the "disease genotype," in which case he is called *nonpenetrant*, to perhaps manifesting only a minor, yet still diagnostic, feature of the disease, in which case he would demonstrate the variable expressivity of the disease.

4.3 Polygenic Ion Channel Disease

The situation described above is magnified in the polygenic diseases, common diseases of ill-defined pathogenesis, such as hypertension, diabetes, and epilepsy, where typically one mutant gene is insufficient to create the disease phenotype, most of the genes still remain to be defined, and the pattern of inheritance observed is simply familial clustering, not the clear segregation patterns of recessive or dominant inheritance seen in the Mendelian disorders. While many gene loci contribute together to produce the common polygenic diseases, both common polymorphisms and rare mutant alleles seem to participate, and while major-effect genes play a very minor role in the population, not uncommonly rare Mendelian forms of the disease phenotype do occur, their etiological gene providing a peak into the pathogenic mechanisms likely underlying the common form of the disease, such as the monogenic channelopathies recognized to cause epilepsy in benign familial neonatal convulsions (BFNC), hypertension in Liddle syndrome, and diabetes in permanent neonatal diabetes (all to be discussed in detail below). Most polygenic diseases additionally have a significant environmental component, hence their more comprehensive title of *complex polygenic diseases*. The environmental component of these diseases can be recognized since even identical monozygotic twins, sharing all of their genes in common, are discordant for the disease phenotype a high percentage of the time.

4.4 Pharmacogenetic Ion Channel Disease

Finally, some diseases appear to be caused by environmental agents. However, we now recognize that only in the face of an appropriate "at risk" genotype is an appropriate physiological target available such that the agent can be a toxin. Should the toxin be fairly rare in the environment, such as a general anesthetic, it may appear initially to cause an idiosyncratic environmental disease, such as the lethal rare adverse reaction to anesthetics referred to as *malignant hyperthermia* (MHS). Much later it was recognized that specific heritable ion channel mutations cause one to be vulnerable to this syndrome, an individual inheriting an MHS allele being perfectly healthy, just at risk for malignant hyperthermia, and not developing the lethal disease phenotype until exposed to the inducing anesthetic. The recognition of major genetic and environmental components in the disease classifies it as a *pharmacogenetic* disease, and likely many of the idiosyncratic drug reactions we observe, either toxic or therapeutic, reflect this same phenomenon. On the other

hand, should the environmental agent be ubiquitous, such as phenylalanine, an amino acid found in all foods, a disease caused by an environmental agent may appear to be a genetic disease, in this case phenylketonuria (PKU), the archetypical "genetic" disease for which every newborn is screened, but which produces no disease phenotype if phenylalanine exposure is controlled.

4.5 Ion Channel Toxins

That ion channels are exquisitely sensitive, finely tuned, evolutionarily highly conserved mechanisms essential to an organism's physiology has not been lost to the combatants on Evolution's fierce battleground. Nearly all of the most potent toxins in Nature target ion channels, most with nanomolar to picomolar affinity. Tetrodotoxin from the puffer fish, saxitoxin from red tide dinoflagellates, and a huge number of scorpion and coelenterate peptide toxins target the voltage-gated sodium channels; ω-conotoxin from the cone snails and a host of spider venom toxins target the voltage-gated calcium channels; a number of scorpion, insect, and coelenterate toxins target families of potassium channels; α-bungarotoxin, from snake venom, and the Amazon blow-dart alkaloid toxin curare target the nicotinic acetylcholine receptor and the plant alkaloid picrotoxin targets the GABA receptor channel. Most toxins act by relying upon intimate, high-affinity contacts within the channel molecule, and for this reason many have proven to be essential tools in defining classes of ion channels and their mechanisms of action. Since most of these toxins are quite rare in the environment and none seem to have participated in major human evolutionary bottlenecks, nearly all human genotypes are vulnerable to these toxins, despite the fact that in the laboratory mutations can be selected to convey toxin resistance. For this reason, a huge array of toxins can be recognized to produce quite pure forms of environmental or toxic channelopathies.

4.6 Mechanism of Ion Channel Disease Pathogenesis

Ion channel diseases span the gamut of pathogenic mechanisms discussed above. Furthermore, the pathophysiological mechanisms that lead from a defective ion channel protein to a disease phenotype are complex and multifaceted. That Nature requires a genetic investment in the huge array of subtly differing channels leads one to expect that subtle changes in ion channels matter, foreshadowing the wide range of phenotypic abnormalities different mutants of the same gene might produce. It is therefore not surprising that ion channel diseases as apparently distinct as migraine and ataxia, paralysis, and myotonia or even as having dominant versus recessive inheritance, are found to be allelic, caused by different mutations in the same gene.

A *periodic* disturbance of rhythmic function is the cardinal feature of ion channel disease of excitable tissues. In the heart, this produces a fatal *arrhythmia*, observed as a loss of the normal synchronous contraction of the muscle composing the pumping chambers secondary to a disruption in the rhythmic action potentials arising in this electrically coupled tissue. In skeletal muscle, abnormal ion channel function produces periodic alterations in contractility, ranging from the inability to contract – *paralysis*, to the inability to relax – *myotonia*. And in the central nervous system (CNS), the rhythmic disturbance

is observed as an abnormally synchronous electrical discharge, a seizure. The remarkable finding in all of these syndromes is the absence of an overt functional abnormality the vast majority of the time. In the extreme, this results in the tragic situation where no phenotype is obvious until the moment of premature death. While many insights have been garnered as to how a variety of stresses serve to create the decompensations that allow these periodic phenotypes to become manifest, the mechanisms that serve to compensate for a constitutionally defective channel such that healthful homeostatic function is maintained the majority of the time, remain to be defined.

Ion channel phenotypes in nonexcitable tissues are more diverse because of the very different roles channels play in such tissues, participating in signaling involved in endocrine secretion, the function of cytosolic compartments and complex epithelial secretory and resorptive function, which in the kidney, produce secondary changes in systemic electrolyte balance and blood pressure.

It is simplest to develop the pathophysiology of ion channel disease by starting with the cardiac phenotype of arrhythmia and the Long QT (LQT) syndromes, demonstrating how the ion channels in the membrane work together in tissue function and how component parts of the membrane potential are dissected by the genetic lesions. It further provides specific paradigms for how subunit interactions can be manifest, how loss-of-function and gain-of-function phenotypes occur and how dominance can arise through either haploinsufficiency, gain-of-function, or in a "dominant negative" fashion. The muscle and CNS phenotypes, while not lending themselves to as simple a dissection of function are, however, easily built by extrapolation from this paradigm. They further illustrate the power of gene families in identifying candidate disease genes and reinforce the intrinsic polygenic nature of ion channel phenotypes. Finally, the nonexcitable tissue phenotypes can be used to show the range of functions ion channels carry out.

5 Ion Channel Diseases of the Heart

5.1 Function of Ion Channels in the Heart

Cardiac cells possess an intrinsic instability of their membrane potential, such that spontaneous depolarizing action potentials arise (as described at the end of Sect. 2 above), that can be observed as contractions in dissociated cardiac cells in culture. Muscle contraction is directly coupled to the propagation of an action potential, a phenomenon referred to as *excitation-contraction* (EC) *coupling*, since the calcium ions that enter the cell, amplified by other calcium sources inside the cell, produce the biochemical activation of the contractile proteins. Briefly, this entails calcium binding to troponin and releasing tropomyosin's block on myosin's cycling cross-bridge formation with actin. The electrical *pacemaker* function in a healthy heart only has the opportunity to manifest itself at the sinoatrial (SA) node, the physiological pacemaker of the heart. Since all of the other cardiac cells have a slower endogenous rate of spontaneous depolarization (pacing), and since all cells in the heart are electrically coupled, the fastest spontaneously depolarizing pacemaker, the SA node, is the origin of a propagated action potential that drives though the remaining cells of the organ,

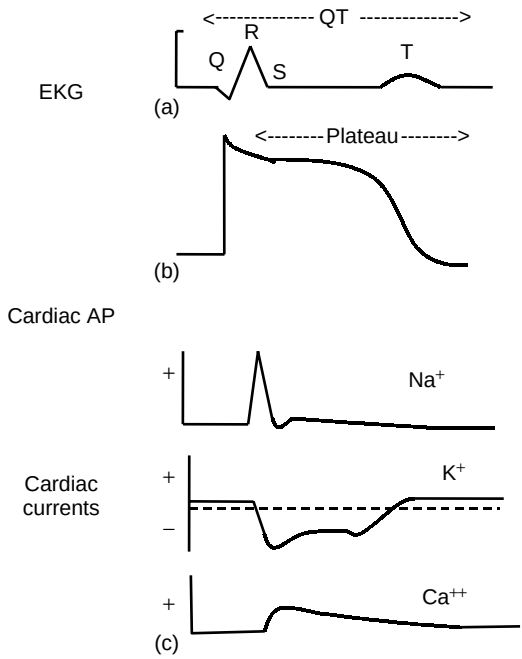

Fig. 7 Synchronized EKG, cardiac action potential and component currents. (a) EKG trace, (b) cardiac ventricle myocyte action potential, (c) resolved ionic currents underlying ventricular myocyte action potential. From Gargus, J.J. (2003) Unraveling monogenic channelopathies and their implications for complex polygenic disease.

forcing them to depolarize before they would do so spontaneously. Should the fastest pacemaker cells be lost, the next fastest pacemakers progressively take over, a very reassuring redundancy in a complicated process that must be repeated several billion times, failure rapidly resulting in death. The directional flow of the propagated action potential through the heart muscle creates the vector of the surface current revealed by the EKG traces, and produces a synchronous rhythmic coordinated contraction of the muscle cells composing the pumping chambers of the heart. If the cardiac cells of a pumping chamber all contract synchronously, the chamber volume rapidly shrinks, forcefully pumping blood that is directed by the cardiac valves. Without synchrony there is no pumping, only ineffectual fibrillation of individual cardiocytes and death to the individual. The cardiac pump cycle thus can be followed with the EKG. The QT interval measured on the EKG is the long interval of the cardiac cycle during which the ventricles, the major pumping chambers, repolarize, their conducted depolarizing action potential being reflected by the QRS waves (Fig. 7).

5.2 Ion Channel Dysfunction in Arrhythmia

A design feature engineered into the ventricular action potential by the mix of ion channels in its membrane is a prolonged plateau phase of depolarization, a phase *not* present in the classic rapid action potential of nerve (Fig. 1a). LQT syndrome is caused by mutations in the cardiac ion channel genes contributing to this component of the cardiac action potential. The plateau phase is a time during which calcium enters ventricular myocytes to produce contraction, but as importantly, the depolarized plateau phase serves to hold

sodium channels in the "inactive" nonexcitable state until the inciting depolarizing wave has spread to all of the electrically coupled myocardium. In this manner, it can be assured that all ventricular myocytes will repolarize and regain the ability to again depolarize in synchrony. If this mechanism fails, and one group of cells fires an action potential that cannot be conducted through a still-recovering inactive area, those inactive cells may subsequently have sufficient time to develop their endogenous pacemaker-like depolarization and asynchrony in the tissue is initiated as an endlessly looping futile depolarization wave spreads through any newly excitable domains of the tissue. Such asynchrony is lethal since unless the entire ventricular muscle depolarizes and contracts together, it does not pump blood, but only fibrillates. During the plateau phase, the membrane potential struggles between the depolarizing influence of sodium and calcium currents and the repolarizing influence of potassium currents, with the slowly developing potassium currents finally overcoming the depolarizing currents and restoring the resting membrane potential (Fig. 7). Mutations that serve to either enhance the sodium current or to reduce the potassium currents predictably prolong the plateau phase and are associated with LQT syndrome. A long plateau phase is a dangerous situation since prolonging it opens the opportunity for asynchrony between the cells and susceptibility to arrhythmia, as described above.

5.3 Long QT Syndrome

Classically the LQT syndromes are divided into those with a dominant pattern of inheritance and a phenotype limited to the heart, the Romano–Ward syndrome (RW), and those having recessive inheritance and a phenotype that includes sensorineural hearing loss, the Jervell and Lange-Nielsen syndrome (JLN). In addition, there are now recognized acquired forms of the disease, primarily developing in a failing heart, and pharmacogenetic forms, a serious complicating side effect for a wide spectrum of drugs. As the name suggests, LQT prolongs the QT interval measured on the EKG and predisposes to a fatal arrhythmia; however, typically the pedigree may only reveal early sudden death of no known etiology. There are no diagnostic pathological findings and commonly no symptoms before death such that the disease is commonly not diagnosed until repeated deaths in a family have occurred. Diagnosis of the syndrome, on the other hand, does not depend upon a clinically recognized phenotype; rather, it depends only upon the EKG findings. There is a characteristic context of death observed in the syndrome that further illuminates the critical balancing act performed by multiple cardiac channels in normal physiology. The context is that of excess adrenergic outflow, as occurs in high emotion or exertion. This is a risk since the heart rate dramatically increases during the "fight or flight" responses these transmitters mediate. While we may initially focus on this response in terms of the β-adrenergic receptor increasing the rate at which the cardiac pacemaker depolarizes and the heart contracts, it is clear that to depolarize more rapidly, it must also *repolarize* more rapidly, and in fact specific cAMP-dependent channel regulatory mechanisms downstream of the β-adrenergic receptor assure this coordination, effectively shortening the QT as the heart rate rises. This explains why the rational way to evaluate a QT is in the context of the underlying heart rate, the QTc. In a patient with an intrinsic LQT defect

that lengthens the QT, this adjustment of the rate of repolarization can not keep pace, the QT can not sufficiently shorten at high heart rates, and an arrhythmia ensues.

5.4 Nature of LQT Genetic Loci and Alleles

LQT mutations have been found in six cardiac ion channel genes. One LQT gene encodes the cardiac sodium channel and five encode the primary and auxiliary subunits of three different potassium channels (Fig. 8). While this provides a nearly complete dissection of the major channels involved during the critical plateau phase, one major actor, the cardiac calcium channel, encoded by *CACNA1C*, has yet to yield a pathogenic allele, perhaps being an early embryonic lethal. However, the most recently discovered LQT locus, LQT4, encoding ankyrin-B, while being unique in *not* encoding an ion channel protein, regulates a complex of transporters in the cardiac membrane mediating calcium homeostasis. *LQT1*, the first LQT gene identified, was found using a strategy involving linkage and positional cloning. All of the other LQT genes were identified as functional candidates within a mapped chromosomal interval.

5.5 The Dominant LQT Alleles

LQT1 was located on chromosome 11p 15.5, and it was ultimately demonstrated that mutations in a potassium channel α-subunit gene, *KCNQ1*, were causal. When expressed *in vitro*, *KCNQ1* produces a current that is not found in the heart; however, when expressed together with its β-subunit, encoded by *KCNE1*, they can be recognized to produce the I_{Ks} channel underlying the slow delayed rectifier K^+ current that participates in repolarization (Fig. 8). Over 35 different pathogenic alleles have been reported in this gene and they are the most common cause of RW syndrome. They primarily confer a dominant-negative phenotype when studied *in vitro*. Defective subunits coassemble with wild type copies producing, through combinatorials, a supermajority of defective channel tetramers, and hence reduced repolarizing current.

LQT2 was the first LQT gene cloned, taking advantage of a candidate gene approach within the region of chromosome 7q35–36 where a second locus was identified in families not mapping to chromosome 11. The gene, *KCNH2*, was a strong functional candidate based upon its homology to a fly gene with a proven ability to create a rhythm disorder phenotype, ether-a-go-go. Like *KCNQ1*, coexpression of *KCNH2* with its β-subunit, *KCNE2*, was required to produce a current that can be recognized in the heart. The channel they form is I_{Kr}, the rapidly activating delayed rectifier K^+ current (Fig. 8). Over 16 alleles of *KCNH2* have been reported, and like *KCNQ1*, many confer a "dominant-negative" phenotype when expressed *in vitro*. However, other dominant alleles appear to be simple loss-of-function alleles, suggesting that the membrane current conducted by this channel is so finely tuned that a haploinsufficiency mechanism is adequate to produce dominance.

LQT3 is the only sodium channel locus involved in the disease, and it is caused by mutations in the *SCN5A* gene found at chromosome 3p21–24. It encodes the cardiac-specific voltage-gated sodium channel that underlies the rapid depolarization phase of the ventricular action potential that produces the QRS complex and ventricular contraction (Fig. 8). Over 30 alleles of this gene have been

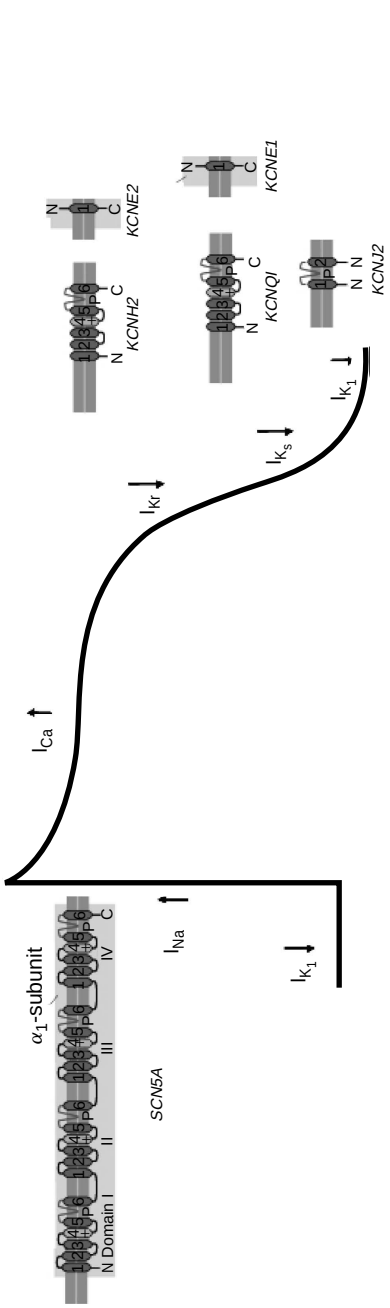

Fig. 8 Cardiac ion channel LQT genes. Each channel subunit gene product shown carries known LQT mutations. ↑ signifies depolarizing current, excess leading to LQT; ↓ signifies hyperpolarizing current, deficiency leading to LQT.

reported. Channels encoded by pathogenic alleles have delayed or decreased inactivation after opening, leaving excess inward depolarizing current during the plateau phase, delaying the time at which potassium currents can bring about repolarization. They are thus dominant "gain-of-function" mutations.

A distinct set of *SCN5A* alleles produce the *opposite* effect on inactivation from the *LQT3* alleles; they produce a rapid recovery from inactivation. They produce a different dominant arrhythmia syndrome, Brugada syndrome. The myocardium of these individuals contains a mixed population of sodium channels that are no longer locked in synchrony by a common period of inactivation, forming an ideal substrate for arrhythmia. Some alleles, surprisingly, are capable of producing either sodium channel phenotype within a family, demonstrating the delicate balancing act between excitation and inactivation carried out by the sodium channel that will be more clearly illustrated in the skeletal muscle disorders, and hinting at modifier genes that alter this balance, and therefore polygenic disorders.

LQT4 is the exception that proves the rule that LQT is an ion channel disease. The *LQT4* locus, found at 4q25 encodes ankyrin-b, a scaffold protein that colocalizes and regulates the Na/K-ATPase, the Na/Ca exchanger and the IP3 receptor. It is not a channel, but pathogenic loss-of-function alleles disturb calcium homeostasis through the functional loss of these critical mechanisms. It thus creates a dominant arrhythmia syndrome that includes LQT, but is not restricted to LQT.

LQT5 and *LQT6* are two adjacent loci on chromosome 21q22.1, and their genes, *KCNE1* and *KCNE2*, respectively, encode the two highly homologous potassium channel β-subunits discussed above as the accessory subunits that interact with their α-subunit partner to form I_{Ks} and I_{Kr}, respectively (Fig. 8). Both are proteins with only a single transmembrane α-helix and both function only to modify the behavior of the α-subunit with which they multimerize. Dominant pathogenic alleles are missense and effectively achieve dominance by altering channel gating to produce less current.

LQT7 is an alternative name for Andersen syndrome, a dominant multisystem disorder that includes long QT, but also extracardiac findings such as periodic paralysis and, more surprisingly, dysmorphology. Over nine different dominant-negative alleles of the potassium channel gene *KCNJ2* are responsible for this syndrome. The *KCNJ2* channel subunits multimerize to form the inwardly rectifying potassium channel that governs the resting membrane potential of the cardiac myocyte (Fig. 8). Current through this channel participates at the very end of repolarization, and hence dominant-negative loss of its function prolongs the process. The extracardiac findings imply that it also contributes in a significant way to membrane signaling in muscle and perhaps to the developmental processes underlying the dysmorphology.

5.6 The Recessive LQT Alleles

The recessive JLN syndrome differs from the dominant RW syndrome not only in its pattern of inheritance but also in its extracardiac manifestation of a constitutive sensorineural hearing loss. While in both syndromes the cardiac arrhythmia phenotype is lethal and intermittent, the striking differences between the two syndromes would never lead you to guess they might be allelic. In fact, the JLN alleles

are strong loss-of-function or null alleles in the same genes that cause dominant *LQT1* and *LQT5*, *KCNQ1* and *KCNE1*, the two channel genes encoding the I_{Ks} channel (Fig. 8). Surprisingly the same channel is expressed selectively in the heart and the inner ear, where its abundant expression is involved in the production of the potassium-rich endolymph fluid that fills this chamber and participates in the transduction of auditory sensation. Only complete loss of the channel significantly disturbs this secretory process. It is of note that another inner ear *LQT1* homolog, encoded by *KCNQ4,* has dominant alleles that produce only deafness without LQT (*DFNA2*).

5.7 Acquired LQT Channelopathies

While heart failure is the very common end-stage phenotype of a dilated and poorly contractile myocardial pump, and is produced by a variety of common insults to the heart, most deaths in this disorder arise from a fatal arrhythmia, and it likely reflects one of the most common forms of the LQT syndrome. The failing myocytes downregulate expression of potassium currents, perhaps in an attempt to enhance contractility by prolonging the period of calcium entry, the QT interval, but doing so predisposes them to the arrhythmia discussed above.

5.8 Pharmacogenetic and Polygenic Arrhythmia Syndromes

Initial studies suggested that LQT was a highly penetrant Mendelian phenotype with "strong" alleles that, while clinically silent, would dependably be revealed by EKG. Not surprisingly, it has become clear that there are both "weak" and "subclinical" alleles of these genes as well and likely polygenic interactions between them. Additionally, there is clearly an environmental component to the disease, seen as a pharmacogenetic syndrome, which greatly broadens the scope of the diseases involving these cardiac ion channel genes. These more subtle features of the disease became most apparent in studies on families initially considered to have a case of "sporadic" LQT. While half of the "sporadic" probands in the study had the predicted new dominant mutation in one of the LQT genes, the other half were found *not* to have new mutations, but rather to have families segregating a "weak" pathogenic LQT allele. This allele in these families had many silent carriers with no EKG abnormality, or they had carriers who only expressed the abnormality once taking a medication with potassium channel blocking activity. The silent carriers presumably lacked a phenotype because they lacked other unidentified susceptibility alleles at other loci carried by the proband. It is easy to imagine, based upon the descriptions above of the interactions between the multiple normally and abnormally functioning cardiac ion channels in maintaining and terminating the plateau phase of the cardiac action potential, how minor functional variants in these same genes could sum with one another to produce a major alteration in the plateau phase and hence produce a polygenic LQT, arrhythmia or sudden death phenotype. It is likewise easy to imagine how they might sum with a drug effect that acts like one of these mutations by blocking a potassium channel.

The pharmacological induction of the phenotype in an individual such as those in the Priori study suggests that environmental or genetic liability can sum with that

contributed by the weak LQT alleles. Since the drugs that unmask the phenotype produce potassium channel block, similar to known LQT mutation, so presumably do the hidden unmasking liability alleles. The pharmacogenetic LQT syndrome is clinically quite important; causing the recall of several otherwise valuable drugs, and is now recognized as a potentially lethal side effect of many common medications. These include antihistamines, such as Seldane, antibiotics and cisapride, as well as the more predictable antiarrhythmics. Most of the implicated drugs share a common mechanism of action that involves the block of I_{Kr} potassium channels, and a number of cases now explicitly demonstrate that weak or subclinical alleles in the LQT genes, encoding both Na^+ and K^+ channels, are contributory to the pharmacogenetic disease.

A special consideration of pathogenic weak LQT alleles is raised by the contribution of these genes to sudden infant death syndrome (SIDS). In a landmark prospective study spanning an 18-year period, Schwartz and coworkers from nine large maternity hospitals performed EKGs on all healthy newborns between the third to fourth day of life, studying a total of over 34 000 neonates. At the 1-year follow-up evaluation, it could be observed that infants who had died of SIDS had a longer QTc interval *at birth* than did survivors or infants dying of other causes, even though none had a positive family history for LQT. Half of the infants dying of SIDS (12 of 24) had a QTc at birth greater than two standard deviations above the mean for the cohort, a finding that alone predicted greater than a 40-fold increase in the odds of dying of SIDS. However, because SIDS is so rare, despite the finding's dramatic odds ratio, only 2% of the neonates found to have a long QT die of SIDS. More recent studies have begun to add a molecular dimension to these cases, and for a few cases, a new mutation in one of the LQT genes, producing an allele previously seen in a family with the classical syndrome, can be identified. It remains to be determined whether weak inherited mutations, or even polymorphisms, in two or more of the LQT genes contribute to the rest.

6 Ion Channel Diseases of Muscle

6.1 Function of Ion Channels in Muscle

In many ways, the skeletal muscle AP is similar to the initial and final portions of that in the heart, but lacks the long plateau phase (Fig. 7). It is a fast depolarization, driven by sodium and calcium currents, followed immediately by a rapid repolarization driven by potassium currents, and in this regard is much like that of nerve (Fig. 1a). Instead of arising endogenously within the tissue, as at the cardiac sinoatrial node, the inciting muscle depolarization arrives *synaptically* via the quantal release of a chemical neurotransmitter, acetylcholine (ACh), from a motor nerve terminal bouton that forms a specialized neuromuscular junction with the *motor endplate* of the muscle plasma membrane, or *sarcolemma*. As the motor nerve's propagating AP reaches its terminus, cytosolic calcium rises, triggering a few hundred *synaptic vesicles*, each holding several thousand ACh molecules, to fuse into the presynaptic neuronal membrane (Fig. 9). As they fuse, each vesicle releases, in a quantal fashion, all its contents. This released ACh diffuses, in a matter of milliseconds, across the *synaptic cleft* to the muscle, where it is degraded by acetyl

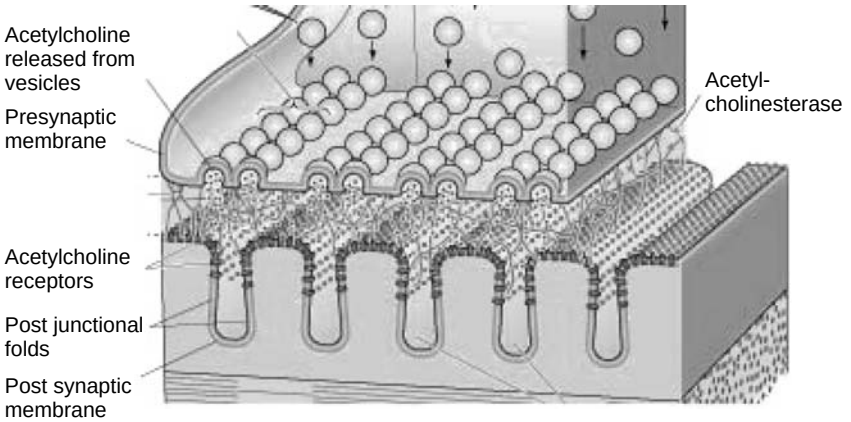

Fig. 9 Neuromuscular junction.

cholinesterase. A ligand-gated cation channel found on the postjunctional folds of the sarcolemmal endplate membrane, the heteropentameric nicotinic ACh receptor (nAChR), is activated by ACh binding (Figs. 5 and 9). Each receptor activated progressively brings about an initial incremental muscle membrane depolarization, and these events, referred to as *miniature endplate potentials* (MEPPs), were critical in uncovering the quantal nature of synaptic transmission. This receptor is the target of autoimmune and genetic myasthenia syndromes, and these quantal synaptic events are disturbed by these diseases, which are discussed below. In the physiological condition, a large number of tiny receptor-mediated MEPPs summate, producing an endplate potential that brings the membrane potential to the *threshold* at which a few sodium channels begin to open (Fig. 10). The sodium channels in turn further depolarize the membrane, causing more adjacent sodium channels to open in a reinforcing cycle until all sodium channels in the muscle have explosively opened while rapidly spreading a depolarizing wave across the surface of the muscle. At this point, all of the sodium channels are rendered "inactive" and nonexcitable, beginning the process of repolarization, a process completed when more slowly activating potassium channels restore the resting potential and return the sodium channels to the "closed" state. Unlike the situation in cardiac muscle, the rapid recovery of skeletal muscle excitability allows rapid repetitive contractions, an advantage in performing fine precision movements. The major subunit of muscle sodium channels is the muscle-specific isoform α_4, encoded by *SCN4A*, a relative of the gene involved in *LQT3*. Alleles of this one gene produce the diverse array of muscle phenotypes that will be discussed below.

As the depolarization wave of the muscle AP spreads across the sarcolemmal plasma membrane, it extends into specialized membrane invaginations, called *T-tubules*, which dive deep into the cytoplasm and ramify near the contractile proteins (Fig. 9). Their function is to facilitate calcium-dependent activation of contraction throughout the large muscle fiber. In the muscle T-tubule, as in the motor nerve bouton, the depolarization wave activates the voltage-gated calcium channel, which is similar in structure to

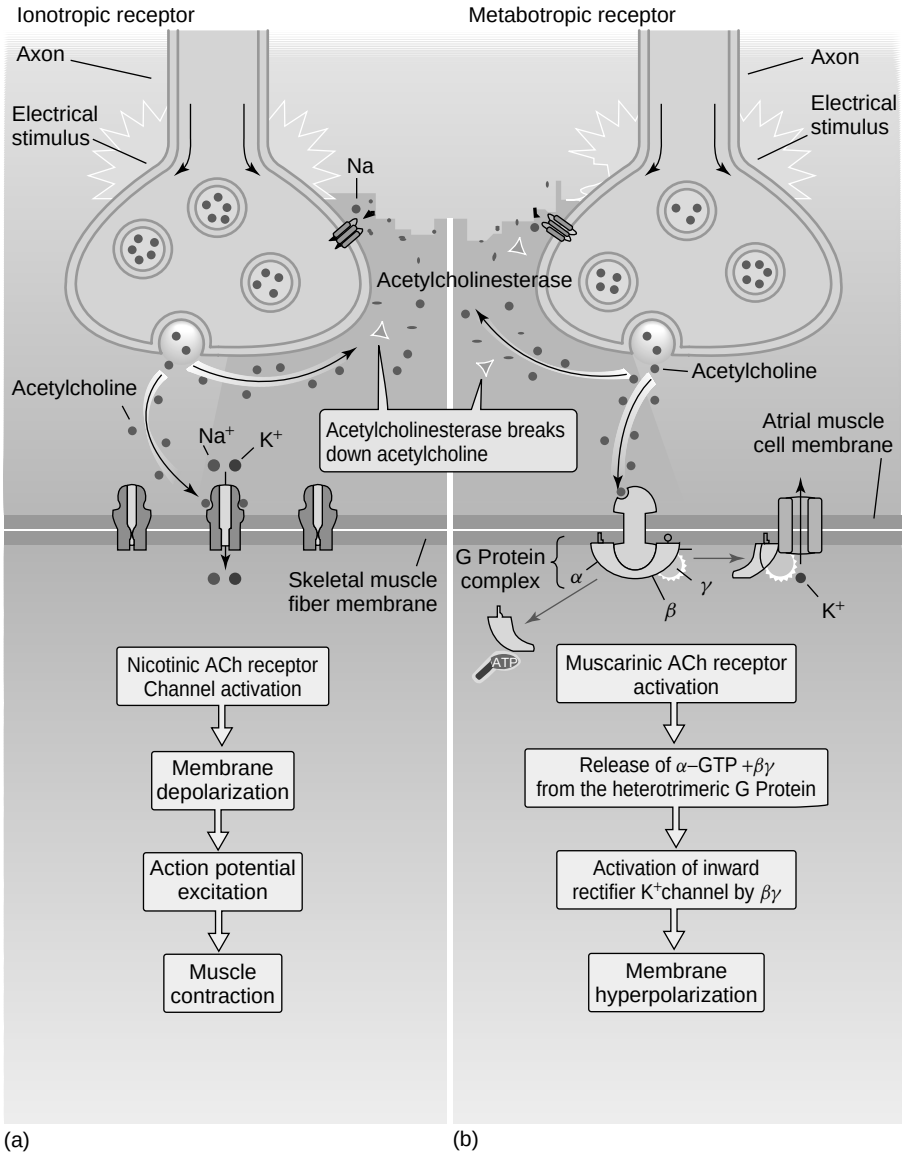

Fig. 10 Mechanisms of acetylcholine receptor activation.

the sodium channel, discussed above. The major, α_1, subunit is pseudotetrameric and contains the ion-conducting pore. Its function is modified by auxiliaryβ, γ and α_2/δ subunits, the skeletal muscle calcium channel's isoforms being encoded by *CACNA1S, CACNB1, CACNG1,* and *CACNA2D1,* respectively (Fig. 5). This channel is responsible for rapidly immersing the contractile proteins throughout the large muscle cell in elevated concentrations of ionized calcium, thereby

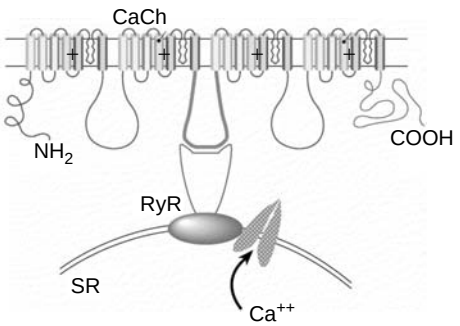

Fig. 11 Ca channel and Ryanodine receptor.

bringing about concerted contraction, much as was described for the cardiac contractile machinery. It does this in part by allowing the passage of extracellular calcium through its pore, down its electrochemical gradient into the cytoplasm. But additionally a large cytosolic loop in the skeletal muscle α_1-subunit contacts a *different* calcium channel located in a closely opposed *different* membrane compartment, the sarcoplasmic reticulum (SR) (Figs. 5 and 11). The SR membrane contains calcium pumps that rapidly sequester intracellular calcium after each AP, creating a releasable calcium store and restoring muscle relaxation. Through the physical interaction with its cytosolic loop, the voltage-gated calcium channel directly gates the SR calcium release channel. The release channel is called the *ryanodine receptor*, encoded by *RYR1*, and opening it spills the SR intracellular calcium stores into the cytoplasm. After activation, both the sodium and calcium voltage-gated channels spontaneously enter the nonconducting "inactive" conformation, beginning the process of membrane repolarization, a process completed by potassium channels that hyperpolarize the membrane, resetting the sodium and calcium channels into the "closed" conformation. The sodium and calcium channels then are kept closed by the large stabilizing current of the chloride channels, encoded by *CLCN1* (Fig. 5).

6.2 Ion Channel Dysfunction in Paralysis and Myotonia

The two major skeletal muscle phenotypes produced by ion channel disorders reflect either an inability to contract the muscle, ranging from weakness to *paralysis*, or an inability to relax the muscle, *myotonia*. While a third muscle phenotype, malignant hyperthermia (MHS), appears quite different, it is surprisingly closely related to the paralysis syndromes, in fact, being allelic. Two major themes in the ion channel diseases begin to develop in discussing the skeletal muscle diseases. First, it becomes clear that the LQT syndrome provides a powerful paradigm for understanding the types of alleles that produce disease and the mechanisms by which they act. Second, the proclivity of certain ion channel gene families to produce disease begins to emerge as does their ability to produce distinctly different clinical phenotypes with subtly differing alleles (Table 1).

While the pathophysiology of paralysis and hyperthermia is complex, that of myotonia is fairly straightforward. Since the high chloride conductance of *CLCN1* in the muscle membrane serves to stabilize

Tab. 1 Channelopathy gene families.

Gene family	*Tissue*	*Disease phenotype*	*Pathogenic mutation*
KCNE1	Heart/ear	LQT5/JLN	Dominant negative/recessive loss-of-function
KCNE2 (100%)[a]	Heart	LQT6	Dominant negative
KCNE3	Muscle	HOKPP	Dominant negative
KCNQ1	Heart/ear	LQT1/JLN	Dominant negative/recessive loss-of-function
KCNQ2	CNS	BFNC1	Dominant loss-of-function
KCNQ3 (80%)[a]	CNS	BFNC2	Dominant loss-of-function
KCNQ4	Ear	DFNA2	Dominant missense
KCNQ5	CNS	6q14	
RYR1	Muscle	MHS1	Missense
RYR2 (66%)[a]	Heart	Ventricular tachycardia	Missense
RYR3	CNS	15q14	
SCN1A	CNS	GEFS + /SMEI	Gain-of-function/loss-of-function
SCN2A	CNS	GEFS+	Gain-of-function
SCN3A	CNS	2q24	
SCN4A (40%)[a]	Muscle	HYPP/HOKPP/MHS/etc.	Gain-of-function
SCN5A	Heart	LQT3/Brugada S	Gain-of-function
SCN6A and *SCN7A*	Neuronal	2q23	
SCN8A	Motor endplate	12q13	
SCN9A	Neuroendo	2q24	
SCN10A	Nerve/muscle	3p22	
SCN11A and *SCN12A*	Sensory neurons	3p24	
SCN1B	CNS	GEFS+	Gain-of-function
SCN2B	Neuronal	11q23	
CACNA1S	Muscle	HOKPP/MHSS	Missense
CACNA1A	CNS	FHM1/EA2/SCA6/seizures	Many types
CACNA1F	Retina	Night blindness	Hemizygous loss-of-function
CHRNA1	Muscle	SCCMS	Gain-of-function

Sources: Gargus, J.J. (2003) Unraveling monogenic channelopathies and their implications for complex polygenic disease, *Am. J. Hum. Genet.* **72**(4), 785–803.
[a] Percent of gene family members already associated with disease.

the membrane potential, thereby preventing "echo" contractions after an induced contraction, its functional loss is associated with the inability to normally relax the muscle, and hence the slow muscle relaxation and muscle stiffness of myotonia. Both dominant and recessive mutations of this gene, which decrease channel current, produce myotonia and a characteristic "dive bomber" EMG record of repetitive contractions. The repetitive volleys of contraction can arise in skeletal muscle because unlike in the heart, where sodium channels are held in the inactive state for

a prolonged period to prevent arrhythmia, in skeletal muscle, sodium channels must be ready to accept a rapid train of controlled activations from the motor nerve. Problems arise, however, in the absence of a strong chloride stabilizing current, since instability of the membrane potential during recovery from a depolarization leaves a high probability that a secondary "echo" depolarization will reach threshold and trigger a repeated activation of the sodium channels.

The recessive loss-of-function alleles of *CLCN1* produce Becker myotonia, with a phenotype characterized by stiffness that warm-up exercise improves, and paradoxically weak hypertrophied muscles. The dominant-negative alleles leave more residual current and produce Thomsen myotonia, having a similar but milder phenotype without weakness. The hypertrophy arises because the muscles are constantly "exercising" even at rest secondary to their electrical instability, and the weakness arises because of incomplete recovery from sodium channel inactivation, as will be described for the periodic paralysis syndromes below.

6.3 Periodic Paralysis Syndromes

To understand how the array of periodic paralysis syndromes arise, it is essential to understand how very subtle "tweaks" in sodium channel function give rise to macroscopic phenotypes that are quite distinct – superficially the antithesis of one another – yet are so closely related that a given patient can transition from one state to the other. All of these pathogenic muscle sodium channel alleles appear to share with the cardiac sodium channel *LQT3* alleles a dominant gain-of-function mechanism of pathogenesis – *delayed and incomplete inactivation*. The essential biophysical feature of sodium channel function that makes this interpretable is the fact that while a depolarization will make it more probable that the voltage-sensitive gate of the closed channel will open, when sustained, depolarization also leaves behind more of the channels in an "inactive" nonexcitable conformation. As the sodium channel is the major mechanism for propagating APs in muscle (as it is in most tissues), this delicate balancing act has the potential to move these tissues from a state of hyperexcitability to a state of inexcitability, manifested as weakness and paralysis. Plasma potassium levels are commonly altered during these attacks of paralysis, and, while incompletely understood, probably has to do with the fact that the vast majority of the potassium in the body, >95%, is found in the muscle. The anxiety that surrounds clinical management of such changes is that very small amounts of potassium dramatically change the plasma potassium concentration and this is a critical determinant of the potassium Nernst potential (hence the membrane potential of all cells) and such changes in the heart pose the risk of fatal arrhythmia.

The ranges of phenotypes gain-of-function *LQT3*-like incompletely inactivating *SCN4A* alleles produce include: (1) hyperkalemic periodic paralysis (HYPP), (2) hypokalemic periodic paralysis (HOKPP), (3) paramyotonia congenita, a cold exacerbated myotonia and (4) potassium-activated myotonia.

HYPP alleles are the most common, and the disorder is characterized by short, mild frequent attacks of profound weakness beginning in infancy and provoked by rest just after exercise or stress. Commonly, the individual performs significant physical exertion, then sits to rest and finds he is unable to move for an hour

or more (although respiration is not compromised). There is often myotonia between attacks. During an attack, plasma potassium levels rise to pathological levels, probably via release from the depolarized muscle. The alleles are characteristically missense mutations altering amino acids in the transmembrane spans of the channel protein. Andersen syndrome, discussed above as caused by *LQT7* alleles, additionally has the muscle phenotypes of HYPP and that of MHS, discussed below, suggesting an intrinsic relation between the two. Furthermore, some HYPP alleles of *SCN4A* additionally produce MHS, and the gene is probably also the *MSH2* locus, a rare cause of the phenotype.

HOKPP is characterized by an onset in the second decade of infrequent, profound painless episodes of weakness that can last for days. During an attack, the plasma potassium can fall to dangerous levels, presumably being driven into the muscles. The factors known to incite the attacks are known to physiologically have this activity. They include insulin or glucose intake (which stimulates insulin release) and the hormonal changes that occur upon awakening. Two other loci, *CACNA1S* and *KCNE3*, also have alleles that produce HOKPP. Missense dominant alleles of *CACNA1S*, encoding the major α_1, subunit of the muscle voltage-gated calcium channel, were the first recognized and remain the most common cause of hypokalemic periodic paralysis. Most recently mutations in one of the skeletal muscle potassium channels involved in the rapid repolarization of the membrane, encoded by *KCNE3*, a relative of the *LQT5* and *LQT6* genes also have been shown to cause HOKPP. With this discovery, all of the members of the fully ascertained *KCNE* gene family are proven pathogenic loci, suggesting such familial relationships may be useful in promoting candidate genes for evaluation in disease association (Table 1).

6.4 Malignant Hyperthermia

The pathophysiology of malignant hyperthermia is typical of a pharmacogenetic syndrome; it involves an inciting agent, typically an inhalation anesthetic such as halothane, and an individual with a predisposing genotype. The pathogenic mechanism arises from a dramatic rise in muscle cytoplasmic calcium as the hypersensitized SR stores are discharged. This then secondarily produces contracture of the muscles and a hyper-metabolic state of glycogenolysis, ATP depletion, and lactic acidosis driven by the attempt to resequester calcium. This process evolves to produce hyperthermia (not fever) from activity of the contractile machinery and mitochondrial respiration, and disruption of the muscle membranes, releasing myoglobin, creatine kinase, and potassium, and leading to renal failure, disseminated intravascular coagulation and death The predisposing genotypes enhance the sensitivity of the calcium release process, which primarily involves the sarcolemmal calcium channel and the SR calcium release channel. Missense alleles altering either the cytosolic loop of the *CACNA1S*-encoded α_1 calcium channel subunit or its contact domains on the ryanodine receptor render this complex hypersensitive and capable of triggering massive calcium release, muscle activation, and malignant hyperthermia (Fig. 11). Over 50% of the cases are caused by dominant *MHS1* alleles of *RYR1* and the next most common cause is dominant *MHS5* alleles of *CACNA1S*. Some of the HYPP alleles of *SCN4A* additionally seem

to produce this syndrome, and it probably reflects *MHS2*. Three other *MHS* loci have been mapped but susceptibility alleles have yet to be identified. Additionally, this phenotype is also a component of Andersen syndrome, discussed above, affecting the potassium channel that controls the resting membrane potential in muscle and heart, *KCNJ2* (Fig. 8). Finally, alleles of *RYR1* also cause central core disease (CCD), a congenital myopathy characterized by nonprogressive proximal muscle weakness and distinctive muscle fiber histology. While individuals with CCD do have a susceptibility to anesthetic-induced MHS, the remainder of the disease phenotype is constitutive and does not require pharmacological induction, and therefore the disorder is illustrative of the continuum between genetic and environmental disease.

6.5 Other Monogenic Muscle Ion Channel Syndromes

Dominant gain-of-function mutations in four of the five subunits that compose the nAChR found at the neuromuscular junction and encoded by *CHRNA1*, *CHRNB1*, *CHRND* and *CHRNE* produce slow-channel congenital myasthenic syndrome (SCCMS), a phenotype characterized by muscle weakness. Acetylcholine dissociates slowly from these mutant receptors, leaving them persistently activated, depolarizing the membrane. One might guess that the prolonged receptor activation the mutant alleles produce will give a hyperexcitable muscle; however, as was discussed above in the periodic paralysis syndromes, the phenotype seen is produced by the loss of sodium channels into the inactive state in this chronically depolarized tissue.

6.6 Autoimmune Channelopathies in Muscle

Myasthenia gravis (MG) is one of the classic autoimmune diseases, first placed on a molecular footing in the early 1970s, well before the era of the monogenic channelopathies. It develops in mid-life, produces facial, respiratory, and limb muscle weakness and is associated with tumors of the thymus and antibodies to the nAChR. It is clear that the antibodies produce the disease, since they can passively transfer the symptoms and their removal is therapeutic. The autoantibodies accelerate the destruction of the nAChR, reducing their abundance in the synapse and hence the efficacy of synaptic transmission. This was first observed as a decrease in the size, but not number, of MEPPs, suggesting that the ACh-filled vesicles were normally releasing quanta of neurotransmitter, but that they had a smaller postsynaptic effect, causing weakness. Enhancing the synaptic concentration of ACh by inhibiting acetyl cholinesterase rapidly improves strength, and is used diagnostically and therapeutically (Fig. 9). Note that the muscle phenotype is very similar to that created by SCCMS alleles of the nAChR subunit genes, discussed above, but that those genetic syndromes do not have thymus tumors or autoantibodies, confirming that receptor dysfunction does not cause those findings in MG, the tumor appearing to be primary and causal in autoimmunity.

Lambert-Eaton myasthenia syndrome (LEMS) is another autoimmune disorder that produces a clinical picture very similar to that of MG, but preferentially targeting proximal limb muscles and being associated with small-cell carcinoma of the lung. It additionally produces symptoms referable to the autonomic nervous

system, such as dry mouth, so it is clear that the epitope targeted by the antibodies is not strictly limited to muscle. In fact, the antibodies are directed at the P/Q-type voltage-gated calcium channels of the presynaptic neuron. This channel is encoded by *CACNA1A* and is similar to its family members expressed in heart and skeletal muscle, but it participates in excitation-*secretion* coupling (not EC coupling), serving as a calcium entry pathway that couples the propagated AP to the biochemical pathways underlying neurotransmitter release at the presynaptic bouton, to be discussed below. It is distinguished from its family members by its high affinity for ω-conotoxin. Antibodies do not cross the blood-brain barrier, so symptoms in the autoimmune disorders are confined to effects in the periphery. As will be seen below, very different symptoms arise from genetic lesions in *CACNA1A* where CNS symptoms predominate.

7 Ion Channel Diseases of Nerve

7.1 Function of Ion Channels in Neurons

The neuronal AP is very similar to that of muscle: typically, activation is initiated across a *synapse*, neuroreceptors and their endogenous or coupled channels are activated, the supra-threshold summated postsynaptic currents in turn trigger a rapidly spreading depolarization wave by voltage-gated sodium and calcium channel currents, this is rapidly followed by a repolarization driven by potassium currents, and the AP couples to biochemical effector pathways through a rise in the cytosolic calcium concentration. Despite this gross oversimplification of all of the diversity of neuronal structure and function, and their vastly more complex geometry, the major distinctions from muscle lie at the afferent and efferent ends of the neuronal transduction pathway, not in the propagated AP in between.

First, neurons have a vastly more complex variety of activating, as well as inhibitory, neuroreceptors and synapses than muscle, and they are primarily found at the branching *dendritic*, or afferent, ends of the prototypical neuron (Fig. 12). The receptors produce both depolarizing, excitatory, as well as hyperpolarizing, inhibitory, *postsynaptic potentials* (EPSPs and IPSPs) that, like the MEPPs of muscle, summate over time, but also through the complex spacial geometry of the arborized dendritic tree. This summation serves the "computational" function of the neuron; some combinations of dendritic inputs will drive the soma, or neuronal cell body, to initiate an AP that is propagated down the axon, while others that differ slightly in timing, spacial geometry of the activated synapses, or the mix of incoming neurotransmitters will not.

Second, when the neuron does "decide" to produce a propagated AP, the elevated cytosolic calcium wave does not participate in EC coupling, but rather in excitation-secretion (ES) coupling, the controlled release of neurotransmitter into another synapse (Fig. 10), as was briefly discussed for LEMS, above. In the neuron, the calcium-coupled neurosecretion events occur after the AP has been conducted down the *axon*, the long cellular projection that serves as the efferent and effector side of the neuron. There a wide variety of small-molecule neurotransmitters can be released into a synapse to initiate in the downstream neuron the process just described above (Fig. 12).

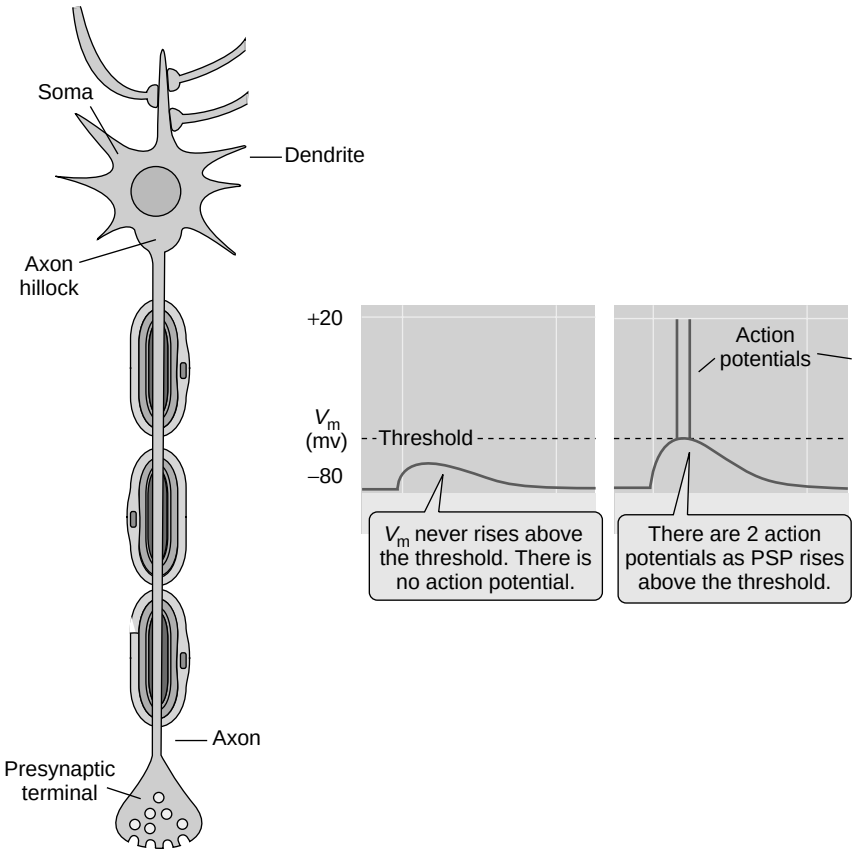

Fig. 12 Neuron structure and initiation of neuronal AP.

7.2 Ion Channel Dysfunction in Seizures

As opposed to the situation in heart and muscle, synchrony of electrical activity in the CNS is actively prevented. An abnormally synchronous discharge of neurons in the CNS produces a *seizure*, a pathological state characterized by a stereotyped alteration in behavior. Seizures are the most common monogenic ion channel phenotype in the CNS and in many ways are similar to the arrhythmia phenotype that arises in the hyperexcitable LQT heart. Both are periodic disorders in constitutionally hyperexcitable tissues in which the normal rhythmic electrical activity of the tissue is temporarily lost. However, whereas the heart is typified by its extremely homogeneous set of responding cells, the CNS displays maximal tissue complexity both in terms of the number of different cell types present and in the number of different ways they connect with and influence one another in stimulatory or inhibitory fashions. Therefore, while one can visualize the very specific and individual contribution each channel type makes to components of an action potential perturbed in LQT in the heart or in the skeletal muscle syndromes, one can much less specifically interpret how

Tab. 2 Unlike heart, must still guess WHERE neuronal channels act.

Seizures:		
BFNC1	*KCNQ2*	Haploinsufficiency, decrease I_K
BFNC2	*KCNQ3*	Haploinsufficiency, decrease I_K
ADNFLE	*CHRNA4*	GOF increase activation, increase I_{Na}
ADNFLE	*CHRNB2*	GOF increase activation, increase I_{Na}
GEFS+	*SCN1B*	GOF decrease inactivation, increase I_{Na}
GEFS+	*SCN1A*	GOF decrease inactivation, increase I_{Na}
GEFS+	*SCN2A*	GOF decrease inactivation, increase I_{Na}
GEFS+	*GABRG2*	Dominant missense, decrease I_{Cl}
JME	*GABRA1*	Dominant missense, decrease I_{Cl}
JME	*CACNB4*	Dominant missense, I_{Ca}
SMEI	*SCN1A*	Haploinsufficiency, decrease I_{Na}
...[f]	*KCNA1*	Haploinsufficiency, decrease I_K
...[f]	* *CACNA1A*	Dominant negative, decrease I_{Ca}
Ataxia		
...[f]	* *CACNA1A*	Dominant negative, decrease I_{Ca}
SCA6	* *CACNA1A*	Dominant CAG expansion, decrease I_{Ca}
EA2	* *CACNA1A*	Haploinsufficiency, decrease I_{Ca}
EA1	*KCNA1*	Haploinsufficiency, decrease I_K
Migraine:		
FHM1	* *CACNA1A*	Dominant missense, decrease I_{Ca}
FHM2	*ATP1A2*	Dominant loss-of-function

Sources: Gargus, J.J. (2003) Unraveling monogenic channelopathies and their implications for complex polygenic disease, *Am. J. Hum. Genet.* **72**(4), 785–803.
* Different manifestations of P/Q channel dysfunction;
f – case reports, discussed in text.

or even where the channel dysfunction underlying seizures occurs. Also, rhythm disorders in the heart are often a once in a lifetime event; in the CNS they are usually a lifelong chronic condition, *epilepsy*. Epilepsy is also a useful phenotype to serve as a springboard to understanding the other major monogenic CNS channelopathy phenotypes, *ataxia* and *migraine*, as they can be allelic.

While the complexity of neuroanatomy obscures functional analysis of the mutations causing monogenic epilepsy syndromes, extrapolation from the cardiac and muscle syndromes provides a simplified predictive framework of molecular pathology: K^+ and Cl^- channels, which physiologically stabilize excitable tissue, will likely have pathological lesions that diminish their current, and Na^+ and Ca^{++} channels, which physiologically excite the tissue, will likely have gain-of-function lesions (Table 2).

7.3 Hereditary Epilepsy Syndromes

Nearly 40% of all epilepsy is associated with no macroscopic pathological lesion, and is therefore classically referred to as *idiopathic epilepsy*, a common disorder that affects ~1% of the population. It is overwhelmingly a genetic disease, with monozygotic twins greater than 95% concordant for the phenotype; however, nearly all cases are polygenic and most of

those genes still remain to be identified. Obviously, since ion channels control electrical activity in the CNS, as in the heart and muscle, they are strong functional candidates for this disorder. In man, they were the first identified, and remain nearly the exclusive, proven genetic causes of epilepsy. Only a handful of rare monogenic epilepsy syndromes are recognized, but predictably success at pathogenic allele definition has thus far come only from these rare monogenic syndromes.

7.4 LQT-like Loci and Alleles in Neuronal Syndromes

The first rare monogenic seizure syndrome for which the major etiological genes were identified was benign familial neonatal convulsions (BFNC), a rare autosomal dominant disorder characterized by a brief period of seizures in the neonatal period, generally resolving in weeks, but 10% having persistent adult epilepsy. The BFNC loci were mapped to chromosomes 20 and 8. The *BFNC1* gene was positionally localized within chromosome 20q13.3 taking advantage of a family with the syndrome and a microdeletion chromosome. The interrupted locus contained a promising candidate gene, *KCNQ2*, a member of the *LQT1* gene family that was predominantly expressed in neurons. Unlike the *LQT1* alleles that demonstrate dominant-negative interactions *in vitro*, this first *BFNC1* allele was a null, and most subsequent alleles are simply loss-of-function. The relevant neuronal currents mediated by this channel must be so critically tuned that pathogenic alleles can achieve dominance simply via haploinsufficiency. Heterozygous null mice display a milder inducible pharmacogenetic syndrome, showing no basal seizures, but only an increased sensitivity to seizure-inducing drugs. Presumably, the strain's polygenic background raises their seizure threshold making them appear much like silent-carrier family members of the "sporadic" LQT syndrome probands who have disease produced by weak alleles.

Since the *KCNQ* family had the demonstrated ability to produce disease, other family member genes were sought by homology. *KCNQ3* was thus identified, and mapping to chromosome 8q, was a tempting candidate for the second BFNC locus, a hypothesis proven by finding loss-of-function alleles segregating in a family. While both KCNQ subunits produced channels when expressed *in vitro*, neither subunit alone produced a recognizable current. However, KCNQ2 and KCNQ3 were subsequently recognized to heteromultimerize to form the "M current," a long-sought signature potassium current activated by muscarinic acetylcholine receptors (mAChR) (Fig. 10). Therefore, both their gene-family relationship and their functional subunit interaction help explain the common phenotype that mutations in either gene produce. The final member of this *KCNQ* family, *KCNQ5*, mapping to chromosome 6q14, and the only member still to be associated with a disease, is expressed in neurons and can also interact with KCNQ3 to produce M current, making it an extremely strong functional candidate disease gene.

7.5 Other Monogenic Ion Channel Epilepsy Syndromes

The gene *CHRNA4* was the first ion channel gene demonstrated to contribute pathogenic alleles to an epilepsy syndrome. It is found on chromosome 20q13 and encodes the most abundant neuronal

isoform of the major subunit, α_4, of the nAChR. Unlike the mAChR, discussed immediately above, which couples to channel gating via G-protein activation (a G-protein coupled receptor, GPCR), the nAChR, like that found in muscle, is a heteropentameric receptor/channel that is *itself* a ligand-gated nonselective cation channel, its activation depolarizing the membrane (Fig. 10). It is closely related to the muscle nAChR, which participates in the phenotype of SCCMS congenital myasthenia. In the nerve, it is composed of two α- and three β-homologous subunits, each with four transmembrane α-helices (Fig. 5). The rare seizure syndrome, autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE), producing a phenotype with brief clusters of frontal seizures that occur at night, was mapped to this region in a family and missense alleles identified. To date, three alleles have been recognized, all altering the channel pore region and receptor function but having no obvious common effect on *in vitro* channel behavior. More recently, mutations in the most abundant neuronal β-subunit, β_2, were also recognized to cause this syndrome, suggesting that the channel isoform relevant to this phenotype is an $\alpha_4\beta_2$ pentamer. These missense alleles in *CHRNB2*, located on chromosome 1q21, produced the common *in vitro* effect of receptor activation, suggesting that some depolarizing gain of function is probably the mechanism of dominant pathogenesis, like that found in the SCCMS muscle pathogenic homologs. The vulnerable cell in the CNS whose function is perturbed by these mutations, however, remains to be defined.

Fever, a common accompaniment of many illnesses, lowers the seizure threshold so it is not surprising that febrile seizures are by far the most common polygenic seizure disorder, affecting 3% of children worldwide. While these susceptibility genes remain to be defined, a rare Mendelian dominant seizure syndrome that includes febrile seizures provides the first lead to their identification. This febrile seizure syndrome additionally evolves to include a variety of afebrile seizures, and is called *generalized epilepsy with febrile seizures plus* (GEFS+). The first locus responsible for GEFS+ was mapped in a large family to chromosome 19q13.1 and a missense mutation was identified in an auxiliary subunit of the voltage-gated sodium channel. This subunit, the β_1 subunit encoded by the gene *SCN1B*, has a single transmembrane helix and functions only to modify the activity of the large α-subunit (Fig. 5). Subsequently, mutations producing GEFS+ or closely related syndromes were identified in two different adjacent neuronal sodium channel α-subunit genes, *SCN1A* and *SCN2A*, on chromosome 2q24. Not until the α_1 subunit alleles were coexpressed with their auxiliary β-subunits did a reproducible picture of defective channel inactivation and sustained sodium current appear. This is a picture similar to that observed with the *LQT3* and HYPP and HOKPP alleles of the gene-family members *SCN5A* and *SCN4A*, respectively. As in the heart, most simply, a persistent sodium channel current can be viewed as a gain-of-function dominant lesion that favors a hyperexcitable state; however, the periodic paralysis lesions caution that hyperexcitability can easily transition into inexcitability, through the inactivation mechanism, and this dichotomous nature of sodium channel lesions is reinforced by the recent observations that a rare, very severe dominant seizure syndrome, initially associated with febrile seizures, called *severe myoclonic epilepsy of infancy* (SMEI), proved to be allelic with GEFS+. All of

dozens of known alleles were shown to be new mutations in *SCN1A*, and their nature (e.g. frameshift, nonsense) suggests they are predominantly functional nulls. While the physiological interpretation of these findings is clear, together they are puzzling; haploinsufficiency for the major subunit of the sodium channel produces a very severe dominant *loss*-of-function phenotype via some critically tuned cell type; a milder syndrome being caused by the dominant *gain*-of-function GEFS+ alleles.

The remaining GEFS+ locus, *GABRG2*, found within a cluster of GABA receptor genes on chromosome 5q31, encodes the γ_2 subunit of the inhibitory $GABA_A$ receptor. The $GABA_A$ receptor, like the nAChR discussed above, is a pentameric ligand-gated channel, but it differs in that instead of conducting a depolarizing cationic current, it conducts a stabilizing chloride current (Fig. 5). This receptor is the major mechanism through which GABA, the principle *inhibitory* neurotransmitter in the brain, functions. It is at this site that the benzodiazepine drugs act to treat seizures, by potentiating GABA effects, and that toxins, such as picrotoxin, act to induce seizures, by blocking GABA. So physiologically the inhibitory, neuronal silencing, activity of GABA participates in preventing seizures. The GEFS+ allele produces a decrease in GABA-induced chloride current *in vitro*. Therefore, like the seizure-inducing toxins, seizure-inducing mutations alter receptor/channel function to produce *less* GABA-induced stabilizing current.

While some human "seizure" genes have not recapitulated that phenotype in mice (for instance, mice homozygous for null alleles of *CHRN4A* and *CHRNB2* – the two loci thus far recognized with alleles causing ADNFLE in man – have a phenotype that alters the pain response, not seizures), spontaneous mutations producing a seizure phenotype in mice have led to the recognition of an important new category of ion channel seizure disorders in man, those altering voltage-gated calcium channel function. More importantly, those loci have proven to be particularly informative for their ability to tie the various neurological phenotypes together in an interpretable fashion. As discussed above, the voltage-gated calcium channels are similar to the voltage-gated sodium channel, and the major pore-containing α_1 subunit that dictates the channel's subtype is pseudotetrameric with its function modified by the auxiliary β, γ, and α_2/δ subunits (Fig. 5). Alleles in all four different types of neuronal voltage-gated calcium channel subunit genes produce a monogenic absence seizure phenotype in mice. The "lethargic," "stargazer," and "ducky" mice are recessive phenotypes produced by mutations altering neuronal calcium channel auxiliary subunits, *CACNB4* encoding a β-subunit, *CACNG2* encoding a γ-subunit, and *CACNA2D2* encoding a protein cleaved to produce the α_2 and δ-subunit, respectively. Of these three genes, a human phenotype has been reported only for mutations in the first. The human *CACNB4* gene maps to human chromosome 2q22-q23 and two dominant alleles, a truncation and missense allele, were found in families with a dominant seizure syndrome. The human *CACNG2* maps to chromosome 22q13.1 and the human *CACNA2D2* gene maps to 3p21.3, both remaining only promising seizure candidate genes in man. The "tottering" and "leaner" mice are recessive seizure phenotypes produced by alleles of *CACNA1A*, the gene encoding the major α_1 subunit of the neuronal P/Q-type calcium channel. The P/Q-type calcium channel plays a major

role in calcium entry underlying synaptic release of the major *excitatory* neurotransmitter, glutamate, and while this process is decreased in the mutants, suggesting they are hypomorphic alleles, it is unclear how this produces a hyperexcitable seizure phenotype.

7.6 Monogenic Ataxia and Migraine Syndromes

While the "tottering" and "leaner" mice are recessive seizure phenotypes produced by alleles of *CACNA1A*, a seizure phenotype is *not* the major phenotype caused by mutations in this gene in man. Human mutations in *CACNA1A*, which maps to chromosome 19p13, have been shown to cause three apparently distinct and different late-onset neurological disease phenotypes: episodic ataxia type 2 (*EA2*), familial hemiplegic migraine type 1 (*FHM1*), and spinocerebellar ataxia type 6 (*SCA6*). The *EA2* alleles are predominantly truncations, with over 22 known alleles, including frameshift and splice site mutations, but also five missense alleles, some altering conserved pore residues producing complete loss of function. It is therefore presumably a haploinsufficiency dominant loss-of-function syndrome. Familial hemiplegic migraine is caused by at least nine known missense alleles producing no obvious uniform functional change in calcium current *in vitro*; therefore, the relevant functional change they share in common is presumably subtle. Finally, SCA6 is a progressive degenerative phenotype primarily caused by trinucleotide expansion alleles. The polymorphic CAG repeat encodes a C-terminal polyglutamine repeat of 5 to 20 residues in unaffected individuals. Pathogenic alleles encode 21 to 30 glutamines. The only point mutant allele recognized to produce the syndrome is G293R, which changes a conserved pore residue. An intermediate-length CAG repeat allele with only 20 repeats produced the milder phenotype of EA2.

Thus, the four allelic *CACNA1A* diseases in man and mouse are clearly closely related at the molecular level. The phenotypes also can blend one into the other, for instance, many individuals with FHM have ataxia as do most of the murine seizure mutants, and recently a dominant-negative truncation allele has been reported to produce a syndrome of progressive ataxia and seizures in a child. In man, the mildest alleles appear to be the missense mutations producing FHM1. On a phenotypic continuum, these appear to be hypomorphic since the truncations, pore mutations, and intermediate repeat expansions produce EA2, apparently a haploinsufficiency syndrome. Expanding the polyglutamine repeat by only a few additional residues yields the potent long-repeat alleles producing the progressive SCA6 syndrome, which can also be produced by one specific missense pore mutation. While the mechanism of action of all pathogenic polyglutamine expansion alleles are controversial, if the three human *CACNA1A* phenotypes are truly the continuum suggested by the EA2 intermediate repeat phenotype, these repeat expansion alleles should be dominant negative to eliminate more of the P/Q current than haploinsufficiency and to keep the progression of allele potency parallel to the severity of the symptoms and pathology. While all of the human alleles produce the phenotype in heterozygotes, the mouse seizure alleles, which are missense, like FHM1, and demonstrated to be hypomorphic, produce no phenotype in heterozygotes, including knockout null alleles; decreased synaptic glutamate neurotransmitter release and

seizures being produced only in the homozygotes.

Another strong clue of the relationship between ataxia and seizures comes from the other known episodic ataxia locus, *EA1*. EA1 is produced by dominant loss-of-function alleles of perhaps the most "famous" potassium channel, "shaker", named after its mutant phenotype in fly, and the paradigmatic first K^+ channel cloned. It is a voltage-gated potassium channel formally called *KCNA1*. Knockout null alleles in the mouse produce a recessive seizure disorder, with the heterozygotes lacking a phenotype, just like the calcium channel mutants, but like them showing a reduced threshold for induced seizures, hinting at a potential mechanism for pharmacogenetic neurological syndromes in man. While the vast majority of EA1 patients have only ataxia, a few cases have been reported to additionally have seizures.

Just as examining both loci that produce episodic ataxia, *EA1* and *EA2*, informs the underlying pathophysiology of the indistinguishable diseases they produce, so too does examining both loci that cause familial hemiplegic migraine, *FHM1* and *FHM2*. Whereas *FHM1* alleles are subtle missense mutations of the calcium channel-encoding *CACNA1A* gene, recently *FHM2* alleles were discovered to be mutations in *ATP1A2*, the gene encoding the $\alpha 2$ subunit of the Na/K-ATPase, one of the major ion pump mechanisms responsible for establishing the ion gradients tapped by ion channels (Fig. 6). This is the first disease that is established to be caused by mutations in the Na/K pump, a central physiological mechanism. All FHM2 alleles identified to date are missense and alter narrowly conscribed domains of the pump protein. While loss-of-function and haploinsufficiency mechanisms are more suggestive of deletions and other forms of more readily produced null alleles than the observed constrained missense, this has been suggested to be the mechanism of some pathogenic alleles; however, detailed enzyme kinetics have more recently shown others to achieve dominance by gain of function, producing altered kinetics of the pump. It is still not clear what biophysical property of the neuron is perturbed in common by all of the FHM lesions, although a common denominator seems to be membrane depolarization and elevated cytosolic calcium.

7.7 Other Monogenic Neuronal Ion Channel Syndromes

One of the first monogenic CNS ion channel disorders for which the underlying pathogenic alleles were identified is the rare disorder hyperekplexia (HE), more commonly called *Startle syndrome*. The phenotype includes hypertonia and a severe startle reaction that simulates a drop seizure. It is produced by dominant and recessive mutations in *GLRA1*, the gene encoding the $\alpha 1$ subunit of the glycine receptor/channel. Since like the inhibitory GABA – and excitatory nAChR – receptor/channels, described above, the chloride-conducting inhibitory glycine receptor is a ligand-gated ion channel, sharing with other members of this channel family a subunit structure with four membrane-spanning helices and a characteristic heteropentameric composition (Fig. 5), it is not surprising that recessive mutations in the *GLRB* gene, encoding the glycine receptor β-subunit, also cause the same HE phenotype. Further, it is interesting that pathogenic alleles in all three types of ligand-gated channels, causing the discussed syndromes

of HE, SCCMS, ADNFLE, and GEFS+, are missense mutations primarily altering residues in the extracellular loop between the second and third transmembrane helix that controls gating, suggesting a related molecular pathogenesis of these macroscopically distinct diseases.

Many of the mutations that underlie hereditary defects in the special senses, such as vision and hearing, are found in ion channel genes. X-linked congenital stationary night blindness type 2 is caused by defects in the retinal-specific close relative of both the neuronal calcium channel "seizure-ataxia-migraine" gene *CACNA1A* and the skeletal muscle "HOKPP periodic paralysis" gene *CACNA1S*, discussed above. The syndrome is caused by several loss-of-function mutations in *CACNA1F*. Color vision requires, in addition to the three cone opsin pigment genes, mutations in which cause loss of color detected by one cone type, only the genes encoding the two subunits of a specific cyclic nucleotide-gated heteromultimeric ion channel found in the photoreceptor membrane (Fig. 5). Defects in these channel subunits cause color blindness due to dysfunction of all cones. Rod monochromacy is produced by at least 10 recessive missense alleles of *CNGA3* and Achromatopsia type 3 is produced by missense loss-of-function alleles in *CNGB3*.

Nonsyndromic deafness (NSD) is produced by mutations in dozens of genes, including those potassium channel genes that contribute alleles to produce Jervell and Lange-Nielsen recessive LQT syndrome, discussed above. Additionally, the genes encoding subunits of a very different type of channel, the connexin gap junctions (Figs. 5 and 13), are the most common source of alleles for NSD, being polymorphic in many populations. Gap junctions serve to connect all types of cells via a large nonselective pore that mediates the transit between cells of not only ions but also small macromolecules and metabolites. They are functionally dimers contributed by each member of a pair of cells, each cell membrane having a hexameric hemichannel that seals together with its partner. Each hemichannel is composed of the family of connexin subunits. Alleles of *GJB2*, encoding connexin 26, are the most common cause of NSD. There are a few dominant alleles, but several dozen recessive alleles, including one specific polymorphic pathogenic allele, 35delG, which accounts for as many as 4% of the European Caucasian alleles. Alleles of *GJB6* and *GJB3* also contribute to this syndrome. The mechanism by which they act appears to be similar to that of the Jervell and Lange-Nielsen LQT alleles; they perturb the formation of the potassium-rich endolymph fluid in the inner ear that is required for signal transduction. Curiously, some alleles of all three connexin genes also cause distinct hyperkeratotic skin syndromes with or without deafness.

7.8 Acquired CNS Ion Channel Syndromes

Just as cardiac failure produces an altered pattern of ion channel gene expression, and thereby creates an acquired channelopathy arrhythmia syndrome (see Sect. 5.7), experimental forms of induced temporal lobe epilepsy, the most common form of adult epilepsy, greatly reduce expression of the *KCND2* potassium channel gene, producing a downregulation of the neuronal A-type potassium current and a reduced threshold for seizure initiation in the hippocampus, a common origin for such seizures. Thus, seizure activity

produces a reduced threshold for subsequent seizures.

7.9 Polygenic Syndromes of Altered CNS Excitability

While most common seizure, migraine, and ataxia syndromes behave as complex polygenic traits, *not* as the simple monogenic Mendelian traits discussed above, and none of the genes involved in the polygenic forms of the disease have yet been convincingly demonstrated, a definition of the shared pathophysiology of these monogenic syndromes is likely to provide the best leads to the nature of those genes and to novel therapeutics for these common diseases. Likewise, while the definition of specific pharmacogenetic syndromes of altered neural excitability is only in its infancy, largely limited at this time to the definition of those individuals with altered rates of metabolism of seizure or antipsychotic medications, one would anticipate a dramatic impact on neuropsychopharmacology just by defining the genotypic basis of the wide variance now observed in patient response to current neuropsychiatric drugs. Additionally, it is likely that polygenic ion channel disorders produce a wider range of phenotypes in the CNS than those already revealed by the monogenic disorders. The salient feature of the CNS ion channel disorders, as well as those in the heart and muscle, was that they arose from a *periodic* disturbance in normal rhythmic activity. They were intrinsically episodic disorders, although many progressed to become constant. Many of the common complex polygenic neuropsychiatric disorders share this character. Further, several of the drugs used to treat the recognized channel phenotypes are additionally used to treat these disorders (e.g. valproate is used to treat both seizures and bipolar disease). Ion channel genes are thus important "functional candidates" to pursue within the broad chromosomal regions shown through linkage or linkage disequilibrium to be associated with these diseases, and are particularly intriguing should alleles be recognized that are similar in activity to those seen in the monogenic channel disorders.

8 Ion Channel Diseases of Epithelia

8.1 Structure and Function of Epithelial Tissues

Epithelia are sheets of *polarized* cells (sometimes consisting of many layers but only one cell layer of which is functionally significant) that participate in the transport of ions, water, and metabolites, not merely in and out of the cell, but fully *across* the entire cell sheet (Fig. 13). The epithelial cells are not excitable, like heart, muscle, and nerve, but they are electrically and metabolically coupled horizontally by gap junctions, described above, and the epithelial sheet therefore functions nearly as though it was two sheets of parallel membrane with cytosol between. One side of the epithelial sheet, called the *basolateral membrane*, faces the blood and is relatively unspecialized; the other, called the *apical membrane*, lines a nonblood compartment, often the lumen of a tubule, and the membrane is often elaborated into a brush border and microvilli (Fig. 13). It is primarily in the apical membranes that are found the many special channels and transporters that distinguish one epithelium from another. The point of demarcation between the two faces of

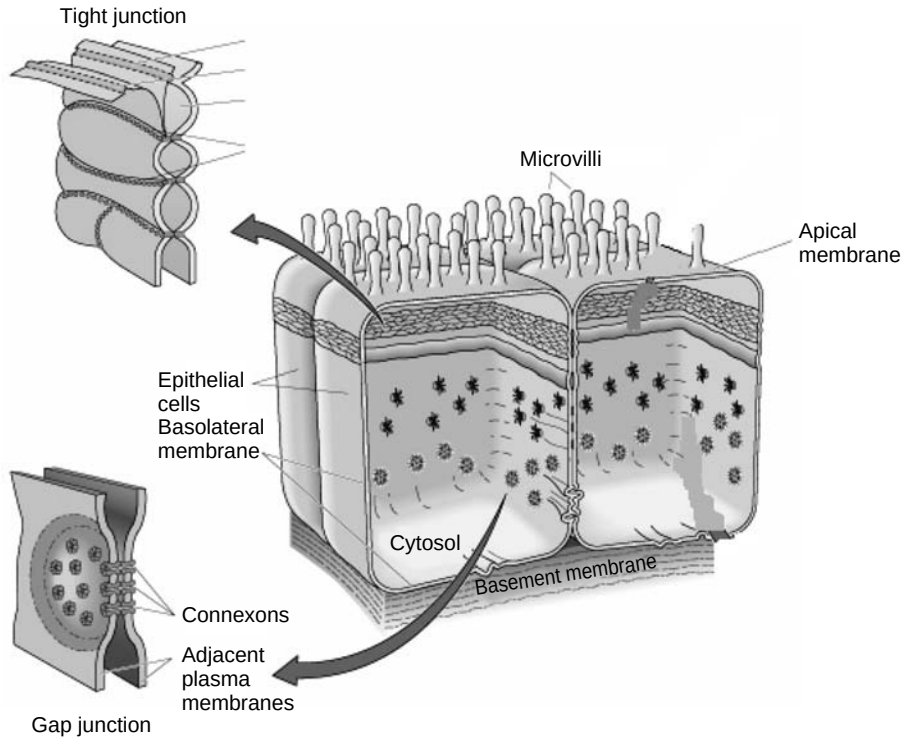

Fig. 13 Structure of epithelia.

an epithelial cell is the *tight junction*, the special barrier-forming connection made between adjacent cells in the epithelium (Fig. 13).

Epithelia serve two essential functions in physiology that can be perturbed in disease: to transport salt, water and other substances from one compartment to another, and to maintain gradients of salt, water, or other substances between compartments. While all epithelia share a fundamental organization and have in common a basic set of transport functions, various epithelia are optimized to perform one or the other of these two principle functions, and this is primarily conferred by the nature of the tight junctions between cells in the sheet. Those optimized for rapid transepithelial transport have ''leaky'' tight junctions and large *paracellular* pathways between cells providing a shunt parallel to the *transcellular* pathway across the sheet, and those optimized to form large electrical or chemical gradients across the membrane sheet have ''tight'' tight junctions that do not allow the passage of ions and water except through the cells.

8.2 Function of Ion Channels in Epithelia

Even though various epithelia transport a wide spectrum of compounds, there are a number of basic membrane properties that they all share, and these shared properties even extend to most nonpolarized mammalian cells, as introduced in Sect. 2. This

generalization is most particularly true for epithelial basolateral membranes, which are much less specialized and more like those of other cells. First there is a Na/K-ATPase ion pump located in the basolateral membrane (Figs. 6 and 14). This ATPase is oriented so as to extrude Na^+ from the cytoplasm in exchange for the uptake of K^+. Second, the electrical potential difference across the basolateral membrane is electrically negative (cell interior with respect to the blood) and on the order of 50–60 mV, like most mammalian cells. Third, the intracellular ion concentrations reflect the activity of the Na/K pump, with a high internal K^+, a low internal Na^+ and a Cl^- concentration near its Nernst electrochemical equilibrium potential (Fig. 6). Fourth, the passive permeability of the basolateral membrane is predominated by that for K^+, with low permeabilities to Na^+ and Cl^-. Since K^+ is not in electrochemical equilibrium across the K^+-permeable basolateral membrane, the −60 mV resting membrane potential is then a K^+ diffusion potential attenuated by low finite Na^+ and Cl^- permeabilities. Finally, the properties of the apical membrane of epithelia are very diverse, but commonly are predominated by a high Na^+ permeability. The specifics of these apical Na^+ permeabilities and their properties, however, vary such that broad generalizations can not be made. Major breakthroughs in understanding epithelial function occurred as Ussing defined how simply opening a passive Na^+ channel in the apical membrane of an epithelial sheet with the five cardinal characteristics described above allowed transepithelial transport of sodium, and later Curran defined how such vectorial transport of sodium could be coupled to the iso-osmotic vectorial transepithelial flow of water and other solutes. To allow these mechanisms to create a net flux, anion had to be allowed to passively accompany the transported Na^+ since macroscopic electroneutrality had to be maintained on both sides of the epithelium, as it required an enormous amount of energy to separate charge across a membrane. As will be discussed below, if there is no pathway for chloride movement, as occurs in cystic fibrosis, transepithelial salt and water flow are greatly impeded.

8.3 Ion Channel Dysfunction and Diseases in Epithelia

On one hand, the macroscopic phenotypes produced by defects in epithelial channels can be seen to reflect the wide range of specific uses to which Nature has put epithelia: failure of lung, kidney, digestive and other internal organs, deafness, systemic fluid and electrolyte abnormalities, and even abnormalities of morphogenesis. On the other hand, these phenotypes all ultimately can be recognized to arise from defects in either of the two cardinal functions of epithelia, transepithelial transport, or maintenance of transepithelial gradients.

8.4 Cystic Fibrosis

Cystic fibrosis is the most common simple genetic disease in Caucasians, nearly 5% of the members of those populations carrying one pathogenic allele for this lethal autosomal recessive disease, with one common allele, Δ F508, reflecting 70%, but at least four other pathogenic alleles also reaching polymorphic gene frequencies (~1%) and over 1000 alleles now being known. It is also the first genetic disease for which the etiological gene and pathogenic alleles were discovered strictly

on the basis of only its chromosomal position, without benefit of a known protein or a pathogenic allele created by a chromosomal rearrangement. It was additionally a landmark since it represented the first genetic channelopathy. The clinical phenotype of this disease – predominated by failure of lung, pancreas, and intestinal function – suggests nothing to bring to mind ion channel processes in the context discussed for the excitable tissues, and only slowly is the full molecular pathophysiology of this disease coming to light.

The gene, *CFTR*, standing for "cystic fibrosis transmembrane regulator," encodes a large transmembrane protein with 12 transmembrane helical spans, two cytosolic ATP-binding domains, and a large cytoplasmic regulatory domain that has multiple phosphorylation sites (Fig. 5). It is itself an ATP-regulated chloride channel, activated by the combination of both direct ATP binding and ATP-dependent phosphorylation via a cAMP-dependent kinase.

How does loss of function of this epithelial chloride channel produce the multifaceted phenotype of CF? The key features of the disease reflect epithelial dysfunction and include very thick dehydrated mucus secretions in the respiratory tree, which become purulent with chronic infections and progress to destruction of the lung parenchyma and therefore chronic respiratory failure. Additionally, a majority of patients have malabsorption caused by pancreatic insufficiency, caused by plugging of the secretory ducts resulting in activation of the digestive enzymes within the ducts and eventually pancreatic fibrosis (hence CF's old names "mucoviscidosis" and "cystic fibrosis of the pancreas"). As medical treatment allowed a longer lifespan, it was recognized that nearly all males at adulthood have infertility caused by plugged atretic vasa deferens and that a few percent of the patients first presented as newborns with meconium ileus, an obstruction of the newborn's intestines caused by cement-like intestinal contents. The physiology of these complex secretory processes was insufficiently understood to lead to an understanding of this disease or its diagnosis. However, the modern scientific underpinnings of CF began during a New York heat wave in 1953, when it was astutely observed that many of the children who were hospitalized for dehydration and salt depletion came to be diagnosed with CF. The dehydration arose because they secreted a salt-rich sweat. This observation led di Sant'Agnese and coworkers to the development of the diagnostic "sweat chloride test" based upon the recognition that the high concentration of chloride in sweat of CF homozygotes, well before they became symptomatic, could be distinguished from unaffected controls or carriers. Not until 30 years later did patch-clamp electrophysiology studies on the sweat gland by Quinton and coworkers establish that a defect in the function of an epithelial chloride channel prevented the physiological resorption of salt from sweat during its transit through the duct. Therefore, the cardinal aspects of transepithelial transport, salt and water movements, were impeded by the CF chloride channel defect. As the sweat gland was a *resorptive epithelium*, salt resorption was prevented, so salt was wasted in sweat. The same molecular lesion in the *secretory epithelium* of the respiratory mucus gland, the pancreatic acinus, and the vas deferens resulted in concretions of dehydrated product that plugged and scarred the ducts, ultimately destroying the tissues.

Despite the complicated name of the CFTR channel it is clear that this protein *itself* functions as an ion channel, not just an indirect regulator of other channels. This

point is most simply made by classes of mutant alleles that alter ion selectivity and single-channel conductance, functions intrinsic to a channel and *not* subject to its regulation. There are also informative classes that alter channel gating; however, this *is* a process subject to regulation, so such alleles do not prove intrinsic channel activity. Likewise, the vast majority of the mutant alleles are even less informative in that they merely alter the membrane protein's biosynthesis, maturation, and stability, telling little about the protein's function, but perhaps pointing to vulnerabilities that can be targeted in therapy.

8.5 Renal Channelopathies

The range of the diversity of apical sodium entry pathways available in epithelial designs is reflected in channelopathies of the kidney tubules, but mutations in the subunits of the heterotrimeric EnaC epithelial sodium channels, encoded by *SCNN1A, SCNN1B*, and *SCNN1G*, provide the simplest entry to the pathophysiology of these diseases. The α-, β-, and γ-subunits of this channel are homologous to each other and to the degenerins that participate as an olfaction and neurodegeneration channel in *Caenorhabditis elegans*. Each gene encodes a simple ion channel subunit that contains cytoplasmic termini and only two transmembrane helices (Figs. 5 and 6). Expression of the α-subunit alone produced ion channel activity *in vitro*, but it did not reflect a recognized native current. Neither the β- nor γ-subunits alone produced any channel activity. All three subunits expressed together, however, could be recognized to produce the amiloride-sensitive, aldosterone-regulated sodium channel best characterized in the principal cell of the distal nephron (Fig. 14d), but also found in the colon, lung, and exocrine glands, many places CFTR was expressed. Amiloride is a diuretic drug, and drugs we call diuretics are in fact naturetics in that they cause sodium wasting by inhibiting a renal sodium resorption mechanism. Each class of clinically useful diuretic targets a different renal epithelial apical sodium entry pathway, and, remarkably, mutations have been found targeting each of these pathways, simulating the effects of the diuretic drugs furosemide, thiazide, and amiloride.

Just as without a rudimentary understanding of transepithelial transport one would be unlikely to guess the phenotype of CFTR mutations, it takes at least a superficial appreciation of renal physiology to derive the phenotype produced by these mutations: alterations in blood pressure. These days we tend to forget that salt is a precious essential ingredient for life and that loss of body salt (much more than loss of body water, as will be discussed below) results in blood volume depletion and a syndrome of shock similar to that produced by a severe hemorrhage. While the sweat gland is a relatively simple structure that does a fair job of resorbing some salt from the small volume of blood ultrafiltrate that gives rise to the excreted fluid, the kidneys are Olympic performers in this field in that they cope with fully one-fifth of the cardiac output, sending one-fifth of that across the glomerular filter (G) into the epithelial nephron tubules, where in excess of 99.5% of the salt can be reabsorbed by crossing the luminal apical membrane and then the basolateral membrane back into the blood, leaving by default unspecified waste to be lost in the urine (Fig. 14). The epithelial sheet that forms the nephron tubule changes in character along its length, and classical renal tubule

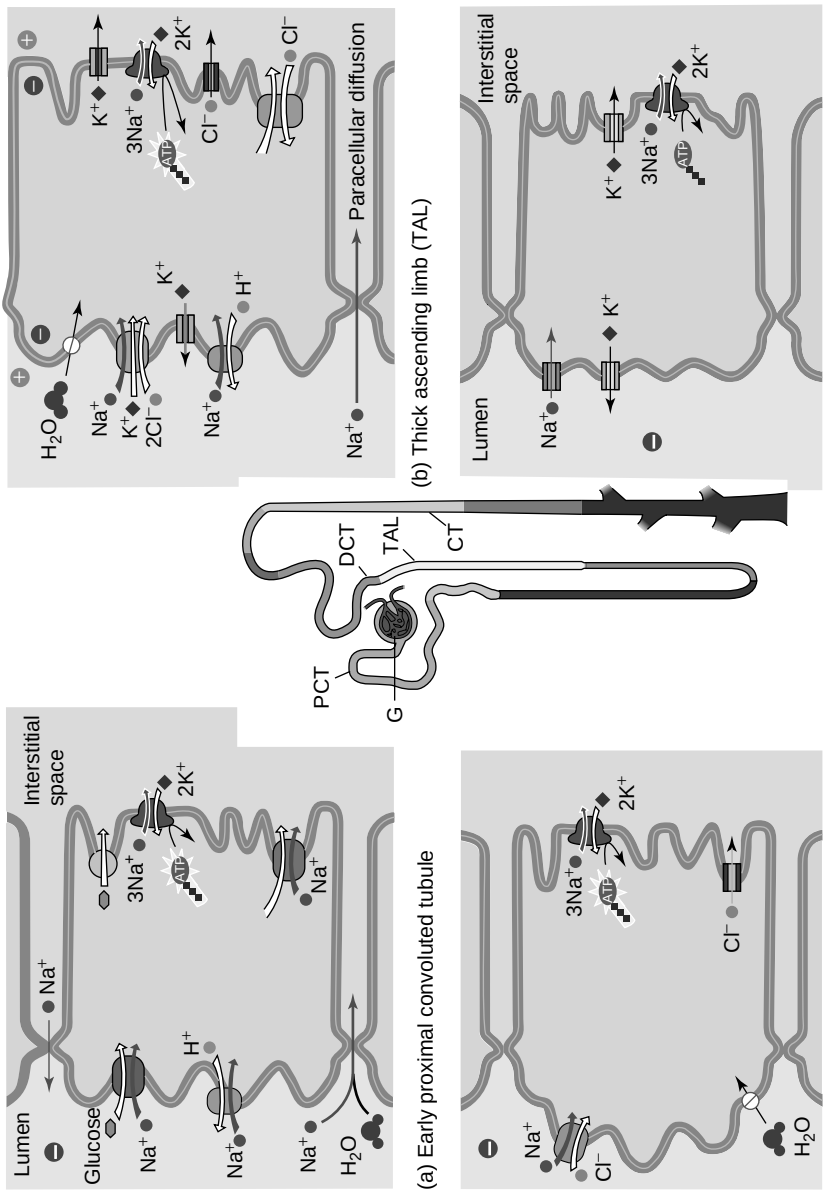

Fig. 14 Renal epithelia.

segment names have now come to be associated with specific molecular transport mechanisms. For instance, 66% of the filtered salt is resorbed in first segment, the proximal tubule (PCT), 25% in the thick ascending limb (TAL) of the loop of Henle via the furosemide-sensitive apical Na-K-Cl cotransporter, 7% in the distal convoluted tubule (DCT) via the thiazide-sensitive apical Na-Cl cotransporter, and the final 2% via the amiloride-sensitive, aldosterone-regulated heterotrimeric EnaC epithelial sodium channels in the collecting tubule (CT) (Fig. 14).

The kidney is capable of regulating sodium resorption over a wide range to compensate for changes in intake, using only the final 2% of the filtered load. This is all achieved by control of the abundance of the EnaC epithelial sodium channels in the principal cells of the CT by the adrenal steroid hormone aldosterone (Fig. 14d). Aldosterone is secreted in response to angiotensin II, which in turn is produced from angiotensin I by angiotensin converting enzyme (ACE), an important pharmacological target for the control of blood pressure. The entire cascade leading to aldosterone is initiated in response to sensors that detect aspects of blood volume, critical ones being the intrarenal and central baroreceptors that trigger specialized cells in the nephron to release an enzyme, called *renin*, which cleaves the inactive angiotensinogen precursor to angiotensin I. Therefore, hemorrhage causes a low blood volume and pressure to be sensed, triggering renin release, leading to the production of angiotensin and aldosterone, the insertion of excess EnaC epithelial sodium channels in the principal cells of the CT, enhanced distal nephron sodium resorption into the blood and a physiological trend to restoring blood volume. On the other hand, artificially producing high blood pressure will homeostatically decrease renin release, leaving angiotensin and aldosterone levels low, reducing the abundance of EnaC epithelial sodium channels and distal nephron sodium resorption, wasting sodium in urine and diminishing blood volume. Because the transepithelial electrical potential across the "tight" CT epithelium is used to drive the secretion of potassium and hydrogen into the nephron lumen, and that electrical potential results from EnaC sodium channel activation, a low abundance of EnaC sodium channels results in reduced K^+ and H^+ wasting in the urine, and hence elevated plasma levels of both. Conversely, high levels of EnaC sodium channels enhance K^+ and H^+ wasting and low plasma levels of both.

8.6 Pseudohypoaldosteronism and Liddle Syndrome

Loss-of-function mutations, primarily produced by deletions, in *SCNN1A*, *SCNN1B*, and *SCNN1G* genes, encoding subunits of the EnaC channel, produce the phenotype of autosomal recessive pseudohypoaldosteronism type 1. This is a severe disease that presents in infancy with life-threatening failure to thrive caused by salt wasting and hypovolemia. They additionally have hyperkalemia and metabolic acidosis. As the name implies, the phenotype appears like that produced by an absence of aldosterone; however, in this case aldosterone (as well as angiotensin and renin) levels are very high. One allele, *SCNN1B G37S*, replaces a conserved glycine residue of the channel with serine, thereby altering gating and reducing the channel open probability. This establishes that channel function *itself* is critical to the phenotype, not some other aspect of

the protein. This phenotype powerfully defines the physiology: all of the complex and apparently multifaceted actions of aldosterone can be understood to emanate from increasing the EnaC channel current, since loss of the hormone and loss of the channel appear indistinguishable. The sodium current conducted by the channel is the effector of aldosterone action; *a classical endocrine and hormonal pathway is truly a complex means of ion channel regulation, its diseases truly channelopathies.*

Gain-of-function dominant mutations in *SCNN1B* and *SCNN1G*, altering the C-terminus of the β- or γ-subunits of EnaC channels, cause a pseudo-hyperaldosterone state called *Liddle syndrome*. The phenotype is hypertension, with low renin, angiotensin and aldosterone levels, suggesting primary excess sodium retention and volume expansion. They also have hypokalemia and metabolic alkalosis, secondary to excessive renal K^+ and H^+ wasting. The mutations mainly result in increased numbers of EnaC channels in the CT; however, some alleles do increase the channel open probability, again showing the sodium current, and not some other aspect of the protein, to be critical. Since the diuretic amiloride blocks the EnaC channel, the drug is useful in treating the disease. Spironolactone, a drug that inhibits aldosterone action, however, has no effect, since the mutant channels are constitutively active and no longer under hormonal control. Not only does amiloride block the excessive sodium resorption, allowing sodium wasting, volume loss, and a normalization of blood pressure, but it also indirectly blocks excessive K^+ and H^+ wasting and corrects those plasma disorders. EnaC does *not* itself transport K^+ or H^+. Blocking the sodium channel in the CT prevents K^+ and H^+ secretion there via effects on the CT transepithelial membrane potential. For this reason, amiloride is a "potassium-sparing" diuretic, in contrast to all of the others to be discussed below.

8.7 Bartter and Gitelman Syndrome

Amiloride, with its ability to block EnaC channels and therefore resorption of only the final 2% of the filtered sodium load, is a relatively weak diuretic. The most potent diuretics, such as furosemide, target the apical Na-K-Cl cotransporter found in the TAL where 25% of the filtered sodium load is resorbed (Fig. 14b). The cotransporter is a solute carrier protein with 12 transmembrane helical spans encoded by *SLC12A1*. It has the ability to couple the electroneutral movement of four ions across the membrane, $1Na^+$, $1K^+$, and $2Cl^-$ (Fig. 6). The carrier's movement is passive in that no ATP energy is directly consumed in transport, but because the flows of all four ions are coupled and must take place in concert for any movement to occur, flux coupling, as was mentioned above, allows the cotransporter to capture the potential energy available in the sodium electrochemical gradient to move K^+ and Cl^- ions along with it as it enters the TAL cell across the apical membrane. This happens despite the fact that K^+ and Cl^- move uphill against their individual electrochemical gradients since it is the net electrochemical potential of the entire complex that dictates the direction of the coupled movements. This is a process called *secondary active transport*, a bioenergetic process coupled to an ion's electrochemical potential, and it is contrasted to *primary active transport*, a process coupled to the chemical potential of ATP hydrolysis and exemplified by the Na/K-ATPase, discussed above.

Loss-of-function alleles of *SLC12A1* were first recognized to cause the severe autosomal recessive disease phenotype of Bartter syndrome (BS), characterized by prenatal polyhydramnios, infantile polyuria, and polydipsia, renal Na^+, K^+, and H^+ wasting with subsequent hypokalemia and hypochloremic metabolic alkalosis, and very high renin, angiotensin, and aldosterone levels. It looked in every way as if these infants were on high doses of furosemide since conception, with the massive renal salt wasting producing the polyhydramnios, polyuria, and volume depletion that the homeostatic aldosterone signaling pathway was mightily struggling to restore. The high aldosterone was forcing a maximal EnaC channel response, but this low-capacity system could not handle the huge load of sodium the dysfunctional TAL had left behind, so sodium continued to be wasted. However, the electrical effect of maximal EnaC channel activation in the CT is present, hence the massive K^+ and H^+ wasting. Just as the EnaC channel has a major impact on the transepithelial electrical potential in the CT, and hence the flux of other ions in that renal segment, the SLC12A1 cotransporter has a major role in creating the TAL transepithelial potential, a very atypical potential in that it is lumen positive, not negative like nearly every other epithelial sheet (Fig. 14b). Physiologically, the lumen positive potential, created by the secondary active transport of Cl^- by the cotransporter, repels cations, particularly the divalent cations Ca^{++} and Mg^{++}, favoring their paracellular resorption and being the major site in the nephron at which they are resorbed. With the loss of SLC12A1 function, the divalent cations are wasted in the urine and the blood levels fall.

Since the SLC12A1 cotransporter carries an electroneutral load of two cations and two anions, it appears puzzling it should be able to create the unusual lumen positive condition. This arises because of two types of critical ion channels in the TAL apical and basolateral membranes (Fig. 14b). Mutations in these channels also produce the BS phenotype, BS type 1 used to specifically designate the disease caused by *SLC12A1* mutations. Since the TAL cells have a high internal K^+, the energetics of cotransport are favored by allowing K^+ to futilely recycle back into the lumen across the apical membrane. This apical potassium channel, the ATP-sensitive ROMK channel encoded by *KCNJ1*, a member of the channel family causing Andersen syndrome (Figs. 5 and 8), was shown to be essential to TAL transepithelial transport since loss-of-function mutations in this gene produced the Bartter syndrome phenotype, BS type 2. The high abundance of ROMK channels in the apical membrane of the TAL contributed half of the explanation for the unusual positive transepithelial potential; abundant K^+ channels gave the apical membrane a hyperpolarized membrane potential nearer the K^+ Nernst potential (Fig. 14b). The last part of the puzzle is the presence of abundant Cl^- channels in the basolateral membrane (Fig. 14b). They depolarize the basolateral membrane potential, and therefore accentuate the difference in potential between the two faces of the TAL cell. They are there to provide a pathway for Cl^- to leave the cell and passively reenter the blood, accompanying Na^+ and effecting net salt resorption in the TAL – the classical "diluting segment" of the tubule where hypotonic luminal fluid is produced. The basolateral Cl^- channel is composed of subunits from the large CLC channel family (Fig. 5) that includes the muscle chloride channel underlying Becker and Thomsen myotonia, discussed

above. A crystal structure has recently been obtained for this family of chloride channels, showing them to function as "double-barreled" dimers with each subunit, composed of 18 helical segments and both termini on the cytoplasmic face, having its own ion conduction pathway. The specific subunits found in the kidney are encoded by *CLCNKB* and *CLCNKA*, which are inserted in the membrane accompanied by barttin, their small β-subunit having two transmembrane helices and encoded by *BSND*. Recessive loss-of-function mutations in *CLCNKB* cause classical Bartter syndrome, BS type 3. Bartter syndrome, type 4 is caused by homozygosity for loss-of-function alleles of *BSND*, but this phenotype additionally includes sensorineural deafness. This same set of phenotypes is produced as a *digenic* syndrome with loss-of-function alleles of *both CLCNKB* and *CLCNKA*.

Channelopathies as a cause of deafness were briefly mentioned in the context of recessive cardiac LQT syndromes in Sect. 5.6. Surprisingly, the same channel is expressed selectively in the heart and the inner ear epithelium, where its abundant expression is involved in the production of the potassium-rich endolymph fluid that fills this chamber and participates in the transduction of auditory sensation. Only complete loss of the channel significantly disturbs this secretory process. It is of note that another inner ear *LQT1* homolog, encoded by *KCNQ4*, has dominant alleles that produce only deafness without LQT (*DFNA2*). Just as even a small amount of LQT channel function, such as provided by the dominant LQT alleles, was sufficient for production of the KCl-rich fluid of the inner ear compartment – deafness only occurring with a profound homozygous loss of function – only complete loss of inner ear epithelium chloride channel function causes deafness. The chloride that accompanies potassium in endolymph production passes through the *CLCNKB*- and *CLCNKA*-encoded channel. Without barttin, neither the *CLCNKB*- nor *CLCNKA*-encoded channel can be inserted into either the TAL or inner ear epithelial basolateral membrane. Either channel alone is sufficient to maintain the endolymph production for hearing, but only the *CLCNKB*-encoded channel can support the higher demands of TAL function.

Gitelman syndrome is the result of homozygosity for loss-of-function mutations in *SLC12A3*. It encodes a close relative of the Na-K-Cl cotransporter, called the *thiazide-sensitive Na-Cl cotransporter* to reflect its sensitivity to the thiazide class of diuretics and the K^+ insensitivity of the transport mechanism (Figs. 6 and 14c). This apical cotransporter is found in the DCT of the nephron and is responsible for the resorption of 7% of the filtered sodium load (Fig. 14c). Like Bartter syndrome, the disease presents with renal Na^+ wasting and a secondary wasting of K^+ and H^+ as well as resultant hypokalemia and hypochloremic metabolic alkalosis, as aldosterone and the EnaC channels struggle to correct the volume loss, which ultimately is only minor. Renin, angiotensin, and aldosterone are all elevated; however, the disease is much milder than Bartter syndrome, since the DCT resorbs a much smaller fraction of filtered sodium. Therefore, it classically presents in adolescence, but the presenting picture is *not* one of volume depletion, it is rather a picture predominated by consequences of the chronic plasma electrolyte abnormalities and their impact on excitable tissues. This is commonly manifested as muscle weakness, abdominal pain, and neuromuscular irritability. Patients commonly have hypomagnesemia and accompanying tetany,

although the pathogenesis of this abnormality remains to be elucidated.

It should be noted that one would expect hypomorphic alleles in the five genes contributing to Bartter and Gitelman syndrome and the three genes contributing to the EnaC channel to be protective against hypertension. Conversely, hypermorphic alleles in these same eight genes have the potential to contribute to hypertension. Therefore, polymorphic variants of these genes are promising candidates for evaluation in the heritable polygenic trait of blood pressure.

8.8 Hypomagnesemia

Autosomal recessive primary hypomagnesemia has played an important role in beginning to reveal the mechanisms underlying the paracellular pathways of epithelial ion transport, revealing them to be far more complex than simply an unmediated leak between cells (Fig. 14b). The disease presents in early childhood with symptoms similar to Gitelman syndrome, but additionally polyuria and renal stones. It is caused by homozygosity for loss-of-function alleles of *CLDN16*. This gene encodes Paracellin-1, a transmembrane protein with four helical spans found in the TAL and DCT tight junctions where it appears to contribute to a calcium and magnesium selective paracellular path. This gene family appears to encode proteins designated for such paracellular epithelial transport pathways since *CLDN4* and *CLDN15* also influence ion selectivity in other tight junctions.

Autosomal dominant hypomagnesemia appears to arise by a very different mechanism. It was shown in a large extended family to cosegregate with a heterozygous missense mutation in *FYXD2*, the gene encoding the small single membrane-spanning proteolipid γ-subunit of the Na/K-ATPase, a modulatory subunit found associated with the functional $\alpha\beta$ subunits only when the pump is expressed in basolateral membranes of the distal nephron. This subunit alters the transport kinetics of the ion pump, the prime mover setting the gradients and energetics for all coupled transport in these tight epithelia. The pathological *FYXD2 G41R* allele must function as a dominant negative gain-of-function allele, since haploinsufficiency produced by heterozygous deletions spanning the locus does not produce the phenotype. The mutation alters charge in the middle of the transmembrane helix and causes the γ-subunit to fail to traffic beyond the golgi, allowing functional $\alpha\beta$ subunit insertion into the membrane alone. As this is the only pathogenic allele reported to date, the mechanism by which mutations in this locus produce magnesium wasting remains unclear.

Hypomagnesemia with secondary hypocalcemia is caused by homozygosity for loss-of-function mutations in the *TRPM6* gene, over a dozen pathogenic alleles already known. This gene, and other members of its large extended family, are cousins of the genes encoding the pore-containing hexahelical channel subunits of voltage-gated K^+ channels, described in Sect. 3; however, the TRP branch of the family encode an array of selective and nonselective tetrameric cation channels that are coming to be recognized as playing important roles in sensory transduction and epithelial function (Fig. 5). The TRPM6 and TRPM7 channel subunits, each of which contains a C-terminal protein kinase domain, appear to heteromultimerize to form a Mg^{++} channel and pathogenic alleles of *TRPM6* prevent

subunit association and channel function. No pathogenic alleles, however, have yet been reported in *TRPM7*.

8.9 Polycystic Kidney Disease

Autosomal dominant polycystic kidney disease (ADPKD) is one of the most common genetic conditions, affecting over 1/800 births and responsible for 10% of chronic dialysis. Mutations – primarily null alleles caused by deletions, nonsense, frameshifts or splicesite alterations – in a relative of the TRP family of genes, PKD2 – encoding polycystin-2 – and in PKD1 – encoding polycystin-1, its homologous heteromultimerizing partner – cause this disease, and produce cystic dilation of kidney tubules, hepatic cysts, and aneurysms in blood vessels. Most of the nephrons appear normal, suggesting a "two-hit" pathogenic model reflecting a germline loss-of-function mutation and probabilistic somatic loss-of-heterozygosity (LOH), similar to that which has proved so powerful in understanding dominant cancers. Coexpression of PKD1 and PKD2 *in vitro* produce functional calcium-permeable nonselective cation channels, but expression of either PKD1 or PKD2 alone does not, and pathogenic alleles prevent heteromultimerization and channel function. The channels are found localized in association with the cilium, a structure of the apical membrane, perhaps transducing mechanical sensory signals it detects. While it is clear that calcium signaling is induced by the cilium and perturbed by the mutant alleles, it is unclear how either alters morphogenesis producing cyst formation. However, recently a clear example of morphogenetic pattern formation, left-right axis asymmetry, was shown to require epithelial PKD2 function.

8.10 Glucosuria

The notion that energy conserved in the sodium gradient across the apical membrane of an epithelial cell can be used in secondary active transport was introduced above in discussing Bartter and Gitelman syndrome. Sodium-coupled transport, however, is used in the gut and nephron in the resorption of many uncharged metabolites, not only ions (Fig. 6). A clear picture of how such energy coupling occurs is provided by two autosomal recessive glucose-wasting disorders. The sodium gradient across the apical membrane arises as a consequence of Na/K pump activity at the basolateral membrane. Secondary active transport of a substance, in this case glucose, occurs "uphill," against its electrochemical gradient by being obligatorily coupled on a carrier protein to the movement of sodium down its electrochemical gradient into the cell, such that the net movement of the coupled system obeys thermodynamics and remains "downhill" (Fig. 14a). A tremendous amount of energy is stored in the sodium gradient. A simple estimation reveals that cellular sodium is about 1/10th that found in the renal tubular fluid and that the membrane potential is ~ -60 mV – the Nernst equation revealing that translates to the equilibrium diffusion potential set up by a 10-fold gradient of a monovalent ion (*see* Sect. 2). Both the 10-fold chemical gradient and the equivalent 10-fold electrical gradient favor movement of Na^+ into the cell, and those potentials are multiplicative. This should allow 1 : 1 coupling of sodium:glucose to drive cellular glucose to a concentration 100 times

that found in tubular fluid. Should one obligatorily couple the movement of two sodium ions to cotransport with one glucose, the multiplicative properties suggest one could achieve the 100-fold enhancement for *each* Na^+ site, and hence a 10 000-fold gradient favoring glucose absorption. These are the properties of SLC5A2, the low affinity 1 : 1 apical sodium-glucose cotransporter and SLC5A1, the high affinity 2 : 1 apical sodium-glucose cotransporter, respectively.

Isolated renal glucosuria is caused by homozygous loss-of-function mutations in *SLC5A2*. The high-capacity, low-affinity cotransporter it encodes is glucose-selective and is abundantly expressed in the apical membrane of the initial segment of the proximal tubule of the kidney, the site where the majority of the glucose filtered across the glomerulus is reabsorbed (Fig. 14a). Physiologically, no glucose remains in the urine unless blood levels become so high that they overwhelm the transport capacity of this transporter, as occurs in diabetes mellitus. Therefore, the symptoms of this transport defect are massive glucosuria, and the resultant diuresis of water obligated by this large osmotic load.

Homozygous loss-of-function alleles of *SLC5A1* cause a severe neonatal dehydration syndrome; however, this is because of diarrhea, not the renal loss it produces. The disease caused by these mutations is glucose-galactose malabsorption, and it is characterized by loss of high affinity glucose transport in the kidney and gut. This high-affinity cotransporter is expressed in the apical membrane of the final segment of the proximal tubule where it scavenges the last of the glucose missed by the SLC5A2 cotransporter. Its low capacity, however, is easily overwhelmed, so it cannot compensate for loss of *SLC5A2*. Unlike SLC5A2, it transports galactose in addition to glucose, and is the mechanism that predominates intestinal absorption of these sugars – essential for a newborn dependent on the lactose of milk. Just as sugar left behind in the lumen of the nephron osmotically obligates water loss, so too does this unresorbed intestinal sugar; however, providing the infant a formula of fructose, a sugar that uses a different transporter, is life-saving.

8.11 Nephrogenic Diabetes Insipidus

Ancient physicians recognized that not only could diabetes, characterized by excessive urination and thirst – polyuria and polydipsia – be accompanied by sugary-sweet tasting urine – diabetes mellitus – but could alternatively be accompanied by tasteless urine – diabetes insipidus. Autosomal nephrogenic diabetes insipidus (NDI) results from defects in water channels, called *aquaporins*, and a recent Nobel Prize was awarded for the solution of their nature. Aquaporins reflect a large family of water channels and their function – allowing the flow of water while excluding that of ions, is the mirror image of that of ion channels, which allow the flow of ions without water. Loss-of-function recessive and dominant-negative alleles of *AQP2* result in NDI and in a loss of regulated water channels from the apical membrane of the principal cell of the CT (Fig. 14d), the site of regulated water resorption in the kidney. The peptide hormone alternatively called *vasopressin* or antidiuretic hormone (ADH), and long known to be responsible for the production of concentrated urine is tonically secreted by the pituitary in response to small changes in plasma osmolarity. It was shown to control insertion of the

AQP2 channel into the apical membrane of the principal cell; in the absence of hormone, the aquaporin was sequestered in vesicles beneath the membrane and hormone receptor signaling through cAMP triggered insertion of the channels. The receptor receiving the ADH signal, called the *V2 vasopressin receptor*, is encoded by the X-linked gene *AVPR2* and mutations in this gene were shown to cause the X-linked, more common form of nephrogenic diabetes insipidus. The phenotype in NDI, lethargy progressing to coma, results from effects on excitable tissues of euvolumic hypernatremic hyperosmolarity, as water is wasted by the kidney in excess of salt. Note that both major hormones involved in salt and water balance act independently by directing the insertion of different apical channels into the principal cell of the CT: aldosterone regulates sodium balance and hence plasma volume via the insertion of epithelial sodium channels and ADH regulates water balance and hence plasma osmolarity via the insertion of aquaporin-2.

8.12 Hereditary Kidney Stone

Kidney stone formation, nephrolithiasis, is a common polygenic trait with a lifetime risk of up to 10%. Stones are crystal-like structures that precipitate from urine within the kidney. Their origin is now recognized to be an abnormal urine environment, and recently defects in renal ion channels have been shown to contribute to this phenotype. Above it was briefly mentioned that stone formation was a component of the phenotype in primary hypomagnesemia. Nephrolithiasis is most commonly associated with hypercalciuria and calcium salts are a key component of most stones. Loss-of-function alleles of *CLCN5*, encoding a voltage-gated chloride channel within the family of those causing Bartter syndrome and myotonia (Fig. 5), described above, were recognized to cause a Mendelian X-linked form of hypercalciuria and stone formation, called *Dent disease*, a phenotype that additionally included proteinuria and progressed to renal failure. Additional alleles in the same gene were found to cause two other diseases that had not previously been recognized to be related, X-linked nephrolithiasis and X-linked hypophosphatemic rickets – a phenotype characterized by demineralization of bone. This chloride channel is localized to endosomes within epithelial cells of the proximal tubule, *demonstrating that ion channels play a critical part not only in the function of cellular membranes but also of subcellular compartments*. Defects in the function of the endosomal Cl^- channel impair the process of endocytosis from the apical membrane, secondarily altering the regulation of various apical transporters, but the exact mechanism of pathogenesis remains to be defined. It is of note that homozygous loss-of-function and heterozygous dominant-negative mutations in its close relative *CLCN7*, while causing a distinctly different phenotype, osteopetrosis, ultimately produce the phenotype as a result of a failure in endosomal function. Further, homozygous loss-of-function alleles in *TCIRG1*, encoding a subunit of the endosomal vacuolar ATPase proton pump (Fig. 6), also produce osteopetrosis. Osteopetrosis is a phenotype that includes very dense bone since bone remodeling by the degradative bone osteoclasts fails because endosomes fail to function since they can not acidify. Together, the two osteopetrosis diseases loci suggest that the endosomal Cl^- channels provide a shunt pathway that allows the endosomal H^+-ATPase to pump a proton gradient without stalling because

of electrostatic charge accumulation, and suggest a similar role for *CLCN5* in the kidney.

9
Ion Channel Disease of Other Nonexcitable Tissue

In addition to major phenotypes produced by defects directly assignable to abnormalities in the function of heart, muscle, nerve, and epithelia, ion channel defects are already known to produce a wide array of ancillary phenotypes, some of which were mentioned above as additional components to a syndrome, for instance, the skin findings produced by alleles producing deafness in the three gap junction loci, mentioned at the end of Sect. 7.7. Further, new channel phenotypes are coming to be recognized as arising from a failure of internal membranous organelle function, and these organelles play similar functions in most cell types. In these other tissues, channels seem to play pathophysiological roles similar to those played in diseases of excitable tissues or epithelia. For instance, the mirror-image pancreatic β-cell endocrine channelopathies of hyperinsulinism in infancy (HI) and permanent neonatal diabetes (PND) are very similar to neuronal receptor/channel disease, once the special mechanism of glucose sensing by the ATP-sensitive sulfonylurea receptor/channel is taken into account. This receptor/channel contains K^+ channel subunits, encoded by *KCNJ11* (from the family of those involved in Andersen and Bartter syndromes, discussed above) (Figs. 5 and 8), and a regulatory ATP-binding cassette subunit, encoded by *ABCC8,* named *SUR* for the sulfonylurea (SU) drugs that interact with it. Blocking this K^+ channel depolarizes the membrane, activating voltage-activated Ca^{++} channels and a calcium influx that triggers excitation-secretion coupling (as described in Sect. 7.1) with the fusion of insulin-laden vesicles into the β-cell membrane, and hence insulin secretion into the blood. Glucose induces insulin release through its metabolism in the β-cell, the ATP produced serving to block the channel; the SU drugs induce insulin release by directly blocking the channel's conductance. Not surprisingly, HI, which simulates the effect of SU, is caused by homozygosity for loss-of-function mutations in either of the two channel subunits. PND, where insulin can not be released, results from heterozygous dominant gain-of-function missense alleles of *KCNJ11,* which produce constitutively active channels that cannot respond to physiological ATP block induced by glucose, but can respond with insulin secretion when blocked by an SU. Notice how these alleles and the mirror-image phenotypes they produce are reminiscent of the mirror-image "endocrine" EnaC renal channelopathies, discussed in Sect. 8.6. As a second example, osteopetrosis, a defect in bone morphogenesis caused by defects in Cl^- channels and ultimately endosomal function, was briefly mentioned above as an endosomal channelopathy, an initial case of a disease caused by defects in membrane compartments *within* cells. Other membrane compartment channelopathies are beginning to emerge: endoplasmic reticulum channelopathies and MHS, caused by alleles of *RYR1,* discussed above in Sect. 6.4; lysosomal channelopathies and the lysosomal storage disease mucolipidosis type IV, caused by loss-of-function alleles in *MLN1,* encoding a lysosomal TRP-like nonselective cation channel; and mitochondrial channelopathies and severe obesity with diabetes, caused by

loss-of-function alleles of *UCP3*, encoding a mitochondrial proton channel that shuts protons across the membrane and thereby uncouples respiration from ATP synthesis, physiologically making mitochondria less efficient, and the disease appearing to result from hyper-efficient mitochondria.

10
Ion Channels as Targets for Drug Therapy in Common Diseases

For the discovery of a pathogenic allele to have full impact in modern medicine, it is necessary to understand its function and to find small molecules able to alter that function – such "rationally designed" novel pharmaceuticals being one of the major hopes for the postgenomic era. While often positionally identified disease genes have been slow to yield either their function or a "rational drug," ion channels are highly amenable and proven targets. For example, discovering ion channel participants in a disease, such as the Mendelian seizure disorder caused by alleles of *KCNQ2* and *KCNQ3*, suggests not only a novel *type* of target for therapy (e.g. suggesting use of acetazolamide – a drug proven effective in a wide range of ion channel diseases), but these channels themselves can further serve as the guides to *finding* entirely novel selective pharmaceuticals. Because the function of channel mechanisms is so well understood, and it is now clear that the *full* range of pathogenic abnormalities in the huge ensemble of channel diseases discussed above is either too much or too little conductance through the channel (not *anything* else), one does not need to guess what has to be changed to remedy the gene defect. It is clear that the therapeutic goal must be to discover a channel closer/blocker or a channel opener/activator, and a number of drugs of both types are already in current use. *In vitro* expression assays of channel function can be used in high-throughput drug screening and in drug optimization. An example of this kind of breakthrough is retigabine, a novel class of seizure medication proven to activate the KCNQ2/KCNQ3 "M current" channel suggested as a seizure drug target by the role of these genes in BFNC. Since the channel mechanisms perturbed in such disorders are likely to be critical participants in producing the relevant membrane potentials leading to disease vulnerability, the drugs targeting them should be useful for treating the disease even in those *not* having an intrinsic defect in the targeted mechanism (i.e. even if a patient has the disease because of a pathologically depolarized critical membrane because of some *other* mechanism, hyperpolarizing the implicated membrane via *any* mechanism should be helpful). Another advantage of pharmacology targeted to channels is the expectation, discussed above, that *a narrow range of tissues express the channel in a physiologically important context, therefore drug specificity is achieved and unwanted side effects are minimized*. Therefore, in this postgenomic world rich in potential therapeutic targets, the potential channels hold for rapid novel drug discovery warrants their prioritization.

Most of the diseases discussed in this chapter are rare channelopathies and despite being promising targets for pharmacological intervention would likely remain "orphan" diseases were it not for the fact that they produce a *phenocopy* of a common disease, and perhaps provide the only molecular lead to a new drug urgently needed by a major market. However, it is likely that many of the loci critical to the development of the

common complex polygenic diseases *cannot* be revealed by major-effect Mendelian alleles, such alleles perhaps being early embryonic lethals. An increasingly important approach to the identification of genes underlying complex polygenic disease is to identify "functional candidates" within broad chromosomal regions implicated in the disease and to test disease association (i.e. linkage disequilibrium) with polymorphic markers within these candidates. Ion channel candidates in such diseases could be suggested by the known physiology, pharmacology and pathophysiology, their appropriate or extremely narrow tissue distribution, their membership in a demonstrably pathogenic gene family, contribution to a demonstrably vulnerable component of the action potential, or a salient disease-associated motif. The existence of alleles sharing features common to known pathogenic alleles underlying the monogenic channelopathies, such as dominant-negative inhibition of expressed channel function or incomplete inactivation, would be particularly suggestive.

See also Receptor Biochemistry; Receptor Targets in Drug Discovery.

Bibliography

Books and Reviews

Agre, P., Kozono, D. (2003) Aquaporin water channels: molecular mechanisms for human diseases, *FEBS Lett.* **555**(1), 72–78.

Gargus, J.J. (2003) Unraveling monogenic channelopathies and their implications for complex polygenic disease, *Am. J. Hum. Genet.* **72**(4), 785–803.

Hille, B. (2001) *Ionic Channels of Excitable Membranes.* 2nd edition, Sinauer Associates, Sunderland, Massachusetts.

Jurkat-Rott, K., Lerche, H., Lehmann-Horn, F. (2002) Skeletal muscle channelopathies, *J. Neurol.* **249**, 1493–1502.

Lang, B., Vincent, A. (2003) Autoantibodies to ion channels at the neuromuscular junction, *Autoimmun. Rev.* **2**(2), 94–100.

Lerche, H., Jurkat-Rott, K., Lehmann-Horn, F. (2001) Ion channels and epilepsy, *Am. J. Med. Genet.* **106**, 146–159.

Lifton, R.P., Gharavi, A.G., Geller, D.S. (2001) Molecular mechanisms of human hypertension, *Cell* **104**, 545–556.

MacKinnon, R. (2003) Potassium channels, *FEBS Lett.* **555**(1), 62–65. '

Marban, E. (2002) Cardiac channelopathies, *Nature* **415**, 213–218.

Schultz, S.G. (1998) A century of (epithelial) transport physiology: from vitalism to molecular cloning, *Am. J. Physiol.* **274**(1 Pt 1), C13–C23.

Vander, A.J. (1995) *Renal Physiology*, 5th edition, McGraw-Hill, New York.

Primary Literature

Abbott, G.W., Butler, M.H., Bendahhou, S., Dalakas, M.C., Ptacek, L.J., Goldstein, S.A.N. (2001) MiRP2 forms potassium channels in skeletal muscle with Kv3.4 and is associated with periodic paralysis, *Cell* **104**, 217–231.

Abbott, G.W., Sesti, F., Splawski, I., Buck, M.E., Lehmann, M.H., Timothy, K.W., Keating, M.T., Goldstein, S.A. (1999) MiRP1 forms IKr potassium channels with HERG and is associated with cardiac arrhythmia, *Cell* **97**, 175–187.

Argyropoulos, G., Brown, A.M., Willi, S.M., Zhu, J., He, Y., Reitman, M., Gevao, S.M., Spruill, I., Garvey, W.T. (1998) Effects of mutations in the human uncoupling protein 3 gene on the respiratory quotient and fat oxidation in severe obesity and type 2 diabetes, *J. Clin. Invest.* **102**, 1345–1351.

Barclay, J., Balaguero, N., Mione, M., Ackerman, S.L., Letts, V.A., Brodbeck, J., Canti, C., Meir, A., Page, K.M., Kusumi, K., Perez-Reyes, E., Lander, E.S., Frankel, W.N., Gardiner, R.M., Dolphin, A.C., Rees, M. (2001) Ducky mouse phenotype of epilepsy and ataxia is associated with mutations in the Cacna2d2 gene and decreased calcium channel current in cerebellar Purkinje cells, *J. Neurosci.* **21**, 6095–6104.

Barhanin, J., Lesage, F., Guillemare, E., Fink, M., Lazdunski, M., Romey, G. (1996) K(v)LQT1 and IsK (minK) proteins associate to form the I(Ks) cardiac potassium current, *Nature* **384**, 78–80.

Baulac, S., Huberfeld, G., Gourfinkel-An, I., Mitropoulou, G., Beranger, A., Prud'homme, J.-F., Baulac, M., Brice, A., Bruzzone, R., LeGuern, E. (2001) First genetic evidence of GABA(A) receptor dysfunction in epilepsy: a mutation in the gamma-2-subunit gene, *Nat. Genet.* **28**, 46–48.

Bennett, P.B., Yazawa, K., Makita, N., George, A.L. Jr. (1995) Molecular mechanism for an inherited cardiac arrhythmia, *Nature* **376**, 683–685.

Bernard, C., Anderson, A., Becker, A., Poolos, N.P., Beck, H., Johnston, D. (2004) Acquired dendritic channelopathy in temporal lobe epilepsy, *Science* **305**(5683), 532–535.

Birkenhager, R., Otto, E., Schurmann, M.J., Vollmer, M., Ruf, E.-M., Maier-Lutz, I., Beekmann, F., Fekete, A., Omran, H., Feldmann, D., Milford, D.V., Jeck, N., Konrad, M., Landau, D., Knoers, N.V.A.M., Antignac, C., Sudbrak, R., Kispert, A., Hildebrandt, F. (2001) Mutation of BSND causes Bartter syndrome with sensorineural deafness and kidney failure, *Nat. Genet.* **29**, 310–314.

Browne, D.L., Gancher, S.T., Nutt, J.G., Brunt, E.R., Smith, E.A., Kramer, P., Litt, M. (1994) Episodic ataxia/myokymia syndrome is associated with point mutations in the human potassium channel gene, KCNA1, *Nat. Genet.* **8**, 136–140.

Bulman, D.E., Scoggan, K.A., van Oene, M.D., Nicolle, M.W., Hahn, A.F., Tollar, L.L., Ebers, G.C. (1999) A novel sodium channel mutation in a family with hypokalemic periodic paralysis, *Neurology* **53**, 1932–1936.

Burgess, D.L., Jones, J.M., Meisler, M.H., Noebels, J.L. (1997) Mutation of the Ca(2+) channel beta subunit gene Cchb4 is associated with ataxia and seizures in the lethargic (lh) mouse, *Cell* **88**, 385–392.

Caddick, S.J., Wang, C., Fletcher, C.F., Jenkins, N.A., Copeland, N.G., Hosford, D.A. (1999) Excitatory but not inhibitory synaptic transmission is reduced in lethargic (Cacnb4(lh)) and tottering (Cacna1atg) mouse thalami, *J. Neurophysiol.* **81**, 2066–2074.

Canessa, C.M., Horisberger, J.D., Rossier, B.C. (1993) Epithelial sodium channel related to proteins involved in neurodegeneration, *Nature* **361**(6411), 467–470.

Canessa, C.M., Schild, L., Buell, G., Thorens, B., Gautschi, I., Horisberger, J.D., Rossier, B.C. (1994) Amiloride-sensitive epithelial Na^+ channel is made of three homologous subunits, *Nature* **367**(6462), 463–467.

Cascio, M. (2004) Structure and function of the glycine receptor and related nicotinicoid receptors, *J. Biol. Chem.* **279**(19), 19383–19386.

Chandy, K.G., Fantino, E., Wittekindt, O., Kalman, K., Tong, L.L., Ho, T.H., Gutman, G.A., Crocq, M.A., Ganguli, R., Nimgaonkar, V., Morris-Rosendahl, D.J., Gargus, J.J. (1998) Isolation of a novel potassium channel gene hSKCa3 containing a polymorphic CAG repeat: a candidate for schizophrenia and bipolar disorder? *Mol. Psychiatry.* **3**, 32–37.

Chang, S.S., Grunder, S., Hanukoglu, A., Rosler, A., Mathew, P.M., Hanukoglu, I., Schild, L., Lu, Y., Shimkets, R.A., Nelson-Williams, C., Rossier, B.C., Lifton, R.P. (1996) Mutations in subunits of the epithelial sodium channel cause salt wasting with hyperkalaemic acidosis, pseudohypoaldosteronism type 1, *Nat. Genet.* **12**, 248–253.

Charlier, C., Singh, N.A., Ryan, S.G., Lewis, T.B., Reus, B.E., Leach, R.J., Leppert, M. (1998) **A** pore mutation in a novel KQT-like potassium channel gene in an idiopathic epilepsy family, *Nat. Genet.* **18**, 53–55.

Chen, Q., Kirsch, G.E., Zhang, D., Brugada, R., Brugada, J., Brugada, P., Potenza, D., Moya, A., Borggrefe, M., Breithardt, G., Ortiz-Lopez, R., Wang, Z., Antzelevitch, C., O'Brien, R.E., Schulze-Bahr, E., Keating, M.T., Towbin, J.A., Wang, Q. (1998) Genetic basis and molecular mechanism for idiopathic ventricular fibrillation, *Nature* **392**, 293–295.

Chubanov, V., Waldegger, S., Mederos y Schnitzler, M., Vitzthum, H., Sassen, M.C., Seyberth, H.W., Konrad, M., Gudermann, T. (2004) Free in PMC Disruption of TRPM6/TRPM7 complex formation by a mutation in the TRPM6 gene causes hypomagnesemia with secondary hypocalcemia, *Proc. Natl. Acad. Sci. U.S.A.* **101**(9), 2894–2899.

Claes, L., Del-Favero, J., Ceulemans, B., Lagae, L., Van Broeckhoven, C., De Jonghe, P. (2001) De novo mutations in the sodium-channel gene SCN1A cause severe myoclonic

epilepsy of infancy, *Am. J. Hum. Genet.* **68**, 1327–1332.

Cleiren, E., Benichou, O., Van Hul, E., Gram, J., Bollerslav, J., Singer, F.R., Beaverson, K., Aledo, A., Whyte, M.P., Yoneyama, T., deVernejou, M.-C., Van Hul, W. (2001) Albers-Schonberg disease (autosomal dominant osteopetrosis, type II) results from mutations in the ClCN7 chloride channel gene, *Hum. Mol. Genet.* **10**, 2861–2867.

Cummings, C.J., Zoghbi, H.Y. (2000) Trinucleotide repeats: mechanisms and pathophysiology, *Annu. Rev. Genomics Hum. Genet.* **1**, 281–328.

Curran, M.E., Splawski, I., Timothy, K.W., Vincent, G.M., Green, E.D., Keating, M.T. (1995) A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome, *Cell* **80**, 795–803.

Curran, P.F., McIntosh, J.R. (1962) A model system for biological water transport, *Nature* **193**, 347–348.

De Fusco, M., Marconi, R., Silvestri, L., Atorino, L., Rampoldi, L., Morgante, L., Ballabio, A., Aridon, P., Casari, G. (2003) Haploinsufficiency of ATP1A2 encoding the Na^+/K^+ pump alpha-2 subunit associated with familial hemiplegic migraine type 2, *Nat. Genet.* **33**, 192–196.

Deen, P.M.T., Verdijk, M.A.J., Knoers, N.V.A.M., Wieringa, B., Monnens, L.A.H., van Os, C.H., van Oost, B.A. (1994) Requirement of human renal water channel aquaporin-2 for vasopressin-dependent concentration of urine, *Science* **264**, 92–94.

Delmas, P. (2004) Polycystins: from mechanosensation to gene regulation, *Cell* **118**(2), 145–148.

Di sant'agnese, P.A., Darling, R.C., Perera, G.A., Shea, E. (1953) Abnormal electrolyte composition of sweat in cystic fibrosis of the pancreas; clinical significance and relationship to the disease, *Pediatrics* **12**(5), 549–563.

Doyle, D.A., Morais Cabral, J., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T., MacKinnon, R. (1998) The structure of the potassium channel: molecular basis of K^+ conduction and selectivity, *Science* **280**, 69–77.

Drachman, D.B. (1978) Myasthenia gravis (first of two parts), *N. Engl. J. Med.* **298**(3), 136–142.

Dror, V., Shamir, E., Ghanshani, S., Kimhi, R., Swartz, M., Barak, Y., Weizman, R., Avivi, L., Litmanovitch, T., Fantino, E., Kalman, K., Jones, E.G., Chandy, K.G., Gargus, J.J., Gutman, G.A., Navon, R. (1999) hKCa3/KCNN3 potassium channel gene: association of longer CAG repeats with schizophrenia in Israeli Ashkenazi Jews, expression in human tissues and localization to chromosome 1q21, *Mol. Psychiatry.* **4**, 254–260.

Ducros, A., Denier, C., Joutel, A., Cecillon, M., Lescoat, C., Vahedi, K., Darcel, F., Vicaut, E., Bousser, M.G., Tournier-Lasserve, E. (2001) The clinical spectrum of familial hemiplegic migraine associated with mutations in a neuronal calcium channel, *N. Engl. J. Med.* **345**, 17–24.

Duggal, P., Vesely, M.R., Wattanasirichaigoon, D., Villafane, J., Kaushik, V., Beggs, A.H. (1998) Mutation of the gene for IsK associated with both Jervell and Lange-Nielsen and Romano-Ward forms of long-QT syndrome, *Circulation* **97**, 142–146.

Dunne, M.J., Cosgrove, K.E., Shepherd, R.M., Aynsley-Green, A., Lindley, K.J. (2004) Hyperinsulinism in infancy: from basic science to clinical disease, *Physiol. Rev.* **84**(1), 239–275.

Dutzler, R., Campbell, E.B., Cadene, M., Chait, B.T., MacKinnon, R. (2002) X-ray structure of a ClC chloride channel at 3.0 a reveals the molecular basis of anion selectivity, *Nature* **415**(6869), 287–294.

Dutzler, R., Campbell, E.B., MacKinnon, R. (2003) Gating the selectivity filter in ClC chloride channels, *Science* **300**(5616), 108–112.

Engel, A.G., Ohno, K., Wang, H.-L., Milone, M., Sine, S.M. (1998) Molecular basis of congenital myasthenic syndromes: mutations in the acetylcholine receptor, *The Neuroscientist* **4**, 185–194.

Escayg, A., De Waard, M., Lee, D.D., Bichet, D., Wolf, P., Mayer, T., Johnston, J., Baloh, R., Sander, T., Meisler, M.H. (2000a) Coding and noncoding variation of the human calcium-channel beta(4)-subunit gene CACNB4 in patients with idiopathic generalized epilepsy and episodic ataxia, *Am. J. Hum. Genet.* **66**, 1531–1539.

Escayg, A., MacDonald, B.T., Meisler, M.H., Baulac, S., Huberfeld, G., An-Gourfinkel, I., Brice, A., LeGuern, E., Moulard, B., Chaigne, D., Buresi, C., Malafosse, A. (2000b) Mutations of SCN1A, encoding a neuronal sodium channel, in two families with GEFS + 2, *Nat. Genet.* **24**, 343–345.

Estevez, R., Boettger, T., Stein, V., Birkenhager, R., Otto, E., Hildebrandt, F., Jentsch,

T.J. (2001) Barttin is a Cl- channel beta-subunit crucial for renal Cl- reabsorption and inner ear K^+ secretion, *Nature* **414**, 558–561.

Eunson, L.H., Rea, R., Zuberi, S.M., Youroukos, S., Panayiotopoulos, C.P., Liguori, R., Avoni, P., McWilliam, R.C., Stephenson, J.B., Hanna, M.G., Kullmann, D.M., Spauschus, A. (2000) Clinical, genetic, and expression studies of mutations in the potassium channel gene KCNA1 reveal new phenotypic variability, *Ann. Neurol.* **48**, 647–656.

European Polycystic Kidney Disease Consortium. (1994) The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16, *Cell* **77**, 881–894.

Fambrough, D.M., Drachman, D.B., Satyamurti, S. (1973) Neuromuscular junction in myasthenia gravis: decreased acetylcholine receptors, *Science* **182**(109), 293–295.

Fletcher, C.F., Lutz, C.M., O'Sullivan, T.N., Shaughnessy, J.D. Jr., Hawkes, R., Frankel, W.N., Copeland, N.G., Jenkins, N.A. (1996) Absence epilepsy in tottering mutant mice is associated with calcium channel defects, *Cell* **87**, 607–617.

Forbush, B. III, Kaplan, J.H., Hoffman, J.F. (1978) Characterization of a new photoaffinity derivative of ouabain: labeling of the large polypeptide and of a proteolipid component of the Na, K-ATPase, *Biochemistry* **17**(17), 3667–3676.

Frattini, A., Orchard, P.J., Sobacchi, C., Giliani, S., Abinun, M., Mattsson, J.P., Keeling, D.J., Andersson, A.-K., Wallbrandt, P., Zecca, L., Notarangelo, L.D., Vezzoni, P., Villa, A. (2000) Defects in TCIRG1 subunit of the vacuolar proton pump are responsible for a subset of human autosomal recessive osteopetrosis, *Nat. Genet.* **25**, 343–346.

Gloyn, A.L., Pearson, E.R., Antcliff, J.F., Proks, P., Bruining, G.J., Slingerland, A.S., Howard, N., Srinivasan, S., Silva, J.M., Molnes, J., Edghill, E.L., Frayling, T.M., Temple, I.K., Mackay, D., Shield, J.P., Sumnik, Z., van Rhijn, A., Wales, J.K., Clark, P., Gorman, S., Aisenberg, J., Ellard, S., Njolstad, P.R., Ashcroft, F.M., Hattersley, A.T. (2004) Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes, *N. Engl. J. Med.* **350**(18), 1838–1849.

Gomez, C.M., Maselli, R.A., Vohra, B.P., Navedo, M., Stiles, J.R., Charnet, P., Schott, K., Rojas, L., Keesey, J., Verity, A., Wollmann, R.W., Lasalde-Dominicci, J. (2002) Novel delta subunit mutation in slow-channel syndrome causes severe weakness by novel mechanisms, *Ann. Neurol.* **51**(1), 102–112.

Goodenough, D.A., Paul, D.L. (2003) Beyond the gap: functions of unpaired connexon channels, *Nat. Rev. Mol. Cell Biol.* **4**(4), 285–294.

Grant, A.O., Carboni, M.P., Neplioueva, V., Starmer, C.F., Memmi, M., Napolitano, C., Priori, S. (2002) Long QT syndrome, Brugada syndrome, and conduction system disease are linked to a single sodium channel mutation, *J. Clin. Invest.* **110**, 1201–1209.

Grunder, s., Firsov, D., Chang, S.S., Jaeger, N.F., Gautschi, I., Schild, L., Lifton, R.P., Rossier, B.C. (1997) A mutation causing pseudohypoaldosteronism type 1 identifies a conserved glycine that is involved in the gating of the epithelial sodium channel, *EMBO J.* **16**, 899–907.

Hamill, O.P., Marty, A., Neher, E., Sakmann, B., Sigworth, F.J. (1981) Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches, *Pflugers Arch.* **391**(2), 85–100.

Hanaoka, K., Qian, F., Boletta, A., Bhunia, A.K., Piontek, K., Tsiokas, L., Sukhatme, V.P., Guggino, W.B., Germino, G.G. (2000) Coassembly of polycystin-1 and -2 produces unique cation-permeable currents, *Nature* **408**, 990–994.

Hansson, J.H., Schild, L., Lu, Y., Wilson, T.A., Gautschi, I., Shimkets, R., Nelson-Williams, C., Rossier, B.C., Lifton, R.P. (1995) A de novo missense mutation of the beta subunit of the epithelial sodium channel causes hypertension and Liddle syndrome, identifying a proline-rich segment critical for regulation of channel activity, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11495–11499.

Hansson, J.H., Nelson-Williams, C., Suzuki, H., Schild, L., Shimkets, R., Lu, Y., Canessa, C., Iwasaki, T., Rossier, B., Lifton, R.P. (1995) Hypertension caused by a truncated epithelial sodium channel gamma subunit: genetic heterogeneity of Liddle syndrome, *Nat. Genet.* **11**, 76–82.

Hevers, W., Luddens, H. (1998) The diversity of GABAA receptors. Pharmacological and electrophysiological properties of GABAA channel subtypes, *Mol. Neurobiol.* **18**, 35–86.

Hodgkin, A.L., Huxley, A.F. (1952) The components of membrane conductance in the giant axon of Loligo, *J. Physiol.* **116**(4), 473–496.

Ishikawa, K., Tanaka, H., Saito, M., Ohkoshi, N., Fujita, T., Yoshizawa, K., Ikeuchi, T., Watanabe, M., Hayashi, A., Takiyama, Y., Nishizawa, M., Nakano, I., Matsubayashi, K., Miwa, M., Shoji, S., Kanazawa, I., Tsuji, S., Mizusawa, H. (1997) Japanese families with autosomal dominant pure cerebellar ataxia map to chromosome 19p13.1-p13.2 and are strongly associated with mild CAG expansions in the spinocerebellar ataxia type 6 gene in chromosome 19p13.1, *Am. J. Hum. Genet.* **61**, 336–346.

Jentsch, T.J. (2000) Neuronal KCNQ potassium channels: physiology and role in disease, *Nat. Rev. Neurosci.* **1**, 21–30.

Jentsch, T.J. (2002) Chloride channels are different, *Nature* **415**(6869), 276–277.

Jentsch, T.J., Stein, V., Weinreich, F., Zdebik, A.A. (2002) Molecular structure and physiological function of chloride channels, *Physiol. Rev.* **82**, 503–568.

Jiang, Y., Ruta, V., Chen, J., Lee, A., MacKinnon, R. (2003) The principle of gating charge movement in a voltage-dependent K^+ channel, *Nature* **423**(6935), 42–48.

Jiang, Y., Lee, A., Chen, J., Cadene, M., Chait, B.T., MacKinnon, R. (2002) The open pore conformation of potassium channels, *Nature* **417**(6888), 523–526.

Jodice, C., Mantuano, E., Veneziano, L., Trettel, F., Sabbadini, G., Calandriello, L., Francia, A., Spadaro, M., Pierelli, F., Salvi, F., Ophoff, R.A., Frants, R.R., Frontali, M. (1997) Episodic ataxia type 2 (EA2) and spinocerebellar ataxia type 6 (SCA6) due to CAG repeat expansion in the CACNA1A gene on chromosome 19p, *Hum. Mol. Genet.* **6**, 1973–1978.

Jouvenceau, A., Eunson, L.H., Spauschus, A., Ramesh, V., Zuberi, S.M., Kullmann, D.M., Hanna, M.G. (2001) Human epilepsy associated with dysfunction of the brain P/Q-type calcium channel, *Lancet* **358**, 801–807.

Jun, K., Piedras-Renteria, E.S., Smith, S.M., Wheeler, D.B., Lee, S.B., Lee, T.G., Chin, H., Adams, M.E., Scheller, R.H., Tsien, R.W., Shin, H.S. (1999) Ablation of P/Q-type Ca(2+) channel currents, altered synaptic transmission, and progressive ataxia in mice lacking the alpha(1A)-subunit, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 15245–15250.

Kaunisto, M.A., Harno, H., Vanmolkot, K.R., Gargus, J.J., Sun, G., Hamalainen, E., Liukkonen, E., Kallela, M., van den Maagdenberg, A.M., Frants, R.R., Farkkila, M., Palotie, A., Wessman, M. (2004) A novel missense ATP1A2 mutation in a Finnish family with familial hemiplegic migraine type 2, *Neurogenetics* **5**(2), 141–146.

Keating, M., Atkinson, D., Dunn, C., Timothy, K., Vincent, G.M., Leppert, M. (1991) Linkage of a cardiac arrhythmia, the long QT syndrome, and the Harvey RAS-1 gene, *Science* **252**, 704–706.

Koefoed-Johnsen, V., Ussing, H.H. (1958) The nature of the frog skin potential, *Acta Physiol. Scand.* **42**(3–4), 298–308.

Kohl, S., Marx, T., Giddings, I., Jagle, H., Jacobson, S.G., Apfelstedt-Sylla, E., Zrenner, E., Sharpe, L.T., Wissinger, B. (1998) Total colourblindness is caused by mutations in the gene encoding the alpha-subunit of the cone photoreceptor cGMP-gated cation channel, *Nat. Genet.* **19**, 257–259.

Konrad, M., Weber, S. (2003) Recent advances in molecular genetics of hereditary magnesium-losing disorders, *J. Am. Soc. Nephrol.* **14**(1), 249–260.

Konrad, M., Schlingmann, K.P., Gudermann, T. (2004) Insights into the molecular nature of magnesium homeostasis, *Am. J. Physiol. Renal Physiol.* **286**(4), F599–F605.

Kornak, U., Kasper, D., Bosl, M.R., Kaiser, E., Schweizer, M., Schulz, A., Friedrich, W., Delling, G., Jentsch, T.J. (2001) Loss of the ClC-7 chloride channel leads to osteopetrosis in mice and man, *Cell* **104**, 205–215.

Kors, E.E., van den Maagdenberg, A.M., Plomp, J.J., Frants, R.R., Ferrari, M.D. (2002) Calcium channel mutations and migraine, *Curr. Opin. Neurol.* **15**, 311–316.

Kubisch, C., Schroeder, B.C., Friedrich, T., Lutjohann, B., El-Amraoui, A., Marlin, S., Petit, C., Jentsch, T.J. (1999) KCNQ4, a novel potassium channel expressed in sensory outer hair cells, is mutated in dominant deafness, *Cell* **96**, 437–446.

Langman, C.B. (2004) The molecular basis of kidney stones, *Curr. Opin. Pediatr.*. **16**(2), 188–193.

Leppert, M., Anderson, V.E., Quattlebaum, T., Stauffer, D., O'Connell, P., Nakamura, Y., Lalouel, J.M., White, R. (1989) Benign familial neonatal convulsions linked to genetic

markers on chromosome 20, *Nature* **337**, 647–648.

Lerche, C., Scherer, C.R., Seebohm, G., Derst, C., Wei, A.D., Busch, A.E., Steinmeyer, K. (2000) Molecular cloning and functional expression of KCNQ5, a potassium channel subunit that may contribute to neuronal M-current diversity, *J. Biol. Chem.* **275**, 22395–22400.

Letts, V.A., Felix, R., Biddlecome, G.H., Arikkath, J., Mahaffey, C.L., Valenzuela, A., Bartlett, F.S.I.I., Mori, Y., Campbell, K.P., Frankel, W.N. (1998) The mouse stargazer gene encodes a neuronal Ca(2+)-channel gamma subunit, *Nat. Genet.* **19**, 340–347.

Lewis, T.B., Leach, R.J., Ward, K., O'Connell, P., Ryan, S.G. (1993) Genetic heterogeneity in benign familial neonatal convulsions: identification of a new locus on chromosome 8q, *Am. J. Hum. Genet.* **53**, 670–675.

Lin, S.H., Cheng, N.L., Hsu, Y.J., Halperin, M.L. (2004) Intrafamilial phenotype variability in patients with Gitelman syndrome having the same mutations in their thiazide-sensitive sodium/chloride cotransporter, *Am. J. Kidney Dis.* **43**(2), 304–312.

Lipscombe, D., Pan, J.Q., Gray, A.C. (2002) Functional diversity in neuronal voltage-gated calcium channels by alternative splicing of Ca(v)alpha1, *Mol. Neurobiol.* **26**(1), 21–44.

Liu, B.A., Juurlink, D.N. (2004) Drugs and the QT interval – caveat doctor, *N. Engl. J. Med.* **351**(11), 1053–1056.

Lloyd, S.E., Pearce, S.H.S., Fisher, S.E., Steinmeyer, K., Schwappach, B., Schelnman, S.J., Harding, B., Bolino, A., Devoto, M., Goodyer, P., Rigden, S.P.A., Wrong, O., Jentsch, T.J., Craig, I.W., Thakker, R.V. (1996) A common molecular basis for three inherited kidney stone diseases, *Nature* **370**, 445–449.

Lossin, C., Wang, D.W., Rhodes, T.H., Vanoye, C.G., George, A.L. Jr. (2002) Molecular basis of an inherited epilepsy, *Neuron* **34**, 877–884.

Makita, N., Horie, M., Nakamura, T., Ai, T., Sasaki, K., Yokoi, H., Sakurai, M., Sakuma, I., Otani, H., Sawa, H., Kitabatake, A. (2002) Drug-induced long-QT syndrome associated with a subclinical SCN5A mutation, *Circulation* **106**, 1269–1274.

Marubio, L.M., del Mar Arroyo-Jimenez, M., Cordero-Erausquin, M., Lena, C., Le Novere, N., de Kerchove d'Exaerde, A., Huchet, M., Damaj, M.I., Changeux, J.-P. (1999) Reduced antinociception in mice lacking neuronal nicotinic receptor subunits, *Nature* **398**, 805–810.

McClatchey, A.I., McKenna-Yasek, D., Cros, D., Worthen, H.G., Kuncl, R.W., DeSilva, S.M., Cornblath, D.R., Gusella, J.F., Brown, R.H. Jr. (1992) Novel mutations in families with unusual and variable disorders of the skeletal muscle sodium channel, *Nat. Genet.* **2**, 148–152.

McGrath, J., Somlo, S., Makova, S., Tian, X., Brueckner, M. (2003) Two populations of node monocilia initiate left-right asymmetry in the mouse, *Cell* **114**(1), 61–73.

McRory, J.E., Hamid, J., Doering, C.J., Garcia, E., Parker, R., Hamming, K., Chen, L., Hildebrand, M., Beedle, A.M., Feldcamp, L., Zamponi, G.W., Snutch, T.P. (2004) The CACNA1F gene encodes an L-type calcium channel with unique biophysical properties and tissue distribution, *J. Neurosci.* **24**(7), 1707–1718.

Meij, I.C., Koenderink, J.B., van Bokhoven, H., Assink, K.F.H., Groenestege, W.T., de Pont, J.J.H.H.M., Bindels, R.J.M., Monnens, L.A.H., van den Heuvel, L.P.W.J., Knoers, N.V.A.M. (2000) Dominant isolated renal magnesium loss is caused by misrouting of the Na^+,K(+)-ATPase gamma-subunit, *Nat. Genet.* **26**, 265–266.

Meisler, M.H., Kearney, J., Ottman, R., Escayg, A. (2001) Identification of epilepsy genes in human and mouse, *Annu. Rev. Genet.* **35**, 567–588.

Meyer-Kleine, C., Steinmeyer, K., Ricker, K., Jentsch, T.J., Koch, M.C. (1995) Spectrum of mutations in the major human skeletal muscle chloride channel gene (CLCN1) leading to myotonia, *Am. J. Hum. Genet.* **57**, 1325–1334.

Miller, M.J., Rauer, H., Tomita, H., Rauer, H., Gargus, J.J., Gutman, G.A., Cahalan, M.D., Chandy, K.G. (2001) Nuclear localization and dominant-negative suppression by a mutant SKCa3 N-terminal channel fragment identified in a patient with schizophrenia, *J. Biol. Chem.* **276**, 27753–27756.

Mochizuki, T., Wu, G., Hayashi, T., Xenophontos, S.L., Veldhuisen, B., Saris, J.J., Reynolds, D.M., Cai, Y., Gabow, P.A., Pierides, A., Kimberling, W.J., Breuning, M.H., Constantinou Deltas, C., Peters, D.J.M., Somlo, S. (1996) PKD2, a gene for polycystic kidney disease

that encodes an integral membrane protein, *Science* **272**, 1339–1342.

Mohler, P.J., Splawski, I., Napolitano, C., Bottelli, G., Sharpe, L., Timothy, K., Priori, S.G., Keating, M.T., Bennett, V. (2004) A cardiac arrhythmia syndrome caused by loss of ankyrin-B function, *Proc. Natl. Acad. Sci. U.S.A.* **101**(24), 9137–9142.

Mohler, P.J., Schott, J.J., Gramolini, A.O., Dilly, K.W., Guatimosim, S., duBell, W.H., Song, L.S., Haurogne, K., Kyndt, F., Ali, M.E., Rogers, T.B., Lederer, W.J., Escande, D., Le Marec, H., Bennett, V. (2003) Ankyrin-B mutation causes type 4 long-QT cardiac arrhythmia and sudden cardiac death, *Nature* **421**(6923), 634–639.

Monnier, N., Procaccio, V., Stieglitz, P., Lunardi, J. (1997) Malignant-hyperthermia susceptibility is associated with a mutation of the alpha-1-subunit of the human dihydropyridine-sensitive L-type voltage-dependent calcium channel receptor in skeletal muscle, *Am. J. Hum. Genet.* **60**, 1316–1325.

Morais-Cabral, J.H., Zhou, Y., MacKinnon, R. (2001) Energetic optimization of ion conduction rate by the K^+ selectivity filter, *Nature* **414**, 37–42.

Morimoto, T., Nagao, H., Yoshimatsu, M., Yoshida, K., Matsuda, H. (1993) Pathogenic role of glutamate in hyperthermia-induced seizures, *Epilepsia* **34**, 447–452.

Moslehi, R., Langlois, S., Yam, I., Friedman, J.M. (1998) Linkage of malignant hyperthermia and hyperkalemic periodic paralysis to the adult skeletal muscle sodium channel (SCN4A) gene in a large pedigree, *Am. J. Med. Genet.* **76**, 21–27.

Mulders, S.M., Bichet, D.G., Rijss, J.P.L., Kamsteeg, E.-J., Arthus, M.-F., Lonergan, M., Fujiwara, M., Morgan, K., Leijendekker, R., van der Sluijs, P., van Os, C.H., Deen, P.M.T. (1998) An aquaporin-2 water channel mutant which causes autosomal dominant nephrogenic diabetes insipidus is retained in the Golgi complex, *J. Clin. Invest.* **102**, 57–66.

Nielsen, S., Chou, C.-L., Marples, D., Christensen, E.I., Kishore, B.K., Knepper, M.A. (1995) Vasopressin increases water permeability of kidney collecting duct by inducing translocation of aquaporin-CD water channels to plasma membrane, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1013–1017.

Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Kikyotani, S., Furutani, Y., Hirose, T., Takashima, H., Inayama, S., Miyata, T., Numa, S. (1983) Structural homology of Torpedo californica acetylcholine receptor subunits, *Nature* **302**, 528–532.

Ohmori, I., Ouchida, M., Ohtsuka, Y., Oka, E., Shimizu, K. (2002) Significant correlation of the SCN1A mutations and severe myoclonic epilepsy in infancy, *Biochem. Biophys. Res. Commun.* **295**, 17–23.

Ophoff, R.A., Terwindt, G.M., Vergouwe, M.N., van Eijk, R., Oefner, P.J., Hoffman, S.M.G., Lamerdin, J.E., Mohrenweiser, H.W., Bulman, D.E., Ferrari, M., Haan, J., Lindhout, D., van Ommen, G.-J.B., Hofker, M.H., Ferrari, M.D., Frants, R.R. (1996) Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca(2+) channel gene CACNL1A4, *Cell* **87**, 543–552.

Parkinson, N., Brown, S.D. (2002) Focusing on the genetics of hearing: you ain't heard nothin' yet, *Genome Biol.*. **3**(6), COMMENT2006.

Patsalos, P.N. (2000) Antiepileptic drug pharmacogenetics, *Ther. Drug Monit.* **22**(1), 127–130.

Phillips, H.A., Favre, I., Kirkpatrick, M., Zuberi, S.M., Goudie, D., Heron, S.E., Scheffer, I.E., Sutherland, G.R., Berkovic, S.F., Bertrand, D., Mulley, J.C. (2001) CHRNB2 is the second acetylcholine receptor subunit associated with autosomal dominant nocturnal frontal lobe epilepsy, *Am. J. Hum. Genet.* **68**, 225–231.

Piwon, N., Gunther, W., Schwake, M., Bosl, M.R., Jentsch, T.J. (2000) CIC-5 Cl(-)-channel disruption impairs endocytosis in a mouse model for Dent's disease, *Nature* **408**, 369–373.

Preisig-Muller, R., Schlichthorl, G., Goerge, T., Heinen, S., Bruggemann, A., Rajan, S., Derst, C., Veh, R.W., Daut, J. (2002) Heteromerization of Kir2.x potassium channels contributes to the phenotype of Andersen's syndrome, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7774–7779.

Priori, S.G., Napolitano, C., Schwartz, P.J. (1999) Low penetrance in the long-QT syndrome: clinical impact, *Circulation* **99**, 529–533.

Ptacek, L.J., George, A.L. Jr., Barchi, R.L., Griggs, R.C., Riggs, J.E., Robertson, M., Leppert, M.F. (1992) Mutations in an S4 segment of the adult skeletal muscle sodium channel cause paramyotonia congenita, *Neuron* **8**, 891–897.

Ptacek, L.J., George, A.L. Jr., Griggs, R.C., Tawil, R., Kallen, R.G., Barchi, R.L., Robertson,

M., Leppert, M.F. (1991) Identification of a mutation in the gene causing hyperkalemic periodic paralysis, *Cell* **67**, 1021–1027.

Ptacek, L.J., Tawil, R., Griggs, R.C., Engel, A.G., Layzer, R.B., Kwiecinski, H., McManis, P.G., Santiago, L., Moore, M., Fouad, G., Bradley, P., Leppert, M.F. (1994) Dihydropyridine receptor mutations cause hypokalemic periodic paralysis, *Cell* **77**, 863–868.

Pu, H.X., Scanzano, R., Blostein, R. (2000) Distinct regulatory effects of the Na,K-ATPase gamma subunit, *J. Biol. Chem.* **277**(23), 20270–20276.

Quane, K.A., Healy, J.M., Keating, K.E., Manning, B.M., Couch, F.J., Palmucci, L.M., Doriguzzi, C., Fagerlund, T.H., Berg, K., Ording, H., Bendixen, D., Mortier, W., Linz, U., Muller, C.R., McCarthy, T.V. (1993) Mutations in the ryanodine receptor gene in central core disease and malignant hyperthermia, *Nat. Genet.* **5**, 51–55.

Quinton, P.M., Bijman, J. (1983) Higher bioelectric potentials due to decreased chloride absorption in the sweat glands of patients with cystic fibrosis, *N. Engl. J. Med.* **308**(20), 1185–1189.

Ratjen, F., Doring, G. (2003) Cystic fibrosis, *Lancet* **361**(9358), 681–689.

Raychowdhury, M.K., Gonzalez-Perrett, S., Montalbetti, N., Timpanaro, G.A., Chasan, B., Goldmann, W.H., Stahl, S., Cooney, A., Goldin, E., Cantiello, H.F. (2004) Molecular pathophysiology of mucolipidosis type IV: pH dysregulation of the mucolipin-1 cation channel, *Hum. Mol. Genet.* **13**(6), 617–627.

Rees, M.I., Lewis, T.M., Kwok, J.B.J., Mortier, G.R., Govaert, P., Snell, R.G., Schofield, P.R., Owen, M.J. (2002) Hyperekplexia associated with compound heterozygote mutations in the beta-subunit of the human inhibitory glycine receptor (GLRB), *Hum. Mol. Genet.* **11**, 853–860.

Rho, J.M., Szot, P., Tempel, B.L., Schwartzkroin, P.A. (1999) Developmental seizure susceptibility of kv1.1 potassium channel knockout mice, *Dev. Neurosci.* **21**, 320–327.

Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.L., et al. (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA, *Science* **245**, 1066–1073.

Rosenthal, W., Seibold, A., Antaramian, A., Lonergan, M., Arthus, M.-F., Hendy, G.N., Birnbaumer, M., Bichet, D.G. (1992) Molecular identification of the gene responsible for congenital nephrogenic diabetes insipidus, *Nature* **359**, 233–235.

Roux, A.F., Pallares-Ruiz, N., Vielle, A., Faugere, V., Templin, C., Leprevost, D., Artieres, F., Lina, G., Molinari, N., Blanchet, P., Mondain, M., Claustres, M. (2004) Molecular epidemiology of DFNB1 deafness in France, *BMC Med. Genet.* **5**(1), 5.

Rowntree, R.K., Harris, A. (2003) The phenotypic consequences of CFTR mutations, *Ann. Hum. Genet.* **67**(Pt 5), 471–485.

Sanders, D.B. (2003) Lambert-eaton myasthenic syndrome: diagnosis and treatment, *Ann. N.Y. Acad. Sci.* **998**, 500–508.

Sanguinetti, M.C., Curran, M.E., Spector, P.S., Keating, M.T. (1996) Spectrum of HERG K^+-channel dysfunction in an inherited cardiac arrhythmia, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2208–2212.

Sansone, V., Griggs, R.C., Meola, G., Ptacek, L.J., Barohn, R., Iannaccone, S., Bryan, W., Baker, N., Janas, S.J., Scott, W., Ririe, D., Tawil, R. (1997) Andersen's syndrome: a distinct periodic paralysis, *Ann. Neurol.* **42**, 305–312.

Schlingmann, K.P., Konrad, M., Jeck, N., Waldegger, P., Reinalter, S.C., Holder, M., Seyberth, H.W., Waldegger, S. (2004) Salt wasting and deafness resulting from mutations in two chloride channels, *N. Engl. J. Med.* **350**(13), 1314–1319.

Schlingmann, K.P., Weber, S., Peters, M., Niemann Nejsum, L., Vitzthum, H., Klingel, K., Kratz, M., Haddad, E., Ristoff, E., Dinour, D., Syrrou, M., Nielsen, S., Sassen, M., Waldegger, S., Seyberth, H.W., Konrad, M. (2002) Hypomagnesemia with secondary hypocalcemia is caused by mutations in TRPM6, a new member of the TRPM gene family, *Nat. Genet.* **31**(2), 166–170.

Schultz, S.G., Curran, P.F. (1970) Coupled transport of sodium and organic solutes, *Physiol. Rev.* **50**(4), 637–718.

Schwartz, P.J., Priori, S.G., Dumaine, R., Napolitano, C., Antzelevitch, C., Stramba-Badiale, M., Richard, T.A., Berti, M.R., Bloise, R. (2000) A molecular link between the sudden infant death syndrome and the long-QT syndrome, *N. Engl. J. Med.* **343**, 262–267.

Schwartz, P.J., Priori, S.G., Bloise, R., Napolitano, C., Ronchetti, E., Piccinini, A., Goj, C.,

Breithardt, G., Schulze-Bahr, E., Wedekind, H., Nastoli, J. (2001) Molecular diagnosis in a child with sudden infant death syndrome, *Lancet* **358**, 1342–1343.

Schwartz, P.J., Stramba-Badiale, M., Segantini, A., Austoni, P., Bosi, G., Giorgetti, R., Grancini, F., Marni, E.D., Perticone, F., Rosti, D., Salice, P. (1998) Prolongation of the QT interval and the sudden infant death syndrome, *N. Engl. J. Med.* **338**, 1709–1714.

Seeburg, P.H., Single, F., Kuner, T., Higuchi, M., Sprengel, R. (2001) Genetic manipulation of key determinants of ion flow in glutamate receptor channels in the mouse, *Brain Res.* **907**(1–2), 233–243.

Segall, L., Scanzano, R., Kaunisto, M.A., Wessman, M., Palotie, A., Gargus, J.J., Blostein, R. (2004) Kinetic alterations due to a missense mutation in the Na,K-ATPase alpha 2 subunit cause familial hemiplegic migraine type 2, *J. Biol. Chem.* **279**(42), 43692–43696.

Sesti, F., Abbott, G.W., Wei, J., Murray, K.T., Saksena, S., Schwartz, P.J., Priori, S.G., Roden, D.M., George, A.L. Jr., Goldstein, S.A. (2000) A common polymorphism associated with antibiotic-induced cardiac arrhythmia, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10613–10618.

Shakkottai, V.G., Regaya, I., Wulff, H., Fajloun, Z., Tomita, H., Fathallah, M., Cahalan, M.D., Gargus, J.J., Sabatier, J.M., Chandy, K.G. (2001) Design and characterization of a highly selective peptide inhibitor of the small conductance calcium-activated K^+ channel, SkCa2, *J. Biol. Chem.* **276**(46), 43145–43151.

Shiang, R., Ryan, S.G., Zhu, Y.Z., Hahn, A.F., O'Connell, P., Wasmuth, J.J. (1993) Mutations in the alpha 1 subunit of the inhibitory glycine receptor cause the dominant neurologic disorder, hyperekplexia, *Nat. Genet.* **5**(4), 351–358.

Shieh, C.C., Coghlan, M., Sullivan, J.P., Gopalakrishnan, M. (2000) Potassium channels: molecular defects, diseases, and therapeutic opportunities, *Pharmacol. Rev.* **52**, 557–594.

Shimkets, R.A., Warnock, D.G., Bositis, C.M., Nelson-Williams, C., Hansson, J.H., Schambelan, M., Gill, J.R. Jr., Ulick, S., Milora, R.V., Findling, J.W., Canessa, C.M., Rossier, B.C., Lifton, R.P. (1994) Liddle's syndrome: heritable human hypertension caused by mutations in the beta subunit of the epithelial sodium channel, *Cell* **79**, 407–414.

Simon, D.B., Karet, F.E., Hamdan, J.M., Di Pietro, A., Sanjad, S.A., Lifton, R.P. (1996a) Bartter's syndrome, hypokalemic alkalosis with hypercalciuria, is caused by mutations in the Na-K-2Cl cotransporter NKCC2, *Nat. Genet.* **13**, 183–188.

Simon, D.B., Karet, F.E., Rodriguez-Soriano, J., Hamdan, J.H., DiPietro, A., Trachtman, H., Sanjad, S.A., Lifton, R.P. (1996b) Genetic heterogeneity of Bartter's syndrome revealed by mutations in the K^+ channel, ROMK, *Nat. Genet.* **14**, 152–156.

Simon, D.B., Nelson-Williams, C., Bia, M.J., Ellison, D., Karet, F.E., Molina, A.M., Vaara, I., Iwata, F., Cushner, H.M., Koolen, M., Gainza, F.J., Gitelman, H.J., Lifton, R.P. (1996c) Gitelman's variant of Bartter's syndrome, inherited hypokalaemic alkalosis, is caused by mutations in the thiazide-sensitive Na-Cl cotransporter, *Nat. Genet.* **12**, 24–30.

Simon, D.B., Lu, Y., Choate, K.A., Velazquez, H., Al-Sabban, E., Praga, M., Casari, G., Bettinelli, A., Colussi, G., Rodriguez-Soriano, J., McCredle, D., Milford, D., Sanjad, S., Lifton, R.P. (1999) Paracellin-1, a renal tight junction protein required for paracellular Mg(2+) resorption, *Science* **285**, 103–106.

Simon, D.B., Bindra, R.S., Mansfield, T.A., Nelson-Williams, C., Mendonca, E., Stone, R., Schurman, S., Nayir, A., Alpay, H., Bakkaloglu, A., Rodriguez-Soriano, J., Morales, J.M., Sanjad, S.A., Taylor, C.M., Pilz, D., Brem, A., Trachtman, H., Griswold, W., Richard, G.A., John, E., Lifton, R.P. (1997) Mutations in the chloride channel gene, CLCNKB, cause Bartter's syndrome type III, *Nat. Genet.* **17**, 171–178.

Singh, N.A., Charlier, C., Stauffer, D., DuPont, B.R., Leach, R.J., Melis, R., Ronen, G.M., Bjerre, I., Quattlebaum, T., Murphy, J.V., McHarg, M.L., Gagnon, D., Rosales, T.O., Peiffer, A., Anderson, V.E., Leppert, M. (1998) A novel potassium channel gene, KCNQ2, is mutated in an inherited epilepsy of newborns, *Nat. Genet.* **18**, 25–29.

Smart, S.L., Lopantsev, V., Zhang, C.L., Robbins, C.A., Wang, H., Chiu, S.Y., Schwartzkroin, P.A., Messing, A., Tempel, B.L. (1998) Deletion of the K(V)1.1 potassium channel causes epilepsy in mice, *Neuron* **20**, 809–819.

Splawski, I., Timothy, K.W., Vincent, G.M., Atkinson, D.L., Keating, M.T. (1997a) Molecular basis of the long-QT syndrome associated with deafness, *New. Eng. J. Med.* **336**, 1562–1567.

Splawski, I., Tristani-Firouzi, M., Lehmann, M.H., Sanguinetti, M.C., Keating, M.T. (1997b) Mutations in the hminK gene cause long QT syndrome and suppress IKs function, *Nat. Genet.* **17**, 338–340.

Steinlein, O.K., Mulley, J.C., Propping, P., Wallace, R.H., Phillips, H.A., Sutherland, G.R., Scheffer, I.E., Berkovic, S.F. (1995) A missense mutation in the neuronal nicotinic acetylcholine receptor alpha-4 subunit is associated with autosomal dominant nocturnal frontal lobe epilepsy, *Nat. Genet.* **11**, 201–203.

Stoffel, M., Jan, L.Y. (1998) Epilepsy genes: excitement traced to potassium channels, *Nat. Genet.* **18**, 6–8.

Strom, T.M., Nyakatura, G., Apfelstedt-Sylla, E., Hellebrand, H., Lorenz, B., Weber, B.H.F., Wutz, K., Gutwillinger, N., Ruther, K., Drescher, B., Sauer, C., Zrenner, E., Meitinger, T., Rosenthal, A., Meindl, A. (1998) An L-type calcium-channel gene mutated in incomplete X-linked congenital stationary night blindness, *Nat. Genet.* **19**, 260–263.

Sugawara, T., Mazaki-Miyazaki, E., Fukushima, K., Shimomura, J., Fujiwara, T., Hamano, S., Inoue, Y., Yamakawa, K. (2002) Frequent mutations of SCN1A in severe myoclonic epilepsy in infancy, *Neurology* **58**, 1122–1124.

Sugawara, T., Tsurubuchi, Y., Agarwala, K.L., Ito, M., Fukuma, G., Mazaki-Miyazaki, E., Nagafuji, H., Noda, M., Imoto, K., Wada, K., Mitsudome, A., Kaneko, S., Montal, M., Nagata, K., Hirose, S., Yamakawa, K. (2001) A missense mutation of the Na^+ channel alpha II subunit gene Na(v)1.2 in a patient with febrile and afebrile seizures causes channel dysfunction, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6384–6389.

Sundin, O.H., Yang, J.M., Li, Y., Zhu, D., Hurd, J.N., Mitchell, T.N., Silva, E.D., Maumenee, I.H. (2000) Genetic basis of total colourblindness among the Pingelapese islanders, *Nat. Genet.* **25**, 289–293.

Tanabe, T., Beam, K.G., Powell, J.A., Numa, S. (1988) Restoration of excitation-contraction coupling and slow calcium current in dysgenic muscle by dihydropyridine receptor complementary DNA, *Nature* **336**, 134–139.

Tempel, B.L., Papazian, D.M., Schwarz, T.L., Jan, Y.N., Jan, L.Y. (1987) Sequence of a probable potassium channel component encoded at Shaker locus of Drosophila, *Science* **237**, 770–775.

Thomas, P., Ye, Y., Lightner, E. (1996) Mutation of the pancreatic islet inward rectifier Kir6.2 also leads to familial persistent hyperinsulinemic hypoglycemia of infancy, *Hum. Mol. Genet.* **5**, 1809–1812.

Thomas, P.M., Cote, G.J., Wohllk, N., Haddad, B., Mathew, P.M., Rabl, W., Aguilar-Bryan, L., Gagel, R.F., Bryan, J. (1995) Mutations in the sulfonylurea receptor gene in familial persistent hyperinsulinemic hypoglycemia of infancy, *Science* **268**, 426–429.

Tomita, H., Shakkittai, V.G., Gutman, G.A., Sun, G., Bunney, W.E., Cahalan, M.D., Chandy, K.G., Gargus, J.J. (2003) Novel truncated isoform of SK3 potassium channel is a potent dominant-negative regulator of SK currents: implications in schizophrenia, *Mol. Psychiatry.* **8**(5), 524–535,

Towbin, J.A., Vatta, M. (2001) Molecular biology and the prolonged QT syndromes, *Am. J. Med.* **110**, 385–398.

Tristani-Firouzi, M., Jensen, J.L., Donaldson, M.R., Sansone, V., Meola, G., Hahn, A., Bendahhou, S., Kwiecinski, H., Fidzianska, A., Plaster, N., Fu, Y.H., Ptacek, L.J., Tawil, R. (2002) Functional and clinical characterization of KCNJ2 mutations associated with LQT7 (Andersen syndrome), *J. Clin. Invest.* **110**, 381–388.

Turk, E., Zabel, B., Mundlos, S., Dyer, J., Wright, E.M. (1991) Glucose/galactose malabsorption caused by a defect in the Na(+)/glucose cotransporter, *Nature* **350**, 354–356.

van den Heuvel, L.P., Assink, K., Willemsen, M., Monnens, L. (2002) Autosomal recessive renal glucosuria attributable to a mutation in the sodium glucose cotransporter (SGLT2), *Hum. Genet.* **111**, 544–547.

Van Den Maagdenberg, A.M., Kors, E.E., Brunt, E.R., Van Paesschen, W., Pascual, J., Ravine, D., Keeling, S., Vanmolkot, K.R., Vermeulen, F.L., Terwindt, G.M., Haan, J., Frants, R.R., Ferrari, M.D. (2002) Episodic ataxia type 2 Three novel truncating mutations and one novel missense mutation in the CACNA1A gene, *J. Neurol.* **249**, 1515–1519.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Walder, R.Y., Landau, D., Meyer, P., Shalev, H., Tsolia, M., Borochowitz, Z., Boettger, M.B., Beck, G.E., Englehardt, R.K., Carmi, R., Sheffield, V.C. (2002) Mutation of TRPM6 causes familial hypomagnesemia with secondary hypocalcemia, *Nat. Genet.* **31**(2), 171–174.

Wallace, R.H., Marini, C., Petrou, S., Harkin, L.A., Bowser, D.N., Panchal, R.G., Williams, D.A., Sutherland, G.R., Mulley, J.C., Scheffer, I.E., Berkovic, S.F. (2001a) Mutant GABA(A) receptor gamma2-subunit in childhood absence epilepsy and febrile seizures, *Nat. Genet.* **28**, 49–52.

Wallace, R.H., Wang, D.W., Singh, R., Scheffer, I.E., George, A.L. Jr., Phillips, H.A., Saar, K., Reis, A., Johnson, E.W., Sutherland, G.R., Berkovic, S.F., Mulley, J.C. (1998) Febrile seizures and generalized epilepsy associated with a mutation in the Na(+)-channel beta-1 subunit gene SCN1B, *Nat. Genet.* **19**, 366–370.

Wallace, R.H., Scheffer, I.E., Barnett, S., Richards, M., Dibbens, L., Desai, R.R., Lerman-Sagie, T., Lev, D., Mazarib, A., Brand, N., Ben-Zeev, B., Goikhman, I., Singh, R., Kremmidiotis, G., Gardner, A., Sutherland, G.R., George, A.L. Jr., Mulley, J.C., Berkovic, S.F. (2001b) Neuronal sodium-channel alpha-1-subunit mutations in generalized epilepsy with febrile seizures plus, *Am. J. Hum. Genet.* **68**, 859–865.

Wang, H.-S., Pan, Z., Shi, W., Brown, B.S., Wymore, R.S., Cohen, I.S., Dixon, J.E., McKinnon, D. (1998) KCNQ2 and KCNQ3 potassium channel subunits: molecular correlates of the M-channel, *Science* **282**, 1890–1893.

Wang, Q., Shen, J., Splawski, I., Atkinson, D., Li, Z., Robinson, J.L., Moss, A.J., Towbin, J.A., Keating, M.T. (1995) SCN5A mutations associated with an inherited cardiac arrhythmia, long QT syndrome, *Cell* **80**, 805–811.

Wang, Q., Curren, M.E., Splawski, I., Burn, T.C., Millholland, J.M., VanRaay, T.J., Shen, J., Timothy, K.W., Vincent, G.M., de Jager, T., Schwartz, P.J., Towbin, J.A., Moss, A.J., Atkinson, D.L., Landes, G.M., Connors, T.D., Keating, M.T. (1996) Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias, *Nat. Genet.* **12**, 17–23.

Warmke, J.W., Ganetzky, B. (1994) A family of potassium channel genes related to eag in Drosophila and mammals, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3438–3442.

Watanabe, H., Nagata, E., Kosakai, A., Nakamura, M., Yokoyama, M., Tanaka, K., Sasai, H. (2000) Disruption of the epilepsy KCNQ2 gene results in neural hyperexcitability, *J. Neurochem.* **75**(1), 28–33.

Watnick, T.J., Torres, V.E., Gandolph, M.A., Qian, F., Onuchic, L.F., Klinger, K.W., Landes, G., Germino, G.G. (1998) Somatic mutation in individual liver cysts supports a two-hit model of cystogenesis in autosomal dominant polycystic kidney disease, *Mol. Cell* **2**, 247–251.

Wilson, P.D. (2004) Polycystic kidney disease, *N. Engl. J. Med.* **350**(2), 151–164.

Wickenden, A.D., Yu, W., Zou, A., Jegla, T., Wagoner, P.K. (2000) Retigabine, a novel anticonvulsant, enhances activation of KCNQ2/Q3 potassium channels, *Mol. Pharmacol.* **58**, 591–600.

Wollnik, B., Schroeder, B.C., Kubisch, C., Esperer, H.D., Wieacker, P., Jentsch, T.J. (1997) Pathophysiological mechanisms of dominant and recessive KVLQT1 K^+ channel mutations found in inherited cardiac arrhythmias, *Hum. Mol. Genet.* **6**, 1943–1949.

Wright, E.M., Turk, E. (2004) The sodium/glucose cotransport family SLC5, *Pflugers Arch.* **447**(5), 510–518.

Yang, P., Kanki, H., Drolet, B., Yang, T., Wei, J., Viswanathan, P.C., Hohnloser, S.H., Shimizu, W., Schwartz, P.J., Stanton, M., Murray, K.T., Norris, K., George, A.L. Jr., Roden, D.M. (2002) Allelic variants in long-QT disease genes in patients with drug-associated torsades de pointes, *Circulation* **105**, 1943–1948.

Yue, Q., Jen, J.C., Nelson, S.F., Baloh, R.W. (1997) Progressive ataxia due to a missense mutation in a calcium-channel gene, *Am. J. Hum. Genet.* **61**, 1078–1087.

Zhuchenko, O., Bailey, J., Bonnen, P., Ashizawa, T., Stockton, D.W., Amos, C., Dobyns, W.B., Subramony, S.H., Zoghbi, H.Y., Lee, C.C. (1997) Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha(1A)-voltage-dependent calcium channel, *Nat. Genet.* **15**, 62–69.

Zuberi, S.M., Eunson, L.H., Spauschus, A., De Silva, R., Tolmie, J., Wood, N.W., McWilliam, R.C., Stephenson, J.P., Kullmann, D.M., Hanna, M.G. (1999) A novel mutation in the human voltage-gated potassium channel gene (Kv1.1) associates with episodic ataxia type 1 and sometimes with partial epilepsy, *Brain* **122**, 817–825.

Zwingman, T.A., Neumann, P.E., Noebels, J.L., Herrup, K. (2001) Rocker is a new variant of the voltage-dependent calcium channel gene Cacna1a, *J. Neurosci.* **21**, 1169–1178.

Receptors, Vitamin: *see* Vitamin Receptors

Recombination and Genome Rearrangements

Hannah L. Klein
New York University Medical Center, New York, NY

1 Homologous Recombination Events 3
1.1 Homologous Recombination (HR) Mechanisms 3
1.1.1 Double-strand Break Repair 4
1.1.2 Synthesis-dependent Strand Annealing 5
1.1.3 Break-induced Replication 7
1.1.4 Single-strand Annealing 8
1.2 Nonhomologous End Joining 9

2 Steps in Homologous Recombination 11
2.1 Initiation 11
2.2 Propagation 11
2.3 Resolution 12

3 Regulation of Recombination 12
3.1 Meiosis versus Mitosis 13
3.2 Sister Chromatid Recombination versus Homolog Recombination 15
3.3 Regulation by Cell Cycle DNA Damage Checkpoint Factors 16
3.4 Alternative Repair Pathways 17

4 Consequences of Defects in Homologous Recombination 17
4.1 Rearrangements Associated with Mutations in HR Genes 17
4.2 Meiotic Defects Associated with Mutations in HR Genes 18

5 Concluding Remarks 18

Acknowledgment 19

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

Bibliography 19
Books and Reviews 19
Primary Literature 19

Keywords

Crossing-over
A recombination that involves physical exchange of DNA between nonsister strands of homologous chromosomes, resulting in a nonparental combination of linked markers.

Gene Conversion
Nonreciprocal recombination between homologous chromosomes with one allele being the donor of information to convert the recipient allele to the donor sequence.

Heteroduplex
DNA hybrid formed from single strands of two nonsister chromatids; an intermediate in recombination.

Holliday Junction
Crossed-strand structure consisting of two DNA duplexes joined by a bridge as an intermediate in recombination.

Mismatch Repair
Repair of mismatched bases in heteroduplex DNA by a mismatch repair system, resulting in homoduplex DNA.

Nonhomologous End Joining
Joining of blunt ends or ends with microhomologies of usually 2 to 4 bases.

Resolution
Cutting of the Holliday junction by a resolvase to disconnect the recombining DNA duplexes and restoring the parental configuration of markers or resulting in a nonparental configuration of markers termed crossovers.

Strand Exchange
One of the processes to form heteroduplex DNA during recombination in which a strand from a homoduplex is displaced and a strand from a nonsister duplex is transferred to form a hybrid region on the chromatid.

Transformation
Addition of exogenous DNA to a cell to change the genotype of the recipient cell.

Homologous recombination (HR) is an essential activity of all cells as it is one of two major pathways used to repair potentially lethal double-strand breaks (DSBs). DSBs arise from many sources. They may be induced by exogenous agents such as ionizing radiation or chemicals that modify the DNA bases of the DNA backbone, such as methylmethane sulfonate, bleomycin, or mitomycin C. Agents that inhibit DNA topoisomerases, such as cisplatin, also cause DSBs. Endogenous DSBs may occur from programmed breaks, such as those involved in immune system V(D)J recombination of vertebrates, faulty topoisomerases, or oxygen free radicals. When the DNA replication apparatus encounters bulky lesions or a nicked template strand, DSBs may form, collapsing the replication fork. Alternatively, the replication fork may stall with single-strand DNA at the fork. This is a substrate for HR, or further processing to form a DSB. These DSBs are repaired by HR or nonhomologous DSB rejoining events. Therefore, although mitotic HR occurs at significantly reduced rates compared to meiotic HR, it is nonetheless an essential aspect of vegetative growth.

HR is essential for maintaining integrity of the genome, for preventing chromosome rearrangements, for preventing changes in chromosome number, for telomere maintenance when the normal telomere replication pathway is defective, and for preventing the occurrence of alternate double-strand break repair pathways that may result in loss of heterozygosity (LOH) events. Since spontaneous DSBs occur in every replication cycle in vertebrate cells, HR is an essential function of a normal replication cycle. Meiotic HR may also be linked to repairing premeiotic DNA replication errors, but its main purpose lies in ensuring genetic variation and proper chromosome segregation. Meiotic HR gives genetic variation to future generations by providing new combinations of alleles in the gametes. The most important function of meiotic HR is to provide a mechanism for proper segregation of homologous chromosomes at the first meiotic division. In the absence of crossing-over, chromosomes do not disjoin at the first meiotic division and give rise to aneuploid meiotic products. In this chapter, the current models for double-strand break repair by HR are reviewed, differences between mitotic and meiotic HR are discussed, and the consequences of defective HR are considered.

1 Homologous Recombination Events

1.1 Homologous Recombination (HR) Mechanisms

Homologous recombination (HR) uses information from a homologous DNA sequence, usually the sister chromatid or the homolog chromosome, to accurately repair a double-strand break (DSB), and in the process HR can transfer information from the donating DNA duplex to the recipient DNA duplex. Traditionally, HR has been thought of as reciprocal exchange of nonsister chromatids. Crossing-over, often referred to as *reciprocal recombination*, involves the physical joining of two parental molecules of homologous chromosomes. Genetically this is detected as recombinant progeny. If one parent has the linked markers AB and the other

parent has the alternate alleles, ab, the crossover products have the genotypes Ab and aB (Fig. 1a). Gene conversions, often referred to as *nonreciprocal recombination*, involve the transfer of information from one parental molecule to another parental molecule. The genotype of the donor molecule remains unchanged while the genotype of the recipient is converted to the donor genotype (Fig. 1b).

Nonreciprocal recombination or gene conversion is best studied in the fungi where all four products of a single meiosis can be recovered. Gene conversion does not result from mutation. Genetic and molecular studies have shown that the gene conversion product has the same DNA sequence as the donor sequence. Occasionally, a meiotic product is recovered that has genetic information from both parents. This mosaic product is called a *postmeiotic segregation event*. In wild-type strains, gene conversion events occur at a much higher frequency than postmeiotic segregation events. Postmeiotic segregation events result from the failure to repair a DNA mismatch in heteroduplex DNA. Mutants that are defective for mismatch repair show higher PMS frequencies and reduced gene conversion frequencies, indicating that heteroduplex formation is an intermediate in gene conversion.

1.1.1 Double-strand Break Repair

The first clues that HR is initiated by DSBs came from studies showing that

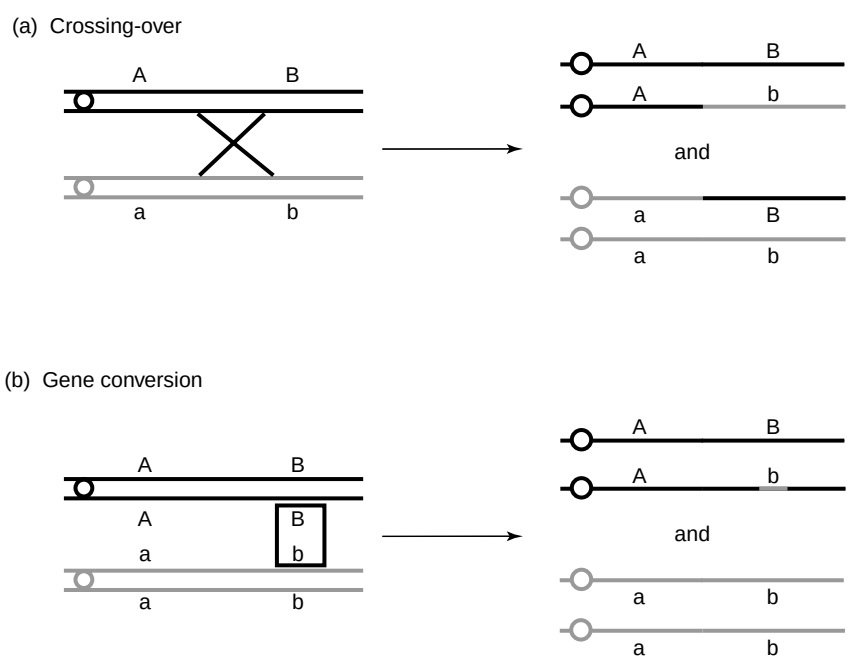

Fig. 1 Types of homologous recombination events. (a) Crossing-over, indicated by the X, generates two recombinant chromosomes and two nonrecombinant chromosomes. (b) Gene conversion, indicated by the box, yields only one recombinant chromosome. This chromosome has only a small region of substitution, at the B gene, from the donor chromosome. The donor chromosome remains unchanged, and is of the parental genotype.

HR was stimulated by ionizing radiation. This idea was further strengthened by experiments of transformation in the yeast *Saccharomyces cerevisiae*. When intact plasmid molecules that carry a yeast gene are applied to yeast cells, the cells take up the plasmid and express the yeast gene or marker located on the plasmid after it is integrated into the yeast genome. The site of integration is not random; instead, it occurs by homologous crossing-over. Thus, transformation in yeast was viewed as a model for HR. Although transformation using intact plasmid molecules is relatively efficient, the frequency could be greatly increased if the plasmid molecules were first treated with a restriction enzyme that cuts within the yeast gene located on the plasmid.

When yeast cells are treated with linear plasmid molecules, transformation still occurs by homologous crossing-over. If the plasmid molecule contains two unlinked yeast genes A and B, the plasmid molecule will integrate into either the A gene or the B gene with approximately equal efficiency in transformation experiments. However, if the plasmid is first treated with a restriction enzyme that cuts only in the A gene, then most transformation events will be integrations at the A gene, demonstrating that DSBs are highly recombinogenic.

When the linear transforming plasmid contained a gap in a yeast gene, it was repaired using the chromosomal information. This is a gene conversion event and shows that DSBs are recombinogenic for both gene conversion and crossing-over events. Transformation using gapped molecules has also been performed using molecules that contain a sequence that acts as an origin of DNA replication. As these molecules can replicate autonomously, they are not obliged to integrate, although the gap must be repaired. Transformation with such molecules was associated with repair (gene conversion) and homologous crossing-over (integration) approximately 50% of the time and with simple repair (gene conversion without crossing-over) for the remainder 50% of the time. This result gave further support for the double-strand break repair model as a general model for meiotic recombination.

The general features of the double-strand break repair (DSBR) model for HR are shown in Fig. 2. After DSB formation, the ends of the break are resected to give 3′ single-strand tails of up to 600 nucleotides in length. The single-strand tails invade a homologous DNA sequence, and then primer synthesis from the 3′ end until synthesis fills in the gap on both strands. Ligation to the 5′ ends of the break creates a joint molecule with two Holliday Junctions (HJs). The joint molecule contains heteroduplex DNA, which is a target for the mismatch repair system. Resolution of the HJs results in crossover or non-crossover molecules. Gap repair gives rise to the gene conversion events.

1.1.2 Synthesis-dependent Strand Annealing

Although the DSBR model nicely explained why DSBs promoted HR, how a DSB could be repaired to give a gene conversion, and how crossing-over could be mechanistically linked to gene conversion via formation and resolution of the HJs, the model did not satisfactorily explain some aspects of mitotic HR. First, in other experimental systems, gap repair of a plasmid molecule was not often associated with a crossing-over event. Second, in yeast, mitotic gene conversion was not frequently associated with crossing-over. Third, mating type switching in yeast, which is DSB-promoted via the HO endonuclease,

was rarely associated with crossing-over in both haploid and diploid cells.

These observations promoted a revision of the DSBR model to form a new model that retained the features of DSB repair and gene conversion through gap filling, but did not result in intermediates with HJs. Hence, the gene conversion events were not mechanistically linked to crossing-over. The key feature of this model, called

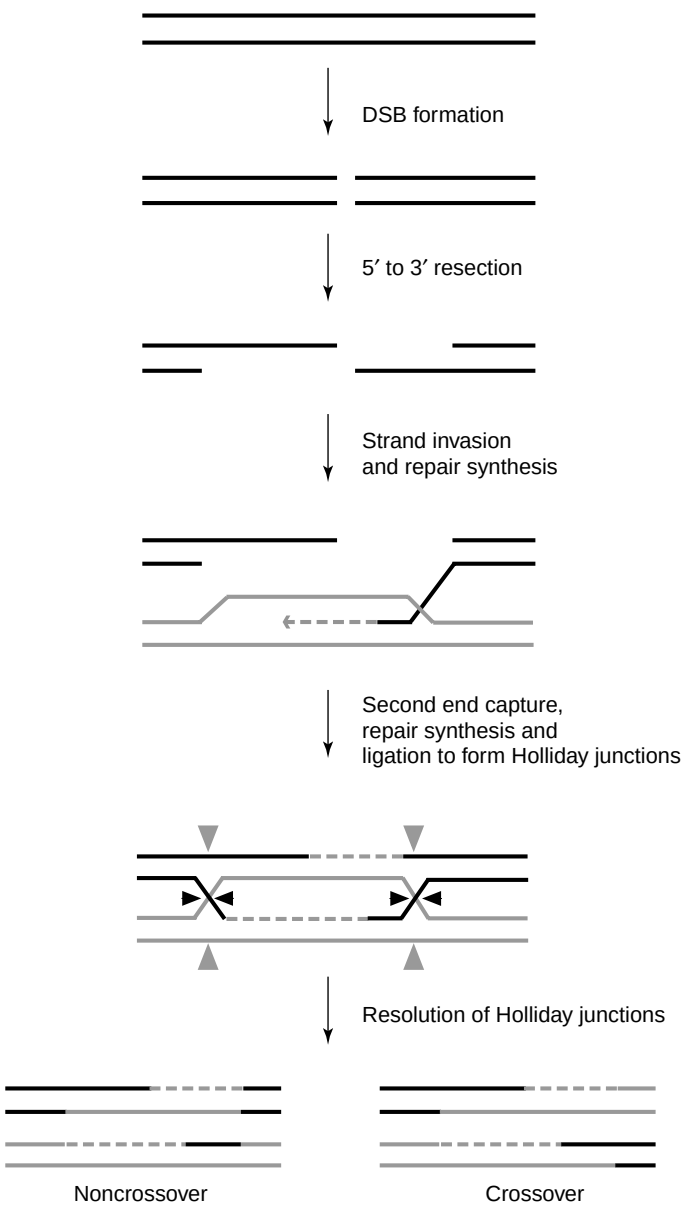

the *synthesis-dependent strand annealing* or *SDSA* model, is a migrating D-loop that does not capture the second DSB end (Fig. 3). As in the DSBR model, the DSB is resected to give 3′ tails, one of which invades into homologous sequences on paired DNA molecule. The 3′ end primes DNA synthesis, and as synthesis proceeds, the D-loop migrates. However, the second 3′ end never is paired with the D-loop. Instead, the invading 3′ end eventually is displaced from the template duplex, and pairs with the single-strand region from the other end of the DSB. Repair synthesis fills in the gaps, resulting in DSB repair to give a gene conversion not associated with crossing-over.

SDSA has also been proposed to operate in meiosis to form all of the noncrossover gene conversion events. This will be discussed further in Sect. 3.1.

1.1.3 Break-induced Replication

The DSBR model is a two-ended strand invasion, where both 3′ ends of the DSB are involved in strand invasion into a homologous DNA duplex. In contrast, the SDSA model is a one-ended invasion, where only one of the two 3′ ends is involved in a strand invasion, and the other end participates in the annealing reaction. Both of these repair models are initiated from a DSB that is flanked by sequence on both sides. However, there are also DSBs that come from essentially terminal deletions, where only one side of the DBS can pair with homologous sequence (Fig. 4). Repair of this DSB by HR of necessity uses a one-ended invasion pathway. Repair is proposed to occur by processing of the DSB to give a 3′ end and invasion into homologous sequence. Repair is completed by synthesis to the end of the chromosome. This type of repair is called *break-induced replication* or *BIR*.

Evidence for BIR first came from studies in bacteria, where replication of the *Escherichia coli* chromosome was dependent on both recombination factors and replication factors. BIR differs from DSBR and SDSA in that very long tracts of DNA are replicated in a semiconservative manner. BIR was first observed in yeast from transformation with broken chromosomes. The repaired DNA fragments were linear, and contained a centromere. One chromosome arm ended in a normal telomere sequence, while the other chromosome arm was truncated internally without a telomere. This truncated arm was repaired by strand invasion into an intact homologous chromosome followed by synthesis to the end of the chromosome. This regenerated the deleted telomere and chromosome arm. The acquired sequences were identical to those from the intact chromosome arm used as template for semiconservative replication. Thus, BIR results in

Fig. 2 Double-strand break repair (DSBR) model of homologous recombination. After induction of a double-strand break (DSB), both ends are resected to yield 3′ ends. The 3′ end invades a homologous sequence and primes repair DNA synthesis using the 3′ end and the invaded strand as template. As synthesis proceeds, the second 3′ end of the DSB can be captured by either strand invasion or annealing to the displaced homologous DNA sequence. Ligation of the ends results in two-crossed strand Holliday Junctions. Resolution of the Holliday Junctions can yield noncrossover products (resolution of both Holliday Junctions by cutting at the black arrow heads) or crossover products (resolution of one Holliday Junction by cutting at the black arrow head and the second Holliday Junction by cutting at the gray arrow head).

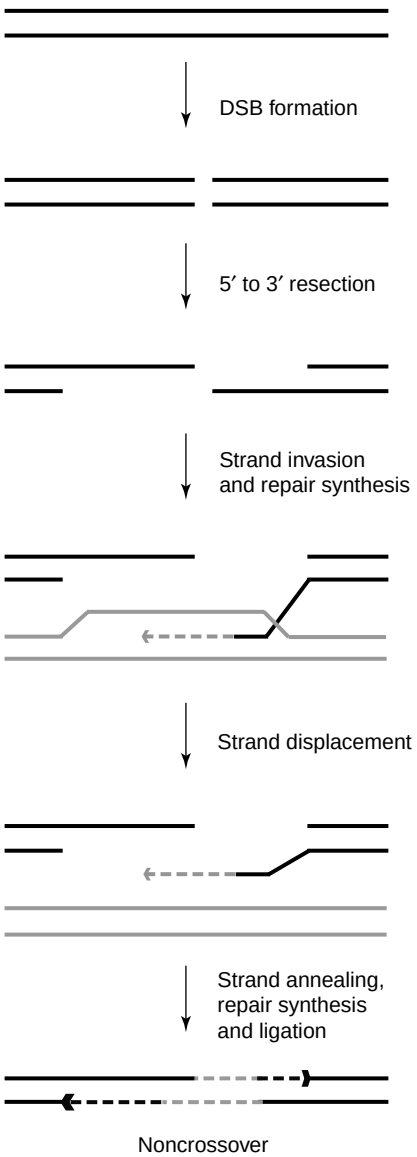

Fig. 3 Synthesis-dependent strand annealing (SDSA) model of homologous recombination. After induction of a double-strand break (DSB), both ends are resected to yield 3′ ends. The 3′ end invades a homologous sequence and primes repair DNA synthesis using the 3′ end and the invaded strand as template. After synthesis, the invading strand becomes displaced and anneals with the other 3′ single-strand tail of the processed DSB. This process differs from the DSBR in that the second end never is captured into the donor homologous duplex. The gaps in the annealed strand structure are filled in by repair synthesis, using the original strand as template. The reaction is completed by ligation of the nicks.

LOH events, and genetically are identical to products from mitotic crossing-over. The BIR process in yeast was initially called *break copy duplication*, based on the results from transformation with chromosome fragments. BIR may function in cells to maintain telomeres when the telomere replicating enzyme telomerase is absent. BIR can also result in nonreciprocal translocations if the broken end invades into homologous repeated sequences located on a different chromosome.

1.1.4 **Single-strand Annealing**

The last DSB repair process that uses homology to be discussed is single-strand annealing (SSA). SSA is thought to occur between directly repeated sequences (Fig. 5), but could also occur between sequences located on different chromosomes. In the SSA model, the ends of the DSB are resected to give 3′ tails, but these are not engaged in strand invasion of homologous sequences. Rather, the 3′ ends anneal to each other at the complementary single-stranded regions that have been revealed by the resection. The 3′ tails emanating from the annealed duplex are removed by endonucleases, and the nicks are then sealed by DNA ligase. SSA events are considered error-prone as they result in deletions between direct repeats, or potentially translocations and other rearrangements if sequences

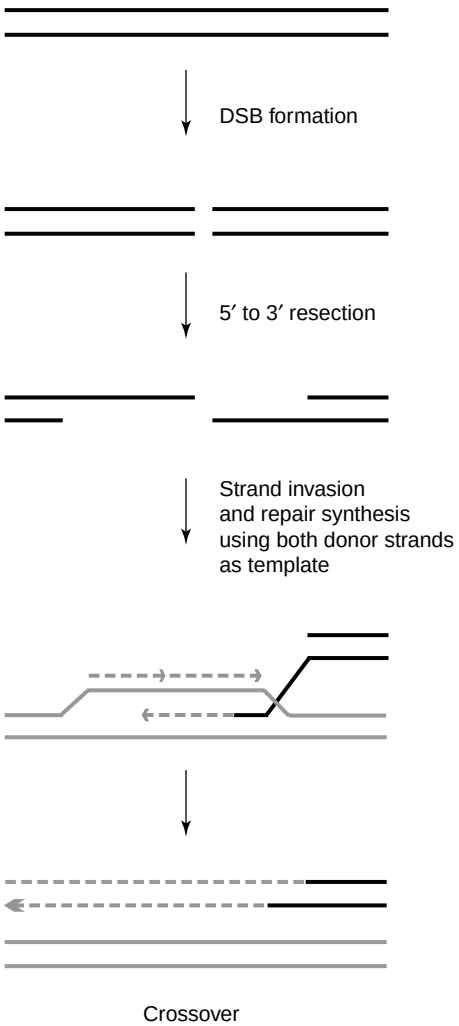

Fig. 4 Break-induced replication (BIR) model of homologous recombination. After induction of a double-strand break (DSB), both ends are resected to yield 3′ ends. The 3′ end invades a homologous sequence and primes repair DNA synthesis using the 3′ end and the invaded strand as template. The second end of the DSB is never engaged in the DSB repair reaction. DNA synthesis is established in the opposite direction, using the displaced D-loop as template. Synthesis continues to the end of the donor chromosome, copying both strands to repair the lost DNA sequence. The resultant product looks like half of a crossover reaction (see Fig. 1).

on different chromosomes anneal. Since SSA cannot occur without some type of chromosome rearrangement, it most likely is rare and usually involves direct repeats.

1.2 Nonhomologous End Joining

In addition to DSB repair using homology, cells also possess pathways to repair DSBs by nonhomologous methods. This type of repair is called *nonhomologous end joining* (*NHEJ*). NHEJ may use microhomologies of 1 to 2 nucleotides at the DSB ends, or there may be no homology at all. Ends are brought together by proteins specific for NHEJ, called *Ku70 and Ku80*. Additional factors essential for NHEJ include the protein kinase DNA-PKcs. Binding of these proteins to the DSB ends helps to bring the ends together. The ends are then ligated by a special ligase called *ligase IV* and its partner *XRCC4*. NHEJ is often error-prone and small deletions or additions occur at the ligated ends.

Although NHEJ is frequently error-prone, it is the major DSB end joining process used in mitotic cells in the G1 phase of the cell cycle. In G1, the only available homologous partner is the homolog chromosome, and pairing by a search for homology may be rate limiting when chromosomes are not naturally paired. Since only about 5% of the mammalian genome sequence encodes ORFs, error-prone end joining often may not be deleterious. During S phase and G2, the sister chromatid is available for repair

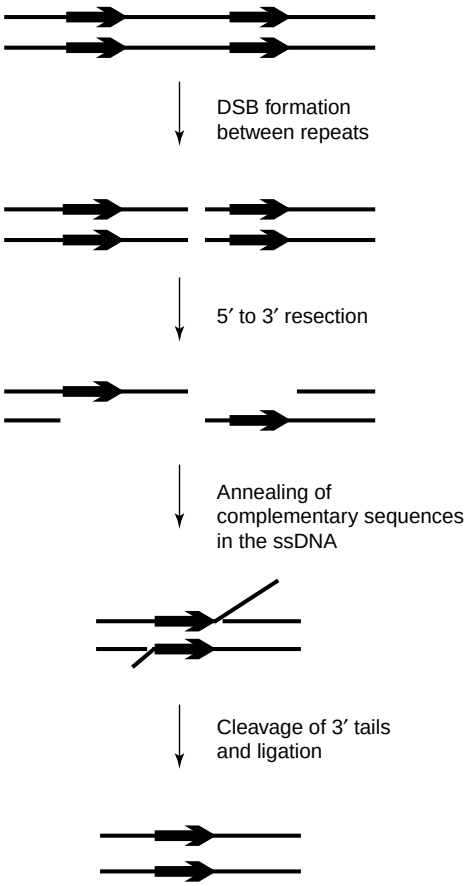

Fig. 5 Single-strand annealing (SSA) model of DSB repair. After induction of a double-strand break (DSB) between direct repeats, both ends are resected past the repeats to yield 3′ ends that have the repeat sequence in single-strand configuration. The complementary strand anneal at the repeat sequences, generating 3′ single-strand tails. The tails are cleaved by the structure specific Rad1/10 endonuclease (XPF/ERCC1 in humans), and the nicks are sealed by ligation. The SSA reaction results in a deletion of one copy of the repeat and the DNA sequence located between the repeats.

through HR and is held in close proximity through the cohesion complexes, which hold the sister chromatids together until mitosis.

NHEJ appears to be the preferred methods for repair of DSBs in vertebrate somatic cells. However, DSBs generated by sequence-specific endonucleases are often repaired by HR. Most HR proteins are essential and knockout mice for these proteins die as early embryos. This shows that HR is essential for repair of spontaneous damage during embryogenesis. The damage most likely occurs during replication, resulting in stalled or collapsed replication forks. Vertebrate cells that have nonlethal mutations in HR genes, such as Rad54, some of the Rad51 paralogs, or BRCA1, are sensitive to irradiation, showing that HR is also used to repair damage resulting from irradiation. NHEJ appears to be avoided in meiotic cells by downregulating the Ku proteins.

In yeast, NHEJ is used in the G1 phase of the cell cycle to repair DSBs while HR is used during late S and G2. NHEJ is also regulated through sequestration of the yeast XRCC4-like protein Lif1 to the cytoplasm in diploid cells that are able to undergo meiosis.

2
Steps in Homologous Recombination

In this section, the molecular mechanisms of the DSBR and SDSA models will be discussed, as these have been the most extensively studied of the HR mechanisms and are the most common DSB repair processes by HR. The proteins discussed are those used in eukaryotic systems. Most of these proteins are conserved from yeast to human. SSA usually occurs between direct repeats and appears to involve nucleolytic processing of the DSB, annealing of complementary strand promoted by the annealing activity of the Rad52 protein and RPA protein, processing of any nonannealed DNA tails by the ERCC1/XPF endonuclease, and sealing of the nicks by DNA ligase.

2.1
Initiation

Once a DSB is formed, to prepare the ends for HR, they must be resected to reveal 3′ single-strand tails. The resection most likely occurs through the action of the MRN complex and additional nuclease and DNA helicase activities. Once single-strand tails with 3′ ends are formed, these become coated with RPA protein to protect the single-stranded DNA and to remove any secondary structure. Next, RPA is displaced and Rad51 protein coats the single-strand DNA to form a DNA-protein filament. Rad51 loading is facilitated by a set of proteins related to Rad51 called the *Rad51 paralogs*, plus Rad52, and Rad54 proteins. Additional proteins related to Rad52 and Rad54 may also be involved. There are also a number of proteins that interact with Rad51 and may aid in the initiation stage of loading Rad51 or in the search for homology with a DNA partner. One important interactor is the Brca2 protein. Brca2 is mutated in some heritable human breast cancers and is clearly involved in HR. The DNA helicases WRN and BLM also are required in HR. Some models place these early in the initiation step, while other models propose that these helicases act during the later stage of resolution. The helicases could also act in the extension step, to open up the D-loop for extension of the invading 3′ end by replication.

The second step in initiation involves invasion of the intact donor DNA duplex by the 3′ Rad51 filament tail of the recipient DNA at a point of DNA homology, signaling the end of the search for homology and the beginning of strand invasion. Strand invasion is promoted by Rad54 protein and may also require Rad52 protein at the 3′ ends of the DSB. During strand invasion, the Rad51 protein is displaced from the 3′ single-strand region as this sequence enters into duplex-paired DNA. In the DSBR model, the displaced D-loop sequence is enlarged by extension of the 3′ end of the invading strand by replication. This requires leading strand DNA polymerases. Extension continues until the 3′ end completes the gap repair and can be ligated to the recessed 5′ end of the second DSB end. The second DSB end can also engage in a strand invasion of the D-loop, or be captured by extension of the D-loop until the displaced single strand of the D-loop can anneal with the other 3′ end of the DSB. In the SDSA model, the displaced D-loop sequence from the strand invasion is small and migrates, propagated by DNA synthesis.

2.2
Propagation

Propagation can be considered in two parts, extension of the 3′ end by replication,

and branch migration of the entire joint molecule. DNA synthesis from the invading 3′ strand appears to require the replicative DNA polymerases δ and ε (POL2 and POL3). It is not clear whether any of the error-prone translesion synthesis polymerases are also involved. HR is considered to be error-free as the DNA synthesis in replication uses the replicative DNA polymerases. However, some recent studies in yeast suggest that DSB repair by HR is more error-prone than replicative DNA synthesis. As described above, in the DSBR model, synthesis extends from the 3′ ends of the DSB until the gap is repaired and the ends can be ligated to the resected 5′ ends of the DSB, forming two HJs. In the SDSA model, synthesis proceeds only from the invading 3′ end, and continues until it is displaced from the template donor duplex. The newly replicated portion of the DSB is now able to base pair with the 3′ end from the other tail of the DSB, thus providing a template for synthesis from this 3′ end. When synthesis reaches the resected 5′ ends of the DSB, ligation can occur.

There is genetic evidence that a cross-strand DNA molecule with HJs can branch migrate. *In vitro*, the *E. coli* RecG DNA helicase and the *E. coli* RuvB helicase in association with RuvA can branch migrate a HJ. Both activities are required for normal HR in *E. coli*. Similar proteins have not been found in eukaryotes, although genetic evidence suggests that symmetric heteroduplex tracts can form during HR, and their formation is best explained by branch migration of HJs.

2.3 Resolution

Resolution in HR refers to resolution of the HJs. In the SDSA model, no HJs are formed and the repair products are always noncrossover. The DSBR model proposes the formation of double HJs. The search for proteins with endonuclease activity specific for HJ structures, producing products that can be ligated without errors has been ongoing for many years. In *E. coli*, the proteins that fulfill these criteria are the RuvC and RusA proteins. *In vitro*, RuvC cleaves HJs and *in vivo* it is required for HR. In eukaryotes, an endonucleolytic activity that is specific for HJs has been found to act on mitochondrial DNA. Loss of this activity has no effect on mitotic HR in chromosomal DNA or on meiotic HR. Recent studies point to a role for some of the Rad51 paralogs in HJ resolution activity. Alternatively, resolution of HJs could occur through the combined action of DNA topoisomerases and DNA helicases.

However resolution occurs, it must be able to resolve the HJs in a manner that allows the ends to be ligated together after cutting and the interlinked DNA strand to be untwined. Resolution may give crossover or noncrossover products, depending on which strands are cut, as shown in Fig. 2.

3 Regulation of Recombination

HR and NHEJ are the major DSB repair pathways. NHEJ is preferred in mammalian cells for DSB repair while HR is preferred in yeast. The consequences of deletion of NHEJ-specific factors in yeast are minimal, but in mammalian cells, loss of NHEJ results in increased HR. NHEJ and HR repair DSBs by different mechanisms, but have compensatory roles in maintenance of genomic integrity. Vertebrate cells defective in both NHEJ

and HR have high rates of genome instability and cannot be maintained as stable cell lines. There is greatly increased chromosomal aberrations and cell death. These observations show that spontaneous DSBs occur frequently enough to be lethal if unrepaired and that NHEJ and HR have overlapping roles in genome stability maintenance. Nonetheless, the use of NHEJ or HR to repair DSBs is regulated on several levels, including cell cycle regulation. NHEJ acts primarily in G1-early S for repair of both induced and spontaneous damage, while HR acts in late S-G2. This compartmentalization of the mode DSB repair helps to ensure that most HR DSB repair involves sister chromatid HR instead of interhomolog HR.

3.1 Meiosis versus Mitosis

Meiotic HR differs from mitotic HR in several aspects. These include the nature of the HR initiation events, the frequency of HR, meiotic-specific proteins involved in HR, the presence of the synaptonemal complex, the requirement for crossover HR events, preferential use of the homolog for HR, and the phenomenon of interference, which limits the number of crossovers per chromosome arm in meiotic HR.

While mitotic HR is essential for replication restart after fork stalling or collapse, meiotic HR is essential for proper chromosome segregation in meiosis I, through the occurrence of crossovers between homologous chromosomes. Thus, the ability of most organisms to produce euploid and viable meiotic products is dependent on HR. Most mitotic HR is initiated from spontaneous or exogenously induced DSBs. There are a few instances of programmed DSBs, including mating type switching in the fungi and V(D)J recombination in the immune system of vertebrates. Repair of these breaks occurs though a special program for each DSB system. In meiosis, HR occurs at induced DSBs, induced by action of the type II topoisomerase-like protein called *Spo11*. Spo11 DSBs occur at preferred chromosomal regions, which often have transcription initiation sites, but there is no specific DNA sequence that is recognized by any of the Spo11 proteins. Spo11-induced DSBs occur frequently enough such that the meiotic HR rate is at least 100-fold higher than the mitotic HR rate, and often is 1000-fold higher.

Mitotic HR occurs in the G2 phase of the cell cycle. Although there does not seem to be a specific structure required for mitotic HR, some chromosome scaffold or structure is required. Recent experiments show that DSB repair in G2 requires that chromosome cohesion, cohesion between sister chromatids, be established during S phase and be present in G2. Whether the cohesion complex helps direct most DSB repair by HR to interactions with the sister chromatid instead of the homolog is not known. In meiosis, in most eukaryotic organisms, the chromosomes become associated with a protein structure called the *synaptonemal complex* (*SC*). SC forms between homologous chromosomes. In yeast, DSB formation by Spo11 occurs prior to SC and the SC is thought to aid in homolog alignment. Recombination by the DSBR model occurs in the zygotene and early pachytene stages of meiosis. By pachytene, chromosomes are condensed and paired along their entire lengths, with SC. The SC dissolves in diplotene as chromosomes come apart in preparation for the first meiotic chromosome division.

The link between SC and HR is complicated as in some organisms SC formation is dependent on meiotic DSB formation, while in other organisms SC can form in the absence of DSBs. Nonetheless, mutations in SC components reduce meiotic HR, with resulting chromosome missegregation.

In addition to the meiotic-specific Spo11 protein and SC components, there are proteins that are required for meiotic HR that are expressed only in meiosis. Some of these include the homolog pairing proteins Hop1, Hop2, and Red1, a meiosis-specific Rad51 paralog called *Dmc1*, mismatch repair-like proteins called *Msh4* and *Msh5*, and a meiotic-specific checkpoint kinase called *Mek1*. These proteins are essential for meiotic HR, but have no detectable function in mitosis. Dmc1 does not substitute for Rad51 in the HR step of strand exchange, but does have strand exchange activity *in vitro*. Loss of Dmc1 results in arrest in meiosis during pachytene, without completion of HR. The arrest suggests that there is a checkpoint signal for incomplete HR. Some of the mitotic DNA damage checkpoint factors function in meiosis to ensure efficient repair of the meiotic DSBs by HR and arrest cells when DSBs are not repaired by HR. The Mek1 kinase appears to regulate DSB repair through Dmc1. Msh4 and Msh5 are related to the mismatch repair proteins MutS of *E. coli* and Msh2 of yeast. However, Msh4 and Msh5 do not function in mismatch repair. Rather, they are essential for crossing-over in meiosis and loss of Msh4 or Msh5 in mice results in aberrant meioses and sterility. Msh4 and Msh5 act with the mismatch repair protein Mlh1 in meiosis in this pathway.

Most mitotic DSB repair that produces genetic recombinant products is of the gene conversion type without any associated crossing-over. This observation led to the development of the SDSA model. In contrast, about half of the meiotic genes conversions are associated with crossing-over. This observation was crucial in the development of the DSBR model, with resolution of the HJs as crossover or noncrossover. However, recent studies on the timing of molecular intermediates in meiotic HR in yeast have shown that noncrossovers occur before crossovers. This has led to the suggestion that there are two types of meiotic HR pathways. The first is dedicated to noncrossovers and occurs by the SDSA model for gene conversions. The second pathway occurs slightly later through the DSBR model. Resolution of the HJs results in crossovers. The suggestion is that all noncrossovers come from SDSA HR while crossovers come from DSBR, which has HJs. Moreover, the HJs are always resolved to give crossovers. Both the SDSA and DSBR models involve formation of heteroduplex DNA, which is a substrate for the mismatch repair proteins. Mismatch repair occurs whenever DNA mismatches arise. In mitotic growth, most mismatches occur from errors during DNA replication where an incorrect nucleotide is used, resulting in a base-pair mismatch. If the mismatch is not corrected by the editing function of the DNA polymerase, it becomes a target for the mismatch repair proteins. Mismatch repair is an essential step in recombination, correcting the heteroduplex that has formed between two recombining DNA strands. In the absence of mismatch repair, mutations will accumulate and the strains will have a mutator phenotype. The strains are able to undergo recombination, but maturation of the heteroduplex is defective.

3.2 Sister Chromatid Recombination versus Homolog Recombination

Sister chromatid recombination can only occur in late S or G2, after chromosomes are replicated. As the recombination occurs between identical DNA duplexes, there is no genetic signal for its occurrence. Sister chromatid recombination can be detected genetically when it occurs out of register within a duplication or a multiple

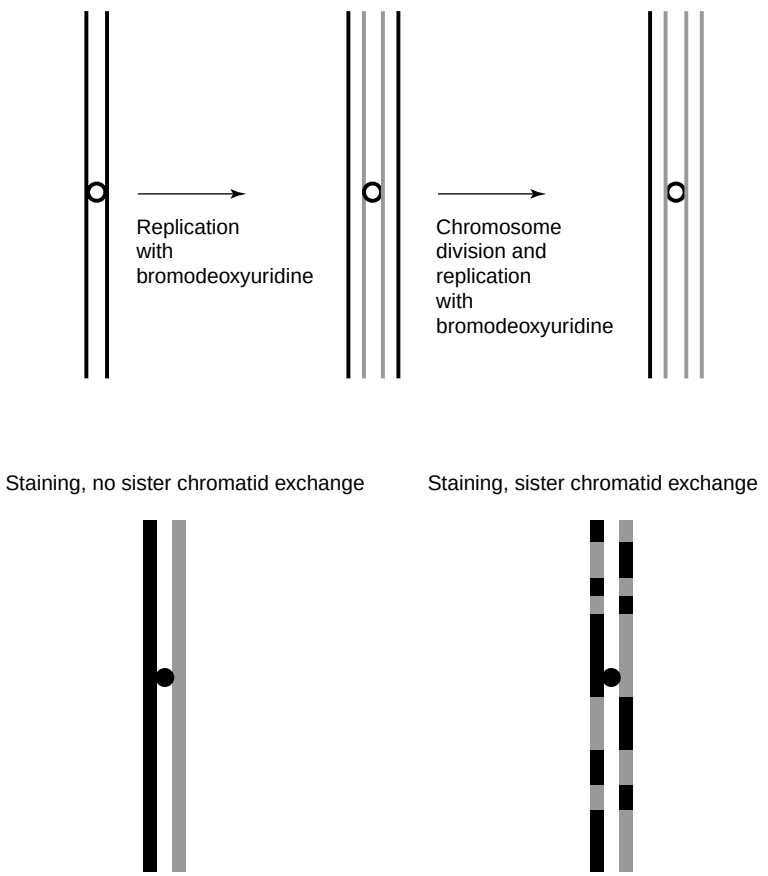

Fig. 6 Cytological detection of sister chromatid exchange. Cell is grown for two rounds of replication in the presence of the nucleotide analog bromodeoxyuridine. After one round of replication, the chromatids are hybrid, with one substituted and one unsubstituted DNA strand. After the second round of replication, cells are arrested in mitosis. The chromosomes have one fully substituted chromatid and one hemi-substituted chromatid. Differential staining of the fully substituted and hemi-substituted chromatids results in a microscopic light/dark pattern, with the fully substituted chromatid being light colored. If no sister chromatid exchange has occurred, then one chromatid is fully dark while the attached sister chromatid is fully light. If sister chromatid exchange has occurred, each chromatid will show an alternating pattern of light and dark regions, and the two attached chromatids are mirror images of each other.

tandem repeat. Assays using duplications have been devised to detect unequal sister chromatid exchange (Fig. 6). Alternatively, sister chromatid recombination can be detected visually in vertebrate cells using chromosome staining methods. Cells are allowed to go through two rounds of replication with a nucleotide substitute of bromodeoxyuridine for thymidine. Cells are then arrested in mitosis with a microtubule inhibitor, so that cells will have replicated chromosomes with both sister chromatids still attached at the centromere. The chromosomes are stained with a fluorescent dye, which reveals the bromodeoxyuridine incorporation, and Giemsa stain for the chromosome regions that do no have bromodeoxyuridine. If no sister chromatid exchange has occurred, the chromatid with both strand containing bromodeoxyuridine will be light colored while the chromatid with one substituted strand will be dark colored. Where sister chromatid exchange has occurred, a single chromatid will contain both light and dark sections, and this alternates with the light/dark staining pattern of the sister chromatid. Such chromosomes are called *harlequin chromosomes*.

Using the harlequin assay, it has been found that spontaneous sister chromatid exchange requires HR proteins, but not NHEJ proteins. DNA damage induced sister chromatid exchanges also required HR proteins, showing that sister chromatid HR is important for DSB repair and damage avoidance. In fact, a hallmark of some human diseases that are defective in the correct response to DNA damage is increased sister chromatid exchange. The defect probably lies in an inability to regulate repair by nonrecombinogenic pathways at stalled replication forks.

Since most mitotic HR between homologs is gene conversion that is not associated with crossing-over, one might expect the same to be true for sister chromatid recombination. Using a recombination substrate consisting of a marked duplication with a site-specific DSB to induce the HR, it has been found that most DSB repair uses the sister chromatid as a template for repair, and that most sister chromatid recombination is in fact gene conversion, not exchange. This is consistent with an SDSA model for sister chromatid recombination and consistent with the observation that crossing-over, either between homolog chromosomes or sister chromatids is infrequent in mitotic cells, even in the presence of DSB damage.

If most DSB repair occurs between sister chromatids, how is the meiotic HR program modified to promote interhomolog HR? Part of the answer must lie in meiotic-specific proteins that promote interhomolog HR. Some of these include the meiotic-specific chromosome cohesion protein Rec8, the SC structure itself, and a newly recognized protein called *Mnd1*, which is proposed to promote close homolog juxtaposition in preparation for strand exchange.

3.3 Regulation by Cell Cycle DNA Damage Checkpoint Factors

DNA damage is sensed by cell cycle checkpoint factors that recognize unrepaired DSBs. One important signal in the checkpoint response is single-strand DNA coated with the single-strand DNA binding protein RPA. DNA damage checkpoint factors bind to unrepaired DSBs and set off a cascade of signal transduction through protein kinases that modify components of the replication and recombination repair machinery and arrest cells in the G2 phase of the cell cycle until damage has

been repaired. At this point, the arrest is reversed and cells resume growth and proceed through mitosis to the next cell cycle. In mammalian cells, apoptosis is also a response to unrepaired DSBs. Key proteins in the checkpoint response are the ATM (ataxia-telangiectasia mutated) and ATR (ATM and RAD3-related) protein kinases, which transduce the damage signal and phosphorylate many proteins involved in DNA repair, recombination, replication transcription, and cell cycle progression.

3.4 Alternative Repair Pathways

As has been discussed in Sects. 1.1 and 1.2, there are multiple DSB repair pathways in eukaryotic cells, some of which do not involve HR. However, not all spontaneous damage originates as DSBs. DSBs may arise from processing of stalled replication forks, but the initial damage may be single-strand DNA gaps at the fork. There is overlap in the pathways used for repair of spontaneous damage. Evidence for this comes from several sources. First, genetic studies in yeast have revealed several genes that function in alternative repair pathways for damage associated with stalled replication forks. One of these pathways involving the yeast *MPH1* gene also requires HR gene functions. However, loss of Mph1 function does not affect HR. Mph1 functions in error-free bypass of DNA lesions, in a manner that also requires HR. This most likely involves sister chromatid interactions, possibly some type of template switching, which requires strand invasion of the sister chromatid to synthesize past DNA lesions.

The second set of observations on repair pathways that overlap with HR comes from studies of chicken DT40 cells. The HR mutant RAD54 (−/−) is viable, but sensitive to ionizing radiation and defective in DSB-promoted HR assays. RAD18 is a key gene in the translesion synthesis repair, to fill in gaps on the daughter strands caused by lesions in the template strands during replication. The mutant RAD18 (−/−) is viable, but sensitive to UV radiation. The mutant has a great increase in sister chromatid exchange events, suggesting that damage normally repaired by RAD18 is now repaired by the HR pathway. The double mutant Rad54 (−/−) RAD18 (−/−) is lethal. These types of experiments highlight the cross talk between different repair modes of the cell.

4 Consequences of Defects in Homologous Recombination

4.1 Rearrangements Associated with Mutations in HR Genes

In yeast, the HR gene mutants are viable. Haploid HR mutants, in addition to reduced HR rates, are associated with increased genomic instability and deletions, translocations, and telomere fusions. Many of the rearrangements have breakpoints with no homology or microhomology. These rearrangements are similar to those observed in cancer cells, suggesting that defective HR results in the types of chromosome aberrations seen in tumor cells. Diploid HR mutants have chromosome loss as the most frequent type of chromosome aberration, although the aberrations listed above also arise, but at much lower frequencies than chromosome loss. These findings show that HR is essential for genome stability maintenance.

In vertebrate cells, many HR mutants are not viable. However, examination of

early embryonic cells has revealed massive chromosome rearrangements. Mutants in the RAD51 paralog genes are associated with chromosome and chromatid breaks. In the chick DT40 system, it has been possible to make conditional RAD51 mutant cells. These mutants have very high levels of spontaneous chromosome breaks and have increased sensitivity to DNA damaging agents.

BRCA2-deficient cells accumulate several different types of chromosome rearrangements and aberrations. Most prevalent are chromosome breaks, chromatid breaks, end-to-end fusions of chromosomes, translocations, and deletions, and chromosomes with multiple arms that appear to result from multiple aberrant recombination reactions. Similar chromosome aberrations have been seen in RAD51-deficient cells. The overlap in phenotypes is expected given that BRCA2 and RAD51 function together at an early stage in HR. The finding that the rearrangements occur spontaneously underscores the importance of HR in maintaining genomic stability in growing cells, repairing lesions that occur during DNA replication.

4.2 Meiotic Defects Associated with Mutations in HR Genes

Meiotic phenotypes associated with defects in HR genes have been seen only in mutants of genes that are either only expressed in meiosis or genes that have milder HR consequences in mitosis. In yeast, the HR mutants are mitotically viable, although they are associated with decreased HR and increased genomic instability and chromosome rearrangements. In meiosis, the same mutants fail to complete meiosis and give no viable meiotic products. Thus, these mutants are meiotic-inviable, but mitotic-viable. There exists another set of mutants in meiotic-specific genes such as *SPO11*, *DMC1*, *MSH4*, and *MSH5* plus genes encoding subunits of the synaptonemal complex. Mutants in these genes give no viable meiotic products, or a reduced number of viable meiotic products with aneuploid karyotypes. In the mouse, *SPO11*−/− mutants are infertile, have synapsis defects cells undergoing meiosis, and apoptosis is increased in the spermatocytes and ooctyes cells.

In vertebrates, most HR mutants are not mitotically viable. Most work has focused on the meiotic-specific genes. Mutants in these genes are associated with aberrant chromosome structures in meiosis and infertility, consistent with the essential role of HR in meiotic chromosome segregation.

5 Concluding Remarks

In the past few years, our understanding of HR models, proteins, and regulation has been elevated by quantum leaps. It is now appreciated that HR occurs by distinct mechanisms, even with the same initiating substrate of a DSB. Most mitotic HR occurs by SDSA, and in fact most mitotic HR probably occurs between sister chromatids, not between homologs. Meiotic HR occurs by SDSA and DSBR, with the two types of HR occurring with different timing in a meiotic program that carefully regulates the number and spacing of crossovers. Most meiotic HR occurs between homologs, in contrast to mitotic HR.

Although some of the HR proteins are related, such as RAD51 and DMC1, and RAD54 and RAD54B, each protein has a distinct role in the HR process. The

number of accessory proteins seems to be increasing to include DNA damage checkpoint factors. One aspect of HR that remains controversial is the eukaryotic resolvase. While resolvase activity has been identified in cell extracts and found to be missing from mutant cell extracts, the resolvase protein has not yet been definitively identified.

The importance of HR is underscored by the finding that HR is essential for all dividing cells. Given its essential role, it is not surprising that there are few human diseases associated with mutations in the genes required for the HR process itself, such as the *RAD51* gene. However, there are rare cases of mutations in genes that encode proteins that are mediators of HR or are checkpoint proteins for the DNA damage response pathway. These include ATM, leading to ataxia-telangiectasia, NBS1, leading to Nijmegen breakage syndrome, ATR, leading to Sekel syndrome, BLM, leading to Bloom syndrome, WRN, leading to Werner syndrome, MRE11, leading to AT-like disease, the FA genes, leading to Fanconi anemia, BRCA1 and BRCA2, leading to breast cancer. Most of these diseases are associated with increased cancer risk, and on the cellular level, there is increased genomic instability and defective repair of DNA damage, including DSBs. Understanding how these HR defects result in genomic instability is the first step in being able to limit the damage caused by genomic instability.

Acknowledgment

Work in the Klein laboratory is supported by grants from the National Institutes of Health (GM53738 and GM070711).

See also Genetics, Molecular Basis of; RecA Superfamily Proteins; Repair and Mutagenesis of DNA.

Bibliography

Books and Reviews

Jasin, M. (2002) Homologous repair of DNA damage and tumorigenesis: the BRCA connection, *Oncogene* **21**, 8981–8993.

Keeney, S. (2001) Mechanism and control of meiotic recombination initiation, *Curr. Top. Dev. Biol.* **52**, 1–53.

Kolodner, R.D., Putnam, C.D., Myung, K. (2002) Maintenance of genome stability in *Saccharomyces cerevisiae*, *Science* **297**, 552–557.

Lieber, M.R., Ma, Y., Pannicke, U., Schwarz, K. (2003) Mechanism and regulation of human nonhomologous DNA end-joining, *Nat. Rev. Mol. Cell. Biol.* **4**, 712–720.

Mills, K.D., Ferguson, D.O., Alt, F.W. (2003) The role of DNA breaks in genomic instability and tumorigenesis, *Immunol. Rev.* **194**, 77–95.

Paques, F., Haber, J.E. (1999) Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae, *Microbiol. Mol. Biol. Rev.* **63**, 349–404.

Richardson, C., Horikoshi, N., Pandita, T.K. (2004) The role of the DNA double-strand break response network in meiosis, *DNA Repair (Amst)* **3**, 1149–1164.

Shivji, M.K., Venkitaraman, A.R. (2004) DNA recombination, chromosomal stability and carcinogenesis: insights into the role of BRCA2, *DNA Repair (Amst)* **3**, 835–843.

Symington, L.S. (2002) Role of *RAD52* epistasis group genes in homologous recombination and double-strand break repair, *Microbiol. Mol. Biol. Rev.* **66**, 630–670.

Yamazoe, M., Sonoda, E., Hochegger, H., Takeda, S. (2004) Reverse genetic studies of the DNA damage response in the chicken B lymphocyte line DT40, *DNA Repair (Amst)* **3**, 1175–1185.

Primary Literature

Allers, T., Lichten, M. (2001) Differential timing and control of noncrossover and crossover

recombination during meiosis, *Cell* **106**, 47–57.

Allers, T., Lichten, M. (2001) Intermediates of yeast meiotic recombination contain heteroduplex DNA, *Mol. Cell* **8**, 225–231.

Baudat, F., Manova, K., Yuen, J.P., Jasin, M., Keeney, S. (2000) Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11, *Mol. Cell* **6**, 989–998.

Chen, C., Kolodner, R.D. (1999) Gross chromosomal rearrangements in *Saccharomyces cerevisiae* replication and recombination defective mutants, *Nat. Genet.* **23**, 81–85.

Chen, C., Umezu, K., Kolodner, R.D. (1998) Chromosomal rearrangements occur in *S. cerevisiae rfa1* mutator mutants due to mutagenic lesions processed by double-strand-break repair, *Mol. Cell* **2**, 9–22.

Couedel, C., Mills, K.D., Barchi, M., Shen, L., Olshen, A., Johnson, R.D., Nussenzweig, A., Essers, J., Kanaar, R., Li, G.C., Alt, F.W., Jasin, M. (2004) Collaboration of homologous recombination and nonhomologous end-joining factors for the survival and integrity of mice and cells, *Genes Dev.* **18**, 1293–1304.

Davies, A.A., Masson, J.Y., McIlwraith, M.J., Stasiak, A.Z., Stasiak, A., Venkitaraman, A.R., West, S.C. (2001) Role of BRCA2 in control of the RAD51 recombination and DNA repair protein, *Mol. Cell* **7**, 273–282.

Deans, B., Griffin, C.S., O'Regan, P., Jasin, M., Thacker, J. (2003) Homologous recombination deficiency leads to profound genetic instability in cells derived from Xrcc2-knockout mice, *Cancer Res.* **63**, 8181–8187.

Dronkert, M.L., Beverloo, H.B., Johnson, R.D., Hoeijmakers, J.H., Jasin, M., Kanaar, R. (2000) Mouse *RAD54* affects DNA double-strand break repair and sister chromatid exchange, *Mol. Cell. Biol.* **20**, 3147–3156.

Edelmann, W., Cohen, P.E., Kneitz, B., Winand, N., Lia, M., Heyer, J., Kolodner, R., Pollard, J.W., Kucherlapati, R. (1999) Mammalian MutS homologue 5 is required for chromosome pairing in meiosis, *Nat. Genet.* **21**, 123–127.

Edelmann, W., Cohen, P.E., Kane, M., Lau, K., Morrow, B., Bennett, S., Umar, A., Kunkel, T., Cattoretti, G., Chaganti, R., Pollard, J.W., Kolodner, R.D., Kucherlapati, R. (1996) Meiotic pachytene arrest in MLH1-deficient mice, *Cell* **85**, 1125–1134.

Essers, J., van Steeg, H., de Wit, J., Swagemakers, S.M., Vermeij, M., Hoeijmakers, J.H., Kanaar, R. (2000) Homologous and nonhomologous recombination differentially affect DNA damage repair in mice, *EMBO J.* **19**, 1703–1710.

Gangloff, S., Soustelle, C., Fabre, F. (2000) Homologous recombination is responsible for cell death in the absence of the Sgs1 and Srs2 helicases, *Nat. Genet.* **25**, 192–194.

Henderson, K.A., Keeney, S. (2004) Tying synaptonemal complex initiation to the formation and programmed repair of DNA double-strand breaks, *Proc. Natl. Acad. Sci. U. S. A* **101**, 4519–4524.

Johnson, R.D., Jasin, M. (2000) Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells, *EMBO J.* **19**, 3398–3407.

Keeney, S., Giroux, C.N., Kleckner, N. (1997) Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family, *Cell* **88**, 375–384.

Kneitz, B., Cohen, P.E., Avdievich, E., Zhu, L., Kane, M.F., Hou, H. Jr., Kolodner, R.D., Kucherlapati, R., Pollard, J.W., Edelmann, W. (2000) MutS homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice, *Genes Dev.* **14**, 1085–1097.

Kraakman-van der Zwet, M., Overkamp, W.J., van Lange, R.E., Essers, J., van Duijn-Goedhart, A., Wiggers, I., Swaminathan, S., van Buul, P.P., Errami, A., Tan, R.T., Jaspers, N.G., Sharan, S.K., Kanaar, R., Zdzienicka, M.Z. (2002) Brca2 (XRCC11) deficiency results in radioresistant DNA synthesis and a higher frequency of spontaneous deletions, *Mol. Cell. Biol.* **22**, 669–679.

Krejci, L., Van Komen, S., Li, Y., Villemain, J., Reddy, M.S., Klein, H., Ellenberger, T., Sung, P. (2003) DNA helicase Srs2 disrupts the Rad51 presynaptic filament, *Nature* **423**, 305–309.

Libby, B.J., De La Fuente, R., O'Brien, M.J., Wigglesworth, K., Cobb, J., Inselman, A., Eaker, S., Handel, M.A., Eppig, J.J., Schimenti, J.C. (2002) The mouse meiotic mutation mei1 disrupts chromosome synapsis with sexually dimorphic consequences for meiotic progression, *Dev. Biol.* **242**, 174–187.

Lim, D.S., Hasty, P. (1996) A mutation in mouse *rad51* results in an early embryonic lethal that

is suppressed by a mutation in *p*53, *Mol. Cell. Biol.* **16**, 7133–7143.

Lisby, M., Rothstein, R., Mortensen, U.H. (2001) Rad52 forms DNA repair and recombination centers during S phase, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8276–8282.

Lisby, M., Mortensen, U.H., Rothstein, R. (2003) Colocalization of multiple DNA double-strand breaks at a single Rad52 repair centre, *Nat. Cell Biol.* **5**, 572–577.

Lomonosov, M., Anand, S., Sangrithi, M., Davies, R., Venkitaraman, A.R. (2003) Stabilization of stalled DNA replication forks by the BRCA2 breast cancer susceptibility protein, *Genes Dev.* **17**, 3017–3022.

Mahadevaiah, S.K., Turner, J.M., Baudat, F., Rogakou, E.P., de Boer, P., Blanco-Rodriguez, J., Jasin, M., Keeney, S., Bonner, W.M., Burgoyne, P.S. (2001) Recombinational DNA double-strand breaks in mice precede synapsis, *Nat. Genet.* **27**, 271–276.

Metzler-Guillemain, C., de Massy, B. (2000) Identification and characterization of an *SPO11* homolog in the mouse, *Chromosoma* **109**, 133–138.

Mills, K.D., Ferguson, D.O., Essers, J., Eckersdorff, M., Kanaar, R., Alt, F.W. (2004) Rad54 and DNA Ligase IV cooperate to maintain mammalian chromatid stability, *Genes Dev.* **18**, 1283–1292.

Morrison, C., Takeda, S. (2000) Genetic analysis of homologous DNA recombination in vertebrate somatic cells, *Int. J. Biochem. Cell Biol.* **32**, 817–831.

Moynahan, M.E., Pierce, A.J., Jasin, M. (2001) BRCA2 is required for homology-directed repair of chromosomal breaks, *Mol. Cell* **7**, 263–272.

Myung, K., Chen, C., Kolodner, R.D. (2001) Multiple pathways cooperate in the suppression of genome instability in *Saccharomyces cerevisiae*, *Nature* **411**, 1073–1076.

Nichols, M.D., DeAngelis, K., Keck, J.L., Berger, J.M. (1999) Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11, *EMBO J.* **18**, 6177–6188.

Parisi, S., McKay, M.J., Molnar, M., Thompson, M.A., van der Spek, P.J., van Drunen-Schoenmaker, E., Kanaar, R., Lehmann, E., Hoeijmakers, J.H., Kohli, J. (1999) Rec8p, a meiotic recombination and sister chromatid cohesion phosphoprotein of the Rad21p family conserved from fission yeast to humans, *Mol. Cell. Biol.* **19**, 3515–3528.

Patel, K.J., Yu, V.P., Lee, H., Corcoran, A., Thistlethwaite, F.C., Evans, M.J., Colledge, W.H., Friedman, L.S., Ponder, B.A., Venkitaraman, A.R. (1998) Involvement of Brca2 in DNA repair, *Mol. Cell* **1**, 347–357.

Petukhova, G., Stratton, S., Sung, P. (1998) Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins, *Nature* **393**, 91–94.

Richardson, C., Jasin, M. (2000a) Frequent chromosomal translocations induced by DNA double-strand breaks, *Nature* **405**, 697–700.

Richardson, C., Jasin, M. (2000b) Coupled homologous and nonhomologous repair of a double-strand break preserves genomic integrity in mammalian cells, *Mol. Cell. Biol.* **20**, 9068–9075.

Romanienko, P.J., Camerini-Otero, R.D. (2000) The mouse *Spo11* gene is required for meiotic chromosome synapsis, *Mol. Cell* **6**, 975–987.

Sehorn, M.G., Sigurdsson, S., Bussen, W., Unger, V.M., Sung, P. (2004) Human meiotic recombinase Dmc1 promotes ATP-dependent homologous DNA strand exchange, *Nature* **429**, 433–437.

Sharan, S.K., Morimatsu, M., Albrecht, U., Lim, D.S., Regel, E., Dinh, C., Sands, A., Eichele, G., Hasty, P., Bradley, A. (1997) Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking Brca2, *Nature* **386**, 804–810.

Sigurdsson, S., Van Komen, S., Bussen, W., Schild, D., Albala, J.S., Sung, P. (2001) Mediator function of the human Rad51B-Rad51C complex in Rad51/RPA-catalyzed DNA strand exchange, *Genes Dev.* **15**, 3308–3318.

Sonoda, E., Takata, M., Yamashita, Y.M., Morrison, C., Takeda, S. (2001) Homologous DNA recombination in vertebrate cells, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8388–8394.

Stark, J.M., Jasin, M. (2003) Extensive loss of heterozygosity is suppressed during homologous repair of chromosomal breaks, *Mol. Cell. Biol.* **23**, 733–743.

Sung, P. (1997) Function of yeast Rad52 protein as a mediator between replication protein A and the Rad51 recombinase, *J. Biol. Chem.* **272**, 28194–28197.

Sung, P., Robberson, D.L. (1995) DNA strand exchange mediated by a RAD51-ssDNA

nucleoprotein filament with polarity opposite to that of RecA, *Cell* **82**, 453–461.

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J., Stahl, F.W. (1983) The double-strand-break repair model for recombination, *Cell* **33**, 25–35.

Takata, M., Sasaki, M.S., Sonoda, E., Fukushima, T., Morrison, C., Albala, J.S., Swagemakers, S.M., Kanaar, R., Thompson, L.H., Takeda, S. (2000) The Rad51 paralog Rad51B promotes homologous recombinational repair, *Mol. Cell. Biol.* **20**, 6476–6482.

Takata, M., Sasaki, M.S., Tachiiri, S., Fukushima, T., Sonoda, E., Schild, D., Thompson, L.H., Takeda, S. (2001) Chromosome instability and defective recombinational repair in knockout mutants of the five Rad51 paralogs, *Mol. Cell. Biol.* **21**, 2858–2866.

Tarsounas, M., Munoz, P., Claas, A., Smiraldo, P.G., Pittman, D.L., Blasco, M.A., West, S.C. (2004) Telomere maintenance requires the RAD51D recombination/repair protein, *Cell* **117**, 337–347.

Tsuzuki, T., Fujii, Y., Sakumi, K., Tominaga, Y., Nakao, K., Sekiguchi, M., Matsushiro, A., Yoshimura, Y., Morita, T. (1996) Targeted disruption of the *Rad51* gene leads to lethality in embryonic mice, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 6236–6240.

Veaute, X., Jeusset, J., Soustelle, C., Kowalczykowski, S.C., Le Cam, E., Fabre, F. (2003) The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments, *Nature* **423**, 309–312.

Wiese, C., Pierce, A.J., Gauny, S.S., Jasin, M., Kronenberg, A. (2002) Gene conversion is strongly induced in human cells by double-strand breaks and is modulated by the expression of BCL-x(L), *Cancer Res.* **62**, 1279–1283.

Yamashita, Y.M., Okada, T., Matsusaka, T., Sonoda, E., Zhao, G.Y., Araki, K., Tateishi, S., Yamaizumi, M., Takeda, S. (2002) RAD18 and RAD54 cooperatively contribute to maintenance of genomic stability in vertebrate cells, *EMBO J.* **21**, 5558–5566.

Yu, V.P., Koehler, M., Steinlein, C., Schmid, M., Hanakahi, L.A., van Gool, A.J., West, S.C., Venkitaraman, A.R. (2000) Gross chromosomal rearrangements and genetic exchange between nonhomologous chromosomes following BRCA2 inactivation, *Genes Dev.* **14**, 1400–1406.

Regulation by Protein Phosphoylation: *see* Biological Regulation by Protein Phosphorylation

Regulation of Gene Expression

Anil Kumar
Devi Ahilya University, Indore, MP, India

1 Regulation of Gene Expression in Prokaryotes 27
1.1 Induction and Repression 27
1.2 The Operon 28
1.3 The Lactose Operon (*Lac* Operon) 29
1.4 The Histidine Operon 33
1.5 The Tryptophan Operon 34
1.6 Attenuation: the Leader Sequence 36
1.7 Positive and Negative Control 37
1.8 The Arabinose Operon (*Ara* Operon) 39
1.9 Catabolite Repression 40
1.10 Cyclic AMP Receptor Protein 40
1.11 Guanosine-5′-Diphosphate, 3′-Diphosphate 43

2 Regulation of Gene Expression in Eukaryotes 43
2.1 Transcriptionally Active Chromatin 46
2.2 Regulation of Gene Expression at the Initiation of Transcription 48

3 RNA Splicing 52
3.1 Nuclear Splicing 53
3.2 Spliceosome and Small Nuclear RNAs 55
3.3 Group I and Group II Introns 56
3.4 Cis- and Trans-splicing Reactions 57
3.5 tRNA Splicing 57
3.6 Alternate Splicing 60

4 Posttranscriptional Regulation of mRNA 61

5 Transport of Processed mRNA to the Cytoplasm 62

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

6 Regulation of Gene Expression at the level of Translation 62

Acknowledgments 63

Bibliography 64
Books and Reviews 64
Primary Literature 65

Keywords

Alternate Splicing
When a single gene after splicing gives rise to more than one mRNA sequences. It may be due to joining of exons in different series. Sometimes, HnRNA may splice differently (a portion of sequence may act as intron in one case and as exon in other case).

Attenuation
A mechanism to control RNA polymerase to read through an attenuator, an intrinsic termination sequence present at the beginning of the transcription unit. This type of control is present in some prokaryotic operons.

CAAT Box
A conserved sequence located nearly 75 nucleotides upstream of the start point of transcription units found in eukaryotes, also called as −75 box sequence. It is recognized by certain transcription factors. It has the consensus sequence GGCAATCT. It plays an important role in increasing the promoter strength.

Cyclic AMP Receptor Protein (CRP or CAP)
A regulatory protein activated by 3′,5′ cyclic AMP (cAMP). In prokaryotes, transcription of many genes gets activated after binding of this protein (in the form of CRP–cAMP complex) on a specific site in the DNA. Two molecules of cyclic AMP bind with one molecule of CRP.

Exon
A segment of interrupted gene having coding region and present in mature mRNA.

Gratuitous Inducer
A substance that induces the transcription of a gene(s) but is not a substrate for its enzyme protein product. Generally, it is an analog of the substrate, a normal inducer.

Intron or Intervening Sequence
A segment of interrupted gene found in eukaryotes. Intron gets transcribed but do not code for a protein product. Intron sequence gets removed during maturation of primary transcript, a process called as RNA splicing.

Inducer
A small molecule that triggers biosynthesis of RNA by binding to the cytoplasmic repressor (product of a regulatory gene). It is generally the substrate of the enzyme protein product of the structural gene.

Induction
The ability of bacteria or yeast to synthesize certain enzymes only when their substrates are present. The inducer binds to the cytoplasmic repressor preventing it from binding to the operator region. If cytoplasmic repressor is already bound with the operator, it gets detached from the operator region after binding with the inducer.

Lariat
An intermediate formed during RNA splicing where a circular structure with a tail is formed by a 5′,2′ bond.

Leader Sequence
A nontranslated sequence at the 5′ end of mRNA preceding the initiation codon.

Operator
A DNA sequence to which cytoplasmic repressor (protein product of regulatory gene) binds specifically.

Polyadenylation
The addition of Poly A sequence to the 3′ end of an eukaryotic RNA (a posttranscriptional change).

Polycistronic mRNA
An mRNA having the information for more than one protein. It forms after transcription of more than one gene present in a cluster (operon).

Promoter
The region of DNA involved in the binding of RNA polymerase to start RNA biosynthesis.

Regulatory Gene
This gene codes for an RNA or a protein that controls the expression of other genes.

Repression
Inhibition of enzyme biosynthesis by a product of the metabolic pathway. Generally, inhibition is at the level of transcription. The product of the regulatory gene (cytoplasmic repressor) and the product of the metabolic pathway (corepressor) complex binds to the operator region on the DNA.

Ribozyme
RNA as an enzyme. Some RNA molecules are capable of self-RNA splicing without involvement of any protein. This type of RNA is called as ribozyme.

SnRNAs (Small Nuclear RNAs)
These are small RNAs present in the nucleus and are considered to be involved in RNA splicing/other processing reactions.

SnRNPs
Are small nuclear ribonucleoproteins. In these, SnRNAs are associated with proteins.

Splicing
Removal of introns and joining of exons in RNA.

TATA box
A conserved sequence found about 25 nucleotides upstream from the start point of eukaryotic RNA polymerase II transcription unit. It is considered to be involved in positioning RNA polymerase II for correct initiation.

Telomerase
Is an enzyme resembling reverse transcriptase. Telomerase enzyme adds telomeres to chromosome ends.

Telomere
Are the specialized structures at the ends of linear eukaryotic chromosomes. They generally have many tandem copies of a short oligonucleotide sequence, T_aG_b in one strand and C_bA_a in the complementary strand, where a and b are on an average 1 to 4.

Upstream
The sequences found at the 5′ end of and beyond the region of expression.

The central dogma in gene expression is ''DNA makes the RNA, a process called as transcription; and RNA makes the protein, called as translation.'' Gene expression can be regulated at the stage of transcription, RNA processing (post-transcriptional changes), and translation. Organisms have been classified into two groups – Prokaryotes whose cells do not have distinct well-defined nucleus, and eukaryotes whose cells have distinct well-defined nucleus. Examples of prokaryotes are bacteria and the blue green algae and of eukaryotes are animals, plants and fungi. In prokaryotes, the RNA primary product may itself be the target of regulation. In eukaryotic cells, due to compartmentation, transport of mRNA from the nucleus to the cytoplasm could be an additional target for regulation. However, bacterial mRNA is directly available for protein biosynthesis soon after its synthesis. Regulation of transcription usually occurs at the stage of initiation. The regulators of initiation of transcription may be a protein or RNA. In prokaryotes, on–off of transcription is the main regulatory control of the gene expression, whereas, in eukaryotes, more complex regulatory mechanism of transcription takes place. In addition, RNA splicing also plays major role in the regulation of gene expression. Here, it is pertinent to mention that eukaryotic genes have been found to have coding as well

as noncoding sequences. In fact, as per recent human genome sequence data, more than 50% sequences are noncoding sequences whose function is also not clear. These are called *introns* or junk sequences. The coding sequences are called *exons*. Of course, there are regulatory elements too. The primary transcript of DNA has complementary sequences of both exons and introns and is called heterogenous RNA (HnRNA). The HnRNA gets spliced removing introns and ligating exons. Regulation of gene expression in both prokaryotes and eukaryotes is important since it decides whether a particular protein should be synthesized and in which quantity.

1 Regulation of Gene Expression in Prokaryotes

1.1 Induction and Repression

In prokaryotes, induction and repression especially in enzyme proteins is the prominent way of regulation at the level of transcription. There are certain proteins that are synthesized at constant rate at all the times. There are certain proteins, especially the enzymes, that are produced in larger amounts in the presence of certain substances. The substance, which is generally the substrate of the enzyme protein, enhances an enzyme's synthesis and is called an *inducer*, the enzyme is called as *inducible enzyme*, and the process is known as *enzyme induction*. It is not necessary for the inducer to be the substrate of the enzyme. The inducer may resemble the natural substrate and need not necessarily be affected by it. An inducer that is not the substrate of the enzyme is called *gratuitous inducer*. An inducer may induce more than one enzyme whose genes are arranged in cluster.

An inducible enzyme is normally present only in trace amounts in a bacterial cell, but its concentration can quickly increase even a thousand fold or more when its substrate is present in the medium particularly when this substrate is the only carbon source of the cell. Under these conditions, the induced enzyme is required to transform the substrate into a metabolite that can be utilized by the cell directly. A well-studied example of inducible enzyme is of β-galactosidase from *Escherichia coli*. The *E. coli* cells having wild type β-galactosidase gene do not utilize lactose if glucose is also present in the medium. If only lactose is present as sole carbon source or after the complete utilization of glucose, within one to two minutes the cells synthesize β-galactosidase enzyme and start utilizing lactose. Simultaneously, there is synthesis of β-galactoside permease, which is required for the transfer of β-galactoside inside the bacterial cell and also of β-thiogalactoside transacetylase. If the induced bacterial cells are transferred into a medium deficient in lactose, synthesis of β-galactosidase (along with β-galactoside permease and β-thiogalactoside transacetylase) ceases immediately and the previously induced enzyme also gets declined to a normal level. Induction of a group of related enzymes or proteins to the same extent by a single inducing agent is known as *coordinate induction*. Here, induction of β-galactosidase along with β-galactoside permease and β-thiogalactoside transacetylase is an example of coordinate induction.

Earlier, two hypotheses were given to explain the mechanism of enzyme induction using the β-galactosidase system:

1. Activation of preexisting protein occurred.
2. *De novo* synthesis of the protein occurred.

The first possibility was discounted because of the reason that, prior to induction, no protein could be detected having antigenic property similar to β-galactosidase. This observation suggested that actual synthesis of the enzyme protein occurs after addition of the inducer. At that time, scientists were knowing that specific protein will not be synthesized except in the presence of gene which dictates its primary structure (amino acid sequence). From the observations, it was also clear that *E. coli* cells carry structural gene for β-galactosidase enzyme protein. Thereafter, the question remained as to why β-galactosidase enzyme protein synthesis did not occur in the absence of inducer. For this question, again the following two explanations were considered:

1. It may be considered that inducer acts as some sort of template that triggers the enzyme protein synthesis. It was thought that because of this reason the inducer resembles the substrate.
2. It is possible that enzyme protein is not synthesized because of inhibition by some agent(s). In this case, the inducer itself may act as some sort of inhibitor, which inhibits the activity of inhibiting agent(s). This possibility seems to be very complex, but afterward, evidence became available in support of this hypothesis and now we know that this is what actually occurs.

Similarly, in the presence of a substance that is the product of a particular enzyme reaction, the synthesis of an enzyme protein can be reduced. This phenomenon is called *enzyme repression* and the substance is called *corepressor*.

Induction and repression are two complementary phenomena. Generally biosynthetic pathways have been found under the control of repression. Many amino acids biosynthetic pathways are found to be under the control of repression. For example, if histidine is added to *E. coli* growth medium, all the enzymes involved in its biosynthesis no longer are synthesized since cells do not require to synthesize histidine. In the presence of histidine, all the enzymes required for the biosynthesis of histidine starting from ATP phosphoribosyl transferase, which catalyzes the first reaction in the histidine biosynthetic pathway (biosynthesis of phosphoribosyl ATP from phosphoribosyl pyrophosphate and ATP), are repressed. Such repression of the synthesis of a group of enzymes by a single corepressor is called *coordinate repression*. Coordinate repression is generally caused by the end product of the biosynthetic pathway, and, because of this reason, it is also called *end product repression*.

1.2 The Operon

Jacob and Monod in 1961 gave the concept of operon. They explained that genes encoding proteins whose functions are related, such as the consecutive genes proteins in a pathway, may be organized in a cluster that is transcribed into a polycistronic mRNA from a single operator and control of this operator regulates gene expression of the entire structural genes in the operon. The unit of regulation containing structural genes, regulator gene(s), and cis-acting elements is called *operon*. The genes may be classified into

two groups depending on their coded protein functions: *structural genes* whose protein products are directly involved in the metabolic activity (act as enzymes) or act as constituent of an organelle and the *regulatory genes* whose products regulate the transcription of the structural genes. The activity of an operon is controlled by the regulatory gene(s) whose protein product interacts in its control elements. Although many operons have been studied in detail, one of the most-studied examples of operon is *lac* operon of *E. coli*. In *lac* operon, *lac* i is the regulatory gene whose protein product called as cytoplasmic repressor is involved solely in regulation, while *lac* Z, *lac* Y, and *lac* A are structural genes which code for the enzymes, β-galactosidase, β-galactoside permease, and β-thiogalactoside transacetylase, respectively. The *lac* operon is described below.

1.3 The Lactose Operon (*Lac* Operon)

If *E. coli* cells are grown in a medium containing lactose as a sole carbon and energy source, the cells will synthesize the enzymes, β-galactosidase (which catalyzes the hydrolysis of lactose into glucose and galactose), β-galactoside permease (which allows lactose to enter the cell) and β-thiogalactoside transacetylase (which catalyzes the transfer of acetyl group from acetyl CoA on to 6 position of β-thiogalactoside to form 6-acetyl β-thiogalactoside). Glucose is easily metabolized and directly enters the glycolytic pathway. Galactose must be converted into glucose before it enters the glycolytic pathway. Studies with the *lac* operon showed that it consisted of three adjacent structural genes *lac* Z, *lac* Y, and *lac* A. Preceding these genes are *lac* O, *lac* P, and *lac* i, involved in regulation. *Lac* i codes for a protein called as cytoplasmic repressor. *Lac* P is the promoter site on which RNA polymerase binds. *Lac* O is the controlling site on which cytoplasmic repressor binds and after binding, there is switch off of transcription of the structural genes (Fig. 1). All the three structural genes are transcribed as a single polycistronic messenger RNA carrying genetic information for all the three enzyme proteins. Besides lactose, many other analogs such as isopropyl-β-thiogalactoside (IPTG), methyl- β-thiogalactoside, and mellibiose also act as inducers. *In vitro*, the most commonly used inducer for *lac* operon is IPTG.

The genes located in the *lac* operon are:

Gene Z^+ (*lac* Z^+): this gene in mutant condition results in the loss of ability to synthesize active β-galactosidase either in the presence or absence of inducer.

Gene Y^+ (*lac* Y^+): this gene in mutant condition results in the loss of ability to synthesize active β-galactoside permease either in the presence or absence of inducer.

Gene A^+ (*lac* A^+): this gene in mutant condition results in the loss of ability to synthesize active β-thiogalactoside transacetylase either in the presence or absence of inducer.

Gene i^+ (*lac* i^+): this gene causes changes in the influence of inducer on the synthesis of β-galactosidase, β-galactoside permease, and β-thiogalactoside transacetylase.

Many i mutants synthesize large amount of the enzymes in the absence of inducer. Using different combinations of wild type and mutant genes of *lac* Z, *lac* Y and *lac* i genes, different types of genetic structures

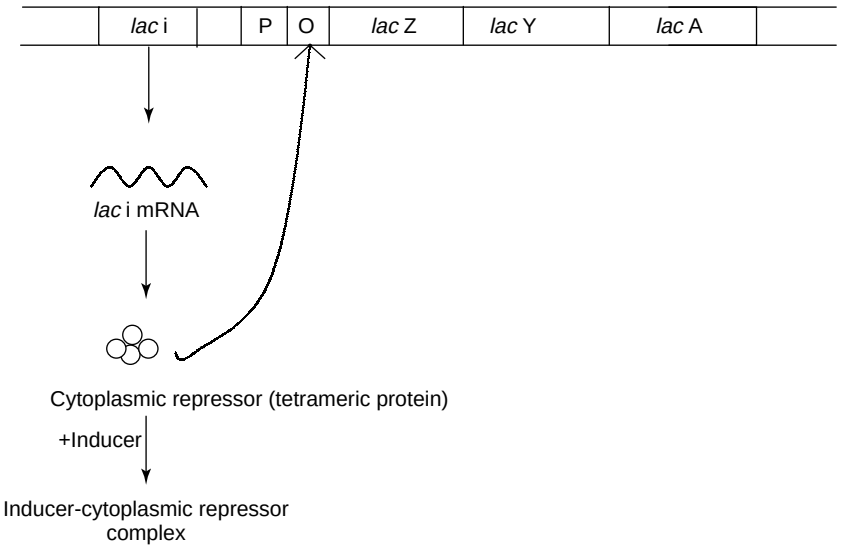

Fig. 1 Line diagram of *lac* operon: P, promoter; O, operator; *lac* i, regulatory gene; *lac* Z, Y, and A, structural genes for β galactosidase, β galactoside permease, and β thiogalactoside transacetylase, respectively. RNA polymerase binds on the promoter site. Cytoplasmic repressor binds on the operator site and represses transcription of the structural genes (negative control). In the presence of the inducer, there is formation of inducer-cytoplasmic repressor complex, which is not capable of binding on the operator site. If already cytoplasmic repressor is bound on the operator site, it gets detached from there after making a complex with the inducer, resulting in transcription of the structural genes.

Tab. 1 Various genotypes of the *E. coli* lactose system.

Genotype	***Noninducible***		***Inducible***	
	β-galactosidase	***β-galactoside permease***	***β-galactosidase***	***β-galactoside permease***
$Z^+ Y^+ i^+$	−	−	+	+
$Z^- Y^+ i^+$	−	−	−	+
$Z^+ Y^- i^+$	−	−	+	−
$Z^- Y^- i^+$	−	−	−	−
$Z^+ Y^+ i^-$	+	+	+	+
$Z^- Y^+ i^-$	−	+	−	+
$Z^+ Y^- i^-$	+	−	+	−
$Z^- Y^- i^-$	−	−	−	−

of *lac* region of *E. coli* chromosome may be assumed (Table 1).

Regulatory gene(s) need not necessarily be located near the structural genes. Its action is due to the biosynthesis of some intracellular substance or substances. Study of mutation in the regulatory gene provided insights into the main mechanism

of induction and repression. Mutated *lac* i gene is called as *lac* i^-. The *E. coli* cells containing i^+ produce β-galactosidase only in the presence of inducer whereas cells containing mutated *lac* i can produce β-galactosidase both in the presence and absence of inducer. These results indicate that *lac* i^+ is dominant and *lac* i^- is recessive.

Regulatory gene, *lac* i^+ codes for a protein called as cytoplasmic repressor or *lac* repressor. It is a tetrameric protein having four identical subunits of the molecular weight, around 37 000. This protein specifically binds on the operator region. Each subunit is a chain of 347 amino acids. The N-terminal amino acid is methionine and C-terminal is glutamine. The tetrameric protein may be dissociated in the presence of sodium dodecyl sulfate. Each individual subunit has one binding site for inducer. The individual subunits also have the ability to bind inducer but are not able to bind on the operator region. The binding constant for IPTG has been calculated to be about 10^{-6} M. The cytoplasmic repressor binds very tightly with the operator region. The equilibrium constant of the complex is about 10^{-13} M. The rate constants for association and dissociation are $7 \times 10^9\ M^{-1}\ sec^{-1}$ and $6 \times 10^{-4}\ sec^{-1}$, respectively. Although, there is normal content of various amino acids, tryptophan content is comparatively low. There are only two tryptophanyl residues out of 347 amino acids in a subunit. The $OD_{280}^{1\%}$ is 0.59 (comparatively low value due to low contents of tryptophan).

Using circular dichroism and optical rotatory dispersion studies, an estimate of about 33 to 40% α-helix contents and 18 to 42% β-structure has been made. Using Chou Fasman model, from the primary sequence, a prediction of 37% α-helix contents and 35% β-structure has been done. Both the data are in the same range.

Using electron microscopy after negative staining and powder X-ray diffraction analysis, an asymmetric dumbbell-shaped tetramer of the dimensions about 45 Å by 60 Å with four tetramers contained in one unit cell of 91 Å by 117 Å has been found. Powder X-ray diffraction patterns indicated that the third unit cell dimension, which is not seen in electron micrographs, is 140 Å. Four tetramers can easily be packed into this cell, in a manner that accounts for the stain distribution observed in electron micrographs. The molecule extends the full length of the 140 Å cell and gives the tetramer an elongated shape having approximately molecular dimensions of 140 Å by 60 Å by 45 Å. On the basis of these data, a model has been proposed in which the subunits are related by 222 symmetry and are placed at the corners of a rectangular plane. This suggests the existence of two operator binding sites per tetramer if the repressor maintains perfect 222 symmetry.

However, the shape of the *lac* repressor in solution appears quite different. In solution, the repressor tetramer appears as a square structure with dimensions of about 105 Å by 95 Å, and the shape as well as dimensions of the molecule does not change in the presence of IPTG. The subunits can be distinguished in tetrameric structure, however, due to poor resolution, no decision on the geometry of their arrangement could be taken.

It looks that there should be two different types of binding sites in the tetrameric structure: the first for the small molecular-weight effectors and the second for the *lac* operator. There are indications that both inducers and antiinducers bind to the same site, or at least to overlapping sites. *O*-Nitrophenyl β-D-fucoside has been

found to act as anti-inducer for *lac* operon. Similarly, the operator binding site involves the same region of the tetrameric repressor that binds nonoperator DNA. Each repressor subunit looks to have an effector binding site besides contributing some interactions with operator DNA. On the basis of experiments, it has been concluded that the effector binding site and operator binding site are distinct and nonoverlapping. On the basis of trypsin limited digestion studies which splits 59 amino terminal residues and 20 carboxyl terminal residues of each subunit, leaving a tetrameric trypsin resistant core protein made up of subunits having 60 to 327 residues, it has been shown that the amino and carboxyl terminal residues are neither required for the binding of inducer nor for the folding of the subunits into proper tetrameric conformation. However, it could not be confirmed whether the interaction between the subunits of the core is identical to the interaction of the subunits in the native repressor. There are indications that terminal regions are involved in operator binding. There are also indications that effector binding changes the affinity of the repressor for its operator. Both amino and carboxyl terminal regions have hydroxyl group containing amino acids (threonyl residues at position 5, 19, 34, 315, 316, 321, 323; seryl residues at position 16, 21, 28, 31, 309, 325, 332, 341, 345) which could contribute their hydroxyl groups of side chains to form hydrogen bonds with specific groups of the bases in the *lac* operator. Out of total 8 tyrosyl residues in the chain, there are four residues in the 59 residues amino terminal region (at position 7, 12, 17, 47), which offer the possibility of interactions with DNA either by providing additional hydroxyl groups or by intercalation between the bases. Besides, eight positively charged amino acids have been found in the 59 amino terminal region and six positively charged residues in the 36 carboxyl terminal region. It is the higher content of positively charged amino acids in the terminal regions (14.7%) compared to total percentage in the molecule (10.7%). It has been considered that the combination of the amino and carboxyl terminal regions constitute a basic region which through electrostatic interactions make contact with the negatively charged DNA and contribute to its specific binding to operator region. The region between 215 and 324 amino acyl residues has few charged residues (only 3 each positive and negatively charged residues) and is enriched in hydrophobic residues, therefore, it may involve in the stabilization of both the native tetramer and its tryptic core. This region may act as a hydrophobic nucleus resistant to trypsin attack.

The *lac* repressor can make four types of interactions:

1. Specific interaction between *lac* repressor and its operator
2. Nonspecific interaction between *lac* repressor and any DNA
3. Specific interaction between *lac* repressor and its small molecular weight effectors, which include inducers, antiinducers
4. The effector (inducer) may also interact with the *lac* repressor already bound to the operator to form an intermediate ternary complex.

One molecule of inducer IPTG is found to be sufficient to release the *lac* repressor from its specific operator. After binding of the inducer, there is nearly 1000-fold decrease in the affinity of the repressor for its operator. *Lac* repressor shows a single emission maximum at

338 nm in fluorescence spectrum, which is characteristic of tryptophan. On addition of inducer in saturating concentrations, there is a shift in emission maximum to 330 nm with no change in peak shape and also fluorescence intensity. This change in the emission maximum suggests that at least one tryptophanyl residue per subunit becomes less accessible to the solvent upon inducer binding. No change in the shape of the fluorescence spectrum indicates that either both tryptophanyl residues of the subunit are similarly affected or that only one contributes to the change in emission maximum. At least two sequential steps appear to be involved in the binding of IPTG to the *lac* repressor. The first step is bimolecular and much slower than expected for a diffusion-controlled reaction. The second step is monomolecular and may be attributed to a conformational change in the protein. The circular dichroism studies showed that no major changes in the overall geometry of the peptide backbone of the repressor occurred upon binding of the inducer. The sedimentation coefficient studies indicated compactness in the protein molecule upon inducer binding. The glycerol perturbation spectra indicated that fewer aromatic residues are available to solvent in the presence of the inducer than in the protein alone or in the presence of antiinducers. It has been predicted that the repressor undergoes a conformational change upon binding to its operator and a major change for induction may be taking place upon interaction of inducer with the repressor-operator complex.

After the availability of *lac* repressor in pure form, use of it was made to isolate the operator region. The DNA having *lac* region was fragmented into smaller fragments of nearly 1000 nucleotides. In the fragmented mixture, *lac* repressor was added and after incubation for some time, it was filtered through a cellulose nitrate membrane. DNA fragments without bound *lac* repressor passed through the filter whereas fragment bound with *lac* repressor remained tightly bound with the membrane. The bound DNA was released from the cellulose nitrate filter by adding IPTG. Afterward, *lac* repressor was added and treated the fragment with deoxynuclease. The operator region gets protected from digestion by deoxynuclease after binding the *lac* repressor. Nucleotide sequence determination revealed that the repressor protected 27 nucleotides and this sequence has a *dyad* symmetry. It has been shown that this *dyad* symmetry is important for specific binding of the *lac* repressor with its operator. The symmetrical sequence of the *lac* operator region is given below:

5′ TGGAATTGTGAGCGGATAACAATT 3′

3′ ACCTTAACACTCGCCTATTGTTAA 5′

1.4 The Histidine Operon

The histidine operon is an example of enzyme repression. In *Salmonella*, most of the structural genes encoding the enzymes required for histidine biosynthesis are arranged in the same order as the sequence of chemical reactions catalyzed by respective enzymes except one or two reactions. The sequence of the genes in the histidine operon is as follows where O denotes its operator:

EIFAHBCDGO

There are nine genes that specify the structure of nine proteins involved in histidine biosynthesis. Biosynthesis of

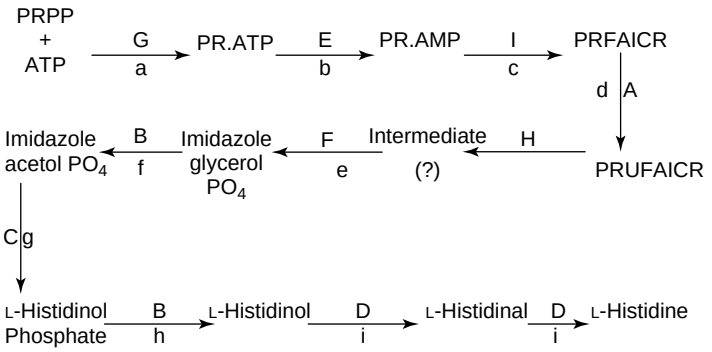

PRPP – Phosphoribosyl pyrophosphate
PRATP – Phosphoribosyl adenosine triphosphate
PRAMP – Phosphoribosyl adenosine monophosphate
PRFAICR – Phosphoribosyl formimino amino imidazole carboxamide ribonucleotide
PRUFAICR – Phosphoribulosyl formimino amino imidazole carboxamide ribonucleotide

a - ATP phosphoribosyl transferase
b - Pyrophosphohydrolase
c - Phosphoribosyl adenosine monophosphate cyclohydrolase
d - Phosphoribosyl formimino 5-amino-imidazole-4-carboxamide ribonucleotide isomerase
e - Glutamine amido transferase
f - Imidazole glycerol-3-phosphate dehydratase
g - L-Histidinol phosphate amino transferase
h - Histidinol phosphate phosphatase
i - Histidinol dehydrogenase

Fig. 2 Histidine biosynthetic pathway. The capital letters on the arrows denote the genes encoding the enzyme catalyzing the biochemical steps.

histidine starting from phosphoribosyl pyrophosphate (PRPP) and ATP along with the genes coding for the enzymes involved has been shown in Fig. 2. Two enzymes in the pathway have been found to be bifunctional. As written above, with the exception of 1 or 2 genes, the arrangement of genes on the chromosome is related to their position in the pathway *in vivo*. This indicates that chromosome contains a remarkable amount of information not only of the sequence of the amino acids in the enzyme proteins but also about the metabolic pathway catalyzed by these enzymes.

1.5 The Tryptophan Operon

The tryptophan operon consists of five structural genes that code for the enzymes involved in the biosynthesis of tryptophan. Tryptophan biosynthesis, in addition to enzyme repression, is also controlled by feed back inhibition with tryptophan inhibiting the activity of first enzyme unique to the tryptophan biosynthetic pathway. Besides, tryptophan operon is also controlled by attenuation. The line diagram of tryptophan operon showing the structural genes, regulatory genes and other regulatory elements is shown in Fig. 3.

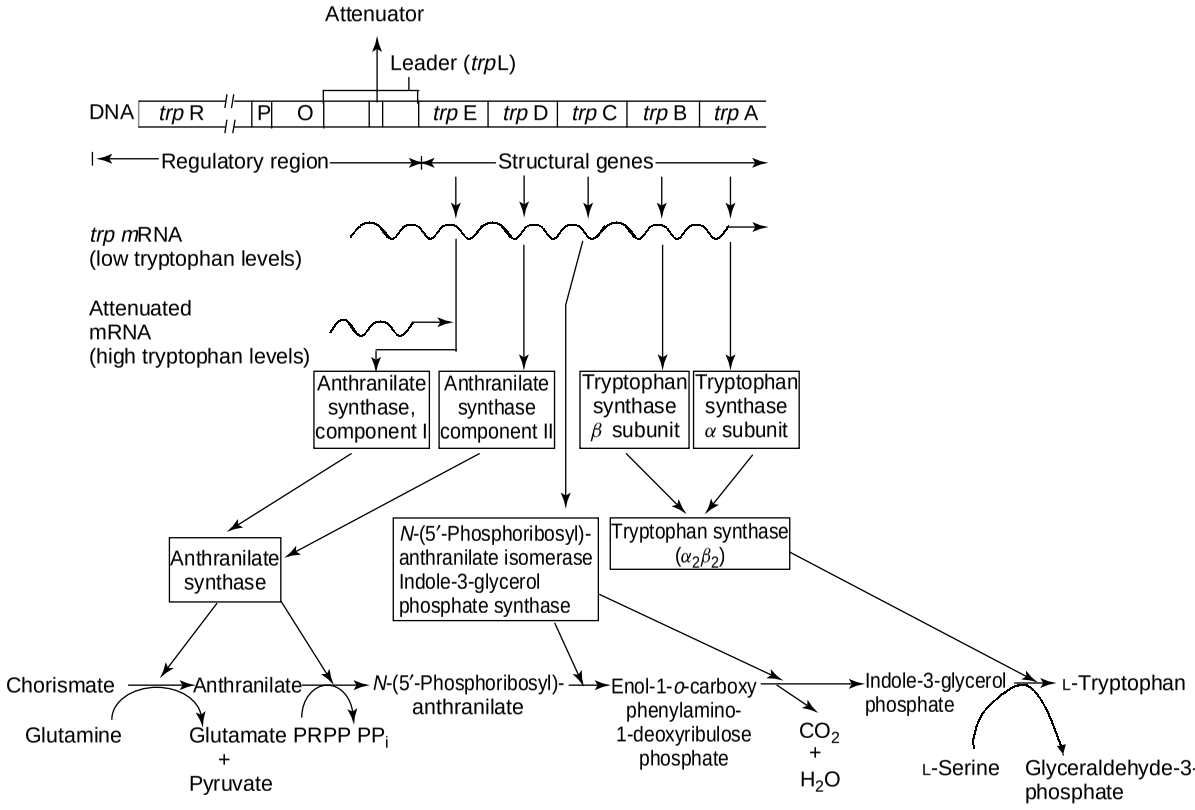

Fig. 3 Line diagram of tryptophan operon *Trp* R-regulatory gene; P, promoter; O, operator; between O and *trp* E is the leader sequence used in attenuation control. *Trp* E codes for anthranilate synthase component I and *trp* D for component II. Both components I and II on combination form active anthranilate synthase. *Trp* C codes for N-(5′-phosphoribosyl)-anthranilate isomerase Indole-3-glycerol phosphate synthase; *trp* B codes for the β subunit of tryptophan synthase and *trp* A for the α subunit of tryptophan synthase. $\alpha_2\beta_2$ complex forms active tryptophan synthase.

There is a regulatory gene called as *trp* R whose product cytoplasmic repressor is a dimeric protein having two identical subunits. Each subunit is of the molecular weight nearly 12 500 and has 107 amino acids. As in all cases of enzyme repression, in the absence of tryptophan (corepressor), cytoplasmic repressor does not bind with the operator region. In the presence of tryptophan, there is formation of cytoplasmic repressor–corepressor complex that binds with the operator region. The operator region is partially overlapped with the promoter region. The points contacted by the cytoplasmic repressor lie symmetrically and occupy the region from positions −23 to −3. The operator has the region of *dyad* symmetry, which also includes the consensus sequence of the promoter at −10. As a result, RNA polymerase is not capable of binding with the promoter region repressing the transcription of the structural genes. It is clear that the different needs of induction and repression are accomplished in almost similar manner. The difference is that the effector molecule modulates the operator binding specificity of the cytoplasmic repressor differently.

In case of tryptophan operon, deprivation of tryptophan results in approximately 70-fold increase in the frequency of initiation events at the tryptophan promoter. Even under the repressing conditions, transcription of the structural genes remains continued at a low level. In case of *lac* operon, the basal level of synthesis is only about one thousandth of the induced level. It indicates that the efficiency of repression in tryptophan operon is much lower than that seen in the *lac* operon.

1.6 Attenuation: the Leader Sequence

In the regulation of amino acid operons, generally the end product (amino acid) acts as a corepressor for repressing the transcription of the structural genes. On the basis of mechanism of enzyme repression, it was thought that regulatory gene deleted operon or operon having mutant regulatory gene, should not be under the transcriptional repression. However, in already derepressed *trp* R^- mutants, tryptophan synthesis could be stimulated by deprival of tryptophan and also by the internal deletion of the region between the operator and first structural gene. On the basis of these findings, a second mechanism of regulation involving variable, premature termination of transcription in this region was elucidated. This process is called as *attenuation*. On analysis of the sequence of early mRNA of the tryptophan operon, a part of this sequence was found to code for a short leader peptide. Variation in the translation of leader peptide is dependent on the supply of tryptophan. Tryptophan influences the frequency of termination of transcription at the attenuator site, which lies still further ahead.

Attenuation controls the ability of RNA polymerase to read through an attenuator, an intrinsic terminator located at the beginning of a transcription unit. The common feature of attenuator systems from different operons is that some external event controls the formation of the hairpin needed for intrinsic termination. If the hairpin is allowed to form, termination prevents RNA polymerase to transcribe the structural genes. If the hairpin is prevented from forming, RNA polymerase elongates through the terminator, and the structural genes are expressed. Control by attenuation requires a precise timing of events that control termination. Translation of the leader peptide must occur at the same time when RNA polymerase approaches the terminator site. The RNA polymerase remains paused till translation

of leader peptide on the ribosome occurs. Thereafter, the RNA polymerase is released and moves toward the attenuation site. By providing a mechanism to sense the inadequacy of the supply of Trp-tRNA, attenuation responds directly to the need of the cell for tryptophan in protein biosynthesis and uses attenuation as a control mechanism.

In the tryptophan operon, the attenuator lies within the transcribed leader sequence of 162 nucleotides that precedes the initiation codon of *trp* E. It has *rho* independent termination site and is a barrier for transcription. A short G C rich palindrome sequence is followed by eight successive U residues. RNA polymerase terminates at this site producing only a 140 nucleotide mRNA. The leader region sequence contains ribosome binding site and has an open reading frame for coding a peptide of 14 amino acids called as leader peptide. The leader peptide is unstable. The sequence of the leader peptide is as follows:

Met-Lys-Ala-Ile-Phe-Val-Leu-Lys-

Gly-Trp-Trp-Arg-Thr-Ser

As is clear from the sequence of the leader peptide that out of 14 amino acid residues, two are tryptophanyl residues. Tryptophan is considered to be a rare amino acid in proteins; therefore its abundance in leader peptide has some significance. When the amount of tryptophan in the cell is deficient, biosynthesis of the leader peptide on the ribosome gets stopped when it reaches the *trp* codons. The sequence of the mRNA suggests that this ribosome stalling may in turn influence termination at the attenuator. Pairing of the regions generates the hairpin that precedes the oligo U sequence, a termination signal for transcription. The position of ribosome can determine which structure is formed. When tryptophan is deficient in the cell, ribosomes stall at the trp codons, which are part of region 1. This region 1 gets sequestered within the ribosome and cannot base-pair with region 2. Under the conditions, region 2 will base-pair with region 3 compelling region 4 to remain in a single-stranded form. In the absence of terminator hairpin, RNA polymerase continues transcription after the attenuator. When tryptophan is present in the cell, biosynthesis of the leader peptide occurs through the *trp* codons. Synthesis remains continued along the leader section of the mRNA to the UGA codon which lies between regions 1 and 2. Under the conditions, ribosomes extend over region 2 and prevent it from base-pairing with region 3. Now, region 3 remains available to base-pair with region 4 generating a termination hairpin. This results in the termination of transcription at the attenuator (Fig. 4).

Regulation by attenuation mechanism has been found in many amino acid operons of *E. coli*, for example, *His*, *Phe*, *Leu*, *Thr*, *ilv*.

1.7 Positive and Negative Control

Positive and negative control systems are distinguished on the basis of mode of action of the cytoplasmic repressor. Genes under the negative control are unable to get transcribed in the presence of the product of the regulatory gene (cytoplasmic repressor) but are transcribed in the absence of cytoplasmic repressor. It indicates that cytoplasmic repressor switches off the transcription. The cytoplasmic repressor either binds to DNA to prevent RNA polymerase from initiating transcription or binds to mRNA to prevent a ribosome from

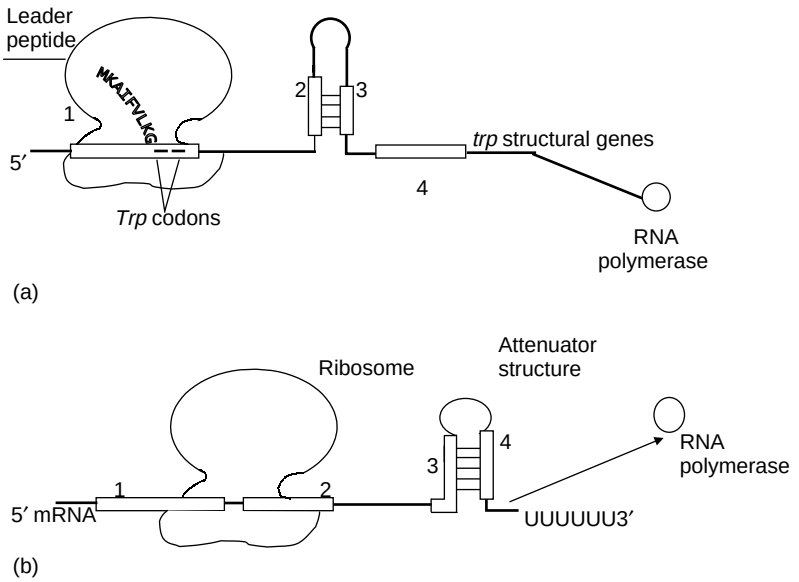

Fig. 4 Ribosomal stalling. (a) When tryptophan levels are low, biosynthesis of the leader peptide on the ribosome gets paused at the *trp* codons in region 1. The region 1 sequence gets sequestered within the ribosome and cannot base-pair with region 2. Therefore, region 2 base-pairs with region 3, compelling region 4 to remain in a single-stranded form. (b) When tryptophan levels are high, biosynthesis of leader peptide occurs through *trp* codons. Synthesis remains continued along the leader section of the mRNA to the UGA codon present between regions 1 and 2. The ribosome extends over region 2 and prevents it from base-pairing with region 3. Now region 3 base-pairs with region 4 generating a termination hairpin.

initiating translation. In fact, negative control provides a fail-safe mechanism. The *lac* operon described above is an example of negative control.

The tryptophan operon described above is another example of negative control. The level of tryptophan in the cell regulates both the activity and synthesis of the tryptophan synthesizing enzymes. Tryptophan inhibits the catalytic activity of the first enzyme of the pathway. Thus, tryptophan inhibits the synthesis of further molecules of the amino acid by inhibiting the beginning of the pathway. Tryptophan also acts as corepressor that activates the product of *trp* R gene. When tryptophan is present, tryptophan operon gets repressed by binding cytoplasmic repressor (product of *trp* R gene)-tryptophan complex to the operator region.

Genes under the positive control are expressed only when an active regulatory protein is present. The regulatory protein acts to switch on transcription and is thus an activator protein. Such activator proteins are also called as *positive control factors*. The regulatory protein interacts with DNA and with RNA polymerase to assist the initiation. A positive control factor that responds to a small molecule is called an *activator*. However, the activator alone cannot bind to the operon. It requires another molecule to be bound to the activator protein, which increases the

DNA-binding ability. An example is cyclic AMP activated cyclic AMP receptor protein (CRP) which activates arabinose operon. You will find in the arabinose operon, that it is an example of both negative and positive control.

1.8 The Arabinose Operon (*Ara* Operon)

The arabinose operon in *E. coli* consists of three structural genes that code for the enzymes involved in the utilization of arabinose. *E. coli* can utilize arabinose as a carbon source. This operon is an example of both positive and negative control. A line diagram of *ara* operon showing structural genes, regulatory gene, and regulatory elements is shown in Fig. 5. The *ara* C is the regulatory gene whose product is called Ara C protein. Biosynthesis of Ara C protein is self-regulatory. Ara C protein regulates its own biosynthesis after binding with *ara* O_1 operator. It represses its *ara* C gene transcription. Generally, about 40 copies of Ara C protein are present in the cell. The Ara C protein acts as positive and negative regulator for the transcription of structural genes, *ara* B, *ara* A, and *ara* D which code for L-ribulose kinase, L-arabinose isomerase and L-ribulose-5-phosphate epimerase, respectively. Some regulatory DNA sequences exert their effect from a distance. These sequences are not always contiguous with promoters. These distant DNA sequences come close together by DNA looping mediated by specific protein–protein and protein-DNA interactions.

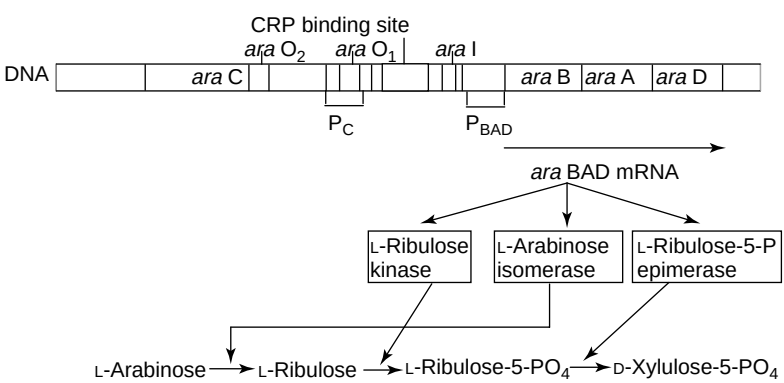

Fig. 5 Line diagram of arabinose operon: *ara* C, regulatory gene; *ara* O_2, *ara* O_1 and *ara* I, regulatory elements where *ara* C gene product may bind; Pc, promoter for *ara* C gene; P_{BAD}, promoter for BAD genes; *ara* B, *ara* A and *ara* D, structural genes for L-ribulose kinase, L-arabinose isomerase, and L-ribulose-5-phosphate epimerase, respectively. Ara C protein regulates its own synthesis after binding on *ara* O_1 resulting in the repression of transcription of *ara* C gene. Ara C protein acts as positive as well as negative regulator for *ara* BAD genes. If arabinose is absent and glucose present, ara C protein binds with *ara* O_2 as well as *ara* I forming a DNA loop and there is repression of *ara* BAD genes. If arabinose is present, cAMP–CRP becomes abundant and binds on the site adjacent to the *ara* I site (CRP binding site). Arabinose also binds to the ara C protein altering its conformation, DNA loop is opened and ara C protein bound on *ara* I acts as activator for transcription of *ara* BAD genes.

When glucose is present and arabinose is not present, Ara C protein binds to *ara* O_2 and *ara* I both, forming a DNA loop of about 210 nucleotides. In this condition, there is repression of the transcription of the structural genes. If glucose is not present and arabinose present, cyclic AMP and CRP become abundant and complex of cyclic AMP and CRP binds to its site adjacent to *ara* I. Arabinose binds with Ara C protein changing its conformation. Now after the binding of arabinose to Ara C protein, DNA loop is opened and Ara C protein bound on *ara* I acts as activator and in concurrence with cyclic AMP–CRP complex, induce transcription of the structural genes. If arabinose and glucose both are present, then also there is repression of transcription, perhaps due to catabolite repression concentration of cyclic AMP is lowered.

1.9 Catabolite Repression

Glucose is the most easily utilizable sugar for energy purpose, therefore, is preferred by *E. coli* as a carbon source. If *E. coli* cells are grown in a medium containing both glucose and lactose, there is no induction of *lac* operon. After utilization of glucose, cells start utilizing lactose and now there is induction of *lac* operon for the biosynthesis of β-galactosidase. In the presence of glucose, the lactose operon is not induced. The inhibitory molecule is not glucose itself but some unknown catabolite derived from glucose which acts by preventing expression of several operons, including lactose, galactose and arabinose operons. This effect is called as *catabolite repression*. It represents a general coordinating system that exercises a preference for glucose by inhibiting the expression of the operons that code for the enzymes of alternative metabolic pathways.

Catabolite repression does not work by preventing the entry of sugars but by inhibition of transcription. In the presence of glucose, there is much decrease in the concentration of cyclic AMP in the cell. The unknown catabolite derived from glucose stimulates phosphodiesterase thereby, converts cyclic AMP into 5′ AMP. There is no decrease in the activity of adenylyl cyclase enzyme responsible for converting ATP into cyclic AMP. Cyclic AMP binds with CRP and the complex stimulates transcription of catabolite sensitive operons. CRP is also called as *catabolite gene activator protein* (CAP). It is a positive control factor whose presence is essential to initiate transcription at CAP dependent promoters. CRP without cyclic AMP is unable to stimulate transcription. Cyclic AMP–CRP binding to DNA changes its structure in some way that assists RNA polymerase to bind.

1.10 Cyclic AMP Receptor Protein

Cyclic AMP plays an important role in controlling the catabolic activity of both prokaryotic and eukaryotic cells. In prokaryotes, cyclic AMP modulates transcription through CRP, also known as CAP, whereas, in eukaryotes, cyclic AMP modulates the enzyme activity by covalent modulation through cyclic AMP dependent protein kinases.

CRP does not bind to DNA without prior binding of cyclic AMP. Among the genes that are activated in bacteria in response to an increase in cyclic AMP are that encode the enzymes for the catabolism of lactose, arabinose, galactose, maltose, and so on. Presence of cyclic AMP is necessary for activation of transcription in bacteria.

This has been demonstrated by mutating the gene coding for adenylyl cyclase (*cya*$^-$) which converts ATP into cyclic AMP. If cyclic AMP is added externally in such system, then there is activation of the transcription. Promoters of the operons whose expression depends on cyclic AMP and CRP contain specific sites for binding cyclic AMP–CRP complex. *In vitro* transcription of DNA fragments containing cyclic AMP–CRP dependent promoters is also activated by cyclic AMP–CRP. In some cases, mutant promoters have been isolated at which cyclic AMP–CRP is unable to bind. At these promoters, cyclic AMP–CRP fails to activate transcription both in intact cells and *in vitro*.

It has been found that one CRP dimer (after binding two cyclic AMP molecules) binds at the specific site in the operon whose transcription is activated by cyclic AMP–CRP. With the help of CRP protection experiments against chewing by deoxynuclease, it has been found that approximately 25 bp are protected. The bases protected by cyclic AMP–CRP from chemical attack and the mutations that prevent CRP binding are located within these sequences. The experiments indicated that CRP makes major contacts in two successive grooves of the DNA. The most-conserved sequence that binds CRP is 5′TGTGA3′. All the evidences indicate that 5′TGTGA3′ sequence is critical for CRP binding. At the *gal* and *lac* sites, point mutations known to prevent stable CRP binding are located in this sequence. At the *ara* site, deletion experiments showed the importance of this sequence for CRP binding. There is an another sequence 6 bp downstream of the TGTGA motif and has inverted repeat. Although in many cases, it is not exactly inverted repeat sequence. This second motif irrespective of the symmetry of the sequence has also been shown to be required for efficient CRP binding. Many types of experiments indicated that the two subunits of CRP recognize two zones of sequences separated by 6 bp. As written above, one of these zones contains the sequence 5′TGTGA3′. The other stretch contains either a symmetrically arranged second version of this sequence or another type of sequence. However, affinity of CRP for DNA appears to be more when 6-bp downstream sequence of 5′TGTGA3′ is symmetrical compared to when it is not symmetrical. The distance between the transcription start point and the CRP binding site is different at different promoters. At some promoters like those for *ara* operon and *mal* operon, an additional protein, the Ara C protein or the Mal T protein is also required for activation of the transcription. These activator proteins also bind to the promoter. In some cases (for example *lac*, *cat*), two CRP binding sites have been found. The secondary sites bind CRP less tightly. It is considered that these secondary sites might help in the search of CRP for the primary sites.

Although, the main function of CRP is to activate transcription, in some cases, binding of CRP has been shown to repress transcription. At the *gal* promoter, there are two promoters, P1 and P2. When CRP binds at the P1, it activates transcription. However, when CRP binds at P2, it represses transcription. This is due to the fact that at P2, CRP binds near the −35 region blocking the binding of RNA polymerase at the P2 promoter. CRP also acts as a repressor of transcription of its own promoter *in vitro*. CRP also inhibits transcription of the gene for the major outer membrane protein, Omp A, again by binding near the −35 region of the promoter.

CRP is a dimeric protein having two identical subunits, each containing 210

amino acids. Complete amino acids sequence is known as deduced from the nucleotide sequence of the gene. Equilibrium dialysis studies indicated that two molecules of cyclic AMP could bind per CRP dimer. CRP has a two-domain structure as indicated by the resolution crystal structure of CRP complexed with cyclic AMP. A large amino terminal domain extending from residue 1 to 135 is separated by a cleft from a smaller carboxy terminal domain that extends from residue 136 to 210. The amino terminal domain of each subunit contains one cyclic AMP molecule buried in the interior of the protein. Residues from both subunits are involved in the binding of cyclic AMP. The 6-amino group of the adenine ring in cyclic AMP interacts with Thr^{127} on one subunit and Ser^{128} on the other subunit.

Amino terminal domain of CRP in the region of 30 to 89 residue, exhibits sequence homology with the regulatory subunit of the protein kinase of eukaryotes. Regulatory subunit of protein kinase also has two cyclic AMP binding sites.

The carboxyl terminal domain of each CRP subunit consists of three α-helices connected by short β-sheet structures. In each subunit, one of the α-helices protrudes from the surface of the CRP dimer. These two α-helices are considered to be involved in DNA binding. The other DNA binding proteins like Cro and cI proteins also have α-helices but in the amino terminal region. It has been shown that all the DNA-binding proteins have helix turn helix domain essential for interactions with DNA. However, *E. coli* fnr protein, which is essential for anaerobic respiratory metabolism, also has helix turn helix domain in carboxyl terminal region. Additional homology is also found in the amino terminal regions of the two proteins. Although fnr protein does not bind cyclic AMP, it has a somewhat similar function as CRP. The fnr protein is a pleiotropic activator for a series of genes that are turned on under limiting aerobic conditions. The sequence comparison indicated that *fnr* gene might have derived by duplication either from the *CRP* gene itself or from a common ancestor.

Interactions between RNA polymerase and promoters may be described as a two-step event. First, the enzyme binds to the promoter to form a closed complex. This binding step is reversible and is characterized by an association constant, K_B. In the second step, the closed complex isomerizes to give rise an open complex. This isomerization includes a localized unwinding of the DNA over a distance of approximately 12 bp near the transcription start is generally irreversible and the corresponding rate constant, K_f, is slow. Strong promoters have high values of both K_B and K_f, whereas, weak promoters have low values for both the constants. Addition of cyclic AMP and CRP has two effects on *lac* promoter. First, it enhances the rate of open complex formation by increasing the value of K_B without affecting the value of K_f. Second, the presence of cyclic AMP–CRP also increases the binding of RNA polymerase on P1 promoter. This increased binding of RNA polymerase on P1 is due to inhibition of RNA polymerase binding on other secondary sites in the presence of cyclic AMP–CRP. The structure of the CRP-DNA complex is interesting. The DNA has a bend. Proteins may distort the double helical structure of DNA when they bind and several regulatory proteins induce a bend in the axis. Therefore, there is a dramatic change in the organization of the DNA double helix after the binding of CRP.

1.11 Guanosine-5′-Diphosphate, 3′-Diphosphate

The *rel* A gene required for the synthesis of guanosine-5′-diphosphate, 3′-diphosphate (ppGpp) has been shown to enhance the transcription of *lac* Z gene and *glg* genes (genes coding for glycogen biosynthetic enzymes). There are indications that ppGpp interacts directly with RNA polymerase to directly alter the transcription of various genes. A small protein has been shown to mediate the effect of ppGpp on *lac* Z gene under certain conditions. Nitrogen regulatory proteins C and A (Ntr C and Ntr A) are also shown to activate *gln* promoter. The *ntr* C and *ntr* A genes encode a specific DNA-binding protein and an alternate sigma factor for RNA polymerase, respectively. However, Ntr C and Ntr A did not increase the synthesis of glycogen biosynthetic enzymes.

2 Regulation of Gene Expression in Eukaryotes

Compared to prokaryotes, eukaryotes have more complex regulatory mechanism of transcription. In addition, RNA splicing also plays important role in the regulation of gene expression. Besides, activation of gene structure, polyadenylation of mRNA, capping of mRNA, transport of mRNA to cytoplasm, translation of mRNA are the potent control points in the process of regulation of gene expression. Figure 6 shows five potential control points for regulation of gene expression in eukaryotes. The most important way to control gene expression is by regulating the initiation of transcription (interaction of RNA polymerase with the promoter region). The regulation of initiation of transcription may be demonstrated by a technique called as *run off transcription*. In this technique, nuclei are isolated from the cells and are incubated with radiolabeled nucleoside triphosphates. Under suitable conditions, unfinished transcripts are completed, but no new transcripts are synthesized. Therefore, the RNA labeled by this method is derived from genes that have started transcription at the time of isolation of the nuclei. The labeled RNA is used to probe DNA from a clone of the genes under investigation. The absence of any hybridization between the labeled RNA and the cloned DNA indicates that the DNA is not transcribed in the tissue. This technique has been used to examine

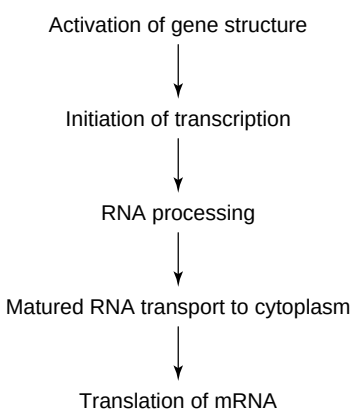

Fig. 6 Regulation of gene expression. Gene regulation may take place in a gene-specific way at any of the several sequential steps. However, there are five potential control steps.

a number of genes. The absence of gene expression does, indeed, result from the absence of transcription.

Now a days, *DNA microarray technology*, more commonly known as "*gene chip technology*" is also being used to examine the presence of complementary sequences. In fact, microarray technology is a genetic revolution of today and may be compared with the microprocessors to the computer revolution 25 years ago. The microarray technology has changed the job of screening the genetic information into an automatic routine that exploits the tendency for the molecule that carries the template for synthesizing mRNA, protein, to bind to the DNA that produces it. The microarrays contain thousands of probes, each imbued with a different nucleic acid from known and unknown genes to bind with mRNA. The resulting bonded molecules fluoresce under different colors of laser light showing which complementary sequence is present. The microarrays can measure the incidence of genes and their expression. After the human genome sequence availability, importance of single nucleotide polymorphism (SNP) has also been appreciated. The SNP is minor variations in DNA that define differences among people, predispose a person to disease, and influence a patient's response to a drug. With the genetic makeup of human broadly known, it is now possible to build microarrays that can target individual SNP variations as well as making deeper comparisons across the genome. All these studies may help in finding the roots of many diseases. Nowadays, specific software are also available to design microarrays with much number of probes.

Real-time Polymerase Chain Reaction (RT PCR) has also been used to measure the level of gene expression. It is a highly sensitive and convenient technology and its approaches are a natural complement of transcriptone analysis, either when tuning of array results is necessary or when array sensitivity limit is reached for low-level transcripts of interest. In RT PCR, sensitive quantification of PCR products relies on detection of a fluorescent signal proportional to the amount of product. PCR products can be measured in real time by the use of a double-stranded DNA-binding dye which does not bind with single-stranded DNA, or labeled oligonucleotides that specifically bind to the PCR products.

Cells of multicellular organisms are genetically homogeneous but structurally and functionally heterogeneous because of differential expression of genes. Many of these differences in gene expression occur during development and are retained through mitosis. Stable alterations of this type are called to be *epigenetic*, because they are heritable in the short term but do not involve mutations of the DNA itself. Researches have focused on two molecular mechanisms that mediate epigenetic phenomena namely DNA methylation and histone modifications. It has been shown that epigenetic effects by means of DNA methylation have an important role in development but can also arise stochastically as animal age. Identification of proteins that mediate these effects has helped in elucidating the epigenetic effects. When it gets perturbed, diseases may occur. External factors on epigenetic processes are found in the diet on long-term diseases like cancer. It has been indicated that epigenetic mechanisms allow an organism to respond to the environment through changes in gene expression.

There are many genes that are transcribed in one tissue or organ but not in others. In fact, this explains the cell differentiation in eukaryotes. Some genes

are expressed under the influence of certain signaling agents such as substrates of specific enzymes, hormones, regulatory nucleotides, and so on. Expression under the influence of certain signaling agents has also been considered as *induction*. Although the phenomena of induction is not that much prominent in eukaryotes as in prokaryotes. In eukaryotes, it takes more time and the amount stimulated is only about 10 to 20 times in contrast to bacteria where stimulation to the extent of thousands times has been found as a result of induction. In eukaryotes, monocistronic mRNA are found in contrast to polycistronic RNAs in prokaryotes, therefore, coordinate induction has also not been reported in eukaryotes. Longtime back, a model named as Britten Davidson model was proposed for explaining induction phenomena in eukaryotic genes. According to this model, eukaryotic genome contains a large number of sensor sites that recognize specific molecular signaling agents such as hormones, substrates of specific enzymes, and so on. Each sensor site is adjacent to an **integrator gene**. When sensor site is activated after binding of a signaling agent, the integrator gene is transcribed to form its complementary RNA called as *activator RNA*. The activator RNA in turn is recognized by one or more receptor sites located elsewhere in the genome. The receptor sites may be on the same or some other chromosome. When activator RNA binds on the receptor site, an adjacent structural gene is transcribed (Fig. 7).

Although, the phenomena of coordinate expression in eukaryotes was considered to be nonexistent owing to monocistronic mRNAs, using DNA microarrays, global expression analysis showed highly coordinate expression of genes functioning in common processes in eukaryotes. It has been called as "*syn-expression*." The syn-expression in eukaryotes has been considered comparable to operons in prokaryotes. It is also proposed that it may be a key determinant facilitating evolutionary change leading to animal diversity. Using DNA microarrays, simultaneous monitoring of thousands of transcripts is possible. These studies gave global insights into gene expression. The expression data revealed a high-degree of order in the genetic program, a tight coordination of the expression of groups of genes functioning in a common process.

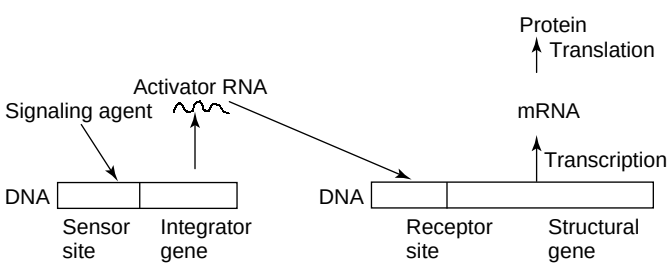

Fig. 7 Britten Davidson model of gene regulation: when a signaling agent such as hormone binds on sensor site, there is transcription of integrator gene forming its complementary RNA called as activator RNA. The activator RNA is recognized by the receptor site located elsewhere in the genome. When activator RNA binds on the receptor site, adjacent structural gene is transcribed and afterward translated.

Frenster while working with mammalian cells infected with SV40 virus suggested a derepressor model for gene regulation in eukaryotes. On the basis of experimental data indicating the ability of exogenous DNA or RNA to derepress specific loci on the host cellular genome, the model suggested a close relationship to the normal mechanisms of gene regulation in animal cells, which may be subverted to allow reexpression of otherwise repressed embryonic information. The derepression model accounts for selective gene transcription that is locus and strand specific but does not discuss gene–gene interactions. Afterward, Frenster proposed a Mated Model of gene regulation in eukaryotes.

According to the Mated Model, a derepressor RNA (dRNA) binds to anticoding strand of operator locus permitting transcription of operator and structural gene loci. The derepressor RNA of an operon is complementary in base sequences to its operator portion of the direct transcription product. Afterward, direct transcription product gets split into mRNA and operator RNA (oRNA). The cleavage may occur directly or after formation of heterometric duplex RNA by base-pairing of derepressor RNA with the operator portion. Thereafter, derepressor RNA is selectively removed from the operator locus providing feedback inhibition of transcription of the operon. With consumption of mRNA and with degradation of operator DNA, derepressor RNA is released from the duplex providing positive feedback derepression of transcription of the operon. Different structural genes may share operators with common base sequences, and thereby are equally sensitive to given species of derepressor RNA both during gene transcription and during its selective inhibition (Fig. 8).

In eukaryotes, normally cell division is highly regulated with the help of growth factors that cause the cells to undergo cell division and in some cases, cell differentiation. Some growth factors are specific for certain type of cells due to specific receptors present at the surface of the cell. Some growth factors are general rather than specific in their effects. Some growth factors controlling the cell division are epidermal growth factor, nerve growth factor, platelet derived growth factor, fibroblast growth factor, lymphokines, and so on. In fact, sometimes due to failure of control of the cell division by growth factors, there may be uncontrolled cell division causing tumors.

In prokaroyotes, negative control of transcription plays an important role (fail safe mechanism). However, in eukaryotes, positive control plays a more important role. It is due to the simple reason that in a large genome, it is more efficient. If so many genes have to be negatively controlled, each cell would have to synthesize the same number of different repressors in sufficient amount to permit specific binding of each. Besides, nonspecific DNA binding of regulatory proteins (repressors) is also an important problem in much larger genomes of higher eukaryotes. The chance of specific-binding sequence to be present randomly at inappropriate sites also increases with genome size.

2.1 Transcriptionally Active Chromatin

The major part of the eukaryotic genome is sequestered in the nucleus where it is surrounded by a nuclear membrane, safeguarding it from exposure to the cytoplasm. Transcription of the genes also occurs within the nucleus. However,

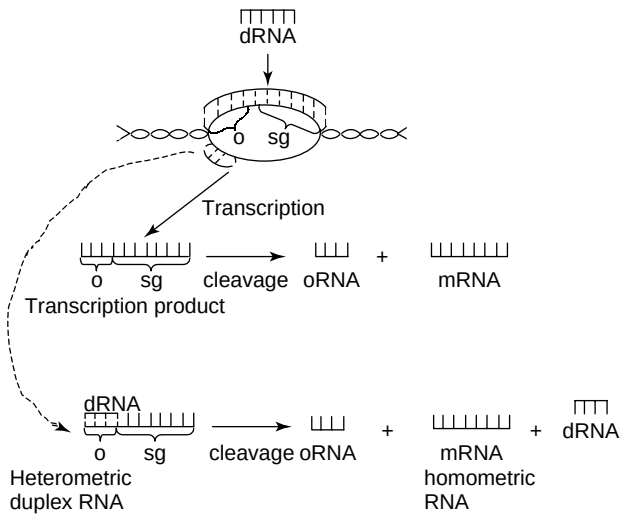

Fig. 8 Mated model for gene regulation: o, operator; sg, structural gene; dRNA, derepressor RNA. Derepressor RNA (dRNA) binds to anticoding strand of operator locus permitting transcription of operator and structural gene loci. The transcription product gets split into mRNA and operator RNA (oRNA). The cleavage may occur directly or after formation of heterometric duplex RNA.

translation occurs mainly in the cytoplasm. Therefore, coupling of transcription and translation cannot occur. The chromosomes of eukaryotes are more complex than of the bacteria. Each eukaryotic chromosome contains a double helix DNA molecule that may be more than twenty times larger than that of bacterial chromosomes. The DNA is tightly complexed with histone proteins that are considered to have structural and protective functions. Additionally, there are some loosely bound nonhistone proteins. These are mostly present in smaller quantities than histones. The functions of nonhistone proteins are not exactly known, these may have role(s) in transcription and/or in replication. Mitochondrial and chloroplast DNAs are small double-stranded molecules. Plant mitochondrial DNA is larger than that of animals. All plants seem to have similarly sized chloroplast DNA. Mitochondrial and chloroplast DNAs resemble bacterial DNA and unlike eukaryotic nuclear DNA are not associated with histones.

It has been seen that chromosomal regions that have been activated for transcription are more sensitive to DNase degradation indicating lesser protection by histones. It has also been found that actively transcribed regions have sequences that have high sensitivity to DNase, and these have been called as *hypersensitive sites*. These hypersensitive sites are generally not more than 200-bp long and are generally found with in the 1000 bp flanking the 5′ ends of transcribed genes. In some cases, these hypersensitive sites are found farther from the 5′ end, near the 3′ end or within the gene itself. Many hypersensitive sites have

been found to be the binding sites for regulatory proteins.

Telomeres are the specialized structures at the ends of linear eukaryotic chromosomes. They generally have many tandem copies of a short oligonucleotide sequence, T_aG_b in one strand and C_bA_a in the complementary strand where a and b are 1 to 4. The structure of telomere poses a biological problem. DNA replication requires a primer but in linear DNA molecules it is impossible to synthesize an RNA primer starting at the end nucleotide. The problem is solved by telomerase, an enzyme resembling reverse transcriptase. It adds telomeres to chromosome ends. Telomerase has both RNA and protein in its structure. The RNA part is about 150 nucleotides long and has about 1.5 copies of C_bA_a telomere repeat. This telomere repeat part acts as a template for synthesis of the T_aG_b strand of the telomere. Telomerase like reverse transcriptase synthesizes only a segment of DNA that is complementary to an internal RNA molecule.

The DNA in transcriptionally active chromatin has been found to be less methylated. Besides, nucleosomes have also not been found in transcriptionally active regions (at least in some cases). Chromatin has been classified into two groups – *heterochromatin*, a highly condensed chromatin which is transcriptionally inert, and *euchromatin*, a loosely packed chromatin and transcriptionally active.

2.2 Regulation of Gene Expression at the Initiation of Transcription

Regulation of gene expression at the initiation stage of transcription (binding of RNA polymerase onto the promoter) has been evidenced. At present, there is no evidence for control of gene expression at the subsequent stages of transcription. Three RNA polymerases have been found in eukaryotic cells for biosynthesis of different classes of RNAs. Biosynthesis of HnRNA takes place in the presence of RNA polymerase II. Initiation of transcription by RNA polymerase II is regulated by a series of DNA elements that may be divided into the core promoter element consisting of TATA box, the transcription initiation site and the upstream activating sequences (UAS). The UAS are usually found upstream of the core promoter sequence but in some cases, have been found downstream of the transcription start site (Fig. 9). A specific protein gets bound on each UAS resulting in a positive or negative effect on the core promoter activity. TATA box is found generally 25-bp upstream from the transcription initiation site. Although TATA box is common in eukaryotic genes, few genes have been shown to get expressed without TATA box. Besides, TATA box, another sequence called as CAAT box has been found at the −75 position from the initiation site. It functions in either orientation and plays an important role in increasing the promoter strength. It has the consensus sequence GGCAATCT. A GC box at −90 position from the initiation site has also been found. It occurs in either orientation and is a common component of the promoters of house-keeping genes. The GC box has the consensus sequence GGGCCGGG.

Transcription factor D for RNA polymerase II (TFIID) has been shown to play an important role in the initiation of transcription by binding on TATA box sequences. After binding of TFIID, it facilitates the binding of the RNA polymerase II on to the promoter. The assembly of an initiation complex and RNA polymerase

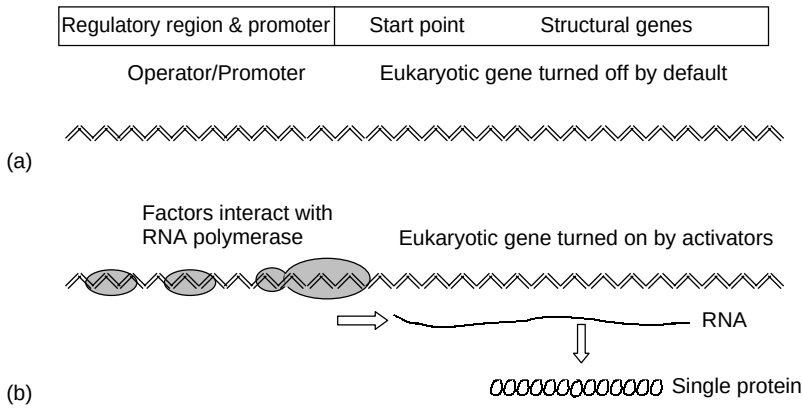

Fig. 9 Regulation of eukaryotic gene expression at transcriptional level. Eukaryotic control is often positive: trans-acting factors bind to cis-acting sites in order for RNA polymerase to initiate transcription at the promoter. (a) Transcription gets turned off by default (if proper initiation factors are not bound in the regulatory region). (b) Transcription gets turned on after the binding of initiation factors on the regulatory region, after which RNA polymerase binds on the promoter region.

at the promoter is a complex process that requires participation of many other initiation factors. TFIID has two components, TATA box binding protein (TBP) and another protein called as TBP associated factor (TAF). TBP associated factor is also important for regulating the transcription. TBP, also called as the "commitment factor," binds to DNA in the minor groove. The inner surface of TBP binds to DNA and the outer surface is available to extend contacts to other proteins. The DNA binding sites consist of sequences that are conserved between species. The variable N-terminal tail is exposed to interact with other proteins. TBP is the only transcription factor that makes contacts with the specific sequences in DNA. The activity of TFIID has also been shown to be regulated by inhibitory proteins that interact with TBP. Such inhibitory proteins may serve an important regulatory role by keeping genes that have been removed from inactive chromatin in a repressed but rapidly inducible state.

There are other factors too that play an important role in the regulation of initiation of transcription. Transcription Factor II A (TFIIA) joins the initiation complex after TFIID. It has two subunits in yeast and three in mammals. After joining of TFIIA, TFIID becomes able to protect a region extending further upstream. Addition of Transcription Factor II B (TFIIB) gives further protection to the region of the template in the vicinity of the start point from −1 to +10 bp. The TFIIF is a dimeric protein. The larger subunit has ATP-dependent DNA helicase activity that could be involved in opening the DNA during initiation. The smaller subunit has been found to be equivalent of sigma factor of *E. coli*. The TFIIF brings RNA polymerase II to the assembling transcription complex and provides the means for its binding. Interaction with TFIIB may be important when TFIIF joins the complex. Polymerase binding extends the sites that are protected downstream to +15 on the template strand and +20

on the nontemplate strand. TFIIE binds at the upstream boundary. Two more factors, TFIIH and TFIIJ also join the complex after TFIIE. TFIIH has kinase activity that may phosphorylate the CTD (carboxy terminal domain) tail of RNA polymerase II, which consists of multiple repeats of the consensus sequence, Tyr-Ser-Pro-Thr-Ser-Pro-Ser. This sequence is unique to RNA polymerase II. Phosphorylation of the tail (at seryl or threonyl residue) is required to release the enzyme from the transcription factors, so that it can leave promoter region to start elongation. The TATA box determines the location of the start point. The general initiation process of transcription is the same as in bacteria. The enzyme RNA polymerase generates a closed complex and afterward is converted into an open complex where DNA strands get separated. In the process of open-complex formation, there is removal of TFIIE.

The CAAT box is recognized by the proteins of the CTF (CCAAT-binding transcription factor) family which are generated by the alternative splicing from a single gene. The CP1 factor binds to the CAAT boxes of α-globin and CP2 binds CAAT box in a β-fibrinogen gene. Other proteins also bind CAAT boxes. The ACF (albumin CCAAT factor) protein binds CAAT in the albumin promoter. The CAAT box may also be a regulatory point. In embryonic tissues, a protein called CDP (CAAT displacement protein) binds to the CAAT boxes preventing the transcription factor from recognizing them. In testes, the promoter is bound by transcription factor at the TATA box, CAAT box and the octamer sequences. In embryonic tissue, exclusion of CAAT binding factor from the promoter prevents a transcription complex being assembled. This behavior is analogous to the effect of bacterial repressor.

The GC box is recognized by the factor SP1, a monomeric protein which makes contacts on one strand of the DNA over a −20 bp binding site including GC box. In SV40 promoter, the multiple boxes between −70 and −110 all bind this factor protecting the whole region. However, in thymidine kinase promoter, SP1 interacts with a factor at the CAAT box on one side and with TFIID bound at the TATA box on the other side.

Additional regulatory sequence elements are enhancers in higher eukaryotes and UAS in yeast. For enhancer, the location and orientation of sequences relative to transcription start site are relatively unimportant. The enhancers exert their regulatory effects even when moved experimentally and they may naturally occur thousands of base pairs away from the gene being regulated. The enhancers have no promoter activity of their own but may stimulate transcription over considerable distances. The enhancers may be involved in regulation of gene expression during the development of the organism like the immunoglobulin enhancer which only functions in B-lymphocytes.

Regulatory role of enhancer activity has been found in the transcription of genes responsive to steroid hormones. The steroid gets bound on to a soluble protein, which in turn binds the enhancers for the steroid responsive genes. Transcriptional activation is also accompanied by the decondensation of the chromatin in the regions containing the genes which is evident by the fact that the region becomes more sensitive to DNase I digestion and subsequent binding of transcriptional factors to the promoter regions. An enhancer may also provide an entry site, a point at which RNA polymerase/other essential protein associates with chromatin. It involves the

same type of interaction with the basal apparatus as the interactions promoted by the upstream promoter elements.

UAS are found in yeast, which are analogous to enhancers of higher eukaryotes. The UAS are located upstream of the gene in a region having two identical sequences of 72 bp each repeated at tandem 200-bp upstream of the initiation start site. The −72 bp repeat is located in a hypersensitive site of chromatin.

RNA polymerase I transcribes the genes for ribosomal RNA from a single type of promoter. The promoters for RNA polymerase I have the least diversity in the eukaryotic genome. The promoter has been found located 70-bp downstream of a control element called as UCE (upstream control element). The promoter consists of a bipartite sequence in the region preceding the start site which surrounds the start site extending from the −45 to +20 bp and is able to start the transcription. The UCE that is located at −180 to −107 bp increases the efficiency of the promoter. Both regions are enriched in GC contents. Two initiation factors are required for initiation of transcription by RNA polymerase I. The upstream binding factor 1 (UBF 1) binds in a sequence-specific manner to related sequences in the core promoter and UCE. Another factor called SL1 binds cooperatively to UBF 1 to extend the region of DNA. After binding of both the factors, RNA polymerase is able to initiate transcription after binding with the promoter region. The SL1 is specific species wise, for example, mouse SL1 cannot work on a human DNA. It consists of four proteins including one called TBP required for initiation of transcription by RNA polymerase II and III too. The SL1 has been considered analogous to sigma factor of bacteria.

RNA polymerase III transcribes DNA coding for 5S rRNA, tRNA, and many small nuclear RNAs. The DNAs transcribed by RNA polymerase III are all smaller in size, generally less than 300 nucleotides. The regulation of oocyte 5S rRNA synthesis has been studied. It requires three transcription factors for initiation called as TF IIIA, TF IIIB, TF IIIC. The TF IIIA is a member of zinc finger proteins whereas TF IIIB consists of TBP and two other proteins. The TF IIIC is still used as a partially purified preparation containing 5 subunits. It has been shown that TF IIIA is specific for oocyte 5S rRNA whereas, TF IIIB and TF IIIC are required for transcription of all the DNA by RNA polymerase III.

Promoters recognized by RNA polymerase III are of two types, lying upstream and lying downstream from the initiation site. These promoters are recognized by different initiation factors. The promoters for 5S rRNA and tRNA genes lie downstream to start site, whereas promoters for small nuclear RNAs gene located upstream from the start point. The promoters for 5S rRNA gene are located between −55 to +80 within the gene.

The promoters for RNA polymerase III have bipartite structure. The two short sequences are separated by a variable sequence. Type I promoter consists of box A sequence separated from a box C sequence and type II promoter consists of a box A sequence separated from a box B sequence. In Type II promoters, TF IIIC recognizes box B but binds to a region involving box A as well as box B. In Type I promoters, TF IIIA binds on box C. In both Type I and II promoters, binding of TF IIIC facilitates the binding of TF IIIB to the sequence surrounding the start site. It has been shown that TF IIIB is the main initiation factor for RNA polymerase

III, whereas TF IIIA and TF IIIC assist TF IIIB to bind at the correct site. The efficiency of the transcription by RNA polymerase is found to get increased by the presence of proximal sequence element (PSE). All the transcription factors for RNA polymerase bind at the promoter region forming preinitiation complex before the binding of RNA polymerase III on to the promoter.

3
RNA Splicing

Most eukaryotic genes have been found to have noncoding sequences (introns) in addition to coding sequences (exons). Introns are present in DNA and in the primary transcription product of the gene (HnRNA). The introns are removed by RNA splicing before mature mRNA is transported to the cytoplasm. The number of introns varies between the genes. The *dystrophin* gene has 70 introns whereas, α-interferon gene has no intron. The size of the intron also varies from nearly 100 nucleotides to 200 000 nucleotides.

There are at least four types of reactions of RNA splicing: splicing of nuclear introns, group I and II introns, and tRNA introns. Each reaction carries a change within the individual RNA molecule, therefore, is considered a cis-acting event. In RNA splicing, only very short consensus sequences are required. The required sequences are the end sequences of the intron, GT-AG. In yeast, a branch sequence UACUAAC is also required which is less conserved sequence in mammals. The ends of the introns are identified by RNA–RNA base pairing between the HnRNA and uridine rich small nuclear ribonucleoprotein particles called as snRNPs. Since the conserved splice site sequences are short and not precisely conserved between introns, and occur frequently in the primary sequences of many HnRNAs, this allows the spliceosomes to combine different 5′ and 3′ splice sites in HnRNA to produce several alternatively spliced mRNAs from a single nuclear gene. Therefore, due to the process of alternate splicing, multiple proteins with different primary amino acid sequences and biological activities may be produced from a single gene. The alternatively spliced mRNAs have been shown to be regulated in either at temporal, development, or tissue-specific manner in many cases. An alternatively RNA splice site choice is shown to be used to regulate expression of somatic sex determination pathway in *Drosophila*. The sex lethal and the transformer 1 proteins have been shown to control the maintenance of sex by regulating *Drosophila* gene expression at the level of alternative RNA splicing.

The RNA splicing starts with at 5′ splicing site and formation of a lariat occurs by joining GU end of the intron to the A position of the branch sequence via a 5′,2′ linkage. Afterward, 3′-OH end of the exon attacks the 3′ splicing site in such a way that ligation of exons occurs releasing intron as a lariat. Both reactions involve trans-esterification in which bonds are conserved. At several stages, ATP hydrolysis occurs probably for the conformational changes in the RNA and/or proteins. Lariat formation is responsible for 3′ splicing site. Nuclear splicing requires formation of a spliceosome, which contains various snRNPs and splicing factors. SnRNPs recognize consensus sequences and also share some interacting proteins. The U1, U2 and U5 snRNPs each contain a single snRNA and several proteins, whereas U4/U6 snRNP contains two snRNAs and several proteins. The U1 snRNP base pairs

with the 5′ splicing site, U2 base-pairs with the branch sequence, U5 snRNP acts at the 5′ splicing site. From U4/U6, there is cleavage of U4 and afterward, U6 base-pairs with U2. All this creates the catalytic center for splicing. Group I and Group II introns perform RNA splicing as a self-catalyzed property of RNA. In Group I introns, required for attack at the 5′ exon–intron junction is provided by a free guanine nucleotide whereas in Group II introns, it is provided by an internal 2′-OH position. These introns also follow GT-AG rule, but form a characteristic secondary structure that holds the splice sites in the appropriate position. The tRNA splicing in yeast has been shown to involve separate endonuclease and ligase reactions. The endonuclease recognizes the secondary structure of the pre-tRNA and cleaves both ends of the intron. The two halves-tRNAs released by the removal of introns can be ligated using the enzyme RNA ligase in the presence of ATP.

3.1
Nuclear Splicing

In eukaryotes, majority of genes have introns. Because of the presence of introns (noncoding sequences) in the gene, there is much discrepancy in size between the nuclear genes and their corresponding mRNAs. The average size of nuclear RNA (HnRNA) was found to be much larger and more complex than mRNA. The HnRNP has also been found to be a ribonucleoprotein in which HnRNA is bound by proteins such that it looks like a bead connected by a fiber. The beads are globular-shaped RNA associated with six common proteins, A1, A2, B1, B2, C1, C2, called as *core proteins*. The size of the core proteins ranges from 34 000 to 120 000 daltons. The exact structure of the HnRNP and the function of RNAs packaging in the form of beads are not clear.

The RNA splicing and other posttranscriptional changes occur in the nucleus and the substrate for these processes is HnRNP. The transcript is capped at the 5′ end, introns removed, polyadenylation at the 3′ end. As such all these reactions are called RNA processing. The RNA after processing is transported through the nuclear pores to the cytoplasm where it is available for translation.

There are many types of splicing systems (Fig. 10):

1. Introns are removed from the nuclear RNAs using spliceosome. The reaction requires a large splicing system.
2. Excision of certain introns is an autonomous property of the RNA itself. The ability of RNA to act as an enzyme is seen in the self-cleavage of viroid RNAs and in the catalytic activity of RNase P.
3. The removal of introns in yeast pre-tRNAs involves endonuclease and RNA ligase whose dealings with pre-tRNA seem to resemble those of the RNA-processing enzymes. A critical feature is the conformation of pre-tRNA.

Nuclear RNA splicing junctions are interchangeable but are read in pairs. There is no extensive homology or complementarity between the two ends (5′ GU – AG3′) of an intron. As written above, the junctions have well-conserved consensus sequences. The really high conservation is found only within the introns at the presumed junctions. The 5′ and 3′ end dinucleotides sequence define the left (or 5′) and right (or 3′) splicing sites, and these are also called donor and acceptor sites. It has been shown that there is a common mechanism for nuclear HnRNA splicing.

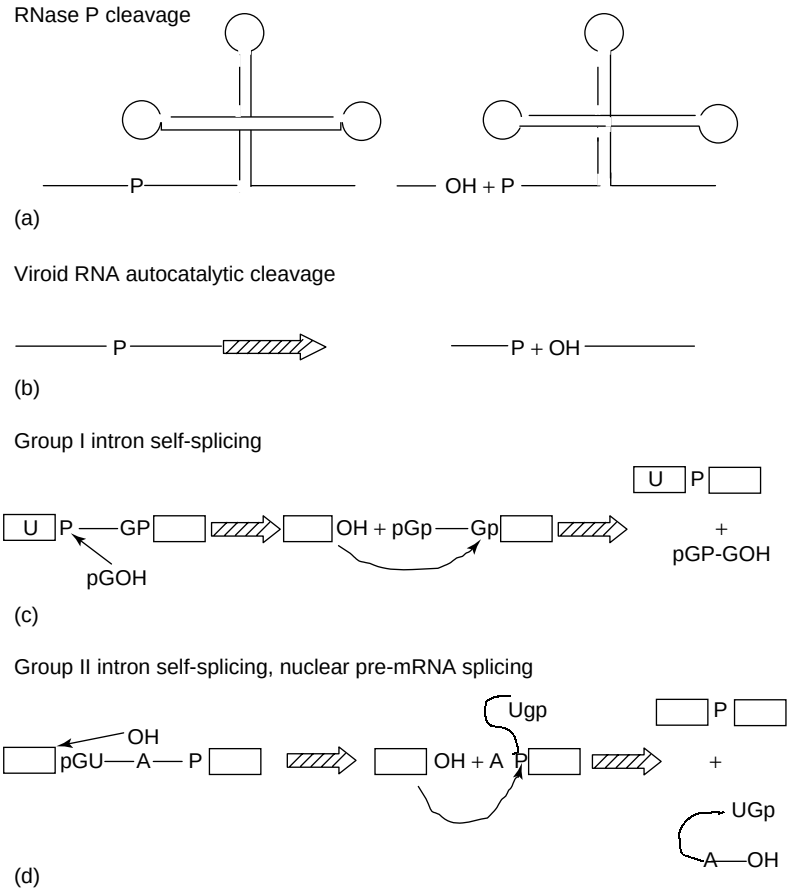

Fig. 10 Different types of RNA catalyzed intron splicing reactions.

However, the consensus is not applied to the introns of mitochondria, chloroplasts and pre-tRNA introns.

To ensure splicing of the correct pairs of junctions, the following two points may be applicable:

1. It could be the intrinsic property of the RNA to connect the sites at the ends of a particular intron, because of the base-pairing involving these regions.
2. All the 5′ sites may be functionally equivalent and all 3′ sites may be similarly indistinguishable. The splicing could follow rules ensuring that 5′ site is always connected to the 3′ site that comes next in the RNA.

The splicing sites are generic; they do not have specific individual RNA precursors. Besides, the apparatus (spliceosome) for splicing is not tissue specific. RNA may be spliced by any cell and the conformation of the RNA influences the accessibility of the splicing sites. The reaction also does not proceed sequentially along the precursor. RNA splicing is also independent of modifications to the RNA. *In vitro*, a cut is first made at the 5′ end of the intron

separating the left exon and the right intron–exon molecule. The left exon takes the form of a linear molecule whereas the right intron–exon is not a linear molecule. The 5′ terminus generated at the left end of the intron becomes linked by a 5′-2′ bond to the A in the branch site located 30 nucleotides upstream of the 3′ end of the intron. This linkage keeps the intron in the form of a structure called as lariat. Afterward, cutting at the 3′ end releases the free intron in a lariat form and right exon gets ligated with the left exon. The lariat is then debranched to give a linear excised intron, which is rapidly degraded.

3.2 Spliceosome and Small Nuclear RNAs

Many small RNAs have been found in the nucleus and cytoplasm of eukaryotic cells, called as small nuclear RNAs (snRNAs) or snurps and small cytoplasmic RNAs (scRNAs) or scrps, respectively. A small nuclear ribonucleoprotein particle (snRNP) generally contains one snRNA and about 10 proteins. Some of the proteins are common in all the snRNPs while some of the proteins are unique to a particular snRNP. The common proteins are recognized by an autoimmune antiserum (anti-Sm) and are considered to be involved in the autoimmune reaction. Many snRNPs are involved in RNA splicing. The snRNAs present in these snRNPs have sequences complementary to the 5′ or 3′ splicing sites or to the branching sequence. It is considered that base-pairing between snRNA and HnRNA or between snRNAs plays an important role in splicing.

The spliceosome consists of many snRNPs and many additional proteins often called as splicing factors (Fig. 11). The snRNPs are U1, U2, U5, U4/U6 named according to the snRNPs present in the spliceosome. The snRNPs present in spliceosome together have about 40 proteins. Some of these proteins may be directly involved in splicing and others may have structural roles for assembly or for interaction with the snRNPs.

In U1 snRNP, a region of 5′ terminal 11 nucleotides that is single-stranded and has a stretch complementary to the consensus sequence at the 3′ site of the exon is considered to be directly involved in splicing. The intact U1 snRNP can bind to a 5′ splicing site *in vitro*. Only snRNA of U1 cannot bind with the 5′ splicing site. The U1 RNP first binds at the 5′ splice site, then it also binds to the branch site. How U1 RNP recognizes to the branch site is not known. The U2 RNA binds to the branch site by recognizing base-pairing interaction. However, for binding of U2 RNP, prior binding of U1 RNP is essential. Although interaction with U1 snRNP is responsible for recognizing the splicing site, it does not control the cleavage. Initially, U5 snRNA binds close to exon sequences at the 5′ splice site, then it changes its position to the vicinity of the intron. From the results, it has been suggested that snRNA components of snRNPs interact both among themselves and with the substrate RNA by base-pairing interactions, and these interactions allow for changes in structure that may bring reaction groups into opposition and create catalytic center.

A series of loci containing genes potentially coding for splicing factors were originally thought to code RNA, but are now known to encode pre-RNA processing proteins (PRP). Some of the PRP are components of snRNPs and others may function as independent factors. A protein, PRP16, an ATP-dependent helicase has been shown to be involved

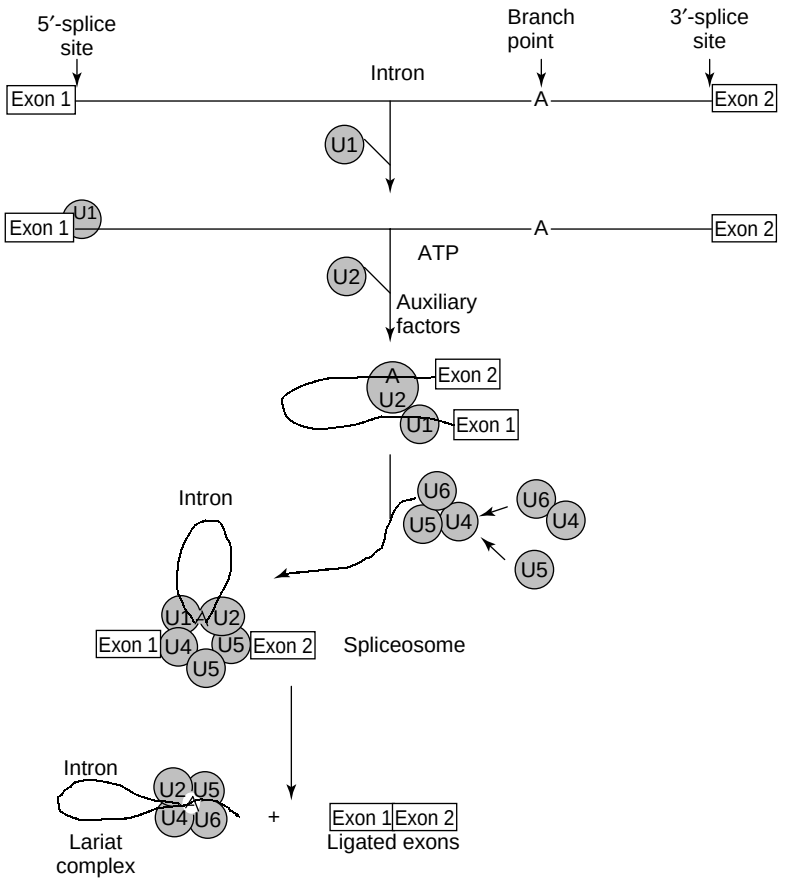

Fig. 11 Pre-mRNA splicing using spliceosome. The spliceosome formation involves the interaction of component that recognizes the consensus sequences. U1, U2, U3, U4, U5 & U6 are different small ribonuclear proteins.

in the second catalytic step of RNA splicing. Another PRP22, another ATP-dependent helicase has been shown to be required to release mRNA from the spliceosome.

3.3
Group I and Group II Introns

The splicing is a two-stage reaction: 5′ exon-intron junction is attacked by a hydroxyl group and secondly the hydroxyl group that has been transferred at the 5′ exon end attacks the 3′ intron–exon junction.

Group I introns (where hydroxyl group is provided by a free guanine nucleotide) are more common than Group II introns (where hydroxyl group is provided by an internal 2′-OH position). Both Group I and II introns can perform the splicing by itself without requiring enzymic activities provided by the proteins. Group II mitochondrial introns have splicing sites resembling the nuclear introns. They are also spliced by the same mechanism as

nuclear HnRNA. Two trans-esterification reactions can be performed by the Group II introns. The number of phosphodiester bonds is conserved in the reaction; therefore, external supply of energy is not required. This could have been an important feature in the evolution of splicing. In autocatalytic splicing, RNA folds into a specific conformation or series of conformations and splicing occurs in cis-conformation. In contrast, snRNAs act in trans form upon HnRNA.

Cech and Group while working with *Tetrahymena thermophila* for the first time found that RNA molecules are capable of self-RNA splicing without the involvement of any protein. Such type of RNA was called as *ribozyme* by them, which means RNA as an enzyme. They showed that RNA could catalyze its own splicing. The self-splicing of RNA in *Tetrahymena thermophila* has been shown in Fig. 12. The enzymes act on other molecules other than on themselves, that's why they coined the term "Ribozyme". Cech and Group afterward also showed that ribozyme could act on a slightly different form of the same RNA; therefore, it is an enzyme in true sense. Cech also suggested that because of the fact that RNA acts as catalyst as well as informational molecule, when life originated RNA may have functioned without DNA or proteins.

3.4 Cis- and Trans-splicing Reactions

Splicing occurs generally as a cis-reaction. It is an intramolecular reaction where controlled deletion of the introns take place. Introns can be removed from RNA molecule allowing exons of the RNA molecule to be spliced together. There is also intermolecular splicing where exons present in different RNA molecules can be spliced (ligated) in one molecule, it is called trans-splicing (Fig. 13). The trans-reactions are rare and never occur between pre-mRNA transcripts of the same gene.

The trans-splicing occurs *in vivo* in some special conditions. In trypanosome, there is presence of a 35 nucleotide leader sequence at the end of many mRNAs. This RNA is called as spliced leader RNA (SL RNA). It donates the 5′ exon for trans-splicing. The SL RNAs found in species of trypanosomes and nematodes have common features. They fold into a common secondary structure having three stem loops and a single-stranded region resembling snRNPs. Trypanosomes possess the U2, U4, and U6 snRNAs but do not have U1 and U3 snRNAs. SL RNA functions without participation of the recognition of the 3′ splicing site and depends directly on RNA.

Some chloroplast genes are also trans-spliced. The *psa* gene of *Chlamydomonas* chloroplast has three widely separated exons. Exon 1 is located 50 kb away from Exon 2 and Exon 2 is 90 kb away from Exon 3. Many other genes lie between these exons. These cannot be transcribed as a common transcript since Exon 1 is in reversed orientation from Exon 2 and Exon 3. Several other genes are required for one or the other of trans-splicing reactions, so the process of splicing this mRNA together is quite complex.

3.5 tRNA Splicing

All the genes coding for tRNAs do not have noncoding sequences (introns) in their structures. Only about 40 out of nearly 400 nuclear tRNA gene products in yeast are interrupted. In each interrupted tRNA gene product, only one intron present just one nucleotide beyond the 3′ side

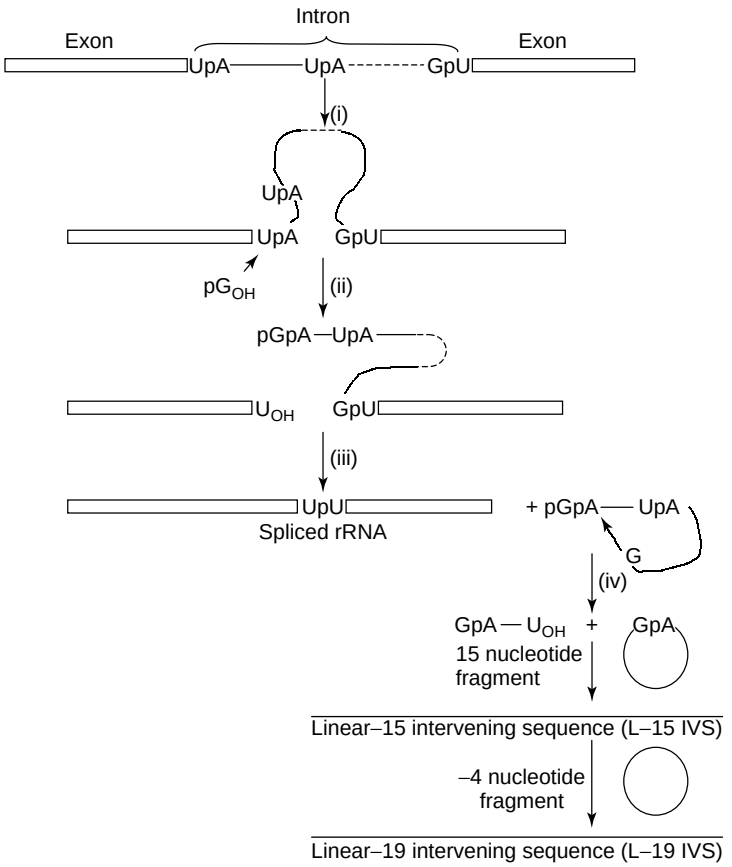

Fig. 12 Self-cleavage of rRNA intron: the intron of pre-rRNA of *tetrahymena* is cleaved by autocatalytic splicing. (1) Folding of RNA to act as ribozyme; (2) a hydroxyl group attached to GTP attacks the 5′ end phosphate of the intron and the phosphodiester bond between exon and intron is broken and a new bond is formed between the guanine nucleotide and the intron; (3) the hydroxyl group at the left exon attacks at the 3′ end of intron. The bond is broken and the exons are ligated releasing the intron; (4) a similar reaction enables the intron to form a circle, snipping 15 nucleotides off its end in the process. The circle opens into a linear molecule and then closes with the removal of four nucleotides. The final open form is called L-19 IVS (linear minus 19 intervening sequence).

of the anticodon has been found. The size of these introns varies from 14 to 46 nucleotides. No consensus sequence has been found in these introns. RNA splicing in the primary transcript of tRNA gene occurs in a different way. There is a separate cleavage and ligation reaction (Fig. 14). The same mode of splicing as in yeast has been reported to the nuclear tRNA gene products of plants, amphibians, and mammals. All the introns in tRNA gene products have a sequence complementary to the anticodon of the tRNA. This is an alternative

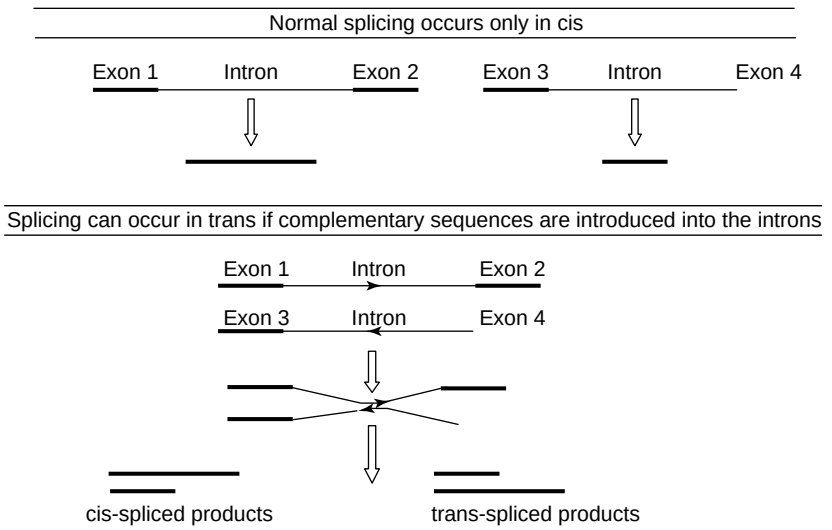

Fig. 13 Schematic diagram showing cis-splicing and trans-splicing reactions.

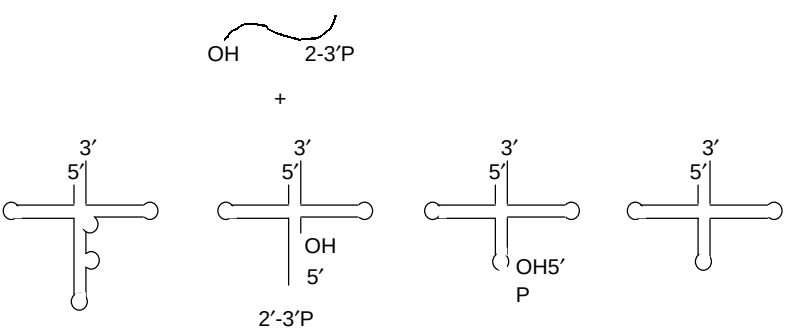

Fig. 14 Splicing of tRNA. It involves cleavage of exon-intron boundaries by endonuclease to generate 2′-3′ cyclic phosphate and 5′-OH termini. The cyclic phosphate is opened to generate 3′-OH and 2′ phosphate group. The 5′-OH is phosphorylated and after releasing the intron, the tRNA half molecule folds into a tRNA-like structure that now has a 3′-OH and 5′ phosphate which is then joined by RNA ligase.

conformation for the anticodon arm in which the anticodon is base-paired to form an extension of usual arm. Splicing of tRNA gene product depends primarily on recognition of a common secondary structure in tRNA. No common sequence in the introns has been reported.

In tRNA gene product splicing, there is cleavage of phosphodiester bond with the help of an endonuclease and no ATP hydrolysis is required. Afterward, an enzyme called as RNA ligase is required for bond formation and ligase reaction requires energy in the form of ATP hydrolysis. Generation of a 2′, 3′ cyclic phosphate bond also occurs during splicing in plants and mammals. On cleavage of phosphodiester bond as a result

of endonuclease reaction, first there is generation of 2′, 3′ cyclic phosphate and 5′-OH termini. The cyclic phosphate is opened to form 3′-OH and 2′ phosphate groups and 5′-OH is phosphorylated. After releasing the intron, the tRNA half molecules fold into a tRNA like structure that has a 3′-OH, 5′-phosphate which gets sealed by RNA ligase.

Ribonuclease P is a tRNA processing enzyme found both in bacteria and eukaryotes. It is a nucleoprotein. Earlier, it was shown that both the RNA and protein components are required for the nuclease to cut the tRNA precursor at a specific point. However, afterward, it was shown that only RNA part of Ribonuclease P may cut the pre-tRNA molecule at a specific site. On the other hand, the protein part alone could not do this indicating that RNA catalyzes the splicing of pre-tRNA molecule.

3.6 Alternate Splicing

When a single gene gives more than one mRNA sequences, it is called alternate splicing. In some cases, use of different start point (5′ splicing site) and/or 3′ splice site alters the pattern of splicing. As written above, this occurs because of only dinucleotide sequences at the ends of the introns. As a result, it is possible that the same portion of the gene may act as exon in one mRNA and as intron in other mRNA. Alternate splicing also occurs because of substitution, addition or deletion of internal exons. For example, if a gene has exon number 1,2,3,4,5,6,7,8, these may be ligated in different ways like 1,2,3,4; 2,3,5,7; 1,2,4,6; 1,2,3,7; 1,3,5,6, and so on. The multiple products are made in the same cell but in others the process is regulated so that particular splicing patterns occur only under certain conditions. Proteins that intervene to bias the use of alternate splice sites have been identified in some cases. In recently sequenced human genome, the number of genes has been estimated in the range of 26,000 but total number of proteins in different stages of development has been estimated to be nearly 500 000. This situation has been explained on the basis of alternate splicing. It has been estimated that from one gene under different conditions, nearly 20 proteins may be coded.

In *Drosophila melanogaster*, native splicing may be caused by mutations in the genes. In case of T/t antigens, the 5′ site for the T antigen removes a termination codon present in t antigen mRNA. That's why the size of T antigen is more than of the t antigen. In E1A transcripts, one of the 5′ sites connect to the last exon in a different open-reading frame which causes change in the C-terminal region of the protein.

Drosophila flies with one X chromosome and two sets of autosomes are male while with equal number of chromosomes and sets of autosomes are female. The X : A ratio activates the so called Sxl which controls positively its own expression as well as that of three other genes. This provides a mechanism so that a single X chromosome of males is transcribed into as much RNA as the two X chromosomes of females. The *Sxl* gene produces male- or female-specific spliced transcripts. They have identical 5′ ends but differ in the presence or absence of a small male specific exon that inserts a stop codon into the transcript. The protein encoded by *Sxl* gene has an eighty amino acid RNA binding domain. The same motif has been reported in many other RNA

binding proteins. It perhaps provides the clue of control of its own processing and that of further proteins in the regulatory cascade.

4
Posttranscriptional Regulation of mRNA

Although regulation of gene expression in eukaryotes at the level of initiation of transcription is considered to be most important, regulation at the level of post-transcription has also been noted in many cases. Capping at the 5′ end of the eukaryotic mRNA is considered to be important. However, polyadenylation at the 3′ end has not been found in all the mRNAs. Whether inhibition of polyadenylation is specifically used to block expression of particular genes is not known. Some genes have multiple putative polyadenylation sites which may be used for alternate splicing (formation of more than one mRNA from one gene). Besides, the choice of polyadenylation site may vary during development of a cell. Switching of splicing patterns may occur in a developmentally significant way.

Polyadenylation does not occur at the extreme 3′ end of mRNA. In fact, between 10 to 30 nucleotides are transcribed preceding the polyadenylation signal which is having the sequence 5′AAUAAA3′ or a variant of it. These terminal nucleotides are cleaved with the help of an endonuclease producing an intermediate 3′ end to which the polyA tail is subsequently attached by the enzyme poly A polymerase. For polyadenylation, there is requirement of a specificity factor too that recognizes the 5′AAUAAA3′ sequence. The specificity factor has three subunits which together bind specifically to RNA containing the sequence 5′AAUAAA3′. The poly A polymerase first synthesizes nearly 10 residue oligo A at the 3′ end of mRNA in the presence of specificity factor. Afterward, this oligo A tail is extended to nearly 200 residues in the presence of an another factor which recognizes oligo A tail and directs poly A polymerase to extend it. As written above, polyadenylation of mRNA is not essential for further translation. It is considered that it may affect the stability of mRNA in the cell. The poly A tail is associated with a particular protein called as poly A binding protein (PABP). It is believed that binding of poly A with PABP is essential for protection of mRNA against degradation by nucleases.

Besides, the stability of mRNA may be regulated in the cytoplasm. This may result in changes in its concentration. It has been found that estrogen hormone not only induces the transcription of vitellogenin gene but also increases the stability of its mRNA in the cytoplasm increasing its half-life to 300 h from 16 h.

All the mRNAs in eukaryotes are found to have 5′ cap. Although the exact significance of capping is not clear, it is considered that it is a recognition point for the attachment of a ribosome at the outset of translation. It is considered equivalent to Shine Dalgarno sequence, GGAGGC, found in prokaryotes to which the small subunit of ribosome attaches to start protein biosynthesis. The ribosome recognizes the cap structure as its binding site, and after attachment migrates along the mRNA until it reaches the initiation codon. It is also found that mRNAs that are translated on the cytoplasmic ribosomes are capped. There is no capping in mitochondrial and chloroplasts mRNAs.

In eukaryotes, a mRNA after splicing at the 5′ end has 5′pppPuNp----3′, where Pu is a purine residue, N is the sugar component of the nucleotide and p represents the phosphate group. However, mature mRNA (after posttranscriptional changes) at the 5′ end has 5′-7-mGpppPuNp-----3′, where 7 mG is 7-methyl guanosine which is attached after transcription and is called a *cap*. During capping, there is cleavage of the terminal phosphate group of the first nucleotide catalyzed by a phosphohydrolase. Afterward, a guanylyl residue is transferred at the 5′ end from GTP by the enzyme, guanylyl transferase and thereafter, modified to 7-methyl guanylyl residue by the enzyme mRNA, guanylyl-7-methyl transferase in the presence of *S*-adenosyl methionine which acts as methyl donor. The newly added guanylyl residue is in the reverse orientation compared to other nucleotides present in mRNA. A cap containing single methyl group is known as *cap o*. If there is addition of another methyl group on the second nucleotide, which was in fact the first nucleotide in the original mRNA, it is called *cap 1*. This occurs only if it is adenine nucleotide. The methyl group is added on the N^6 position of adenine nucleotide. This reaction is catalyzed by the enzyme 2′-*O*-methyl transferase and methyl group donor is same, that is *S*-adenosyl methionine. In some cases, even there is addition of another methyl group to the third nucleotide. The substrate for this reaction is cap 1 mRNA, and acceptor of the methyl group is ribosyl moiety at 2′position. This is called as *cap 2*. The number of caps is considered to be characteristic of the organism. In addition, a low frequency of internal methylation (1 in 1000 nucleotides) also occurs in the mRNA of higher eukaryotes.

In prokaryotic mRNA, generally posttranscriptional changes do not occur. Because of no compartmentation, there may be coupled translation (as soon as mRNA biosynthesis occurs, or even in progress, it may bind with the ribosome to start translation).

5
Transport of Processed mRNA to the Cytoplasm

After posttranscriptional modifications, matured mRNA is transported out of the nucleus in a very rapid process. After coming in the cytoplasm, mRNA binds on the cytoplasmic ribosomes for translation within 1 to 5 min after leaving the nucleus. It is considered that some specific proteins help mature mRNA in transporting it out of the nucleus but exactly how it occurs is not known.

6
Regulation of Gene Expression at the level of Translation

Most of the regulation of gene expression takes place at the level of transcription. Production of mRNA requires many steps, several of which like promoter utilization, RNA splicing, polyadenylation, are known to be regulated. Pre-mRNA stability and the transport of mRNA from the nucleus to the cytoplasm provide a very rapid control of gene expression. Sometimes, the level of translation is manipulated by changing essential components of the translational machinery of the cell. The phosphorylation of ribosomal components (particularly 5S rRNA in the 40S ribosomal subunit) has been correlated with higher polysome levels after

the presence of different growth factors in the mammalian cells. Similarly, in the brine shrimp egg, upon fertilization, an all of a sudden appearance of a previously missing translational initiation factor takes place that is involved in the formation of polysome.

Besides, in mammalian reticulocytes, the control of protein synthesis by hemin is mediated by the formation of a potent inhibitor of polypeptide chain initiator called as hemin controlled repressor (HCR) from a presynthesized precursor (prorepressor). These cells have lost their nuclei, but retain high levels of stable mRNAs encoding mostly hemoglobin chains. In reticulocyte lysates, protein synthesis occurs at a high rate but declines rapidly if hemin is absent. In cells, hemin is synthesized in the mitochondria, which are not present in the lysate. In fact, in the absence of heme, HCR gets activated, whereas, in the presence of heme, HCR gets inhibited. Although the mode of action of HCR was mysterious for many years, now it has been found that it acts as a specific kinase for the phosphorylation of the α subunit of translation initiation factor 2. Because of the initiation factor 2 (eIF2), GTP-GDP exchange cycle, phosphorylation of even a fraction of eIF 2 is sufficient to stop initiation of protein synthesis. Apparently, all of the eIF2B, which is present in lower amount than the amount of eIF2, becomes sequestered into eIF2–eIF2B complexes and is no longer available to recycle the remaining unphosphorylated eIF2.

A translational inhibitor has also been reported from Friend Leukemia cells. The inhibitor has been characterized as heat labile, sulfhydryl reagent insensitive protein of the molecular weight nearly 214 000 daltons. It inhibits initiation of protein synthesis by preventing initiation factor dependent binding of methionyl-tRNA to 40S ribosomal subunit. However, it does not interfere with the formation of ternary complex between the eIF2, methionyl-tRNA and GTP. The inhibitor in fact acts as protein kinase, which phosphorylates the α subunit of eIF2. This inhibitor has been considered analogous to HCR of reticulocyte cells.

An enzyme protein having phosphoprotein phosphatase activity capable of releasing the phosphate group from the phosphorylated α subunit of eIF 2 has also been reported from the rabbit reticulocyte cells. Therefore, it may restore the activity of eIF 2 lost after phosphorylation. Activity of this phosphoprotein phosphatase has been shown to be stimulated about threefolds by optimal concentration of Mn^{+2} ions, but is not stimulated by Ca^{+2} or Mg^{+2} ions. The enzyme is strongly inhibited by Fe^{+2} and by purine nucleoside diphosphates.

During posttranslational modifications, many proteins are modified by phosphorylation, acetylation, hydroxylation, and so on at the side chains of the amino acids. In many proteins, there is conjugation of nonprotein component(s).

Other strategies that act at or before the translation initiation step are alteration in the inherent variability in the life span of eukaryotic mRNA and mRNA stability in response to certain agents.

Acknowledgments

The author acknowledges the facilities of the Distributed Bioinformatics Sub-Centre of the Department of Biotechnology, Government of India, New Delhi presently within the School of Biotechnology, Devi Ahilya University, Indore used in the preparation of this article.

The author also thanks Horizon Scientific Press, Wymondham, England for granting permission to use some material and figures from the author's own book "Advanced Topics in Molecular Biology" published by it, whose copyright is with the publisher.

See also Alternatively Spliced Genes; Biological Regulation by Protein Phosphorylation; Heterochromatin and Eurochromatin – Organization, Packaging, and Gene Regulation; Plant Gene Expression, Regulation of; Preimplantation Mammalian Development, Regulation of Gene Expression in; Regulome; Spatial and Temporal Expression Patterns in Animals; Translation of RNA to Protein; Variation and Conservation of the Development of Animal Bodies: Evolutionary Developmental Biology.

Bibliography

Books and Reviews

Adams, M.D., Rudner, D.Z., Rio, D.C. (1996) Biochemistry and regulation of pre-mRNA splicing, *Curr. Opin. Cell Biol.* **8**, 331–339.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J. (1994) *Molecular Biology of the Cell*, Garland, New York.

Anonymous (2002) The race to computerise biology, in The Economist print Edition, Dec. 12, 2002.

Bourgeois, S., Pfahl, M. (1976) Repressor, *Adv. Protein Chem.* **30**, 1–99.

Cech, T.R., Bass, B.L. (1986) Biological catalysis by RNA, *Annu. Rev. Biochem.* **55**, 599–629.

Crombrugghe, B.D., Busby, S., Buc, H. (1984) Cyclic AMP receptor protein: role in transcription activation, *Science* **224**, 831–838.

Gesteland, R.F., Atkins, J.F. (1993) *The RNA World*, Cold Spring Harbor Laboratory Press, New York.

Herstein, P.R., Frenster, J.H. (1972) Mated Models of Gene Regulation in Eukaryotes, in: Anderson, N.G., Coggin, J.H. (Eds.) *Embryonic and Fetal Antigens in Cancer*, Vol. 2, National Technical Information Service, U.S. Dept. Commerce, Springfield, Virginia, 5–7.

Kumar, A., Srivastava, A.K. (2001) *Advanced Topics in Molecular Biology*, Horizon Scientific Press, Wymondham, UK.

Latchman, D.S. (1993) *Eukaryotic Transcription Factors*, Academic Press, New York.

Lehninger, A.L., Nelson, D.L., Cox, M.M. (1993) *Principles of Biochemistry*, CBS Publishers & Distributors, Delhi, India

Lewin, B. (1994) *Genes V*, Oxford University Press, Oxford, UK.

Maldonado, E., Reinberg, D. (1995) News on initiation and elongation of transcription by RNA polymerase II, *Curr. Opin. Cell Biol.* **7**, 352–361.

Mcknight, S.L., Yamamoto, K.R. (1992) *Transcriptional Regulation*, Cold Spring Harbor Laboratory Press, New York.

Miller, J.H., Reznikoff, W.S. (1978) *The Operon*, Cold Spring Harbor Laboratory, New York.

Morrissey, J.P., Tollervery, D. (1995) Birth of the snoRNPs – the evolution of RNase MRP and the eukaryotic pre-rRNA-processing system, *Trends Biochem. Sci.* **20**, 78–82.

Nagai, K. (1996) RNA-protein complexes, *Curr. Opin. Struct. Biol.* **6**, 53–61.

Raiband, O., Schwartz, M. (1984) Positive control of transcription initiation in bacteria, *Annu. Rev. Genet.* **18**, 173–206.

Romeo, T., Preiss, J. (1989) Genetic regulation of glycogen biosynthesis in Escherichia coli: In vitro effects of cyclic AMP and guanosine-5′-diphosphate-3′-diphosphate and analysis of in vitro transcripts, *J. Bacteriol.* **171**, 2773–2782.

Sharp, P.A. (1987) Splicing of mRNA precurssors, *Science* **253**, 766–771.

Singer, M., Berg, P. (1991) *Genes and Genome*, University Science Press, Mill Valley, CA.

Smith, H.C., Snowden, M.P. (1996) Base modification mRNA editing through deamination – the good, the bad and the unregulated, *Trends Genet.* **12**, 418–424.

Tijan, R. (1995) Molecular machines that control genes, *Sci. Am.* **272**, 38–45.
Wahle, E., Keller, W. (1992) Polyadenylation, *Annu. Rev. Biochem.* **61**, 419–440.
Wahle, E., Keller, W. (1996) The Biochemistry of Polyadenylation, *Trends Biochem. Sci.* **21**, 247–250.
Wolberger, C. (1996) Homeodomain interactions, *Curr. Opin. Cell Biol.* **6**, 62–68.
Yamamoto, K.R. (1985) Steroid receptor regulated transcription of specific genes and gene networks, *Annu. Rev. Genet.* **19**, 209–252.
Young, R.A. (1991) Eukaryotic RNA polymerases, *Annu. Rev. Biochem.* **60**, 689–715.

Primary Literature

Abound, M., Pastan, I. (1975) Activation of transcription by guanosine-5′-diphosphate-3′-diphosphate, transfer ribonucleic acid and a novel protein from Escherichia coli, *J. Biol. Chem.* **250**, 2189–2195.
Akhtar, A. (2003) Dosage compensation: an intertwined world of RNA and chromatin remodelling, *Curr. Opin. Genet. Dev.* **13**, 161–169.
Britten, R.J., Davidson, E.H. (1969) Gene regulation for higher cells: a theory, *Science* **165**, 349–357.
Darnell, J. (1982) Variety in the level of gene control in eukaryotic cell, *Nature* **297**, 365–371.
Grankawski, N., Lehmusvirta, D., Kramer, G., Hardesty, B. (1980) Partial purification and characterization of reticulocyte phosphatase with activity for phosphorylated peptide initiation factor-2, *J. Biol. Chem.* **255**, 310–317.
Gross, M. (1979) Control of protein synthesis by hemin: Evidence that the human controlled translational repressor inhibits formation of 80S initiation complexes from 48S intermediate initiation complexes, *J. Biol. Chem.* **254**, 2370–2377.
Jaenisch, R., Bird, A. (2003) Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals, *Nat. Genet.* **33**, 245–254.
Laemmli, U.K., Tijan, R. (1996) Nucleus and gene expression – a nuclear traffic jam – unraveling multicomponent machines and compartments, *Curr. Opin. Cell Biol.* **8**, 299–302.
Leckie, M.P., Tieber, V.L., Porter, S.E., Roth, W.G., Dietzler, D.N. (1985) Independence of cyclic AMP and rel A gene stimulation of glycogen synthesis in intact Escherichia coli cells, *J. Bacteriol.* **161**, 133–140.
Maniatis, T., Goodbours, S., Fischer, J.A. (1987) Regulation of inducible and tissue specific gene expression, *Science* **236**, 1237–1244.
Niehrs, C., Pollet, N. (1999) Synexpression groups in eukaryotes, *Nature* **402**, 483–487.
Padgett, R.A. (1985) Splicing messenger RNA precurssor: branch site and lariat RNAs, *Trends Biochem. Sci.* **10**, 154–157.
Pinphanichakarn, P., Kramer, G., Hardesty, B. (1977) Partial purification and characterization of a translational inhibitor from Friend Leukemia cells, *J. Biol. Chem.* **252**, 2106–2112.
Primakoff, P., Artz, S.W. (1979) Positive control of lac operon expression *in vitro* by guanosine-5′-diphosphate-3′-diphosphate, *Proc. Natl. Acad. Sci* **76**, 1726–1730.
Ross, J. (1996) Control of mRNA stability in higher eukaryotes, *Trends Genet.* **12**, 171–175.
Scott, W.G., Klug, A. (1996) Ribozymes: structure and mechanism of RNA catalysis, *Trends Biochem. Sci.* **21**, 220–224.
Tuschl, T., Thomson, J.B., Eckstein, F. (1995) RNA cleavage by small catalytic RNAs, *Curr. Opin. Struct. Biol.* **5**, 296–302.
Wasyly, K.B. (1988) Enhancers and transcription factors in control of gene expression, *Biochim. Biophys. Acta* **951**, 17–35.

Regulation of Stem Cell Populations

Daniela Drummond-Barbosa
Vanderbilt University School of Medicine, Nashville, TN

1 Introduction to Stem Cell Biology 69

2 Adult Stem Cells 69
2.1 Germline Stem Cells 69
2.1.1 Mammalian Testis 69
2.1.2 Mammalian Ovary 73
2.1.3 Drosophila Testis 73
2.1.4 Drosophila Ovary 74
2.1.5 *Caenorhabditis Elegans* Gonad 76
2.2 Epithelial Stem Cells 77
2.2.1 Mammalian Epidermis and Hair Follicles 77
2.2.2 Mammalian Gut Crypts 80
2.2.3 Drosophila Follicle Cells 81
2.3 Neural Stem Cells 82
2.4 Hematopoietic Stem Cells 84

3 Embryonic Stem Cells 86

4 Themes of Stem Cell Regulation 88
4.1 Local Regulation – Stem Cell Niches 88
4.2 Cell-intrinsic Factors 90
4.3 Tissue-extrinsic Factors 90

5 Implications 92

Bibliography 93
Books and Reviews 93
Primary Literature 93

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

Keywords

Basement Membrane
Thin layer of extracellular matrix, often rich in collagen type IV, which separates epithelia from connective tissue; also called *basal lamina*.

Germline
The lineage of cells that give rise to gametes, which carry the genetic material that will be transmitted to the next generation.

Knockout Mouse
Genetically modified mouse in which a particular gene has been deleted (in the case of a conditional knockout mouse, the gene is selectively deleted within particular cells or tissues of an otherwise wild-type animal), in a time and/or spatially controlled manner.

Model Organisms
Organisms that have characteristics, such as short life cycles, ready availability, and/or tractability, which make them particularly well suited for laboratory study of biological processes. Because of the great degree of evolutionary conservation of molecular pathways and cellular processes, the large amount of information derived from the study of model organisms can often be applied to the analysis of human development, physiology, disease, and evolution.

Niche
Term used to describe the cellular microenvironment in which stem cells reside and which provides a unique combination of signaling molecules essential for the maintenance of stem cells; the niche hypothesis was first proposed by Schofield in 1978 to explain the regulation of hematopoietic stem cells.

Self-renewal
The ability of a stem cell to indefinitely replenish itself through cell divisions that are followed by the maintenance of stem cell character in at least one of the daughter cells.

Transgenic Animal
Animal carrying genetic material (transgene) introduced into its genome using genetic engineering techniques.

Transit Amplifying Cells
Stem cell progeny that have a limited potential for mitotic divisions and are not yet fully differentiated.

Stem cells are essential for the health of multicellular organisms, as they give rise to all cells of the body during embryonic development, and in adults, maintain the integrity and function of many tissues. Stem cells are functionally defined as cells that have the extended ability to self-renew and to give rise to various cell types present in a given tissue. The behavior of stem cells is remarkably context dependent, underscoring both the existence of complex regulatory mechanisms involved in their maintenance and function, as well as the inherent difficulties of studying their biology in a nonadulterating manner. Despite these difficulties, intense research has revealed that their regulation is achieved via the orchestration of cell-intrinsic factors, locally secreted factors, and extrinsic signals impinging on the tissue. As knowledge of the mechanisms of stem cell regulation expands through the combined study of powerful model systems, we will continue to gain valuable insights into how to manipulate stem cells for the welfare of human kind.

1 Introduction to Stem Cell Biology

Stem cells have been main characters in media reports for the past several years, reflecting their awe-inspiring properties and potential for clinical applications. Despite their importance and the attention they have received from many basic, applied, and clinical scientists, however, stem cells remain difficult to recognize, to study, and even to define. Morphological and molecular markers are scarce and not available for all types of stem cells, and studies so far indicate that no intrinsic stem-specific marker exists that unifies all stem cells and sets them apart from all other cells in the body. The scarcity or absence of markers coupled with the relatively low numbers at which stem cells normally occur leads to their being defined on a functional basis, namely, that they are cells with the potential to undergo a very large number of divisions to self perpetuate and give rise to different cell types in an embryo or tissue. Many layers of regulation are required for maintaining stem cell properties and tailoring their functions to the needs of the tissues in which they reside. Although much is still unknown, considerable progress has been made in recent years in unraveling the mechanisms of stem cell regulation, largely resulting from the use of many different systems of study, ranging from invertebrate model systems to human stem cells. Continued investigation of the regulation of stem cells will provide the pillars on which to build our basic knowledge of stem cell biology, from which informed strategies for their clinical application will be derived.

2 Adult Stem Cells

2.1 Germline Stem Cells

2.1.1 Mammalian Testis

During embryonic development, pluripotent primordial germ cells are set aside and associate with the developing gonad, and a fraction of them gives rise to germline stem cells (GSCs). In males, GSCs are

active throughout adult life, leading to the constant production of sperm within seminiferous tubules. The proper function of GSCs requires a signaling environment (or niche) provided by closely associated Sertoli cells, which are large somatic cells (Fig. 1a). Although Sertoli cells extend processes radially from the basement membrane to the lumen, contacting all germ cells, their asymmetric secretion of signals creates gradients along their extended processes. The interstitial area of the testis,

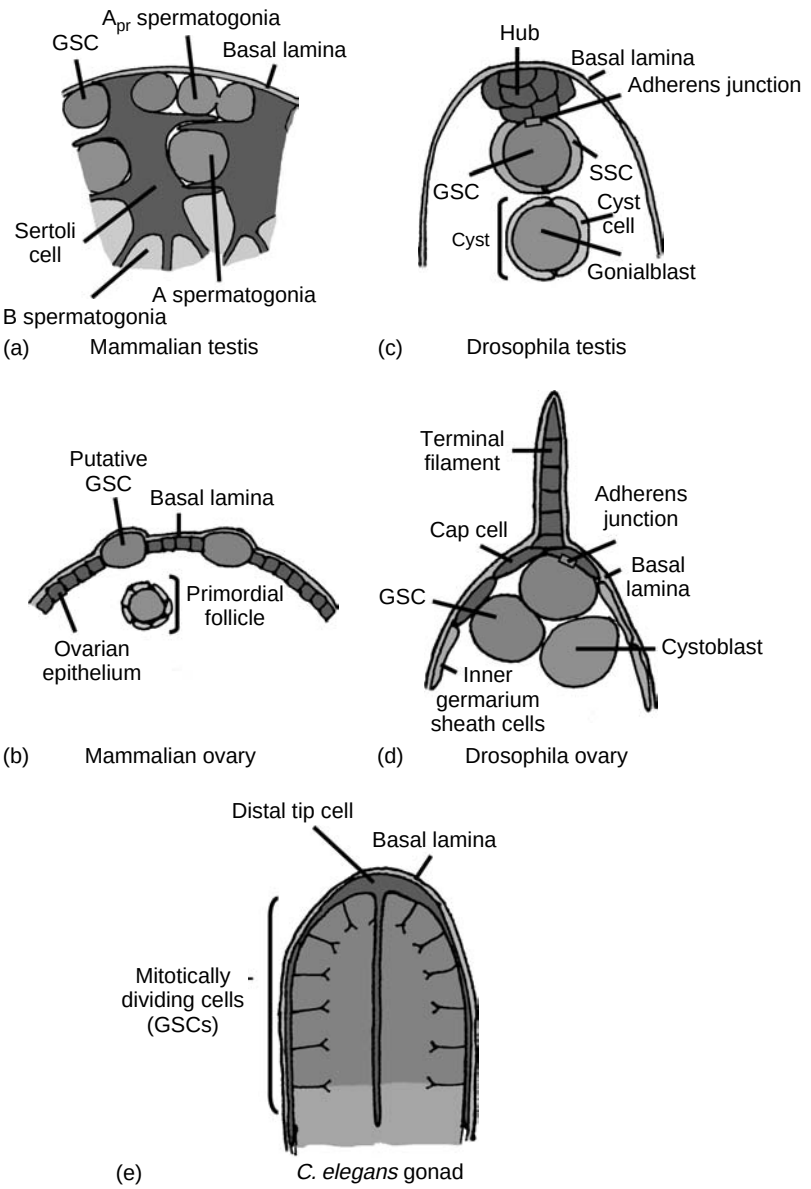

surrounding the basement membrane of the seminiferous tubules, also contains steroid-secreting Leydig cells as well as myoid cells.

Premeiotic male germ cells are called spermatogonia, and these are divided into undifferentiated A-type spermatogonia and differentiating spermatogonia. GSCs correspond to the earliest undifferentiated stage, A_{single} (A_s) spermatogonia, located at the periphery of seminiferous tubules in low numbers. A_s spermatogonia divide to self-renew and give rise to daughter cells that undergo four rounds of divisions remaining in clusters of 2 cells (A_{paired} [A_{pr}] spermatogonia), and subsequently 4, 8, and 16 cells (collectively called $A_{aligned}$ [A_{al}] spermatogonia). The spermatogonia subsequently differentiate through A1, A2, A3, and A4 intermediate stages into B spermatogonia, which undergo meiosis and form sperm. As they progress through spermatogenesis, germline cells move to the center of the tubules, culminating in the final release of sperm into the lumen.

Male GSCs, or A_s spermatogonia, can be recognized by their location and behavior. A_s spermatogonia show longer cell cycles and asynchronous divisions because they are not present in clusters. They are also present near the basement membrane, closely associated with Sertoli cells. The gold standard for the functional identification of stem cells is the ability of transplanted cells to repopulate testes of an infertile host, thereby restoring long-term fertility. On the basis of transplantation assays, the number of GSCs in normal testes was estimated at 2 per 10^4 germ cells. For transplantation assays, testis cells from fertile mice or rats are harvested, and the cell suspension is microinjected into seminiferous tubules of a host treated

Fig. 1 Germline stem cells. (a) In the mammalian testis, low numbers of germline stem cells (GSCs), also called A_{single} (A_s) spermatogonia, are located at the periphery of seminiferous tubules near the basal lamina and in close contact with large somatic sertoli cells. GSCs renew themselves and also give rise to daughter cells that form cysts of 2 (A_{paired}, A_{pr}), 4, 8, and 16 cells ($A_{aligned}$, A_{al}). They subsequently differentiate through A and B spermatogonia, which undergo meiosis and form sperm. (b) In the mammalian ovaries, putative GSCs reside in the periphery of the ovary, within the ovarian epithelium, juxtaposed to the basal lamina. These putative GSCs are thought to renew themselves and give rise to daughter cells that associate with follicle cells to form new primordial follicles in adult females. The existence of these mammalian female GSCs, however, is highly controversial. (c) The Drosophila testis is surrounded by a basal lamina and contains a signaling center, called the hub, at its apical tip. Five to nine GSCs are attached to the hub cells via adherens junctions, and each GSC is also in close association with a pair of surrounding somatic stem cells (SSCs). GSCs divide asymmetrically, perpendicular to the hub, to give rise to a new GSC (near the hub) and a gonialblast (away from the hub). The gonialblast is associated with a pair of somatic cyst cells derived from the SSCs, and it divides to form germline cysts that differentiate into sperm. (d) In each germarium within the Drosophila ovary, two to three GSCs are present in a microenvironment or niche, formed by somatic terminal filament and cap cells in close association with a basal lamina. GSCs are in close association with cap cells via adherens junctions, and they divide asymmetrically to renew themselves and generate cystoblasts. In a region surrounded by inner germarium sheath cells, each cystoblast divides to give rise to a germline cyst containing a future oocyte. (e) In the distal end of the *C. elegans* gonad, the distal tip cells extend long processes and are in close contact with syncytial germline cells that divide mitotically to renew themselves, implying that they include GSCs. In addition, germline cells generated in this distal region move more proximately, away from the distal tip cell, and eventually undergo meiosis to form oocytes.

with busulfan (a chemotherapeutic drug that depletes testes of GSCs) or carrying a sterility-causing mutation. Injected GSCs populate sites along the basement membrane of the seminiferous tubules and establish colonies of donor GSC-derived cells supporting sperm production. Subpopulations of germline cells can also be isolated based on the expression of combinations of proteins by fluorescence-activated cell sorting, and subsequently subjected to transplantation assays to determine whether they include GSCs, allowing the identification of molecular markers enriched in A_s spermatogonia.

Although no marker specifically labeling A_s spermatogonia has been identified to date, $\alpha 6\beta 1$ integrins (which may help attach GSCs to the basement membrane), and Neurogenin 3 (a class B basic helix-loop-helix (bHLH) transcription factor) are enriched in A_s spermatogonia, which also lack c-kit expression. Recently, analysis of the classic *luxoid* mutant showed that A_s spermatogonia renewal is impaired and germ cells are lost; mutation of the gene *Zfp145*, which encodes the zinc finger–containing transcription factor Plzf, was found to be responsible for the phenotype. Plzf is coexpressed with the transcription factor Oct 4 in undifferentiated spermatogonia.

Sertoli cells produce a variety of signals, such as glial-cell-line-derived neurotrophic factor (GDNF), insulin-like growth factor 1 (IGF-1), transforming growth factors (TGFs) α and β, epidermal growth factor (EGF), and Steel factor. GDNF, however, is the only secreted factor produced by Sertoli cells known to be required for A_s spermatogonia maintenance. Decreased levels of GDNF, caused by gene targeting of one copy of the GDNF gene, result in Sertoli cell-only seminiferous tubules, reflecting a depletion of germ cells. Conversely, transgenic mice overexpressing GDNF under a testis-specific promoter, display large clusters of undifferentiated spermatogonia resulting from a block in differentiation. Studies using a serum-free culture system for GSC showed that GDNF acts directly on undifferentiated spermatogonia, which express the receptor for GDNF (composed of GDNF-family receptor $\alpha 1$ and c-Ret receptor tyrosine kinase), to stimulate GSC self-renewal. In addition, bone morphogenetic protein (BMP) 8B is expressed in early germ cells, and knockout mouse experiments suggest that it is required for the establishment of GSCs as well as their proliferation and maintenance.

Other secreted factors are required for germline differentiation. Steel factor, which is produced by Sertoli cells, acts on the c-kit receptor of spermatogonia to promote proliferation of differentiating type A spermatogonia. Pituitary luteinizing hormone (LH) and follicle stimulating hormone (FSH) influence the differentiation of spermatogonia, acting indirectly via supporting somatic cells, but no direct effects of these hormones on undifferentiated spermatogonia are known. In response to LH, Leydig cells secrete testosterone, which in turn acts on Sertoli cells. FSH plus testosterone stimulate secretion by Sertoli cells of signals required for germ cell differentiation.

The mammalian testis is a powerful system to understand the regulation of stem cells. The availability of transplantation assays, in which the ability of donor GSCs to reconstitute the ablated germline of the host is tested, provide a rigorous system for dissecting the *in vivo* requirement of different factors for GSC function. Insights obtained from this system have direct implications for human fertility.

2.1.2 Mammalian Ovary

For over half a century, it has been widely accepted that adult mammalian ovaries do not contain self-renewing GSCs. This dogmatic view, however, was recently challenged by experiments suggesting that GSCs may in fact be present in adult mouse ovaries. Putative GSCs with large ovoid shape that reside in the surface epithelial cell layer covering the ovary are thought to divide mitotically to generate germline cells that enter meiosis and give rise to new follicles in adult females (Fig. 1b). These putative GSCs may play a central role in replenishing the follicular population of the ovary as follicles undergo a high rate of atresia or degeneration. Similar to GSCs in the testis, the ovarian GSCs are sensitive to busulfan-treatment, and they also appear capable of contributing to follicle formation upon transplantation of ovarian fragments from one female to another. These findings may hold true in human females as well, and this is an area of investigation. The possible existence of GSCs in female adults suggests that the decreased fertility observed in older females may be because of stem cell senescence and/or impaired function.

The data suggesting the presence of GSCs in mammalian ovaries, however, remains highly controversial. Definitive demonstration that female GSCs exist will require lineage tracing experiments to show that single putative GSCs can indeed give rise to daughter cells that contribute to follicle formation in adult females. In addition, it will be very important to understand what cell types and molecules are required for establishing, maintaining, and regulating GSCs, and to determine their mechanism of function.

2.1.3 Drosophila Testis

Drosophila germline precursor cells (''pole cells'') are formed early during embryogenesis and later populate the gonads. A fraction of the pole cells give rise to GSCs, which can be identified by their morphology, location, and behavior. The Drosophila testis is a simple tube with 5 to 9 GSCs located in a circle at its closed apical end that are in close contact with a compact, dome-shaped structure called the hub, which is composed of 8 to 16 somatic cells (Fig. 1c). Each GSC is enclosed by a pair of somatic stem cells (SSCs) known as cyst progenitor cells that send out cytoplasmic projections that contact the hub. Lineage tracing experiments have confirmed that the GSCs indeed function as stem cells. GSCs divide asymmetrically to self-renew and give rise to a gonialblast, which is surrounded by a pair of cyst cells resulting from the asymmetric division of the SSCs to form a germline cyst. As the germline cells progress through spermatogenesis, they remain surrounded by the pair of postmitotic cyst cells. The gonialblast divides four times with incomplete cytokinesis to form a cyst of primary spermatocytes that enter meiosis to form spermatids and, ultimately, sperm.

Drosophila GSCs, gonialblasts, and early germline cysts have a specialized subcellular structure called the fusome, which is rich in the cytoskeletal proteins Hu li tai shao (Hts, an adducin-like protein), α-spectrin, β-spectrin, and ankyrin. In GSCs and gonialblasts, the fusome is round and often referred to as a spectrosome, whereas in germline cysts, it becomes elongated and branched. GSCs and gonialblasts do not express Bam (encoded by the *bag-of-marbles* gene), a differentiation factor expressed in two-cell to early 16-cell germline cysts.

The hub is an important component of the GSC niche in the Drosophila testis, providing important signals regulating GSC maintenance and function. As GSCs divide, mitotic spindles are oriented perpendicular to the hub such that one daughter cell (the GSC) remains in contact with the hub, whereas the other (the gonialblast) is placed in a distinct environment away from the hub. The orientation of GSC division requires the Drosophila homolog of the adenomatous polyposis coli (APC) tumor suppressor, presumably via the interaction with DE-cadherin and Armadillo (the homolog of β-catenin) for the formation of adherens junctions between hub cells and GSCs.

Genetic analysis has identified several factors required for GSC maintenance and function. An important signal secreted by the hub is the ligand Unpaired (Upd). GSCs unable to respond to Upd due to a lack of *stat92E* (encoding Drosophila Signal Transducer and Activator of Transcription [STAT]) function are not maintained as stem cells, but instead differentiate. Conversely, overexpression of Upd in the testis results in accumulation of undifferentiated, GSC-like cells away from the hub. These data demonstrate that Upd acts on neighboring GSCs to activate the Janus kinase (JAK)-STAT pathway, resulting in maintenance of GSC status, while gonialblasts, which are not exposed to high levels of Upd, can undergo differentiation.

Other signals are also important for GSC maintenance and differentiation. Hub cells and somatic cyst cells express the BMP molecules Glass bottom boat (Gbb, also known as 60A) and Decapentaplegic (Dpp), and these signaling molecules together activate the BMP pathway directly in male GSCs to control their maintenance, at least in part, via the repression of Bam expression. In addition, disruption of EGF signaling in somatic cyst cells disrupts differentiation of gonialblasts, resulting in accumulation of undifferentiated germ cells in the testis. Somatic cyst cells also signal to germline cysts via a TGF β-like signaling pathway to limit the number of mitotic divisions during cyst formation. Finally, proliferation of germ cells is also controlled by the germline-expressed genes *bam* and *benign gonial cell neoplasm* (*bgcn*).

Remarkably, when the GSC niche is depleted of GSCs via the transient inactivation of the JAK-STAT pathway, dividing germline cysts can dedifferentiate into GSCs that repopulate the niche. In these experiments, a temperature-sensitive allele of *stat92E* was employed; at the nonpermissive temperature, loss of *stat92E* function resulted in loss of GSCs because of their differentiation into gonialblasts and subsequent formation of germline cysts. Restoration of JAK-STAT signaling by returning GSC-depleted males to the permissive temperature resulted in a breakdown of dividing germline cysts into single cells that repopulated the hub as GSCs. These experiments revealed a great degree of plasticity in this system.

The knowledge of the precise location of GSCs and surrounding niche cells in the Drosophila testis, as well as the ability to follow their behavior and apply powerful genetic tools to the study of GSCs, makes this an ideal experimental system to further examine their maintenance and regulation.

2.1.4 Drosophila Ovary

Drosophila females produce eggs throughout most of their adult lives, and egg production is supported by both germline and somatic stem cells. Each ovary is composed of 15 to 20 ovarioles, which are strings of progressively older egg chambers that contain the developing eggs.

Somatic and GSCs reside at the anterior-most tip of each ovariole in a structure called the germarium, within which the egg chambers are formed (Fig. 1d). The division of a GSC generates a daughter cell called a cystoblast, which undergoes four rounds of division with incomplete cytokinesis to form a 16-cell cyst. The posterior-most cell within a cyst will become the oocyte, and the remaining 15 cells will become nurse cells that support the growth and development of the oocyte. As the cyst progresses through the germarium, it acquires a layer of follicle cells derived from the somatic stem cells. A round cyst surrounded by a monolayer of follicle cells buds off from the germarium as an egg chamber, which continues to develop as it progresses posteriorly through the ovariole.

GSCs are located near several populations of somatic cells, namely, terminal filament cells, cap cells, and inner germarium sheath cells. These cells create a niche crucial for stem cell maintenance and function. GSCs contact cap cells and inner sheath cells, and they divide asymmetrically to generate a stem cell in its original position and a cystoblast in a slightly posterior position, such that the cystoblast contacts inner sheath cells, but not cap cells. GSCs and cystoblasts contain a round spectrosome, whereas older cysts within the germarium contain branched fusomes. GSCs do not express Bam, whereas cystoblasts and early dividing germline cysts express this differentiation factor.

Signals from the niche control GSC maintenance and function. Adherens junctions formed by DE-cadherin and armadillo are present between cap cells and GSCs. Genetic analysis suggests that this molecular "glue" is important for the maintenance of GSCs, presumably by physically keeping them within the niche environment. Dpp is a pivotal signal expressed in somatic niche cells. Ectopic Dpp expression results in increased numbers of GSC-like cells within the germarium. Dpp acts directly on GSCs as evidenced by the fact that elimination of the function of signaling pathway components (e.g. Dpp receptors encoded by *punt* and *tkv* or downstream transcriptional factors encoded by *Mad* and *Med*) specifically within the germline disrupts GSC maintenance and division. Gbb is also expressed in somatic cells near GSCs, but overexpression of Gbb is not sufficient to induce GSC tumors within the germarium. Reduced Gbb levels, however, result in reduced activation of the BMP pathway in GSCs, suggesting that Gbb may cooperate with Dpp for optimal levels of GSC activation. Dpp and Gbb activate a signal transduction cascade in GSCs that results in direct binding of the downstream Mad-Med Smad protein complex and repression of expression of the Bam differentiation factor. Repressing Bam expression in GSCs is crucial: if Bam expression is forced in these cells using a heat-shock inducible *bam* transgene, they undergo differentiation. In addition, Bam and the ubiquitin protein ligase Smurf, appear to play redundant roles in restricting the domain of Dpp receptor activation, suggesting the existence of multiple mechanisms that ensure that only GSCs respond to the Dpp signal.

Several other factors are also required for GSC maintenance. Nanos is required to establish and maintain GSCs by preventing their differentiation, as removal of Nanos function in the germline of adult females results in loss of GSCs. *Pumilio* mutant flies also show loss of GSCs. On the basis of their functions during early embryogenesis, Nanos and Pumilio presumably form

a complex that binds to the 3′ untranslated region of target mRNAs to repress their translation. The *piwi* gene encodes a conserved protein expressed in niche cells as well as in GSCs. Expression of *piwi* within somatic cells is required for GSC maintenance, whereas it is required within the germline to control GSC division and germline development. The *fs(1)Yb* gene is required to maintain *piwi* expression in terminal filament and cap cells.

As in the Drosophila testis, dedifferentiation of germline cysts into GSCs has been reported in the ovary. In these experiments, a pulse of Bam expression caused larval germ cells, which are present in a Dpp-rich environment, to form cysts. After removal of Bam, four- and eight-cell cysts present in second instar larval ovaries were found to break down and form single germline cells that gave rise to normal GSCs, supporting oogenesis in the resulting adult females. Similarly, a pulse of Bam expression in adult ovarioles overexpressing Dpp resulted in GSC differentiation and cyst formation, and these cysts were also capable of breaking down into single GSCs. These results, coupled with those in the Drosophila male testis, suggest that, under specific conditions (e.g. a Dpp-rich environment), differentiated cells may serve as a source of stem cells.

The investigation of GSC regulation in the Drosophila ovary has greatly advanced the field of stem cell regulation. The power of genetics, the easy identification of GSCs and surrounding niche cells, and the ability to measure the behavior of GSCs and their progeny will continue to make this an excellent system to establish new paradigms of GSC regulation. Insights from the Drosophila ovary will be instrumental in forging ahead with the investigation of the putative GSCs present in the mammalian ovary.

2.1.5 *Caenorhabditis Elegans* Gonad

In the *Caenorhabditis elegans* hermaphrodite, the adult gonad is a U-shaped tube containing mitotically dividing germ cells at the distal ends of the anterior and posterior arms. The germline forms a syncytium, such that each cell is connected via a cytoplasmic bridge to a core of cytoplasm (rachis) that is present along the center of the distal gonad arm (Fig. 1e). These mitotically dividing germline cells self-renew and give rise to cells that progress through meiosis to produce oocytes; however, a stereotypic stem cell has not yet been identified. At a ~20-cell diameter distance from the distal tips of the gonad, cells enter meiosis and progress into meiotic divisions as they move proximally within the gonad arms. A distal tip cell is in close contact with each distal end of the gonad, providing a niche environment that regulates the balance between mitosis and meiosis, while sheath cells surround more proximal regions of each gonad arm.

The distal tip cell is required for the establishment and maintenance of GSCs, because distal tip cell ablation results in meiosis and differentiation of all germline cells, while its relocation results in the coordinate ectopic location of mitotically dividing germline cells. The distal tip cell produces LAG-2: a Notch receptor ligand. This signal is received by germline cells at the distal end of the gonad, activating the GLP-1/Notch signaling pathway in distally located germline cells, which causes them to remain in mitosis. Constitutive activation of the GLP-1/Notch pathway in the germline inhibits exit from mitosis and entry into meiosis, resulting in a tumorous germline, whereas mutations in Notch pathway components lead to premature entry into meiosis. Activation of the

GLP-1/Notch pathway inhibits accumulation of the RNA-binding protein GLD-1 via the translational inhibitors NOS-3 (similar to Drosophila Nanos) and FBF-2 (similar to Drosophila Pumilio). GLD-1 levels are important in controlling the decision between proliferation (low GLD-1) or meiosis (high GLD-1). GLP-1/Notch signaling also inhibits the *gld-3* gene, a homolog of *Bicaudal-C*, via the FBF-2 translational repressor. GLD-3 binds to the poly-A polymerase GLD-2, and together they help promote the transition from mitosis to meiosis.

Despite the current view that the distal mitotic germline population as a whole maintains stem cell properties, it is not clear whether all mitotically dividing cells within that population are equivalent, or whether only specific cells have stem cell properties. Lineage analysis will be required to distinguish between these mechanisms and represents an important step towards a full understanding of how these syncytial GSCs are regulated.

2.2 Epithelial Stem Cells

2.2.1 Mammalian Epidermis and Hair Follicles

In the mammalian epidermis and hair follicles, differentiated cells are continuously lost and replaced by cells derived from stem cells. In the interfollicular epidermis, dividing keratinocytes are found in the basal layer attached to the basement membrane (Fig. 2a). As keratinocytes move away from the basement membrane towards the surface of the skin, they stop dividing, undergo differentiation into spinous and granular layer cells, and finally die, thereby forming the heavily keratinized stratum corneum. In mouse skin, the entire differentiation process from the basal layer to the keratinized stratum corneum takes 10 to 14 days, whereas in human skin this process takes longer.

Rare stem cells are thought to reside in the basal layer of the epidermis, in regions that express high levels of the Delta1 ligand for Notch and of $\alpha 6\beta 1$ integrins, suggesting that interactions with the extracellular matrix may be important for their regulation. Indeed, $\alpha 6$ and $\beta 1$ RNAs are found to be upregulated in stem cell-enriched populations relative to other mitotically dividing cells, whereas $\beta 1$ integrin can be used as a marker during cell sorting procedures for enrichment for stem cells. Stem cells give rise to actively dividing transit amplifying cells that subsequently undergo differentiation through the layers of the epidermis. Although the stem cell population renews itself, there is no evidence for invariant asymmetric divisions, suggesting that each stem cell division can give rise to any combination of stem cell and transit amplifying cells. In addition, long-term maintenance of this stem cell population is likely to require mobilization of bulge stem cells, which reside in the hair follicles (see below).

The hair follicle is composed of concentric layers: the outer root sheath (ORS), which is an extension of the epidermis, the inner root sheath (IRS), and the hair shaft (Fig. 2a). The IRS and hair shaft are derived from relatively undifferentiated matrix cells. Matrix cells are in close contact with the dermal papilla, which is composed of mesenchymal cells. Although matrix cells are actively dividing, their proliferation potential is limited. The stem cells that sustain long-term hair follicle growth cycles are a distinct population thought to reside in a structure called the bulge. Bulge cells are undifferentiated

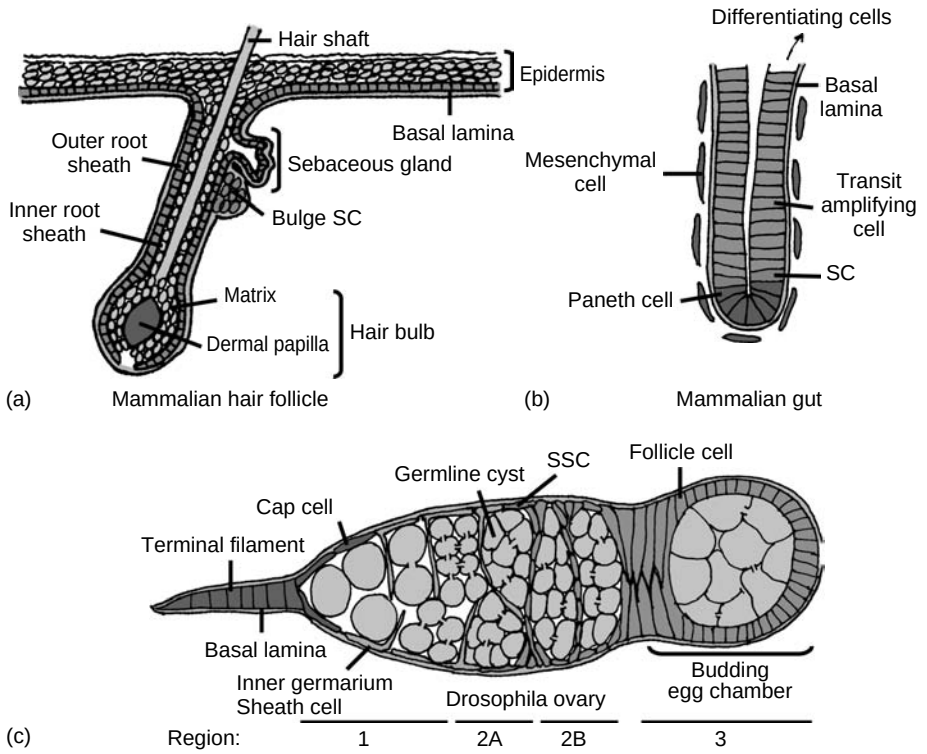

Fig. 2 Epithelial stem cells. (a) The mammalian hair follicle contains multipotent stem cells (SCs) in a region called the bulge, below the sebaceous gland, and the entire follicle is surrounded by a basal lamina. Bulge SCs give rise to transit amplifying matrix cells, which have a limited proliferation potential and give rise to differentiating cells that produce the hair shaft. Matrix cells surround the dermal papilla, which is an important signaling center for the bulge SCs. During the catagen phase of the hair follicle cycle, the lower part of the follicle undergoes extensive cell death, and as a result, the dermal papilla moves toward the bulge. After the telogen rest period, signals from the dermal papilla mobilize SCs from the bulge to regenerate the matrix and reinitiate the anagen growth phase. (b) The basal region of the mammalian gut crypt contains stem cells (SCs) above differentiated Paneth cells, and a basal lamina and mesenchymal cells surround the entire crypt. New cells produced from the basal SCs migrate up and, once they exit the crypt, begin differentiating into various intestinal cell types. (c) The Drosophila ovary germarium contains two somatic stem cells (SSCs) that contact a basal lamina as well as inner germarium sheath cells. In addition, SSCs respond to signals produced by the terminal filament and cap cells. As SSCs divide, they renew themselves and also generate daughter cells that further divide and surround newly formed 16-cell germline cysts, supporting oocyte development.

cells that lie at the base of the epithelial portion of the hair follicle and are present throughout the hair follicle growth cycles. Bulge cells are thought to have a large proliferation potential on the basis of the size of colonies formed in culture. *In vivo*, however, the proliferation of bulge cells is tightly regulated, and they rarely divide. Because bulge stem cells are slow cycling, they can be experimentally identified as

"label-retaining cells." In this procedure, all cells in the epithelium are initially labeled by a continuous supply of tritiated thymidine or bromodeoxyuridine (BrdU), which becomes incorporated into the DNA as cells divide. After a long chase period, the label is diluted and lost from all rapidly dividing cells (e.g. matrix transit amplifying cells), whereas the slowly dividing bulge stem cells retain the label. This strategy has been used both in lineage tracing experiments and to obtain highly purified bulge stem cell populations.

Each hair follicle growth cycle includes an active growth phase (anagen), a destruction phase (catagen), and a rest period (telogen), with the length of different phases varying according to the type of hair follicle and the specific organism. During anagen, the active proliferation of matrix cells gives rise to cells that move upward and differentiate to form IRS and hair shaft cells. During catagen, the lower part of the follicle undergoes extensive cell death and, as a result, the dermal papilla moves towards the bulge. After the quiescence of the telogen, signals from the dermal papilla leads to recruitment of cells from the bulge, which gives rise to matrix cells that reinitiate the cycle.

Bulge-residing stem cells give rise not only to hair follicle cells, but also to cells of the epidermis and sebaceous glands, in a highly regulated process. In particular, bulge stem cells are mobilized to repopulate interfollicular epidermis upon wounding of the epidermis, and it is likely that they also give rise to the interfollicular epidermal stem cells in normal situations. Evidence for the multipotency of stem cells in the bulge comes from both transplantation assays and lineage tracing analysis. Recent experiments have shown that two distinct populations of bulge-residing stem cells exist. One population maintains contact with the basement membrane, the other is suprabasal and arises later, after the start of the first postnatal hair cycle. Although these populations show different gene-expression profiles, they behave similarly within the niche, have the ability to self-renew *in vitro*, and can form epidermis and hair when grafted.

Although the signals regulating bulge stem cell activity are incompletely understood, recent genetic profiling of purified bulge stem cells performed by two groups suggests possible mechanisms for their quiescent, multipotential character. Bulge stem cells express elevated levels of cell cycle regulatory proteins and keratinocyte growth inhibitors such as TGF β signaling components. Indeed, TGFβ receptor activation, as measured by the levels of nuclear phosphorylated Smad2 (a downstream effector), appears to be elevated in bulge cells relative to their progeny. Conversely, transcripts for proliferation-associated proteins are downregulated in bulge stem cells. Bulge stem cells also express high levels of fibroblast growth factor (FGF) 1 and its receptor, suggesting a role for the FGF1 pathway in maintenance of stem cell character.

Wnt signals appear to play a dual role in the transition from bulge to matrix cell type. Both bulge cells and matrix cells receive Wnt signals, resulting in stabilization and increased levels of β catenin; however, the cellular outcome of Wnt signaling is different in bulge versus matrix cells, presumably due to expression of distinct β catenin binding partners. Bulge cells express Tcf3, which normally acts as a transcriptional repressor when complexed with β catenin, suggesting a Wnt-repressed niche for bulge stem cells. In contrast, matrix cells express Lef1, which, when associated with β catenin

results in transcriptional activation of downstream targets. Downstream targets of β catenin – Lef1 complexes involved in switching from rarely dividing bulge cells to actively dividing matrix cells include c-Myc and cyclin D.

Wnt signaling via Lef1 is also required for hair shaft differentiation of matrix cells, and establishment of Lef1 expression in matrix cells requires noggin, a BMP-inhibitor secreted by the dermal papilla. These observations suggest that transient repression of BMP signaling is required to make matrix cells responsive to Wnt signals. Subsequently, however, BMP signaling via the BMP receptor IA is required for stabilization of β-catenin in response to Wnt stimulation such that both pathways act together to induce the expression of hair differentiation–inducing genes.

As the complex roles of different signaling pathways are unraveled and new players are identified, our understanding of the hair follicle stem cell system will continue to expand. In addition, new technologies are currently being developed to allow the introduction of transgenes into hair follicle cells for potential therapeutic use.

2.2.2 Mammalian Gut Crypts

The inner lining of the small intestine and colon is a simple columnar epithelium that forms invaginations called crypts of Lieberkuhn and projections called villi. The intestinal epithelial cells are turned over at a high rate, with new cells being produced from stem cells present at the base of the crypts (Fig. 2b). Stem cell-derived transit amplifying cells migrate up, and once they exit the crypt, begin differentiating into absorptive epithelial columnar cells, goblet cells, and enteroendocrine cells. Differentiated cells continue to migrate up and are shed from the tips of the villi, taking 5 to 7 days from crypt base to villus tip. Paneth cells originate from the same precursor cells as do other differentiated cells; however, Paneth cells migrate toward the bottom of the crypt instead.

Crypts are composed mostly of undifferentiated cells, and on the basis of lineage tracing studies and label-retaining properties, 4 to 5 stem cells have been shown to reside at the fourth or fifth cell positions from the base of each crypt in the small intestine, above differentiated Paneth cells. Paneth cells are long lived and produce a number of factors such as EGF and tumor necrosis factor α; however, ablation of Paneth cells does not appear to have any effect on crypt cell production. Stem cells are present in regions enriched for $\beta 1$ integrin (crypts express $\alpha 2\beta 1$, in contrast to villi, which express $\alpha 3\beta 1$ integrins), suggesting that adhesion to the extracellular matrix is important.

The Wnt pathway, acting via Tcf4/β catenin complexes, plays an important role in stimulating proliferation in the gut. Normal proliferating cells accumulate β catenin in the nucleus. In Tcf4 null mice, all epithelial cells differentiate, and no proliferation occurs. In addition, if expression of the secreted Dkk1 inhibitor of Wnt molecules is targeted to the proliferative compartment of the mouse intestine, epithelial proliferation is greatly reduced, resulting in the loss of crypts. Conversely, mutations in APC or β-catenin that stabilize β-catenin-Tcf4 complexes occur in the vast majority of human colon cancers and are sufficient to cause intestinal adenomatous polyp formation in mice.

Among the downstream factors of the Wnt pathway in the intestine are c-myc, SOX9 and Eph/Ephrin. c-myc pathway components are enriched in stem cells,

and c-myc is thought to act via the inhibition of the $p21^{CIP/WAF1}$ cell cycle inhibitor. SOX9 is expressed in proliferating, undifferentiated cells, and appears to repress the differentiation-inducing genes CDX2 and MUC2. The ephrin receptor EphB2 is expressed in proliferating cells, decreasing in a gradient towards the top of the crypts, whereas EphB3 expression is restricted to all cells below the stem cells, including Paneth cells. Ephrin-B1 and ephrin-B2 are expressed at the crypt-villus junction, decreasing toward the base of the crypt. This compartmentalization of Eph/Ephrin expression is thought to control the localization of Paneth cells as well as the separate placement of proliferating versus differentiated cells along the crypt-villus axis. In *EphB3* null mice, Paneth cells no longer migrate downwards, whereas in *EphB2/EphB3* double mutants, the proliferative and differentiated cells intermingle.

BMP signals also have a role in homeostasis of the intestine via inhibition of Wnt signaling. BMP4 is expressed in mesenchymal cells, including those adjacent to the stem cells, whereas the BMP receptor IA is expressed along the epithelial cells in the crypt-villus axis, being highly expressed in the stem cell region but excluded from the proliferating zone in the upper crypt region. Noggin, a negative regulator of BMP, is expressed in a region adjacent to the crypt bottom and in some cells in the stem cell region. Transgenic mice constitutively expressing *Xenopus* Noggin in the intestine have increased proliferation within crypts and *de novo* crypt formation, showing similar intestinal histopathology to human patients with the cancer predisposition syndrome, juvenile polyposis. Conditional inactivation of BMP receptor IA in adult mice results in inactivation of PTEN (a tumor suppressor gene-encoded inhibitor of phosphatidyl inositol [PI]-3 kinase) and localization of β catenin to the nucleus (where β catenin forms a complex with Tcf proteins to activate transcriptional targets), also with a resulting expansion of the proliferative region. Similarly, treatment with Noggin also results in inactivation of PTEN and nuclear transport of β-catenin in *ex vivo* cultures, and a PI-3 kinase inhibitor antagonizes this effect. These results suggest that BMP inhibits proliferation of intestinal cells via the inhibition of β-catenin activity downstream of Wnt signaling.

Other signals from the underlying mesenchyme are also important for stem cell function in the gut. For example, mutations in *Fkh6*, a gene expressed in the mesenchyme underlying mid- and hindgut, result in altered crypt morphology and overproliferation. In addition, keratinocyte growth factor (KGF, a member of the FGF family) is thought to stimulate proliferation of gut epithelial cells, and FGF receptor-3 is expressed in undifferentiated, proliferating cells during crypt formation and regeneration.

2.2.3 Drosophila Follicle Cells

Somatic cells called follicle cells are crucial for germline development in females. In Drosophila, follicle cells are produced throughout adult life from somatic stem cells (SSCs) located in the germarium (Fig. 2c). Lineage tracing studies have demonstrated that two SSCs are located near the junction between regions 2a (corresponding to the location of early 16-cell cysts) and 2b (corresponding to the location of lens-shaped cysts spanning the entire width of the germarium). As these stem cells divide, they give rise to daughter cells that envelop newly formed 16-cell germline cysts to form egg chambers. The follicle cells continue to divide until stage

7 of oogenesis with a doubling time of approximately 10 h, and with their division rate coordinated with the division of SSCs.

The Hedgehog (Hh) signaling molecule plays an essential role in SSC regulation. Hh is produced in terminal filament and cap cells in the anterior portion of the germarium, and directly stimulates SSC division. Inactivation of Hh using a temperature-sensitive *hh* allele disrupts egg chamber budding and, as a result, germline cysts accumulate within the germaria. Conversely, excessive Hh signaling induced by overexpression of an *hh* transgene or inactivation of the Hh receptor Patched (a negative regulator of the Hh signaling cascade) results in excessive generation of somatic cells. Lineage analysis suggests that the excessive generation of somatic cells results from an increased number of SSCs.

The Wingless (Wg) signal, also produced in terminal filament and cap cells, is required for SSC maintenance as well. Specific inactivation of downstream components of the Wg signaling pathway within SSCs results in stem cell loss, indicating that direct activation of stem cells by Wg molecules is necessary for stem cell maintenance. Conversely, overexpression of Wg or inactivation of its negative regulators results in excessive somatic cell production. In addition, overexpression of Wg results in a variety of differentiation defects.

Other factors are also required for SSC maintenance. *fs(1)Yb* is a gene involved in regulation of somatic and germline stem cell divisions via the regulation of *hh* expression. SSCs also require DE-cadherin for their maintenance, and they are in close contact with some of the inner germarial sheath cells, suggesting that these may form a stem cell niche.

2.3 Neural Stem Cells

The mammalian brain was thought to be an entirely postmitotic organ until regions containing proliferating cells were identified and shown to contain neural stem cells (NSCs). NSCs are present in the subventricular zone of the lateral ventricle and in the dentate gyrus subgranular zone of the hippocampus (Fig. 3). Surprisingly, astrocytes, originally thought to be in the astroglial lineage, function as NSCs within these neurogenic regions. Subventricular zone astrocytes give rise to transit amplifying cells, which in turn give rise to committed neuroblasts (Fig. 3a). These neuroblasts migrate to the olfactory bulb, where they differentiate into interneurons. Subgranular zone astrocytes give rise to progenitor cells that mature into granule cells (Fig. 3b), which integrate in the dentate gyrus granule cell layer and may play a role in learning and memory.

The behavior of NSCs is regulated by a microenvironment or niche composed of a basal lamina (rich in laminin and collagen-1), endothelial cells from the vasculature and their associated perivascular fibroblasts and macrophages, and, in the subventricular zone, ependymal cells lining the lateral ventricle. High levels of $\beta 1$ integrin are expressed on the surface of NSCs within neurospheres (aggregate cultures derived from single NSCs), suggesting that NSCs may associate with the basement membrane via an interaction between $\beta 1$ integrin and laminin. Interestingly, astrocytes also serve as niche cells, producing signals necessary to maintain their own NSC properties. This adult niche retains signaling molecules that are used during development of the nervous system.

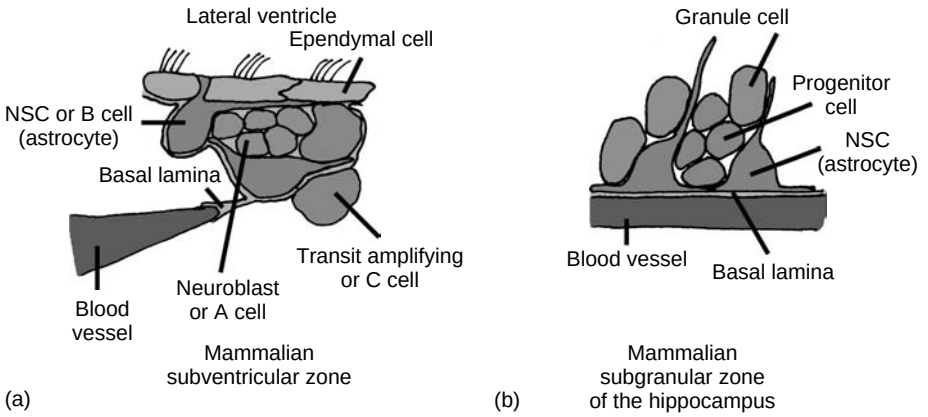

Fig. 3 Neural stem cells. (a) The subventricular zone of the adult mammalian brain contain astrocytes that function as neural stem cells (NSC), also called B cells. Some of these B cells contact the ventricle lumen, which is filled with cerebrospinal fluid and lined with ependymal cells. B cells generate transit amplifying C cells. C cells give rise to neuroblasts or A cells that migrate to the olfactory bulb and differentiate into interneurons. NSCs are regulated by a microenvironment created by astrocytes themselves (which have a dual role as NSC and niche cell), endothelial cells from local blood vessels, as well as associated fibroblasts and macrophages, and an intimately associated basal lamina. (b) The subgranular zone of the adult mammalian hippocampus also contains astrocytes that function as NSC. These astrocytes are in close proximity to blood vessels, which are presumably surrounded by an analogous basal lamina to the one found in the subventricular zone. Astrocytes divide to give rise to progenitor cells, which differentiate into granule cells that integrate nearby into the dentate gyrus.

NSCs are concentrated around blood vessels in the brain, and recent evidence points to endothelial cells, previously thought of as an exclusively structural component of blood vessels, as an NSC niche signaling component. *In vitro* studies have shown that endothelial cells produce diffusible factors that can promote self-renewal of NSCs, as well as prime them for differentiation into neurons and interneurons upon removal of such signals.

Many different signaling molecules are thought to contribute to maintaining the stem cell character of NSCs and regulating their neurogenic potential. Vascular endothelial growth factor (VEGF) receptors are expressed in neurogenic regions, and intraventricular VEGF infusions increase neuronal precursor proliferation. LeX, a carbohydrate moiety also known as CD15, is enriched in blood vessels and in a subpopulation of astrocytes in the subventricular and subgranular zones, and its expression is associated with subventricular zone cells that grow as neurosphere stem cells *in vitro*. Subventricular zone astrocytes also express ephrin B2/3 ligands, and intraventricular infusion of the EphB2 ligand results in increased astrocyte numbers. TGFα, an EGF receptor ligand, is expressed in the subventricular zone, and reduced TGFα activity results in decreased neuronal precursor proliferation, while activation of EGF receptors (by intraventricular EGF infusion) increases proliferation of transit amplifying cells. Genetically altered mice lacking the extracellular matrix protein tenascin C, which is normally highly expressed in the subventricular zone, have defective expression of EGF receptors in the subventricular zone

due to alterations in the response of NSCs to FGF2 and BMP4, and they also have altered numbers of NSCs that are prone to differentiation in culture.

Manipulation of Sonic hedgehog (Shh) activity *in vivo* and *in vitro* also suggests a role for this molecule in maintaining proliferation and neurogenic potential of either NSCs or the transit amplifying population. BMP signaling also occurs in regions with Shh activity, and it may also be important for maintenance of NSCs. Notch1 and its Jagged ligand are expressed in the subventricular and subgranular zones, and expression of an activated form of Notch in subventricular zone cells suppresses proliferation and differentiation of neuronal precursors. BMP and Notch activate the inhibitor of differentiation (Id) proteins in neuroepithelial cells, and this may be how they inhibit neural differentiation. Coculture of NSCs with endothelial cells stimulates neuroepithelial cell contact, activating Notch and Hes1 to promote self-renewal, suggesting that this mechanism may also operate *in vivo*. Ependymal cells in the subventricular zone produce Noggin (a negative regulator of BMP signaling), which helps regulate the niche. Wnt molecules, on the other hand, play a role in neuronal differentiation, as is the case during embryonic stem cell differentiation.

Niche signaling molecules presumably regulate intrinsic factors necessary for NSC activity. Bmi-1 and TLX are transcriptional factors that may play such a role. These proteins are expressed in the nervous system, and elimination of the function of either results in decreased proliferation of adult NSCs. Small, noncoding double-stranded (ds) RNAs play a critical role in neuronal differentiation. For instance, the NRSE/RE1 dsRNA binds to the NRSF/REST protein, which is a transcriptional repressor that inhibits neuronal gene expression in NSCs. This dsRNA-protein interaction alleviates this transcriptional repression, resulting in expression of neuronal genes and differentiation of NSCs. It is, however, unclear how the combination of signaling pathways regulates these and other factors.

2.4 Hematopoietic Stem Cells

Hematopoietic stem cells (HSCs) are among the best-studied stem cell systems. HSCs reside within the bone marrow inside the "major" bones (Fig. 4). About 0.05% of mouse bone marrow cells correspond to a lineage of long-term self-renewing HSCs (the "true" stem cells), which give rise to short-term self-renewing HSCs (equivalent to transit amplifying cells) that form multipotent progenitors without apparent self-renewing properties. Multipotent progenitors give rise to common lymphoid progenitors and common myeloid progenitors, which respectively give rise to all lymphoid and myeloid cells of the blood. Single long-term HSCs isolated from mice and humans are capable of regenerating the hematopoietic system of irradiated hosts through self-renewal and multilineage differentiation, and have been extensively used for therapeutic purposes.

Long-term HSCs are in close association, likely mediated by N-cadherin and β-catenin, with spindle-shaped osteoblasts lining the inner surface of trabecular bone in the marrow cavity (Fig. 4). The demonstration that these osteoblasts are a key component of the HSC niche came from the fact that variations in the number of osteoblasts directly correlate with variations in the number of HSCs. Inactivation of

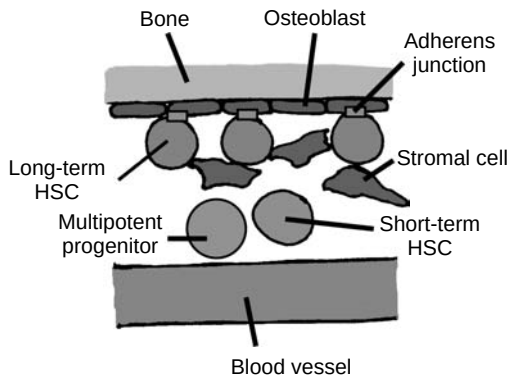

Fig. 4 Hematopoietic stem cells. Long-term hematopoietic stem cells (HSC), short-term HSCs, and multipotent progenitors reside in the bone marrow. Long-term HSCs are in close association with osteoblasts, present along the inner surface of the bone, and are surrounded by nearby stromal cells. Long-term HSC give rise to short-term HSC, which are analogous to transit amplifying cells. Short-term HSC have a limited self-renewal potential, and they generate multipotent progenitors that do not self-renew. Multipotent progenitors give rise to differentiated blood cells of the lymphoid and myeloid lineages.

BMP signaling in adult mice osteoblasts, by the conditional knockout of the BMP receptor type IA, results in a threefold increase in osteoblast number, presumably via reduced BMP-induced apoptosis. This increase in osteoblast number is accompanied by a twofold increase in HSC number, as demonstrated by functional transplantation assays, suggesting that BMP signaling controls the number of HSCs by limiting niche size.

Osteoblasts are a signaling center for HSCs. Transgenic mice expressing a constitutively active parathyroid hormone (PTH)/PTH-related protein receptor have increased numbers of osteoblasts producing elevated levels of the Notch ligand Jagged 1. In these transgenic mice, HSCs show high levels of Notch activation and are increased in numbers as well. These effects can be mimicked by treatment with PTH hormone. In culture, Jagged-1 activation of Notch results in increased progenitor activity, suggesting that the Notch pathway stimulates self-renewal and/or maintenance of multipotency.

Osteoblasts also express the ligand Angiopoietin-1 (Ang-1) that activates the receptor tyrosine kinase Tie2 on the surface of closely associated HSCs. Transplantation experiments in which levels of activity of Tie2 in HSCs are manipulated by Tie2 mutations or expression of Ang-1, demonstrate that increased Tie2 activity inhibits cell division and enhances adhesion of HSCs to bone. Tie2 signaling is thought to maintain long-term HSC self-renewing properties by inhibiting their proliferation, which is consistent with the fact that negative cell cycle regulators and polycomb group gene products such as Bmi-1 and Edr-1 may also be important regulators of HSC self-renewal.

Other signals are also important for the biology of HSCs. Their self-renewal depends on JAK2 signaling as well as activation of either c-kit (upon release of its ligand by the metalloprotease MMP9) or Flt-3. Shh is also thought to contribute to self-renewal of HSCs. Moreover, lipid modified Wnt molecules are expressed in bone marrow, and overexpression of activated β catenin, a downstream component of the Wnt signaling pathway, results in increased numbers of HSCs. Purified, lipid-modified Wnt molecules can also stimulate self-renewal of HSCs in culture. Conversely, overexpression of

axin, a negative regulator of Wnt signaling, results in inhibition of HSC proliferation and increased death. Cell death is thought to normally regulate the homeostasis of HSC, as supported by the fact that expression of bcl-2, a negative regulator of cell death, results in increased numbers of HSCs *in vivo*.

3 Embryonic Stem Cells

Between fertilization and implantation, the zygote will undergo many rounds of cell division to form an embryo containing three layers, namely, the trophectoderm, the primitive endoderm, and the inner cell mass (ICM) (Fig. 5). Embryonic tissues are exclusively derived from the ICM, whereas extraembryonic tissues arise from all three layers. The ICM cells, therefore, can undergo a great number of divisions and are pluripotent, giving rise to all tissues during embryonic development as they are exposed to the appropriate differentiation signals.

Remarkably, the ICM can be removed from early preimplantation mouse or human embryos and dissociate into what are called embryonic stem (ES) cells (Fig. 5). Under the appropriate signals, ES cells can be expanded in culture indefinitely (i.e. remain immortal), which is a unique property. For example, primary cultures of genetically normal skin fibroblasts can be established, but are short lived, whereas transformed cells are long lived and can be perpetually propagated in culture, but have many genetic aberrations.

ES cells also retain their pluripotency such that, if injected into a developing embryo, they can give rise to all different cell types and tissues. The ability to isolate and culture ES cells has revolutionized the use of mouse as a model system. Methodology was also developed for genetic modification of ES cells in culture that can be reintroduced into developing embryos. Because they can potentially contribute to any tissue in the animal, including the germline, this allowed the creation of genetically modified mouse strains, widely used in scientific research to study basic biological questions, to mimic and investigate human disease conditions, and to test drug targets.

ES cells can also be induced to differentiate into multiple cell types in culture. For instance, embryoid bodies containing

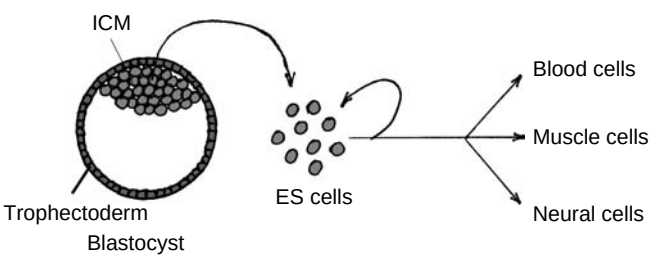

Fig. 5 Derivation of embryonic stem cells. Embryonic stem (ES) cells are derived from the inner cell mass (ICM) of the preimplantation embryo in the blastocyst stage. ES cells can be propagated in culture indefinitely as undifferentiated, pluripotent cells. If introduced back into embryos, ES cells can give rise to all cell types, and they can also be induced to differentiate in culture.

different embryonic tissues can form spontaneously if leukemia inhibitory factor (LIF) is removed. The ability to force ES cells to differentiate into embryoid bodies, or, in some cases, specific differentiated cells, has allowed the development of more cost-effective and faster assays for research, in addition to opening new avenues for the design of clinical approaches for tissue replacement and treatment of diseases. Mechanisms underlying directed differentiation of ES cells, however, are poorly understood.

Several intrinsic factors are known to be required for the extended life and pluripotency of ES cells, and these requirements reflect the natural embryonic environment from which these cells originate. Oct-4 is a transcription factor present in pluripotent cells, including the early preimplantation embryo (from which ES cells are derived), and primordial germ cells. It is essential for early embryonic development and becomes repressed during differentiation. Oct-4 requires cofactors such as the coactivator SOX2, and FGF4 and Hand1 are among its potential targets. The homeodomain-containing protein Nanog has a similar expression pattern to Oct-4, being expressed in preimplantation embryos (including ES cells) and primordial germ cells, and it is believed to be a pluripotency factor. A cluster of mouse ES cell-specific microRNAs, namely, miR-290, miR-291-as, miR-292-as, miR-293, miR-294, and miR-295 has recently been identified. The expression of these microRNAs is repressed during differentiation of ES cells into embryoid bodies, and they are also absent from adult mouse tissues. Similarly regulated microRNAs that are human ES cell-specific have also been identified, suggesting that mouse and human microRNAs may have a role in maintaining pluripotency of ES cells and regulating early embryonic development.

Recently, global gene-expression analyses of many human and mouse ES cell lines have been performed in an attempt to identify a pattern of gene expression that is common to all stem cells. Although these studies suggest that each ES line has a unique overall gene-expression profile, all ES lines were enriched for expression of Oct4 and Nanog. In addition, LeftyA and LeftyB, which encode members of the TGF-β superfamily, were also enriched in ES cells; further analysis will be required to establish whether they play a role in ES cell biology.

Extrinsic factors are also required for maintenance of ES cell properties. LIF is a cytokine that acts via the LIF receptor (LIFR) β coupled to the gp130 receptor, resulting in activation of the JAK-STAT pathway. Coupled with serum, LIF can stimulate long-term self-renewal of undifferentiated, Oct4-positive, mouse ES cells, but not human ES cells. In the absence of serum, cells undergo neural differentiation. Recently, BMPs 2 and 4 have been shown to promote self-renewal in combination with LIF, partially overcoming the serum requirement. BMPs act via activation of the Smad pathway and stimulation of Id proteins, which are negative bHLH factors that block neural differentiation.

As the regulation of ES cells is better understood, manipulation of ES cells will become technically less challenging and the balance between self-renewal and differentiation into specific paths more easily controllable to facilitate experimental as well as potentially clinical procedures. Importantly, given that differences are found between the regulation of mouse versus human ES cells in culture, it will be essential to carefully study human ES cell regulation and define optimal conditions for their

manipulation such that they can eventually be used for therapeutic purposes in the most effective and safe manner.

4
Themes of Stem Cell Regulation

Although no universally conserved signaling mechanism or intrinsic factor that controls stem cell identity and behavior has been found, common themes have emerged. For example, stem cells exist within specialized microenvironments where they are exposed to a unique combination of signals critical for regulating intrinsic factors to ensure their maintenance and proper division. The retention of stem cells within these niches is often achieved through the establishment of adherens junctions between niche cells and stem cells and/or via stem cell interactions with the extracellular matrix. Finally, stem cells are also subject to regulation by signals extrinsic to tissues in which they reside, such as hormones, nutrition, or physical insults.

4.1
Local Regulation – Stem Cell Niches

In 1978, Schofield first proposed on theoretical grounds the hypothesis that stem cells reside in specialized microenvironments, or niches, which generate an optimized set of signals that maintain stem cell properties. More recently, experimental data obtained from several different stem cell systems have demonstrated the existence of such niches and have begun to reveal their structural, organizational and mechanistic features that are essential for the maintenance and function of stem cells.

The details of how niches function vary from system to system; however, common themes exist (Fig. 6). Stem cells reside near a basal lamina with supporting cells forming the niche, such as Sertoli cells in the mammalian testis, hub cells in the Drosophila testis, cap cells and terminal filament cells in the Drosophila ovary, spindle-shaped osteoblasts in the bone marrow, and endothelial cells and astrocytes in the brain. Specialized junctions ensure the retention of stem cells within this niche. For instance, DE-cadherin- and β-catenin-containing adherens junctions are present between hub cells and GSCs in the Drosophila testis and between cap cells and GSCs in the Drosophila ovary, whereas N-cadherin- and β-catenin-containing adherens junctions occur between spindle-shaped osteoblasts and HSCs. In the cases of the mammalian gut crypt stem cells, epidermal stem cells, and GSCs in the mammalian testis, attachment to the basement membrane occurs via β1 integrins expressed on the stem cell surface.

Residence within niches endows stem cells with special properties due to their exclusive exposure to a unique combination of signals produced by niche cells. These signals activate evolutionarily conserved pathways that often operate, either similarly or in a distinct way, in multiple stem cell systems as illustrated by the BMP family of signaling molecules, which are part of the TGF β superfamily. BMP2 and 4 are required for self-renewal of ES cells via the stimulation of Id proteins, which are inhibitors of differentiation. Proliferation and maintenance of GSCs in the mammalian testis is stimulated by both BMP8B and Sertoli cell-secreted GDNF, which is a more distantly related TGF-β family member. Elevated levels of BMP signaling are also found in hair follicle

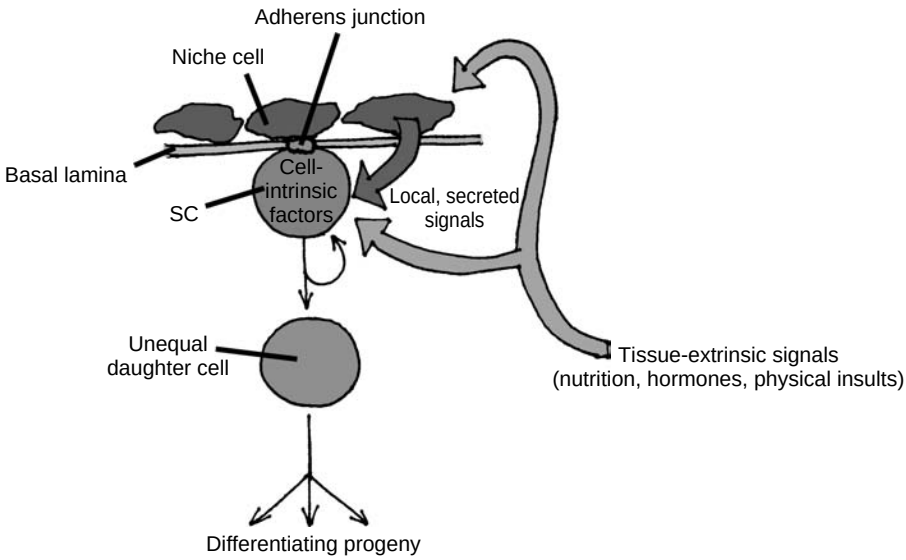

Fig. 6 Multilayered regulation of adult stem cells. Multiple intrinsic and extrinsic factors are required for the maintenance of stem cell (SC) properties and regulation of their function. Although the specific molecules may vary from system to system, common themes of SC regulation can be easily discerned. The local microenvironment (or niche) produces a set of signaling molecules that regulate SCs. SCs physically associate (for example, via adherens junctions) with cellular or extracellular components of the niche to ensure that they are permanently attached to and exposed to the unique signaling microenvironment that bestows them with SC properties. The locally secreted signals provided by the niche in turn regulate cell-intrinsic factors, which often represent global regulators of expression. The expression or silencing of specific intrinsic factors within SCs ensures that these cells keep their extended self-renewal capacity as well as their multipotency. The combined action of locally secreted and cell-intrinsic factors, therefore, ensures that SCs renew themselves and give rise to unequal daughter cells that differentiate into multiple cell types of a given tissue, ensuring the tissue's integrity and normal function. A tissue's need for new cells, however, is not always kept at steady state levels, and signals extrinsic to the tissue may change this demand. Such external signals may impinge on the normal regulation of SCs, either directly or via neighboring niche cells, to modify their rate of division and cell production. In addition, tissue-extrinsic signals may also modify the rate of division and death of the stem cell progeny.

bulge stem cells, as indicated by high levels of phosphorylated Smad2, and may also be important for NSC function. The BMP-like molecules Gbb and Dpp are produced by niche cells, and mediate GSC maintenance and proliferation in the Drosophila testis and ovary via inhibition of the Bam differentiation factor.

Other signals have important recurring roles in stem cell function. The LIF cytokine is essential for self-renewal of ES cells, an effect mediated via JAK-STAT activation, and JAK2 signaling is involved in regulating HSCs. The hub-produced Upd ligand also promotes GSC maintenance in the Drosophila testis via activation of the JAK-STAT pathway within these cells. The Notch signaling pathway is involved in maintenance of NSCs, HSCs, and GSCs within the distal end of the *C. elegans* gonad. The Hh pathway has conserved roles in stem cell function in

SSCs of the Drosophila ovary, NSCs in the mammalian brain, and HSCs. The Wnt pathway also has conserved roles in stem cell function. In the mammalian gut crypt, Wnt stimulates proliferation and inhibits differentiation of stem cells, and the Wg molecule plays a similar role in SSCs in the Drosophila ovary. In the mammalian hair follicle, Wnt signaling has a dual role, depending on the specific downstream effector. In bulge stem cells, increased Wnt signaling results in formation of a complex between stabilized β-catenin and Tcf3, a transcriptional repressor complex, whereas in the matrix, β-catenin and Lef1 form a transcriptional activator complex, resulting in distinct cellular behaviors. Multiple receptor tyrosine kinases also play roles in many stem cell systems.

4.2 Cell-intrinsic Factors

No specific and universal molecular marker for stem cells has been identified; however, several intrinsic factors have been shown to be required for stem cell maintenance and function (Fig. 6). The regulation of intrinsic factors by signals from the stem cell microenvironment provides a mechanism through which these signals control stem cell maintenance and function. These intrinsic factors are usually regulators of gene expression such as transcription factors or translational regulators, and the stem cell-related functions of some of these factors have been conserved across stem cell systems. For instance, the Oct-4 transcription factor is present in ES cells and undifferentiated spermatogonia in the mammalian testis, while Bmi-1 is involved in regulation of NSCs and HSCs. Upon activation of BMP signaling, the Mad/Med complex acts as a transcriptional repressor of the differentiation factor Bam in GSCs within Drosophila ovaries and testes. The translational inhibitors Nanos/NOS-3 and Pumilio/FBF-2 are important for GSC maintenance in Drosophila ovaries and *C. elegans* gonads. MicroRNAs are also emerging as intrinsic factors associated with mouse and human ES cells.

4.3 Tissue-extrinsic Factors

Despite the unquestionable importance of signals generated by the local stem cell environment in regulating stem cell properties and behavior, the activity of stem cells may also be influenced by stimuli originating outside the tissues in which they reside (Fig. 6). If a change in the demand for cells occurs because of external factors, such as, for instance, an injury or changes in nutrient levels, signals must be generated that are interpreted by stem cells, resulting in a change in their behavior.

Injury can have a dramatic effect on stem cell activity. Normally, skin homeostasis is maintained by the proliferation of stem cells within the basal layer of the epidermis, whereas hair follicle bulge stem cells generate daughter cells that migrate down to repopulate the matrix when signals from the dermal papilla are brought into close proximity because of the hair follicle cycle. In contrast, in severely burned patients who have lost the entire epidermis in large portions of their body, cells originating from the hair follicle can instead migrate up to reestablish the epidermis. Moreover, in mice carrying small area, full-thickness wounds, lineage tracing studies have demonstrated that bulge stem cells generate a large number of cells that migrate to the epidermis

to allow wound repair to take place. Similarly, brain injury (as well as mental and physical activity) results in increased neurogenesis in the dentate gyrus of the hippocampus. It remains unclear, however, what signals mediate a change in stem cell activity in response to physical stimuli in these systems.

Diet also has reported effects in several stem cell systems. Increased neurogenesis in response to caloric restriction has been reported in the dentate gyrus of the hippocampus of rats and mice, although this effect appears to be due to decreased death of newly produced cells rather than from increased proliferation. These effects appear to be partially mediated by brain-derived neurotrophic factor. Similarly, an increased number of long-term HSCs, as assessed by transplantation assays, have been reported in mice subjected to dietary restriction, although it is unclear how this effect is mediated or whether it occurs via modulation of cell death or proliferation.

The proliferation of stem cells can be controlled by nutritional status, as has been well documented in the Drosophila ovary for both germline and somatic stem cells. In females exposed to a high-protein diet, germline and somatic stem cells up-regulate their proliferation rates relative to those on a protein-poor diet in a coordinated fashion with their dividing progeny, and this process is reversible. In this case, similar levels of carbohydrates were provided to both groups of flies and they were, therefore, not under caloric restriction, which may explain the distinct effects of diet variations. The nutritional signal is relayed to the Drosophila ovary via insulin-like peptides, as demonstrated by a diminished response to nutritional status by mutants that are defective in insulin signaling. Parallels can be drawn to other systems. IGF-1 appears to stimulate HSC function as measured by transplantation assays. Mice in which diabetes is induced and which, therefore, have reduced levels of circulating insulin, show reduced levels of proliferation in the dentate gyrus subgranular zone. Furthermore, diabetes patients have reduced rates of wound healing, an effect that could be due in part to impaired stem cell behavior or proliferation.

Hormonal signals may also influence the activity of stem cells or differentiation of their progeny. In the bone marrow, PTH hormone appears to up-regulate the numbers of osteoblasts and the production of Jagged 1, thereby increasing long-term HSC numbers. In addition, growth hormone also promotes hematopoiesis via IGF-1 *in vivo*. Growth hormone also promotes higher rates of wound repair in patients with severe burns. Transgenic mice overexpressing growth hormone, however, show delayed and abnormal wound healing. In either case, it is unclear whether growth hormone exerts any direct effect on stem cells. In the mammalian testis, LH and FSH, produced in the pituitary gland, affect the local production of differentiation signals, although no direct effects on stem cells have been described. Administration of estrogen to diabetic mice restores proliferation in the dentate gyrus, suggesting that estrogens may normally play a role in NSC regulation.

Despite the many indications that tissue-extrinsic signals can modify stem cell activity, it is the least understood layer of regulation of stem cells. The identification of all extrinsic signals that modulate stem cells and elucidation of their mechanisms of action, which may involve directly signaling to stem cells or via their niches, remain areas of limited research and investment. A more complete understanding

of how external signals influence stem cell activity will be crucial, not only for establishing a thorough body of basic knowledge about stem cell regulation, but also to design better therapeutic approaches involving endogenous stem cells.

5 Implications

The study of stem cells in many diverse systems is the best approach for achieving a full understanding of their regulation. Many themes of stem cell regulation are conserved across tissues and organisms, and different systems offer distinct advantages, such as the detailed knowledge of the location of stem cell and niche components, a readily identifiable stem cell population, the availability of functional transplantation assays, the ability to monitor the behavior of individual stem cells *in vivo*, and/or the ability to recapitulate stem cell function *in vitro*. In recent years, immense progress has been made towards identifying stem cells and dissecting their regulation, and this work has broad implications for human health.

The study of normal stem cells has led to a revised model of cancer, in which a key event of tumorigenesis involves disruption of the regulation of stem cell self-renewal. In leukemias and solid tumors, only a small fraction of cancerous cells have extensive proliferative capacity that can give rise to new tumors, and these cells behave as hyperfunctioning stem cells. Indeed, several genes encoding proteins involved in the normal regulation of stem cells, such as Bmi-1, Notch, Wnt, and Hedgehog, are also involved in cancer formation. Thus, the knowledge obtained from further study of stem cell systems will likely continue to help elucidate the etiology of cancer.

Stem cells can also potentially be used therapeutically to replace damaged tissues in humans. This approach has been fairly successful for the replacement of long-term HSCs in cancer patients who have undergone high-dose chemotherapy. Bone marrow transplants containing stem cells can restore blood formation in patients and in lethally irradiated mice. The ability to enrich the population of cells introduced into patients or mice for HSCs on the basis of cell surface markers can greatly enhance the efficacy of this procedure such that approximately 200-fold fewer cells are needed. In addition, in cases of autologous bone marrow transplantation (i.e. when the bone marrow of a cancer patient is harvested prior to chemotherapy and later reintroduced into the patient to reestablish blood formation), the purification of HSCs away from potentially contaminating tumor cells can greatly enhance survival of the patients by reducing the recurrence of cancers. Likewise, therapeutic approaches using fetal tissues to replace damaged brain cells (e.g. dopaminergic neurons of Parkinson's disease patients) have been tried with limited success, underscoring the need for better knowledge of stem cell systems. Many other stem cell–based therapies are currently being developed; continued advancement in the stem cell field will be extremely valuable for their success.

Finally, the recent explosion in stem cell research has generated ethical debates. Some question the use of human embryonic stem cells in light of the fact that embryos are destroyed during the generation of ES cells and that other adult stem cell systems are available. Others argue that strict guidelines should indeed govern the generation and use of ES cells

(although not to the point of hindering research progress and its potential therapeutic benefits) and that, until effective stem cell-based therapies are available, all avenues of research involving embryonic and adult stem cells should be exploited.

See also Female Reproduction System, Molecular Biology of; Male Reproductive System: Testis Development and Spermatogenesis; Regulation of Stem Cell; Stem Cells; Stem Cells, Embryonic.

Bibliography

Books and Reviews

Alvarez-Buylla, A., Lim, D.A. (2004) For the long run: maintaining germinal niches in the adult brain, *Neuron* **41**, 683–686.

Brinster, R.L. (2002) Germline stem cell transplantation and transgenesis, *Science* **296**, 2174–2176.

Crittenden, S.L., Eckmann, C.R., Wang, L., Bernstein, D.S., Wickens, M., Kimble, J. (2003) Regulation of the mitosis/meiosis decision in the Caenorhabditis elegans germline, *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **358**, 1359–1362.

Fuchs, E., Tumbar, T., Guasch, G. (2004) Socializing with the neighbors: stem cells and their niche, *Cell* **116**, 769–778.

Gage, F.H. (2000) Mammalian neural stem cells, *Science* **287**, 1433–1438.

Kondo, M., Wagers, A.J., Manz, M.G., Prohaska, S.S., Scherer, D.C., Beilhack, G.F., Shizuru, J.A., Weissman, I.L. (2003) Biology of hematopoietic stem cells and progenitors: implications for clinical application, *Annu. Rev. Immunol.* **21**, 759–806.

Marshak, D.R., Gardner, R.L., Gottlieb, D. (2001) *Stem Cell Biology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Marshman, E., Booth, C., Potten, C.S. (2002) The intestinal epithelial stem cell, *Bioessays* **24**, 91–98.

Oving, I.M., Clevers, H.C. (2002) Molecular causes of colon cancer, *Eur. J. Clin. Invest.* **32**, 448–457.

Pyle, A.D., Donovan, P.J., Lock, L.F. (2004) Chipping away at 'stemness.', *Genome Biol.* **5**, 235–1–235–5.

Sadlon, T.J., Lewis, I.D., D'Andrea, R.J. (2004) BMP4: its role in development of the hematopoietic system and potential as a hematopoietic growth factor, *Stem Cells* **22**, 457–474.

Spradling, A., Drummond-Barbosa, D., Kai, T. (2001) Stem cells find their niche, *Nature* **414**, 98–104.

Weissman, I.L. (2000) Translating stem and progenitor cell biology to the clinic: barriers and opportunities, *Science* **287**, 1442–1446.

Wurmser, A.E., Palmer, T.D., Gage, F.H. (2004) Neuroscience. Cellular interactions in the stem cell niche, *Science* **304**, 1253–1255..

Primary Literature

Arai, F., Hirao, A., Ohmura, M., Sato, H., Matsuoka, S., Takubo, K., Ito, K., Koh, G.Y., Suda, T. (2004) Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche, *Cell* **118**, 149–161.

Asaoka, M., Lin, H. (2004) Germline stem cells in the Drosophila ovary descend from pole cells in the anterior region of the embryonic gonad, *Development* **131**, 5079–5089.

Batlle, E., Henderson, J.T., Beghtel, H., van den Born, M.M., Sancho, E., Huls, G., Meeldijk, J., Robertson, J., van de Wetering, M., Pawson, T., Clevers, H. (2002) Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB, *Cell* **111**, 251–263.

Blanpain, C., Lowry, W.E., Geoghegan, A., Polak, L., Fuchs, E. (2004) Self renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche, *Cell* **118**, 635–648.

Brawley, C., Matunis, E. (2004) Regeneration of male germline stem cells by spermatogonial dedifferentiation in vivo, *Science* **304**, 1331–1334.

Buaas, F.W., Kirsh, A.L., Sharma, M., McLean, D.J., Morris, J.L., Griswold, M.D., de Rooij, D.G., Braun, R.E. (2004) Plzf is required in adult male germ cells for stem cell self-renewal, *Nat. Genet.* **36**, 647–652.

Bunt, S.M., Hime, G.R. (2004) Ectopic activation of Dpp signalling in the male Drosophila germline inhibits germ cell differentiation, *Genesis* **39**, 84–93.

Calvi, L.M., Adams, G.B., Weibrecht, K.W., Weber, J.M., Olson, D.P., Knight, M.C., Martin, R.P., Schipani, E., Divieti, P., Bringhurst, F.R., Milner, L.A., Kronenberg, H.M., Scadden, D.T. (2003) Osteoblastic cells regulate the haematopoietic stem cell niche, *Nature* **425**, 841–846.

Campos, L.S., Leone, D.P., Relvas, J.B., Brakebusch, C., Fassler, R., Suter, U., ffrench-Constant, C. (2004) Beta1 integrins activate a MAPK signalling pathway in neural stem cells that contributes to their maintenance, *Development* **131**, 3433–3444.

Casanueva, M.O., Ferguson, E.L. (2004) Germline stem cell number in the Drosophila ovary is regulated by redundant mechanisms that control Dpp signaling, *Development* **131**, 1881–1890.

Chan, E.F., Gat, U., McNiff, J.M., Fuchs, E. (1999) A common human skin tumour is caused by activating mutations in beta-catenin, *Nat. Genet.* **21**, 410–413.

Chen, D., McKearin, D. (2003a) Dpp signaling silences bam transcription directly to establish asymmetric divisions of germline stem cells, *Curr. Biol.* **13**, 1786–1791.

Chen, D., McKearin, D.M. (2003b) A discrete transcriptional silencer in the bam gene determines asymmetric division of the Drosophila germline stem cell, *Development* **130**, 1159–1170.

Chen, J., Astle, C.M., Harrison, D.E. (2003) Hematopoietic senescence is postponed and hematopoietic stem cell function is enhanced by dietary restriction, *Exp. Hematol.* **31**, 1097–1103.

Cox, D.N., Chao, A., Lin, H. (2000) piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells, *Development* **127**, 503–514.

Crittenden, S.L., Bernstein, D.S., Bachorik, J.L., Thompson, B.E., Gallegos, M., Petcherski, A.G., Moulder, G., Barstead, R., Wickens, M., Kimble, J. (2002) A conserved RNA-binding protein controls germline stem cells in Caenorhabditis elegans, *Nature* **417**, 660–663.

Cruise, B.A., Xu, P., Hall, A.K. (2004) Wounds increase activin in skin and a vasoactive neuropeptide in sensory ganglia, *Dev. Biol.* **271**, 1–10.

DasGupta, R., Fuchs, E. (1999) Multiple roles for activated LEF/TCF transcription complexes during hair follicle development and differentiation, *Development* **126**, 4557–4568.

Deng, W., Lin, H. (2001) Asymmetric germ cell division and oocyte determination during Drosophila oogenesis, *Int. Rev. Cytol.* **203**, 93–138.

Doetsch, F., Caille, I., Lim, D.A., Garcia-Verdugo, J.M., Alvarez-Buylla, A. (1999) Subventricular zone astrocytes are neural stem cells in the adult mammalian brain, *Cell* **97**, 703–716.

Dor, Y., Brown, J., Martinez, O.I., Melton, D.A. (2004) Adult pancreatic beta-cells are formed by self-duplication rather than stem cell differentiation, *Nature* **429**, 41–46.

Drummond-Barbosa, D., Spradling, A.C. (2001) Stem cells and their progeny respond to nutritional changes during Drosophila oogenesis, *Dev. Biol.* **231**, 265–278.

Drummond-Barbosa, D., Spradling, A.C. (2004) Alpha-endosulfine, a potential regulator of insulin secretion, is required for adult tissue growth control in Drosophila, *Dev. Biol.* **266**, 310–321.

Eckmann, C.R., Crittenden, S.L., Suh, N., Kimble, J. (2004) GLD-3 and control of the mitosis/meiosis decision in the germline of Caenorhabditis elegans, *Genetics* **168**, 147–160.

Eriksson, P.S., Perfilieva, E., Bjork-Eriksson, T., Alborn, A.M., Nordborg, C., Peterson, D.A., Gage, F.H. (1998) Neurogenesis in the adult human hippocampus, *Nat. Med.* **4**, 1313–1317.

Forbes, A.J., Lin, H., Ingham, P.W., Spradling, A.C. (1996a) hedgehog is required for the proliferation and specification of ovarian somatic cells prior to egg chamber formation in Drosophila, *Development* **122**, 1125–1135.

Forbes, A.J., Spradling, A.C., Ingham, P.W., Lin, H. (1996b) The role of segment polarity genes during early oogenesis in Drosophila, *Development* **122**, 3283–3294.

Garcion, E., Halilagic, A., Faissner, A., ffrench-Constant, C. (2004) Generation of an environmental niche for neural stem cell development by the extracellular matrix molecule tenascin C, *Development* **131**, 3423–3432.

Gat, U., DasGupta, R., Degenstein, L., Fuchs, E. (1998) De novo hair follicle morphogenesis and hair tumors in mice expressing a truncated beta-catenin in skin, *Cell* **95**, 605–614.

Gilboa, L., Forbes, A., Tazuke, S.I., Fuller, M.T., Lehmann, R. (2003) Germline stem cell differentiation in Drosophila requires gap junctions and proceeds via an intermediate state, *Development* **130**, 6625–6634.

Hansen, D., Hubbard, E.J., Schedl, T. (2004a) Multi-pathway control of the proliferation versus meiotic development decision in the Caenorhabditis elegans germline, *Dev. Biol.* **268**, 342–357.

Hansen, D., Wilson-Berry, L., Dang, T., Schedl, T. (2004b) Control of the proliferation versus meiotic development decision in the C. elegans germline through regulation of GLD-1 protein accumulation, *Development* **131**, 93–104.

Hirabayashi, Y., Itoh, Y., Tabata, H., Nakajima, K., Akiyama, T., Masuyama, N., Gotoh, Y. (2004) The Wnt/beta-catenin pathway directs neuronal differentiation of cortical neural precursor cells, *Development* **131**, 2791–2801.

Houbaviy, H.B., Murray, M.F., Sharp, P.A. (2003) Embryonic stem cell-specific MicroRNAs, *Dev. Cell* **5**, 351–358.

Johnson, J., Canning, J., Kaneko, T., Pru, J.K., Tilly, J.L. (2004) Germline stem cells and follicular renewal in the postnatal mammalian ovary, *Nature* **428**, 145–150.

Kai, T., Spradling, A. (2003) An empty Drosophila stem cell niche reactivates the proliferation of ectopic cells, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4633–4638.

Kai, T., Spradling, A. (2004) Differentiating germ cells can revert into functional stem cells in Drosophila melanogaster ovaries, *Nature* **428**, 564–569.

Kauffman, T., Tran, J., DiNardo, S. (2003) Mutations in Nop60B, the Drosophila homolog of human dyskeratosis congenita 1, affect the maintenance of the germ-line stem cell lineage during spermatogenesis, *Dev. Biol.* **253**, 189–199.

Kawase, E., Wong, M.D., Ding, B.C., Xie, T. (2004) Gbb/Bmp signaling is essential for maintaining germline stem cells and for repressing bam transcription in the Drosophila testis, *Development* **131**, 1365–1375.

Kayahara, T., Sawada, M., Takaishi, S., Fukui, H., Seno, H., Fukuzawa, H., Suzuki, K., Hiai, H., Kageyama, R., Okano, H., Chiba, T. (2003) Candidate markers for stem and early progenitor cells, Musashi-1 and Hes1, are expressed in crypt base columnar cells of mouse small intestine, *FEBS Lett.* **535**, 131–135.

Kiger, A.A., White-Cooper, H., Fuller, M.T. (2000) Somatic support cells restrict germline stem cell self-renewal and promote differentiation, *Nature* **407**, 750–754.

Kiger, A.A., Jones, D.L., Schulz, C., Rogers, M.B., Fuller, M.T. (2001) Stem cell self-renewal specified by JAK-STAT activation in response to a support cell cue, *Science* **294**, 2542–2545.

King, F.J., Lin, H. (1999) Somatic signaling mediated by fs(1)Yb is essential for germline stem cell maintenance during Drosophila oogenesis, *Development* **126**, 1833–1844.

King, F.J., Szakmary, A., Cox, D.N., Lin, H. (2001) Yb modulates the divisions of both germline and somatic stem cells through piwi- and hh-mediated mechanisms in the Drosophila ovary, *Mol. Cell* **7**, 497–508.

Kobielak, K., Pasolli, H.A., Alonso, L., Polak, L., Fuchs, E. (2003) Defining BMP functions in the hair follicle by conditional ablation of BMP receptor IA, *J. Cell Biol.* **163**, 609–623.

Kraemer, B., Crittenden, S., Gallegos, M., Moulder, G., Barstead, R., Kimble, J., Wickens, M. (1999) NANOS-3 and FBF proteins physically interact to control the sperm-oocyte switch in Caenorhabditis elegans, *Curr. Biol.* **9**, 1009–1018.

Kubota, H., Avarbock, M.R., Brinster, R.L. (2004) Growth factors essential for self-renewal and expansion of mouse spermatogonial stem cells, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16489–16494.

Kuwabara, T., Hsieh, J., Nakashima, K., Taira, K., Gage, F.H. (2004) A small modulatory dsRNA specifies the fate of adult neural stem cells, *Cell* **116**, 779–793.

Lamont, L.B., Crittenden, S.L., Bernstein, D., Wickens, M., Kimble, J. (2004) FBF-1 and FBF-2 regulate the size of the mitotic region in the C. elegans germline, *Dev. Cell* **7**, 697–707.

Lavoie, C.A., Ohlstein, B., McKearin, D.M. (1999) Localization and function of Bam protein require the benign gonial cell neoplasm gene product, *Dev. Biol.* **212**, 405–413.

Lee, J., Duan, W., Long, J.M., Ingram, D.K., Mattson, M.P. (2000) Dietary restriction

increases the number of newly generated neural cells, and induces BDNF expression, in the dentate gyrus of rats, *J. Mol. Neurosci.* **15**, 99–108.

Margolis, J., Spradling, A. (1995) Identification and behavior of epithelial stem cells in the Drosophila ovary, *Development* **121**, 3797–3807.

McKearin, D., Ohlstein, B. (1995) A role for the Drosophila bag-of-marbles protein in the differentiation of cystoblasts from germline stem cells, *Development* **121**, 2937–2947.

McKearin, D.M., Spradling, A.C. (1990) bag-of-marbles: a Drosophila gene required to initiate both male and female gametogenesis, *Genes Dev.* **4**, 2242–2251.

Meng, X., Lindahl, M., Hyvönen, M.E., Parvinen, M., de Rooij, D.G., Hess, M.W., Raatikainen-Ahokas, A., Sainio, K., Rauvala, H., Lakso, M., Pichel, J.G., Westphal, H., Saarma, M., Sariola, H. (2000) Regulation of cell fate decision of undifferentiated spermatogonia by GDNF, *Science* **287**, 1489–1493.

Merkle, F.T., Tramontin, A.D., Garcia-Verdugo, J.M., Alvarez-Buylla, A. (2004) Radial glia give rise to adult neural stem cells in the subventricular zone, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17528–17532.

Morris, R.J., Liu, Y., Marles, L., Yang, Z., Trempus, C., Li, S., Lin, J.S., Sawicki, J.A., Cotsarelis, G. (2004) Capturing and profiling adult hair follicle stem cells, *Nat. Biotechnol.* **22**, 411–417.

Ohlstein, B., McKearin, D. (1997) Ectopic expression of the Drosophila Bam protein eliminates oogenic germline stem cells, *Development* **124**, 3651–3662.

Ohlstein, B., Lavoie, C.A., Vef, O., Gateff, E., McKearin, D.M. (2000) The Drosophila cystoblast differentiation factor, benign gonial cell neoplasm, is related to DExH-box proteins and interacts genetically with bag-of-marbles, *Genetics* **155**, 1809–1819.

Otero, J.J., Fu, W., Kan, L., Cuadra, A.E., Kessler, J.A. (2004) Beta-catenin signaling is required for neural differentiation of embryonic stem cells, *Development* **131**, 3545–3557.

Palmer, T.D., Takahashi, J., Gage, F.H. (1997) The adult rat hippocampus contains primordial neural stem cells, *Mol. Cell. Neurosci.* **8**, 389–404.

Park, I.K., He, Y., Lin, F., Laerum, O.D., Tian, Q., Bumgarner, R., Klug, C.A., Li, K., Kuhr, C., Doyle, M.J., Xie, T., Schummer, M., Sun, Y., Goldsmith, A., Clarke, M.F., Weissman, I.L., Hood, L., Li, L. (2002) Differential gene expression profiling of adult murine hematopoietic stem cells, *Blood* **99**, 488–498.

Park, I.K., Qian, D., Kiel, M., Becker, M.W., Pihalja, M., Weissman, I.L., Morrison, S.J., Clarke, M.F. (2003) Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells, *Nature* **423**, 302–305.

Potten, C.S., Booth, C., Tudor, G.L., Booth, D., Brady, G., Hurley, P., Ashton, G., Clarke, R., Sakakibara, S., Okano, H. (2003) Identification of a putative intestinal stem cell and early lineage marker; musashi-1, *Differentiation* **71**, 28–41.

Qiao, D., Zeeman, A.M., Deng, W., Looijenga, L.H., Lin, H. (2002) Molecular characterization of hiwi, a human member of the piwi gene family whose overexpression is correlated to seminomas, *Oncogene* **21**, 3988–3999.

Raghavan, S., Bauer, C., Mundschau, G., Li, Q., Fuchs, E. (2000) Conditional ablation of beta1 integrin in skin. Severe defects in epidermal proliferation, basement membrane formation, and hair follicle invagination, *J. Cell Biol.* **150**, 1149–1160.

Reya, T., Duncan, A.W., Ailles, L., Domen, J., Scherer, D.C., Willert, K., Hintz, L., Nusse, R., Weissman, I.L. (2003) A role for Wnt signalling in self-renewal of haematopoietic stem cells, *Nature* **423**, 409–414.

Saravia, F., Revsin, Y., Lux-Lantos, V., Beauquis, J., Homo-Delarche, F., De Nicola, A.F. (2004) Oestradiol restores cell proliferation in dentate gyrus and subventricular zone of streptozotocin-diabetic mice, *J. Neuroendocrinol.* **16**, 704–710.

Schonhoff, S.E., Giel-Moloney, M., Leiter, A.B. (2004) Neurogenin 3-expressing progenitor cells in the gastrointestinal tract differentiate into both endocrine and nonendocrine cell types, *Dev. Biol.* **270**, 443–454.

Schulz, C., Kiger, A.A., Tazuke, S.I., Yamashita, Y.M., Pantalena-Filho, L.C., Jones, D.L., Wood, C.G., Fuller, M.T. (2004) A misexpression screen reveals effects of bag-of-marbles and TGF beta class signaling on the Drosophila male germ-line stem cell lineage, *Genetics* **167**, 707–723.

Shen, Q., Goderie, S.K., Jin, L., Karanth, N., Sun, Y., Abramova, N., Vincent, P., Pumiglia, K., Temple, S. (2004) Endothelial

cells stimulate self-renewal and expand neurogenesis of neural stem cells, *Science* **304**, 1338–1340.

Shivdasani, A.A., Ingham, P.W. (2003) Regulation of stem cell maintenance and transit amplifying cell proliferation by tgf-beta signaling in Drosophila spermatogenesis, *Curr. Biol.* **13**, 2065–2072.

Song, X., Xie, T. (2002) DE-cadherin-mediated cell adhesion is essential for maintaining somatic stem cells in the Drosophila ovary, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14813–14818.

Song, X., Xie, T. (2003) Wingless signaling regulates the maintenance of ovarian somatic stem cells in Drosophila, *Development* **130**, 3259–3268.

Song, H., Stevens, C.F., Gage, F.H. (2002a) Astroglia induce neurogenesis from adult neural stem cells, *Nature* **417**, 39–44.

Song, X., Zhu, C.H., Doan, C., Xie, T. (2002b) Germline stem cells anchored by adherens junctions in the Drosophila ovary niches, *Science* **296**, 1855–1857.

Song, X., Wong, M.D., Kawase, E., Xi, R., Ding, B.C., McCarthy, J.J., Xie, T. (2004) Bmp signals from niche cells directly repress transcription of a differentiation-promoting gene, bag of marbles, in germline stem cells in the Drosophila ovary, *Development* **131**, 1353–1364.

Stappenbeck, T.S., Mills, J.C., Gordon, J.I. (2003) Molecular features of adult mouse small intestinal epithelial progenitors, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1004–1009.

Suh, M.R., Lee, Y., Kim, J.Y., Kim, S.K., Moon, S.H., Lee, J.Y., Cha, K.Y., Chung, H.M., Yoon, H.S., Moon, S.Y., et al. (2004) Human embryonic stem cells express a unique set of microRNAs, *Dev. Biol.* **270**, 488–498.

Taylor, G., Lehrer, M.S., Jensen, P.J., Sun, T.T., Lavker, R.M. (2000) Involvement of follicular stem cells in forming not only the follicle but also the epidermis, *Cell* **102**, 451–461.

Tazuke, S.I., Schulz, C., Gilboa, L., Fogarty, M., Mahowald, A.P., Guichet, A., Ephrussi, A., Wood, C.G., Lehmann, R., Fuller, M.T. (2002) A germline-specific gap junction protein required for survival of differentiating early germ cells, *Development* **129**, 2529–2539.

Tramontin, A.D., Garcia-Verdugo, J.M., Lim, D.A., Alvarez-Buylla, A. (2003) Postnatal development of radial glia and the ventricular zone (VZ): a continuum of the neural stem cell compartment, *Cereb. Cortex* **13**, 580–587.

Tulina, N., Matunis, E. (2001) Control of stem cell self-renewal in Drosophila spermatogenesis by JAK-STAT signaling, *Science* **294**, 2546–2549.

Tumbar, T., Guasch, G., Greco, V., Blanpain, C., Lowry, W.E., Rendl, M., Fuchs, E. (2004) Defining the epithelial stem cell niche in skin, *Science* **303**, 359–363.

van de Wetering, M., Sancho, E., Verweij, C., de Lau, W., Oving, I., Hurlstone, A., van der Horn, K., Batlle, E., Coudreuse, D., Haramis, A.P., Tjon-Pon-Fong, M., Moerer, P., van den Born, M., Soete, G., Pals, S., Eilers, M., Medema, R., Clevers, H. (2002) The beta catenin/ TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells, *Cell* **111**, 241–250.

Wang, Z., Lin, H. (2004) Nanos maintains germline stem cell self-renewal by preventing differentiation, *Science* **303**, 2016–2019.

Wang, L., Eckmann, C.R., Kadyk, L.C., Wickens, M., Kimble, J. (2002) A regulatory cytoplasmic poly(A) polymerase in Caenorhabditis elegans, *Nature* **419**, 312–316.

Willert, K., Brown, J.D., Danenberg, E., Duncan, A.W., Weissman, I.L., Reya, T., Yates, J.R., Nusse, R. 3rd (2003) Wnt proteins are lipid-modified and can act as stem cell growth factors, *Nature* **423**, 448–452.

Xie, T., Spradling, A.C. (1998) decapentaplegic is essential for the maintenance and division of germline stem cells in the Drosophila ovary, *Cell* **94**, 251–260.

Xie, T., Spradling, A.C. (2000) A niche maintaining germline stem cells in the Drosophila ovary, *Science* **290**, 328–330.

Yamashita, Y.M., Jones, D.L., Fuller, M.T. (2003) Orientation of asymmetric stem cell division by the APC tumor suppressor and centrosome, *Science* **301**, 1547–1550.

Ying, Q.L., Nichols, J., Chambers, I., Smith, A. (2003) BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3, *Cell* **115**, 281–292.

Yoshida, S., Takakura, A., Ohbo, K., Abe, K., Wakabayashi, J., Yamamoto, M., Suda, T., Nabeshima, Y. (2004) Neurogenin3 delineates

the earliest stages of spermatogenesis in the mouse testis, *Dev. Biol.* **269**, 447–458.

Zhang, Y., Kalderon, D. (2001) Hedgehog acts as a somatic stem cell factor in the Drosophila ovary, *Nature* **410**, 599–604.

Zhang, J., Niu, C., Ye, L., Huang, H., He, Xi., Tong, W.-G., Ross, J., Haug, J., Johnson, T., Feng, J.Q., Harris, S., Wiedemann, L.M., Mishina, Y., Li, L. (2003) Identification of the haematopoietic stem cell niche and control of the niche size, *Nature* **425**, 836–841.

Regulome

Miguel A. Andrade and Michael A. Rudnicki
Ottawa Health Research Institute, Ottawa, Canada

1 Introduction 100

2 Definition 101

3 Components 103

4 A Proposal for a High-throughput Analysis of the Regulome 104

5 Computational Challenges 105

6 Conclusions 105

Bibliography 106
Books and Reviews 106
Primary Literature 106

Keywords

Chromatin Immunoprecipitation
A technique for identifying specific DNA sequences that are bound, *in vivo*, to proteins of interest.

Cloned Mice
Mice produced asexually from a single cell.

DNA Binding
Selective interaction with DNA.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

Genomics
The systematic study of the complete DNA sequences (genome) of organisms.

High-throughput Techniques
Techniques that produce a large amount of data in a relatively short amount of time.

Proteomics
The systematic study of the complete complement of proteins (proteome) of organisms.

Transcription Factors
Proteins required to initiate or regulate transcription.

Transcriptional Complexes
Specific structures involved in transcribing the genome.

Transcriptional Regulation
Any process that modulates the frequency, rate, or extent of global transcription.

The regulome is defined as the complete set of regulons, where a regulon is the complete set of genes positively and negatively regulated by a specific regulatory protein. We propose the study of the regulome in mouse by analysis of 2500 known or predicted transcription factors, by generating one clone mouse with each TAP-tagged TF, followed by the study of the DNA binding sites of the TF by chromatin immunoprecipitation (ChIP) and the protein partners by proteomics, in several cellular types *in vitro*, and in different tissues and developmental stages *in vivo*. The integration of those data should provide the complete plan of mammalian transcriptional regulation.

1 Introduction

The completion of the human genome sequence is fundamental to paving the way for the complete understanding of human biology, but it is not the last step. The development of the concepts of transcriptome (the gene expression levels of all genes in the cell) and proteome (the concentrations of all proteins in the cell) were derived to define further layers that need to be described in order to explain a given cellular state. However, the mere description of the state of a cell does not explain how it works. A new layer integrating the levels of the genome, transcriptome, and proteome is needed to explain their interplay, which results in their dynamic change over time. This is the regulome, a concept coined in the year 2003 to describe the global set of regulatory molecular interactions that result in spatial–temporal variations of the cellular typology and state. The regulome includes the description of the interactions among genome, nucleotide and protein sequences, and metabolites,

occurring in different localizations of a cell, in different tissues, developmental stages, and possible states (e.g. normal versus pathological).

Ultimately, the description of the regulome of an organism would allow us to perfectly predict its transcriptome and proteome variation in different time periods, when the cellular initial state and the cellular environment are known. This accomplishment might be seen with skepticism today given the current state of the techniques and knowledge in molecular biology and notwithstanding possible limitations to the deterministic nature of the definition of regulome. However, the creation of the term regulome that, akin to those of transcriptome and proteome, is only putting preexisting concepts in a holistic perspective, is required by a technological revolution: the concept of regulome is born parallel to the possibility of technically measuring regulatory interactions at a genomic scale, the same way the concepts of transcriptome and proteome were born when high-throughput techniques to measure mRNA and protein concentrations were first possible. In this entry, we will describe our views about these developing techniques that will allow cellular regulatory relations to be measured at a scale that was never possible before. These new techniques will soon be modifying the way research is done in molecular biology the same way as proteomics and transcriptomics did in the 2000s, and genomics in the 1990s.

2 Definition

In prokaryotes, genes are organized and regulated in operons. Such structures do not exist in eukaryotes that rely on coregulation by the binding of thousands of transcription factors (TFs) to tens of thousands of binding sites in the genome. The complexity of transcriptional regulation grows in parallel with the complexity of the organism along with evolution. The more functions that have to be carried out, the more the elements that have to be controlled, that is, more compartments, developmental constraints, stem cells (tissue renewal and maintenance is necessary as the life cycle of the organism increases), interactions in a multicellular environment, and so on. All these elements appear as organism complexity increases from bacteria, to vertebrates, to mammals, and require a similar increment in the complexity of transcriptional regulation.

One of the main features of transcriptional regulation is the interaction of proteins (TFs) to DNA to regulate the transcription of nearby genes. Although it is possible to describe the patterns of the sequences to which TFs bind, it is obvious that we do not as yet understand the specificity of the interaction very well since such sequences exist often in high numbers all over the genome but, TFs do not bind everywhere. Recent experiments indicate that chromatin organization is dynamic and is subject to regulatory mechanisms that enforce the transcriptional potential of the genome during cellular commitment and differentiation. Chromatin is remodeled into transcriptionally permissive or repressive conformations by complexes that either covalently modify histones, or act in an ATP-dependent manner to reposition nucleosomes along DNA. Hence, chromatin structure restricts the availability of TF binding sites within different cell types at different developmental stages. Therefore, TF binding sites must be experimentally determined within each cell type rather than mapped using *in silico* analysis of genomic DNA sequence information.

Another element of transcriptional regulation is the interaction of TFs with other proteins in transcriptional complexes that modulate their actions. Therefore, understanding of the protein–protein interactions of TFs is also very important. Again, computational predictions of these interactions are not good enough and experimental analysis is needed.

Transcriptional regulation also involves other more indirect mechanisms, such as protein posttranslational modifications, small RNAs with capabilities of blocking or guiding the transport of coding mRNA, yet the most important point is controlling the production of mRNA from genes and also controlling which mRNA is produced (alternative splicing) (See Fig. 1). We think that the regulome can be described by analysis of just the TFs (as opposed to, for example, posttranslational modifications, or metabolites) because transcription is the hub through which all transcriptional regulation goes. Metabolites can be considered as effectors, and some proteins (such as receptors) as sensors of that process. Eventually, proteins are implied in those related processes and their production too is controlled by transcriptional regulation.

In this entry, we will present one proposal for studying the regulome. It is only one of many possibilities, and emanates from the thoughts revolving around the generation of the International Regulome Consortium (IRC), a liaison of research groups in Canada, UK, France, Singapore,

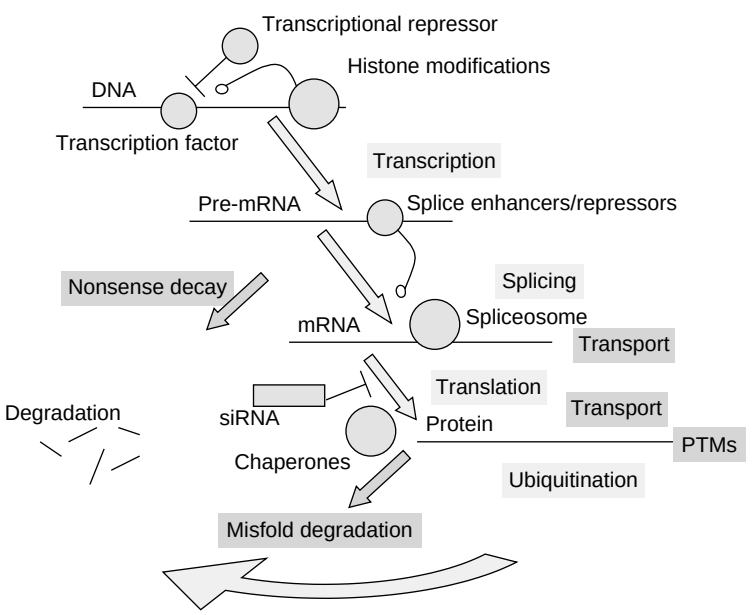

Fig. 1 The regulation of transcription is just one piece in the control of the expression of a genome in a phenotype, but it is situated right at the top of it. For example, posttranslational modifications of proteins control their function and cellular location, but those modifications are carried out by other proteins that are generated from genes ultimately controlled by TFs. We distinguish the control mechanism (transcription regulation) from its effectors (e.g. PTMs) or sensors (e.g. ion concentration).

and Australia, created in 2004 with the objective of joining efforts in the study of the regulome. For additional information concerning the mission and history of the IRC, please see the IRC Web portal http://www.internationalregulomeconsortium.ca/ and the IRC's 2004 white paper (http://www.internationalregulomeconsortium.ca/documents/IRC%20White%20Paper%20final%20V1.pdf).

3 Components

Although it is virtually impossible to check all possible conditions and combinations under which cells perform, and the system responding to them, a cell with its collection of biologically active macromolecules is finite. A number of measurements on the elements of the system with a large enough number of conditions should be enough to reveal their dependencies. The question is which of those molecules describe the regulome.

We propose to center the study on mouse as a model organism and on TFs. First, we have to define the genes that will be the object of the study. Since most genes, including all protein-coding genes, are transcribed by RNA polymerase II (RNAPII), the regulome can be defined by the RNAPII TFs that are expressed and the sites to which they bind in specific differentiated cell types and tissues.

There are a number of protein domains that have been seen as indicative of transcription factors, for example, DNA binding domains, but there are others too. Preliminary sequence analysis that was done by Timothy Hughes at the University of Toronto by scanning mouse genes for these protein domains against the Pfam and SMART databases of protein domains revealed more than 2200 mouse genes. Not all will be TFs but they give us a starting point.

We contrasted this list with microarray data from mouse cells displaying particularly rich patterns of expression: embryonic stem cells and their derivatives. Around 400 genes were not included in the set of genes detected by the microarray. Another 1000 were found to be expressed. Therefore, we estimated a number of 1400 targets.

How many binding sites are we considering? Any prediction will be an underestimation since it is not possible to discover absolutely all relevant binding sites in all possible conditions (e.g. pathological cases add further aberrant binding sites). However, to put our proposal in the context of current knowledge, we can carry out an estimation on the basis of well-known sets. For example, in a study of the evolution of TF binding sites, the number of binding sites described in the promoters of 50 genes is 186, so a guess of an average of 4 binding sites per gene looks reasonable. Given 30 000 genes in the mouse genome (another guess) we have 120 000 binding sites and possibly an average of 100 binding sites per TF. For a comparison, the current public version of the TRANSFAC database (16 Feb 2005, release 6.0) contains 911 mouse binding sites and 718 TFs, which is slightly above one binding site per TF.

How many transcriptional complexes are there? We examined the set of protein complexes in mouse annotated as DNA binding and implied in transcription in the BIND database of protein–protein interactions. Only 31 complexes are currently described (as of 16 Feb 2005) with a total of 103 partners. The average of subunits per complex is just above three. If we assume at least one complex per TF, it

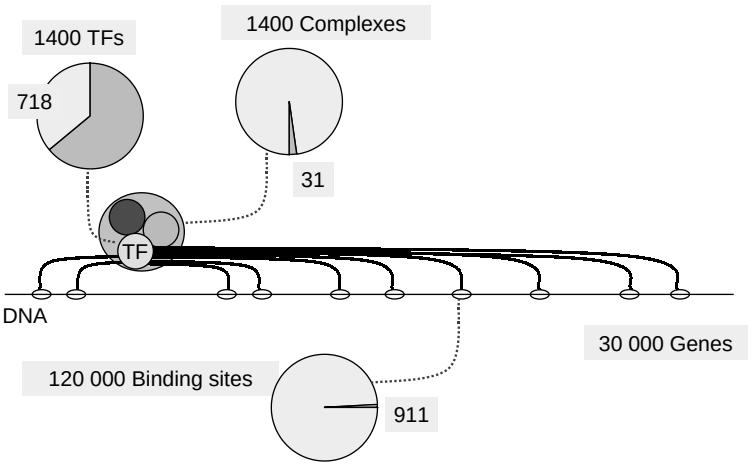

Fig. 2 Estimation of the amount of elements integrating the mouse regulome and of the fraction of those currently identified (darker sectors in pie charts).

means that we would be looking at 1400 protein complexes, describing about 4200 protein-to-protein interactions.

Additionally, in order to be able to have a good survey of TF modes of action and interactions, it is not realistic to scan for less than a hundred conditions (tissues, cells, developmental stages) per TF. Note that if we take TRANSFAC and BIND as a representation of the current knowledge of TFs in mouse, it means that only half of all TFs is characterized (718 of 1400), only 2% of the complexes they form (31 of 1400), and less than 1% of their binding sites (911 of 120 000). This indicates that there is a long way to go in the study of the regulome and, together with the need for testing hundreds of conditions, justifies a high-throughput analysis (see Fig. 2).

4 A Proposal for a High-throughput Analysis of the Regulome

Given a set of putative TFs in the mouse genome we propose to: create null mutations in all TFs in mouse embryonic stem cells (mESC) and characterize the phenotype and gene expression profile *in vivo* and *in vitro*; create mESC lines with dual ''knocked-in'' affinity tags at the C terminals of 1400 RNAPII TFs; affinity purify the tagged proteins from undifferentiated and differentiated mESC and various adult mouse tissues in which they are expressed; identify interacting proteins by SDS-PAGE, MALDI-TOF MS, and LC-MS/MS; and identify TF binding sites by chromatin immunoprecipitation (ChIP).

We propose to generate a clone mouse line for each of the 1400 target TFs. Each target gene will be inserted in a bacterial artificial chromosome (BAC) subsequently targeted in mESC. Hybrid mESC lines will be used to generate targeted clones. Use of hybrid mESC lines dramatically increases the frequency of survival to adulthood of mice generated by mESC-tetraploid embryo aggregation. Following excision of the selection cassette, the gene will be modified by the insertion of the tag. We propose the generation of mice for *in vivo* analyses,

using mESC-tetraploid embryo aggregation by fusion of two-cell-stage embryos and aggregation to targeted hybrid mESC following standard protocols. Embryos that develop into blastocysts are then transferred to the uterus of surrogate mothers and mice derived entirely from the targeted mESC are born 17 days later.

The association of TFs or TF complexes with DNA *in vivo* can be analyzed by using the ChIP procedure. Briefly, proteins are cross-linked to DNA by treating living cells with formaldehyde and then antibodies are used to precipitate specific proteins from sonicated chromatin preparations. In such experiments, specific genes or regions of genes that are coprecipitated with the protein are amplified and quantified by PCR. Depending on the number of PCRs, the approach can locate the binding site for a given factor on a given gene to within 50 bp or less.

TFs are regulated by forming part of transcriptional complexes with other proteins. For example, they may need to bind a protein in order to bind to DNA. Therefore, an important part of assessing the regulome is the elucidation of the protein partners of TFs. Of the many experimental methods that have been developed for the identification of protein–protein interactions, one of the most appropriate for systematic analysis involves expressing a tagged protein at its normal concentration *in vivo* so as to avoid nonphysiological interactions. This is used to retrieve and purify the complexes interacting with the tagged protein. Then, standard proteomic methods can be applied to detect the protein components of the complex. In particular, in the TAP method, two affinity tags, protein A from *Staphylococcus aureus* and a calmodulin-binding peptide (CBP) are placed at the C terminals of the proteins to be analyzed.

5 Computational Challenges

The approach presented here to the study of the regulome will produce details of the DNA binding sites and protein partners of thousands of TFs in a variety of cells and conditions. There are three major computational challenges posed by that data. The first two are the storage/dissemination of the data and the integration of the data with existing databases. These are the basis for the third and most important computational task, which is the translation of the data into predictions evaluated with a score indicating the confidence of the prediction. This requires the interpretation of the data produced above and to model the results according to a regulatory network. Such a step is necessary in order to suggest experiments to verify the most interesting predictions (according to their novelty, relevancy, or confidence).

6 Conclusions

The generation of data of complete regulomes of several organisms will happen in the following years. This is likely to be the next biological data revolution, which is more experimentally complex to obtain than sequencing, or measuring mRNA levels by DNA microarrays, but also richer in content.

The total knowledge of a biological entity brings an increase in the possibilities for the use of that information that is orders of magnitude richer than a partial or even nearly complete knowledge of that entity. This is seen with the sequencing of complete genomes, which allows the knowledge of complete sets of genes in

bacterial organisms and to elaborate complex inferences about their metabolism, even in the absolute absence of any experimental characterization on any protein of that organism. Identically, the knowledge of the regulome in a number of organisms will result in the complete understanding of the cellular control, including maintenance, development, and the origin of pathologic states.

The complete knowledge of the regulome will unleash the next data revolution. This will probably be the use of dynamic models of complete eukaryotic cells that now must work on partial knowledge of the regulatory network and therefore with a very limited predictive power. This is a necessary step before the modeling of multicellular ensembles, and the modeling of tissue differentiation, stem cell, and cancer-related processes at cellular scales.

The data obtained on the regulome rely on many pieces of information that are prone to change over time, such as methods of data interpretation using databases and those databases themselves. This requires the storage of the data and its analysis in a dynamic form in yet another database in which the data is not just collected but continuously reevaluated, as databases and methods improve over time. The challenge is to convince molecular biologists to get used to looking at data in these repositories, and to display the data in ways that make it easy for them to understand the data and get insights for the design of relevant experiments. In this way, we can expect that the study of the regulome will stimulate fundamental and translational research and provide new insights in understanding the genomics of cell function that are critical for developing novel therapies that pave the way to improved health benefits for society.

See also Genetics, Molecular Basis of; Heterochromatin and Eurochromatin – Organization, Packaging, and Gene Regulation; Plant Gene Expression, Regulation of; RecA Superfamily Proteins; Spatial and Temporal Expression Patterns in Animals; Variation and Conservation of the Development of Animal Bodies: Evolutionary Developmental Biology.

Bibliography

Books and Reviews

Li, E. (2002) Chromatin modification and epigenetic reprogramming in mammalian development, *Nat. Rev. Genet.* **3**(9), 662–673.

Nagy, A. (2003) *Manipulating the Mouse Embryo; A Laboratory Manual*, 3rd edition, Cold Spring Harbor Laboratory Press, New York.

Woodcock, C.L., Dimitrov, S. (2001) Higher-order structure of chromatin and chromosomes, *Curr. Opin. Genet. Dev.* **11**(2), 130–135.

Primary Literature

Alfarano, C., et al. (2005) The biomolecular interaction network database and related tools 2005 update, *Nucleic Acids Res.* **33**, D418–D424.

Bateman, A., et al. (2004) The Pfam protein families database, *Nucleic Acids Res.* **32**(Database issue), D138–D141.

Dermitzakis, E.T., Clark, A.G. (2002) Evolution of transcription factor binding sites in mammalian gene regulatory regions: conservation and turnover, *Mol. Biol. Evol.* **19**(7), 1114–1121.

Eggan, K., et al. (2001) Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation, *Proc. Natl. Acad. Sci. U.S.A.* **98**(11), 6209–6214.

Harbison, C.T., et al. (2004) Transcriptional regulatory code of a eukaryotic genome, *Nature* **431**(7004), 99–104.

Horn, P.J., Peterson, C.L. (2002) Molecular biology. Chromatin higher order folding: wrapping up transcription, *Science* **297**(5588), 1824–1827.

Jenuwein, T., Allis, C.D. (2001) Translating the histone code, *Science* **293**(5532), 1074–1080.

Krogan, N.J., et al. (2002) RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach, *Mol. Cell. Biol.* **22**(20), 6979–6992

Krogan, N.J., et al. (2003a) A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1, *Mol. Cell* **12**(6), 1565–1576.

Krogan, N.J., et al. (2003b) Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II, *Mol. Cell. Biol.* **23**(12), 4207–4218.

Letunic, I., et al. (2004) SMART 4.0: towards genomic data integration, *Nucleic Acids Res.* **32**(Database issue), D142–D144.

Matys, V., et al. (2003) TRANSFAC: transcriptional regulation, from patterns to profiles, *Nucleic Acids Res.* **31**(1), 374–378.

Rigaut, G., et al. (1999) A generic protein purification method for protein complex characterization and proteome exploration, *Nat. Biotechnol.* **17**(10), 1030–1032.

Schwenk, F., et al. (2003) Hybrid embryonic stem cell-derived tetraploid mice show apparently normal morphological, physiological, and neurological characteristics, *Mol. Cell. Biol.* **23**(11), 3982–3989.

Wade, J.T., Struhl, K. (2004) Association of RNA polymerase with transcribed regions in Escherichia coli, *Proc. Natl. Acad. Sci. U.S.A.* **101**(51), 17777–17782.

Wang, Z., Jaenisch, R. (2004) At most three ES cells contribute to the somatic lineages of chimeric mice and of mice produced by ES-tetraploid complementation, *Dev. Biol.* **275**(1), 192–201.

Yugi, K., Tomita, M. (2004) A general computational model of mitochondrial metabolism in a whole organelle scale, *Bioinformatics* **20**(11), 1795–1796.

Repair and Mutagenesis of DNA

Raymond Devoret
Institute Curie, Orsay, France

1 The Structure of DNA Determines its Functions in Repair and Mutagenesis 111
1.1 When Donated by Both Parents, the CFTR Mutations are Deleterious 112

2 DNA-damaging Agents are Mutagens 114
2.1 A DNA Lesion is Different from a Mutation 114

3 Characteristic Mutation Types 114

4 Silent Base Substitutions 117
4.1 Synonymous Mutations are Neutral 118

5 Four Main Mechanisms that Generate Spontaneous Mutations 119

6 Induced Mutations 120

7 Mutagenesis in Bacteria and Carcinogenesis in Mammals 121

8 Carcinogens Leave No Definite Mutational Signatures on DNA 122

9 Humans Use Mutagenesis to Produce their Immunity Repertoire 122

Acknowledgments 123

Bibliography 123
Books and Reviews 123
Primary Literature 123

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

Keywords

DNA-damaging Agents
Physical or chemical agents, for example, such as Xrays or aflatoxin B_1, which affect DNA functions by distorting its structure. The newly formed structure is often termed *DNA lesion*, which often precedes the appearance of a mutation. Note that the transmogrification of a DNA lesion may give rise to a mutation. Yet, note that a lesion is not a mutation.

DNA Repair
A collective term that encompasses at least three dozens of biochemical processes that are involved in restoring the structure of DNA after it has sustained damage.

Gene
A segment of DNA, the expression of which results in the production of a messenger RNA (mRNA) that translates into a protein.

Genetic
Qualifies what pertains to DNA, whereas, *hereditary* specifies a trait passed from a parent to its progeny. The Down syndrome is a genetic disease, resulting from a triplication of human chromosome 21. In contrast, mucoviscidosis is a hereditary disease caused by a mutation located in human chromosome 7.

Genome
Designates the whole set of chromosomes present in a cell.

Mutagens
Physical or chemical agents that alter DNA, increasing the occurrence of mutations.

Mutagenesis
Designates the overall process that results in the induction of a mutation.

Mutation
A DNA change that is recorded durably and passed on to the offspring. *Induced mutations* result from the error-prone repair of DNA lesions or from DNA replication errors. *Spontaneous mutations* correspond to naturally occurring DNA changes.

SOS Mutagenic Repair
In prokaryotes, it is an induced error-prone DNA replication, involving a specialized polymerase that causes the synthesis of a base across a lesion (termed TLS, standing for trans-lesion DNA synthesis). TLS is a major SOS function regulated by the activation of RecA protein.

During their lifetime, all the cells of a living organism are exposed constantly to harmful agents such as radiations and chemicals. DNA (deoxyribonucleic acid), a crucial cell constituent, is highly sensitive to deleterious agents. A large majority of cells survive when exposed chronically at low doses since DNA damage induces cell enzymes that repair DNA lesions. Repair acts very efficiently when DNA stops replicating after the occurrence of lesions. After repair has taken place, DNA replication resumes, followed by cell division. While various repair processes concur to restore cell survival, each of them is never completely efficient and lesions remain in the DNA, acting as stepping-stones. Resumption of DNA synthesis on the strand opposite the remaining lesions permits the bypassing of these obstacles. A process, called *trans-lesion DNA synthesis* is one of the most efficient processes generating mutations. Mutations come into view in cells that survive the DNA-damaging treatment.

In short, a mutation is a DNA change that is recorded durably and passed on to the offspring after replication. A number of mutations in *germ* cells have *harmful* effects for they often cause hereditary diseases. Similarly, mutations that accumulate in dividing body cells (*somatic cells*) can generate cancers. Yet, most of the occurring mutations behave as *neutral* in their usual environment. Selection forces the genome to retain the *beneficial* mutations.

1 The Structure of DNA Determines its Functions in Repair and Mutagenesis

DNA is the essential component of chromosomes in all cells. Each chromosome is made of an extremely long DNA molecule with two intertwined *complementary strands* (a double helix). Each strand consists of a poly(phosphate sugar) backbone from which project Adenine (A), Cytosine (C), Guanine (G), and Thymine (T) bases at each sugar. The two strands form a *duplex* maintained by hydrogen bonds between complementary bases: A binds to T and G to C. The four bases A, C, G, or T may combine into a linear chain that constitutes a *genetic sequence* (see Fig. 1).

DNA of the whole set of chromosomes in a cell is called the *genome*. A bacterium like *Escherichia coli*, a natural resident of the colon of mammals, is made of one chromosome, whereas human cells carry 23 chromosome pairs. In general, the cells of complex organisms like mammals have more chromosomes per cell than lower organisms like microbes.

The double-stranded structure of DNA permits its duplication, one strand providing the template for the making of its copy. Not only has such a structure the possibility of replicating but also of repairing itself. If, for example, ultraviolet light hits a strand of the DNA double helix, the damaged piece undergoes: (1) excision of the damaged DNA piece that leaves a gap; and (2) DNA resynthesis to close the gap, using the undamaged strand as a template to make a correct new strand.

Repair processes are many, involving more than fifty identified proteins. The three major processes that repair DNA are (1) *excision*; (2) *recombination*; and (3) *mismatch repair*.

```
(a)   AAGGCAAACCTACTGGTCTTATGT   original sequence

(b)   AAGGCAAATCTACTGGTCTTATGT   transition

(c)   AAGGCAAACCTACTGCTCTTATGT   transversion

          Del(ACCTA)
(d)   AAGGCAA^CTGGTCTTATGT   deletion of ACCTA

                    Ins(AAAGC)
(e)   AAGGCAAACCTACT^AAAGC^GGTCTTATGT insertion of AAAGC
```

Fig. 1 Types of mutations: (a) wild type original sequence; (b) transition from C to **T** (underlined bold T); (c) transversion from G to **C** (underlined bold C); (d) deletion (Del) of the sequence **ACCTA**, the ^ sign indicates from where it has been taken out; and (e) insertion (Ins) of the sequence **AAAGC**, the two ^ signs indicate where the sequence has been put in.

There are two types of excision repair depending on the length of excised DNA: nucleotide excision repair (NER) or base excision repair (BER). Various different types of recombination repair exist depending on whether the recombining DNAs are *homologous, nonhomologous,* or *nonhomologous end joining*. More complicated repair processes use a combination of all of these. Repair, by definition, should lead to the full restoration of the original DNA structure. In fact, almost all repair processes leave a low percentage of errors, some of which give rise to mutations.

The main function of DNA is to code for the hereditary traits that are passed from parents to progeny. Recently, geneticists have identified the long string of bases present in many genomes, from bacteria to man. The mapping of the human genome has resulted in building a compendium of the number and positioning of the 7000 or so genes, whose mutations are responsible for human hereditary diseases.

When a mutation occurs within a gene, the protein encoded by the gene is often altered (see Figs. 1b, c and 2c); this alteration may produce a visible change (*phenotype*) of the organism under study. The actual mutation itself (*genotype)* is invisible to the naked eye, whereas the technique of DNA sequencing will locate it precisely.

Since a mutation is a DNA change, one should define it in relation to a preceding, parental DNA sequence (Fig. 1a). Consequently, geneticists have agreed to define a *standard genetic sequence* for each organism studied. The adopted standard is usually the most widespread type that fits in its environment.

Mutations fall into three classes termed *deleterious, neutral,* or *beneficial* according to their effects on the organism in which they arose. These qualifications are only valid in relation to a given environment.

1.1 When Donated by Both Parents, the CFTR Mutations are Deleterious

Human cells carry 22 pairs of chromosomes called *autosomes* plus a pair of sex chromosomes. The probability for a child to suffer from *cystic fibrosis* is one out of 1600 children. This disease is dramatic; the fluidity of lung secretions disappear, stifling normal pulmonary aeration. The cause of the disease is genetic: the affected child has received from the two

```
(a) Ile  Cys  Ile  Lys  Ala  Leu  Val  Leu  Leu  Thr amino
    acid  sequence
    ATA  TGT  ATA  AAG  GCA  CTG  GTC  CTG  TTA  ACA DNA
    sequence

    ATA  TGT  ATA  AAG  GCA  CTG  GTA  CTG  TTA  ACA base
    substitution
    Ile  Cys  Ile  Lys  Ala  Leu  Val  Leu  Leu  Thr synonymous
    amino acid

(b) Ile  Cys  Ile  Lys  Ala  Asn  Val  Leu  Leu  Thr amino
    acid sequence
    ATA  TGT  ATA  AAG  GCA  AAC  GTC  CTG  TTA  ACA DNA
    sequence

    ATA  TGT  ATA  AAG  GCA  AAC  TTC  CTG  TTA  ACA base
    substitution
    Ile  Cys  Ile  Lys  Ala  Asn  Phe  Leu  Leu  Thr missense
    amino acid

(c) Ile  Cys  Ile  Lys  Ala  Asn  Val  Leu  Leu  Thr amino
    acid  sequence
    ATA  TGT  ATA  AAG  GCA  AAC  GTC  CTG  TTA  ACA DNA
    sequence

    ATA  TGT  ATA  TAG  GCAAACGTCCTGTTAACA  stop codon
    Ile  Cys  Ile  Ter  polypeptide termination
```

Fig. 2 Types of substitutions in a protein-coding region: (a) synonymous, (b) missense, and (c) nonsense. In each case, the top sequence is wild type while the bottom sequence is the mutated one.

parental chromosomes 7, a mutated cystic fibrosis CFTR gene. In Europe, the frequency of people carrying only one CFTR mutation is 1/40. Such heterozygotes show a decrease in the fluidity of mucus excretion that may pass unnoticed. Geneticists have postulated that the selection of one CFTR mutation results from past cholera epidemics. People, ill with cholera, suffer from an almost 15 L of diarrhea per day; they mostly die of dehydration more than from infection sickness. Individuals with only one mutated cystic fibrosis CFTR mutation leak much less water and can overcome the infection with cholera germs. In countries where

cholera is endemic, individuals with only one CFTR mutation seem to be resistant to cholera.

The probability for a child to receive two CFTR mutations is 1/1600. This fraction shows the high frequency of cystic fibrosis in the human population. Resistance to cholera in a human population is proportional to the frequency of occurrence of spontaneous CFTR mutations.

2
DNA-damaging Agents are Mutagens

Naturally, mutations arise at low frequency. To increase the frequency of appearance of mutations, Herman Muller thought of irradiating Drosophila eggs with X rays. He discovered that X rays increased the mutation frequency as a function of the absorbed dose. Moreover, chemicals such as mustard gas (yperite) showed a similar dose response. Agents that produce mutations are denoted *mutagens*.

2.1
A DNA Lesion is Different from a Mutation

The practical knowledge of DNA lesions that we have is that after too much bathing in the sun, our skin becomes red and itching. Solar radiations, named UVA (320–400 nm), produce various *DNA lesions* within our skin cells. Photobiologists have identified more than 20 lesions, depending on their 3D-structure and shapes.

Why is it that with so many diverse lesions produced, we observe that 85% of them are single base pair changes, in just a DNA base? For example, Adenine (A) may change into Cytosine (C), Guanine (G), or Thymine (T). Benzopyrene, a chemical inhaled with tobacco smoke, when metabolized, gives rise to a chemical derivative, benzopyrene dihydrodiol that sticks to DNA. In our body, this compound produces a single base mutation: an A or a T or a C or a G.

What is the relationship between a DNA lesion and a mutation? In the preceding example, a benzopyrene-dihydrodiol-base complex bound to one DNA strand is taken out by excision, which leaves a gap of around 10 bases (in bacteria) or 30 bases (in human cells). DNA resynthesis will fill the gap, sometimes in an error-prone way, giving rise to a mutation. Recall that a DNA lesion is not a mutation even if nowadays one can synthesize a piece of DNA with an altered base *in vitro*.

Biochemical reactions, namely, *trans-lesion DNA-synthesis*, proceed on the DNA to restore a functional sequence at the site where there was a lesion. Most unexcised DNA lesions block DNA replication. SOS-induced-repair polymerases bypass the DNA lesions, allowing the cells to survive. At least 50 repair proteins concur to cell survival. Repair enzymes act all the more efficiently as DNA lesions are not too many.

Interestingly, the lesions that trigger an SOS response in bacteria induce a cascade of repair processes. Structural restoration of a DNA double helix is a primordial step permitting replication, cell division, and cell survival to take place.

3
Characteristic Mutation Types

By the use of various techniques, biologists proved that all living species possess a similar DNA structure. In all genomes, there are four nucleic bases, A, C, G, T, whether in microorganisms or mammals.

1. *Point mutations*: The most frequent are base substitutions. Substitutions

happen between purines, Adenine (A) and Guanine (G), or between pyrimidines, Cytosine (C) and Thymine (T) (Fig. 1b); they are called transitions (A capital letter for each base name is used to stress what base the symbol stands for). An AT base pair may give rise to a GC (AT → GC) and, conversely, a GC base pair to an AT (GC → AT); the two-way reactions are illustrated as AT ↔ GC.

Transversions are substitutions of a purine for a pyrimidine and *vice versa* (Fig. 1c). There are four reciprocal base pair changes: AT ↔ CG; AT ↔ TA; GC ↔ TA; GC ↔ CG. Base substitutions that occur in protein-coding regions affect the structure of the expressed protein with the exception of the third base of a codon when the change does not modify the amino acid (see neutral or synonymous mutation, Fig. 2a).

2. *Deletions and insertions*: Examples of deletions and insertions are seen in Fig. 1(d) and (e). Note that an unequal crossing-over between two chromosomes may result in a deletion of a DNA segment in one chromosome and in an addition in the other.
3. *Frameshifts*: A small deletion or insertion within a gene, whose number of bases is not a multiple of three, causes a shift in the reading frame, denoted frameshift (Fig. 3a). Then, the coding sequence downstream displays a wrong phase. The new phasing changes the nature of the encoded amino acids; or else a new stop codon may be brought upstream into phase, producing a shortened protein. Figure 3(a) shows a protein sequence that may not have normal activity. Figure 3(b) also illustrates the fact that an addition of a base results into a +1 frameshift that may remove a preexisting stop signal giving rise to an elongated protein.
4. *DNA rearrangements*: Deletions of a segment of a gene or of sets of genes may reduce or eliminate protein functions

```
(a)  Lys Ala Leu Val Leu Leu Thr Ile Cys Ile Ter

     AAG GCA CTG GTC CTG TTA ACA ATA TGT ATA TAA TACCATCGCAATAGGG

              Del(G)

     AAG GCA CTG ᵛTCC TGT TAA CAATATGTATATAATACCATCGCAATAGGG

     Lys Ala Leu  Phe Cys Ter

(b)  Lys Ala Asn Val Leu Leu Thr Ile Cys Ile Ter

     AAG GCA AAC GTC CTG TTA ACA ATA TGT ATA TAA TACCATCGCAATAGGG

               Ins(G)

     AAG GCA AAC ᵛGᵛGT CCT GTT AAC AAT ATG TAT ATA ATA CCA TCG CAA AGT GG

     Lys Ala Asn  Gly Pro Val Asn Asn Met Tyr Ile Ile Pro Ser Gln Ter
```

Fig. 3 Examples of frameshifts caused by deletion or insertion. (a) A deletion of a G causing premature termination. (b) An insertion of a G obliterates a stop codon. Termination codons are underlined. In each case, the top sequence is wild type while the bottom sequence is the mutated one.

(Fig. 3). In addition, insertions of large DNA segments can occur, as can duplications, inversions, and other more complicated DNA rearrangements. All these processes are the likely results of recombination between DNA sequences sharing homology. Agents that break chromosomes like X rays or yperite, facilitate DNA rearrangements.

Effect of a Bcl2 rearrangement: When a piece of human chromosome 14, carrying the *Bcl2* gene, and thus expressing a function preventing cell death, is translocated to chromosome 18, then the *Bcl2* gene gets expressed constantly. This leads to the production of a modified immunoglobulin, normally eliminated from the body by cell death. The modified immunoglobulin is kept alive here because of the *Bcl2* translocation. Thus, a pool of unhealthy cells builds gradually; point mutations accumulate in those unhealthy cells and will eventually give rise to cancer cells.

5. *Insertion of transposable genetic elements*: Transposable genetic elements, also called transposons or jumping genes, are present in most organisms. They are highly mobile around chromosomes. If a transposon inserts into a gene encoding a protein, it may cut the gene into two pieces and the protein may not be expressed as a whole. This is the negative property of transposon insertion. On the contrary, if a transposon with active promoters inserts itself into the regulatory region of a gene, it may strongly increase the expression of downstream genes.
6. *Mutation reversions: true- or pseudoreversion of a mutation*: Sometimes after a few generations, the phenotype of a mutant organism seems to have reversed to the wild type original and a question arises whether the mutation reverted to the original parental type? Or else, has the mutation remained in the DNA while another mutation had occurred elsewhere, counterbalancing the first one?
7. *Mutation hotspots*: Mutations may occur at random along the genome, but more often than not, they crop up at some sites designated hotspots. There is such a hotspot in the *E. coli* chromosome: the dinucleotide CpG, in which one finds the cytosine frequently methylated, changes to TpG, introducing an error.

   Another specific mechanism can take place, designated *induced activation of cytosine deaminase* (AID). The deamination of methyl cytosine gives rise to thymine in the absence of replication. This process converts cytosine to uracil and adenine to hypoxanthine. It may be the main cause of immunoglobin hypermutation, which may result from the conversion of cytosine to uracil with the AID enzyme.

   The dinucleotide TpT is a mutation hotspot in prokaryotes (but not in eukaryotes). The particular 3-D shape of a DNA region generates some targets for specific enzymes like methylases. In bacteria, DNA regions containing short *palindromes* like 5′GCCGGC3′, 5′GGCGCC3′, and 5′GGGCCC3′ are prone to mutate more frequently than other regions. In eukaryotic genomes, short tandem repeats, resulting from *slipped-strand mispairing* are often hotspots (Fig. 4a,b).
8. *Mutations at regulatory sequences*: DNA encodes sequences that control the expression of proteins, such as promoters

```
                    Normal pairing
                 during DNA replication

                 5'—AATCCTAGTATATA          3'
                     :::::::::::::
                 3'—TTAGGATCATATATGTGCTTAA—5'
```

```
Slipped-strand mispairing

         AG
        T |T
        C |A
5'—AATC    TATA            3'
    ::::    ::::
3'—TTAGGATCATATATGTGCTTAA— 5'

Replication continues
inserting TA repeat unit

         AG
        T |T
        C |A    >>>>>>>>>>_
5'—AATC    TATATACACGAATT— 3'
    ::::    :::::::::::::
3'—TTAGGATCATATATGTGCTTAA— 5'
```

```
Slipped-strand mispairing

5' —AATCCTAGTATATA        3'
     ::::    ::::**
3' —TTAG    ATATGTGCTTAA—5'
        G  T|
        A  A|
         TC

Replication continues after
excision of unrepaired TA
repeat unit

                  >>>>>>>>
5' —AATCCTAGTATACACGAATT—3'
     ::::    ::::::::::::
3' —TTAG    ATATGTGCTTAA—5'
        G  T|
        A  A|
         TC
```

(a)

Fig. 4 Generation of duplications or deletions by slipped-strand mispairing between contiguous repeats (underlined). Small arrows indicate the direction of DNA synthesis. Dots indicate base pairing. (a) *A two-base slippage in a TA repeat during DNA replication*. Slippage in the 3′ → 5′ direction results in the insertion of one TA unit (left panel). Slippage in the other direction results in the deletion of one repeat unit (right panel). The deletion shown in the right panel results from excision of the unpaired repeat unit (asterisks) at the 3′ end of the growing strands, presumably by the 3′ → 5′ exonuclease activity of DNA polymerase. (b) *A two-base slippage in a TA repeat in nonreplicating DNA*. Mismatched regions form single-stranded loops, which may be targets of excision or mismatch repair. The outcome (a deletion or an insertion) will depend on which strand is excised or repaired and which strand is used as template in the DNA repair process. (Modified scheme from Levinson and Gutman).

and operators. Mutations in these sequences may alter the cell concentration of proteins encoded downstream of the regulatory sequence. Overexpression or underexpression of proteins may sharply modify the function of a gene in an organism. Since many proteins work in a concerted action, a sharp increase or decrease in the cell concentration of a protein may disturb the very existence of an organism.

4 Silent Base Substitutions

A mutation that does not cause any amino acid change in the expressed protein is

Fig. 4 (*Continued*)

```
     Normal pairing
 of intact chromosomal DNA

5'—AATCCTAGTATATACACGAATT—3'
    :::::::::::::::::::::
3'—TTAGGATCATATATGTGCTTAA—5'

Slipped-strand mispairing

         AG
        T  T
        C  A
5'—AATC    TATACACGAATT — 3'
    ::::    ::::::::::::
3'—TTAGGATCATAT    TTAA — 5'
                A  C
                T  G
                 GT

Excision or mismatch repair inserts
          TA repeat unit

        AG
       T |T
       C |A    >>>>>>>>>>
5'—AATC    TATATACACGAATT—3'
   ::::    ::::::::::::::
3'—TTAGGATCATATATGTGCTTAA—5'
```

(b)

dubbed *synonymous* and may become *silent*. In Fig. 2(a), the protein codes for the valine amino acid but the DNA sequence reveals a base change.

A *nonsynonymous* mutation alters an amino acid, resulting in a *missense* or a *nonsense* mutation. A missense mutation specifies an amino acid different from the one originally encoded, for instance, valine (Val) becomes phenylalanine (Phe) (Fig. 2b). A nonsense mutation changes a codon into one of the three termination codons TAG, TAA, or TGA. Then, a truncated protein is produced since lysine (Lys) becomes Stop (Fig. 2c).

4.1 Synonymous Mutations are Neutral

In a DNA sequence that codes for a protein, each of the sense codons can mutate to nine other codons by means of a single base substitution. For example, the triplet CCU that codes for proline (Pro) can give rise to three *synonymous* substitutions CCC, CCA, or CCG, which also code for proline (Pro). In a different order, the triplet can give rise to six *nonsynonymous* substitutions, which code for different amino acids, UCU for serine (Ser), ACU for threonine (Thr), GCU for alanine (Ala), CUU for leucine (Leu), CAU for histidine (His), or CGU for arginine (Arg).

Since the genetic code defines 61 sense codons, it is reckoned that the possible codon substitutions are $61 \times 9 = 549$. If we assume that base substitutions may occur at random and that all codons are equally frequent in regions coding for proteins, we can compute the expected proportion of the different types of base substitutions. Because of the structure of the genetic code, *synonymous* substitutions occur mainly at the *third* position of codons. Indeed, 69% of all the possible *synonymous* changes in a codon are at the third-base position.

In contrast, substitutions at the *second* position of codons are *nonsynonymous*, and so are the vast majority of base changes at the *first* position (96%).

Synonymous amino acid replacements do not alter either protein structure or function. From an evolutionary viewpoint, these mutations are *neutral* since they do not seem to affect the physiology of the mutated organisms. Such synonymous mutations accumulate with time since there is no counterselection for their emergence. Synonymous mutations indeed generate *change at the DNA level*, as indicated by the use of restriction enzymes and sequencing techniques. Over time, accumulation of synonymous mutations increases DNA divergence between individuals within a species. Then, a new species can emerge by segregation. Chimpanzee and man share 98.5% of the structure of their proteins. DNA sequences in the two species have diverged upon evolution from a common ancestor. The consequence of the divergence is a gradual loss of DNA homology that hinders recombination between the two species and thus the formation of hybrids.

A species is defined by the property of the individuals within a group to mate and give rise to a progeny since the parental DNAs can recombine. They can do so if, and only if, the DNAs are homologous. If genomic DNA accumulates third-base changes with time, the DNA of the mating pair turns out to be too divergent with respect to their respective ancestors. The resulting diverged individuals may attempt to mate, but they cannot give rise to progeny.

Since the act of mating is followed by DNA recombination, the latter is successful only if it arises between two homologous DNAs. Mating is doomed to fail if it occurs between two heterologous DNAs. The first step in recombination is the pairing of two pairs of DNA strands. When two heterologous DNA chains attempt to pair, base-to-base *mismatches* arise between bases that should otherwise complement. Mismatches are removed by the induction of a very efficient repair process called *mismatch repair*. Radman and associates have demonstrated that recombination between two heterologous DNAs aborts through the action of *mismatch repair*.

5
Four Main Mechanisms that Generate Spontaneous Mutations

Spontaneous mutations may result from four main mechanisms: (1) DNA polymerization errors, (2) spontaneous oxidative DNA damage, (3) loss of a purine base, or less frequently, loss of a pyrimidine base, and (4) deamination of cytosine or of adenine.

1. *DNA polymerization errors*: Point mutations, usually transitions, arise from two successive faulty processes: (1) an error in chromosome replication resulting in a mismatch; and (2) followed with inefficient repair.

For example, DNA polymerase III may incorporate *in vitro* an incorrect base at a frequency of 10^{-6} to 10^{-7} per base pair per cell per generation. In *E. coli* bacteria, mismatch repair will remove the incorrect base, reducing the overall rate of replication and postreplication errors down from 10^{-9} to 10^{-10} per base pair per cell per generation.

Replication slippage, also denoted *slipped-strand mispairing*, is the major mechanism that accounts for frameshift mutations occurring at contiguous short repeats (Fig. 4).

2. *Incorporation of an oxidized base*: Exposed to oxygen, guanine is subject to oxidation that produces 8-oxo-G. This abnormal base has a tendency to slip into DNA, generating transversions. Fortunately, an elaborate repair system comes into play that either removes 8-oxo-G or decreases the consequences of its incorporation. Such a repair system is ubiquitous, protecting DNA from oxidative lesions.

   Note that the oxygen we breathe is far from harmless. Oxygen is a DNA-damaging agent and may have a role in the aging process. Ames and associates have correlated the number of oxidative hits to DNA with the amount of DNA breakdown products released in rat urine and human urine. They have shown that the number of oxidative hits to DNA per cell per day is around 100 000 in rats and about 10 000 in humans. The 10-fold difference may result in part from the elevated metabolic rate of rodents, at least five times higher than that of mammals.

   The mutagenic action of 8-oxo-G implies that there is no clear-cut difference between "natural" and "artificial" DNA-damaging agents. The distinction between spontaneous and induced mutations is also vague since the air we breathe does also insert oxides into DNA.

3. *Loss of a nucleic base*: It is the most frequently arising spontaneous lesion. If the chemical bond between sugar and base breaks, releasing a free base and leaving a gap in the sequence of bases, such an event generates an abasic site (apurinic or apyrimidinic). Loss of a purine base is more frequent than loss of a pyrimidine. A pyrimidine base strengthens DNA, which is in general sturdier than RNA.

   Replication over an abasic site may produce a point mutation since the polymerase has no way of determining the identity of the missing base it has to copy. Adenine (A) seems frequently placed opposite the missing base. When T is the missing base, the placement of an A restores the normal coding.

4. *Deamination*: This process may convert cytosine to uracil and adenine to hypoxanthine. The former conversion seems to be involved as the main cause of immunoglobin hypermutation. This process would account for the creation of half a million antibodies, a process permitting our immunity repertoire to be constituted.

6 Induced Mutations

They may result from three different mechanisms: (1) the replacement of a base in the DNA by a base analog that confuses the polymerase; (2) an imperfect fit between a chemically altered base with another base during replication; and (3) the replacement of a base with a gap that lacks coding information.

1. *Base replacement by a base analog*: 2-Aminopurine, a base analog that mimics a normal base, causes frequent mispairing when incorporated into DNA.
2. *Mispairing by alkylation of bases*: Alkylation with ethylmethanesulfonate or with MNNG (*N*-methyl-*N*′-nitro-*N*-nitrosoguanidine) causes mispairing. Although these two chemicals modify different bases at various positions, alkylation at the O^6 position of guanine and at the O^4 position of thymine causes specific mispairing with T and with G, leading to GC ↔ AT and AT ↔ GC transitions. *In vivo*, O^6-alkylguanine and O^4-alkylthymine are two chemicals that cause the most mutations through alkylation.
3. *Noncoding lesions*: Ultraviolet light C (256 nm) generates pyrimidine dimers and 6–4 pyrimidine-pyrimidone compounds in DNA. Such lesions sharply bend DNA, distorting the precise pairing of bases upon replication.

7
Mutagenesis in Bacteria and Carcinogenesis in Mammals

Bruce Ames asked an interesting question: do mutagens cause cancer? If they do, it should be possible practically to replace the usual lengthy cancer tests on mice, lasting months and years, with a much shorter test involving bacteria instead of animals. Ames and collaborators devised a mutagenicity test using a *Salmonella* tester strain (*his*$^-$) that requires histidine for growth. The *his*$^-$ tester bacteria treated with various carcinogens produced *his*$^+$ mutants that were able to grow on a medium devoid of histidine.

To make the test even more efficient, the bacterial tester strains were engineered in order to: (1) prevent the removal of the DNA lesions produced by the chemical tested; (2) increase the mutagenic rate by introducing a mutator plasmid in the tester strains; and (3) render the tester strains permeable to chemical compounds in changing its cell wall structure.

In the experimental procedure, one mixes the chemical with rat liver enzymes to reproduce the metabolization of the compound by a human liver. In the test, a transient physiological chimera combines rat enzymes with permeable bacteria to reproduce the human situation.

The results of the tests have shown that more than 65% of the 3000 chemical compounds tested, which are rodent carcinogens also cause mutations in bacteria. The converse is true: most mutagens are active carcinogens. Using a bacterial test to detect potential carcinogens has saved time, money, and animal suffering.

Note that even though bacteria are evolutionarily distant from rodents and humans, they provide an excellent tool for establishing that mutagenicity correlates with carcinogenicity. This is because DNA damage is the origin of both processes.

Because the *Salmonella* mutagenicity test is rapid, accurate, and relatively simple, many government agencies and biological laboratories have tested new chemicals for their mutagenicity and their potential carcinogenicity all over the world. In the European Union, the bacterial mutagenicity test has become mandatory. The overall result is that even though bacteria are evolutionarily more than a billion years distant from rodents and humans, they provide an excellent tool for measuring the mutagenic potential of chemical compounds. Government agencies and biological laboratories are testing newly found chemicals for their mutagenicity and potential carcinogenicity

all over the world. In the European Union, the *Salmonella* mutagenicity test has become mandatory.

8
Carcinogens Leave No Definite Mutational Signatures on DNA

In human cells, a gene called *p53* expresses a protein that inhibits the occurrence of cancers; wild type protein P53 looks as an *anti-oncogene*. Measurements of the occurrence of p53 gene mutations demonstrated that in more than half of all cancers taken together, 90% of the p53 mutations are point mutations, two-third of them being transitions GC → AT.

In bronchial cancers, one finds that 45% of p53 mutations are transversions GC → TA. Benzopyrene, a mutagen produced by cigarettes or car exhausts, gives rise to a low level of GC → TA transversions.

Is the presence in cancer cells of a GC → TA transversion in the *p53* gene a case for making benzopyrene responsible for the production of the mutation? So many carcinogens and environmental pollutants are causing GC → TA transversions. How is it possible to incriminate benzopyrene itself in causing a tumor? Moreover, a point mutation is so common and so minute a DNA change that it is hard to take it as an indelible mark left over by a carcinogen.

In order to provide a legal proof for carcinogenicity of a mutated DNA sequence, a sequence must be long and varied enough to offer a valuable correlation. Moreover, correlation may indicate a causality link if the correlation index is extremely high.

The probability of finding an identical copy of a given DNA fingerprint is below 10^{-10}. Since the human population has not reached 10^{10} yet, fingerprinting provides a valid legal proof of identification. There is a large difference in the probability of occurrence of a given microsatellite with that of a transversion.

Few chemicals are legally recognized carcinogens. For instance, asbestos, a mineral, causes a unique and almost exclusive type of tumor called *mesothelioma*. Another compound, radioactive iodine, ingested by Ukrainian children in the aftermath of the Chernobyl nuclear accident is instrumental in breaking a specific chromosome in thyroid cancer cells. That type of causal relationship between mutations induced by radioactive iodine and a specific tumor is rather infrequent in practice. The probability for carcinogens to leave a ''mutational signature'' on DNA is so low that this concept has never been accepted as evidence in court.

9
Humans Use Mutagenesis to Produce their Immunity Repertoire

Mutations in DNA sequences in the immunoglobulin locus are involved in *class switch recombination* (CSR). These mutations are base substitutions, deletions, or duplications. Such mutations are primarily dependent upon the *induced activation of cytosine deaminase* or AID. One finds them during *V gene somatic hypermutation*. A correlation between a pattern of mutations and a mutational complex might generate them. This is due to the observation of the V gene mutation pattern of XP-V patients afflicted with the cancer-prone *Xeroderma pigmentosum variant syndrome*. The patients are deficient in DNA polymerase η (polη), suggesting that this enzyme is responsible for a large part of the mutations occurring at A–T bases. After analysis of switched memory B cells from two XP-V

patients, it has been deduced that polymerase polη is also an A–T mutator during CSR, in both the switch region of tandem repeats as well as upstream of it. This suggests that the same error-prone translesional polymerase is involved, together with AID, in *CSR*.

Acknowledgments

It is a pleasure to acknowledge Li and Graur for their figure models. I also thank Pr. Joann Sweasy (Yale University) and Dr. Evelyne Sage (Institut Curie) for their insightful comments, and also thank Mrs. Melanie Pierre for her kind and indefatigable help.

See also Aging and Sex, DNA Repair in; DNA Damage Induced by Ultraviolet Radiation; DNA Repair in Yeast; Ionizing Radiation Damage to DNA; Nucleic Acids (DNA) Damage and Repair; RecA Superfamily Proteins; Recombination and Genome Rearrangements.

Bibliography

Books and Reviews

Ames, B.N. (1979) Identifying environmental chemicals causing mutations and cancer, *Science* **204**, 587–593.

Devoret, R. (1979) Bacterial tests for potential carcinogens, *Sci. Am.* **241**, 40–50.

Echols, H., Goodman, M.F. (1991) Fidelity mechanisms in replication, *Annu. Rev. Biochem.* **60**, 477–511.

Friedberg, E.C., Walker, G.C., Siede, W. (1995) *DNA Repair and Mutagenesis*, American Society for Microbiology, Washington, DC.

Griffiths, A.J.F., Miller, J.H., Suzuki, D.T., Lewontin, R.C., Gelbart, W.M. (1993) *An Introduction to Genetic Analysis*, W. H. Freeman and Company, New York.

Li, W.H., Graur, D. (1991) *Fundamentals of Molecular Evolution*, Sinauer Associates, Sunderland.

Radman, M. (1975) SOS Repair Hypothesis, Phenomenology of an Inducible DNA Repair Which is Accompanied by Mutagenesis, in: Hanawalt, P., Setlow, R. (Eds.) *Molecular Mechanisms for Repair of DNA*, Plenum Publishers, New York.

Radman, M., Wagner, R. (1988) The high fidelity of DNA duplication, *Sci. Am.* **259**, 24–30.

Walker, G.C. (1984) Mutagenesis and inducible responses to deoxyribonucleic acid damage in Escherichia coli, *Microbiol. Rev.* **48**, 60–93.

Primary Literature

Bailone, A., Sommer, S., Knezevic, J., Dutreix, M., Devoret, R. (1991) A RecA protein mutant deficient in its interaction with the UmuDC complex. *Biochimie* **73**, 479–484.

Becherel, O.J., Fuchs, R.P.P., Wagner, J. (2002) Pivotal role of the β-clamp in translesion DNA synthesis and mutagenesis in E. coli cells, *DNA Repair* **1**, 703–708.

Blanco, M., Herrera, G., Collado, P., Rebollo, J.E., Botella, L.M. (1982) Influence of RecA protein on induced mutagenesis, *Biochimie* **64**(8–9), 633–636.

Boudsocq, F., Campbell, M., Devoret, R., Bailone, A. (1997) Quantitation of the inhibition of Hfr × F^- recombination by the mutagenesis complex UmuD′C, *J. Mol. Biol.* **270**, 201–211.

Bruck, I., Woodgate, R., McEntee, K., Goodman, M.F. (1996) Purification of a soluble UmuD′C complex from Escherichia coli. cooperative binding of UmuD′C to single-stranded DNA, *J. Biol. Chem.* **271**, 10767–10774.

Burnouf, D.Y., Olieric, V., Wagner, J., Fujii, S., Reinbolt, J., Fuchs, R.P., Dumas, P. (2004) Structural and biochemical analysis of sliding clamp/ligand interactions suggest a competition between replicative and translesion DNA polymerases, *J. Mol. Biol.* **335**, 1187–1197.

Cordonnier, A.M., Fuchs, R.P. (1999) Replication of damaged DNA, molecular defect in xeroderma pigmentosum variant cells, *Mutat. Res.* **435**, 111–119.

Dalrymple, B.P., Kongsuwan, K., Wijffels, G., Dixon, N.E., Jennings, P.A. (2001) A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11627–11632.

Dickerson, S.K., Market, E., Besmer, E., Papavasiliou, F.N. (2003) AID mediates hypermutation by deaminating single stranded DNA, *J. Exp. Med.* **197**, 1291–1296.

Dorner, T., Lipsky, P.E. (2001) Smaller role for pol eta? *Nat. Immunol.* **2**, 982–984.

Dutreix, M., Moreau, P.L., Bailone, A., Galibert, F., Battista, J.R., Walker, G.C., Devoret, R. (1989) New recA mutations that dissociate the various RecA protein activities in Escherichia coli provide evidence for an additional role for RecA protein in UV mutagenesis, *J. Bacteriol.* **171**, 2415–2423.

Faili, A., Aoufouchi, S., Gueranger, Q., Zober, C., Leon, A., Bertocci, B., Weill, J.-C., Reynaud, C.-A. (2002) AID dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line, *Nat. Immunol.* **3**, 815–821.

Faili, A., Aoufouchi, S., Weller, S., Vuillier, F., Stary, A., Sarasin, A., Reynaud, C.-A., Weill, J.-C. (2004) DNA Polymerase eta is involved in hypermutation occurring during immunoglobulin class switch recombination, *J. Exp. Med.* **199**, 265–270.

Friedberg, E.C., Wagner, R., Radman, M. (2004) Specialized DNA polymerases, cellular survival, and the genesis of mutations, *Science* **296**(5573), 1627–1630.

Goodman, M.F. (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes, *Annu. Rev. Biochem.* **71**, 17–50.

Goodman, M.F., Tippin, B. (2000) Sloppier copier DNA polymerases involved in genome repair, *Curr. Opin. Genet. Dev.* **10**, 162–168.

Kannouche, P., Fernandez de Henestrosa, A.R., Coull, B., Vidal, A.E., Gray, C., Zicha, D., Woodgate, R., Lehmann, A.R. (2002) Localization of DNA polymerases eta and iota to the replication machinery is tightly co-ordinated in human cells, *EMBO J.* **21**, 6246–6256.

Lang, T., Maitra, M., Starcevic, D., Li, S.X., Sweasy, J.B. (2004) A DNA polymerase beta mutant from colon cancer cells induces mutations, *Proc. Natl. Acad. Sci. U.S.A.* **101**(16), 6074–6079.

Lehmann, A.R., Kirk-Bell, S., Arlett, C.F., Paterson, M.C., Lohman, P.H., de Weerd-Kastelein, E.A., Bootsma, D. (1975) Xeroderma pigmentosum cells with normal levels of excision repair have a defect in DNA synthesis after UV-irradiation, *Proc. Natl. Acad. Sci. U.S.A.* **172**, 219–223.

Livneh, Z. (2001) DNA damage control by novel DNA polymerases, translesion replication and mutagenesis, *J. Biol. Chem.* **276**, 25639–25642.

Maor-Shoshani, A., Livneh, Z. (2002) Analysis of the stimulation of DNA polymerase V of Escherichia coli by processivity proteins, *Biochemistry* **41**, 14438–14446.

Masutani, C., Kusumoto, R., Iwai, S., Hanaoka, F. (2000) Mechanisms of accurate translesion synthesis by human DNA polymerase eta, *EMBO J.* **19**, 3100–3109.

Matic, I., Babic, A., Radman, M. (2003) 2-aminopurine allows interspecies recombination by a reversible inactivation of the Escherichia coli mismatch repair system, *J. Bacteriol.* **185**(4), 1459–1461.

Matic, I., Taddei, F., Radman, M. (2000) No genetic barriers between Salmonella enterica serovar typhimurium and Escherichia coli in SOS-induced mismatch repair-deficient cells, *J. Bacteriol.* **182**(20), 5922–5924.

Nagaoka, H., Muramatsu, M., Yamamura, N., Kinoshita, K., Honjo, T. (2002) Activation-induced deaminase (AID)-directed hypermutation in the immunoglobulin Smu region, implication of AID involvement in a common step of class switch recombination and somatic hypermutation, *J. Exp. Med.* **195**, 529–534.

Napolitano, R., Janel-Bintz, R., Wagner, J., Fuchs, R.P. (2000) All three SOS-inducible DNA polymerases (Pol II, Pol IV and Pol V) are involved in induced mutagenesis, *EMBO J.* **19**, 6259–6265.

Ohmori, H., Friedberg, E.C., Fuchs, R.P., Goodman, M.F., Hanaoka, F., Hinkle, D. et al (2001) The Y-family of DNA polymerases, *Mol. Cell* **8**, 7–8.

Pages, V., Fuchs, R.P. (2002) How DNA lesions are turned into mutations within cells? *Oncogene* **21**, 8957–8966.

Pham, P., Bransteitter, J., Petruska, J., Goodman, M.F. (2003) Processive AID-catalysed cytosine deamination on single stranded DNA simulates somatic hypermutation, *Nature* **424**, 103–107.

Pham, P., Bertram, J.G., O'Donnell, M., Woodgate, R., Goodman, M.F. (2001) A model for

SOS-lesion targeted mutations in Escherichia coli, *Nature* **409**, 366–370.

Reina-San-Martin, B., Difilippantonio, S., Hanitsch, L., Masilamani, R.F., Nussenzweig, A., Nussenzweig, M.C. (2003) H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation, *J. Exp. Med.* **197**, 1767–1778.

Reuven, N.B., Arad, G., Maor-Shoshani, A., Livneh, Z. (1999) The mutagenesis protein UmuC is a DNA polymerase activated by UmuD′, RecA, SSB and is specialized for translesion replication, *J. Biol. Chem.* **274**, 31763–31766.

Reuven, N.B., Arad, G., Stasiak, A.Z., Stasiak, A., Livneh, Z. (2001) Lesion bypass by the Escherichia coli DNA polymerase V requires assembly of a RecA nucleoprotein filament, *J. Biol. Chem.* **276**, 5511–5517.

Reynaud, C.-A., Aoufouchi, S., Faili, A., Weill, J.-C. (2003) What role for AID, mutator, or assembler of the immunoglobulin mutasome? *Nat. Immunol.* **4**, 631–638.

Roca, A.I., Cox, M.M. (1997) RecA protein, structure, function, and role in recombinational DNA repair, *Prog. Nucleic Acid Res. Mol. Biol.* **56**, 129–223.

Sarasin, A. (2003) An overview of the mechanisms of mutagenesis and carcinogenesis, *Mutat. Res.* **544**(2–3), 99–106.

Sassanfar, M., Roberts, J.W. (1990) Nature of the SOS-inducing signal in Escherichia coli, the involvement of DNA replication, *J. Mol. Biol.* **212**, 79–96.

Sommer, S., Bailone, A., Devoret, R. (1993) The appearance of the UmuD′C protein complex in Escherichia coli switches repair from homologous recombination to SOS mutagenesis, *Mol. Microbiol.* **10**, 963–971.

Sommer, S., Boudsocq, F., Devoret, R., Bailone, A. (1998) Specific RecA amino acid changes affect RecA-UmuD′C interaction, *Mol. Microbiol.* **28**, 281–291.

Starcevic, D., Dalal, S., Sweasy, J.B. (2004) Is there a link between DNA polymerase beta and cancer? *Cell Cycle* **3**, 998–1001.

Sutton, M.D., Walker, G.C. (2001) Managing DNA polymerases, coordinating DNA replication, DNA repair, and DNA recombination, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8342–8349.

Sutton, M.D., Narumi, I., Walker, G.C. (2002) Posttranslational modification of the umuD-encoded subunit of Escherichia coli DNA polymerase V regulates its interactions with the beta processivity clamp, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5307–5312.

Tang, M., Shen, X., Frank, E.G., O'Donnell, M., Woodgate, R., Goodman, M.F. (1999) $UmuD_2C$ is an error-prone DNA polymerase, Escherichia coli pol V, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8919–8924.

Tang, M., Pham, P., Shen, X., Taylor, J.S., O'Donnell, M., Woodgate, R., Goodman, M.F. (2000) Roles of E. coli DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis, *Nature* **404**, 1014–1018.

Wagner, J., Etienne, H., Janel-Bintz, R., Fuchs, R.P.P. (2002) Genetics of mutagenesis in E. coli, various combinations of translesion polymerases (Pol II, IV and V) deal with lesion/sequence context diversity, *DNA Repair* **1**, 159–167.

Wagner, J., Fujii, S., Gruz, P., Nohmi, T., Fuchs, R.P. (2000) The beta clamp targets DNA polymerase IV to DNA and strongly increases its processivity, *EMBO Rep.* **1**, 484–488.

Wagner, J., Gruz, P., Kim, S.-R., Yamada, M., Matsui, K., Fuchs, R.P.P., Nohmi, T. (1999) The dinB gene encodes a novel E. coli DNA polymerase, DNA PolIV, involved in mutagenesis, *Mol. Cell* **4**, 281–286.

Yavuz, S., Yavuz, A.S., Kraemer, K.H., Lipsky, P.E. (2002) The role of polymerase eta in somatic hypermutation determined by analysis of mutations in a patient with xeroderma pigmentosum variant, *J. Immunol.* **169**, 3825–3830.

Zeng, X., Winter, D.B., Kasmer, C., Kraemer, K.H., Lehmann, A.R., Gearhart, P.J. (2001) DNA polymerase eta is an A-T mutator in somatic hypermutation of immunoglobulin variable genes, *Nat. Immunol.* **2**, 537–541.

Repeat Explansion Disorders: *see* Triplet Repeat Diseases

Repetitive Elements, Human: *see* Anthology of Human Repetitive DNA

Repetitive Sequence and Noncoding DNA Evolution in Eukaryotes

Josep M. Comeron
The University of Iowa, Iowa City, IA

1 The C-value Paradox 128

2 Noncoding Genome Size: Passive versus Selective Trait 129

3 Deletion Bias (DB) Based on Small Indels 130
3.1 DB in Drosophila 130
3.2 What is the True DB in Drosophila? 130
3.3 DB in Other Eukaryotes 131
3.4 Overall Rate of DNA Loss due to Small Indels and Genome Size 131

4 Transposable Elements (TE) and DNA Gain 132

5 Variation across and among Genomes: The Influence of Recombination 132

6 Noncoding DNA is not Free of Genetic Information or Evolutionary Effect 133

7 Summary 134

Acknowledgment 135

Bibliography 135
Book and Reviews 135
Primary Literature 135

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

Keywords

Deletion Bias
The study of small insertions and deletions (indels) in most eukaryotes shows an excess of deletions over insertions, forecasting DNA loss.

Effective Population Size (*Ne*)
A theoretical number of individuals that would explain the amount of genetic variation observed within a species in the absence of selection. *Ne* is somewhat related to, but always smaller than, the actual number of individuals. *Ne* is positively associated with the effectiveness of selection and varies among species as well as across genomes.

The amount of haploid nuclear DNA (or C-value) among eukaryotes varies more than 4 orders of magnitude, with brewer's yeast *Saccharomyces cerevisiae* (1.2×10^7 bp) or the parasitic microsporidium *Encephalitozoon intestinalis* (3×10^6 bp) as prototypical examples of the smallest genomes and several amoebae mentioned as the eukaryotes with largest genomes with $>6 \times 10^{11}$ bp (e.g. *Amoeba dubia* and *Chaos chaos*). Differences in the amount of noncoding DNA are also observed within a given taxonomic group, with algae varying more than 5000-fold, more than 1000-fold among flowering plants or invertebrates, more than 300-fold among vertebrates, and 100-fold among amphibians and insects. An apparent ''paradox'' arises when the size of the nuclear genome is compared to organismal complexity (morphological or developmental) if we assume that the amount of genetic material (genome size) should somehow correspond to the amount of genetic information. The so-called C-value paradox or enigma was coined to describe this lack of correspondence. Later findings revealed that coding sequences account for only a small proportion of the genomic DNA in most eukaryotes. This observation does not solve the C-value paradox, but shifts the debate from the number of genes responsible for a given degree of complexity (the G-value paradox) to the amount of noncoding DNA and the causes for its tremendous variation. In fact, recent studies based on fully sequenced genomes and expression data in model eukaryotes, including humans, have shown that complex gene regulation and the number of different transcripts – with a high fraction of genes having multiple transcripts – and not the number of genes, would better represent biological ''complexity,'' hence explaining the G-value paradox.

1
The C-value Paradox

Many cause/consequence relationships have been proposed on the basis of phenotypic correlations. The variation in noncoding genome size among eukaryotes has been associated with biological factors, both cellular and organismal: cell and nucleus size, cell division rate, metabolic rate, developmental rate, developmental complexity, and overall gene expression.

Indeed, eukaryotes with large genomes tend to have larger nucleus and cells, and species with high developmental rate to developmental complexity ratio tend to have reduced genomes; other associations are only observed in particular lineages.

However, an aspect not always considered is that many biological factors are associated with species or population parameters such as generation time, population size, and effective population size (*Ne*). This indirect association between cellular or developmental factors and *Ne* is pertinent in the analysis of any genomic characteristic (or any biological feature for that matter) because *Ne* is a main parameter influencing natural selection. Population genetic theory predicts and evolutionary analyses confirm that *Ne* is positively associated with the effectiveness of selection. (Increased effectiveness of selection will favor the fixation of advantageous mutations and the removal of deleterious mutations.) As a consequence, it is not direct to distinguish between cause, effect, and indirect or coincidental association; many proposed causal associations illustrate this difficulty. For instance, a causal association between fast developmental rate (and short generation time) and small genomes can be proposed. However, many species with short generation time also have large population sizes and *Ne*, and therefore the observed small genome is likely the result of more efficient selective constraints favoring small genomes while the negative relationship between developmental rate and genome size would be secondary.

Moreover, several cases of closely related species – with equivalent cellular and developmental parameters – exemplify that the size of noncoding DNA can change very rapidly in evolutionary terms independent of biological factors. For instance, closely related Drosophila species show severalfold differences in noncoding DNA and there is significant variation even at intraspecific level. Another piece of information comes from genomic analyses that have shown that several features of noncoding DNA such as the amount of intergenic DNA (or gene density), and intron presence and size also show a heterogeneous distribution across genomes. Thus, an integral explanation to the presence and size of noncoding DNA should be able to explain differences between distant groups or taxa, differences between closely related species with equivalent cellular/developmental features, as well as differences across a given genome.

2
Noncoding Genome Size: Passive versus Selective Trait

One possible explanation to the size of noncoding DNA and genome size in eukaryotes is that the amount of noncoding DNA is the passive result of mutational tendencies only; the manifestation of a mutation-drift equilibrium between insertions and deletions. Under this mutational-only scenario, variation in genome size is explained by differences in the patterns of insertion and deletion of ''selfish'' DNA or due to variation in the intrinsic rates of DNA loss through small insertions and deletions (indels). Special attention has been paid recently to the latter process, with several eukaryotes showing an excess of small deletions over insertions (the so-called deletion bias; DB). In particular, it has been proposed that differences in the rate of DNA loss caused by small indels are a major factor influencing genome size.

Under a selective scenario, genome size represents the equilibrium of mutational

tendencies to either increase or decrease the amount of DNA counterbalanced by selective tendencies: a mutational selection–drift equilibrium. Mutational tendencies can certainly vary among species, with differences in the amount of repetitive DNA and its tendency to expand in the genome (mostly associated with transposable elements; TE), as well as variation in DB. It follows, however, that a mutational-only hypothesis – which assumes that the presence of noncoding DNA has no fitness costs – would predict either the collapse of noncoding DNA or the unlimited expansion of noncoding DNA. A more likely proposal assumes that both mutational tendencies (DNA expansion by TE amplification and DNA loss by DB) are balanced as a result of selective constraints.

The selective causes acting on genome size can be varied and be particular to groups or taxa according to specific cellular or developmental requirements (e.g. fast cell division), targeting replication costs, or biological effects of bulk DNA. As indicated earlier, the effectiveness of selection acting on noncoding DNA would also depend on parameters such as *Ne* and recombination (see the following) and so the outcome of the equilibrium would be species specific. Moreover, both *Ne* and recombination vary across genomes. Thus, a mutation selection–drift equilibrium is expected to be highly variable among taxa, between closely related species, and also vary across a given genome.

3
Deletion Bias (DB) Based on Small Indels

3.1
DB in Drosophila

The struggle to determine the mutational DB on the basis of small indels is best exemplified in Drosophila, in which three different approaches give three different results. The first significant attempt to investigate DB in Drosophila, studied nonallelic copies of non long–terminal repeat retrotransposable elements, so-called "dead-on-arrival" (DOA) elements. The study of *Helena* DOA elements gives a DB of 8.7, suggesting a very high rate of DNA loss. The second approach investigated defective TEs, generating a DB of 3.6. The third approach is the study of indels at polymorphic level in DNA sequences likely free of selective constraints (intergenic regions and introns), with a DB that ranges between 1.4 and 1.9, evidencing a significant but moderate rate of DNA loss.

3.2
What is the True DB in Drosophila?

As indicated by B. Charlesworth, the study of nonallelic DOA elements (that are either fixed or segregate at high frequencies within a species) could bias the estimates of DB if selection favors deletions in recently inserted elements, hence causing an overestimation of DB. The rational for arguing selection against inserted repetitive elements (favoring deletions and shorter versions of the elements) is based on the notion that shorter elements would reduce the probability of ectopic recombination and its deleterious effects. On the other hand, population genetic theory predicts that polymorphic studies will estimate mutational parameters that can only be closer to the true values compared to analyses that use mutations that are either fixed or segregate at high frequencies within a species. If indels are neutral, all approaches should give equivalent results, but if indels are under weak/moderate

selection, the study of indels at polymorphic level is least influenced by selection. Strong selective constraints on the sequences used to investigate polymorphic indels, however, could also reduce the number of deletions observed at polymorphic level, hence causing underestimates in DB.

Several observations suggest that the estimates of DB based on different polymorphic studies is close to the true DB. First, an *ad hoc* theoretical model presuming that long deletions in introns are under strong selection would forecast a lower DB, but even in this case it cannot explain the observed data. Second, the study of polymorphic indels in intergenic regions also produces a very low DB (1.7) and, therefore, the argument that the observed DB is biased because of selection is not likely to be correct. And third, a recent study on the frequency of several families of TEs in populations has generated a key result to understand the observed discrepancy in DB based on long repetitive elements. This study shows that there is a negative correlation between the length of TEs within a family and the frequency in the population, in agreement with the ectopic recombination model that claims that deletions would be selectively favored in long repetitive elements. The consequence of this observation is that estimates of DB based on long repetitive elements will be biased, showing an excess of deletions than those expected by mutational tendencies only.

3.3
DB in Other Eukaryotes

Measures of DB have been obtained in other eukaryotes also on the basis of different approaches. One of the first studies to detect an excess of small deletions over insertions was applied to mammals by analyzing processed pseudogenes. Large-scale analyses of pseudogenes in humans and murids show an average DB of 2.74. The study of pseudogenes in the insect *Podisma pedestris* and in *Caenorhabditis elegans* shows DB of 2.7 and 3.8, respectively. The study of DOA *Lau1* elements in the Hawaiian crickets of the genus *Laupala* and *Maui* elements in two pufferfish species suggests a DB of 2.7 and 1.3–2.0, respectively.

3.4
Overall Rate of DNA Loss due to Small Indels and Genome Size

The rate of DNA loss depends on DB as well as the size of indels, and in most species deletions are longer than insertions. In order of increasing genome size, the overall rate of DNA loss observed in the different eukaryotes is the following: 1.8 in *Drosophila melanogaster* (based on polymorphic indels in intergenic regions and introns), 9 to 15 in Pufferfish, 3.8 in Laupala, 2.6 in rat/mouse, 1.7 in humans, and 3.6 in Podisma. Conversely, *C. elegans* shows a net gain of DNA based on small indels. As T.R. Gregory concluded recently, the use of updated estimates of DNA loss for mammals and Drosophila raise considerable doubts about the strength of any correlation between DNA loss rate by small indel bias and genome size. Therefore, although the mutational rate of DNA loss caused by small indels can vary evolutionarily, and might itself be the result of selection, it cannot explain the observed differences in noncoding DNA content and genome size between species. Moreover, this mutational explanation cannot account also for the observed differences across genomes.

4
Transposable Elements (TE) and DNA Gain

TE elements are widespread in eukaryotic genomes and comprise in many cases a substantial portion of euchromatic genome size, including humans where TEs explain 45% of intergenic and intronic sequences. The recent invasion/amplification of particular TE families causes rapid increases in genome size, temporarily altering any specific equilibrium among selective and long-term mutational tendencies. Two examples of such effect are maize, in which genome size has recently doubled due to TE invasion, and Drosophila, with substantial differences in TE presence between closely related species, as well as among populations of the same species. The opposite trend is also detected, with species that have recently lost repetitive DNA showing a sudden reduction in genome size albeit having similar mutational tendencies associated with small indels. For instance, tetraodontid (smooth) and diodontid (spiny) pufferfish have similar DB but smooth pufferfish shows a twofold reduction in genome size because of a recent reduction in repetitive elements. This latter case also illustrates that any study of mutational tendencies toward DNA gain or loss should include small indels (i.e. DB), as well as the rate of large insertions and TE amplification.

It is possible to argue that a generic mutational tendency of TE to replicate and increase genome size might have favored a mutational mechanism associated with DNA repair of small indels with a tendency toward DNA loss, as observed in most eukaryotes. In this regard, small differences in DB among taxa would reflect long-term tendencies associated with TE amplification in the genome. The additional influence of selective forces that can be species specific and/or vary across genomes would generate the final outcome.

5
Variation across and among Genomes: The Influence of Recombination

Many genomic analyses have revealed heterogeneity of noncoding DNA presence across eukaryotic genomes. In Drosophila, the amount of intergenic DNA is higher in genomic regions with low recombination (and crossing-over) rates. Introns are also longer and more frequent in genes located in regions with severely reduced rates of crossing-over. An equivalent trend is observed in humans, with longer intergenic sequences, and introns in isochores are associated with reduced rates of crossing-over. *S. cerevisiae* also shows an influence of recombination on the presence of noncoding DNA. A single exception to this tendency is observed in *C. elegans,* in which an atypical pattern of crossing-over along chromosomes has been reported.

Interestingly, a negative relationship between recombination rates and the amount of noncoding DNA is observed among species. The study of nine distant eukaryotes with well-characterized genetic and physical maps (Fig. 1) reveals a significant negative relationship between recombination rate and a measure of the amount of noncoding DNA once the number of genes is taken into account. (Albeit the study is based on a small number of species, the observed relationship is not caused by any single species.) This observation suggests that similar, or at least overlapping, forces are shaping the size of noncoding DNA among species and

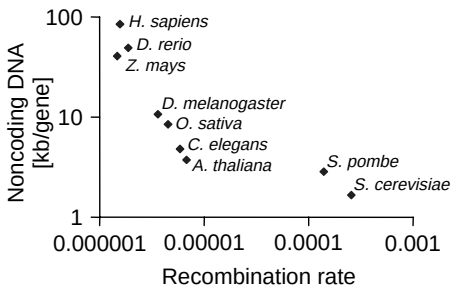

Fig. 1 Relationship between recombination rate and gene density among evolutionary distant model eukaryotes. The nine eukaryotes with well-characterized genetic and physical maps are: *Homo sapiens*, *Danio rerio* (zebra fish), *Zea mays* (maize), *Oryza sativa* (rice), *D. melanogaster*, *C. elegans*, *Arabidopsis thaliana*, *S. cerevisiae*, and *Schizosaccharomyces pombe* (fission yeast). Recombination rate indicates the average percentage of meiotic recombination (cM) between adjacent base pairs (total genetic map/total physical map), taking also into account the number of chromosomes to appraise the recombination caused by the random segregation of chromosomes. Although *A. thaliana* is a predominantly selfing plant, this feature has only recently been acquired, with all closely related species being outcrossing and, therefore, the recombination rate used for *A. thaliana* reflects the recent ancestral outcrossing condition. Nonparametric Spearman's rank correlation $R = -0.95$ ($p = 0.00009$). The least significant relationship after removing any one species is $R = -0.929$ ($p = 0.0009$). The removal of both distantly related yeasts (*S. cerevisiae* and *S. pombe*) from the analysis does not influence the outcome either ($R = -0.893$, $p = 0.007$).

across genomes. These equivalent findings across and among genomes could be explained if the mechanism of mismatch repair associated with recombination instigates DNA loss. Although this possibility needs to be investigated thoroughly in different species, in *D. melanogaster* there are similar patterns of indels in regions with normal and reduced rates of crossing-over.

An alternative explanation, invoking natural selection, emerges when we consider the influence that recombination has on the effectiveness of selection. Indeed, theoretical and experimental investigations predict that recombination will have an impact on the power or effectiveness of selection. In particular, most models of selection forecast that recombination will raise the effectiveness of selection. The total length of genetic map in a species as well as the number of chromosomes will play a role in the total meiotic recombination rate. Recombination rates and *Ne* (mentioned earlier) are two parameters that have been usually limited to population genetics studies but they should be included in any general explanation to genome architecture, the amount of noncoding DNA and genome size due to their influence to the effectiveness of selection. Moreover, both recombination and *Ne* are influenced by species-specific factors as well as vary across genomes.

6 Noncoding DNA is not Free of Genetic Information or Evolutionary Effect

Noncoding DNA in eukaryotes has been often labeled as "junk" DNA, useless (with no genetic information) as well as with no evolutionary effect (neutral to selection). In fact, however, noncoding DNA can be evolutionarily relevant by containing genetic information as well by its only presence. The number of known noncoding genes and especially noncoding RNA genes (ncRNA) has recently increased rapidly. Many of these ncRNA function directly as RNA while others play an essential role in gene regulation, with antisense RNAs being an important class of these genes. Moreover, noncoding DNA,

intergenic or intronic, harbors most of the information on gene regulation.

On the other hand, the presence of DNA with no genetic information between genes or between exons should not be assumed to be devoid of evolutionary effect. For instance, the presence of introns and their length can have a direct fitness effect on transcription costs and be a trait under selection. However, the presence of DNA sequences with no direct fitness effects can also be a selective trait. Note that the significant parameter associated with recombination is the total recombination between genes or exons (i.e. clusters of sites under selection) and this is influenced by the rate per physical unit as well as by the physical distance. Therefore, under realistic conditions for most eukaryotes, an organism might increase the effectiveness of selection by increasing the amount of noncoding DNA, intergenic or intronic. Under this scenario, the presence of "junk" DNA can be viewed as a modifier of recombination with a definite evolutionary long-term effect and can itself be a selective trait. The long-term fitness effects of increasing the physical distance between clusters of sites under selection (genes, exons, or regulatory regions) and recombination, is expected to differ among species (according to the overall recombination rate) and across genomes (according to variation in recombination rates along chromosomes).

7 Summary

The study of noncoding DNA is no longer restricted to phenotypic correlations and all recent observations support the notion that noncoding genome size in eukaryotes is the result of a highly dynamic balance between mutational and selective forces. Population parameters that influence levels and patterns of polymorphisms also influence the effectiveness of selection and have the potential to influence genomic features. Therefore, the evolution of noncoding DNA (intergenic and intron presence, and size) ought to be investigated from a population genetics perspective (i.e. the so-called *population genomics* approach). The presence of regulatory sequences and noncoding genes (e.g. ncRNA genes) would add genetic information required for organismal "complexity," but it can explain only a fraction of noncoding DNA.

Mutational tendencies are strongly influenced by TE invasion and amplification, which can be species specific and cause sudden changes in genome size. This "one-way ticket to genome obesity" is somewhat counteracted by a moderate tendency toward DNA loss due to small indels observed in most eukaryotes. The widespread presence of TEs in eukaryotes might have caused a selective pressure on mechanisms of DNA repair favoring DNA loss. The study of small indels only cannot produce an accurate picture of mutational tendencies. On the other hand, selective forces on the presence of noncoding DNA can vary in different species and/or taxa and be associated with replication costs, nuclear architecture, or other biological effects of bulk DNA. Importantly, the actual impact of these selective tendencies is determined by recombination rates and *Ne* parameters that can vary across genomes, as well as among species and should be included in any general explanation to genome evolution. Future studies on variation of recombination rates, indel evolution, and *Ne* in many eukaryotes will give us a broader and more

precise picture on the selective and mutational forces influencing the evolution of noncoding DNA.

Acknowledgment

The author thanks Dr. A. Llopart for her critical review of the manuscript. Josep M. Comeron is funded by National Science Foundation (NSF), DEB03-44209.

See also Noncoding Tandemly Repeated DNA Sequences; Recombination and Genome Rearrangements.

Bibliography

Book and Reviews

Cavalier-Smith, T. (1982) Skeletal DNA and the evolution of genome size, *Annu. Rev. Biophys. Bioeng.* **11**, 273–302.

Cavalier-Smith, T. (1985) *The Evolution of Genome Size*, John Wiley & Sons, New York.

Comeron, J.M. (2001) What controls the length of noncoding DNA? *Curr. Opin. Genet. Dev.* **11**, 652–659.

Gregory, T.R. (2002) Genome size and developmental complexity, *Genetica* **115**, 131–146.

Hartl, D.L. (2000) Molecular melodies in high and low C, *Nat. Rev. Genet.* **1**, 145–149.

Ohno, S. (1972) So much 'junk' DNA in Our Genome, in: Smith, H.H. (Ed.) *Evolution of Genetic Systems*, Gordon and Breach, New York, pp. 366–370.

Primary Literature

Awadalla, P. (2003) The evolutionary genomics of pathogen recombination, *Nat. Rev. Genet.* **4**, 50–60.

Bennett, M.D. (1971) The duration of meiosis, *Proc. R. Soc. Lond., B, Biol. Sci.* **178**, 259–275.

Bensasson, D., Petrov, D.A., Zhang, D.X., Hartl, D.L., Hewitt, G.M. (2001) Genomic gigantism: DNA loss is slow in mountain grasshoppers, *Mol. Biol. Evol.* **18**, 246–253.

Biemont, C. (1999) Distribution of transposable elements in Drosophila species, *Genetica* **105**, 43–62.

Blumenstiel, J.P., Hartl, D.L., Lozovsky, E.R. (2002) Patterns of insertion and deletion in contrasting chromatin domains, *Mol. Biol. Evol.* **19**, 2211–2225.

Carvalho, A.B., Clark, A.G. (1999) Intron size and natural selection, *Nature* **401**, 344.

Castillo-Davis, C.I., Mekhedov, S.L., Hartl, D.L., Koonin, E.V., Kondrashov, F.A. (2002) Selection for short introns in highly expressed genes, *Nat. Genet.* **31**, 415–418.

Cavalier-Smith, T. (1978) Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox, *J. Cell Sci.* **34**, 247–278.

Charlesworth, B. (1996) The changing sizes of genes, *Nature* **384**, 315–316.

Claverie, J.M. (2000) What if there are only 30,000 human genes? *Science* **291**, 1255–1257.

Comeron, J.M. (2004) Selective and mutational patterns associated with gene expression in humans: influences on synonymous composition and intron presence, *Genetics* **167**, 1293–1304.

Comeron, J.M., Kreitman, M. (2000) The correlation between intron length and recombination in Drosophila. Dynamic equilibrium between mutational and selective forces, *Genetics* **156**, 1175–1190.

Comeron, J.M., Kreitman, M. (2002) Population, evolutionary and genomic consequences of interference selection, *Genetics* **161**, 389–410.

Comeron, J.M., Kreitman, M., Aguade, M. (1999) Natural selection on synonymous sites is correlated with gene length and recombination in Drosophila, *Genetics* **151**, 239–249.

Crow, J.F., Kimura, K. (1970) *An Introduction to Population Genetics Theory*, Harper and Row, New York.

Day, T., Gloor, G.B. (1997) Homology requirements for targeting heterologous sequences

during P-induced gap repair in Drosophila melanogaster, *Genetics* **147**, 689–699.

Devos, K.M., Brown, J.K.M., Bennetzen, J.L. (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis, *Genome Res.* **12**, 1075–1079.

Doolitle, W.F., Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution, *Nature* **284**, 601–603.

Duret, L., Mouchiroud, D., Gautier, C. (1995) Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores, *J. Mol. Evol.* **40**, 308–317.

Evgen'ev, M., Zelentsova, H., Mnjoian, L., Loluectova, H., Kidwell, M.G. (2000) Invasion of Drosophila virilis by the Penelope transposable element, *Chromosoma* **109**, 350–357.

Graur, D., Shuali, Y., Li, W.-H. (1989) Deletions in processed pseudogenes accumulate faster in rodents than in humans, *J. Mol. Evol.* **28**, 279–285.

Gregory, T.R. (2004) Insertion-deletion biases and the evolution of genome size, *Gene* **324**, 15–34.

Hey, J. (1998) Selfish genes, pleiotropy and the origin of recombination, *Genetics* **149**, 2089–2097.

Hey, J., Kliman, R.M. (2002) Interactions between natural selection, recombination and gene density in the genes of Drosophila, *Genetics* **160**, 595–608.

Hill, W.G., Robertson, A. (1966) The effect of linkage on limits to artificial selection, *Genet. Res.* **8**, 269–294.

Kliman, R.M., Irving, N., Santiago, M. (2003) Selection conflicts, gene expression, and codon usage trends in yeast, *J. Mol. Evol.* **57**, 98–109.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Mirsky, A.E., Ris, H. (1951) The desoxyribonucleic acid content of animal cells and its evolutionary significance, *J. Gen. Physiol.* **34**, 451–462.

Neafsey, D.E., Palumbi, S.R. (2003) Genome size evolution in pufferfish: a comparative analysis of diodontid and tetraodontid pufferfish genomes, *Genome Res.* **13**, 821–830.

Parsch, J. (2003) Selective constraints on intron evolution in Drosophila, *Genetics* **165**, 1843–1851.

Petrov, D.A. (2002a) DNA loss and evolution of genome size in Drosophila, *Genetica* **115**, 81–91.

Petrov, D.A. (2002b) Mutational equilibrium model of genome size evolution, *Theor. Popul. Biol.* **61**, 531–544.

Petrov, D.A., Hartl, D.L. (1998) High rate of DNA loss in the Drosophila melanogaster and Drosophila virilis species groups, *Mol. Biol. Evol.* **15**, 293–302.

Petrov, D.A., Hartl, D.L. (2000) Pseudogene evolution and natural selection for a compact genome, *J. Hered.* **91**, 221–227.

Petrov, D.A., Lozovskaya, E.R., Hartl, D.L. (1996) High intrinsic rate of DNA loss in Drosophila, *Nature* **384**, 346–349.

Petrov, D.A., Chao, Y.C., Stephenson, E.C., Hartl, D.L. (1998) Pseudogene evolution in Drosophila suggests a high rate of DNA loss, *Mol. Biol. Evol.* **15**, 1562–1567.

Petrov, D.A., Sangster, T.A., Johnston, J.S., Hartl, D.L., Shaw, K.L. (2000) Evidence for DNA loss as a determinant of genome size, *Science* **287**, 1060–1062.

Petrov, D.A., Aminetzach, Y.T., Davis, J.C., Bensasson, D., Hirsh, A.E. (2003) Size matters: non-LTR retrotransposable elements and ectopic recombination in Drosophila, *Mol. Biol. Evol.* **20**, 880–892.

Powell, J.R. (1997) *Progress and Prospects in Evolutionary Biology. The Drosophila Model*, Oxford University Press, New York.

Ptak, S.E., Petrov, D.A. (2002) How intron splicing affects the deletion and insertion profile in Drosophila melanogaster, *Genetics* **162**, 1233–1244.

Robertson, H.M. (2000) The large srh family of chemoreceptor genes in Caenorhabditis nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses, *Genome Res.* **10**, 192–203.

San Miguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., Bennetzen, J.L. (1998) The paleontology of intergene retrotransposons of maize, *Nat. Genet.* **20**, 43–45.

Schaeffer, S.W. (2002) Molecular population genetics of sequence length diversity in the Adh region of Drosophila pseudoobscura, *Genet. Res.* **80**, 163–175.

Smit, A.F. (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes, *Curr. Opin. Genet. Dev.* **9**, 657–663.

Thomas, C.A. (1971) The genetic organization of chromosomes, *Annu. Rev. Genet.* **5**, 237–256.

Vinogradov, A.E. (1995) Nucleotypic effect in homeotherms: body-mass-corrected basal metabolic rate of mammals is related to genome size, *Evolution* **49**, 1249–1259.

Wright, S. (1931) Evolution in Mendelian populations, *Genetics* **16**, 97–159.

Zuckerkandl, E. (1976) Gene control in eukaryotes and the c-value paradox ''excess'' DNA as an impediment to transcription of coding sequences, *J. Mol. Evol.* **9**, 73–104.

Zuckerkandl, E. (2002) Why so many noncoding nucleotides? The eukaryote genome as an epigenetic machine, *Genetica* **115**, 105–129.

Replication and the Cell Cycle

Helena E. Richardson[1], *Leonie M. Quinn*[1], *Nancy Amin*[1], *Anabel Herr*[1], *Julie Secombe*[2], *Anthony M. Brumby*[1], *Kathryn M. Doyle*[1], *and Patrick O. Humbert*[1]
[1] *Peter MacCallum Cancer Centre, Melbourne, Australia*
[2] *Fred Hutchinson Cancer Center, Seattle, Washington, DC*

1 Overview of the Cell Cycle: An Historical Perspective 143
1.1 Stages of the Cell Cycle 143
1.2 Existence of Cell Cycle Controls 144
1.3 Approaches to Analyze the Cell Cycle 144

2 The Molecular Basis of G1-S Phase Regulation 146
2.1 Events of the G1-S Phase Transition 146
2.2 Key Regulators of the G1-S Phase Transition 146
2.2.1 Cyclins and Cyclin-dependent Kinases (Cdks) 146
2.2.2 Cks (Suc1, p13) 155
2.3 The Retinoblastoma-E2F Pathway – A Target of Cyclin/Cdks in G1-S Phase Progression 155
2.3.1 Rb Regulates the G1-S Restriction Point 155
2.3.2 Current Model of Cell Cycle Control by Rb and E2F Family Complexes 156
2.3.3 Regulation of Rb, E2F, and DP Function by Cdks 158
2.3.4 Target Gene Regulation by E2F and Rb 158
2.3.5 Functional Analysis of Rb, E2F, and DP in Drosophila 160
2.3.6 Functional Analysis of Mammalian Rb, E2F, and DP: Lessons from Mutant Mice 161
2.4 Regulation of G1-S Cyclin/Cdk Activity 164
2.4.1 Regulation of Cdk Activity by Phosphorylation 164
2.4.2 Regulation of G1-S Cyclin/Cdk Activity by Cdk Inhibitors 165
2.4.3 Regulation of G1-S by Ubiquitin-mediated Proteolysis 168
2.5 Regulation of Cyclin and Cdk Inhibitors by Extracellular Signals 172
2.5.1 Growth-factor Signaling Pathways in G1-S Regulation 172

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

2.5.2 Regulation of G1-S Phase Progression by Anchorage, Contact Inhibition, and Tissue Architecture 184
2.6 Regulation of G1-S Phase by the DNA-damage and DNA-replication Checkpoint: p53 187
2.7 G1-S Cyclin/Cdks and Progression into S Phase 189
2.7.1 Centrosomal Duplication 189
2.7.2 Initiation of DNA Replication 190

3 Summary and Future Directions 197

Acknowledgments 197

Bibliography 197
Books and Reviews 197
Primary Literature 199

Keywords

Adherens Junctions
Cell–cell junction, mediated by the cadherins, which are transmembrane proteins that form contacts with each other extracellularly and anchor to the actin cytoskeleton via catenins, intracellularly.

Anchorage Dependence
The dependence for cells to attach to an extracellular matrix, mediated by integrins, before progression from G1 to S phase can occur.

Cdk Inhibitor (Cki)
Proteins of two different classes that act to inhibit Cdks; the CIP(KIP)-family proteins ($p21^{CIP1}$, $p27^{KIP1}$ and $p57^{KIP2}$) bind to and inhibit Cyclin/Cdk activity, while the INK4 family ($p16^{INK4A}$, $p15^{INK4B}$, $p18^{INK4C}$ and $p19^{INK4D}$) bind to Cdk4(6) and prevent Cyclin binding to the Cdk.

Cell Growth
The increase in cell mass due to increased rates of mRNA and protein synthesis that is distinct from cell proliferation, which is an increase in cell numbers due to passage through the cell cycle.

Cell Polarity
The structure or shape of a cell, which is determined by the specific location of membrane-localized protein complexes along the apical–basal axis.

Centrosome (Microtubule-organizing Center, MTOC)
An organelle that is duplicated at the onset of S phase, and functions in mitosis to nucleate microtubules to form the mitotic spindle.

Checkpoint
A surveillance mechanism preventing cell cycle progression until a particular event is completed. G1-DNA-damage checkpoint arrests cells in G1 until DNA damage can be repaired. DNA-replication checkpoint, blocks cell cycle progression into mitosis when DNA replication is inhibited.

Contact Inhibition
The arrest of the cell cycle in G1 phase that occurs when cell density increases to confluence in culture, mediated by cell–cell contact via cadherins at the adherens junction.

Cop9 Signalosome (CSN)
A complex that is required to cleave Nedd8 (a ubiquitin-like protein) from Cullin proteins in the SCF and that is important for SCF activity.

Cyclin
The regulatory subunit required for the activity of cyclin-dependent kinases, many of which are unstable proteins, regulated by ubiquitin-mediated proteolysis.

Cyclin-dependent Protein Kinase (Cdk)
A family of serine–threonine protein kinases, which require the binding of a cyclin subunit for activity. Cdk1, 2, 3, 4, and 6 are required for progression through specific cell cycle phases.

Extracellular Matrix
An organized array of extracellular proteins, such as laminin and fibronectin that acts as anchors and receptors for the basally localized transmembrane proteins, integrins.

Mitogen
An extracellular growth factor that stimulates cell growth and proliferation.

Prereplication Complex (Pre-RC)
A protein complex, composed of ORC1-6, Cdc6, Cdt1, and MCM2–7, that binds to origins of DNA replication in G1 phase, and is required to set the origins for the initiation of DNA replication.

Restriction Point
A point late in G1, where a cell becomes committed to enter and complete an entire cell cycle independent of the presence of growth factors. Defined by the inactivation of the retinoblastoma tumor suppressor protein by Cyclin/Cdk phosphorylation.

SCF
A complex containing, Skp1, Cullin, and an F-box protein, and Rbx1 (or Roc1, Hrt1, a ring-finger protein) that acts as an E3 ubiquitin ligase. The F-box protein (such as Skp2, Ago/Fbw7, or Slimb) confers substrate specificity.

Septate Junctions
Drosophila cell–cell junctions, located basal to the adherens junctions, where Scribble and Dlg are located.

Tissue Architecture
The three-dimensional arrangement of individual cells to each other, which creates the organized structure of a tissue in a whole organism. It is mediated by the polarity of individual cells and the linkage of adjacent cells via intercellular junctions.

Tumor Suppressor
A protein that acts to prevent cell proliferation and tumorigenesis, and is inactivated in cancers.

Ubiquitin Ligase (E3)
A protein complex that transfers ubiquitin (a 76-residue protein) from an E2 ubiquitin-conjugating enzyme to the substrate, thereby tagging it for ubiquitin-mediated degradation.

It is critical to strictly regulate cell proliferation for the faithful duplication of DNA and the transmission of chromosomes to the next generation. This fundamental cellular process is known as the *cell cycle*. The cell cycle has been extensively studied since the 1970s, and many seminal studies have elucidated cell cycle regulatory mechanisms and how these controls fail in proliferative disorders, such as cancer. This review focuses on the underlying mechanisms that regulate the progression of cells into the DNA-replication phase (the G1-S phase transition), since it is at this stage in mammalian cells where cell proliferation is predominantly regulated. To enter into S phase, cells must pass a stage known as the *restriction point* where the cell monitors the relative level of positive versus negative extracellular signals, before it commits to progress into and completes an entire cell cycle. Key protein kinases, the cyclin/Cdks, are critical in progression past the restriction point. It is at this stage that cyclin/Cdks phosphorylate and inactivate the tumor suppressor protein, retinoblastoma (Rb), which acts to block the activity of the E2F S-phase gene transcription factor. The accumulation and activity of cyclin/Cdks is determined by the level of cyclin transcription, translation, and proteolysis, and in turn affect the production of Cdk inhibitors that block Cdk activity. Extracellular signals, via growth factors and from contacts between cells or the extracellular matrix, influence the accumulation and activity of cyclin/Cdks and are critical in determining whether

the cell will progress into S phase. Environmental assaults also impact on the G1-S phase progression, and lead to the upregulation of the p53 tumor suppressor protein, which by upregulation of Cdk inhibitors results in G1-S phase arrest until DNA damage or DNA-replication defects can be repaired. In S phase, a cell needs to duplicate the centrosome (an organelle required for the separation of chromosomes in mitosis) and to initiate DNA replication, which depends upon cyclin/Cdk phosphorylation of specific targets. The establishment of DNA-replication origins and the process of DNA-replication initiation is a highly regulated process, to ensure that the DNA is completely replicated once, and only once, per cell cycle. However, in particular, developmental situations in Drosophila and mammals, there are different modes of DNA replication, known as the *endoreplication cycles*, which result in amplification of the genome or parts of the genome without mitosis. The switch from the canonical cell cycle to the endoreplication cycles is also tightly regulated. Thus, the G1-S phase transition is a strictly regulated process, whose control within a whole organism is essential for limiting cell proliferation, and which when dysregulated can lead to cancer.

1 Overview of the Cell Cycle: An Historical Perspective

1.1 Stages of the Cell Cycle

The cell cycle is the process by which cells reproduce. For a eukaryotic cell, this requires the duplication of DNA and the faithful segregation of chromosomes into two daughter cells. When researchers first examined dividing cells under the microscope, they noted that most of the time little activity could be detected, but within a short time period the cell exhibited a dramatic rearrangement of cellular architecture where chromosomes were segregated into daughter cells. The cyclic stage of chromosome segregation was termed *mitosis*, and the rest of the cell cycle was termed *interphase*. It is now known, that within interphase, the continuous process of cell growth (the accumulation of cell mass) occurs, and at a defined stage, DNA replication occurs. To monitor DNA replication, cultures of eukaryotic cells could be synchronized at specific stages of the cell cycle (by nutrient or serum starvation or by using drugs to block DNA replication or mitosis) and cells labeled with the nucleotide analogs, H^3-thymidine or bromodeoxyuridine. These studies showed that DNA replication occurred at a precise stage in the cell cycle, termed *S phase*.

Thus, the eukaryotic cell cycle can be divided into four major phases; G1, S phase, G2, and mitosis (M phase). During the replication period (S phase), the genetic information is duplicated and then distributed into the two daughter cells during mitosis (M phase). These two phases are separated by so-called *Gap phases*, the G1 phase and the G2 phase. In most eukaryotic cell cycles, the G1 phase is where the most critical regulatory decisions are made. It is at this stage where, depending upon extracellular signals, the cell makes the decision to

commit to another complete cell division or to enter G0 (quiescence). The cell also needs to grow, to accumulate cell mass (including the synthesis of new RNA, proteins, ribosomes, organelles, and membranes) to acquire a critical cell size before committing to enter the cell cycle. In most mammalian cultured cells, the cycle time is 24 h, consisting of G1 (12 h), S phase (6 h), G2 (6 h), and mitosis (0.5 h). A notable exception are cultured embryonic stem cells which have a modified cell cycle circuitry that enables a 6 h cell cycle. Similarly, the cell cycle time in other organisms can be considerably shorter. For example, in budding yeast (*Saccharomyces cerevisiae*) or fission yeast (*Schizosaccharomyces pombe*) the cell cycle time is ~2 h, while in the developing embryos of the vinegar fly *Drosophila melanogaster*, it can be as little as 8 min. The rapid cell cycles, in yeast are due to their smaller genome size and the presence of abundant nutrient supplies, while in the Drosophila embryo, maternal contribution of all critical proteins allows DNA replication and cell division to occur rapidly, without the need for cell growth.

1.2 Existence of Cell Cycle Controls

Cell fusion studies carried out in the 1970s by Rao and Johnson, first revealed that progression through the cell cycle is a regulated process. These studies utilized the fusion of cells that were at different points in the cell cycle to demonstrate that there is an S phase–inducing activity present in S-phase cells, and an M phase–inducing activity in mitotic cells. In addition, these studies showed that there are *checkpoint controls* (surveillance mechanisms that block cell cycle transitions until a particular event is complete) that prevent rereplication of DNA and entry into mitosis prior to the completion of DNA replication.

Subsequent studies revealed additional checkpoint controls in cell cycle regulation. In the budding yeast, there is a checkpoint that monitors cell size (to ensure adequate growth), the presence of appropriate levels of nutrients and the absence of extracellular mating factors (a or α-pheromones), before allowing the cell to enter the cell cycle. This critical point, where a cell becomes committed to the cell cycle, is called *start* in yeast, while in more complex eukaryotes this checkpoint is called the *restriction point* (defined as the point, that once passed, a cell becomes committed to enter and complete a cell cycle without the need for growth factors). In multicellular eukaryotes, nutrient levels are less important, while the presence of positive or negative extracellular signals, such as growth factors and signals from neighboring cells or the extracellular matrix (ECM), determines whether a cell becomes committed to enter into the cell cycle, or exits the cell cycle into a quiescent (G0) state or differentiates.

1.3 Approaches to Analyze the Cell Cycle

The first molecular approaches to the analysis of the cell cycle came from two different camps; (1) the biochemical approach using mammalian cultured cells and the oocytes of marine organisms, and (2) the genetic approach using the budding yeast (*S. cerevisiae*) or fission yeast (*S. pombe*). From the biochemical approach, two seminal discoveries were made. First, the first members of a family of important regulatory proteins, the

cyclins, were purified from the eggs of marine organisms. These cyclins were so called because they varied dramatically in abundance during the cell cycle, and also because of their cyclic synthesis and destruction. Second, the factor important in driving the cell cycle (mitosis promoting factor, MPF) was identified and purified, and its molecular nature determined by protein sequencing. The genetic approach was based on the isolation of conditional lethal mutants (temperature-sensitive mutants that were viable at 20 °C but lethal at 37 °C) that affected the cell cycle, and their subsequent phenotypic characterization enabled the identification of important cell cycle functions. Genetic analysis revealed hierarchical relationships between genes in cell cycle control (see Fig. 1). Moreover, complementation cloning and DNA sequencing enabled the identification of the cell cycle gene at a molecular level.

In the late 1980s, studies from the genetic and biochemical perspectives converged to reveal that the proteins identified to be key drivers of cell cycle progression in marine organisms and in mammalian cultured cells, were the same as those identified by genetic studies in yeast. Thereby, the cell cycle engine components were revealed to be a family of serine–threonine protein kinases, the cyclin-dependent protein kinases (Cdks) and their regulatory subunits, cyclins. The cyclin-Cdk complex, in its various different flavors, was subsequently shown to be conserved in all eukaryotes and to drive progression from G1 to S phase and from G2 to M phase.

This review will focus on the G1-S phase transition and the initiation of DNA replication. In this review, we will primarily focus on analyses of mammalian cultured cells and *in vivo* studies using multicellular animal model systems, the mouse and the fly, Drosophila. We will describe the key events of G1-S phase, the regulators of the G1-S phase progression, modes of regulation of the key regulators, the signaling pathways that regulate G1-S phase progression, the control of

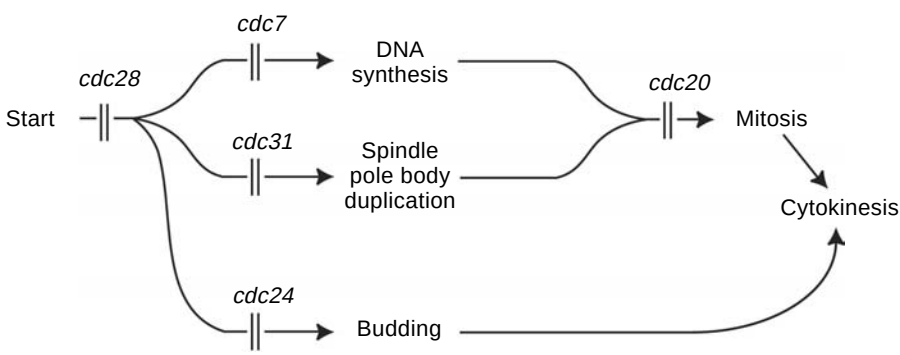

Fig. 1 Genetic analysis of yeast cell cycle mutants. Genetic analysis of budding yeast (*S. cerevisiae*) cell cycle mutants defined a cell cycle pathway. Mutants in specific genes blocked particular aspects of the cell cycle, thus revealing the cell cycle function of many genes. Mutants in *cdc7* blocked DNA synthesis, in *cdc31* blocked duplication of the spindle pole body (the yeast equivalent of the centrosome), in *cdc24* blocked cell budding (an event specific to the biology of budding yeast), and in *cdc20* blocked mitosis. Importantly, mutants in *cdc28* (*cdc2*, *cdk*) blocked all events of the G1-S-phase transition. Thus, the function of Cdc28, a serine–threonine cyclin-dependent kinase (Cdk), defines the G1-S phase transition.

the initiation of DNA replication, and regulation of alternative modes of DNA replication.

2 The Molecular Basis of G1-S Phase Regulation

2.1 Events of the G1-S Phase Transition

In all eukaryotes, the G1-S phase transition is not only marked by the initiation of DNA replication but by another key event also, that of centrosome duplication. The centrosome, or the microtubule-organizing center, is an organelle required for the nucleation of microtubules in mitosis, and undergoes duplication at the G1-S phase transition. In addition, in budding yeast, the initiation of bud formation is another key event that characterizes passage into S phase. Genetic studies in yeast, have revealed mutations that block either DNA replication, centrosome duplication or budding, but mutation of only one gene, *cdc28*, encoding the yeast Cdk, blocks all three of these events (see Fig. 1). Thus, in budding yeast, the Cdk, Cdc28, is the key regulator of all the events of the G1-S phase transition, and therefore is said to be the regulator of *start*. In higher eukaryotes, while more complexity exists (see the following), Cdks are also critical for passage through the *restriction point* and entry into S phase.

In G1, the cell responds to the presence of nutrients in the case of yeast, and to the presence of positive growth factors in higher eukaryotes, by activation of signaling pathways that result in the accumulation of G1-phase cyclins. In addition, at the G1-S phase transition, transcriptional programs are initiated, which lead to the coordinated expression of genes required for DNA replication, the so-called *S-phase genes*. The regulators of the events of the G1-S phase transition and the molecular mechanisms controlling them will now be discussed.

2.2 Key Regulators of the G1-S Phase Transition

2.2.1 Cyclins and Cyclin-dependent Kinases (Cdks)

As discussed earlier, the key molecules involved in progression through the cell cycle are the cyclin-dependent protein kinases (Cdks) and their regulatory component, cyclins. Cyclins are characterized by a sequence motif of about 100 amino acids called the *cyclin box*, which is sufficient for association with Cdk proteins. The structural changes brought about by the binding of a cyclin to a Cdk have been investigated by X-ray crystallography using human Cyclin A and Cdk2. Monomeric Cdk remains inactive due to an α-helical region (αL12) that constrains the ATP-binding site and a "T-loop" domain that blocks the catalytic cleft. The binding of cyclin to the Cdk shifts the inhibitory T-loop domain and "melts" the αL12 helical region, allowing substrates to bind to and be phosphorylated by the Cdk.

Specific activation of Cdk activity during the cell cycle is achieved by the cyclic synthesis and destruction of cyclins. Different cyclins peak in abundance at different points in the cell cycle, which is due to both transcriptional regulation of *cyclin* gene expression and to cell cycle regulated ubiquitin-mediated proteolysis of the cyclin protein.

In yeast, there is only one Cdk (Cdc28 in *S. cerevisiae* or Cdc2 in *S. pombe*), which

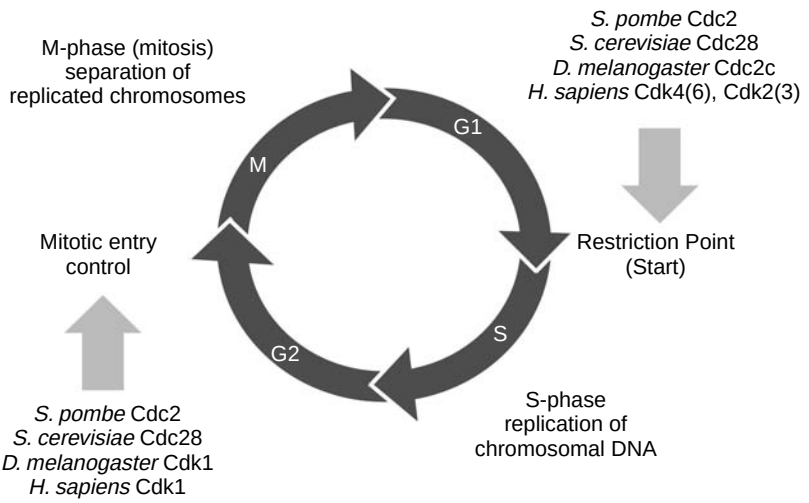

Fig. 2 Role of Cdks in cell cycle progression. The G1-S phase transition is driven by Cdc28 in *S. cerevisiae*, Cdc2 in *S. pombe*, Cdc2c in Drosophila, and by Cdk4(6) and Cdk2(3) in mammalian cells, while the G2-M phase transition is driven by Cdc28 in *S. cerevisiae*, Cdc2 in *S. pombe*, Cdk1 in Drosophila, and by Cdc2 (also termed *Cdk1*) in mammalian cells. The restriction point (or start in yeast) is a point late in G1 where cells become committed to enter and complete an entire cell cycle.

by association with different cyclins, regulate progression through the cell cycle (see Fig. 2). While in more complex eukaryotes, there are multiple Cdk family members that in association with different cyclins, have specific roles in G1-S or G2-M transitions (Fig. 2). In mammalian cells, Cdk1 is the primary protein kinase involved in the G2-M phase transition, while Cdk2 and Cdk3, and Cdk4 and Cdk6 play redundant roles in driving the G1-S transition.

In budding yeast, there are three groups of cyclins, Cln1-3, Clb5 and 6, and Clb1-4, which, in association with the Cdk, Cdc28, have specific roles during the cell cycle. Clb1-4 are important in the G2-M phase transition, while the Clns and Clb5(6) are essential for the G1-S phase transition. The Clns are important in initiation of bud formation while Clb5(6) are important for the initiation of DNA replication and for duplication of the centrosome (termed the *spindle pole body* in yeast). In addition, translational control of Cln3, in response to nutrient levels, has a role in coupling cell growth to cell division. In more complex eukaryotes, there are four major groups of cell cycle specific cyclins, Cyclin A1 and A2, B1-3, D1-3, and E1 and E2. These cyclins are more related to the yeast Clbs than the Clns; however, another cyclin, Cyclin C with a less defined role in cell cycle regulation (see the following), exhibits greater homology to yeast Clns. In mammalian cells, there are 11 Cdks that show strong homology to the yeast cell cycle Cdks; however, only Cdk1, 2, 3, 4, and 6 have specific roles in the G1-S and G2-M-phase transitions (Fig. 2). Mammalian cyclins associate with specific Cdks: Cyclin D associates with Cdk4 and Cdk6, Cyclin E with Cdk2 and Cdk3, Cyclin A with Cdk2 and Cdk1, and Cyclin B with Cdk1. Cdk1 functions at the G2-M phase transition,

while Cdk4(6) Cdk2(3) primarily function at G1-S and S phase. Cyclin A is unique in that it is able to interact with Cdk1 or Cdk2, and therefore has roles in G1-S and G2-M phase.

Studies in mammalian cultured cells have shown that Cdk2 has an essential role in the G1 to S-phase transition. In mammalian cells, Cyclins D, E, and A have different, essential roles in G1-S phase. Depletion of Cyclin D1, E, or A in cell culture by antibody injection or by using antisense results in G1 arrest, whereas overexpression of Cyclins D1, D2, E, or A promotes premature passage from G1 into S phase, showing that these cyclins are essential and rate limiting for G1-S progression. Furthermore, Cyclins D, E, and A have different roles during G1 and S phase, as revealed by the following results: (1) depletion of Cyclin D1 or E during G1 prevents entry into S phase, while depletion close to the initiation of DNA replication does not, while depletion of Cyclin A during S phase is able to inhibit DNA replication; (2) coexpression of Cyclin E and Cyclin D results in a greater acceleration of G1-S progression than expression of either cyclin alone. Thus, Cyclins D, E, and A act at different times during G1-S, and are likely to regulate different pathways during G1-S.

Specific aspects of the function of Cyclin D, E, and A will now be discussed, highlighting *in vivo* studies in Drosophila and the mouse.

2.2.1.1 **Cyclin A/Cdk2(3).** Cyclin A was first discovered by biochemical purification of a protein that showed dramatic fluctuation in the embryos of marine organisms, and was subsequently found to be conserved in all metazoans analyzed. In mammalian cells, *cyclin A* mRNA, protein, and associated kinase activity is present from late G1-M phase. As discussed earlier, Cyclin A can form complexes with Cdk1 and Cdk2 in mammalian cells, and can therefore have roles in the progression through the G1-S and G2-M phases.

While in mammalian cells there is evidence that Cyclin A/Cdk2 is required for entry into or progression through S phase, an S-phase role for Cyclin A in Drosophila is less clear. Drosophila Cyclin A interacts specifically with the mitotic Cdk, Cdk1, and not with the G1-S Cdk, Cdc2c, and *cyclin A* mutants arrest in G2 phase during embryogenesis (see Table 1). *Cyclin A* mutant embryos then enter ectopic endoreplicative Cyclin E-mediated S phases, indicating that Cyclin A is not normally required for S phase in Drosophila. There is, however, evidence that Cyclin A can function in the G1 to S phase transition, as ectopic expression of *cyclin A* can induce G1 cells to enter S phase.

In mammals, there are two Cyclin A homologs, Cyclin A1 and A2. In the mouse, Cyclin A1 is specifically expressed in the male germline and knockout of this gene results in male infertility (see Table 1). Cyclin A2 is more ubiquitously expressed in mouse development, and as such a mouse knockout of Cyclin A2 is early embryonic lethal (see Table 1), demonstrating the essential and nonredundant function of Cyclin A in cell proliferation *in vivo*. Conversely, overexpression of Cyclin A in mice leads to overproliferation and the development of cancer. The importance of appropriate control of Cyclin A *in vivo* is highlighted by the identification of Cyclin A as an oncogene in Hepatitis B-mediated hepatocellular carcinomas in humans.

2.2.1.2 **Cyclin E/Cdk2(3).** Human cyclin E was identified by its ability to rescue

Tab. 1 Phenotypes of knockouts/mutants of G1-S phase regulators in Drosophila and the mouse

	Drosophila (mutants)[a]	*Mouse (knockouts)*[b]		*Summary*
Cyclin D	Not yet characterized	$D1^{-/-}$	Viable with neurological defects, hypoplastic retina, and mammary glands	Dispensable for most embryonic cell cycles in the mouse.
		$D2^{-/-}$	Viable with fertility defects	
		$D3^{-/-}$	Viable with defects in T lymphocytes	
		$D1^{-/-}$ $D2^{-/-}$	Lethal neonatal with lack of cerebellar development	
		$D1^{-/-}$ $D3^{-/-}$	Lethal neonatal with neurological defects	
		$D2^{-/-}$ $D3^{-/-}$	Lethal neonatal with megaloblastic anemia	
		$D1^{-/-}$ $D2^{-/-}$ $D3^{-/-}$	Lethal E12.5-E16.5 with heart abnormalities and defects in hematopoietic stem cell expansion	
Cyclin E	Embryonic lethal. Essential for all cell cycles – maternal contribution until embryonic cycle 17	$E1^{-/-}$	Viable	Essential for G1-S in Drosophila, but dispensable for most cell cycles in mouse embryonic development.
		$E2^{-/-}$	Viable with reduced male fertility	
		$E1^{-/-}$ $E2^{-/-}$	Lethal E10.5-E11.5 due to endoreplication defects in placenta trophectoderm	
Cyclin A	Embryonic lethal – maternal contribution to cycle 16	$A1^{-/-}$	Viable with male infertility	Perform essential function(s) that cannot be compensated for by other Cyclins.
		$A2^{-/-}$	Lethal E5.5	
Cdk4/6	Viable with growth defects and reduced female fertility	$Cdk4^{-/-}$	Viable with proliferative defects in some endocrine cells	Dispensable for most cell cycles in Drosophila and the mouse. Mainly a regulator of cell growth in Drosophila.
		$Cdk6^{-/-}$	Viable with mild hematopoietic defects	
		$Cdk4^{-/-}$ $Cdk6^{-/-}$	Lethal late embryonic/neonatal with erythroid defects	
Cdk2	Larval lethal – maternal contribution during embryonic development	Viable with gonad atrophy and meiotic defects resulting in sterility		Consistent with requirement for G1-S in Drosophila and dispensability for most cell cycles in mouse development.

(*continued overleaf*)

Tab. 1 (*Continued*)

	Drosophila (mutants)[a]	*Mouse (knockouts)*[b]		*Summary*
P16-Ink4 family	Not present in Drosophila	$p15^{-/-}$	Viable with lymphoproliferative disorders and angiosarcomas	Not present in *Drosophila,* but play nonredundant roles in control of cell cycle exit and differentiation in multiple mammalian tissues.
		$p16^{-/-}$	Viable with thymic hyperplasia and increased tumor incidence	
		$p18^{-/-}$	Viable with increased body size and overgrowth of some organs, lymphoproliferative disorders, multiple cysts in kidney and mammary glands, pituitary tumors, and other malignancies	
		$p19^{-/-}$	Viable and fertile with males displaying testicular atrophy and progressive deafness	
		$p18^{-/-}$ $p19^{-/-}$	Viable with overlap of individual phenotypes and impaired spermatogenesis and infertility	
P21-Cip1 family	Embryonic lethal – maternal contribution of Dacapo until cycle 17; cells progress through another complete cell cycle.	$p21^{-/-}$	Viable and normal, cultured cells arrest inefficiently upon DNA damage	Essential for control of G1-S progression in Drosophila and important for control of cell cycle exit and differentiation in mammalian tissue.
		$p27^{-/-}$	Viable with increased body size and overgrowth of organs, female sterility, retinal dysplasia, and pituitary tumors with high penetrance	
		$p57^{-/-}$	Lethal neonatal with severe developmental defects including shortened limbs due to deregulated chondrocyte proliferation	
		$p21^{-/-}$ $p27^{-/-}$	Phenotype not characterized. Cultured cells defective in the formation of stable cyclin D–Cdk4 complexes	
		$p21^{-/-}$ $p57^{-/-}$	Lethal neonatal with skeletal muscle and lung alveoli defects	
		$p27^{-/-}$ $p57^{-/-}$	Lethal E12.5-E16.5 with hyperproliferation and differentiation defects in lens and trophoblast cells	

Rb	Rbf1$^{-/-}$	Larval lethal – maternal contribution during embryonic development	Rb$^{-/-}$	Lethal E13.5-E15.5 with defects in erythroid, neuronal, and lens development	pRb is important for cell cycle exit and differentiation of many tissues. P107 and p130 perform a redundant role in cell cycle exit and differentiation of chondrocytes and other tissues.
	Rbf2$^{-/-}$	Not yet characterized	p107$^{-/-}$	Viable	
			p130$^{-/-}$	Viable	
			p107$^{-/-}$ p130$^{-/-}$	Lethal neonatal – deregulated chondrocyte proliferation	
			Rb$^{-/-}$ p107$^{-/-}$	Lethal early embryonic	
			Rb$^{-/-}$ p130$^{-/-}$	Lethal early embryonic	
E2F	E2F1$^{-/-}$	Larval lethal	E2f1$^{-/-}$	Viable with thymic and testicular atrophy and increased tumor incidence	Dispensable for most cell cycles in Drosophila. Mouse activating E2F1/2/3 perform an essential function when repressive E2F4/5 are present. E2F6 is dispensable for cell cycle but has differentiation defects.
			E2f2$^{-/-}$	Viable with T cell defects; develop late onset autoimmune disease	
			E2f3$^{-/-}$	Lethal – neonatal, heart defects	
			E2f1$^{-/-}$ E2f2$^{-/-}$	Viable with defects in multiple hematopoietic lineages	
			E2f1$^{-/-}$ E2f3$^{-/-}$	Lethal – early embryonic	
			E2f2$^{-/-}$ E2f3$^{-/-}$	Lethal – early embryonic	
			E2f1$^{-/-}$ E2f2$^{-/-}$ E2f3$^{-/-}$	Lethal – early embryonic. Embryonic fibroblasts unable to enter S phase	
	E2F2$^{-/-}$	Weakly viable with reduced fertility	E2f4$^{-/-}$	Lethal neonatal, developmental defects including decreased size, intestinal, and hematopoietic defects	
			E2f5$^{-/-}$	Lethal neonatal, hydrocephaly	
			E2f4$^{-/-}$ E2f5$^{-/-}$	Lethal E16.5	
			E2f6$^{-/-}$	Viable with homeotic transformations of the axial skeleton	
	E2F1$^{-/-}$/E2F2$^{-/-}$	Pupal lethal			
DP	Larval/pupal lethal		Dp1$^{-/-}$	Lethal E9.5-E12.5 due to defects in extraembryonic (trophectoderm) development	Dispensable for most cell cycles in Drosophila and the mouse.
			Dp2$^{-/-}$	Not yet characterized	

[a]Many Drosophila genes have a large maternal contribution, and the phenotypes listed are at the stage of depletion of the maternal supply.
[b]Common mouse strains used for gene targeting (including C57Bl/6 and 129/sv) have an inactivating mutation in *cdk3*, therefore *cdk2* knockouts are also deficient in *cdk3*.

an yeast Cln-deficient strain, and thus was capable of substituting for the function of these cyclins in the G1-S phase transition in yeast. Mammalian Cyclin E mRNA, protein, and associated Cdk2(3) kinase activity peaks at late G1 phase and is rapidly degraded during S phase. As described earlier, in mammalian cultured cells, Cyclin E plays an essential role in G1 to S phase progression, and further studies have shown that downregulation of Cyclin E is crucial for the G1 arrest that must occur in terminally differentiating cells. Conversely, abnormal increases in expression of Cyclin E leads to premature entry of G1 cells into S phase and furthermore can lead to chromosome instability, one of the hallmarks of cancer.

The Drosophila homolog of *cyclin E*, *DmcycE* encodes two proteins with a common C-terminal and unique N-terminal regions. These two different Cyclin E variants are differentially expressed in proliferating cells during embryogenesis. *DmcycEII* mRNA is supplied maternally and is present during early embryonic division cycles, whereas *DmcycEI* mRNA is supplied zygotically and is expressed in all proliferating cells. Analysis of the *DmcycEI* promoter revealed the presence of multiple independent tissue and stage-specific regulatory elements, suggesting that *DmcycE* transcription is regulated by a variety of developmental cues. Drosophila Cyclin E interacts with the Drosophila homolog of the mammalian Cdk2 protein, Cdc2c, and is necessary and sufficient for the G1 to S phase transition. *DmcycE* loss-of-function mutants arrest in G1 of cycle 17 during embryonic development, due to the depletion of maternal supplies at this stage (see Table 1). Conversely, ectopic overexpression of DmcycE in G1-arrested cells drives these cells into a premature S phase. Furthermore, in Drosophila, overexpression of Cyclin E leads to defects in expression of morphogens and morphological movements during embryonic development, and in the eye imaginal disc, ectopic expression of Cyclin E disrupts normal pattern formation. Thus, Cyclin E is a critical regulator of the G1-S transition in Drosophila and its downregulation is important for normal development.

In mammals, there are two *cyclin E* genes, *Ccne1* and *Ccne2*. Recently, single and compound mutations of the *cyclin E* genes, and a knockout of *Cdk2* have revealed unexpected results (see Table 1). Mice singly deficient for *cyclin E1* or *E2* are viable with no apparent defects, with *cyclin E2*$^{-/-}$ males displaying decreased fertility indicating a specific requirement for Cyclin E2 in meiosis. Surprisingly, compound loss of *cyclin E1* and *E2*, results in a relatively late embryonic lethality at E10.5–E11.5 of development. By analysis of the tetraploid chimera, lethality at this stage was shown to be due to an extraembryonic requirement for *cyclin E* in the endoreplicating trophoblasts of the placenta. Provision of wild-type placenta to *cyclin E1*$^{-/-}$ *E2*$^{-/-}$ embryos resulted in survival to term of many embryos with the rest succumbing to cardiac abnormalities. Importantly, all other major organs examined appeared normal in *cyclin E1*$^{-/-}$ *E2*$^{-/-}$ embryos, indicating that Cyclin E activity is largely dispensable for normal mammalian development. However, a number of tissues are critically dependent on Cyclin E levels in the mouse. These included trophoblast giant cells required for placental function and megakaryocytes, both of which undergo repeated rounds of DNA synthesis without intervening mitoses (endoreplication) (see Sect. 2.7.2.2). These data highlight an important role for mammalian Cyclin E in driving the endoreplication cycles. Mouse embryonic

fibroblasts (MEFs) derived from the *cyclin E1*$^{-/-}$ *E2*$^{-/-}$ mutant mice, showed that *cyclin E* is dispensable for continual proliferation, but not for cell cycle reentry. Cell cycle reentry in the absence of Cyclin E activity resulted in defective DNA replication due to defects in the loading of the MCM protein complex onto DNA-replication origins (see Sect. 2.7.2).

Likewise, the analysis of Cdk2-deficient mice suggested that Cdk2 is dispensable for mitotic divisions and development (Table 1). *Cdk2*$^{-/-}$ mice are born viable with no gross anatomical or behavioral abnormalities; however, these mice are sterile due to defects in the meiotic cell cycle. Moreover, MEFs derived from these mice, proliferate normally in the absence of Cdk2. This is consistent with the results from the depletion of *cdk2* in human cancer cell lines. This surprising finding is unlikely to represent compensation by Cdk3, the other kinase bound by Cyclin E, as the mice used in these studies carry a predicted inactivating mutation in *Cdk3*. It is possible that Cdk1 can compensate for Cdk2, since Cyclin A/Cdk1 and Cyclin B/Cdk1 can drive cells through S phase and *Cdk2*$^{-/-}$ embryos contain unchanged levels of Cyclin A-associated kinase activity, presumably due to Cdk1. Therefore, Cyclin A/Cdk1 may compensate for S-phase functions normally mediated by Cyclin E/Cdk2 and Cyclin A/Cdk2 kinase. This is in contrast to the situation in Drosophila, where Cyclin E and Cdk2 are essential for cell cycle progression and normal development (Table 1). Interestingly, unlike the *cyclin E* knockout mouse phenotype, *Cdk2* mutants did not exhibit defects in the extraembryonic tissue, suggesting that Cyclin E may have Cdk2(3)-independent functions.

Mouse transgenic studies have shown that inappropriate expression of *cyclin E* can lead to tumorigenesis. Indeed, high levels of Cyclin E expression are associated with a broad spectrum of human malignancies, and in breast cancers, truncated versions of Cyclin E that are more stable and have increased associated kinase activity have been observed. Moreover, expression levels of Cyclin E often correlate with prognosis and the progression stage of human cancers, with higher levels associated with later stages of tumorigenesis.

2.2.1.3 **Cyclin D/Cdk4(6).** Cyclin D1 was also discovered as a gene capable of rescuing Cln-deficient yeast, suggesting that it can complex with the yeast Cdk, Cdc28, to drive the G1-S phase transition in yeast. As discussed earlier, in mammalian cells, Cyclin D is a family of three related proteins, D1, D2, and D3, which form a complex with Cdk4(6). D-type cyclins accumulate in response to mitogenic signals and have been proposed to provide the link between extracellular signaling pathways and the core cell cycle machinery.

In Drosophila, there is only one *cyclin D* and one *cdk4* homolog, and mutations have been now described in both genes (Table 1). Cyclin D-Cdk4 function is dispensable for survival, as flies that lack parental and zygotic Cdk4 function are semiviable. In addition, overexpression of Cyclin D-Cdk4 in embryos, has minimal effect on Cyclin E activity, suggesting that in Drosophila, Cyclin D has little role in titrating away inhibitors from the Cyclin E-Cdk2 complex. However, depending on the tissue or differentiation state, the role of Cyclin D-Cdk4 expression differs. In differentiating cells of the third instar larval eye imaginal disc, overexpression of Cyclin D-Cdk4 leads to increased cell growth (larger-sized cells), but in the proliferating cells of the wing disc, leads to coordinated cell growth and cell division.

In the mouse, the three D-type cyclins are expressed in a tissue-specific manner. Single knockouts of either *cyclin D* gene resulted in viable mice; however, hypoproliferative defects were observed in those tissues where the gene was maximally expressed (Table 1). Recently, mice have been generated that lack all D-type cyclins or Cdk4/6. Both *cyclin D*-deficient embryos and *cdk4/6*-deficient embryos survive until E15.5 of development, but display a severe anemia with megaloblastic features. Both *cyclin D*-deficient and $cdk4^{-/-}$ $cdk6^{-/-}$ fetal livers, show defects in multiple hematopoietic lineages. The similarity in phenotypes of the *cyclin D*-deficient and $cdk4^{-/-}$ $cdk6^{-/-}$ mice, indicate that the major function of Cyclin D in development is to activate Cdk4/6. Analysis of MEFs showed that Cyclin D and Cdk4/6 were not essential for cell cycle entry. However, Cyclin D-deficient and $cdk4^{-/-}$ $Cdk6^{-/-}$ MEFs were sensitive to loss of Cdk2 activity. Thus, while the majority of mouse tissues do not need Cyclin D or Cdk4(6), it appears that the loss of Cyclin D/Cdk4(6) can be compensated for by Cyclin E (or A)/Cdk2 complexes.

Transgenic mice overexpressing D-type *cyclins* show hyperplasia and increased incidences of tumors. Conversely, MEFs deficient in *cyclin D* or *Cdk4/6* were refractory to transformation by oncogenes. This oncogene-mediated transformation resistance also occurred with MEFs deficient in *cyclin E*, and to a lesser extent in Cdk2, suggesting that cells may require a certain Cdk threshold to enable transformation. Consistent with this, both Cyclin E and Cyclin D upregulation have been implicated in human cancers. Indeed, the oncogenes, *Prad1* (resulting in parathyroid adenoma) and *Bcl1* (resulting in B-cell lymphoma) are due to the translocation of the *cyclin D1* gene resulting in ectopic expression in parathyroid or B cells, respectively.

2.2.1.4 **Other cyclin-Cdks in G1-S progression.** There are several other classes of cyclins that are present in mammals that have less well-defined roles in G1-S phase regulation. These are Cyclin C, G, and F (discussed below). Other cyclin-Cdks have more general roles in the cell cycle, such as Cyclin H/Cdk7, which has Cdk activating kinase (CAK) activity and activates core cell cycle cyclin/Cdk5 during G1-S and G2-M phase by phosphorylating a critical threonine residue (T160 in Cdk2) (see Sect. 2.4.1). Cyclin H/Cdk7 also has a role in RNA polymerase II transcription, as does Cdk9 in association with T-type Cyclins and Cyclin K, and Cyclin C/Cdk8. Since these do not have specific roles in the G1-S phase transition, we will not discuss these further. Current knowledge of the role of Cyclin C, G, and F is summarized in the following.

Cyclin C was originally identified (along with Cyclin D and E) as a protein that when expressed in yeast was capable of rescuing a Cln-deficient strain, and has highest homology of all the mammalian cyclins to the Cln proteins. Cyclin C was also isolated from Drosophila by its ability to rescue the yeast *cln* mutant stain. Cyclin C associates with Cdk8 in both mammalian cells, and in Drosophila, and plays a role in RNA polymerase II transcription. Furthermore, Cyclin C/Cdk8 has been shown to phosphorylate specific transcription factors at their promoters, leading to their degradation via ubiquitin-mediated proteolysis (see Sect. 2.4.3). Recently, however, a complex between Cyclin C and Cdk3 has been observed and shown to play a role in passage from quiescence (G0) into the cell cycle. Indeed, Cyclin C/Cdk3 targets the same critical protein, as Cyclin D and

Cyclin E-Cdks, to mediate passage through G0-G1 (see Sect. 2.3).

Cyclin G was discovered as a target of the p53 transcription factor, which is a tumor suppressor that plays a critical role in DNA replication and DNA-damage checkpoint response (see Sect. 2.6). Further studies have shown that Cyclin G plays a role in negatively regulating the p53 pathway. Cyclin G has also been shown to promote cellular growth, and to form a complex with the protein kinase, GAK, which appears to play a role in modulating transmembrane receptor signaling pathways.

Cyclin F was discovered by its ability to rescue yeast mutants in Cdc4, which is a component of the S-phase ubiquitin-ligase complex, SCF (see Sect. 2.4.3). Cyclin F is closely related to Cyclin A, but also contains an F-box motif, which is a sequence present in Cdc4 that has a role in substrate recognition (see Sect. 2.4.3). Cyclin F expression is highest in S and G2 phases and is phosphorylated and rapidly degraded in mitosis. Recently, *cyclin F* (*cycF*) knockout mice have been generated, and show cellular defects suggesting a role in progression from G0 into S phase. While $cycF^{-/-}$ mice die at E10.5 due to defects in the proliferation of the chorionic trophoblast cells of the placenta, $cycF^{-/-}$ MEFs, although viable, are unable to reenter the cell cycle from quiescence (G0). In $cycF^{-/-}$ cells, the expression of Cyclin E was delayed, suggesting that Cyclin F may play a role in the control of Cyclin E expression; however, the precise role of Cyclin F in the progression from G0 into S phase remains to be determined.

2.2.2 Cks (Suc1, p13)

Another evolutionary conserved core protein present in the Cyclin/Cdk complex is Cks (also known as Suc1 or p13). Cks can promote the assembly and activation of yeast G1-S–cyclin/Cdk complexes *in vitro*. However, the function of Cks is not specific to the G1-S phase, and in yeast, appears to play a primary role in promoting mitosis through transcriptional regulation of an important gene, *cdc20*, which is required for proteolysis in mitosis. In mammalian cells, there are two homologs of Cks, Cks1, and Cks2, which play redundant roles in development. However, knockout of both genes in the mouse leads to early embryonic lethality, which precludes detailed analysis of specific cell cycle defects. Mammalian Cks1 also has a Cdk-independent role in promoting degradation of the cell cycle inhibitors, p27 and p130, via the SCF ubiquitin ligase (see Sect. 2.4.3), thereby promoting G1-S phase progression.

2.3 The Retinoblastoma-E2F Pathway – A Target of Cyclin/Cdks in G1-S Phase Progression

While much is known about the regulation of cyclin/Cdk activity, few targets of G1 cyclin/Cdk complexes have been described. In mammalian cells, the most extensively characterized target of cyclin/Cdk complexes is the retinoblastoma tumor suppressor (Rb) family of "pocket proteins", Rb, p107 and p130.

2.3.1 Rb Regulates the G1-S Restriction Point

The phosphorylation state of Rb proteins varies through the cell cycle, and is correlated with activity. Dephosphorylated Rb proteins are present in quiescent cells and cells in early G1, while phosphorylated forms of these proteins are present from late G1 until the end of mitosis. The growth inhibitory effects of the Rb family of proteins are associated with the hypophosphorylated forms that interact with

a family of transcription factors known as E2Fs. Hyperphosphorylated Rb proteins are unable to interact with E2Fs, allowing E2F-mediated activation of S phase and G1-S transition genes. Thus, the phosphorylation state of Rb proteins regulates the G1-S transition. Indeed, Rb is thought to define the *restriction point*: Once Rb becomes phosphorylated and inactivated, cells are committed to enter and complete a cell cycle driven through the E2F transcriptional program. E2Fs are required for the activation of genes important for the initiation and regulation of DNA synthesis such as *PCNA*, *RNR2*, *Cdc6*, *dihydrofolate reductase*, *DNA polymerase* α, and *thymidine kinase*, as well as genes required for the G1-S phase transition, such as *cyclin E*, *cyclin A*, and *cdk2*. In addition to being the main transcriptional controllers of G1 to S phase progression, E2Fs have now also been shown to regulate the expression of genes required for DNA repair, cell cycle checkpoint activation and maintenance and mitosis. Furthermore, E2Fs can also regulate genes involved in apoptotic signals as well as differentiation.

2.3.2 Current Model of Cell Cycle Control by Rb and E2F Family Complexes

E2Fs comprise a group of at least seven related proteins that can be divided into distinct subsets, on the basis of differences in both their transcriptional properties and their interaction with the Rb-family members, Rb, p107, and p130 (see Fig. 3). One subgroup consists of E2Fs that are potent transcriptional activators (E2F1, E2F2, E2F3a) while another subgroup behaves as transcriptional repressors (E2F3b, E2F4, E2F5, E2F6, E2F7). The *E2F3* locus gives rise to two protein isoforms (E2F3a and E2F3b) that differ only in their N-terminal region but have opposite transcriptional activities. With the exception of E2F7 that homodimerizes, binding of E2F moiety to one of two DP partners (DP1 or DP2) allows competent DNA binding to promoter regions of target genes at a consensus E2F binding site, TTTSSCGC (where S refers to C or G).

Extensive *in vitro* biochemical studies have provided a concise model for how various E2F/Rb complexes regulate cell cycle progression (see Fig. 4). Early in the cell cycle (G0 and G1), E2F4 and E2F5 (and presumably E2F3b) regulate transcriptional repression at E2F-dependent promoters by binding to p107 and p130, leading to the recruitment of a corepressor complex containing histone deacetylases (HDAC), histone methyltransferases (HMT), Sin3 and members of the SWI(SNF) family. At the G1-S transition, transcriptional repression is relieved through the phosphorylation of Rb-related proteins by cyclin-dependent kinases, which releases the "repressor" E2F complexes. This allows the "activator" E2Fs, E2F1, 2 and 3a, and coactivator complex that includes p300(CBP) to occupy the promoter and leads to transcriptional activation of E2F-dependent promoters. In this manner, E2Fs can participate in both the repression and activation of E2F-responsive genes, dependent on their association with Rb-family members. Although this model is well supported by chromatin immunoprecipitation (ChIP) studies, identifying the E2F species bound to promoters of E2F target genes through the cell cycle of *in vitro* cell lines, it must be noted that *in vivo* studies indicate that this model may be somewhat simplistic and the transcriptional properties for a given E2F may be modified depending on the developmental context (see Sect. 2.3.6).

In contrast to E2F1–5, E2F6 and E2F7 are not bound by Rb-family proteins and are therefore regulated differently. E2F6 does not contain an Rb-family

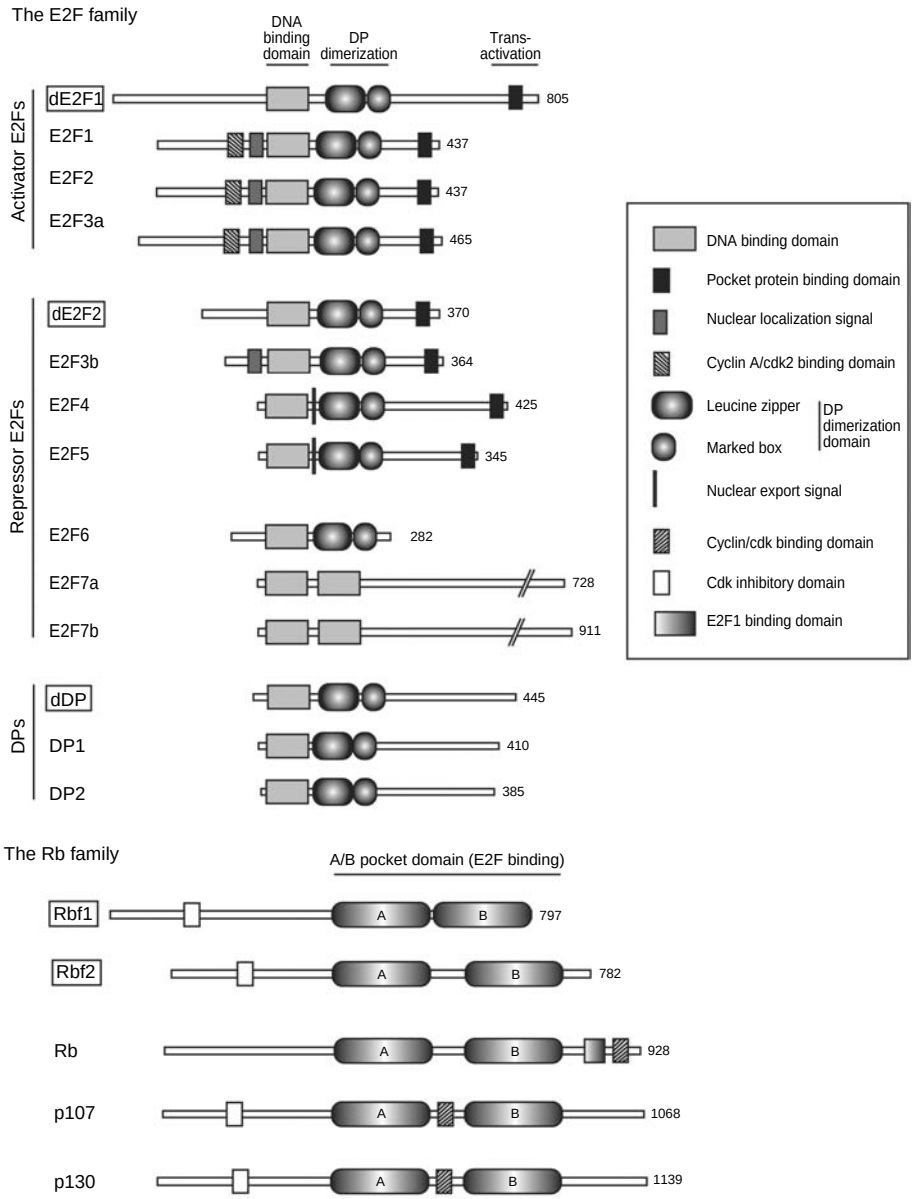

Fig. 3 Structure of the E2F, DP, and Rb family of proteins. The structure of the E2F, DP, and Rb-family members from mammalian cells and Drosophila are compared. Drosophila protein names are boxed.

binding domain and acts to repress transcription by recruiting the Polycomb group of chromatin-remodeling factors to promoters. Recent studies indicate that it may play a redundant role with E2F4 in regulating the G1-S and G2-M transitions.

E2F7 functions independently of a DP partner and of Rb-family members and is thought to repress transcription and cell cycle progression by competing with the activating E2F members at a subset of E2F promoters.

2.3.3 Regulation of Rb, E2F, and DP Function by Cdks

Rb has 15 consensus G1 cyclin/Cdk phosphorylation sites that are conserved among a wide variety of organisms. *In vitro*, human Cyclin D/Cdk4(6) and Cyclin E/Cdk2 are able to phosphorylate Rb at different sites, with neither being sufficient to phosphorylate all sites observed *in vivo*, indicating that both of these kinases are involved in the phosphorylation of Rb. Indeed, by selectively inhibiting Cdk2 or Cdk4(6) function, it has been shown that Cyclin D complexes are only able to partially phosphorylate Rb, and this phosphorylation is required for the action of Cyclin E/Cdk2 on Rb. These data have given rise to a model of sequential phosphorylation of Rb, initially by Cyclin D/Cdk4(6) followed by Cyclin E/Cdk2. There is also strong evidence that Rb is the only essential target of Cyclin D/Cdk complexes, as the G1 arrest induced by injection of anti-Cyclin D antibodies is dependent on wild-type Rb function. By contrast, Cyclin E is required for entry into S phase in Rb-deficient cells, indicating that Cyclin E/Cdk2 has additional essential targets in the G1 to S phase transition. Although these additional targets are likely to include the Rb-related proteins p107 and p130, Cyclin E is still required for the G1 to S phase transition after collective inactivation of all pocket proteins using the SV40 DNA tumor virus large T antigen. Pertinent to this, activated Cyclin E/Cdk2 allows firing of DNA-replication origins and coordinated centrosome duplication (see Sect. 2.7), and regulates the activity of both chromatin-remodeling factors and the NPAT transcription factor, which is required for the expression of histone genes. A role for other Cdk complexes has also been proposed in regulating Rb function. Overexpression of Cyclin A in cell culture leads to phosphorylation of Rb, and Cyclin A/Cdk2 can phosphorylate Rb *in vitro*. It remains unknown, however, whether phosphorylation of Rb by Cyclin A/Cdk2 is important for Rb inactivation. Finally, recent studies have shown that Cyclin C/Cdk3 can also phosphorylate Rb in the transition from G0 into G1 (see Sect. 2.2.1.4).

Cdk phosphorylation can also regulate E2F/DP transcriptional activity directly. Cyclin E/Cdk2 has been shown to phosphorylate E2F5, resulting in the recruitment of the p300(CBP) family of coactivators to promoters, thereby stimulating transcription and cell cycle progression. Importantly, phosphorylation of the DP subunit of E2F by CyclinA/Cdk2 in S phase leads to loss of E2F heterodimer DNA binding activity and is thought to be one of the major mechanisms by which E2F transcriptional activity is switched off following the G1-S transition.

2.3.4 Target Gene Regulation by E2F and Rb

As indicated earlier, hypophosphorylated Rb, in a complex with E2F and DP, acts to repress gene transcription by recruiting global transcriptional repressors, such as HDACs and chromatin-remodeling complexes to E2F-target promoters (see Fig. 4). Phosphorylation of Rb promotes its release from E2F/DP, and replacement of the HDACs with histone acetylases such as p300(CBP) allowing the activation of E2F-responsive promoters. To date, the regulation of the "classical" E2F target genes *cyclin E1* and *cyclin A1* have been the

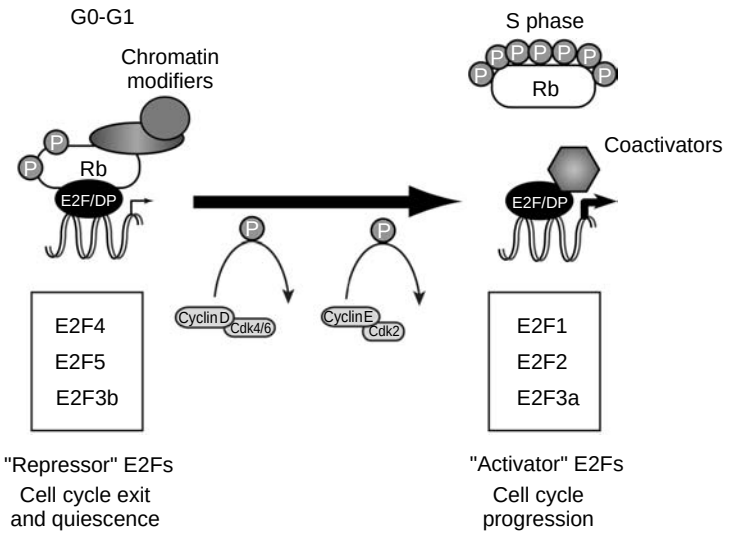

Fig. 4 E2F, DP, and Rb family of proteins in transcriptional regulation of S-phase genes. Repressor E2F proteins act with Rb-family members and chromatin-remodeling factors to repress S-phase genes at the G0-G1 phase. In the G1-S phase transition, Rb-family proteins are phosphorylated and inactivated by Cyclin/Cdks. Activating E2Fs then bind to S-phase gene promoters, along with coactivators, to transactivate transcription. Not shown are E2F6 and E2F7 proteins, which are thought to mediate transcriptional repression in G0-G1 (and hence are also classed as repressor E2Fs) but do not bind to and are not regulated by Rb and family members.

most thoroughly studied. In mammalian cells, *cyclin E1* transcription is activated when Rb is hyperphosphorylated. The human *cyclin E1* promoter contains an E2F binding site, and at this site a complex of E2F4, Rb, Brg1 (SWI or SNF) chromatin-remodeling complex and HDAC has been detected. Phosphorylation of Rb by Cyclin D/Cdk4 disrupts the association with HDAC, leading to the expression of *cyclin E1*. *Cyclin A1* transcription is also repressed by an E2F4/Rb/Brg1/HDAC complex, but this complex is still active upon the loss of HDAC by Cyclin D/Cdk4 mediated phosphorylation of Rb. Since Cyclin E/Cdk2 is known to phosphorylate and inactivate Brg1, only when Cyclin E/Cdk2 levels increase in late G1 phase can repression at the *cyclin A1* promoter be relieved.

Recently, microarray expression profiling of mammalian E2F targets together with ChIP studies, have revealed that in addition to regulating G1-S transition and S-phase genes, E2F also regulates genes involved in the DNA damage checkpoint, DNA repair, chromatin assembly or condensation, chromosome segregation, and the mitotic spindle checkpoint, suggesting that E2F activity is important for both G1-S and G2-M phase progression. Likewise in the Drosophila system (see below), expression profiling of E2F-Rb pathway targets has revealed both G1-S and G2-M targets, but unexpectedly also showed that different specific E2F-Rb complexes repress the expression of target genes required for differentiation that are transcribed in a developmentally regulated,

rather than a cell cycle–regulated manner. This result provokes speculation that in mammalian cells, the E2F-Rb pathway may also not be exclusive to the regulation of cell cycle genes. Consistent with this hypothesis, genome-wide analysis in mammalian cells have identified a number of E2F targets involved in differentiation and development including *HOX* cluster genes, tissue-specific transcription factors and growth factors such as *BMP2* and *TGFα*.

2.3.5 Functional Analysis of Rb, E2F, and DP in Drosophila

The study of Drosophila has been key to understanding the complex interplay between E2F, DP, and Rb function due to its reduced complexity. Drosophila homologs of the retinoblastoma protein (RBF1 and RBF2), two E2F proteins (dE2F1 and dE2F2), and one DP (dDP) protein have been identified (see Fig. 3). Drosophila RBF1 and RBF2 are equally similar to the mammalian Rb, p107, and p130 proteins. Similar to mammalian Rb-family proteins, RBF1 is phosphorylated by Drosophila Cyclin E/Cdk2, which is important for its inactivation. dE2F1, a potent transcriptional activator, is specifically bound by RBF1, which is required to prevent E2F1-dependent transcription of S-phase genes. Conversely, RBF2 binds exclusively to dE2F2, which functions as a transcriptional repressor. RBF1 can also bind to dE2F2; however, dE2F2 can function as a repressor independent of RBF1, since inactivation of RBF1 by expression of Drosophila Cyclin E/Cdk2 does not override repression by dE2F2. Interestingly, microarray analysis suggest that RBF1 and dE2F1 coordinately regulate expression of genes involved in cell cycle progression, while RBF2 and dE2F2 repress a variety of tissue-specific genes involved in differentiation. In the absence of dE2F1, dE2F2 occupies dE2F1-regulated promoters. The defects that occur in *dE2F1* mutants are therefore due to unchecked activity of the dE2F2 repressor since they can be suppressed by mutating *dE2F2*.

Drosophila embryos lacking maternal and zygotic *RBF1* expression show ectopic expression of the dE2F target genes *PCNA* and *RNR2*. Despite the ectopic expression of *PCNA* and *RNR2*, epidermal cells of *RBF1* mutant embryos appear to transiently arrest in G1, although these cells subsequently enter ectopic S phases. Thus, RBF is required for the maintenance of the G1-arrested state. Mutant analysis of Drosophila *dE2F1* and *dDP* showed that dE2F1 is required for the G1-S transcriptional program, as embryos homozygous for null mutations in *dE2F1* or *dDP* do not express dE2F target genes such as *RNR2* and *PCNA*. In some tissues in *dE2F1* or *dDP* mutants, transcription of *cyclin E* was also affected, although Cyclin E/Cdk2 function still remained. Thus, while dE2F1 is required for normal levels of *cyclin E* expression in some tissues, dE2F-independent mechanisms for *cyclin E* transcription also exist. Consistent with this finding, there is evidence that *cyclin E* transcription is regulated by developmental signals in many tissues (see Sect. 2.5).

Surprisingly, analysis of Drosophila lacking all dE2F family members (dE2F1 and dE2F2), suggests that dE2F regulation is not required for cell proliferation (Table 1), raising questions as to the function of dE2F in cell cycle control. Recent studies in Drosophila have addressed this issue by revealing a striking role for dE2F in G1-S and G2-M phase timing. It was shown that perturbations to progression through either G1-S or G2-M were compensated for by adjustment of

the time spent in the alternate phase of the cell cycle. dE2F1/dDP was important for this coordination, via transcriptional upregulation of *cyclin E* in G1-S and of the G2 regulator, *cdc25* (*string*) (see Sect. 2.4.1) in G2-M. Conversely, dE2F protein levels are downregulated in S-phase cells by the SCF complex using the F-box protein, Slimb, via phosphorylation by Cdk1 and Cdk2 (see Sect. 2.4.3). As mammalian E2Fs have also been shown to regulate genes important for G1-S and G2-M progression, analogous cell cycle compensation mechanisms may operate in mammalian tissues.

2.3.6 Functional Analysis of Mammalian Rb, E2F, and DP: Lessons from Mutant Mice

Although biochemical approaches originally identified E2F and Rb proteins in mammalian cells, genetic studies in Drosophila have been key to the formulation of hypotheses, as to the mechanism of action and diversity of E2F and Rb function. Genetic studies in mammalian systems have confirmed current views but also revealed unexpected roles for Rb and E2F families in development and tissue homeostasis (Table 1).

2.3.6.1 **Analysis of Rb-family single and compound mutant mice.** Rb-deficient mice die at E13.5–E15.5 of development with defects in ocular lens development, neurogenesis, and erythropoiesis (see Table 1). Analysis of these tissues revealed extensive ectopic proliferation, abnormal differentiation, and apoptosis consistent with the proposed role of Rb in controlling the *restriction point* and cell cycle exit during terminal differentiation. However, subsequent analysis of mice chimeric for Rb-deficient cells indicated that *Rb*-null cells could contribute to the majority of tissues of these animals with an apparent absence of defects in Rb-deficient tissues. These studies therefore suggested that the neuronal and erythroid phenotype resulting from loss of Rb function may not be cell intrinsic. Consistent with this notion, recent analysis has shown that these defects are in part due to a role for Rb in placental development. As observed for *cyclin E*, Rb as well as E2F3 and DP1 play an important role in controlling the endoreplication of the extraembryonic trophoblast cells (see Sect. 2.7.2.2). Provision of a wild-type placenta to Rb-deficient embryos largely rescued the embryonic defects, indicating that many of the *Rb*-null cell phenotypes are noncell intrinsic. Rb, like Cyclin E, is also required for megakaryocyte maturation, another cell type that undergoes polyploidization during differentiation. This is consistent with studies in Drosophila, where Rbf, dE2F1, and dDP, function together at DNA-replication origins to regulate endoreplication of *chorion* genes in Drosophila follicle cells (see Sect. 2.7.2.2).

Although placental dysfunction in *Rb*-null embryos can account for their lethality, conditional deletion of Rb specifically in erythroid cells or neuronal cells nevertheless indicates that Rb is required in a cell-intrinsic manner for cell cycle exit and differentiation of these cell lineages. For example, acute loss of Rb in erythroid cells leads to enucleation defects, ectopic proliferation, and accumulation of cells in G2-M presumably due to engagement of DNA damage checkpoint. Similarly specific deletion of Rb in the central nervous system (CNS) *in vivo* results in inappropriate S-phase entry. However, in contrast to the *Rb*-null embryo situation, no apoptosis is observed in the CNS. The non-cell intrinsic apoptosis observed in *Rb*-null embryos has therefore been proposed to be a

direct result of the hypoxic environment caused by deficient placental function of *Rb*-null embryos. Using similar tissue-specific deletion of Rb approaches *in vivo*, a requirement for Rb has now been observed for the proper development of a wide variety of additional tissues including, skin, breast, lung, and muscle.

Similar to Rb, loss-of-function mutations in the mouse have indicated a requirement for p107 and p130 in cell cycle exit and differentiation. Indeed, although individual loss of p107 or p130 in the mouse result in viable and mostly normal adults, compound loss of p107 and p130 has revealed an important and overlapping role for these proteins in bone, and more recently, in skin development. p107, p130-doubly null mice die shortly after birth from unknown causes, but showed decreased limb size associated with defective endochondrial bone development and deregulated chondrocyte proliferation. Similarly, loss of p107 and p130 in skin epithelium results in impaired terminal differentiation and overproliferation of interfollicular keratinocytes *in vivo*. The partial redundancy of p107 and p130 in development has also been observed for Rb. Acute or chronic loss of Rb results in upregulation of p107 protein levels and embryos generated with a compound loss of p107 and Rb, or p130 and Rb, show similar phenotype to *Rb*-null embryos, but die two days earlier. Although animals deficient in all three Rb-family members have not yet been generated, the analysis of triple-deficient cells show them to be unresponsive to G1 arrest following DNA damage or growth-factor withdrawal and *in vivo* fail to differentiate and to exit the cell cycle.

In view of their role in controlling cell cycle exit and linking growth signals to the cell cycle, it may be expected that mutations in Rb-family proteins would lead to deregulated proliferation and predispose to tumor formation. Indeed, loss or inactivation of Rb is a frequent event in human tumors. In contrast, p130 is less frequently lost and p107 inactivation has not been reported. $Rb^{+/-}$ mice develop tumors of endocrine origin including pituitary, thyroid, and adrenal medulla tumors. Although loss of p107 or p130 alone has no effect on tumor susceptibility, combined deficiency of Rb and p107 causes retinoblastoma in the mouse indicating that these proteins partially overlap in tumor suppression. Importantly, tumor susceptibility in $Rb^{+/-}$ mice is thought to be in part a result of deregulated E2F activity as loss of E2F1, E2F3, or E2F4 can suppress tumorigenesis in $Rb^{+/-}$ mice. Finally, although in this chapter we have focused on the cell cycle roles of Rb family members related to regulation of E2F function, it must be noted that many studies have described non-E2F mediated roles for Rb, in particular in the control of differentiation, senescence and apoptosis. In the case of differentiation, Rb has been described to bind and regulate a number of tissue-specific transcription factors such as MyoD in muscle cells, CBFA1 in bone lineage cells, and C/EBPα in adipose tissue, and by so doing directly regulate the differentiation programs of these lineages. It is still unclear at this stage to what extent these other functions of Rb-family members contribute to tumour suppression or the developmental phenotypes of Rb-family mutant animals.

2.3.6.2 **Analysis of E2F/DP family single and compound mutant mice.** The presence of multiple mammalian E2F and DP family members has made analysis of their role in development complex due to issues of redundancy (see Table 1). Nevertheless,

analysis of mice singly and compound mutant for E2Fs have provided novel and often surprising insights into the functional requirement of these proteins in development and tumorigenesis.

Mice deficient for E2f1 are viable and fertile and show excessive T-cell production and testicular atrophy. In addition, and paradoxically in view of its proposed role in driving proliferation, *E2f1*$^{-/-}$ mice show increased tumor predisposition in aged mice. E2f1 has been shown to be able to induce apoptosis (cell death) and hence this property may provide an explanation for the increased tumor predisposition of *E2f1*$^{-/-}$ mice. Alternatively, recent experiments have also implicated E2f1 in regulating the DNA damage response pathway therefore predisposing *E2f1*$^{-/-}$ mice to acquire somatic mutations. *E2f2*$^{-/-}$ mice are viable and fertile; however, they develop late onset autoimmune disease that is thought to result from decreased activation threshold of *E2f2*$^{-/-}$ T cells. In this context, and contrary to its proposed role in driving proliferation, E2f2 has been proposed to act as a transcriptional repressor of genes required for S phase, in particular, E2f1. In contrast, E2f3 is required for normal development with the majority of E2f3-deficient mice dying *in utero* and the adult survivors succumbing to heart failure by 12 months of age. Compound E2f mutant mice have also revealed a level of redundancy between E2f1, E2f2, and E2f3. *E2f1*$^{-/-}$ *E2f2*$^{-/-}$ mice display defects in multiple hematopoietic lineages including B, T, and erythroid cells resulting in megaloblastic anemia. These defects appear to be due to impaired S-phase progression and decreased expression of E2F targets such as *cyclin A2*. In addition, compound loss of E2f1 and E2f2 leads to increased incidence of T-cell tumors as well as nonimmune related diabetes. Finally, compound loss of E2f2 or E2f1 and E2f3 exacerbates the respective developmental defects of each E2F mutant mouse.

In contrast to the "activating" E2F family members, loss of function of the "repressor" E2f3, E2f4, E2f5, or E2f6 appears to give rise to differentiation related phenotypes. *E2f4*$^{-/-}$ mice show growth retardation and die neonatally due to upper respiratory tract infections. They display fetal megaloblastic anemia and defects in gut development. Of note is the fact that recent analysis of the erythroid defect of *E2f4*-deficient mice has indicated an unexpected role for E2f4 in promoting (rather than inhibiting) transcriptional activation and cell proliferation of the erythroid lineage. Loss of E2f5 results in perinatal death due to hydrocephaly in what is thought to be a non-cell cycle related E2f5 function. E2f5 is highly expressed in the choroid plexus and has been proposed to regulate cerebral spinal fluid (CSF) protein secretion. Finally, *E2f6*$^{-/-}$ mice are viable; however, they show homeotic transformation of their axial skeleton reminiscent of mice with mutations in the Polycomb group of chromatin-remodeling proteins. This is consistent with the known ability of E2F6 to recruit Polycomb group proteins to direct repression of target genes and suggests a function for E2F6 in the direct regulation of developmental transcriptional programs.

In summary, the *in vivo* analysis of E2F function in mouse models, suggest that the convenient classification of E2Fs into activating and repressive E2Fs may be oversimplistic. E2F function will differ depending on cellular context and the expression of other E2Fs and Rb-family members. Although in many cases, E2Fs and Rb-family members appear to regulate cell cycle progression, in particular, the G1-S transition, it is also increasingly

clear that they can also regulate functions as disparate as apoptosis, DNA damage repair, differentiation, and developmental cues. It will therefore be important to conduct biochemical, functional, and gene expression analysis in specific tissues and cell lineages to obtain a thorough perspective of the function and crosstalk of these proteins families in the regulation of cell cycle progression and differentiation.

2.4 Regulation of G1-S Cyclin/Cdk Activity

In addition to Cdk activation by the binding of a cyclin, cyclin/Cdk activity is also regulated by phosphorylation, the binding of inhibitory proteins and proteolysis of the cyclin subunit.

2.4.1 Regulation of Cdk Activity by Phosphorylation

Two modes of phosphorylation have been documented in the regulation of cyclin/Cdk activity (see Fig. 5). The first of these modes of phosphorylation is stimulatory while the second is inhibitory. Stimulatory phosphorylation of the Cdk occurs on a critical threonine residue of the Cdk (Thr160/161 for Cdk2) in a region termed the *T-loop*, by CAK. CAK itself is composed of a cyclin/Cdk complex, Cyclin H/Cdk7 and a ring-finger protein, Mat1. CAK phosphorylation of Cdks is important to increase the binding affinity of the cyclin to Cdk and to increase the kinase activity. As described earlier (Sect. 2.2.1.4), analysis of the Cdk7 complex may be complicated as a result of its expected role in RNA polymerase II transcription. However, analysis of *Mat*1 mutant mice revealed defects in cell proliferation of embryonic cells and impaired endoreplication of extraembryonic trophoblast cells, but mutant cells did not show any apparent defects in general transcription. Thus, *in vivo* it appears that the primary role for mammalian Cyclin H/Cdk7/Mat1 complex is as a CAK, rather than in general transcriptional regulation. In Drosophila, the Cyclin H/Cdk7 complex is conserved and has CAK activity toward Drosophila Cdk1 and human Cdk2 *in vitro*. Drosophila temperature-sensitive *cdk7* mutants display defects in mitosis; however, the ovarian endoreplication cycles (see Sect. 2.7.2.2) were not affected. Thus, in Drosophila, the importance of Cdk7 in G1-S regulation is unclear. In

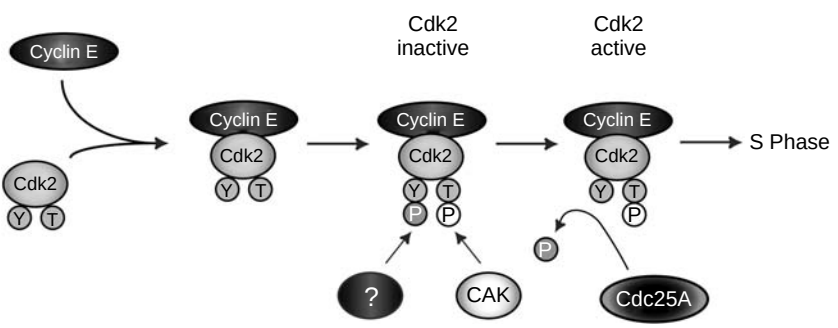

Fig. 5 Regulation of Cyclin/Cdk2 activity by phosphorylation. Mammalian Cyclin E/Cdk2 is regulated by two modes of phosphorylation. Phosphorylation at Thr160 on Cdk2 by CAK is required for activation, while phosphorylation at Tyr15 on Cdk2, by an unknown kinase, fixes Cyclin E/Cdk2 in an inactive state. It is only after the Cdc25A phosphatase is upregulated and dephosphorylates the Tyr15 of Cdk2 that Cyclin E/Cdk2 is activated.

contrast to the mouse studies, analysis of Drosophila *cdk7* mutants revealed defects in RNA polymerase II transcription. Taken together, these studies in Drosophila, suggest that Cyclin H/Cdk7/Mat1 complex may have roles as a CAK and in transcription in the fly.

The second mode of control of Cdk activity is through inhibitory phosphorylation. Inhibitory phosphorylation has been well documented for the mitotic Cdk, Cdk1, by the Wee1 kinase. Wee1 phosphorylates tyrosine and threonine residues (Tyr15, Thr14) on the N-terminus of Cdk1, which prevents cyclin/Cdk kinase activity. Removal of these inhibitory phosphates on Cdk1 by the Cdc25 phosphatase can be a rate-limiting step for entry into mitosis. In mammalian cells, the G1-S Cdks, Cdk2, and Cdk4, are also phosphorylated on the N-terminal tyrosine (Tyr15 and 17, respectively), which has been shown to be inhibitory to their activities. There are three Cdc25 phosphatases in mammalian cells, but Cdc25A is thought to be the major Cdc25 required for G1-S progression. Cdc25A is specifically expressed in late G1-S cells and is able to accelerate G1-S progression when ectopically expressed by dephosphorylating Cyclin E/Cdk2. Drosophila also has homologs of Wee1 and Cdc25, and studies of the Drosophila homolog of Cdc25, String, have shown that it is essential and rate limiting for the G2-M phase transition during Drosophila development. However, in Drosophila, this regulation does not extend to the G1-S Cdk, Cdk2, since mutations of the Cdc25/Wee1 regulatory site in Drosophila Cdk2 do not impair Cdk2 function in Drosophila.

2.4.2 Regulation of G1-S Cyclin/Cdk Activity by Cdk Inhibitors

In mammalian cells, the activity of cyclin-Cdk complexes is negatively regulated by direct interactions with proteins called Cdk inhibitors (Ckis). These proteins fall into two families on the basis of their structural and functional properties: INK4 family ($p15^{INK4B}$, $p16^{INK4A}$, $p18^{INK4C}$, $p19^{INK4D}$) and the CIP(KIP) family ($p21^{CIP1}$, $p27^{KIP1}$, $p57^{KIP2}$). The INK4 (inhibitor of Cdk4) class of proteins are composed of multiple ankyrin repeats, and bind Cdk4(6) preventing it from associating with Cyclin D, thereby preventing this complex from promoting G1-S phase progression (Fig. 6). The CIP(KIP) family of Cdk inhibitors act more broadly to inhibit the kinase activity of Cyclin E/Cdk2 and Cyclin A/Cdk2. Consistent with CIP(KIP)-family proteins inhibiting G1 Cyclin/Cdk activity, overexpression of any member of this family results in G1 phase arrest of cultured cells. Paradoxically, CIP(KIP) family play a dual role, since they can also act as positive regulators of Cyclin D/Cdk4 by stabilizing the formation of the Cyclin D/Cdk4 complex.

In Drosophila, there are no obvious INK4-family members and just a single CIP(KIP)-family protein, Dacapo (Dap). Dap is expressed at low levels in proliferating cells and is transiently expressed at high levels as cells exit from the mitotic cycle, consistent with a role in cell cycle arrest prior to terminal differentiation. Like mammalian CIP(KIP)-family proteins, Dap interacts with Cyclin E/Cdk2 *in vivo*, and is able to inhibit Cdk2-associated kinase activity *in vitro*. It is not currently known whether Dap also interacts with Cyclin D complexes. However, Dap does not inhibit Cyclin A-associated kinase activity, consistent with the data that Drosophila Cyclin A associates exclusively with Cdk1, and not Cdk2. Unlike mammalian inhibitors, which are not essential for viability, mutations in *dap* result in lethality. Epidermal cells of embryos

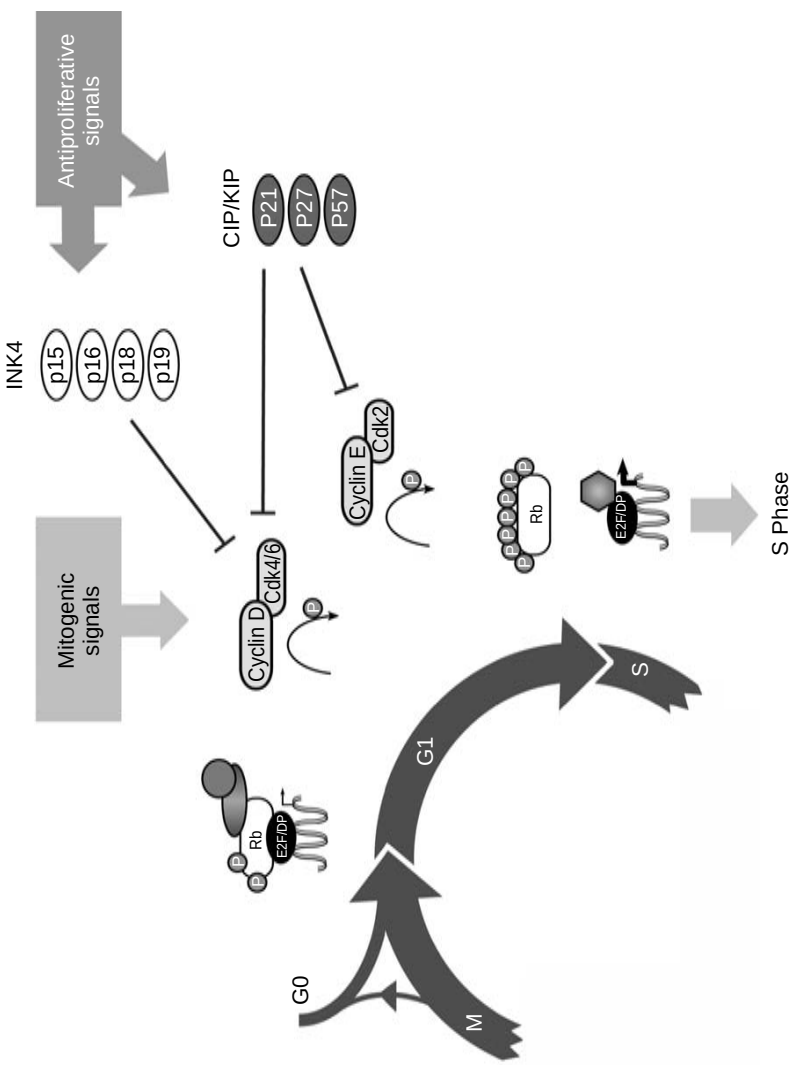

Fig. 6 Regulation of Cyclin/Cdk activity by Cdk inhibitors. Mitogens (such as EGF) lead to upregulation of Cyclin D expression, which acts to phosphorylate and inactivate Rb allowing G1-S phase progression. Antiproliferative signals (such as TGFβ) induce the expression of Cdk inhibitors (Ckis) of the p16^{INK4A} and p21^{CIP1} (p27^{KIP1}) families. p16^{INK4A} family members act to specifically inhibit Cyclin D/Cdk4(6), while p21^{CIP1} (p27^{KIP1}) proteins act more generally to block Cyclin D/Cdk4(6) and Cyclin E/Cdk2.

homozygous for mutations in *dap* undergo an additional cell cycle during embryogenesis, indicating that Dap is required for the establishment of the G1 arrest normally seen in these cells. In addition, ectopic expression of *dap* is sufficient to induce G1 arrest, showing that Dap is clearly required for timely exit from the cell cycle during Drosophila development.

Analyses of mice that are deficient for CIP(KIP)–family proteins indicate an important role for these Ckis in regulation of G1 progression and cell cycle exit *in vivo* (see Table 1). Ablation of $p27^{KIP1}$ in the mouse results in increased body size, overgrowth of organs, retinal dysplasia, female sterility, and pituitary and parathyroid tumors, consistent with a role in controlling proliferation during organ development. In contrast, mice deficient in $p21^{CIP1}$ are viable and normal and do not show increased tumor predisposition. Further analysis, however, has revealed a role for $p21^{CIP1}$ in hematopoietic stem cell maintenance, with loss of $p21^{CIP1}$ resulting in increased cycling of hematopoietic stem cells leading to exhaustion of the stem cell pool and eventually to hematopoietic failure. In mammalian cell culture, $p21^{CIP1}$ has been implicated in muscle cell differentiation. The transcription factor MyoD activates the expression of $p21^{CIP1}$, in addition to inducing the transcription of muscle-specific genes. Although the induction of $p21^{CIP1}$ was presumed to induce cell cycle arrest required for muscle differentiation, $p21^{CIP1}$ knockout mice do not show any muscle differentiation defects. Thus, $p21^{CIP1}$ does not play a critical role in muscle differentiation *in vivo*.

While most Ckis are not required for viability, loss of $p57^{KIP2}$ leads to neonatal death and severe developmental defects with phenotypic characteristics similar to the human disease, Beckwith–Wiedemann syndrome. These include cleft palate, defects in gut development, proliferative defects in the lens, and impaired cell cycle exit, during chondrocyte differentiation resulting in defective bone formation. These endochondrial bone development defects are very similar to those observed in mice compound mutant for the Rb-family members, p107 and p130. Likewise, compound loss of $p27^{KIP1}$ and $p57^{KIP2}$ gives rise to lens and trophoblast defects remarkably similar to that seen in Rb-deficient mice and compound loss of $p21^{CIP1}$ and $p57^{KIP2}$ result in deregulated proliferation and defective differentiation of skeletal muscle, again similar to that observed in Rb-deficient skeletal muscle. As Rb-family proteins are one of the primary targets of G1-S Cdks (see Sect. 2.3.3), the phenotypic similarities highlight a role for the CIP(KIP) family of Ckis in the regulation of Rb, p107, and p130 activity *in vivo*.

In contrast to the CIP(KIP) family, mutations in INK4 family members, primarily $p16^{INK4A}$ and $p15^{INK4B}$, have been consistently reported both in mouse and human tumors. The generation and analysis of single and doubly deficient mice for individual INK4 family members has increased our understanding of how these genes may be important in development and tumorigenesis (Table 1). Loss of $p15^{INK4B}$ results in lymphoproliferative disorder as well as small increased susceptibility to tumor development, primarily angiosarcomas. Because the INK4A locus also encodes the overlapping gene for the $p19^{ARF}$ tumor suppressor (see Sect. 2.6), it is only recently that the specific role of $p16^{INK4A}$ has been able to be tested *in vivo* by the generation of a mouse mutant for $p16^{INK4A}$, but retaining functional $p19^{ARF}$ protein. Only $p16^{INK4A}$ deficient mice are born viable

and normal; however, they develop a range of tumors including soft tissue sarcomas, lymphomas, and melanomas. In addition, p16^{INK4A} deficient mice show thymic hyperplasia and T-cell hyperproliferation following stimuli consistent with a role in regulating cell cycle entry. Mice deficient for p*18*^{INK4C} exhibit phenotypes similar to p27^{KIP1}-deficient mice including enlarged organs, tissue hyperproliferation, and a variety of tumors, including pituitary adenomas. Furthermore, *p18*$^{INK4C-/-}$ mice exhibit lymphoproliferative disease as well as epithelial proliferation in the kidney leading to the formation of cysts. Cyst formation is exacerbated in mice mutant for both *p18*$^{INK4C-/-}$ and *p15*$^{INK4B-/-}$, indicating some level of redundancy for these Ckis in epithelial tissues. Finally, loss of p19^{INK4D} gives rise to viable mice; however, they develop age-related testicular atrophy and progressive hearing loss. The hearing loss is due to aberrant cell cycle reentry of sensory hair cells in the organ of Corti consistent with a role for p19^{INK4D} in maintaining the postmitotic state. Redundant roles for p19^{INK4D} and p18^{INK4C} have been established in spermatogenesis, as compound loss of p19^{INK4D} and p18^{INK4C} enhances the *p19*$^{INK4D-/-}$ mice testicular defects, now resulting in sterility. The loss of fertility is primarily due to delayed exit of spermatogonia from the mitotic cycle, leading to apoptosis and decreased spermatocyte production.

The analysis of mouse models for INK4 and CIP(KIP) CDK inhibitors have revealed a large level of redundancy built into the mechanism required to maintain cells in quiescence as well as cell cycle progression *in vivo*. However, all experiments to date have demonstrated a requirement for these Cdk inhibitors for normal development, tissue homeostasis, and tumor suppression.

2.4.3 Regulation of G1-S by Ubiquitin-mediated Proteolysis

Protein degradation plays an important role in the G1-S phase transition. An ubiquitin ligase (E3) known as the SCF (Skp1/Cullin/F-box protein) complex, targets both Cdk inhibitors and cyclins to ubiquitin-mediated degradation (see Fig. 7). The catalytic core components of the SCF include Skp1, Cul1, and Rbx1, which assemble with a variety of substrate specific binding proteins known as F-box proteins to transfer ubiquitin (a 76-residue protein) onto substrates. Phosphorylation of targets, leads to recognition by the SCF complex and subsequent action of a ubiquitin-conjugating enzyme (E2), followed by a ubiquitin-activating enzyme (E1) to promote the addition of ubiquitin-residues to the substrate, thereby tagging it for degradation by the 26S Proteasome.

2.4.3.1 Regulation of Cki proteolysis by the SCF complex. The Cdk inhibitor p27^{KIP1} is targeted for ubiquitin-mediated proteolysis by the SCF, after it is phosphorylated by Cyclin E/Cdk2. The F-box protein, Skp2, is important for recognizing phosphorylated p27^{KIP1}. Skp2 is synthesized specifically at the G1-S phase boundary, as is Cyclin E. This enables active Cyclin E/Cdk2 complexes (that are free of p27^{KIP1}) to accumulate, to phosphorylate p27^{KIP1}, and then Skp2-SCF mediated degradation of p27^{KIP1}, allowing a burst of Cyclin E/Cdk2 activity in cells committed to entering S phase. Skp2 also requires the small Cdk-binding protein, Cks1 (see Sect. 2.2.2), to bind to p27^{KIP1} as well as some other substrates, such as the Rb-family member p130. The Cdk inhibitor p21^{CIP1} is also ubiquitinated by Skp2-SCF upon phosphorylation at the G1-S phase transition; however, p21^{CIP1}

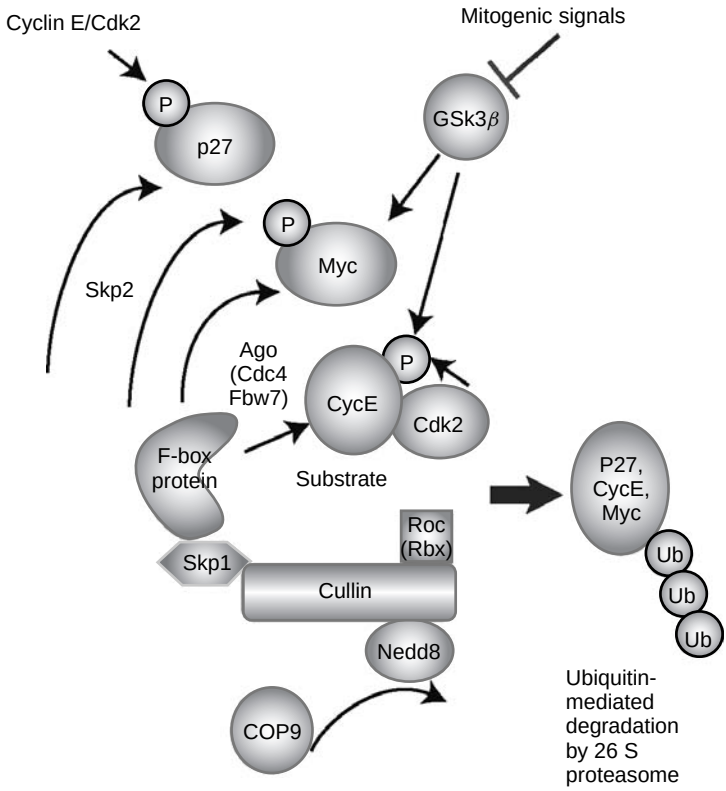

Fig. 7 Regulation of the G1-S phase by the SCF. The ubiquitin–ligase complex, SCF (composed of Skp1, Cullin, Rbx1, and an F-box protein) recognizes phosphorylated targets and ubiquitinates them, tagging them for ubiquitin-mediated degradation by the 26S proteosome. The Cullin in the SCF is modified by a ubiquitin-like protein, Nedd8, and cyclic removal of Nedd8 by the COP9 signalosome is required for SCF activity. The F-box, Skp2 (together with Cks1, not shown), targets phosphorylated $p27^{KIP1}$, which is phosphorylated by Cyclin E/Cdk2. SCF-Skp2 is also important in basal ubiquitin-mediated degradation of Myc. SCF-Fwb7(Ago) targets phosphorylated Cyclin E and Myc for ubiquitin-mediated degradation. Cyclin E is phosphorylated by the associated Cdk2 kinase. GSK3β also phosphorylates Cyclin E and Myc, tagging them for recognition by SCF-Fwb7(Ago). Mitogen signaling leads to the inactivation of GSK3β enabling progression from G1 into S phase.

can also be degraded by the proteasome via an ubiquitin-independent mechanism.

By contrast, the INK4 family of inhibitors are degraded throughout the cell cycle via ubiquitin-mediated proteolysis and their accumulation upon negative growth conditions is mostly achieved transcriptionally (see Sect. 2.5).

2.4.3.2 **Regulation of G1-S phase cyclins proteolysis by the SCF complex.** In human cells, Cyclin E is degraded by

the ubiquitin/proteasome pathway, which requires phosphorylation at Thr380 in an Len-Thr-Pro-Pro (LTPP) motif in Cyclin E by the associated Cdk2 protein kinase. The LTPP destruction sequence is conserved in Xenopus, mouse, and Drosophila Cyclin E, although the functional significance of this motif has not been demonstrated in these systems. Recently, GSK3β has also been shown to facilitate the efficient phosphorylation of Cyclin E at multiple sites to trigger ubiquitin-mediated degradation. GSK3β is a kinase responsive to the PI3K-Akt pathway (see Sect. 2.5.1.4) and therefore, specific phosphorylation of Cyclin E in response to mitogenic stimuli provides a means for developmental regulation of Cyclin E turnover.

Degradation of Cyclin E is dependent on the SCF ubiquitin ligase, in association with another F-box protein, Ago(Cdc4,Fbw7) (see below). The importance of the SCF in Cyclin E degradation is exemplified by the high levels of Cyclin E present in *Cul*1$^{-/-}$ mutant mice (where the SCF does not form). In addition to this mechanism, free Cyclin E, in a phosphorylation-independent manner is also degraded by a non-SCF mechanism, involving a Cul1 related protein, Cul3.

Degradation of Cdk4-associated Cyclin D1 also depends on phosphorylation at a threonine residue at the C-terminus (Thr268) by GSK3β, although it is not known whether this is SCF dependent. However, free Cyclin D1 is degraded via an ubiquitin-mediated mechanism in a phosphorylation-independent manner. Conversely, Cyclin A, which also has roles in the G2-M phase transition, is targeted for degradation by a different ubiquitin-dependent ligase complex, the APC (anaphase promoting complex), which is active in mitosis.

2.4.3.3 **Cyclin E and Ago/Fbw7.** Genetic screens in Drosophila designed to isolate genes that confer proliferative advantage identified *ago* (*hCDC4* in humans and *Fbw7* in the mouse) as the F-box protein that targets Cyclin E for degradation. Loss of *ago* results in excessive cell proliferation and delayed exit from the cell cycle. *ago* mutant cells exhibit elevated levels of Cyclin E protein without affecting its transcription. Ago protein binds to and ubiquitinates phosphorylated Cyclin E and thereby targets it for ubiquitin-mediated degradation.

Fbw7 (hCDC4, Ago)-deficient mice have now been generated that provide parallels with Drosophila *ago* mutants. Although these studies confirmed the role of Fbw7 in Cyclin E regulation, they have also demonstrated that Fbw7 targets a number of other important developmental proteins for degradation. Fbw7-null animals die at E10.5 gestation displaying defects in somitogenesis, heart chamber formation, hematopoiesis, and vascularization in both yolk sac and the embryo proper. These were associated with increased expression of Notch intracellular domains (ICD) and overexpression of Notch targets in the embryo, consistent with what has been observed in Drosophila (see Sect. 2.5.1.7). Importantly, the vascular defects phenocopy that of mice with loss of Notch1 and Notch4, or overexpression of Notch4, further suggesting a role for Fbw7 in control of Notch pathway in mouse development. As would be expected from previous studies, Cyclin E was found to be stabilized in both placenta and embryo and its high expression in *Fbw7*$^{-/-}$ trophoblast giant cells is thought to result in the observed increased DNA synthesis in these trophoblast cells. Of note, *Cul1*$^{-/-}$ mice also show Cyclin E stabilization and increased DNA synthesis

of trophoblast giant cells. These studies suggest the SCF^{Fbw7} complex plays an important role in regulating endocycles in extraembryonic tissue through regulation of Cyclin E levels.

2.4.3.4 **Myc, another G1-S targets of Ago/Fbw7.** Studies in Drosophila have also identified Myc (dMyc, a homolog of c-Myc, a transcription factor involved in regulating of cell growth and proliferation, see Sect. 2.5.1.3) as a new regulatory target for Ago/Fbw7. Overgrowth of *ago* clones in Drosophila suggested that targets other than Cyclin E exist to regulate cell growth. Mass spectrometric analysis and yeast-2-hybrid studies identified d-Myc as a binding partner for Ago. dMyc protein is elevated in *ago* mutant clones. Ago physically binds to dMyc, and halving the dosage of *ago* can rescue the body size and fertility defects of a hypomorphic *dmyc* mutant, although these effects were through increases in cell numbers rather than cell size. Parallel experiments in mammalian cells, revealed that Fbw7 was also required for c-Myc degradation. Consistent with the effect of Fbw7 on c-Myc stability, knockdown of Fbw7 increases cell size, as expected when c-Myc levels are elevated. Similar to Cyclin E, the c-Myc phosphorylation site that conforms to the Cdc4 phosphobinding motif (Thr28), is also phosphorylated by GSK3β and this phosphorylation is required for c-Myc ubiquitination by Fbw7-SCF and subsequent degradation. Interestingly, Thr28 is the most common site of c-Myc mutation in lymphoma. A model has been proposed where Skp2, the other F-box protein known to target c-Myc, regulates basal turnover of c-Myc, while Fbw7 regulates signal-induced degradation of c-Myc. Again, differential phosphorylation provides another strategy for the tight developmental regulation of c-Myc levels. The original studies describing the human homolog of Ago(Fbw7), revealed its mutation in a number of tumors. Since Ago(Fbw7) regulates multiple pathways, such as Myc, Cyclin E, Notch (Sect. 2.5.1.7), and also Jun, its mutation is likely to result in severe disruption to signaling and cell proliferation control and therefore is consistent with its importance in human cancer. Contrary to this notion, however, Fbw7 heterozygous mice do not show increased tumor predisposition. Sensitized backgrounds and/or use of conditional alleles of mouse Fbw7 may clarify its role in tumorigenesis.

2.4.3.5 **Cop9 signalosome in Cyclin E regulation.** Recent studies in the mouse and Drosophila have identified an important role for the COP9 signalosome (CSN) in the regulation of Cyclin E degradation. CSN consists of eight highly conserved proteins, and is important for the modification and activation of Cul1. Cul1 is modified by the conjugation of an ubiquitin-like protein, Nedd8, onto a specific Lysine residue of Cul1 by UBC12. Removal of Nedd8 (deneddylation), by CSN decreases the association of the E2 with Cul1. However, genetic studies in the mouse have shown that CSN is a positive regulator of Cullin-ubiquitin ligases (SCFs), and thus the dynamic cycling of Nedd8 from Cul1 appears to be important for SCF activity. In the mouse, mutation of the CSN components, *Csn2* or *Csn3,* result in destabilization of the CSN and early embryonic lethality. $Csn2^{-/-}$ blastocysts show severe proliferation defects. Consistent with the proposed role for the CSN in Cul1 regulation, $Csn2^{-/-}$ proliferation defects are associated with abnormal Cul1 deneddylation, resulting in stabilization of Cyclin E,

similar to the phenotype observed in *Cul1* mutant mouse tissue. In parallel studies in Drosophila, Csn4 and Csn5 were shown to be required for oogenesis. In particular, *csn5* and *csn4* mutants showed defects in deneddylation of Cul1 and accumulation of Cyclin E. As overexpression of Cyclin E in Drosophila phenocopies CSN mutants, and *csn5* and *cyclin E* mutations suppress each other, it is likely that Cyclin E is a major target for the CSN in early oogenesis.

Thus, in summary, we have discussed three tiers of regulation involved in the accumulation and activity of the G-S phase Cyclin/Cdk complexes; (1) phosphorylation of the Cdk, (2) binding of Cdk inhibitors to regulate kinase activity, and (3) proteolysis to regulate availability of the Cyclin or levels of Cdk inhibitors (see Fig. 8). In addition, Cyclins and Cdk inhibitors are regulated transcriptionally and translationally, and all of these regulatory tiers are impacted upon by inputs from extracellular signals as discussed below.

2.5 Regulation of Cyclin and Cdk Inhibitors by Extracellular Signals

2.5.1 Growth-factor Signaling Pathways in G1-S Regulation

Mitogenic signaling pathways trigger cellular growth (increase in cell mass, by increasing protein synthesis) as well as cell cycle progression. The main mitogenic pathways are the insulin receptor (Inr)-PI3 kinase (PI3K) pathway and the EGFR-Ras pathway, which leads to the upregulation of the transcription factor, Myc. However, signaling pathways that have major roles in tissue patterning can also influence cell growth and proliferation. These include the Hh, Wnt(Wg), TGFβ(Dpp), and Notch pathways. These pathways can induce cell growth and cell proliferation, however, in some contexts the TGFβ(Dpp) and Wg pathways can inhibit cell proliferation. Cells are normally dependent upon mitogenic growth factors for survival and proliferation; however, one of the hallmarks of cancer is growth-factor independence, due to mutations resulting in the activation of signaling pathways.

In the absence of mitogenic stimuli, the levels of free $p27^{KIP1}$ levels are high and thus inhibit the activity of Cyclin E/Cdk2 (see Fig. 9). On the induction of mitogenic pathways, such as the EGFR-Ras and Myc pathways, Cyclin D is transcriptionally activated and as cells accumulate Cyclin D, $p27^{KIP1}$ is sequestered in Cyclin D/Cdk4 complexes, thereby facilitating Cyclin E/Cdk2 activation. Via this mechanism known as Cki exchange, Cyclin E/Cdk2 is activated leading to the phosphorylation of $p27^{KIP1}$ and its ubiquitin-mediated degradation via the SCF (Skp1/Cullin/F-box protein) ubiquitin–ligase complex (see Sect. 2.4.3). In this manner, there is a rapid burst of Cyclin E/Cdk2 activity and cells proceed into S phase. By contrast, antiproliferative (antimitogenic) signals push the equilibrium into the reverse direction, leading to inhibition of Cyclin E/Cdk2 and Cyclin D/Cdk4 activity and arrest in G1 (see Fig. 9).

Specific aspects of these major mitogenic and antimitogenic signaling pathways in regulation of the G1-S-phase regulators will now be discussed.

2.5.1.1 **TGFβ pathway.** The transforming growth factor β (TGFβ; Dpp in Drosophila) signals through the heterodimeric transmembrane receptors, Transforming Growth Factor Receptor (TGFR) type I (Thickvein or Saxophone in Drosophila) and type II (Punt in Drosophila) (see Fig. 10). This results in the phosphorylation of receptor-associated

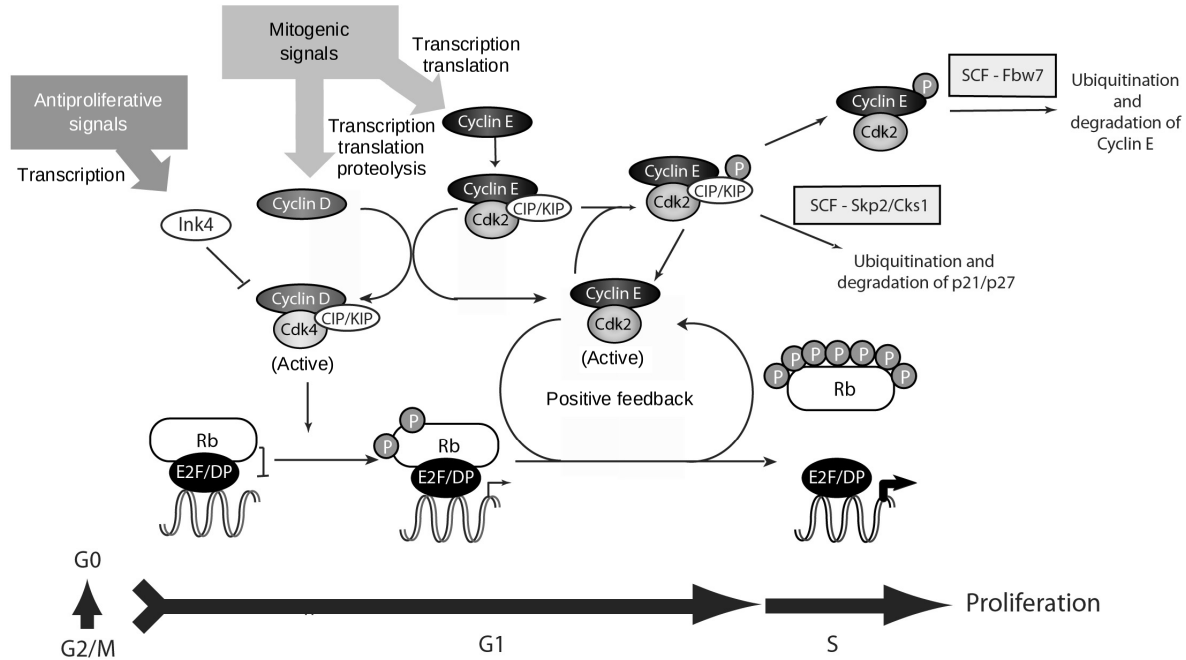

Fig. 8 Summary of G1-S phase progression. Activation of Cyclin D/Cdk4 and Cyclin E/Cdk2 leads to the phosphorylation of Rb and the alleviation of Rb inhibition. Free E2F/DP activates the transcription of genes required for progression into S phase. One of the targets of E2F/DP is *cyclin E*, creating a positive feedback loop, allowing the accumulation of active Cyclin E/Cdk2 kinase at the G1 to S-phase transition. The activity of the Cyclin/Cdk complexes is regulated by the $p16^{INK4A}$ and $p21^{CIP1}$ ($p27^{KIP1}$) family of proteins. $p16^{INK4A}$ binds Cdk4 and prevents binding of Cyclin D, while both $p21^{CIP1}$ and $p27^{KIP1}$ can inhibit Cyclin E/Cdk2 kinase activity, but can act as positive regulators of Cyclin D/Cdk4 assembly. $p27^{KIP1}$ is phosphorylated by Cyclin E/Cdk2, making it a target for the SCF-Skp2/Cks1 ubiquitin–ligase complex and subsequent ubiquitin-dependent degradation by the 26S proteosome complex. Later in S phase, Cyclin E is phosphorylated by the associated Cdk2 protein kinase and by GSK3β, targeting it for ubiquitination by the SCF-Fbw7 ubiquitin ligase followed by ubiquitin-dependent degradation.

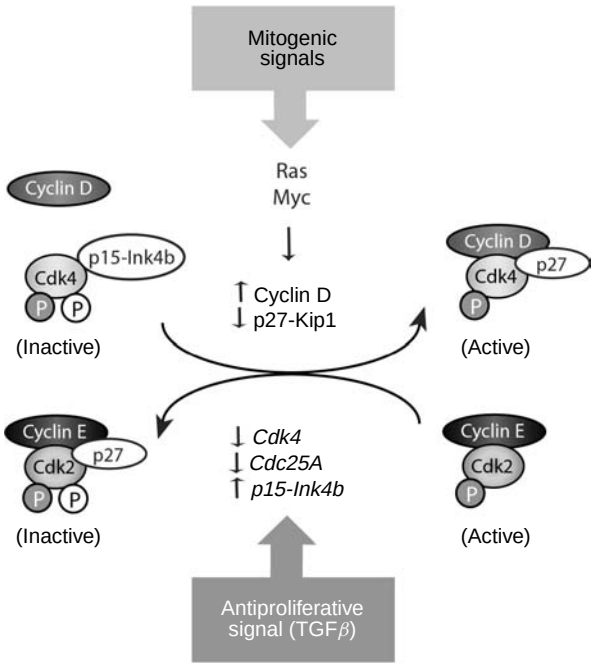

Fig. 9 Cki exchange. During G1, Cyclin/Cdks are in equilibrium between active and inactive states. When the Cdk inhibitor $p27^{KIP1}$ is associated with Cyclin D/Cdk4, it acts to stabilize rather than inhibit this complex. As a consequence, Cyclin E/Cdk2 is freed from inhibition by $p27^{KIP1}$, allowing kinase activity. Mitogenic signaling via Ras and Myc, allows upregulation of *cyclin D* and repression of $p27^{KIP1}$ transcription, further promoting the activation of Cyclin E/Cdk2 allowing progression into S phase. Antiproliferative signals (such as TGFβ) lead to the transcriptional repression of *cdk4*, and *cdc25A*. Since Cdc25A is required to dephosphorylate Cdk2 and Cdk4 (on Tyr15 and 17, respectively) for kinase activation, downregulation of *cdc25A* pushes the equilibrium toward the inactive state. TGFβ also induces expression of $p15^{INK4B}$, which binds specifically to Cdk4 preventing Cyclin D and $p27^{KIP1}$ from binding and inactivating the kinase. Free $p27^{KIP1}$ can then bind to Cyclin E/Cdk2 and inhibit its activity.

Smad transcription factors R-Smads (Smad1, 2, 3, 5, and 8; Mad in Drosophila) enabling their interaction with the Co-Smad (Smad4; Medea in Drosophila). These complexes then translocate to the nucleus and transactivate target genes. Another class of Smads, inhibitory Smads (Smad6 and 7; Dad in Drosophila) antagonize the effect of R-Smads and block TGFβ signaling.

In mammalian epithelial cells, TGFβ negatively regulates cell growth and proliferation. TGFβ induces G1 arrest in epithelial cells via inhibition of Cyclin D/Cdk4(6) and Cyclin E/Cdk2 activity. In the absence of TGFβ signaling, as cells

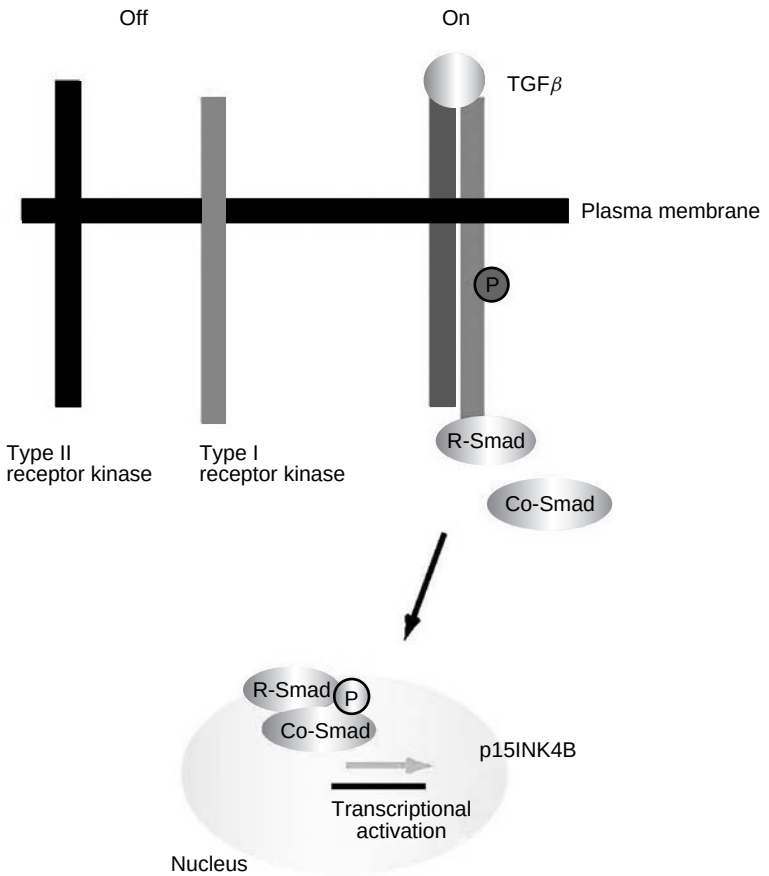

Fig. 10 TGFβ signaling pathway. TGFβ signals though the type I and type II heterodimeric receptor serine–threonine protein kinases. This leads to the activation of the receptor-associated Smad transcription factor (R-Smad) that then binds to another class of Smads (Co-Smad) and enters into the nucleus where it promotes transcription of target genes. The p15^{INK4B} gene is a well characterized downstream target of Smad and TGFβ signaling (see text for details).

progress through the G1-S transition, D-type Cyclins accumulate and Cyclin D/Cdk4(6) sequesters p21^{CIP1}(p27^{KIP1}) away from Cyclin E/Cdk2, thereby allowing its activation. TGFβ signaling induces expression of p15^{INK4B}, a member of the Cdk4(6) class of inhibitors. p15^{INK4B} binds to Cdk4(6), leading to the dissociation of Cyclin D and the release of p27^{KIP1} from this complex. The "free" p27^{KIP1} then acts, via Cki exchange, to bind to and inhibit Cyclin E/Cdk2 activity (see Fig. 9). TGFβ therefore leads to G1 arrest by inhibiting both Cyclin D and Cyclin E-associated kinase activity. However, this is not the only mechanism by which TGFβ induces cells to arrest in G1, as cells derived from p27^{KIP1} knockout mice are

still sensitive to TGFβ. TGFβ also results in downregulation of *cdc25A* transcription, thereby reducing levels of active Cyclin E/Cdk2 and Cyclin D/Cdk4 complexes. In addition, TGFβ leads to downregulation of *cdk4* transcription, which serves to free more p27^{KIP1} from the Cyclin D complexes, allowing exchange to Cyclin E/Cdk2. TGFβ also downregulates *c-myc* expression, which is not only a regulator of cell growth but also of *cyclin D1* transcription, thereby serving to further promote exchange of p27^{KIP1} to Cyclin E/Cdk2. Finally, Smads have also been shown to cooperate with a member of the Forkhead class of transcription factors, FoxO, to upregulate p21^{CIP1} transcription. Thus, TGFβ acts via p27^{KIP1} dependent and independent mechanisms to mediated G1 arrest.

The Drosophila TGFβ superfamily member Decapentaplegic (Dpp) acts to stimulate cell growth and proliferation in some contexts, but to inhibit cell proliferation in other contexts. In early development, and in third instar larval wing imaginal discs, loss of *dpp* expression results in reduced tissue size. However, in the third instar larval eye imaginal disc, *dpp* mutant clones exhibit ectopic cell proliferation, while increasing Dpp signaling leads to G1 arrest. Recent studies have shown that the mechanism for this cell cycle inhibition by Dpp is likely to involve both the Drosophila Retinoblastoma-E2F pathway and the Drosophila p21^{CIP1}(p27^{KIP1}) homolog, Dap.

2.5.1.2 **EGFR–Ras pathway.** The small GTPase, Ras, is a key signaling protein that is regulated by numerous receptor tyrosine kinases. One major modulator is the epidermal growth-factor receptor (EGFR), which is a transmembrane protein that interacts with different ligands, to activate Ras and the MAP Kinase (MAPK or ERK) pathway (see Fig. 11). The Ras GTPase is active in its Guanosine 5'-triphosphate (GTP)-bound state, but inactive when bound to Guanosine 5'-diphosphate (GDP), and when active, triggers the MAPK pathway. Oncogenic (activated) forms of Ras, are due to mutations that lock Ras in the GTP-bound state (e.g. Ras Valine-12 mutation) and contribute to ~25% of all cancers. The MAPK pathway, involves the serial activation of Raf (MAPKKK), MAPKK (MEK), and MAPK (ERK) kinase cascade leading to different outcomes depending on the cell context. Ras can also upregulate other signaling pathways, such as the PI3K pathway (see below) and lead to the activation of another small GTPase, Ral. In mammalian cells, the importance of Ras in cell proliferation is exemplified by the G1 arrest observed when neutralizing antibodies to Ras, or dominant-negative versions of Ras (blocked in the GDP-bound state) are injected into growth-factor stimulated cells, and the prevalence of activated Ras mutations in cancer. In Drosophila, EGFR–Ras signaling also regulates cell cycle progression, as well as cell growth, survival, differentiation, and patterning during development.

The Ras–MAPK cascade acts via phosphorylation of numerous effectors, including transcription factors of the Ets family, and Jun and Fos(Fra), which dimerize to generate the AP1 transcription factor. Likewise, in Drosophila, Ras signaling activates the Ets-family transcription factors, Pointed and Yan, and also the Jun/Fos transcription factor. In mammalian cells, an important function of the Ras–MAPK pathway is to phosphorylate and stabilize the c-Myc transcription factor, which plays an important role in cell growth and proliferation (see Sect. 2.5.1.3). In Drosophila,

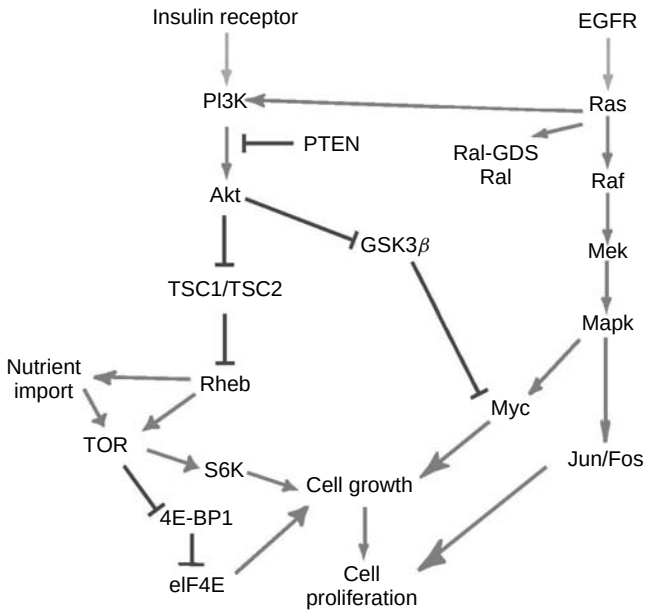

Fig. 11 Mitogenic signaling pathways – Ras, Myc, and Insulin-PI3K EGFR-Ras signaling activates the MAPK cascade leading to upregulation of transcription factors, such as Jun/Fos (AP1), and acts to phosphorylate and stabilize Myc. Ras and Myc regulate cell growth by increasing the transcription of translation initiation factors and ribosomal components, and promote the cell cycle by positively regulating Cyclin E levels or Cyclin E/Cdk2 activity. Signaling through the insulin receptor pathway occurs via PI3K and AKT, which phosphorylates and inactivates the Tsc1/Tsc2 complex, leading to activation of Tor and S6 kinase. Activation of S6 kinase results in increased cellular growth and promotes the G1-S phase transition. AKT activation also leads to phosphorylation of GSK3β, which acts to antagonize Myc and core cell cycle components, such as Cyclin E by phosphorylating them and targeting them for ubiquitin-mediated degradation. See text for further details.

activation of Ras also leads to upregulation of the Myc transcription factor, by an unknown mechanism.

Ras-mediated G1-S progression requires the inactivation of Rb, which depends upon the upregulation of *cyclin D1* transcription via the activation of Jun and Fos(Fra) transcription factors, which bind to the AP1 site on the Cyclin D1 promoter. Ras also acts to promote translation of Cyclin D1, to increase Cyclin D1 protein stability via activation of the PI3K pathway (and inhibition of GSK3β-mediated phosphorylation, which triggers ubiquitin-dependent degradation of Cyclin D; see Sect. 2.4.3 and the following). Activation of Ras also leads to decreased $p27^{KIP1}$ transcription through repression of the Forkhead transcription factors, decreased $p27^{KIP1}$ protein synthesis, and decreased $p27^{KIP1}$ protein stability through both Cdk2-dependent and independent

mechanisms. Finally, Ras signaling, via the MAPK pathway, directly affects cell growth by leading to the phosphorylation of the Mnk protein kinase that phosphorylates the eukaryotic initiation factor eIF4E and enhances its activity. Thus, Ras signals through many pathways to promote cell cycle progression.

Conversely, in some contexts high levels of Ras signaling can trigger upregulation of $p21^{CIP1}$, $p15^{INK4B}$ and $p16^{INK4A}$ and downregulation of *cdk4*, leading to cell cycle arrest. Upregulation of other pathways, such as the Rho GTPase pathway, which blocks $p21^{CIP1}$ expression, is required to overcome the G1 arrest caused by high levels of activated Ras signaling. Recent studies have also shown that in cells with high levels of Ras signaling, the transcription factor, NF-κB, is responsible for inhibiting *cdk4* transcription, and if NF-κB activity is blocked, then activated Ras-expressing cells show unrestrained cell proliferation.

2.5.1.3 **Myc.** c-Myc is a basic-helix-loop-helix-zipper (bHLHZ) transcription factor that dimerizes with a related protein, Max, and is upregulated in response to many growth-factor signaling pathways. c-Myc/Max complexes regulate genes involved in cell growth and metabolism, such as translation initiation factors and ribosomal mRNA and proteins. Myc also promotes the G1 to S phase progression in both mammalian cells and Drosophila. Drosophila *dmyc* mutations produce small adults, suggesting a role in cell and tissue growth. Indeed, Drosophila wing disc cells that express higher levels of dMyc compared with surrounding cells "outgrow" the neighboring cells by exhibiting faster rates of cell growth and cell proliferation, as well as inducing apoptosis in the neighboring cells. dMyc increases transcription of *Cyclin E* and the activity of Cyclin D/Cdk4, as well as increasing the levels of Cyclin E posttranscriptionally. In mammalian cells, c-Myc has also been shown to induce the transcription of D-type Cyclins, and in association with a zinc-finger transcription factor, Miz1, acts to repress transcription of the $p21^{CIP1}$ and $p15^{INK4B}$ genes.

c-Myc is normally rapidly turned over via ubiquitination and proteasomal degradation (see Sect. 2.4.3.4). The EGFR-Ras-MAPK signaling pathway (as seen earlier) regulates the level of c-Myc via phosphorylation of the Ser62 residue, resulting in c-Myc stabilization (Fig. 11). In Drosophila, the Ras pathway also leads to upregulation of dMyc, although the mechanism has not yet been elucidated. While phosphorylation of c-Myc at Ser62 stabilizes it, this event allows the subsequent phosphorylation at Thr58 by GSK3β, which leads to ubiquitin-mediated degradation of c-Myc (see Sect. 2.5.1.3). This mechanism provides tight control of c-Myc levels, which is important for strict regulation of G1-S phase progression. The importance of controlling Myc is highlighted by the prevalence of Myc overexpression in cancers.

2.5.1.4 **Insulin receptor-PI3 kinase pathway.** The insulin receptor (Inr) pathway (see Fig. 11) acts through the phospholipid kinase PI3-Kinase (PI3K), the protein kinase AKT (PKB), the protein kinase Tor, and S6 kinase (S6K), which together lead to increased cellular growth by induction of the protein translation machinery. The protein kinase Tor, also responds to nutrient levels to induce ribosome biogenesis. In mammalian cells, activation of PI3K stimulates S phase via increased Cyclin D/Cdk4 and Cyclin E/Cdk2 activity.

The PI3K pathway acts through AKT to phosphorylate and inactivate GSK3β, which normally acts to inhibit G1-S progression via phosphorylation of Cyclin D, E, and c-Myc, thereby promoting their ubiquitin-mediated degradation (see Sect. 2.4.3). AKT also downregulates Cdk inhibitors by several mechanisms: AKT phosphorylates p27^{KIP1} and p21^{CIP1}, sequestering them to the cytoplasm; leads to upregulation of Skp2, the F-box component of the SCF, which targets phospho-p27^{KIP1} and p21^{CIP1} for ubiquitin-mediated degradation; and phosphophorylates and inactivates Forkhead transcription factors thereby preventing the transcription of p27^{KIP1} and p21^{CIP1}. AKT also facilitates the assembly of CyclinD/Cdk4/p27^{KIP1} complexes, thereby allowing activation of Cyclin E/Cdk2 via Cki exchange.

In Drosophila, the Tubulous Sclerosis Complex (TSC1/TSC2) complex, a downstream component of the InR pathway (see Fig. 11), acts to repress S6K activity and decrease Cyclin E levels. In the absence of Inr activity, the TSC1/TSC2 complex, acts to repress the Rheb GTPase, which is required to activate TOR and S6K. Activation of Inr inhibits TSC1/TSC2, therefore relieving inhibition of Rheb and allowing S6K and Tor activation and subsequent phosphorylation of the ribosomal protein S6. S6 phosphorylation stimulates translation of a specific class of mRNAs, the so-called TOP (terminal oligopyrimidine tract in the 5′UTR) mRNAs that encode ribosomal proteins and elongation factors. Tor also phosphorylates and inactivates 4E-BP1, an inhibitor of the eIF4E translational initiation factor, thereby generally increasing translation. Mutational analysis of *tsc1/tsc2* in Drosophila shows that loss of this complex leads to increased S phases and Cyclin E levels.

2.5.1.5 **Hedgehog pathway.** The Hedgehog (Hh) signal is transmitted through two transmembrane proteins, Patched (Ptc) that accumulates in intracellular vesicles (endosomes) and Smoothened (Smo) that predominantly localizes to the plasma membrane (see Fig. 12). The zinc-finger transcription factor Gli (or Ci in Drosophila) acts in an Hh-dependent manner, to repress or activate target genes. In the absence of Hh signaling, Smo is inhibited indirectly by Ptc and directly by the kinesin-related protein Costal2 (Cos2), which causes internalization of Smo into endosomes, a process dependent upon actin, microtubules, and endocytosis proteins. Internalization of Smo, leads to the degradation of Ci (Gli) into its repressor form (Ci^{75} in Drosophila), which requires PKA phosphorylation and Nedd8 modification of the Cul1-based SCF^{Slimb} complex. At the same time, full-length Gli (Ci^{155} in Drosophila) is tethered to microtubules by Cos2, Fused and Suppressor of Fused (Su(Fu)), and further regulated by constitutive export from the nucleus. In Drosophila, the Ci^{75} repressor form localizes to the nucleus, where it inhibits transcription of targets such as *hh* and the TGFβ family member, *dpp*. In the presence of the Hh signal, Ptc inhibition of Smo is overcome, resulting in Smo activation and translocation from internal membranes to the cell surface, in a Cos2-dependent manner. This leads to the stabilization of full-length Ci^{155}, which translocates to the nucleus and activates targets, such as *hh*, *ptc*, *dpp* in Drosophila.

In Drosophila, Hh signaling was first implicated in cell proliferation when *ptc* mutants were identified in a genetic screen for regulators of Retinoblastoma (RBF1). Furthermore, *smo* mutants prevent cells in the posterior region of the third instar larval eye imaginal disc from entering

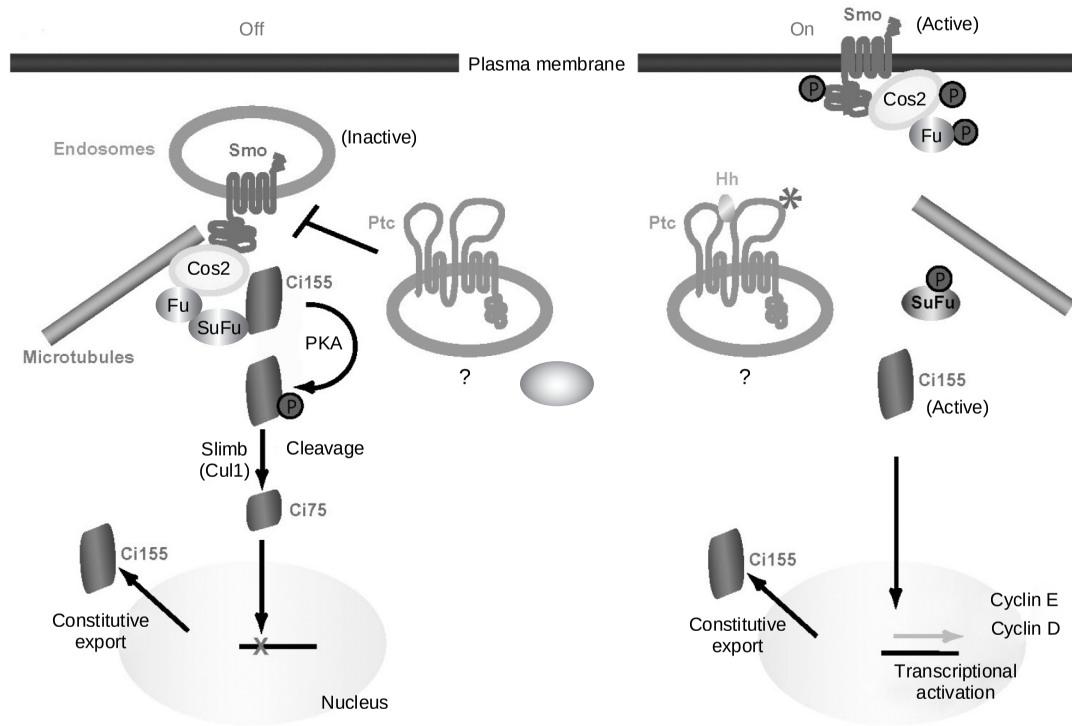

Fig. 12 Hh signaling pathway. In the absence of the Hedgehog signal, the transmembrane protein Patched localizes to endosomes and inhibits the activity of Smoothened via an unknown mechanism. Inactive Smoothened localizes to endosomes and is anchored to the cytoplasm via interaction with the Cos2/Fu/Su(Fu)/Ci complex which binds to microtubules. Formation of the Cos2/Fu/Su(Fu)/Ci complex allows the interaction with PKA and the subsequent phosphorylation of Ci. Phosphorylated Ci undergoes Slimb/Cul1 mediated degradation that yields a transcriptional repressor form of Ci. In the presence of the Hedgehog signal, Patched binds Hedgehog and relieves its repression upon Smoothened. Signaling by Hedgehog also leads to phosphorylation of Fu, Cos-2 and Su(Fu) via an unknown mechanism. The Cos2/Fu/Su(Fu)/Ci complex dissociates, and activated Smoothened moves to the plasma membrane in a Cos2-dependent manner. Free Ci is able to enter the nucleus and activate the transcription of its targets. Both *cyclin E* and *cyclin D* have been reported to be downstream targets of Ci and the Hedgehog pathway (see text for details).

S phase. Conversely, overexpression of Ci^{155}, results in ectopic S-phase entry in cells normally in G1 in the third instar larval eye and wing imaginal discs and the brain lobes. Moreover, Ci^{155}-induced S-phase entry is Rb-independent, since ectopic S phases are still observed in the presence of a nonphosphorylatable and constitutively active version of Rb, $RBF1^{280}$ (see Sect. 2.3.5). However, upregulation of Ci in the third instar larval eye imaginal disc, results in ectopic Cyclin E expression, which is due to the direct binding of Ci^{155} to the *cyclin E* promoter and induction of *cyclin E* transcription. This is consistent with the observation that progression into S phase in the posterior region of the eye disc is an dE2F1-independent process.

The Hh pathway also regulates Cyclin D expression in the Drosophila eye imaginal disc. Normally, Cyclin D is expressed in the posterior region of the Drosophila eye disc, with the highest level of expression coinciding with Ci expression in cells undergoing the G1-S phase transition. Consistent with Hh signaling normally being required for Cyclin D expression, reduced levels of Cyclin D are observed in *smo* mutant cells, and overexpression of Ci induces high levels of *cyclin D* transcript in the third instar larval eye and wing imaginal discs. Furthermore, overexpression of Patched reduces cell growth and cell proliferation, which can be overcome by coexpression of Cyclin D/Cdk4, indicating that Cyclin D is an important downstream target of Hh. Thus Hh signaling upregulates the two cyclins required for cell growth and cell proliferation in the Drosophila eye disc. It remains to be determined how general this mechanism is in Drosophila cell cycles and whether this is also the case in mammals Indeed, recent studies have shown that Hh signaling, like Dpp may be important in the developmental F1 arrest during Drosophila eye development.

2.5.1.6 **Wnt (Wingless) pathway.** Wnt or Wingless are extracellular morphogens that act through Frizzled receptors, to mediate signaling through several different pathways. In the canonical pathway (which we consider here), Wnt signaling leads to the transcriptional activation by β-catenin, an important transcriptional regulator of cell cycle genes (see Fig. 13). In the absence of Wnt signaling, β-catenin is localized to the cytoplasm in a complex with the two scaffolding proteins, adenomatous polyposis coli tumor suppressor protein (Apc) and Axin, which mediate phosphorylation of β-catenin by GSK3β, and subsequent degradation. Apc can also shuttle into the nucleus, and when overexpressed, acts in the nucleus to block G1-S progression. In the presence of Wnt, activation of the canonical Wnt signaling pathway results in stabilized, nuclear-localized β-catenin, which interacts with the HMG-box transcription factor (TCF), to activate transcription of cell cycle target genes including *c-myc* and *cyclin D1*.

In Drosophila wing development, Wingless (Wg) signaling is required for cell survival, as cells deprived of Wg activity die at a high frequency. However, inhibition of cell death in cells deprived of Wg activity, results in increased proliferation correlating with increased dE2F1 activity. Conversely, activation of the Wg pathway leads to reduced cell proliferation correlating with reduced dE2F1 activity. In contrast to observation in mammalian cells (as seen earlier), upregulation of Wg signaling in Drosophila inhibits *dmyc* expression in the third instar larval presumptive wing margin cell (referred to as

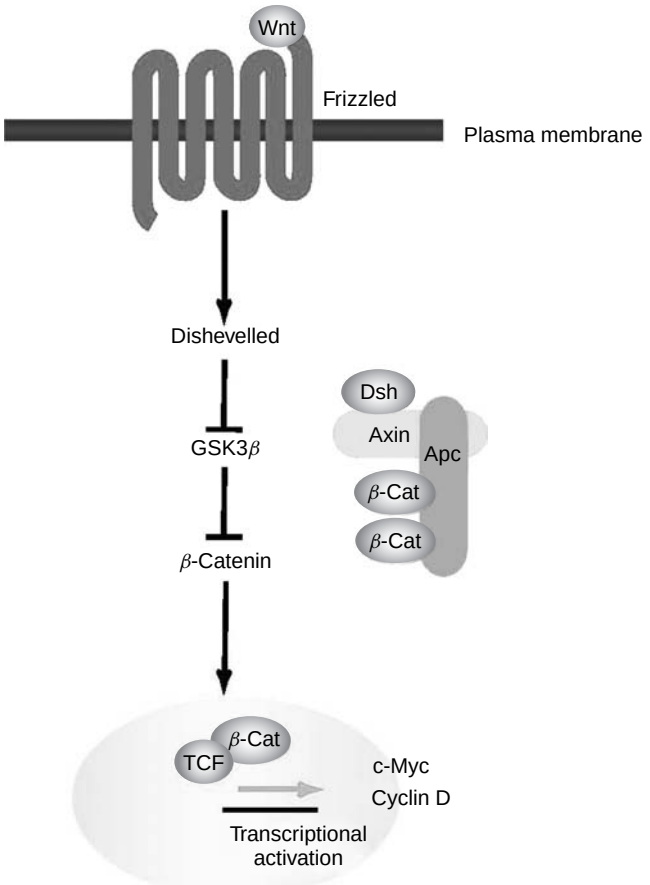

Fig. 13 Canonical Wnt (Wingless) signaling pathway. Wnt, or Wingless (Wg) in Drosophila, signals through the Frizzled transmembrane receptor and the PDZ-containing adaptor protein, Disheveled, to lead to the inactivation of the serine/threonine kinase GSK3β. In the absence of Wnt in mammalian cells, β-catenin is normally bound by a complex containing adenomatous polyposis coli protein (Apc), Axin and GSK3β, which phosphorylates and targets β-catenin for degradation. In the presence of the Wnt signal, Disheveled binds to Axin, excludes GSK3β and allows β-catenin to accumulate. Stable β-catenin translocates to the nucleus where it forms a complex with the HMG-box transcription factor, TCF (LEF1), and transactivates expression of target genes. Both *c-myc* and *cyclin D* have been reported to be downstream transcriptional targets of the Wnt pathway (see text for details).

the zone of nonproliferating cells, ZNC). In these cells, Wg-mediated downregulation of *dmyc* leads to an RBF1-mediated G1 arrest. However, in other tissues, such as the eye imaginal disc, upregulation of Wg signalling promotes cell proliferation.

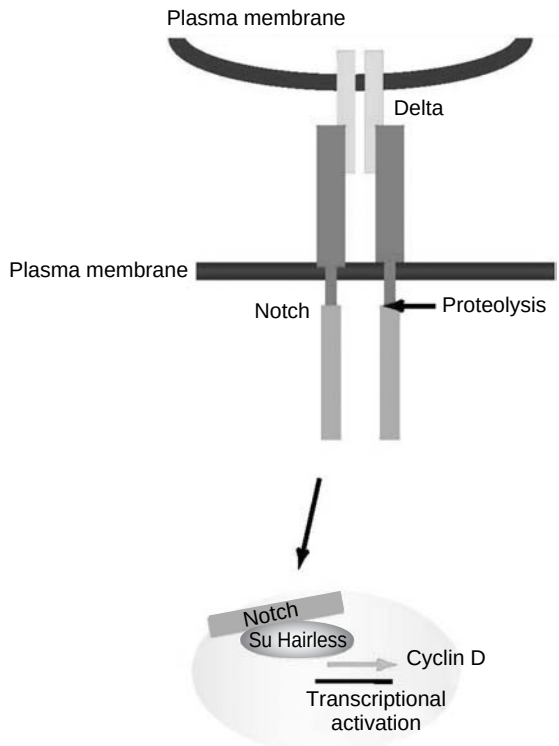

Fig. 14 Notch signaling pathway. In the Notch signaling pathway, the sending cell expresses Delta or Serrate, which are ligands for the Notch receptor at the receiving cell. Activation of Notch leads to its cleavage and its localization to the nucleus, where it associates with the coactivator Su(Hairless) to regulate transcription of its targets. *Cyclin D* has been reported to be a downstream target of the Notch pathway (see text for details).

2.5.1.7 **Notch pathway.** Notch is a family of large transmembrane proteins that mediate multiple cellular processes, including proliferation, differentiation, and apoptosis, by interacting with the transmembrane protein Delta in the adjacent cell. Notch signaling leads to cleavage of Notch and release of the Notch intracellular domain, which can enter the nucleus and bind to the transcription factor, Su(H), activating transcription of target genes (Fig. 14). In mammalian cells, Notch signaling leads to the induction of *cyclin D1* transcription and the activation of the Cdk2 protein kinase, promoting entry into S phase. In Drosophila eye development, Notch regulates neural cell differentiation, but also promotes cell proliferation. Recently, Notch signaling has been shown to play an important role in driving the developmental synchronized S phases during Drosophila development. While in other studies ectopic Notch signaling has been shown to promote cell proliferation in the developing eye by upregulation of the Jak kinase-Stat transcription factor signaling pathway and the Eyg transcription factor. Interestingly, Cyclin D/Cdk4 appears to act downstream of Jak, and plays a role in the stability of the Stat transcription factor. Furthermore, in Drosophila oogenesis, Notch signaling has been shown to have a role in the downregulation of the CIP(KIP) homolog, Dacapo, which is required to alter the cell cycle mode (see Sect. 2.7.2.2). However, in other contexts such as the developing wing Notch signaling can act to inhibit cell proliferation.

In conclusion, extracellular signaling via growth-factor receptors play a major role in regulating the G1-S phase transition through affects on transcription, translation, and proteolysis of cyclins, Cdks, and Cdk regulators. There are, however, additional mechanisms by which extracellular factors can regulate G1-S phase progression, which is by the interaction of cells with their neighbors and via their tissue context (see below).

2.5.2 Regulation of G1-S Phase Progression by Anchorage, Contact Inhibition, and Tissue Architecture

From the earliest days of growing cells in tissue culture, it was noted that cell proliferation was influenced by two events. First, cells depended upon attachment to the culture dish in order to enter into the cell cycle, a phenomena known as *anchorage dependence*. Second, it was observed that when cells grew to confluency, cells arrested in G1 phase, a phenomenon known as *contact inhibition*. These controls were known to be important for restraining cell proliferation, since loss of anchorage dependence and contact inhibition are hallmarks of cancer. It is now known that both of these events are due to communication between a cell and the neighboring cell or between the cell and the extracellular matrix (ECM, a complex ordered network of proteins, such as laminin and fibronectin).

Anchorage-dependent proliferation requires signaling through integrins after they have bound to their substrates in the ECM. Integrin-mediated adhesion via upregulation of the tyrosine kinases, Focal adhesion kinase (FAK) and Src, can lead to the upregulation of many signaling pathways, such as the Rac GTPase, PI3K (Sect. 2.5.1.4), Ras–MAPK (Sect. 2.5.1.2) and the Jun kinase pathways. Activation of these pathways, can lead to the activation of the Jun/Fos (AP1) and β-catenin transcription factors that are important for *cyclin D* transcription (see Sects. 2.5.1.2 and 2.5.1.5). Moreover, activation of the PI3K pathway, via AKT, leads to upregulation of Cyclin D translation and protein accumulation, and downregulation of Cdk inhibitors, and thereby leads to activation of Cyclin E/Cdk2 (see Sect. 2.4.2). Anchorage-dependent proliferation also requires the transcription of the *cyclin A* gene, via p107/E2F4 regulation, at the onset of S phase. Thus, by signaling through multiple pathways after contact with ECM components, integrins promote the upregulation of cyclin/Cdk activity and entry into S phase. By contrast, in some contexts, contact of specific integrins with particular components of the ECM can act to limit cell proliferation; however, the mechanism by which this occurs is not well understood.

Contact inhibition acts primarily through adherens junctions, which are specialized junctional complexes involved in cell–cell connection and signaling. Key proteins at adherens junctions, are the cadherins, such as E-cadherin and N-cadherin, which are large transmembrane proteins that form complexes with the cytoskeletal protein, α-catenin, and with the transcription factor, β-catenin (see Sect. 2.5.1.5). By sequestering β-catenin to the membrane, cadherins, prevent β-catenin from entering the nucleus, thereby preventing transcription of *c-myc* and *cyclin D*. The cadherin complex also regulates other pathways, such as the Rac GTPase, which can inhibit signaling via another GTPase, Rho. Rho activation in mammalian cells has been shown to promote cell cycle progression by leading to downregulation of the cyclin/Cdk inhibitors $p21^{CIP1}$ and $p27^{KIP}$, and upregulation of Cyclin D

and Cyclin E. Within a whole organism, contact inhibition via the adhesion junctions also appears to be important in limiting cell proliferation. In mouse models and in human disease conditions, disruption of the integrity of adherens junctions by mutations in the junctional proteins, α-catenin or E-cadherin, leads to hyperproliferation and tumorigenesis. For example, ablation of E-cadherin and α-catenin in mouse skin leads to hyperproliferation. The epithelial hyperplasia in mice mutant for α-catenin, is associated with ectopic Ras-MAPK signaling. Ectopic activation of signaling pathways, has also been observed in other studies when cellular architecture is disrupted, and was shown to be due to the breakdown of cell polarity, which normally acts to separate ligands and their receptors to exclusive cellular compartments.

Within a whole organism, the architecture of the tissue, the arrangement of individual cells to each other, can also be an important factor in restraining inappropriate cell proliferation. Epithelial cells, in particular, exhibit a highly organized cell shape (cell polarity), which requires the formation of specialized junctional complexes, the Bazooka complex, Crumbs complex, Dlg complex, and the adherens junctions, along their apical–basal axis (Fig. 15). The importance of cell polarity and integrity of junctional complexes to cell proliferation control, is dramatically exemplified by the Drosophila neoplastic tumor suppressor genes, *dlg,* encoding a Membrane Associated Guanylate Kinase (MAGUK) protein, *scrib* (encoding a leucine rich repeat, LRR and PDZ domain protein) and *lgl* (encoding a WD40 domain-containing protein). Dlg and Scrib localize to the septate junctions, and genetically interact with Lgl, which is cytoplasmically and cortically localized. In the absence of any of these proteins, correct polarization of epithelial tissues and neuroblasts is disrupted, the tissue fails to properly differentiate, and cells upregulate Cyclin E and overproliferate to form large invasive tumors. The ability of Scrib, Dlg, and Lgl to regulate cell polarity and proliferation is likely to be conserved in the mouse, since mouse *lgl1* mutant exhibit defects in neural progenitor cell polarity associated with overproliferation. A second Lgl homolog, Lgl2, also exists and exposing redundant functions will require the generation of an Lgl2 knockout and analysis of the double mutant. Four Dlg homologs exist in the mouse, the overlapping functions of which are yet to be explored *in vivo*. However, Scrib is represented by a single gene (*Scrb1*), mouse mutants of which exhibit dramatic developmental defects in neural tube closure, although cell proliferation defects have not yet been reported. Interestingly, however, both Dlg and Scrib proteins contain PDZ domains and are targeted for degradation by the E6 onco-protein of high-risk human papilloma viruses, which appears to be important for E6's ability to promote epithelial hyperplasia.

A central issue raised by studies of *scrib, lgl,* and *dlg* mutants is whether excessive cell proliferation is a consequence of disrupting cell polarity or due to cell polarity-independent functions of these proteins. At this stage, very little is known of the molecular mechanism by which, Lgl, Dlg, or Scrib connect to the cell cycle. Studies in Drosophila suggest that the polarity and proliferation defects of *lgl* mutants can be suppressed by reducing the dosage of atypical protein kinase C (aPKC), a component of the Bazooka apical complex). However, it is not clear whether this reflects a positive cell cycle role for aPKC downstream of Lgl, or an inhibitory

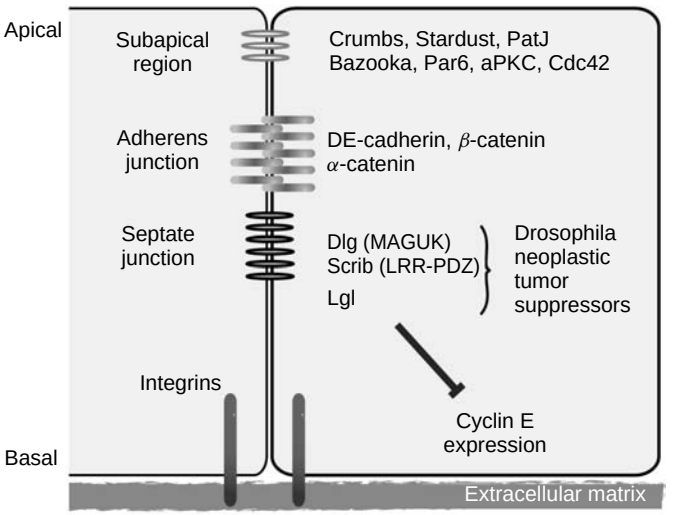

Fig. 15 Cell polarity and G1-S regulation. A polarized epithelial cell exhibits distinct domains of membrane associated-protein complexes along its apical–basal axis. At the apical surface, known as the subapical region, there are two complexes: The Crumbs complex (Crumbs/Stardust/Patj) and the Bazooka complex (Bazooka/Atypical Protein Kinase C (aPKC)/Dpar-6/Cdc42). Below the subapical region lies the adherens junction where, E-cadherin, β-catenin and α-catenin form a complex. Basal to the adherens junction, is the septate junction in Drosophila (or the basal region in mammalian cells), where Dlg and Scribble (Scrib) are located. Lgl can also concentrate at the septate junctions and genetically interacts with Scrib and Dlg. At the basal surface of the cell, the transmembrane Integrin proteins localize and form connections to components of the extracellular matrix. In Drosophila Dlg, Scrib, and Lgl are neoplastic tumor suppressors and mutation in any one of these proteins leads to loss of polarity and uncontrolled cell proliferation. Dlg, Scrib, and Lgl work in a common pathway to regulate cell polarity and inhibit G1-S progression by preventing Cyclin E expression.

role for aPKC on Lgl function. In the *lgl1* mouse knockout, the hyperproliferation of neural cells is correlated with mislocalization of Numb, a protein involved in asymmetric division, and ectopic Notch signaling. Consistent with this, there are strong similarities between the hyperproliferation defects of the *lgl1* knockout and the conditional double knockout of Numb and Numblike. In this context, therefore, it seems likely that defects in the regulation of cell asymmetry are primarily responsible for the inappropriate activation of a signaling pathway promoting ectopic cell proliferation.

Intriguingly, a specific connection between the regulation of Cyclin E and cellular architecture has recently emerged from a genetic screen in Drosophila for dominant suppressors of a *cyclin E* hypomorphic allele. Unexpectedly, within this sensitized background, adhesion and cytoskeleton genes (including *scrib*, *dlg*, and *lgl*) were identified as negative regulators of

cyclin E. This suggests that there are direct links between the polarized cell architecture and the cell cycle machinery *in vivo*. Indeed, analysis of mutants and dissection of domains for Drosophila Dlg and Scrib, suggest that their cell proliferation role is separable from their role in cell polarity, although how these genes regulate Cyclin E levels is not known.

In conclusion, progression into S phase depends upon extracellular inputs, both from diffusible growth factors, and through contact with neighboring cells and the ECM. Moreover, the structure (shape) of a cell is very important in enabling appropriate contacts with adjacent cells and with the ECM, as well as for correct localization of growth-factor receptors to provide strict control of cell proliferation. When these controls fail, inappropriate cell proliferation can ensue, leading to cancer.

2.6 Regulation of G1-S Phase by the DNA-damage and DNA-replication Checkpoint: p53

Another important control of G1-S phase progression occurs in response to DNA damage or DNA-replication errors. These controls are known as the DNA-damage and DNA-replication G1 checkpoints. When DNA is damaged or DNA replication is halted in response to environmental assaults, monitoring proteins sense the damage or impairment and lead to cell cycle arrest until the damage can be repaired or apoptosis (regulated cell death) is triggered to eliminate the damaged cell. This checkpoint control is very important to maintain genomic integrity, and if this mechanism fails the accumulated DNA damage and/or chromosome loss (aneuploidy) is highly conducive to the development of cancer. The tumor suppressor, p53, is a major player in the DNA-damage and DNA-replication checkpoint mechanism, and its importance in this mechanism is highlighted by its mutation in most human cancers.

In G1 or S phase cells, p53 is activated in response to DNA damage and DNA-replication errors by a signaling cascade involving the ATM (ataxia telangiectasia mutated) or ATR (ATM-related) and the Chk2 protein kinases (see Fig. 16). ATM is activated by double-strand DNA breaks due to ionizing radiation, while ATR is activated by DNA base damage due to ultraviolet radiation and by unreplicated DNA. The p53 transcription factor is normally highly unstable due to ubiquitin-mediated degradation, mediated by the Mdm2 ubiquitin ligase. Upon phosphorylation by ATM or ATR and Chk2 protein kinases, p53 is stabilized and is able to transactivate targets involved in cell cycle arrest, DNA repair, and apoptosis. The critical p53 target required for G1 phase arrest is the $p21^{CIP1}$ Cdk inhibitor. By upregulation of $p21^{CIP1}$, the DNA damage or DNA-replication checkpoint pathways lead to inactivation of cyclin-Cdks and arrest in G1 or S phase, until the damage can be repaired. In Drosophila, the p53 pathway is also conserved and acts in response to DNA damage; however, the outcome of this pathway in Drosophila is focused toward upregulation of the apoptotic pathway, and not upregulation of the CIP(KIP) homolog, Dacapo.

An additional layer of regulation on p53 activity is mediated by the $p19^{ARF}$ protein, which is encoded by an alternative reading frame in the $p16^{INK4A}$ locus. In response to high levels of mitogenic signaling and upregulation of the Ras, Myc, and E2F pathways, transcriptional upregulation of both $p16^{INK4A}$ and

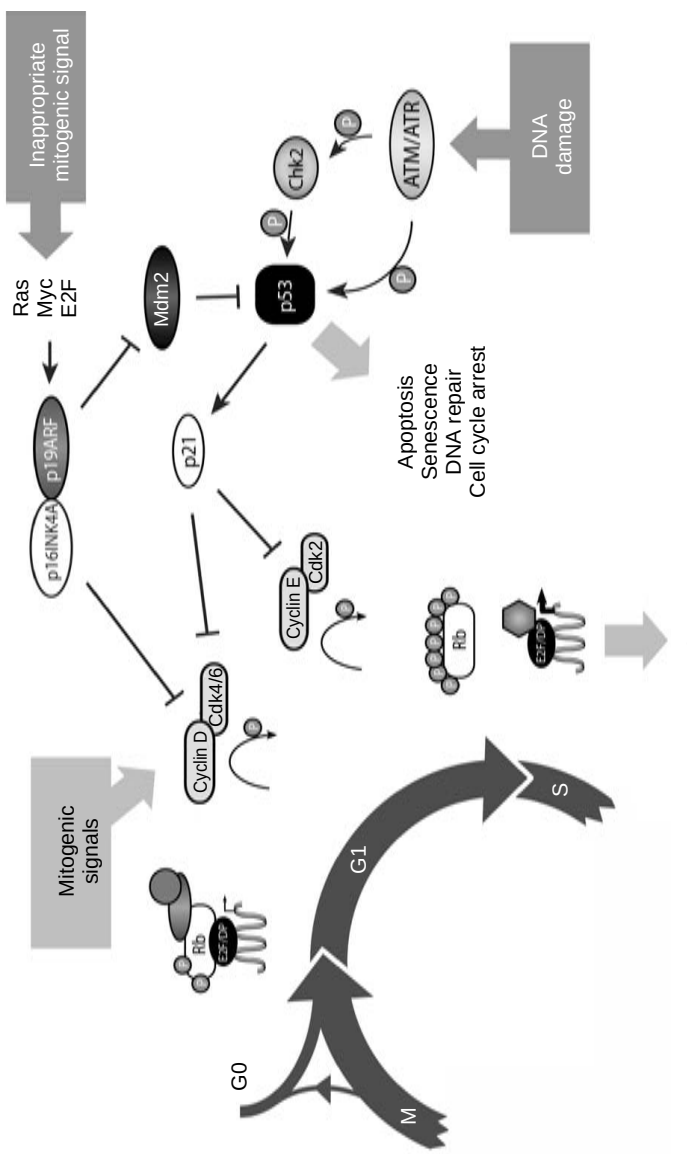

Fig. 16 DNA damage and DNA replication G1-S checkpoint: p53. Upon DNA damage or perturbations to DNA replication, the protein kinases ATM and ATR are induced. Activation of ATM (ATR) leads to phosphorylation and activation of the Chk2 protein kinase and phosphorylation of the tumor suppressor transcription factor, p53. Phosphorylation of p53 stabilizes it, allowing it to activate target genes, such as $p21^{CIP1}$, resulting in inhibition of Cyclin/Cdks and G1-S phase arrest. p53 can also trigger DNA repair, apoptosis and senescence. In addition, inappropriate mitogenic signaling through Ras, Myc, and/or E2F results in transcriptional induction of the *INK4A* locus, that encodes both the Cki $p16^{INK4A}$ and the tumor suppressor $p19^{ARF}$. The INK4A locus therefore controls two pathways necessary for cell cycle arrest, $p16^{INK4A}$ inhibits Cyclin D/Cdk4 assembly and $p19^{ARF}$ downregulates the p53 ubiquitin-ligase Mdm2 resulting in stabilization of p53, transcription of $p21^{CIP1}$ and inhibition of Cyclin E/Cdk2.

p19^{ARF} occurs. While p16^{INK4A} inhibits cyclin/Cdk4 activity, p19^{ARF} acts to stabilize p53 by inhibiting the Mdm2 ubiquitin ligase. The convergence of the p16^{INK4A}-Rb and p19^{ARF}-p53 pathways at the *INK4A* locus remains an evolutionary puzzle as it provides a doubly effective target for mutation. Indeed, the *INK4A* locus is frequently mutated in human cancer, particularly in melanoma.

In addition to their role in upregulating p53, ATM and ATR also have a p53-independent role in regulating the firing of DNA-replication origins (see Sect. 2.7.2). Upon perturbations to DNA replication, such as inhibition of DNA polymerase by the drug aphidocolin, ATM (ATR) trigger activation of Chk1 and Chk2, which then inhibit the Cdc25A phosphatase, thereby leading to the inhibition of Cyclin E/Cdk2 and preventing DNA-replication initiation. In addition, ATM and ATR directly phosphorylate components of the DNA-replication initiation complex, the MCM proteins (see Sect. 2.7.2), which is important for these checkpoints to function. The importance of this mechanism is highlighted by the premature DNA-replication origin firing that occurs when ATM (ATR) are inhibited by the drug, caffeine.

In conclusion, the G1 DNA damage and DNA-replication checkpoints provide additional tiers of regulation to the G1-S phase transition, by monitoring the status of the cell so that DNA damage can be repaired and DNA replication can occur faithfully.

2.7 G1-S Cyclin/Cdks and Progression into S Phase

Once Cyclin E/Cdk2 activity has been fully activated and Rb has been phosphorylated and inactivated, the cell becomes committed to enter S phase. Two important events need to occur during S phase: centrosome duplication and DNA replication. The regulation of centrosomal duplication by Cyclin E/Cdk2 and the events occurring at the initiation of DNA replication are discussed in the following.

2.7.1 Centrosomal Duplication

The centrosome (microtubule-organizing center, MTOC) is important for segregation of chromosomes in mitosis, and is duplicated in S phase. The centrosome consists of a central centriole (clusters of microtubules) surrounded by protein matrix, termed the *pericentriolar material*. For the centrosome to duplicate, the centriole first needs to duplicate. The daughter centriole forms in an orthogonal orientation to the mother centriole, by an unknown mechanism. The mother and daughter centrioles then separate and form a pair of centrosomes from the pericentriolar material. Centrosome duplication in vertebrates requires Cyclin E/Cdk2 activity. Constitutive activation of Cyclin E or mutation of p21-Cip1 results in an increase in centrosome numbers. How Cyclin E/Cdk2 results in centrosomal duplication in mammalian cells is presently unclear, but the centrosomal proteins, Nucleophosmin (NPM/B23) and CP110, which are required for centrosome duplication, have been shown to be substrates of Cyclin E/Cdk2. However, a Cdk2-independent role for Cyclin E in centrosomal duplication, by specific localization of Cyclin E at the centrosome has recently been described. This finding is consistent with roles for Cyclin E, distinct from Cdk2 that have been revealed by mouse knockout studies (see Sect. 2.2.1.2). In Drosophila, however, while mutations in SkpA (homolog of Skp1) of the SCF ubiquitin–ligase complex, have supernumerary

centrosomes, this is apparently not due to upregulation of Cyclin E/Cdk2 activity, since *skpA, cyclin E* double mutants still exhibit overduplication of centrosomes. Thus, it appears that there may be differences in regulation of centrosome duplication between Drosophila and mammalian cells.

2.7.2 Initiation of DNA Replication

Studies in the budding yeast, *S. cerevisiae*, have generated a basic model for how duplication of chromosomes is controlled during the cell cycle (see Fig. 17). Genetic analysis in *S. cerevisiae* has identified well-defined origins of DNA replication to which a large prereplicative protein complex (pre-RC) is recruited during the M-G1 phase transition of the cell cycle. Components of the pre-RC have been identified, by genetic and biochemical approaches, and include six origin recognition complex, ORC proteins, six minichromosome maintenance MCM (MCM2–7) proteins, Cdc6, and Cdc45 (see Fig. 17). The ORC proteins remain associated with the replication origin throughout the cell cycle, while other components of the preinitiation complex are sequentially loaded onto origins during G1 phase. Cdc6 is first recruited to origins by an ORC-dependent mechanism. Cdc6 then recruits the MCM protein complex onto DNA-replication origins. In late G1, recruitment of Cdc45 to the pre-RC allows the binding of DNA polymerase-α. Cdc6 is a member of a large superfamily of ATPases known as the AAA^+ family and it exhibits significant sequence similarities to subunits of clamp-loading proteins in eukaryotes (replication factor C complex) and prokaryotes (γ-complex), and may be the loading factor that clamps the MCM protein complex to the DNA. The MCM protein complex has recently been shown to have helicase activity *in vivo* and is required for unwinding the DNA helix at the replication fork during DNA-replication elongation.

The mechanism by which origins are activated is not fully understood, but recent studies have shown that phosphorylation of Cdc45 by Cdk and Cdc7/Dbf4 protein kinases is required for the loading of Cdc45 onto replication origins. Once DNA replication is initiated and a bidirectional replication fork is generated, disassembly of the pre-RC complex occurs, thereby preventing further DNA polymerase-α recruitment and rereplication. In yeast, reinitiation is prevented by mitotic Cyclin B/Cdk activity by maintaining MCMs in a hyperphosphorylated state and inhibiting new synthesis of Cdc6, which is degraded by ubiquitin-mediated proteolysis after origin firing. At the completion of mitosis, inactivation of mitotic cyclin-associated Cdk activity marks the end of inhibitory signals that block new pre-RC formation and allows the pre-RC to reform. Thus, Cdk activity in yeast is responsible for origin firing as well as restricting initiation to once per cell cycle.

Most regulators of replication in yeast such as ORCs, MCMs, Cdc45, and Cdc6 have counterparts in animal cells including Drosophila, although the regulation has not yet been rigorously defined. As in yeast, Drosophila Cdc6 is recruited to the origins in an ORC-dependent process followed by the minichromosome maintenance 2–7 (MCM2–7) protein complex. Later in G1, the recruitment of Cdc45 allows the binding of the DNA polymerase-α at the origin. Recruitment of Cdc45 is dependent on phosphorylation by Dbf4/Cdc7 and in Xenopus the replication initiation factor MCM10 is required after pre-RC assembly and to load Cdc45 onto chromatin. Cyclin E/Cdk2 is subsequently required to allow replication initiation. Indeed, in

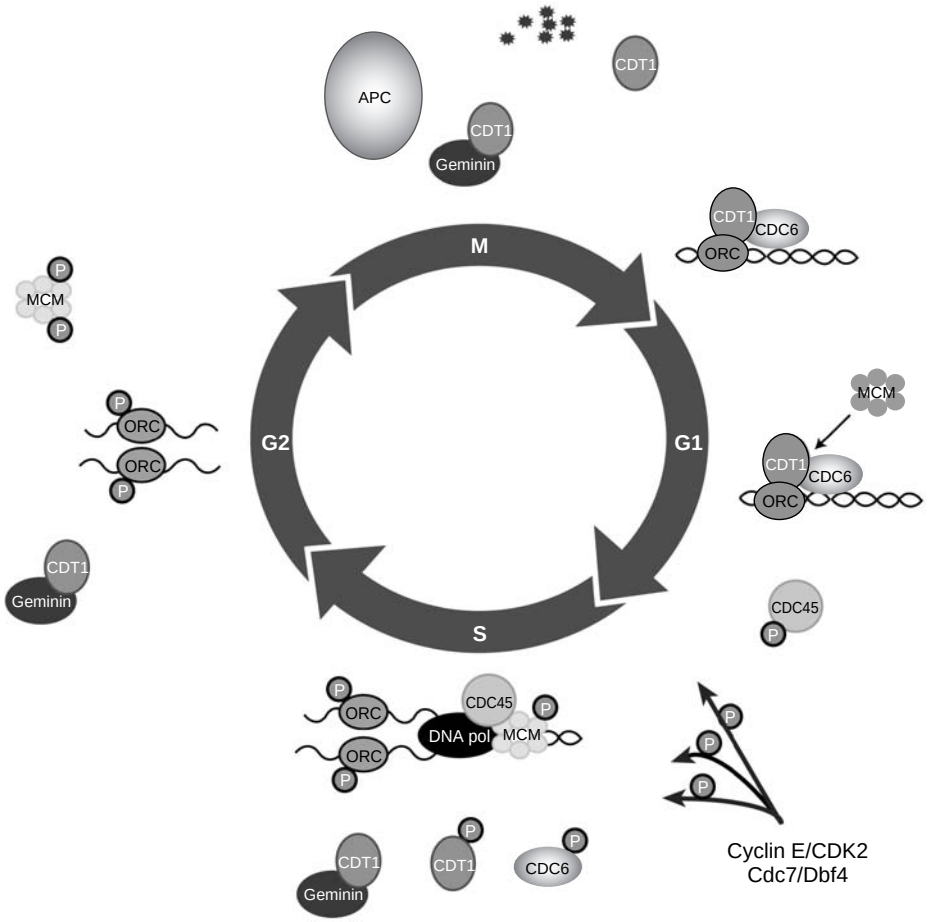

Fig. 17 The initiation of DNA replication. In late G1, origins are first recognized and bound by a six-subunit origin recognition complex (ORC), which remains associated with the origin of replication throughout the cell cycle. In an ORC-dependent process, Cdc6 and Cdt1 are recruited to the origins and are necessary to recruit the MCM2-7 protein complex. Cdc45 can then binds and assemble the prereplication (pre-RC) complex, which allows the binding of the DNA polymerase-α. The CyclinE/Cdk2 and Dbf4/Cdc7 protein kinases are then recruited to the pre-RC complex and trigger DNA replication. After initiation of replication, Cdks phosphorylate Cdc6 and Cdt1, thereby preventing re-replication during S, G2, and M phase. In metazoans, Geminin prevents re-replication during S, G2, and M phase by binding and sequestering Cdt1. Geminin is degraded via the anaphase promoting complex (APC) at the meta-anaphase transition. In this way, Cdt1 allows MCM binding to occur from late mitosis to early S phase.

Xenopus, Cyclin E binds to Cdc6, thereby localizing the Cdk2 complex to chromatin, which is important for DNA-replication initiation. Components of the SCF ubiquitin–ligase complex, including Skp1 and Cul1, are also recruited to chromatin and the preinitiation complex through Cyclin E/Cdk2, and are important in degrading the p27^{KIP1} inhibitor bound to Cyclin E/Cdk2, thereby allowing activation of

Cyclin E/Cdk2 and DNA-replication origin firing.

While most proteins are conserved between yeast and animals, there appears to be considerable differences in regulation of DNA-replication initiation between organisms. One example is the loading of Cdc45 onto pre-RCs. In extracts from the oocytes of the frog, *Xenopus laevis*, phosphorylation of Cdc45 by Dbf4/Cdc7 protein kinase recruits Cdc45 to the pre-RC and Cyclin E/Cdk2 functions after this step to allow initiation, while in yeast, Cdk activity is required before Dbf4/Cdc7. Another example includes differences in regulation of Cdc6. While in yeast, Cdc6 activity is regulated by degradation, its regulation in animal cells is more complex. In human cells, Cdc6 activity requires Cdk phosphorylation in G1, while Cyclin A/Cdk phosphorylation in early S phase results in nuclear-cytoplasmic export. This Cyclin A-dependent event, results in the degradation of free Cdc6, while pre-RC-bound Cdc6 remains stable. Phosphorylated Cdc6 is targeted for ubiquitination by the Skp1, Cul1, F-box (SCF) ubiquitin E3 ligase and subsequently degraded by the 26S proteasome. Therefore, activity of soluble Cdc6 is controlled through G1-S cyclin/Cdk2 activity, while chromatin-bound Cdc6 is constant for most of the cell cycle, suggesting that DNA-replication initiation is not primarily controlled by the presence of Cdc6 on chromatin. In contrast to these findings, *in vitro* studies in Xenopus, suggest that phosphorylation of Cdc6 is not important for either DNA-replication initiation or blocking rereplication.

To explain how DNA replication is limited to once per cell cycle in metazoans, a factor termed *licencing factor* was proposed to permit DNA to become replication-competent after mitosis, but then to be inactivated once DNA replication was initiated, until resynthesized again after mitosis. Recently this licensing factor has been shown to be Cdt1. Cdt1, together with Cdc6, promotes loading of MCMs onto replication origins in M-G1 phase and is inactivated after origin firing in S phase. The majority of Cdt1 protein is rapidly degraded at the G1-S transition. Evidence from several organisms suggests that Cdt1 is targeted for proteasome-mediated degradation by two ubiquitin–ligase complexes, an SCF ubiquitin ligase that contains the specificity subunit Skp2, and an SCF-like ubiquitin ligase that is based on Cul4. The SCF-dependent proteolytic pathway is activated at the G1-S transition, and SCF targets are marked for degradation by Cdk-dependent phosphorylation. In Drosophila, the Cdt1 homolog, Double-parked (Dup) is abundant in G1 when origins are licensed, but is then rapidly degraded when Cyclin E appears at the G1-S transition. This degradation appears to be at least partly regulated through Cyclin E/Cdk2 phosphorylation; however, it seems that additional mechanisms are needed to ensure complete Dup degradation. In human cells, the decrease in hCdt1 protein levels closely mirrors the accumulation of Cyclin A. Indeed, hCdt1 has a consensus site for Cdk phosphorylation, suggesting that hCdt1 may be phosphorylated by a Cyclin A–Cdk complex and targeted for SCF-dependent proteolysis in S phase. The oscillation of hCdt1 protein levels, similar to that of *S. pombe* Cdc18 (the Cdc6 homolog) and Cdt1, may therefore be important for preventing origin refiring in S and G2 phase in human cells.

In metazoans, it has recently been discovered that an additional tier of regulation exists. A novel DNA-replication inhibitor, Geminin, has been identified

in Xenopus, Drosophila, and mammalian cells. Geminin functions as a negative regulator of DNA-replication initiation by binding to Cdt1 and thereby inhibiting the loading of the MCMs. Consistent with Geminin functioning to inhibit Cdt1 and prevent rereplication, Geminin is high in G2 cells, but is degraded in mitosis. In this way, Cdt1-induced MCM binding can only occur when Geminin is absent, at the end of mitosis until early S phase.

2.7.2.1 **Transcriptional regulation of DNA replication.** A key question in understanding DNA replication in metazoans is what determines the timing of DNA-replication origin firing. Global genomic and expression analysis in Drosophila have revealed a strong correlation between gene expression and timing of DNA replication, where early replicating regions correlate with expressed genes and late replicating regions generally correlate with silent genes. Indeed many studies have found correlations between the chromatin state and the firing of DNA-replication origins, suggesting that chromatin-remodeling complexes may play an important role in regulating DNA replication. Recent studies in Drosophila have shown that histone acetylation is an epigenetic event that regulates DNA replication. Acetylated histone H4 colocalizes with ORC2 at times of active initiation of replication in *chorion* genes and is enriched specifically at the origins. Loss of the histone deacetylase (HDAC), Rpd3 by mutation (or by use of an HDAC inhibitor) results in genome reduplication independent of transcription. In addition, Rpd3 and the chromatin-remodeling protein, Polycomb (Pc), locally inhibit DNA replication origin activity, whereas the histone acetylase, HAT1 (Hbo1 in Drosophila) activated origin firing. In mammalian cells, HDAC and the Sin3 HDAC-recruiting complex have been found to associate with the DNA replication and repair protein, Proliferating cell nuclear antigen (PCNA), Replication factor C (RFC), which acts as a clamp loader for PCNA, or with the DNA polymerase-ε complex involved in DNA-replication elongation. Together these data indicate that widespread acetylation may open chromatin and allow DNA-replication origin proteins to gain access to additional binding sites, thereby imparting origin activity in both Drosophila and mammalian cells.

How are these chromatin modifiers recruited to origins? In Drosophila RBF1, dE2F1 and dDP function together at origins of replication to regulate replication of the *chorion* genes during the amplification endoreplication cycles of the ovarian follicle cells (see the following). Indeed, dE2F1 can bind directly to ORC2 at these *chorion* gene origins, and the localization of ORC2 to these origins is prevented in *dE2F1* and *dDP* female sterile mutants. Furthermore, activating mutations of dE2F1 have been identified that have increased replication at the *chorion* gene loci. These mutations remove the binding site for the Drosophila retinoblastoma homolog, RBF1, but also remove the transactivation domain of dE2F1. This shows that dE2F1 has a DNA replication–promoting role at the *chorion* gene loci that is independent of dE2F1s role as a transcriptional activator. This mechanism is not limited to these specialized cycles, since transcription-independent roles for dE2F in inducing S phase have also been documented in the Drosophila eye imaginal disc. Furthermore, this mechanism appears to be conserved as a similar nontranscriptional role in DNA-replication inhibition has been observed

for Rb in mammalian cells. In mammalian cells, Rb binds to MCM7 and inhibits DNA replication and Rb localizes to DNA-replication foci. These studies suggest that the E2F/DP complex plays a direct role at DNA-replication origins which is antagonized by Rb. Finally, mammalian Rb can bind to HDACs and the chromatin-remodeling HDAC complex is recruited with Rb to sites of DNA-replication initiation in mammalian cells. Since the SWI(SNF) chromatin-remodeling complex (Brg1 complex) has been implicated in Rb/E2F-mediated transcriptional regulation with HDACs (see Sect. 2.4), these findings described earlier, now beg the question of whether the SWI(SNF) complex may also be involved in DNA-replication initiation.

The transcription factor DmMyb is another protein that binds to and regulates the activity of *chorion* gene origins in Drosophila. Loss of DmMyb in follicle cells leads to impaired amplification of *chorion* origins despite the presence of the prereplication complex (pre-RC) at the origins. Importantly, recent studies have demonstrated striking similarities between the phenotypes of mutants of DmMyb and its interacting protein Mip130 and that of mutants in dE2F1 and dE2F2. Furthermore, DmMyb and associated proteins are in tight association with RBF1, RBF2, dE2F2 and dDP, and the HDAC, Rpd3. Therefore, DmMyb and dE2F provide a means to differentially recruit chromatin modifiers at *chorion* origins and regulate their activity. These complexes have further been found in Drosophila at stages undergoing genomic DNA replication. It is possible that similar mechanisms also act to regulate genomic replication in mammalian cells.

2.7.2.2 **Alternative modes of DNA replication.** Analyses of Drosophila development have revealed a number of variations to the canonical view of G1-S phase regulation and DNA replication (see Fig. 18). In Drosophila, the early embryonic cell cycles are driven by maternally supplied proteins and mRNA, and occur very rapidly without need for cell growth. Cyclin E becomes limiting at cycle 17, and most cells in the embryo arrest in G1. The upregulation of the CIP(KIP) inhibitor, Dacapo, and the presence of retinoblastoma, RBF1, is also required for the exit of these cells from the cell cycle. After this first G1 phase arrest occurs in embryogenesis, cell growth, and transcription of Cyclin E is required for all further DNA-replicative cycles, such as those of the neuroblasts and imaginal disc cells and to drive a modified replicative cycle, termed *endoreplication*.

Endoreplication, consisting of rounds of S phases without mitosis, occurs in the gut, salivary glands, and other tissues of the developing embryo and larvae and in the adult ovarian nurse and follicle cells. During endoreplication cycles, replication is not complete, leading to amplification of the euchromatic regions of the genome. Endoreplication cycles are driven by *cyclin E* pulses and regulated by dE2F1 and CIP(KIP) homolog, Dacapo, expression. During the follicle cell endocycles, a specialized type of endoreplication occurs, known as *DNA amplification*, where specific regions, such as the *chorion* gene loci are massively amplified. This requires a specialized DNA-replication origin, known as the amplification control element (ACE), a 300-bp sequence that responds to developmental signals to trigger repeated rounds of replication. These endoreplication and amplification cycles are important to allow increased protein production for either rapid larval growth or production of an oocyte in the adult

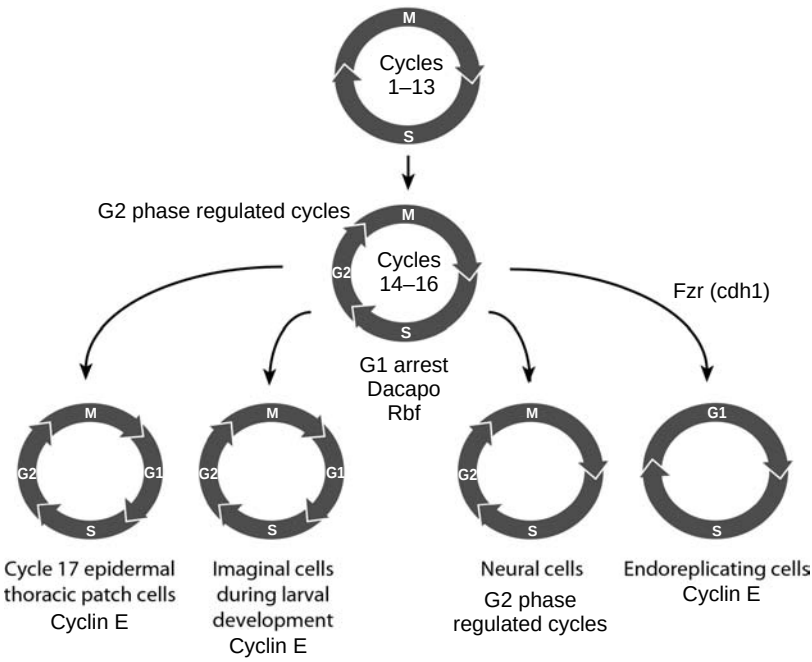

Fig. 18 Alternative modes of DNA replication in Drosophila development During Drosophila development, there are different modes of cell cycle regulation at different stages and in different tissues. During early embryogenesis, cell cycles are very rapid due to the high levels of maternally supplied proteins. Later in embryogenesis, G2 regulators become limiting and cell cycles are controlled by the cyclic synthesis of G2 regulators. Cyclin E becomes limiting at cycle 17, and most cells in the embryo then arrest in G1. The upregulation of Retinoblastoma (Rbf) and the $p21^{CIP1}$ ($p27^{KIP1}$) homolog Dacapo is also required for the timely arrest of these cells in G1. Cyclic transcription of cyclin E then drives cell cycle progression of certain cell types, such as those of the cycle 17 patch and imaginal discs. Cyclin E is also essential for switching from the canonical cell cycle to endoreplication, which consists of cycles of alternate DNA replication phases without mitoses. Upregulation of Frz (Fizzy-related/Cdh1), a regulator of the anaphase promoting complex (APC) ubiquitin ligase, is required to degrade G2 regulators and for switching the mitotic cycles to endoreplication cycles.

female. In the mouse, endoreplication cycles occur in the trophoblast cells of the placenta during development, which is important for the function of the placenta. These cycles are also highly dependent on Cyclin E, although Cyclin E can be redundant in most other cell cycles (see Sect. 2.2.1.2).

Studies of the follicle cell epithelium of the Drosophila ovary have shown that the Notch signaling pathway controls the follicle cell mitotic-to-endocycle transition by stopping the mitotic cycle and promoting the endocycle. Downregulation of the G2-M phosphatase Cdc25(String), occurs in response to Notch signaling and is required for the mitotic-endocycle transition. In a screen for genes whose transcription was markedly changed at the mitotic-endocycle transition, the growth regulator

dMyc, the CDK inhibitor Dacapo and the APC ubiquitination complex regulator, Fizzy-related (Fzr, Hec1, Cdh1) were identified and shown also to be important for the mitotic-endocycle transition. Upregulation of Fzr leads to degradation of G2-M phase regulators, such as Cyclin A, Cyclin B, and Cdc25(String), thereby preventing cells that have replicated their DNA from entering mitosis. Although not required for mitosis, Fzr is essential for endocycles and when expressed prematurely during mitotic stages, Fzr is sufficient to stop the mitotic cycle and can promote ectopic endocycles. Similarly, in embryogenesis, for endocycles to initiate, G2-M cell cycle components need to be degraded, which is mediated by the upregulation of Fzr. Downregulation of Dacapo at the mitotic-endocycle transition is consistent with the observation that higher levels of Dacapo are required in the oocyte during the early stages of oogenesis, to ensure that the oocyte stays in meiotic prophase. Indeed, ectopic overexpression of Dacapo prevents entry into the endocycle and *dacapo* mutant oocytes precociously enter the endocycle.

Consistent with the involvement of the Notch pathway in the mitotic cycle-endocycle switch, Notch signaling is required for downregulation of Dacapo and for upregulation of Fzr expression. Indeed, persistent expression of *dacapo* occurs in the Notch signaling mutant, *Delta*. Therefore, Notch signaling regulates the mitotic-to-endocycle switch by coordinating all three major cell cycle transitions; first by repression of *cdc25* (*string*) expression, next by activation of Fzr leading to degradation of Cyclin A and Cyclin B thereby blocking mitosis, and lastly by repression of Dacapo which allows Cyclin E/Cdk2 to drive S phase entry.

dMyc was also identified as being transcriptionally upregulated at the mitotic-endocycle transition and, consistent with this, a recent study has shown a role for dMyc in regulation of ovarian follicle cell endoreplication. A role for dMyc in regulating larval endoreplication has also recently been reported. In addition, Archipelago (Ago), the F-box component of the SCF ubiquitin ligase, was found to be essential for Cyclin E degradation and endocycle progression in the follicle epithelium. Ago also plays a role in dMyc degradation (see Sect. 2.4.3), suggesting that ubiquitin-mediated proteolysis via the Ago-SCF may regulate dMyc protein levels in the mitotic-to-endocycle transition.

In summary, during Drosophila development, the canonical DNA-replicative cycle can be modulated by altering expression of key cell cycle regulators. These variations in the cell cycle regulatory modes have not been well studied in tissues known to undergo endoreplication in mammals; however, it is expected that similar types of regulation will occur. Indeed the variation in G1-S phase regulation in mammalian cells is only now being appreciated. The best characterized example to date is the Embryonic Stem (ES) cell cycle. The ES cell cycle is one of the most rapid cycle so far described in mammals with ES cells undergoing cycle in under 6 hours in vitro and for their equivalent population in vivo, the inner cell mass of the blastocyst, under 4.5 hours. This cell cycle is unusual in that it is characterized by extremely high levels of cyclinE/Cdk activity that lack normal cell cycle periodicity, resulting in low activity of Rb-family proteins and the constitutive activation of E2F-dependent gene transcription thus ensuring a rapid G1-S transition. Periodicity of the cell cycle is ensured by regulated activation of cyclin B/cdc2 activity.

3
Summary and Future Directions

We have highlighted here, the circuits of controls required to regulate the transition from G1 into S phase. The cell cycle needs to be tightly regulated to ensure proper duplication and segregation of chromosomes into daughter cells. This is highly important, since loss of cell cycle regulation can lead to genomic instability and promote chromosome loss (aneuploidy), which can result in the development of tumors.

Studies in yeast and mammalian cultured cells have provided a detailed picture of cell cycle control in single cells; however, it is becoming increasingly clear in multicellular organisms that cell cycle regulators can play redundant roles and different regulatory steps are emphasized at different stages of development. Thus, to fully elucidate the myriad of regulatory circuits that occur *in vivo*, and to more clearly understand proliferative disorders, such as cancer, future research needs to focus more on whole animal genetic studies and use of more sophisticated animal models.

Acknowledgments

Helena E. Richardson is supported by an Australian National Health and Medical Research Senior Research Fellowship, Patrick O. Humbert and Leonie M. Quinn by Australian National Health and Medical Research Career Development Fellowships, Julie Secombe by the Lady Tata Memorial Trust for Leukemia research, Anthony M. Brumby by NIH grant #1R21CA098997-01, Nancy Amin and Kathryn M. Doyle by Australian National Health and Medical Research grants and Anabel Herr by an Australian Research Council grant. We thank Dr. David Thomas (Peter MacCallum Cancer Centre) for critical comments on this review.

Bibliography

Books and Reviews

Aladjem, M.I., Fanning, E. (2004) The replicon revisited: an old model learns new tricks in metazoan chromosomes, *EMBO Rep.* **5**, 686–691.

Anderson, J.M. (1996) Cell signalling: MAGUK magic, *Curr. Biol.* **6**, 382–384.

Bell, S.P., Dutta, A. (2002) DNA replication in eukaryotic cells, *Annu. Rev. Biochem.* **71**, 333–374.

Ben-Ze'ev, A. (1997) Cytoskeletal and adhesion proteins as tumor suppressors, *Curr. Opin. Cell Biol.* **9**, 99–108.

Bilder, D. (2004) Epithelial polarity and proliferation control: links from the *Drosophila* neoplastic tumor suppressors, *Genes Dev.* **18**, 1909–1925.

Bracken, A.P., Ciro, M., Cocito, A., Helin, K. (2004) E2F target genes: unraveling the biology, *Trends Biochem. Sci.* **29**, 409–417.

Bregman, D.B., Pestell, R.G., Kidd, V.J. (2000) Cell cycle regulation and RNA polymerase II, *Front. Biosci.* **5**, D244–D257.

Brenner, M. (2000) To be or notch to be, *Nat. Med.* **6**, 1210–1211.

Cam, H., Dynlacht, B.D. (2003) Emerging roles for E2F: beyond the G1/S transition and DNA replication, *Cancer Cell* **3**, 311–316.

Christofori, G., Semb, H. (1999) The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene, *Trends Biochem. Sci.* **24**, 73–76.

Classon, M., Harlow, E. (2002) The retinoblastoma tumour suppressor in development and cancer, *Nat. Rev. Cancer* **2**, 910–917.

Coelho, C.M., Leevers, S.J. (2000) Do growth and cell division rates determine cell size in multicellular organisms? *J. Cell Sci.* **113**(Pt 17), 2927–2934.

Coleman, M.L., Marshall, C.J., Olson, M.F. (2004) RAS and RHO GTPases in G1-phase cell-cycle regulation, *Nat. Rev. Mol. Cell Biol.* **5**, 355–366.

Danen, E.H., Yamada, K.M. (2001) Fibronectin, integrins, and growth control, *J. Cell. Physiol.* **189**, 1–13.

Delattre, M., Gonczy, P. (2004) The arithmetic of centrosome biogenesis, *J. Cell Sci.* **117**, 1619–1630.

Edgar, B.A., Orr-Weaver, T.L. (2001) Endoreplication cell cycles: more for less, *Cell* **105**, 297–306.

Ekholm, S.V., Reed, S.I. (2000) Regulation of G(1) cyclin-dependent kinases in the mammalian cell cycle, *Curr. Opin. Cell Biol.* **12**, 676–684.

Fisher, D., Mechali, M. (2004) Sleeping policemen for DNA replication? *Nat. Cell Biol.* **6**, 576–577.

Freeman, M. (1998) Complexity of EGF receptor signalling revealed in *Drosophila*, *Curr. Opin. Genet. Dev.* **8**, 407–411.

Frolov, M.V., Dyson, N.J. (2004) Molecular mechanisms of E2F-dependent activation and pRB-mediated repression, *J. Cell Sci.* **117**, 2173–2181.

Garriga, J., Grana, X. (2004) Cellular control of gene expression by T-type cyclin/CDK9 complexes, *Gene* **337**, 15–23.

Grandori, C., Cowley, S.M., James, L.P., Eisenman, R.N. (2000) The Myc/Max/Mad network and the transcriptional control of cell behavior, *Annu. Rev. Cell Dev. Biol.* **16**, 653–699.

Harper, J.W., Elledge, S.J. (1998) The role of Cdk7 in CAK function, a retro-retrospective, *Genes Dev.* **12**, 285–289.

Hinchcliffe, E.H., Sluder, G. (2001) It takes two to tango: understanding how centrosome duplication is regulated throughout the cell cycle, *Genes Dev.* **15**, 1167–1181.

Humbert, P., Russell, S., Richardson, H. (2003) Dlg, Scribble and Lgl in cell polarity, cell proliferation and cancer, *BioEssays* **25**, 542–553.

Humbert, P.O., Brumby, A.M., Quinn, L.M., Richardson, H.E. (2004) New tricks for old dogs: unexpected roles for cell cycle regulators revealed using animal models, *Curr. Opin. Cell Biol.* **16**, 614–622.

Hunter, T., Pines, J. (1994) Cyclins and cancer. II: Cyclin D and CDK inhibitors come of age, *Cell* **79**, 573–582.

Ingham, P.W. (1998) Transducing Hedgehog: the story so far, *EMBO J.* **17**, 3505–3511.

Kato, J. (1999) Induction of S phase by G1 regulatory factors, *Front. Biosci.* **4**, D787–D792.

Leclerc, V., Leopold, P. (1996) The cyclin C/Cdk8 kinase, *Prog. Cell Cycle Res.* **2**, 197–204.

Lee, L.A., Orr-Weaver, T.L. (2003) Regulation of cell cycles in *Drosophila* development: intrinsic and extrinsic cues, *Annu. Rev. Genet.* **37**, 545–578.

Liang, J., Slingerland, J.M. (2003) Multiple roles of the PI3K/PKB (Akt) pathway in cell cycle progression, *Cell Cycle* **2**, 339–345.

Lum, L., Beachy, P.A. (2004) The Hedgehog response network: sensors, switches, and routers, *Science* **304**, 1755–1759.

Malumbres, M., Barbacid, M. (2001) To cycle or not to cycle: a critical decision in cancer, *Nat. Rev. Cancer* **1**, 222–231.

Malumbres, M., Carnero, A. (2003) Cell cycle deregulation: a common motif in cancer, *Prog. Cell Cycle Res.* **5**, 5–18.

Malumbres, M., Ortega, S., Barbacid, M. (2000) Genetic analysis of mammalian cyclin-dependent kinases and their inhibitors, *Biol. Chem.* **381**, 827–838.

Massague, J. (2004) G1 cell-cycle control and cancer, *Nature* **432**, 298–306.

McGowan, C.H., Russell, P. (2004) The DNA damage response: sensing and signaling, *Curr. Opin. Cell Biol.* **16**, 629–633.

Mittnacht, S. (1998) Control of pRB phosphorylation, *Curr. Opin. Genet. Dev.* **8**, 21–27.

Mulligan, G., Jacks, T. (1998) The retinoblastoma gene family: cousins with overlapping interests, *Trends Genet.* **14**, 223–229.

Murray, A.W. (2004) Recycling the cell cycle: cyclins revisited, *Cell* **116**, 221–234.

Nollet, F., Berx, G., van Roy, F. (1999) The role of the E-cadherin/catenin adhesion complex in the development and progression of cancer, *Mol. Cell Biol. Res. Commun.* **2**, 77–85.

Nojima, H. (2004) G1 and S-phase checkpoints, chromosome instability, and cancer, *Methods Mol. Biol.* **280**, 3–49.

Pawlak, G., Helfman, D.M. (2001) Cytoskeletal changes in cell transformation and tumorigenesis, *Curr. Opin. Genet. Dev.* **11**, 41–47.

Planas-Silva, M.D., Weinberg, R.A. (1997) The restriction point and control of cell proliferation, *Curr. Opin. Cell Biol.* **9**, 768–772.

Reed, S.I. (2003) Ratchets and clocks: the cell cycle, ubiquitylation and protein turnover, *Nat. Rev. Mol. Cell Biol.* **4**, 855–864.

Ritzi, M., Knippers, R. (2000) Initiation of genome replication: assembly and disassembly of replication-competent chromatin, *Gene* **245**, 13–20.
Sage, J. (2004) Cyclin C makes an entry into the cell cycle, *Dev. Cell* **6**, 607–608.
Sauer, K., Lehner, C.F. (1995) The role of cyclin E in the regulation of entry into S phase, *Prog. Cell Cycle Res.* **1**, 125–139.
Secombe, J., Pierce, S.B., Eisenman, R.N. (2004) Myc: a weapon of mass destruction, *Cell* **117**, 153–156.
Sherr, C.J. (1993) Mammalian G1 cyclins, *Cell* **73**, 1059–1065.
Sherr, C.J., Roberts, J.M. (1999) CDK inhibitors: positive and negative regulators of G1-phase progression, *Genes Dev.* **13**, 1501–1512.
Silver, D.L., Montell, D.J. (2003) A new trick for Cyclin-Cdk: activation of STAT, *Dev. Cell* **4**, 148–149.
Stevaux, O., Dyson, N.J. (2002) A revised picture of the E2F transcriptional network and RB function, *Curr. Opin. Cell Biol.* **14**, 684–691.
Trimarchi, J.M., Lees, J.A. (2002) Sibling rivalry in the E2F family, *Nat. Rev. Mol. Cell Biol.* **3**, 11–20.
Vincan, E. (2004) Frizzled/WNT signalling: the insidious promoter of tumour growth and progression, *Front. Biosci.* **9**, 1023–1034.
Vleminckx, K., Kemler, R. (1999) Cadherins and tissue formation: integrating adhesion and signaling, *BioEssays* **21**, 211–220.
Wanzel, M., Herold, S., Eilers, M. (2003) Transcriptional repression by Myc, *Trends Cell Biol.* **13**, 146–150.
Weinberg, R.A. (1995) The retinoblastoma protein and cell cycle control, *Cell* **81**, 323–330.
Winey, M. (1999) Cell cycle: driving the centrosome cycle, *Curr. Biol.* **9**, R449–R452.
Zhang, H.S., Dean, D.C. (2001) Rb-mediated chromatin structure regulation and transcriptional repression, *Oncogene* **20**, 3134–3138.

Primary Literature

Aggarwal, B.D., Calvi, B.R. (2004) Chromatin regulates origin activity in *Drosophila* follicle cells, *Nature* **430**, 372–376.
Akli, S., Zheng, P.J., Multani, A.S., Wingate, H.F., Pathak, S., Zhang, N., Tucker, S.L., Chang, S., Keyomarsi, K. (2004) Tumor-specific low molecular weight forms of cyclin E induce genomic instability and resistance to p21, p27, and antiestrogens in breast cancer, *Cancer Res.* **64**, 3198–3208.
Albanese, C., Johnson, J., Watanabe, G., Eklund, N., Vu, D., Arnold, A., Pestell, R.G. (1995) Transforming p21ras mutants and c-Ets-2 activate the cyclin D1 promoter through distinguishable regions, *J. Biol. Chem.* **270**, 23589–23597.
Anderson, L.A., Perkins, N.D. (2002) The large subunit of replication factor C interacts with the histone deacetylase, HDAC1, *J. Biol. Chem.* **277**, 29550–29554.
Asano, M., Nevins, J.R., Wharton, R.P. (1996) Ectopic E2F expression induces S phase and apoptosis in *Drosophila* imaginal discs, *Genes Dev.* **10**, 1422–1432.
Aza-Blanc, P., Ramirez-Weber, F.A., Laget, M.P., Schwartz, C., Kornberg, T.B. (1997) Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor, *Cell* **89**, 1043–1053.
Baeg, G.H., Matsumine, A., Kuroda, T., Bhattacharjee, R.N., Miyashiro, I., Toyoshima, K., Akiyama, T. (1995) The tumour suppressor gene product APC blocks cell cycle progression from G0/G1 to S phase, *EMBO J.* **14**, 5618–5625.
Bai, C., Richman, R., Elledge, S.J. (1994) Human cyclin F, *EMBO J.* **13**, 6087–6098.
Baker, N.E., Yu, S.Y. (2001) The EGF receptor defines domains of cell cycle progression and survival to regulate cell number in the developing *Drosophila* eye, *Cell* **104**, 699–708.
Bakiri, L., Lallemand, D., Bossy-Wetzel, E., Yaniv, M. (2000) Cell cycle-dependent variations in c-Jun and JunB phosphorylation: a role in the control of cyclin D1 expression, *EMBO J.* **19**, 2056–2068.
Baonza, A., Casci, T., Freeman, M. (2001) A primary role for the epidermal growth factor receptor in ommatidial spacing in the *Drosophila* eye, *Curr. Biol.* **11**, 396–404.
Beall, E.L., Manak, J.R., Zhou, S., Bell, M., Lipsick, J.S., Botchan, M.R. (2002) Role for a *Drosophila* Myb-containing protein complex in site-specific DNA replication, *Nature* **420**, 833–837.
Beall, E.L., Bell, M., Georlette, D., Botchan, M.R. (2004) Myb mutant lethality in *Drosophila* is dependent upon Mip130: positive and negative regulation of DNA replication, *Genes Dev.* **18**, 1667–1680.
Ben-Ze'ev, A., Shtutman, M., Zhurinsky, J. (2000) The integration of cell adhesion with

gene expression: the role of beta-catenin, *Exp. Cell Res.* **261**, 75–82.

Berthet, C., Aleem, E., Coppola, V., Tessarollo, L., Kaldis, P. (2003) Cdk2 knockout mice are viable, *Curr. Biol.* **13**, 1775–1785.

Betschinger, J., Mechtler, K., Knoblich, J.A. (2003) The Par complex directs asymmetric cell division by phosphorylating the cytoskeletal protein Lgl, *Nature* **422**, 326–330.

Bilder, D., Li, M., Perrimon, N. (2000) Cooperative regulation of cell polarity and growth by *Drosophila* tumor suppressors, *Science* **289**, 113–116.

Baonza, A., Freeman, M. (2005) Control of cell proliferation in the Drosophila eye by notch signaling, *Dev. Cell* **8**, 529–539.

Bortner, D.M., Rosenberg, M.P. (1995) Overexpression of cyclin A in the mammary glands of transgenic mice results in the induction of nuclear abnormalities and increased apoptosis, *Cell Growth Differ.* **6**, 1579–1589.

Bortner, D.M., Rosenberg, M.P. (1997) Induction of mammary gland hyperplasia and carcinomas in transgenic mice expressing human cyclin E, *Mol. Cell. Biol.* **17**, 453–459.

Bosco, G., Du, W., Orr-Weaver, T.L. (2001) DNA replication control through interaction of E2F-RB and the origin recognition complex, *Nat. Cell Biol.* **3**, 289–295.

Brumby, A.M., Richardson, H.E. (2003) Scribble mutants cooperate with oncogenic Ras or Notch to cause neoplastic overgrowth in *Drosophila*, *EMBO J.* **22**, 5769–5779.

Brumby, A., Secombe, J., Horsfield, J., Coombe, M., Amin, N., Coates, D., Saint, R., Richardson, H. (2004) A genetic screen for dominant modifiers of a cyclin E hypomorphic mutation identifies novel regulators of S-phase entry in *Drosophila*, *Genetics* **168**, 227–251.

Burke, R., Basler, K. (1996) Dpp receptors are autonomously required for cell proliferation in the entire developing *Drosophila* wing, *Development* **122**, 2261–2269.

Botz, J., Zerfass-Thome, K., Spitkovsky, D., Delius, H., Vogt, B., Eilers, M., Hatzigeorgiou, A., Jansen-Durr, P. (1996) Cell cycle regulation of the murine cyclin E gene depends on an E2F binding site in the promoter, *Mol. Cell. Biol.* **16**, 3401–3409.

Brugarolas, J., Chandrasekaran, C., Gordon, J.I., Beach, D., Jacks, T., Hannon, G.J. (1995) Radiation-induced cell cycle arrest compromised by p21 deficiency, *Nature* **377**, 552–557.

Calvi, B.R., Lilly, M.A., Spradling, A.C. (1998) Cell cycle control of chorion gene amplification, *Genes Dev.* **12**, 734–744.

Calzada, A., Sanchez, M., Sanchez, E., Bueno, A. (2000) The stability of the Cdc6 protein is regulated by cyclin-dependent kinase/cyclin B complexes in Saccharomyces cerevisiae, *J. Biol. Chem.* **275**, 9734–9741.

Cayirlioglu, P., Bonnette, P.C., Dickson, M.R., Duronio, R.J. (2001) *Drosophila* E2f2 promotes the conversion from genomic DNA replication to gene amplification in ovarian follicle cells, *Development* **128**, 5085–5098.

Chao, J.L., Tsai, Y.C., Chiu, S.J., Sun, Y.H. (2004) Localized Notch signal acts through eyg and upd to promote global growth in *Drosophila* eye, *Development* **131**, 3839–3847

Chen, X. (2002) Cyclin G: a regulator of the p53-Mdm2 network, *Dev. Cell* **2**, 518–519.

Chen, C.R., Kang, Y., Siegel, P.M., Massague, J. (2002a) E2F4/5 and p107 as Smad cofactors linking the TGF beta receptor to c-myc repression, *Cell* **110**, 19–32.

Chen, Z., Indjeian, V.B., McManus, M., Wang, L., Dynlacht, B.D. (2002b) CP110, a cell cycle-dependent CDK substrate, regulates centrosome duplication in human cells, *Dev. Cell* **3**, 339–350.

Chen, C.H., von Kessler, D.P., Park, W., Wang, B., Ma, Y., Beachy, P.A. (1999) Nuclear trafficking of Cubitus interruptus in the transcriptional regulation of Hedgehog target gene expression, *Cell* **98**, 305–316.

Chen, X., Oh, S.W., Zheng, Z., Chen, H.W., Shin, H.H., Hou, S.X. (2003) Cyclin D-Cdk4 and cyclin E-Cdk2 regulate the Jak/STAT signal transduction pathway in *Drosophila*, *Dev. Cell* **4**, 179–190.

Chen, P., Zindy, F., Abdala, C., Liu, F., Li, X., Roussel, M.F., Segil, N. (2003) Progressive hearing loss in mice lacking the cyclin-dependent kinase inhibitor Ink4d, *Nat. Cell Biol.* **5**, 422–426.

Cheng, M., Sexl, V., Sherr, C.J., Roussel, M.F. (1998) Assembly of cyclin D-dependent kinase and titration of p27Kip1 regulated by mitogen-activated protein kinase kinase (MEK1), *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1091–1096.

Cheng, M., Olivier, P., Diehl, J.A., Fero, M., Roussel, M.F., Roberts, J.M., Sherr, C.J. (1999) The p21(Cip1) and p27(Kip1) CDK 'inhibitors' are essential activators of cyclin D-dependent kinases in murine fibroblasts, *EMBO J.* **18**, 1571–1583.

Cheng, T., Rodrigues, N., Shen, H., Yang, Y., Dombkowski, D., Sykes, M., Scadden, D.T. (2000) Hematopoietic stem cell quiescence maintained by p21cip1/waf1, *Science* **287**, 1804–1808.

Ciemerych, M.A., Kenney, A.M., Sicinska, E., Kalaszczynska, I., Bronson, R.T., Rowitch, D.H., Gardner, H., Sicinski, P. (2002) Development of mice expressing a single D-type cyclin, *Genes Dev.* **16**, 3277–3289.

Clarke, A.R., Maandag, E.R., van Roon, M., van der Lugt, N.M., van der Valk, M., Hooper, M.L., Berns, A., te Riele, H. (1992) Requirement for a functional Rb-1 gene in murine development, *Nature* **359**, 328–330.

Cloud, J.E., Rogers, C., Reza, T.L., Ziebold, U., Stone, J.R., Picard, M.H., Caron, A.M., Bronson, R.T., Lees, J.A. (2002) Mutant mouse models reveal the relative roles of E2F1 and E2F3 in vivo, *Mol. Cell. Biol.* **22**, 2663–2672.

Clurman, B.E., Sheaff, R.J., Thress, K., Groudine, M., Roberts, J.M. (1996) Turnover of cyclin E by the ubiquitin-proteasome pathway is regulated by cdk2 binding and cyclin phosphorylation, *Genes Dev.* **10**, 1979–1990.

Cobrinik, D., Lee, M.H., Hannon, G., Mulligan, G., Bronson, R.T., Dyson, N., Harlow, E., Beach, D., Weinberg, R.A., Jacks, T. (1996) Shared role of the pRB-related p130 and p107 proteins in limb development, *Genes Dev.* **10**, 1633–1644.

Coller, H.A., Grandori, C., Tamayo, P., Colbert, T., Lander, E.S., Eisenman, R.N., Golub, T.R. (2000) Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signaling, and adhesion, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3260–3265.

Cope, G.A., Deshaies, R.J. (2003) COP9 signalosome: a multifunctional regulator of SCF and other cullin-based ubiquitin ligases, *Cell* **114**, 663–671.

Cortez, D., Glick, G., Elledge, S.J. (2004) Minichromosome maintenance proteins are direct targets of the ATM and ATR checkpoint kinases, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10078–10083.

Coverley, D., Pelizon, C., Trewick, S., Laskey, R.A. (2000) Chromatin-bound Cdc6 persists in S and G2 phases in human cells, while soluble Cdc6 is destroyed in a cyclin A-cdk2 dependent process, *J. Cell Sci.* **113**(Pt 11), 1929–1938.

Dajee, M., Lazarov, M., Zhang, J.Y., Cai, T., Green, C.L., Russell, A.J., Marinkovich, M.P., Tao, S., Lin, Q., Kubo, Y., Khavari, P.A. (2003) NF-kappaB blockade and oncogenic Ras trigger invasive human epidermal neoplasia, *Nature* **421**, 639–643.

Daksis, J.I., Lu, R.Y., Facchini, L.M., Marhin, W.W., Penn, L.J. (1994) Myc induces cyclin D1 expression in the absence of de novo protein synthesis and links mitogen-stimulated signal transduction to the cell cycle, *Oncogene* **9**, 3635–3645.

Dannenberg, J.H., van Rossum, A., Schuijff, L., te Riele, H. (2000) Ablation of the retinoblastoma gene family deregulates G(1) control causing immortalization and increased cell turnover under growth-restricting conditions, *Genes Dev.* **14**, 3051–3064.

Datar, S.A., Jacobs, H.W., de la Cruz, A.F., Lehner, C.F., Edgar, B.A. (2000) The *Drosophila* cyclin D-Cdk4 complex promotes cellular growth, *EMBO J.* **19**, 4543–4554.

de Bruin, A., Maiti, B., Jakoi, L., Timmers, C., Buerki, R., Leone, G. (2003) Identification and characterization of E2F7, a novel mammalian E2F family member capable of blocking cellular proliferation, *J. Biol. Chem.* **278**, 42041–42049.

de Nooij, J.C., Letendre, M.A., Hariharan, I.K. (1996) A cyclin-dependent kinase inhibitor, Dacapo, is necessary for timely exit from the cell cycle during *Drosophila* embryogenesis, *Cell* **87**, 1237–1247.

Dealy, M.J., Nguyen, K.V., Lo, J., Gstaiger, M., Krek, W., Elson, D., Arbeit, J., Kipreos, E.T., Johnson, R.S. (1999) Loss of Cul1 results in early embryonic lethality and dysregulation of cyclin E, *Nat. Genet.* **23**, 245–248.

Deng, W.M., Althauser, C., Ruohola-Baker, H. (2001) Notch-Delta signaling induces a transition from mitotic cell cycle to endocycle in *Drosophila* follicle cells, *Development* **128**, 4737–4746.

Deng, C., Zhang, P., Harper, J.W., Elledge, S.J., Leder, P. (1995) Mice lacking p21CIP1/WAF1 undergo normal development, but are defective in G1 checkpoint control, *Cell* **82**, 675–684.

Di Stefano, L., Jensen, M.R., Helin, K. (2003) E2F7, a novel E2F featuring DP-independent repression of a subset of E2F-regulated genes, *EMBO J.* **22**, 6289–6298.

Dimberg, A., Oberg, F. (2003) Retinoic acid-induced cell cycle arrest of human myeloid cell lines, *Leuk. Lymphoma* **44**, 1641–1650.

Dimova, D.K., Stevaux, O., Frolov, M.V., Dyson, N.J. (2003) Cell cycle-dependent and cell cycle-independent control of transcription by the *Drosophila* E2F/RB pathway, *Genes Dev.* **17**, 2308–2320.

Dominguez, M., de Celis, J.F. (1998) A dorsal/ventral boundary established by Notch controls growth and polarity in the *Drosophila* eye, *Nature* **396**, 276–278.

Dominguez, M., Wasserman, J.D., Freeman, M. (1998) Multiple functions of the EGF receptor in *Drosophila* eye development, *Curr. Biol.* **8**, 1039–1048.

Dominguez, M., Ferres-Marco, D., Gutierrez-Avino, F.J., Speicher, S.A., Beneyto, M. (2004) Growth and specification of the eye are controlled independently by Eyegone and Eyeless in *Drosophila* melanogaster, *Nat. Genet.* **36**, 31–39.

Donovan, P.J., Reed, S.I. (2003) Germline exclusion of Cks1 in the mouse reveals a metaphase I role for Cks proteins in male and female meiosis, *Cell Cycle* **2**, 275–276.

Doronkin, S., Djagaeva, I., Beckendorf, S.K. (2003) The COP9 signalosome promotes degradation of Cyclin E during early *Drosophila* oogenesis, *Dev. Cell* **4**, 699–710.

Du, W. (2000) Suppression of the rbf null mutants by a de2f1 allele that lacks transactivation domain, *Development* **127**, 367–379.

Du, W., Dyson, N. (1999) The role of RBF in the introduction of G1 regulation during *Drosophila* embryogenesis, *EMBO J.* **18**, 916–925.

Du, W., Xie, J.E., Dyson, N. (1996) Ectopic expression of dE2F and dDP induces cell proliferation and death in the *Drosophila* eye, *EMBO J.* **15**, 3684–3692.

Duman-Scheel, M., Johnston, L.A., Du, W. (2004) Repression of dMyc expression by wingless promotes Rbf-induced G1 arrest in the presumptive *Drosophila* wing margin, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3857–3862.

Duman-Scheel, M., Weng, L., Xin, S., Du, W. (2002) Hedgehog regulates cell growth and proliferation by inducing Cyclin D and Cyclin E, *Nature* **417**, 299–304.

Duronio, R.J., Bonnette, P.C., O'Farrell, P.H. (1998) Mutations of the *Drosophila* dDP, dE2F, and cyclin E genes reveal distinct roles for the E2F-DP transcription factor and cyclin E during the G1-S transition, *Mol. Cell. Biol.* **18**, 141–151.

Duronio, R.J., O'Farrell, P.H., Xie, J.E., Brook, A., Dyson, N. (1995) The transcription factor E2F is required for S phase during *Drosophila* embryogenesis, *Genes Dev.* **9**, 1445–1455.

Elsasser, S., Chi, Y., Yang, P., Campbell, J.L. (1999) Phosphorylation controls timing of Cdc6p destruction: A biochemical analysis, *Mol. Biol. Cell* **10**, 3263–3277.

Emmerich, J., Meyer, C.A., de la Cruz, A.F., Edgar, B.A., Lehner, C.F. (2004) Cyclin d does not provide essential cdk4-independent functions in *Drosophila*, *Genetics* **168**, 867–875.

Evans, T., Rosenthal, E.T., Youngblom, J., Distel, D., Hunt, T. (1983) Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division, *Cell* **33**, 389–396.

Ewen, M.E., Sluss, H.K., Whitehouse, L.L., Livingston, D.M. (1993b) TGF beta inhibition of Cdk4 synthesis is linked to cell cycle arrest, *Cell* **74**, 1009–1020.

Ewen, M.E., Sluss, H.K., Sherr, C.J., Matsushime, H., Kato, J., Livingston, D.M. (1993a) Functional interactions of the retinoblastoma protein with mammalian D-type cyclins, *Cell* **73**, 487–497.

Fero, M.L., Rivkin, M., Tasch, M., Porter, P., Carow, C.E., Firpo, E., Polyak, K., Tsai, L.H., Broudy, V., Perlmutter, R.M., Kaushansky, K., Roberts, J.M. (1996) A syndrome of multiorgan hyperplasia with features of gigantism, tumorigenesis, and female sterility in p27(Kip1)-deficient mice, *Cell* **85**, 733–744.

Field, S.J., Tsai, F.Y., Kuo, F., Zubiaga, A.M., Kaelin, W.G., Livingston, D.M., Orkin, S.H., Greenberg, M.E. Jr. (1996) E2F-1 functions in mice to promote apoptosis and suppress proliferation, *Cell* **85**, 549–561.

Firth, L.C., Baker N.E. (2005) Extracellular signals responsible for spatially regulated proliferation in the differentiating Drosophila eye, *Dev. Cell.* **8**, 541–551.

Franklin, D.S., Godfrey, V.L., O'Brien, D.A., Deng, C., Xiong, Y. (2000) Functional collaboration between different cyclin-dependent kinase inhibitors suppresses tumor growth with distinct tissue specificity, *Mol. Cell. Biol.* **20**, 6147–6158.

Franklin, D.S., Godfrey, V.L., Lee, H., Kovalev, G.I., Schoonhoven, R., Chen-Kiang, S., Su, L., Xiong, Y. (1998) CDK inhibitors

p18(INK4c) and p27(Kip1) mediate two separate pathways to collaboratively suppress pituitary tumorigenesis, *Genes Dev.* **12**, 2899–2911.

Frolov, M.V., Huen, D.S., Stevaux, O., Dimova, D., Balczarek-Strang, K., Elsdon, M., Dyson, N.J. (2001) Functional antagonism between E2F family members, *Genes Dev.* **15**, 2146–2160.

Frolov, M.V., Stevaux, O., Moon, N.S., Dimova, D., Kwon, E.J., Morris, E.J., Dyson, N.J. (2003) G1 cyclin-dependent kinases are insufficient to reverse dE2F2-mediated repression, *Genes Dev.* **17**, 723–728.

Fryer, C.J., White, J.B., Jones, K.A. (2004) Mastermind recruits CycC:CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover, *Mol. Cell* **16**, 509–520.

Fujita, M., Yamada, C., Goto, H., Yokoyama, N., Kuzushima, K., Inagaki, M., Tsurumi, T. (1999) Cell cycle regulation of human CDC6 protein. Intracellular localization, interaction with the human mcm complex, and CDC2 kinase-mediated hyperphosphorylation, *J. Biol. Chem.* **274**, 25927–25932.

Furstenthal, L., Kaiser, B.K., Swanson, C., Jackson, P.K. (2001a) Cyclin E uses Cdc6 as a Chromatin-associated receptor required for DNA replication, *J. Cell Biol.* **152**, 1267–1278.

Furstenthal, L., Swanson, C., Kaiser, B.K., Eldridge, A.G., Jackson, P.K. (2001b) Triggering ubiquitination of a CDK inhibitor at origins of DNA replication, *Nat. Cell. Biol.* **3**, 715–722.

Gaubatz, S., Lindeman, G.J., Ishida, S., Jakoi, L., Nevins, J.R., Livingston, D.M., Rempel, R.E. (2000) E2F4 and E2F5 play an essential role in pocket protein-mediated G1 control, *Mol. Cell* **6**, 729–735.

Geng, Y., Eaton, E.N., Picon, M., Roberts, J.M., Lundberg, A.S., Gifford, A., Sardet, C., Weinberg, R.A. (1996) Regulation of cyclin E transcription by E2Fs and retinoblastoma protein, *Oncogene* **12**, 1173–1180.

Geng, Y., Yu, Q., Sicinska, E., Das, M., Schneider, J.E., Bhattacharya, S., Rideout, W.M., Bronson, R.T., Gardner, H., Sicinski, P. (2003) Cyclin E ablation in the mouse, *Cell* **114**, 431–443.

Giangrande, P.H., Zhu, W., Schlisio, S., Sun, X., Mori, S., Gaubatz, S., Nevins, J.R. (2004) A role for E2F6 in distinguishing G1/S- and G2/M-specific transcription, *Genes Dev.* **18**, 2941–2951.

Hateboer, G., Wobst, A., Petersen, B.O., Le Cam, L., Vigo, E., Sardet, C., Helin, K. (1998) Cell cycle-regulated expression of mammalian CDC6 is dependent on E2F, *Mol. Cell. Biol.* **18**, 6679–6697.

Hazlehurst, L.A., Damiano, J.S., Buyuksal, I., Pledger, W.J., Dalton, W.S. (2000) Adhesion to fibronectin via beta1 integrins regulates p27kip1 levels and contributes to cell adhesion mediated drug resistance (CAM-DR), *Oncogene* **19**, 4319–4327.

He, T.C., Sparks, A.B., Rago, C., Hermeking, H., Zawel, L., da Costa, L.T., Morin, P.J., Vogelstein, B., Kinzler, K.W. (1998) Identification of c-MYC as a target of the APC pathway, *Science* **281**, 1509–1512.

Heikkila, R., Schwab, G., Wickstrom, E., Loke, S.L., Pluznik, D.H., Watt, R., Neckers, L.M. (1987) A c-myc antisense oligodeoxynucleotide inhibits entry into S phase but not progress from G0 to G1, *Nature* **328**, 445–449.

Henderson, B.R. (2000) Nuclear-cytoplasmic shuttling of APC regulates beta-catenin subcellular localization and turnover, *Nat. Cell Biol.* **2**, 653–660.

Herber, B., Truss, M., Beato, M., Muller, R. (1994) Inducible regulatory elements in the human cyclin D1 promoter, *Oncogene* **9**, 2105–2107.

Herbig, U., Griffith, J.W., Fanning, E. (2000) Mutation of cyclin/cdk phosphorylation sites in HsCdc6 disrupts a late step in initiation of DNA replication in human cells, *Mol. Biol. Cell* **11**, 4117–4130.

Heriche, J.K., Ang, D., Bier, E., O'Farrell, P.H. (2003) Involvement of an SCFSlmb complex in timely elimination of E2F upon initiation of DNA replication in *Drosophila*, *BMC Genet.* **4**, 9.

Hinchcliffe, E.H., Li, C., Thompson, E.A., Maller, J.L., Sluder, G. (1999) Requirement of Cdk2-cyclin E activity for repeated centrosome reproduction in Xenopus egg extracts, *Science* **283**, 851–854.

Hinds, P.W., Mittnacht, S., Dulic, V., Arnold, A., Reed, S.I., Weinberg, R.A. (1992) Regulation of retinoblastoma protein functions by ectopic expression of human cyclins, *Cell* **70**, 993–1006.

Hong, A., Lee-Kong, S., Iida, T., Sugimura, I., Lilly, M.A. (2003) The p27cip/kip ortholog dacapo maintains the *Drosophila* oocyte in prophase of meiosis I, *Development* **130**, 1235–1242.

Horsfield, J., Penton, A., Secombe, J., Hoffman, F.M., Richardson, H. (1998) decapentaplegic is required for arrest in G1 phase during *Drosophila* eye development, *Development* **125**, 5069–5078.

Hough, C.D., Woods, D.F., Park, S., Bryant, P.J. (1997) Organizing a functional junctional complex requires specific domains of the *Drosophila* MAGUK Discs large, *Genes Dev.* **11**, 3242–3253.

Hubalek, M.M., Widschwendter, A., Erdel, M., Gschwendtner, A., Fiegl, H.M., Muller, H.M., Goebel, G., Mueller-Holzner, E., Marth, C., Spruck, C.H., Reed, S.I., Widschwendter, M. (2004) Cyclin E dysregulation and chromosomal instability in endometrial cancer, *Oncogene* **23**, 4187–4192.

Hubbard, E.J., Wu, G., Kitajewski, J., Greenwald, I. (1997) sel-10, a negative regulator of lin-12 activity in Caenorhabditis elegans, encodes a member of the CDC4 family of proteins, *Genes Dev.* **11**, 3182–3193.

Huet, X., Rech, J., Plet, A., Vie, A., Blanchard, J.M. (1996) Cyclin A expression is under negative transcriptional control during the cell cycle, *Mol. Cell. Biol.* **16**, 3789–3798.

Humbert, P.O., Verona, R., Trimarchi, J.M., Rogers, C., Dandapani, S., Lees, J.A. (2000b) E2f3 is critical for normal cellular proliferation, *Genes Dev.* **14**, 690–703.

Humbert, P.O., Rogers, C., Ganiatsas, S., Landsberg, R.L., Trimarchi, J.M., Dandapani, S., Brugnara, C., Erdman, S., Schrenzel, M., Bronson, R.T., Lees, J.A. (2000a) E2F4 is essential for normal erythrocyte maturation and neonatal viability, *Mol. Cell* **6**, 281–291.

Ishidate, T., Matsumine, A., Toyoshima, K., Akiyama, T. (2000) The APC-hDLG complex negatively regulates cell cycle progression from the G0/G1 to S phase, *Oncogene* **19**, 365–372.

Ito, N., Rubin, G.M. (1999) Gigas, a *Drosophila* homolog of tuberous sclerosis gene product-2, regulates the cell cycle, *Cell* **96**, 529–539.

Jacks, T., Fazeli, A., Schmitt, E.M., Bronson, R.T., Goodell, M.A., Weinberg, R.A. (1992) Effects of an Rb mutation in the mouse, *Nature* **359**, 295–300.

Johnson, R.T., Rao, P.N. (1970) Mammalian cell fusion: induction of premature chromosome condensation in interphase nuclei, *Nature* **226**, 717–722.

Johnston, L.A., Edgar, B.A. (1998) Wingless and Notch regulate cell-cycle arrest in the developing Drosophila wing, *Nature* **394**, 82–84.

Johnston, L.A., Gallant, P. (2002) Control of growth and organ size in *Drosophila*, *BioEssays* **24**, 54–64.

Johnston, L.A., Sanders, A.L. (2003) Wingless promotes cell survival but constrains growth during *Drosophila* wing development, *Nat. Cell Biol.* **5**, 827–833.

Johnston, L.A., Prober, D.A., Edgar, B.A., Eisenman, R.N., Gallant, P. (1999) *Drosophila* myc regulates cellular growth during development, *Cell* **98**, 779–790.

Jones, L., Richardson, H., Saint, R. (2000) Tissue-specific regulation of cyclin E transcription during *Drosophila* melanogaster embryogenesis, *Development* **127**, 4619–4630.

Kanaoka, Y., Kimura, S.H., Okazaki, I., Ikeda, M., Nojima, H. (1997) GAK: a cyclin G associated kinase contains a tensin/auxilin-like domain, *FEBS Lett.* **402**, 73–80.

Karlseder, J., Rotheneder, H., Wintersberger, E. (1996) Interaction of Sp1 with the growth- and cell cycle-regulated transcription factor E2F, *Mol. Cell. Biol.* **16**, 1659–1667.

Kato, J., Matsushime, H., Hiebert, S.W., Ewen, M.E., Sherr, C.J. (1993) Direct binding of cyclin D to the retinoblastoma gene product (pRb) and pRb phosphorylation by the cyclin D-dependent kinase CDK4, *Genes Dev.* **7**, 331–342.

Kennedy, B.K., Barbie, D.A., Classon, M., Dyson, N., Harlow, E. (2000) Nuclear organization of DNA replication in primary mammalian cells, *Genes Dev.* **14**, 2855–2868.

Keyomarsi, K., Tucker, S.L., Buchholz, T.A., Callister, M., Ding, Y., Hortobagyi, G.N., Bedrosian, I., Knickerbocker, C., Toyofuku, W., Lowe, M., Herliczek, T.W., Bacus, S.S. (2002) Cyclin E and survival in patients with breast cancer, *N. Engl. J. Med.* **347**, 1566–1575.

Kitagawa, M., Higashi, H., Jung, H.K., Suzuki-Takahashi, I., Ikeda, M., Tamai, K., Kato, J., Segawa, K., Yoshida, E., Nishimura, S., Taya, Y. (1996) The consensus motif for phosphorylation by cyclin D1-Cdk4 is different from that for phosphorylation by cyclin A/E-Cdk2, *EMBO J.* **15**, 7060–7069.

Kiyokawa, H., Kineman, R.D., Manova-Todorova, K.O., Soares, V.C., Hoffman, E.S., Ono, M., Khanam, D., Hayday, A.C., Frohman, L.A., Koff, A. (1996) Enhanced growth of mice lacking the cyclin-dependent kinase

inhibitor function of p27(Kip1), *Cell* **85**, 721–732.

Klezovitch, O., Fernandez, T.E., Tapscott, S.J., Vasioukhin, V. (2004) Loss of cell polarity causes severe brain dysplasia in Lgl1 knockout mice, *Genes Dev.* **18**, 559–571.

Klippel, A., Escobedo, M.A., Wachowicz, M.S., Apell, G., Brown, T.W., Giedlin, M.A., Kavanaugh, W.M., Williams, L.T. (1998) Activation of phosphatidylinositol 3-kinase is sufficient for cell cycle entry and promotes cellular changes characteristic of oncogenic transformation, *Mol. Cell. Biol.* **18**, 5699–5711.

Knoblich, J.A., Sauer, K., Jones, L., Richardson, H., Saint, R., Lehner, C.F. (1994) Cyclin E controls S phase progression and its downregulation during *Drosophila* embryogenesis is required for the arrest of cell proliferation, *Cell* **77**, 107–120.

Knudsen, E.S., Buckmaster, C., Chen, T.T., Feramisco, J.R., Wang, J.Y. (1998) Inhibition of DNA synthesis by RB: effects on G1/S transition and S-phase progression, *Genes Dev.* **12**, 2278–2292.

Koff, A., Ohtsuki, M., Polyak, K., Roberts, J.M., Massague, J. (1993) Negative regulation of G1 in mammalian cells: inhibition of cyclin E-dependent kinase by TGF-beta, *Science* **260**, 536–539.

Kohn, M.J., Bronson, R.T., Harlow, E., Dyson, N.J., Yamasaki, L. (2003) Dp1 is required for extra-embryonic development, *Development* **130**, 1295–1305.

Korenjak, M., Taylor-Harding, B., Binne, U.K., Satterlee, J.S., Stevaux, O., Aasland, R., White-Cooper, H., Dyson, N., Brehm, A. (2004) Native E2F/RBF complexes contain Myb-interacting proteins and repress transcription of developmentally controlled E2F target genes, *Cell* **119**, 181–193.

Korsisaari, N., Rossi, D.J., Paetau, A., Charnay, P., Henkemeyer, M., Makela, T.P. (2002) Conditional ablation of the Mat1 subunit of TFIIH in Schwann cells provides evidence that Mat1 is not required for general transcription, *J. Cell Sci.* **115**, 4275–4284.

Kozar, K., Ciemerych, M.A., Rebel, V.I., Shigematsu, H., Zagozdzon, A., Sicinska, E., Geng, Y., Yu, Q., Bhattacharya, S., Bronson, R.T., Akashi, K., Sicinski, P. (2004) Mouse development and cell proliferation in the absence of D-cyclins.

Krek, W., Xu, G., Livingston, D.M. (1995) Cyclin A-kinase regulation of E2F-1 DNA binding function underlies suppression of an S phase checkpoint, *Cell* **83**, 1149–1158.

Lacey, K.R., Jackson, P.K., Stearns, T. (1999) Cyclin-dependent kinase control of centrosome duplication, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2817–2822.

Lai, A., Kennedy, B.K., Barbie, D.A., Bertos, N.R., Yang, X.J., Theberge, M.C., Tsai, S.C., Seto, E., Zhang, Y., Kuzmichev, A., Lane, W.S., Reinberg, D., Harlow, E., Branton, P.E. (2001) RBP1 recruits the mSIN3-histone deacetylase complex to the pocket of retinoblastoma tumor suppressor family proteins found in limited discrete regions of the nucleus at growth arrest, *Mol. Cell. Biol.* **21**, 2918–2932.

Lane, M.E., Sauer, K., Wallace, K., Jan, Y.N., Lehner, C.F., Vaessin, H. (1996) Dacapo, a cyclin-dependent kinase inhibitor, stops cell proliferation during *Drosophila* development, *Cell* **87**, 1225–1235.

Lane, M.E., Elend, M., Heidmann, D., Herr, A., Marzodko, S., Herzig, A., Lehner, C.F. (2000) A screen for modifiers of cyclin E function in *Drosophila* melanogaster identifies Cdk2 mutations, revealing the insignificance of putative phosphorylation sites in Cdk2, *Genetics* **155**, 233–244.

Larochelle, S., Pandur, J., Fisher, R.P., Salz, H.K., Suter, B. (1998) Cdk7 is essential for mitosis and for in vivo Cdk-activating kinase activity, *Genes Dev.* **12**, 370–381.

Larochelle, S., Chen, J., Knights, R., Pandur, J., Morcillo, P., Erdjument-Bromage, H., Tempst, P., Suter, B., Fisher, R.P. (2001) T-loop phosphorylation stabilizes the CDK7-cyclin H-MAT1 complex in vivo and regulates its CTD kinase activity, *EMBO J.* **20**, 3749–3759.

Latres, E., Malumbres, M., Sotillo, R., Martin, J., Ortega, S., Martin-Caballero, J., Flores, J.M., Cordon-Cardo, C., Barbacid, M. (2000) Limited overlapping roles of P15(INK4b) and P18(INK4c) cell cycle inhibitors in proliferation and tumorigenesis, *EMBO J.* **19**, 3496–3506.

Leclerc, V., Tassan, J.P., O'Farrell, P.H., Nigg, E.A., Leopold, P. (1996) *Drosophila* Cdk8, a kinase partner of cyclin C that interacts with the large subunit of RNA polymerase II, *Mol. Biol. Cell* **7**, 505–513.

Lee J.D., Treisman J.E. (2001) The role of Wingless signaling in establishing the

anteroposterior and dorsoventral axes of the eye disc, *Development* **128**, 1519–1529.

Lee, E.Y., Chang, C.Y., Hu, N., Wang, Y.C., Lai, C.C., Herrup, K., Lee, W.H., Bradley, A. (1992) Mice deficient for Rb are nonviable and show defects in neurogenesis and haematopoiesis [see comments], *Nature* **359**, 288–294.

Lee, M.H., Williams, B.O., Mulligan, G., Mukai, S., Bronson, R.T., Dyson, N., Harlow, E., Jacks, T. (1996) Targeted disruption of p107: functional overlap between p107 and Rb, *Genes Dev.* **10**, 1621–1632.

Li, Q.J., Pazdera, T.M., Minden, J.S. (1999) *Drosophila* embryonic pattern repair: how embryos respond to cyclin E-induced ectopic division, *Development* **126**, 2299–2307.

Li, F.X., Zhu, J.W., Hogan, C.J., DeGregori, J. (2003a) Defective gene expression, S phase progression, and maturation during hematopoiesis in E2F1/E2F2 mutant mice, *Mol. Cell. Biol.* **23**, 3607–3622.

Li, F.X., Zhu, J.W., Tessem, J.S., Beilke, J., Varella-Garcia, M., Jensen, J., Hogan, C.J., DeGregori, J. (2003b) The development of diabetes in E2f1/E2f2 mutant mice reveals important roles for bone marrow-derived cells in preventing islet cell loss, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12935–12940.

Li, X., Zhao, Q., Liao, R., Sun, P., Wu, X. (2003b) The SCF(Skp2) ubiquitin ligase complex interacts with the human replication licensing factor Cdt1 and regulates Cdt1 degradation, *J. Biol. Chem.* **278**, 30854–30858.

Li, H.S., Wang, D., Shen, Q., Schonemann, M.D., Gorski, J.A., Jones, K.R., Temple, S., Jan, L.Y., Jan, Y.N. (2003a) Inactivation of Numb and Numblike in embryonic dorsal forebrain impairs neurogenesis and disrupts cortical morphogenesis, *Neuron* **40**, 1105–1118.

Liao, C., Wang, X.Y., Wei, H.Q., Li, S.Q., Merghoub, T., Pandolfi, P.P., Wolgemuth, D.J. (2001) Altered myelopoiesis and the development of acute myeloid leukemia in transgenic mice overexpressing cyclin A1, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6853–6858.

Lilly, M.A., Spradling, A.C. (1996) The *Drosophila* endocycle is controlled by Cyclin E and lacks a checkpoint ensuring S-phase completion, *Genes Dev.* **10**, 2514–2526.

Lindeman, G.J., Dagnino, L., Gaubatz, S., Xu, Y., Bronson, R.T., Warren, H.B., Livingston, D.M. (1998) A specific, nonproliferative role for E2F-5 in choroid plexus function revealed by gene targeting, *Genes Dev.* **12**, 1092–1098.

Liu, H., Dibling, B., Spike, B., Dirlam, A., Macleod, K. (2004) New roles for the RB tumor suppressor protein, *Curr. Opin. Genet. Dev.* **14**, 55–64.

Logan, N., Delavaine, L., Graham, A., Reilly, C., Wilson, J., Brummelkamp, T.R., Hijmans, E.M., Bernards, R., La Thangue, N.B. (2004) E2F-7: a distinctive E2F family member with an unusual organization of DNA-binding domains, *Oncogene* **23**, 5138–5150.

Lukas, J., Bartkova, J., Rohde, M., Strauss, M., Bartek, J. (1995) Cyclin D1 is dispensable for G1 control in retinoblastoma gene-deficient cells independently of cdk4 activity, *Mol. Cell. Biol.* **15**, 2600–2611.

Lum, L., Zhang, C., Oh, S., Mann, R.K., von Kessler, D.P., Taipale, J., Weis-Garcia, F., Gong, R., Wang, B., Beachy, P.A. (2003) Hedgehog signal transduction via Smoothened association with a cytoplasmic complex scaffolded by the atypical kinesin, Costal-2, *Mol. Cell* **12**, 1261–1274.

Lundberg, A.S., Weinberg, R.A. (1998) Functional inactivation of the retinoblastoma protein requires sequential modification by at least two distinct cyclin-cdk complexes, *Mol. Cell. Biol.* **18**, 753–761.

Luo, R.X., Postigo, A.A., Dean, D.C. (1998) Rb interacts with histone deacetylase to repress transcription, *Cell* **92**, 463–473.

Lykke-Andersen, K., Schaefer, L., Menon, S., Deng, X.W., Miller, J.B., Wei, N. (2003) Disruption of the COP9 signalosome Csn2 subunit in mice causes deficient cell proliferation, accumulation of p53 and cyclin E, and early embryonic death, *Mol. Cell. Biol.* **23**, 6790–6797.

Maines, J.Z., Stevens, L.M., Tong, X., Stein, D. (2004) *Drosophila* dMyc is required for ovary cell growth and endoreplication, *Development* **131**, 775–786.

Maiorano, D., Moreau, J., Mechali, M. (2000) XCDT1 is required for the assembly of pre-replicative complexes in Xenopus laevis, *Nature* **404**, 622–625.

Malumbres, M., Sotillo, R., Santamaria, D., Galan, J., Cerezo, A., Ortega, S., Dubus, P., Barbacid, M. (2004) Mammalian cells cycle without the D-type cyclin-dependent kinases Cdk4 and Cdk6, *Cell* **118**, 493–504.

Mantel, C., Braun, S.E., Reid, S., Henegariu, O., Liu, L., Hangoc, G., Broxmeyer, H.E. (1999)

p21(cip-1/waf-1) deficiency causes deformed nuclear architecture, centriole overduplication, polyploidy, and relaxed microtubule damage checkpoints in human hematopoietic cells, *Blood* **93**, 1390–1398.

Marheineke, K., Hyrien, O. (2004) Control of replication origin density and firing time in Xenopus egg extracts: role of a caffeine-sensitive, ATR-dependent checkpoint, *J. Biol. Chem.* **279**, 28071–28081.

Masui, Y., Markert, C.L. (1971) Cytoplasmic control of nuclear behavior during meiotic maturation of frog oocytes, *J. Exp. Zool.* **177**, 129–145.

Matsumoto, Y., Maller, J.L. (2004) A centrosomal localization signal in cyclin E required for Cdk2-independent S phase entry, *Science* **306**, 885–888.

Matsumoto, Y., Hayashi, K., Nishida, E. (1999) Cyclin-dependent kinase 2 (Cdk2) is required for centrosome duplication in mammalian cells, *Curr. Biol.* **9**, 429–432.

McGarry, T.J., Kirschner, M.W. (1998) Geminin, an inhibitor of DNA replication, is degraded during mitosis, *Cell* **93**, 1043–1053.

Mendez, J., Stillman, B. (2000) Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis, *Mol. Cell. Biol.* **20**, 8602–8612.

Methot, N., Basler, K. (1999) Hedgehog controls limb development by regulating the activities of distinct transcriptional activator and repressor forms of Cubitus interruptus, *Cell* **96**, 819–831.

Methot, N., Basler, K. (2000) Suppressor of fused opposes hedgehog signal transduction by impeding nuclear accumulation of the activator form of Cubitus interruptus, *Development* **127**, 4001–4010.

Methot, N., Basler, K. (2001) An absolute requirement for Cubitus interruptus in Hedgehog signaling, *Development* **128**, 733–742.

Mettouchi, A., Klein, S., Guo, W., Lopez-Lago, M., Lemichez, E., Westwick, J.K., Giancotti, F.G. (2001) Integrin-specific activation of Rac controls progression through the G(1) phase of the cell cycle, *Mol. Cell* **8**, 115–127.

Meyer, C.A., Jacobs, H.W., Lehner, C.F. (2002) Cyclin D-cdk4 is not a master regulator of cell multiplication in *Drosophila* embryos, *Curr. Biol.* **12**, 661–666.

Meyer, C.A., Jacobs, H.W., Datar, S.A., Du, W., Edgar, B.A., Lehner, C.F. (2000) *Drosophila* Cdk4 is required for normal growth and is dispensable for cell cycle progression, *EMBO J.* **19**, 4533–4542.

Milutinovic, S., Zhuang, Q., Szyf, M. (2002) Proliferating cell nuclear antigen associates with histone deacetylase activity, integrating DNA replication and chromatin modification, *J. Biol. Chem.* **277**, 20974–20978.

Mimura, S., Takisawa, H. (1998) Xenopus Cdc45-dependent loading of DNA polymerase alpha onto chromatin under the control of S-phase Cdk, *EMBO J.* **17**, 5699–5707.

Moberg, K.H., Bell, D.W., Wahrer, D.C., Haber, D.A., Hariharan, I.K. (2001) Archipelago regulates Cyclin E levels in *Drosophila* and is mutated in human cancer cell lines, *Nature* **413**, 311–316.

Moberg, K.H., Mukherjee, A., Veraksa, A., Artavanis-Tsakonas, S., Hariharan, I.K. (2004) The *Drosophila* F box protein archipelago regulates dMyc protein levels in vivo, *Curr. Biol.* **14**, 965–974.

Moore, J.D., Kirk, J.A., Hunt, T. (2003) Unmasking the S-phase-promoting potential of cyclin B1, *Science* **300**, 987–990.

Moroy, T., Geisen, C. (2004) Cyclin E, *Int. J. Biochem. Cell Biol.* **36**, 1424–1439.

Morris, L., Allen, K.E., La Thangue, N.B. (2000) Regulation of E2F transcription by cyclin E-Cdk2 kinase mediated through p300/CBP co-activators, *Nat. Cell Biol* **2**, 232–239.

Morris, M.C., Kaiser, P., Rudyak, S., Baskerville, C., Watson, M.H., Reed, S.I. (2003) Cks1-dependent proteasome recruitment and activation of CDC20 transcription in budding yeast, *Nature* **423**, 1009–1013.

Murdoch, J.N., Henderson, D.J., Doudney, K., Gaston-Massuet, C., Phillips, H.M., Paternotte, C., Arkell, R., Stanier, P., Copp, A.J. (2003) Disruption of scribble (Scrb1) causes severe neural tube defects in the circletail mouse, *Hum. Mol. Genet.* **12**, 87–98.

Murga, M., Fernandez-Capetillo, O., Field, S.J., Moreno, B., Borlado, L.R., Fujiwara, Y., Balomenos, D., Vicario, A., Carrera, A.C., Orkin, S.H., Greenberg, M.E., Zubiaga, A.M. (2001) Mutation of E2F2 in mice causes enhanced T lymphocyte proliferation, leading to the development of autoimmunity, *Immunity* **15**, 959–970.

Murphy, T.D. (2003) *Drosophila* skpA, a component of SCF ubiquitin ligases, regulates centrosome duplication independently of cyclin E accumulation, *J. Cell Sci.* **116**, 2321–2332.

Nakayama, K., Ishida, N., Shirane, M., Inomata, A., Inoue, T., Shishido, N., Horii, I., Loh, D.Y. (1996) Mice lacking p27(Kip1) display increased body size, multiple organ hyperplasia, retinal dysplasia, and pituitary tumors, *Cell* **85**, 707–720.

Nateri, A.S., Riera-Sans, L., Da Costa, C., Behrens, A. (2004) The ubiquitin ligase SCFFbw7 antagonizes apoptotic JNK signaling, *Science* **303**, 1374–1378.

Nguyen, M.L., Nguyen, M.M., Lee, D., Griep, A.E., Lambert, P.F. (2003a) The PDZ ligand domain of the human papillomavirus type 16 E6 protein is required for E6's induction of epithelial hyperplasia in vivo, *J. Virol.* **77**, 6957–6964.

Nguyen, M.M., Nguyen, M.L., Caruana, G., Bernstein, A., Lambert, P.F., Griep, A.E. (2003b) Requirement of PDZ-containing proteins for cell cycle regulation and differentiation in the mouse lens epithelium, *Mol. Cell. Biol.* **23**, 8970–8981.

Nielsen, S.J., Schneider, R., Bauer, U.M., Bannister, A.J., Morrison, A., O'Carroll, D., Firestein, R., Cleary, M., Jenuwein, T., Herrera, R.E., Kouzarides, T. (2001) Rb targets histone H3 methylation and HP1 to promoters, *Nature* **412**, 561–565.

Nishitani, H., Lygerou, Z., Nishimoto, T., Nurse, P. (2000) The Cdt1 protein is required to license DNA for replication in fission yeast, *Nature* **404**, 625–628.

Nougarede, R., Della Seta, F., Zarzov, P., Schwob, E. (2000) Hierarchy of S-phase-promoting factors: yeast Dbf4-Cdc7 kinase requires prior S-phase cyclin-dependent kinase activation, *Mol. Cell. Biol.* **20**, 3795–3806.

Ogawa, H., Ishiguro, K., Gaubatz, S., Livingston, D.M., Nakatani, Y. (2002) A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells, *Science* **296**, 1132–1136.

Ogris, E., Rotheneder, H., Mudrak, I., Pichler, A., Wintersberger, E. (1993) A binding site for transcription factor E2F is a target for trans activation of murine thymidine kinase by polyomavirus large T antigen and plays an important role in growth regulation of the gene, *J. Virol.* **67**, 1765–1771.

Ohtani, K., DeGregori, J., Nevins, J.R. (1995) Regulation of the cyclin E gene by transcription factor E2F1, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 12146–12150.

Ohtani, K., Tsujimoto, A., Ikeda, M., Nakamura, M. (1998) Regulation of cell growth-dependent expression of mammalian CDC6 gene by the cell cycle transcription factor E2F, *Oncogene* **17**, 1777–1785.

Ohtsubo, M., Theodoras, A.M., Schumacher, J., Roberts, J.M., Pagano, M. (1995) Human cyclin E, a nuclear protein essential for the G1-to-S phase transition, *Mol. Cell. Biol.* **15**, 2612–2624.

Okuda, M. (2002) The role of nucleophosmin in centrosome duplication, *Oncogene* **21**, 6170–6174.

Okuda, M., Horn, H.F., Tarapore, P., Tokuyama, Y., Smulian, A.G., Chan, P.K., Knudsen, E.S., Hofmann, I.A., Snyder, J.D., Bove, K.E., Fukasawa, K. (2000) Nucleophosmin/B23 is a target of CDK2/cyclin E in centrosome duplication, *Cell* **103**, 127–140.

Oldham, S., Montagne, J., Radimerski, T., Thomas, G., Hafen, E. (2000) Genetic and biochemical characterization of dTOR, the *Drosophila* homolog of the target of rapamycin, *Genes Dev.* **14**, 2689–2694.

Orian, A., van Steensel, B., Delrow, J., Bussemaker, H.J., Li, L., Sawado, T., Williams, E., Loo, L.W., Cowley, S.M., Yost, C., Pierce, S., Edgar, B.A., Parkhurst, S.M., Eisenman, R.N. (2003) Genomic binding by the *Drosophila* Myc, Max, Mad/Mnt transcription factor network, *Genes Dev.* **17**, 1101–1114.

Ortega, S., Prieto, I., Odajima, J., Martin, A., Dubus, P., Sotillo, R., Barbero, J.L., Malumbres, M., Barbacid, M. (2003) Cyclin-dependent kinase 2 is essential for meiosis but not for mitotic cell division in mice, *Nat. Genet.* **35**, 25–31.

Pagano, M., Tam, S.W., Theodoras, A.M., Beer-Romero, P., Del Sal, G., Chau, V., Yew, P.R., Draetta, G.F., Rolfe, M. (1995) Role of the ubiquitin-proteasome pathway in regulating abundance of the cyclin-dependent kinase inhibitor p27, *Science* **269**, 682–685.

Parisi, T., Beck, A.R., Rougier, N., McNeil, T., Lucian, L., Werb, Z., Amati, B. (2003) Cyclins E1 and E2 are required for endoreplication in placental trophoblast giant cells, *EMBO J.* **22**, 4794–4803.

Pearson, B.E., Nasheuer, H.P., Wang, T.S. (1991) Human DNA polymerase alpha gene: sequences controlling expression in cycling and serum-stimulated cells, *Mol. Cell. Biol.* **11**, 2081–2095.

Peeper, D.S., Upton, T.M., Ladha, M.H., Neuman, E., Zalvide, J., Bernards, R., DeCaprio, J.A., Ewen, M.E. (1997) Ras signalling linked to the cell-cycle machinery by the retinoblastoma protein, *Nature* **386**, 177–181.

Pelizon, C., Madine, M.A., Romanowski, P., Laskey, R.A. (2000) Unphosphorylatable mutants of Cdc6 disrupt its nuclear export but still support DNA replication once per cell cycle, *Genes Dev.* **14**, 2526–2533.

Perkins, G., Diffley, J.F. (1998) Nucleotide-dependent prereplicative complex assembly by Cdc6p, a homolog of eukaryotic and prokaryotic clamp-loaders, *Mol. Cell* **2**, 23–32.

Petersen, B.O., Lukas, J., Sorensen, C.S., Bartek, J., Helin, K. (1999) Phosphorylation of mammalian CDC6 by cyclin A/CDK2 regulates its subcellular localization, *EMBO J.* **18**, 396–410.

Pierce, S.B., Yost, C., Britton, J.S., Loo, L.W., Flynn, E.M., Edgar, B.A., Eisenman, R.N. (2004) dMyc is required for larval growth and endoreplication in *Drosophila*, *Development* **131**, 2317–2327.

Plant, P.J., Fawcett, J.P., Lin, D.C., Holdorf, A.D., Binns, K., Kulkarni, S., Pawson, T. (2003) A polarity complex of mPar-6 and atypical PKC binds, phosphorylates and regulates mammalian Lgl, *Nat. Cell Biol.* **5**, 301–308.

Posakony, L.G., Raftery, L.A., Gelbart, W.M. (1990) Wing formation in *Drosophila* melanogaster requires decapentaplegic gene function along the anterior-posterior compartment boundary, *Mech. Dev.* **33**, 69–82.

Prober, D.A., Edgar, B.A. (2000) Ras1 promotes cellular growth in the *Drosophila* wing, *Cell* **100**, 435–446.

Prokhorova, T.A., Mowrer, K., Gilbert, C.H., Walter, J.C. (2003) DNA replication of mitotic chromatin in Xenopus egg extracts, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13241–13246.

Quinn, L.M., Herr, A., McGarry, T.J., Richardson, H. (2001) The *Drosophila* Geminin homolog: roles for Geminin in limiting DNA replication, in anaphase and in neurogenesis, *Genes Dev.* **15**, 2741–2754.

Rao, P.N., Johnson, R.T. (1970) Mammalian cell fusion: studies on the regulation of DNA synthesis and mitosis, *Nature* **225**, 159–164.

Rayman, J.B., Takahashi, Y., Indjeian, V.B., Dannenberg, J.H., Catchpole, S., Watson, R.J., te Riele, H., Dynlacht, B.D. (2002) E2F mediates cell cycle-dependent transcriptional repression in vivo by recruitment of an HDAC1/mSin3B corepressor complex, *Genes Dev.* **16**, 933–947.

Reis, T., Edgar, B.A. (2004) Negative regulation of dE2F1 by cyclin-dependent kinases controls cell cycle timing, *Cell* **117**, 253–264.

Rempel, R.E., Saenz-Robles, M.T., Storms, R., Morham, S., Ishida, S., Engel, A., Jakoi, L., Melhem, M.F., Pipas, J.M., Smith, C., Nevins, J.R. (2000) Loss of E2F4 activity leads to abnormal development of multiple cellular lineages, *Mol. Cell* **6**, 293–306.

Ren, S., Rollins, B.J. (2004) Cyclin C/cdk3 promotes Rb-dependent G0 exit, *Cell* **117**, 239–251.

Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R.A., Dynlacht, B.D. (2002) E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints, *Genes Dev.* **16**, 245–256.

Reynard, G.J., Reynolds, W., Verma, R., Deshaies, R.J. (2000) Cks1 is required for G(1) cyclin-cyclin-dependent kinase activity in budding yeast, *Mol. Cell. Biol.* **20**, 5858–5864.

Rialland, M., Sola, F., Santocanale, C. (2002) Essential role of human CDT1 in DNA replication and chromatin licensing, *J. Cell Sci.* **115**, 1435–1440.

Richardson, H.E., O'Keefe, L.V., Reed, S.I., Saint, R. (1993) A *Drosophila* G1-specific cyclin E homolog exhibits different modes of expression during embryogenesis, *Development* **119**, 673–690.

Richardson, H., O'Keefe, L.V., Marty, T., Saint, R. (1995) Ectopic cyclin E expression induces premature entry into S phase and disrupts pattern formation in the *Drosophila* eye imaginal disc, *Development* **121**, 3371–3379.

Robanus-Maandag, E., Dekker, M., van der Valk, M., Carrozza, M.L., Jeanny, J.C., Dannenberg, J.H., Berns, A., te Riele, H. (1998) p107 is a suppressor of retinoblastoma development in pRb-deficient mice, *Genes Dev.* **12**, 1599–1609.

Robbins, D.J., Nybakken, K.E., Kobayashi, R., Sisson, J.C., Bishop, J.M., Therond, P.P. (1997) Hedgehog elicits signal transduction by means of a large complex containing

the kinesin-related protein costal2, *Cell* **90**, 225–234.

Rolls, M.M., Albertson, R., Shih, H.P., Lee, C.Y., Doe, C.Q. (2003) *Drosophila* aPKC regulates cell polarity and cell proliferation in neuroblasts and epithelia, *J. Cell Biol.* **163**, 1089–1098.

Ronchini, C., Capobianco, A.J. (2001) Induction of cyclin D1 transcription and CDK2 activity by Notch(ic): implication for cell cycle disruption in transformation by Notch(ic), *Mol. Cell. Biol.* **21**, 5925–5934.

Rossi, D.J., Londesborough, A., Korsisaari, N., Pihlak, A., Lehtonen, E., Henkemeyer, M., Makela, T.P. (2001) Inability to enter S phase and defective RNA polymerase II CTD phosphorylation in mice lacking Mat1, *EMBO J.* **20**, 2844–2856.

Royzman, I., Whittaker, A.J., Orr-Weaver, T.L. (1997) Mutations in *Drosophila* DP and E2F distinguish G1-S progression from an associated transcriptional program, *Genes Dev.* **11**, 1999–2011.

Royzman, I., Austin, R.J., Bosco, G., Bell, S.P., Orr-Weaver, T.L. (1999) ORC localization in *Drosophila* follicle cells and the effects of mutations in dE2F and dDP, *Genes Dev.* **13**, 827–840.

Royzman, I., Hayashi-Hagihara, A., Dej, K.J., Bosco, G., Lee, J.Y., Orr-Weaver, T.L. (2002) The E2F cell cycle regulator is required for *Drosophila* nurse cell DNA replication and apoptosis, *Mech. Dev.* **119**, 225–237.

Ruiz, S., Segrelles, C., Bravo, A., Santos, M., Perez, P., Leis, H., Jorcano, J.L., Paramio, J.M. (2003) Abnormal epidermal differentiation and impaired epithelial-mesenchymal tissue interactions in mice lacking the retinoblastoma relatives p107 and p130, *Development* **130**, 2341–2353.

Sage, J., Mulligan, G.J., Attardi, L.D., Miller, A., Chen, S., Williams, B., Theodorou, E., Jacks, T. (2000) Targeted disruption of the three Rb-related genes leads to loss of G(1) control and immortalization, *Genes Dev.* **14**, 3037–3050.

Sander, E.E., ten Klooster, J.P., van Delft, S., van der Kammen, R.A., Collard, J.G. (1999) Rac downregulates Rho activity: reciprocal balance between both GTPases determines cellular morphology and migratory behavior, *J. Cell Biol.* **147**, 1009–1022.

Sauer, K., Knoblich, J.A., Richardson, H., Lehner, C.F. (1995) Distinct modes of cyclin E/cdc2c kinase regulation and S-phase control in mitotic and endoreduplication cycles of *Drosophila* embryogenesis, *Genes Dev.* **9**, 1327–1339.

Schaeffer, V., Althauser, C., Shcherbata, H.R., Deng, W.M., Ruohola-Baker, H. (2004) Notch-dependent Fizzy-related/Hec1/Cdh1 expression is required for the mitotic-to-endocycle transition in *Drosophila* follicle cells, *Curr. Biol.* **14**, 630–636.

Scheper, G.C., Morrice, N.A., Kleijn, M., Proud, C.G. (2001) The mitogen-activated protein kinase signal-integrating kinase Mnk2 is a eukaryotic initiation factor 4E kinase with high levels of basal activity in mammalian cells, *Mol. Cell. Biol.* **21**, 743–754.

Schraml, P., Bucher, C., Bissig, H., Nocito, A., Haas, P., Wilber, K., Seelig, S., Kononen, J., Mihatsch, M.J., Dirnhofer, S., Sauter, G. (2003) Cyclin E overexpression and amplification in human tumours, *J. Pathol.* **200**, 375–382.

Schubeler, D., Scalzo, D., Kooperberg, C., van Steensel, B., Delrow, J., Groudine, M. (2002) Genome-wide DNA replication profile for *Drosophila* melanogaster: a link between transcription and replication timing, *Nat. Genet.* **32**, 438–442.

Schulze, A., Zerfass, K., Spitkovsky, D., Middendorp, S., Berges, J., Helin, K., Jansen-Durr, P., Henglein, B. (1995) Cell cycle regulation of the cyclin A gene promoter is mediated by a variant E2F site, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11264–11268.

Schwartz, B.E., Larochelle, S., Suter, B., Lis, J.T. (2003) Cdk7 is required for full activation of *Drosophila* heat shock genes and RNA polymerase II phosphorylation in vivo, *Mol. Cell. Biol.* **23**, 6876–6886.

Sears, R., Leone, G., DeGregori, J., Nevins, J.R. (1999) Ras enhances Myc protein stability, *Mol. Cell.* **3**, 169–179.

Sears, R., Nuckolls, F., Haura, E., Taya, Y., Tamai, K., Nevins, J.R. (2000) Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability, *Genes Dev.* **14**, 2501–2514.

Seoane, J., Le, H.V., Massague, J. (2002) Myc suppression of the p21(Cip1) Cdk inhibitor influences the outcome of the p53 response to DNA damage, *Nature* **419**, 729–734.

Seoane, J., Le, H.V., Shen, L., Anderson, S.A., Massague, J. (2004) Integration of Smad and forkhead pathways in the control

of neuroepithelial and glioblastoma cell proliferation, *Cell* **117**, 211–223.

Shanahan, F., Seghezzi, W., Parry, D., Mahony, D., Lees, E. (1999) Cyclin E Associates with BAF155 and BRG1, components of the mammalian SWI-SNF complex, and alters the ability of BRG1 to induce growth arrest, *Mol. Cell. Biol.* **19**, 1460–1469.

Sharpless, N.E., Bardeesy, N., Lee, K.H., Carrasco, D., Castrillon, D.H., Aguirre, A.J., Wu, E.A., Horner, J.W., DePinho, R.A. (2001) Loss of p16Ink4a with retention of p19Arf predisposes mice to tumorigenesis, *Nature* **413**, 86–91.

Shcherbata, H.R., Althauser, C., Findley, S.D., Ruohola-Baker, H. (2004a) The mitotic-to-endocycle switch in *Drosophila* follicle cells is executed by Notch-dependent regulation of G1/S, G2/M and M/G1 cell-cycle transitions, *Development* **131**, 3169–3181.

Shcherbata, H.R., Althauser, C., Findley, S.D., Ruohola-Baker, H. (2004b) The mitotic-to-endocycle switch in *Drosophila* follicle cells is executed by Notch-dependent regulation of G1/S, G2/M and M/G1 cell-cycle transitions, *Development* **131**, 3169–3181.

Shechter, D., Costanzo, V., Gautier, J. (2004a) ATR and ATM regulate the timing of DNA replication origin firing, *Nat. Cell Biol.* **6**, 648–655.

Shechter, D., Ying, C.Y., Gautier, J. (2004b) DNA unwinding is an Mcm complex-dependent and ATP hydrolysis-dependent process, *J. Biol. Chem.* **279**, 45586–45593.

Sicinska, E., Aifantis, I., Le Cam, L., Swat, W., Borowski, C., Yu, Q., Ferrando, A.A., Levin, S.D., Geng, Y., von Boehmer, H., Sicinski, P. (2003) Requirement for cyclin D3 in lymphocyte development and T cell leukemias, *Cancer Cell* **4**, 451–461.

Skowyra, D., Craig, K.L., Tyers, M., Elledge, S.J., Harper, J.W. (1997) F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex, *Cell* **91**, 209–219.

Spike, B.T., Dirlam, A., Dibling, B.C., Marvin, J., Williams, B.O., Jacks, T., Macleod, K.F. (2004) The Rb tumor suppressor is required for stress erythropoiesis, *EMBO J.* **23**, 4319–4329.

Spruck, C.H., Won, K.A., Reed, S.I. (1999) Deregulated cyclin E induces chromosome instability, *Nature* **401**, 297–300.

Spruck, C., Strohmaier, H., Watson, M., Smith, A.P., Ryan, A., Krek, T.W., Reed, S.I. (2001) A CDK-independent function of mammalian Cks1: targeting of SCF(Skp2) to the CDK inhibitor p27Kip1, *Mol. Cell* **7**, 639–650.

Staller, P., Peukert, K., Kiermaier, A., Seoane, J., Lukas, J., Karsunky, H., Moroy, T., Bartek, J., Massague, J., Hanel, F., Eilers, M. (2001) Repression of p15INK4b expression by Myc through association with Miz-1, *Nat. Cell Biol.* **3**, 392–399.

Stead, E., White, J., Faast, R., Conn, S., Goldstone, S., Rathjen, J., Dhingra, U., Rathjen, P., Walker, D., Dalton, S. (2002) Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities, *Oncogene* **21**, 8320–8333.

Sterner, J.M., Dew-Knight, S., Musahl, C., Kornbluth, S., Horowitz, J.M. (1998) Negative regulation of DNA replication by the retinoblastoma protein is mediated by its association with MCM7, *Mol. Cell. Biol.* **18**, 2748–2757.

Stevaux, O., Dimova, D., Frolov, M.V., Taylor-Harding, B., Morris, E., Dyson, N. (2002) Distinct mechanisms of E2F regulation by *Drosophila* RBF1 and RBF2, *EMBO J.* **21**, 4927–4937.

Storre, J., Elsasser, H.P., Fuchs, M., Ullmann, D., Livingston, D.M., Gaubatz, S. (2002) Homeotic transformations of the axial skeleton that accompany a targeted deletion of E2f6, *EMBO Rep.* **3**, 695–700.

Strausfeld, U.P., Howell, M., Descombes, P., Chevalier, S., Rempel, R.E., Adamczewski, J., Maller, J.L., Hunt, T., Blow, J.J. (1996) Both cyclin A and cyclin E have S-phase promoting (SPF) activity in Xenopus egg extracts, *J. Cell Sci.* **109**(Pt 6), 1555–1563.

Tada, S., Li, A., Maiorano, D., Mechali, M., Blow, J.J. (2001) Repression of origin assembly in metaphase depends on inhibition of RLF-B/Cdt1 by geminin, *Nat. Cell Biol.* **3**, 107–113.

Taipale, J., Cooper, M.K., Maiti, T., Beachy, P.A. (2002) Patched acts catalytically to suppress the activity of Smoothened, *Nature* **418**, 892–897.

Takahashi, Y., Rayman, J.B., Dynlacht, B.D. (2000) Analysis of promoter binding by the E2F and pRB families in vivo: distinct E2F proteins mediate activation and repression, *Genes Dev.* **14**, 804–816.

Tam, S.W., Theodoras, A.M., Shay, J.W., Draetta, G.F., Pagano, M. (1994) Differential expression and regulation of Cyclin D1 protein in normal and tumor human cells: association with Cdk4 is required for Cyclin D1 function in G1 progression, *Oncogene* **9**, 2663–2674.

Tao, Y., Kassatly, R.F., Cress, W.D., Horowitz, J.M. (1997) Subunit composition determines E2F DNA-binding site specificity, *Mol. Cell. Biol.* **17**, 6994–7007.

Tapon, N., Ito, N., Dickson, B.J., Treisman, J.E., Hariharan, I.K. (2001) The *Drosophila* tuberous sclerosis complex gene homologs restrict cell growth and cell proliferation, *Cell* **105**, 345–355.

Tarapore, P., Okuda, M., Fukasawa, K. (2002) A mammalian in vitro centriole duplication system: evidence for involvement of CDK2/cyclin E and nucleophosmin/B23 in centrosome duplication, *Cell Cycle* **1**, 75–81.

Tedesco, D., Lukas, J., Reed, S.I. (2002) The pRb-related protein p130 is regulated by phosphorylation-dependent proteolysis via the protein-ubiquitin ligase SCF(Skp2), *Genes Dev.* **16**, 2946–2957.

Tetsu, O., McCormick, F. (1999) Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells, *Nature* **398**, 422–426.

Tetzlaff, M.T., Bai, C., Finegold, M., Wilson, J., Harper, J.W., Mahon, K.A., Elledge, S.J. (2004a) Cyclin F disruption compromises placental development and affects normal cell cycle execution, *Mol. Cell. Biol.* **24**, 2487–2498.

Tetzlaff, M.T., Yu, W., Li, M., Zhang, P., Finegold, M., Mahon, K., Harper, J.W., Schwartz, R.J., Elledge, S.J. (2004b) Defective cardiovascular development and elevated cyclin E and Notch proteins in mice lacking the Fbw7 F-box protein, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3338–3345.

Thomer, M., May, N.R., Aggarwal, B.D., Kwok, G., Calvi, B.R. (2004) *Drosophila* double-parked is sufficient to induce re-replication during development and is regulated by cyclin E/CDK2, *Development* **131**, 4807–4818.

Thommes, P., Kubota, Y., Takisawa, H., Blow, J.J. (1997) The RLF-M component of the replication licensing system forms complexes containing all six MCM/P1 polypeptides, *EMBO J.* **16**, 3312–3319.

Tinkle, C.L., Lechler, T., Pasolli, H.A., Fuchs, E. (2004) Conditional targeting of E-cadherin in skin: insights into hyperproliferative and degenerative responses, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 552–557.

Tsunematsu, R., Nakayama, K., Oike, Y., Nishiyama, M., Ishida, N., Hatakeyama, S., Bessho, Y., Kageyama, R., Suda, T., Nakayama, K.I. (2004) Mouse Fbw7/Sel-10/Cdc4 is required for notch degradation during vascular development, *J. Biol. Chem.* **279**, 9417–9423.

Vasioukhin, V., Bauer, C., Degenstein, L., Wise, B., Fuchs, E. (2001) Hyperproliferation and defects in epithelial polarity upon conditional ablation of alpha-catenin in skin, *Cell* **104**, 605–617.

Vermeer, P.D., Einwalter, L.A., Moninger, T.O., Rokhlina, T., Kern, J.A., Zabner, J., Welsh, M.J. (2003) Segregation of receptor and ligand regulates activation of epithelial growth factor receptor, *Nature* **422**, 322–326.

Wada, M., Miyazawa, H., Wang, R.S., Mizuno, T., Sato, A., Asashima, M., Hanaoka, F. (2002) The second largest subunit of mouse DNA polymerase epsilon, DPE2, interacts with SAP18 and recruits the Sin3 co-repressor protein to DNA, *J. Biochem. (Tokyo)* **131**, 307–311.

Walter, J.C. (2000) Evidence for sequential action of cdc7 and cdk2 protein kinases during initiation of DNA replication in Xenopus egg extracts, *J. Biol. Chem.* **275**, 39773–39778.

Walter, J., Newport, J. (2000) Initiation of eukaryotic DNA replication: origin unwinding and sequential chromatin association of Cdc45, RPA, and DNA polymerase alpha, *Mol. Cell* **5**, 617–627.

Wang, G., Jiang, J. (2004) Multiple Cos2/Ci interactions regulate Ci subcellular localization through microtubule dependent and independent mechanisms, *Dev. Biol.* **268**, 493–505.

Wang, J., Chenivesse, X., Henglein, B., Brechot, C. (1990) Hepatitis B virus integration in a cyclin A gene in a hepatocellular carcinoma, *Nature* **343**, 555–557.

Wang, Y., Penfold, S., Tang, X., Hattori, N., Riley, P., Harper, J.W., Cross, J.C., Tyers, M. (1999) Deletion of the Cul1 gene in mice causes arrest in early embryogenesis and accumulation of cyclin E, *Curr. Biol.* **9**, 1191–1194.

Welcker, M., Orian, A., Grim, J.A., Eisenman, R.N., Clurman, B.E. (2004a) A nucleolar isoform of the Fbw7 ubiquitin ligase regulates c-Myc and cell size, *Curr. Biol.* **14**, 1852–1857.

Welcker, M., Orian, A., Jin, J., Grim, J.A., Harper, J.W., Eisenman, R.N., Clurman, B.E. (2004b) The Fbw7 tumor suppressor regulates glycogen synthase kinase 3 phosphorylation-dependent c-Myc protein degradation, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9085–9090.

Welcker, M., Singer, J., Loeb, K.R., Grim, J., Bloecher, A., Gurien-West, M., Clurman, B.E.,

Roberts, J.M. (2003) Multisite phosphorylation by Cdk2 and GSK3 controls cyclin E degradation, *Mol. Cell* **12**, 381–392.

Whittaker, A.J., Royzman, I., Orr-Weaver, T.L. (2000) *Drosophila* double parked: a conserved, essential replication protein that colocalizes with the origin recognition complex and links DNA replication with mitosis and the down-regulation of S phase transcripts, *Genes Dev.* **14**, 1765–1776.

Wingate, H., Bedrosian, I., Akli, S., Keyomarsi, K. (2003) The low molecular weight (LMW) isoforms of cyclin E deregulate the cell cycle of mammary epithelial cells, *Cell Cycle* **2**, 461–466.

Winston, J.T., Chu, C., Harper, J.W. (1999) Culprits in the degradation of cyclin E apprehended, *Genes Dev.* **13**, 2751–2757.

Wohlschlegel, J.A., Dhar, S.K., Prokhorova, T.A., Dutta, A., Walter, J.C. (2002) Xenopus Mcm10 binds to origins of DNA replication after Mcm2-7 and stimulates origin binding of Cdc45, *Mol. Cell* **9**, 233–240.

Wohlschlegel, J.A., Dwyer, B.T., Dhar, S.K., Cvetic, C., Walter, J.C., Dutta, A. (2000) Inhibition of eukaryotic DNA replication by geminin binding to Cdt1, *Science* **290**, 2309–2312.

Won, K.A., Reed, S.I. (1996) Activation of cyclin E/CDK2 is coupled to site-specific autophosphorylation and ubiquitin-dependent degradation of cyclin E, *EMBO J.* **15**, 4182–4193.

Woods, D.F., Hough, C.D., Peel, D., Callaini, G., Bryant, P.J. (1996) Dlg protein is required for junction structure, cell polarity, and proliferation control in *Drosophila* epithelia, *J. Cell Biol.* **134**, 1469–1482.

Wu, L., Timmers, C., Maiti, B., Saavedra, H.I., Sang, L., Chong, G.T., Nuckolls, F., Giangrande, P., Wright, F.A., Field, S.J., Greenberg, M.E., Orkin, S., Nevins, J.R., Robinson, M.L., Leone, G. (2001) The E2F1-3 transcription factors are essential for cellular proliferation, *Nature* **414**, 457–462.

Wu, L., de Bruin, A., Saavedra, H.I., Starovic, M., Trimboli, A., Yang, Y., Opavska, J., Wilson, P., Thompson, J.C., Ostrowski, M.C., Rosol, T.J., Woollett, L.A., Weinstein, M., Cross, J.C., Robinson, M.L., Leone, G. (2003) Extra-embryonic function of Rb is essential for embryonic development and viability, *Nature* **421**, 942–947.

Xin, S., Weng, L., Xu, J., Du, W. (2002) The role of RBF in developmentally regulated cell proliferation in the eye disc and in Cyclin D/Cdk4 induced cellular growth, *Development* **129**, 1345–1356.

Yada, M., Hatakeyama, S., Kamura, T., Nishiyama, M., Tsunematsu, R., Imaki, H., Ishida, N., Okumura, F., Nakayama, K., Nakayama, K.I. (2004) Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7, *EMBO J.* **23**, 2116–2125.

Yamasaki, L., Jacks, T., Bronson, R., Goillot, E., Harlow, E., Dyson, N.J. (1996) Tumor induction and tissue atrophy in mice lacking E2F-1, *Cell* **85**, 537–548.

Yan, J., Walz, K., Nakamura, H., Carattini-Rivera, S., Zhao, Q., Vogel, H., Wei, N., Justice, M.J., Bradley, A., Lupski, J.R. (2003) COP9 signalosome subunit 3 is essential for maintenance of cell proliferation in the mouse embryonic epiblast, *Mol. Cell. Biol.* **23**, 6798–6808.

Yasmeen, A., Berdel, W.E., Serve, H., Muller-Tidow, C. (2003) E- and A-type cyclins as markers for cancer diagnosis and prognosis, *Expert Rev. Mol. Diagn.* **3**, 617–633.

Ye, X., Zhu, C., Harper, J.W. (2001) A premature-termination mutation in the Mus musculus cyclin-dependent kinase 3 gene, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1682–1686.

Yeh, E., Cunningham, M., Arnold, H., Chasse, D., Monteith, T., Ivaldi, G., Hahn, W.C., Stukenberg, P.T., Shenolikar, S., Uchida, T., Counter, C.M., Nevins, J.R., Means, A.R., Sears, R. (2004) A signalling pathway controlling c-Myc degradation that impacts oncogenic transformation of human cells, *Nat. Cell Biol.* **6**, 308–318.

Zarkowska, T., Mittnacht, S. (1997) Differential phosphorylation of the retinoblastoma protein by G1/S cyclin-dependent kinases, *J. Biol. Chem.* **272**, 12738–12746.

Zarkowska, T., Harlow, E., Mittnacht, S. (1997) Monoclonal antibodies specific for underphosphorylated retinoblastoma protein identify a cell cycle regulated phosphorylation site targeted by CDKs, *Oncogene* **14**, 249–254.

Zeitler, J., Hsu, C.P., Dionne, H., Bilder, D. (2004) Domains controlling cell polarity and proliferation in the *Drosophila* tumor suppressor Scribble, *J. Cell Biol.* **167**, 1137–1146.

Zhang, L., Gjoerup, O., Roberts, T.M. (2004) The serine/threonine kinase cyclin G-associated

kinase regulates epidermal growth factor receptor signaling, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10296–10301.

Zhang, P., Wong, C., DePinho, R.A., Harper, J.W., Elledge, S.J. (1998) Cooperation between the Cdk inhibitors p27(KIP1) and p57(KIP2) in the control of tissue growth and development, *Genes Dev.* **12**, 3162–3167.

Zhang, P., Wong, C., Liu, D., Finegold, M., Harper, J.W., Elledge, S.J. (1999) p21(CIP1) and p57(KIP2) control muscle differentiation at the myogenin step, *Genes Dev.* **13**, 213–224.

Zhang, H.S., Gavin, M., Dahiya, A., Postigo, A.A., Ma, D., Luo, R.X., Harbour, J.W., Dean, D.C. (2000) Exit from G1 and S phase of the cell cycle is regulated by repressor complexes containing HDAC-Rb-hSWI/SNF and Rb-hSWI/SNF, *Cell* **101**, 79–89.

Zhang, P., Liegeois, N.J., Wong, C., Finegold, M., Hou, H., Thompson, J.C., Silverman, A., Harper, J.W., DePinho, R.A., Elledge, S.J. (1997) Altered cell differentiation and proliferation in mice lacking p57KIP2 indicates a role in Beckwith-Wiedemann syndrome, *Nature* **387**, 151–158.

Zhao, J., Kennedy, B.K., Lawrence, B.D., Barbie, D.A., Matera, A.G., Fletcher, J.A., Harlow, E. (2000) NPAT links cyclin E-Cdk2 to the regulation of replication-dependent histone gene transcription, *Genes Dev.* **14**, 2283–2297.

Zhu, A.J., Zheng, L., Suyama, K., Scott, M.P. (2003) Altered localization of *Drosophila* smoothened protein activates Hedgehog signal transduction, *Genes Dev.* **17**, 1240–1252.

Zindy, F., den Besten, W., Chen, B., Rehg, J.E., Latres, E., Barbacid, M., Pollard, J.W., Sherr, C.J., Cohen, P.E., Roussel, M.F. (2001) Control of spermatogenesis in mice by the cyclin D-dependent kinase inhibitors p18(Ink4c) and p19(Ink4d), *Mol. Cell. Biol.* **21**, 3244–3255.

Zou, L., Stillman, B. (1998) Formation of a preinitiation complex by S-phase cyclin CDK-dependent loading of Cdc45p onto chromatin, *Science* **280**, 593–596.

Zou, L., Stillman, B. (2000) Assembly of a complex containing Cdc45p, replication protein A, and Mcm2p at replication origins controlled by S-phase cyclin-dependent kinases and Cdc7p-Dbf4p kinase, *Mol. Cell. Biol.* **20**, 3086–3096.

Zou, L., Mitchell, J., Stillman, B. (1997) CDC45, a novel yeast gene that functions with the origin recognition complex and Mcm proteins in initiation of DNA replication, *Mol. Cell. Biol.* **17**, 553–563.

Replication and Transcription: *see* DNA Replication and Transcription

Reporter Gene System: Green Fluorescent Protein

Boon Chuan Low and Zhiyuan Gong
National University of Singapore, Singapore

1 The Historic Background of GFP 217
1.1 Jellyfish and GFP 217
1.2 Cloning of GFP cDNA 219

2 Biochemical and Physical Properties of GFP 220
2.1 Physical Characteristics of GFP 220
2.2 Three-dimensional Structure of GFP 220
2.3 Chromophore of GFP 221
2.4 Genetically Modified GFP and its Natural Homologs 223

3 Applications of GFP Family Proteins 227
3.1 GFP as a Reporter of Gene and Protein Expression 228
3.2 GFP-based FRET for Studies of Protein–Protein Interactions and Signal Activation 230
3.3 GFP-fusion Protein for Imaging Cell Dynamics, Subcellular Localization, and Cell Signaling 233
3.4 GFP for Tracing Microbial Gene Expression and Host-pathogen Interactions 234
3.5 GFP Imaging of Tumor Growth, Metastasis and Angiogenesis 236

4 GFP and Transgenic Organisms 236
4.1 Analysis of Gene Promoters by Transient Transgenic Assay 237
4.2 Recapitulation of Gene Expression Programs in Stable Transgenic Lines 237
4.3 Formation of GFP Fusion Proteins 238
4.4 Use of GFP as Transgenic Marker and Labeling of Cells by GFP Expression 239
4.5 Mutant Generation and Characterization in Developmental Analysis 240

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

4.6 Cell Sorting for *in vitro* Culture and Cell Type Specific cDNA Library Construction 240
4.7 Development of Transgenic Systems for Biomonitoring 240
4.8 Transgenic Ornamentals 241

5 Future of GFP 241

Acknowledgment 242

Bibliography 242
Books and Reviews 242
Primary Literature 242

Keywords

Cell Signaling
A group of molecular mechanisms that relay external stimuli to molecular machinery within the cells. This can be achieved via interactions between similar or different classes of proteins, or between proteins, peptides, ions, or even nucleic acids.

Chromophore
A chemical moiety that absorbs light in some, but not all, regions of visible spectrum, thereby creating color appearance. If the absorbed light energy is subsequently reemitted also in the form of light (i.e. fluorescence occurs), the moiety is called fluorophore.

FACS
Fluoroscence-activated cell sorter, a device that separates different cells on the basis of their fluorescence.

Fluorescence
The ability of a chemical to absorb light at one wavelength (the excitation wavelength) and to reemit the light (fluoresce) at another specific but longer wavelength.

FRET
Fluorescence resonance energy transfer, a radiationless transfer of excitation energy from one (donor) fluorophore to another (acceptor) fluorophore. This transfer is possible due to overlap of donor's emission and acceptor's excitation spectra and the close proximity of the two fluorophores.

Transgenic Organisms
Animals or plants that carry artificially introduced gene(s) that are stably integrated into the host genome and passed to future generations.

Green fluorescent protein (GFP) possesses unique biochemical and biophysical properties that make it a popular and powerful tool for both basic and applied research. Originally isolated from jellyfish, its gene can be introduced to various host species and results in a stable protein that can form matured chromophore without the need of other cofactors and enzymes. Various wildtype and mutant forms of GFP have been identified through protein engineering or the search for similar proteins in other species such as reef corals and their relatives. With unique properties, they offer excellent markers for monitoring various biological processes in a single cell or whole organism. These include viewing of the internal structure and protein movement as well as studies of protein–protein interaction. As a transgene, it is also very useful for monitoring developmental process and as a promising source of genetically modified ornamentals. Coupled with advanced bioimaging methods, GFP and its variants are expected to continue to provide new insights into many biological processes.

1 The Historic Background of GFP

1.1 Jellyfish and GFP

During the course of evolution, some species such as bacteria, mushrooms, dinoflagellates, jellyfish, earthworms, squids, fireflies, shrimps, and flashlight fish, have all adopted various forms of bioluminescence to serve certain biological functions that are important for their own survival. Bioluminescence is the process by which visible light is emitted by an organism as a result of chemical reaction after the oxidation of a substrate (luciferin) by an enzyme (luciferase) that usually requires oxygen as the oxidant. Such a mechanism is employed as a part of the organism's strategies in defense (to deter predators), offense (to help in predation), or communications (such as during courtship or mating). Understanding the biochemistries for different classes of such compounds had set the stage for a key revolution in science.

Studies on the bioluminescence of the jellyfish *Aequorea victoria* (also known as *Aequorea aequorea* or *Aequorea forskalea*) dated back to the mid 1950s. Then, Davenport and Nicol reported that the light-producing cells of the jellyfish fluoresced green when the animal was irradiated with ultraviolet light. Five years later, Frank Johnson and Osamu Shimomura developed a protocol to extract the active light emitting particle of the jellyfish, a substance they later named *aequorin*. In their earlier classic work, they reported on the ability of aequorin to emit blue light in the presence of Ca^{2+} and exhibited a broad spectrum with peak at 460 nm. During the purification of aequorin, they noticed the presence of yet another protein that, according to their report, was "..... exhibiting a very bright, greenish fluorescence in the ultraviolet of a Mineralite.... " Such *green protein*, having a sharp spectrum instead and with an emission peak at 508 nm was termed *green fluorescent protein* (GFP). This demonstrated that light organs of the jellyfish contain the aequorin and GFP, perhaps allowing the jellyfish to be more

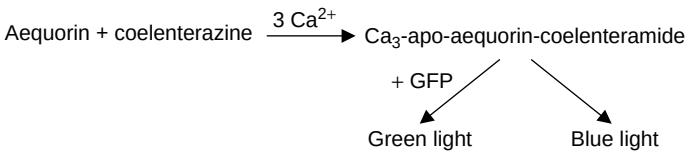

efficient in triggering visual responses as the green corresponds to the region of maximum sensitivity of the eyes of the animals adapted to living in coastal waters. The most interesting findings, then, were the observations that aequorin could exist as a complex with GFP upon binding with Ca^{2+} and the oxidation of its luciferin, coelenterazine, although it itself was not directly bound to the GFP. *In vitro*, such complex of Ca^{2+}-apo-aequorin-coelenteramide emits blue light, but *in vivo*, it is believed that this complex undergoes radiationless energy transfer to GFP that gives off green fluorescence as indicated above:

Subsequently, GFP homologs have been found in numerous organisms and many mutant forms have been generated. Since the GFP from the jellyfish *A. victoria* is best studied at the biochemical and molecular levels, and is also the most commonly used for the tracer studies, it serves as the focus of our discussion. However, due to its diverse forms and applications, we have used the word "GFP" loosely to include the natural form, mutants, or its homologs that may produce different spectra. Since the cloning of the GFP cDNA from *A. victoria* in 1992 and the introduction of GFP cDNA as a living color-reporter gene in both prokaryotic (*Escherichia coli*) and eukaryotic (*Caenorhabditis elegans*) cells in 1994, the scientific research on GFP and its applications have rapidly expanded, as witnessed by the dramatic increase of literature containing the key words *green fluorescent protein* or GFP (Fig. 1). It provides an important tool in many disciplines of life science research in both prokaryotes and eukaryotes – from cellular and molecular biology to neurobiology; from developmental biology and molecular medicine to food industry. Several major breakthroughs are highlighted (Table 1) some of which will be covered in the text below. Coupled with

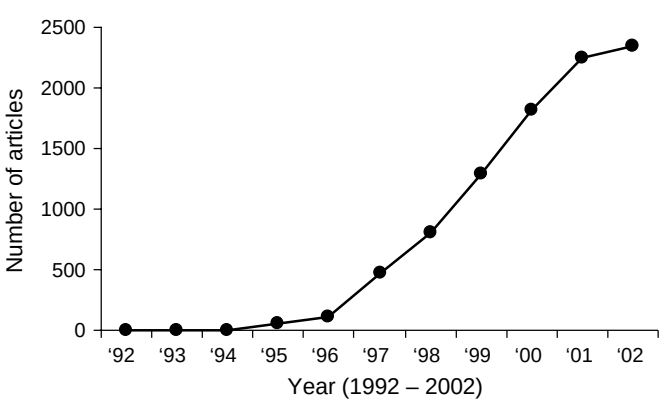

Fig. 1 Number of articles reporting on GFP, retrieved by *green fluorescent protein or GFP* query search in science citation index.

Tab. 1 Chronicle of research on GFP and its applications.

Year	*Events*
1955	Observation that *Aequorea* fluoresced green when irradiated with ultraviolet light
1969	One of the earliest biochemical characterization leading to coining the word "green fluorescent protein"
1979	Chemical structure of chromophore deduced
1992	Cloning of GFP cDNA and deduced GFP primary structure
1993	Chemical structure of GFP chromophore confirmed
1994	GFP as a reporter gene in *E. coli* and *C. elegans*
	Expression of GFP fusion protein in *Drosophila*
1994–1996	Mutations on GFP that generates different spectral properties, including the Blue Fluorescent Protein (BFP), Cyan Fluorescent Protein (CYP) and Yellow Fluorescent Protein (YFP)
1995	GFP transgenic plant, zebrafish and mouse
1996	First crystal structures of wildtype and mutant GFP solved
1997	Use of GFP to recapitulate gene expression patterns in living transgenic zebrafish
1997–1998	FRET-based GFP reporter as biosensors and tools for studies of protein–protein interaction
1998	pH-sensitive mutant "pHluorin" used in visualization of secretion and synaptic transmission;
	GFP knock-in mice
	Destabilized GFP as transcriptional reporter
1999–2001	Discovery of fluorescent proteins from nonbioluminescent corallimorph *Discosoma sp.* (DsFP583/DsRed) and from the sea anemone *Heteractis crispa* with far-red-emitting mutant of nonfluorescent blue chromoprotein (hcriGFP/HcRed); multispectral imaging is feasible. Mutant form of DsFP583 is used as "fluorescent timer."
2002–2003	Development of "photoactivatable probes." Use in live-cells kinetics imaging for signaling metabolites or proteins in single cell or larger population, and for cell lineage studies.

advances in bioimaging methods and the discoveries of more variants with readily distinguishable spectra and improved properties, many more experimental approaches are expected to provide new insights into many biological processes.

1.2 Cloning of GFP cDNA

The first wave of GFP *Green Revolution* was initiated by the cloning of the complementary DNA (cDNA) from *A. victoria* and its deduced primary sequence by Prasher et al.. It encodes a protein of 238 amino acids, with an estimated molecular weight of 27 kDa. This has allowed scientists to carry out extensive molecular and structural biology work to elucidate the structure–function relationship for GFP that laid the foundation for the development of GFP technology. In 1994, Chalfie et al. had introduced the GFP cDNA into heterologous systems and produced a fluorescent product in both bacteria and *C. elegans*. Their findings demonstrated that single polypeptide was sufficient to confer fluorescence without exogenous substrates and cofactors, thus confirming the earlier notion that autocatalytic cyclization of the chromophore formation was involved. More importantly, their results elegantly showed that GFP expression could be used to monitor gene expression and protein

localization in living organisms, which are later extensively used either as a coexpression marker or fusion proteins. This had created tremendous excitement among the scientific community and a "star" was truly born.

2 Biochemical and Physical Properties of GFP

2.1 Physical Characteristics of GFP

We have seen rapid expansion of GFP applications. To provide further improvements by protein engineering, one needs to understand the biochemical and physical properties of the GFP molecule and its chromophore. This will facilitate our hunt for better versions of GFP that promises enhanced contrast in fluorescence detection either by means of changing the biophysical properties or improvement in terms of their gene expression, protein synthesis and folding, or its photostability. This is also important so as to provide a sound basis for the quantitative and qualitative interpretations of data generated in its applications.

Cloning of GFP cDNA and its successful expression as a recombinant protein has allowed production of high amounts of pure form of GFP for extensive biochemical and physical studies, including its structural studies. As expected, most of the biochemical and physical studies were performed on this *Aquorea wildtype* form, *Aequorea victoria* green fluorescent protein (AvGFP) and the sea pansy, *Renilla reniformis* (RrGFP), as well as their mutants, while minimal information can be obtained from 20 other naturally occurring GFPs – usually with nothing more than just the wavelength of peak emission.

All known GFPs are acidic, compact, and globular with monomer molecular weight of 27 kDa. Some, for example, RrGFP, could exist as a nondissociable dimer unless denatured or diluted in aqueous solution whereas AvGFP forms a weak dimer with the dissociation constant of 0.11 mM. They are generally very stable, an obvious advantage as a reporter of gene expression. The core of the protein is highly resistant against proteolysis, but the carboxyl terminus of the AvGFP (His-Gly-Met-Asp-Glu-Tyr-Lys) is quite susceptible to attack by carboxypeptidases. Furthermore, the presence of two basic residues (histidine and lysine) and two acidic residues (aspartic acid and glutamic acid) can cause partial proteolytic cleavage that generates mixture of isoforms.

The T_m (temperature where half of the endogenous fluorescence is lost) for AvGFP and RrGFP is 76 °C and 70 °C respectively. While the fluorescence of RrGFP is unaffected over a wide pH range and is stable at its isoelectric point (pH 5.34), AvGFP is less stable at its isoelectric point (pH 4.7 to 5.1). Furthermore, AvGFP is different from RrGFP when placed under alkaline conditions such that there is a drastic shift in its excitation spectra. Interestingly, there are no significant changes in the excitation or emission spectra of GFP after treatment with anionic, cationic, zwitterionic, or nonionic detergents indicating a general rigid tertiary conformation. Overall, GFP is very resistant to denaturation, requiring treatment with 6 M guanidinium hydrochloride at 90 °C or under very acidic (pH < 4) or basic (pH > 12) conditions.

2.2 Three-dimensional Structure of GFP

The crystal structure of recombinant wildtype and Ser65Thr (substitution of

the amino acid serine by threonine at position 65) mutant of AvGFP was solved to a resolution of 1.9 Å by Yang et al., and Ormö et al. respectively. They revealed various structural aspects of GFP that explain some of the known biological behavior of this molecule and would allow a better rational design of mutants for future development of this tracer molecule. Overall, GFP presents a new protein fold termed *beta-can* (Fig. 2a, 2b).

The cylinder has a diameter of about 30 Å and a length of about 40 Å. There are 11 antiparallel beta strands forming the core of the very compact cylinder inside which there is a helix bearing the chromophore. At the end of the *can* are several short helixes. Such a compact single-domain structure with the chromophore centrally located protects itself from bulk solvent. This could explain why GFP is relatively stable and inaccessible to classical quenching agents such as acrylamide and molecular oxygen, resulting in the high quantum yield of fluorescence. At both the ends of the *can* are short loops and distorted α-helices. Interestingly, deletion of more than the methionine at N-terminus or more than seven amino acids from C-terminus leads to a severe loss of fluorescence without the typical spectrum of absorption, probably due to the improper folding of the *can*. Furthermore, two GFP proteins can also pack closely together to form a dimer in the crystal or in solution at low ionic strengths. Dimerization can affect the excitation spectra and energy transfer of GFP, thus the identification of the dimer contacts could allow mutagenic control of the state of assembly of the protein.

Since the first reports, structures for various mutant forms as well as the homologs of GFP from other species have also been solved, each presenting a conserved beta-can structure. They have provided important information on the key residues that affect the general or specific properties of GFP family proteins so as to allow further fine-tuning through genetic and protein engineering.

2.3 Chromophore of GFP

Colors originate from differential scattering of light or display of colored molecule – either by biochrome or pigments, or both. Their syntheses require multiple steps of sequential reactions catalyzed by a specific enzyme(s) and require certain cofactors. GFP represents a unique system where it is both the substrate and enzyme. Its chromophore is a product of modification of its own backbone and sidechains and is covalently bound within the protein. Indeed, the lack of dependency on multienzyme systems provides this *all-in-one* enzyme a unique advantage as an inconspicuous indicator of protein position in cells.

The first small peptide fragment containing the chromophore of GFP was isolated from a papain digest of heat-denatured GFP. The structure proposed, which was subsequently confirmed, comprises an internal Phe64-Ser-Tyr-Gly-Val-Gln69 sequence that is posttranslationally modified via an intramolecular autocatalytic cyclization to form a 4-(*p*-hydroxybenzylidene)-imidazolid-5-one structure (Fig. 2d). The formation of the chromophore involves a sequential, autocatalytic mechanism. The reaction is initiated by a rapid cyclization between Ser65 and Gly67 to form an imidazolin-5-one intermediate that is followed by a much slower rate-limiting oxygenation of the Tyr66 side chain by O_2 on a timescale of hours. Gly67 is required for the formation of chromophore;

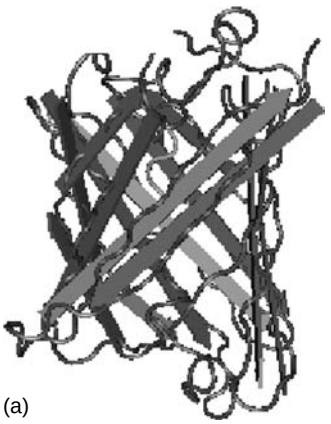

(a)

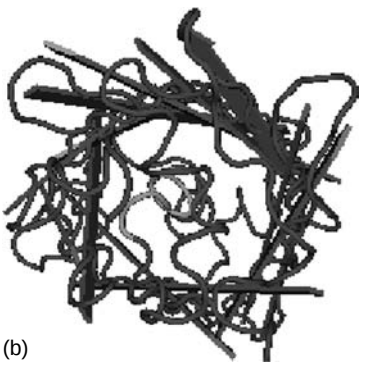

(b)

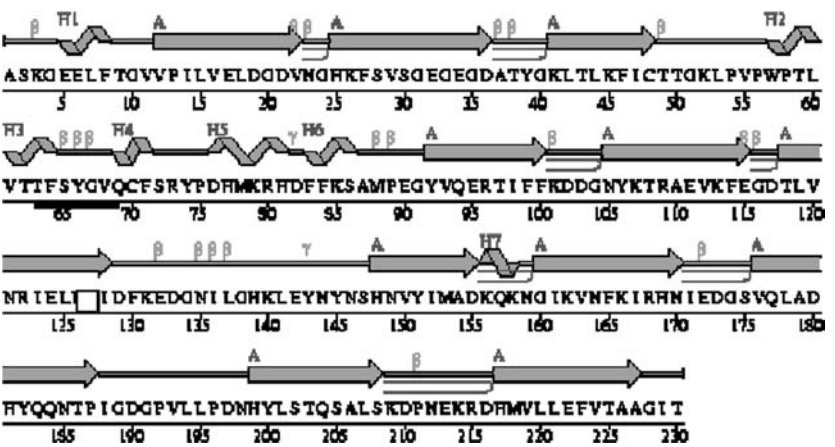

(c)

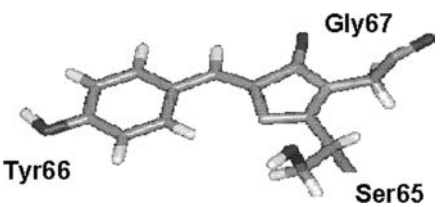

(d)

```
Aequorea victoria CTTG---KLPVPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQ 94
Ptilosarcus sp.   VTKG--GPLPFAFDIVSIAFQYGNRTFTKYPDDIA--DYFVQSFPAGFFY 96
Renilla muelleri  VTKG--APLPFAFDIVSPAFQYGNRTFTKYPNDIS--DYFIQSFPAGFMY 96
Anemonia sulcata  VIEG--GPLPFAFHILSTSCMYGSKTFIKYVSGIP--DYFKQSFPEGFTW 90
Discosoma striata VTKG--GPLPFGWHILCPQFQYGNKAFVHHPDNIH--DYLKLSFPEGYTW 93
Zoanthus sp.      VIEG--GPLPFSEDILSAGFKYGDRIFTEYPQDIV--DYFKNSCPAGYTW 93
Anemonia majano   VTMANGGPLAFSFDILSTVFKYGNRCFTAYPTSMP--DYFKQAFPDGMSY 95
                  .. :   .*::   .:.  : ** : *. :   :   *::  : * *
```

(e)

no other amino acid can replace glycine in this role, as seen by its conservation in different members of the GFP family (Fig. 2e). Oxygen molecules are then needed for the formation of the double bond between two carbons on the tyrosine to form an extended aromatic system. However, despite its requirement, the oxygen must be excluded from interacting with the chromophore itself to avoid collisional quenching of the fluorescence. Two resonant forms of the chromophore can be assumed: one with a partial negative charge on the benzyl oxygen of tyrosine and the other with the charge on the carbonyl oxygen of the imidazolidone ring. They can be stabilized by some of the positive charges developed by His148, Gln94, and Arg96 in the GFP itself. The fluorescence is not an intrinsic property of the Ser-Tyr-Gly tripeptide as this sequence can be found in quite a number of other proteins, none of which has the fluorescence.

With the discovery of several GFP-like molecules, it is a pleasant surprise to learn that there exist variations in the mechanisms of reactions. Thus far, at least 4 different color classes of GFP-like proteins can be distinguished from each other: green, yellow, orange-red and nonfluorescent purple-blue or chromoproteins. Interestingly, each class can apparently synthesize its own chromophore with several possible variations of pathways but none of these naturally occurring proteins has utilized any of the spectral-shifting solutions identified in mutant variants of GFP from *A. victoria* (see Sect. 2.4). Phylogenetic analyses also indicate that these chromophore structures are very recent entities and not the products of prolonged evolutionary fine-tuning. Coupled to the overall conservation of the beta-can folding of the GFP-like polypeptides, it is believed that the catalytic environment in GFP family proteins is *flexible* to accommodate switches between alternative pathways of chromophore formation during the course of evolution.

2.4 Genetically Modified GFP and its Natural Homologs

The fluorescent behavior of GFP depends greatly on the microenvironment surrounding the chromophore. Most of the polar residues in the pocket form a hydrogen-bonding network on the side of Tyr66 whereas atoms in the side chains of Thr203, Glu222, and Ile167 are in van

Fig. 2 Structure of GFP. (a) Side view of the GFP *beta can* with the chromophore (indicated in yellow) embedded in the center; the figure was generated by Cn3D view of wildtype *A. victoria* GFP, based on 1gfl at Protein Data Bank, accessible through NCBI website (http://www.ncbi.nlm.nih.gov/Structure/) (see color plate) (b) Top view for the *beta can* structure from figure A. (c) Summary of the PDB structure for 1gfl as available from http://www.biochem.ucl.ac.uk/bsm/pdbsum/1gfl/; note that the first methionine and the last eight residues (His-Gly-Met-Asp-Glu-Leu-Tyr-Lys) are not represented here. The chromophore site is indicated by heavy underline. (d) Structure of the chromophore of GFP. (e) Multiple sequence alignment of GFPs from different species as performed using the ClustalW (http://www2.ebi.ac.uk/clustalw/), showing invariant glycine residue (indicated by arrow) in the chromophore (underlined). Amino acids were color-coded to represent basic (magenta), acidic (blue), small or hydrophobic (red) and, hydroxyl or amine (green). * indicate residues that are conserved among all members, whereas : and. represent highly and moderately conserved residues respectively (see color plate p. xxvi).

der Waals contact with Tyr66 . Therefore, any mutations of these residues could have direct steric effects on the chromophore, thus the absorption and emission spectra of the GFP protein. Some mutations have significant wavelength shifts and most suffer a loss of fluorescence intensity.

To date, many random and directed point mutations have been introduced into AvGFP and have resulted in various changes in its fluorescent behavior (Table 2). Some of these improved characteristics are combined in the GFP variants known as enhanced GFP (EGFP) and are exceptionally bright and stable and are widely used in many applications nowadays. Subsequently, other mutants of AvGFP such as Blue Fluorescent Protein (BFP), Cyan Fluorescent Protein (CYP) and Yellow Fluorescent Protein (YFP), were derived with further improvement and distinctive properties useful for multicolor imaging as well as Fluorescence Resonance Energy Transfer (FRET) studies (see Sect. 3.2). Some mutants have increased fluorescence with their major excitation peak red-shifted to 490 nm with the emission staying at 509 nm. This is better for use with standard optical filter sets, as the mutant GFPs have excitation range more compatible with the commonly available optical filters. This is also more useful in confocal laser scanning applications.

Efforts to further red-shift AvGFP spectra to produce additional partners for multispectral experiments have had little success. Fortunately, some marine organisms have provided alternative fluorescent proteins to the red-wavelength range. These include a class of Anthozoan naturally fluorescent proteins that share 26 to 30% identity to the AvGFP. Among them is DrFP583 (better known as *DsRed* commercially) from the corallimorph *Discosoma* and the DsRed mutants. The sea anemone *Heteractis crispa* also offers another important source of alternative, the HcRed, which itself is a fluoresecent mutant of an original nonfluorescent blue chromoprotein cloned from this species. Each of them has emission peaks ranging from 579 to 645 nm, making multispectral imaging very feasible (Fig. 3). For instance, the HcRed that exhibited bright emission at 645 nm could be excitable by 600-nm dye laser, thus promoting new detection channels for multicolor flow cytometry applications. Furthermore, successful attempts were made in decreasing dimerization of these GFP variants at high concentrations. To date, about 30 significantly different members of *GFP-like proteins* have been identified and more discoveries are expected to follow. Despite a modest degree of sequence identity, all of them are likely to share similar β-can folding and possess intrinsic chromophores. While most of these GFP-like proteins are themselves fluorescent proteins, some are chromoproteins that display only intense absorption without fluorescence emission.

Following structural and mutagenic studies that have provided powerful insights into not just the mechanism of its chemistry but its potential for further development of more versatile GFPs, we have recently seen several new forms of GFPs or the GFP-like homologs being introduced, each capitalizing on the improved or unique labeling performances well as optimized for expression in eukaryotic system. They include, the *Ace-Green* expression vectors, encoding novel bright green fluorescent protein AceGFP (mammalian, *Dictyostelium* and *C. elegans* codon usage optimized); the *HcRED tandem* expression vectors, encoding the first commercially available nonoligomeric far-red

Tab. 2 Spectral properties of GFP and selected variants from *Aequorea victoria* or homologs from *Aequorea coerulescens, Discosoma* sp. and *Heteractis crispa*.

Fluorescent protein	*Source*	*Excitation/emission (nm)*	*Main properties and key mutations*
avGFP (wildtype)	*Aequorea victoria*	395[a], 475 / 507	The first GFP cloned and characterized. The template from which many other variants are derived.
EGFP (Enhanced GFP)	*Aequorea victoria*	489 / 507	Phe64Leu, Ser65Thr, red-shifted, accelerates speed of chromophore formation and elevates temperature sensitivity and humanized codon usage; increases fluorescence by 40-fold in mammalian cells; excitation peak nears the 488 nm laser line for confocal microscopy.
Destabilized GFP	*Aequorea victoria*	489 / 507	Reduced halflife with the PEST degradation domain of mouse ornithine decarboxylase at the C-terminal end of an EGFP; used to directly correlate gene induction with biochemical change
pH-sensitive GFP (pHluorin)	*Aequorea victoria*	395, 475 / 507	Two classes of mutants, one with maximum excitation peak that shifts to 475 nm as pH decreases (ratiometric pHluorin) and the other gradually loses fluorescence as pH decreases (ecliptic pHluorin). Both classes contain a common Ser202His mutation plus other unique composite of mutant residues; used for visualizing secretion and synaptic transmission; also used in plants for biomonitoring
AceGFP	*Aequorea coerulescens*	280, 390[a] / 460, 505	Derived from a colorless GFP-like protein from *Aequorea coerulescens* (acGFPL); a novel mutant (Glu222Gly substitution) for UV-induced photoconversion from an immature nonfluorescent form into a green fluorescent form.
BFP (Blue FP)	*Aequorea victoria*	384 / 448	Tyr66His, first used for multicolor imaging but it is dim and readily photobleached
CFP (Cyan FP)	*Aequorea victoria*	433 / 475	Tyr66Trp, blue-shifted, brighter and greater photostability than BFP; good dual color-imaging and FRET partner with YFP
YFP (Yellow FP)	*Aequorea victoria*	514 / 527	Ser65Gly and Thr203Tyr, red-shifted, brighter than EGFP but more sensitive in low pH and high halide concentrations, and photobleaching

(*continued overleaf*)

Tab. 2 (*Continued*)

Fluorescent protein	*Source*	*Excitation/emission (nm)*	*Main properties and key mutations*
DrFP583 (DsRed)	*Discosoma sp.*	558 / 583	Extending to red spectral range; slow maturation (>10 h) and self-associated to tetramers. A mutant form is used as "fluorescent timer" where initially it emits green fluorescence followed by red fluorescence
mRFP1	*Discosoma sp.*	584 / 607	Less bright than DsRed and photobleaches more easily, but faster maturation and no self-association
HcRed	*Heteractis crispa*	592 / 645	Derived from nonfluorescent chromoprotein and has far-red shifted spectra. Optimized for high solubility and rapid maturation through mutations, hence early detection. Tetramer in nature but mutant can form dimers that are preventable by expression in tandem; very useful for multicolor analyses when combined with cyan, green, yellow fluorescent proteins
PA-GFP	*Aequorea victoria*	400 (preactivation)/515 502 (postactivation)/517	Val163Ala and Thr203His. Improving photoconversion from neutral to anionic species. Photoactivable to give 100-fold increases in green fluorescence
Kaede	*Trachyphyllia geoffroyi*	508 (preactivation)/518 572 (postactivation)/582	Kaede exhibits remarkable 2000-fold increase in red-to-green fluorescence. However it requires around 400 nm light for photoactivation and it could form oligo-tetramers
KFP1	*Anemonia sulcata*	580 (preactivation)/600 580 (postactivation)/600	Mutant generated from a chromoprotein. Red fluorescence increases by 30-fold upon green irradiation – "kindling" with reversible photoactivation. It can also self-associate.

[a]Indicates major peak.

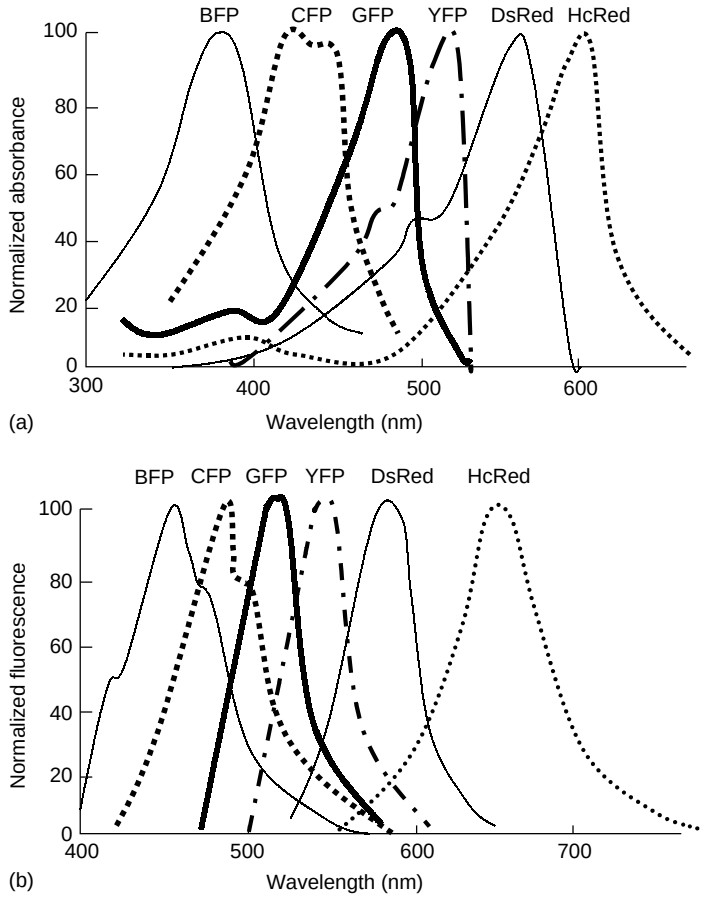

Fig. 3 The normalized absorbance (a) and emission (b) BFP, CFP, GFP, YFP, DsRed and HcRed fluorescent proteins.

fluorescent protein t-HcRed especially designed for protein localization and interaction studies. The other one, Kindling-RED *in vivo* tracking vectors, encodes unique photoactivated red kindling fluorescent protein (KFP-RED) that may be switched from nonfluorescent to fluorescent form by light. Some of the photoactivatable fluorescent proteins and their properties are described and discussed in more detail in Sect. 3.1 below. Some of the properties for these GFP-like proteins were recently reviewed by Miyawaki and Matz et al., highlighting the emerging prospects of GFP-like proteins in the wide spectrum of biotechnological applications, some of which will be discussed within the context of GFP applications below.

3 Applications of GFP Family Proteins

The wide applications of GFP family proteins are mainly attributed to their ability to form the chromophore via an internal autocatalytic cyclization without

external substrates and cofactors. This also bypasses the regulation by multiple enzymes. Furthermore, its relatively stable structure and inert behavior such as low or nontoxicity in cells and the lack of interference in the fusion proteins' activity and mobility makes it an even more attractive tool to use. With the availability of various mutants and homologs, some of them have been engineered to be optimized for use in certain hosts (through codon usage), while others can be made to change in their half-lives and fluorescence spectra with time (as fluorescent *timers*). These and other unique advantages have made GFP and other GFP-like proteins as some of the most versatile tools used in the modern day research. Some of their applications are given in the sections that follow.

3.1 GFP as a Reporter of Gene and Protein Expression

Understanding the timing of gene expression and the protein dynamics are very crucial for assigning the physiological function of a protein, especially when we want to understand how its expression and activity relate to the presence of other molecules in the cell. *GFP* was first used as a reporter gene in transgenic *C. elegans*. Various forms of GFP are now available for such studies, including mutant variants with different spectral properties (see Sect. 2.4) or other homologs such as the DsRed from the corallimorph *Discosoma*. Use of variants allows simultaneous visualization of multiple fusion proteins in a cell or multiple cell types in tissues. In gene-expression profiling, multicolor fluorescent proteins help to detect different aspects of gene expression. For example, when the peroxisome-targeted cyan fluorescent proteins (CFP) cDNA was linked to the *lac* operator site and coexpressed with the yellow fluorescent protein (YFP)-fused *lac* repressor, were able to observe the concerted transcription, marked by visible decondensation of YFP-associated chromatin and translation, as seen with an accumulation of CFP signals in the peroxisomes.

However, the relatively stable structure of GFP makes it difficult to accurately monitor the temporal sequence of reaction in cells or tissues as residual signals from the earlier process could mask the subsequent events. To decrease the half-life of GFP, Li et al. inserted a PEST-containing protein domain to the carboxyl terminus of GFP and showed that the new fusion could be rapidly degraded. While such *destabilized* GFP reporters are more difficult to detect, they do provide an excellent means to effectively quantify some processes such as the ubiquitin-mediated proteolytic process.

Another excellent example where modified GFP is used for temporal studies involves a fluorescent protein whose spectral properties change with time. This was generated by random mutations of the red fluorescent protein DrFP583 (DsRed). In contrast with the analogous GFP-based techniques, multiphoton bleaching of DsRed does not abolish its fluorescence completely but, instead, converts it from red to green, thus providing an additional way to regulate the process of fluorescence redistribution. Such a *fluorescent timer* produces green fluorescence initially but hours later, an oxygen-dependent autocatalytic reaction drives the formation of chromophore that emits red fluorescence. Therefore, by measuring the ratios of red to green one could determine whether the

expression is nascent (yellow to orange appearance) or has already stopped (appears entirely red).

Several other promising approaches to studying protein lifespans and turnover rates involve the use of photoactivatable fluorescent proteins, or the use of photobleaching-based kinetic microscopy techniques. Photoactivatable fluorescent proteins exhibit little initial fluorescence under excitation at the imaging wavelength but their fluorescence increase following the activation by irradiation at a different wavelength. The outcomes are direct display of specific populations of molecules in the cells, where the behavior and lifespans of the selective photoactivated molecules are studied independently of other newly synthesized proteins. The short time required for photoactivation (in some cases, less than 1 second) makes it potentially useful for investigating the movement and compartmental residency time of proteins.

To date, three forms of photoactivatable fluorescent systems have been developed; each one is based on different proteins and with different principles. One of these, PA-GFP was developed by improving the wildtype AvGFP photoconversion from a neutral to anionic species by the Thr203His mutation. It was based on the same excitation spectra change effect (increase in the 488-nm peak at the expense of a decrease in the 400-nm peak) with a remarkable 100-fold improvement in contrast. Furthermore, PA-GFP was developed using the monomeric GFP that is useful as a fusion tag for *in vivo* labeling. Such photoactivatable GFPs were tested both as a free protein to measure protein diffusion across the nuclear envelope and as a chimera with a lysosomal membrane protein to demonstrate rapid interlysosomal membrane exchange.

The second system was developed by Ando et al. who cloned a gene that encoded a fluorescent protein from a stony coral *Trachyphyllia geoffroyi* that emits green, yellow, and red light. The protein, which was serendipitously discovered as a result of someone accidentally leaving a test tube of the fluorescent protein overnight that turned red the following day, was named *Kaede*, the Japanese word for "maple leaf". Kaede includes a tripeptide, His-Tyr-Gly, which acts as a green chromophore that can be converted into red under UV or violet light (350–400 nm). It acts by exciting the protonated form of the chromophore. The extensive diffusion of Kaede allowed them to examine a single neuron in a dense culture where visualization of contact sites between the red and green neurons of interest becomes possible. Although this system allows bright and stable emission of red fluorescence, it however has a disadvantage by employing damaging UV light. Moreover, the tetrameric nature of the Kaede protein as well as the use of both green and red fluorescence makes it less versatile for studies of fusion proteins and multicolor labeling. Recently, Campbell et al. have converted DsRed into a monomeric red fluorescent protein (mRFP1) in a stepwise fashion. On the basis of the homology between Kaede and DsRed, similar semirational attempts are being made to develop a monomer of the Kaede for use in Kaede-fused proteins to monitor protein trafficking.

Both PA-GFP and Kaede protein do not allow reversible photoactivation that is ideal for certain applications, especially in fast processes. More recently, Chudakov et al. generated a mutant of *Anemonia sulcata* chromoprotein asCP that is capable of photoconversion from the nonfluorescent to a stable bright-red fluorescent

form (*kindling*). This *kindling fluorescent protein* (KFP1) allows rapid, repeated reversible kindling involving less harmful green light (compared to UV and blue light) and can be used together with other fluorescent markers for FRET analyses or multicolor imaging. For example, when used concurrently with PA-GFP, the distinctive differences in the photoactivation light required for them (blue for PA-GFP and green for KFP1) and their fluorescence products (green and red, respectively) can be used for precise in *vivo* photolabeling to track the movements of cells, organelles, and proteins. KFP1 was already used for *in vivo* cell labeling in mRNA microinjection assays to monitor *Xenopus* embryonic development and to track mitochondrial movement in mammalian cells. Using site-directed mutagenesis also showed that amino acids spatially surrounding Tyr(66) in the chromophore are crucial for kindling, probably involving chromophore-turning or cis–trans isomerization. They have also obtained novel variants that get kindled upon blue-light irradiation. The diversity of photoactivated fluorescent proteins discovered through site-directed mutagenesis is certainly poised for wide applications in the future.

As in photoactivatable fluorescent proteins, photobleaching also allow selected pools of fluorescent proteins to be distinguished from others and be monitored as they reequilibrate, without affecting protein pathways or creating protein gradients. Two techniques are used: the fluorescence recovery after bleaching (FRAP), and the fluorescence loss in photobleaching (FLIP). During FRAP, an area of the cell is photobleached under an intense laser pulse and the movement of unbleached molecules from the neighboring zones into the bleached area is recorded by time-lapse microscopy. FLIP, on the other hand, allows repeated photobleaching fluorescence to be performed in one area of the cell while collecting images of the entire cell. By monitoring the nonphotobleached areas, the movement of fluorescent-tagged proteins can be observed and the continuity of the cellular environments can be studied.

The various GFP-based tools of monitoring kinetics of gene and protein expression described above are summarized in the schematic diagram below (Fig. 4).

3.2 GFP-based FRET for Studies of Protein–Protein Interactions and Signal Activation

Fluorescence resonance energy transfer is the radiationless transfer of excited-state energy from an initially excited donor to an acceptor where the latter reemits light at another wavelength. It depends on the degree of overlap of the emission spectra of the donor and excitation spectra of the acceptor, their distance from each other and the relative orientation of the chromophore's transition dipoles. The efficiency of FRET is inversely proportional to the distance between two chromophores. It is proportional to the inverse sixth power of the intermolecular separation and it takes place only when they are separated by less than 10 nm apart. This means that compatible GFP-tagged proteins undergo FRET only when they are associated with direct binding or as part of protein complex. FRET imaging, therefore, provides spatial information about protein localization at a resolution unparalleled by any optical microscopes. The most commonly used GFP variant pairs for FRET experiments are the CFP (as the donor) and YFP (the acceptor) (Fig. 5). There are various ways to detect

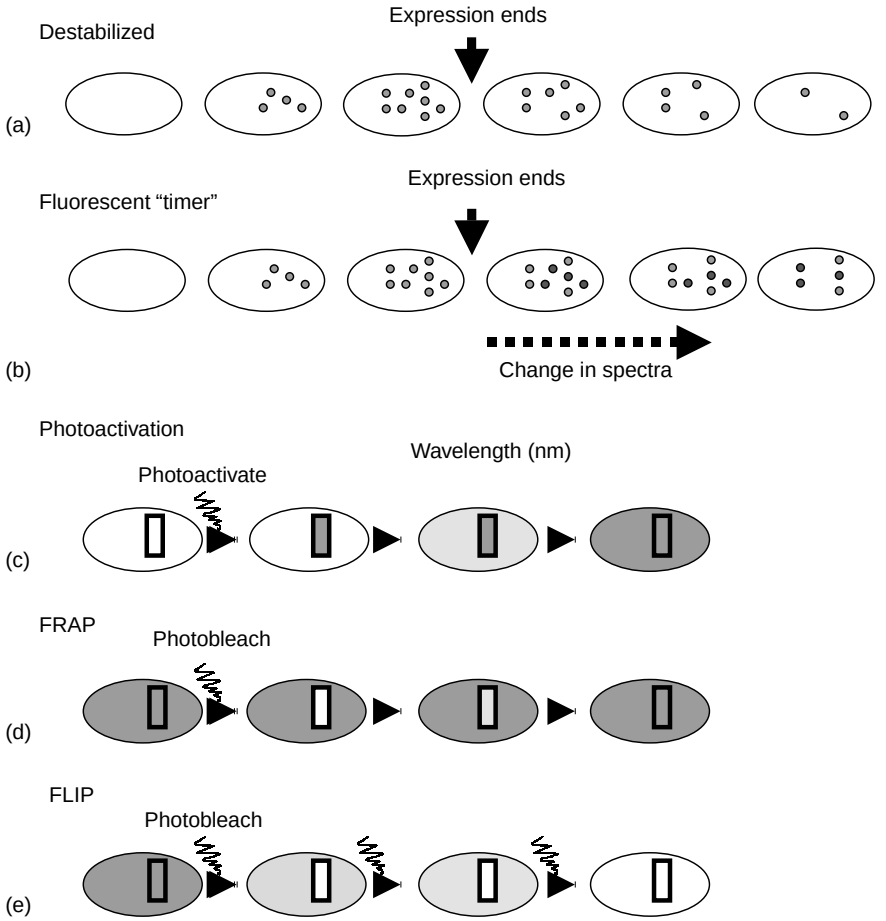

Fig. 4 GFP-based kinetics studies of gene and protein expression. (a) Destabilized fluorescent protein that is degraded as expression ends. (b) Fluorescent *timer* that changes its spectra with time. (c) Photoactivation allows selected irradiation and activation of certain pool of GFP (indicated by rectangle) such that their distribution out of this region can be monitored. (d) In Fluorescence Recovery After Photobleaching (FRAP), the recovery of fluorescent molecules into the region that is preirradiated and photobleached can be assessed to determine diffusion coefficients and mobile fractions. (e) In FLIP, a region of the cell is repeatedly photobleached and that movement of new fluorescent molecules into this region results in loss of fluorescence outside the target area. This helps assess the boundaries of protein's diffusional movement within the cell. Figures (a) and (b) are modified from van Roessel, P., Brand, A.H. (2002) Imaging into the future: visualizing gene expression and protein interactions with fluorescent proteins, *Nat. Cell Biol.* **4**, E15–E20 and figures (c), (d) and (e) are modified from Lippincott-Schwartz, J. and Patterson, G.H. (2003) Development and use of fluorescent protein markers in living cells, *Science*, **300**, 87–91 (see color plate p. xxv).

and measure FRET, which is dependent on the degree of association between two fusion proteins. One can either use the ratiometric method to determine the ratio of acceptor to donor fluorescence, or use the spectral measurement approach where the emission profiles of interacting FRET pair can be interpreted as additive mixture of emission spectra of two component chromophores. The third way is to use fluorescent lifetime imaging microscopy (FLIM) where energy transfer from donor to acceptor results in a decrease in the donor fluorescence lifetime. This method is useful for chromophores that are nearly identical in their emission spectra, and for donors that are prone to photobleaching.

FRET system has also been used between GFP and other chromophores, such as fluorescently labeled antibodies or luciferase. Bioluminescence resonance energy transfer (BRET) has been used in studies of protein complex where excitation illumination is an disadvantage to the sample. Examples for this include use of *Renilla* luciferase with enhanced YFP for protein interactions in circadian clock complex in cyanobacteria.

In addition to studies of protein–protein interactions, several important applications of GFP-based FRET in bioimaging have also been developed. Some of these fluorescent reporters are used as indicators of concentrations of second messengers or activation of certain biological

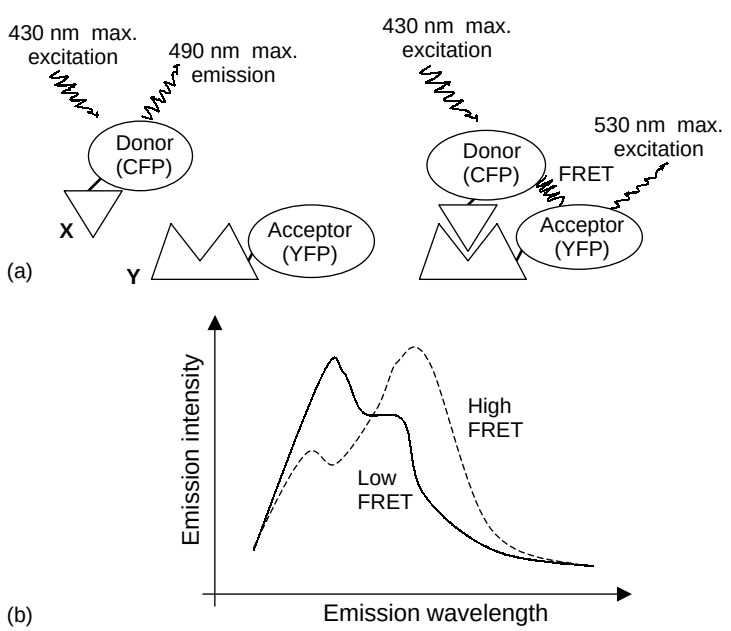

Fig. 5 Principles of fluorescence resonance energy transfer (FRET) with fluorescent proteins. (a) Schematic representation of FRET. (b) Schematic spectra of two fluorescent proteins undergoing FRET. Changes in the spectra reflect the extents of FRET, hence proximity between two interacting partners. Figure is (Adapted from van Roessel, P., Brand, A.H. (2002) Imaging into the future: visualizing gene expression and protein interactions with fluorescent proteins, *Nat. Cell Biol.* **4**, E15–E20.)

pathways. The first sensor for intracellular Ca^{2+} concentration, termed *CAMeleons*, incorporated a calmodulin domain and their ability to change color in the presence of different Ca^{2+} levels. Improved versions of this biosensor have since been reported that saw decreased pH and chloride sensitivity but increased photostability. Beyond ion-sensing, GFP-based FRET are used to examine various downstream signaling targets, including the cAMP, cGMP and the Ras family G-protein activation. To visualize Ras signaling in live cells, a single-gene construct encoding CFP-Ras and the Raf Ras-binding domain (RBD)-YFP was fused with a protein linker domain. After targeting the cell membrane by farnesyl moiety and being activated by the growth factor, RBD binds Ras resulting in a direct readout of Ras-activation patterns. Intriguingly, there was a wave of activation from cell membrane to nucleus – something unexpected based on previous biochemical data. Since the interacting partners were synthesized as a single-construct sensor, the Ras-GFP is very unlikely to have affected endogenous Raf and interfere with its signaling. Spatial and temporal dynamics of other signaling pathways have also been elucidated recently using the same approach. As cells respond to stimuli, certain pathways become activated or inactivated through series of specific protease actions, for example, involvement of caspases in apoptosis. To provide a direct *in vivo* marker for such a cleavage reaction, caspase-3-sensitive DEVD amino acid sequence was placed between BFP and GFP, and FRET was monitored for the caspase activity. One could envisage that by coupling the GFP-based FRET, various activation processes inside the cell can now be more readily monitored and quantified to provide a more dynamic picture of the *in vivo* setting.

3.3 GFP-fusion Protein for Imaging Cell Dynamics, Subcellular Localization, and Cell Signaling

One of the earliest and most exciting applications of GFP is the ability to track the movement of its fusion proteins in the cells *in vivo*. They have been very widely used in various experimental systems including viruses, plants, *Drosophila*, *Xenopus*, zebrafish, *C. elegans*, mouse, and other mammalian cells. Examples include studies of protein movement in many subcellular compartments or organelles such as targeting of GFP fusion protein domains to signaling metabolites such as phosphoinositides on the membrane. Using the same strategy, the pleckstrin homology (PH) domains of the cytosolic regulator of adenylyl cyclase (CRAC) and the general receptor for phosphoinositides (GRP1) were shown to be localized to the leading edge of the cell moving chemotactically towards a gradient of cyclic AMP. However, while the PH domain of CRAC concentrated solely on the leading edge during the chemotaxis, the GRP1 counterpart appeared to spread more posterior than CRAC. Indeed, various PH domains of varying binding affinity for phosphoinositides are also used as selective sensors for studies of signaling in insulin or growth factors, phagocytosis and macropinocytosis. Likewise, this approach could be used for studies of protein translocation from cytoplasm to nucleus and intracellular trafficking. For example, in signals leading to activation of transcriptional process, it was shown that the GFP-tagged STAT transcriptional factor clearly translocated to nucleus upon stimulation whereas the trafficking of ADP-ribosylation factor was demonstrated to be important for

maintenance of endoplasmic reticulum morphology in the plant *Arabidopsis*.

Cells perform dynamic process where subcellular components are continuously broken down and resynthesized and through local remodeling of cytoskeleton, they could undergo various morphological changes during different stages of growth; for example, cell division, cell migration, endocytosis, phagocytosis, or differentiation. Tagging of actin-binding proteins with GFP revealed detailed process of local synthesis and breakdown of actin cytoskeleton and by varying the ratio of labeled and unlabelled monomers in filaments of microtubule and actin microfilaments, it is now possible to observe them as *beads on a string*. Coupled with the FRET (see Sect. 3.2), the activation states of various small G-proteins of the Ras and Rac family can be monitored where Rac is activated as expected in the leading and trailing edges of the chemotactically moving cells, where active remodeling of the cytoskeleton takes place. Lately, the FLAP system has been employed with CYP and YFP of actin (Fig. 6a) and the dynamics of actin transport has been delineated during cell protrusion at the speed of 5 $\mu m\ s^{-1}$ in a process that is likely to involve active transport rather than diffusion. The dynamics of cell shape changes in response to the effects of a novel apoptosis-inducing Bnip2 and Cdc42GAP Homology (BCH) domain where cell shrinkage follows an earlier protrusion was also recorded. In other closely related proteins, extensive cell elongation was observed with the unique distribution of the protein on the leading edge (Fig. 6b). With the availability of CFP, YFP, and DsRed, triple labeling for more precise studies of subcellular localization of proteins of interest can now be obtained (Fig. 6c).

In addition to imaging of intracellular events, various forms of protein movement between cells *in vivo* or in tissue development have also been recorded using the GFP-fusion system such as those seen during the formation of *Drosophila* wing imaginal discs or during the *Drosophila* dorsal closure. Similarly, movement across synapses between cultured cortical neurons and the formation of dendrites or synapses in developing zebrafish embryos have also been recorded with the use of GFP-based reporter system. Likewise, the order of cell differentiation in developing eye disc of *Drosophila* could be demonstrated by double labeling that incorporated an enhanced mutant of DsRed as a developmental timer (Fig. 6d). In plant cell culture, a secretable form of GFP by addition of signal peptide is also being developed as a reporter for bioprocess applications.

3.4 GFP for Tracing Microbial Gene Expression and Host-pathogen Interactions

GFP are also popular among microbiologists. As for eukaryotes, GFP is used frequently to address the issue of bacterial gene expression and its regulation both at real-time and at single-cell level. An elegant example of transcriptional regulation in bacterial system is shown in flagella biosynthesis and chemotaxis where the expression of 14 flagella operons was monitored in real time using GFP on a low copy plasmid. Each of the 14 strains carried a different promoter-GFP construct and was grown in parallel. Cluster analyses of the fluorescence level detected at regular intervals revealed unexpected findings that the order of class 2 operon expression corresponded to spatial position of

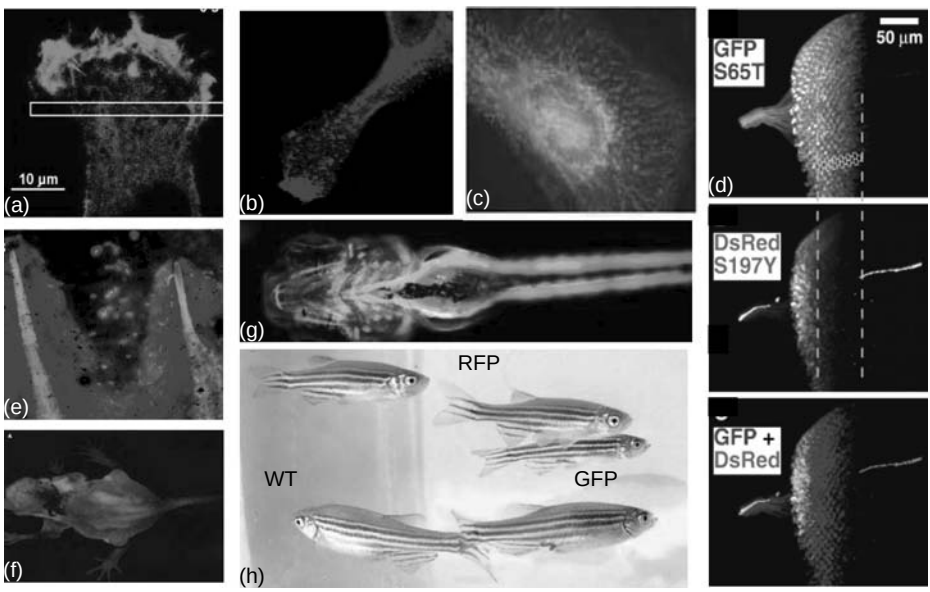

Fig. 6 Applications of GFP and variants in cellular and organismic systems. (a) CYP- image of rapid actin transport in the leading lamella (Reprinted with permission from Zicha, D., Dobbie, I.M., et al. (2003) Rapid actin transport during cell protrusion, *Science* **300**, 142–145. Copyright 2003 American Association for the Advancement of Science). (b) Localization of a GFP-tagged BNIP-2 and Cdc42GAP Homology (BCH) domain–containing protein at the leading edge of a protruding epithelial cell (BC Low; unpublished data). (c) Triple labeling of intracellular markers (Reprinted with permission from Becton, Dickinson and Company).
(d) Double labeling of developing eye disc of *Drosophila* (Reprinted with permission from Verkhusha, V.V., Otsuna, H., et al. (2001) An enhanced mutant of red fluorescent protein DsRed for double labeling and developmental timer of neural fiber bundle formation, *J. Biol. Chem.* **276**, 29621–29624. Copyright 2001 American Society for Biochemistry and Molecular Biology). Note that the DsRed signal was labeled *magenta* (used instead of *red* to ensure visibility for most color-blind readers). *Yellow hexagons* and *lines* indicate the rows of ommatidia cells that emit GFP signal but not DsRed (see color plate). (e) Invasion of *E. tarda* to the gills of fish (Reprinted with permission from Ling, S.H., Wang, X.H., et al. (2001) Green fluorescent protein-tagged *Edwardsiella tarda* reveals portal of entry in fish, *FEMS Microbiol Lett.* **194**, 239–243. Copyright 2001 Elsevier).
(f) Tumor growth and metastasis (lymphoma dissemination in eμ-myc-GFP transgenic mouse) (Reprinted with permission from Hoffman, R. (2002) Green fluorescent protein imaging of tumor growth, metastasis, and angiogenesis in mouse models, *Lancet Oncol.* **3**, 546–556. Copyright 2002 Elsevier).
(g) Double-color transgenic zebrafish: GFP expression under a skin-specific promoter and RFP expression under a muscle-specific promoter. (h) Ornamental transgenic zebrafish expressing GFP or RFP. WT, wildtype nontransgenic zebrafish. Please refer to the text for more detailed descriptions (see color plate p. xxiv).

the gene products in the flagella motor complex.

By incorporating fluoroscence-activated cell sorter (FACS), GFP can be employed in high-throughput screening of libraries of cells, for example, those that are expressed preferentially in the host. Similarly, GFP coupled with confocal microscopy allows probing into the structure of biofilms and also various aspects of cell dynamics including cell division, chromosome duplication and segregation, protein localization, and bacterial chemotaxis.

In other cases, the entry of microbes or its sub-particles such as human immunodeficiency virus, bacteria (& Fig. 6e;) or plant viruses to the host cells or tissues were monitored with relatively good resolution and provides important insights on the potential mechanism of pathogenesis. Finally, GFP is also being used as a screening tool during food preparation process. For examples, monitoring the *Lactobacillus sakei* in fermented meat products or the food colonization by *Aspergillus flavus*.

3.5 GFP Imaging of Tumor Growth, Metastasis and Angiogenesis

Unlike *in vitro* cellular system, one of the most difficult hurdles in studies of whole-animal physiology or pathophysiology is the ability to trace with confidence the behavior of cells or tissues and monitor them for a certain period of time. One good example is in the field of tumor biology where following the natural course or impediment of tumor progression and metastasis requires high specificity and high resolution of signal for the markers used. Recently, various groups have tried using GFP system for the purpose and the results are promising. Some have expressed GFP at high levels in cancer cells both *in vivo* and *in vitro* and transplanted them into animals and visualized *in situ* in fresh tissues where they have shown that GFP expression does not affect the metastatic behavior. Others have developed GFP models of metastatic disease for cancers in lung, prostate, melanoma, colon, breast, ovary, and brain. In some models, visualization of tumor growth, angiogenesis, dormancy, dissemination, invasion, and metastasis were made using the quantitative transcutaneous whole-body fluorescence imaging (Fig. 6f). The increased contrast between brightly fluorescent tumor tissue and blood vessels *in vivo* also greatly facilitates testing of antiangiogenic agents during metastatic development. In related work, Ahmed et al. used the mouse mammary virus promoter to directly drive the expression of GFP in the mammary gland for the analyses of transgenic mammary tumors and monitoring metastases *in vivo* at the single cell level of resolution. For clinical applications, selectively targeted tumor cells with retroviral GFP hoping to identify future occurrence of metastasis. Other prospects include labeling tumor-infiltrating lymphocytes and bone marrow that confer chemoresistance for the similar purposes. It might one day become a regular regime to use GFP transgenic rodents as vehicles for antidrug testing.

4 GFP and Transgenic Organisms

The pioneer work of to introduce the *GFP* gene into transgenic *C. elegans* not only demonstrated a new reporter gene in transgenic research but also revolutionized the transgenic field in many aspects. Compared to the traditional reporter genes, such as *chloramphenicol acetyl transferase* (*CAT*), *β-galactosidase*, and *luciferase*, *GFP* allows gene expression to be directly observed in real time in

living embryos/organisms and thus a dynamic event can be recorded. The use of GFP also eliminates the need of sacrifice of living samples for *in vitro* analyses. Despite that GFP is less sensitive than other reporters due to the lack of enzyme amplification effect, generally a single *GFP* transgene is sufficient to produce visible green fluorescence for detection. With the later development of many improved GFP variants and new species of fluorescent protein reporters, both the detection sensitivity and the fluorescent color spectra have been greatly improved.

So far, *GFP* has been used to produce transgenic organisms for many multicellular species, ranging from *C. elegans, Drosophila*, mosquitos, silkworms, sea urchins, ascidan, zebrafish, medaka fish, *Xenopus*, mice, rats, cows, pigs, monkeys to several species of plants. With the rapid progress of *GFP* transgenics, new applications of *GFP* transgenic organisms are constantly explored and reported. Now *GFP* transgenic approach has become a standard assay for characterization of gene regulatory elements for spatial and temporal expression, for analysis of gene expression program *in vivo*, for labeling cells to facilitate many experimental operations, and for formation of fusion proteins to determine the subcellular localization of a target protein and use as a tag to study the function of a target protein. In addition, *GFP* transgenic organisms also have numerous specialized applications, such as mutagenesis and mutant analysis, drug screening, biomonitoring, and ornamental. The following is a brief summary of these applications.

4.1 Analysis of Gene Promoters by Transient Transgenic Assay

In general, there are two types of transgenic systems: one is a transient transgenic assay in which analyses of gene expression are carried out when the introduced foreign gene remains in the extrachromosomal form; the other is a stable transgenic line where the transgene has been stably integrated into the host genome. The transient transgenic assay provides a rapid and convenient way to analyze a gene promoter, which generally contains many distinct binding sites called *cis*-elements that interact with specific proteins factors to control where, when, and how strongly the gene is expressed. By splicing different regions of a promoter to a reporter gene, analysis of its temporal and spatial expression allows a thorough investigation into the activity of a promoter. Compared to other commonly used reporter genes, *GFP* is particularly useful for dissection of *cis*-elements responsible for temporal and spatial patterns of gene expression. For example, have introduced into zebrafish embryos a series of deletions of the zebrafish *GATA-2* promoter ligated to the *GFP* reporter gene and was able to identify three distinct tissue-specific elements that enhance gene expression specifically in blood cell precursors, the enveloping layer of skin epithelia, or the central nervous system. The disadvantage of the *GFP* reporter gene in promoter analysis is the difficulty of quantifying reporter gene activity. This can be partially overcome by counting the number of expressing cells and by arbitrary classification of GFP expression into different intensity groups.

4.2 Recapitulation of Gene Expression Programs in Stable Transgenic Lines

Generation of stable *GFP* transgenic organisms under a tissue-specific promoter has become an established assay to analyze the expression program of a target gene. In

general, the GFP expression pattern faithfully reflects that of the endogenous gene. This approach is particularly powerful for the species with external development, such as invertebrate species, fish and amphibians, as GFP expression can be obtained in real time throughout embryogenesis and development. The zebrafish is probably the best model for this type of studies because of its rapid development as well as its transparent embryos to allow all major organs and tissues to be easily observed. The first stable line of *GFP* transgenic zebrafish under a zebrafish tissue-specific promoter was reported by. In this study, a 5.6-kb erythroid-specific *GATA-1* promoter was linked to the *GFP* reporter gene. Analyses of expression in the transgenic progeny showed that GFP was specifically expressed in erythroid cells from the progenitors to adult red blood cells, similar to the endogenous *GATA-1* expression. Thus, the *GATA-1:GFP* transgenic line provide an excellent model for a real-time study of the development of erythrocyte cell lineage. We have also created a double-color transgenic zebrafish line in which GFP is expressed in skin cells while RFP in muscle cells (Fig. 6(g)).

However, depending on the length and region of the promoters used, sometimes only a subset of cells that express the endogenous gene are able to express transgenic GFP because of the lack of certain gene regulatory elements in the promoters used. In mice, this limitation has been overcome by use of the gene knock-in technology, by which a functional gene like *GFP* is inserted, through embryonic stem cell–mediated homologous recombination, into a target gene locus and thus GFP transcription is completely under the control of the target gene promoter. Therefore, the expression pattern of the endogenous target gene can be fully recapitulated by visible GFP expression. For example, have knocked-in the *GFP* reporter gene into the *GATA-2* gene locus and the knock-in transgenic line faithfully recapitulated the endogenous expression profile of *GATA-2*, displaying green fluorescence widely in various tissues. In contrast, the expression profile in the regular *GFP* transgenic line by random insertion of up to 9-kb *GATA-2* 5′-flanking region linked to the *GFP* cDNA revealed that *GATA-2* expression was only partially recapitulated by transgenic GFP expression. For example, GFP expression in the urogenital tissues and in the aortic endothelial cells was observed in *GFP* knock-in transgenic line but not in the random insertion transgenic line.

Because of tissue-specific GFP expression in transgenic organisms, specific sets of cells are labeled by GFP expression and the development of these cells can be traced from the beginning of GFP expression. This is particularly useful to trace cell lineage, cell migration and axon growth of neurons during development. Another example of using *GFP* transgenic organisms to study cell lineage is use of an inducible promoter. In a *hsp70:GFP* transgenic zebrafish line generated by Halloran et al., the activation of GFP expression can be induced not only by heat shock but also by laser microbeam. Thus, by focusing the laser microbeam on specific cells, cells can be selectively activated for GFP labeling and these cells can then be traced during development.

4.3 Formation of GFP Fusion Proteins

In many cases, GFP can be fused to a complete or partial target protein and the

fusion proteins generally have dual functionality. GFP serves as a visible marker to monitor the cellular localization of target protein. To study cellular structures and dynamics in living cells of multicellular organisms, have carried out a large-scale project on random insertion of *GFP*-cDNA fusions into *Arabidopsis* genome and they were able to identify a variety of cellular localizations. Another potential of GFP fusion protein is to investigate both the expression and function of target protein simultaneously. For example, have generated a *C. elegans* transgenic line to express GFP-Elt (a GATA specific factor) fusion protein under its own gut-specific *Elt* promoter and found that the fusion protein was specifically expressed in gut nuclei and rescued the *elt-2* null mutant.

4.4 Use of GFP as Transgenic Marker and Labeling of Cells by GFP Expression

GFP is also commonly used as an *in vivo* reporter to assess frequency of transient and stable transformation. In plant cells, GFP has been used as a marker for selection of transformed cells to replace antibiotics markers. This also has practical applications to address the concern of using antibiotics resistant genes in field-released genetically modified crops. In mice, using *GFP* transgenic cells in embryonic chimera, GFP expression has been used as a preimplantation screen marker to improve the success rate of transgenic embryos. For transgenesis, *GFP* gene construct can be cointroduced into the transgenic organism and expression of visible GFP can be conveniently used to identify transgenic individuals. For example, in *Xenopus*, it has been proposed to coinject a crystalline promoter driven *GFP* construct with a target gene construct and thus expression of GFP in lens become a convenient marker for screen of transgenic *Xenopus*.

Targeted GFP expression in transgenic organisms also aids many other experimental operations such as cell and nuclear transplantation and microsurgery. Transgenic GFP labeled cells serve as an excellent donor or recipient for cell or nuclear transplantation, as contribution of the GFP expressing cells to the hosts (or a wildtype cell to a *GFP* transgenic host) can be easily identified in the developing embryos. *GFP* transgenic lines had been used to characterize the *eyeless* mutation in medaka in which either the wild-type transgenic fish donated the cells to the mutant host or *vice versa*. Study of the interaction between donor and host cells leads to the conclusion that the *eyeless* functions in a cell-autonomous manner. Recently, by aiming a laser beam at the GFP labeled retinal fiber in the *Shh:GFP* transgenic zebrafish, selectively ablated the tectum to study visuomotor behaviors in larval zebrafish. Likewise, *GFP* transgenic organisms also provide useful information on the dynamics of cells *in vivo*, for example, the elucidation of GFP-labelled synapses being formed by asynchronous process of branch retraction and elimination.

Owing to the ease of monitoring of gene expression, *GFP* transgenic organisms are also ideal systems to test new experimental tools in gene regulation. For example, *GFP* transgenic organisms have been widely used to test the efficiency of RNA interference (RNAi) in almost all model organisms including *C. elegans, Drosophila*, zebrafish, mouse, plants etc. Similarly, GFP transgenic embryos are also used to develop another gene knockdown technology using the morpholino antisense oligonucleotides. In addition, GFP has been frequently used to demonstrate the

functionality of some binary transgenic organisms such as the Cre-loxP recombination system and the GAL4-UAS activator–effector binary transgenic systems.

4.5 Mutant Generation and Characterization in Developmental Analysis

There are two areas that *GFP* transgenic organisms can serve as a valuable tool in mutant studies. First, the transgenic lines with restricted GFP expression in specific tissues/organs can be used for mutagenesis screen as the GFP labeled tissues/organs are readily recognizable. Both genetic and chemical (drug) screens can benefit from *GFP* transgenic lines. In the second area, *GFP* transgenic lines may help to characterize existing mutants after introduction of the *GFP* transgene into relevant mutants through standard breeding. Because of the ability to observe a dynamic GFP expression in living embryos, a high resolution of characterization of mutant can be achieved. For example, when transferred the *GFP* transgene from the gutGFP line into *ntl* (*no tail*) mutant and discovered the presence of the liver rudiment in $ntl^{-/-}$ embryos, though it was previously regarded to be absent from the $ntl^{-/-}$ embryos.

Furthermore, *GFP* transgenic organisms also provide an ideal experimental system to investigate the roles of upstream regulatory genes and signals in regulation of development of specific tissues. For example, in a fast muscle-specific *GFP* transgenic zebrafish line, GFP expression is specifically in fast skeletal muscles but not in slow muscles. By manipulating the Sonic hedgehog signals, GFP expression was inhibited because these GFP-labeled fast muscle cells were differentiated into non-GFP expressing slow muscles. Thus, the direct visualization of GFP labeled cells/tissues greatly facilitates the analysis of molecular events involving in the development of these labeled cells/tissues and to screen upstream molecules affecting their development.

4.6 Cell Sorting for *in vitro* Culture and Cell Type Specific cDNA Library Construction

Tissue-specific GFP expression in transgenic lines is also helpful in purification of rare cell types. These GFP expressing cells can be purified by FACS and used for *in vitro* cell culture and for construction of cell type–specific cDNA libraries, which will be an important source for isolation of transcripts only expressed in a single cell type. This approach is especially desirable if the cells of interest are too limited in number, a successful application of this technique in *GFP* transgenic fish has been reported by Long et al., who constructed a cDNA library using purified *GATA-1*-positive cells from *GATA-1:GFP* transgenic zebrafish embryos and isolated a novel hematopoietic death receptor gene, *ZH-DR*, from this cDNA library. Tomomura et al. have purified the rare cerebellar Purkinje cells from a transgenic mouse line that expresses GFP only in Purkinje cells in an adult brain.

4.7 Development of Transgenic Systems for Biomonitoring

GFP may have another potential applications in biomonitoring of environmental pollution if GFP expression is under control by some inducible promoters such as heavy metals and estrogens. This type

of studies has been carried out in *C. elegans* and fish. Small aquarium fish such as the zebrafish and medaka are particularly interesting for the development of a transgenic system for biomonitoring in which the transgenic fish change to fluorescent colors in the presence of environmental inducers (pollutants). There are many responsible DNA elements that can be considered for addition to the promoter driving the *GFP* reporter gene. Examples include aromatic hydrocarbon responsive elements (AHREs), electrophilic responsive elements (EPREs), metal responsive elements (MREs), estrogen responsive elements (EREs) and retinoic acid responsive elements (RARE, RXRE). These transgenic fish could hopefully be used in the detection of environmental pollutants based on the activated reporter gene expression. Recently, we made a *GFP* transgenic medaka line under the estrogen-responsive vitellogenin promoter and demonstrated that GFP can be turned on by treatment of estrogens. In addition to animals, plants are also being developed to express pH-sensitive derivative of GFP for rapid, environmentally induced changes in cytoplasmic pH in roots.

4.8 Transgenic Ornamentals

The constant discovery of new uses of GFP has now moved beyond science. The vivid fluorescent colors of GFP and other fluorescent proteins prompt us to develop transgenic organisms for ornamental purpose. This is particularly interesting for ornamental fish and flowers. To investigate the feasibility, we have developed several fluorescent transgenic zebrafish lines by using a strong muscle-specific promoter to drive *GFP, RFP*, or *YFP* reporter gene. These transgenic zebrafish can express these fluorescent proteins to a level of 3 to 17% of total muscle proteins and thus the fluorescent colors can be directly observed to unaided eyes even under normal daylight (Fig. 6h).

5 Future of GFP

"Seeing" is believing. When GFP was first introduced about 10 years ago as a marker of gene expression in heterologous host organisms, it gave the whole scientific community an astonishing ride into the internal world of mysterious cellular processes. Its overwhelming acceptance by the scientific community is a testimony for the rapid evolution in the development and improvement on the molecule and its variants as well as the advancement of the technology for its detection and signal interpretations. Today, they do not merely provide a spectacular view of internal structure of cells and lively protein movement within and between cells but also promise to enable more sophisticated studies of protein dynamics such as protein–protein interaction and second messenger cascades during intracellular as well as intercellular signaling.

In the postgenomic era, there is an urgent need to address the issue of gene and protein functions in the context of single cell system as well as the whole animal physiology. With many aspects of applications already described above, GFP can certainly fulfill a significant role in this *functional genomics* endeavor. The key question here is, can we expect another big wave of GFP green revolution? Or, simply, has it reached its full utility? The biggest challenge still lies in two major camps: the hardware and software development for

signal detection, analyses, and interpretation, as well as more refinements towards the biochemical properties of GFP and its variants, including other GFP-like proteins from the Anthozoa animals. While engineers are geared with more advanced imaging systems that are more sensitive with filters for detection of different chromophores and with quicker cameras for image capturing, biologists and chemists too are to place their creativity in tackling some of the limitations of GFP and arrive at more versatile and diverse collection of variants for dynamic bioassays, either by protein engineering or by hunting through nature. In the meanwhile, many of us will likely still be amazed and grateful by the power of GFP.

Acknowledgment

The authors' research was supported by National University of Singapore and Biomedical Research Council of Singapore.

See also Flow Cytometry; Gene Mapping and Chromosome Evolution by Fluorescence–Activated Chromosome Sorting; Nucleic Acid and Protein Single Molecule Detection and Characterization.

Bibliography

Books and Reviews

Chalfie, M., Kain, S. (1998) in: Martin, C., Steven, K. (Eds.) *GFP Green Fluorescent Protein: Properties, Applications and Protocols*, Wiley-Liss, Canada.

Gong, Z., Ju, B., Wan, H. (2001) Green fluorescent protein (GFP) transgenic fish and their applications, *Genetica* **111**, 213–225.

Lippincott-Schwartz, J., Patterson, G.H. (2003) Development and use of fluorescent protein markers in living cells, *Science* **300**, 87–91.

Matz, M.V., Lukyanov, K.A., Lukyanov, S.A. (2002) Family of the green fluorescent protein: journey to the end of the rainbow, *BioEssays* **24**, 953–959.

Miyawaki, A. (2002) Green fluorescent protein-like proteins in reef Anthozoa animals, *Cell Struct. Funct.* **27**, 343–347.

van Roessel, P., Brand, A.H. (2002) Imaging into the future: visualizing gene expression and protein interactions with fluorescent proteins, *Nat. Cell Biol.* **4**, E15–E20.

Weijer, C.J. (2003) Visualizing signals moving in cells, *Science* **300**, 96–100.

Yu, Y.A., Oberg, K., Wang, G., Szalay, A.A. (2003) Visualization of molecular and cellular events with green fluorescent proteins in developing embryos: a review, *Luminescence* **18**, 1–18.

Zimmer, M. (2002) Green fluorescent protein (GFP): applications, structure, and related photophysical behavior, *Chem. Rev.* **102**, 759–781.

Primary Literature

Ahmed, F., Wyckoff, J., Lin, E.Y., Wang, W., Wang, Y., Hennighausen, L., Miyazaki, J., Jones, J., Pollard, J.W., Condeelis, J.S., Segall, J.E. (2002) GFP expression in the mammary gland for imaging of mammary tumor cells in transgenic mice, *Cancer Res.* **62**, 7166–7169.

Amsterdam, A., Lin, S., Hopkins, N. (1995) The *Aequorea victoria* green fluorescent protein can be used as a reporter in live zebrafish embryos, *Dev. Biol.* **171**, 123–129.

Ando, R., Hama, H., Yamamoto-Hino, M., Mizuno, H., Miyawaki, A. (2002) An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12651–12656.

Baird, G.S., Zacharias, D.A., Tsien, R.Y. (2000) Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11984–11989.

Balla, T., Varnai, P. (2002) Visualizing cellular phosphoinositide pools with GFP-fused protein-modules, *Sci. STKE*, **2002**(125), PL3.

Beckman, M. (2003) Play-by-play imaging rewrites cells' rules, *Science* **300**, 76–77.

Bevis, B.J., Glick, B.S. (2002) Rapidly maturing variants of the Discosoma red fluorescent protein (DsRed), *Nat. Biotechnol.* **20**, 83–87.

Blackman, L.M., Boevink, P., Cruz, S.S., Palukaitis, P., Oparka, K.J. (1998) The movement protein of cucumber mosaic virus traffics into sieve elements in minor veins of *nicotiana clevelandii*, *Plant Cell* **10**, 525–538.

Campbell, R.E., Tour, O., Palmer, A.E., Steinbach, P.A., Baird, G.S., Zacharias, D.A., Tsien, R.Y. (2002) A monomeric red fluorescent protein, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7877–7882.

Carvan, M.J. III, Dalton, T.P., Stuart, G.W., Nebert, D.W. (2000) Transgenic zebrafish as sentinels for aquatic pollution, *Ann. N.Y. Acad. Sci.* **919**, 133–147.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., Prasher, D.C. (1994) Green fluorescent protein as a marker for gene expression, *Science* **263**, 802–805.

Chudakov, D.M., Belousov, V.V., Zaraisky, A.G., Novoselov, V.V., Staroverov, D.B., Zorov, D.B., Lukyanov, S., Lukyanov, K.A. (2003a) Kindling fluorescent proteins for precise *in vivo* photolabeling, *Nat. Biotechnol.* **21**, 191–194.

Chudakov, D.M., Feofanov, A.V., Mudrik, N.N., Lukyanov, S., Lukyanov, K.A. (2003b) Chromophore environment provides clue to 'kindling fluorescent protein' riddle, *J. Biol. Chem.* **278**, 7215–7219.

Cody, C.W., Prasher, D.C., Westler, W.M., Prendergast, F.G., Ward, W.W. (1993) Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein, *Biochemistry* **32**, 1212–1218.

Cole, N.B., Smith, C.L., Sciaky, N., Terasaki, M., Edidin, M., Lippincott-Schwartz, J. (1996) Diffusional mobility of Golgi proteins in membranes of living cells, *Science* **273**, 797–801.

Cormack, B.P., Valdivia, R.H., Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP), *Gene* **173**, 33–38.

Costerton, J.W., Stewart, P.S., Greenberg, E.P. (1999) Bacterial biofilms: a common cause of persistent infections, *Science* **284**, 1318–1322.

Cross, L.M., Cook, M.A., Lin, S., Chen, J.N., Rubinstein, A.L. (2003) Rapid analysis of angiogenesis drugs in a live fluorescent zebrafish assay, *Arterioscler. Thromb. Vasc. Biol.* **23**(5), 911–912.

Cruz, S.S., Chapman, S., Roberts, A.G., Roberts, I.M., Prior, D.A., Oparka, K.J. (1996) Assembly and movement of a plant virus carrying a green fluorescent protein overcoat, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6286–6290.

Cutler, S.R., Ehrhardt, D.W., Griffitts, J.S., Somerville, C.R. (2000) Random GFP::cDNA fusions enable visualization of subcellular structures in cells of Arabidopsis at a high frequency, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3718–3723.

Dantuma, N.P., Lindsten, K., Glas, R., Jellne, M., Masucci, M.G. (2000) Short-lived green fluorescent proteins for quantifying ubiquitin/proteasome-dependent proteolysis in living cells, *Nat. Biotechnol.* **18**, 538–543.

Davenport, D., Nicol, J.A.C. (1955) Luminescence of hydromedusae, *Proc. R. Soc. London, Ser.* **B144**, 399–411.

Day, R.N. (1998) Visualization of Pit-1 transcription factor interactions in the living cell nucleus by fluorescence resonance energy transfer microscopy, *Mol. Endocrinol.* **12**, 1410–1419.

Du, W., Huang, Z., Flaherty, J.E., Wells, K., Payne, G.A. (1999) Green fluorescent protein as a reporter to monitor gene expression and food colonization by *Aspergillus flavus*, *Appl. Environ. Microbiol.* **65**, 83483–83486.

Ellenberg, J., Lippincott-Schwartz, J., Presley, J.F. (1998) Two-color green fluorescent protein time-lapse imaging, *Biotechniques* **25**, 838–846.

Ellenberg, J., Lippincott-Schwartz, J., Presley, J.F. (1999) Dual-colour imaging with GFP variants, *Trends Cell Biol.* **9**, 52–56.

Ellson, C.D., Anderson,K.E., Morgan, G., Chilvers, E.R., Lipp, P., Stephens, L.R., Hawkins, P.T. (2001) Phosphatidylinositol 3-phosphate is generated in phagosomal membranes, *Curr. Biol.* **11**, 1631–1635.

Feng, G., Mellor, R.H., Bernstein, M., Keller-Peck, C., Nguyen, Q.T., Wallace, M., Nerbonne, J.M., Lichtman, J.W., Sanes, J.R. (2000) Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP, *Neuron* **28**, 41–51.

Field, H.A., Ober, E.A., Roeser, T., Stainier, D.Y.R. (2003) Formation of the digestive system. I. Liver morphogenesis, *Dev. Biol.* **253**, 279–290.

Fradkov, A.F., Verkhusha, V.V., Staroverov, D.B., Bulina, M.E., Yanushevich, Y.G., Martynov, V.I., Lukyanov, S., Lukyanov, K.A. (2002) Far-red fluorescent tag for protein labelling, *Biochem. J.* **368**(Pt 1), 17–21.

Fu, L., Buchholz, D., Shi, Y.B. (2002) Novel double promoter approach for identification of transgenic animals: a tool for in vivo analysis of gene function and development of gene-based therapies, *Mol. Reprod. Dev.* **62**, 470–476.

Fukushige, T., Hendzel, M.J., Bazett-Jones, D.P., McGhee, J.D. (1999) Direct visualization of the elt-2 gut-specific GATA factor binding to a target promoter inside the living *Caenorhabditis elegans* embryo, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11883–11888.

Gadella, T.W.J. Jr., Jovin, T.M., Clegg, R.M. (1993) Fluorescence lifetime imaging microscopy (FLIM)- spatial resolution of microstrucutres on the nanosecond time-scale, *Biophys. Chem.* **48**, 221–239.

Gardiner, E.M., Pestonjamasp, K.N., Bohl, B.P., Chamberlain, C., Hahn, K.M., Bokoch, G.M. (2002) Spatial and temporal analysis of Rac activation during live neutrophil chemotaxis, *Curr. Biol.* **12**, 2029–2034.

Gong, Z., Wan, H., Tay, T.L., Wang, H., Chen, M., Yan, T. (2003) Development of transgenic fish for ornamental and bioreactor by strong expression of fluorescent proteins in the skeletal muscle, *Biochem. Biophys. Res. Commun.* **388**, 58–63.

Godwin, A.R., Stadler, H.S., Nakamura, K., Capecchi, M.R. (1998) Detection of targeted GFP-Hox gene fusions during mouse embryogenesis, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13024–13027.

Gory, L., Montel, M.C., Zagorec, M. (2001) Use of green fluorescent protein to monitor *Lactobacillus sakei* in fermented meat products, *FEMS Microbiol. Lett.* **194**, 127–133.

Gurskaya, N.G., Fradkov, A.F., Terskikh, A., Matz, M.V., Labas, Y.A., Martynov, V.I., Yanushevich, Y.G., Lukyanov, K.A., Lukyanov, S.A. (2001) GFP-like chromoproteins as a source of far-red fluorescent proteins, *FEBS Lett.* **507**, 16–20.

Halloran, M.C., Sato-Maeda, M., Warren, J.T., Su, F., Lele, Z., Krone, P.H., Kuwada, J.Y., Shoji, W. (2000) Laser-induced gene expression in specific cells of transgenic zebrafish, *Development* **127**, 1953–1960.

Harry, E.J. (2001) Bacterial cell division: regulating Z-ring formation, *Mol. Microbiol.* **40**, 795–803.

Hasegawa, S., Yang, M., Chishima, T., Miyagi, Y., Shimada, H., Moossa, A.R., Hoffman, R.M. (2000) *In vivo* tumor delivery of the green fluorescent protein gene to report future occurrence of metastasis, *Cancer Gene Ther.* **7**, 1336–1340.

Hastings, J.W., Morin, J.G. (1969) Comparative biochemistry of calcium-activated photoproteins from the ctenophore, Mnemiopsis and the coelenterates *Aequorea, Obelia, Pelagia, Biol. Bull.* **137**, 402.

Heim, R., Tsien, R.Y. (1996) Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer, *Curr. Biol.* **6**, 178–182.

Heim, R., Cubitt, A.B., Tsien, R.Y. (1995) Improved green fluorescence, *Nature* **373**, 663–6644.

Heim, R., Prasher, D.C., Tsien, R.Y. (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12501–12504.

Hoffman, R. (2002) Green fluorescent protein imaging of tumor growth, metastasis, and angiogenesis in mouse models, *Lancet Oncol.* **3**, 546–556.

Ikawa, M., Kominami, K., Yoshimura, Y., Tanaka, K., Nishimune, Y., Okabe, M. (1995) Green fluorescent protein as a marker in transgenic mice, *Dev. Growth Diff.* **37**, 455–459.

Ikawa, M., Kominami, K., Yoshimura, Y., Tanaka, K., Nishimune, Y., Okabe, M. (1995) A rapid and noninvasive selection of transgenic embryos before implantation using green fluorescent protein (GFP), *FEBS Lett.* **375**, 125–128.

Jacinto, A., Woolner, S., Martin, P. (2002) Dynamic analysis of dorsal closure in Drosophila: from genetics to cell biology, *Dev. Cell* **3**, 9–19.

Ju, B., Chong, S.W., He, J., Wang, X., Xu, Y., Wan, H., Tong, Y., Yan, T., Korzh, V., Gong, Z. (2003) Recapitulation of fast skeletal muscle development in zebrafish by transgenic expression of GFP under the mylz2 promoter, *Dev. Dyn.* **227**, 14–26.

Kalir, S., McClure, J., Pabbaraju, K., Southward, C., Ronen, M., Leibler, S., Surette, M.G., Alon, U. (2001) Ordering genes in a flagella

pathway by analysis of expression kinetics from living bacteria, *Science* **292**, 2080–2083.

Keller-Peck, C.R., Walsh, M.K., Gan, W.B., Feng, G., Sanes, J.R., Lichtman, J.W. (2001) Asynchronous synapse elimination in neonatal motor units: studies using GFP transgenic mice, *Neuron* **31**, 381–394.

Kerbel, R.S. (1999) What is the optimal rodent model for antitumor drug testing? *Cancer Metastasis Rev.* **17**, 301–304.

Kohara, K., Kitamura, A., Morishima, M., Tsumoto, T. (2001) Activity-dependent transfer of brain-derived neurotrophic factor to postsynaptic neurons, *Science* **291**, 2419–2423.

Labas, Y.A., Gurskaya, N.G., Yanushevich, Y.G., Fradkov, A.F., Lukyanov, K.A., Lukyanov, S.A., Matz, M.V. (2002) Diversity and evolution of the green fluorescent protein family, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4256–4261.

Lee, M.H., Min, M.K., Lee, Y.J., Jin, J.B., Shin, D.H., Kim, D.H., Lee, K.H., Hwang, I. (2002) ADP-ribosylation factor 1 of *Arabidopsis* plays a critical role in intracellular trafficking and maintenance of endoplasmic reticulum morphology in *Arabidopsis*, *Plant Physiol.* **129**, 1507–1520.

Li, X., Zhao, X., Fang, Y., Jiang, X., Duong, T., Fan, C., Huang, C.C., Kain, S.R. (1998) Generation of destabilized green fluorescent protein as a transcription reporter, *J. Biol. Chem.* **273**, 34970–34975.

Ling, S.H., Wang, X.H., Lim, T.M., Leung, K.Y. (2001) Green fluorescent protein-tagged *Edwardsiella tarda* reveals portal of entry in fish, *FEMS Microbiol. Lett.* **194**, 239–243.

Liu, S., Bugos, R.C., Dharmasiri, N., Su, W.W. (2001) Green fluorescent protein as a secretory reporter and a tool for process optimization in transgenic plant cell cultures, *J. Biotechnol.* **87**, 1–16.

Long, Q., Huang, H., Shafizadeh, E., Liu, N., Lin, S. (2000) Stimulation of erythropoiesis by inhibiting a new hematopoietic death receptor in transgenic zebrafish, *Nat. Cell Biol.* **2**, 549–552.

Long, Q., Meng, A., Wang, H., Jessen, J.R., Farrell, M.J., Lin, S. (1997) GATA-1 expression pattern can be recapitulated in living transgenic zebrafish using GFP reporter gene, *Development* **124**, 4105–4111.

Majoul, I., Straub, M., Hell, S.W., Duden, R., Soling, H.D. (2001) KDEL-cargo regulates interactions between proteins involved in COPI vesicle traffic: measurements in living cells using FRET, *Dev. Cell* **1**, 139–153.

Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L., Lukyanov, S.A. (1999) Fluorescent proteins from nonbioluminescent Anthozoa species, *Nat. Biotechnol.* **17**, 969–973.

McDonald, D., Vodicka, M.A., Lucero, G., Svitkina, T.M., Borisy, G.G., Emerman, M., Hope, T.J. (2002) Visualization of the intracellular behavior of HIV in living cells, *J. Cell Biol.* **159**, 441–452.

McDonald, D., Wu, L., Bohks, S.M., KewalRamani, V.N., Unutmaz, D., Hope, T.J. (2003) Recruitment of HIV and its receptors to dendritic cell-T cell junctions, *Science* **300**, 1295–1297.

Meng, A., Tang, H., Ong, B.A., Farrell, M.J., Lin, S. (1997) Promoter analysis in living zebrafish embryos identifies a *cis*-acting motif required for neuronal expression of GATA-2, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6267–6272.

Miesenbock, G., De Angelis, D.A., Rothman, J.E. (1998) Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins, *Nature* **394**, 192–195.

Minegishi, N., Suzuki, N., Yokomizo, T., Pan, X., Fujimoto, T., Takahashi, S., Hara, T., Miyajima, A., Nishikawa, S.I., Yamamoto, M. (2003) Expression and domain-specific function of GATA-2 during differentiation of the hematopoietic precursor cells in midgestation mouse embryos, *Blood*; **102**, 896–905.

Miyawaki, A. (2003) Visualization of the spatial and temporal dynamics of intracellular signaling, *Dev. Cell* **4**, 295–305.

Miyawaki, A., Llopis, J., Heim, R., McCaffery, J.M., Adams, J.A., Ikura, M., Tsien, R.Y. (1997) Fluorescent indicators for $Ca2^{+}$ based on green fluorescent proteins and calmodulin, *Nature* **388**(6645), 882–887.

Mochizuki, N., Yamashita, S., Kurokawa, K., Ohba, Y., Nagai, T., Miyawaki, A., Matsuda, M. (2001) Spatio-temporal images of growth factor-induced activation of Ras and Rap1, *Nature* **411**, 1065–1068.

Moseyko, N., Feldman, L.J. (2001) Expression of pH-sensitive green fluorescent protein in *Arabidopsis thaliana*, *Plant Cell Environ.* **24**, 557–563.

Nagai, Y., Miyazaki, M., Aoki, R., Zama, T., Inouye, S., Hirose, K., Iino, M., Hagiwara, M. (2000) A fluorescent indicator for visualizing

cAMP-induced phosphorylation *in vivo*, *Nat. Biotechnol.* **18**, 313–316.

Nasevicius, A., Ekker, S.C. (2000) Effective targeted gene 'knockdown' in zebrafish, *Nat. Genet.* **26**, 216–220.

Niedz, R.P., Sussman, M.R., Satterlee, J.S. (1995) Green fluorescent protein – an in vivo reporter of plant gene-expression, *Plant Cell Rep.* **14**, 403–406.

Ormo, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y., Remington, S.J. (1996) Crystal structure of the *Aequorea victoria* green fluorescent protein, *Science* **273**, 1392–1395.

Patterson, G.H., Lippincott-Schwartz, J. (2002) A photoactivatable GFP for selective photolabeling of proteins and cells, *Science* **297**, 1873–1877.

Prasher, D.C., Eckenrode, V.K., Ward, W.W., Prendergast, F.G., Cormier, M.J. (1992) Primary structure of the *victoria* green-fluorescent protein, *Gene* **111**, 229–233.

Reits, E.A., Neefjes, J.J. (2001) From fixed to FRAP: measuring protein mobility and activity in living cells, *Nat. Cell Biol.* **3**, E145–E147.

Roeser, T., Baier, H. (2003) Visuomotor behaviors in larval zebrafish after GFP-guided laser ablation of the optic tectum, *J. Neurosci.* **23**, 3726–3734.

Sato, M., Hida, N., Ozawa, T., Umezawa, Y. (2000) Fluorescent indicators for cyclic GMP based on cyclic GMP-dependent protein kinase Ialpha and green fluorescent proteins, *Anal. Chem.* **72**, 5918–5924.

Shimomura, O (1998) in: Chalfie, M., Kain, S. (Eds.) *Green Fluorescent Protein: Properties, Applications and Protocols*, Wiley-Liss, New York, pp. 3–16.

Shimomura, O., Johnson, F.H., Saiga, Y. (1962) Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan *Aequorea*, *J. Cell. Comp. Physiol.* **59**, 223–239.

Small, J.V., Kaverina, I. (2003) Microtubules meet substrate adhesions to arrange cell polarity, *Curr. Opin. Cell Biol.* **15**, 40–47.

Southward, C.M., Surette, M.G. (2002) The dynamic microbe: green fluorescent protein brings bacteria to light, *Mol. Microbiol.* **45**, 1191–1196.

Suzuki, Y., Yasunaga, T., Ohkura, R., Wakabayashi, T., Sutoh, K. (1998) Swing of the lever arm of a myosin motor at the isomerization and phosphate-release steps, *Nature* **396**, 380–383.

Teleman, A.A., Cohen, S.M. (2000) Dpp gradient formation in the *Drosophila* wing imaginal disc, *Cell* **103**, 971–980.

Terskikh, A., Fradkov, A., Ermakova, G., Zaraisky, A., Tan, P., Kajava, A.V., Zhao, X., Lukyanov, S., Matz, M., Kim, S., Weissman, I., Siebert, P. (2000) "Fluorescent timer": protein that changes color with time, *Science* **290**, 1585–1588.

Tomomura, M., Rice, D.S., Morgan, J.I., Yuzaki, M. (2001) Purification of Purkinje cells by fluorescence-activated cell sorting from transgenic mice that express green fluorescent protein, *Eur. J. Neurosci.* **14**, 57–63.

Tsien, R.Y. (1998) The green fluorescent protein, *Annu. Rev. Biochem.* **67**, 509–544.

Tsukamoto, T., Hashiguchi, N., Janicki, S.M., Tumbar, T., Belmont, A.S., Spector, D.L. (2000) Visualization of gene activity in living cells, *Nat. Cell Biol.* **2**, 871–878.

Valdivia, R.H., Falkow, S. (1997) Fluorescence-based isolation of bacterial genes expressed within host cells, *Science* **277**, 2007–2011.

Verkhusha, V.V., Otsuna, H., Awasaki, T., Oda, H., Tsukita, S., Ito, K. (2001) An enhanced mutant of red fluorescent protein DsRed for double labeling and developmental timer of neural fiber bundle formation, *J. Biol. Chem.* **276**, 29621–29624.

Wan, H., He, J., Ju, B., Yan, T., Lam, T.J., Gong, Z. (2002) Generation of two-color transgenic zebrafish using the green and red fluorescent protein reporter genes, *gfp* and *rfp*, *Mar. Biotech.* **4**, 146–154.

Wang, S., Hazelrigg, T. (1994) Implications for bcd mRNA localization from spatial distribution of exu protein in *Drosophila* oogenesis, *Nature* **368**, 400–403.

Ward, W.W. (1998) in: Chalfie, M., Kain, S. (Eds.) *Green Fluorescent Protein: Properties, Applications and Protocols*, Wiley-Liss, New York, pp. 45–75.

Ward, W.W., Bokman, S.H. (1982) Reversible denaturation of *Aequorea* green-fluorescent protein: physical separation and characterization of the renatured protein, *Biochemistry* **21**, 4535–4540.

Ward, W.W., Prentice, H.J., Roth, A.F., Cody, C.W., Reeves, S.C. (1982) Sprectral

perturbations of the Aequora green-fluorescent protein, *Photochem. Photobiol.* **35**, 803–808.

Webb, C.D., Graumann, P.L., Kahana, J.A., Teleman, A.A., Silver, P.A., Losick, R. (1998) Use of time-lapse microscopy to visualize rapid movement of the replication origin region of the chromosome during the cell cycle in *Bacillus subtilis*, *Mol. Microbiol.* **28**, 883–892.

Wheeler, R.T., Shapiro, L. (1999) Differential localization of two histidine kinases controlling bacterial cell differentiation, *Mol. Cell* **4**, 683–694.

Wianny, F., Zernicka-Goetz, M. (2000) Specific interference with gene function by double-stranded RNA in early mouse development, *Nat. Cell Biol.* **2**, 70–75.

Winkler, S., Loosli, F., Henrich, T., Wakamatsu, Y., Wittbrodt, J. (2000) The conditional medaka mutation eyeless uncouples patterning and morphogenesis of the eye, *Development* **127**, 1911–1919.

Xu, Y., Piston, D.W., Johnson, C.H. (1999) A bioluminescence resonance energy transfer (BRET) system: application to interacting circadian clock proteins, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 151–156.

Xu, X., Gerard, A.L., Huang, B.C., Anderson, D.C., Payan, D.G., Luo, Y. (1998) Detection of programmed cell death using fluorescence energy transfer, *Nucleic Acids Res.* **26**, 2034–2035.

Yang, F., Moss, L.G., Phillips, G.N. Jr. (1996) The molecular structure of green fluorescent protein, *Nat. Biotechnol.* **14**, 1246–1251.

Zacharias, D.A., Violin, J.D., Newton, A.C., Tsien, R.Y. (2002) Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells, *Science* **296**, 913–916.

Zhang, C.L., Chen, D.F., McCormac, A.C., Scott, N.W., Elliott, M.C., Slater, A. (2001) Use of the GFP reporter as a vital marker for *Agrobacterium*-mediated transformation of sugar beet (*Beta vulgaris L.*), *Mol. Biotechnol.* **17**, 109–117.

Zhou, Y.T., Soh, U.J., Shang, X., Guy, G.R., Low, B.C. (2002) The BNIP-2 and Cdc42GAP homology/Sec14p-like domain of BNIP-Sα is a novel apoptosis-inducing sequence, *J. Biol. Chem.* **277**, 7483–7492.

Zicha, D., Dobbie, I.M., Holt, M.R., Monypenny, J., Soong, D.Y., Gray, C., Dunn, G.A. (2003) Rapid actin transport during cell protrusion, *Science* **300**, 142–145.

Zolotukhin, S., Potter, M., Hauswirth, W.W., Guy, J., Muzyczka, N.A. (1996) ''Humanized'' green fluorescent protein cDNA adapted for high-level expression in mammalian cells, *J. Virol.* **70**, 4646–4654.

Repressor-Operator Recognition: *see* DNA, Repressor–Operator Recognition

Reproduction of Plants

Jaroslava Dubová[1], *Jan Hejátko*[1,3], *and Jiří Friml*[2]
[1] *Masaryk University, Kotlářská 2, Brno, Czech Republic*
[2] *ZMBP University of Tübingen, Auf der Morgenstelle 3, Tübingen, Germany*
[3] *Academy of Sciences of the Czech Republic, Institute of Biophysics, Brno, Czech Republic*

1 Asexual (Vegetative) Reproduction – Amixis 252

2 Sexual Reproduction – Amphimixis 252
2.1 Plant Life Cycle, Alternation of Sporophyte/Gametophyte Generations in Plant Evolution 252
2.2 Structure and Development of Plant Sexual Organs 253
2.2.1 Transition from Vegetative to Reproductive Growth 256
2.2.2 Floral Meristem Identity Acquisition 259
2.2.3 Flower Formation 260
2.3 Structure and Formation of Male and Female Gametes 264
2.3.1 Male Gamete Formation 264
2.3.2 Female Gamete Formation 269
2.4 Angiosperm Fertilization 275
2.4.1 Progamic Phase 275
2.4.2 Syngamic Phase: Karyogamy and Plasmagamy, Double Fertilization 278
2.4.3 Prezygotic and Postzygotic Barriers to Fertilization 279
2.4.4 *In Vitro* Fertilization 280
2.5 Embryo and Endosperm Development 280
2.5.1 Endosperm Development 280
2.5.2 Embryo Development 283
2.5.3 Interregional Communication During Seed Development 287
2.6 Seed and Fruit Formation 288
2.6.1 Seed and Fruit Maturation 288
2.6.2 Dormancy and Seed Germination 288

3 Apomixis 290

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

Acknowledgment 291

Bibliography 291
Books and Reviews 291
Primary Literature 291

Keywords

Amixis
Asexual propagation of plants by fragmentation, cuttings from stems, leaves, and roots, by rhizomes or bulbils. Offspring have identical copies of the parental genome.

Amphimixis
Sexual reproduction including formation of haploid gametes and their fusion.

Apomixis
The asexual formation of a seed from the maternal tissues of the ovule, avoiding the processes of meiosis and fertilization.

Auxin Gradients
Differential (graded) distribution of the plant signaling molecule indole-3-acetic acid, IAA, which mediates various patterning processes.

Double Fertilization
Fertilization in angiosperms when one sperm fuses with the egg cell and the second sperm fuses with the two polar nuclei. These two events usually result in the formation of the diploid zygote and the triploid primary endosperm nucleus. Ploidy level of the endosperm is determined by the ploidy level of the fusion nucleus, which depends on the type of embryo sac formation.

Embryogenesis
Process that transforms a single-celled zygote into the embryo comprising all the basic pattern elements of the future plant.

Endosperm
Special tissue in angiosperms originating from the triploid primary endosperm cell. The roles of the endosperm in the embryo development are both nutritive and regulatory. Developmental types of endosperm (nuclear, cellular, helobial) are distinguished by differences in nuclear division and cell wall formation. Endosperm can function as a storage tissue of the mature seed (albuminous seeds) or as nutritive tissue for the developing embryo (exalbuminous seeds).

Female Gamete
Egg cell.

Gametophyte
Haploid generation producing male and/or female gametes that form a zygote upon fertilization.

Incompatibility
Inability of a fertile hermaphrodite flowering plant to produce a zygote after pollination. Various types of the self-incompatibility prevent inbreeding; cross-incompatibility presents a barrier for remote pollination.

Incongruity
Incompleteness of space and time coordination in the plant reproduction process caused by evolutionary divergence.

Male Gamete
Sperm cell.

Seedling
Juvenile form of plant, which consists of a shoot apical meristem, embryonic leaves (cotyledons), hypocotyl, and root with root apical meristem. Seedling is a product of seed germination.

Somatic Embryogenesis
Formation of embryo-like structures originating from somatic cells. Very rarely spontaneous *in planta*, more often induced in plant tissue cultures.

Sporophyte
Diploid generation producing haploid spores.

Zygote
Product of gamete (egg cell and sperm cell) fusion, the first cell of a new sporophyte generation.

Plant reproduction involves both *asexual propagation* (with various types of vegetative reproduction) and *sexual reproduction*. Alternation of haploid and diploid phases is typical for sexual reproduction. Processes leading to the formation of haploid male and female gamete and their subsequent fusion restoring the diploid state are parts of one plant life cycle. Various types of *apomixis* represent modified forms of asexual reproduction by which seeds are produced without fertilization.

1
Asexual (Vegetative) Reproduction – Amixis

Asexual (vegetative) propagation of plants by fragmentation, bulbils, rhizomes, or underground organs like roots and tubers produce offspring that have identical copies of the parental genes. Such propagation allows rapid colonization of a new site and enables even isolated individuals to reproduce. This ability of vegetative plant reproduction is being exploited in horticultural practice for multiplication of desired clones for a long time. Nevertheless, the development of plant tissue cultures amplified the rate of multiplication and extended the list of organs and tissues, in which induction of organogenesis can be initiated. Furthermore, this approach also permitted the development of structures resembling various stages of regular zygotic embryogenesis from somatic cells and/or pollen grains. *SOMATIC EMBRYO RECEPTOR KINASE1* (*SERK1*) is expressed transiently in a small population of embryo competent cells up to the globular stage and seems to be a highly specific signal transduction system of somatic embryogenesis. Most other genes cloned from somatic embryos of carrot (*DC8, DC59, and EMB1*) are homologous to *LATE EMBRYOGENESIS ABUNDANT* (*LEA*) genes. Pollen embryogenesis *in vitro* can be induced by various signals, such as heat-shock treatment or starvation, which can affect organization of cytoskeletal elements and lead to a symmetrical cell division.

2
Sexual Reproduction – Amphimixis

Almost all eukaryotic organisms have evolved a mode of sexual reproduction not only to permit the maintenance of the species but also to increase genetic diversity, which is the basis for adaptation and evolutionary potential. Sexual reproduction is brought about by specialized structures in which the haploid gametes as well as structures and/or mechanisms that ensure the fusion of these gametes are produced. Haploid *male* or *female gametes* are formed in special flower tissues, called *sporangia*, during the processes of *sporogenesis* (formation of haploid spores by meiotic reduction division) and subsequent *gametogenesis* (differentiation of the gametes) (see Sect. 2.3). Somatic chromosome number is restored after gamete fusion – *fertilization* – that results in *zygote* formation. The zygote is the first cell of the next sporophytic generation. Mitotic activity of the zygote and differentiation of these cells leads to *embryo* formation. The mature embryo enclosed within the seed is the structural link between the completed gametophytic generation and the future sporophytic generation of the plant.

2.1
Plant Life Cycle, Alternation of Sporophyte/Gametophyte Generations in Plant Evolution

The eukaryotic life cycle includes events leading to the formation of gametes as well as the development ensuring their fusion. Two alternating generations can be recognized in the plant life cycle: a haploid phase, known as the *gametophytic* and a diploid phase, known as the *sporophytic generation*. The gametophyte reproduces by means of *gametes*, but it does not reproduce itself directly. Instead, the zygote resulting after gamete fusion develops into the sporophyte. The

sporophyte, similarly, is not reproduced directly, but forms reproductive cells known as *spores*, which develop into gametophytes. Thus, the gametophytic and the sporophytic generations alternate and reproduce each other. The relative lengths of the sporophyte and the gametophyte generations have changed during evolution from a dominant autotrophic (self-feeding) gametophyte and a nutritionally dependent sporophyte to a dependent gametophyte and a dominant autotrophic sporophyte. In mosses (Fig. 1a), both the embryo and the mature sporophyte are dependent on the photosynthetic gametophyte for nutrition. Meiosis within the capsule of the sporophyte yields haploid spores that are released and germinate to form a male or female green gametophyte. Differentiation of the gametophyte produces *antheridia* with sperm cells in males and *archegonia* with eggs in females. Fertilization occurs after transfer of sperm cells in drops of water, and the sporophyte generation develops in the sporangium that remains attached to the gametophyte. Both the gametophyte and the sporophyte of ferns (Fig. 1b) photosynthesize and are thus autotrophic; the shift to a dominant sporophyte generation is apparent. Further reduction of the gametophyte generation during evolution led to the highly reduced gametophyte of angiosperms (Fig. 1c and Sect. 2.3), where the *male gametophyte* is represented only by the *pollen grain* consisting of the *vegetative cell* that develops a pollen tube and the *generative cell*, which produces *male gametes* – the *sperm cells*. The *female gametophyte* of angiosperms is represented by the *embryo sac*, which often contains eight haploid cells, the most important of which is the *female gamete* – the *egg cell* (EC).

2.2 Structure and Development of Plant Sexual Organs

Molecular mechanisms of flower development have recently been extensively investigated in the model plant *Arabidopsis thaliana*. Although other model plants have been used to study the molecular genetics of flower development (e.g. *Petunia* and *Antirrhinum*), this brief overview will focus on recent models of flower development in *Arabidopsis*.

During vegetative growth, the aerial portion of the vegetative plant organs (*shoot*) is a product of the *shoot apical meristem* (SAM). In indeterminate species (like *Arabidopsis* and similar plants with raceme inflorescence), flowers are formed in *inflorescence*, growing out from the rosette of pedestrian leaves in *Arabidopsis*. Thus, induction of flowering in *Arabidopsis* involves two transitions: (1) formation of inflorescence (otherwise known as *bolting*) from *inflorescence apical meristem* and (2) formation of individual flowers from *floral meristems* growing out from the inflorescence. In determinate species, the inflorescence meristem differentiates into a single floral meristem leading to formation of a single terminal flower.

Multiple endogenous and environmental stimuli lead to the switch from vegetative to reproductive growth (see the following). After bolting is induced, secondary inflorescence meristems, *axillary meristems*, are formed giving rise to lateral inflorescences (coflorescences) on the axils of 2 to 5 *cauline leaves*. It is still not completely clear whether SAM directly switches to the inflorescence apical meristem or whether there is some interphase between the vegetative and reproductive phases. The first model predicts that cauline leaves and the axillary meristems are produced from

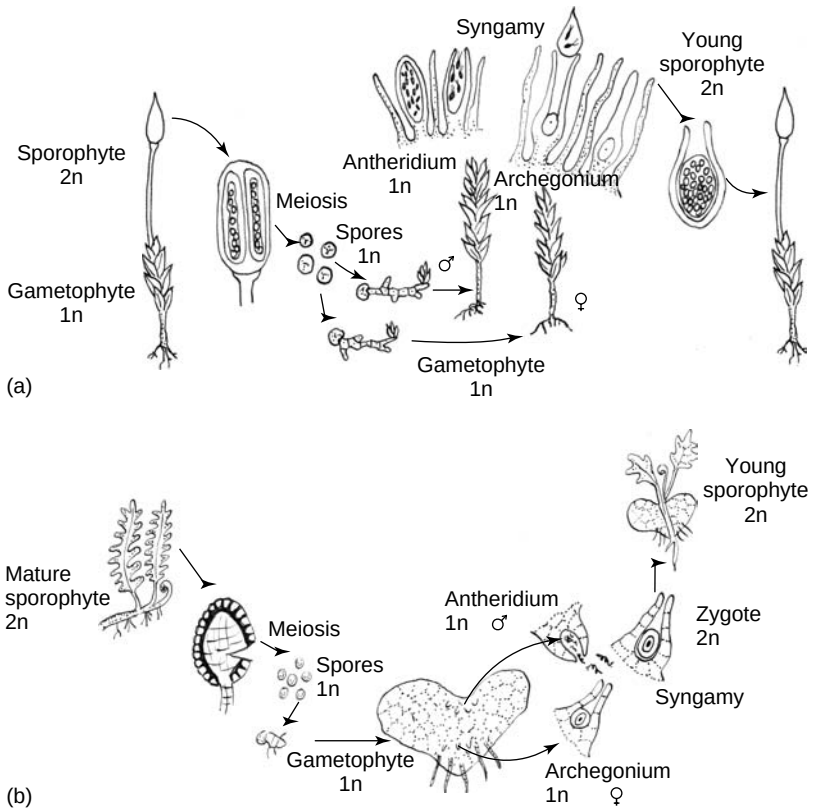

Fig. 1 Schematic representation of the alternation of sporophyte/gametophyte generations in plant evolution. (a) Life cycle in mosses: the sporophyte generation in mosses is reduced to only a sporangium within which haploid spores are formed during meiosis. Germinating spores form a male or female green gametophyte. Differentiation of the gametophyte produces antheridia with sperm cells in males and archegonia with eggs in females. Fertilization occurs after transfer of the sperm cell in a drop of water to the egg cell within the archegonium. The next sporophyte generation develops in the archegonium and remains attached to the gametophyte. (b) Life cycle in a fern: most (but not all) ferns are homosporous and produce only one type of spore within a structure called the *sporangium*. The gametophyte develops from the spores and can produce both male and female sex organs. Water facilitates fertilization like in mosses. The biggest contrast between the mosses and the ferns is that both the gametophyte and the sporophyte of the fern are autotrophic (According to Fosket, D.E. (1994) *Plant Growth and Development: A Molecular Approach*, Academic Press, San Diego, CA). (c) Life cycle in the angiosperms: the gametophyte generation of the angiosperms is reduced to just a few cells. The male gametophyte is formed after meiosis in microsporangia located in anthers and is represented by the pollen grains, which are complexes of two or three cells. The female gametophyte – the embryo sac – containing the egg cell, synergids, antipodals, and the central cell, is formed within the megasporangium (ovule). Angiosperm fertilization no longer relies on water like, in mosses or ferns. Production of a seed coat, which forms a layer of protection around the embryo, and a further protective layer, the fruit, are unique features of angiosperms. PMCs, pollen mother cells, MMC, megaspore mother cell.

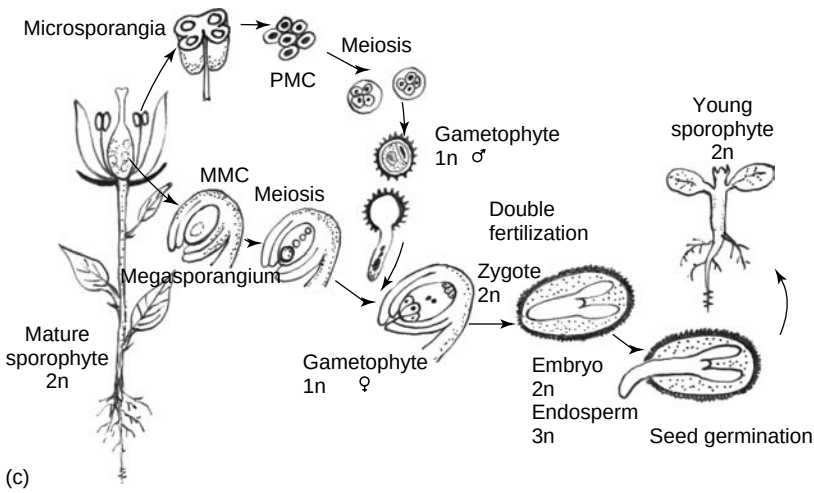

(c)

Fig. 1 (*Continued*)

the inflorescence apical meristem simultaneously with the first floral meristems. The later one proposes that cauline leaves and the axillary meristems are products of the "interphase" apical meristem. Recent evidence suggests that both genetic and environmental factors control the process and favor the latter possibility under normal conditions, suggesting the existence of coflorescence-producing *inflorescence interphase* (I phase) between the *vegetative*, leaf-producing *phase* (V phase) and the *reproductive*, flower-producing *phase* (F phase). According to this model, not two but three developmental switches would occur during flowering induction in *Arabidopsis*.

Induction of flowering is controlled by *flowering time genes*, which control the activity of *meristem identity genes*. There are two subclasses of the meristem identity genes in *Arabidopsis*: *shoot meristem identity genes* and *floral meristem identity genes*. The activity of shoot meristem identity genes (e.g. *TERMINAL FLOWER1* (*TFL1*)) specifies the inflorescence apical meristem as indeterminate and nonfloral. The second subclass, floral meristem identity genes (e.g. *LEAFY* (*LFY*) and *APETALA1* (*AP1*)), mediate transformation of the lateral ends of the inflorescence apical meristem into floral meristems, where *floral organ identity genes* are activated (Fig. 2). The floral organ identity genes promote differentiation of floral organ primordia and induce formation of floral organs by activation of organ building genes. Floral organs in *Arabidopsis* (i.e. *sepals, petals, stamens,* and *pistil*) are organized in four concentric whorls (see Fig. 3). Formation of different organs in each of the whorls reflects interactions of spatially specific gene activities according to the so-called *ABCE model* (Fig. 4) as described below (Sect. 2.2.4.1).

Developmental processes leading to flower formation could be divided into four basic steps occurring in a temporal sequence: (1) switch from vegetative to reproductive growth leading to inflorescence meristem formation, (2) floral meristem identity acquisition by activation of floral meristem identity genes, (3) activation of floral organ identity genes, and finally, (4) activation of flower "organ building" genes. Molecular control of all of these

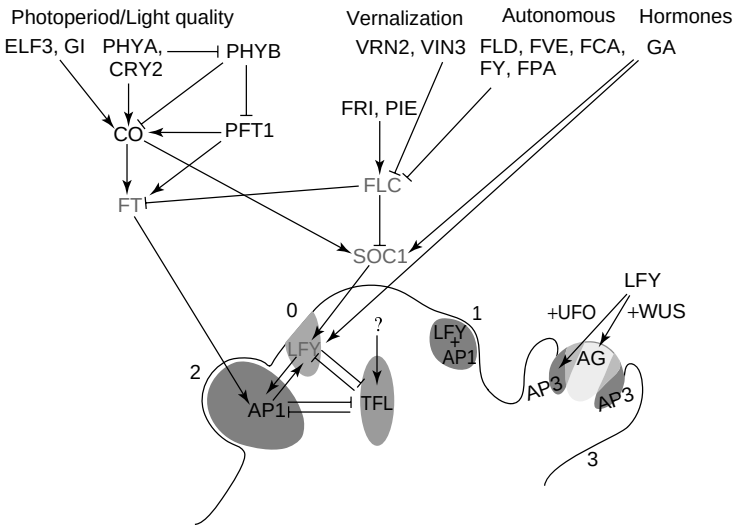

Fig. 2 Schematic representation of gene interactions during flower development in *Arabidopsis*. Selected genes acting in four major pathways of flowering induction and their interactions are shown. Both direct and indirect as well as positive (arrows) and negative interactions are marked; floral pathway integrators are in red. PHYA and CRY2 have similar effects on flowering, but they act via independent pathways. However, for the sake of simplicity, they are depicted as sharing the same one. The colored areas in the schematically depicted inflorescence apical meristem represent approximate areas of the respective gene activities (expression patterns): *AP1* (red), *LFY* (blue), *TFL* (green), *AP3* (brown), *AG* (yellow). *LFY* is first activated in the so-called floral anlagen (floral stage 0) and it precedes the expression of *AP1*. Later on, the expression patterns of *AP1* and *LFY* are overlaid (stages 1 and 2). On the left-hand side, developing floral meristem (stage 2) is shown, on the right-hand side, there are developing flowers at stages 1 and 3. Numbers label the individual stages of flower development (0–3). Overlapping activities of B and C gene classes *AP3* and *AG*, respectively, in the whorl 3 are schematically shown (light brown). At that stage of flower development, expression patterns of *UFO* and *AP3* are similar; *LFY* is expressed all over the developing flower at stage 3 (not shown). *AP1* is expressed all over the flower meristem at stages 1 till early stage 3 and after *AG* induction at stage 3, *AP1* expression is repressed in *AG* domain (whorls 3 and 4), but persist in whorls 1 and 2 (not shown) (See color plate p. xxiii).

processes integrates outputs of many signaling pathways.

2.2.1 Transition from Vegetative to Reproductive Growth

The switch from vegetative to reproductive growth is regulated by flowering time genes via four major signaling pathways: *long-day photoperiod/light quality* and *vernalization pathways*, which mediate the influence of environmental factors and the intrinsic *autonomous* and *hormone-dependent pathways*.

2.2.1.1 **Long-day period and light quality.** *Arabidopsis* is a facultative long-day period flowering plant, which means that *Arabidopsis* comes into flower more rapidly

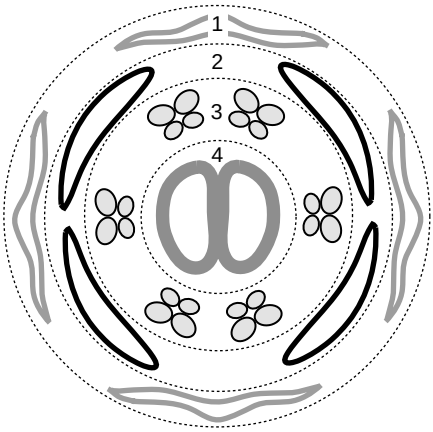

Fig. 3 Scheme of the Arabidopsis flower. *Arabidopsis* flower organs are organized in four concentric whorls. Whorl 1 contains four *sepals*, whorl 2 four *petals*, whorl 3 six *stamens* (four medial and two lateral), and whorl 4 two fused *carpels*, which form the female reproductive organ, the *gynoecium* (pistil).

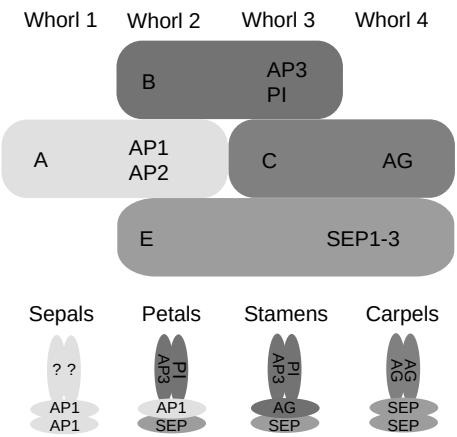

Fig. 4 ABCE model: the overlapping spatial specificity of individual gene classes in the flower whorls 1–4 in *Arabidopsis* is depicted. In the lower part, there are putative complexes of the flower organ identity genes, whose interactions are supposed to be involved in the acquisition of floral organ identity according to the quartet model (According to Jack, T. (2004) Molecular and genetic mechanisms of floral control, *Plant Cell* **16**(Suppl.), S1–S17).

when the light period reaches 14 to 16 h. However, even under short day conditions (light 8–10 h), flowering is induced, suggesting the existence of flowering promotion pathways that are independent of the length of the day period (see the following). Molecules involved in the perception of the day length are light receptors (e.g. *PHYTOCHROME A* (PHYA) and *CRYPTOCHROME2* (CRY2)), and components of the circadian clock (e.g. *GIGANTEA* (GI) and *EARLY FLOWERING3* (ELF3)). Both of these groups of molecules cooperate by the detection of the day length period and the information is transmitted via cyclic expression of *CONSTANS* (*CO*), encoding a putative component of a transcription activation complex that regulates the activity of downstream target genes.

Apart from day length, quality of light also influences initiation of flowering. Under shading, a series of responses known as "shade-avoidance syndrome," including stem elongation and accelerated flowering can be observed. In crowded plant populations, the light is depleted in red wavelengths by chlorophyll of neighboring plants and consequently the ratio of red to far-red light is reduced. In general, red light represses flowering in

Arabidopsis, whereas far-red light has a flowering promotion effect. The change in light quality is perceived by red/far-red receptors, so-called phytochromes. PHYTOCHROME B (PHYB) acts predominantly in red-light perception and mediates inhibition of flowering via repression of *PHYTOCHROME AND FLOWERING TIME1* (*PFT1*), encoding a nuclear protein and putative transcriptional factor (see also Sect. Integration of flowering promotion signals). PFT1 might promote flowering by indirect activation of both *CO* and *FLOWERING LOCUS T* (*FT*, see Fig. 2). PHYA acts primarily in perception of far-red light and together with the blue/ultraviolet A light receptor CRY2, activates flowering. Both PHYA and CRY2 repress *PHYB* and activate *CO* via independent pathways.

2.2.1.2 **Vernalization.** Vernalization, that is, extended cold treatment, mimics wintering under natural conditions and induces flowering in most of the *Arabidopsis* accessions in a quantitative manner. It was found that vernalization leads to changes in chromatin structure or DNA methylation state and subsequent transcriptional inactivation of one of the major floral pathway integrators *FLOWERING LOCUS C* (*FLC*, see the following). After cold treatment, *FLC* repression persists even at subsequent higher growing temperatures. This "memory of vernalization" involves the participation of VERNALIZATION2 (VRN2), a putative Polycomb-group protein. Another molecule necessary for the negative regulation of *FLC* by vernalization is VERNALIZATION-INSENSITIVE3 (VIN3), a plant homeodomain protein, which seems to be involved in the measurement of the cold period.

2.2.1.3 **Autonomous flowering promotion pathway.** The autonomous pathway controls flowering time independently of environmental factors, that is, vernalization and photoperiod/light quality. The central role in the integration of different flowering promotion pathways (see the following) is played by *FLC*. *FLC* encodes a MADS-box transcription factor (for more details on MADS-box proteins see Sect. Acquisition of Floral Organ Identity) and the level of *FLC* product is negatively correlated with the time of the flowering induction. Members of the autonomous pathway induce flowering by repression of *FLC*. *FLOWERING LOCUS D* (*FLD*) encodes a protein similar to the component of human histone deacetylase complex and it is supposed to act in the transcriptional control of *FLC* by histone H4 deacetylation. Another member of the autonomous pathway is FVE, a putative retinoblastoma-associated protein, and like FLD regulates *FLC* activity by histone H4 deacetylation. FCA, a protein with two RNA binding domains and a protein binding domain is supposed to interact with FY, a protein with similarity to yeast RNA processing proteins. FCA/FY complex might be involved in posttranscriptional repression of *FLC*. Similarly, *FPA*, another gene with a negative effect on the accumulation of FLC, encodes an RNA binding protein. However, the role of RNA processing in the regulation of *FLC* remains unclear.

2.2.1.4 **Plant hormones.** The plant hormone *gibberellic acid* (GA) promotes flowering in *Arabidopsis* under short day conditions. In the absence of GA, *Arabidopsis* is an obligate long-day flowering plant. Genes involved in the biosynthesis (e.g. *GA REQUIRING1* (*GA1*)) and perception of GA (e.g. *FLOWERING PROMOTIVE FACTOR* (*FPF1*)) are known to be involved

in the gibberellin-dependent regulation of flowering time in *Arabidopsis*. *Abscisic acid* (ABA), another plant hormone, plays a role opposite to that of GA in flowering promotion, that is, it delays flowering. Mutants with lowered biosynthesis of abscisic acid and mutants in genes involved in abscisic acid signaling (e.g. *ABA INSENSITIVE1* (*ABI1*) and *ABI2*) flower earlier under noninductive conditions. Other plant hormones (e.g. cytokinins and ethylene) are also implicated in the regulation of flowering induction; however, the molecular mechanism of their action in flowering induction and interaction with other flowering promoting factors is largely unclear.

2.2.1.5 **Integration of flowering promotion signals.** All of the above-mentioned signals are integrated in a complex network (Fig. 2), leading to the final regulation of downstream targets, the meristem identity genes (e.g. *LFY, AP1, TFL1*). There is quite a small group of factors in *Arabidopsis*, integrating multiple pathway outputs, so-called *floral pathway integrators* (e.g. *FLC*, *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* (*SOC1*), *FT* and *LFY*, see above and Fig. 2). *LFY* plays a special role as one of the flower pathway integrators and together with *AP1* belongs to the floral meristem identity genes (see below). *FLC* integrates outputs mediated by both flowering repressing factors (e.g. *FRIGIDA* (*FRI*), *PHOTOPERIOD INDEPENDENT EARLY FLOWERING1* (*PIE1*)) as well as flowering promoting factors (e.g. *FLD*, *FCA*, see above and Fig. 2). *Arabidopsis* accessions with functional *FRI* allele over winter in the vegetative state and after vernalization they flower in spring. Allelic variations found in *FRI* correspond to the variations in flowering time found among different accessions of *Arabidopsis*. *FRI* encodes a protein, which is suggested to be involved in the interactions with DNA or other proteins and FRI inhibits flowering by upregulating *FLC*. This effect cannot be overcome even by the strong flowering promoting effect of a long-day photoperiod. PIE1, a protein with similarity to ATP-dependent chromatin remodeling proteins, also delays flowering by positive regulation of *FLC*. *SOC1*, encoding a MADS-box protein and a flowering activator, is directly repressed by FLC via a MADS binding sequence in the *SOC1* promoter. *SOC1* is also activated by CO and gibberellic acid. The antagonistic effects of FLC and CO on *SOC1* are mediated by the same 351-bp promoter sequence of *SOC1*. *LFY* integrates the outputs from long-day period and GA pathways by separated *LFY* promoter elements. The effects of long-day period on *LFY* might be mediated by SOC1 or another MADS-box protein, AGAMOUS-LIKE24 (AGL24). GA might influence *LFY* via activation of *AtMYB33*, encoding an MYB-like transcription factor that binds to the 8-bp sequence identified in the *LFY* promoter. Finally, *FT* integrates positive control from the long-day photoperiod pathway mediated by *CO* and negative control from multiple pathways integrated by *FLC*. The inhibition of flowering under shading by the PHYB pathway mentioned in Sect. 2.2.1.1 is also integrated by *FT*, as PHYB downregulates *FT* via repression of *PFT1*, encoding a nuclear activator of *FT* expression. Expression of *FT* could also be affected by GA; that process might be repressed by EARLY BOLTING IN SHORT DAYS (EBS), a protein suggested to be involved in chromatin remodeling. For a scheme summarizing the above-mentioned interactions, see Fig. 2.

2.2.2 Floral Meristem Identity Acquisition

The final downstream targets of the above-mentioned interactions of flowering

time genes are *LFY* and *AP1*, the floral meristem identity genes. The regulation of *TFL1*, a member of the shoot meristem identity group of genes, is still unclear. *LFY* and *AP1* are necessary to specify a meristem as floral, but they do not act independently of each other. *LFY* directly activates *AP1* and *AP1* is necessary for the function of *LFY* in floral promotion. There is a mutual repression between *LFY* and *AP1* on one hand and *TFL1* on the other. This is consistent with the apparently opposite role of both groups of the genes. *TFL1* maintains the shoot meristem indeterminate and nonfloral and negatively regulates flowering, whereas *LFY1* and *AP1* promote the formation of floral meristem and flowering. LFY1 and AP1 are transcription factors, which might directly regulate *TFL1*. TFL1 is similar to FT (see above), but the molecular mechanism of their opposite functions is unclear.

2.2.3 Flower Formation

2.2.3.1 Acquisition of floral organ identity. Individual flowers are formed by floral meristems growing out from the lateral parts of the inflorescence meristem. Once flower meristem identity is acquired, the prerequisite for proper flower development is spatially delimited expression of flower organ identity genes. This finding led to the postulation of the so-called *ABCE model*, which explains the proper formation of *Arabidopsis* floral organs as a result of interactions among ABCE gene classes (see Fig. 4). *Arabidopsis* floral organs are organized in four concentric whorls. There are four sepals in the most outer whorl 1, four petals in whorl 2, six stamens (four longer medial and two shorter lateral) in whorl 3 and two carpels in whorl 4; the two carpels fuse to form a bilaterally symmetrical gynoecium (pistil, see Fig. 3). The ABCE model postulates that activity of class A genes alone specifies sepals in a flower whorl 1, A + B specifies petals in whorl 2, B + C specifies stamens in whorl 3 and C alone specifies carpels in whorl 4. Genes of class E were additionally found to be necessary for the proper identity of petals, stamens, and carpels (see Fig. 4). Function of ABC genes is specific to the flower meristem; however, in combination with class E genes, they are able to transform even vegetative organs (e.g. leaves) into floral organ-like structures. With the exception of APETALA2 (AP2), a member of a small plant-specific family of transcription factors, products of ABCE genes (see Table 1) are MADS-transcription factors. The MADS-transcription factors are analogs of animal homeotic genes, which, instead of homeodomain, contain the so-called *MADS-box*, found in MCM1 (yeast), AGAMOUS (*Arabidopsis*), DEFICIENS (*Antirrhinum*), and SERUM RESPONSIBLE FACTOR (SRF, human). Additionally, there are D-class genes [e.g. *SEEDSTICK* (*STK*), *SHATTERPROOF1* (*SHP1*), and *SHP2*], which are important for specification of ovule organ identity (see Sect. Ovule identity).

2.2.3.2 Activation of floral organ identity genes. The major activators of floral organ identity genes are *LFY* and *AP1*. As mentioned earlier, *LFY* and *AP1* are flower meristem identity genes, which are necessary for flower meristem identity acquisition. In combination with other factors, LFY and AP1 also act as activators of the flower meristem identity genes *AGAMOUS* (*AG*, C class) and *APETALA3* (*AP3*, B class). LFY and UNUSUAL FLOWER ORGANS (UFO) coactivate *AP3*: LFY is able directly to bind to the *AP3* promoter *in vitro*; UFO is an F-box

Tab. 1 Genes involved in the Arabidopsis flower development.

Gene group	*Gene subgroup*	*Function*	*Selected members of the subgroup*
Flowering Time Genes	Long-day period/light quality pathway	Regulation of flowering by day period and light quality	*PHYA*, *PHYB*, *CRY2*, *GI*, *ELF3*, *CO*, *PFT1*
	Vernalization pathway	Induction of flowering by cold	*VRN2*, *VIN3*
	Autonomous pathway	Induction of flowering independent of environmental conditions	*FCA*, *FY*, *FLD*, *FVE*, *FPA*
	Hormone-dependent pathway	Gibberellic acid and abscisic acid dependent regulation of flowering	*GA1*, *FPF1*, *AtMYB33*, *EBS*, *ABI1*, *ABI2*
	Floral pathway integrators	Integration of outputs from various flowering time regulation pathways	*FLC*, *SOC1*, *FT*, *LFY*
Meristem Identity Genes	Shoot meristem identity genes	Specification of the inflorescence apical meristem as indeterminate and nonfloral	*TFL1*
	Flower meristem identity genes	Specification of the meristem as floral	*LFY*, *AP1*
Floral Organ Identity Genes	A	Specification of sepals in whorl 1, repression of class C genes in whorls 1 and 2	*AP1*, *AP2*
	B	Together with A, specification of petals in whorl 2	*AP3*, *PI*
	C	Together with class B genes, specification of stamens in whorl 3, repression of A class genes in whorls 3 and 4 and specification of carpels in whorl 4; repression of meristem proliferation	*AG*
	D	Specification of ovule organ identity	*STK*, *SHP1*, *SHP2*
	E	Cooperation with ABC genes in specification of petals, stamens, and carpels in whorls 2–4, development of sepals	*SEP1-4*

protein and it functions as a part of the SKP1-cullin-F-box (SCF) complex that targets proteins for ubiquitin-mediated degradation. A recent model suggests a role for UFO in the degradation of the putative *AP3* repressor in whorls 2 and 3 (see Fig. 4). Proper function of B-class genes in petal formation is also dependent on *AP1*, probably via AP1 activation of *UFO* by promoting *UFO* mRNA accumulation at the base of petals in whorl 2 late during flower development (stage 5, see Sect. 2.2.3.4). The central role of LFY in the activation of floral organ identity genes is further supported by its role in coactivation of class C gene *AG*. Together with the homeodomain protein WUSCHEL (WUS), LFY binds to the regulatory region of *AG* in its second intron. The expression of *WUS* in precursor cells of whorls 3 and 4 seems to be responsible for spatial specificity of the LFY/WUS-mediated *AG* activation (see Fig. 4). After *AG* is activated by LFY/WUS, AG in turn represses *WUS* in whorls 3 and 4. WUS promotes the proliferative state of meristem cells and absence of the feedback in *ag* mutants leads to the indeterminate flower phenotype.

2.2.3.3 **Interactions among ABCE genes.** There are two major types of interactions among ABCE genes during flower formation. First, they cooperate in the specification of precursor cells to form individual flower organs as mentioned in previous sections. Second, after initial activation of the floral organ identity genes, the fine tuning of their expressional patterns is achieved by mutual repressive interactions between A- and C-class genes; the territorial competition is also called *cadasteral activity*. The molecular nature of those two types of interactions will be briefly described in following sections.

Multifactorial MADS-box Protein Complexes The vast majority of ABCE genes encode type II MADS-box proteins. Besides an N-terminal MADS-box, a domain with DNA binding, nuclear targeting, and dimerization functions, there are two additional conserved domains in most of the ABCE proteins: a K-domain, responsible for protein–protein interactions and a C-domain, involved in transcriptional activation. The recent *quartet model* suggests formation of tetramers, composed of two pairs of MADS-box proteins, homo- or heterodimers, each of which binds to a single MADS binding site. There are two possible models, explaining the molecular mechanisms of functionality of those complexes. One suggests synergistic effects of individual subunits on the affinity of the entire complex to individual MADS binding sites, the other considers functional complementation of MADS proteins lacking transcription activating domains (e.g. AP3) by the transcription activating domain of interacting partner(s) (e.g. AP1). The quartet model implies formation of tetramers consisting of AP3/PISTILLATA (PI) and SEPALLATA (SEP)/AP1 heterodimers in petals whorl 2 (there are four members of the *SEP* gene family in *Arabidopsis*, SEP1-4), AP3/PI and SEP/AG heterodimers in stamens whorl 3 and AG/AG and SEP/SEP homodimers in whorl 4 (see Fig. 4). Whether AP1 alone (possibly as a homotetramer) or in combination with any other product of a class A gene is responsible for sepal specification in whorl 1 remains unclear. Similar interactions among floral organ identity genes were found in *Petunia* and *Antirrhinum*. It could suggest that the above-mentioned principles might be a unifying dogma of flower organ development, with some species-specific modifications, for example, only class B

and C genes were identified in *Antirrhinum*. However, it should be noted here that the *in planta* nature of MADS–protein complexes remains to be characterized.

Cadasteral activity of A and C genes *AP1* and *AP2* are the earliest class A genes expressed in the floral meristem. *AP1* is expressed throughout the meristem to specify the meristem as floral. Later on (stage 3), the expression of the class C gene *AG* is activated by LFY and WUS (see Sect. 2.2.3.2) in whorls 3 and 4, where the expression of *AG* suppresses transcription of *AP1*. Vice versa, *AP2*, another class A gene, inhibits expression of *AG* in whorls 1 and 2. Interestingly, in the case of *AP2*, posttranscriptional regulation via microRNA (miRNA) leads to the localization of AP2 protein to whorls 1 and 2 (*AP2* mRNA is detectable throughout the flower). Besides class A genes, there are also other, non-MADS genes, working as repressors of *AG* (e.g. *CURLY LEAF* (*CLF*), encoding a homolog of the *Drosophila* Polycomb-group protein *Enhancer of Zeste*, *LEUNIG* (*LUG*), which encodes protein with multiple WD-repeats, *SEUSS* (*SEU*), encoding a protein with a glycine-rich domain and showing similarity to animal LIM domain binding coregulators, *STERILE APETALA* (*SAP*) and *AINTEGUMENTA* (*ANT*), encoding DNA binding protein related to AP2, see also Sect. Ovule patterning). Like *AP2*, *LUG*, *SEU*, and *ANT* repress *AG* in whorls 1 and 2.

2.2.3.4 **Development of floral organs.** Development of the *Arabidopsis* flower was classified into 12 stages, based on the successive appearance of individual landmarks. Stages 1 to 3 are schematically depicted in Fig. 2. Flower development starts when a flower buttress arises (stage 1) and proceeds to the next stage when the floral buttress is separated by grooves (stage 2). In stage 3, sepal primordia arise that overlie the primordium at stage 4. During stage 5, the petal and stamen primordia become apparent and are enclosed by the sepals at stage 6. During stage 6, the gynoecium starts growth as an open-ended tube. At stage 7, primordia of middle stamens become constricted toward their base, thus defining the future stamen stalks (filaments) and develop locules (stage 8). At stage 9, petal primordia become stalked and all organs undergo rapid elongation. When the petals reach the length of lateral stamens, stage 10 begins. Papillae of stigma appear at stage 11 and the stage 12 starts when petals reach the height of medial stamens. Stage 12 ends with the opening of the flower.

Recently, the molecular factors regulating floral organ development have been identified using expressional studies and mutant analysis. Expression of *TAP1* and *FILAMENTOUS FLOWER1* (*FIL1*) encoding a secreted protein and a candidate cell wall protein respectively were identified in anthers of *Antirrhinum*. In *Arabidopsis*, *NAC-LIKE ACTIVATED BY AP3/PI* (*NAP*) represents the downstream target of B-class genes in petals and stamens. *SHOOT MERISTEMLESS* (*STM*), *CUP-SHAPED COTYLEDONS* (*CUC1*), and *CUC2* are expressed in the so-called medial domain, located at the interface of the fused lateral carpel margins in the pistil. These genes are supposed to be important for maintaining the meristematic state of the medial domain, where later on the *placenta* tissue differentiates and from which the ovule primordia arise. *CRABS CLAW* (*CRC*), *SPATULA* (*SPT*), and *TOUSLED* (*TSL*), (see also Sect. Integument initiation and growth) are expressed in the medial domain of

the pistil and might affect the production of ovules.

2.3 Structure and Formation of Male and Female Gametes

In higher plants, male and female gametes, the *sperm* and *egg cells*, are formed in special flower tissues, so-called sporangia. *Microsporangia*, located in *anthers* (Fig. 5), produce a large number of *microspores* that differentiate into mature *male gametophytes* (*microgametophytes*), the *pollen grains*. *Megasporangia*, so-called *ovules*, are located in the ovary of the pistil (gynoecium). In contrast to microsporangia, during megasporogenesis in a majority of plant species, only one of four megaspores differentiates into a functional *megaspore* that develops into a *female gametophyte* (*megagametophyte*), the *embryo sac*. Two *sperm cells* develop from each pollen grain, whereas there is only one *egg cell* in each embryo sac. Both micro- and megagametophytes (i.e. pollen grains and embryo sacs) are multicellular structures that undergo specific developmental processes.

2.3.1 Male Gamete Formation

2.3.1.1 Microsporogenesis.

Archesporial cell number and differentiation The formation of pollen in *A. thaliana* is schematically shown in Fig. 5. In each of four developing locules (microspororangia) formed in the anther, one *archesporial cell* differentiates. The archesporial cell is specified from one of the subepidermal cells originating

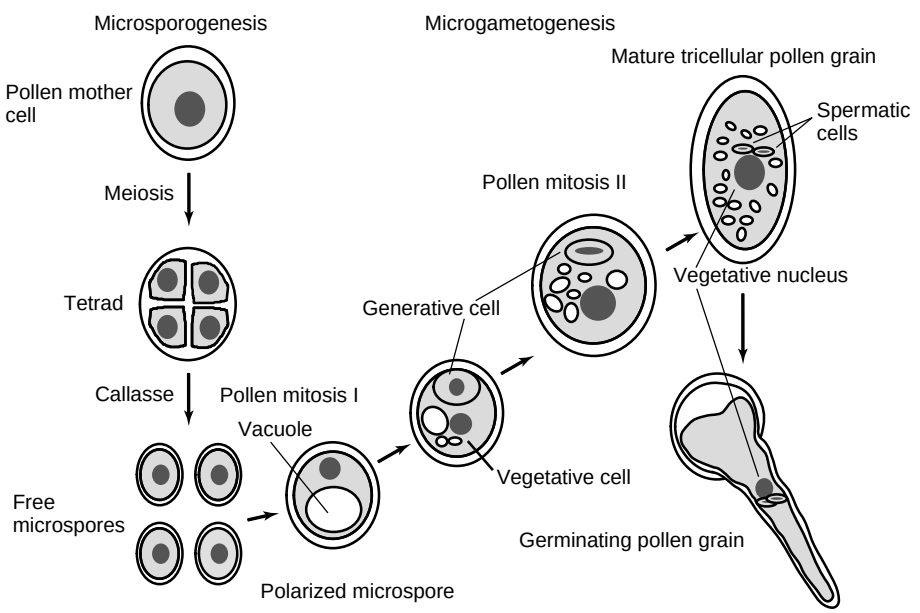

Fig. 5 Male gametophyte development: formation of mature pollen grain from pollen mother cell in *Arabidopsis* is schematically depicted. Cytoplasm is in yellow, nuclei are in red, and vacuoles are in white. For details see text (Sect. 2.3.1). (See color plate p. xxvii).

from the L2 layer. In *Arabidopsis NOZZLE/SPOROCYTELESS* (*NZZ/SPL*) encodes a putative transcription factor similar to MADS-box proteins, necessary for the specification of archesporial cell. *NZZ/SPL* was recently shown to be under direct control of AG (see also Sect. 2.2.3.2). The number of archesporial cells in *Arabidopsis* seems to be determined by the activity of *EXTRASPOROGENOUS CELLS/EXCESS MICROSPOROCYTES1* (*EXS/EMS1*) encoding a putative Leu-rich receptor (LRR) kinase; in *exs/ems1* plants, multiple archesporial cells per microsporangium are formed. A similar phenotype was found in mutants in *TAPETAL DETERMINANT1* (*TPD1*), encoding a putative secreted protein. Whether TPD1 is involved in the same signaling pathway, perhaps as a ligand of EXS/EMS1, remains to be clarified. The ortholog of *EXS/EMS1 MULTIPLE SPOROCYTE1* (*MSP1*) was identified in rice and mutants in maize *MULTIPLE ARCHESPORIAL CELLS1* (*MAC1*) also resemble *exs/ems1* plants. The archesporial cell divides to form a *primary parietal cell* and a *primary sporogenous cell*. The primary parietal cell divides to give rise to the *tapetum, endothecium*, and the *middle layer* of the anther. The primary sporogenous cell divides and the progeny differentiate into *microsporocytes – pollen mother cells* (PMCs) that undergo meiosis, leading to the formation of tetrads of haploid *microspores*.

Meiosis The basic processes during meiotic division I and II in plants do not seem to differ from other eukaryotes. However, in contrast to animals, there are differences in some of the cellular processes accompanying meiosis. In plant microsporocytes, the reduction of rRNA and mRNA together with plastid and mitochondria dedifferentiation and division were observed. The reason for that "purging of sporophytic information" is unclear. It might facilitate later gametophytic development or free the developing microspores of detrimental RNA-like viruses and silencing elements.

There are many genes found to affect plant meiosis, but only a few of them have been cloned and characterized in detail. Mutants in *SWITCH* (*SWI1*)/*DYAD* are defective in both male and female meiosis. Interestingly, male meiosis is unaffected in *swi1-1* and *dyad* mutant alleles, which might suggest distinct processes involved in the male and female germlines during meiosis in *Arabidopsis*. *SWI1* was cloned, but the putative amino acid sequence of SWI1 does not reveal similarity to any known protein. On the basis of the phenotype of *swi1* male meiocytes (absence of bivalents during meiosis), the protein is suggested to be involved in establishment of chromatid cohesion and chromosome structure during meiosis. *AtDMC1*, the *Arabidopsis* homolog of the yeast *DMC1* gene was identified and its expression was found to be specifically associated with meiosis in both PMCs in anthers and megaspore mother cells (MMC) in ovules. DMC1 in yeast, the homolog of *E. coli* recA, plays a role in homology-dependent pairing of chromosomes and chiasmata formation during meiosis. The *Arabidopsis Atdmc1* mutant shows defects in meiosis; however, in contrast to mouse and yeast, meiosis is not completely arrested; ten univalents instead of five bivalents were observed in *Atdmc1* mutant meiocytes. The absence of bivalents results in a formation of nonviable gametes with an aberrant number of chromosomes. The occurrence of double-strand breaks (DSB) is a consequence of

the *dmc1* mutation in yeast. However, the presence of DSB was not observed in *Arabidopsis dmc1* mutants. This might be explained by either efficient DNA repair systems or that DSBs do not occur in *Arabidopsis* mutant meiocytes. *arabidopsis homolog pairing2* (*ahp2*) represents another recently identified *Arabidopsis* meiosis-defective mutant. The *ahp2* mutant plants are sterile, with defects in both male and female gametophyte development. AHP2 is a homolog of the fission yeast MEU 13^+. MEU 13^+ is supposed to be involved in chromosome pairing during early prophase I and monitoring of the sequence homology of the paired chromosomes to either promotes homolog pairing or destabilizes pairing between nonhomologous ones. In good concert with that, the *ahp2* male meiocytes are characterized by the absence of bivalents during prophase I and by the presence of chromosome fragments, chromatin bridges, and unbalanced segregation in anaphase I and II.

Microspore cell wall formation After each meiotic division in species with successive cytokinesis (e.g. most of monocots) or after meiosis II in species with simultaneous cytokinesis (e.g. most of eudicots including *Arabidopsis*), the cell wall is formed between individual microspores. The cell wall is composed of *callose*, a β-1,3-glucan. Control of the division plane in plant male meiosis is achieved by an array of radial microtubules surrounding newly formed nuclei. At the interface of those microtubule arrays, the callose wall is formed either centrifugally (in case of successive cytokinesis) or centripetally (in simultaneous cytokinenesis). In *Arabidopsis*, *TETRASPORE/STUD* (*TES/STD*) gene was identified, encoding a putative kinesin. In *tes/std* mutants, formation of radial microtubule arrays at the end of meiosis is hampered and all the microspores develop in a common cytoplasm. In *tes/std* plants, few functional pollen grains are produced revealing different aberrations, for example, polyploid sperm cells. Interestingly, the recessive nature of *tes/std* mutation, together with the expression of *TES/STD* throughout the premeiotic anther, suggests sporophytic control of meiotic cytokinesis in *Arabidopsis*. Recently, the *Arabidopsis* gene *GLUCANSYNTHASE-LIKE5* (*AtGSL5*) with similarity to yeast β-1,3-glucan synthase was identified that is supposed to be involved in callose wall formation.

Individual microspores are released by the action of the enzyme *callase*, comprising endo- and exoglucanase activities, which is produced by the tapetum. This step was shown to be essential for male fertility in some species (e.g. *Petunia*). *Arabidopsis* gene *A6* is expressed in tapetum and encodes a protein similar to yeast β-1,3-glucanase and represents a candidate that may function in microspore release. Nevertheless, callose degradation was shown to be necessary but not sufficient to release the microspores in *Arabidopsis*. During microsporogenesis in *quartet* mutants, in spite of normal callose wall formation and dissolution, permanent tetrads are formed, probably as a result of persistent pectic components in the parent cell wall of the tetrad. However, in contrast to *Arabidopsis*, other species (e.g. members of *Juncaceae* and *Oenotheraceae*) produce permanent tetrads that are fully functional.

2.3.1.2 **Microgametogenesis.** Formation of haploid microspores completes *microsporogenesis,* which is followed by

microgametogenesis. The individual microspores enlarge and a single large vacuole is formed, which is associated with polarization of the microspore that subsequently undergoes two rounds of mitosis (see Fig. 5). During the first, asymmetric cell division also called *pollen mitosis I*, the nucleus is attached to the cell wall and the spindle orientation leads to the formation of a large vegetative and a small generative cell. The nucleus of the vegetative cell generally maintains irregular shape with dispersed chromatin, while the nucleus of the generative cell is highly condensed, often spindle shaped. The generative cell detaches from the cell wall (leading to the formation of the "cell in cell" structure) and undergoes a further round of mitosis, *pollen mitosis II*, which gives rise to the mature, three-celled male gametophyte consisting of the large *vegetative cell* and two small *sperm cells*. In species with bicellular pollen (e.g. tobacco), pollen mitosis II occurs after pollen grain germination within the pollen tube (see Sect. 2.4). In the *Arabidopsis* mutant *gemini pollen1* (*gem1*), microgametogenesis is arrested after the first mitotic division. The initial microspore polarization and phragmoplast formation seems to be affected, leading to the formation of two cells that are both able to express vegetative cell-specific markers. *GEM1* encodes a microtubule-associated protein, which was shown to bind microtubules *in vivo* and is possibly involved in the correct positioning of the phragmoplast during pollen mitosis I. The vegetative nucleus and two sperm cells after pollen mitosis II are physically connected, forming a functional assemblage, the male germ unit (MGU). The connection of vegetative nucleus is disrupted in male gametophytic mutants of two classes, *germ unit malformed* (*gum*) and *germ unit displaced* (*gud*). The penetrance of both of these male-specific gametophytic mutations is complete, suggesting functional importance of the MGU.

Pollen grain wall formation The wall of pollen grains is a complex, multilayered structure consisting of the innermost pectocellulosic *intine* and outer *exine*. The outer exine is further divided into the two-layered *nexine* (nexine I and II) with nexine I adjacent to the intine, and outer *sexine* that consists of *columella* and the outermost *tectum*. The exine is composed of sporopollenin comprising long-chain fatty acids and a minor component of phenolic compounds. Sporopollenin is polymerized from precursors secreted by the microspore and (after microspore is released from tetrad) by the tapetum. The exine forms a rigid, 3D-coat structure, providing protection against physical, chemical, and biological attacks and allowing attachment to animal pollinators and stigma surface. The exine does not develop over the entire surface of the pollen grain, creating precisely positioned germination apertures or lacunae. The number and positions of apertures could be used as taxonomic characters, for example, in fossil records. During pollen wall development, the individual layers of the pollen wall are added from the outermost layer inwards, so that the ephemeral callose wall is first deposited on the surface of the developing pollen, followed by the primexine (a precursor of the sexine), the nexine, and finally the intine. There is some evidence that genes involved in pollen wall patterning are transcribed premeiotically. However, there are very few mutations identified up to now that affect the patterning of pollen wall formation. In *Arabidopsis*, the *DEFECTIVE IN EXINE FORMATION1* (*DEX1*) gene was identified, encoding a putative membrane-associated calcium-binding protein, which

is proposed to act as a nucleation site for sporopollenin polymerization.

2.3.1.3 **Interactions of mitochondrial and nuclear genes during male gametophyte development.** The mitochondria seem to play a key role in male gametophyte development. There are many loci identified in the mitochondrial genome that are associated with so-called *cytoplasmic male sterility* (CMS). The defects cosegregating with CMS loci occur in different stages of the male gametophyte development and range from anther degeneration and a lack of fertile pollen (e.g. *Petunia*) to absence of exerted anthers (e.g. maize) and homeotic transformation of anthers into petal- or bract-like structures (e.g. carrot). Interesting feature of the CMS phenomenon in plants is the existence of nuclear located genes, so-called *fertility restorers* (referred to as Rf of Fr). Specific alleles of these genes are able to restore male fertility in CMS lines, suggesting importance of nuclear to mitochondrial genome communication during male gametophyte development. Molecular nature of this communication remains to be elucidated. However, in the last few years, substantial progress has been achieved in understanding the CMS phenomenon in plants.

Mitochondrial loci associated with CMS and mechanism of their action Most of the loci identified in CMS are the result of (often multiple) recombination events in the mitochondrial genome, leading to the formation of new open reading frames (ORFs) that could lose regulatory association with their original regulatory sequences. In CMS loci identified up to now, genes for ATP synthase subunits were often found to be involved in such recombination events. As an example, one of the best-understood CMS systems in maize could be mentioned in detail. In mitochondrial genome of maize CMS-T (Texas) male sterile line, the existence of *urf13* gene was identified, encoding a 13-kD protein. The locus is a result of multiple recombination events, leading to the fusion of primarily noncoding sequences normally found in 3′ to 23S rRNA gene that now comes under the control of a regulatory sequence of the gene encoding the 6th subunit of ATP synthase (*atp6*). At the 3′ proximity to *urf13*, cotranscribed *orf221* gene was identified, encoding a protein similar to *Arabidopsis* F_0 subunit of ATP synthase. In addition to ATP synthase subunits 4, 6, 8, and 9 coding sequences, other known sequences (e.g. cytochrome oxidase subunit coding regions) and sequences with unknown function were identified in various CMS-associated loci. The molecular nature of the defects caused by CMS-associated loci is mostly unknown. It has been found that many of the CMS-associated loci encode putative transmembrane domains. A recent hypothesis suggests that mitochondrial membrane could be a potential target for those proteins, as it has been shown in case of URF13, which forms a pore in the mitochondrial inner membrane in the presence of T-toxin produced by fungus *Cochliobolus heterostrophus*. Whether there is a naturally occurring biosynthetic product analogous to fungal T-toxin in young anthers of maize is not known.

Fertility restorers, nature, and mechanisms of action The nuclear genes found to act as fertility restorers are mostly loci found to be involved in regulation of mitochondrial gene expression; however, there are also genes that seem to work through metabolic effects. The regulation of mtDNA expression could be achieved

at both transcriptional and posttranscriptional levels, either by regulation of the mitochondrial transcription profile or protein accumulation. Good examples are *Rf1/Rf2* genes found in the restoration of the above-mentioned CMS-T in maize. *Rf1* acts as a primary CMS-T restorer and affects processing of *urf13* transcripts. *Rf2* is unable by itself to restore fertility in CMS-T lines, but it is supposed to remove residual sterility effect of *urf13* in the presence of *Rf1*. *Rf2* encodes a protein with aldehyde dehydrogenase activity and the role of RF2 could be to reduce toxic levels of aldehyde, potentially present as a result of residual *urf13* expression. In the sequence of other cloned fertility restorer genes, a pentatricopeptide (PPR) motif was often identified (e.g. *Rf* genes from *Petunia*, *Brassica*, *Raphanus*, and *Oryza*). PPR proteins were implicated in organelle gene expression in both mitochondria and chloroplasts and they seem to have different functions, for example, in RNA processing and translation or accumulation of plastid ribosomes.

2.3.2 Female Gamete Formation

There are more than 15 different patterns in the development of female gametophyte in angiosperms based on the variations in cytokinesis during meiosis, the various number of functional megaspores and the number and pattern of subsequent mitotic divisions and cellularization patterns. In more then 70% of flowering plant species, the *monosporic, Polygonum type* of female gametophyte occurs. This type of female gametophyte development can be found in many biologically and economically important species like *Zea* (maize), *Triticum* (wheat), *Gossypium* (cotton), Brassica (e.g. cabbage, kohlrabi, cauliflower, rape, mustard), and *Arabidopsis*. The basics of the *Polygonum* type of the female gametophyte development will be summarized here with an emphasis on understanding the entire process at the molecular level.

2.3.2.1 **Ovule development.** In angiosperms, the key structure where the pre- and postzygotic development takes place is the *ovule*. The female gametophyte is represented by the so-called *embryo sac*, the multicellular, mostly haploid and highly differentiated tissue, surrounded by diploid mother (sporophytic) tissue of the ovule. In *Arabidopsis*, typically 60 ovules are placed along the *replum* in the female reproductive organ, the gynoecium (the pistil). Early in flower development, *ovule primordia* differentiate from the *placenta*. During ovule development, three anatomically distinguishable and functionally distinct parts of the ovule are formed (see Fig. 6). The *nucellus*, located at the distal pole of the ovule, contains the sporogenic tissue. The growth of *inner* and *outer integuments* is initiated in the *chalaza*, located approximately in the middle of the ovule primordium. Soon after integument differentiation, elongation growth is initiated and the integuments cover the nucellus of the mature ovule except a small part at the distal pole, forming a gate for the pollen tube entry, the *micropyle*. The extended growth of the outer integument at the outer (posterior) side of the ovule results in a characteristic bending and finally the placement of the micropyle along the base (the stalk) of the ovule (Fig. 6). The position of the integument outgrowth defines the most proximal part of the ovule, which differentiates into the tissue connecting the maternal placenta tissue, the *funiculus*.

Ovule identity It seems that there are parallels between ovule and flower patterning and that those two processes

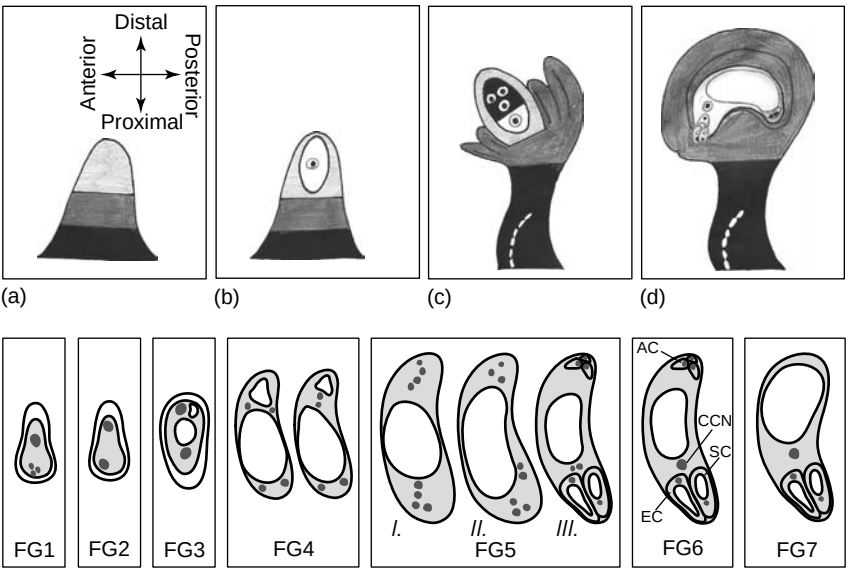

Fig. 6 Female gametophyte development. Ovule development (a–d) and individual stages of female gametogenesis (FG1–FG7) in *Arabidopsis* are schematically depicted. (a) Primary, nondifferentiated protrusion of the placenta; the proposed diversification of three future ovule regions (nucellus, chalaza, and funiculus) is depicted by different gray grades. (b) Megaspore mother cell (MMC) is differentiated. (c) Inner and outer integuments are developed; asymmetrical growth of integuments on abaxial and adaxial sides is depicted. Three of four megaspores at the distal pole will degenerate, the most proximal one will form the functional megaspore; the developing vascular tissue of funiculus is schematically shown. (d) Fully developed ovule with two synergid cells and egg cell at the micropylar (distal) pole. Drawings in a–d by Romana Dobešová (According to Grossniklaus, U., Schneitz, K. (1998) The molecular and genetic basis of ovule and megagametophyte development, *Semin. Cell Dev. Biol.* **9**, 227–238). FG1–FG7 megagametogenesis in Arabidopsis. Cytoplasm is in yellow, nuclei are in red and vacuoles are in white. EC, egg cell; SC, synergid cell; CCN, central cell nucleus; AC, antipodal cells. Only one of the two synergid cells is shown. For details of female gametophyte development see Sect. 2.3.2 (See color plate p. xxxiv).

share common factors (see also Sect. 2.2). Similar to flower organ identity acquisition, the identity of ovule primordia and ovule structures (e.g. integuments) is genetically determined and it seems to be a result of complex interactions of factors that might resemble both cadasteral and cooperative interactions found among ABCE genes during flower patterning (see Sect. 2.2.3.3). For example, the balanced activity of *AG*, the potent promoter of carpel identity in *Arabidopsis* and its homologs in *Brassica* and *Nicotiana*, was found to be necessary for proper identity acquisition of ovule primordia. AG was also shown to control *NZZ/SPL*, involved in ovule primordia patterning (see the following). In *Arabidopsis*, activity of *BELL1 (BEL1)*, encoding a homeodomain protein is crucial for the determination of ovule integument identity, possibly via negative regulation of *AG*. In this model, the AG-dependent, carpel-promoting pathway overrides effects of

integument identity genes in the absence of BEL1. BEL1, together with members of a monophyletic clade of *AG*, the MADS-box genes *SHP1, SHP2,* and *STK,* is necessary for integument identity acquisition and in the absence of these factors, integuments are transformed into carpelloid structures, possibly as a result of the above-mentioned dominance of *AG*. *SHP1, SHP2,* and *STK* have apparently redundant functions in the promotion of integument identity and *SHP1* and *SHP2* can also substitute for *AG* in carpel specification. In *Petunia*, homologs of *Arabidopsis STK FLORAL BINDING PROTEIN7* (*FBP7*) and *FBP11* share a similar expression pattern with *STK, SHP1,* and *SHP2* and in plants with lowered expression of both *FBP7* and *FBP11*, the conversion of ovules into carpel-like structures was observed. In *Arabidopsis*, the E genes *SEP1-3* collaborate with B and C genes in the acquisition of floral organ identity, probably via formation of multimeric complexes (see Sect. Multifactorial MADS-box protein complexes). Similar interactions might be involved in the ovule identity acquisition. Expression of *SEP* genes was identified in developing ovules and *SEP1/sep1, sep2, sep3* mutants show phenotype alterations similar to phenotype of *stk, shp, shp2* triple mutants (i.e. homeiotic transformation of integuments into carpelloid structures). STK, AG, SHP1, and SHP2 were shown to interact with SEP proteins in multimeric complexes, suggesting the existence of a molecular mechanism resembling those identified in acquisition of flower organ identity. In line with that, a similar situation was found in *Petunia*, where STK orthologs FBP7 and FBP11 interact with FBP2 and FBP5, putative orthologs of SEP proteins in *Arabidopsis*.

Ovule patterning The developing ovule primordium differentiates into three zones: the most distal nucellus, the central chalaza, where the integuments initiate and the most proximal funiculus. These three pattern units are distinguished very early in ovule development after the non-differentiated protrusion of the mother placental tissue appears. Recently, the first molecular evidence has become available that might explain how the supposed positional information distributed along the proximal–distal axis is interpreted into cell-fate determination. In *Arabidopsis*, the position of the inner integument outgrowth seems to be dependent on the spatial interactions of expression domains of *ANT*, encoding an AP2 domain transcription factor and *NZZ/SPL*, whose role was already mentioned in male archesporial cell differentiation (see Sect. Archesporial cell number and differentiation). The expression of *WUS*, encoding a homeodomain protein was also found to be involved in the positional specification of integument outgrowth (see also Sect. 2.2.3.2). According to a recent model, the distal expansion of proximally located expression domain of *ANT* is negatively regulated by activity of *NZZ*, and the spatial relationship between *WUS* expression in the nucellus and *ANT* in the chalaza determines the position of inner integument outgrowth. However, as *NZZ* is expressed throughout the developing ovule primordium, the spatial specificity mechanism of its negative regulation of *ANT* is so far unclear.

Integument initiation and growth Multiple factors, involved in outer integument initiation, regulation of differential integument growth, and integument extension were identified. *INNER NO OUTER* (*INO*) encoding a YABBY protein transcription

factor seems to be a specific determinant of the outer integument outgrowth. Spatial specificity of its expression depends on *ANT*, whose role was discussed in inner integument initiation (see previous section). The expression of *INO* not only seems to regulate outer integument outgrowth along the proximal–distal axis but also along the anterior–posterior (or adaxial–abaxial, recently also called *gynoapical–gynobasal*) axis (see Fig. 6). *INO* is expressed only on the posterior/abaxial/gynobasal side of the ovule, where the outer integument is initiated and exhibits extended growth, leading to the typical ovule curvature (Fig. 6). *SUPERMAN* (*SUP*) encoding a putative zinc finger transcription factor was identified to be necessary for this posterior specific location of *INO* expression domain, probably via negative regulation of *INO* positive autoregulation. However, the factors involved in the initial induction of the asymmetrical *INO* expression remain unknown. Additionally, *INO* seems to be necessary for the abaxial identity of the outer integument and its growth.

Several other factors were identified in the coordination of the inner and outer integument growth in *Arabidopsis*. For example *TSL*, encoding a Ser/Thr protein kinase (see also Sect. Development of floral organs), represents a good candidate for the molecule mediating signaling between apparently functional and evolutionarily different tissues of inner and outer integuments. Also, factors influencing growth of both inner and outer integuments were identified, for example, *DICER-LIKE1* (*DCL1*, former *SHORT INTEGUMENTS1*) encodes RNase III/RNA helicase, which might be involved in the production of small RNAs associated with the translational regulation of developmental genes during integument cell elongation.

2.3.2.2 **Megasporogenesis.** In the early phases after *Arabidopsis* ovule primordial growth initiation, an *archesporial cell* of subepidermal origin directly differentiates into the megasporocyte, the MMC. The ovule is tenuinucellate, which means that the MMC is in direct contact with the nucellar epidermis. The archesporial cell is originally established as a nonpolar, rather large cell (~17 μm in diameter), while the MMC just before meiosis is already polarized, with the majority of the organelles located at the functional pole. In a majority of species including *Arabidopsis*, the functional pole points toward the distal pole of the ovule primordium. The polarity of the megasporocyte is also reflected by the preferential accumulation of callose at the functional pole. However, the polarity is reversed in some species, for example, *Oenothera* and *Endymion*, (see also below). Activity of *NZZ/SPL* was shown to be crucial for archesporial cell differentiation in both male and female germlines (see also previous sections).

Cytokinesis in *Arabidopsis* occurs after meiosis is completed, which results in the formation of a multiplanar or linear tetrad. Molecular control of meiosis, which is mostly common for both male and female meiocytes was already discussed previously (see Sect. Meiosis). The most proximal product of the MMC meiosis survives in *Arabidopsis*, forming the *functional megaspore*. In contrast to that, in *Oenothera* and *Endymion*, the most distal megaspore survives and in some other species the surviving megaspore is not determined by position at all. The mechanisms of the functional megaspore selection are unknown, possibly both cytoplasmic determinants and

cellular interactions with neighboring cells might contribute to this.

2.3.2.3 **Megagametogenesis.** Megagametogenesis follows megasporogenesis, which ends with the formation of the functional megaspore (see previous section). Recently, megagametogenesis in Arabidopsis was described in detail. On the basis of the occurrence of anatomically distinguishable traits, the entire process was classified into seven distinct steps FG1–FG7 (see Fig. 6). During the first step, FG1 (the **F**emale **G**ametophyte 1), a functional megaspore is already differentiated; the remnants of the three degenerating meiotic products are still present. In the next development, the functional megaspore undergoes three rounds of mitosis, constituting the eight-nuclear embryo sac. In FG2, the first round of mitosis results in a two-nuclear female gametophyte, with many small vacuoles all around the cytoplasm. In FG3, the large central vacuole is formed and usually another (smaller) one at the proximal (chalazal) pole is also present. In FG4, the second round of mitosis results in the formation of a four-nuclear gametophyte, with a well-developed central vacuole. The four nuclei are organized in two pairs, located at opposite poles. The chalazal nuclei change their positions during FG4, as the line between them, originally orthogonal to the micropylar-chalazal axis, is flipped-over to become parallel with the micropylar-chalazal axis at the end of FG4. After the third (and final) round of mitosis, the eight-nuclear gametophyte is formed, defining the beginning of the most complex stage, FG5. This stage is characterized by dramatic changes in the position of nuclei; cellularization and cell differentiation takes place during FG5. FG5 was originally considered to be a rather uniform stage; however, on the basis of our results we suggested its further division into three subphases FG5 *I*–FG5 *III*. In the first subphase FG5 *I*, the embryo sac is a syncytium of eight nuclei in two groups of four, separated by the central vacuole. In FG5 *II*, the translocation of nuclei can be observed, resulting in the formation of two groups of three on both poles, and two *polar nuclei* migrating toward the position of the future central cell nucleus (CCN). The polar nucleus of chalazal origin migrates faster, which results in the asymmetric location of the polar nuclei that underlines the already preestablished polarity of the developing embryo sac. Both these stages are rather short, as the frequency of the ovules found in both FG5 *I* and FG5 *II* is very low. In FG5 *III*, cellularization and cell differentiation occurs. At the end of the phase, the two *synergid cells* (SCs) with well-developed vacuoles can be distinguished surrounding an *egg cell* on the *micropylar* (distal) *pole*. At the opposite, *chalazal* (proximal) *pole*, three *antipodal cells* (ACs) are distinguishable; the *central cell* is still not fully differentiated. The end of FG5 is defined by the fusion of polar nuclei that leads to the formation of the diploid nucleus of the central cell. In FG6, the antipodal cells undergo cell death, thus forming the four-celled mature female gametophyte (FG7), where double fertilization triggers embryo and endosperm development, giving rise to a fully functional seed (see Sects. 2.4 and 2.5). In some species (e.g. maize), the antipodal cells do not degenerate but proliferate, forming a cluster of up to 40 cells at the chalazal pole (ChP).

2.3.2.4 **Molecular nature of female gametophyte development.** In spite of the crucial importance of the female gametophyte tissue in the entire plant life

cycle, the factors necessary for proper formation and function of mature embryo sac are mostly still unclear. Just in the last decade, extensive mutation screens specifically directed to identify female gametophyte mutations were performed. On the basis of the process affected, female gametophytic mutations could be classified into six phenotypic categories and very recently, the molecular nature of the involved factors is being discovered. Interestingly, the largest group of mutants that has been defective in female gametophyte development or function seems to affect postfertilization processes. This fact suggests importance of maternal control of postfertilization processes in *Arabidopsis* and possibly also in other plant species (see also Sect. 2.5).

First gametophytic mutation with impact on female gametophyte development, which was characterized at the molecular level, is *gametophytic factor2* (*gfa2*). The *GFA2* gene encodes a J-domain containing protein targeted to mitochondria. *GFA2* was shown to partially complement the yeast mutation *mitochondrial DNAJ homolog1 (mdj1)*, a gene for a chaperone in the mitochondrial matrix and *GFA2* is supposed to be an *Arabidopsis* ortholog of the yeast *Mdj1p*. The *GFA2* gene was found to affect both female and male gametophyte. The penetrance of *gfa2* in the female gametophyte is almost complete and a substantial part of the male gametes is affected (99 and 86%, respectively). Female gametophyte development in *gfa2* embryo sacs is terminated at FG5. Synergids of the mature female gametophyte were suggested to be a source of the diffusive attractants mediating signaling that directs the pollen tube into the micropyle (see also Sect. 2.4.1). In *Arabidopsis*, cell death of one of the synergids occurs as a consequence of pollen tube entry into the micropyle or just before pollen tube arrival (for more details see Sect. 2.4.2). However, pollen tube entry did not induce synergid cell death in *gfa2* mutants. In spite of incomplete megagametogenesis that prematurely terminates at FG5, the *gfa2* mutant ovules were still able to attract pollen tubes. This might suggest functional independence of synergid cells in the final stages of megagametogenesis, at least in terms of their functionality in pollen tube attraction. The other female gametophyte mutant that is well characterized at the molecular level is *nomega*. The ovules in *nomega* mutants were found to be arrested at the FG2 phase of megagametogenesis. The *NOMEGA* gene was found to encode a homolog of the APC6/Cell Division Cycle (CDC) 16 subunit of the Anaphase Promoting Complex/Cyclosome (APC/C), which is a part of the APC/C complex, and which together with the SCF pathway represents a key regulatory component in cell cycle regulation by ubiquitin-dependent proteolysis in yeast. The APC/C complex was found to be involved in the degradation of Cyclins A and B in clam and *Xenopus*. In the *nomega* mutant embryo sacs, the absence of Cyclin B degradation was shown. The APC/C complex was also found to be involved in the earlier checkpoint of the cell cycle, the degradation of the anaphase inhibitor securin, thus facilitating the separation of chromatids. Therefore, a functional APC/C complex seems to be necessary for cell exit from mitosis. The ability of the *nomega* mutant embryo sacs to accomplish the first mitosis is explained by the residual activity of the APC/C originating from the MMC, which is sufficient for metaphase to anaphase progression. The later checkpoint allowing exit from first mitosis and leading to the formation of the two-nuclear

embryo sac might be overcome by inactivation of the CDK1 mitotic complex by phosphorylation. The absence of central vacuole formation in the *nomega* mutants remains unclear.

2.3.2.5 **Sporophyte to gametophyte communication during female gametophyte development.** In contrast to fucoid algae, where the polarity of the naked zygote is dependent on external environmental signals, embryo sac polarity in *Arabidopsis* is fully position dependent. On the basis of current genetic data, an interregional interaction between the developing female gametophyte and surrounding sporophytic tissue representing two consecutive generations seems to be crucial for proper embryo sac assembly and development. Presence of strong mutant alleles of *bel1, sin 1*, and *ant* genes, involved in the formation of the ovule sporophyte (see also Sect. Ovule patterning), results in the early arrest of embryo sac development in *Arabidopsis*. Further, despite the fact that the ovule sporophyte and female gametophyte tissue development have been genetically dissected, the presence of at least inner integument seems to be a prerequisite for successful embryo sac development. This might suggest either functional dependency and/or common check point, respectively, linking female gametophyte and ovule sporophyte development. A role for positional cues transmitted by sporophytic tissue adjacent to the developing embryo sac was suggested in the regulation of female gametophyte development.

Mutations in *CYTOKININ INDEPENDENT1* (*CKI1*) result in the abortion of megagametogenesis at transition from FG5/FG6. Degeneration of chalazal nuclei at FG4 and distortion and later degradation of the central vacuole, coupled with mispositioning of some of the nuclei at FG5 *III* represents one of the first anatomically distinguishable traits in the *cki1* mutant embryo sacs. In terminal *cki1* mutant phenotypes, pleiotropic effects were observed, including different degrees of embryo sac degradation and increased number of nuclei. Thus, the *cki1* mutant phenotypes might imply importance of the central vacuole for the timing and spatial distribution of specific cellular events, for example, mitosis during megagametogenesis. CKI1 is a putative receptor histidine kinase that belongs to the two-component signaling system and it is presumed to act in communication between ovule sporophyte and female gametophyte generations in *Arabidopsis*. However, despite various experimental efforts, the signaling molecule recognized by CKI1 remains unknown.

2.4 Angiosperm Fertilization

Double fertilization in flowering plants involves a complex series of interactions between gametophytes and sporophyte, which results in the fusion of sexual cells and nuclei.

2.4.1 Progamic Phase

2.4.1.1 **Sporophyte–gametophyte interactions.** In angiosperms, the *pistil* (gynoecium) in the central part of the flower represents the pollen-accepting organ, which has multiple functions in plant reproduction: it decides whether to accept or reject the pollen, sustains pollen tube growth, and forms and protects the female gametophyte inside the ovule. The pistil is composed of one or more fused *carpels* that bear the *ovules*. Carpel fusion occurs very early in pistil development. Inner tissues of the developing upper pistil part differentiate to form the specialized secretory zone

of the stigma at the top of the style and the transmitting tissue within the central cylinder of the style. In some species, the style remains hollow, with only one layer of secretory tissue lining the inner surface of the cylinder. At flower maturity, when pollination takes place, the pistil is fully developed and composed of the *stigma*, the *style*, and the *ovary*. The morphology of anthers and pistils coevolved with the mode of pollen dispersal. Different mechanisms of pollen release, pollen transfer, and its deposition to the female sexual organs have developed in plants during evolution.

Numerous *interactions* between the sporophyte and gametophyte include signaling and nutrition. The first interactions take place on the stigma surface after pollen grain landing, where the pollen grain hydrates and germinates forming a *pollen tube*, followed by the penetration of the pollen tube through the specialized tissues of the pistil, – the stigma, and the style. Finally, the pollen tube bearing the two sperm cells reaches the female gametophyte, the *embryo sac*, located inside the *ovule* in the *ovary*. Pollen tube guidance at this stage is influenced by interactions between pollen tube and *micropyle exudates* and the *filiform structures* of the *synergids*.

2.4.1.2 **Stigma types and their role in pollen attachment and hydration.** Appropriate temporal coordination of stigma development is crucial for receptivity, which is defined as the ability to attach pollen by adhesion, to facilitate pollen hydration and subsequent germination, leading to the formation of a pollen tube. Two basic types of the stigma influence the mode and place of the first sporophytic cell–gametophytic cell interaction. *Dry stigmas* are characteristic with surface cells protruding as papillae, which are covered by a primary cell wall, a waxy cuticle and proteinaceous pellicle. During the receptive stage, the cuticle is broken either by visiting insects, due to the activity of some enzymes such as esterases or cutinases or by increasing stigma cell turgidity. The surface of *wet stigmas* is covered by a viscous layer of exudates containing lipids, proteins, polysaccharides, and pigments produced by the surface cells or by cells of the secretory zone of the stigma. A thin layer of water is found beneath this viscous layer.

Pollen grains are shed either in a bicellular or in a tricellular condition and are transferred to stigmas by pollinators either biotic (such as insects, birds, or animals) or abiotic (mainly wind, and less frequently, water). The first interactions of the proteins and lipids of the pollen coat and proteins of the stigma surface follow pollen capture mediated by the structure of exine. A cysteine-rich protein with similarity to lipid transfer proteins called STIGMA/STYLAR *Cys*-RICH ADHESIN (SCA) and pectic polysaccharide are required for pollen tube adhesion in lily. In Brassica with a dry stigma, *S-LOCUS-RELATED PROTEIN* (*SLR1*) is a stigma-expressed gene related to the *S*-glycoprotein, which is involved in self-incompatibility (SI) (see Sect. 2.4.3).

Recently, general similarities were found in the pollination of dry and wet stigmas; the presence of lipids modulates water transfer to the desiccated pollen and tube penetration, while highly diverse proteins and peptides mediate self and foreign pollen recognition as well as pollen tube growth. Hydration of pollen grains, which are dehydrated and metabolically quiescent when they are released from anthers, seems to be actively regulated on stigmas. Hydration induces pollen grain activation, subsequent germination, and pollen tube formation. Ultrastructural

rearrangements during pollen activation are mainly connected with the endomembrane systems, such as expansion of ER, appearance of dictyosomes and secretory vesicles. Polarized migration of organelles to the domain of pollen aperture is observed just before pollen tube emergence. The cytoplasm of the vegetative cell exhibits a rotational or streaming movement and passes into the tip of pollen tube followed by the movement of the MGU (see Sect. 2.3.1.1). Pollen tube penetration of the stigma surface is controlled by means of various enzymes. Activities of numerous hydrolases, such as acid phosphatase, esterases (e.g. cutinase), pectate lyase, ribonuclease, and protease, were localized to the walls of pollen grains and tubes; while cutinase inhibitors that can significantly reduce the ability of the pollen tube to penetrate the stigma were detected in stigma cells. Enzymatic penetration of the stigma is strictly controlled, which demands communication between pollen tube and the stigma. Receptor kinases are supposed to be mediators of such communication (e.g. *Lycopersicum esculentum* PROTEIN RECEPTOR KINASE1 (LePRK1) and (LePRK2)). The pollen specific cysteine-rich protein LATE ANTHER TOMATO52 (LAT52), which is essential for pollen hydration and germination *in vitro* and for normal tube growth, interacts *in vivo* with LePRK2. Other proteins like PISTIL POLLEN ALLERGEN-LIKE (PPAL) protein with similarity to β-expansin, LIPID TRANSFER PROTEIN (LTP) or SCA were detected in stigma exudates. These proteins are members of a signaling system that regulates pollen adhesion, germination, and pollen tube growth.

2.4.1.3 **Pollen tube growth.** Pollen tubes are among the most rapidly extending cells. These growth rates are possible due to a highly polarized apical fusion of vesicles, which transports cell wall components to the growing tip. A tip-focused cytoplasmic Ca^{2+} gradient is known to play a central role in the regulation of pollen tube growth. Modulation of the cytoplasmic Ca^{2+} concentration results in changes in the rate and direction of pollen tube growth. It was found that increases in GTP levels could promote pollen tube growth. Rop GTPases have been implicated in the regulation of pollen tube growth probably by interaction with the tip-focused Ca^{2+} gradient.

The growth of pollen tubes through the style is promoted by products of the secretory cells in the transmitting tissue of the central core of solid (closed) styles or those of the canal lining in hollow (open) styles. The cells of the transmitting tissue form a continuous connection with the cells of the stigma secretory zone. Longitudinal walls of the transmitting tissue cells are separated by extracellular matrix that contains a mixture of sugars, proteins, and phenolic compounds. Hyp-rich glycoproteins (HRGPs) like arabinogalactan proteins, PISTIL-SPECIFIC EXTENSION-LIKE PROTEIN (PELP), TRANSMITTING TISSUE-SPECIFIC (TTS) and GALACTOSE-RICH STYLE GLYCOPROTEIN (GaRSGP) form an intercellular matrix that serve as a source of nutritive and regulative compounds for tobacco pollen tube growth. A gradient of TTS protein glycosylation in the styles of tobacco or a gradient of γ-aminobutyric acid (GABA) in *Arabidopsis* pistil could function as possible mechanisms for the control of pollen tube growth. Pollination of the stigma also triggers a cascade of responses in distal flower organs. It was found that ethylene and its biosynthetic precursor 1-aminocyclopropane-1-carboxylate (ACC)

play an important role in the regulation of these responses. Analysis of ACC metabolism showed that pollination induces high ACC–oxidase transcript levels in all cells of the transmitting tissue, while the ACC-synthase transcripts accumulate in a wavelike pattern in which the peak coincides with the front of the ingrowing pollen tube tips. This wave of ACC-synthase expression can also be induced by incongruous pollination and partially by wounding indicating that pollen tube invasion displays similar features to wounding and such effects might be recruited for evoking the postpollination response. Barriers preventing incorrect pollination negatively affect pollen tube growth in the pistil (see Sect. 2.4.3.).

Within the ovary, the pollen tube is guided on the surface of the placenta and the funiculus toward the micropyle of the *receptive ovule*. The most common way of pollen tube entry to the embryo sac located inside the ovule is through the opening in the ovule integuments, the micropyle. Attraction signals for pollen tube growth could include calcium accumulation in the synergid during embryo sac maturation, as well as other chemotropic signals created near micropylar region by various ovule structures (e.g. micropylar cells of the nucellus, the inner integument or the filiform apparatus of synergids). Presence of proteins, RNA, and carbohydrates has been demonstrated in the micropylar region. Thus, it can be concluded that numerous checkpoints exist along the whole pollen tube pathway, which provide active signaling that guides the pollen tube to its destination, the embryo sac.

2.4.2 Syngamic Phase: Karyogamy and Plasmagamy, Double Fertilization

The pollen tube enters the embryo sac through the *receptive synergid*. The signs of the synergid receptivity are: (1) cytoskeleton reorganization, (2) Ca^{2+} accumulation, and (3) degeneration of organelles and plasma membrane. The pollen tube develops an aperture on the tip and discharges its contents into the receptive synergid or in the space previously occupied by this synergid (depending on the time of the synergid degeneration). Pollination initiates disintegration of one of two synergids, which were identical before pollination. The mechanism of this process initiation at the molecular level is still unclear. In some species, synergid cell death occurs even in the absence of the pollination, suggesting that the process might be an intrinsic feature of megagametogenesis.

DNase degrades both the vegetative cell nucleus and the nucleus of the synergid and the sperm cells are transferred to the egg and central cells. Bundles of actin filaments form two corona-shaped structures; one is located near the egg nucleus and the other near the central cell. Coordinated actions of actin and myosin on the sperm cell surface mediate sperm cell transport to the position near the egg cell and the central cell nuclei. Regulation of the actin-corona formation is not yet fully known. The cell walls of the egg cell and the central cell are modified and surrounded only by the cell membrane that facilitate delivery of the sperm nuclei with the male cytoplasm. This delivery is initiated by the apposition of plasma membranes of the sperm and the egg and those of the second sperm cell and the central cell. *Double fertilization* includes both: two nuclear fusions (*karyogamy*) and the coalescence of male and female cytoplasm (*plasmagamy*). The first karyogamy (syngamy) is the fusion of the egg and sperm cell nuclei,

which produces the diploid nucleus of the *zygote*. Measurements of calcium concentration during *in vitro* fertilization of maize egg demonstrated that fusion of gametes triggers a calcium influx, which is then followed by an increase of cytoplasmic calcium (similarly as in animal systems). The second karyogamy, occurring only in angiosperms, is the fusion of the (usually) diploid nucleus of the central cell with the second sperm cell nucleus forming the triploid *primary endosperm nucleus*. Plasmagamy can be conditioned by the female gametophyte as, in some cases, male cytoplasm is excluded from the egg cell. It seems that despite morphologic uniformity of the two sperm cells, some differences on their surface may allow specific trafficking either to the egg or to the central cell. Formation of MGU could significantly influence sperm cell traffic.

2.4.3 Prezygotic and Postzygotic Barriers to Fertilization

Various physiological or genetic barriers can preclude successful fertilization. Male or female sterility, which arrest normal development of gametes, is one example of such a genetic barrier. Furthermore, other mechanisms called *incompatibility* exist, which are defined as the inability of functional gamete fusion. Many types of intraspecific and interspecific incompatibility have developed in plants to prevent inbreeding on the one hand (self-incompatibility) and on the other to maintain stability of species preventing gene flow between unrelated species (cross-incompatibility, incongruity). These reproductive strategies can function either as prezygotic barriers that inhibit pollen tube growth or as postzygotic barriers causing embryo abortion.

2.4.3.1 **Self-incompatibility.** The widely distributed mechanism of the prezygotic barrier named *self-incompatibility* (SI) is an inherited phenomenon, by which plants can recognize and reject their own pollen by inhibiting pollen hydration, germination, and/or tube invasion and thus block the delivery of sperm cells to the ovary and prevent inbreeding. Variants of self-incompatible systems were classified according to different aspects. On the basis of flower morphological differences, heteromorphic, and homomorphic types of SI, are distinguished. *Heteromorphic SI* is characterized by floral polymorphism. Flowers of species with heteromorphic SI vary in position, size, or shape of the reproductive floral organs. Nevertheless, most of the self-incompatible plants are of a *homomorphic type*. Plants with this type of SI produce morphologically identical flowers. Homomorphic SI is further divided into sporophytic and gametophytic types according to determination of successful pollen phenotype. In *sporophytic SI*, the pollen phenotype is determined by the genotype of parent; while in *gametophytic SI*, the pollen phenotype is determined by its own genotype. A single locus with multiple alleles named the *S-gene* (*STERILITY*) represents the genetic basis of self-incompatibility. Later, two separate genes at the S-locus were distinguished that control male and female specificities. In sporophytic SI, pollen rejection occurs on the stigma surface when the pollen parent has one or both S-haplotypes in common with the pistil, independently of the S-haplotype of the pollen grain. Inhibition of pollen tube growth in case of more frequent gametophytic self-incompatibility occurs mainly in the upper part of the style in the case in which S-haplotype of the pollen matches either one of the two S-haplotypes in the

pistil. Gametophytic SI in the *Poaceae* family is genetically controlled by two unlinked multiallelic loci S and Z. Alleles of both S and Z need to match in the pollen and pistil to reject pollen.

2.4.3.2 **Mechanisms of self-incompatibility.** It is now known that the common phenomenon of self-incompatibility is not realized uniformly by all flowering plants, since several biochemically distinct mechanisms of S-locus expression were revealed among different families. (1) Gametophytic SI mechanism in Solanaceae, *Rosaceae*, and *Scrophulariaceae* involves *S-RNase*-mediated degradation of RNA in own pollen tubes. Several variants of a receptor or inhibitor models explaining this mechanism were proposed. (2) The *Papaveraceae* gametophytic SI mechanism represents *S-protein*-mediated signal transduction cascade via Ca^{2+}, phosphoinositides, protein kinase, and phosphatases in pollen tubes. Actin cytoskeleton was found to be a target for signaling pathways involved in this SI response. (3) Mechanism of the sporophytic SI in *Brassicaceae* is based on the different genes of S-locus that control male and female specificity. Stigma specific proteins S-LOCUS-SPECIFIC GLYCOPROTEIN (SLG) and S-LOCUS RECEPTOR KINASE (SRK) are associated with self-incompatibility in the pistil, while S-LOCUS CYSTEIN-RICH/S-LOCUS PROTEIN-11 (SCR/SP11) represents the pollen specific S protein.

2.4.3.3 **Incongruity.** Interspecific incompatibility caused by evolutionary divergence is based on the absence of information that leads to an incomplete spatial and temporal coordination in the plant reproduction process.

2.4.4 *In Vitro* Fertilization

Experimental approaches for the study of fertilization or overcoming incompatibility involve methods of *in vitro* pollination/fertilization such as intraovarian pollination, test tube pollination of segments of the placenta with ovules, and *in vitro* fusion of isolated, single gametes.

2.5 Embryo and Endosperm Development

Fertilization of the female gametophyte in angiosperm plants initiates a process of coordinated development of embryo, endosperm, and seed coat, which ensures the production of a viable seed. It is more than a century since *double fertilization* was discovered. After pollination, two sperm cells, delivered by a pollen tube, fuse with the *egg* and *central cell*, which results in the formation of the *embryo* and endosperm, respectively. The embryo and endosperm both develop within the confines of the maternal tissue (seed coat), but each follows a different developmental program. Within the embryo, the basic body plan of the mature plant is laid down, whereas the "life-history" of the endosperm is far shorter and limited to the seed development stage. Nevertheless, the endosperm fulfills essential functions in nourishing the developing embryo and controlling the whole seed development process. It is also important to note that the endosperms of plants, including cereal species such as rice, maize, wheat, and barley represent one of the most important renewable sources of food and raw materials.

2.5.1 Endosperm Development

Endosperm development is initiated by the union of the haploid sperm cell and the diploid central cell. Three main types of

triploid endosperm can be distinguished depending on coupling between nuclear and cell divisions. The most common form of endosperm is the so-called *nuclear type*, which involves a phase of free-nuclear (syncytial) division that is not accompanied by cytokinesis. On the other hand, the *cellular type* of endosperm develops via a series of mitotic divisions always coupled with cytokinesis. A third possibility is provided by *helobial type*, in which the proendosperm divides to give rise to two daughter cells, one of which undergoes cellular type of development, whereas the other undergoes nuclear type of development.

2.5.1.1 **Evolutionary aspects of endosperm development.** Interestingly, gymnosperms instead of developing true endosperms possess characteristically enlarged female gametophytes that accumulate nutrients before fertilization. Thus, from an evolutionary point of view, the endosperm can be viewed as a functional homolog of the gymnosperm female gametophyte. Many aspects of nuclear-type endosperm development appear to have been repeatedly sequestered from preexisting mechanisms found in female gametophytes of land plants. Both female gametophytes and nuclear endosperms undergo multiple rounds of free-nuclear divisions forming a multinucleate *syncytium* and later arrangement of nuclei into nucleoplasmic domains via the establishment of radial microtubule arrays. Finally, *cellularization* (formation of cell walls around the nuclei and surrounding regions) is achieved through the formation of tubelike structures termed *alveolus* (ALV). Genetic evidence also supports a common mechanism in the regulation of endosperm and female gametophyte since a number of mutants have been isolated, which are affected in both these processes.

2.5.1.2 **Development of nuclear endosperm.** The nuclear-type endosperm is the most common type in angiosperms including dicot *Arabidopsis* and monocot cereals. Owing to availability of molecular markers for different endosperm compartments and developmental stages, most of the information on nuclear endosperm development has been obtained from studies in *Arabidopsis*. The young *Arabidopsis* endosperm can be divided into three *domains*: the *micropylar endosperm* (MCE), which surrounds the embryo, the *peripheral endosperm* (PEN), which fills most of the seed volume, and the *chalazal endosperm* (CZE) in the nucellar region.

Nuclear endosperm development is characterized by stages of *syncytial nuclear divisions* (Fig. 7a) followed by *cellularization* (Fig. 7b). The whole syncytial endosperm stage can be divided into nine substages, each representing one of eight rounds of mitotic divisions. After the initial three synchronous division cycles, the mitotic activity of MCE, PEN, and CZE occurs independently, with nuclei dividing synchronously within domains. Nuclear divisions have never been observed directly in the CZE after the three synchronous rounds of division. Increased signal intensity of the labeled histone in the CZE nuclei together with the absence of nuclear divisions in CZE suggests that these nuclei undergo endoreduplication. At the final stage, the syncytial endosperm contains 200 nuclei.

The cellularization process starts as a wave in MCE, progressing through PEN and CZE at different rates and with significant variations between the different parts. When the cellularization around the embryo in MCE is completed, the PEN is

still syncytial, whereas CZE remains syncytial until late stages of seed maturation. Cellularization in *Arabidopsis* as well as cereals involves repeated rounds of formation of a *radial microtubule system* (RMS) emanating from nuclei and enclosure of nuclei with RMS in a tubular cell wall structure (*alveolus*) (Fig. 7b). In PEN, the process of cellularization starts with the occurrence of RMS at the nuclear surface and subsequent formation of cytoplasmic *phragmoplasts* in interzones between RMS. The phragmoplasts mediate formation of tubular alveolus, which encloses the nucleus with its RMS. Nuclei within the alveoli undergo synchronous periclinal division leading to formation of peripheral cells and internal alveoli. The cellularization process in MCE also occurs via RMS and cytoplasmic phragmoplasts, but because of spatial constraints, alveoli typically do not form. The cellularization process for *Arabidopsis* results in a completely cellular endosperm except for a small area in the CZE adjacent to the chalazal cyst (ChC).

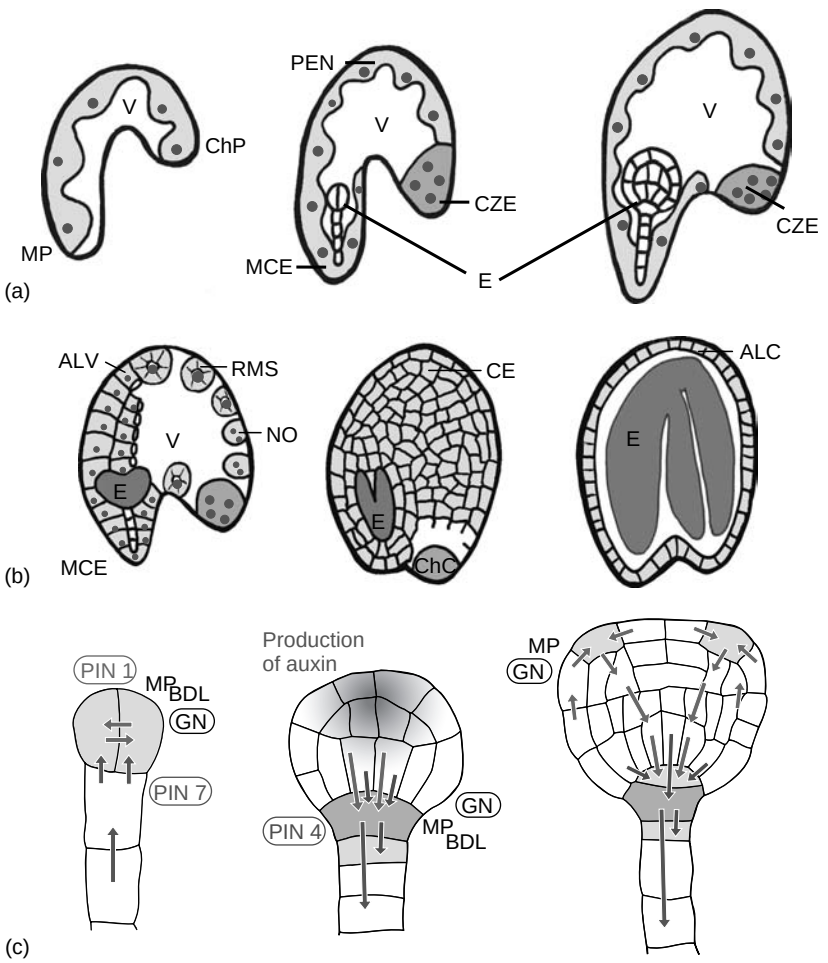

As the *Arabidopsis* embryo grows, the cellular endosperm is gradually depleted, which contrasts with the persistent endosperm of the cereals. In the mature *Arabidopsis* seed, a massive embryo fills the ovule and only a single peripheral layer (the *aleurone layer*) with unknown function persists. It is assumed that the nonpersistent endosperm supports the developing embryo and this support function is later taken over by the cotyledons during germination.

Despite obvious differences in some aspects of endosperm development between *Arabidopsis* and many monocots including cereals, the cellularization processes reveal striking similarities. Importantly, the process of cellularization by RMS formation also occurs during the cellularization process in a number of other systems, including sporogenesis in lower plants, male and female gametophyte development in gymnosperms and angiosperms, and embryogenesis in gymnosperms. Thus, the findings obtained from studies on *Arabidopsis* and the endosperm of cereals have more general validity in our understanding of plant development.

2.5.2 Embryo Development

Embryogenesis is a process that transforms a single-celled zygote into the embryo containing all the basic pattern elements of the future plant. Unlike in higher animals, where the embryo is a miniature variant of the adult, the juvenile form of the

Fig. 7 Development of the endosperm coenocyte in the *Arabidopsis* embryo sac. (a) Endosperm nuclei migrate from the micropylar pole (MP) toward the chalazal pole (ChP) of the embryo sac. A large central vacuole (V) fills up most of the volume and is surrounded by cytoplasm. Cytoplasm is in yellow, nuclei are in red and vacuole is in white. As development progresses, the coenocyte forms three distinct domains: (1) the region surrounding the embryo (MCE); (2) the peripheral endosperm (PEN); and (3) the region of the chalazal endosperm (CZE), which contains the chalazal cyst (ChC). At the end of the globular embryo stage, the embryo (E) becomes completely surrounded by cytoplasm (Modified according to Olsen, O.A. (2004) Nuclear endosperm development in cereals and *Arabidopsis thaliana, Plant Cell* **16**(Suppl. 1), S214–S227). (b) Endosperm cellularization in Arabidopsis: the endosperm at the late globular embryo stage shows a cellular endosperm in MP, a gradient of stages in the alveolation process in PEN, and endosperm nodules (NO) as well as chalazal cyst (ChC) formation in ChP. During alveolation, radial microtubule system (RMS) emanates from the nuclear membranes, cell walls form tubes or alveoli (ALV) around each nucleus with open ends toward the central vacuole (V). Later, endosperm becomes completely cellular (CE). The endosperm is consumed during seed maturation, leaving only the peripheral aleurone-like cell (ALC) layer in a mature seed (Modified according to Olsen, O.A. (2004) Nuclear endosperm development in cereals and *Arabidopsis thaliana, Plant Cell* **16**(Suppl. 1), S214–S227). (c) Model for a role of auxin distribution in embryo patterning: two-cell stage embryo – specification of the apical cell. Auxin accumulates (depicted in green) in the proembryo by PIN7-dependent transport (red arrows) through the suspensor. Auxin response (*mp, bdl*) and transport (*gn*) mutants show defects in the establishment of the apical cell. Young globular embryo – specification of the root pole. Auxin production is established in the apical part (purple) and auxin transport routes reverse. Auxin accumulates, in a PIN1- and PIN4-dependent manner (blue and purple arrows) in the hypophysis and is further transported through the suspensor. *mp*, *bdl*, and *gn* show defects in root pole establishment. Early heart (triangular) stage – establishment of cotyledons. Most auxin is still transported to the basal part of the embryo and then further through the suspensor. New sites of auxin accumulation emerge at the tips of forming cotyledons. MP and GN are also involved in cotyledon establishment. (See color plate p. xxx).

plant (*seedling*) shows only basic pattern elements, which can be strongly modified by elaborated postembryonic development. The seedlings of different higher plant species are remarkably uniform. They display a main *apical–basal axis* of polarity, with the shoot meristem at the apical end and root meristem at the basal end and the *radial pattern* of tissue layers, which are arranged concentrically. This basic embryonic pattern is then postembryonically elaborated mainly by the activity of the *meristems* (permanent stem cell populations) and *de novo* organogenesis, which can lead to a great increase in architectural complexity and which generates most species-specific features of the adult plant.

2.5.2.1 **Pattern formation in *Arabidopsis* embryogenesis.** In the progression of plant embryogenesis, three main patterning events can be distinguished: the generation of the apical–basal axis, formation of radial pattern and establishment of bilateral symmetry. In different higher plant species, the mature embryos show the same basic patterning elements; however, the early pattern of cell division during embryogenesis is not uniform between species. Therefore, it is unclear how important the early patterning events are. By far, most of the information on mechanisms of embryonic patterning comes from the studies in *Arabidopsis* and other *Brassicaceae*. These species display an exceptionally regular and reproducible pattern of cell divisions during embryo development.

The apical–basal pattern in *Arabidopsis* has been traced back to the earliest stages of embryogenesis. The zygote has an asymmetrically positioned nucleus with a vacuole, which elongates in a polar manner at the apical end. The zygote divides asymmetrically and produces a small *apical cell* and a larger *basal cell*. The apical cell divides vertically and generates the "*proembryo*," which later gives rise to most regions of the seedling including the shoot meristem. The basal cell continues to divide horizontally and produces the *suspensor* – a single file of cells that attaches the proembryo to maternal tissue. At the early globular stage, the uppermost cell is specified to become the *hypophysis* – the founder of the root meristem. Thus, both apical and basal ends of the future seedling are established. This initial apical–basal subdivision is also reflected molecularly: for example, genes encoding the transcription factors AtML1 or WUS are expressed exclusively in the apical cell lineage, whereas specific arabinogalactan epitopes or a regulator of auxin transport PIN-FORMED7 (PIN7) were detected in the basal cell lineage of early embryos.

The formation of the radial pattern initiates at the eight-cell stage of the proembryo. These cells divide periclinally and set apart the primordial epidermis (*protoderm*). At the next stage, the inner cells divide vertically producing outer ground tissue and inner provascular cells. Thus, the globular embryo is composed of a radial pattern consisting of three concentric cell layers that later develop into epidermis, ground tissue (endodermis and cortex), and vascular tissue of the seedling. The early marker of embryonic ground tissue is the transcription factor SCARECROW (SCR), which has been shown in roots to be required for endodermis specification. Interestingly, SCR transcription is regulated by a homologous factor SHORTROOT (SHR), which moves into the endodermis from the cells of vascular bundle, where it is expressed.

At the early heart (triangular) stage, early patterning is completed with the

establishment of bilateral symmetry, which is manifested by the initiation of two symmetrically positioned embryonic leaves – *cotyledons*. The initiation of cotyledon primordia is preceded by production of transcription factor ANT, which is required for maintaining the proliferative cell state during various organogenesis processes.

The early patterning is followed by a phase of growth and elongation of organs that merely elaborate already prelaid architecture. The next and last phase of embryo development (embryo maturation) is characterized by accumulation of storage reserves that include lipid, starch, and protein bodies within embryo cells. Hereafter, the seed dries and becomes separated from the mother plant. After seed dormancy is overcome, germination occurs and the embryo gives rise to a seedling.

2.5.2.2 **Lessons from embryo mutant screens.** The mechanisms that govern patterning processes in *Arabidopsis* embryogenesis have been addressed mainly using genetic approaches. In mutant screens for genes encoding the regulators of embryo development, mutants were categorized by seedling phenotypes and by early embryonic division patterns.

Morphogenetic embryo mutants Many of the mutants recovered from these screens were found to be involved in basic cellular processes that only subsequently affect pattern formation. Cytokinesis mutants such as those in the *KNOLLE* (*KN*) gene (encodes a cytokinesis-specific syntaxin) or its interactor *KEULE* (*KEU*) (the yeast Sect. 1 homolog) fall into this category. Both these mutants are affected in vesicle fusion in the plane of cell division, which results in incomplete cell walls and multinucleate cells.

Other mutants are defective in the proper orientation of the plane of division rather than in cell division *per se*. These are, for example, *fass* mutant embryos, which are made up of irregularly shaped and enlarged cells. *FASS/TONNEAU2* (*FASS/TON2*) encodes a type 2a protein phosphatase and seems to be required for the formation of ordered microtubular arrays that make up the plant cytoskeleton. Also members of the *pilz* and *titan* (*ttn*) groups of mutants have abnormally formed embryos as a result of cytoskeletal defects that interfere with mitosis and cytokinesis. The *pilz* mutant embryos contain actin but lack microtubules, which results in a mushroom-shaped embryo with one or a few large cells. *TTN5* encodes a small G-protein Arl2 with a possible role in regulation of vesicle trafficking, whereas other genes of the *pilz* group encode orthologs of mammalian tubulin-folding cofactors.

Another class of mutants that shows more pleiotropic defects during embryogenesis is defective in genes involved in sterol biosynthesis such as *STEROL METHYLTRANSFERASE1/CEPHALOPOD* (*SMT1/CPH*), *FACKEL/HYDRA2* (*FK/HYD2*), and *HYDRA1* (*HYD1*). Early embryo defects in cell division and expansion in these mutants result in embryos with abnormally shaped cotyledons and reduced central and basal regions; however, the exact role of sterols in embryogenesis is unclear.

In summary, molecular analysis of morphogenetic mutants did not provide much insight into plant patterning mechanisms, but provided important entry points into molecular studies of basic cellular functions such as vesicle trafficking and cytoskeleton arrangement.

Pattern formation embryo mutants Most of the mutants specifically affected in embryo patterning processes were identified by loss or alterations in apical–basal or radial pattern elements at the seedling stage. In contrast to the many transcription factors revealed by similar screens in animal systems such as *Drosophila*, only a few transcriptional regulators were found in *Arabidopsis*.

Several mutants with apical-basal patterning defects, such as *monopteros* (*mp*), *bodenlos* (*bdl*), *auxin resistant6* (*axr6*), or *gnom* (*gn*), have been isolated. All these mutants show deletion of basal structures (root and occasionally hypocotyl) and variable defects in formation of apically positioned cotyledons. *MP* encodes the transcriptional activator AUXIN RESPONSE FACTOR 5 (ARF5) and *BDL* the transcriptional repressor from Aux/IAA family (IAA12), both of which are involved in auxin response. Consistent with the similar early embryo phenotypes, *mp* and *bdl* interact genetically, and yeast two-hybrid studies also indicate that a physical protein interaction can occur. *AXR6* encodes the *Arabidopsis CULLIN1* (AtCUL1), a component of the SCF (*SKIP*, *CDC53/CULLIN*, and *F-box protein*) TIR1 complex. The $SCF^{\mathbf{TIR1}}$ is a complex with ubiquitin ligase function, which by binding and ubiquitination of Aux/IAA proteins (e.g. BDL) mediates their degradation. The involvement of BDL, MP, and AXR6 in the same patterning processes suggests that BDL and MP form a complex *in vivo* that prevents MP from activating target genes. In this scenario, auxin promotes interaction between BDL and SCF^{TIR1} complex, which ultimately leads to BDL degradation. Consequently, derepressed MP is able to mediate auxin-dependent gene expression and thereby initiate root formation.

On the other hand, the *gn* mutant, which also displays severe aberrations in apical-basal pattern, is defective in the guanine-nucleotide exchange factor on ADP-ribosylation factor G proteins (ARF-GEF). Endosomal membrane-associated GNOM ARF-GEF acts as an activator of ARF G proteins, which play a role in the recruitment of vesicle coats required for their budding and cargo selection. Here, the mechanistic connection to auxin signaling and transport was established through a role for GN in vesicle trafficking that mediates proper subcellular targeting of PIN auxin transport components.

Thus, the molecular analysis of embryo patterning mutants suggested a prominent role for the phytohormone *auxin* (indole-3-acetic acid, IAA) in the regulation of this process.

2.5.2.3 **Differential auxin distribution in embryonic axis formation.** Complementary evidence for the involvement of auxin in embryo development has been obtained through chemical manipulation of auxin homeostasis in excised *Brassica* embryos. The main breakthrough in understanding the role of auxin in embryo patterning was the discovery of dynamic *auxin gradients* during embryogenesis. Cellular distribution of auxin and its response was indirectly visualized using anti-IAA antibodies and an auxin-responsive reporter *DR5::GFP*. Immediately after the division of the zygote, auxin accumulates in the apical cell, which is being specified. During subsequent development, the maximum auxin activity persists in the proembryo. Approximately at the 32-cell stage, when the basal embryo pole is being specified, the gradient of auxin accumulation suddenly reverses and forms a new maximum in the uppermost suspensor cells, including the hypophysis. At later stages

of embryogenesis, additional auxin accumulation foci appear in the tips of the developing cotyledons. The next important question was, by what mechanism are these auxin gradients established and maintained. Conceptually, cellular auxin levels could be regulated by different levels of synthesis/degradation and conjugation/deconjugation or by intercellular transport.

Auxin is unique among plant hormones, being actively distributed within the plant by cell-to-cell movement. This movement is strictly directional (polar) and involves auxin influx and efflux carrier proteins. The direction of auxin flow is supposed to be determined by the asymmetric cellular localization of efflux carriers. The plant-specific PIN proteins have so far been the best-characterized candidates for facilitators of auxin efflux. Since both chemical and genetic (*gn* and *pin* mutants) inhibition of auxin transport interferes with the embryonic auxin gradients, it seems that polar auxin transport is a major determinant for establishing and maintaining auxin gradients. Furthermore, inhibition of auxin transport leads to severe embryo developmental defects, ranging from cup-shaped embryos with misspecified apical structures and a nonfunctional root pole, to ball-shaped embryos without any discernible apical–basal axis. These findings indicate that auxin transport–dependent auxin gradients are required for proper embryo development. Analysis of expression and localization of PIN auxin transport components finally completed the model, describing the role of PIN-dependent auxin distribution in the establishment of the embryonic apical–basal axis (Fig. 7c).

In the early stages, auxin is actively provided to the apical cell from the adjacent basal cell by the action of basal cell-specific, apically-localized PIN7. In subsequent stages, the cells of the suspensor continue to localize PIN7 on their apical side, while in the proembryo, another protein, PIN1, is produced without apparent polarity. But after the 32-cell stage, PIN1 becomes localized to the basal membranes of the provascular cells and PIN7 polarity in suspensor cells reverses, suggesting downward transport toward the region of the future root pole and out of the embryo. *PIN4* expression then starts at the basal pole of the embryo, supporting the action of both PIN1 and PIN7. As a result of these changes in auxin flow, the auxin gradient reverses, displaying its new maximum in the uppermost suspensor cell, which in response to auxin is specified to become the *hypophysis* – the founder of the future root meristem.

Thus, developmentally regulated changes in polarity of PIN proteins result in the redirection of auxin fluxes for local auxin accumulation, which is required first for specification of the apical and later the basal pole of the embryonic apical–basal axis. Interestingly, it seems that a similar mechanism utilizing PIN-dependent auxin distribution also operates in postembryonic development such as organogenesis. Regardless of the type of organ primordia, the new direction of PIN-mediated auxin transport determines the growth axis of the developing organ.

2.5.3 Interregional Communication During Seed Development

It seems obvious that seed development requires the coordination of embryo and endosperm development and therefore requires communication between these two regions. The available evidence for this communication relies

mostly on the analysis of mutant phenotypes in *Arabidopsis*. For example, mutants in two genes, *FERTILIZATION-INDEPENDENT ENDOSPERM* (*FIE*) and *FERTILIZATION INDEPENDENT SEED2* (*FIS2*) have similar phenotypes to the *Arabidopsis medea* (*mea* mutant). In the absence of double fertilization, continual divisions of polar nuclei are followed by cellularization, and overproliferation of the endosperm is accompanied by aberrant embryo development. Specifically, analysis of the *haiku* mutant confirmed that endosperm development plays a role in late embryo development. Further elegant studies using conditional expression of diphtheria toxin in different regions of developing seed allowed selective ablation of the endosperm. This leads to clear embryo and seed development defects demonstrating an important role for the endosperm in embryo development and confirming the communication between these seed regions. However, the molecular nature and mechanism of this communication remains unknown.

2.6 Seed and Fruit Formation

Double fertilization activates, in parallel with embryo and endosperm development inside the embryo sac, transformation of an ovule into a *seed* and transformation of the ovary into a *fruit*. In cases in which pollination and fertilization do not occur, flower abscission follows.

2.6.1 Seed and Fruit Maturation

Both seed and fruit formation include processes of growth and maturation (ripening). General changes during fruit ripening include (1) changes of color through chlorophyll degradation and carotenoid or flavonoid accumulation, (2) textural modifications caused by changes of cell wall structure and cell turgor, (3) qualitative as well as quantitative changes in sugars, acids, and other substances that affect flavor, aroma, and nutritional quality, and (4) increase of respiration and ethylene synthesis during ripening. Such an increase is typical only for the so-called climacteric fruits such as the fleshy fruits of tomato, apple, and banana. Ethylene production is not involved in ripening of nonclimacteric fruits in which exogenous ethylene can cause increased respiration while natural ripening is not promoted. Multigene families of ACC synthase and ACC oxidase control regulation of ethylene synthesis during climacteric ripening. Genes for both synthesis and degradation of proteins are activated during the ripening process.

2.6.2 Dormancy and Seed Germination

The period of seed maturation is rounded off by transition to a stage of reduced growth, which can be either *quiescence* or *dormancy*. Quiescent seeds can germinate in appropriate physiological conditions, while dormant seeds require additional hormonal and environmental regulation mechanisms for initiation of germination. Seed dormancy can represent a benefit either during embryo formation, when it prevents precocious germination, or as a regulator of seed germination in unfavorable conditions; otherwise, seed dormancy can be an inconvenient factor in agricultural crops, where rapid germination is required. Processes leading to seed dormancy are a part of the normal developmental pathway of seed formation. They include (1) gradual desiccation, (2) differentiation of the seed coat, (3) changes in gene transcription, and

(4) reduction of cellular metabolism. Regulation of these processes involves activation of the transcription of new genes. Some of them can shut off transcription of genes regulating cellular metabolism, while others may be involved in regulation of seed formation. The phytohormone abscisic acid (ABA) is involved in the regulation

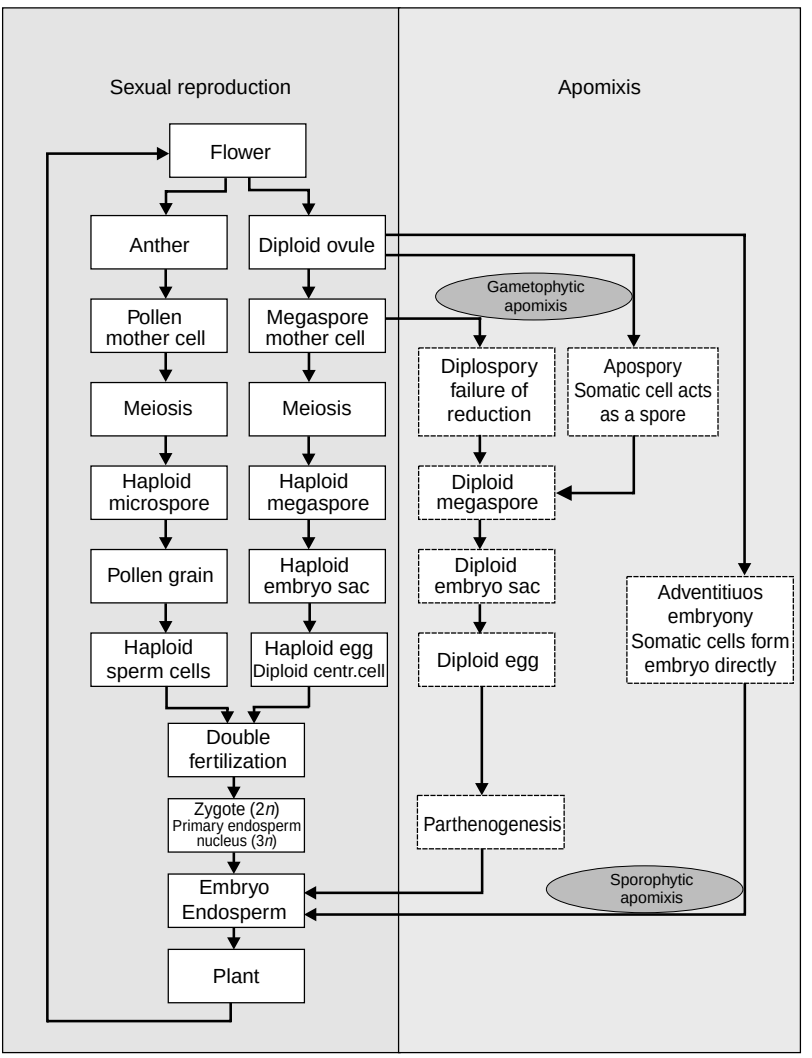

Fig. 8 Comparison of apomixis and sexual life cycle of flowering plants. Apomictic reproduction by seed occurs when the sexual life cycle is shortened. Sporophytic apomixis (adventitious embryony) occurs when the unreduced cells of the ovule directly give rise to an embryo. Gametophytic apomixis occurs by the formation of an unreduced megaspore either due to failure of megaspore mother cell reduction (diplospory) or by the formation of an unreduced megaspore from the somatic cells of a diploid ovule (apospory). (See color plate p. xxix).

of the onset and in maintenance of dormancy. ABA accumulation in seeds is low during early embryogenesis and then increases during storage-reserve biosynthesis and declines when the seed undergoes maturation. Lowering of water potential is the signal for ABA accumulation. The course of dormancy is regulated by differential gene activation. *LEA* genes encode proteins more abundant in mature embryos than in young or germinated embryos. Expression of *LEA* genes can be induced by ABA.

Overcoming seed dormancy can be achieved by various pretreatments like scarification, prolonged washing, and storage at low or high temperature enabling additional ripening of seeds, or exposure to light or dark conditions. Seed *germination* begins with events such as water uptake, resumption of metabolic activity, degradation of LEA proteins, production of new proteins, and mobilization of stored food reserves. Termination of seed germination is manifested by the elongation of the embryonic axis, which is completed by penetration of a root pole through the seed envelope. Information for synthesis of the first proteins that are essential for germination and root cell elongation comes from mRNA stored in the embryo during embryo development.

3 Apomixis

Apomixis is a naturally occurring process of asexual reproduction through seeds. It involves elimination of key developmental processes including meiosis and double fertilization. In nature, apomixis is widespread but infrequent: it occurs in 40 families of flowering plants, but only in about 400 species, which represents 1% of the species that make up those families. Apomixis is most frequent in *Poaceae*, *Asteraceae*, and *Rosaceae* families. Occurrence of various forms of apomixis in different angiosperm taxa can suggest that it has evolved independently many times. Classification of numerous mechanisms of apomixis into two main subgroups is based on the origin and further development of the unreduced cells (Fig. 8.). *Sporophytic apomixis*, also called *adventitious embryony*, is a process in which the embryo arises spontaneously directly from the cells of the nucellus or the integuments of the ovule. More common *gametophytic apomixis* occurs when the apomictic embryo arises from the cells of the unreduced megagametophyte. Gametophytic mechanisms are further divided according to the origin of the cells that give rise to the apomictic embryo. In *diplospory*, the MMC that failed in meiosis is the progenitor of unreduced embryo sac. *Apospory* in which the unreduced embryo sac arises from the somatic cells of the ovule is the most common mechanism of apomixis. Pollen of apomictic plants is often viable. Sometimes, development of both sexual and asexual seeds occurs in the same flower, although precocious apomictic embryo formation before flower opening or synthesis of complete cell wall around the unreduced egg cell can prevent fertilization in apomictic species. Apomixis as a mechanism for rapid generation and multiplication of desired genotypes through seeds is a very attractive idea for plant breeders. Artificial introduction of apomixis into crops as a reproductive alternative is the aim of many attempts as it could represent a unique opportunity to simplify breeding schemes and genetically perpetuate any desired heterozygous genotype.

Acknowledgment

The authors would like to thank Boris Vyskot for critical comments on the manuscript and Kiran Nagavalli Subbana for language editing. This work was supported by grant number 0021622415 (Ministry of Education of the Czech Republic).

See also Molecular Motors in Plant Cells; Plant Gene Expression, Regulation of; Plant Pathology, Molecular; Transgenic Plants for Food Use.

Bibliography

Books and Reviews

Cresti, M., Blackmore, S., Van Went, J.L. (1992) *Atlas of Sexual Reproduction in Flowering Plants*, Springer-Verlag, Berlin, Germany.

Davies, P.J. (Ed.) (2005) *Plant Hormones: Biosynthesis, Signal Transduction, Action!* Kluwer Academic Publishers, Dordrecht, Netherlands.

De Nettancourt, D. (2001) *Incompatibility and Incongruity in Wild and Cultivated Plants*, Springer-Verlag, Berlin, Germany.

Fosket, D.E. (1994) *Plant Growth and Development: A Molecular Approach*, Academic Press, San Diego, CA.

Howell, S. (1998) *Molecular Genetics of Plant Development*, Cambridge University Press, Cambridge, UK.

Johri, B.M., Ambegaokar, K.B., Srivastava, P.S. (1992) *Comparative Embryology of Angiosperms*, Springer-Verlag, Berlin, Germany.

Raghavan, V. (1997) *Molecular Embryology of Flowering Plants*, Cambridge University Press, Cambridge, UK.

Raghavan, V. (2000) *Developmental Biology of Flowering Plants*, Springer, New York.

Thorpe, T.A. (Ed.) (1995) *In Vitro Embryogenesis in Plants*, Kluwer Academic Publishers, Dordrecht, Netherlands.

Primary Literature

Antoine, A.F., Faure, E.J., Cordeiro, S., Dumas, C., Rouigier, M., Feijó, J.A. (2000) A calcium influx is triggered and propagates in the zygote as a wavefront during *in vitro* fertilization of flowering plants, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10643–10648.

Assaad, F.F., Huet, Y., Mayer, U., Jürgens, G. (2001) The cytokinesis gene *KEULE* encodes a Sec1 protein that binds the syntaxin KNOLLE, *J. Cell Biol.* **152**, 531–543.

Baker, S.C., Robinson-Beers, K., Villanueva, J.M., Gaiser, J.C., Gasser, C.S. (1997) Interaction among genes regulating ovule development in *Arabidopsis thaliana*, *Genetics* **145**, 1109–1124.

Baroux, C., Spillane, C., Grossniklaus, U. (2002) Evolutionary origins of the endosperm in flowering plants, *Genome Biol.* **30**, 1026.

Bastow, R., Mylne, J.S., Lister, C., Lippman, Z., Martienssen, R.A., Dean, C. (2004) Vernalization requires epigenetic silencing of *FLC* by histone methylation, *Nature* **427**, 164–167.

Benková, E., Michniewicz, M., Sauer, M., Teichmann, T., Seifertová, D., Jürgens, G., Friml, J. (2003) Local, efflux-dependent auxin gradients as a common module for plant organ formation, *Cell* **115**, 591–602.

Berger, F. (2003) Endosperm: the crossroad of seed development, *Curr. Opin. Plant Biol.* **6**, 42–50.

Bewley, J.D. (1997) Seed germination and dormancy, *Plant Cell* **9**, 1055–1066.

Bicknell, R.A., Koltunow, A.M. (2004) Understanding apomixis: recent advances and remaining conundrums, *Plant Cell* **16**(Suppl. 1), S228–S245.

Boss, P.K., Bastow, R.M., Mylne, J.S., Dean, C. (2004) Multiple pathways in the decision to flower: enabling, promoting and resetting, *Plant Cell* **16**(Suppl.), S18–S31.

Brownlee, C., Bouget, F.Y. (1998) Polarity determination in *Fucus*: from zygote to multicellular embryo, *Semin. Cell Dev. Biol.* **9**, 179–185.

Camilleri, C., Azimzadeh, J., Pastuglia, M., Bellini, C., Grandjean, O., Bouchez, D. (2002) The *Arabidopsis TONNEAU2* gene encodes a putative novel protein phosphatase 2A regulatory subunit essential for the control of the cortical cytoskeleton, *Plant Cell* **14**, 833–845.

Carman, J.G. (1997) Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony, *Biol. J. Linn. Soc. Lond.* **61**, 51–94.

Cerdán, P.D., Chory, J. (2003) Regulation of flowering time by light quality, *Nature* **423**, 881–885.

Christensen, C.A., King, E.J., Jordan, J.R., Drews, G.N. (1997) Megagametogenesis in *Arabidopsis* wild type and the Gf mutant, *Sex. Plant Reprod.* **10**, 49–64.

Christensen, C.A., Gorsich, S.W., Brown, R.H., Jones, L.G., Shaw, J.M., Drews, G.N. (2002) Mitochondrial GFA2 is required for synergid cell death an *Arabidopsis*, *Plant Cell* **14**, 2215–2232.

Couteau, F., Belzile, F., Horlow, C., Grandjean, O., Vezon, D., Doutriaux, M.P. (1999) Random chromosome segregation without meiotic arrest in both male and female meiocytes of *dmc1* mutant of *Arabidopsis*, *Plant Cell* **11**, 1623–1634.

Dixit, R., Rizzo, C., Nasrallah, M., Nasrallah, J. (2001) *Brassica MIP-MOD* gene encodes a functional water channel that is expressed in the stigma epidermis, *Plant Mol. Biol.* **45**, 51–62.

Drews, G.N., Yadegari, R. (2002) Development and function of the angiosperm female gametophyte, *Annu. Rev. Genet.* **36**, 99–124.

Drews, G.N., Yadegari, R. (2004) Female gametophyte development, *Plant Cell* **16**(Suppl. 1), S133–S141.

Edlund, A.F., Swanson, R., Preuss, D. (2004) Pollen and stigma structure and function: the role of diversity in pollination, *Plant Cell* **16**(Suppl.), S84–S97.

Erdelská, O., Dubová, J. (2000) Double fertilization of angiosperms 1889–2000 (The origin and the significance of flowering plants double fertilization), *Biologia (Bratisl)* **55**, 311–319.

Faure, J.E., Rotman, N., Fortune, P., Dumas, C. (2002) Fertilization in *Arabidopsis thaliana* wild type: developmental stages and time course, *Plant J.* **30**, 481–488.

Friml, J. (2003) Auxin transport – shaping the plant, *Curr. Opin. Plant Biol.* **6**, 1–6.

Friml, J., Vieten, A., Sauer, M., Weijers, D., Schwarz, H., Hamann, T., Offringa, R., Jürgens, G. (2003) Efflux-dependent auxin gradients establish the apical-basal axis of *Arabidopsis*, *Nature* **426**, 147–153.

Garcia, D., Saingery, V., Chambrier, P., Mayer, U., Jürgens, G., Berger, F. (2003) *Arabidopsis haiku* mutants reveal new controls of seed size by endosperm, *Plant Physiol.* **131**, 1661–1670.

Gasser, C.S., Broadhvest, J., Hauser, B.A. (1998) Genetic analysis of ovule development, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 1–24.

Geitmann, A., Snowman, B.N., Emons, A.M., Franklin-Tong, V.E. (2000) Alterations in the actin cytoskeleton of pollen tubes are induced by the self-incompatibility reaction in *Papaver rhoeas*, *Plant Cell* **12**, 1239–1251.

Gray, W., Kepinski, S., Rouse, D., Leyser, O., Estelle, M. (2001) Auxin regulates SCF(TIR1)-dependent degradation of AUX/IAA proteins, *Nature* **414**, 271–276.

Grossniklaus, U., Schneitz, K. (1998) The molecular and genetic basis of ovule and megagametophyte development, *Semin. Cell Dev. Biol.* **9**, 227–238.

Grossniklaus, U., Nogler, G.A., van Dijk, P.J. (2001) How to avoid sex: the genetic control of gametophytic apomixis, *Plant Cell* **13**, 1491–1498.

Hamann, T., Benková, E., Bäurle, I., Kientz, M., Jürgens, G. (2002) The *Arabidopsis BODENLOS* gene encodes an auxin response protein inhibiting MONOPTEROS-mediated embryo patterning, *Genes Dev.* **16**, 1610–1615.

Hanson, M.R., Bentolila, S. (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development, *Plant Cell* **16**(Suppl.), S154–S169.

Hejátko, J., Pernisová, M., Eneva, T., Palme, K., Brzobohatý, B. (2003) The putative sensor histidine kinase CKI1 is involved in female gametophyte development in *Arabidopsis*, *Mol. Genet. Genomics* **269**, 443–453.

Herrero, M. (2001) Ovary signals for directional pollen tube growth, *Sex Plant Reprod.* **14**, 3–7.

Higashiyama, T. (2002) The synergid cell: attractor and acceptor of the pollen tube for double fertilization, *J. Plant Res.* **115**, 149–160.

Higashiyama, T., Yabe, S., Sasaki, N., Nishimura, Y., Miyagishima, S.-Y., Kuroiwa, H., Kuroiwa, T. (2001) Pollen tube attraction by the synergid cell, *Science* **293**, 1480–1483.

Hiscock, S.J., McInnis, S.M. (2003) The diversity of self-incompatibility systems in flowering plants, *Plant Biol.* **5**, 23–32.

Ito, T., Wellmer, F., Yu, H., Das, P., Ito, N., Alves-Ferreira, M., Riechmann, J.L., Meyerowitz, E.M. (2004) The homeiotic protein

AGAMOUS controls microsporogenesis by regulation of *SPOROCYTELESS*, *Nature* **430**, 356–360.

Jack, T. (2004) Molecular and genetic mechanisms of floral control, *Plant Cell* **16**(Suppl.), S1–S17.

Jürgens, G. (2001) Apical-basal pattern formation in *Arabidopsis* embryogenesis, *EMBO J.* **20**, 3609–3616.

Kachroo, A., Nasrallah, M.E., Nasrallah, J.B. (2002) Self-incompatibility in *Brassicaceae*: receptor-ligand signaling and cell-to-cell communication, *Plant Cell* **14**(Suppl.), S227–S238.

Kao, T., Tsukamoto, T. (2004) The molecular and genetic bases of S-RNase-based self-incompatibility, *Plant Cell* **16**(Suppl.), S72–S83.

Klimyuk, V.I., Jones, J. (1997) *AtDMC1*, the *Arabidopsis* homologue of the yeast *DMC1* gene: characterization, transposon-induced allelic variation and meiosis-associated expression, *Plant J.* **11**, 1–14.

Koltunow, A.M., Grossniklaus, U. (2003) Apomixis: a developmental perspective, *Annu. Rev. Plant Biol.* **54**, 547–574.

Koltunow, A.M., Bicknell, R.A., Chaudhury, A.M. (1995) Apomixis: molecular strategies for the generation of genetically identical seeds without fertilization, *Plant Physiol.* **108**, 1345–1352.

Kranz, E., Kumlehn, J. (1999) Angiosperm fertilization, embryo and endosperm development *in vitro*, *Plant Sci.* **142**, 183–197.

Kwee, H.S., Sundaresan, V. (2003) The *NOMEGA* gene required for female gametophyte development encodes the putative APC6/CDC16 component of the Anaphase Promoting Complex in *Arabidopsis*, *Plant J.* **36**, 853–866.

Lalanne, E., Twell, D. (2002) Genetic control of male germ unit organization in *Arabidopsis*, *Plant Physiol.* **129**, 865–875.

Lauber, M.H., Waizenegger, I., Steinmann, T., Schwarz, H., Mayer, U., Hwang, I., Lukowitz, W., Jürgens, G. (1997) The *Arabidopsis* KNOLLE protein is a cytokinesis-specific syntaxin, *J. Cell Biol.* **139**, 1485–1493.

Laux, T., Würschurm, T., Breuninger, H. (2004) Genetic regulation of embryonic pattern formation, *Plant Cell* **16**(Suppl.), S190–S202.

Li, H., Lin, Y., Heath, R.M., Zhu, M.X., Yang, Z. (1999) Control of pollen tube tip growth by a Rop GTPase dependent pathway that leads to tip-localised calcium influx, *Plant Cell* **11**, 1731–1742.

Lin, Ch. (2000) Photoreceptors and regulation of flowering time, *Plant Physiol.* **123**, 39–50.

Lohmann, J.U., Weigel, D. (2002) Building beauty: the genetic control of floral patterning, *Dev. Cell* **2**, 135–142.

Lord, E.M., Russel, R.D. (2002) The mechanisms of pollination and fertilization in plants, *Annu. Rev. Cell Dev. Biol.* **18**, 81–105.

Luu, D.T., Marty-Mazars, D., Trick, M., Dumas, C., Heizmann, P. (1999) Pollen-stigma adhesion in *Brassica* spp involves SLG and SLR1 glycoproteins, *Plant Cell* **11**, 251–262.

Ma, H. (1998) To be, or not to be, a flower – control of floral meristem identity, *Trends Genet.* **14**, 26–32.

Mayer, U., Torres Ruiz, R.A., Berleth, T., Misera, S., Jürgens, G. (1991) Mutations affecting body organization in the *Arabidopsis* embryo, *Nature* **353**, 402–407.

McCormick, S. (2004) Control of male gametophyte development, *Plant Cell* **16**(Suppl. 1), S142–S153.

Mercier, R., Vezon, D., Bullier, E., Motamayor, J.C., Sellier, A., Lefevre, F., Pelletier, G., Horlow, C. (2001) SWITCH (SWI1): a novel protein required for the establishment of sister chromatid cohesion and for bivalent formation at meiosis, *Genes Dev.* **15**, 1859–1871.

Mól, R., Idzikowska, K., Dumas, C., Mattys-Rochon, E. (2000) Late steps of egg cell differentiation are accelerated by pollination in *Zea mays* L., *Planta* **210**, 749–759.

Motamayor, J.C., Vezon, D., Bajon, C., Sauvanet, A., Grandjean, O., Marchand, M., Bechtold, N., Pelletier, G., Horlow, C. (2000) *Switch* (*swi*), an *Arabidopsis thaliana* mutant affected in the female meiosis switch, *Sex Plant Reprod.* **12**, 209–218.

Muschietti, J., Eyal, Y., McCormick, S. (1998) Pollen tube localization implies a role in pollen-pistil interactions for the tomato receptor-like protein kinases LePRK1 and LePRK2, *Plant Cell* **10**, 319–330.

Nakajima, K., Sena, G., Nawy, T., Benfey, P.N. (2001) Intercellular movement of the putative transcription factor SHR in root patterning, *Nature* **413**, 307–311.

Nasrallah, J.B. (2000) Cell-cell signalling in the self-incompatibility response, *Curr. Opin. Plant Biol.* **3**, 368–373.

Negrutiu, I., Vyskot, B., Barbacar, N., Georgiev, S., Moneger, F. (2001) Dioecious plants:

A key to the early events of sex chromosome evolution, *Plant Physiol.* **127**, 1418–1424.

Ng, M., Yanofsky, M.F. (2000) Three ways to learn ABCs, *Curr. Opin. Plant Biol.* **3**, 47–52.

Olsen, O.A. (2004) Nuclear endosperm development in cereals and *Arabidopsis thaliana*, *Plant Cell* **16**(Suppl. 1), S214–S227.

Otegui, M.S., Mastronarde, D.N., Kang, B.-H., Bednarek, S.Y., Staehelin, L.A. (2001) Three-dimensional analysis of syncytial-type cell plates during endosperm cellularization visualized by high resolution electron tomography, *Plant Cell* **13**, 2033–2051.

Pagnussat, G.C., Yu, H.J., Ngo, Q.A., Rajani, S., Mayalagu, S., Johnson, C.S., Capron, A., Xie, L.F., Ye, D., Sundaresan, V. (2005) Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*, *Development* **132**, 603–614.

Pezzotti, M., Feron, R., Mariani, C. (2002) Pollination modulates expression of the PPAL gene, a pistil-specific beta-expansin, *Plant Mol. Biol.* **49**, 187–197.

Pischke, M.S., Jones, L.G., Otsuga, D., Fernandez, D.E., Drews, G.N., Sussman, M.R. (2002) An *Arabidopsis* histidine kinase is essential for megagametogenesis, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15800–15805.

Raghavan, V. (2003) Some reflections on double fertilization, from its discovery to the present, *New Phytol.* **159**, 565–583.

Ratcliffe, O.J., Amaya, I., Vincent, C.A., Rothstein, S., Carpenter, R., Coen, E.S., Bradley, D.J. (1998) A common mechanism controls the life cycle and architecture of plants, *Development* **125**, 1605–1615.

Rhee, S.J., Somerville, C.R. (1998) Tetrad pollen formation in quartet mutants of *Arabidopsis thaliana* is associated with persistence of pectic polysaccharides of the pollen mother cell wall, *Plant J.* **15**, 79–88.

Rotman, N., Rozier, F., Boavida, L., Dumas, C., Berger, F., Faure, J.-E. (2003) Female control of male gamete delivery during fertilization in *Arabidopsis thaliana*, *Curr. Biol.* **13**, 432–436.

Rubery, P.H., Sheldrake, A.R. (1974) Carrier-mediated auxin transport, *Planta* **188**, 101–121.

Russel, S.D. (1996) Attraction and transport of male gametes for fertilization, *Sex. Plant Reprod.* **9**, 337–342.

Sanchez, A.M., Bosch, M., Bots, M., Nieuwland, J., Feron, R., Mariani, C. (2004) Pistil factors controlling pollination, *Plant Cell* **16**(Suppl.), S98–S106.

Schmidt, E.D.L., Guzzo, F., Toonen, M.A.J., de Vries, S.C. (1997) A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos, *Development* **124**, 2049–2062.

Schneitz, K., Hülskamp, M., Pruit, R.E. (1995) Wild type ovule development in *Arabidopsis thaliana*: a light microscope study of cleared whole-mount tissue, *Plant J.* **7**, 731–749.

Schommer, C., Beven, A., Lawrenson, T., Shaw, P., Sablowski, R. (2003) *AHP2* is required for bivalent formation and for segregation of homologous chromosomes in *Arabidopsis* meiosis, *Plant J.* **36**, 1–11.

Scott, R.J., Spielman, M., Dickinson, H.G. (2004) Stamen structure and function, *Plant Cell* **16**(Suppl. 1), S46–S60.

Skinner, D.J., Hill, T.A., Gasser, Ch.S. (2004) Regulation of ovule development, *Plant Cell* **16**(Suppl.), S32–S45.

Souter, M., Topping, J., Pullen, M., Friml, J., Palme, K., Hackett, R., Grierson, D., Lindsey, K. (2002) *Hydra* mutants of *Arabidopsis* are defective in sterol profiles and auxin and ethylene signaling, *Plant Cell* **14**, 1017–1031.

Spielman, M., Vinkenoog, R., Scott, R.J. (2003) Genetic mechanisms of apomixis, *Philos. Trans. R. Soc. London, Ser. B* **358**, 1095–1103.

Steinborn, K., Maulbetsch, C., Priester, B., Trautmann, S., Pacher, T., Geiges, B., Kuttner, F., Lepiniec, L., Stierhof, Y., Schwarz, H., Jürgens, G., Mayer, U. (2002) The *Arabidopsis* PILZ group genes encode tubulin-folding cofactor orthologs required for cell division but not cell growth, *Genes Dev.* **16**, 959–971.

Steinmann, T., Geldner, N., Grebe, M., Mangold, S., Jackson, C.L., Paris, S., Gälweiler, L., Palme, K., Jürgens, G. (1999) Coordinated polar localization of auxin efflux carrier PIN1 by GNOM ARF GEF, *Science* **286**, 316–318.

Stone, S.L., Goring, D. (2001) The molecular biology of self-incompatibility systems in flowering plants, *Plant Cell Tissue Organ. Cult.* **67**, 93–114.

Suh, S.S., Choi, K.R., Lee, I. (2003) Revisiting phase transition during flowering in *Arabidopsis*, *Plant Cell Physiol.* **44**, 836–843.

Sung, S., Amasino, R.M. (2004) Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3, *Nature* **427**, 159–164.

Tang, W., Ezcura, I., Muschietti, J., McCormick, S. (2002) A cysteine-rich extracellular protein, LAT52, interacts with the extracellular domain of the pollen receptor kinase LePRK2, *Plant Cell* **14**, 2277–2287.

Thomas, S., Osman, K., de Graaf, B.H.J., Schevchenko, G., Wheeler, M.J., Franklin, F.C.H., Franklin-Tong, V.E. (2003) Investigating mechanisms involved in the self-incompatibility response in *Papaver rhoeas*, *Philos. Trans. R. Soc. London, Ser. B* **358**, 1033–1036.

Twell, D., Park, S.K., Lalanne, E. (1998) Assymetric division and cell-fate determination in developing pollen, *Trends Plant Sci.* **3**, 305–310.

Weijers, D., Van Hamburg, J.P., Van Rijn, E., Hooykaas, P.J., Offringa, R. (2003) Diphtheria toxin-mediated cell ablation reveals interregional communication during *Arabidopsis* seed development, *Plant Physiol.* **133**, 1882–1892.

Western, T.L., Haughn, G.W. (1999) *BELL1* and *AGAMOUS* genes promote ovule identity in *Arabidopsis thaliana*, *Plant J.* **18**, 329–336.

Weterings, K., Russel, S.D. (2004) Experimental analysis of the fertilization process, *Plant Cell* **16**(Suppl.), S107–S118.

Weterings, K., Pezzotti, M., Cornelissen, M., Mariani, C. (2002) Dynamic 1-amino-cyclopropane-1-carboxylate-synthase and -oxidase transcript accumulation patterns during pollen tube growth in tobacco styles, *Plant Physiol.* **130**, 1190–1200.

Wilhelmi, L.K., Preuss, D. (1997) Blazing new trails. Pollen tube guidance in flowering plants, *Plant Physiol.* **113**, 307–312.

Willemse, M., van Lammeren, A. (2003) Fertilization, from attraction to embryo, *Biologia (Bratisl)* **57**, 13–22.

Willemsen, V., Friml, J., Grebe, M., van den Toorn, A., Palme, K., Scheres, B. (2003) Cell polarity and PIN protein positioning in *Arabidopsis* require Sterol methyltransferase1 function, *Plant Cell* **15**, 612–625.

Yadegari, R., Drews, G.N. (2004) Female gametophyte development, *Plant Cell* **16**(Suppl. 1), S133–S141.

Zheng, Z.L., Yang, Z.B. (2000) The Rop GTPase switch turns on polar growth in pollen, *Trends Plant Sci.* **5**, 298–303.

Respiratory Syncytial Virus Vaccine

Edward E. Walsh
University of Rochester School of Medicine and Dentistry, Rochester, NY

1 Introduction and History 299

2 Molecular Biology of RSV 299
2.1 Classification 299
2.2 Physical Structure of RSV 300
2.3 Genomic and Messenger RNA 301
2.4 Genetic and Antigenic Diversity among RSV Isolates (Strains) 301
2.5 Structure and Function of the RSV Proteins 302
2.5.1 Attachment Protein (G) 302
2.5.2 Fusion Protein (F) 303
2.5.3 Small Hydrophobic (SH) Protein 305
2.5.4 Matrix Protein (M) 305
2.5.5 Nucleocapsid Protein (N), Phosphoprotein (P), Polymerase (L), and the M2 Proteins (M2-1, M2-2) 305
2.5.6 Nonstructural Proteins (NS1 and NS2) 306

3 Clinical Disease and Epidemiology of RSV Infection 306

4 Immunology of RSV Infection 307
4.1 Transplacentally Acquired Maternal Antibody 307
4.2 Innate Immunity 308
4.3 Mucosal Immune Mechanisms 309
4.4 Humoral Immunity 309
4.5 Cellular Immunity 310
4.6 Disease Pathogenesis 312

5 Prior RSV Vaccine Trials 312

6 Future Approaches to an RSV Vaccine 315
6.1 Nonreplicating Subunit Vaccines 315

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

6.2 Live Recombinant Subunit Vaccines 316
6.3 Live Attenuated RSV Vaccines 316

7 Conclusions 317

Bibliography 317
Books and Reviews 317
Primary Literature 317

Keywords

Bronchiolitis
Clinical pediatric respiratory illness manifested by wheezing, cough, and hypoxia.

Paramyxovirus
The virus family to which RSV belongs.

Pneumonia
Inflammatory infectious process of the lower airway and alveoli of lung.

Pneumovirus
The virus genus to which RSV belongs.

Respiratory Syncytial Virus
An RNA virus causing respiratory illness in humans.

RNA Virus
A virus whose genome is composed of RNA.

Respiratory syncytial virus (RSV), an enveloped single-stranded negative-sense RNA virus, is an important cause of serious respiratory tract illness in infants, older adults, and persons with underlying cardiopulmonary disease. Winter time epidemics occur annually in all parts of the world and account for significant morbidity and mortality. Immunity to infection is incomplete and reinfections are common, often associated with serious consequences especially in frail elderly persons and the severely immunocompromised. There is substantial experimental evidence that the clinical characteristics are primarily influenced by the nature of the innate and cellular immune responses to infection. Development of a successful vaccine requires an understanding of the relationship between protective and pathogenic immune responses.

1 Introduction and History

Human respiratory syncytial virus (RSV) was originally isolated in 1956 from a laboratory chimpanzee with respiratory symptoms and tentatively named *chimpanzee coryza agent*. The next winter, a similar virus was recovered from two infants hospitalized with respiratory symptoms and was renamed to reflect the giant syncytia that develop in cell culture. In the ensuing decade, RSV was proven to be the single most important cause of serious lower respiratory tract infection in infants and young children. Among this age group, it is estimated that in the United States >120 000 hospitalizations and ~235 deaths annually are attributable to pneumonia and bronchiolitis caused by RSV and a substantially greater number in developing countries. In addition, RSV infection in infancy has been causally associated with recurrent wheezing during childhood. Perhaps most vexing of all, immunity to RSV is incomplete and frequent reinfection occurs throughout life. In older age groups and those with underlying cardiopulmonary diseases, RSV probably accounts for at least 80 000 hospitalizations and 13 000 deaths each winter. Certain immunocompromised adults, such as bone marrow transplant recipients, are also at risk of severe RSV infection with high mortality rates.

Despite remaining a high priority target for vaccine development since its discovery, attempts at active immunoprophylaxis, spanning nearly 40 years, have proven unsuccessful. Parenterally administered inactivated whole virion, parenteral live virus, and intranasally administered live attenuated vaccines have each been evaluated and abandoned for various reasons. The first vaccine trial using a formalin-inactivated whole virus preparation produced in the 1960 s was disastrous. Not only did the vaccine fail to protect young infants, but it was associated with more severe illness in vaccinees who became infected during the subsequent RSV season with an 80% hospitalization rate among RSV-infected vaccinees. Animal studies have provided insight into the immune mechanisms underlying this vaccine-induced enhanced disease, although a full understanding may never be available. Shortly thereafter, a parenterally administered live virus preparation also failed to protect young children from infection, although enhanced disease was not noted. Interspersed with these trials, a series of intranasally administered live attenuated virus vaccines were tested and rejected because they were either overly attenuated, poorly immunogenic, or retained excessive virulence. More recently, recombinant live attenuated vaccine with specific genetic alterations have been tested in humans following initial studies in rodents and nonhuman primates. Perhaps with a better understanding of protective immunity, molecular virology, and disease pathogenesis we can improve on the earlier efforts that were largely based on empiric approaches successfully used with viruses such as polio and mumps. This chapter will review the molecular biology, epidemiology, and immunology of RSV infection to provide a framework for assessing this task.

2 Molecular Biology of RSV

2.1 Classification

Human RSV is an enveloped nonsegmented negative-sense RNA virus,

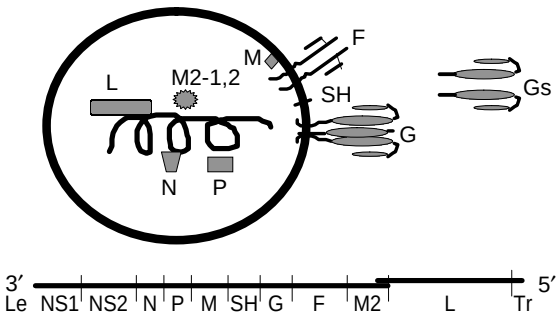

Fig. 1 Schematic structure of respiratory syncytial virus and genome organization.

classified within the genus *Pneumovirus* of the family *Paramyxoviridae*. Other closely related viruses include bovine and caprine RSV, the newly identified human metapneumovirus, turkey rhinotracheitis virus, and pneumonia virus of mice. The *paramyxoviridae* also include two other genera; *morbillivirus*, of which measles virus is the principal human pathogen, and *paramyxovirus*, containing the parainfluenza viruses of humans (PIV 1–4) and animals (Sendai virus) and mumps virus. All of these viruses share similarities of genomic organization, virion structure, and replication strategies, although RSV is distinguished by gene number and order and the absence of hemagglutinin and neuraminidase activity.

2.2 Physical Structure of RSV

By electron microscopy, RSV is a pleomorphic spherical or filamentous virus of variable diameter (80–350 nm) and length (up to 10 μm). Replication is cytoplasmic, and mature virus buds from the apical cell membrane of polarized respiratory epithelial cells. A significant proportion of viral particles remain cell associated, and analysis of purified virus is complicated by contamination with cellular material. Eight structural and three nonstructural proteins (NS1, NS2, and Gs) have been characterized in detail (Fig. 1). There are three transmembrane glycoproteins; F (fusion protein) and G (attachment protein) form trypsin sensitive 11–20 nm spikes projecting from the viral membrane, and trace amounts of a small hydrophobic protein (SH) whose function is unknown. Within the envelope, the tightly coiled helical ribonucleoprotein capsid contains genomic RNA (vRNA) encased in the viral nucleocapsid protein (N) and associated with the less abundant phosphoprotein (P), M2-1 and M2-2 proteins, and the polymerase (L) protein. The M protein is a membrane associated matrix protein and the M2-1 and M2-2 proteins, products of the M2 gene, are important in viral transcription and replication.

The virus is relatively labile under most conditions, with 99% of viral infectivity lost after 10 min at 55 °C, 48 h at 37 °C, and 2 weeks at 4 °C. Inactivation is retarded by addition of 1 M $MgSO_4$ (yielding a half-life of 12 weeks at 4 °C), rapid freezing in a dry ice-alcohol bath, and preservation at −70 °C with added glycerin or 25% sucrose. The virus has a density of 1.18 g cm^{-3} on sucrose gradients, a method commonly employed for purification.

2.3 Genomic and Messenger RNA

The single-stranded negative-sense genome contains 15 222 nucleotides (nt) and contains 10 genes flanked by noncoding leader (Le) and trailer (Tr) sequences in the following order: 3′ Le-NS1-NS2-N-P-M-SH-G-F-M2-L-Tr 5′ (Fig. 1). Messenger RNA is transcribed directly from vRNA, while synthesis of new vRNA requires a full-length positive-sense genomic template. The 44-nt leader and terminal portion of the 155-nt trailer sequence have a high degree of complementarity, a feature common to paramyxoviruses. The leader sequence is thought to promote encapsidation of the vRNA by N and serves as the sole promoter for both mRNA and the positive-sense replicative intermediate of vRNA.

Each gene begins with a highly conserved 9 nt gene-start signal (3′CCCCGUU UA) and ends with a semiconserved 12–13 nucleotide gene-stop sequence, comprised of a conserved 3′..UCAAU..5′ and a poly U_{4-7} tract with an intervening nonconserved 1–4 nt region. Nonconserved intergenic regions of 1–52 nt separate the genes. Each gene specifies a single monocistronic mRNA except for the M2 gene, which has two open reading frames that are transcribed into mRNA for the M2-1 and M2-2 proteins. In addition, a second initiation site in the transmembrane region of G allows for the production of Gs, a soluble secreted form of G protein. Transcription is initiated at a single 3′ polymerase entry site with sequential transcription of the downstream genes using a "stop–start" mechanism in which the polymerase must terminate transcription of the upstream gene before progressing across the intergenic sequence to initiate transcription of the subsequent gene. With each stop–start, there is inefficiency resulting in a decreasing gradient of transcription products in a polar fashion inversely related to the distance from the 3′ genome end. Thus, the largest amount of mRNA is noted for early genes (NS1, NS2, N) and the least for downstream ones (F, M2, L). The gene-stop sequence also regulates transcription efficiency at each gene junction. Single nucleotide changes to any of the three regions significantly alters transcription termination efficiency and amounts downstream gene mRNA.

2.4 Genetic and Antigenic Diversity among RSV Isolates (Strains)

Human RSV isolates are divided into two major groups, A and B, each with multiple subgroups based upon reactivity with monoclonal antibodies and confirmed by genetic sequence data. Monoclonal antibodies have identified antigenic differences both between RSV groups and within groups for several proteins, including F, G, P, and N with the greatest sequence divergence noted in the G protein. Genetic analysis reveals even greater heterogeneity among virus isolates (Table 1). The F, N, P, M2, NS1, and NS2 proteins are highly conserved (>78% nucleotide homology and >87% amino acid homology), as are the noncoding leader, trailer, and gene-start sequences. In contrast, the attachment protein G, followed by SH, have the greatest degree of genetic variability. Analysis of multiple isolates indicates that there is ~50% amino acid sequence variability between group A and B virus G genes, and as much as 20% within groups. Concordantly, there is low antigenic relatedness (~3–7%) between the G proteins of group A and B virus, determined by cross neutralization with G

Tab. 1 Genetic and amino acid homology between and within RSV groups.

Protein	*Group A*	*Group B*	*Homology [%]*	
			Nucleotide	*Amino Acid*
F_0	A2	18 537	79	89
F_1	A2	18 537		93
F_2	A2	18 537		83
G	A2	18 537	67	53
G	Within group			80
G		Within group		88
SH	A2	18 537	78	76
N	A2	18 537	86	96
P	A2	18 537	80	90
L	A2	18 537	ND	ND
M	A2	18 537	ND	ND
M2	A2	18 537	78	92
NS2	A2	18 537	78	92
NS1	A2	18 537	78	87
Intergenic	A2	18 537	Little homology	NA

ND = not reported
NA = not applicable.

specific polyclonal antisera or by EIA with postprimary infection human sera. A high proportion of the nucleotide changes resulting in amino acid substitutions occur in hypervariable regions that contain neutralizing epitopes and supports the notion that sequence changes may be the result of antibody mediated immune pressure. Detailed analysis of the evolution of G protein during the past 40 years indicates slowly accumulating (0.25% per year) amino acid substitutions. Variability in the G protein can be due to single nucleotide substitutions, frameshifts, and changes in location of termination codons.

2.5 Structure and Function of the RSV Proteins

2.5.1 Attachment Protein (G)

The G protein is the primary viral attachment protein, although replication competent viruses lacking the entire G gene have been described. G has several unusual features that distinguish it from the more common hemagglutinin-neuraminidase and hemagglutinin attachment proteins of other paramyxoviruses (Fig. 2). G is ~298 amino acids in length which predicts a 33 kDa size, although numerous carbohydrate side chains account for the final 90 kDa size. G is a type II envelope protein with a hydrophobic transmembrane region 38 residues from the amino terminus. Striking features include a high serine and threonine content (30.6%) that provide numerous acceptor sites (>70) for short O-linked carbohydrate side chains. There are also four predicted N-linked glycosylation sites. This glycosylation pattern, rare for viral glycoproteins, is characteristic of mucinous proteins secreted by respiratory epithelium. A high proline content (13% in the ectodomain)

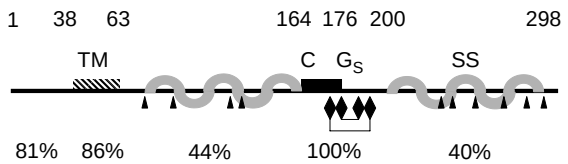

Fig. 2 Schematic representation of the RSV G protein. The amino acid numbers are indicated at the top, and the degree of homology between group A and B viruses for each of the regions are shown at the bottom. The transmembrane region is indicated by the hatched bar, and the central conserved region of 13 amino acids by the black bar. The conserved cysteine noose (CX3C motif) is represented by diamonds with intermolecular disulfide bonds by the lines. The mucin-like O-linked glycosylation sites are represented by the gray lines. Potential N-linked sites are indicated by the triangles.

may contribute to a nonglobular structure. The intracytoplasmic and transmembrane regions are well conserved (84%) between group A and B strains, while the ectodomain contains variable and hypervariable regions that flank a conserved 13 aa cysteine rich region (aa 163–176). This later region has structural and functional analogy to a cysteine rich motif found in the CX3C chemokine fractalkine (see Sect. 4.2). During translation, the 33 kDa G protein undergoes stepwise modification by the addition of N-linked complex high mannose carbohydrate side chains, bringing the molecular weight to 45 kDa, followed by the addition of O-linked carbohydrate yielding the mature 90-kDa protein. The carbohydrate residues play important roles in the synthesis of G, viral infectivity, and immunogenicity. Transport of G to the cell surface is diminished only when both N- and O-linked sugars are absent and enzymatic removal of carbohydrate from purified virus reduces infectivity by 98%. Finally, many Mabs react only with fully glycosylated forms of G, although neutralizing and protective Mabs, which react with the nonglycosylated 33 kDa G protein, have been described. Genetic analysis of Mab escape mutants and reactivity of Mabs and human serum with synthetic peptides have identified several discrete neutralizing sites on G, most of which are group or subgroup specific, although at least one neutralizing site in the conserved cysteine rich area has been described.

In addition to the membrane anchored G found in virions, a truncated G is secreted from infected cells (Gs), the result of translation starting at Met48 in the transmembrane region. The role of Gs in viral growth and disease pathogenesis is unknown, although it has been shown to bind the fractalkine (CX3CR) receptor and have chemokine-like properties (see Sect. 4.2).

Although G is the principal RSV attachment protein and binds to cell membrane heparin-like glycosaminoglycans and to CX3CR, naturally occurring and engineered replication competent viruses lacking the G gene have been described. F protein alone apparently can provide the cell attachment function. However, these deletant viruses replicate more slowly and to lower titer than virus expressing all three surface glycoproteins.

2.5.2 Fusion Protein (F)

The RSV F protein is a type I transmembrane glycoprotein that, in its mature proteolytically cleaved trimeric form, is responsible for pH-independent fusion of the viral envelope with the cell plasma membrane and facilitates cell entry of the nucleocapsid complex. During replication, F induces cell-to-cell fusion that

produces the characteristic syncytia seen in cell culture, and facilitates spread of virus to adjacent cells. The importance of cell–cell spread in human infection is unknown although microscopic examination of lungs from infants who succumb to RSV have demonstrable giant cell formation. F is 574 amino acids in length and has a calculated molecular weight of 63 kDa (Fig. 3). Three of five potential N-linked carbohydrate acceptor sites are used, which increases the molecular weight to 70 kDa. Similar to the fusion proteins of other paramyxoviruses, F is synthesized as an inactive precursor and posttranslationally modified in the Golgi by simultaneous addition of N-linked complex carbohydrate side chains (Fo) and cleaved by furin-like proteases into disulfide linked 48-kDa (F_1) and 23-kDa (F_2) subunits to form an active fusion protein. Upon cleavage, a long hydrophobic region on the amino terminus of F_1 is exposed, which, by analogy to other paramyxoviruses, is probably important in membrane fusion. Interestingly, the F_2 fragment, and not the G protein, determines host cell tropism. Like the G protein, the F protein binds to cell-associated heparin sulfate and may contribute up to 25% of the virus-cell binding. Two heptad repeats, at either end of the F_1 fragment, are important in fusion activity, and participate in secondary and tertiary structure of the final fusion permissive trimeric F protein that has a coiled coil structure. Although cells infected with recombinant vectors expressing F protein alone form syncytia, coexpression of G and SH are required for maximal fusogenic activity. The F protein also interacts with RhoA, a surface expressed cellular signaling protein associated with many cellular functions relevant to RSV pathogenesis.

The F protein contains major conserved neutralizing epitopes as defined by Mab binding studies and analysis of neutralization escape mutants. Both fusion-inhibiting and nonfusion-inhibiting neutralizing sites have been described, most of which are conserved in group A and B RSV isolates. Fusion-inhibiting neutralizing Mabs provide maximal protection in passive protection studies in rodents and one of these Mabs has been humanized (palivizumab; Synagis®) and licensed for prophylactic administration to high-risk infants. The binding sites of neutralizing Mabs have been mapped to regions on F by analysis of antibody binding to synthetic peptides, trypsin digest fragments of F_1, or genetically cloned fragments of F, as well as by sequence analysis of neutralization escape mutants. The majority bind to the N-terminal portion of F_1, between amino acids 200–230 or 250–290. At least one of the neutralizing, fusion-inhibiting Mabs binds to amino acids 283–315. In addition to B-cell epitopes, T-helper cell

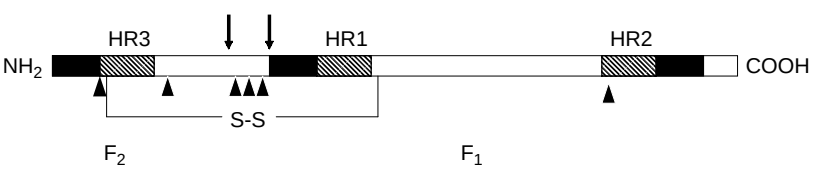

Fig. 3 Schematic representation of RSV fusion protein F. The signal peptide, cleavage activation region and transmembrane regions are shown in dark bars. The three heptad repeats are identified by hatched bars. Triangles indicate potential N-linked glycosylation sites, and the arrows indicate furin-like protease cleavage sites. The disulfide linkage of F_1 and F_2 fragments is indicated.

epitopes map to the F_1 fragment between amino acids 233–278 and 328–355. F also contains cytotoxic T-cells (CTL) epitopes although they are not well defined.

2.5.3 Small Hydrophobic (SH) Protein

The smallest of the RSV transmembrane proteins, SH, is analogous to a similar protein found in some paramyxoviruses. It is of unknown function, but may be important for efficient fusion activity of the F protein. The SH gene sequence indicates a protein of 64 amino acids, with a predicted size of 7.5 kDa. Multiple forms of SH, including 13–15 and 21–40 kDa molecules, are abundantly present in infected cells although only small amounts are found in purified virions. SH is characterized by a hydrophobic membrane spanning region from residue 14–42, an external C-terminus, and two potential glycosylation sites. Neutralizing epitopes have not been described; however, T- and B-cell epitopes have been identified. Genetically engineered RSV constructs lacking the SH gene (RSVΔSH) are fully replication competent.

2.5.4 Matrix Protein (M)

The RSV matrix protein is loosely associated with the inner surface of the viral membrane and by analogy, to other paramyxoviruses, which is important in assembly of viral proteins into budding virions at the cell membrane. M is 256 amino acids in length, has a 25 amino acid hydrophobic region near the C-terminus presumably for membrane interaction, and is phosphorylated.

2.5.5 Nucleocapsid Protein (N), Phosphoprotein (P), Polymerase (L), and the M2 Proteins (M2-1, M2-2)

The vRNA is encapsidated by the 44 kDa N protein, a requirement for transcription and genomic replication. The 32-kDa phosphoprotein (P) and the polymerase (L) comprise the RNA-dependent RNA polymerase, which directs transcription and vRNA replication. Transcriptional activity of the nucleocapsid complex peaks at 14 h postinfection and is severely restricted by 30 h when vRNA replication begins. The cytoskeletal protein actin is also required for transcription, since actin-specific antibody is inhibitory *in vitro*. Phosphorylation of P by cellular casein kinase II at a conserved serine residue (Ser^{237}) is critical for efficient transcription since wild-type unphosphorylated P or mutant P with an amino acid substitution at residue 237 have minimal transcriptional activity. Binding of P to the encapsidated vRNA, however, is not dependent upon Phosphorylation since mutant P competes with wild-type P binding.

The RSV L gene comprises 43% of the vRNA and encodes a protein of 2165 amino acids with a molecular weight of 250 kDa. The L nucleotide sequence defines four polymerase motifs common to other paramyxovirus polymerases. One conserved region contains an ATP-binding motif required for protein kinase activity. *In vitro* experiments have defined the optimum molar ratio of the N, P, and L proteins for transcription at 12 : 5 : 1. Notably, many of the nucleotide changes associated with temperature sensitivity during virus growth have been localized to the L gene.

The M2 gene encodes two products, M2-1 and M2-2, that are important in viral replication. M2-1, a product of the first M2 open reading frame, is an antitermination protein that enhances transcriptional processivity. Early unsuccessful attempts to generate recombinant virus with cDNA and plasmids expressing N, P, and L yielded incomplete mRNAs. The addition

of an M2-1 expressing plasmid allowed complete transcription of mRNA and production of recombinant virus. The 90 aa M2-2 protein functions as a regulatory factor involved in the balance between transcription and RNA replication, favoring the latter by way of a positive-sense vRNA intermediate.

2.5.6 Nonstructural Proteins (NS1 and NS2)

Unique among paramyxoviruses, RSV has two nonstructural proteins, which are synthesized in abundance due to the 3′ promoter proximal location of their corresponding genes. They function as antagonists of cellular interferons (IFN-α and -β), which are induced by RSV in respiratory epithelial cells as part of the innate immune response to infection (see Sect. 4.2). Recombinant viruses lacking these genes grow poorly in cells capable of interferon synthesis, and are attenuated in animal models of RSV.

3 Clinical Disease and Epidemiology of RSV Infection

RSV causes annual epidemics from late fall through early spring, and has been found in every locale in which it is sought. Overall, RSV accounts for ~250 000 total hospitalizations and ~10–14 000 deaths in all age groups annually in the United States. The virus is uncommon during the summer months except in tropical climates, where it coincides with the rainy season. RSV is the dominant cause of serious respiratory tract infection in the young, and accounts for nearly half of all pneumonia, 75 to 90% of bronchiolitis cases, and 10% of the group that requires hospitalization in the first year of life. During their first winter, 70% of infants are infected with the remainder infected the following season. Although most infants under a year of age have mild upper or lower respiratory tract symptoms, 1 to 5% are hospitalized depending on the presence of underlying medical conditions. The mean age at hospitalization is 2 to 3 months, a time when transplacentally derived antibody levels would have fallen to approximately one-tenth of those at birth. Importantly, infants appear to be spared serious infection in the first few weeks of life when antibody levels are highest. Risk factors for severe disease are present in approximately 25 to 50% of all hospitalized infants and include premature birth (<36 weeks gestation), congenital cardiopulmonary disease (with mortality rates of 37%), bronchopulmonary dysplasia, cystic fibrosis, congenital neurologic defects, and immunodeficiency states.

RSV is transmitted principally by large particle fomites, rather than by small particle aerosol-like influenza, with an incubation period of 2 to 5 days before the onset of rhinnorhea, cough, and fever. Lower respiratory symptoms of wheezing, tachypnea, and hypoxia will develop in approximately 30% of infants with first infections. It is estimated that >100 000 hospitalizations and ~250 deaths occur annually in the United States, although worldwide estimates are substantially greater, especially in underdeveloped and developing countries.

Reinfections occur throughout life, underscoring the imperfect immunity induced by natural infection. In the day care setting, three quarters of exposed, previously infected infants become reinfected. Generally, repeat infection is less severe than primary infection, and children above age three typically manifest mild upper respiratory tract symptoms. Adults are also at risk for reinfection and

47% of parents become infected when the virus is introduced into the home. In healthy working adults, illness is generally mild to moderate in severity, although lower respiratory tract signs and symptoms develop in 26% and 38% miss work. Among healthy elderly persons >65 years of age, 32% will seek medical attention during RSV infection. Overall, attack rates are ~5% in adult populations. In certain high-risk adult populations, RSV reinfections can cause severe illness. Nosocomial or community-acquired infection in bone marrow transplant recipients progresses to pneumonia in half of infected persons with 30 to 78% mortality. Residents of long-term care facilities or adults of any age with underlying cardiopulmonary disease may also become seriously ill with pneumonia and bronchitis. This age group represents an expanding population that may benefit from vaccination against RSV. Longitudinal studies in Rochester, New York, noted that RSV was associated with wintertime hospitalization of high-risk or elderly persons for acute cardiopulmonary symptoms and is nearly as frequent as influenza A virus infection (9.6 vs 10.4%, respectively). Overall, RSV probably accounts for >150 000 hospitalizations and ~14 000 deaths annually in elderly and high-risk adults in the United States.

During epidemics, virus strains representing both group A and B can be recovered. During a 15-year period (1974–1990) in Rochester, New York, group A viruses accounted for 71% of isolates overall, and only in four winters did group B isolates dominate. In contrast, during an eight-year period (1982–1990) in Caen, France, group B dominated four years and, overall, accounted for 64% of the isolates. The ratio of A to B viruses from 14 separate cities representing all regions of the United States during two consecutive winters varied considerably, thus suggesting that outbreaks of RSV are local in nature rather than national or global. In addition, there is regional and temporal variations in the circulation of viral groups, and subgroups. The clinical importance of antigenic variation on disease severity or in immunity has not been conclusively demonstrated, but is considered by many investigators to be relevant. Following primary infection, the antibody response to G is group specific and even subgroup specific. In addition, analysis of repeat infections in two small studies concluded that homotypic immunity may influence the RSV group responsible for secondary infection. Although not definitively proven, it is thought that the G protein antigenic changes are the results of antibody mediated immune pressure.

4 Immunology of RSV Infection

Although immunity is clearly incomplete, prior infections do reduce the frequency and severity of subsequent infection. Notably, there is substantial evidence, primarily from animal studies with inference to humans, that severity of infection and disease manifestations of RSV, in large part, are a result of the immune responses to the virus.

4.1 Transplacentally Acquired Maternal Antibody

Although the effect of passively acquired neutralizing antibody is limited owing to its rapid decline, most studies demonstrate reduced infection rates and severity of primary infection with higher levels

of antibody. Animal experiments support these conclusions. More definitively, prophylactic administration of high-titer RSV gammaglobulin or F protein specific monoclonal antibody to high-risk infants provides benefit. In these trials, severity of illness, but not infection rates, were reduced by half in treated infants. In experimental animals, protection is directly related to the serum neutralizing titer, with nearly complete suppression of virus growth at titers exceeding 1 : 380 dilution of serum. Murine monoclonal antibodies to either F or G protein are protective in animal models, in contrast to antibodies to internal viral proteins (N, P, M) (Table 2). Antibody dependent cell-mediated cytotoxicity and complement are not necessary for protection since F(ab_2) fragments can also inhibit RSV replication *in vivo*.

4.2 Innate Immunity

The earliest immune response to RSV involves nonadaptive innate immune mechanisms. Immediately following virus exposure, cells express and secrete a large number of proinflammatory and immunomodulatory cytokines and chemokines. Some have direct antiviral and inflammatory effects, while others also affect the subsequent T-cell adaptive immune response. Substances induced from upper and lower respiratory tract epithelial cells *in vitro* include the cytokines interleukin (IL)-1, tumor necrosis factor (TNF)-α, and IL-6, the antiviral proteins IFN-α and -β, the CXC chemokine IL-8 (neutrophil chemoattractant), the CC chemokines RANTES, macrophage inflammatory protein (MIP)-1α, macrophage chemoattractant protein (MCP)-1, and the CX3C chemokine fractalkine. Additionally, intracellular adhesion molecules, major histocompatibility complex class I antigens and antigen transporter (TAP)-1 are also induced in infected cells. Some responses are triggered by the interaction of RSV with Toll-like receptor (TLR)-4 present on cells. The CD14-TLR-4 complex, a member of a family of proteins structurally related to Drosophila Toll proteins, directly interacts with the RSV F protein with subsequent expression of proinflammatory cytokines. In addition, it has been shown that RSV infection is associated with stimulation of natural killer (NK) cells, which also plays a role in the adaptive immune response.

Both RSV G protein or Gs bind directly to the fractalkine receptor (CX3CR) present on immune cells and, similar

Tab. 2 Immune responses to RSV.

Protein	*Antibody response to natural infection*	*CTL epitopes*	
		Mouse	*Human*
F	Neutralizing, fusion inhibiting	+	+
G	Neutralizing	−	−
SH	Nonneutralizing	−	+
N	Nonneutralizing	+	++
P	Nonneutralizing	+	+
M	Nonneutralizing	−	+
M2	Nonneutralizing	++	+

to fractalkine, can act as a chemoattractant. However, G also competitively blocks fractalkine binding to CX3CR, and thus the net effect on immune response is unclear. RSV also directly induces secretion of the tachykinin neuropeptide substance P (SP), a product of efferent neurons, dendritic cells, and lymphocytes. SP may exacerbate inflammation by mediating vasodilatation. *In vivo* studies in mice demonstrate that inhibition of SP reduces pulmonary inflammatory infiltrates during RSV infection. A similar role of SP in humans has not been confirmed.

4.3 Mucosal Immune Mechanisms

One of the earliest adaptive immune response to infection is the appearance of RSV-specific secretory IgA in nasal secretions, which coincides with disappearance of virus. In experimental animals, the IgA response to F and G proteins is relatively brief and the protective capacity and durability of IgA in humans is unknown. Some but not all adult challenge studies indicate that nasal wash neutralizing capacity correlates with resistance to infection. Similarly, the ability to infect infants and children with live attenuated viruses is not consistently correlated to nasal IgA. Possibly the greatest indictment of the protective capacity of nasal IgA is that healthy adults are readily infected experimentally just weeks after natural infection. In animal experiments, nasal virus replication is inhibited to a greater extent by local secretory IgA than serum IgG, although IgA-mediated immunity is brief (<8 months), in contrast to more durable pulmonary immunity from serum IgG (>18 months). Animal studies also suggest that T-cell immunity, especially $CD8^+$ cells, play a role in mucosal immunity.

4.4 Humoral Immunity

The infant's humoral response to primary and repeat infections has been extensively investigated with most attention given to anti-F, anti-G, and neutralizing antibody responses. Newly synthesized IgG is detected 14 days after primary infection, peaks at 1 to 2 months, and then slowly declines sometimes reaching undetectable levels by one year. Infants above the age of 8 months are more likely to have responses to F (100%) or G (78%) than younger infants (62 and 48%, respectively), perhaps due to the suppressive effects of preexisting maternally derived antibody and immunological immaturity. Not surprisingly, the antibody response to G is group specific, consistent with the strain-specific amino acid differences, although moderate cross reactivity to the heterologous G is noted when the infecting strain is of group A. Analysis of antibody response using peptides representing subgroup specific hypervariable regions also demonstrates subgroup specific responses. Although the F response is highly cross-reactive, postinfectious neutralizing capacity is greatest against the homologous virus group with primary infection. Possibly of importance in vaccine development, group A RSV infection induces better cross neutralization than group B infection. The IgG subclass response to F and G following primary infection is predominantly IgG_1 and IgG_3, subclasses usually induced by protein antigens. This was unexpected since carbohydrate antigens (such as the heavily glycosylated G protein) usually stimulate IgG_2. Following three or more infections, IgG levels of F and G are boosted, reaching adult levels after the third infection. Interestingly, young adults with repeat infection have

rather weak antibody responses despite significant symptoms. In contrast, elderly adults with RSV infection have very strong EIA and neutralizing serum antibody responses to infection, exceeding those of young adults. The mechanism behind this difference is unknown. Serologic studies in elderly adults have confirmed that both severity and illness rates are less in persons with higher levels of neutralizing antibody.

The frequent presence of wheezing during primary RSV infection prompted measurement of RSV-specific IgE in serum and nasal secretions. Wheezing infants are more likely to have virus-specific IgE, histamine, and eosinophilic cationic protein in nasal secretions. The relationship of RSV-specific IgE in nasal secretions during acute RSV to future episodes of wheezing or development of asthma is uncertain although several studies implicate RSV IgE with increased wheezing during the first 4 to 5 years of life. In addition, bronchoconstrictive leukotrienes are released by macrophages infected *in vitro* in the presence of RSV IgG, a finding which strengthens the notion that symptoms are related to the host immune response. Interestingly, RSV infection in HIV-infected infants, develop diffuse interstitial pneumonia rather than bronchiolitis.

4.5 Cellular Immunity

Cell-mediated immunity is the primary mechanism for recovery from most virus infections, and there is data which suggests that the specific nature of the T-cell response determines the clinical manifestations of infection. Furthermore, the T-cell response to various RSV vaccine strategies appears to dictate the clinical outcome of subsequent RSV exposure. In general, replicating virus stimulates T-helper cells ($CD4^+$) with a Th_1 cytokine profile (interferon-γ) in contrast to inactivated or nonreplicating antigens that stimulate a Th_2 cytokine profile (IL-4, IL-5, IL-13). Th_1 responses are associated with processing of replicating antigen in context with class I histocompatibility proteins resulting in production of neutralizing IgG_{2a} (mouse system) or IgG_1 (human system) antibody and $CD8^+$ CTL. In contrast, Th_2 responses develop when nonreplicating antigens are processed in association with class II histocompatibility proteins resulting in synthesis of IgG_1 and IgE (mouse system) or IgG_4 and IgE (human system) without CTL induction.

In the murine model, clearance of RSV is unaffected by B-cell deficiency, whereas T-cell depletion results in persistent virus shedding. Virus eradication correlates with induction of $CD4^+$ cells with Th_1 cytokine characteristics resulting in the expected IgG_{2a} and CTL. $CD8^+$ CTL are lytic for target cells infected with either group A and B virus and are detected in the lungs of mice on day 7 postchallenge when clearance of virus occurs. Recent evidence also suggests that RSV may persist in the lungs of infected mice, as evidenced by detection of vRNA for up to 3 months following challenge. The significance of this is unknown but provocative data from patients with chronic obstructive pulmonary disease suggests that persistence may also occur in humans.

Infusions of RSV-specific $CD8^+$ CTL lines or clones clears virus in normal or persistently infected immunodeficient mice. Using target cells infected with recombinant vaccinia viruses (rvv) expressing RSV proteins has identified the N, F, and M2 proteins and possibly P as CTL targets in the mouse system. Mice

immunized with rvv-N or rvv-M2 are protected from RSV challenge, but protection wanes by day 28. Immunization with rvv-M2 confers protection from both group A and B challenge, perhaps due to a specific CTL epitope that has been mapped to a sequence (amino acids 82–90) conserved between both major RSV groups (see Sect. 2.5.4). Immunization of mice with a DNA vaccine expressing this short peptide induces high-level resistance to virus challenge, which can persist for several months.

Important for vaccine development, the priming immunization significantly alters T-helper cytokine mRNA expression in the lungs of mice when they are subsequently challenged with virus. Live virus immunization stimulates a Th_1 pattern and following virus challenge, replication is restricted and the pulmonary infiltrate is mild with predominantly PMNs and lymphocytes. In contrast, intramuscular (im) injection of FI-RSV or purified F protein (PFP) induces a Th_2 pattern with more extensive infiltration of lymphocytes and eosoinophils. Notably, vaccination with a live vector expressing F induces a Th_1 response similar to live virus. In contrast, immunization with purified G protein, rvv-G, or rvv-Gs all induce a Th_2 cytokine profile without CTL, uses the $V\beta14^+$ T-cell receptor in processing of G, and the pulmonary infiltrate following challenge is predominately eosinophils in nature. Although the response induced by G is similar to FI-RSV, the G protein is not a requirement for FI-RSV induced eosinophilia. These findings are analogous to the eosinophilic infiltrates seen in infants who died of RSV infection in the earlier FI-RSV vaccine trials.

The respective roles of $CD4^+$ and $CD8^+$ T cells in virus clearance has been investigated by passive transfer or depletion of specific cell types prior to virus challenge. Depletion of either $CD4^+$ or $CD8^+$ cells results in higher and slightly prolonged virus shedding, while depletion of both subsets results in prolonged, high titered virus shedding. Passive transfer of either RSV-specific $CD4^+$ or $CD8^+$ cells reduces virus lung growth, but a dose dependent increase in pulmonary pathology, characterized by lung hemorrhage and polymorphonuclear infiltration is noted when large numbers of $CD8^+$ T cells are infused, consistent with the concept that both $CD4^+$ and $CD8^+$ cells contribute in individual ways to the inflammatory process. The mechanism employed by $CD8^+$ CTL may also influence outcome, as perforin mediated cytotoxicity is more efficient and targeted than Fas ligand mediated killing in which uninfected cells are also lysed. This is the evidence that IL-4 (a type 2 cytokine) can stimulate the latter response with resultant nonspecific cell death and delayed virus clearance. The clinical manifestations of illness in the murine model (ruffled fur, weight loss) is also linked to TNF-α secretion. Overall, the results from the mouse model reinforces the concept that immune responses to RSV influence clinical illness.

Considerably less is known about the cell-mediated immune response in humans, especially infants. There is *in vitro* evidence of impaired lymphocyte proliferation responses to RSV; specifically, stimulation of peripheral blood mononuclear cells *in vitro* with live virus suppresses lymphocyte proliferation in contrast to inactivated RSV or live influenza virus. It has been speculated that this may contribute to suboptimal immune responses and contribute to reinfection. Technical limitations have impeded study of the infant's CTL response. Nevertheless, 4 of 22

infants, ages 9 days to 11 months hospitalized with RSV bronchiolitis, had measurable CTL activity in peripheral blood mononuclear cells within the first week of infection. In another study, 5 of 13 infants under 6 months of age and 8 of 12 infants over 6 months of age developed a CTL response, generally within the first week after infection, which rapidly dissipates in the ensuing months. In another study of RSV-infected hospitalized infants, both Th_1 and Th_2 cytokine profiles could be detected by flow cytometry. The specific antigens involved in the human CTL response has only been examined in adults; N, SH, F, M, M2, and NS1 have been identified as CTL targets (Table 2). Peripheral blood mononuclear cells from adults and infants produce both Th_1 cytokines (γ-interferon) and Th_2 cytokines (IL-4, IL-5) when stimulated *in vitro*.

4.6 Disease Pathogenesis

The clinical manifestations of RSV infection and severity of disease are multifactorial. There is evidence that severity can be related to the RSV group or subgroup, with several reports suggesting that group A virus infections are more severe. However, the primary contributor to the inflammatory response and subsequent clinical manifestations are thought to be immune responses to the virus, attributable to specific RSV proteins. The innate and humoral immune responses have also been implicated in disease severity in either animal models or in humans. Host genetic variability may also contribute substantially to disease manifestations. Several investigators have analyzed cytokine gene polymorphism in infants with RSV infection and found mutations in IL-4, IL-4 receptor, IL-6, IL-9, IL-10, and TNF-α genes to be associated with more severe disease. In addition, genetic variability in Toll-like receptor-4 (TLR4) also has been identified as a factor that influences disease severity. Thus, it appears that multiple genetic factors influence the innate and adaptive immune responses to RSV infection and influence clinical severity.

5 Prior RSV Vaccine Trials

Four major approaches to RSV vaccination have been assessed in infants or adults: (1) parenterally administered inactivated whole virus, (2) intranasally administered live attenuated viruses, (3) parenterally administered live virus, and (4) parenterally administered subunit immunization.

In 1969, a large multicenter field trial of a formalin-inactivated whole virion vaccine (FI-RSV) in infants was published. The results of this study have had profound effects on subsequent RSV vaccine research and development. The vaccine virus (Bernett strain; group A virus) was grown in vervet monkey kidney cells, formalin inactivated and concentrated by centrifugation and adsorption to alum. Children, ages 3 months to 9 years, received 1 to3 immunizations over 3 months followed by careful surveillance. The humoral immune response, measured by an insensitive complement fixation assay and by neutralization titers, appeared satisfactory. In the ensuing winter, RSV attack rates in controls and vaccinees were similar; however, illness was consistently more severe among recipients of the RSV vaccine compared to unvaccinated controls or controls immunized with a similarly prepared trivalent parainfluenza vaccine. The youngest infants who were RSV naïve, were most adversely affected, although

children 12 to 24 months of age who may have already experienced an RSV infection, also had more severe illness. The clinical manifestations were indistinguishable from a typical RSV illness, but lower respiratory tract involvement was considerably more common (69 vs 9%), and hospitalization was 5 to 16 times more frequent among vaccine recipients. Peripheral eosinophilia was reported in 56% of RSV vaccinees compared to 11% of controls at one of the centers, although not at another. The adverse outcome, since designated as vaccine-associated enhanced RSV disease, has been likened to the atypical measles syndrome, which followed use of a similarly prepared formalin-inactivated measles virus vaccine. Two decades later, stored serum from these infants was found to contain binding antibody to F and G, which lacked functional activity in virus neutralization and fusion inhibition assays.

The pathogenesis of enhanced RSV disease associated with the FI-RSV vaccine has been meticulously scrutinized in various animal models, with the hope that the safety of future vaccines can be better predicted. Cotton rats immunized with original FI-RSV vaccine (lot 100) and challenged with live RSV develop more peribronchiolar lymphocytic and alveolar neutrophilic infiltration than controls immunized with im or intranasal live virus. Similar to the clinical trials, binding antibody to F and G proteins were induced, but neutralizing and fusion-inhibiting activity were diminished. Passive transfer of antibody from FI-RSV immunized mice did not reproduce the findings; thus, antibody alone is insufficient to cause enhanced disease.

As noted earlier, FI-RSV induces a Th_2 cytokine profile from $CD4^+$ T cells without demonstrable CTL activity that is associated with an eosinophilic pulmonary infiltrate. Depletion of $CD4^+$ cells prior to virus challenge of FI-RSV immunized mice reduces pulmonary infiltrates to a greater extent than $CD8^+$ depletion. Furthermore, FI-RSV induced histologic changes are diminished by *in vivo* depletion of IL-4 and IL-10 (Th_2 cytokines) but not by depletion of γ-interferon or IL-2 (Th_1 cytokines). These results have led to the suggestion that a Th_1 $CD4^+$ response with CTL induction is the desired result of an RSV vaccine, while a Th_2 response, in which CTL is lacking, may predispose vaccine recipients to enhanced disease.

Concurrent with the FI-RSV trials, live attenuated vaccines were sought by adapting virus to preferentially grow in the cooler nasal tissues rather than in the warmer lower respiratory tract by repeated passage at low temperature or by randomly inducing genetic mutations and searching for temperature-sensitive (*ts*) virus. The first such viruses, reported in 1968, were cold-passaged (*cp*) variants of the A2 strain. Viruses adapted to 34 and 28 °C successfully infected a high proportion of adult volunteers, but retained virulence. Further attenuation by repeated passage at 26 °C produced a candidate that was less infectious (8 of 45 challenged adults) and less virulent, but when tested in children, significant upper respiratory tract illness developed in three who had not been previously infected with RSV. Next, temperature-sensitive (*ts*) mutants, produced by growing A2 strain in 5-fluourouracil were tested. One mutant, identified as *ts*-1, replicated exclusively below 38 °C in tissue culture and was restricted to nasal tissues in hamsters. Initial trials in 9 adults found that although only 3 shed virus or had a serologic response, 6 of 7 who were challenged with wild-type A2 virus were protected. In children,

this vaccine was highly infectious (34 of 39 inoculated) and immunogenic (21 of 39 humoral responses and 31 of 39 nasal antibody responses), but nonimmune infants again developed upper respiratory illness. Equally disturbing, reversion to wild-type growth patterns was found. A 100-fold dose reduction eliminated symptoms but revertant virus was again recovered and protective efficacy from natural RSV was not discernible. Further alteration of *ts*-1 by nitrosoguanidine treatment resulted in overattenuation. Another mutant (*ts*-2) with deficient syncytia formation was poorly infectious and also had reversion.

The final early vaccine to undergo field testing was a subcutaneously administered live RSV preparation derived from a clinical isolate. Animal studies provided experimental evidence of both upper and lower respiratory tract protection. Following pilot studies in over 400 infants and children in whom immunogenicity and safety profiles were favorable, a placebo controlled efficacy study was performed in 510 infants above age 6 months. Serologic status and age at time of vaccination appeared to determine immunogenicity. Overall, 68% of seronegative infants responded in contrast to only 16% of seropositive infants. Seronegative infants >12 months of age had the highest response rates (90%). Follow-up during the two subsequent RSV seasons documented neither protection nor enhanced illness attributable to vaccination, regardless of the serologic response to immunization.

More recent efforts to develop live attenuated vaccines have utilized reverse genetics to construct and selectively tailor virus vaccine candidates. Genetic analysis of the above noted *ts* and *ca* mutants allowed investigators to combine growth limiting mutations in different genes in order to further attenuate the virus and inhibit reversion. Several have been tested in adults, older children, and infants, but the same difficulties of overattenuation and poor immunogenicity, or excessive virulence have been noted. The most promising of these newer vaccine candidates, known as *cpts* 248/404 (designating mutations in the N, F M2, and L genes), was judged to be only minimally virulent in the youngest of infants, and preliminary evidence suggested it may be protective.

In addition to live attenuated viruses, subunit vaccines, which often rely on the common F antigen have been assessed in specific populations. These vaccines are generally not considered to be viable candidates for RSV naïve infants, due to their propensity to induce Th2 responses in mice, with the associated inflammatory infiltrates upon virus challenge. One such vaccine of PFP adsorbed to alum has been studied in phase I and II trials in older high-risk children with cystic fibrosis and in healthy elderly and frail nursing home subjects. Although safe and moderately immunogenic when administered intramuscularly, further studies have not been carried out. PFP has also been given during pregnancy to boost maternal transfer of antibody to newborns and ~35 pregnant women demonstrated reasonable immunogenicity and there was increased antibody titers in newborns. Another subunit vaccine composed of a conserved region of the G protein (aa 130–230) linked to the albumin-binding domain of streptococcal protein G as a carrier (designated BBG2Na). Although protective in animals, this vaccine stimulates only low levels of neutralizing antibody. Unfortunately, when tested in a phase II clinical study some patients developed a type II hypersensitivity reaction characterized as *vasculitic purpura*. Finally, a subunit vaccine composed of purified F, G, and M

proteins has been found to be highly immunogenic in adults with induction of neutralizing antibody and to be without significant side effects. Efficacy trials in high-risk adults have not been carried out at this time.

Beginning in 1995, further attempts to attenuate a previous *ca*RSV strain by further mutagenesis yielded a series of genetically well-characterized vaccine candidates. These strains, containing nucleotide substitutions resulting in the *ca* phenotype (N, F, L genes) and *ts* phenotype (M2 and L genes), were restricted to growth in the upper respiratory tract of mice and nonhuman primates and were protected from wild-type RSV challenge. Several were tested sequentially in adults, older children, and seronegative children >6 months of age. Although two, identified as *cpts* 248/955 and 530/1009, were not sufficiently attenuated in seronegative infants, one strain (*cpts* 248/404) was evaluated in infants as young as one month. Unfortunately, it caused nasal congestion, which affected feeding and sleeping in a few of the vaccines. Although this strain also was abandoned, it was capable of inducing serum neutralizing antibody in children above age 6 months and possibly protective serum neutralizing IgA in the youngest infants.

6 Future Approaches to an RSV Vaccine

Development of a protective and safe RSV vaccine remains a formidable task. In choosing potential vaccine candidates, the target population must be considered carefully. Considering current knowledge, the ideal vaccine would at least mimic, if not improve upon, the immune response to natural infection. This includes stimulation of mucosal IgA, humoral neutralizing antibodies to both major RSV groups, and a Th_1 type cellular immune response with CTLs recognizing virus from both groups. The optimal vaccine would be immunogenic in infants less than 2 months of age and overcome the immunosuppression of maternal antibody and the generally immature immune system of the very young infant. Research to determine those viral components that contribute to virulence may also allow specific deletion of the offending region of the protein or even entire genes. In contrast, parenterally or perhaps mucosally administered subunit vaccines may be effective in previously infected children and adults to prevent reinfection, and for maternal immunization to boost transplacental delivery of antibody in pregnant women. Since natural infection only provides partial immunity, a conservative, but realistic, goal of immunization is reduction of illness severity rather than prevention of infection. In addition to safety, the successful vaccine must carefully navigate a fine line between beneficial and detrimental immune stimulation.

6.1 Nonreplicating Subunit Vaccines

Given the current state of knowledge and technical tools available, single or combined viral protein subunit vaccines, comprised of entire proteins or smaller peptides, can currently be devised and evaluated. In addition to choosing the appropriate viral antigen, route of administration (parenteral, mucosal) and inclusion of the proper adjuvant are important considerations in attempting to induce the correct immune response. Because of the

catastrophic events clouding the formalin-inactivated vaccine trials, clinical trials with inactivated subunit vaccines for infants would proceed very slowly, if at all. Adjuvants and/or immune modulators such as cytokines capable of stimulating Th_1 responses may ultimately prove useful in shifting the response to nonreplicating antigens from Th_2 to Th_1 in nature.

Another more indirect way to protect infants is vaccination of women during pregnancy, an approach being considered for a variety of infant infections. Modest increases in neutralizing antibody may be sufficient to lessen the severity of RSV infection or delay infection until the child is older and less susceptible to severe disease. The other major vaccine target group, the elderly and high-risk adults, may also benefit from immunization with a subunit vaccine.

Alternate routes of administration and newer adjuvants can significantly alter the immune response to these subunit vaccines. Intranasal administration of F or an FG chimeric protein with the mucosal adjuvant cholera toxin (CT) or its purified B subunit (CTB) induces protective mucosal and systemic immunity in mice. Although CTB is unlikely to be approved for use in humans, other mucosal adjuvants may provide similar capabilities. Mucosal immunization is also theoretically possible via the oral route, and is attractive because of its simplicity. There has been one report of an immunogenic G peptide fragment synthesized in a plant, with the ultimate goal of high yield inexpensive production with the possibility of an oral vaccine.

In addition to entire viral proteins, peptide fragments of F and G have also been proposed as potential vaccines. Regions of F, which represent neutralizing or CTL epitopes, can be synthesized. One F peptide, representing the neutralizing site from amino acids Ile221 to Glu232, elicited low level neutralizing antibody in mice but protection has not been demonstrated. In addition, a peptide representing amino acids 174–187 from the G protein provided complete protection from virus challenge, which was relatively durable. The principle advantage of a synthetic peptide vaccine is that production of large quantities is relatively simple. On the other hand, vaccination of a heterogeneous population may be difficult, if not impossible, due to the limited immune response to single epitopes.

6.2 Live Recombinant Subunit Vaccines

Replicating vectors expressing the desired RSV protein(s) may provide a viable approach to immunization. Replication incompetent vectors such as the VEE system, have the advantage of potentially inducing a Th_1 response with appropriate humoral and cellular immunity to envelope glycoproteins. Although vaccinia virus and adenovirus vectors have been developed and tested in mice and primates, they did not progress to human trials. Recently, a bovine parainfluenza type 3 virus expressing the RSV surface glycoproteins has been constructed and tested in animals. Finally, DNA vaccination is another approach to subunit immunization with a replicating antigen, although it has not been fully explored.

6.3 Live Attenuated RSV Vaccines

Despite prior frustrations, there remains interest in intranasal live attenuated vaccines. The availability of reverse genetics for production of vaccines allows virtual unlimited choices. Deletions of entire

nonessential genes is feasible and several (designated RSV ΔNS1, ΔNS2, ΔSH, ΔG) have been tested in animals and some in humans. Until the specific viral proteins or regions of proteins invoked in disease pathogenesis are identified, live vaccine development will remain relatively empiric since predictions of attenuation based upon growth *in vitro* or in experimental animals is not always precise. It may be possible that important immunogenic proteins can be left unaltered or even enhanced, yet viral growth and inflammatory responses are limited by alterations in other genes. For instance, the F and/or G proteins, which carry dominant protective epitopes, can be moved proximal to the promoter region, thus producing a virus with greater expression of F and G. Alternatively, substitutions in the intergenic sequences could enhance or diminish expression of downstream genes.

7 Conclusions

Important advances have been made in our understanding of the virology, immunology, and pathogenic mechanisms of RSV infection during the past decade. Despite rapid development in methods of vaccine preparation, advances in immune modulators and mechanisms of vaccine delivery, the prospects for an RSV vaccine remain tempered by the prior failures.

See also Immunology; Nucleic Acid Packaging of RNA Viruses; Vaccination, Genetic; Vaccine Research and Development Methodology; Viral Inhibitors and Immune Response Mediators: The Interferons.

Bibliography

Books and Reviews

Collins P. (1991) The Molecular Biology of Human Respiratory Syncytial Virus (RSV) of The Genus *Pneumovirus*, in: Kingsbury, D.W. (Ed.) *The Paramyxoviruses*, Plenum Publishing, 103–162.

Collins, P.L., Murphy, B.R. (2002) Respiratory syncytial virus: reverse genetics and vaccine strategies, *Virology* **296**, 204–211.

Durbin, A.P., Karron, R.A. (2003) Progress in the development of respiratory syncytial virus and parainfluenza vaccines, *Clin. Infect. Dis.* **37**, 1668–1677.

Hall, C.B. (2001) Respiratory syncytial virus and parainfluenza virus, *N. Engl. J. Med.* **334**, 1917–1928.

Polack, F.P., Karron, R.A. (2004) The future of respiratory syncytial virus vaccine development, *Pediatr. Infect. Dis. J.* **23**, S65–S73.

Walsh, E.E., Graham, B.S. (1999) Respiratory Syncytial Virus, in: Dolin, R. Wright, P.F. (Eds.) *Lung Biology in Health and Disease: Viral Infections of the Respiratory Tract*, Marcel Dekker, New York, pp. 161–204.

Primary Literature

Alwan W., Record F., Openshaw, P. (1992) CD4+ T cells clear virus but augment disease in mice infected with respiratory syncytial virus. Comparison with the effects of $CD8^+$ T cells, *Clin. Exp. Immunol.* **88**, 527–536.

Anderson, L., Heilman, C.A. (1995) Protective and disease-enhancing immune responses to respiratory syncytial virus, *J. Infect. Dis.* **171**, 1–7.

Anderson, L.J., Hendry, R.M., Pierik, L.T., Tsou, C., McIntosh, K. (1991) Multicenter study of strains of respiratory syncytial virus, *J. Infect. Dis.* **163**, 687–692.

Bangham, C.R.M., McMichael, A. (1986) Specific human cytotoxic T cells recognize B-cell lines persistently infected with respiratory syncytial virus, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9183–9187.

Bangham, C.R.M., Openshaw, P., Ball, L.A., King, A.M.W., Wertz, G.W., Askonas, B.A.

(1986) Human and murine cytotoxic T cells specific to respiratory syncytial virus recognize the viral nucleoprotein (N), but not the major glycoprotein (G), expressed by vaccinia virus recombinants, *J. Immunol.* **137**, 3973–3977.

Bartholdy, C., Olszewska, W., Stryhn, A., Thomsen, A.R., Openshaw, P.J.M. (2004) Gene-gun DNA vaccination aggravates respiratory syncytial virus-induced pneumonitis, *J. Gen. Virol.* **85**, 3017–3026.

Belanger, H., Fleysh, N., Cox, S., Bartman, G., Deka, D., Trudel, M., Koprowski, H., Yusibov, V. (2000) Human respiratory syncytial virus vaccine antigen produced in plants, *FASEB J.* **14**, 2323–2328.

Belshe, R.B., Newman, F.K., Anderson, E.L. et al. (2004) Evaluation of combined live, attenuated respiratory syncytial virus and parainfluenza 3 virus vaccines in infants and young children, *J. Infect. Dis.* **190**, 2096–2103.

Bembridge, G.P., Rodriguez, N., Garcia-Beato, R., Nicolson, C., Melero, J.A., Taylor, G. (2000) Respiratory syncytial virus infection of gene gun vaccinated mice induces Th2-driven pulmonary eosinophilia even in the absence of sensitisation to the fusion (F) or attachment (G) protein, *Vaccine* **19**, 1038–1046.

Bermingham, A., Collins, P.L. (1999) The M2-2 protein of human respiratory syncytial virus is a regulatory factor involved in the balance between RNA replication and transcription, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11259–11264.

Brock, S.C., Goldenring, J.R., Crowe, J.E. Jr. (2003) Apical recycling systems regulate directional budding of respiratory syncytial virus from polarized epithelial cells, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15143–15148.

Cane, P.A. (1997) Analysis of linear epitopes recognized by the primary human antibody response to a variable region of the attachment (G) protein of respiratory syncytial virus, *J. Med. Virol.* **51**, 297–304.

Cane, P.A., Pringle, C. (1997) Evolution of subgroup A respiratory syncytial virus respiratory syncytial virus: evidence for progressive accumulation of amino acid changes in the attachment protein, *J. Virol.* **69**, 2918–2925.

Cannon, M.J., Stott, E.J., Taylor, G., Askonas, B.A. (1987) Clearance of persistent respiratory syncytial virus infections in immunodeficient mice following transfer of primed T cells, *Immunology* **62**, 133–138.

Cartee, T.L., Megaw, A.G., Oomens, A.G.P., Wertz, G.W. (2003) Identification of a single amino acid change in the human respiratory syncytial virus L protein that affects transcriptional termination, *J. Virol.* **77**, 7352–7360.

Chanock, R., Roizman, B., Myers, R. (1957) Recovery from infants with respiratory illness of a virus related to chimpanzee coryza agent (CCA), *Am. J. Hyg.* **66**, 281–290.

Cherrie, A.H., Anderson, K., Wertz, G.W., Openshaw, P.J.M. (1992) Human cytotoxic T cells stimulated by antigen on dendritic cells recognize the N, SH, F, M 22 K, and 1b proteins of respiratory syncytial virus, *J. Virol.* **66**, 2102–2110.

Collins, P.L., Huang, Y.T., Wertz, G.W. (1984) Nucleotide sequence of the gene encoding the fusion (F) glycoprotein of human respiratory syncytial virus, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7683–7687.

Collins, P.L., Hill, M.G., Christina, J., Grosfeld, H. (1996) Transcription elongation factor of respiratory syncytial virus, a nonsegmented negative-strand RNA virus, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 81–85.

Collins, P.L., Hill, M.G., Camargo, E., Grosfeld, H., Chanock, R.M., Murphy, B.R. (1995) Production of infectious human respiratory syncytial virus from cloned cDNA confirms an essential role for the transcription elongation factor from the 5′ proximal opening reading frame of the M2 mRNA in gene expression and provides a capability for vaccine development, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11563–11567.

Connors, M., Collins, P.L., Firestone, C.-Y., Murphy, B.R. (1991) Respiratory syncytial virus (RSV) F, G, M2 (22K) and N proteins each induced resistance to RSV challenge, but resistance induced by M2 and N proteins is relatively short-lived, *J. Virol.* **65**, 1634–1637.

Connors, M., Kulkarni, A.B., Firestone, C.-Y., et al. (1992) Pulmonary histopathology induced by respiratory syncytial virus (RSV) challenge of formalin-inactivated RSV-immunized BALB/c mice is abrogated by depletion of $CD4^+$ T cells, *J. Virol.* **66**, 7444–7451.

Crowe J.E. Jr., Bui P.T., London W.T., et al. (1994) Satisfactorily attenuated and protective mutants derived from a partially attenuated cold-passaged respiratory syncytial virus mutant by introduction of additional

attenuating mutations during chemical mutagenesis, *Vaccine* **12**, 691–699.

Dowell S.F., Anderson L.J., Gary H.E. Jr., et al. (1996) Respiratory syncytial virus is an important cause of community-acquired lower respiratory infection among hospitalized adults, *J. Infect. Dis.* **174**, 456–462.

Falsey, A.R., Walsh, E.E. (1996) Safety and immunogenicity of a respiratory syncytial virus subunit vaccine (PFP-2) in ambulatory adults over age 60, *Vaccine* **14**, 1214–1218.

Falsey, A.R., Walsh, E.E. (1997) Safety and immunogenicity of a respiratory syncytial virus subunit vaccine (PFP-2) in the institutionalized adult, *Vaccine* **15**, 1130–1132.

Falsey, A.R., Walsh, E.E. (1998) Relationship of serum antibody to risk of respiratory syncytial virus infection in elderly adults, *J. Infect. Dis.* **177**, 463–466.

Falsey, A.R., Walsh, E.E. (2000) Respiratory syncytial virus infection in adults, *Clin. Microbiol. Rev.* **13**, 371–384.

Feldman, S.A., Hendry, R.M., Beeler, J.A. (1999) Identification of a linear heparin binding domain for human respiratory syncytial virus attachment glycoprotein G, *J. Virol.* **73**(8), 6610–6617.

Fernie, B.F., Gerin, J. (1980) The stabilization and purification of respiratory syncytial virus using MgSO4, *Virology* **106**, 141–144.

Garofalo, R.P., Patti, J., Hintz, K.A., Ogra, P.L., Welliver, R.C. (2001) Macrophage inflammatory protein-1 alpha (not T helper type 2 cytokines) is associated with severe forms of respiratory syncytial virus bronchiolitis, *J. Infect. Dis.* **184**(4), 393–399.

Glezen, W.P., Taber, L.H., Frank, A., Kasel, J. (1986) Risk of primary infection and reinfection with respiratory syncytial virus, *Am. J. Dis. Child.* **140**, 543–546.

Goetsch, L., Plotnicky-Gilquin, H., Aubry, J.P., et al. (2001) BBG2Na an RSV subunit vaccine candidate intramuscularly injected to human confers protection against viral challenge after nasal immunization in mice, *Vaccine* **19**, 4036–4042.

Graham, B.S., Henderson, G., Tang, Y., Lu, X., Neuzil, K., Colley, D. (1993) Priming immunization determines T helper cytokine mRNA expression patterns in lungs of mice challenged with respiratory syncytial virus, *J. Immunol.* **151**, 2032–2040.

Hall, C.B., Schnabel, K.C., Gieman, J.M., Douglas, R.G. (1981) Infectivity of respiratory syncytial virus by various routes of inoculation, *Infect. Immun.* **33**, 779–783.

Hall, C.B., Walsh, E.E., Schnabel, K.C., et al. (1990) Occurrence of groups A and B of respiratory syncytial virus over 15 years: associated epidemiologic and clinical characteristics in hospitalized and ambulatory children, *J. Infect. Dis.* **192**, 1283–1290.

Hancock, G.E., Smith, J.D., Heers, K.M. (2000a) Serum neutralizing antibody titers of seropositive chimpanzees immunized with vaccines coformulated with natural fusion and attachment proteins of respiratory syncytial virus, *J. Infect. Dis.* **181**, 1768–1771.

Hancock, G.E., Smith, J.D., Heers, K.M. (2000b) The immunogenicity of subunit vaccines for respiratory syncytial virus after co-formulation with aluminum hydroxide adjuvant and recombinant interleukin-12, *Viral Immunol.* **13**(1), 57–72.

Haynes, L.M., Moore, D.D., Kurt-Jones, E.A., Finberg, R.W., Anderson, L.J., Tripp, R.A. (2001) Involvement of toll-like receptor 4 in innate immunity to respiratory syncytial virus, *J. Virol.* **75**, 10730–10737.

Heidema, J., de Bree, G.J., de Graff, P.M.A. et al. (2004) Human CD8+ T cell responses against five newly identified respiratory syncytial virus-derived epitopes, *J. Gen. Virol.* **85**, 2365–2374.

Hendry, R.M., Burns, J.C., Walsh, E.E. et al. (1988) Strain-specific serum antibody responses in infants undergoing primary infection with respiratory syncytial virus, *J. Infect. Dis.* **61**, 3163–3166.

Hendry, R.M., Talis, A.L., Godfrey, E., Anderson, L.J., Fernie, B.F., McIntosh, K. (1986) Concurrent circulation of antigenically distinct strains of respiratory syncytial virus during community outbreaks, *J. Infect. Dis.* **153**, 291–297.

Hull, J., Thomson, A., Kwiatkowski, D. (2000) Association of respiratory syncytial virus bronchiolitis with the interleukin 8 gene region in UK families, *Thorax* **55**(12), 1023–1027.

Johnson, T.R., Teng, M.N., Collins, P.L., Graham, B.S. (2004) Respiratory syncytial virus (RSV) G glycoprotein is not necessary for vaccine-enhanced disease induced by immunization with formalin-inactivated RSV, *J. Virol.* **78**, 6024–6032.

Johnson, P.R., Spriggs, M.K., Olmsted, R.A., Collins, P.L. (1987) The G glycoprotein of human respiratory syncytial viruses of subgroups A and B: extensive sequence divergence between antigenically related proteins, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5625–5629.

Johnson, P.R., Olmsted, R.A., Prince, G.A., et al. (1987) Antigenic relatedness between glycoproteins of human respiratory syncytial virus subgroups A and B: evaluation of the contributions of F and G glycoproteins to immunity, *J. Virol.* **61**, 3163–3166.

Johnson, T.R., Varga, S.M., Braciale, T.J., Graham, B.S. (2004) Vβ14$^+$ T cells mediate the vaccine-enhanced disease induced by immunization with respiratory syncytial virus (RSV) G glycoprotein but not with formalin-inactivated RSV, *J. Virol.* **78**, 8753–8760.

Karron R.A., Wright P.F., Crowe J.E. Jr., et al. (1997) Evaluation of two live, cold-passaged, temperature-sensitive respiratory syncytial virus vaccines in chimpanzees and in human adults, infants and children, *J. Infect. Dis.* **176**, 1428–1436.

Karron, R., Buonagurio, D., Georgiu, A., et al. (1997) Respiratory syncytial virus (RSV) SH and G proteins are not essential for viral replication in vitro: clinical evaluation and molecular characterization of a cold-passaged, attenuated RSV subgroup B mutant, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13961–13966.

Kim, H.W., Canchola, J.G., Brandt, C.D., et al. (1969) Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine, *Am. J. Epidemiol.* **89**, 422–434.

Kurt-Jones, E.A., Popova, L., Kwinn, L., Haynes, L.M., Jones, L.P., Tripp, R.A., Walsh, E.E., Freeman, M.W., Golenbock, D.T., Anderson, L.J., Finberg, R.W. (2000) Pattern recognition receptors TLR4 and CD14 mediate response to respiratory syncytial virus, *Nat. Immunol.* **1**(5), 398–401.

Lee, F.E.-H., Walsh, E.E., Falsey, A.R., Betts, R.F., Treanor, J.J. (2004) Experimental infection of humans with A2 respiratory syncytial virus, *Antiviral Res.* **63**, 191–196.

Matthews, J.M., Young, T.F., Tucker, S.P., MacKay, J.P. (2000) The core of the respiratory syncytial virus fusion protein is a trimeric coiled coil, *J. Virol.* **74**, 5911–5920.

McKay, E., Higgins, P., Tyrell, D., Pringle, C. (1988) Immunogenicity and pathogenicity of temperature sensitive modified respiratory syncytial virus in adult volunteers, *J. Med. Virol.* **75**, 411.

Melero, J.A., Garcia-Barreno, B., Martinez, I., Pringle, C.R., Cane, P.A. (1997) Antigenic structure, evolution and immunobiology of human respiratory syncytial virus attachment (G) protein, *J. Gen. Virol.* **78**, 2411–2418.

Moudy, R.M., Harmon, S.B., Sullender, W.M., Wertz, G.W. (2003) Variations in transcription termination signals of human respiratory syncytial virus clinical isolates affect gene expression, *Virology* **313**, 250–260.

Mufson, M.A. (1987) Subgroup characteristics of respiratory syncytial virus strains recovered from children with two consecutive infections, *J. Clin. Microbiol.* **25**, 1535–1539.

Munoz, F.M., Piedra, P.A., Glezen, W.P. (2003) Safety and immunogenicity of respiratory syncytial virus purified fusion protein-2 vaccine in pregnant women, *Vaccine* **21**, 3465–3467.

Murphy, B.R., Walsh, E.E. (1988) Formalin-inactivated respiratory syncytial virus vaccine induces antibodies to the fusion glycoprotein that are deficient infusion-inhibiting activity, *J. Clin. Microbiol.* **26**, 1595–1597.

Nicolson, K.G. (1996) Impact of influenza and respiratory syncytial virus on mortality in England and Wales from January 1975 to December 1990, *Epidemiol. Infect.* **116**, 51–63.

Piedmonte, G., King, K.A., Holmgren, N.L., Bertrand, P.J., Rodriguez, M.M., Hirsch, R.L. (2000) A humanized monoclonal antibody against respiratory syncytial virus (Palivizumab) inhibits RSV-induced neurogenic-mediated inflammation in rat airways, *Pediatr. Res.* **47**, 351–356.

Polack, F.P., Teng, M.N., Collins, P.L., et al. (2002) A role for immune complexes in enhanced respiratory syncytial virus disease, *J. Exp. Med.* **196**, 859–865.

Prince, G.A., Horswood, R.L., Chanock, R.M. (1985) Quantitative aspects of passive immunity to respiratory syncytial virus infection in infant cotton rats, *J. Virol.* **55**, 517–520.

Schlender, J., Zimmer, G., Herrier, G., Conzelmann, K.-K. (2003) Respiratory syncytial virus (RSV) fusion protein subunit F2, not attachment protein G, determines the specificity of RSV infection, *J. Virol.* **77**, 4609–4616.

Schwarze, J., O'Donnell, D.R., Rohwedder, A., Openshaw, P.J.M. (2004) Latency and

persistence of respiratory syncytial virus despite T cell immunity, *Am. J. Respir. Crit. Care Med.* **169**, 801–805.

Sullender, W.M. (2000) Respiratory syncytial virus genetic and antigenic diversity, *Clin. Microbiol. Rev.* **13**(1), 1–15.

Sutherland, K.A., Collins, P.L., Peeples, M.E. (2001) Synergistic effects of gene-end signal mutations and the M2-1 protein on transcription termination by respiratory syncytial virus, *Virology* **288**, 295–307.

Tang, R.S., MacPhail, M., Schickli, J.E., et al. (2004) Parainfluenza virus type 3 expressing the native or soluble fusion (F) protein of respiratory syncytial virus (RSV) confers protection from RSV infection in African green monkeys, *J. Virol.* **78**, 11198–11207.

Teng, M.N., Collins, P.L. (2002) The central conserved cysteine noose of the attachment G protein of human respiratory syncytial virus is not required for efficient viral infection in vitro or in vivo, *J. Virol.* **76**, 6164–6171.

Teng, M.N., Whitehead, S.S., Collins, P.L. (2001) Contribution of the respiratory syncytial virus G glycoprotein and its secreted and membrane-bound forms to virus replication in vitro and in vivo, *Virology* **289**, 283–296.

Tripp, R.A., Dakhama, A., Jones, L.P., Barksey, A., Gelfand, E.W., Anderson, L.J. (2003) The G glycoprotein of respiratory syncytial virus depresses respiratory rates through the CX3C motif and substance P, *J. Virol.* **77**, 6580–6584.

Tripp, R.A., Jones, L.P., Haynes, L.M., HaoQiang, Z., Murphy, P.M., Anderson, L.J. (2001) CX3C chemokine mimicry by respiratory syncytial virus G glycoprotein, *Nat. Immunol.* **2**(6), 732–738.

Tripp, R.A., Moore, D., Barksey, A., Jones, L., Moscatiello, C., Keyserling, H., Anderson, L.J. (2002) Peripheral blood mononuclear cells from infants hospitalized because of respiratory syncytial virus infection express T helper-1 and T helper-2 cytokines and CC chemokine messenger RNA, *J. Infect. Dis.* **185**, 1388–1394.

Tristram, D.A., Welliver, R., Mohar, C.K., Hogerman, D.A., Hildreth, S.W., Paradiso, P. (1993) Immunogenicity and safety of respiratory syncytial virus subunit vaccine in seropositive children 18–36 months old, *J. Infect. Dis.* **167**, 191–195.

Walsh, E.E., Falsey, A.R. (2004) Humoral and mucosal immunity in protection from natural respiratory syncytial virus infection in adults, *J. Infect. Dis.* **190**, 373–378.

Walsh, E.E., Falsey, A.R., Hennessey, P.A. (1999) Respiratory syncytial virus and other infections in persons with chronic cardiopulmonary disease, *Am. J. Respir. Crit. Care Med.* **160**, 791–795.

Walsh, E.E., Peterson, D.R., Falsey, A.R. (2004) Risk factors for severe respiratory syncytial virus infection in elderly adults, *J. Infect. Dis.* **189**, 233–238.

Walsh, E.E., McConnochie, K.M., Long, C.E., Hall, C.B. (1997) Severity of respiratory syncytial virus infection is related to virus strain, *J. Infect. Dis.* **175**, 814–820.

Walsh, E.E., Hruska, J. (1983) Monoclonal antibodies to respiratory syncytial virus proteins: identification of the fusion protein, *J. Virol.* **47**, 171–177.

Walsh, E.E., Brandriss, M.W., Schlesinger, J.J. (1985) Purification and characterization of the respiratory syncytial virus fusion protein, *J. Gen. Virol.* **66**, 409–415.

Walsh, E.E., Schlesinger, J.J., Brandriss, M. (1986) Purification and characterization of GP90, one of the envelope glycoproteins of respiratory syncytial virus, *J. Gen. Virol.* **65**, 761–767.

Welliver, R.C., Duffy, L. (1993) The relationship of RSV-specific immunoglobulin E antibody response in infancy, recurrent wheezing, and pulmonary function at age 7–8 years, *Pediatr. Pulmonol.* **15**, 19–27.

Wertz, G.W., Moudy, R.M. (2004) Antigenic and genetic variation in human respiratory syncytial virus, *Pediatr. Infect. Dis. J.* **23**, S19–S24.

Wertz, G.W., Collins, P.L., Huang, Y., Gruber, C., Levine, S., Ball, L.A. (1985) Nucleotide sequence of the G protein gene of human respiratory syncytial virus reveals an unusual type of viral membrane protein, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4075–4079.

Wright, P.F., Karron, R.A., Belshe, R.B. et al. (2000) Evaluation of a live, cold-passaged, temperature-sensitive, respiratory syncytial virus vaccine candidate in infancy, *J. Infect. Dis.* **182**, 1331–1342.

Wright, P.F., Gruber, W.C., Peters, M., Reed, G., Zhu, Y., Robinson, F., Coleman-Dockery, S., Graham, B.S. (2002) Illness severity, viral shedding, and antibody responses in infants hospitalized with bronchiolitis caused by

respiratory syncytial virus, *J. Infect. Dis.* **185**, 1011–1018.

Zambon, M.C., Stockton, J.D., Clewley, J.P., Fleming, D.M. (2001) Contribution of influenza and respiratory syncytial virus to community cases of influenza-like illness: an observational study, *Lancet* **358**, 1410–1416.

Restriction Endonuclease and DNA-Modification Methyltransferases

Albert Jeltsch[1] *and Richard I. Gumport*[2]
[1] *International University Bremen, Bremen, Germany*
[2] *University of Illinois, Urbana, IL*

1 Introduction 325
1.1 Restriction-modification Systems 325

2 Types of Restriction-modification Systems 327
2.1 Type II Systems 327
2.2 Type I and III Systems 328

3 Mechanisms of Action of ENases and MTases 328
3.1 Common Features 328

4 Catalytic Mechanism of ENases 329
4.1 Structures of ENases 329
4.2 Mechanism of Target-site Location 330
4.3 Mechanism of DNA Recognition 331
4.4 Mechanism of DNA Cleavage 332

5 Catalytic Mechanism of DNA MTases 334
5.1 Mechanism of DNA Methylation 334
5.2 Reaction Mechanism of C5-MTases 335
5.3 Mechanism of N-MTases 336
5.4 Structures of MTases 337
5.5 Mechanism of DNA Recognition 337
5.6 Mechanism of Target-site Location and Lack of Processivity of Methylation 338

Bibliography 339
Books and Reviews 339
Primary Literature 340

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

Keywords

Bacteriophage
A virus that infects bacteria and ultimately reproduces itself usually leading to the death of the bacterium.

Endonuclease
An enzyme that cleaves the internal phosphodiester bonds of a polynucleotide, in this article of a double-strand DNA.

Heterocycle
A molecule with a ring structure in which not all the atoms are carbon. The bases A, C, G, and T in DNA are aromatic heterocyles with nitrogen in place of some of the carbon atoms.

Homodimer
A protein comprising two identical polypeptide chains.

Metaphosphate
The anion PO_3^-.

Methylsulfonium Bond
A covalent bond between a divalent sulfur atom and a methyl group that has a formal positive charge on the sulfur and is thermodynamically activated for transfer of the methyl group to an acceptor molecule.

Methyltransferase
An enzyme that transfers methyl groups from S-adenosyl-L-methionine to acceptor molecules.

Motif
A sequence of amino acids that is shared (conserved) among different proteins.

Nucleophile
An atom or group with excess electrons that is capable of reacting with another atom or group that is deficient in electrons.

Palindrome
A sequence of nucleotides on one strand of a DNA duplex that is identical to the sequence in the opposite direction on its complementary strand, for example, GAATTC in one strand is the same sequence on the antiparallel, base-paired strand.

Restriction Endonuclease
An endonuclease that cleaves duplex DNA with very high sequence specificity.

Stoichiometry
The quantitative relationship between the substances in a compound or in a molecular complex.

Restriction-modification systems occur in many prokaryotes. Each consists of a restriction endonuclease that cleaves DNA after recognizing a specific sequence and a DNA methyltransferase that modifies the bacterial DNA at the recognized sequence. Methylation protects the bacterial genome from cleavage by the endonuclease. Since infecting bacteriophage DNA usually is not methylated, it is cleaved by the restriction enzyme before it can be methylated by the bacterial methyltransferase. Thus, the bacteriophage DNA is inactivated and the restriction-modification system protects the bacterial cell from bacteriophage infection. In this review, we summarize the properties of different types of restriction-modification systems and the enzymology of the enzymes involved. Our focus is primarily on the type II systems.

1 Introduction

1.1 Restriction-modification Systems

Restriction-modification (RM) systems were discovered in the 1950s by S. E. Luria and M. L. Human as well as by G. Bertani and J. J. Weigle. These groups found that after growing bacteriophages on one bacterial strain, the progeny phages could reinfect this same strain with high efficiency, but they could not infect some other bacterial strains. These observations demonstrated that the phage carries a ''label'' imparted by the strain on which it was grown. A decade later, the molecular basis of this phenomenon was elucidated by Arber and colleagues who showed that RM systems consist of two enzyme activities: a restriction endonuclease (ENase) that specifically recognizes short, often palindromic, sequences and then cleaves the DNA, and a methyltransferase (MTase) that methylates the same sequence and renders it refractory to cleavage by the ENase. The activity of the ENase is responsible for the restriction phenotype because it digests phage DNA entering a bacterial cell. In the 1970s, H. O. Smith and coworkers and H. W. Boyer and coworkers were the first to publish the purification of functional type II ENases (*Hind*II from *Haemophilus influenza* and *Eco*RI from *Escherichia coli*, respectively) and the H. O. Smith laboratory identified the first DNA recognition sequence of an ENase (*Hind*II).

The DNA of a bacterial strain carrying an RM system is protected from cleavage by the resident ENase as a result of covalent modification carried out by a DNA methyltransferase. As indicated, the MTase modifies the DNA within the recognition sequence of the partner ENase and thereby prevents endonuclease action (Fig. 1). The first RM system MTase was described in Arber's laboratory. Methylation of the DNA uses S-adenosyl-L-methionine (AdoMet) as the methyl

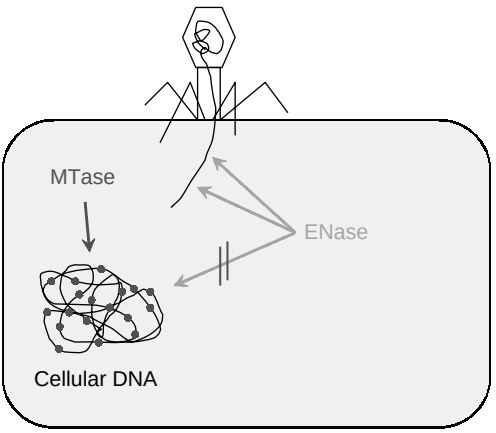

Fig. 1 Biological function of restriction/modification systems. ENases cleave unmethylated phage DNA, the bacterial genome is protected against the endonucleolytic attack by the DNA methyltransferase.

group donor and occurs at the cytosine-C5, cytosine-N4, or adenine-N6 positions of DNA depending on the particular MTase. It represents a minimal change to the molecular structure of the DNA and does not interfere with the Watson/Crick base pairing of the double helix. In all three positions of methylation, the methyl groups are positioned in the major groove of the DNA, a region where the restriction endonucleases recognize the sequence of the DNA by formation of specific contacts to the edges of the base pairs. Methylation of the recognition sequence interferes with the ENase interaction with the DNA and prevents ENase cleavage. Therefore, the activity of the ENases can be controlled by the methylation state of the DNA (see below).

Since different RM systems have different recognition sequences, the methylation pattern is, in effect, a bar code on the DNA that allows an ENase to distinguish between its own and foreign DNA. Given that DNA from bacteriophages is often not modified at the sites of the resident bacterial RM systems, RM systems efficiently protect bacteria against bacteriophage infection. However, phages grown in a bacterial cell containing an RM system will have DNA modified by the resident MTase, which explains why they can easily grow in the same strain, but not in other strains containing different RM systems that recognize different DNA sequences. RM systems control the uptake and incorporation of foreign DNA from any source into the bacterium. They are widely distributed in the bacterial and archaeal kingdoms with up to 5% of the total genome of some bacteria devoted to encoding components of RM systems. Because RM systems distinguish between self DNA and nonself DNA and protect bacteria from foreign DNA infection, they can be compared in terms of biological function to the immune system of higher eukaryotes. Like the immune system, RM systems are also involved in a molecular arms race between the invader and the host: bacteriophages have evolved many antirestriction systems such as antirestriction factors that directly inhibit the ENase or multispecific MTases that modify the phage genome at many sites and thereby provide a broad protection against the endonucleases of different RM systems.

Historically, RM systems and, in particular, ENases have been important enzymes for both the basic and applied biomolecular

sciences. The discovery of the molecular basis of RM systems was a hallmark in the molecular understanding of genetic mechanisms that was recognized with a Nobel Prize in 1978 to W. Arber, H. Smith, and D. Nathans. ENases have served as excellent model systems to study the site-directed interactions between enzymes and DNA. Many principles that later could be generalized to other systems were first recognized as a result of studies of ENases. ENases were the first tool available to manipulate DNA in a directed manner and were key players in the recombinant DNA revolution that occurred in molecular biology in the 1980s and 1990s. In addition, they enabled the first analyses of the genomes of bacteria and eukaryotes. Later, ENases made possible the first insights into the relation of DNA sequence diversity and human disease, because as a result of restriction length polymorphism studies, the cleavage patterns of DNA of different individuals could be compared and the differences correlated with pathogenic phenotypes.

2 Types of Restriction-modification Systems

2.1 Type II Systems

Three major types of RM systems, called type I, II, or III, are distinguished on the basis of their structural and functional properties. In type II systems (e.g., *Eco*RI, recognition sequence: GAATTC), the nuclease and methyltransferase activities reside in discrete enzymes, whereas they are combined in multienzyme complexes in type I and type III systems. Type II systems usually have a homodimeric ENase that recognizes a palindromic DNA recognition sequence. These homodimers carry an active center on each subunit that together catalyze DNA cleavage in both strands of the DNA. The type II methyltransferases are either monomeric or dimeric and do not interact with the restriction enzyme. The ENase and MTase bind to the DNA independently and their amino acid DNA recognition modules are different from one another. Type II systems are subdivided into many classes on the basis of functional differences, for example, whether the ENases form tetramers or whether the DNA is cut within or at some distance from the recognition site. Some type II enzymes require the interaction with two recognition sites; some will cleave both, others just one of the sites.

One of the largest subgroups of type II ENases is that of the type IIS enzymes, which interact with nonpalindromic recognition sites and cleave the DNA at a defined distance from the recognition site. The enzymes consist of two domains, one DNA recognition domain that binds to the DNA as a monomer and identifies the target sequence, and a cleavage domain, that dimerizes with a second monomer on the DNA and the dimer catalyzes the cleavage of the DNA in both strands. This dimerization of the cleavage domains is facilitated if two recognition sequences are present on the DNA. Therefore, many type IIS enzymes show an increased activity on substrates containing more than one recognition sequences. Studies with type IIS enzymes suggest that only small changes in the amino acid sequence or the acquisition or deletion of a domain can be responsible for significant changes in the specific biochemical properties of the enzyme. Consequently, many such small changes could create a continuous spectrum of ENases, many of which cannot

easily be assigned solely to one of the presently defined group subtypes.

2.2 Type I and III Systems

Type I and III systems form large supermolecular complexes containing the restriction (R) and modification (M) function. In type I systems (e.g. *Eco*B, recognition sequence: TGAN$_8$TGTC), DNA recognition for methylation and for cleavage is mediated by one S (specificity) subunit. Complexes of M_2S stoichiometry are active MTases, while M_2R_2S complexes have methyltransferase and endonuclease activities. In type III systems (e.g., *Eco*P15, recognition sequence: CTGAAG), RM complexes are active as both methyltransferase and endonuclease. The choice of reaction actually catalyzed by the multimeric enzyme depends on the methylation state of the target sequence: unmodified sites are cleaved; hemimethylated sites that are formed typically during DNA replication are methylated. Systems of both types I and III need interaction with two recognition sequences, which are asymmetric in either case. After initial binding of one enzyme complex to an unmethylated target site, the enzyme remains bound to that site and pulls the DNA using an ATP-dependent motor domain to form a characteristic looped structure. Upon collision of two enzyme molecules that are translocating along the DNA in opposite directions, cleavage is initiated. The cleavage occurs at random locations between the two target sites in the case of type I systems. Type III enzymes cleave the DNA at defined distances of up to 20 bp outside the recognition sequence. As a consequence of their different architectures, type I systems contain two motordomains and translocate along the DNA from both sides of the target sequence, whereas type III systems translocate away from one side only. Therefore, type III systems can cleave DNA only if it contains two target-site sequences in opposite orientation so that the translocating enzymes will collide. DNA cleavage by type I and III systems is stimulated or dependent on the presence of AdoMet, which is the methyl donating cofactor of the MTase subunit.

3 Mechanisms of Action of ENases and MTases

3.1 Common Features

In the following sections, we shall briefly describe some mechanistic features of the type II ENases and DNA MTases that explain how these fascinating enzymes carry out their respective biological functions. Since both classes of enzymes perform their catalytic function on DNA in a sequence-specific manner, common mechanistic elements are found in both. These include the mechanisms of DNA recognition and target-site location. The general catalytic cycle is similar in both cases (Fig. 2). The enzymes bind initially to the DNA in a nonspecific reaction, driven by electrostatic attraction between basic amino acid residues in the DNA-binding site of the enzyme and the anionic DNA. The enzymes then scan the DNA for the presence of the specific site by a process called *facilitated diffusion* in which they stay in close spatial contact with the DNA. After having identified a specific site, the formation of several base-specific contacts usually triggers characteristic conformational changes of the enzyme and the DNA. This mutually induced fit activates

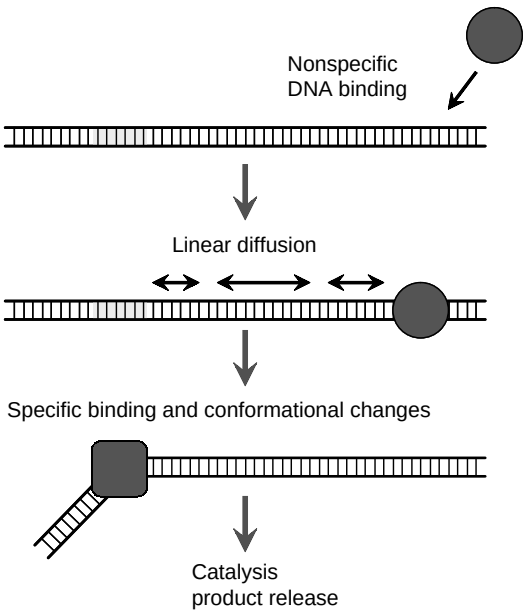

Fig. 2 Catalytic cycle of restriction endonucleases and DNA methyltransferases. The yellow box indicates the target site (see color plate p. xxxi).

the catalytic center leading to the sequence-specific DNA cleavage or the methylation reaction. After DNA cleavage or methylation, release of the modified DNA is often the rate-limiting step of the reaction under multiple-turnover conditions.

4 Catalytic Mechanism of ENases

4.1 Structures of ENases

Most structural information is available for homodimeric type II ENases (see Fig. 3 for an example). The comparison of the structures of protein–DNA complexes of several ENases revealed significant differences in their three-dimensional structures. However, most contain some conserved structural elements in the core of the enzymes that are characterized by a mixed five-stranded ß-sheet flanked by α-helices. The active site amino acid residues are located on this structural core at equivalent positions in many structures. The active site of many restriction enzymes is characterized by a weakly conserved PD. . .(D/E)XK amino acid sequence motif. The PD part of the active site is located in a turn; the remaining residues lie on the edge of a ß-strand. The three-dimensional structure of the conserved element brings together the two conserved acidic residues of the motif such that they can form the binding site for the catalytic magnesium ion that is an essential cofactor for these enzymes. The lysine in the active site motif contributes to the catalytic mechanism as well (see below). Interestingly, other nucleases including λ-exonuclease, MutH and Vsr have a similar fold constituting their active site.

Most of the type II ENases encircle, to a greater or lesser extent, the DNA in their specific DNA complex, allowing them to form an extensive set of interactions both to the major and minor grove of the DNA

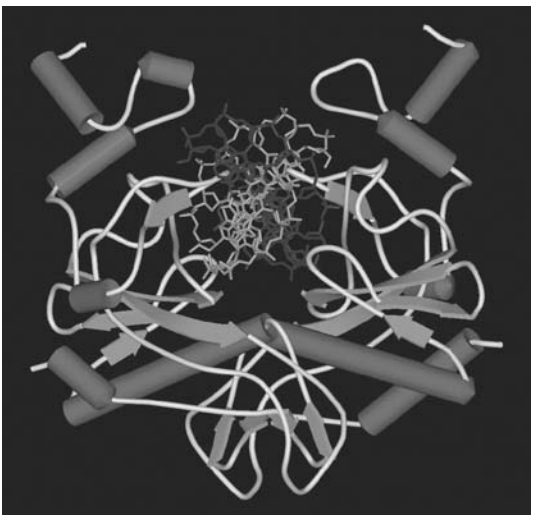

Fig. 3 Crystal structure of the *Eco*RV restriction enzyme in complex with specific DNA. The homodimeric protein is shown in a schematic view, colored according to the secondary structure, with α-helices in red, β-sheets in blue, and loops in gray. The DNA strands, which are viewed perpendicular to the helix axis, are colored blue and orange (see color plate p. xxxii).

as well as to the phosphodiester backbone (Fig. 3). On the basis of the orientation and structure of the conserved secondary structure elements, ENases can be subdivided into two large groups, represented by the first two enzymes whose structures were solved: *Eco*RI and *Eco*RV. With structures becoming available for more and more ENases, some turn out to be closely related, for example, the *Eco*RI and *Mun*I enzymes that recognize GAATTC and CAATTG, respectively.

4.2 Mechanism of Target-site Location

Since restriction endonucleases and DNA methyltransferases are present at the same time in the cell, they compete kinetically for their target site. Destruction of the incoming bacteriophage DNA dictates that the restriction enzyme finds its target site faster than the methyltransferase. Localization of specific recognition sites in the sea of nonspecific sites is a challenging process for an ENase (A six-base-pair-recognizing enzyme, for example, has just one target site every $4^6 = 4096$ sites in a DNA of 50% A + T composition.). Target sites can be discriminated from nontarget sites only during formation of the sequence-specific contacts between enzyme and DNA. However, this is not possible during the initial phase of complex formation that is mainly driven by ionic interactions. Consequently, most binding events between ENases and DNA will result in binding to a nontarget site, that is, to nonspecific DNA.

After the initial binding event, the enzyme searches for its target site by a one-dimensional movement on the DNA, which accelerates target site location by two mechanisms. First, it reduces the search space from three dimensions to one, and second, it abrogates the necessity for the enzyme to dissociate from the DNA after each binding event, which is particularly important, because DNA release is a slow step in protein–DNA interactions. Restriction enzymes were among the earliest examples in which facilitated diffusion was shown to be effective in target-site location. During facilitated diffusion, the

enzymes accurately scan the DNA sequence. The actual mechanism of this process includes a sliding movement of the enzyme on the DNA and a "hopping" process composed of repeated micro dissociation/reassociation reactions. The contribution of both processes varies from enzyme to enzyme and depends on the average distance traveled on the DNA. Scanning of short distances is most efficiently performed by a sliding movement, whereas at longer distances hopping becomes effective.

In addition, tetrameric enzymes can make use of intersegmental transfer, a process during which two binding sites on the protein interact simultaneously with two different parts of the DNA. One binding site can release the DNA and the other remains bound to another site without causing complete dissociation of the enzyme from the DNA. Thereby, the enzyme can make efficient long-range "jumps" on the DNA, which in combination with short-range scanning in the vicinity of each site, produces an efficient target-site searching strategy. This mechanism may explain why some ENases function as tetramers.

4.3 Mechanism of DNA Recognition

DNA recognition by ENases has been investigated by examination of the structures of ENases in complexes with their specific DNA. In most cases, the base pairs of the target site are contacted by several specific interactions, including hydrogen bonds to the edges of the bases in the major and minor grove (Fig. 4) and hydrophobic interactions to the bases and the methyl group in thymidine. In addition, the phosphodiester backbone is contacted by numerous hydrogen and ionic bonds. Both kinds of ionic bonds to the bases and to the backbone are sometimes mediated by interstitial water molecules that bridge the amino acid residue and the DNA. In many, but not all, cases, specific complex formation is accompanied by more-or-less dramatic changes of the DNA structure, which may involve bending, unstacking, or unwinding. Thereby, in a mutually induced fitting of enzyme to DNA, a tight complex of complementary surfaces is formed. Specific DNA binding is accompanied by the release of bound water molecules and counterions

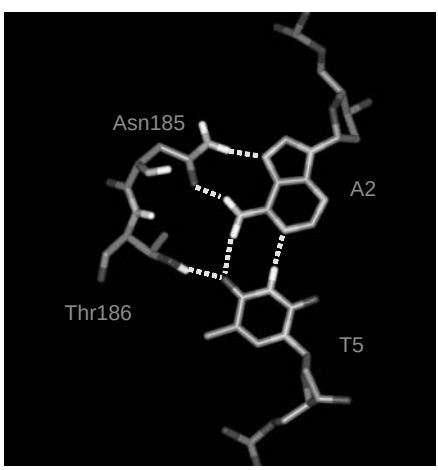

Fig. 4 Example of the specific contacts of the *Eco*RV restriction enzyme to its DNA. Asn185 and Thr186 contact the A2 and T5 base pair in the GATATC/GATATC recognition sequence by specific hydrogen contacts.

from the ENase and the DNA. The methylation of the target base by the partner MTase interferes sterically with this close approximation of the ENase and the DNA, explaining the energetic basis for the lack of activity of ENases at methylated target sites.

The numerous interactions (direct contacts to bases and to backbone and water-mediated contacts to both) in the network of bonding that occurs between ENases and specific DNA guarantee that each base pair is contacted by several specific noncovalent bonds and provides the structural basis of the exquisite sequence specificity of ENases that ensures they do not kill their host cells through nonspecific cleavage. Typically, these enzymes cleave any sequence differing by one base pair from the canonical site at rates that are orders of magnitude reduced; sites differing at more than one base pair are refractory to cleavage. The mechanistic basis for this extreme sequence specificity is that the scissile phosphodiester bond enters the active site only after the conformational change of DNA and enzyme that occurs in the specific complex. Therefore, DNA cleavage can occur only after binding of the enzyme to a specific site on the DNA.

The evolutionary history of this high degree of sequence specificity of ENases can be understood, because cellular DNA is protected against cleavage by the MTase only at the specific sites. Therefore, nonspecific double-strand cleavage of the bacterial DNA by ENases is toxic and has been strongly counterselected during evolution. In addition, ENases usually introduce only a single-strand break at nontarget sites, which can be easily repaired by DNA ligase, in contrast to their double-strand cut at target sites. Because of their very high degree of specificity, ENases have been refractory to all attempts to change their specificity by protein design. In particular, unsuccessful trials to design new contacts between ENases and DNA by site-directed mutagenesis have demonstrated that enzyme variants that carry amino acid exchanges at critical positions usually are catalytically inactive. This property ensures that erroneous proteins that are formed in the cell after rare misincorporations of amino acid residues during protein biosynthesis will not act as nonspecific nucleases or ENases with altered recognition specificity and kill the cell.

It is interesting to note that the sequence-specific contacts are formed by different segments of the protein in the various structural families of ENases. For example, in *Eco*RI a structure similar to a ß-strand is located in the major groove of the DNA and in *Eco*RV two loops of protein mediate sequence recognition. On the other hand, families of enzymes exist that use similar structural approaches to recognize related sites, for example, the *Eco*RI/*Mun*I pair or in the large subfamily of ENases containing *Ngo*MIV (GCCGGC) that have a GG dinucleotide in their target sequence and use a conserved RXXR motif for its recognition. In contrast, the interaction of *Bam*HI (GGATCC) and *Bgl*II (AGATCT) with DNA is different, although both enzymes recognize the same GATC central tetranucleotide.

4.4 Mechanism of DNA Cleavage

Despite being studied for decades, the detailed catalytic mechanism of ENases remains incompletely resolved. One obvious source of uncertainty is that different crystal structures revealed one, two, or

three divalent cations bound at or near their active sites. These observations led to models requiring one or two metal ions for catalytic function. Furthermore, the relationships of the metal ions observed in various structures to the activity of the enzymes in solution is not clear, since most crystallizations were carried out at much higher concentrations of the divalent metal ions than is commonly used for cleavage in solution.

DNA cleavage by ENases always produces a 5′ phosphate and a 3′ OH. It is initiated by the in-line attack of a water molecule on the scissile phosphodiester group (Fig. 5). Chemically, this nucleophilic attack is difficult, since the phosphodiester group carries negative charge and efficient neutralization of this charge is an important catalytic requirement. All ENases appear to recruit a Mg^{2+} ion for that purpose along with basic residues that contribute to charge neutralization – most frequently the lysine often found in the PD...(D/E)XK catalytic motif whose aspartic or glutamic acid residues function in positioning of the metal cation. The second important function of the enzyme is to accelerate the hydrolytic reaction by stabilization of the leaving 3′-O^- alcoholate group by protonation. In most models, this role is attributed to a water molecule bound in the ligand sphere of the Mg^{2+} ion, which explains the absolute requirement for this metal ion that in most cases can be replaced by Mn^{2+}, sometimes also by Co^{2+}, but usually not by other ions.

Whereas this consensus view on the mechanism is widely accepted, other details are unclear. For example, an S_N2-like in-line attack would also require activation of the attacking water. This function has been attributed to another metal ion, the neighboring phosphodiester group, lysine, arginine, or aspartic acid residues in the different mechanistic models. However, for the hydrolysis of a phosphodiester bond, a more S_N1-like dissociative mechanism could also hold in which the bond to the leaving group is almost completely broken before the attacking nucleophile comes in. In this mechanism, a highly reactive metaphosphate intermediate that readily reacts with water is formed. Therefore, activation of the attacking water might not be necessary, if a dissociative mechanism functions. In this model, the proton leaving the attacking water during the reaction is transferred to bulk solvent. To yield the stereochemistry observed upon ENase cleavage, the metaphosphate that would be produced during a dissociative mechanism would

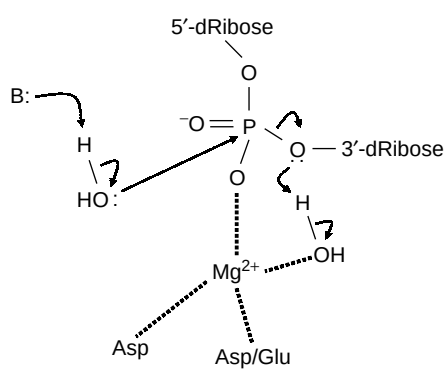

Fig. 5 Model of the catalytic mechanism of restriction endonucleases. The catalytic metal ion is bound by two acidic amino acid residues located in the PD...D/EXK motif. The nature of the base responsible for the deprotonation of the attacking nucleophile is not known and may vary between different ENases.

have to be positioned by the enzyme in a stereochemically defined manner such that an inversion of configuration always occurred.

5 Catalytic Mechanism of DNA MTases

5.1 Mechanism of DNA Methylation

All DNA MTases use the coenzyme S-adenosyl-L-methionine (AdoMet) as the donor of an activated methyl group whose activated methylsulfonium bond can be easily attacked by a nucleophile to achieve the methylation. However, direct methylation of either the exocyclic amino groups of adenine and cytosine or of the C5-position of cytosine is a difficult chemical task because the N1 and N3 positions of adenine and cytosine, respectively, are the favored sites of nucleophilic attack and the C5 of cytosine is intrinsically resistant to nucleophilic attack. For adenine and cytosine in duplex DNA, only the N3 position of adenine is readily alkylated by nucleophilic reagents. To overcome these intrinsic lacks of reactivity, the two families of DNA MTases, N-Mtases, and C-MTases, have evolved highly sophisticated reaction mechanisms. Both mechanisms have in common the need for close contact between the enzyme active site and the target base. Such proximity is not possible with the target base buried within the DNA double helix. In a seminal structural study, X. Cheng, R. Roberts, S. Klimasauskas, and coworkers demonstrated that the cytosine-C5 MTase HhaI completely flips its target base out of the DNA double helix during catalysis – an unprecedented, dramatic conformational change in the DNA that solves the accessibility problem (Fig. 6).

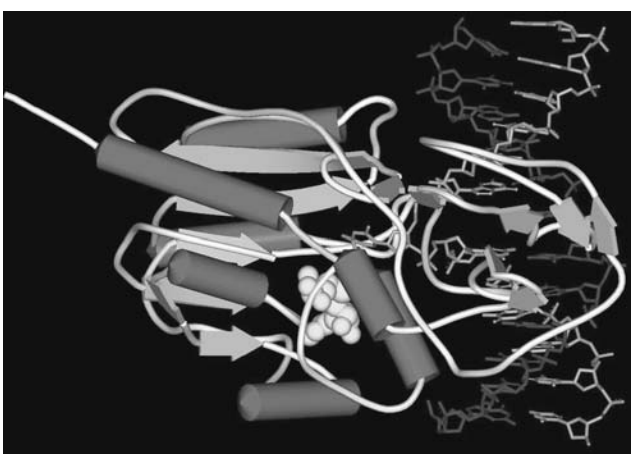

Fig. 6 Structure of the HhaI DNA-(cytosine-C5)-MTase in complex with specific DNA. The flipped target cytosine is readily visible in this view of the enzyme. The protein is shown in a schematic view, colored according to the secondary structure, the DNA strands are colored blue and orange and the AdoMet coenzyme is shown in yellow (see color plate p. xxxii).

5.2 Reaction Mechanism of C5-MTases

Methylation of the C5-position of cytosine follows a two-step reaction mechanism first proposed by Daniel Santi (Fig. 7). In the first step, a cysteine residue in the active site of the DNA MTase performs a nucleophilic attack on the C6-position of the cytosine that is flipped out of the DNA and bound in a hydrophobic pocket of the enzyme. Thereby, a transient carbanion is generated, which is stabilized by a conserved glutamic acid residue that hydrogen bonds to the flipped base and protonates the N3 atom of the base. Since the carbanion has a high negative charge density at position 5, the activated base can attack the methyl group of the cofactor yielding a stable intermediate comprising a covalently linked complex of the methylated cytosine and the MTase. The catalytic cysteine and glutamic acid residues are located in two highly conserved amino acid motifs found in Cytosine-C5 MTases, the PCQ motif and the ENV motif also called *motif IV* and *motif VI*, respectively. The second step of the reaction resolves the covalent intermediate. It is

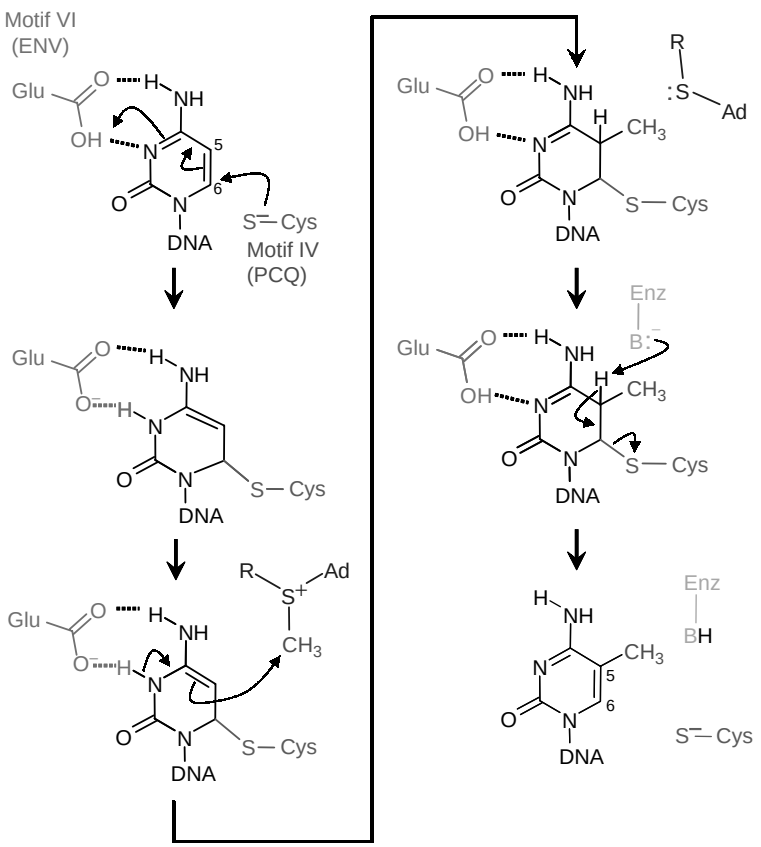

Fig. 7 Catalytic mechanism of DNA-(cytosine-C5)-MTases. (See color plate p. xxxv).

initiated by deprotonation of C5 catalyzed by an unknown proton acceptor, leading to the elimination of the cysteine residues and re-establishment of aromaticity.

The mechanism is supported by several lines of evidence. The covalent complex has been detected in crystal structures and in biochemical studies, the catalytic cysteine and glutamic acid residues are highly conserved in all cytosine-C5 Mtases, and mutational analyses have supported their involvement in catalysis. In addition to these elements of covalent and acid-base catalysis, MTases facilitate the methylation reaction by proper positioning of the substrate and the cofactor in the active site and by tight binding to the transition state of the reaction.

5.3 Mechanism of N-MTases

Methylation of either adenine and cytosine residues at their exocyclic amino group, in contrast to their N1 or N3 positions, respectively, presents a difficult chemical problem because in each case the substrate amino group is part of an electron poor, heterocyclic aromatic system. The free electron pair formally present at the exocyclic amino groups is delocalized into the aromatic system and is not as readily available for nucleophilic attack on the AdoMet as are the heterocyclic ring nitrogens. The catalytic center of the adenine-N6 and cytosine-N4 MTases (N-MTase) is characterized by a (D/N/S)PP(Y/F) motif. Involvement of this amino acid sequence element in catalysis was demonstrated in several site-directed mutagenesis studies carried out with N-MTases. The reaction mechanism of the N-MTases is illustrated by the structure of the TaqI adenine-N6 MTase. The side chain of the active site asparagine and the backbone carbonyl group of the proline residue next to the asparagine of the active site tetrapeptide (NPPY in M.TaqI) contact the exocyclic amino group of the flipped adenine target (Fig. 8). Since the hydrogen-bond acceptors are presented in a tetrahedral geometry, the hybridization

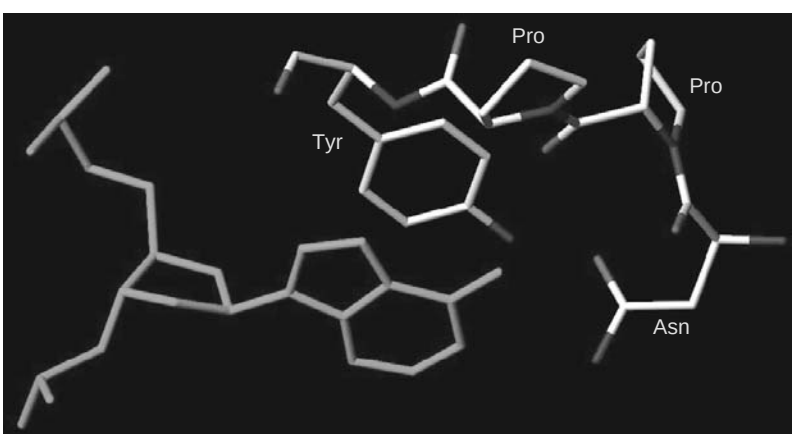

Fig. 8 Crystal structure of the *Taq*I adenine-N6 MTase. The flipped base (shown in green) is stacked on the tyrosine residue of the conserved (D/N/S)PP(Y/F) motif. The exocyclic amino group is contacted by two hydrogen-bond donors (the Asn and Pro residue in the active site motif) (see color plate p. xxviii).

of the amino group changes from sp^2 to sp^3, which leads to the localization of the free electron pair at the nitrogen atom. The flipped base is stabilized by a π-stacking interaction to the aromatic amino acid residue in the (NPPY) motif. Proper positioning of the DNA-flipped base and the AdoMet is critical for catalysis in N-MTases as it is in C5-MTases.

Consistent with the similarity of their active site motifs and the chemistry of the methylation reaction, the amino MTases do not show absolute specificity for the flipped base that they methylate. A cytosine positioned in the DNA recognition sequence of some adenine-N6-MTases can be methylated at the N4 position. Analogously, an appropriately placed adenine can sometimes be methylated at the N6 position by a cytosine-N4-MTase. These reactions occur with reduced efficiencies, but they indicate that the catalytic mechanisms of adenine-N6 and cytosine-N4 MTases are similar. In addition, the HemK enzymes, found across the phylogenetic tree of N-MTases, contain this motif and methylate glutamine at the N5 position, which also contains an amino group conjugated to an electron poor π-system. These findings suggest that the (D/N/S)PP(Y/F) sequence is a versatile motif for the methylation of different target molecules. One can envisage that other enzymes containing this motif might have even more, as yet undescribed, substrates, for example, arginine.

5.4 Structures of MTases

All DNA MTases comprise two domains, one large and one small. The large domain consists of a central mixed ß-sheet, flanked by α-helices. It is folded into two subdomains, each of which forms a hydrophobic pocket. One of these pockets is used as the coenzyme-binding site and the other as the binding site for the flipped target base. The AdoMet binding site is characterized by conserved amino acid motifs that are found in most AdoMet-interacting enzymes, with conserved motif I [(D/E/S)XFXGXG] contributing important contacts to the AdoMet. Cytosine-C5-MTases and N-MTases show a striking similarity in the structure of their catalytic domains, with important catalytic residues [motif IV: PCQ in the case of cytosine-C5 MTases and (D/N/S)PP(Y/F) in the case of N-MTases] occupying equivalent positions in the 3D-structures of both MTase families. The second, smaller domain of DNA MTases is more heterogeneous in size and structure. It contains most of the residues that mediate sequence-specific recognition of the DNA. Therefore, the variability of the target sequence is reflected by the loss of sequence conservation of the small DNA MTase domains.

5.5 Mechanism of DNA Recognition

DNA recognition by MTases, as with ENases, arises from sequence-specific contacts of the enzymes in the major and minor grooves of the DNA as well as to its phosphodiester backbone. The small domain of the MTases contacts the DNA from the major groove side, and the body of the large domain interacts with the minor groove. In addition, loops of the catalytic domain contribute to sequence recognition in some enzymes. Depending on the particular enzyme, specific complex formation sometimes, but not always, leads to strong bending of the DNA. Unfortunately, structural information is available presently only for enzymes that do not induce bending of the DNA.

Although only a few have been studied, the sequence specificity of different DNA MTases appears to vary dramatically. Some MTases display a high specificity that may be comparable to that of restriction enzymes, whereas others interact with the DNA in a much more relaxed manner such that sites differing from the consensus target site are modified at reasonable rates. It should be noted that inaccurate MTases are much less toxic to the cell than are inaccurate ENases. Therefore, the major evolutionary pressure on MTases is to maintain an effective modification of the bacterial DNA at the ENase recognition site. Occasional methylation of nontarget sites usually does not have major, immediate effects. As a consequence, RM systems are steadily evolving and are regularly changing or modifying their recognition sites by broadening the specificity of their MTases and occasionally acquiring by horizontal gene transfer a new ENase whose specificity matches the broadened specificity of the MTase. Most likely, MTases with relaxed specificity have been caught in the act of molecular evolution and have not yet optimized selectivity for their new target site.

5.6 Mechanism of Target-site Location and Lack of Processivity of Methylation

Like the restriction enzymes, the DNA MTases of RM systems use facilitated diffusion to accelerate the search for target sites. In this mechanism, the enzyme first binds to the DNA at a nonspecific site in a very fast reaction that often is close to the theoretical limit for a bimolecular association process between macromolecules. Then, the enzyme follows the DNA in a one-dimensional random movement.

One interesting aspect in the target-acquisition mechanism of the DNA MTases is the way in which several target sites may be modified when they reside on the same DNA molecule. Enzymes can react with substrates containing several target sites in two principal ways – processively or distributively. In a processive reaction, after one turnover, the enzyme remains bound to, or closely associated with, the same substrate molecule, moves to the next target site, and modifies it a second time, and so on. Therefore, in a completely processive reaction, a given DNA molecule is modified at all its target sites before the enzyme dissociates completely from the DNA. In contrast, in a distributive reaction the enzyme releases the DNA after each turnover and then rebinds to the same DNA or to a different DNA molecule. As a consequence of these two different mechanisms, the distribution of substrates, methylation intermediates, and products in a reaction mixture are completely different: methylation intermediates are not observed in a processive multiple-turnover reaction, whereas intermediates are necessarily formed in a distributive reaction. Recently, experiments with different MTases, demonstrate that the solitary M.SssI, the CcrM MTase, *E. coli* dam, and the mammalian Dnmt1 enzymes methylate DNA in a highly processive reaction. In contrast, the RM enzymes M.*Hha*I, M.*Hpa*I, *Eco*RI, and *Eco*RV react distributively. This difference is particularly striking in the case of the *Eco*RV-*E. coli* dam pair, because these enzymes share high sequence similarity and show overlapping target-site specificity, GATATC and GATC, respectively, where the common sequence is underlined.

The processive and distributive mechanism reflects the biological roles played by the respective enzymes. RM systems

protect cells from bacteriophage infections by endonuclease cleavage of the invading phage DNA. The RM MTase prevents cleavage of the host genome by methylation. However, if the MTase modified the incoming DNA too rapidly, the phage DNA would be protected and lethal. The distributive reaction mechanism of the RM MTases helps prevent complete methylation of any one molecule thereby leaving it vulnerable to the restriction enzyme, since cleavage of one site only is usually sufficient for phage inactivation. In agreement with this model, all RM MTases examined show a distributive reaction. Mechanisms such as compartmentalization or tight control of the expression levels of both types of enzymes could also ensure proper RM MTase function. Because rapid methylation of the chromosomal DNA after replication is desirable, the solitary MTases would not be selected against by evolution.

The difference in the order of DNA and AdoMet binding provides an explanation for the difference in processivity between *Eco*RV and *E. coli* dam. *Eco*RV first binds to the cofactor and then to the DNA, whereas *E. coli* dam prefers the opposite binding order. Microscopic reversibility dictates that the order of DNA and cofactor release must reflect their order of binding. *Eco*RV will release the DNA first and the cofactor second. The obligate release of DNA first precludes *Eco*RV from modifying DNA in a processive fashion, because the enzyme cannot exchange the product AdoHcy with AdoMet while bound to the DNA. The exchange of the cofactor in the ternary complex of M.*Eco*RV may be prevented by slight conformational changes that could block a channel needed for cofactor exchange. This effect of bound DNA on the cofactor-exchange order illustrates how seemingly subtle effects on the structure and mechanism of an enzyme can have profound effects on the function, for example, in this case, ensuring a distributive reaction rather than a processive one.

See also Protein and Nucleic Acid Enzymes; Serial Analysis of Gene Expression.

Bibliography

Books and Reviews

Cheng, X. (1995) Structure and function of DNA methyltransferases, *Annu. Rev. Biophys. Biomol. Struct.* **24**, 293–318.

Cheng, X. (1995) DNA modification by methyltransferases, *Curr. Opin. Struct. Biol.* **5**, 4–10.

Cheng, X., Blumenthal, R.M. (Eds.) (1999) *S-Adenosylmethionine-Dependent Methyltransferases: Structures and Functions*, World Scientific Publishing, Singapore.

Cheng, X., Roberts, R.J. (2001) AdoMet-dependent methylation, DNA methyltransferases and base-flipping., *Nucleic Acids Res.* **29**, 3784–3795.

Jeltsch, A. (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases, *Chem. BioChem.* **3**, 274–293.

Jeltsch, A., Gumport, R.I. (2004) DNA Methyltransferases, Bacterial, in: Lennarz, W.J., Lane, M.D. (Eds.) *Encyclopedia of Biological Chemistry*, Vol. 1, Elsevier, Oxford, pp. 644–651.

Lanio, T., Jeltsch, A., Pingoud, A. (2002) Evolutionary Generation Versus Rational Drug Design of Restriction Endonucleases with Novel Specificity, in: Brakmann, S., Johnsson, K. (Eds.) *Directed Molecular Evolution of Proteins: or How to Improve Enzymes for Biocatalysis*, Wiley-VCH, Weinheim, Germany, pp. 309–328.

Low, D.A., Weyand, N.J., Mahan, M.J. (2001) Roles of DNA adenine methylation in regulating bacterial gene expression and virulence, *Infect. Immun.* **69**, 7197–7204.

Pingoud, A. (Ed.) (2004) *Restriction Endonucleases*, Springer-Verlag, New York.

Pingoud, A., Jeltsch, A. (1997) Recognition and cleavage of DNA by type II restriction endonucleases, *Eur. J. Biochem.* **246**, 1–22.

Pingoud, A., Jeltsch, A. (2001) Structure and function of type II restriction endonucleases, *Nucleic Acids Res.* **29**, 3705–3727.

Reisenauer, A., Kahng, L.S., McCollum, S., Shapiro, L. (1999) Bacterial DNA methylation: a cell cycle regulator? *J. Bacteriol.* **181**, 5135–5139.

Roberts, R.J., Cheng, X. (1998) Base flipping, *Annu. Rev. Biochem.* **67**, 181–198.

Roberts, R.J. et al. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes, *Nucl. Acids Res.* **31**, 1805–1812.

Primary Literature

Aiken, C.R., Fisher, E.W., Gumport, R.I. (1991) The specific binding, bending, and unwinding of DNA by RsrI endonuclease, an isoschizomer of EcoRI endonuclease, *J. Biol. Chem.* **266**, 19063–19069.

Aiken, C.R., McLaughlin, L.W., Gumport, R.I. (1991) The highly homologous isoschizomers RsrI endonuclease and EcoRI endonuclease do not recognize their target sequence identically, *J. Biol. Chem.* **266**, 19070–19078.

Allan, B.W., Beechem, J.M., Lindstrom, W.M., Reich, N.O. (1998) Direct real time observation of base flipping by the EcoRI DNA methyltransferase, *J. Biol. Chem.* **273**, 2368–2373.

Allan, B.W., Reich, N.O. (1996) Targeted base stacking disruption by the EcoRI DNA methyltransferase, *Biochemistry* **35**, 14757–14762.

Allet, B., Jeppesen, P.G.N., Katagiri, K.J., Delius, H. (1973) Mapping the DNA fragments produced by cleavage of lambda DNA with endonuclease RI, *Nature* **241**, 120–123.

Alves, J., Ruter, T., Geiger, R., Fliess, A., Maass, G., Pingoud, A. (1989) Changing the hydrogen-bonding potential in the DNA binding site of EcoRI by site-directed mutagenesis drastically reduces the enzymatic activity, not the preference of this restriction endonuclease for cleavage within the site GAATTC, *Biochemistry* **28**, 2678–2684.

Alves, J., Urbanke, C., Fliess, A., Maass, G., Pingoud, A. (1989) Fluorescence stopped-flow kinetics of the cleavage of synthetic oligodeoxynucleotides by the EcoRI restriction endonuclease, *Biochemistry* **28**, 7879–7888.

Arber, W. (1965) Host specificity of DNA produced by Escherichia coli V. The role of methionine in the production of host specificity, *J. Mol. Biol.* **11**, 247–256.

Arber, W., Dussoix, D. (1962) Host specificity of DNA produced by Escherichia Coli: I. Host controlled modification of bacteriophage lambda, *J. Mol. Biol.* **5**, 18–36.

Athanasiadis, A., Vlassi, M., Kotsifaki, D., Tucker, P.A., Wilson, K.S., Kokkinidis, M. (1994) Crystal structure of PvuII endonuclease reveals extensive structural homologies to EcoRV, *Nat. Struct. Biol.* **1**, 469–475.

Baldwin, G.S., Vipond, I.B., Halford, S.E. (1995) Rapid reaction analysis of the catalytic cycle of the EcoRV restriction endonuclease, *Biochemistry* **34**, 705–714.

Beck, C., Cranz, S., Solmaz, M., Roth, M., Jeltsch, A. (2001) How does a DNA interacting enzyme change its specificity during molecular evolution? A site-directed mutagenesis study at the DNA binding site of the DNA-(adenine-N(6))-methyltransferase EcoRV, *Biochemistry* **40**, 10956–10965.

Beck, C., Jeltsch, A. (2002) Probing the DNA interface of the EcoRV DNA-(adenine-N6)-methyltransferase by site-directed mutagenesis, fluorescence spectroscopy, and UV cross-linking, *Biochemistry* **41**, 14103–14110.

Bertani, G., Weigle, J.J. (1953) Host controlled variation in bacterial viruses, *J. Bacteriol.* **65**, 113–121.

Bheemanaik, S., Chandrashekaran, S., Nagaraja, V., Rao, D.N. (2003) Kinetic and catalytic properties of dimeric KpnI DNA methyltransferase, *J. Biol. Chem.* **278**, 7863–7874.

Cal, S., Connolly, B.A. (1997) DNA distortion and base flipping by the EcoRV DNA methyltransferase, *J. Biol. Chem.* **272**, 490–496.

Cheng, X., Balendiran, K., Schildkraut, I., Anderson, J.E. (1994) Structure of PvuII endonuclease with cognate DNA, *EMBO J.* **13**, 3927–3935.

Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W., Roberts, R.J. (1993) Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine, *Cell* **74**, 299–307.

Colandene, J.D., Topal, M.D. (1998) The domain organization of NaeI endonuclease: separation of binding and catalysis, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3531–3536.

Deibert, M., Grazulis, S., Sasnauskas, G., Siksnys, V., Huber, R. (2000) Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA, *Nat. Struct. Biol.* **7**, 792–799.

Dussoix, D., Arber, W. (1962) Host specificity of DNA produced by Escherichia Coli: II. Control over acceptance of DNA from infecting phage lambda, *J. Mol. Biol.* **5**, 37–49.

Ellis, D.J., Dryden, D.T., Berge, T., Edwardson, J.M., Henderson, R.M. (1999) Direct observation of DNA translocation and cleavage by the EcoKI endonuclease using atomic force microscopy, *Nat. Struct. Biol.* **6**, 15–17.

Embleton, M.L., Vologodskii, A.V., Halford, S.E. (2004) Dynamics of DNA loop capture by the SfiI restriction endonuclease on supercoiled and relaxed DNA, *J. Mol. Biol.* **339**, 53–66.

Firman, K., Szczelkun, M.D. (2000) Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement, *EMBO J.* **19**, 2094–2102.

Ghosh, S.S., Eis, P.S., Blumeyer, K., Fearon, K., Millar, D.P. (1994) Real time kinetics of restriction endonuclease cleavage monitored by fluorescence resonance energy transfer, *Nucleic Acids Res.* **22**, 3155–3159.

Goedecke, K., Pignot, M., Goody, R.S., Scheidig, A.J., Weinhold, E. (2001) Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog, *Nat. Struct. Biol.* **8**, 121–125.

Gowers, D.M., Bellamy, S.R., Halford, S.E. (2004) One recognition sequence, seven restriction enzymes, five reaction mechanisms, *Nucleic Acids Res.* **32**, 3469–3479.

Gowers, D.M., Halford, S.E. (2003) Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling, *EMBO J.* **22**, 1410–1418.

Gowher, H., Jeltsch, A. (2000) Molecular enzymology of the EcoRV DNA-(Adenine-N6)-methyltransferase: kinetics of DNA binding and bending, kinetic mechanism and linear diffusion of the enzyme on DNA, *J. Mol. Biol.* **303**, 93–110.

Halford, S.E., Marko, J.F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052.

Halford, S.E., Welsh, A.J., Szczelkun, M.D. (2004) Enzyme-mediated DNA looping, *Annu. Rev. Biophys. Biomol. Struct.* **33**, 1–24.

Hattman, S., Fukasawa, T. (1963) Host-induced modification of T-even phages due to defective glucosylation of their DNA, *Proc. Natl. Acad. Sci. U.S.A.* **50**, 297–300.

Heithoff, D.M., Sinsheimer, R.L., Low, D.A., Mahan, M.J. (1999) An essential role for DNA adenine methylation in bacterial virulence, *Science* **284**, 967–970.

Heitman, J., Model, P. (1990) Mutants of the EcoRI endonuclease with promiscuous substrate specificity implicate residues involved in substrate recognition, *EMBO J.* **9**, 3369–3378.

Holz, B., Dank, N., Eickhoff, J.E., Lipps, G., Krauss, G., Weinhold, E. (1999) Identification of the binding site for the extrahelical target base in N6-adenine DNA methyltransferases by photo-cross-linking with duplex oligodeoxyribonucleotides containing 5-iodouracil at the target position, *J. Biol. Chem.* **274**, 15066–15072.

Holz, B., Klimasauskas, S., Serva, S., Weinhold, E. (1998) 2-Aminopurine as a fluorescent probe for DNA base flipping by methyltransferases, *Nucleic Acids Res.* **26**, 1076–1083.

Hopkins, B.B., Reich, N.O. (2004) Simultaneous DNA binding, bending, and base flipping: evidence for a novel M.EcoRI methyltransferase-DNA complex, *J. Biol. Chem.* **279**, 37049–37060.

Horton, J.R., Ratner, G., Banavali, N.K., Huang, N., Choi, Y., Maier, M.A., Marquez, V.E., MacKerell, A.D., Cheng, X. Jr. (2004) Caught in the act: visualization of an intermediate in the DNA base-flipping pathway induced by HhaI methyltransferase, *Nucleic Acids Res.* **32**, 3877–3886.

Huai, Q., Colandene, J.D., Chen, Y., Luo, F., Zhao, Y., Topal, M.D., Ke, H. (2000) Crystal structure of NaeI – an evolutionary bridge between DNA endonuclease and topoisomerase, *EMBO J.* **19**, 3110–3118.

Huai, Q., Colandene, J.D., Topal, M.D., Ke, H. (2001) Structure of NaeI-DNA complex reveals dual-mode DNA recognition and complete dimer rearrangement, *Nat. Struct. Biol.* **8**, 665–669.

Jeltsch, A., Alves, J., Oelgeschlager, T., Wolfes, H., Maass, G., Pingoud, A. (1993) Mutational analysis of the function of Gln115 in the EcoRI restriction endonuclease, a critical amino acid

for recognition of the inner thymidine residue in the sequence-GAATTC-and for coupling specific DNA binding to catalysis, *J. Mol. Biol.* **229**, 221–234.

Jeltsch, A., Alves, J., Wolfes, H., Maass, G., Pingoud, A. (1993) Substrate-assisted catalysis in the cleavage of DNA by the EcoRI and EcoRV restriction enzymes, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8499–8503.

Jeltsch, A., Alves, J., Wolfes, H., Maass, G., Pingoud, A. (1994) Pausing of the restriction endonuclease EcoRI during linear diffusion on DNA, *Biochemistry* **33**, 10215–10219.

Jeltsch, A., Christ, F., Fatemi, M., Roth, M. (1999) On the substrate specificity of DNA methyltransferases. adenine-N6 DNA methyltransferases also modify cytosine residues at position N4, *J. Biol. Chem.* **274**, 19538–19544.

Jeltsch, A., Pingoud, A. (1996) Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems, *J. Mol. Evol.* **42**, 91–96.

Jeltsch, A., Pingoud, A. (1998) Kinetic characterization of linear diffusion of the restriction endonuclease EcoRV on DNA, *Biochemistry* **37**, 2160–2169.

Jeltsch, A., Roth, M., Friedrich, T. (1999) Mutational analysis of target base flipping by the EcoRV adenine-N6 DNA methyltransferase, *J. Mol. Biol.* **285**, 1121–1130.

Jeltsch, A., Wenz, C., Stahl, F., Pingoud, A. (1996) Linear diffusion of the restriction endonuclease EcoRV on DNA is essential for the in vivo function of the enzyme, *EMBO J.* **15**, 5104–5111.

Jen-Jacobson, L., Lesser, D., Kurpiewski, M. (1986) The enfolding arms of the EcoRI endonuclease: role in DNA binding and cleavage, *Cell* **45**, 619–629.

Jones, P.A., Gonzalgo, M.L. (1997) Altered DNA methylation and genome instability: a new pathway to cancer? *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2103–2105.

Kettling, U., Koltermann, A., Schwille, P., Eigen, M. (1998) Real-time enzyme kinetics monitored by dual-color fluorescence cross-correlation spectroscopy, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1416–1420.

Kim, Y., Grable, J.C., Love, R., Greene, P.J., Rosenberg, J.M. (1990) Refinement of EcoRI endonuclease crystal structure: a revised protein chain tracing, *Science* **249**, 1307–1309.

Klimasauskas, S., Kumar, S., Roberts, R.J., Cheng, X. (1994) Hhal methyltransferase flips its target base out of the DNA helix, *Cell* **76**, 357–369.

Klimasauskas, S., Roberts, R.J. (1995) M.HhaI binds tightly to substrates containing mismatches at the target base, *Nucleic Acids Res.* **23**, 1388–1395.

Klimasauskas, S., Szyperski, T., Serva, S., Wuthrich, K. (1998) Dynamic modes of the flipped-out cytosine during HhaI methyltransferase-DNA interactions in solution, *EMBO J.* **17**, 317–324.

Lacks, S.A., Mannarelli, B.M., Springhorn, S.S., Greenberg, B. (1986) Genetic basis of the complementary dpni and dpnii restriction systems of S. pneumoniae: an intercellular cassette mechanism, *Cell* **46**, 993–1000.

Lagunavicius, A., Sasnauskas, G., Halford, S.E., Siksnys, V. (2003) The metal-independent type IIs restriction enzyme BfiI is a dimer that binds two DNA sites but has only one catalytic centre, *J. Mol. Biol.* **326**, 1051–1064.

Lanio, T., Jeltsch, A., Pingoud, A. (2000) On the possibilities and limitations of rational protein design to expand the specificity of restriction enzymes: a case study employing EcoRV as the target, *Protein Eng.* **13**, 275–281.

Lesser, D.R., Kurpiewski, M.R., Jen-Jacobson, L. (1990) The energetic basis of specificity in the EcoRI endonuclease-DNA interaction, *Science* **250**, 776–786.

Li, L., Wu, L.P., Chandrasegaran, S. (1992) Functional domains in FokI restriction endonuclease, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4275–4279.

Liebert, K., Hermann, A., Schlickenrieder, M., Jeltsch, A. (2004) Stopped-flow and mutational analysis of base flipping by the Escherichia coli Dam DNA-(adenine-N6)-methyltransferase, *J. Mol. Biol.* **341**, 443–454.

Lukacs, C.M., Kucera, R., Schildkraut, I., Aggarwal, A.K. (2000) Understanding the immutability of restriction enzymes: crystal structure of BglII and its DNA substrate at 1.5A resolution, *Nat. Struct. Biol.* **7**, 134–140.

Luria, S.E., Human, M.L. (1952) A nonhereditary, host-induced variation of bacterial viruses, *J. Bacteriol.* **64**, 557–569.

Malone, T., Blumenthal, R.M., Cheng, X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyl-transferases, and suggests a

catalytic mechanism for these enzymes, *J. Mol. Biol.* **253**, 618–632.

Malygin, E.G., Lindstrom, W.M., Zinoviev, V.V., Evdokimov, A.A., Schlagman, S.L., Reich, N.O., Hattman, S. Jr. (2003) Bacteriophage T4Dam (DNA-(adenine-N6)-methyltransferase): evidence for two distinct stages of methylation under single turnover conditions, *J. Biol. Chem.* **278**, 41749–41755.

Malygin, E.G., Lindstrom, W.M.J., Schlagman, S.L., Hattman, S., Reich, N.O. (2000) Pre-steady state kinetics of bacteriophage T4 Dam DNA-[N6-adenine] methyltransferase: interaction with native (GATC) or modified sites, *Nucleic Acids Res.* **28**, 4207–4211.

Malygin, E.G., Sclavi, B., Zinoviev, V.V., Evdokimov, A.A., Hattman, S., Buckle, M. (2004) Bacteriophage T4Dam DNA-(Adenine-N6)-methyltransferase: comparison of pre steady-stase and single turnover methylation of 40mer duplexes containing two (un)modified target sites, *J. Biol. Chem.* **279**, 50012–50018.

Malygin, E.G., Zinoviev, V.V., Evdokimov, A.A., Lindstrom, W.M., Reich, N.O., Hattman, S. Jr. (2003) DNA (cytosine-N4-)-and-(adenine-N6-)-methyltransferases have different kinetic mechanisms but the same reaction route. A comparison of M.BamHI and T4 Dam, *J. Biol. Chem.* **278**, 15713–15719.

Meisel, A., Bickle, T.A., Kruger, D.H., Schroeder, C. (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage, *Nature* **355**, 467–469.

Meisel, A., Mackeldanz, P., Bickle, T.A., Kruger, D.H., Schroeder, C. (1995) Type III restriction endonucleases translocate DNA in a reaction driven by recognition site-specific ATP hydrolysis, *EMBO J.* **14**, 2958–2966.

Meselson, M., Yuan, R. (1968) DNA restriction enzyme from Escherichia coli, *Nature* **217**, 1110–1114.

Modrich, P., Zabel, D. (1976) EcoRI endonuclease. Physical and catalytic properties of the homogeneous enzyme, *J. Biol. Chem.* **251**, 5866–5874.

Mordasini, T., Curioni, A., Andreoni, W. (2003) Why do divalent metal ions either promote or inhibit enzymatic reactions? The case of BamHI restriction endonuclease from combined quantum-classical simulations, *J. Biol. Chem.* **278**, 4381–4384.

Mucke, M., Grelle, G., Behlke, J., Kraft, R., Kruger, D.H., Reuter, M. (2002) EcoRII: a restriction enzyme evolving recombination functions? *EMBO J.* **21**, 5262–5268.

Murray, N.E., Murray, K. (1974) Manipulation of restriction targets in phage lambda to form receptor chromosomes for DNA fragments, *Nature* **251**, 476–481.

Nagaraja, V., Shepherd, J.C.W., Bickle, T.A. (1985) A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity, *Nature* **316**, 371–372.

Newman, M., Lunnen, K., Wilson, G., Greci, J., Schildkraut, I., Phillips, S.E.V. (1998) Crystal structure of restriction endonuclease BglI bound to its interrupted DNA recognition sequence, *EMBO J.* **17**, 5466–5476.

Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I., Aggarwal, A.K. (1994) Structure of restriction endonuclease BamHI and its relationship to EcoRI, *Nature* **368**, 660–664.

Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I., Aggarwal, A.K. (1995) Structure of BamHI endonuclease bound to DNA: partial folding and unfolding on DNA binding, *Science* **269**, 656–663.

O'Gara, M., Horton, J.R., Roberts, R.J., Cheng, X. (1998) Structures of HhaI methyltransferase complexed with substrates containing mismatches at the target base, *Nat. Struct. Biol.* **5**, 872–877.

Osipiuk, J., Walsh, M.A., Joachimiak, A. (2003) Crystal structure of MboIIA methyltransferase, *Nucleic Acids Res.* **31**, 5440–5448.

Parry, D., Moon, S.A., Liu, H.H., Heslop, P., Connolly, B.A. (2003) DNA recognition by the EcoRV restriction endonuclease probed using base analogues, *J. Mol. Biol.* **331**, 1005–1016.

Pingoud, V., Conzelmann, C., Kinzebach, S., Sudina, A., Metelev, V., Kubareva, E., Bujnicki, J.M., Lurz, R., Luder, G., Xu, S.Y., Pingoud, A. (2003) PspGI, a type II restriction endonuclease from the extreme thermophile Pyrococcus sp.: structural and functional studies to investigate an evolutionary relationship with several mesophilic restriction enzymes, *J. Mol. Biol.* **329**, 913–929.

Pues, H., Bleimling, N., Holz, B., Wolcke, J., Weinhold, E. (1999) Functional roles of the conserved aromatic amino acid residues at position 108 (Motif IV) and position 196 (Motif VIII) in base flipping and catalysis by the N6-adenine DNA methyltransferase

from Thermus aquaticus, *Biochemistry* **38**, 1426–1434.

Reddy, Y.V., Rao, D.N. (2000) Binding of EcoP15I DNA methyltransferase to DNA reveals a large structural distortion within the recognition sequence, *J. Mol. Biol.* **298**, 597–610.

Reich, N.O., Mashhoon, N. (1993) Presteady state kinetics of an S-adenosylmethionine-dependent enzyme, *J. Biol. Chem.* **268**, 9191–9193.

Reinisch, K.M., Chen, L., Verdine, G.L., Lipscomb, W.N. (1995) The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing, *Cell* **82**, 143–153.

Roberts, R.J., Wilson, G.A., Young, F.E. (1977) Recognition sequence of specific endonuclease BamHI from Bacillus amyloliquefaciens H, *Nature* **265**, 82–84.

Roth, M., Helm-Kruse, S., Friedrich, T., Jeltsch, A. (1998) Functional roles of conserved amino acid residues in DNA methyltransferases investigated by site-directed mutagenesis of the EcoRV adenine-N6-methyltransferase, *J. Biol. Chem.* **273**, 17333–17342.

Roth, M., Jeltsch, A. (2001) Changing the target base specificity of the EcoRV DNA methyltransferase by rational de novo protein-design, *Nucleic Acids Res.* **29**, 3137–3144.

Rubin, R.A., Modrich, P. (1977) EcoRI methylase physical and catalytic properties of the homogeneous enzyme, *J. Biol. Chem.* **252**, 7265–7272.

Russell, D.W., Zinder, N.D. (1987) Hemimethylation prevents DNA replication in E. coli, *Cell* **50**, 1071–1079.

Santi, D.V., Garrett, C.E., Barr, P.J. (1983) On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs, *Cell* **33**, 9–10.

Saravanan, M., Bujnicki, J.M., Cymerman, I.A., Rao, D.N., Nagaraja, V. (2004) Type II restriction endonuclease R.KpnI is a member of the HNH nuclease superfamily, *Nucleic Acids Res.* **32**, 6129–6135.

Sasnauskas, G., Halford, S.E., Siksnys, V. (2003) How the BfiI restriction enzyme uses one active site to cut two DNA strands, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6410–6415.

Scavetta, R.D., Thomas, C.B., Walsh, M.A., Szegedi, S., Joachimiak, A., Gumport, R.I., Churchill, M.E.A. (2000) Structure of RsrI methyltransferase, a member of the N6-adenine beta class of DNA methyltransferases, *Nucleic Acids Res.* **28**, 3950–3961.

Schulze, C., Jeltsch, A., Franke, I., Urbanke, C., Pingoud, A. (1998) Crosslinking the EcoRV restriction endonuclease across the DNA-binding site reveals transient intermediates and conformational changes of the enzyme during DNA binding and catalytic turnover, *EMBO J.* **17**, 6757–6766.

Serva, S., Weinhold, E., Roberts, R.J., Klimasauskas, S. (1998) Chemical display of thymine residues flipped out by DNA methyltransferases, *Nucleic Acids Res.* **26**, 3473–3479.

Sethmann, S., Ceglowski, P., Willert, J., Iwanicka-Nowicka, R., Trautner, T.A., Walter, J. (1999) M. PhiBssHII, a novel cytosine-C5-DNA-methyltransferase with target-recognizing domains at separated locations of the enzyme, *EMBO J.* **18**, 3502–3508.

Spyridaki, A., Matzen, C., Lanio, T., Jeltsch, A., Simoncsits, A., Athanasiadis, A., Scheuring-Vanamee, E., Kokkinidis, M., Pingoud, A. (2003) Structural and biochemical characterization of a new Mg^{2+} binding site near Tyr94 in the restriction endonuclease PvuII, *J. Mol. Biol.* **331**, 395–406.

Stahl, F., Wende, W., Jeltsch, A., Pingoud, A. (1996) Introduction of asymmetry in the naturally symmetric restriction endonuclease EcoRV to investigate intersubunit communication in the homodimeric protein, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6175–6180.

Stanford, N.P., Szczelkun, M.D., Marko, J.F., Halford, S.E. (2000) One-and three-dimensional pathways for proteins to reach specific DNA sites, *EMBO J.* **19**, 6546–6557.

Studier, F.W., Bandyopadhyay, P.K. (1988) Model for how type I restriction enzymes select cleavage sites in DNA, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4677–4681.

Surby, M.A., Reich, N.O. (1992) The role of nonspecific DNA in the site-location kinetics of the EcoRI DNA methyltransferase, *Biochemistry* **31**, 2208.

Szegedi, S.S., Gumport, R.I. (2000) DNA binding properties in vivo and target recognition domain sequence alignment analyses of wild-type and mutant RsrI [N6-adenine] DNA methyltransferases, *Nucleic Acids Res.* **28**, 3972–3981.

Szegedi, S.S., Reich, N.O., Gumport, R.I. (2000) Substrate binding in vitro and kinetics of RsrI

[N6-adenine] DNA methyltransferase, *Nucleic Acids Res.* **28**, 3962–3971.

Thomas, C.B., Scavetta, R., Churchill, M.E.A., Gumport, R.I. (2002) A tale of three tails, a mutant, and a dimer: structural studies of RsrI methyltransferase, *FASEB J.* **16**, A923–A924.

Thomas, C.B., Scavetta, R.D., Gumport, R.I., Churchill, M.E.A. (2003) Structures of liganded and unliganded RsrI N6-adenine DNA methyltransferase: a distinct orientation for active cofactor binding, *J. Biol. Chem.* **278**, 26094–26101.

Townson, S.A., Samuelson, J.C., Vanamee, E.S., Edwards, T.A., Escalante, C.R., Xu, S.Y., Aggarwal, A.K. (2004) Crystal structure of BstYI at 1.85A resolution: a thermophilic restriction endonuclease with overlapping specificities to BamHI and BglII, *J. Mol. Biol.* **338**, 725–733.

Trautner, T.A., Pawlek, B., Behrens, B., Willert, J. (1996) Exact size and organization of DNA target-recognizing domains of multispecific DNA-(cytosine-C5)-methyltransferases, *EMBO J.* **15**, 1434–1442.

Urig, S., Gowher, H., Hermann, A., Beck, C., Fatemi, M., Humeny, A., Jeltsch, A. (2002) The Escherichia coli Dam DNA methyltransferase modifies DNA in a highly processive reaction, *J. Mol. Biol.* **319**, 1085–1096.

van Aalten, D.M.F., Erlanson, D.A., Verdine, G.L., Joshua-Tor, L. (1999) A structural snapshot of base-pair opening in DNA, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11809–11814.

Venclovas, C., Siksnys, V. (1995) Different enzymes with similar structures involved in Mg^{2+}-mediated polynucleotidyl transfer, *Nat. Struct. Biol.* **2**, 838–841.

Viadiu, H., Aggarwal, A.K. (1998) The role of metals in catalysis by the restriction endonuclease BamHI, *Nat. Struct. Biol.* **5**, 910–916.

Wenz, C., Jeltsch, A., Pingoud, A. (1996) Probing the indirect readout of the restriction enzyme EcoRV, *J. Biol. Chem.* **271**, 5565–5573.

Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B., Trautner, T.A. (1988) Sequential order of target-recognizing domains in multispecific DNA-methyltransferases, *EMBO J.* **7**, 2601–2609.

Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K., Wilson, K.S. (1993) The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments, *EMBO J.* **12**, 1781–1795.

Xu, Q.S., Kucera, R.B., Roberts, R.J., Guo, H.C. (2004) An asymmetric complex of restriction endonuclease MspI on its palindromic DNA recognition site, *Structure (Cambridge)* **12**, 1741–1747.

Yang, Z., Horton, J.R., Zhou, L., Zhang, X.J., Dong, A., Zhang, X., Schlagman, S.L., Kossykh, V., Hattman, S., Cheng, X. (2003) Structure of the bacteriophage T4 DNA adenine methyltransferase, *Nat. Struct. Biol.* **10**, 849–855.

Zinoviev, V.V., Evdokimov, A.A., Malygin, E.G., Schlagman, S.L., Hattman, S. (2003) Bacteriophage T4 Dam DNA-(N6-adenine)-methyltransferase. Processivity and orientation to the methylation target, *J. Biol. Chem.* **278**, 7829–7833.

Restriction Landmark Genomic Scanning Method: *see* Genome Scanning Method; Restriction Landmark Genomic Scanning (RLGS)

Retinoblastoma, Molecular Genetics of

Vivette D. Brown[1] *and Brenda L. Gallie*[1,2,3]
[1] *Princess Margaret Hospital, Toronto, Canada*
[2] *Hospital for Sick Children, Toronto, Canada*
[3] *University of Toronto, Toronto, Canada*

1 Retinoblastoma, Cancer of Infants 349

2 Genetic Principles of Cancer Discovered through Retinoblastoma 350
2.1 Genetic Locus of the Retinoblastoma Gene 350
2.2 Retinoma 351

3 Spectrum of *RB1* Mutant Alleles and Risk Prediction 351

4 The Retinoblastoma Gene and Protein 352
4.1 RB Protein Structure 353
4.2 RB Family Members 354
4.3 RB Protein Phosphorylation 355
4.4 RB and E2F Protein Families in Cell Cycle Progression 357
4.5 Other Binding Proteins 358
4.6 Differentiation by Retinoblastoma Protein 359

5 RB Family Regulation of Murine Development 360
5.1 *RB1* in Murine Development 360
5.2 p107 and p130 in Murine Development 360
5.3 E2F in Murine Development 362

6 Retinoblastoma in Mice 362
6.1 Viral Oncoprotein-induced Retinoblastoma 362
6.2 Chimeric Knockout Retinoblastoma 363
6.3 Targeted $RB1^{-/-}$ Knockout Retinoblastoma 363

7 RB Protein and Apoptosis 364

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

8 Additional Genomic Changes in Retinoblastoma 365
8.1 Genomic Gains and Losses 365
8.2 Candidate M3-Mn Genes 366
8.3 Causes of Drug Resistance in Retinoblastoma 366

9 Importance of Retinoblastoma 367

Bibliography 368
Books and Reviews 368
Primary Literature 368

Keywords

Apoptosis
The process of programmed cell death.

E2F
A transcription factor that regulates S-phase genes.

LOH
Loss of heterozygosity.

Proband
The very first individual in a family to be diagnosed.

RB1
Retinoblastoma tumor susceptibility gene.

Retinoblast
A precursor cell that becomes a retinal cell.

SV40
Simian virus 40 DNA tumor virus.

TAg
Simian virus 40 large T antigen oncoprotein.

Tumor Suppressor
A gene that prevents a cell from becoming a tumor.

Retinoblastoma is a rare malignant tumor initiated by the inactivation of both alleles of the retinoblastoma tumor susceptibility gene, *RB1*, in a developing human retinal cell. Surprisingly, the protein (pRB) encoded by *RB1* is an important transcription factor. Cell cycle control by pRB is mainly accomplished by transcriptional repression of genes required for cell cycle progression. Control of differentiation by pRB is achieved by the activation of transcription. Almost every kind of tumor has disruption in the retinoblastoma pathway associated with tumor progression, but germline mutation of the *RB1* gene predisposes children to a 95% specific risk of developing retinoblastoma and a significant but much less risk of second primary tumors, such as osteosarcoma and melanoma. Ongoing intense efforts to define the features of the retinal cell that is sensitive to *RB1* loss have potential to once again broadly influence the understanding of cancer in general.

1
Retinoblastoma, Cancer of Infants

Retinoblastoma most commonly is diagnosed only when tumor fills one or both eyes, and the parents notice a white gleam from the pupil of the baby's eye. Often, the diagnosis is definitive based on clinical examination and CT Scan that reveals calcification within the tumor. When only one eye is affected, the preferred treatment is almost always removal of the eye, which is usually curative. When both eyes are involved, the choice of treatment for each eye depends on the size and extent of tumor, the risk of tumor spreading outside the eye, and the risk to vision of the treatment. The new International Intraocular Retinoblastoma Classification (IIRC), which stages each eye with retinoblastoma and predicts the likelihood of treatment curing the eye, is being validated through a global Internet survey.

Eyes with advanced tumor at risk to spread outside the eye, and virtually no possibility of useful vision even if the tumor is controlled (IIRC Group E), are removed. Eyes with medium or large tumors (IIRC Groups B, C, D), no risk of extraocular tumor and some possibility of vision, are now commonly treated by systemic chemotherapy combined with laser and cryotherapy to destroy physically any surviving or chemotherapy-resistant tumor cell. Eyes with small tumors not immediately adjacent to the center of vision or the optic nerve (IIRC Group A) are treated with laser and cryotherapy. Cure is usually obtained within 2 to 4 years. Late recurrences are rare. The fraction of children who lose both eyes to retinoblastoma is now small. Most children end up with vision in atleast one eye.

Retinoblastoma tumor cells may escape from the eye and spread through the optic nerve to the brain or through the blood to grow preferentially in the bone marrow. Until recently, metastatic retinoblastoma was incurable. There are now several reports of long remissions and, possibly, cure following intensive systemic chemotherapy, chemotherapy given into the meningeal space, and megadoses of chemotherapy to destroy any surviving tumor cells followed by stem cell transplant to reconstitute the bone marrow.

Radiation therapy is very effective in controlling retinoblastoma tumor growth.

However, radiation of children with *RB1* mutant alleles greatly increases the risk of a second primary cancer later in life, from the already high risk that the mutant *RB1* allele imposes. Chemotherapy also carries a risk of induction of leukemia, but the magnitude of this risk has not been established.

2 Genetic Principles of Cancer Discovered through Retinoblastoma

The tendency for retinoblastoma to be transmitted in families as an autosomal dominant trait was recognized in the early 1900s. The location of the predisposing gene on chromosome 13q was suggested by the observation in the early 1960s that large deletions involving chromosome 13q14 were associated with retinoblastoma and other abnormalities. However, bilateral retinoblastoma was thought to be a unique special case of cancer in which genetics played some role, while nongenetic mechanisms were presumed, erroneously, to underlie common forms of cancer.

In 1971, Alfred Knudson proposed, from simple analysis of age of diagnosis of bilateral versus unilateral retinoblastoma, that both hereditary (usually bilateral) and nonhereditary (always unilateral) retinoblastoma are initiated by two mutational events, M1 and M2. In 1973, Comings proposed the possibility that these two events could be the two alleles of one gene whose normal function would be suppression of tumors.

2.1 Genetic Locus of the Retinoblastoma Gene

Proof that retinoblastoma tumors had mutation of both alleles of the 13q14 predisposing gene was obtained when genetic markers could distinguish the two alleles. Although the *RB1* gene was not yet cloned, markers near the RB chromosomal locus showed that the heterozygous 13q14 allele associated with the presence of retinoblastoma in members of a family was often the only retained allele in retinoblastoma tumors (loss of heterozygosity, LOH). LOH for 13q14 was observed in both hereditary and unilateral, likely not hereditary, retinoblastoma tumors, proving that Coming's predictions are correct. Both hereditary and nonhereditary retinoblastomas are initiated by mutation of the same tumor suppressor gene. The retinoblastoma studies opened the way to find other tumor suppressor genes in which mutation initiates a wide variety of other cancers. Through the 1980s, numerous papers reported the chromosomal localization of disease-specific tumor suppressor genes by searching for LOH in tumors. Cancer was clearly demonstrated to be the most common somatic genetic disease.

In sporadic bilateral retinoblastoma patients, the M1 occurs either *de novo* in a parent and is transmitted via the germ cells or during embryonal development. Retinoblastoma is initiated when the remaining normal allele is inactivated (M2) in a retinoblast cell carrying M1. Sporadic unilateral retinoblastoma occurs in children with normal *RB1* germline alleles, who develop rare somatic mutations of both alleles (M1/M2) in a developing retinal cell.

Bilaterally affected patients all have one germline *RB1* mutant allele, which results in autosomal dominant transmission of the predisposition to retinoblastoma. Unilaterally affected patients have only a 15% chance that they carry a germline *RB1* mutant allele. Therefore, the standard

approach for unilateral patients is to test the retinoblastoma tumor DNA to identify the two mutant *RB1* alleles and then test the blood DNA for each mutant allele. In this way, a negative result generally rules out heritable retinoblastoma in the unilateral proband, although mosaicism is never excluded. Antecedent relatives are excluded from risk, and future offspring of the unilateral patient can be tested for the mutant alleles found in the tumor of their parent, just in case the parent had undetected germline mosaicism.

The majority of children with heritable retinoblastoma have newly acquired *RB1* mutations. Most frequently, it is the paternal *RB1* allele that is mutated, presumably because of the increased risk of mutation of sperm relative to ova. About 10% of new *RB1* germline mutations can be shown to be mosaic, present in less than 50% of the *RB1* DNA and in only a fraction of the cells of the proband. Clinically, mosaicism may become obvious when a parent has more than one child with retinoblastoma but does not show in blood the mutant *RB1* allele present in the children. Mosaicism cannot be inherited since the predisposing mutation must occur during the development of the embryo.

2.2 Retinoma

Nonprogressing retinal lesions were originally considered to represent "spontaneous regression" of retinoblastoma. Retinoma is a preferable name, which does not imply knowledge of a mechanism. A retinoma is typically a translucent-gray retinal mass, frequently associated with calcification and hyperplasia of the retinal pigment epithelium, unlike the actively proliferating retinoblastoma that shows calcification but appears opaque-white. Pathology of retinoma (also called *retinocytoma*) shows a varied arrangement of neuronal cells with photoreceptor differentiation, axons, Müller glia cells, and astrocytes.

Although retinomas are "benign," they can undergo malignant progression to retinoblastoma. Particularly in a young child, such tumors must be watched regularly for indication of transition to retinoblastoma. *RB1* germline mutations predispose to both retinoma and retinoblastoma in the same patient and in other members of a family who have the mutant *RB1* allele. Although no molecular analysis of retinoma has been reported, germline *RB1* mutations have been identified in several patients with bilateral or unilateral multifocal retinomas as their only clinical manifestation.

3 Spectrum of *RB1* Mutant Alleles and Risk Prediction

RB1 mutations are spread throughout the gene including the promoter, most exons, and splicing regions of introns (Fig. 1). More than 50% of *RB1* mutations have been found only once, and the majority of germline mutations originate in the proband. Molecular identification of *RB1* mutations supports accurate determination of the risk of retinoblastoma in relatives.

RB1 mutations are most frequently "null" mutations that are out of frame and result in premature STOP codons (Fig. 1). Premature truncation of translation in turn triggers nonsense-mediated decay, unstable mutant mRNA, and no detectable pRB. The only significant hot spots for *RB1* mutations are the "recurrent" mutations,

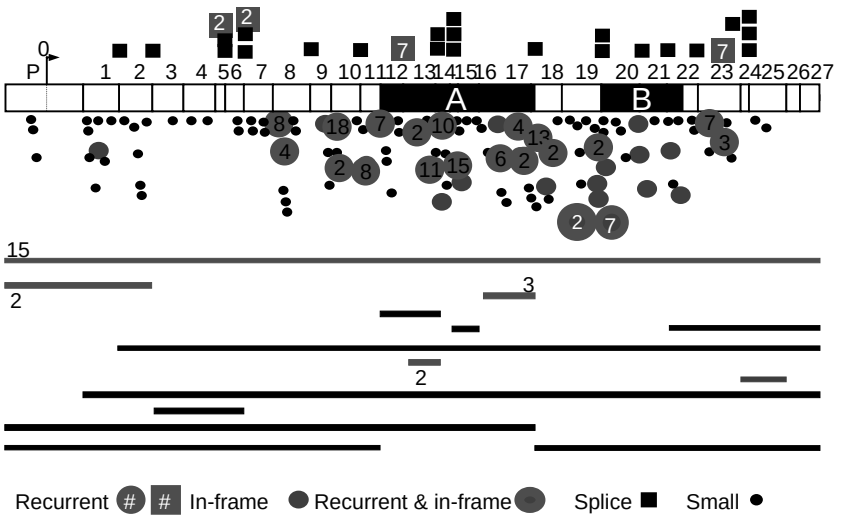

Fig. 1 Retinoblastoma-initiating mutations are distributed throughout the *RB1* gene. Of the 230 germline mutations, 53% have been reported only once. Recurrent mutations are shown in red, with the number of occurrences indicated in the box. In-frame mutations are shown in blue. Boxes indicate splice mutations. Circles indicate small point or few nucleotide mutations. Lines indicate lengths of deletions. (See color plate p. xxxiii).

most commonly CGA (arginine) mutating to TGA (STOP) because CpG codons are prone to mutate to TG. Missense mutations almost invariably involve the A/B "pocket" domain of pRB, which is critical in many protein interactions of pRB.

In bilateral probands, 93% of the germline mutations are null, 6% are in-frame, and 1% are promoter. In the uncommon unilateral germline probands (both familial and sporadic), the ratios of null (57%), in-frame (40%), and promoter (3%) mutations are very different. The presence of unilateral tumor indicates reduced expressivity of the germline *RB1* mutation, manifesting as fewer than expected tumors, reducing the likelihood of a null allele. Reduced penetrance/expressivity *RB1* mutant alleles retain enough activity to prevent some, but not all, retinoblastoma tumors. The types of mutant alleles identified in retinoblastoma tumors, presumably of somatic origin (88% null, 3% in-frame, and 8% promoter methylation), are different again.

The second *RB1* allele (M2) is lost by LOH in 52% of sporadic retinoblastoma tumors. The likelihood that the M2 event is LOH also varies with the type of M1 event. While 50% of the tumors with M1 whole gene and exonic deletions show LOH, 88% of tumors with methylation of the promoter show LOH. The biological reasons for these differences are unknown.

4
The Retinoblastoma Gene and Protein

The *RB1* gene was cloned by identification of a chromosome 13 DNA clone that was totally deleted from a retinoblastoma tumor. This was correctly interpreted to be LOH for a deleted allele, and the clone

turned out to be located within an exon of the *RB1* gene. The 190-kb *RB1* gene transcribes a 4.7-kb RNA in 27 exons. Comparison of the human and mouse *RB1* genes shows 80% homology in the first 235-bp upstream of the initiating methionine, 89% homology in the coding sequence, and 77% homology in the 3′ untranslated region. The human and mouse promoters contain binding sites for the transcriptional elements ATF, SP1, and E2F, but are devoid of TATA and CAAT boxes. The human and mouse proteins show 91% identity, suggesting that the function of this protein is highly conserved. The retinoblastoma protein (pRB) is a nuclear phosphoprotein of 928 amino acids that runs as a doublet of molecular weight 110–114 kDa on SDS polyacrylamide gel electrophoresis. The slower migrating band corresponds to the phosphorylated species of pRB.

RB1 function is abrogated either by deletion or mutation in retinoblastoma. In most other tumors, *RB1* is mutated or functionally inactivated, confirming that *RB1* is a tumor suppressor gene. The reintroduction of *RB1* into certain cells that have nonfunctional *RB1* alleles can suppress proliferation. For example, retroviral introduction of *RB1* into the retinoblastoma cell line WERI-Rb27, the osteosarcoma cell line Saos2, the prostate carcinoma cell line DU145, and the bladder carcinoma cell line HTB9 resulted in morphological changes, reduction in growth rate, reduced colony formation in soft agar, and reduced tumorigenicity in nude mice. However, introduction of pRB into the U2OS osteosarcoma cell line, which contains a normal retinoblastoma allele, or the C33A cervical carcinoma cell line that contains a mutant pRB, had no effect. The ability of pRB to reverse the malignant phenotype of tumor cells may depend on what other genetic alterations have occurred in the cell.

4.1 RB Protein Structure

Functional domains of the coding region of pRB were delineated using deletion mutant analysis. The protein is divided into four regions, the N terminus, the A and B domains separated by a spacer region, and the C terminus (Fig. 2). Disruption of growth suppression by pRB requires the A and B domains as well as portions of the C-terminal domain. The nuclear c-Abl protein binds to the C terminus of pRB, promoting proliferation. The C terminus is required for growth suppression along with the A and B domains.

The A/B region of pRB binds to the LXCXE (Leucine-X-Cysteine-X-Glutamic acid, where X is any amino acid) epitope of the viral oncoproteins, simian virus 40 largeT antigen (TAg), adenovirus E1A, and human papilloma virus-16 E7. The importance of the A/B domain is evident by the large number of endogenous proteins binding to it and the number of tumorigenic missense mutations affecting it (Fig. 1). X-ray crystallography shows that the A domain is required for stable folding of the B region, the site of binding of LXCXE proteins. Even when the A and B regions of pRB are expressed on separate proteins, they are able to form a repressor motif that is regulated by phosphorylation. Another highly conserved region is the intervening sequence between the A and B regions, suggesting an additional protein binding site, or a critical role in folding. Indeed, the E2F transcription factor binds to the highly conserved interface between the A and B regions of pRB.

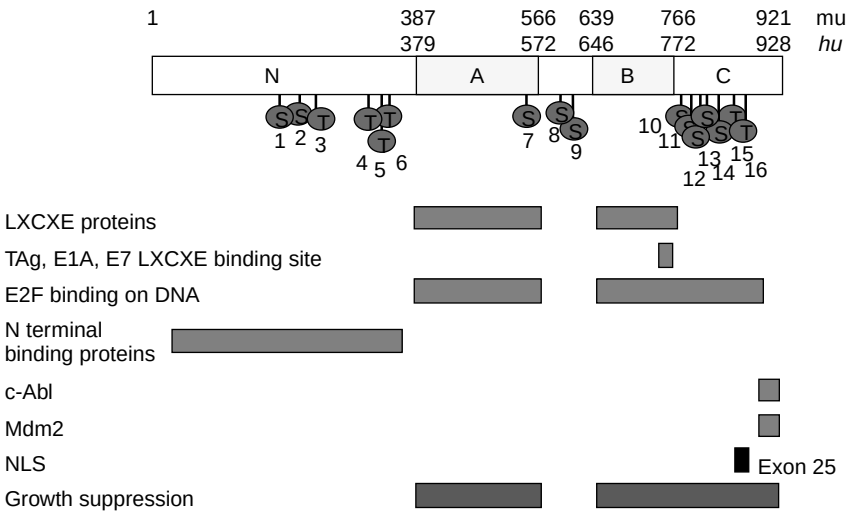

Fig. 2 Structure of pRB. Numbers above the schematic designate the amino acids. Mu corresponds to murine pRB (normal numbers) and hu to human pRB (italics). The N and C termini and the A and B regions are required for binding the SV40 large T antigen. Positions of the 16 phosphorylation sites in murine pRB are shown with S (serine) (red) and T (threonine) (blue). Areas of pRB required for different functions are indicated below the schematic. (See color plate p. xxxiii).

The ability of pRB to enter the nucleus is regulated by a bipartite nuclear localization signal located in the C terminus. The underphosphorylated (active) form of pRB binds the nucleus, following salt extraction, more effectively than the phosphorylated (inactive) form. A/B region mutants are poorly retained in the nucleus, and the A/B region can experimentally translocate a cytoplasmic protein to the nucleus.

4.2 RB Family Members

Two genes/proteins homologous to *RB1*, p107, and p130, which had previously been observed to bind E1A, were identified by sequence comparison. Some antibodies that were developed against pRB also recognized p107 and p130, suggesting structural homology. All three RB family proteins are found in both hypo- and hyperphosphorylated forms. However, while p107 and p130 bind to cyclin A and cyclin E, pRB does not. Cyclin A binds to the A/B spacer region of p107, which is highly conserved between p107 and p130.

The RB family members express differently during the cell cycle. During quiescence, p130 is expressed, but disappears as cells are stimulated to reenter the cell cycle. Quiescent cells do not express p107, but as cells reenter the cell cycle, p107 accumulates. In contrast, pRB levels are generally constant throughout the cell cycle, although there is a small increase in the level of pRB following growth stimulation.

While tumor suppression was defined first by the powerful function of pRB, there is not much evidence that the other RB family members can independently act as tumor suppressors. However, to

some extent, the RB family compensate for each other. Exogenous overexpression of p107 can suppress the growth of C33A cells, which have a mutant pRB, even though wild-type pRB is unable to. Growth suppression in C33A cells by p107 is mediated through its cyclin/cdk binding domain in addition to the A/B and C domains. p130 has been shown to suppress the growth of only one cell line, the human glioblastoma cell line T98G, which is refractory to suppression by pRB or p107. Few mutations of p107 or p130 have been observed in human tumors.

4.3 RB Protein Phosphorylation

One of the first clues to the function of pRB came following the discovery that the phosphorylation status of pRB depends on the cell cycle. pRB is underphosphorylated in G_0/G_1 phase and hyperphosphorylated in the S- and G_2/M phases in all cell types. pRB phosphorylation occurs on serine and threonine residues. No phosphotyrosine is evident. Following mitogenic stimulation, pRB is phosphorylated in three stages during cell cycle entry and progression: mid-G_1-phase, S-phase, and G_2/M (Fig. 3). Dephosphorylation of pRB occurs after the beginning of mitosis, and may be regulated by differentiation signals.

Because of its significance in growth regulation, great attention has been paid to the regulation of pRB by phosphorylation. Large T antigen specifically binds unphosphorylated (active) pRB to transform cells, possibly by releasing cellular-binding proteins. Inactivation of pRB by phosphorylation occurs through the action of the cyclin-dependent kinases (cdks). The cdk family member, $p34^{cdc2}$, was first found to bind to and phosphorylate pRB. The cdc2 kinase is the active subunit of the M-phase promoting factor, which phosphorylates specific serines or threonines in its consensus site Z-S/T-P-X-Z (where X is polar and Z is generally basic). pRB contains 16 potential cdc2 phosphorylation sites, of which at least nine have been shown to be phosphorylated *in vivo* (Fig. 2). Phosphopeptide analysis shows that *in vitro* phosphorylation by cdc2 corresponds to *in vivo* phosphorylation of pRB. However, cdc2 is activated in late G_2-and M-phases and a homologous kinase, cdk2, may first bind to and phosphorylate pRB, consistent with phosphorylation of pRB to promote exit from G_1.

Cdks are activated following binding to cyclins and phosphorylation of a conserved threonine by the Cdk-activating kinase. The expression of cyclins varies during the cell cycle, with cyclin D and E expressed during G_1, cyclin A during S-phase, and cyclin B during mitosis. Cyclins E and A form a functional unit by binding to cdk2 in G_1- or S-phase respectively, whereas cyclin D forms a complex with cdk4 or cdk6 in G_1. Cyclin B binds to cdc2 in late G_2.

While complexes of each of these cyclins can phosphorylate pRB, only cyclins A, D, and E are able to reverse the ability of pRB to block cells in G_1. Although initial reports suggested that phosphorylation of pRB occurred on the same residues whether phosphorylated by cyclin D-cdk4, cyclin E-cdk2, or cyclin A-cdk2 complexes, it is now recognized that different cyclin-cdk complexes regulate different phosphorylation sites, in a collaborative effort to completely inactivate pRB. In yeast, three G_1 cyclins, Cln1, Cln2, and Cln3, regulate the activity of Cdc28 throughout G_1. Mammalian cyclin E is able to substitute for a *cln2* mutant and cyclin D1 for a *cln3* mutant. Phosphorylation of pRB in yeast requires a combination of Cln3 and either Cln1

or Cln2, suggesting that cyclins D and E must collaborate to completely phosphorylate pRB. Study of cyclin D-cdk4/6 and cyclin E-cdk2 kinase inhibitors activities in mammalian cells confirms these results. In U2OS cells, pRB is partially phosphorylated by cyclin D kinase activity, with complete phosphorylation requiring cyclin E activity. Phosphorylation of pRB only by cyclin D kinase did not release pRB from its interaction with E2F1. Continued phosphorylation of pRB on specific sites in S-phase by cyclin A-cdk2 activity may also be important for the completion of DNA synthesis.

pRB is also regulated by another kinase, cyclin C-cdk3, at the G_0/G_1 transition. Phosphorylation of pRB at specific substrate sites is required for cells to exit G_0 efficiently (Fig. 3). Cyclin C directs phosphorylation of pRB in a temporal pattern that precedes pRB phosphorylation by cyclin D or cyclin E complexes. Therefore, cyclin C plays an important role in regulating the G_0-to-G_1 transition in a pRB-dependent manner.

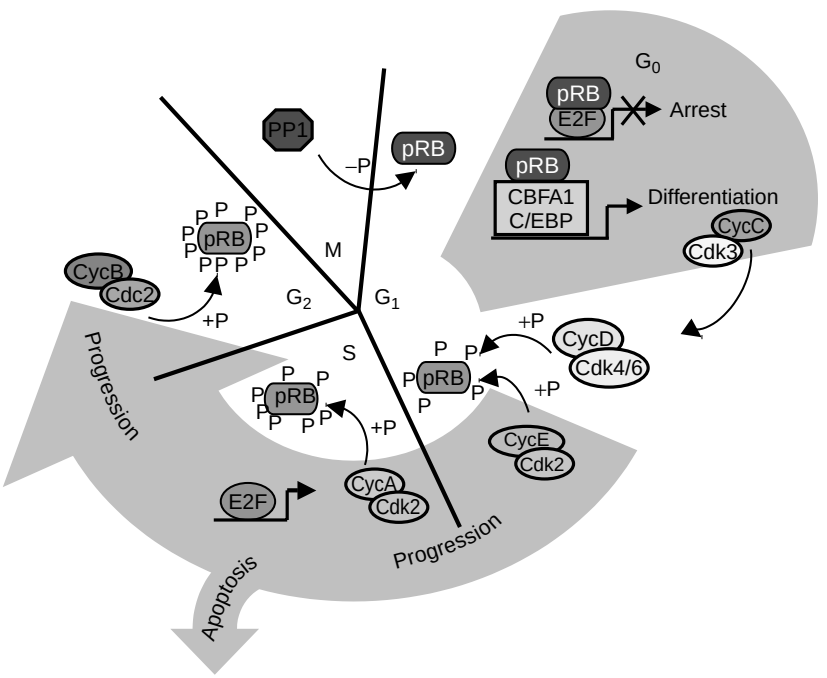

Fig. 3 pRB regulates cell cycle progression, differentiation, and apoptosis. Hypophosphorylated pRB represses genes necessary for S-phase entrance by repressing transcriptional activation by E2F, causing cell cycle arrest, and activates genes necessary for differentiation by binding cell type–specific transcription factors. Phosphorylation of pRB by cyclin C-cdk3 kinase stimulates exit from G_0, and in late G_1 phosphorylation by cyclin D-cdk4/6 and cyclin E-cdk2 kinases releases E2F, allowing cell cycle progression. Continual phosphorylation of pRB occurs by cyclin A-cdk2 and cyclin B-cdc2 complexes in S and G_2 phases. pRB is activated in late mitosis by dephosphorylation. Inappropriate entry into S-phase in the absence of pRB induces apoptosis.

Dephosphorylation (reactivation) of pRB is dependent on the activity of phosphoprotein phosphatase type 1 (PP1), which associates with pRB *in vivo* during mitosis. Microinjection of PP1 into either the nucleus or the cytoplasm of cells synchronized in G_1 increases the resistance of pRB to extraction from the nucleus and correlates with inhibition of S-phase progression.

4.4 RB and E2F Protein Families in Cell Cycle Progression

Viral oncoproteins induce proliferation of mammalian cells by interfering with binding of pRB to cellular proteins. The first such cellular protein to be identified was the transcription factor E2F. The E2F transcription factor regulates the activity of promoters containing an E2F binding site, originally identified as the sequence element in the adenovirus E2 promoter that allowed transcription in mammalian cells. The E2F binding site (TTTSSCGC where S is C or G) has been identified in promoters of genes encoding S-phase-regulatory proteins such as DNA polymerase α, thymidylate synthase (TS), proliferating cell nuclear antigen (PCNA) and ribonucleotide reductase (RR), and cell cycle progression genes such as cyclin A, cyclin E, cdc2, and B-myb. The *E2F1* and *RB1* genes are also regulated by E2F activity, suggesting the existence of an autoregulatory loop in the pRB/E2F pathway. pRB mutants that bind free E2F but do not repress transcription are unable to block S-phase entry. Coexpression of *E2F1* with pRB releases the G_1 block and pushes cells into S-phase. Thus, the primary way that pRB blocks the cell cycle in the G_1-phase is through repression of E2F-responsive genes.

The E2F family of proteins consists of seven members (E2F1, -2, -3, -4, -5, -6, and -7). E2F1-5 proteins each contain a DNA binding domain, a transactivation domain, and a C terminus region that binds pRB or other pRB family members. The ability of E2F proteins to regulate transcription is also dependent on binding to the DP family of proteins. Two mammalian DP proteins bind E2F to enhance interaction with E2F binding sites and activate E2F-responsive promoters. However, not all genes with an E2F site are activated by E2F. *In vivo,* E2F is released from the E2F binding site of the B-myb promoter at the same time that B-myb transcription is activated, suggesting that the E2F complex suppresses this promoter. E2F6 contains neither the transactivation domain nor the C terminus pRB binding region, but represses transcription by complexing with DP proteins to either actively repress or replace other E2F/DP complexes. E2F7 has two DNA binding domains but binds the E2F binding consensus site independently of DP proteins. It too does not have a transactivation domain or the pRB binding region, but functions as a repressor of certain E2F-regulated genes.

Overexpression of E2F1 pushes cells into untimely S-phase, followed immediately by induction of apoptosis. pRB protects cells from apoptosis by complexing with E2F1. However, while a transactivation-defective E2F1 mutant induced apoptosis, a DNA-binding defective E2F1 mutant did not, suggesting that the pRB/E2F1 complex may repress genes necessary for induction of apoptosis.

Different RB family proteins bind E2F family proteins at different phases of the cell cycle. pRB binds E2F1- 4, whereas p107 and p130 bind only E2F4 and 5. E2F4 or 5 are predominantly bound to p130 in quiescent G_0 cells and to p107

in proliferating cells. pRB binds and regulates E2F1, 2, and 3 in G_1 but binds to E2F4 when cells are induced to reenter the cell cycle.

The activity of E2F is also regulated in a cell cycle–dependent manner by degradation of free E2F protein. E2F1 and 4 have C-terminal sequences that target them for degradation via the ubiquitin-proteasome pathway. Binding of pRB to E2F1 or p107 to E2F4 increases the half-life of the E2F protein. Endogenous E2F1 is more stable in G_0/G_1 when bound to pRB than in S-phase. This suggests that the RB family may maintain cells in G_1 by assuring that E2F is present to repress promoters. E2F activity is also stabilized by acetylation. pRB also blocks recruitment of the preinitiation complex to the promoter by E2F.

4.5 Other Binding Proteins

A diversity of action is evident from the large number of proteins that bind to pRB with different functions (Table 1). In addition to E2F, pRB binds a number of general transcription factors (ATF-2 and SP1) and cell type–specific transcription factors (Elf-1, PU-1, MyoD, and NF-IL6). Generally, the binding of pRB represses transcriptional activity; however, binding to factors such as c-Jun, MyoD, C/EBP, and NF-IL6 has a positive effect on differentiation of specific cell types,

Tab. 1 Some of the proteins that bind pRB.

Function	*pRB-binding Proteins*
Cdk inhibitor	p21
Chromatin modulator	BRG1
Chromosome scaffold	H-Nuc
Corepressor	CtIP, RbAp48, RBP1, HBP1
Cyclins	Cyclins D1, 2, 3
Histone deacetylase	HDAC1
Molecular chaperone	hsp75
Nuclear matrix	Lamin A, p84
Phosphatase	PP-1a2
Pol I transcription factor	UBF
Pol II transcription factor	TAFII250
Pol III transcription factor	TFIIIB
Potential tumor suppressor	Prohibitin, BRCA1
Protein kinase	RbK
Regulator of p53 stability	MDM2
Replication licensing factor	MCM7
Ser/Thr kinase	cdc2, cdk2
Signaling molecule	Raf-1
Transcription factor	ATF-2, c-Jun, c-Myc, N-Myc, C/EBP, E2F-1, Elf-1, hBRM, Id-2, MyoD, Myogenin, NF-IL6, PU.1, Sp1, Pax-3, PHox, Chx10, RIZ
Transformation inducer	Bog
Tyrosine kinase	c-Abl
Viral oncoprotein	E1A, E7, Large T Ag
Viral transcription factor	EBNA-5, HCMV IE2

although the precise mechanism(s) of activation is not yet known.

pRB may repress the activity of factors such as E2F1 by recruitment of histone deacetylase (HDAC1) to the promoter. HDAC1 facilitates the formation of nucleosome structures that block access of positive transcription factors to the promoter. In addition, pRB binds a number of corepressors including RbAp48, RBP1, and C-terminal interacting protein (CtIP), which function by directly repressing transcription or by recruiting HDAC1. However, pRB also represses transcription of promoters that are not sensitive to histone deacetylase activity, by unknown mechanisms. Thus, pRB can actively repress transcription factors, block access of transcription factors to basal transcriptional machinery, or directly inhibit the basal transcriptional machinery.

In addition to regulation of polymerase II transcription, pRB regulates polymerase I genes by binding and inhibiting the ability of an upstream binding factor to bind the ribosomal DNA promoter. pRB binds the general polymerase III transcription factor TFIIIB to regulate genes such as tRNA, 5s RNA, and TATA-box containing genes such as U6. Polymerase III transcription is also regulated by binding of p107 and p130 to the BRF subunit of TFIIIB.

The signaling molecule Raf-1 binds pRB within 30 min of serum stimulation, directly linking mitogenic signaling and cell cycle regulation, which may be one mechanism of transducing Ras cell surface signals to the cell cycle. Inactivation of Ras by a monoclonal antibody in $RB1^{+/+}$ cells efficiently inhibits DNA synthesis. In $RB1^{-/-}$ cells, inhibition of DNA synthesis is defective. Thus, pRB is required for cell cycle progression downstream of Ras.

4.6 Differentiation by Retinoblastoma Protein

Cell culture systems have revealed a role of pRB in determining the fate of the cell. pRB positively regulates terminal adipocyte differentiation of murine embryonic fibroblasts by binding to and activating the CCAAT/enhancer-binding proteins (C/EBPs), a family of transcription factors necessary for adipocyte differentiation. When the cell is induced to differentiate, pRB binds to C/EBP. When U937 cells respond to phorbol 12-myristate 13-acetate by differentiation into monocyte/macrophage lineages, pRB binds NF-IL6, another member of the C/EBP family. Production and maintenance of a terminally differentiated muscle cell phenotype requires binding of pRB to MyoD both *in vitro* and *in vivo*. Differentiating human keratinocytes contain pRB/c-Jun complexes that are absent from nonsynchronized cycling cells, in part by facilitating binding of c-Jun to an AP-1 consensus site. In bone development, pRB interacts with an osteoblast transcription factor, CBFA1, resulting in synergistic transactivation of an osteoblast-specific reporter. This finding is significant since osteosarcoma is the second most common tumor after retinoblastoma in $RB1^{+/-}$ individuals, and loss of pRB occurs in 60% of sporadic osteosarcomas, suggesting a bone-specific tumor suppressor function of pRB. These culture experiments are consistent with the idea that differentiation involves activation by pRB of genes that are needed for cell type–specific differentiation.

5
RB Family Regulation of Murine Development

5.1
RB1 in Murine Development

RB1$^{-/-}$ mice show normal early retinal development up to embryonic days (E)13–15, but then die with abnormalities in neurogenesis and erythropoiesis due to placental failure (Table 2). Wild-type trophoblast rescues *RB1*$^{-/-}$ embryos until birth, when a muscle defect, previously recognized in partial rescue of *RB1*$^{-/-}$ embryos, is lethal. The cause of early embryonic lethality is excessive proliferation of trophoblast cells and a severe disruption of the normal labyrinth architecture in the placenta, resulting in decreased vascularization and reduction in placental transportation function. *RB1*$^{-/-}$ mice that are reconstituted with a functionally normal placenta lack many of the phenotypes of *RB1*$^{-/-}$ mice and show histologically normal dorsal root ganglia, brain, retina, spinal cord, and liver, but they still die at birth or shortly thereafter with marked defects in skeletal muscle and apoptosis of peripheral nervous system and lens.

Tagged *RB1*$^{-/-}$ cells in the brains of chimeric embryos exhibited extensive S-phase entry. However, unlike the widespread apoptosis in the germline *RB1*$^{-/-}$ mice, the majority of the *RB1*$^{-/-}$ cells in the chimera survived and differentiated into neurons. The developing retina of chimeric embryos showed ectopic mitoses and substantial cell degeneration, while the contribution of *RB1*$^{-/-}$ cells to the adult retina was much reduced. Therefore, in the CNS, cell-autonomous cell cycle regulation is dependent on functional pRB, but neuronal cell survival and fate may be non-cell-autonomous. In the chimeric CNS, an exception was the *RB1*$^{-/-}$ Purkinje neurons that were lost. Thus, pRB function may be required in a cell-autonomous manner for cell cycle regulation in most tissues but suppression of apoptosis and differentiation may be either cell-autonomous or non-cell-autonomous, depending on the tissue involved. These data suggest that the role of pRB in development and differentiation is more important for some cell types than others.

5.2
p107 and p130 in Murine Development

The importance of the pRB family in differentiation is also demonstrated by *p107*$^{-/-}$; *p130*$^{-/-}$ mice, which are viable until birth and show deregulated chondrocyte growth and defective endochondral bone formation (reduced rib cage size and shortened limbs). The importance of each pRB family member may be dependent on the level of expression in a particular cell type, and may also depend on the strain of mouse, suggesting the existence of strain-specific modifier genes. Deletion of *p130* in Balb/cJ mice resulted in embryonic lethality, while the same deletion in C57BL/6J mice produced viable mice with a reduction in the number of spinal cord and dorsal root ganglia neurons and myocytes, and neural tube and brain apoptosis with increased proliferation.

Balb/cJ and C57BL/6J mice lacking *p107* were viable like their counterparts but the Balb/cJ mice displayed impaired growth, reaching only 50% of the weight of their counterparts by 21 days of age. There was also myeloid hyperplasia in the spleen and liver and a twofold acceleration of the cell cycle in *p107*$^{-/-}$ fibroblasts and myoblasts, suggesting a role for *p107* in negatively regulating the overall length of the cell cycle.

Tab. 2 *RB1* phenotypes in mice.

Mouse	***Phenotype***
No retinoblastoma	
RB1$^{+/-}$	*RB1*$^{-/-}$ pituitary and thyroid tumors
RB1$^{-/-}$	Lethal E13-E15, abnormalities in neurogenesis and erythropoeisis, apoptosis, and inappropriate S-phase entry in CNS, PNS, and lens
RB1$^{-/-}$ with normal placenta	Die at birth with marked defects in muscle and lens, no abnormalities in neurogenesis and erythropoeisis
RB1$^{-/-}$; *RB1*$^{+/+}$ chimeric	Normal, apoptosis of *RB1*$^{-/-}$ cells in retina
NesCre1-Cre; *RB1*$^{loxP/loxP}$	Apoptosis in lens and PNS but not in CNS, inappropriate S-phase entry in CNS, PNS, and lens
IRBP-Cre; *RB1*$^{loxP/loxP}$	Neuroendocrine tumors of the pineal and pituitary gland
p107$^{-/-}$	Normal on C57B/6J, 50% smaller on Balb/cJ
p130$^{-/-}$	Normal on C57B/6J, die at E11.5 on Balb/cJ
RB1$^{+/-}$; *p107*$^{-/-}$	Retinal dysplasia; *RB1*$^{-/-}$; *p107*$^{-/-}$ pituitary and thyroid tumors
RB1$^{-/-}$; *p107*$^{-/-}$	Lethal E11.5
RB1$^{+/-}$; *E2F1*$^{-/-}$	Reduced frequency of pituitary and thyroid tumors, increased life span compared to *RB1*$^{+/-}$ mouse
RB1$^{-/-}$; *E2F1*$^{-/-}$	Lethal E17
RB1$^{+/-}$; *p53*$^{-/-}$	Retinal dysplasia, *RB1*$^{-/-}$; *p53*$^{-/-}$ pituitary and thyroid tumor
RB1$^{+/-}$; *Arf*$^{-/-}$	Accelerated pituitary tumor development compared to *RB1*$^{+/-}$ mouse
RB1$^{-/-}$; *Apaf1*$^{-/-}$	No apoptosis in CNS and lens, reduced apoptosis in PNS and skeletal muscles
RB1$^{-/-}$; *Id2*$^{-/-}$	Survive to term with no defects in neurogenesis and hematopoiesis seen in the *RB1*$^{-/-}$ mice, die at birth with severe reduction in muscle tissue
MMTV-RB$^{\Delta K}$	Dysplasia in mammary tissue, breast cancer
Papillomavirus E7	Apoptosis in the retina
Viral oncoprotein-induced retinoblastoma	
SV40 TAg transgene (retina)	Retinoblastoma, brain tumor
Papillomavirus E7 & E6 transgene (retina)	Retinoblastoma
Papillomavirus E7, p53$^{-/-}$	Retinoblastoma
Chimeric knockout retinoblastoma	
RB1$^{-/-}$; *p107*$^{-/-}$ (chimera)	Retinoblastoma
Targeted RB1$^{-/-}$ *knockout retinoblastoma*	
Pax6α-Cre; *RB1*$^{loxP/loxP}$; *p107*$^{-/-}$	Retinoblastoma
NesCre1-Cre; *RB1*$^{loxP/loxP}$; *p130*$^{-/-}$	Maternal inheritance results in retinoblastoma

The expression of *p107* and *p130* mRNA helps explain some of these data. While *p130* is expressed at very low levels throughout embryogenesis, there is a notable increase in expression in developing bone, which would explain the bone phenotype in the $p107^{-/-}$; $p130^{-/-}$ knockout mice. *p107*, on the other hand, is expressed highly in the heart, lung, kidney, and intestines but there is no detectable phenotype in these tissues in the $p107^{-/-}$ mice. There is also high expression of *p107* in the CNS and liver, which would explain the accelerated apoptosis seen in the liver and CNS in the $RB1^{-/-}$; $p107^{-/-}$ mice compared to $RB1^{-/-}$ mice.

5.3 E2F in Murine Development

pRB's ability to block cell growth has always been associated with its ability to block E2F activity. Therefore, it was hypothesized that the loss of E2F in the mouse should have a profound effect on development. Surprisingly, the E2F1 knockout mouse was viable but highly tumor prone and contained some tissue atrophy. This suggested that E2F1 was not vital for proliferation of all cell types, possibly because the other E2Fs can compensate, but is required to block proliferation in some cells while promoting proliferation in other cells. Of particular interest is the enlargement of the thymus and lymph nodes as a result of lack of E2F1. This expansion is a result of an increase in the number of mature CD4/CD8 cells and appears to be due to a lack of apoptosis, consistent with the role of E2F1 in apoptosis. However, E2F1 appears to be an important target of pRB in the regulation of tumorigenesis and development. The loss of E2F1 in the *RB1* heterozygous mouse resulted in reduced frequency of pituitary and thyroid tumors and increased life span of the $RB1^{+/-}$; $E2F1^{-/-}$ mouse compared to $RB1^{+/-}$ mouse. Loss of both *RB1* and E2F1 in the $RB1^{-/-}$; $E2F1^{-/-}$ mouse resulted in embryonic lethality but the mouse lived longer than the $RB1^{-/-}$ mouse (E17 compared with E13.5), suggesting that some functions of pRB during development occur through E2F1. Loss of DP1 leads to embryonic lethality due to failure of extra-embryonic lineage development, suggesting a role for the pRB/E2F pathway in placental development.

6 Retinoblastoma in Mice

The generation of $RB1^{-/-}$ mice was expected to provide information regarding development of retinoblastoma. However, embryonic lethality in $RB1^{-/-}$ mice precluded studies of later development of the retina. Unlike humans, $RB1^{+/-}$ mice do not develop retinoblastoma but instead develop pituitary tumors in which both *RB1* alleles are inactivated. The extreme sensitivity of developing human retina to loss of *RB1* has stimulated intense efforts to mimic human retinoblastoma in mice. The features of the susceptible retinal cell, when defined, may well be the features shared by cancer-susceptible cells in many tissues.

6.1 Viral Oncoprotein-induced Retinoblastoma

The first transgenic mouse model of retinoblastoma arose when the simian virus 40 large T antigen (TAg), expressed from the human luteinizing hormone beta subunit (LHβ) promoter in order to

produce pituitary tumors, was serendipitously expressed in a specific cell of retina with malignant potential. This transgenic line of mice develops focal bilateral retinal tumors that are very similar to human retinoblastoma, because of high expression of TAg in the retina. Viral oncoproteins such as HPV E7 and SV40 TAg inactivate pRB when they bind the A/B domain of pRB. TAg binds the active, hypophosphorylated, but not the inactive, hyperphosphorylated form of pRB and also inactivates p107, p130, p53, and other proteins. However, expression of TAg or E7 from the retina-specific promoter interphotoreceptor retinoid-binding protein (IRBP) or opsins does not result in tumors like retinoblastoma. However, IRBP-driven HPV E7 results in retinoblastoma in p53$^{-/-}$ mice.

6.2 Chimeric Knockout Retinoblastoma

The retina was normal in both embryonic *RB1*$^{-/-}$ mice and chimeric *RB1*$^{-/-}$; *RB1*$^{+/+}$ mice, and *RB1*$^{-/-}$ cells contributed to most tissues. However, *RB1*-deficient cells were excluded from the outer and inner nuclear layer of the retina by apoptosis, suggesting that pRB is required during specific stages of postnatal retinal differentiation. While both the *p107*$^{-/-}$ and *p130*$^{-/-}$ mice are viable with no retinal dysplasia, loss of *p107* and one copy of pRB in the *RB1*$^{+/-}$; *p107*$^{-/-}$ resulted in retinal dysplasia. Although the *RB1*$^{-/-}$; *p107*$^{-/-}$ mice die at E11.5, preventing evaluation of the developing retina, loss of both p107 and pRB in the retinoblasts of *RB1*$^{-/-}$; *p107*$^{-/-}$ chimeric mice resulted in retinoblastoma. Using expression markers for terminally differentiated retinal cell types, the authors concluded that the retinoblastoma originated in cells committed to the amacrine cell compartment of the inner nuclear layer (INL) but not in cells committed to the outer nuclear layer (ONL).

6.3 Targeted *RB1*$^{-/-}$ Knockout Retinoblastoma

Conditional inactivation of *RB1* by expression of Cre recombinase in IRBP-expressing photoreceptor cells (IRBP-Cre; *RB1*$^{loxP/loxP}$ mice) failed to cause retinoblastoma, even when *p107* and *TP53* were also inactivated. Conditional inactivation of *RB1* by expression of Cre from the Pax6α enhancer in *p107*$^{-/-}$; *RB1*$^{loxP/loxP}$ mice resulted in ectopic proliferation of all retinal precursors but only the precursors of horizontal, amacrine, and Müller glia cells survived and terminally differentiated. The Pax6α enhancer is active in peripheral retinal progenitors at E10. Sixty-eight percent of these mice developed tumors in the peripheral retina as early as P8, with larger tumors filling the vitreous at later stages. Markers for amacrine (NeuroD and Math3) and Müller glia (Hes5) were present in the P8 tumor, with few markers for photoreceptor (Crx), bipolar (Chx10), and ganglion cells (Brn3B). P30 tumors expressed markers for amacrine (NeuroD, Math3, Pax6, syntaxin, and calretinin) and Müller glia (Hes5, GFAP, and CRALBP) only.

Conditional inactivation of *RB1* by expression of Cre from the *nestin* promoter (*NesCre1*) results in mosaic expression of Cre through all layers of the retina with maternal NesCre1 inheritance. Breeding these mice with p130 knockout mice resulted in retinoblastoma tumor formation in five out of

five mice, of which four developed bilateral retinoblastoma. The tumors expressed amacrine markers (syntaxin and calretinin) and a few Müller glia cells (GFAP expressing) were also present. However, no tumor stained for the photoreceptor-specific gene IRBP, the bipolar cell marker protein kinase C-α (PKC-α), or the retinal ganglion cell marker Brn3b. A potential amacrine cell origin of these mouse tumors is consistent with the earlier mouse model. However, these mouse tumors contain Homer–Wright rosettes but not the classical Flexner–Wintersteiner rosettes pathognomonic of human retinoblastoma, nor any evidence of photoreceptor differentiation.

Examination of the tumor for markers may define some features of the cell of origin, but may be misleading: progenitor cells destined to follow a pathway that is dependent on pRB may adopt a different fate in the absence of pRB. Retinal progenitors go through a series of changes in intrinsic properties that control their competence to make different cell types. If a pRB-related pathway is disturbed, a cell may follow a default pRB-independent pathway, defined by extrinsic cues from their neighbors. This could explain why the majority of human retinoblastomas show characteristics of photoreceptors (Flexner–Wintersteiner rosettes), while much evidence suggests that they originate in the INL and not in the ONL.

The timing of inactivation of *RB1* may play an important role in the development of retinoblastoma. Unlike adult mice, pRB is expressed in all retinal cells at different times during development. Therefore, loss of pRB at different developmental times could result in different retinal phenotypes. This could explain why mice expressing TAg from two different promoters necessary for photoreceptor development have different phenotypes (no tumors with the opsin promoter versus the entire photoreceptor layer developing tumors with the IRBP promoter).

7
RB Protein and Apoptosis

Among the diverse functions of pRB is the ability to protect cells from apoptosis. $RB1^{-/-}$ embryos die at E13.5 with apoptosis in the CNS, PNS, lens, and other tissues, but rescue of these mice by a normal placenta allows the animals to survive to birth, when a muscle defect is lethal. Similarly, $RB1^{-/-}$; $Id2^{-/-}$ mice show no defects in neurogenesis and hematopoiesis, but die at birth from severe reduction in muscle tissue. Id2 plays a role in regulating cell death in differentiated neurons, and may also protect against the loss of *RB1* in placenta.

Comparison of $RB1^{-/-}$; $p53^{-/-}$ double knockout mice with $RB1^{-/-}$ mice show that the CNS and lens apoptosis are p53 dependent. However, the apoptosis in the PNS was p53 independent. Cells in the CNS of the $RB1^{-/-}$; $p53^{-/-}$ mouse continue to ectopically enter S-phase but do not die. The *p53*-dependent cell death in the CNS but not the lens is due to hypoxia as a result of defective erythropoiesis. However, conditional knockout of *RB1* in the CNS, PNS, and lens in the presence of normal erythropoiesis show similar levels of apoptosis in the PNS and lens compared to the $RB1^{-/-}$ mice, suggesting that the apoptosis in these tissues is a direct result of *RB1* loss. The *NesCre1* mice with *RB1* knockout only in retina show high levels of *p53*-independent retinal apoptosis in late development.

Deregulated expression of E2F1 in fibroblasts results in S-phase entry followed by apoptosis, which is blocked by expression of high levels of pRB. Apoptosis induced following loss of pRB is dependent on the activity of E2F1, since apoptosis does not occur in the lens of the double knockout *RB1*$^{-/-}$; *E2F1*$^{-/-}$ mice. Additionally, mutant mouse embryos lacking E2F3 and *RB1* survive longer than the *RB1*$^{-/-}$ mice and display a significant reduction in apoptosis and inappropriate S-phase entry.

E2F1 induces cell death in a p53-dependent manner by increasing the stability of the p53 protein. In part, this is due to transcriptional activation of p14^{ARF} (p19^{ARF} in mouse), which promotes the rapid degradation of MDM2 in MDM2-p53 complexes, blocking the degradation of p53 by MDM2. However, p53-dependent apoptosis in the lens of the *ARF*$^{-/-}$; *RB1*$^{-/-}$ mice is only partially suppressed and p53-dependent apoptosis and inappropriate S-phase entry in the CNS are similar to that in *RB1*$^{-/-}$ mice.

E2F1 also induces transcription of the p53 homolog, *p73*, which can induce apoptosis. A dominant negative p73 mutant can suppress E2F1-induced apoptosis in cells lacking p53. Additionally, E2F1-induced apoptosis is severely impaired in *p73*$^{-/-}$ MEFs. Another transcriptional target of E2F1 is *APAF1*, which plays an important role in stress- and oncogene-induced apoptosis. Levels of *APAF1* are increased in *RB1*$^{-/-}$ mice, and apoptosis in *APAF1*$^{-/-}$ MEFs is impaired in response to increased E2F1 activity. Analysis of *RB1*$^{-/-}$; *Apaf1*$^{-/-}$ mice shows that *APAF1* is absolutely required for the apoptosis in the CNS and lens of the *RB1*$^{-/-}$ mouse, but only partially rescues the apoptosis in the lens and muscle.

The transactivation domain of E2F1 plays a role in p53-dependent apoptosis but not in p53-independent apoptosis. E2F1 may also play a role in p53-independent apoptosis. Transcriptionally impaired E2F1 mutants induce apoptosis in Saos2 cells, which lack p53 and pRB, while a DNA-binding defective mutant fails to induce apoptosis. The p53-independent induction of apoptosis by E2F1 can be inhibited by direct binding of pRB to E2F1, but is not inhibited by MDM2. Thus, induction of p53-independent and p53-dependent apoptosis by E2F1 occurs through different mechanisms: p53-dependent apoptosis may rely on relief from repression of promoters and p53-independent apoptosis may involve transcriptional activation by E2F1.

pRB may also protect cells from p53-independent apoptosis induced by ionizing radiation, transforming growth factor β1 (TGF-β1), and γ-interferon (IFN-γ). Following DNA damage, cells can arrest in both G_1/S and G_2/M phases of the cell cycle. pRB is required for the G_1/S phase arrest but not for the G_2/M phase arrest. This phenomenon is characteristic of pRB alone. The other members of the pRB family are not able to block cells in G_1/S following DNA damage.

8 Additional Genomic Changes in Retinoblastoma

8.1 Genomic Gains and Losses

Although mutation of both copies of the *RB1* gene is required for formation of retinoblastoma, all retinoblastoma tumors have additional genetic alterations. There is no evidence that any gene besides *RB1*

initiates retinoblastoma; however, other genomic changes may be required for full malignant transformation. Karyotypic analysis and comparative genomic hybridization (CGH) show genomic gains and losses in retinoblastoma tumors. The minimal regions most frequently gained are 1q31 (52%), 6p22 (44%), 2p24-25 (30%), and 13q32-34 (12%) and the minimal region most frequently lost is 16q22 (14%). Some evidence suggests that gains at 1q, 2p, and loss of 16q are restricted to more advanced tumors in older children, possibly correlating with tumor progression. Almost unique to retinoblastoma is a specific pattern of 6p gain, i(6p), identified in 60% of retinoblastoma tumors. This marker chromosome results in four copies of genes on chromosome 6p, or in low-level amplification. The gains and losses may point to potential oncogenes or tumor suppressor genes necessary for progression to malignant retinoblastoma. Identification of these cancer genes is important as they may form the basis of therapy to prevent the progression of small retinoblastoma tumors.

8.2 Candidate M3-Mn Genes

Narrowing the region of 6p gain to a minimal 0.6-Mb region of 6p22 using quantitative multiplex polymerase chain reaction (QM-PCR) for sequence-tagged sites (STS) identified RBKIN, a novel human kinesin-like gene, as a potential oncogene in retinoblastoma. RBKIN expression is increased in retinoblastoma compared to normal human retina, and inhibition of RBKIN in retinoblastoma cell lines using antisense oligonucleotides blocks cell proliferation, consistent with a role for RBKIN as an oncogene. Chromosome 6p gain is also common in bladder cancer and is associated with an elevated risk of progression.

Cadherin-11 (CDH11) was identified as a potential tumor suppressor gene using LOH and QM-PCR to define the minimal region of 16q22 genomic loss in retinoblastoma. The intact form of CDH11 is either completely lost or decreased, while the variant form is present in 50% of retinoblastoma. CDH11 is a cell adhesion molecule whose loss may promote progression of some $RB1^{-/-}$ retinal cells to malignancy. Unusually low levels of the intact form of CDH11 are present in osteosarcoma, while the variant form was highly expressed in invasive breast cancer and was associated with promotion of invasiveness.

8.3 Causes of Drug Resistance in Retinoblastoma

Ling and coworkers first described the multidrug resistance (MDR) phenotype in human cancer, due to upregulation of the *MDR1* gene with increased expression of P-glycoprotein. P-glycoprotein expression is increased in response to environmental toxicity in many organisms, and broadly reduces intracellular levels of vinca alkaloids (vincristine, vinblastine), epipodophyllotoxins (etoposide, teniposide), and other natural-product antineoplastic agents by functioning as an ATP-dependent plasma membrane pump that expels drugs from human cancer cells. Platinum compounds are not substrates of the P-glycoprotein drug efflux mechanism.

Expression of *MDR1* in retinoblastoma at initial diagnosis and increased expression at relapse after failed treatment may account for frequent chemotherapy failures. It is not possible to correlate P-glycoprotein levels before therapy with the

outcome of chemotherapy in intraocular retinoblastoma, since biopsies incur risk of systemic spread. In metastatic retinoblastoma, P-glycoprotein expression before therapy correlates with failure of therapy, and initially undetectable P-glycoprotein correlates with long-term remission.

In order to save eyes without radiation, which increases the lifelong risk of secondary malignancies and other long-term side effects, chemotherapy has become the primary therapy for intraocular retinoblastoma in many centers worldwide. One protocol with excellent success to cure eyes without use of radiation or removal of the eye includes a direct approach to counter P-glycoprotein MDR. High-dose cyclosporine A (CSA) is delivered simultaneously with the chemotherapy to block the ability of P-glycoprotein to protect tumor cells from chemotherapy. An international Phase II Clinical Trial of CSA-modified chemotherapy is underway.

Intraocular retinoblastoma that failed chemotherapy without CSA showed increased P-glycoprotein when examined by immunohistochemistry. Retinoblastoma that failed chemotherapy despite administration of CSA did not express P-glycoprotein but expressed instead the Multidrug Resistance Protein (MRP). MRP belongs to the same ATP-dependent membrane transporter superfamily as P-glycoprotein, and transfection of the MRP gene confers a similar broad-spectrum pattern of drug resistance to antineoplastic agents. Although CSA inhibits P-glycoprotein *in vitro*, there is presently no effective inhibitor of MRP.

Several chemotherapy drugs induce apoptosis in $RB1^{-/-}$ cells and growth arrest in $RB1^{+/-}$ cells. It is proposed therefore that continued expression of pRB in some tumors may contribute to drug resistance. These drugs might selectively destroy $RB1^{-/-}$ tumors, while pRB-expressing adjacent normal cells might be protected by pRB.

9 Importance of Retinoblastoma

The rare eye cancer, retinoblastoma, continues to contribute much more than its "share" to understanding of cancer, biological processes, and use of scientific knowledge to optimize health outcomes. *RB1* is the first tumor suppressor gene discovered and revealed the process (search for loss of heterozygosity) that was used to discover other genes underlying human cancer. The RB gene family members are key cell cycle regulators, so that their mutation alters development and deregulates the cell cycle to induce cancer in susceptible cells. The many ways that *RB1* can be damaged and the high mutation rate have forced creative technology development to meet the challenges of clinical diagnostics, much of which has been widely generalized to other genetic disease challenges. At a health service level, it is clearly shown that molecular determination of individual *RB1* mutant alleles contributes importantly to health quality at less cost.

Analysis of genetic changes in retinoblastoma subsequent to loss of both *RB1* alleles has revealed several potential novel genes involved in tumor progression and resistance to therapy, which will clearly become the targets for potential prevention and therapeutic strategies. Recognition of the key role of the *MDR* gene in retinoblastoma provided the rational basis for a therapeutic strategy to counter drug resistance that appears to improve the cure rate of intraocular retinoblastoma and

that is currently the basis of an international multicenter clinical trial. Although basic studies clearly show an important role of *MDR* in many cancers, only in retinoblastoma has a therapeutic strategy to alter *MDR* effect shown positive clinical outcome.

Much more is yet to come. The focus on defining the unique features of the exact cell in developing human retina that is susceptible to cancer by loss of pRB is key to understanding ''cancer stem cells.'' Retinoblastoma is one of the best examples of human disease indicating that only rare cells at a specific developmental and proliferative stage may be at risk to initiate tumors. And the rarity of retinoblastoma has been turned again to advantage by the world participation of major retinoblastoma treatment centers in an Internet survey to validate a new clinical classification, showing that global cooperation can reach out to future optimal opportunities for all families afflicted with this rare, unique cancer.

See also Cancer Chemotherapy, Theoretical Foundations of; Epigenetic Mechanisms in Tumorigenesis; Intracellular Signaling in Cancer; Growth Factors and Oncogenes in Gastrointestinal Cancers; Oncology, Molecular; Oncogenes.

Bibliography

Books and Reviews

Brown, V.D., Chan, H.S.L., Gallie, B.L. (2003) A Cascade of Genomic Changes Leads to Retinoblastoma, in: Albert, D., Polans, A. (Eds.) *Ocular Oncology*, Marcel Dekker Inc., New York and Basel, pp. 103–128.

Classon, M., Dyson, N. (2001) p107 and p130: Versatile proteins with interesting pockets, *Exp. Cell Res.* **264**, 135–147.

Classon, M., Harlow, E. (2002) The retinoblastoma tumor suppressor in development and cancer, *Nat. Rev. Cancer* **2**, 910–917.

DiCiommo, D., Gallie, B.L., Bremner, R. (2000) Retinoblastoma: the disease, gene and protein provide critical leads to understand cancer, *Semin. Cancer Biol.* **10**, 255–269.

Dyson, N. (1998) The regulation of E2F by pRB-family proteins, *Genes Dev.* **12**, 2245–2262.

Kaelin, W.G. Jr. (1999) Functions of the retinoblastoma protein, *BioEssays* **21**, 950–958.

Mittnacht, S. (1998) Control of pRB phosphorylation, *Curr. Opin. Genet. Dev.* **8**, 21–27.

Morris, E.J., Dyson, N.J. (2001) Retinoblastoma protein partners, *Adv. Cancer Res.* **82**, 1–54.

Weinberg, R.A. (1995) The retinoblastoma protein and cell cycle control, *Cell* **81**, 323–330.

Zacksenhaus, E., Bremner, R., Jiang, Z., Gill, R.M., Muncaster, M., Sopta, M., Phillips, R.A., Gallie, B.L. (1993) Unraveling the function of the retinoblastoma gene, *Adv. Cancer Res.* **61**, 115–141.

Primary Literature

Akiyama, T., Ohuchi, T., Sumida, S., Matsumoto, K., Toyoshima, K. (1992) Phosphorylation of the retinoblastoma protein by cdk2, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7900–7904.

al-Ubaidi, M.R., Font, R.L., Quiambao, A.B., Keener, M.J., Liou, G.I., Overbeek, P.A., Baehr, W. (1992) Bilateral retinal and brain tumors in transgenic mice expressing simian virus 40 large T antigen under control of the human interphotoreceptor retinoid-binding protein promoter, *J. Cell Biol.* **119**, 1681–1687.

Bagchi, S., Weinmann, R., Raychaudhuri, P. (1991) The retinoblastoma protein copurifies with E2F-1, an E1A-regulated inhibitor of the transcription factor E2F, *Cell* **65**, 1063–1070.

Bandara, L.R., Thangue, N.B.L. (1991) Adenovirus E1a prevents the retinoblastoma gene product from complexing with a cellular transcription factor, *Nature* **351**, 494–497.

Buchkovich, K., Duffy, L.A., Harlow, E. (1989) The retinoblastoma protein is phosphorylated during specific phases of the cell cycle, *Cell* **58**, 1097–1105.

Chen, D., Gallie, B.L., Squire, J.A. (2001) Minimal regions of chromosomal imbalance in retinoblastoma detected by comparative

genomic hybridization, *Cancer Genet. Cytogenet.* **129**, 57–63.

Chen, D., Livne-bar, I., Vanderluit, J.L., Slack, R.S., Agochiya, M., Bremner, R. (2004) Cell-specific effects of RB or RB/p107 loss on retinal development implicate an intrinsically death-resistant cell-of-origin in retinoblastoma, *Cancer Cell* **5**, 539–551.

Chen, D., Pajovic, S., Duckett, A., Brown, V.D., Squire, J.A., Gallie, B.L. (2002) Genomic amplification in retinoblastoma narrowed to 0.6 Mb on chromosome 6p containing a kinesin-like gene, RBKIN, *Cancer Res.* **63**, 967–971.

Chellappan, S., Hiebert, S., Mudryi, M., Horowitz, J.M., Nevins, J.R. (1991) The E2F transcription factor is a cellular target for the RB protein, *Cell* **65**, 1053–1061.

Chellappan, S., Kraus, V.B., Kroger, B., Munger, K., Howley, P.M., Phelps, W.C., Nevins, J.R. (1992) Adenovirus E1A, simian virus 40 tumor antigen, and human papillomavirus E7 protein share the capacity to disrupt the interaction between transcription factor E2F and the retinoblastoma gene product, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4549–4553.

Clarke, A.R., Maandag, E.R., vanRoon, M., vanderLugt, N.M.T., vanderValk, M., Hooper, M.L., Berns, A., teRiele, H. (1992) Requirement for a functional Rb-1 gene in murine development, *Science* **359**, 328–330.

Cobrinik, D., Lee, M.-H., Hannon, G., Mulligan, G., Bronson, R.T., Dyson, N., Harlow, E., Beach, D., Weinberg, R.A., Jacks, T. (1996) Shared role of the pRB-related p130 and p107 proteins in limb development, *Genes Dev.* **1996**, 1633–1644.

Comings, D.E. (1973) A general theory of carcinogenesis, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3324–3328.

de Bruin, A., Wu, L., Saavedra, H.I., Wilson, P., Yang, Y., Rosol, T.J., Weinstein, M., Robinson, M.L., Leone, G. (2003) *Rb* function in extraembryonic lineages suppresses apoptosis in the CNS of *Rb*-deficient mice, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6546–6551.

DeCaprio, J.A., Fusuke, Y., Ajchenbaum, F., Griffin, J.D., Livingston, D.M. (1992) The retinoblastoma-susceptibility gene product becomes phosphorylated in multiple stages during cell cycle entry and progression, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1795–1798.

DeCaprio, J.A., Ludlow, J.W., Lynch, D., Furukawa, Y., Griffin, J., Piwnica-Worms, H., Huang, C.-M., Livingston, D.M. (1989) The product of the retinoblastoma susceptibility gene has properties of a cell cycle regulatory element, *Cell* **58**, 1085–1095.

DeCaprio, J.A., Ludlow, J.W., Figge, J., Shew, J.-Y., Huang, C.-M., Lee, W.-H., Marsilio, E., Paucha, E., Livingstone, D.M. (1988) SV40 large tumor antigen forms a specific complex with the product of the retinoblastoma susceptibility gene, *Cell* **54**, 275–283.

DeGregori, J., Kowalik, T., Nevins, J.R. (1995) Cellular targets for activation by the E2F1 transcription factor include DNA synthesis- and G_1/S-regulatory genes, *Mol. Cell. Biol.* **15**, 4215–4224.

Dowdy, S.F., Hinds, P.W., Louie, K., Reed, S.I., Arnold, A., Weinberg, R.A. (1993) Physical interaction of the retinoblastoma protein with human D cyclins, *Cell* **73**, 499–511.

Dunn, J.M., Phillips, R.A., Zhu, X., Becker, A., Gallie, B.L. (1989) Mutations in the RB1 gene and their effects on transcription, *Mol. Cell. Biol.* **9**, 4596–4604.

Durfee, T., Becherer, K., Chen, P.-L., Yeh, S.-H., Yang, Y., Kilburn, A.E., Lee, W.-H., Elledge, S.J. (1993) The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit, *Genes Dev.* **7**, 555–569.

Dyson, N., Howley, P.M., Munger, K., Harlow, E. (1989) The human papilloma virus-16 E7 oncoprotein is able to bind the retinoblastoma gene product, *Science* **243**, 934–937.

Eagle, Jr., R.C., Shields, J.A., Donoso, L., Milner, R.S. (1989) Malignant transformation of spontaneously regressed retinoblastoma, retinoma/retinocytoma variant, *Ophthalmology* **96**, 1389–1395.

Ewen, M., Xing, Y., Lawrence, J.B., Livingston, D.M. (1991) Molecular cloning, chromosomal mapping, and expression of p107, a retinoblastoma gene product-related protein, *Cell* **66**, 1155–1164.

Ewen, M.E., Sluss, H.K., Sherr, C.J., Matsushime, H., Kato, J.-Y., Livingston, D.M. (1993) Functional interactions of the retinoblastoma protein with mammalian D-type cyclins, *Cell* **73**, 487–497.

Field, S.J., Tsai, F.-Y., Kuo, F., Zubiaga, A.M., Kaelin, W.G., Livingston, D.M., Orkin, S.H., Greenberg, M.E. (1996) E2F-1 functions in mice to promote apoptosis and suppress proliferation, *Cell* **85**, 549–561.

Flemington, E.K., Speck, S.H., Kaelin, W.G. (1993) E2F-1-mediated transactivation is inhibited by complex formation with the retinoblastoma susceptibility gene product, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6914–6918.

Friend, S.H., Bernards, R., Rogelj, S., Weinberg, R.A., Rapaport, J.M., Albert, D.M., Dryja, T.P. (1986) A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma, *Nature* **323**, 643–646.

Gallie, B.L., Ellsworth, R.M., Abramson, D.H., Phillips, R.A. (1982) Retinoma: spontaneous regression of retinoblastoma or benign manifestation of the mutation? *Br. J. Cancer* **45**, 513–521.

Gallie, B.L., Phillips, R.A., Ellsworth, R.M., Abramson, D.H. (1982) Significance of retinoma and phthisis bulbi for retinoblastoma, *Ophthalmology* **89**, 1393–1399.

Gallie, B.L., Campbell, C., Devlin, H., Duckett, A., Squire, J.A. (1999) Developmental basis of retinal-specific induction of cancer by RB mutation, *Cancer Res.* **59**, 1731s–1735s.

Goodrich, D.W., Wang, N.P., Qian, Y., Lee, E.Y., Lee, W.-H. (1991) The retinoblastoma gene product regulates progression through the G1 phase of the cell cycle, *Cell* **67**, 293–302.

Haas-Kogan, D.A., Kogan, S.C., Levi, D., Dazin, P., T'Ang, A., Fung, Y.-K.T., Israel, M.A. (1995) Inhibition of apoptosis by the retinoblastoma gene product, *EMBO J.* **14**, 461–472.

Hamel, P., Gill, R.M., Phillips, R.A., Gallie, B.L. (1992) Transcriptional repression of the E2-containing promoters EIIaE, c-myc, and RB1 by the product of the RB1 gene, *Mol. Cell. Biol.* **12**, 3431–3438.

Hannon, G.J., Demetrick, D., Beach, D. (1993) Isolation of the Rb-related p130 through its interaction with cdk2 and cyclins, *Genes Dev.* **7**, 2378–2391.

Hatakeyama, M., Brill, J.A., Fink, G.R., Weinberg, R.A. (1994) Collaboration of G1 cyclins in the functional inactivation of the retinoblastoma protein, *Genes Dev.* **8**, 1759–1771.

Helin, K., Harlow, E., Fattaey, A. (1993) Inhibition of E2F-1 transactivation by direct binding of the retinoblastoma protein, *Mol. Cell. Biol.* **13**, 6501–6508.

Hiebert, S.W. (1993) Regions of the retinoblastoma gene product required for its interaction with the E2F transcription factor are necessary for E2 promoter repression and pRB-mediated growth suppression, *Mol. Cell. Biol.* **13**, 3384–3391.

Hiebert, S.W., Chellappan, S.P., Horowitz, J.M., Nevins, J.R. (1992) The interaction of RB with E2F coincides with an inhibition of the transcriptional activity of E2F, *Genes Dev.* **6**, 177–185.

Hinds, P.W., Mittnacht, S., Dulic, V., Arnold, A., Reed, S.I., Weinberg, R.A. (1992) Regulation of retinoblastoma protein functions by ectopic expression of human cyclins, *Cell* **70**, 993–1006.

Horton, L.E., Qian, Y., Templeton, D.J. (1995) G1 cyclins control the retinoblastoma gene product growth regulation activity via upstream mechanisms, *Cell Growth Differ.* **6**, 395–407.

Howes, K.A., Lasudry, J.G., Albert, D.M., Windle, J.J. (1994) Photoreceptor cell tumors in transgenic mice, *Invest. Ophthalmol. Vis. Sci.* **35**, 342–351.

Hsieh, J.-K., Fredersdorf, S., Kouzarides, T., Martin, K., Lu, X. (1997) E2F1-induced apoptosis requires DNA binding but not transactivation and is inhibited by the retinoblastoma protein through direct interaction, *Genes Dev.* **11**, 1840–1852.

Hu, Q., Dyson, N., Harlow, E. (1990) The regions of the retinoblastoma protein needed for binding to adenovirus E1A or SV40 large T antigen are common sites for mutation, *EMBO J.* **9**, 1147–1155.

Hu, Q., Lees, J.A., Buchkovich, K.J., Harlow, E. (1992) The retinoblastoma protein physically associates with the human cdc2 kinase, *Mol. Cell. Biol.* **12**, 971–980.

Huang, H.-J.S., Yee, J.-K., Shrew, J.-Y., Chen, P.-L., Bookstein, R., Friedmann, T., Lee, E.Y.-H.P., Lee, W.-H. (1988) Suppression of the neoplastic phenotype by replacement of the RB gene in human cancer cells, *Science* **242**, 1563–1566.

Jacks, T., Fazeli, A., Schmitt, E.M., Bronson, R.T., Goodell, M.A., Weinberg, R.A. (1992) Effects of an Rb mutation in the mouse, *Science* **359**, 295–300.

Jiang, Z., Zacksenhaus, E., Gallie, B.L., Phillips, R.A. (1997) The retinoblastoma gene family is differentially expressed during embryogenesis, *Oncogene* **14**, 1789–1797.

Johnson, D.G., Ohtani, K., Nevins, J.R. (1994) Autoregulatory control of E2F1 expression in

response to positive and negative regulators of cell cycle progression, *Genes Dev.* **8**, 1514–1525.

Johnson, D.G., Schwarz, J.K., Cress, W.D., Nevins, J.R. (1993) Expression of transcription factor E2F1 induces quiescent cells to enter S phase, *Nature* **365**, 349–352.

Kaelin, W.G., Pallas, D.C., DeCaprio, J.A., Kaye, F.J., Livingston, D.M. (1991) Identification of cellular proteins that can interact specifically with the T/E1A-binding region of the retinoblastoma gene product, *Cell* **64**, 521–532.

Kato, J.-Y., Matsushime, H., Hiebert, S.W., Ewen, M.E., Sherr, C.J. (1993) Direct binding of cyclin D to the retinoblastoma gene product (pRb) and pRb phosphorylation by the cyclin D-dependent kinase CDK4, *Genes Dev.* **7**, 331–342.

Knudson, J.A.G. (1971) Mutation and cancer: statistical study of retinoblastoma, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 820–823.

Krek, W., Livingston, D.M., Shirodkar, S. (1993) Binding to DNA and the retinoblastoma gene product promoted by complex formation of different E2F family members, *Science* **262**, 1557–1560.

LeCouter, J.E., Kablar, B., Whyte, P.F., Ying, C., Rudnicki, M.A. (1998) Strain-dependent embryonic lethality in mice lacking the retinoblastoma-related p130 gene, *Development* **125**, 4669–4679.

LeCouter, J.E., Kablar, B., Hardy, W.R., Ying, C., Megeney, L.A., May, L.L., Rudnicki, M.A. (1998) Strain-dependent myeloid hyperplasia, growth deficiency, and accelerated cell cycle in mice lacking the Rb-related *p107* gene, *Mol. Cell. Biol.* **18**, 7455–7465.

Lee, E.Y.-H.P., Chang, C.-Y., Hu, N., Wang, Y.-C.J., Lai, C.-C., Herrup, K., Lee, W.-H., Bradley, A. (1992) Mice deficient for Rb are nonviable and show defects in neurogenesis and haematopoiesis, *Science* **359**, 288–294.

Lee, J.-O., Russo, A.A., Pavletich, N.P. (1998) Structure of the retinoblastoma tumor-suppressor pocket domain bound to a peptide from HPV E7, *Nature* **391**, 859–865.

Lee, M.-H., Williams, B.O., Mulligan, G., Mukai, S., Bronson, R.T., Dyson, N., Harlow, E., Jacks, T. (1996) Targeted disruption of *p107*: functional overlap between *p107* and *Rb*, *Genes Dev.* **10**, 1621–1632.

Lees, J.A., Buchkovich, K.J., Marshak, D.R., Anderson, C.W., Harlow, E. (1991) The retinoblastoma protein is phosphorylated on multiple sites by human cdc2, *EMBO J.* **10**, 4279–4290.

Li, Y., Graham, C., Lacy, S., Duncan, A.M.V., Whyte, P. (1993) The adenovirus E1a-associated 130-kD protein is encoded by a member of the retinoblastoma gene family and physically interacts with cyclins A and E, *Genes Dev.* **7**, 2366–2377.

Lin, B.T.-Y., Gruenwald, S., Morla, A.O., Lee, W.-H., Wang, J.Y.J. (1991) Retinoblastoma cancer suppressor gene product is a substrate of the cell cycle regulator cdc2 kinase, *EMBO J.* **10**, 857–864.

Lipinski, M.M., Macleod, K.F., Williams, B.O., Mullaney, T.L., Crowley, D., Jacks, T. (2001) Cell-autonomous and non-cell-autonomous functions of the Rb tumor suppressor in developing central nervous system, *EMBO J.* **20**, 3402–3413.

Lohmann, D.R. (1999) RB1 gene mutations in retinoblastoma, *Hum. Mutat.* **14**, 283–288.

Lohmann, D.R., Gallie, B.L. (2004) Retinoblastoma: Revisiting the model prototype of inherited cancer, *Am. J. Med. Genet.* **129C**, 23–28.

Ludlow, J.W., Glendening, C.L., Livingston, D.M., DeCaprio, J.A. (1993) Specific enzymatic dephosphorylation of the retinoblastoma protein, *Mol. Cell. Biol.* **13**, 367–372.

Ludlow, J.W., Shon, J., Pipas, J.M., Livingston, D.M., DeCaprio, J.A. (1990) The retinoblastoma susceptibility gene product undergoes cell cycle-dependent dephosphorylation and binding to and release from SV40 large T, *Cell* **60**, 387–396.

Ludlow, J.W., DeCaprio, J.A., Huang, C., Lee, W.-H., Paucha, E., Livingston, D.M. (1989) SV40 large T antigen binds preferentially to an underphosphorylated member of the retinoblastoma susceptibility gene product family, *Cell* **56**, 57–65.

Lundberg, A.S., Weinberg, R.A. (1998) Functional inactivation of the retinoblastoma protein requires sequential modification by at least two distinct cyclin-cdk complexes, *Mol. Cell. Biol.* **18**, 753–761.

Maandag, E.C., vanderValk, M., Vlaar, M., Feltkamp, C., O'Brien, J., vanRoon, M., vanderLugt, N., Berns, A., teRiele, H. (1994) Developmental rescue of an embryonic-lethal mutation in the retinoblastoma gene in chimeric mice, *EMBO J.* **13**, 4260–4268.

Macleod, K.F., Hu, Y., Jacks, T. (1996) Loss of *Rb* activates *p53*-dependent and independent

cell death pathways in the developing mouse nervous system, *EMBO J.* **15**, 6178–6188.

MacPherson, D., Sage, J., Crowley, D., Trumpp, A., Bronson, R.T., Jacks, T. (2003) Conditional mutation of *Rb* causes cell cycle defects without apoptosis in the central nervous system, *Mol. Cell. Biol.* **23**, 1044–1053.

MacPherson, D., Sage, J., Kim, T., Ho, D., McLaughlin, M.E., Jacks, T. (2004) Cell type-specific effects of *Rb* deletion in the murine retina, *Genes Dev.* **18**, 1681–1694.

Marchong, M.N., Chen, D., Corson, T.W., Lee, C., Harmandayan, M., Bowles, E., Chen, N., Gallie, B.L. (2004) Minimal 16q genomic loss implicates CDH11 in retinoblastoma, *Mol. Cancer Res.* **2**, 495–503.

Mittnacht, S., Lees, J.A., Desai, D., Harlow, E., Morgan, D.O., Weinberg, R.A. (1994) Distinct sub-populations of the retinoblastoma protein show a distinct pattern of phosphorylation, *EMBO J.* **13**, 118–127.

Morgenbesser, S.D., Williams, B.O., Jacks, T., DePinho, R.A. (1994) p53-dependent apoptosis produced by Rb-deficiency in the developing mouse lens, *Nature* **371**, 72–74.

Peeper, D.S., Upton, T.M., Ladha, M.H., Neuman, E., Zalvide, J., Bernards, R., DeCaprio, J.A., Ewen, M.E. (1997) Ras signalling linked to the cell-cycle machinery by the retinoblastoma protein, *Nature* **386**, 177–181.

Qian, Y., Luckey, C., Horton, L., Esser, M., Templeton, D.J. (1992) Biological function of the retinoblastoma protein requires distinct domains for hyperphosphorylation and transcription factor binding, *Mol. Cell. Biol.* **12**, 5363–5372.

Qin, X.-Q., Chittenden, T., Livingston, D.M., Kaelin, W.G. (1992) Identification of a growth suppression domain within the retinoblastoma gene product, *Genes Dev.* **6**, 953–964.

Qin, X.-Q., Livingston, D.M., Kaelin, W.G., Adams, P.D. (1994) Deregulated transcription factor E2F-1 expression leads to S-phase entry and p53-mediated apoptosis, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10918–10922.

Richter, S., Vandezande, K.N.C., Zhang, K., Sutherland, J., Anderson, J., Han, L., Panton, R., Branco, P., Gallie, B. (2003) Sensitive and efficient detection of RB1 gene mutations enhances care for families with retinoblastoma, *Am. J. Hum. Genet.* **72**, 253–269.

Robanus-Maandag, E., Dekker, M., vanderValk, M., Carrozza, M.L., Jeanny, J.-C., Dannenberg, J.-H., Berns, A., teRiele, H. (1998) p107 is a suppressor of retinoblastoma development in pRb-deficient mice, *Genes Dev.* **12**, 1599–1609.

Sellers, W.R., Novitch, B.G., Miyake, S., Heith, A., Otterson, G.A., Kaye, F.J., Lassar, A.B., Kaelin, W.G. (1998) Stable binding to E2F is not required for the retinoblastoma protein to activate transcription, promote differentiation, and suppress tumor cell growth, *Genes Dev.* **12**, 95–106.

Shan, B., Lee, W.-H. (1994) Deregulated expression of E2F-1 induces S-phase entry and leads to apoptosis, *Mol. Cell. Biol.* **14**, 8166–8173.

Shan, B., Durfee, T., Lee, W.-H. (1996) Disruption of RB/E2F-1 interaction by single point mutations in E2F-1 enhances S-phase entry and apoptosis, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 679–684.

Squire, J.A., Gallie, B.L., Phillips, R.A. (1985) A detailed analysis of chromosomal changes in heritable and non-heritable retinoblastoma, *Hum. Genet.* **70**, 291–301.

Templeton, D.J., Park, S.H., Lanier, L., Weinberg, R.A. (1991) Nonfunctional mutants of the retinoblastoma protein are characterized by defects in phosphorylation, viral oncoprotein association, and nuclear tethering, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3033–3037.

Tsai, K.Y., Hu, Y., Macleod, K.F., Crowley, D., Yamasaki, L., Jacks, T. (1998) Mutation of *E2f-1* suppresses apoptosis and inappropriate S phase entry and extends survival of *Rb*-deficient mouse embryos, *Mol. Cell.* **2**, 293–304.

Vooijs, M., Berns, A. (1999) Developmental defects and tumor predisposition in *Rb* mutant mice, *Oncogene* **18**, 5293–5303.

Weintraub, S.J., Prater, C.A., Dean, D.C. (1992) Retinoblastoma protein switches the E2F site from positive to negative element, *Nature* **358**, 259–261.

Whyte, P., Buchkovich, K.J., Horowitz, J.M., Friend, S.H., Raybuck, M., Weinberg, R.A., Harlow, E. (1988) Association between an oncogene and an anti-oncogene: the adenovirus E1A proteins bind to the retinoblastoma gene product, *Nature* **334**, 124–129.

Williams, B.O., Remington, L., Albert, D.M., Mukai, S., Bronson, R.T., Jacks, T. (1994) Cooperative tumorigenic effects of germline mutations in *Rb* and p53, *Nat. Genet.* **7**, 480–484.

Williams, B.O., Schmitt, E.M., Remington, L., Bronson, R.T., Albert, D.M., Weinberg, R.A., Jacks, T. (1994) Extensive contribution of *Rb*-deficient cells to adult chimeric mice with limited histopathological consequences, *EMBO J.* **13**, 4251–4259.

Windle, J.J., Albert, D.M., O'Brien, J.M., Marcus, D.M., Disteche, C.M., Bernards, R., Mellon, P.L. (1990) Retinoblastoma in transgenic mice, *Nature* **343**, 665–669.

Wu, L., Bruin, Ad., Saavedra, H.I., Starovic, M., Trimboli, A., Yang, Y., Opavska, J., Wilson, P., Thompson, J.C., Ostrowski, M.C., Rosol, T.J., Woollett, L.A., Weinstein, M., Cross, J.C., Robinson, M.L., Leone, G. (2003) Extra-embryonic function of Rb is essential for embryonic development and viability, *Nature* **421**, 942–947.

Yamasaki, L., Bronson, R., Williams, B.O., Dyson, N.J., Harlow, E., Jacks, T. (1998) Loss of E2F-1 reduces tumorigenesis and extends the lifespan of *Rb1*(+/−) mice, *Nat. Genet* **18**, 360–364.

Yamasaki, L., Jacks, T., Bronson, R., Goillot, E., Harlow, E., Dyson, N.J. (1996) Tumor induction and tissue atrophy in mice lacking E2F-1, *Cell* **85**, 537–548.

Zacksenhaus, E., Jiang, Z., Chung, D., Marth, J.D., Phillips, R.A., Gallie, B.L. (1996) pRb controls proliferation, differentiation, and death of skeletal muscle cells and other lineages during embryogenesis, *Genes Dev.* **10**, 3051–3064.

Zarkowska, T., Mittnacht, S. (1997) Differential phosphorylation of the retinoblastoma protein by G1/S cyclin dependent kinases, *J. Biol. Chem.* **272**, 12738–12746.

Zhang, J., Gray, J., Wu, L., Leone, G., Rowan, S., Cepko, C.L., Zhu, X., Craft, C.M., Dyer, M.A. (2004) Rb regulates proliferation and rod photoreceptor development in mouse retina, *Nat. Genet.* **36**, 351–360.

Zhu, L., vanderHeuvel, S., Helin, K., Fattaey, A., Ewen, M., Livingston, D., Dyson, N., Harlow, E. (1993) Inhibition of cell proliferation by p107, a relative of the retinoblastoma protein, *Genes Dev.* **7**, 1111–1125.

Retinoids and the Chemical Biology of the Visual Cycle

Robert R. Rando
Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA

1 Introduction 376

2 Overview of Isomerization Pathway 378

3 Individual Components of the Three-component System 382
3.1 Lecithin Retinol Acyltransferase 382
3.1.1 C161, H60, and Y154 as a Catalytic Triad in LRAT Action 382
3.2 Molecular Function of RPE65 385
3.3 The Isomerization Reaction 388
3.3.1 tREs as IMH Substrates in Rod- and Cone-dominated Species 388
3.3.2 A Specific Inactivator of IMH and an Approach to IMH Identification 390

4 Possible Medical Significance of Studies on Retinoid Processing in the Visual System 391

Acknowledgment 393

Bibliography 393
Books and Reviews 393
Primary Literature 394

Keywords

Photoreceptor
The highly specialized retinal cells in which light perception and phototransduction occur.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

Phototransduction
The sequence of biochemical reactions that are initiated by the photoisomerization of rhodopsin and lead to the hyperpolarization of photoreceptor membranes.

Retinal Pigment Epithelial Cells (RPE)
Epithelial cells that are anatomically placed behind the photoreceptors in the eye. Virtually all of the biochemical reactions in the visual cycle occur in the RPE.

Retinoids
Diterpene polyene derivative in the vitamin A family that includes all-*trans*-retinol (vitamin A), all-*trans*-retinal, all-*trans*-retinyl esters and their 11-*cis*-congeners.

Visual Cycle
The biochemical pathway that begins with the photoisomerization of the 11-*cis*-retinal Schiff base chromophore of rhodopsin to all-*trans*-retinal and ends with the resynthesis of 11-*cis*-retinal.

In the 1930s, Wald discovered that the photoconversion of 11-*cis*-retinal to all-*trans*-retinal is the primary event in vision. Vision would be a once in a lifetime event if the 11-*cis*-retinal chromophore of rhodopsin were not resynthesized. In this review, the biosynthetic pathway through which vitamin A is transformed into the visual chromophore is described.

1 Introduction

Light perception in all sighted animals depends on a photoisomerization event of a derivative of the retinoid vitamin A. The photoisomerization reaction in question is the *cis* → *trans* isomerization of the 11-*cis*-retinal Schiff base (the chromophore) of rhodopsin into its all-*trans* congener (Fig. 1). This simple double-bond isomerization reaction in the retinoid polyene underlies human vision. As vision proceeds, the all-*trans*-retinal photoisomerization product must be enzymatically reisomerized into 11-*cis*-retinal in order for vision to continue. This turns out to be by far the most intricate and complicated chemical process found in all of retinoid biology. The chemical process through which an all-*trans*-retinoid is isomerized into an 11-*cis*-retinoid is the subject of this review.

The vertebrate visual cycle is comprised of those biochemical reactions that begin with the photoisomerization/hydrolysis of the 11-*cis*-retinal Schiff base chromophore of rhodopsin in the photoreceptors to all-*trans*-retinal, and ends with the processing of all-*trans*-retinyl esters (tREs) in the retinal pigment epithelium (RPE) to form 11-*cis*-retinol (al) (cROL(AL))(Fig. 1). The bleaching of rhodopsin to initiate

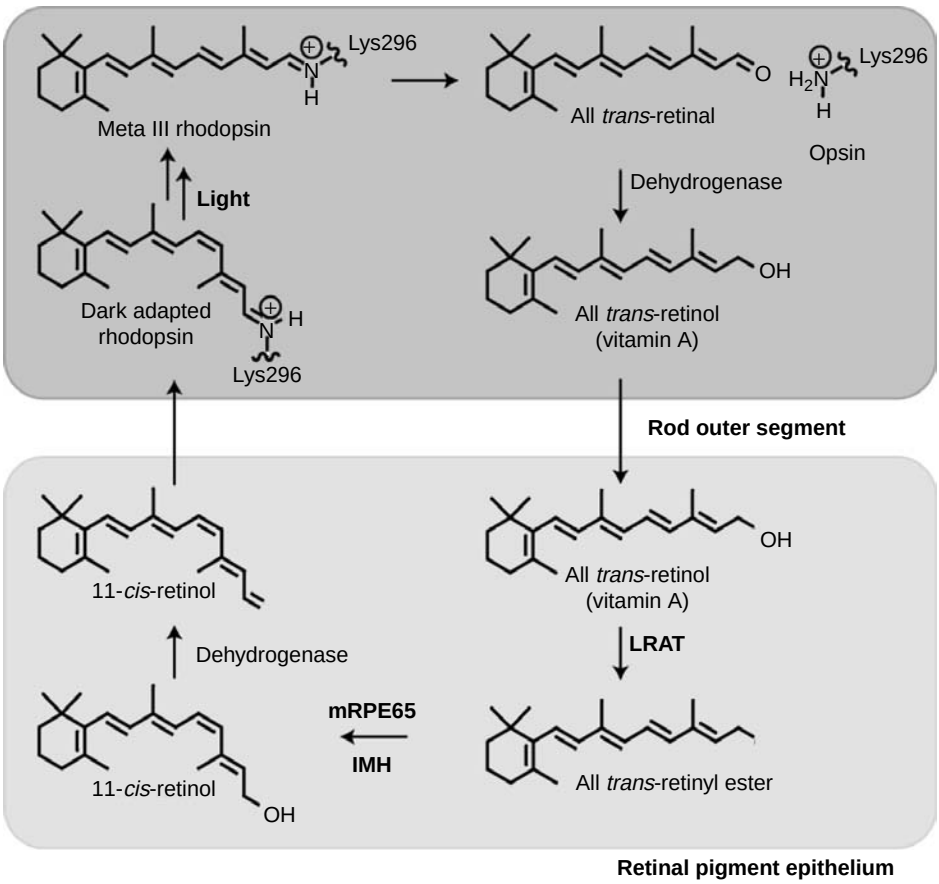

Fig. 1 The visual cycle. The critical retinoid isomerization process is carried out by a three-component system comprised of LRAT, mRPE65, and IMH, which is located in the retinal pigment epithelium.

the phototransduction cascade has been intensively studied over the years. The key event here is the photochemical *cis–trans* photoisomerization, leading to the functionally active conformer(s) of rhodopsin (Meta II). Meta II initiates the cascade of biochemical events leading to photoreceptor hyperpolarization.

Since the photoisomerization/hydrolysis of the 11-*cis*-retinal Schiff base chromophore to all-*trans*-retinal is irreversible at low levels of illumination, biochemical processes must be in place to resynthesize 11-*cis*-retinal in the dark from an all-*trans*-congener. A limitation on any mechanism postulated for regeneration is the fact that 11-*cis*-retinoids are thermodynamically unstable, accounting for only one part per thousand in an equilibrium mixture of retinoids. Thus, an energy source must be posited for the enzymatic *trans-cis* isomerization required in the biosynthesis of 11-*cis*-retinol. This thermodynamic constraint is central to any mechanism forwarded concerning the operation of the visual cycle.

2
Overview of Isomerization Pathway

A central issue in the operation of the visual cycle to address is the nature of the isomerization substrate. Studies of this type required an *in vitro* system capable of producing 11-*cis*-retinoids from all-*trans*-retinoids. The first *in vitro* system able to enzymatically process an all-*trans*-retinoid to an 11-*cis*-retinoid was established in 1987. It was shown that all-*trans*-retinol (vitamin A) (tROL) added to the retinal pigment epithelial cells (RPE) is converted into tREs (chiefly palmitate) and to cROL (al) and esters. The studies established that the RPE contained the biochemical machinery required for the processing of tROL into 11-*cis*-retinoids. In rod-dominated eyes, the only biochemical reaction known to occur in the photoreceptors is the reduction of all-*trans*-retinal (tRAL) into tROL. After the tROL is transported across the interphotoreceptor space, probably by the binding protein interphotoreceptor retinol-binding protein (IRBP), the remaining enzymatic processing to cRAL occurs in the RPE. The visual chromophore, cRAL, is returned to the photoreceptor, again probably bound to IRBP, where it is able to regenerate rhodopsin from opsin. It should be noted that all discussions here are concerned with rod regeneration, although it is clear that cone light independent regeneration obeys the same molecular logic found in rod regeneration.

Since all-*trans* and 11-*cis*-retinoids exist as retinols, retinals, and retinyl esters, it is not immediately obvious what the substrate and product of isomerization are; nine possible routes formally exist. To simplify analysis, double labeling experiments with 15-^{3}H, 15-^{14}C-tROL were performed. If isomerization occurred through the retinal as an intermediate, then the product of isomerization should lose one-half of its ^{3}H. The results are clear-cut and showed minimal loss of ^{3}H during isomerization, demonstrating that isomerization occurs at the alcohol level of oxidation. This leaves four possible isomerization routes, with retinyl esters and alcohols being possible isomerization substrates and products.

Retinyl esters are generated in the RPE by lecithin-retinol acyltransferase (LRAT), which catalyzes the reversible esterification of tROL using lecithin (phosphatidylcholine) as the acyl donor (Fig. 2). The transfer occurs specifically from the *sn*-1 position of lecithin, producing a retinyl ester and a lysophospholipid as products. This is the only known route of retinyl ester synthesis in RPE membranes. It should be noted that LRAT had been identified, not by traditional enzyme purification techniques, but by using specific affinity-labeling agents, such as all-*trans*-retinyl bromoacetate (tRBA) and *N*-Boc-L-biocytinyl-11-aminoundecane chloromethyl ketone (BACMK) (Fig. 2). LRAT, a minor membrane-bound protein, proved to be exceedingly difficult to fully purify. While the protein could be readily solubilized, it rapidly loses activity during a variety of purification schemes. In order to actually identify LRAT, the procedure of biotin affinity labeling was used. In this method, a biotin moiety is incorporated into a specific affinity-labeling agent directed toward a particular target protein(s). The enormous affinity and specificity of avidin–biotin interactions were used both to detect the labeled protein on gels and to purify the protein. In the particular instance discussed here, biotin is attached to a tRBA-like molecule (BACMK) and

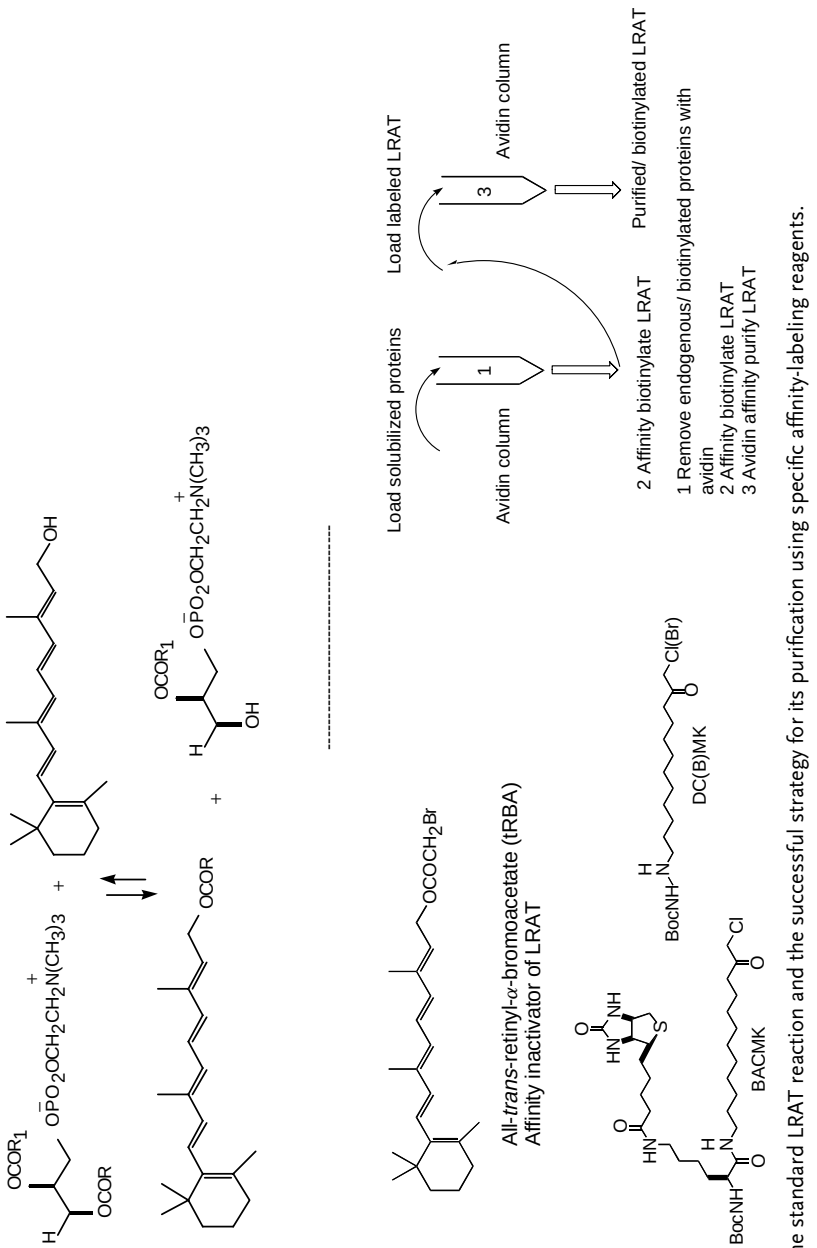

Fig. 2 The standard LRAT reaction and the successful strategy for its purification using specific affinity-labeling reagents.

the partially purified protein is specifically labeled (Fig. 2). Using Avidin chromatography, it was possible to obtain enough biotinylated protein to carry out partial sequence analysis. This partial sequence analysis showed that LRAT is a novel protein. This enabled the cloning and expression of human LRAT. LRAT has also been studied kinetically and mechanistically. Interestingly, LRAT is the founder member of a new family of proteins, which include class II tumor suppressors, viral proteases, and a protein that mediates morphogenesis.

As mentioned above, highly specific irreversible enzyme inactivators of LRAT have been designed and they alkylate the active-site cysteine residue of LRAT (C161). These inactivators have been essential for sorting out the substrate and product of isomerization. Two of these inactivators, BACMK, alluded to above, and (tRBA) are shown in Fig. 2. The use of tRBA as a specific inactivator of LRAT in RPE membranes revealed the substrate for isomerization and its product. First, pretreatment of RPE membranes with tRBA abolishes both tRE synthesis and 11-*cis*-retinoid biosynthesis. However, if tRE synthesis is allowed to occur first from tROL, before the applications of tRBA to block LRAT, cROL formation occurs as before. This means that isomerization occurs via an isomerohydrolase (IMH), which converts tREs into cROL. The isomerohydrolase reaction is shown in Fig. 3. The teleological reason for this unusual route is that it allows for the coupling of the free energy of hydrolysis of acyl esters to the thermodynamically uphill *trans* to *cis* isomerization. This unusual process of coupled isomerization/hydrolysis (IMH) for cROL biosynthesis in the RPE has been independently verified using different approaches. Moreover, recent studies on an LRAT knockout mouse are completely consistent with the earlier biochemical experiments and show that tRE synthesis is essential for 11-*cis*-retinoid generation. While these mice have tROL in the RPE, they cannot make 11-*cis*-retinoids. This is exactly what one would predict because tREs are the isomerization substrates. If there were a tROL isomerase, then the knockout animals would have synthesized cROL from it.

The studies described above identify two of the three known elements in the minimally three-component system required for the biosynthesis of cROL in the RPE, namely, LRAT and IMH. Of course, other elements are important as well in the full visual cycle. A specific RPE membrane–bound dehydrogenase (cROL dehydrogenase) carries out the oxidation of cROL to the visual chromophore cRLA. A stereospecific cROL(AL)) binding protein cellular retinaldehyde–binding protein (cRALBP) is also found in the RPE. *In vitro* studies show that this protein kinetically facilitates isomerohydrolase (IMH) action. Since the same result can be achieved with bovine serum albumin, a decidedly nonstereospecific retinoid-binding protein

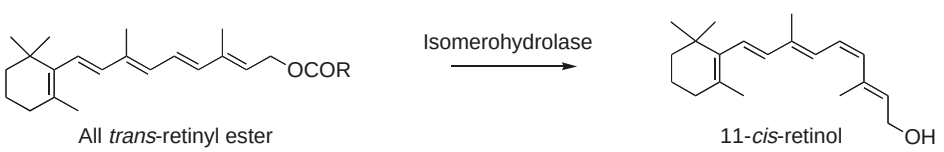

Fig. 3 Isomerohydrolase reaction.

(RBP), specific interactions with cRALBP are not essential for isomerization. It has been established that 11-*cis*-retinoids inhibit IMH so that the interaction of these retinoids with binding proteins promote isomerization by removing the 11-*cis*-retinoids. cRALBP knockout studies in mice also show that while cRALBP (−) mice appear to undergo rhodopsin regeneration more slowly than WT, the knockouts still regenerate rhodopsin, suggesting an important kinetic, although not essential, role for cRALBP in rhodopsin regeneration. In addition to cRALBP, several other RBPs are known to operate in the visual cycle and are important for its function. None of these RBPs bind the highly hydrophobic tREs, which brings up the question as to how these molecules are mobilized for processing by IMH.

RPE65 plays a central role in tRE binding, mobilization, and processing. RPE65 is a major protein expressed in the retinal pigment epithelium (RPE). This protein, first discovered in 1991, plays an important role in visual cycle function. This conclusion arises from knockout studies in mice, which show that RPE65 knockouts are unable to substantially produce 11-*cis*-retinoids, but accumulate hydrophobic tREs. RPE65 is not IMH and has also not been shown to possess any other enzymatic activity. While RPE65 shows very strong sequence homology to β-carotene-15, 15′ dioxygenase, an enzyme that oxidizes β-carotene into tRAL, RPE65 does not show this enzymatic activity. However, the fact that RPE65 and β-carotene dioxygenase are strongly homologous, suggests the possibility that RPE65 may be a RBP. Studies in this laboratory using retinoid-based affinity biotinylation of RPE65 have demonstrated a tRE-binding ability for RPE65. These experiments reveal that RPE65 is an RBP specific for tREs. This study was followed up with quantitative fluorescence measurements, which further revealed that the membrane-associated form of RPE65 (mRPE65) stereospecifically and saturably binds all-*trans*-retinyl palmitate (tRP) with a 47-nM Kd. tROL is only weakly bound, as are 11-*cis*-retinoids. These studies leave little doubt that mRPE65 is a tRE-binding protein.

As mentioned above, RPE65 knockout mice inappropriately accumulate tREs in oil droplets. REs are exceedingly hydrophobic, with little solubility in aqueous solutions. This can easily be seen in measurements of intermembranous transfer of retinoids. While retinol (al) undergo interliposomal transfer at exceedingly rapid rates, REs are inert to transfer. This would suggest that REs might require binding proteins to mobilize them if they are to productively enter the visual cycle. Since binding proteins are required stoichiometrically, they are required in substantial quantities. The substantial level of RPE65 in RPE membranes, coupled with its ability to bind tREs, suggests the possibility that mRPE65 might have a role in binding to and mobilizing tREs for processing in the visual cycle by IMH. Functional experiments leave little doubt that the role of mRPE65 in the visual cycle is to specifically bind to and mobilize tREs so that they can be processed by IMH. This conclusion suggests a minimally three-component system as the *sine qua non* of visual-pigment regeneration (Fig. 4). In the next section, the chemical biology of the individual components is discussed.

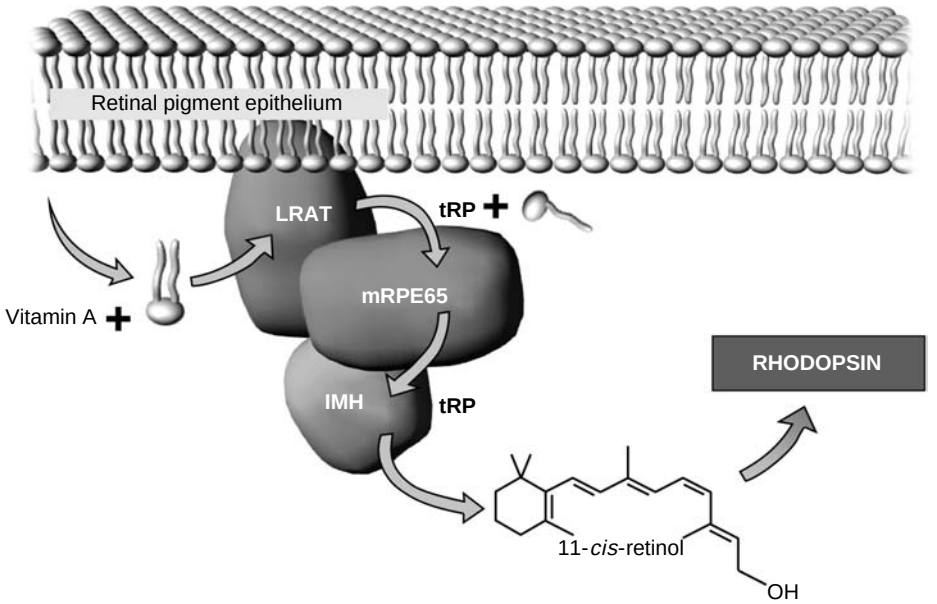

Fig. 4 The three-component system required for 11-*cis*-retinal biosynthesis.

3 Individual Components of the Three-component System

3.1 Lecithin Retinol Acyltransferase

The discovery of LRAT led to an appreciation of the fact that it is a novel enzyme, which, at the time of its discovery, was unrelated to other enzymes of known function. Initial experiments on the enzyme were directed toward mechanistic studies inasmuch as these studies reveal the mechanistic class to which LRAT belongs. Given the nature of LRAT's function as a catalyst of a transesterification reaction, it is likely that the enzyme contains an active-site nucleophile. Initial site-specific mutagenic experiments focused on the four cysteine residues found in human RPE LRAT as candidates. These experiments revealed C161 as the most likely candidate for a catalytically essential active-site cysteine residue. Further experiments were directed at the elaboration of the remaining catalytically essential amino acid residues of LRAT.

3.1.1 C161, H60, and Y154 as a Catalytic Triad in LRAT Action

As described above, LRAT has putative active-site cysteine and histidine residues. Because of this, it is possible that LRAT might operate via a cysteine protease-like mechanism. LRAT shows two pKa's in the pH versus rate profile at pH = 8.3 and 10.8. This however contrasts with thiol proteases, which generally show pKa's of 4.0(abnormal Cys) and 8.5(His). Therefore, it was of substantial interest to determine in what sense LRAT might be included in a thiol protease type family. Figure 5 shows a partial multiple sequence alignment for the LRAT family, in which one can see that the C161 of LRAT is conserved in the remaining family. In

Human-LRAT	D----VLEVPRTHLTHYGIYLGDNRVAHMMPDILLALTDDMGRTQKVVSN	94
Bovine-LRAT	D----VLEVPRTHLTHYGIYLGDNRVAHMMPDILLALTDDKGRTQKVVSN	94
Mouse-LRAT	D----VLEVSRTHFIHYGIYLGENRVAHLMPDILLALTNDKERTQKVVSN	94
Rat-LRAT	D----VLEVSRTHFTHYGIYLGDNRVAHLMPDILLALTSDKERTQKVVSN	94
XL-LRAT	D----LLEVPRTLFVHFGIYLGNNKVAHLMPDILPALSDDTCLIRRVVTN	92
ZF-LRAT	D----LLEVPRTLFTHFGIYLGDNKVAHLMPDILPVLTSNKSHLQNVVTN	91
HRev107-Mouse	D----LIEIFRPMYRHWAIYVGDGYVIHLAPPSEVAGAGAAS--------	49
HREV107-RAT	D----LIEIFRPMYSHWAIYVGDGYVIHLAPPSEIPGAGAAS--------	47
HRev107-3	D----LIEIFRPFYRHWAIYVGDGYVVHLAPPSEVAGAGAAS--------	49
TIG3	D----LIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSS--------	49
Echovirus23	K----LAYLDRGFYKHYGIIVG-GYVYQLDSDDIFKTA-----------	86
HumanParechovirus1	K----LAYLDRGFYKHYGIIVG-DHVYQLDSDDIFKTA-----------	78
HumanParechovirus2	K----LAYLDRGFYKHYGIVVG-DDVYQLDSDDIFKTA--------	77
Avian-encephalomyelitis-virus	D----IVSCSGEKAKHFGVYVG-DGVVHVDPEGNATNWFMKR--------	76
EGL26	DEVFCEVNVSGVKFYHSGIYAGDGMCYHFVCDAQESESFADALAVFSG--	198
Aichivirus	DDRVYIVRAQRPTYVHWAIRKVAPDGSAKQISLSRSGIQALV--------	48
Human-LRAT	AEKLLG-FTPYSLLWNNCEHFVTYCRYGTPISPQSDKFCETVKIIIRDQR	193
Bovine-LRAT	AEKLLG-ITPYSLLWNNCEHFVTYCRYGTPISPQADKFCENVKIIIRDQR	193
Mouse-LRAT	AEQQLG-LTPYSLLWNNCEHFVTYCRYGSRISPQAEKFYDTVKIIIRDQR	193
Rat-LRAT	AEQQLG-LTPYSLLWNNCEHFVTYCRYGSPISPQAEKFHETVKILIRDQR	193
XL-LRAT	AEKLVG-STPY-LLWNNCEHFVTYCRYGMPVSFQTEKFCETVKKIIRDRR	190
ZF-LRAT	AEKLVG-HFTYSLMWNNCEHFVTYCRYGTAVSLQTDQFCESLKSIIRDQR	190
HRev107-Mouse	AERLVGQEVLYRLTSENCEHFVNELRYGVPRSDQ----------VRDAV	134
HREV107-RAT	AEELVGQEVLYRLTSENCEHFVNELRYGVPRSDQ----------VRDTV	132
HRev107-3	AEELVGQEVLYKLTSENCEHFVNELRYGVARSDQ----------VRDVI	134
TIG3	AKEMVGQKMKYSIVSRNCEHFVAQLRYGKSRCKQ----------VEKAK	134
Echovirus23	----------VDSNCETIAKDIFGTHTLSQH----------QAI	153
HumanParechovirus1	----------VDKNCETIAKDIFGTHTLSQH----------QAI	145
HumanParechovirus2	----------VDNNCETIAKDIFGSHSLSQH----------QQI	144
Avian-encephalomyelitis-virus	----------IFVKNCETYARGIASGDYGTKEGEK--------WKTLL	147
EGL26	AEHLQRELENYDIRRCNCQHFSSECSTGVPFSYD----------MTSNF	282
Aichivirus	A-------T------NNCTHFVSSI-TGESLPN----------T	99

Fig. 5 Partial sequence alignment of LRAT family[a] sequence homology of LRAT family. Multiple sequence alignment of LRAT family was performed by using CLUSTAL W 1.82(http://www.ebi.ac.uk/clustalw/). Fully conserved His, Tyr, and Cys are in red. Accession numbers are as follows: human LRAT (GP:AF 071510), bovine LRAT (GP:AF 275344), mouse LRAT (GP:AF 255061), rat LRAT (GP:AF 255060), mouse Hrev107 (AAH24581), rat Hrev107 (X76453), human Hrev107-3 (P53816), tazarotene-induced gene protein (TIG3) (AF060228), echovirus23, human parechovirus1 (L02971), human parechovirus2 (AJ005695), Avian-encephalomyelitis-virus (AJ225173), EGL26 (NP493652), Aichivirus (AB010145). H60, Y154, and C161 are in red (see color plate p. xxxvi).

addition, though, H60, but not H57, is also conserved. Further experiments on site-specific mutagenesis on LRAT were performed using the readily expressible truncated lecithin-retinol acyltransferase (tLRAT), in which the N- and C-terminal transmembrane domains are removed to increase the solubility of the enzyme. In these experiments, it was clearly revealed that H60 is an essential histidine residue in the LRAT mechanism. In addition, to further explore whether LRAT might be a member or closely associated with the histidine-dependent protease family, it was also determined whether there are catalytically essential aspartate or glutamine residues in tLRAT. These two amino acids often show up as being important for thiol protease catalysis. However, of the glutamine and aspartic acid residues in the LRAT family, only Gln177 and Asp67 are largely conserved in the homologs. We studied the largely homologous aspartate and glutamine amino acids in tLRAT to

determine if they are essential for catalysis. The mutants produced were D49N, D67A, D76N, D111N, D128N, D191N, Q177E, and Q177S (92). The results of these experiments demonstrated unequivocally that neither the aspartate nor glutamine residues are essential for LRAT catalysis.

As mentioned above, we also show that H60 is the essential histidine of LRAT. Previous studies had strongly suggested that C161 is the active-site cysteine. To prove this beyond a shadow of a doubt, tLRAT was inactivated with (3-(*N*-Boc-lys-(biotinyl-0-)-all-*trans*-retinyl chloroacetate (BRCA), a specific inactivator designed for studying tRE binding proteins. The structure of this analog is shown in Fig. 6, compound 1. This analog is a pseudo first-order specific affinity-labeling agent for LRAT and also has the advantage of containing a biotinyl residue, which allows for the tryptic cleavage of the labeled protein followed by the ready purification of the labeled peptide. Moreover, because it has base cleavable ester moieties, the labeled peptide can easily be eluted from an Avidin column by simply raising the pH to 11, thereby generating the active-site peptide labeled with a carboxymethylene moiety. This easily identifies C161 as the active-site nucleophilic group of LRAT. Therefore, C161 is unequivocally the active-site nucleophile, as determined both by site-specific mutagenesis experiments and affinity-labeling experiments. In addition, H60 is the essential histidine residue in the LRAT family of proteins.

An important issue related to the role of C161 in catalysis is to identify the two pKa values found in the pH versus rate profile for LRAT. As mentioned before, there are two pKa values of approximately 8 and 10 observed in the pH versus rate profiles for LRAT. In the tLRAT series, the operant pKa's are 8.22 and 9.95 as compared to 8.5 and 10.4 in full-length LRAT. On the basis of known ionic equilibria, it is assumed that the pKa of approximately 8 is ascribable to C161. However, it is also known that amino acid residues at the active sites of enzymes may have pKa values that are markedly different from those of the amino acids themselves in solution. Therefore, to determine whether C161 is responsible for the pKa ~8, we carried out the pH versus rate profile for the titration of tLRAT with BRCA to determine whether C161 is responsible for the pKa of 8 in LRAT. When the titration of tLRAT was carried out with BRCA, a pKa value of 8.3 was observed. This independent titration experiment leaves very little doubt that C161 is responsible

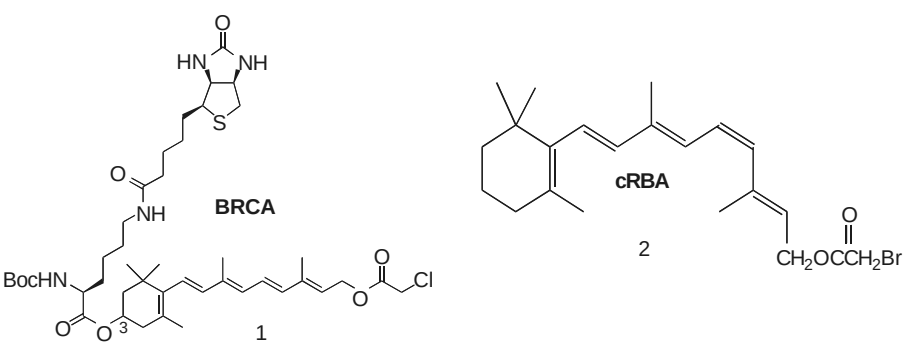

Fig. 6 Retinoid affinity-labeling agents.

for the ascending limb of the pH versus rate profile for LRAT.

With respect to the higher pKa ~10, this could be either due to an essential lysine or a tyrosine residue. These are the only amino acids with pKa's that reasonably could be ascribed to approximately 10. Site-specific mutagenic experiments on the relatively conserved lysine residues in LRAT (K90, K95, K104, K133, K134, K147, K180, and K186) (Fig. 5) were performed and showed that none of the lysine residues were essential for catalysis of tLRAT. Therefore, lysines can be eliminated as being responsible for the pKa ~10. The conserved tyrosine residues were also studied as candidates. Of these residues, tyrosine 154 is the most highly conserved of all of the tyrosine residues, as shown in Fig. 5. Indeed, of the relatively highly conserved tyrosine residues, only the Y154F mutation showed no activity. For comparison sake, the Y64F, Y118F, and the Y167F mutations were also studied, and all of those had substantial, if not full, activity. Therefore, we ascribe the higher pKa to tyrosine 154 and this suggests that LRAT operates via a catalytic triad in which C161, H60, and Y154 are essential for catalysis. Further inspection of Fig. 5 reveals that there is a bifurcation in the LRAT family. All of the LRAT family contains a cognate H60 and C161, but only a distinct subset contains a cognate of Y154. The enzymes that do not contain the Y154 are likely to be in the thiol protease wing of the LRAT superfamily. We, of course, have not done any mechanistic experiments on the protease family, and therefore do not know which other amino acids are essential for catalysis. A catalytic mechanism involving the catalytic C161, H60, and Ty154 in LRAT action is shown in Fig. 7.

3.2 Molecular Function of RPE65

The general role of RPE65 as a tRE chaperone is described in the Introduction section. The first indication of this was made by showing that a retinoid affinity-labeling agent (Fig. 6, compound 1) labeled RPE65. The all-*trans*-analogs (bromo and chloro esters) containing a biotin moiety attached to the 3-position of the retinoid were studied. The probes have two advantages over other biotinylated probes currently in use. A major advantage is the inclusion of a base cleavable moiety. Proteins labeled with these biotinylated derivatives are bound to tetrameric Avidin columns. Because the biotinylated moieties are adducted to the protein via base-sensitive ester linkages, we can wash the nonlabeled proteins from the Avidin matrix, and then simply raise the pH to 11 to cleave the esters and isolate the labeled protein with a carboxymethylene group adducted to an active-site residue. This makes our studies extremely straightforward in terms of the identification of labeled proteins. In addition, since base-sensitive linkages are imbedded in the probe, we are able to elute the proteins from high-affinity tetrameric Avidin columns with close to 100% yield. This labeling technology proved to be exceedingly useful in the discovery that RPE65 is a tRE-binding protein. Essentially, two proteins in RPE membranes were labeled. The minor protein is LRAT, which is expected since tRBA is known to label LRAT. Unexpected was the profound labeling of RPE65 by the retinoid reagent. RPE65 is labeled by the derivative in a time-dependent, and saturable fashion. The labeling was blocked by tRE, but not 11-*cis*-retinyl ester (cRE), oleyl alcohol or cROL, showing strong specificity in the binding reaction. Biochemical analysis

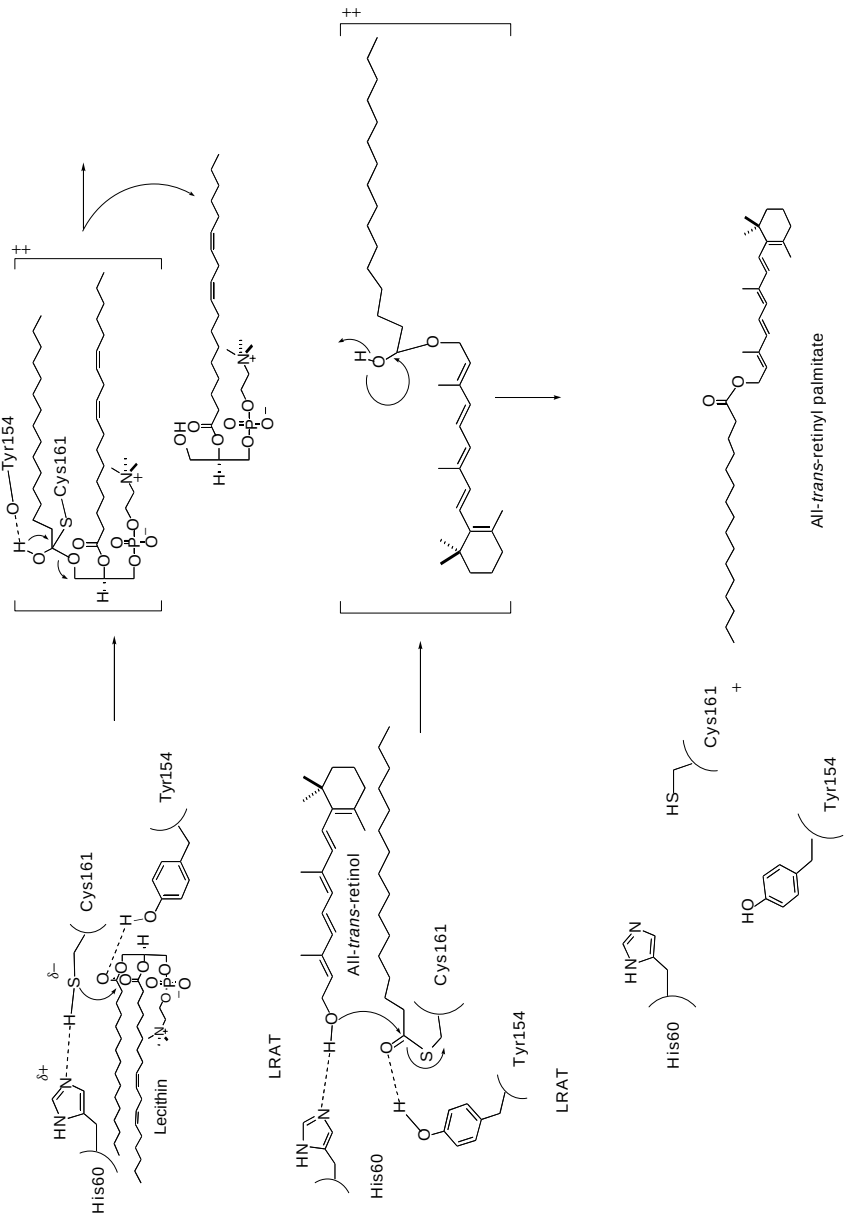

Fig. 7 LRAT mechanism showing the involvement of H60, Y154, and C161.

showed that two cysteine residues were labeled: C231 and C448. It is expected that these two sites may be involved in the active-site binding of tREs. These experiments were the first indication that RPE65 is a tRE-binding protein.

Following these labeling experiments, direct equilibrium fluorescence binding measurements were carried out, which showed that tREs bind in a saturable and specific fashion. We investigated the binding of all-*trans*- and 11-*cis*-retinyl esters, along with their congeneric alcohol derivatives, to freshly prepared, membrane-bound bovine RPE65 (mRPE65). Binding of tREs was of substantially higher affinity than for cREs and for tROL and cROL. The K_D for tRP is 47 nm. The binding constant for 11-*cis*-retinyl palmitate (cRP) is 2546 nM. The binding constant for tROL is 1463 nM and for cROL it is 2171 nM. We suggest that mRPE65 is a chaperone for highly hydrophobic tREs and mobilizes them as they enter the visual cycle for processing by IMH. The role of RPE65 in vision goes beyond its simple role as a tRE-binding protein.

Studies on mRPE65 suggest that it is part of the rate-limiting step(s) in the visual cycle. For example, mutations in RPE65 have been found which diminish rates of regeneration in WT mice and A2E formation in *Abcr* (*Abca4*) −/− mice (106). The mutational data are supported by small-molecule inhibitor data from our laboratory on retinoic acid. Accutane, 13-*cis*-retinoic acid, is a drug known to be effective in the treatment of certain dermatological diseases, including acne. Since night blindness is a major side effect of Accutane, we wondered whether the retinoic acids might specifically bind to mRPE65 and inhibit its function. Using fluorescence methods, we showed that tRP specifically binds to mRPE65. Accutane (13-*cis*-retinoic acid) also binds to mRPE65 with a $K_D = 195$ nm, while all-*trans*-retinoic acid specifically binds with a $K_D = 109$ nm. This is to be compared with the binding of tRP with a K_D of 47 nm. Since 13-*cis*-retinoic acid and all-*trans*-retinoic acid are equilibrated *in vivo*, the binding of both retinoic acids to mRPE65 is of interest. We went on to further study the interactions of the retinoic acids on tRE processing in the visual cycle and found that the competition of retinoic acid with tREs was evidenced by a diminution of the formation of cROL in the *in vitro* system. These studies demonstrate that the retinoic acids are competitive with tRP, both in direct competition assays, and they also compete with tRP visual cycle function. Thus, we suggest that at least one target for the retinoic acids is mRPE65. It had previously been shown that Accutane inhibits cROL dehydrogenase, but this inhibition does not account for the effect of Accutane on overall inhibition biosynthesis of 11-*cis*-retinoid biosynthesis in the visual cycle. We believe that this inhibition of RPE65 by the retinoic acids explains limiting of 11-*cis*-retinoid biosynthesis by these drugs. In addition, these experiments provide evidence for the rate-limiting nature of RPE65 function in the visual cycle.

The central role of RPE65 in visual cycle performance is further underscored by its likely importance in the regulation of retinoid fluxes through the cycle. Earlier we demonstrated one element of regulation of the visual cycle, namely, the apparent feed back inhibition of IMH by cROL. Chemical biological studies on RPE65 have uncovered an additional mode of regulation. There are two functionally distinct forms of RPE65, one membrane bound (mRPE65) and one soluble

(sRPE65). We found that mRPE65 is at least triply palmitoylated and sRPE65 is largely unpalmitoylated. Palmitoylation of RPE65 occurs at C231, C329, and C330. This palmitoylation explains the membrane association of mRPE65. We had already shown that mRPE65 binds to tRP with a $K_D = 47$ nM and to tROL with a $K_D = 1463$ nM. In the present study, we found that sRPE65 binds tROL with a $K_D = 65$ nM and tRP with a $K_D = 1235$ nM. In other words, mRPE65 and sRPE65 show opposite binding specificities. Most interestingly, LRAT uses mRPE65 as an acyl donor and acylates tROL to tRP. The kinetics of palmitoyl transfer to tROL with mRPE65 is sigmoidal, suggesting a regulatory role for this process. Overall, mRPE65 acts as a peptidomimetic of lecithin. In the process of forming tRP, mRE65 is functionally converted into sRPE as the product shows retinoid-binding behavior identical to sRPE65. It was also found that sRPE65 acts as a RBP and facilitates the esterification of tROL by LRAT. Furthermore, in the presence of mRPE65, cROL becomes an excellent substrate for LRAT in the formation of cRP. This suggests that the previously described feed back inhibition of IMH by cROL is at least partly due to the conversion of mRPE65 → sRPE65. This is directly revealed by studying the processing of ^{3}H-cROL biosynthesis in RPE membranes pre-incubated with cold cROL. The cold cROL is photochemically destroyed before incubation with ^{3}H-tROL. Here, a pronounced lag is observed, suggesting that the sRPE65 → mRPE65 reaction is partly rate limiting.

Overall, these experiments reveal a new role for LRAT, which acts as a molecular switch. mRPE65 is a palmitoyl donor, revealing a new acyl carrier protein role for palmitoylated proteins. When chromophore synthesis is not required, mRPE65 is converted into sRPE65 by LRAT, and further chromophore synthesis is blocked. These studies reveal new roles for palmitoylated proteins as molecular switches and LRAT as a palmitoyl transferase whose role is to catalyze the mRPE65 → sRPE65 conversion (Fig. 8).

3.3 The Isomerization Reaction

3.3.1 tREs as IMH Substrates in Rod- and Cone-dominated Species

As described earlier, an IMH mechanism is central to the operation of the visual cycle. In this mechanism, tREs are the substrates for isomerization and cROL is the product. The initial experiments in support of an IMH route relied on the use of tRBA, a selective LRAT inactivator. Inactivation of LRAT with tRBA produces an *in vitro* phenotype in which tREs are not formed from tROL, and hence isomerization cannot proceed. Similar experiments with additional inactivators produce the same result. The LRAT knockout mouse provides an *in vivo* cognate of these experiments, and, indeed, 11-*cis*-retinoids are not synthesized here either. However, it has been suggested that a tROL isomerase route may provide an alternative pathway to cROL biosynthesis. To assess whether such a pathway is quantitatively significant, a version of a pulse-chase experiment was performed.

The central problem in addressing whether the isomerization substitute is tROL or tRE is dealing with the rapid interconversion of the two putative substrates, as compared to the rate of isomerization. Thus, rapid interconversion makes

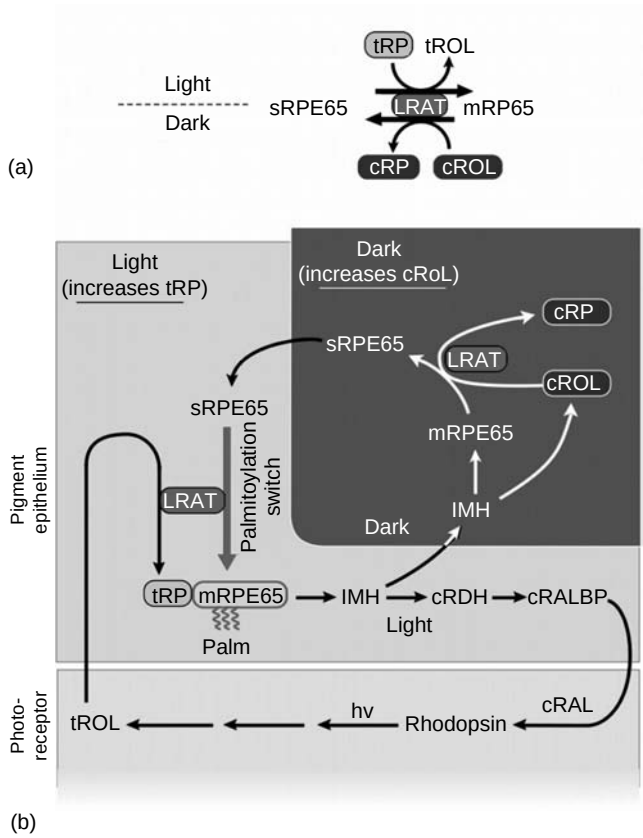

Fig. 8 Flow of retinoids in the visual cycle. (a) How the palmitoyl switch would behave in light and dark. (b) The flux of retinoids in the visual cycle and the role of the palmitoyl switch in directing their flow.

it impossible to carry out a standard pulse-chase experiment with either tROL or tRE because the pools will interconvert prior to isomerization. However, a pulse-chase experiment can be performed if the interconversion of tRE and tROL is blocked. This is readily done using tRBA to inhibit tRE synthesis and ebalactone A/B to inhibit tRE hydrolysis. Under these conditions, we can start with either high specific activity (^{3}H)-tRE/cold tROL or high specific activity (^{3}H)-tROL/cold tRE. The isomerization reaction is initiated and the specific activity of the isomerization product cROL is measured. The result could not have been more clear-cut. When the tRE pool is radioactively labeled, the resulting cROL is labeled with the same specific activity as the precursor tRE. The converse is true with tROL. These data unambiguously establish tRE as the substrate for isomerization.

Using this pulse-chase technique along with other approaches, we explored the mechanism of regeneration in a cone-dominated animal (chick). There

are known differences in the cone regeneration process compared to rod regeneration. For one thing, 11-*cis*-retinoid regeneration can occur in cone Müeller cells, although the quantitative importance of processing here as compared to flux through the RPE is unknown. Within experimental error, 11-*cis*-retinoid regeneration in rod-dominated animals is confined to the RPE. Secondly, tRE synthesis in cone-dominated species is stimulated by palmitoyl CoA, suggesting an atypical LRAT-type mechanism, or a novel tRE synthetase activity altogether. Given these observed differences, we explored the nature of the isomerization substrate and found it to be tRE, just as in rod-dominated species. The cone regeneration pathway was established using three complimentary approaches. First, no isomerization occurred after preincubation of chick RPE/retina with tRBA, which blocks LRAT, followed by addition of tROL that prevented isomerization. This is consistent with tRE being the isomerization substrate. Secondly, pulse-chase experiments, along the lines discussed above, showed that only tRE is processed into cROL. Finally, 11-*cis*-retinyl bromoacetate (cRBA), a specific affinity-labeling agent for IMH (Fig. 6, compound 2), blocked the processing of tRE into cROL. These data leave little doubt that cone and rod regeneration share the same overall molecular logic. An interesting additional issue here, though, is the nature of the RE synthetase. Western blotting, with anti-LRAT antibody shows a 21-kDa protein, as does labeling with BRCA. The standard LRAT molecular weight is approximately 25 kDa. Furthermore, chick RE synthase is also inactivated with tRBA, an affinity-labeling agent of LRAT. These data suggest an LRAT-type mechanism. However, RE synthesis in chick is stimulated by palmitoyl CoA (Palm CoA), but not by dipalmitoyl phosphatidyl choline (DPPC), an acyl donor for LRAT.

3.3.2 A Specific Inactivator of IMH and an Approach to IMH Identification

The identification and cloning of LRAT was made possible by the specific affinity-labeling agents tRBA and biotinylated derivatives thereof. IMH also presents a difficult problem with respect to identification owing to its lability in detergent and miniscule amount. Hence, a specific affinity-labeling agent directed at IMH was sought. Because IMH is strongly inhibited by 11-*cis*-retinoids, we prepared the 11-*cis*-retinoid-based affinity-labeling agent 11-*cis*-retinyl bromoacetate (cRBA) (Fig. 6, compound 2).

cRBA was designed as an affinity-labeling agent for IMH by analogy to the reagent used (tRBA) for the identification of LRAT. This molecule (cRBA) is a relatively high-affinity ($K_i = 0.1\ \mu M$) and a pseudo first-order inactivator (kinh $= 1.9 \times 10^{-3}\ s^{-1}$) of IMH action. The molecule shows saturable kinetics and a pseudo first-order mode of inactivation. This demonstrates a highly specific mode of inactivation of IMH action. Preliminary data on the use of ^{3}H-cRBA reveals the labeling of proteins largely in the 45- and 35-kDa region, and in the approximately 200-kDa region. We believe that the latter labeling occurs because IMH is likely to be found in a multienzyme complex whose interactions may partly survive SDS gel electrophoresis.

As mentioned earlier, IMH action can occur in one or two steps. As long as ester hydrolysis occurs subsequent to isomerization, the same thermodynamic imperative is met. The first step could involve *trans* → *cis* isomerization at the ester stage, and then the second step would involve specific hydrolysis of the

cRE. cRE hydrolase can be assayed in RPE membranes, and we were able to show that cRBA had no effect on general cRE hydrolase activity. So, if there should be a two-step reaction (isomerase/hydrolase) and the commonly assayed RPE cRE hydrolase activity is relevant in the visual cycle, then we know that cRBA is likely to be directed at the isomerase. The discovery of cRBA as a powerful and specific inactivator of IMH may be important in the identification and cloning of this protein.

Direct purification approaches to the identification and cloning of isomerohydrolase have not proved fruitful over the years, suggesting that alternative approaches will be required. One approach, of course, involves cloning by analogy. However, since the IMH reaction is a relatively unique reaction, this approach may be difficult to implement. Indirect approaches, such as the affinity-labeling approach described above, may provide an alternative approach to the identification of IMH. In this functional proteomic approach, IMH is specifically labeled with specific affinity-labeling reagents, such as cRBA, in mixtures of proteins and the labeled protein(s) is harvested for sequence analysis. Cloning and expression of the protein is then used to determine whether the correct protein has been identified. This approach was successful in the identification and cloning of LRAT and is being applied to the identification of IMH.

4 Possible Medical Significance of Studies on Retinoid Processing in the Visual System

Every time the visual cycle is traversed a molecule of rhodopsin is bleached and a molecule of all-*trans*-retinal (tRAL) is generated (Fig. 1). Like other aldehydes, tRAL is chemically reactive and needs to be detoxified enzymatically by reduction to form vitamin A. The tRAL that is generated can partition productively and re-enter the visual cycle, or it can chemically react with biological amines to generate retinotoxic products (Fig. 9). In these reactions, a tRAL forms a Schiff base with the amine followed by enamine formation and then a series of reactions that generate the toxic compounds irreversibly. These reactions are similar in kind to those found in diabetes, where excess glucose aldehyde chemically reacts with proteins by reversible Schiff base formation followed by further chemical reactions, which irreversibly secures the glucose moiety to the protein. These chemical reactions are thought to be of primary importance in the etiology of diabetes.

In the case of tRAL reactions, the salient biological amine in question is photoreceptor phosphatidyl ethanolamine, which reacts with tRAL to generate the toxic lipofuscins of RPE cells (Fig. 9). The best-characterized RPE lipofuscin fluorophores is the diretinal conjugate A2E, its photoisomer iso-A2E, and other minor *cis*-isomers of A2E. These fluorophores are generated from sequential reactions between all-*trans*-retinal and phosphatidylethanolamine (2 : 1). The A2E family perturbs cell membranes, confers susceptibility to blue light-induced apoptosis and alters lysosomal function. In light of this adverse behavior, there is considerable interest in retarding A2E formation as a means to prevent vision loss in retinal degenerative diseases including Stargardt's disease and perhaps age-related macular degeneration. In Stargardt's disease, a juvenile form of macular degeneration, excessive accumulation of A2E occurs because of the presence of

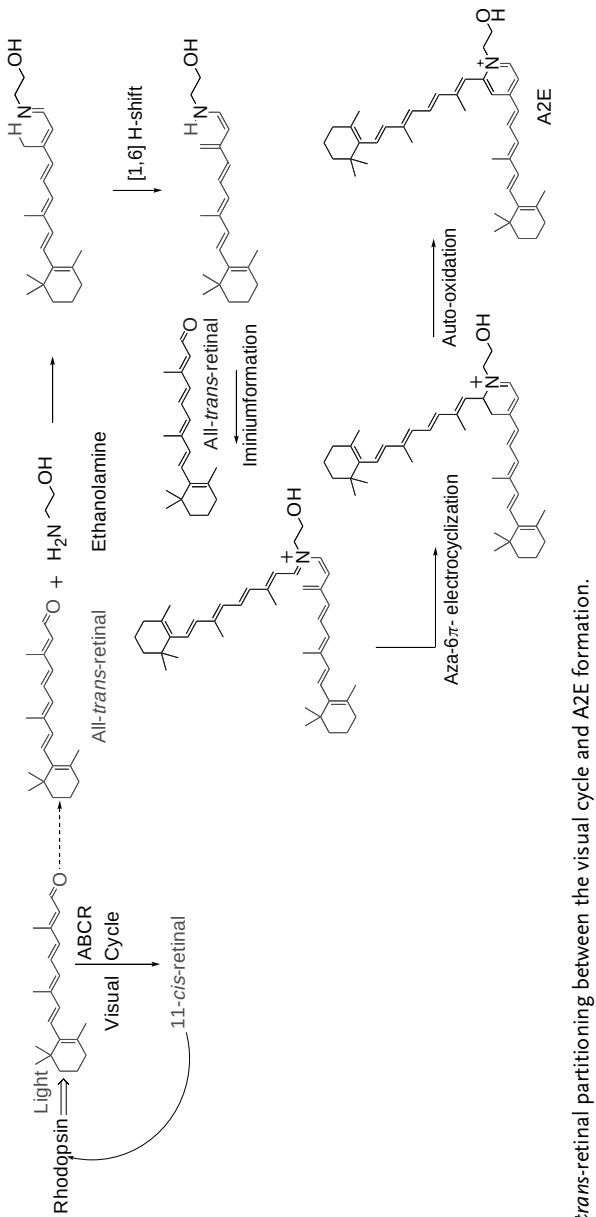

Fig. 9 All-*trans*-retinal partitioning between the visual cycle and A2E formation.

inactivating mutations in a pump that removes precursors to A2E before they proceed to form A2E and conjugates. Because accumulation of A2E, either through age-related processes or because of mutations, as in Stargardt's disease, it has become increasingly clear that small-molecule inhibitors, which limit the visual cycle, and thus limit the formation of A2E, may be of therapeutic value in the treatment of visual disorders. By limiting fluxes through the visual cycle by inhibiting key enzymes and retinoid-binding proteins thereby reducing all-*trans*-retinal formation, it should be possible to decrease formation of A2E. Inhibiting the visual cycle would probably only affect scotopic vision (night blindness) in a marginal way, but would not affect cone vision because the cones preferentially take up 11-*cis*-retinoids. Moreover, even though rods account for >95% of total photoreceptors, rod function is not very important for most people in industrial societies as a consequence of the advent of artificial light.

To effectively inhibit the visual cycle, a target must be chosen, which is at least partly rate limiting. RPE65, described above, may be such a target. At least one mutation in RPE65 is known to inhibit retinoid flux through the visual cycle, while also inhibiting A2E accumulation. Moreover, a drug, known to cause night blindness in humans and A2E accumulation in mice, interferes with RPE65 function. This again suggests that RPE65 is part of the rate-limiting step. The drug in question, Accutane™ (13-*cis*-retinoic acid), which is used to treat serious dermatological disorders, binds powerfully to mRPE65 and displaces all-*trans*-retinyl esters, making the latter unavailable for processing by IMH. While it is highly unlikely that the highly toxic retinoic acids are serious candidates for prophylactic treatment of macular degeneration, the studies suggest the realistic possibility that appropriate nontoxic small-molecule antagonists can be found to limit the visual cycle and treat retinal degenerative diseases.

Acknowledgment

Figures 1 and 8 are reprinted from *Cell*, Vol. 117, L. Xue, D. R. Gollapalli, P. Maiti, W. J. Jahng, and R. R. Rando: A palmitoylation switch mechanism in the regulation of the visual cycle, 761–771, 2004, with permission from Elsevier.

See also Bacteriorhodopsin, Molecular Biology of.

Bibliography

Books and Reviews

Lamb, T.D., Pugh, E.N. Jr. (2004) Dark adaptation and the retinoid cycle of vision, *Prog. Retin. Eye Res.* **23**, 307–380.

Makino, C.L., Wen, X.H., Lem, J. (2003) Piecing together the timetable for visual transduction with transgenic animals, *Curr. Opin. Neurobiol.* **13**, 404–412.

Rodieck, R.W. (1998) How Photoreceptors Work, *The First Steps in Seeing*, Sinauer Associates, Sunderland, MA, pp. 159–194.

Sakmar, T.P., Menon, S.T., Marin, E.P., Awad, E.S. (2002) Rhodopsin: insights from recent structural studies, *Annu. Rev. Biophys. Biomol. Struct.* **31**, 443–484.

Sparrow, J.R., Cai, B., Fishkin, N., Jang, Y.P., Krane, S., Vollmer, H.R., Zhou, J., Nakanishi, K. (2003a) A2E a fluorophore of RPE lipofuscin: can it cause RPE degeneration? *Adv. Exp. Med. Biol.* **533**, 205–211.

Sparrow, J.R., Fishkin, N., Zhou, J., Cai, B., Jang, Y.P., Krane, S., Itagaki, Y., Nakanishi, K. (2003b) A2E, a byproduct of the visual cycle, *Vision Res.* **43**, 2983–2990.

Thompson, D.A., Gal, A. (2003) Vitamin A metabolism in the retinal pigment epithelium: genes, mutations, and diseases, *Prog. Retin. Eye Res.* **22**, 683–703.

Primary Literature

Adler, A.J., Evans, C.D., Stafford, W.F. III (1985) Molecular properties of bovine interphotoreceptor retinol-binding protein, *J. Biol. Chem.* **260**, 4850–4855.

Barry, R.J., Cañada, F.J., Rando, R.R. (1989) Solubilization and partial purification of retinyl ester synthetase and retinoid isomerase from bovine ocular pigment epithelium, *J. Biol. Chem.* **264**, 9231–9238.

Batten, M.L., Imanishi, Y., Maeda, T., Tu, D.C., Moise, A.R., Bronson, D., Possin, D., Van Gelder, R.N., Baehr, W., Palczewski, K. (2004) Lecithin-retinol acyltransferase is essential for accumulation of all-trans-retinyl esters in the eye and in the liver, *J. Biol. Chem.* **279**, 10422–10432.

Baumann, C., Zeppenfeld, W. (1981) Effect of pH on the formation and decay of the metarhodopsins of the frog, *J. Physiol.* **317**, 347–364.

Bavik, C.O., Eriksson, U., Allen, R.A., Peterson, P.A. (1991) Identification and partial characterization of a retinal pigment epithelial membrane receptor for plasma retinol-binding protein, *J. Biol. Chem.* **266**, 14978–14985.

Baylor, D.A., Burns, M.E. (1998) Control of rhodopsin activity in vision, *Eye* **12**(Pt. B), 521–525.

Ben-Shabat, S., Parish, C.A., Vollmer, H.R., Itagaki, Y., Fishkin, N., Nakanishi, K., Sparrow, J.R. (2002) Biosynthetic studies of A2E, a major fluorophore of retinal pigment epithelial lipofuscin, *J. Biol. Chem.* **277**, 7183–7190.

Bernstein, P.S., Rando, R.R. (1986) In vivo isomerization of all-*trans*- to 11-*cis*-retinoids in the eye occurs at the alcohol oxidation-state, *Biochemistry* **25**, 6473–6478.

Bernstein, P.S., Law, W.C., Rando, R.R. (1987) Isomerization of all-*trans*-retinoids to 11-*cis*-retinoids in vitro, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1849–1853.

Boorjian, S., Tickoo, S.K., Mongan, N.P., Yu, H., Bok, D., Rando, R.R., Nanus, D.M., Scherr, D.S., Gudas, L.J. (2004) Reduced lecithin: retinol acyltransferase expression correlates with increased pathologic tumor stage in bladder cancer, *Clin. Cancer Res.* **10**, 3429–3437.

Cañada, F.J., Law, W.C., Rando, R.R., Yamamoto, T., Derguini, F., Nakanishi, K. (1990) Substrate specificities and mechanism in the enzymatic processing of vitamin A into 11-cis-retinol, *Biochemistry* **29**, 9690–9697.

Chader, G.J., Wiggert, B., Lai, Y.-L., Lee, L., Fletcher, R.T. (1983) in: Osborne N.N., Chodes G.J. (Eds.) *Progress in Retinal Research*, Vol. 2, Pergamon Press, Oxford, pp. 163–184.

Choo, D.W., Cheung, E., Rando, R.R. (1998) Lack of effect of RPE65 removal on the enzymatic processing of all-*trans*-retinol into 11-*cis*-retinol in vitro, *FEBS Lett.* **440**, 195–198.

Crabb, J.W., Nie, Z., Chen, Y., Hulmes, J.D., West, K.A., Kapron, J.T., Ruuska, S.E., Noy, N., Saari, J.C. (1998) Cellular retinaldehyde-binding protein ligand interactions. Gln-210 and Lys-221 are in the retinoid binding pocket, *J. Biol. Chem.* **273**, 20712–20720.

Crabb, J.W., Carlson, A., Chen, Y., Goldflam, S., Intres, R., West, K.A., Hulmes, J.D., Kapron, J.T., Luck, L.A., Horwitz, J., Bok, D. (1998) Structural and functional characterization of recombinant human cellular retinaldehyde-binding protein, *Protein Sci.* **7**, 746–757.

Cunliffe, W.J. (1986) Retinoids in dermatology, *Biochem. Soc. Trans.* **14**, 943–945.

Das, S., Bhardwaj, N., Kjeldbye, H., Gouras, P. (1992) Muller cells of chicken retina synthesize 11-*cis*-retinol, *Biochem. J.* **285**, 907–913.

Deigner, P.S., Law, W.C., Cañada, F.J., Rando, R.R. (1989) Membranes as the energy source in the endergonic transformation of vitamin A to 11-*cis*-retinol, *Science* **244**, 968–971.

Eldred, G.E., Lasky, M.R. (1993) Retinal age pigments generated by self-assembling lysosomotropic detergents, *Nature* **361**, 724–726.

Farrens, D.L., Khorana, H.G. (1995) Structure and function in rhodopsin. Measurement of the rate of metarhodopsin II decay by fluorescence spectroscopy, *J. Biol. Chem.* **270**, 5073–5076.

Gollapalli, D.R., Rando, R.R. (2003a) All-*trans*-retinyl esters are the substrates for isomerization in the vertebrate visual cycle, *Biochemistry* **42**, 5809–5818.

Gollapalli, D.R., Rando, R.R. (2003b) Molecular logic of the biosynthesis of 11-*cis*-retinoids in a cone-dominated species, *Biochemistry* **42**, 14921–14929.

Gollapalli, D.R., Rando, R.R. (2003c) Specific inactivation of isomerohydrolase activity by 11-*cis*-retinoids, *Biochim. Biophys. Acta.* **1651**, 93–101.

Gollapalli, D.R., Rando, R.R. (2004) The specific binding of retinoic acid to RPE65 and approaches to the treatment of macular degeneration, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10030–10035.

Gollapalli, D.R., Maiti, P., Rando, R.R. (2003) RPE65 operates in the vertebrate visual cycle by stereospecifically binding all-*trans*-retinyl esters, *Biochemistry* **42**, 1824–11830; erratum in: *Biochemistry* **43**, 7226.

Groenendjk, G.W., Jacobs, C.W., Bonting, S.L., Daemen, F. (1980) Dark isomerization of retinals in the presence of phosphatidylethanolamine, *Eur. J. Biochem.* **106**, 119–128.

Hamel, C.P., Tsilou, E., Haris, E., Pfeffer, B.A., Hooks, J.J., Detrick, B., Redmond, T.M. (1993) A developmentally regulated microsomal protein specific for the pigment epithelium of the vertebrate retina, *J. Neurosci.* **34**, 414–415.

Hamm, H.E. (1991) Molecular interactions between the photoreceptor G protein and rhodopsin, *Cell. Mol. Neurobiol.* **11**, 563–578.

Hanna-Rose, W., Han, M. (2002) The Caenorhabditis elegans EGL-26 protein mediates vulval cell morphogenesis, *Dev. Biol.* **241**, 247–258.

Ho, M.T., Massey, J.B., Pownall, H.J., Anderson, R.E., Hollyfield, J.G. (1989) Mechanism of vitamin A movement between rod outer segments, interphotoreceptor retinoid-binding protein, and liposomes, *Proc. Nutr. Soc.* **24**, 170–172.

Holz, F.G., Schutt, F., Kopitz, J., Eldred, G.E., Kruse, F.E., Volcker, H.E., Cantz, M. (1999) Inhibition of lysosomal degradative functions in RPE cells by a retinoid component of lipofuscin, *Invest. Ophthalmol. Vis. Sci.* **40**, 737–743.

Hubbard, R. (1956) Retinene isomerase, *J. Gen. Physiol.* **39**, 935–962.

Hughes, P.J., Stanway, G. (2000) The 2A proteins of three diverse picornaviruses are related to each other and to the H-rev107 family of proteins involved in the control of cell proliferation, *J. Gen. Virol.* **81**, 201–207.

Hurley, J.B. (1992) Signal transduction enzymes of vertebrate photoreceptors, *J. Bioenerg. Biomembr.* **24**, 219–226.

Jahng, W.J., Xue, L., Rando, R.R. (2003) Lecithin retinol acyltransferase is a founder member of a novel family of enzymes, *Biochemistry* **42**, 12805–12812.

Jahng, W.J., David, C., Nesnas, N., Nakanishi, K., Rando, R.R. (2003) A cleavable affinity biotinylating agent reveals a retinoid binding role for RPE65, *Biochemistry* **42**, 6159–6168.

Kennedy, B.N., Huang, J., Saari, J.C., Crabb, J.W. (1998) Molecular characterization of the mouse gene encoding cellular retinaldehyde-binding protein, *Mol. Vis.* **4**, 14.

Kim, S.R., Fishkin, N., Kong, J., Nakanishi, K., Allikmets, R., Sparrow, J.R. (2004) RPE65 Leu450Met variant is associated with reduced levels of the retinal pigment epithelium lipofuscin fluorophores A2E and iso A2E, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11668–11672.

Knowles, J.R. (1976) The intrinsic pK_a-values of functional groups in enzymes: improper deductions from the pH-dependence of steady-state parameters, *CRC Crit. Rev. Biochem.* **4**, 165–173.

Koenig, B., Arendt, A., McDowell, J.H., Kahlert, M., Hargrave, P.A., Hofmann, K.P. (1989) Three cytoplasmic loops of rhodopsin interact with transducin, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6878–6882.

Kühne, W. (1878) The photochemistry of the retina and visual purple, *Unters. Physiol. Inst. Heidelberg* **1**, 341–369.

Law, W.C., Rando, R.R. (1989) The molecular basis of retinoic acid induced night blindness, *Biochem. Biophys. Res. Commun.* **161**, 825–829.

Liu, J., Itagaki, Y., Ben-Shabat, S., Nakanishi, K., Sparrow, J.R. (2000) The biosynthesis of A2E, a fluorophore of aging retina, involves the formation of the precursor, A2-PE, in the photoreceptor outer segment membrane, *J. Biol. Chem.* **275**, 29354–29360.

Loew, A., Gonzalez-Fernandez, F. (2002) Crystal structure of the functional unit of interphotoreceptor retinoid binding protein, *Structure* **10**, 43–49.

Mata, N.L., Weng, J., Travis, G.H. (2000) Biosynthesis of a major lipofuscin fluorophore in mice and humans with ABCR-mediated retinal and macular degeneration, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7154–7159.

Mata, N.L., Radu, R.A., Clemmons, R.S., Travis, G.H. (2002) Isomerization and oxidation of vitamin A in cone-dominant retinas: a novel pathway for visual-pigment regeneration in daylight, *Neuron* **36**, 69–80.

Mata, N.L., Moghrabi, W.N., Lee, J.S., Bui, T.V., Radu, R.A., Horwitz, J., Travis, G.H. (2004) RPE65 is a retinyl ester binding protein that presents insoluble substrate to the isomerase in retinal pigment epithelial cells, *J. Biol. Chem.* **279**, 635–643.

Mitchell, D.C., Niu, S.-L., Litman, B.J. (2001) Optimization of receptor-G protein coupling by bilayer lipid composition I. Kinetics of rhodopsin-transducin binding, *J. Biol. Chem.* **276**, 42801–42806.

Moiseyev, G., Crouch, R.K., Goletz, P., Oatis, J. Jr., Redmond, M., Ma, J.-X. (2002) Retinyl esters are the substrate for isomerohydrolase, *Biochemistry* **42**, 2229–2238.

Mondal, M.S., Ruiz, A., Bok, D., Rando, R.R. (2000) Lecithin retinol acyltransferase contains cysteine residues essential for catalysis, *Biochemistry* **39**, 5215–5220.

Nesnas, N., Rando, R.R., Nakanishi, K. (2002) Synthesis of biotinylated retinoids for cross-linking and isolation of retinol binding proteins, *Tetrahedron* **58**, 6577–6584.

Okada, T., Ernst, O.P., Palczewski, K., Hofmann, K.P. (2001) Activation of rhodopsin: new insights from structural and biochemical studies, *Trends Biochem. Sci.* **26**, 318–324.

Okajima, T.I., Wiggert, B., Chader, G.J., Pepperberg, D.R. (1994) Retinoid processing in retinal pigment epithelium of toad (Bufo marinus), *J. Biol. Chem.* **269**, 21983–21989.

Parish, C.A., Hashimoto, M., Nakanishi, K., Dillon, J., Sparrow, J. (1998) Isolation and one-step preparation of A2E and iso-A2E, fluorophores from human retinal pigment epithelium, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14609–14613.

Pepe, I.M., Panfoli, I., Hamm, H.E. (1989) Visual transduction in vertebrate photoreceptors, *Cell. Biophys.* **14**, 129–137.

Rando, R.R. (1990) The chemistry of vitamin A and vision, *Angew. Chem., Int. Ed. Engl.* **29**, 461–480. Also published in German as Die chemie des Vitamins A und des Sehvorgangs. *Angew. Chem.* **102**, 507–526.

Rando, R.R. (1991a) Membrane phospholipids as an energy source in the operation of the visual cycle, *Biochemistry* **30**, 595–602.

Rando, R.R. (1991b) Membrane phospholipids and the dark side of vision, *J. Bioenerg. Biomembr.* **23**, 133–146.

Rando, R.R. (1994) Isomerization reactions of retinoids in the visual system, *Pure Appl. Chem.* **66**, 989–994.

Rando, R.R. (2001) The biochemistry of the visual cycle, *Chem. Rev.* **101**, 1881–1896.

Rando, R.R., Bangerter, F.W. (1982) The rapid intermembranous transfer of retinoids, *Biochem. Biophys. Res. Commun.* **104**, 430–436.

Rando, R.R., Chang, A. (1983) Studies on the catalyzed interconversions of vitamin A derivatives, *J. Am. Chem. Soc.* **105**, 2879–2882.

Rattner, A., Smallwood, P.M., Nathans, J. (2000) Identification and characterization of all-*trans*-retinol dehydrogenase from photoreceptor outer segments, the visual cycle enzyme that reduces all-*trans*-retinal to all-*trans*-retinol, *J. Biol. Chem.* **275**, 11034–11043.

Redmond, T.M., Yu, S., Lee, E., Bok, D., Hamasaki, D., Chen, N., Goletz, P., Ma, J.X., Crouch, R.K., Pfeifer, K. (1998) RPE65 is necessary for production of 11-*cis*-vitamin A in the retinal visual cycle, *Nat. Genet.* **20**, 344–351.

Ripps, H., Peachey, N.S., Xu, X., Nozell, S.E., Smith, S.B., Liou, G.I. (2000) The rhodopsin cycle is preserved in IRBP "knockout" mice despite abnormalities in retinal structure and function, *Vis. Neurosci.* **17**, 97–105.

Ruiz, A., Winston, A., Lim, Y.-H., Gilbert, B.A., Rando, R.R., Bok, D. (1999) Molecular and biochemical characterization of lecithin retinol acyltransferase, *J. Biol. Chem.* **274**, 3834–3841.

Saari, J.C., Bredberg, D.L. (1989) Lecithin retinol acyltransferase in retinal pigment epithelial microsomes, *J. Biol. Chem.* **264**, 8636–8640.

Saari, J.C., Nawroot, M., Kennedy, B.N., Garwin, G.G., Hurley, J.B., Huang, J., Possin, D.E., Crabb, J.W. (2001) Visual cycle impairment in cellular retinaldehyde binding protein (CRALBP) knockout mice results in delayed dark adaptation, *Neuron* **29**, 739–748.

Schutt, F., Davies, S., Kopitz, J., Holz, F.G., Boulton, M.E. (2000) Photodamage to human RPE cells by A2-E, a retinoid component of lipofuscin, *Invest. Ophthalmol. Vis. Sci.* **41**, 2303–2308.

Shi, Y.Q., Hubacek, I., Rando, R.R. (1993) Kinetic mechanism of lecithin retinol acyl transferase, *Biochemistry* **32**, 1257–1263.

Shi, Y.-Q., Furuyoshi, S., Hubacek, I., Rando, R.R. (1993) Affinity labeling of lecithin retinol acyltransferase, *Biochemistry* **32**, 3077–3080.

Sieving, P.A., Chaudhry, P., Kondo, M., Provenzano, M., Wu, D., Carlson, T.J., Bush, R.A., Thompson, D.A. (2001) Inhibition of the visual cycle in vivo by 13-cis retinoic acid protects from light damage and provides a mechanism

for night blindness in isotretinoin therapy, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1835–1840.

Simon, A., Hellman, U., Wernstedt, C., Eriksson, U. (1995) The retinal pigment epithelial-specific 11-*cis* retinol dehydrogenase belongs to the family of short chain alcohol dehydrogenases, *J. Biol. Chem.* **270**, 1107–1112.

Sparrow, J.R. (2003) Therapy for macular degeneration: insights from acne, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4353–4354.

Sparrow, J.R., Cai, B. (2001) Blue light-induced apoptosis of A2E-containing RPE: involvement of caspase-3 and protection by Bcl-2, *Invest. Ophthalmol. Vis. Sci.* **42**, 1356–1362.

Sparrow, J.R., Nakanishi, K., Parish, C.A. (2000) The lipofuscin fluorophore A2E mediates blue light-induced damage to retinal pigmented epithelial cells, *Invest. Ophthalmol. Vis. Sci.* **41**, 1981–1989.

Sparrow, J.R., Parish, C.A., Hashimoto, M., Nakanishi, K. (1999) A2E, a lipofuscin fluorophore, in human retinal pigmented epithelial cells in culture, *Invest. Ophthalmol. Vis. Sci.* **40**, 2988–2995.

Sparrow, J.R., Vollmer-Snarr, H.R., Zhou, J., Jang, Y.P., Jockusch, S., Itagaki, Y., Nakanishi, K. (2003) A2E-epoxides damage DNA in retinal pigment epithelial cells. Vitamin E and other antioxidants inhibit A2E-epoxide formation, *J. Biol. Chem.* **278**, 18207–18213.

Stecher, H., Gelb, M.H., Saari, J.C., Palczewski, K. (1999) Preferential release of 11-cis-retinol from retinal pigment epithelial cells in the presence of cellular retinaldehyde-binding protein, *J. Biol. Chem.* **274**, 8577–8585.

Storer, A.C., Menard, R. (1994) Catalytic mechanism in papain family of cysteine peptidases, *Methods Enzymol.* **244**, 486–500.

Struthers, M., Oprian, D.D. (2000) Mapping tertiary contacts between amino acid residues within rhodopsin, *Methods Enzymol.* **315**, (Vertebrate Phototransduction and the Visual Cycle, Pt. A), 130–143.

Stryer, L. (1986) Cyclic GMP cascade of vision, *Annu. Rev. Neurosci.* **9**, 87–119.

Suter, M., Reme, C., Grimm, C., Wenzel, A., Jaattela, M., Esser, P., Kociok, N., Leist, M., Richter, C. (2000) Age-related macular degeneration. The lipofusion component N-retinyl-N-retinylidene ethanolamine detaches proapoptotic proteins from mitochondria and induces apoptosis in mammalian retinal pigment epithelial cells, *J. Biol. Chem.* **275**, 39625–39630.

Trehan, A., Cañada, F.J., Rando, R.R. (1990) Inhibitors of retinyl ester formation also prevent the biosynthesis of 11-cis-retinol, *Biochemistry* **29**, 309–312.

Wald, G. (1953) The biochemistry of vision, *Annu. Rev. Biochem.* **22**, 497–526.

Wenzel, A., Reme, C.E., Williams, T.P., Hafezi, F., Grimm, C. (2001) The RPE65 Leu450Met variation increases retinal resistance against light-induced degeneration by slowing rhodopsin regeneration, *J. Neurosci.* **21**, 53–58.

Winston, A., Rando, R.R. (1998) Regulation of isomerohydrolase activity in the visual cycle, *Biochemistry* **37**, 2044–2050.

Xue, L., Rando, R.R. (2004) Roles of cysteine 161 and tyrosine 154 in the lecithin-retinol acyltransferase mechanism, *Biochemistry* **43**, 6120–6126.

Xue, L., Gollapalli, D.R., Maiti, M., Jahng, W.J., Rando, R.R. (2004) A palmitoylation switch mechanism in the regulation of the visual cycle, *Cell* **117**, 761–771.

Znoiko, S.L., Crouch, R.K., Moiseyev, G., Ma, J.-X. (2002) Identification of the RPE65 protein in mammalian cone photoreceptors, *Invest. Ophthalmol. Vis. Sci.* **43**, 1604–1609.

Retrotransposons and Transposons: *see* Horizontal Gene Transfer

Retrovirus Integration, Biological and Biochemical Machineries for

Duane Grandgenett and Sibes Bera
Institute for Molecular Virology, St. Louis, MO

1 Retrovirus Replication and Integration Overview 401
1.1 Replication Pathway 401
1.2 Consequences of Integration 403

2 Retrovirus Preintegration Complexes 403
2.1 Physical and Molecular Properties 403
2.2 Host Chromosomes Target Site Selection *In Vivo* 406
2.3 Integration Properties of Isolated PIC *In Vitro* 406

3 Synaptic Complexes Capable of Full-site Integration 407
3.1 Reconstitution of Synaptic Complexes with IN and Viral DNA Substrates Capable of Efficient Full-site Integration 407

4 Retrovirus Integrase 408
4.1 Origin and Structure of IN 408
4.2 Viral DNA–IN Synaptic Complexes 409

5 Human Gene Therapy 410
5.1 Success and Failure 410

6 Ubiquitous Nature of Mobile DNA Elements 410

7 Perspectives 410

Acknowledgment 411

Bibliography 411
Books and Reviews 411
Primary Literature 411

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

Keywords

Attachment Sites (att)
The sequences located at the viral DNA ends that are necessary and sufficient for IN binding and full-site integration activity.

Full-site Integration
The IN-mediated concerted insertion of two viral DNA ends into cellular DNA *in vivo* or into an exogenously provided DNA target *in vitro*.

Human Gene Therapy
The retrovirally mediated transduction of a therapeutic gene into the human chromosome.

Human Immunodeficiency Virus
Human retrovirus that is the causative agent of HIV-1/AIDS.

Integrase (IN)
The viral enzyme that possesses 3′ OH nuclease and strand transfer activities, both of which are essential for proper insertion of the retrovirus DNA genome into host DNA.

Mobile DNA Elements
Small DNA elements similar to retroviral DNA in structure that serve to provide self-mobility within and between cells. Retrotransposons contain a DNA structure that is similar to retroviral DNA. Many other mobile DNA elements dissimilar to retroviruses also exist in nature.

Preintegration Complexes (PIC)
Cytoplasmic viral DNA-protein complexes formed after reverse transcription of the viral RNA genome. PIC are transported into the cell nucleus allowing integration of the viral DNA genome into host chromosomes.

Provirus
The term used to describe the integrated viral DNA genome in cell DNA.

Retrovirus
Retroviruses comprise a large family of viruses that contain RNA as their genome. In a retrovirus, a cell-derived membrane envelope encases the genomic RNA along with viral enzymes and structural proteins. The replication cycle requires reverse transcription of the RNA genome into DNA followed by integration of this DNA by viral integrase into host chromosomes.

Synaptic Complex
The assembly of IN together with the linear viral DNA genome; a nucleoprotein complex capable of mediating full-site integration.

Retroviruses are unique in nature in that they require the integration of their viral DNA genome into host chromosomes for replication. This article summarizes and compares the integration machineries of three retroviruses that have increased our understanding of HIV-1/AIDS and the possibility of developing a safe retrovirus vector for human gene therapy. HIV-1, a retrovirus, is a significant pathogen for humans. After virus entry into the cell, reverse transcription of the retroviral RNA into DNA allows the formation of cytoplasmic preintegration complexes (PIC) that are subsequently transported into the nucleus. Though the pathway of infection is similar, the biological machineries required for nuclear transport of PIC are significantly different between HIV-1, murine leukemia, and avian retroviruses. These transport mechanism variances have aided the development of retrovirus vectors for delivery of genes into nonproliferating cells. The viral integrase (IN) within the PIC integrates the viral DNA into host chromosomes. The consequences of viral DNA integration are multifaceted. Sequencing of the human genome and viral DNA integration sites in the genome has shown that integration by HIV-1 favors actively transcribing genes. Murine leukemia virus prefers to integrate into or near transcription start sites while avian retrovirus shows no significant bias for transcription start sites and only a weak preference for active genes. The two- to threefold site selection preferences suggest that cellular proteins may influence the PIC to integrate into some host target sites. However, all three retroviruses can integrate their DNA into many different sites in all of the chromosomes. Reconstitution of synaptic complexes using purified IN and viral DNA substrates, possessing similar integration capabilities as PIC, has provided and will provide further molecular insights into the PIC. Understanding the biochemical machineries within PIC contributes valuable information for developing inhibitors to the HIV-1 IN. In particular, combinational drug therapy against different HIV-1 enzymatic targets have been successful at the clinical level and these therapies should be enhanced by the addition of IN inhibitors.

1 Retrovirus Replication and Integration Overview

1.1 Replication Pathway

Retroviruses are animal viruses ubiquitously present in nature. They comprise a large group of viruses that share a common mode of replication. Retroviruses attach to a specific receptor(s) on the cell membrane prior to entry into the cell (Fig. 1). After fusion and penetration of the virus into the cell, the viral single-stranded RNA genome located in the virion core undergoes reverse transcription yielding a double-stranded DNA copy of the viral genome. The viral DNA genome contained in the PIC (preintegration complexes) is transported into the nucleus and undergoes IN-mediated integration into the cellular genome. The integrated viral DNA genome is transcribed by cellular RNA polymerase II into messenger RNA that codes for viral proteins and also serves as virion genomic RNA. The virion RNA

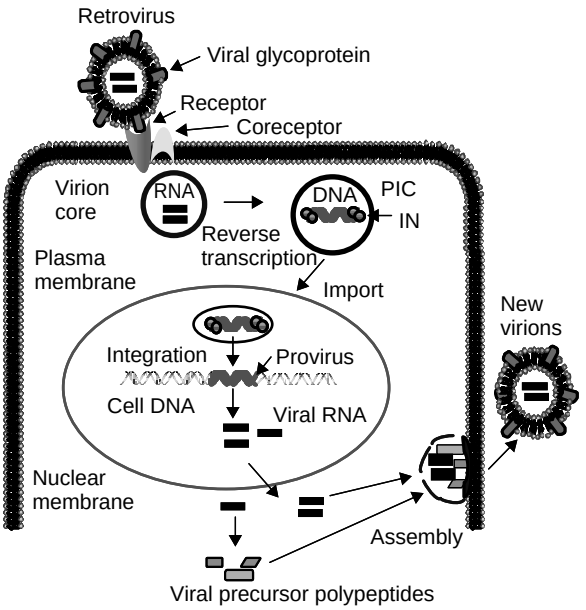

Fig. 1 Retrovirus replication cycle. The viral envelope glycoproteins (red structures) on retroviral particles attach to a specific receptor (multiple colors) on the cell surface leading to fusion and entry of the virion core into the cytoplasm. Sometimes viruses, like HIV-1, use a coreceptor (yellow) for fusion and entry into the cell. The viral RNA (two identical copies) in the virion core is reverse transcribed into double-stranded DNA. PIC are formed in the cytoplasm. IN (dimer, green circles) is located at the ends of the viral DNA in the PIC. These nucleoprotein complexes are transported into the nucleus by one of several different mechanisms depending on the particular retrovirus. The viral DNA is integrated by IN into the cell DNA, which is subsequently transcribed into viral RNA. The full-length viral RNA and spliced messenger RNAs are transported into the cytoplasm for protein translation. Viral proteins along with full-length RNA are assembled together as immature virus particles at the plasma membrane prior to release of new infectious virions. (See color plate p. xxxvii.)

is transported along with viral precursor polypeptides and cellular proteins to the cell membrane where immature, noninfectious virus particles are formed. The immature virus particles incorporate the plasma membrane from the cell as part of its membrane structure. During release of the immature particles, the viral protease processes the viral precursor polypeptides allowing the production of new infectious virus particles. These virions infect adjacent cells and the replication cycle is repeated. HIV-1, the infectious human retrovirus that causes AIDS, follows the earlier-mentioned pathway. Variations of the replication pathway exist within the large family of retroviruses, but all include reverse transcription of the RNA and

integration of the viral DNA genome into cellular DNA.

1.2 Consequences of Integration

Following integration of the viral DNA genome into the host chromosomes, the parasitic retrovirus utilizes host-cell machinery to self-replicate. In model systems, some proviruses do not adversely affect cellular genes because either the viral DNA integrates into nonessential regions of the chromosomes or because cellular mechanisms silence the provirus. In addition to being responsible for the production of new virus particles, the integrated provirus can exert multiple effects onto the infected cell. The integration process is essentially irreversible, although on occasion, host-cell mediated homologous recombination between the ends of the integrated DNA eliminates most or all of the viral internal genes. The remaining viral regulatory signals in the cell DNA can influence the transcription of cellular genes just as they do with the intact provirus.

Proviruses can produce mutations in cellular genes that affect the growth properties of the cell. Mutations in these genes result in either overproduction or alterations of normal growth control proteins modifying the replication properties of cells in culture. These abnormal cells can cause cancers when introduced into animal model systems. In some cases, integration of the viral DNA allows the simultaneous transcription of viral and cellular genes together into one messenger RNA that is subsequently packaged into virion particles. This "capturing" of cellular growth regulatory genes into infectious virions results in the transduction of cellular genes from infected cells to normal cells. Further discussion of the consequences resulting from viral DNA integration into various regions of the human chromosome will be presented later.

In summary, studying the genetic consequences of retroviral DNA integration into the host chromosomes of cells and the use of various animal model systems has enabled scientists to further their pursuit to understand human cancers. The initial discovery of the "capture" of regulatory cellular genes by retroviruses has prompted scientists to identify many cellular regulatory genes.

2 Retrovirus Preintegration Complexes

2.1 Physical and Molecular Properties

The virus particle is fused with the plasma membrane and it enters the cytoplasm of the cell (Fig. 1). During and following reverse transcription, the viral IN (integrase) associates with viral DNA to form a cytoplasmic nucleoprotein complex termed PIC (Fig. 1). They can be purified from the cytoplasm of virus-infected cells. The PIC are physically large having a sedimentation velocity of ~80 to 120 S. Within the PIC, IN first removes two nucleotides from the 3′ OH blunt-ended termini of the linear DNA genome. IN forms a stable association with the 3′ OH recessed ends of the viral DNA. PIC are then transported into the nucleus and IN inserts the two 3′ OH recessed DNA ends in a concerted fashion into the host genome (Figs. 2 and 3). A successful insertion of both viral DNA ends into the host DNA is termed *full-site integration*. This event can be mimicked *in vitro* by introducing purified PIC to an exogenously supplied DNA

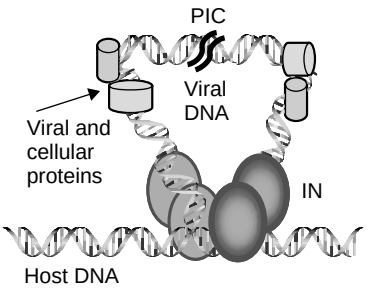

Fig. 2 Pictorial representation of how IN holds the two viral DNA ends together within PIC at the time of integration into the host genome. The entire linear viral DNA (10 kbp) is not drawn to scale as indicated by the dark lines in the viral DNA. The ends of the DNA are inserted by IN (illustrated as dimers) into the host or target DNA. IN recognizes specific *att* site sequences at the viral 3′ OH recessed DNA ends (Fig. 5). To produce the appropriate host-site duplications upon integration into the target DNA, the distances between the 3′ OH ends held together by IN in PIC must be in the ~16 to 20 Å range (Fig. 3). The number of IN subunits required at the ends of the viral DNA for full-site integration is unknown. A tetramer is shown for illustration purposes only. The three separate domains on these simple models of IN are also not drawn. Several cellular and viral proteins associated with the PIC are represented as gray and light blue objects. This model displays the viral DNA ends in a parallel fashion although the ends could also exist in an antiparallel fashion. (See color plate p. xxviii.)

target (Fig. 2). Molecular analysis shows that the 3′ OH ends act as nucleophiles that form a covalent linkage between the viral ends and the 5′ ends of the target DNA at staggered sites of insertion. The host-site DNA gaps and the 5′ two base-pair viral DNA that overhang at the viral-host junctions undergo repair by cellular enzymes, producing an integrated provirus (Fig. 3). The small size host duplications created at the integration site range from 4 to 6 bp and is therefore a function of the specific retrovirus IN (Fig. 3).

The mode of transport of cytoplasmic PIC across the nuclear membrane is different among retroviruses and thus affects their ability to integrate their DNA in either dividing or nondividing cells (Fig. 1). In addition to reverse transcriptase, IN, and the viral DNA genome, PIC contains several viral proteins. IN and these proteins, along with one or more cellular proteins (Fig. 2), appear to orchestrate the transport of PIC into the nucleus. HIV-1 and avian retroviruses have active transport mechanism(s) because they integrate their DNA in both nondividing and dividing mammalian cells. Avian retroviruses integrate their DNA less efficiently in nondividing mammalian cells suggesting a less efficient transport mechanism than HIV-1. Murine leukemia viruses can only efficiently transport their PIC in dividing cells when the nuclear membrane breaks down during cell division (passive transport).

In summary, the efficiency with which the earlier-mentioned three viruses infect nondividing cells correlates with their respective transport mechanisms across the nuclear membrane. Understanding the basis for these restrictions is of particular importance in the development of retroviral vectors for human gene therapy.

Retroviral vectors containing only the essential elements required for reverse transcription and integration can be constructed *in vitro*. These vectors have the same integration properties of normal retroviruses implying that they produce structurally analogous PIC. Most vectors (replication-defective) obtain their viral proteins needed for the initial transduction of cells from a retroviral packaging cell in trans, thus preventing additional rounds of infection. Some vectors have self-inactivating features that eliminate unwanted transcription from cellular genes. Pseudotyping vectors with either specific or nonspecific viral glycoproteins (Fig. 1)

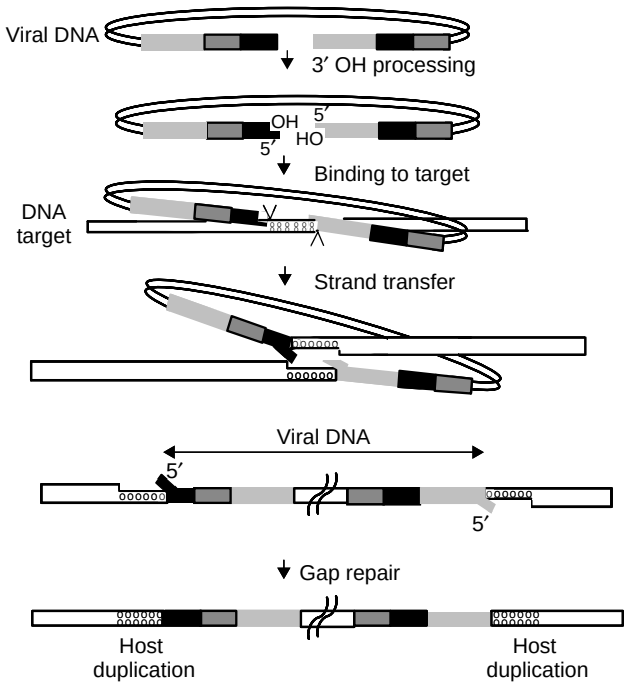

Fig. 3 Steps necessary for integration of the linear viral DNA genome into the host chromosome. The 10-kbp linear blunt-ended viral DNA is represented at the top. The different shaped areas on the viral DNA represent different viral regulatory regions including the terminal attachment sites for IN recognition. IN cleaves two nucleotides from the 3′ OH termini prior to binding to the target DNA. The target DNA represents either the host genome or exogenously added target. IN inserts the 3′ OH recessed ends into the target DNA at stagger sites upon integration. Host repair enzymes remove the 5′ 2-bp overhang on the viral DNA and fill in the gap between the host and viral DNAs. For avian retrovirus IN, the stagger sites are 6 bp apart on opposite strands of the target. The host-site duplications for avian, HIV-1, and murine leukemia retroviruses are 6, 5, and 4 bp, respectively. (Adapted from Brown, in: Coffin, J.M., Hughes, S.H., Varmus, H.E. (1997) *Retroviruses*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.)

can bestow on these vectors the ability to infect a variety of mammalian cells including human cells. Integration of a gene of interest carried by a retrovirus vector into nondividing or quiescent human cells would have a significant advantage over a vector requiring rapidly replicating cells. For instance, HIV-1 vectors can transduce differentiated neurons, quiescent hematopoietic stem cells, and macrophages in contrast to murine leukemia virus–based vectors, which require dividing cells.

In summary, the different integration machineries of retroviruses have an influence in which cells of a specific retrovirus

can efficiently replicate via integration of their DNA. Hopefully, further progress in the development of safe retrovirus vectors will allow treatment of certain genetic defects in humans.

2.2 Host Chromosomes Target Site Selection *In Vivo*

Understanding how retroviruses preferentially infect either dividing or nondividing cells is critical for selecting a particular retrovirus vector that is best suited for a specific protocol or experimental approach. Is it possible to direct the insertion of a gene of interest by a retrovirus vector into a specific region of the human chromosome? Sequencing of the human genome has provided a platform through which it is determined whether retroviruses integrate their DNA genomes into distinctive chromosomal regions as observed in yeast retrotransposons. These retrotransposons like retroviruses have reverse transcriptase and IN enzymes and therefore employ similar modes of replication. Recent findings demonstrate that avian retroviruses and murine leukemia viruses as well as HIV-1 do indeed integrate their DNA preferentially into distinct regions of the human chromosomes. HIV-1 favors actively transcribing genes, murine leukemia virus into or near transcription start sites, and avian retroviruses show no significant preference for transcription start sites but only a weak preference for active genes. However, proviruses exist in all of the human chromosomes in various regions even though these retroviruses have integration site preferences (~two to threefold) for certain regions. Therefore, it is currently not possible to construct a retrovirus vector that directs integration of a gene of interest specifically into a desired chromosomal site.

The details of retroviral integration selectivity remain obscure. For some yeast retrotransposons, tethering their PIC complexes to specific transcription factors directs the integration machinery to a specific transcription site in the yeast chromosome. Several cellular proteins interact with HIV-1 IN in the PIC. Although some of these have been identified, there is little genetic proof that one or more of these proteins participate in or mediate integration selectivity. The macroscale pattern (chromosomal location) consisting of the net distribution of integrated provirus arises from the microscale (DNA sequence level) selecting of IN-mediated integration. This microscale selectivity is likely due, at least in part, to assistance from cellular proteins. *In vitro* target site selectivity data suggests that purified IN may favor certain physical DNA structures and possibly sequences for integration. Analyses of hundreds of viral DNA integration sites at the sequence level will be necessary to provide a more accurate assessment of which mechanism(s) may direct the selectivity and specificity of the PIC. Finally, integration selectivity may have cell (like nondividing cells) and tissue type specific requirements, which may be discovered by further analysis of viral DNA integration sites.

2.3 Integration Properties of Isolated PIC *In Vitro*

Isolation and purification of cytoplasmic PIC from murine leukemia virus-infected cells first provided an idealized model through which to understand the biochemistry of retrovirus integration. Again, IN within purified PIC catalyzes the insertion

of the two viral DNA ends in a concerted fashion into an exogenously supplied target with high efficiency and fidelity (Figs. 2 and 3). Other genetic and biochemical studies of PIC have enhanced our appreciation of the complex structural and functional nature of the PIC. For example, PIC captures a variety of cellular proteins that appear to allow the PIC to traverse the cytoplasm and to be imported into the nucleus. Further studies of the PIC are also necessary for elucidating the roles of IN and these cellular proteins in selecting DNA sites for integration into the host chromosomes. For instance, recent studies demonstrate that when purified HIV-1 PIC select host target sites *in vitro* (Figs. 2 and 3), the ~two to threefold preference for active transcribing genes selectivity observed *in vivo* is not evident. The results suggest that the cellular environment and the unique interactions of the PIC with cellular proteins may be critical for integration selectivity.

3 Synaptic Complexes Capable of Full-site Integration

3.1 Reconstitution of Synaptic Complexes with IN and Viral DNA Substrates Capable of Efficient Full-site Integration

It is difficult to obtain sufficient quantities of PIC from virus-infected cells to manipulate at the molecular and biochemical levels. The reconstitution of synaptic complexes possessing properties associated with PIC requires purified IN, viral DNA substrates, and possibly other viral and cellular proteins. It is easy to mimic the viral genome in size and sequence through the synthesis of model linear viral DNA substrates using recombinant DNA technologies. Isolation of purified IN from virions has only been accomplished with avian myeloblastosis virus. Recombinant HIV-1, murine leukemia virus, and Rous sarcoma virus (RSV) IN have been expressed in bacteria and other expression systems. To date, only avian myeloblastosis virus and recombinant RSV IN have the capacity to efficiently assemble synaptic complexes holding the two viral DNA ends in close juxtaposition to catalyze full-site integration *in vitro* (Fig. 2). No other viral or cellular proteins are required with avian retrovirus IN for efficient full-site integration. The fidelity for producing the avian host-site duplications (6 bp) by purified IN (Fig. 3) is greater than 95% with the reconstituted synaptic complexes closely matching the fidelity observed *in vivo* with PIC.

Numerous laboratories using recombinant HIV-1, murine leukemia virus, and other retrovirus species IN have demonstrated that these recombinant IN essentially lack the ability to efficiently assemble synaptic complexes (Fig. 2) capable of full-site integration. Numerous nonspecific integration reactions occur whereby IN either inserts the viral DNA substrates into other viral DNA substrates (autointegration) or inserts only one end into the added target DNA (half-site integration). It appears that recombinant IN is for the most part defective in forming the necessary protein–protein interactions required to form a stable synaptic complex *in vitro*. The possibility exists that other cellular proteins are required for HIV-1 IN to form stable synaptic complexes *in vitro* (Fig. 2). Biochemical studies with isolated PIC from HIV-1 and murine leukemia virus highly suggested that LEDGF/75, a transcription activator, and BAF (autointegration barrier protein) are critical

components responsible for nuclear transport and stability of the PIC, respectively. Genetic evidence is necessary to firmly establish if these and other identified cellular proteins may have a role in the direct integration process of the three retroviruses described in this chapter.

In summary, the study of isolated PIC and reconstitution of synaptic complexes for studying full-site integration have limitations in comparison to the process *in vivo*. However, further molecular insights into retrovirus integration will also require study of these integration machineries at the biochemical level.

4 Retrovirus Integrase

4.1 Origin and Structure of IN

IN was first identified in virion core structures isolated from purified avian myeloblastosis virus as a homodimer of subunits having a molecular weight of 32 000 Da each (Figs. 1 and 4). In virus particles, IN is proteolytically processed by a viral protease from the carboxy-terminal domain of the precursor polypeptide for reverse transcriptase. HIV-1 IN is similar in size to avian IN, while murine leukemia IN has a molecular weight of 42 000 Da. Avian and HIV-1 IN contain three structural domains (Fig. 4). The N-terminal domain (~50 residues) contains a zinc binding region, promotes multimerization of IN, and is necessary for full-site integration. The central or catalytic domain (~200 residues) contains a highly conserved triad of amino acids comprising the conserved D,D(35)E motif, which is involved in aligning divalent metal ions for enzymatic activities. The catalytic domain binds to both, the very terminal viral *att* sequences (Fig. 5) and target DNA. The C-terminal domain (~35 residues) also binds to *att* site sequences (~7–10 nucleotides from the end) to DNA target sequences, as well as being involved in multimerization of IN subunits. The multiple modes of DNA binding to both viral *att* and target sequences by IN implies that multiple

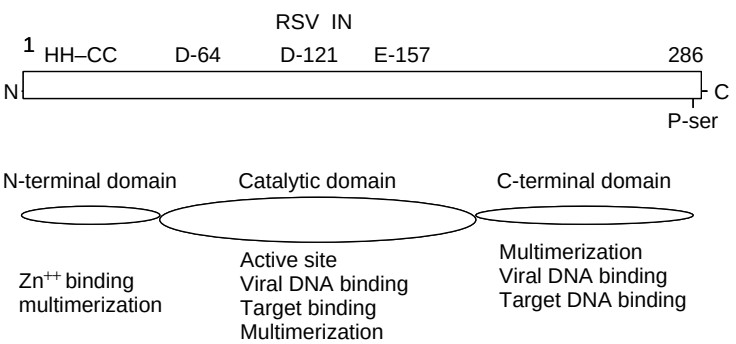

Fig. 4 Schematic representations of avian retrovirus IN. The Rous sarcoma virus (RSV) IN is 286 residues in length (top). The N- and C-termini are marked. The HH–CC (N-terminal domain), the conserved D,D(35)E motif (catalytic domain), and the C-terminal domain are indicated below the box. The functions associated with each domain are listed below. RSV IN is phosphorylated at Ser282. HIV-1 IN has a similar domain structure but is not posttranslationally modified.

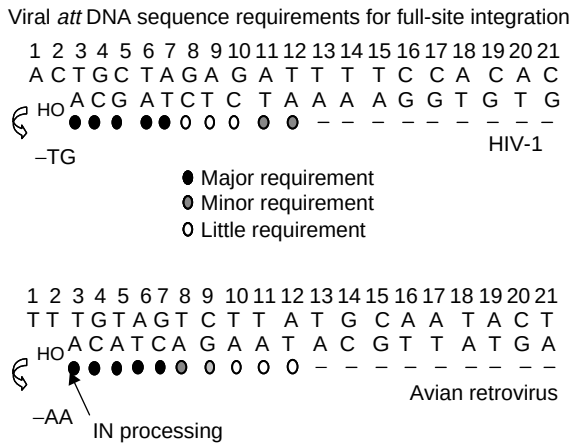

Fig. 5 Retrovirus *att* DNA sequence requirements for full-site integration. For simplicity, only one each of either the U5 or the U3 terminal sequences presence on linear viral DNA are shown. Both U3 and U5 *att* sites are necessary for full-site integration *in vivo*. HIV-1 U5 terminal sequences are shown on the top line. Below these sequences are circles indicating either a major, minor, or little requirement for IN enzymatic activities observed both *in vivo* and *in vitro*. The dash below each nucleotide indicates no requirement for IN activities. Murine leukemia virus terminal *att* sequences (not shown) are different than those shown for HIV-1. Similar requirements exist for murine sequences at these positions as shown for HIV-1 IN. The bottom line contains the U3 terminal sequences of avian retroviruses with the indicated requirement circles below each nucleotide. The arrow demonstrates IN cleaving two nucleotides from the blunt-ended viral DNA end. This cleavage occurs with all retrovirus IN.

subunits participate in full-site integration (Fig. 2).

The atomic structures of the three domains of IN have been determined individually or in two domain pairs. In solution as well as in crystal structures, the domains exhibit dimeric forms possessing huge surface areas for binding viral and target DNA, as well as for protein–protein interactions between the paired domains. Neither the structure of full-length IN nor IN bound to viral DNA *att* sites have been obtained.

4.2
Viral DNA–IN Synaptic Complexes

Investigators have exerted considerable effort to understand IN-viral DNA interactions in the hopes of designing drug molecules that inhibit HIV-1 IN. Investigations into the multidomain structures of IN (Fig. 4) provided molecular insights into IN activities, including 3′ OH processing, host target site selection, and strand transfer activities (full-site and half-site integration) (Fig. 3). For instance, IN cleaves specifically at the highly conserved CA dinucleotide in the viral *att* site (Fig. 5). IN exhibits no apparent binding preference for blunt-ended viral DNA ends over ends with other DNA sequences. Specificity appears to occur through conformational changes in IN that occur upon binding with the 3′ OH recessed ends. These conformational changes lock the complex into its proper "strand transfer configuration," permitting subsequent catalysis (Fig. 5). For full-site integration, IN specifically holds the two *att* viral DNA ends in close juxtaposition (Fig. 2) requiring only the very terminal *att* sites as shown in Fig. 5. The point of joining in the target DNA by two active sites on different dimers of avian IN are known to be spaced 6 bp apart on opposite strands (Fig. 3), that is, ~18 Å for helical DNA. On the basis of *in vitro* experiments, several models have been proposed for IN–DNA complexes. Most of the models are composed of an IN tetramer (dimer of dimers) with two ends of viral DNA with only two

active sites out of four involved in full-site integration. An octamer of IN has also been proposed for holding together the viral DNA ends in the presence of target DNA. In summary, intensive investigation of the biochemical machinery has revealed the complex associations of IN with viral DNA *att* sites both within PIC and reconstituted synaptic complexes.

5 Human Gene Therapy

5.1 Success and Failure

HIV-1, murine leukemia, and avian retroviruses have preferences in selecting host sites for integration into host chromosomes. This preference by PIC to integrate viral DNA into different regions of the human chromosomes has added to our understanding of host-site selection, possibly permitting the insertion of genes of interest using retrovirus vectors. Considerable caution is necessary as illustrated by the use of murine leukemia virus as a vector for insertion of a gene to correct human X-linked severe combined immunodeficiency. Several years after gene therapy, two of the eleven successfully treated children developed T-cell leukemia because the vector inserted next to a human T-cell regulatory gene (*LMO2*), led to aberrant transcription and expression of the gene. These human cases of gene therapy–induced leukemia have long been understood to occur in animal model systems where different types of cancers have been induced by various retroviruses, called *insertional mutagenesis*. Currently, besides human safety concerns, the size of the gene of interest capable of being packaged within retrovirus vectors is limited to ~8 to 10 kbp of sequences. In summary, future studies may permit the development of gene therapy devices that harness the unique power of retroviruses, providing safe treatment of a variety of genetic diseases in humans.

6 Ubiquitous Nature of Mobile DNA Elements

Retroviruses are not the only DNA elements that have the potential to transduce their genetic information from cell to cell. As previously stated, yeast retrotransposons have the ability to move within the same cell using a similar replication and integration cycle as retroviruses. Retrotransposons exist also in the genomes of a large variety of fungi, plants, and vertebrates. For example in plants, the genome of maize (corn) is composed of ~50% retrotransposon elements. Many mobile DNA elements that are either similar or dissimilar to retroviruses can modify chromosome structure and function. Some of these elements are active, although most are inactive and comprise a large percentage of animal and plant genomes. These elements move within the cell and do not require an exterior envelope structure associated with retroviruses (Fig. 1). We are now just beginning to understand the ubiquitous presence of mobile DNA elements in nature relative to the evolutionary paradigm.

7 Perspectives

The study of retrovirus integration has far-reaching effects on human health. Retroviruses are of particular concern

and interest because of the obvious HIV-1/AIDS pandemic, and because they can potentially serve as tools for human gene therapy. The current clinically proven therapy for HIV-1/AIDS consists of combinations of drugs known as *highly active antiretroviral therapy* (HAART) to the viral catalytic enzymes. The use of HAART directed against the HIV-1 reverse transcriptase and protease has effectively reduced the virus loads in HIV-1 infected individuals with concurrent reduction of diseases and death rates. The emergence of HIV-1 strains that are resistant to these inhibitors necessitates development of unique antiviral drugs. Inhibitors to IN are in human clinical trials. The biological and biochemical studies of PIC and IN of all three retroviruses have aided scientists in identifying these new IN inhibitors. Hopefully, the future will bring retrovirus vectors that can function safely in human gene therapy. More information on human gene therapy including retroviral and other viral vectors is available at http://www4.od.nih.gov/oba/rac/clinicaltrial.htm.

Acknowledgment

We would like to thank Jacob Zahm, Sapna Sinha, and Elena Lomonosova for their editorial comments.

See also Nucleic Acid Packaging of DNA Viruses; Vaccine Approaches for Protection Against HIV; Vaccine Research and Development Methodology; Vector Targeting in Gene Therapy; Vectors and Gene Therapy; Virus-Cell Interactions.

Bibliography

Books and Reviews

Bushman, F. (2002) *Lateral DNA Transfer, Mechanisms and Consequences*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Coffin, J.M., Hughes, S.H., Varmus, H.E. (1997) *Retroviruses*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Craig, N.L., Craigie, R., Gellert, M., Lambowitz, A.M. (2002) *Mobile DNA II*, American Society for Microbiology, Washington, DC.

Flint, S.J., Enquist, L.W., Racaniello, V.R., Skalka, A.M. (2004) *Principles of Virology; Molecular Biology, Pathogenesis, and Control of Animal Viruses*, American Society for Microbiology, Washington, DC.

Primary Literature

Brown, P.O., Bowerman, B., Varmus, H.E., Bishop, J.M. (1987) Correct integration of retroviral DNA *in vitro*, *Cell* **49**, 347–356.

Chiu, R., Grandgenett, D.P. (2003) Molecular and genetic analysis of Rous sarcoma virus integrase for concerted DNA integration, *J. Virol.* **77**, 6482–6492.

Hazuda, D., Anthony, N., Gomez, R., Jolly, S., Wai, J., Zhaung, L., Fisher, T., Embrey, M., Guare, J., Egbertson, M., Vacca, J., Huff, J., Felock, P., Witmer, M., Stillmock, K., Danovich, R., Grobler, J., Miller, M., Espesth, A., Jin, L., Chen, I., Lin, J., Kassahum, K., Ellis, J., Wong, B., Xu, W., Pearson, P., Schleif, W., Cortese, R., Emini, E., Summa, V., Holloway, K., Young, S. (2004) A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11233–11238.

Mitchell, R.S., Beitzel, B.F., Schroder, A.R.W., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., Bushman, F.D. (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences, *PLoS Biol.* **2**, 1127–1137.

Turlure, F., Devroe, E., Silver, P.A., Engelman, A. (2004) Human cell proteins and human immunodeficiency virus DNA integration, *Front. Biosci.* **9**, 3187–3208.

Vora, A.C., Bera, S., Grandgenett, D.P. (2004) Structural organization of avian retrovirus integrase in assembled intasomes mediating full-site integration, *J. Biol. Chem.* **279**, 18670–18678.

Revascularization of Ischemic Tissue: *see* Therapeutic Angiogenesis

Rhythmic and Temporal Processes in Biology

Russell G. Foster[1], *Stuart Peirson*[1], *and David Whitmore*[2]
[1] *Charing Cross Hospital, London, UK*
[2] *University College London, London, UK*

1 Introduction 418

2 The Properties of Circadian Rhythms 420

3 The Mechanisms of Circadian Rhythmicity 423
3.1 Drosophila 423
3.1.1 Entrainment 424
3.1.2 Oscillator 426
3.1.3 Output 431
3.2 Mouse 432
3.2.1 Entrainment 434
3.2.2 Oscillator 436
3.2.3 Output 444
3.3 Zebrafish 446
3.3.1 Entrainment 446
3.3.2 Oscillator 448
3.3.3 Output 449
3.4 Neurospora 451
3.4.1 Entrainment 451
3.4.2 Oscillator 453
3.4.3 Output 455
3.5 Arabidopsis 455
3.5.1 Entrainment 455
3.5.2 Oscillator 459
3.5.3 Output 461
3.6 Cyanobacteria 462
3.6.1 Entrainment 462

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

3.6.2 Oscillatorff 463
3.6.3 Output 467

4 General Conclusions and Future Perspectives 468

Bibliography 470
Books and Reviews 470
Primary Literature 472

Keywords

Action Spectrum
A method of analyzing the effect of different wavelengths of light on a biological response to characterize its spectral sensitivity and photopigment type.

Arabidopsis
Arabidopsis thaliana (mustard plant). A plant widely used in genetics research.

bHLH
Basic helix-loop-helix, a structural feature of many proteins capable of binding DNA.

Chromophore
The light-absorbing moiety of a photopigment.

Circadian Rhythm
A biological rhythm that persists under constant conditions with a period length of around 24 h. From the Latin "*circa*" and "*diem*," about a day.

Circadian Time (CT)
The subjective internal time of an organism under constant conditions. By convention, CT 12 corresponds to activity onset for a nocturnal species, whereas CT 0 designates activity onset for a diurnal species.

Clock Gene
A general term used to denote a gene involved in the normal functioning of the core oscillator of the circadian system (see text for details).

Clock-controlled Gene
A gene whose expression is regulated directly by the core oscillator mechanism.

Cyanobacteria
Photosynthetic bacteria, often called *blue-green algae*.

DD
Abbreviation for constant darkness.

Diurnal
An activity or process that occurs during the daytime (light).

Drosophila
Drosophila melanogaster (fruit fly). A common model organism in genetics research.

E-box
A nucleotide motif, CACGTG, involved in enhancing the expression of many clock genes.

Eclosion
Hatching or emergence of an adult insect from its pupal case.

Entrainment
The process by which a biological oscillator is phase locked to a rhythmic environmental stimulus such as the light–dark cycle.

Free-running
The endogenous rhythm exhibited by a circadian system under constant conditions.

LD
Abbreviation for a lighting regime composed of alternating periods of light and dark.

LL
Abbreviation for constant light.

Masking
The phenomena whereby an external factor directly affects the expression of an overt circadian rhythm.

Neurospora
Neurospora crassa (bread mold). A filamentous fungus commonly used in genetic research.

Opsin
The protein component of animal photopigments. Opsins utilize a vitamin A chromophore and possess a characteristic absorbance spectrum.

Oscillator
A system capable of producing a regular fluctuation of an output around a mean. In circadian biology, an oscillator often refers to the molecular mechanism within a cell capable of generating a self-sustained rhythm with a period ~24 h.

Overt Rhythm
An observable rhythm that is directly or indirectly regulated by the circadian clock.

Pacemaker
An anatomical correlate of an oscillator, capable of both maintaining its own oscillation and entraining additional oscillators.

PAS Domain
Sequence motif found in many clock proteins, often associated with protein–protein interactions.

Period
The regular time interval between phase points (reference points) on an overt rhythm.

Peripheral Oscillator
A term that has been used to refer to an oscillator located in tissues that are capable of regulating local circadian rhythms. Also referred to as a *slave oscillator*.

Phase
A specific reference point in the cycle of a rhythm; for example, the daily onset of sleep.

Phase shift
A single, persistent change in phase brought about by the action of a zeitgeber.

Photoentrainment
The entrainment of the circadian system by a light–dark cycle.

Photoperiod
The period of light and dark over a specific period, usually 24 h.

Photopigment
A molecule that is capable of transducing the absorbance of a photon into an intracellular response.

SCN
Suprachiasmatic nuclei. Paired nuclei within the basal hypothalamus that function as the circadian pacemaker in mammals.

State Variable
A term used in mathematical modeling of oscillators, referring to a quantity that changes with time. More generally, a term used to denote a variable essential for defining the state of a system. Sometimes applied to the essential components (genes/proteins) required to generate a circadian oscillator.

***Tau* (τ)**
The endogenous period of a free-running biological rhythm.

Temperature Compensation
A characteristic of circadian rhythms, whereby changes in temperature do not significantly alter the period. A Q_{10} close to 1.

Zeitgeber
"Time-giver," from the German, referring to an environmental signal capable of entraining a clock-driven rhythm.

Zeitnehmer
"Time-taker," from the German, referring to an output from the clock that can feed back to regulate the input pathway. The action of light on the clock is often gated by the clock itself. The rhythmic regulation of an input pathway can generate a rhythmic entraining stimulus even under constant conditions.

We are ruled by time and we need to know the time to tell us what to do. But the clocks that instruct us to wake, eat, and go to bed are unnatural. In our modern world, we are being driven by the energy states of an electron in the cesium atom and use machines to hack our day into hours, minutes, and seconds. Yet, despite the atomic clock, our bodies answer to another more persistent beat that probably started to tick shortly after life evolved. Embedded within the genes of us, and almost all life on earth, are the instructions for a biological clock that marks the passage of approximately 24 h. Biological clocks drive or alter our sleep patterns, alertness, mood, physical strength, blood pressure, and every other aspect of our physiology and behavior. Under normal conditions, we experience a 24-h pattern of light and dark, and our clock uses this signal to align biological time to the day and night. The clock is then used to anticipate the differing demands of the 24-h day and "fine-tune" physiology and behavior in advance of the changing conditions. Body temperature drops, blood pressure decreases, and tiredness increases in anticipation of going to bed. While, before dawn, metabolism is geared up in anticipation of increased activity when we wake. The past decade has witnessed remarkable progress in the understanding of circadian rhythms in many different organisms. Much of what we know of the molecules that make up our biological clockwork has been learnt from the fruit fly (*Drosophila*) and mouse. It is remarkable that species separated by hundreds of millions of years of evolutionary history are united by the ways in which they generate biological time. The understanding of the molecular basis of circadian rhythms is one of the first success stories arising from the genome sequencing projects, and is currently one of the best examples we have of how genes and their protein products give rise to complex behaviors.

1
Introduction

The term "chronobiology" is derived from the Greek (*chronos* for time, *bios* for life, and *logos* for study), and is used to denote the study of "biological clocks" within living organisms. These internal clocks exist in all eukaryotes and have been identified in some prokaryotes. Indeed, the possession of such a clock has been regarded by many as one of the signatures of a living organism. Biological clocks provide an organism with an internal representation of a timed and predictable environmental event such as the 24-h rotation of the earth and the consequent rising or setting of the sun, or the annual change of the seasons. Knowledge of these predictable events allows an organism to prepare in advance of the changed circumstances, and so avoid the dangers and exploit the advantages of the "new" environment. A biological clock can be considered to "fine-tune" an organism to its temporal niche. This advanced warning provided by the clock is necessary because different physiological or behavioral states cannot be switched instantly. For example, the transition from sleep to wake requires several hours in most mammals and involves the complete realignment of almost every biological system and process. If an organism waited to trigger its biology in response to a major change in environment, then valuable time would be wasted while physiology was realigned to generate an optimal performance. Peak performance is demanded by natural selection, where even modest advantages are translated into long-term evolutionary success. Thus, almost every cellular, physiological, or behavioral output is modulated to a greater or lesser extent by a chronobiological process. For the experimentalist, this presents a practical problem. We often undertake our studies at particular times of the day or night, and as a result, the data collected may not provide an accurate representation of events over the whole day. Furthermore, data collected at one time of day may not be comparable with data collected at other times. As shown in Fig. 1, body temperature, hormone levels, reaction times, alertness, and performance, and even our responses to medical treatment all show marked differences depending upon the time of day.

The constant temporal adjustment of physiology by an internal clock is somewhat at odds with the original concept of homeostasis as proposed by Claude Bernard and Walter Cannon. For them, and the several generations that followed, constancy was everything and deviations away from a "set-point" represented a breakdown in a homeostatic mechanism rather than an adaptive response to a changing environment. The teaching of biology and medicine has been dominated by the concept of an unchanging internal environment. Medical students have been expected to learn a vast range of physiological parameters that described the "average human" over a narrow temporal range. Blood pressure in a healthy individual is considered to be around 120/70, and yet at 06 : 00, it can drop as low as 60/30 in some healthy individuals. Similarly, body temperature is not always set at 37 °C, it reaches its peak around 20 : 00 and then falls by ~1 °C or more between 04 : 00 and 06 : 00 in the morning. Certain hormones, such as melatonin or growth hormone are very low during the day but are elevated at night, and the incidence of epileptic seizures increases dramatically at night (Fig. 1).

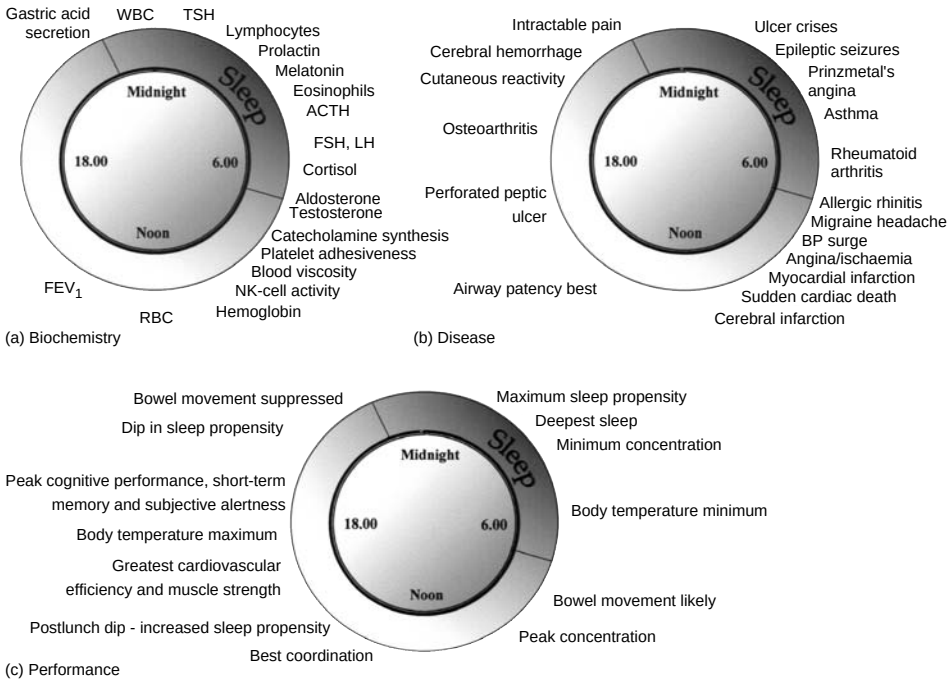

Fig. 1 Illustration of circadian rhythms in various aspects of human physiology, including (a) biochemistry, (b) disease incidence, and (c) performance. Times of peak incidence are approximate. Data taken from Smolensky, M.H. (1996) Chronobiology and chronotherapeutics. Applications to cardiovascular medicine, *Am. J. Hypertens.* **9**, 11S–21S and Carrier, J., Monk, T.H. (2000) Circadian rhythms of performance: new trends, *Chronobiol. Int.* 17, 719–732.

These lists of daily variations are endless, and even amusing, the preferred time for human sexual behavior in a recent survey reported in a popular magazine was 22 : 00. But what is often missed is that these lists contain critical information relating to the potential health and wealth of our species. Jürgen Aschoff, in his pioneering studies on human biological rhythms, made the point that an awareness of these changes is not only important for the design of specific clinical treatment paradigms but can also be used to help gauge the likelihood of accidents and performance-related mistakes in groups as diverse as nightshift workers, nurses, and fighter pilots.

After almost 100 years of detailed investigation, the study of biological clocks has come of age, and clock-driven changes in physiology are being incorporated into the mainstream teaching of biology and medicine. Although chronobiology has not yet been entirely embraced by the medical profession, clock-driven changes are no longer ignored. The concept of time-dependent variables is replacing the "fixed set-points" of former years, and while this knowledge is welcomed at one level, the life of the health care worker has become even more demanding. Another dimension of the body has to be taken into consideration – body time.

2
The Properties of Circadian Rhythms

The first person to study biological rhythms in an experimental context was the French astronomer Jean Jacques Ortus de Marian. His interest in the rotation of the earth led him to ask why the leaves of plants move up and down during each day in time with the 24-h rotation of the earth. In 1727, he placed a Mimosa plant in a cupboard and isolated it from the daily light–dark cycle, and noted that the plant still moved its leaves so that they drooped when it was night and were raised up during the day even though the plant had no direct contact with the light–dark cycle. He suggested that the plant had to have some internal or endogenous knowledge of the time of day. Some 30 years later, Henri Louis Duhamel reasoned that the Mimosa plants might have been responding to the daily change in temperature and so repeated the experiment in the constant light and temperature conditions of a salt mine. Duhamel found the same responses as de Marian, rhythmicity persisted under constant conditions. Approximately 100 years later in 1832, a Swiss botanist Alphonse de Candolle made the observation that under constant darkness the period of leaf movement was not exactly 24 h and varied slightly between different plants. Although this early work on *Mimosa*, and the movement of other plants, was pioneering, the rigorous investigation of biological rhythms began in Germany, with the studies of Erwin Bünning in the 1930s. Bünning worked on the bean plant *Phaseolus* and showed that leaf movement oscillated with a mean period of 24.4 h in constant darkness, and established one of the critical properties of biological clocks; when isolated from environmental cues, they persist with a period close to but never exactly 24 h. This near-24-h oscillation under constant conditions is known as the *free-running* rhythm of the clock. For example, the free-running period of a mouse is approximately 23.7 h, and so its activity will start a little earlier and earlier every 24 h under constant conditions. From the 1930s, free-running rhythms close to 24 h were demonstrated in very many different animals and plants, so that by 1959, a collective term was needed to describe such phenomena. Franz Halberg suggested that they should be called circadian (circa about, diem a day) rhythms, and the term has stuck. Circadian rhythms are not the only biological clock–driven rhythms observed in nature. Additional rhythms have been defined on the basis of their period under constant conditions and relatedness to environmental cycles. For example, circannual rhythms have periods of approximately one year, and have been best described in hibernating mammals such as the golden mantled ground squirrel, and circalunar or circatidal rhythms have periods of approximately 30 days and are found in many intertidal organisms. Our understanding of these other "circa" rhythms extends a little beyond the knowledge of their existence, and so all further discussion of biological clocks in this article will be confined to those generated by a circadian clock.

In parallel with the demonstration of free-running rhythms, work in the 1950s by Karl von Frisch who studied bees and Gustav Kramer working on migratory birds showed that animals must utilize an internal clock. Both studied time-compensated sun compass navigation. Bees and birds will orientate their movements by using the sun as a navigational reference point. Bees will forage for nectar and locate different plants

at specific times of the day, and birds will migrate at different times of the year. On average, the sun moves across the sky at 15° per hour, and both groups of animals were shown to consult an internal clock to compensate for this movement. For example, in the case of birds migrating north in the spring, they will orientate so that the sun is 90° to the left at 09 : 00, straight ahead at 12:00, and 90° to their right at 15:00. For reasons that are not exactly clear, these studies showing time-compensated orientation had much more of an impact on biologists, convincing of the existence of endogenous clocks than the demonstration of free-running rhythms. By the mid-1950s, the stage was set for a more formal description of circadian rhythmicity, and onto this scene emerged a man of remarkable intellect, Colin S. Pittendrigh. He not only pioneered the study of circadian rhythms in the fruit fly *Drosophila* but also applied his reasoning to a broad range of different animal models. Pittendrigh's work examined the nature of free-running rhythms, but his major contribution was that he went on to characterize/identify two further fundamental properties of all circadian clocks; entrainment and temperature compensation.

Biological clocks are temperature compensated, that is, their period is more or less constant irrespective of environmental temperatures. The clock of ectotherms would be useless if it were not. Circadian clocks do not obey the Q_{10} rule, a rise of 10 °C does not double the period of the oscillation. This issue of temperature compensation was a major point of controversy in the early days of biological clock research. Researchers such as Frank Brown argued that no biological process could possibly act independently of temperature and the constancy of the circadian period under different temperature conditions must mean that these rhythms were driven by some unknown environmental factor related to the earth's rotation. One of Pittendrigh's experiments on this topic showed that *Drosophila* would always emerge from their pupal case (eclosion) at dawn irrespective of environmental temperatures. Although his observation has now been duplicated in many different species using multiple circadian outputs, our understanding of how temperature compensation is brought about is little more than guesswork. By contrast, the mechanisms that adjust biological clock time to environmental time (entrainment) are far better understood.

A biological clock is of any use only if internal clock time is adjusted to environmental time. The classic example of a mismatch between biological and environmental time is "jet lag." In the natural environment, many factors could set or "entrain" circadian rhythms to a specific phase of the 24-h rotation of the earth. Light, temperature, food availability, or even social contact could indicate the time of day. However, only those factors that provide the most reliable indicator will be selected by evolution to entrain circadian clocks, and for most species, the 24-h change in the quality of light at dawn and dusk provides this signal or zeitgeber (time signal). In the absence of light, other signals such as temperature or food can act as zeitgebers, but if animals are exposed to conflicting signals such as light and temperature, light will invariably be selected as the zeitgeber.

Pittendrigh was one of the first to make a systematic study of entrainment. He explored the effects of short pulses of light on free-running rhythms in a

variety of animals, including *Drosophila*, maintained in constant darkness. He showed that light had different effects on the clock at different phases of the circadian cycle. As the animals were kept under constant darkness, he used the phase of activity of the animal to determine the position of the clock. In a nocturnal species, activity onset was considered to indicate the beginning of the night and designated circadian time (CT) 12 (the subjective night would span CT 12–24), while in a diurnal species, the start of activity was considered to signal dawn and designated CT 0 (the subjective day would span CT 0–12). Pittendrigh observed that light pulses given to free-running animals during their subjective day would have no marked effect on the clock, while light during the first half of the night (CT 12–18) would delay the clock – the animal would start activity later the next day. By contrast, light given during the second part of the night (CT 18–24) would advance the clock – activity would start earlier the next day. Pittendrigh studied a range of both nocturnal and diurnal species, and remarkably, all showed the same basic "phase response curve" (PRC) to light (Fig. 2), although the precise shape would vary between species.

In all organisms studied to date, light around dusk will delay the clock and light around dawn will advance the clock, and in this way, activity is broadly aligned to the expanding and contracting light–dark cycle throughout the year. The PRC model of entrainment has been helpful in searching for entrainment mechanisms and the molecular components of the circadian system (see below). However, it does not provide a complete description of entrainment. We do not fully understand how light intensity, wavelength, the phase of exposure, the length of the free-running period, the magnitude of the advances and delays, or even how photic and nonphotic cues might interact to generate stable entrainment.

The existence of circadian rhythms and their basic properties (persistence under constant conditions, temperature compensation, and entrainment) were established by the 1970s, and interest began to shift to more mechanistic questions. If there is a clock, where is it? How does it work? How does light synchronize the clock to local time? Is there just one clock? How did they evolve? The answers to these questions are still emerging and are outlined in the following pages. Progress over the past few years has been spectacularly rapid, and currently, there is perhaps no better example of how patterns of gene expression and protein interactions generate a specific set of behaviors in highly diverse groups of organisms.

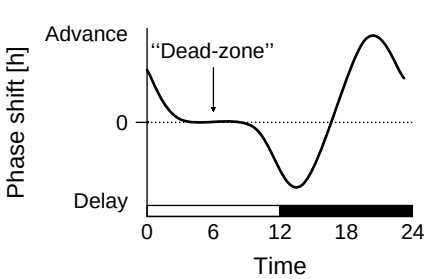

Fig. 2 Phase response curve of the circadian system, illustrating the response to light administration at different phases of the circadian cycle. Advances correspond to an earlier activity onset, whereas delays correspond to a later activity onset. The "dead zone" represents a period of relative insensitivity to light.

3
The Mechanisms of Circadian Rhythmicity

Early on in the hunt for circadian mechanisms, a conceptual framework emerged that compartmentalized the circadian system of all organisms into three basic elements: the input or entrainment pathway, the core circadian oscillator, and an output pathway (Fig. 3). While this depiction has turned out to be overly simplistic, largely because different molecules often span an input/oscillator or oscillator/output function, it has nevertheless provided a valuable model to develop experimental questions and compare the various elements of circadian systems in different organisms. For this reason, we have organized the information below using the entrainment > oscillator > output format, as diagrammed in Fig. 3, and have considered the circadian system of six model organisms, *Drosophila*, mouse, zebrafish, *Neurospora, Arabidopsis,* and cyanobacteria. We have selected these species to provide taxonomic breadth to the discussion of how the molecular clock might be generated and because rightly or wrongly, these species are the focus of genomic analysis.

3.1
Drosophila

Drosophila has proved to be a critical organism in the study of many biological processes, but few more so than the circadian clock. Studies in the fly have formed the historical backbone of most of what we know about circadian clocks, and, as such, no review can cover all of the contributions made by this organism. The power of *Drosophila* forward genetics and the use of mutant screens to identify molecules involved in a complex biological process, about which one has no *a priori* knowledge, is best exemplified by the use of the fly to examine clock function. Most of the critical clock genes in animals have been identified in this organism, as the following sections demonstrate. Furthermore, once a gene has been isolated, research on *Drosophila* has helped frame a set of questions that have been used by the whole field of circadian biology to help establish whether a gene encodes a real component of the clock. These questions can be summarized as follows: (1) Do mutations in the gene cause changes in period or arrhythmicity of output rhythms? (2) Does the abundance or activity of the gene oscillate with the same period as the overt rhythm? (3) Does the constitutive expression of the gene result in arrhythmicity of the overt rhythm? (4) Do entrainment signals modify levels or patterns of gene expression? (5) Can the protein product alter the expression of its own gene? As we shall discuss below, many "clock genes" do not fulfill all of these criteria, but addressing these questions has provided a useful

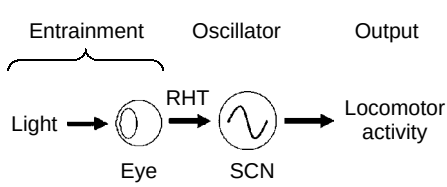

Fig. 3 Basic model of circadian system. A schematic depiction of the components of a circadian system, including an environmental input, an oscillator, and clock-controlled output. The mammalian system is shown as an example, with light-entrainment signals detected by the eye, providing an input to the SCN oscillator via the retinohypothalamic tract, and the SCN driving circadian patterns of locomotor activity.

experimental framework for defining their possible function(s).

3.1.1 Entrainment

A clock is not a clock unless it can be set to local time. So how does the *Drosophila* clock detect light? In a larger animal, such as a rodent, one would begin to answer such a question by performing lesion studies, physically removing the structures likely to be involved. In a sense, this is also exactly what one does with *Drosophila*, but instead of using a scalpel, one uses the identification of specific mutants. The identification of mutant flies that lack or have modified circadian light responses is a very powerful approach. This approach is made even stronger when combined with another powerful tool for the examination of light responses in general, the generation of action spectra. All photopigments have distinct light absorbance characteristics, peaking at a particular wavelength, and acting as a "spectral fingerprint" that can be used to distinguish between the various types. By determining the specific wavelengths that optimally regulate/activate a particular response, usually entrainment or phase shifting in a circadian experiment, an action spectrum is generated and compared to the absorbance profiles of known photopigments. This comparison is a critical step in identifying the light-detecting molecules involved in a particular pathway.

Action spectra have been generated in *Drosophila* from as early as 1969 in *D. pseudoobscura* through to 2001. Over this 32-year period, a range of responses have been examined, including phase shifts of eclosion and locomotor behavior, and a surprising range of peak photosenstivities have been reported. Some of the reasons for this variation will become apparent later in this section. An apparently straightforward question might be, does *Drosophila* use an opsin/vitamin A- or non-opsin-based photopigment? An examination of phase delays in the eclosion rhythm showed a number of peak photosenstivities, mainly in the "blue" at 457 nm, but also around 375, 435, and 473 nm. The lack of a typical "smooth" opsin/vitamin A action spectrum was more suggestive of a flavoprotein response. This argument was supported by studies in which flies, deprived of carotenoids in their diet and showing a marked decline in the sensitivity of normal visual light responses, failed to show any significant alteration of their circadian eclosion. By contrast, when action spectra were generated for phase shifts of the locomotor rhythms, peak sensitivities were recorded around 500 nm, and these curves extended into the "red" region of the spectrum. These data fit more precisely with results that would be predicted if an opsin/vitamin A–based photopigment was involved in the circadian light response.

The results for the wild-type fly are surprisingly complex, with data supporting a role for both opsin- and non-opsin-based pigments. Some of this complexity might be due to the wide range of light-detecting structures found in insects, and specific mutants, lacking particular photic structures, will be required to clarify this situation. Insects like *Drosophila* possess two large compound eyes, as well as a variety of extraocular photoreceptive structures including several ocelli, and the relatively recently described H-B eyelets (after Hofbauer and Buchner). It may be likely that additional, as yet unidentified, light-responsive structures also exist. The H-B eyelets are of considerable interest. These structures, each composed of four photoreceptor-like cells and expressing the *Drosophila* rhodopsin gene *Rh6*, as well as

arrestin, have direct projections to specific ventral lateral neurons (LNv) within the brain of the fly. These LNv cells are of particular importance in the *Drosophila* circadian system, as they are clearly responsible for generating the rhythm in locomotor activity, and as such are seen as the "master" pacemaker within the fly (see below). This direct link between master clock and photoreceptive structure, though correlative, is strongly suggestive of the fact that these H-B eyelets may play a role in clock entrainment.

The results from several *Drosophila* mutants help confirm the importance of these H-B eyelets. The mutant fly *sine oculis*^{1} lacks compound eyes, but retains the extraocular photoreceptive structures, the H-B eyelets. The action spectrum from these mutants revealed two distinct peaks of sensitivity at 420 and 480 nm. An additional loss of function mutant called *glass*^{60j} (as well as the double mutant *sine oculis*^{1}:*glass*^{60j}) produces an animal that lacks all compound eyes and H-B eyelets. Action spectra performed on these mutants produced only one peak of sensitivity at 420 nm. These experiments clearly demonstrate that the H-B eyelets are involved, in part, in light-induced phase shifts, responding to light primarily at 480 nm. This peak sensitivity is quite close to the absorbance of 515 nm reported for the *rhodopsin 6* (*Rh6*) gene found to be expressed in these cells. Remarkably, however, the glass mutant is still able to show light-dependent phase shifts in activity, even though it lacks all known ocular and extraocular photoreceptive structures. There must be additional light-responsive cells in the fly and these must have a peak of sensitivity in the region of 420 nm.

An apparent answer to this problem was solved by the discovery of the blue light–absorbing proteins, the cryptochromes, and the isolation of another *Drosophila* mutant *cry*^{baby} (*cry*^{b}). The cryptochromes are flavoproteins, using a pterin or flavin (FAD) as a cofactor, which appears to act as a primary chromophore, absorbing light primarily in the "blue" wavelengths. These molecules are closely related to the DNA repair enzymes, the photolyases, and as such have not only an ancient evolutionary history but are expressed widely in a number of cell types. The *cry*^{b} mutant fly was discovered as an arrhythmic animal in terms of *period* and *timeless* gene expression (see Sect. 3.1.2), as measured in *period*-luciferase transgenic animals. This lack of rhythmicity could represent a defect in their central clock mechanism, or a lack of entrainment (due to lack of photosensitivity), and subsequent asynchronous clock gene oscillations within the animal. These flies also cannot phase shift in response to short light pulses and fail to become arrhythmic in constant light, as do wild-type flies. Furthermore, these behavioral defects in the *cry*^{b} flies can be partially overcome by specifically overexpressing wild-type *cry*$^{+}$ in the lateral neurons. This important result shows that these neurons themselves are directly light responsive, and that cryptochrome can act intracellularly as part of a local photoentrainment pathway. However, these mutant flies are still capable of entraining to a light–dark cycle, as are the molecular oscillations within the lateral neurons. Also, cryptochrome clearly appears to have a role in the mechanism of peripheral oscillators outside of the lateral neurons (see below). Cryptochrome is clearly not the only element of the photosensory pathway leading to entrainment of the circadian clock. Only when the double mutant *glass*^{60j}:*cry*^{b} is generated, are totally clock-blind animals produced.

These double mutants are not capable of showing phase shifts or entrainment, as measured by a variety of circadian clock parameters. The amazing conclusion is that *Drosophila* use a range of independent mechanisms to set the circadian clock. Clearly, a cryptochrome-dependent mechanism within the lateral neurons is capable of responding to light and adjusting the clock. Also, the H-B eyelets respond to light, probably using an opsin/vitamin A–based molecule (*Rh6*). Furthermore, there also appears to be an influence of photoreceptors within the compound eyes on resetting. These multiple pathways into the clock give *Drosophila* great flexibility in terms of their circadian light response, and may extract much more information about the environmental lighting conditions than we initially expected. Different pathways appear to entrain different aspects of the circadian system, with, for example, the cryptochrome-dependent pathway being mainly involved in entraining the evening peak of activity, whereas retinal and extra retinal structures primarily influence the morning peak. One major advantage of such a system is that these animals may be able to determine more than simply the presence or absence of light, but actually sense whether they are seeing "morning" or "evening" light, a very useful feature for a clock-related photosensory system.

3.1.2 **Oscillator**

The first mutant screens in *Drosophila* were started as far back as 1968 in Seymour Benzer's laboratory, and over the following 35 years, forward genetic efforts have provided us with most of what we know about the clock mechanism. The first circadian fly mutant was isolated in a pupal eclosion screen and identified as flies that either hatched randomly with respect to the light–dark cycle (aperiodic) or emerged earlier or later than expected from their pupal cases. These mutations mapped to the same locus, named *period* or *per*, and when placed into constant darkness, three distinct mutant alleles were identified, per^S (Short) with a free-running period of ~19 h, per^L (Long) with a period of ~29 h, and per^0, which are aperiodic.

It was a further 16 years until the *period* (*per*) gene itself was isolated and a couple more until the sequence was published. The PERIOD (PER) protein contains what has become known as a PAS domain, named after the proteins in which they were first identified: *Drosophila* PER, the human aryl hydrocarbon receptor nuclear translocator (ARNT) and the *Drosophila* single-minded protein (SIM), and is a stretch of amino acids thought to be involved in protein–protein interactions. The PER protein, therefore, has a partner as we shall see below. PER in *Drosophila* lacks a basic helix-loop-helix (bHLH) DNA binding domain, and as such, does not appear to be a DNA binding/transcription factor. Oscillations of protein and mRNA were observed with an interesting 4- to 6-h delay between the peak of transcript and the subsequent peak in PER protein. Furthermore, constitutive overexpression of the PER protein inhibited cycling of the period transcript, and so a negative feedback loop between protein and gene expression was strongly suggested.

PER was clearly an important and central clock component, and it remained the only one for quite some time. Then, finally, the protein partner of PER was discovered in 1995. Gekakis and colleagues isolated a protein called *TIMELESS* (TIM) through use of a yeast two-hybrid screen, in which PER-interacting proteins could be identified. At about the same time, the *timeless* (*tim*) gene was positionally cloned

and sequenced by a second group as part of an analysis of the *timeless* mutant fly, which showed a loss of behavioral and *per* gene rhythmicity. Then there were two, but how the clock might work still required the isolation of at least three more critical genes.

For once, *Drosophila* lagged behind the mouse in the identification of novel clock components. In 1997, the first vertebrate clock gene, *Clock*, was cloned following a large-scale mutant screen in the mouse (see Sect. 3.2.2). Within a year, however, a homolog of mouse *Clock, dClock (Clk)*, had been isolated as part of an ongoing large-scale clock mutant screen in *Drosophila*. The homozygous *dClock* mutant flies were arrhythmic in terms of locomotor behavior and eclosion, and perhaps, more interestingly, showed very low levels of both PER and TIM protein. As in the mouse CLOCK protein, dCLOCK (CLK) is a bHLH transcription factor, which also possesses a PAS domain. A protein partner had already been discovered for mouse CLOCK, called BMAL1, and it was not long before an equivalent gene was identified in *Drosophila*. The *Drosophila* homolog is known as *cycle* (*cyc*) and again was identified as an arrhythmic fly in a *Drosophila* locomotor screen. As with mammalian BMAL1, CYCLE (CYC) is a bHLH PAS domain–containing protein that forms a heterodimer with CLK. The *cyc* mutant also shows reduced levels of TIM and PER protein expression. Considerable similarity is clearly apparent between the putative clock mechanism of both *Drosophila* and mammals (see Sect. 3.2). CLK and CYC bind to regulatory elements, called E-boxes (CACGTG), within the upstream promoter regions of *per* and *tim*. This positive transcriptional activation increases the transcript levels of both *per* and *tim*, and explains the low levels of both seen in the *Clk* and *cyc* mutants.

One additional *Drosophila* clock mutant was also discovered in 1998. Three alleles of the *double-time* (*dbt*) mutant fly were identified, with a short- and long-period form (dbt^S and dbt^L). The third allele dbt^P was lethal in the pupal stage, but eliminated oscillations of PER and TIM proteins in the larva. Mapping and cloning of this mutation revealed it to be a homolog of mammalian casein kinase I epsilon. In the dbt^P mutation, PER protein accumulated in a hypophosphorylated form, suggesting that DOUBLE-TIME (DBP) might well be involved in PER phosphorylation and potentially in its degradation.

How might the clock components, described so far, actually fit together to generate a circadian oscillation? Though additional clock mutants have been discovered, some of which will be discussed below, the core mechanism can be deduced from research on the elements described above (Fig. 4). The *per* gene appears to be activated in the afternoon, as a consequence of the binding of the positive transcriptional activators CLK and CYC, acting as a heterodimer on E-box elements within the *per* promoter. However, the PER protein does not begin to accumulate for some time. The reason for this is believed to be a consequence of the action of the casein kinase, DBT, which phosphorylates PER monomers and targets them for proteolytic degradation. Recently, an F-box/WD40-repeat protein, called *SLIMB*, has been shown to interact in a preferential manner with hyperphosphorylated PER and target it for ubiquitin-proteosomal degradation. PER is protected from this degradation only when sufficient amounts of the TIM protein begin to accumulate. The PER-TIM heterodimer then begins to

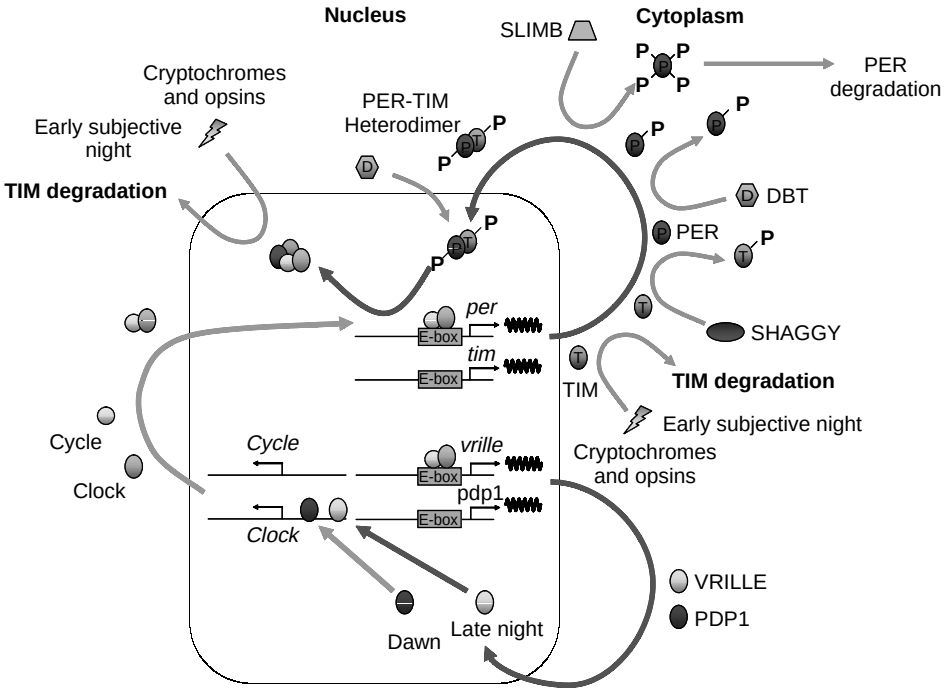

Fig. 4 The *Drosophila* circadian clock appears to be composed of two interacting molecular loops. The clock proteins, CLOCK and CYCLE, act positively on E-boxes within the promoters of the genes *period* and *timeless*. PER and TIM proteins begin to accumulate in the cytoplasm but with an approximate 6-h delay between the peaks in transcript levels and those of the corresponding proteins. PER monomers are phosphorylated by the casein kinase, DOUBLE-TIME, which targets these proteins for degradation. With the increase in TIM protein and the formation of PER-TIM heterodimers, PER is stabilized, and the amount of both proteins begins to accumulate. TIM is phosphorylated by SHAGGY (GSK3), which may play a role in regulating the nuclear entry of PER and TIM. Upon nuclear entry, PER (as a protein complex) binds to CLOCK-CYCLE, repressing their positive transcriptional activity. In this way, PER and TIM feed back negatively to inhibit their own expression. Expression of the *clock* gene also oscillates in *Drosophila*. This oscillation is thought to be due to the positive action of the PAR domain protein PDP1 and repression by the protein VRILLE. The phasing of the rhythms of PDP1 and VRILLE differs, allowing for the generation of a *clock* transcript oscillation. In turn, *vrille* and *pdp1* promoters contain E-boxes, and so these genes are rhythmically expressed as a consequence of the action of CLOCK and CYCLE. The interaction of these proteins and genes represents a second but linked circadian molecular oscillation in *Drosophila*. Entrainment of the molecular clock occurs through both opsin and cryptochrome-dependent pathways. Light exposure results in the degradation of TIM (see text for details).

accumulate in the cytoplasm of the clock-containing cells.

TIM protein itself undergoes rhythmic phosphorylation. The kinase thought to be responsible for this posttranslational modification is *Drosophila* SHAGGY (SGG), an ortholog of the mammalian glycogen synthase kinase-3 (GSK3). GSK3 has been shown to have an essential role in mammalian development, but overexpression of *Drosophila shaggy* (*sgg*) in particular tissues shortens the circadian

free-running period. Mutations in *sgg*, reducing its levels, cause a lengthening of period and reduced TIM phosphorylation. This change in period is thought to be brought about by an alteration in the next critical event in the circadian cycle, the nuclear entry of the PER:TIM heterodimer. The overexpression of SGG clearly advances the timing of nuclear entry of this complex and may explain the "faster" running clock. A similar role may also be played by the phosphorylation of PER by casein kinase, as the dbt^S mutation delays the apparent nuclear accumulation of PER protein. In the mid-subjective night, the PER:TIM heterodimers enter the nucleus, possibly in combination with DBT and perhaps other proteins, as a large protein complex.

The nuclear entry of this complex is a critical step in the negative feedback part of the circadian loop, as it now brings the positive transcriptional regulators, CLK and CYC, into close physical proximity with the negative regulators, PER and TIM. The negative repressive action of PER:TIM is thought to occur as a consequence of direct physical interaction between the two protein complexes, which reduces the ability of CLK:CYC to bind to the E-boxes within the *per:tim* promoters. Of course, for the positive action of CLK:CYC to be reactivated and a new cycle to begin, this nuclear repression must be removed. How this occurs is not yet totally clear. However, it is known that the removal of TIM from the negatively acting protein complex is a critical step and that this may involve the action of SGG in some manner. The TIM protein may inhibit the direct action of DBT in phosphorylating PER, and so a reduction in TIM would lead to an increase in PER phosphorylation. This increased phosphorylation may then allow for degradation of PER protein in the nucleus and a removal of the repression on CLK:CYC. With the repression removed, these positive transcriptional regulators can act again back on the DNA to increase *per* and *tim* transcription, beginning the cycle anew. As light reduces PER and TIM levels via cryptochrome (see below), this results in a derepression of *per* and *tim* transcription, providing a mechanism by which light may influence the phase of the molecular clock. These processes are summarized in Fig. 4.

The above description is, in some ways, a simplified description of what is thought to occur in constant darkness (DD), but circadian clocks are rarely exposed to DD in living animals. The rhythmic environment, and exposure to light, is obviously essential for setting the clock, as has been discussed above. However, circadian oscillations are not only set to LD cycles but appear more "robust" with the molecular components showing a higher amplitude of oscillation. Light probably provides a "tonic drive" to these molecular clocks as well as setting their phase. But the critical question remains: how is the *Drosophila* molecular clock entrained to a light–dark cycle? The main action of light on central clock components is the light-dependent degradation of TIM via the proteosomal degradation pathway. As TIM is required for the stabilization of PER, it seems likely that the reduction in TIM levels by light will increase PER phosphorylation (as discussed above) and its targeted degradation.

Also, as discussed under "entrainment" (Sect. 3.1.1), CRY appears to play a critical role in the light-dependent degradation of TIM. This can be clearly shown in cry^b mutant *Drosophila*, lacking functional cryptochrome, in which high levels of stable TIM protein can be detected under an LD cycle. The typical light-dependent

TIM degradation is abolished in such mutants. Thus, the current model for molecular entrainment and an explanation for the PRC (Fig. 2) suggests that in the late subjective night, a pulse of light will lead to the degradation of TIM within the nucleus at an advanced circadian phase. This TIM degradation, along with that of destabilized PER, should remove the negative regulation on CLK and CYC. These positive transcriptional activators are then free to act again on the promoters of *tim* and *per* at an earlier phase and as such cause a phase advance of the circadian rhythm. The explanation for the generation of phase delays is a little less convincing, but the same light-dependent degradation of cytoplasmic TIM occurs in response to light pulses given in the early subjective night. However, at this time in the cycle, transcript levels for *tim* are high and so new TIM protein rapidly replaces that which is degraded, an event that is thought to generate a delay in the timing of the rhythm by extending the inhibition of CLK and CYC. What is clear, however, is that a complex balance of protein levels and protein interactions, between positive and negative acting factors, is essential to determine the resulting phase of the clock following exposure to light.

The interaction of CRY and TIM, as well as that of CRY and PER, in response to light pulses, is itself a complex event. It has been proposed that a currently unknown nuclear factor exists that binds to CRY in the dark, with the effect of preventing an interaction between CRY and the two central clock proteins. Upon light exposure, this nuclear factor is displaced from CRY, allowing for the critical interaction with TIM, which eventually leads to TIM degradation. If there was only one "unknown" nuclear factor, we might all sleep easy, but in reality, many more components of the light-entrainment pathway undoubtedly remain to be discovered in *Drosophila*.

The *Drosophila* clock seems also to have a second transcriptional loop that runs in parallel with the PER/TIM loop described above. In *Drosophila, Clk* transcript levels show a circadian oscillation, unlike the situation for *Clk* in the mouse. This raises the obvious question, how might *Clk* transcription itself be regulated? The oscillation in *Clk* transcript peaks in the early day, and is antiphasic to that of *per* and *tim*. A particular mutation in *Drosophila* appears to directly influence the levels of *Clk* expression. A loss of one copy of the gene *vrille* (*vri*) causes a shortening of circadian period, while overexpression tends to lengthen period. The VRILLE (VRI) protein itself oscillates, with a peak in expression corresponding to the trough in *Clk* transcript levels, and overexpression of *vri* causes suppression of *Clk* transcription through direct action on the *Clk* promoter (Fig. 4). VRI is the active repressor of *Clk* transcription. However, a related protein PAR domain protein 1 (PDP1) appears to be the positively acting transcriptional activator of *dclock* transcription. Heterozygous *pdp1* null mutant flies show a lengthening of circadian period, and amounts of *Clk* transcript fall to very low levels in homozygous null *pdp1* mutant flies. Crucially, the timing of the rhythms of the repressor VRI and the activator PDP1 at both the RNA and protein levels are temporally separated, allowing for differential temporal effects on *Clk* expression. It is interesting that both genes appear to be positively regulated by CLK:CYC themselves. So, a second circadian transcription–translation negative feedback loop appears to exist in *Drosophila* in parallel but linked to the PER/TIM loop described above. In this second loop,

CLK:CYC increases the transcription of *vri* and *pdp1*. VRI then feeds back negatively to shut down expression of *Clk*, while PDP1 acts on the *Clk* promoter to stimulate its transcription. The functional reasons for a second loop are not immediately clear. However, a consequence of such a system may give the circadian oscillation an added robustness and accuracy over the length of many circadian cycles. Of course, it also provides additional areas for clock modulation and entrainment possibilities through a range of environmental stimuli.

A molecular basis for temperature compensation in the clock mechanism has been suggested by a polymorphism identified in the *Drosophila per* gene. The two major variants demonstrate a different geographical distribution related to the animal's ability to maintain a circadian period at different temperatures. Furthermore, the identification of a thermosensitive splicing event that occurs in the 3′ untranslated region of the *per* gene, suggests a temperature-sensitive switch in the regulation of *per* expression. These studies provide fascinating clues to the molecular mechanisms underlying the temperature compensation of circadian rhythms.

3.1.3 Output

The two assays used in most *Drosophila* mutant screens to identify novel clock genes have been locomotor activity and eclosion from the pupal case. Both of these behaviors are under strong circadian clock control. In the case of locomotor behavior, it is clear that the clocks residing within the lateral neurons of the fly brain are critical for the timing of these events. Outputs from these lateral neurons are clearly essential for providing the correct temporal drive to locomotor centers within the fly brain. One gene involved in the timing of eclosion is the clock-controlled gene *lark*. Mutations in this gene cause very early emergence from the pupal case, advanced eclosion. However, no effect is detected during the period of circadian locomotor behavior. LARK protein oscillates, even in constant darkness, but not in *per* null mutant flies, demonstrating the importance of the clock in its regulation. Expression is restricted to a number of neuropeptide-containing neurons in the ventral nervous system, which are believed to play a critical role in the regulation of pupal eclosion.

One clock-controlled gene of particular interest is *takeout* (*to*). This gene is activated in flies under starvation conditions but not in flies lacking a functional clock. In some way, the circadian clock is necessary for this stress response to lack of food, and these mutants not only die more rapidly than wild-type flies but also show altered locomotor rhythms. There is clearly an interplay between the clock mechanism and the metabolic condition of the fly. This, of course, makes a great deal of sense. If the fly is in a state of metabolic stress (starvation), then obviously the individual will need to alter its locomotor pattern in order to search for food. The relationship between circadian clock and metabolic state is undoubtedly of great importance, an idea that is supported by studies on food restriction and entrainment of peripheral clocks in the mouse (see Sect. 3.2.3).

As we shall see in almost all of the other clock model systems discussed below, the use of modern genomic approaches, microarray, and DNA "chips" has been applied to the study of clock-regulated genes in *Drosophila*. Several groups have performed such analysis, and a wide range of rhythmic genes have been identified. Microarray studies in *Drosophila* have demonstrated very little agreement, highlighting technical problems with transcriptome

analysis. The inconsistency between these studies may be due to factors such as the specificity of the tissue under analysis, lack of biological replication, as well as the different levels of stringency applied during data analysis. An interesting finding of microarray studies on clock gene mutants (*dClock*, *dper*, and *dtim*) is that the expression level of many nonrhythmic genes is also affected. This suggests that these genes may also possess nonclock functions, rather than being dedicated solely to circadian rhythm generation. However, a wide range of cellular and physiological processes within the fly have been implicated as being under clock control, including aspects of metabolism, olfaction (sensory function), learning and memory, immune function, and so on. One of the great pleasures of circadian clock research is being able to torment one's colleagues in other fields by asking them to repeat all of their experiments in the middle of the night!

Drosophila not only possess circadian clocks in the lateral neurons of the brain but apparently light-responsive clocks also exist throughout the tissues of their body, a fact that is also true in zebrafish (see Sect. 3.3 and 5.0). What these peripheral clocks actually regulate in the fly is not absolutely clear. However, one excellent example of the independent function of a peripheral pacemaker is in the control of the olfactory system in the antennae of the fly. An electrophysiological analysis of the antennal response to food odors showed a circadian rhythm in sensitivity. This response continued in cultured antennae, demonstrating its independence from central clock control. Of some interest is that this oscillation is lost in cry^b mutant flies. As described in the beginning of Sect. 3.1.1, cryptochrome is thought to play a role in light detection and entrainment of the central circadian clock (amongst numerous other functions). However, in the periphery, the role of cryptochrome appears to be much more as a clock component rather than a photopigment, as is the case in the mouse. This strongly suggests that there are interesting differences in the clock mechanism between central and peripheral pacemakers within the fly. The antennal response is only one example of a clock-controlled physiological event in *Drosophila*. Over the next few years, as our focus turns away from the mechanism of the central clock, it is undoubtedly true that many more interesting aspects of clock-controlled processes will be discovered.

3.2 Mouse

The identification of the master circadian pacemaker of the mammalian circadian system has followed the classical approach in the neurosciences, starting with anatomy, followed by lesions, then transplantation to restore function, and finally *in vitro* and *in vivo* analysis. One of the major goals of neuroscience is to identify the anatomical substrates of specific behavioral control systems. This has proven frustratingly difficult in many cases, but the achievement of locating the mammalian master pacemaker is one of the success stories.

Curtis Richter's seminal contribution to chronobiology was to help find the location of the mammalian clock. He conducted a heroic series of experiments to confirm first of all that it was the brain that housed the clock and second, to try and establish where exactly it was in the brain. Richter removed adrenals, gonads, pituitary, thyroid, pineal, and pancreas. It was only when he turned to the brain that he

was able to produce arrhythmia. He made over 200 studies, investigating what happened if he damaged the rat brain and measured the effect of the lesion on the rat's locomotion, feeding, and drinking behaviors. In 1967, Richter reported that lesions in the front part of the hypothalamus of the brain eliminated multiple behavior rhythms. But he could not get any closer. In the early 1970s, Friedrich Stephan and Irving Zucker linked the observations that the estrus cycle was dependent on a small part of the brain in the anterior hypothalamus and that there was circadian control in the timing of hormone release. By careful lesioning in the frontal part of the hypothalamus, they identified a small paired cluster of cells known as the *suprachiasmatic nuclei* (*SCN*) as critical elements in organizing rhythms of behavior such as drinking and locomotion. At the same time that Stephan and Zucker were performing lesions, Robert Moore was using a different approach to find the clock. Many studies had shown that there was an intimate relationship between the clock and light. Moore's approach was to follow the light path as it came in through the eye and see where it went. In 1972, he and his coworkers injected radioactive amino acids into the eyes of rats. They followed the tracer molecules as they traveled from the eye, through the optic nerves, and concentrated in the SCN as well as the known visual structures. The newly found pathway from the eye to the SCN was called the *retinohypothalamic tract* (*RHT*). When Moore lesioned the SCN, there was a loss of the circadian rhythm of the hormone corticosterone. Between them, Richter, Moore, Stephan, and Zucker had established that removing the whole of the SCN (but not just a part of it) destroyed circadian rhythms in both behavioral and endocrine outputs. Just 20 000 or so cells seemed to be responsible for controlling the timing of a mammal's endogenous rhythms. Support came from Bill Schwartz who decided to study the metabolism of the SCN. Schwartz injected 2-deoxyglucose (2DG) and found that the SCN is metabolically active during the light phase of a 12 h:12 h light–dark cycle and relatively inactive during the dark phase. Significantly, no other brain region showed such a dramatic rhythm.

By the middle of the 1980s, several laboratories had shown that circadian activity rhythms in SCN-lesioned adult rats and hamsters could be restored by transplanting fetal SCN. The final proof that the SCN contained the "master clock" came from the chance discovery of a mutant golden hamster in Michael Menaker's laboratory. Martin Ralph found that one of the hamsters in a routine shipment of animals had a circadian activity period of 22 h, which was far below anything seen before. This hamster was named the *tau* mutation, and subsequent breeding experiments showed that this was a single-gene defect and two alleles of the *tau* gene resulted in a 20-h circadian period. Ralph, Menaker, and colleagues then used these animals as a tool to see whether the SCN was indeed the key oscillator. They asked the critical question: If the SCN was taken from a 20-h mutant and transplanted into a wild-type SCN lesioned host, would the recipient show a circadian period of the donor (20 h) or of the recipient (24 h)? The results were unambiguous. In every case, the restored rhythm showed a period close to 20 h. Likewise, when the transplantation was done the other way round, from normal to mutant, again it was the period of the donor that was restored. This specificity of the period was the unambiguous proof

needed to satisfy even the severest skeptic. The hypothesis that the SCN contained a circadian oscillator was correct.

Soon after the *tau* mutant transplantation studies, other researchers found that even individual cells in the SCN show a circadian rhythm. A SCN was placed into culture and the individual cells were then dissociated from each other while supported on a multielectrode array. SCN neurons fire at a frequency between 0.3 and 9.9 Hz. However, the overall firing rate of the rat SCN neurons showed a marked circadian rhythmicity averaging 24.35 h, with peak rates of firing occurring during the subjective day. So, not only was the SCN the anatomical site of the oscillator but it was composed of individual cells that themselves oscillated and somehow co-coordinated their individual firing to give an overall rhythm. The hunt for the master circadian pacemaker of mammals dominated chronobiology for approximately 80 years, and establishing a central role for this structure represents one of the major achievements in the field. But as we shall discuss below, the view of a master clock within the mammalian SCN simply driving peripheral rhythmicity in physiology and behavior is an oversimplification. Other tissues possess the molecular components to generate a circadian oscillation independently of the SCN, and in many respects show patterns of gene expression that resemble the molecular clock of individual SCN cells. The uniqueness of the SCN, however, is that this tissue alone seems to be able to restore rhythmicity when transplanted into an SCN-lesioned host.

3.2.1 Entrainment

It was assumed that in humans and other mammals the same light-sensitive cells in the eye that are used for vision, the rods and cones, are also used to adjust the clock. But in the early 1990s, unexpected experimental results began to challenge this assumption. It had been known since the 1980s that the circadian system of mammals was relatively insensitive to light, needing both relatively bright and long duration exposure to light to bring about entrainment. For example, the visual system of a rodent is some 200 times more sensitive to light than the circadian system, and rodents will not entrain to a light stimulus shorter than 30 s. The human circadian system is more than a thousand times less responsive to light than our visual system, and again needs a long duration signal to entrain. It also became clear that there was a marked anatomical separation between the projections from the eye to the visual and circadian structures within the brain. The master circadian clock in mammals, the SCN (discussed below), receives its retinal projections from the RHT, which is formed from a small number of distinct retinal ganglion cells (RGCs). These RGCs (around 1% of the total) tend to be distributed evenly over the entire retina and send an unmapped or random projection to the SCN. By contrast, the ganglion cells of the visual system send a highly mapped projection to the visual centers of the brain, such that a point on the retina maps precisely to a group of cells in the visual cortex. The visual system is thus able to deduce both how much light and where it occurs in specific regions of the environment, whereas the SCN receives only information about the general brightness of environmental light. Thus, the mammalian eye has parallel outputs, providing both image and brightness information.

Eye loss (like constant darkness) results in free-running rhythms in all mammals studied, including humans. However, claims have been made from time to time that mammals have nonocular photoreceptors. There has even been a recent suggestion that humans have photoreceptors behind the knee, as bright light shone in this region apparently shifted human circadian rhythms. It was an amusing thought but unfortunately nobody could replicate the findings. All the experimental evidence shows that the circadian system of mammals is entrained exclusively via photoreceptors within the eye. Although ocular, the remarkable finding in recent years has been that neither the rods nor cones are needed for this task and there exists another photoreceptor within the eye. The identification of this nonrod, noncone ocular photoreceptor has resulted from a series of experiments over a 10-year period. The first experiments studied mutant mice that lacked all rod and most of their cone photoreceptors. This mutant strain of mouse, known as the *retinal degeneration* or *rd/rd* mouse, is visually blind. But despite the massive (but not complete) loss of their rods and cones, these animals showed normal circadian responses to light. Blocking light reaching the eye abolished entrainment, confirming that the photoreceptors were indeed located within the eye. At the very least, these studies in mice and later studies in other rodents and humans showed that the processing of light by the eye for vision was very different from the way the eye processes light information for the clock. However, suggestive as these early studies were, they did not conclusively demonstrate the existence of a new ocular photoreceptor. Although the rods and cones were massively reduced in these mammals, they were never completely eliminated. There was the possibility that even a small number of photoreceptors may still be sufficient to maintain normal circadian responses to light. As a result, a new model, the *rd/rd cl* mouse, was developed that lacked all its rods and cones. Remarkably, these mice still showed normal circadian responses to light. Again, by blocking light reaching the eye, the effects of light on the circadian system were abolished, so there had to be a novel ocular photoreceptor. Because these mice lacked an outer retina, the new photoreceptor cells were assumed to be located within the inner retina, and the search was on for these novel sensory cells using a range of different approaches.

The spectral responses or action spectra for adjusting the circadian clock of *rd/rd cl* mice demonstrated the involvement of an opsin/vitamin A–based photopigment with a λ_{max} of ~480 nm (OP^{480}). The known mouse photopigments peak at ~360 nm (UV cone), ~498 nm (rod), and ~508 nm (green cone) and do not show any significant fit to the 480-nm action spectrum in *rd/rd cl* mice. These results argued that a novel photopigment contributes to the regulation of circadian response to light in the mouse. Parallel studies in rodents showed that at least some of the retinal ganglion cells that project to the SCN are intrinsically light sensitive. This was demonstrated by recording from single RGCs from the normal rat retina either surgically or pharmacologically isolated from the rods and cone or by monitoring light-induced changes in calcium concentration in the RGCs of the whole retina from *rd/rd cl* mice. Significantly, these light-sensitive RGCs express the photopigment melanopsin. Evidence that melanopsin plays an important role in the transduction of light information in the intrinsically photosensitive RGCs comes from several sets of

experiments. The first set of experiments showed that melanopsin knockout mice have attenuated phase-shifting response to light and that the photosensitive RGCs fail to respond to light in these melanopsin knockout mice. Most recently, mice in which rods, cones, and melanopsin photoreceptors have all been ablated failed to show any responses to light, arguing that these three classes of photoreceptors account for all light detection within the eye. These data support the conclusion that unlike the cryptochromes of *Drosophila*, mammalian cryptochromes play no role in circadian photoreception (see Sect. 3.2.2). Indeed, Russell Van Gelder, at Washington University School of Medicine, who has been the main proponent for a photosensory role for the cryptochromes stated recently: "It would thus appear that the primary photopigment in non-visual photoreception is melanopsin-dependent, but not cryptochrome-dependent."

The light signal from these different photoreceptors is transmitted to the SCN by the RHT. Two neurotransmitters, glutamate and pituitary adenylate cyclase-activating polypeptide (PACAP) that are stored in the projections of the RHT signal the information. If glutamate is applied to the SCN, then it will mimic the effects of light, while substances that block the action of glutamate will block light-induced phase shifts of the clock. Light or glutamate will delay the clock in the early night by increasing the levels of calcium (Ca^{2+}) within SCN neurons. But during the late night, light or glutamate seems to act via a different signaling cascade to induce phase advances. PACAP seems to modulate the action of glutamate. The differential response of these neurotransmitters acting at the level of individual SCN neurons to light provides the molecular basis for the delay and advance portions of the PRC (Fig. 2). The light input pathway to the clock interacts with the circadian oscillator so as to synchronize it to the solar cycle. Understanding how this happens at the molecular level takes us to the heart of how the circadian clock is generated.

Establishing that there are photoreceptors other than the classic rods and cones also has practical applications. The studies in retinally degenerate mice encouraged studies of circadian function in blind people. Josephine Arendt and her group at the University of Surrey and Charles Czeisler's group at Harvard both identified blind individuals with eyes but who lacked conscious light perception. Despite this, some of these individuals were still able to regulate their circadian responses to light. One practical result is that every attempt is now made to preserve an intact eye in people suffering from certain forms of eye disease so that it can perform its circadian function.

3.2.2 Oscillator

Before we consider the complex nature of the molecular clock in mammals (Fig. 5), the key players will be formally introduced and have been summarized in Table 1. Forward and reverse genetics have been the approaches used to identify the components of the mammalian molecular clock. Forward genetics represents the classical approach for identifying gene function, as exemplified in *Drosophila*, which typically starts with an abnormal phenotype and attempts to isolate the gene responsible. By contrast, reverse genetics starts with a gene of interest (often due to homology) and investigates the effects of mutations or deletions. Forward genetic approaches in circadian biology were originally thought to be impossible in mammals due to the labor and animal-intensive nature of such screening programs. Put bluntly, the cost

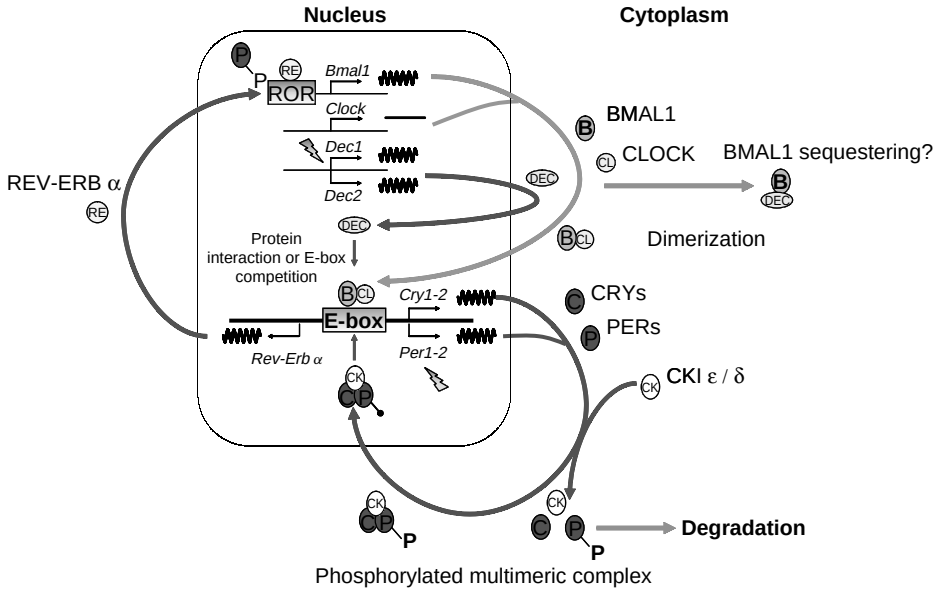

Fig. 5 The mammalian molecular clock is based upon a transcriptional drive produced by CLOCK:BMAL1 heterodimers. *Clock* is constitutively expressed, whereas *Bmal1* is rhythmically expressed. The CLOCK:BMAL1 heterodimers bind to E-box enhancers in the *period* and *cryptochrome* promoters, producing rhythmic *Per* and *Cry* expression. The resulting PER proteins are phosphorylated by CK1ε/δ and interact with CRY proteins to form a phosphorylated multimeric complex. This complex reenters the nucleus and produces a negative feedback by inhibiting CLOCK:BMAL1-mediated transcription. An additional loop is produced via *Rev-erbα*, which also possesses an E-box enhancer activated by CLOCK:BMAL1. *Rev-erbα* acts via an ROR element in the *Bmal1* promoter to inhibit *Bmal1* transcription, thus feeding back to remove the positive drive produced by CLOCK and BMAL1. As the PER/CRY/ CK1ε/δ complex reenters the nucleus and inhibits the CLOCK:BMAL1 drive on the E-box, *Rev-erbα* expression is also reduced. This leads to a disinhibition (activation) of *Bmal1*, thus restarting the molecular cycle. The newly identified components *Dec1* and *Dec2* may modulate the CLOCK:BMAL1 drive by competing for E-box binding or sequestering BMAL1.

and time required for such a mutagenesis screen seemed prohibitive. However, it was in fact a forward genetics approach that isolated the first clock gene (*Clock*) in mammals after the screening of only 25 animals! Despite this success, the majority of the mammalian *clock* genes have been identified using reverse genetic approaches. Mammalian homologs of the *Drosophila clock* genes (*mPer1–3* and *mCry1–2*) have highlighted a remarkable conservation of clock components between insects and mammals (due to the similarities found between species, where necessary we shall use the prefix "d" to indicate *Drosophila*, "m" to indicate mouse, "h" to indicate human and "zf" to indicate zebrafish). In addition, the use of interaction screens with previously identified clock components has been essential in identifying further elements involved in the mammalian clock (*Bmal1*, *Rev-Erbα*, *Dec1*, and *Dec2*). Common features of many of the proteins involved in the mammalian oscillator are bHLH motifs and PAS domains. As discussed

Tab. 1 Table depicting the peak expression of both mRNA and protein of the major components of the mammal (mouse) molecular clock. Time of day and the relative light–dark cycle are indicated. The approximate peaks of mRNA and protein are shown in light and dark gray respectively. The method of identification has been listed as either homology to *Drosophila*; interaction screens with previously identified clock proteins; or reverse genetics employing mutational screening. The impact of specific gene mutations upon the circadian phenotype is shown along with the free-running period (τ)

	SCN expression profile	Isolation & Analysis	Mutation	
Time	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24	**Identification**	**Phenotype**	**τ**
LD cycle				
mPer1*		Homology	Short period, then arrhythmic	22.6
mPer2*		Homology	Short period, then arrhythmic	22.1
mPer3		Homology	Short period	23.27
mCry1		Homology	Short period	22.51
mCry2		Homology	Long period	24.63
Bmal1		Interaction	Arrhythmic	-
Clock	Not rhythmic	Reverse	Long period, then arrhythmic	27.26
CK1ε	Not rhythmic (?)	Reverse	Short period	20-22
CK1δ	Not rhythmic (?)	Homology	-	-
Rev-erbα		Interaction	Short period	23.38
Dec1*		Interaction	-	-
Dec2		Interaction	-	-
Tim	Not rhythmic – ortholog of *time-out*	Homology	Embryonic lethal	

*Asterisks indicate genes that are upregulated by light.

above (Sect. 3.1.2), bHLH motifs occur in a variety of proteins capable of binding DNA, while PAS domains are amino acid sequences involved in protein:protein dimerization.

The first clock gene to be isolated in mammals was named *Circadian Locomotor Output Cycles Kaput*, or less torturously *Clock*, by using chemical mutagenesis coupled with a screen for abnormal circadian phenotypes. The *Clock* mutation results in a lengthened free-running period, and in homozygous mutants, a loss of circadian rhythms after several days in constant darkness. By use of positional cloning and transgenic BAC rescue, *Clock* was characterized. The *Clock* gene is a large transcriptional unit of around 100 000 bp, with 24 exons, encoding an 855 amino acid bHLH PAS domain–containing protein. *Clock* mRNA is widely expressed throughout the brain, with highest levels found in the SCN and pyriform cortex. The mutant *Clock* allele possesses a truncation of its transcriptional activation domain, providing a molecular basis for its loss of function. In addition, overexpression of *Clock* can lead to a period shortening. The expression of *Clock* mRNA appears to be nonrhythmic within the SCN, although recent studies have suggested that *Clock* mRNA may be weakly rhythmic.

The search for a binding partner for CLOCK was facilitated by the identification of additional bHLH PAS proteins, including Brain muscle arnt-like 1 (BMAL1) and five additional members of the PAS superfamily (MOPs 1–5). BMAL1 (MOP3) was independently identified in screens for bHLH PAS proteins involved in dioxin-signaling pathways and central nervous system development, and its importance to circadian rhythmicity was not initially appreciated. A series of clever molecular screens demonstrated that BMAL1 is capable of forming transcriptionally active complexes with CLOCK, which specifically bind E-box enhancers. From studies in *Drosophila*, E-box enhancers were known to be important in driving expression of *dper* and were also found to be present in the mammalian homolog (see below). In addition, the previously identified mutant CLOCK protein was found to be capable of dimerizing with BMAL1 and binding DNA, but failed to activate *dper* transcription. The importance of BMAL1 in the clock mechanism was highlighted when targeted deletion of this gene was found to result in an immediate loss of behavioral rhythms in constant darkness and reduced activity under normal light–dark cycles. *Bmal1* mRNA is expressed in the SCN in antiphase to the *Period* and *Cryptochrome* genes (see below), with a peak expression around CT 18. BMAL1 protein also appears to be rhythmic, with a peak expression between CT 0 and 8.

The identification of the mammalian *Period* (*Per*) genes followed closely on the heels of *Clock*. The presence of a mammalian homolog of the *Drosophila per* gene was suggested by an early study, demonstrating staining of the SCN with a *Drosophila* PER antibody. However, the isolation of the mammalian *Per* gene was not reported until 1997, when two groups independently isolated human and mouse sequences. The *Per* (or *Rigui* as it was alternatively named) gene encodes a PAS domain–containing protein with 44% homology to *Drosophila per*. *mPer* is expressed in a circadian manner within the SCN under constant conditions, and peaks ~CT 4–6 under a light–dark cycle. In addition to *Per* (hereafter termed *Per1*), further homologs were soon identified and named *Per2* and *Per3*, respectively. An additional *Per* gene, *Per4* was also isolated

in humans, although this appeared to be a pseudogene caused by retrotransposition of an ancestral *Per3*, as it contains no introns and numerous insertions, deletions, and frame shifts. The PER proteins all possess PAS domains, but do not appear to be capable of DNA binding. All three *Per* genes are rhythmically expressed in the SCN, with *Per1* and *Per3* peaking around CT 4–6 and *Per2* slightly later at CT 9–12 (see Table 1). *In vitro*, all appear capable of modestly inhibiting the CLOCK:BMAL1 transcriptional drive. However, *in vivo*, this negative feedback appears much reduced. The *Per1* and *Per2* genes play an essential role in the clock mechanism, as simultaneous deletions of both genes result in arrhythmicity. *Per1* and *Per2* also appear important in the light input to the molecular clock. In the SCN, both genes show increased expression following a nocturnal light pulse.

Targeted deletion of the *mPer1* gene results in a shortening of circadian period by around an hour, with reduced precision and stability of the cycle. However, in extended periods of constant darkness, *mPer1* knockouts become arrhythmic. The phase-shifting ability of *mPer1* mutants appears normal, and induction of *c-fos* in the SCN in response to light is no different to wild-type animals. Furthermore, deletion of *mPer1* does not appear to alter the molecular phenotype, as rhythms of *mPer2, mCry1,* and *Bmal1* are unaffected. Mice homozygous for a mutation of the *mPer2* gene arising from a deletion of 87 residues of the PAS domain ($mPer2^{Brdm1}$) exhibit a shorter free-running period, followed by a loss of rhythmicity in DD. The molecular phenotype of these animals is surprising, as *mPer1* and *mCry1* mRNA levels are still rhythmic in the SCN in DD, although the amplitude of expression is reduced. Furthermore, the peak expression of *Bmal1* is shifted, suggesting that *mPer2* may in fact play a role in a positive feedback mechanism. Levels of mPER2 protein are also reduced in *mCry*-deficient animals, even though *mPer2* mRNA is expressed at constant mid- to high-range levels, suggesting that mCRY proteins may be important for the stabilization of mPER2. The mechanism of the positive drive of mPER2 on *Bmal1* remains unclear, but may involve driving *Bmal1* transcription when *Rev-Erbα* is inhibited.

Mice with a targeted disruption of *mPer3* present a more subtle circadian phenotype than their *mPer1* and *mPer2* counterparts. The circadian period of these animals is shortened but maintained in DD, and the expression of other clock genes appears largely unaffected. Moreover, double mutants of *mPer1/mPer3* and *mPer2/mPer3* display the same phenotype as single *mPer1* and *mPer2* mutants, respectively. This has led to the suggestion that *mPer3* may in fact be involved in the output of the clock, rather than as part of the central mechanism.

The two mammalian *Cryptochrome* genes (*Cry1* and *Cry2*) were originally isolated through their homology with *Drosophila* (6–4) photolyase and plant cryptochromes.Owing to their homology with plant and *Drosophila* cryptochromes, it was not long before the *Crys* were suggested to be involved in mediating the light input to the circadian system. However, in mammals, the evidence for this was always weak, and recent studies have shown that the CRYs do not contribute to circadian light detection. Expression of *Cry1* mRNA in the SCN is rhythmic, whereas results on *Cry2* have suggested either no rhythmicity or weak rhythmicity within the SCN.

The peak of CRY1 protein expression follows some 2–3 h after the mRNA peak. Evidence for an involvement of *Crys* in the molecular clock emerged following targeted deletion studies, where *mCry1* knockouts were found to have a free-running period approximately 1 h shorter than wild-type animals. In contrast, *mCry2* knockouts exhibited a period of approximately 1 h longer. The most surprising result of these studies was that double *mCry* knockouts were totally arrhythmic in DD, suggesting that the mCRYs play an essential role in the molecular oscillator. In the double *mCry* knockouts, levels of *mPer1* mRNA in the SCN were found to be constitutively high and nonrhythmic, whereas *mPer2* levels were rhythmic under LD but showed constitutive mid-level expression in DD. These data suggested that the negative limb of the molecular clock in mCRY knockouts had been disrupted. This was confirmed by *in vitro* studies, which demonstrated that both CRY proteins were capable of inhibiting CLOCK:BMAL1 mediated transcription.

An additional component of the mammalian molecular clock was provided by the analysis of the *tau* mutation (see Sect. 3.2). The absence of genetic resources for the hamster proved a major setback in identifying the molecular basis of the *tau* mutation. However, by using a region of conserved synteny to the human and mouse genomes, the *tau* locus was shown to encode casein kinase 1ε (CK1ε), a serine/threonine protein kinase, and a mammalian ortholog of *Drosophila double-time* (*dbt*). The *tau* mutation results in a nonconserved substitution at residue 178, resulting in an enzyme that can bind PER1 and PER2, but has a decreased ability to phosphorylate these proteins. The isolation of CK1ε also led to the identification of CK1δ as a candidate clock regulator. This protein has 76% identity to CK1ε, and will also bind and phosphorylate PER1 *in vitro*. CK1ε lacks any rhythmic expression either at the mRNA or protein level.

Despite the isolation of a mammalian ortholog of *Drosophila timeless* early on in the dissection of the mammalian clock, a role for *mTim* and its protein product in the circadian mechanism remains uncertain. Expression of *mTim* is not rhythmic, nor is it light induced. Targeted disruption of *mTim* results in a normal circadian phenotype in heterozygotes, but homozygous deletion is embryonic lethal. A twist in the *Timeless* tale has come to light, however, as the *mTim* originally isolated appears not to be the true ortholog of *Drosophila tim* but instead is the ortholog of a more recently described gene, *time-out* (Sect. 3.1.3). A true mammalian ortholog of *dtim* may therefore exist, and a role within the circadian mechanism cannot be excluded.

Genome-wide complex trait analysis has suggested that in addition to those genes already identified, numerous uncharacterized loci contribute to the genetic complexity underlying the circadian phenotype. Vital pieces of the clock puzzle are undoubtedly missing and at the time of writing this review, the most promising candidates were *Rev-Erbα* and *Dec1* and *Dec2*. The nuclear orphan receptor *Rev-Erbα* is a transcription factor that binds to retinoic acid receptor-related orphan receptor (ROR) elements in the 5′ promoter of *Bmal1* to inhibit its expression. *Rev-Erbα* was isolated by analysis of the *Bmal1* promoter for known transcription factors and independently by screening transcription factor binding sites in genes shown to be rhythmic by microarray analysis. *Rev-Erbα* is expressed in a circadian manner, and knockouts display reduced

amplitude rhythms of both *Clock* and *Bmal1*. Despite these effects, constant conditions do not result in arrhythmicity in these animals. However, the free-running period is slightly shorter in knockouts, and light pulses produce dramatically increased phase shifts. DEC1 and DEC2 are additional bHLH proteins that appear to either directly interact with BMAL1 or compete for E-box binding, and thus modulate the positive drive of the clock. *Dec1* and *Dec2* are expressed in the SCN in a circadian manner, and the expression of *Dec1* is induced by light during the subjective night. By contrast, *Dec2* is unresponsive to light stimuli. The precise role played by these factors in the circadian clock *in vivo* remains to be clarified, and the effects of targeted deletion of the *Dec* genes is unknown. Another PAS domain protein that has attracted attention is the neuronal PAS-2 protein (NPAS2), also known as MOP4. However, NPAS2 is not expressed in either the adult or neonatal SCN and as such does not appear to be involved in the core clock mechanism. NPAS2 has recently been suggested to play a role in hormone-mediated regulation of peripheral vascular clocks, where it shows robust rhythmic expression. Furthermore, DNA binding by NPAS2 appears, *in vitro*, to be regulated by both redox state and cellular gas concentration (e.g. CO_2) via interactions with heme.

As can be seen, the mammalian molecular clock appears to be the emergent product of multiple interacting transcriptional–translational feedback loops, comprising both positive and negative elements. Studies in *Drosophila* have provided an essential framework for the understanding of the mammalian clock, allowing the assembly of genetic networks of ever-increasing complexity. The following is an attempt to summarize a current model of how all these interactions fit together to yield the molecular mechanism that enables the mammalian clock to tick (see Table 1 and Fig. 5).

The driving force of the mammalian oscillator appears to be the rhythmic transcriptional enhancement provided by CLOCK:BMAL1 heterodimers. *Bmal1* transcription yields rhythmic BMAL1 protein that heterodimerizes with constitutively expressed CLOCK. The CLOCK:BMAL1 complex binds to E-box enhancers, driving rhythmic transcription of the *Per 1–3* and *Cry 1–2* genes. The PER and CRY proteins form homo- and heterodimers and translocate back to the nucleus, where CRY forms the negative limb of the clock by interacting with CLOCK and/or BMAL1 to inhibit *Per* and *Cry* transcription. A positive drive also occurs, and indeed, may be an essential component for sustained rhythmicity, providing the instability required for sustained oscillations. PER2 has been suggested to provide a positive drive on *Bmal1* transcription, and *in vitro* assays have revealed that transcription of *Bmal1* is enhanced by both CRYs and PER2, but is repressed by CLOCK:BMAL1, suggesting the presence of both positive and negative drives on *Bmal1* transcription. These findings can be explained, at least in part, by the action of the recently identified nuclear orphan receptor *Rev-Erbα* (Fig. 5). As well as driving the expression of PERs and CRYs, CLOCK:BMAL1 also enhance expression of *Rev-Erbα* via an E-box, and, thus, REV-ERBα protein synthesis. REV-ERBα then feeds back to inhibit *Bmal1* transcription by acting through ROR response elements in the *Bmal1* promoter. The inhibition produced by REV-ERBα is removed by the CRY negative feedback loop on the CLOCK:BMAL1 heterodimer, which leads to a disinhibition (activation) of *Bmal1*. The central drive of the clock may

in turn be modulated by DEC1 and DEC2, as these proteins are capable of inhibiting CLOCK:BMAL1 mediated transcriptional activation *in vitro*, through either direct protein–protein interactions or competition for E-box binding. An additional level of regulation has also been suggested on the basis of independent evening (E) and morning (M) oscillators. The evidence for this model stems from the differing roles of *Per* and *Cry* genes within the clock and the differences in light responses and control of output pathways at different times of the circadian cycle.

While translational delays are expected to occur between mRNA and protein synthesis (especially with posttranscriptional modification of large polyintronic mRNAs), these delays by themselves appear insufficient to provide a cycle of 24 h. The discovery of a range of posttranslational mechanisms that affect nuclear entry, protein interactions, phosphorylation states, and degradation of the clock proteins will all build time delays into the molecular mechanism to produce the 24-h output. For example, as the mutation in Ck1ε in the *tau* mutant hamster results in a shortened circadian period, and diminished rate of phophorylation of PER *in vitro*, phosphorylation has been proposed as one mechanism for providing time delays. However, this view would not be supported by *in vivo* data from the *tau* mutant, which shows that the degree of PER phosphorylation is unaffected. Nevertheless, phosphorylation does appear to be critically important. The degree of phosphorylation of PER has been shown to affect its cytoplasmic retention, possibly by masking nuclear localization signals. Furthermore, the half-life of human PER1 is reduced by phosphorylation, and phosphorylation of PER1 and PER3 by CK1 results in rapid degradation via the ubiquitin-proteasome pathway.

In the context of protein–protein interactions and the generation of the circadian period, heterodimerization of CRY and PER appears essential for nuclear translocation. *Cry* double knockouts have lower levels of the PER1, PER2, and CK1ε in the nucleus (Fig. 5). The formation of PER:CRY heterodimers also appears to be important for preventing the breakdown of PER2 by the ubiquitin-proteasome pathway. As well as PER:CRY heterodimers, PER proteins will dimerize with each other as well as with CK1ε. PER proteins may act as a scaffold to enable the simultaneous binding of CRYs and CK1ε, and the phosphorylation of CRY is dependent upon both proteins being bound to PER. The absolute levels of CRY and CK1ε protein are in excess of PER, suggesting that PER abundance is normally rate limiting for PER:CRY heterodimerization, which is in turn necessary for nuclear entry of the protein complex (Fig. 5). Further evidence for the role of PER in nuclear localization comes from the subcellular localization of CRY proteins in PER1/PER2 double mutants. In these animals, CRY was concentrated in the cytoplasm, but only weakly detectable in the nucleus. Furthermore, the synchronous nuclear localization of PERs, CRYs, and CK1ε appears to be dependent upon PER rhythmicity. The circadian variation in protein interactions results in phosphorylated multimeric complexes of PER, CRY, and CK1ε in the nucleus, which are bound to DNA during the negative feedback phase of the molecular cycle. Despite the rhythmic expression of BMAL1 protein, CLOCK:BMAL1 heterodimers are bound to DNA throughout the circadian cycle. Thus, it is the rhythmic presence of the phosphorylated multimeric complex

that appears to inhibit the CLOCK:BMAL1 transcriptional drive.

Cellular metabolism may also be linked with the clock mechanism. *In vitro*, the DNA binding activity of BMAL1:CLOCK and BMAL1:NPAS2 heterodimers are regulated by the redox state of the respiratory chain electron carrier nicotinamide adenine nucleotide (NAD). The reduced form, NADPH, stimulates DNA binding, whereas the oxidized form, $NADP^+$, inhibits DNA binding. Consequently, the NADPH:$NADP^+$ ratio, which is affected by the metabolic state of the cell, may act to influence the central drive to the circadian clock. Evidence for such a role *in vivo* is, however, lacking and given the presence of CLOCK:BMAL1 heterodimers bound to DNA throughout the circadian cycle may not play a major role.

3.2.3 Output

While a quantum leap has occurred over the last decade in our understanding of the inputs and mechanisms of the mammalian circadian system, the links between these cellular oscillators and overt behavioral rhythms, while undeniably present, still remain poorly understood. In short, how does the SCN transduce rhythmicity to the rest of the animal? Although direct neural projections from the SCN to target brain regions have been demonstrated, factors secreted from the SCN play a critical role. A role for humoral signals in SCN output was suggested following the functional rescue of circadian rhythms in SCN-lesioned animals by transplantation of SCN tissue from mutant animals (see Sect. 3.2). These transplants were encapsulated in a semipermeable membrane to prevent neural outgrowth, therefore demonstrating that the factor(s) responsible for rhythmicity were humoral in origin. Two neuropeptides, transforming growth factor α (TGFα) and prokinectin-2 (PK2) have been recently isolated, and form strong candidates for providing this humoral signal.

A number of microarray studies have recently highlighted the extent of circadian regulation of the transcriptome in a variety of tissues, including liver, heart, SCN as well as fibroblasts, with estimates varying from 3 to 9% of genes demonstrating circadian modulation. What we know is that within the SCN, several first-order clock-controlled genes (ccg) have been identified that appear to play a role in transducing the output signal from the molecular oscillator to effector systems. Moreover, several of these genes appear to be involved in similar roles in peripheral clocks. Both mRNA and protein levels of the neuropeptide arginine vasopressin (AVP) display robust circadian rhythms. The rhythmic expression of AVP appears to be a product of BMAL1:CLOCK binding to an E-box enhancer within its promoter. Vasopressin acts locally to regulate the amplitude of SCN firing, as well as transmitting circadian signals to other hypothalamic targets. D-element binding protein (DBP) is a bZIP transcription factor, also driven by BMAL1:CLOCK E-box binding. DBP has been shown to regulate key enzymes in metabolic processes within the liver, and may play a similar role in regulating second-order clock-controlled genes in the SCN.

While early studies recognized the expression of clock genes outside of the SCN, the finding was typically overlooked because of the essential role of the SCN in providing rhythmicity. After all, SCN-lesioned animals are arrhythmic! However, it has become clear over the last few years that the broad expression patterns of the clock genes appear to

truly reflect oscillatory mechanisms that are independent of the SCN. The field of chronobiology was truly shocked when in 1998, Uli Schibler at the University of Geneva found that a cell line of fibroblasts that had been cultured for about 30 years could be induced to show a 24-h pattern of clock gene expression. He was able to induce 24-h cycles of expression by simply treating the cultured cells with serum. Somehow, one of the many ingredients in the serum (or a combination) was capable of initiating a 24-h oscillation of clock gene expression for a few cycles before the oscillation flattened and damped out. The rhythm could be induced once again with a further exposure to serum.

The possibility of peripheral oscillators in mammals made sense of a range of somewhat anomalous findings. Early studies had shown that when food was available only for a limited time each day, rats increased their locomotor activity 2 to 4 h before the onset of food availability and will even behave this way after their SCN had been lesioned. Recent work in Mike Menaker's laboratory has shown that circadian rhythms of clock genes within the liver can function independently of both the SCN and the light–dark cycle. Providing food for a 4-h period in the middle of the rat's subjective day had little effect on its general nocturnal behavior, but it showed increased locomotor and clock gene expression in the period before feeding. This suggests that food intake in itself generates an entraining signal for the liver and perhaps other digestive organs. The research team suggests that "the hierarchical method is still at least partially valid: the SCN does generate rhythmic neural and hormonal signals that influence rhythms in other brain areas, in peripheral endocrine organs such as the pineal and in behavior. Nevertheless, circadian oscillations in the liver (and perhaps in other peripheral organs) may respond more directly to the environment. This revised view of mammalian circadian organization emphasizes the possible role of behavior in maintaining internal temporal structure".

The view of the circadian system in mammals has changed from a central clock driving all outputs to a hierarchy of multiple oscillators, including tissues such as the liver, heart, spleen, kidney, lung, and testes. As such, the differences between central and peripheral oscillators have remained difficult to clarify, and synchronization of these oscillators may represent an inevitable consequence of multiple oscillatory phenomena, whereby any stimuli capable of transferring temporal information from one oscillator to another may function as a potential zeitgeber. The importance of the SCN may simply reflect its privileged position of receiving light input from the retina and its ability to modulate peripheral rhythms directly via hormonal and autonomic outputs as well as indirectly through driving locomotor activity (Fig. 2). Eighty years after Richter started to study endogenous activity in the rat, we return to one of his central themes. He showed that an animal maintains its internal stability through its general behavior as well as through specific and automatic physiological processes. The recent studies on food restriction and liver activity suggest that there is a possible role for behavior in maintaining internal temporal structure. Thus, models for circadian organization in mammals should include feedback loops from outputs, such as feeding, which may act as zeitgeber signals to assist in the overall regulation of the circadian system.

3.3 Zebrafish

At present, zebrafish represent the only alternative vertebrate genetic model to the mouse. To phrase this differently, there are only two model systems currently available in vertebrates, in which one can perform a forward genetic or mutant screening–based approach for determining the genes involved in a biological problem. The zebrafish system certainly is not nearly as advanced in genetic terms as the mouse or *Drosophila*, but progress is being rapidly made by research institutes such as the Sanger Center in terms of sequencing the genome.

Zebrafish made a relatively late arrival on the circadian scene. Consequently, relatively little is known in explicit detail about clock entrainment, central mechanism, or the output processes. However, the fish system has provided interesting novel information on developmental aspects of the circadian clock (developmental biology of course being one of its main applications) and surprising data regarding clock organization at the whole animal level.

The cloning of a zebrafish homolog of the *Clock* gene followed shortly after the identification of this gene in the mouse. Not surprisingly, the initial efforts examining the expression profile of this gene focused on the eye and the pineal gland, the two structures known in the lower vertebrates to contain a circadian clock. Of some interest was the finding that transcript levels for *zfClock* showed a robust circadian oscillation, something that is lacking for the mRNA of *mClock* in the SCN. Control studies were performed on other tissues, which we believed lacked a pacemaker, with the certainty that no clock expression or at least no rhythm would be detected. It generated considerable surprise, therefore, when a robust rhythm was detected in these body tissues, which included the samples taken from whole brain, heart, kidney, gill, fin, and so on. This raised two possibilities (1) that these peripheral oscillations are driven by a central master circadian clock or (2) that each tissue within the animal contains an independent, endogenous circadian pacemaker. To test this, a simple tissue culture procedure was established and organs such as heart and kidney were cultured for multiple days. The results from these culture experiments revealed a damped but continuing circadian oscillation in *clock* gene expression. Each organ in the fish must, therefore, contain its own circadian clock. Additional studies using primary cell lines generated from 24-h zebrafish embryos revealed similar oscillations. The current view is that most, if not all, cells within the zebrafish body contain a functional circadian clock. In this regard, zebrafish bears considerable similarity to data previously described in *Drosophila* and the emerging results from mammals. An obvious but central question is, how are these peripheral and cell line clocks set to local time? This issue is discussed below.

3.3.1 Entrainment

How are these tissue and cell clocks entrained? The simplest hypothesis was that these peripheral clocks would be set by some form of hormonal cue secreted by a "master" clock (possibly a pineal or retinal pacemaker), and as such this hormonal signal would maintain internal synchrony. One might postulate that such a signal could be melatonin, or possibly even a glucocorticoid hormone. However, the reality is a little more bizarre. A group of dissected zebrafish hearts was placed

into illuminated incubators on either a matching light cycle to the individuals from which they were removed (LD cycle) or on a reversed (DL cycle). By examining the levels of clock gene expression over the next 5 days in culture, it became clear that the circadian clocks within these hearts re-entrained to the reversed light cycle very rapidly, in fact within two cycles of exposure to the new lighting regime. This result has also been seen with other zebrafish tissues, as well as cell lines. The organs and tissue within the fish not only contain a clock but are also directly light responsive. All of these cells must, therefore, contain the relevant circadian photopigments and signal transduction pathways to enable this completely decentralized form of clock entrainment.

The obvious question of course is, what are these tissue photopigments and signaling events? The answer is not yet clear, but some candidate molecules have been proposed. The groups of Sancar and Sassone-Corsi performed an action spectrum on a zebrafish embryonic cell line in an attempt to determine optimal wavelengths for clock gene induction. Only one wavelength produced a significant response at 380 nm. On the basis of this single peak in the UV/blue wavelengths, it has been proposed that the relevant photopigment is not an opsin but in fact a cryptochrome. Cryptochromes have of course been implicated in *Drosophila* entrainment as discussed earlier (Sect. 3.1.1). In reality, the shape of this action spectrum neither fits that predicted for an opsin or a cryptochrome. The fact that these experiments were performed on cells in suspension (zebrafish cell lines grow as attached cells) and using RNase protections as a means of assaying RNA levels, which works badly at submaximal light intensities, raises some concerns regarding the validity of these data.

Cryptochromes are certainly expressed throughout most tissues in zebrafish, but as they are likely to play a central role in the core clock mechanism, this cannot be used as an argument for their role in light detection. There are six cryptochromes in the zebrafish, and the role of this class of molecule is likely to be quite complex. The isolation and identification of a widely expressed opsin, however, does generate a novel candidate for the role of a circadian photopigment. The Foster group identified a novel class of opsin by screening a *Fugu* (pufferfish) gridded genomic cosmid DNA library with a probe designed against Salmon VA opsin. The resulting novel opsin gene has been called *Teleost Multiple Tissue* opsin or TMT opsin, and is unique, in part, due to its unusual expression profile. Unlike most opsins, which are expressed primarily in the retina, pineal gland, or a few highly localized photoreceptive brain regions, TMT opsin is found throughout adult brain, in tissues such as heart and kidney, and in clock-containing embryonic cell lines, but not in cell lines lacking a clock. In other words, the expression pattern of this opsin parallels the localization of directly light-entrainable clocks. Though, of course, this correlation cannot be taken as proof of function (such studies are underway), TMT opsin is certainly the best candidate to date for the role of zebrafish circadian photopigment.

The signaling events that link the photopigment to the circadian clock are also not well elucidated. A possible role for the MAP kinase pathway has been proposed, but again no convincing functional tests of this pathway have yet been conducted. The potential roles of CREB and its phosphorylation, or the induction of *c-fos*

transcription, as seen in the SCN of mammals have not been examined. Nor has the role of intracellular calcium changes been explored to date; but this work is currently underway. One can reasonably state that, at this time, there are more question marks than candidates on the zebrafish input or light-signaling pathway.

3.3.2 Oscillator

Again, owing to the recent arrival of zebrafish on the circadian scene, not a great deal is known about the central clock mechanism. This fact, though, is certain to change in the near future, and a number of laboratories have isolated circadian clock mutant animals. When the genes mutated in these animals have been identified, it is likely that novel clock components will be discovered. The availability of numerous clock-containing, light-responsive primary cell lines in the zebrafish, which are not available in any other animal model, will also aid greatly in the functional testing and biochemical analysis of how these clock components fit together.

To date, all of the known clock genes in zebrafish have been cloned by homology from *Drosophila* and mouse. The first zebrafish clock gene to be cloned was *zf-Clock*, the zebrafish homolog of the first vertebrate clock gene, *mClock*. There was considerable similarity between the two genes, both possessing two distinct PAS domains, a DNA binding domain, and an extremely long glutamine-rich transactivation domain in the C-terminus of the protein. The major and most interesting difference between zebrafish and mouse *Clock* was the strong circadian rhythm seen in *zfClock* expression. Zebrafish *clock* shows a large peak in expression in the early night, whereas mouse *Clock* does not appear to oscillate, at least at the mRNA level.

The existence of two PAS domains suggested that *clock* had a protein partner that acted as a dimer within the cell. A yeast two-hybrid screen was, therefore, performed with zfCLOCK to identify novel partners. This screen yielded two interacting proteins, both very similar in sequence to each other and to the protein BMAL1. BMAL1 in mouse had also just been identified as a partner of the CLOCK protein in the mammalian system, and so perhaps it was no surprise that zebrafish were similar to mouse in this respect. The two *BMAL1* genes in zebrafish show a slightly different phasing, that is, timing of peak expression relative to the light–dark cycle, but the biological meaning of this is not yet clear. The CLOCK:BMAL1 dimer acts as a positive transcription regulator, activating the transcription of target genes via (generally speaking) E-box elements found in the promoter regions of these genes. The positive element of the circadian feedback loop, as represented by the action of CLOCK and BMAL, appears to be conserved between *Drosophila*, mouse, and zebrafish.

Over the last two years, the issue of genome duplication has arisen in the context of the zebrafish circadian system. This has been an issue in the area of zebrafish development, in particular, regarding *hox* gene duplication. It is clear that at some point in evolution zebrafish underwent a duplication of the genome in some areas, if not all, leading to a state of polyploidy. So, for example, three *Clock* genes and three *bmal1* genes have now been identified in zebrafish. There are also at least, but undoubtedly more than, three *per* genes. However, the greatest increase in gene number, as mentioned above, can be seen with the cryptochromes, of which there are currently six known distinct forms. Such genome multiplications have

been described in *Xenopus laevis*, and in this system, there is some evidence that the gene replication simply leads to the production of redundant copies. Do these multiple gene copies in zebrafish simply reflect redundancy in the clock mechanism? Though there is no absolute answer yet, the evidence suggests that these multiple genes are not simply redundant copies or pseudogenes as in the case of *hPer4* (see Sect. 3.2.2). In the case of the cryptochromes, different individual types show differences in the timing of their expression, as well as differences in the nature of their light induction. Similar differences exist for the *bmal1* genes, as described above, but this is also true for the *per* genes. Each of these gene replications appears to have had sufficient "evolutionary time" to have evolved distinct circadian characteristics. If these differences also reflect distinct functional roles, then the clock in fish may be appallingly complex, and one might be able to say that 15 clock genes have been identified to date in this species. Providing functional evidence for each of these genes will be quite a challenge. As a result, attempts to provide a model of the zebrafish molecular clock would be premature.

3.3.3 Output

With circadian clocks present in apparently most, if not all, tissues and cells, an obvious question is what are all of these tissue clocks regulating? At present, very few rhythmic clock output genes have been identified in the fish, although it is likely that there will be broad similarity between clock-controlled genes in mouse and zebrafish peripheral tissues. As discussed above, a wide range of cellular processes are regulated by the clock in tissues such as the mouse liver, including genes related to steroid biosynthesis, structural genes, glycolysis, and metabolism. One aspect of these mammalian studies, which take advantage of DNA microarray technologies, is that these organs are dissected and immediately processed for RNA at particular circadian times. In other words, the oscillations in gene expression seen in these tissues are occurring *in vivo* in the living whole animal. These data are, of course, of considerable interest, but from such studies it is not possible to determine whether these downstream oscillations are the result of peripheral clock control or central SCN clock control via neuronal or humoral regulation. The zebrafish should prove to be an excellent model system for the separation of local clock-controlled genes from those that might be regulated by a central master pacemaker. Such studies will be greatly facilitated by the relative ease with which it is possible to maintain zebrafish tissues in culture for many days. Oscillations in clock-controlled genes can then be compared between *in vitro* and *in vivo* samples, and some determination of the relative roles played by central versus peripheral clocks in regulating animal physiology can be ascertained.

One clear aspect of physiology under clock control, however, is the pineal and retinal release of melatonin. The clock regulation of these events appears to occur very early in development. The pineal gland is one of the first structures to differentiate in the developing zebrafish embryo, and is clearly present 26 h post-fertilization. The first release of melatonin can also be detected at this time. Whether this early developmental production of melatonin has any impact on development has not explicitly been examined. However, it is likely that this hormone may play a role in pigmentation changes (and subsequent camouflage) in early larvae,

which are prone to high levels of predation during the initial stages of life.

Rhythmic locomotor activity is the standard measure of clock output. Zebrafish larvae have been video monitored and do show circadian rhythms in twitching movements or rotation within the chorion (egg) and so a clock does seem to regulate at least this aspect of rhythmic behavior. Adult fish are a little less reliable, with most animals being day active. However, some 73% of individuals are rhythmic in constant darkness at 21 °C. This level of noise in the behavior has made adult locomotor activity a relatively poor assay of clock function and, therefore, an undesirable method for identifying clock mutants. The level of adult activity does, however, seem to be very dependent upon the level of nutrition, and varies between fish facilities. Fish in certain animal facilities appear more circadian in terms of day/night activity than individuals maintained under different conditions.

The examination of genes involved in embryo development is the major use of zebrafish as a model system. The advantages are of course that (1) large numbers of embryos are fertilized *ex vivo* and develop outside of the mother, (2) the embryos are largely transparent, (3) transgenic approaches have been well established, and (4) mutational analysis as well as sequencing of the genome are well underway. All of these advantages also make this an excellent system with which the development of the circadian clock can be studied. Yet, having said this, there is considerable confusion about exactly when the clock starts and whether developing embryos need to experience a rhythmic environment or not in order to establish a functional clock. One report claims that the zebrafish embryos directly inherit a sense of time from their mothers. Certain clock genes appear to be maternally inherited in that mRNA encoding some of these clock genes is present in the zebrafish egg and early embryo prior to the onset of zygotic transcription. Experiments looking at clock gene oscillation (in particular, *zfper3*) claim that rhythmic expression of this gene begins almost immediately in development and when embryos are raised in constant darkness. This claim that the oscillation starts without exposure to a rhythmic light–dark environment (or supposedly any other entraining signals) forces the conclusion that these embryos must inherit the "sense of time" or phasing of this rhythm from their mother. However, studies from other groups draw the opposite conclusion. In order for melatonin rhythms or larval locomotor rhythms to be expressed, early embryos appear to require exposure to a light–dark cycle for at least two or more days. Embryos raised in constant darkness fail to show any rhythmic clock outputs. If the circadian clock is maternally inherited, then it does not appear to provide any functional clock outputs. Clearly, more studies are required to address and clarify this issue.

In comparison to the other model systems discussed in this issue and because of its recent arrival on the circadian scene, relatively little is still known about the circadian system in the zebrafish. However, homologs of most of the known clock genes have been identified in this system, and many aspects of the molecular machinery seem similar. We can expect an increase in knowledge and greater contribution from the fish when some of the genes mutated in ongoing clock screens have been identified. To date, perhaps the most interesting contribution of the zebrafish has been to help re-evaluate the role of central "master" clocks versus peripheral pacemakers in

a vertebrate system. The fish system was the first to reveal the presence of independent peripheral oscillators in a vertebrate, and the generation of multiple clock-containing cell lines may prove of real value in future biochemical studies.

3.4 Neurospora

Along with *Drosophila*, the fungus *Neurospora crassa* provided some of the earliest insight into how a molecular clock might be generated. Like *Drosophila*, this species has several significant advantages; powerful genetics, well-established molecular tools, and a robust and easy-to-measure circadian rhythm in the form of asexual spore production (conidiation). In nature, *Neurospora* grow as a syncytium of undifferentiated hyphae in which many nuclei are distributed throughout the cytoplasm. In the laboratory, cultures are grown in specialized hollow glass tubes called *race tubes* (~40-cm long and 16 mm in diameter). These tubes are turned up at each end, and along the entire length of the tube, a thin layer of agar medium provides the substrate for fungal growth. *Neurospora* spores are inoculated at one end of the tube and the fungus grows within the thin layer of agar along the tube. Under a light–dark cycle, every 24 h, a tuft of hyphae grow up and out of the agar (aerial hyphae) from which an orange, carotenoid-rich, conidium full of asexual spores develop. This rhythm in conidiation, noted early on by Pittendrigh, shows all the characteristics of a circadian generated behavior. Under constant darkness, the rhythm persists and freeruns with a period of approximately 22 h. The period of the rhythm is temperature compensated over a broad range of temperatures ($Q_{10} \sim 1$ between 18 and 30 °C), and the rhythm can be entrained by both light and temperature cycles.

In the Introduction, we argued that the circadian system can be divided into a conceptual framework of input, oscillator, and output elements. Although we stressed that this depiction is flawed (Fig. 3), it has provided the field (at least initially) with a useful way to both represent and search for the elements of the circadian system in all organisms. But as we shall see, the key clock genes of *Neurospora* (*frq*, *wc-1*, and *wc-2*) span multiple functions. As a result, the division of the *Neurospora* circadian system into input > oscillator > output works far less successfully than for the other species that we have considered in this review. But for the sake of consistency, we have followed the same general structure in this section, and forewarn the reader that the information is not easily contained within these headings. As a result, we suggest looking at the summary of the molecular clock of *Neurospora* in Fig. 6 before proceeding.

3.4.1 Entrainment

Until very recently, the photoreceptors entraining the circadian system of *Neurospora* remained largely uncharacterized, although two genes were implicated fairly early on. Mutant screens in *Neurospora* strains defective in light detection and phototransduction for a range of light responses, including phototropism, carotenoid biosynthesis, and circadian entrainment, would consistently identify two distinct genes: *white collar-1* (*wc-1*) and *white collar-2* (*wc-2*). In the context of the present discussion, the absence of *wc-1* or *wc-2* would completely abolish the light induction of the *frequency* (*frq*) gene of *Neurospora*, and as discussed below, the *frq* mRNA and FRQ protein are central to the

molecular clock in this species. The finding that light will induce rapidly *frq* provided one of the key insights in unraveling the clock mechanism.

In constant darkness, there is a circadian oscillation of *frq*/FRQ that peaks in the late morning and mid-afternoon respectively. After light exposure and induction of *frq*/FRQ, the phase of the clock corresponds to the *frq*/FRQ peak in the late morning/mid-afternoon. In constant light, *frq*/FRQ are constitutively elevated and arrhythmicity results. After the transfer from light to dark, levels of *frq* and FRQ decline rapidly, and FRQ becomes extensively phosphorylated and is degraded. The cloning of *wc-1* and *wc-2* indicated that WC-1 and WC-2 encode transcriptional activators that can bind via their GATA type zinc-finger DNA binding domains. The *frq* gene has two cis-acting GATA elements, called *light response elements* (LREs), which if mutated will either attenuate or abolish light induction of *frq* and entrained patterns of conidiation. Furthermore, electrophoretic mobility shift assays (EMSAs) were used to show that WC-1 and WC-2 are the trans-acting factors that interact with the *frq* LREs. WC-1 and WC-2 heterodimerize using their PAS domains and form the white-collar complex (WCC). Upon light exposure, this WCC has been shown to bind to these LREs and initiate transcription of *frq*.

An action spectrum of binding of the WCC to the LREs has implicated the involvement of a flavin-based photopigment, with a wavelength of maximum photosensitivity at ~465 nm. These *in vitro* studies showed a marked similarity to previously published *in vivo* results on clock responses to light, suggesting that the *in vitro* EMSA analysis was a real reflection of the circadian photosensitivity in *Neuorospora*. But what is the photopigment? Could it be WC-1 or WC-2 or even both? WC-1 contains a LOV (light, oxygen, voltage) domain, which has been shown in characterized photopigments to bind a flavin adenine dinucleotide (FAD) chromophore. Significantly, the LOV domain in WC-1 also binds FAD, and disruption of the LOV domain in WC-1 disrupts FAD binding and circadian photosensitivity.

The model emerging for photoentrainment of the *Neurospora* clock suggests that in the dark WC-1 and WC-2 form a WCC and bind the LREs on *frq* and this causes *frq*/FRQ production (remember, in the dark, there is a circadian oscillation of *frq*/FRQ). After light exposure, WC-1 (using FAD as the chromophore) acts as the blue light photopigment and undergoes a conformation change such that the WCC is then able to strongly activate the LREs on the *frq* promoter and activate *frq* transcription (Fig. 6). This light- dependent response can help explain the broad characteristics of a phase response curve to light (Fig. 2). For example, light will rapidly activate *frq* transcription at any time of the day. Thus, at dawn, when *frq* transcription is normally low, a pulse of light will drive *frq* to levels that would normally occur at mid-morning and hence advance the clock/molecular cycle. Conversely, a light pulse at dusk (when *frq* levels are normally falling under the dim-light conditions at the end of the day) will upregulate *frq* to levels encountered previously during the bright light conditions of daylight and hence delay the clock/molecular cycle.

Although WC-1 is an FAD-based photopigment, it is not alone in modulating the effects of light on the clock. A rhythmically expressed protein called *VIVID* (VVD), although not needed for clock function, appears to act as a negative regulator

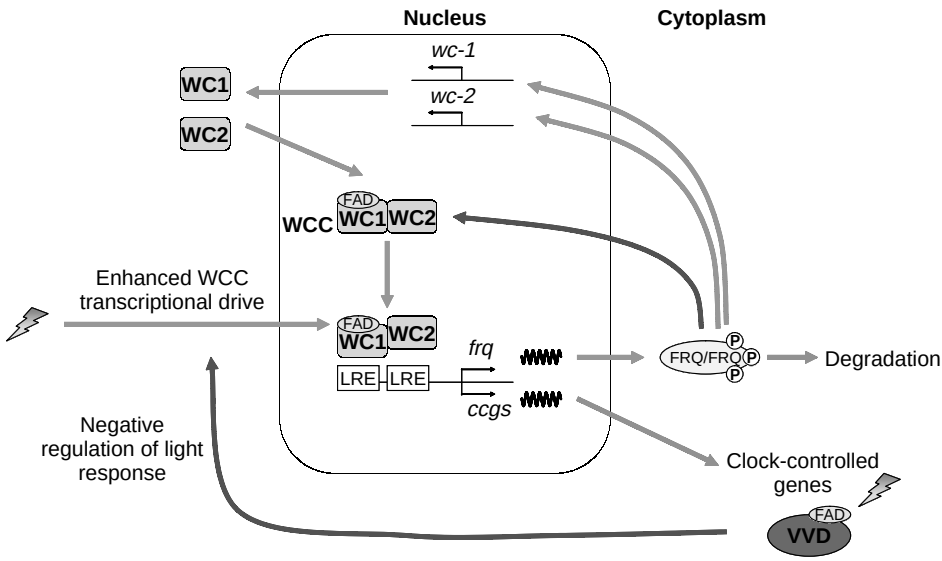

Fig. 6 The molecular clock of *Neurospora crassa* defies a simple compartmentalization into input > oscillator > output. The positive drive to the *Neurospora* clock is provided by *wc-1* and *wc-2* that form a white-collar complex (WCC) that enhances *frq* expression. The production of long and short forms of FRQ protein lead to the formation of homo- and heterodimers, which not only feed back to inhibit the WCC drive but also positively regulate the expression of *wc-1* and *wc-2*. FRQ stability determines period length. Light input is mediated through an FAD chromophore bound to WC-1 via an LOV domain. Light absorption results in a conformational change to the WCC, resulting in greatly enhanced *frq* expression. Clock-controlled genes also play a role in gating input signals to the clock. The protein VVD also contains an LOV domain and can bind the FAD chromophore in a similar manner to WC-1. VVD is rhythmically synthesized, and acts as a negative regulator of the light-induced WCC transcriptional drive.

of light responses. For example, *frq* transcription is more sensitive to light in *vvd* mutants. How VVD acts remains unclear but probably via an interaction with the WCC. VVD, like WC-1, has a LOV domain and binds FAD and is thought to be a blue light photopigment. VVD itself is light regulated, with levels low in the dark and high in the light. In addition, *vvd* expression requires the WCC. Thus, VVD not only acts as a negative regulator of the WCC in the light but also requires the presence of the WCC and light for its own expression (Fig. 6). The rhythmic expression of *vvd* presumably comes about as a result of the negative feedback of FRQ on the WCC (see below). As such, VVD provides an example of a clock-controlled gene that is also capable of regulating the input to the core oscillator. As VVD is influenced by an output from the clock, and VVD acts on the input pathway, it can be called a *zeitnehmer*.

3.4.2 Oscillator

Early in the field of molecular chronobiology, the *frq* gene was identified as a critical component of the clock. Mutations in *frq* would shorten, lengthen, or abolish free-running rhythms of conidiation. Significantly, these mutations appeared to cause no obvious morphological or

physiological defects, suggesting that *frq* acted specifically within the *Neurospora* clock. A chromosomal walk finally cloned the gene and the functional analysis of *frq*/FRQ began. It was shown that *frq* mRNA and FRQ protein levels cycle with a 22-h period in free-running wild-type strains and that the period of these oscillations matched the period of conidiation found in long and short mutant strains. A fusion gene, consisting of the *frq*-coding region attached to an inducible promoter, showed that the constitutive expression of *frq* would abolish circadian rhythms of conidiation. In addition, by driving levels of *frq* from high to low (removing the inducible agent) at any time of the day would reset the clock to dusk, which corresponds to the low point of *frq* mRNA in wild-type strains (*frq* levels reach their peak some 4–5 h after dusk). A negative feedback by FRQ on *frq* transcription was supported by the finding that defective FRQ protein would result in constant high levels of *frq*. Indeed, more recently, nuclear localization has shown to be critical for both behavioral and molecular oscillations in *Neurospora*. Thus, it was established fairly early on that *frq*/FRQ appears to be part of a molecular feedback loop that is essential for the generation of circadian rhythmicity. But, of course, *frq* is not the only gene involved in the fungal clock.

As discussed above, the WCC acts in the dark as the positive element of the *frq*/FRQ negative feedback loop and is part of the entrainment pathway under a light–dark cycle. Two forms of FRQ (long or short) are transcribed and associate as either homodimers or heterodimers to form a protein complex. When the FRQ complex reaches a certain level, it enters the nucleus and inhibits activation of *frq* transcription by interfering with WCC-driven transcription, and hence closes the negative feedback loop. Phosphorylation of FRQ occurs immediately after translation and targets FRQ for eventual degradation. Several kinases have been implicated in this process including casein kinase I, casein kinase II, and a calcium-/calmodulin-dependent kinase. The stability of FRQ in turn determines the period length of the oscillation. FRQ has also been shown to positively regulate expression of both WC-1 and WC-2, probably via different mechanisms (Fig. 6).

Thus, *Neurospora*, like other molecular clocks we have encountered, has both positive and negative transcriptional drives for the generation of the circadian oscillator. Although FRQ is an essential clock element, suggestions that *frq* was not the sole oscillatory component of the circadian system in *Neurospora* go back to 1986. Returning to this observation, and using *frq* null mutants, Merrow and colleagues employed traditional circadian entrainment protocols ("T-cycles") in an attempt to clarify the role of *frq* within the circadian system. They found strong evidence for a temperature, but not light, entrainable oscillator in *Neurospora*. The nature of this or these addition feedback loops remain uncharacterized, and their identification forms an ongoing focus for the *Neurospora* research community.

Neurospora has also provided insights into the underlying mechanisms of temperature compensation. Temperature effects in *Neurospora* appear to be mediated through the amount and type of FRQ protein made, with FRQ oscillating around higher levels at higher temperatures. As noted above, two forms of FRQ protein are produced, and temperature also appears to regulate the relative levels of these long and short forms by favoring differing initiation codons at different temperatures. When either of these initiation codons is

removed, the range of temperatures over which the clock will function is reduced.

3.4.3 Output

As discussed above, *vvd* is one example of a clock-controlled gene, but a large number of additional genes are emerging that are under clock regulation. These genes are associated with every aspect of *Neurospora's* biology, ranging across reproduction, development, behavior, and metabolism. One of the most intensely studied has been glyceraldehydes-3-phosphate dehydrogenase (GAPDH). GADPH acts as a catalyst in the first step of glycolysis. It is a critical enzyme in the glycolytic cycle and couples the oxidation of the glyceraldehydes-3-phosphate substrate to its phosphorylation. The oscillation in GADPH and other critical housekeeping genes is substantial, but a robust oscillation does not necessarily mean that the gene is a clock-controlled gene. The point that has often been made is that it is difficult to address whether rhythmic genes are regulated by the clock directly or merely as a by-product of rhythmic changes in the cellular environment. As with the other genetic models, microarray analysis is being used to assess the global pattern of rhythmically expressed genes in *Neurospora*, and like *Drosophila*, mice, and *Arabidopsis*, approximately 2–6% of the genome appears to be under some form of clock regulation.

3.5 Arabidopsis

The common theme in this manuscript has been to describe the major impact that forward genetics, and the identification of the particular mutated genes, has had on progress in understanding the biological clock in a wide range of species. It should be of no surprise then that in plants the primary model system is one renowned for its genetic accessibility, *Arabidopsis*. Our understanding of plant clocks is mainly due to progress in this species (Fig. 7).

3.5.1 Entrainment

To plants, of course, light is more than simply a signal of dawn and dusk, but critically the source of their "energy" through photosynthesis, and a regulator of their reproduction through flowering, on an annual or photoperiodic basis. As such, plants have evolved a large number of photopigments with which they take maximum advantage of the light source to which they are exposed. It is not surprising, therefore, that a large number of photopigments have been implicated in the regulation of the circadian clock.

The classic manner of examining the light characteristics of the circadian system is to provide a series of short, discreet light pulses of varying wavelengths and intensities and then determine the change in some clock parameter (circadian phase, acute gene induction, etc.), or in other words, to generate an action spectrum. This short term or acute action of light on clock phase can be seen as being equivalent to nonparametric entrainment, that is, the ability of short light pulses to mimic the action of long, sustained light effects (parametric entrainment). Phase shifting and the generation of action spectra to short light pulses have not yet been performed in *Arabidopsis*. However, extensive data has been produced regarding the effects of sustained, constant light treatments of specific wavelengths and intensities (fluence rate). This is made possible by the fact that the *Arabidopsis* circadian clock continues to run under constant light treatments and shows very reproducible changes in period length, depending upon the wavelength and intensity of light. In

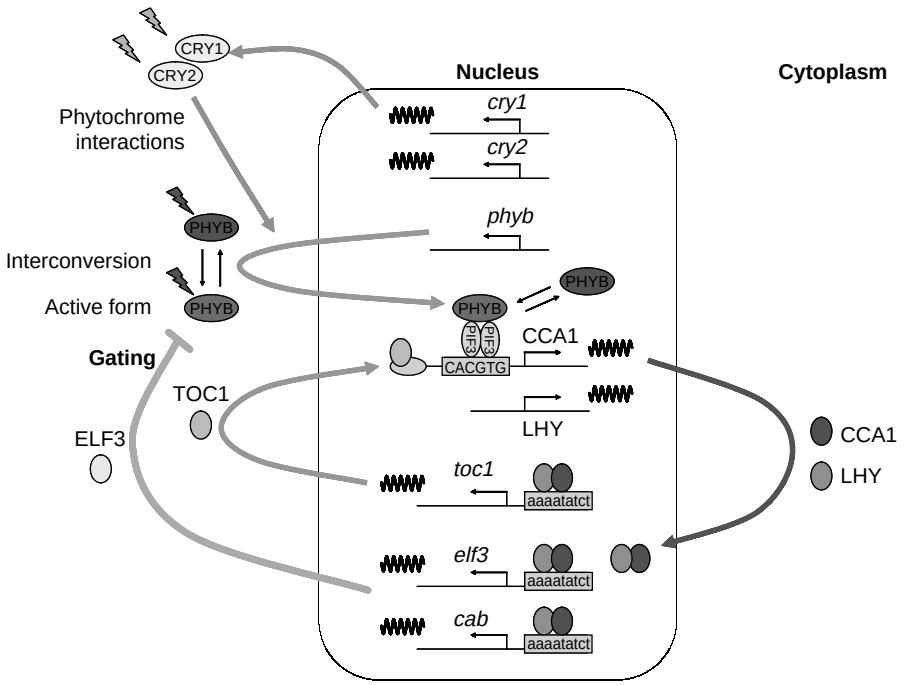

Fig. 7 Model of the molecular clock of *Arabidopsis thaliana*. Entrainment of the *Arabidopsis* circadian clock is thought to occur through a combination of the actions of cryptochromes and phytochromes. Cryptochromes function primarily in the blue region of the spectrum, with phytochromes acting in the red–far red region of the spectrum. Phytochrome undergoes nuclear entry following absorption of an activating photon of red light (converting it to the active far-red absorbing form). In the nucleus, phytochrome interacts with the bHLH transcription factor PIF3. This complex then increases the transcription of downstream genes containing the PIF3 DNA binding site, a G-box (CACGTG). The circadian-regulated output protein, ELF3 is thought to feed back on to the clock input pathway, and, as a consequence, gate the clock response to light to particular times of day. Two rhythmic gene products are essential components of the negative limb of the circadian oscillator. Transcripts for the genes *lhy* and *cca1* peak around the late subjective night/early dawn. Transcript levels for the gene *toc1* show oscillations 12 h antiphase to this, peaking at the end of the late subjective day. LHY and CCA1 proteins are capable of binding directly to a DNA motif in the promoter of the *toc1* gene. This DNA binding motif, termed an *evening element* (AAAATATCT), is not only found in the promoter of *toc1* but also in a number of clock output genes. TOC1 is thought to be the positive element of the circadian oscillator, acting to increase the transcriptional rates of *cca1* and *lhy*. These three proteins compose the central, core elements of the *Arabidopsis* circadian clock.

fact, the application of constant red and blue light caused a shortening of the circadian period, as measured in a transgenic *Arabidopsis*, which contained the luminescent reporter gene luciferase linked to the rhythmic promoter of the chlorophyll a/b binding protein gene (*cab::luc*). Further studies, taking advantage of specific mutant seedlings with null mutations in identified photoreceptive molecules, has allowed the roles of these particular photopigments to be determined even further. The results show a wide and complex arrangement of photosensory systems in the

plant, which allow for clock entrainment in response to many wavelengths of light (see below).

Arabidopsis contain five types of phytochrome, from PHYA to PHYE, which, generally speaking, absorb light in the red and far-red region of the spectrum. They are quite similar molecules of about 124 kDa and possess a covalently linked chromophore molecule consisting of a linear tetrapyrrole. Phytochromes exist within the cell in two states that can be interconverted from one to the other, depending upon the wavelengths of light they absorb. Upon initial synthesis, phytochrome is in an inactive red absorbing form (Pr), but upon absorption of a photon of light converts to the active far-red absorbing form (Pfr), in which it can regulate a wide range of physiological responses. This number of phytochromes provides the cell with a range of detection sensitivities. For example, PHYA is thought to function primarily in response to low-level red light, while the other phytochromes function under more intense red light conditions. *Arabidopsis* also contains two members of the cryptochrome family of proteins (CRY1 and CRY2). These photopigments absorb light in the blue/UV region of the spectrum using two chromophores, a light-harvesting pterin and a catalytic flavin.

Somers and colleagues crossed the transgenic *cab::luc* reporter plants to individuals with null mutation in two of the phytochromes, *phyA* and *phyB*, as well as to mutants for both *cry1* and *cry2*. The resulting mutant reporter seedlings were then tested under constant light conditions of varying wavelengths and intensities. The *phyA* mutant seedlings showed a loss of response to low intensities (low fluence rates) of both red and blue light. By contrast, the *phyB* mutants lack normal responses to high intensities of red light. The presence of these two pigments clearly provides the plant with a detection system that works optimally at differing intensities of light. A double mutant lacking both *phyA* and *phyB* showed that these two photopigment systems normally work additively. The *cry1* mutant seedlings lack a wild-type response to both high and low fluence levels of blue light, but respond normally to intermediate intensity levels. *cry2* mutants showed little alteration in the light response, but most likely act in a redundant manner with *cry1* at intermediate fluency levels. CRY1 is apparently the primary blue light–responsive photopigment at high light intensities, with CRY2 undergoing light-dependent degradation.

Interestingly, PHYA and CRY1 have been shown to interact physically at the protein level in yeast. At low light intensities, it is now believed that CRY1 acts downstream of PHYA, as a component of the signal-transduction cascade for PHYA light detection in the red wavelengths. This is not the only interaction that occurs between phytochromes and cryptochromes, as PHYB has been shown to interact with CRY2 by immunoprecipitation. This complex appears to form "nuclear speckles" in a light-dependent manner, but the exact functional meaning of this is not yet clear. This light-dependent nuclear localization is a critical event for light regulation of downstream gene expression (see below). In both tobacco and *Arabidopsis*, GFP protein fusions with PHYB show light-dependent nuclear entry. A nuclear localization signal exists in the carboxy terminus of this protein in *Arabidopsis*, and red light exposure promoted cytoplasmic to nuclear translocation. In tobacco, both PHYA and B undergo such light-driven movement, though PHYB translocation is inhibited by far-red light in contrast

to PHYA. The remaining phytochromes, PHYD and E, all appear to act in a redundant manner with PHYB to detect light of higher irradiances in the red region of the spectrum. Several triple mutants, including *phyA*, *phyB*, and *phyE*, showed a reduced response at high intensities of red light. However, all triple mutants examined so far still showed responses at low intensities of red light, possibly arguing for an additional, but as yet unidentified, role for *phyC*. Nevertheless, consensus implies that PHYB is certainly a predominant photopigment in red light detection for many aspects of light-regulated physiology in *Arabidopsis*.

Several additional regulators of light responsiveness in *Arabidopsis* have recently been identified, again through use of mutant seedlings. Mutations in the gene *sensitivity to red light reduced* (*srr1*) showed a reduced sensitivity to red light, as the name would suggest. A range of phenotypes in this mutant showed great similarity to defects seen in *phyB* mutations, suggesting a strong link between the roles played by both molecules. The acute induction of *cab* gene expression, which normally occurs in response to red light pulses, is in fact, increased in *srr1* mutants. The actual mRNA levels of *srr1* are themselves increased in response to red light, but not far-red light, pulses in an acute manner, with levels raised within 2 h of the pulse. However, transcript levels do not appear to oscillate across the circadian cycle when plants are monitored in constant light. This might suggest that the *srr1* is not involved in the central clock mechanism itself, but by measuring the period of the clock with *ccr2::luc* transgenic reporter lines in *srr1* mutants, it was apparent that there was a shortening of some 1.6 h in mutant versus wild-type plants. *srr1* clearly has a role not only within the light-detection machinery but also the clock mechanism itself. One interesting aspect of *srr1* is the fact that homologous genes appear to be present in a wide range of species, including mammals and flies. It will be interesting to see if this gene plays a role in clock function or light-entrainment pathways in these species as well as *Arabidopsis*.

As mentioned above, a critical aspect of phytochrome B red light signaling appears to be its light-dependent nuclear translocation. PHYB has been shown to directly bind to a bHLH factor known as *PIF3*, but only after conversion to its active far-red form (PfrB). PIF3 is a nuclear protein and binds to a DNA regulatory motif known as a *G-box*. It is of considerable interest that a G-box recognition sequence, CACGTG, is of course identical to the "classical" E-box sequence believed to be critical for circadian clock gene regulation in *Drosophila*, mammals, and zebrafish. The binding of photoactivated PHYB to PIF3 is thought to activate PIF3, and so turn on the transcription of genes containing this G/E-box motif. It is through this nuclear translocation of activated PHYB, binding to PIF3 and then activation of genes containing G-box recognition sites, that red light exposure can regulate downstream gene activation.

A defining feature of circadian clocks is their required ability to entrain to light–dark cycles and the additional fact that identical light pulses can either phase advance, phase delay, or have no effect on the phase of the clock, depending upon the time of day that they are given (Fig. 2). The nature or shape of this phase response curve (PRC) is critical for stable entrainment of the clock. There are also a number of examples where a given light pulse can acutely induce the transcription of a given gene at one time of day but not another.

As discussed in Sect. 3.4.1, the action of light on the clock can be gated by the clock itself, such that a clock output can feed back to regulate the light input pathway. Such an event has been termed a *zeitnehmer*. An example of a gene thought to be involved in such a process is *elf3* in *Arabidopsis*. The *elf3–1* (*early flowering3–1*) mutation was first discovered in a mutant screen for altered flowering phenotypes, but possessed several distinct circadian phenotypes also. Rhythms in the gene *cold circadian regulated 2* (*ccr2*), as measured by bioluminescence of a *ccr2::luc* transgenic plant, become arrhythmic in *elf3* mutant plants grown in constant light, but continue to oscillate in constant darkness, without a change in clock period. *elf3* cannot be a central component of the clock therefore, as the pacemaker continues to run normally in the mutant background. However, *elf3::luc* plants show that *elf3* itself oscillates in constant darkness, and so must be an output of the clock under rhythmic transcriptional control. Additional studies using *elf3* overexpression showed that the period length of the clock lengthened under constant white light conditions but remained unaltered in constant darkness. The ELF3 protein must, therefore, be feeding back to regulate the circadian clock response to light. This was confirmed by examining the shape and amplitude of the phase response curve to short, 1-h pulses of either red or blue light. The *ccr2::luc* transgenic plant made an examination of nonparametric entrainment, entrainment to short light pulses, possible for the first time in *Arabidopsis* due to its sustained rhythmicity under constant dark conditions. Overexpression, as well as the null mutation in *elf3*, caused distinct changes in the PRC amplitude. Overexpression of *elf3* caused a loss in the "sharpness" of the breakpoint at which light pulses change from causing delays to when they cause advances. Furthermore, overexpression tended to reduce the size of phase shifts, especially in red light, whereas the null mutation tended to increase the size of shifts or lead to arrhythmia. The conclusion to these data is that *elf3* is a clock output that is capable of feeding back to regulate the pacemakers response to light pulses or constant light conditions in a negative manner.

3.5.2 Oscillator

Mutations in the plant cryptochromes and phytochromes do not abolish circadian function, and so these light-detecting molecules themselves do not appear to be the central components of the clock. Clearly, this distinguishes the role of the cryptochromes in plants from the central clock role in, for example, mammals. Though there are still many gaps in our understanding of how the *Arabidopsis* clock functions, again much has been learned from the analysis of specific mutants.

A mutation in the gene *late elongated hypocotyl* (*lhy*) generated a number of aberrant circadian phenotypes, including a strong tendency for loss of rhythmicity in leaf movements, as well as the rhythmic circadian regulation of some gene expression, including that for both *cab* and *ccr2* genes. Furthermore, this mutation caused flowering to occur independently of the photoperiod upon which the plants were grown. The *lhy* gene encodes a MYB-DNA binding protein, and mRNA levels show a circadian oscillation with peak expression around dawn. This oscillation is lost in the mutant, while overexpression of the gene in a transgenic leads to a reduction in endogenous transcript levels. Further analysis of additional *lhy* mutant alleles showed a distinct period shortening in the oscillation of certain genes, such as

the gene *circadian clock associated* (*cca1*) by about 2–3 h (see below). These data together strongly suggest that LHY can feed back negatively to suppress its own transcription, a key and common characteristic of many central clock components, but that it may not function alone in this role (Fig. 7).

A similar MYB-like DNA binding protein, circadian clock associated (CCA1), shows a similar oscillatory profile, with both mRNA and protein oscillating. Overexpression of CCA1 also leads to a repression of endogenous transcript levels, again arguing for its role as a component of a negative feedback loop. Mutations in the *cca1* gene show a shortened period compared to wild types. However, these oscillations do continue to run with this altered period for several days, and so, as with LHY, CCA1 cannot be acting alone as a negative transcription regulator. Data from double mutants, lacking both LHY and CCA1 proteins, would be essential to determine if these two molecules were acting in a redundant manner. Existing data suggests this, as double mutant plants show rapid dampening of circadian rhythms in constant light and dark conditions. CCA1 and LHY appear to be partially redundant, but are essential proteins if the clock is able to continue running in a sustained manner. They are critical players in the negative limb of the circadian clock mechanism in *Arabidopsis*, but it would not be surprising if more components were yet to be discovered.

If CCA1 and LHY are critical negative regulators of the plant clock, what are the positive clock regulators? In 1995, a number of *Arabidopsis* short- and long-period mutants were identified in a screen using *cab::luc* luminescent transgenic plants. One mutant, called *timing of cab expression* (*toc1*), was a semi-dominant mutation, leading to a shortening of both the *cab* gene expression rhythm and the normal oscillation in primary leaf movement. Five years later, the *toc1* gene was cloned and was shown to encode a protein containing several previously identified domains; an atypical response regulator receiver domain, found in some bacterial two-component signal transduction systems, and an acidic domain as well as a region found in several CONSTANS family transcription factors. TOC1 was likely to play a critical role as a transcription factor, and the fact that its mRNA levels oscillated with a peak at the end of the subjective day supported its role as a central clock component.

The fact that *cca1* and *lhy* transcripts oscillate 12 h out of phase with TOC1 transcript levels strongly suggests that a form of oscillatory feedback loop is likely to exist between these genes. Further analysis of *toc1* mRNA in plants overexpressing CCA1 or LHY showed that TOC1 transcript levels ceased to oscillate and declined to below levels seen at the trough point in matching wild-type strains. CCA1 and LHY are, therefore, likely to act negatively on the transcription rate of TOC1. This was confirmed by an analysis of the *toc1* promoter, where binding of CCA1 and LHY directly to regulatory elements was shown to occur. This element, known as the *evening element* (*EE*) was found to be common to a number of genes that showed peak expression in the late subjective day. Two mutant alleles of *toc1* have been identified. The *toc1-1* allelic form was the first mutant identified, and is a semi-dominant mutation, which retains some TOC1 function. *toc1-2* is a recessive mutation, and most likely represents a loss of function of TOC1. An examination of *cca1* and *lhy* mRNA levels in *toc1-2* mutants revealed a dramatic decrease in the peak amplitude of both

rhythms. It has been proposed, therefore, that TOC1 represents a component of the positive limb of the circadian clock mechanism, acting as a positive transcriptional regulator of both *cca1* and *lhy* expression. However, the fact that both *cca1* and *lhy* both continue to oscillate in the *toc1-2* null mutant must mean that other components that perform this role also exist. A schematic of how these molecules may fit together can be seen in Fig. 7.

3.5.3 Output

Perhaps, the "oldest" of all circadian outputs, or at least the first described, can be attributed to higher plants, with the description of leaf movements in constant darkness by de Mairan in 1727. This assay is still used regularly today in mutant screens of *Arabidopsis* seedlings. A wide range of plant physiological processes are under clock control, including stomatal opening, CO_2 assimilation, and the timing of flower opening in some species. Plants that grow in desert environments are under very strong selective pressure-to-time aspects of their physiology to certain times of day, as is true of course for desert-dwelling animals as well. Crassulacean acid metabolism, found in what are generally termed *CAM* plants, is a particular adaptation to such environments and enables such plants to fix CO_2 during the night while keeping stomata closed during the day. The temporal regulation of such stomatal closure is obviously critical if water loss is to be minimized. However, the exact balance of clock regulation, versus that of acute environmental cues, on CAM function is not yet completely clear, especially as this type of metabolism cannot continue under constant light conditions.

An analysis of clock-regulated gene expression, using microarray technology or "DNA chips," has been performed in *Arabidopsis* by several groups. Estimates of between 2 and 6% of all genes are believed to be under clock control, though the final number is likely to be much higher. Perhaps, not surprisingly, genes involved in the light-harvesting reactions of photosynthesis showed strong circadian regulation, as did the genes involved in starch mobilization, nitrogen metabolism, and many other such processes. The study of Harmer and colleagues revealed the presence of the "evening element" within the upstream promoter regions of some 31 cycling genes that showed peak expression in the late subjective day. This "evening element" is composed of a remarkably conserved 9-nucleotide motif, AAAATATCT, which appears to be important for clock regulation of these rhythmic genes. This global clock regulation of gene expression bears remarkable similarity in scale, if not detail, to that seen in a number of mammalian tissues and shows how central the circadian clock is to the physiology of most organisms.

Of course, one of the most remarkable seasonal responses of many plant species is the annual regulation of flowering and reproduction. Such processes are regulated by measurement of day length as a photoperiodic response. Though this topic is beyond the scope of this review, it is interesting to note that several clock genes have been implicated in the regulation of flowering in *Arabidopsis*. This is particularly true for the *toc1-1* mutant discussed above. *toc1-1* plants are unresponsive to photoperiod when grown on light–dark cycles of 24 h (a normal day length). However, when grown on cycles of 21 h, which matches the natural free-running period of this mutant, individuals can distinguish between short and long days. In addition, overexpression of the

circadian negative regulators CCA1 and LHY will delay flowering on long days. While the *elf3* mutant also influences the time of flowering, two additional mutants that alter flowering time are of particular interest. Both ZEITLUPE (ZTL) and FKF1 represent a similar class of molecule that contains a LOV domain, a subgroup of PAS domain proteins, as well as F-box motifs. F-boxes have been strongly implicated in targeting proteins for degradation via the ubiquitin pathway and LOV domains have been implicated in events relating to blue light photoreception. Mutations in both of these genes have a robust effect on *Arabidopsis* flowering events. Mutations in *ztl* also lengthen the free-running period of genes under clock control. So, it is clear, as one would expect, that there is a strong and intimate link between genes involved in the generation and regulation of the daily or circadian clock and the expression of annual rhythms, such as flowering.

3.6 Cyanobacteria

The cyanobacteria are the first prokaryotes that have been shown to possess a circadian pacemaker. This was an important step forward in expanding our understanding of the evolution of clocks and an added example of how incredibly widespread the phenomena of circadian oscillations really are. It was long assumed that prokaryotes, with their generally very fast (shorter than 24 h) cell cycles, would not possess or have a need for a clock, but this is clearly not true. Nor is the presence of a nucleus (a new form of biological snobbery) required either.

With the recent sequencing of many bacterial genomes, it has become apparent that the presence of one central cyanobacterial clock gene, *kaiC*, is more widespread than simply within the cyanobacteria, being found also in the Archaea and Proteobacteria (though only the cyanobacteria contain the complete family of the *kai* genes; see below). This is of course not proof that clocks are widely distributed amongst many bacteria but certainly demonstrates the very early evolution of genes that are involved in the clock mechanism. The first *kai* genes may have appeared some 3500–3800 million years ago, before the presence of an oxygen atmosphere on the earth. Cyanobacteria, of course, tend to be photosynthetic, and as such the potential value of a circadian clock is clear. It is not yet known how widespread functional clocks may be amongst prokaryotes. However, as many human disease states are caused by bacterial infection and other prokaryotic organisms, the presence of a clock and the timing of certain physiological processes to a particular time of day may prove to be a very valuable clinical tool and have considerable medical implications.

As we shall see, most of the molecules of the cyanobacteria, and how they may work, appear quite alien (and unique) to individuals more used to mammalian transcriptional regulation. However, many of the basic aspects of how one generates an oscillation are similar, including the critical importance of negative feedback with delay. The cyanobacteria provide a fascinating and unique view of alternative ways of generating a circadian oscillation.

3.6.1 Entrainment

Relatively little is currently known about light detection and the details of the input pathway to the clock in cyanobacteria. Changes in the lighting regime (timing

of the LD cycle) are capable of reentraining the clock in *Synechococcus* RF-1, but few, if any, typical short light pulse, phase shifting experiments have been performed in this species. However, recently, Golden and colleagues did isolate a mutant (*Synechococcus* sp. PCC7942) that showed an impaired ability to phase shift in response to 5-h dark treatments. The gene underlying this mutation was isolated and called *circadian input kinase* (*cikA*). *cikA* encodes for a bacterial phytochrome, containing histidine kinase motifs, which raised the possibility that this molecule might be acting as a putative bacterial photopigment. The importance of phytochromes in the circadian system of higher plants added support to this idea. However, sequence comparison shows that *cikA* lacks a critical amino acid residue thought to be essential for binding of a required chromophore. In addition, the optimal wavelengths for phase shifting the clock appear to be in both the blue and red regions of the spectrum, with green and far red having little or no effect. This spectral response is not quite what one might predict for a typical phytochrome response nor is it similar to that one would expect for a photosynthetic response. At present, the photopigment responsible for *Synechococcus* entrainment remains unknown, and further data will be required to confirm the exact role of *cikA* in the clock input pathway.

Within the last year, a new class of cryptochrome flavoproteins (called *cryptochrome DASH*) has been isolated in the related cyanobacteria, *Synechocystis*. This gene has considerable sequence homology to photolyases but has the capacity to act as a transcriptional repressor. Whether this bacterial cryptochrome will have a central clock role or potential photoreceptive role as in other circadian systems still remains to be tested. However, the presence of cryptochrome in cyanobacteria is interesting, if not least for the fact that it reveals how early this class of molecule appeared on an evolutionary timescale.

It is perhaps important, especially when dealing with an unusual system like cyanobacteria, to keep an open-mind about the possible mechanisms of light detection and not get locked in a very mammalian, classical signal transduction perspective. Light, for example, has profound effects on the RNA stability of a number of cyanobacterial transcripts. These include transcripts like *psbA2*, a molecule centrally involved in the function of photosystem II in the photosynthetic machinery and believed to be a rhythmic output of the circadian clock. A short stretch of the 5′-untranslated region (5′UTR) of this transcript contains an "AU-box" (UAAAUAAA), which, when deleted, removes light-dependent regulation of *psbA2* expression. This region of RNA appears to increase transcript instability in the dark and may aid accumulation in the light via putative positive and negative cis-acting factors. Light is acting in a posttranscriptional manner to influence transcript levels and presumably subsequent protein levels. At this time, no evidence for such a mechanism in the circadian input pathway, or in stabilizing central clock transcripts, has been demonstrated.

3.6.2 Oscillatorff

Identification of central clock genes in cyanobacteria has progressed through the use of mutant screens/forward genetics, as in most other circadian model systems. The prerequisite for such screening is the availability of a reliable wild-type phenotype against which to compare potential mutants. Though there are a number of rhythmic outputs in cyanobacteria, including nitrogen fixation, photosynthetic

activity, and cell cycle regulation, these are not necessarily the easiest of assays to use with which to identify mutant strains. One major advantage of bacteria, of course, is the relative ease with which one can introduce genes and stably integrate them in the genome via homologous recombination. The generation of a luminescent reporter strain, with luciferase under the control of a clock gene promoter, would provide a very clean and relatively easy assay for circadian rhythmicity and the subsequent isolation of mutants. To be more exact, the upstream promoter region of the gene *psbA1* (another molecule involved in the function of photosystem II) was fused to luciferase (*luxAB*) and the luminescent reporter strain AMC149 of *Synechococcus* sp. PCC7942 was generated. This strain showed robust circadian oscillations, showing that the promoter of *psbA1* was under clock control.

Interestingly, the random insertion of a luciferase reporter, in a "promoter trap" style experiment, revealed something quite unexpected. Of approximately 3000 randomly transformed colonies, 800 were sufficiently bright to allow for imaging by use of a cooled CCD camera system. When examined for circadian cycling, almost all of these 800 transformants showed circadian oscillations in bioluminescence with a range of amplitudes and waveforms. Some 80% of these colonies had the same phase relationship to the light–dark cycle as *psbA1*. The majority, if not all, of the genes in *Synechococcus* are under global clock control. The default state for gene transcription in this system is to oscillate, and so it appears that there must be some non-gene-specific regulation of the transcriptional machinery. This idea was extended by the fact that even heterologous promoters, in particular the promoter of *conII* from *Escherichia coli*, show robust circadian oscillations when inserted into cyanobacteria.

Additional evidence that supports the idea of a global circadian regulation of transcriptional state includes the fact that when the recently identified clock genes (*kaiBC*) in cyanobacteria were placed under the control of a heterologous promoter, the transcripts from these genes continued to oscillate. In other words, the exact nature of the promoter, in terms of sequence specifics, did not seem to matter in terms of regulating rhythmicity, and so the regulation must be at a more general level. Furthermore, one of the new clock genes, *kaiC*, bore considerable resemblances to bacterial recombinases and helicases, that is, molecules involved in the general winding and unwinding of chromosomal DNA. Though this was not a proof of a novel transcriptional mechanism, this certainly supported the idea of a global regulation of transcription by changing the supercoiled state of the bacterial chromosomal DNA. Johnson and colleagues have proposed a model, therefore, where this global rhythm in the transcription of apparently all cyanobacterial genes is controlled by large-scale changes in the state of the bacterial chromosome, with regulation occurring as a change in the supercoiled nature of the DNA. The state of the whole "nucleiod" changing rhythmically and so being termed (somewhat disagreeably) the *oscilloid*. Similar changes in gross DNA state have been reported in the chromosomal DNA of chloroplasts found in *Chlamydomonas*, and as chloroplasts are thought to have originated from the inclusion of endosymbiotic cyanobacteria, this also tends to support the rhythmic supercoiling regulation of DNA as a means of achieving rhythmic gene transcription.

The generation of a luminescent reporter strain of cyanobacteria allowed for a

large scale, semiautomated mutant screen to be performed. Approximately 500 000 individual clones were mutagenized by EMS, and then screened for period length defects or other circadian phenotypes. A large number of mutants were isolated, of which 12 displayed altered period ranging from 16 to 60 h and a number of others showed complete arrhythmia. The exact genes mutated in such a screen could be identified by rescuing the given mutants through complementation with libraries constructed from wild-type *Synechococcus* DNA, with more than 30 mutants rescued so far in this manner. This rapid and compelling method of identifying mutant genes is a far cry from the laborious gene mapping approaches required in vertebrates, and allowed for the efficient isolation of the critical genes involved in rhythm generation.

The majority of these mutations revealed three critical clock genes in *Synechococcus*, called *kaiA*, *kaiB*, and *kaiC*, which exist in a tight gene cluster referred to simply as *kai* (kai meaning *rotation* or *cycle* in Japanese). Nineteen of the mapped mutations can be localized to these three genes. The *kai* genes appear to be unique clock molecules in that, based on sequence comparison, they bear no resemblance to clock genes identified in other species (and as such may represent the independent evolution of a circadian clock). All three genes appear to be critical for normal clock function, with *kaiB* and *kaiC* being cotranscribed and thus forming a bacterial operon. Whereas *kaiA* transcription appears to be under independent regulation, *kaiC* is the largest of the three genes and as such is the site of most mutations affecting period length. Sequence comparison places kaiC as a potential member of the RecA/DnaB family of proteins, suggesting, as mentioned above, a potential role in the regulation of DNA structure. RecA proteins act as ATP-dependent DNA recombinases and DnaB proteins function as replication fork helicases. kaiC is able to bind ATP through the presence of two Walker A motifs (regions involved in ATP binding). Mutation of these Walker A regions disrupts ATP binding and has a seriously disruptive effect on the generation of circadian oscillations. This ATP binding event appears to be essential for the function of kaiC proteins, as in the presence of ATP kaiC associates to form homohexamers, which are visible as rings with a central pore when examined under an electron microscope. As such, kaiC appears to be able to bind directly to forked DNA, again supporting the idea of a role in regulating the global status of DNA supercoiling in *Synechococcus*.

All three genes show a peak in expression at around CT 8, when maintained in continuous light, with the peak in kaiB and kaiC protein expression occurring some 8 h later at CT 16. Such large delays between peak transcript levels and protein levels have been seen in systems such as *Drosophila*, and may be an essential requirement for the generation of long, 24-h periods, as seen in the circadian clock. Such a large lag requires considerable posttranscriptional regulation of the nascent RNA, with presumably mechanisms in place to delay translational activation, as well as processes involved in regulating RNA stability. The delays supposedly produced by controlled nuclear entry clearly cannot occur in a system lacking a nucleus.

The three *kai* genes form an interactive feedback loop responsible for the generation of the circadian rhythm (Fig. 8). All three proteins appear to be able to interact with each other when tested in a yeast protein interaction assay. If kaiC is

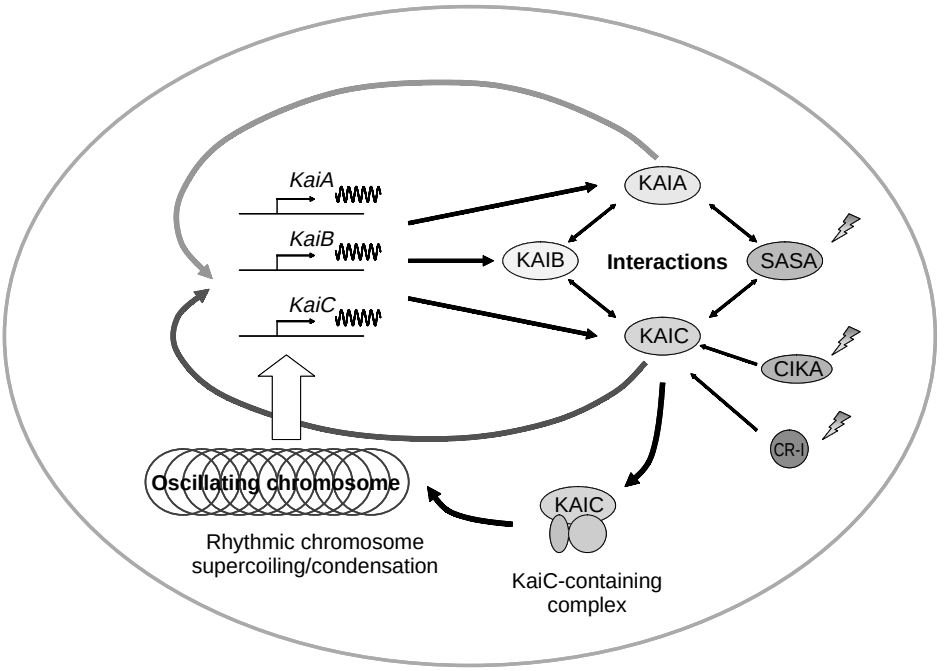

Fig. 8 Model of the molecular clock of Cyanobacteria. The central clock mechanism in cyanobacteria appears to be based on the Kai gene complex and proteins (*kaiA, kaiB,* and *kaiC*), as well as the potential rhythmic coiling and uncoiling of the bacterial chromosome. This global change in chromatin structure may underlie the circadian rhythmicity seen in all genes studied to date. *kaiB* and *kaiC* are cotranscribed as part of a bacterial operon, but all three genes are rhythmic with a peak in expression in the mid-subjective day. Protein levels peak some 8 h later, around CT16 in the subjective night, showing the typical transcription–translation delay seen in many circadian systems. KAIC appears to be capable of binding forked DNA, and as such may play a role in regulating the global genes seen in chromatin structure. The three Kai proteins can interact to form a Kai protein complex, possibly including the protein SASA. KAIC feeds back negatively to repress transcription of itself and *kaiB*, thus representing the core negative feedback element of the clock. KAIA appears to act positively on the transcription of *kaiB* and *kaiC*, along with SASA, and so these proteins represent the positive limb of the circadian clock mechanism. Entrainment of this core oscillator may occur through the action of the bacterial phytochrome *cikA*, but further data will be required to confirm this. SASA appears to play some role in adjusting the sensitivity of the oscillator to changes in light intensity. A mutation in *Synechococcus* called *CR-1* shows alteration in entrainment to light–dark cycles, and so a CR-1 protein may play a role in clock resetting. However, at this time, the photopigment underlying cyanobacterial entrainment has not been definitively identified.

overexpressed, then the subsequent levels of kaiB and kaiC are reduced and the oscillation suppressed. Overexpression of *kaiA* acts to increase the transcription of *kaiBC,* while disruption of *kaiA* lowers levels of *kaiBC*. Therefore, kaiC appears to act as a repressor of its own transcription, forming the negative feedback part of the circadian cycle, with kaiA acting as an activator and the positive limb of the circadian mechanism. Yeast two-hybrid screening was performed to

isolate additional proteins capable of interacting with kaiC. This screen led to the isolation of *sasA*, believed to be similar to sensory histidine kinases. Disruption of *sasA* dramatically lowered the amplitude of the circadian rhythm, and lowered *kaiBC* expression levels, while overexpression completely eliminated the rhythm. sasA is believed to be an additional positive regulator of *kaiBC* transcription (possibly though binding with kaiC protein), and as such forms a second positive feedback loop, which appears important for the generation of robust circadian oscillations, especially when *Synechococcus* is placed on an LD cycle. It seems likely, from previous discussions, that this negative and positive transcriptional regulation (especially that of kaiC) acts through a global change in the coiled status and structural state of the bacterial chromosome (Fig. 8). It may be of interest that the recent discussion regarding the regulation of the mammalian circadian clock, especially by Reppert and colleagues, has focused on the importance of chromatin remodeling as a critical event within the central mechanism of the mammalian clock.

3.6.3 Output

The cyanobacteria are made up of a wide range of individual species, with varying biochemical and physiological responses, and as such one should be cautious about overgeneralization. However, two clear clock-regulated outputs in these cells are the processes of photosynthesis and nitrogen fixation. These processes have conflicting physiological conditions to be performed optimally, as nitrogen fixation occurs optimally in low oxygen/anaerobic conditions. The temporal separation of these biochemical processes has a clear selective advantage, and so it is not surprising that in *Synechococcus*, therefore, photosynthesis takes place during the day and nitrogen fixation during the night. This is not true, however, for all cyanobacteria, as *Trichodesmium* shows peaks in both activities during the day. The rhythm in nitrogen fixation appears to be regulated by a nocturnal increase in nitrogenase mRNA levels in *Synechococcus*.

Perhaps, a more surprising phenomenon is that the timing of cell division also appears to be clock regulated, or to be more exact clock gated. Under conditions of rich, unlimiting media, these cells can divide as rapidly as once in every 6–10 h (so up to three times per circadian cycle). When the rhythm in bioluminescence is measured in these "subconfluent cultures," amazingly, a strong circadian rhythm in gene expression can still be measured. The amplitude of this rhythm increases during the exponential phase of cell growth (probably due to an increase in cell number) and then stabilizes during the sustained phase of growth. As these cells are dividing at a much higher rate than the period of the clock, the exact phase of the rhythm must be inherited by daughter cells, which keep the oscillation running in perfect synchrony. Moreover, when the stage of the cell cycle is examined by flow cytometry, it becomes clear that the exact timing of cell division is restricted or gated to certain times of the day. On an LD cycle, *Synechococcus* cells only divide during the day and not at night. When placed into continuous light, cell division is blocked during the early subjective night, but can occur during the late subjective night and subjective day. A direct link between the clock machinery and the regulation of the cell cycle mechanism must, therefore, occur to allow for such "gated" regulation. It is

tempting to speculate, of course, that the influence of kaiC on DNA structure may play an important role in the regulation of this event.

The number of mutants and the ease with which bacteria can be grown has allowed for an interesting series of experiments to be performed, examining fitness under differing lighting regimes. Three strains of *Synechococcus* were used, wild type with a period of 25 h, C28a with a long period of around 30 h, and C22a with a short period of 23 h. Importantly, when grown as isolated cultures, no difference in growth rate between the differing strains was detectable. The three strains were then mixed together and placed on one of the three differing T-cycles: either an 11L:11D, 22-h day; a 12L:12D, 24-h day; or a 15L:15D, 30-h day. In all the situations tested, the strain that grew optimally and outcompeted the other strains was the one whose endogenous period most closely resembled that of the environmental light–dark cycle. So, C22a was the most competitive on a 22-h day, growing better than wild type, while C28a (with an endogenous period of 30 h) was most successful when grown on a 30-h day. The mutant phenotypes were clearly not generally impaired varieties, being much better adapted than wild types to the abnormal lighting conditions. Exactly how this improved fitness or competitive advantage is conveyed to each strain is not clear. Though, speculation includes the fact that the expression of *psbA1* is mistimed (incorrectly phased) when on deviant light–dark cycles and that this suboptimal phase angle might lead to relatively less efficient photosynthesis and, therefore, growth rates. The importance of an endogenous clock, even to organisms as "simple" as bacteria, has finally been demonstrated.

4 General Conclusions and Future Perspectives

The study of circadian rhythms has benefited from the interest of researchers with diverse backgrounds who have studied different groups of organisms. This broadly comparative approach has allowed results in one species to guide the development of experiments in other groups, and as a result, progress on the study of circadian rhythms has advanced across a fairly broad taxonomic front. This has been particularly true for the study of the molecular basis of the clock. As discussed above, several of the genes involved in the circadian clock in the mouse and *Drosophila* have been conserved between these highly divergent organisms. This conservation has not only aided the hunt for additional clock molecules but has had important implications for understanding clock evolution. The fact that *Drosophila* and mice share homologous clock genes suggests that there was not only a clock some 700 million years ago in the common ancestor of insects and mammals but that this clock was built along broadly similar lines to the circadian clocks we find in animals today. If we look back beyond the animal branch of the evolutionary tree and examine the molecular basis of the clock in plants, fungi, or even photosynthetic bacteria, the genes identified thus far bear no real resemblance to any of the clock genes in animals. They are not homologous, leading to the suggestion that the biological clock may have evolved multiple times during the course of evolution. It is important to stress, however, that although different sets of genes seem to generate the clock in animals, plants, fungi, and bacteria, they use the same fundamental mechanism – an "autoregulatory negative

feedback loop'' involving several genes. It is these circadian feedback loops that have formed the central focus of the discussion in this article.

Although the basic mechanisms of the molecular clock seem to be in place, the devil is in the details. The very complexity of the models that are emerging, with their cycles and epicycles of positive and negative drives and feedback loops, resemble to some the ungainly Ptolemaian versions of the solar system! There are some very serious questions about the orthodox explanations of the circadian mechanisms that we have described. For example, clock genes probably have other functions; indeed, the primary role of some of the genes we have described may not be as elements of the clock. It has been assumed that clocks have been built from special proteins dedicated to time keeping. This may not be the case. Roenneberg argues that it is critical to consider that before clocks there had to be complex metabolic networks that were driven by light and temperature and that these metabolic pathways may still be closely associated with or even embedded within the clock. Mutant screens would never identify such metabolic genes as ablation or mutation would result in death. The orthodox models that we have described may not be wrong; the genes are clearly critical for clock function, but we cannot exclude the possibility that the clock is the product of interacting metabolic elements and the genes that we have described are epiphenomena and part of a fine-tuning mechanism.

Freedom is something that most humans strive for in one form or another. In advanced industrialized societies, the freedom to sleep, work, eat, drink, or travel when we want is becoming a real option for many, and generally applauded. But the freedom to do whatever we want, when we want, has brought an appreciation that we are not free of the ticking of our biological clock. The biological order that the clock imparts means that we are unable to perform with the same efficiency throughout the 24-h day and we ignore our biology at our peril (Fig. 1). The emerging fields of chronotoxicology and chronopharmacology provide some of the most striking examples of the relevance of circadian physiology to our daily lives. In studies of a chemical food additive tested in nocturnal mice for potential carcinogenicity, the additive appeared to be perfectly safe and approved for human use. However, the test were undertaken at 12:00 h, corresponding to the midresting phase in the mice, and at which time the rate of DNA synthesis in the liver was at its lowest and least susceptible to a potential carcinogen. When the same dose was administered at 24:00 h, the midactivity phase of mice and crucially the time of maximum DNA synthesis in the liver, 40% of the animals developed liver cancers. Because physiology is dynamic, the response of an organism to any manipulation must take the time of day into account. This knowledge is already being exploited in the treatment of cancer, where differences in the timing of cell division of host and tumor cells may be exploited to maximize treatment efficacy and minimize adverse side effects. The importance of circadian rhythms in cancer therapy is reinforced by a recent link between clock genes and cell-growth control in long-term studies of mPer2 mutant mice. These animals appear unusually cancer prone as well as susceptible to γ-radiation. A number of genes involved in cell cycle regulation and tumor suppression appear deregulated in mPer2 mutants, suggesting that this component of the molecular

clock may be vital in DNA damage responsive pathways.

A recent study by Toh et al. has provided an additional connection between the seemingly disparate worlds of molecular clocks and medicine, with the identification of a mutation in *hPer2* associated with Familial advanced sleep phase syndrome (FASPS). This condition is characterized by a shortened circadian period and an advanced sleep phase. Individuals go to bed on average at 21:00 h and wake at 05:00 h. It is remarkable that this dramatic change in behavior appears to arise from a single serine to glycine substitution in the human *Per2* gene, which yields a protein much less effectively phosphorylated by CK1ε. Similarly, two recent studies have suggested that diurnal preferences may be linked to *hPer3*, either via a repeat motif or a polymorphism within this gene. This latter study also located a polymorphism in the *NPAS2* gene, which appears to provide a link between circadian rhythms and seasonal affective disorder. Recent advances in the study of narcolepsy have provided additional clues as to how sleep–wake behavior may be driven from molecular components. The discovery of a class of neuropeptides termed *orexins (hypocretins)* involved in both regulating food intake and wakefulness may provide an additional target of clock output. In narcoleptics, deficiencies of the orexin system lead to excessive daytime sleepiness, cataplexy (spontaneous sleep episodes, often, at times of heightened emotional arousal), and hypnagogic hallucinations (presleep dreamlike episodes).

Our emerging understanding of the molecular basis of circadian rhythmicity is in identifying compelling links between a variety of human illnesses and the molecular components of the clock. This understanding will not only provide us with the means to correct these defects but also the means to manipulate normal circadian rhythms to allow us to function under abnormal temporal conditions. Edward O. Wilson stated: ''Homo sapiens, the first truly free species, is about to decommission natural selection, the force that made us. . .. Soon we must look deep within ourselves and decide what we wish to become.'' Shortly, we will have to make some moral choices. We are a diurnal species living in a 24-h world and subject to our chronobiology. We need to ask the question: should we use pharmacological or light treatments to alter patterns of gene and protein expression to manipulate the biological consequences of working at night, and embrace fully the 24/7 society, or should we reject this trend and use the knowledge of our internal temporal structure to become ''time-wise''. Should temporal freedom be something we should strive for or reject?

See also Behavior Genes; *Drosophila* Genome; Thale Cress (*Arabidopsis thaliana*) Genome; Zebrafish (*Danio rerio*) Genome and Genetics.

Bibliography

Books and Reviews

Albrecht, U. (2002) Invited review: regulation of mammalian circadian clock genes, *J. Appl. Physiol.* **92**, 1348–1355.

Ashmore, L.J., Sehgal, A. (2003) A fly's eye view of circadian entrainment, *J. Biol. Rhythms* **18**, 206–216.

Aschoff, J. (1981) *Biological Rhythms*, Plenum Press, New York, London.

Bell-Pedersen, D. (2000) Understanding circadian rhythmicity in Neurospora crassa: From behavior to genes and back again, *Fungal Genet. Biol.* **29**, 1–18.

Balsalobre, A. (2002) Clock genes in mammalian peripheral tissues, *Cell Tissue Res.* **309**, 193–199.

Buijs, R.M., Kalsbeek, A. (2001) Hypothalamic integration of central and peripheral clocks, *Nat. Rev. Neurosci.* **2**, 521–526.

Burns, E.R. (2000) Biological time and *in vivo* research: A field guide to pitfalls, *Anat. Rec.* **261**, 141–152.

Carrier, J., Monk, T.H. (2000) Circadian rhythms of performance: new trends, *Chronobiol. Int.* **17**, 719–732.

Dodd, A.N., Borland, A.M., Haslam, R.P., Griffiths, H., Maxwell, K. (2002) Crassulacean acid metabolism: plastic, fantastic, *J. Exp. Bot.* **53**, 569–580.

Duffield, G.E. (2003) DNA microarray analyses of circadian timing: the genomic basis of biological time, *J. Neuroendocrinol.* **15**, 911–1002.

Dunlap, J.C. (1999) Molecular bases for circadian clocks, *Cell* **96**, 271–290.

Dunlap, J.C., Loros, J.J., DeCoursey, P.J. (2004) *Chronobiology; Biological Timekeeping*, Sinauer, Sunderland, MA.

Ebling, F.J. (1996) The role of glutamate in the photic regulation of the suprachiasmatic nucleus, *Prog. Neurobiol.* **50**, 109–132.

Foster, R.G. (1998) Shedding light on the biological clock, *Neuron* **20**, 829–832.

Foster, R.G. (2001) Photoreceptors, Clocks & Criteria, in: Coohill, T.P., Valenzeno, D.P. (Eds.) *Photobiology for the 21st Century*, Valdenmar Publishing Company, Overland Park, Kansas, pp. 183–198.

Foster, R.G., Kreitzman, L. (2004) *Rhythms of Life: The biological Clocks that Control the Daily Lives of Every Living Thing*, Profile Books, London, UK.

Foster, R.G., Provencio, I. (1999) The Regulation of Vertebrate Biological Clocks by Light, in: Archer, S.N., Djamgoz, M.B.A., Loew, E.R., Partridge, J.C., Vallerga, S. (Eds.) *Adaptive Mechanisms in the Ecology of Vision*, Kluwer Academic Publishers, Dordrecht, Netherlands, pp. 223–243.

Fu, L., Lee, C.C. (2003) The circadian clock: pacemaker and tumour suppressor, *Nat. Rev. Cancer* **3**, 350–361.

Gillette, M.U., Tischkau, S.A. (1999) Suprachiasmatic nucleus: the brain's circadian clock, *Recent Prog. Horm. Res.* **54**, 33–58. discussion 58–39.

Helfrich-Förster, C., Engelmann, W. (2001) Photoreceptors for the Circadian Clock of the Fruitfly, in: Kumar, V. (ed.) *Circadian Rhythms*, Narosa Publishing House, New Delhi, India.

Hrushesky, W.J. (2001) Tumor chronobiology, *J. Control. Release* **74**, 27–30.

Johnson, C.H. (2001) Endogenous timekeepers in photosynthetic organisms, *Annu. Rev. Physiol.* **63**, 695–728.

Johnson, C.H., Golden, S.S. (1999) Circadian programs in cyanobacteria: adaptiveness and mechanism, *Annu. Rev. Microbiol.* **53**, 389–409.

Lakin-Thomas, P.L., Cote, G.G., Brody, S. (1990) Circadian rhythms in Neurospora crassa: biochemistry and genetics, *Crit. Rev. Microbiol.* **17**, 365–416.

Lucas, R.J., Foster, R.G. (1999) Circadian Clocks: A *cry* in the dark? *Curr. Biol.* **9**, 825–828.

Moore-Ede, M.C., Sulzman, F.M., Fuller, C.A. (1982) *The Clocks that Time Us; Physiology of the Circadian Timing System*, Harvard University Press, Cambridge, MA.

Mori, T., Johnson, C.H. (2000) Circadian control of cell division in unicellular organisms, *Prog. Cell Cycle Res.* **4**, 185–192.

Mori, T., Johnson, C.H. (2001) Circadian programming in cyanobacteria, *Semin. Cell Dev. Biol.* **12**, 271–278.

Mormont, M.C., Levi, F. (2003) Cancer chronotherapy: principles, applications, and perspectives, *Cancer* **97**, 155–169.

Mrosovsky, N. (1990) *Rheostasis The Physiology of Change*, Oxford University Press, New York, Oxford.

Pittendrigh, C.S. (1981) Circadian Systems: Entrainment, in: Aschoff, J. (Ed.) *Handbook of Behavioral Neurobiology: Biological Rhythms*, Plenum Press, New York, London, pp. 95–124.

Pittendrigh, C.S. (1993) Temporal organization: reflections of a Darwinian clock-watcher, *Annu. Rev. Physiol.* **55**, 17–54.

Prince, V. (2002) The Hox Paradox: More complex(es) than imagined, *Dev. Biol.* **249**, 1–15.

Rajaratnam, S.M., Arendt, J. (2001) Health in a 24-h society, *Lancet* **358**, 999–1005.

Reppert, S.M., Weaver, D.R. (2002) Coordination of circadian timing in mammals, *Nature* **418**, 935–941.

Richter, C.P. (1967b) Psychopathology of periodic behavior in animals and man, *Proc. Annu. Meet. Am. Psychopathol. Assoc.* **55**, 205–227.

Ripperger, J.A., Schibler, U. (2001) Circadian regulation of gene expression in animals, *Curr. Opin. Cell Biol.* **13**, 357–362.

Roenneberg, T., Foster, R.G. (1997) Twilight times: light and the circadian system, *Photochem. Photobiol.* **66**, 549–561.

Roenneberg, T., Merrow, M. (1999) Circadian systems and metabolism, *J. Biol. Rhythms* **14**, 449–459.

Roenneberg, T., Merrow, M. (2001) Circadian systems: different levels of complexity, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **356**, 1687–1696.

Rosbash, M., Takahashi, J.S. (2002) Circadian rhythms: the cancer connection, *Nature* **420**, 373–374.

Shand, J., Foster, R.G. (1999) The Extraretinal Photoreceptors of Non-mammalian Vertebrates, in: Archer, S.N., Djamgoz, M.B.A., Loew, E.R., Partridge, J.C., Vallerga, S. (Eds.) *Adaptive Mechanisms in the Ecology of Vision*, Kluwer Academic Publishers, Dordrecht, Netherlands, pp. 197–222.

Skrabanek, L., Wolfe, K.H. (1998) Eukaryote genome duplication – where's the evidence? *Curr. Opin. Genet. Dev.* **8**, 694–700.

Smolensky, M.H. (1996) Chronobiology and chronotherapeutics. Applications to cardiovascular medicine, *Am. J. Hypertens.* **9**, 11S–21S.

Taheri, S., Zeitzer, J.M., Mignot, E. (2002) The role of hypocretins (orexins) in sleep regulation and narcolepsy, *Annu. Rev. Neurosci.* **25**, 283–313.

Willie, J.T., Chemelli, R.M., Sinton, C.M., Yanagisawa, M. (2001) To eat or to sleep? Orexin in the regulation of feeding and wakefulness, *Annu. Rev. Neurosci.* **24**, 429–458.

Wilson, E.O. (1998) in: Alfred, A. (Ed.) *Consilience*, Knopf, New York.

Young, M.W., Kay, S.A. (2001) Time zones: a comparative genetics of circadian clocks, *Nat. Rev. Genet.* **2**, 702–715.

Primary Literature

Akashi, M., Tsuchiya, Y., Yoshino, T., Nishida, E. (2002) Control of intracellular dynamics of mammalian period proteins by casein kinase I epsilon (CKIepsilon) and CKIdelta in cultured cells, *Mol. Cell. Biol.* **22**, 1693–1703.

Alabadi, D., Oyama, T., Yanovsky, M.J., Harmon, F.G., Mas, P., Kay, S.A. (2001) Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock, *Science* **293**, 880–883.

Albrecht, U., Sun, Z.S., Eichele, G., Lee, C.C. (1997) A differential response of two putative mammalian circadian regulators, mper1 and mper2, to light, *Cell* **91**, 1055–1064.

Allada, R., White, N.E., So, W.V., Hall, J.C., Rosbash, M. (1998) A mutant *Drosophila* homolog of mammalian Clock disrupts circadian rhythms and transcription of period and timeless, *Cell* **93**, 791–804.

Antoch, M.P., Song, E.J., Chang, A.M., Vitaterna, M.H., Zhao, Y., Wilsbacher, L.D., Sangoram, A.M., King, D.P., Pinto, L.H., Takahashi, J.S. (1997) Functional identification of the mouse circadian Clock gene by transgenic BAC rescue, *Cell* **89**, 655–667.

Archer, S.N., Robilliard, D.L., Skene, D.J., Smits, M., Williams, A., Arendt, J., von Schantz, M. (2003) A length polymorphism in the circadian clock gene Per3 is linked to delayed sleep phase syndrome and extreme diurnal preference, *Sleep* **26**, 413–415.

Aronson, B.D., Johnson, K.A., Loros, J.J., Dunlap, J.C. (1994) Negative feedback defining a circadian clock: autoregulation of the clock gene frequency, *Science* **263**, 1578–1584.

Bae, K., Jin, X., Maywood, E.S., Hastings, M.H., Reppert, S.M., Weaver, D.R. (2001) Differential functions of mPer1, mPer2, and mPer3 in the SCN circadian clock, *Neuron* **30**, 525–536.

Ballario, P., Vittorioso, P., Magrelli, A., Talora, C., Cabibbo, A., Macino, G. (1996) White collar-1, a central regulator of blue light responses in Neurospora, is a zinc finger protein, *EMBO J.* **15**, 1650–1657.

Balsalobre, A., Damiola, F., Schibler, U. (1998) A serum shock induces circadian gene expression in mammalian tissue culture cells, *Cell* **93**, 929–937.

Bao, S., Rihel, J., Bjes, E., Fan, J.Y., Price, J.L. (2001) The *Drosophila* double-timeS mutation delays the nuclear accumulation of period protein and affects the feedback regulation of period mRNA, *J. Neurosci.* **21**, 7117–7126.

Barkai, N., Leibler, S. (2000) Circadian clocks limited by noise, *Nature* **403**, 267–268.

Bellingham, J., Foster, R.G. (2002) Opsins and mammalian photoentrainment, *Cell Tissue Res.* **309**, 57–71.

Benna, C., Scannapieco, P., Piccin, A., Sandrelli, F., Zordan, M., Rosato, E., Kyriacou, C.P., Valle, G., Costa, R. (2000) A second

timeless gene in *Drosophila* shares greater sequence similarity with mammalian tim, *Curr. Biol.* **10**, R512–R513.

Berson, D.M., Dunn, F.A., Takao, M. (2002) Phototransduction by retinal ganglion cells that set the circadian clock, *Science* **295**, 1070–1073.

Blau, J., Young, M.W. (1999) Cycling vrille expression is required for a functional *Drosophila* clock, *Cell* **99**, 661–671.

Brudler, R., Hitomi, K., Daiyasu, H., Toh, H., Kucho, K., Ishiura, M., Kanehisa, M., Roberts, V.A., Todo, T., Tainer, J.A., Getzoff, E.D. (2003) Identification of a new cryptochrome class. Structure, function, and evolution, *Mol. Cell.* **11**, 59–67.

Bunger, M.K., Wilsbacher, L.D., Moran, S.M., Clendenin, C., Radcliffe, L.A., Hogenesch, J.B., Simon, M.C., Takahashi, J.S., Bradfield, C.A. (2000) Mop3 is an essential component of the master circadian pacemaker in mammals, *Cell* **103**, 1009–1017.

Bunney, W.E., Bunney, B.G. (2000) Molecular clock genes in man and lower animals: possible implications for circadian abnormalities in depression, *Neuropsychopharmacology* **22**, 335–345.

Cahill, G.M., Hurd, M.W., Batchelor, M.M. (1998) Circadian rhythmicity in the locomotor activity of larval zebrafish, *Neuroreport* **9**, 3445–3449.

Campbell, S.S., Murphy, P.J. (1998) Extraocular circadian phototransduction in humans, *Science* **279**, 396–399.

Carter, D.A., Murphy, D. (1992) Nuclear mechanisms mediate rhythmic changes in vasopressin mRNA expression in the rat suprachiasmatic nucleus, *Brain Res. Mol. Brain Res.* **12**, 315–321.

Cashmore, A.R., Jarillo, J.A., Wu, Y.J., Liu, D. (1999) Cryptochromes: blue light receptors for plants and animals, *Science* **284**, 760–765.

Cermakian, N., Whitmore, D., Foulkes, N.S., Sassone-Corsi, P. (2000) Asynchronous oscillations of two zebrafish CLOCK partners reveal differential clock control and function, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4339–4344.

Cermakian, N., Monaco, L., Pando, M.P., Dierich, A., Sassone-Corsi, P. (2001) Altered behavioral rhythms and clock gene expression in mice with a targeted mutation in the Period1 gene, *EMBO J.* **20**, 3967–3974.

Cermakian, N., Pando, M.P., Thompson, C.L., Pinchak, A.B., Selby, C.P., Gutierrez, L., Wells, D.E., Cahill, G.M., Sancar, A., Sassone-Corsi, P. (2002) Light induction of a vertebrate clock gene involves signaling through blue-light receptors and MAP kinases, *Curr. Biol.* **12**, 844–848.

Chen, D., Buchanan, G.F., Ding, J.M., Hannibal, J., Gillette, M.U. (1999) Pituitary adenylyl cyclase-activating peptide: a pivotal modulator of glutamatergic regulation of the suprachiasmatic circadian clock, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13468–13473.

Cheng, M.Y., Bullock, C.M., Li, C., Lee, A.G., Bermak, J.C., Belluzzi, J., Weaver, D.R., Leslie, F.M., Zhou, Q.Y. (2002) Prokineticin 2 transmits the behavioural circadian rhythm of the suprachiasmatic nucleus, *Nature* **417**, 405–410.

Cheng, P., He, Q., Yang, Y., Wang, L., Liu, Y. (2003) Functional conservation of light, oxygen, or voltage domains in light sensing, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5938–5943.

Claridge-Chang, A., Wijnen, H., Naef, F., Boothroyd, C., Rajewsky, N., Young, M.W. (2001) Circadian regulation of gene expression systems in the *Drosophila* head, *Neuron* **32**, 657–671.

Collett, M.A., Garceau, N., Dunlap, J.C., Loros, J.J. (2002) Light and clock expression of the Neurospora clock gene frequency is differentially driven by but dependent on WHITE COLLAR-2, *Genetics* **160**, 149–158.

Covington, M.F., Panda, S., Liu, X.L., Strayer, C.A., Wagner, D.R., Kay, S.A. (2001) ELF3 modulates resetting of the circadian clock in Arabidopsis, *Plant Cell* **13**, 1305–1315.

Crosthwaite, S.K., Dunlap, J.C., Loros, J.J. (1997) *Neurospora* wc-1and wc-2: Transcription, photoresponses, and the origins of circadian rhythmicity, *Science* **276**, 763–769.

Crosthwaite, S.K., Loros, J.J., Dunlap, J.C. (1995) Light-induced resetting of a circadian clock is mediated by a rapid increase in frequency transcript, *Cell* **81**, 1003–1012.

Cyran, S.A., Buchsbaum, A.M., Reddy, K.L., Lin, M.C., Glossop, N.R., Hardin, P.E., Young, M.W., Storti, R.V., Blau, J. (2003) vrille, Pdp1, and dClock form a second feedback loop in the *Drosophila* circadian clock, *Cell* **112**, 329–341.

Czeisler, C.A., Shanahan, T.L., Klerman, E.B., Martens, H., Brotman, D.J., Emens, J.S., Klein, T., Rizzo III, J.F. (1995) Suppression of melatonin secretion in some blind patients by exposure to bright light, *N. Engl. J. Med.* **332**, 6–11.

Daan, S., Albrecht, U., van der Horst, G.T., Illnerova, H., Roenneberg, T., Wehr, T.A., Schwartz, W.J. (2001) Assembling a clock for all seasons: are there M and E oscillators in the genes? *J. Biol. Rhythms* **16**, 105–116.

Delaunay, F., Thisse, C., Marchand, O., Laudet, V., Thisse, B. (2000) An inherited functional circadian clock in zebrafish embryos, *Science* **289**, 297–300.

Devlin, P.F., Kay, S.A. (2000) Cryptochromes are required for phytochrome signaling to the circadian clock but not for rhythmicity, *Plant Cell* **12**, 2499–2510.

Dioum, E.M., Rutter, J., Tuckerman, J.R., Gonzalez, G., Gilles-Gonzalez, M.A., McKnight, S.L. (2002) NPAS2: a gas-responsive transcription factor, *Science* **298**, 2385–2387.

Dvornyk, V., Vinogradova, O., Nevo, E. (2003) Origin and evolution of circadian clock genes in prokaryotes, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2495–2500.

Eide, E.J., Vielhaber, E.L., Hinz, W.A., Virshup, D.M. (2002) The circadian regulatory proteins BMAL1 and cryptochromes are substrates of casein kinase Iepsilon, *J. Biol. Chem.* **277**, 17248–17254.

Etchegaray, J.P., Lee, C., Wade, P.A., Reppert, S.M. (2003) Rhythmic histone acetylation underlies transcription in the mammalian circadian clock, *Nature* **421**, 177–182.

Feldman, J.F., Hoyle, M.N. (1973) Isolation of circadian clock mutants of *Neurospora crassa*, *Genetics* **75**, 605–613.

Feldman, J.F., Atkinson, C.A. (1978) Genetic and physiological characteristics of a slow-growing circadian mutant of Neurospora crassa, *Genetics* **88**, 255–265.

Foster, R.G., Lucas, R.J. (1999) Clocks, Criteria & Critical Genes, *Nat. Genet.* **22**, 217–219.

Foster, R.G., Hankins, M.W. (2002) Non-rod, non-cone photoreception in the vertebrates, *Prog. Retin. Eye Res.* **21**, 507–527.

Foster, R.G., Helfrich-Forster, C. (2001) The regulation of circadian clocks by light in fruitflies and mice, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **356**, 1779–1789.

Foster, R.G., Provencio, I., Hudson, D., Fiske, S., DeGrip, W., Menaker, M. (1991) Circadian photoreception in the retinally degenerate mouse (*rd/rd*), *J. Comp. Physiol. [A]* **169**, 39–50.

Freedman, M.S., Lucas, R.J., Soni, B., von Schantz, M., Munoz, M., David-Gray, Z.K., Foster, R.G. (1999) Regulation of mammalian circadian behavior by non-rod, non-cone, ocular photoreceptors, *Science* **284**, 502–504.

Froehlich, A.C., Liu, Y., Loros, J.J., Dunlap, J.C. (2002) White Collar-1, a circadian blue light photoreceptor, binding to the frequency promoter, *Science* **297**, 815–819.

Fu, L., Pelicano, H., Liu, J., Huang, P., Lee, C. (2002) The circadian gene Period2 plays an important role in tumor suppression and DNA damage response in vivo, *Cell* **111**, 41–50.

Gekakis, N., Saez, L., Delahaye-Brown, A.M., Myers, M.P., Sehgal, A., Young, M.W., Weitz, C.J. (1995) Isolation of timeless by PER protein interaction: defective interaction between timeless protein and long-period mutant PERL, *Science* **270**, 811–815.

Gekakis, N., Staknis, D., Nguyen, H.B., Davis, F.C., Wilsbacher, L.D., King, D.P., Takahashi, J.S., Weitz, C.J. (1998) Role of the CLOCK protein in the mammalian circadian mechanism, *Science* **280**, 1564–1569.

Gotter, A.L., Manganaro, T., Weaver, D.R., Kolakowski, L.F. Jr., Possidente, B., Sriram, S., MacLaughlin, D.T., Reppert, S.M. (2000) A time-less function for mouse timeless, *Nat. Neurosci.* **3**, 755–756.

Green, R.M., Tobin, E.M. (1999) Loss of the circadian clock-associated protein 1 in Arabidopsis results in altered clock-regulated gene expression, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 4176–4179.

Hannibal, J., Moller, M., Ottersen, O.P., Fahrenkrug, J. (2000) PACAP and glutamate are co-stored in the retinohypothalamic tract, *J. Comp. Neurol.* **418**, 147–155.

Hardin, P.E., Hall, J.C., Rosbash, M. (1990) Feedback of the *Drosophila* period gene product on circadian cycling of its messenger RNA levels, *Nature* **343**, 536–540.

Harmer, S.L., Hogenesch, J.B., Straume, M., Chang, H.S., Han, B., Zhu, T., Wang, X., Kreps, J.A., Kay, S.A. (2000) Orchestrated transcription of key pathways in Arabidopsis by the circadian clock, *Science* **290**, 2110–2113.

Hattar, S., Lucas, R.J., Mrosovsky, N., Thompson, S., Douglas, R.H., Hankins, M.W., Lem, J., Biel, M., Hofmann, F., Foster, R.G., Yau, K.W. (2003) Melanopsin and rod-cone photoreceptive systems account for all major accessory visual functions in mice, *Nature* **424**(6944), 76–81. Epub 2003 Jun 15.

Helfrich-Forster, C., Winter, C., Hofbauer, A., Hall, J.C., Stanewsky, R. (2001) The circadian clock of fruit flies is blind after elimination

of all known photoreceptors, *Neuron* **30**, 249–261.

Helfrich-Forster, C., Edwards, T., Yasuyama, K., Wisotzki, B., Schneuwly, S., Stanewsky, R., Meinertzhagen, I.A., Hofbauer, A. (2002) The extraretinal eyelet of *Drosophila*: development, ultrastructure, and putative circadian function, *J. Neurosci.* **22**, 9255–9266.

Herzog, E.D., Takahashi, J.S., Block, G.D. (1998) Clock controls circadian period in isolated suprachiasmatic nucleus neurons, *Nat. Neurosci.* **1**, 708–713.

Hogenesch, J.B., Gu, Y.Z., Jain, S., Bradfield, C.A. (1998) The basic-helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5474–5479.

Hogenesch, J.B., Chan, W.K., Jackiw, V.H., Brown, R.C., Gu, Y.Z., Pray-Grant, M., Perdew, G.H., Bradfield, C.A. (1997) Characterization of a subset of the basic-helix-loop-helix-PAS superfamily that interacts with components of the dioxin signaling pathway, *J. Biol. Chem.* **272**, 8581–8593.

Honma, S., Ikeda, M., Abe, H., Tanahashi, Y., Namihira, M., Honma, K., Nomura, M. (1998) Circadian oscillation of BMAL1, a partner of a mammalian clock gene Clock, in rat suprachiasmatic nucleus, *Biochem. Biophys. Res. Commun.* **250**, 83–87.

Honma, S., Kawamoto, T., Takagi, Y., Fujimoto, K., Sato, F., Noshiro, M., Kato, Y., Honma, K. (2002) Dec1 and Dec2 are regulators of the mammalian molecular clock, *Nature* **419**, 841–844.

Hsu, D.S., Zhao, X., Zhao, S., Kazantsev, A., Wang, R.P., Todo, T., Wei, Y.F., Sancar, A. (1996) Putative human blue-light photoreceptors hCRY1 and hCRY2 are flavoproteins, *Biochemistry* **35**, 13871–13877.

Hunter-Ensor, M., Ousley, A., Sehgal, A. (1996) Regulation of the *Drosophila* protein timeless suggests a mechanism for resetting the circadian clock by light, *Cell* **84**, 677–685.

Hurd, M.W., Cahill, G.M. (2002) Entraining signals initiate behavioral circadian rhythmicity in larval zebrafish, *J. Biol. Rhythms.* **17**, 307–314.

Hurd, M.W., Debruyne, J., Straume, M., Cahill, G.M. (1998) Circadian rhythms of locomotor activity in zebrafish, *Physiol. Behav.* **65**, 465–472.

Ikeda, M., Nomura, M. (1997) cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix/PAS protein (BMAL1) and identification of alternatively spliced variants with alternative translation initiation site usage, *Biochem. Biophys. Res. Commun.* **233**, 258–264.

Ishiura, M., Kutsuna, S., Aoki, S., Iwasaki, H., Andersson, C.R., Tanabe, A., Golden, S.S., Johnson, C.H., Kondo, T. (1998) Expression of a gene cluster kaiABC as a circadian feedback process in cyanobacteria, *Science* **281**, 1519–1523.

Iwasaki, H., Taniguchi, Y., Ishiura, M., Kondo, T. (1999) Physical interactions among circadian clock proteins KaiA, KaiB and KaiC in cyanobacteria, *EMBO J.* **18**, 1137–1145.

Iwasaki, H., Williams, S.B., Kitayama, Y., Ishiura, M., Golden, S.S., Kondo, T. (2000) A kaiC-interacting sensory histidine kinase, SasA, necessary to sustain robust circadian oscillation in cyanobacteria, *Cell* **101**, 223–233.

Jackson, F.R., Bargiello, T.A., Yun, S.H., Young, M.W. (1986) Product of per locus of *Drosophila* shares homology with proteoglycans, *Nature* **320**, 185–188.

Jin, X., Shearman, L.P., Weaver, D.R., Zylka, M.J., de Vries, G.J., Reppert, S.M. (1999) A molecular mechanism regulating rhythmic output from the suprachiasmatic circadian clock, *Cell* **96**, 57–68.

Kaneko, M., Hall, J.C. (2000) Neuroanatomy of cells expressing clock genes in *Drosophila*: transgenic manipulation of the period and timeless genes to mark the perikarya of circadian pacemaker neurons and their projections, *J. Comp. Neurol.* **422**, 66–94.

Kazimi, N., Cahill, G.M. (1999) Development of a circadian melatonin rhythm in embryonic zebrafish, *Brain Res. Dev. Brain Res.* **117**, 47–52.

Keesler, G.A., Camacho, F., Guo, Y., Virshup, D., Mondadori, C., Yao, Z. (2000) Phosphorylation and destabilization of human period I clock protein by human casein kinase I epsilon, *Neuroreport* **11**, 951–955.

King, D.P., Zhao, Y., Sangoram, A.M., Wilsbacher, L.D., Tanaka, M., Antoch, M.P., Steeves, T.D., Vitaterna, M.H., Kornhauser, J.M., Lowrey, P.L., Turek, F.W., Takahashi, J.S. (1997) Positional cloning of the mouse circadian clock gene, *Cell* **89**, 641–653.

Kircher, S., Kozma-Bognar, L., Kim, L., Adam, E., Harter, K., Schafer, E., Nagy, F. (1999) Light quality-dependent nuclear import

of the plant photoreceptors phytochrome A and B, *Plant Cell* **11**, 1445–1456.

Kloss, B., Price, J.L., Saez, L., Blau, J., Rothenfluh, A., Wesley, C.S., Young, M.W. (1998) The *Drosophila* clock gene double-time encodes a protein closely related to human casein kinase Iepsilon, *Cell* **94**, 97–107.

Ko, H.W., Jiang, J., Edery, I. (2002) Role for Slimb in the degradation of *Drosophila* Period protein phosphorylated by Doubletime, *Nature* **420**, 673–678.

Koike, N., Hida, A., Numano, R., Hirose, M., Sakaki, Y., Tei, H. (1998) Identification of the mammalian homologues of the *Drosophila* timeless gene, Timeless1, *FEBS Lett.* **441**, 427–431.

Kondo, T., Mori, T., Lebedeva, N.V., Aoki, S., Ishiura, M., Golden, S.S. (1997) Circadian rhythms in rapidly dividing cyanobacteria, *Science* **275**, 224–227.

Kondo, T., Tsinoremas, N.F., Golden, S.S., Johnson, C.H., Kutsuna, S., Ishiura, M. (1994) Circadian clock mutants of cyanobacteria, *Science* **266**, 1233–1236.

Kondo, T., Strayer, C.A., Kulkarni, R.D., Taylor, W., Ishiura, M., Golden, S.S., Johnson, C.H. (1993) Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria, *Proc. Natl. Acad. Sci. U. S. A.* **90**, 5672–5676.

Konopka, R.J., Benzer, S. (1971) Clock mutants of *Drosophila* melanogaster, *Proc. Natl. Acad. Sci. U. S. A.* **68**, 2112–2116.

Kramer, A., Yang, F.C., Snodgrass, P., Li, X., Scammell, T.E., Davis, F.C., Weitz, C.J. (2001) Regulation of daily locomotor activity and sleep by hypothalamic EGF receptor signaling, *Science* **294**, 2511–2515.

Krishnan, B., Dryer, S.E., Hardin, P.E. (1999) Circadian rhythms in olfactory responses of *Drosophila* melanogaster, *Nature* **400**, 375–378.

Krishnan, B., Levine, J.D., Lynch, M.K., Dowse, H.B., Funes, P., Hall, J.C., Hardin, P.E., Dryer, S.E. (2001) A new role for cryptochrome in a *Drosophila* circadian oscillator, *Nature* **411**, 313–317.

Kume, K., Zylka, M.J., Sriram, S., Shearman, L.P., Weaver, D.R., Jin, X., Maywood, E.S., Hastings, M.H., Reppert, S.M. (1999) mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop, *Cell* **98**, 193–205.

Lakin-Thomas, P.L., Brody, S. (2000) Circadian rhythms in Neurospora crassa: lipid deficiencies restore robust rhythmicity to null frequency and white-collar mutants, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 256–261.

Lavery, D.J., Lopez-Molina, L., Margueron, R., Fleury-Olela, F., Conquet, F., Schibler, U., Bonfils, C. (1999) Circadian expression of the steroid 15 alpha-hydroxylase (Cyp2a4) and coumarin 7-hydroxylase (Cyp2a5) genes in mouse liver is regulated by the PAR leucine zipper transcription factor DBP, *Mol. Cell. Biol.* **19**, 6488–6499.

Lee, C., Etchegaray, J.P., Cagampang, F.R., Loudon, A.S., Reppert, S.M. (2001) Posttranslational mechanisms regulate the mammalian circadian clock, *Cell* **107**, 855–867.

Lee, K., Dunlap, J.C., Loros, J.J. (2003) Roles for WHITE COLLAR-1 in Circadian and General Photoperception in Neurospora crassa, *Genetics* **163**, 103–114.

Lee, K., Loros, J.J., Dunlap, J.C. (2000) Interconnected feedback loops in the Neurospora circadian system, *Science* **289**, 107–110.

Liu, Y. (2003) Molecular mechanisms of entrainment in the Neurospora circadian clock, *J. Biol. Rhythms.* **18**, 195–205.

Liu, Y., Loros, J., Dunlap, J.C. (2000) Phosphorylation of the Neurospora clock protein FREQUENCY determines its degradation rate and strongly influences the period length of the circadian clock, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 234–239.

Liu, Y., Tsinoremas, N.F., Johnson, C.H., Lebedeva, N.V., Golden, S.S., Ishiura, M., Kondo, T. (1995) Circadian orchestration of gene expression in cyanobacteria, *Genes. Dev.* **9**, 1469–1478.

Lockley, S.W., Skene, D.J., Arendt, J., Tabandeh, H., Bird, A.C., Defrance, R. (1997) Relationship between melatonin rhythms and visual loss in the blind, *J. Clin. Endocrinol. Metab.* **82**, 3763–3770.

Loros, J.J., Feldman, J.F. (1986) Loss of temperature compensation of circadian period length in the *frq*-9 mutant of *Neurospora crassa*, *J. Biol. Rhythms* **1**, 187–198.

Lowrey, P.L., Shimomura, K., Antoch, M.P., Yamazaki, S., Zemenides, P.D., Ralph, M.R., Menaker, M., Takahashi, J.S. (2000) Positional syntenic cloning and functional characterization of the mammalian circadian mutation tau, *Science* **288**, 483–492.

Lucas, R.J., Douglas, R.H., Foster, R.G. (2001) Characterization of an ocular photopigment capable of driving pupillary constriction in mice, *Nat. Neurosci.* **4**, 621–626.

Lucas, R.J., Freedman, M.S., Munoz, M., Garcia-Fernandez, J.M., Foster, R.G. (1999) Regulation of the mammalian pineal by non-rod, non-cone, ocular photoreceptors, *Science* **284**, 505–507.

Lucas, R.J., Hattar, S., Takao, M., Berson, D.M., Foster, R.G., Yau, K.W. (2003) Diminished pupillary light reflex at high irradiances in melanopsin-knockout mice, *Science* **299**, 245–247.

Martinek, S., Inonog, S., Manoukian, A.S., Young, M.W. (2001) A role for the segment polarity gene shaggy/GSK-3 in the *Drosophila* circadian clock, *Cell* **105**, 769–779.

Martinez-Garcia, J.F., Huq, E., Quail, P.H. (2000) Direct targeting of light signals to a promoter element-bound transcription factor, *Science* **288**, 859–863.

Maywood, E.S., O'Brien, J.A., Hastings, M.H. (2003) Expression of mCLOCK and other circadian clock-relevant proteins in the mouse suprachiasmatic nuclei, *J. Neuroendocrinol.* **15**, 329–334.

McDonald, M.J., Rosbash, M. (2001) Microarray analysis and organization of circadian gene expression in *Drosophila*, *Cell* **107**, 567–578.

McNamara, P., Seo, S.P., Rudic, R.D., Sehgal, A., Chakravarti, D., FitzGerald, G.A. (2001) Regulation of CLOCK and MOP4 by nuclear hormone receptors in the vasculature: a humoral mechanism to reset a peripheral clock, *Cell* **105**, 877–889.

McNeil, G.P., Zhang, X., Genova, G., Jackson, F.R. (1998) A molecular rhythm mediating circadian clock output in *Drosophila*, *Neuron* **20**, 297–303.

McWatters, H.G., Bastow, R.M., Hall, A., Millar, A.J. (2000) The ELF3 zeitnehmer regulates light signalling to the circadian clock, *Nature* **408**, 716–720.

Merrow, M., Brunner, M., Roenneberg, T. (1999) Assignment of circadian function for the Neurospora clock gene frequency, *Nature* **399**, 584–586.

Merrow, M., Franchi, L., Dragovic, Z., Gorl, M., Johnson, J., Brunner, M., Macino, G., Roenneberg, T. (2001) Circadian regulation of the light input pathway in Neurospora crassa, *EMBO J.* **20**, 307–315.

Millar, A.J., Carre, I.A., Strayer, C.A., Chua, N.H., Kay, S.A. (1995a) Circadian clock mutants in Arabidopsis identified by luciferase imaging, *Science* **267**, 1161–1163.

Millar, A.J., Straume, M., Chory, J., Chua, N.H., Kay, S.A. (1995b) The regulation of circadian period by phototransduction pathways in Arabidopsis, *Science* **267**, 1163–1166.

Miyamoto, Y., Sancar, A. (1998) Vitamin B_2-based blue-light photoreceptors in the retinohypothalamic tract as the photoactive pigments for setting the circadian clock in mammals, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6097–6102.

Mizoguchi, T., Wheatley, K., Hanzawa, Y., Wright, L., Mizoguchi, M., Song, H.R., Carre, I.A., Coupland, G. (2002) LHY and CCA1 are partially redundant genes required to maintain circadian rhythms in Arabidopsis, *Dev. Cell* **2**, 629–641.

Moore, R., Lenn, N. (1972) A retinohypothalamic projection in the rat, *J. Comp. Neurol.* **146**, 1–14.

Moore, R.Y., Eichler, V.B. (1972) Loss of a circadian adrenal corticosterone rhythm following suprachiasmatic lesions in the rat, *Brain Res.* **42**, 201–206.

Mori, T., Binder, B., Johnson, C.H. (1996) Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10183–10188.

Moutsaki, P., Whitmore, D., Bellingham, J., Sakamoto, K., David-Gray, Z.K., Foster, R.G. (2003) Teleost multiple tissue (*tmt*) opsin: a candidate photopigment regulating the peripheral clocks of zebrafish? *Brain Res. Mol. Brain Res.* **112**, 135–145.

Myers, M.P., Wager-Smith, K., Wesley, C.S., Young, M.W., Sehgal, A. (1995) Positional cloning and sequence analysis of the *Drosophila* clock gene, timeless, *Science* **270**, 805–808.

Nelson, D., Takahashi, J. (1991) Sensitivity and integration in a visual pathway for circadian entrainment in the hamster (*Mesocricetus auratus*), *J. Physiol.* **439**, 115–145.

Nelson, D.C., Lasswell, J., Rogg, L.E., Cohen, M.A., Bartel, B. (2000) FKF1, a clock-controlled gene that regulates the transition to flowering in Arabidopsis, *Cell* **101**, 331–340.

Newby, L.M., Jackson, F.R. (1993) A new biological rhythm mutant of *Drosophila* melanogaster that identifies a gene with an

essential embryonic function, *Genetics* **135**, 1077–1090.

Ni, M., Tepperman, J.M., Quail, P.H. (1999) Binding of phytochrome B to its nuclear signalling partner PIF3 is reversibly induced by light, *Nature* **400**, 781–784.

Nowrousian, M., Duffield, G.E., Loros, J., Dunlap, J.C. (2003) The *frequency* gene is required for temperature-dependent regulation of many *clock-controlled genes* in *Neurospora crassa*, *Genetics* **164**, 923–933.

Oishi, K., Sakamoto, K., Okada, T., Nagase, T., Ishida, N. (1998) Antiphase circadian expression between BMAL1 and period homologue mRNA in the suprachiasmatic nucleus and peripheral tissues of rats, *Biochem. Biophys. Res. Commun.* **253**, 199–203.

Okamura, H., Miyake, S., Sumi, Y., Yamaguchi, S., Yasui, A., Muijtjens, M., Hoeijmakers, J.H., van der Horst, G.T. (1999) Photic induction of mPer1 and mPer2 in cry-deficient mice lacking a biological clock, *Science* **286**, 2531–2534.

Ouyang, Y., Andersson, C.R., Kondo, T., Golden, S.S., Johnson, C.H. (1998) Resonating circadian clocks enhance fitness in cyanobacteria, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 8660–8664.

Panda, S., Sato, T.K., Castrucci, A.M., Rollag, M.D., DeGrip, W.J., Hogenesch, J.B., Provencio, I., Kay, S.A. (2002) Melanopsin (Opn4) requirement for normal light-induced circadian phase shifting, *Science* **298**, 2213–2216.

Panda, S., Provencio, I., Tu, D.C., Pires, S.S., Rollag, M.D., Castrucci, A.M., Pletcher, M.T., Sato, T.K., Wiltshire, T., Andahazy, M., Kay, S.A., Van Gelder, R.N., Hogenesch, J.B. (2003) Melanopsin is required for non-image-forming photic responses in blind mice, *Science* **301**(5632), 525–527. Epub 2003 Jun 26.

Pittendrigh, C.S., Caldarola, P.C., Cosbey, E.S. (1973) A differential effect of heavy water on temperature-dependent and temperature-compensated aspects of circadian system of *Drosophila* pseudoobscura, *Proc. Natl. Acad. Sci. U. S. A.* **70**, 2037–2041.

Pittendrigh, C.S., Daan, S. (1976) A functional analysis of circadian pacemakers in nocturnal rodents: IV Entrainment: Pacemaker as clock, *J. Comp. Physiol. A.* **106**, 333–355.

Pittendrigh, C.S., Bruce, V.G., Rosenweig, N.S., Rubin, M.L. (1959) A biological clock in *Neurospora*, *Nature* **184**, 169–170.

Plautz, J., Kaneko, M., Hall, J., Kay, S. (1997) Independent photoreceptive circadian clocks throughout *Drosophila*, *Science* **278**, 1632–1635.

Preitner, N., Damiola, F., Lopez-Molina, L., Zakany, J., Duboule, D., Albrecht, U., Schibler, U. (2002) The orphan nuclear receptor REV-ERBalpha controls circadian transcription within the positive limb of the mammalian circadian oscillator, *Cell* **110**, 251–260.

Price, J.L., Blau, J., Rothenfluh, A., Abodeely, M., Kloss, B., Young, M.W. (1998) double-time is a novel *Drosophila* clock gene that regulates PERIOD protein accumulation, *Cell* **94**, 83–95.

Ralph, M.R., Menaker, M. (1988) A mutation of the circadian system in golden hamsters, *Science* **241**, 1225–1227.

Ralph, M.R., Foster, R.G., Davis, F.C., Menaker, M. (1990) Transplanted suprachiasmatic nucleus determines circadian period, *Science* **247**, 975–978.

Ramsdale, M., Lakin-Thomas, P.L. (2000) sn-1,2-diacylglycerol levels in the fungus Neurospora crassa display circadian rhythmicity, *J. Biol. Chem.* **275**, 27541–27550.

Richter, C.P. (1967a) Sleep and activity: their relation to the 24-hour clock, *Res. Publ. Assoc. Res. Nerv. Ment. Dis.* **45**, 8–29.

Richter, C.P. (1971) Inborn nature of the rat's 24-hour clock, *J. Comp. Physiol. Psychol.* **75**, 1–4.

Roeder, P.E., Sargent, M.L., Brody, S. (1982) Circadian rhythms in Neurospora crassa: oscillations in fatty acids, *Biochemistry* **21**, 4909–4916.

Roenneberg, T., Merrow, M. (2002) Life before the clock: modeling circadian evolution, *J. Biol. Rhythms* **17**, 495–505.

Roenneberg, T., Merrow, M. (2003) The network of time: understanding the molecular circadian system, *Curr. Biol.* **13**, R198–R207.

Rosato, E., Codd, V., Mazzotta, G., Piccin, A., Zordan, M., Costa, R., Kyriacou, C.P. (2001) Light-dependent interaction between *Drosophila* CRY and the clock protein PER mediated by the carboxy terminus of CRY, *Curr. Biol.* **11**, 909–917.

Ruby, N.F., Brennan, T.J., Xie, X., Cao, V., Franken, P., Heller, H.C., O'Hara, B.F. (2002) Role of melanopsin in circadian responses to light, *Science* **298**, 2211–2213.

Rutila, J.E., Suri, V., Le, M., So, W.V., Rosbash, M., Hall, J.C. (1998) CYCLE is a second bHLH-PAS clock protein essential for circadian rhythmicity and transcription of *Drosophila* period and timeless, *Cell* **93**, 805–814.

Rutter, J., Reick, M., Wu, L.C., McKnight, S.L. (2001) Regulation of clock NPAS2 DNA binding by the redox state of NAD cofactors, *Science* **293**, 510–514.

Sangoram, A.M., Saez, L., Antoch, M.P., Gekakis, N., Staknis, D., Whiteley, A., Fruechte, E.M., Vitaterna, M.H., Shimomura, K., King, D.P., Young, M.W., Weitz, C.J., Takahashi, J.S. (1998) Mammalian circadian autoregulatory loop: a timeless ortholog and mPer1 interact and negatively regulate CLOCK-BMAL1-induced transcription, *Neuron* **21**, 1101–1113.

Sato, T.K., Panda, S., Kay, S.A., Hogenesch, J.B. (2003) DNA arrays: applications and implications for circadian biology, *J. Biol. Rhythms* **18**, 96–105.

Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M., Wisman, E. (2001) Microarray analysis of diurnal and circadian-regulated genes in Arabidopsis, *Plant Cell* **13**, 113–123.

Schaffer, R., Ramsay, N., Samach, A., Corden, S., Putterill, J., Carre, I.A., Coupland, G. (1998) The late elongated hypocotyl mutation of Arabidopsis disrupts circadian rhythms and the photoperiodic control of flowering, *Cell* **93**, 1219–1229.

Schmitz, O., Katayama, M., Williams, S.B., Kondo, T., Golden, S.S. (2000) CikA, a bacteriophytochrome that resets the cyanobacterial circadian clock, *Science* **289**, 765–768.

Schwartz, W.J., Gainer, H. (1977) Suprachiasmatic nucleus: use of 14C-labeled deoxyglucose uptake as a functional marker, *Science* **117**, 1090–1091.

Sehgal, A., Price, J.L., Man, B., Young, M.W. (1994) Loss of circadian behavioral rhythms and per RNA oscillations in the *Drosophila* mutant timeless, *Science* **263**, 1603–1606.

Sekaran, S., Foster, R.G., Lucas, R.J., Hankins, M.W. (2003) Calcium imaging reveals a network of intrinsically light sensitive inner retinal neurones, *Curr. Biol.* **13**, 1290–1298.

Shearman, L.P., Zylka, M.J., Reppert, S.M., Weaver, D.R. (1999) Expression of basic helix-loop-helix/PAS genes in the mouse suprachiasmatic nucleus, *Neuroscience* **89**, 387–397.

Shearman, L.P., Jin, X., Lee, C., Reppert, S.M., Weaver, D.R. (2000a) Targeted disruption of the mPer3 gene: subtle effects on circadian clock function, *Mol. Cell. Biol.* **20**, 6269–6275.

Shearman, L.P., Zylka, M.J., Weaver, D.R., Kolakowski, L.F., Reppert, S.M. (1997) Two period homologs: Circadian expression and photic regulation in the suprachiasmatic nuclei, *Neuron* **19**, 1261–1269.

Shearman, L.P., Sriram, S., Weaver, D.R., Maywood, E.S., Chaves, I., Zheng, B., Kume, K., Lee, C.C., van der Horst, G.T., Hastings, M.H., Reppert, S.M. (2000b) Interacting molecular loops in the mammalian circadian clock, *Science* **288**, 1013–1019.

Shibato, J., Agrawal, G.K., Kato, H., Asayama, M., Shirai, M. (2002) The 5'-upstream cis-acting sequences of a cyanobacterial psbA gene: analysis of their roles in basal, light-dependent and circadian transcription, *Mol. Genet. Genomics* **267**, 684–694.

Shigeyoshi, Y., Taguchi, K., Yamamoto, S., Takekida, S., Yan, L., Tei, H., Moriya, T., Shibata, S., Loros, J.J., Dunlap, J.C., Okamura, H. (1997) Light-induced resetting of a mammalian circadian clock is associated with rapid induction of the mPer1 transcript, *Cell* **91**, 1043–1053.

Shimomura, K., Low-Zeddies, S.S., King, D.P., Steeves, T.D., Whiteley, A., Kushla, J., Zemenides, P.D., Lin, A., Vitaterna, M.H., Churchill, G.A., Takahashi, J.S. (2001) Genome-wide epistatic interaction analysis reveals complex genetic determinants of circadian behavior in mice, *Genome Res.* **11**, 959–980.

Silver, R., Lesauter, J., Tresco, P.A., Lehman, M.N. (1996) A diffusable coupling signal form the transplanted suprachiasmatic nucleus controlling circadian locomotor rhythms, *Nature* **382**, 810–813.

Siwicki, K.K., Schwartz, W.J., Hall, J.C. (1992) An antibody to the *Drosophila* period protein labels antigens in the suprachiasmatic nucleus of the rat, *J. Neurogenet.* **8**, 33–42.

Siwicki, K.K., Eastman, C., Petersen, G., Rosbash, M., Hall, J.C. (1988) Antibodies to the period gene product of *Drosophila* reveal diverse tissue distribution and rhythmic changes in the visual system, *Neuron* **1**, 141–150.

Somers, D.E., Devlin, P.F., Kay, S.A. (1998a) Phytochromes and cryptochromes in the

entrainment of the Arabidopsis circadian clock, *Science* **282**, 1488–1490.

Somers, D.E., Schultz, T.F., Milnamow, M., Kay, S.A. (2000) ZEITLUPE encodes a novel clock-associated PAS protein from *Arabidopsis*, *Cell* **101**, 319–329.

Somers, D.E., Webb, A.A., Pearson, M., Kay, S.A. (1998b) The short-period mutant, toc1-1, alters circadian clock regulation of multiple outputs throughout development in Arabidopsis thaliana, *Development* **125**, 485–494.

Staiger, D., Allenbach, L., Salathia, N., Fiechter, V., Davis, S.J., Millar, A.J., Chory, J., Fankhauser, C. (2003) The Arabidopsis SRR1 gene mediates phyB signaling and is required for normal circadian clock function, *Genes Dev.* **17**, 256–268.

Stanewsky, R., Kaneko, M., Emery, P., Beretta, B., Wager-Smith, K., Kay, S.A., Rosbash, M., Hall, J.C. (1998) The cry^b mutation identifies cryptochrome as a circadian photoreceptor in *Drosophila*, *Cell* **95**, 681–692.

Stephan, F.K., Zucker, I. (1972a) Circadian rhythms in drinking behavior and locomotor activity of rats are eliminated by hypothalamic lesions, *Proc. Natl. Acad. Sci. U. S. A.* **69**, 1583–1586.

Stephan, F.K., Zucker, I. (1972b) Rat drinking rhythms: central visual pathways and endocrine factors mediating responsiveness to environmental illumination, *Physiol. Behav.* **8**, 315–326.

Stokkan, K.A., Yamazaki, S., Tei, H., Sakaki, Y., Menaker, M. (2001) Entrainment of the circadian clock in the liver by feeding, *Science* **291**, 490–493.

Strayer, C., Oyama, T., Schultz, T.F., Raman, R., Somers, D.E., Mas, P., Panda, S., Kreps, J.A., Kay, S.A. (2000) Cloning of the Arabidopsis clock gene TOC1, an autoregulatory response regulator homolog, *Science* **289**, 768–771.

Sun, Z.S., Albrecht, U., Zhuchenko, O., Bailey, J., Eichele, G., Lee, C.C. (1997) RIGUI, a putative mammalian ortholog of the *Drosophila* period gene, *Cell* **90**, 1003–1011.

Takumi, T., Matsubara, C., Shigeyoshi, Y., Taguchi, K., Yagita, K., Maebayashi, Y., Sakakida, Y., Okumura, K., Takashima, N., Okamura, H. (1998a) A new mammalian period gene predominantly expressed in the suprachiasmatic nucleus, *Genes Cells* **3**, 167–176.

Takumi, T., Nagamine, Y., Miyake, S., Matsubara, C., Taguchi, K., Takekida, S., Sakakida, Y., Nishikawa, K., Kishimoto, T., Niwa, S., Okumura, K., Okamura, H. (1999) A mammalian ortholog of *Drosophila* timeless, highly expressed in SCN and retina, forms a complex with mPER1, *Genes Cells* **4**, 67–75.

Takumi, T., Taguchi, K., Miyake, S., Sakakida, Y., Takashima, N., Matsubara, C., Maebayashi, Y., Okumura, K., Takekida, S., Yamamoto, S., Yagita, K., Yan, L., Young, M.W., Okamura, H. (1998b) A light-independent oscillatory gene mPer3 in mouse SCN and OVLT, *EMBO J.* **17**, 4753–4759.

Tei, H., Okamura, H., Shigeyoshi, Y., Fukuhara, C., Ozawa, R., Hirose, M., Sakaki, Y. (1997) Circadian oscillation of a mammalian homologue of the *Drosophila* period gene, *Nature* **389**, 512–516.

Thresher, R.J., Vitaterna, M.H., Miyamoto, Y., Kazantsev, A., Hsu, D.S., Petit, C., Selby, C.P., Dawut, L., Smithies, O., Takahashi, J.S., Sancar, A. (1998) Role of mouse cryptochrome blue-light photoreceptor in circadian photoresponses, *Science* **282**, 1490–1494.

Todo, T., Ryo, H., Yamamoto, K., Toh, H., Inui, T., Ayaki, H., Nomura, T., Ikenaga, M. (1996) Similarity among the *Drosophila* (6–4) photolyase, a human photolyase homolog, and the DNA photolyase-blue-light photoreceptor family, *Science* **272**, 109–112.

Toh, K.L., Jones, C.R., He, Y., Eide, E.J., Hinz, W.A., Virshup, D.M., Ptacek, L.J., Fu, Y.H. (2001) An hPer2 phosphorylation site mutation in familial advanced sleep phase syndrome, *Science* **291**, 1040–1043.

Ueda, H.R., Chen, W., Adachi, A., Wakamatsu, H., Hayashi, S., Takasugi, T., Nagano, M., Nakahama, K., Suzuki, Y., Sugano, S., Iino, M., Shigeyoshi, Y., Hashimoto, S. (2002) A transcription factor response element for gene expression during circadian night, *Nature* **418**, 534–539.

van der Horst, G.T., Muijtjens, M., Kobayashi, K., Takano, R., Kanno, S., Takao, M., de Wit, J., Verkerk, A., Eker, A.P., van Leenen, D., Buijs, R., Bootsma, D., Hoeijmakers, J.H., Yasui, A. (1999) Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms, *Nature* **398**, 627–630.

Vielhaber, E., Eide, E., Rivers, A., Gao, Z.H., Virshup, D.M. (2000) Nuclear entry of the circadian regulator mPER1 is controlled by mammalian casein kinase I epsilon, *Mol. Cell. Biol.* **20**, 4888–4899.

Vitaterna, M.H., King, D.P., Chang, A., Kornhauser, J.M., Lowrey, P.L., McDonald, J.D., Dove, W.F., Pinto, L.H., Turek, F.W., Takahashi, J.S. (1994) Mutagenesis and mapping of a mouse gene, clock, essential for circadian behavior, *Science* **264**, 719–725.

Vitaterna, M.H., Selby, C.P., Todo, T., Niwa, H., Thompson, C., Fruechte, E.M., Hitomi, K., Thresher, R.J., Ishikawa, T., Miyazaki, J., Takahashi, J.S., Sancar, A. (1999) Differential regulation of mammalian period genes and circadian rhythmicity by cryptochromes 1 and 2, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 12114–12119.

Wang, Z.Y., Tobin, E.M. (1998) Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression, *Cell* **93**, 1207–1217.

Welsh, D.K., Logothetis, D.E., Meister, M., Reppert, S.M. (1995) Individual neurons dissociated from rat suprachiasmatic nucleus express independently phased circadian firing rhythms, *Neuron* **14**, 697–706.

Whitmore, D., Foulkes, N.S., Sassone-Corsi, P. (2000) Light acts directly on organs and cells in culture to set the vertebrate circadian clock, *Nature* **404**, 87–91.

Whitmore, D., Foulkes, N.S., Strahle, U., Sassone-Corsi, P. (1998) Zebrafish Clock rhythmic expression reveals independent peripheral circadian oscillators, *Nat. Neurosci.* **1**, 701–707.

Wright, K.P. Jr., Czeisler, C.A. (2002) Absence of circadian phase resetting in response to bright light behind the knees, *Science* **297**, 571.

Yagita, K., Tamanini, F., Yasuda, M., Hoeijmakers, J.H., van der Horst, G.T., Okamura, H. (2002) Nucleocytoplasmic shuttling and mCRY-dependent inhibition of ubiquitylation of the mPER2 clock protein, *EMBO J.* **21**, 1301–1314.

Yamaguchi, R., Nakamura, M., Mochizuki, N., Kay, S.A., Nagatani, A. (1999) Light-dependent translocation of a phytochrome B-GFP fusion protein to the nucleus in transgenic Arabidopsis, *J. Cell Biol.* **145**, 437–445.

Yu, W., Nomura, M., Ikeda, M. (2002) Interactivating feedback loops within the mammalian clock: BMAL1 is negatively autoregulated and upregulated by CRY1, CRY2, and PER2, *Biochem. Biophys. Res. Commun.* **290**, 933–941.

Zeng, H., Hardin, P.E., Rosbash, M. (1994) Constitutive overexpression of the *Drosophila* period protein inhibits period mRNA cycling, *EMBO J.* **13**, 3590–3598.

Zhang, Q., Piston, D.W., Goodman, R.H. (2002) Regulation of corepressor function by nuclear NADH, *Science* **295**, 1895–1897.

Zheng, B., Larkin, D.W., Albrecht, U., Sun, Z.S., Sage, M., Eichele, G., Lee, C.C., Bradley, A. (1999) The mPer2 gene encodes a functional component of the mammalian circadian clock, *Nature* **400**, 169–173.

Zheng, B., Albrecht, U., Kaasik, K., Sage, M., Lu, W., Vaishnav, S., Li, Q., Sun, Z.S., Eichele, G., Bradley, A., Lee, C.C. (2001) Nonredundant roles of the mPer1 and mPer2 genes in the mammalian circadian clock, *Cell* **105**, 683–694.

Zhou, H., Watts, J., Aebersold, R. (2001) A systematic approach to the analysis of protein phosphorylation, *Nat. Biotechnol.* **19**, 375–378.

Zimmerman, W.F., Goldsmith, T.H. (1971) Photosensitivity of the circadian rhythm and of visual receptors in carotenoid-depleted *Drosophila* , *Science* **171**, 1167–1169.

Zylka, M.J., Shearman, L.P., Weaver, D.R., Reppert, S.M. (1998b) Three period homologs in mammals: differential light responses in the suprachiasmatic circadian clock and oscillating transcripts outside of brain, *Neuron* **20**, 1103–1110.

Zylka, M.J., Shearman, L.P., Levine, J.D., Jin, X., Weaver, D.R., Reppert, S.M. (1998a) Molecular analysis of mammalian timeless, *Neuron* **21**, 1115–1122.

Ribosome, High Resolution Structure and Function

Christiane Schaffitzel and Nenad Ban
Institute for Molecular Biology and Biophysics, Swiss Federal Institute of Technology, Zürich, Switzerland

1 A Macroscopic View of Peptide Synthesis 486

2 The 30S Ribosome Structure 489
2.1 The Path of the mRNA Through the 30S Subunit 489
2.2 Interactions of the 30S Subunit with tRNA 491
2.3 The Decoding Center 491
2.4 Modes of Antibiotic Inhibition 495
2.5 Translation Initiation – Interactions with Initiation Factors 495

3 The 50S Ribosome Structure 496
3.1 The Molecular Mechanism of Peptidyl Transfer 497
3.2 The Exit Tunnel 499
3.3 Antibiotics Targeting the Large Subunit 501
3.4 Regulatory Nascent Chains 502
3.5 The GTPase Factor-Binding Center 503
3.6 Structural Mimicry 503

4 The 70S Ribosome Structure 503
4.1 The 30S–50S Interface 504
4.2 Translocation – The Concerted Movement of tRNAs and mRNA 504
4.3 Movement of tRNAs through the Ribosome 506
4.4 Cotranslational Folding 506
4.5 Protein Secretion and Membrane Insertion 507

5 Perspectives 508

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

Bibliography 509
Books and reviews 509
Primary Literature 509

Keywords

Codon and Anticodon
The specific sequence of three nucleotides on the mRNA (codon) and the complementary sequence on the tRNA (anticodon).

Decoding
The mechanism by which the ribosome finds the correct aminoacyl-tRNA based on codon–anticodon interaction.

Elongation Factors
Proteins involved in the elongation of the growing peptide chain.

Exit Tunnel
The tunnel in the large ribosomal subunit for the exit of the nascent peptide chain.

Initiation Factors
Proteins involved in the translation initiation steps, that is, the initial association of the two ribosomal subunits and the binding of the first tRNA.

Molecular Chaperones
Proteins that help other proteins to fold and/or prevent aggregation.

Nascent Peptide
The growing peptide still connected to the ribosome as peptidyl-tRNA.

Peptidyl Transferase Reaction
The elongation of the polypeptide chain by one amino acid.

Release Factors
Proteins required for the release of the nascent peptide when a stop codon is encountered on the mRNA.

Translocation
The movement of the mRNA and the tRNAs through the ribosome during translation.

Translocation Machinery
Accomplishes the transport of the nascent polypeptide through the membrane or its integration in the membrane.

Abbreviations

A, P, E sites:	aminoacyl, peptidyl, exit sites
CryoEM:	cryo-electron microscopy
EF-G:	elongation factor G
EF-Tu:	elongation factor Tu
GTP:	guanosine triphosphate
mRNA:	messenger RNA
rRNA:	ribosomal RNA
SRP:	signal recognition particle
tRNA:	transfer RNA

The code for the amino acid sequence of proteins is translated by the ribosome using messenger RNA as the template. Ribosomes are RNA-protein complexes built of two subunits. In prokaryotes, the ribosome consists of a large 50S and a small 30S ribosomal subunit. Both subunits together form the 70S ribosome, the prokaryotic translation machinery. Structures of both ribosomal subunits have been solved at near-atomic resolution. The large ribosomal subunits are from the halophilic archaebacterium *Haloarcula marismortui* and from the eubacterium *Deinococcus radiodurans*. The small subunit has been solved from the thermophilic eubacterium *Thermus thermophilus*. Structures of the ribosomal subunits were also determined in complex with various antibiotics, substrate analogs, and two translation initiation factors. Finally, the entire 70S ribosome from *T. thermophilus* with mRNA and tRNAs is available at 5.5 Å as a molecular model. This wealth of structural data allows for a detailed understanding of the molecular mechanism of peptide synthesis by the ribosome.

The two most important steps in protein synthesis are the recognition of the cognate aminoacyl-tRNA and the peptidyl transfer reaction. These two functions are allotted to the two subunits of the ribosome. The small subunit mediates the interactions between tRNAs and the mRNA and selects for the correct tRNA in the decoding center. The large subunit comprises the peptidyl transferase center and provides the exit tunnel for the growing nascent polypeptide chain. The crystal structures of the separated subunits reveal many mechanistic details of the peptidyl transfer and decoding. The structures show unequivocally that mainly ribosomal RNA is present in the peptidyl transferase center as well as in the decoding center, indicating that it is in fact ribosomal RNA that is responsible for all aspects of peptide synthesis. In other words, the ribosome is a ribozyme.

Here, we analyze the crystal structures of prokaryotic 50S and 30S subunits and the intact 70S ribosome with respect to their function and describe how our understanding of the mechanism of translation has been influenced by these recent structures. Furthermore, we outline the impact of cryo-electron microscopic studies (cryoEM) on our knowledge of ribosome conformational flexibility and ribosome complexes with translation factors.

1
A Macroscopic View of Peptide Synthesis

Aminoacyl-tRNAs are the substrates for protein synthesis. They all have an L-shaped structure with the anticodon loop at the end of the long L-arm. The anticodon sequence is part of the anticodon loop and complementary to the cognate mRNA base triplet (codon). At the 3′ end of the short L-arm, the acceptor stem has a terminal CCA sequence to which individual amino acids are attached by aminoacyl-tRNA synthetases. In the cell, a specific aminoacyl-tRNA synthetase exists for each amino acid, catalyzing the formation of an ester bond between the carboxyl group of the amino acid and the 2′ or 3′ hydroxyl group of the 3′ adenosine of tRNA with the appropriate anticodon.

Both ribosomal subunits have three binding sites for tRNA, called the *A* (*aminoacyl*), *P* (*peptidyl*) *and E* (*exit*) *sites*. During protein synthesis, the tRNA anticodons interact with the mRNA and the small ribosomal subunit, while the respective aminoacyl-acceptor stems extend into the large subunit where the peptidyl transfer reaction occurs (Fig. 1). In the *decoding center* of the small ribosomal subunit,

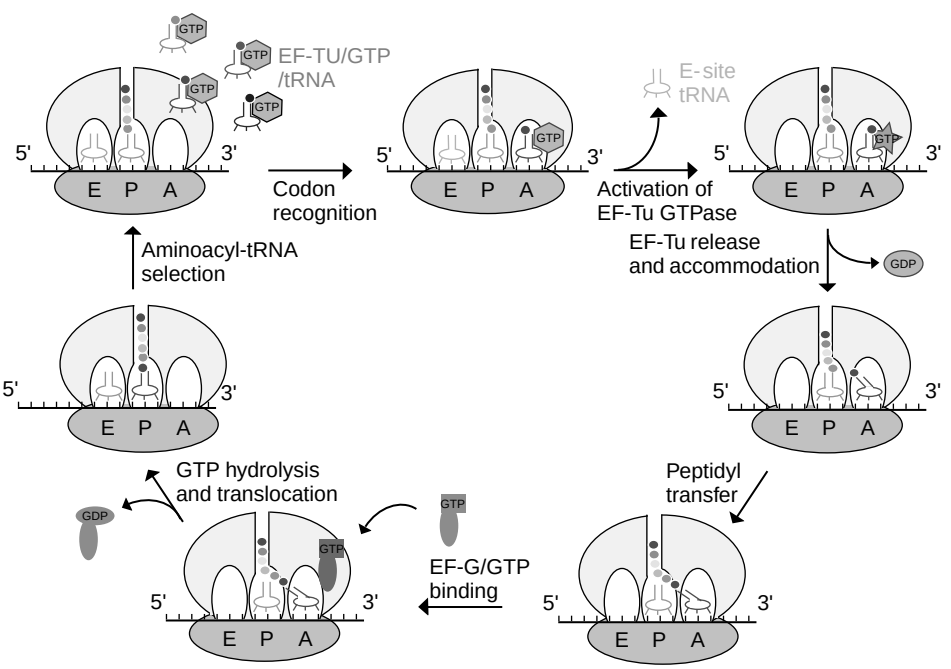

Fig. 1 Overview of the translation cycle. The cognate aminoacyl-tRNA is selected in the A site of the small subunit. The tRNA binds as a ternary complex together with EF-Tu and GTP. Binding of the cognate aminoacyl-tRNA induces GTPase activity of EF-Tu. The release of EF-Tu and GDP is followed by a conformational change of the aminoacyl-tRNA such that it binds in the A site of the peptidyl transferase center. The elongating chain is then transferred from the P-site peptidyl-tRNA to the amino group of the A-site aminoacyl-tRNA. The peptide chain leaves the peptidyl transferase center through the ribosomal tunnel, which spans the large subunit. EF-G catalyzes GTP hydrolysis and thereby drives translocation, that is, transport of the mRNA, the deacylated tRNA from the P to the E site and the peptidyl-tRNA from the A to the P site.

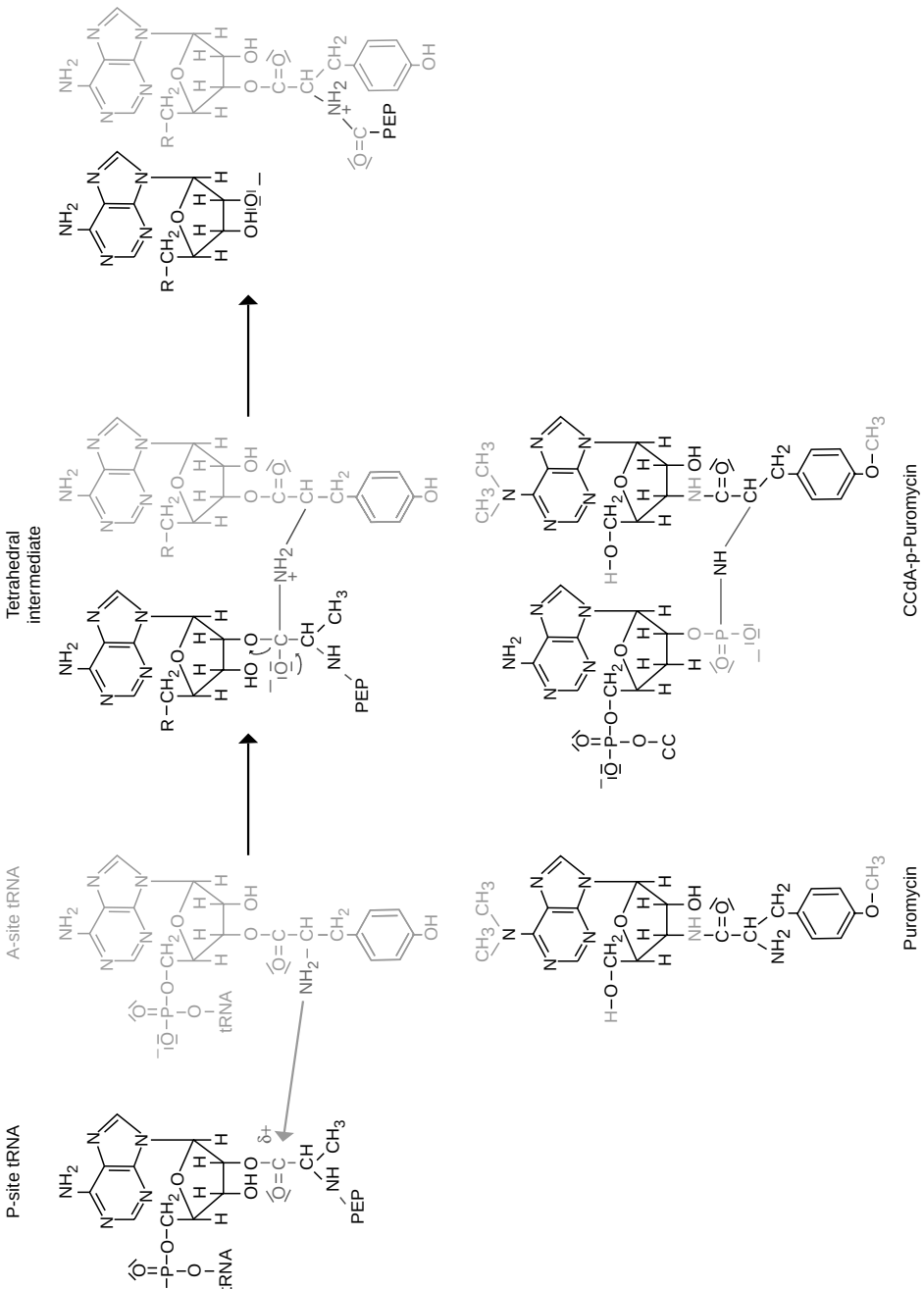

Fig. 2 (See the legend on next page)

selection of the correct aminoacyl-tRNA occurs based on the interaction of the mRNA base triplet with the anticodon triplet of the aminoacyl-tRNA. Aminoacyl-tRNAs bind to the ribosome in complex with elongation factor Tu (EF-Tu) and GTP. These ternary complexes bind tightly to the ribosome only if the anticodon matches the mRNA codon (*initial selection*). Then, GTP is hydrolyzed, resulting in release of phosphate and EF-Tu in complex with GDP (Fig. 1). Concomitantly, a large conformational change of the A site aminoacyl-tRNA occurs, referred to as *accommodation*. This positions the 3′ end of the aminoacyl-tRNA in the *peptidyl transferase center* adjacent to the P site–bound peptidyl-tRNA. Subsequently, the α-amino group of the aminoacyl-tRNA bound in the A site attacks the ester group of the peptidyl- (or N-formyl methionine initiator) tRNA in the P site at its carbonyl atom (Fig. 2). This nucleophilic attack results in a tetrahedral intermediate. A new peptide bond is formed by displacement of the P-site tRNA hydroxyl group from this intermediate (Fig. 2). As a result, the tRNA in the A site carries the nascent peptide chain, which is extended by one amino acid. The synthesized nascent peptide chain diffuses from the peptidyl transferase center into the *tunnel* and exits at the opposite side of the large ribosomal subunit (Fig. 1).

Before the next cycle of peptidyl transfer can occur, the peptidyl-tRNA has to be transported from the ribosomal A site to the P site. Furthermore, the deacylated tRNA in the P site is moved to the E site of the ribosome. This complex concerted movement of tRNAs and mRNA through the ribosome is termed *translocation*. It occurs with high accuracy and involves both subunits. The energy for translocation is provided by GTP hydrolysis catalyzed by elongation factor G (EF-G, Fig. 1). The tRNA bound in the E site interacts still with both ribosomal subunits; it is only released when the next aminoacyl-tRNA is accommodated in the ribosomal A site. A cartoonlike overview of the catalytic centers of both ribosomal subunits is given in Fig. 3.

Fig. 2 Mechanism of peptide bond formation catalyzed in the peptidyl transferase center of the large subunit. (a) The ribosomal peptidyl transferase reaction. The substrates of the peptidyl-transferase reaction are aminoacyl-tRNA and peptidyl-tRNA binding to the ribosomal A site and P site respectively. The α-amino group of the aminoacyl-tRNA in the A site of the ribosome attacks the carbonyl carbon atom of the ester bond that links the nascent polypeptides to P site–bound tRNAs. Thereby, a tetrahedral intermediate is formed at the carbonyl group with a chiral carbon atom (D) and a negatively charged oxygen. The 2′ (or 3′) OH group of the tRNA 3′-adenosine is the leaving group of the tetrahedral intermediate, and a new peptide bond (amide bond) is generated, resulting in a peptidyl-tRNA in the A site extended by one amino acid and a deacylated tRNA in the P site.
(b) Puromycin resembles the 3′ terminus of tyrosyl-tRNA, and it binds to the A site of the ribosome without the need for any elongation factors. Subsequently, the peptide is transferred to the puromycin and the nascent chain is released. (c) The Yarus inhibitor (CCdA-p-Puromycin) is formed by coupling the 3′ OH group of CCdA (in the P site) to the amino group of puromycin (in the A site) via a phosphoramide group, which mimics the tetrahedral carbon intermediate. The inhibitor binds tightly to the ribosome and inhibits peptidyl transferase activity. Important differences between the CCdA-p-Puromycin and the tetrahedral carbon intermediate formed during translation are the nonchiral phosphorus atom, the delocalized negative charge, and the missing 2′ OH group of the adenine in the P site.

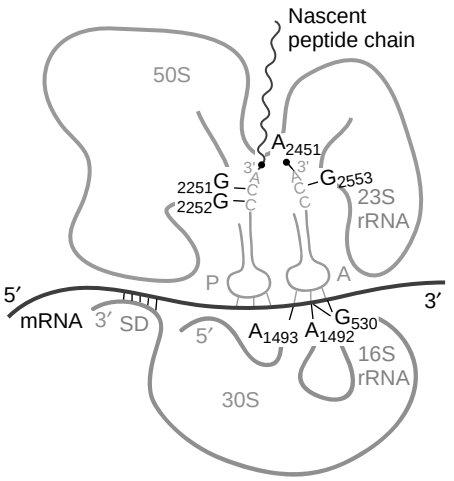

Fig. 3 Schematic overview of the ribosomal active sites. The Shine-Dalgarno (SD) sequence of the mRNA forms a helix with the complementary 3′ terminal tail of 16S rRNA. The anticodons of the A- site and P- site tRNA pair with the mRNA. In the A site decoding center of the small subunit, the codon–anticodon interaction is monitored by 16S rRNA via tertiary interactions. The tRNA acceptor stems point into the peptidyl transferase center of the large subunit. The 3′ CCA ends of the acceptor stems base-pair with 23S rRNA A-site and P-site RNA loops. The adenine of the central loop of 23S rRNA is suggested to play a role in peptidyl transfer catalysis. The nascent peptide chain is extending into the exit tunnel of the large subunit.

2
The 30S Ribosome Structure

The small subunit of the ribosome is responsible for binding mRNA and selection of the correct aminoacyl-tRNA. The structure of the small ribosomal subunit from a thermophilic bacterium *T. thermophilus* was recently determined. On the basis of electron microscopic studies, the 30S subunit is classically divided into head, body, neck, shoulder, and platform. It consists of one 16S rRNA chain and 20 ribosomal proteins (S1–S20) (Fig. 4). The shape of the small subunit is mostly determined by ribosomal RNA. The proteins are distributed over the top, sides, and back of the 30S subunit, while interface with the 50S subunit is mostly formed by rRNA (Fig. 4). Only protein S12 is located near the decoding site, whereas other proteins are found at the periphery of the subunit interface allowing them to contact the 50S subunit. Besides a globular domain many ribosomal proteins have extensions or tails that penetrate far into the rRNA core and closely interact with it, thereby stabilizing the RNA domains. The mRNA binding site and the 3′ end of 16S rRNA are situated on the platform. The tRNA binding sites are located in a cleft formed by the platform and the head (Fig. 4).

2.1
The Path of the mRNA Through the 30S Subunit

The exact path of mRNA through the ribosome was determined by X-ray crystallography of 70S ribosomes at 7 Å resolution. About 30 nucleotides of mRNA are shown to be in contact with the ribosome, predominantly with 16S rRNA. The mRNA wraps around the neck of the 30S subunit, starting with its 5′ end between head and platform and exiting on the opposite site between head and shoulder. The Shine-Dalgarno sequence, that is, the ribosome binding side provides for proper reading of the mRNA by forming a helix with the complementary 3′ terminal tail of 16S rRNA. The Shine-Dalgarno helix is located in the 30S subunit cleft between the head

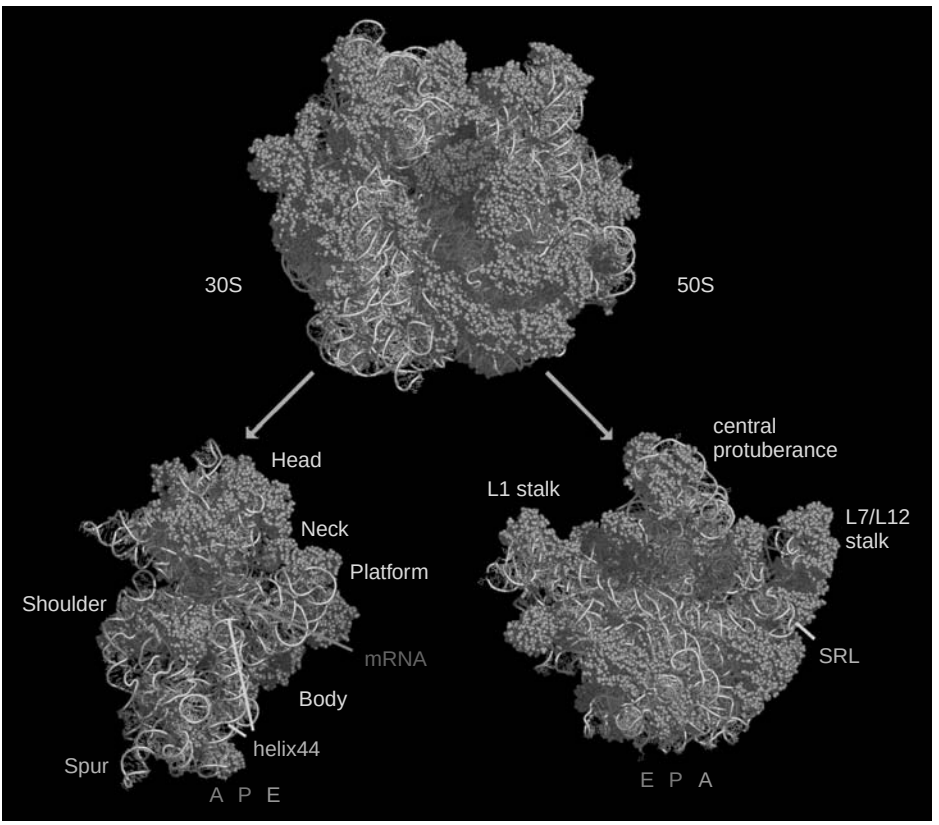

Fig. 4 Complete 70S ribosome is modeled by docking high-resolution 50S and 30S structures onto the low-resolution 70S structure. The 30S structure is from *T. thermophilus* and 50S structure from *H. marismortui* with tRNA positions determined from lower resolution 70S studies. Ribosomal RNA is depicted in gray and proteins in yellow. The 70S is shown from the side with the 30S subunit on the left and the 50S subunit on the right. In the subunit interface cavity, the tRNAs are visible. The A-site tRNA is depicted in blue, the P-site tRNA in magenta and the E-site tRNA in green. The anticodon arms of the tRNAs point to the 30S subunit, while the acceptor stems point into a large cleft in the large subunit. On the left-hand side, the 30S subunit is shown from the interface with the A-, P- and E-tRNA anticodon stem-loops. The anticodon loops of the A and P tRNAs contact the mRNA. Architectural characteristics and important features are labeled. On the right-hand side, the large subunit with tRNAs docked onto the *Haloarcula* 50S structure is shown from the interface as well. The acceptor stems of A- and P-site tRNAs point into the peptidyl transferase center. The sarcin–ricin loop (SRL) is a central part of the GTPase factor binding center. The three characteristic protuberances of the large subunit are labeled. (See color plate p. xxxviii).

and the back of the platform. From there, the mRNA enters a short tunnel, which is situated between head and platform. This region comprises the E site, which does not have full access to the interface. As a consequence, codon–anticodon interaction is not possible to the full extent for E site tRNA. Downstream of the E site, the mRNA reaches the interface by a sharp turn. On the interface, only about eight

nucleotides including the A- and P-site codons are presented to the tRNAs. There is a sharp kink in the mRNA between the codons in the A site and in the P site. This kink could play a role in maintaining the reading frame during translocation. A- and P-site codons occupy the major groove of the penultimate helix 44, which is the longest helix of the small subunit. It expands from the head to the bottom of the body and protrudes to interact with the large subunit (Fig. 4). Downstream of the A site responsible for decoding, the mRNA passes through a tunnel between head and shoulder of the subunit. Subsequently, it reaches a ring formed by proteins S3, S4, and S5 that functions probably as a helicase to remove mRNA secondary structure before entering the decoding region. During translation initiation, binding of mRNA to the small subunit requires an opening and closing of at least one of the 30S subunit tunnels at the 3′ or 5′ site of the mRNA.

2.2 Interactions of the 30S Subunit with tRNA

The three tRNAs are aligned on the 30S subunit with the anticodon loops bound in the 16S rRNA groove between the head, body, and platform (Fig. 4). In the *T. thermophilus* 30S ribosome structure, the interaction of mRNA and tRNA in the P site is mimicked through fortuitous interactions between neighboring subunits in the crystal. The 3′ end of 16S rRNA, which is complementary to the Shine-Dalgarno sequence, folds back and imitates mRNA in the P site while a stem-loop structure (the spur, Fig. 4) of a neighboring molecule in the crystal mimics the anticodon loop of the tRNA. As mentioned above, the P-site codon of the mRNA binds to the major groove of a well-conserved region of helix 44. The anticodon stem-loop of the tRNA forms several hydrogen bonds to 16S rRNA. Furthermore, the ribosomal proteins S13 and S9 interact with the anticodon loop of P-site tRNA.

The *E site*, unlike the A and P sites, consists mostly of ribosomal proteins as visualized in the *T. thermophilus* 70S structure. The E site tRNA is tightly bound and interacts with 16S rRNA as well as with a number of ribosomal proteins, mainly S7 and S11. Codon–anticodon interactions are not observed in the E site.

Recently, 30S subunit structures have been solved in complex with mRNA and cognate as well as near-cognate tRNA anticodon stem-loops in the *A site*, revealing important aspects of the decoding mechanism. The A site is much shallower and wider than the P and E sites, consistent with lower affinity for tRNA. It is mainly composed of 16S rRNA with a small contribution by ribosomal proteins S13 and S12, which contact the mRNA. Importantly, parts of the 16S rRNA, that is, the 3′ major domain and helix 44 (the penultimate helix), are responsible for the functioning of the decoding center.

2.3 The Decoding Center

The error rate of translation is 10^{-3} to 10^{-4}. Thus, the ribosome must be able to discriminate cognate from near and noncognate tRNAs by screening all three codon positions for correct interactions. This high translation fidelity cannot be explained by the difference in free energy between correct and incorrect base pairing of the codon–anticodon helix, since the free energy of some mismatches is very similar to the base pairing of cognate tRNAs. For instance, the gain of free energy of GU base pairing is very similar to the gain of AU pairing. In addition, AU

base pairing is energetically less favorable than GC pairing, that is, the gain of free energy is not the same for all cognate codon–anticodon interactions.

Moreover, the difference in free energy does not explain why the recognition of the third base pair is sometimes less stringently recognized than the preceding two. Consequently, a *wobble* can be observed in the pairing of the third base of the codon, giving rise to part of the degeneracy of the genetic code. Therefore, some tRNAs can recognize two or three codons with differences in the third base. It is clear that the *difference in free energy* of correct and incorrect base pairing may well contribute to the accuracy of translation, but by itself it cannot provide a satisfactory explanation.

As a second source of selectivity, the ribosome is capable of recognizing deviations from the Watson-Crick *base-pair geometry*, that is, the suboptimal shape of the sugar-phosphate backbone or the minor and major grooves in mismatched pairs. This sensitivity of the ribosome with regards to base-pair geometry is in analogy to enzyme *substrate recognition*.

The complex between 30S subunit, mRNA and the tRNA anticodon stem-loop switches from the open to the closed conformation if the codon–anticodon interaction is cognate. This conformational change involves flipping out of two adenines, A1492 and A1493, of helix 44. Theses adenines then tightly pack in the minor groove of the first two base pairs of the codon–anticodon helix. Both adenines form hydrogen bonds with the 2′ ribose hydroxyl groups of the codon–anticodon base pairs: a motif termed "A-minor" (Fig. 5). These types of tertiary interactions involving the two adenines are not possible if the codon–anticodon base pairing is noncognate, that is, not a Watson-Crick pair. Similarly, DNA polymerase and RNA polymerase use specific arginine and glutamine residues in their active center, which recognize the shape of Watson-Crick base pairs via the minor groove. A third residue, guanine530, is positioned on a loop that belongs to the 30S subunit shoulder and interacts with the second and the wobble-position base pair (Fig. 5). Consistent with the wobble base pairing in this position, the interaction of G530 with the third base pair is less sensitive to mismatches than the interaction described above involving the two adenines. The structure of 30S subunit, mRNA, and a cognate anticodon stem-loop shows that upon binding of the cognate tRNA, G530 flips from a *syn* to an *anti* conformation (Fig. 6). Near-cognate tRNA (with a single mismatch) does not cause flipping out of the adenines or G530. Consistent with their important role in decoding, A1492, A1493, and G530 are universally conserved and essential for cell viability.

In summary, the ribosome precisely monitors the minor groove geometry of the codon–anticodon interaction at the first two positions in the A site. Therefore, codons that differ in the first or second base but encode the same amino acid must pair with different tRNAs. The third position allows mismatches, which explains the frequent appearance of inosine in the anticodon at this position. Inosine can pair with cytosine, adenine, and uridine, thus maximizing the number of codons recognized by the specific tRNA molecules.

The close contacts of the cognate codon–anticodon pair with A1492, A1493, and G530 induce a characteristic closing of the 30S subunit around the A site. This locking up involves conformational changes of the overall 30S architecture. In particular, these are rotations of the

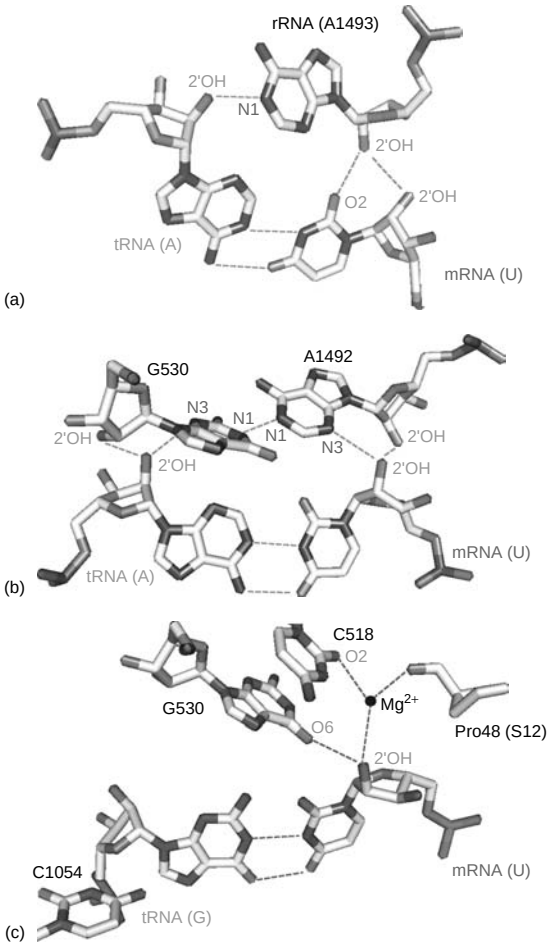

Fig. 5 Monitoring of the codon–anticodon interactions in the decoding center. (a), (b), and (c) show the minor groove recognition at the first (a), second (b) and third base position (c) respectively. (a) Adenine1493 binds into the minor groove of the first AU base pair. It forms hydrogen bonds to the 2′OH group of the tRNA with its N1 atom and to the 2′ OH group of the mRNA base with its 2′OH group. An additional hydrogen bond is formed between the 2′OH group of A1493 and the O2 group of uracil in the mRNA. (b) In the second position, A1492 and G530 both bind into the minor groove of the codon–anticodon helix and form hydrogen bonds with their N1 atoms. A1492 interacts with the 2′OH group of the mRNA via its N3 and 2′OH group. G530 also forms two hydrogen bonds with its 2′OH and N3 and the 2′OH group of the tRNA respectively. (c) The third base pair is monitored less stringently (wobble position). O6 of G530 forms a hydrogen bond to the 2′OH group of the mRNA. A second, magnesium ion-mediated interaction occurs between the 2′OH group of the mRNA and O2 of C518, as well as the main chain carbonyl of proline48 of protein S12. The GU mismatch is not selected against at this third position.

head toward the shoulder and the subunit interface, as well as movement of the shoulder toward the interface and the platform. Particularly well characterized is the "helix 27 accuracy switch" during the transition to the closed conformation. In this case, the packing of helix 27 to helix 44 changes because of a disrupted RNA turn. Thus, the structure of the decoding center is altered. Noncognate and near-cognate tRNA binds to the open conformation and cannot induce closing of the 30S subunit. In the closed conformation, the aminoacyl-tRNA cannot leave the A site. In conclusion, *induced fit* is a further source of accuracy, in addition to the A1492-, A1493-, and G530-based proofreading mechanism (substrate recognition) and free energy gains of codon–anticodon interaction.

During *initial selection,* the specific aminoacyl-tRNA is still bound by EF-Tu, and the acceptor stem of the aminoacyl-tRNA is not in contact with the peptidyl transferase center of the large ribosomal subunit. In cryoEM studies of the ternary complex bound to the 70S ribosome stalled with kirromycin resembling the

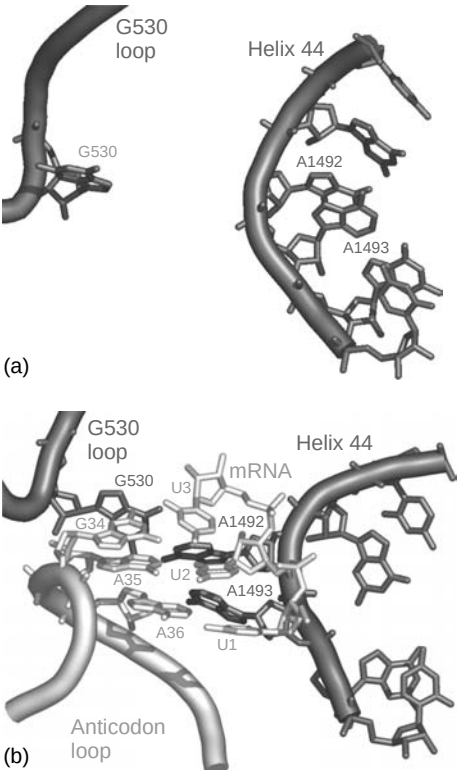

Fig. 6 (a) The empty decoding center of the native 30S subunit. The two 16S rRNA loops in the catalytic center are depicted in gray. The bases A1492 and A1493 are stacked in the helix 44 loop interior. G530 is in the *syn* conformation. (b) The decoding center in the presence of mRNA and the cognate tRNA anticodon loop. The mRNA and the tRNA anticodon loop are shown in white. A1492 and A1493 are now flipped out of helix 44 and pack tightly into the minor grove of the codon–anticodon helix monitoring the first two codon positions. G530 changes its conformation to *anti* and binds in the minor groove of the codon–anticodon helix monitoring the second and the third codon position.

GTPase activated state, tRNA adopts an unusual conformation that appears to facilitate codon–anticodon interaction. Cognate tRNA induces the closed conformation of the 30S subunit and triggers GTP hydrolysis by EF-Tu bound to the large subunit. Subsequently, the 3′CCA end of the aminoacyl-tRNA is released by the EF-Tu-GDP complex and can then rotate into the A site of the peptidyl transferase center. This *accommodation* process needs the full-size tRNA molecule, indicating that tRNA may play an active role in intersubunit signaling. Indeed, in the cryoEM studies of the kirromycin stalled ternary complex, the elbow region of the aminoacyl-tRNA could contact the GTPase associated center of the 50S subunit (while EF-Tu binds the sarcin–ricin loop) during initial selection and thus could induce GTP hydrolysis by a conformational change. However, since the data is at low resolution, it can also support a different model of GTPase activation triggered by conformational changes in both ribosomal subunits transmitted from the 30S subunit to the 50S subunit through intersubunit bridges. In fact, a significant movement of the two subunits relative to each other is observed during accommodation Accommodation can lead to dissociation of the aminoacyl-tRNA from the ribosome. This provides a *second proofreading step* during translation because the interactions between the ribosome and the codon–anticodon minor groove must be sufficiently strong to prevent the dissociation of the A site tRNA in the absence of EF-Tu. In summary, two selection stages

exist, which are separated by irreversible GTP hydrolysis.

Clearly, there is a sensitive equilibrium between the open and closed conformation of the 30S subunit during decoding, and stabilization of one or the other conformation seriously affects translation accuracy. The transition into the closed conformation reduces polar interactions present at the S4/S5 interface. Destabilizing mutations in S4 and S5 residing between the shoulder and platform facilitate the transition to the closed conformation. This results in a higher affinity for tRNA and thus in an *error-prone ribosome*. In the closed conformation, S12 can form new salt bridges to helices 27 and 44. Destabilizing mutations of the involved residues result in reduced affinity for tRNA. Thus, the open conformation is favored and the ribosome is in a *hyperaccurate state*.

2.4 Modes of Antibiotic Inhibition

Structures of antibiotics bound to both ribosomal subunits provided detailed information on antibiotic action and the mechanistic basis of antibiotic resistance. In addition, further insights into the translation process were derived from such analyses. About 20 structures of antibiotics bound to the ribosome are available to date. Most of the antibiotics examined bind to functionally important parts of the 30S subunit at or near the decoding center or at locations that are important for conformational changes during translocation. Here, we focus on two 30S subunit binding antibiotics, paromomycin and streptomycin, because the mechanism of their action can be explained with the obtained structures.

The binding of the error-inducing aminoglycoside antibiotic *paromomycin* has been structurally characterized by both, NMR using a fragment of 16S rRNA and by crystallography of the 30S subunit in complex with paromomycin, mRNA, and a near-cognate tRNA anticodon stem-loop. Paromomycin binds to helix 44 inducing a conformational change where crucial bases of the decoding center, A1492 and A1493, are flipped out of the helix and interact directly with the minor groove of the codon–anticodon helix. This is the same conformation otherwise observed upon binding of cognate tRNAs. Thus, paromomycin decreases ribosome's selectivity.

Streptomycin acts during the accommodation step by stabilization of the overall closed conformation of 30S subunit, leading to error-prone translation. Streptomycin binds tightly to helix 27, helix 44, the 530 loop, and protein S12. Resistance to streptomycin is conferred by mutations in S12 that destabilize the closed conformation of the 30S subunit, leading to a hyperaccurate state.

2.5 Translation Initiation – Interactions with Initiation Factors

The initiation factors IF1 and IF3 bind to the small ribosomal subunit. Their structures have been solved in complex with the 30S subunit. IF1 prevents tRNA binding to the A site on the 30S subunit. In the IF1–30S complex, IF1 binds to a site that includes the A site of the small subunit, that is, helix 44, the 530 loop, and protein S12. The ribosomal binding site of IF1 contains many basic residues whereas the solvent-exposed site is rich in negative charges. A loop of IF1 interacts with the minor groove of helix 44. The two adenine bases, A1492 and A1493, are flipped out of helix 44 upon IF1 binding and are buried in IF1. In addition, IF1

induces a conformational change of helix 44 and shifts the relative orientations of the 30S domains. The head, platform, and the shoulder rotate toward the A site, adopting a conformation that probably resembles an intermediate between subunit association and dissociation.

IF2 binds to the acceptor stem of the initiator tRNA, and it interacts with IF1. The IF2 GTPase domain presumably binds to the GTPase factor binding site on the 50S subunit, but direct location of IF2 has not been determined yet.

IF3 binds tightly to the 30S subunit and interferes with its association to the 50S subunit. In addition, it helps in selecting the *N*-formyl methionine-tRNA by excluding other aminoacyl-tRNAs from the P site. IF3 has two distinct domains, of which the C-terminal domain has been shown to be sufficient for 70S dissociation. The location of the C-terminal domain of IF3 in the X-ray structure of the 30S–IF3 complex is not the same as observed by cryoEM nor is it as derived from biochemical data obtained by hydroxyl radical cleavage. In the cryoEM structure, the N-terminal domain of IF3 binds to the tRNA binding cleft of the 30S subunit, whereas the C-terminal domain is found at the interface side of the platform and the neck of the 30S subunit, which is the E site. Thus, IF3 is spanning the interface region from the platform to the neck of the 30S subunit. At this position, the C-terminal domain IF3 would be adjacent to the P site and would directly interfere with subunit association. IF3 has been shown to bind to three base pairs of the anticodon stem-loop that are unique to initiator tRNA. Therefore, the C-terminal domain is probably responsible for recognizing and selecting the initiator tRNA without interfering with the codon–anticodon interaction in the P site. In contrast to biochemical and cryoEM data, IF3 was visualized at the upper end of the platform on the solvent-exposed side of the 30S subunit in the X-ray structure of the 30S–IF3 complex. In that case, the negative effect of IF3 on subunit association would have to be much more indirect, for instance, by influencing the conformational state of the 30S subunit.

In summary, IF1 binds to the A site on the small subunit, IF2 binds over the A site in complex with the initiator tRNA which occupies the P site, and IF3 binds to the ribosomal E site. Thus, in the initiation complex, all tRNA binding sites are occupied and IF2 and IF3 both select for the correct initiator tRNA.

3 The 50S Ribosome Structure

The large ribosomal subunit is mostly composed of 23S and 5S rRNA and contains approximately 31 to 35 proteins (L1–L35); the exact amount of protein is species dependent. The ribosome core is well conserved in all species and consists mainly of ribosomal RNA. The 50S structure shows three characteristic protuberances: the L1 stalk, the central protuberance, and the L7/L12 stalk (Fig. 4). The proteins are not distributed equally over the ribosome but cluster at the solvent-exposed surface. Their extensions play an important role in the stabilization of the RNA tertiary structure. Similar to the structure of the small subunit, proteins are largely absent from the functionally important sites, namely, the peptidyl transferase center and the interface. 23S rRNA forms a compact, single hemispherical domain, which is the body of the large subunit (Fig. 4). The protuberances consist of 23S rRNA, proteins L1, L7, L10, L11, and L12

and 5S rRNA in case of the central protuberance. The stalks are involved in bridges to the small subunit or interact with tRNAs or GTP-binding translation factors and thus are likely to be dynamic elements of the 50S subunit.

The first 50S X-ray structure obtained from the archaebacterium *H. marismortui* was resolved to 2.4 Å resolution (Fig. 4). The structure highlighted two paramount features of the large subunit, the peptidyl transferase center and the exit tunnel. It provided a detailed overview of the overall architecture and important information about RNA-protein interactions. Furthermore, water and metal ion positions as well as base modifications could be identified, all of which may contribute to the folding and stabilization of the ribosome. Subsequently, the *D. radiodurans* 50S structure was solved at 3.1 Å resolution. While being very similar to the archaeal large subunit, this structure provided additional information on some regions that were disordered in the *Haloarcula* 50S subunit. The disordered regions were proteins L1, L10, L11, L12, that is, components of the two lateral protuberances, the L11 RNA region, and some RNA stem-loops contacting the small ribosomal subunit. These regions are likely to be flexible in the isolated 50S subunit and could adopt a more defined conformation upon association with their binding partner. Stabilizing elements of the large ribosomal subunit are proteins, water, metal ions, and tertiary RNA interactions named A-minor motifs (as observed in the decoding center).

3.1
The Molecular Mechanism of Peptidyl Transfer

The peptidyl transferase center is located at the bottom of a large cleft in the center of the subunit interface (Fig. 4). Although there was mounting evidence that ribosomal proteins are not involved in peptidyl transfer, unambiguous demonstration that the peptidyl transferase center is entirely composed of 23S rRNA was provided with the *Haloarcula* 50S structure in complex with CCdA-p-Puromycin (the Yarus inhibitor). In fact, there is no ribosomal protein present in the proximity of the peptidyl transferase center within a distance of 18 Å. The catalytic site is formed by the highly conserved central loop (the peptidyl transferase loop) of domain V, which is located at the bottom of the cleft.

Once the cognate aminoacyl-tRNA enters the ribosomal A site of the large subunit, the peptidyl transfer occurs spontaneously, that is, without additional energy input. On the basis of the structure of the 50S subunit in complex with CCdA-p-Puromycin and a number of substrate and product analogs, several roles of the ribosome in catalysis were suggested.

First, the ribosome closely brings the two tRNAs together and orients them precisely (*substrate orientation*). The acceptor stems of the aminoacyl-tRNA (mimicked by an aminoacylated minihelix) and the peptidyl-tRNA (mimicked by CCA phenylalanine caproic acid biotin (CCApcb)) point into the cleft, and their 3′CCA ends are proximate to each other (Fig. 7). The 3′CCA sequence of the A-site tRNA interacts with the A loop (helix 92 of 23S rRNA) in the catalytic center, and the 3′CCA sequence of the P-site tRNA base-pairs with the P loop (helix 80 of 23S rRNA) (Fig. 7). The CCA sequence alone can bind to the large subunit, indicating that this is the most important interaction of tRNA with the 50S subunit. By base-pairing with the CCA ends, both RNA loops position the two acylated tRNAs precisely such that the

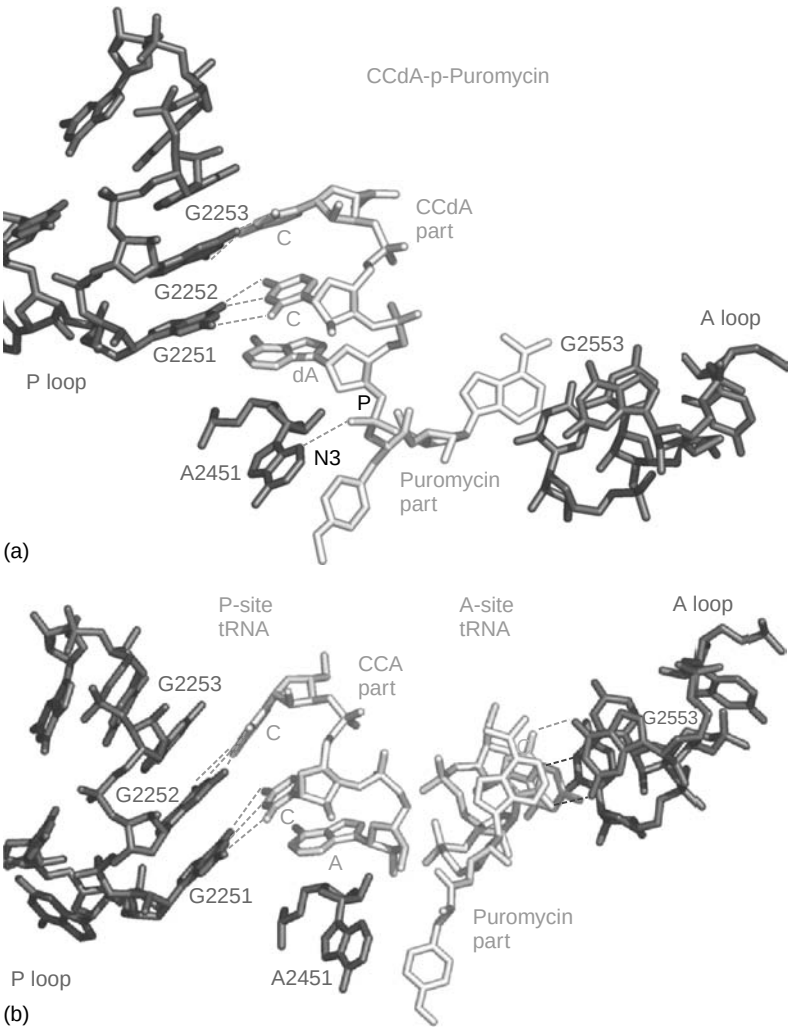

Fig. 7 (a) The structure of the peptidyl transferase center of the 50S ribosomal subunit. The A and P loop of 23S rRNA are shown in gray, the CCdA-p-Puromycin in white. The inhibitor occupies both the tRNA binding sites. The phosphoramide of the inhibitor is close to N3 of adenine2451. A2451 is part of the central loop in the peptidyl transferase center and may play a role in catalysis. (b) The peptidyl transferase center with a P-site and an A-site substrate analogs. In the P site, the cytosines of the CCA pair with two guanines G2251 and G2252 of the P loop. In the A site, the second base of CCA, the cytosine pairs with guanine 2553 of the A loop. The two substrate analogs mimic the situation in the active center immediately after peptidyl transfer.

α-amino group of the A-site aminoacyl-tRNA is close enough to the carbonyl carbon of the P-site peptidyl-tRNA for the nucleophilic attack (Fig. 2).

Second, *transition state stabilization* could lower the activation energy barrier as suggested by the 50S structure in complex with the CCdA-Puromycin inhibitor, which resembles a reaction intermediate. CCdA-p-Puromycin (Fig. 2) binds tightly to the P and A site loops and inhibits peptidyl transfer (Fig. 7a). Nevertheless, this compound is not a good analog because of the nonchiral phosphorus atom, the delocalized negative charge, and the missing 2′OH group on the CCdA part.

The third contribution to catalysis, suggested based on the *Haloarcula* structure with CCdA-p-Puromycin, could be the N3 of A2451 (*E. coli* numbering), which may participate in *acid–base catalysis*. This adenine is closest to the tetrahedral intermediate mimic of CCdA-p-Puromycin (Fig. 7a). Thus, N3 of A2451 could hydrogen-bond and abstract a proton from the amino group of the tetrahedral intermediate facilitating the release of the leaving group.

The pH-dependence of the peptidyl transfer reaction was analyzed by quench flow assays in which purified ribosomal complexes containing radioactive peptidyl-tRNA in the P site were incubated with puromycin at saturating concentrations. At pH 7.5, reaction rates of 50 s^{-1} were obtained, comparable with *in vivo* protein synthesis. At lower pH, the peptidyl transfer was 100-fold slower. Taking into account the pK_a of puromycin, the data indicated the existence of a single titratable group in the catalytic center with a pK_a of 7.5. The pH effect disappeared when an A2451U mutant was analyzed and the reaction rate was decreased by a factor of 130, consistent with the idea that A2451 acts as a general base. Nevertheless, alternative explanations involving protonation of neighboring residues are also possible.

Although the role of A2451 in transition state stabilization and acid–base catalysis is unclear, it certainly plays an important role in peptidyl transfer by substrate orientation and hydrogen bond formation to the α-amino group of aminoacyl-tRNA.

3.2 The Exit Tunnel

The structure of the large ribosomal subunit reveals a tunnel spanning its entire body through which the peptide emerges while still connected to the peptidyl transferase center (Fig. 8). The tunnel is approximately 100 Å in length and has a diameter of about 15 Å. The polypeptide exit tunnel is largely formed by 23S rRNA domains I and II but shows significant contributions from ribosomal proteins L4 and L22 (Fig. 8). Approximately 28-Å apart from the tunnel, exit proteins L4 and L22 fence the most constricted part of the tunnel with a diameter of 10 Å (Fig. 9). The opening of the exit tunnel is at the bottom of the back side of the 50S subunit. The exit is encircled by proteins L19, L22, L23, L24, L29, and L31 (Fig. 10).

The tunnel can accommodate 30 to 40 amino acid residues of a growing polypeptide chain and can therefore protect it from protease digestion or immunoprecipitation with antibodies specific for a nascent polypeptide. In fact, these initial experiments led to the proposal of the existence of a ribosomal tunnel about 35 years ago. Little is known about the conformation of the nascent polypeptide in the tunnel. Since the diameter of the tunnel is 10 Å at its narrowest part, the peptide could adopt an α-helical structure,

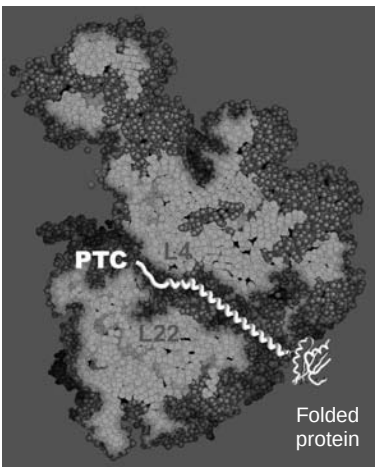

Fig. 8 The polypeptide exit tunnel. The 50S subunit is cut in half such that the tunnel is shown in its entire length. The ribosome atoms are shown in a space-filling representation. All RNA atoms that do not contact solvent are shown in white, and surface atoms are color-labeled with oxygen red, carbon yellow, and nitrogen blue. Proteins are depicted in green, and surface residues are in the same color code as RNA surface atoms. A model of a nascent chain polypeptide passing the tunnel is shown in white ribbon. The polypeptide is still connected to the peptidyl transferase center and can adopt an α-helical conformation in the tunnel. Outside the tunnel, the polypeptide chain can fold cotranslationally adopting its tertiary structure. The narrowest part of the tunnel is constricted by the two proteins L4 and L22. (See color plate p. xxxix).

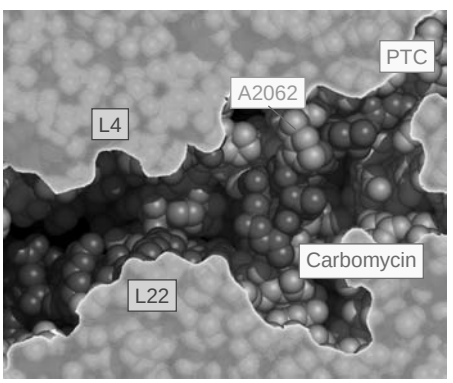

Fig. 9 A section of the ribosomal tunnel with Carbomycin A bound. The first and most constricted half of the tunnel is shown in a longitudinal section. All rRNA atoms are shown in yellow, and protein atoms of L4 and L22 are colored blue in a space-filling representation. The cutting plane is gray. The macrolide antibiotic Carbomycin A (red) is bound to the 50S tunnel blocking the passage of the nascent chain, and its disaccharide isobutyrate extension reaches to the peptidyl transferase center (PTC). A covalent bond is formed between N6 of A2062 of 23S rRNA and an ethylaldehyde substituent of Carbomycin A. (See color plate p. xxxiv).

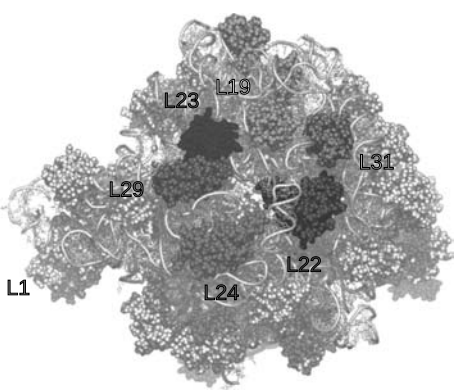

Fig. 10 The tunnel exit of the 50S subunit. Proteins at the rim of the exit tunnel are shown in the different colors. Some of the proteins are suggested to interact with the translocation machinery. L23 and L29 are implied to play a role in protein folding and transport across membranes based on their interaction with chaperones and SRP. The 23S rRNA and other ribosomal proteins are shown in white. (See color plate p. xxxvi).

which would be approximately 6 to 10 Å in diameter (Fig. 8). However, it appears to be unlikely that tertiary protein folding can occur within the tunnel although it widens toward the exit. 23S rRNA can be cross-linked to the nascent chains, although the surface of the tunnel should prevent tight binding of unfolded nascent polypeptides. The tunnel surface is largely composed of hydrophilic, noncharged groups, but no large hydrophobic patches that could form extensive interactions with exposed hydrophobic amino acids are found. Thus, because of selective pressure, the tunnel should allow the passage of any nascent protein with naturally occurring sequence.

3.3 Antibiotics Targeting the Large Subunit

The large ribosomal subunit is targeted by many antibiotics and substrate analogs. These compounds bind to the functionally important regions of the subunit such as the GTPase factor binding center, the peptidyl transferase center and the nascent polypeptide exit tunnel. Because of structural studies, detailed information of their mode of function is currently available for some of the peptidyl transferase inhibitors and the macrolide class of antibiotics.

Chloramphenicol prevents the accurate binding of aminoacyl-tRNA to the ribosomal A site. In contrast to puromycin that acts as an aminoacyl-tRNA analog, chloramphenicol interferes with the exact positioning of the aminoacyl moiety in the A site. Therefore, the formation of the tetrahedral transition state during the peptidyl transfer reaction may be hampered. In addition, the electronegative dichloromethyl group of chloramphenicol is close to the amino group of A site–bound aminoacyl-tRNA, and this may interfere with its nucleophilic attack on the carbonyl carbon atom of the ester bond that links the nascent polypeptide to the P-site anchored tRNA. The other antibiotic that interferes with substrate binding is *Clindamycin*. Binding of this compound overlaps with both A- and P site–bound substrates.

Macrolide antibiotics such as erythromycin block the diffusion of the nascent polypeptide into the tunnel. The exact binding site is found between the peptidyl transferase center and the narrowest part of the tunnel fenced by protein L4 and L22 (Fig. 9). The lactone ring of erythromycin binds with its hydrophobic face directed toward the tunnel wall forming van der Waals contacts with the tunnel surface, whereas its hydrophilic side is exposed. The sugar moieties of the macrolide form hydrogen bonds to the rRNA. Erythromycin binding reduces the diameter of the tunnel from 18 to 19 Å to less than 10 Å, thereby interfering with the path of the elongating peptide. Conversely, if the nascent chain is already in the tunnel, protein synthesis is not inhibited by erythromycin. Mutations in L4 and L22 causing an enlargement or a further constriction of the tunnel confers resistance to erythromycin. These mutations allow the passage of the polypeptide even though erythromycin is bound, or they prevent macrolide binding through conformational changes in the rRNA. Erythromycin binding does not inhibit the peptidyl transferase reaction, and the ribosome can synthesize short tri- or tetrapeptides in the presence of erythromycin. There is a clear correlation between the length of the macrolide substituent at the C5 position of the lactone ring extending toward the active center and the maximal length of the polypeptide that can still be translated. *Carbomycin A*, a macrolide antibiotic with an extended disaccharide isobutyrate

substituent on the lactone ring, interferes directly with the peptidyl transfer reaction since its extension reaches to the peptidyl transferase center. In addition to the van der Waals interactions and the hydrogen bonds, some macrolides like carbomycin A carrying an ethylaldehyde substituent on the C8 position can form a reversible covalent bond with N6 of A2062 (Fig. 9).

3.4 Regulatory Nascent Chains

It was recently discovered that certain polypeptide chains interact with the ribosomal exit tunnel. This contradicts the idea that the tunnel is a neutral environment for the exit of polypeptides and indicates that there is interaction between certain nascent peptides and the tunnel, which can regulate the rate of translation. Tight interaction between the nascent chain and the ribosomal tunnel leads to stalling of protein synthesis.

Such a stalling sequence motif was discovered at the C-terminus of the SecM (secretion monitor) protein of *E. coli* by mutational studies. It was determined that a particular sequence motif of 17 amino acids FxxxxWIxxxxGIRAGP is sufficient to cause ribosomal stalling. A genetic screen for stalling suppressor mutants revealed that the SecM peptide interacts tightly with 23S rRNA at the entrance of the exit tunnel and with protein L22 at the narrowest constriction of the tunnel, which is interestingly the same region in which macrolide antibiotics bind (Fig. 9). Accordingly, the stalling motif can be transferred from SecM to other proteins and causes translational arrest as well.

The real function of SecM is regulation of SecA expression through its N-terminal export signal. In the presence of abundant SecA, an ATPase, which is part of the translocation machinery, the stalled SecM peptide is "pulled" out of the ribosomal tunnel and delivered to the translocon. The basis of this regulation is a dicistronic mRNA encoding SecM and SecA, with a strong RNA stem-loop occluding the ribosome binding site of the secA gene. When SecA is missing, translation of the upstream secM gene causes ribosomal stalling, which resolves mRNA stem-loop and permits initiation of secA translation. In the presence of abundant SecA, translational arrest is abolished and the Shine-Dalgarno sequence of the secA gene is inaccessible for the ribosome, preventing SecA synthesis. Remarkably, interactions between the stalling sequence and the ribosome is so strong that overexpression of the SecM stalling sequence without the N-terminal signal sequence, recognized by SecA, is lethal for *E. coli*.

These results demonstrate that the exit tunnel is not a passive environment for the nascent chain. In fact, there are increasing number of effector sequences that either interfere with protein elongation or translation termination. The individual effector sequences share little sequence homology, although several sequences have positively charged residues, which could interact with ribosomal RNA, and the C-terminal residue is frequently a proline.

A distinct mechanism of translational arrest is mediated by binding of the signal recognition particle (SRP). When the translated polypeptide chain has a signal sequence at the N terminus, this sequence is recognized by SRP. The binding of the SRP to the ribosome-nascent chain complex causes a translation arrest, which is resolved only when the peptide binds to the translocon. The stalling effect could either result from a conformational change of the active site of the GTPase factor binding center upon SRP binding to the

ribosome, or it could be transmitted in a more subtle way through the peptide itself.

3.5 The GTPase Factor-Binding Center

The elongation factors EF-Tu and EF-G, as well as the other GTP-binding translation factors such as initiation factor IF2 and release factor RF3 interact with the large ribosomal subunit. The location of the GTPase domain binding site, named factor binding center, is known from cryoEM studies. The factor binding center includes the highly conserved *sarcin–ricin loop* (stem-loop 95 of 23S rRNA, Fig. 4), which is absolutely essential for factor binding. Ribosomes can be inactivated by the cleavage of a single covalent bond in the sarcin–ricin stem-loop. The rest of the factor binding region is formed by proteins L14, L23, and L6 on the body of the 50S subunit and proteins L11, L10, and L7/L12 on the flexible stalk. Conformational changes of the ribosome and the elongation factors upon GTP-hydrolysis were visualized by cryoEM. Sequence comparisons of EF-Tu, EF-G, IF2, and RF3 suggest that they all contain two domains with a similar fold that are responsible for binding to the GTPase activating center. This implies that they all bind to the ribosome in a similar fashion.

3.6 Structural Mimicry

Crystallographic studies revealed a *structural mimicry* of the ternary complex comprising tRNA, EF-Tu and GTP and EF-G : GTP. The tRNA shape is almost perfectly mimicked by the three domains of EF-G. This mimicry was further supported by cryoEM data of EF-G bound to the A site on the ribosome.

In fact, several proteins have been identified with a similar shape as the tRNA including release factors eRF1 and RF2, the ribosome recycling factor (RRF), and the tetracycline resistance protein. Clearly, there is also *functional mimicry* between aminoacyl-tRNAs and release factors RF1 and RF2, since both bind to the ribosomal A site, recognize specific mRNA codons, and simultaneously interact with the peptidyl transferase center. In prokaryotic RF1 and RF2, triplets of aminoacids have been identified that are responsible for the RF specificity. In eukaryotic eRF1, which recognizes all three stop codons, a highly conserved sequence motif has been identified as the "anticodon" serving region. It is assumed that the decoding regions of RFs directly interact with the mRNA codon in the ribosome. However, the detailed decoding mechanism of stop codons and the discrimination against sense codons by RFs is still unclear. The catalytic mechanism of peptide release is another unsolved question. Genetic studies suggest that the mechanism of peptide release is distinct from peptide formation but involves 23S rRNA as well.

The final step in translation termination is the disassembly of the ribosomal complex and recycling of the ribosomal subunits. This step is supported by EF-G and RRF, which is a nearly perfect mimic of tRNA.

4 The 70S Ribosome Structure

The *T. thermophilus* 70S structure at 5.5 Å, containing mRNA and three tRNAs, sheds light on the interactions between the 50S and 30S subunit (Fig. 4). In comparison to the isolated subunits, 70S ribosomes show certain conformational differences, most

of them stem directly from intersubunit association. Some flexible regions of the 50S structure are ordered in the 70S ribosome, indicating that these elements may adopt a defined geometry upon 30S association.

4.1
The 30S–50S Interface

The 70S structure reveals a network of intermolecular bridges involving RNA–RNA and RNA–protein interactions as well as a single protein–protein interaction. Mobile *intersubunit bridges* are suggested to play a major role in translocation because they allow movement of the ribosomal subunits with respect to each other. Viewed from the interface, ribosomal proteins are mainly located at the periphery of the 30S and 50S subunits (Fig. 4). The interface consists predominantly of RNA: twelve individual intersubunit bridges were identified, most of which are RNA helices. These contacts mostly involve minor groove–minor groove interactions. The RNA bridges are located in the center of the interface, on the 30S platform and helix 44, adjacent to the tRNA binding sites. Contacts involving proteins are peripheral to the functional sites. One of the most important intersubunit contacts seems to be the centrally located flexible bridge element formed by interaction of helix 69 of 23S rRNA with helix 44 of 16S rRNA at the base of the decoding center. Breakage and formation of many of these interface bridges probably accompanies translocation.

4.2
Translocation – The Concerted Movement of tRNAs and mRNA

After peptidyl transfer, the empty P-site tRNA and the peptidyl-tRNA in the A site have to move in concert with the mRNA. Several models have been proposed for translocation. Widely accepted is the *hybrid state model*, which involves independent tRNA movement with respect to the two subunits (Fig. 11). Pursuant to this model, tRNAs can occupy different tRNA binding sites on the small and large subunit, corresponding to a hybrid state. On the small subunit, the mRNA moves in complex with the tRNA anticodon stem-loops, while the acceptor stems of the tRNA are fixed in their respective binding sites on the large subunit (Fig. 11).

Chemical footprinting and cryoEM was used to monitor the movement of tRNAs through the ribosome. Following peptidyl transfer, it was found that the 50S A-site tRNA footprint disappeared and a 50S E-site tRNA footprint emerged instead. At the same time, the tRNA footprints on the small subunit did not change. The hybrid state model suggests a transition into a state in which the former A-site tRNA (A/A) still occupies the A site on the small subunit but its acceptor stem moved into the P site on the large subunit (A/P hybrid state). Concomitantly, the former P-site tRNA (P/P) still forms codon–anticodon contacts in the P site of the small subunit but, at the same time, moves away from the peptidyl transferase center on the large subunit into the E site (P/E hybrid state). This rearrangement is independent of EF-G. In cryoEM, only a small movement of the A-site acceptor stem toward the P site is visible after the peptidyl transfer reaction. In the *Haloarcula* 50S structure, the peptidyl-tRNA remained bound to the A loop in the peptidyl transferase center (the loop that binds A-site aminoacyl-tRNA, Fig. 7b). The P/E hybrid state has been observed by cryoEM, but only under stringent buffer conditions.

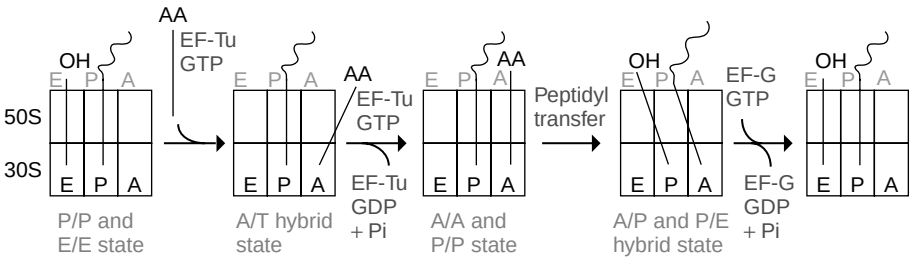

Fig. 11 The hybrid state model for translocation. The tRNAs move independently within the active sites of the two ribosomal subunits. Translocation from the A/T hybrid state into the A/A state depends on GTP hydrolysis by EF-Tu. Translocation into the P/E and A/P hybrid state occurs spontaneously after peptidyl transfer. In contrast, the translocation of the tRNAs and mRNA within the small subunit from the P/E : A/P hybrid states to the E/E : P/P state requires GTP hydrolysis by EF-G.

Upon incubation of the A/P and P/E hybrid state with EF-G and GTP, the tRNAs move to the P/P (occupying the P site on both subunits) and E/E site (Fig. 11), as evidenced by chemical footprinting. This means that upon GTP hydrolysis, EF-G induces tRNA movement with respect to the 30S subunit whereas movement with respect to the 50S subunit occurs spontaneously after peptidyl transfer. Curiously, GTPase activity of EF-G is not essential since translocation can still occur in the presence of nonhydrolyzable GTP analogs, although at a slower rate.

A third type of hybrid state (the *A/T hybrid state*, T for EF-Tu) is observed when the ternary complex of EF-Tu, GTP, and tRNA binds to the ribosome (Fig. 11). In chemical footprinting studies, this complex of tRNA EF-Tu and GTP-protected 16S rRNA in the 30S A site, but there was no tRNA footprint in the 50S A site. Instead, bases in the sarcin–ricin loop in the GTPase factor binding center of the 50S subunit were protected in a similar fashion as upon EF-G binding. After EF-Tu GTP hydrolysis and accommodation, the 50S A site becomes occupied by tRNA.

A prerequisite for a hybrid state is the mobility of the two subunits relative to each other or at least of parts of the subunits. Indeed, it has been observed by cryoEM that the two subunits can rotate 6° relative to each other upon EF-G binding. Following EF-G dependent GTP hydrolysis, translocation occurs, coupled with EF-G release and simultaneous back-rotation of the 30S subunit relative to the 50S subunit. In addition to subunit rotation, regions of the ribosomal subunits, such as intersubunit bridges, could move together with the tRNAs during translocation.

What exactly could be the driving force for the translocation in the 50S subunit? First, the peptidyl transferase reaction could directly provide the free energy for spontaneous translocation from the A/A state to the A/P state. Second, the tRNA binding sites on the large subunit could have different substrate specificities. In that case, the A site would have highest affinity for aminoacyl-tRNA, the P site a higher affinity for peptidyl-tRNA and the E site for deacylated tRNA. The latter possibility would explain the directionality of the process. There is evidence that peptidyl transfer and translocation occur in a sequential manner on the large subunit. Contrariwise, the translocation of tRNAs and mRNA on the small subunit

requires energy input. The movement of the tRNA from the A/A state to the A/P hybrid state could decrease the affinity of the tRNA–mRNA helix for the small subunit. This would result in lower activation energy for translocation on the small subunit, provided that the affinity of tRNAs and the bound mRNA for 16S rRNA really presents an energy barrier for the movement between the tRNA binding sites on the small subunit. Moreover, if EF-G binding and GTP hydrolysis depend on a hybrid state of the ribosome, the coupling of peptidyl transfer reaction and translocation would be ensured.

4.3 Movement of tRNAs through the Ribosome

In the 70S structure, the tRNAs are in A/A, P/P, and E/E states rather than in any hybrid state. The large subunit makes contacts with the conserved structures of the tRNA, that is, the tRNA D-stem, elbow, and acceptor arm. The A and P site tRNAs are closest to each other at their acceptor stems, which are within 5-Å vicinity. In the small subunit, the anticodon ends of the tRNAs and the mRNA move only 20 to 28 Å between the A and P site or P and E site respectively. At the same time, the acceptor stem moves over much larger distances in the large subunit. For instance, the E site acceptor stem is rotated away from the P-site tRNA acceptor stem and directed into a separate cleft nearly 50 Å away from the peptidyl transferase cavity (Fig. 4). The L1 stalk and protein S7 seem to lock the deacylated tRNA in the E site as evidenced by the 70S X-ray structure. In the free 50S structure, the L1 stalk has an open conformation, in which it would not prevent release of E-site tRNA. A large rotational movement of tRNA is also necessary during its accommodation from the A/T hybrid state into the A site of the peptidyl transferase center.

4.4 Cotranslational Folding

Upon leaving the tunnel, the peptide chain can fold into its tertiary structure while still being bound to the ribosome. Cotranslational folding has been demonstrated, *in vitro* and *in vivo*, as well as in eukaryotic and prokaryotic cells. For large multidomain proteins, folding in a sequential manner may be beneficial to the folding process to avoid interference between individual protein domains during early folding steps. In the tunnel, the nascent polypeptide can adopt α-helical conformation (Fig. 8), but the diameter of the tunnel does not allow significant tertiary structure formation beyond that. Depending on the amino acid sequence of the nascent polypeptide, a different extent of secondary structure in the tunnel is acquired as indicated by different patterns of cross-linking to 23S rRNA. Moreover, depending on the sequence, different lengths of nascent polypeptides (30–72 amino acids) are protected by the ribosome, suggesting a varying extent of secondary structure formation. The opening of the tunnel is surrounded by ribosomal proteins L19, L22, L23, L24, L29, and L31 and regions of 23S rRNA. These elements may assist in protein folding (Fig. 10).

During translation, the ribosome-nascent peptide chain complexes can interact with various proteins involved in protein targeting and folding, for instance, molecular chaperones like the trigger factor, DnaK/DnaJ/GrpE in *E. coli*, and their homologs in eukaryotes. *Trigger factor*, a cis/trans prolyl isomerase that has protein-folding activity, was shown to be associated

with the large ribosomal subunit when binding the nascent chain polypeptide. The N-terminal domain of trigger factor was identified as a ribosome binding domain. It could be cross-linked to proteins L23 and L29, which are located adjacent to each other at the tunnel exit site, exactly where the nascent chain emerges from the tunnel (Fig. 10). By mutational studies, it was shown that ribosome association is essential for trigger factor chaperone activity *in vitro* and *in vivo* and that an exposed, conserved glutamate (E18) of L23 protein is critical for the binding. Thus, ribosomal protein L23 directly links protein synthesis with chaperone-assisted folding.

4.5 Protein Secretion and Membrane Insertion

The ribosome participates directly in the transport of nascent chains into or across membranes to the periplasm or endoplasmic reticulum. Most of the secreted proteins contain an N-terminal hydrophobic signal sequence recognized by the *SRP*. Helicity as well as hydrophobicity of the *signal sequence* are necessary for the productive interaction with SRP and the translocation machinery. SRP binds and targets the ribosome-nascent chain complex to its receptor in the membrane. Subsequently, SRP hands over the ribosomal complex to its receptor, which coordinates the release of SRP from the ribosome-nascent chain complex and the insertion of the nascent chain into the translocation channel, which is built up by the Sec61 complex in eukaryotes and by SecYEG in prokaryotes.

In eukaryotes, SRP is composed of a 7S RNA and six proteins. The 54 kD protein SRP54 has been identified to bind the signal peptide and to interact with the signal receptor. SRP54 can be cross-linked to the large ribosomal subunit in the presence and absence of a nascent chain. Thus, this is clearly the ribosome binding domain of SRP. The two adjacent ribosomal proteins L23 and L35 (L29 in *E. coli*), both located at the rim of the tunnel exit, are the binding partners of SRP54, (Fig. 10). SRP proteins SRP9 and SRP14 are required for the SRP induced translational arrest (as discussed earlier). The nascent chain is released when SRP54 binds GTP and interacts with its receptor. The SRP receptor induces changes in the SRP binding to the ribosome and permits access of the translocation channel to the ribosome-nascent chain complex, such that the signal peptide is handed over to the Sec61 complex and translation continues. In the presence of SRP, direct binding of the Sec61 complex to the ribosomal complexes is blocked.

The structure of the eukaryotic ribosome-nascent chain complex and the eukaryotic *translocation machinery* Sec61 has been studied by cryoEM. The Sec61 complex consists of three transmembrane subunits α, β, and γ, comprising 12 transmembrane helices altogether. Three or four Sec61 complexes assemble to form the functional ringlike translocon with a central pore or indentation. The ringlike Sec61 complex binds to the large ribosomal subunit such that its central pore is aligned with the opening of the protein exit tunnel. Thereby, a path for the nascent chain is formed from the peptidyl transferase center directly into the endoplasmic reticulum. Four connections were identified between the Sec61 complex and the ribosome-nascent chain complexes in the 15-Å cryoEM structure. They involve mostly ribosomal RNA (28S) but also involve the eukaryotic homologs of the ribosomal proteins L19,

L23, L24 and L29 (Fig. 10); the region next to the tunnel exit is relatively conserved in prokaryotes and eukaryotes. The most substantial connection is found for proteins L23, L29, and helix7 of the rRNA. It is interesting to note that L23 and L29 are also implied to play a role in trigger factor and SRP binding, and this may be the reason why deletion of L23 is lethal.

The binding of the translocation machinery to the ribosome leaves a gap between ribosome and the transmembrane channel of at least 15 Å. This finding contradicts earlier postulations that the translocon/ribosome interface should form an ion-tight seal. The cryoEM structure suggests that the channel itself is the permeability barrier between cytoplasm and lumen.

Cotranslational folding also occurs during translocation of proteins into the endoplasmatic reticulum. The minimum distance between the folded domain and the peptidyl transferase center was determined to be 64 amino acids. This agrees well with the protection of 70 amino acids from proteolysis and the minimal peptide length of 70 amino acids needed for glycosylation by the glycosylase in the ER at the exit of the translocon pore. All biochemical and structural data support the idea that the translocon channel extends the ribosomal exit tunnel, even though there appears to be a gap between them. The translocon is probably a dynamic complex, since it allows lateral diffusion of the nascent chain into the membrane bilayer, which probably requires opening of the channel ring.

5 Perspectives

Recent crystal structures improved our understanding of the mechanism of protein synthesis. The peptidyl transferase reaction and the accuracy of aminoacyl-tRNA selection can now be studied with detailed structure-based experiments (Fig. 3). However, many questions remain to be addressed on a molecular level. Especially since information about intersubunit interactions, conformational changes during translation and interaction with translation factors originate from low-resolution structures. Ultimately, all relevant intermediates, complexes, and conformational states of the ribosome should be characterized at high resolution to completely understand translation and associated processes like cotranslational folding and translocation across the membrane.

Functionally important regions of the ribosome including the decoding center, the peptidyl transferase center, and the tunnel are highly conserved. Therefore, it is likely that the fundamental mechanism of translation is the same in all kingdoms of life. Most differences in the architecture of the eukaryotic 80S ribosome compared to the 70S ribosomes are generated by additional proteins and RNA insertions into the periphery of the structure. Nevertheless, eukaryotic translation is much more complex and subjected to a more diversified regulation. Extracting structural information on the eukaryotic ribosome will be undoubtedly difficult, but the task is well worth the extraordinary effort.

See also Expression Systems for DNA Processes; Plant Gene Expression, Regulation of; Protein Expression by Expansion of the Genetic Code; RNA Methodologies; Translation of RNA to Protein.

Bibliography

Books and reviews

Frank, J. (2001) Cryo-electron microscopy as an investigative tool: the ribosome as an example, *BioEssays* **23**, 725–732.

Kramer, G., Ramachandiran, V., Hardesty, B. (2001) Cotranslational folding–omnia mea mecum porto? *Int. J. Biochem. Cell Biol.* **33**, 541–553.

Moore, P.B., Steitz, T.A. (2002) The involvement of RNA in ribosome function, *Nature* **418**, 229–235.

Nakamura, Y., Ito, K. (2003) Making sense of mimic in translation termination, *Trends Biochem. Sci.* **28**, 99–105.

Nissen, P., Kjeldgaard, M., Nyborg, J. (2000) Macromolecular mimicry, *EMBO J.* **19**, 489–495.

Noller, H.F. (1991) Ribosomal RNA and translation, *Annu. Rev. Biochem.* **60**, 191–227.

Noller, H.F., Yusupov, M.M., Yusupova, G.Z., Baucom, A., Cate, J.H. (2002) Translocation of tRNA during protein synthesis, *FEBS Lett.* **514**, 11–16.

Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation, *Cell* **108**, 557–572.

Rodnina, M.V., Wintermeyer, W. (2001) Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms, *Annu. Rev. Biochem.* **70**, 415–435.

Tenson, T., Ehrenberg, M. (2002) Regulatory nascent peptides in the ribosomal tunnel, *Cell* **108**, 591–594.

Primary Literature

Ban, N., Nissen, P., Hansen, J., Moore, P.B., Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution, *Science* **289**, 905–920.

Bashan, A., Agmon, I., Zarivach, R., Schluenzen, F., Harms, J., Berisio, R., Bartels, H., Franceschi, F., Auerbach, T., Hansen, H.A.S., Kossoy, E., Kessler, M., Yonath, A. (2003) Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression, *Mol. Cell* **11**, 91–102.

Beckmann, R., Spahn, C.M., Eswar, N., Helmers, J., Penczek, P.A., Sali, A., Frank, J., Blobel, G. (2001) Architecture of the protein-conducting channel associated with the translating 80S ribosome, *Cell* **107**, 361–372.

Blobel, G., Sabatini, D.D. (1970) Controlled proteolysis of nascent polypeptides in rat liver cell fractions. I. Location of the polypeptides within ribosomes, *J. Cell Biol.* **45**, 130–145.

Carter, A.P., Clemons, W.M., Brodersen, D.E., Morgan-Warren, R.J., Wimberly, B.T., Ramakrishnan, V. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics, *Nature* **407**, 340–348.

Carter, A.P., Clemons, Jr., W.M., Brodersen, D.E., Morgan-Warren, R.J., Hartsch, T., Wimberly, B.T., Ramakrishnan, V.V. (2001) Crystal structure of an initiation factor bound to the 30S ribosomal subunit, *Science* **4**, 4.

Choi, K.M., Brimacombe, R. (1998) The path of the growing peptide chain through the 23S rRNA in the 50S ribosomal subunit; a comparative cross-linking study with three different peptide families, *Nucleic Acids Res.* **26**, 887–895.

Dallas, A., Noller, H.F. (2001) Interaction of translation initiation factor 3 with the 30S ribosomal subunit, *Mol. Cell* **8**, 855–864.

Frank, J., Agrawal, R.K. (2000) A ratchet-like inter-subunit reorganization of the ribosome during translocation, *Nature* **406**, 318–322.

Hansen, J.L., Schmeing, T.M., Moore, P.B., Steitz, T.A. (2002) Structural insights into peptide bond formation, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11670–11675.

Hansen, J.L., Ippolito, J.A., Ban, N., Nissen, P., Moore, P.B., Steitz, T.A. (2002) The structures of four macrolide antibiotics bound to the large ribosomal subunit, *Mol. Cell* **10**, 117–128.

Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., Yonath, A. (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium, *Cell* **107**, 679–688.

Katunin, V.I., Muth, G.W., Strobel, S.A., Wintermeyer, W., Rodnina, M.V. (2002) Important contribution to catalysis of peptide bond formation by a single ionizing group within the ribosome, *Mol. Cell* **10**, 339–346.

Kowarik, M., Kung, S., Martoglio, B., Helenius, A. (2002) Protein folding during co-translational translocation in the endoplasmic reticulum, *Mol. Cell* **10**, 769–778.

Kramer, G., Rauch, T., Rist, W., Vorderwulbecke, S., Patzelt, H., Schulze-Specking, A., Ban, N., Deuerling, E., Bukau, B. (2002) L23 protein functions as a chaperone docking site on the ribosome, *Nature* **419**, 171–174.

Malkin, L.I., Rich, A. (1967) Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding, *J. Mol. Biol.* **26**, 329–346.

McCutcheon, J.P., Agrawal, R.K., Philips, S.M., Grassucci, R.A., Gerchman, S.E., Clemons, Jr., W.M., Ramakrishnan, V., Frank, J. (1999) Location of translational initiation factor IF3 on the small ribosomal subunit, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4301–4306.

Menetret, J.F., Neuhof, A., Morgan, D.G., Plath, K., Radermacher, M., Rapoport, T.A., Akey, C.W. (2000) The structure of ribosome-channel complexes engaged in protein translocation, *Mol. Cell* **6**, 1219–1232.

Nakatogawa, H., Ito, K. (2002) The ribosomal exit tunnel functions as a discriminating gate, *Cell* **108**, 629–636.

Nissen, P., Hansen, J., Ban, N., Moore, P.B., Steitz, T.A. (2000) The structural basis of ribosome activity in peptide bond synthesis, *Science* **289**, 920–930.

Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B., Steitz, T.A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4899–4903.

Ogle, J.M., Murphy, F.V., Tarry, M.J., Ramakrishnan, V. (2002) Selection of tRNA by the ribosome requires a transition from an open to a closed form, *Cell* **111**, 721–732.

Ogle, J.M., Brodersen, D.E., Clemons, Jr., W.M., Tarry, M.J., Carter, A.P., Ramakrishnan, V. (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit, *Science* **292**, 897–902.

Piepenburg, O., Pape, T., Pleiss, J.A., Wintermeyer, W., Uhlenbeck, O.C., Rodnina, M.V. (2000) Intact aminoacyl-tRNA is required to trigger GTP hydrolysis by elongation factor Tu on the ribosome, *Biochemistry* **39**, 1734–1738.

Pioletti, M., Schluenzen, F., Harms, J., Zarivach, R., Gluehmann, M., Avila, H., Bashan, A., Bartels, H., Auerbach, T., Jacobi, C., Hartsch, T., Yonath, A., Franceschi, F. (2001) Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3, *EMBO J.* **20**, 1829–1839.

Pool, M.R., Stumm, J., Fulga, T.A., Sinning, I., Dobberstein, B. (2002) Distinct modes of signal recognition particle interaction with the ribosome, *Science* **297**, 1345–1348.

Schlunzen, F., Zarivach, R., Harms, J., Bashan, A., Tocilj, A., Albrecht, R., Yonath, A., Franceschi, F. (2001) Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria, *Nature* **413**, 814–821.

Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., Yonath, A. (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution, *Cell* **102**, 615–623.

Schmeing, T.M., Seila, A.C., Hansen, J.L., Freeborn, B., Soukup, J.K., Scaringe, S.A., Strobel, S.A., Moore, P.B., Steitz, T.A. (2002) A pre-translocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits, *Nat. Struct. Biol.* **4**, 4.

Stark, H., Rodnina, M.V., Wieden, H.J., Zemlin, F., Wintermeyer, W., van Heel, M. (2002) Ribosome interactions of aminoacyl-tRNA and elongation factor Tu in the codon-recognition complex, *Nat. Struct. Biol.* **9**, 849–854.

Valle, M., Sengupta, J., Swami, N.K., Grassucci, R.A., Burkhardt, N., Nierhaus, K.H., Agrawal, R.K., Frank, J. (2002) Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process, *EMBO J.* **21**, 3557–3567.

Welch, M., Chastang, J., Yarus, M. (1995) An inhibitor of ribosomal peptidyl transferase using transition-state analogy, *Biochemistry* **34**, 385–390.

Wimberly, B.T., Brodersen, D.E., Clemons, Jr., W.M., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit, *Nature* **407**, 327–339.

Yusupova, G.Z., Yusupov, M.M., Cate, J.H., Noller, H.F. (2001) The path of messenger RNA through the ribosome, *Cell* **106**, 233–241.

Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H., Noller, H.F. (2001) Crystal structure of the ribosome at 5.5 A resolution, *Science* **292**, 883–896.

Ribosome Preparation and Protein Synthesis Techniques: *see* Ribosome, High Resolution Structure and Function

Ribozymes

Joyce E. Heckman
The University of Vermont, Burlington, VT

1 RNA Can Encode Genetic Information, Participate in Translating it, and also Catalyze Biochemical Reactions 515

2 RNA Structure and Catalysis 515

3 New Understanding of the Importance of RNA in Biological Systems 516

4 Self-splicing RNA 517

5 Self-cleaving Catalytic RNA 518

6 Ribonucleoprotein Enzymes: The Ribosome is a Ribozyme! 519

7 *In Vitro* Selection 520

Bibliography 523
Books and Reviews 523
Primary Literature 523

Keywords

Exon
The sequences within a gene or an RNA transcript that are represented in the final mature RNA product of a gene. Before RNA splicing, exons are separated by introns.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

Intron
An internal sequence of nucleotides within a gene or an RNA transcript that is removed by RNA splicing before the formation of the functional RNA product.

Polymerase Chain Reaction (PCR)
A method utilizing a DNA polymerase and primers complementary to each end of a targeted DNA region to produce multiple copies of a double-stranded DNA sequence.

Reverse Transcriptase
An enzyme, derived from retroviruses, that copies RNA sequences to produce DNA.

Ribonucleoprotein
A macromolecular complex comprising both RNA and protein.

Ribozyme
An RNA molecule that is endowed with catalytic activity; for *ribo*nucleic acid en*zyme*.

RNA Processing
The biochemical reactions the primary RNA transcript of a gene must undergo before it can become a mature, functional RNA.

RNA Splicing
The removal of intron sequences and joining of exons to generate a mature RNA molecule.

Self-cleaving RNA
An RNA molecule that is capable of cleaving itself in a site-specific manner.

Ribozymes are RNA molecules that can function to catalyze specific intracellular chemical reactions without the obligatory participation of proteins. Until recently, all biochemical reactions were believed to be catalyzed by protein enzymes. Since 1982, a number of RNA molecules have been found to be endowed with catalytic activity. For example, group I ribozymes take the form of introns, which can mediate their own excision from a self-splicing precursor RNA. Other ribozymes are derived from self-cleaving RNA structures that are essential for the replication of viral RNA molecules. Like protein enzymes, ribozymes can fold into complex three-dimensional structures that provide specific binding sites for substrates as well as cofactors, such as metal ions. Ribozymes have been developed into highly specific enzymes that are capable of recognizing and catalytically cleaving targeted RNA molecules. A wide variety of activities can be elicited from ribozymes by selection, and targeted ribozymes have been used scientifically and therapeutically to cleave viral RNAs and messenger RNAs associated with disease.

1
RNA Can Encode Genetic Information, Participate in Translating it, and also Catalyze Biochemical Reactions

RNA is a unique molecule in the biological world because it serves several fundamentally distinct functions. It has long been known that RNA, like DNA, can serve as an informational molecule. For example, messenger RNA transcripts carry the information encoded by genes to the cell's protein synthesis machinery. Many viruses, like HIV and influenza, use RNA as their genetic molecule. Transfer RNAs help decode the information from messenger RNAs, and ribosomal RNAs are large, essential functioning components of the ribosomes, the cellular machines that produce proteins by translating messenger RNAs.

More recently, investigators have found that some RNA molecules resemble protein enzymes in that they can catalyze specific biochemical reactions. The discovery that RNA can be a catalyst has revolutionized our view of RNA and its importance, and it has led to the speculation that catalytic RNA may have played a critical role in the prebiotic evolution of self-replicating systems.

2
RNA Structure and Catalysis

The catalytic activity of ribozymes results from the three-dimensional folded structure of RNA. Because RNA is a single-stranded molecule, it can fold into structures much more complex than the monotonous structure of the DNA double helix. Much is known about the secondary structure of ribozymes (Fig. 1), which generally consists of a number of short intramolecular helices composed of Watson–Crick G–C and A–U base pairs, separated by nonhelical segments. There are computer applications that can predict likely secondary structures for given RNA sequences. It is clear that the nonhelical segments of ribozymes are very important for catalytic activity; these nonhelical nucleotides interact with one another to form a tertiary structure. They may also provide functional groups that help catalyze biochemical reactions.

Recent advances in X-ray crystallography of RNA have provided three-dimensional crystal structures for several small ribozymes. In at least two of these cases, the crystal structures correlate well with biochemical data, showing how important residues of the ribozyme are positioned very close to the cleavable linkage of the target RNA, in such a fashion that functional groups from the ribozyme could be helping to catalyze the chemical reaction. These crystal structures also help explain why some catalytic RNAs can function without divalent ions like magnesium. Magnesium helps in folding RNA to its tertiary structure, but some ribozymes function well in high concentrations of monovalent ions like sodium. This is important because it shows that these catalytic RNAs do not function merely by positioning a magnesium ion to catalyze a reaction. Instead, the catalytic function is performed by chemical groups on the RNA itself.

RNA differs from DNA in the presence of a 2′-hydroxyl group on the sugar of each nucleotide. This functional group may be important for catalytic activity in three respects. First, it makes RNA much more chemically labile than DNA. Second, it participates in essential tertiary interactions within ribozymes that are essential for catalytic activity. Third, it

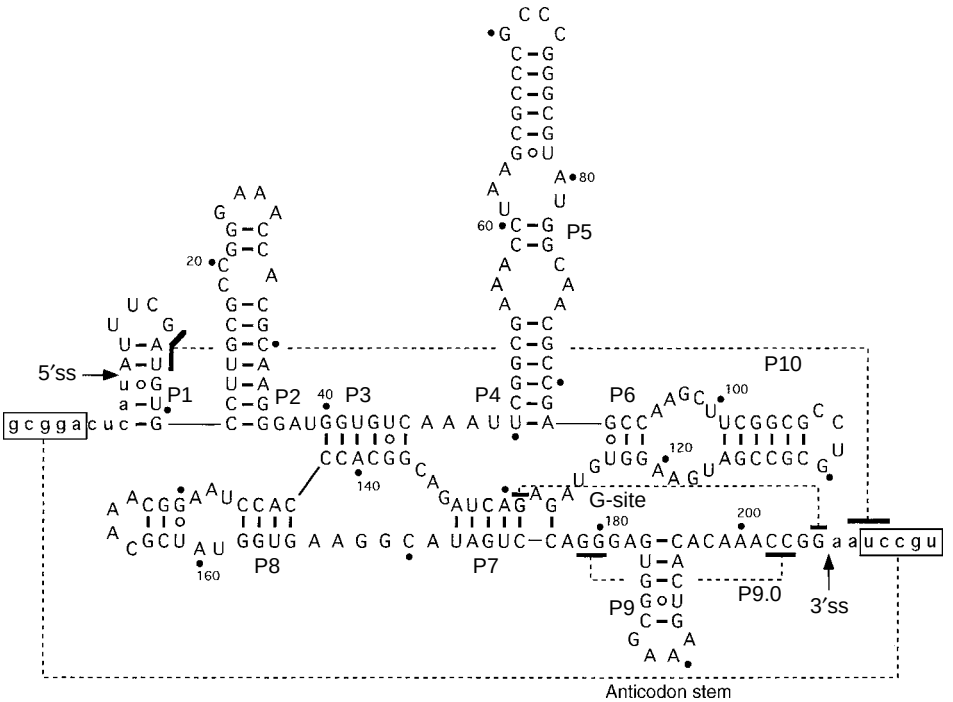

Fig. 1 The secondary structure of a self-splicing group I intron within a transfer RNA precursor of the bacterium *Azoarcus*: arrows indicate splice sites, intron and exon nucleotides are indicated in uppercase and lowercase letters respectively. All group I introns have very similar RNA folding patterns. The RNA folding is responsible for catalytic activity and for the precise identification of the reaction sites.

may participate directly in the chemical reactions that take place at the active site of the ribozyme.

3 New Understanding of the Importance of RNA in Biological Systems

In recent years, investigators have discovered important new biological roles carried out by RNA in cells and organisms. In eukaryotes, small interfering RNAs (siRNAs) only about twenty nucleotides in length, derived from double-stranded RNAs, serve as guides to target enzymatic degradation of foreign RNA and viral messenger RNAs. Small cellular species called *micro RNAs* (miRNAs) interact with messenger RNAs to modulate their translation during development.

In bacterial systems, Breaker and colleagues discovered RNA motifs called *riboswitches* at the leading ends of important messenger RNAs. These motifs bind specific small molecules, often the products of the enzymes produced from the messenger RNA. Binding of this metabolic product to the RNA motif changes its folding pattern and decreases its expression, either in transcription or translation, creating a negative feedback control at the level of RNA. One of these riboswitches is actually a new small ribozyme, which,

upon binding its specific small ligand, becomes activated to cleave its own messenger RNA.

Recent studies of ribosomes (see the following) have shown that the actual chemical reaction joining amino acids to form a new protein is almost certainly carried out by the RNA component of the ribosome rather than by protein, as previously assumed. Moreover, there is good preliminary evidence that RNA may also catalyze the splicing steps that remove introns from messenger RNAs in eukaryotic nuclei. In summary, RNA has more active and important roles in biology than were previously understood. Ribozymes, rather than being evolutionary oddities, are central players in the biochemistry of life.

4
Self-splicing RNA

The first catalytic RNA molecules discovered were self-splicing RNA molecules containing group I introns. In the absence of proteins, these molecules can undergo rapid and precise structural rearrangements to excise the intron and covalently join the exons. Group I introns contain characteristic internal nucleotide sequences and have very similar RNA folding patterns (Fig. 1); these common structural features are essential for catalytic activity and for recognition of the correct reaction sites in the precursor RNA.

Splicing of precursor RNAs containing group I introns occurs by a two-step reaction pathway (Fig. 2). In the first step, GTP attacks the 5′ splice site, becoming covalently linked to the intron and liberating the 5′ exon. In the second step, the 5′ exon attacks the 3′ splice site, joining the exons together and liberating the intron RNA. It is clear that the intron RNA itself contains the catalytic center, because it can catalyze additional reactions on internal and external RNA substrates. A second class of self-splicing introns, the group II introns, is known. These introns share common folded structures and reaction pathways that are distinct from

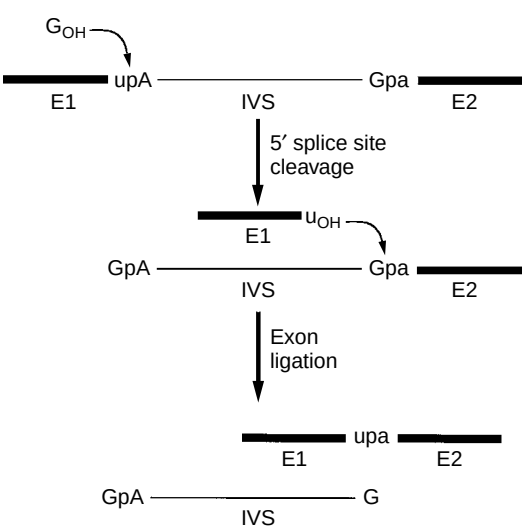

Fig. 2 RNA splicing pattern for group I introns. First, GTP or another guanosine attacks the 5′ splice site, liberating the 5′ exon and becoming covalently linked to the intron. Second, the 5′ exon attacks the 3′ splice site in a reaction that generates ligated exons and liberates the intron. Each of these transesterification reactions can take place in the presence or in the absence of proteins.

those of group I introns. Interestingly, group II introns appear to be more closely related to the nuclear pre-mRNA introns, whose splicing is catalyzed by complex nuclear particles called *spliceosomes*. As mentioned in Sect. 3, there is some evidence that the chemistry carried out by the spliceosome may be catalyzed by one or more of the small nuclear RNAs contained in it.

5 Self-cleaving Catalytic RNA

In addition to their importance for RNA splicing, catalytic RNA molecules are essential for RNA replication in some viral systems. For example, a circular RNA molecule associated with tobacco ringspot virus is replicated by a rolling-circle mechanism, as shown in Fig. 3. Site-specific RNA cleavage and joining reactions are essential for RNA replication. Instead of using protein ribonucleases and RNA ligases as catalysts, these reactions are carried out by the RNA molecules themselves.

That is, during replication the RNA molecules contain sites at which the cleavage and joining reactions happen, as well as domains that recognize those sites and catalyze the reactions. A similar replication pathway is followed by the RNA genome of an important human pathogen, hepatitis delta virus.

Deletion of the nonessential RNA sequences serves to isolate the catalytic domains and the sequences that are recognized and cleaved. The catalytic domains function as ribozymes to cleave the substrate in a catalytic manner; that is, each ribozyme molecule can potentially cleave a large number of substrate RNA molecules.

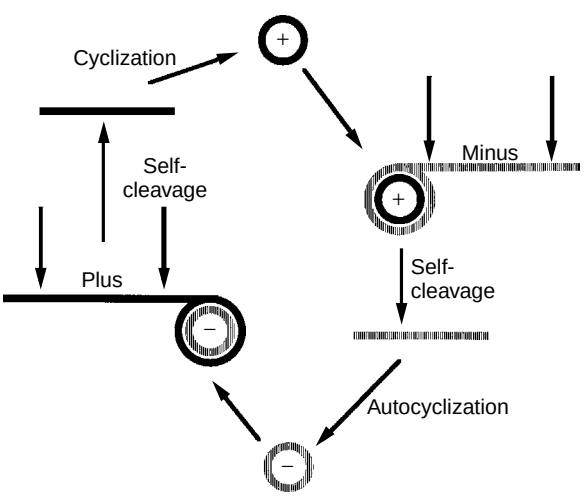

Fig. 3 Self-cleaving RNA in RNA replication: the replication pathway for the satellite RNA associated with tobacco ringspot virus. Rolling-circle replication gives rise to linear negative- and positive-stranded concatemers that must be cleaved and cyclized at specific sites. The self-cleaving domains in the positive and negative strands are known as the hammerhead ribozyme and the hairpin ribozyme, respectively.

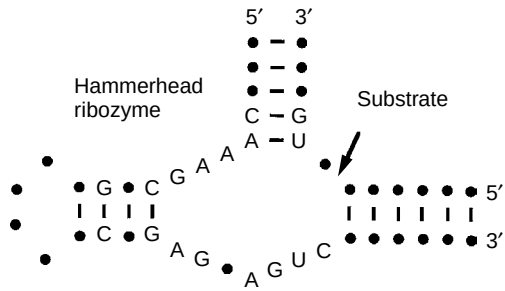

Fig. 4 The secondary structure of the complex between the hammerhead ribozyme and its substrate: arrow indicates cleavage site; essential bases are indicated by letters; one dot indicates that any base can function at the specified position; two dots joined by a line indicates that any base pair will function. The substrate may be a short RNA oligonucleotide for laboratory biochemical studies, or it may be a part of a long RNA molecule (e.g. a cellular mRNA or viral genomic RNA molecule).

The hammerhead ribozyme is one of the smallest ribozymes derived from self-cleaving RNA molecules. The secondary structure of the ribozyme–substrate complex (Fig. 4) consists of one intramolecular duplex within the ribozyme domain and two intermolecular duplexes between the ribozyme and substrate. Because the sequence of the intermolecular duplexes can vary in any manner that maintains base pairing, the hammerhead, and other small ribozymes can be engineered to bind and cleave new RNA targets such as messenger RNAs or viral RNAs. They have proven useful as tools for inhibiting expression of specific genes, and some have reached clinical trials as antiviral therapeutics.

Three of the small ribozymes, the hammerhead, hairpin, and hepatitis delta ribozymes, have been characterized by X-ray crystallography. For two of them, the hairpin ribozyme and hepatitis delta ribozyme, the crystal structures seem to represent active conformations. The ribozymes fold into complex three-dimensional structures that force the cleavable bond into an orientation favorable for catalysis, and they position functional groups of important nucleotides in the ribozyme close enough to help facilitate the cleavage reaction.

6 Ribonucleoprotein Enzymes: The Ribosome is a Ribozyme!

Many cellular RNAs are bound to specific proteins, forming ribonucleoprotein (RNP) complexes. These RNPs are responsible for catalyzing a number of critical cellular functions. For instance, ribosomes are RNP complexes responsible for synthesizing proteins from a messenger RNA template. The spliceosome is an RNP complex that splices intron-containing messenger RNA precursors within the eukaryotic nucleus.

The bacterial ribonuclease P (RNase P) is a particularly interesting and well-studied example of an RNP enzyme. This enzyme cleaves the 5′ extension from precursor transfer RNA molecules (Fig. 5), yielding the mature 5′ end of all the different tRNAs in the cell. RNase P consists of a fairly large RNA subunit plus a small basic protein. *In vitro*, in high salt, the RNA subunit can carry out the correct cleavage reaction within the protein.

Investigators have gained much insight into the structure of the RNase P RNA through phylogenetic analysis – that is, through comparison of the nucleotide sequences of this RNA isolated from many different bacterial species. Many sequence positions vary between species, but often

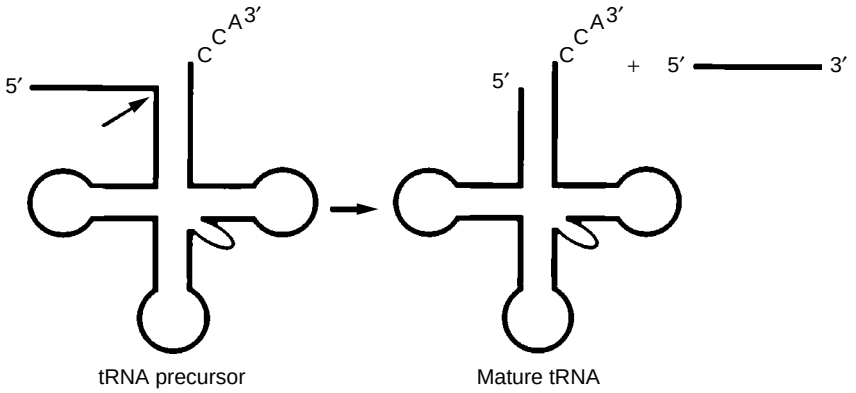

Fig. 5 Reaction catalyzed by Rnase P. Newly transcribed tRNA precursors contain 5′ extensions. Rnase P cleaves these extensions from all tRNA precursors, generating the mature 5′ end for the tRNA molecules.

they "covary," as necessary to maintain the base pairing of a stem. Other sequence positions essential for activity may be conserved in many different species. Such phylogenetic comparisons of RNase P RNA have yielded a "consensus" sequence (Fig. 6), which shows the minimal essential secondary structure of the RNA plus conserved sequences that participate in tertiary interactions required for enzymatic activity.

The story of the RNA-mediated activity of the ribosome is a finding particularly exciting for biological scientists. Ribosomes consist of two subunits, "large" and "small," each containing at least one long RNA molecule and scores of proteins. The large ribosomal subunit was known to carry peptidyl transferase, the activity that transfers amino acids from their transfer RNA carriers to the growing peptide chain in protein synthesis. Recently, the large ribosomal subunit was crystallized, and the structure of the enormous complex was characterized by X-ray diffraction to very fine resolution. The region shown to be involved in peptidyl transferase activity consists entirely of RNA, with no ribosomal proteins located close enough to be participants. This and other results confirm that the chemical step of protein synthesis is catalyzed by RNA: in other words, the ribosome is a ribozyme. This finding reinforces the possibility that the early stages in the evolution of life involved an "RNA world," in which RNA served both to carry genetic information and to catalyze many of the chemical reactions essential for a self-replicating system.

7 *In Vitro* Selection

A recently developed set of methods provides powerful tools for analyzing the structure–function relationships of existing ribozymes and also makes it possible to create and "evolve" new ribozymes that carry out desired chemical reactions. These methods are analogous to the procedures of genetic selection, but they can be carried out on RNA without the need for a host cell.

The first step in *in vitro* selection (Fig. 7) is the production of a large pool, or

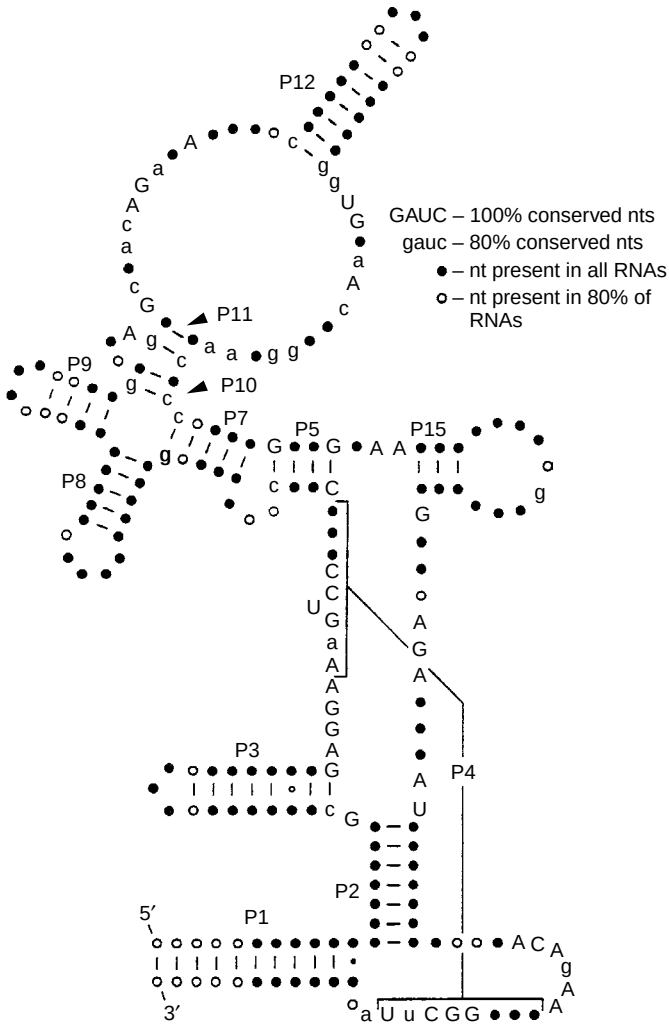

Fig. 6 Phylogenetic minimum bacterial consensus Rnase P RNA: helices are labeled (P1, P2 etc.); base pairs in P2 are shown as brackets and lines. Only the structural elements present in all known bacterial RNAse P RNAs are included in the minimum structure. Universally conserved nucleotides are in uppercase. Nucleotides not conserved in identity but present in all sequences are indicated by solid circles; those present in at least 80% of sequences, but absent in at least one, are indicated by open circles.

population, or RNA molecules with a region of randomized sequence. If the molecules are chemically synthesized, a position can be randomized by providing a mixture of all four nucleotide precursors at synthesis instead of one particular nucleotide. Then this population of molecules is provided with the opportunity to catalyze a chemical reaction. The reaction is designed to permit

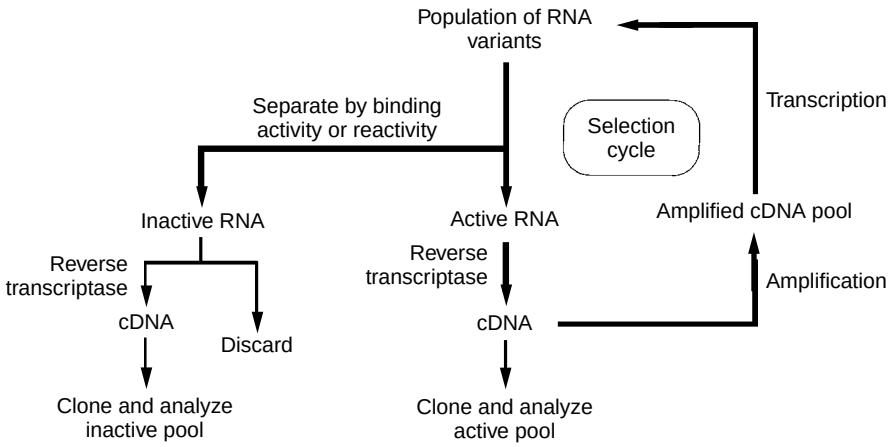

Fig. 7 *In vitro* selection of RNA molecules from a pool of variants. A population of RNA sequence variants is separated into active and inactive molecules on the basis of binding or catalytic activity. Active molecules are copied with reverse transcripts to generate double-stranded cDNA. The cDNA is then amplified, usually through use of the polymerase chain reaction (PCR), making a large pool of the subset of selected sequences. This pool can then be sampled and cloned for analysis or transcribed into a new "generation" of RNA molecules and subjected to further rounds of selection.

the few "active" molecules that succeed to be in some way separated from the large number that fail. For instance, an RNA molecule that ligates, or joints itself to another RNA, can be separated from inactive molecules by gel electrophoresis because it is now larger than the inactive molecules.

Once the few active RNA molecules have been selected, they can be copied into DNA by the enzyme reverse transcriptase. This complementary DNA (cDNA) can then be amplified manyfold using the polymerase chain reaction (PCR). The amplified cDNA can be used to generate RNA by transcription. Thus, the RNA can be selected for a desired characteristic and then replicated, just as in a genetic selection procedure.

The pool of amplified cDNA sequences can be sampled, cloned, and analyzed to determine the sequences of active molecules. It can also be transcribed to produce a new "generation" of RNAs. These molecules can then be sent through the selection procedure again, enriching even more for active RNA sequences (Fig. 7). Multiple rounds of this *in vitro* selection can result in RNA populations containing only a few sequences best capable of catalyzing the desired reaction. This approach has been used to generate novel engineered ribozymes capable of catalyzing many different biochemical reactions not known to be carried out by RNA in modern biological systems.

See also Alternatively Spliced Genes; Nucleic Acid Packaging of RNA Viruses; Protein and Nucleic Acid Enzymes; RNA Interference; RNA Secondary Structures; RNA Three-Dimensional Structures, Computer Modeling of.

Bibliography

Books and Reviews

Bagheri, S., Kashani-Sabet, M. (2004) Ribozymes in the age of molecular therapeutics, *Curr. Mol. Med.* **4**, 489–506.

Cech, T.R. (2002) Ribozymes, the first 20 years, *Biochem. Soc. Trans.* **30**, 1162–1166.

Joyce, G.F. (2004) Directed evolution of nucleic acid enzymes, *Annu. Rev. Biochem.* **73**, 791–836.

Orgel, L.E. (2004) Prebiotic chemistry and the origin of the RNA world, *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123.

Steitz, T.A., Moore, P.B. (2003) RNA, the first macromolecular catalyst: the ribosome is a ribozyme, *Trends Biochem. Sci.* **28**, 411–418.

Primary Literature

Ban, N., Nissen, P., Hansen, J., Moore, P.B., Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution, *Science* **289**, 905–920.

Bartel, D.P., Szostak, J.W. (1993) Isolation of new ribozymes from a large pool of random sequences [see comment], *Science* **261**, 1411–1418.

Beaudry, A.A., Joyce, G.F. (1992) Directed evolution of an RNA enzyme [see comments], *Science* **257**, 635–641.

Been, M.D., Perrotta, A.T. (1991) Group I intron self-splicing with adenosine: evidence for a single nucleoside-binding site, *Science* **252**, 434–437.

Bergeron, L.J., Perreault, J.P. (2005) Target-dependent on/off switch increases ribozyme fidelity, *Nucleic Acids Res.* **33**, 1240–1248.

Brown, T.S., Chadalavada, D.M., Bevilacqua, P.C. (2004) Design of a highly reactive HDV ribozyme sequence uncovers facilitation of RNA folding by alternative pairings and physiological ionic strength, *J. Mol. Biol.* **341**, 695–712.

Cai, Z., Tinoco, I. Jr. (1996) Solution structure of loop A from the hairpin ribozyme from tobacco ringspot virus satellite, *Biochemistry* **35**, 6026–6036.

Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R., Doudna, J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing, *Science* **273**, 1678–1685.

Collins, R.A., Olive, J.E. (1993) Reaction conditions and kinetics of self-cleavage of a ribozyme derived from Neurospora VS RNA, *Biochemistry* **32**, 2795–2799.

Deschenes, P., Ouellet, J., Perreault, J., Perreault, J.P. (2003) Formation of the P1.1 pseudoknot is critical for both the cleavage activity and substrate specificity of an antigenomic trans-acting hepatitis delta ribozyme, *Nucleic Acids Res.* **31**, 2087–2096.

Ferre-D'Amare, A.R., Doudna, J.A. (2000) Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module, *J. Mol. Biol.* **295**, 541–556.

Garriga, G., Lambowitz, A.M. (1984) RNA splicing in neurospora mitochondria: self-splicing of a mitochondrial intron in vitro, *Cell* **39**, 631–641.

Guerrier-Takada, C., Altman, S. (1984) Catalytic activity of an RNA molecule prepared by transcription in vitro, *Science* **223**, 285–286.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme, *Cell* **35**, 849–857.

Hampel, A., Cowan, J.A. (1997) A unique mechanism for RNA catalysis: the role of metal cofactors in hairpin ribozyme cleavage, *Chem. Biol.* **4**, 513–517.

Koch, J.L., Boulanger, S.C., Dib-Hajj, S.D., Hebbar, S.K., Perlman, P.S. (1992) Group II introns deleted for multiple substructures retain self-splicing activity, *Mol. Cell. Biol.* **12**, 1950–1958.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena, *Cell* **31**, 147–157.

Kuiper, M.T., Lambowitz, A.M. (1988) A novel reverse transcriptase activity associated with mitochondrial plasmids of Neurospora, *Cell* **55**, 693–704.

Lafontaine, D.A., Norman, D.G., Lilley, D.M. (2002) The global structure of the VS ribozyme, *EMBO J.* **21**, 2461–2471.

Luptak, A., Ferre-D'Amare, A.R., Zhou, K., Zilm, K.W., Doudna, J.A. (2001) Direct pK(a) measurement of the active-site cytosine in a

genomic hepatitis delta virus ribozyme, *J. Am. Chem. Soc.* **123**, 8447–8452.

Michel, F., Dujon, B. (1983) Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members, *EMBO J.* **2**, 33–38.

Murray, J.B., Seyhan, A.A., Walter, N.G., Burke, J.M., Scott, W.G. (1998) The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone, *Chem. Biol.* **5**, 587–595.

Nakano, S., Bevilacqua, P.C. (2001) Proton inventory of the genomic HDV ribozyme in Mg(2+)-containing solutions, *J. Am. Chem. Soc.* **123**, 11333–11334.

Nakano, S., Cerrone, A.L., Bevilacqua, P.C. (2003) Mechanistic characterization of the HDV genomic ribozyme: classifying the catalytic and structural metal ion sites within a multichannel reaction mechanism, *Biochemistry* **42**, 2982–2994.

Nissen, P., Hansen, J., Ban, N., Moore, P.B., Steitz, T.A. (2000) The structural basis of ribosome activity in peptide bond synthesis, *Science* **289**, 920–930.

Noah, J.W., Lambowitz, A.M. (2003) Effects of maturase binding and Mg2+ concentration on group II intron RNA folding investigated by UV cross-linking, *Biochemistry* **42**, 12466–12480.

Oyelere, A.K., Kardon, J.R., Strobel, S.A. (2002) pK(a) perturbation in genomic hepatitis delta virus ribozyme catalysis evidenced by nucleotide analogue interference mapping, *Biochemistry* **41**, 3667–3675.

Peebles, C.L., Zhang, M., Perlman, P.S., Franzen, J.S. (1995) Catalytically critical nucleotide in domain 5 of a group II intron, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4422–4426.

Piccirilli, J.A., Vyle, J.S., Caruthers, M.H., Cech, T.R. (1993) Metal ion catalysis in the tetrahymena ribozyme reaction, *Nature* **361**, 85–88.

Pinard, R., Lambert, D., Walter, N.G., Heckman, J.E., Major, F., Burke, J.M. (1999) Structural basis for the guanosine requirement of the hairpin ribozyme, *Biochemistry* **38**, 16035–16039.

Pinard, R., Hampel, K.J., Heckman, J.E., Lambert, D., Chan, P.A., Major, F., Burke, J.M. (2001) Functional involvement of G8 in the hairpin ribozyme cleavage mechanism, *EMBO J.* **20**, 6434–6442.

Pley, H.W., Flaherty, K.M., McKay, D.B. (1994) Three-dimensional structure of a hammerhead ribozyme, *Nature* **372**, 68–74.

Prody, G.A., Bakos, J.T., Buzayan, J.M., Schneider, I.R., Breuning, G. (1986) Autolytic processing of dimeric plant virus satellite RNA, *Science* **321**, 1577–1580.

Rupert, P.B., Ferre-D'Amare, A.R. (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis, *Nature* **410**, 780–786.

Saville, B.J., Collins, R.A. (1990) A site-specific self-cleavage reaction performed by a novel RNA in neurospora mitochondria, *Cell* **61**, 685–696.

Scott, W.G., Finch, J.T., Klug, A. (1995) The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage, *Cell* **81**, 991–1002.

Sood, V.D., Collins, R.A. (2002) Identification of the catalytic subdomain of the VS ribozyme and evidence for remarkable sequence tolerance in the active site loop, *J. Mol. Biol.* **320**, 443–454.

Walter, N.G., Burke, J.M., Millar, D.P. (1999) Stability of hairpin ribozyme tertiary structure is governed by the interdomain junction, *Nat. Struct. Biol.* **6**, 544–549.

Walker, S.C., Avis, J.M., Conn, G.L. (2003) General plasmids for producing RNA in vitro transcripts with homogeneous ends, *Nucleic Acids Res.* **31**, e82.

RNA Chromatography: *see* Nucleic Acid Chromatography

RNA Clinical Applications, Emerging: *see* Pharmacology and Emerging Clinical Application of RNA

RNA Interference

John G. Doench[1] *and Carl Novina*[2]
[1] *Massachusetts Institute of Technology, Cambridge, MA*
[2] *Cancer Immunology and AIDS, Dana-Farber Cancer Institute, and Department of Pathology, Harvard Medical School, Boston, MA*

1 Discovery of RNAi 529

2 Mechanisms and Components of RNAi 530
2.1 Dicer 530
2.2 siRNAs 532
2.3 Asymmetry of siRNAs 533
2.4 RISC 535
2.4.1 The Argonaute Family of Proteins 535
2.4.2 Target Selection and Cleavage 535
2.4.3 Off-target Effects of siRNAs 536
2.5 RNA-dependent RNA Polymerases 536

3 Endogenous Roles for RNAi 537
3.1 Cosuppression 537
3.2 Silencing of Invading Genetic Elements 537
3.3 Transcriptional Silencing 539
3.4 Generation of the Silencing Signal 540

4 MicroRNAs 540
4.1 Dual Discovery of a Large Class of Genes 541
4.2 Mechanism of miRNAs 542
4.3 Biogenesis of miRNAs 543
4.4 Proteins Associated with miRNAs: the miRNP 543
4.5 Overlap and Distinctions between RNAi and Translational Repression 545
4.6 MicroRNAs and their Targets 545

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

5 RNAi as a Technology 546
5.1 Methods of Triggering RNAi 546
5.2 RNAi as a Potential Therapeutic 547

Acknowledgment 548

Bibliography 548
Books and Reviews 548
Primary Literature 548

Keywords

Dicer
An RNase III enzyme responsible for processing siRNAs from long dsRNA, as well as processing mature miRNAs from pre-miRNA hairpins.

microRNAs (miRNA)
Short single strand RNA species that are ~22 nucleotides long and that are encoded by an endogenous class of genes. miRNAs guide the microribonucleoprotein complex (miRNP) to the 3′ untranslated region (UTR) of mRNAs and base pair with imperfect complementarity, resulting in translational repression.

miRNP
Microribonucleoprotein complex; the complex that contains endogenous miRNAs. The miRNP is not as well defined as RISC, either in terms of its protein components or its relevant biochemical activities.

RISC
RNA-induced silencing complex; the complex that contains a single strand of an siRNA, used as a guide to find target mRNAs. RISC catalyzes the cleavage of mRNAs complementary to one strand of the siRNA, resulting in degradation of the mRNA. A main component of RISC in all species is a member of the Argonaute family of proteins.

Short Interfering RNA (siRNA)
An intermediate in the RNAi pathway, processed by Dicer from long double-strand RNA (dsRNA). siRNAs are composed of two ~21 nucleotide RNA strands that are duplexed: there is a region of 19 nucleotide duplex RNA with recessed and phosphorylated 5′ ends and two unpaired nucleotides at the 3′ ends. One strand of the siRNA is incorporated into RISC, guiding it to mRNAs with perfect or near-perfect complementarity, resulting in cleavage of the mRNA.

A revolution has recently swept through biology called *RNA interference*. RNAi serves as an umbrella term, encompassing several phenomena discovered in a diversity of eukaryotes, all of which are mediated by short RNA species. RNAi has been implicated in regulating gene expression by several mechanisms, from transcriptional silencing to mRNA degradation and translational repression. The central role of RNAi processes in gene expression is underscored by its strong conservation across almost all eukaryotes, from plants to fungi to mammals. This review will focus on the discovery of the classical RNAi pathway before considering the molecular details of this pathway. RNAi will then be viewed from a broader perspective, by examining other small RNA directed pathways in a variety of organisms. One pathway in particular will receive ample attention, that of the endogenous microRNA genes, a class of small RNAs that has arisen in multicellular plants and animals. Finally, the adoption of RNAi as a powerful technology will be discussed, including prospects for RNAi-based therapeutic applications.

1 Discovery of RNAi

Of the many remarkable aspects of RNA interference, perhaps the most striking from a historical perspective is that this phenomenon, central to the regulation of gene expression in a wide variety of eukaryotes, evaded the detection of scientists for so long. The initial characterization of this process by Andrew Fire, Craig Mello, and colleagues brought together disparate, and largely unexplained, observations from several model systems and in doing so unlocked an entire field of research that continues to grow at an astounding pace.

As biology marched into the genomic age in the middle of the 1990s, researchers were in the position of having the DNA sequence of genes but no clue as to how these genes might function. Classical genetics (also known as *reverse genetics*) relied on first having a mutant and then working back to find which of the many genes in the genome caused the mutant phenotype. The utility of the approach was beyond doubt, but now, with so much DNA sequence available and more on the way, forward genetic techniques were needed; biologists wanted to start with the DNA sequence of a gene, functionally impair the gene, and then observe the phenotype.

One conceptually simple solution to this problem was antisense technology. This approach attempted to stop gene expression after the DNA had been transcribed into mRNA, but before it was translated into the language of proteins. A researcher would make a nucleic acid with a sequence antisense to the gene of interest, thereby allowing it to base pair with the cognate mRNA. It was assumed that this now double-stranded mRNA would not be competent for translation and the gene would be functionally knocked out. Early experiments in *Caenorhabditis elegans* showed that *in vitro* transcribed antisense RNA, when injected into the worm, would mimic the phenotype of a true genetic knockout. There was a major caveat, however, that was left unexplained

at that time: sense RNA, used as a control because it could not base pair to the mRNA, oftentimes gave the same result!

Fire and Mello had two key insights. First, they noticed that this antisense technique resulted in systemic silencing of the knocked out gene. Even though the worm was injected with the antisense (or sense) RNA in a discreet location, the gene would be silenced essentially throughout the organism. Also, this silencing effect was heritable over several generations. These observations led them to hypothesize that these effects were not a simple antisense effect, which would in all likelihood be local and short lived, but instead that the injected RNA was triggering an endogenous biological response. The second insight was the (educated) guess that the molecular trigger for this process was actually a low level of double-strand RNA (dsRNA) made during the *in vitro* transcription. Fire and Mello reasoned that the RNA polymerase sometimes transcribed the wrong (i.e. sense) strand of the DNA template, and thus the RNA preparations were contaminated with low levels of dsRNA. Experimental proof of this beautifully simple idea was quickly obtained. Annealing sense and antisense RNA before injection resulted in an incredibly potent silencing, and the term *RNA interference* was coined to describe this phenomenon.

RNAi is an evolutionarily ancient process that has been found in all multicellular eukaryotes that have been investigated to date. Versions of RNAi also appear in many single-celled eukaryotes including *Trypanosoma brucei*, *Tetrahymena thermophila*, and the fission yeast *Schizosaccharomyces pombe* (interestingly, there is no evidence for RNAi in the budding yeast *Saccharomyces cerevisiae*). This article focuses on the metazoan RNAi pathway, as defined largely by work done in worm, fly, and mammalian systems, although results from other organisms are discussed when appropriate. Furthermore, RNAi is closely related to a body of research on endogenous small RNAs known as microRNAs (miRNAs), and the biology, implications, and horizons of this field also receive attention. Lastly, while RNAi is a burgeoning field of biology in its own right, RNAi has also become an essential technology. This review covers both the use of RNAi in molecular biology as well as its potential applications in human health.

2 Mechanisms and Components of RNAi

Canonical RNAi (see Fig. 1a) begins when long dsRNA enters the cytoplasm of a cell, where it is processed by the enzyme called *Dicer*. Dicer successively cleaves the dsRNA into smaller fragments known as *short interfering RNA*s (siRNAs). The siRNA is unwound and one of the two strands is incorporated into a multi-subunit protein complex known as the *RNA-induced silencing complex* (RISC). RISC then uses the siRNA as a guide to find and then cleave complementary mRNAs. This section discusses the discovery of these major protein and RNA species and provides further insight into the mechanism of RNAi.

2.1 Dicer

As soon as it was appreciated, in both animal and plant systems, that the long dsRNA which triggered RNAi was processed into siRNAs, the hunt was on for protein responsible for this step. Two characteristics of the siRNA product, the 5′ phosphate and the two unpaired

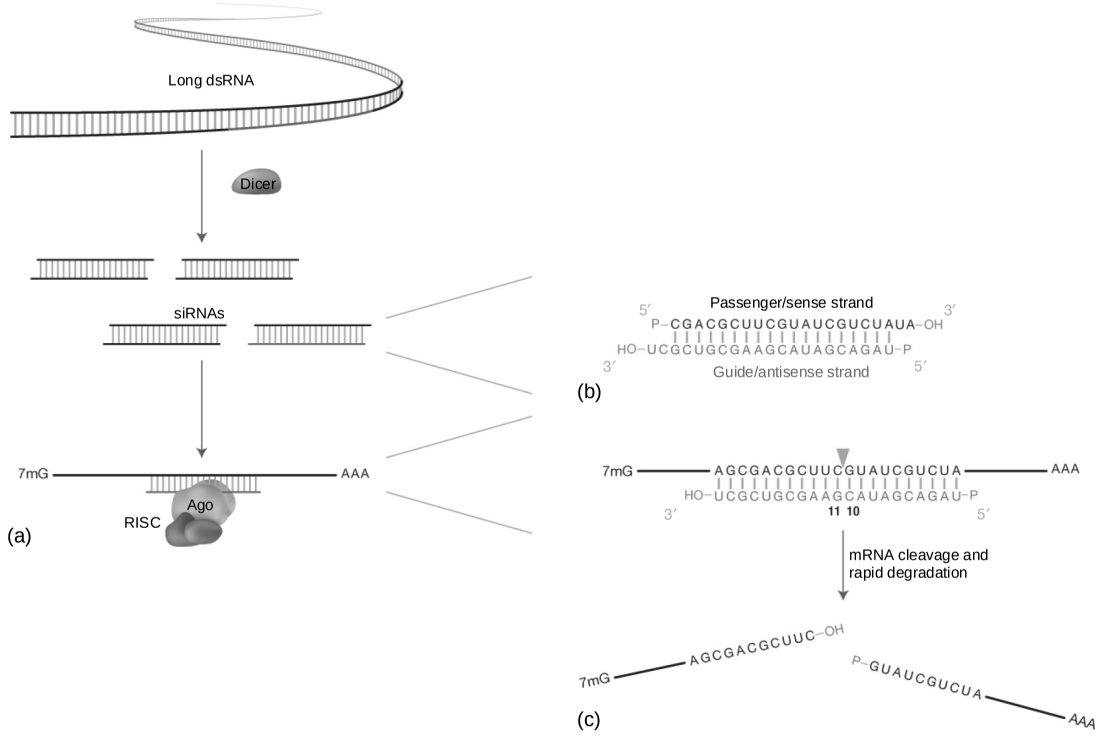

Fig. 1 The RNAi pathway (a) Long dsRNA in the cytoplasm of a cell encounters the enzyme Dicer, which processes the long dsRNA into siRNAs. These siRNAs are unwound and one strand is incorporated into RISC. This guide strand of the siRNA allows RISC to identify target mRNAs and then cleave them, resulting in their degradation. (b) Strand selection of an siRNA. In this example, the strand that will be incorporated into RISC is shown in red, and is designated the guide or antisense strand. This strand is chosen because the first four nucleotides, counting from the 5′ end of the antisense strand of the siRNA, have less thermal stability (A–U rich) than the other strand, which is G–C rich in its first four nucleotides. A helicase activity unwinds this duplex, and hands off one strand to RISC. (c) Target cleavage. The site of the RISC-mediated cleavage is very precise, cleaving between the nucleotides opposite the 10th and 11th bases of the siRNA, counting from the 5′ end of the antisense strand of the siRNA. The RNase H-like cleavage leaves a free hydroxyl on the 5′ region (i.e. the capped end) of the mRNA, and a phosphate on the 3′ half (i.e. the polyadenylated end) (see color plate p. lii).

nucleotides at the 3′ end (see Fig. 1b), suggested that an RNase III enzyme would likely be responsible. A candidate gene approach was undertaken to identify the responsible enzyme, and a bioinformatic search for proteins with a dsRNA binding domain and an RNaseIII domain resulted in a very manageable number of candidates. Affinity-tagged versions of these proteins were expressed and purified. Researchers assayed for the ability to process radiolabeled dsRNA into siRNAs. One of the tagged proteins had this activity, and was named *Dicer*. To confirm this observation, researchers performed an "RNAi–on–RNAi" experiment, to knockdown the expression of Dicer, and thus RNAi activity, via RNAi, with dsRNA homologous to Dicer. While such an approach could not logically be expected to generate a full knockdown of Dicer activity, the attenuation of RNAi activity could still be readily detected, thus proving that Dicer is an essential protein in the RNAi pathway.

A Dicer protein has been found in all organisms that perform RNAi or RNAi-like phenomena. The genomes of *S. Pombe*, *C. elegans*, and mammals reveal only a single Dicer homolog, while Drosophila has two Dicers and *Arabidopsis* has at least four. As might be expected in the latter cases, different Dicers appear to have evolved specialized functions. For example, in Drosophila, Dicer-2 appears to function predominantly in the RNAi pathway, while Dicer-1 is utilized in miRNA biogenesis (see Sect. 4.3). Knockout models reveal that Dicer is essential: in *C. elegans*, homozygous null alleles of Dicer are sterile and survive to adulthood only because of maternal rescue, while a Dicer knockout mouse dies during early embryonic development.

Dicer is necessary and sufficient to cleave long dsRNA into siRNAs, although Dicer interacts with several other proteins *in vivo*. On a molecular level, the best-studied Dicer is mammalian Dicer, which has two RNaseIII domains and a single dsRNA binding motif (dsRBM), as well as a PAZ domain (see Fig. 2). Site-directed mutagenesis studies have made clear the role of each of these domains in the process of siRNA formation. The PAZ domain binds at the free end of the dsRNA, and crystal structures have revealed that the PAZ domain recognizes the two-nucleotide overhang of siRNA-like molecules. Certainly, this domain is also important for recognition during miRNA processing (see Sect. 4.3). Dicer is capable of processing blunt-ended dsRNA, albeit at a decreased efficiency, as would be the case when exogenous dsRNA is used to trigger RNAi (after the first cleavage of a long dsRNA, Dicer leaves a two-nucleotide overhang, thus making subsequent cleavage events more favorable). Binding of the PAZ domain to the end of the dsRNA positions the two RNaseIII domains ~21 nucleotides upstream of the end, with one on each strand of the dsRNA. Using conserved aspartic acid and glutamic acid residues, each RNaseIII domain then makes a single cut in its substrate strand, resulting in the liberation of an siRNA.

2.2 siRNAs

Initial biochemical study of the RNAi pathway was performed in Drosophila embryo lysates. After establishing that dsRNA would lead to destruction of target mRNA in this *in vitro* system, researchers asked what happened to the trigger dsRNA. By labeling their input dsRNA, incubating

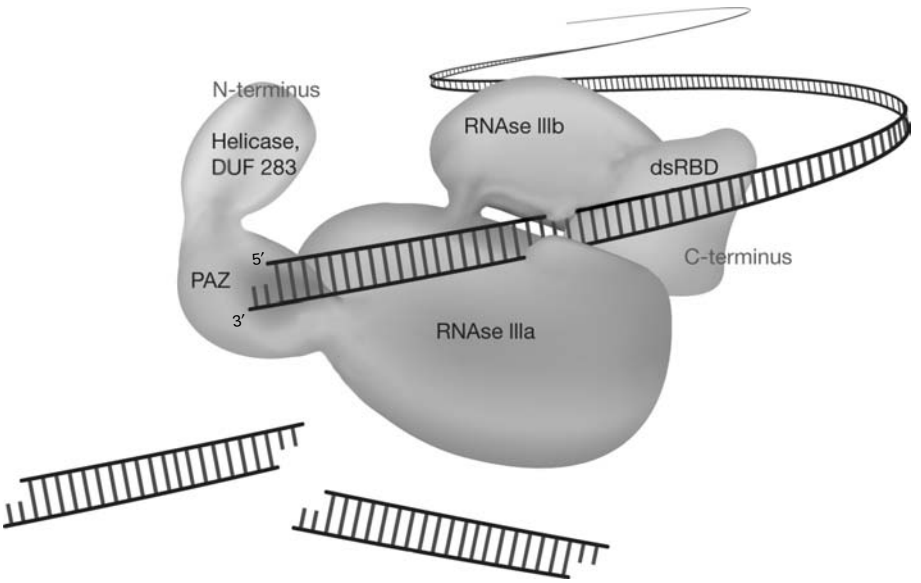

Fig. 2 Recognition and cleavage of dsRNA substrates by Dicer. This schematic illustrates how the different domains of Dicer bind to a dsRNA substrate. The PAZ domain recognizes the two unpaired nucleotides at the 3′ end resulting from a previous Dicer cleavage, in the case of RNAi, or from Drosha cleavage, in the case of miRNA processing. The two RNaseIII domains are then positioned ~21 nucleotides away from the base, and each cleaves one strand of the RNA. From Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E., Filipowicz, W. (2004) Single processing center models for human Dicer and bacterial RNase III, *Cell* **118**, 57–68.

it in the lysate, and then running the products on a polyacrylamide gel, they observed that the long, input dsRNA was processed into a smaller RNA species, of ~21 nucleotides in length (see Fig. 3). This short RNA species was then cut out of the gel to see if it alone could potentiate the degradation of target mRNA, and indeed it could, meaning that this RNA is a true intermediate in the pathway. Chemical characterization revealed that these 21 nucleotide species are double stranded and each strand has a 5′ phosphate. While each strand is 21 nucleotides, the strands are duplexed in a staggered manner, with 19 nucleotides of duplex and two unpaired nucleotides at the 3′ end. This species was originally named short interfering RNA or siRNA, although the literature quickly began interchanging ''short'' and ''small.''

2.3 Asymmetry of siRNAs

A question often posed to RNAi researchers by scientists outside the field attempting to use the technology of RNAi was ''why doesn't my siRNA work?'' Initially, the only hard and fast rule about siRNAs was that they worked, except when they did not. A great deal of unsuccessful siRNA-based experiments were explained when one lab hit upon the hypothesis that perhaps nonfunctional siRNAs were nonfunctional because the wrong strand of the siRNA was being incorporated into RISC. In other words, when researchers design

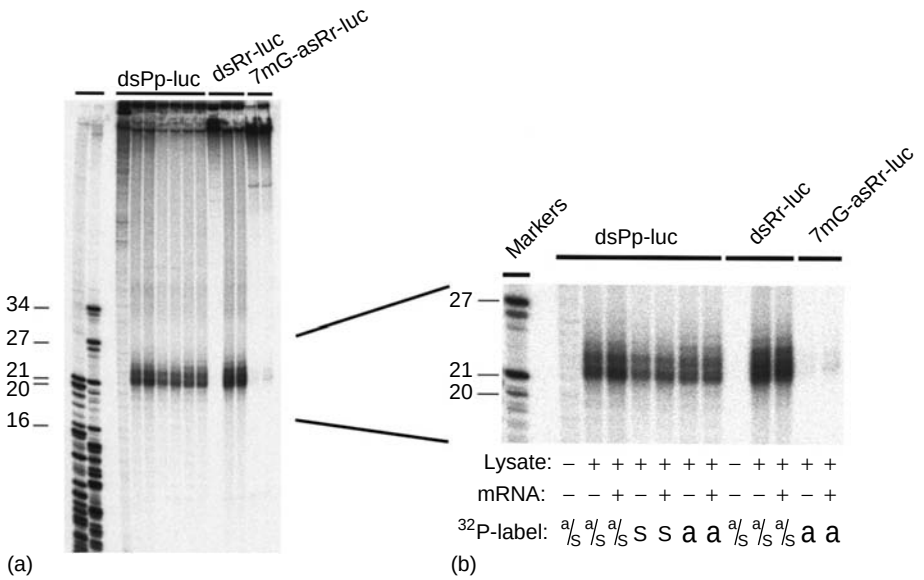

Fig. 3 Identification of siRNAs as intermediates in the RNAi pathway. Radiolabeled dsRNA was incubated in a Drosophila embryo lysate, and the subsequent polyacrylamide gel showed the appearance of a small RNA species. Two different dsRNAs were used, dsPp-luc and dsRr-luc, showing that this was not specific to a given dsRNA sequence. Furthermore, single-stranded RNA (7 mG-asRr-luc) did not yield small RNA species. Part B, a zoom-in of part A, shows that these species are in fact double strand, as either strand of the input dsRNA could be labeled and give rise to the small RNA. The small RNA species appears regardless of the presence of a target mRNA. From Zamore, P.D., Tuschl, T., Sharp, P.A., Bartel, D.P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals, *Cell* **101**, 25–33.

siRNAs to target genes of interest, they know which strand of this geometrically symmetric molecule they want to function as the guide strand of RISC, but RISC itself, of course, does not. Phil Zamore and colleagues hypothesized that a helicase activity must unwind the double-strand siRNA and hand off one strand to RISC, and that perhaps the helicase has a preferred end of the molecule from which to unwind.

Examination of siRNAs used in the lab revealed that functional siRNAs tended to have an A or a U as the first nucleotide of the guide strand, while nonfunctional siRNAs tended to start with a G or C. Thermodynamically, an A–U base pair has less stability than a G–C base pair, and thus the A–U base-paired end would be easier for the helicase to unwind. The necessary experiments were performed to confirm this hypothesis, showing that the thermodynamic stability of the first four nucleotides on each side were very predictive of which strand would be incorporated into RISC (see Fig. 1b). These data are supported by experiments analyzing hundreds of functional and nonfunctional siRNAs synthesized by a leading industrial provider of siRNAs. This strand preference should be followed in designing siRNAs, but there are certainly

more criteria, both known and unknown, that further contribute to the functionality of siRNAs.

2.4 RISC

RISC is a multi-subunit complex that carries out the final step in the RNAi pathway, cleavage of the target mRNA. This definition is rather strict, as there are almost certainly many RISC-like complexes in cells that are not exclusively involved in canonical RNAi. For example, miRNAs involved in translational repression associate with a related complex known simply as the *microribonucleoprotein* complex (miRNP) (see Sect. 4.4). Although different organisms and purification schemes have yielded different accessory proteins, one constant is the central role of an Argonaute family member.

2.4.1 The Argonaute Family of Proteins

Argonaute proteins are defined by both a PAZ domain (see Sect. 2.1) and a PIWI domain, which forms an RNase H-like fold. Argonaute proteins are further divided into two subfamilies, the Piwi subfamily and the Ago subfamily. Members of the Piwi subfamily play a role in silencing phenomena. For example, the Drosophila protein piwi is needed for both transcriptional and posttranscriptional gene silencing (see Sect. 3.4). The Ago subfamily has been directly implicated in canonical RNAi. In Drosophila, Ago-2 was quickly identified as a core component of RISC, and biochemical purification from human cells confirmed that human Ago-2 (originally named *eIF2c2*) is the main component responsible for siRNA-directed mRNA cleavage. Interestingly, the human genome has four well-defined Ago proteins, and all four bind to siRNAs and miRNAs, yet only Ago-2 is capable of cleaving mRNA targets. Certainly, some if not all of the Ago proteins are involved in the miRNA pathway, and it will be interesting to see what different roles the Ago proteins play in small RNA silencing pathways. Although Argonaute proteins are core components of RISC, they appear to work at many steps in the RNAi pathway. One of the original RNAi deficient worms, *rde-1*, encodes for an Argonaute protein that works upstream of the generation of siRNAs, while two other Argonaute proteins in worms, *alg-1* and *alg-2*, are important for miRNA biogenesis.

2.4.2 Target Selection and Cleavage

The details of siRNA-directed mRNA cleavage have been established in two biochemically tractable systems, mammalian and Drosophila extracts. Experiments have shown that the siRNA guides RISC to the mRNA and cleaves the mRNA between the bases opposite bases 10 and 11 of the siRNA, counting from the 5′ end. This cleavage leaves a 3′ hydroxyl on the 5′ end of the cleaved mRNA and a 5′ phosphate on the 3′ end (see Fig. 1c). *In vitro* analysis of this reaction demonstrated that this cleavage is catalytic and conforms to Michaelis–Menten kinetics, as a single RISC molecule can cleave multiple mRNAs.

As might be inferred from the cleavage products, RISC does not mediate cleavage via the mechanisms employed by common RNases such as RNase A, which produce a 5′ hydroxyl and a 2′3′ cyclic phosphate. For quite some time, researchers pursued the elusive "Slicer" protein that catalyzed the mRNA cleavage, but the crystal structure of an Argonaute protein revealed that Ago itself is responsible for the cleavage. The PIWI domain of Ago adopts an RNase H fold, an observation that is consistent

with the known cleavage products of RISC. As mentioned, only mammalian Ago-2 is capable of catalyzing mRNA cleavage, yet the other three mammalian Ago family members have a conserved PIWI domain with the appropriate residues at the catalytic center of the enzyme. Ago-1, 3, and 4, however, have mutated glutamine and histidine residues outside the proposed active site that eliminate the ability of these three Ago proteins to cleave mRNA. It is not clear at this point how these two mutations disrupt the cleavage activity, but it has been proposed that these two residues coordinate metal ions needed for catalysis.

2.4.3 Off-target Effects of siRNAs

For RNAi to be useful to an organism, the signal that it generates must be specific for the intended RNA. Likewise, specificity is an important parameter for RNAi to be a successful tool, to ensure that any observations are due to knockdown of the gene of interest rather than other genes. Biochemical studies in various *in vitro* systems have shown that the guide strand of the siRNA does not need to base pair throughout the target, but that more-limited complementarity can produce mRNA cleavage, albeit at a greatly reduced rate. Importantly, it is not clear if this limited cleavage has a phenotypic effect in live cells or whole organisms. To assess the effects of siRNAs in a genome-wide context, extensive microarray analyses have been performed. These results show that, on average, a given siRNA will modestly downregulate approximately 50 unintended genes. Most of the genes downregulated do in fact have some stretches of complementarity to the siRNA, especially in the 5′ region, and thus are likely to be cleaved via an RNAi-like mechanism.

This problem of less-than-perfect specificity has been addressed in several ways. First, one potential source of off-target effects is the entry of the wrong strand of the siRNA into RISC. With a better understanding of siRNA asymmetry (see Sect. 2.3), this problem can now be avoided. Second, siRNAs of different sequence can be used to confirm that a phenotype is due to knockdown of the gene of interest rather than an off-target effect; because the siRNAs have different sequence, the off-target profile would be different. It is also important to bear in mind that RNAi is a natural process, and that evolution has selected for its presence. Therefore, the usefulness of RNAi to the host organisms throughout evolution strongly outweighed the potentially detrimental effects of silencing unintended genes.

2.5 RNA-dependent RNA Polymerases

A significant difference between RNAi in plants, worms, and fungi from that in flies and mammals is the requirement for an RNA-dependent RNA polymerase (RdRP) activity in the former. RdRPs amplify the RNAi signal, and RdRP proteins are required for productive RNAi in organisms that have them. siRNAs generated from Dicer cleavage can go down either of two paths. First, the guide strand of the siRNA can enter RISC and perform mRNA target cleavage. Alternatively, the guide strand of an siRNA can function as a primer, binding to the mRNA target and allowing polymerization of an RNA strand antisense to the mRNA. This now-dsRNA can serve as a substrate for Dicer, producing a second generation of siRNAs, thereby amplifying the RNAi effect.

RdRP activity was formally proven to be a part of the RNAi pathway in an experiment in *C. elegans* utilizing a *pha-4*:GFP fusion transgene, such that the mRNA would code for *pha-4* in its 5′ half and GFP in its 3′ half. As expected, dsRNA homologous only to GFP silenced this transgene. This same dsRNA also silenced endogenous *pha-4*, implying that siRNAs generated against the GFP were then extended by an RdRP activity to make dsRNA homologous to *pha-4*. These siRNAs could now target the endogenously encoded *pha-4* mRNAs, thus silencing *pha-4* in trans and leading to the term *transitive RNAi*. Importantly, if the experiment were reversed and *pha-4* were the 3′ half of the fused transgene, dsRNA against GFP would not cause silencing of endogenous *pha-4*, implying a directionality to the RdRP activity.

Bioinformatics has shown that no RdRP homologs exist in either the mammalian or Drosophila genomes. Biochemical analysis has shown that siRNAs with modifications on their 3′ ends, such that the 3′ end could not serve as a primer for extending an RNA transcript, were still able to support RNAi. Thus, it can be concluded that RdRP is not necessary for robust RNAi in these organisms. Why this RdRP activity was apparently lost during insect and vertebrate development is unclear.

3 Endogenous Roles for RNAi

While one can never know for sure the selective pressures guiding evolution, it is unlikely that RNAi evolved as a means for researchers to knockout their favorite genes. As work in plants and *C. elegans* has made clear, RNAi serves as a defense mechanism against invading genetic elements. Furthermore, the RNAi machinery is used not only as an immediate posttranscriptional response, but also silences genes for the long term, on the transcriptional level.

3.1 Cosuppression

In plant systems, it had been known for some time that transgenes introduced into plants with the purpose of overexpressing a gene would, sometimes, actually produce a knockdown of both the transgene and any endogenous copies (see Fig. 4a). Initially termed *cosuppression*, a clear mechanistic understanding of this process was lacking. A Northern blot provided a key piece of data, showing the appearance of 20 to 25 nucleotide RNAs homologous to the now-silenced gene, while these species were absent in the plants that were not suppressing gene expression (see Fig. 4b). This seminal observation made clear that an RNAi-like process was occurring in cosuppression. Cosuppression is also observed in *C. elegans*, Drosophila, and Neurospora (where it is termed *quelling*) and in all cases, RNAi proteins and small RNAs have been implicated in the process.

3.2 Silencing of Invading Genetic Elements

The genomes of higher organisms are littered with potentially deleterious genetic elements, such as transposons, LINEs, and SINEs, and RNAi serves as a defense mechanism to keep them in check. Early screens for worms defective in the RNAi pathway showed a connection between RNAi and transposon silencing, as some worms that had already been identified as mutant for silencing transposons (designated as *mut* for mutator, as the decreased ability to silence transposons resulted in

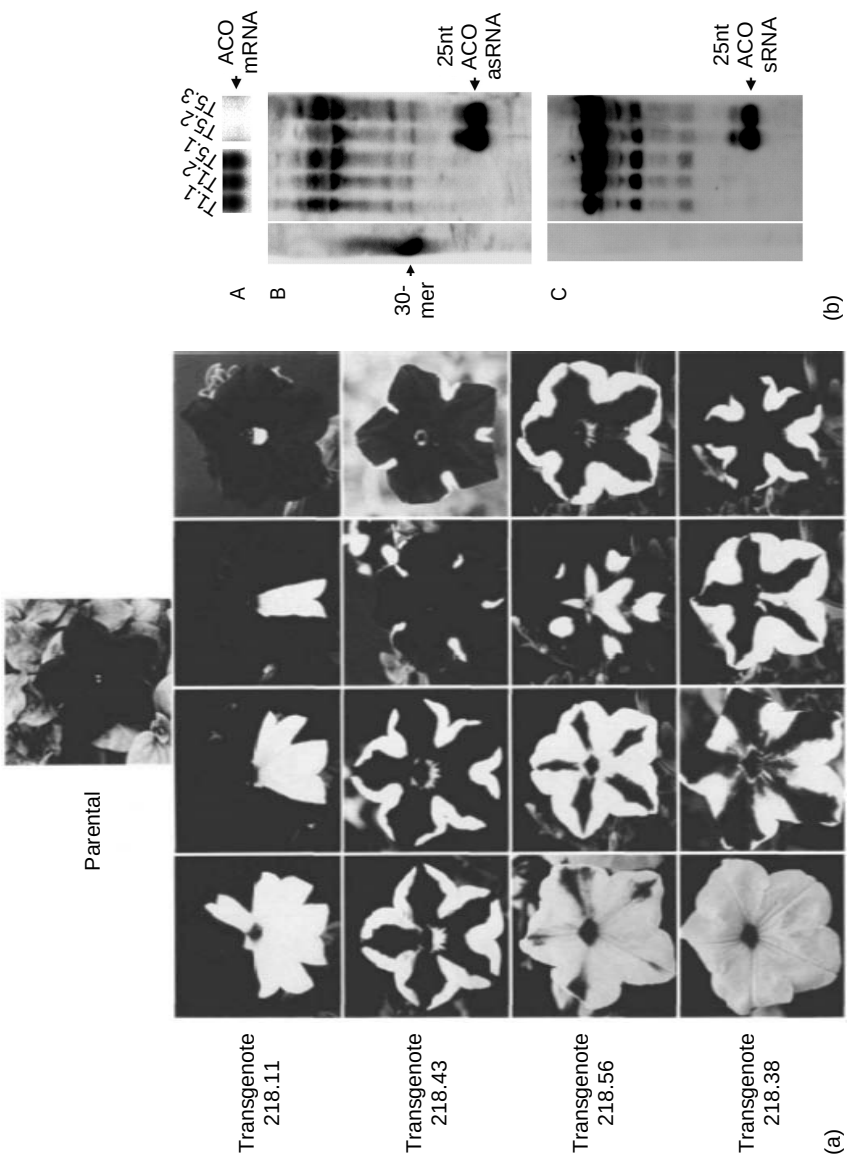
Parental
Transgenote 218.11
Transgenote 218.43
Transgenote 218.56
Transgenote 218.38
(a)
A
T1.1
T1.2
T5.1
T5.2
T5.3
ACO mRNA
B
30-mer
25nt ACO asRNA
C
25nt ACO sRNA
(b)

heightened mutation frequency in these worms) were also deficient in RNAi. Likewise, some of the worms identified in a screen for RNAi deficiency (*rde* mutants) also showed mutator phenotypes. An interesting example of RNAi used to regulate an endogenous gene occurs in Drosophila, where the *Stellate* locus needs to be silenced in order for proper meiosis. A locus known to suppress *Stellate*, Su(Ste), is actually a series of repeats that triggers the RNAi pathway by producing both sense and antisense RNA, which are homologous to *Stellate* and silence it via the RNAi pathway. An Argonaute family member, *aubergine*, is required for this silencing as is *Spindle-E*, an RNA helicase.

In plants, it was quickly appreciated that RNAi constituted an immune system against viruses. When a plant is infected with a virus, it uses the RNAi machinery first to silence the virus in the infected cells. Plants also mount a second response, sending a signal to other cells of the plant, a phenomenon known as *spreading*. This signal renders those other cells immune to a second attack by the same virus. As would be expected in the evolutionary arms race between pathogen and host, some viruses have evolved mechanisms to defend against the endogenous RNAi response. For example, tombusvirus encodes the P19 protein that binds tightly to siRNAs, thus titrating them out of the cell and allowing viral replication.

3.3 Transcriptional Silencing

RNAi was first recognized as a posttranscriptional phenomenon, but even from the beginning, there were many tantalizing hints of a transcriptional arm of the general term that is RNAi, both via DNA and histone modifications. The results were most clear in the plant field. Researchers showed that a plant would trigger canonical RNAi if dsRNA homologous to the coding sequence of a gene was introduced. dsRNA designed against the promoter region of a gene, however, also resulted in silencing of the gene. Since the dsRNA was not homologous to an expressed RNA, something distinct from RNAi was occurring. Further experimentation showed that the DNA homologous to the dsRNA is methylated, a common way for many organisms to silence gene expression at the level of transcription. Further research has

Fig. 4 Cosuppression in plants. (a) Initial observation of cosuppression phenomena. The parental strain of petunia is purple due to the expression of endogenous chalcone synthase. When transgenic lines were made to express more chalcone synthase, many of the plants did not become more purple as would be expected, but in fact lost chalcone synthase expression, resulting in a white color. From Napoli, C., Lemieux, C., Jorgensen, R. (1990). Introduction of a chimeric chalcone synthase gene into petunia results in reversible cosuppression of homologous genes in trans. *Plant Cell* **2**, 279–289. (b) Cosuppression and disappearance of mRNA correlates with appearance of small RNAs. In a similar situation as that shown in 4a, a transgene with the 1-aminocyclopropane-1-carboxylate oxidase (ACO) gene, an endogenous tomato gene, was transformed into tomatoes. Part A is a Northern blot for the ACO mRNA, and 3 of the 5 transgenic plants still express ACO, while two have now silenced ACO. Part B shows the appearance of small RNAs antisense to ACO appearing in the silenced strains but not the ACO-expressing strains. Part C shows that these small RNA species are double stranded, as they also hybridize to the antisense transcript of ACO. From Hamilton, A.J., Baulcombe, D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants, *Science* **286**, 950–952.

indicated that the RNAi pathway can also alter chromatin structure in plants, leading to the methylation of histone 3 lysine 9 (H3K9) and thus the condensation and silencing of those regions of DNA.

The fission yeast *S. pombe* has one homolog each of three RNAi proteins: Dicer, Argonaute, and RdRP. Knockout strains of each individual protein revealed the same phenotype, a defect in centromere definition due to inappropriate transcription. In this organism, the RNAi machinery generates a small RNA species that is used to guide the RITS complex (RNA-induced initiation of transcriptional gene silencing) to modify chromatin structure. RITS recruits the histone methyltransferase Clr4 to methylate H3K9, which then recruits Swi6, the yeast ortholog of metazoan HP1. This type of silencing appears important in defining both the centromeric region and the mating-type locus.

Plants, flies, and budding yeast have provided the three best-characterized examples of transcriptional silencing directed by short RNAs to date, but new research continues to point to transcriptional silencing in other organisms as well. For example, researchers have shown that siRNAs transfected into mammalian cells targeted against promoter sequences can indeed induce transcriptional silencing. While this observation has not yet been tied to an endogenous process, the role of methylation in mammalian biology is well established, and it is tempting to speculate that small RNAs play a role in providing specificity to transcriptional silencing phenomena.

3.4 Generation of the Silencing Signal

Given that a dsRNA trigger is needed to provoke RNAi, a key question centers on how dsRNA forms from an endogenous source to provoke any of the phenomena discussed earlier. One possible model hypothesizes that the transcripts arising from these loci are somehow aberrant, and thus recognized by RdRPs and converted into dsRNA. This explanation is unsatisfyingly vague and cannot account for organisms that do not have an RdRP in their genome. Perhaps, an increase in the copy number of a gene increases the chance of read-through transcription from another locus, thereby potentially leading to antisense transcripts that could form dsRNA. This hypothesis has experimental support, as increasing the copy number of a transgene was shown to provoke both transcriptional and posttranscriptional silencing in Drosophila in a manner dependent on the Argonaute protein piwi. Many of the repetitive elements in the human genome are remnants of retroviruses that contain promoters in their long terminal repeats (LTRs). Indeed, it has recently been appreciated that there is a great deal of antisense transcription in the genome, perhaps serving as a means of quality control, whereby the wrong proportion of sense and antisense would lead to dsRNA formation and thus trigger RNAi.

4 MicroRNAs

It was not long after the discovery of siRNAs in the original Drosophila lysate experiments that an obvious question arose: are there endogenous siRNAs? These experiments led to the discovery of a large class of endogenously encoded small RNAs that were termed *microRNAs* or miRNAs. miRNAs have the interesting distinction of essentially being discovered twice, as work in worms had previously identified a small RNA that remained an

isolated observation for some time. The *lin-4* RNA would turn out, however, to be the founding member of a large class of very interesting genes.

4.1 Dual Discovery of a Large Class of Genes

The cloning approaches used to look for siRNAs selected for these species by two key properties. First, siRNAs are a defined size, so one step of cloning involves size fractionation of RNA on a denaturing polyacrylamide gel, to isolate RNAs of approximately 21 nucleotides. Of course, by running a denaturing gel, any siRNAs in the population are separated into the sense and antisense strands. Second, siRNAs contain a 5′ phosphate and a 3′ hydroxyl group, whereas most RNA degradation products do not. To take advantage of this property, a linker is ligated onto the 3′ end of the RNA population, in an ATP-independent reaction to prevent circularization of the small RNA molecules. A second linker is ligated onto the 5′ end, in a manner dependent on the presence of a 5′ phosphate. The resulting product is amplified by reverse transcription and polymerase chain reaction, then cloned into bacteria and sequenced.

The results of initial cloning experiments were not, as expected, large numbers of RNA sequences that could form perfect complements with two-nucleotide overhangs with other RNA sequences in the population, that is, endogenous siRNAs. Instead, the researchers found sequences that appeared to exist predominately as single-stranded RNA in the cell. Interestingly, when these RNA sequences were found in their genomic context, the transcript that would contain these sequences would invariably form a hairpin structure. Northern blots for these species showed not only the single-stranded ~22 nucleotide form that was cloned but also a band at ~60–70 nucleotides, corresponding to the hairpin structure. As further evidence that these cloned products were "real" and not simply random degradation products, several of the sequences were conserved from worms to flies to mammals. The first cloning efforts yielded several dozen of these small RNA species in each organism, and this new class of RNA species was dubbed microRNAs.

Concurrent with these cloning efforts, researchers in the RNAi field began to appreciate some previous work on very interesting small, endogenous RNAs. In the early 1980s, a screen was performed in *C. elegans* for genes involved in lineage defects (*lin* mutants). During the late 1980s and up through the 1990s, one of these mutants became a decade-long research effort spearheaded by Gary Ruvkun and Victor Ambros. The *lin-4* mutant showed an interesting phenotype, whereby certain cells were retarded, meaning that instead of differentiating during the life of the worm, they continually divided into the same cell type. Another mutant, the *lin-14* mutant, showed the opposite phenotype; a precocious mutant, it skipped over early developmental stages. It was determined that the cause of the *lin-14* mutant phenotype was not a mutation in the coding sequence of the protein, but rather in the 3′ untranslated region (UTR) of the mRNA. Owing to their opposite phenotypes, it was hypothesized that the *lin-4* gene product regulated *lin-14*. Mapping of the *lin-4* mutation was incredibly laborious, and the *lin-4* locus became winnowed down to a smaller and smaller region, making it less and less likely that *lin-4* could encode a protein. Finally, it was determined that *lin-4* did not code for a protein but rather for a

small RNA species. Furthermore, when the sequence of the ~22 nucleotide *lin-4* RNA was determined, it was immediately noticed that it shared a great deal of complementarity to several regulatory regions in the *lin-14* 3′ UTR. Further experimentation proved the importance of this RNA:RNA interaction, and *lin-4* was called a *small temporal RNA* (stRNA) for its important role in proper developmental timing.

For almost a decade, *lin-4* remained an isolated case, a strange "worm-thing." In the late 1990s, however, when RNAi was discovered, this "other" small RNA started to receive more notice, and researchers also found another example in the worm of a small RNA gene controlling development, *let-7*. To complete the connection to RNAi, genetic, and biochemical studies implicated Dicer in processing the hairpin precursor to the ~22 nucleotide mature RNA, and several Argonaute proteins were also needed for their biogenesis and function. Not surprisingly, cloning efforts from *C. elegans* for siRNAs identified both *let-7* and *lin-4*, and thus *lin-4* became the founding member of this very exciting new class of genes, the microRNAs.

Initial efforts resulted in a few dozen miRNAs identified in worms, flies, and humans, and the original results are revealing. Some miRNAs are highly conserved across evolution, while other are specific to a given species. Furthermore, within a species, it is clear that some miRNAs are related, as they have similar sequences, especially in the 5′ region of the miRNA. Like *lin-4*, many miRNAs have unique expression patterns, appearing at certain times in development or in certain tissues. As cloning efforts have expanded and computational methods are used to predict miRNAs, more miRNAs have been discovered, and there appear to be 200 to 1000 miRNAs in the human genome. Given that there are 25 000 to 30 000 genes in the human genome, miRNAs are indeed a large class of genes.

4.2 Mechanism of miRNAs

Almost all that is known about the function of miRNAs comes from studies performed on the *lin-4*:*lin-14* interaction. The *lin-4*-mediated repression of *lin-14*, unlike RNAi, does not occur via targeting the *lin-14* mRNA for cleavage and degradation. Northern blot analysis shows that *lin-14* levels remain the same when *lin-4* expression increases during the L1 to L2 developmental transition. There are central bulges in all of the several *lin-4*:*lin-14* interactions, and RNAi requires essentially perfect duplexes for full cleavage activity. Because mRNA levels do not appear to be changed, this is not a posttranscriptional phenomenon like RNAi, but instead has been dubbed "translational repression." It is possible that *lin-4* does not affect the process of translation *per se*, but instead destabilizes the newly made protein; nevertheless, this type of repression is currently known as *translational repression*.

Biochemical analysis of the *lin-4*:*lin-14* interaction revealed that *lin-4* does not cause *lin-14* to lose ribosomes, as the mRNA remains in the polysome fraction of a sucrose gradient. This result is interpreted as indicating that translation initiates properly on the repressed mRNA; otherwise, the *lin-14* mRNA would be associated with fewer ribosomes, and thus would fractionate differently. This same result has been seen with the *lin-28* mRNA, another target of *lin-4*. These are but two studies, and both were performed in *C. elegans*. Future work in this system and in other organisms will be needed to gain

a deeper understanding of how miRNAs actually lead to a decrease in steady state protein levels. Indeed, an immediate alternative to this pathway is found in plants. Plant miRNAs, unlike metazoan miRNAs, oftentimes have extensive complementarity to known mRNAs, and function in an RNAi-like fashion, cleaving the target mRNA. And while most animal miRNAs have limited complementarity to any mRNA, miR-196 bears almost exact complementarity to the *HOXB8* gene, which appears to be cleaved in an RNAi-like fashion. Clearly, the world of miRNAs and siRNAs are overlapping (see Sect. 4.4).

4.3 Biogenesis of miRNAs

MicroRNAs are found in diverse locations in the genome (see Fig. 5). Some are entirely on their own, whereas others are grouped in clusters, probably to allow coordinate expression. Some miRNAs are found far away from any known protein-coding genes, while a good number (~25% in humans) are actually found within introns. miRNAs are not expressed as 22 nucleotide transcripts but rather as very long primary transcripts (pri-miRNA), and these appear to be transcribed by RNA polymerase II, the same polymerase that transcribes protein-coding mRNAs. Another RNase III enzyme, Drosha, recognizes the hairpin that contains the miRNA and cuts it out of the primary transcript, to produce a 60 to 70 nucleotide hairpin known as the precursor-miRNA (pre-miRNA). Like Dicer, the Drosha cut generates the telltale 5′ phosphate and two-nucleotide 3′ overhang. This structure is then recognized by Exportin 5, and the pre-miRNA is exported into the cytoplasm.

In the cytoplasm, the pre-miRNA encounters Dicer. In worms and mammals, it appears that the same Dicer is responsible for both RNAi and miRNA processing, but in Drosophila, there are two Dicers, one for each pathway. The PAZ domain of Dicer recognizes the 5′ phosphate and two unpaired nucleotides, and uses this to position the two RNase III domains, one on each strand of the hairpin stem. Dicer then catalyzes two cleavages, to produce a molecule that bears much similarity to a canonical siRNA. However, as mentioned before, miRNAs are found largely as single-stranded species. Correlative evidence suggests that the same rules underlying siRNA strand selection (see Sect. 2.3) apply to miRNAs, and this siRNA-like intermediate is soon unwound, with one strand being stabilized in an miRNP, whereas the other strand is rapidly degraded.

4.4 Proteins Associated with miRNAs: the miRNP

The ribonucleoprotein complex that contains miRNAs, known simply as the miRNP, is not as well-defined biochemically as RISC, largely because what miRNAs do is largely unknown and thus the miRNP does not yet lend itself to biochemical assays (see Sect. 4.2). It is known that the miRNP contains an Argonaute protein, as an antibody against human Ago-2 immunoprecipitates many miRNAs. Furthermore, overexpressed, tagged versions of human Ago-1, 2, 3, and 4 all associate with endogenous miRNAs, although only Ago-2 is capable of using the miRNA to catalyze RNA cleavage (see Sect. 2.4.2). In Drosophila, Ago-1 associates preferentially with miRNAs while Ago-2 associates with siRNAs, although unlike in mammals, both Ago proteins can catalyze mRNA cleavage.

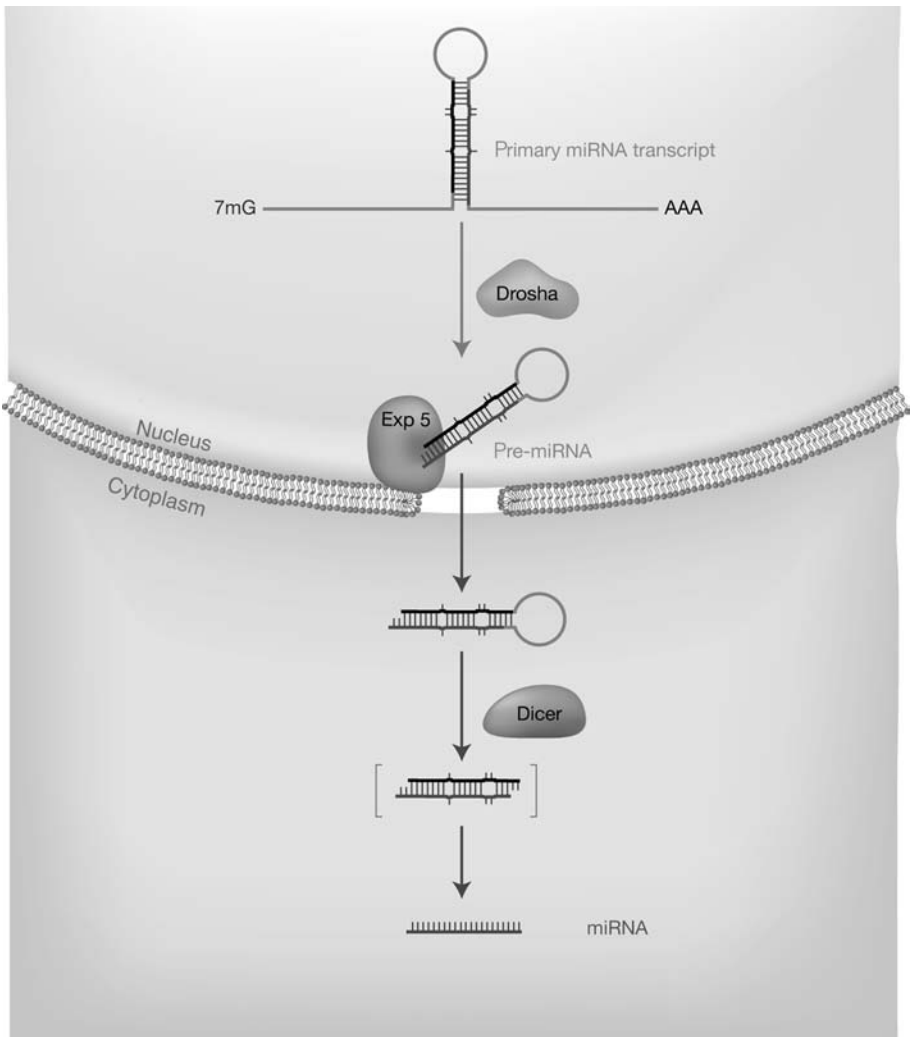

Fig. 5 MicroRNA biogenesis pathway.

In human cells, Gemin 3 and Gemin 4 are found associated with Ago-2 and with miRNAs. Gemin 3, a DEAD-box RNA helicase, and Gemin 4, a novel protein, were previously found associated with the Survival of Motor Neuron (SMN) complex, which is predominantly nuclear and is important for biogenesis of many ribonucleoproteins. SMN is a complex distinct from the miRNP, as several other Gemins associate with SMN that are not found as part of the miRNP; likewise, Ago-2 does not associate with these other Gemins. The association with Ago-2 appears to be a separate function of Gemin 3 and Gemin 4, although their roles are not yet known. Purification of the minimal protein components of RISC activity does not yield Gemin 3 or 4, arguing that they are not necessary for mRNA cleavage

and instead function predominantly in the miRNA pathway.

An important protein in neurobiology has been found associated with Ago-2, the Drosophila homolog of the human Fragile X Mental Retardation Protein (FMRP). Human genetic studies have shown that disruption of this protein leads to one of the most common forms of inherited mental retardation. FMRP is known to associate with mRNAs in neurons, and loss of FMRP alters plasticity at synapses. Identified in a biochemical purification of RISC activity, FMRP is not necessary for mRNA cleavage, and at this point, it is not clear what role FMRP may play in RISC or the miRNP. Many researchers have speculated that FMRP provides additional specificity to miRNA-containing complexes, as FMRP is known to bind to the 3′ UTR of mRNAs and regulate their translation.

4.5 Overlap and Distinctions between RNAi and Translational Repression

Certainly, there is a great deal of overlap between RNAi and translational repression, as target genes are silenced by short RNAs in both cases. However, these silencing pathways are qualitatively different, as the former results in a reduction of mRNA through cleavage, whereas the latter results in stable amounts of mRNA and translational repression occurs by mechanisms that are incompletely understood. The miRNP that contains an miRNA is capable of mRNA cleavage, if the mRNA target has sufficient complementarity. Conversely, exogenous siRNAs are capable of translationally repressing a target mRNA with imperfectly complementary binding sites. Furthermore, Dicer is needed in both pathways for generation of the small RNA that guides the protein complex to the target mRNA.

These pathways are certainly distinct, however. A knockout mouse of Ago-2 is embryonic lethal, but mouse embryonic fibroblasts (MEFs) can be derived. As expected, these MEFs are no longer capable of performing RNAi with a perfectly complementary RNA:RNA interaction, but translational-repression activity is still intact. This result implies that, in mammals, at least one of the other Argonautes (the human genome has four closely related Argonautes) can perform translational repression. The decision between mRNA cleavage and translational repression appears to depend on the extent of complementarity between small RNA and target mRNA, although future work will be needed to see if additional factors also modulate this decision.

4.6 MicroRNAs and their Targets

The discovery of miRNAs as an abundant class of regulatory RNAs immediately prompted investigation into the potential target mRNAs. The first examples of regulation were all found via classical genetic approaches, whereby mutagenesis happened to hit the miRNA (or its binding sites in the 3′ UTR) and thus disrupted its function. In addition to the *C. elegans* examples of *lin-4* regulating *lin-14* and *lin-28*, and *let-7* regulating *lin-41*, several other interactions have been found. For example, the *bantam* locus in Drosophila was simply a mutant locus and there was no clear sense of how this mutation caused the phenotype of premature apoptosis. With the discovery of miRNAs, it was soon appreciated that the *bantam* locus contained an miRNA, and experiments

showed that *bantam* repressed the expression of the proapoptotic gene *hid*. Classical genetic approaches are likely of limited use for finding miRNA targets, however, as an miRNA is a relatively small target for random mutagenesis as compared to a protein-coding gene. Furthermore, evidence suggests that miRNA family members can function redundantly. An alternative approach to this problem is computational biology.

A few general rules have been developed for miRNA:mRNA target interactions, derived both from observations of known targets and from direct experimentation. Of course, these rules are open for revision as more interactions are verified. Furthermore, what follows is a discussion applying to animal miRNAs; plant miRNA targets have been much easier to predict computationally, as they bear significantly more complementarity to their target. These algorithms generally assume that the miRNA targets the 3′ UTR of the mRNA, as is the case for *lin-14*, *lin-28*, *lin-41*, and *hid*. Second, an miRNA will likely bind to the target mRNA multiple times, although coordinate regulation, whereby several different miRNAs bind to a single mRNA, has experimental support, and certainly cannot be discounted. Third, the 5′ region of the miRNA, sometimes referred to as the *seed* region, is a critical determinant of activity, and there is both computational and experimental evidence to support this assumption. Lastly, some computational approaches rely on conservation of a binding site across several species. Certainly this criterion helps eliminate false-positive predictions, but also loses miRNA targets that are specific for a given species.

Several attempts have been made to define miRNA targets in this computational approach, and the results are, perhaps expectedly, quite varied, depending on the exact nature of the algorithm in terms of the relative weight of the different assumptions and parameters discussed earlier. What is clear, however, is that miRNAs have the potential to regulate many genes, a conclusion bolstered by direct experimentation showing that as few as ~8 nucleotides of complementarity between miRNA and target mRNA can give translational-repression activity. Do miRNAs regulate almost every transcript in a cell, fine-tuning the expression of each mRNA, serving as "micromanagers" of gene expression? Or perhaps there are as-yet-unknown specificity determinants, such as protein components of the miRNP that help guide miRNAs to a subset of mRNAs.

5 RNAi as a Technology

The ability to significantly reduce the expression of theoretically any gene, in a wide variety of organisms, has made RNAi a powerful component of the biologists' toolkit. This section will examine how researchers use RNAi as a tool and the promise of RNAi as a potential therapeutic.

5.1 Methods of Triggering RNAi

Depending on the organism under investigation, RNAi is triggered via different means. In *C. elegans,* the first demonstration of RNAi relied on injection of *in vitro* transcribed dsRNA into the worm, a somewhat technically demanding process that does not readily lend itself to high-throughput approaches. Fortunately,

several easier approaches have been developed. One involves a plasmid expressing dsRNA in the bacteria that also serve as the worms' food. When the worm ingests the bacteria, enough of the dsRNA gets into the worm to trigger RNAi. Clearly, this approach does lend itself to high-throughput screens, as a cDNA library can simply be inserted into the plasmid to create an RNAi library. Remarkably, RNAi can also be induced by simply soaking the worm in a solution of dsRNA. Soaking also works in a standard Drosophila-derived S2 cell line, although this approach does not work in mammalian cell lines.

RNAi as a tool has greatly impacted researchers working in mammalian systems. Previously, mammalian biologists had to construct transgenic mice to study the knockout phenotype of a gene of interest, a costly and laborious process. With the advent of RNAi, however, researchers working both in mammalian cell culture systems and in organisms have finally been able to perform reverse genetics. In mammalian systems, the standard way of inducing RNAi is not with long dsRNA, as this molecule would trigger the interferon response, an antiviral mechanism that results in cell death. Instead, siRNAs are used to enter the pathway at an intermediate step, as siRNAs are short enough to avoid the interferon response. These siRNAs are usually chemically synthesized at a relatively inexpensive cost, and are delivered to cells via cationic lipid transfection or electroporation. Maximal gene knockdown on the RNA level occurs within 48 to 72 h after transfection in normally dividing cells, and after that the knockdown effect begins to dilute out as the cell divides.

While siRNA transfection has been widely used, there is certainly an interest in making the RNAi effect permanent, not only in cell culture systems but also in mouse models. For whole-organism studies, the RNA needs to be continuously transcribed from DNA that has been stably integrated into the genome. To do this, researchers have made constructs, which, when transcribed, form a short hairpin RNA (shRNA). This shRNA is processed by Dicer in the cytoplasm to make an siRNA, and the RNAi pathway is triggered. The shRNA constructs lend themselves to large-scale, high-throughput approaches, and several hairpin libraries have been constructed. One such library has been used in a genetic screen to find new members of the p53 pathway, proving that loss-of-function screens utilizing RNAi are not only feasible but also capable of yielding interesting and relevant biological results. shRNA constructs have also been successfully introduced into mouse ES cells and are used to make an RNAi-based knockout mouse.

5.2 RNAi as a Potential Therapeutic

The ability to knockdown a specific gene in somatic cells makes RNAi an attractive approach for therapeutics. Previous nucleic acid–based therapies, such as ribozymes and antisense technology, have largely failed, but RNAi appears to surmount obstacles that have hindered previous methods. Because RNAi is an endogenous process, an entire pathway already dedicated to utilizing siRNAs allows exogenously introduced siRNAs to work at very low concentrations. Especially since drug delivery is a common problem for all molecular therapies, the nanomolar concentrations needed for siRNAs to work (as compared to micromolar concentrations

required for antisense to work) make siRNAs an attractive option. Furthermore, as siRNAs are natural molecules, they greatly reduce toxic side effects, such as is common with chemotherapeutics.

Early work in various model systems showed that RNAi holds the potential for therapeutic use. For example, CCR5 is a coreceptor for HIV, and there is a population of humans who carry a CCR5 allele with a truncation in the protein that confers protection from HIV infection. In cell culture systems, RNAi-mediated knockdown of CCR5 is also able to confer resistance to HIV, thus serving as a proof of principle that RNAi can be used as a successful therapeutic. This idea can also be reversed in the sense of using RNAi not to knockdown a human gene but instead use it to target viral genes that are important for its life cycle, and indeed, experimentation has shown this to be the case for several infectious pathogens, including HIV and influenza. Of course, the theoretical usefulness of RNAi is not limited to viral infections, as any human disease that involves an aberrant or overexpressed gene would be amenable to an RNAi-based treatment.

RNAi is not yet ready to make the leap from bench to bedside, but progress is ongoing in this arena. Several companies are pursuing RNAi-based therapeutics, and mouse models have shown preliminary but intriguing results suggesting that, with time, RNAi will be a powerful means of treating human disease. One recently published example shows that an siRNA chemically modified with a cholesterol moiety can be injected intravenously into mice to silence apoliprotein B. This exciting result illustrates that problems relating to stability, bioavailability, and cellular transport are beginning to be addressed.

Acknowledgment

We wish to thank Helen Cargill for her assistance in preparation of the figures.

See also Genetics, Molecular Basis of; Pharmacology and Emerging Clinical Application of RNA; Ribozymes; RNA Methodologies.

Bibliography

Books and Reviews

Ambros, V. (2004) The functions of animal microRNAs, *Nature* **431**, 350–355.

Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function, *Cell* **116**, 281–297.

Baulcombe, D. (2004) RNA silencing in plants, *Nature* **431**, 356–363.

Dykxhoorn, D.M., Novina, C.D., Sharp, P.A. (2003) Killing the messenger: short RNAs that silence gene expression, *Nat. Rev. Mol. Cell Biol.* **4**, 457–467.

Lee, R., Feinbaum, R., Ambros, V. (2004) A short history of a short RNA, *Cell* **116**, S89–S92, 81 p following S96.

Ruvkun, G., Wightman, B., Ha, I. (2004) The 20 years it took to recognize the importance of tiny RNAs, *Cell* **116**, S93–S96, 92 p following S96.

Zamore, P.D. (2002) Ancient pathways programmed by small RNAs, *Science* **296**, 1265–1269.

Primary Literature

Abrahante, J.E., Daul, A.L., Li, M., Volk, M.L., Tennessen, J.M., Miller, E.A., Rougvie, A.E. (2003). The Caenorhabditis elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs. *Dev. Cell* **4**, 625–637.

Allen, E., Xie, Z., Gustafson, A.M., Sung, G.H., Spatafora, J.W., Carrington, J.C. (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in arabidopsis thaliana, *Nat. Genet.* **36**, 1282–1290.

Ambros, V., Lee, R.C. (2004) Identification of microRNAs and other tiny noncoding RNAs by cDNA cloning, *Methods Mol. Biol.* **265**, 131–158.

Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., Gvozdev, V.A. (2001) Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline, *Curr. Biol.* **11**, 1017–1027.

Bartel, D.P., Chen, C.Z. (2004) Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs, *Nat. Rev. Genet.* **5**, 396–400.

Berns, K., Hijmans, E.M., Mullenders, J., Brummelkamp, T.R., Velds, A., Heimerikx, M., Kerkhoven, R.M., Madiredjo, M., Nijkamp, W., Weigelt, B., Agami, R., Ge, W., Cavet, G., Linsley, P.S., Beijersbergen, R.L., Bernards, R. (2004) A large-scale RNAi screen in human cells identifies new components of the p53 pathway, *Nature* **428**, 431–437.

Bernstein, E., Caudy, A.A., Hammond, S.M., Hannon, G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference, *Nature* **409**, 363–366.

Bhuyan, P.K., Kariko, K., Capodici, J., Lubinski, J., Hook, L.M., Friedman, H.M., Weissman, D. (2004) Short interfering RNA-mediated inhibition of herpes simplex virus type 1 gene expression and function during infection of human keratinocytes, *J. Virol.* **78**, 10276–10281.

Bonnet, E., Wuyts, J., Rouze, P., Van de Peer, Y. (2004) Detection of 91 potential conserved plant microRNAs in Arabidopsis thaliana and Oryza sativa identifies important target genes, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11511–11516.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., Cohen, S.M. (2003) bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila, *Cell* **113**, 25–36.

Brummelkamp, T.R., Bernards, R., Agami, R. (2002) A system for stable expression of short interfering RNAs in mammalian cells, *Science* **296**, 550–553.

Cai, X., Hagedorn, C.H., Cullen, B.R. (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs, *RNA* **10**, 1957–1966.

Caudy, A.A., Myers, M., Hannon, G.J., Hammond, S.M. (2002) Fragile X-related protein and VIG associate with the RNA interference machinery, *Genes Dev.* **16**, 2491–2496.

Caudy, A.A., Ketting, R.F., Hammond, S.M., Denli, A.M., Bathoorn, A.M., Tops, B.B., Silva, J.M., Myers, M.M., Hannon, G.J., Plasterk, R.H. (2003) A micrococcal nuclease homologue in RNAi effector complexes, *Nature* **425**, 411–414.

Chalfie, M., Horvitz, H.R., Sulston, J.E. (1981) Mutations that lead to reiterations in the cell lineages of C. elegans, *Cell* **24**, 59–69.

Chang, S., Johnston, R.J., Jr., Frokjaer-Jensen, C., Lockery, S., Hobert, O. (2004) MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode, *Nature* **430**, 785–789.

Chiu, Y.L., Rana, T.M. (2002) RNAi in human cells: basic structural and functional features of small interfering RNA, *Mol. Cell* **10**, 549–561.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., Hannon, G.J. (2004) Processing of primary microRNAs by the microprocessor complex, *Nature* **432**, 231–235.

Doench, J.G., Sharp, P.A. (2004) Specificity of microRNA target selection in translational repression, *Genes Dev.* **18**, 504–511.

Doench, J.G., Petersen, C.P., Sharp, P.A. (2003) siRNAs can function as miRNAs, *Genes Dev.* **17**, 438–442.

Dunoyer, P., Lecellier, C.H., Parizotto, E.A., Himber, C., Voinnet, O. (2004) Probing the microRNA and small interfering RNA pathways with virus-encoded suppressors of RNA silencing, *Plant Cell* **16**, 1235–1250.

Elbashir, S.M., Lendeckel, W., Tuschl, T. (2001b) RNA interference is mediated by 21- and 22-nucleotide RNAs, *Genes Dev.* **15**, 188–200.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., Tuschl, T. (2001c) Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate, *Embo J.* **20**, 6877–6888.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T. (2001a) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells, *Nature* **411**, 494–498.

Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., Marks, D.S. (2003) MicroRNA targets in Drosophila, *Genome. Biol.* **5**, R1.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C. (1998) Potent and

specific genetic interference by double-stranded RNA in Caenorhabditis elegans, *Nature* **391**, 806–811.

Fukagawa, T., Nogami, M., Yoshikawa, M., Ikeno, M., Okazaki, T., Takami, Y., Nakayama, T., Oshimura, M. (2004) Dicer is essential for formation of the heterochromatin structure in vertebrate cells, *Nat. Cell Biol.* **6**, 784–791.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., Shiekhattar, R. (2004) The microprocessor complex mediates the genesis of microRNAs, *Nature* **432**, 235–240.

Griffiths-Jones, S. (2004) The microRNA registry, *Nucleic Acids Res.* 32 Database issue, D109–D111.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., Mello, C.C. (2001) Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. Elegans developmental timing, *Cell* **106**, 23–34.

Ha, I., Wightman, B., Ruvkun, G. (1996) A bulged lin-4/lin-14 RNA duplex is sufficient for Caenorhabditis elegans lin-14 temporal gradient formation, *Genes Dev.* **10**, 3041–3050.

Haley, B., Zamore, P.D. (2004) Kinetic analysis of the RNAi enzyme complex, *Nat. Struct. Mol. Biol.* **11**, 599–606.

Hamilton, A.J., Baulcombe, D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants, *Science* **286**, 950–952.

Hammond, S.M., Bernstein, E., Beach, D., Hannon, G.J. (2000) An RNA-directed nuclease mediates posttranscriptional gene silencing in Drosophila cells, *Nature* **404**, 293–296.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., Hannon, G.J. (2001) Argonaute2, a link between genetic and biochemical analyses of RNAi, *Science* **293**, 1146–1150.

Houbaviy, H.B., Murray, M.F., Sharp, P.A. (2003) Embryonic stem cell-specific MicroRNAs, *Dev. Cell* **5**, 351–358.

Hutvagner, G., Zamore, P.D. (2002) A microRNA in a multiple-turnover RNAi enzyme complex, *Science* **297**, 2056–2060.

Hutvagner, G., Simard, M.J., Mello, C.C., Zamore, P.D. (2004) Sequence-specific inhibition of small RNA function, *PLOS Biol.* **2**, E98.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., Zamore, P.D. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA, *Science* **293**, 834–838.

Ishizuka, A., Siomi, M.C., Siomi, H. (2002) A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins, *Genes Dev.* **16**, 2497–2508.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., Linsley, P.S. (2003) Expression profiling reveals off-target gene regulation by RNAi, *Nat. Biotechnol.* **21**, 635–637.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., Marks, D.S. (2004) Human MicroRNA targets, *PLOS Biol* **2**, e363.

Jones-Rhoades, M.W., Bartel, D.P. (2004) Computational Identification of plant MicroRNAs and their targets, including a stress-induced miRNA, *Mol. Cell.* **14**, 787–799.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., Plasterk, R.H. (1999) Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD, *Cell* **99**, 133–141.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., Plasterk, R.H. (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans, *Genes Dev.* **15**, 2654–2659.

Khvorova, A., Reynolds, A., Jayasena, S.D. (2003) Functional siRNAs and miRNAs exhibit strand bias, *Cell* **115**, 209–216.

Kiriakidou, M., Nelson, P.T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., Hatzigeorgiou, A. (2004) A combined computational-experimental approach predicts human microRNA targets, *Genes Dev.* **18**, 1165–1178.

Knight, S.W., Bass, B.L. (2001) A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans, *Science* **293**, 2269–2271.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs, *Science* **294**, 853–858.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., Tuschl, T. (2002) Identification of tissue-specific microRNAs from mouse, *Curr. Biol.* **12**, 735–739.

Lai, E.C. (2002) Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative posttranscriptional regulation, *Nat. Genet.* **30**, 363–364.

Lau, N.C., Lim, L.P., Weinstein, E.G., Bartel, D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans, *Science* **294**, 858–862.

Lee, R.C., Ambros, V. (2001) An extensive class of small RNAs in caenorhabditis elegans, *Science* **294**, 862–864.

Lee, R.C., Feinbaum, R.L., Ambros, V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14, *Cell* **75**, 843–854.

Lee, Y., Jeon, K., Lee, J.T., Kim, S., Kim, V.N. (2002) MicroRNA maturation: stepwise processing and subcellular localization, *Embo J* **21**, 4663–4670.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., Kim, V.N. (2004a) MicroRNA genes are transcribed by RNA polymerase II, *Embo J* **23**, 4051–4060.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., Carthew, R.W. (2004b) Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways, *Cell* **117**, 69–81.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., Kim, V.N. (2003) The nuclear RNase III Drosha initiates microRNA processing, *Nature* **425**, 415–419.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., Burge, C.B. (2003) Prediction of mammalian microRNA targets, *Cell* **115**, 787–798.

Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., Bartel, D.P. (2003a) Vertebrate microRNA genes, *Science* **299**, 1540.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., Bartel, D.P. (2003b) The microRNAs of Caenorhabditis elegans, *Genes Dev.* **17**, 991–1008.

Lin, S.Y., Johnson, S.M., Abraham, M., Vella, M.C., Pasquinelli, A., Gamberi, C., Gottlieb, E., Slack, F.J. (2003) The C elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target, *Dev. Cell* **4**, 639–650.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., Hannon, G.J. (2004) Argonaute2 is the catalytic engine of mammalian RNAi, *Science* **305**, 1437–1441.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., Kutay, U. (2004) Nuclear export of microRNA precursors, *Science* **303**, 95–98.

Mallory, A.C., Reinhart, B.J., Jones-Rhoades, M.W., Tang, G., Zamore, P.D., Barton, M.K., Bartel, D.P. (2004) MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5′ region, *Embo J.* **23**, 3356–3364.

Martinez, J., Tuschl, T. (2004) RISC is a 5′ phosphomonoester-producing RNA endonuclease, *Genes Dev.* **18**, 975–980.

Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., Tuschl, T. (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi, *Cell* **110**, 563–574.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs, *Mol. Cell* **15**, 185–197.

Moss, E.G., Lee, R.C., Ambros, V. (1997) The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA, *Cell* **88**, 637–646.

Moss, E.G., Tang, L. (2003) Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites, *Dev. Biol.* **258**, 432–442.

Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., Dreyfuss, G. (2002) miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs, *Genes Dev.* **16**, 720–728.

Napoli, C., Lemieux, C., Jorgensen, R. (1990). Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell* **2**, 279–289.

Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D., Grewal, S.I. (2004) RITS acts in cis to promote RNA interference-mediated transcriptional and posttranscriptional silencing, *Nat. Genet.* **36**, 1174–1180.

Novina, C.D., Murray, M.F., Dykxhoorn, D.M., Beresford, P.J., Riess, J., Lee, S.K., Collman, R.G., Lieberman, J., Shankar, P., Sharp, P.A. (2002) siRNA-directed inhibition of HIV-1 infection, *Nat. Med.* **8**, 681–686.

Nykanen, A., Haley, B., Zamore, P.D. (2001) ATP requirements and small interfering RNA structure in the RNA interference pathway, *Cell* **107**, 309–321.

Okamura, K., Ishizuka, A., Siomi, H., Siomi, M.C. (2004) Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways, *Genes Dev.* **18**, 1655–1666.

Olsen, P.H., Ambros, V. (1999) The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation, *Dev. Biol.* **216**, 671–680.

Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J., Conklin, D.S. (2002) Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells, *Genes Dev.* **16**, 948–958.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P., et al. (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA, *Nature* **408**, 86–89.

Pfeffer, S., Zavolan, M., Grasser, F.A., Chien, M., Russo, J.J., Ju, J., John, B., Enright, A.J., Marks, D., Sander, C., Tuschl, T. (2004) Identification of virus-encoded microRNAs, *Science* **304**, 734–736.

Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W., Sontheimer, E.J. (2004) A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila, *Cell* **117**, 83–94.

Pillai, R.S., Artus, C.G., Filipowicz, W. (2004) Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis, *RNA* **10**, 1518–1525.

Poy, M.N., Eliasson, L., Krutzfeldt, J., Kuwajima, S., Ma, X., Macdonald, P.E., Pfeffer, S., Tuschl, T., Rajewsky, N., Rorsman, P., Stoffel, M. (2004) A pancreatic islet-specific microRNA regulates insulin secretion, *Nature* **432**, 226–230.

Rajewsky, N., Socci, N.D. (2004) Computational identification of microRNA targets, *Dev. Biol.* **267**, 529–535.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., Ruvkun, G. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans, *Nature* **403**, 901–906.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., Bartel, D.P. (2002) MicroRNAs in plants, *Genes Dev.* **16**, 1616–1626.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., Bartel, D.P. (2002) Prediction of plant microRNA targets, *Cell* **110**, 513–520.

Rubinson, D.A., Dillon, C.P., Kwiatkowski, A.V., Sievers, C., Yang, L., Kopinja, J., Rooney, D.L., Ihrig, M.M., McManus, M.T., Gertler, F.B. et al. (2003) A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference, *Nat. Genet.* **33**, 401–406.

Sarot, E., Payen-Groschene, G., Bucheton, A., Pelisson, A. (2004) Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene, *Genetics* **166**, 1313–1321.

Saxena, S., Jonsson, Z.O., Dutta, A. (2003) Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells, *J. Biol. Chem.* **278**, 44312–44319.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., Zamore, P.D. (2003) Asymmetry in the assembly of the RNAi enzyme complex, *Cell* **115**, 199–208.

Schwarz, D.S., Hutvagner, G., Haley, B., Zamore, P.D. (2002) Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways, *Mol. Cell* **10**, 537–548.

Schwarz, D.S., Tomari, Y., Zamore, P.D. (2004) The RNA-induced silencing complex is a Mg2+-dependent endonuclease, *Curr. Biol.* **14**, 787–791.

Seggerson, K., Tang, L., Moss, E.G. (2002) Two genetic circuits repress the caenorhabditis elegans heterochronic gene lin-28 after translation initiation, *Dev. Biol.* **243**, 215–225.

Shen, B., Goodman, H.M. (2004) Uridine addition after microRNA-directed cleavage, *Science* **306**, 997.

Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R., Ruvkun, G. (2000) The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor, *Mol. Cell* **5**, 659–669.

Song, E., Lee, S.K., Dykxhoorn, D.M., Novina, C., Zhang, D., Crawford, K., Cerny, J.,

Sharp, P.A., Lieberman, J., Manjunath, N., Shankar, P. (2003) Sustained small interfering RNA-mediated human immunodeficiency virus type 1 inhibition in primary macrophages, *J. Virol.* **77**, 7174–7181.

Song, J.J., Smith, S.K., Hannon, G.J., Joshua-Tor, L. (2004) Crystal structure of Argonaute and its implications for RISC slicer activity, *Science* **305**, 1434–1437.

Souret, F.F., Kastenmayer, J.P., Green, P.J. (2004) AtXRN4 degrades mRNA in arabidopsis and its substrates include selected miRNA targets, *Mol Cell* **15**, 173–183.

Soutschek, J., Akinc, A., Bramlage, B., Charisse, K., Constien, R., Donoghue, M., Elbashir, S., Geick, A., Hadwiger, P., Harborth, J. et al. (2004) Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs, *Nature* **432**, 173–178.

Stark, A., Brennecke, J., Russell, R.B., Cohen, S.M. (2003) Identification of Drosophila MicroRNA targets, *PLOS Biol.* **1**, 1–13.

Stewart, S.A., Dykxhoorn, D.M., Palliser, D., Mizuno, H., Yu, E.Y., An, D.S., Sabatini, D.M., Chen, I.S., Hahn, W.C., Sharp, P.A., et al. (2003) Lentivirus-delivered stable gene silencing by RNAi in primary cells, *RNA* **9**, 493–501.

Suh, M.R., Lee, Y., Kim, J.Y., Kim, S.K., Moon, S.H., Lee, J.Y., Cha, K.Y., Chung, H.M., Yoon, H.S., Moon, S.Y., et al. (2004) Human embryonic stem cells express a unique set of microRNAs, *Dev. Biol.* **270**, 488–498.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., Mello, C.C. (1999) The rde-1 gene, RNA interference, and transposon silencing in C. elegans, *Cell* **99**, 123–132.

Tang, G., Zamore, P.D. (2004) Biochemical dissection of RNA silencing in plants, *Methods Mol. Biol.* **257**, 223–244.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., Zamore, P.D. (2004) A protein sensor for siRNA asymmetry, *Science* **306**, 1377–1380.

Vaucheret, H., Vazquez, F., Crete, P., Bartel, D.P. (2004) The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development, *Genes Dev.* **18**, 1187–1197.

Vella, M.C., Choi, E.Y., Lin, S.Y., Reinert, K., Slack, F.J. (2004) The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3′UTR, *Genes Dev.* **18**, 132–137.

Ventura, A., Meissner, A., Dillon, C.P., McManus, M., Sharp, P.A., Van Parijs, L., Jaenisch, R., Jacks, T. (2004) Cre-lox-regulated conditional RNA interference from transgenes, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10380–10385.

Wang, J.F., Zhou, H., Chen, Y.Q., Luo, Q.J., Qu, L.H. (2004) Identification of 20 microRNAs from Oryza sativa, *Nucleic Acids Res.* **32**, 1688–1695.

Wightman, B., Burglin, T.R., Gatto, J., Arasu, P., Ruvkun, G. (1991) Negative regulatory sequences in the lin-14 3′-untranslated region are necessary to generate a temporal switch during Caenorhabditis elegans development, *Genes Dev.* **5**, 1813–1824.

Wightman, B., Ha, I., Ruvkun, G. (1993) Post-transcriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans, *Cell* **75**, 855–862.

Xu, P., Vernooy, S.Y., Guo, M., Hay, B.A. (2003) The Drosophila microRNA Mir-14 suppresses cell death and is required for normal fat metabolism, *Curr. Biol.* **13**, 790–795.

Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., Zhou, M.M. (2003) Structure and conserved RNA binding of the PAZ domain, *Nature* **426**, 468–474.

Yekta, S., Shih, I.H., Bartel, D.P. (2004) MicroRNA-directed cleavage of HOXB8 mRNA, *Science* **304**, 594–596.

Yi, R., Qin, Y., Macara, I.G., Cullen, B.R. (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs, *Genes Dev.* **17**, 3011–3016.

Zamore, P.D., Tuschl, T., Sharp, P.A., Bartel, D.P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals, *Cell* **101**, 25–33.

Zeng, Y., Cullen, B.R. (2003) Sequence requirements for micro RNA processing and function in human cells, *RNA* **9**, 112–123.

Zeng, Y., Cullen, B.R. (2004) Structural requirements for pre-microRNA binding and nuclear export by Exportin 5, *Nucleic Acids Res.* **32**, 4776–4785.

Zeng, Y., Wagner, E.J., Cullen, B.R. (2002) Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells, *Mol. Cell* **9**, 1327–1333.

Zeng, Y., Yi, R., Cullen, B.R. (2003) MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9779–9784.

Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E., Filipowicz, W. (2004) Single processing center models for human Dicer and bacterial RNase III, *Cell* **118**, 57–68.

RNA Methodologies

Robert E. Farrell, Jr.
Penn State University, York, PA

1 Introduction 559

2 Subpopulations of RNA 559
2.1 Total Cellular RNA 560
2.2 Total Cytoplasmic RNA 560
2.3 Nuclear RNA 560
2.4 Poly(A)$^+$ RNA 560
2.5 Poly(A)$^-$ RNA 561

3 Goals in the Purification of RNA (adapted, in part, from Farrell, 1993; 1998; 2005) 561

4 Methods of Cellular Disruption 563
4.1 Gentle Lysis Buffers 563
4.2 Harsh Lysis Buffers 564

5 Inhibition of Ribonuclease Activity 565
5.1 Preparation of Equipment and Reagents 565
5.2 Inhibitors of RNase 566

6 Methods for the Analysis of RNA 567
6.1 Northern Analysis 567
6.2 Nuclease Protection Assay 569
6.3 Dot Blot Analysis 572
6.4 Transcription Rate Assays 573
6.5 Synthesis of cDNA 575
6.6 RT-PCR 575
6.7 RNAi 576
6.8 *In Vitro* Translation 578

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

7 Summary 578

Bibliography 579
Books and Reviews 579
Primary Literature 579

Keywords

Chaotropic
Biologically disruptive. Chaotropic lysis buffers disrupt cell and organelle membranes and destroy enzymatic activity on contact.

Complementary DNA (cDNA)
DNA synthesized *in vitro* from an RNA template by an enzyme known as a *reverse transcriptase*. cDNA can be either single- or double-stranded, and is used for RT-PCR, nucleic acid probe synthesis, or library construction. Because cDNA can only be made from transcripts present at the moment of cellular disruption, it is a permanent biochemical record of the cell.

Diethyl Pyrocarbonate (DEPC)
A chemical used to purge reagents of nuclease activity. DEPC is carcinogenic and should be handled with extreme care. See text for details.

Dot Blot (also Slot Blot)
A membrane-based technique for the quantification of specific RNA or DNA sequences in a sample. The sample is usually ''dot'' configured onto a filter by vacuum filtration through a manifold. Dot blots lack the qualitative component associated with electrophoretic assays.

Heterogeneous nuclear RNA (hnRNA)
The primary product of RNA polymerase II transcription in eukaryotic cells. hnRNA alone can be processed and mature into mRNA, which, in turn, is able to support the synthesis of proteins.

Hybridization
The formation of hydrogen bonds between two nucleic acid molecules, which are complementary to some degree. The specificity of hybridization is a direct function of the stringency of the system in which the hybridization is being conducted.

Messenger RNA (mRNA)
The mature product of RNA polymerase II transcription. mRNA is derived from heterogeneous nuclear RNA (hnRNA) and, in conjunction with the protein translation apparatus, is capable of directing the translation of the encoded polypeptide.

Northern Blot Analysis
A technique for transferring electrophoretically chromatographed RNA from a gel matrix, usually agarose, onto a filter paper for subsequent immobilization and hybridization. The information gained from Northern blot analysis is used to qualitatively and quantitatively assess the expression of specific genes.

Nuclear Runoff Assay
A method for labeling nascent RNA molecules in isolated nuclei. The rate at which specific RNAs are being transcribed can then be assayed based upon the degree of label incorporation. See entry for steady-state RNA for comparison.

Nuclease Protection Assay
A method for mapping and/or quantifying RNA transcripts. In general, hybridization between probe and target RNA takes place in solution, followed by nuclease digestion (with S1 nuclease or RNase) of all molecules or parts thereof, which do not actually participate in duplex formation. Nucleic acid molecules, which are locked up in a double-stranded configuration, are relatively safe or protected from nuclease degradation. The undigested RNA:RNA or RNA:DNA hybrids are then precipitated and/or electrophoresed for quantification.

Poly(A)$^+$Tail
A tract of adenosine residues enzymatically added to the 3′ terminus of mRNA by the nuclear enzyme poly(A) polymerase. The addition of a poly(A) tail involves cleavage at the 3′ terminus of the primary transcript followed by polyadenylation.

Polymerase Chain Reaction (PCR)
Primer-mediated enzymatic amplification of cDNA or genomic DNA sequences. This technology revolutionized molecular biology in the early and mid-1990s.

Primer
A short single-stranded nucleic acid molecule, or portion thereof, which can base pair with a complementary sequence and which provides a free 3′-OH for any of a variety of primer extension-related reactions.

Probe
Usually labeled nucleic acid molecules, either DNA or RNA, which is used to hybridize to complementary sequences present among the complexity of different target sequences in a nucleic acid sample, as in the Northern analysis, Southern analysis, or nuclease protection analyses and DNA library screening.

Proteome
The full complement of proteins produced by a cell at a particular time. Proteome maps are typically generated and assessed by 2-D electrophoresis and other techniques designed to identify, quantify, and characterize the products of translation.

Real-Time PCR
A state-of-the-art method for measuring PCR product accumulation as it is produced in each cycle, rather than measuring final product mass at the end of the reaction. Real-time PCR is widely regarded as the premier quantitative PCR-based assay.

Relative Abundance
The quantity of a particular RNA transcript relative to some other transcript in the same sample **or** relative to the amount of the same transcript in other experimentally related samples. This determination is made by hybridization with a complementary nucleic acid (see above).

Ribonuclease
A family of resilient enzymes, which rapidly degrade RNA molecules. Control of ribonuclease activity is a key consideration in all manipulations involving RNA.

RNA (Ribonucleic acid)
A polymer of ribonucleoside monophosphates, synthesized by an RNA polymerase. RNA is the product of transcription.

RNA Interference (RNAi)
RNAi is also known as *posttranscriptional gene silencing*.

RT-PCR
The technology by which RNA molecules are converted into their complementary DNA (cDNA) sequences by any one of several reverse transcriptases, followed by the amplification of the newly synthesized cDNA by standard PCR amplification.

Steady State RNA
The final accumulation of RNA in the cell/cytoplasm. For example, measurement of the cytoplasmic prevalence of a particular species of mRNA in no way relates to the **rate** of transcription or RNA degradation in the cell (see nuclear runoff assay).

Transcription
The process by which complementary RNA molecules are synthesized from a DNA template.

Transcriptome
The complete set of mRNA molecules produced by a particular cell under a particular set of circumstances.

Xenobiotic
A chemical foreign to a biological system.

The cellular biochemistry is reflected in the overall abundance of nuclear and cytoplasmic RNA species, which, inevitably, drive the phenotype of the cell. In order to more readily understand the cellular response to experimental or environmental challenges, various subpopulations of RNA are harvested and characterized to gain insight into the differential expression of genes and possibly the cellular level at which these genes are modulated. The expedient isolation of high quality RNA is essential to support all downstream applications, and the method by which this end is accomplished is dictated by the nature of the biological source material. The goal in a study such as this is to quantify gene expression by measuring the prevalence of one or more RNA species, which are of immediate interest to the investigator, and how these data relate to a control population of RNAs. Methods for RNA analysis are rather diverse and each of these provides a glimpse of some aspect of gene regulation with a characteristic level of sensitivity. Often several RNA methodologies are utilized to provide a more complete picture of the up- and downregulation of specific genes and gene families.

1 Introduction

The isolation and characterization of ribonucleic acid (RNA) from cells and tissue samples is a central and recurrent theme in molecular biology. In particular, the purification of chemically stable and biologically functional RNA is the starting point for the systematic evaluation of the cellular biochemistry by standard molecular techniques such as Northern analysis, nuclease protection (S1 and ribonuclease protection assays), dot blot analysis, nuclear runoff assay, the synthesis of representative cDNA libraries, and all forms of RT-PCR (The polymerase chain reaction (PCR) process is covered by patents currently owned by Roche Molecular Systems, Inc. and F. Hoffmann-LaRoche Ltd. Use of PCR requires a limited license under their patents. PCR must also be performed in an authorized thermal cycler. For specific details contact the Director of Licensing, Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404.). Because messenger RNA (mRNA) abundance is a useful parameter of gene expression, the expedient recovery of RNA from a biological source is a critical first step for the derivation of meaningful data. Difficulties in the purification, handling, and storage of RNA are intrinsic to the labile chemical nature of these molecules. These difficulties are further compounded by the aggressive character of resilient ribonucleases (RNase), the apparent ubiquity of which is undisputed. Indeed, the novice quickly learns of the absolute requirement for management of RNase activity at each level of RNA isolation and characterization. Failure to do so will almost certainly compromise the integrity of the resulting RNA and its probable utility.

2 Subpopulations of RNA

Prior to the onset of cellular disruption, the investigator must determine which RNA subpopulation is of experimental interest.

For example, the precise questions being asked of a particular set of experiments may require characterization of total cellular RNA, cytoplasmic RNA alone, nuclear RNA alone, poly(A)$^+$ RNA, or poly(A)$^-$ RNA. Transcriptional activity is generally assayed using one of the methods described below, for example, Northern analysis, and the data is then validated using another method, such as nuclease protection or RT-PCR.

2.1 Total Cellular RNA

Total cellular RNA is generally isolated when it is necessary to assess the transcriptional activity of a gene (is the gene transcribed and to what extent?) by any of the methods delineated below. The isolation of total cellular RNA is usually easier and more rapid compared to other isolation strategies. Because purified total cellular RNA consists of both nuclear RNA and cytoplasmic RNA, it is impossible to determine the relative contribution of each. Most of the kits and reagents, which favor rapid isolation in less than 30 min, yield either total cellular RNA or total poly(A)$^+$ RNA (both mRNA and hnRNA). The disadvantage of working with populations of total cellular RNA is that the investigator is generally unable to determine the relative contribution of RNA from the nucleus and the cytoplasm. This distinction may or may not be important in the context of a particular investigation.

2.2 Total Cytoplasmic RNA

Total cytoplasmic RNA (predominantly rRNA, tRNA, mRNA) is generally isolated when assessment of posttranscriptional regulation of genes is being investigated. Purification of cytoplasmic RNA can also be a method of enrichment, that is, by noninclusion of the nuclear RNA, the percentage of the total mass of RNA in a sample that mRNA represents is increased. Further enrichment can be accomplished via the selection of poly(A)$^+$ mRNA, as described below.

2.3 Nuclear RNA

Nuclear RNA is often studied in conjunction with the independent characterization of cytoplasmic RNA as a means of assessing the level (transcriptional vs posttranscriptional) and degree of regulation of various genes. It is well documented that a large mass of transcribed RNA is degraded in the nucleus; this precursor RNA never matures into mRNA capable of supporting translation in the eukaryotic cytoplasm. By comparing the nuclear abundance and cytoplasmic abundance of a particular RNA, a cause–effect relationship may be discerned between an experimental manipulation and the regulation of gene expression in that system with respect to RNA biogenesis because hnRNA, produced by RNA polymerase II, matures into mRNA. The analysis of nuclear RNA may also be performed in order to study the rate at which genes are transcribed (nuclear runoff assay, below), as opposed to the assay of steady-state RNA levels, data that can be used to clarify the assessment of the level of regulation of gene expression.

2.4 Poly(A)$^+$ RNA

Poly(A)$^+$ RNA can be purified from previously isolated cellular RNA, cytoplasmic RNA, or directly from whole cell lysates.

This approach is generally taken for one of two fundamental reasons. First, a sample of poly(A)$^+$ mRNA is often assumed to reflect the biochemical and physiological status of the cell. By eliminating rRNA, tRNA, and large, unspliced nuclear RNA species, improved sensitivity and resolution in the assay of specific mRNAs is greatly favored. Secondly, purified poly(A)$^+$ mRNA is the template material for *in vitro* synthesis of complementary DNA (cDNA). While cDNA can be synthesized from total RNA, exclusion of the highly abundant rRNA and tRNA species from the synthesis reactions generates cDNA which is far more representative of the cellular biochemistry at the moment of cellular disruption. Poly(A)$^+$ mRNA may be purified by affinity chromatography using oligo(dT)$_{12-18}$ linked to one of several popular matrices, including paramagnetic beads, biotin, cellulose beads, or microcrystalline cellulose. It is important to note, however, that transcript enrichment performed to increase sensitivity may actually be counterproductive in some cases because the loss of some mRNA during the enrichment procedure may serve only to further underrepresent very low-abundance mRNA. Due in no small measure to the power of PCR and the clever design of the required primers, poly(A)$^+$ selection is viewed by many investigators as unnecessary for most contemporary applications.

2.5 Poly(A)$^-$ RNA

Poly(A)$^-$ RNA is that subpopulation of RNA lacking the tract of adenosine residues, which characterizes the 3′ terminus of a majority of eukaryotic mRNAs. The predominant members of this class include rRNA, tRNA, as well as a small family of naturally nonpolyadenylated mRNAs, of which, the histone mRNAs are a noteworthy example. In instances where poly(A)$^-$ mRNA might not be detected due to exclusion from a sample, either the poly(A)$^-$ fraction or a sample of total RNA from the same biological source will contain these naturally nonadenylated transcripts for assay, assuming their respective genes are transcriptionally active. Moreover, depletion of poly(A)$^+$ mRNA from a sample renders the resulting poly(A)$^-$ fraction an excellent negative control in the assay of poly(A)$^+$ mRNA species.

3 Goals in the Purification of RNA (adapted, in part, from Farrell, 1993; 1998; 2005)

Proper and thoughtful planning prior to the onset of lab work is an absolute requirement for the recovery of high quality RNA capable of supporting further biochemical analysis. In the preliminary stages, an experimental design for the purification of nucleic acids in general must address five specific goals, the successful achievement of each of these, in order, will influence profoundly the yield, quality, and utility of the sample.

Goal 1: Select an appropriate method for membrane solubilization. The first decision factoring into an RNA isolation strategy is based upon which population of RNA or subcellular compartment the investigator wishes to study. For example, an investigator may wish to determine whether the observed modulation of gene expression in a model system is regulated transcriptionally or by some posttranscriptional event(s). In such an instance, selection of the method of cellular disruption and subsequent RNA isolation must permit

analysis of salient nuclear transcripts independently of those localized in the cytoplasm. The method of cell lysis will determine the extent of subcellular disruption in a sample, and is a direct function of the lysis buffer. For example, a lysis buffer that is used successfully with tissue culture cells may be entirely inappropriate for whole tissue samples. Further, the method by which membrane solubilization is accomplished will dictate which additional steps will be required to remove DNA and protein from the RNA preparation, and whether compartmentalized nuclear RNA and cytoplasmic RNA species can be purified independently of one another. While DNA can be purged from an RNA preparation with minimal fanfare, it is virtually impossible to determine the relative contribution of transcripts from the nucleus and from the cytoplasm once RNA from these two subcellular compartments have mingled and copurified. A particular lysis procedure must likewise demonstrate compatibility with ensuing protocols. The take-home lesson is to always think two steps ahead: the proper method of solubilization is dependent upon the plans for the RNA after purification and the questions being asked of a particular study.

Goal 2: Ensure total inhibition of nuclease activity. The imperative for controlling nuclease activity is nonnegotiable. This includes purging RNase from reagents and equipment (extrinsic sources of nuclease activity) and controlling RNase activity in a cell lysate (intrinsic source of nuclease activity). While some lysis reagents inhibit nuclease activity in their own right (harsh lysis buffers, below), other lysis reagents require additional nuclease inhibitors (gentle lysis buffers, below) to safeguard the RNA during the isolation procedure. With respect to long-term storage of purified RNA, it is incumbent upon the investigator to ensure that added ribonuclease inhibitors will not interfere with any subsequent manipulations and/or reactions involving the RNA.

Goal 3: Remove contaminating proteins from the sample. The complete removal of protein from a cellular lysate is of paramount importance in the isolation of both RNA and DNA. Meticulous attention to this detail is required both for accurate quantification and precision in hybridization. The removal of proteins from nucleic acid samples may be accomplished by

1. protein hydrolysis with proteinase K;
2. salting-out of proteins;
3. solubilizing proteins in guanidinium-based buffers;
4. repeated extraction with mixtures of phenol and chloroform;
5. any combination of the above.

RNA molecules are not as fragile as high molecular weight DNA; thus, more aggressive methods can be employed for removal of proteins, including the use of phenol:chloroform. While deproteinization is itself a means of controlling ribonuclease activity, purified RNA samples will be once again susceptible to nuclease degradation following removal of the protein denaturant.

Goal 4: Concentrate the sample. This is the final step in nearly all RNA purification schemes. The most versatile method for concentrating nucleic acids is precipitation using various combinations of salt and alcohol, such as the very common method of adding sodium acetate and ethanol. Nucleic acids and the salt that drives their precipitation form complexes that have greatly reduced solubility in high concentrations of alcohol. Further, the rate of precipitation appears to be temperature dependent. Unlike the precipitation

of genomic DNA, the precipitation of RNA often requires longer incubation periods at −20°, a phenomenon that is a direct function of the complexity of cellular RNA. In general, greater g-forces must be applied in order to completely recover an RNA precipitate for subsequent analysis. Other concentration procedures include the use of commercially available concentrating devices, dialysis, centrifugation under vacuum, and binding to silica matrices in high salt. These methods are less frequently employed for RNA work because the degree of success is much more dependent upon the skill of the investigator and equipment availability than is the case with simple salt and alcohol concentration.

Goal 5: Select the proper storage conditions for the purified RNA. Because of the naturally labile character of RNA, improper storage of excellent RNA samples will often result in degradation in a relatively short time. There are many opinions as to the proper temperature, buffer, and storage form for RNA. As a general rule, RNA is most stable as an ethanol precipitate at −80°. Large samples or RNA stocks should be stored in convenient aliquots in order to avoid repeated freeze–thaw cycles.

4 Methods of Cellular Disruption

As suggested above, in order to select a suitable method for cellular disruption or "solubilization," consideration must be given as to which subpopulation of RNA is desired for study as well as the nature of the biological material (i.e. cells grown in tissue culture vs whole tissues). Lysis buffers are generally categorized as either "gentle" or "harsh."

4.1 Gentle Lysis Buffers

Gentle lysis buffers are used when a specific subpopulation of RNA is desired (e.g. cytoplasmic RNA alone) and nuclear integrity must be maintained. Gentle lysis buffers, often slightly hypotonic, frequently contain the nonionic detergent NP-40 (Nonidet™ P-40 (Nonidet P-40 is a registered trademark of the Shell Company.); now known as *Igepal CA-630*). Because osmotic lysis is the gentlest form of cellular disruption, NP-40 lysis buffers are ideal for the solubilization of the plasma membrane alone, and the inclusion of low concentrations of magnesium helps maintain nuclear integrity. Thus, the nucleus and its contents (DNA and nuclear RNA) can be separated from the cytosol by differential centrifugation. The resulting supernatant is rich in cytoplasmic RNA and proteins, the latter of which are easily removed by repeated extraction with phenol: chloroform. If desired, the nuclear pellet may be processed separately for the recovery of nuclear transcripts. This method of cellular disruption is ideally suited for cells harvested from tissue culture. However, owing to the complex geometry and formidable nature of whole tissue samples, NP-40-containing lysis buffers do not work well with tissue samples unless coupled with limited, non-shearing homogenization.

The obvious advantage of this isolation strategy is that the material precipitated at the end of the procedure is cytoplasmic RNA. A disadvantage to using this approach is that the lysis buffer alone is not sufficiently inhibitory toward RNase. Upon cellular lysis, normally sequestered ribonucleases are liberated, the activity of which will greatly compromise the integrity of the RNA even as the investigator

is working diligently to purify the same. It is very helpful to keep the sample on ice at all times, unless the protocol specifically dictates otherwise; it is likewise helpful to use ice-cold reagents and tubes, which have been prechilled on ice prior to use. If desired, an exogenous RNase inhibitor (below) can be added to the lysis buffer.

4.2 Harsh Lysis Buffers

There is probably no better way to deal with seemingly recalcitrant ribonucleases than to disrupt cells in guanidinium lysis buffer. Upon contact, guanidinium-containing buffers distort the tertiary folding of ribonucleases, resulting in the inactivation of these enzymes (other chaotropic lysis buffers, which contain high concentrations of SDS have also been described). The inclusion of additional ribonuclease inhibitors to such lysis buffers is not necessary, and these procedures for RNA isolation are usually performed at room temperature.

In the presence of chaotropic agents, organelle lysis accompanies disruption of the plasma membrane. Thus, nuclear RNA, genomic DNA, and mitochondrial DNA all copurify with the cytoplasmic RNA, requiring additional steps to remove DNA from the sample. In the past, the most prevalent of these methods was isopycnic centrifugation, involving centrifugation through a cesium chloride (CsCl) gradient or cesium trifluoroacetate (CsTFA) gradient. Isopycnic separation is possible because of the differing buoyant densities of DNA (1.5–1.7 g/mL) versus RNA (1.7–2.0 g/mL).

The differential partitioning of DNA, RNA, and protein by acid-phenol extraction was first described by Chomczynski and Sacchi (1987), changing the way the RNA and DNA are separated from each other in a whole-cell lysate or tissue homogenate. The investigator can fully exploit the salient chemical differences between RNA and DNA in an acidic environment and by the judicious blending of organic solvents. Organic extraction of nucleic acids at acidic pH causes DNA partitioning to the interphase and organic phase while RNA remains in the aqueous phase. This approach precludes the requirement for ultracentrifugation and greatly reduces the required amount of hands-on time, to the obvious benefit of the investigator. The popularity of this general approach has resulted in the development of a number of nucleic acid isolation reagents, which support the unceremonious purification of RNA from both tissue and tissue culture cells alike.

In order to take full advantage of the disruptive nature of the guanidinium isolation procedures and still maintain the subcellular compartmentalization of RNA, one worthwhile strategy is to begin the isolation procedure with gentle NP-40 lysis, followed by recovery of intact nuclei, which are then lysed with guanidinium buffer. The purification of nuclear (or cytoplasmic) RNA then proceeds as if working with intact cells. This approach is particularly suited for the isolation of nuclear RNA for Northern analysis.

The principal drawback to these chaotropic methods is that the investigator is not able to discriminate between cytoplasmic and nuclear RNA, as there is no method by which to separate nuclear RNA from mRNA, once mixing has occurred (size fractionation may result in a partial separation). Moreover, it is unfortunate that many seasoned investigators begin to show signs of sloppiness with respect to the control of RNase activity when routinely working with guanidinium buffers. While

it is true that RNA is safe from nuclease degradation in the presence of these agents, purified RNA is once again susceptible to nuclease degradation.

5 Inhibition of Ribonuclease Activity

The difficulties associated with the isolation of full-length, intrinsically labile RNA are further compounded by the ubiquitous RNase activity. RNases are a family of enzymes, which degrade RNA molecules through both endonucleolytic and exonucleolytic activity. Particularly resilient are ribonucleases of the pancreatic variety. These small, remarkably stable enzymes resist denaturation under conditions that would easily purge other enzyme activities, including pH extremes and autoclaving. It is clearly incumbent upon the investigator to ensure that both equipment and reagents are purged of nucleases from the onset of an experiment. For most RNA-minded molecular biologists, saying that a reagent or apparatus is sterile is more than likely a statement that it is ribonuclease free.

The method selected for controlling ribonuclease activity must, first and foremost, demonstrate compatibility with the cell lysis procedure. Nuclease inhibitors are most often added to gentle lysis buffers when subcellular organelles (nuclei especially) must be purified intact, as in the independent examination of nuclear RNA and cytoplasmic RNA. Secondly, the method of nuclease inhibition must support the integrity of the RNA throughout the subsequent fractionation or purification steps, some of which may require several hours. Thirdly, the reagents used to inhibit ribonuclease activity must be easily removed from purified RNA so as not to interfere with subsequent manipulations. In all cases, and especially when characterizing a new cell type or model system for the first time, actions taken to control nuclease activity should be aggressive if not compulsive. Failure to do so is likely to yield a useless sample of degraded RNA.

5.1 Preparation of Equipment and Reagents

Rule number one when working with RNA is to wear gloves during the preparation of reagents and equipment and especially during the actual RNA extraction procedure. Finger greases are notoriously rich in ribonuclease and are generally accepted as the single greatest source of ribonuclease contamination. Further, one should not hesitate to change gloves several times during the course of an RNA-related experiment; door knobs, micropipettes and refrigerator door handles, containers in which chemicals are packaged, nonsterile deionized H_2O, gel electrophoresis boxes and combs, and microbial/fungal growth are also potential sources of nuclease contamination.

With respect to laboratory consumables, any plasticware that is certified as being tissue-culture sterile is always preferred for RNA work. This includes individually wrapped serological pipettes and conical 15 mL and 50 mL tubes. In any event, these items should be handled only with gloves. Bulk packed polypropylene products (microcentrifuge tubes and micropipette tips) are potential sources of nuclease contamination, mainly due to handling and distribution with ungloved hands from a single bag. It is best to purchase these consumables certified as being both DNase- and RNase-free. Any plastic product or other implement that will come into contact with an RNA sample at any time, either

directly or indirectly, and which can withstand autoclaving should be so treated and set aside exclusively for RNA work.

When plasticware is avoided and glassware must be used, as with organic reagents such as phenol and chloroform, individually wrapped borosilicate glass pipettes are strongly preferred. Glassware, which must be reused, should be set aside for RNA work and not be in general circulation in the laboratory. Contrary to popular belief, the temperature and pressure generated during the autoclaving cycle are usually not sufficient to eliminate all ribonuclease activity. Fortunately, ribonucleases can be destroyed quite effectively by baking in a dry heat oven: rinse glassware with ribonuclease-free water (see below) and then bake for 3 to 4 h at 200°. Baking pertains to glassware alone; questions regarding the heating of plastics or other materials can usually be answered by the manufacturers' technical assistance department.

5.2 Inhibitors of RNase

Endogenous ribonuclease activity varies tremendously from one biological source to the next, and the degree to which action must be taken to inhibit nuclease activity is a direct function of cell type. Knowledge of the extent of intrinsic nuclease activity is derived from two principal sources, namely, the salient literature and personal experience.

The method of ribonuclease inhibition is to a great extent a function of the type of lysis buffer. Whereas nondenaturing, osmotic lysis buffers require the inclusion of a nuclease inhibitor, strongly denaturing (chaotropic) lysis buffers generally do not. Such chaotropic compounds include guanidinium thiocyanate, guanidinium HCl, sarcosyl, SDS, 8-hydroxyquinoline, CsCl, CsTFA, and/or various formulations of organic solvents.

Vanadyl ribonucleoside complexes (*VDR; VRC*) were at one time a popular addition to lysis buffers, which alone are ineffective in the control of ribonuclease, as in the popular nonionic NP-40 method. These complexes function as transition state analogs. In the absence of VDR, RNase-mediated cleavage of the phosphodiester backbone of RNA results in the transient formation of a dicyclic transition state intermediate that is subsequently opened by reaction with a water molecule. In its capacity as an RNA analogue, the vanadium complex forms a highly stable dicyclic species to which the enzyme remains irreversibly bound. Thus, nuclease activity is eliminated by locking ribonuclease and "pseudosubstrate" in the transition state. VDR binds tightly to a broad spectrum of cellular ribonucleases including RNase A and RNase T1, but not RNase H, and is compatible with a variety of cell fractionation methods. VDR should be used selectively, however, as even trace carryover quantities are sufficient to inhibit *in vitro* translation of purified mRNA and can also interfere with reverse transcriptase activity, thereby excluding its use with any RT-PCR applications. For this reason, the use of VDR as an RNase inhibitor has fallen out of favor with most molecular biologists.

RNasin® (Promega) may be used to inhibit nuclease activity and circumvent some of the problems commonly associated with the use of VDR. RNasin® inactivates RNase A, RNase B, and RNase C, but not RNase T1, S1 nuclease, nor RNase from *Aspergillus*, and is compatible with a variety of *in vitro* reactions. Care must be taken to avoid strongly denaturing

conditions, which can cause the uncoupling of RNase-RNasin® complexes and restoration of RNase degradative activity.

Diethyl Pyrocarbonate (*DEPC*) is an efficient, nonspecific inhibitor of ribonuclease. It is used to purge reagents of nuclease activity because of the unreliability of autoclaving alone. DEPC is typically added to stock solutions and buffers to a final concentration of 0.05 to 0.1% (v/v); buffers are then shaken or stirred for at least 2 h. Finally, DEPC must be completely destroyed by autoclaving or by maintaining the solutions at 60° overnight. The complete removal of DEPC from buffers and reagents is important because trace amounts of DEPC will result in chemical modification of adenine residues, thereby changing many of its physical properties, including strong inhibition of the polymerase chain reaction. Precautions indicated by the manufacturer must be observed when handling this material, as it is carcinogenic and potentially explosive.

DEPC should *never* be added to any buffer containing tris or mercaptans, with which DEPC is reactive. Instead use DEPC-treated water (add DEPC to water; autoclave to remove DEPC) to make up the tris-containing solution and then autoclave again. Alternatively, buffers consisting of chemicals, which demonstrate DEPC incompatibility, can be twice filtered through a nitrocellulose membrane to remove RNase activity and other trace proteins. Moreover, an electrophoresis apparatus (gel comb, casting tray, or gel box) should never be exposed to DEPC, as acrylic is not DEPC-resistant. Alternatively, various types of laboratory plastics can be rinsed briefly in a 3% solution of H_2O_2, and then rinsed with copious volumes of sterile H_2O. As a powerful oxidizing agent, hydrogen peroxide very effectively destroys RNase activity on many common laboratory surfaces. The use of DEPC has fallen out of favor due to the widespread availability of nuclease-free reagents and stock solutions from most vendors.

Many ribonucleases manage to renature following removal of many denaturing reagents; therefore it is prudent to maintain separate containers of chemicals and stock solutions for exclusive use as RNA reagents. Chemical solids should be weighed out with a ribonuclease-free spatula and stock solutions should be aliquoted into suitable volumes. Aliquots used once should be discarded. While such actions may at first seem to be excessive, they may well preclude the accidental introduction of ribonucleases and facilitate recovery of the highest possible quality RNA.

6 Methods for the Analysis of RNA

Evaluation of the cellular biochemistry by hybridization of RNA is possible in conformity with many of the parameters associated with the analysis of DNA. Methods range from the traditional to the contemporary and each of these has applicability under a defined set of experimental conditions. For the convenience of the reader, the relative merits of three of the standard methods are presented in Table 1.

6.1 Northern Analysis

The quintessential method for the assay of gene expression is a method properly known as *Northern analysis* (also known colloquially as *Northern blotting*, the Northern blot analysis and/or RNA blot analysis). In this technique, purified RNA is resolved

Tab. 1 Comparison of the traditional Northern analysis, nuclease protection assay, nuclear runoff assay, and RT-PCR.

	Northern analysis	***Nuclease protection assay***	***Nuclear runoff assay***	***RT-PCR***
Advantages	Provides a qualitative component to RNA analysis. Supports several rounds of hybridization with different probes. Is compatible with total, cytoplasmic, or poly(A)$^+$ RNA. RNA is relatively stable on filter. Is able to assess integrity of the sample.	Much higher sensitivity than Northern analysis. Requires less handling of RNA than other types of analysis. Is tolerant of partially degraded RNA. Solution hybridization is more quantitative than filter hybridization. Can be used for steady-state or transcription rate assays.	Characterizes relative rate of transcription. Natural geometry of the chromatin is maintained. Permits simultaneous study of several genes. Can be used to discern transcriptional vs posttranscriptional gene regulation when used in conjunction with data from Northern analysis.	Provides unparalleled sensitivity when properly designed. Provides unparalleled resolution. Supersedes many of the classical techniques. Minimizes the amount of handling of the RNA. Very rapid technique. Favors research productivity.
Disadvantages	Is the least sensitive assay. Denaturants can be toxic. Requires extensive handling of RNA. Is a time-consuming process. Provides ample opportunity for RNase degradation. Characterizes only steady-state RNA.	Protected fragment is smaller than native RNA. Nucleases, especially S1, can be difficult to control. Assay is more sensitive to exact hybridization parameters than other assays. Double-stranded probes can compromise quantitativeness of the assay if reannealing occurs.	Nuclear isolation requires a fair amount of skill. Probe complexity is very large. Unlabeled endogenous RNA can compete with labeled RNA during hybridization. Mechanics of the assay support transcript elongation, and not initiation, during labeling.	Much more sensitive to the precise reaction components and conditions than the other assays. Exquisitely sensitive to contaminants, especially genomic DNA. Carryover contamination must be addressed. Optimization can be time-consuming and costly.

From Farrell Jr., R.E. (2005) *RNA Methodologies: A Laboratory Guide for Isolation and Characterization*, 3rd edition, page 389. With permission from Elsevier.

electrophoretically under denaturing conditions and subsequently transferred or "blotted" onto a solid filter support (the so-called "filter membrane"). The RNA on the blot is then hybridized to an appropriately labeled nucleic acid probe, which will support subsequent detection by autoradiography or by chemiluminescence. Because samples of RNA are electrophoresed prior to hybridization, the Northern analysis provides both a quantitative and qualitative biochemical profile of the sample; denaturation of the RNA prior to electrophoresis is necessary to ensure that migration of the sample through the gel occurs only with respect to molecular weight and is not distorted by the formation of the secondary structure commonly associated with single-stranded molecules so that the overall length of the transcript(s) can be determined, a datum, which cannot be discerned using other methods.

The objective of Northern analysis is to quantify gene expression by detecting the relative abundance of those mRNAs in the sample, which are of immediate interest to the investigator. Whereas in the Southern analysis, the resulting data pertains to the structure and organization of genes, data derived by Northern analysis reflects the transcriptional activity of genes.

The principal concern of many in the assessment of Northern blot data is the limited sensitivity of the assay. Few would argue against the notion that the physical application and immobilization of an RNA sample onto a filter membrane renders some of those molecules incapable of/unavailable for base pairing to a complementary nucleic acid probe. Moreover, the Northern analysis is by no means intended to discern the absolute mass of RNA in the cell. Such data may be measured far more accurately by solution hybridization-based methods, especially real-time PCR. Hence, data derived from the Northern analysis must be interpreted in the context of the relative abundance of a particular RNA amongst all of the samples involved.

The electrophoresis of RNA is itself an important diagnostic tool. A great deal can be learned about the integrity and probable utility of an RNA sample upon examination of a representative aliquot. RNA has a very characteristic profile on a denaturing gel (Fig. 1). The appearance of the predominant species, the 28S and 18S ribosomal RNAs (rRNA) is an indicator of the integrity of the sample. Ideally, a very light smearing above, between, and just barely below the rRNAs indicates that the sample is enriched in undergraded RNA and is probably capable of supporting nucleic acid hybridization. Heavier smearing, especially below the level of the 18S rRNA, is quite ominous, being indicative of partially degraded RNA. The complete absence of the ribosomal RNAs indicates a completely degraded sample. As it is obviously desirable to ascertain the integrity of a sample before moving on to sophisticated and often time-consuming techniques, a brief period of electrophoresis to assess the quality of the sample should be the standard operating procedure in any molecular biology setting.

6.2 Nuclease Protection Assay

The intrinsic shortcomings of the Northern analysis mandate a different format for the assay of gene expression when very exacting quantitative data is required. In contrast to the assay method of the Northern analysis, at the heart of assay by nuclease protection is high stringency hybridization of target and probe

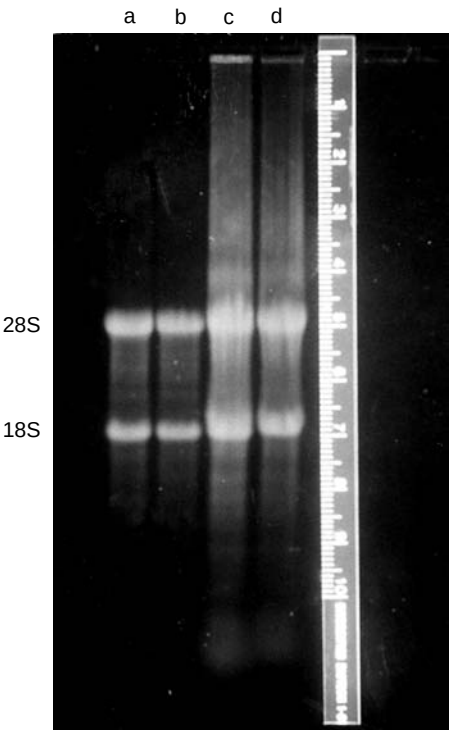

Fig. 1 Assessment of the integrity of RNA samples. The sharp definition of the 28S and 18S rRNA species in lanes a and b demonstrates the integrity of the sample. RNA in lanes c and d is also high quality, although an excessive amount of RNA was applied to these lanes. Lanes a and b: 20 μg of total cytoplasmic RNA prepared by NP-40 lysis. Lanes c and d: 25 μg of total cellular RNA (nuclear and cytoplasmic), prepared by guanidinium-acid-phenol extraction. Note the higher molecular weight nuclear RNA species in the sample. (From Farrell Jr., R.E. (1993) *RNA Methodologies: A Laboratory Guide for Isolation and Characterization*, Academic Press, San Diego, CA), p. 60. Reproduced with permission.

molecules, which are both free-floating in solution (solution hybridization) as opposed to having the target mRNA fixed on the filter paper (mixed phase hybridization). Driving solution hybridization are random molecular collisions, the kinetics of which are related directly to the total mass of nucleic acid in the reaction tube (probe + target + carrier = total mass). Presumably, all complementary nucleic acid molecules are capable of hybridization. The S1 nuclease protection assay and the ribonuclease (RNase) protection assay are methods of greatly enhanced sensitivity and resolution, and are universally considered to be more quantitative than Northern analysis.

The best nucleic acid probes for these assays are substantially shorter than the target mRNA. Upon molecular hybridization, a short double-stranded region is generated; the 5′ and 3′ regions of the target molecule flanking the double-stranded area remain single stranded. The enzyme S1 nuclease or a combination of RNases is then used to digest all nucleic acid molecules, which did not participate in nucleic acid hybridization. Only double-stranded nucleic acid molecules are resistant to nuclease attack and remain intact during incubation with these enzymes (Fig. 2). The resulting product of this assay, the so-called protected fragment, is then resolved by electrophoresis. By virtue of the mechanics of this assay, the size of the protected fragment is expected to be similar to the size of the probe sequence itself, which is often substantially shorter than the native RNA target. The protected fragment is then visualized

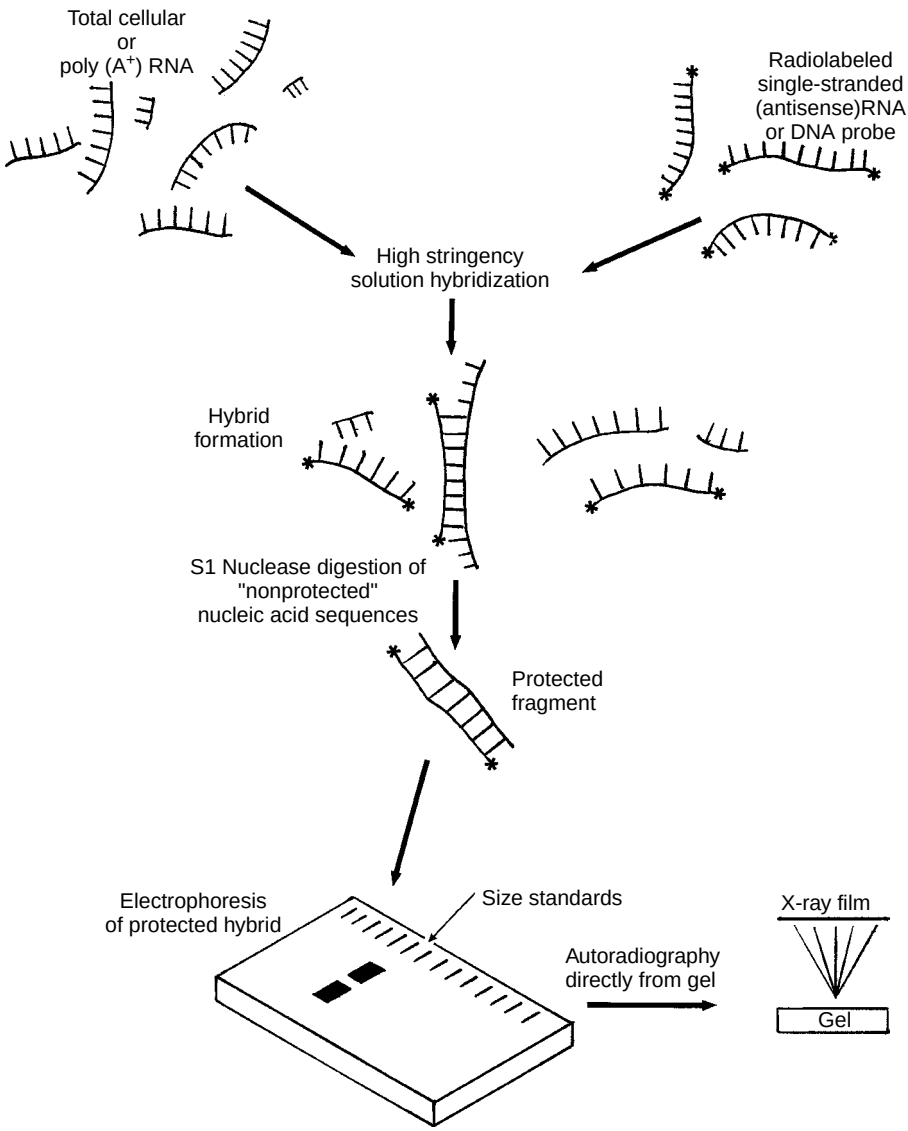

Fig. 2 Schematic of the S1 nuclease assay for the quantitation of specific RNA species. Purified RNA is hybridized in solution with a labeled, complementary sequence to form thermodynamically stable hybrid molecules. Any RNA and probe species that do not participate in the formation of hybrid molecules are digested away by the single-strand-specific nuclease S1, followed by electrophoresis of the intact hybrid molecules. In the case of an antisense RNA probe, an RNase cocktail can be used to produce the protected fragment (RNase protection assay). The size and abundance of protected RNAs are then deduced by autoradiography directly from the gel. (From Farrell Jr., R.E. (1998) *RNA Methodologies: A Laboratory Guide for Isolation and Characterization*, 2nd edition, Academic Press, San Diego, CA, p. 387). Reproduced with permission.

by autoradiography. As a direct result of solution hybridization and digestion of all nonhybridized nucleic acid molecules, the investigator can expect at least a tenfold enhancement in sensitivity compared to Northern analysis, particularly when performed using antisense RNA probes.

6.3 Dot Blot Analysis

The isolation of high quality RNA from tissue culture cells and whole cell samples is merely the first (although the most critical) step in the evaluation of a model system. Northern analysis, nuclease protection analysis, conversion into cDNA, and so forth can be a time-consuming and expensive road to travel and should not immediately be deemed necessary. When evaluating a model system, cell type, or experimental regimen for the first time, it may be worthwhile to quantify mRNAs of interest by dot blot analysis. This simple technique allows the investigator to make definitive statements about the biochemical composition of a sample without investing excessive man-hours and laboratory resources.

In order to dot blot RNA, denatured samples are applied directly onto a membrane under vacuum using a multiwell dot blot (Fig. 3) filtration manifold. Samples are immobilized onto the surface of the membrane, followed by nucleic acid hybridization. Thus, dot blots and a closely related variant known as *slot blots*, permit rapid detection of the relative amount of a particular RNA in a sample. Salient information can be obtained from purified RNA samples or whole cell lysates without electrophoresis, cDNA synthesis, or synthesis mediated by the polymerase chain reaction. These methods are reserved for later, more exacting analysis of a system after preliminary information has been derived.

Dot blotting is an excellent method for the analysis of gene expression when large numbers of samples are to be evaluated simultaneously, such as an experiment requiring numerous time points. If sample dilutions are desired, they may be arranged either vertically or horizontally. The degree of hybridization is then assessed using image analysis software.

Dot blot analysis yields purely quantitative data. The main drawback of this

Fig. 3 Minifold I dot blot apparatus. Sample dilutions are applied under vacuum directly to the surface of the filter membrane resting beneath the faceplate. (Photo courtesy of Schleicher & Schuell, Inc., Keene, NH).

approach is that it lacks the qualitative component, which accompanies electrophoresis. In order to be truly reliable, dot blot analysis must include excellent positive and negative controls in order to demonstrate hybridization specificity and to gauge nonspecific probe binding to the filter membrane. Moreover, good internal controls are always in order: one should observe equally intense signals from wells into which equal amounts of positive control target was applied. When attempting this type of blot analysis for the first time or with a new system, it is strongly suggested that dilutions of the positive control target material be made in order to determine the linear range of the assay. For example, it would be useless, quantitatively speaking, if the hybridization signals were too intense to be accurately measured on X-ray film, which also has a defined linear range.

6.4 Transcription Rate Assays

The modulation of key regulatory molecules is an integral cellular response to both intracellular and extracellular challenge. One fundamental goal in the assessment of any biological model system is the elucidation of the level of gene modulation. While potential levels of regulation are infinite, they are broadly categorized as transcriptional or due to some posttranscriptional event. The initial characterization of these systems commonly involves isolation, hybridization, and subsequent detection of specific RNA species by dot blot, Northern blot, and nuclease protection analysis. While these approaches may furnish excellent qualitative and quantitative data with respect to steady-state levels of message, that is, the final accumulation of RNA in the cell, RNA prepared by total cellular lysis does not provide information about the rate of transcription, the intracellular compartmentalization (nuclear or cytoplasmic) of the RNA under investigation, or the translatability the RNA in the cytoplasm. Knowledge of this aspect of the cellular biochemistry is necessary to elucidate the level of gene regulation because many mRNAs naturally have much longer half-lives than others, and because the half-life of many mRNA species can be modified in response to a particular xenobiotic regimen or environmental stimulus.

In order to address these questions, two basic approaches have been employed to study the mechanism of transcription and the processing of the resulting transcripts in eukaryotic cells. In one approach, the rate of transcription is measured in intact nuclei by the incorporation of labeled precursor nucleotides into RNA transcripts initiated on endogenous chromatin at the time of nuclear isolation. Elongated, labeled nuclear RNA is then purified for hybridization to complementary, membrane-bound DNA sequences. This widely used technique is known as the *nuclear runoff assay* (Fig. 4). It is a superbly sensitive method for measuring transcription rate as a function of cell state. Because the RNA transcripts are radiolabeled, rather than the probes used to quantify their abundance, the basic format of this assay may be likened to a "reverse dot blot" since the probe is membrane-bound and nonradiolabeled. Another technique utilizes specific cloned DNA fragments as templates for transcription and is used to dissect the factors (cellular, ionic, etc.) and sequences necessary for faithful transcription *in vitro* using whole cell extracts.

The principal advantage of the nuclear runoff assay is that labeling occurs whilst

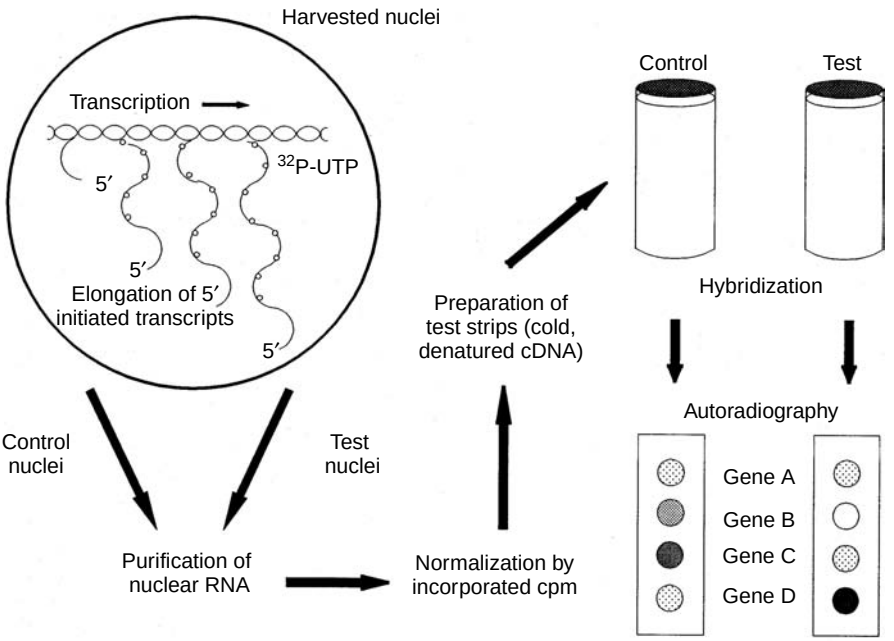

Fig. 4 Nuclear runoff assay. Relative rate of transcription of all genes can be assessed by incubation of intact nuclei with an NTP cocktail containing labeled UTP. Elongated, radiolabeled transcripts are then hybridized to cold cDNA probes immobilized on a nylon filter. Upon autoradiography, the intensity of the signal from each dot is indicative of the degree of label incorporation and thus the relative rate of transcription of specific genes under a define set of experimental conditions. (From Farrell Jr., R.E. (1998) *RNA Methodologies: A Laboratory Guide for Isolation and Characterization*, 2nd edn., p. 411.) With permission from Elsevier.

maintaining the natural geometry of the transcription apparatus. The mechanics and reaction conditions of the nuclear runoff assay promote the elongation of initiated transcripts, but are not believed to support new initiation events. The degree of labeling of any particular RNA species (and hence, the relative transcription rate of specific genomic sequences) may then be assessed by Cerenkov counting (a specific type of radioactive detection) coupled with autoradiography. These data correlate directly with the number of RNA polymerase molecules engaged in transcribing a specific gene and indirectly with the transcriptional efficiency of regulatory sequences associated with the gene under a defined set of experimental conditions. When used in conjunction with the Northern blot analysis of cytoplasmic RNA species (above), data from the nuclear runoff assay may be used to assess whether observed gene modulation is a result of a change in the synthesis (transcriptional control) or a change in splicing/nucleocytoplasmic transport/mRNA stability (posttranscriptional control).

The nuclear runoff assay permits the simultaneous analysis of several genes, all of which are presumably transcribed in the same relative amounts in isolated nuclei as in intact cells. The most critical parameter by far is the preparation of nuclei prior to labeling: success of the nuclear runoff assay is almost entirely

dependent upon the speed with which nuclei are harvested and radiolabeled with precursor UTP. Failure to generate high specific activity RNA probe is usually a direct result of inexperienced handling of the nuclei prior to the labeling step. Nuclei are most often isolated by incubation of cells in NP-40-containing, hypotonic lysis buffer, in iso-osmotic sucrose buffer containing Triton X-100, or by nonaqueous methods. Nuclear purification must also be carried out in such a way as to preserve RNA polymerase activity and nuclear structure during the isolation from cells cultured *in vitro* or, if absolutely necessary, from tissue. Harvested nuclei are then labeled immediately or may be stored frozen in liquid nitrogen or at $-70°$ for several months without significant loss of labeling potential. As with the standard nuclease protection assay format, hybridization between probe and target sequences occurs entirely in solution and for that reason constitutes a very quantitative assay.

6.5 Synthesis of cDNA

The synthesis of complementary DNA (cDNA) is an enzymatic, *in vitro* method by which RNA (usually mRNA) is converted stepwise into single- or double-stranded cDNA. First-strand cDNA is synthesized directly from naturally occurring mRNA templates by an enzyme known as *reverse transcriptase*. Thus, the synthesis of cDNA is dependent on the integrity of the mRNA template and precisely complementary to it. Because nucleic acids are only synthesized $5' \rightarrow 3'$, a primer must be provided, which is complementary to sequences near or at the $3'$ end of the RNA template in order to support reverse transcriptase-mediated nucleic acid synthesis. The second strand cDNA can be synthesized only if the original RNA is removed from the RNA:cDNA hybrid. This may be accomplished using older chemical methods (incubation at alkaline pH), enzymatically through the combined activities of RNase H and DNA polymerase I, or by RT-PCR (below).

As a consequence of the very labile character of RNA, the synthesis of stable, double-stranded cDNA is a very important method for amplification of sequences and their long-term storage. Only those transcribed RNAs, which are purified from the cell can be converted into cDNA. Different tissues, even from the same biological source, will yield different cDNA products; as such, cDNA is best thought of as a permanent biochemical record of the cell. cDNA is a means by which the cellular biochemistry can be studied in great detail over a period of months or years, much longer than the storage life of RNA in purified form. An in-depth discussion of the mechanics of cDNA synthesis and cloning is found elsewhere in this volume.

6.6 RT-PCR

The polymerase chain reaction (PCR) is a primer-mediated, enzymatic method for the quasiexponential amplification of nucleic acid sequences. This revolutionary method exploits the natural $5' \rightarrow 3'$ polymerase activity of various thermostable DNA polymerase, among which *Taq* DNA polymerase is the quintessential example. Two short oligonucleotides function as nucleic primer sequences; their orientation facing each other supports the enzymatic amplification of that sequence, which lies between (is framed by) their respective $5'$ ends.

RT-PCR is a twofold process. First, high-quality RNA must be converted to first-strand cDNA by reverse transcription. The components and mechanics of this reaction are nearly identical to any other first-strand cDNA synthesis reaction. Second, the products of the first strand synthesis reaction are diluted in buffer, which provides all of the cofactors necessary to support amplification of these products by PCR (Newer methods for performing RT-PCR require only one enzyme in a single reaction tube format (one-tube RT-PCR)). The resulting cDNA synthesis products are thus amplified as would be DNA from any other source. This approach is predicated upon the availability of an appropriate set of primers. PCR-amplified cDNA products can then be cloned, subcloned, and otherwise manipulated or quantified in order to characterize some aspect of the cellular biochemistry. In addition to its obvious utility for the quantification of gene expression, the judicious design of primers permits RT-PCR to be used to map the 5′ and 3′ ends of transcripts (5′ RACE and 3′ RACE, respectively), detect alternative splicing of transcripts, and identify induced and repressed genes under a defined set of environmental conditions (subtraction–suppression PCR).

6.7 RNAi

A remarkable new tool has emerged that is able to suppress expression of specific endogenous genes through the use of double-stranded RNA (dsRNA). In the few short years since the demonstration of the power of this technique in *Caenorhabditis elegans* and subsequently in human and other mammalian cells, refinements in the methodology have had the same level of impact on research that PCR proffered. This new and emerging technology, known as *RNA interference* (RNAi). RNA interference is a patented process. Commercial use this process may require licensing form the Carnegie Institution of Washington. For information contact the Director of Administration and Finance, Carnegie Institution of Washington, 1530 P Street, N.W., Washington, DC 20005. Telephone: 202-939-1118. RNAi is also frequently referred to as *gene knockdown* or *posttranscriptional gene silencing*.

RNAi is a natural phenomenon that was considered an oddity when it was first observed in petunias. It is now known to exist in many organisms as a means of protecting against viruses and other molecular invaders that would otherwise plague a host genome and wreak havoc. This method of protecting the integrity of the eukaryotic genome is highly evolutionarily conserved, and dsRNA entering the cell is targeted for immediate destruction. Natural processes mediated by RNAi certainly include the turnover of both wild-type and mutant mRNAs and translational regulation during the development of an organism, and undoubtedly other regulatory mechanisms in the cell that have not yet come to light.

RNAi is an endogenous catalytic pathway that is triggered by dsRNA. The ''trigger'' can occur naturally, as in the case of a cellular infection by a dsRNA virus, or by the intentional introduction of dsRNA to induce user-directed degradation of the homologous transcript(s). The net result of RNAi is the downregulation of specific genes by destruction of their mRNA(s). This method of studying the effects of gene expression in a cell or in an organism is occasionally loosely referred to as *reverse genetics*, the goal of which is to determine the consequences for a cell when a protein

is *not* produced. This has profound ramifications in the realm of developmental biology, the treatment of infectious diseases, and other diseases that result from inappropriate protein expression (gain-of-function mutations). The major strength of RNAi is that it permits the study of the function of one gene at a time, and does so over an extended period.

RNAi has emerged as an important mainstream tool for both basic and applied research, and as already revolutionized what has become known as *functional genomics*. RNAi is ubiquitous in eukaryotes and currently is the preferred tool for investigating the regulation of gene expression in plants, animals, and fungi. This method is becoming increasingly popular because of its wide-ranging applicability in research owing to its compatibility with cell culture as well as *in vivo* models. Comparatively speaking, RNAi is much faster and much more economical than creating knockout animals in order to study the function of specific genes. The resulting precision silencing of specific genes also makes RNAi an attractive platform for the discovery and development of life-saving pharmaceuticals. There are multiple approaches by which RNAi can be induced, each of which has several mechanistic permutations. RNAi is, fundamentally, a two-step process (Fig. 5). The first step involves one of the master enzymes in the RNAi process, a type III endoribonuclease aptly named *Dicer*. This enzyme, a ubiquitous member of the eukaryotic proteome, is involved in the digestion of long dsRNA into 21- to 23-bp short inhibitory RNA (siRNA) molecules with characteristic 3′ dinucleotide overhangs on both strands. In addition, the siRNA manifest a 5′-phosphate and a 3′-OH terminus. The long dsRNA can be introduced

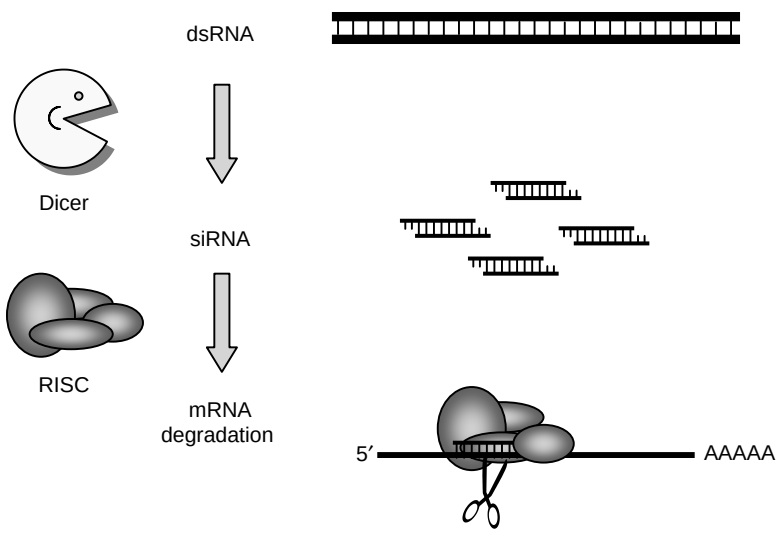

Fig. 5 Major steps in the RNAi process. dsRNA from any of a number of sources and in any of a number of configurations is cut by the enzyme Dicer into siRNA, which, in turn, become part of the multicomponent RISC. This ultimately leads to destruction of the target mRNA and concomitant downregulation of the associated gene. (From Farrell Jr., R.E. (2005) *RNA Methodologies: A Laboratory Guide for Isolation and Characterization*, 3rd edn., p. 607.) With permission from Elsevier.

into the cell where it will become a substrate for Dicer; otherwise, recombinant Dicer can be used to generate siRNA *in* vitro, after which these short molecules are introduced into the cell by transfection. In the second step, siRNA, regardless of the source, becomes part of a multicomponent nuclease-containing RNA-induced silencing complex (RISC). Part of RISC is an ATP-dependent helicase that unwinds the double-stranded siRNA, the antisense component of which is now able to base pair with the mRNA to be silenced. The formation of a double-stranded region between the antisense component of the silencing RNA and the cognate mRNA dooms it. RISC cuts the mRNA close to the middle of this transiently formed double-stranded region. The mRNA is then further degraded, preventing any level of interaction with the cellular translation machinery. The subtleties of RNAi are described in much greater detail elsewhere in this volume.

6.8 *In Vitro* Translation

The classical methods for the analysis of RNA include all of the standard techniques described above and many others. Another aspect of the RNA story, however, is the destiny of those transcripts, which emerge as mature mRNA molecules in the cytoplasm. mRNA biogenesis guarantees neither translation nor translatability; a gene is really not "expressed" until a functional peptide is produced. In order to more fully characterize gene expression, the extent of translation may be assessed by Western analysis, and the resulting data correlated to the transcriptional activity of the corresponding gene(s).

The translation of RNA into protein *in vitro* goes hand in hand with message quantification as a parameter of gene expression. Briefly, naturally synthesized mRNA or message synthesized by *in vitro* transcription is added to a whole cell lysate containing all of the components needed to support the *in vitro* translational of mRNAs added to the system. Common *in vitro* translation systems include the reticulocyte lysate system and the wheat germ extract system, and, in some cases, microinjection of message directly into a living cell. Pretreatment of the *in vitro* translation system lysates with micrococcal nuclease destroys endogenous mRNA, ensuring that the synthesis of *de novo* protein products harvested from the system was supported exclusively by the introduction of experimental mRNA. The same *in vivo* requirements for 5′ cap, initiation codon, and polyadenylation apply here if experimental introduced message is to be translated with any efficiency at all. Protein products are then analyzed as usual by Western analysis, biological assay, immunopurification, and so forth. This particular aspect of biotechnology is an attractive technique in a number of different applications, including the study of naturally occurring animal and plant mRNAs, transfection experiments, characterization of mRNA products, protein engineering, screening and analysis of mutants, and DNA, RNA, or protein-binding studies.

7 Summary

The characterization of ribonucleic acid (RNA) from biological sources is a central act of profound significance in molecular biology. Gene expression is frequently analyzed by blot analysis, by solution hybridization methods such as nuclease protection assay, and by the synthesis and

study of cDNA by RT-PCR. More and more, many of these classical methods are becoming streamlined or even circumvented by utilization/exploitation of the polymerase chain reaction. While no one would argue against the fact that PCR has revolutionized many aspects of biotechnology, it is important to realize that PCR is not a reason, but a tool used to achieve an end. In the absence of a clear understanding of experimental goals and, certainly, without high quality RNA as starting material, data intended to reflect the cellular biochemistry and provide insight into gene regulation are hardly reliable.

See also DNA Libraries; RNA Secondary Structures.

Bibliography

Books and Reviews

Blackburn, P., Moore, S. (1982) in: Boyer, P.D. (Ed.) *The Enzymes*, 3rd edition, Academic Press, New York, pp. 317–433.

Cooper, T.G. (1977) *The Tools of Biochemistry*, John Wiley & Sons, Inc., New York, pp. 309–354.

Farrell Jr., R.E. (1993) *RNA Methodologies: A Laboratory Guide for Isolation and Characterization*, Academic Press, San Diego, CA.

Farrell Jr., R.E. (1998) *RNA Methodologies: A Laboratory Guide for Isolation and Characterization*, 2nd edition, Academic Press, San Diego, CA.

Farrell Jr., R.E. (2005) *RNA Methodologies: A Laboratory Guide for Isolation and Characterization*, 3rd edition, Academic Press, San Diego, CA.

Frohman, M.A. (1995) Rapid Amplification of cDNA Ends, in: Dieffenbach, C.W., Dveksler, G.S. (Eds.) *PCR Primer*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Laskey, R.A. (1990) Radioisotope Detection Using X-ray Film, in: Slater, R.J. (Ed.) *Radioisotopes in Biology: A Practical Approach*, IRL Press, Washington, DC, pp. 87–108.

Lewin, B. (Ed.) (2004) *Genes VIII*, Pearson Education, Inc., Upper Saddle River, NJ.

Marzluff, W.F., Huang, R.C.C. (1984) Transcription of RNA in Isolated Nuclei, in: Hames, B.D., Higgins, S.J. (Eds.) *Transcription and Translation: A Practical Approach*, IRL Press, Washington, DC, pp. 89–129.

Sambrook, J., Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Primary Literature

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., Rutter, W.J. (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease, *Biochemistry* **18**, 5294.

Chomczynski, P., Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction, *Anal. Biochem.* **162**, 156.

Elbashir, S.M., Lendeckel, W., Tuschl, T. (2001b) RNA interference is mediated by 21- and 22-nucleotide RNAs, *Genes Dev.* **15**, 188–200.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T. (2001a) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells, *Nature* **411**, 494–498.

Ernest, M.J., Schultz, G., Feigelsen, P. (1976) RNA synthesis in isolated hen oviduct nuclei, *Biochemistry* **15**, 824.

Favaloro, J., Triesman, R., Kamen, R. (1980) Transcriptional maps of polyoma virus-specific RNA: analysis by two-dimensional S1 gel mapping, *Methods Enzymol.* **65**, 718.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C. (1998) Potent and specific interference by double-stranded RNA in *Caenorhabditis elegans*, *Nature* **391**, 806–811.

Frohman, M.A., Dush, M.K., Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer, *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8998.

Glisin, V., Crkvenjakov, R., Byus, C. (1974) Ribonucleic acid purified by cesium chloride centrifugation, *Biochemistry* **13**, 2633.

Hammond, S.M., Bernstein, E., Beach, D., Hannon, G.J. (2000) An RNA-directed nuclease mediates posttranscriptional gene silencing in *Drosophila* cells, *Nature* **404**, 293–296.

Lund, E., Paine, P.L. (1990) Nonaqueous isolation of transcriptionally active nuclei from *Xenopus oocytes*, *Methods Enzymol.* **181**, 36.

Napoli, C., Lemieux, C., Jorgensen, R. (1990) Introduction of a chalcone synthase gene into *Petunia* results in reversible cosuppression of homologous genes in trans, *Plant Cell* **2**, 279–289.

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A., Lukyanov, S.A. (1995) An improved method for walking in uncloned genomic DNA, *Nucleic Acids Res.* **23**, 1087–1088.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis, *J. Mol. Biol.* **98**, 503.

van der Krol, A.R., Mur, L.A., Beld, M., Mol, J.N., Stuitje, A.R. (1990) Flavonoid genes in Petunia: addition of a limited number of gene copies may lead to a suppression of gene expression, *Plant Cell* **2**, 291–299.

Zamore, P.D., Tuschl, T., Sharp, P.A., Bartel, D.P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals, *Cell* **101**, 25–33.

Zarlenga, D.S., Gamble, H.R. (1987) Simultaneous isolation of preparative amounts of RNA and DNA from *Trichinella spiralis* by cesium trifluoroacetate isopycnic centrifugation, *Anal. Biochem.* **162**, 569.

RNA Secondary Structures

Ivo L. Hofacker[1] *and Peter F. Stadler*[1,2]
[1] *University of Vienna, Wien, Austria*
[2] *University of Leipzig, Leipzig, Germany*

1 Introduction 583

2 Basic Concepts 585
2.1 Representations of Secondary Structures 585
2.2 Combinatorics and Typical Structures 585
2.3 Energy Model 587
2.4 Pseudoknotted Structure 588

3 RNA Structure Prediction 589
3.1 Structure Prediction by Energy Minimization 589
3.2 Suboptimal Folding and Pair Probabilities 590
3.2.1 Available Programs and Web Services 591
3.2.2 Prediction Accuracy 591
3.3 Well-defined Regions and Reliability 591
3.4 Structure Prediction using Sequence Covariation 592
3.4.1 Search for ncRNA 594

4 Current Research in RNA Secondary Structures 595

Bibliography 597
Books and Reviews 597
Primary Literature 598

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

Keywords

Base Pair
Base pairs are formed by specific hydrogen bonds between two nucleobases. The canonical Watson–Crick pairs between guanosine and cytosine (GC), and between adenosine and uracil (AU), as well as the wobble pairs between guanosine and uracil (GU) regularly appear in RNA structures. The nucleobases of the two paired nucleotides form a planar structure.

Dynamic Programming
A general computational technique that is based on solving a combinatorial optimization problem recursively, while tabulating intermediate results. In the case of RNA folding, this means composing large structures from the foldings of shorter subsequences.

Helix or Stem
Helices or stems are formed by stacking base pairs on top of each other. The interaction of the parallel planar arrangements of the nucleobases stabilizes the three-dimensional structure. The geometry of helices is nearly sequence independent, as long as they contain only canonical base pairs, whence energy contributions of stacked pairs are nearly independent of each other.

Loops
Loops are formed by those parts of the molecule that are not part of a helix. Different types of loops (bulges, interior loops, and multibranch loops) can be distinguished depending on their structure. Loops destabilize the structure. In the context of RNA folding algorithms, one considers adjacent stacked pairs as a special type of (stabilizing) loop. In this language, each secondary structure can be decomposed completely into loops, which form the basis of the empirical energy model.

Noncoding RNA
An RNA molecule with any function other than encoding proteins.

Pseudoknot
A structural feature composed of crossing base pairs. Pseudoknots are in most cases excluded from computational approaches for technical reasons.

RNomics
The emerging science of RNA molecules and their interactions in the cell.

Secondary Structure
The secondary structure is the collection of *base pairs* or, equivalently, of *loops* and *helices*. In a stricter sense, secondary structures are defined as crossing-free collections of base pairs. All other base pairs are then considered as part of a higher-order (tertiary) structure.

The three-dimensional structures of nucleic acids, RNA, and DNA, are dominated by double-helical regions that are formed by canonical Watson–Crick and wobble (GU) base pairs, which collectively are referred to as the *secondary structure of the molecule*. In the case of RNA, this level of description is of particular interest because it captures the thermodynamics of RNA folding quite well and allows a self-consistent description of folding dynamics. Furthermore, secondary structures are often evolutionarily well conserved, evolving much slower than the underlying sequences. From a bioinformatics perspective, RNA secondary structures are a convenient representation because efficient exact algorithms are known to enumerate the structures that can be formed by a given sequence, to solve the folding problem, and to compute the partition function (and hence any thermodynamic quantity of interest) from a well-measured set of empirical energy parameters.

1 Introduction

Secondary structures, that is, patterns of base pairing, not only form the most important distinguishing characteristics of the various classes of RNA molecules but they also provide us with a unique coarse-grained level of description that naturally lends itself to computational studies of RNA structure and evolution.

RNA has moved from a fringe existence to the center stage of research in molecular biology following the discovery of novel classes of small regulatory RNAs. While textbooks still speak of *genes and their encoded protein products*, thousands of human genes produce transcripts that exert their function without ever producing proteins. The list of functional noncoding RNAs (ncRNAs) includes key players in the biochemistry of the cell. Many of them have characteristic secondary structures that are highly conserved in evolution. Databases (referenced in the Table 1) collect the most important classes.

In addition, there is a diverse and growing list of ncRNAs with sometimes enigmatic function. We give just a few examples: The 17-kb *Xist* RNA of humans and the smaller roX RNAs of Drosophila play a key role in dosage compensation and X-chromosome inactivation. Several large ncRNAs are expressed from imprinted regions. Many of these are cis-antisense RNAs that overlap coding genes on the other genomic strand. An RNA (meiRNA) regulates the onset of meiosis in fission yeast. Human vaults are cytoplasmic ribonucleoprotein particles believed to be involved in multidrug resistance. The complex contains several small untranslated RNA molecules. No precise function is known at present for the human H19 transcript, the hrsω transcript induced by heat shock in Drosophila, or the *Escherichia coli* 6S RNA.

Even though the sequence of the human DNA is known by now, the contents of about half of it remains unknown. The diversity of sequences, sizes, structures, and functions of the known ncRNAs strongly suggests that we have seen only a small fraction of the functional RNAs. Most of the ncRNAs are small, they do not have translated ORFs, and they are not polyadenylated. Unlike protein-coding genes, ncRNA gene sequences do

Tab. 1 Major classes of functional RNAs

Type	*Size*	*Organisms*	*Function* –
miRNA	22	Multicellular?	RNA interference
tRNA	70–80	All	Transfer RNA
rRNA 5S	130	All	Part of ribosome
RNase P	260–360	All	tRNA maturation
tmRNA	300–400	Bacteria	Tags protein for proteolysis
snRNA	50–300	Eukaryotes	Part of spliceosome
snoRNA	80–600	Eukaryotes Archea	RNA modification
SRP RNA	300–400	All	Signal recognition particle
rRNA 16S	1500	All	Part of ribosome
rRNA 23S	3300	All	Part of ribosome

not seem to exhibit a strong common statistical signal; hence, a reliable general-purpose computational genefinder for ncRNA genes has been elusive. It is quite likely therefore that a large class of genes has gone relatively undetected so far because they do not encode proteins.

Another level of RNA function is presented by functional motifs within protein-coding RNAs. We list a few of the best-understood examples of structurally conserved RNA motifs in viral RNAs:

- An *IRES* (internal ribosomal entry site) region is used instead of a CAP to initialize translation by Picornaviridae, some Flaviviridae including Hepatitis C virus, and a small number of mRNAs.
- The TAR hairpin structure in HIV and related retroviruses is the target for viral transactivation.
- The RRE structure of retroviruses serves as a binding site for the Rev protein and is essential for the viral replication. The RRE is a characteristic five-fingered structural motif.
- The CRE hairpin in Picornaviridae is vital for replication.

A textbook example of a functional RNA secondary structure is the *Rho*-independent termination in *E. coli*. The newly synthesized mRNA forms a hairpin in the 3′UTR that interacts with the RNA polymerase causing a change in conformation and the subsequent dissociation of the enzyme–DNA–RNA complex.

Only part of the mature mRNA is translated into protein. At the 5′ terminus of the mRNA, just behind the cap, is a noncoding sequence, the so-called leader sequence (10–200 nt) that may be followed by another noncoding sequence of up to 600 nt. An increasing number of functional features in the untranslated regions of eukaryotic mRNA have been reported in recent years.

An extreme example is the Early Noduline gene. Enod40, which is expressed in the nodule primordium developing in the root cortex of leguminous plants after infection by symbiotic bacteria, codes for an RNA of about 700 nt that gives rise to two short peptides of 13 and 27 amino acids, respectively. The RNA structure itself exhibits significant conservation of secondary structure motifs, and might take part in localization of mRNA translation,

as in the case of the bicoid gene *bcd* of Drosophila.

2 Basic Concepts

2.1 Representations of Secondary Structures

A *secondary structure* Ψ is a special type of contact structure, represented by a list of base pairs (i, j) with $i < j$ on a sequence x, such that for any two base pairs $[i, j]$ and $[k, l]$ with $i \leq k$ holds:

1. $i = k$ if and only if $j = l$, and
2. $k < j$ implies $i < k < l < j$.

The first condition simply means that each nucleotide can take part in at most one base pair. The second condition forbids knots and pseudoknots. While pseudoknots are important in many natural RNAs, they can be considered part of the tertiary structure for our purposes. We will therefore neglect them for the most part of this presentation. The restriction to knot-free structures is necessary for the efficient dynamic programming algorithms discussed below.

The two conditions above imply that secondary structures form a special type of graph. In particular, a secondary structure graph is *subcubic* (i.e. the vertex degree is at most three) and *outer-planar*. The latter property means that the structure can be drawn in the plane in such a way that all vertices (which represent the nucleotides) are arranged on a circle (the molecule's backbone), and all edges (which represent the bases pairs) lie inside the circle and do not intersect, see Fig. 1.

Secondary structures can also be stored compactly in strings consisting of dots and matching brackets: For any pair between positions i and j ($i < j$) we place an open bracket "(" at position i and a closed bracket ")" at j, while unpaired positions in the molecule are represented by a dot ("."), see Fig. 1c. Since base pairs may not cross, the representation is unambiguous.

Although secondary structure graphs are always planar, it is not trivial to find nonoverlapping and visually pleasing layouts for large structures. A representation that works well for large structures and is well suited for comparing structures is the so-called mountain representation. In the mountain representation, a single secondary structure is represented in a two-dimensional graph, in which the x-coordinate is the position k of a nucleotide in the sequence and the y-coordinate is the number $m(k)$ of base pairs that enclose nucleotide k.

Often it is desirable to present not just one structure, but an ensemble of structural possibilities. Mountain plots can be adapted for structural ensembles by using the *mean* number of base pairs $m(k)$. A more detailed representation is given by so-called dot plots, where each possible pair (i, j) is represented by a dot on square grid. The size or color of the dot is used to indicate the probability p_{ij} of the pair, or the best possible energy of a structure containing this pair. Dot plots can provide an excellent overview of possible alternative foldings. Fig. 4 gives examples of dot plots and mountain plots.

2.2 Combinatorics and Typical Structures

The basic studies into the combinatorics of RNA secondary structures goes back to the work of M. Waterman in the late 1970s.

We begin our exposition by counting the secondary structures that can be formed by a given sequence $x = (x_1, x_2, \ldots, x_n)$

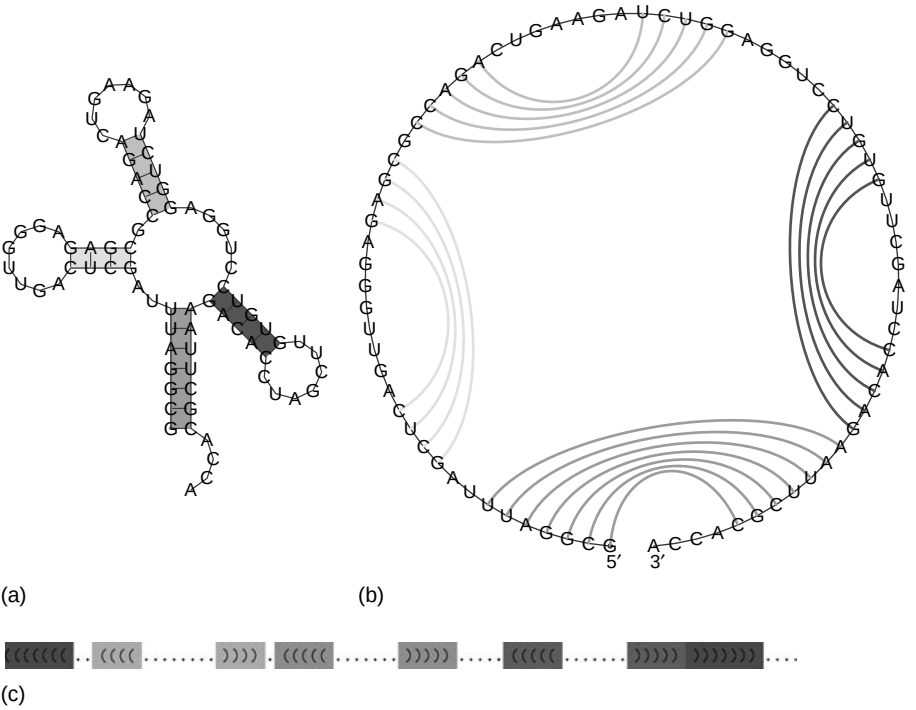

Fig. 1 Secondary structure of phenylalanine-tRNA from yeast as (a) conventional drawing, (b) in circular representation, and (c) in dot-bracket notation. Colors are used to highlight pairs belonging to the same helix. Note that the chords in the circular representation must not cross in secondary structure graphs. (See color plate p. xl).

of length n. We will simply write "(i, j) pairs" to mean that the nucleotides x_i and x_j *can* form a Watson–Crick or a wobble pair, that is, $x_i x_j$ is one of **GC**, **CG**, **AU**, **UA**, **GU**, or **UG**. The basic idea behind all dynamic programming algorithms for RNA folding is the observation that a structure on n nucleotides can be formed in only two distinct ways from shorter structures: Either a structure on $n-1$ nucleotides is extended by an unpaired base, or the nth nucleotide is paired. In the latter case, it has a pairing partner, say j such that the (j, n) pair encloses a secondary structure on the subsequence from $j+1$ to $n-1$ since base pairs must not cross by definition. The remainder, the interval from 1 to $j-1$ is of course also a secondary structure:

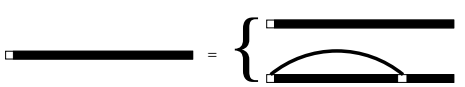

It is now easy to compute the number N_{ij} of secondary structures on the subsequence $x[i..j]$ from positions i to j:

$$N_{ij} = N_{i+1,j} + \sum_{k,(i,k)\text{pairs}} N_{i+1,k-1} N_{k+1,j} \tag{1}$$

The first term accounts for the case in which position i is unpaired, the

terms in the sum consider the base pairs from i to some position k. Because of the "no-pseudoknots" condition, both the part of the sequence that is enclosed by the pair (i, k) and the part beyond the base pair form secondary structures that are completely independent of each other: thus, we may simply multiply their numbers.

The counting recursion can also be used to produce uniformly distributed random structures, or adapted to derive typical structural features of RNA molecules such as expected helix length or distribution of loop types. One should note, however, that such random structures differ significantly from the typical structures obtained from folding random or natural sequences. In contrast to folded structures, random structures contain many loops connected by short helices or even isolated base pairs. Qualitative features, such as an average length of helical regions independent of the sequence lengths or the distribution of branching orders of loops is shared between random and biological structures.

2.3 Energy Model

The physicochemical basis for the coarse-grained secondary structure model is the fact that a dominating part of the energy of structure formation is contained in the stacking of the aromatic nucleobases that gives rise to helical base-paired regions with a spatial structure that can be considered as almost sequence independent. This distinction between paired and unpaired regions allows us to approximate the free energy of structure formation given the sequence and the list of base pairs. It is important to note that a secondary structure as defined in the previous section corresponds to an *ensemble* of conformations of the molecule at atomic resolution restricted to a certain base pairing (hydrogen bonding) pattern. For example, no information is assumed about the spatial conformation of unpaired regions. The entropic contributions of these restricted conformations have to be taken into account, and hence we are dealing with (temperature dependent) free energies.

This free energy of an RNA secondary structure is assumed to be the sum of the energy contributions of all "loops," that is, the faces of the planar drawing of the structure. This decomposition has a solid graph theoretical foundation: the loops form the unique minimal cycle basis of the secondary structure graph. More importantly, however, a large number of careful melting experiments have shown that the energy of structure formation (relative to the random coil state) is indeed additive to a good approximation. Usually, only Watson–Crick (**AU**, **UA**, **CG**, and **GC**) and wobble pairs (**GU**, **UG**) are allowed in computational approaches since nonstandard base pairs have, in general, context-dependent energy contributions that do not fit into the "nearest-neighbor model". Individual nonstandard base pairs are therefore treated as special types of interior loops in the most recent parameter sets.

Qualitatively, there are two major energy contributions: stacking of base pairs and loop entropies. Stacking energies can be computed for molecules in the vacuum by means of standard quantum chemistry approaches. The secondary structure model, however, considers only energy differences between folded and unfolded states in an aqueous solution with rather high salt concentrations. As a consequence, one has to rely on empirical energy parameters. Loops are destabilizing: the closing base pair restricts the possible conformations

of the sequence in the loop relative to the conformations that could be formed by the same sequence segments in a random coil, resulting in an entropy loss and thus an increase in free energy.

A collection of energy parameters is maintained by the group of David Turner. These standard parameters are measured in a buffer of 1 M NaCl at 37 °C. Since both entropies and enthalpies are available, the parameters can be extrapolated to other temperatures. As examples we list the free energies for stacked pairs in Table 2. Note that a single additional base pair can stabilize a structure by up to -3.4 kcal mol^{-1}.

2.4 Pseudoknotted Structure

The definition of secondary structure relegates pseudoknots to the realm of a tertiary structure. Often this is justified purely by convenience, since the simple dynamic programming algorithms, which will be described in some detail in the following section, cannot handle pseudoknotted structures. It is worth mentioning, however, that many pseudoknotted base pairing patterns are not sterically feasible, while any knot-free secondary structure can be realized in 3D.

Tab. 2 Free energies for stacked pairs in kcal mol^{-1}.

	CG	GC	GU	UG	AU	UA
CG	–2.4	–3.3	–2.1	–1.4	–2.1	–2.1
GC	–3.3	–3.4	–2.5	–1.5	–2.2	–2.4
GU	–2.1	–2.5	–1.3	–0.5	–1.4	–1.3
UG	–1.4	–1.5	–0.5	0.3	–0.6	–1.0
AU	–2.1	–2.2	–1.4	–0.6	–1.1	–0.9
UA	–2.1	–2.4	–1.3	–1.0	–0.9	–1.3

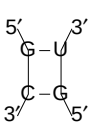

–2.5 kcal mol^{-1}

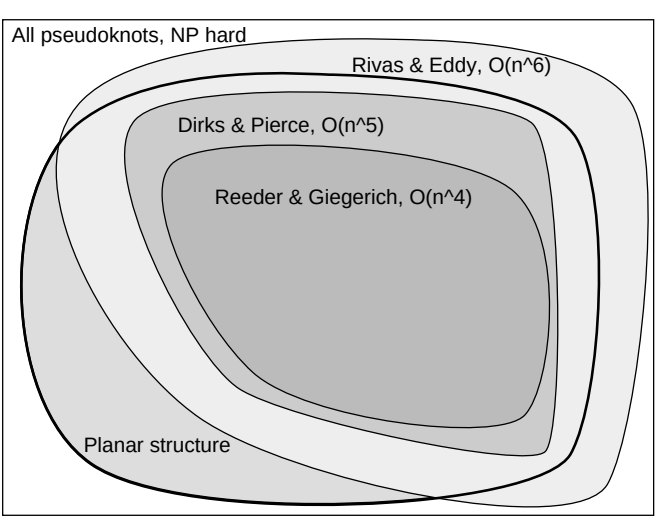

Fig. 2 Classes of allowed pseudoknots in different algorithms.

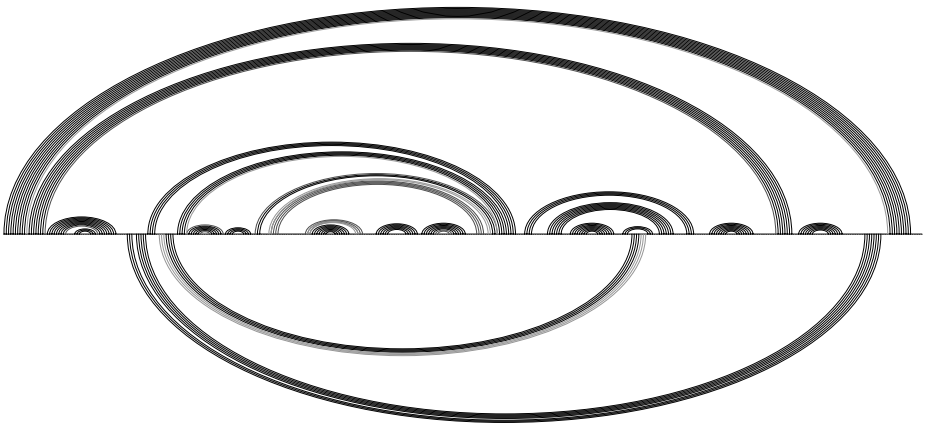

Fig. 3 Bisecondary structure of RNase P RNA as predicted by HXMATCH compared to the reference structure. Black: correct predicted base pairs; green: base pairs not predicted; red: predicted pairs not present in reference structure (See color plate p. xxxix).

Comparative sequence analysis revealed conserved pseudoknots in many important classes of RNA, such as rRNAs, RNase P RNAs, and tmRNA. Consequently, there is significant interest in computational methods for their prediction. While thermodynamic structure prediction with a loop-based energy model is NP-complete, in general, restricted classes of pseudoknots can be dealt with by polynomial algorithms (Fig. 2). Fortunately, most known pseudoknots are relatively simple. With few exceptions, they can be drawn as a superposition of just two knot-free secondary structures, that is, as a bisecondary structure. Recently a number of polynomial time algorithms have been proposed that predict pseudoknotted structures. Depending on how restrictive the set of allowed pseudoknots is chosen, their time complexity ranges between $\mathcal{O}(n^6)$ and $\mathcal{O}(n^4)$.

The practical applicability of these algorithms, however, is limited not only by their time and memory requirements, but most severely by our poor knowledge of pseudoknot energetics. Useful energy parameters have so far only been collected for simple H-type pseudoknots. In the absence of good scoring functions, expanding the set of possible structures will often lead to predictions that are worse than knot-free predictions.

The problem can be somewhat mitigated in methods that work with multiple sequence and exploit the covariance information in an alignment, such as `ILM` and `HXMATCH`. An example is shown in Fig. 3.

3 RNA Structure Prediction

3.1 Structure Prediction by Energy Minimization

Given an energy model the simplest approach to structure prediction is to determine the optimal structure with respect to its free energy, that is, the minimum free energy (MFE) structure. For pseudoknot free RNA secondary structures, the optimal structure can

be computed efficiently using dynamic programming.

If the energy of a structure could be modeled as a sum of contributions from base pairs, an optimal structure could be computed from a simple variant of the counting recursion 1,

$$E_{ij} = \min\{E_{i+1,j} + \min_{k,\,(i,k)\text{ pairs}} \times \{E_{i+1,k-1} + E_{k+1,j} + \beta_{ik}\}\}, \quad (2)$$

where β_{ik} is the energy contribution of a pair between positions i and k. The algorithm requires $\mathcal{O}(n^2)$ memory for storing the E matrix while cpu timescales as $\mathcal{O}(n^3)$. While this was the first approach to structure prediction used by Nussinov, a loop-based energy model as described above is necessary to obtain reasonable accuracies. Using a loop-based energy model requires somewhat more involved recursions, but still exhibits the same time and space complexity.

As is typical for dynamic programming, the algorithm first computes the best possible energy of a secondary structure, then "backtracks" to find a structure with this optimal energy. In the simplest case, this procedure returns a single solution, the MFE structure. This is unsatisfactory not only because inaccuracies in the energy parameters will lead to errors in the predicted structure but also because significantly different structures may be needed to represent the molecule in thermodynamic equilibrium.

3.2 Suboptimal Folding and Pair Probabilities

The most common strategy for generating additional suboptimal structures is the algorithm Michael Zuker, which considers for each possible base pair the best structure containing that pair. The number of structures in the output is further reduced by considering only structures within some energy interval of the mfe and filtering out structures that are too similar to others. The method usually returns a short list of possible foldings that form a representative sample. Occasionally, however, important alternatives will be missed.

A more rigorous approach is the computation of the partition function and base pairing probabilities using McCaskill's algorithm. For every possible base pair (i, j), the algorithm yields the probability p_{ij} that the base pair will be formed, that is, the sum of the probabilities of all structures containing that pair. The partition function can also be used to calculate heat capacities and thus characterize melting transitions.

Base pair probabilities can be nicely represented in dot plots, where we plot a square with area p_{ij} for each pair (i, j). Similarly, Zuker's suboptimal folding algorithm can produce energy dot plots, where instead of the probability, the best possible energy in structures containing (i, j) is plotted.

The *complete* suboptimal folding algorithm of Wuchty et al. can generate *all* suboptimal structures in a predefined energy range above the mfe. For small molecules, it can be illuminating to look at the exhaustive list of structural possibilities. For larger molecules, the information quickly becomes overwhelming. With further postprocessing, however, these data can be used for detailed analysis of RNA energy landscapes and the dynamics of the folding process.

Finally, the partition function algorithm can be enhanced with a stochastic backtracking procedure, which samples suboptimal structures from the Boltzmann distribution. This can be used to compute the average value of any structural feature

simply by sampling a sufficiently large number of suboptimal structures.

3.2.1 Available Programs and Web Services

Zuker's suboptimal folding algorithm is implemented in his popular `mfold` program for UNIX, as well as in David Mathew's `RNAstructure` for Windows. Partition function folding, complete suboptimal folding, and stochastic backtracking are all available as part of the `Vienna RNA Package`.

Users who need to do structure predictions only occasionally will find several of the above algorithms as a Web service on Michael Zuker's mfold server at http://bioinfo.math.rpi.edu/~zuker/rna/, as well as the Vienna RNA server http://rna.tbi.univie.ac.at.

3.2.2 Prediction Accuracy

Prediction accuracies can be evaluated by comparison to structures inferred using phylogenetic methods (see the following). For short RNAs such as tRNAs, one can expect accuracies (in terms of known base pairs that are correctly predicted) of around 70%. As pointed out recently by Doshi et al., long-range pairs are generally predicted poorly, resulting in a mean accuracy of only 41% for the large 16S and 23S rRNAs. While predicted mfe structure may have less then 20% correct pairs in unfortunate cases, good structures are still found in the vicinity of the mfe structure by suboptimal folding. Moreover, even when overall prediction accuracy is low, pairs that are predicted with high probability are usually correct.

Often a small number of constraints can dramatically improve prediction accuracy. Most folding programs allow constraints, such as specifying positions as (un)paired, to be specified. Such constraints can be obtained without too much effort from chemical probing experiments.

3.3 Well-defined Regions and Reliability

Pair probability and energy dot plots can also give a good visual impression of the quality of prediction and well-defined regions. A dot plot cluttered by many alternatives may not only indicate structural flexibility but may also make the prediction of a single mfe structure less reliable. Well-defined structures are likely to be correctly predicted, since they will be robust with respect to small variations of the energy parameters.

Several quantitative measures of well definedness are being used. In the simplest case, the well definedness of the prediction can be quantified by the difference between the MFE and the free energy of the best suboptimal structure. A more robust measure is the difference between the mfe and the ensemble free energy $G = -RT\ln(Q)$ where Q is the partition function. The latter is equivalent to the probability of the mfe structure in the ensemble given by Boltzmann's law $p(mfe) = \exp(-E_{\min}/RT)/Q = \exp(-(E_{\min} - G)/RT)$.

Even more useful are position-wise measures that help identify credible parts of the prediction. One can, for example, compute from the pair probabilities p_{ik} the positional entropy

$$S_k = -\sum_i p_{ik} \ln p_{ik} \qquad (3)$$

where p_{ii} is defined as the probability that i does not pair $p_{ii} = 1 - \sum_{j \neq i} p_{ij}$. A useful application of this measures is the annotation of structure drawings (Fig. 4).

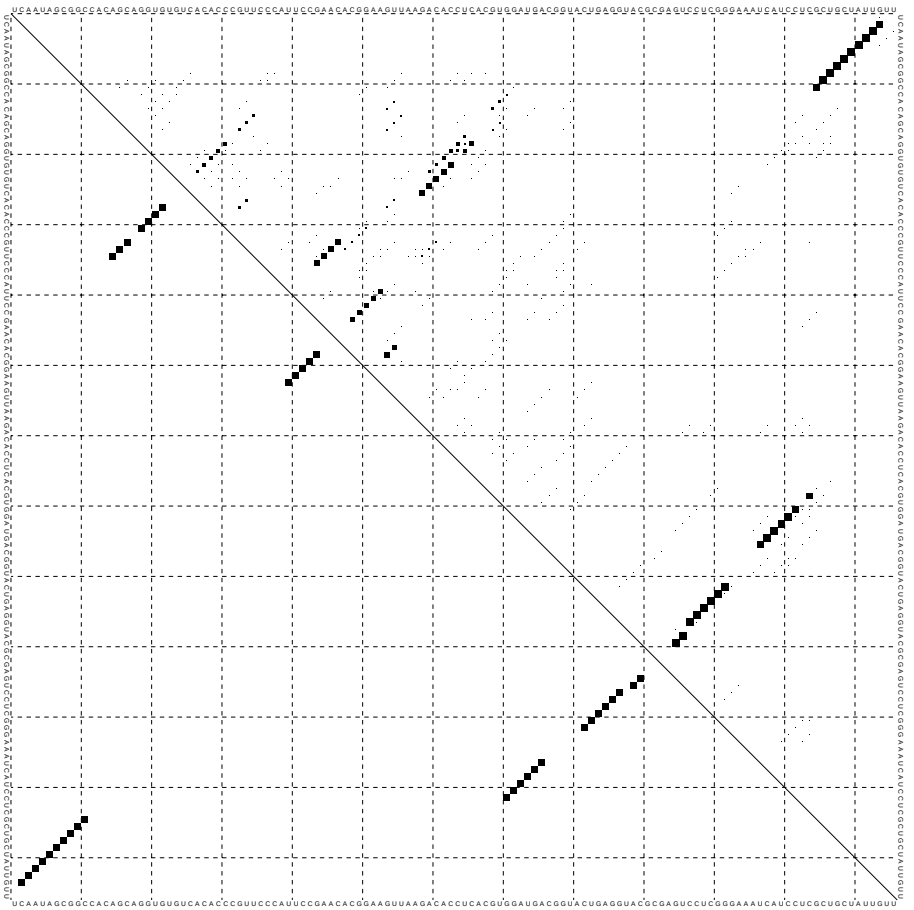

(a)

Fig. 4 Predicted structure for an 5S rRNA. (a) base pair probabilities as computed by `RNAfold -p -noLP`; (b) predicted mfe structure annotated with positional entropy; (c) Mountain plot. Note that all regions with low entropy (red) are correctly predicted, while high entropy regions (blue) deviate from the reference structure (See color plate p. xli).

3.4 Structure Prediction using Sequence Covariation

If several sequences are known to fold into (almost) the same structure, their common structure can be inferred from sequence covariation, typically measured as mutual information between two columns of a multiple sequence alignment. The approach requires a large number of related sequences and has therefore been mostly limited to ribosomal RNAs and tRNAs. However, where applicable, these phylogenetic methods produce secondary structure models of high quality, and can even elucidate some tertiary interactions.

Recently, a number of methods have appeared that combine thermodynamic prediction with covariation analysis, in order to achieve accurate predictions with only a few related sequences. Usually, these

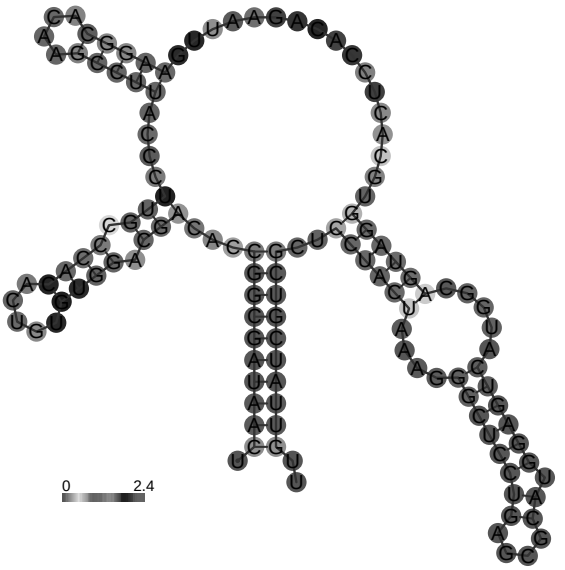

(b)

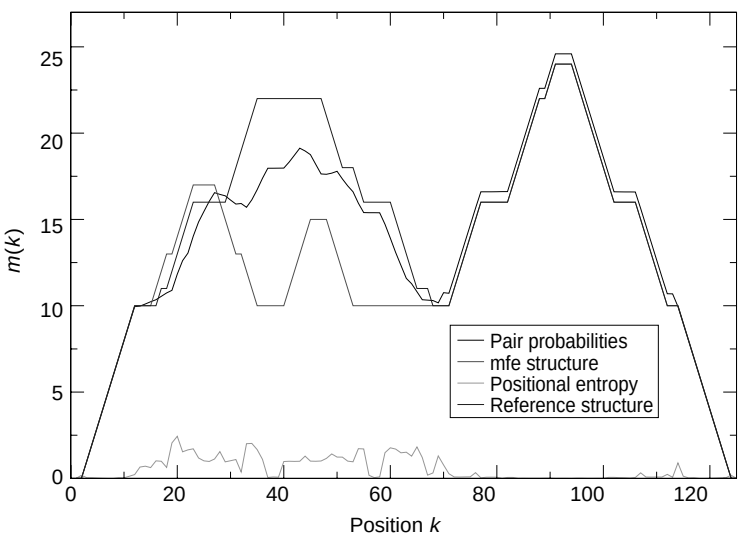

(c)

Fig. 4 (*Continued*)

methods start from a conventional multiple sequence alignment. One approach, as exemplified by the `alidot` and `ConStruct` tools, is to predict structures of individual sequences, combine them using the sequence alignment, and then find significantly conserved structure motifs. This is particularly useful when searching for locally conserved structure motifs in larger sequences.

Another approach that is suitable for predicting globally conserved structures

works by modifying the folding algorithm itself so that an optimal structure is predicted for a sequence profile, or a multiple sequence alignment, instead of a single sequence. The two best-known implementations of this approach are `pfold` and `RNAalifold`. `pfold` is based on a stochastic context-free grammar, and thus uses parameters derived from a training set; it also makes explicit use of a predicted phylogenetic tree. `RNAalifold`, on the other hand, uses the standard energy model for RNA secondary structures, augmented with a covariation term that rewards consistent and compensatory mutations. Thus, for identical sequences, it gives the same result as the single sequence prediction from `RNAfold`. With a few (or even just two) related sequences, these programs achieve prediction accuracies much higher than prediction methods for single sequences.

Methods starting from a multiple sequence alignment are of course limited by the accuracy of the input alignment. Typically, this becomes a problem when pairwise sequence identities drop below 60%. An alternative is to start by aligning predicted mfe structures using `RNAforester` or `MARNA`. The most rigorous way is to use a variant of the Sankoff algorithm, which computes the alignment and consensus structure simultaneously. Notable implementations are `foldalign`, `dynalign`, `pmcomp/pmmulti`, and `dart`. The Sankoff algorithm is, however, computationally very expensive ($\mathcal{O}(n^6)$ in the unrestricted case). The above algorithms therefore use various restrictions to improve speed.

3.4.1 Search for ncRNA

With the realization that a significant fraction of the genomic DNA of higher organisms might code for functional RNAs rather than protein-coding genes, the development of techniques for the detection and classification of RNAs in large sequences has become an active field of research. Generally, we can distinguish between experimental and computational methods to search for ncRNA. Experimental methods, as ligation of 5′ and 3′ adapters to size fractionated isolated RNA, followed by reverse transcription, cDNA cloning and sequencing, or simply using Northern blot analysis for sequence-specific detection of ncRNAs are at present restricted to the detection of abundant ncRNA. In this chapter, we will therefore concentrate on computational methods.

Some ncRNAs can be found by searching for likely transcripts that do not contain an open reading frame. A survey of the *E. coli* genome for DNA regions that contain a $\sigma 70$ promotor within a short distance of a *Rho*-independent terminator, for instance, resulted in 144 novel possible ncRNAs. This approach is limited, however, to functional RNAs that are transcribed in the "usual" manner by means of pol-II. For many ncRNAs, however, the mode of transcription is unknown.

Neural networks or support vector machines have also been successfully trained to recognize all known ncRNA in the *E. coli* genome. The survey conducted resulted in 370 new candidates, which have not been verified experimentally. The number seems to be quite reasonable, in particular when we keep in mind that false positives are a common problem with most computational searches for ncRNA so far.

Some structure-based searches use the known secondary structure of the major classes of functional RNAs. Programs such as `RNAmot`, `tRNAscan`, `HyPa`, `RNAMotif`, `bruce`, and many others exploit this

avenue. An interesting variant that makes use of evolutionary computation is described by Fogel. Nevertheless, all these approaches are restricted to searching for new members of the few well-established families. Especially for miRNAs, being quite short with a characteristic stem-loop structure ending up in a hairpin, several genome-wide surveys have been undertaken. The initially very large number of candidates is filtered in a postprocessing process according to other structural and sequential elements.

Comparative approaches such as the program QRNA can detect novel structural RNA genes in a pair of aligned homologous sequences by deciding whether the substitution pattern fits better with (1) synonymous substitutions, which are expected in protein-coding regions; (2) the compensatory mutations consistent with some base-paired secondary structure; or (3) uncorrelated mutations. Genome-wide scans for new ncRNA have already been accomplished with QRNA.

Another approach tries to determine functional RNAs by means of structure prediction. The basic assumption is that functional and hence conserved structures will be thermodynamically more stable. While such procedures are capable of detecting some particularly stable features, a recent study concludes that "although a distinct, stable secondary structure is undoubtedly important in most noncoding RNAs, the stability of most noncoding RNA secondary structures is not sufficiently different from the predicted stability of a random sequence to be useful as a general genefinding approach." Nevertheless, in some special cases such as hyperthermophilic organisms, **GC**-content (and hence thermodynamic stability) proved sufficient.

When thermodynamic stability is combined with comparative methods, structural ncRNAs can indeed be recognized with high specificity and sensitivity. Two such approaches were developed recently building on consensus structure predictions using RNAalifold. The alifoldz method compares the folding energy of the native alignment with those of randomized alignments, using a somewhat costly shuffling procedure. The even faster RNAz method combines a measure for RNA secondary structure conservation on the basis of comparing the consensus folding energies to folding energies of individual sequences, with a measure for thermodynamic stability, which is normalized with respect to both sequence length and base composition. This method is fast and reliable enough to scan even large vertebrate genomes.

4 Current Research in RNA Secondary Structures

Despite the recent progress in detection algorithms for ncRNAs in comparative genomics approaches, a reliable *annotation* of functional RNAs in genomic data is still a largely unsolved problem for a variety of reasons. Most importantly, annotation is almost always limited to comparative methods; at present there are no generally applicable methods that could be used to determine the function of an RNA from its sequence and/or (predicted) structure. So far, structure and function of only a very small fraction of ncRNAs and other functional RNA motifs have been characterized. In fact, the functions of a number of very well conserved, evolutionary ancient ncRNAs such as Y

RNAs or vault RNAs have remained in the dark.

Even the recognition of members of the large classes of known ncRNAs is less straightforward than one might expect: the sequences of snoRNAs, snRNAs, and many others evolve relatively rapidly, their secondary structures are thus the main features that can be used to identify them. Note that it is much easier to identify additional members of a family of homologous snoRNA and microRNAs than to recognize a new class of homologous ncRNAs, for which no member has been described experimentally, as a novel family of microRNAs or snoRNAs. For the first class of problems, simple sequence comparison or general pattern matching techniques that combine sequence and secondary structure information, such as `ERPIN`, `HyPa`, or `PATSearch`, can be used in many cases. While search tools exist for such pattern description languages, the inference of characteristic patterns from a set of aligned (or, even more demanding, nonaligned) sequences is still an open problem.

In the case of microRNAs, a characteristic pattern of mutations in the precursor hairpin is a reliable signal provided a sufficient number of homologous sequences can be found. The motifs of box H/ACA and box C/D snoRNAs, as well as the short sequences for some snRNAs, are less informative, however. In this case, knowledge about their potential targets, rRNAs, snRNAs, and some mRNAs, can be used to improve the quality of the prediction. In the case of snRNAs, conserved promoter and enhancer sequences could potentially be used. However, efficient tools for RNA classes besides tRNAs that could be used in genomic surveys are at present not available.

The identification of the target genes of microRNAs is a hot topic at present. While microRNAs bind their target mRNAs almost complementarily in plants, the mechanism appears to be much more complicated in animals. As a consequence, none of the approaches published to date, such as `miRanda`, `targetscan`, or `RNAhybrid` provide a complete solution to the problem. From an algorithmic point of view, the microRNA target problem has stimulated research into methods for computing RNA–RNA interactions, and to consider RNA–protein interactions in a more systematic way.

A related open problem concerns the classification of the ncRNAs that do *not* belong to one of the known classes. While it is unlikely that the functions of these molecules can be elucidated without further experimental data, it is of interest to determine whether there are additional larger classes of ncRNAs that have escaped our attention. Of course, the same question arises for functional RNA motifs in mRNAs.

Structure-function studies of small RNAs as well as the analysis of artificial RNA sequences obtained from SELEX experiments depend upon the ability to compute structure-based alignments. As mentioned earlier, Sankoff's algorithm solves the RNA folding and pairwise alignment problem simultaneously, but is often computationally too expensive. The alternative, obtaining pairwise alignments through tree editing or tree alignment, is problematic since it depends on predicted, possibly incorrect, structures. In all cases, heuristics are used to obtain multiple alignments from pairwise alignments. Open questions in this area concern both algorithmic improvements and more realistic cost functions; for instance, tree alignments are currently available only with linear gap costs.

RNA structure design can be regarded as the inverse of the RNA folding problem: for a given secondary structure graph, find one or more sequences that have this structure as their MFE structure. An enumeration of sequences that fold into a given structure is infeasible in general because of their large number. A heuristic that turns this problem into a combinatorial optimization problem has already been described. There is, however, no exact algorithm that could determine whether there is at least one sequence that folds into a prescribed secondary structure graph.

RNA switches exhibit two competing conformations, whose equilibrium can be shifted easily by molecular events such as the binding of another molecule. Such elements have recently been identified as important regulators of gene expression, in particular, in bacteria. A theoretical study shows that RNAs that have very different secondary structures with near-groundstate energy, that is, potential riboswitches, are relatively frequent and easily accessible in evolution. The secondary structure model furthermore allows estimates of energy barriers and even entire folding trajectories. An impressive experimental study demonstrates that a single sequence can have two distinct functional ribozyme structures. Heuristic algorithms can be used to design sequences with two near groundstates. The case of more than two alternative structures is only partially understood.

Small subunit ribosomal RNAs are probably the most frequently used data sets in molecular phylogenetics. Most studies assume that individual sequence positions evolve at least approximately independently of each other. This is, however, at odds with the fact that rRNAs form stable secondary structures that are very well preserved over very long timescales. Selection for stabilizing RNA structure is seen in many types of structural RNA sequences. Nucleotides in stem regions evolve in strong correlation with their pairing counterpart. More elaborate scoring schemes and even associated maximum likelihood techniques acting directly on RNA secondary structures are algorithmically relatively unproblematic but will require detailed knowledge about the dynamics of RNA structure evolution. While the structures of many RNA families change very little, for example within vertebrates, significant changes of the structure, however, are evident at timescales of the divergence of the major metazoan phyla. As a consequence, not only fixed secondary structures but the evolution of the structures themselves need to be taken into account. Secondary structures have rarely been used in molecular phylogenetics so far. An exception is the investigation into the history of RNase P and RNase MRP RNAs by Davin Penny and coworkers, which demonstrates that "RNA secondary structure is useful for evaluating evolutionary relatedness, even with sequences that cannot be aligned with confidence". More recently, cladistic analyses based on RNA secondary structure have demonstrated this point convincingly, in particular at the level of deep phylogenies.

See also RNA Methodologies; RNA Three-Dimensional Structures, Computer Modeling of.

Bibliography

Books and Reviews

Eddy, S.R. (2002) Computational genomics of noncoding RNA genes, *Cell* **109**, 137–140.

Flamm, C., Hofacker, I.L., Stadler, P.F. (1999) RNA in silico: the computational biology of RNA secondary structures, *Adv. Complex Syst.* **2**, 65–90.

Higgs, P.G. (2000) RNA secondary structure: physical and computational aspects, *Q. Rev. Biophys.* **33**, 199–253.

Zuker, M. (2000) Calculating nucleic acid secondary structure, *Curr. Opin. Struct. Biol.* **10**, 303–310.

Primary Literature

Abfalter, I., Flamm, C., Stadler, P.F. (2003) Design of multi-stable nucleic acid sequences, in: Mewes, H.-W., Heun, V., Frishman, D., Kramer, S. (Eds.) *Proceedings of the German Conference on Bioinformatics. GCB 2003*, Vol. 1, Belleville Verlag Michael Farin, München, Germany, pp. 1–7.

Accardo, M.C., Giordano, E., Riccardo, S., Digilio, F.A., Iazzetti, G., Calogero, R.A., Furia, M. (2004) A computational search for box C/D snoRNA genes in the D. melanogaster genome, *Bioinformatics* **20**, 3293–3301.

Adai, A., Johnson, C., Mlotshwa, S., Archer-Evans, S., Manocha, V., Vance, V., Sundaresan, V. (2005) Computational prediction of miRNAs in Arabidopsis thaliana, *Genome Res.* **15**, 78–91.

Akutsu, T. (2001) Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots, *Discr. Appl. Math.* **104**, 45–62.

Andronescu, M., Fejes, A.P., Hutter, F., Hoos, H.H., Condon, A. (2004) A new algorithm for RNA secondary structure design, *J. Mol. Biol.* **336**, 607–624.

Argaman, L., Vogel, J., Bejerano, G., Wagner, E.G., Margalit, H., Altuvia, S. (2001) Novel small RNA-encoding genes in the intergenic regions of Escherichia coli, *Curr. Biol.* **11**, 941–950.

Avner, P., Heard, E. (2001) X-chromosome inactivation: counting, choice, and initiation, *Nat. Rev. Genet.* **2**, 59–67.

Brown, J. (1999) The ribonuclease P database, *Nucleic Acids Res.* **27**, 314–314.

Brown, J.W., Nolan, J.M., Haas, E.S., Rubio, M.A.T., Major, F., Pace, N.R. (1996) Comparative analysis of ribonuclease P RNA using gene sequences from natural microbial populations reveals tertiary structural elements, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 3001–3006.

Caetano-Anollés, G. (2002a) Evolved RNA secondary structure and the rooting of the universal tree, *J. Mol. Evol.* **54**, 333–345.

Caetano-Anollés, G. (2002b) Tracing the evolution of RNA structure in ribosomes, *Nucleic Acids Res.* **30**, 2575–2587.

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Müller, K.M., Pande, N., Shang, Z., Yu, N., Gutell, R.R. (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs, *BMC Bioinformatics* **3**, 2, http://www.rna.icmb.utexas.edu.

Carter, R., Dubchak, I., Holbrook, S. (2001) A computational approach to identify genes for functional RNAs in genomic sequences, *Nucleic Acids Res.* **29**, 3928–3938.

Chen, S., Lesnik, E., Hall, T., Sampath, R., Griffey, R., Eker, D., Blyn, L. (2002) A bioinformatics based approach to discover small RNA genes in the Escherichia coli genome, *Biosystems* **65**, 157–177.

Collins, L.J., Moulton, V., Penny, D. (2000) Use of RNA secondary structure for studying the evolution of RNase P and RNase MRP, *J. Mol. Evol.* **51**, 194–204.

d'Aubenton Carafa, Y., Brody, E., Thermes, C. (1990) Prediction of rho-independent Escherichia coli transcription terminators. a statistical analysis of their RNA stem-loop structures, *J. Mol. Biol.* **216**, 835–858.

Dayton, E., Konings, D., Powell, D., Shapiro, B., Butini, L., Maizel, J., Dayton, A. (1992) Extensive sequence-specific information throughout the CAR/RRE, the target sequence of the human immunodeficiency virus type 1 Rev protein, *J. Virol.* **66**, 1139–1151.

Dimitrov, R.A., Zuker, M. (2004) Prediction of hybridization and melting for double-stranded nucleic acids, *Biophys. J.* **87**, 215–226.

Ding, Y., Lawrence, C.E. (2003) A statistical sampling algorithm for RNA secondary structure prediction, *Nucleic Acids Res.* **31**, 7180–7301.

Dirks, R.M., Pierce, N.A. (2003) A partition function algorithm for nucleic acid secondary structure including pseudoknots, *J. Comput. Chem.* **24**, 1664–1677.

Doshi, K., Cannone, J., Cobaugh, C., Gutell, R. (2004) Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary

structure prediction, *BMC Bioinformatics* **5**(1), 105.

Eddy, S. (2001) Non-coding RNA genes and the modern RNA world, *Nat. Genet.* **2**, 919–929.

Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., Marks, D.S. (2003) MicroRNA targets in Drosophila, *Genome Biol.* **5**(1), Article R1.

Erdmann, V., Szymanski, M., Hochberg, A., de Groot, N., Barciszewski, J. (1999) Collection of mRNA-like non-coding RNAs, *Nucleic Acids Res.* **27**, 192–195.

Erdmann, V., Barciszewska, M., Hochberg, A., de Groot, N., Barciszewski, J. (2001) Regulatory RNAs, *Cell. Mol. Life Sci.* **58**, 960–977.

Farris, A.D., Koelsch, G., Pruijn, G.J., van Venrooij, W.J., Harley, J.B. (1999) Conserved features of Y RNAs revealed by automated phylogenetic secondary structure analysis, *Nucleic Acids Res.* **27**, 1070–1078.

Flamm, C., Fontana, W., Hofacker, I., Schuster, P. (2000a) RNA folding kinetics at elementary step resolution, *RNA* **6**, 325–338.

Flamm, C., Hofacker, I.L., Maurer-Stroh, S., Stadler, P.F., Zehl, M. (2000b) Design of multi-stable RNA molecules, *RNA* **7**, 254–265.

Fogel, G., Porto, V., Weekes, D., Fogel, D., Griffey, R., McNeil, J., Lesnik, E., Ecker, D., Sampath, R. (2002) Discovery of RNA structural elements using evolutionary computation, *Nucleic Acids Res.* **30**, 5310–5317.

Fontana, W., Konings, D., Stadler, P., Schuster, P. (1993) Statistics of RNA secondary structures, *Biopolymers* **33**, 1389–1404.

Franke, A., Baker, B. (2000) Dosage compensation rox! *Curr. Opin. Cell Biol.* **12**, 351–354.

Gardner, P.P., Giegerich, R. (2004) A comprehensive comparison of comparative RNA structure prediction approaches, *BMC Bioinformatic* **5**, 140.

Gautheret, D., Lambert, A. (2001) Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles, *J. Mol. Biol.* **313**, 1003–1011.

Gautheret, D., Major, F., Cedergren, R. (1990) Pattern searching/alignment with RNA primary and secondary structures: an effective descriptor for tRNA, *Comput. Appl. Biosci.* **6**, 325–331.

Gong, H., Liu, C.-M., Liu, D.-P., Liang, C.-C. (2005) The role of small RNAs in human diseases: potential troublemaker and therapeutic tools, *Med. Res. Rev.* **25**, 361–381, Epub 6 Jan 2005.

Gorodkin, J., Heyer, L., Stormo, G. (1997) Finding Common Sequences and Structure Motifs in a Set of RNA Molecules, in: Gaasterland, T., Karp, P., Karplus, K., Ouzounis, C., Sander, C., Valencia, A., eds. *Proceedings of the ISMB-97*, AAAI Press, Menlo Park, CA, pp. 120–123.

Gorodkin, J., Stricklin, S.L., Stormo, G.D. (2001b) Discovering common stem-loop motifs in unaligned RNA sequences, *Nucleic Acids Res.* **29**(10), 2135–2144.

Gorodkin, J., Knudsen, B., Zwieb, C., Samuelsson, T. (2001a) SRPDB (signal recognition particle database), *Nucleic Acids Res.* **29**, 169–170.

Grad, Y., Aach, J., Hayes, G., Reinahrt, B., Church, G., Ruvkun, G., Kim, J. (2003) Computational and experimental identification of C. elegans microRNAs, *Mol. Cell* **11**, 1253–1263.

Gräf, S., Strothmann, D., Kurtz, S., Steger, G. (2001) HyPaLib: a database of RNAs and RNA structural elements defined by hybrid patterns, *Nucleic Acids Res.* **29**, 196–198.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., Eddy, S.R. (2003) Rfam: an RNA family database, *Nucleic Acids Res.* **31**, 439–441.

Gultyaev, A.P., van Batenburg, F., Pleij, C. (1999) An approximation of loop free energy values of RNA H-pseudoknots, *RNA* **5**, 609–617.

Gutell, R.R., Power, A., Hertz, G.Z., Putz, E.J., Stormo, G.D. (1992) Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods, *Nucleic Acids Res.* **20**, 5785–5795.

Hackermüller, J., Meisner, N.-C., Auer, M., Jaritz, M., Stadler, P.F. (2005) The effect of RNA secondary structures on RNA-ligand binding and the modifier RNA mechanism: a quantitative model, *Gene* **345**, 3–12. doi:10.1016/j.gene.2004.11.043.

Harris, J.K., Haas, E.S., Williams, D., Frank, D.N. (2001) New insight into RNase P RNA structure from comparative analysis of the archaeal RNA, *RNA* **7**, 220–232. Hernandez, N. (2001) Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription, *J. Biol. Chem.* **276**, 26733–26736.

Hobza, P., Šponer, J. (2002) Towards true DNA base-stacking energies: MP2, CCSD(T), and

complete basis set calculations, *J. Am. Chem. Soc.* **124**, 11802–11808.

Höchsmann, M., Töller, T., Giegerich, R., Kurtz, S. (2003) Local similarity in RNA secondary structures, Proceedings of the Computational Systems Bioinformatics Conference, Stanford, CA, Aug. 2003 (CSB 2003), pp. 159–168.

Hofacker, I.L. (2003) The Vienna RNA secondary structure server, *Nucleic Acids Res.* **31**, 3429–3431.

Hofacker, I., Stadler, P. (1999) Automatic detection of conserved base pairing patterns in RNA virus genomes, *Comput. Chem.* **23**, 401–414.

Hofacker, I.L., Schuster, P., Stadler, P.F. (1998b) Combinatorics of RNA secondary structures, *Discr. Appl. Math.* **88**, 207–237.

Hofacker, I., Fekete, M., Stadler, P. (2002) Secondary structure prediction for aligned RNA sequences, *J. Mol. Biol.* **319**, 1059–1066.

Hofacker, I.L., Bernhart, S.H.F., Stadler, P.F. (2004) Alignment of RNA base pairing probability matrices, *Bioinformatics* **20**, 2222–2227.

Hofacker, I., Fontana, W., Stadler, P., Bonhoeffer, S., Tacker, M., Schuster, P. (1994) Fast folding and comparison of RNA secondary structures, *Monatsh. Chem.* **125**, 167–188.

Hofacker, I., Fekete, M., Flamm, C., Huynen, M., Rauscher, S., Stolorz, P., Stadler, P. (1998a) Automatic detection of conserved RNA structure elements in complete RNA virus genomes, *Nucleic Acids Res.* **26**, 3825–3836.

Holmes, I. (2004) A probabilistic model for the evolution of RNA structure, *BMC Bioinformatics* **5**, 166.

Huez, I., Créancier, L., Audigier, S., Gensac, M., Prats, A., Prats, H. (1998) Two independent internal ribosome entry sites are involved in translation initiation of vascular endothelial growth factor mRNA, *Mol. Cell. Biol.* **18**, 6178–6190.

Hüttenhofer, A., Kiefmann, M., Neier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J., Brosius, J. (2001) Rnomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse, *EMBO J.* **20**, 2943–2953.

Huynen, M., Perelson, A., Viera, W., Stadler, P. (1996) Base pairing probabilities in a complete HIV-1 RNA, *J. Comput. Biol.* **3**, 253–274.

Jacobs, G., Rackham, O., Stockwell, P., Tate, W., Brown, C.M. (2002) Transterm: a database of mRNAs and translational control elements, *Nucleic Acids Res.* **30**, 310–311.

Klein, R., Misulovin, Z., Eddy, S. (2002) Noncoding RNA genes identified in AT-rich hyperthermophiles, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7542–7547.

Knudsen, B., Hein, J. (1999) Using stochastic context free grammars and molecular evolution to predict RNA secondary structure, *Bioinformatics* **15**, 446–454.

Knudsen, B., Hein, J. (2003) Pfold: RNA secondary structure prediction using stochastic context-free grammars, *Nucleic Acids Res.* **31**, 3423–3428.

Kool, E.T. (1997) Preorganization of DNA: design principles for improving nucleic acid recognition by synthetic oligonucleotides, *Chem. Rev.* **97**, 1473–1487.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs, *Science* **294**, 853–857.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., Tuschl, T. (2002) Identification of tissue specific microRNAs from mouse, *Curr. Biol.* **12**, 735–739.

Laslett, D., Canback, B., Andersson, S. (2002) BRUCE: a program for the detection of transfer-messenger RNA genes in nucleotide sequences, *Nucleic Acids Res.* **30**, 3449–3453.

Lau, N., Lim, L., Weinstein, E., Bartel, D. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans, *Science* **294**, 858–862.

Le, S.-Y., Chen, J.-H., Currey, K., Maizel, J. (1988) A program for predicting significant RNA secondary structures, *Comput. Appl. Biosci.* **4**, 153–159.

Lee, R.C., Ambros, V. (2001) An extensive class of small RNAs in Caenorhabditis elegans, *Science* **294**, 862–864.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., Burge, C.B. (2003) Prediction of mammalian microRNA targets, *Cell* **115**, 787–798.

Leydold, J., Stadler, P.F. (1998) Minimal cycle basis, outerplanar graphs, *Elec. J. Comb.* **5**, R16, See http://www.combinatorics.org.

Lim, L., Lau, N., Weinstein, E., Abdelhakim, A., Yekta, S., Rhoades, M., Burge, C., Bartel, P. (2003) The microRNAs of Caenorhabditis elegans, *Genes Dev.* **17**, 991–1008.

Lowe, T., Eddy, S. (1997) tRNAscan-SE: a program for improved detection of transfer

RNA genes in genomic sequence, *Nucleic Acids Res.* **25**, 955–964.

Lück, R., Steger, G., Riesner, D. (1996) Thermodynamic prediction of conserved secondary structure: application to the RRE element of HIV, the tRNA-like element of CMV, and the mRNA of prion protein, *J. Mol. Biol.* **258**, 813–826.

Lück, R., Gräf, S., Steger, G. (1999) ConStruct: a tool for thermodynamic controlled prediction of conserved secondary structure, *Nucleic Acids Res.* **27**, 4208–4217.

Lyngsø, R.B., Pedersen, C.N.S. (2000) RNA pseudoknot prediction in energy based models, *J. Comput. Biol.* **7**(3/4), 409–428.

Macdonald, P. (1990) Bicoid mRNA localization signal: phylogenetic conservation of function and RNA secondary structure, *Development* **110**, 161–171.

MacIntosh, G., Wilkerson, C., Green, P. (2001) Identification and analysis of arabidopsis expressed sequence tags characteristic of non-coding RNAs, *Plant Physiol.* **127**, 765–776.

Macke, T., Ecker, D., Gutell, R., Gautheret, D., Case, D., Sampath, R. (2001) RNAMotif, an RNA secondary structure definition and search algorithm, *Nucleic Acids Res.* **29**, 4724–4735.

Maidak, B., Cole, J., Lilburn, T., Parker, C. Jr., Saxman, P., Farris, R., Garrity, G., Olsen, G., Schmidt, T., Tiedje, J. (2001) The RDP-II (ribosomal database project), *Nucleic Acids Res.* **29**, 173–174.

Mathews, D.H. (2004) Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization, *RNA* **10**, 1178–1190.

Mathews, D., Turner, D.H. (2002) Dynalign: an algorithm for finding the secondary structure common to two RNA sequences, *J. Mol. Biol.* **317**, 191–203.

Mathews, D., Sabina, J., Zuker, M., Turner, D. (1999) Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure, *J. Mol. Biol.* **288**, 911–940.

Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., Turner, D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7287–7292.

McCaskill, J. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure, *Biopolymers* **29**, 1105–1119.

McCutcheon, J., Eddy, S. (2003) Computational identification of non-coding RNAs in Saccharomyces cerevisiae by comparative genomics, *Nucleic Acids Res.* **31**, 4119–4128.

Merino, E., Yanofsky, C. (2002) *Regulation by Termination-Antitermination: A Genomic Approach*, ASM Press, pp. 323–336.

Nussinov, R., Piecznik, G., Griggs, J., Kleitman, D. (1978) Algorithms for loop matching, *SIAM J. Appl. Math.* **35**(1), 68–82.

Ohno, M., Mattaj, I. (1999) Meiosis: MeiRNA hits the spot, *Curr. Biol.* **28**, R66–R69.

Oleynikov, Y., Singer, R. (1998) RNA localization: different zipcodes, same postman? *Trends Cell Biol.* **8**, 381–383.

Omer, A., Lowe, T., Russel, A., Ebhardt, H., Eddy, S., Dennis, P. (2000) Homologs of small nucleolar RNAs in Archaea, *Science* **288**, 517–522.

Ornstein, R.L., Rein, R., Breen, D.L., MacElroy, R.D. (1978) An optimized potential function for the calculation of nucleic acid interaction energies. I. Base stacking, *Biopolymers* **17**, 2341–2360.

Otsuka, J., Sugaya, N. (2003) Advanced formulation of base pair changes in the stem regions of ribosomal RNAs; its application to mitochondrial rRNAs for resolving the phylogeny of animals, *J. Theor. Biol.* **222**, 447–460.

Pesole, G., Liuni, S., D'Souza, M. (2000) PatSearch: a pattern matcher software that finds functional elements in nucleotide and protein sequences and assesses their statistical significance, *Bioinformatics* **16**(5), 439–450.

Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F., Sabino, L. (2001) Structural and functional features of eukaryotic mRNA untranslated regions, *Gene* **276**, 73–81.

Peterson, K., Eernisse, D.J. (2001) Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S DNA gene sequences, *Evol. Dev.* **3**, 170–205.

Reeder, J., Giegerich, R. (2004) Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics, *BMC Bioinformatics* **5**(1), 104.

Rehmsmeier, M., Steffen, P., Höchsmann, M., Giegerich, R. (2004) Fast and effective prediction of microRNA/target duplexes, *RNA* **10**, 1507–1517.

Rivas, E., Eddy, S. (1999) A dynamic programming algorithm for RNA structure prediction including pseudoknots, *J. Mol. Biol.* **285**, 2053–2068.

Rivas, E., Eddy, S. (2000) Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs, *Bioinformatics* **16**, 583–605.

Rivas, E., Eddy, S. (2001) Noncoding RNA gene detection using comparative sequence analysis, *BMC Bioinformatics* **2**(8), 19.

Rivas, E., Klein, R., Jones, T., Eddy, S. (2001) Computational identification of noncoding RNAs in E. coli by comparative genomics, *Curr. Biol.* **11**, 1369–1373.

Ruan, J., Stormo, G.D., Zhang, W. (2004) An iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots, *Bioinformatics* **20**, 58–66.

Rueckert, R. (1996) Picornaviridae: The Viruses and their Replication, in: Fields, N., Knipe, D., Howley, P. (Eds.) *Virology*, Vol. 1, 3rd edition, Lippincott-Raven, Philadelphia, NY, pp. 609–654.

Samarsky, D., Fournier, M. (1999) A comprehensive database for the small nucleolar RNAs from Saccharomyces cerevisiae, *Nucleic Acids Res.* **27**, 161–164.

Sankoff, D. (1985) Simultaneous solution of the RNA folding, alignment, and proto-sequence problems, *SIAM J. Appl. Math.* **45**, 810–825.

Savill, N.J., Hoyle, D.C., Higgs, P.G. (2001) RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods, *Genetics* **157**, 399–411.

Schattner, P. (2002) Searching for RNA genes using base composition statistics, *Nucleic Acids Res.* **30**, 2076–2082.

Schöninger, M., von Haeseler, A. (1999) Towards assigning helical regions in alignments of ribosomal RNA and testing the appropriateness of evolutionary models, *J. Mol. Evol.* **49**, 691–698.

Schultes, E.A., Bartel, D.P. (2000) One sequence, two ribozymes: Implications for the emergence of new ribozyme folds, *Science* **289**, 448–452.

Shapiro, B.A. (1988) An algorithm for comparing multiple RNA secondary structures, *Comput. Appl. Biosci.* **4**, 387–393.

Siebert, S., Backofen, R. (2003) MARNA a server for multiple alignment of RNAs, in: Mewes, H.-W., Heun, V., Frishman, D., Kramer, S., (Eds.) *Proceedings of the German Conference on Bioinformatics. GCB 2003*, Vol. 1, Belleville Verlag Michael Farin, München, Germany pp. 135–140.

Sousa, C., Johansson, C., Charon, C., Manyani, H., Sautter, C., Kondorosi, A., Crespi, M. (2001) Translational and structural requirements of the early nodulin gene enod40, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex, *Mol. Cell. Biol.* **21**, 354–366.

Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., Steinberg, S. (1998) Compilation of tRNA sequences and sequences of tRNA genes, *Nucleic Acids Res.* **26**, 148–153.

Szymanski, M., Barciszewska, M., Barciszewski, J., Erdmann, V. (2000) 5S ribosomal RNA database Y2K, *Nucleic Acids Res.* **28**, 166–167.

Tacker, M., Stadler, P.F., Bornberg-Bauer, E.G., Hofacker, I.L., Schuster, P. (1996) Algorithm independent properties of RNA structure prediction, *Eur. Biophy. J.* **25**, 115–130.

Tang, T., Bachellerie, J., Rozhdestvensky, T., Bortolin, M., Huber, H., Drungowski, M., Elge, T., Brosius, J., Húttenhofer, A. (2002) Identification of 86 candidates for small non-messenger RNAs from the archeon Archeoglobus fulgidus, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7536–7541.

Van de Peer, Y., De Rijk, P., Wuyts, J., Winkelmans, T., De Wachter, R. (2000) The European small subunit ribosomal RNA database, *Nucleic Acids Res.* **28**, 175–176.

van Zon, A., Mossink, M., Schoester, M., Scheffer, G., Scheper, R., Sonneveld, P., Wiemer, E. (2001) Multiple human vault RNAs. Expression and association with the vault complex, *J. Biol. Chem.* **276**, 37715–37721.

Vitreschak, A.G., Rodionov, D.A., Mironov, A.A., Gelfand, M.S. (2004) Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet.* **20**(1), 44–50.

Walter, A., Turner, D., Kim, J., Lyttle, M., Müller, P., Mathews, D., Zuker, M. (1994) Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9218–9222.

Wang, X.J., Reyes, J.L., Chua, N.H., Gaasterland, T. (2004) Prediction and identification of *Arabidopsis thaliana* micrornas and their mRNA targets, *Genome Biol.* **5**, R65, [Epub].

Washietl, S., Hofacker, I.L. (2004) Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics, *J. Mol. Biol.* **342**, 19–39.

Washietl, S., Hofacker, I.L., Stadler, P.F. (2005) Fast and reliable prediction of noncoding rnas, *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 2454–2459.

Wassarman, K., Repoila, F., Rosenow, C., Storz, G., Gottesman, S. (2001) Identification of novel small RNAs using comparative genomics and microarrays, *Genes Dev.* **15**, 1637–1651.

Waterman, M.S. (1978) Secondary structure of single-stranded nucleic acids, *Adv. Math. Suppl. Studies* **1**, 167–212.

Waterman, M.S., Smith, T.F. (1978) Combinatorics of RNA hairpins and cloverleaves, *Stud. Appl. Math.* **60**, 91–96.

Westhof, E., Jaeger, L. (1992) RNA pseudoknots, *Curr. Opin. Struct. Biol.* **2**, 327–333.

Witwer, C., Hofacker, I.L., Stadler, P.F. (2004) Prediction of consensus RNA secondary structures including pseudoknots, *IEEE/ACM Trans. Comp. Biol. Bioinf.* **1**, 65–77.

Witwer, C., Rauscher, S., Hofacker, I., Stadler, P. (2001) Conserved RNA secondary structures in picornaviridae genomes, *Nucleic Acids Res.* **29**, 5079–5089.

Wolfinger, M.T., Svrcek-Seiler, W.A., Flamm, C., Hofacker, I.L., Stadler, P.F. (2004) Exact folding dynamics of RNA secondary structures, *J. Phys. A: Math. Gen.* **37**, 4731–4741.

Wuchty, S., Fontana, W., Hofacker, I., Schuster, P. (1999) Complete suboptimal folding of RNA and the stability of secondary structures, *Biopolymers* **49**, 145–165.

Wuyts, J., De Rijk, P., Van de Peer, Y., Winkelmans, T., De Wachter, R. (2001) The European large subunit ribosomal RNA database, *Nucleic Acids Res.* **29**, 175–177.

Xia, T., SanatLucia J. Jr., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C., Turner, D.H. (1998) Parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick pairs, *Biochemistry* **37**, 14719–14735.

Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule, *Science* **244**, 48–52.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Res.* **31**, 3406–3415.

Zuker, M., Sankoff, D. (1984) RNA secondary structures and their prediction, *Bull. Math. Biol.* **46**, 591–621.

Zuker, M., Jacobson, A.B. (1998) Using reliability information to annotate RNA secondary structures, *RNA* **4**, 669–679.

Zwieb, C. (1996) The uRNA database, *Nucleic Acids Res.* **24**, 76–79.

Zwieb, C., Wower, J. (2000) tmRDB (tmRNA database), *Nucleic Acids Res.* **28**, 169–170.

Zwieb, C., Wower, I., Wower, J. (1999) Comparative sequence analysis of tmRNA, *Nucleic Acids Res.* **27**(10), 2063–2071.

RNA Translation to Protein: *see* Translation of RNA to Protein

RNA Three-Dimensional Structures, Computer Modeling of

François Major and Philippe Thibault
Université de Montréal, Montréal, Québec, Canada

1 Goal 608

2 Structural Knowledge 609
2.1 Nucleotide Conformations 609
2.2 Nucleotide Interactions 612
2.2.1 Base Stacking 612
2.2.2 Base Pairing 612
2.2.3 Base Pairing Nomenclatures 613
2.2.4 Structure Annotation 615
2.2.5 Isosteric Base Pairs 616
2.3 Motifs 616
2.3.1 Motifs and Function 616
2.3.2 Sarcin/Ricin Motif 617

3 Structure Building 618
3.1 Constraint Satisfaction Problem 618
3.2 Search Space 619
3.2.1 Nucleotide Frames 619
3.2.2 Variable Assignments 620
3.2.3 Building a GNRA Tetraloop 620
3.3 Backtracking 621
3.3.1 Experimental Constraints 621
3.3.2 Search Tree and Time Complexity 622
3.3.3 Backbone Construction 622
3.3.4 Cycles of Interactions 622
3.3.5 Probabilistic Search 623
3.3.6 Building by Fragments 623
3.4 *MANIP*: Interactive Computer Graphics 624

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

4 Refinement 625

5 Applications 626
5.1 *MC-Sym* 626
5.1.1 Yeast tRNA-Phe (1993) 626
5.1.2 Ribonuclease P RNA (1996–1999) 628
5.1.3 Hairpin Ribozyme (1999) 628
5.1.4 Leadzyme (1998–2001) 628
5.1.5 *Escherichia Coli* 16S rRNA Fragments (2001) 631
5.1.6 Prohead RNA (1998–2002) 631
5.1.7 Red Clover Necrotic Mosaic Virus Bimolecular Complex (2004) 631
5.2 *MANIP* 631
5.2.1 Group II Intron (2000) 631
5.2.2 Hammerhead Ribozyme (2003) 632

6 Perspectives 633

Bibliography 633
Books and Reviews 633
Primary Literature 633

Keywords

Base Pairing
Two nitrogen bases stabilized by the formation of hydrogen bonds. Two complementary RNA regions are connected by base pairing and form a double helix.

Base Stacking
Two nitrogen bases stabilized by induced dipole.

Graph
Data structure consisting of nodes and edges between pairs of nodes.

Leadzyme
A catalytic RNA that cleaves a specific ribophosphodiester bond in the presence of lead ions.

MC-Sym
Macromolecular conformations by symbolic programming.

Motif
A recurring structural element of biological significance.

Spanning Tree
A tree that includes every node of a graph.

Abbreviations

bp:	Base pair
CSP:	Constraint satisfaction problem
3-D:	Three-dimensional
HB:	Hydrogen bond
LSU:	Large ribosomal subunit
NDB:	Nucleic acid database
NMR:	Nuclear magnetic resonance spectroscopy
PDB:	Protein Data Bank
rRNA:	Ribosomal RNA
tRNA:	Transfer RNA
RMSD:	Root mean square deviation

The knowledge of the 3-D structure of an RNA molecule gives us indications about which other molecules it can interact with, and how its biochemical function is achieved or can be modified. Unfortunately, direct structure determination methods, such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and electron microscopy, cannot be applied at the rate of the genomics era yet, limiting the deciphering of RNA folding and structure. The recent resolution of the large ribosomal subunit (LSU) crystal structure represented a leaping step, but it has mainly pointed out the weaknesses of the current methods for analyzing and inspecting RNA 3-D structures. At this time, RNA sequences of predetermined function cannot be designed and folded properly. Further analysis and interpretations of the available 3-D structures are thus needed to bring additional insights into RNA function. Here, we present state-of-the-art data structures and algorithms that were developed in the last 10 years to analyze and inspect efficiently and objectively RNA 3-D structures. As you will realize by reading the following sections, they are the same as those employed in the development of the best computer modeling methods of RNA 3-D structures.

1
Goal

During the past 10 years, computational studies of RNA 3-D structures were focused on the gathering of experimental data and knowledge, as well as on the development of computer formalisms to represent structural data precisely and accurately. Formalized structural data and knowledge can be exploited by using computer modeling tools to generate precise and unbiased RNA 3-D structures. Computer generated 3-D structures give new insights about the function of a given RNA, which can be challenged experimentally. Again, the data obtained from experimental verification are sent for computer analysis; from which keener structural hypotheses will be made.

The flowchart in Fig. 1 indicates the principal components of computer modeling. The respective human modeler and computer contributions are shown separately. The arrows highlight the principal activities of computer modeling. The role of computer programs, delimited by a shaded box in Fig. 1, is to generate systematic and objective 3-D models. The modeler's contribution consists in developing computer formalisms and programs. In the context of a given data formalism, computer programs model 3-D structures: search structural motifs (Sect. 2), and build (Sect. 3) and refine (Sect. 4) 3-D structures. The resulting RNA 3-D models are interpreted by the modeler and help improve the computer formalisms and programs for the next development cycle. Symbolic annotation is another computer program that can assist the modeler in interpreting the 3-D models, by automatically identifying and formalizing structural data from 3-D models.

The long-term goal of computer modeling is to drive precise and efficient wet laboratory experiments by generating the most plausible 3-D models, and by supporting structural hypotheses that can be tested experimentally. Thus, the use of computer modeling of RNA 3-D structures is a tool in the hands of the experimentalists to catalyze the production of precise 3-D models (Sect. 5).

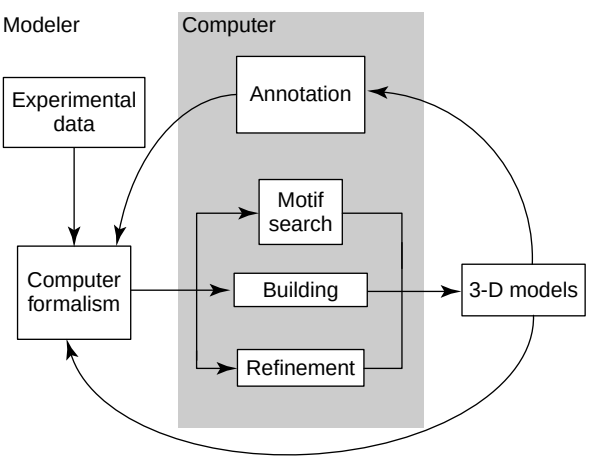

Fig. 1 Computer modeling development cycle. Boxes: data and computer programs; arrows: research activities and data flow.

2 Structural Knowledge

In this section, we introduce three levels of abstraction that are relevant to the representation of RNA 3-D structural knowledge and, thus, to the building of RNA 3-D computer modeling systems. The mathematical and computer formalisms were deduced from the observation of high-resolution RNA structures (X-ray crystal structures of 3 Å resolution or better), and serve as first-class objects in the description of 3-D structure building algorithms. It is mandatory for computer systems to describe formally the structural information to drive the building engine accurately. In Sect. 2.1, we present the RNA polymer and nucleotide conformations. In Sect. 2.2, we introduce the nucleotide noncovalent interactions, base pairing, and base stacking. Finally, in Sect. 2.3, a computational formalism to encode arrangements of nucleotide interactions is developed.

2.1 Nucleotide Conformations

RNAs are polymers of linked ribonucleotides (or nucleotides for short). As shown in Fig. 2(a), each nucleotide in the chain contains a ribose sugar, a phosphate group (P attached to four O atoms), and a nitrogen base (or simply base) that is one of adenine (A), guanine (G), cytosine (C), or uracil (U) (Fig. 2b). The ribose sugar is linked to the phosphate group through the C5′−O5′ diester bond. The O5′ is covalently linked to the P, which in turn is linked to the O3′ of the next nucleotide. The O3′ forms a second diester bond with the C3′, which is why the connection between two nucleotides in RNAs is referred to as a *phosphodiester linkage*. The ribose sugar is linked to the base through the C1′−N glycosidic bond (the N in the base is N9 in the purines A and G, and N1 in the pyrimidine C and U).

In general, computer modeling system developers define data structures that maintain at all time the atomic coordinates of all atoms, from which it is easy to compute the covalent bond lengths and angles, as well as the torsion angles (see Fig. 2c). As a fair approximation, the covalent bond lengths and angles are initialized at optimal and predetermined values that are kept fixed during modeling. The torsion angles, on the other hand, parameterize nucleotide conformations and are free to change. Six torsions are located along the phosphodiester linkage (α, β, γ, δ, ε, and ζ); five within the ribose sugar (θ_0 to θ_4), and one around the glycosidic bond (χ) (see Fig. 2a). The glycosidic torsion is measured using the atoms O4′−C1′−N9−C4 in purines, and using the atoms O4′−C1′−N1−C2 in pyrimidines. All torsions along the backbone, as well as the glycosidic torsion, are free to rotate almost independently. However, the ribose ring stereochemistry imposes a relation among the θ_0 to θ_4 torsions that can be expressed by a cosine function,

$$\theta_j = \theta_{\max} \cos(\rho + j\varphi) \qquad (1)$$

where $j = 0$ to 4 and $\varphi = 144^\circ$ ($720^\circ/5$). When $j = 0$, we have:

$$\theta_0 = \theta_{\max} \cos(\rho) \qquad (2)$$

In equation (1) and (2), ρ is the pseudorotation of the ribose ring. By varying ρ from 0 to 360° by steps of 90°, θ_0 goes from $\theta_{\max}$ to 0, to $-\theta_{\max}$, back to 0, and finally back to $\theta_{\max}$. The $\theta_{\max}$ value is reached twice, at the initial conformation ($\rho = 0$, as $\cos(0) = 1$) and at $\rho = 360^\circ$. At each

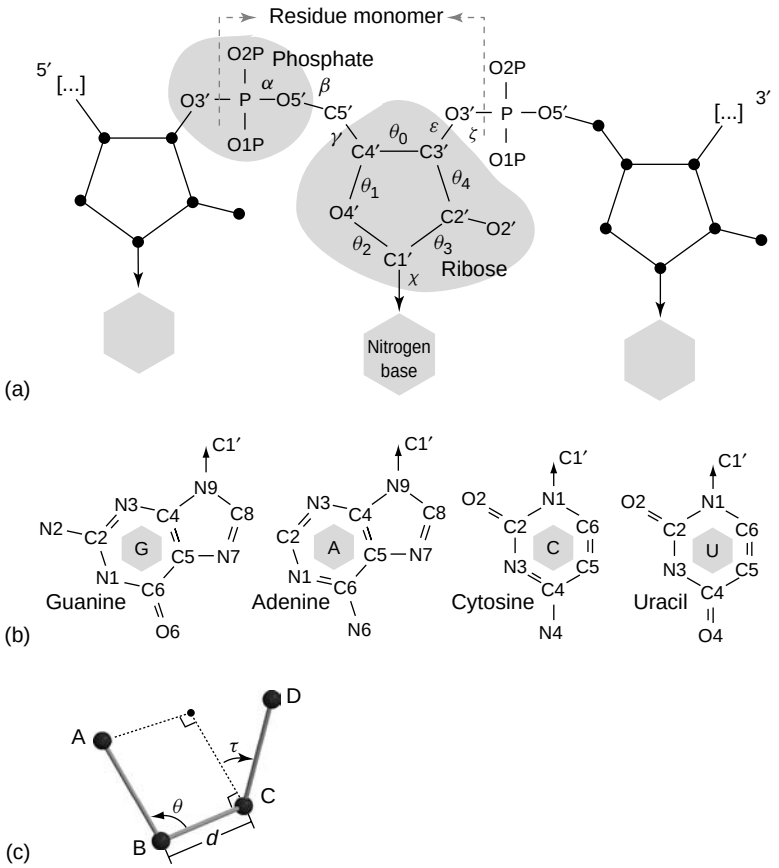

Fig. 2 Nucleotide. (a) Atomic numbering and torsion angles nomenclature presented on a suite of residue monomers, delimited by dashed arrows. The residue's subgroups are highlighted in light gray: phosphate group, ribose, and one of the four possible nitrogenous bases. (b) The four nitrogenous bases and their atomic numbering. (c) Definition of bond length (d), angle (θ) and torsion (τ) on four consecutively covalently bonded atoms A, B, C, and D.

step of $\rho + 180°$, the sign of all torsions is inversed, corresponding to the mirror image of the conformation at ρ. A useful equation is derived from Equation (1), which determines ρ:

$$\tan \rho = \frac{(\theta_2 + \theta_4) - (\theta_1 + \theta_3)}{2\theta_0(\sin 36° + \sin 72°)} \qquad (3)$$

Equation (2) determines θ_{max}.

The pseudorotation quantifies the puckering mode of the ribose, traditionally labeled by its "envelope" and "twist" (half-chair) geometries. The "envelope" geometry is observed when one atom is located over or below the plane formed by the four others, whereas the "twist" geometry is observed when one atom is over and another is below the plane formed by the three others. When the plane is horizontal and the C5′ atom is on the left side, the atoms located over the plane are said to be endo to the C5′, and those located below

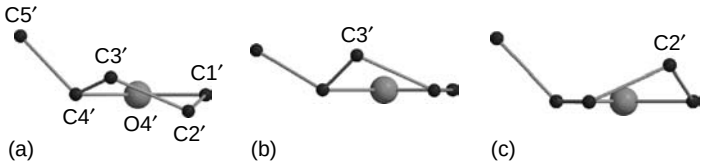

Fig. 3 Reference and main ribose pucker modes. (a) The $^2T_3$ twist shape C2′-*exo*–C3′-*endo*. (b) The ^3E envelope shape C3′-*endo*. (c) The ^2E envelope shape C2′-*endo*.

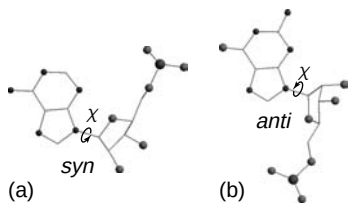

Fig. 4 The main RNA glycosidic bond torsions. The sphere's radius is proportional to the represented atom's proton mass: nitrogen, oxygen, and phosphate in ascending order. Carbons were omitted. The bond torsion is χ. (a) The syn torsion aligns the base and the backbone in the same direction. (b) The anti torsion aligns the base and the backbone in opposite direction.

the plane are said to be exo to the C5′. A convenient nomenclature is to use the letters "E" for *envelope* and "T" for *twist* conformations, and the atom number as an exponent for the endo, and as an index for the exo atoms. As described earlier, the convention is to choose $\rho = 0°$ when θ_0 is positive and maximum, which results in the C2′-*exo*–C3′-*endo* conformation, or $^3T_2$ using the nomenclature shown in Fig. 3a'. Note the mirror image of the initial conformation is $^2T_3$, reached at $\rho = 180°$.

RNAs are characterized by two principal "envelope" puckering modes, the C3′-*endo* (^3E) typical of the A-form helix (see Fig. 3b) and the C2′-*endo* (^2E) typical of the B-form helix (Fig. 3c). The B-form helix is prevented from forming because of steric conflicts created by the 2′-OH group of the ribose, although isolated B-form conformations may be found at the extremities of A-form helices, such as U_7 in the yeast tRNA-Phe, and within loops. RNAs are also dominated by two favored glycosidic bond torsions. In the syn conformation (Fig. 4a), $0° < \chi < 90°$, the bulk of the base is oriented toward the sugar group. In the anti conformation (Fig. 4b), $-120° < \chi < 180°$, the bulk of the base is in the direction opposite to the sugar group. The nucleotide conformations in both the A-form and B-form helices adopt the anti conformation.

Many attempts were made to classify the nucleotides conformations according to the six phosphodiester linkage torsion angles. In the 1970s, Olson presented a statistical analysis where the six phosphodiester torsions were reduced to two virtual angles spanning the two C4′–C–O–P segments. They reported a correlation among the single torsions showing that they are not completely independent. Gautheret, Major, and Cedergren have superimposed the P_n–$O3'_{n+1}$ atoms of dinucleotides (adjacent nucleotides) and used root mean square deviation (RMSD) to reduce the classification problem to one dimension. They were able to form a small number of dinucleotide clusters according to

the orientation of the P−P vector (backbone direction), and used them in early versions of their *MC-Sym* RNA modeling computer system to sample the conformational space of RNAs. They noticed a great variability of the torsion values in each cluster, also known as the *crankshaft* effect. More recently to analyze the ribosomal RNA (rRNA) structures, Duarte and Pyle have taken an approach similar to Olson's and simplified the six backbone torsions to two pseudorotations, $\eta(C4'_{n-1}-P_n-C4'_n-P_{n+1})$ and $\theta(P_n-C4'_n-P_{n+1}-C4'_{n+1})$. They reported torsion motifs found in consecutive nucleotides. Others have classified the torsions by using unsupervised machine learning approaches. Hershkovitz and coworkers have proposed discrete torsion values on the basis of Gaussian distribution data fitting of the four backbone torsions α, γ, δ, and ζ. They identified 37 nucleotide "signatures" (or conformers) in the large ribosomal subunit (LSU) crystal structure of *Haloarcula marismortui*.

Others have proposed the analysis of three torsions at a time in order to visualize the results of clustering in 3-D. Murray and her colleagues have identified 42 nucleotide conformers by applying quality filtering techniques to high-resolution crystal structures. They grouped the torsions in two triplets: (α, β, γ) and (δ, ε, ζ). The 3-D distributions of each triplet were plotted, and their peaks were visualized and aligned to define the nucleotide conformers. Each nucleotide conformer represents a high quality reference nucleotide conformation. Schneider and coworkers have analyzed the torsion angles of dinucleotides by Fourier averaging of six selected 3-D distributions. They found 18 dinucleotide conformers not belonging to the overrepresented A-type conformation family. Similar to Gautheret, Major and Cedergren, they concluded the structural conformational space of RNA 3-D structures, which could be sampled by a small number of dinucleotide conformers.

2.2 Nucleotide Interactions

In addition to the covalent linkage, two nucleotides can be related in 3-D space by base stacking and base pairing.

2.2.1 Base Stacking

Base stacking involves London dispersion between two bases that favor an arrangement where one base is stacked on top of the other. Base stacking can be evaluated by defining and comparing base planes. In pyrimidines, a base plane normal vector can be defined using a right-handed rotation around the base ring atoms N1, C2, N3, C4, C5, and C6 (see Fig. 5a). If the same atoms in the same sequence are used in purines, then the base plane normal vector is defined by a left-handed rotation. Considering the normal to the base plane vector orientations, then base stacking can adopt four different arrangements (Fig. 5b).

2.2.2 Base Pairing

Base pairing involves the formation of hydrogen bonds (HB) between hydrogen donor groups (mainly NH and NH_2) and acceptor groups (mainly CO and N) located along the base edges. The formation of two or more HBs induces a quasi-coplanar arrangement of the two interacting bases. Canonical G=C and A−U Watson–Crick bps (see Fig. 6) are the most frequent and dominate RNA folding. Stretches of stacked Watson–Crick bps adopt the A-RNA double-helix conformation, and are

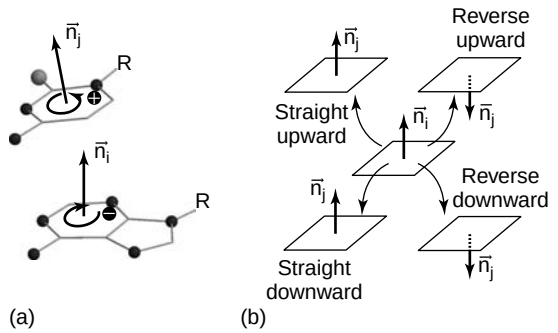

Fig. 5 Base stacking. (a) Straight upward base stacking between a purine (below) and a pyrimidine (above). A base plane in purines is defined in a left-handed coordinate system (−), whereas it is defined in a right-handed coordinate system (+) in pyrimidines. (b) The four possible relative orientations of base stacking according to the direction of the normal vectors to the base plane, n_i (origin base) and n_j (destination base). The bases are represented by squares and the origin to destination direction is indicated by the arrows.

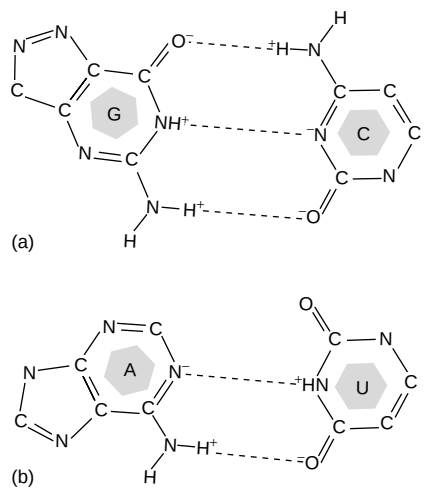

Fig. 6 Canonical Watson–Crick base pairs. The dotted lines indicate HBs. (a) G=C. (b) A−U.

called *stems* or simply *helices*; they form the RNA secondary structure. Dynamic programming is used to determine the optimal secondary structure of an RNA sequence, although the optimal and actual structures can differ considerably. In general, for simplification, helices are considered as rigid objects in computer modeling.

Non–Watson–Crick bps, often qualified as noncanonical, are also present in RNA 3-D structures. All sixteen combinations with various HB patterns have been observed. Their number varies, from 20% of the yeast tRNA-Phe crystal structure bps to as high as 50% of the LSU bps. The nucleotides not involved in the secondary structure have a tendency to stack in roughly helical conformation and to form noncanonical bps.

2.2.3 Base Pairing Nomenclatures

The classical and highly referred bp catalog of Saenger contains 28 base pairing patterns, which were labeled by roman numbers. However, roman numbers are not convenient to human discussions, and common names were thus given to the most frequent ones (cf. "wobble G−U",

"Hoogsteen A−U", "reverse Hoogsteen A−U", "sheared G−A", etc).

Leontis and Westhof proposed a more formal and systematic base pairing nomenclature (*LW*) to simplify discussions and to suggest bp geometries directly from the names. They defined for each base three contact edges: the Watson–Crick (*W*), Hoogsteen (*H*), and Sugar (*S*) (see Fig. 7a). The nomenclature is rapidly inferred from a quick observation of the 3-D structure of a bp, as the edges involved in

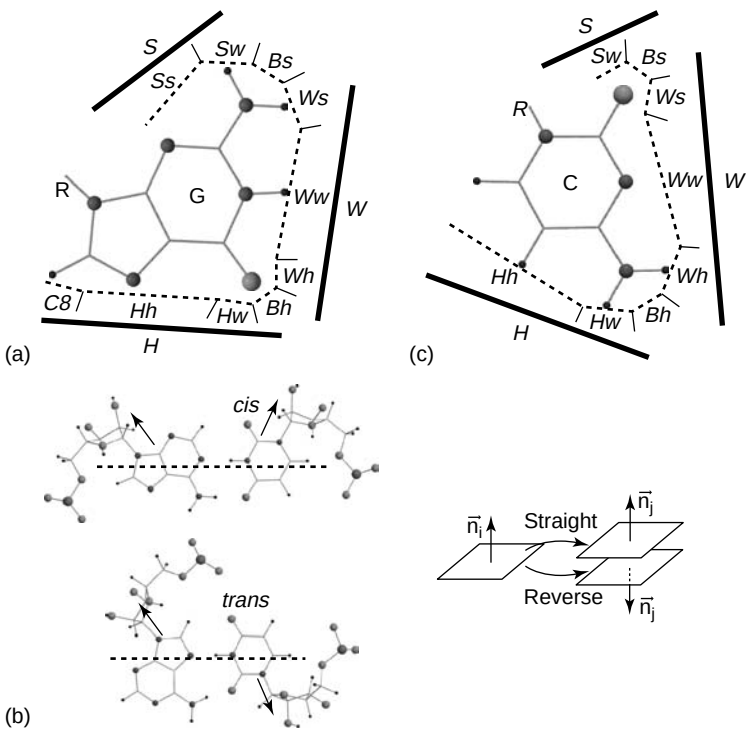

Fig. 7 *LW*+ base pair nomenclature. (a) The interacting faces and edges of guanine (G) and cytosine (C) (the same faces and edges were also assigned to adenosine and uracil). "*S*" sugar; "*W*" Watson–Crick; "*H*" Hoogsteen. The sphere's radius is proportional to the represented atom's proton mass: hydrogen, nitrogen, and oxygen in ascending size order. Carbons are omitted. The backbone connects at "R." The three *LW* edges are highlighted in dark bands, forming roughly a triangular shape around the base. The dotted lines delimit the *LW*+ faces. "*B*" stands for a bifurcated hydrogen bond, and "*C8*" is used to label an additional possible interaction that occurs only in purines. (b) The backbone relative orientation. The dotted line represents a plane perpendicular to an imaginary bp plane. As indicated by the arrows, cis: the backbone of both bases enters the same side of the perpendicular plane; trans: the backbones point in opposite sides of the perpendicular plane. (c) The base pair relative orientation. The curved arrows indicate HBs. The straight arrows indicate the normal vectors to the base planes, n_i and n_j (defined in Fig 5). Straight: the normal vectors point to the same direction; Reverse: the normal vectors are in opposite directions.

HB formation can easily be identified. The vectors formed by N1 (in pyrimidines) or N9 (in purines) to C1′ in each base complement the nomenclature by the assignment of the cis or trans relative orientation of the bp (see Fig. 7b). Using the *LW* nomenclature, a standard Watson–Crick bp is referred to as a "*W/W cis*" bp.

To increase the precision in computer modeling, Lemieux and Major slightly extended the *LW* base pairing nomenclature by subdividing each edge in faces (*LW*+), and resolved small ambiguities introduced by the formation of single HB bps (see Fig. 7a). For instance, in *LW*, a GU wobble bp is named "*W/W cis*" (the same as Watson–Crick), whereas in *LW*+ it is more precisely named *Ww/Ws cis*. If the relative orientation between each base plane normal vectors is added ("straight" is used for the same orientations, and "reverse" is used otherwise; see Fig. 7c), the bp geometry described by the nomenclature is completely unambiguous.

Lee and Gutell proposed an alternative base pairing nomenclature (*LG*), after visually examining the bps present in the LSU crystal structures of *Thermus thermophilus* and *H. marismortui*. They presented a topological relationship that starts at the canonical C=G or A−U *cis* Watson–Crick bps, or at the *cis* wobble G−U bp, and that extends to 14 families by successive manipulations of the base planes and glycosidic bond relative orientations: shearing, flipping, reversing, paralleling, or slipping. The 14 families are the Watson–Crick (*WC*), wobble (*Wb*), slipped Watson–Crick (*sWC*), slipped wobble (*sWb*), reversed Watson–Crick (*rWC*), reversed wobble (*rWb*), Hoogsteen (*H*), reversed Hoogsteen (*rH*), sheared (*S*), reversed sheared (*rS*), flipped sheared (*fS*), parallel flipped sheared (*pfS*), parallel sheared (*pS*), and reversed parallel sheared (*rpS*).

2.2.4 Structure Annotation

Structure annotation is an activity consisting of symbolic labeling of the nucleotide conformations and interactions from an input RNA 3-D structure. Annotated 3-D structures are simpler to analyze than 3-D coordinates. The annotation of nucleotide conformation is done by inferring the sugar pucker modes and the glycosidic torsion directly from the nucleotide torsion angles. The annotation of base stacking is done by computing and comparing the normal base plane vectors. Finally, the annotation of bps requires the computation of the HBs followed by the appropriate name assignment from a chosen nomenclature.

Structure annotation can be done by interactive computer graphics, by looking at a structure and describing the conformations and interactions. Structure annotation can also be automated by computer programs. Using the JUMNA system (see Sect. 4), Harvey and coworkers have implemented molecular mechanics restraints to identify bases that stack and pair, but they did not implement any labeling. Gendron, Lemieux, and Major have developed a computer program, *MC-Annotate*, which identifies geometrically and labels the bases involved in stacking and pairing. Bases that stack are identified by applying the Gabb and coworkers geometrical method. The bps are identified by applying an unsupervised machine learning approach in which three HB geometric parameters were defined and modeled using a series of Gaussian distributions. Potential HBs are identified and are assigned probabilities. For each identified bp, the probabilities are used to determine the most likely HB pattern that applies. It is then easy to assign the appropriate bp name according to any of the proposed nomenclatures; although the current implementation uses the *LW*+

nomenclature. The *MC-Annotate* program has been made available on the Web, and can be queried online (currently at www.lbit.iro.umontreal.ca/mcannotate).

Westhof and coworkers have proposed a simpler annotation system that was implemented in the *RNAView* program. Their annotation is fully based on geometrical features, and predetermined cutoff values are used to accept or reject a bp. The labeling is made according to the general *LW* nomenclature. *RNAView* is accessible on the Web (currently stored at www.ndbserver.rutgers.edu/services).

2.2.5 Isosteric Base Pairs

Two bps that preserve a local 3-D structure are said to be isosteric, and can appear in evolutionary related structures to preserve function. Leontis and coworkers have superimposed the geometry of all possible bps according to their C1′−C1′ distances and cis/trans base orientations. They then mapped all 16 combinations to the 12 families of their *LW* nomenclature, which resulted in isostericity matrices that can be consulted on the Web (currently stored at www.bgsu.edu/departments/chem/RNA/pages). It was shown clearly from these studies that all standard Watson–Crick combinations are isosteric, and the wobble G−U bp is isosteric to the protonated A−C bp.

Walberer and coworkers proposed to quantify the isostericity between two bps by a value reflecting the overlapping of the four glycosidic bonds, a measure more representative than the C1′−C1′ distances. They observed high isostericity values not only in the helical bps but also in several all purine and all pyrimide base pairing overlaps. So far, base pairing studies have supported the hypothesis that evolution maintains RNA structure more tightly than sequence. Consequently, covariations reflecting bp isostericity must be evaluated when comparing RNA sequences. Accurate sequence comparison requires precise sequence alignment to ensure that the positions compared are indeed homologous. Despite the practical problems, sequence comparison, from which covariations can be identified, greatly benefits secondary structure determination.

2.3 Motifs

The function of RNA is established by the presence of specific and recurrent arrangements of base interactions, other than the canonical stems. Researchers in the field refer to these arrangements as *motifs*. Prior to the resolution of the LSU, many of these motifs were predicted from comparative sequence analysis. Short oligomers containing the predicted motifs have been experimentally studied, and X-ray crystallography and nuclear magnetic resonance (NMR) assays have revealed their 3-D structures. Later, the motif structures were confirmed to form in the crystal structure of the LSU as well.

2.3.1 Motifs and Function

As RNA motifs point to function, the search for new motifs in the LSU, and in other previously determined 3-D structures, has thus become the focus of several theoretical studies. RNA motifs are thermodynamically stable and fold similarly in many different environments, whether in short or large structures, suggesting they might be considered to be the fundamental RNA building blocks. Consequently, the determination and prediction of RNA motifs could help study the RNA folding mechanisms, as well as predict the native structure of RNA directly from its sequence of nucleotides.

There are a limited number of published RNA motif studies, and they are generally made by teams of researchers that have introduced computer formalisms for nucleotide interactions (see Sect. 2.2).

Inside a computer, a data structure that is convenient to represent RNA motifs is the graph. An RNA graph has nucleotide nodes that are connected by interaction edges. Consequently, the RNA motif searching problem can be reduced to matching RNA subgraphs (putative motifs) to full RNA graphs (structures). Major and his coworkers have implemented a motif searching engine, *MC-Search*, which uses the annotations described earlier from the augmented Leontis and Westhof nomenclature, *LW+*, and the computer program *MC-Annotate*. The input of *MC-Search* is the target RNA subgraph (motif) and the RNA structures to be scanned.

2.3.2 Sarcin/Ricin Motif

For example, searching for the generic sarcin/ricin motif (see Fig. 8) in the *H. marismortui* 23S and 5S rRNA structures generates five hits, which share a minimal RMSD of 0.73 Å (An RMSD of 1 Å or less is considered very close for structures of this size.).Note that no coordinates are given in the input, as the subgraph for the motif contains only the symbols

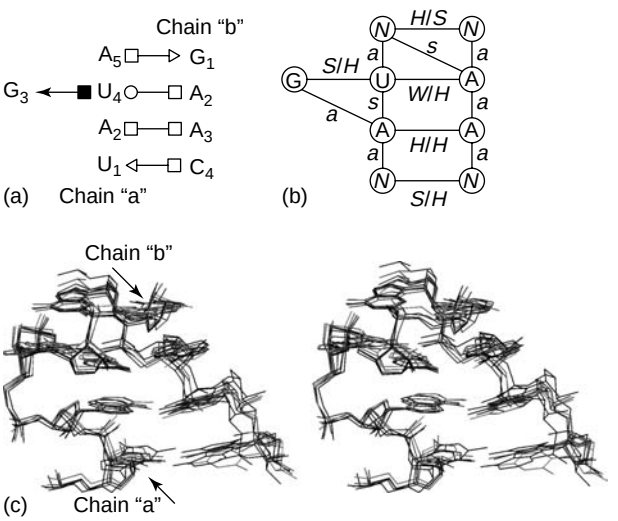

Fig. 8 Sarcin/ricin motif. (a) Base interactions. The *LW* annotation is used. The circles indicate the *W* edge, the squares the *H* edge, and the triangles the *S* edge. The white circles, squares, and triangles indicate the trans backbone orientation, whereas the black symbols indicate the cis backbone orientation. (b) RNA graph used for the input of *MC-Search* to find the sarcin/ricin motifs in *H. marismortui* 23S and 5S rRNA. "*a*" indicates sequence adjacency and "*s*" indicates base stacking. The letters that label the edges correspond to the *LW* base edges. "*N*" stands for any base. (c) Stereo view of five superimposed motifs matched by *MC-Search*. The hydrogen atoms were omitted.

representing each nucleotide interaction. These motifs were also found by Leontis and his coworkers by visual examination. Interestingly, bp S/H $U_{a1}C_{b4}$ can be substituted by an S/H C−C or G–A bps, as seen in other instances of the motif; represented by the generic S/H N−N interaction in the RNA graph of the motif (see Fig. 8b). These three bps belong to the same isosteric family, as discussed in Sect. 2.2.5, and as confirmed by the very small RMSD. Furthermore, if we remove the nonadjacent stacking interactions, $A_{a2}U_{a4}$ and $A_{a5}A_{b2}$, from the target motif, a sixth motif occurrence is found. The new motif instance, which does not include the nonadjacent stacking interactions, shares a RMSD of 5.2 Å with the five others, strongly suggesting the nonadjacent stacking interactions are a requirement of the sarcin/ricin motif.

The SCOR (structural classification of RNA) database contains several examples of published and visually identified RNA motifs and their references to the appropriate 3-D structures.

3 Structure Building

There are two major approaches to RNA 3-D structure modeling: folding and assemblage. Folding methods start with an initial model, usually extended or randomized. The initial model is then successively modified until a final and satisfactory model is identified. The conformational space of folding methods is defined by the number of models that can be reached from the modifications.

Molecular mechanics is a folding method that uses an objective function to evaluate the succeeding models and guide the modifications toward optimal states. Harvey and his coworkers have developed an objective function for RNA that is mainly composed of experimentally determined nucleotide interaction constraint terms in order to emphasize their formation in the final models. They defined in their computer modeling system, *YAMMP*, one to five points per nucleotide to simplify the folding operations, as they wanted to apply it to large RNAs. Other terms in the objective function were present to ensure the appropriate stereochemistry of the models. *YAMMP* has been used by Harvey and coworkers to model the subunits of the ribosomal RNA in the early 1990s.

Assemblage methods start from scratch. The components of the RNA are assembled using construction operators that place them in 3-D space, relatively to the other already positioned components. In this section, we describe automated and interactive assemblage methods, as they are the two variants that were mainly employed in the last 10 years to propose RNA 3-D structures. We will describe *MC-Sym*, a constraint satisfaction problem solver, and *MANIP* an interactive computer graphics system.

In Sect. 3.1, the constraint satisfaction problem is introduced. In Sect. 3.2, the search space of RNA 3-D structures is defined. In Sect. 3.3, the backtracking algorithm to solve the RNA CSP (constraint satisfaction problem) is presented, and, in Sect. 3.4, the interactive computer graphics approach is described.

3.1 Constraint Satisfaction Problem

The RNA 3-D structure building problem can be described in terms of a more general problem, well known by computer scientists: the constraint satisfaction

problem or CSP. The CSP is defined by a set of variables that need to be assigned discrete values from finite domains to satisfy a set of predetermined constraints. The number of variables and the size of the domains define the CSP search space. Solving the CSP consists in finding one, many, or all possible assignments that respect the constraints. The constraints are applicable to one or more variables and restrict the number of possible assignments.

A classical CSP is the 8-queen problem: 8 queens must be positioned on a standard 8×8 checker board in such a way that each queen is safe from the others. In the 8-queen problem, we have 8 variables (the queens), the 64 possible positions for each queen, and three constraints; two queens cannot be positioned in the same row, column, or diagonal. There are 92 solutions to this problem; 92 ways of positioning 8 queens on a regular 8×8 checker board so that no queen can attack another.

3.2 Search Space

3.2.1 Nucleotide Frames

To help visualize RNA search space, consider an RNA structure as a set of nucleotides that are all related to each other by spatial transformations composed of rotations and translations. Major and coworkers have parameterized nucleotide transformations by defining a local base referential, or frame, whose origin is placed on the terminal nitrogen atom (N9 for purines and N1 for pyrimidines), and *XY* plane aligned to the base plane (see Fig. 9a).

The frame of a base, A, is represented by the 4×4 homogeneous transformation matrix, ${}^{O}T_{A}$, that encodes the transformation (rotation and translation), which applied to the global origin, O, results in the frame of base A. Symmetrically, applying the inverse matrix of A's local base frame, ${}^{O}T_{A}{}^{-1}$, results in O. From the local base frames, one can compute the transformation matrices that relate any base interactions in an RNA 3-D structure. Consider the bp in Fig. 9a, the relative transformation from frame A to frame B is expressed by the product ${}^{A}T_{B} = {}^{O}T_{A}{}^{-1} \times {}^{O}T_{B}$.

Now, to reproduce an observed relation, ${}^{M}T_{N}$, when M is already positioned, we apply the transformation ${}^{O}T_{N} = {}^{O}T_{M} \times {}^{M}T_{N}$ to the coordinates of base N. First, we need to place N in M's frame, by using the transformation ${}^{M}T_{N}$, and, second, since N is now local to M, we move N by the global frame of M, ${}^{O}T_{M}$ (see Fig. 9b). The coordinates of base N in the previous example are assumed to be expressed in the global O, a task

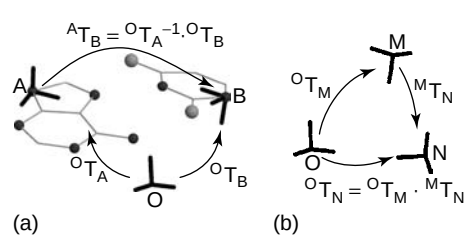

Fig. 9 Base frame transformation. (a) Base coordinate frame transformation. The base A and B local referentials (frames) and the origin global referential are shown using small bold axis systems. The frame transformation, ${}^{A}T_{B}$, allows us to express the B to A relation independently of the position of base A and of the global referential, O, and can be computed directly from base A and B frames, ${}^{O}T_{A}$ and ${}^{O}T_{B}$. (b) Application of base frame transformations. Base N is first put in the local referential of base M, by using the base frame transformation ${}^{M}T_{N}$, and then to the global referential by using the base frame transformation ${}^{O}T_{M}$.

that can be precomputed to accelerate building.

3.2.2 Variable Assignments

The building procedure consists in positioning n bases using $n-1$ transformations, and can be represented by an ordered spanning tree of the RNA graph (see Fig. 10). The $n-1$ transformations are mapped to the edges of the spanning tree, which define the CSP variables. In Major's laboratory, a computer program has been designed to extract the frames of all nucleotides in available RNA structures, and then using the frame manipulation given above, to produce a database of RNA base relative transformations. The list of 4×4 matrices (mainly obtained from the structures of high resolution in the PDB and NDB) is mapped to each type of nucleotide interactions described in Sect. 2.2.

The RNA CSP search space is thus defined by the Cartesian product of the transformation matrices from the database that can be assigned to each edge in the spanning tree of the RNA graph. In practice, the relevant characteristics of each interaction are specified by the user. The computer modeling program synchronizes with the database to extract a set of transformations that optimizes the search space according to the user's characteristics. The other task of the user is to define the ordered spanning tree, as only one tree is solved by run. If several trees need to be considered, then they must be defined and searched independently.

3.2.3 Building a GNRA Tetraloop

Consider modeling the 3-D structure of the 5′-CGGGAG-3′ sequence. The task of the user, as mentioned earlier, is to introduce an ordered spanning tree of the RNA graph

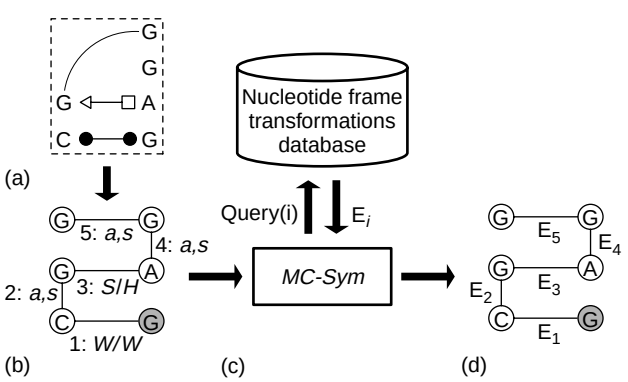

Fig. 10 *MC-Sym* search space for the GNRA motif. (a) Base interactions. The annotations follow *LW* (as in Fig. 8). (b) A user chosen ordered spanning tree and edge characteristics for (a). "*a*" stands for sequence adjacency, and "*s*" for stacking. The shaded base is the spanning tree root, or modeling reference. (c) *MC-Sym* database communication. The transformation set for each edge in the spanning tree is extracted from the nucleotide frame transformations database. The ith edge translates in query(i), which returns transformation set E_i. (d) Search space. The Cartesian product of the transformation sets defines the search space, and *MC-Sym* is ready for backtracking.

of Fig. 10(a), by employing the symbols described in Sect. 2.2. Figure 10(b) shows one possible ordered spanning tree of this RNA graph. The input spanning tree instructs *MC-Sym* to obtain from the database the transformations that match the input characteristics of each interaction edge in the tree (Fig. 10c), defining this modeling problem's search space.

Here the user selected the G in the Watson–Crick bp as the root of the spanning tree. Figure 10(d) shows the user chosen order in which *MC-Sym* will build the structures. The C of the Watson–Crick bp is placed from the G using *W*/*W* cis transformations. The edges are labeled E_1 to E_5. E_1: given C is placed, the G of the G–A bp is positioned by applying adjacent/stack transformations from the C in the Watson–Crick bp. E_2: the G of the G–A bp is positioned by applying adjacent/stack transformations from the C in the Watson–Crick bp. E_3: the A of the G–A bp is positioned by applying S/H cis transformations. E_4: the G that stacks on top of the A in the G–A bp is placed by applying adjacent/stack transformations. Finally, E_5: adjacent/stack transformations are also applied to get the final G in the tree.

3.3 Backtracking

In the *MC-Sym*'s search space defined earlier, one transformation assigned to each edge in the spanning tree places the bases of the structure in 3-D space. The chosen transformations are applied systematically to the edges of the ordered spanning tree. However, what happens if an assignment creates atomic clashes in the structure? *MC-Sym* includes atomic clashes as constraints to be satisfied, typically by defining a distance threshold of 1 Å on all atoms. Consequently, the constructions containing atoms positioned closer than 1 Å apart are eliminated.

When a constraint is not satisfied or all transformations for a given edge have been tried, *MC-Sym* returns (backtracks) to the last assigned edge (variable). The next transformation is selected and the search continues until all transformations have been tried for each edge.

3.3.1 Experimental Constraints

User constraints that address the nucleotide conformations, such as sugar pucker modes, glycosidic torsions, atomic distances, bond angles and torsions, or base interactions can also be added to the constraints to ensure the generated structures include the structural features that were observed experimentally or predicted theoretically.

Theoretical bps can be derived from secondary structure, covariation and tertiary interaction predictions. Experimental bps can be derived from site-directed mutagenesis, a palliative to the lack of sequence variation. Monitoring of the RNA activity while specific and strategic bases are replaced can detect base interactions. The use of modified nucleotides can detect the participation of specific chemical groups, such as hydroxyls, phosphates, oxygens, amino, and imino, in the activity and interactions. The use of chemical modifications can provide enough data for the construction of accurate local sites, and in particular catalytic sites. Chemical and enzymatic probes cleave at specific sites or attack specific groups that are exposed to the solvent. The results of such probing are highly qualitative rather than quantitative, but, nevertheless, have been used to confirm secondary and tertiary structure elements of several RNAs. Cross-linking regions of an RNA by creating artificial covalent bonds, using for

instance UV irradiation, allows one to establish the proximities of some atoms and nucleotides that can then be used as powerful constraints in RNA modeling. NMR is a powerful technique that provides precise interatomic distances that can then be translated easily in nucleotide conformation and interaction characteristics.

3.3.2 Search Tree and Time Complexity

The time complexity of the backtracking algorithm is exponential in both the number of edges (N) and the number of transformations for each edge (M). The backtracking procedure develops a search tree, where each node is a partial assignment. The depth of the search tree is determined by N, and its width is determined by M. When a partial assignment is invalidated, the portion of the search tree under the current assignment node is pruned, as no valid solution can exist in that portion of the search tree. Assuming the complexity for checking partial assignments is constant, the time for completing a search is reduced proportionally to the efficiency of pruning the search tree. Consequently, in practice *MC-Sym* applies to small or highly constrained RNAs.

3.3.3 Backbone Construction

How is the backbone handled? So far, we have paid attention to the bases and their interactions. Here we turn the focus to the backbone. In the original versions of *MC-Sym*, the whole nucleotide (base, phosphate group, and ribose) was moved with the base transformations. In this context, the nucleotide was considered a rigid body sampled by various atomic coordinates, corresponding to the sugar puckers modes and glycosidic torsions specified by the user (see Sect. 2.1). The nucleotide conformation sampling increased the search space. The use of rigid nucleotides created an additional constraint to be verified, the length of the O3′-P bond, which connects two consecutive nucleotides in the sequence.

However, the results of recent studies on RNA backbones by Thibault and Major have shown that a more efficient approach can be adopted, where no sampling of the nucleotide conformation database is needed. The motivation came from the hypothesis that the forces involved in the formation of the bps drive RNA folding and constrain the backbone bends and twists to adopt conformations that satisfy the base pairing arrangement. In other words, the backbone conformation is a consequence of base pairing.

The new RNA 3-D structure building engine positions only the actual rigid components of the nucleotides: the bases and phosphate groups. The coordinates of the interconnecting riboses are generated from an optimization process in constant time. The lengths of the bonds that link the phosphate groups to the ribose replaces the former O3′-P adjacency constraint during the backtracking. The absence of backbone sampling considerably reduces the size of the CSP search space, while simplifying the task of the user since no nucleotide conformation characteristics need to be input. Note, however, that constraints on the backbone can still be described and are checked.

3.3.4 Cycles of Interactions

The introduction of the spanning tree to generate RNA 3-D structures implies a conceptual pitfall. In the RNA graph, the structural information in the edges that are not covered by the spanning tree is not contributing to the building.

Take for example the construction of two-stacked Watson–Crick bps. One spanning tree would employ the two base pairing edges and only one stack edge. The solution structures could then include Watson–Crick bps that stack on one side only. User constraints need to be introduced to make sure the stack information of the dropped edge is satisfied. However, *ad hoc* user constraints to represent sophisticated geometrical concepts such as base stacking or pairing are difficult to define.

Lemieux and Major have designed a structure building approach that addresses cycles of base interactions to overcome this pitfall. They realize that successive applications of edge transformations in a cycle results in the identity matrix. Consequently, a solution to the cyclic structure building problem is defined by the assignment of a transformation to each edge in the RNA graph so that for each cyclic subgraph the product of the transformations is close to the identity matrix.

3.3.5 Probabilistic Search

Exhaustive searches for all valid RNA 3-D structures are useful to analyze the possible alternative folding of an RNA. However, sometimes only a single valid structure is desired rather than all possible ones. Recent works on *MC-Sym* have shown that a probabilistic search algorithm generates valid structures more rapidly and with an increased diversification rate than the deterministic backtracking. The probabilistic algorithm successively selects a transformation at random. If a partial construction is invalid, then a fixed size backtracking search is launched. The backtracking runs until the constraints are satisfied, or until a maximum number of backtracking steps is reached. In the former case, the random assignments resume to the next edge in the RNA spanning tree. In the latter case, the current valid substructure is lost and the probabilistic search restarts at the first edge.

Benchmark results indicated that the probabilistic approach accelerates the generation of valid RNA 3-D structures if the search space is increased up to a critical point. This particular feature of the probabilistic search allows us to efficiently explore fine-grained conformational spaces. Exhaustive exploration of a coarse-grained search space, or nonexhaustive exploration of a fine-grained search space, is always a choice the modeler can make.

3.3.6 Building by Fragments

As in many problems, the "divide and conquer" paradigm has proven useful in RNA 3-D structure building. The divide-and-conquer algorithm splits a complex problem in many smaller and simpler to resolve subproblems. The solutions of the smaller problems are then joined together to compose the solutions of the larger problem. RNA structures can be easily split into smaller fragments that can be modeled independently, and then the small models for each fragment can be merged to form the larger structure.

MC-Sym can manipulate a fragment in the same way it manipulates a single base, by selecting one base of the fragment to define its local frame as seen in Sect. 3.2.1. The fragment coordinates can either be preconstructed by *MC-Sym* modeling, or extracted from a structure database. Fragment building of RNA 3-D structures is a direct application of RNA 3-D motif searching (see Sect. 2.3). In Major's laboratory, different ways of fragmenting RNA graphs are studied. They will eventually put together their results to provide a database of RNA fragments that

will fit an automated fragmentation of any RNA 3-D graph.

Consider the example in Fig. 11. The sarcin/ricin RNA graph can be fragmented in five independent and cyclic subgraphs. Each cyclic fragment can be solved independently by the cyclic RNA 3-D structure building approach. The final structure is the result of merging the independent solutions by applying transformations that superimpose the common edges.

3.4 *MANIP*: Interactive Computer Graphics

An alternative to automated RNA 3-D structure building is to use interactive computer graphic systems, where the

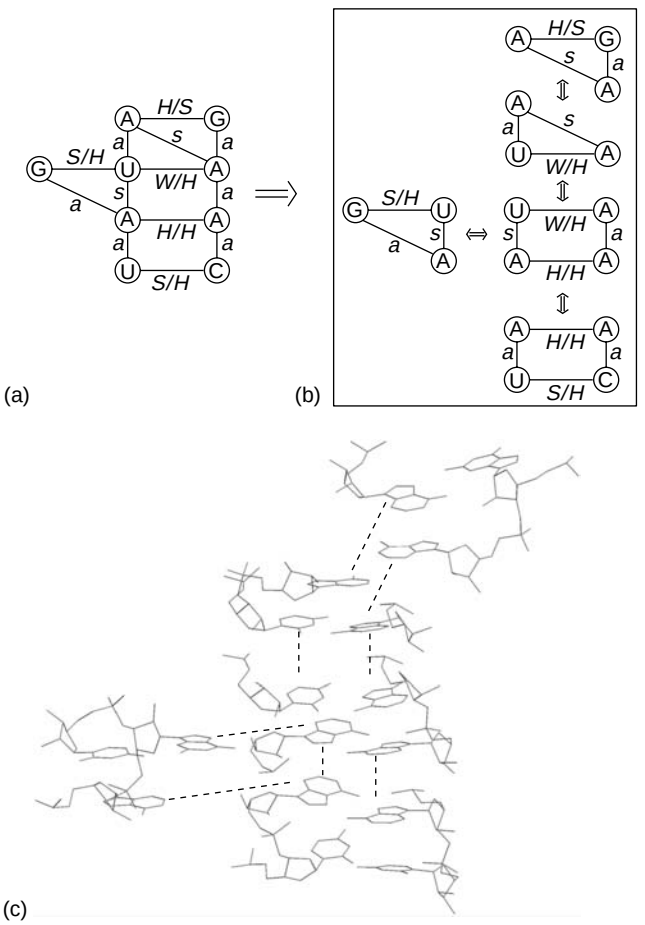

Fig. 11 Sarcin/ricin motif revisited. (a) RNA graph. See caption of Figure 8b. (b) Cyclic subgraphs. The fragmentation of the sarcin/ricin RNA graph results in five independent cyclic subgraphs. (c) 3-D structure. The independent results for each fragment are merged together to produce the motif 3-D structure. The dotted lines indicate the nucleotides shared by two fragments, and that need to be superimposed to construct the final structure.

user can interactively manipulate RNA structures and fragments in 3-D space. Such systems also allow the user to build and assemble RNA fragments in complete and larger RNA structures. Stereo chemical constraint and force fields are dynamically applied to limit the user to valid manipulations and buildings.

A popular system is *MANIP*, developed by Massire and Westhof. *MANIP*'s graphical interface gives the user the flexibility that is needed to create and move fragments by using the desktop keyboard and mouse, or more sophisticated input devices such as dials and wheels. The user's "artworks" are refined by applying integrated molecular mechanics minimization, here NUCLIN–NUCLSQ. Minimization ensures a correct stereochemistry at the nucleotide conformation level, removes the bad atomic contacts, and corrects the bad bond angles and torsions. The force fields implemented in NUCLIN–NUCLSQ was especially developed for RNAs.

The refinement step also allows for the identification of potential HBs and bps by matching their geometries to a database of known bps. Systems like *MANIP* are limited by the content of their database, here predefined fragments and bps. The major drawback of such interactive systems is the need for in-depth knowledge of RNA 3-D structure combined with strong experience in RNA 3-D modeling. But still, the *MANIP* system has proven useful in the context of sketch modeling to provide initial insights and starting points for further sharper modeling and structure determination projects.

ERNA-3D is another interactive system that was developed by Mueller and Brimacombe. It was applied to the modeling of the 3-D folding of the *Escherichia coli* 16S rRNA that fitted a 3-D electron microscopic map at a resolution of 20 Å.

4 Refinement

The structure building methods presented in Sect. 3 are mainly based on geometrical parameters and use rigid and discrete partitioning of RNA atomic structures. There is a need for further refinement of the resulting structures, and to fix the RNA internal geometry such as covalent bond lengths, angles, and torsions. Molecular dynamics fill this need.

The classical approach to simulate the molecular dynamics of RNA is to model at the atomic level the force fields that apply. This is achieved by developing an empirical model of potential energy function describing the known physical properties of RNA structures. The Newton's classical motion equations are applied and solved over discrete time steps from the evaluation of derivatives of the potential energy function. The equations can be solved by numerical optimization or heuristic methods. As shorter time steps are needed to obtain a reliable simulation, more steps need to be evaluated to reach energy minima. The input RNA 3-D structures of molecular dynamics need to be of good precision, which is not always possible due to the approximations made in modeling. Molecular dynamics can only be used as an RNA 3-D structure building procedure when a large number of precise constraints are known, such as to resolve NMR structures, and is rarely used as a modeling tool otherwise.

The critical part of a molecular dynamics simulation system lies in force field modeling. Three energetic interactions prime the force field: (1) a bonded interaction term, (2) a nonbonded charged interaction

term, and (3) a nonbonded noncharged interaction term. Bonded interactions result from distortions in covalently bonded atoms, and are measured by bond length, angle, and torsion variations from empirical data. Nonbonded charged interactions result from Coulomb electrostatic forces between charged atoms. Finally, a dispersion–repulsion force occurs between nonbonded and noncharged atoms by induced dipole–dipole interactions, and is often modeled by a Lennard–Jones potential, which is lightly attractive as two uncharged molecules approach from a distance, but is strongly repulsive when they are pushed close to each other.

Amber (assisted model building with energy refinement) is a widely used force field implementation. It was initially parameterized for proteins, but more recently a generalization has made it convenient for any organic molecule. The molecular dynamics simulation program *JUMNA* (JUnction Minimization of Nucleic Acids), has been developed specifically for nucleic acids. *JUMNA* has two additional terms to its potential energy function to restrain base pairing and stacking. Lavery and his colleagues have implemented a harmonic potential energy function to consider explicitly the geometrical properties of base stacking and pairing. The base center-to-center distances and angles are measured between the base normal vectors, which parameterize stacking interactions. HB lengths and the donor acceptor angle are used to parameterize base pairing interactions.

5 Applications

This section is dedicated to applications of computer modeling of RNA 3-D structure. Section 5.1 introduces applications that employed *MC-Sym*, whereas Sect. 5.2 discusses applications of *MANIP*. The applications in each subsection are introduced chronologically.

5.1 *MC-Sym*

5.1.1 Yeast tRNA-Phe (1993)

The classical benchmark of computer modeling of RNA 3-D structure is the yeast tRNA-Phe. In 1993, Major and coworkers demonstrated some capabilities of their computer modeling system *MC-Sym* by modeling the benchmark with great atomic precision. They generated 26 different structures sharing RMSD with the X-ray crystal structure from 3.1 to 3.8 Å, prior to their postprocessing refinement by energy minimization. They introduced other structural information that was available in the 1960s: the secondary structure, and the U8–A14, G9 = C56, G10–G45, U54–A58 bps. They also assumed stacking between the amino-acceptor and D-stems, as well as between the anticodon and T stems. The loops were modeled independently and only the best loop structures (smallest RMSD with the crystal) were kept and appended to the general construction. A key constraint to obtain L-shaped structures was bp U8–A14, as determined by UV irradiation cross-linking. The theoretical size of the search tree defined by the input of the tRNA constraints is 10^{26} nodes. The constraints pruned the search tree efficiently and only about 10^5 nodes were explored by the backtracking algorithm. In the 1960s, the most successful model was built by Levitt who proposed parallel coaxial stems (amino-acceptor stem and D-stem stack on each other and are positioned parallel to the anticodon stem and T stem that stack on each other)

rather than the typical L-shape of the crystal structure. The efficiency of the *MC-Sym* search was astonishing, and could be explained by the fact that a limited number of transformations were available, including a large fraction from the tRNA; the only large RNA structure at that time.

Nowadays, with the availability of many more high-resolution RNA structures, in particular the LSU, the transformation database (see Sect. 3.2.2) has grown to the 10^4 entries, ideal to test the new probabilistic version of the modeling program. Therefore, using the same input as in 1993, the probabilistic search method found 80 different 3-D structures exploring a search tree of 10^{38} nodes during seven days. The RMSD between the generated structures and the crystal structure range from 5.2 to 12.6 Å and were much higher than for those obtained by Major and coworkers in 1993. However, if the RMSD is evaluated among the stem regions only (loops removed), then the value range is 2.3 to 9.9 Å (see Fig. 12). Thus, the main structural differences with the 1993

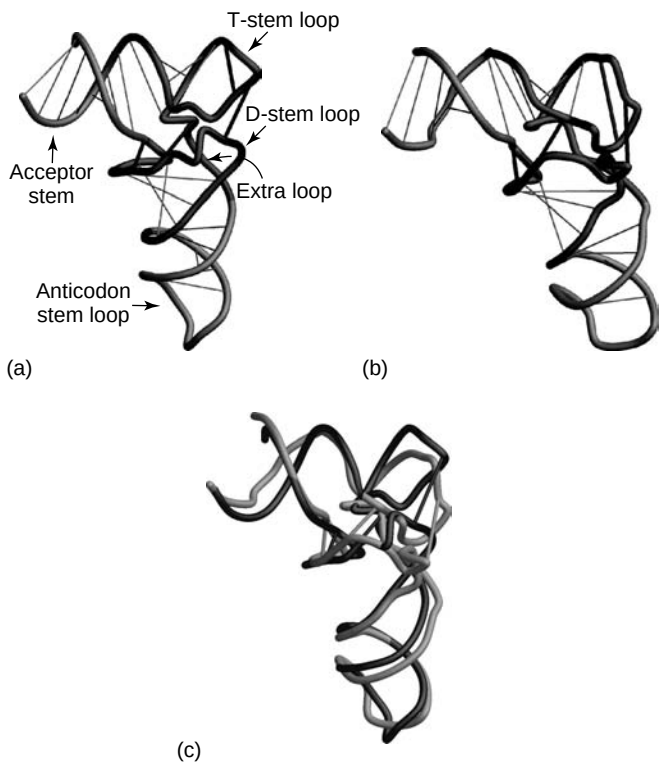

Fig. 12 Yeast tRNA-Phe. (a) Backbone representation of the crystal structure. The thin lines indicate secondary structure bps, and the thick lines indicate tertiary bps. (b) Backbone representation of the model, as generated by *MC-Sym* (probabilistic version). (c) Superimposition of the crystal structure and model. Dark backbone: crystal; light backbone: model. All atom but H in all nonloop nucleotide RMSD: 2.3 Å.

models are in the loops. No constraints were specified in the probabilistic search, whereas in the 1993 models, the loops were fine-modeled prior to the global search. The probabilistic search method found better solutions and generated a greater diversification of structures. Observations of the generated structures revealed that the base pairing between U8 and A14 is most represented among the tertiary interactions, supporting the 1993 claim that this base pairing is sufficient to determine the L-shaped tRNAs.

5.1.2 Ribonuclease P RNA (1996–1999)

Nolan in Pace's laboratory modeled the L14-P8-L18 (GNRA–helix–GNRA) junction in the RNase P RNA. The base pairing geometry linking the three fragments was inferred from aligning and comparing sequences of a large naturally occurring microbial population of RNase P RNA–encoding genes. Easterwood and Harvey extended the RNase P RNA modeling. In *E. coli*, they proposed structures of the 15/16 bulge, and in *Bacillus subtilis*, they built models of the binding between RNase P RNA P15 and a pre-tRNA. Schmitt applied the results of comparative sequence analysis and chemical modifications to generate 3-D structures of the RNase mitochondrial RNA processing (MRP). The modeled MRP RNA structure reveals two main structural domains, one highly conserved among all MRPs and RNase P RNAs, and the other only present in MRPs.

5.1.3 Hairpin Ribozyme (1999)

The hairpin ribozyme is a small catalytic RNA, which acts as an endonuclease that catalyzes a reversible sequence-specific cleavage reaction within a substrate RNA. Lambert in Major's laboratory initiated the modeling of the hairpin from a previously *MANIP* built 3-D model provided by Westhof and coworkers, as well as preliminary structural data provided by the laboratory of Burke. Preliminary structures generated by *MC-Sym* revealed the formation of a possible base triple (a single base simultaneously paired to two other bases) involving a nucleotide in the substrate that was considered to be immutable. The presence of the triple was verified and explained the immutability of the substrate base. In order to change the base in the substrate and to preserve the triple formation that was necessary to the activity, the two other nucleotides would have to be changed simultaneously to preserve the isostericity of the triple. Active hairpins were made by mutagenesis experiments that included isosteric triples. Other mutants that contained nonisosteric combinations were not active, confirming the hypothesis. Further theoretical and experimental modeling predicted the D-shaped conformation of the hairpin's substrate, as well as the complete hairpin structure and its precise chemical reaction mechanism. Recently, the same research team has determined a new active minimal hairpin by modifications and deletions of key helical elements within the two independent folding domains using cleavage activity experimental monitoring and *MC-Sym* modeling.

5.1.4 Leadzyme (1998–2001)

The leadzyme is an RNA molecule that is issued from *in vitro* selection of autocatalytic cleavage with lead ions using a library of yeast tRNA-Phe fragments. The leadzyme contains an asymmetric interior loop composed of six nucleotides: $C_1C_2A_3G_4$ on the 5′ side and G_7A_8 on 3′ side (see Fig. 13a). The leadzyme autocleaves in the presence of lead ions. Starting with the secondary structure and

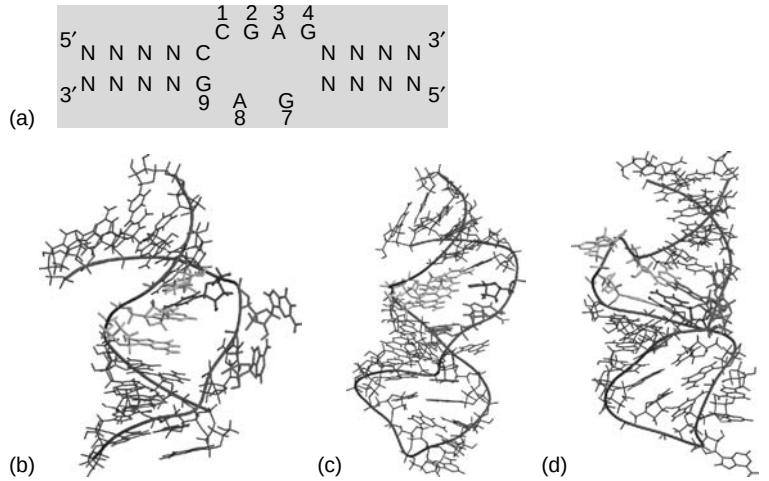

Fig. 13 Leadzyme. (a) Secondary structure. The nucleotides forming the catalytic site are numbered and colored. The scissile phosphate is the one of C_1 (red). G_2 and A_3 are not involved in base pairing (green). G_4 participates in the triple formation (black). The three nucleotides on the 3′ side of the catalytic core are colored yellow. (b) Crystal structure. (c) NMR structure. (d) Model. The colors defined in (a) apply to (b), (c), and (d). All atoms but H are shown. The phosphodiester chains are shown by a line tracing the P atoms. (See color plate p. xl.)

eight leadzyme variants (slightly mutated sequences), Lemieux and coworkers had first proposed an active conformation of the leadzyme that included C_1–A_8 and G_4–G_7 bps. They applied a modeling protocol based on the intersection of conformational space of functionally related RNA sequences. The protocol computes the intersection of all consistent 3-D structures of the active sequences, and then works by elimination under the assumption that all leadzyme sequences should have the same active conformation.

Chemical modification assays addressing the specific atomic groups involved in the base interactions predicted in the model were performed, and they confirmed base pairing involvement of three out of the four bases. None of the chemical groups on A_8 were found to be important to the catalytic activity of the leadzyme, whereas the chemical groups of C_1 normally involved in Watson–Crick interactions were found to be important. Consequently, a new Watson–Crick partner was needed for C_1. Interestingly, the predicted G_4–G_7 bp left the Watson–Crick groups of G_7 exposed to the interior loop, ready to form partnership with C_1. The base triple C_1–G_7–G_4 was thus proposed to be formed in a step of the cleavage activity, as its structure allows all atoms thought to be involved in the catalytic reaction in the right configuration for an SN2 catalytic reaction mechanism (see Fig. 14). David and coworkers then compared Lemieux and coworkers' 3-D model of the catalytic leadzyme with the X-ray crystal structure of Wedekind and McKay, and the NMR structure of Hoogstraten and coworkers (see Fig. 13). In the model, all catalytic core nucleotides are C3′-*endo anti*, except for the C_1 C2′-*endo anti* conformation. The crystal structure agrees on the

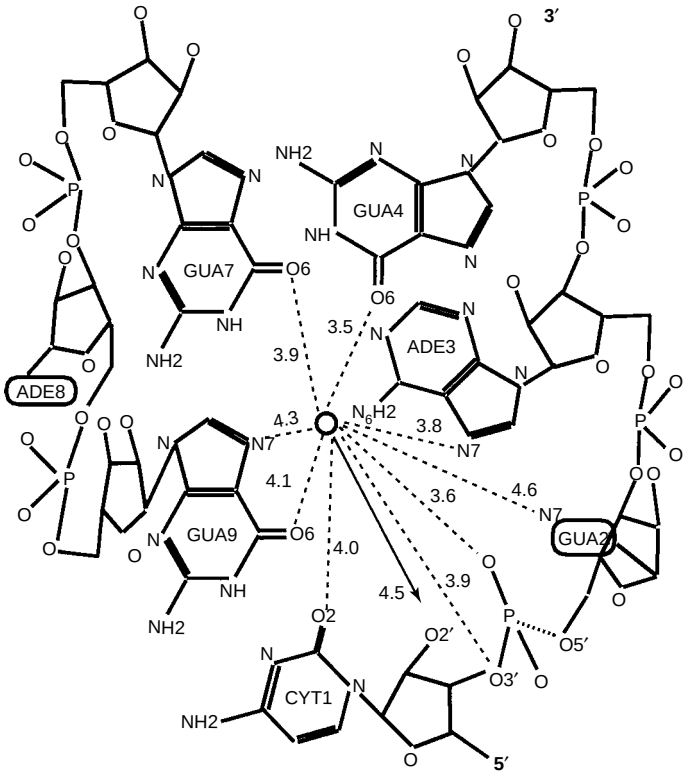

Fig. 14 Leadzyme catalytic core. The circle in the middle indicates the predicted position of the lead ion. The arrow points to the attacked 2′OH. The dotted lines indicate lead ion coordinations.

nucleotide conformations with the model, except for the G_4 C1′-*exo syn* conformation. The NMR structure disagrees with the crystal and model for all nuleotides on the 5′ side of the catalytic core: C_1 C3′-*endo anti*, G_2 C2′-*endo syn*, A_3 C2′-*exo anti*, and G_4 C1′-*exo anti*. The three structures agree on the stacking of A_3/G_9, as well as on the nonstacking between G_7 and A_8. The main difference in the base pairing between the model and the experimental structures is the base triple C_1–G_7–G_4, supported by chemical modification data. The experimental structures pair C_1 to A_8, even though chemical modification data did not identify any important chemical group for the catalytic activity on A_8, and did not show any base pairing between G_4 and G_7, in which chemical groups were shown to be important for the catalytic activity.

The model and crystal structures were found to be more stable to maintain their 3-D structures, as determined by molecular dynamics. A fitness function to evaluate SN2 propensity can be defined with the angle formed between O2′–P–O5′ and the distance between the O2′ and P atoms. An angle of 180° and a distance of 3 Å represent the structure with an optimal rate of transesterification. Again, the crystal structure and the

model maintain the SN2 configuration during molecular dynamics, although the theoretical model fit slightly better than the SN2 conformation.

5.1.5 *Escherichia Coli* 16S rRNA Fragments (2001)

Starting from the secondary structure, Dolan and coworkers modeled 58 different fragments of *E. coli* 16S rRNA using *MC-Sym*. They then added structural information inferred by chemical modification and covariation analysis. They assumed and favored the A-form helix conformation in nonhelical regions, and where no base pairing was inferred. The *MC-Sym* generated structures were refined by energy minimization using the *Insight II* suite of molecular modeling programs. The resulting fragment structures confirmed several structural inferences that were previously made in other rRNA studies, such as the presence of a central pseudoknot joining the three major domains of the ribosome's small subunit.

5.1.6 Prohead RNA (1998–2002)

A 174-nucleotide RNA has been shown to play an essential role in the DNA packaging of bacteriophage Φ29 in *B. subtilis*. From these 174 nucleotides, only a subsequence of 120 is necessary for efficient packaging. A preliminary secondary structure was proposed from comparative sequence analysis of sequences related to *B. subtilis* phages, and confirmed by directed mutagenesis experiments. Nucleotides 22 to 84 play an important role in prohead binding, as was determined by ribonuclease footprinting. The subunits were experimentally observed by electrophoresis gels and were shown to exist in hexamers. In order to further understand the specific packaging mechanism of this bacteriophage, *MC-Sym* was used to propose 3-D structures of this RNA.

The hexamer structure was determined to be circular, and composed of six copies of the same monomer structure. The monomers interact in the hexamer conformation by kissing nucleotides, as determined by comparative sequence analysis. By defining a transformation matrix, M, from an overlapping nucleotide in two consecutive monomers, Major and coworkers were able to select possible monomer structures that allowed for the formation of a cyclic hexamer, those satisfying the constraint $M^6 \approx I$. Other structure determination attempts include the application of electron microscopy of the hexamer, of X-ray crystallography of crystals adopting pentamer conformations, and of computer modeling of the dimer.

5.1.7 Red Clover Necrotic Mosaic Virus Bimolecular Complex (2004)

Guenther and coworkers at the North Carolina State University have studied the RNA–RNA structural interaction between the Red clover necrotic mosaic virus and the capsid protein subgenomic protomer that enables subgenomic RNA synthesis transactivation. This interaction was predicted by mutagenesis, but its precise 3-D structure was still unknown. Starting from NMR data of an *in vitro* RNA oligomer, they applied *MC-Sym* to solve the NMR constraints and to build 3-D models of the RNA–RNA interaction. They found a G–U wobble bp needs to be broken in the viral RNA structure to stabilize base pairing with the binding RNA.

5.2 *MANIP*

5.2.1 Group II Intron (2000)

Group II introns are known to recognize their 5′ exon by binding two exon

sequences (ESB1 and ESB2) by 6 bps respectively, and with two regions at the 3′ end of the 5′ exon (ISB1 and ISB2). Costa, Michel, and Westhof have modeled the complex between the exons, their binding sites, and the intron's domain V using structural data obtained from chemical footprinting, comparative sequence analysis, and the application of the *MANIP* modeling program. It was shown that reverse splicing of the group II intron is largely facilitated in the presence of a base pairing between the nucleotide 5′ of the ESB1 sequence (δ) and the 3′ exon (δ'). Applying computer modeling, Costa and her colleagues have proposed an alternative interaction pattern, where δ is base paired with the 5′ side of an interior loop within domain I, while the former δ', renamed IBS3 for clarity, was suggested to base pair with the 3′ side of the same interior loop (ESB3). Kinetic analysis showed that the new δ–δ' interaction is still needed to facilitate binding of the 5′ exon with the intron, while the ISB3–ESB3 interaction ensures an efficient exon ligation.

5.2.2 Hammerhead Ribozyme (2003)

The hammerhead ribozyme is a small RNA whose structure was extensively studied, mainly because of potential therapeutic properties. The hammerhead has a conserved catalytic core structured by the three-way junction of three helices, which confers its cleavage activity in high Mg^{2+} concentrations. Using site-directed mutagenesis data, Khvorova and coworkers identified additional conserved sequence elements outside the core that were necessary to initiate cleavage in low Mg^{2+} concentration. Different hammerhead sequences were in alignment showing that these critical nucleotides were present in the hammerhead's hairpin loops I and II, capping two of the core helices. Using this information and a crystal structure of the hammerhead, a 3-D model was built with *MANIP*.

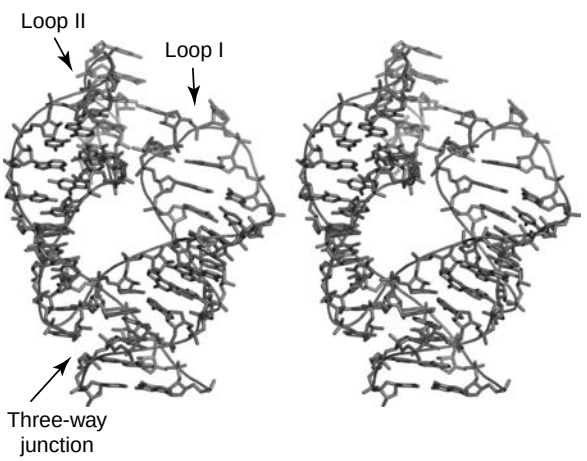

Fig. 15 Stereo view of a 3-D model of the hammerhead ribozyme. The model was generated by using *MANIP*. All atoms but H are shown. The phosphodiester chain is shown by a line tracing the P atoms. The modeled interaction between loop I and II can be seen. (Courtesy E. Westhof.)

The model showed H/H and S/S bps between the conserved nucleotides in loops I and II (see Fig. 15). They concluded that these noncanonical loop–loop interactions were necessary to activate the cleavage mechanism in low Mg^{2+} concentrations.

6 Perspectives

In the last 10 years, we have seen more modeling than ever, produced mainly from two RNA 3-D structure computer modeling systems: *MC-Sym* and *MANIP*. With *MC-Sym*, we are now close to a generation of computer programs that produce 3-D structures from a simple description of RNA structural data. In the near future, it is expected that RNA structure specialists will agree on more formalism to describe and unify all RNA abstractions unambiguously. Such formalisms will come from the result of multidisciplinary efforts involving computer scientists, biophysicists, and molecular biologists. As our knowledge of RNA 3-D structure increases, our ability to develop computer modeling systems to address the RNA theoretical flexibility will become better. After all, the success obtained with *MANIP* strongly suggests the existence of a precise and accurate formalism, in which automated modeling should produce high-precision models in efficient ways.

See also Protein Modeling; Ribozymes; RNA Secondary Structures.

Bibliography

Books and Reviews

Cedergren, R., Major, F. (1996) Modeling RNA Tertiary Structure, in: Simons, R., Simons, W.R., Grunberg-Manago, M. (Eds.) *RNA Structure and Function*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 37–76.

Paul, R.P. (1981) *Robot Manipulators: Mathematics, Programming and Control*, MIT Press, Cambridge, MA.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

Van Hentenrick, P. (1989) *Constraint Satisfaction in Logic Programming*, MIT Press, Cambridge, MA.

Westhof, E., Auffinger, P. (2000) RNA Tertiary Structure, in: Meyers, R.A. (Ed.) *Encyclopedia of Analytical Chemistry*, John Wiley & Sons, Chichester, pp. 5222–5232.

Primary Literature

Altona, C., Sundaralingam, M. (1972) Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation, *J. Am. Chem. Soc.* **94**, 8205–8212.

Auffinger, P., Vaiana, A.C. (2005) Molecular Dynamics Simulations of RNA Systems, in: Hartmann R.K., et al. (Eds.) *Handbook of RNA Biochemistry*, Wiley-VCH, pp. 560–575.

Barrette, I., Poisson, G., Gendron, P., Major, F. (2001) Pseudoknots in prion protein mRNAs confirmed by comparative sequence analysis and pattern searching, *Nucleic Acids Res.* **29**, 753–758.

Batey, R.T., Rambo, R.P., Doudna, J.A. (1999) Tertiary motifs in RNA structure and folding, *Angew. Chem. Int. Ed. Engl.* **38**, 2326–2343.

Bergman, N.H., Lau, N.C., Lehnert, V., Westhof, E., Bartel, D.P. (2004) The three-dimensional architecture of the class I ligase ribozyme, *RNA* **10**, 176–184.

Bourassa, N., Major, F. (2002) Implication of the prohead RNA in phage Φ29 DNA packaging, *Biochimie* **84**, 945–951.

Brown, J.W., Nolan, J.M., Haas, E.S., Rubio, M.A.T., Major, F., Pace, N.R. (1996) Comparative analysis of ribonuclease P RNA using gene

sequences from natural microbial populations reveals tertiary structural elements, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 3001–3006.

Cheatham, T.E. III, Young, M.A. (2001) Molecular dynamics simulation of nucleic acids: successes, limitations, and promise, *Biopolymers* **56**, 232–256.

Costa, M., Michel, F., Westhof, E. (2000) A three-dimensional perspective on exon binding by a group II self-splicing intron, *EMBO J.* **19**, 5007–5018.

David, L., Lambert, D., Gendron, P., Major, F. (2001) Leadzyme, in: Nicholson, A.W. (Ed.) *Ribonucleases, Methods in Enzymology*, Vol. 341, Academic Press, New York, pp. 518–540.

Dolan, M.A., Babin, P., Wollenzien, P. (2001) Construction and analysis of base-paired regions of the 16S rRNA in the 30S ribosomal subunit determined by constraint satisfaction molecular modeling, *J. Mol. Graph. Model.* **19**, 495–513.

Duarte, C.M., Wadley, L.M., Pyle, A.M. (2003) RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space, *Nucleic Acids Res.* **31**, 4755–4761.

Easterwood, T.R., Harvey, S.C. (1997) Ribonuclease P RNA: models of the 15/16 bulge from *Escherichia coli* and the P15 stem loop of *Bacillus subtilis*, *RNA* **3**, 577–585.

Gabb, H.A., Sanghani, S.R., Robert, C.H., Prévost, C. (1996) Finding and visualizing nucleic acid base stacking, *J. Mol. Graph.* **14**, 6–11.

Gendron, P., Lemieux, S., Major, F. (2001) Quantitative analysis of nucleic acid three-dimensional structures, *J. Mol. Biol.* **308**, 919–936.

Guenther, R.H., Sit, T.L., Gracz, H.S., Dolan, M.A., Townsend, H.L., Liu, G., Newman, W.H., Agris, P.F., Lommel, S.A. (2004) Structural characterization of an intermolecular RNA-RNA interaction involved in the transcription regulation of a bipartite plant virus, *Nucleic Acids Res.* **32**, 2819–2828.

Harvey, S.C., Wang, C., Teletchea, S., Lavery, R. (2002) Motifs in nucleic acids: molecular mechanics restraints for base pairing and base stacking, *J. Comput. Chem.* **24**, 1–9.

Hermann, T., Patel, D.J. (1999) Stitching together RNA tertiary architectures, *J. Mol. Biol.* **294**, 829–849.

Hershkovitz, E., Tannenbaum, E., Howerton, S.B., Sheth, A., Tannenbaum, A., Williams, L.D. (2003) Automated identification of RNA conformational motifs: theory and application to the HM LSU 23S rRNA, *Nucleic Acids Res.* **31**, 6249–6257.

Hoffmann, B., Mitchell, G.T., Gendron, P., Major, F., Andersen, A.A., Collins, R.A., Legault, P. (2003) NMR structure of the active conformation of the Varkud satellite ribozyme cleavage site, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7003–7008.

Khvorova, A., Lescoute, A., Westhof, E., Jayasena, S.D. (2003) Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity, *Nat. Struct. Biol.* **10**, 708–712.

Lavery, R., Zakrzewska, K., Sklenar, H. (1995) JUMNA (junction minimization of nucleic acids), *Comput. Phys. Commun.* **91**, 135–158.

Lee, J.C., Gutell, R.R. (2004) Diversity of base-pair conformations and their occurrence in rRNA structure and RNA structural motifs, *J. Mol. Biol.* **344**, 1225–1249.

Legault, P., Hoogstraten, C.G., Metlitzky, E., Pardi, A. (1998) Order, dynamics and metal-binding in the lead-dependent ribozyme, *J. Mol. Biol.* **284**, 325–335.

Lemieux, S., Major, F. (2002) RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire, *Nucleic Acids Res.* **30**, 4250–4263.

Lemieux, S., Oldziej, S., Major, F. (1998) Nucleic Acids: Qualitative Modeling, in: Schleyer, P.V.R., Allinger, N.L., Clark, T., Gasteiger, J., Kollman, P.A., Schaefer, III, H.F., Schreiner, P.R. (Eds.) *Encyclopedia of Computational Chemistry*, John Wiley & Sons, West Sussex, England.

Lemieux, S., Chartrand, P., Cedergren, R., Major, F. (1998) Modeling active RNA structures using the intersection of conformational space: application to the lead-activated ribozyme, *RNA* **4**, 739–749.

Leontis, N.B., Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs, *RNA* **7**, 499–512.

Leontis, N.B., Westhof, E. (2002) The annotation of RNA motifs, *Comp. Funct. Genomics* **3**, 518–524.

Leontis, N.B., Westhof, E. (2003) Analysis of RNA motifs, *Curr. Opin. Struct. Biol.* **13**, 300–308.

Leontis, N.B., Stombaugh, J., Westhof, E. (2002a) The non-Watson-Crick base pairs and their associated isostericity matrices, *Nucleic Acids Res.* **30**, 3479–3531.

Leontis, N.B., Stombaugh, J., Westhof, E. (2002b) Motif prediction in ribosomal RNAs. Lessons and prospects for automated motif prediction in homologous RNA molecules, *Biochimie* **84**, 961–973.

Major, F. (2003) Building three-dimensional ribonucleic acid structures, *IEEE Comput. Sci. Eng.* **sept.–oct.**, 44–53.

Major, F., Griffey, R. (2001) Computational methods for RNA structure determination, *Curr. Opin. Struct. Biol.* **11**, 282–286.

Major, F., Gautheret, D., Cedergren, R. (1993) Reproducing the three-dimensional structure of a tRNA molecule from structural constraints, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9408–9412.

Major, F., Turcotte, M., Gautheret, D., Lapalme, G., Fillion, E., Cedergren, R. (1991) The combination of symbolic and numerical computation for three-dimensional modeling of RNA, *Science* **253**, 1255–1260.

Malhotra, A., Gabb, H.A., Harvey, S.C. (1993) Modeling large nucleic acids, *Curr. Opin. Struct. Biol.* **3**, 241–246.

Massire, C., Westhof, E. (1998) MANIP: an interactive tool for modeling RNA, *J. Mol. Graph. Model.* **16**, 197–205.

Mueller, F., Brimacombe, R. (1997) A new model for the three-dimensional folding of *Escherichia coli* 16S ribosomal RNA. I. Fitting the RNA to a 3D electron microscopic map at 20 A., *J. Mol. Biol.* **271**, 524–544.

Murray, L.J.W., Arendall, W.B. III, Richardson, D.C., Richardson, J.S. (2003) RNA backbone is rotameric, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13904–13909.

Olson, W.K. (1980) Configurational statistics of polynucleotide chains. An updated virtual bond model to treat effects of base stacking, *Macromolecules* **13**, 721–728.

Olson, W.K., Bansal, M., Burley, S.K., Dickerson, R.E., Gerstein, M., Harvey, S.C., Heinemann, U., Lu, X.J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C.S., Westhof, E., Wolberger, C., Berman, H.M. (2001) A standard reference frame for the description of nucleic acid base-pair geometry, *J. Mol. Biol.* **313**, 229–237.

Pinard, R., Lambert, D., Pothiawala, G., Major, F., Burke, J.M. (2004) Modifications and deletions of helices within the hairpin ribozyme-substrate complex: an active ribozyme lacking helix 1, *RNA* **10**, 395–402.

Pinard, R., Lambert, D., Heckman, J.E., Esteban, J.A., Gundlach, B., Click, G., Walter, N.G., Major, F., Burke, J.M. (2001) The hairpin ribozyme substrate binding domain: a highly constrained D-shaped conformation, *J. Mol. Biol.* **307**, 51–65.

Rangan, P., Masquida, B., Westhof, E., Woodson, S.A. (2004) Architecture and folding mechanism of the *Azoarcus* group I pre-tRNA, *J. Mol. Biol.* **339**, 41–51.

Schmitt, M.E. (1999) Molecular modeling of the three-dimensional architecture of the RNA component of yeast RNase MRP, *J. Mol. Biol.* **292**, 827–836.

Schneider, B., Morávek, Z., Berman, H.M. (2004) RNA conformational classes, *Nucleic Acids Res.* **32**, 1666–1677.

Tamura, M., Hendrix, D.K., Klosterman, P.S., Schimmelman, N.R.B., Brenner, S.E., Holbrook, S.R. (2004) SCOR: structural classification of RNA, version 2.0, *Nucleic Acids Res.* **32**, D182–D184.

Wadley, L.M., Pyle, A.M. (2004) The identification of novel RNA structural motifs using COMPADRES: an automated approach to structural discovery, *Nucleic Acids Res.* **32**, 6650–6659.

Walberer, B.J., Cheng, A.C., Frankel, A.D. (2003) Structural diversity and isomorphism of hydrogen-bonded base interactions in nucleic acids, *J. Mol. Biol.* **327**, 767–780.

Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., Case, D.A. (2004) Development and testing of a general Amber force field, *J. Comput. Chem.* **25**, 1157–1174.

Wedekind, J.E., McKay, D.B. (1999) Crystal structure of a lead-dependent ribozyme revealing metal binding sites relevant to catalysis, *Nat. Struct. Biol.* **6**, 261–268.

Yang, H., Jossinet, F., Leontis, N.B., Chen, L., Westbrook, J., Berman, H., Westhof, E. (2003) Tools for the automatic identification and classification of RNA base pairs, *Nucleic Acids Res.* **31**, 3450–3460.

Yu, E.T., Zhang, Q., Fabris, D. (2004) Untying the FIV frameshifting pseudoknot structure by MS3D, *J. Mol. Biol.* **345**, 69–80.

Sense of Smell

Robert P. Lane[1], *Gregory S. Smutzer*[2,3], *and Richard L. Doty*[3]
[1] *Wesleyan University, Middletown, CT, USA*
[2] *Temple University, Philadelphia, PA, USA*
[3] *University of Pennsylvania Medical Center, Philadelphia, PA, USA*

1 Introduction 640

2 Basic Structure of the Main Olfactory System 642
2.1 Olfactory Neuroepithelium 642
2.1.1 Cell Types 642
2.1.2 Receptor Cell Cilia 645
2.1.3 Neurogenesis in the Neuroepithelium 645
2.2 Olfactory Bulb 648
2.2.1 Olfactory Glomeruli 648
2.2.2 Centrifugal Inputs into the Bulb 650
2.2.3 Olfactory Bulb Tuning of Afferent Signals 651
2.2.4 Neurogenesis in the Bulb 653
2.3 Olfactory Cortex 654

3 Basic Structure of the Accessory Olfactory System 656

4 Early Concepts of Olfactory Signal Transduction 658

5 The Modern Era of Olfactory Signal Transduction 658
5.1 Odorant Binding Proteins 660
5.2 Olfactory Receptors 661
5.2.1 Identification of Ligand/Receptor Complexes 665
5.3 Second Messengers 666
5.3.1 Adenosine 3′,5′-cyclic Phosphate (Cyclic AMP or cAMP) 666
5.3.2 Guanosine 3′,5′-cyclic Monophosphate (cyclic GMP or cGMP) 669
5.3.3 Inositol-1,4,5-trisphosphate (IP_3) 670
5.4 Signal Termination and Desensitization 672

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

6 Olfactory System Receptor Gene Repertoires in Vertebrates 673
6.1 Main Olfactory System 673
6.2 Accessory Olfactory System 676

7 Basic Principles of Olfactory System Organization 677
7.1 Monogenic Expression of Olfactory Receptor Genes 677
7.2 Convergence of Olfactory Receptor-expressing Neurons in the Olfactory Bulb 680

8 The Combinatorial System of Odor Representation 683

9 Transcriptional Regulation of Odor Receptor Genes 683
9.1 Singular Olfactory Receptor Gene Activation 684
9.2 Olfactory Receptor Regulatory Elements 687

10 Human Diseases Associated with Genetic Defects and Smell Dysfunction 690
10.1 Congenital Disorders Associated with Olfactory Dysfunction 690
10.1.1 Bardet–Biedl Syndrome 691
10.1.2 Down Syndrome 691
10.1.3 Friedreich Ataxia 691
10.1.4 Kallmann Syndrome 692
10.1.5 Pseudohypoparathyroidism and Pseudopseudohypoparathyroidism 692
10.2 Neurodegenerative Diseases 692
10.2.1 Alzheimer's Disease 693
10.2.2 Huntington's Disease 693
10.2.3 Multiple Sclerosis 693
10.2.4 Parkinson's Disease 694

11 Summary 695

Acknowledgment 695

Bibliography 695
Books and Reviews 695
Primary Literature 696

Keywords

Anosmia
Inability to detect environmental odors. Smell loss can reflect genetic or environmental factors, and can be transient or permanent.

Gene Repertoire
The members of a gene family or gene superfamily found in the genome of an organism that produce functional messenger RNA transcripts and functionally active gene products.

Odor Coding
The process by which the main olfactory system of vertebrates recognizes and differentiates among volatile chemicals to produce the perception of odor quality.

Olfaction
The sensory system responsible for smell and other aspects of flavor sensation, whose receptors are located within the superior or dorsal sectors of the nasal cavity.

Receptor
A specialized molecule on the surface or within the cytoplasm of a cell that serves as a recognition or ligand binding site for sensory stimuli, protein signals, neurotransmitters, steroids, antigens, antibodies, or other molecules found in the extracellular environment. After ligand binding, these activated receptors can either stimulate or inhibit a biochemical change in cellular activity.

Second Messenger
A short-lived chemical signal that is formed in the cytosol, relays and amplifies an externally derived message from a cell's surface to its interior, and subsequently stimulates or inhibits a biochemical change in cellular activity. Calcium ions, nitric oxide, cyclic AMP, cyclic GMP, and IP_3 can all function as second messengers in cells.

Smell
The psychological percept caused by activation of the olfactory system.

Olfactory receptor genes devoted to the sense of smell compose the largest multigene family in mammalian genomes. Although the number of receptor genes and related olfactory-specific genes vary in different species, up to 1% of the mammalian genome may be devoted exclusively to the sense of smell. This number reflects the importance of olfaction throughout vertebrate evolution, and the role of olfaction in maintaining a wide range of life-supporting activities, including feeding, nutrition, enhancement of gustatory responses, safety, predator/prey detection, and esthetics. In humans, ~400 different genes encode functional olfactory receptor proteins that localize to the ciliary membranes of ~6 000 000 receptor cells in the olfactory neuroepithelium only the visual system of vertebrates prossesses with more receptor cells.

In this chapter, we review basic aspects of olfactory system anatomy, expression of odorant receptors within sensory neurons, and pathways of mammalian and invertebrate olfactory transduction. We focus primarily on the main (i.e. nonvomeronasal organ) olfactory system of vertebrates. In particular, specific

attention is paid to the (1) chromosomal organization of olfactory receptor genes; (2) monogenic expression of olfactory receptor genes; (3) regulation of odorant receptor gene transcription within sensory neurons; (4) role of olfactory receptor gene expression in regulating the projection of functionally distinct olfactory sensory neurons from the olfactory epithelium to the olfactory bulb; (5) initial biochemical transduction events that include the molecular mechanisms responsible for stimulus transduction and generation of second messengers for signal amplification; and (6) microanatomy and function of olfactory bulb circuits. The plasticity of the olfactory system, as reflected by neurogenesis both within the olfactory neuroepithelium and the olfactory bulb is discussed in detail, as are a number of inherited and idiopathic human disorders associated with olfactory dysfunction in humans.

1 Introduction

Most organisms, whether they live in water, air, or in or on the ground, depend upon environmental chemicals for survival. For this reason, chemosensory systems have evolved in all animal phyla to detect such chemicals and to establish their meaning to the organism. Within the chordates – in particular the vertebrates – such systems have achieved great complexity, culminating in chemosensory systems within not only the general integument (e.g. free nerve endings within the skin that respond to chemical irritants), but numerous specialized receptor regions as well (e.g. taste buds). In mammals – the major focus of this chapter – the predominant intranasal chemosensory system is that of olfaction. Nearly all mammals, with the exception of whales and some other cetaceans, such as dolphins and porpoises, have a functional main olfactory system whose receptor cells are located in the nasal vault. The axons of these cells project through holes in the bony plate separating the nasal and brain cavities (termed the *foramina* of the cribriform plate of the ethmoid bone) to targets within the olfactory bulb, a complex ovoid structure located on the ventral surface of each frontal lobe on the dorsal side of the cribriform plate. These axons, which number in humans between 6 and 10 million, collectively comprise Cranial Nerve I (CN I).

A number of mammals also possess an "accessory" olfactory system whose neural projections are quite separate from those of the main olfactory system. The receptors of this system are found within a tube-like structure, termed the *vomeronasal organ*, located near the base of the nasal cavity. Openings into this organ occur, depending upon the species, either within the nasal cavity or through the vomer bone in the roof of the mouth. The receptor cells within this structure project axons to an "accessory olfactory bulb" (AOB) located behind the normal olfactory bulb. In rodents and ungulates, this system has been closely associated with reproductive processes. In a number of other vertebrates, including snakes, this system is involved in food finding and in both predator and prey identification. While a rudimentary vomeronasal sac is found in humans and some other Old World primates, these species have no AOB, and the vomeronasal organ has no neural connections to the brain. A diagram of the main and accessory

olfactory systems of the mouse is presented in Fig. 1.

The importance of olfaction to our own species is underappreciated, despite the fact that decreased smell function adversely influences the flavor of foods and beverages, places individuals at risk from hazards such as leaking natural gas, fire, and spoiled food, and significantly alters quality of life. Among 750 patients presenting to the University of Pennsylvania Smell and Taste Center with largely olfactory problems, more than 68% experienced altered quality of life, 46% described changes in appetite or body weight, and 56% reported influences in daily living or psychological well-being. Loss of smell, particularly when it occurs in adulthood, can result in significant psychological disruption and, in some cases, may even generate feelings of physical and social vulnerability and victimization. The

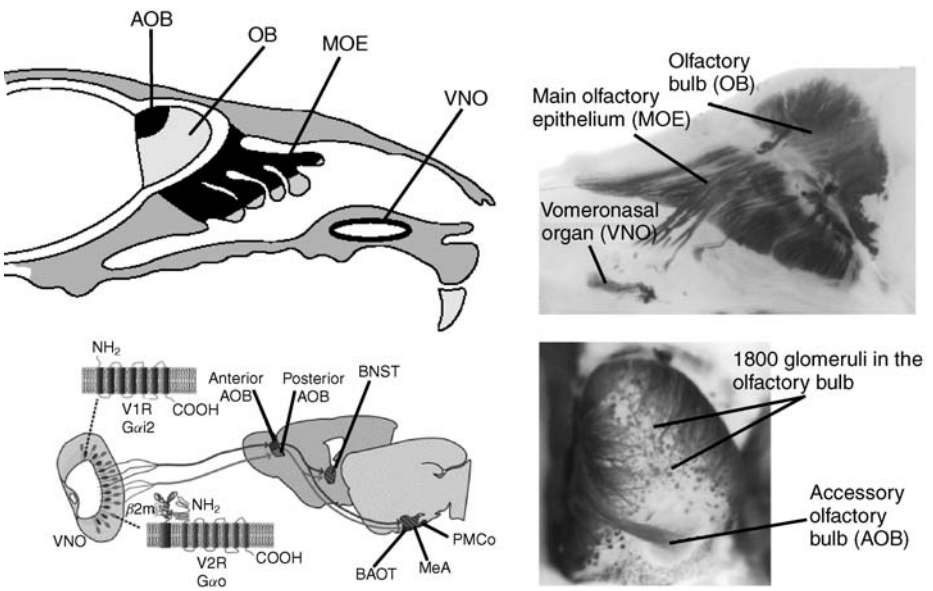

Fig. 1 Rodent olfactory system. Upper left: Schematic drawing depicting the positioning of the olfactory cavity; the main olfactory epithelium (MOE), which contains the olfactory sensory neurons; the olfactory bulb (OB) to which the olfactory sensory neurons project their axons; the vomeronasal organ (VNO), which is a second chemosensory organ located above the palate in the mouth; and the accessory olfactory bulb (AOB) to which the sensory neurons of the VNO project their axons. Reprinted from ''Seminars in cell and Developmental Biology, Vol. 8, C. Dulac'', Molecular biology of pheromone perception in mammals, pages 197–205 Copyright 1997, with permission from Elsevier. Right: Sagittal section (top) and dorsal view (bottom) of a mouse brain in which all mature olfactory sensory neurons are labeled with *Lac*Z receptor gene. The dorsal view shows the termination of olfactory sensory neurons within glomeruli. Figure from Mombaerts, P. (2001). How smell develops. *Nat. Neurosci Suppl.* **4**, 1192–8. Lower left: Drawing depicting the apical (V2R-expressing) and basal (V1R-expressing) compartments of the rodent vomeronasal organ, which projects sensory neuronal axons to the accessory olfactory bulb (AOB). Reprinted from Hormones and Behavior, Vol. 46. Breunan, ''The nose knows who's who: Chemosensory indiduality and mate recognition in mice'' pages 231–40. Copyright, with permission from Elsevier.

importance of this all-too-often underappreciated sensory system is reiterated by the consequences of its loss in those who depend on it for their livelihood (e.g. chefs, homemakers, firefighters, plumbers, fragrance and wine merchants, chemical plant workers, etc.).

In this chapter, we review advances in understanding olfactory function in mammals, emphasizing the signal transduction pathways, the evolution of odorant receptors, and olfactory coding at both the receptor and olfactory bulb level. Initially we summarize the basic anatomy and physiology of the mammalian olfactory system, with an emphasis on the olfactory epithelium, bulb, and cortex. Neurogenesis within the olfactory epithelium and bulb is then described. The next section examines early concepts of olfactory coding, followed by an elucidation of current views of mammalian odorant signal transduction. Included in this section are details of olfactory receptors, second messengers, and modes of terminating signal transduction. Subsequent sections describe olfactory receptor gene families, the organization and expression of mammalian olfactory genes, and the topographic organization of the olfactory pathways. The combinational system of odor representation and transcriptional regulation of olfactory receptor genes is then presented, followed by human clinical studies of gene-related olfactory disturbances.

2 Basic Structure of the Main Olfactory System

2.1 Olfactory Neuroepithelium

In humans, the receptors of the main olfactory system are located in the upper recesses of the nasal chambers within a layer of cells (termed the *olfactory neuroepithelium*) that lines the cribriform plate and sectors of the superior turbinate, middle turbinate, and septum Turbinates are shelf-like, cartilaginous structures that are covered by mucous membranes and protrude into the nasal airway to help warm, filter, and humidify inhaled air on its way to the lungs. In amphibia, the neuroepithelium is flat and planar, whereas in carnivores (e.g. dogs, cats), rodents, and many other mammals, it is distributed over a very convoluted surface dictated by the complexity of the turbinal folds that arise from the ethmoid bone. The number of ethmoidal turbinates differs among species, extending from one to three in primates to over a dozen in the spiny anteater (Echidna). A picture of the ethmoidal turbinates of the pig is shown in Fig. 2.

2.1.1 Cell Types

The olfactory neuroepithelium is comprised of a half dozen morphologically and biochemically distinct cell types, although additional classes of less well-defined microvilli-containing cells have been noted prenatally and postnatally (Fig. 3). The Microvilli are minute, brush-like, cylindrical processes on the cell surface that increase the surface size of the cell. The first cell type – the *bipolar sensory receptor neuron* – was first described by Schultze in the middle of the nineteenth century. This neuron contains a small (5–7 μm wide) cell body generally located within the lower two-thirds of the epithelium. Its single dendrite – that portion of the cell that is at the receiving end of information input – often takes a tortuous path within the mucosa around other dendrites, supporting cells, and adjacent receptor cell bodies

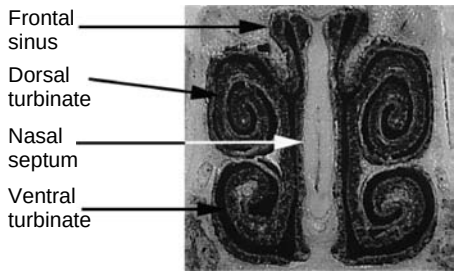

Fig. 2 Cross section of the nose of the adult pig. Note two sets of turbinates designated dorsal and ventral, the complex turbinal folds, and the thick nasal septum.

to reach the mocosal surface, where it extends, from a knob-like structure, odorant receptor-containing cilia into the mucus. Olfactory cilia contain the olfactory receptors and differ structurally from cilia in other segments of the respiratory tract, as described in detail later in this chapter. These cilia contain the olfactory receptors. A belt-like tight junctional complex attaches the dendrite to adjacent supporting cells at the apical surface, forming a transmembrane barrier characteristic of most epithelial cells. The receptor cell extends a single unmyelinated axon, which also takes a complex path to its target, in this case the olfactory bulb of the brain. The major neurotransmitter released from axon terminals of the olfactory receptor cells is glutamate.

The second cell type is the *supporting (sustentacular) cell*. These cells, which extend microvilli, rather than cilia, into the mucus layer, insulate the bipolar receptor cells from one another, and may help regulate the chemical composition of the mucus, secreting, for example, immunoglobulins into the mucus. Sustentacular cells also likely help to deactivate odorants, as well as to protect the epithelium from damage from xenobiotics. The supporting cells contain xenobiotic-metabolizing enzymes (e.g. cytochrome P-450s) that catabolize organic chemicals and other potentially damaging molecules that enter the nasal cavity. Such enzymes are also found within the acinar and duct cells of Bowman's glands, the major source of mucus in the olfactory epithelium.

The third cell type is the poorly understood *microvillar cell* located at the surface of the epithelium. Microvillar cells, which look similar to the so-called *brush cells* found throughout the upper and lower airways of many species, also have microvillae at their apical surfaces. While there has been suggestion that these cells extend axon-like processes to the olfactory bulb, convincing support for this notion is lacking. In the human, microvillar cells occur in about a 1 : 10 ratio with bipolar receptor cells. It seems unlikely that these cells play any direct role in chemosensory functioning, as *in vitro* patch clamping has failed to find any odorant-related responses.

The fourth cell type is the *cell that lines the Bowman's glands and ducts*, whereas the fifth and six cell types are the *globose (light) basal cell* and *horizontal (dark) basal cell* – cells located near the basement membrane from which most of the other cell types arise.

Recent data suggest that, under conditions of marked damage to the olfactory epithelium, the same type of basal cell, most likely the spherically shaped globose cell, has the potential for giving rise to neurons and nonneural cells, implying considerable multipotency. This population of stem cells is maintained in the

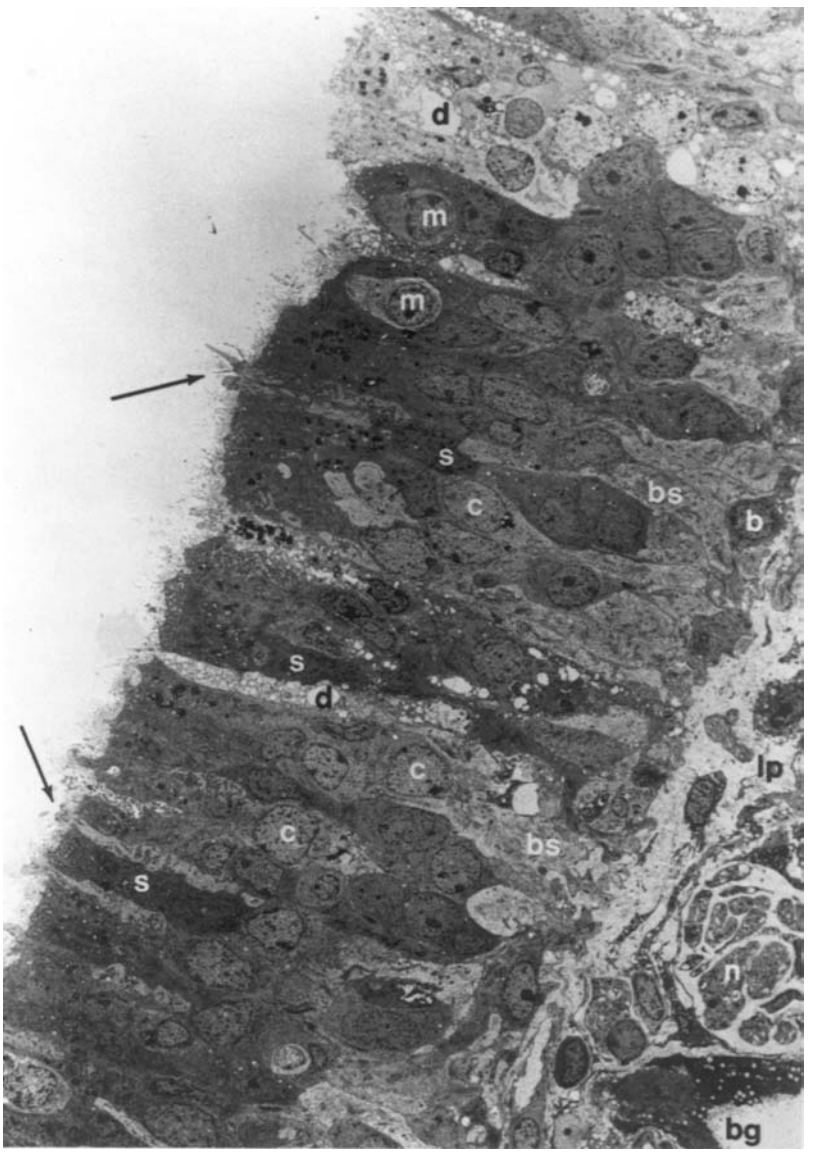

Fig. 3 Low power electron photomicrograph of cross section of the human olfactory neuroepithelium depicting the four major types of cells: bipolar receptor cell (arrows point to cilia at dendritic knob; c, cell body), microvillar cells (m), sustentacular cells (s), and basal cells (b); bg, Bowman's gland; lp, lamina propria; n, collection of axons within an ensheathing cell; d, degenerating cells; bs, basal cell undergoing mitosis. Photo courtesy of Dr. David Moran, Longmont, Colorado.

olfactory epithelium as these cells divide and give rise to new receptor neurons.

2.1.2 Receptor Cell Cilia

The bipolar receptor cells cilia differ from the cilia of the nonneuronal, respiratory epithelial cells, in being much longer and in lacking dynein arms required for intrinsic motility. Collectively, the surface area of the cilia is quite large, exceeding, for example, 7 m^2 in the German shepherd dog and 22 cm^2 in the human. This variation reflects not only the number of receptor cells but also the shape of the dendritic knob and the number and length of cilia that extend from it. The number of cilia per cell depends upon the species, with some species, such as the human, having from 10 to 30 cilia per cell, and others, such as the dog, having over a hundred per cell.

The olfactory cilia contain the cell surface receptors that interact with incoming odorants and have a $9 + 2$ microtubule fiber pattern similar to that found in other vertebrate cilia. Ultrastructural studies show that the cilia are devoid of endoplasmic reticulum (ER). Thus, little or no Ca^{2+} storage in internal membranes occurs in these cilia. This implies that, following odorant binding to receptors, depolarization of ciliary membranes occurs by cation influx from outside the cell.

2.1.3 Neurogenesis in the Neuroepithelium

Olfactory receptor cells, which are derived from ectoderm, the outermost layer of three primary germ layers of the embryo can regenerate to some degree after being damaged. This phenomenon, which was first observed in 1940, was in conflict with the dogma that neurons in the adult animal are irreplaceable, and suggested that the olfactory system may contain the key to producing neural regeneration in a variety of neural systems. In the 1960s, it was discovered that mitotic cells, young sensory cells, mature sensory cells, and dying cells coexist within the olfactory epithelium. This observation suggested that olfactory receptor cells were continually being replaced and were relatively short-lived. However, in the 1980s it was found that a number of olfactory receptor cells of rats reared in a pathogen-free environment survived for at least 12 months. This led to the concepts that (1) olfactory nerve cell turnover involves recently formed or immature receptor cells that fail to establish synaptic connections with the olfactory bulb; (2) environmental agents likely play an important role in dictating which elements of the receptor sheet become replaced; and (3) the rate of regeneration of olfactory receptor cells is not genetically predetermined. Today, questions still remain as to why metaplastic respiratory epithelium often replaces damaged regions of the olfactory epithelium and why, when such metaplasia occurs, the epithelium in that region may never convert back to olfactory epithelium. Nonetheless, studies employing toxic agents that damage the epithelium do find that the type of repair correlates with the degree or extent of the initial epithelial damage. Thus, when the basilar layer of the mucosa is completely damaged, replacement with a respiratory-like epithelium occurs. When the damage is not marked or the toxic insult is not sustained, regeneration, usually with fewer or irregularly arranged cells, occurs. The observation that improvement in olfactory function after cessation of chronic cigarette smoking occurs over a period of years and is dose-related suggests that either (1) turnover of the olfactory epithelial cell complement takes a much longer time

than previously supposed or (2) growth of olfactory epithelium into damaged areas is relatively slow and dependent on the extent of prior trauma, or both.

Numerous studies, some performed on olfactory epithelia grown in culture, have shown that homeostatic mechanisms regulate the development of olfactory receptor neurons so that the appropriate number of cells at each developmental stage is continuously maintained in the sensory epithelium. Such mechanisms include autocrine and paracrine signals, similar to those that function during embryonic development (When a factor, such as a hormone, is produced by the same cell that receives it, it is termed *an autocrine factor*. When such a factor is produced by a cell different than the one that receives it, it is termed *a paracrine factor*.). In addition, sensory neurons are trophically dependent on the olfactory bulb for their prolonged survival. Thus, when olfactory bulbs are lesioned in rodents, most olfactory neurons perish within a few weeks after lesioning. Such observations suggest that olfactory receptor neurons require trophic factors from the bulb at a critical stage in their development, and will not survive to maturity in the absence of such factors.

Growth factors, small proteins that bind to receptors on the surfaces of cells, are necessary for cell proliferation. A number of growth factors localize to the mammalian olfactory epithelium (Table 1). Among those implicated in olfactory epithelial neurogenesis are several members of the neurotrophin family and their cognate tropomyosin-related tyrosine kinase (Trk) receptors. Examples include brain-derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3), neurotrophin-4/5 (NT-4/5), and high-affinity nerve growth factor (NGF). NGF is a 130-kDa protein that is composed of five noncovalently linked subunits, and signals through TrkA receptors. This protein promotes and regulates the survival of neurons in the developing peripheral nervous system, and is produced by peripheral tissue for guidance and development of outgrowing sensory neurons. NGF is critical for both the survival and differentiation of sensory neurons, possibly by activation of the transcription factor called nuclear factor kappaB (NF-κB) or activation of receptor protein tyrosine kinases and activation of phosphatidylinositol 3-kinase (PI 3-kinase) (Nuclear factor-kappa B (NF-κB) transcription factors are required for regulating cell survival and differentiation, and are critical for the development and maintenance of the nervous system.) NGF also stimulates neurite outgrowth, and presumably stimulates axon formation in olfactory receptor neurons. Other receptor tyrosine kinases and their ligands may also participate in the regeneration of olfactory neurons within the adult rat olfactory mucosa. For example, NT-3 and NT-4/5 localize to the olfactory epithelium. NT-3 is the extracellular ligand for TrkC receptor, while BDNF or NT-4/5 can bind to, and activate, TrkB receptors. The targeted deletion of the NT-3 gene in mice increases the number of immature olfactory neurons in the epithelium. Interestingly, NGF immunoreactivity is upregulated in sensory neurons following bulbectomy.

Olfactory receptor neurons of the epithelium sequentially express TrkA, B, and C receptors during neuronal division, commitment, and differentiation. Upon binding neurotrophin ligands, Trk receptors undergo autophosphorylation, and create docking sites for signaling modules such as PI 3-kinases. Trk A is primarily detected in sparse populations of basal cells, and this receptor is highly expressed in differentiating neurons and immature

Tab. 1 Growth factors and receptors in the olfactory epithelium.

Growth factor family	***Ligands***	***Receptors***
Cytokines	CNTF	Unknown
	LIF	Unknown
Dopamine	DA	D_2
EGF family	TGFα	EGFR
	TGFβ-1	
	TGFβ-2	
	NDF	ErbB2
		ErbB3
		ErbB4
FGF family	FGF2	FGFR1
		FGFR1b,c
		FGFR2
		FGFR2b,c
		FGFR3b,c
GDNF family	GDNF	Ret
		GFRα1
IGF family	IGF-I	IGFR-1
	IGF-II	IGFBP-2
		IGFBP-3
		IGFBP-4
		IGFBP-5
Neurotrophins	NGF	TrkA
	BDNF	TrkB
		TrkC
PDGF	PDGFA	PDGFRα
TGFβ family	BMP2, BMP4, BMP7	BMPR-1b
		ActR-1b

Note: ActR-Ib: Activin type Ib receptor; BMP: bone morphogenic protein; BMPR: bone morphogenic protein receptor; CNTF: ciliary neurotrophic factor; DA: dopamine; D_2, D_2 dopamine receptor; EGF: epidermal growth factor; EGFR: epidermal growth factor receptor; ErbB: retrovirus-associated DNA sequences isolated from, or related to, avian erythroblastosis virus; GDNF: glial cell line-derived neurotrophic factor; GFRα1: growth factor receptor, α1; IGF: insulin-like growth factor; IGFBP: insulin-like growth factor binding protein; IGFR: insulin-like growth factor receptor; LIF: leukemia inhibitory factor; NDF: Neu differentiation factor; PDGF: platelet-derived growth factor; PDGFR: platelet-derived growth factor receptor; Ret: rearranged during transformation tyrosine kinase receptor, TGFβ: transforming growth factor-β; Trk: tropomyosin-related tyrosine kinase receptor.

neurons. TrkC is highly expressed only in mature olfactory receptor neurons. Therefore, TrkA-positive basal cells function as neuronal precursor cells, and undergo mitosis to produce TrkB-positive immature neurons that, in turn, develop to become TrkC-positive mature neurons in the olfactory epithelium. Since the TrkC receptor is expressed in the last phase of development, growth factors associated

with this receptor may aid in the survival of olfactory neurons.

The neurotrophin NT-3 is another potential survival factor for olfactory sensory neurons. NT-3 localizes to discrete subsets of olfactory receptor neurons within the olfactory epithelium, as well as to the olfactory bulb. This trophic factor may function as a target-derived factor in olfactory receptor neuron survival and homeostasis, playing a role in coordinating proliferation, maturation, and survival of neurons essential to the maintenance of neuronal population homeostasis. *In vitro* studies have demonstrated that the addition of NT-3 to culture media prevents apoptosis of a select population of olfactory neurons.

Several fibroblast growth factors (FGFs) may also stimulate neurogenesis in the olfactory epithelium. FGFs comprise a large family of polypeptide growth factors, and the FGF family currently has over 20 distinct members. All FGFs bind with high affinity to at least one of four distinct receptor tyrosine kinases. The affinity of membrane-bound FGF receptors for ligands is determined by different sequences in the third extracellular immunoglobulin-like domain of the receptor, and occurs through alternative splicing of messenger RNA. FGF receptor-1 and FGF receptor-2 mRNA have been identified in the olfactory epithelium of rodents. Several reports suggest an effect of FGFs on neurogenesis in the olfactory epithelium. For example, FGF2 appears to have the largest proliferative effect on the olfactory epithelium although FGF1, FGF4, and FGF7 may also significantly increase cell proliferation.

2.2 Olfactory Bulb

In humans, the axons of receptor cells coalesce into 30 to 40 fascicles, termed the *olfactory fila*, formed by ensheathing glial cells. The fila leave the olfactory epithelium to traverse the cribriform plate and pia matter, and the axons form connections within the olfactory bulbs.

The olfactory bulbs are the first place where considerable synthesis of incoming sensory information occurs, and contain a remarkable number of neurotransmitters, including dopamine, GABA, and glutamate. Each bulb is composed of a number of cell types, including neurons, afferent and efferent nerve fibers, microglia, astrocytes, and blood vessels, all surrounded by a thin layer of pia-arachnoid cells. The cellular elements of the bulb are arranged in six concentric layers: the olfactory nerve layer (ONL), the glomerular layer (GL), the external plexiform layer (EPL), the mitral cell layer (MCL), the internal plexiform layer (IPL), and the granule cell layer (GCL). The GCL is the largest of these layers, making up about half the volume of the entire bulb (Fig. 4). Cells within this layer include axonless granule cells that extend a superficial process radially that ramifies and terminates among tufted or mitral cell dendrites within the EPL, and deep short-axon cells, whose axons extend into the EPL.

2.2.1 Olfactory Glomeruli

The first synapse of a CN I bipolar neuron occurs in one or two of the olfactory bulb's *glomeruli*, which are a defining feature of both vertebrate and invertebrate olfactory systems (Fig. 4(a–c)). Although the axons of the incoming olfactory neurons do not branch before reaching the glomeruli, once inside they branch repeatedly and make as many as 18 *en passant* and terminal bouton appositions. The glomeruli are an excellent example of the nervous system's grouping of neural elements and synapses into anatomically

defined modules, and are analogous to the multineuronal "columns" or "barrels" of the cerebral cortex. An olfactory bulb of young adult humans has ~8000 of these 50- to 200-μm structures, arranged in single or double layers within the GL. The development and maintenance of the glomeruli depend on trophic influences exerted by the receptor cells. With age, the number and integrity of the glomeruli greatly decrease, making them nearly absent in a number of persons over the age of 80 years. Considerable convergence of information occurs within the glomeruli, such that many more neurons enter than exit from each glomerulus. For example, in the rabbit, the convergence rates are 25 000 receptor cells onto 100 mitral cells and 500 tufted cells. Importantly, a genetically distinct receptor cell projects to only one or two glomeruli, and a given glomerulus receives most of its input from a restricted zone of the neuroepithelium, at least in the mouse. The topographical organization of glomeruli relative to receptor expression in neurons is discussed in detail in Sect. 7.

During odor stimulation, glutamate is released from axon terminals of olfactory receptor neurons onto the apical processes of mitral and tufted cells within the

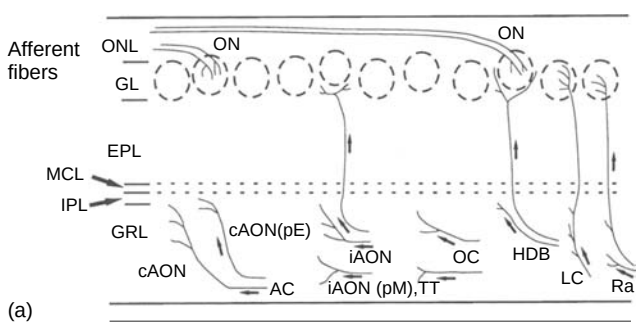

Fig. 4 Neural elements of the mammalian olfactory bulb, grouped according to subdivision into (a) afferent fibers, (b) principal cells, and (c) local interneurons. Diagrams based on various studies using the Golgi staining method and horse radish peroxidase (HRP) tracing. Abbreviations for layers: EPL, external plexiform layer; GL, glomerular layer; GRL, granule cell layer; IPL, internal plexiform layer; MCL, mitral cell body layer; ONL, olfactory nerve layer. (a) ON indicates afferent olfactory nerve fibers. Centrifugal afferents are from the contralateral anterior olfactory nucleus (cAON), ipsilateral olfactory nucleus (iAON), tenia tecta (TT), olfactory cortex (OC), horizontal limb of the diagonal band (HDB), locus coeruleus (LC), and raphe nucleus (Ra). pE, pars externa of the AON; pM, pars medialis of the AON. (b) The dendrites and axon collaterals of the mitral cell (M), an internal tufted cell (T_i, or displaced mitral cell, M_d), a middle tufted cell (T_m), and an external tufted cell (T_e), a, axon; d, dendrite; LOT, lateral olfactory tract. (c) Three types of granule cells (GI, GII, GIII); PG, periglomerular cell; SA(B) Blanes' cell; SA(C), Clandins' cell; SA(G), Golgi cell; SA(H) Hensen's cell; SA(S), Schwann cell; SA(V), van Gehuchten cell. Reprinted with permission from Shepherd, G.M. (2004) *The Synaptic Organization of the Brain*, 5th edition, Oxford University Press, Oxford, p. 167. Copyright 2004 Oxford University Press.

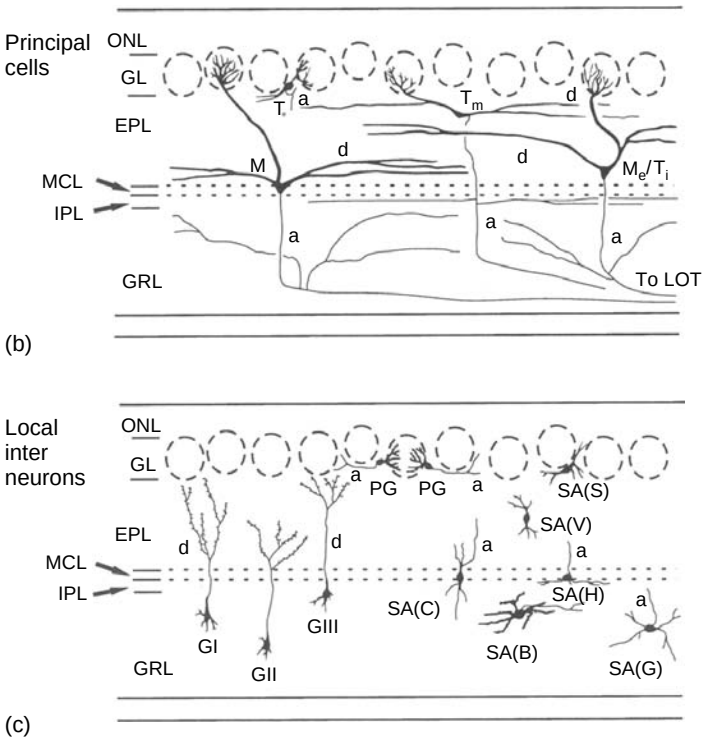

Fig. 4 (*Continued*)

glomeruli, as well as onto dendrites of periglomerular cells that function as interneurons. This glutamate is thought to bind to α-2-amino-3-hydroxy-5-methyl-4-isoxazole-propionic acid (AMPA) kaniate receptors on the dendrites of second-order neurons. As described in the next section, mitral and tufted cell activity is also influenced, usually in an inhibitory manner, by interactions with interneurons and centrifugal fibers primarily within the external plexiform layer of the bulb.

2.2.2 Centrifugal Inputs into the Bulb

The olfactory bulb's neural activity reflects influences from efferent as well as afferent processes. Among central structures that project axons or axon collaterals to the bulb are the pyramidal neurons of the olfactory cortex, the anterior olfactory nucleus, the nucleus of the horizontal limb of the diagonal band (a cholinergic center in the basal forebrain), the locus coeruleus, and the raphe nucleus (Fig. 4a, 8). Most of these structures send fibers to the granule cell layer of the bulb, and several, such as the locus coeruleus and the raphe nucleus, also send fibers to periglomerular regions, presumably inducing inhibitory responses on the primary output neurons of the bulb (i.e. the tufted and mitral cells). Although the exact function of such inputs is not known, they may participate in increasing sensitivity to selected odors based on bodily state (e.g. hunger, sexual arousal). Clearly, as noted in the next section, central brain structures have direct influences on the modulation of olfactory bulb output.

2.2.3 Olfactory Bulb Tuning of Afferent Signals

Dendritic activation of a particular subset of primary bulbar afferent neurons, that is, the mitral and tufted cells, is a selective process. Not all mitral or tufted cells reach the axon potential threshold at any moment in time. In some cases, this reflects too little odorant ligand and, hence, too little receptor cell activity to release an adequate amount of transmitter to induce an action potential. In other cases, the lack of firing of the mitral or tufted cell reflects inhibition from surrounding cells that raise the membrane potential. In this case, the cell cannot fire, even when synaptic input from the incoming receptor cells would, under other circumstances, be effective.

The basis of such tuning, which presumably sharpens contrasts between adjacent channels and leads ultimately to a synthesis of the odor construct, is complex, reflecting associations between local interneurons and dendritic processes of the mitral and tufted cells within and outside of the glomeruli. The activity of the local interneurons is regulated, in many cases, by afferent or efferent circuits, including the centrifugal processes noted in Sect. 2.2.2.

As shown in Fig. 4, the neural elements of the olfactory bulb can be grouped into three major categories: afferent fibers (receptor cell axons, centrifugal inputs), the principal output neurons (mitral and tufted cells), and local interneurons (e.g. granule cells, periglomerular cells). To understand the nature of such interactions, it is instructive to focus on the primary output neuron of the bulb, the "mitral cell." Mitral cells typically have a single primary dendrite that has multiple "tufts" at their terminations within glomeruli. Such cells also have secondary dendrites that radiate horizontally (laterally) over long distances within the EPL. These secondary dendrites are linked by reciprocal synapses to dendritic spines on granule and periglomerular cells. Granule cells, which are GABAergic, form reciprocal dendrodendritic synapses with mitral and tufted cell secondary dendrites that connect with multiple glomeruli. Many of these cells receive the inputs from higher brain regions mentioned in Sect. 2.2.2. Periglomerular cells, which include both GABAergic and dopaminergic subtypes, largely mediate interactions among cells associated with the same glomerulus. They are a heterogeneous group of neurons that synapse, within the glomeruli, to the primary dendrites of both mitral and tufted cells, as well as to axons of receptor cells. These cells also make axodendritic connections outside of the glomeruli (Fig. 5).

The influence of granule cells upon mitral cells is profound. This reflects, in part, the fact that there are up to 100 granule cells synapsing on the dendrites of each mitral cell, and that each granule cell has 50 to 100 spines in those segments that enter into the EPL. As a result of such interactions, the mitral cell's resting potential is raised or lowered, increasing or decreasing its ability to produce an action potential and to influence surrounding bulbar regions. The net effect is that neuronal activity in one area of the bulb can produce a dampening effect on other areas of the bulb. In addition to shaping the odor activity pattern by enhancing or decreasing contrast across segments of the bulb, these circuits also play a critical role in establishing precise timing relationships between the spike activities of different output neurons, thereby increasing the efficiency of the temporal coding of odorants in the bulb.

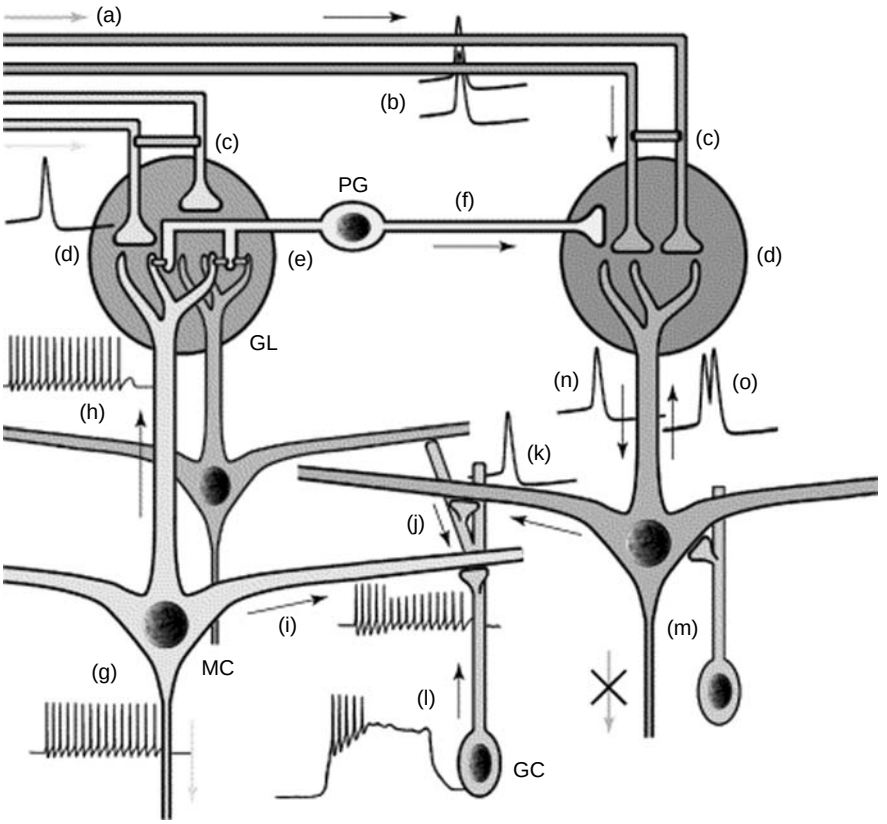

Fig. 5 Electrical signaling mechanisms in the olfactory bulb. (a) Action potentials propagate from olfactory receptor cells to glomeruli (GL), along axons segregated according to expressed odorant receptor genes (e.g. blue or red). Spikes may already be phase locked presynaptically (b) by ephaptic (point of contact where two or more nerve cell processes touch but do not form a typical synaptic contact), and (c) electrical coupling in nerve fascicles. (d) When action potentials invade olfactory nerve terminals, glutamate is released onto apical tufts of mitral cells (MC) and dendrites of periglomerular cells (PG). Tufts that share a glomerulus are coupled electrically by gap junctions, (e) perhaps through periglomerular cells, (f) which may relay timing information among glomeruli. (g) Glutamate evoked excitatory postsynaptic potentials in the tuft can depolarize the mitral soma and axon initial segment, which fires spike trains. (h) Spike output backpropagates up the primary dendrite to trigger glutamate release from tufts, and (e, f) coordinate firing in other mistral cells by electrical coupling. (I) Spikes also propagate laterally into secondary dendrites, in which they are subject to local inhibition by GABA release from granule cells (GC). (j) Lateral inhibition may spread locally among mitral cells through domains in the granule cell apical dendrite. (k) Domains are depolarized by glutamate released from other mitral cell dendrites with propagating spikes. (l) Strong depolarization drives ramped somatic spiking of granule cells, and these spikes back-propagate and release GABA throughout the granule dendrite. (m) Granule cell inhibition of the mitral soma can shift spike initiation to (n) the distal primary dendrite. (o) It can also veto mitral cell output, or generate compound waveforms composed of orthorhombic and back-propagating spikes. Reprinted from Lowe, G. (2003) Electrical signaling in the olfactory bulb. *Curr. Opin. Neurobiol.* **13**, 476–481. Copyright 2003 with permission from Elsevier. (See color plate p. xliii.)

2.2.4 Neurogenesis in the Bulb

It is of interest that cells arising in the subventricular zone of the brain regularly replace a number of neuronal populations within the olfactory bulbs of rodents. Although studies as early as 1912 found mitotic figures within the walls of the lateral ventricle, it was not until the 1960s that thymidine autoradiography confirmed the presence of DNA synthesis and the presence of proliferating cells within the subventricular zone, which lines, lining segments of the lateral ventricles. The progenitor cells that replenish interneurons of the olfactory bulb are localized in the anterior subventricular zone (aSVZ), a remnant of the lateral ganglionic eminence. Within this stem cell niche, cells with structural and molecular characteristics similar to astrocytes function as neural stem cells. These adult stem cells are thought to be derived from radial glial cells, and generate large numbers of neuroblasts. These migrating neuroblasts (see A cells in Fig. 6) express the neuron-specific class III β-tubulin (TUJ1), and the highly polysialated neural cell adhesion molecule (NCAM) that aids in migration. A subpopulation of these neuroblasts then undergoes restricted chain migration along a path known as *the rostral migratory stream*. This migration from the aSVZ terminates in the center of the olfactory bulb and occurs without guidance from radial glia or astrocytes. Reelin acts as a detachment signal for the migrating interneuron precursors, and this protein mediates chain dispersal. After release, differentiating neuroblasts migrate radially (outward) along glial processes, terminating within the periglomerular and granular layers where these neuroblasts mature into interneurons and form synaptic connections. Some of these new neurons become periglomerular cells that surround each glomerulus. A subpopulation of these new periglomerular cells known as juxtaglomerular neurons express enzymes such as tyrosine hydroxylase for synthesis of dopamine, and contain the inhibitory neurotransmitter γ-aminobutyric acid (GABA). In fact, over two-thirds of interneurons in the glomerular layer colocalize dopamine and GABA. Although the significance of neurogenesis of dopaminergic interneurons in the bulb is unclear, exposure of an animal to odorants has a significant effect on stimulating the expression of tyrosine hydroxylase, the rate-limiting enzyme of dopamine biosynthesis. Recent reports suggest that this neurotransmitter may inhibit olfactory sensory neuron activation of olfactory bulb neurons, including juxtaglomerular neurons, by olfactory receptor cells.

The migrating neuroblasts from the aSVZ also repopulate the granule cell layer where they form granular cells that integrate with existing neurons, including mitral cells. In NCAM-deficient mice that exhibit deficits in neuroblast migration to the bulb, the width of the granule cell layer is significantly reduced, and odor discrimination is impaired. Hence, new granule cells likely replace dying neurons, and granule cell activity is critical for normal odor discrimination. Presently, these migrating progenitor cells undergo differentiation by an unknown mechanism.

Factors that regulate neurogenesis in the SVZ are only now becoming apparent. One such factor that has received considerable study is *Vax*1, the ventral anterior homeobox containing gene 1. This gene product directs development of the basal forebrain, the visual system, and neuron turnover in the olfactory bulb. In particular, it regulates basal forebrain development by mediating

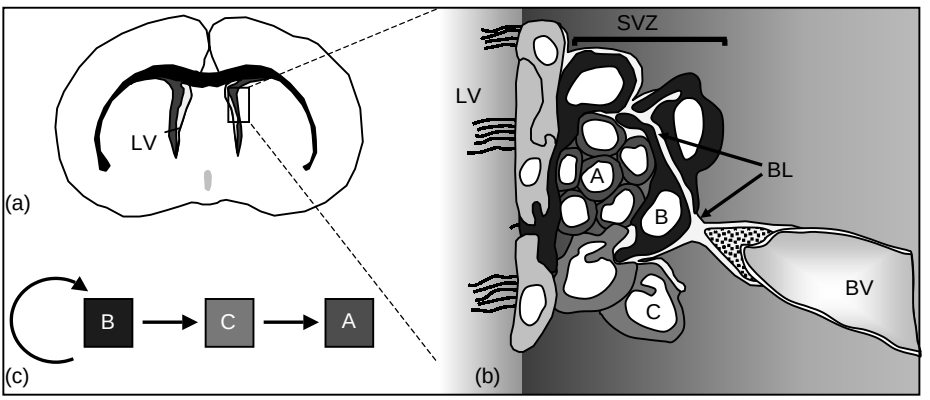

Fig. 6 Adult SVZ Neurogenesis in Rodents. (a) Coronal section through the adult mouse brain. Light blue shows the lateral ventricle (LV) space filled with cerebrospinal fluid. Boxed area is shown enlarged in B. (b) Architecture of the SVZ. B cells (dark blue) are the astrocytes that are stem cells of the SVZ, and also serve as niche cells. Some of the B cells contact the ventricle lumen and contain a single cilium (shown). C cells (green) are rapidly dividing transit-amplifying cells derived from the B cells. C cells give rise to A cells (red), neuroblasts that migrate to the olfactory bulb, where they become local interneurons. A blood vessel (BV, pink) is shown with a perivascular macrophage (dotted fill); a basal lamina (BL, yellow) extends from the BV and interdigitates extensively with the SVZ cells. Ciliated ependymal cells (gray) line the ventricle walls and have been shown to produce the organizer protein Noggin, which is important for this niche. Noggin, BMPs, Sonic hedgehog, Notch, TGFβ, ephrins, ephrin receptors, and vascular endothelial growth factor play roles in regulation of this neurogenic niche. (c) SVZ lineage. C cells can behave as stem cells under the influence of excess EGF signaling. Reprinted from Alvarez-Buylla, A., Lim, D.A. (2004) *Neuron*, **41**, 683–686. Copyright 2004 with permission from Elsevier. (See color plate p. xliv.)

axon guidance and major tract formation in the developing forebrain. *Vax*1 is an important regulator of SVZ organization and neural stem cell proliferation in rodents, where it is highly expressed in the developing and postnatal SVZ and rostral migratory stream. In *Vax1* mutant mice, the SVZ is significantly underdeveloped, and the differentiation and subsequent migration of precursor cells from the SVZ to the bulb is impaired. Since the SVZ supplies neurons to the bulb, olfactory bulbs in mutant mice are depleted of periglomerular neurons.

These stem cell phenomena, which are only now beginning to receive widespread attention within the chemical senses community, are of considerable significance, as they indicate that the plasticity of the olfactory system extends beyond the olfactory neuroepithelium. What role such regeneration plays in odor perception is not entirely clear, although it appears likely that it is an integral component of the olfactory process.

2.3 Olfactory Cortex

After leaving the olfactory bulb, the axons of primary projection neurons form the lateral olfactory tract and eventually synapse on structures collectively termed the *primary olfactory cortex* (Figs. 7 and 8). Although commonly divided into "lateral" and "medial" olfactory tracts in textbooks

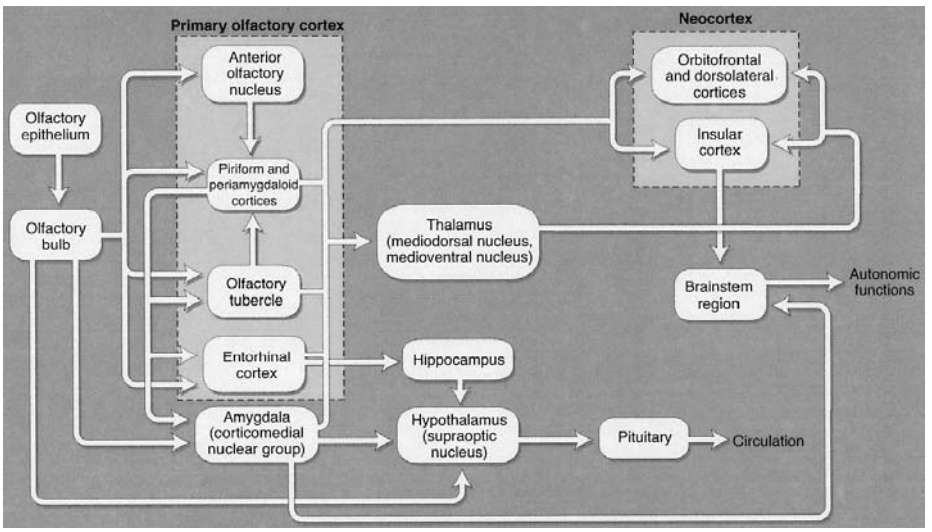

Fig. 7 Major afferent neural connections of the olfactory system. Structures common to both the olfactory and limbic systems include the olfactory tubercle, the entorhinal cortex, the amygdala, the hippocampus, and the hypothalamus. Although it is well established that the olfactory fibers project from the amygdala to the ''feeding center'' of the hypothalamus (especially the ventromedial and ventrolateral nuclei), it is unclear whether direct olfactohypothalamic routes exist in humans, as may be the case in some other vertebrates. Copyright © 2001 Richard L. Doty.

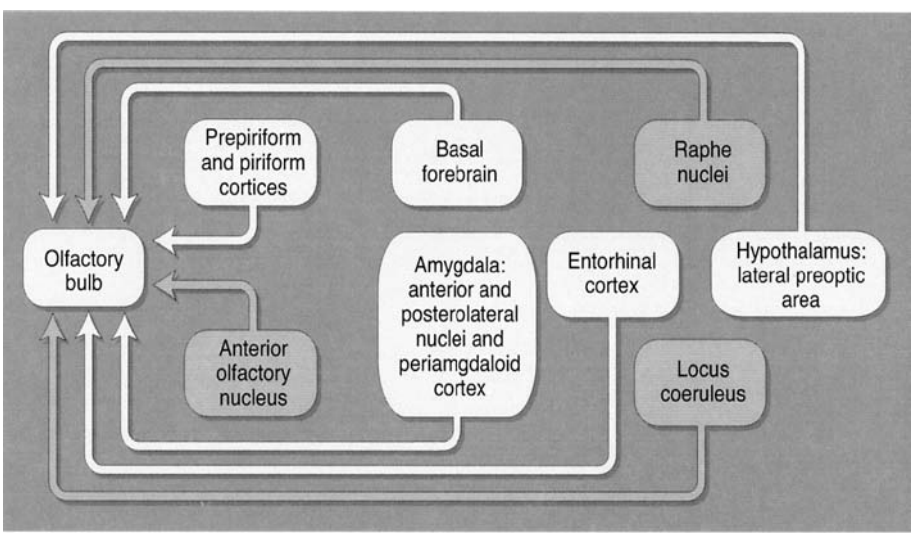

Fig. 8 Major neural structures from which centrifugal fibers project to cells within the olfactory bulb. The anterior olfactory nucleus, locus coeruleus, and raphe nuclei are known to have projections to both olfactory bulbs. The other depicted structures are presumed to have primarily, if not solely, ipsilateral projections to the bulbs. Copyright © 2001 Richard L. Doty.

of anatomy, there is no meaningful medial olfactory tract in primates. In the human, the olfactory tract is relatively flat posteriorly and becomes the olfactory trigone, a small triangular area just rostral to the anterior perforated substance (so named because of the many small holes for blood vessels found throughout this region). It is generally believed that the main olfactory system is unique among sensory systems in sending fibers directly to cortical regions without synapsing in the thalamus.

Although definitions vary, the following structures are commonly considered primary olfactory cortex: (1) the anterior olfactory nucleus (AON), (2) the olfactory tubercle (poorly developed in humans), (3) the piriform cortex, (4) the anterior cortical nucleus of the amygdala, (5) the periamygdaloid complex, and (6) the rostral entorhinal cortex. In many species, the AON extends through the olfactory tract to the rostral part of the olfactory peduncle near the anterior perforated substance. This structure contains pyramidal cells whose dendrites receive synapses from mitral and tufted cells as well as from the contralateral AON (via the anterior commissure (AC)) and numerous central brain structures. The axons of its pyramidal cells project to ipsilateral bulb neurons (mainly granule cells) as well as to the contralateral AON, bulb, or rostral olfactory cortex via the AC. Although termed a nucleus, the AON has a cortical structure with subdivisions based on cellular architecture and connectional patterns.

Like the AON, the other components of the olfactory cortex have rich (and reciprocal) relations with a number of higher brain structures. One such structure is, the entorhinal cortex, which supplies afferent fibers along the entire length of the hippocampus, including the dentate gyrus. Because of these connections, the olfactory system has the most direct access of all sensory systems to the hippocampus. Extensive interactions occur between the cells comprising the superficial and deeper laminae of each component of the olfactory cortex as well as among the components themselves.

In humans, most of the projections from the mitral and tufted cells extend to the more rostral elements of the primary olfactory cortex. It is presently believed that specific sites within the bulb do not preferentially communicate with specific sites within the primary olfactory cortex; that is, no obvious point-to-point topography exists. Thus, small areas of the bulb can project to large areas of the olfactory cortex and vice versa. Among the brain regions that exhibit reciprocal connections with the olfactory cortex are the orbitofrontal cortex, the dorsomedial and submedial thalamic nuclei, the lateral hypothalamus, the amygdala, and the hippocampus. A major pathway from the olfactory cortex to the orbitofrontal cortex is via the mediodorsal nucleus of the thalamus.

3 Basic Structure of the Accessory Olfactory System

As noted in Sect. 1, receptors for the accessory olfactory system are found within the vomeronasal organ (VNO; Fig. 1). The accessory olfactory system consists of the VNO, the vomeronasal duct that opens into the VNO, seromucous glands that secrete into the lumen of the organ, paravomeronasal ganglia lying adjacent to the VNO neuroepithelium, blood vessels adjacent to the epithelium, vomeronasal nerve bundles, the vomeronasal cartilage,

the accessory olfactory bulbs, and central connections of the bulbs. The underdevelopment or lack of development of any one of these structures results in a nonfunctional VNO, as evidenced in a number of comparative studies. Surprisingly, considerable variability exists among a number of tetrapods in the presence of all elements of this system.

In most mammals that have been examined, the VNO is sexually dimorphic in the adult (larger in males than in females), reflecting the organizational action of sex steroids during the early postnatal period. The left and right VNOs are filled with fluid from vomeronasal glands, and are separated from one another by the nasal septum. The VNO neuroepithelium, like that of the main olfactory system, is comprised of neural, supporting, microvillar, and basal cells. As in the case of the main olfactory epithelium, chemicals are sensed by bipolar receptor cells embedded in the epithelium lining the medial concave side of the organ. Unlike the olfactory receptor cells, however, the VNO receptor cells of adult organisms lack cilia, instead containing microvillae, which harbor seven-transmembrane domain receptors (see Sect. 5.2 for a description of such receptors). The large blood vessels and sinuses along the lateral wall of the VNO, whose engorgement is controlled largely by the autonomic nervous system, can induce a pump-like action for bringing materials into the organ. Depending upon the species, entrance of chemicals into the VNO occurs via ducts from the anterior nasal or oral cavities. Stimulus access to the VNO is enhanced in some forms by distinct behaviors, such as the *flehmen* response – a characteristic lip curling and snorting behavior, in which the external nares are closed. This is frequently seen in cows, horses, deer, and sheep as the male samples the female's urine stream prior to estrus.

The axons from the VNO's sensory neurons project through the vomeronasal nerve, which typically runs along the base of the olfactory bulbs to the AOB, the first relay of the system. Recent data suggest the VNO system, unlike the main olfactory system, may be active only postnatally, at least in mice, since transneuronal tracers do not label mitral cells within the AOB prior to that time. Like the VNO proper, the AOB is sexually dimorphic, being larger in males than in females.

Dendritic processes of mitral cells within the AOB communicate, via glomeruli, with the axonal terminals of bipolar receptor cells. Axons of a given class of receptors converge onto numerous glomeruli with the AOB. The axons of the mitral cells project via a fiber bundle to the "vomeronasal amygdala," consisting of the medial and posteromedial cortical amygdaloid nuclei and the nucleus of the accessory olfactory tract. The terminations within the medial nuclei are adjacent, but separate, from those from CN I projections, and this nucleus connects through the stria terminalis to the medial preoptic region of the hypothalamus.

Despite the fact, as mentioned earlier, that nearly all adult humans possess paired VNO-like structures at the base of the septum and paired VNO ducts ~15 to 20 mm from the posterior aspect of the external nares, this organ is nonfunctional. During embryogenesis, the human VNO develops a pseudostratified epithelium and loses receptor cells and their associated neural elements in the second trimester of pregnancy. The vestigial adult human VNO has a comparatively homogeneous epithelium along both its medial and lateral aspects, in contrast to the receptor-rich medial and receptor-free (but vascular rich)

lateral elements of the VNOs of other forms with functional systems. Adult human VNOs contain cilia (which are present only in the early embryonic stages of VNO development in most species) and short microvillae, unlike the elongated microvillae typical of other VNOs. Importantly, the transient receptor potential (Trp)-2 ion channel, required for transduction in the rodent VNO, is a nonfunctional pseudogene in humans. The lack of a VNO nerve and associated accessory olfactory bulb in postnatal humans is in stark contrast to the clearly delineated VNO nerves and accessory olfactory bulbs present in mammals with functioning VNOs, as described in more detail in Sect. 5.2.

4 Early Concepts of Olfactory Signal Transduction

Until the late 1970s, notions of olfactory coding largely reflected structure/activity associations, where activity was indexed by the psychological quality of the odorants to humans. Although physicochemical properties are much more quantitative than measures of odor quality, they too were found wanting. Since many physicochemical properties are highly correlated (e.g. solubility and molecular weight), equivalent prediction of odorant quality could be made using quite different physical parameters, limiting the explanatory value of any set of parameters. What was missing from such theories, in effect, were biological correlates and constraints. Among the more sophisticated of such theories was that proposed by Amoore, in which it was hypothesized that all odorant qualities arose from the relative activity of several "primary odors" in a manner analogous to the mixing of the color primaries of red, yellow, and blue to obtain spectral hues. The notion of primary odors is an ancient one, although Amoore brought to the fore better quantitative psychophysical assessment; that is, the use of threshold testing to establish the existence of the primaries. Thus, he proposed that some individuals have "specific anosmias" to certain odorants and that such anosmias are indicative of the involved primaries. Although some question the existence of specific anosmias (e.g. in many instances hyposmia, rather than anosmia, is present and repeated exposure to an odor can result in transforming a nonsmeller to a smeller), this approach to understanding underlying receptor mechanisms is logical and may still shed some light on the nature of the receptive elements of the main olfactory system. Examples of physicochemical-based theories of olfaction from 1870 to modern times are shown in Table 2.

5 The Modern Era of Olfactory Signal Transduction

During the last 25 years, monumental strides have been made in understanding the initial events of olfactory transduction. It is now known, as described in the following, serious that olfactory receptors are members of a well-identified family of receptors used in a number of sensory, neurotransmitter, and endocrine systems. The involved second messenger systems are well described, and topographic patterns of the convergence of receptor cells harboring individual receptor types are now well mapped, at least in mice.

Most odorants are relatively small (less than 300 Da), volatile, lipophilic molecules that are carried as vapors into the nose. The binding affinities for volatile odorants

Tab. 2 Representative theories and concepts of olfactory function from 1870 to early 1960s.[a]

Author	*Date*	*General class*	*Salient features of smell stimulus*
Ogle	1870	Vibrational	Vibrations affected nasal pigment, which gave out heat that excited the olfactory cells.
Woker	1906	Chemical	Unsaturation main cause of odor; not essential if the substance is very volatile.
Fabre	1911	Vibrational	Limited to insects. Human olfaction due to material particles.
Marchand	1915	Chemical	Unsaturation. Two points of unsaturation reduces odor.
Henning	1916	Chemical	Osmophore groups are important, but their relative position determines the odor quality.
Heyninx	1917	Vibrational	Vibrations causing absorption in the ultraviolet band also cause odor.
Backman	1917	Chemical	Water solubility and lipid solubility are essential.
Teudt	1919	Vibrational	Electronic vibrations of sensory nerves is increased by resonance with similar vibrations of odorants.
Durrans	1920	Chemical	Residual affinity. Addition reaction on the receptors.
Heller	1920	Chemical	Direct chemical action on nerve endings.
Ruzicka	1920	Chemical	Osmophore and osmoceptor.
Tschirch	1921	Chemical	Substance must be soluble in air. Loose compound formed with the plasma of the olfactory cell.
Zwaardemaker	1922	Chemical–Vibrational	Possess odoriphore, be volatile, low surface tension, lipid soluble. Odoriphore depends upon molecular vibrations.
Ungerer & Stoddard	1922	Vibrational	Intramolecular vibrations within definite frequency range. Unsaturation helpful. Interference and resonance effects.
Delage	1922	Chemical	Unsaturation.
Missenden	1926	Chemical	Intensity depends on number of molecules; quality on nature of reaction between molecules and lipoid tissues.
Pirrone	1929	Chemical	Osmophore group determines odor type and variety.
Niccolini	1933	Chemical	Volatility, solubility in nasal mucosa, and oxidizability.
Krisch	1934	Vibrational	Insects.
Müller	1936	Physical	Odorous substances are dipolar, irritate the molecular fields of the osmoceptor in the nose.
Dyson	1937	Vibrational	Volatility. Lipoid solubility. Raman shift between 1400 and 3500 cm^{-1}.
Beck and Miles	1947	Vibrational	Infrared radiation from receptors absorbed by odorants.
McCord and Witheridge	1949	Electrochemical	Change in bonding angle of odorant molecules on solution in mucosa.
Baradi and Bourne	1951	Enzyme	Inhibition of enzyme action by odorants.
Hainer	1953	Information	Thirty levels of intensity; 24 types of primary odors.
Wright	1954	Vibrational	Raman shift of frequency lower than 800–1000 cm^{-1}.
Davies	1954	Physicochemical	Puncturing of cell membrane and exchange of Na^+ and K^+.
Moncrieff	1961	Physical	Volatility, adsorbability, and customary absence from olfactory region.

[a]Modifed from Moncrief, R.W. (1967) *The Chemical Senses*, 3rd edition, London, Leonard Hill, pp. 597–598.

are thought to be in the micromolar range. As described in detail later in this section, the smell produced by an odorant depends on its interaction with a specific receptor protein in the plasma membranes of olfactory cilia. Functional studies of amphibian and rodent olfactory receptors have demonstrated that multiple olfactory receptor proteins recognize individual odorants. Conversely, an individual olfactory receptor can recognize a variety of structurally different odorants, and may allow different response intensities to these odorants. These and other observations suggest that most olfactory receptors are broadly tuned, and have overlapping odor response profiles. This tuning allows for the identification and discrimination of a large number of different odorant molecules.

It is important to emphasize that odorants can exhibit a dual effect on olfactory receptor cells. They can either evoke an inward depolarizing (stimulatory) current, or they can suppress this current in such cells. Also, certain odorants may act as antagonists at the receptor level where they suppress the signaling initiated by structurally related compounds. For example, in the mouse, a derivative of isoeugenol has been shown to act as a potent competitive antagonist for the receptor that binds eugenol. This compound inhibits the physiological response of eugenol in HEK293 cells expressing mouse olfactory receptor "EG." In potential support of such findings are human psychophysical studies that find the perceived intensity of an odorant mixture is often smaller than the sum of the intensities of its components.

Presently, no theory adequately explains how a specific odorant molecule causes the perception of a particular smell. A correlation between the chemical structure of an odorant, the odorant receptor combination it activates, and the perceived odor is not readily apparent. Structure–activity relationship studies suggest that the ligand binding site of an olfactory receptor can discriminate subtle differences in chemical structures of odorants, such as the D- and L-stereoisomers of carvone, while other receptors may be insensitive to large differences in molecular structure. Current evidence suggests that individual odorant receptors may recognize one or several specific structural motifs, or molecular features, that comprise an odorant molecule, and then map these different features to glomeruli in the olfactory bulb. Some theorists have suggested that olfactory receptors respond to intramolecular vibrations, perhaps via inelastic electron tunneling. Whatever the proximate basis of odorant receptor activation, it is clear that a unique combination of activated olfactory receptors and glomeruli encode the critical information passed on to higher brain centers that allow for the perception of an odor.

5.1 Odorant Binding Proteins

Odorant molecules are largely hydrophobic in nature. However, in many cases they must traverse the hydrophilic mucus layer before reaching the receptors of the olfactory cilia. To aid in this process, it has been hypothesized that many odorants bind to hydrophilic proteins that carry them through the aqueous media. Such proteins, termed *odorant binding proteins* (OBPs), have been identified within the nasal secretions at millimolar concentrations in a number of mammals, as well as in some invertebrates. To date, four types of vertebrate OBPs have been reported in mice, three in rabbits, and two in cattle.

The first such protein to be discovered, *odorant binding protein I* (OBP I), is a soluble homodimeric protein of 38 kDa that reportedly binds odorants with affinities in the micromolar range in a manner similar to retinol-binding protein in the visual system. The latter protein conveys retinol from the pigmented retinal epithelium to the rods and cones, where it is incorporated into rhodopsin. OBP I is a member of the lipocalin superfamily of proteins, a family that includes a large set of distantly related proteins that act as carriers for small, lipophilic molecules. This protein is highly concentrated in the tubuloacinar cells of the respiratory epithelium of the nasal mucosa. A second OBP was identified in 1991. This protein, known as *odorant binding protein II* (OBP II), is a secretory protein with DNA sequence homology to OBP I. The OBP II gene is expressed in the lateral nasal gland of rodents. OBP I and II are thought to interact with different classes of odorants.

Although it is generally accepted that OBPs are involved in odorant transport and perhaps in removal of odorants from the receptors, some recent studies are not entirely in accord with this notion. For example, expression studies have found high-affinity binding of unbound pig OBP I to a human olfactory receptor, an effect that would not be expected if OMB I is, in fact, a ligand carrier. Also, the crystal structure of bovine OBP I in complex with odorant molecules at a resolution of 2.0 Å was determined in 2004. No significant conformational changes were detected by structural comparison of crystal structures of unliganded and liganded OBP I, throwing into question whether, in fact, OBP I binds odors in a way analogous to retinol-binding protein in the retina.

5.2 Olfactory Receptors

Odorant molecules, carried in the inspired air into the nose, bind with olfactory receptors located on the cilia of receptor cells, initiating the transduction process. Parameters of airflow, including direction, velocity, and volume, determine the numbers of molecules that reach the receptors. In the human, from 5 to 15% of the inhaled airstream is diverted to the receptor region, depending upon the strength of sniffing and idiosyncratic aspects of nasal cavity anatomy, turbinate engorgement, and mucus secretion and rheology, the latter being affected by adrenergic, cholinergic, and peptidergic neural influences.

Hypothetising that the candidate odorant receptor proteins were members of a G-protein-coupled receptor (GPCR) superfamily, Linda Buck and Richard Axel in 1991, astutely employed a degenerate PCR strategy to amplify complementary DNA (cDNA) sequences from olfactory epithelial tissue encoding well-conserved GPCR amino acid motifs. Degenerate PCR was used to amplify odorant receptor sequences using primers that anneal to common GPCR motifs. This approach was highly successful, and the first 18 members of a novel GPCR gene family, the odorant receptor (OR) gene family, were identified. This landmark research was the impetus for the 2004 Nobel Prize in Medicine or Physiology.

Like other GPCR proteins, OR proteins are predicted to have seven transmembrane (TM) domains with a stereotyped topology (Fig. 9). OR proteins can be distinguished from other GPCR proteins on the basis of uniquely conserved amino acid motifs found at various positions in this predicted protein structure. The functions of these conserved amino acid motifs are

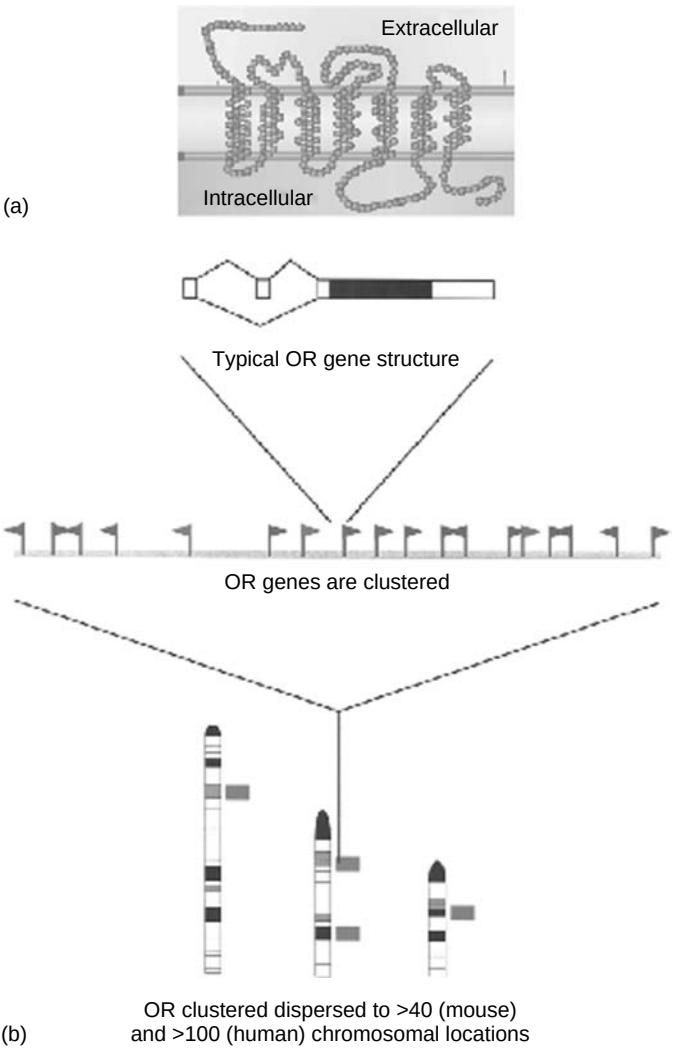

Fig. 9 OR genes. (a) Schematic drawing depicting the stereotyped structure of a GPCR protein, with seven membrane-spanning α-helices. Figure from Mombaerts, 2001. (b) Top: Gene structure and genome organization of OR genes. A typical OR gene consists of upstream noncoding exons that are commonly alternatively spliced, a single coding exon of approximately 1 kb (black), and a short 3′ untranslated region. Middle: OR genes are clustered in groups of 1 to over 100 genes (green arrows) and pseudogenes (red arrows) in both transcriptional orientations. Bottom: OR clusters are dispersed to more than 40 mouse and 100 human chromosomal locations. From Young, J.M., Trask, B.J. (2002) The sense of smell: genomics of vertebrate odorant receptors. *Hum. Mol. Genet.* **11**, 1153–60, by permission of Oxford University Press. (See color plate p. xlvi.)

not yet well understood, with the exception of the absolutely conserved "DRY" motif (aspartic acid, arginine, tyrosine consensus sequence) at the end of the third TM domain, which is required for activation and coupling of the G-protein.

Odorant receptor proteins probably interact with their odorant ligands using the internal TM domains (e.g. TM 3, 4, 5). This hypothesis stems from the characterization of analogous GPCR-ligand interactions, such as the β_2 adrenergic receptor-ligand interaction. It has also been noted that amino acid sequences within TM domains 3, 4, and 5 of odorant receptors are hypervariable in evolution, consistent with a role in interfacing with an enormous repertoire of varied odorants (Fig. 10a). Recently, a structural model of a specific OR protein was reported which was predictive with respect to experimentally established affinities to a panel of odorant ligands (Fig. 10b). In this model, residues in TM domains 3, 5, and 6 were predicted to be directly involved in odorant interactions. The receptor interacts with two different odorants with high affinity, and these interactions are hypothesized to involve different amino acid residues of TM domains. While this model awaits functional or structural studies for validation, these insights point to the possibility that OR proteins may not have fixed ligand pockets. The evolutionary success of the seven TM structure might be due to flexible ligand pockets, capable of many possible ligand interfaces involving different amino acid residues, which nevertheless transduce these diverse interactions into a common G-protein-mediated signaling effect in the cell.

The involvement of G-proteins in odorant transduction stemmed from studies in the early 1980s that reported activation of olfactory cilia by odorants required the presence of GTP. In 1987, a G-protein from an olfactory complementary DNA library was cloned that had 88% amino acid sequence identity to the stimulatory G-protein, G_s, and was strongly and almost exclusively expressed in olfactory cilia. This G-protein was designated $G\alpha_{olf}$ or G_{olf} and was found to be the predominant signaling G-protein in olfactory receptor cells. More recently, gene targeting has been used to examine the functional roles of the G-α subunit, G_{olf}, and type III adenylyl cyclase in olfactory signal transduction. Mice homozygous for a null mutation in G_{olf} exhibit a large reduction in the electrophysiological response of primary olfactory receptor neurons to a wide variety of odorants, verifying the important role of G_{olf} in mediating mammalian odorant signal transduction. Disruption of the gene for a critical subunit of type III adenylyl cyclase resulted in anosmia, as measured behaviorally and electrophysiologically, to all odorants tested, despite the presence of other adenylyl cyclases in the cilia. These results supported previous biochemical and genetic studies that implied that adenylyl cyclase and cAMP are essential for odorant signal transduction in vertebrates.

Although G-proteins other than G_{olf} (e.g. G_{i2} and G_o) have subsequently been identified in olfactory receptor neurons, they appear not to be involved in early transduction events, presumably assisting in such processes as axonal signal propagation, axon sorting, and target innervation. However, G_{i2} and G_o appear to play primary roles in vomeronasal sensory transduction. Interestingly, although G_{olf} was originally thought to be an olfactory-specific protein, we now know it is also expressed in the basal ganglia.

Relatively little is known about the expression of odorant receptor proteins in sensory neurons. Using polyclonal

antibodies raised against some odorant receptors, receptor protein visualization has been reported. Expression occurs as early as embryonic day 14, and is restricted to the cilia and dendritic knobs of receptor cells. However, such studies are not without their limitations. Thus, despite the general usefulness of antisera in biochemical and immunohistochemical studies, the extreme size of the odorant receptor gene family makes it very difficult to demonstrate the specificity of an antibody to a given receptor protein.

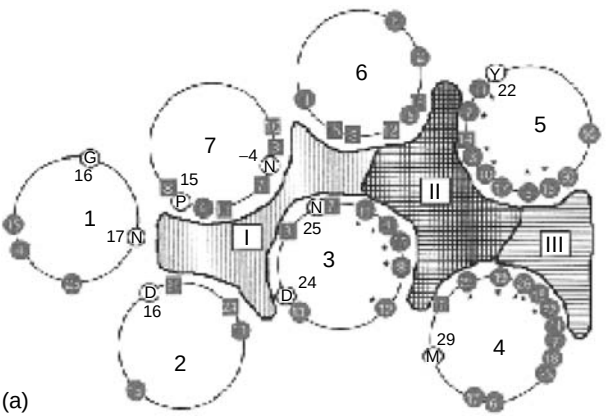

Fig. 10 Predicted odorant-binding pockets of OR proteins. (a) Schematic drawing depicting an overhead view of the seven transmembrane-spanning barrels of an OR protein modeled against the rhodopsin GPCR. Residues conserved among all GPCRs are shown in open circles. Colored squares and circles represent positions of conserved and variable residues, respectively, in OR proteins. Residues that align with ligand contact residues in other GPCRs are colored green and residues that do not align with such residues are colored red. The residues in each helix are numbered separately, according to the predicted TM boundaries. Hypervariable residues (putative odorant-binding residues) are indicated by asterisks. Area II denotes a hypervariable pocket that corresponds to the ligand binding pocket in other GPCRs Figure modified from Pilpel, Y., Lancet, D. (1999). The variable and conserved interfaces of modeled olfactory receptor proteins. *Prot. Sci.* **8**, 969–77. (b) Model depicting the predicted binding pocket for the hexanol ligand (purple) to the S25 OR protein (side view). Each TM and inter-TM loop is labeled. The membrane is represented in yellow. This computed model of the S25 OR protein atoms successfully predicts that relative affinities to a panel of odorants, including hexanol, which is predicted to interact with residues in TMs 3, 5, and 6. Figure modified from Floriano, W.B., Vaidehi, N., Goddard, W.A., Singer, M.S., Shepherd, G.M. (2000) Molecular mechanisms underlying differential odor responses of a mouse olfactory receptor. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 10712–16. Copyright 2000 National Academy of Sciences, U.S.A. (See color plate p. xlv.)

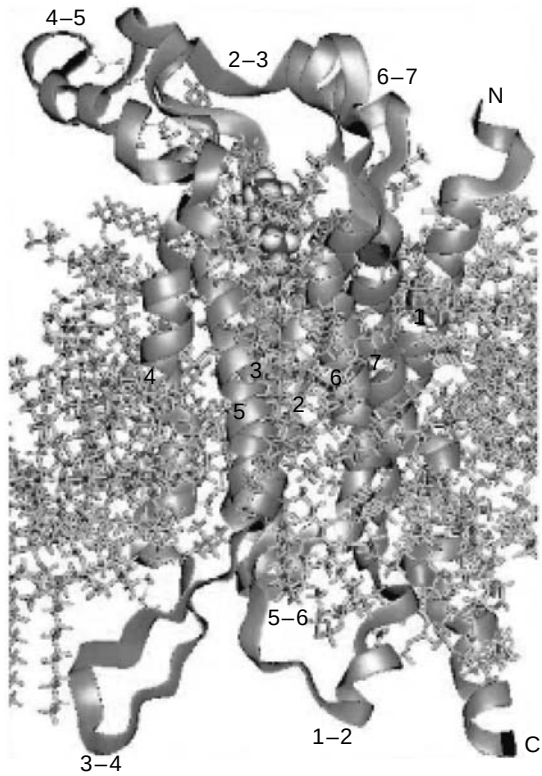

(b)

Fig. 10 (*Continued*)

5.2.1 Identification of Ligand/Receptor Complexes

The inability to efficiently process, translocate, and functionally insert olfactory receptor proteins into the plasma membrane, as well as the low level of olfactory receptor expression in heterologous systems, has impeded the identification of most odorants with their cognate receptors. When expressed in cell lines, mammalian olfactory receptors generally exhibit improper receptor targeting to the plasma membrane, possibly due to improper folding and subsequent retention in the ER. Despite such difficulties, several investigators have overcome these problems and have identified a small number of ligand–receptor interactions.

The discovery of the first odorant/receptor pair occurred in 1997 in the nematode *Caenorhabditis elegans*, a worm that uses either a cGMP-gated or a Trp-gated channel for odorant signal transduction. The *odr-10* gene encodes a putative G-protein-coupled receptor whose expression is regulated by the *odr-7* gene product. In order to identify the function of *odr-10, odr-10* mutants were identified in a genetic screen. Then, mutant worms were exposed simultaneously to the volatile chemoattractants diacetyl and pyrazide. These two odors are sensed by AWA olfactory neurons in *C. elegans*. These worms showed defective chemotaxis only to the volatile odorant diacetyl (butanedione). A human cell line that

expresses the chemosensory receptor protein ODR-10 from *C. elegans* was found to display high ligand selectivity, and to exhibit a transient elevation in intracellular Ca^{2+} levels after addition of the odorant diacetyl. Untransfected cells did not show this response. These results implicated ODR-10 as the receptor for the odorant diacetyl.

In 1998, the first functional expression of an odorant receptor was made in a homologous system where adenovirus-mediated infection of neurons was used to identify which ligands would bind to an orphan olfactory receptor. Olfactory receptor I7 was expressed in mature sensory neurons of the rat olfactory epithelium by adenovirus-mediated gene transfer. For this study, recombinant adenovirus was prepared that encoded the I7 odorant receptor fused to green fluorescent protein (for localization), and introduced into the nasal cavity of rats. After adenovirus transfection of olfactory receptor neurons, *in vivo* electrophysiological recordings (i.e. changes in the summated extracellular field potential measured at the surface of the epithelium, termed the *electro-olfactogram or EOG*), were then used to functionally identify the expressed receptor and its ligands. These transfected sensory neurons responded to the compound *n*-octanol with elevated EOG responses. Smaller electrophysiological responses were found for structurally related aldehydes such as heptyl (C_7), nonyl (C_9), and decyl (C_{10}) aldehydes. However, negligible increases were detected for aldehydes with less than seven, or greater than 10, carbon atoms. In addition, an odorant with some antagonist activity was identified for receptor I7.

Calcium imaging has been employed to identify olfactory neurons that respond to odorants. Then the expressed receptor is identified after cloning. In some cases, a combination of gene targeting and Fura-2 Ca^{2+} imaging has been used to functionally analyze olfactory receptor neurons that express a defined olfactory receptor. For example, acetophenone and benzaldehyde were identified as agonists for mouse and rat M71 olfactory receptor in sensory neurons isolated from mice using a gene-targeted modification that caused coexpression of mouse olfactory receptor 71 and green fluorescent protein. In another experiment, a candidate mouse odorant receptor identified as MOR23, which is expressed in the olfactory epithelium and testis, was cloned from an olfactory neuron that displayed an odorant-induced increase in Ca^{2+}. Then, adenovirus-mediated expression of the cloned receptor gene was carried out in the olfactory epithelium, again with green fluorescent protein as a marker. Calcium imaging demonstrated that the adenovirus-infected neurons elicited Ca^{2+} entry upon stimulation with the odorant lyral. In addition, MOR23 also recognized compounds structurally similar to lyral. Therefore, these virally infected neurons elicited external Ca^{2+} entry when exposed to the same odorant that was originally used to identify the receptor gene.

5.3 Second Messengers

Second messengers are small intracellular molecules that change in concentration in response to environmental signals. Such change amplifies and conveys information inside the cell, which results in a metabolic change.

5.3.1 Adenosine 3′,5′-cyclic Phosphate (Cyclic AMP or cAMP)

A key signaling component of G-protein-coupled receptor activation is the enzyme

adenylyl cyclase, also known as *adenylate cyclase* (Fig. 11). During signal transduction, adenylyl cyclase functions as an effector that catalyzes the breakdown of cytosolic ATP into the second messenger 3′,5′-cyclic AMP, and pyrophosphate. This integral membrane protein contains twelve presumed TM domains, and two large cytoplasmic domains that form the catalytic component of the enzyme. It was discovered in the 1970s that the enzyme adenylyl cyclase, which is usually activated by a stimulatory G-protein, is highly active in olfactory tissue. Thus, adenylyl cyclase activity was found to increase, typically in the presence of GTP, in ciliary preparations following application of a number of olfactory ligands. Subsequent studies from the mid- to late-1980s reported a strong positive correlation between an odorant's ability to activate adenylyl cyclase activity in a frog ciliary preparation and both its perceived odor intensity to humans and the magnitude of the EOG response it produces in frog epithelia. These studies suggested that a functional role exists between the amount of activated adenylyl cyclase and intensity of odor perception, and that this relationship is highly conserved across a wide range of mammalian species. Such studies provided a mechanistic basis by which an odorant-stimulated increase in cyclic nucleotide concentration could lead to increased membrane conductance, to membrane depolarization, and generation of an action potential.

In the mid-1980s, the application of sophisticated patch-clamp techniques allowed for the study of voltage-gated membrane currents, single-channel recordings,

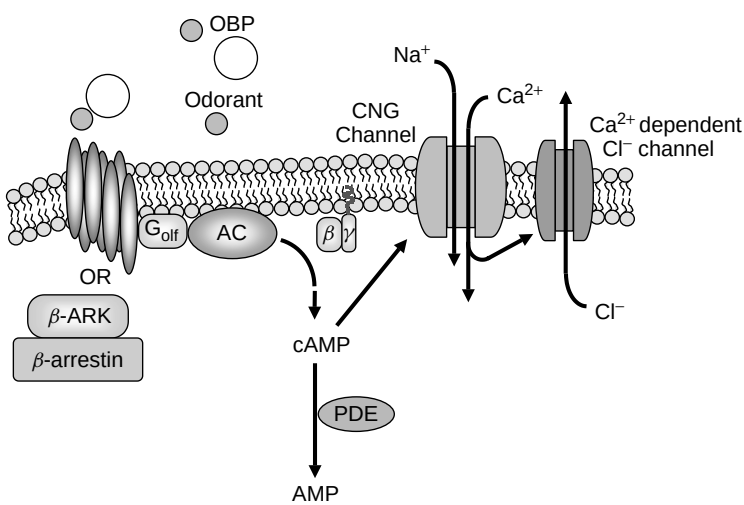

Fig. 11 Model of odorant signal transduction in mammals via the cyclic AMP pathway (see text for details). This signaling cascade mediates the initial steps in odor detection. Abbreviations: AC, adenylyl cyclase; AMP, 5′-adenosine monophosphate; β-ARK, beta-adrenergic receptor kinase; β-arrestin, beta-adrenergic receptor protein; $\beta\gamma$, beta and gamma subunits of GTP binding protein; cAMP, 3′5′-cyclic adenosine monophosphate; CNG channel, cyclic nucleotide-gated channel, G_{olf}, alpha subunit of olfactory GTP binding protein; OBP, odorant binding protein; OR, olfactory receptor; PDE, phosphodiesterase. Copyright © 2005, Nabil Sayed.

and odorant-induced currents in both vertebrate and invertebrate olfactory receptor neurons. For example, using a patch-clamp procedure, a cAMP-gated channel in excised patches of cilia from toad olfactory neurons was identified in 1987. These ciliary patches contained a conductance that was gated directly not only by cAMP, but also by cGMP with a dissociation constant of 20 μM for cAMP and 5 μM for cGMP. At least three molecules of cyclic nucleotides must bind for activation of the membrane channel. Inhibitors of cAMP-gated channels quench odorant responses to this channel. In addition, cAMP, as measured by a rapid quench-flow technique, was rapidly and transiently produced in response to most odors, and some odorants appeared to directly suppress the cAMP-gated conductance by an unknown mechanism. These odorant-stimulated currents are similar to currents induced by flash photolysis of caged cAMP in olfactory neurons. Overall, these findings suggested a functional analogy to the cGMP-gated channel of vertebrate rod and cone photoreceptors. Cyclic nucleotide-gated (CNG) channel of mammalian olfactory neurons shares a 65% sequence identity with the cGMP channel of vertebrate photoreceptors.

Six distinct cyclic nucleotide-gated channel genes have been identified in mammals, and these genes are grouped into two subfamilies known as *CNGA,* which encodes the α-subunit of the channel, and *CNGB* which encodes the β-subunit. Each subunit contains six TM domain segments, a highly variable N-terminal domain, and a 130-amino acid domain at the C terminus. Mammalian olfactory receptor neurons express the three genes *CNGA2, CNGA4,* and *CNGB1b,* and vertebrate cyclic nucleotide-gated channels are heterotetramers that contain ion-conducting α-subunits, and modulatory β-subunits. Functional tetramers assemble by incorporating two subunits of CNGA2, and single subunits of CNGA4 and CNGB1b is highly permeable to Ca^{2+}. Ligand binding occurs at an intracellular cyclic nucleotide-binding domain in each subunit, causing a conformational change that opens the channel to cation influx. The binding domain is linked to the pore region by a "C linker" of approximately 90 amino acids in length. The olfactory cyclic nucleotide-gated channel is regulated by negative-feedback modulation by Ca^{2+}/calmodulin, since the binding of Ca^{2+}/calmodulin results in a decrease in the apparent binding affinity of cyclic nucleotide to these channels.

Recently, a targeted deletion of the cAMP-gated channel in mice showed that these mice were anosmic to all odorants tested when assessed by the electro-olfactogram (EOG). In addition, a mutation in a cyclic nucleotide-gated channel in *C. elegans* disrupted the response to selected odorants in these worms. Taken together, these studies suggest that the CNG channel is required for odorant signal transduction, and provide conclusive evidence that a CNG channel mediates most excitatory olfactory signal transduction in vertebrates such as mice. These results further suggest that cAMP (and possibly cGMP in a subset of sensory neurons) is the second messenger that mediates excitatory odorant signal transduction in mammals via a CNG channel.

The opening of the CNG channel allows an influx of Na^+ and Ca^{2+} into the cell. As Ca^{2+} enters the cytoplasm of olfactory cilia through the channel, this cation activates a secondary depolarizing receptor current that mediates an outward Cl^- conductance in olfactory receptor neurons. Owing to the high Cl^- concentration within the

olfactory receptor neuron, this conductance causes an elevated reversal potential for Cl^-. This increases (depolarizes) the positive charge across the membrane so that an action potential is created. This second current thus amplifies the primary cationic current. The outward Cl^- current also allows a steep sigmoidal dependence on cAMP concentration in both rat and amphibian olfactory neurons. Finally, the sensitivity of this Cl^- current improves the signal-to-noise ratio of vertebrate odorant signal transduction. At present, the Ca^{2+}-dependent Cl^- channel that stimulates this conductance has not yet been identified at the molecular level in mammalian olfactory cilia.

5.3.2 Guanosine 3′,5′-cyclic Monophosphate (cyclic GMP or cGMP)

The fact that some olfactory receptor neurons do not express adenylyl cyclase suggests the existence of other transduction pathways. It is noteworthy that the CNG channel is responsive to cGMP, as well as cAMP, consistent with the hypothesis that cGMP may act as an alternative second messenger.

Two distinct classes of guanylyl cyclases have been identified that may regulate cGMP levels within the olfactory receptor cell: soluble guanylyl cyclase and membrane-bound guanylyl cyclase. The former is activated by gaseous messengers such as nitric oxide (NO) or carbon monoxide (CO), whereas the latter is activated by Ca^{2+} or specific extracellular ligands. Although cGMP is the primary second messenger in visual signal transduction, in olfaction this molecule has a relatively slow and sustained time course of activation. These results imply that cGMP does not play a role in initial olfactory signal transduction. The fact that both guanylyl cyclases are expressed in olfactory receptor cells suggests, however, a complex regulation of cGMP levels occurs within the olfactory receptor neurons.

Soluble guanylyl cyclase is activated by gaseous molecules after binding to the heme iron of protoporphyrin IX, which is a heme oxygenase (HO) inhibitor. NO and CO are generated by nitric oxide synthase (NOS) and heme oxygenase (HO), respectively. NOS most likely plays a role in the development of olfactory receptor cells, since it is expressed primarily in the embryo and is significantly reduced postnatally. In contrast, marked expression of HO only occurs postnatally, suggesting it plays a role in mature receptor neurons. HO-2, one of the two known forms of HO, is highly expressed in mature olfactory neurons, as well as in the outermost (neuronal) and innermost (granule cell) layers of the olfactory bulb. Odorant-induced increases in cGMP levels in the receptor cells are inhibited by Zn protoporphyrin IX, but not by an NOS inhibitor. However, since inhibition of HO does not completely deplete cGMP within the olfactory receptor cells, membrane-bound guanylyl cyclases likely contribute to the production of cGMP.

Considerable evidence suggests that olfactory receptor neurons contain multiple cGMP pathways that mediate both delayed and sustained cGMP odor-induced responses. At least two membrane-bound guanylyl cyclases, a low K_m and a high K_m isoform, have been found in rat olfactory cilia. Since elevation in ciliary cGMP occurs in a Ca^{2+}-dependent manner, and intracellular Ca^{2+} fluctuates in response to odorants, a Ca^{2+}-binding protein, such as guanylyl cyclase activating protein (GCAP), has been hypothesized to regulate particulate guanylyl cyclase within olfactory receptors. A similar pathway occurs in the visual transduction pathway.

Support for this hypothesis comes from immunohistochemical studies that find GCAP1 to be localized to olfactory cilia and to regulate odorant-induced cGMP responses in a Ca^{2+}-dependent manner.

A subset of olfactory neurons that expresses guanylyl cyclase also expresses the cyclic GMP-stimulated phosphodiesterase 2(PDE2), an enzyme that catalyzes the breakdown of cyclic GMP to GMP in these cells. Interestingly, a subunit of the CNG channel previously found in cone photoreceptors is also present in these neurons, suggesting that they may contain a unique CNG channel. Neurons that express PDE2 and membrane-bound guanylyl cyclase D (an olfactory-specific cyclase), are thought to project to a distinct group of glomeruli, localized in the posterior aspects of the olfactory bulb that are similar in appearance to "necklace glomeruli." Necklace glomeruli are important because these structures are a subset of "atypical" glomeruli that receive heavy cholinergic innervation, and may function in mother–infant interactions. Thus, selective compartmentalization of different phosphodiesterases and cyclases is involved in the regulation of odorant signal transduction in olfactory neurons. This selective compartmentalization may modulate the sensitivity of a subpopulation of sensory neurons to specific odorants.

It would appear that cGMP and cAMP produce distinctly different Ca^{2+} signaling kinetics and different cytosolic concentrations for CNG channel activation. Thus, cGMP reaches significantly lower levels in olfactory receptor neurons when compared to cAMP. These differences affect the rate and amount of cation influx into the cytosol of olfactory neurons. This influx may occur through a common CNG channel, or through functionally similar CNG channels with different subunit compositions. The second messenger signal is transient in the case of cAMP, and is more persistent in the case of cGMP. cGMP may thus participate in long-term cellular events in vertebrate olfactory receptor neurons such as desensitization, sensory adaptation, or neuronal activity-dependent transcription.

5.3.3 Inositol-1,4,5-trisphosphate (IP_3)

The PI signaling pathway generates the second messengers IP_3 and (1,2)-diacylglycerol (DAG) by hydrolysis of the membrane lipid phosphatidyl inositol 4,5-bisphosphate via cell surface receptors that couple to stimulatory G-proteins such as $G\alpha_q/G\alpha_{11}$, which then activate phospholipase C-β isoenzymes. DAG activates protein kinase C isoforms, while IP_3 binds to a membrane-bound tetrameric receptor/channel. The IP_3 receptor/channel is composed of a family of structurally related proteins with both distinct and overlapping tissue distributions. Upon binding of IP_3, this receptor/channel stimulates the transient release of Ca^{2+} from non-mitochondrial internal membrane stores (Fig. 12).

IP_3 receptors are highly enriched in ER membranes and sarcoplasmic reticulum membranes of muscle cells. Since odorant signal transduction occurs in cilia, and olfactory cilia contain no ER for Ca^{2+} storage and release, an IP_3 pathway in olfactory neurons, if indeed present, must utilize a plasma membrane IP_3 receptor. Therefore, Ca^{2+} influx would have to occur by this mechanism.

Although IP_3 has been implicated in the odorant response of both vertebrates and invertebrates, its role in mammalian olfactory transduction is controversial. The weight of the evidence suggests that IP_3 is not involved in the initial steps in mammalian odorant signal transduction. A plasma membrane IP_3 receptor/channel,

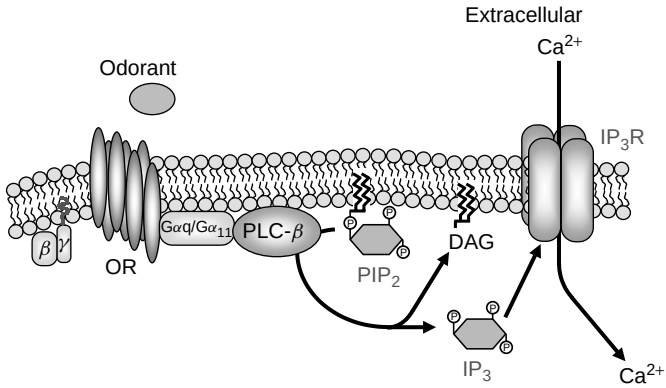

Fig. 12 Model of odorant signal transduction in invertebrates via the IP_3 pathway (see text for details). Abbreviations: $\beta\gamma$, beta and gamma subunits of GTP binding protein; DAG, 1,2-diacylglycerol; $G\alpha_q/G\alpha_{11}$, alpha subunits of Gq class of GTP binding proteins; IP_3, inositol-1,4,5-trisphosphate; IP_3R, plasma membrane inositol-1,4,5-trisphosphate receptor, OR, olfactory receptor; PLC-β, beta isoform of phospholipase C; PIP_2, phosphatidylinositol-4, 5-bisphosphate. Copyright © 2005, Nabil Sayed.

as well as a PI 3-kinase, is involved in odor transduction in the lobster, where considerable research has been carried out. In this crustacean, the olfactory organ consists of an array of hairlike sensilla located on the antennae. Each of these sensilla contain from two to several hundred primary olfactory receptor neurons that are easily recorded electrophysiologically.

In "cultured" lobster olfactory receptor neurons, IP_3 selectively evokes an excitatory current and directly activates two types of channels. Furthermore, odors can activate dual signal transduction pathways in lobster antennules where they both stimulate and inhibit the same olfactory neuron. Both cAMP and IP_3 levels are increased by odor stimulation in the outer dendritic membranes, where IP_3 activates stimulatory current, and cAMP activates the inhibitory current. When IP_3 and cAMP are applied to the inner face of cell-free patches of the outer dendritic membrane of lobster olfactory receptor neurons, both second messenger-gated plasma membrane channels are found in the same patch. In accord with these observations, IP_3 receptors and the PLC-β enzyme localize to the outer dendrites of lobsters, where PLC-β associates with G-proteins in response to odorants.

One possible target of this PI pathway is a Na^+-gated nonselective cation channel that generates a large component of the depolarizing receptor potential. Pharmacological studies have suggested that this Na^+-gated channel in lobsters may belong to the Trp family of plasma membrane receptors, conceivably similar to those reported in *C. elegans* Trp channels are Ca^{2+}-permeable channels, which have been proposed to function as store-operated channels in a variety of cell types in the presence of Na^+. Increased Ca^{2+} levels activate this nonselective cation channel. In addition, this Na^+-gated channel is modulated by addition of exogenous phosphoinositides in cell-free membrane

patches. In such neurons, this plasma membrane channel may amplify the olfactory receptor current.

5.4 Signal Termination and Desensitization

A number of processes are involved in signal termination and desensitization. Such processes include (1) phosphodiesterase action to break down cyclic nucleotides, as previously mentioned for cGMP, (2) phosphorylation of receptors with occupied ligand binding sites, (3) modulation of CNG channels, and (4) conversion into derivatives that do not function as second messengers.

The relative amount of cyclic nucleotide in an olfactory neuron is dependent on both the synthesis and breakdown of that second messenger. Thus, a critical process for olfactory transduction is termination of the odorant-induced signal in the receptor cells. In order to terminate the signal, a rapid decline in cAMP or cGMP formation occurs after odorant stimulation. This decline is catalyzed by cyclic nucleotide phosphodiesterases, which are responsible for the hydrolysis of both cAMP and cGMP to pyrophosphate. At present, three different phosphodiesterases have been identified in olfactory neurons. Of these three enzymes, two phosphodiesterases regulate the hydrolysis of cAMP to 5′-AMP in olfactory cilia for signal termination. A Ca^{2+}/calmodulin-dependent phosphodiesterase IC (PDE1C2), and a Ca^{2+}/calmodulin-independent high-affinity phosphodiesterase (PDE4A) specific for cAMP are both expressed at high levels in many olfactory neurons, with the high-affinity PDE4A being highly expressed in axons.

Many G-protein-coupled receptors show decreased ability to couple to G-proteins after prolonged agonist stimulation, a phenomenon known as *agonist-mediated desensitization*. G-protein-coupled TM receptors, and second messenger-dependent kinases may be involved in this desensitization process. This desensitization is initiated by receptor phosphorylation from G-protein-coupled receptor kinases that function to uncouple the olfactory receptor from its stimulatory G-protein. The phosphorylation and subsequent desensitization of olfactory receptors may occur by a specific receptor kinase named *β-adrenergic receptor kinase* (β-ARK). β-ARK is a heparin-sensitive serine–threonine kinase involved in homologous desensitization of G-protein-coupled receptors. However, the complete quenching of odorant signal transduction at the receptor level may further require binding of the protein cofactor β-arrestin to the phosphorylated olfactory receptor. The binding of β-arrestin results in uncoupling of the receptor from its stimulatory G-protein, and inhibition of cAMP formation. Isoforms of β-adrenergic receptor kinase and β-arrestin have been localized to olfactory neurons, but their function has not yet been demonstrated.

In olfactory neurons, the second messenger-dependent kinase involved in desensitization could also include protein kinase A, which is a serine–threonine kinase that requires cAMP for activation. In olfactory tissue, antagonists of protein kinase A prolong the production of cAMP by odorants. Although odorant-induced phosphorylation of ciliary proteins by protein kinase A has been demonstrated in olfactory tissue, the phosphorylated proteins have not yet been identified. Therefore, some cAMP that is formed in response to odorant stimulation may activate a negative-feedback pathway that

may inhibit the odorant response by phosphorylating specific target proteins.

Another mechanism for desensitization involves the CNG channel, where modulation of cAMP sensitivity of the olfactory CNG channel by Ca^{2+}/calmodulin occurs. When Ca^{2+} ions enter the olfactory cilia through the opened CNG channel, these cations mediate decreased responsivity of the sensory response by providing a negative-feedback signal that modulates the olfactory signal transduction machinery. One important mechanism for this rapid decline in sensitivity to odors is the Ca^{2+}/calmodulin-mediated reduction in cAMP sensitivity of the CNG channel. The modulatory CNGA4 subunit of the channel is thought to accelerate the Ca^{2+}-mediated negative feedback in olfactory signaling, and may allow rapid adaptation in the vertebrate olfactory system. When this gene is deleted, deficiencies in Ca^{2+}/calmodulin-dependent inhibition of the CNG channel occur. This inhibition results in a reduction in adaptation of odor-induced responses in the olfactory neurons. For example, channels in excised membrane patches from the null mouse exhibited slower Ca^{2+}/calmodulin-mediated channel desensitization.

Desensitization may also occur via PI 3-kinase, an enzyme that phosphorylates inusitol phospholipids in the plasma membrane. These lipids then become docking sites for other intracellular signaling proteins. In dissociated rat olfactory receptor neurons, the blocking of PI 3-kinase activity increased an odorant-evoked, cyclic nucleotide-dependent elevation of cytosolic Ca^{2+}. This finding suggests that 3-phosphoinositide signaling may inhibit cyclic nucleotide-dependent excitation of the cells during vertebrate olfactory transduction. Whether this signal transduction pathway acts in concert with the cAMP pathway, or functions to desensitize the cAMP pathway in olfactory receptor neurons, remains to be determined (Fig. 11).

Finally, the second messenger IP_3, is inactivated when this molecule is converted into derivatives that do not open, IP_3 receptor/cannels IP_3 may be converted to inositol by phosphates or phosphorylated to 1,3,4,5-tetrakis phosphate which is then converted to inositol.

6 Olfactory System Receptor Gene Repertoires in Vertebrates

6.1 Main Olfactory System

The biological challenges of the olfactory system are similar to those of the immune system. Both systems must be able to recognize a broad range of environmental chemicals, even ones that the animal has never before been exposed to in its evolutionary past. The breadth and adaptability in antigen recognition of the immune system is based on an elegant combinatorial strategy: a limited number of gene segments are recombined to generate an enormous diversity of antigen receptor combinations. Novel receptor protein combinations permit recognition of novel antigens. Thus, adaptation in the immune system (i.e. the ability to respond to an unpredicted antigen) is a probabilistic outcome of this combinatorial strategy; that is, there is a certain probability that a novel receptor protein combination will be represented in the immune cell population that will be able to bind the novel antigen with high affinity.

However, receptor diversity in the olfactory system is not accomplished by developmentally regulated gene recombination

like in the immune system; rather, it has been accomplished by receptor repertoire expansion and diversification during evolution. The odorant receptor gene family is one of the largest gene families known and it is comprised of ~1000 genes in a typical mammal. The odorant receptor gene repertoire in mouse, for example, is over 1500 genes, which represents >5% of the entire coding content in the mouse genome. Humans and other primates have a smaller OR repertoire (see the following).

OR genes are organized in large clusters in the genome (Fig. 13a). In mouse, clusters range in size from one to more than 100 genes, although cluster sizes of approximately 30 ORs are more typical. Mouse OR clusters have been identified at >40 chromosomal locations and on all chromosomes but 12 and Y (Fig. 13b). Within clusters, the individual mouse OR genes are tightly spaced – a typical intergenic distance is ~20 kb. Accordingly, the OR gene structure is compact, consisting of a single coding exon, ~1 kb in size, and short 5′ and 3′ untranslated exons that are usually separated from the coding exon by 1 to 2 small introns. OR gene clusters are rarely, if ever, interrupted by non-OR genes, and in general, the regions of the genome housing OR clusters are gene-poor.

The functional significance of OR gene clustering, if any, is unknown. However, clustering is at least in part a product of frequent tandem gene duplication events that have taken place in evolution. Since the common ancestor of fish and mammals, OR repertoires have expanded by an order of magnitude (~100 genes in fish vs ~1000 genes in mammals). One wave of OR gene expansion occurred around the time when vertebrates were adapting to land (Fig. 13c). Approximately during the period when amphibians evolved ~370 MYA, a period of radiation occurred in which entire OR clusters duplicated to new locations in the genome. This expansion likely provided the basis for important adaptations in the olfactory system that permitted terrestrial vertebrates to recognize a repertoire of airborne chemicals with their noses.

Following this period of OR cluster radiation early in vertebrate evolution, the OR repertoire has evolved and expanded dynamically by local and nonlocal duplications of OR-containing gene blocks. These duplication events have occurred frequently and in some cases, very recently. For example, a subset of OR duplications near human telomeres have occurred so frequently that homologous blocks are dispersed on >10 chromosomes, and so recently that the copy number and locations of these OR blocks are polymorphic in the human population (Fig. 13a). Duplicated OR genes evolved new amino acid sequences, and thus presumably novel functions, by a number of evolutionary mechanisms. Duplicated ORs tend to be subject to frequent gene conversion events, and/or adaptive selection for amino acid change, all of which are mechanisms that tend to rapidly produce novel protein functions. Like the major histocompatibility (MHC) alleles, OR allelic variation in the human population appears to be unusually large, another indication that ORs have been recently adapting (Allelic variation refers to the numbers of different, or "polymorphic," alleles represented in a population. Genes subject to adaptive selection tend to have high allelic variation.). The ongoing rapid evolution of OR gene repertoires in terrestrial vertebrates presumably reflects selective pressures to adapt to the specific olfactory requirements of individual species in their diverse and rapidly changing niches.

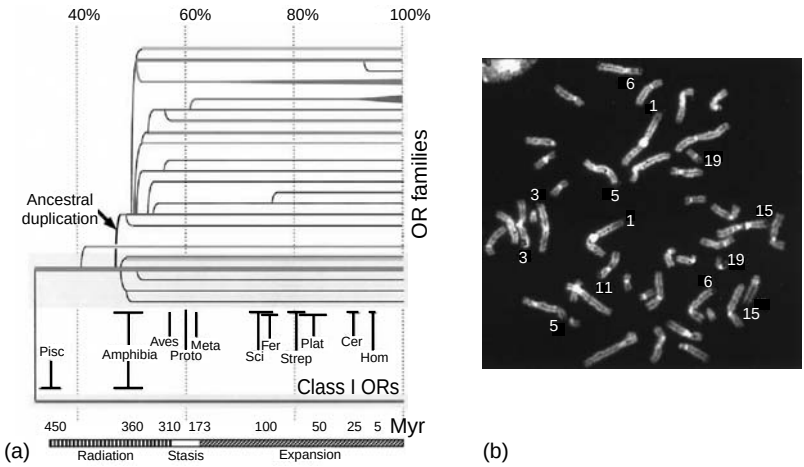

Fig. 13 OR evolution. (a) Application of a molecular clock based on subfamily sequence divergences (% divergence indicated above the tree) to infer periods of OR radiation and expansion during vertebrate evolution. According to this methodology, OR subfamily diversity arose predominantly during amphibian radiation, beginning ~370 million years ago (MYA). So-called "Class I" OR genes are described as *fish-like*, suggesting that these ORs are remnants of the preamphibian radiation. Taxonomy: Pisc = fish; Amphibia = frogs, salamander; Aves = chick; Proto = platypus; Meta = koala; Sci = marmot, mouse, rat; Fer: pig, dolphin, dog; Strep = lemurs, squirrel monkeys, Plat = marmoset; Cer = baboon, macaques; Hom = chimpanzee, gorilla. Figure modified from Glusman, G., Yanai, I., Rubin, I., Lancet, D. (2001) The complete human olfactory subgenome. *Genome Res.* **11**, 685–702. Copyright 2001 Cold Spring Harbor Laboratory Press. (b) Fluorescence *in situ* hybridization (FISH) on human chromosomes using a labeled probe (red) that recognizes a human subtelomeric OR sequence that is dispersed on seven human chromosomes (1, 3, 5, 6, 11, 15, 19) due to frequent duplication and translocation. The copy number and chromosomal distribution pattern of this OR locus is polymorphic in the human population, suggesting very recent duplications. Figure from Trask, B.J., Friedman, C., Martin-Gallardo, A., Rowen, L., Akinbami, C., Blankenship, J., Collins, C., Giorgi, D., Iadonato, S., Johnson, F., Kuo, W.-L., Massa, H., Morrish, T., Naylor, S., Nguyen, O.T.-H., Rouquier, S., Smith, T., Wong, D.J., Youngblom, J., van den Engh, G. (1998) Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum. Mol. Genet.* **7**, 13–26, by permission of Oxford University Press. (c) OR gene tree depicting evolutionary relationships between mouse (red) and human (blue) OR genes. Although the human OR repertoire is less than half the size of the mouse repertoire, the diversity of the human OR repertoire is comparable to that of the mouse repertoire. Therefore, it is expected that mice have better discriminatory power, but not necessarily a broader range of olfaction. Figure from Young et al., 2002 by permission of Oxford University Press. (d) Gene tree depicting species-specific expansion of a mouse and rat V1R subfamily. Most of the duplications occurred over a short period of evolution just after mouse and rat diverged. Figure modified from Lane, R.P., Young, J., Newman, T., Trask, B.J. (2004) Species specificity in rodent pheromone receptor repertoires. *Genome Res.* **14**, 603–608. Copyright 2002 Cold Spring Harbor Laboratory Press. (See color plate p. xlvii.)

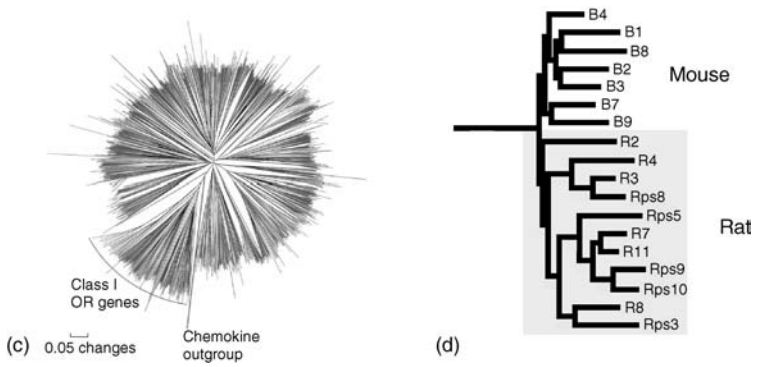

Fig. 13 (*Continued*)

The functional OR repertoire in many primates, including humans, is diminished. While the total number of human OR genes is estimated between 700 and 900 genes, more than half of these genes lack open reading frames that encode seven TM receptor proteins and, therefore, are probably pseudogenes. The rapid decline of the olfactory system in primates probably began shortly before the separation of hominoids and Old World monkeys, concurrent with the advent of trichromatic color vision. This development might have accelerated the displacement of the olfactory system with the visual system as the primary sensory modality in these animals. The evolution of upright posture could have also facilitated this olfactory decline, since such a development would distance the nose from the source of many important smells on the ground. However, humans still have >300 functional OR genes. The human OR repertoire is about as diverse as the mouse repertoire (Fig. 13c), suggesting that humans probably do not have reduced breadth or range of odor detection; rather, that humans may have reduced discriminatory power because of possessing fewer copies of similar genes. It is also important to note that since olfaction is considered a "subliminal" sense (the nose communicates with the subconscious limbic system), we humans may underestimate the influence of our olfactory systems on our own behavior. In later sections of this chapter, clinical implications of the olfactory system in human psychology/behavior are discussed.

6.2 Accessory Olfactory System

Like the main olfactory system, the sensory neurons in thevomeronasal organ (VNO) transduce signals via a G-protein-mediated pathway and predictably, the odorant receptors in the VNO, like the ORs in the main nose, are members of the seven TM/GPCR superfamily. There are two different populations of VNO sensory neurons, each of which express a different GPCR family. In mouse, the sensory neurons of the apical compartment of the VNO express the V1R receptor family, and the neurons of the basal compartment of the VNO express the V2R receptor family (Fig. 1). The ORs of the main nose, and the V1R and V2R genes of the VNO are all GPCRs. However, they are distinct gene families that do not share sequence homology, coupled to different G-proteins, and are expressed in mutually

exclusive populations of neurons. It is likely that the OR, V1R, and V2R receptor gene families represent an example of convergent evolution; that is, independent recruitment of chemosensory functions, as opposed to these three families sharing a common olfactory ancestry.

The V1Rs, like the ORs, have a simple gene structure of an intronless, ~1 kb coding sequence. Also, like the ORs, V1Rs are clustered in the genome. In mouse, there are over 300 V1R sequences (including 165 presumably functional intact genes) distributed on eight mouse chromosomes. In rat, there are over 200 V1R sequences (106 intact) distributed on six rat chromosomes. The large repertoire of intact V1Rs in these two rodent species, however, may be exceptional; recently, the canine V1R repertoire was estimated to consist of only eight intact V1R genes. The human and chimpanzee repertoires are estimated to be less than five intact V1R genes each, and, as noted earlier in this chapter, many primates, including humans, do not possess a functional VNO. Other mammalian species, such as pig, sheep, and ferret, also seem to have reduced VNO function as compared to rodents. Therefore, the rodent VNO might possess a greatly expanded range of specialized functions as compared to other mammals. Alternatively, the VNO in nonrodents might rely more heavily on contributions from the V2R gene family. To date, little is known about the repertoire sizes, genome organization, or biological functions of the mammalian V2R genes.

If V1R receptors are involved in mediating largely conspecific behaviors, as functional studies suggest, then species specificity in these receptors would be expected. A comparison of the mouse and rat V1R repertoires provides striking examples of species-specific expansions or contractions of specific V1R subfamilies (Fig. 13d). Moreover, the duplication events that gave rise to expanded subfamilies in the rodent genomes have been nonuniform, occurring predominantly over a short period of evolution that is coincident with the period when mouse and rat diverged. It is tempting to speculate that some of these species-specific expansions of particular V1R subfamilies helped reinforce mating/social barriers as these two rodent species diverged. However, the functions of individual V1R receptors and their ligands have not yet been elucidated, so the behavioral significance of these striking species-specific events in V1R evolution is unknown.

7 Basic Principles of Olfactory System Organization

How does the olfactory system coordinate the large repertoires of rapidly evolving receptor proteins in order to make sense of an odor? The basic organizing principle of the olfactory system was elucidated by two important discoveries: the discovery that each sensory neuron in the main olfactory epithelium expresses only ~1 OR gene, and that all neurons expressing any given OR gene extend their axons to a common target in the olfactory bulb. As discussed in the following sections, these two attributes of the olfactory system permit a combinatorial representation of smells in the brain – combinatorial patterns of neural activity that the brain can recognize and interpret as particular smells.

7.1 Monogenic Expression of Olfactory Receptor Genes

Several studies indicate that probably one OR gene is transcribed in each

olfactory sensory neuron. The evidence for this hypothesis was initially based on *in situ* hybridization studies conducted on rodent olfactory epithelium sections using individual OR gene probes. Each OR gene is expressed in one of four topological zones within the olfactory epithelium, and within a zone, each OR gene is transcribed in a "punctate" pattern (Fig. 14c). The fraction of cells transcribing RNA of any given OR gene is approximately one in a thousand. This expressing fraction (~0.1%) of olfactory sensory neurons is predicted by the one receptor-one neuron hypothesis, that is, if each olfactory sensory neuron expresses one of the ~1000 OR genes, and each OR is represented approximately equally in the population of olfactory sensory neurons, then any given OR gene should be expressed in ~1/1000 of the neurons. Yet, these studies are inconclusive due to numerous unknowns, such as, the frequency with which a particular OR probe cross-reacts with other similar OR sequences. Furthermore, some OR genes are expressed in more than 10 times the number of cells than other OR genes; therefore, not all OR probes are expected to exhibit a one-in-a-thousand expression pattern.

Another method to directly examine how many ORs are expressed in a single olfactory sensory neuron is to use single-cell reverse transcriptase polymerase chain reaction (RT-PCR) RT-PCR uses the reverse transcriptase enzyme to synthesize DNA templates complementary to RNA templates (cDNA). The cDNA templates can then be amplified by polymerase chain reaction (PCR) using degenerate OR primers to investigate the number of different ORs present in the template sample.

Degenerate PCR primers, designed against well-conserved amino acid sequences in the olfactory receptor protein are successful at amplifying ~50 to 70% of all OR templates in a single reaction. The complexity of OR degenerate PCR products can then be analyzed in bulk by restriction digestions, or by sequencing individual template products from the reaction. A few publications report the amplification of only one OR template in individual sensory neurons, consistent with the one receptor-one neuron hypothesis. However, the false negative rate with this technique is high – more than half of the cells tested do not produce detectable RT-PCR product of any OR cDNA – raising a question as to whether detecting a single OR product might also be an underestimation of the actual OR representation in that cell.

The most compelling evidence for mutually exclusive expression of OR genes is based on the fact that OR alleles are never, or rarely, coexpressed. This was first demonstrated for the two parental alleles of a single OR gene. In these experiments, allele-specific RT-PCR demonstrated that pools of olfactory sensory neuronal cells express one or the other allele, but not both. More recently, it was demonstrated that even identical OR transgenes are not coexpressed in the same neuron (Fig. 14d). (A transgene is a DNA segment typically constructed with a copy of an entire gene (including regulatory regions)). Thus, this principle of mutually exclusive OR expression is not limited to just endogenous alleles situated at homologous native positions on the two parental chromosomes; rather, it seems to apply to pairs of alleles situated anywhere in the genome. While monoallelic expression does not prove monogenic expression, the most parsimonious interpretation of these experiments is that, in general, only one OR locus is active per cell.

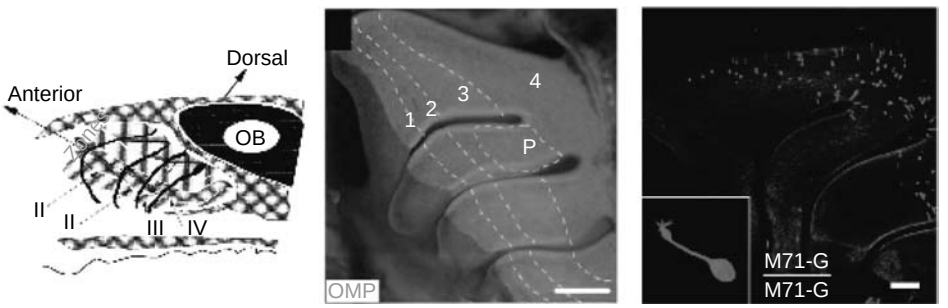

(a) (b) (c)

Fig. 14 OR expression patterns. (a) Schematic of the olfactory epithelium depicting four endoturbinate bones (II, II′, III, IV) and four topological OR expression zones along the dorsal axis (hatchings). Figure modified from Scott, J.W., Shannon, D.E., Charpentier, J., Davis, L.M., Kaplan, C. (1997) Spatially organized response zones in rat olfactory epithelium. *J. Neurophysiol.* **77**, 1950–1962. (b) Fully labeled olfactory epithelium using a construct that reports olfactory marker protein, OMP, expression within olfactory sensory neurons. The four topological zones are labeled (1–4); the patch, P, is a medial zone where some OR genes are expressed. Scale bar is 500 µm. Reprinted from Neuron, Vol. 4, Vassalli, A., Rothman, A., Feinstein, P., Zapotocky, M., Mombaerts, P. (2002). Minigenes impart odorant receptor-specific axon guidance in the olfactory bulb, pages 681–696. Copyright 2002, with permission from Elsevier. (c) Punctate expression of the M71 olfactory receptor gene using a construct in which M71 activation is reported by green fluorescence protein, GFP. Inset, fluorescence image of a single M71 marked olfactory sensory neuron. Scale bar is 250 µm. Figure from Bozza, T., Feinstein, P., Zheng, C., Mombaerts, P. (2002) Odorant receptor expression defines functional units in the mouse olfactory system. *J. Neurosci.* **22**, 3033–3043. Copyright 2002 by the Society for Neuroscience. (d) Coronal section showing mutually exclusive expression of a tagged MOR28 transgene (red) and the endogenous MOR28, tagged with GFP (green). Inset, higher magnification showing individual sensory neurons expressing one color or the other, but not both. Scale bar is 50 µm. Modified from Serizawa, S., Ishii, T., Nakatani, H., Tsuboi, A., Nagawa, F., Asano, M., Sudo, K., Sakagami, J., Sakano, H., Ijiri, T., Matsuda, Y., Suzuki, M., Yamanori, T., Iwakura, Y., Sakano, H. (2000) Mutually exclusive expression of odorant receptor transgenes. *Nat. Neurosci.* **3**, 687–693. Reprinted with permission by Nature Publishing Group. (e) Expression of an alternative OR gene in cells that select a GFP-marked transgene in which the MOR28 coding region is deleted. *In situ* hybridization using non-MOR28 probes (red) colabel the same sensory neuron with the MOR28-deleted transgene (green), producing a merged yellow color (coexpression of the red and green markers). The pie chart summarizes coexpression data for a panel of 13 different OR probes (different colors in chart); ~77% of the cells labeled with GFP (expressing the MOR28-deleted transgene) exhibit alternative expression of one of the OR probes tested. These data suggest that failure to express a functional OR gene results in the neuron choosing another OR locus to express. Reprinted with permissions from Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., Sakano, H. (2003) Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science* **302**, 2088–2094. Copyright 2003 AAAS. Permission from AAAS required for all other uses. (See color plate p. xlix.)

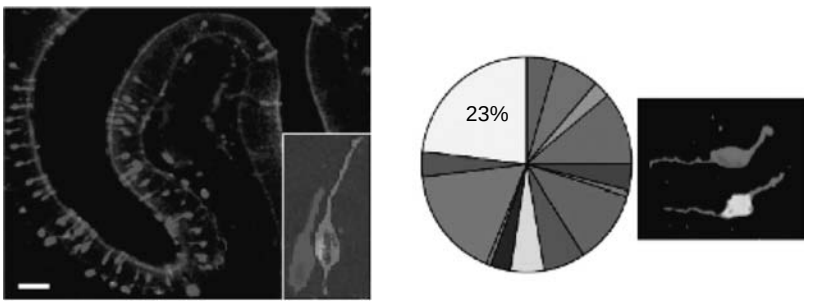

(d) (e)

Fig. 14 (*Continued*)

7.2 Convergence of Olfactory Receptor-expressing Neurons in the Olfactory Bulb

As discussed earlier, the expression pattern of any particular OR gene is "punctate;" that is, it is expressed in a seemingly random ~0.1% fraction of sensory neurons in the olfactory epithelium of the nose. Remarkably, despite originating from various positions throughout a given zone of the olfactory epithelium, each of the neurons that express a particular OR gene extend their axons to a common target in the olfactory bulb (Fig. 15a). The targets, where olfactory sensory neurons synapse with other central neurons, are called *glomeruli*. Since there are two bilaterally symmetrical olfactory bulbs per animal, the set of olfactory sensory neurons expressing a given OR gene projects axons to a pair of symmetric glomeruli per mouse.

Insights about the development of olfactory sensory neuronal projections to the olfactory bulb have come from *knockin* transgenic mouse lines in which specific endogenous ORs have been engineered to report all neurons expressing that OR in the animal. *Knockin* transgenic lines are lines in which an endogenous gene locus has been modified. OR *knockins* typically modify an endogenous OR locus so that it reports the expression of that locus. In 1998, this approach was used to show that the odorant receptor protein itself governs the convergence of like-expressing neurons to common glomeruli. When a marked OR locus has its coding sequence deleted, the neurons that report the activation of this locus fail to converge (Fig. 15a). Thus, by as yet undescribed mechanisms, an olfactory sensory neuron seems to select an olfactory receptor gene to express, and this choice governs the axon convergence of this neuron to a specific glomerulus dedicated to that particular OR.

In mice, there are ~2400 glomeruli (~1200 in each bulb), which will be formed with precise input from specific olfactory sensory neuronal populations. How does this topological map develop? How precise is the map? Does the map change with developmental time or with postnatal neural activity? Some important insights into the role of the OR protein itself in glomerulus targeting have come from "receptor swapping" experiments. Receptor "swapping" experiments replace the coding region of an OR locus with the coding region of another OR gene. These experiments test

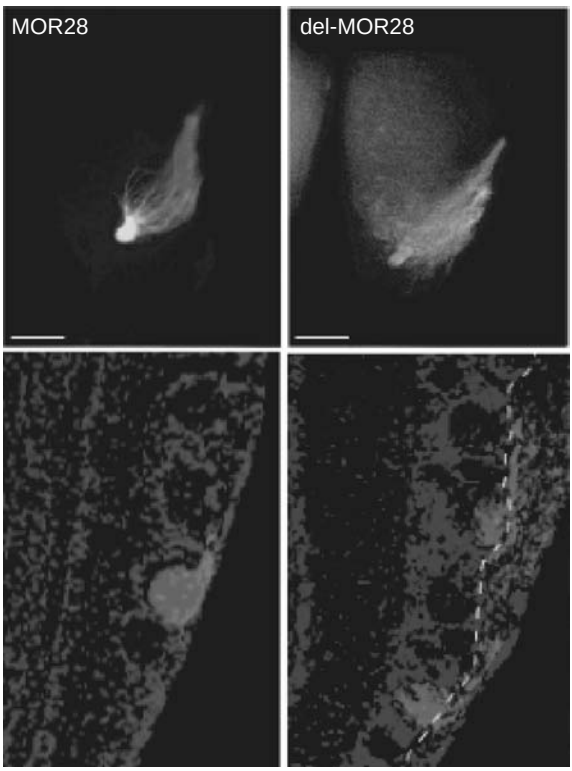

(a)

Fig. 15 Convergence of olfactory sensory neurons expressing a particular OR protein. (a) As first demonstrated by Wang et al. (1998), axons that express a particular OR gene, such as MOR28, converge their axons onto a specific glomerulus target in the olfactory bulb. The left panel shows whole-mount (top) and coronal (bottom) views of multiple convergent GFP-tagged MOR28-expressing axons in the bulb. The right panel illustrates that this convergence is dependent on the expression of the OR gene itself. Whole-mount (top) and coronal (bottom) views of GFP-tagged neurons in which the MOR28 coding region is deleted do not converge. As discussed in Figure 6, these neurons select an alternative OR gene to express, and thus, these axons are likely converging onto other glomeruli specified by this alternative OR choice. Scale bars are 500 μm. Reprinted with permissions from Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., Sakano, H. (2003) Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science* **302**, 2088–2094. Copyright 2003 AAAS. Permission from AAAS is required for all other uses. (b) Model depicting the central organizing principle of the olfactory system. Each olfactory sensory neuron in the olfactory epithelium expresses ~1 OR protein; a given OR gene expression is confined to one of four zones. All sensory neurons expressing a particular OR protein converge their axons onto a common glomerulus in the olfactory bulb. Thus, the stimulation of a subset of OR proteins from a particular smell causes neural activity at a specific combination of glomeruli, which is the neural signature of that smell. Figure from Mombaerts, P. (2001). How smell develops. *Nat. Neurosci Suppl.* **4**, 1192–1198. Reprinted with permission by Nature Publishing group.

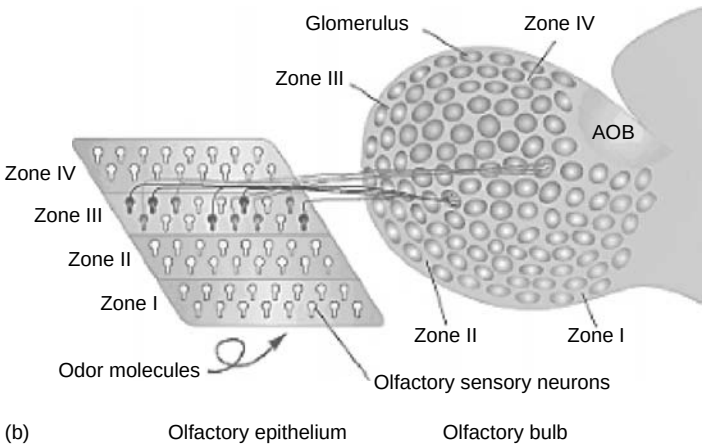

Fig. 15 (*Continued*)

whether the locus (i.e. the surrounding regulatory region) or the coding region of an OR gene are responsible for specifying the targeting of a sensory neuron to its correct glomerulus target. Replacement of the coding region of one OR gene with the coding region of another related OR gene expressed in the same epithelial zone causes neurons expressing the replaced OR to target glomeruli according to the donor OR. Thus, OR coding regions (and presumably the OR protein itself) is sufficient to direct glomerulus targeting. However, replacement of an OR coding region with an *OR* gene normally expressed in a different epithelial zone results in targeting to neither the donor nor the recipient glomerulus, rather, to an altogether unique glomerulus target. These data suggest that developmental context is important in the formation of the glomeruli map (i.e. OR influence on glomerulus targeting depends on the time/place of birth of the olfactory sensory neuron). Remarkably, even single amino acid substitutions in OR coding regions can result in novel glomerulus targeting. Moreover, a non-OR GPCR coding sequence (the $\beta 2$ adrenergic receptor) is able to substitute for an OR and form a glomerulus in the bulb. These data suggest that the sorting of olfactory sensory neurons in the bulb is exquisitely sensitive to OR sequence, yet able to accommodate almost any GPCR sequence.

Sensory neurons expressing a given OR start to become fasciculated as they enter the inner olfactory nerve layer, prior to entering the glomeruli neuropil. The path to a target glomerulus can be convoluted, yet determined, through a maze of other glomeruli. Many molecular guidance cues are likely to play a role in this remarkable projection. One promising candidate set of guidance cues are the *ephrin A* ligands and their *Eph* receptors. The *Eph* receptor tyrosine kinases have been implicated in the formation of the precise projection of retinal ganglion neurons in the visual system. In the visual system, the visual map is preserved between retina and target tissue (the optic tectum or superior colliculus) by a precise one-to-one spatial projection pattern in which retinal ganglion neurons maintain their relative positions. To accomplish this, a gradient of *EphA* receptors on the retinal axons and a gradient of two *ephrin* ligands in the target tissue specify nasotemporal

target positions in the visual map. The olfactory system may also use *EphA* receptors to specify geographic coordinates within the olfactory bulb. Olfactory sensory neurons express differing levels of two *ephrin A* ligands and target cells in the olfactory bulb express *Eph* receptors. Mutant mice lacking the two *ephrin A* ligands, or having increased *ephrin A* expression in subsets of neurons, mistarget their olfactory sensory neurons along the posterior–anterior axis. These data suggest that *ephrin A*s and their receptors are part of the code specifying target positions within the olfactory bulb.

The initial projection of olfactory sensory neurons to target glomeruli seems to be imperfect. For example, the sensory neurons expressing the *P2* OR gene project axons to the glomerular layer at embryonic day 14.5–15.5; however, these early projections are not perfectly converged and mistargeting is observed; clean glomerulus targeting is not observed until well after the mouse is born. Studies with sensory neurons expressing a pair of similar ORs, *M71* and *M72*, suggest that glomerulus maturation requires sensory activity during a sensitive period specific for each glomerulus (i.e. this period is asynchronous for various OR-expressing populations) in order to achieve postnatal refinement of the map. Thus, maturation of the glomeruli map seems to be an activity-dependent process, as it is in the development of other sensory systems like the visual system.

8 The Combinatorial System of Odor Representation

With these two organizing principles – dedication of individual sensory neurons to the expression of a particular OR and the dedication of individual glomeruli to only sensory neurons expressing a given OR – the olfactory system is equipped to make sense of smells in the environment. The complex chemical quality of a smell is thus reduced to a precise combination of sensory neurons that are activated by that odorant blend, which in turn results in a precise combination of active glomeruli in the bulb (Fig. 15b). For example, if a mouse smells cheese, this means that the cheese is emitting a specific set of chemical compounds the mouse must recognize as a cheese smell. Some subset of OR proteins in the mouse nose will specifically recognize this set of compounds, leading to action potentials in the subset of neurons that express these ORs. Because like-neurons wire to like-glomeruli, these action potentials will be sent to a specific combination of glomeruli dedicated to this subset of OR-containing sensory neurons. The result is a pattern of glomeruli activity unique to the smell of cheese. This pattern of glomeruli activity would then be the digital representation of the smell of cheese to the mouse brain. Novel smells should produce novel combinations of activated glomeruli, a pattern of activity that could be learned/remembered by the animal if an association is made between the novel smell and its biological significance.

9 Transcriptional Regulation of Odor Receptor Genes

Thus far, we have described how OR genes have evolved in dynamic ways to modulate gene repertoires on timescales that permit individual species to adapt to their particular niches and lifestyles. Indeed, large OR repertoires, distributed widely in the

genome, seem to be continuously evolving by a robust gene birth/death process. This chapter also describes how the olfactory system is organized around the dedication of individual sensory neurons to individual OR proteins; that is, each individual olfactory sensory neuron transcribes only ~1 OR gene of the entire OR gene repertoire. With these two considerations, a bewildering enigma emerges: how does the transcriptional machinery in an olfactory sensory neuron transcribe only ~1 of ~1000 ORs, while keeping the remaining 99.9% of the OR genes silenced? Such a coregulatory system must operate genome-wide to accommodate the genome-wide distribution of OR genes, must be able to distinguish among OR genes that are tightly clustered and highly similar in sequence, and must be able to "keep track" of the full OR gene repertoire that is large in number and ever changing in evolution.

9.1 Singular Olfactory Receptor Gene Activation

In order for an olfactory sensory neuronal cell to execute one OR transcriptional event, it seems likely that there must be only one OR transcriptional assembly formed in that cell. At least four feasible mechanistic models fit this criterion (Fig. 16): (a) combinatorial promoters, (b) recombination into an active locus, (c) single activation complex, (d) and feedback inhibition.

In the *combinatorial promoters model* (Fig. 16a), each OR promoter could have a unique combination of transcription factor binding sites and each cell expresses a specific combination of these transcription factors. OR activation would then be due to some minimal combination of transcription factors assembled on one of the OR promoters, which will occur uniquely on the OR promoter that has binding sites for the specific combination of factors expressed in that cell. This model has been dismissed because of an experiment described previously (Fig. 16d). Identical transgenes, with identical cis regulatory sequences, are expressed in a mutually exclusive manner. Therefore, mutual exclusive transcription of OR genes is not due to unique transcription factor combinations assembled on a promoter. If it were, then the identical *cis* elements in these two transgenes would be expected to direct the expression of both transgenes in the same cell (since both transgenes would be expected to assemble exactly the same transcription factors on their identical promoter sequences).

In the *recombination into an active locus model* (Fig. 16b), each OR gene could have flanking signal sequences that permit the recombination ("cut and paste") or gene conversion ("copy and paste") of an OR locus into a single, mutually exclusive locus in the genome that is transcriptionally competent (i.e. in this model, native OR loci are transcriptionally incompetent; the OR transcription factor complex only forms at the promoter of the single active locus). Only one of the repertoire of OR genes in the genome can occupy the active locus in any given cell. This model has also been dismissed on several experimental grounds. First, RNA *in situ* hybridization studies on chromosomes from single cells demonstrate that OR RNA is produced from native loci and not from a common, nonnative locus in the genome. Second, two groups have successfully cloned a mouse from a single OR-expressing neuron, and in each case, the cloned mouse produces a

full complement of OR-expressing olfactory sensory neurons. This result argues that a cell's commitment to transcribe an OR gene is not irreversible, as would be predicted by changes in the actual DNA sequence.

In the *single activation complex model* (Fig. 16c), each OR promoter could compete for a single transcriptional complex that is assembled in one physical position in the nucleus. Only one of the repertoire of OR genes in the genome is able to stably associate with this complex in a given cell. Currently, there is no evidence to support or refute this model; however, it is a compelling model based on a process of elimination – it seems to be the best remaining model that would ensure exclusive OR activation. There is a biological precedent for this model. The "trypanosome" parasite mutually exclusively expresses a variable surface glycoprotein (VSG) – only one of a large repertoire of these VSG genes is expressed at a time. Monogenic expression is in part accomplished by an exclusive association of one VSG gene with a single transcriptional complex formed in the nucleus.

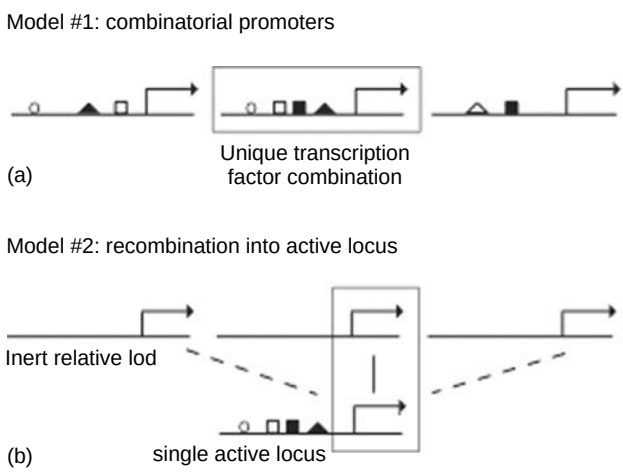

Fig. 16 Models to account for mutually exclusive OR expression. (a) In model #1, each OR promoter has a unique combination of cis elements and each olfactory sensory neuron expresses a combination of transcription factors. Only one OR promoter has this combination of binding sites and is thus, uniquely active in that cell. In the example shown, the binding of four factors is sufficient to activate the middle gene. (b) In model #2, inert OR loci compete to recombine their coding regions into a single active locus in the genome. This model is conceptually similar to the yeast mating type locus. (c) In model #3, each OR promoter competes to associate with a unique transcription factor complex in the nucleus. This model is conceptually similar to the "pol body" utilized by trypanosomes to express only one variable surface glycoprotein (reference). (d) In model #4, transcriptional activation is inefficient such that ~1 OR reaches a superthreshold level of activation first, and this winning gene then inhibits all other loci, thereby securing its unique expression status in the cell. Reprinted with permission from Elsevier.

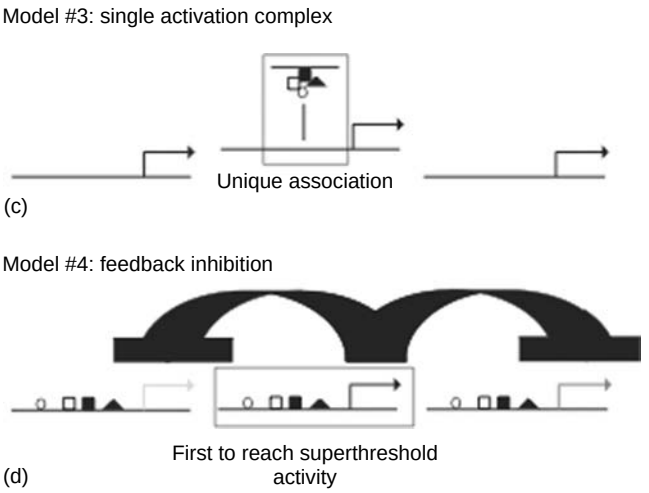

Fig. 16 (*Continued*)

In the *feedback inhibition model* (Fig. 16d), each OR gene could be competent to be transcribed; however, only ~1 OR gene manages to establish an above-threshold activation state. Once this above-threshold activation state is established, this "winning" OR gene is able to negatively feed back onto all other loci, thus closing their opportunity to reach this same level of activation. This model has recently gained momentum on the basis of two independent studies that demonstrate that OR gene selection can take place more than once during a cell's maturation. Specifically, an OR locus can be initially selected; however, if the protein product of this OR gene fails to be produced, the neuron will select another OR locus (Fig. 14e). This alternative OR choice does not appear to be constrained to one or a small number of alternative options: the population of labeled neurons that initially selected the transgene locus is now able to select among a diverse set of secondary OR options and accordingly, project axons to a diverse set of target glomeruli (Fig. 15a).

Notably, the presence of OR protein, as opposed to just RNA, is required to stabilize the OR choice, because expression of transgenes with frameshift mutations (not capable of translating a functional OR protein but still capable of transcription of the RNA) leads to the cell making an alternative choice. Such a mechanism may be the reason why so many OR pseudogenes are tolerated in the genome – if a pseudogene is haphazardly expressed (because of a still active promoter), the absence of a functional protein would, by this feedback mechanism, eventually lead to a functional selection. The molecular details of this feedback pathway are still unknown; however, there is a biological precedent in the immune system that could provide hints. In B cells of the immune system, the pre-B receptor, once stably expressed, signals via a tyrosine kinase pathway to prohibit further activation of B-cell receptor loci.

The feedback inhibition model (Fig. 16d) and the single activation complex (Fig. 16c) are not mutually exclusive models. It is possible that singular activation is accomplished initially by an exclusive association

with a unique transcription factor complex (Fig. 16c), but stabile commitment to this choice requires feedback that the selected OR locus is producing a functional OR protein. Alternatively, singular OR activation prior to feedback inhibition could be a probabilistic outcome of OR transcriptional activation being a rare or inefficient process. In such a scenario, exclusive OR activation would not necessarily be absolute – the "winning" OR gene has presumably reached a superthreshold state of activation first, but there is not an inherent physical barrier in the model to prevent more than one OR gene reaching this same threshold of activation. The probability of reaching a state of activation must be low so that, on average, only one of a thousand genes reaches this status first. This presumed inefficiency is likely to be at the transcriptional level, since by RT-PCR and other molecular methods, more than one abundant OR mRNA in a cell is generally not observed. Weak promoters, limited enhancement, and/or repressive chromatin could contribute to protracted kinetics of OR activation.

9.2 Olfactory Receptor Regulatory Elements

The expression patterns of olfactory receptor genes (Fig. 9) indicate several regulatory features presumably governed by transcription factors interacting with local regulatory elements. First, the olfactory receptor must be activated in the correct cell type and at the correct developmental time. While there are reports of expression of OR genes in nonolfactory tissue, most notably male germ cells in which some OR proteins have been proposed to play a role in chemoreception during sperm–egg communication, all OR gene promoters are presumably interacting with transcription factors expressed in mature olfactory sensory neurons. Second, subsets of olfactory receptor genes are activated within particular topological zones in the olfactory epithelium. Four major geographical zones along the dorsal–ventral axis have been described (Fig. 14); most OR genes examined are expressed widely within only one of these four zones. Some OR genes, however, do not conform to this stereotypic pattern of expression; for example, some ORs are reported to have restricted expression medially between the two dorsalmost zones. Third, as discussed previously, each OR gene must contain regulatory elements that govern mutually exclusive expression. Such elements could bind factors contained with a singular activation complex (see Fig. 16c) or repressors that could be part of a feedback inhibition pathway (see Fig. 16d).

Inspection of mouse genomic sequences have not revealed any universal motifs shared among OR gene promoters, even among subsets of OR genes known to be expressed in the same topological zones. Transgenic approaches have been most fruitful at identifying important *cis* elements. Such experiments aim to elucidate the minimum genomic fragment that, when placed upstream of an OR transgene, can report an expression pattern that recapitulates the *in vivo* expression pattern of the endogenous OR gene. Since stable activation of a single OR gene requires the presence of the OR protein itself (Fig. 14e), OR transgenes are typically designed to produce a bi-cistronic mRNA: both a reporter protein (e.g. *β-galactosidase*) and the OR protein are produced from a single transcript. This is accomplished by inserting an internal ribosomal entry site (IRES) upstream of the second reporter coding region in the transgene construct, so that the ribosome will initiate translation both

at the 5′ end of the transcript as usual (to translate the OR protein), as well as at an internal site just upstream of the reporter coding region (see Fig. 17A). These transgenes are integrated randomly into the mouse genome to produce a transgenic mouse, which can be bred to homozygosity.

Recently, researchers at Rockefeller University produced a minimal OR "minigene" (<10 kb) that recapitulated the expression pattern of the endogenous copy of this OR (Fig. 17A). Approximately 400 bp upstream of the transcription start site was necessary to visualize any expression of the transgene reporter. Given how little upstream sequence is required for transgene expression, the *cis* regulatory region of a typical OR gene seems to be quite proximal. Several additional transgene experiments (with one exception, discussed in the following sections) support the hypothesis that OR regulation is governed locally, presumably within the genomic space between ORs in a cluster. Removal of the 5′ introns of the minigene construct resulted in additional expression outside the normal zonal boundary, implying that zonal restriction is due to inhibitory mechanisms. Sequences 3′ to the coding region do not seem to be important for the transcriptional regulation of this minigene, as well as a panel of other transgenes so far examined. With the putative regulatory territory narrowing, several researchers are now scrutinizing these sequences and using various methods to begin to identify transcription factors that interact with these OR regulatory regions. Other researchers have identified transcription factors that interact with conserved sequences proximal to transcription start sites of a set of closely related OR genes. These experiments yielded six putative OR regulators, including the O/E-2(OS) helix–loop–helix transcription factor, already thought to drive expression of other olfactory neuronal-specific genes, and several homeobox factors that are involved in pattern formation in other tissues.

The paradigm that OR genes are autonomously regulated by proximal elements is supported by several examples of successful mini-transgenes. However, there is one compelling counterexample that indicates some ORs may be coregulated by distant locus control regions (LCRs). In 2000, the Sakano group began characterizing the regulatory structures of a cluster of OR genes that are adjacent to the V-α gene segments of the T-cell receptor (TCR) locus on mouse chromosome 14. In this case, only large ~200-kb OR transgene constructs were successfully expressed. The group later identified a distant ~2-kb sequence, termed the *H region*, required for activation of OR genes in the cluster (Fig. 17B). This enhancer is situated ~75 kb away from the OR cluster and between the OR and TCR gene clusters. Thus, at least this one OR locus seems to be regulated by an LCR, reminiscent of the LCR-mediated coregulation of clustered globin genes. H-region homology has not been identified at other OR loci, suggesting it may be a special feature of this particular OR locus. It is possible that this cluster of OR genes has weak local promoters that require additional enhancement in order to be able to compete with other OR loci for activation. One of the functions postulated for the globin LCR is to open the chromatin structure within the globin locus so that transcription factors are able to access individual globin gene promoters. Similarly, it is possible that extra enhancement from the H region might be needed at this OR locus to counterbalance repressive position effects from the adjacent T-cell

receptor locus, which is transcriptionally silenced in olfactory neurons.

In summary, ~1 OR gene is expressed per mature sensory neuron, and this mutually exclusive expression permits individual neurons to be dedicated to respond to specific odorants. An activated OR locus that produces an intact OR protein is able to inhibit the activation of other OR loci. Each OR gene is expressed in a seemingly random set of sensory neurons, generally within one of four spatial zones in the olfactory epithelium. The sequences responsible for governing OR expression in the correct zone and in a mutually exclusive way appear to be proximal and autonomous, and thus, OR genes do not generally appear to be dependent on their native locus or on clustered relationships for correct expression. The one well-characterized counterexample, the locus-dependent coregulation of an OR gene cluster near the T-cell receptor locus, hints that some OR clusters require additional enhancement, possibly because of having weak local promoters or because

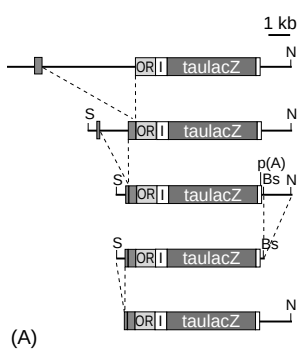

Expr. OSNs	Zone 4	Ecto. zone	Mut. excl.	Co-conv.	Ecto. glom.
+	+	–	+	+	–
+	+	–	+	+	–
+	+	+[b]	+	+	+
+	+	+	+	+	+
–[a]	NA	NA	NA	NA	NA

Fig. 17 Local and distant OR regulatory sequences. (A) mOR23 minigene constructs and resulting expression patterns. In each case, the OR coding region (yellow) is constructed with a downstream internal ribosome entry sequence (i) and reporter gene (*tauLac*Z). The wild-type construct (top) contains 400 bp upstream of the transcription start site, the two 5′ untranslated exons (orange) and introns, and >1 kb 3′ noncoding region. This construct reports the wild-type expression pattern of the mOR23 gene in olfactory sensory neurons. The table reports expression patterns for various deleted constructs; that is, whether the deleted transgene is expressed at detectable levels in olfactory sensory neurons (OSNs), whether its expression is within epithelial zone 4 (the normal zone of expression for mOR23), whether it expresses ectopically outside zone 4 (Ecto. zone), whether the mOR23 transgene exhibits mutually exclusive (Mut. excl.) expression with the endogenous mOR23 gene, whether the OSNs expressing the construct exhibit convergence onto the endogenous mOR23 glomerulus in the olfactory bulb (Co-conv.), and/or whether the OSNs expressing the construct extend axons to multiple glomeruli (+glom.). Deletion of the 400 bp just upstream of the transcription start site of mOR23 abolishes expression of the transgene (bottom construct, a). Deletion of 5′ introns results in transgene expression outside zone 4 (b). Deletion of sequences 3′ to the coding region do not seem to have an effect. Figure modified from Neuron, Vol. 4, Vassalli, A., Rothman, A., Feinstein, P., Zapotocky, M., Mombaerts, P. (2002) Minigenes impart odorant receptor-specific axon guidance in the olfactory bulb, pages 681–696. Copyright 2002, with permission from Elsevier. (B) Expression of OR genes from the MOR28 locus on mouse chromosome 14 requires a distant enhancer, the "H region." Various truncated constructs in which the H region is missing do not report MOR28 expression. The H region is situated ~75 kb upstream of the *MOR28* gene and between the MOR28 OR locus and the T-cell receptor (TCR) Vα gene segments. Inset, dot matrix plot showing conservation of the H region between mouse and human (diagonal trace within box). Reprinted with permission from Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., Sakano, H. (2003) Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science* **302**, 2088–2094. Copyright 2003 AAAS. (See color plate p. l.)

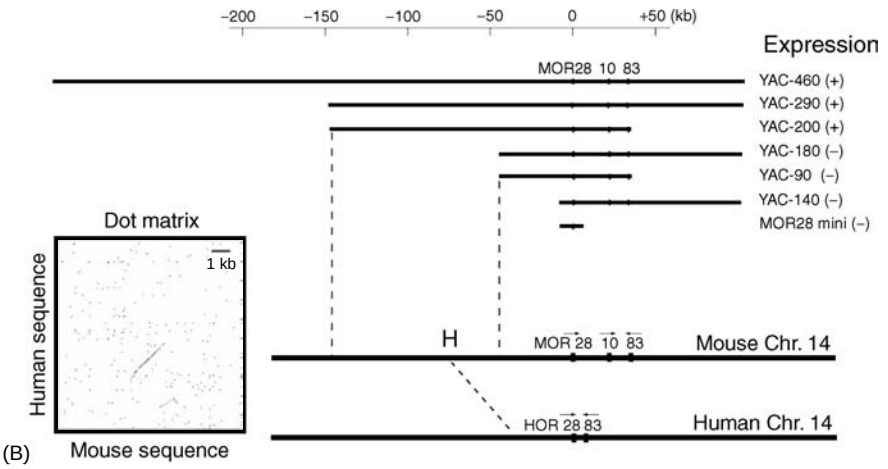

Fig. 17 (*Continued*)

of positioning in an especially repressed region of the genome. The problem of how a large and rapidly evolving repertoire of olfactory receptor genes is so exquisitely regulated remains one of the more intriguing puzzles in molecular biology.

10 Human Diseases Associated with Genetic Defects and Smell Dysfunction

Despite the aforementioned advances in understanding factors associated with signal transduction and the chemotopographic anatomy of the olfactory pathways, such knowledge has shed little light on cases of smell loss and distortion that plague the human population. Most cases of olfactory dysfunction are acquired in adulthood, reflecting damage to the olfactory neuroepithelium secondary to insults from various infections (e.g. viral, bacterial), head trauma, and exposure to xenobiotic agents (e.g. pollutants, industrial dusts, metals, and volatiles). Nevertheless, the advances being made in understanding the processes of olfactory receptor cell regeneration will undoubtedly, in the not-too-distant future, have clinical applications.

That being said, there are numerous inherited diseases associated with olfactory dysfunction. Some of these disorders relate to the abnormal development of the olfactory bulbs and tracts and have an anosmic phenotype. In other cases, the olfactory losses are subtle and of unknown physiologic origin. Alterations in olfaction are only now being discovered in a number of diseases, largely as a result of the development of easy-to-use commercially available quantitative tests of olfactory function. Listed in the following sections are major diseases with at least some heritable component for which quantitative olfactory testing has documented the dysfunction. This listing is representative, and not inclusive.

10.1 Congenital Disorders Associated with Olfactory Dysfunction

There are numerous congenital diseases for which olfactory dysfunction, usually

anosmia, is a key marker. In most cases, a genetic linkage is clear, although in others it is only indirectly inferred and difficult to differentiate from neurodevelopmental problems of unknown cause that can intervene during fetal development.

10.1.1 Bardet–Biedl Syndrome

Bardet–Biedl syndrome (BBS), also known as the *Laurence–Moon–Bardet–Biedl syndrome*, is caused by mutations in at least eight loci, seven of which are now identified. Three of these loci (BBS4, BBS5, and BBS8) encode gene products that are localized to the basal body of cultured cells at ciliated tissue borders. Half of BBS are anosmic or markedly hyposmic. Interestingly, in the study defining the olfactory loss in this syndrome, 4 of 6 patients with mutations in BBS1 (the most common BBS-associated mutation), and 3 of 3 patients with mutations in BBS8, had normal olfactory function, implying multiple gene involvement in the olfactory loss. Mice with the BBS4 gene knocked out have less Type III adenylyl cyclase and $G\gamma_{13}$ signal transduction components in their atypically short cilia. The cilia within the respiratory epithelium are normal. Such mice are anosmic or hyposmic, as measured by reduced EOG amplitudes to a wide range of odorants.

10.1.2 Down Syndrome

Down syndrome (DS) is the most common form of human mental retardation. This trisomy 21 disorder accounts for ~17% of the mentally challenged population. Persons with DS have difficulty in smelling. Given that the average smell loss observed in DS is very close to that observed in Alzheimer's disease (AD), and those individuals with DS who live into early adulthood inevitably develop the clinical and neuropathological features of AD, the question arises as to whether DS and AD share a common olfactory neuropathology. While this is possible, there are several lines of evidence that suggest this may not be the case. First, IQ matched non-DS retarded persons exhibit the same degree of smell dysfunction as seen in DS. Second, the smell loss in DS occurs earlier than that of the neuropathology, at least in terms of senile plaques and diffuse amyloid deposits. The latter generally begin to occur around the age of 19, with significant numbers occurring in the 20s. However, juveniles with DS as young as the age of 11 exhibit the same level of olfactory dysfunction as older DS individuals. Hence, the cause of the olfactory loss remains obscure.

10.1.3 Friedreich Ataxia

Friedreich ataxia is the most common inherited form of ataxia (diseases associated with extreme loss of muscle and movement coordination). This disorder is caused by an abnormal expansion of a GAA trinucleotide repeat sequence in intron 1 of the *FRDA* gene. This sequence is genetically unstable and polymorphic. At the FRDA locus, disease-causing expanded (E) alleles, containing 66–1700 triplets, interfere with gene transcription, resulting in frataxin deficiency and Friedreich's ataxia. The cause of the relatively mild loss of olfactory function in this disorder is unknown, although it is of interest that there is evidence that the cerebellum, per se, may play a role in higher-order olfactory processing. Support for this concept includes (1) early reports of olfactory deficits in patients with tumors in or near the cerebellum; (2) evidence that the staggered mutant mouse, a mouse with functional deficits

in the olivocerebellar pathway, is hyposmic; and (3) functional imaging studies of normal subjects demonstrating marked odor-induced hypermetabolic activity independent of sniffing within regions of the cerebellum. Despite such findings, the actual cause of the olfactory disturbance is not understood.

10.1.4 Kallmann Syndrome

Perhaps the best-known congenital anosmic condition is Kallmann syndrome (KS). In KS, the olfactory bulbs and tracts fail to develop or exhibit significant malformation. Additionally, this disorder is characterized by hypogonadotropic hypogonadism, often first noted as the failure to achieve puberty. Deficits in a gene (*FGFR1*) involved in olfactory development and neuron chimiotactism are believed to be the basis for the malformation of the bulbs and tracts in this disorder. The hypogonadism of this KS appears to reflect mutations of the gonadotrophin releasing factor (GnRH) receptor via the *GPR54* gene.

10.1.5 Pseudohypoparathyroidism and Pseudopseudohypoparathyroidism

Prior to the discovery of G_{olf}, support for the hypothesis that another G-protein, $G_{s\alpha}$, plays a major role in the initial phases of olfactory transduction in humans came from findings of variably decreased olfactory ability in Type Ia pseudohypoparathyroidism (PHP). PHP is a syndrome in which generalized hormone resistance is associated with a deficiency of $G_{s\alpha}$, as measured in red blood cells. However, these patients have other problems that might cause or contribute to their olfactory dysfunction, including an unusual constellation of skeletal and developmental deficits termed *Albright hereditary osteodystrophy* (AHO). More recent research, while confirming that PHP Type Ia patients have defective olfaction, also found that patients with Type Ib PHP, who have no AHO, no generalized hormone resistance, and normal $G_{s\alpha}$ activity exhibit olfactory dysfunction relative to matched controls. Furthermore, this work found that patients with pseudopseudohypoparathyroidism (PPHP), who have AHO, no generalized hormone resistance, and deficient $G_{s\alpha}$ protein activity, have relatively normal olfactory function. These observations do not support the hypothesis that the olfactory dysfunction associated with PHP is the result of generalized $G_{s\alpha}$ protein deficiency, and imply that other mechanisms are responsible for the olfactory deficits of this disorder. Whether G_{olf} is deficient in the olfactory epithelia of PHP patients has not been determined.

10.2 Neurodegenerative Diseases

Some clearly inherited neurodegenerative diseases in addition to DS are associated with smell dysfunction, the classic example being Huntington disease. The genetic involvement in a number of other neurodegenerative disorders, such as multiple sclerosis and most forms of Alzheimer's disease and idiopathic Parkinson's disease, is less clear. Moreover, the causes of the olfactory losses in nearly all neurodegenerative diseases are poorly understood, and only recently have some forms of such diseases been linked to specific genes. In many cases, the olfactory phenotype appears prior to the onset of other symptoms of the disorder. Inferences based upon twin studies, the fact that relatives are more susceptible to the disease, and associations with risk factors such as the *APOE-4* gene

imply heritable elements, although interactions with environmental factors cannot be ruled out.

10.2.1 Alzheimer's Disease

AD, the most common form of dementia, is an idiopathic slowly developing degenerative disease most notable for its memory loss. Smell dysfunction is present in a large proportion of AD patients (85–90%) bilaterally early in the disease process, apparently preceding classic cognitive elements of the disorder. Some inherited forms of AD exhibit smell loss, and the loss observed in idiopathic AD correlates with a family history of dementia, implying genetic linkage. Importantly, there is an association between the AD-related olfactory dysfunction and presence of the APOE-4 allele. For example, in 1999 one group of investigators administered a standardized 12-item smell identification test to 1604 nondemented community-dwelling senior citizens 65 years of age or older. Over a subsequent 2-year time period, the scores on this test were found to be a better predictor of cognitive decline than scores on a global cognitive test. Persons who were anosmic and possessed at least one APOE-4 allele had 4.9 times the risk of having cognitive decline than normosmic persons not possessing this allele. This contrasts with the 1.23 times greater risk for cognitive decline in normosmic individuals possessing at least one such APOE allele. When the data were stratified by sex, women who were anosmic and possessed at least one APOE-4 allele had an odds ratio of 9.71, compared to an odds ratio of 1.90 for women who were normosmic and possessed at least one allele. The corresponding odds ratios for men were 3.18 and 0.67, respectively.

10.2.2 Huntington's Disease

Huntington's disease (HD) is an inherited autosomal dominant neurodegenerative disorder whose phenotype typically manifests between the ages of 30 and 40, although onset can occur at any age. Studies evaluating olfactory function in HD patients have found no dysfunction in gene carriers before the time of the expression of the classical phenotypes of the disease, although it is not yet known whether the smell loss precedes such expression by a year or two, as is seen in AD and Parkinson's disease (PD). The locus for HD is located on the short (p) arm of chromosome 4 at position 16.3. Patients with this disease have from 40 to over 100 trinucleotide repeats (cytosine-adenine-guanine), which is considerably higher than the repeat number normally present (less than 30). There is a rough inverse correlation between the abnormal number of CAG repeats and the age of onset of phenotypic expression, although this is not a clinically useful predictor of either the age of symptom onset or the rate of progression in individual patients.

10.2.3 Multiple Sclerosis

Approximately a quarter of MS patients will exhibit, at any one time, some degree of olfactory dysfunction and, in rare instances such dysfunction can be the presenting symptom of MS. The olfactory loss is intimately related to the plaque-related activity in the subtemporal and subfrontal regions of the brain, such that an inverse relation exists between the number of such plaques and scores on an odor identification test. Such an association is not present with plaque numbers or volumes outside of these olfaction-related brain regions. Longitudinal assessment of smell function in individual MS patients demonstrates a waxing and waning of their olfactory

test scores that occurs in concert with the exacerbation and remission of plaque activity within the subtemporal and subfrontal brain regions. Thus, the olfactory dysfunction is directly associated with the poorly understood cycle of inflammatory processes associated with the disease proper.

10.2.4 Parkinson's Disease

In recent years, it has become apparent that idiopathic PD is not solely a disease of motor dysfunction, but is accompanied by subtle alterations in vision and hearing, and not-so-subtle alterations in the ability to smell. Indeed, like AD, bilateral smell loss is a major component of this disease, with its prevalence being from 85 to 90% of all tested cases – a prevalence greater than that of tremor and rigidity, two of PD's cardinal diagnostic signs. It is noteworthy that James Parkinson, for whom this disorder is named, indicated in his classic description of this disorder that no sensory anomalies were present. It has been speculated that had the smell dysfunction of PD been a visual dysfunction, Parkinson may well have been forced to classify the disorder primarily as a visual one with secondary motor deficits.

The olfactory deficit of PD is not caused by a deficiency in dopamine, as treatment with dihydroxyphenylalanine (L-DOPA) and dopamine agonists have no influence on the dysfunction, which occurs as severely in nonmedicated or never-medicated patients as in medicated ones. As in AD, total anosmia is not the rule for PD, despite its severity and high prevalence. Thus, in one study only 13% of 38 patients who received an odor detection threshold test were unable to detect the highest odorant concentration presented, and only 38.3% of 81 patients had odor identification scores suggestive of anosmia. Moreover, all but one of 41 patients who were asked if an odor was present on each of 40 odorant items answered affirmatively to 35 or more of the items, even though the majority were unable to identify most of the odors, or felt that the perceived sensation did not correspond to the response alternatives. Nonetheless, olfactory testing can easily discriminate between persons with and without PD, including persons with other forms of parkinsonism. The sensitivity and specificity of the 40-item smell test in differentiating between clinically diagnosed PD patients and normal controls was reported in one study as 91 and 80%, respectively, in males $\leq$ 60 years of age. Another study reported the overall sensitivity and specificity of such testing in distinguishing between older persons with PD (mean = 72.6 yrs) and vascular parkinsonism (mean = 74.1 yrs) to be 85.7 and 88.9%, respectively. When the latter data were divided into two categories on the basis of the patient's age (65–75 and 76–88 yrs), the sensitivity and specificity values were 100 and 85.7%, and 85.7 and 80%, respectively. Olfactory testing has been shown to clearly differentiate idiopathic PD from a number of other neurodegenerative diseases with motor symptoms, including disorders often misdiagnosed as PD such as progressive supranuclear palsy (PSP), MPTP-induced PD, and essential tremor.

The genetic contribution to the olfactory loss seen in PD is not clear. However, a recent study found patients with early onset autosomal recessive juvenile parkinsonism associated with the *PARK2* mutation (termed *parkin disease*) have odor identification test scores that do not differ from normal controls. In contrast, individuals with early onset parkinsonism not associated with this mutation exhibit olfactory dysfunction similar to that of adult-onset PD, suggesting that parkin disease may

be a quite different entity than early onset parkinsonism.

11 Summary

The sense of smell is critically important for a broad range of organisms, including our own species. In this chapter, we have described an array of processes involved in peripheral elements of olfactory transduction and how evolution has shaped the receptor repertoire of mammals. Although there is yet much to learn, it is apparent that major advances have been made during the last few decades in our understanding of the mammalian olfactory system and that this field is one of the most active fields of neuroscience. It is also apparent that the complexity of this exquisite sensory system rivals that of any other sensory system.

Among the major advances of the last 20 years are (1) the identification of gene families that express olfactory receptors, (2) the elucidation of membrane mechanisms responsible for stimulus transduction and second messenger signaling, (3) mapping of the topography of receptor-specific pathways through the system, (4) establishing the nature and location of synaptic microcircuits within the olfactory bulb and higher olfaction-related brain regions, and (5) understanding and articulating the general plasticity of the olfactory system, including cell replacement within the olfactory neuroepithelium and olfactory bulb. Such advances, along with those in establishing the causes of smell dysfunction in a variety of genetically determined clinical disorders, make it clear that the next few decades will be exciting ones for further understanding of the sense of smell.

Acknowledgment

Supported, in part, by a grant from the University of Pennsylvania Research Foundation (GSS) and by the following grants from the National Institutes of Health, Bethesda, MD, USA: RO1-DC006267 (RPL), RO1-DC04278 (RLD), and RO1-AG17496 (RLD). We thank the following individuals for providing assistance, including insightful suggestions and comments, on this work: Andrew Chess (Whitehead Institute), Alexis Jackman (University of Pennsylvania), Rarijo Kambere (Wesleyan University), Igor Kratskin (University of Pennsylvania), Harry Rappaport (Temple University), and Nabil Sayed (Temple University).

See also Alzheimer's Disease; Down Syndrome, Molecular Genetics of; Neuron Chemistry.

Bibliography

Books and Reviews

Brennan, P.A., Keverne, E.G. (2003) The vomeronasal organ, in: Doty, R.L. (Ed.) *Handbook of Olfaction and Gustation*, Marcel Dekker, New York, pp. 967–980.

Doty, R.L. (1997) Studies of human olfaction from the University of Pennsylvania Smell & Taste Center, *Chem. Senses* **22**, 565–586.

Doty, R.L. (2001) Olfaction, *Annu. Rev. Psychol.* **52**, 423–452.

Doty, R.L. (2003) Mammalian pheromones: fact or fantasy? in: Doty, R.L. (Ed.) *Handbook of Olfaction and Gustation*, Marcel Dekker, New York, pp. 345–383.

Laurent, G. (2002) Olfactory network dynamics and the coding of multidimensional signals, *Nat. Rev. Neurosci.* **3**, 884–895.

Lennington, J.B., Yang, Z., Conover, J.C. (2003) Neural stem cells and the regulation of adult neurogenesis, *Reprod. Biol. Endocrinol.* **1**, 99–105.

Lledo, P.-M., Gheusi, G., Vincent, J.-D. (2005) Information processing in the mammalian olfactory system, *Physiol. Rev.* **85**, 281–317.

Lowe, G. (2003) Electrical signaling in the olfactory bulb, *Curr. Opin. Neurobiol.* **13**, 476–481.

Mackay-Sim, A. (2003) Neurogenesis in the adult olfactory epithelium, in: Doty, R.L. (Ed.) *Handbook of Olfaction and Gustation*, Marcel Dekker, New York, pp. 93–113.

Mombaerts, P. (2004) Genes and ligands for odorant, vomeronasal and taste receptors, *Nat. Rev. Neurosci.* **5**, 263–278.

Moon, C., Ronnett, G.B. (2003) Molecular neurobiology of olfactory transduction, in: Doty, R.L. (Ed.) *Handbook of Olfaction and Gustation*, Marcel Dekker, New York, pp. 75–91.

Moulton, D.G. (1977) Minimum odorant concentrations detectable by the dog and their implications for olfactory receptor sensitivity, in: Müller-Schwarze, D., Mozell, M.M. (Eds.) *Chemical Signals in Vertebrates*, Plenum Publishing, New York, pp. 455–464.

Shepherd, G.M. (2004) *The Synaptic Organization of the Brain*, 5th edition, Oxford University Press, Oxford.

Smutzer, G.S., Doty, R.L., Arnold, S.E., Trojanowski, J.Q. (2003) Olfactory system neuropathology in alzheimer's disease, Parkinson's disease, and schizophrenia, in: Doty, R.L. (Ed.) *Handbook of Olfaction and Gustation*, Marcel Dekker, New York, pp. 503–523.

Young, J.M., Trask, B.J. (2002) The sense of smell: genomics of vertebrate odorant receptors, *Hum. Mol. Genet.* **11**, 1153–1160.

Primary Literature

Anderson, P.A.V., Ache, B.W. (1985) Voltage- and current-clamp recordings of the receptor potential in olfactory receptor cells *in situ*, *Brain Res.* **338**, 273–280.

Bakalyar, H.A., Reed, R.R. (1990) Identification of a specialized adenylyl cyclase that may mediate odorant detection, *Science* **250**, 1403–1406.

Baker, H., Cummings, D.M., Munger, S.D., Margolis, J.W., Franzen, L., Reed, R.R., Margolis, F.L. (1999) Targeted deletion of a cyclic nucleotide-gated channel subunit (OCNC1): biochemical and morphological consequences in adult mice, *J. Neurosci.* **19**, 9313–9321.

Baker, H., Liu, N., Chun, H.S., Saino, S., Berlin, R., Volpe, B., Son, J.H. (2001) Phenotypic differentiation during migration of dopaminergic progenitor cells to the olfactory bulb, *J. Neurosci.* **21**, 8505–8513.

Bastianelli, E., Polans, A.S., Hidaka, H., Pochet, R. (1995) Differential distribution of six calcium-binding proteins in the rat olfactory epithelium during postnatal development and adulthood, *J. Comp. Neurol.* **354**, 395–409.

Belluscio, L., Gold, G.H., Nemes, A., Axel, R. (1998) Mice deficient in G(olf) are anosmic, *Neuron* **20**, 69–81.

Berendse, H.W.B. (2001) Subclinical dopaminergic dysfunction in asymptomatic Parkinson's disease patients' relatives with a decreased sense of smell, *Ann. Neurol.* **50**, 34–41.

Bhatnagar, K.P., Meisami, E. (1998) Vomeronasal organ in bats and primates: extremes of structural variability and its phylogenetic implications, *Microsc. Res. Tech.* **43**, 465–475.

Bhatnagar, K.P., Kennedy, R.C., Baron, G., Greenberg, R.A. (1987) Number of mitral cells and the bulb volume in the aging human olfactory bulb: a quantitative morphological study, *Anat. Rec.* **218**, 73–87.

Bobkov, Y.V., Ache, B.W. (2004) Pharmacological properties and functional role of a TRP-related ion channel in lobster olfactory receptor neurons, *Neurophysiol.* **93**, 1372–1380.

Boekhoff, I., Michel, W.C., Breer, H., Ache, B.W. (1994) Single odors differentially stimulate dual second messenger pathways in lobster olfactory receptor cells, *J. Neurosci.* **14**, 3304–3309.

Bonigk, W., Bradley, J., Muller, F., Sesti, F., Boekhoff, I., Ronnett, G.V., Kaupp, U.B., Frings, S. (1999) The native rat olfactory cyclic nucleotide-gated channel is composed of three distinct subunits, *J. Neurosci.* **19**, 5332–5347.

Borisy, F.F., Hwang, P.M., Ronnett, G.V., Snyder, S.H. (1993) High affinity cyclic AMP phosphodiesterase and adenosine localized in sensory organs, *Brain Res.* **610**, 199–207.

Bozza, T., Feinstein, P., Zheng, C., Mombaerts, P. (2002) Odorant receptor expression defines functional units in the mouse olfactory system, *J. Neurosci.* **22**, 3033–3043.

Bradley, J., Bonigk, W., Yau, K.W., Frings, S. (2004) Calmodulin permanently associates with rat olfactory CNG channels under native conditions, *Nat. Neurosci.* **7**, 705–710.

Breer, H., Boekhoff, I., Tareilus, E. (1990) Rapid kinetics of second messenger formation in olfactory transduction, *Nature* **345**, 65–68.

Brennan, P.A. (2004) The nose knows who's who: chemosensory individuality and mate recognition in mice, *Horm. Behav.* **46**, 231–240.

Brunet, L.J., Gold, G.H., Ngai, J. (1996) General anosmia caused by a targeted disruption of the mouse olfactory cyclic nucleotide-gated cation channel, *Neuron* **17**, 681–693.

Buck, L., Axel, R. (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition, *Cell* **65**, 175–187.

Bylsma, F.W., Moberg, P.J., Doty, R.L., Brandt, J. (1997) Odor identification in Huntington's disease patients and their offspring with and without the genetic mutation for HD, *J. Neuropsychiatry Clin. Neurosci.* **9**, 598–600.

Calof, A.L., Chikaraishi, D.M. (1989) Analysis of neurogenesis in a mammalian neuroepithelium: proliferation and differentiation of an olfactory neuron precursor in vitro, *Neuron* **3**, 115–127.

Chen, T.-Y., Yau, K.-W. (1994) Direct modulation by Ca^{2+}-calmodulin of cyclic nucleotide-activated channel of rat olfactory receptor neurons, *Nature* **368**, 545–548.

Chen, S., Lane, A.P., Bock, R., Leinders-Zufall, T., Zufall, F. (2000) Blocking adenylyl cyclase inhibits olfactory generator currents induced by "IP(3)-odors", *J. Neurophysiol.* **84**, 575–580.

Chen, J., Tu, Y., Moon, C., Nagata, E., Ronnett, G.V. (2003) Heme oxygenase-1 and heme oxygenase-2 have distinct roles in the proliferation and survival of olfactory receptor neurons mediated by cGMP and bilirubin, respectively, *J. Neurochem.* **85**, 1247–1261.

Chess, A., Simon, I., Cedar, H., Axel, R. (1994) Allelic inactivation regulates olfactory receptor gene expression, *Cell* **78**, 823–834.

Choi, E.J., Xia, Z., Storm, D.R. (1992) Stimulation of the type III olfactory adenylyl cyclase by calcium and calmodulin, *Biochemistry* **31**, 6492–6498.

Coccia, V.J., Cote, R.H. (1994) Regulation of intracellular cyclic GMP concentration by light and calcium in electropermeabilized rod photoreceptors, *J. Gen. Physiol.* **103**, 67–86.

Cohen-Tannoudji, J., Lavenet, C., Locatelli, A., Tillet, Y., Signoret, J.P. (1989) Non-involvement of the accessory olfactory system in the LH response of anoestrous ewes to male odour, *J. Reprod. Fertil.* **86**, 135–144.

Cooper, J.G., Bhatnagar, K.P. (1976) Comparative anatomy of the vomeronasal organ complex in bats, *J. Anat.* **122**, 571–601.

Cutforth, T., Moring, L., Mendelsohn, M., Nemes, A., Shah, N.M., Kim, M.M., Frisen, J., Axel, R. (2003) Axonal ephrin-As and odorant receptors: coordinate determination of the olfactory sensory map, *Cell* **114**, 311–322.

Dal Monte, M., Andreini, I., Revoltella, R., Pelosi, P. (1991) Purification and characterization of two odorant-binding proteins from nasal tissue of rabbit and pig, *Comp. Biochem. Physiol., B* **99**, 445–451.

Dawson, T.M., Arriza, J.L., Jaworsky, D.E., Borisy, F.F., Attramadal, H., Lefkowitz, R.J., Ronnett, G.V. (1993) Beta-adrenergic receptor kinase-2 and beta-arrestin-2 as mediators of odorant-induced desensitization, *Science* **259**, 825–829.

Deems, D.A., Doty, R.L., Settle, R.G., Moore-Gillon, V., Shaman, P., Mester, A.F., Kimmelman, C.P., Brightman, V.J., Snow, J.B. Jr (1991) Smell and taste disorders: a study of 750 patients from the University of Pennsylvania Smell and Taste Center (1981–1986), *Arch. Otolaryngol. Head Neck Surg.* **117**, 519–528.

Del Punta, K., Leinders-Zufall, T., Rodriguez, I., Jukam, D., Wysocki, C.J., Ogawa, S., Zufall, F., Mombaerts, P. (2002) Deficient pheromone responses in mice lacking a cluster of vomeronasal receptor genes, *Nature* **419**, 70–74.

Dhallan, R.S., Yau, K.W., Schrader, K.A., Reed, R.A. (1990) Primary structure and functional expression of a cyclic nucleotide-activated channel from olfactory neurons, *Nature* **347**, 184–187.

Dorries, K.M., Adkins-Regan, E., Halpern, B.P. (1997) Sensitivity and behavioral responses to the pheromone androstenone are not mediated by the vomeronasal organ in domestic pigs, *Brain Behav. Evol.* **49**, 53–62.

Doty, R.L., Deems, D.A., Stellar, S. (1988) Olfactory dysfunction in parkinsonism: a general deficit unrelated to neurologic signs, disease stage, or disease duration, *Neurology* **38**, 1237–1244.

Doty, R.L., Kreiss, D.S., Frye, R.E. (1990) Human odor intensity perception: correlation with frog epithelial adenylate cyclase activity and transepithelial voltage response, *Brain Res.* **527**, 130–134.

Doty, R.L., Bromley, S.M., Stern, M.B. (1995) Olfactory testing as an aid in the diagnosis of Parkinson's disease: development of optimal discrimination criteria, *Neurodegeneration* **4**, 93–97.

Doty, R.L., Singh, A., Tetrude, J., Langston, J.W. (1992) Lack of olfactory dysfunction in MPTP-induced parkinsonism, *Ann. Neurol.* **32**, 97–100.

Doty, R.L., Li, C., Mannon, L.J., Yousem, D.M. (1999) Olfactory dysfunction in multiple sclerosis: relation to longitudinal changes in plaque numbers in central olfactory structures, *Neurology* **53**, 880–882.

Doty, R.L., Riklan, M., Deems, D.A., Reynolds, C., Stellar, S. (1989) The olfactory and cognitive deficits of Parkinson's disease: evidence for independence, *Ann. Neurol.* **25**, 166–171.

Doty, R.L., Stern, M.B., Pfeiffer, C., Gollomp, S.M., Hurtig, H.I. (1992) Bilateral olfactory dysfunction in early stage treated and untreated idiopathic Parkinson's disease, *J. Neurol. Neurosurg. Psychiatr.* **55**, 138–142.

Doty, R.L., Fernandez, A.D., Levine, M.A., Moses, A., McKeown, D.A. (1997) Olfactory dysfunction in type I pseudohypoparathyroidism: dissociation from $G_{s\alpha}$ protein deficiency, *J. Clin. Endocrinol. Metab.* **82**, 247–250.

Doty, R.L., Shaman, P., Applebaum, S.L., Giberson, R., Sikorsky, L., Rosenberg, L. (1984) Smell identification ability: changes with age, *Science* **226**, 1441–1443.

Drinnan, S.L., Hope, B.T., Snutch, T.P., Vincent, S.R. (1991) G_{olf} in the basal ganglia, *Mol. Cell. Neurosci.* **2**, 66–70.

Dulac, C., Axel, R. (1995) A novel family of genes encoding putative pheromone receptors in mammals, *Cell* **83**, 195–206.

Dwyer, N.D., Troemel, E.R., Sengupta, P., Bargmann, C.I. (1998) Odorant receptor localization to olfactory cilia is mediated by ODR-4, a novel membrane-associated protein, *Cell* **93**, 455–466.

Eggan, K., Baldwin, K., Tackett, M., Osborne, J., Gogos, J., Chess, A., Axel, R., Jaenisch, R. (2004) Mice cloned from olfactory sensory neurons, *Nature* **428**, 44–49.

Ehlers, A., Beck, S., Forbes, S.A., Trowsdale, J., Volz, A., Younger, R., Ziegler, A. (2000) MHC-linked olfactory receptor loci exhibit polymorphism and contribute to extended HLA-OR-haplotypes, *Genome Res.* **10**, 1968–1978.

Emes, R.D., Beatson, S.A., Ponting, C.P., Goodstadt, L. (2004) Evolution and comparative genomics of odorant- and pheromone-associated genes in rodents, *Genome Res.* **14**, 591–602.

Ennis, M., Zhou, F.M., Ciombor, K.J., Aroniadou-Anderjaska, V., Hayar, A., Borrelli, E., Zimmer, L.A., Margolis, F., Shipley, M.T. (2001) Dopamine D2 receptor-mediated presynaptic inhibition of olfactory nerve terminals, *J. Neurophysiol.* **86**, 2986–2997.

Fadool, D.A., Ache, B.W. (1992) Plasma membrane inositol 1,4,5-trisphosphate-activated channels mediate signal transduction in lobster olfactory receptor neurons, *Neuron* **9**, 907–918.

Fadool, D.A., Estey, S.J., Ache, B.W. (1995) Evidence that a Gq-protein mediates excitatory odor transduction in lobster olfactory receptor neurons, *Chem. Senses* **20**, 489–498.

Feinstein, P., Mombaerts, P. (2004) A contextual model for axonal sorting into glomeruli in the mouse olfactory system, *Cell* **117**, 817–831.

Feinstein, P., Bozza, T., Rodriguez, I., Vassalli, A., Mombaerts, P. (2004) Axon guidance of mouse olfactory sensory neurons by odorant receptors and the beta2 adrenergic receptor, *Cell* **117**, 833–846.

Fijimoto, T., Nakade, S., Miyawaki, A., Mikoshiba, K., Ogawa, K. (1992) Localization of inositol 1,4,5-trisphosphate receptor-like protein in plasmalemmal caveolae, *J. Cell Biol.* **119**, 1507–1513.

Firestein, S., Werblin, F.S. (1989) Odor-induced membrane currents in vertebrate olfactory receptor neurons, *Science* **244**, 79–82.

Firestein, S., Shepherd, G.M., Werblin, F.S. (1990) Time course of the membrane current underlying sensory transduction in salamander olfactory receptor neurones, *J. Physiol.* **430**, 135–158.

Firestein, S., Darrow, B., Shepherd, G.M. (1991) Activation of the sensory current in salamander olfactory receptor neurons depends on a G protein-mediated cAMP second messenger system, *Neuron* **6**, 825–835.

Floriano, W.B., Vaidehi, N., Goddard, W.A., Singer, M.S., Shepherd, G.M. (2000) Molecular mechanisms underlying differential odor responses of a mouse olfactory receptor, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10712–10716.

Frye, R., Schwartz, B., Doty, R.L. (1990) Dose-related effects of cigarette smoking on olfactory function, *JAMA* **263**, 1233–1236.

Fulle, H.J., Vassar, R., Foster, D.C., Yang, R.B., Axel, R., Garbers, D.L. (1995) A receptor guanylyl cyclase expressed specifically in olfactory sensory neurons, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3571–3575.

Gaillard, I., Rouquier, S., Chavanieu, A., Mollard, P., Giorgi, D. (2004) Amino-acid changes acquired during evolution by olfactory receptor 912–93 modify the specificity of odorant recognition, *Hum. Mol. Genet.* **13**, 771–780.

Garibotti, M., Navarrini, A., Pisanelli, A.M., Pelosi, P. (1997) Three odorant-binding proteins from rabbit nasal mucosa, *Chem. Senses* **22**, 383–390.

Gawel, M.J., Das, P., Vincent, S., Rose, F.C. (1981) Visual and auditory evoked responses in patients with Parkinson's disease, *J. Neurol. Neurosurg. Psychiatry* **44**, 227–232.

Gimelbrant, A.A., Stoss, T.D., Landers, T.M., McClintock, T.S. (1999) Truncation releases olfactory receptors from the endoplasmic reticulum of heterologous cells, *J. Neurochem.* **72**, 2301–2311.

Glusman, G., Clifton, S., Roe, B., Lancet, D. (1996) Sequence analysis in the olfactory receptor gene cluster on human chromosome 17: recombinatorial events affecting receptor diversity, *Genomics* **37**, 147–160.

Glusman, G., Yanai, I., Rubin, I., Lancet, D. (2001) The complete human olfactory subgenome, *Genome Res.* **11**, 685–702.

Godfrey, P.A., Mainic, B., Buck, L.B. (2004) The mouse olfactory receptor gene family, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2156–2161.

Goldstein, B.J., Schwob, J.E. (1996) Analysis of the globose basal cell compartment in rat olfactory epithelium using GBC-1, a new monoclonal antibody against globose basal cells, *J. Neurosci.* **16**, 4005–4016.

Goulding, E.H., Ngai, J., Kramer, R.H., Colicos, S., Axel, R., Siegelbaum, S.A., Chess, A. (1992) Molecular cloning and single-channel properties of the cyclic nucleotide-gated channel from catfish olfactory neurons, *Neuron* **8**, 45–58.

Graves, A.B., Bowen, J.D., Rajaram, L., McCormick, W.C., McCurry, S.M., Schellenberg, G.D., Larson, E.B. (1999) Impaired olfaction as a marker for cognitive decline: interaction with apolipoprotein E epsilon4 status, *Neurology* **53**, 1480–1487.

Graziadei, P.P.C. (1973) Cell dynamics in the olfactory mucosa, *Tissue Cell.* **5**, 113–131.

Graziadei, P.P.C., Metcalf, J.F. (1971) Autoradiographic and ultrastructural observation on the frog's olfactory mucosa, *Z Zellforsch Mikrosk Anat* **116**, 305–318.

Graziadei, P.P.C., Monti-Graziadei, G.A. (1979) Neurogenesis and neuron regeneration in the olfactory system of mammals, *J. Neurocytol.* **8**, 1–18.

Grus, W.E., Zhang, J. (2004) Rapid turnover and species-specificity of vomeronasal pheromone receptor genes in mice and rats, *Gene* **340**, 303–312.

Hack, I., Bancila, M., Loulier, K., Carroll, P., Cremer, H. (2002) Reelin is a detachment signal in tangential chain-migration during postnatal neurogenesis, *Nat. Neurosci.* **5**, 939–945.

Hamana, H., Hirono, J., Kizumi, M., Sato, T. (2003) Sensitivity-dependent hierarchical receptor codes for odors, *Chem. Senses* **28**, 87–104.

Hanoune, J., Defer, N. (2001) Regulation and role of adenylyl cyclase isoforms, *Annu. Rev. Pharmacol. Toxicol.* **41**, 145–174.

Hansel, D.E., Eipper, B.A., Ronnett, G.V. (2001) Neuropeptide Y functions as a neuroproliferative factor, *Nature* **410**, 940–944.

Hatt, H., Ache, B.W. (1994) Cyclic nucleotide- and inositol phosphate-gated ion channels in lobster olfactory receptor neurons, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6264–6268.

Hawkes, C.H., Shephard, B.C. (1993) Selective anosmia in Parkinson's disease? *Lancet* **341**, 435–436.

Holley, A., Duchamp, A., Revial, M.F., Juge, A. (1974) Qualitative and quantitative discrimination in the frog olfactory receptors: analysis from electrophysiological data, *Ann. N. Y. Acad. Sci.* **237**, 102–114.

Hoppe, R., Frank, H., Breer, H., Strotmann, J. (2003) The clustered olfactory receptor gene family 262: genomic organization, promoter elements, and interacting transcription factors, *Genome Res.* **13**, 2674–2685.

Hsu, P., Yu, F., Feron, F., Pickles, J.O., Sneesby, K., Mackay-Sim, A. (2001) Basic fibroblast growth factor and fibroblast growth factor receptors in adult olfactory epithelium, *Brain Res.* **896**, 188–197.

Huard, J.M., Youngentob, S.L., Goldstein, B.J., Luskin, M.B., Schwob, J.E. (1998) Adult

olfactory epithelium contains multipotent progenitors that give rise to neurons and non-neural cells, *J. Comp. Neurol.* **400**, 469–486.

Hudson, R., Distel, H. (1986) Pheromonal release of suckling in rabbits does not depend on the vomeronasal organ, *Physiol. Behav.* **37**, 123–128.

Iino, S., Kobayashi, S., Okazaki, K., Hidaka, H. (1995) Neurocalcin-immunoreactive receptor cells in the rat olfactory epithelium and vomeronasal organ, *Neurosci. Lett.* **191**, 91–94.

Ingi, T., Ronnett, G.V. (1995) Direct demonstration of a physiological role for carbon monoxide in olfactory receptor neurons, *J. Neurosci.* **15**, 8214–8222.

Ingi, T., Cheng, J., Ronnett, G.V. (1996) Carbon monoxide: an endogenous modulator of the nitric oxide-cyclic GMP signaling system, *Neuron* **16**, 835–842.

Ishii, T., Serizawa, S., Kohda, A., Nakatani, H., Shiroishi, T., Okumura, K., Iwakura, Y., Nagawa, F., Tsuboi, A., Sakano, H. (2001) Monoallelic expression of the odorant receptor gene and axonal projection of olfactory sensory neurones, *Genes Cells* **6**, 71–78.

Jaworsky, D.E., Matsuzaki, O., Borisy, F.F., Ronnett, G.V. (1995) Calcium modulates the rapid kinetics of the odorant-induced cyclic AMP signal in rat olfactory cilia, *J. Neurosci.* **15**, 310–318.

Jones, D.T., Reed, R.R. (1987) Molecular cloning of five GTP-binding protein cDNA species from rat olfactory neuroepithelium, *J. Biol. Chem.* **262**, 14241–14249.

Jones, D.T., Reed, R.R. (1989) G_{olf}: An olfactory neuron specific-G protein involved in odorant signal transduction, *Science* **244**, 790–795.

Juilfs, D.M., Fulle, H.J., Zhao, A.Z., Houslay, M.D., Garbers, D.L., Beavo, J.A. (1997) A subset of olfactory neurons that selectively express cGMP-stimulated phosphodiesterase (PDE2) and guanylyl cyclase-D define a unique olfactory signal transduction pathway, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3388–3395.

Kajiya, K., Inaki, K., Tanaka, M., Haga, T., Kataoka, H., Touhara, K. (2001) Molecular bases of odor discrimination: reconstitution of olfactory receptors that recognize overlapping sets of odorants, *J. Neurosci.* **21**, 6018–6025.

Kaneko, N., Debski, E.A., Wilson, M.C., Whitten, W.K. (1980) Puberty acceleration in mice: II. Evidence that the vomeronasal organ is a receptor for the primer pheromone in male mouse urine, *Biol. Reprod.* **22**, 873–878.

Katzenschlager, R., Zijlmans, J., Evans, A., Watt, H., Lees, A.J. (2005) Olfactory function distinguishes vascular parkinsonism from Parkinson's disease, *J. Neurol. Neurosurg. Psychiatry* **75**, 1749–1752.

Kevetter, G.A., Winans, S.S. (1981) Connections of the corticomedial amygdala in the golden hamster. I. Efferents of the "vomeronasal amygdala", *J. Comp. Neurol.* **197**, 81–98.

Kirk, J.M., Grant, D.B., Savage, M.O., Besser, G.M., Bouloux, P.M. (1994) Identification of olfactory dysfunction in carriers of X-linked Kallmann's syndrome, *Clin. Endocrinol.* **41**, 577–580.

Klar, A.J., Foqel, S., Lusnak, K., (1979) Gene conversion of the mating-type locus in *Saccharomyces cerevisiae*, *Genetics* **92**, 777–782.

Kleene, S.J. (1993) Origin of the chloride current in olfactory transduction, *Neuron* **11**, 123–132.

Kleene, S.J., Gesteland, R.C. (1991) Calcium-activated chloride conductance in frog olfactory cilia, *J. Neurosci.* **11**, 3624–3629.

Koistinaho, J., Miettinen, S., Keinanen, R., Vartiainen, N., Roivainen, R., Laitinen, J.T. (1996) Long-term induction of haem oxygenase-1 (HSP-32) in astrocytes and microglia following transient focal brain ischaemia in the rat, *Eur. J. Neurosci.* **8**, 2265–2272.

Komatsu, H., Mori, I., Rhee, J.S., Akaike, N., Ohshima, Y. (1996) Mutations in a cyclic nucleotide-gated channel lead to abnormal thermosensation and chemosensation in *C. elegans*, *Neuron* **17**, 707–718.

Koshimoto, H., Katoh, K., Yoshihara, Y., Nemoto, Y., Mori, K. (1994) Immunohistochemical demonstration of embryonic expression of an odor receptor protein and its zonal distribution in the rat olfactory epithelium, *Neurosci. Lett.* **169**, 73–76.

Krautwurst, D., Yau, K.W., Reed, R.R. (1998) Identification of ligands for olfactory receptors by functional expression of a receptor library, *Cell* **95**, 917–926.

Kurahashi, T., Yau, K.-W. (1993) Co-existence of cationic and chloride components in odorant-induced current of vertebrate olfactory receptor cells, *Nature* **363**, 71–74.

Kurahashi, T., Lowe, G., Gold, G.H. (1994) Suppression of odorant responses by odorants in olfactory receptor cells, *Science* **265**, 118–120.

Lane, R.P., Young, J., Newman, T., Trask, B.J. (2004) Species specificity in rodent pheromone receptor repertoires, *Genome Res.* **14**, 603–608.

Lane, R.P., Cutforth, T., Axel, R., Hood, L., Trask, B.J. (2002) Sequence analysis of mouse vomeronasal receptor gene clusters reveals common promoter motifs and a history of recent expansion, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 291–296.

Lazard, D., Zupko, K., Poria, Y., Nef, P., Lazarovits, J., Horn, S., Khen, M., Lancet, D. (1991) Odorant signal termination by olfactory UDP-glucuronosyl transferase, *Nature* **349**, 790–793.

Leinders-Zufall, T., Shepherd, G.M., Zufall, Z. (1996) Modulation by cyclic GMP of the odour sensitivity of vertebrate olfactory receptor cells, *Proc. R. Soc. Biol. Lond., B.,* **263**, 803–811.

Leypold, B.G., Yu, C.R., Leinders-Zufall, T., Kim, M.M., Zufall, F., Axel, R. (2002) Altered sexual and social behaviors in trp2 mutant mice, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6376–6381.

Li, J., Ishii, T., Feinstein, P., Mombaerts, P. (2004) Odorant receptor gene choice is reset by nuclear transfer from mouse olfactory sensory neurons, *Nature* **428**, 393–399.

Liman, E.R., Buck, L.B. (1994) A second subunit of the olfactory cyclic nucleotide-gated channel confers high sensitivity to cAMP, *Neuron* **13**, 611–621.

Lobel, D., Marchese, S., Krieger, J., Pelosi, P., Breer, H. (1998) Subtypes of odorant-binding proteins–heterologous expression and ligand binding, *Eur. J. Biochem.* **254**, 318–324.

Lowe, G., Gold, G.H. (1993a) Contribution of the ciliary cyclic nucleotide-gated conductance to olfactory transduction in the salamander, *J. Physiol.* **462**, 175–196.

Lowe, G., Gold, G.H. (1993b) Nonlinear amplification by calcium-dependent chloride channels in olfactory receptor cells, *Nature* **366**, 283–286.

Lowe, G., Nakamura, T., Gold, G.H. (1989) Adenylate cyclase mediates olfactory transduction for a wide variety of odorants, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5641–5645.

Luskin, M.B. (1993) Restricted proliferation and migration of postnatally generated neurons derived from the forebrain subventricular zone, *Neuron* **11**, 173–189.

Malnic, B., Hirono, J., Sato, T., Buck, L.B. (1999) Combinatorial receptor codes for odors, *Cell* **96**, 713–723.

Markopoulou, K., Larsen, K.W., Wszolek, E.K., Denson, M.A., Lang, A.E., Pfeiffer, R.F., Wszolek, Z.K. (1997) Olfactory dysfunction in familial parkinsonism, *Neurology* **49**, 1262–1267.

Matsunami, H., Buck, L.B. (1997) A multigene family encoding a diverse array of putative pheromone receptors in mammals, *Cell* **90**, 775–784.

McClintock, T.S., Sammeta, N. (2003) Trafficking prerogatives of olfactory receptors, *Neuroreport* **14**, 1547–1552.

McKeown, D., Doty, R.L., Perl, D.P., Frye, R.E., Sims, I., Mester, A.F. (1996) Olfactory dysfunction in Down's syndrome, *J. Neurol. Neurosurg. Psychiatry* **61**, 412–416.

Meisami, E., Bhatnagar, K.P. (1998) Structure and diversity in mammalian accessory olfactory bulb, *Microsc. Res. Tech.* **43**, 476–499.

Menco, B.P. (1997) Ultrastructural aspects of olfactory signaling, *Chem. Senses* **22**, 295–311.

Menco, B.P., Jackson, J.E. (1997a) A banded topography in the developing rat's olfactory epithelial surface, *J. Comp. Neurol.* **388**, 293–306.

Menco, B.P., Jackson, J.E. (1997b) Cells resembling hair cells in developing rat olfactory and nasal respiratory epithelia, *Tissue Cell* **29**, 707–713.

Menco, B.P., Cunningham, A.M., Qasba, P., Levy, N., Reed, R.R. (1997) Putative odour receptors localize in cilia of olfactory receptor cells in rat and mouse: a freeze-substitution ultrastructural study, *J. Neurocytol.* **26**, 297–312.

Meredith, M., Marques, D.M., O'Connell, R.O., Stern, F.L. (1980) Vomeronasal pump: significance for male hamster sexual behavior, *Science* **207**, 1224–1226.

Meyer, M.R., Angele, A., Kremmer, E., Kaupp, U.B., Muller, F. (2000) A cGMP-signaling pathway in a subset of olfactory sensory neurons, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10595–10600.

Mombaerts, P. (2001) How smell develops, *Nat. Neurosci.* **4**(Suppl.), 1192–1198.

Mombaerts, P. (2004) Odorant receptor gene choice in olfactory sensory neurons: the one receptor-one neuron hypothesis revisited, *Curr. Opin. Neurol.* **14**, 31–36.

Mombaerts, P., Wang, F., Dulac, C., Chao, S.K., Nemes, A., Mendelsohn, M., Edmondson, J., Axel, R. (1996) Visualizing an olfactory sensory map, *Cell* **87**, 675–686.

Montgomery, E.B. Jr., Lyons, K., Koller, W.C. (2000) Early detection of probable idiopathic Parkinson's disease: II. A prospective application of a diagnostic test battery, *Mov. Disord.* **15**, 474–478.

Montgomery, E.B. Jr., Baker, K.B., Lyons, K., Koller, W.C. (1999) Abnormal performance on the PD test battery by asymptomatic first-degree relatives, *Neurology* **52**, 757–762.

Moon, C., Sung, Y., Reddy, R., Ronnett, G.V. (1999) Odorants induce the phosphorylation of the cAMP response element binding protein in olfactory receptor neurons, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14605–14610.

Moon, C., Jaberi, P., Otto-Bruc, A., Baehr, W., Palczewski, K., Ronnett, G.V. (1998) Calcium-sensitive particulate guanylyl cyclase as a modulator of cAMP in olfactory receptor neurons, *J. Neurosci.* **18**, 3195–3205.

Munger, S.D., Gleeson, R.A., Aldrich, H.C., Rust, N.C., Ache, B.W., Greenberg, R.M. (2000) Characterization of a phosphoinositide-mediated odor transduction pathway reveals plasma membrane localization of an inositol 1,4,5-trisphosphate receptor in lobster olfactory receptor neurons, *J. Biol. Chem.* **275**, 20450–20457.

Murphy, C., Jinich, S. (1996) Olfactory dysfunction in Down's Syndrome, *Neurobiol. Aging* **17**, 631–637.

Nakamura, T., Gold, G.H. (1987) A cyclic nucleotide-gated conductance in olfactory receptor cilia, *Nature* **325**, 442–444.

Navarro, M., Gull, K. (2001) A pol I transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*, *Nature* **414**, 759–763.

Negus, V. (1958) *The Comparative Anatomy and Physiology of the Nose and Paranasal Sinuses*, E & S Livingstone, Edinburgh.

Oka, Y., Nakamura, A., Watanabe, H., Touhara, K. (2004) An odorant derivative as an antagonist for an olfactory receptor, *Chem. Senses* **29**, 815–822.

Ottoson, D. (1956) Analysis of the electrical activity of the olfactory epithelium, *Acta Physiol. Scand.* **122**, 1–83.

Pace, U., Lancet, D. (1986) Olfactory GTP-binding protein: signal-transducing polypeptide of vertebrate chemosensory neurons, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4947–4951.

Pace, U., Hanski, E., Salomon, Y., Lancet, D. (1985) Odorant-sensitive adenylate cyclase may mediate olfactory reception, *Nature* **316**, 255–258.

Parkinson, J. (1817) *An Essay on the Shaking Palsy*, Sherwood, Neely and Jones, London.

Peppel, K., Boekhoff, I., McDonald, P., Breer, H., Caron, M.G., Lefkowitz, R.J. (1997) G protein-coupled receptor kinase 3 (GRK3) gene disruption leads to loss of odorant receptor desensitization, *J. Biol. Chem.* **272**, 25425–25428.

Pes, D., Pelosi, P. (1995) Odorant-binding proteins of the mouse, *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **112**, 471–479.

Pevsner, J., Sklar, P.B., Snyder, S.H. (1986) Odorant-binding protein: localization to nasal glands and secretions, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4942–4946.

Pevsner, J., Trifiletti, R.R., Strittmatter, S.M., Snyder, S.H. (1985) Isolation and characterization of an olfactory receptor protein for odorant pyrazines, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3050–3054.

Pevsner, J., Hou, V., Snowman, A.M., Snyder, S.H. (1990) Odorant-binding protein. Characterization of ligand binding, *J. Biol. Chem.* **265**, 6118–6125.

Pilpel, Y., Lancet, D. (1999) The variable and conserved interfaces of modeled olfactory receptor proteins, *Protein Sci.* **8**, 969–977.

Potter, S.M., Zheng, C., Koos, D.S., Feinstein, P., Fraser, S.E., Mombaerts, P. (2001) Structure and emergence of specific olfactory glomeruli in the mouse, *J. Neurosci.* **21**, 9713–9723.

Pyrski, M., Xu, Z., Walters, E., Gilbert, D.J., Jenkins, N.A., Copeland, N.G., Margolis, F.L. (2001) The *OMP-lacZ* transgene mimics the unusual expression pattern of OR-Z6, a new odorant receptor gene on mouse chromosome 6: implication for locus-dependent gene expression, *J. Neurosci.* **21**, 4637–4648.

Qasba, P., Reed, R.R. (1998) Tissue and zonal-specific expression of an olfactory receptor transgene, *J. Neurosci.* **18**, 227–236.

Reed, R.R. (2004) After the Holy Grail: establishing a molecular basis for mammalian olfaction, *Cell* **116**, 329–336.

Ressier, K.J., Sullivan, S.L., Buck, L.B. (1993) A zonal organization of odorant receptor gene expression in the olfactory epithelium, *Cell* **73**, 597–609.

Ressler, K.J., Sullivan, S.L., Buck, L.B. (1994) Information coding in the olfactory system: evidence for a stereotyped and highly

organized epitope map in the olfactory bulb, *Cell* **79**, 1245–1255.

Restrepo, D., Okada, Y., Teeter, J.H., Lowry, L.D., Cowart, B., Brand, J.G. (1993) Human olfactory neurons respond to odor stimuli with an increase in cytoplasmic Ca^{2+}, *Biophys. J.* **64**, 1961–1966.

Rhein, L.D., Cagan, R.H. (1980) Biochemical studies of olfaction: Isolation, characterization, and odorant binding activity of cilia from rainbow trout olfactory rosettes, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4412–4416.

Rhein, L.D., Cagan, R.H. (1983) Biochemical studies of olfaction: binding specificity of odorants to a cilia preparation from rainbow trout olfactory rosettes, *J. Neurochem.* **41**, 569–577.

Rodnitzky, R.L. (1998) Visual dysfunction in Parkinson's disease, *Clin. Neurosci.* **5**, 102–106.

Rodriguez, I., Feinstein, P., Mombaerts, P. (1999) Variable patterns of axonal projections of sensory neurons in the mouse vomeronasal system, *Cell* **97**, 199–208.

Rodriguez, I., Del Punta, K., Rothman, A., Ishii, T., Mombaerts, P. (2002) Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors, *Nat. Neurosci.* **5**, 134–140.

Ronnett, G.V., Parfitt, D.J., Hester, L.D., Snyder, S.H. (1991) Odorant-sensitive adenylate cyclase: rapid potent activation and desensitization in primary olfactory neuronal cultures, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2366–2369.

Ronnett, G.V., Cho, H., Hester, L.D., Wood, S.F., Snyder, S.H. (1993) Odorants differentially enhance phosphoinositide turnover and adenylyl cyclase in olfactory receptor neuronal cultures, *J. Neurosci.* **13**, 1751–1758.

Roskams, A.J.I., Cai, X., Ronnett, G.V. (1998) Expression of neuron-specific beta-III tubulin during olfactory neurogenesis in the embryonic and adult rat, *Neuroscience* **83**, 191–200.

Roskams, A.J., Bethel, M.A., Hurt, K.J., Ronnett, G.V. (1996) Sequential expression of Trks A, B, and C in the regenerating olfactory neuroepithelium, *J. Neurosci.* **16**, 1294–1307.

Royal, S.J., Key, B. (1999) Development of P2 olfactory glomeruli in P2-internal ribosome entry site-*Tau-Lac*Z transgenic mice, *J. Neurosci.* **19**, 9856–9864.

Ryba, N.J., Tirindelli, R. (1995) A novel GTP-binding protein gamma-subunit, G gamma 8, is expressed during neurogenesis in the olfactory and vomeronasal neuroepithelia, *J. Biol. Chem.* **270**, 6757–6767.

Sajjadian, A., Doty, R.L., Gutnick, D.N., Shirugi, R.J., Sivak, M., Perl, D. (1994) Olfactory dysfunction in amyotrophic lateral sclerosis, *Neurodegeneration* **3**, 1–5.

Sam, M., Vora, S., Malnic, B., Ma, W., Novotny, M.V., Buck, L.B. (2001) Neuropharmacology: odorants may arouse instinctive behaviours, *Nature* **412**, 142.

Schandar, M., Laugwitz, K.L., Boekhoff, I., Kroner, C., Gudermann, T., Schultz, G., Breer, H. (1998) Odorants selectively activate distinct G protein subtypes in olfactory cilia, *J. Biol. Chem.* **273**, 16669–16677.

Schleicher, S., Boekoff, I., Arriza, J., Lefkowitz, R.J., Breer, H. (1993) A β-adrenergic receptor kinase-like enzyme is involved in olfactory signal termination, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1420–1424.

Schultze, M.J.S. (1856) Über die Endigungsweise des Geruchsnerven und die Epithelialgebilde der Nasenschleimhaut, *Verhandl. Akad. Wissensch. Berlin* **21**, 504–514.

Schultze, M.J.S. (1863) Untersuchungen Über den Bau der Nasenschleimhaut, namentlich die Structur und Endigungsweise der Geruchsnerven bei dem Menschen und den Wirbeltieren, *Abhandl. d. Nat. Ges. zu Halle* **7**, 1–99.

Schwob, J.E. (2002) Neural regeneration and the peripheral olfactory system, *Anat. Rec.* **269**, 33–49.

Scott, J.W., Shannon, D.E., Charpentier, J., Davis, L.M., Kaplan, C. (1997) Spatially organized response zones in rat olfactory epithelium, *J. Neurophysiol.* **77**, 1950–1962.

Segovia, S., Guillamón, A. (1982) Effects of sex steroids on the development of the vomeronasal organ in the rat, *Dev. Brain Res.* **14**, 209–291.

Sengupta, P., Chou, J.H., Bargmann, C.I. (1996) odr-10 encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl, *Cell* **84**, 899–909.

Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., Sakano, H. (2003) Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse, *Science* **302**, 2088–2094.

Serizawa, S., Ishii, T., Nakatani, H., Tsuboi, A., Nagawa, F., Asano, M., Sudo, K., Sakagami, J., Sakano, H., Ijiri, T., Matsuda, Y., Suzuki, M., Yamanori, T., Iwakura, Y., Sakano, H. (2000) Mutually exclusive expression of odorant receptor transgenes, *Nat. Neurosci.* **3**, 687–693.

Sharon, D., Glusman, G., Pilpel, Y., Khen, M., Gruetzner, F., Haaf, T., Lancet, D. (1999) Primate evolution of an olfactory receptor cluster: diversification by gene conversion and recent emergence of pseudogenes, *Genomics* **61**, 24–36.

Shykind, B.M., Rohani, C., O'Donnell, S., Nemes, A., Mendelsohn, M., Sun, Y., Axel, R., Barnea, G. (2004) Gene switching and the stability of odorant receptor gene choice, *Cell* **117**, 801–815.

Simpson, P.J., Wang, E., Moon, C., Matarazzo, V., Cohen, D.R., Liebl, D.J., Ronnett, G.V. (2003) Neurotrophin-3 signaling maintains maturational homeostasis between neuronal populations in the olfactory epithelium, *Mol. Cell. Neurosci.* **24**, 858–874.

Sklar, P.B., Anholt, R.R., Snyder, S.H. (1986) The odorant-sensitive adenylate cyclase of olfactory receptor cells. Differential stimulation by distinct classes of odorants, *J. Biol. Chem.* **261**, 15538–15543.

Smith, T.D., Siegel, M.I., Bonar, C.J., Bhatnagar, K.P., Mooney, M.P., Burrows, A.M., Smith, M.A., Maico, L.M. (2001) The existence of the vomeronasal organ in postnatal chimpanzees and evidence for its homology with that of humans, *J. Anat.* **198**, 1–82.

Soria, J.M., Taglialatela, P., Gil-Perotin, S., Galli, R., Gritti, A., Verdugo, J.M.G., Bertuzzi, S. (2004) Defective postnatal neurogenesis and disorganization of the rostral migratory stream in absence of the *Vax*1 homeobox gene, *J. Neurosci.* **24**, 11171–11181.

Spehr, M., Wetzel, C.H., Hatt, H., Ache, B.W. (2002) 3-Phosphoinositides modulate cyclic nucleotide signaling in olfactory receptor neurons, *Neuron* **33**, 731–739.

Stoltzfus, J.R., Horton, W.J., Grotewiel, M.S. (2003) Odor-guided behavior in *Drosophila* requires calreticulin, *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* **189**, 471–483.

Stowers, L., Holy, T.E., Meister, M., Dulac, C., Koentges, G. (2002) Loss of sex discrimination and male-male aggression in mice deficient for trp2, *Science* **295**, 1493–1500.

Takami, S., Getchell, M.L., Chen, Y., Monti-Bloch, L., Berliner, D.L., Stensaas, L.J., Getchell, T.V. (1993) Vomeronasal epithelial cells of the adult human express neuron-specific molecules, *Neuroreport* **4**, 375–378.

Touhara, K., Sengoku, S., Inaki, K., Tsuboi, A., Hirono, J., Sato, T., Sakano, H., Haga, T. (2001) Functional identification and reconstitution of an odorant receptor in single olfactory neurons, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4040–4045.

Trask, B.J., Friedman, C., Martin-Gallardo, A., Rowen, L., Akinbami, C., Blankenship, J., Collins, C., Giorgi, D., Iadonato, S., Johnson, F., Kuo, W.-L., Massa, H., Morrish, T., Naylor, S., Nguyen, O.T.-H., Rouquier, S., Smith, T., Wong, D.J., Youngblom, J., van den Engh, G. (1998) Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes, *Hum. Mol. Genet.* **7**, 13–26.

Treloar, H.B., Feinstein, P., Mombaerts, P., Greer, C.A. (2002) Specificity of glomerular targeting by olfactory sensory axons, *J. Neurosci.* **22**, 2469–2477.

Trotier, D. (1986) A patch-clamp analysis of membrane currents in salamander olfactory receptor cells, *Pflügers Arch.* **407**, 589–595.

Trotier, D., Eloit, C., Wassef, M., Telmain, G., Bensimon, J.L., Doving, K.B., Ferrand, J. (2000) The vomeronasal cavity in adult humans, *Chem. Senses* **25**, 369–380.

Turin, L. (2002) A method for the calculation of odor character from molecular structure, *J. Theor. Biol.* **216**, 367–385.

Vassalli, A., Rothman, A., Feinstein, P., Zapotocky, M., Mombaerts, P. (2002) Minigenes impart odorant receptor-specific axon guidance in the olfactory bulb, *Neuron* **35**, 681–696.

Vassar, R., Ngai, J., Axel, R. (1993) Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium, *Cell* **74**, 309–318.

Vassar, R., Chao, S.K., Sitcheran, R., Nunez, J.M., Vosshall, L.B., Axel, R. (1994) Topographic organization of sensory projections to the olfactory bulb, *Cell* **79**, 981–991.

Wang, F., Nemes, A., Mendelsohn, M., Axel, R. (1998) Odorant receptors govern the formation of a precise topographic map, *Cell* **93**, 47–60.

Weiler, E., Apfelbach, R., Farbman, A.I. (1999) The vomeronasal organ of the male ferret, *Chem. Senses* **24**, 127–136.

Wong, S.T., Trinh, K., Hacker, B., Chan, G.C., Lowe, G., Gaggar, A., Xia, Z., Gold, G.H., Storm, D.R. (2000) Disruption of the type III adenylyl cyclase gene leads to peripheral and behavioral anosmia in transgenic mice, *Neuron* **27**, 487–497.

Wood, S.F., Ronnett, G.V., Snyder, S.H. (1990) Activation of inositol-phosphate metabolism in primary olfaction cell cultures, *Chem. Senses* **15**, 252.

Wu, A.Y., Tang, X.B., Martinez, S.E., Ikeda, K., Beavo, J.A. (2004) Molecular determinants for cyclic nucleotide binding to the regulatory domains of phosphodiesterase 2A, *J. Biol. Chem.* **279**, 37928–37938.

Wysocki, C.J., Beauchamp, G.K. (1984) Ability to smell androstenone is genetically determined, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4899–4902.

Yan, C., Zhao, A.Z., Bentley, J.K., Loughney, K., Ferguson, K., Beavo, J.A. (1995) Molecular cloning and characterization of a calmodulin-dependent phosphodiesterase enriched in olfactory sensory neurons, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9677–9681.

Young, J.M., Trask, B.J. (2002) The sense of smell: genomics of vertebrate odorant receptors, *Hum. Mol. Genet.* **11**, 1153–1160.

Young, J.M., Kambere, M., Trask, B.J., Lane, R.P. (2005) Divergent V1R repertoires in five species: amplification in rodents, decimation in primates, and a surprisingly small repertoire in dogs, *Genome Res.* **15**, 231–240.

Young, J.M., Friedman, C., Williams, E.M., Ross, J.A., Tonnes-Priddy, L., Trask, B.J. (2002) Different evolutionary processes shaped the mouse and human olfactory receptor gene families, *Hum. Mol. Genet.* **11**, 535–546.

Young, J.M., Shykind, B.M., Lane, R.P., Tonnes-Priddy, L., Ross, J.A., Walker, M., Williams, E.M., Trask, B.J. (2003) Odorant receptor expressed sequence tags demonstrate olfactory expression of over 400 genes, extensive alternative splicing and unequal expression levels, *Genome Biol.* **4**, R71.

Zhang, X., Firestein, S. (2002) The olfactory receptor gene superfamily of the mouse, *Nat. Neurosci.* **5**, 124–133.

Zhang, J., Webb, D.M. (2003) Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8337–8341.

Zhang, Y., Chou, J.H., Bradley, J., Bargmann, C.I., Zinn, K. (1997) The *Caenorhabditis elegans* seven-transmembrane protein ODR-10 functions as an odorant receptor in mammalian cells, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12162–12167.

Zhao, H., Ivic, L., Otaki, J.M., Hashimoto, M., Mikoshiba, K., Firestein, S. (1998a) Functional expression of a mammalian odorant receptor, *Science* **279**, 237–242.

Zhao, H., Ivic, L., Otaki, J.M., Hashimoto, M., Mikoshiba, K., Firestein, S. (1998b) Functional expression of a mammalian odorant receptor, *Science* **279**, 237–241.

Zou, D.-J., Feinstein, P., Rivers, A.L., Mathews, G.A., Kim, A., Greer, C.A., Mombaerts, P., Firestein, S. (2004) Postnatal refinement of peripheral olfactory projections, *Science* **304**, 1976–1979.

Zufall, F., Leinders-Zufall, T. (1997) Identification of a long-lasting form of odor adaptation that depends on the carbon Monoxide/cGMP second-messenger system, *J. Neurosci.* **17**, 2703–2712.

Separating and Purifying Proteins: *see* Protein Purification

Sequence Analysis: *see* Nucleic Acid and Protein Sequence Analysis and Bioinformatics

Serial Analysis of Gene Expression

Jacques Marti[1] *and Jean-Marc Elalouf*[2]
[1] *Institut de Génétique Humaine, Montpellier, France*
[2] *Commissariat à l'Energie Atomique/Saclay, France*

1 Methodology 709
1.1 Rationale for SAGE 709
1.2 How SAGE Libraries are Built: The Original Method and Further Improvements 709
1.3 Processing Sequence Data 714
1.4 Statistical Analysis of Digital Transcriptomes 714
1.5 Tag-to-gene Mapping 715

2 Making Sense of SAGE Data 717
2.1 Unmatched Tags and Novel Transcripts: Using SAGE to Annotate the Genome and the Transcriptome 717
2.2 From SAGE to Functional Genomics: Linking High-throughput Expression Analysis, Studies on Genome Organization and Complex Biological Networks 718
2.3 SAGE, Molecular Classification of Cell Types and Cell Physiology 719
2.4 SAGE and Disease-induced Alterations of the Transcriptome 719

3 Perspectives 720

Bibliography 721
Books and Reviews 721
Primary Literature 722

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 12
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30649-8

Keywords

Expressed Sequence Tag (EST)
A partial, usually 200–800 nucleotides long fragment of cDNA, obtained by single-pass sequencing of a randomly selected clone isolated from a cell or tissue cDNA library.

Microarray
A glass slide that contains hundreds to several thousands ordered cDNA probes to measure gene expression levels by hybridization analysis of labeled cDNAs.

Polymerase Chain Reaction
Referred to as PCR, it consists of an *in vitro* enzymatic amplification of specific DNA sequences.

Single Nucleotide Polymorphism
A DNA sequence variation that occurs when a single nucleotide in the genome sequence is changed to another one.

Transcriptome
The molecular phenotype of a cell, as defined by the diversity and abundance of mRNAs, differs between normal, developmental, and disease conditions.

Type IIS Restriction Endonuclease
A restriction enzyme that cleaves at a defined distance away from its asymmetric recognition site.

Serial Analysis of Gene Expression (SAGE) is a method designed to measure the abundance of a large number of mRNAs in tissues and cells to characterize their transcriptome. By the end of the twentieth century, when determining the entire genomic sequence of higher organisms began to be perceived as a realistic and attainable goal, the interest for genome-wide analysis of gene expression rapidly increased. One of the first strategies was to generate expressed sequence tags (ESTs), that is, partial sequences of cDNA clones randomly selected in libraries prepared from various tissue samples. The EST approach was highly valuable for gene discovery. Moreover, as long as the abundance of a given mRNA could be correlated with the frequency of its EST, it might also provide quantitative information. However, this approach evaluated only a limited number of genes per library. Obviously, getting larger quantitative expression profiles required to increase the rate of acquisition of experimental data. SAGE rests on the sequencing of short diagnostic tags, and as such increased by more than 1 order of magnitude the number of transcripts analyzed by current sequencing methods. Sets of 100 000 SAGE tags are now routinely sequenced, making possible the quantification of low copy number mRNAs.

1 Methodology

1.1 Rationale for SAGE

A major principle of serial analysis of gene expression (SAGE) was to reduce expressed sequence tags (ESTs) to their minimal size by using a type IIS restriction enzyme (Fig. 1). Type IIS endonucleases cleave DNA at a defined distance from their asymmetric recognition site. *Bsm* FI was chosen to generate the conventional SAGE 14-bp tags from a defined (anchoring) site created by a conventional restriction enzyme near the 3′ end of each cDNA. Following the four common nucleotides of this anchoring site, SAGE tags differ by the next 10 variable nucleotides. Associated with the positional information, this size is usually sufficient for identifying each transcript, considering that 1 048 576 (4^{10}) different sequences can be distinguished. However, it may not be sufficient to map tags directly onto the complex genome of higher vertebrates: another endonuclease, *Mme* I, generating 21-bp tags, was preferred when using SAGE to annotate the genome by the so-called LongSAGE method.

One of the most appealing features of SAGE was the strategy used to amplify the whole population of tags while preserving the same quantitative distribution as in the initial mRNA population. Another characteristic of SAGE was to assemble tags into concatemers before DNA sequence analysis so that multiple tags (20–30 or more) could be read from each sequencing lane. Specific computer tools were developed for extracting the list of tags from sequence files, counting the number of times each one is observed and identifying the corresponding genes.

A SAGE analysis eventually results in a large sequence file in which tag sequences are punctuated by the four nucleotides of each anchoring site. Because experimental datasets are obtained *de novo* and do not rely on the calibration of any physical instrument, revocable external data or historical context, SAGE data can be considered as definitively acquired. Absolute transcript numbers are directly provided in a digital format, and data from multiple laboratories are easily assembled in comprehensive databases (Table 1). Tag-to-gene assignment can be updated at any time to keep in pace with progress in genome annotation, and an increasing body of functional annotations can be registered in the database. SAGE has now become one of the leading methods for gene expression profiling.

1.2 How SAGE Libraries are Built: The Original Method and Further Improvements

Poly(A) RNAs are preferably isolated using oligo$(dT)_{25}$ covalently linked to paramagnetic beads (Fig. 1). Other methods such as oligo(dT)-cellulose chromatography provide highly variable yields of mRNAs contaminated by substantial amounts of ribosomal RNAs (see Sect. 1.5). Thus, it is not surprising that the very first SAGE protocol, described in 1995, required 5 μg poly(A) RNA (i.e. several millions of mammalian cells). The need to work on small biological samples prompted several investigators to adapt the method to downsized extracts as in the SADE, microSAGE, and miniSAGE protocols. It then became possible to build SAGE libraries starting from 1 μg total RNA and to analyze biological samples containing 10^5 cells or even less. It is still possible to lower sample size and to start from 2500 cells or as little as

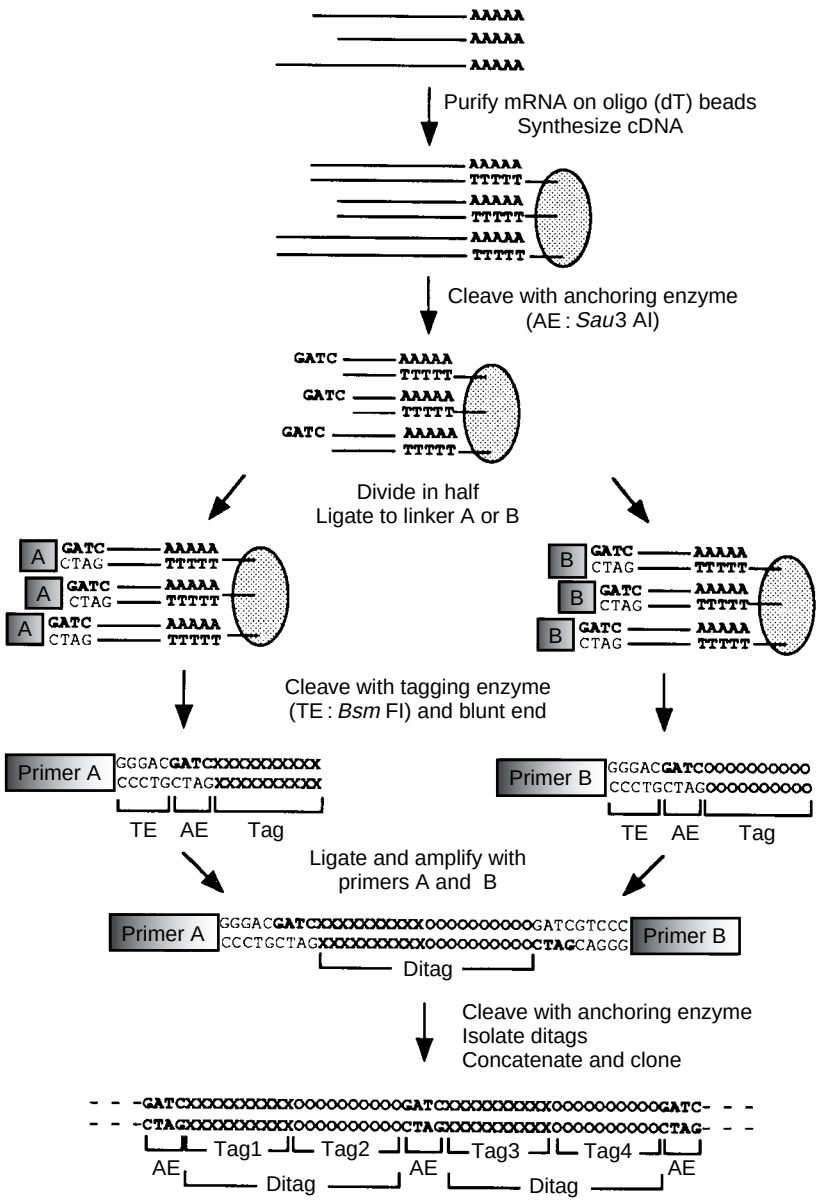

Fig. 1 Outline of procedures for constructing SAGE libraries. Poly(A) RNAs are isolated from tissue lysate using oligo$(dT)_{25}$ covalently linked to paramagnetic beads, and cDNA is synthesized under solid-phase condition. Bold face characters correspond to biologically relevant sequences, whereas light characters represent linker-derived sequences. The anchoring enzyme (AE) is *Sau* 3AI, whereas the tagging enzyme (TE) is *Bsm* FI. See text for details (Reprinted with permission from Cheval, L., Virlon, B., Elalouf, J.-M. (2000) SADE: A Microassay for Serial Analysis of Gene Expression, in: Hunt, S.P., Livesey, F.J. (Eds.) *Functional Genomics: A Practical Approach*, Oxford University Press, Oxford).

Tab. 1 Example of SAGE output (reproduced by courtesy of R. Quere and G. Bertrand, PhD theses, Montpellier University). Comparison between the expression profiles of (**A**) proliferating acute promyelocytic leukemia cells (NB4 human cell line), (**B**) cells incubated for 24 h in the presence of 1 μM retinoic (inducing cell growth arrest and granulocytic differentiation), and (**C**) normal blood granulocytes. SAGE was performed, using *Sau*3AI as anchoring enzyme. Comparison with (**D**) the sum of tag counts for the same genes in a compilation of 260 published SAGE libraries, built from various human tissues using *Nla* III as anchoring enzyme was also performed. These 36 entries represent about 0.7% of the total number of genes (>5000) revealed in this experiment. Examples of ambiguities in tag assignment are provided by *B2M* and *CAPN10*, sharing the same *Sau* 3AI-anchored tag but distinguished with *Nla* III, and by *COX7C* identified only with *Sau* 3AI since its transcript is devoid of *Nla* III restriction site. The upper rows illustrate the detection of differentiation markers. The lowest rows show the case of components of the translation machinery, ribosomal proteins, and elongation factor transcripts, abundant in leukemia cells, but barely detectable in terminally differentiated normal cells.

Tag (Sau 3AI)	*A*	*B*	*C*	*Tag (Nla III)*	*D*	*Gene name/ symbol/ accession number*
GATCATGCTCCATC	0	2	27	CATGCTCCATCCAG	359	Colony stimulating factor 3 receptor (granulocyte)/ CSF3R/NM_156039
GATCATCTTTCCTG	0	6	55	CATGGTGCGCTGAG	3923	Major histocompatibility complex, class I, C/HLA-C/NM_002117
GATCACATAAAACA	0	2	131	CATGGCCTTAACAA	1040	Pre-B-cell colony-enhancing factor/PBEF/NM_005746
GATCACAGTGGCCA	0	7	551	CATGGTGGCCACGG	5094	S100 calcium binding protein A9 (calgranulin B)/S100A9/ NM_002965
GATCCAGGTTTTTA	0	8	73	CATGACTCAGCCCG	750	Tumor necrosis factor, α-induced protein 2/TNFAIP2/NM_006291
GATCTTCTTTATAA	1	9	251	CATGGTTGTGGTTA	33 742	β-2-microglobulin/B2M/NM_004048
GATCTTCTTTATAA	1	9	251	CATGAGCGCTGATG	543	Calpain 10/CAPN10/NM_023087
GATCAAAGTTGGCG	1	5	27	CATGTTAGGGAGGA	233	Intercellular adhesion molecule3/ICAM3/NM_002162
GATCCACAAGTCCT	1	3	28	CATGTGGAAGCACT	2464	Interleukin 8/IL8/NM_000584
GATCCATGGATGTA	1	7	117	CATGGATGTAGTGG	273	Myeloid cell nuclear differentiation antigen/ MNDA/NM_002432
GATCACACTGACCT	2	20	38	CATGCTGACCTGTG	5927	Major histocompatibility complex, class I, B/HLA-B/NM_005514
GATCCTGTACAGCC	2	19	18	CATGGTGCTGAATG	14 044	Myosin, light polypeptide 6, alkali,../MYL6/NM_079425
GATCTTCATGCCCT	3	10	74	CATGCCCTGGGTTC	16 750	Ferritin, light polypeptide/ FTL/NM_000146
GATCACCTGGGTTT	3	3	28	CATGTCACCTTAGG	2424	Integral membrane protein 2B/ ITM2B/NM_021999
GATCTCTGGGCTGG	6	4	32	CATGACGATGGCCG	319	Glia maturation factor, γ/GMFG/NM_004877

(*continued overleaf*)

Tab. 1 (*Continued*)

Tag (Sau 3AI)	*A*	*B*	*C*	*Tag (Nla III)*	*D*	*Gene name/ symbol/ accession number*
GATCAACAAGAAAT	8	17	62	CATGGGCTGGGGGC	12 427	Profilin 1/ PFN1/NM_005022
GATCCAAGTGTTTA	9	10	49	CATGTTAACCCTCT	1877	H3 histone, family 3B (H3.3B)/H3F3B/NM_005324
GATCAGGAATCTTG	9	2	0	CATGTGAATCTGGG	1222	SET translocation (myeloid leukemia-associated)/SET/NM_003011
GATCGAGCTCGCCT	14	14	13	CATGGTACTGTGGC	1847	Chloride intracellular channel 1/CLIC1/NM_001288
GATCAAGTTTAAAT	16	37	279	CATGTTGGTGAAGG	24 592	Thymosin, β4, X chromosome/TMSB4X/NM_021109
GATCACACTGACCT	20	38	2	CATGCTGACCTGTG	5927	Major histocompatibility complex, class I,B/HLA-B/NM_005514
GATCAGTTCAGCCA	20	21	3	CATGGCGGTTGTGG	2308	Lysosomal-associated multispanning membrane protein-5/LAPTM5/NM_006762
GATCTATCACCTGT	24	12	16	CATGTAGGTTGTCT	43 934	Tumor protein, translationally controlled 1/TPT1/NM_003295
GATCTCATTCCATG	28	26	20	CATGGGAAAAAAAA	9029	ATP synthase, H+ transporting, mitochondrial F1 complex/ ATP5E/NM_006886
GATCCCCTCAGAAG	28	3	33	CATGTGAGGGAATA	9867	Triosephosphate isomerase 1/TPI1/NM_000365
GATCTGGGAGTGGC	31	85	29	CATGGCCATAAAAT	1740	Proteoglycan 1, secretory granule/ PRG1/NM_002727
GATCAAGAATTTGG	37	49	62	CATGTCTTTACTTG	1209	Actin related protein 2/3 complex, subunit 3, 21 kDa/ARPC3/NM_005719
GATCAAACTAGAAC	43	36	2	none		Cytochrome c oxidase subunit VIIc/COX7C/NM_001867
GATCTGCGCTTCCA	50	57	144	CATGTTTTTAATGT	4434	H3F3A histone/H3F3A/NM_002107
GATCCACTACCGGA	100	73	1	CATGCTCCTCACCT	9461	Ribosomal protein L13a/RPL13A/NM_012423
GATCGGGTAGCTCA	142	95	1	CATGCTGCTATACG	7957	Ribosomal protein L5/RPL5/NM_000969
GATCTGTGTTTGCT	148	123	2	CATGTACAAGAGGA	6980	Ribosomal protein L6/RPL6/NM_000970
GATCCAGTTCTAAG	198	165	1	CATGGTTAACGTCC	4635	Ribosomal protein L36a/RPL36a/NM_021029
GATCCTACTCTCTT	257	238	1	CATGGGAGTGGACA	4556	Ribosomal protein L18/ RPL18/BF677228
GATCAGACTCTGAA	448	263	3	CATGGCCGTGTCCG	15 586	Ribosomal protein S6/RPS6/AK027187
GATCGCCGTTCTGG	643	592	0	CATGTGTGTTGAGA	46 979	Eukaryotic translation elongation factor 1α1/EEF1A1/NM_001402

50 ng of RNA, by adding a loop of RNA amplification in the SAGE protocol as in SAR-SAGE or aRNA-longSAGE. It must be kept in mind that modifications aimed at reducing the amount of starting material may introduce representation biases. When this risk is carefully evaluated, it is possible to build a SAGE library starting from minute amounts of material.

The synthesis of double-stranded cDNA is performed according to classical methods. In the original protocol, a soluble biotinylated oligo(dT) primer was used to initiate the reaction and cDNA fragments were purified at a later step by capture on streptavidin-coated beads. In more recent protocols, oligo(dT) covalently linked to paramagnetic beads were used to bind poly(A) RNA. Performing all subsequent reactions on this solid phase greatly improved the overall yield.

The next step consists of cleaving cDNA by a restriction enzyme with a 4-bp recognition site (SAGE-anchoring enzyme). Only the most 3′ fragment of each cDNA remains on magnetic beads and the other fragments are eliminated. Since the enzyme cleaves every 256 bp (4^4) on average, it is expected to generate a 3′ fragment from the vast majority of cDNAs. The choice of an anchoring enzyme (AE) is limited by the necessity to create a protruding 4-bp single strand, further used to bind a synthetic linker. The most commonly used one is *Nla* III recognizing CATG sites, but *Sau* 3AI (GATC) works equally well. The bead suspension, bearing the purified 3′ cDNA fragments, is then divided in half. Each aliquot is ligated to one of two linkers containing the cognate sequence for the type IIS restriction endonuclease (*Bsm* FI or *Mme* I), used as a tagging enzyme (TE). The enzyme releases DNA fragments, each containing a SAGE tag associated to a synthetic linker.

The two pools of released fragments are ligated to form ditags. For being properly coupled tail to tail, DNA fragments must be blunted on their "biological" side and chemically blocked on their "synthetic" side. Ditags then serve as templates for amplification by polymerase chain reaction (PCR) with primers specific to each linker. Since there is a small probability of any two tags being assembled twice in the same combination, amplifying randomly assembled ditags rather than separate tags eliminates most PCR biases. Following amplification, the 4-bp restriction enzyme initially used to generate anchoring sites is now used to release the synthetic linkers. Purified ditags, now bearing sticky ends, are concatenated by ligation and the concatemers inserted in a cloning vector.

The SAGE library now consists of a collection of bacterial clones, each one harboring a unique random collection of tags. Before submitting them to sequence analysis, it is advisable to evaluate the average size of concatemers. It is indeed rather difficult to control ligation kinetics and concatemer elongation. Although it has no incidence on individual tag detection and does not alter the overall quality of the analysis, it has a direct effect on the cost and rate of acquisition of data. In some cases, it may be wiser preparing a new reaction rather than analyzing very small concatemers.

Technically, building a SAGE library was perceived as a heavy task, involving multiple steps critical for the success of the overall project. Several laboratories have devoted specific studies to the solution of various problems: efficiency of SAGE adaptor ligation, efficiency of *Bsm*FI digestion, prevention of amplification biases, improved *Nla*III digestion of 102-bp ditags,

synthesis of concatemers and cloning efficiency, easier clone selection. Eventually, the commercial availability of dedicated reagents (Invitrogen Corporation, Carlsbad, CA) greatly facilitated the preparation of SAGE libraries. With all these improvements, the SAGE technology is now being considered in many laboratories as a standard for gene expression profiling.

1.3 Processing Sequence Data

Raw SAGE data are provided as computer files containing the sequences of numerous clones. Merging these files into a unique one, usually in FASTA format, provides a permanent record of the analysis. A dedicated computer program is then needed for parsing this file and identifying individual tag boundaries. Since the first software created at the inception of SAGE in the laboratory of K. Kinzler (Johns Hopkins University, Baltimore, MD, USA), other parsers were developed with more or less extended functions. They usually include the detection of linker artifacts and sequence ambiguities, and must provide a measure of ditag average length. Indeed, type IIS restriction enzymes such as *Bsm* FI yield fragments of variable length. Since tags are ligated tail to tail, delineation of tag ends becomes so difficult in short ditags that they are preferably rejected. This problem has been discussed elsewhere. Ditags with twice the expected length can also be detected in sequence files. Their frequency may help estimate the sequencing error rate, assuming that the apparent loss of the 4-bp anchoring enzyme site defining ditag boundaries results from sequencing errors. If obtained early enough during the sequencing process, this indication may prompt the investigators to improve the quality of the sequencing machine or base calling software. When dealing with already obtained SAGE data, correcting the sequencing error noise requires more complex treatments. Identical ditags can also be observed. In principle, considering the complexity of the transcriptome in eukaryotic cells expressing thousands of genes, there is a low probability, which can be evaluated, to find the same tags associated several times in identical ditags. A low transcriptome complexity, increasing the number of identical ditag, may be observed in terminally differentiated, highly specialized cells. Alternatively, a high frequency of identical ditags may indicate that the cDNA input was not large enough to reproduce the true complexity of the natural mRNA population: this problem may be encountered when sampling minute amounts of biological material. The decision to discard or include repeatedly observed ditags in the calculation of tag frequencies eventually depends on a biologist's appraisal.

While in any case, the ultimate goal is to generate a reliable list of tags along with the number of times they occur, sequence data contain additional information, which, when extracted by appropriate software, provides elements of diagnosis for assessing the quality of SAGE libraries.

1.4 Statistical Analysis of Digital Transcriptomes

Methods for the quantitative analysis of gene expression can be broadly classified into two categories. "Analog" methods, using homemade microarrays or commercial DNA chips, are based on the measurement of analogical signals, correlated with the extent of hybridization of labeled transcripts to arrayed cDNAs

or oligonucleotides. "Digital" methods are based on tag counting in large collections of ESTs, SAGE, or Massively Parallel Signature Sequencing libraries (MPSS). The quantitative value of both kinds of methods has been thoroughly investigated and it is possible to delineate the limits within which each experimental dataset can be used. Digital profiles can now be interpreted on a sound theoretical framework. In analysis of SAGE data, the first assumption is usually that the number of tags is proportional to the number of cognate transcripts in the original biological sample, implying that each mRNA copy has the same chance of ending up as a tag in the library. The second assumption is that, when prepared for sequence characterization, SAGE tags are picked at random in a highly redundant, unbiased library. The problem of distinguishing biological values from technical errors becomes to set the limits between genuine variation and sampling fluctuation. Focusing on this sampling variability, several tests have been published giving essentially the same values when applied to the comparison of two SAGE libraries. Assessing fluctuations between groups of libraries requires more complex treatments.

1.5 Tag-to-gene Mapping

Identification of the gene transcripts revealed by SAGE tags is independent of experimental data acquisition. It might thus be conceivable to initiate a SAGE project on a given species before having an extensive knowledge of its genome, and indeed, investigations on human cells began several years before completion of the human genome sequencing project. Even when the full-length genome sequence is available, mapping SAGE tags directly to chromosomes is possible only for species with a small compact genome, such as yeast and *Arabidopsis thaliana*. For mammalian species, the size of noncoding regions is such that only LongSAGE (21-bp) tags can be mapped on to the genome. When dealing with conventional 10-bp tags, it is better to start from collections of expressed sequences such as UniGene, maintained at the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/). This experimental system organizes GenBank sequences into a nonredundant set of clusters, each one containing sequences that represent a unique gene. A sequence representative of each cluster can be retrieved from UniGene, as well as related information about the biological properties of the corresponding gene.

Using such a collection, an efficient strategy is to prepare a list of putative tags, predicted by performing virtual SAGE *in silico*. Lists of already predicted tags can be retrieved from SAGEmap, another public resource of NCBI dedicated to the dissemination of SAGE data. SAGEmap gives both tag-to-gene and gene-to-tag mapping, and calculates a score indicating the quality of the connection. Lists of predicted and experimental tags are easily recorded in the table format of current relational database management systems, either from commercial (Microsoft Access) or open source (MySQL), so that all tag-to-gene connections can be performed at once. The rate of success obviously depends on the number of well-annotated genes in the species under investigation. In human samples, more than 50% of tag sequences can now be connected immediately to known genes, representing about 70% of the total number of tag counts. The discrepancy between these two values simply reflects the fact that

abundant transcripts correspond more often to already well-known genes.

A small number of genes, which can be predicted, will systematically escape SAGE analysis because their transcript lacks the 4-bp anchoring site required to adapt SAGE linkers. Obviously, they must be studied by other means. A solution would be to prepare two libraries in parallel using each of the two enzymes (*Nla* III and *Sau* 3AI) already selected for their efficiency as anchoring enzymes. Table 1, as well as computer simulation on UniGene sequences, show that identification would then be nearly complete, but with a rather limited benefit compared to the twofold increase of the cost for sequencing analysis.

Another problem is that different genes may share the same tag. Although this risk seems statistically low, real-life examples show that multiple matches are actually observed at a rather high rate. In human samples, this problem results essentially from the presence of retrotransposed repetitive elements (mostly Alu sequences) inserted in the 3′ end of otherwise unrelated transcripts. These sequences can be retrieved from a dedicated database such as RepBase Update (http://www.girinst.org/Repbase_Update.html), and the corresponding tag and gene sequences registered in specific files.

Conversely, a gene may have more than one tag, either as the consequence of alternative splicing or because longer mRNA species encompassing the first polyadenylation site are actually present in cell extracts. The existence of alternative polyadenylation sites increases the number of tags and this phenomenon must be taken into account when analyzing SAGE data.

Upon completion of the tag-to-gene mapping procedure, it is tempting to consider that unmatched tags actually signal yet unknown transcript. However, before developing strategies for their identification, it is necessary to purge SAGE datasets from spurious sequences generated by technical artifacts. Tags matching the mitochondrial genome are sometimes considered as contaminants. However, it is worth mentioning that transcription of the mitochondrial genome produces two polycistronic RNAs (one from each strand) giving rise to a series of poly(A) RNAs. Tags matching the mitochondrial genome are therefore biologically relevant. By contrast, ribosomal RNA fragments may be considered as contaminants, because mRNAs represent a low percentage of the overall nucleic acid cell content. These fragments amounted to up to 1% of the total tag population in libraries generated using the initial SAGE protocol. With more stringent conditions being used for the purification of poly(A) RNAs, their abundance was reduced to less than 0.01% in most libraries. Finally, a significant number of incorrect tags are created by sequencing errors. Typically, SAGE tags are obtained by single-pass sequencing. With 1% of nucleotides determined incorrectly, about 1 of 10 tags will be erroneous. However, since the probability of the same real tag being erroneously sequenced several times is low, the major effect is an increasing number of tags observed at low to moderate frequency. With a small number of SAGE libraries at hand, authors usually discarded tags observed only once. Now that a large number of libraries are publicly available, it is possible to distinguish artifacts from genuine tags of rare transcripts by checking their occurrence in multiple samples. Taken together, technical artifacts essentially increase the apparent number of different tag sequences. However, because their individual frequencies

are low, they alter only marginally the quantification of authentic mRNAs.

As a whole, tag-to-gene mapping can be perceived as a sieving process, in which the list of experimental tags is percolated through multiple filters. At the end of this process, investigators obtain a list of tags identifying authentic mRNAs and a list of tags that are susceptible to revealing the existence of novel transcripts.

2 Making Sense of SAGE Data

2.1 Unmatched Tags and Novel Transcripts: Using SAGE to Annotate the Genome and the Transcriptome

In species with a well-annotated genome, the vast majority of known mRNAs are identified without ambiguity. However, in spite of progress in genome annotation, SAGE libraries often contain unassigned tags susceptible to reveal rarely expressed genes, new alternate transcripts of known genes, genetic variants, or polyadenylated transcripts of noncoding genes.

As could be expected, the characterization of new transcripts proved to be more efficient when using the LongSAGE method instead of the classical one, as shown for the human and more recently for the mouse genome. When dealing with a conventional 10-bp short tag, its assignment to entirely new loci will be easier if it can be matched with a yet unassigned EST, providing an extended probe for experimental investigations and computer searches. When such information is lacking, it is still possible to use the tag itself as a probe, and to amplify an extended cDNA stretch starting from the 10-bp SAGE tag.

With respect to the size of the genome, evaluating the size of the transcriptome remains a controversial issue. Alternative splicing and alternative polyadenylation, each concerning perhaps half the human genes, contribute to increasing the number of transcripts. SAGE can detect alternative exons in the very 3′ end of mRNAs, but does not make distinction between splice variants differing in more 5′ sites. On the contrary, SAGE is highly sensitive to alternative polyadenylation. The existence of transcripts of variable length, seldom registered in current transcript database, has been clearly documented by using dedicated computer programs. These results suggest that nearly half the human genes might generate alternate transcripts, depending on tissue- or disease-specific regulatory mechanisms.

SAGE is sensitive to genetic polymorphism. Alternative tags can be generated by single nucleotide polymorphism (SNP), either because an SNP disrupts the anchoring site of the most commonly observed tag, creates a new anchoring site, or changes the sequence usually registered in databases. A database of human SNP-associated alternative SAGE tags has recently been integrated in SAGE Genie, improving the interpretation of SAGE data.

Detailed analyzes of chromosomes 21 and 22 by massively parallel hybridization on genomic DNA probes revealed a mammalian transcriptome larger than previously estimated. In fact, just as SAGE itself, these strategies made no distinction between conventional mRNAs and noncoding polyadenylated transcripts. The ability of SAGE to reveal such untranslated transcripts deserves interest, considering the regulatory function they may exert.

An initially intriguing observation was that of a subpopulation of tags mapping on to the opposite strand of known mRNAs. For a part, they might be assigned to repetitive elements inserted in opposite orientations in multiple mRNAs. More than one million copies of the approximate 300-bp Alu elements are interspersed in the human genome and a fraction of them are inserted in transcribed 3′-UTRs. Other pairs of tags were found associated with unique pairs of transcripts. Some of them are encoded in trans, the corresponding genes being found on remote loci, while other pairs of tags reveal cis-encoded genes. These observations are corroborated by recent computer searches at the genome level, identifying a large number of contiguous, oppositely oriented genes, potentially able to express complementary mRNAs. SAGE is well adapted for investigating whether such genes are expressed simultaneously, a condition obviously required for them for producing functional double-stranded RNAs.

The detection of genuine tags that were left unpredicted by the analysis of conventional transcript sequences contributes to improving the quality of genome annotation. SAGE can now be considered as more sensitive than EST sequencing for detecting low-abundantly and rarely expressed genes in mammalian genomes.

2.2 From SAGE to Functional Genomics: Linking High-throughput Expression Analysis, Studies on Genome Organization and Complex Biological Networks

The first comprehensive view of a transcriptome was obtained by representing the frequency of SAGE tags on the physical map of yeast chromosomes. More recently, a human transcriptome map, based on the computer-assisted representation of SAGE tags mapping on to chromosomes, has illustrated the power of SAGE for investigating the organization of chromatin and the distribution of genes in actively transcribed regions.

A contemporary challenge is to establish relationships between global variations of gene expression profiles and changes in complex metabolic or signaling networks. Until now, findings in molecular and cell biology were reported in textual form, using an uncontrolled vocabulary, and disseminated in a million publications. The heterogeneity of biological knowledge and, even more critically, the fuzziness of essential concepts including the concept of biological function itself, make the task highly difficult. Collective efforts are being made for calling genes unequivocally (HUGO Gene Nomenclature Committee, http://www.gene.ucl.ac.uk/nomenclature/) and to control the vocabulary used for defining the semantic links that describe functional relations among biological entities (Gene Ontology™ (GO) Consortium: http://www.geneontology.org/). Lists of descriptors provided by this consortium can be linked to SAGE data, so that tags may be automatically classified according to the functional properties of their corresponding genes. Most algorithms for large-scale gene expression analysis were developed to facilitate the interpretation of microarray data. In fact, methods such as association-rules discovery or clustering analysis can be used as well for analyzing SAGE data.

Several interactive Web sites provide synthetic views of an increasing number of networks. For instance, by connecting to the Web site of the Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/), it is possible to automatically stain a large number

of metabolic charts according to the results of SAGE and to get in a few hours a synthetic view of events occurring in most metabolic and some signaling pathways.

Interpretation of the set of data generated by each individual laboratory is usually not extensive. Since SAGE data are cumulative, comparisons with remote data may enrich the interpretation of local results. For this purpose, the Gene Expression Omnibus collects expression datasets issued from SAGE and microarray technologies, making them accessible for further analysis and reinterpretation. More specifically, SAGE Genie, a bioinformatic project dedicated to the management of large amounts of experimental SAGE data provides, among other intuitive tools, visualization displayers relating gene expression and tissue localization.

2.3 SAGE, Molecular Classification of Cell Types and Cell Physiology

Conceptually, a given cell type might be defined by its transcriptome, considered as the whole set of genes responsible for its phenotype. However, cells are under constant evolution. They may express functions dictated by proteins whose mRNAs were expressed at a former stage of development, but already catabolized at the time of sampling. Conversely, the presence of a transcript does not always mean that it was being translated at the time of sampling. A gene expression profile must therefore be considered as a snapshot and interpreted as such. With this reservation, transcriptome analysis possesses an advantage over other high-throughput strategies, such as proteome analysis, which, because proteins cannot be amplified, is less sensitive and identifies a smaller number of cell components. Evaluation of SAGE data from a variety of normal and diseased tissues soon provided an enlarged view of the human transcriptomes.

The possibility to work on tiny cell samples was decisive for defining the profile of discrete cell subtypes. Analysis of microdissected anatomical structures provided panoramic views of complex tissues such as kidney (Fig. 2), retina, or brain. Homogeneous samples of circulating cells are more easily purified than cells from solid tissues and numerous blood cell types have now been described by SAGE at the molecular level.

Most biological response modifiers do not act only on rapid cell kinetics through the activation of signaling pathways, but also exert long-lasting effects by changing gene expression patterns. Large-scale transcriptome analysis thus added a new dimension to studies involving the experimental induction of a signaling pathway and the effects of cytokines, hormones, or differentiation inducers.

2.4 SAGE and Disease-induced Alterations of the Transcriptome

It was readily perceived that SAGE offered new possibilities for characterizing global expression profile alterations in infectious diseases, for instance, upon exposition to bacterial toxins or viral infection. However, the major incentive for implementing the SAGE technology came from cancer research. Molecular classification of cancer cells, expected for many years, is now coming of age, opening new perspectives for cancer therapies. The power of SAGE for identifying molecular markers of cancer was rapidly demonstrated, and since the pioneering works, a large number of SAGE libraries have been obtained

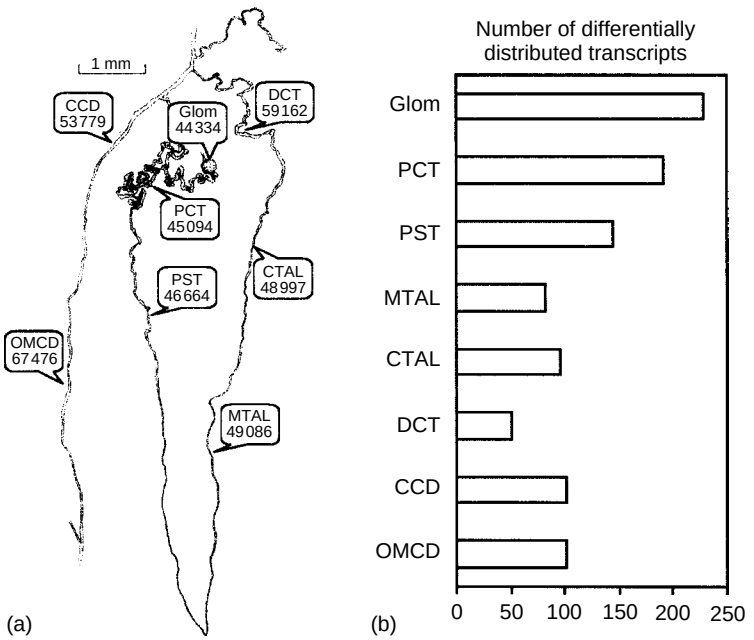

Fig. 2 Transcriptome characterization of individual nephron portions of the human kidney. (a) A microdissected human nephron. Structures analyzed for trancriptome characterization, together with the number of SAGE tags sequenced are indicated. Glom, glomerulus; PCT, proximal convoluted tubule; PST, proximal straight tubule; MTAL, medullary thick ascending limb of Henle's loop; CTAL, cortical thick ascending limb of Henle's loop; DCT,distal convoluted tubule; CCD, cortical collecting duct; OMCD, outer medullary collecting duct. (b) Overview of mRNA differential distribution. The number of mRNAs displaying a significant differential distribution ($p < 0.01$ as compared with three nephron portions) is indicated for each structure. The overall number of heterogeneously distributed transcripts is 998. Data from Chabardes-Garonne, D., Mejean, A., Aude, J.C., et al. (2003) A panoramic view of gene expression in the human kidney, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13710–13715.

with the purpose of comparing normal cells and their malignant counterparts. SAGE was also used to investigate tumor angiogenesis or drug sensitivity of malignant cells. It is not possible to summarize here all the projects developed since the inception of SAGE. A comprehensive view can be obtained from SAGEmap (http://www.ncbi.nlm.nih.gov/SAGE) and SAGE Genie (http://cgap.nci.nih.gov/SAGE), two components of the Cancer Genome Anatomy Project intended to provide a central database for comparing gene expression profiles in normal and malignant tissues.

3 Perspectives

SAGE was developed at the same time as DNA chip and microarrays technologies.

More recently, MPSS offered an alternative approach to SAGE, but the need for specific equipment limited its use to a small number of laboratories. Each technology has its specific advantages. Methods based on hybridization are well adapted for the comparison of multiple samples, but are limited to the study of the genes represented in the array. SAGE and MPSS are more laborious, but do not require *a priori* knowledge on gene sequences. Moreover, their results are cumulative and comparison between independent datasets are easily performed, while exchanging microarray data requires a complex standardization protocol. Rather than a competing methodology with other global gene expression approaches, it thus seems better considering SAGE as a complementary approach, and the decision to initiate a new SAGE project as depending on the biological problem is under investigation.

SAGE is still receptive to methodological improvements. It is obvious that it will benefit from any progress in DNA sequencing technology that increases the rate of acquisition and reduces the cost of analysis. Until now, the conventional method identified mRNAs by tagging their 3′ end. However, there was still a need for a global method of also identifying transcripts through their 5′ end. Protocols enabling the high-throughput extraction of tags from mRNA 5′ ends have now been published. The 5′-end SAGE is expected to improve the characterization of transcriptional start sites and the functional annotation of poorly known genes.

It is worth noticing that SAGE has already produced a huge amount of experimental results from which latent biological knowledge may be extracted by mining the existing dataset. With the progress of computational biology, the integration of SAGE data in larger information systems, seamlessly coupling data from various sources, opens the way to global analyses on the regulation of gene expression at a higher level of organization. Comparative analysis between man and animal models will be highly useful, and facilitated by the development of new databases such as the Mouse SAGE site (http://mouse.biomed.cas.cz/sage/).

SAGE has already contributed, in conjunction with other methodologies, in achieving a more comprehensive and quantitative picture of transcriptomes. It will probably continue to provide new data until their complete disclosure.

See also DNA Libraries; Genetics, Molecular Basis of; Genomic DNA Libraries, Construction and Applications; Proteomics; Restriction Endonuclease and DNA-Modification Methyltransferases.

Bibliography

Books and Reviews

Blackshaw, S., Kim, J.B., St. Croix, B., Polyak, K. (2002) Serial Analysis of Gene Expression, in: Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (Eds.) *Current Protocols in Molecular Biology*, Wiley, NY.

Cheval, L., Virlon, B., Elalouf, J.-M. (2000) SADE: A Microassay for Serial Analysis of Gene Expression, in: Hunt, S.P., Livesey, F.J. (Eds.) *Functional Genomics: A Practical Approach*, Oxford University Press, London.

Lorkowski, S., Cullen, P. (2003) *Analysing Gene Expression; A Handbook of Methods Possibilities and Pitfalls*, Wiley-VCH, Weinheim, Germany.

Shimkets, R.A. (2004) *Gene Expression Profiling: Methods and Protocols (Methods in Molecular Biology)*, Humana Press, Totowa, NJ.

Wang, S.M. (2005) *SAGE Technologies: Current Innovations and Future Trends*, Horizon Scientific Press, Norwich, CT.

Primary Literature

Adams, M.D., Kelley, J.M., Gocayne, J.D., et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project, *Science* **252**, 1651–1656.

Akmaev, V.R., Wang, C.J. (2004) Correction of sequence-based artifacts in serial analysis of gene expression, *Bioinformatics* **20**, 1254–1263.

Allinen, M., Beroukhim, R., Cai, L., et al. (2004) Molecular characterization of the tumor microenvironment in breast cancer, *Cancer Cell* **6**, 17–32.

Angelastro, J.M., Klimaschewski, L.P., Vitolo, O.V. (2000) Improved NlaIII digestion of PAGE-purified 102-bp ditags by addition of a single purification step in both the SAGE and microSAGE protocols, *Nucleic Acids Res.* **28**, E62.

Angelastro, J.M., Ryu, E.J., Torocsik, B., Fiske, B.K., Greene, L.A. (2002) Blue-white selection step enhances the yield of SAGE concatemers, *Biotechniques* **32**, 484–486.

Angelastro, J.M., Klimaschewski, L., Tang, S., Vitolo, O.V., Weissman, T.A., Donlin, L.T., Shelanski, M.L., Greene, L.A. (2000) Identification of diverse nerve growth factor-regulated genes by serial analysis of gene expression (SAGE) profiling, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10424–10429.

Audic, S., Claverie, J.M. (1997) The significance of digital gene expression profiles, *Genome Res.* **7**, 986–995.

Baggerly, K.A., Deng, L., Morris, J.S., Aldaz, C.M. (2004) Overdispersed logistic regression for SAGE: modelling multiple groups and covariates, *BMC Bioinformatics* **5**, 144.

Beaudoing, E., Gautheret, D. (2001) Identification of alternate polyadenylation sites and analysis of their tissue distribution using EST data, *Genome Res.* **11**, 1520–1526.

Becquet, C., Blachon, S., Jeudy, B., Boulicaut, J.F., Gandrillon, O. (2002) Strong-association-rule mining for large-scale gene-expression data analysis: a case study on human SAGE data, *Genome Biol.* **3**, RESEARCH0067.

Beissbarth, T., Hyde, L., Smyth, G.K., Job, C., Boon, W.M., Tan, S.S., Scott, H.S., Speed, T.P. (2004) Statistical modeling of sequencing errors in SAGE libraries, *Bioinformatics* **20**(Suppl. 1), I31–I39.

Bertrand, G., Coste, J., Segarra, C., Schved, J.F., Commes, T., Marti, J. (2004) Use of serial analysis of gene expression (SAGE) technology reveals new granulocytic markers, *J. Immunol. Methods* **292**, 43–58.

Bonafoux, B., Lejeune, M., Piquemal, D., Quere, R., Baudet, A., Assaf, L., Marti, J., Aguilar-Martinez, P., Commes, T. (2004) Analysis of remnant reticulocyte mRNA reveals new genes and antisense transcripts expressed in the human erythroid lineage, *Haematologica* **89**, 1434–1438.

Boon, K., Caron, H.N., Van Asperen, R., et al. (2001) N-myc enhances the expression of a large set of genes functioning in ribosome biogenesis and protein synthesis, *EMBO J.* **20**, 1383–1393.

Boon, K., Osorio, E.C., Greenhut, S.F., et al. (2002) An anatomy of normal and malignant gene expression, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11287–11292.

Brazma, A., Hingamp, P., Quackenbush, J., et al. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data, *Nat. Genet.* **29**, 365–371.

Brenner, S., Johnson, M., Bridgham, J., et al. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays, *Nat. Biotechnol.* **18**, 630–634.

Buckhaults, P., Zhang, Z., Chen, Y.C., et al. (2003) Identifying tumor origin using a gene expression-based classification map, *Cancer Res.* **63**, 4144–4149.

Cai, L., Huang, H., Blackshaw, S., Liu, J.S., Cepko, C., Wong, W.H. (2004) Clustering analysis of SAGE data using a Poisson approach, *Genome Biol.* **5**, R51.

Caron, H., van Schaik, B., van der Mee, M., et al. (2001) The human transcriptome map: clustering of highly expressed genes in chromosomal domains, *Science* **291**, 1289–1292.

Chabardes-Garonne, D., Mejean, A., Aude, J.C., et al. (2003) A panoramic view of gene expression in the human kidney, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13710–13715.

Chen, J., Sun, M., Lee, S., Zhou, G., Rowley, J.D., Wang, S.M. (2002) Identifying novel transcripts and novel genes in the human

genome by using novel SAGE tags, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12257–12262.

Chen, J., Sun, M., Kent, W.J., Huang, X., Xie, H., Wang, W., Zhou, G., Shi, R.Z., Rowley, J.D. (2004) Over 20% of human transcripts might form sense-antisense pairs, *Nucleic Acids Res.* **32**, 4812–4820.

Datson, N.A., van der Perk-de Jong, J., van den Berg, M.P., de Kloet, E.R., Vreugdenhil, E. (1999) MicroSAGE: a modified procedure for serial analysis of gene expression in limited amounts of tissue, *Nucleic Acids Res.* **27**, 1300–1307.

de Chaldee, M., Gaillard, M.C., Bizat, N., Buhler, J.M., Manzoni, O., Bockaert, J., Hantraye, P., Brouillet, E., Elalouf, J.M. (2003) Quantitative assessment of transcriptome differences between brain territories, *Genome Res.* **13**, 1646–1653.

de Jonge, R.R., Vreijling, J.P., Meintjes, A., Kwa, M.S., van Kampen, A.H., van Schaik, I.N., Baas, F. (2003) Transcriptional profile of the human peripheral nervous system by serial analysis of gene expression, *Genomics* **82**, 97–108.

DeRisi, J.L., Iyer, V.R., Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale, *Science* **278**, 680–686.

Divina, P., Forejt, J. (2004) The ouse SAGE Site: database of public mouse SAGE libraries, *Nucleic Acids Res.* **32**, D482–D483.

Du, Z., Scott, A.D., May, G.D. (2003) Amplification of high-quantity serial analysis of gene expression ditags and improvement of concatemer cloning efficiency, *Biotechniques* **35**, 66–67.

El-Rifai, W., Moskaluk, C.A., Abdrabbo, M.K., Harper, J., Yoshida, C., Riggins, G.J., Frierson, H.F., Powell, S.M. (2002) Gastric cancers overexpress S100A calcium-binding proteins, *Cancer Res.* **62**, 6823–6826.

Fizames, C., Munos, S., Cazettes, C., et al. (2004) The arabidopsis root transcriptome by serial analysis of gene expression. Gene identification using the genome sequence, *Plant Physiol.* **134**, 67–80.

Gnatenko, D.V., Dunn, J.J., McCorkle, S.R., Weissmann, D., Perrotta, P.L., Bahou, W.F. (2003) Transcript profiling of human platelets using microarray and serial analysis of gene expression, *Blood* **101**, 2285–2293.

Hashimoto, S., Nagai, S., Sese, J., et al. (2003) Gene expression profile in human leukocytes, *Blood* **101**, 3509–3513.

Hashimoto, S., Suzuki, T., Dong, H.Y., Yamazaki, N., Matsushima, K. (1999) Serial analysis of gene expression in human monocytes and macrophages, *Blood* **94**, 837–844.

Hashimoto, S., Suzuki, Y., Kasai, Y., Morohoshi, K., Yamada, T., Sese, J., Morishita, S., Sugano, S., Matsushima, K. (2004) 5′-end SAGE for the analysis of transcriptional start sites, *Nat. Biotechnol.* **22**, 1146–1149.

Heidenblut, A.M., Luttges, J., Buchholz, M., et al. (2004) aRNA-longSAGE: a new approach to generate SAGE libraries from microdissected cells, *Nucleic Acids Res.* **32**, e131.

Hough, C.D., Sherman-Baust, C.A., Pizer, E.S., Montz, F.J., Im, D.D., Rosenshein, N.B., Cho, K.R., Riggins, G.J., Morin, P.J. (2000) Large-scale serial analysis of gene expression reveals genes differentially expressed in ovarian cancer, *Cancer Res.* **60**, 6281–6287.

Hu, Y., Sun, H., Drake, J., et al. (2004) From mice to humans: identification of commonly deregulated genes in mammary cancer via comparative SAGE studies, *Cancer Res.* **64**, 7748–7755.

Iseli, C., Stevenson, B.J., de Souza, S.J., Samaia, H.B., Camargo, A.A., Buetow, K.H., Strausberg, R.L., Simpson, A.J., Bucher, P., Jongeneel, C.V. (2002) Long-range heterogeneity at the 3′ ends of human mRNAs, *Genome Res.* **12**, 1068–1074.

Jurka, J. (2000) Repbase update: a database and an electronic journal of repetitive elements, *Trends Genet.* **16**, 418–420.

Kampa, D., Cheng, J., Kapranov, P., et al. (2004) Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22, *Genome Res.* **14**, 331–342.

Kenzelmann, M., Muhlemann, K. (1999) Substantially enhanced cloning efficiency of SAGE (Serial Analysis of Gene Expression) by adding a heating step to the original protocol, *Nucleic Acids Res.* **27**, 917–918.

Kenzelmann, M., Muhlemann, K. (2000) Transcriptome analysis of fibroblast cells immediate-early after human cytomegalovirus infection, *J. Mol. Biol.* **304**, 741–751.

Lal, A., Lash, A.E., Altschul, S.F., et al. (1999) A public database for gene expression in human cancers, *Cancer Res.* **59**, 5403–5407.

Lash, A.E., Tolstoshev, C.M., Wagner, L., Schuler, G.D., Strausberg, R.L., Riggins, G.J., Altschul, S.F. (2000) SAGEmap: a public gene expression resource, *Genome Res.* **10**, 1051–1060.

Lee, S., Chen, J., Zhou, G., Wang, S.M. (2001) Generation of high-quantity and quality tag/ditag cDNAs for SAGE analysis, *Biotechniques* **31**, 348–350.

Lee, J.Y., Eom, E.M., Kim, D.S., Ha-Lee, Y.M., Lee, D.H. (2003) Analysis of gene expression profiles of gastric normal and cancer tissues by SAGE, *Genomics* **82**, 78–85.

Liang, P. (2002) SAGE genie: a suite with panoramic view of gene expression, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11547–11548.

Lockhart, D.J., Dong, H., Byrne, M.C., et al. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays, *Nat. Biotechnol.* **14**, 1675–1680.

Margulies, E.H., Kardia, S.L., Innis, J.W. (2001) Identification and prevention of a GC content bias in SAGE libraries, *Nucleic Acids Res.* **29**, E60.

Menssen, A., Hermeking, H. (2002) Characterization of the c-MYC-regulated transcriptome by SAGE: identification and analysis of c-MYC target genes, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6274–6279.

Muschen, M., Lee, S., Zhou, G., Feldhahn, N., Barath, V.S., Chen, J., Moers, C., Kronke, M., Rowley, J.D., Wang, S.M. (2002) Molecular portraits of B cell lineage commitment, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10014–10019.

Pauws, E., van Kampen, A.H., van de Graaf, S.A., de Vijlder, J.J., Ris-Stalpers, C. (2001) Heterogeneity in polyadenylation cleavage sites in mammalian mRNA sequences: implications for SAGE analysis, *Nucleic Acids Res.* **29**, 1690–1694.

Pauws, E., Veenboer, G.J., Smit, J.W., de Vijlder, J.J., Morreau, H., Ris-Stalpers, C. (2004) Genes differentially expressed in thyroid carcinoma identified by comparison of SAGE expression profiles, *FASEB J.* **18**, 560–561.

Piquemal, D., Commes, T., Manchon, L., Lejeune, M., Ferraz, C., Pugnere, D., Demaille, J., Elalouf, J.M., Marti, J. (2002) Transcriptome analysis of monocytic leukemia cell differentiation, *Genomics* **80**, 361–371.

Polyak, K., Riggins, G.J. (2001) Gene discovery using the serial analysis of gene expression technique: implications for cancer research, *J. Clin. Oncol.* **19**, 2948–2958.

Porter, D.A., Krop, I.E., Nasser, S., Sgroi, D., Kaelin, C.M., Marks, J.R., Riggins, G., Polyak, K. (2001) A SAGE (serial analysis of gene expression) view of breast tumor progression, *Cancer Res.* **61**, 5697–5702.

Powell, J. (1998) Enhanced concatemer cloning-a modification to the SAGE (Serial Analysis of Gene Expression) technique, *Nucleic Acids Res.* **26**, 3445–3446.

Quere, R., Manchon, L., Lejeune, M., Clement, O., Pierrat, F., Bonafoux, B., Commes, T., Piquemal, D., Marti, J. (2004) Mining SAGE data allows large-scale, sensitive screening of antisense transcript expression, *Nucleic Acids Res.* **32**, e163.

Rinn, J.L., Euskirchen, G., Bertone, P., et al. (2003) The transcriptional activity of human Chromosome 22, *Genes Dev.* **17**, 529–540.

Robert-Nicoud, M., Flahaut, M., Elalouf, J.M. et al. (2001) Transcriptome of a mouse kidney cortical collecting duct cell line: effects of aldosterone and vasopressin, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2712–2716.

Ruijter, J.M., Van Kampen, A.H., Baas, F. (2002) Statistical evaluation of SAGE libraries: consequences for experimental design, *Physiol. Genomics* **11**, 37–44.

Saha, S., Sparks, A.B., Rago, C., Akmaev, V., Wang, C.J., Vogelstein, B., Kinzler, K.W., Velculescu, V.E. (2002) Using the transcriptome to annotate the genome, *Nat. Biotechnol.* **20**, 508–512.

Seth, P., Krop, I., Porter, D., Polyak, K. (2002) Novel estrogen and tamoxifen induced genes identified by SAGE (Serial Analysis of Gene Expression), *Oncogene* **21**, 836–843.

Sharon, D., Blackshaw, S., Cepko, C.L., Dryja, T.P. (2002) Profile of the genes expressed in the human peripheral retina, macula, and retinal pigment epithelium determined through serial analysis of gene expression (SAGE), *Proc. Natl. Acad. Sci. U.S.A.* **99**, 315–320.

Silva, A.P., De Souza, J.E., Galante, P.A., Riggins, G.J., De Souza, S.J., Camargo, A.A. (2004) The impact of SNPs on the interpretation of SAGE and MPSS experimental data, *Nucleic Acids Res.* **32**, 6104–6110.

So, A.P., Turner, R.F., Haynes, C.A. (2004) Increasing the efficiency of SAGE adaptor

ligation by directed ligation chemistry, *Nucleic Acids Res.* **32**, e96.

St Croix, B., Rago, C., Velculescu, V. et al. (2000) Genes expressed in human tumor endothelium, *Science* **289**, 1197–1202.

Stein, W.D., Litman, T., Fojo, T., Bates, S.E. (2004) A serial analysis of gene expression (SAGE) database analysis of chemosensitivity: comparing solid tumors with cell lines and comparing solid tumors from different tissue origins, *Cancer Res.* **64**, 2805–2816.

Strausberg, R.L., Simpson, A.J., Old, L.J., Riggins, G.J. (2004) Oncogenomics and the development of new cancer therapies, *Nature* **429**, 469–474.

Sun, M., Zhou, G., Lee, S., Chen, J., Shi, R.Z., Wang, S.M. (2004) SAGE is far more sensitive than EST for detecting low-abundance transcripts, *BMC Genomics* **5**, 1.

Suzuki, T., Hashimoto, S., Toyoda, N., Nagai, S., Yamazaki, N., Dong, H.Y., Sakai, J., Yamashita, T., Nukiwa, T., Matsushima, K. (2000) Comprehensive gene expression profile of LPS-stimulated human monocytes by SAGE, *Blood* **96**, 2584–2591.

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W. (1995) Serial analysis of gene expression, *Science* **270**, 484–487.

Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Hieter, P., Vogelstein, B., Kinzler, K.W. (1997) Characterization of the yeast transcriptome, *Cell* **88**, 243–251.

Velculescu, V.E., Madden, S.L., Zhang, L., et al. (1999) Analysis of human transcriptomes, *Nat. Genet.* **23**, 387–388.

Vencio, R.Z., Brentani, H., Patrao, D.F., Pereira, C.A. (2004) Bayesian model accounting for within-class biological variability in serial analysis of gene expression (SAGE), *BMC Bioinformatics* **5**, 119.

Versteeg, R., van Schaik, B.D., van Batenburg, M.F., Roos, M., Monajemi, R., Caron, H., Bussemaker, H.J., van Kampen, A.H. (2003) The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes, *Genome Res.* **13**, 1998–2004.

Vilain, C., Libert, F., Venet, D., Costagliola, S., Vassart, G. (2003) Small amplified RNA-SAGE: an alternative approach to study transcriptome from limiting amount of mRNA, *Nucleic Acids Res.* **31**, e24.

Virlon, B., Cheval, L., Buhler, J.M., Billon, E., Doucet, A., Elalouf, J.M. (1999) Serial microanalysis of renal transcriptomes, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 15286–15291.

Wei, C.L., Ng, P., Chiu, K.P., Wong, C.H., Ang, C.C., Lipovich, L., Liu, E.T., Ruan, Y. (2004) 5′ Long serial analysis of gene expression (LongSAGE) and 3′ LongSAGE for transcriptome characterization and genome annotation, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11701–11706.

Yamamoto, M., Wakatsuki, T., Hada, A., Ryo, A. (2001) Use of serial analysis of gene expression (SAGE) technology, *J. Immunol. Methods* **250**, 45–66.

Ye, S.Q., Zhang, L.Q., Zheng, F., Virgil, D., Kwiterovich, P.O. (2000) miniSAGE: gene expression profiling using serial analysis of gene expression from 1 microg total RNA, *Anal. Biochem.* **287**, 144–152.

Yu, J., Zhang, L., Hwang, P.M., Rago, C., Kinzler, K.W., Vogelstein, B. (1999) Identification and classification of p53-regulated genes, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14517–14522.

Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., Vogelstein, B., Kinzler, K.W. (1997) Gene expression profiles in normal and cancer cells, *Science* **276**, 1268–1272.

Zhou, G., Chen, J., Lee, S., Clark, T., Rowley, J.D., Wang, S.M. (2001) The pattern of gene expression in human CD34(+) stem/progenitor cells, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13966–13971.

Sex Determination, Molecular Biology of: *see* Gonad Differentiation Genetics

Sex Hormones (Male): Analogs and Antagonists

Robert W. Brueggemeier
The Ohio State University, Columbus, OH, USA

1 Introduction 4

2 Historical 4

3 Endogenous Male Sex Hormones 5
3.1 Occurrence and Physiological Roles 5
3.2 Biosynthesis 6
3.3 Absorption and Distribution 10
3.4 Metabolism 11
3.4.1 Reductive Metabolism 11
3.4.2 Oxidative Metabolism 14
3.5 Mechanism of Action 17

4 Synthetic Androgens 22
4.1 Current Drugs on the Market 22
4.2 Therapeutic Uses and Bioassays 22
4.3 Structure–Activity Relationships for Androgens 25
4.3.1 Early Modifications 25
4.3.2 Methylated Derivatives 25
4.3.3 Ester Derivatives 25
4.3.4 Halo Derivatives 26
4.3.5 Other Androgen Derivatives 26
4.3.6 Summary of Structure–Activity Relationships 27
4.4 Absorption, Distribution, and Metabolism 29
4.5 Toxicities 30

5 Anabolic Agents 30
5.1 Current Drugs on the Market 30
5.2 Therapeutic Uses and Bioassays 30
5.3 Structure–Activity Relationships for Anabolics 33

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

5.3.1 19-Nor Derivatives 33
5.3.2 Dehydro Derivatives 33
5.3.3 Alkylated Analogs 35
5.3.4 Hydroxy and Mercapto Derivatives 36
5.3.5 Oxa, Thia, and Aza Derivatives 37
5.3.6 Deoxy and Heterocyclic-fused Analogs 38
5.3.7 Esters and Ethers 40
5.3.8 Summary of Structure–Activity Relationships 41
5.4 Absorption, Distribution, and Metabolism 43
5.5 Toxicities 43

6 Androgen Antagonists 43

6.1 Current Drugs on the Market 44
6.2 Antiandrogens 44
6.2.1 Therapeutic Uses 44
6.2.2 Structure–Activity Relationships for Antiandrogens 44
6.2.3 Selective Androgen Receptor Modulators (SARMs) 47
6.2.4 Absorption, Distribution, and Metabolism 48
6.2.5 Toxicities 48
6.3 Enzyme Inhibitors 48
6.3.1 5α-Reductase Inhibitors 49
6.3.2 17,20-Lyase Inhibitors 52
6.3.3 C_{19}Steroids as Aromatase Inhibitors 53
6.3.4 Other Agents 55

7 Summary 56

Acknowledgment 57

Bibliography 57
Books and Reviews 57
Primary Literature 58

Keywords

Anabolics
Compounds that demonstrate a marked retention of nitrogen through an increase of protein synthesis and a decrease in protein catabolism in the body.

Androgens
Steroid hormones responsible for the primary and secondary sex characteristics of the male, including the development of the vas deferens, prostate, seminal vesicles, and penis.

Antiandrogens
Agents that compete with endogenous androgens for the hormone binding site on the androgen receptor and blocks androgen action.

Dihydrotestosterone
The C_{19} steroid hormone that is the 5α-reduced metabolite of testosterone produced in certain androgen target tissues and is the most potent endogenous androgen.

5α-Reductase Inhibitors
Compounds that inhibit the conversion of testosterone to its more active metabolite dihydrotestosterone.

Selective Androgen Receptor Modulators
Agents that may act as an androgen antagonist or weak agonist in one tissue but act as strong androgen agonist in another tissue type.

Testosterone
The C_{19} steroid hormone that is the predominant circulating androgen in the bloodstream and is produced mainly by the testis.

The steroid testosterone is the major circulating sex hormone of the male and serves as the prototype for the androgens, the anabolic agents, and androgen antagonists. The endogenous androgens are biosynthesized from cholesterol in various tissues in the body; the majority of the circulating androgens are made in the testes under the stimulation of the gonadotropin LH. The reduction of testosterone to dihydrotestosterone is necessary for the androgenic actions of testosterone in androgen target tissues such as the prostate; the oxidation of testosterone by the enzyme aromatase produces estrogens. The androgenic actions of testosterone are due to the binding of dihydrotestosterone to its nuclear receptor, followed by dimerization of the receptor complex and binding to a specific DNA sequence. This binding of the homodimer to the androgen response element leads to gene expression, stimulation of the synthesis of new mRNA, and subsequent protein biosynthesis. The synthetic androgens and anabolics were prepared to impart oral activity to the androgen molecule, to separate the androgenic effects of testosterone from its anabolic effects and to improve upon its biological activities. Drug preparations are used for the treatment of various androgen-deficient diseases, for the therapy of diseases characterized by muscle wasting and protein catabolism, for postoperative adjuvant therapy, and for the treatment of certain hormone-dependent cancers. The two major categories of androgen antagonists are the antiandrogens, which block interactions of androgens with the androgen receptor, and the inhibitors of androgen biosynthesis and metabolism. Such compounds have therapeutic potential in the treatment of acne, virilization in women, hyperplasia and neoplasia of the prostate, and baldness, and for male contraception.

1 Introduction

Androgens are a class of steroids that are responsible for the primary and secondary sex characteristics of the male. In addition, these steroids possess potent anabolic or growth-promoting properties. The general chemical structure of androgens is based upon the androstane C_{19} steroid, consisting of the fused four-ring steroid nucleus (17 carbon atoms, rings A–D) and the two axial methyl groups (carbons 18 and 19) at the A/B and C/D ring junctions. The hormone testosterone (**1**) is the predominant circulating androgen and is produced mainly by the testis. 5α-Dihydrotestosterone (DHT) (**2**) is a 5α-reduced metabolite of testosterone produced in certain androgen target tissues and is the most potent endogenous androgen.

These two steroids and other endogenous androgens influence not only the development and maturation of the male genitalia and sex glands but also affect other tissues such as kidney, liver, and brain. This chapter will discuss the endogenous androgens, synthetic analogs, various anabolic agents, and the androgen antagonists employed in clinical practice or animal husbandry in the United States and elsewhere. Modified androgens that have found use as biochemical or pharmacological tools also are included. More extensive presentations of the topic of androgens, anabolics, and androgen antagonists have appeared in several treatises published over the past three decades.

2 Historical

The role of the testes in the development and maintenance of the male sex characteristics, and the dramatic physiological effects of male castration have been recognized since early time. Berthold was the first to publish in 1849 a report that gonadal transplantation prevented the effects of castration in roosters, suggesting that the testis produced internal secretions exhibiting androgenic effects. However, the elucidation of the molecules of testicular origin responsible for these actions took almost another century. The first report of the isolation of a substance with androgenic activity was made by Butenandt in 1931. The material, isolated in very small quantities from human male urine, was named *androsterone* (**3**). A second weakly androgenic steroid hormone was isolated from male urine in 1934. This substance was named *dehydroepiandrosterone* (**4**) because of its ready chemical transformation and structural similarity to androsterone. A year later, the isolation of the testicular androgenic hormone, testosterone (**1**), was reported, which was 10 times as potent as androsterone in promoting capon comb growth. Shortly after this discovery, the first chemical synthesis of testosterone

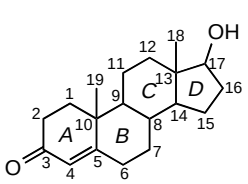

Testosterone (**1**)

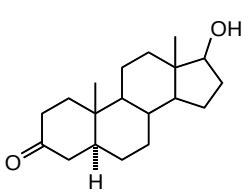

5α-Dihydrotestosterone (**2**)

was reported by Butenandt and Hanisch and confirmed by Ruzicka.

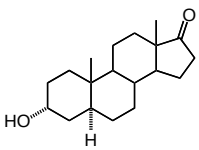

Androsterone (**3**)

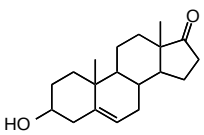

Dehydroepiandrosterone (**4**)

For many years it was believed that testosterone was the active androgenic hormone in man. In 1968, however, 5α-DHT (**2**), also referred to as *stanolone*, was demonstrated to be the active androgen in target tissues, such as the prostate and seminal vesicles and was formed from testosterone by a reductase present in these tissues. Shortly thereafter a soluble receptor protein was isolated and demonstrated to be specific for DHT and related structures.

The anabolic action of the androgens was first documented by Kochakian and Murlin in 1935. In their experiments, extracts of male urine caused a marked retention of nitrogen when injected into dogs fed a constant diet. Soon afterward testosterone propionate was observed to produce a similar nitrogen-sparing effect in humans. Subsequent clinical studies demonstrated that testosterone was capable of causing a major acceleration of skeletal growth and a marked increase in muscle mass. This action on muscle tissue has been referred to more specifically as the *myotrophic effect*.

The first androgenic-like steroid used for its anabolic properties in humans was testosterone. Unfortunately, its use for this purpose was limited by the inherent androgenicity and the need for parenteral administration. 17α-Methyltestosterone (**5**) was the first androgen discovered to possess oral activity, but it too did not show any apparent separation of androgenic and anabolic activity. The promise of finding a useful, orally effective, anabolic agent free from androgenic side effects prompted numerous clinical and biological studies.

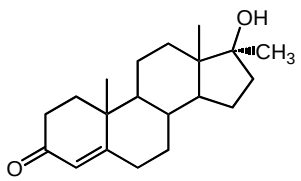

17α-Methyltestosterone (**5**)

3 Endogenous Male Sex Hormones

3.1 Occurrence and Physiological Roles

The hormone testosterone affects many organs in the body. Its most dramatic effects are observed on the primary and secondary sex characteristics of the male. These actions are first manifested in the developing male fetus when the embryonic testis begins to secrete testosterone. Differentiation of the Wolffian ducts into the vas deferens, seminal vesicles, and epididymis occurs under this early androgen influence, as does the development of external genitalia and the prostate. The reductive metabolism of testosterone to 5α-DHT is critical for virilization during this period of fetal development, as dramatically demonstrated in patients with a 5α-reductase deficiency.

At puberty, further development of the sex organs (prostate, penis, seminal vesicles, and vas deferens) is again evident and under the control of androgens. Additionally, the testes now begin to produce mature spermatozoa. Other effects of testosterone, particularly on the secondary sex characteristics, are observed.

Hair growth on the face, arms, legs, and chest is stimulated by this hormone during younger years; in later years, testosterone is responsible for thinning of the hair and recession of the hairline. The larynx develops and a deepening of the voice occurs. The male's skin at puberty thickens, the sebaceous glands proliferate, and the fructose content in human semen increases. Testosterone influences sexual behavior, mood, and aggressiveness of the male at the time of puberty.

In addition to these androgenic properties, testosterone also exhibits anabolic (myotrophic) characteristics. General body growth is initiated, including increased muscle mass and protein synthesis, a loss of subcutaneous fat, and increased skeletal maturation and mineralization. This anabolic action is associated with a marked retention of nitrogen brought about by an increase of protein synthesis and a decrease of protein catabolism. The increase in nitrogen retention is manifested primarily by a decrease in urinary rather than fecal nitrogen excretion and results in a more positive nitrogen balance. For example, intramuscular administration of 25 mg of testosterone propionate twice daily can produce a nitrogen retention to appear within 1 to 3 days, reaching a maximum in about 5 to 8 days. This reduced level of nitrogen excretion may be maintained for at least a month and depends on the patient's nutritional status and diet.

Androgens influence skeletal maturation and mineralization, which is reflected in an increase in skeletal calcium and phosphorus. In various forms of osteoporosis, androgens decrease urinary calcium loss and improve the calcium balance in patients. This effect is not as noticeable in normal patients. Moreover, the various androgen analogs differ markedly in their effects on calcium and phosphorus balance in man. Androgens and their 5β-metabolites (e.g. etiocholanolone) markedly stimulate erythropoiesis, presumably by increasing the production of erythropoietin and by enhancing the responsiveness of erythropoietic tissue to erythropoietin. The effects of androgens on carbohydrate metabolism appear to be minor and secondary to their primary protein anabolic property. The effects on lipid metabolism, on the other hand, seem to be unrelated to this anabolic property. Weakly androgenic metabolites such as androsterone have been found to lower serum cholesterol levels when administered parenterally.

3.2 Biosynthesis

The androgens are secreted not only by the testis but also by the ovary and adrenal cortex. Testosterone is the principal circulating androgen and is formed by the Leydig cells of the testes. Other tissues, such as liver and human prostate, form testosterone from precursors, but this contribution to the androgen pool is minimal. Since dehydroepiandrosterone and androstenedione (see Fig. 2) are secreted by the adrenal cortex and ovary, they indirectly augment the testosterone pool because they can be rapidly converted to testosterone by peripheral tissues.

Plasma testosterone levels for men usually range between 0.61 and 1.1 μg/100 mL and are 5 to 100 times female values. The circulating level of DHT in normal adult men is about one-tenth the testosterone level. Daily testosterone production rates have been estimated to be 4 to 12 mg for young men and 0.5 to 2.9 mg for young women. Although attempts have been made to estimate

the secretion rates for testosterone, these studies are hampered by the number of tissues capable of secreting androgens and the considerable interconversion of the steroids concerned.

The synthesis of androgens in the Leydig cells of the testes is regulated by the gonadotropic hormone, luteinizing hormone (LH), also called *interstitial cell-stimulating hormone* (ICSH). The other pituitary gonadotropin, follicle-stimulating hormone (FSH), acts primarily on the germinal epithelium and is important for sperm development. Both of these pituitary gonadotropins are under the regulation of a decapeptide hormone produced by the hypothalamus. This hypothalamic hormone is luteinizing hormone-releasing hormone (LHRH), also referred to as *gonadotropin-releasing hormone* (GnRH). In adult males, pulsatile secretion of LHRH, and subsequently LH and FSH, occurs at a frequency of 8 to 14 pulses in 24 h. The secretions of these hypothalamic and pituitary hormones are, in turn, regulated by circulating testosterone levels in a negative feedback mechanism. Testosterone will decrease the frequency and amplitude of pulsatile LH secretion, whereas, testosterone and a gonadal peptide, inhibin, are both involved in suppressing the release of FSH.

Our understanding of steroidogenesis in the endocrine organs has advanced considerably during the past three decades, based largely on initial investigations with the adrenal cortex and subsequent studies in the testis and ovary as well. Figure 1 outlines the following sequence of events known to be involved with steroidogenesis in the Leydig cells. LH binds to its receptor located on the surface of the Leydig cell and, via a G-protein-mediated process, activates adenylyl cyclase to result in an increase in intracellular concentrations of cyclic adenosine monophosphate (cAMP). cAMP activates a cAMP-dependent protein kinase,

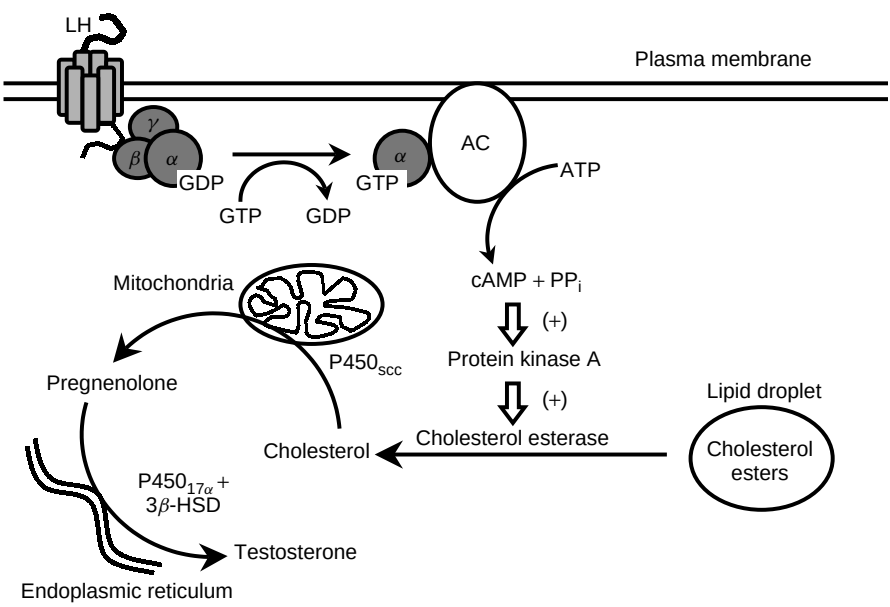

Fig. 1 Cellular events in steroidogenesis in Leydig cell.

which subsequently phosphorylates and activates several enzymes involved in the steroidogenic pathway, including cholesterol esterase and cholesterol side-chain cleavage. Cholesterol esters (present in the cell as a storage form) are converted to free cholesterol by cholesterol esterase, and free cholesterol is translocated to mitochondria. A cytochrome P-450 mixed-function oxidase system, termed *cholesterol side-chain cleavage,* converts cholesterol to pregnenolone. Several nonmitochondrial enzymatic transformations then convert pregnenolone to testosterone, which is secreted.

The conversion of cholesterol (**6**) to pregnenolone (**7**) has been termed the *rate-limiting step* in steroid hormone biosynthesis. The reaction requires NADPH and molecular oxygen and is catalyzed by a cytochrome P-450 enzyme complex termed *cholesterol side-chain cleavage*. This enzyme complex consists of three proteins – cytochrome $P450_{SCC}$ (also called cytochrome P450 11A), adrenodoxin, and adrenodoxin reductase. Three moles of NADPH and oxygen are required to convert one mole of cholesterol into pregnenolone (Fig. 2).

Tracer studies have shown that two major pathways known as the "4-ene" and "5-ene" pathways are involved in the conversion of pregnenolone to testosterone. Both these pathways and the requisite enzymes are shown in Fig. 2. Earlier studies tended to favor the "4-ene" pathway, but more subsequent work has disputed this view and suggests that the "5-ene" pathway is quantitatively more important in man. The unconjugated steroids identified in normal males are listed in Table 1 along with the concentration in micrograms per hundred milliliters. All the intermediates of the "5-ene" pathway were identified,

Tab. 1 Mean concentration of steroids in spermatic venous plasma from five normal males.

Compound	*Concentration [μg 100 mL^{-1}]*
5-Androstene-3β,17α-diol	4.0
5-Androstene-3β,17β-diol	18.5
Dehydroepiandrosterone	2.2
Androstenedione	2.5
Testosterone	74.0
Pregnenolone	4.8
17α-Hydroxypregnenolone	3.9
17α-Hydroxyprogesterone	6.2

but progesterone (**8**), an important intermediate of the "4-ene" pathway, was not found. In addition, sulfate conjugates were present in significant quantities, especially androst-5-ene-3β,17β-diol 3-monosulfate. The data strongly suggest that this intermediate and its unconjugated form constitute an important precursor of testosterone in man. This view, however, was not supported by a kinetic analysis of the metabolism of androst-5-ene-3β,17β-diol (**12**) in man. Further evidence that the predominant pathway appears to be the "5-ene" pathway was provided by *in vitro* studies in human testicular tissues.

Another important step is the conversion of the C-21 steroids to the C-19 androstene derivatives. Whereas the enzymes for side-chain cleavage are localized in mitochondria, those responsible for cleavage of the $C_{17}-C_{20}$ bond ($C_{17}-C_{20}$ lyase) reside in the endoplasmic reticulum of the cell. Early studies implicated 17α-hydroxypregnenolone (**9**) or 17αhydroxyprogesterone (**10**) as an obligatory intermediate in testosterone biosynthesis, and the $C_{17}-C_{20}$ bond was subsequently cleaved by a second enzymatic process to produce the C-19

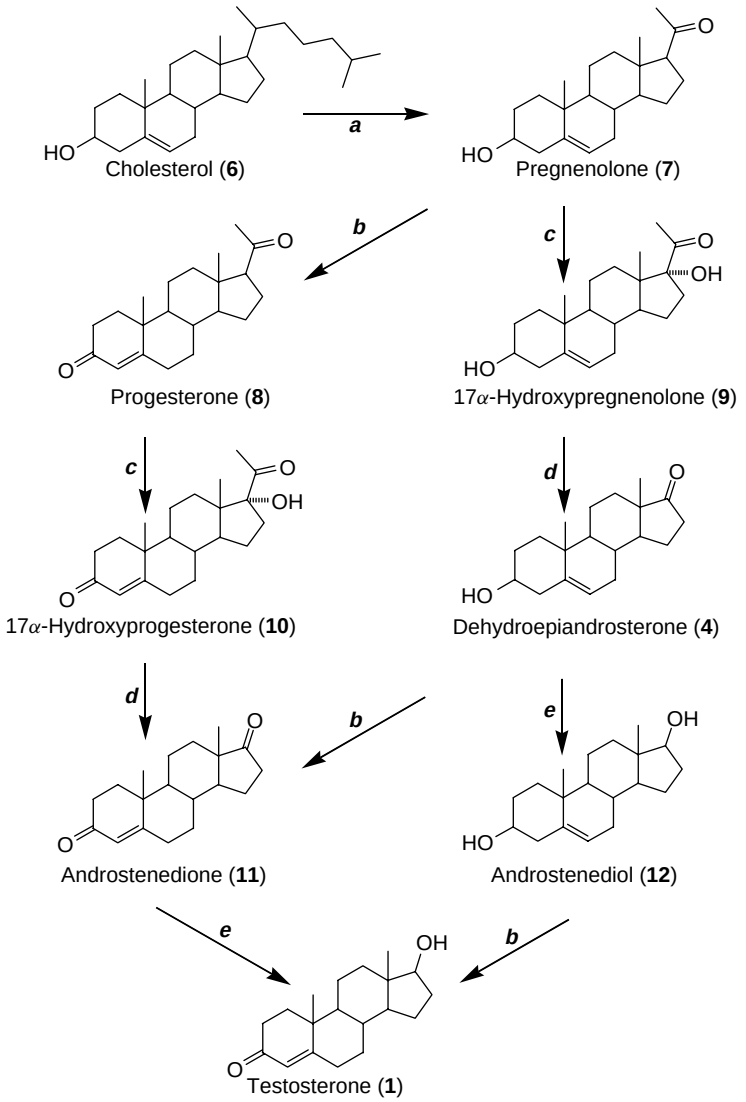

Fig. 2 Enzymatic conversion of cholesterol to testosterone.

androstene molecule. This view of the involvement of two separate enzymes in the conversion of C-21 to C-19 steroids existed until the purification of the proteins in the 1980s. The 17α-hydroxylase/17,20-lyase cytochrome P450 (abbreviated cytochrome P450 17 or cytochrome $P450_{17\alpha}$) was first isolated from neonatal pig testis microsomes, and this cytochrome $P450_{17\alpha}$ possessed both 17α-hydroxylase and 17,20-lyase activity when reconstituted with cytochrome P450 reductase and phospholipid. Identical full-length human cytochrome $P450_{17\alpha}$ cDNA sequences were independently isolated and reported in 1987.

Two additional enzymes are necessary for the formation of testosterone from dehydroepiandrosterone. The first is the 3β-hydroxysteroid dehydrogenase/$\Delta^{4,5}$-isomerase complex, which catalyzes the oxidation of the 3β-hydroxyl group to the 3-ketone and the isomerization of the double bond from C_5-C_6 to C_4-C_5. Again, these processes were originally thought to involve two different enzymes, but purification of the enzymatic activity demonstrated that a single enzyme catalyzes both reactions. The final enzyme in the pathway is the 17β-hydroxysteroid dehydrogenase, which catalyzes the reduction of the 17-ketone to the 17β-alcohol.

3.3 Absorption and Distribution

Although considerable research has been devoted to the biochemical mechanism of action of the natural hormones and the synthesis of modified androgens, little is known about the absorption of these substances. It is well recognized that a steroid hormone might have high intrinsic activity but exerts little or no biological effects because its physicochemical characteristics prevent it from reaching the site of action. This is particularly true in humans, where slow oral absorption or rapid inactivation may greatly reduce the efficacy of a drug. Even though steroids are commonly given by mouth, little is known of their intestinal absorption. One study in rats showed that androstenedione (**11**) was absorbed better than testosterone or 17α-methyltestosterone, and conversion of testosterone to its acetate enhanced absorption. Results with other steroids indicated that lipid solubility was an important factor for intestinal absorption. This may explain the oral activity of certain ethers and esters of testosterone.

Once in the circulatory system by either secretion from the testis or absorption of the administered drug, testosterone and other androgens will reversibly associate with certain plasma proteins, the unbound steroid being the biologically active form. The extent of this binding is dependent on the nature of the proteins and the structural features of the androgen.

The first protein to be studied was albumin, which exhibited a low association constant for testosterone and bound less polar androgens such as androstenedione to a greater extent.

α-Acid glycoprotein (AAG) was shown to bind testosterone with a higher affinity than albumin. A third plasma protein that binds testosterone is corticosteroid-binding α-globulin (CBG). However under normal physiological conditions these plasma proteins are not responsible for extensive binding of androgens in plasma.

A specific protein termed *sex steroid binding β-globulin* (SBG) or testosterone estradiol binding globulin (TEBG) was found in plasma that bound testosterone with a very high affinity. The SBG–sex hormone complex serves several functions, such as transport or carrier system in the bloodstream, storage site or reservoir for the hormones, and protection of the hormone from metabolic transformations. SBG has been purified and contains high affinity, low capacity binding sites for the sex hormones. Dissociation constants of approximately 1×10^{-9} M have been reported for the binding of testosterone and estradiol to SBG and are 2 orders of magnitude less than values reported for the binding of the hormone to the cytosolic receptor protein. The plasma levels of SBG are regulated by the thyroid hormones and remain fairly constant throughout adult

life in both the male and female. SBG is not present in the plasma of all animals. For example, SBG-like activity is notably absent in the rat, and testosterone may be bound in the rat plasma to CBG.

The specificity of the binding of steroids to human SBG has been extensively studied. The presence of a 17β-hydroxyl group is essential for binding to SBG. In addition to testosterone, DHT, 5α-androstane-3β,17β-diol (**19**), and 5α-androstane-3α,17β-diol (**20**) bind with high affinity, and these steroids compete for a common binding site. Binding to SBG is decreased by 17α-substituents such as 17α-methyl and 17α-ethinyl moieties and by unsaturation at C-1 or C-6. Also, 19-nortestosterone derivatives have lower affinity. SBG has been purified to homogeneity by affinity chromatography using a DHT-agarose adsorbent.

Another extracellular carrier protein exhibiting high affinity for testosterone, found in seminiferous fluid and the epididymis, originates in the testis and is called *androgen binding protein* (ABP). This protein is produced by the Sertoli cells on stimulation by FSH and has very similar characteristics to those of plasma SBG produced in the liver.

The absorption of androgens and other steroids from the blood by target cells was usually assumed to occur by a passive diffusion of the molecule through the cell membrane. However, studies in the early 1970s using tissue cultures or tissue slices suggested entry mechanisms for the steroids. Estrogens, glucocorticoids, and androgens exhibit a temperature-dependent uptake into intact target cells, suggesting a protein-mediated process. Among the androgens, DHT exhibited a greater uptake than testosterone in human prostate tissue slices, and it was found that estradiol or androstenedione interfered with this uptake mechanism. In addition, cyproterone competitively inhibited androstenedione, testosterone, and DHT entry, whereas cyproterone acetate enhanced the uptake of these androgens. Little is known about the exit of steroids from target cells; the only reported research has dealt with an active transport of glucocorticoids out of cells.

3.4 Metabolism

For decades, the primary function of metabolism was thought to be the inactivation of testosterone, the increase in hydrophilicity, and the mechanism for the excretion of the steroid into the urine. However, the identification of metabolites of testosterone formed in peripheral tissues and the potent and sometimes different biological activities of these products emphasized the importance of the metabolic transformations of androgens in endocrinology. Two active metabolites of testosterone have received considerable attention – the reductive metabolite 5α-DHT and the oxidative metabolite estradiol (Fig. 3).

3.4.1 Reductive Metabolism

The metabolism of testosterone in a variety of *in vitro* and *in vivo* systems has been reviewed. The principal pathways for reductive metabolism of testosterone in man appear in Fig. 4. Human liver produces a number of metabolites, including androstenedione (**11**), 3β-hydroxy-5α-androstan-17-one (**16**), 5α-androstane-3β,17β-diol (**19**), and 5α-androstane-3α,17β-diol (**20**). Human adrenal preparations, on the other hand, gave 11β-hydroxytestosterone as the major metabolite. (Intestinal metabolism of testosterone is similar to transformations in the

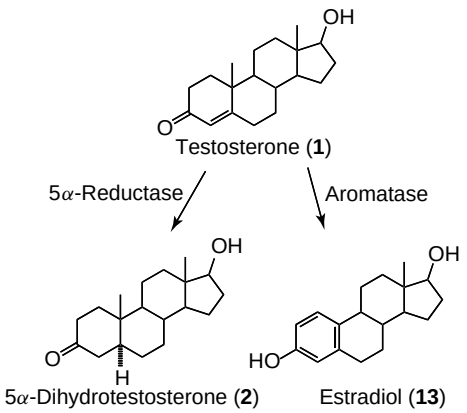

Fig. 3 Enzymatic conversion of testosterone to biologically active metabolites, 5α-dihydrotestosterone and estradiol.

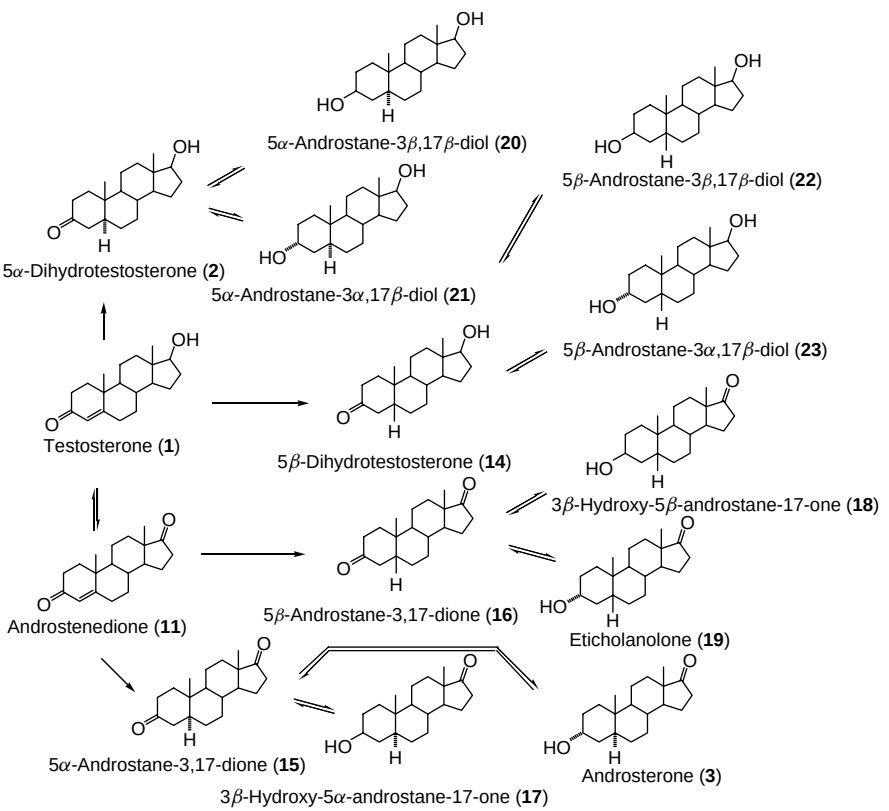

Fig. 4 Reductive metabolites of testosterone.

liver; the major metabolite in lung is androstenedione).

Studies on testosterone metabolism since the late 1960s have centered on steroid transformations by prostatic tissues. Normal prostate, benign prostatic hypertrophy (BPH), and prostatic carcinoma all contain 3α-, 3β-, and 17β-hydroxysteroid dehydrogenases, and 5α- and 5β-reductases, capable of converting testosterone to various metabolites. Prostatic carcinoma metabolizes testosterone more slowly than does BPH or normal prostate. In addition, increased levels of androgens are found in BPH. K. D. Voigt et al. have extensively studied *in vivo* metabolic patterns of androgens in patients with BPH. Tritiated androgens were injected intravenously into these patients 30 min before prostatectomy, and prostatic tissue, tissue from surrounding skeletal muscle, and blood plasma were analyzed for metabolites. The major metabolite of testosterone found in BPH tissues was DHT, with minor amounts of diols isolated. Skeletal muscle and plasma contained primarily unchanged testosterone.

Table 2 lists the urinary metabolites that have been identified following the administration of testosterone to humans (see Fig. 4). These products are excreted as such or in the form of their glucuronide or sulfate conjugates. Androsterone (**3**) and etiocholanolone (**18**), the major urinary metabolites, are excreted predominantly as glucuronides, and only about 10% as sulfates. These conjugates are capable of undergoing further metabolism. Testosterone glucuronide, for example, is metabolized differently from testosterone in man, giving rise mainly to 5β-metabolites. Only a relatively small amount of the urinary 17-ketosteroids is derived from testosterone metabolism. In man at least 67% and in women about 80% or more of the urinary 17-ketosteroids are metabolites of adrenocortical steroids. This explains why a significant increase in testosterone secretion associated with various androgenic syndromes does not usually lead to elevated levels of 17-ketosteroid excretion.

Tab. 2 Urinary metabolites of testosterone.

Metabolite	*Approximate Conversion [%]*
Androsterone	25–50
Etiocholanolone	
5β-Androstane-3α,17β-diol	2
5α-Androstane-3α,17β-diol	1
5α-Androstan-3β-ol-17-one	1
Androst-16-en-3α-ol	0.4
3α,18-Dihydroxy-5β-androstan-17-one	0.3
3α,7β-Dihydroxy-5β-androstan-17-one	Trace
11β-Hydroxytestosterone	Trace
6α-, 6β-Hydroxytestosterone	Trace

Although androsterone and etiocholanolone are the major excretory products, the exact sequence whereby these 17-ketosteroids arise is still not clear. Studies with radiolabeled androst-4-ene-3β,17β-diol and the epimeric 3α-diol in humans showed that oxidation to testosterone was necessary before reduction of the A-ring. Moreover, in rats 5β-androstane-3α,17β-diol (**22**) was the major initial liver metabolite, but this decreased with time with the simultaneous increase of etiocholanolone. This formation of saturated diols agrees with studies using human liver and provides evidence that the initial step in testosterone metabolism is reduction of the α,β-unsaturated ketone to a mixture of diols followed by oxidation to the 17-ketosteroids.

Until 1968, it was generally thought that the excretory metabolites of testosterone were physiologically inert. Subsequent work has shown, however, that etiocholanolone has thermogenic effects when administered to man. Moreover, the hypocholesterolemic effects of parenterally administered androsterone have been described.

The conversion of testosterone to DHT by 5α-reductase has major importance in the mechanism of action of the hormone. This enzymatic activity has been found in the endoplasmic reticulum and in the nuclear membrane of androgen-sensitive cells. In addition, the levels of 5α-reductase are under the control of testosterone and DHT; 5α-reductase activity decreases after castration and can be restored to normal levels of activity with testosterone or DHT administration.

Early biochemical studies of 5α-reductase were performed using a microsomal fraction from rat ventral prostate. The irreversible enzymatic reaction catalyzed by 5α-reductase requires NADPH as a cofactor, which provides the hydrogen for carbon-5. The 5α-reductase from rat ventral prostate tissues exhibited a broad range of substrate specificity for various C_{19} and C_{21} steroids; this broad specificity was also observed in inhibition studies. However, more detailed studies of the enzyme were limited because of the extreme hydrophobic nature of the protein, its instability upon isolation, and its low concentrations in androgen-dependent tissues.

Investigations on the molecular biology of 5α-reductase resulted in the demonstration of two different genes and two different isozymes of the enzyme. The first cDNA isolated and cloned that encoded 5α-reductase was designated type 1, and the second was designated type 2. The gene encoding type 1 is located on chromosome 5, while the gene encoding type 2 is located on chromosome 2. The two human 5α-reductases have approximately 60% sequence homology. These two isozymes differ in their biochemical properties, tissue location, and function. Type 1 5α-reductase exhibits an alkaline pH optimum (6–8.5) and has micromolar affinities for steroid substrates. On the other hand, type 2 5α-reductase has a sharp pH optimum at 4.7–5.5, has higher affinity (lower apparent K_m) for testosterone, and is more sensitive to inhibitors than the type 2 isozyme. The type 2 isozyme is expressed primarily in androgen target tissues, the liver expresses both types, and the type 1 form is expressed in various peripheral tissues. Type 2 5α-reductase appears to be essential for masculine development of the fetal urogenital tract and the external male phenotype, whereas the type 1 isozyme is primarily a catabolic enzyme. In certain cases of human male pseudohermaphroditism, mutations in the type 2 5α-reductase gene are observed and results in significant decreases in DHT levels needed for virilization.

3.4.2 Oxidative Metabolism

Another metabolic transformation of androgens leading to hormonally active compounds is their conversion to estrogens (see Table 3). Estrogens are biosynthesized in the ovaries and placenta and, to a lesser extent, in the testes, adrenals, and certain regions of the brain. The enzyme complex that catalyzes this biosynthesis is referred to as *aromatase*, and the enzymatic activity was first identified by Ryan in the microsomal fraction from human placental tissue. The elucidation of the mechanism of the aromatization reaction began in the early 1960s and continues to receive extensive study. It is a

cytochrome P-450 enzyme complex and requires 3 mol of NADPH and 3 mol of oxygen per mole of substrate. Aromatization proceeds via three successive steps, with the first two steps being hydroxylations. 19-Hydroxyandrostenedione (**23**) was a more active precursor of estrone (**25**) than the substrate androstenedione led to its postulated role in estrogen biosynthesis. This report and numerous studies that followed led to the currently accepted pathway for aromatization, as shown in Fig. 5.

The first two oxidations occur at the C_{19} position, producing the 19-alcohol (**23**) and then the 19-*gem*-diol (**24**), originally isolated as the 19-aldehyde (**25**). The exact mechanism of the last oxidation remains to be fully determined. The final oxidation results in the stereospecific elimination of the 1β and 2β-hydrogen atoms and the concerted elimination of the oxidized C_{19} moiety as formic acid. Investigations using $^{18}O_2$ and isotopically labeled steroidal intermediates failed to show incorporation of the 2β-hydroxyl group into formic acid

Tab. 3 Relative substrate activity in aromatization.

Substrate	***Activity [%]***
C_{19} –Steroids	
Androst-4-ene-3,17-dione	100
19-Hydroxyandrost-4-ene-3,17-dione	184, 133
17β-Hydroxyandrost-4-en-3-one	100
3β-Hydroxyandrost-5-en-17-one	66
5α-Androst-1-ene-3,17-dione	0
5α-Androstane-3,17-dione	0
1α-Hydroxyandrost-4-ene-3,17-dione	0
17β-Hydroxy-1α-methylandrost-4-ene-3-one	0
2β-Hydroxyandrost-4-ene-3,17-dione	15
2α-Hydroxyandrost-4-ene-3,17-dione	0
17β-Hydroxy-2β-methylandrost-4-en-3-one	0
11β-Hydroxyandrost-4-ene-3,17-dione	0
11α-Hydroxyandrost-4-ene-3,17-dione	100
17β-Hydroxy-17α-methylandrost-4-en-3-one	44
6β-Hydroxyandrost-4-ene-3,17-dione	21
6α-Fluoro-17β-hydroxyandrost-4-ene-3,17-dione	0
6β-Fluoro-17β-hydroxyandrost-4-ene-3,17-dione	0
9α-Fluoroandrosta-1,4-diene-3,17-dione	55
Androsta-1,4-diene-3,17-dione	22, 35
Androsta-4,6-diene-3,17-dione	0
Androsta-1,4,6-triene-3,17-dione	0
Androst-4-ene-3,11,17-trione	0
C_{18} Steroids	
Estr-4-ene-3,17-dione	21
17β-Hydroxyestr-4-en-3-one	20
17β-Hydroxy-5α,10β-estr-3-one	0
C_{12} Steroids	
Pregn-4-ene-3,20-dione	0
17α,19,21-Trihydroxypregn-4-ene-3,20-dione	0

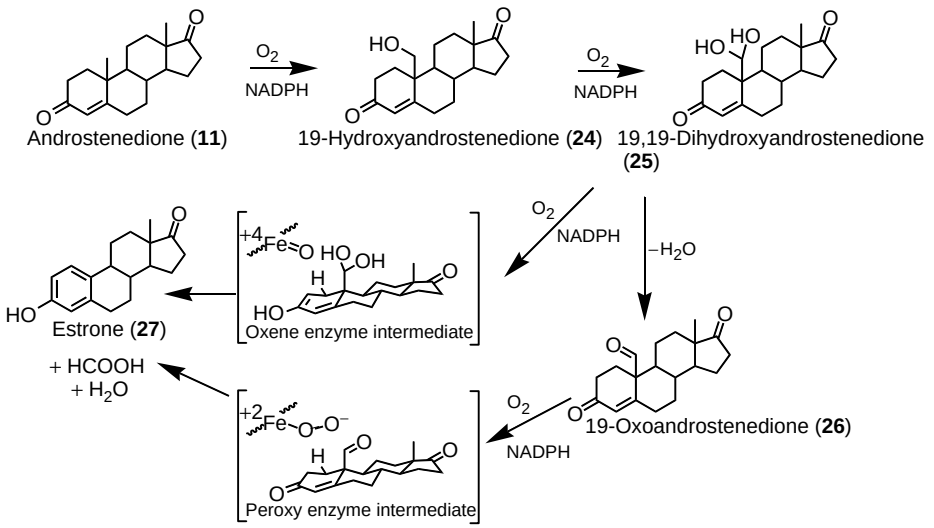

Fig. 5 Aromatization of androgens.

Tab. 4 Comparison of the androgenic and myotrophic activities of testosterone derivatives in the chick comb and castrated male rat assays following subcutaneous administration.

Testosterone modification	*Increase in weight [%]*		
	Chick comb	*Rat ventral prostate*	*Rat levator ani*
None	100	100	100
19-Nor	81	42	90
7α-Methyl	11	97	135
7α-Methyl-19-nor	75	218	226
14-Dehydro	128	54	8
14-Dehydro-19-nor	320	69	133
14-Dehydro-7α-methyl-19-nor	435	352	330

under enzymatic or nonenzymatic condition and did demonstrate that the oxygen atoms from the first and third oxidation steps are incorporated into formic acid. These results led to the proposal that the last oxidation step is a peroxidative attack at the C_{19} position. Recent computational chemistry approaches of this third step of aromatase catalysis using density functional theory (DFT) and molecular mechanical methods support the final catalytic step involving the 19-*gem*-diol substrate, a cytochrome P450 oxene intermediate, 1β-hydrogen atom abstraction, and release of formic acid.

Incubation of a large number of testosterone analogs with human placental tissue has provided some insight into the structural requirements for aromatization (see Table 4). Whereas androstenedione was converted rapidly to estrone, the 1-dehydro and 19-nor analogs were

metabolized slowly, and the 6-dehydro isomer and saturated 5α-androstane-3,17-dione remained unchanged. Hydroxyl and other substituents at lα, 2β, and 11β interfered with aromatization, whereas similar substituents at 9α and 11α seemingly had no effect. Of the stereoisomers of testosterone, only the 8β,9β,10β-isomer is aromatized, in addition to compounds having the normal configuration (8β, 9α, 10β). Thus, the substrate specificity of aromatase appears to be limited to C_{19} steroids with the 4-en-3-one system. Inhibition studies with various steroids have provided additional insights into the structural requirements for the enzyme; steroidal aromatase inhibitors are described later in Sect. 6.3.3.

Recent research in aromatase has focused on the biochemistry, molecular biology, and regulation of the aromatase protein. Aromatase is a membrane-bound cytochrome P-450 monooxygenase consisting of two proteins – aromatase cytochrome P-450 (P-450_{arom}) and NADPH-cytochrome P-450 reductase. Cytochrome P-450_{arom} is a heme protein that binds the steroid substrate and molecular oxygen and catalyzes the oxidations. The reductase is a flavoprotein, is found ubiquitously in endoplasmic reticulum, and is responsible for transferring reducing equivalents from NADPH to cytochrome P-450_{arom}. Purification of cytochrome P-450_{arom} proved to be very difficult because of its membrane-bound nature, instability, and low tissue concentration. The combination of hydrophobic chromatography using phenyl Sepharose, nonionic detergent, and the presence of micromolar amounts of substrate androstenedione yielded a highly purified and active cytochrome P-450_{arom}, with the highest specific content of 11.5 nmol of cytochrome P-450 per milligram of protein reported. Reconstitution of this cytochrome P-450_{arom} with NADPH-cytochrome P-450 reductase and phospholipid resulted in complete conversion of androstenedione to estrone, thus, demonstrating that one cytochrome P-450 protein catalyzes all three oxidation steps. Knowledge of the molecular biology of aromatase has advanced greatly in the past five years. A full-length cDNA complementary to mRNA encoding cytochrome P-450_{arom} was sequenced and the open reading frame encodes a protein of 503 amino acids. This cDNA sequence was inserted into COS1 monkey kidney cells, and aromatase mRNA and aromatase enzymatic activity were detected in transfected cells. The entire human cytochrome $P450_{arom}$ gene is greater than 70 kb in size and is located on chromosome 15. Clones have been utilized to examine the regulation of aromatase in ovarian, adipose, and breast tissues.

The metabolism of androgens by the mammalian brain has also been investigated under *in vitro* conditions. In rat brain, testosterone is converted to DHT, androstenedione, 5α-androstane-3,17-dione, and 5α-androstane3β,17β-diol. The aromatization of androgens to estrogens was also found to occur in the hypothalamus and the pituitary gland. The full significance of these metabolites on various neuroendocrine functions, such as regulation of gonadotropin secretion and sexual behavior, is not yet fully understood.

3.5 Mechanism of Action

It would indeed be impossible to explain all the varied biological actions of testosterone by one biochemical mechanism. Androgens, as well as the other steroid hormones adrenocorticoids, estrogens, and progestins, exert potent physiological effects on sensitive tissues and yet are

present in the body in only extremely low concentrations (e.g. 0.1–1.0 nM). The majority of investigations concerning the elucidation of the mechanisms of action of androgens have dealt with actions in androgen-dependent tissues and, in particular, the rat ventral prostate. The results of these studies indicate that androgens primarily act in regulating gene expression and protein biosynthesis by formation of a hormone–receptor complex, analogous to the mechanisms of action of estrogens and progestins. Extensive research activities directed at elucidation of the general mechanism of steroid hormone action have been performed for over three decades, and several reviews have appeared on this subject.

Investigations of a selective uptake of androgens by target cells performed in the early 1960s were complicated by low specific activity of the radiolabeled hormones and the rapid metabolic transformations. Nonetheless, it was noted that target cells retained primarily unconjugated metabolites, whereas conjugated metabolites were present in nontarget cells such as blood and liver. With the availability of steroids with high specific activity, later studies demonstrated the selective uptake and retention of androgens by target tissues. In addition, DHT was found to be the steroidal form selectively retained in the nucleus of the rat ventral prostate. This discovery led to the current concept that testosterone is converted by 5α-reductase to DHT, which is the active form of cellular androgen in androgen-dependent tissues.

The rat prostate has been the most widely examined tissue, and current hypotheses on the mode of action of androgens are based largely on these studies (see Fig. 6). The lipophilic steroid hormones are carried in the bloodstream, with the majority of the hormones reversibly bound to serum carrier proteins and a small amount of free steroids. The androgens circulating in the bloodstream are the sources of steroid hormone for androgen action in target tissues. Testosterone, synthesized

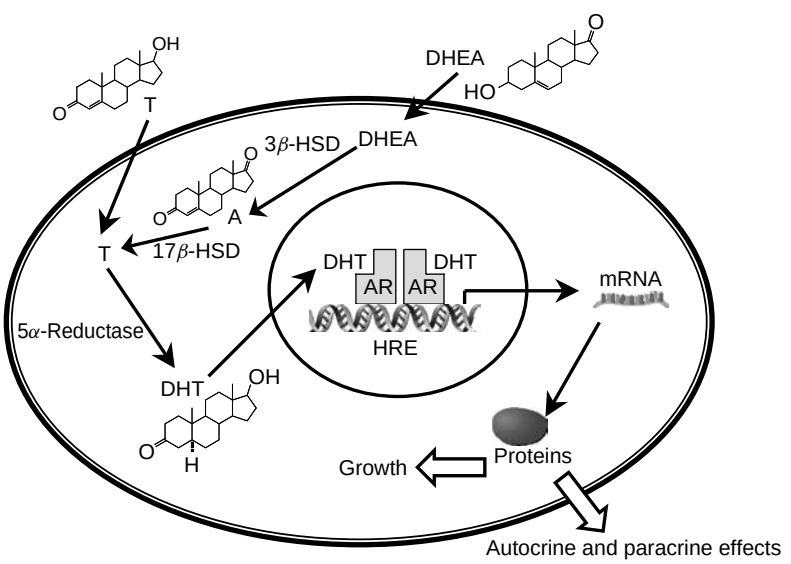

Fig. 6 Mechanism of action of 5α-dihydrotestosterone.

and secreted by the testis, is the major androgen in the blood stream and the primary source of androgen for target tissues in men. Dehydroepiandrosterone (DHEA) and androstenedione also circulate in the bloodstream and are secreted by the adrenal gland under the regulation of adrenocorticotropic hormone (ACTH). DHEA and androstenedione supplement the androgen sources in normal adult men, but these steroids are the important circulating androgens in women. The free circulating androgens steroids passively diffuse through the cell membrane and are converted to the active androgen 5α-DHT within the target cells.

The androgens act on target cells to regulate gene expression and protein biosynthesis via the formation of steroid–receptor complexes. Those cells sensitive to the particular steroid hormone (referred to as *target cells*) contain high-affinity steroid receptor proteins capable of interacting with the steroid. The binding of DHT with the receptor protein is a necessary step in the mechanism of action of the steroid in the prostate cell. Early studies suggested that the steroid receptor proteins were located in the cytosol of target cells and, following formation of the steroid–receptor complex, the steroid–receptor complex translocated into the nucleus of the cell. More recent investigations on androgen action indicate that active, unoccupied receptor proteins are present only in the nucleus of the cell. This nuclear localization of receptor is also observed for estrogens and progestins, while the majority of the glucocorticoid receptor is located in the cytoplasm. In the current model, DHT is formed in the cytoplasm, then enters the nucleus of the cell, and binds to the nuclear steroid receptor protein.

The binding of DHT to the nuclear androgen receptor (AR) initiates a conformational change or activation of the steroid nuclear–receptor complex and results in the formation of a homodimer. The homodimer interacts with particular regions of the cellular DNA, referred to as *androgen-responsive elements* (AREs), and with various nuclear transcriptional factors. Binding of the nuclear steroid–receptor complex to DNA initiates transcription of the DNA sequence to produce messenger RNA (mRNA). Finally, the elevated levels of mRNA lead to an increase in protein synthesis in the endoplasmic reticulum; these proteins include enzymes, receptors, and/or secreted factors that subsequently result in the steroid hormonal response regulating cell function, growth, and differentiation.

Extensive structure–function studies on the androgen receptor or AR have identified regions critical for hormone action. The androgen receptor is encoded by the AR gene located on the X chromosome, and the AR gene is comprised of 8 exons. The human AR contains approximately 900 to 920 amino acids, and the exact length varies because of polymorphisms in the NH_2 terminal of the protein. The primary amino acid sequences of AR, as well as of the various steroid hormone receptors, were deduced from cloned complementary DNAs (cDNAs). The calculated molecular weight of AR is approximately 98 000 kDa based upon amino acid composition; however, the AR is a phosphoprotein and migrates higher at approximately 110 kDa in sodium dodecyl sulfate (SDS) gel electrophoresis. The steroid receptor proteins are part of a larger family of nuclear receptor proteins that also include receptors for vitamin D, thyroid hormones, and retinoids. The overall structure of the androgen receptor (shown in Fig. 7) has

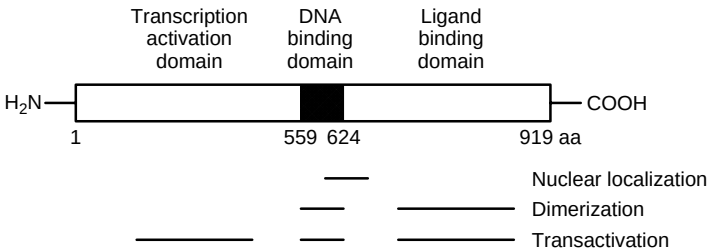

Fig. 7 Schematic diagram of androgen receptor.

strong similarities to the other steroid hormone receptors, with the proteins containing regions that bind to the DNA and bind to the steroid hormone ligand. A high degree of homology (sequence similarities) in the steroid receptors is found in the DNA binding region that interacts with the hormone response elements (HREs). The DNA binding region is rich in cysteine amino acids and chelate zinc ions, forming fingerlike projections called *zinc fingers* that bind to the DNA. The hormone binding domain (or ligand binding domain, LBD) is located on the COOH terminal of the protein. Structure–function studies of cloned receptor proteins also identify regions of the molecules that are important for nuclear localization of the receptor, receptor dimerization, interactions with nuclear transcriptional factors, and activation of gene transcription. Importantly, two regions of the androgen receptor protein are identified as transcriptional activation domains. The domain on the NH_2-terminal region may interact with both coactivators and corepressors, and the COOH-terminal domain initiates transcriptional activation only upon binding of an agonist such as 5α-DHT. The interactions necessary for formation of the steroid–receptor complexes and subsequent activation of gene transcription are complicated, involve multiple protein partners referred to as *coactivators*, and leave many unanswered questions.

Although the tertiary structure of the entire androgen receptor has not been determined, the crystallographic structure of the LBD has recently been reported. The AR LBD consists of an α-helical sandwich, similar to the LBDs reported for other nuclear receptors, and AR LBD contains only 11 helices (no helix 2) and four short β-strands (Fig. 8). Minor differences in the two reported crystallographic structures are likely due to limits of experimental resolution, differences in data interpretation, and the use of different ligands for crystallization. The endogenous ligand DHT (**2**) interacts with helices 3, 5, and 11, and the DHT-bound AR LBD has a single, continuous helix 12 (207). Similar interactions are observed with metribolone (methyltrienolone, **53**); however, helix 12 is split into two shorter helical segments.

Additional information on receptor structure–function has been obtained by analyzing androgen receptor mutations in patients with various forms of androgen resistance and abnormal male sexual development. Two polymorphic regions have been identified in the NH_2-terminal region, encoding a polyglycine repeat and a polyglutamate tract. Currently, these polymorphic regions have not been shown to significantly alter AR levels, stability, or transactivation. These repeats are useful in pedigree analysis of patients. Mutations in the androgen receptor have been identified

Fig. 8 Ribbon diagram of androgen receptor ligand binding domain (LBD) bound with 5α-dihydrotestosterone. Atomic coordinates were obtained from the Protein Data Bank (PDB ID code 1I37; www.rcsb.org) and displayed using ViewerLite 4.2, Accerlys Inc. 5α-Dihydrotestosterone is illustrated as a stick structure behind Helix 12.

in patients with either partial or full androgen insensitivity syndrome (AIS), with the majority of the mutations identified in exons 4 through 8 encoding the DNA binding domain and the hormone binding domain. Finally, studies with the human LNCaP prostate cancer cell line have provided interesting results regarding receptor protein structure and ligand specificity. The LNCaP cells exhibited enhanced proliferation in the presence of androgens, but these cells also unexpectedly proliferated in the presence of estrogens, progestins, cortisol, or the antiandrogen flutamide. Analysis of the cDNA for the LNCaP androgen receptor revealed that a single base mutation in the LBD was present and resulted in the increased affinity for progesterone and estradiol. The crystallographic structures of the LBD with the T877A mutation confirm that the mutated AR LBD can accommodate larger structures at the C-17 position.

The ultimate action of androgens on target tissues is the stimulation of cellular growth and differentiation through regulation of protein synthesis, and numerous androgen-inducible proteins have been identified. One of the prominent androgen-inducible proteins is prostate-specific antigen (PSA), a serine protease expressed by secretory prostate epithelial cells and utilized as blood test in screening for possible prostate diseases such as prostate cancer. Three AREs have been identified in the promoter regions of the PSA gene. Another androgen-regulated gene examined extensively in rats is the gene encoding the protein probasin, a 20-kDa secretory protein from the rat dorsolateral prostate structurally similar to serum globulins. Other proteins induced by androgens include spermine-binding protein, keratinocyte growth factor or fibroblast growth factor (KGF or FGF-7), androgen-induced growth factor (AIGF or FGF-8), nerve-growth factor, epidermal growth factor, c-myc, protease D, β-glucuronidase, and α_{2u}-globulin. Studies of these proteins suggest that the androgens act by enhancing transcription and/or translation of specific RNAs for the proteins. Also, the androgen receptor represses gene expression of certain proteins

such as glutathione S-transferase, TRMP-2 involved in apoptosis, and cytokines such as interleukin-4, interleukin-5, and γ-interferon.

While most biochemical studies focused on the rat ventral prostate, some researchers began to investigate the presence of cellular receptor proteins in other androgen-sensitive tissues. Androgen receptors have been reported in seminal vesicles, sebaceous gland, testis, epididymis, kidney, submandibular gland, pituitary and hypothalamus, bone marrow, liver, and androgen-sensitive tumors. Although DHT is the active androgen in rat ventral prostate, it is not the only functioning form in other androgen-sensitive cells. In ventral prostate and seminal vesicles, DHT is readily formed. It is metabolized only slowly, however, and therefore can accumulate and bind to receptors. Also, comparison of binding kinetics for testosterone and DHT demonstrated that testosterone dissociates faster, implying extended retention of DHT by the androgen receptor. In other tissues, such as brain or kidney, DHT is not readily formed and is metabolized quickly compared to testosterone. Species variations have also been demonstrated. The most striking example is the finding that 5α-androstane-3α, 17α-diol interacts specifically with cytosolic receptor protein from dog prostate and may be the active androgen in this species. Apparently, the need for a 17β-hydroxyl is not essential in all species.

Thus, current findings indicate that androgen receptor proteins vary in steroid specificity among different tissues from the same species as well as among different species. Nevertheless, the basic molecular mechanism of action of the androgens in androgen-sensitive tissues is consistent with the results of the studies on rat ventral prostate.

The manner whereby the androgens exert their anabolic effects has not been as extensively studied. The conversion of testosterone to DHT has been shown to be insignificant in skeletal and levator ani muscles, suggesting that the androgen-mediated growth of muscle is due to testosterone itself. Classical steroid receptors for testosterone are found in the cytoplasm of the levator ani and quadriceps muscles of the rat. Unlike prostate receptor protein, DHT had a lower affinity than testosterone for this protein. Androgen receptors have also been identified in other muscle tissues as well, including cardiac muscle.

4 Synthetic Androgens

4.1 Current Drugs on the Market

See Table 5.

4.2 Therapeutic Uses and Bioassays

The primary uses of synthetic androgens are the treatment of disorders of testicular function and of cases with decreased testosterone production. Several types of clinical conditions result from testicular dysfunction. Information on the biochemistry and mechanism of action of testosterone that has accumulated over the past 20 years has greatly aided in the elucidation of the underlying pathophysiology of these diseases.

Hypogonadism arises from the inability of the testis to secrete androgens and can be caused by various conditions. These hypogonadal diseases can, in many cases, result in disturbances in

Tab. 5

Generic name (structure)	*Trade name*	*US manufacturer*	*Chemical class*	*Dose*
Testosterone enanthate (**31**)	Delatestryl	BTG Pharm	Androstane	Injection: 200 mg mL^{-1}
Testosterone cypionate (**33**)	Depo-Testosterone	Pharmacia	Androstane	Injection: 100 mg mL^{-1}; 200 mg mL^{-1}
Testosterone pellets (**1**)	Testopel	Bartor Pharmacal	Androstane	Pellets: 75 mg
Testosterone transdermal system (**1**)	Testoderm	Alza	Androstane	Transdermal: 10 mg; 15 mg
	Testoderm TTS	Alza	Androstane	Transdermal: 328 mg
	Testoderm with Adhesive	Alza	Androstane	Transdermal: 15 mg
	Androderm	Watson Pharma	Androstane	Transdermal: 12.2 mg; 24.3 mg
Testosterone gel (**1**)	AndroGel 1%	Unimed Pharm	Androstane	Gel: 1% testosterone
Methyltestosterone (**5**)	Methyltestosterone	Various suppliers	Androstane	Tablets: 10 mg; 25 mg; Tablets (buccal): 10 mg
	Testred	ICN Pharm	Androstane	Capsules: 10 mg
	Android	ICN Pharm	Androstane	Capsules: 10 mg
	Virilon	Star	Androstane	Capsules: 10 mg
	Virilon IM	Star	Androstane	Injection: 200 mg mL^{-1}
Fluoxymesterone (**36**)	Halotestin	Pharmacia	Androstane	Tablets: 2 mg; 5 mg; 10 mg
	Fluoxymesterone	Various suppliers	Androstane	Tablets: 10 mg
Danazol (**77**)	Danazol	Various suppliers	Androstane	Capsules: 50 mg; 100 mg; 200 mg
	Danocrine	Sanofi-Synthelabo	Androstane	Capsules: 50 mg; 100 mg; 200 mg
Testolactone (**38**)	Teslac	Bristol-Myers Squibb	Androstane	Tablets: 50 mg

sexual differentiation and function and/or sterility. Primary hypogonadism is the result of a basic disorder in the testes, while secondary hypogonadism results from the failure of pituitary and/or hypothalamic release of gonadotropins and thus diminished stimulation of the testis. Usually primary hypogonadism is not recognized in early childhood (with the exception of cryptorchidism) until the expected time of puberty. This testosterone deficiency is corrected by androgen treatment for several months, at which time the testes are evaluated for possible development. Long-term therapy is necessary if complete testicular failure is present. Patients with Klinefelter's syndrome, a disease in which a genetic male has an extra X chromosome, have low testosterone levels and can also be treated by androgen replacement.

Male pseudohermaphroditism are disorders in which genetically normal men do not undergo normal male development. One type, testicular feminization, is observed in patients who have normal male XY chromosomes, but the male genitalia and accessory sex glands do not develop. Rather, the patients have female external genitalia. These patients are unresponsive to androgens and have defective androgen receptors. Another type of male pseudohermaphroditism results from a deficiency of the enzyme 5α-reductase. Since DHT is necessary for early differentiation and development, the patients again develop female genitalia; later, some masculinization can occur at the time of puberty due to elevated testosterone levels in the blood. A third disorder is Reifenstein syndrome, an incomplete pseudohermaphroditism. In these patients, the androgen levels are normal, 5α-reductase is present, and elevated LH levels are found. Partially deficient androgen receptors are present in these patients. In most cases of male pseudohermaphroditism, androgen replacement has little or no effect and thus steroid treatment is not recommended.

Deficiencies of circulating gonadotropins lead to secondary hypogonadism. This condition can be caused by disorders of the pituitary and/or hypothalamus resulting in diminished secretions of neurohormones. The lack of stimulation of the seminiferous tubules and the Leydig cells due to the low levels of these neurohormones decreases androgen production. Drugs such as the neuroleptic phenothiazines and the stimulant marijuana can also interfere with release of gonadotropins. The use of androgens in secondary hypogonadism is symptomatic.

Synthetic androgens have also been used in women for the treatment of endometriosis, abnormal uterine bleeding, and menopausal symptoms. However, their utility is severely limited by the virilizing side effects of those agents. Two weak androgens, calusterone, and 1-dehydrotestolactone, are used clinically in the treatment of mammary carcinoma in women. The mode of action of these drugs in the treatment of breast cancer is unknown, and is not simply related to their androgenicity. More recent evidence on the ability of these compounds to inhibit estrogen biosynthesis catalyzed by aromatase suggests that they effectively lower estrogen levels *in vivo*.

Traditionally, androgens have been assayed by the capon comb growth method and by the use of the seminal vesicles and prostate organs of the rodent. An increase in weight and/or growth of the capon comb have been used to denote androgenic activity following injection or topical application of a solution of the test compound in oil. A number of minor modifications of this test have been

described. The increase in weight of the seminal vesicles and the ventral prostate of the immature castrated male rat has provided another measure of androgenic potency. The test compound is administered either intramuscularly or orally, and the weight of the target organs is compared with those of control animals. *In vitro* evaluation of the relative affinity of potential androgens for the androgen receptor has also become an important tool in assessing biological activity of androgens.

4.3 Structure–Activity Relationships for Androgens

4.3.1 Early Modifications

Most of the early structure–activity relationship studies concerned minor modifications of testosterone and other naturally occurring androgens. Studies in animals and humans showed the 17β-hydroxyl function to be essential for androgenic and anabolic activity. In certain cases, esterification of the 17β-hydroxyl group not only enhanced but also prolonged the anabolic and androgenic properties. The 1-dehydro isomer of testosterone (**28**) and related compounds are potent androgenic and anabolic steroids.

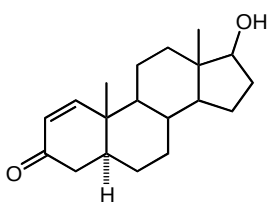

(**28**)

Reduction of the A-ring functional groups has variable effects on activity. For example, conversion of testosterone to DHT has little effect or may increase potency in a variety of bioassay systems. On the other hand, changing the A/B trans stereochemistry of known androgens such as androsterone (**3**) and DHT to the A/B *cis*-etiocholanolone (**18**) and 17β-hydroxy-5β-androstan-3-one (**13**), respectively, drastically reduces both the anabolic and androgenic properties. These observations established the importance of the A/B trans ring juncture for activity.

4.3.2 Methylated Derivatives

The discovery that C-17α-methylation conferred oral activity on testosterone prompted the synthesis of additional C-17α-substituted analogs. Increasing the chain length beyond methyl invariably led to a decrease in activity. As a result of these studies, however, 17α-methylandrost-5-ene-3β,17β-diol (methandriol, **29**) was widely evaluated in humans as an anabolic agent. Clinical studies showed no advantage of methandriol over 17α-methyltestosterone (**5**).

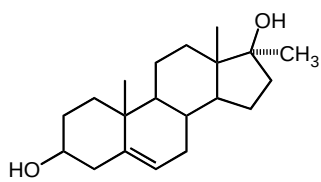

(**29**)

4.3.3 Ester Derivatives

As early as 1936 it was known that esterification of testosterone markedly prolonged the activity of this androgen when it was administered parenterally. This modification enhances the lipid solubility of the steroid and, after injection, permits a local depot effect. The acyl moiety is usually derived from a long-chain aliphatic or arylaliphatic acid such as propionic, heptanoic

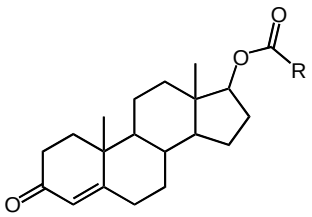

R = CH_2CH_3 (**30**) Propionate

$(CH_2)_5CH_3$ (**31**) Heptanoate (enanthate)

$(CH_2)_8CH_3$ (**32**) Decanoate

CH_2CH_2 — (**33**) Cyclopentylpropionate (cypionate)

CH_2CH_2 — (**34**) β-Phenylpropionate

(enanthoic), decanoic, cyclopentylpropionic (cypionic), or β-phenylpropionic acid (**30–34**).

4.3.4 Halo Derivatives

In general, the preparation of halogenated testosterone derivatives has been therapeutically unrewarding. 4-Chloro-17β-hydroxyandrost-4-en-3-one (chlorotestosterone, **35**) and its derivatives are the only chlorinated androgens that have been used clinically, albeit sparingly. The introduction of a 9α-fluoro and an 11β-hydroxy substitutent (analogous to synthetic glucocorticoids) gives 9α-fluoro-11β,17β-dihydroxy-17α-methylandrost-4-en-3-one (fluoxymesterone, Halotestin, **36**), which is an orally active androgen exhibiting approximately fourfold greater oral activity than 17α-methyltestosterone. Early clinical studies with fluoxymesterone indicated an anabolic potency that was 11 times that of the unhalogenated derivative. Nitrogen balance studies, however, revealed an activity only 3 times that of 17α-methyltestosterone. Because of the lack of any substantial separation of anabolic and androgenic activity, halotestin is used primarily as an orally effective androgen, particularly in the treatment of mammary carcinoma.

4.3.5 Other Androgen Derivatives

Several synthetic steroids having weak androgenic activity are being utilized in patients. 7β,17α-Dimethyltestosterone (calusterone, **37**) and 1-dehydrotestolactone (Testlac, **38**) are very weak androgenic agents that have been used in the treatment of advanced metastatic breast cancer.

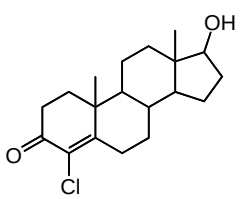

Chlorotestosterone (**35**)

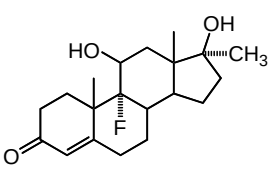

Fluoxymesterone (**36**)

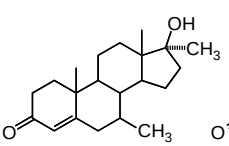

Calusterone (**37**)

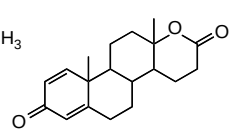

1-Dehydrotestolactone (**38**)

4.3.6 Summary of Structure–Activity Relationships

As with other areas of medicinal chemistry, the desire to relate chemical structure to androgenic activity has attracted the attention of numerous investigators. Although it is often difficult to interrelate biological results from different laboratories, androgenicity data from the same laboratory afford useful information. In evaluating the data, one must be careful to note not only the animal model employed, but also the mode of administration. For example, marked differences in androgenic activity can be found when compounds are evaluated in the chick comb assay (local application) as opposed to the rat ventral prostate assay (subcutaneous or oral). The chick comb assay measures "local androgenicity" and is believed to minimize such factors as absorption, tissue distribution, and metabolism, which complicate the interpretation of *in vivo* data in terms of hormone–receptor interactions.

Furthermore, although the rat assays correlate well with what one eventually finds in humans, few studies of comparative pharmacology have been performed. Indeed, DHT may not be the principal mediator of androgenicity in all species. For example, a cytosol receptor protein has been found in normal and hyperplastic canine prostate that is specific for 5α-androstane-3α,17α-diol.

Since the presence of the 17β-hydroxyl group was demonstrated very early to be an important feature for androgenic activity in rodents, most investigators interested in structure–activity relationships maintained this function and modified other parts of the testosterone molecule. Three observations can be made on the basis of these studies: (1) The 1-dehydro isomer of testosterone is at least as active as testosterone; (2) the 1- and 4-keto isomers of testosterone and DHT have variable activity; (3) the 2-keto isomers of testosterone and DHT consistently lack appreciable activity.

The oxygen function at position 3 could be removed from testosterone with little reduction in androgenic activity, while removal of the hydroxyl group from position 17 sharply reduced the androgenicity. As a continuation of these studies, the hydrocarbon nucleus, 5α-androstane (**39**), was synthesized. It too was found to possess androgenicity when applied topically or given intramuscularly in the chick comb assay, albeit at high doses. On the other hand, it was learned later that the 19-nor analog, 5α-estrane (**40**), had less than 1% of the androgenic activity of testosterone propionate in castrated male rats.

The relative androgenicity of the isomeric A-ring olefins of 3-deoxy testosterones was the order $\Delta^1 > \Delta^2 > \Delta^3 > \Delta^4$. The Δ^2-isomer displayed the greatest anabolic activity and the best anabolic-to-androgenic ratio. On the basis that sulfur is bioisosteric with CH=CH, the thia, seleno, and tellurio androstanes were synthesized, which displayed androgenic activity. When the heteroatom was oxygen, however, the compound (**41**) was essentially devoid of

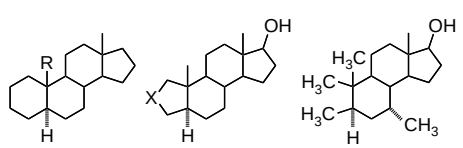

(**39**) R = CH_3
(**40**) R = H

(**41**)
X = O

(**42**)

androgenicity. The oxygen analog was said to be inactive because oxygen is isosteric with CH_2 rather than $CH_2{=}CH_2$. Thus, a minimum ring size was found to be required for activity. When the oxygen atom was introduced as part of a six-membered A-ring, an active androgen resulted.

As with the case of the double-bond isomers, the position of the oxygen atom was found to be important. The substitution of oxygen at C-2 gives rise to the most active compound, and the order of activity was $2 > 3 \gg 4$. The full steroid nucleus is not essential for androgen activity, since 7α-methyl 1,4-seco-2,3-bisnor-5α-androstan-17β-ol (**42**) has 50% of the anabolic activity of testosterone.

Both 14-dehydrotestosterone and the corresponding 19-nor analog were found to be potent androgens when applied topically. An extension of this series ascertained the effect of introducing a 7α-methyl. The results of this study are listed in Table 4, in terms of percentage increases in the weights of chick combs, rat ventral prostates, and rat levator ani induced by the test compounds as related to a similar dose of testosterone, the responses to the latter being described as 100%.

The effects of either 7α-methyl or 14- dehydro modification are more pronounced for 19-nortestosterone than for testosterone. The 14-dehydro modification had a greater effect on local androgenicity, whereas 7α-methylation had a more positive effect on systemic androgenicity. A marked synergism resulted when both the 14-dehydro and 7α-methyl modifications were present. The resultant compound, 7α-methyl-14-dehydro-19-nortestosterone, represents one of the most potent androgens reported to date.

The characterization of a specific receptor protein in androgen target tissues has made it possible to directly analyze the receptor affinity of various testosterone analogs. Table 6 shows the receptor affinity and androgenic activity of a variety of androgens relative to DHT. The ability of the various steroids to be retained by prostate nuclei is also indicated. As would be expected, the receptor affinity data did not necessarily correlate with the systemic androgenicity. In some cases, such as with 7α-methyl-19-nortestosterone, there was good agreement. Such was not the case, however, for 19-nortestosterone. Table 7 summarizes their findings. Whereas the importance of the A/B trans ring fusion

Tab. 6 Relative androgenicity and receptor binding capacity of various androgens.

Steroid	*Androgenicity in rat (s.c.)*	*Cytosol receptor*	*Nuclear retention*
DHT	1.0	1.0	1.0
Testosterone (T)	0.4	<0.1	0.7
5α-Androstanedione	0.2	0.0	0.2
19-NorDHT	0.1	0.5	0.6
19-NorT	0.2	0.9	0.7
7α-CH_3-DHT	1.2	0.4	0.4
7α-CH_3-T	0.4	0.2	0.2
7β-CH_3-T	0.1	<0.1	<0.1
7α-CH_3−19-NorDHT	0.3	0.6	0.4
7α-CH_3−19-NorT	2.6	2.6	1.8

Tab. 7 Binding affinity of various androgens for rat ventral prostate receptor protein.

Steroid	K_B $[M^{-1}]$
5α-DHT	6.9×10^8
5β-DHT	6.4×10^7
17β-Testosterone	4.2×10^8
17α-Testosterone	2.1×10^7
Androstenedione	1.3×10^7
5α-Androstanedione	3.5×10^7
19-Nortestosterone	8.6×10^8
14-Dehydrotestosterone	4.4×10^8
14-Dehydro-19-nortestosterone	5.9×10^8
7α-CH_3-14-Dehydro-19-nortestosterone	5.0×10^8

and 17β-hydroxyl prevailed, the data failed to demonstrate the potency previously noted for 7α-methyl-14-dehydro-19-nortestosterone. Moreover, 19-nortestosterone displayed a receptor affinity greater than that of DHT, yet its androgenicity is much less than DHTs.

These differences in correlations between receptor assays and *in vivo* data should not cloud the importance of the receptor studies. The receptor assays measure affinity for the receptor protein, and this property is shared by androgens as well as antiandrogens. Moreover, such assays cannot predict the disposition and metabolic fate of an androgen following administration. Figure 9 contains a summary of the structure–activity relationships for androgens.

4.4 Absorption, Distribution, and Metabolism

Numerous factors are involved in the absorption, distribution, and metabolism of the synthetic androgens and the physicochemical properties of these steroids greatly influence the pharmacokinetic parameters. The lipid solubility of a synthetic steroid is an important factor in its intestinal absorption. The acetate ester of testosterone demonstrated enhanced absorption from the gastrointestinal tract over both testosterone and 17α-methyltestosterone. Injected solutions of testosterone in oil result in the rapid absorption of the hormone from the injection site; however, rapid metabolism greatly decreases the biological effects of the injected testosterone. The esters of testosterone are much more nonpolar and, when injected intramuscularly, are absorbed more slowly. As a result, the commercial preparations of testosterone propionate are administered

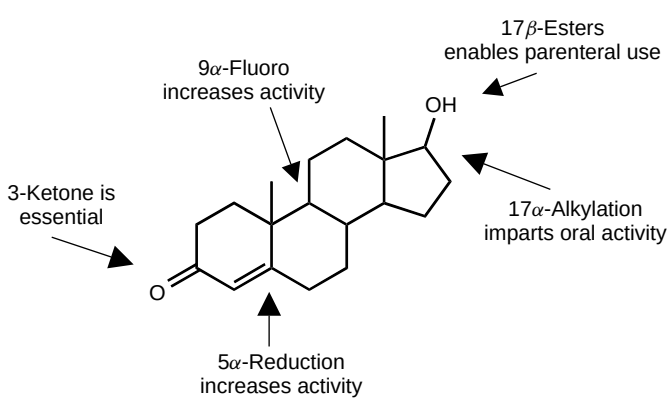

Fig. 9 Summary of structure–activity relationships for androgens.

every 2 or 3 days. Increasing the size of the ester functionality enables the testosterone esters such as the ethanate or cypionate to be given in a depot injection lasting 2 to 4 weeks.

Once absorbed, the steroids are transported in the circulation primarily in a protein-bound complex. Testosterone and other androgens are reversibly associated with certain plasma proteins and the unbound fraction can be absorbed into target cells to exert its action. The structure-binding relationships of the natural and synthetic androgens to sex SBG have been extensively investigated. A 17α-hydroxyl group is essential for binding and the presence of a 17α-substituent such as the 17α-methyl moiety decreases its affinity. The 5α-reduced androgens bind with the highest affinity. A much smaller quantity of the androgen is bound to other plasma proteins, principally albumin and CBG or transcortin.

The metabolism of the synthetic androgens is similar to that of testosterone. The introduction of the 17α-methyl group greatly retards the metabolism, thus providing oral activity. Reduction of the 4-en-3-one system in synthetic androgens to give the various α- and β-isomers occurs *in vivo*. Finally, aromatization of the A-ring can also occur. One analog that demonstrates an alternate metabolic pattern is 4-chlorotestosterone, which in humans gave rise to an allylic alcohol, 4-chloro-3α-hydroxyandrost-4-en-17-one. A number of other halogenated testosterone derivatives subsequently were found to take this abnormal reduction path *in vitro*. It was proposed that fluorine or chlorine substituents at the 2-, 4-, or 6- position in testosterone interfere with the usual α,β-unsaturated ketone resonance so that the C-3 carbonyl electronically resembles a saturated ketone.

4.5 Toxicities

The use of androgens in women and children can often result in virilizing or masculinizing side effects. In boys, an acceleration of the sexual maturation is seen, while in girls and women growth of facial hair and deepening of the voice can be observed. These effects are reversible when medication is stopped; however, prolonged treatment can produce effects that are irreversible. Inhibition of gonadotropin secretion by the pituitary can also occur in patients receiving androgens.

Both males and females experience salt and water retention resulting in edema. This edema can be treated by either maintaining a low-salt diet or by using diuretic agents. Liver problems are also encountered with some of the synthetic androgens. Clinical jaundice and cholestasis can develop after the use of the 17α-alkylated products. Various clinical laboratory tests for hepatic function such as bilirubin concentrations, sulfobromophthalein retention, and glutamate transaminase and alkaline phosphatase activities are affected by these androgen analogs.

5 Anabolic Agents

5.1 Current Drugs on the Market

See Table 8.

5.2 Therapeutic Uses and Bioassays

Many synthetic analogs of testosterone were prepared in order to separate the anabolic activity of the C_{19} steroids from

Tab. 8

Generic name (structure)	*Trade name*	*US manufacturer*	*Chemical class*	*Dose*
Nandrolone decanonate (**43**)	Deca-Durabolin	Organon	Estrane	Injection (in oil): 100 mg mL^{-1} 200 mg mL^{-1}
Oxandrolone (**65**)	Oxandrin	Gynex	Androstane	Tablets: 2.5 mg
Oxymetholone (**63**)	Anadrol-50	Syntex	Androstane	Tablets: 50 mg
Stanozolol (**75**)	Winstrol	Winthrop Pharm	Androstane	Tablets: 2 mg

their androgenic activity. Although the goal of a pure synthetic anabolic that retains no androgenic activity has not been accomplished, numerous preparations are now available on the market that have high anabolic/androgenic ratios.

The primary criterion for assessing anabolic activity of a compound is the demonstration of a marked retention of nitrogen. This nitrogen-retaining effect is the result of an increase in protein synthesis and a decrease in protein catabolism in the body. Thus, the urinary nitrogen excretion, particularly urea excretion, is greatly diminished. The castrated male rat serves as the most sensitive animal model for nitrogen retention, although other animals have been utilized. Another bioassay for anabolic activity involves examination of the increase in mass of the levator ani muscle of the rat upon administration of an anabolic agent. This measure of myotrophic effect correlates well with the nitrogen retention bioassay and the two are usually performed in the determination of anabolic activity.

Anabolic steroids exert other effects on the body as well. Skeletal mineralization and bone maturation are enhanced by androgens and anabolics. These agents will decrease calcium excretion by the kidney and result in increased deposition of both calcium and phosphorus in bone. Androgenic and anabolic agents also can influence red blood cell formation. Two mechanisms of action of this erythropoiesis-involving increased erythropoietin production and enhanced responsiveness of the tissue have been described.

These various biological activities of the anabolics have prompted the use of these agents in treatment protocols, with varying success. Clinical trials have demonstrated the effectiveness of the anabolic steroids in inducing muscle growth and development in some diseases. Anabolic steroids are effective in the symptomatic treatment of various malnourished states due to their ability to increase protein synthesis and decrease protein catabolism. Treatment of diseases such as malabsorption, anorexia nervosa, emaciation, and malnutrition as a result of psychoses includes dietary supplements, appetite stimulants, and anabolics. Improved postoperative recovery with adjunctive use of anabolic agents has been demonstrated in numerous clinical studies. However, their usefulness in other diseases such as muscular dystrophies and atrophies and in geriatrics has not been observed.

Anabolic steroids also have the ability to lower serum lipid levels *in vivo*. The most

widely studied agent is oxandrolone, which dramatically lowers serum triglycerides and, to a lesser extent, cholesterol levels at pharmacological doses. The proposed mechanism of this hypolipidemic effect includes both an inhibition of triglyceride synthesis and an increased clearance of the triglycerides. The androgenic side effects of the anabolics and the lack of superiority over conventional hypolipidemic agents have curtailed its use for treatment of these conditions.

The stimulation of erythropoiesis by anabolics has resulted in the use of these agents for the treatment of various anemias. Anemias arising from deficiencies of the bone marrow are particularly responsive to pharmacological doses of anabolic agents. Treatment of aplastic anemia with anabolics and corticosteroids has been proven effective. Secondary anemias resulting from inflammation, renal disease, or neoplasia are also responsive to anabolic steroid administration. Synthetic anabolics have been prescribed for women with osteoporosis and for children with delayed growth. These applications have produced limited success; however, the virilizing side effects severely limit their usefulness, particularly in children.

The methods employed to determine the anabolic or myotrophic properties of steroids are based on an increase in nitrogen retention and/or muscle mass in various laboratory animals. The castrated male rat is presently the most widely used and most sensitive laboratory animal for nitrogen balance studies. Dogs and ovariectomized monkeys have also been employed. Although it is generally agreed that variations in urinary nitrogen excretion relate to an increase or decrease in protein synthesis, nitrogen balance assays are not without their limitations. This is partly because such studies fail to describe the shifts in organ protein and measure only the overall status of nitrogen retention in the animal.

The easily accessible levator ani muscle of the rat has provided a valuable index for measuring the myotrophic activity of steroidal hormones. By comparing the weight of levator ani muscle, seminal vesicles, and ventral prostate with controls, one can obtain a ratio of anabolic-to-androgenic activity. There also appears to be some correlation between the levator ani response and urinary nitrogen retention. A modification of this muscle assay utilizes the parabiotic rat and allows for the simultaneous measurement of pituitary gonadotropic inhibition and myotrophic activity. The suitability of the levator ani assay has been questioned for the possibility that its growth is more a result of androgenic sensitivity than of any steroid-induced myotrophic effect. Thus, this assay is usually performed in conjunction with nitrogen balance studies or acceleration of body growth.

Finally, the myotrophic effects have also led to the use and the abuse of these agents by athletes. Conflicting reports on the effectiveness of anabolics to increase strength and power in healthy males have resulted from clinical trials. Several groups reported no significant differences between groups of male college-age students receiving anabolics and weight training and those groups receiving placebo plus the weight training in double-blind studies. Other reports cited some improvement in strength and power, but they utilized small numbers of subjects or were only single blind studies. ''Supraphysiologic'' doses of anabolics have demonstrated enhancement of muscle size and strength. Overall, the consensus of these various studies are that anabolic steroids provide only limited improvement in strength and

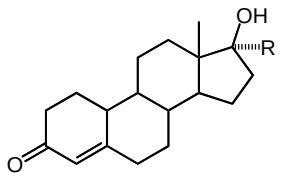

R = H (**43**) Nandrolone
R = CH_3 (**44**) Normethandrone
R = CH_2CH_3 (**45**) Norethandrone

power in healthy males. Anabolic steroids also exhibit an anticatabolic effect, that is, reversing the catabolic effects of glucocorticoids released in response to stress. Such effects would enable individuals to recover more quickly following strenuous workouts. An alarming percentage of amateur and professional athletes utilize anabolic steroids, which are readily available "on the street." The use of these steroids for increasing strength and power is banned in intercollegiate and international sports, and very sensitive assays (radioimmunoassay (RIA), gas chromatography-mass spectrometry (GS-MS)) have been developed for measuring anabolic levels in urine and blood. Nonetheless, the recent discovery of "designer steroids" such as tetrahydrogestrinone (THG), a new chemical entity synthesized specifically for illicit anabolic use to avoid standard antidoping tests, is an alarming trend and highlights the importance of aggressive detection approaches for monitoring drug doping in sports.

5.3 Structure–Activity Relationships for Anabolics

5.3.1 19-Nor Derivatives

An important step toward developing an anabolic agent with minimal androgenicity occurred when 19-nortestosterone (17β-hydroxyestr-4-en-3-one, nandrolone, **43**) was found to be as myotrophic but only about 0.1 as androgenic as testosterone. This observation prompted the synthesis and evaluation of a variety of 19-norsteroids, including the 17α-methyl (normethandrone, **44**) and the 17α-ethyl (norethandrolone, **45**) homologs of 19-nortestosterone.

Nandrolone in the form of a variety of esters (such as decanoate and β-phenylpropionate) and norethandrolone have been widely used clinically. The latter, under the name Nilevar®, was the first agent to be marketed in the United States as an anabolic steroid. Androgenic and progestational side effects, however, led to its eventual replacement by other agents.

Nonetheless, these studies did stimulate the synthesis of other norsteroids. Interestingly, both 18-nortestosterone and 18,19-bisnortestosterone were essentially devoid of both androgenic and anabolic properties. Contraction of the B-ring led to B-norsteroids, which were also lacking in androgenicity, but unlike the foregoing, this modification led to compounds with antiandrogenic activity.

Of the number of homoandrostane derivatives (those having one or more additional methylene groups included in normal tetracyclic ring system) that have been synthesized, only B-homo and D-homodihydrotestosterone have shown appreciable androgenic activity. A D-bishomo analog (**46**) was reported to be weakly androgenic.

5.3.2 Dehydro Derivatives

The marked enhancement in biological activity afforded by introduction of a double bond at C_1 of cortisone and hydrocortisone

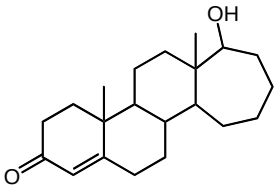

(**46**)

prompted similar transformations in the androgens. The acetate of 17β-hydroxyandrosta-1,4-dien-3-one (**47**) was as myotrophic as testosterone propionate but was much less androgenic. Furthermore, 17α-methyl-17β-hydroxyandrosta-1, 4-dien-3-one (methandrostenolone, **48**) had 1 to 2 times the oral potency of 17α-methyltestosterone in the rat nitrogen retention and levator ani muscle assays. In clinical studies, methandrostenolone produced a marked anabolic effect when given orally at doses of 1.25 to 10 mg daily and was several times more potent than 17α-methyltestosterone.

In contrast with the 1-dehydro analogs, introduction of an additional double bond at the 6-position (**49**) markedly decreased both androgenic and myotrophic activity in the rat. Moreover, removal of the C_{19}-methyl, inversion of the configuration at C_9 and C_{10} and at C_8 and C_{10}, and reduction of the C_3-ketone failed to improve the biological properties.

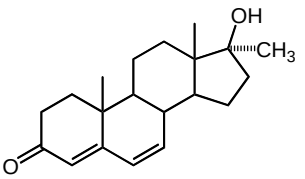

(**49**)

On the other hand, introduction of unsaturation into the B-, C-, and D-rings has given rise to compounds with significant androgenic or anabolic activity. Ethyldienolone (**50**), for example, displayed an anabolic-to-androgenic ratio of 5 and was slightly more active than methyltestosterone when both were given orally. Introduction of a 14–15 double bond (**51**) increased androgenicity when compared with testosterone by local application in the chick comb assay. On the other hand, there was a 25% decrease in androgenicity when measured by the rat ventral prostate following subcutaneous administration. Conversion to the 19-nor analog (**52**) increased

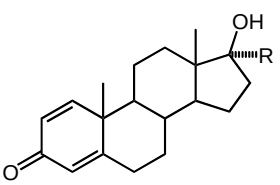

R = H (**47**)
R = CH_3 (**48**)

Methandrostenolone

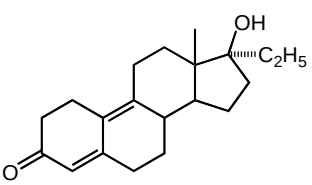

(**50**)

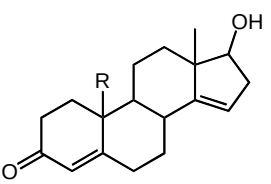

(**51**) R = CH_3
(**52**) R = H

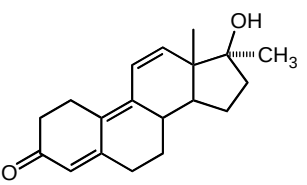

(**53**)

androgenicity, but the anabolic activity was significantly enhanced.

Of a variety of triene analogs of testosterone that have been tested, only 17α-methyl-17β-hydroxyestra-4,9,11-trien-3-one (methyltrienolone, **53**) showed significant activity in rats. Surprisingly, this compound had 300 times the anabolic and 60 times the androgenic potency of 17α-methyltestosterone when administered orally to castrated male rats. In this instance, however, the potent hormonal properties on rats did not correlate with later studies in humans. One study in patients with advanced breast cancer found methyltrienolone to have weak androgenicity and to produce severe hepatic dysfunction at very low doses.

5.3.3 Alkylated Analogs

An extensive effort has been directed toward assessing the physiological effect of replacing hydrogen with alkyl groups at most positions of the steroid molecule. Although methyl substitution at C_3, C_4, C_5, C_6, C_{1l}, and C_{16} has generally led to compounds with low anabolic and androgenic activity, similar substitutions at C_1, C_2, C_7, and C_{18} have afforded derivatives of clinical significance.

1-Methyl-17β-hydroxy-5α-androst-l-en-3-one (**54**) as the acetate (methenolone acetate) was about 5 times as myotrophic, but only 0.1 as androgenic, as testosterone propionate in animals. In addition, this compound or the free alcohol represented one of the few instances of a C_{17} nonalkylated steroid that possessed significant oral anabolic activity in animals and in man. This effect may be related to the slow *in vivo* oxidation of the 17β-hydroxyl group when compared with testosterone. At a daily dose of 300 mg, methenolone acetate caused little virilization. By contrast, the dihydro analog, 1α-methyl-17β-hydroxy-5α-androstan-3-one (mesterolone, **55**), was found to possess significant oral androgenic activity in the cockscomb test and in clinical assays. A comparison of the anabolic and androgenic activity of (**54**) with its A-ring congeners revealed that the double bond was necessary at C_1 for anabolic activity. For example, 1α-methyl-17β-hydroxyandrost-4-en-3-one had a much lower activity. Furthermore, either reduction of the C_3 carbonyl group of (**55**) or removal of the C_{19} methyl group greatly reduced both anabolic and androgenic activity in this series.

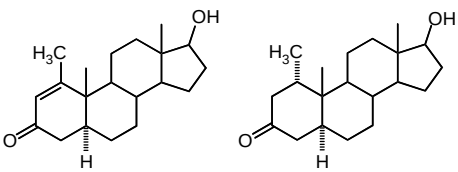

Methenolone (**54**) Mesterolone (**55**)

Among the C_2-alkylated testosterone analogs, 2α-methyl-5α-androstan-17β-ol-3-one (drostanolone, **56**) and its 17α-methylated homolog (**57**) have displayed anabolic activity both in animals and in man. In contrast, 2,2-dimethyl and 2-methylenetestosterone or their derivatives showed only low anabolic or androgenic activity in animals. The most interest in drostanolone has been in relation to its potential as an antitumor agent with decreased masculinizing propensity.

(**56**) R = H
(**57**) R = CH3

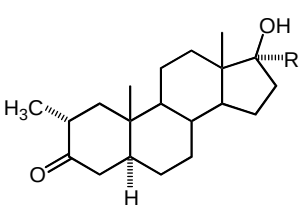

7α,17α-Dimethyltestosterone (bolasterone, **58**) had 6.6 times the oral anabolic

potency of 17α-methyltestosterone in rats. Similar activity was observed in man at 1 to 2 mg day^{-1} without many of the usual side effects. Moreover, the corresponding 19-nor derivative was 41 times as active as 17α-methyltestosterone as an oral myotrophic agent in the rat. 7α-Methyl-14-dehydro-19-nortestosterone (**59**) was found to be approximately 1000 times as active as testosterone in the chick comb assay and about 100 times as active as testosterone in the ventral prostate assay.

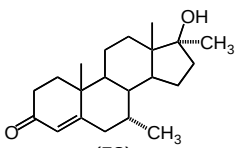

(**58**)

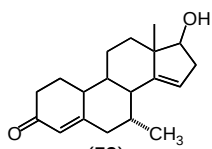

(**59**)

Certain totally synthetic 18-ethylgonane derivatives possessed pronounced anabolic activity. Similar to other 19-norsteroids, 13β,17α-diethyl-17β-hydroxygon-4-en-3-one (norbolethone, **60**) was found to be a potent anabolic agent in animals and in man. Since it is prepared by total synthesis, the product is isolated and marketed as the racemic DL-mixture. The hormonal activity resides in the D-enantiomer.

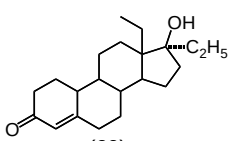

(**60**)

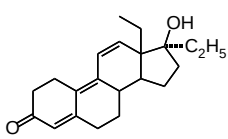

(**61**)

Tetrahydrogestrinone or THG (**61**), a "designer steroid" for illicit anabolic use, is synthesized from the progestin, gestrinone, by reduction of the 17α-ethinyl group. THG demonstrates potent androgenic and progestational activity in *in vitro* receptor bioassay systems. No animal or human studies with THG are known, since this agent was never prepared as a pharmaceutical for medical or veterinary use.

5.3.4 Hydroxy and Mercapto Derivatives

Testosterone has been hydroxylated at virtually every position on the steroid nucleus. For the most part, nearly all these substances possess no more than weak myotrophic and androgenic properties. Two striking exceptions to this, however, are 4-hydroxy- and 11β-hydroxytestosterones. 4-Hydroxy-17α-methyltestosterone (oxymesterone, **62**), for instance, had 3 to 5 times the myotrophic and 0.5 times the androgenic activity of 17α-methyltestosterone in the rat. In clinical studies, oxymesterone produced nitrogen retention in adults at a daily dose of 20 to 40 mg, and no adverse liver function was observed. Introduction of an 11β-hydroxyl group in many instances resulted in a favorable effect on biological activity. 11β-Hydroxy-17α-methyltestosterone (**63**) was more anabolic in the rat than 17α-methyltestosterone and 1.5 times as myotrophic in humans.

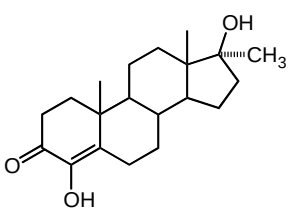

(**62**)

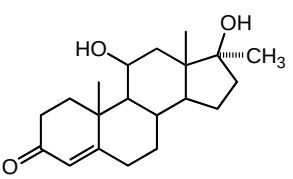

(**63**)

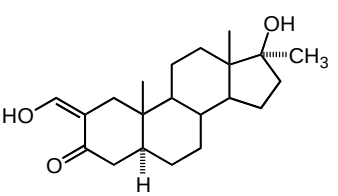

(**64**)

One of the most widely studied anabolic steroids has been 2-hydroxymethylene-17α-methyl-5α-androstan-17β-ol-3-one (oxymetholone, **64**). In animals, it was found to be 3 times as anabolic and 0.5 times as androgenic as 17α-methyltestosterone. Clinical studies confirmed these results.

The substitution of a mercapto for a hydroxyl group has generally resulted in decreased activity. However the introduction of a thioacetyl group at C_1 and C_7 of 17α-methyltestosterone afforded 1α,7α-bis(acetylthio)-17α-methyl-17β-hydroxyandrost-4-en-3-one (thiomesterone, **65**), a compound with significant activity. Thiomesterone was 4.5 times as myotrophic and 0.6 times as androgenic as 17α-methyltestosterone in the rat and has been used clinically as an anabolic agent.

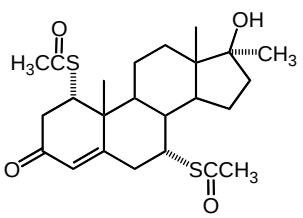

(**65**)

Moreover, numerous 7α-alkylthio androgens have exhibited anabolic-androgenic activity similar to that of testosterone propionate when administered subcutaneously. Even though no clinically useful androgen resulted, similar 7α-substitutions were advantageous in the development of radioimmunoassays now employed in clinical laboratories. In addition, certain 7α-arylthioandrost-4-ene-3,17-diones are effective inhibitors of estrogen biosynthesis (see Sect. 6.3.3).

5.3.5 Oxa, Thia, and Aza Derivatives

A number of androgen analogs in which an oxygen atom replaces one of the methylene groups in the steroid nucleus have been synthesized and biologically evaluated. Of these derivatives, 17β-hydroxy-17αmethyl-2-oxa-5α-androstan-3-one (oxandrolone, **66**) was 3 times as anabolic and only 0.24 times as androgenic as 17α-methyltestosterone in the oral levator ani assay. By contrast, only minimal responses were obtained following intramuscular administration. The 2-thia and 2-aza analogs were essentially devoid of activity by both routes. The 3-aza-A-homoandrostene derivative (**67**) displayed only 5% the anabolic-to-androgenic activity of methyltestosterone.

The clinical anabolic potency of oxandrolone was considerably more active than 17α-methyltestosterone and provided perceptible nitrogen sparing at a dose as low as 0.6 mg day^{-1}. Moreover, at dosages of 0.25 to 0.5 mg kg^{-1}, oxandrolone was effective as a growth-promoting agent without producing the androgenically induced bone maturation. Because of this

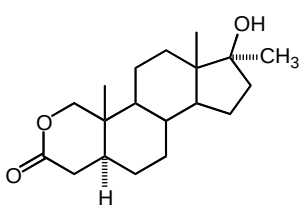

(**66**)

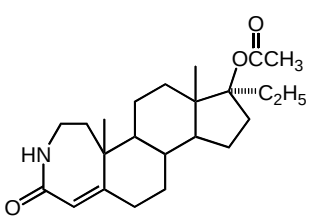

(**67**)

favorable separation of anabolic from androgenic effects, oxandrolone has been one of the most widely studied anabolic steroids. Its potential utility in various clinical hyperlipidemias was discussed in Sect. 5.2.

The significant hormonal activity noted for estra-4,9-dien-3-ones such as (**50**) (see Sect. 5.3.2) prompted the synthesis of the 2-oxa bioisosteres in this series. Despite the lack of a 17α-methyl group, (**68**) had 93 times the oral anabolic activity of 17α-methyltestosterone. It was also 2.7 times as androgenic. As might be expected, the corresponding 17α-methyl derivative, (**69**), was the most active substance in this series. It had 550 times the myotrophic and 47 times the androgenic effect of 17α-methyltestosterone. These two compounds differed dramatically in progestational activity, however. The activity of (**68**) was only 0.1 times, whereas the activity of (**69**) was 100 times that of progesterone in the Clauberg assay. The pronounced oral activity of (**68**) suggests that it is not a substrate for the 17β-alcohol dehydrogenase and represents an interesting finding.

(**68**) R = H
(**69**) R = CH_3

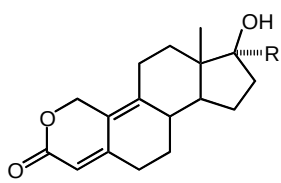

5.3.6 Deoxy and Heterocyclic-fused Analogs

Early studies indicated that the 17β-hydroxyl group and the 3-keto group were essential for maximum androgenic activity. On the basis of this observation, the C_3 oxygen function was removed in the hope of decreasing the androgenic potency while maintaining anabolic activity. Unfortunately, the results failed to substantiate the rationale, and 17α-methyl-5α-androstan-17β-ol (**70**) was found to be a potent androgen in animals and humans. However, this substance is extensively metabolized to the 3-keto derivative by rabbit liver homogenate. Other deoxy analogs of testosterone have been synthesized and tested. A 19-nor derivative, 17α-ethylestr-4-en-17β-ol (estrenol, **71**) had at least 4 times the anabolic and 0.25 times the androgenic activity of 17α-methyltestosterone in animals and was effective in humans at a daily dose of 3 to 5 mg. In addition, 17α-methyl-5α-androst-2-en-17β-ol (**72**) offered a good separation of anabolic from androgenic activity.

Since sulfur is considered to be isosteric with $-CH{=}CH-$, the 2-thia-A-nor-5α-androstane derivatives such as (**73**) possessed high androgenic and anabolic activity and served to verify that steric rather than electronic factors are important in connection with the structural requirements at C-2 and/or C-3 in androgens. Interestingly, the selenium and tellurium isosteres in the same series were found to

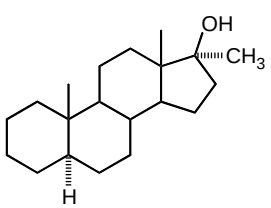

(**70**)

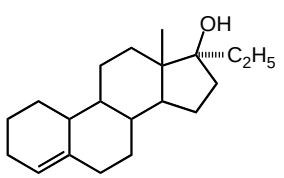

(**71**)

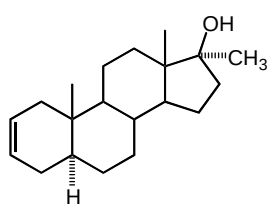

(**72**)

have good androgenic activity. Moreover, experiments with a ^{75} Se-labeled analog have shown (**74**) to selectively bind with the specific cytosol receptor for DHT in rat prostate.

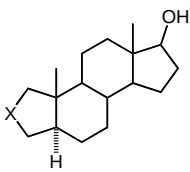

(**73**) X = S
(**74**) X = Se

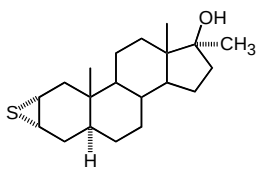

(**75**)

The high biological activity noted for the 3-deoxy androstanes prompted numerous investigators to fuse various systems to the A-ring. The simplest such changes were 2,3-epoxy, 2,3-cyclopropano, and 2,3-epithioandrostanes. The 2α,3α-cyclopropano-5α-androstan-17β-ol was as active as testosterone propionate as an anabolic agent. While the epoxides had little or no biological activity, certain of the episulfides possessed pronounced anabolic/androgenic activity. For example, 2,3α-epithio-17α-methyl-5α-androstan17β-ol (**75**) was found to have approximately equal androgenic and 11 times the anabolic activity of methyltestosterone after oral administration to rats. The 2,3-β-episulfide, on the other hand, was much less active. 2,3α-Epithio-5α-androstan17β-ol has been shown to have long-acting antiestrogenic activity, as well as some beneficial effects in the treatment of mammary carcinoma.

Other heterocyclic androstane derivatives have included the pyrazoles. Thus, 17β-hydroxy-17α-methylandrostano[3,2-*c*] pyrazole (stanozolol, **76**) was 10 times as active as 17α-methyltestosterone in improving nitrogen retention in rats. The myotrophic activity, however, was only twice that of 17α-methyltestosterone. Stanozolol at a dose of 6 mg day^{-1} produced an adequate anabolic response with no lasting adverse side effects.

The high activity of the pyrazoles instigated the synthesis of other heterocyclic fused androstane derivatives including isoxazoles, thiazoles, pyridines, pyrimidines, pteridines, oxadiazoles, pyrroles, indoles, and triazoles. One of the most potent was 17α-methylandrostan-17β-ol[2,3-*d*]isoxazole (androisoxazol, **77**), which exhibited an oral anabolic-to-androgenic ratio of 40. The corresponding 17α-ethynyl analog (danazol, **78**) has been of most interest clinically. This compound has impeded androgenic activity and inhibits pituitary gonadotropin secretion. Since it depresses blood levels of androgens and gonadotropins, it has been studied as an antifertility agent in males. At doses of 200 or 600 mg daily, danazol lowered plasma testosterone and androstenedione levels, and this effect was dose related. In addition to an inhibition in gonadotropin release, a direct inhibition of Leydig cell

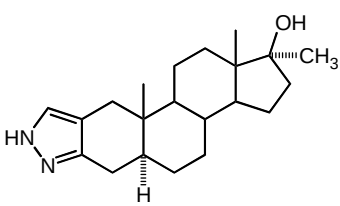

(**76**)

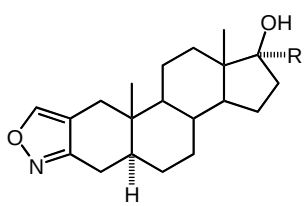

(**77**) R = H
(**78**) R = C≡CH

androgen synthesis was observed. Other studies have shown danazol to be effective for the treatment of endometriosis, benign fibrocystic mastitis, and precocious puberty. Several reports have appeared relating to its disposition and metabolic fate.

5.3.7 Esters and Ethers

Since esterification of testosterone markedly prolonged its activity, it was only natural that this approach to drug latentiation would be extended to the anabolic steroids. The acyl moiety is usually derived from a long-chain aliphatic or arylaliphatic acid such as heptanoic (enanthoic), decanoic, cyclopentylpropionic, and β-phenylpropionic. For example, no less than 12 esters of 19-nortestosterone (nandrolone) have been used clinically as long-acting anabolic agents.

In the case of nandrolone, the duration of action and the anabolic-to-androgenic ratio increased with the chain length of the ester group. The decanoate and laurate esters, for instance, were active 6 weeks after injection. Clinically, nandrolone decanoate appeared to be the most practical, since a dose of 25 to 100 mg/week produced marked nitrogen retention.

Since the 17α-alkyl group has been implicated as the cause of the hepatotoxic side effects of oral preparations, the effect of esterification on oral efficacy has attracted attention. For example, esterification of DHT with short chain fatty acids resulted in oral anabolic and androgenic activity in rats. Moreover, esters of methenolone possessed appreciable oral anabolic activity. Unfortunately, follow-up studies in humans have not been reported.

The manner in which the steroid esters evoke their enhanced activity and increased duration of action has puzzled investigators for many years. The classical concept has been that esterification delays the absorption rate of the steroid from the site of injection, thus preventing its rapid destruction. Other factors must be involved, however, since the potency and prolongation of action vary markedly with the nature of the esterifying acid.

The effect of various aliphatic esters of testosterone on rat prostate and seminal vesicles were studied and correlated androgenicity with lipophilicity and rate of ester hydrolysis by liver esterase. The peak androgenic response was observed with the butyrate ester, which was also the most readily hydrolyzed. The more lipophilic valerate ester was slightly less androgenic in a quantitative sense, but its action was longer lasting. It was concluded that the ease of hydrolysis controls the weight of the target organs, whereas lipophilicity was responsible for the duration of androgenic effect. These results explain the low androgenic activity previously noted for hindered trimethylacetate (pivalate) esters, which would be expected to be resistant to *in vivo* hydrolysis.

The effect of etherification on anabolic or androgenic activity has been studied less rigorously. Replacement of the 17β-OH with 17β-OCH_3 markedly reduced androgenic activity but did not greatly affect the ability to counteract cortisone-induced adrenal atrophy in male rats. A series of 17β-acetals, alkyl ethers, and 3-enol ethers, however, showed significant activity when given orally. The cyclohexyl enol ether of 17α-methyltestosterone, for example, was orally 5 times as myotrophic as 17α-methyltestosterone.

Other ethers such as the tetrahydropyranol and trimethylsilyl have oral anabolic and androgenic activity in animals. The

trimethylsilyl ether of testosterone (silandrone) had protracted activity following injection and orally had twice the anabolic and androgenic activities of 17α-methyltestosterone.

5.3.8 Summary of Structure–Activity Relationships

Synthetic modifications of C_{19} steroids have resulted in the enhancement of anabolic activity, even though a pure synthetic anabolic agent that retains no androgenic activity has not been accomplished. Structural changes in two regions of the testosterone molecule have resulted in the greatest enhancement of the anabolic/androgenic ratio. The first region is the C-17 position of the testosterone molecule. Introduction of the 17α-alkyl functionality, such as a 17α-methyl or a 17α-ethyl group, greatly increases the metabolic stability of the anabolic and decreases *in vivo* conversion of the 17β-alcohol to the 17-ketone by 17β-hydroxysteroid dehydrogenases. In addition, esterification of the 17β-alcohol enhances the lipid solubility of the steroids and provides injectable preparations for depot therapy.

The A-ring of testosterone is the second region in which structural modifications can be made to increase anabolic activity. Removal of the C-19 methyl group results in the 19-nortestosterone analogs, which have slightly higher anabolic activity. A major impact on the structure–activity relationships of anabolics can be observed with modifications at the C-2 position. Bioisosteric replacement of the carbon atom at position 2 with an oxygen provides a threefold increase in anabolic activity, as is seen with oxandrolone. Finally, the greatest effects were observed with the addition of heterocyclic rings fused at positions 2 and 3 of the A-ring. The two heterocycles that have lead to the greatest changes are the pyrazole and the isoxazole rings, as seen in stanozolol and androisoxazole, respectively. In these anabolics, the 3-ketone of testosterone is replaced by the bioisosteric 3-imine. Stanozolol, which contains the pyrazole ring at C-2 and C-3, shows the greatest increases when compared to testosterone. Table 9 compares the anabolic activities of nitrogen retention and myotrophic activity for several common anabolics. Figure 10 contains a summary of the structure–activity relationships for anabolic agents.

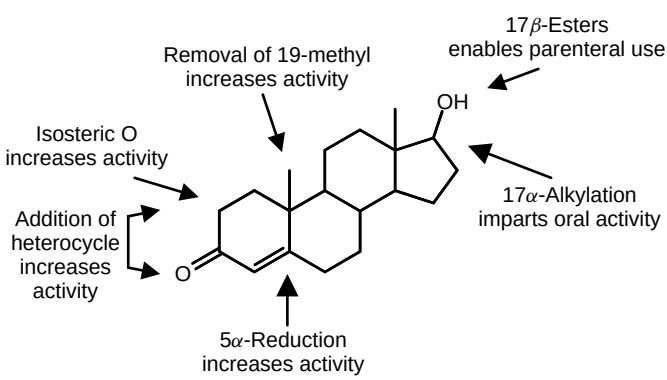

Fig. 10 Summary of structure–activity relationships for anabolic agents

Tab. 9 Comparison of anabolic activities.

Compound	*Number*	*Trade names*	*Anabolic activity*	
			Nitrogen retention	*Myotrophic activity*
Testosterone	(**1**)	Android-T Malestrone Oreton Primotest Virosterone	1.0	1.0
19-Nortestosterone Nandrolone	(**43**)	Nerobolil Nortestonate	0.8	1.0
Normethandrone	(**44**)	Methalutin Orgasteron	4.0	4.5
Norethandrolone	(**45**)	Nilevar Solevar	3.9	4.0
Methandrostenolone Methandienone	(**48**)	Danabol Dianabol Nabolin Nerobil	0.6	1.4
Drostanolone	(**56**)	Drolban Masterone	–	1.3
Oxymetholone	(**63**)	Adroyd Anadrol Anadroyd Anapolon Anasterone Nastenon Protanabol Synasteron	2.75	2.8
Oxandrolone	(**65**)	Anavar Provita	3.0	3.0
Estrenol	(**70**)	Duraboral-O Maxibolin Orabolin Orgaboral Orgabolin	1.7	2.0
Stanozolol	(**75**)	Stanozol Winstrol Tevabolin	10.0	7.5
Androisoxazole	(**76**)	Androxan Neo-ponden	1.5	1.7

5.4 Absorption, Distribution, and Metabolism

The absorption, distribution, and metabolism of the various anabolic steroids is quite similar to those pharmacokinetic properties of the endogenous and synthetic androgens discussed earlier in the chapter. Again, lipid solubility is critical for the absorption of these agents following oral or parenteral administration. The 17α-methyl group retards the metabolism of the compounds and provides orally active agents. Other anabolics such as methenolone are orally active without a 17α-substituent, indicating that these steroids are poor substrates for 17α-hydroxysteroid dehydrogenase. Reduction of the 4-en-3-one system in synthetic anabolics to give the various α- and β-isomers occurs *in vivo*. The 3-deoxy agent 17α-methyl-5α-androstan-17α-ol was shown to be extensively converted to the 3-keto derivative by liver homogenate preparations. The metabolic fates of stanozolol and danazol have been reported, with the major metabolites being heterocycle-ring opened derivatives and their deaminated products. Finally, both the unchanged anabolics and their metabolites are primarily excreted in the urine as the glucuronide or sulfate conjugates.

5.5 Toxicities

The major side effect of the anabolic steroids is the residual androgenic activity of the molecules. The virilizing actions are undesirable in adult males as well as in females and children. In addition, many anabolic steroids can suppress the release of gonadotropins from the anterior pituitary and lead to lower levels of circulating hormones and potential reproductive problems. Headaches, acne, and elevated blood pressure are common in individuals taking anabolics. The salt and water retention induced by these agents can produce edema.

The most serious toxicities resulting from the use of anabolic steroids are subsequent liver damages. Liver damage including jaundice and cholestasis can occur after use of the 17α-alkylated C_{19} steroids. Also, individuals who have received anabolic agents over an extended period have developed hepatic adenocarcinomas. Such clinical reports serve to underscore the inherent risks associated with anabolic steroid use in amateur athletes for no demonstrable benefits.

6 Androgen Antagonists

A majority of the recent research efforts in the area of androgens has concentrated on the preparation and biological activities of androgen antagonists. An androgen antagonist is defined as a substance that antagonizes the actions of testosterone in various androgen-sensitive target organs and, when administered with an androgen, blocks or diminishes the effectiveness of the androgen at various androgen-sensitive tissues. Androgen antagonists may act to block the action of testosterone at several possible sites. First, such compounds could interfere with the entrance of the androgen into the target cell. A second site of action of androgen antagonists may be to block the conversion of testosterone to its more active metabolite dihydrotestosterone or DHT. Third, competition for the high-affinity binding sites on the androgen receptor molecule may account for antiandrogenic effects. Finally, certain agents such as LHRH agonists can act in the

pituitary to lower gonadotropin secretion via receptor downregulation and thus diminish the production of testosterone by the testis. The substances described in this section act through at least one of these mechanisms.

6.1 Current Drugs on the Market

See Table 10.

6.2 Antiandrogens

6.2.1 Therapeutic Uses

Antiandrogens are agents that compete with endogenous androgens for the hormone binding site on the androgen receptor. These agents have therapeutic potential in the treatment of acne, virilization in women, hyperplasia and neoplasia of the prostate, baldness and male contraception, and clinical studies have demonstrated the potential therapeutic benefits of the antiandrogens. The applications of antiandrogens for the treatment of prostatic carcinoma and for the treatment of BPH have also been investigated. Antiandrogens are effective for the treatment of prostate cancer when combined with androgen ablation, such as surgical (orchiectomy) or medical (LHRH agonist) castration.

6.2.2 Structure–Activity Relationships for Antiandrogens

6.2.2.1 **Steroidal agents.** Several steroidal and nonsteroidal compounds with demonstrated antiandrogenic activity have been utilized clinically. The first compounds used as antiandrogens were the estrogens and progestins. Steroidal estrogens and diethylstilbesterol are used in the treatment of prostatic carcinoma and exert their action via suppression of the release of pituitary gonadotropins. Progestational compounds have also been utilized for antiandrogenic actions with limited success. The inherent hormonal activities of these compounds and the development of more selective antiandrogens have limited the clinical applications of estrogens and progestins as antiandrogens.

A modified progestin that is a potent antiandrogen and has minimal progestational activity is the agent cyproterone acetate (**79**). This compound was originally prepared in search of orally active

Tab. 10

Generic name (structure)	*Trade name*	*US manufacturer*	*Chemical class*	*Dose*
Antiandrogens				
Flutamide (**91**)	Eulexin	Schering-Plough	Nonsteroidal	Tablets: 125 mg
Bicalutamide (**97**)	Casodex	AstraZeneca	Nonsteroidal	Tablets: 50 mg
Nilutamide (**93**)	Nilandron	Aventis	Nonsteroidal	Tablets: 50 mg 150 mg
Cyproterone acetate (**78**)	Androcur	Schering AG	Pregnane	
5α-Reductase inhibitors				
Finasteride (**104**)	Proscar	Merck	Androstane	Tablets: 5 mg
	Propecia	Merck	Androstane	Tablets: 1 mg

progestins, but was quickly recognized for its ability to suppress gonadotropin release. It was later demonstrated that this compound also bound with high affinity to the androgen receptor and thus competed with DHT for the binding site. Cyproterone acetate has received the most clinical attention in antiandrogen therapy. Cyproterone acetate has produced quite satisfactory results in the treatment of acne, seborrhea, and hirsutism. Therapeutic effectiveness of this agent in the treatment of prostatic carcinoma has been reported. Cyproterone acetate was reported to be a good alternative to estrogens for the treatment of prostate cancer when combined with androgen ablation. However, this combination did not improve disease-free survival or overall survival when compared to castration alone.

Other pregnane compounds that exhibit antiandrogenic actions by binding to the androgen receptor are chlormadinone acetate (**80**), medroxyprogesterone acetate (**81**), medrogesterone (**82**), A-norprogesterone (**83**) and gestonorone capronate (**84**). In addition, medrogesterone exerts antiandrogenic effects by inhibiting 5α-reductase and thus preventing the formation of DHT. Gestonorone capronate interferes with the uptake process in target cells.

Several androstane derivatives demonstrate antiandrogenic properties. 17α-Methyl-B-nortestosterone (**85**) was prepared and first tested for antihormonal

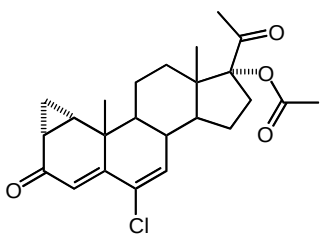

(**79**)

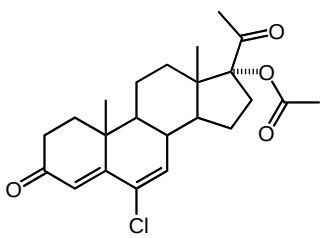

(**80**)

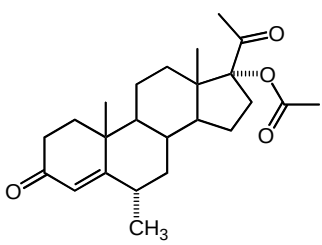

(**81**)

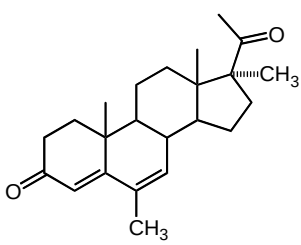

(**82**)

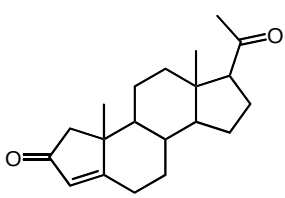

(**83**)

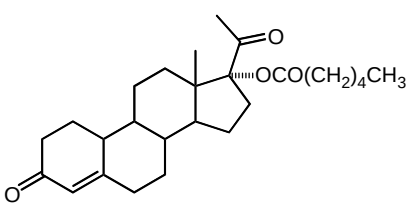

(**84**)

activity. Several other androstane analogs were found to possess antiandrogenic activity, including BOMT (**86**), R2956 (**87**), and oxendolone (**88**). As expected, the mechanism of antiandrogenic action of these synthetic steroids is the competition with androgens for the binding sites on the receptor molecule. Numerous A- and B-ring modified steroids were examined for antiandrogenic activity and the ability to bind to the androgen receptor, demonstrating that the structural requirements of receptor binding site can accommodate some degree of flexibility in the A- and/or B-rings of antiandrogenic molecules. Heterocyclic-substituted A-ring antiandrogens such as zanoterone (WIN 49,596) (**89**) further support these conclusions on the structure-activity relationships of steroidal antiandrogens. Additional A-ring heterocycles identified as novel antiandrogens are the thiazole (**90**) and oxazole (**91**). The optimal substitutions on the A-ring heterocyclic androstanes for *in vivo* antiandrogenic activity are the methylsulfonyl group at the N-1′ position and a 17α-substituent (e.g. 17α-methyl or 17α-ethinyl).

6.2.2.2 **Nonsteroidal agents.** The absolute requirement of a steroidal compound for interaction with the androgen receptor was invalidated when the potent nonsteroidal antiandrogen flutamide (Eulexin, **92**) was introduced. Subsequent receptor studies demonstrated that this compound competed with DHT for the binding sites. The side chain of flutamide allows sufficient flexibility for the molecule to assume a structure similar to an androgen. In addition, a hydroxylated metabolite (**93**) has been identified, which is a more

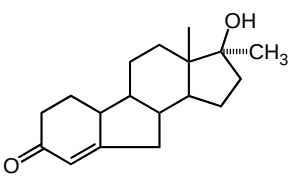

(**85**)

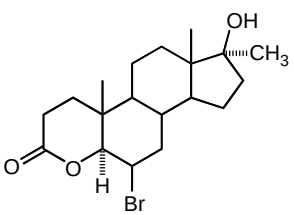

(**86**)

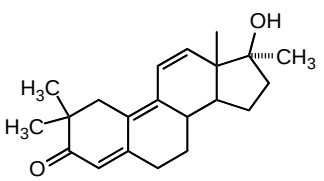

(**87**)

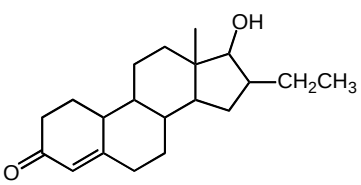

(**88**)

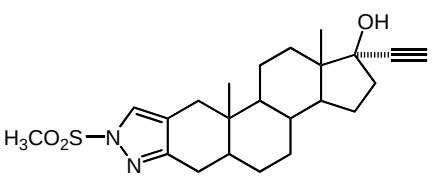

(**89**)

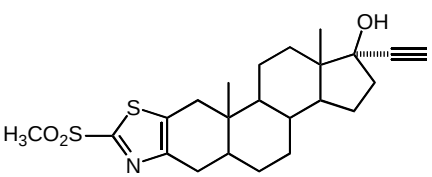

(**90**)

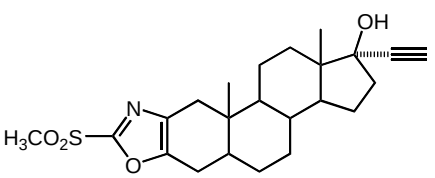

(**91**)

(**92**) X = H
(**93**) X = OH

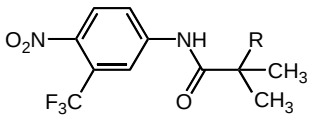

powerful antiandrogen *in vivo*, and has a higher affinity for the receptor than the parent compound. Important factors in the structure–activity relationships of flutamide and analogs are the presence of an electron-deficient aromatic ring and a powerful hydrogen-bond donor group.

Flutamide has been extensively evaluated for the treatment of prostate cancer. Large double-blind studies in prostate cancer patients using a combination of flutamide with an LHRH agonist (as a medical castration) resulted in an increased number of favorable responses and increased overall survival when compared to an LHRH agonist or surgical castration.

Nilutamide (Anandron, **94**), and related nilutamide analogs (**95**–**97**), and bicalutamide (Casodex, **98**) are other nonsteroidal antiandrogens with a similar electron-deficient aromatic ring and

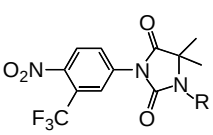

(**94**) R = H
(**95**) R = $(CH_2)_4OH$

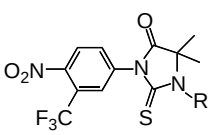

(**96**) R = CH_3
(**97**) R = $(CH_2)_4OH$

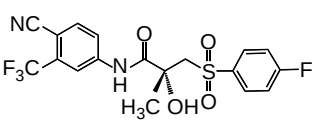

(**98**)

have been shown to interact with the androgen receptor to varying degrees. Nilutamide and bicalutamide are pure antiandrogens and are effective in suppressing testosterone-stimulated cell proliferation. Both nilutamide and bicalutamide have demonstrated effectiveness against prostate cancer.

Other aryl substituted nonsteroidal compounds have also been identified as antiandrogens. DIMP (**99**) is a phthalimide derivative that showed weak affinity for the androgen receptor and poor *in vivo* activity. A series of tetrafluorophthalimides such as (**100**) demonstrated moderate activity as antiandrogens in cell proliferation assays.

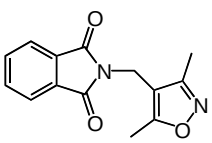

DIMP (**99**)

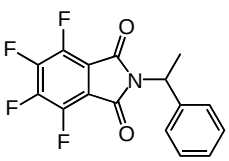

(**100**)

6.2.3 Selective Androgen Receptor Modulators (SARMs)

Current research efforts in the medicinal chemistry and pharmacology of androgens and antiandrogens have focused on the development of ligands that produce tissue-selective interactions with the androgen receptor. These ligands, referred to as *selective androgen receptor modulators* (SARMs), may act as antagonists or weak agonists in one tissue but act as strong agonists in another tissue type. Such SARMs would theoretically enable effective treatment of certain disorders affecting bone or muscle with little or no effect on the prostate.

A series of 1,2-dihydropyridono[5,6-*g*]quinolines were identified as novel nonsteroidal antiandrogens on the basis of a cell-based screening approach. Several analogs (**101**–**104**) demonstrated excellent *in vivo* activity, reducing rat ventral prostate weight without affecting serum

gonadotropins and serum testosterone levels.

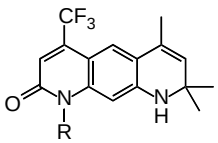

(**101**) R = H
(**102**) R = CH_3

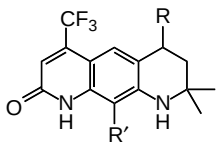

(**103**) R,R′ = H
(**104**) R,R′ = CH_3

Structural modifications of the antiandrogen bicalutamide have resulted in molecules (**105**–**107**) that produced weak partial agonist activities in the prostate and seminal vesicles (androgen-responsive tissues) and full agonist activity in the levitor ani muscle (anabolic effect) in male rats. Furthermore, these agents did not affect serum gonadotropins and serum testosterone levels, again emphasizing the tissue selectivity for their actions.

6.2.4 Absorption, Distribution, and Metabolism

The steroidal antiandrogens exhibit similar pharmacokinetic properties to the androgens and anabolic agents. The lipophilicity of the compounds influences absorption both orally and from injection sites. Reduction of the 3-ketone and 4,5 double bond are common routes of metabolism. An unusual metabolite of cyproterone acetate, 15α-hydroxycyproterone acetate, was isolated and identified in both animals and man. The nonsteroidal antiandrogen flutamide is rapidly absorbed and extensively metabolized *in vivo*. As described earlier, the hydroxy metabolite (**93**) of flutamide is a more potent antiandrogen. The major metabolite of bicalutamide is the sulfone, which has comparable *in vivo* activity. Finally, the antiandrogens are primarily excreted as the glucuronide and sulfate conjugates in the urine.

6.2.5 Toxicities

Side effects of these agents have been identified from various clinical trials. Testicular atrophy and decreased spermatogenesis have been observed during treatment with cyproterone acetate. Certain antiandrogens such as cyproterone and medrogesterone also exhibit inherent progestational activity, suppress corticotropin release, and have some androgenic effects. No hormonal activities were observed for the nonsteroidal antiandrogens, such as flutamide. On the other hand, many nonsteroidal antiandrogens exhibit other endocrine side effects, such as elevated serum gonadotropins and serum testosterone levels. Gynecomastia, nausea, diarrhea, and liver toxicities have been observed in patients on nonsteroidal antiandrogens. Also, resistance to antiandrogen therapy has been observed in prostate cancer patients.

6.3 Enzyme Inhibitors

Enzymes involved in the biosynthesis and metabolism of testosterone are attractive targets for drug design and drug development. Suppression of the synthesis

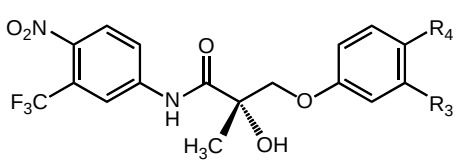

(**105**) R_3 = H, R_4 = F
(**106**) R_3 = H, R_4 = $NHC(O)CH_3$
(**107**) R_3 = F, R_4 = Cl

of androgenic hormones and androgen precursors is a viable therapeutic approach for the treatment of various androgen-mediated disease processes and an important endocrine treatment for prostate cancer. Potent inhibition of type 2 5α-reductase in androgen target tissues and the resultant decrease in DHT levels will provide selective interference with androgen action within those target tissues and no alterations of other effects produced by testosterone, other structurally related steroids, and other hormones such as corticoids and progesterone. The cytochrome $P450_{17\alpha}$ enzyme complex displays two enzymatic activities – 17α-hydroxylation to produce 17α-hydroxysteroids and $C_{17}-C_{20}$ bond cleavage (17,20-lyase activity) to produce androgens. In the male, this enzyme is found in both testicular and adrenal tissues, with these organs providing circulating androgens in the blood. Effective inhibition of this microsomal enzyme complex would eliminate both testicular and adrenal androgens and remove the growth stimulus to androgen-dependent prostate carcinoma. Synthetic androgen analogs that inhibit the oxidative metabolism of androgens testosterone and androstenedione to estrogens estradiol and estrone can serve as potential therapeutic agents for controlling estrogen-dependent diseases such as hormone-dependent breast cancer.

6.3.1 5α-Reductase Inhibitors

The most extensively studied class of 5α-reductase inhibitors is the 4-azasteroids, which includes the drug finasteride (Proscar, **108**). Finasteride is the first 5α-reductase inhibitor approved in the United States for the treatment of BPH. This drug has approximately a 100-fold greater affinity for type 2 5α-reductase than for the type 1 enzyme, demonstrating an IC_{50} value of 4.2 nM for type 2 5α-reductase. In humans, finasteride decreases prostatic DHT levels by 70 to 90% and reduces prostate size, while testosterone tissue levels increased. Clinical trials demonstrated sustained improvement in BPH disease and reduction in PSA levels. Related analogs (**109–111**)

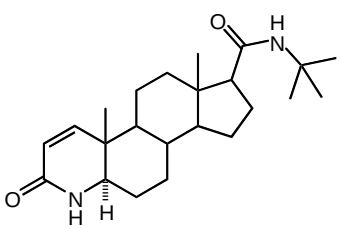

Finasteride (**108**)

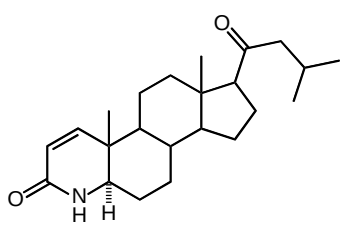

(**109**)

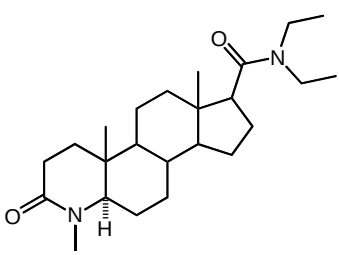

(**110**)

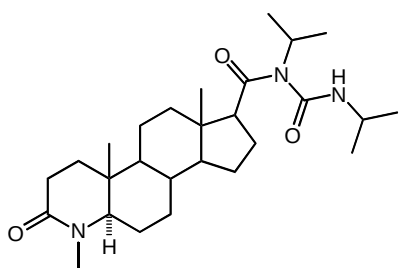

(**111**)

have also demonstrated effectiveness *in vitro* and *in vivo*. These agents were originally designed to mimic the putative 3-enolate intermediate of testosterone and serve as transition-state inhibitors. Subsequently, finasteride was shown to produce time-dependent enzyme inactivation and function as a mechanism-based inactivator. The structure–activity relationships for the 4-azasteroids illustrate the stringent requirements for inhibition of human type 2 5α-reductase. The 5α-reduced azasteroids are preferred, a 1,2-double bond can be tolerated, and the nitrogen can be substituted with only hydrogen or small lipophilic groups. Lipophilic amides or ketones are preferred as substituents at the C-17β position.

Several 6-azasteroids, such as **(112)** and**(113)**, were prepared as extended mimics of the enolate transition state and have also demonstrated potent inhibition of 5α-reductase. The 6-azasteroids are more effective inhibitors of type 2 5α-reductase, but some analogs also exhibit good inhibition of type 2 5α-reductase. Alkylation of the nitrogen can be tolerated; however, a 1,2-double bond decreases inhibitory activity in this series. The best inhibitors contain large lipophilic substituents at the C-17β position. Figure 11 contains a summary

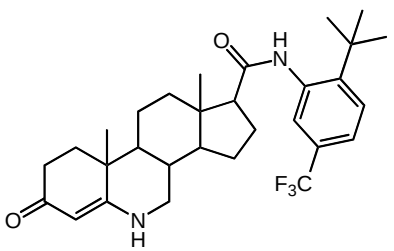

(**112**)

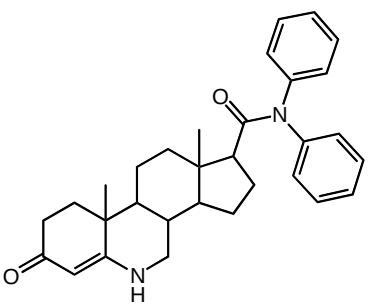

(**113**)

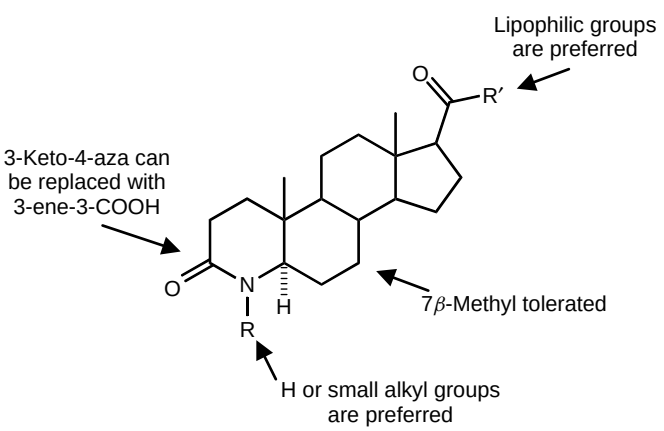

Fig. 11 Summary of structure–activity relationships for 5α-reductase inhibitors.

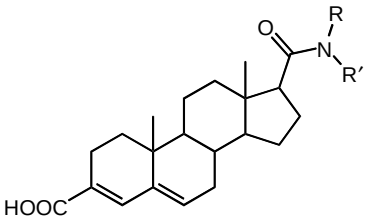

(**114**) R = H; R′ = $C(CH_3)_3$
(**115**) R = R′ = $CH(CH_3)_2$

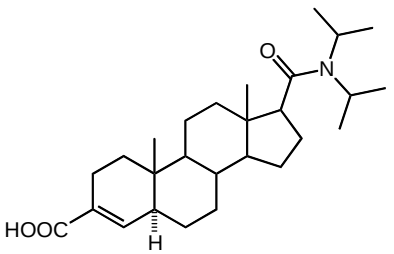

(**116**)

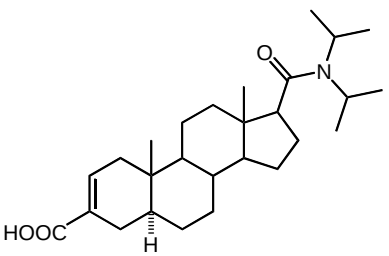

(**117**)

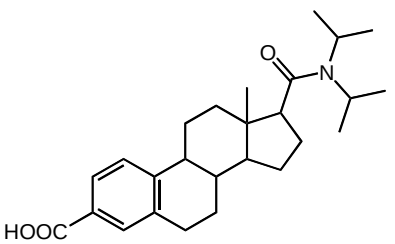

(**118**)

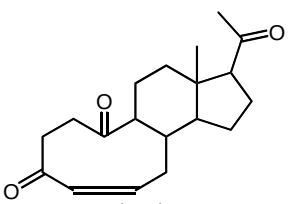

(**119**)

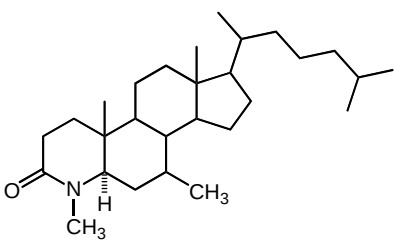

(**120**)

of the structure–activity relationships for steroidal 5α-reductase inhibitors.

Androstadiene 3-carboxylic acids **(114)** and **(115)** were recently designed as transition-state inhibitors and have demonstrated potent uncompetitive inhibition of type 2 5α-reductase. Epristeride (SK&F 105,657, **114**) has demonstrated the ability to lower serum DHT levels by 50% in clinical trials. Other analogs with acidic functionalities at the C-3 position include other androstene carboxylic acids (**116, 117**) and estratriene carboxylic acids (**118**). The allenic secosteroid (**119**) has demonstrated potent irreversible inhibitor of 5α-reductase, even though it was originally developed as an irreversible inhibitor of 3β-hydroxysteroid dehydrogenase/$\Delta^{4,5}$-isomerase. Finally, selective and potent inhibitors of type 1 5α-reductase were developed on the basis of the 4-azacholestane MK-386 (**120**).

Several nonsteroidal 5α-reductase inhibitors have been developed on the basis of the azasteroid molecule or from high-throughput screening methods. Examples of these nonsteroidal inhibitors include the benzoquinolinone (**121**), an aryl carboxylic acid (**122**), and FK143 (**123**).

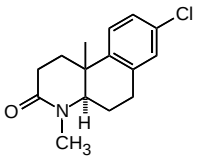

(**121**)

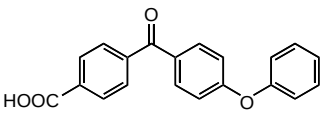

(**122**)

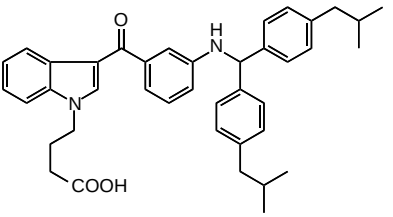

FK143 (**123**)

6.3.2 17,20-Lyase Inhibitors

Both nonsteroidal and steroidal agents have been examined as inhibitors of 17α-hydroxylase/17,20-lyase. The nonsteroidal agents studied most extensively are aminoglutethimide (**124**) and ketoconazole (**125**), both *in vitro* and in clinical trials. Objective response rates for treatment of prostate cancer in relapsed patients were observed with high doses of aminoglutethimide and high doses of ketoconazole, but both agents produce frequent side effects. A third nonsteroidal agent that has received extensive preclinical evaluation is the benzimidazole analog, liarozole (**126**), which produced reduction in plasma testosterone and androstenedione levels *in vivo*. Other nonsteroidal agents reported to exhibit 17α-hydroxylase/17,20-lyase inhibitory activity *in vitro* include other imidazole analogs, nicotine, bifluranol analogs, and pyridylacetic acid esters. In general, high doses of nonsteroidal agents are needed to produce significant *in vitro or in vivo* activity. Another potential problem with these agents is nonspecific inhibition of other cytochrome P450 enzymes involved in either steroidogenesis or liver metabolism.

A few studies of steroidal inhibitors of 17α-hydroxylase/17,20-lyase have been reported. An extensive analysis of the specificity of steroid binding to testicular microsomal cytochrome P450 identified several steroids exhibiting binding affinity. One of these, promegestrone (**127**), has been utilized in kinetic analysis of purified cytochrome $P450_{17\alpha}$. An affinity label inhibitor, 17-bromoacetoxyprogesterone (**128**), alkylates a unique cysteine residue on purified cytochrome $P450_{17\alpha}$. Potential mechanism-based inhibitors include 17β-(cyclopropylamino)-5-androsten-3β-ol (**129**) and 17β-vinylprogesterone (**130**). To date, all of these inhibitors exhibit apparent K_i's in the micromolar (μM) range, while the apparent K_m for progesterone is 140 nM. The 17β-aziridinyl analog (**131**) and 17β-pyridyl derivative (**132**) also exhibited similar inhibitory activity.

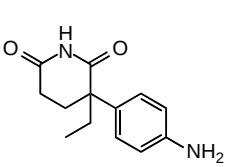

Aminoglutethimide
(**124**)

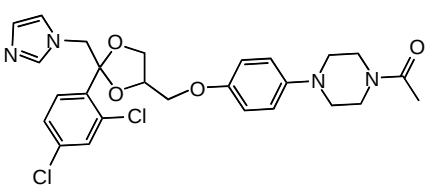

Ketoconazole
(**125**)

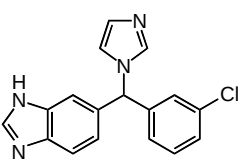

Liarozole
(**126**)

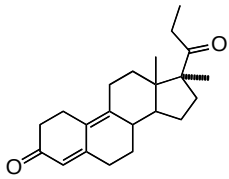

(**127**)

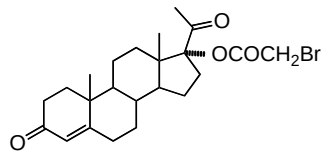

(**128**)

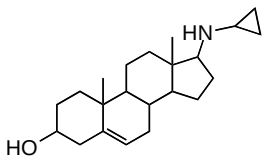

(**129**)

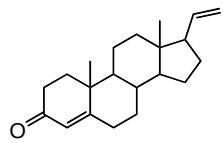

(**130**)

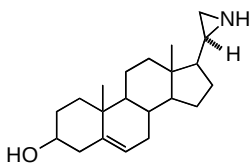

(**131**)

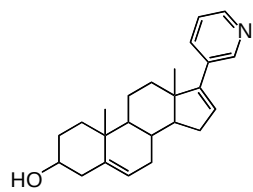

(**132**)

6.3.3 C_{19} Steroids as Aromatase Inhibitors

Aromatase is the enzyme complex that catalyzes the conversion of androgens into the estrogens. This enzymatic process is the rate-limiting step in estrogen biosynthesis and converts C_{19} steroids, such as testosterone and androstenedione, into the C_{18} estrogens, estradiol, and estrone, respectively. Inhibition of aromatase has been an attractive approach for examining the roles of estrogen biosynthesis in various physiological or pathological processes. Furthermore, effective aromatase inhibitors can serve as potential therapeutic agents for controlling estrogen-dependent diseases such as hormone-dependent breast cancer. Investigations on the development of aromatase inhibitors began in the 1970s and have expanded greatly in the past two decades.

Steroidal inhibitors that have been developed to date build upon the basic androstenedione nucleus and incorporate chemical substitutents at varying positions on the steroid. These inhibitors bind to the aromatase cytochrome P450 enzyme in the same manner as the substrate androstenedione. Even though the steroidal aromatase inhibitors are C_{19} steroids, these agents exhibit no significant androgenic activity. A limited number of effective inhibitors with substituents on the A-ring have been reported. Several steroidal aromatase inhibitors contain modifications at the C-4 position, with 4-hydroxyandrostenedione (**133**) (4-OHA; formestane) being the prototype agent. Initially, 4-OHA was thought to be a competitive inhibitor, but was later demonstrated to produce enzyme-mediated inactivation. *In vivo*, 4-OHA inhibits reproductive process and causes regression of hormone-dependent mammary rat tumors. 4-OHA is effective in the treatment of advanced breast cancer

in postmenopausal women, and this drug is approved in the United Kingdom for breast cancer therapy. Thus, the spacial requirements of the A-ring for binding of the steroidal inhibitor to aromatase are rather restrictive, permitting only small structural modifications to be made. Incorporation of the polar hydroxyl group at C-4 enhances inhibitory activity. 1-Methyl-1,4-androstadiene-3,17-dione **(134)** is a potent inhibitor of aromatase *in vitro* and *in vivo*; on the other hand, bulky substitutents at the 1α-position are poor inhibitors. At the C-3 position, replacement of the ketone with a methylene provided effective inhibitors.

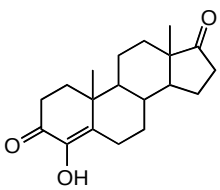

(**133**)

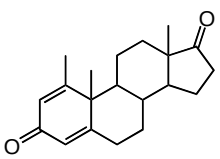

(**134**)

More extensive structural modifications may be made on the B-ring of the steroid nucleus. Bulky substitutions at the C-7 position of the B-ring have provided several very potent aromatase inhibitors. 7α-(4′-Amino)phenylthio-4-androstene-3,17-dione (**135**) (7α-APTA) is a very effective competitive inhibitor, with an apparent K_i of 18 nM. This inhibitor has also demonstrated effectiveness in inhibiting aromatase in cell cultures and in treating hormone-dependent rat mammary tumors. Evaluation of various substituted aromatic analogs of 7α-APTA provided no correlation between the electronic character of the substituents and inhibitory activity. Investigations of various seven-substituted 4,6-androstadiene-3,17-dione derivatives suggest that only those derivatives that can project the 7-aryl substitutent into the 7α pocket are effective inhibitors. More metabolically stable inhibitors were synthesized by replacing the thioether linkage at the 7α-position with a carbon–carbon linkage, such as 7α-phenethylandrost-4-ene-3,17-dione (**136**). Overall, the most effective B-ring modified aromatase inhibitors are those with 7α-aryl derivatives, with several analogs having 2 to 10 times greater affinity for the enzyme than the substrate. These results suggest that additional interactions occur between the phenyl ring at the 7α-position and amino acids at or near the enzymatic site of aromatase, resulting in enhanced affinity.

Numerous modified androstenedione analogs have been developed as effective mechanism-based inhibitors of aromatase. The first compound designed as a mechanism-based inhibitor of aromatase was 10-propargyl-4-estrene-3,17-dione **(137)** (PED; MDL 18,962). MDL 18,962 has an electron-rich alkynyl function on the C-19 carbon atom, the site of aromatase-mediated oxidation of the

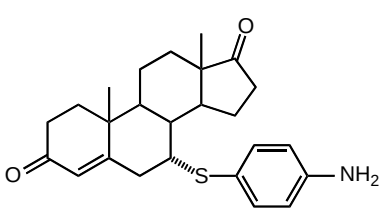

(**135**)

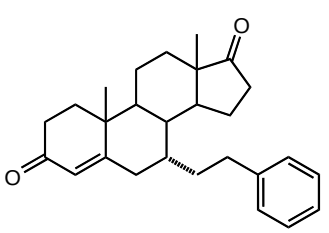

(**136**)

substrate. Although the identity of the reactive intermediate formed is not known, an oxirene and a Michael acceptor have been suggested. This agent is an effective inhibitor *in vitro* and *in vivo*. Other approaches to C-19 substituted mechanism-based inhibitors containing latent chemical groups have provided a limited number of inhibitors. These agents include the difluoromethyl analog **(138)** and a thiol **(139)**. Another series of mechanism-based inhibitors have developed from more detailed biochemical investigations of several inhibitors originally thought to be competitive inhibitors. These inhibitors can be grouped into the general categories of 4-substituted androst-4-ene-3,17-diones, such as 4-hydroxyandrostenedione **(133)**, and substituted androsta-1,4-diene-3,17-diones, such as 7α-(4′-amino)-phenylthioandrosta-4,6-diene-3,17-dione (7α-APTADD; **140**), 7α-phen-propylandrosta-1,4-diene-3,17-dione (**141**), and 6-methyleneandrost-4-ene-3,17-dione (exemestane; **142**). Exemestane (Aromasin) is marketed as second-line therapy for the treatment of breast cancer patients who failed tamoxifen and is effective as first-line therapy in women with advanced breast cancer.

6.3.4 Other Agents

An interesting natural product that lowers circulating androgen levels *in vitro* is the nonsteroidal agent gossypol. Gossypol, (2,2′-binaphthalene)-8,8′-dicarboxaldehyde-1,1′,6,6′,7,7′-hexahydroxy-5,5′-diisopropyl-3,3′-dimethyl **(143)** is a

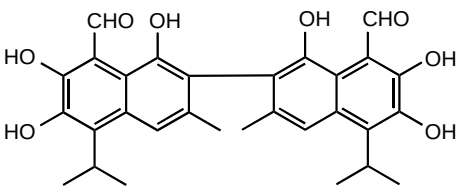

(**143**)

polyphenolic compound contained in the pigment of cottonseed. This natural product has been used in fertile men in China as an effective male contraceptive agent for many years; its antifertility effects on reproductive endocrine tissues are observed at 1000-fold lower doses than its toxic effects in other tissues. Gossypol has been shown to disrupt spermatogenesis by inhibiting lactate dehydrogenase-X (LDH-X), to interfere with steroidogenesis in testicular Leydig cells, and to hinder the function of primary cultures of Leydig and Sertoli

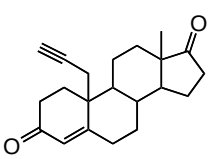

(**137**)

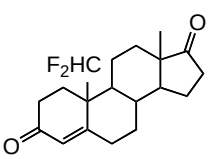

(**138**)

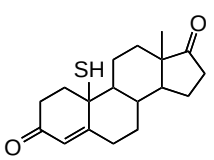

(**139**)

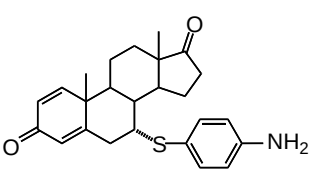

(**140**)

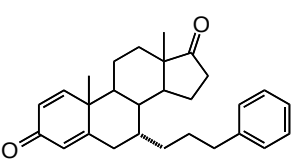

(**141**)

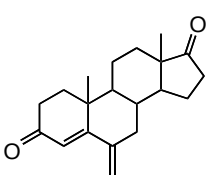

(**142**)

cells. In addition, gossypol is also capable of altering steroid biosynthesis in the female reproductive systems. The antisteroidogenic effect of gossypol in cultured bovine luteal cells involves suppression of the activities of adenylate cyclase and 3β-hydroxysteroid dehydrogenase (3β-HSD). Reproductive endocrine tissues are also sensitive to metabolites of gossypol, such as gossypolone. Gossypolone, a major metabolite formed by gossypol oxidation, inhibits both 3β-HSD and cytochrome P450scc activities in cultured bovine luteal cells and suppresses adrenocorticotropic hormone-induced corticosterone secretion in cultured rat adrenocortical cells at a similar potency as gossypol. Gossypol metabolites have also demonstrated inhibitory action on hCG-induced testosterone production in young male rats. The major side effects of gossypol therapy fatigue, gastrointestinal upset, weakness, and hypokalemia thus limit its therapeutic usefulness.

7 Summary

The steroid testosterone is the major circulating sex hormone of the male and serves as the prototype for the androgens, the anabolic agents, and androgen antagonists. The endogenous androgens are biosynthesized from cholesterol in various tissues in the body; the majority of the circulating androgens are made in the testes under the stimulation of the gonadotropin LH. A critical aspect of testosterone and its biochemistry is that this steroid is converted in various cells to other active steroidal agents. The reduction of testosterone to dihydrotestosterone is necessary for the androgenic actions of testosterone in androgen target tissues such as the prostate. On the other hand, oxidation of testosterone by the enzyme aromatase to yield the estrogens is crucial for certain CNS actions. Investigations of these enzymatic conversions of circulating testosterone continue to be a fruitful area of biochemical research on the roles of the steroid hormones in the body. Additionally, the elucidation of the mechanism of action of the androgens in various target tissues receives ongoing emphasis. The androgenic actions of testosterone are due to the binding of DHT to its nuclear receptor, followed by dimerization of the receptor complex and binding to a specific DNA sequence. This binding of the homodimer to the androgen response element leads to gene expression, stimulation of the synthesis of new mRNA, and subsequent protein biosynthesis. Other actions of testosterone, particularly the anabolic actions, appeared to be mediated through a similar nuclear receptor-mediated mechanism. Many of the intricate biochemical events that occur during the action of the androgens in their target cells remain for further clarification. Nevertheless, receptor studies of new agents are an important biological tool in the evaluation of the compounds for later, in-depth pharmacological testing.

The synthetic androgens and anabolics were prepared to impart oral activity to the androgen molecule, to separate the androgenic effects of testosterone from its anabolic effects and to improve upon its biological activities. These research efforts have provided several effective drug preparations for the treatment of various androgen-deficient diseases, for the therapy of diseases characterized by muscle wasting and protein catabolism, for postoperative adjuvant therapy, and for the treatment of certain hormone-dependent cancers. The synthetic anabolics have also resulted in the abuse of these agents in

athletics. Finally, the most recent area of research emphasis is the development of the androgen antagonists, both steroidal and nonsteroidal agents. The two major categories of these antagonists are the antiandrogens, which block interactions of androgens with the androgen receptor, and the inhibitors of androgen biosynthesis and metabolism. Such compounds have therapeutic potential in the treatment of acne, virilization in women, hyperplasia and neoplasia of the prostate, and baldness and for male contraception. The current focus on the development of SARMs enables tissue-selective activation of the androgen receptor, such that these ligands may act as strong agonists in one tissue while being antagonists or weak agonists in another tissue type. A number of androstane derivatives are also being developed as inhibitors of aromatase for the treatment of hormone-dependent breast cancer. Thus, the numerous biological effects of the male sex hormones testosterone and dihydrotestosterone and the varied chemical modifications of the androstane molecule have resulted in the development of effective medicinal agents for the treatment of androgen-related diseases.

Acknowledgment

The present text is a modified and updated version of the chapter ''Male Sex Hormones, Analogs, and Antagonists'' published in ''Burger's Medicinal Chemistry and Drug Discovery'' (ISBN 0-471-37029-0) edited by D.J. Abraham in 2003.

See also Male Reproductive System: Testis Development and Spermatogenesis; Medicinal Chemistry; Receptor Biochemistry.

Bibliography

Books and Reviews

Bhasin, S., Gabelnick, H.L., Spieler, J.M., Swerdloff, R.S., Wang, C., Kelly, C. (Eds.) (1996) *Pharmacology, Biology, and Clinical Applications of Androgens*, Wiley-Liss, New York.

Brueggemeier, R.W. (1986) Androgens, Anabolics, and Antiandrogens, in: Verderame, M. (Ed.) *CRC Handbook of Hormones, Vitamins, and Radiopaques*, CRC Press, Boca Raton, pp. 1–49.

Brueggemeier, R.W. (2003) Male Sex Hormones, Analogs, and Antagonists, in: Abraham, D. (Ed.) *Burger's Medicinal Chemistry & Drug Discovery*, Vol. 3, 6th edition, John Wiley & Sons, New York, pp. 679–746.

Brueggemeier, R.W. (2004) Aromatase inhibitors: new endocrine treatment of breast cancer, *Semin. Reprod. Med.* **22**, 31–43.

Chang, C. (Ed.) (2002) *Androgens and Androgen Receptor: Mechanisms, Functions, and Clinical Applications*, Kluwer Academic Publishers, Boston.

Eik-Nes, K.B. (1970) *The Androgens of the Testes*, Dekker, New York.

Griffin, J.E., Wilson, J.D. (1989) The Androgen Resistance Syndromes: 5α-Reductase Deficiency, Testicular Feminization, and Related Syndromes in: Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D. (Eds.) *The Metabolic Basis of Inherited Diseases*, 6th edition, McGraw-Hill, New York, p. 1919.

Griffin, J.E., Wilson, J.D. (1992) Disorders of the Testes and the Male Reproductive Tract, in: Wilson, J.D., Foster, D.W. (Eds.) *William's Textbook of Endocrinology*, 9th edition, WB Saunders, Philadelphia, p. 819.

Heinlein, C.A., Chang, C. (2004) Androgen receptor in prostate cancer, *Endocr. Rev.* **25**, 276–308.

Kochakian, C.D. (1976) *Anabolic-Androgenic Steroids*, Springer-Verlag, New York.

Martini, L., Motta, M. (Eds.) (1977) *Androgens and Antiandrogens*, Raven Press, New York.

Miller, W.L. (1988) Molecular biology of steroid hormone synthesis, *Endocr. Rev.* **9**, 295–318.

Mooradian, A.D., Morley, J.E., Korenman, S.G. (1987) Biological actions of androgens, *Endocr. Rev.* **8**, 1.

Munson, P.L., Diczfalusy, E., Glover, J., Olsen, R.E. (Eds.) (1975) *Vitamins and Hormones*, Vol. 33, Academic Press, New York.
Rasmusson, G.H., Torrey, J.H. (1994) Therapeutic control of androgen action, *Annu. Rep. Med. Chem.* **29**, 225.
Rogozkin, V.A. (1991) *Metabolism of Anabolic Androgenic Steroids*, CRC Press, Boca Raton.
Santen, R.J., Harvey, H.A. (1999) Use of aromatase inhibitors in breast carcinoma, *Endocr. Relat. Cancer* **6**, 75–92.
Singh, S.M., Gauthier, S., Labrie, F. (2000) Androgen receptor antagonists (antiandrogens): structure-activity relationships, *Curr. Med. Chem.* **7**, 211–247.
Snyder, P.J. (2000) Androgens, in: Hardman, J.G., Limbird, L.E. (Eds.) *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, 10th edition, McGraw-Hill Medical Publishing, New York, pp. 1635–1648.
Vida, J.A. (1969) *Androgens and Anabolic Agents. Chemistry and Pharmacology*, Academic Press, New York.
Wilson, J.D., Griffin, J.E., Russell, D.W. (1993) Steroid 5 alpha-reductase 2 deficiency, *Endocr. Rev.* **14**, 577–593.
Yanase, T., Simpson, E.R., Waterman, M.R. (1991) 17 alpha-hydroxylase/17,20-lyase deficiency: from clinical investigation to molecular definition, *Endocr. Rev.* **12**, 91–108.
Zeelen, F.J. (1990) *Medicinal Chemistry of Steroids*, Elsevier, Amsterdam.
Zhi, L., Martinborough, E. (2001) Selective androgen receptor modulators (SARMs), *Annu. Rep. Med. Chem.* **36**, 169.

Primary Literature

Abe, O., Herraneu, H., Dorfman, R.I. (1962) *Proc. Soc. Exp. Biol. Med.* **111**, 706.
Adachi, K., Kano, M. (1972) *Steroids* **19**, 567.
Akhtar, M., Skinner, S.J.M. (1968) *Biochem. J.* **109**, 318.
Akhtar, M., Njar, V.C.O., Wright, J.N. (1993) *J. Steroid Biochem. Mol. Biol.* **44**, 375.
Akhtar, M., Calder, M.R., Corina, D.L., Wright, J.N. (1981) *J. Chem. Soc., Chem. Commun.* 129.
Akhtar, M., Calder, M.R., Corina, D.L., Wright, J.N. (1982) *Biochem. J.* **201**, 569.
Albanese, A.A., Lorenze, E.J., Orto, L.A. (1963) *N. Y. State J. Med.* **63**, 80.
Alibrandi, A., Bruni, G., Ercoli, A., Gardi, R., Meli, A. (1960) *Endocrinology* **66**, 13.
Ammedick, U., Konrad, R.M., Gotzen, E. (1968) *Med. Ernähr.* **9**, 121.
Anderson, I.A. (1962) *Acta Endocrinol. (Copenh.)* **63**(Suppl.), 54.
Anderson, K.M., Liao, S. (1968) Selective retention of dihydrotestosterone by prostaticnuclei, *Nature (London)* **219**, 277.
Andersson, S., Bishop, R.W., Russell, D.W. (1989) *J. Biol. Chem.* **264**, 16249.
Anderson, S., Berman, D.M., Jenkins, E.P., Russell, D.W. (1991) *Nature* **354**, 150.
Angelastro, M.R., Laughlin, M.E., Schatzman, G.L., Bey, P., Blohm, T.R. (1989) *Biochem. Biophys. Res. Commun.* **162**, 1571.
Ariel, G. (1973) *J. Sports Med. Phys. Fitness* **13**, 187.
Ariel, G., Saville, W. (1972) *J. Appl. Physiol.* **32**, 795.
Arigoni, D., Battaglia, R., Akhtar, M., Smith, T. (1975) *J. Chem. Soc., Chem. Commun.* 185.
Arnold, A., Potts, G.O., Beyler, A.L. (1963a) *Endocrinology* **72**, 408.
Arnold, A., Potts, G.O., Beyler, A.L. (1963b) *J. Endocrinol.* **28**, 87.
Audet, P., Nurcombe, H., Lamb, Y., et al. (1993) *Clin. Pharmacol. Ther.* **53**, 231.
Ayub, M., Levell, M.J. (1989) *J. Steroid Biochem.* **32**, 515.
Azadian-Boulanger, G., Bonne, C., Sechi, J., Raynaud, J.P. (1974) *J. Pharmacol. (Paris)* **5**, 509.
Backle, R.M. (1959) *Br. Med. J.* **1**, 1378.
Bakshi, R.K., Patel, G.F., Rasmusson, G.H., Baginsky, W.F., Cimis, G., Ellsworth, K., Chang, B., Bull, H., Tolman, R.L., Harris, G.S. (1994) *J. Med. Chem.* **37**, 3871–3874.
Baldratti, G., Arcari, G., Clini, V., Tani, F., Sala, G. (1959) *Sperimentale* **109**, 383.
Baran, J.S. (1963) *J. Med. Chem.* **6**, 329.
Bardin, C.W., Bullock, L.P., Sherins, R.J. (1973) *Recent Prog. Horm. Res.* **29**, 65.
Barnes, L.E., Stafford, R.O., Guild, M.E., Thole, L.C., Olson, K.J. (1954) *Endocrinology* **55**, 77.
Barrie, S.E., Rowlands, M.G., Foster, A.B., Jarman, M. (1989) *J. Steroid Biochem.* **33**, 1191.
Barthe, P.L., Bullock, L.P., Mowszowicz, I. (1974) *Endocrinology* **95**, 1019.
Baulieu, E.E., Jung, I. (1970) *Biochem. Biophys. Res. Commun.* **38**, 599.
Beato, M. (1989) *Cell* **56**, 335.

Becker, H., Kaufmann, J., Klosterhalfen, H., Voigt, K.D. (1972) *Acta Endocrinol. (Copenh.)* **71**, 589.
Bednarski, P.J., Nelson, S.D. (1989) *J. Med. Chem.* **32**, 203–213.
Bednarski, P.J., Porubek, D.J., Nelson, S.D. (1985) *J. Med. Chem.* **28**, 775–779.
Bellino, F.L., Gilani, S.S.H., Eng, S.S., Osawa, Y., Duax, W.L. (1976) *Biochemistry* **15**, 4730.
Berkowitz, D. (1960) *Clin. Res.* **8**, 199.
Berthold, A.A. (1849) Transplantation der Hoden, *Arch. Anat. Physiol. Wiss. Med.* **16**, 42.
Beyler, A.L., Potts, G.O., Arnold, A. (1961) *Endocrinology* **68**, 987.
Bhargava, A.S., Seeger, A., Günzel, P. (1977) *Steroids* **30**, 407.
Bhasin, S., Storer, T.W., Berman, N., Callegari, C., Clevenger, B., Phillips, J., Bunnell, T.J., Tricker, R., Shirazi, A., Casaburi, R. (1996) *N. Engl. J. Med.* **335**, 1–7.
Blaquier, J.A., Calandra, R.S. (1973) *Endocrinology* **93**, 51.
Blumbert, A., Keller, H. (1971) *Schweiz. Med. Wochenschr.* **101**, 1887.
Bonne, C., Raynaud, J.P. (1974) *Mol. Cell Endocrinol.* **2**, 59.
Boris, A., Uskokovic, M. (1970) *Experientia* **26**, 9.
Boris, A., DeMartino, L., Trmal, T. (1971) *Endocrinology* **88**, 1086.
Boris, A., Scott, J.W., DeMartino, L., Cox, D.C. (1973) *Acta Endocrinol. (Copenh.)* **72**, 604.
Borrevang, P. (1962) *Acta Chem. Scand.* **16**, 883.
Bowers, A., Ringold, H.J., Denot, E. (1958) *J. Am. Chem. Soc.* **80**, 6115.
Bowers, A., Cross, A.D., Edwards, J.A., Carpio, H., Calzada, M.C., Denot, E. (1963) *J. Med. Chem.* **6**, 156.
Bracci, U. (1973) *J. Urol. Nephrol.* **79**, 405.
Bracci, U., DiSilverio, F. (1973) *Prog. Med.* **29**, 779.
Bradshaw, K.D., Waterman, M.R., Couch, R.T., Simpson, E.R., Zuber, M.X. (1987) *Mol. Endocrinol.* **1**, 348.
Branda, R.F., Amsden, T.W., Jacob, H.S. (1974) *Clin. Res.* **22**, 607A.
Brinkmann, A.O., Trapman, J. (1992) *Cancer Surv.* **14**, 95.
Brodie, H.J., Possanza, G., Townsley, J.D. (1968) *Biochim. Biophys. Acta* **152**, 770.
Brodie, A.M.H., Schwarzel, W.C., Shaikh, A.A., Brodie, H.J. (1977) *Endocrinology* **100**, 1684–1695.
Brodie, A.M.H., Garrett, W., Hendrickson, J.R., Tsai-Morris, C.-H. (1982) *Cancer Res.* **42**(Suppl.), 3360s–3364s.
Brodie, A.M.H., Garrett, W., Hendrickson, J.R., Tsai-Morris, C.-H., Marcotte, C.H., Robinson, C.H. (1981) *Steroids* **38**, 693–702.
Brooks, J.R., Busch, F.D., Patanelli, D.J., Steelman, S.L. (1973) *Proc. Soc. Exp. Biol. Med.* **143**, 647.
Brown, T. (1996) Androgen Receptor Structure, Function, Regulation, and Dysfunction, in: Bhasin, S., Gabelnick, H.L., Spieler, J.M., Swerdloff, R.S., Wang, C., Kelly, C. (Eds.) *Pharmacology, Biology, and Clinical Applications of Androgens*, Wiley-Liss, New York, p. 45.
Bruchovsky, N., Wilson, J.D. (1968a) *J. Biol. Chem.* **243**, 5953.
Bruchovsky, N., Wilson, J.D. (1968b) *J. Biol. Chem.* **243**, 2012.
Bruchovsky, N., Meakin, J.W. (1973) *Cancer Res.* **33**, 1689.
Bruchovsky, N., Sutherland, D.J.A., Meakin, J.W., Minesita, T. (1975) *Biochim. Biophys. Acta* **381**, 61.
Brueggemeier, R.W., Katlic, N.E. (1987) *Cancer Res.* **47**, 4548–4551.
Brueggemeier, R.W., Li, P.-K. (1988) *Cancer Res.* **48**, 6808–6810.
Brueggemeier, R.W., Floyd, E.E., Counsell, R.E. (1978) *J. Med. Chem.* **21**, 1007–1011.
Brueggemeier, R.W., Li, P.-K., Snider, C.E., Darby, M.V., Katlic, N.E. (1987) *Steroids* **50**, 163–178.
Buchner, H. (1961) *Wien. Med. Wochenschr.* **111**, 576.
Bull, H.G. (1996) *J. Am. Chem. Soc.* **118**, 2359.
Bulun, E.B., Simpson, E.R. (1994) *Breast Cancer Res. Treat* **30**, 19–29.
Burnett, P.C. (1963) *J. Am. Geriatr. Soc.* **11**, 979.
Burton, J.L., Laschet, U., Shuster, S. (1973) *Br. J. Dermatol.* **89**, 487.
Butenandt, A. (1931) *Z. Angew. Chem.* **44**, 905.
Butenandt, A., Tscherning, K. (1934) *Z. Physiol. Chem.* **229**, 167.
Butenandt, A., Hanisch, G. (1935) *Berichte* **68**, 1859; (1935) *Z. Physiol. Chem.* **237**, 89.
Campbell, J.A., Babcock, J.C. (1959) *J. Am. Chem. Soc.* **81**, 4069.
Campbell, J.A., Lyster, S.C., Duncan, G.W., Babcock, J.C. (1963) *Steroids* **1**, 317.
Canick, J.A., Vaccaro, D.E., Ryan, K.J., Leeman, S.E. (1977) *Endocrinology* **100**, 250.
Cardinali, D.P., Nagle, C.A., Rosner, J.M. (1974) *Endocrinology* **95**, 179.

Carson-Jurica, M.A., Schrader, W.T., O'Malley, B. (1990) *Endocr. Rev.* **11**, 201.
Casner, S., Early, R., Carlson, B.R. (1971) *J. Sports Med. Phys. Fitness* **11**, 98.
Caspi, E., Vermeulen, A., Bhat, H.B. (1965) *Steroids* **58**(Suppl. 1), 141.
Caspi, E., Wicha, J., Aninachalam, T., Nelson, P., Spitiller, G. (1984) *J. Am. Chem. Soc.* **106**, 7282.
Castegnaro, E., Sala, G. (1961) *Folia Endocrinol.* **14**, 581.
Catlin, D.H., Sekera, M.H., Ahrens, B.D., Starcevic, B., Chang, Y.C., Hatton, C.K. (2004) *Rapid Commun. Mass Spectrom.* **18**, 1245.
Chan, L., O'Malley, B.W. (1976) *N. Engl. J. Med.* **294**, 1322, 1372, 1430.
Chen, S., Beshman, M.J., Sparkes, R.S., Zollman, S., Klisak, I., Mohandes, T., Hall, P.F., Shively, J.E. (1988) *DNA* **7**, 27.
Christiansen, R.G., Bell, M.R., D'Ambra, T.E., Mallamo, J.P., Herrmann, J.L., Ackerman, J.H., Opalka, C.J., Kullnig, R.K., Winneker, R.C., Snyder, B.W., et al. (1990) *J. Med. Chem.* **33**, 2094–2100.
Cleutjens, K.B., van Eekelen, C.C., van der Korput, H.A., Brinkmann, A.O., Trapman, J. (1996) *J. Biol. Chem.* **271**, 6379–6388.
Cleutjens, K.B., van der Korput, H.A., van Eekelen, C.C., van Rooij, H.C., Faber, P.W., Trapman, J. (1997) *Mol. Endocrinol.* **11**, 148–161.
Cole, P.A., Robinson, C.H. (1988) *J. Am. Chem. Soc.* **110**, 1284.
Colton, F.B. (1959) US Patent 2,874,170.
Colton, F.B., Ray, R.E. (1962) US Patent 3,068,249.
Colton, F.B., Nysted, L.N., Reigel, B., Raymond, A.L. (1957) *J. Am. Chem. Soc.* **79**, 1123.
Cooke, G.M., Robaire, B. (1986) *J. Steroid Biochem.* **24**, 877.
Coombes, R.C., Goss, P., Dowsett, M., Gazet, J.C., Brodie, A.M.H. (1984) *Lancet* **2**, 1237–1239.
Corbin, C.J., Grahan-Lorence, S., McPhaul, M., Mason, J.I., Mendelson, C.R., Simpson, E.R. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8948.
Corvol, P., Michaud, A., Menard, J., Freifeld, M., Mahoudean, J. (1975) *Endocrinology* **97**, 52.
Counsell, R.E. (1966) *J. Med. Chem.* **9**, 263.
Counsell, R.E., Klimstra, P.D. (1963) *J. Med. Chem.* **6**, 736.
Counsell, R.E., Klimstra, P.D., Colton, F.B. (1962) *J. Org. Chem.* **27**, 248.
Covey, D.F., Hood, W.F., Parikh, V.D. (1981) *J. Biol. Chem.* **256**, 1076–1079.
Crawford, E.D., Eisenberger, M.A., McLeod, D.G., Spaulding, J.T., Benson, R., Dorr, F.A., Blumenstein, B.A., Davis, M.A., Goodman, P.J. (1989) *N. Engl. J. Med.* **321**, 419–424.
Cross, A.D., Harrison, I.T. (1965) *Steroids* **6**, 397.
Cross, A.D., Harrison, I.T., Crabbe, P., Kincl, F.A., Dorfman, R.I. (1964) *Steroids* **4**, 229.
Dahbberg, E. (1982) *Biochim. Biophys. Acta* **717**, 65.
Darby, M.V., Lovett, J.A., Brueggemeier, R.W., Groziak, M.P., Counsell, R.E. (1985) *J. Med. Chem.* **28**, 803–807.
David, K., Dingemanse, E., Freud, J., Laqueur, E. (1935) *Hoppe Seyler's Z. Physiol. Chem.* **233**, 281.
Davison, C., Banks, W., Fritz, A. (1976) *Arch. Int. Pharmacodyn. Ther.* **221**, 294.
Dawson, L.A., Chow, E., Morton, G. (1997) *Urology* **49**, 283–284.
Death, A.K., McGrath, K.C., Kazlauskas, R., Handelsman, D.J. (2004) *J. Clin. Endocrinol. Metab.* **89**, 2498.
Dell'acqua, S., Manuso, S., Eriksson, G., Diczfalusy, E. (1966) *Biochim. Biophys. Acta* **130**, 241.
de Lorimier, A.A., Gordan, G.S., Lowe, R.C., Carbone, J.V. (1965) *Arch. Intern. Med.* **116**, 289.
DeMoor, P., Steeno, O., Heyns, W., Van Baelen, H. (1969) *Ann. Endocrinol.* **30**, 233.
Denis, L., Murphy, G.P. (1993) *Cancer* **72**, 3888–3895.
Desaulles, P.A. (1960) *Helv. Med. Acta* **27**, 479.
de Visser, J., Overbeck, G.A. (1960) *Acta Endocrinol. (Copenh.)* **35**, 405.
Dingman, J.F., Jenkins, W.H. (1962) *Metabolism* **11**, 273.
diSalle, E., Guidici, D., Briatico, G., et al. (1993) *J. Steroid Biochem. Mol. Biol.* **46**, 549.
DiSilverio, F., Gagliardi, V. (1968) *Boll. Soc. Urol.* **5**, 198.
Djerassi, C., Riniker, R., Riniker, B. (1956) *J. Am. Chem. Soc.* **78**, 6377.
Djerassi, C., Miramontes, L., Rosenkranz, G., Sondheimer, F. (1954) *J. Am. Chem. Soc.* **76**, 4092.
Dodds, E.C. (1945) *Biochem. J.* **39**, i.
Dorfman, R.I., Dorfman, A.S. (1963) *Acta Endocrinol. (Copenh.)* **42**, 245.
Dorfman, R.I., Kincl, F.A. (1963) *Endocrinology* **72**, 259.

Dorfman, R.I., Dorfman, A.S., Gut, M. (1962) *Acta Endocrinol. (Copenh.)* **40**, 565.
Dowsett, M., Cunningham, D., Nichols, S., Lal, A., Evans, S., Dehennin, L., Hedley, A., Coombes, R.C. (1989) *Cancer Res.* **49**, 1306–1312.
Doyle, A.E., Pinkus, N.B., Green, J. (1974) *Med. J. Aust.* **1**, 127.
Drago, J.R., Santen, R.J., Lipton, A., Worgul, T.J., Harvey, H.A., Boucher, A., Manni, A., Rohner, T.J. (1984) *Cancer* **53**, 1447.
Dray, F., Mowszowicz, I., Ledru, M.J., Crépy, O., Delzant, G., Sebaoun, J. (1969) *Ann. Endocrinol.* **30**, 223.
Drill, V.A., Riegel, B. (1958) *Recent Prog. Horm. Res.* **14**, 29.
Duax, W.L., Erman, M.G., Griffin, J.F., Wolff, M.E. (1976) *Cryst. Struct. Commun.* **5**, 775.
Dunn, J.F., Goldstein, J.L., Wilson, J.D. (1973) *J. Biol. Chem.* **248**, 7819.
Eadon, G., Djerassi, C. (1972) Synthesis and biological activity of D-bishomo steroids, *J. Med. Chem.* **15**, 89.
Edgren, R.A. (1963) *Acta Endocrinol. (Copenh.)* **87**(Suppl.), 3.
Edgren, R.A., Smith, H., Hughes, G.A. (1963) *Steroids* **2**, 731.
Edgren, R.A., Peterson, D.L., Jones, R.C., Nagra, C.L., Smith, H., Hughes, G.A. (1966) *Recent Prog. Horm. Res.* **22**, 305.
Edwards, J.A., Bowers, A. (1961) *Chem. Ind. (London)* 1962.
Eik-Nes, K.B. (1975) *Vitam. Horm.* **33**, 193.
Eisenberg, E., Gordan, G.S. (1950) *J. Pharmacol. Exp. Ther.* **99**, 38.
El Attar, T., Dirscherl, W., Mosebach, K.O. (1964) *Acta Endocrinol. (Copenh.)* **45**, 527.
Emmens, C.W. (1939) *Med. Res. Counc. Spec. Rep. Ser.* **234**, 1.
Engel, L.L., Alexander, J., Wheeler, M. (1958) *J. Biol. Chem.* **231**, 159.
Engelhardt, D., Eisenburg, J., Unterberger, P., Karl, H.J. (1971) *Klin. Wochenschr.* **49**, 439.
Ercoli, A., Gardi, R., Vitali, R. (1962) *Chem. Ind. (London)* 1284.
Ercoli, A., Bruni, G., Falconi, G., Galletti, F., Gardi, R. (1960) *Acta Endocrinol. (Copenh.)* **51**(Suppl.), 857.
Eri, L.M., Tveter, K.J. (1993) *J. Urol.* **150**, 90.
Evans, R.M. (1988) *Science* **240**, 889–895.
Evans, C.R., Pierrepoint, C.G. (1975) *J. Endocrinol.* **64**, 539.
Evans, C.T., Merrill, J.C., Corbin, C.J., Saunders, D., Simpson, E.R., Mendelson, C.R. (1987) *J. Biol. Chem.* **269**, 6914.
Fahey, T.D., Brown, C.H. (1973) *Med. Sci. Sports* **5**, 272.
Fang, S., Liao, S. (1969) *Mol. Pharmacol.* **5**, 420.
Fang, S., Liao, S. (1971) *J. Biol. Chem.* **246**, 16.
Fang, S., Anderson, K.M., Liao, S. (1969) *J. Biol. Chem.* **244**, 6584.
Farnsworth, W.E. (1966) *Steroids* **8**, 825.
Feldman, E.B., Carter, A.C. (1960) *J. Clin. Endocrinol. Metab.* **20**, 842.
Ferrin, J. (1956) *Acta Endocrinol. (Copenh.)* **22**, 303.
Fixson, U. (1963) *Geburtshilfe Frauenheilkd* **23**, 371.
Flores, F., Naftolin, F., Ryan, K.J. (1973) *Neuroendocrinology* **11**, 177.
Flores, F., Naftolin, F., Ryan, K.J., White, R.J. (1973) *Science* **180**, 1074.
Fox, T.O. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4303.
Fox, M., Minot, A.S., Liddle, G. (1962) *J. Clin. Endocrinol. Metab.* **22**, 921.
Frederiksen, D.W., Wilson, J.D. (1971) *J. Biol. Chem.* **246**, 2584.
French, F.S., McLean, W.S., Smith, A.A., Tindall, D.J., Weddington,S.C., Petrusz, P., Nayfeh, S.N., Ritzen, E.M., Hansson, V., Trystad, O. (1974) *Nature (London)* **250**, 387.
Fried, W., Jonasson, O., Lang, G., Schwartz, F. (1973) *Ann. Intern. Med.* **79**, 823.
Frölich, M., Vader, H.L., Walma, S.T., De Rooy, H.A.M. (1980) *J. Steroid Biochem.* **13**, 1097.
Frye, S.V., Haffner, C.D., Maloney, P.R., Mook, R.A. Jr., Dorsey, G.F. Jr., Hiner, R.N., Batchelor, K.W., Bramson, H.N., Stuart, J.D., Schweiker, S.L., et al. (1993) *J. Med. Chem.* **36**, 4313.
Fuhrmann, U., Bengston, C., Repenthin, G., Schillinger, E. (1992) *J. Steroid Biochem. Mol. Biol.* **42**, 787.
Fujimuri, M. (1973) *Cancer* **31**, 789.
Gabbard, R.B., Segaloff, A. (1962) *J. Org. Chem.* **27**, 655.
Gallagher, T.F., Koch, F.C. (1935) *J. Pharmacol. Exp. Ther.* **55**, 97.
Gao, W., Kearbey, J.D., Nair, V.A., Chung, K., Parlow, A.F., Miller, D.D., Dalton, J.T. (2004) *Endocrinology* **145**, 5420.
Gaunt, R., Tuthill, C.H., Antonchak, N., Leathem, J.H. (1953) *Endocrinology* **52**, 407.
Geldof, A.A., Meulenbroek, M.F.A., Dijkstra, I., et al. (1992) *J. Cancer Res. Clin. Oncol.* **118**, 50.

Geller, J., Fruchtman, B., Meyer, C., Newman, H. (1967) *J. Clin. Endocrinol. Metab.* **27**, 556.
Geller, J., Fruchtman, B., Newman, H., Roberts, T., Sylva, R. (1967) *Cancer Chemother. Rep.* **51**, 441.
Gerhards, E., Kolb, K.H., Schulze, P.E. (1965) *Z. Physiol. Chem.* **342**, 40.
Gerhards, E., Gibian, H., Kolb, K.H. (1966) *Arzneim.-Forsch.* **16**, 458.
Giudici, D., Ornati, G., Briatico, G., Buzzetti, F., Lombardi, P., Di Salle, E. (1988) *J. Steroid Biochem.* **30**, 391–394.
Glas, W.W., Lansing, E.H. (1962) *J. Am. Geriatr. Soc.* **10**, 509.
Glovna, R.W., Wilson, J.D. (1969) *J. Clin. Endocrinol. Metab.* **29**, 970.
Glueck, C.J. (1971) *Clin. Res.* **17**, 475.
Glueck, C.J. (1971) *Metabolism* **20**, 691.
Glueck, C.J., Ford, S., Steiner, P., Fallat, R. (1973) *Metabolism* **17**, 807.
Goldberg, I.W., Sicé, J., Robert, H., Plattner, Pl.A. (1947) *Helv. Chim. Acta* **30**, 1441.
Goldenberg, I.S., Waters, N., Randin, R.S., Ansfield, F.J., Segaloff, A. (1973) *J. Am. Med. Assoc.* **223**, 1267.
Golding, L.A., Freydinger, J.E., Fishel, S.S. (1974) *Phys. Sportsmed.* **2**, 39.
Goldman, J.N., Epstein, J.A., Kupperman, H.S. (1957) *Endocrinology* **61**, 166.
Gordan, G.S. (1957) *Arch. Intern. Med.* **100**, 744.
Gordan, G.S., Wessler, S., Avioli, L.V. (1972) *J. Am. Med. Assoc.* **219**, 483.
Gorski, J., Welshons, W.V., Sakai, D., Hansen, J., Walent, J., Kassis, J., Shull, J., Stack, G., Campen, C. (1986) *Recent Prog. Horm. Res.* **42**, 297.
Goss, P.E., Powles, T.J., Dowsett, M., Hutchison, G., Brodie, A.M.H., Gazet, J.-C., Coombes, R.C. (1986) *Cancer Res.* **46**, 4823–4826.
Goto, G., Yoshioka, K., Hiraga, K., Masouka, M., Nakayama, R., Miki, T. (1978) *Chem. Pharm. Bull. (Tokyo)* **26**, 1718.
Greenblatt, R.B., Jungck, E.C., King, G.C. (1964) *Am. J. Med. Sci.* **318**, 99.
Greenwood, A.W., Blyth, J.S.S., Callow, R.K. (1935) *Biochem. J.* **29**, 1400.
Griffin, J.E. (1979) *J. Clin. Invest.* **64**, 1624.
Griffin, J.E., Wilson, J.D. (1980) *N. Engl. J. Med.* **302**, 198.
Griffin, J.E., Punyashthiti, K., Wilson, J.D. (1976) *J. Clin. Invest.* **57**, 1342.
Gu, Y., Lin, Y.C. (1991) *Res. Comm. Chem. Pathol. Pharmacol.* **72**, 27–38.
Gu, Y., Lin, Y.C., Rikihisa, Y. (1990) *Biochem. Biophys. Res. Commun.* **169**, 455–446.
Gu, Y., LI, P.-K., Lin, Y.C., Rikihisa, Y., Brueggemeier, R.W. (1991) *J. Steroid Biochem. Mol. Biol.* **38**, 709–715.
Gual, C., Morato, T., Hayano, M., Gut, M., Dorfman, R.I. (1962) *Endocrinology* **71**, 920.
Guess, H.A., Heyse, J.F., Gormley, G.J. (1993) *Prostate* **22**, 31.
Gustaffson, J.-A., Carlstedt-Duke, J., Poellinger, L., Okret, S., Wikstrom, A.C., Bronnegard, M., Gillner, M., Dong, Y., Fuxe, K., Cintra, A. (1987) *Endocr. Rev.* **8**, 185.
Hackett, J.C., Brueggemeier, R.W., Haddad, C.M. (2005) *J. Am. Chem. Soc.* **127**, 4063–4070.
Hadley, M.A., Lin, Y.C., Dym, M. (1981) *J. Androl.* **2**, 190–199.
Halden, A., Watter, R.M., Gordan, G.S. (1970) *Cancer Chemother. Rep.* **54**, 453.
Hamann, L.G., Higuchi, R.I., Zhi, L., Edwards, J.P., Wang, X.N., Marschke, K.B., Kong, J.W., Farmer, L.J., Jones, T.K. (1998) *J. Med. Chem.* **41**, 623–639.
Hammerstein, J., Cupceancu, B. (1969) *Dtsch. Med. Wochenschr.* **94**, 829.
Hammerstein, J., Meckies, J., Leo-Rossberg, I., Moltz, L., Zielke, F. (1975) *J. Steroid Biochem.* **6**, 827.
Hansson, V., et al. (1974) *Steroids* **23**, 823.
Hansson, V., Tveter, K.J., Unhjem, O., Djoseland, O. (1972) *J. Steroid Biochem.* **3**, 427.
Hansson, V., Trystad, O., French, F.S., McLean, W.S., Smith, A.A., Tindall, D.J., Weddington, S.C., Petrusz, P., Nayfeh, S.N., Ritzen, E.M. (1974) *Nature (London)* **250**, 387.
Harding, B.W., Samuels, L.T. (1962) *Endocrinology* **70**, 109.
Harris, I.H. (1961) *J. Clin. Endocrinol. Metab.* **21**, 1099.
Harris, G., Azzolina, B., Baginsky, W., et al. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10787.
Hartiala, K. (1973) *Physiol. Rev.* **53**, 496.
Hellman, L., Bradlow, H.L., Zumoff, B., Fukushima, D.K., Gallagher, T.F. (1959) *J. Clin. Endocrinol.* **19**, 936.
Henderson, E., Weinberg, M. (1951) *J. Clin. Endocrinol.* **11**, 641.
Henderson, J.T., Richmond, J., Sumerling, M.D. (1973) *Lancet* **1**, 934.
Henderson, D., Norbisrath, G., Kerb, U. (1986) *J. Steroid Biochem.* **24**, 303–306.
Hendler, E.D., Goffinet, J.A., Ross, S., Longnecker, R.E., Bakovic, V. (1974) *N. Engl. J. Med.* **291**, 1046.

Hershberger, L.G., Shipley, E.G., Meyer, R.K. (1953) *Proc. Soc. Exp. Biol. Med.* **83**, 175.
Heusser, H., Herzig, P.T., Furst, A., Plattner, Pl.A. (1950) *Helv. Chim. Acta* **33**, 1093.
Hickson, R., Galessi, T., Kurowski, T., Daniels, D., Chatterton, R. (1983) *J. Steroid Biochem.* **19**, 1705.
Hiraga, K., Tsunehiko, A., Takuichi, M. (1965) *Chem. Pharm. Bull. (Tokyo)* **13**, 1294.
Holt, D.A., Levy, M.A., et al. (1990) *J. Med. Chem.* **33**, 943; (1990) *J. Med. Chem.* **33**, 937.
Holt, D.A., Yamashita, D.S., Konialian-Beck, A.L., Luengo, J.I., Abell, A.D., Bergsma, D.J., Brandt, M., Levy, M.A. (1995) *J. Med. Chem.* **38**, 13–15.
Horst, H.J., Dennis, M., Kaufmann, J., Voigt, K.D. (1975) *Acta Endocrinol. (Copenh.)* **79**, 394.
Hortling, H., Malmio, K., Husi-Brummer, L. (1962) *Acta Endocrinol. (Copenh.)* **39**(Suppl.), 132.
Horton, R., Kato, T., Sherino, R. (1967) *Steroids* **10**, 245.
Howard, R.P., Furman, R.H. (1962) *J. Clin. Endocrinol.* **22**, 43.
Huggins, C., Hodges, C.V. (1941) *Cancer Res.* **1**, 293.
Huggins, C., Jensen, E.V. (1954) *J. Exp. Med.* **100**, 241.
Huggins, C., Stevens, R.E. Jr., Hodges, C.V. (1941) *Arch. Surg.* **43**, 209.
Huggins, C., Jensen, E.V., Cleveland, A.S. (1954) *J. Exp. Med.* **100**, 225.
Hughes, D.W. (1973) *Med. J. Aust.* **2**, 361.
Hümpel, M., Wendt, H., Schulze, P.E., Dogs, G., Weiss, C., Speck, U. (1977) *Contraception* **15**, 579.
Huseman, C., Johanson, A. (1975) *J. Pediatr.* **87**, 946.
Huttenrauch, R. (1964) *Arch. Pharm.* **297**, 124.
Imperato-McGinley, J., Guerrero, L., Gautier, T., Peterson, R.E. (1974) *Science* **186**, 1213.
Irwin, J.F., Lane, S.E., Neuhaus, O.W. (1971) *Biochim. Biophys. Acta* **252**, 328.
Ishii-Ohba, H., Inano, H., Tamaoki, B.-I. (1987) *J. Steroid Biochem.* **27**, 775.
Ismail, A.A., Davidson, D.W., Souka, A.R., Barnes, E.W., Irvine, W.J., Kilimnik, H., Vanderbeeken, Y. (1974) *J. Clin. Endrocrinol. Metab.* **39**, 81.
Jaffe, R.B. (1969) *Steroids* **14**, 483.
Jagarinec, N., Givner, M.L. (1974) *Steroids* **23**, 561.
James, K.C. (1972) *Experientia* **28**, 479.
James, K.C., Nicholl, P.J., Richards, G.T. (1975) *Eur. J. Med. Chem.* **10**, 55.
Janknegt, R.A., Abbou, C.C. Bartoletti, R., Bernstein-Hahn, L., Bracken, B., Brisset, J.M., Da Silva, F.C., Chisholm, G., Crawford, E.D., Debruyne, F.M., et al. (1993) *J. Urol.* **149**, 77–82.
Jensen, E.J., Jacobson, H.I. (1962) *Recent Prog. Horm. Res.* **18**, 387.
Johns, W.F. (1958) *J. Am. Chem. Soc.* **80**, 6456.
Johnson, L.C., Fisher, G., Silvester, L.J., Hofheins, C.C. (1972) *Med. Sci. Sports* **4**, 43.
Johnson, F.L., Feagler, J.R., Lerner, K.G., Majerus, P.W., Siegel, M., Hartman, J.R., Thomas, E.D. (1972) *Lancet* **2**, 1273.
Johnsonbaugh, R.E., Cohen, B.R., McCormick, E.M., et al. (1993) *J. Urol.* **149**, 432.
Johnston, J.O. (1987) *Steroids* **50**, 106–120.
Johnston, J.O., Wright, C.L., Metcalf, B.W. (1984a) *Endocrinology* **115**, 776–785.
Johnston, J.O., Wright, C.L., Metcalf, B.W. (1984b) *J. Steroid Biochem.* **20**, 1221–1226.
Jones, C.D., Audia, J.E., Lawhorn, D.E., McQuaid, L.A., Neubauer, B.L., Pike, A.J., Pennington, P.A., Stamm, N.B., Toomey, R.E., Hirsch, K.S. (1993) *J. Med. Chem.* **36**, 421–423.
Jouan, P., Samperez, S., Thieulant, M.L. (1973) *J. Steroid Biochem.* **4**, 65.
Jouan, P., Samperez, S., Thieulant, M.L., Mercier, L. (1971) *J. Steroid Biochem.* **2**, 223.
Jung, I., Baulieu, E.E. (1972) *Nat. New Biol.* **237**, 24.
Junkmann, K. (1957) *Recent Prog. Horm. Res.* **13**, 389.
Junkmann, K., Suchowsky, G. (1962) *Arzneim.-Forsch.* **12**, 214.
Junkmann, K., Neumann, F. (1964) *Acta Endocrinol. (Copenh.)* **45**(Suppl. 90), 139.
Kalliomaki, J.L., Pirila, A.M., Ruikka, I. (1962) *Acta Endocrinol. (Copenh.)* **63**(Suppl.), 124.
Kaneko, H., Nakamura, K., Yamato, Y., Kurakawa, M. (1969) *Chem. Pharm. Bull. (Tokyo)* **17**, 11.
Kappas, A., Palmer, R.H. (1965) in: Dorfman, R.I. (Ed.) *Methods in Hormone Research*, Vol. 4, Part B, Academic Press, New York, p. 1.
Karr, J.P., Kirdani, R.Y., Murphy, G.P., Sandberg, A.A. (1974) *Life Sci.* **15**, 501.
Katchen, B., Buxbaum, S. (1975) *J. Clin. Endocrinol. Metab.* **41**, 373.
Kato, J. (1975) *J. Steroid Biochem.* **6**, 979.
Kato, T., Horton, R. (1968) *J. Clin. Endocrinol. Metab.* **28**, 1160.

Kaufman, M., Straisfeld, C., Pinsky, L. (1976) *J. Clin. Invest.* **58**, 345.
Kaupp, H.A., Preston, F.W. (1962) *J. Am. Med. Assoc.* **180**, 411.
Kellis, J.K., Vickery, L.E. (1987) *J. Biol. Chem.* **262**, 4413.
Kennedy, B.J. (1958) *N. Engl. J. Med.* **259**, 673.
Kenny, B., Ballard, S., Blagg, J., Fox, D. (1997) *J. Med. Chem.* **40**, 1293–1315.
Keyssner, J., Hauswaldt, C., Uhl, N., Hunstein, W. (1974) *Schweiz. Med. Wochenschr.* **104**, 1938.
Killander, A., lundmark, K., Sjolin, S. (1969) *Acta Paediatr. Scand.* **58**, 10.
Kincl, F.A., Dorfman, R.I. (1964) *Steroids* **3**, 109.
Kincl, F.A., Ringold, H.J., Dorfman, R.I. (1961) *Acta Endocrinol.* **36**, 83.
Klimstra, P.D., Nutting, E.F., Counsell, R.E. (1966) *J. Med. Chem.* **9**, 693.
Kochakian, C.D. (1946) *Am. J. Physiol.* **145**, 549.
Kochakian, C.D. (1952) *Proc. Soc. Exp. Biol. Med.* **80**, 386.
Kochakian, C.D. (1948) *Recent Prog. Horm. Res.* **1**, 177.
Kochakian, C.D., Murlin, J.R. (1935) The effect of male hormones on the protein and energy metabolism of castrate dogs, *J. Nutr.* **10**, 437.
Koga, M., Kasayama, S., Matsumoto, K., Sato, B. (1995) *J. Steroid Biochem. Mol. Biol.* **54**, 1–6.
Kojo, H., Nakayama, O., Hirosumi, J., Chida, N., Notsu, Y., Okuhara, M. (1995) *Mol. Pharmacol.* **48**, 401–406.
Kokontis, J.M., Liao, S. (1999) *Vitam. Horm.* **55**, 219–307.
Konrad, R.M., Ammedick, U., Hupfauer, W., Ringler, W. (1967) *Chirurg* **38**, 168.
Korst, D.R., Bowers, C.Y., Flokstra, J.H., McMahon, F.G. (1963) *Clin. Pharmacol. Ther.* **4**, 734.
Kraft, H.G., Kieser, H. (1964) *Arzneim.-Forsch.* **14**, 330.
Krieg, M., Voigt, K.D. (1976) *J. Steroid Biochem.* **7**, 1005.
Kruskemper, H.L. (1966) *Arzneim.-Forsch.* **16**, 608.
Kruskemper, H.L., Noell, G. (1966) *Steroids* **8**, 13.
Kruskemper, H.L., Moraner, K.D., Noell, G. (1967) *Arzneim.-Forsch.* **17**, 449.
Kuhn-Velten, W.N., Meyer, I., Staib, W. (1989) *J. Steroid Biochem.* **33**, 33.
Kuhn-Velten, W.N., Bunse, T., Forster, M.E.C. (1991) *J. Biol. Chem.* **266**, 6291.
Kumar, M., Roy, A.K., Axelrod, A.E. (1969) *Nature* **223**, 399.
Kundu, N., Sandberg, A.A., Slaunwhite, W.R. Jr. (1965) *Steroids* **6**, 543.
Kupfer, D., Forchielli, E., Dorfman, R.I. (1960) *J. Biol. Chem.* **235**, 1968.
Labrie, C., Cusan, L., Plante, M., Lapointe, S., Labrie, F. (1987) *J. Steroid Biochem.* **28**, 379–384.
Labrie, C., Simard, J., Zhao, H.F., Pelletier, G., Labrie, F. (1990) *Mol. Cell Endocrinol.* **68**, 169–179.
Langecker, H. (1962) *Arzneim.-Forsch.* **12**, 231.
Lawrence, S.U. (1981) *Am. Pharm.* **21**, 57.
Lebeau, M.C., Mercier-Bodard, C., Oldo, J., Bourguon, D., Brécy, T., Raynaud, J.P., Baulieu, E.E. (1969) *Ann. Endocrinol.* **30**, 183.
Lehnert, G., Mucke, W. (1966) *Arzneim.-Forsch.* **16**, 603.
Leibetseder, J., Steininger, K. (1965) *Arzneim.-Forsch.* **15**, 474.
Lennon, H.D., Saunders, F.J. (1964) *Steroids* **4**, 689.
Lephart, E.D. (1993) *Mol. Cell. Neurosci.* **4**, 473.
Lephart, E.D., Andersson, S., Simpson, E.R. (1990) *Endocrinology* **127**, 1121.
Lephart, E.D., Simpson, E.R., McPhaul, M.J., Kilgore, M.W., Wilson, J.D., Ojeda, S.R. (1992) Brain aromatase cytochrome P-450 messenger RNA levels and enzyme activity during prenatal development in the rat, *Mol. Brain Res.* **16**, 187.
Lerner, L.J. (1975) *Pharmacol. Ther. [B]* **1**, 217.
Lerner, L.J., Bianchi, A., Borman, A. (1960) *Proc. Soc. Exp. Biol. Med.* **103**, 172.
Levy, M.A., Brandt, M., Heys, J.R., et al. (1990) *Biochemistry* **29**, 2815.
Levy, M.A., Metcalf, B.W., Brandt, M., et al. (1991) *Bioorg. Chem.* **19**, 245.
Li, P.-K., Brueggemeier, R.W. (1990a) *J. Med. Chem.* **33**, 101–105.
Li, P.-K., Brueggemeier, R.W. (1990b) *J. Enzym. Inhib.* **4**, 113–120.
Liang, T., Mezzetti, G., Chen, C., Liao, S. (1978) *Biochim. Biophys. Acta* **542**, 430–441.
Liao, S. (1975) *Int. Rev. Cytol.* **41**, 87.
Liao, S., Howell, D.K., Chuag, T. (1974) *Endocrinology* **94**, 1205.
Liao, S., Liang, T., Fang, S., Casteneda, E., Shao, T. (1973) *J. Biol. Chem.* **248**, 6154.
Liddle, G.W., Burke, H.A. Jr. (1960) *Helv. Med. Acta* **27**, 504.
Lin, G.C., Erinoff, L. (Eds.) (1990) Anabolic Steroid Abuse *NIDA Research Monograph*, Vol. 102, National Institute on Drug Abuse, Washington, D.C., 29.

Lin, Y.C., Rikihisa, Y. (1987) *N. Y. Acad. Sci.* **513**, 532–534.
Lin, Y.C., Chitcharoenthum, M., Rikihisa, Y. (1987) *Contraception* **36**, 581–592.
Lin, Y.C., Hadley, M.A., Klingener, D., Dym, M. (1980) *Biol Reprod.* **22**, 95A.
Lin, Y.C., Fukaya, T., Rikihisa, Y., Walton, A. (1985) *Life Sci.* **37**, 39–47.
Longcope, C., Femino, A.M., Johnston, J.O. (1988) *Endocrinology* **122**, 2007–2011.
Lukas, S.E. (1993) *Trends Pharmacol. Sci.* **14**, 61.
Lyon, M.F., Hendry, I., Short, R.V. (1973) *Endocrinology* **58**, 357.
Lyster, S.C., Lund, G.H., Stafford, R.O. (1956) *Endocrinology* **58**, 781.
Mahendroo, M.S., Mendelson, C.R., Simpson, E.R. (1993) *J. Biol. Chem.* **268**, 19463.
Mainwaring, W.I.P. (1969a) A soluble androgen receptor in the cytoplasm of rat prostate. *J. Endocrinol.* **45**, 531.
Mainwaring, W.I.P. (1969b) *J. Endocrinol.* **44**, 323.
Mainwaring, W.I.P., Mangan, F.R. (1973) *J. Endocrinol.* **59**, 121.
Mallamo, J.P., Pilling, G.M., Wetzel, J.R., Kowalczyk, P.J., Bell, M.R., Kullnig, R.K., Batzold, F.H., Juniewicz, P.E., Winneker, R.C., Luss, H.R. (1992) *J. Med. Chem.* **35**, 1663–1670.
Mangan, F.R., Mainwaring, W.I.P. (1972) *Steroids* **20**, 331.
Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P. (1995) *Cell* **83**, 835–839.
Manson, A.J., Stonner, F.W., Neumann, H.C., Christiansen, R.G., Clarke, R.L., Ackerman, J.H., Page, D.F., Dean, J.W., Phillips, D.K., Potts, G.O., Arnold, A., Beyler, A.L., Clinton, R.O. (1963) *J. Med. Chem.* **6**, 1.
Marcotte, P.A., Robinson, C.H. (1982a) *Steroids* **39**, 325–344.
Marcotte, P.A., Robinson, C.H. (1982b) *Biochemistry* **21**, 2773–2778.
Marhefka, C.A., Gao, W., Chung, K., Kim, J., He, Y., Yin, D., Bohl, C., Dalton, J.T., Miller, D.D. (2004) *J. Med. Chem.* **47**, 993.
Marquardt, G.H., Fisher, C.I., Levy, P., Dowben, R.M. (1961) *J. Am. Med. Assoc.* **175**, 851.
Marsh, D.A., Brodie, H.J., Garrett, W., Tsai-Morris, C.-H., Brodie, A.M.H. (1985) *J. Med. Chem.* **28**, 788–795.
Massa, R., Stupnicka, E., Kniewald, Z., Martini, L. (1972) *J. Steroid Biochem.* **3**, 385.
Matias, P.M., Donner, P., Coelho, R., Thomaz, M., Peixoto, C., Macedo, S., Otto, N., Joschko, S., Scholz, P., Wegg, A., Basler, S., Schafer, M., Egner, U., Carrondo, M.A. (2000) *J. Biol. Chem.* **275**, 26164–26171.
Matuo, Y., Adams, P.S., Nishi, N., Yasumitsu, H., Crabb, J.W., Matusik, R.J., McKeehan, W.L. (1989) *In Vitro Cell Dev. Biol.* **25**, 581–584.
Max, S.R. (1983) *J. Steroid Biochem.* **18**, 281.
Max, S.R., Mufti, S., Carlson, B.M. (1981) *Biochem. J.* **200**, 77.
McCague, R., Rowlands, M.G., Barrie, S.E., Houghton, J. (1990) *J. Med. Chem.* **33**, 2452.
McConnell, J.D., Wilson, J.D., George, F.W., et al. (1992) *J. Clin. Endocrinol. Metab.* **74**, 505.
McCullagh, D.R., Cuyler, W.K. (1939) *J. Pharmacol. Exp. Ther.* **66**, 379.
McGavack, T.H., Seegers, W. (1958) *Am. J. Med. Sci.* **235**, 125.
McGill, H.C., Anselmo, V.C., Buchanan, J.M., Sheridian, P.J. (1980) *Science* **207**, 775.
McPhaul, M.J. (1999) *J. Steroid Biochem. Mol. Biol.* **69**, 315–322.
McPhaul, M.J., Marcelli, M., Tilley, W.D., Griffin, J.E., Wilson, J.D. (1991) *FASEB J.* **5**, 2910.
Mercier-Bodard, C., Baulieu, E.E. (1968) *Ann. Endocrinol.* **29**, 159.
Mercier-Bodard, C., Alfsen, A., Baulieu, E.E. (1970) *Acta Endocrinol. (Copenh.)* **147**, 204.
Metcalf, B.W., Wright, C.L., Burkhart, J.P., Johnston, J.O. (1981) *J. Am. Chem. Soc.* **103**, 3221–3222.
Meyer, A.S. (1955) *Biochim. Biophys. Acta* **17**, 441.
Michel, M.G., Baulieu, E.E. (1974) *C.R. Acad. Sci. Paris* **279**, 421.
Miescher, K., Tschapp, E., Wettstein, A. (1976) *Biochem. J.* **30**, 1977.
Mietkiewski, K., Malendowicz, L., Lukaszyk, A. (1969) *Acta Endocrinol. (Copenh.)* **61**, 293.
Mingell, J., Valladares, L. (1974) *J. Steroid Biochem.* **5**, 649.
Miyachi, H., Azuma, A., Kitamoto, T., Hayashi, K., Kato, S., Koga, M., Sato, B., Hashimoto, Y. (1997) *Bioorg. Med. Chem. Lett.* **7**, 1483–1488.
Miyairi, S., Fishman, J. (1986) *J. Biol. Chem.* **261**, 6772–6777.
Moore, R.J., Wilson, J.D. (1972) *J. Biol. Chem.* **247**, 958.
Moore, R.J., Wilson, J.D. (1973) *Endocrinology* **93**, 581.
Mullin, W.G., diPillo, F. (1963) *N. Y. State J. Med.* **63**, 2795.
Murphy, B.E.P. (1968) *Can. J. Biochem.* **46**, 299.
Murphy, B.E.P. (1970) *Steroids* **16**, 791.

Myerson, R.M. (1961) *Am. J. Med. Sci.* **241**, 732.
Naess, O., Hansson, V., Djoseland, O., Attramadal, A. (1975) *Endocrinology* **97**, 1355.
Naftolin, F., Ryan, K.J., Petro, Z. (1971) *J. Clin. Endocrinol. Metab.* **33**, 368.
Naftolin, F., Ryan, K.J., Petro, Z. (1972) *Endocrinology* **90**, 295.
Negro-Vilar, A. (1999) *J. Clin. Endocrinol. Metab.* **84**, 3459.
Neri, R.O. (1976) *Adv. Sex Horm. Res.* **2**, 233.
Neri, R.O., Monohan, M. (1972) *Invest. Urol.* **10**, 123.
Neri, R., Perts, E.A. (1975) *J. Steroid Biochem.* **6**, 815.
Neri, R., Florance, K., Koziol, P., van Cleave, S. (1972) *Endocrinology* **91**, 427.
Neri, R.O., Monahan, M.D., Meyer, J.G., Afonso, B.A., Tabachnick, I.I.A. (1967) *Eur. J. Pharmacol.* **1**, 438.
Neumann, F. (1977) *Horm. Metab. Res.* **9**, 1.
Neumann, F., Junkmann, K. (1963) *Endocrinology* **73**, 33.
Neumann, F., Von Berswordt-Wallrabe, R. (1966) *J. Endocrinol.* **35**, 363.
Neumann, F., Wiechert, R., Kramer, M., Raspe, G. (1966) *Arzneim.-Forsch.* **16**, 455.
Neumann, F., von Berswodt-Wallace, R., Elger, W., Steinbeck, H., Hahn, J., Kramer, M. (1970) *Recent Prog. Horm. Res.* **26**, 337.
Nimni, M.E., Geiger, E. (1957a) *Endocrinology* **61**, 753.
Nimni, M.E., Geiger, E. (1957b) *Proc. Soc. Exp. Biol. Med.* **94**, 606.
Njar, V.C., Hector, M., Hartmann, R.W. (1996) *Bioorg. Med. Chem.* **4**, 1447–1453.
Nowakowski, H. (1962) *Acta Endocrinol. (Copenh.)* **63**(Suppl.), 37.
Nowakowski, H. (1965) *Dtsch. Med. Wochenschr.* **90**, 2291.
Nozu, K., Tamaski, B.I. (1974a) *Biochim. Biophys. Acta* **348**, 321.
Nozu, K., Tamaski, B.I. (1974b) *Acta Endocrinol. (Copenh.)* **76**, 608.
Nutting, E.F., Calhoun, D.W. (1969) *Endocrinology* **84**, 441.
Nutting, E.F., Klimstra, P.D., Counsell, R.E. (1966) *Acta Endocrinol. (Copenh.)* **53**, 627, 635.
Odell, W.D., Swerdloff, R.S. (1978) *Clin. Endocrinol.* **8**, 149.
Ofner, P. (1968) in: Harris, R.S., Wool, I.G., Lorraine, J.A. (Eds.) *Vitamins and Hormones*, Vol. 26, Academic Press, New York, p. 237.
Oh, S.S., Robinson, C.H. (1993) *J. Steroid Biochem. Mol. Biol.* **44**, 389.
Ojasoo, T., Delettre, J., Mornon, J.P., Turpin-VanDycke, C., Raynaud, J.P. (1987) *J. Steroid Biochem.* **27**, 255–269.
Okado, H., Matsuyoshi, K., Tokuda, G. (1964) *Acta Endocrinol. (Copenh.)* **46**, 40.
Olgiati, K.L., Toscano, W.A. (1983) *Biochem. Biophys. Res. Commun.* **115**, 180–185.
Olgiati, K.L., Toscano, D.G., Atkins, W.M., Toscano, W.A. (1984) *Arch. Biochem. Biophys.* **231**, 41–48.
O'Malley, B. (1990) *Mol. Endocrinol.* **4**, 363.
Onoda, M., Haniu, M., Kanagibashi, K., Sweet, F., Shively, J.E., Hall, P.F. (1987) *Biochemistry* **26**, 657.
O'Reilly, J.M., Brueggemeier, R.W. (1996) *J. Steroid Biochem. Mol. Biol.* **59**, 93.
O'Reilly, J.M., Li, N., Duax, W.L., Brueggemeier, R.W. (1995) *J. Med. Chem.* **38**, 2842.
Orestano, F., Altwein, J.E., Knapstein, P., Bandhauer, K. (1975) *J. Steroid Biochem.* **6**, 845.
Osawa, Y., Spaeth, D.G. (1971) Estrogen biosynthesis. Stereo-. specific distribution of tritium in testosterone-1a,2a-t. 2, *Biochemistry* **10**, 66.
Overbeck, G.A., de Visser, J. (1961) *Acta Endocrinol. (Copenh.)* **38**, 285.
Overbeck, G.A., Delver, A., deVisser, J. (1962) *Acta Endocrinol. (Copenh.)* **63**(Suppl.), 7.
Pappo, R., Jung, C.J. (1962) *Tetrahedron Lett.* 365.
Partridge, J.W., Boling, L., DeWind, L., Margen, S., Kinsell, L.W. (1953) *J. Clin. Endocrinol. Metab.* **13**, 189.
Pearlman, W.H., Pearlman, M.R.I. (1961) *J. Biol. Chem.* **236**, 1321.
Pearlman, W.H., Crépy, O. (1967) *J. Biol. Chem.* **242**, 182.
Pearlman, W.H., Fong, I.F.F., Tou, K.J. (1969) *J. Biol. Chem.* **244**, 1373.
Peets, E.A., Henson, M.F., Neri, R. (1974) *Endocrinology* **94**, 532.
Pelc, B. (1964) *Collect. Czech. Chem. Commun.* **29**, 3089.
Pelc, B. (1965) *Collect. Czech. Chem. Commun.* **30**, 3408.
Pelc, B., Jodkova, J. (1965) *Collect. Czech. Chem. Commun.* **30**, 3575.
Perez-Palacios, G., Larsson, K., Beyer, C. (1975) *J. Steroid Biochem.* **6**, 999.
Plantier, H.A. (1964) *N. Engl. J. Med.* **270**, 141.
Potter, G.A., Barrie, S.E., Jarman, M., Rowlands, M.G. (1995) *J. Med. Chem.* **38**, 2463–2471.
Potts, G.O., Arnold, A., Beyler, A.L. (1960) *Endocrinology* **67**, 849.

Potts, G.O., Beyler, A.L., Burnham, D.F. (1960) *Proc. Soc. Exp. Biol. Med.* **103**, 383.
Potts, G.O., Beyler, A., Schane, H.P. (1974) *Fertil. Steril.* **25**, 367.
Rasmusson, G.H., Reynolds, G.F., Steinberg, N.G., et al. (1986) *J. Med. Chem.* **29**, 2298.
Rasmusson, G.H., Chen, A., Reynolds, G.F., Patanelli, D.J., Patchett, A.A., Arth, G.E. (1972) *J. Med. Chem.* **15**, 1165.
Ray, C.G., Kirschvink, J.F., Waxman, S.H., Kelley, V.C. (1965) *Am. J. Dis. Child.* **110**, 618.
Rembiesa, R., Holzbauer, M., Young, P.C.M., Birmingham, M.K., Saffran, M. (1967) *Endocrinology* **81**, 1278.
Renzi, A.A., Chart, J.J. (1962) *Proc. Soc. Exp. Biol. Med.* **110**, 259.
Riegman, P.H., Vlietstra, R.J., van der Korput, J.A., Brinkmann, A.O., Trapman, J. (1991) *Mol. Endocrinol.* **5**, 1921–1930.
Ringold, H.J. (1960) *J. Am. Chem. Soc.* **82**, 961.
Ringold, H.J., Graves, J., Hayano, M., Lawrence, H. Jr. (1963) *Biochem. Biophys. Res. Commun.* **13**, 162.
Ritzen, E.M., Nayfeh, S.N., French, F.S., Dobbins, M.C. (1971) *Endocrinology* **89**, 143.
Ritzén, E.M., Nayfeh, S.N., French, F.S., Aronin, P.A. (1972) *Endocrinology* **91**, 116.
Robaire, B., Covey, D.F., Robinson, C.H., Ewing, L.L. (1977) *J. Steroid Biochem.* **8**, 307.
Rongone, E.L., Segaloff, A. (1963) *Steroids* **1**, 179.
Rongone, E.L., Segaloff, A., Gabbard, B., Carter, A.C., Feldman, E.B. (1963) *Steroids* **1**, 664.
Rosi, D., Neumann, H.C., Christiansen, R.G., Schane, H.P., Potts, G.O. (1977) *J. Med. Chem.* **20**, 349.
Rosner, W., Christy, N.P., Kelley, W.G. (1969) *Biochemistry* **8**, 3100.
Rosso, R., Porcile, G., Brema, F. (1975) *Cancer Chemother. Rep.* **59**, 890.
Roy, A.K. (1973) *Endocrinology* **92**, 957.
Roy, A.K., Milin, B.S., McMinn, D.M. (1974) *Biochim. Biophys. Acta* **354**, 213.
Roy, A.K., Lavrovsky, Y., Song, C.S., Chen, S., Jung, M.H., Velu, N.K., Bi, B.Y., Chatterjee, B. (1999) *Vitam. Horm.* **55**, 309–352.
Russell, D.W., Wilson, J.D. (1993) *Annu. Rev. Biochem.* **63**, 25.
Russell, D.W., Berman, D.M., Bryant, J.T., Cala, K.M., Dairs, D.L., Landrum, C.P., Prihada, J.S., Silver, R.I., Thigpen, A.E., Wigley, W.C. (1994) *Recent Prog. Horm. Res.* **49**, 275.
Ruzicka, L. (1935) *J. Am. Chem. Soc.* **57**, 2011.
Ryan, K.J. (1959) *J. Biol. Chem.* **234**, 268.
Sachs, B.A., Wolfman, L. (1968) *Metabolism* **17**, 400.
Sack, J.S., Kish, K.F., Wang, C., Attar, R.M., Kiefer, S.E., An, Y., Wu, G.Y., Scheffler, J.E., Salvati, M.E., Krystek, S.R. Jr., Weinmann, R., Einspahr, H.M. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4904–4909.
Sala, G. (1960) *Helv. Med. Acta* **27**, 519.
Sala, G., Baldratti, G. (1957) *Proc. Soc. Exp. Biol. Med.* **95**, 22.
Sala, G., Cesana, A., Fedriga, G. (1960) *Minerva Med.* **51**, 1295.
Sala, G., Baldratti, G., Ronchi, R., Clini, V., Bertazzoli, C. (1956) *Sperimentale* **106**, 490.
Sansoy, P.M., Naylor, R.A., Shields, L.M. (1971) *Geriatrics* **26**, 139.
Sar, M., Stumpf, W.E. (1973) *Endocrinology* **92**, 251.
Saunders, F.J. (1966) *Proc. Soc. Exp. Biol. Med.* **123**, 303.
Saunders, F.J., Drill, V.A. (1956) *Endocrinology* **58**, 567.
Saunders, H.L., Holden, K., Kerwin, J.F. (1964) *Steroids* **3**, 687.
Schaub, R.E., Weiss, M.J. (1961) *J. Org. Chem.* **26**, 3915.
Schenkein, I., Levy, M., Bueker, E.D. (1974) *Endocrinology* **94**, 840.
Schwartz, J., Laskin, O., Schneider, S., et al. (1993) *Clin. Pharmacol. Ther.* **53**, 231.
Schwarzel, W.C., Kruggel, W.G., Brodie, H.J. (1973) *Endocrinology* **92**, 866.
Schwarzel, W.C., Kruggel, W.G., Brodie, H.J. (1973) *Endocrinology* **92**, 866–880.
Sciarra, F., Toscano, V., Concolino, G., Di Silverio, F. (1990) *J. Steroid Biochem. Mol. Biol.* **37**, 349–362.
Scow, R.O. (1952) *Endocrinology* **51**, 42.
Segaloff, A. (1963) *Steroids* **1**, 299.
Segaloff, A. (1966) *Recent Prog. Horm. Res.* **22**, 351.
Segaloff, A., Bruce Gabbard, R. (1960) *Endocrinology* **67**, 887.
Segaloff, A., Gabbard, R.B. (1963) *Steroids* **1**, 77.
Segaloff, A., Gabbard, R.B. (1973) *Steroids* **22**, 99.
Segaloff, A., Gabbard, B., Carriere, B.T., Rongone, E.L. (1965) The metabolism of 4-14c-17-alpha-methyltestosterone, *Steroids* **58**(Suppl. 1), 419.
Shahidi, N.T. (1973) *N. Engl. J. Med.* **289**, 72.
Shain, S.A., Boesel, R.W. (1975) *J. Steroid Biochem.* **6**, 43.

Sherrins, R.J., Gandy, H.M., Thorsland, T.W., Paulsen, C.A. (1971) *J. Clin. Endocrinol. Metab.* **32**, 522.
Shimazaki, J., Furuya, N., Yamanaka, H., Shida, K. (1969) *Endocrinol. Jpn.* **16**, 163.
Shimazaki, J., Matsushita, I., Furuya, N., Yamanaka, H., Shida, K. (1969) *Endocrinol. Jpn.* **16**, 453.
Sholiton, L.S., Werk, E.E. (1969) *Acta Endocrinol. (Copenh.)* **61**, 641.
Sholiton, L.S., Mornell, R.T., Werk, E.E. (1966) *Steroids* **8**, 265.
Sholiton, L.S., Hall, I.L., Werk, E.E. (1970) *Acta Endocrinol. (Copenh.)* **63**, 512.
Shroff, A.P., Harper, C.H. (1969) *J. Med. Chem.* **12**, 190.
Siiteri, P.K., Thompson, E.A. (1975) *J. Steroid Biochem.* **6**, 317.
Siiteri, P.K., Wilson, J.D. (1970) *J. Clin. Invest.* **49**, 1737.
Simard, J., Singh, S.M., Labrie, F. (1997) *Urology* **49**, 580–586.
Simpson, E.R., Merrill, J.C., Hollub, A.J., Grahan-Lorence, S., Mendelson, C.R. (1989) *Endocr. Rev.* **10**, 136.
Simpson, E.R., Mahendroo, M.S., Means, G.D., Kilgore, M.W., Hinshelwood, M.M., Grahan-Lorence, S., Amarneh, B., Ito, Y., Fisher, C.R., Michael, M.D., Mendelson, C.R., Bulun, S.E. (1994) *Endocr. Rev.* **15**, 342.
Skinner, S.J.M., Akhtar, M. (1969) *Biochem. J.* **114**, 75.
Skiner, R.W.S., Pozderac, R.V., Counsell, R.E., Weinhold, P.A. (1975) *Steroids* **25**, 189.
Skinner, R.W.S., Pozderac, R.V., Counsell, R.E., Hsu, C.F., Weinhold, P.A. (1977) *Steroids* **25**, 189.
Small, E.J., Carroll, P.R. (1994) *Urology* **43**, 408–410.
Smit, A., Westerhof, P. (1963) *Recl. Trav. Chim. Pavs-Bas* **82**, 1107.
Snider, C.E., Brueggemeier, R.W. (1987) *J. Biol. Chem.* **262**, 8685–8689.
Solo, A.J., Bejba, N., Hebborn, P., May, M. (1975) *J. Med. Chem.* **18**, 165.
Sonnenschein, C., Olea, N., Pasanen, M.E., Soto, A.M. (1989) *Cancer Res.* **49**, 3474.
Spence, A.M., Sheppard, P.C., Davie, J.R., Matuo, Y., Nishi, N., McKeehan, W.L., Dodd, J.G., Matusik, R.J. (1989) *Proc. Natl. Acad. Sci. U.S.A* **86**, 7843–7847.
Stafford, R.O., Bowman, B.J., Olson, K.J. (1954) *Proc. Soc. Exp. Biol. Med.* **86**, 322.
Starka, L., Sulcova, J., Broulik, P.D., Joska, J., Fajkos, J., Doskocil, M. (1977) *J. Steroid Biochem.* **8**, 939.
Starka, L., Hanapl, R., Bicikova, M., Cerny, V., Fajkos, J., Kosal, A., Kocovsky, P., Kohout, L., Velgova, H. (1980) *J. Steroid Biochem.* **13**, 455.
Steeno, O., Heyns, W., Van Baelen, H., De Moor, P. (1968) Testosterone binding in human plasma, *Ann. Endocrinol.* **29**, 141.
Steinbach, M. (1968) *Sportarzt. und Sportmedizin* **11**, 485.
Steinkampf, M., Mendelson, C.R., Simpson, E.R. (1987) *Mol. Endocrinol.* **1**, 465.
Steinkampf, M., Mendelson, C.R., Simpson, E.R. (1988) *Mol. Cell. Endocrinol.* **59**, 93.
Stern, J.M., Eisenfield, A.J. (1969) *Science* **166**, 233.
Stern, J., Eisenfeld, A.J. (1971) *Endocrinology* **88**, 1117.
Stevens, J., Jaw, J., Peng, C.T., Halpert, J. (1991) *Biochemistry* **30**, 3649.
Strömme, S.B., Meen, H.D., Aakvaag, A. (1974) *Med. Sci. Sports* **6**, 203.
Stucki, J.C., Forbes, A.D., Northam, J.I., Clark, J.J. (1960) *Endocrinology* **66**, 585.
Stylianou, M.I., Forchielli, E., Tummillo, N.I., Dorfman, R.I. (1961) *J. Biol. Chem.* **236**, 692.
Suchowsky, G.K., Junkmann, K. (1962) *Acta Endocrinol. (Copenh.)* **39**, 68.
Sutherland-Bawlings, E.A.P. (1970) *Br. Med. J.* **111**, 643.
Takayasu, S., Adachi, K. (1975) *Endocrinology* **96**, 525.
Tan, S.Y. (1974) *J. Clin. Endocrinol. Metab.* **39**, 936.
Tanaka, A., Miyamoto, K., Minamino, N., Takeda, M., Sato, B., Matsuo, H., Matsumoto, K. (1992) *Proc. Natl. Acad. Sci. U.S.A* **89**, 8928–8932.
Tanayama, S., Yoshida, K., Kondo, T., Kanai, Y. (1979) *Steroids* **33**, 65.
Tec, L. (1971) *Am. J. Psychiatry* **127**, 1702.
Thompson, E.A., Siiteri, P.K. (1974a) *J. Biol. Chem.* **249**, 5373.
Thompson, E.A., Siiteri, P.K. (1974b) *J. Biol. Chem.* **249**, 5364.
Tindall, D.J., French, F.S., Nayfeh, S.N. (1973) *Biochem. Biophys. Res. Commun.* **49**, 1391.
Townsley, J.D., Brodie, H.J. (1968) Studies on the mechanism of estrogen biosynthesis. 3. The stereochemistry of aromatization of C19 and C18 steroids, *Biochemistry* **7**, 33.

Tremolieres, J., Pequignot, E. (1965) *Presse Med.* **73**, 2655.
Tucker, H., Crook, J.W., Chesterson, G.J. (1988) *J. Med. Chem.* **31**, 954–959.
Tveter, K.J., Attramadal, A. (1968) *Acta Endocrinol. (Copenh.)* **59**, 218.
Tveter, K.J., Unhjem, O. (1969) *Endocrinology* **84**, 963.
Tweedle, D.E.F., Walton, C., Johnston, I.D.A. (1972) *Br. J. Surg.* **59**, 300.
Tyrell, C.J. (1992) *Prostate* **4**, 97.
Valladares, L., Mingell, J. (1975) *Steroids* **25**, 13.
Van Moorselaar, R., Halkes, S.J., Havinga, E. (1965) *Recl. Trav. Chim. Pays-Bas* **84**, 841.
Van Wauwe, J.P., Janssen, P.A. (1989) *J. Med. Chem.* **32**, 2231.
Verhoeven, G., Wilson, J.D. (1976) *Endocrinology* **99**, 79.
Vermeulen, A., Verdonck, L. (1968) *Steroids* **11**, 609.
Vermeulen, A., Rubens, R., Verdonck, L. (1972) *J. Clin. Endocrinol. Metab.* **34**, 730.
Vernon, R.G., Kopec, B., Fritz, I.B. (1974) *Mol. Cell. Endocrinol.* **1**, 167.
Vitali, R., Gardi, R., Falconi, G., Ercoli, A. (1966) *Steroids* **8**, 527.
Voigt, K.D., Horst, H.-J., Krieg, M. (1975) *Vitam. Horm.* **33**, 417.
Voigt, W., Castro, A., Covey, D.F., Robinson, C.H. (1978) *Acta Endocrinol.* **87**, 668.
Wade, N. (1972) *Science* **176**, 1399.
Wakeling, A.E., Furr, B.J.A., Glen, A.T., Hughes, L.R. (1981) *J. Steroid Biochem.* **15**, 355.
Walser, A., Schoenenberger, G. (1962) *Schweiz. Med. Wochenschr.* **92**, 897.
Walsh, P.C., Madden, J.D., Harrod, M.J., Goldstein, J.L., MacDonald, P.C., Wilson, J.D. (1974) *N. Engl. J. Med.* **291**, 944.
Wang, N.G., Guan, M.Z., Lei, H.P. (1987) *J. Ethnopharmacol.* **20**, 45–51.
Weinstein, A., Lindner, H.R., Frielander, A., Bauminger, S. (1972) *Steroids* **20**, 789.
Weisenfeld, S., Akgun, S., Newhouse, S. (1963) *Diabetes* **12**, 375.
Weller, O. (1961) Research on the androgenic action of methyl androsten-ol-on-acetate and -enanthate, *Endokrinologie* **41**, 60.
Weller, O. (1962) Comparative studies on the metabolic effect of testosterone and 1-methyl-delta-androstenolone in cases with primary hypogonadism, *Endokrinologie* **42**, 34.
Weller, O. (1966) *Arzneim.-Forsch.* **16**, 465.
Westaby, D., Ogle, S.J., Paradinas, F.J., Randell, J.B., Murray-Lyon, I.M. (1977) *Lancet* **2**, 261.
Wilding, G., Chen, M., Gelmann, E.P. (1989) *Prostate* **14**, 103.
Williams, G., Kerle, D.J., Ware, H., Doble, A., Dunlog, H., Smith, C., Allen, J., Yeo, T., Bloom, S.R. (1986) *Br. J. Urol.* **58**, 45.
Wilson, E.M., French, F.S. (1976) *J. Biol. Chem.* **251**, 5620.
Wilson, J.D. (1988) *Endocr. Rev.* **9**, 181.
Wilson, J.D., Gloyna, R.E. (1970) *Recent Prog. Horm. Res.* **26**, 309.
Wing, L.Y., Garrett, W., Brodie, A.M.H. (1985) *Cancer Res.* **45**, 2425–2428.
Wolf, D.A., Schulz, P., Fittler, F. (1991) *Br. J. Cancer* **64**, 47–53.
Wolff, M.E., Zanati, G. (1969) *J. Med. Chem.* **12**, 629.
Wolff, M.E., Zanati, G. (1970) *Experientia* **26**, 1115.
Wolff, M.E., Kasuya, Y. (1972) *J. Med. Chem.* **15**, 87.
Wolff, M.E., Ho, W., Kwok, R. (1964) *J. Med. Chem.* **7**, 577.
Wolff, M.E., Zanati, G., Shanmagasundarum, G., Gupte, S., Aadahl, G. (1970) *J. Med. Chem.* **13**, 531.
Wysowski, D.K., Freiman, J.P., Tourtelot, J.B., Horton, M.L. III (1993) *Ann. Intern. Med.* **118**, 860–864.
Yan, G., Fukabori, Y., Nikolaropoulos, S., Wang, F., McKeehan, W.L. (1992) *Mol. Endocrinol.* **6**, 2123–2128.
Yeh, J., Barbieri, R.L., Friedman, A.J. (1989) *J. Steroid Biochem.* **33**, 627.
Yorka, K.V., Truett, W.L., Johnson, W.S. (1962) *J. Org. Chem.* **27**, 4580.
Zaffaroni, A. (1960) *Acta Endocrinol. (Copenh.)* **50**(Suppl.), 139.
Zanati, G., Wolff, M.E. (1971) *J. Med. Chem.* **14**, 958.
Zanati, G., Wolff, M.E. (1973) *J. Med. Chem.* **16**, 90.
Zanati, G., Gaare, G., Wolff, M.E. (1974) *J. Med. Chem.* **17**, 561.
Zeller, F.J. (1971) *J. Reprod. Fertil.* **25**, 125.
Zhou, Z.-X., Wong, C.-L., Sar, M., Wilson, E.M. (1994) *Recent Prog. Horm. Res.* **49**, 249.
Zhuang, L.Z., Philips, D.M., Gunsalus, G.L., Bardin, C.W., Mather, J.P. (1983) *J. Androl.* **4**, 336–344.
Ziminski, S.J., Brandt, M.E., Covey, D.F., Puett, C. (1987) *Steroids* **50**, 135–146.

Shaping Adaptive Immunity against Pathogens: The Contribution of Innate Immune Responses: *see* Pathogens: Innate Immune Reponses

Shotgun Sequencing (SGS)

Jun Yu, Gane Ka-Shu Wong, Jian Wang, and Huanming Yang
Beijing Genomics Institute, Beijing & James D. Watson Institute of Genome Sciences, Hangzhou, China

1 Introduction 73
1.1 A Brief History of DNA Sequencing 76
1.2 Overview of SGS Methodology and Application 79

2 Essentials for SGS 83
2.1 Unbiased Clone Coverage 84
2.2 Sequence Quality Control 86
2.3 Repetitive Contents 87
2.4 Physical and Sequence Gaps 88
2.5 Validation of Sequence Assembly 88
2.6 Integration of Genomic Information 89

3 Technical and Experimental Basics 89
3.1 Genomic DNA Libraries 90
3.1.1 Source DNA 91
3.1.2 Quality Assessment 91
3.2 Sequencing Data Acquisition 92
3.2.1 WG-SGS and CBC-SGC 92
3.2.2 Directed Cloning and Random SGS 93
3.2.3 Process Control 93
3.3 Sequence Assembly 94
3.3.1 An Overview of General Procedures 94
3.3.2 Classification of Repetitive Sequences 94
3.3.3 Software Packages for Sequence Assembly 95
3.3.4 Building Contigs and Scaffolds 95
3.4 Sequence Finishing 96
3.4.1 Improving Sequence Quality 98
3.4.2 Primer-walking and End-sequencing 98

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

3.4.3 Closing Sequence Gaps 99
3.4.4 Physical Mapping and Clone Fingerprinting 100
3.5 Genome Annotation and Analysis 100
3.5.1 Compositional and Structural Dynamics 101
3.5.2 Genome Annotation 102
3.5.3 Sequence Variations 104
3.5.4 Genome Evolution 105

4 Other Useful Resources for SGS 107
4.1 ESTs and Full-length cDNA 107
4.2 Physical and Genetic Maps 108
4.3 Comparative Genome Analyses 109

5 Perspectives 109
5.1 Large Genomes 109
5.2 Genomes with High Repetitive Sequence Content 110
5.3 Polyploidy: Ancient and New 110
5.4 Mixed and Meta Genomes 110
Acknowledgment 111

Bibliography 111
Books and Reviews 111
Primary References 111

Keywords

Large-insert Clones
Clones made as cosmids, fosmids, BACs or YACs, with an insert size around or over 40 Kb.

Minimal-tiling-path Clones
Physically mapped clones that cover their source DNA molecule with a single path and least amount of overlaps.

Physical Gap
Gaps not covered by available clones in a clone assembly.

Physical Map
DNA sequences or chromosome segments delineated by a set of ordered and oriented markers or fingerprinted clones.

Scaffold
A group of ordered and oriented contigs with physical contiguity but without sequence contiguity.

Sequence Contig
A contiguous sequence assembly composed of overlapping sequences.

Sequence Gap
Gaps not covered by sequences in a sequence assembly.

Shotgun Sequencing
A DNA sequencing technique used for achieving sequence contiguity by acquiring random sequences from clone libraries made of the target DNA.

Shotgun sequencing (SGS) is primarily a large-scale sequencing (LSS) technique that does not rely on precise guiding information about the target DNA, which includes large-insert clones and single genomes, ranging from thousands of basepairs (Kb) to billions of basepairs. A mixture of smaller genomes can also be sequenced in a similar way when retrospective means are available to assemble and distinguish them. It provides a fast and cost-effective way of sequencing large genomes regardless of whether the project as a whole takes a ''whole-genome'' (WG) or ''clone-by-clone'' (CBC) approach. The success of SGS essentially depends on random sampling, high-quality data, sufficient sequence coverage, effective assembly, and gap closing procedures. SGS has already demonstrated its tremendous power in sequencing not only microbial genomes but also large eukaryotic genomes, such as those of the cultivated rice and the laboratory mouse. Together with improved sequencing technologies and computing tools, SGS is expected to play a central role in the field of genomics, especially when the latter has to constantly face some major challenges, scientific, managerial, and political. Some of the scientific challenges relate to the effective sequencing of large, polyploid, and mixed genomes, as well as those with highly repetitive sequence contents. A key managerial challenge is for the operators to consistently produce high-quality data while increasing throughput and reducing cost. It is always a tough decision for a steering committee organizing a genome project to choose between WG and CBC, but SGS is always the basic technique of choice.

1 Introduction

DNA sequencing has become one of the essential laboratory methods used in molecular biology, together with DNA cloning, DNA microarray, polymerase chain reaction (PCR), electrophoresis, chromatography, and centrifugation. It has become not only a fundamental method of determining genome sequences but also a basic tool for identifying DNA methylation sites and sequence polymorphisms. In both scale and scope, DNA

sequencing outperforms all other molecular technologies in acquisition of basic genomic information, thereby paving the way for other genome-scale, technology-oriented disciplines and, to use contemporary terminology, the numerous "-omics" (transcriptomics and proteomics being well-accepted examples) that will be leading biologists into the future decades, providing new data, ample information, fresh knowledge, and innovative insights.

At the leading edge of genomics, the principal ideas of DNA sequencing are not fundamentally different from the technical innovations demonstrated in the early days of technology development in molecular biology. When a cloned DNA molecule is larger than one (from one end) or two (from both ends) sequencing readouts (often referred as *sequencing reads*; screenshots of a sequence trace and a sequence alignment are shown in Fig. 1) from an electrophoresis-based instrument, its full-length sequence has to be obtained in one of the following two ways: (1) a stepwise primer-directed "walking" strategy, based

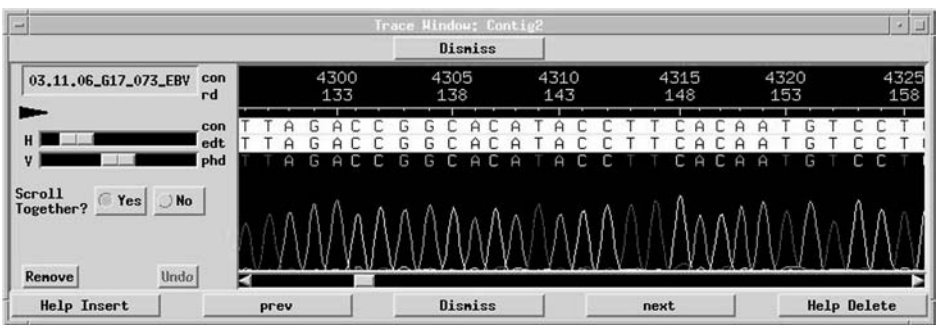

(a)

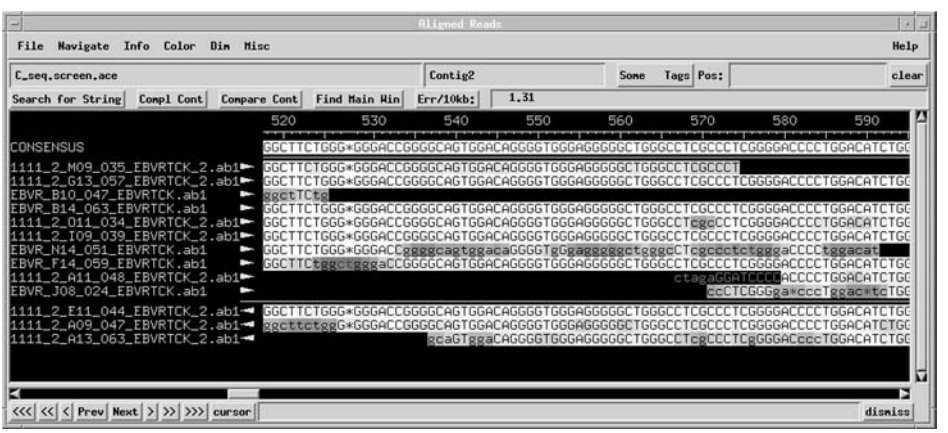

(b)

Fig. 1 Screenshots of a partial sequence read (a) and a sequence assembly with aligned sequences (b). Brightness in (b) indicates the quality of the sequence as estimated and displayed by a software tool kit (sequence assembler), Phrad-Phrep-Consed.

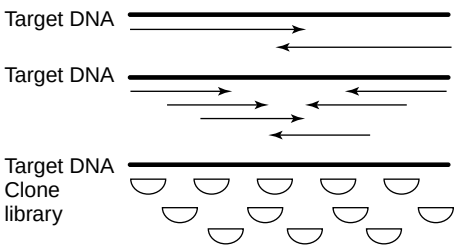

Fig. 2 Two basic ways of sequencing DNA beyond a size of two sequencing reads: (a) to "walk" in a stepwise way from both ends by designing primers based on newly acquired sequence information from a previous walk, and (b) to "shotgun" the target DNA into smaller pieces that are carried by a clone library (or subclones). Random sequencing of the subclones with high redundancy leads to an assembly whose consensus sequence is equivalent to that of the target. Arrows depict sequencing reads and their directions.

on newly acquired sequences; (2) a parallel subcloning strategy, based on directed or "shotgun" cloning techniques by breaking the target DNA into smaller pieces so that a collection of sequencing reads from overlapping "subclones" sufficiently covers the entire target length (Fig. 2). It is obvious that the choice of approaches depends on the capability and throughput of a sequencing operation. Whereas one is managed serially and is therefore limited in scale, the other involves paralleling and is suitable for high-throughput, large-scale operations. The current capability of a typical sequencing instrument limits the sequencing length in a single run to just 500 to 2000 basepairs (bp). Besides solutions involving instrumentation (such as parallelization and incremental increase of read length), the choice of approaches for any large-scale sequencing (LSS) project has to take advantage of SGS for large-insert clones over tens of kilobasepairs (Kb), or genomes in a size range from a few kilobases (such as large viral genomes) to a few megabasepairs (Mb; such as bacterial genomes), right up to a few hundred or even a few thousand Mb in length (such as higher plants and animals).

SGS has been successfully applied for sequencing many large genomes, essentially at two levels, targeting either clones or whole genomes. Large genome-sequencing projects are often initiated by large-insert clone-based physical mapping, especially for important model organisms from representative taxonomic groups; the Human Genome Project (HGP) is the most famous example. These physical mapping techniques include YAC (yeast artificial chromosome)-based or BAC (bacterium artificial chromosome)-based STS (sequence-tagged site) mapping, and restriction-digest mapping with different types of large-insert clones. These clones are subsequently sequenced by SGS, guided by a collection of clones that cover the targeted region with least overlaps, often termed as minimal tiling path (MTP). Physical maps providing anchored clones for sequencing are often referred to as *sequence-ready maps*. In a map-directed process, sequence data are captured at clone level and merged, after data from overlapping clones are completed individually. Although SGS is utilized at clone level, such an approach is often referred as a *clone-by-clone SGS or CBC-SGS*, as opposed to a whole-genome SGS, or WG-SGS. In the second scenario, when sufficient information (including estimated genome size, number of chromosomes, GC-content, repeat contents, and genetic homogeneity) about a genome has been acquired either from the organism itself or from a closely related

one, the WG-SGS approach becomes feasible, and is often the method of choice. In most cases, SGS has its "sweet range" in sequencing genomes, from a few Mb to a few hundred Mb. However, the quality and completeness of a final assembly habitually depends on repetitive sequence content, clonability, and complexity of the genome, coupled with other factors such as funding, project management, and homogeneity of the source DNA (sequence variations among individuals that donate DNA sometimes are known to delay a quality sequence assembly for months or even years). Before long, SGS will have to face even greater challenges, when it begins to be applied to genomes in a size range around thousands of Mb, polyploid genomes, and even a mixture of genomes. Novel approaches, techniques, and computing tools will have to be gradually introduced in order to make SGS a more efficient and robust technique. Two major significant factors are driving this effort:: the ever-increasing desire for more and more sequence information, and the incremental cost reductions, as evidenced in the history of genomics over the past decade.

1.1 A Brief History of DNA Sequencing

The concept and technology of DNA sequencing have been evolving over the past thirty years. The initial phase of DNA sequencing was born with the development of the *basic technology*, fundamentally a series of technology advancements in DNA sequencing and molecular cloning around the early 1970s. The second phase began with the proposals to set up the Human Genome Project in 1983 and 1984. The decade following this, before the advent of LSS of HGP in 1995, was the most critical period for *special technology and concept* development, most notably physical mapping and DNA sequencing. The past 10 years have been the *Large-scale Sequencing Era*, which is undoubtedly going to last at least another decade or even longer. In this period, alongside additional technological and strategic improvements, there has been a collective effort on the part of those involved in sponsored academic research as well as those belonging to the profit-making biotech industry, to make improvements in the areas of policy development, organizational reforms, and management practices. This has resulted in a continual reduction in the cost of sequencing from the earlier standard level of "$1 per base" (calculated as $40 per sequence read on the assumption of 8x coverage of a target genome) when HGP first started, to the real cost of $15 per read when HGP transitioned into the Large-scale Sequencing Era, to a current cost estimate of $1 per read. These are the estimated direct costs for academic operations and not the retail price offered by private companies. Some academic institutions have claimed lower costs in a rather variable range from $0.3 to $0.8. The cost of sequencing is expected to go down continually over the next 5 to 10 years, dropping at a rate of 50 percent every two years (Fig. 3). Meanwhile, of course, some new ground-breaking technology could emerge and dramatically change the current practice.

The DNA polymerase-based sequencing method and a variety of molecular cloning techniques are the centerpieces of contemporary DNA sequencing technology. In the early days, primer-dependent tactics and shotgun principles were already being employed in tackling large pieces of DNA molecules and certain levels of

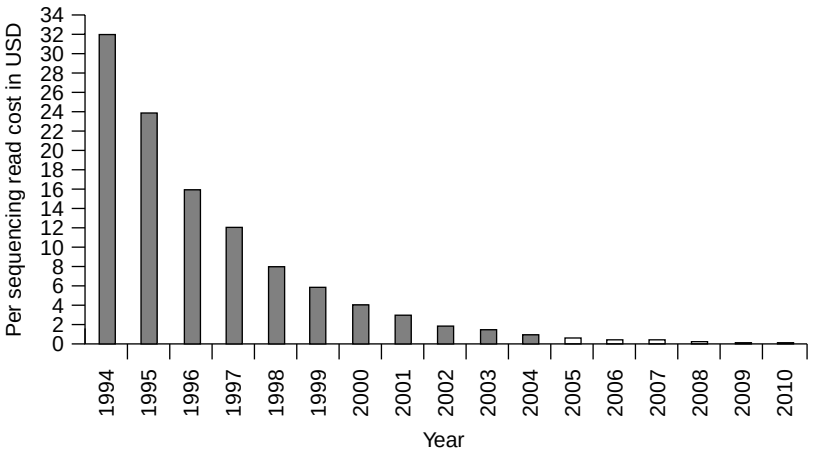

Fig. 3 The trend of estimated cost reductions since the beginning of large-scale sequencing (LSS). Although the legendary estimate, "one dollar per basepair," at the beginning of the HGP should be marked as a starting point, LSS was not launched on a full scale until the middle of 1990s when capillary sequencers and robotics-aided basic sequencing protocols became available. Real academic costs are calculated over a course of 10 years (1995–2004; solid bars). The costs beyond 2004 are estimated according to the trend (open bars). An empirical rule becomes clear, namely, that the sequencing cost is halved every two years.

complexity, such as overhang sequences that are common in bacteriophages. It should be remembered that it took the pioneers over a couple of decades to achieve an increase in sequencing read length from a few tens of basepairs to a couple of hundred basepairs. The first genome sequenced by a "shotgun" design was the bacteriophage phiX174 (5386 bp in length and containing 10 genes) published in 1977. Since then, bacterial cloning systems have leveraged the biology of both bacteriophages and plasmids. They have been designed to achieve higher yields for cloned DNA (such as high-copy plasmid), and equipped with polylinkers composed of multiple cutting sites of commonly used restriction enzymes, as well as several choices in the matter of antibiotic resistance.

The WG-SGS approach was first successfully applied to a microbial genome, *H. influenzae* (1 830 138 bp in size; a bacterium causing upper respiratory infection) in 1995. The success of the WG-SGS approach has opened the floodgates for microbial genome sequencing. As a result, collections of completed microbial genome sequences are updated almost on a daily basis, now number in the hundreds. The first metazoan genome to be sequenced using the WG-SGS method was the *Drosophila* genome with a euchromatic content of slightly over 122 Mb. The SGS data were subsequently combined with those from the public sequencing effort on the same organism, thereby unlocking another information gateway towards insect genomics. Insect genomes, such as mosquitoes and silkworms, were sequenced subsequently by the WG-SGS method in different institutions. The Human Genome Project was carefully planned and initiated with a genome-scale

physical mapping effort, followed by SGS of large-insert clones, including YACs and cosmids, but mostly BACs. It was therefore labeled as an SGS approach, although the large-insert clones were sequenced one by one using SGS. The landmarks for CBC-SGS were *E. coli* K-12 (the first bacterium), *S. cerevisiae* (the first eukaryotic species), and *C. elegans* (the first metazoan species). Private teams later used the WG-SGS approach to interrogate the human and mouse genomes, but the wealth of public data were also carefully merged to arrive at the final assemblies. The mouse genome was the first trial of the WG-SGS strategy by the publicly supported consortium and the rat genome was the next. Following the *Arabidopsis* genome (150 Mb), the first plant genome deciphered with a CBC-SGS approach, the first crop plant genome – the rice genome (450 Mb) – has been sequenced four times. Two of the efforts were led by private companies with the WG-SGS approach. An international consortium took a CBC approach and has completed sequencing several chromosomes of the rice genome from the same cultivar as the one sequenced by the private companies, namely, a *japonica* rice, *Nipponbare*. The fourth rice genome, an *indica* variety, was sequenced by our own institute as part of its Hybrid Rice Genome Project using WG-SGS strategies. The WG-SGS wave has reached new heights. Large genomes of various species, even certain taxonomic groups (such as a suite of vertebrates, mammals, and grasses), are being placed in undergoing sequencing operations, and these are expected to be completed in the next 5 to 10 years. The selection of these targets for sequencing is justified because of their relevance to human biology and diseases, as well as to broader areas of economic interest such as domestic crops, farm animals, and their wild relatives.

From a technical perspective, two major breakthroughs have revolutionized large-scale sequencing. The first is the invention of large-insert clones, YAC being the first and BAC, the second. They offered unprecedented substrates for both LSS and physical mapping. The timely invention of PCR technology also stimulated the new physical mapping concept of STS. Since mapped YACs have to be subcloned into bacterial vectors for sequencing (because YACs are intermingled with host chromosomes, their DNAs are not readily separable by straightforward chemical methods), BACs have now become the preferred substrate. The second breakthrough was the invention of four-colorfluorescence-based methodology and instrumentation, allowing for a high degree of parallelization and miniaturization to come into play. Capillary-based sequencers are the current choice of high-throughput sequencing, although versions of slab-gel-based sequencers emerge from time to time. A central computational support for large-scale sequence acquisition and analysis is the concept of a quality score system for each sequenced base, pioneered by Phil Green and his colleagues. Other emerging technologies such as "single molecule" sequencing technology are being sponsored by certain funding agencies, but these are still in their infancy or at a proof-of-principle stage, and may take years of development and research in order to be introduced to a sequencing equipment assemblage.

On the whole, it is clear that SGS has becoming a mainstay of LSS because of collective experience, sharpened software tools, and increasing knowledge about the structure of target genomes. It is most

likely that many large genomes will be sequenced by the SGS method rather than by the CBC one in the years to come.

1.2 Overview of SGS Methodology and Application

Other than biological significance, two categories of practical issues have to be considered for any large genome project beyond a few Mb in target size: managerial and technical. In general, whereas the total number of sequence reads required to cover a genome length is comparable in the two versions of SGS, namely, WG and CBC, WG-SGS is both money- and time-saving because it largely skips a labor-consuming step: clone library construction based on sequence-ready maps. For a CBC-SGS approach, each of the large-insert clones (often BACs) has to be shotgun cloned into small-insert clones (often plasmids). The larger the genome (or the more the MTP clones), the more the libraries that have to be constructed. For instance, to sequence a 500-Mb genome, over 3500 physically ordered BAC (assuming 150 Kb as an average length for BACs) clones have to be made into subclone libraries, whereas only a few carefully prepared small-insert clone libraries are adequate for WG-SGS. If carefully planned, a genome project could start with randomly picked BACs from a whole-genome library that sufficiently covers an entire length, followed by a low-pass sequence sampling (such as a few hundred sequence reads) and sequence overlap assessment, and finished with parallel procedures to complete each BAC of MTP. Although such a strategy appears feasible, it is as yet untried. It is composed of multiple tasks at different levels, for which the project has to be managed in a serial fashion. In contrast WG-SGS, in its unique path, provides a highly parallel process avoiding managerial pressures often created at intermediate levels. Table 1 lists some completed genome sequences in categories of sequence methods, WG-SGS and CBC-SGS.

SGS has never been a stand-alone technology. It includes many other cloning and computing techniques and strategic approaches along the path of its expanding applications. The first set of technologies has to do with molecular cloning. Currently, the preferred cloning systems are BACs, that are best for cloning large DNA fragments and for clone stability; and plasmids, best as regards high yield and readiness for preparing DNA templates for sequencing. BACs are also easier to fingerprint by six-base cutting enzymes (as opposed to YACs), and the system has a copy number control scheme that is believed to enhance clone stability in bacterial hosts, as compared to cosmids. Other merits of using BACs include their simple, robust DNA preparation procedures for parallel and scaled operations, and the fact that they are better reagents for physical mapping for long-range coverage and contiguity (BACs can be made in an insert size range of 100 to 200 Kb) than the lambda packaging-based cosmids and fosmids, which are limited by an insert size range of 30 to 40 Kb. However, one has to realize that cosmids and fosmids are both suitable for SGS on a limited scale and BAC libraries are not as easy to make in large quantity as the cosmid system. Regardless of which of the large-insert cloning systems is employed, plasmids provide an excellent direct substrate for SGS sequencing, whereas others can only be directly end-sequenced on a limited scale for achieving better contiguity over a variable range. On certain occasions, miniature BACs, or mini-BACs

Tab. 1 Representative sequenced genomes with WG-SGS and CBC-SGS strategies.

Species	*Genome Size [Mb]*	*Strategy*	*# of Contigs*	*# of Scaffolds*
B. mori (Silkworm)	530	WG-SGS	213 289	N/A
			N/A	N/A
P. chrysosporium (White rot fungus)	30	WG-SGS	1323	767
G. gallus (Chicken)	1200	WG-SGS	111 864	N/A
A. mellifera (Honey bee)	300	WG-SGS	30 074	N/A
C. merolae (Red algae)	16.5	WG-SGS	N/A	N/A
R. norvegicus (Rat)	2750	WG-SGS	137 910	N/A
N. crassa (Bread mold)	40	WG-SGS	821	N/A
P. troglodytes (Chimpanzee)	3000	WG-SGS	361 864	37 931
C. briggsae (Nematode)	104	WG-SGS	578	N/A
C. familiaris (Dog)	2500	WG-SGS	1 089 636	N/A
M. musculus (Mouse)	2500	WG-SGS	224 713	N/A
P. falciparum (Malaria parasite)	23	WG-SGS	N/A	N/A
O. sativa (Rice)	450	WG-SGS	103 044	N/A
			N/A	N/A
T. rubripes (Pufferfish)	365	WG-SGS	12 381	20 379
T. nigroviridis (Pufferfish)	350	WG-SGS	108 177	N/A
C. intestinalis (Sea squirt)	160	WG-SGS	2501	2501
A. gambiae (Mosquito)	278	WG-SGS	69 724	8987
H. Sapiens (Human)	3000	CBC-SGS	N/A	N/A
		WG-SGS	221 036	N/A
D. melanogaster (Fruit fly)	165	WG-SGS	2775	N/A
				N/A
A. thaliana (Mouse-ear Cress)	125	CBC-SGS	N/A	N/A
C. elegans (Nematode)	97	CBC-SGS	N/A	N/A
E. coli	4.6	CBC-SGS	N/A	N/A
S. cerevisiae (Yeast)	12.1	CBC-SGS	N/A	N/A

Note: Some of the genomes have not yet been completed, such as *B. mori, R. norvegicus, C. briggsae* and *M. musculus.*

(10 to 20 Kb) are used to compensate some of the systemic drawbacks when insert size and clone stability are of major considerations. Occasionally, mini-insert (100 to 200 bp) clones and YAC clones have to be utilized in coping with low GC-content DNA that is extremely unstable in bacterial hosts, such as is the case in the sequencing of the malaria parasite, *P. falciparum*.

The second set of technologies crucial for SGS comprises various DNA preparation methods for sequencing substrates. Two basic criteria for choosing a practical method are scalability and cost; both demand robust and automatable protocols. The M13 bacteriophage system used in the early days of LSS was soon given up, largely because of its requirement for repeated precipitation steps in isolating phage particles and DNA, which significantly hinders the automation process, although it is capable of producing high-quality data. Being part of the rate-limiting steps in large-scale sequencing, plasmid DNA preparation methods have evolved a great deal over the past few years. Besides alkaline lysis protocols, absorbent-based and isothermal amplification protocols are

also widely used in plasmid DNA preparations. In the absorbent-based protocol, centrifugation steps are mostly reduced to simple pluming, and the isothermal amplification protocol is aimed at eliminating bacterial precipitation and plasmid DNA preparation all together. PCR-based protocols also seem to be convenient in meeting the need for preparation of sequencing templates, but we have yet to see a robust one that yields high-quality data. Despite the difficulty in applying PCR technology to template preparations for large-scale SGS, it is still used in all sequencing reactions, catalyzed by a thermal stable polymerase. It is also vital in primer-directed walking and PCR product sequencing for joining sequence contigs in the finishing phase of genome-sequencing projects.

The most demanding component of all SGS-related technologies is biocomputing, broadly termed as bioinformatics, especially for WG-SGS. Bioinformatic tools are often packaged when applied to almost all aspects of SGS, such as data management, sequence contig assembly, and genome annotation. While being less selective concerning databases and data management systems, SGS relies heavily on sequence assembly and gene-finding algorithms. A WG-SGS project can be initiated only when a qualified sequence assembler is put in place. There are several versions of sequence assembly software packages that are able to handle a target size range of a few tens of Kb to a few Mb (Table 2). By far, the most popular package is the Phred-Phrap-Consed developed by Phil Green and his colleagues. For large genomes in a size range of a few tens of Mb to hundreds of Mb, even thousands of Mb, additional algorithms are required to facilitate the complicated assembly process, including repeat identification and masking, contig and scaffold assembly, and customized annotation. Finally, in most of the cases, a specialized database has to be built to host the integrated information for a sequenced genome, in order to better serve the relevant research community in addition to timely submissions of the generated data to the public databases across continents.

Even though most of the genome projects are aimed at producing a "fully finished" genome sequence (the expected end-product for representative or important genomes), WG-SGS is also capable of producing three intermediate products: a genome survey sequence, targeted genome sequences, and a working draft

Tab. 2 Examples of sequence assembly software packages.

Software package	*Author's name*	*Institute*
PhredPhrap	Phil Green, et al.	University of Washington
CAP	Xiaoqiu Huang[1], Anup Madan[2]	[1]Michigan Technological University, [2]University of Washington
Atlas	Paul Havlak, et al.	Human Genome Sequencing Center
TIGR Assembler	G. Sutton, et al.	The Institute for Genomic Research
RePS	Jun Wang, et al.	Beijing Genomics Institute
ARACHNE	Serafim Batzoglou, et al.	Whitehead Institute/MIT
Celera Assembler	Eugene W. Myers, et al.	Celera Genomics Inc.
Phusion Assembler	James C. Mullikin, Zemin Ning	The Wellcome Trust Sanger Institute
PCAP	Xiaoqiu Huang, et al.	Iowa State University

of the whole genome. Each serves a special purpose and is vital for predicting the outcomes of a project in many circumstances. For instance, the genome survey sequence that covers less than 1x of the genome could provide information about the repeat contents and gene structures. It can also help in gathering polymorphisms when DNA sources from several closely related breeds are compared with a complete or nearly complete reference genome. A noteworthy example is the chicken polymorphism study where DNA sequences from three major domesticated breeds with nearly 1x coverage collectively are able to offer over three million SNPs (single nucleotide polymorphisms). One apparent application for targeted genome sequences is that the sequencing reads from an ongoing project can provide markers for identifying large-insert clones for the target regions before pictures of a genome or a locus become clear. A genome working draft (about 6x coverage of a genome) is the most popular product of many genome projects. It often covers 95 to 99% of the functional content of a genome and also anchors genes and their regulatory regions physically on the chromosomes when physical and genetic maps of the organism become available. Since finishing a genome sequence to a reasonable contiguity is difficult and expensive, sometimes impossible, not all genomes are expected to be finished. In fact, in many cases a finishing standard is debatable and occasionally highly controversial. Therefore, genome working draft sequences of many large eukaryotic organisms are expected to stay as such for a long time, though the community standard for finished prokaryotic genomes remains a gap-free sequence.

When zeroing in on the structural dynamics of large eukaryotic genomes, one will soon realize that there are numerous differences among genomes of taxonomically distant species to be considered before a genome-sequencing project is proposed. The most notable differences are between animal and plant genomes. Among animal genomes, genes spread over almost the entire length of a genome and transposable elements are inserted into introns, enlarging the genes over the evolutionary timescale. Most of the animal genes have more than a few alternatively spliced variants. The conjecture is that animal RNA splicing (spliceosomal) machinery is better established and complex enough to be capable of handling large introns ranging from a few tens of Kb to hundreds of Kb. Plant genomes are different from animal genomes in these same two areas. First, plant genes remain small in size owing to slower growth of the intron size, and transposable elements are mostly inserted into intergenic spaces. Transposable elements jump onto each other, forming gene-free clusters in high densities between gene "islands." As plant genomes evolve, only repetitive contents grow thus enlarging repeat clusters and not the gene islands. This is the case with wheat (1700 Mb) and corn (1200 Mb) as opposed to rice, which has a 500-Mb genome, nearly half of it as repeats of different kinds. Second, alternative splicing among plant genes is not as prevalent as it is among animal genes, perhaps as a result of less capable splicing machineries. Figure 4 illustrates the different genome structural models.

There are some minor issues relating to insect genomes wherein the heterochromatic contents are high. Insect genomes in sizes of hundreds of Mb seem not much different from other animal genomes of different taxonomic groups, but the basic organization of larger ones in the order of thousands of Mb are unknown so far.

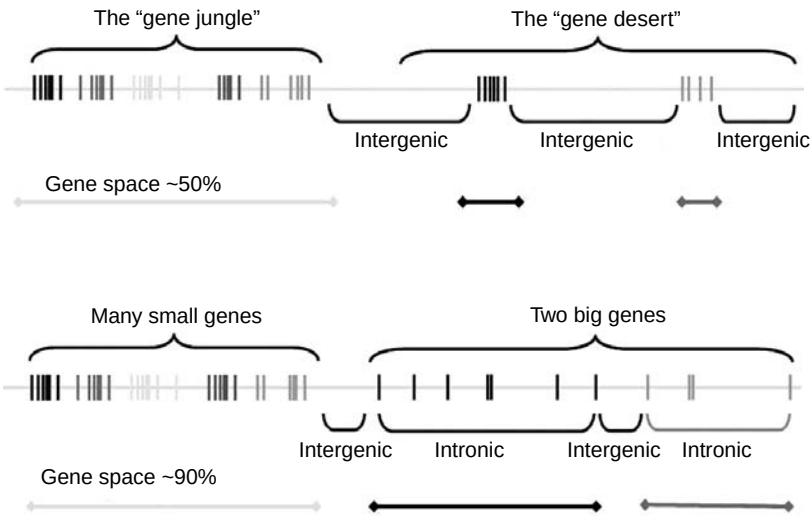

Fig. 4 A schematic illustration explaining major differences in basic genome structures between animals (represented by human) and plants (represented by rice). Genes are depicted as vertical bars in different shades. Gene spaces (diamond-headed lines) are estimates based on real experimental data. Spaces (horizontal brackets) between and within genes are defined as intergenic and intronic sequences, respectively. ''Gene jungles'' and ''gene deserts'' exist in plant genomes but are not known in animal genomes. In animal genomes, most of the chromosomal spaces are occupied by genes at different densities. Gene-poor regions most probably contain large genes or large introns.

Protists have the most diverged genomes studied to date and their genome sizes vary from a few tens of Mb to a few thousand Mb (Fig. 5). The GC-content of these organisms varies dramatically; the most extreme case was found in the malaria parasite, *P. falciparum*, which has a GC-content of 18%. As more genomes of these primitive but unusual organisms are sequenced, we will know better about their genome organizations.

2 Essentials for SGS

Regardless of whether it is associated with WG or CBC, SGS has five irreducible steps (Fig. 6): clone library construction, clone sequencing, sequence assembly, sequence annotation, and data analysis. The rest are either managerial (such as data handling, integration, and release) or political (such as cost accounting and choice of strategies) issues. The first two steps are apparently ''wet-bench'' related (other than data and database management) and the others are mostly bioinformatics, except a hidden ''time bomb'' in sequence assembly – sequence finishing, where close coordination between the two seemingly distinct disciplines is of the essence for completing a project successfully. The wet-bench operators must make sure that there are no obvious technical flaws in the raw data and that the assembled sequence is in agreement with preexisting data and knowledge, and the bioinformatics

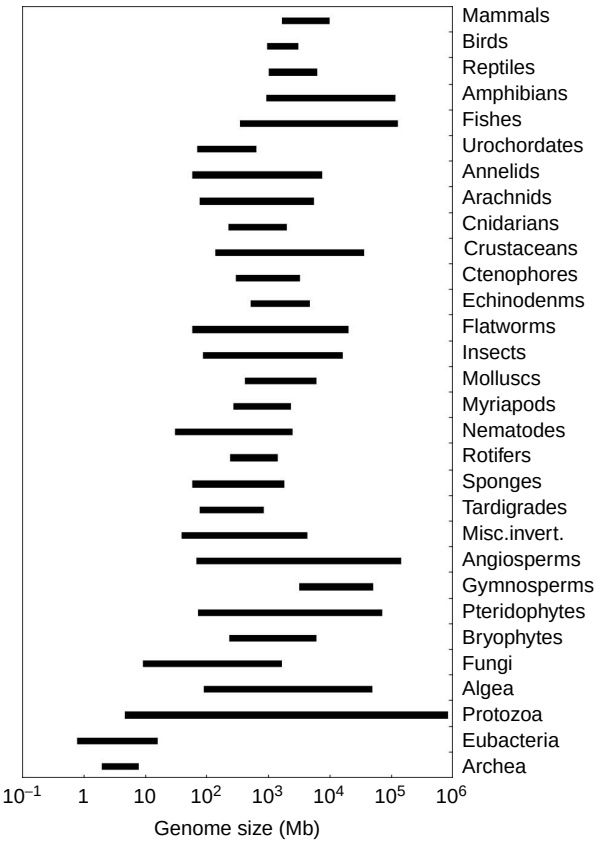

Fig. 5 Genome size distribution from organisms of selected taxonomic groups. The genome size is converted to nucleotide lengths from c-values. Horizontal bars mark the range of genome sizes for a given group. Polyploidy is not factored in to normalize the data, and so that the broader ranges in lower vertebrates, angiosperms, and protists are largely due to this effect. Data are limited to experimentally determined genomes, available from the *D*atabase of *G*enome *S*izes (DOGS).

operators must tell their counterparts if something seems wrong, or when certain problems are not solvable by computers alone. Technical details are useful but principles are extremely important.

2.1 Unbiased Clone Coverage

The essence of achieving unbiased clone coverage lies in making the clone libraries properly, avoiding possible negative effects. In a large SGS project (say for a 500-Mb genome and 6x genome coverage that is calculated as genome size divided by average read length or 6x 500 000 Kb/0.5 Kb = 6 million sequence reads), the first 4x sequence coverage is to cover the target sequence evenly and sufficiently. In this regard, the insert size of clone libraries should be small, usually approximately 1.5 Kb or twice the size

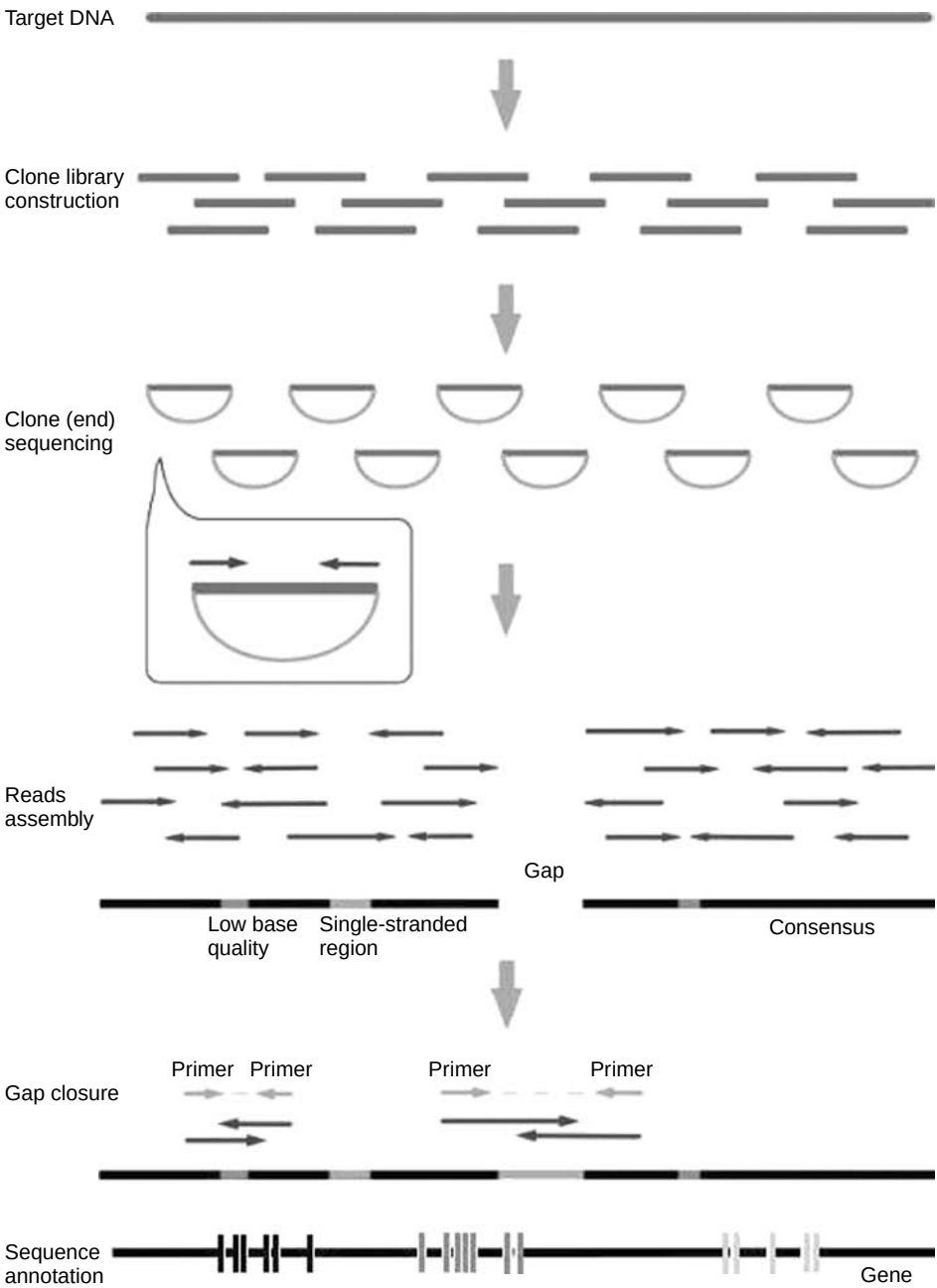

Fig. 6 A schematic illustration of the SGS strategy. Starting from a target DNA, the strategy contains five essential steps: clone library construction, clone (end) sequencing primed by universal primers from vector-borne sequences, reads assembly, gap closure, and sequence annotation. The inset illustrates that the sequencing is done from clone ends with universal primers. Detailed processes are described in the text.

of an average read length. The merits of smaller inserts are multifold. They pose fewer problems for cloning efficiency, randomness, DNA preparation, and sequence assembly. It is also easier to achieve uniformity in size fractionation since DNA fragments of such a size are readily separated in agarose gel electrophoresis. The next 2x is for accomplishing contiguity, and so the insert size should vary according to the repeat contents, including types, ages (in an evolution sense), cluster sizes, and whether there are active transposons that often litter nearly identical repetitive sequences in the genome. When numerous, they create real problems for sequence assembly. The choices of insert size and cloning systems are abundant, anywhere from 5 to 10 Kb for plasmids, from 20 to 30 Kb for mini-BACs, and from 35 to 40 Kb for cosmids or fosmids. BACs are always end-sequenced because in all cases BAC libraries are usually made for building physical maps and targeted gene sequencing, and often constructed before a sequencing project is launched. The only wet-bench concern is that sometimes DNA of large-insert clones is not easily prepared in the same way as small-insert clones and is a poorer substrate for standard sequencing reactions. A subtle computational concern is the "interleaving" problem, where repeat clusters are a few-fold smaller than the size of a large-insert clone and longer than a few sequence read lengths, such that sequence contigs between the two end-sequences could be lost (1.2 and 2.3).

Inadequate clone coverage always results in both sequence and clone gaps. Depending on the compositional dynamics of the genome and the data quality (mostly the average of sequence read lengths), sufficient coverage ranges from 6x to 8x. More often than not, a WG-SGS stops at 4x when contiguity is not a major concern. The Lander–Waterman curves are commonly used for estimating such coverage. The growth of number of contigs can be followed closely while the project is ongoing. The only variable for fitting the curve is the average read length when a genome size is known. The read length can be estimated in several ways. The most popular one is to use the Q values calculated by Phred. The value Q20 cut-off (a base error rate of 1 : 100) is adequate for this purpose although hard-nosed sequencing operators sometimes use Q30 (a base error rate of 1 : 1000) for assembly quality measurement. The reads in a length shorter than 200 bp should not be used for the measurement but they can be incorporated later into the final assembling process. Another way to estimate read length is to calculate the fraction that contributes to the consensus sequence. No matter what is used for the read length estimates, more sequence coverage (redundancy) is always a faithful friend during a project that is pressed for time. Sequence read coverage could be pushed up to 10x for smaller projects, such as a bacterial genome. After redundancy is maximized, anomalies should be carefully inspected in an assembly, not limited to assembled contigs and remaining gaps, physical or sequence, but also leftover singlets (stand-alone sequencing read) or singletons (short contigs that are single-read length and assembled by a coverage lower than average). Finishing strategies should be plotted with special attention to the distribution of physical gaps and the nature of sequence gaps.

2.2 Sequence Quality Control

Poor sequence quality that can be improved experimentally usually comes from

poor DNA preparation, since reagents and equipments are all provided by a supply company and process-optimized. Average read length at Q20 is a superior estimator of daily assessment of data quality, and feedbacks from the analysis to the wet-bench operation are necessary. Read length can be improved in several ways. In addition to better DNA preparation protocols, these include prolonging the separation (run) time, adjusting template concentration, and optimizing sequencing reaction conditions. These should be done and managed separately in a small R&D group, rather than constantly revising pipeline protocols. Poor signal-to-noise ratio is also attributable to poor DNA preparation in most cases, although malfunction of the process prior to DNA preparation may also be to blame, such as poor viability or growth of bacterial cultures. System contamination and malfunction of sequencing reactions can also result in poor signal-to-noise ratios.

2.3 Repetitive Contents

All genomes seem to have repetitive sequences; some are simple and others complex. Repetitive sequences can be defined in two ways: mathematically (mathematically defined-repeat or MDR) and biologically (biologically defined-repeat or BDR). When a sequence of a given length is seen in a genome more than once, it is mathematically defined as a repeat. The only variable is length that depends on the complexity or the size of a genome (assuming a random model for four nucleotides). BDRs are functionally defined and usually refer to *sequence-recognizable transposable elements* that are prevalent in genomes of all life forms. BDRs are usually identified with a software tool (such as RepeatMasker) composed of a database and a search engine. Since BDRs have to be identified by sequence homology, there are several problems in classifying them. First, the discovery of BDRs is very much limited by prior knowledge of their existence although fragments of the repeats are easily defined as MDRs. Second, it is difficult to classify BDRs because they do diverge over time according to a neutral rate of sequence mutation and some of the transposable elements are known to be old, such as LINE (long-interspersed-nuclear-element) in vertebrate genomes. Third, BDRs are often fragmented because of the dynamics of DNA sequencing of a genome (such as mutation and recombination) and the biological nature of various transposons (such as active and nested insertion). Therefore, parameters of a search engine can be adjusted to maximize or to ignore certain diverged BDRs, but to find them all is an unattainable goal. The quantitative relationship among MDRs, BDRs and apparent BDRs (BDR's) is depicted in Fig. 7.

Two basic parameters are always a concern for repetitive sequence analysis: the length of repeats and their complexity in a given genome. There are three additional lengths to be aware of in the analysis: those of a read, a clone, and a targeted DNA fragment (such as a chromosome or a genome). When two sequence reads are nearly identical, their positions in the target sequence become ambiguous, especially when sequencing reads are highly redundant for covering the target. In this sense, the smaller the target size, the easier the assembly process. However, when complexity of a target DNA and density of repeats are both high, it poses a serious problem that often results in sequence gaps in the target sequence. In practice, the

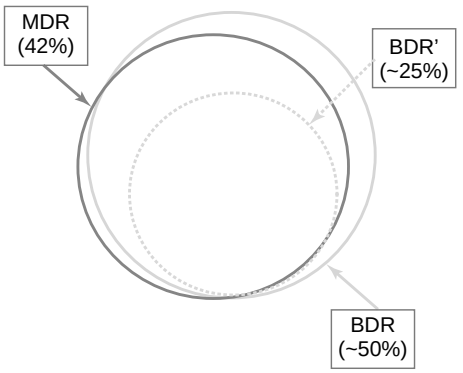

Fig. 7 Distributions of MDRs and BDRs in the rice genome. MDRs can be defined precisely (solid line) by counting exactly repeated elements in a genome. A portion of BDRs (broken line) is defined by degrees of homology (>50%) to known transposon families. The rest is too divergent over the evolutionary timescale to be readily identified but they are clearly present in the genome (gray) and can be estimated through homology analysis with low stringency.

longer the read length (assuming high quality), the less an assembly suffers with sequence gaps. Subclones of SGS project with slightly larger inserts will sometimes help in eliminating sequence gaps but cloning efficiency, size variation due to limited separation power of agarose-based electrophoresis, and other technical details will limit the increase of insert sizes.

2.4 Physical and Sequence Gaps

There are essentially two types of gaps encountered in sequencing a genome. When there is no clone available to bridge two contigs in any libraries for mapping and sequencing, a physical gap is inevitable; it is thus also called a *clone gap*. It is difficult to close such gaps because technology to carry out targeted cloning directly from isolated DNA of a large genome has yet to be developed. It is, however, possible to measure the lengths of gaps by using standard Southern techniques. The reasons for this type of gaps are often complex, but the gaps are usually the results of either inadequate coverage of the library or poor clonability of the missing DNA. Employing directed PCR amplification, adding coverage to the existing library and changing the cloning system, and varying the insert size are all useful remedies, coming from the wet-bench part of a sequence operation. The other type is sequence-related, in which a sequence assembly remains fragmented. In such cases, computer-aided analysis is required to provide clues for the wet-bench workers after they promise to deliver data quality, and have clones of various types in hands. If the nature of the gaps (oftentimes they are either nasty repeats in eukaryotic genomes or toxic genes in microbial genomes) is worked out, they can be left alone after some measurements on the gap sizes. In any case, a finishing plan should be drawn to handle the two types of gaps before they become a problem for sequence analysis, since the more gaps a project has, the more contigs have to be ordered. In a WG-SGS project, at least some effort has to be devoted to sequence finishing because genes or parts of genes may get lost in a massive number of contigs when sequence coverage is low (such as 4x).

2.5 Validation of Sequence Assembly

Multiple estimators should be used to validate a sequence assembly, because oftentimes estimators are biased because

of differences in community interests, emphasizing either functional or targeting aspects preferentially. The available resources are (1) full-length cDNAs and EST clusters (Unigenes); (2) physical maps and their markers (STSs); (3) genetic maps and markers; and (4) close-related genome sequences.

There are three essential parameters to be measured in validating an assembly. The first is average coverage, which can be calculated in a straightforward way such as a percentage of hits of the markers and sequences aligned to the target assembly, although the definition of a hit is always subjective because of the subtle details involved, such as real sequence polymorphisms. The second parameter is to determine how complete those matches are from a study of the markers and sequences to the targeted assembly. The exact matching sequence has to be statistically presented precisely down to each basepair. None of the above parameters would tell much about the correctness of the assembly in terms of long-range contiguity and local misassemblies except in the case of using restriction-digest (fingerprinting)-based physical maps in which the digested fragments are precisely size-measured and even ordered. A retrospective comparison between the sequence assembly and its corresponding physical maps, from a BAC clone or genome-wide, is always an excellent estimator for sequence validation. It is, however, difficult to assess contiguity and accuracy of an assembly when resources for estimation are sparse.

2.6 Integration of Genomic Information

Together with the rapid increase in genome sequences, data integration at multiple levels from genomics-scale studies is of the essence for understanding the basic genomic information and functional annotation of sequenced genomes. In a vertical integration scheme, the top layer consists of genes (their structural and regulatory elements) and other basic genomic elements (repeats, telomeres, and centromeres). The intermediate layer includes gene expression information, from mRNA to protein. The bottom layer is the relationship of a gene product, RNA or protein, and its interacting partners in the context of location (defined by spatial connection and timing), function (defined chemically or physiologically), and mechanisms (defined by specific roles of each of the components and their cooperative functions). Genetic information, including markers and sequence polymorphisms, is usually gene-centric or marker-centric and must be integrated into the top layer as part of the physical map before correlations can be made precisely between genotypes and phenotypes. Database systems are usually built not only to manage and search the raw data. They are also integrated with software packages for visualizing, editing, annotating, and analyzing the data. An integration scheme for rice genomic data is illustrated in Fig. 8.

3 Technical and Experimental Basics

SGS as an operational entity, especially in LSS, is highly specialized, with more than 10 years of experience in process improvement and automation. It is relatively mature in the sense that its basic process cannot change overnight. For instance, sample preparations and all subsequent liquid handling procedures are based on microtiter plates with 96 and 384 wells. The

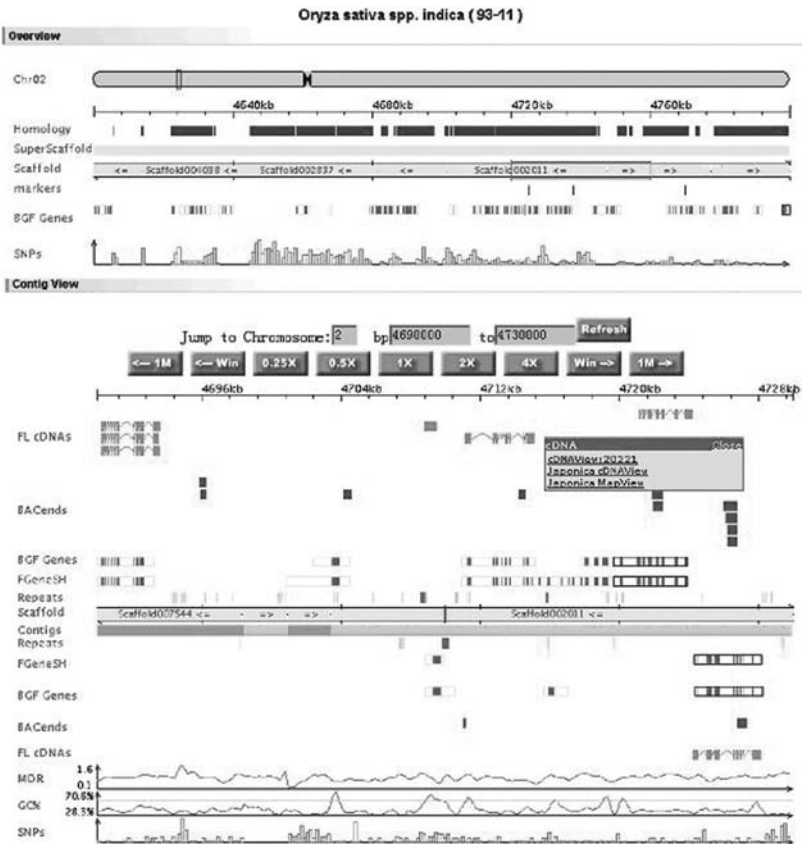

Fig. 8 An example of sequence annotation schemes from the rice genome. Gene-finding results from two gene- finding software packages, FgeneSH and BGF are shown in the annotation.

sequence technology may change some day in the future, but sequencing and cloning systems will live on.

3.1 Genomic DNA Libraries

The preparation of insert DNA is one of the crucial steps in making SGS libraries. For constructing small-insert libraries or subcloning of large-insert clones, physical power, rather than enzymes, is usually preferred for shearing DNA into fragments. There are two ways of doing this – sonication and nebulization. Commercial instruments are readily available and a skillful technician can easily master the technique even for large-scale operations. Partial restriction-digest is not commonly used for preparing insert DNA, especially in large batches. It is, however, very useful when clone libraries with insert sizes ranging from 5 to 20 Kb are constructed for achieving better contiguity.

Size-selection of insert DNA is another important step in securing high-quality libraries, but this is often neglected. Better performance in size-selection results in

not only narrower insert size variation but also lower numbers of "empty" (a consequence of too small an insert) clones, thus improving the standards of library quality control. A more uniform insert size is very helpful in sequence assembly when information on clone ends is used to join contigs (3.3). To achieve a better size-selection, two key factors should be kept in mind: electrophoretic conditions, and appropriate loadings (micrograms of DNA per well in an agarose gel). Rapid electrophoresis, by increasing either currents or voltages of the electric field, reduces separation power, while overloading damages the filtration power of the matrix (agarose). For this simple but key step, an optimized system should be carefully established.

3.1.1 Source DNA

One can never over-emphasize the importance of selecting source DNA for a genome project, regardless of whether WG or CBC was used as the downstream strategy. The potential problem lies in the genetic heterogeneity of DNA samples. It is only a minor problem in large organisms because they have sufficient tissue or body mass, but in small organisms of which tissue samples from multiple individuals are used for extracting genomic DNA, it can be a major problem. In the case of sequencing small-sized genomes, such as those of bacteria or archaea, DNA samples should be prepared starting from single colonies. When different batches of DNA preparations are unavoidable, special care should be given to watch for transposon insertions and segmental duplications. Such a situation pleads for restriction fingerprinting analysis (using pulse-field gel electrophoresis if necessary) to make sure of sequence integrity among the samples. For small multi-cellular organisms such as worms, insects, and small plants, it is crucial to prepare genomic DNA for library construction from a homogenous (inbred) population. It is not unheard of that a WG-SGS project might have to be abandoned because of the impossibility of attaining assembling sequence reads at an identity of 95 to 99%. When a non-laboratory strain has to be used for a project, physical maps should be built with large-insert clones, from which a set of MTP clones is fed into sequence pipelines. It thus becomes a CBC-SGS project. Extreme caution should be exercised lest other invisibly small organisms contaminate the source DNA isolated from a wild species, especially those from open waters, such as large protists, small marine animals (mollusks, sponges, and corals), and miniature plants.

The final concern, though not the least, is about sex chromosomes. In a sexual organism, sex chromosomes are often unevenly distributed between males and females. Other than human males who have a miniature sex chromosome, Y, sex chromosomes in other sexual species are sometimes comparable in size. However, they may not have the same distribution as human sex chromosomes, where XY is male and XX is female. In birds, the male is homozygous, ZZ, and the female is heterozygous, ZW. Oddities are often observed in life forms, with even temperature and crowdedness being determinants for sex and genetic makeup. Whenever possible, a haploid genome provides much better source DNA for WG-SGS.

3.1.2 Quality Assessment

There are a few parameters to be measured experimentally to assess the quality of an SGS clone library: the average insert size, the rate of non-informative clones

(small inserts and background vector self-ligation), and clone recovery rate. The first two parameters can be measured by colony-PCR (using a small number of bacterial cells as PCR templates after heat-denaturation) or by sampling isolated plasmids. To avoid false-negative results of PCR, tested samples should be in duplicates with appropriate controls. The last parameter is a result of several potential tribulations, including mixed or defective clones, and non-optimal growth conditions. It has to be tested by sequence sampling. For a large-scale operation, it might be more economical to just test all the parameters by sequencing random samples.

3.2 Sequencing Data Acquisition

The sequencing data acquisition process is the engine of SGS. It is calibrated in a very subjective and precise way as throughput and cost. The throughput has shaped the way LSS is organized globally, with only a dozen or so academic institutions being focused on such an operation. The cost structure of DNA sequencing is composed of capital equipment, reagent/supply, and labor. At least two-thirds of the total cost is on account of reagents and supplies, so there is still plenty of room for cost reduction. Another important factor is the management of the two major components of sequencing business: one, the experimental process, and the other, computation. Only a short feedback loop between these two components can ensure operational efficiency.

3.2.1 WG-SGS and CBC-SGC

Operationally, there are major differences between CBC-SGS and WG-SGS. First, CBC-SGS is a map-directed operation and it relies heavily on the completion of a physical map containing targeted regions. In most of the multi-institutional operations, work is allocated according to chromosomal regions (in the case of HGP) or simply individual chromosomes (in the case of the international rice genome consortium). A potential danger is that loosely affiliated members may all have their own priorities and agendas, and that their schedules for carrying the project to the finishing line may vary accordingly. As a result, the satellite projects have to be finished and published separately; such an operation not only reduces the scientific impact of a project but also faces the risk of being "scooped." WG-SGS is in a much better position to succeed, since data are to be acquired by the same group or groups with comparable operating styles. Second, CBC-SGS requires an intermediate management layer to handle maps and clones, to construct more shotgun libraries, and to assemble each clone, each of which requires a different sort of attention to complete. In other words, wherever genomes with high repetitive sequence contents are involved, they may never be finished to a similar standard, and therefore CBC-SGS could face a disastrous situation. Large plant genomes are illustrative of such cases, where the repeat contents exceed gene contents. For smaller animal genomes, however, the problem is less hazardous. Third, WG-SGS poses major challenges for sequence assembly when the target genomes are large and contain high repetitive content, but assembling sequences for large-insert clones in typical CBC-SGS are not. In CBC-SGS, sequence contigs are ordered according to clone orientation in physical maps, whereas in WG-SGS, contigs can be ordered within scaffolds or according to physical markers. In some cases of WG-SGS, contig and scaffolds may not be ordered until

both physical and genetic maps become available. WG-SGS data, however, provide a foundation for retrospective physical (gene) and genetic mapping efforts.

3.2.2 Directed Cloning and Random SGS

During the early phase of LSS, there were two essential strategies developed to evaluate the cost and efficiency for an operational process. The first system utilized a random shotgun protocol and M13 phage (single-stranded DNA) cloning system. DNA preparations for M13 cloning, when carried out manually, are usually easy and clean, giving rise to high-quality sequences. It is unfortunate that only one end of the single-stranded DNA can be primed for sequencing. The method was soon replaced because the success of a complex experimental protocol depends on the collective merits of all the procedures. The second system is the plasmid cloning system (double-stranded DNA). Two simple advantages of using plasmids are: sequences from both ends of the insert are obtainable from a single DNA preparation, and the length of DNA inserts is also a piece of powerful information for sequence assembly when their sizes are prepared relatively uniformly. Although other attempts were made in the early days of LSS, such as transposon-directed plasmid sequencing, the current practice is focused on a plasmid-based random cloning system, oftentimes coupled with controlled insert sizes.

3.2.3 Process Control

SGS remains merely a process rather than an approach or strategy when it is not evaluated and compared with other sequencing techniques. It is quite complex, involving integration of sequence acquisition, data processing, and analysis. As a process, it has to be versatile because relevant technologies are still underdeveloped and any part of it can be changed over a short period of time. For instance, capillary sequencers now dominate all large-scale sequencing operations, although the slab gel format may continue for special purposes such as for small-scale operations and for achieving longer read lengths. Of course, the most practical consideration in building a sequencing operation has to be capital investment. Any such investment on a process-unstable technology should be carefully calculated. With an estimated annual budget exceeding $1 billion US dollars globally for large-scale genome sequencing, more than half of which comes from public funding agencies worldwide, LSS is undoubtedly going to stay with us for some time.

SGS creates major managerial challenges for process control. On the wet-bench side, the managers are Ph.D. scientists and skillful technical staff, who are able to master cloning techniques and to have troubleshooting capabilities for all protocols used; each has precise quality-control parameters that are under close surveillance with fast feedback loops. For instance, whether or not a clone library is going into the pipeline depends on its "pass rate," in which contaminations from *E. coli* host and vectors, average lengths of the insert, and identities of the target sequences (such as signature repeats) are all factored in. However, things can go wrong in between too, such as failures in DNA preparation and sequencing reactions, equipment-related errors, and even data tracking mistakes. In most of the cases, a custom-designed software system has to be implemented for a process-controlled operation. The challenge in building a robust system is to decide on the extent

of automation *versus* human intervention, where fully automated systems are impossible to build. On the computing side, things are even worse. Data of different kinds are voluminous and instantly accumulating, including raw trace files (calling for data management and database building) from different projects (calling for project management when target DNAs are from different sources), quality assessment (involving quality management to distinguish usable and unusable data), assembled intermediates (requiring process management dealing with reads, contigs, and scaffolds), and annotations (calling for biological research to interrelate the results by searching against all available data); and even these data are always incomplete. Because automation is being continually expanded, other groups of people have also entered the process, namely, engineers and manufacturers. Three indispensable groups with different cultures have to work together on the same process. When physical mapping is involved, another group of people would have to join this outsized team.

A process should be evaluated by its throughput and efficiency. There are many reasons for there being only a few large genome centers to do the basic jobs, but management and experience matter the most. In addition, scales, automation, and robust protocols are all major concerns for a sizable sequencing operation. As the size of the target genome gets bigger, the scale of an SGS operation goes up. Currently, projects of the size of mammalian genomes (3 billion basepairs) are usually handled by a joint effort of more than one genome center. There are only half a dozen genome centers in the world that are able to handle the sequencing of large genomes independently.

3.3 Sequence Assembly

Sequence assembly is the most time-consuming and computing-intensive part of SGS, especially for WG-SGS. Software tools for sequence assembly, and also for gene-finding and sequence annotation, have to be developed and implemented before a real project starts. In addition, the skills and experience of the operating teams are also important since the entire process is very complex and demanding, as is described below.

3.3.1 An Overview of General Procedures

There are distinct steps in the procedure for sequence assembly. First, repetitive sequences need to be defined, classified, and masked for each sequence read before being placed into an assembly pipeline. Second, low-quality reads (usually at the ends of a sequence read) and contaminated reads (from library hosts and other projects) are removed through different computational filters. Third, computer-assembled sequence contigs are evaluated based on mathematic models (such as the Lander-Waterman model) and examined manually after each finishing protocol is applied (3.4). Fourth, contigs are further linked into scaffolds based on additional information. Finally, contigs and scaffolds are ordered and oriented according to chromosomal organizations based on physical markers or other experimental data, such as those from fluorescent in-situ hybridization (FISH).

3.3.2 Classification of Repetitive Sequences

Almost all eukaryotic genomes contain significant portions of complex repetitive sequences. Despite the fact that these repeats are mostly definable by their

repetitive natures, being created over evolutionary timescales, they have not only jumped in and out of genomes at different points of time (lineage-specific) but are also being mutated and recombined among themselves, making their precise identification an intricate task. Repetitive sequences are often classified into two basic categories, simple and complex. Simple repeats are easily definable and of less concern for sequence assembly, because they usually constitute only a small percentage of the genome sequences in vertebrates and seed plants. They do occasionally pose minor problems for sequencing (3.4.3) and may become a major nuisance for insect genomes, as they could account for a major part of heterochromatins. Complex repeats are mostly defined biologically by detailed evolutionary analyses and by building sequence libraries for featured annotations of discrete taxonomic groups (2.3). Another class of repetitive sequences is the so-called *low-complexity repeats* (LCR). The origin of LCRs is recent segmental duplications, and they may cause significant problems when concentrated in a proximal region of a chromosome. However, they are not very frequent, and experienced researchers can deal with them quite easily with the aid of denser physical markers and unique sequence signatures. The final class of repeats concerns special regions of chromosomes, such as centromeres, telomeres, and their nearby regions, such as subtelomeric regions. Different research groups usually handle these repeats according to their own specific interests.

3.3.3 Software Packages for Sequence Assembly

Sequence information is acquired through an electrophoresis-based instrument that is capable of resolving differences among DNA fragments at single base level. An important parameter is the quality of each resolved base, which is measured as error probability or error rate. In the most popular software package for sequence assembly, Phred-Phrap-Consed developed by Phil Green and his colleagues, Phred is a base caller that is capable of assigning an error rate to each sequenced base from a machine-processed trace file. Using stringent quality measurement, all eligible bases are used for assembly. Phrap carries these quality scores when it assembles consensus sequences from a set of sequencing reads. Consed is essentially a viewer that allows users to examine the results and edit the sequences. Nearly all sequence assemblers have capabilities similar to the Phred-Phrap-Consed package, but some have been outfitted with more user-friendly features. For instance, CAP, another sequence assembly program developed by Xiaoqiu Huang and Anup Madan, is able to clip 5′ and 3′ low-quality regions and uses forward–reverse constraints to correct assembly errors. Phusion, an assembler developed at the Sanger Institute, is competent in assembling mammalian genomes, and was used in the mouse genome project with a WG-SGS dataset. Phusion incorporates end-pairing information and has an ability to group large numbers of sequence reads before performing a parallel assembly protocol. Some of the major software packages for sequence assembly are listed in Table 2.

3.3.4 Building Contigs and Scaffolds

The fundamental nature of sequence assembly software is to define identity among sequences of different reads and to detect their overlaps. Several software packages have been developed for DNA sequence assembly but by far the most

popular one is the Phred-Phrap-Consed package. For large WG-SGS, a special package has to be built in to handle additional information that is not dealt with by Phrap. For the purpose of illustrating the basic principles and methodology, RePS (an assembler, repeat-masked Phrap with scaffolding, developed at Beijing Genomics Institute) is taken as an example (Fig. 9).

The first step of a sequence assembly is to define "repeats," and these are the major obstacles whenever sequence comparison is the engine of an assembler. Repeats are usually defined mathematically as exact matching sequences of oligonucleotides (MDRs), as a series of "oligomers." The lengths are often defined empirically, such as 20mers for large eukaryotic genomes. Repeats are identified from each sequencing read and "masked" before being placed in an assembly process. Some of these reads may be completely masked so that they have little chance to contribute to intermediate assemblies, but they are factored in as anonymous coverage only in the final product. Phrap is often used as an assembly engine for most assemblers although other algorithms have been developed for general sequence comparison and assembly.

In nearly every LSS project, another powerful piece of information is also used, namely, paired sequence reads from the ends of the same clone (acquired from opposite directions), which allow users to order and orient additional contigs that are not obviously overlapping in sequences but definable by clone-length criteria. When the insert sizes of a clone library are not significantly deviated from a median (usually measured experimentally), a uniform size is assigned to all the clones sequenced from both directions. The insert sizes may vary among libraries used for a particular project but not within the libraries. As a result, the ends of a clone can be paired, though sequences of the middle portion of an insert may not be covered from both ends, and this information is utilized to "scaffold" neighboring contigs that are anchored by these end sequences. Figure 9 depicts the assembly process involving the two primary components of RePS: assembly of repeat-masked reads with Phrap, and scaffolding based on clone-end information.

The final process of an assembly is to close the gaps. At this point, two types of gaps are in evidence: artificially created ones due to repeat masking, and physical ones that occur statistically by chance. The former is referred as *repeat gaps* (seen within sequence reads) and the latter is called *Lander–Waterman gaps*. Repeat gaps can be filled after reads are assembled and Lander–Waterman gaps are to be closed experimentally because they often result from inadequate coverage or purely by chance. In practice, both gaps are usually smaller than the lengths of single reads and can be bridged by clone-end information, when paired ends of a clone happen to join their two neighboring contigs.

3.4 Sequence Finishing

Sequence finishing, a process for achieving better contiguity of a target sequence, is always a time-consuming process, and requires the collective experience in LSS built up over a decade. Efforts in automating such a process have had some successes, but have suffered because of the inadequacy of standards as well as funding. There is no question that certain important genomes have to be finished

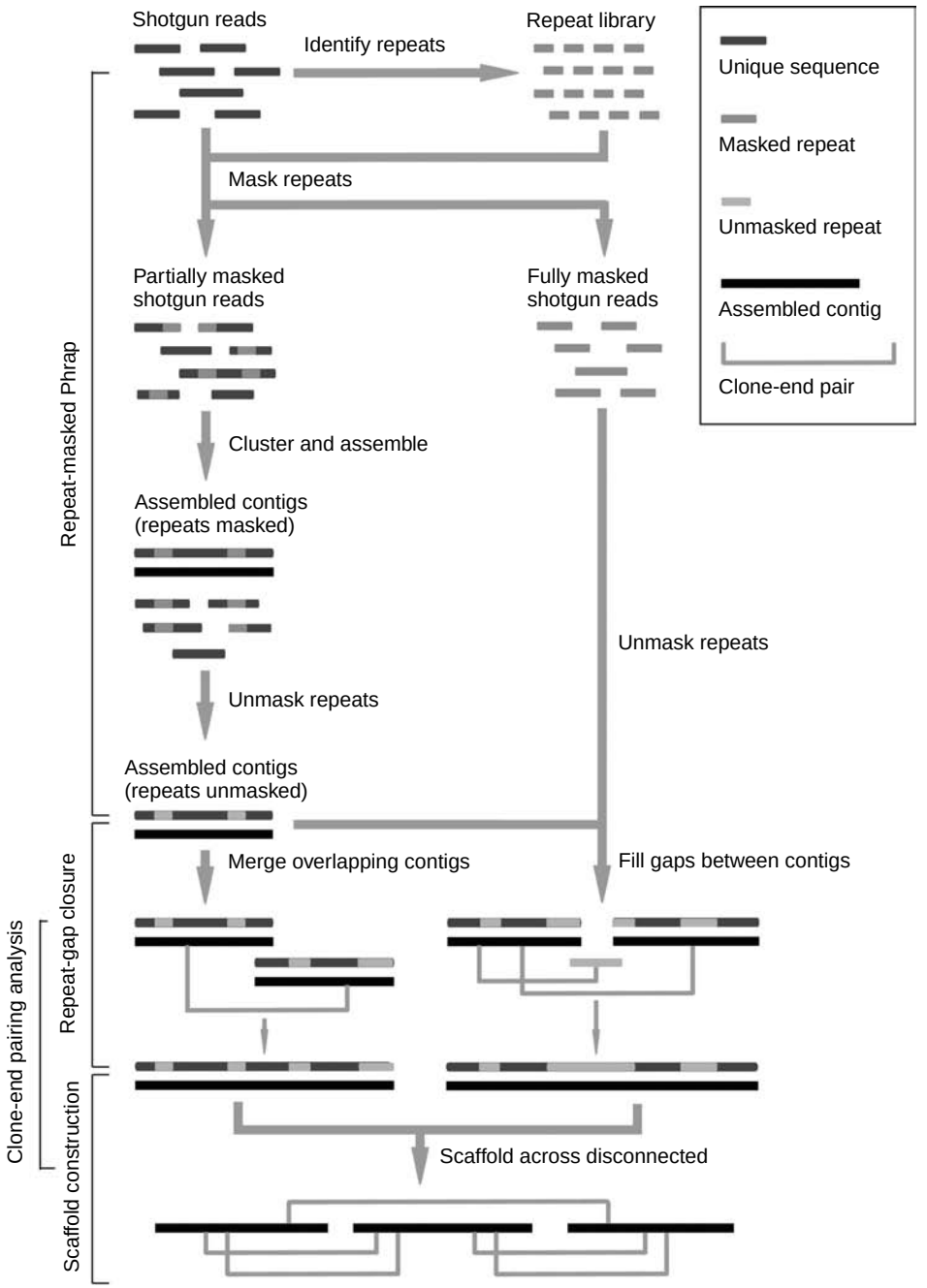

Fig. 9 Schematic flowchart shows the process of sequence assembly.

to a standard that is acceptable to the scientific community. Examples of such genomes are the human and laboratory mouse genomes, but for other genomes a working draft (6–8x coverage) as the final product of a genome project should be satisfactory. With either WG-SGS or CBC-SGS, the basic finishing process for sequence gaps is essentially the same. It includes: (1) improving quality in the weak regions (often defined as those covered by a single read generated from a single type of sequencing chemistry); (2) primer-walking on clones at contig ends; (3) PCR-sequencing to close gaps; and (4) checking consistency guided by physical maps. These activities are normally performed by a team of well-trained technicians and led by a highly experienced scientist with a Ph.D. degree.

3.4.1 Improving Sequence Quality

Before one looks into the relationship among sequence contigs, the first thing to do in the finishing stage is to analyze the weakly covered sequences (poor sequence quality and single-read coverage) that are often the reason for misassemblies and false joints between contigs. Poor reads assembled into contigs should be removed first and repeated experimentally with or without new DNA preparations. Plasmid clones (not M13 clones) can be sequenced from the opposite directions if the poor reads are due to the nature of DNA sequences (such as long simple repeat tracks or extreme GC-contents). This seemingly less impressive practice is in fact very powerful, and at times reduces the number of contigs by half in a project on bacterial genomes of a few Mb in size. The second major effort should be to figure out reasons why certain contigs stop growing; some may be obvious and even straightforward to determine, and others may not. For instance, when contigs repeatedly end with poor-quality reads, the most obvious solution is to add more reads to increase the total sequence coverage. When contigs end with large numbers of reads, higher than the estimated average coverage, and stop sharply with a pile of high-quality reads, the reasons are usually misassemblies due to repetitive sequences. In large genomes, low-copy repeats, which may be either large (such as recent segmental duplications) or small (such as recent transposon insertions and duplications), from time to time create serious headaches for sequence finishers and need to be spotted as early as possible.

3.4.2 Primer-walking and End-sequencing

All sequence contigs end for some particular reason. These reasons include, but are not limited to: (1) low-quality reads or failed sequence in one end of a clone that happens to be in a poorly covered region; (2) high-copy repeats that are too long to be linked by any clones in the assembly; (3) low-copy repeats that are assembled together; and (4) false joints that are made within the contigs. These reasons can be placed in one of two categories: repeat-related and quality related. Quality-related contig-breakers can be eliminated by primer-walking, in which specific primers are designed to improve the local sequence quality. When a PCR product is too long to be "walked" over (for example, if it is more than 3 Kb in size), a clone library can sometimes be made from the PCR product for a small SGS project. Repeat-related contig-breakers are, however, very hard to eliminate. Software tools are required to examine sequence contigs and their contents, and to relate repetitive sequences among contigs. In

the assembly software assembly software package, Phred-Phrap-Consed, there is a useful little tool, called *PhrapView*, allowing users to see distributions of repetitive sequences among contigs within a project. A general remedy for the repeat problem is to add some large-insert and end-sequenced clones into the assembly, such as BACs or cosmids/fosmids. The hope is that these long-range sequences can solve, in part, this type of assembly problem. When repeat-related problems are explicitly understood, editing of software parameters and read assemblies becomes an easy solution.

3.4.3 Closing Sequence Gaps

There are always mixed feelings about closing gaps. On the one hand, it is satisfactory to close some simple gaps, such as by PCR-based methods. On the other hand, when thinking about the nature of these gaps, one may wonder why we have to take so much trouble over closing them. This is because most of the gaps are caused by short (a couple of Kb or less), low-complexity (a few unique copies in the genome), and simple repeats (such as {AT}n and {CCT}n). Time and again, gaps are found due to degenerate simple repeats (such as degenerate repeat tracks of pentamers or hexamers, often classified as MDRs, 2.3) that are poorly maintained by the bacterial hosts. In other words, they are mostly physical gaps caused by cloning failures, and are especially present in animal genomes, such as those of human, mouse, and insects. Short BDRs of high density can also result in similar gaps, especially in plant genomes. When large repeat clusters are encountered, gap closure can be difficult and sometimes impossible. Extreme GC-contents, often caused by simple AT-rich or GC-rich repeats, can also result in gaps because they are tricky to sequence with a standard sequence reaction protocol. Both annealing and extension temperatures of a sequencing reaction have to be adjusted to repeat the experiments. In the genomes of grasses and warm-blooded vertebrates, a negative GC-content gradient along the transcription direction exists in a significant portion of their genes. This is the result of a mutation drive that is believed to originate from the transcription-coupled DNA repair mechanism. When the GC-content of these genes reaches 70%, the 5′-end is occasionally missing in the assembly. Though this is a particular case seen in sequencing the rice genome, the phenomenon may extend to other grass genomes (such as those of wheat and corn). In summary, there are no universal rules about how to close gaps, and what the sequence characteristics of the gaps are. It is, however, worth making an effort to close a few gaps just to know what is out there.

It is still debatable whether all gaps should be closed in a genome sequence and it is up to the research communities and their funding agencies to take a final decision. It is our opinion that, in a WG-SGS project, certain gaps should be closed and their sizes measured where feasible; these should include gaps that separate a coding sequence and that are shorter than a few Kb, within a PCR-reachable distance. On the other hand, when gaps are known to be the result of running against long-range repeats, they should not and cannot be closed with currently available techniques. On an experimental note, when closing gaps that are poorly covered by bacterial clones, one can use isolated YAC or genomic DNAs as PCR templates for gap sizing and filling since they may have resulted from DNA sequences that have poor clonability in bacterial hosts.

3.4.4 Physical Mapping and Clone Fingerprinting

Because both physical mapping and clone fingerprinting are very important techniques for SGS, almost all large-scale sequencing projects are initiated along with a parallel project to make large-insert clone libraries. Physical maps can be constructed by using three basic substrates: chromosomal segments (such as radiation hybrid mapping, in which overlapping chromosomal segments are carried by a cell line originated from different species); large-insert clones; and restriction fragments (or even DNA fragments). Two basic markers are used for physical mapping, namely, STSs, and restriction-digest fingerprints. The essence of physical mapping is to order these markers, clones, and DNA fragments. Large-insert clones are the most useful reagents for sequence-ready physical maps, especially BCAs and cosmids/fosmids.

In CBC-SGS, physical mapping becomes a prerequisite and a sequence-ready map is the starting point for large-scale SGS. Since BACs (100 Kb to 200 Kb) and cosmids (30 Kb to 40 Kb) and also fosmids and phages are better sequencing substrates than YACs, both directly (such as for end-sequencing) and indirectly (their subclones are the direct substrates), they are preferred for constructing a sequence-ready map. YACs are generally useful for long-range physical mapping, such as YAC-based STS mapping, but they have to be subcloned into cosmids or fosmids again in order to make a sequence-ready map. For WG-SGS, it is foreseeable that physical maps will be needed for hanging sequence contigs or scaffolds on chromosomes when projects on large genomes are carried over their SGS phase. With unlimited STS candidates in hands, it is quite feasible to construct a contig-based STS map retrospectively, even if the number of scaffolds may exceed a few hundreds. Another way to order sequence contigs is to use the FISH technique, in which representative large-insert clones can be placed and ordered over chromosomes experimentally.

Clone fingerprints are also useful in validating an assembly, especially for CBC-SGS. When large-insert clones are fingerprinted with high redundancy, each clone and also the genome-wide sequence quality can be systematically checked, aided by software tools such as FPC, which is an integrated program for assembling sequence-ready clones, developed at the Sanger Center.

3.5 Genome Annotation and Analysis

Genome annotation has now become a specialty in genome-sequencing projects. It gets more and more complex because data of different types are being accumulated exponentially and most of them are cross-referenced both vertically and horizontally. In a vertical sense, data of a single species have to be integrated from genome to transcriptome and proteome, as well as information from other levels of functional studies (such as metabolisms, networks, and interactions). In the future, all data that can be cross-referenced will be cross-referenced without any delay. In the horizontal sense, data from closely related species (within a species, genus, and even within the same phylum) have to be compared and analyzed with regard to the details relating to the genes and their functions. The norm is going to be changed over time until species representing all life forms of the extant taxa have been somewhat sequenced. This is not a dream but a future reality.

3.5.1 Compositional and Structural Dynamics

Genome sequences are dynamic entities, deriving from common ancestral genomes, and evolving to become more divergent, complex, while remaining related at different evolutionary scales. Eukaryotic genomes are known to undergo compositional transitions driven by mutation mechanisms that are by and large governed by various DNA polymerases involved in DNA replication and repair. Although eukaryotic genomes with extreme average GC-contents are rare, heterogeneous distributions of DNA composition are regularly observed. Figure 10 shows several examples of such a dynamic nature. Major compositional transitions have been discovered among genomes, between warm- and cold-blooded vertebrates, and between grasses (one of the major monocotyledonous plant groups) and the majority of dicotyledonous plants. Moreover, not only is such an increase in GC-content directional (Fig. 11), from 5′ end of the transcription unit toward its 3′ end, but it also involves significant portions of the genomes and highly expressed genes. The exact molecular mechanism has not yet been demonstrated experimentally, but it is speculated that it could be largely attributable to transcription-coupled DNA repairs. The relevance of genome compositional dynamics to SGS is multifold. Extremes of GC-content (lower than 20% and higher than 60%) often pose problems

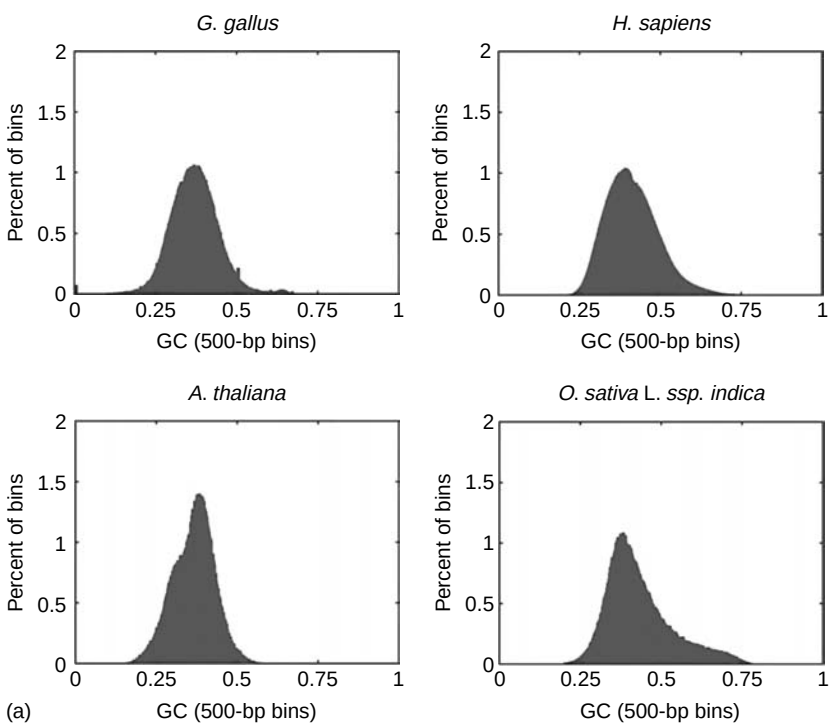

Fig. 10 Genomic GC-content distributions of selected genomes: *G. gallus*, *H. sapiens*, *A. thaliana*, *O. sativa*, *D. rerio*, *X. laevis*, *C. intestinalis*, and *C. elegans*. GC-contents are computed over a bin size of 500 bp.

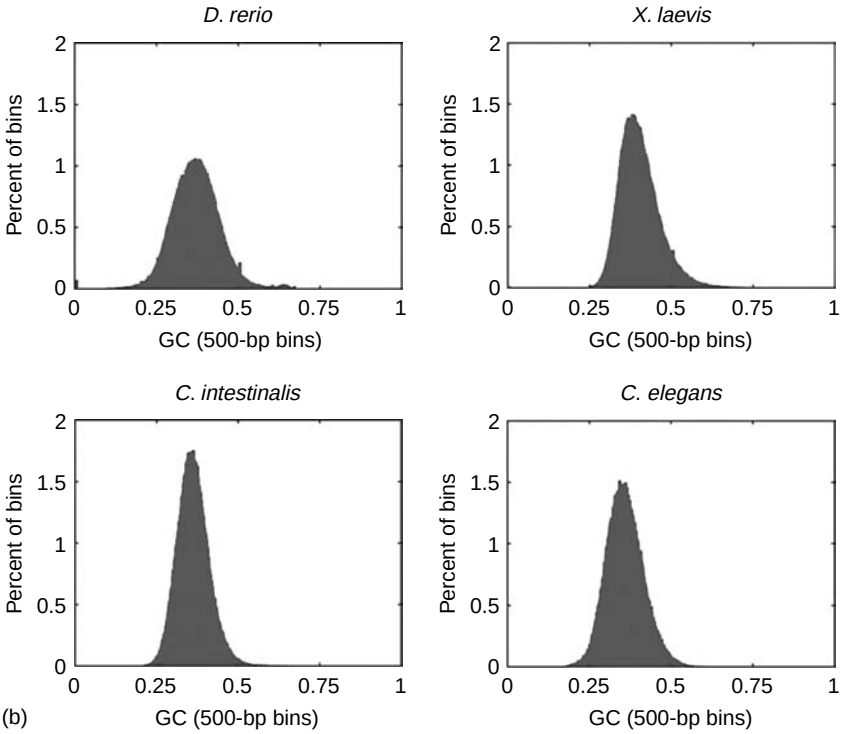

Fig. 10 (*Continued*)

in cloning and sequencing procedures. Sequence heterogeneity interferes with gene-finding algorithms, especially when training data sets are not readily available. It also biases sequence homology comparison, making genome annotation more difficult.

The structural dynamics of evolving genomes are studied at higher levels than single or a few dozen basepairs, generally involving large duplications and deletions, which could occur genome-wide and in the whole genome, in a chromosome segment, and at gene levels. Broadly speaking, structural dynamics within genes and gene families should also be dealt with in a similar way, such as alternative spliced variants and differential gene expression though transcription mechanisms. In-depth understanding of genome and gene structural dynamics is of importance in genome assembly and annotation.

3.5.2 Genome Annotation

The keys in genome annotation are gene-finding and information integration of experimental evidence and, to a lesser extent, repeat identification. For LSS, these are always handled with automated software tools and organized in databases. Table 3 lists some of the major sequence, genome, and annotation databases as well as their relevant information.

Since sequence data are being accumulated much faster than experimental evidence, development of *ab initio* gene-finding software becomes a major research activity for bioinformatics. Table 4

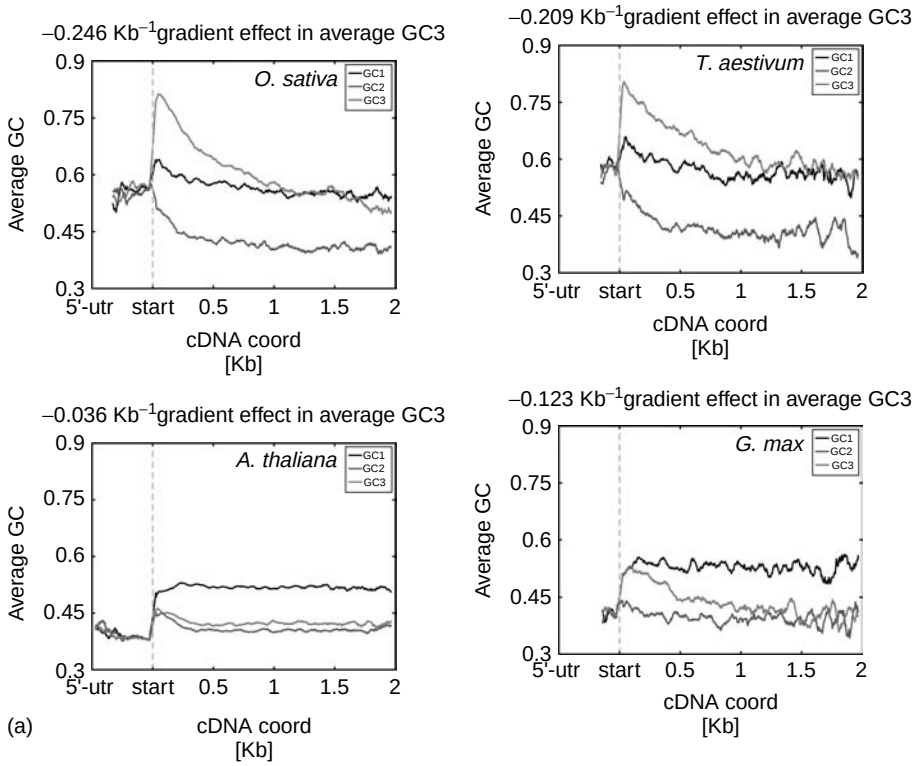

Fig. 11 GC-content increases in cDNAs of plants and animal genomes. GC1, GC2, and GC3 content are plotted as a function of cDNA position, relative to the start of the coding region, and averaged over all cDNAs with a 51-bp sliding window. Phase information is extended into the 5 -UTR. Eight examples are shown here: (a) from plants, *O. sativa*, *T. aetivum*, *A. thaliana* and *G. max*, and (b) from animals, *H. sapiens*, *G. gallus*, *D. rerio*, and *X. tropicalis*.

shows selected examples of such software packages, their methods, and immediate applications. Regardless of which mathematical model (Hidden Markov Models or neural network) is used and what parameters (sequence composition or signal to weight), a common misperception about gene finding is that a universal logarithm must be developed for all genomes. In addition to support from experimental evidence, two common factors that determine the performance of a gene-finding program are sensitivity (a measure of prediction that is equivalent to one minus the false-negative rate) and specificity (a measure of prediction that is equivalent to one minus the false-positive rate). These two parameters often obscure the possibility that there are classes of genes for which the performance is exceptionally better or worse, such as those in different size classes. It is necessary to consider how performance varies as a function of some continuous variables that describe the properties of a test gene set. Gene size as a variable is particularly important in predicting genes for animal genomes because most of them are accounted as gene

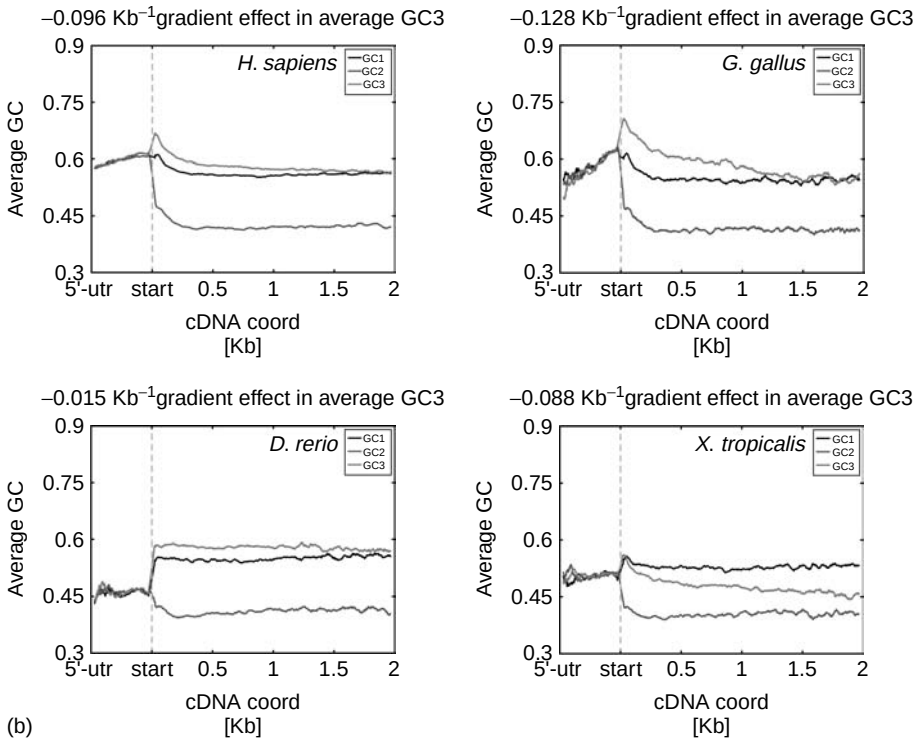

Fig. 11 (*Continued*)

spaces (including exons and introns, 1.2). Not only do mammalian genomes contain large genes that are megabases in size, but insect genomes too are known to have large genes across heterochromatin intervals. Therefore gene predictions should be carefully evaluated and interpreted based on the dynamic features of the target genome. Algorithms that succeed in analyzing one genome should not be blindly used in others, especially when they are by and large unrelated, such as (in an extreme case), between animal and plant genomes.

3.5.3 Sequence Variations

It is becoming increasingly common for a genome project to include an effort to generate sequence variation data through sequence sampling of wild counterparts (the chicken genome project did this in the opposite way, choosing the wild type, the red junglefowl, as the target), diverged breeds and closely related species. The study of sequence variation involves two basic steps: discovery in order to find candidate variations, and validation to acquire frequency information of verified variations in a background population. SGS provides a powerful procedure for the first step and could be used for the second in a rather limited way in consideration of cost differences between sequencing and genotyping. The cost of the latter is believed to be dropping at a much faster pace than that for sequencing.

Tab. 3 Major databases for sequence storage, genome information, and annotation.

Name	***URL***	***Description***
Sequence database		
GenBank	http://www.ncbi.nlm.nih.gov/	Nucleotide sequences
EMBL	http://www.ebi.ac.uk/embl/	Nucleotide sequences
DDBJ	http://www.ddbj.nig.ac.jp/	DNA Data Bank of Japan
SWISS-PROT	http://us.expasy.org/sprot/	Curated protein sequences
PIR	http://pir.georgetown.edu/	Protein sequences
Genome database		
ENSEMBL	http://www.ensembl.org/	A large collection of genomes
UCSC	http://www.genome.ucsc.edu/	A large collection of genomes
AceDB	http://www.acedb.org/	*C. elegans*
FlyBase	http://flybase.net/	*D. melanogaster*
SGD	http://www.yeastgenome.org/	*S. cerevisiae*
WormBase	http://www.wormbase.org/	*Worm genome*
GDB	http://www.gdb.org/	*H. sapiens*
RGD	http://rgd.mcw.edu/	*R. norvegicus*
MGD	http://www.informatics.jax.org/	*M. musculus*
TAIR	http://www.arabidopsis.org/	*A. thaliana*
BGI-RIS	http://rise.genomics.org.cn	*O. sativa*
Major databases for sequence annotation		
dbEST	http://www.ncbi.nlm.nih.gov/dbEST/	ESTs
RefSeq	http://www.ncbi.nlm.nih.gov/RefSeq/	Reference sequences
UniGene	http://www.ncbi.nlm.nih.gov/entrez/	Non-redundant gene clusters
dbSNP	http://www.ncbi.nlm.nih.gov/SNP/	SNPs
COG	http://www.ncbi.nlm.nih.gov/COG/	Cluster of Orthologous Groups of proteins
GO	http://www.geneontology.org/	Gene Ontology
InterPro	http://www.ebi.ac.uk/interpro/	Protein families and domains
Pfam	http://www.sanger.ac.uk/Software/Pfam/	Protein families
PRINTS	http://bioinf.man.ac.uk/dbbrowser/PRINTS/	Protein fingerprints
PROSITE	http://us.expasy.org/prosite/	Protein families and domains
ProDom	http://protein.toulouse.inra.fr/prodom/	Protein domain families
DIP	http://dip.doe-mbi.ucla.edu/	Database of Interacting Proteins
BIND	http://www.blueprint.org/bind/	Biomolecular interaction networks
KEGG	http://www.genome.jp/kegg/	Molecular interaction networks

3.5.4 Genome Evolution

The study of genome evolution in the Genomics Era has focused largely on lineage evolution (such as within vertebrates and fungi) in a vaguely defined context of comparative genomics, but nonetheless on the scale of genomes. Vertebrate genomes appear to be evolving to become more complex, as opposed to insect genomes that are evolving to become more diverged. Complexity has to build upon intricate networks of interacting genes or, more precisely, their products; these intricate networks are in turn defined functionally as molecular mechanisms and cellular processes. It would be of great interest to see how a new mechanism or a new process, created or modified in terms of new genes or

Tab. 4 A list of widely used *ab initio* gene-finding programs and the organisms applied.

Name	***Method***	***Organism***	***Access***
Glimmer	Interpolated Markov model	Microbes	http://www.tigr.org/software/glimmer/
Genie	General Hidden Markov model	Vertebrates	http://www-hgc.lbl.gov/projects/genie.html
GeneID	Perceptron, rules	Vertebrates	http://www.imim.es/GeneIdentification/Geneid/
FgeneSH	Hidden Markov model	Eukaryotes	http://www.softberry.com/
BGF	Hidden Semi Markov model	Eukaryotes	http://61.50.158.108:8080/GeneFinder/
GeneMark	5th-Markov	Eukaryotes	http://genemark.biology.gatech.edu/GeneMark/
GenScan	Semi Markov model	Vertebrates, plants, *C. elegans*	http://genes.mit.edu/GENSCAN.html
GenView	Linear combination	Vertebrates, diptera	http://www.itba.mi.cnr.it/webgene
GRAIL	Neural network	Vertebrates, Plants, *D. melanogaster*, and *E. coli*	http://compbio.ornl.gov/Grail-1.3

their definable variations over evolutionary timescales, results in building new organs, gaining new functions, and giving birth to new species.

As genomes of major domesticated crops and farm animals (oftentimes also their wild counterparts) are being sequenced, learning the rules that underlie many of the evolutionary changes in an artificial process (domestication) over a relatively short time period (10 000 years or so) should command the attention of many research teams, because of its scientific and economic importance. Ample collections of plant germplasms and animal breeds provide invaluable genomic resources, and domesticated animals can also serve as model organisms for human disease research.

Genome sequences evolve in both macro- and micro-scales. At the macro-scale, there are whole-genome duplications and subsequent chromosomal duplications, deletions, and rearrangements. At the micro-scale, there are gene duplications and deletions as well as enormous sequence variations within genes and gene families, and all are complicated by transposable element insertions over a variable evolutional timescale. Genome evolution should be studied in a much broader context than what has been defined by the classic field of molecular evolution.

4 Other Useful Resources for SGS

The ultimate goal of a genome-sequencing project is to be able to present a whole picture of a genome with all genes positioned on its chromosomes. In the pursuit of this goal, one faces at least two major challenges: identifying all the genes, and ordering them. Algorithms for *ab initio* gene identification are not yet mature, as proved by the fact that the numerous versions of such software have had only moderate success thus far. Under current circumstances, independent data are oftentimes required for a better annotation of the WG-SGS data; these include ESTs and cDNAs for validating the presence of gene transcripts, physical and genetic maps for gene ordering and referencing to inheritable traits, and comparative analysis to exploit evolutionary inference.

4.1 ESTs and Full-length cDNA

When compared with a WG-SGS project, EST sequencing requires little effort. For instance, to cover a 500-Mb genome once, it takes one million reads. To acquire significant coverage for a transcriptome for a given metazoan cell, 10 000 reads ought to be enough. It is worthwhile to allot 1 to 10% of the budget of a genome project to this effort, either as part of the main project or independent of it. As an example, the Sino-Danish Swine Genome Project has initiated a "one-million EST" Project, allowing researchers to acquire powerful transcriptomic data from one hundred different tissues and cell types, nearly 10 000 ESTs from each. Table 5 summarizes the top 20 EST collections from the public databases.

Full-length cDNA data are superior, but require much expertise to generate. It involves making full-length cDNA libraries, sequencing cDNA clones that are longer than typical sequencing reads, strenuous management inputs to follow up each clone, and reiterating the libraries for novel clones. Yoshihide Hyashizaki's group at RIKEN and Shoshi Kicuchi's group at NIAS, Japan, have done excellent jobs on

Tab. 5 A list of top 20 EST collections from selected species.

Species	*# of ESTs*	*# of UniGene entries*
H. sapiens	5 657 997	106 934
M. musculus	4 235 142	76 298
R. norvegicus	661 663	39 050
O. sativa	284 007	33 563
T. aestivum	559 149	24 735
B. Taurus	467 943	24 195
X. laevis	434 889	23 754
S. scrofa	328 573	21 947
D. rerio	532 545	20 978
A. thaliana	322 641	20 783
G. gallus	495 089	20 155
C. elegans	298 805	15 942
C. familiaris	154 720	15 665
X. tropicalis	423 107	14 632
Z. mays	415 211	14 179
O. mykiss	157 996	14 178
C. intestinalis	684 280	14 098
A. gambiae	150 042	14 085
D. melanogaster	382 439	13 928

Note: This list is ordered by the total number of UniGene entries/ESTs by 2004-8-20.

this front to have curated adequate among of human, mouse, and rice full-length cDNAs. Full-length cDNAs do provide irreplaceable resources for gene identification and sequence quality validation, and thus demand independent efforts.

4.2 Physical and Genetic Maps

The mapping of information, whether physical (other than sequence maps that are, strictly speaking, also a form of physical map) or genetic, is always a plus point for WG-SGS. Physical maps allow the assembled sequence and genes therein to be referenced to the chromosomes in physical order, and genetic maps link genes associated with inheritable traits to a physical location on a chromosome through polymorphic sequence sites. For large genomes in the size range over a billion basepairs, such information becomes essential. Another yet-to-be tackled challenge is to map genes on small chromosomes that have no obvious chromosome banding patterns, such as the microchromosomes in the chicken genome and numerous minute chromosomes in the genomes of crabs and shrimps. High-quality WG-SGS is capable of producing an assembled genome sequence map within one to a few scaffolds, which can be mapped physically by several methods, such as different versions of FISH and large-insert clone-based physical maps. Since both mapping processes are expensive, time-consuming, and resource-demanding, they can be constructed retrospectively after a WG-SGS, over a period of time. Nevertheless, sequence data, nevertheless, provide ample candidate markers from genes, microsatellites, and even SNPs, for mapping activities. Physical mapping methods are essentially based on three laboratory techniques. The most common technique is restriction digests of large-insert clones, or clone fingerprinting. It relies on redundancy and precise fragment size measurement to determine overlapping clones, and thus achieve ordering. The second technique is large-insert clone-based STS mapping. It leverages PCR technology and calls overlap according to whether the large-insert clones share short unique sequences in a given genome. The third technique is hybridization-based, in which short oligonucleotide sequences (also termed as sequence-tagged connectors generated from end sequences of large-insert clones) serve as probes to tag clones that are gridded in a very high density on reusable nylon filters.

Genetic maps are always needed for genetic and functional studies, where the relationship of each locus is linked by the frequencies of a recombination event as measured in pedigrees or populations. For SGS, genetic maps provide guidance in validating physical maps and sequence assembly, and SGS provides the most powerful tool for sequence polymorphism discovery.

4.3 Comparative Genome Analyses

Finally, one must emphasize the importance of comparative analyses. The NIH initiative of Mammalian Gene Collection (MGC) provides a precedent to direct focus towards this aspect. Representative mammal species from distinct taxonomic groups are all lined up in various sequencing pipelines, if not already sequenced. This is also being extended to other lower vertebrates and species in the root of chordates. Such comparative and systematic studies offer profound and broad information for not only genome annotation but also evolutionary studies on functional aspects of genes, gene families, and gene networks, thereby guiding molecular biologists to explore basic molecular mechanisms among all life forms. The next wave of genome sequencing will unquestionably go deeper and broader into taxa of the extant life forms, prioritized by economic and scientific interests. Comparative genomics offers a guidepost for targets of choice in genome sequencing, deviating from human-centric to species-focused.

5 Perspectives

Looking into the future of SGS, it is not only genomicists who have reason to be excited but also molecular biologists, and even field biologists in all disciplines. Genome sequencing is now well past its infancy and out of its crib. It is exploring new territories, checking out its skills, and learning all the time as it grows. Genomics has already broken out of its original definition and morphed into genome biology or systems biology without losing scale or momentum. What are some of the immediate challenges it has to face?

5.1 Large Genomes

Given the increasing requirements for sequence information, the need to tackle larger genomes is inevitable. The current records for finished genome projects on the basis of genome sizes are the human genome (among animal genomes) and rice (among plant genomes). Can WG-SGS strategy be applied to animal genomes larger than the human genome, or to plant genomes larger than that of rice? The answer is obviously yes but not without caveats. Genomes of different taxonomic groups vary quite significantly (Fig. 5). For instance, the genomes of mammals deviate very minimally from the mean size, about 3000 Mb, and are therefore suitable for simple WG-SGS. The next large plant genome to be sequenced may perhaps be the maize genome that is 2500 Mb in size. It is debatable, however, whether the wheat or barley genomes should be sequenced by the WG-SGS approach, because they pose an unprecedented challenge for sequence assembly on account of their repetitive contents (over 90%), their large sizes, and polyploid nature (in the case of wheat alone). Large insect and crustacean genomes, such as those of grasshoppers and shrimps, and large protist genomes

that have been largely untouched so far, should all be carefully evaluated before placed into a sequencing pipeline.

5.2 Genomes with High Repetitive Sequence Content

The repeat content, especially that of BDRs, varies widely among different taxonomic groups. Most notable are the plant genomes, some of which may have a repeat content as high as 95%. It is not a severe problem when the size of the target genome remains small, because the gene content can be put together easily and the cost is usually low. A balance between demands and costs is achievable within a threshold (say for genomes sized less than 3000 Mb). Large genomes in a size range of a few thousands of Mb have to be handled with great caution. Here, WG-SGC would not be suitable because of its high cost, rendering over 90% of the investment ineffective. Despite the fact that some of the gene content can be enriched experimentally by different methods (such as taking advantage of differential methylation between the gene and repeat contents among plant genomes), the goal of a dedicated genome project is to acquire a complete map of all its genes. In this regard, prioritization becomes a key for acquiring other resources (4) that are of great assistance in sequence assemblies and analyses over a relative prolonged period of time, in comparison with an ordinary genome project.

5.3 Polyploidy: Ancient and New

Polyploidy is very common among the extant species, some of it being ancient (formed naturally), and some recent (formed either naturally or artificially). A straightforward consideration for sequencing targets is to choose a diploid species (preferably the wild counterpart or the ancestor species) rather than a polyploid one, especially in the cases of crop species. However, in most cases, the preference is for the crop species to be sequenced directly first. In an ideal situation, the wild counterpart, the diploid species that gave rise to the modern polyploid crop, and the crop itself should all be sequenced in series. When there is no option, the genome of palepolyploid species should be evaluated first by physical mapping or sequence sampling for sequence divergence and feasibility of whole-genome assembly before a WG-SGS project is launched.

5.4 Mixed and Meta Genomes

In theory, when a mixture of multiple genomes is to be sequenced, it can be regarded as one organism with multiple chromosomes, perhaps in different ratios. For instance, genomes of a microbial community indigenous to a certain habitat are often referred to as *metagenomes*; examples of such habitats are the oral cavity, portions of a ruminant stomach, and even drinking water. However, microbes residing in the same environment are not guaranteed to propagate with the same speed and to respond to nutrient changes in similar ways. As a common characteristic under a defined circumstance, a microbial group is dominated by a limited number of species and contaminated by a large number of minor ones. Therefore, in an SGS project that utilizes DNA extracted from a meta-organism, genomes of the dominant species would be easily assembled, whereas those of the minor ones would have little chance of reaching full coverage. Unless novel strategies and techniques are

developed, such an endeavor provides only a survey of the metagenome and complete sequences of its few dominant members.

Acknowledgment

We apologize to our readers and colleagues for having only a limited number of citations from a recent explosion of genomic literature due to space limits. We would like to thank Ms. Wei Gong and Jing Wang for their valuable assistance in preparing the figures and tables, and Mr. Xin Zhang for proofreading the manuscript. The authors are supported by grants from National Natural Science Foundation (No. 30270748), Ministry of Science and Technology (No. 2002AA229021, 863-306-ZD11-04-1) and Chinese Academy of Sciences (KSCX2-SW-223).

See also DNA Libraries; Genetics, Molecular Basis of; Genomic DNA Libraries, Construction and Applications; Genomic Sequencing (Core Article); Nucleic Acid and Protein Sequence Analysis and Bioinformatics; Pragmatic Computational Biology: Sequence Analysis and Biological Systems Description; Serial Analysis of Gene Expression; Whole Genome Human Chromosome Physical Mapping.

Bibliography

Books and Reviews

Clayton, J., Dennis, C. (2003) *50 years of DNA*, Palgrave Macmillan, Hampshire, England.

Collins, F.S., Green, E.D., Guttmacher, A.E., Guyer, M.S. (2003) A vision for the future of genomics research-A blueprint for the genomic era, *Nature* **422**, 835–847.

Dulbecco, R. (1986) A turning point in cancer research: sequencing the human genome, *Science* **231**, 1055–1056.

Marshall, E. (1995) A strategy for sequencing the genome 5 years early, *Science* **267**, 783–784.

Olson, M.V. (1993) The human genome project, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4338–4344.

Olson, M.V. (1995) A time to sequence, *Science* **270**, 394–396.

Olson, M.V. (2001) The maps. Clone by clone by clone, *Nature* **409**, 816–818.

Sulston, J., Ferry, G. (2002) *The Common Thread: A Story of Science, Politics, Ethics and the Human Genome*, Joseph Henry Press, United States.

Waterston, R.H., Lander, E.S., Sulston, J.E. (2002) On the sequencing of the human genome, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3712–3716.

Watson, J.D. (1990) The human genome project: past, present, and future, *Science* **248**, 44–49.

Primary References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool, *J. Mol. Biol.* **215**, 403–410.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. (2000) The genome sequence of *Drosophila melanogaster*, *Science* **287**, 2185–2195.

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., et al. (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*, *Science* **297**, 1301–1310.

Arratia, R., Lander, E.S., Tavare, S., Waterman, M.S. (1991) Genomic mapping by anchoring random clones: a mathematical analysis, *Genomics* **11**, 806–827.

Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. (1997) The complete genome sequence of *Escherichia coli* K-12, *Science* **277**, 1453–1474.

Burke, D.T., Carle, G.F., Olson, M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors, *Science* **236**, 806–812.

Celniker, S.E., Wheeler, D.A., Kronmiller, B., Carlson, J.W., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, S.P., Frise, E., et al. (2002) Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence, *Genome Biol.* **3**, RESEARCH0079.

Cohen, S.N., Chang, A.C., Boyer, H.W., Helling, R.B. (1973) Construction of biologically functional bacterial plasmids in vitro, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3240–3244.

Coulson, A., Sulston, J., Brenner, S., Karn, J. (1986) Toward a physical map in the genome of the nematode *Caenorhabditis elegans*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7821–7825.

Dehal, P., Satou, Y., Campbell, R.K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D.M., et al. (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins, *Science* **298**, 2157–2167.

Ewing, B., Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities, *Genome Res.* **8**, 186–194.

Ewing, B., Hillier, L., Wendl, M.C., Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment, *Genome Res.* **8**, 175–185.

Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., FitzHugh, W., Ma, L.J., Smirnov, S., Purcell, S., et al. (2003) The genome sequence of the filamentous fungus *Neurospora crassa*, *Nature* **422**, 859–868.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., et al. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*, *Nature* **419**, 498–511.

Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E., et al. (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution, *Nature* **428**, 493–521.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*), *Science* **296**, 92–100.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al. (1996) Life with 6000 genes, *Science* **274**, 546–567.

Gordon, D., Abajian, C., Green, P. (1998) Consed: a graphical tool for sequence finishing, *Genome Res.* **8**, 195–202.

Green, E.D., Olson, M.V. (1990) Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1213–1217.

Gregory, S.G., Sekhon, M., Schein, J., Zhao, S., Osoegawa, K., Scott, C.E., Evans, R.S., Burridge, P.W., Cox, T.V., Fox, C.A., et al. (2002) A physical map of the mouse genome, *Nature* **418**, 743–750.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, J.M., Wides, R., et al. (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*, *Science* **298**, 129–149.

Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K., Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M., et al. (2003) The dog genome: survey sequencing and comparative analysis, *Science* **301**, 1898–1903.

Lander, E.S., Waterman, M.S. (1988) Genomic mapping by fingerprinting random clones: a mathematical analysis, *Genomics* **2**, 231–239.

Martinez, D., Larrondo, L.F., Putnam, N., Gelpke, M.D., Huang, K., Chapman, J., Helfenbein, K.G., Ramaiya, P., Detter, J.C., Larimer, F., et al. (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78, *Nat. Biotechnol.* **22**, 695–700.

Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S.Y., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., et al. (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D, *Nature* **428**, 653–657.

Maxam, A.M., Gilbert, W. (1977) A new method for sequencing DNA, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560–564.

Messing, J., Crea, R., Seeburg, P.H. (1981) A system for shotgun DNA sequencing, *Nucleic Acids Res.* **9**, 309–321.

Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori, H., Namiki, N., Kitagawa, M., Yamashita, H., Yasukochi, Y., et al. (2004) The genome sequence of silkworm, *Bombyx mori*, *DNA Res.* **11**, 27–35.

Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H., Remington, K.A., et al. (2000) A whole-genome assembly of Drosophila, *Science* **287**, 2196–2204.

Olson, M., Hood, L., Cantor, C., Botstein, D. (1989) A common language for physical mapping of the human genome, *Science* **245**, 1434–1435.

Port, E., Sun, F., Martin, D., Waterman, M.S. (1995) Genomic mapping by end-characterized random clones: a mathematical analysis, *Genomics* **26**, 84–100.

Riethman, H.C., Moyzis, R.K., Meyne, J., Burke, D.T., Olson, M.V. (1989) Cloning human telomeric DNA fragments into *Saccharomyces cerevisiae* using a yeast-artificial-chromosome vector, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6240–6244.

Riles, L., Dutchik, J.E., Baktha, A., McCauley, B.K., Thayer, E.C., Leckie, M.P., Braden, V.V., Depke, J.E., Olson, M.V. (1993) Physical maps of the six smallest chromosomes of *Saccharomyces cerevisiae* at a resolution of 2.6 kilobase pairs, *Genetics* **134**, 81–150.

Sanger, F., Coulson, A.R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase, *J. Mol. Biol.* **94**, 441–448.

Sanger, F., Nicklen, S., Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., Hutchison, C.A., Slocombe, P.M., Smith, M. (1977) Nucleotide sequence of bacteriophage phi X174 DNA, *Nature* **265**, 687–695.

Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8794–8797.

Smit, A.F.A., Green, P. RepeatMasker [http://ftp.genome.washington.edu/RM/RepeatMasker.html]

Smith, M., Brown, N.L., Air, G.M., Barrell, B.G., Coulson, A.R., Hutchison, C.A. III, Sanger, F. (1977) DNA sequence at the C termini of the overlapping genes A and B in bacteriophage phi X174, *Nature* **265**, 702–705.

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H., Hood, L.E. (1986) Fluorescence detection in automated DNA sequence analysis, *Nature* **321**, 674–679.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., et al. (2003) The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics, *PLoS Biol.* **1**, E45.

The Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, *Nature* **408**, 796–815.

The *C. elegans* Sequencing Consortium. (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018.

The International Human Genome Sequencing Consortium. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

The Mouse Genome Sequencing Consortium. (2002) Initial sequencing and comparative analysis of the mouse genome, *Nature* **420**, 520–562.

Vieira, J., Messing, J. (1982) The pUC plasmids, an M13MP7-derived system for insertion mutagenesis and sequencing with synthetic universal primers, *Gene* **19**, 259–268.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Whitfield, C.W., Band, M.R., Bonaldo, M.F., Kumar, C.G., Liu, L., Pardinas, J.R., Robertson, H.M., Soares, M.B., Robinson, G.E. (2002) Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee, *Genome Res.* **12**, 555–566.

Wong, G., Yu, J., Thayer, E.C., Olson, M.V. (1997) Multiple-complete-digest restriction fragment mapping: generating sequence-ready maps for large-scale DNA sequencing, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5225–5230.

Wong, G.K.-S., Passey, D.A., Huang, Y.Z., Yang, Z.Y., Yu, J. (2000) Is ''junk'' DNA

mostly intron DNA? *Genome Res.* **10**, 1672–1678.

Wong, G.K.-S., Wang, J., Tao, L., Tan, J., Zhang, J.G., Passey, D.A., Yu, J. (2002) Compositional gradients in *Gramineae* genes, *Genome Res.* **12**, 851–856.

Wu, R. (1970) Nucleotide sequence analysis of DNA. I, Partial sequence of the cohesive ends of bacteriophage lamda and 186 DNA, *J. Mol. Biol.* **51**, 501–521.

Wu, R., Taylor, E. (1971) Nucleotide sequence analysis of DNA. II, Complete nucleotide sequence of the cohesive ends of bacteriophage lambda DNA, *J. Mol. Biol.* **57**, 491–511.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*), *Science* **296**, 79–92.

Signaling in Cancer: *see* Intracellular Signaling in Cancer

Signal Transduction Mediated by Heptahelical Receptors and Heterotrimeric G Proteins

Terence E. Hébert[1], *John K. Northup*[2], *and R. Victor Rebois*[2]
[1] *Université de Montréal, Montréal, PQ, Canada*
[2] *National Institutes of Health, Rockville, MD 20850-3211*

1 Overview of G Protein–mediated Signal Transduction 146

2 Historical Background 146

3 Heptahelical Receptors 147
3.1 Structure of Heptahelical Receptors and Mechanistic Aspects of Activation 147
3.2 Receptor Desensitization and Downregulation 150
3.3 Receptor-mediated Activation of Pathways that Do Not Involve G Proteins 150

4 G Proteins 151
4.1 Structure and Mechanistic Aspects of G Protein Activation 151
4.2 G Protein Regulation that Does not Involve Heptahelical Receptors 152
4.3 Attenuation of G Protein Activity by Regulators of G Protein–signaling Proteins 154

5 Effectors 155
5.1 cGMP Phosphodiesterase 155
5.2 Adenylyl Cyclase 155
5.3 Phosphoinositide 3-Kinase 156
5.4 Phospholipase Cβ 156
5.5 G Protein–coupled Inwardly Rectifying Potassium (Kir3) Channels 157
5.6 N and P/Q Type Calcium Channels 157
5.7 RhoGEFs 158
5.8 Mitogen-activated Protein (MAP) Kinases 159

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

6 The Organization of G Protein-mediated Signal Transduction Systems 161
6.1 The Classical Model of G Protein–mediated Signal Transduction 161
6.2 Compartmentation 163
6.3 Signal Transduction Complexes 163
6.3.1 Scaffolding Proteins 163
6.3.2 Complexes Containing Receptors, G Proteins, and Effectors 165

7 Signal Transduction: Striking A Balance Between Different Organizational Paradigms 168
7.1 Insight 168
7.2 Mammalian Phototransduction: A Tale of Two Signal Transduction Mechanisms 168

8 Clinical Importance 171
8.1 G Protein-mediated Signal Transduction in Diseases 171
8.2 Heptahelical Receptors as Drug Targets 172
8.3 Potential for Therapeutic Treatments that Target G Protein–mediated Signaling Components Other than Heptahelical Receptors 172

9 Future Directions 173

Bibliography 173
Books and Reviews 173
Primary Literature 174

Keywords

Agonist
Any chemical or physical signal that interacts with a specific receptor resulting in receptor activation. Agonists can be subdivided into partial agonists and full agonists depending upon the degree to which they activate the receptor, and the intrinsic, constitutive activity of a receptor can be suppressed by an inverse agonist.

Antagonist
Any chemical signal that interacts with a specific receptor, but neither activates the receptor nor suppresses the intrinsic activity of the receptor. These properties enable an antagonist to block the effects on an agonist or inverse agonist by competing for the receptor binding site.

Effector
A protein whose activity is directly regulated by an activated heterotrimeric G protein.

Heptahelical (7TM) Receptor
Integral membrane proteins that have seven amphipathic helices, and function as receptors for specific ligands. Agonist-mediated activation of heptahelical receptors results in the activation of specific heterotrimeric G proteins.

Heterotrimeric G Protein
A protein consisting of three different subunits referred to as α, β, and γ. The α-subunit has a guanine nucleotide binding site, and intrinsic GTPase activity. Agonist-occupied heptahelical receptors activate heterotrimeric G proteins by facilitating the dissociation of tightly bound GDP, allowing GTP to bind in its place.

Ligand
Any chemical or physical signal that interacts with a specific receptor. This includes full and partial agonists, antagonists, and inverse agonists.

Signal Transduction
A sequential series of biochemical reactions initiated by an environmental stimulus that interacts with a specific receptor located inside a cell or on the cell's surface.

G protein-mediated signal transduction pathways are involved in the responses of organisms and their constituent cells to a wide variety of stimuli. These stimuli include ions, light, gustants, odorants, small peptides, proteins, hormones, neurotransmitters, and lipids. The nature of the response can be equally diverse, varying from changes in gene transcription to altered ion channel kinetics. These pathways are utilized by nearly every species from yeast and fungi to mammals and higher plants. Heptahelical receptors, heterotrimeric guanine nucleotide binding proteins (G proteins) and their proximal downstream effectors constitute the core components of these ubiquitous signaling pathways. Signal transduction is usually initiated when an agonist binds to a selected heptahelical receptor, which in turn activates its cognate G protein. When activated, G proteins are able to regulate the activity of specific effectors such as adenylyl cyclase and ion channels. In addition to the core components, many accessory proteins are involved in the regulation of G protein–mediated signal transduction. Now the challenge lies in elucidating how these pathways with their many constituent components are organized within the cell. G protein–mediated signal transduction pathways are arguably the single most important class of targets for pharmaceuticals. By and large the drugs used for the therapeutic treatment of diseases involving G protein–mediated signaling are ligands that bind to the receptors for these pathways. Although receptor binding sites remain an important target for the development of new drugs, a thorough understanding of the organizational structure of G protein–signaling pathways is likely to reveal additional drug targets that may have therapeutic advantages over those that act as either agonists or antagonists for the receptors themselves.

1
Overview of G Protein–mediated Signal Transduction

Signal transduction refers to the biochemical events involved in the response of a cell to environmental cues. These messages, known as agonists, can take a variety of physical or chemical forms. They initiate their effects by interacting with specific receptor proteins on the cell surface, or in some cases, inside the cell. Receptors fall into four categories: (1) steroid hormone receptors that bind testosterone, estrogen, and other derivatives of cholesterol; (2) receptors that possess an intrinsic enzymatic activity such as receptor tyrosine kinases that respond to agonists such as growth factors and insulin; (3) receptors with an intrinsic ion channel activity such as nicotinic acetylcholine receptors; and (4) receptors that activate heterotrimeric guanine nucleotide binding proteins (G proteins). The latter category accounts for the large majority of receptors. In mammals there are approximately 1500 receptors that activate G proteins of which approximately two-thirds are odorant receptors associated with the sense of smell. A common feature of these receptors is that they have seven amphipathic helices arranged so that each helix fully traverses the membrane in which the receptor is embedded. This topography has given rise to the terms heptahelical receptor or seven transmembrane (7TM) receptor. Agonists for approximately 200 7TM receptors have been identified leaving the majority of 7TM receptors on a list of "orphan" receptors to await the discovery of their agonists.

G proteins that are activated by 7TM receptors are composed of three different subunits referred to as Gα, Gβ, and Gγ. Yeasts have only one of each G protein subunit, while mammals express dozens of different G protein subunit combinations. The α-subunit contains the guanine nucleotide binding site, and in the unactivated G protein, GDP is bound to Gα. When an agonist binds to its 7TM receptor, it facilitates the dissociation of GDP, allowing GTP to bind in its place. This activates the G protein, which in turn regulates the activity of so-called effector proteins. The intrinsic GTPase activity of Gα returns the G protein to its inactive state. As an enzyme, Gα is a relatively slow catalyst, and in some cases catalysis can be accelerated by accessory proteins known as regulators of G protein–signaling (RGS) proteins. This scenario represents the rule and the requisite exceptions are: (1) 7TM receptors can mediate signal transduction without activating G proteins (hence, our choice of the term heptahelical receptors or 7TM receptors rather than the commonly used acronym GPCR for G protein–coupled receptors); (2) G proteins can be activated by proteins other than 7TM receptors; and (3) signal transduction involving mitogen-activated protein kinase cascades that are normally regulated by receptor tyrosine kinases can also be regulated by 7TM receptors and G proteins.

2
Historical Background

During the first few decades of the twentieth century it was discovered that cells react to extracellular chemical substances that we now call hormones and neurotransmitters. The concept of a "receptor" for these environmental messages arose in order to account for the ability of cells to discriminate between them, and respond specifically to them. By mid century, the

idea that hormones and neurotransmitters interacted with specialized receptors in the plasma membrane, and subsequently regulated the activity of "effectors" within the cell, was a prevalent hypothesis despite the fact that none of the cellular components involved in the process had been definitively identified. In the late 1950s, Sutherland and Rall discovered that bathing liver cells with a solution of epinephrine or glucagon lead to the production of adenosine 3′, 5′ monophosphate (cAMP) inside the cell. This observation provided evidence that information in the form of environmental cues could be converted by the cell into an intracellular "second messenger." It was not apparent whether the enzyme responsible for cAMP synthesis, adenylyl cyclase (AC) (previously known as adenyl or adenylate cyclase), was itself the receptor or if there was a separate receptor component. Experimental data eventually revealed that both receptor and effector functions did not reside on the same protein. This raised the question of how information flowed from receptor to effector or, in the words of Martin Rodbell, what the "transducer" process was. The discovery that hormone-mediated stimulation of adenylyl cyclase required guanine nucleotide, that a guanine nucleotide binding component could be separated from the component responsible for catalyzing cAMP synthesis, and that GTP is hydrolyzed during the process led to the conclusion that the transducer was actually a separate component of the pathway. The transducer was eventually purified from retina and liver. The fact that the transducer contained the guanine nucleotide binding site eventually led to its being called a "G protein." The intrinsic GTPase activity of the α-subunit identified the G protein as a molecular switch that allowed the chain of events known as signal transduction to occur when GTP was bound to the G protein, and to shut it off when the GTP was hydrolyzed. Although 7TM receptors, G proteins, and effectors constitute the essential elements of G protein–mediated signaling pathways, many accessory proteins that are critical for the signal transduction process have been discovered and characterized.

3 Heptahelical Receptors

3.1 Structure of Heptahelical Receptors and Mechanistic Aspects of Activation

7TM receptors have seven amphipathic helices, and each helix traverses the cell membrane once. Biochemical evidence showed that the N terminus was on the extracellular side of the plasma membrane leaving the C terminus on the cytoplasm side. The segments connecting the transmembrane domains are commonly referred to as *extracellular* and *intracellular* loops. Many 7TM receptors have a disulfide bond that links the first and second extracellular loops and a fatty acyl moiety covalently bound to their C-terminal cytoplasmic domains. 7TM receptors can be divided into six families (Fig. 1). Structural diversity in the binding site of 7TM receptors enables them to recognize an amazing variety of ligands, but despite this fact all 7TM receptors must activate a heterotrimeric G protein for signal transmission to occur. A bundle of the seven transmembrane domains is the structural element that allows this diverse collection of ligand binding sites to convert agonist binding into conformational changes that are capable of activating the G protein.

Rhodopsin is the best-studied 7TM receptor, and the first and only 7TM receptor to be crystallized. The crystal structure of rhodopsin provided information about a 7TM receptor in its inactive state, thus revealing the arrangement of the transmembrane helices with respect to one another (Fig. 2). Although no crystal structure for the active conformation of the receptor exists, a variety of data have been

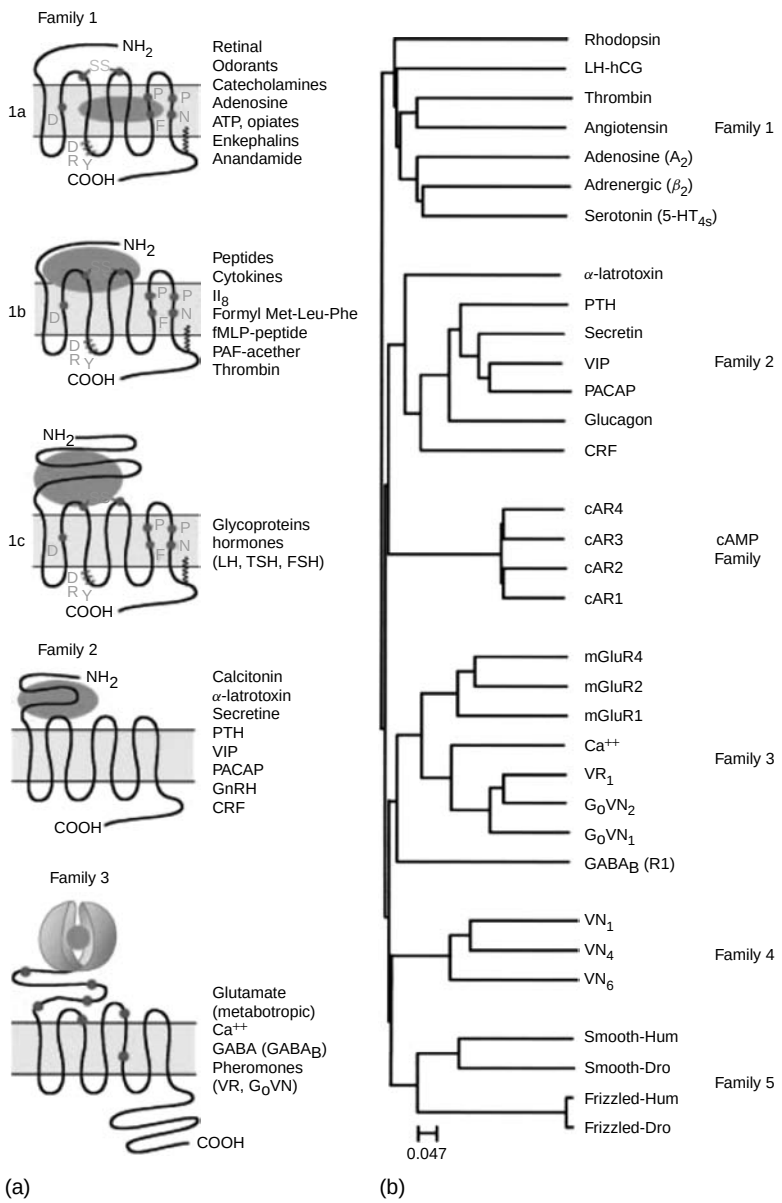

used to predict specific agonist-induced changes in the receptor's conformation. These changes include both rotations and displacements of the transmembrane helices relative to one another that carry through to the cytoplasmic domains of the receptor, thus enabling it to activate its cognate G protein.

Fig. 2 Rhodopsin is depicted as a ribbon model based on the X-ray crystallographic structure of the receptor. Rhodopsin, like all heptahelical receptors, spans the cell membrane so that it is exposed to both the extracellular milieu and the cytoplasm. The model is oriented so that the extracellular and intracellular compartments are above and below the molecule, respectively. The 11-cis retinal that is isomerized to the trans-configuration when struck by a photon is shown as a space-filling model in purple. Each of the seven amphipathic helices is shown in a different color and rhodopsin has an additional short eighth helix near its C terminus that lies on the cytoplasmic surface of the cell membrane in the unactivated state of the receptor. (See color plate p. xxv.)

Fig. 1 7TM receptors can be divided into six families on the basis of sequence homology. (A) Three main families ((1, 2, and 3) can be easily recognized by comparing their amino acid sequences. Family 1 contains most 7TM receptors including odorant receptors. Family 1a contains receptors for small ligands and includes rhodopsin and β-adrenergic receptors. The binding site for this family is located within the seven transmembrane bundle. Family 1b contains receptors for peptides whose binding site includes the N-terminal, the extracellular loops and regions of the transmembrane domains near the extracellular surface. Family 1c receptors have a large extracellular N-terminal domain that together with extracellular loops 1 and 3 form the binding site for glycoprotein hormones. Family 2 7TM receptors have morphology similar to family 1c receptors, but they do not share sequence homology. Agonists for family 2 receptors include glucagon. Family 3 contains metabotropic glutamate receptors (mGluR) and γ-amino butyric acid receptors ($GABA_B$). (B) Family 4 receptors include pheromone receptors; family 5 contains "frizzled" and "smoothened" receptors that are involved in embryogenesis, and the family of cAMP receptors that have only been found in *D. discoideum*. Reprinted from *EMBO Journal*, Vol. 18, J. Bockaert and J. P. Pin, Molecular tinkering of G protein-coupled receptors: An evolutionary success, Pages 1723–1729, Copyright (1999), with permission from Macmillan Publishers Ltd.

3.2 Receptor Desensitization and Downregulation

Signaling by an activated 7TM receptor is rapidly attenuated by a process known as desensitization. Desensitization is initiated by sustained agonist-mediated activation of the receptor, which renders it susceptible to phosphorylation. The observation that rhodopsin was phosphorylated following activation by light led to the discovery of the first 7TM receptor kinase, which was dubbed rhodopsin kinase. Rhodopsin kinase is a member of a family of G protein–coupled receptor kinases (GRKs), and is now known as GRK1. Since its discovery, six additional GRKs (GRK2 through GRK7) have been identified. GRKs catalyze the phosphorylation of serine and threonine residues in the third intracellular loop and the C-terminal cytoplasmic domain of 7TM receptors. The efficacy with which GRK phosphorylates 7TM receptors can be modulated by other kinases. The consequence of GRK-mediated phosphorylation is that the 7TM receptor has an increased affinity for proteins known as arrestins. The prototypical member of this family, visual arrestin (originally known as 48-kD protein, and now called arrestin-1), is found in retinal rod cells. Two other family members, β-arrestin-1 and -2 (otherwise known as arrestin-2 and -3, respectively), are widely distributed, while the fourth member, cone arrestin, is found in retinal cone cells. The binding of arrestin to a GRK-phosphorylated 7TM receptor is believed to directly interfere with the receptors ability to activate its cognate G protein, attenuating signal transduction. Arrestins have also been labeled as "scaffolding" proteins (Sect. 6.3.1) because they bind many proteins in addition to phosphorylated 7TM receptors, and it is through these interactions that arrestin mediates the movement of phosphorylated 7TM receptors to microdomains on the cell surface known as coated pits. Associated with the cytoplasmic surface of coated pits are three proteins, the heavy chain of clathrin, adapter protein 2 (AP2) and N-ethylmaleimide-sensitive factor (NSF), which have been shown to bind β-arrestins. These proteins enable arrestin to drag 7TM receptors into the coated pit so that they can be taken into the cell by endocytosis. The extent to which different 7TM receptors are internalized varies depending upon their ability to engage the endocytotic machinery. In the process of endocytosis, coated pits become coated vesicles that deliver 7TM receptors to organelles called endosomes. Small GTP binding proteins including many Rab family members, ADP ribosylation factor 6, and GIT2 have also been implicated in the trafficking of 7TM receptors to endosomes. Heterotrimeric G proteins are also internalized after signaling, but the fate of the G protein is not known and little is known about the trafficking of effectors following signaling. The fate of the internalized 7TM receptors differs depending upon the type of receptor. Some receptors are resensitized by dephosphorylation and recycled back to the cell surface while others are degraded in a process known as downregulation. Factors governing which pathway a receptor takes remain to be elucidated.

3.3 Receptor-mediated Activation of Pathways that Do Not Involve G Proteins

7TM receptors are now recognized as being able to regulate biochemical pathways independently of heterotrimeric G protein

activation. Although 7TM receptors can activate mitogen-activated protein (MAP) kinase pathways by a G protein–dependent mechanism (Sect. 5.8), it is also clear that these receptors can do so by a G protein-independent mechanism. MAP kinase pathways are so named because of their involvement in the response of cells to mitogenic signals that typically act through receptor tyrosine kinases including epidermal growth factor (EGF) receptors or fibroblast growth factor receptors, among others. The cellular slime mold, *Dictoystelium discoideum*, has a number of 7TM receptors, but only one gene for a Gβ. Disruption of the gene for Gβ prevents activation of all known G protein–dependent effectors in Dictoystelium, but these Gβ null mutants retain the ability to activate MAP kinase pathways after 7TM receptor stimulation. 7TM receptor regulation of MAP kinase pathways also occurs in mammalian cells. The mechanism may involve events that are associated with receptor desensitization, including phosphorylation of the agonist-occupied 7TM receptor by GRK and arrestin binding, but the actual process of endocytosis (Sect. 3.2) is not a prerequisite for MAP kinase activation. In addition to MAP kinase pathways, other G protein–independent physiological responses to 7TM receptor activation have been identified, including alterations in gene transcription and ion fluxes across membranes.

4 G Proteins

4.1 Structure and Mechanistic Aspects of G Protein Activation

The crystal structure of a heterotrimeric protein revealed that Gα has two domains. One domain is replete with α-helices prompting it to be named the α-helical domain, and the other that contains the guanine nucleotide binding site has a structure similar to the small GTP binding protein Ras, thus it is called the Ras-like domain (Fig. 3). The N terminus of Gα subunits is covalently modified by the attachment of myristate, palmitate, or both. Myristate or palmitate is bound by an amide linkage to the N-terminal glycine residue (exposed after removal of the initiator methionine) while palmitate is attached by a thioester linkage to a cysteine near the N terminus.

Gβ has seven similar structural domains that are assembled so that they resemble a seven bladed propeller. The N terminus of Gβ extends beyond the propeller, and is intertwined with the N terminus of Gγ forming a coiled coil structure. The C terminus of Gγ is covalently modified by the attachment of an isoprenoid moiety (either a geranylgeranyl or farnesyl group). The N terminus of Gα, with its covalently bound fatty acyl chains lies close to the isoprenylated C terminus of Gγ and these hydrophobic moieties are important for anchoring the G protein to the cell membrane. The N terminus of Gα also makes contact with amino acid residues in the first blade of the Gβ propeller. In the unactivated G protein, one surface of the Gβ propeller makes additional contact with a region of Gα known as the "switch interface." The switch interface is so named because it contains regions called switches that change conformation when GTP or a GTP analog replaces GDP in the guanine nucleotide binding site of Gα. Activation of the G protein by an agonist-occupied 7TM receptor disrupts the interaction between Gβ and the switch interface of Gα allowing GDP to dissociate so that GTP can bind in its place. In the

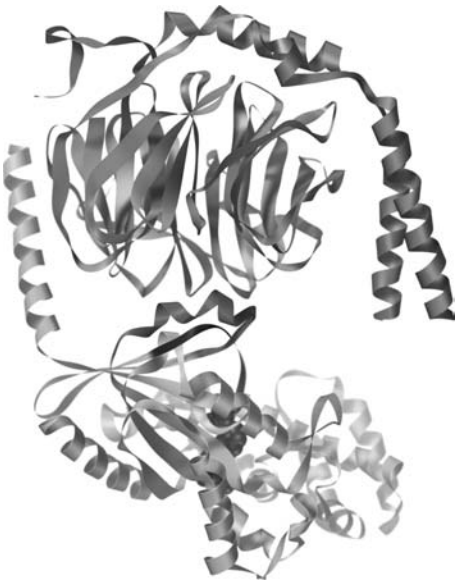

Fig. 3 The unactivated conformation of a heterotrimeric G protein is shown as a ribbon model with bound GDP depicted as a space-filling model (purple) within the Gα subunit. The model is based on the X-ray crystal structure of a protein consisting of a chimeric α-subunit composed predominantly of residues from Gα_t associated with the transducin G$\beta\gamma$ heterodimer (Gβ is shown in blue). The C terminus of Gγ (green) and the N terminus of Gα, which have covalently attached hydrophobic moieties that anchor the G protein to the membrane, lie close to each other at the upper left of the figure. Switches I, II, and III on Gα are shown in progressively lighter shades of red. (See color plate p. xxxii.)

GTP bound state, the conformation of the three switches are significantly altered, thereby preventing Gβ from binding to this surface of Gα. In this way, the G protein is held in an active conformation until GTP is hydrolyzed and the switches can return to their inactive conformations, thus allowing Gβ to again interact with the switch interface on Gα. GTP hydrolysis is a consequence of the intrinsic GTPase activity of Gα (EC 3.6.5.1). The catalytic rates for GTP hydrolysis by G proteins are characteristically slow for an enzyme, but they can be increased by accessory proteins known as *regulators of G protein signaling* or RGS proteins (Sect. 4.3).

There are 12 mammalian genes that code for Gγ subunits, five that code for Gβ subunits, and 16 that code for Gα subunits. The Gα subunits are grouped into four subfamilies (G_s, $G_{i/o}$, $G_{q/11}$, and $G_{12/13}$) on the basis of sequence homology (Table 1). The subtype of G protein responsible for regulating the activity of specific effectors can be found in Sect. 5.

4.2 G Protein Regulation that Does not Involve Heptahelical Receptors

Aside from 7TM receptors, there exists a diverse group of proteins that can interact directly with G proteins, and a number of them are capable of activating G proteins by various mechanisms. The first nonreceptor protein recognized as being able to activate a G protein was cholera toxin produced by the bacteria *Vibrio cholera*. Cholera toxin is an ADP-ribosyl transferase that covalently attaches the ADP-ribose moiety of NAD to an arginine residue (Arg^{201} of bovine Gα_s) in "switch I" of Gα_s. This covalent modification prevents G_s from hydrolyzing GTP once it is bound. As a consequence, G_s becomes irreversibly activated, with important consequences for the progression of the disease cholera. Another toxin, pertussis toxin, from the bacteria *Bordetella pertussis* is also an ADP-ribosyl transferase, but its substrate is a cysteine residue near the C terminus of G

Tab. 1 Heterotrimeric G proteins.

Name	***Gene***	***Expression***
α-Subunits		
$G\alpha_s$ class		
$G\alpha_s$	*Gnas*	Ubiquitous
$G\alpha_{sXL}$	*Gnasxl*	Neuroendocrine
$G\alpha_{olf}$	*Gnal*	Olfactory epithelium, brain
$G\alpha_{i/o}$ class		
$G\alpha_{i1}$	*Gnai1*	Widely distributed
$G\alpha_{i2}$	*Gnai2*	Ubiquitous
$G\alpha_{i3}$	*Gnai3*	Widely distributed
$G\alpha_p$	*Gnao*	Neuronal, neuroendocrine
$G\alpha_z$	*Gnaz*	Neuronal, platelets
$G\alpha_{gust}$	*Gnag*	Taste cells, brush cells
$G\alpha_{t-r}$	*Gnat1*	Retinal rods, taste cells
$G\alpha_{t-c}$	*Gnat2*	Retinal cones
$G\alpha_{q/11}$ class		
$G\alpha_q$	*Gnaq*	Ubiquitous
$G\alpha_{11}$	*Gna11*	Almost ubiquitous
$G\alpha_{14}$	*Gna14*	Kidney, lung, spleen
$G\alpha_{15/16}$	*Gna15*	Hematopoietic cells
$G\alpha_{12/13}$ class		
$G\alpha_{12}$	*Gna12*	Ubiquitous
$G\alpha_{13}$	*Gna13*	Ubiquitous
β-Subunits		
β_1	*Gnb1*	Widely, retinal rods
β_2	*Gnb2*	Widely distributed
β_3	*Gnb3*	Widely, retinal cones
β_4	*Gnb4*	Widely distributed
β_5	*Gnb5*	Mainly brain
γ-Subunits		
γ_1, γ_{rod}	*Gngt1*	Retinal rods, brain
γ_{14}, γ_{cone}	*Gngt2*	Retinal cones, brain
γ_2, γ_6	*Gng2*	Widely
γ_3	*Gng3*	Brain, blood
γ_4	*Gng4*	Brain and other tissues
γ_5	*Gng5*	Widely
γ_7	*Gng7*	Widely
γ_8, γ_9	*Gng8*	
γ_{10}	*Gng10*	Widely
γ_{11}	*Gng11*	Widely
γ_{12}	*Gng12*	Widely
γ_{13}	*Gng13*	Taste buds

The table lists heterotrimeric G protein subunits encoded by mammalian species together with their tissue distribution and the gene symbol. $G\alpha$ subunits are divided into subclasses on the basis of sequence homology. Reprinted from *Progress in Biophysics and Molecular Biology*, Vol. 83, S. Offermanns, G-proteins as transducers in transmembrane signalling, Pages 101–130, Copyright (2003), with permission from Elsevier.

proteins that belong to the G_i subfamily. ADP ribosylation of G_i by pertussis toxin prevents it from being activated by a 7TM receptor. Mastoparan is a tetradecapeptide found in wasp venom that activates G_i by virtue of being a 7TM receptor mimetic. Analogs of mastoparan that activate G_s have also been prepared.

Receptor tyrosine kinases (*e.g.* growth factor receptors) can also use G proteins to regulate the activity of downstream effectors apparently without involving 7TM receptors. Among the receptor tyrosine kinases that fall into this category are platelet-derived growth factor receptors, insulin receptors, and insulin-like growth factor-1 receptors. Other non-7TM receptors have also been implicated in the activation of G proteins. These include erythropoietin receptors and a glycosylphosphatidylinositol-anchored receptor, CK14.

Three mammalian proteins designated AGS1, AGS2 and AGS3 (activators of G protein signaling) were identified by a screen designed to detect proteins that could activate G proteins. Despite the nomenclature, these three proteins are not homologs. AGS1 is identical to DexRas1, a member of the Ras superfamily. AGS1 has been shown to facilitate the binding of a GTP analog to G_i, accounting for its ability to activate G_i. AGS2 is identical to tctex1, the light chain of cytoplasmic motor protein dynein. AGS2 does not cause an exchange of guanine nucleotides but appears to disrupt the interaction between $G\alpha$ and the $G\beta\gamma$ heterodimer, exposing epitopes on the latter that are capable of activating downstream effectors. AGS3 is a mouse protein with homologs in humans (LGN) and the fruit fly *Drosophila melanogaster* (Partner of Inscrutable or Pins). Pins is important for the asymmetric cell division in the *Drosophila* embryo. Although AGS2 and AGS3 were identified on the basis of their ability to activate G proteins in a genetic screen, it is not yet clear that they do so as part of their normal physiological function. However, the ability of AGS proteins to interact with G proteins is certain to be important *in vivo*. Other activators of G proteins include the microtubule protein tubulin, a protein found in the growth cones of neurons called GAP-43, a protein associated with familial Alzheimer's disease known as presenilin-1, G protein activator, and phosphatidylethanolamine-binding protein. Other proteins that do not activate G proteins, but are reported to regulate them include caveolin (Sect. 6.3.1), the signature proteins of lipid microdomains called *caveolae*.

4.3 Attenuation of G Protein Activity by Regulators of G Protein–signaling Proteins

The visual image that you perceive is a consequence of signal transduction mediated by the G protein G_t. The α-subunit of transducin (G_t) hydrolyzes GTP, but the rate of catalysis is extremely slow. The half-life for GTP bound to purified G_t is approximately 15 s, but when you close your eyes, the image disappears without a perceptible delay. How can signal transduction be terminated so quickly in the intact cell if G proteins turn themselves off at such a slow rate? This question prompted the search for GTPase activating proteins (GAPs) that could accelerate the rate of GTP hydrolysis by G proteins, and eventually led to the discovery of a group of proteins known as RGS proteins. RGS proteins have a conserved sequence consisting of approximately 130 amino acids

that is known as the RGS box, and they have been divided into subfamilies on the basis of the homology within this domain. There are at least 20 mammalian RGS proteins, and many of them act as GAPs for the $G_{i/o}$, and $G_{q/11}$ subfamilies of G proteins. Notably, no RGS protein capable of acting as a GAP for G_s has been identified, and not all RGS proteins have GAP activity. RGS proteins that act as GAPs are capable of accelerating the GTPase activity of G proteins by as much as 1000-fold, but they do not directly participate in the hydrolysis of GTP. Phosphorylation of G proteins can alter their susceptibility to the GAP activity of RGS proteins. Several RGS isoforms have palmitate covalently attached to cysteine residues, which provide a hydrophobic moiety that may be important for their association with membranes. Many RGS proteins also have sequences flanking their RGS box that contain various domains important for intermolecular interactions with proteins in addition to $G\alpha$. For example, members of the RGS9 subfamily have a 64 amino acid domain that resembles the C terminus of $G\gamma$ subunits referred to as a $G\gamma$-like (GGL) domain. The GGL domain is responsible for binding to the brain-specific $G\beta_5$ subunit, and *in vivo*, $G\beta_5$ evidently exists as a complex with RGS9. There is evidence that individual RGS proteins regulate selected signaling pathways, though relatively little is known about this aspect of RGS proteins. Two effector proteins, phospholipase $C\beta$ (Sect. 5.4) and RhoGEFs (Sect. 5.7), have been shown to serve as RGS proteins for the respective G proteins which activate them, G_q and $G_{12/13}$, and the effector, cGMP phosphodiesterase (PDE)(Sect. 5.1), greatly enhance the ability of RGS9 to serve as a GAP for the G protein G_t.

5 Effectors

5.1 cGMP Phosphodiesterase

A mechanism for detecting light is one of the indispensable elements needed for vision. In mammals, this function is performed by specialized cells called rods and cones that are located in the retina. The process of detecting light is referred to as phototransduction to reflect the fact that it is a signal transduction process, and it requires the participation of a heterotrimeric G protein. Because the G protein responsible for mediating phototransduction was one of the first proteins to be recognized as fulfilling the role of the "transducer" that was proposed by Martin Rodbell, it came to be called transducin (abbreviated here as G_t). The 7TM receptor responsible for activating G_t is rhodopsin, and the effector that is regulated by G_t is cyclic GMP phosphodiesterase (PDE). The process of phototransduction, including information about the structure, function, and regulation of PDE is discussed in greater detail in Sect. 7.2.

5.2 Adenylyl Cyclase

Adrenergic receptors are among the 7TM receptors that regulate the activity of AC. The importance of adrenergic signaling in cardiovascular function and disease has made AC one of the most intensively studied G protein–regulated effectors. The activity of AC can either be increased or decreased depending upon the G protein–mediated pathway that is activated. 7TM receptors like the prototypical β_2-adrenergic receptor activate G_s leading to stimulation of AC, while 7TM receptors

such as the α_2-adrenergic receptors activate G_i causing inhibition of AC. There are nine different mammalian isoforms of AC and they are all activated by G_s and inhibited by G_i. The ability of G_s to activate ACs was initially attributed entirely to determinants on the α-subunit. However, it is apparent that the activation of some ACs, specifically type II and type IV, also require the presence of the $G\beta\gamma$ heterodimer.

AC is an integral membrane protein with two putative membrane-spanning domains, called M1 and M2. Each of these domains is predicted to possess six membrane-spanning helices. M1 and M2 are connected by a large intracellular domain called C1, and AC has a large intracellular C-terminal domain referred to as C2. Together C1 and C2 form the catalytic domain that is regulated by G proteins. In addition to being regulated by G proteins, ACs are activated by the diterpene forskolin from the root of the plant *Coleus forskohlii*, and inhibited by certain purinergic compounds such as adenosine that bind to the so-called P-site on ACs. Some isoforms of AC, notably types I and VIII, are activated by Ca^{2+}-calmodulin.

When activated, AC catalyzes the production of cAMP, and increased intracellular levels of this second messenger cause a concomitant increase in protein kinase A (PKA) activity. Subsequently, activated PKA phosphorylates many protein substrates, including 7TM receptors, thereby regulating their activity. For each step in this cascade, there is an opposing reaction that enables the pendulum to swing both ways. Desensitization occurs when the active conformation of the receptor is phosphorylated by a GRK leading to arrestin binding and a consequent attenuation of the receptor's ability to activate its cognate G protein (Sect. 3.2). The intracellular concentration of cAMP is reduced when it is hydrolyzed by PDE, and the products of reactions catalyzed by PKA become the substrates for dephosphorylation by protein phosphatases.

5.3 Phosphoinositide 3-Kinase

Phosphoinositide 3-kinases (PI3Ks) phosphorylate the 3′-OH on the inositol ring of phosphoinositides. They are inhibited by the drug LY294002 and by the fungal metabolite wortmannin, and can be categorized as class I, II, or III. Class I PI3Ks is further subdivided into classes IA and IB. PI3Kα, β, and δ constitute class IA and their respective p110α, β, and δ catalytic subunits can bind any of the seven p85 adaptor subunits that range in molecular weight from 50 to 85 kDa. PI3Kγ is the only member of class IB and is most abundant in leukocytes in which it is located on or juxtaposed to the plasma membrane. PI3Kγ differs from its class IA counterparts in that its p110γ catalytic subunit cannot bind to p85 adaptor subunits. Instead, p110γ binds to an adaptor subunit designated p101 that has no known homologs. As with certain isoforms of AC (Sect. 5.2), there is evidence for roles of both $G\alpha$ and $G\beta\gamma$ in the activation of classes IA and IB PI3Ks. PI3Ks in turn regulate the activity of many proteins including protein kinase B (PKB/Akt), protein kinase Cγ, c-Jun N-terminal kinase (JNK) and L type Ca^{2+} channels.

5.4 Phospholipase Cβ

There are 11 different isoforms of phospholipase C (PLC) and they are subdivided into four groups designated β, γ,

δ, and ε. All PLC isoforms consist of a single polypeptide chain that catalyzes the hydrolysis of phosphotidylinositol 4,5-bisphosphate (PIP_2) producing inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG). IP_3 triggers an increase in intracellular Ca^{2+}, and together, DAG and Ca^{2+} activate "classical" protein kinase C (PKC) isoforms, or DAG by itself can activate "novel" PKC isoforms. Many PKC substrates have been identified and phosphorylation by PKC leads to a variety of physiological consequences. Four of the PLC isoforms belong to the β-subfamily (PLCβ1 through PLCβ4), and each PLCβ isoform can be activated by 7TM receptors that couple to G proteins belonging to the $G_{q/11}$ subfamily. PLCβ1 is a GAP for $G_q\alpha$ (Sect. 4.3), which binds to the α-subunit with a K_D of about 2 nM, and increases the rate of GTP hydrolysis by approximately 1000-fold.

5.5 G Protein–coupled Inwardly Rectifying Potassium (Kir3) Channels

Inwardly rectifying K^+ channels belonging to the Kir3 family are critical for the regulation of resting membrane potential in excitable and nonexcitable cells and they are regulated by 7TM receptors. There are four types of Kir3 subunits designated Kir3.1 through Kir3.4. These are assembled as homo- or heterotetramers, though not all combinations produce functional ion channels. Kir3 subunits have two transmembrane domains, and both the N and C terminus are on the cytoplasmic side of the cell membrane. Direct regulation of Kir3 by G proteins is a membrane-delimited event that occurs on a millisecond timescale, and it is mediated by determinants on the G$\beta\gamma$ heterodimer rather that the Gα subunit. A number of studies have demonstrated that little specificity exists *in vitro* when G$\beta\gamma$ heterodimers of different subtype composition are used to regulate effectors such as Kir3. Although Kir3 can be activated by G$\beta\gamma$ alone, the channel also binds Gα. The N-terminal domain of the Kir3 channels binds to both Gα and G$\beta\gamma$, and the channel's C-terminal domain has additional binding sites for G$\beta\gamma$. Kir3 channels are regulated by the G_i subfamily of G proteins, and pertussis toxin (Sect. 4.2) blocks the 7TM receptor-mediated opening of Kir3 channels. 7TM receptors that activate both G_s and G_q can also regulate Kir3 channels. Regulation by receptors that activate G_s may be direct or it may be a consequence of PKA activation brought about by the G_s-mediated stimulation of AC (Sect. 5.2). PIP_2 is important for Kir3 channel activity, and 7TM receptors that activate PLC (Sect. 5.4) via a G_q-mediated mechanism can attenuate channel opening by reducing the concentrations of PIP_2 in the membrane.

5.6 N and P/Q Type Calcium Channels

Voltage-gated Ca^{2+} channels are composed of five different subunits (α_1, α_2, β, γ, and δ) and are subdivided into five different types (L, N, P/Q, R, and T). The α_1 subunit has four similar hydrophobic domains (I–IV) each consisting of six transmembrane helices. Domains I through IV are connected by intracellular loops, and the α_1 subunit has a relatively large intracellular C-terminal domain. The Ca^{2+} channel β-subunit interacts with the intracellular loop between domains I and II of the α_1 subunit as well as the N and C terminus. The β-subunit is important for transport of the channel from the endoplasmic reticulum (ER) to the plasma

membrane, and for both the kinetic and voltage-dependant properties of the channel. N and P/Q type Ca^{2+} channels are also directly inhibited by G_o which antagonize the effects of the Ca^{2+} channel β-subunit. It is the $G\beta\gamma$ heterodimer that is responsible for regulating the activity of the channel, though $G\alpha$ has also been reported to bind to the α_1 channel subunit. There is evidence that the N and C termini of α_1 subunit is involved in Ca^{2+} channel regulation by $G\beta\gamma$, and $G\beta\gamma$ binds to the intracellular loop between domains I and II of α_1. Though this latter interaction does not preclude the binding of the Ca^{2+} channel β-subunit to α_1, it does appear to alter the orientation between the two channel subunits providing a possible clue regarding the mechanism of G protein–mediated regulation of Ca^{2+} channels. Apparently there are also differences in the ability of $G\beta\gamma$ to regulate Ca^{2+} channels depending upon the subtype of the channel β-subunit, and the subtype of $G\beta$ that is involved in regulating the channel.

The α_1 subunit of Ca^{2+} channels can be phosphorylated by PKC (Sect. 5.4). PKC-mediated phosphorylation of the loop between domains I and II of the α_1 subunit relieves the $G\beta\gamma$-induced inhibition of N type Ca^{2+} channels. The ability of PKC to target the α_1 subunit may involve interaction with determinants on the $G\beta\gamma$ that is bound to the channel subunit. Thus, there are two modes by which G proteins are involved in regulating N type Ca^{2+} channels; a direct inhibition of the channel by activated G_o, and an indirect mechanism for alleviating the $G\beta\gamma$-mediated inhibition through 7TM receptors that can activate PKC. Ca^{2+} channels that are not directly regulated by $G\beta\gamma$ subunits (*i.e.* L type) are, however, susceptible to indirect regulation by G protein–mediated pathways. The α_1 subunits of L type Ca^{2+} channels as well as the channel β-subunit are phosphorylated by PKA (Sect. 5.2) resulting in increased channel activity. A protein known as A kinase anchoring protein (AKAP – see Sect. 6.3.1) facilitates modulation of channel activity by PKA. Furthermore, a peptide designed to ablate the binding of PKA to AKAP inhibits the regulation of L type Ca^{2+} channel by 7TM receptor-mediated activation of PKA. These data indicate that localization of PKA by virtue of being associated with other proteins is important for Ca^{2+} channel regulation by the kinase.

5.7 RhoGEFs

Rho is a member of a family of "small" monomeric G proteins that is involved in regulating the architecture of the actin cytoskeleton, and consequently is important for cell motility. Aberrant activation of pathways regulated by Rho can also convert cells from a normal to a transformed phenotype. Rho has a guanine nucleotide binding site, and it is activated when GTP is bound. Whether GDP or GTP is bound to Rho is controlled by guanine nucleotide exchange factors (RhoGEFs) that promote the exchange of GDP for GTP, GTPase activating proteins (GAPs) that accelerate the intrinsic GTPase activity of Rho, and guanine nucleotide dissociation inhibitors (RhoGDIs) that inhibit the dissociation of GDP. The motif in RhoGEFs responsible for catalyzing nucleotide exchange is the Dbl homology (DH) domain. Three RhoGEFs, p115RhoGEF, PDZ-RhoGEF, and leukemia-associated RhoGEF (LARG), are regulated by heterotrimeric G proteins of the $G_{12/13}$ family. Each of these RhoGEFs contains an RGS domain (Sect. 4.3) that serves both as a site for the interaction of $G_{12/13}$ with the RhoGEFs,

and to terminate activation by $G_{12/13}$. The mechanism by which $G_{12/13}$ activates the GEF activity of RhoGEF has not been established. At present the favored hypothesis is that inhibition of the GEF activity is relieved when the activated $G_{12/13}$ interacts with the RGS domain of RhoGEF, allowing the DH domain to catalyze nucleotide exchange on Rho. Rho can also be activated by 7TM receptors that activate G_q, but RGS containing RhoGEFs are not involved in these pathways.

5.8 Mitogen-activated Protein (MAP) Kinases

Receptors for mitogenic substances such as growth factors (*e.g.* EGF, platelet-derived growth factor, nerve growth factor, etc.) have intrinsic tyrosine kinase activity, and are referred to as receptor tyrosine kinases (RTK). RTK function as homodimers composed of subunits that have a single transmembrane domain. Activation occurs when an agonist bound to one RTK subunit causes phosphorylation of tyrosine residues in the cytoplasmic domains of its dimeric partner (trans-autophosphorylation). Phosphotyrosine residues in RTK provide binding sites for proteins with SH2 domains, and this facilitates the assembly of a protein complex that initiates the phosphorylation of numerous proteins including a 42-kDa protein originally designated pp42. The discovery that pp42 was itself a serine/threonine kinase resulted in its being renamed mitogen-activated protein (MAP) kinase. This particular MAP kinase is one member of a family of proteins that now include about 20 different mammalian homologs. Despite the fact that many members of this family play roles in the response of eukaryotic cells to environmental factors other than mitogens, they are all referred to as mitogen-activated protein kinases. MAP kinases can be divided into subfamilies, and the subfamily known as extracellular-signal regulated kinases (ERK) claims ERK1, alias pp42, as a member. MAP kinases are activated when both threonine and tyrosine in a T-X-Y sequence are phosphorylated by a kinase designated MAP kinase kinase (MAPKK). The MAPKK is activated when it is phosphorylated by a MAP kinase kinase kinase (MAPKKK). Together a MAPKKK, a MAPKK, and a MAP kinase constitute a module, and as a rule appear to be held in physical proximity by a scaffolding protein (Sect. 6.3.1).

In addition to regulation by RTKs, MAP kinase modules can be activated by agonists for 7TM receptors. One of the most extensively studied G protein–mediated signaling pathways activates a MAP kinase module in yeast that regulates their mating behavior. Mutations in numerous genes render yeast sterile (*i.e.* unable to mate), and many of these gene products bear the designation "Ste" to reflect this fact. Peptide pheromones secreted by yeast bind to 7TM receptors (Ste2 and Ste3) triggering the fusion of two compatible haploid individuals. These pheromone receptors are coupled to the single G protein expressed in yeast. The G protein in turn interacts directly with several proteins, including a kinase (Ste20) belonging to the family of p21-activated kinases (PAKs), and two proteins, Far1 and Ste5, that serve as scaffolds for the assembly of protein complexes. In addition to interacting with the yeast G protein, Ste20, Far1 and Ste5 interact with each other indirectly through their associations with other proteins, creating a complex network of protein–protein interactions. The extensive interactions between the yeast G protein and downstream components make it difficult to point to a single protein as the effector in this

pathway, but the result is the activation of Ste20 and a MAP kinase module consisting of Ste 11 (MAPKKK), Ste7 (MAPKK), and Fus3/Kss1 (MAP kinases).

In higher eukaryotes, Raf (*i.e.* cRaf-1 or B-Raf), MEK (*i.e.* MEK1 or MEK2), and ERK (*i.e.* ERK1 or ERK2) are the respective MAPKKK, MAPKK, and MAP kinase of a module that is regulated by both 7TM receptors and RTK. Activation by RTK involves the agonist-induced trans-autophosphorylation described earlier, which initiates a cascade of events that leads to ERK activation (Fig. 4). Phosphotyrosine residues on the RTK allow the receptor to bind proteins with SH2 domains such as Shc (Src homology/collagen). The phosphorylation of tyrosine residues in Shc provides a binding site for the SH2 domain of the protein Grb2 (growth factor-receptor bound 2), which brings with it a protein called *Sos (son of sevenless)*. Sos is a guanine nucleotide exchange factor for the small G protein Ras, and the GEF activity of Sos activates Ras by facilitating the exchange of bound GDP for GTP. Activated Ras in turn activates Raf leading to ERK activation. The regulation of ERK by 7TM receptors is less well defined, and apparently can occur via several distinct pathways depending upon the particular 7TM receptor and the type of cell. It is possible for 7TM receptors to regulate MAP kinase without involving G proteins (Sect. 3.3), though the focus here will be on the pathways that do involve G proteins.

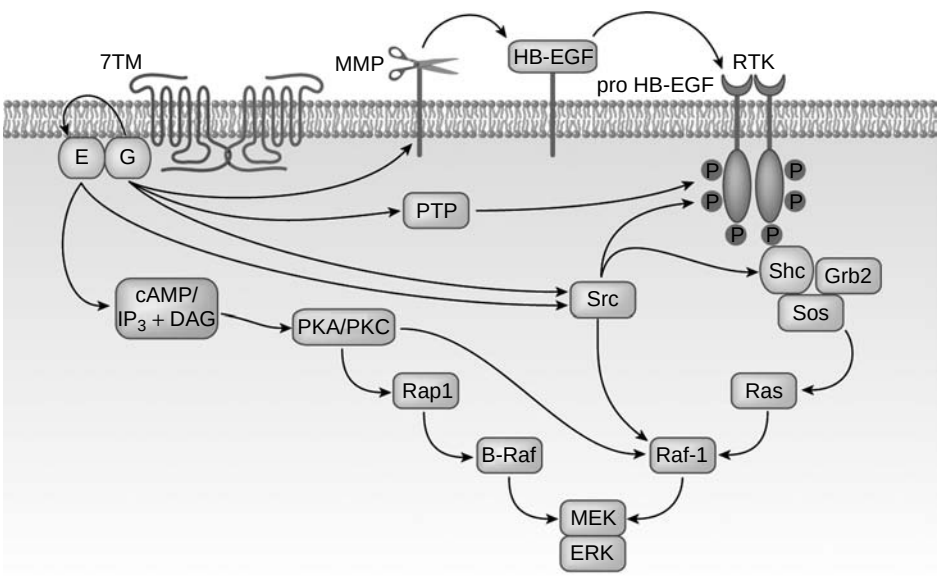

Fig. 4 Pathways leading to the activation of MAP kinase are summarized (see Sect. 5.8 for an explanation). The 7TM receptor is shown as a dimer. cAMP, IP_3, and DAG are second messengers; E, effector; ERK, MAP kinase; G, heterotrimeric G protein; Grb2 and Shc, adaptor proteins; HB-EGF, heparin binding epidermal growth factor; MEK, MAP kinase kinase; MMP, matrix metalloprotease; PKA, protein kinase A; PKC, protein kinase C; PTP, protein tyrosine phosphotase; Raf-1 and B-Raf, MAP kinase kinase kinases; Rap 1 and Ras, small monomeric GTP binding proteins; RTK, receptor tyrosine kinases; Src, tyrosine kinases; Sos, guanine nucleotide exchange factor for Ras.

A simplified scheme for many of these pathways is shown in Fig. 4.

If the regulation of MAP kinase activity by a 7TM receptor requires the participation of an RTK, this process is known as "transactivation." For example, activation of a 7TM receptor may result in the production of an RTK agonist leading to the conventional agonist-mediated regulation of ERK. This involves the activation of a matrix metalloprotease (MMP). There may be more than one pathway by which 7TM receptors activate MMPs, and the intermediate steps have not been elucidated. Once activated, the protease cleaves proHB-EGF to liberate heparin binding EGF (HB-EGF) from the extracellular surface. HB-EGF is an EGF receptor agonist, and binding to its receptor initiates the cascade of events resulting in ERK activation. Alternatively, transactivation can occur by an agonist-independent activation of the RTK. One likely mechanism involves a 7TM receptor-mediated inhibition of protein tyrosine phosphotase (PTP) that effectively favors the activated state of the RTK by preventing dephosphorylation of phosphotyrosine residues in the receptor.

7TM receptors also regulate MAP kinase activity through pathways that involve the Src family of nonreceptor tyrosine kinases. Regulation has been tied to 7TM receptors that activate G_s, G_i, and G_q, and may or may not involve second messenger-dependent events. Phosphorylation of Src increases its kinase activity, and activated Src can phosphorylate Raf, RTK, and Shc. Src-mediated phosphorylation of RTK and Shc enable them to initiate the Ras-dependent activation of Raf, thus contributing to ERK activation, and although GTP-liganded Ras is required for Raf activation, Src-mediated phosphorylation of Raf also contributes to Raf activation.

6 The Organization of G Protein-mediated Signal Transduction Systems

6.1 The Classical Model of G Protein–mediated Signal Transduction

Many of the components involved in G protein–mediated signal transductions are integral or peripheral membrane proteins; thus, the initial steps in signal transduction are membrane-delimited events. Heterotrimeric G proteins were the first of the signaling proteins to be purified, and their activation in solution leads to the dissociation $G\alpha$ from the $G\beta\gamma$ heterodimer. The model that emerged depicted signaling proteins as moving about independently of one another in the membrane with signal transduction being propagated by a series of transient bimolecular interactions. The process is initiated when an agonist binds to its 7TM receptor causing it to adopt an active conformation that can then bind its cognate heterotrimeric G protein, forming a "ternary complex." The activated 7TM receptor enables GDP to dissociate from $G\alpha$ allowing GTP to bind in its place. This leads to the dissociation of the G protein subunits from each other, and from their 7TM receptor. Subsequently, $G\alpha$ and $G\beta\gamma$, either independently or in concert with one another, regulate the activity of specific effector proteins that they encounter as they move about in the plane of the membrane. When the GTP bound to $G\alpha$ is hydrolyzed, the subunit becomes inactive and its affinity for $G\beta\gamma$ is increased allowing the formation of an inactive heterotrimeric G protein, thus completing the cycle.

When the model for the propagation of signals mediated by G proteins were first being developed, the number of different

components known to be involved in the process was relatively small. By 1980, two G proteins (G_t and G_s), a few receptors, and two effectors (PDE and AC) had been identified, and the concept that G protein–mediated signal transduction was propagated as the result of transient bimolecular interactions provided a satisfactory explanation of the mechanics of signal transduction. Subsequently, there was an explosion in the discovery of G protein–signaling components so that now there are hundreds of 7TM receptors, dozens of different G protein subunits, and numerous effector proteins. Along with an appreciation for the large number of components involved in G protein–mediated signal transduction has come the knowledge that individual cells contained multiple G protein–mediated signaling pathways. This fact is exemplified by the rat portal vein myocyte, which has at least 11 different receptors capable of activating G proteins, 7 different $G\alpha$ subunits, 5 different $G\beta$ subunits, 8 different $G\gamma$ subunits, and 6 different effectors whose activities are regulated by G proteins.

Despite the presence of dozens of different signaling components within a single cell, G protein-mediated signal transduction generally displays a high degree of fidelity that cannot be explained by the classical model. For example, the subtype composition of a G protein (*i.e.* the subtype of $G\alpha$, $G\beta$, and $G\gamma$) determines what 7TM receptor a heterotrimeric G protein couples to. It is also known that individually, $G\alpha$ and $G\beta\gamma$ have a low affinity for 7TM receptors; consequently, they must be associated as a heterotrimer before interacting with their cognate receptor. If G protein subunits dissociate from each other and from their receptor when activated, then subunit reassociation must occur before the G protein can interact with the receptor again. $G\alpha$ subunits have comparable affinities for $G\beta\gamma$ heterodimers of any subtype composition. Since most cells have $G\beta\gamma$ heterodimers in many different subtype combinations, it is unlikely that G protein subunits would reassociate to produce a heterotrimer with the same subunit composition as that originally activated by the receptor. This would produce a situation within the cell in which the pool of G protein capable of coupling to a given 7TM receptor would be in constant flux.

Another enigma associated with the classical model involves the activation of effectors by G proteins. Although it was originally thought that only $G\alpha$ subunits were responsible for regulating the activity of effectors, it later became apparent that $G\beta\gamma$ heterodimers were also involved in propagating signals. The first effector recognized as being regulated by $G\beta\gamma$ was the inwardly rectifying potassium channel. Subsequently, it was discovered that a number of other effectors were also regulated by $G\beta\gamma$. In atrial cardiomyocytes, activation of the β_2-adrenergic receptor causes G_s-mediated stimulation of AC (Sect. 5.2). Kir3 channels (Sect. 5.5) in these cells are opened by the $G\beta\gamma$ heterodimer in response to the activation of G_i by M2-muscarinic receptors. If subunits dissociate when G proteins are activated, and signal transduction occurs by a mechanism involving random collisions, then the $G\beta\gamma$ heterodimer produced when G_s is activated by the β_2-adrenergic receptor should also open Kir3 channels. However, this does not happen indicating that the cell has a way of ensuring that there is specificity with regard to G protein–mediated signaling.

The foregoing considerations raise caveats regarding the concept that proteins involved in G protein–mediated signaling

can move freely throughout the plasma membrane. The question now becomes: How does the cell control what signaling proteins interact with each other? The idea that biochemical pathways can be isolated is a long-standing notion. Historically, the term "compartmentation" has been used to describe the containment of a pathway by a subcellular organelle or microdomains within an organelle. Another effective means of compartmentation can be provided by the assembly of complexes through specific protein–protein interactions that include some components while excluding others. Compartmentation of G protein pathways within organelles, microdomains or by assembly into protein complexes is undoubtedly important for G protein–mediated signaling.

6.2 Compartmentation

Membrane microdomains identified as being sites where signaling components are concentrated include detergent-insoluble glycosphingolipid/cholesterol-rich rafts (DIGs), and small cavelike plasma membrane invaginations called caveolae. DIGs and caveolae are functionally defined as lipid microdomains that are resistant to dissolution in detergent. Among the signaling components found in these detergent-resistant membranes (DRM) are 7TM receptors, heterotrimeric G proteins, G protein–regulated enzymes and their substrates, and receptor and nonreceptor tyrosine kinases. In some cases, disruption of DRM can block signal transduction indicating that the integrity of these microdomains can be important in signal transduction. However, whether DRM contribute to the fidelity of signaling by segregating one signal transduction pathway from another remains to be determined.

6.3 Signal Transduction Complexes

6.3.1 Scaffolding Proteins

Colocalization of signaling molecules can occur through protein–protein interactions that produce molecular complexes. One mechanism of forming such complexes is to assemble them on the backs of scaffolding proteins. A distinguishing feature of many scaffold proteins is the presence of characteristic domains, often several of them, that are capable of binding to other proteins. Among the scaffolding domains that are important in the assembly of G protein–mediated signaling pathways are PDZ domains, PH domains, Src homology 2 and 3 (SH2 and SH3) domains, and GGL domains (Sect. 4.3). A number of proteins serve as scaffolding proteins for components of G protein–mediated signaling pathways.

Caveolins are proteins that are found in caveolae. The three isoforms, caveolin-1, -2, and -3, are 20 to 24 kDa integral membrane proteins whose N and C termini project into the cytosol. Caveolins serve as scaffolds that have the ability to bind many signal transduction proteins including G proteins, certain isoforms of AC, GRK2, PKA, PKC, the EGF receptor, Src, and Ras. In addition to scaffolding signaling molecules, caveolins can affect the activity of many of these proteins, and the interaction between caveolins and some signaling molecules can be regulated. The binding of caveolin to G proteins is influenced by the guanine nucleotide that is bound to $G\alpha$, and peptide fragments of caveolin can inhibit the steady state GTPase activity and

GTP binding to purified heterotrimeric G proteins suggesting the possibility of a direct interaction between caveolin and G proteins. Caveolins inhibit the activity of kinases involved in G protein–mediated signaling. Caveolin binds to the GRK N terminus, and inhibits GRK-mediated phosphorylation of 7TM receptors. PKA is also inhibited by caveolin.

The cytoskeleton can serve as a scaffold for proteins involved in G protein–mediated signaling, and numerous studies have demonstrated that disrupting the cytoskeleton affects signaling by G proteins. Tubulin, the principal component of microtubules, constitutes a large percentage of the protein associated with synaptic plasma membranes. The first indication that cytoskeletal proteins interact with components of G protein–mediated signal transduction pathways came with the observation that G_s could be washed from synaptic plasma membranes after treating them with microtubule-disrupting agents. Heterotrimeric G proteins bind to tubulin, as do $G\alpha$ and the $G\beta\gamma$ heterodimer individually, and the binding persists when nonhydrolyzable GTP analogs are bound to $G\alpha$. It was also discovered that tubulin can serve as a source of guanine nucleotides for the activation of G proteins. Cytoskeletal proteins interact with, and in some cases appear to regulate the activity of $G\alpha_s$, $G\alpha_{i1}$, $G_{q/11}$, PLCβ, and GRK. G proteins in turn, can regulate GTP binding to tubulin, GTP hydrolysis by tubulin, and tubulin polymerization.

AKAPs (A kinase anchoring proteins) are a structurally and functionally diverse group of proteins that share the ability to bind PKA (Sect. 5.2). PKAs are a heterotetrameric proteins designated as type I or type II depending upon the kind of regulatory subunit, type RI or RII, respectively, that they contain. Most of the known AKAPs bind type II PKA though several AKAPs capable of binding type I PKA have also been identified. AKAPs serve to target PKA to specific subcellular compartments, and disruption of the AKAP/PKA complex inhibits the cAMP-mediated effects of the kinase. In addition to PKA, many AKAPs bind other proteins that are involved in G protein–mediated signaling, including 7TM receptors, GRK, PKC, phosphatases, and β-arrestin. AKAPs have also been shown to regulate the activity of signaling proteins, for example, suppressing the activities of both phosphatases and kinases, and to anchor signaling complexes to the cytoskeleton as well. The ability of AKAPs to bind many signaling proteins simultaneously makes them ideal candidates for organizing entire pathways. Given their diversity, they may be able to selectively assemble the components of a particular signaling pathway, thus segregating it from others, and in so doing contribute significantly to the specificity of signal transduction. The feasibility of this hypothesis is suggested by data showing that only AKAP79/150 is coprecipitated with the β_2-adrenergic receptor from a detergent extract of rat brain that contain several AKAPs, but AKAP79/150 is not precipitated with α_{2A}-adrenergic receptors, AT_1-angiotensin receptors, or vasoactive intestinal polypeptide receptors.

Recent studies using both genetic and proteomic approaches have shown that many of the components used by *Drosophila* for detecting light are part of a large protein complex. Phototransduction in *Drosophila* is initiated by the light-induced isomerization of retinal converting the photoreceptor rhodopsin to metarhodopsin. This conformational change in the photoreceptor leads to activation of G_q and subsequently PLCβ

(Sect. 5.4). Activated PLCβ eventually triggers the opening of transient receptor potential (TRP) and TRP-like (TRPL) channels, and consequently leads to depolarization of the photoresponsive cell in approximately 20 ms. A number of proteins including PLCβ, calmodulin and a calmodulin binding protein called neither-inactivation or afterpotential C (*NINAC*) play a role in deactivation, which occurs in less than 100 ms. PLCβ, PKC, TRP, TRPL, NINAC, calmodulin, and rhodopsin are assembled into a complex by binding to the scaffolding protein inactivation no afterpotential D (INAD). INAD contains five distinct PDZ domains. Calmodulin binds to a region between the first and second PDZ domain of INAD, and PLCβ, PKC, TRP, TRPL, and NINAC bind to one or more of INAD's PDZ domains. INAD is critical for the efficient activation and deactivation of phototransduction in *Drosophila* and the evidence is consistent with its doing so by virtue of its ability to bring the various relevant signaling molecules into proximity by scaffolding them in what has been referred to as a tranducisome, signalsome or signalplex. There is evidence that the proteins involved in the mammalian phototransduction pathway are also scaffolded. A more extensive discussion of the signaling complex in mammalian visual system is presented in Sect. 7.2.

Juxtaposition of signaling components is important for both efficiency (speed of response) and specificity in the activation of MAP kinase cascades. In yeast, the MAP kinase cascade responsible for triggering mating (Sect. 5.8) is one of several MAP kinase pathways that are segregated from each other by virtue of being scaffolded into macromolecular complexes. In mammals, several proteins (JNK-interacting protein 1, MEK partner 1, and MEKK1) have been identified as candidates for scaffolding proteins that organize MAP kinase cascades. β-arrestin, in addition to its role in 7TM receptor desensitization (Sect. 3.2), plays a pivotal role in 7TM receptor-mediated MAP kinase activation. β-Arrestins scaffolds JNK3 (c-Jun amino-terminal kinase 3, a MAP kinase family member), MKK4 (a MAPKK), and ASK1 (apoptosis signal-regulating kinase 1, a MAPKKK) as well as Src, Raf, and ERK (Sect. 5.8). Indeed, there may be large signaling complexes containing dozens of proteins that include the components of both MAP kinase cascades and G protein–coupled signal transduction systems.

6.3.2 Complexes Containing Receptors, G Proteins, and Effectors

Scaffolding proteins can contribute to the efficacy of signal transduction by colocalizing many of the signaling molecules needed for the propagation of a signal as well as those that are responsible for terminating the signaling once the message has been received. Protein–protein interactions between the individual components of the pathway contribute in a similar way to the efficacy and fidelity of signal transduction. Despite the prevailing hypothesis that 7TM receptors, their cognate heterotrimeric G proteins, and the effectors that they regulate undergo transient bimolecular interactions, the literature of the last 30 years is peppered with data indicating that these proteins associate to form stable complexes.

One concept that has been challenged in recent years is the notion that 7TM receptors are present in cell membranes as monomers. There is now convincing biochemical, pharmacological, and biophysical evidence that 7TM receptors exist as oligomers (probably dimers) *in vivo*. Both homodimeric and heterodimeric complexes of 7TM receptors have been

identified, and heterodimeric complexes between different classes of 7TM receptors are possible. Studies involving the γ-amino butyric acid B ($GABA_B$) receptor revealed one of the functional consequences of receptor dimerization. There are two subtypes of the $GABA_B$ receptor (gb1 and gb2). Cells that express only one subtype do not produce a functional $GABA_B$ receptor, and gb1 when expressed by itself is retained in the ER. In contrast, cells that express both gb1 and gb2 produce functional heterodimeric $GABA_B$ receptors that are correctly transported to the plasma membrane, indicating a role for heterodimerization in the trafficking of receptors. Homodimerization of 7TM receptors may also play a role in the transport of receptors out of the ER. Other consequences of dimerization are that a particular 7TM receptor may display a different pharmacology, or couple to a different G protein depending upon the subtype of its dimeric partner. Not all 7TM receptors form dimers with other 7TM receptors. Two 7TM receptors, the calcitonin receptor and the related calcitonin receptor-like receptor, form heterodimers with receptor activity modifying proteins (RAMPs) that affect their pharmacology.

The concept that subunit dissociation occurs when heterotrimeric G proteins are activated also deserves scrutiny. The hypothesis that G protein subunits dissociate when activated is based almost exclusively on the behavior of purified G proteins in solution of detergent where activation requires incubating them with nonhydrolyzable GTP analogs such as GTPγS and unnaturally high concentrations (10–120 mM) of a divalent cation, usually Mg^{2+}. The high concentrations of Mg^{2+} serve to promote the dissociation of tightly bound GDP from Gα so that a GTP analog can bind in its place. In this way, Mg^{2+} is used to perform the function of a 7TM receptor. In addition to activating the G protein, these conditions cause Gα to dissociate from G$\beta\gamma$. However, high concentrations of Mg^{2+} cause G protein subunit dissociation in the absence of guanine nucleotides as well as in the presence of GDP, which does not activate the G protein, indicating that G protein subunit dissociation is not necessarily a consequence of activation. Activation of G proteins in solution under more nearly physiological conditions does not cause subunit dissociation, and experiments have demonstrated that the physical dissociation of G protein subunits is not necessary in order for them to regulate the activity of effectors *in vivo*. An activated heterotrimer G_s, rather than its dissociated subunits, is responsible for stimulating AC in cell membranes, and G_i subunits remain associated when activated in intact cells. These data together with data indicating that most effectors, including some isoforms of AC, PDE, Kir3, PLCβ, and PI3K, are regulated by or are at least interacting with both Gα and G$\beta\gamma$ make it increasingly likely that subunit dissociation does not occur, as a rule, when G proteins are activated *in vivo*.

There are many examples in which 7TM receptors and G proteins copurify. It is also apparent that a G protein can regulate the activity of an effector while remaining physically associated with its receptor. Studies have demonstrated that 7TM receptors are associated with their cognate heterotrimeric G protein *in vivo*, both before and during signal transduction. There is also evidence that G proteins form a complex with their cognate effectors regardless of whether or not the G protein is activated. Kinetic data support the existence of complexes containing both G proteins and effector *in situ*. Given data

that 7TM receptors form complexes with G proteins and that G proteins form complexes with their effectors, it is not surprising that stable complexes of 7TM receptors and effectors have been detected. The persistent association of G proteins with 7TM receptors, the stable association of G proteins with effectors, evidence that G protein subunits remain associated when the transducer is activated, and data showing that 7TM receptors are part of a complex with their cognate effectors *in vivo* are all consistent with the existence of stable signaling complexes containing 7TM receptors, G proteins, and effectors. These complexes are likely to include scaffolding proteins (Sect. 6.3.1) as well as other proteins that regulate signal transduction. For example, even though RGS proteins (Sect. 4.3) were originally recognized as interacting with certain G proteins, they also have collateral interactions with both receptors and effectors. These interactions are thought to contribute to an increase in signaling efficacy, but may also increase the stability of signaling complexes containing 7TM receptors, heterotrimeric G proteins, and effectors.

With evidence of the existence of signaling complexes has come the question of where these complexes are assembled. In order to be transported from the ER where they are made, to the plasma membrane where they signal, $GABA_B$ receptors must be assembled into heterodimers. A heterodimeric complex of a RAMP and a 7TM receptor (see above) has been reported to occur in the ER, and the formation of homomeric 7TM receptor complexes apparently take place in the ER. Studies indicate that peptide motifs contained in 7TM receptors constitute ER retention signals, and export from the ER requires that these motifs be masked by their association with other proteins. Similar processes appear to be required for the transport of effectors from ER to the plasma membrane. For example, some Kir3 channel subunits combine to form tetramers that are inserted into the plasma membrane, while other combinations are retained in the ER. These observations point to the ER as being the site where at least some assembly takes place. The ER may also be the site of more extensive interactions leading to the formation of signaling complexes containing 7TM receptors, G proteins, and effectors. If the heterotrimeric G proteins are assembled at the same time that receptors and effectors are incorporated into a signaling complex, this may explain how G proteins with a specific combination of $G\alpha$, $G\beta$, and $G\gamma$ subtypes are produced. Although $G\alpha$ subunits, and effectors for that matter, do not discriminate with regard to the subtype composition of the $G\beta\gamma$ heterodimer they interact with, receptors apparently have certain requirements in this regard (Sect. 6.1). Consequently, the presence of a receptor at the time a signaling complex is assembled may be responsible for dictating the subtype composition of the heterotrimeric G proteins that is associated with the signaling complex.

Most 7TM receptor agonists are membrane impermeable so that a response to the agonist requires that the receptors binding site be exposed on the extracellular surface. However, signaling complexes can be found in subcellular compartments other than the plasma membrane. A notable example is the pathway for mammalian phototransduction. Rhodopsin, G_t, and PDE are all located on an organelle within the cell known as the rod outer segment (ROS) disc. Since light can penetrate cell membranes there is no need for rhodopsin to be

in the plasma membrane. Other G protein–mediated signaling pathways occupy intracellular locations. Heterotrimeric G proteins are associated with the cell's mitotic machinery, and are involved in cell division. AGS proteins (Sect. 4.2) appear to be responsible for regulating G protein activity during mitosis though the mechanisms have not been entirely elucidated. The nuclear membrane has been shown to harbor functional G protein–mediated signal transduction pathways that are regulated by 7TM receptors, but it is not clear how, or whether these pathways are regulated by receptor agonists *in vivo*. Signaling complexes may be taken in by endocytosis and transported within the cell after activation as a way of delivering the message to its proper subcellular destination. For example, in neurons activated signaling, complexes translocate within the cell from nerve terminal to nucleus, and these complexes include heterotrimeric G proteins.

7 Signal Transduction: Striking A Balance Between Different Organizational Paradigms

7.1 Insight

Determining the paradigm used by cells to organize G protein-mediated signaling pathways has been an important goal for half a century. Early investigations, highlighted by the seminal work of Alexander Levitzki and Martin Rodbell suggested that these pathways were assembled into complexes of signaling proteins. In contrast, other work indicated that signaling proteins interacted transiently, and these data were predominant in shaping concepts of how signal transduction occurs in the cell. With time, more and more signaling molecules have been discovered, and individual cells are now known to contain many different signaling pathways that employ homologous signaling components with the potential to interact at cross-purposes if signal transduction were to rely entirely on a mechanism involving random collisions between signaling molecules. One consequence has been renewed interest in the hypothesis that G protein-signaling pathways are organized by being assembled into protein complexes, and much of the recent data have supported this hypothesis. Though early evidence for the existence of signaling complexes now seems prophetic, a stable protein complex does not provide the signal amplification that can be attained through transient bimolecular interactions, and this kind of amplification may be necessary for the response to certain stimuli (e.g. light). Thus, the contention between two competing hypothesis persists, and the extent to which transient interactions versus stable protein complexes contribute to G protein-mediated signal transduction remains to be determined. Although G protein-mediated signaling pathways share many characteristics, they must also differ in significant, though perhaps subtle ways including the extent to which they utilize protein complexes and transient interactions to achieve both efficacy and specificity during signaling.

7.2 Mammalian Phototransduction: A Tale of Two Signal Transduction Mechanisms

Within the mammalian retina are found two types of photoreceptor cells called rods and cones. Rod cells serve as sensitive detectors of light when ambient light

levels are low. Cone cells are less sensitive to light, and can be divided into three subtypes on the basis of the 7TM receptors they contain. The receptors in the different cone cells are distinct from one another in that they are activated by different wavelengths of light providing a basis for discerning colors. Rod cells have a structure that is referred to as the rod outer segment (ROS). ROS are easily isolated, and are highly enriched in the components required for phototransduction. These properties helped to make ROS the quintessential experimental system for studying G protein–mediated signal transduction. ROS discs are stacked one on top of another within the ROS, and encapsulated by the plasma membrane of the ROS. By the early 1980s, many of the proteins involved in this signaling pathway had been discovered, and the basic biochemical mechanism of phototransduction was understood. Rhodopsin, G_t, and PDE are bound to the ROS disc membrane so that they are all exposed to the cytoplasm. The ROS plasma membrane contains cyclic GMP-gated cation channels. In the dark, cyclic GMP (cGMP) produced by guanylyl cyclase binds to these channels and opens them allowing an influx of Na^+ ions. In its unactivated state, bovine rhodopsin has 11-cis retinal bound by a Schiff base linkage to lysine residue 296 in the receptor. When a photon causes the isomerization of retinal to the trans-configuration, it triggers a conformational change in rhodopsin (Sect. 3.1) that enables it to activate G_t. G_t subsequently triggers the activation of PDE, which reduces cGMP concentrations below those needed to keep the channels open. This stops the influx of Na^+ (and Ca^{2+}) ions causing the rod cell to become hyperpolarized. When darkness returns, G protein–mediated signaling is terminated so that PDE can no longer be activated by G_t, and the levels of cGMP increase causing the channels to reopen so that the rod cell is depolarized. Several mechanisms have been identified that attenuate the response to light. Some of these are necessary in order for the rod cell to adapt to a wide range of light intensities, while others are important for temporal aspects of the response to light. These mechanisms are not unique to phototransduction, and several have already been described. For example, an RGS protein (RGS9) acts as a GAP for G_t (Sect. 4.3), and activated rhodopsin is susceptible to desensitization by GRK1 and arrestin-1 (Sect. 3.2). Phototransduction is also a representative G protein–mediated signal transduction pathway in that it exhibits the characteristics of two organizational paradigms that appear to be mutually exclusive. On one hand, data supports phototransduction as occurring by a sequence of transient bimolecular interactions, while on the other, there is evidence that the components involved in phototransduction are organized into a signaling complex held together by a web of protein–protein interactions.

The original hypothesis that phototransduction is the consequence of a series of transient bimolecular interactions, first between rhodopsin and G_t and subsequently between G_t and PDE, is based on several observations. Together, rhodopsin, G_t, and PDE in a mole ratio of approximately 100 : 10 : 1 constitute about 90% of the protein present in ROS discs, and may be the only G protein–mediated signal transduction pathway in disc membranes. Thus, there is no need for a mechanism to isolate the phototransduction pathway from other G protein–mediated pathways in order to ensure the fidelity of phototransduction. When ROS discs are exposed to GTP in a low ionic strength buffer, G_t will dissociate from the membranes yielding an almost

pure preparation of the protein. This suggests that the G protein interacts weakly or transiently with the other components involved in phototransduction. Perhaps the most convincing evidence that signaling proteins interact transiently is provided by data showing that a single light-activated rhodopsin molecule can activate G_t at the rate of 10 to 200 s^{-1}.

In contrast, there are data suggesting that the interactions among rhodopsin, G_t, and PDE are not transient. Consider the following: Generally, 7TM receptors interact with heterotrimeric G proteins but only weakly with $G\alpha$ or $G\beta\gamma$ individually (Sect. 6.1). This is true for the interaction of rhodopsin with G_t. Both $G\alpha_t$ and $G\beta\gamma$ together, but not individually, will bind to liposomes or ROS membranes that contain rhodopsin. This is true regardless of whether or not the rhodopsin is light activated. When the nonhydrolyzable GTP analog, Gpp[NH]p, is bound to $G\alpha_t$, it becomes irreversibly activated, yet light-activated rhodopsin can cause the displacement of Gpp[NH]p from $G\alpha_t$ suggesting that $G\alpha_t$-Gpp[NH]p can associate with $G\beta\gamma$ forming a heterotrimer that can also interact with activated rhodopsin. Furthermore, when the nonhydrolyzable GTP analog, GTPγS, is bound to $G\alpha_t$ it can be cross-linked to $G\beta\gamma$ after incorporation into liposomes in the same way as the subunits of the unactivated holo G_t, thus providing additional support for the hypothesis that G_t can exist as an activated heterotrimer in lipid bilayers.

There are data indicating that the interaction of G_t with the effector PDE is not a simple transient interaction. PDE in ROS is a heterotetrameric ($\alpha\beta\gamma_2$) protein consisting of an 88-kDa α, and an 84-kDa β-subunit, each of whose catalytic activities are inhibited by the 13-kDa γ-subunits. Activated G_t binds to PDEγ preventing it from inhibiting the catalytic activity of PDEα and PDEβ. Activated G_t interacts directly with or can be cross-linked to all three of these PDE subunits, and evidence indicates that the $G\alpha_t$-GTP/PDEγ complex remains associated with PDEα and β on the membranes of ROS. $G\alpha_t$-Gpp[NH]p will activate PDE in solution, but its efficacy is greatly enhanced by liposomes or ROS membranes that contain light-activated rhodopsin. Covalent modifications that block $G\alpha_t$ binding to rhodopsin, but do not affect its ability to activate PDE in solution, prevent rhodopsin from enhancing $G\alpha_t$-Gpp[NH]p-mediated activation of PDE. Since the affinity of $G\alpha_t$-Gpp[NH]p for rhodopsin by itself is low, the ability of rhodopsin to enhance the $G\alpha_t$-Gpp[NH]p-mediated activation of PDE is likely to be the result of a simultaneous cooperative interaction between rhodopsin, $G\alpha_t$-Gpp[NH]p and PDE. There is also evidence that unactivated G_t interacts with PDE suggesting that it may be in a complex with PDE in ROS membranes prior to the initiation of phototransduction. The majority of PDE, approximately 70%, is associated with ROS membranes while the remainder is soluble because the C termini of PDEα and PDEβ are associated with PDEδ subunits. Adding recombinant PDEδ causes PDE to be displaced from the ROS membranes with a concomitant loss in light-induced PDE activity, indicating that membrane localization of PDE is important for its activation by G_t.

The regulation of PDE activity is complicated by the fact that deactivation *in situ* occurs much more rapidly than the rate at which purified G_t hydrolyzes GTP. This prompted predictions that a GAP is involved in facilitating the catalytic activity of $G\alpha_t$ (Sect. 4.3). Observations that PDE accelerates the rate of GTP hydrolysis by $G\alpha_t$ led to the suggestion that PDE functions as

a GAP for G_t. Subsequently, it was discovered that a component separate from PDE is needed to accelerate the GTPase activity of $G\alpha_t$ in ROS membranes. The GAP activity of this component, now known to be RGS9, is facilitated by PDEγ, and the evidence suggests that it does so by increasing the affinity of $G\alpha_t$ for RGS9 in an allosteric fashion through binding to nonoverlapping sites on $G\alpha_t$. RGS9 is a tightly bound peripheral membrane protein of ROS, and the GGL domain of RGS9 binds to $G\beta_5$ (Sect. 4.3). RGS9 also binds to, and inactivates ROS guanylyl cyclase. Hence, there is evidence for a complex containing the PDE heterotetramer, $G\alpha_t$, RGS9, $G\beta_5$ and possibly guanylyl cyclase.

The interaction of RGS9 and $G\beta_5$ with PDE and $G\alpha_t$ are an example of collateral interactions that can contribute to complex formation. Other collateral interactions include those between G_t and the protein phosducin that serves to inhibit phototransduction. Phosducin binds holo G_t as well as $G\alpha_t$ and $G\beta\gamma$ individually. When the N-terminal and C-terminal domains of phosducin are produced separately, $G\beta\gamma$ will bind to the N-terminal domain of phosducin unless it is phosphorylated. In the absence of light, PKA mediates phosphorylation of phosducin on serine residue 73. Light-induced reduction of PKA activity leads to dephosphorylation of phosducin. It is proposed that the nonphosphorylated N-terminal portion of phosducin binds $G\beta\gamma$, thus inhibiting G_t-mediated signal transduction. This leads to an attenuation of the cell's response to light. In contrast, the phosphorylated form of phosducin predominates in the dark when PKA is activated, preventing phosducin from interfering with G_t function. The C-terminal portion of phosducin also binds to $G\beta\gamma$, and this site does not overlap with the site on $G\beta\gamma$ that binds the N-terminal regions of $G\alpha_t$, explaining how all three components can interact simultaneously to form a complex of phosducin with heterotrimeric G_t. Whether or not phosducin and G_t form a stable complex within membranes in the intact ROS remains to be determined.

Data support a role for scaffolding proteins in assembling many mammalian phototransduction proteins into a signaling complex. Two glutamic acid–rich proteins (GARPs) referred to as GARP1 and GARP2 as well as a glutamic acid–rich region of the cGMP-gated channel β-subunit referred to as the GARP' part have been identified as playing a role in scaffolding cGMP-gated cation channels, PDE, guanylyl cyclase, ATP-binding cassette transporter (aka. rim protein), and cytoskeletal proteins creating a signaling complex along the rim of the ROS disc where it is close to the plasma membrane. The interactions between PDE, G_t, phosducin, RGS9, $G\beta_5$ and rhodopsin could bring these molecules into a scaffolded complex as well. This signaling complex locates both the activating and attenuating pathways close to the ion channels that they regulate creating a configuration that may function both to prevent high rates of cGMP turnover during periods of intense illumination and to allow for rapid turn-on and turn-off of the response to light.

8 Clinical Importance

8.1 G Protein-mediated Signal Transduction in Diseases

Perturbations in the function of G proteins, their associated receptors and signaling systems have been linked to a number of human diseases including

asthma, heart disease, and cancer among others. Furthermore, recent studies have demonstrated that polymorphic variants (Sect. 9) of many of the proteins involved in G protein–mediated signaling can have consequences in the susceptibility, progression, and prognosis of these diseases. It has also been demonstrated that certain genetic mutations in genes coding for these proteins can result in agonist-independent constitutive signaling, which results in diseases such as precocious puberty, retinitis pigmentosum, and thyroid adenomas.

8.2 Heptahelical Receptors as Drug Targets

G protein–coupled signal transduction systems constitute the largest class of drug targets for the therapeutic treatment of diseases. More than 50% of all prescription drugs on the market today, including 20% of the top 50 selling drugs, directly or indirectly, target these systems, and account for over $20 billion of the pharmaceutical industry's annual sales worldwide. Furthermore, the potential for additional therapeutic drugs is considerable since currently available drugs target pathways that are controlled by only about 5% of the identified human 7TM receptors.

8.3 Potential for Therapeutic Treatments that Target G Protein–mediated Signaling Components Other than Heptahelical Receptors

The majority of drugs that target G protein–coupled signal transduction systems act as either 7TM receptor agonists or antagonists. By and large, attaining a specific therapeutic effect has depended upon the ability to produce ligands that bind selectively to a particular 7TM receptor. Some 7TM receptors have been shown to regulate the activity of more than one kind of effector by activating G proteins with distinct subunit composition. In such cases, targeting a specific pathway may be possible because of a phenomenon known as agonist trafficking. This term was coined to describe the observation that a particular ligand–receptor complex regulates the activity of different effectors with differing efficacies, and the rank order of efficacy for regulating the different effectors depends upon the specific receptor ligand. For example, agonists that bind to the serotonin2C receptor can regulate the activity of both PLC and indirectly phospholipase A_2, and the agonists 3-trifluoromethylphenyl-piperazine and lysergic acid diethylamide (LSD) preferentially activate PLC and phospholipase A_2, respectively. Agonist trafficking is thought to be a consequence of the receptor being able to exist in distinct conformational states that differ in their affinities for a particular ligand. The conformational state preferred by the receptor depends upon what proteins (*e.g.* G protein and effector) it interacts with. In this way, one signaling pathway may be specifically targeted by a particular agonist that would have little effect on other pathways coupled to the same receptor. This raises the possibility of developing pharmaceuticals with specificity for one pathway out of several that may be regulated by the same 7TM receptor.

Although the phenomenon of agonist trafficking offers the possibility of specifically targeting one of several pathways that are controlled by the same 7TM receptor, this goal may be difficult to attain in practice. An alternative approach is to target the interface between components that are unique to a given pathway (*e.g.* the interface between the receptor and the G protein) with peptidic and/or peptidomimetic

drugs. Such drugs may have greater therapeutic specificity, and fewer side effects than drugs that target the receptor binding site. Recent efforts have indeed seen a shift toward therapeutic targets other than the receptor binding site. Evidence supports the feasibility of using substances that either facilitate or inhibit protein–protein interactions as a way of modulating the activity of proteins, and recent studies indicate that G protein – signaling pathways will be amenable to these sorts of manipulations as well. Thus, a primary objective of future research will be to identify peptide motifs that are involved in protein–protein interactions. As the interfaces between the various signaling partners are identified, they will be scrutinized as potential targets for new drugs that can be used to modulate G protein – signaling pathways to therapeutic advantage.

9 Future Directions

If the DNA sequence of two individuals differs by a single nucleotide, and the frequency with which this occurs in a population is relatively high (1% or more), then this difference is called a single nucleotide polymorphism (SNP). It is estimated that a SNP occurs once in every 1000 bases, so the human genome contains several million SNPs. Some polymorphisms have severe consequences such as the SNP responsible for the Glu6Val substitution in hemoglobin that causes sickle-cell anemia, while others are silent causing no change in the gene product. SNPs that occur outside of the coding sequence for a protein may influence gene expression. For many diseases with a genetic component, polymorphisms are likely to be involved in their etiology, and with a few exceptions (*e.g.* sickle-cell anemia), multiple SNPs associated with different genes are liable to contribute. Variation in the response to drugs may also reflect differences in the spectrum of SNPs carried by one individual as compared with another. This interplay of genetics and pharmacology, referred to as pharmacogenomics, has made it important to identify SNPs, and to determine what SNP or combination of SNPs are associated with specific diseases or drug response. The prominence of 7TM receptors as drug targets makes them attractive subjects for probing the pharmacogenomic relevance of SNPs, and numerous SNPs have been identified in 7TM receptors. Many of these SNPs impact the properties of the receptors, and the same is likely to be true for SNPs in other signaling proteins. Knowledge of what genetic variants predispose an individual to disease will provide targets for gene replacement therapies once they become practical. In addition, understanding the correlation between drug response and genetic variation may make it possible to tailor therapeutic treatments to individuals so that maximum benefit can be achieved while avoiding undesirable side effects.

See also Pain Transduction: Gating and Modulation of Ion Channels; Receptor Biochemistry; Receptor Targets in Drug Discovery.

Bibliography

Books and Reviews

Bockaert, J., Fagni, L., Dumuis, A., Marin, P. (2004) GPCR interacting proteins (GIP), *Pharmacol. Ther.* **103**, 203–221.

Cabrera-Vera, T.M., Vanhauwe, J., Thomas, T.O., Medkova, M., Preininger, A., Mazzoni, M.R.,

Hamm, H.E. (2003) Insights into G protein structure, function, and regulation, *Endocr. Rev.* **24**, 765–781.

Gainetdinov, R.R., Premont, R.T., Bohn, L.M., Lefkowitz, R.J., Caron, M.G. (2004) Desensitization of G protein-coupled receptors and neuronal functions, *Annu. Rev. Neurosci.* **27**, 107–144.

Hall, R.A. (2004) β-adrenergic receptors and their interacting proteins, *Semin. Cell Dev. Biol.* **15**, 281–288.

Hall, R.A., Lefkowitz, R.J. (2002) Regulation of G protein-coupled receptor signaling by scaffold proteins, *Circ. Res.* **91**, 672–680.

Kohout, T.A., Lefkowitz, R.J. (2003) Regulation of G protein-coupled receptor kinases and arrestins during receptor desensitization, *Mol. Pharmacol.* **63**, 9–18.

Levitzki, A., Klein, S. (2002) G-protein subunit dissociation is not an integral part of G-protein action, *Chembiochem* **3**, 815–818.

Luttrell, D.K., Luttrell, L.M. (2003) Signaling in time and space: G protein-coupled receptors and mitogen-activated protein kinases, *Assay Drug Dev. Technol.* **1**, 327–338.

Neubig, R.R., Siderovski, D.P. (2002) Regulators of G-protein signalling as new central nervous system drug targets, *Nature. Rev. Drug Discov.* **1**, 187–197.

Offermanns, S. (2003) G-proteins as transducers in transmembrane signalling, *Prog. Biophys. Mol. Biol.* **83**, 101–130.

Rebois, R.V., Hébert, T.E. (2003) Protein complexes involved in heptahelical receptor-mediated signal transduction, *Recept. Channels* **9**, 169–194.

Rodbell, M. (1997) The complex regulation of receptor-coupled G-proteins, *Adv. Enzyme Regul.* **7**, 427–435.

Terrillon, S., Bouvier, M. (2004) Roles of G-protein-coupled receptor dimerization, *EMBO Rep.* **5**, 30–34.

Primary Literature

Akhter, S.A., Luttrell, L.M., Rockman, H.A., Iaccarino, G., Lefkowitz, R.J., Koch, W.J. (1998) Targeting the receptor-G_q interface to inhibit in vivo pressure overload myocardial hypertrophy, *Science* **280**, 574–577.

Alderton, F., Rakhit, S., Kong, K.C., Palmer, T., Sambi, B., Pyne, S., Pyne, N.J. (2001) Tethering of the platelet-derived growth factor β receptor to G-protein-coupled receptors; A novel platform for integrative signaling by these receptor classes in mammalian cells, *J. Biol. Chem.* **276**, 28578–28585.

Angers, S., Salahpour, A., Bouvier, M. (2001) Biochemical and biophysical demonstration of GPCR oligomerization in mammalian cells, *Life Sci.* **68**, 2243–2250.

Angers, S., Salahpour, A., Bouvier, M. (2002) Dimerization: An emerging concept for G protein-coupled receptor ontogeny and function, *Annu. Rev. Pharmacol. Toxicol.* **42**, 409–435.

Angleson, J.K., Wensel, T.G. (1993) A GTPase-accelerating factor for transducin, distinct from its effector cGMP phosphodiesterase, in rod outer segment membranes, *Neuron* **11**, 939–949.

Angleson, J.K., Wensel, T.G. (1994) Enhancement of rod outer segment GTPase accelerating protein activity by the inhibitory subunit of cGMP phosphodiesterase, *J. Biol. Chem.* **269**, 16290–16296.

Antonny, B., Otto-Bruc, A., Chabre, M., Vuong, T.M. (1993) GTP hydrolysis by purified alpha-subunit of transducin and its complex with the cyclic GMP phosphodiesterase inhibitor, *J. Biochem.* **32**, 8646–8653.

Arshavsky, V.Y., Bownds, M.D. (1992) Regulation of deactivation of photoreceptor G protein by its target enzyme and cGMP, *Nature* **357**, 416–417.

Arshavsky, V.Y., Lamb, T.D., Pugh, E.N. (2002) G proteins and phototransduction, *Annu. Rev. Physiol.* **64**, 153–187.

Arshavsky, V.Y., Pugh, J., Edward, N. (1998) Lifetime regulation of G protein-effector complex: Emerging importance of RGS proteins, *Neuron* **20**, 11–14.

Bai, M. (2004) Dimerization of G-protein-coupled receptors: Roles in signal transduction, *Cell. Signal.* **16**, 175–186.

Bardwell, L. (2004) A walk-through of the yeast mating pheromone response pathway, *Peptides* **25**, 1465–1476.

Bar-Sinai, A., Marbach, I., Engelberg, D., Abeliovich, H., Levitzki, A. (1990) *Transmembrane Signalling, Intracellular Messengers and Implication for Drug Development*, John Wiley and Sons.

Basi, N.S., Okuya, S., Rebois, R.V. (1996) GTP binding to Gs does not promote subunit dissociation, *Cell. Signal.* **8**, 209–215.

Bauer, P.H., Bluml, K., Schroder, S., Hegler, J., Dees, C., Lohse, M.J. (1998) Interactions of phosducin with the subunits of G-proteins: Binding to the α as well as the $\beta\gamma$ subunits, *J. Biol. Chem.* **273**, 9465–9471.

Baylor, D.A., Nunn, B.J., Schnapf, J.L. (1984) The photocurrent, noise and spectral sensitivity of rods of the monkey Macaca fascicularis, *J. Physiol.* **357**, 575–607.

Belcheva, M.M., Coscia, C.J. (2002) Diversity of G protein-coupled receptor signaling pathways to ERK/MAP kinase, *Neurosignals* **11**, 34–44.

Bennett, V., Gilligan, D.M. (1993) The spectrin-based membrane skeleton and micron-scale organization of the plasma membrane, *Annu. Rev. Cell Biol.* **9**, 27–66.

Berg, K.A., Maayani, S., Goldfarb, J., Scaramellini, C., Leff, P., Clarke, W.P. (1998) Effector pathway-dependent rRelative efficacy at serotonin type 2A and 2C receptors: Evidence for agonist-directed trafficking of receptor stimulus, *Mol. Pharmacol.* **54**, 94–104.

Bermak, J.C., Li, M., Bullock, C., Zhou, Q.Y. (2001) Regulation of transport of the dopamine D1 receptor by a new membrane- associated ER protein, *Nat. Cell Biol.* **3**, 492–498.

Bernstein, H.G., Keilhoff, C., Reiser, M., Freese, S., Wetzker, R. (1998) Tissue distribution and subcellular localization of a G-protein activated phosphoinositide 3-kinase: An immunohistochemical study, *Cell. Mol. Biol.* **44**, 973–983.

Berstein, G., Blank, J.L., Jhon, D.Y., Exton, J.H., Rhee, S.G., Ross, E.M. (1992) Phospholipase Cβ_1 is a GTPase-activating protein for $G_{q/11}$, its physiological regulator, *Cell* **70**, 411–418.

Bhattacharya, M., Peri, K., Ribeiro-da-Silva, A., Almazan, G., Shichi, H., Hou, X., Varma, D.R., Chemtob, S. (1999) Localization of functional prostaglandin E2 receptors EP3 and EP4 in the nuclear envelope, *J. Biol. Chem.* **274**, 15719–15724.

Biddlecome, G.H., Berstein, G., Ross, E.M. (1996) Regulation of phospholipase Cβ_1 by G_q and m1 muscarinic cholinergic receptor: Steady-state balance of receptor-mediated activation and GTPase-activating protein-promoted deactivation, *J. Biol. Chem.* **271**, 7999–8007.

Birnbaumer, L., Abramowitz, J., Brown, A.M. (1990) Receptor-effector coupling by G-proteins, *Biochim. Biophys. Acta* **1031**, 163–224.

Blumer, J.B., Lanier, S.M. (2003) Accessory proteins for G protein-signaling systems: Activators of G protein signaling and other nonreceptor proteins influencing the activation state of G proteins, *Recept. Channels* **9**, 195–204.

Bockaert, J., Pin, J.P. (1999) Molecular tinkering of G protein-coupled receptors: an evolutionary success, *EMBO J.* **18**, 1723–1729.

Bockaert, J., Roussignol, G., Bécamel, C., Gavarini, S., Joubert, L., Dumuis, A., Fagni, L., Marin, P. (2004) GPCR-interacting proteins (GIPs): Nature and functions, *Biochem. Soc. Trans.* **32**, 851–855.

Bourguignon, L., Walker, G., Huang, H. (1990) Interactions between a lymphoma membrane-associated guanosine 5′- triphosphate-binding protein and the cytoskeleton during receptor patching and capping, *J. Immunol.* **144**, 2242–2252.

Bourne, H.R., Sanders, D.A., McCormick, F. (1990) The GTPase superfamily: A conserved switch for diverse cell functions, *Nature* **348**, 125–132.

Bownds, D., Dawes, J., Miller, J., Stahlman, M. (1972) Phosphorylation of frog photoreceptor membranes induced by light, *Nature New Biol.* **237**, 125–127.

Bray, D. (1998) Signaling complexes: biophysical constraints on intracellular communication, *Annu. Rev. Biophys. Biomol. Struct.* **27**, 59–75.

Breitwieser, G.E. (2004) G protein-coupled receptor oligomerization: Implications for G protein activation and cell signaling, *Circ. Res.* **94**, 17–27.

Brown, D.A., London, E. (1998) Functions of lipid rafts in biological membranes, *Annu. Rev. Cell Dev. Biol.* **14**, 111–136.

Brown, E.J., Frazier, W.A. (2001) Integrin-associated protein (CD47) and its ligands, *Trends Cell Biol.* **11**, 130–135.

Brzostowski, J.A., Kimmel, A.R. (2001) Signaling at zero G: G-protein-independent functions for 7-TM receptors, *Trends Biochem. Sci.* **26**, 291–297.

Bünemann, M., Frank, M., Lohse, M.J. (2003) G_i protein activation in intact cells involves subunit rearrangement rather than dissociation, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 16077–16082.

Can, A., Sayar, K., Friedman, E., Ambrosio, C., Erdemli, E., Gurdal, H. (2000) The effects of agonist stimulation and β_2-adrenergic

receptor level on cellular distribution of $G_s\alpha$ protein, *Cell. Signal.* **12**, 303–309.

Cassel, D., Pfeuffer, T. (1978) Mechanism of cholera toxin action: Covalent modification of the guanyl nucleotide-binding protein of the adenylate cyclase system, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2669–2673.

Cassel, D., Selinger, Z. (1976) Catecholamine-stimulated GTPasse activity in erythrocyte membranes, *Biochim. Biophys. Acta* **452**, 538–551.

Catty, P., Pfister, C., Bruckert, F., Deterre, P. (1992) The cGMP phosphodiesterase-transducin complex of retinal rods: Membrane binding and subunits interactions, *J. Biol. Chem.* **267**, 19489–19493.

Chabre, M. (1985) Trigger and amplification mechanisms in visual phototransduction, *Annu. Rev. Biophys. Biophys. Chem.* **14**, 331–360.

Chan, F.K., Siegel, R.M., Zacharias, D., Swofford, R., Holmes, K.L., Tsien, R.Y., Lenardo, M.J. (2001) Fluorescence resonance energy transfer analysis of cell surface receptor interactions and signaling using spectral variants of the green fluorescent protein, *Cytometry* **44**, 361–368.

Chen, Z., Gibson, T.B., Robinson, F., Silvestro, L., Pearson, G., Xu, B., Wright, A., Vanderbilt, C., Cobb, M.H. (2001) MAP Kinases, *Chem. Rev.* **101**, 2449–2476.

Chidiac, P. (1998) Rethinking receptor-G protein-effector interactions, *Biochem. Pharmacol.* **55**, 549–556.

Claing, A., Chen, W., Miller, W.E., Vitale, N., Moss, J., Premont, R.T., Lefkowitz, R.J. (2001) β-Arrestin-mediated ADP-ribosylation factor 6 activation and β_2-adrenergic receptor endocytosis, *J. Biol. Chem.* **276**, 42509–42513.

Clerc, A., Bennett, N. (1992) Activated cGMP phosphodiesterase of retinal rods. A complex with transducin alpha subunit, *J. Biol. Chem.* **267**, 6620–6627.

Clerc, A., Catty, P., Bennett, N. (1992) Interaction between cGMP-phosphodiesterase and transducin alpha-subunit in retinal rods. A cross-linking study, *J. Biol. Chem.* **267**, 19948–19953.

Colledge, M., Scott, J.D. (1999) AKAPs: From structure to function, *Trends Cell Biol.* **9**, 216–221.

Cong, M., Perry, S.J., Hu, L.A., Hanson, P.I., Claing, A., Lefkowitz, R.J. (2001) Binding of the β_2-adrenergic receptor to N-ethylmaleimide-sensitive factor regulates receptor recycling, *J. Biol. Chem.* **276**, 45145–45152.

Conway, A.-M., Rakhit, S., Pyne, S., Pyne, N.J. (1999) Platelet-derived-growth-factor stimulation of the p42/p44 mitogen-activated protein kinase pathway in airway smooth muscle: role of pertussis-toxin-sensitive G-proteins, c-Src tyrosine kinases and phosphoinositide 3-kinase, *Biochem. J.* **337**, 171–177.

Cook, T.A., Ghomashchi, F., Gelb, M.H., Florio, S.K., Beavo, J.A. (2000) Binding of the δ subunit to rod phosphodiesterase catalytic subunits requires methylated, prenylated C-termini of the catalytic subunits, *J. Biochem.* **39**, 13516–13523.

Cook, T.A., Ghomashchi, F., Gelb, M.H., Florio, S.K., Beavo, J.A. (2001) The δ subunit of type 6 phosphodiesterase reduces light-induced cGMP hydrolysis in rod outer segments, *J. Biol. Chem.* **276**, 5248–5255.

Cooper, D.M.F. (2003) Regulation and organization of adenylyl cyclases and cAMP, *Biochem. J.* **375**, 517–529.

Cote, M., Payet, M.D., Dufour, M.-N., Guillon, G., Gallo-Payet, N. (1997) Association of the G protein α_q/α_{11}-subunit with cytoskeleton in adrenal glomerulosa cells: Role in receptor-effector coupling, *Endocrinology* **138**, 3299–3307.

Couet, J., Sargiacomo, M., Lisanti, M.P. (1997) Interaction of a receptor tyrosine kinase, EGF-R, with caveolins: Caveolin binding negatively regulates tyrosine and serine/threonine kinase activities, *J. Biol. Chem.* **272**, 30429–30438.

Cowan, C.W., Fariss, R.N., Sokal, I., Palczewski, K., Wensel, T.G. (1998) High expression levels in cones of RGS9, the predominant GTPase accelerating protein of rods, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5351–5356.

Craven, S.E., Bredt, D.S. (1998) PDZ proteins organize synaptic signaling pathways, *Cell* **93**, 495–498.

Daaka, Y. (2004) G proteins in cancer: The prostate cancer paradigm, *Sci. STKE* **2004**, re2–re2.

Dalle, S., Ricketts, W., Imamura, T., Vollenweider, P., Olefsky, J.M. (2001) Insulin and insulin-like growth factor I receptors utilize different G protein signaling components, *J. Biol. Chem.* **276**, 15688–15695.

Davare, M.A., Avdonin, V., Hall, D.D., Peden, E.M., Burette, A., Weinberg, R.J., Horne, M.C., Hoshi, T., Hell, J.W. (2001) A β_2-adrenergic receptor signaling complex assembled with the Ca^{2+} channel Cav1.2, *Science* **293**, 98–101.

DeBell, K., Conti, A., Alava, M., Hoffman, T., Bonvini, E. (1992) Microfilament assembly modulates phospholipase C-mediated signal transduction by the TCR/CD3 in murine T helper lymphocytes, *J. Immunol.* **149**, 2271–2280.

DeFea, K.A., Zalevsky, J., Thoma, M.S., Dery, O., Mullins, R.D., Bunnett, N.W. (2000) β-Arrestin-dependent endocytosis of proteinase-activated receptor 2 is required for intracellular targeting of activated ERK1/2, *J. Cell Biol.* **148**, 1267–1282.

Diviani, D., Scott, J.D. (2001) AKAP signaling complexes at the cytoskeleton, *J. Cell Sci.* **114**, 1431–1437.

Doan, T.N., Ali, M.S., Bernstein, K.E. (2001) Tyrosine kinase activation by the angiotensin II receptor in the absence of calcium signaling, *J. Biol. Chem.* **276**, 20954–20958.

Doering, C.J., Kisilevsky, A.E., Feng, Z.-P., Arnot, M.I., Peloquin, J., Hamid, J., Barr, W., Nirdosh, A., Simms, B., Winkfein, R.J., Zamponi, G.W. (2004) A single Gβ subunit locus controls cross-talk between protein kinase C and G protein regulation of N-type calcium channels, *J. Biol. Chem.* **279**, 29709–29717.

Dohlman, H.G., Thorner, J. (2001) Regulation of G protein-initiated signal transduction in yeast: Paradigms and principles, *Annu. Rev. Biochem.* **70**, 703–754.

Dolphin, A.C. (2003) G protein modulation of voltage-gated calcium channels, *Pharmacol. Rev.* **55**, 607–627.

Doupnik, C.A., Davidson, N., Lester, H.A., Kofuji, P. (1997) RGS proteins reconstitute the rapid gating kinetics of gbetagamma- activated inwardly rectifying K+ channels, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10461–10466.

Dowal, L., Elliott, J., Popov, S., Wilkie, T.M., Scarlata, S. (2001) Determination of the contact energies between a regulator of G protein signaling and G protein subunits and phospholipase Cβ_1, *J. Biochem.* **40**, 414–421.

Drmota, T., Novotny, J., Kim, G.-D., Eidne, K.A., Milligan, G., Svoboda, P. (1998) Agonist-induced internalization of the G protein $G_{11}\alpha$ releasing hormone receptors proceed on different time scales, *J. Biol. Chem.* **273**, 21699–21707.

Edwards, A.S., Scott, J.D. (2000) A-kinase anchoring proteins: Protein kinase A and beyond, *Curr. Opin. Cell Biol.* **12**, 217–221.

Eidne, K.A., Kroeger, K.M., Hanyaloglu, A.C. (2002) Applications of novel resonance energy transfer techniques to study dynamic hormone receptor interactions in living cells, *Trends Endocrinol. Metab.* **13**, 415–421.

Elion, E.A. (2001) The Ste5p scaffold, *J. Cell Sci.* **114**, 3967–3978.

Feng, Z.-P., Arnot, M.I., Doering, C.J., Zamponi, G.W. (2001) Calcium channel β subunits differentially regulate the inhibition of N-type channels by individual Gβ isoforms, *J. Biol. Chem.* **276**, 45051–45058.

Ferguson, S.S., Zhang, J., Barak, L.S., Caron, M.G. (1998) Molecular mechanisms of G protein-coupled receptor desensitization and resensitization, *Life Sci.* **62**, 1561–1565.

Ferlito, M., Squadrito, F., Halushka, P.V., Cook, J.A. (2001) Signal transduction events in Chinese hamster ovary cells expressing human CD14: Effect of endotoxin desensitization, *Shock* **15**, 291–296.

Feuilloley, M., Vaudry, H. (1996) Role of the cytoskeleton in adrenocortical cells, *Endocr. Rev.* **17**, 269–288.

Feuilloley, M., Desrues, L., Vaudry, H. (1993) Effect of cytochalasin-B on the metabolism of polyphosphoinositides in andrenocortical cells, *Endocrinology* **133**, 2319–2326.

Foord, S.M., Marshall, F.H. (1999) RAMPs: accessory proteins for seven transmembrane domain receptors, *Trends Pharmacol. Sci.* **20**, 184–187.

Frank, R.N., Cavanagh, H.D., Kenyon, K.R. (1973) Light-stimulated phosphorylation of bovine visual pigments by adenosine triphosphate, *J. Biol. Chem.* **248**, 596–609.

Fraser, I.D., Cong, M., Kim, J., Rollins, E.N., Daaka, Y., Lefkowitz, R.J., Scott, J.D. (2000) Assembly of an A kinase-anchoring protein-β_2-adrenergic receptor complex facilitates receptor phosphorylation and signaling, *Curr. Biol.* **10**, 409–412.

Fredriksson, R., Lagerstrom, M.C., Lundin, L.-G., Schioth, H.B. (2003) The G-protein-coupled receptors in the human genome form five main families: Phylogenetic analysis, paralogon groups, and Fingerprints, *Mol. Pharmacol.* **63**, 1256–1272.

Freeman, J.L.R., Cruz, E.M.D.L., Pollard, T.D., Lefkowitz, R.J., Pitcher, J.A. (1998) Regulation of G protein-coupled receptor kinase 5~ (GRK5) by actin, *J. Biol. Chem.* **273**, 20653–20657.

Fruman, D.A., Meyers, R.E., Cantley, L.C. (1998) Phosphoinositide kinases, *Annu. Rev. Biochem.* **67**, 481–507.

Fujita, S., Inanobe, A., Chachin, M., Aizawa, Y., Kurachi, Y. (2000) A regulator of G protein signalling (RGS) protein confers agonist-dependent relaxation gating to a G protein-gated K^+ channel, *J. Physiol.* **526**(Pt 2), 341–347.

Fujita, T., Toya, Y., Iwatsubo, K., Onda, T., Kimura, K., Umemura, S., Ishikawa, Y. (2001) Accumulation of molecules involved in alpha1-adrenergic signal within caveolae: Caveolin expression and the development of cardiac hypertrophy, *Cardiovasc. Res.* **51**, 709–716.

Fukuhara, S., Chikumi, H., Gutkind, J.S. (2001) RGS-containing RhoGEFs: the missing link between transforming G proteins and Rho?, *Oncogene* **20**, 1661–1668.

Fung, B.K. (1983) Characterization of transducin from bovine retinal rod outer segments I: Separation and reconstitution of the subunits, *J. Biol. Chem.* **258**, 10495–10502.

Fung, B.K., Nash, C.R. (1983) Characterization of transducin from bovine retinal rod outer segments II: Evidence for distinct binding sites and conformational changes revealed by limited proteolysis with trypsin, *J. Biol. Chem.* **258**, 10503–10510.

Fung, B.K.K., Griswold-Prenner, I. (1989) G protein-effector coupling: binding of rod phosphodiesterase inhibitory subunit to transducin, *J. Biochem.* **28**, 3133–3137.

Fung, B.K.K., Hurley, J.B., Stryer, L. (1981) Flow of information in the light-triggered cyclic nucleotide cascade of vision, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 152–156.

Galés, C., Rebois, R.V., Hogue, M., Trieu, P., Schaak, S., Paris, H., Breit, A., Hébert, T.E., Bouvier, M. (2005) Real-time monitoring of receptor-G protein interactions in living cells, *Nat. Biotechnol.* **2**, 177–184.

Ganpat, M.M., Nishimura, M., Toyoshige, M., Okuya, S., Pointer, R.H., Rebois, R.V. (2000) Evidence for stimulation of adenylyl cyclase by an activated G_s heterotrimer in cell membranes: An experimental method for controlling the G_s subunit composition of cell membranes, *Cell. Signal.* **12**, 113–122.

Gao, T., Yatani, A., Dell'Acqua, M.L., Sako, H., Green, S.A., Dascal, N., Scott, J.D., Hosey, M.M. (1997) cAMP-dependent regulation of cardiac L-type Ca^{2+} channels requires membrane targeting of PKA and phosphorylation of channel subunits, *Neuron* **19**, 185–196.

Garrington, T.P., Johnson, G.L. (1999) Organization and regulation of mitogen-activated protein kinase signaling pathways, *Curr. Opin. Cell Biol.* **11**, 211–218.

Gaudet, R., Savage, J.R., McLaughlin, J.N., Willardson, B.M., Sigler, P.B. (1999) A molecular mechanism for the phosphorylation-dependent regulation of heterotrimeric G proteins by phosducin, *Mol. Cell* **3**, 649–660.

Gazi, L., Lopez-Gimenez, J.F., Strange, P.G. (2002) Formation of oligomers by G protein-coupled receptors, *Curr. Opin. Drug Discov. Devel.* **5**, 756–763.

Gilchrist, A., Li, A., Hamm, H.E. (2002) Design and use of C-terminal minigene vectors for studying role of heterotrimeric G proteins, *Meth. Enzymol.* **344**, 58–69.

Gilchrist, A., Bunemann, M., Li, A., Hosey, M.M., Hamm, H.E. (1999) A dominant-negative strategy for studying roles of G proteins in vivo, *J. Biol. Chem.* **274**, 6610–6616.

Gilchrist, A., Vanhauwe, J.F., Li, A., Thomas, T.O., Voyno-Yasenetskaya, T., Hamm, H.E. (2001) Gα minigenes expressing C-terminal peptides serve as specific inhibitors of thrombin-mediated endothelial activation, *J. Biol. Chem.* **276**, 25672–25679.

Gönczy, P. (2002) Mechanisms of spindle positioning: focus on flies and worms, *Trends Cell Biol.* **12**, 332–339.

Goodman, Jr., O.B., Krupnick, J.G., Gurevich, V.V., Benovic, J.L., Keen, J.H. (1997) Arrestin/clathrin interaction: Localization of the arrestin binding locus to the clathrin terminal domain, *J. Biol. Chem.* **272**, 15017–15022.

Gouldson, P.R., Kidley, N.J., Bywater, R.P., Psaroudakis, G., Brooks, H.D., Diaz, C., Shire, D., Reynolds, C.A. (2004) Toward the active conformations of rhodopsin and the β_2-adrenergic receptor, *Proteins: Stuct. Funct. Bioinf.* **56**, 67–84.

Gray-Keller, M.P., Biernbaum, M.S., Bownds, M.D. (1990) Transducin activation in electropermeabilized frog rod outer segments is highly amplified, and a portion equivalent to phosphodiesterase remains membrane-bound, *J. Biol. Chem.* **265**, 15323–15332.

Green, J.M., Zhelesnyak, A., Chung, J., Lindberg, F.P., Sarfati, M., Frazier, W.A., Brown, E.J. (1999) Role of cholesterol in formation and function of a signaling complex involving $\alpha v\beta 3$, integrin-associated protein (CD47), and heterotrimeric G proteins, *J. Cell Biol.* **146**, 673–682.

Guillard, C., Chretien, S., Jockers, R., Fichelson, S., Mayeux, P., Duprez, V. (2001) Coupling of heterotrimeric G_i proteins to the erythropoietin receptor, *J. Biol. Chem.* **276**, 2007–2014.

Hajjar, R.J., Schwinger, R.H.G., Schmidt, U., Kim, C.S., Lebeche, D., Doye, A.A., Gwathmey, J.K. (2000) Myofilament calcium regulation in human myocardium, *Circulation* **101**, 1679–1685.

Hall, R.A., Premont, R.T., Lefkowitz, R.J. (1999) Heptahelical receptor signaling: Beyond the G protein paradigm, *J. Cell Biol.* **145**, 927–932.

Hamilton, S.L., Codina, J., Hawkes, M.J., Yatani, A., Sawada, T., Strickland, F.M., Froehner, S.C., Spiegel, A.M., Toro, L., Stefani, E., Birnbaumer, L., Brown, A.M. (1991) Evidence for direct interaction of $G_s\alpha$ with Ca^{2+} channel of skeletal muscle, *J. Biol. Chem.* **266**, 19528–19535.

Hanoune, J., Defer, N. (2001) Regulation and role of adenylyl cyclase isoforms, *Annu. Rev. Pharmacol. Toxicol.* **41**, 145–174.

Hansen, C.A., Schroering, A.G., Carey, D.J., Robishaw, J.D. (1994) Localization of a heterotrimeric G protein gamma subunit to focal adhesions and associated stress fibers, *J. Cell Biol.* **126**, 811–819.

He, W., Cowan, C.W., Wensel, T.G. (1998) RGS9, a GTPase accelerator for phototransduction, *Neuron* **20**, 95–102.

Hebert, T.E., Bouvier, M. (1998) Structural and functional aspects of G protein-coupled receptor oligomerization, *Biochem. Cell Biol.* **76**, 1–11.

Hebert, T.E., Moffett, S., Morello, J.P., Loisel, T.P., Bichet, D.G., Barret, C., Bouvier, M. (1996) A peptide derived from a β_2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation, *J. Biol. Chem.* **271**, 16384–16392.

Hechter, O., Yoshinaga, K., Halkerston, I.D.K., Cohn, C., Dodd, P. (1965) *Molecular Basis of Some Aspects of Mental Activity*, Walaas, O. (Ed.) Academic Press, New York, pp. 291–346.

Heck, M., Hofmann, K.P. (1993) G-protein-effector coupling: A real-time light-scattering assay for transducin-phosphodiesterase interaction, *J. Biochem.* **32**, 8220–8227.

Hepler, J.R., Kozasa, T., Smrcka, A.V., Simon, M.I., Rhee, S.G., Sternweis, P.C., Gilman, A.G. (1993) Purification from Sf9 cells and characterization of recombinant $G_q\alpha$ and $G_{11}\alpha$: Activation of purified phospholipase C isozymes by G alpha subunits, *J. Biol. Chem.* **268**, 14367–14375.

Hilairet, S., Belanger, C., Bertrand, J., Laperriere, A., Foord, S.M., Bouvier, M. (2001) Agonist-promoted Internalization of a ternary complex between calcitonin receptor-like receptor, receptor activity-modifying protein 1 (RAMP1), and β-arrestin, *J. Biol. Chem.* **276**, 42182–42190.

Ho, Y.K., Fung, B.K. (1984) Characterization of transducin from bovine retinal rod outer segments. The role of sulfhydryl groups, *J. Biol. Chem.* **259**, 6694–6699.

Hooper, N.M. (1999) Detergent-insoluble glycosphingolipid/cholesterol-rich membrane domains, lipid rafts and caveolae, *Mol. Membr. Biol.* **16**, 145–156.

Hou, Y., Chang, V., Capper, A.B., Taussig, R., Gautam, N. (2001) G protein β subunit types differentially interact with a muscarinic receptor but not adenylyl cyclase type II or phospholipase-$\beta 2/3$, *J. Biol. Chem.* **276**, 19982–19988.

Huang, C.-L., Jan, Y.N., Jan, L.Y. (1997) Binding of the G protein $\beta\gamma$ subunit to multiple regions of G protein-gated inward-rectifying K^+ channels, *FEBS Lett.* **405**, 291–298.

Huang, C.-L., Slesinger, P.A., Casey, P.J., Jan, Y.N., Jan, L.Y. (1995) Evidence that direct binding of $G\beta\gamma$ to the GIRK1 G protein-gated inwardly rectifying K+ channel is important for channel activation *Neuron* **15**, 1133–1143.

Huber, A., Sander, P., Gobert, A., Bahner, M., Hermann, R., Paulsen, R. (1996) The transient receptor potential protein (Trp), a putative store-operated Ca^{2+} channel essential for phosphoinositide-mediated photoreception, forms a signaling complex with NorpA, InaC and InaD, *EMBO J.* **15**, 7036–7045.

Hummer, A., Delzeith, O., Gomez, S.R., Moreno, R.L., Mark, M.D., Herlitze, S. (2003) Competitive and synergistic interactions of G protein β_2 and Ca^{2+} Channel β_{1b} subunits with $Ca_v2.1$ channels, revealed by mammalian two-hybrid and fluorescence resonance energy transfer measurements, *J. Biol. Chem.* **278**, 49386–49400.

Hurley, J.B. (1980) Isolation and recombination of bovine rod outer segment cGMP phosphodiesterase and its regulators, *Biochem. Biophys. Res. Commun.* **92**, 505–510.

Hurley, J.H. (1999) Structure, mechanism, and regulation of mammalian adenylyl cyclase, *J. Biol. Chem.* **274**, 7599–7602.

Hurowitz, E.H., Melnyk, J.M., Chen, Y.-J., Kouros-Mehr, H., Simon, M.I., Shizuya, H. (2000) Genomic characterization of the human heterotrimeric G protein α, β, and γ subunit genes, *DNA Res.* **7**, 111–120.

Hynes, T.R., Mervine, S.M., Yost, E.A., Sabo, J.L., Berlot, C.H. (2004) Live cell imaging of G_s and the β_2-adrenergic receptor demonstrates that both α_s and $\beta_1\gamma_7$ Internalize upon stimulation and exhibit similar trafficking patterns that differ from that of the β_2-adrenergic receptor, *J. Biol. Chem.* **279**, 44101–44112.

Ibarrondo, J., Joubert, D., Dufour, M., Cohen-Solal, A., Homburger, V., Jard, S., Guillon, G. (1995) Close association of the α subunits of G_q and G_{11} G proteins with actin filaments in WRK1 cells: Relation to G protein-mediated phospholipase C activation, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8413–8417.

Imamura, T., Vollenweider, P., Egawa, K., Clodi, M., Ishibashi, K., Nakashima, N., Ugi, S., Adams, J.W., Brown, J.H., Olefsky, J.M. (1999) $G\alpha_{q/11}$ protein plays a key role in insulin-induced glucose transport in 3T3-L1 adipocytes, *Mol. Cell. Biol.* **19**, 6765–6774.

Inanobe, A., Morishige, K.I., Takahashi, N., Ito, H., Yamada, M., Takumi, T., Nishina, H., Takahashi, K., Kanaho, Y., Katada, T., Kurachi, Y. (1995) $G\beta\gamma$ directly binds to the carboxyl terminus of the G protein-gated muscarinic K+ channel, GIRK1, *Biochem. Biophys. Res. Commun.* **212**, 1022–1028.

Janmey, P.A. (1998) The cytoskeleton and cell signaling: Component localization and mechanical coupling, *Physiol. Rev.* **78**, 763–781.

Jones, K.A., Borowsky, B., Tamm, J.A., Craig, D.A., Durkin, M.M., Dai, M., Yao, W.J., Johnson, M., Gunwaldsen, C., Huang, L.Y., Tang, C., Shen, Q., Salon, J.A., Morse, K., Laz, T., Smith, K.E., Nagarathnam, D., Noble, S.A., Branchek, T.A., Gerald, C. (1998) GABA(B) receptors function as a heteromeric assembly of the subunits GABA(B)R1 and GABA(B)R2, *Nature* **396**, 674–679.

Jones, P.G., Macdonald, S.G., Cockett, M.I. (1999) Non-binding site modulation of G protein-coupled receptor signalling, *Expert Opin. Ther. Patents* **9**, 1641–1654.

Jordan, B.A., Devi, L.A. (1999) G-protein-coupled receptor heterodimerization modulates receptor function, *Nature* **399**, 697–700.

Kaupmann, K., Huggel, K., Heid, J., Flor, P.J., Bischoff, S., Mickel, S.J., McMaster, G., Angst, C., Bittiger, H., Froestl, W., Bettler, B. (1997) Expression cloning of GABAB receptors uncovers similarity to metabotropic glutamate receptors, *Nature* **386**, 239–246.

Kaupmann, K., Malitschek, B., Schuler, V., Heid, J., Froestl, W., Beck, P., Mosbacher, J., Bischoff, S., Kulik, A., Shigemoto, R., Karschin, A., Bettler, B. (1998) GABA(B)-receptor subtypes assemble into functional heteromeric complexes, *Nature* **396**, 683–687.

Kenakin, T. (1995) Agonist-receptor efficacy II: agonist trafficking of receptor signals, *Trends Pharmacol. Sci.* **16**, 232–238.

Kirstein, S.L., Insel, P.A. (2004) Autonomic nervous system pharmacogenomics: A progress report, *Pharmacol. Rev.* **56**, 31–52.

Kitano, J., Nishida, M., Itsukaichi, Y., Minami, I., Ogawa, M., Hirano, T., Mori, Y., Nakanishi, S. (2003) Direct interaction and functional coupling between metabotropic glutamate receptor subtype 1 and voltage-sensitive Cav2.1 Ca^{2+} channel, *J. Biol. Chem.* **278**, 25101–25108.

Kleuss, C., Krause, E. (2003) G_s is palmitoylated at the N-terminal glycine, *EMBO J.* **22**, 826–832.

Kobrinsky, E., Mirshahi, T., Zhang, H.L., Jin, T.H., Logothetis, D.E. (2000) Receptor-mediated hydrolysis of plasma membrane messenger PIP_2 leads to K^+-current desensitization, *Nat. Cell Biol.* **2**, 507–514.

Kolch, W. (2000) Meaningful relationships: The regulation of the Ras/Raf/MEK/ERK pathway by protein interactions, *Biochem. J.* **351**(Pt 2), 289–305.

Körschen, H.G., Beyermann, M., Müller, F., Heck, M., Vantler, M., Koch, K.W., Kellner, R., Wolfrum, U., Bode, C., Hofmann, K.P., Kaupp, U.B. (1999) Interaction of glutamic-acid-rich proteins with the cGMP signalling pathway in rod photoreceptors, *Nature* **400**, 761–766.

Kozasa, T., Hepler, J.R., Smrcka, A.V., Simon, M., Rhee, S.G., Sternweis, P.C., Gilman, A.G. (1993) Purification and characterization

of recombinant $G_{16\alpha}$ from Sf9 cells: Activation of purified phospholipase C isozymes by G-protein α subunits, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9176–9180.

Kozasa, T., Jiang, X., Hart, M.J., Sternweis, P.M., Singer, W.D., Gilman, A.G., Bollag, G., Sternweis, P.C. (1998) p115 RhoGEF, a GTPase activating protein for $G\alpha_{12}$ and $G\alpha_{13}$, *Science* **280**, 2109–2111.

Kroll, S., Phillips, W.J., Cerione, R.A. (1989) The regulation of the cyclic GMP phosphodiesterase by the GDP-bound form of the alpha subunit of transducin, *J. Biol. Chem.* **264**, 4490–4497.

Kroslak, T., Koch, T., Kahl, E., Hollt, V. (2001) Human phosphatidylethanolamine-binding protein facilitates heterotrimeric G protein-dependent signaling, *J. Biol. Chem.* **276**, 39772–39778.

Kuemmerle, J.F., Murthy, K.S. (2001) Coupling of the insulin-like growth factor-I receptor tyrosine kinase to G_{i2} in human intestinal smooth muscle: $G\beta\gamma$-dependent mitogen-activated protein kinase activation and growth, *J. Biol. Chem.* **276**, 7187–7194.

Kühn, H. (1978) Light-regulated binding of rhodopsin kinase and other proteins to cattle photoreceptor membranes, *J. Biochem.* **17**, 4389–4395.

Kühn, H., Dreyer, W.J. (1972) Light dependent phosphorylation of rhodopsin by ATP, *FEBS Lett.* **20**, 1–6.

Kühn, H., Cook, J.H., Dreyer, W.J. (1973) Phosphorylation of rhodopsin in bovine photoreceptor membranes: Dark reaction after illumination, *J. Biochem.* **12**, 2495–2502.

Kuner, R., Kohr, G., Grunewald, S., Eisenhardt, G., Bach, A., Kornau, H.C. (1999) Role of heteromer formation in GABAB receptor function, *Science* **283**, 74–77.

Kunkel, M.T., Peralta, E.G. (1995) Identification of domains conferring G protein regulation on inward rectifier potassium channels, *Cell* **83**, 443–449.

Kurosu, H., Maehama, T., Okada, T., Yamamoto, T., Hoshino, S.-i., Fukui, Y., Ui, M., Hazeki, O., Katada, T. (1997) Heterodimeric phosphoinositide 3-kinase consisting of p85 and p110β is synergistically activated by the $\beta\gamma$ subunits of G proteins and phosphotyrosyl peptide, *J. Biol. Chem.* **272**, 24252–24256.

Lachance, M., Ethier, N., Wolbring, G., Schnetkamp, P.P.M., Hebert, T.E. (1999) Stable association of G proteins with β_2AR is independent of the state of receptor activation, *Cell. Signal.* **11**, 523–533.

Lambright, D.G., Sondek, J., Bohm, A., Skiba, N.P., Hamm, H.E., Sigler, P.B. (1996) The 2.0 Å crystal structure of a heterotrimeric G protein, *Nature* **379**, 311–319.

Lavine, N., Ethier, N., Oak, J.N., Pei, L., Liu, F., Trieu, P., Rebois, R.V., Bouvier, M., Hebert, T.E., Van Tol, H.H.M. (2002) G protein-coupled receptors form stable complexes with inwardly rectifying potassium channels and adenylyl cyclase, *J. Biol. Chem.* **277**, 46010–46019.

Lee, A., Rana, B.K., Schiffer, H.H., Schork, N.J., Brann, M.R., Insel, P.A., Weiner, D.M. (2003) Distribution analysis of nonsynonymous polymorphisms within the G-protein-coupled receptor gene family, *Genomics* **81**, 245–248.

Lee, N.M., Smith, A.P. (2002) Opioid receptor polymorphisms and opioid abuse, *Pharmacogenomics* **3**, 219–227.

Lee, R.H., Lieberman, B.S., Lolley, R.N. (1990) Retinal accumulation of the phosducin/T$\beta\gamma$ and transducin complexes in developing normal mice and in mice and dogs with inherited retinal degeneration, *Exp. Eye Res.* **51**, 325–333.

Lee, R.H., Ting, T.D., Lieberman, B.S., Tobias, D.E., Lolley, R.N., Ho, Y.K. (1992) Regulation of retinal cGMP cascade by phosducin in bovine rod photoreceptor cells. Interaction of phosducin and transducin, *J. Biol. Chem.* **267**, 25104–25112.

Lefkowitz, R.J. (2000) The superfamily of heptahelical receptors, *Nat. Cell Biol.* **2**, E133–E136.

Leiber, D., Jasper, J.R., Alousi, A.A., Martin, J., Bernstein, D., Insel, P.A. (1993) Alteration in G_s-mediated signal transduction in S49 lymphoma cells treated with inhibitors of microtubules, *J. Biol. Chem.* **268**, 3833–3837.

Lemmon, M.A., Ferguson, K.M. (2000) Signal-dependent membrane targeting by pleckstrin homology (PH) domains, *Biochem. J.* **350**(Pt 1), 1–18.

Leopoldt, D., Hanck, T., Exner, T., Maier, U., Wetzker, R., Nurnberg, B. (1998) $G\beta\gamma$ stimulates phosphoinositide 3-kinase-γ by direct interaction with two domains of the catalytic p110 subunit, *J. Biol. Chem.* **273**, 7024–7029.

Levay, K., Cabrera, J.L., Satpaev, D.K., Slepak, V.Z. (1999) $G\beta_5$ prevents the RGS7-$G\alpha_o$

interaction through binding to a distinct Gγ-like domain found in RGS7 and other RGS proteins, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2503–2507.

Levitzki, A. (1984) Receptor to effector coupling in the receptor-dependent adenylate cyclase system, *J. Recept. Res.* **4**, 399–409.

Levitzki, A. (1986) β-Adrenergic receptors and their mode of coupling to adenylate cyclase, *Physiol. Rev.* **66**, 819–854.

Levitzki, A. (1988a) From epinephrine to cyclic AMP, *Science* **241**, 800–806.

Levitzki, A. (1988b) Transmembrane signalling to adenylate cyclase in mammalian cells and in Saccharomyces cerevisiae, *Trends Biochem. Sci.* **13**, 298–301.

Li, S., Couet, J., Lisanti, M.P. (1996) Src tyrosine kinases, Gα subunits, and H-Ras share a common membrane-anchored scaffolding protein, caveolin: Caveolin binding negatively regulates the auto-activation of Src tyrosine kinases, *J. Biol. Chem.* **271**, 29182–29190.

Li, S., Okamoto, T., Chun, M., Sargiacomo, M., Casanova, J.E., Hansen, S.H., Nishimoto, I., Lisanti, M.P. (1995) Evidence for a regulated interaction between heterotrimeric G proteins and caveolin, *J. Biol. Chem.* **270**, 15693–156701.

Liggett, S.B. (2004) Polymorphisms of β-adrenergic receptors in heart failure, *Am. J. Med.* **117**, 525–527.

Lim, N., Dascal, N., Labarca, C., Davidson, N., Lester, H. (1995) A G protein-gated K channel is activated via β_2-adrenergic receptors and G$\beta\gamma$ subunits in Xenopus oocytes, *J. Gen. Physiol.* **105**, 421–439.

Lin, F., Wang, H., Malbon, C.C. (2000) Gravin-mediated formation of signaling complexes in β_2-adrenergic receptor desensitization and resensitization, *J. Biol. Chem.* **275**, 19025–19034.

Liu, G., Shi, J., Yang, L., Cao, L., Park, S.M., Cui, J., Marx, S.O. (2004) Assembly of a Ca^{2+}-dependent BK channel signaling complex by binding to β_2-adrenergic receptor, *EMBO J.* **23**, 2196–2205.

Liu, Y., Arshavsky, V.Y., Ruoho, A.E. (1999) Interaction sites of the C-terminal region of the cGMP phosphodiesterase inhibitory subunit with the GDP-bound transducin α-subunit, *Biochem. J.* **337**, 281–288.

Logothetis, D.E., Kurachi, Y., Galper, J., Neer, E.J., Clapham, D.E. (1987) The $\beta\gamma$ subunits of GTP-binding proteins activate the muscarinic K+ channel in heart, *Nature* **325**, 321–326.

Lopez-Ilasaca, M., Gutkind, J.S., Wetzker, R. (1998) Phosphoinositide 3-kinase γ is a mediator of G$\beta\gamma$-dependent Jun kinase activation, *J. Biol. Chem.* **273**, 2505–2508.

Lopez-Ilasaca, M., Crespo, P., Pellici, P.G., Gutkind, J.S., Wetzker, R. (1997) Linkage of G protein-coupled receptors to the MAPK signaling pathway through PI 3-kinase γ, *Science* **275**, 394–397.

Lu, D., Yang, H., Shaw, G., Raizada, M.K. (1998) Angiotensin II-induced nuclear targeting of the angiotensin type 1 (AT1) receptor in brain neurons, *Endocrinology* **139**, 365–375.

Luttrell, L.M., Ferguson, S.S., Daaka, Y., Miller, W.E., Maudsley, S., Della Rocca, G.J., Lin, F., Kawakatsu, H., Owada, K., Luttrell, D.K., Caron, M.G., Lefkowitz, R.J. (1999) β-arrestin-dependent formation of β_2-adrenergic receptor-Src protein kinase complexes, *Science* **283**, 655–661.

Ma, D., Zerangue, N., Raab-Graham, K., Fried, S.R., Jan, Y.N., Jan, L.Y. (2002) Diverse trafficking patterns due to multiple traffic motifs in G protein-activated inwardly rectifying potassium channels from brain and heart, *Neuron* **33**, 715–729.

Ma, Y.-C., Huang, J., Ali, S., Lowry, W., Huang, X.-Y. (2000) Src tyrosine kinase is a novel direct effector of G proteins, *Cell* **102**, 635–646.

Macrez, N., Mironneau, C., Carricaburu, V., Quignard, J.-F., Babich, A., Czupalla, C., Nurnberg, B., Mironneau, J. (2001) Phosphoinositide 3-kinase isoforms selectively couple receptors to vascular L-type Ca^{2+} channels, *Circ. Res.* **89**, 692–699.

Macrez-Lepretre, N., Kalkbrenner, F., Morel, J.-L., Schultz, G., Mironneau, J. (1997) G protein heterotrimer $G\alpha_{13}\beta_1\gamma_3$ couples the angiotensin AT1A receptor to increases in cytoplasmic Ca^{2+} in rat portal vein myocytes, *J. Biol. Chem.* **272**, 10095–10102.

Maier, U., Babich, A., Nurnberg, B. (1999) Roles of non-catalytic subunits in G$\beta\gamma$-induced activation of class I phosphoinositide 3-knase isoforms β and γ, *J. Biol. Chem.* **274**, 29311–29317.

Makino, E.R., Handy, J.W., Li, T., Arshavsky, V.Y. (1999) The GTPase activating factor for transducin in rod photoreceptors is the complex between RGS9 and type 5 G

protein β-subunit, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1947–1952.

Malarkey, K., Belham, C.M., Paul, A., Graham, A., McLees, A., Scott, P.H., Plevin, R. (1995) The regulation of tyrosine kinase signalling pathways by growth factor and G-protein-coupled receptors, *Biochem. J.* **309**, 361–375.

Mañes, S., Lacalle, R.A., Gómez-Moutón, C., del Real, G., Mira, E., Martínez-A, C. (2001) Membrane raft microdomains in chemokine receptor function, *Semin. Immunol.* **13**, 147–157.

Margeta-Mitrovic, M., Jan, Y.N., Jan, L.Y. (2000) A trafficking checkpoint controls GABAB receptor heterodimerization, *Neuron* **27**, 97–106.

Marinissen, M.J., Gutkind, J.S. (2001) G-protein-coupled receptors and signaling networks: emerging paradigms, *Trends Pharmacol. Sci.* **22**, 368–376.

Martin-McCaffrey, L., Willard, F.S., Oliveira-dos-Santos, A.J., Natale, D.R.C., Snow, B.E., Kimple, R.J., Pajak, A., Watson, A.J., Dagnino, L., Penninger, J.M. (2004) RGS14 Is a mitotic spindle protein essential from the first division of the mammalian zygote, *Dev. Cell* **7**, 763–769.

McDonald, P.H., Cote, N.L., Lin, F.-T., Premont, R.T., Pitcher, J.A., Lefkowitz, R.J. (1999) Identification of NSF as a β-arrestin1-binding protein. Implications for β_2-adrenergic receptor regulation, *J. Biol. Chem.* **274**, 10677–10680.

McDonald, P.H., Chow, C.W., Miller, W.E., Laporte, S.A., Field, M.E., Lin, F.T., Davis, R.J., Lefkowitz, R.J. (2000) β-arrestin 2: A receptor-regulated MAPK scaffold for the activation of JNK3, *Science* **290**, 1574–1577.

McEnery, M., Snowman, A., Snyder, S. (1994) The association of endogenous $G_o\alpha$ with the purified omega-conotoxin GVIA receptor, *J. Biol. Chem.* **269**, 5–8.

McHenry, Lawrence, C., Jr. (1969) *Garrison's History of Neurology Revised and Enlarged with a Bibliography of Classical, Original and Standard Works in Neurology*, Charles C. Thomas Publisher, Springfield.

Mellado, M., Rodríguez-Frade, J.M., Vila-Coro, A.J., Fernández, S., Martín de Ana, A., Jones, D.R., JTorán, o.L., Martínez-A, C. (2001) Chemokine receptor homo- or heterodimerization activates distinct signaling pathways, *EMBO J.* **20**, 2497–2507.

Michaely, P.A., Mineo, C., Ying, Y.S., Anderson, R.G. (1999) Polarized distribution of endogenous Rac1 and RhoA at the cell surface, *J. Biol. Chem.* **274**, 21430–21436.

Miller, W.E., Lefkowitz, R. (2001) Expanding roles for β-arrestins as scaffolds and adapters in GPCR signaling and trafficking, *Curr. Opin. Cell Biol.* **13**, 139–145.

Miller, W.E., Maudsley, S., Ahn, S., Khan, K.D., Luttrell, L.M., Lefkowitz, R.J. (2000) β-Arrestin1interacts with the catalytic domain of the tyrosine kinase c-Src: Role of β-arrestin1-dependent targeting of c-Src in receptor endocytosis, *J. Biol. Chem.* **275**, 11312–11319.

Milligan, G. (2000) Insights into ligand pharmacology using receptor-G-protein fusion proteins, *Trends Pharmacol. Sci.* **21**, 24–28.

Milligan, G., White, J.H. (2001) Protein-protein interactions at G-protein-coupled receptors, *Trends Pharmacol. Sci.* **22**, 513–518.

Mochizuki, N., Cho, G., Wen, B., Insel, P.A. (1996) Identification and cDNA cloning of a novel human mosaic protein, LGN, based on interaction with $G\alpha i2$, *Gene* **181**, 39–43.

Montell, C. (1998) TRP trapped in fly signaling web, *Curr. Opin. Neurobiol.* **8**, 389–397.

Montell, C. (1999) Visual transduction in *Drosophila*, *Annu. Rev. Cell Dev. Biol.* **15**, 231–268.

Morfis, M., Christopoulos, A., Sexton, P.M. (2003) RAMPs: 5 years on, where to now?, *Trends Pharmacol. Sci.* **24**, 596–601.

Murga, C., Fukuhara, S., Gutkind, J.S. (2000) A novel role for phosphatidylinositol 3-kinase beta in signaling from G protein-coupled receptors to Akt, *J. Biol. Chem.* **275**, 12069–12073.

Ng, G.Y., Clark, J., Coulombe, N., Ethier, N., Hebert, T.E., Sullivan, R., Kargman, S., Chateauneuf, A., Tsukamoto, N., McDonald, T., Whiting, P., Mezey, E., Johnson, M.P., Liu, Q., Kolakowski, L.F. Jr., Evans, J.F., Bonner, T.I., O'Neill, G.P. (1999) Identification of a GABAB receptor subunit, gb2, required for functional GABAB receptor activity, *J. Biol. Chem.* **274**, 7607–7610.

Northup, J.K., Sternweis, P.C., Smigel, M.D., Schleifer, L.S., Ross, E.M., Gilman, A.G. (1980) Purification of the regulatory component of adenylate cyclase, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6516–6520.

Oka, N., Yamamoto, M., Schwencke, C., Kawabe, J., Ebina, T., Ohno, S., Couet, J., Lisanti, M.P., Ishikawa, Y. (1997) Caveolin

interaction with protein kinase C: Isoenzyme-dependent regulation of kinase activity by the caveolin scaffolding domain peptide, *J. Biol. Chem.* **272**, 33416–33421.

Okamoto, T., Schlegel, A., Scherer, P.E., Lisanti, M.P. (1998) Caveolins, a family of scaffolding proteins for organizing "preassembled signaling complexes" at the plasma membrane, *J. Biol. Chem.* **273**, 5419–5422.

Ostrom, R.S., Gregorian, C., Drenan, R.M., Xiang, Y., Regan, J.W., Insel, P.A. (2001) Receptor number and caveolar co-localization determine receptor coupling efficiency to adenylyl cyclase, *J. Biol. Chem.* **276**, 42063–42069.

Ostrowska, A. (2000) The structure and activity of cyclic AMP-dependent protein kinase A, *J. Anim. Feed Sci.* **9**, 213–246.

Ostrom, R.S., Post, S.R., Insel, P.A. (2000) Stoichiometry and compartmentation in G protein-coupled receptor signaling: implications for therapeutic interventions involving G_s, *J. Pharmacol. Exp. Ther.* **294**, 407–412.

Otto-Bruc, A., Antonny, B., Vuong, T.M., Chardin, P., Chabre, M. (1993) Interaction between the retinal cyclic GMP phosphodiesterase inhibitor and transducin. Kinetics and affinity studies, *J. Biochem.* **32**, 8636–8645.

Pages, F., Deterre, P., Pfister, C. (1992) Enhanced GTPase activity of transducin when bound to cGMP phosphodiesterase in bovine retinal rods, *J. Biol. Chem.* **267**, 22018–22021.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Trong, I.L., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M., Miyano, M. (2000) Crystal structure of rhodopsin: A G protein-coupled receptor, *Science* **289**, 739–745.

Papermaster, D.S., Schneider, B.G., Zorn, M.A., Kraehenbuhl, J.P. (1978) Immunocytochemical localization of a large intrinsic membrane protein to the incisures and margins of frog rod outer segment disks, *J. Cell Biol.* **78**, 415–425.

Parma, J., Duprez, L., Van Sande, J., Cochaux, P., Gervy, C., Mockel, J., Dumont, J., Vassart, G. (1993) Somatic mutations in the thyrotropin receptor gene cause hyperfunctioning thyroid adenomas, *Nature* **365**, 649–651.

Pawson, T., Nash, P. (2000) Protein-protein interactions define specificity in signal transduction, *Genes Dev.* **14**, 1027–1047.

Pawson, T., Scott, J.D. (1997) Signaling through scaffold, anchoring, and adaptor proteins, *Science* **278**, 2075–2080.

Payrastre, B., van Bergen en Henegouwen, P., Breton, M., den Hartigh, J., Plantavid, M., Verkleij, A., Boonstra, J. (1991) Phosphoinositide kinase, diacylglycerol kinase, and phospholipase C activities associated to the cytoskeleton: effect of epidermal growth factor, *J. Cell Biol.* **115**, 121–128.

Pfeiffer, M., Koch, T., Schroder, H., Klutzny, M., Kirscht, S., Kreienkamp, H.J., Hollt, V., Schulz, S. (2001) Homo- and heterodimerization of somatostatin receptor subtypes. Inactivation of sst(3) receptor function by heterodimerization with sst(2A), *J. Biol. Chem.* **276**, 14027–14036.

Pfeuffer, T., Helmreich, E.J. (1975) Activation of pigeon erythrocyte membrane adenylate cyclase by guanylnucleotide analogues and separation of a nucleotide binding protein, *J. Biol. Chem.* **250**, 867–876.

Pierce, K.L., Luttrell, L.M., Lefkowitz, R.J. (2001) New mechanisms in heptahelical receptor signaling to mitogen activated protein kinase cascades, *Oncogene* **20**, 1532–1539.

Pierce, K.L., Maudsley, S., Daaka, Y., Luttrell, L.M., Lefkowitz, R.J. (2000) Role of endocytosis in the activation of the extracellular signal-regulated kinase cascade by sequestering and nonsequestering G protein-coupled receptors, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1489–1494.

Pike, L.J., Casey, L. (1996) Localization and turnover of phosphatidylinositol 4,5-bisphosphate in caveolin-enriched membrane domains, *J. Biol. Chem.* **271**, 26453–26456.

Ping, P., Zhang, J., Pierce, W.M., Jr., Bolli, R. (2001) Functional proteomic analysis of protein kinase Cε signaling complexes in the normal heart and during cardioprotection, *Circ. Res.* **88**, 59–62.

Ponting, C.P., Phillips, C., Davies, K.E., Blake, D.J. (1997) PDZ domains: Targeting signalling molecules to sub-membranous sites, *Bioessays* **19**, 469–479.

Popova, J.S., Rasenick, M.M. (2000) Muscarinic receptor activation promotes the membrane association of tubulin for the regulation of G_q-mediated phospholipase Cβ1 signaling, *J. Neurosci.* **20**, 2774–2782.

Popova, J.S., Garrison, J.C., Rhee, S.G., Rasenick, M.M. (1997) Tubulin, G_q, and phosphatidylinositol 4,5-bisphosphate interact

to regulate phospholipase $C\beta_1$ signaling, *J. Biol. Chem.* **272**, 6760–6765.

Posner, B.A., Gilman, A.G., Harris, B.A. (1999) Regulators of G protein signaling 6 and 7: Purification of complexes with $G\beta_5$ and assessment of their effects on G protein-mediated signaling pathways, *J. Biol. Chem.* **274**, 31087–31093.

Premont, R.T., Claing, A., Vitale, N., Perry, S.J., Lefkowitz, R.J. (2000) The GIT family of ADP-ribosylation factor GTPase-activating proteins. Functional diversity of GIT2 through alternative splicing, *J. Biol. Chem.* **275**, 22373–22380.

Prenzel, N., Zwick, E., Daub, H., Leserer, M., Abraham, R., Wallasch, C., Ullrich, A. (1999) EGF receptor transactivation by G-protein-coupled receptors requires metalloproteinase cleavage of proHB-EGF, *Nature* **402**, 884–888.

Quignard, J.-F., Mironneau, J., Carricaburu, V., Fournier, B., Babich, A., Nurnberg, B., Mironneau, C., Macrez, N. (2001) Phosphoinositide 3-kinase gamma mediates angiotensin II-induced stimulation of L-type calcium channels in vascular myocytes, *J. Biol. Chem.* **276**, 32545–32551.

Rana, B.K., Shiina, T., Insel, P.A. (2001) Genteic variations and polymorphisms of G protein-coupled receptors: Functional and Therapeutic Implications, *Annu. Rev. Pharmacol. Toxicol.* **41**, 593–624.

Rasenick, M.M., Wang, N. (1988) Exchange of guanine-nucleotides between tubulin and GTP-binding proteins that regulate adenylate-cyclase-cytoskeletal modification of neuronal signal transduction, *J. Neurochem.* **51**, 300–311.

Rasenick, M.M., Stein, P.J., Bitensky, M.W. (1981) The regulatory subunit of adenylate cyclase interacts with cytoskeletal components, *Nature* **294**, 560–562.

Razani, B., Rubin, C.S., Lisanti, M.P. (1999) Regulation of cAMP-mediated signal transduction via interaction of caveolins with the catalytic subunit of protein kinase A, *J. Biol. Chem.* **274**, 26353–26360.

Rebois, R.V., Warner, D.R., Basi, N.S. (1997) Does subunit dissociation necessarily accompany the activation of all heterotrimeric G proteins?, *Cell. Signal.* **9**, 141–151.

Rhee, S.G. (2001) Regulation of phosphoinositede-specific phospholipase C, *Annu. Rev. Biochem.* **70**, 281–312.

Rios, C.D., Jordan, B.A., Gomes, I., Devi, L.A. (2001) G-protein-coupled receptor dimerization: Modulation of receptor function, *Pharmacol. Ther.* **92**, 71–87.

Robillard, L., Ethier, N., Lachance, M., Hebert, T.E. (2000) $G\beta\gamma$ subunit combinations differentially modulate receptor and effector coupling *in vivo*, *Cell. Signal.* **12**, 673–682.

Robinson, P.R., Cohen, G.B., Zhukovsky, E.A., Oprian, D.D. (1992) Constitutively active mutants of rhodopsin, *Neuron* **9**, 719–725.

Robishaw, J.D., Berlot, C.H. (2004) Translating G protein subunit diversity into functional specificity, *Curr. Opin. Cell Biol.* **16**, 206–209.

Rocheville, M., Lange, D.C., Kumar, U., Patel, S.C., Patel, R.C., Patel, Y.C. (2000) Receptors for dopamine and somatostatin: Formation of hetero-oligomers with enhanced functional activity, *Science* **288**, 154–157.

Rodbell, M. (1972) *Glucagon Molecular Physiology Clinical Therapeutic Implications*, Pergamon Press, Oxford, New York.

Rodbell, M., Birnbaumer, L., Pohl, S.L., Krans, H.M.J. (1971) The glucagon-sensitive adenyl cyclase system in plasma membranes of rat liver V: An obligatory role of guanyl nucleotides in glucagon action, *J. Biol. Chem.* **246**, 1877–1882.

Ross, E.M., Wilkie, T.M. (2000) GTPase-activating proteins for heterotrimeric G proteins: Regulators of G protein signaling (RGS) and RGS-like proteins, *Annu. Rev. Biochem.* **69**, 795–827.

Rossomando, A.J., Payne, D.M., Weber, M.J., Sturgill, T.W. (1989) Evidence that pp42, a major tyrosine kinase target protein, is a mitogen-activated serine/threonine protein kinase, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6940–6943.

Roychowdhury, S., Rasenick, M.M. (1994) Tubulin-G protein association stabilizes GTP binding and activates GTPase: Cytoskeletal participation in neuronal signal transduction, *J. Biochem.* **33**, 9800–9805.

Roychowdhury, S., Rasenick, M.M. (1997) G protein $\beta_1\gamma_2$ subunits promote microtubule assembly, *J. Biol. Chem.* **272**, 31576–31581.

Roychowdhury, S., Wang, N., Rasenick, M.M. (1993) G protein binding and G protein activation by nucleotide transfer involve distinct domains on tubulin: Regulation of signal transduction by cytoskeletal elements, *J. Biochem.* **32**, 4955–4961.

Roychowdhury, S., Panda, D., Wilson, L., Rasenick, M.M. (1999) G protein α subunits activate tubulin GTPase and modulate microtubule polymerization dynamics, *J. Biol. Chem.* **274**, 13485–13490.

Rybin, V.O., Xu, X., Lisanti, M.P., Steinberg, S.F. (2000) Differential targeting of β-adrenergic receptor subtypes and adenylyl cyclase to cardiomyocyte caveolae: A mechanism to functionally regulate the cAMP signaling pathway, *J. Biol. Chem.* **275**, 41447–41457.

Sadja, R., Alagem, N., Reuveny, E. (2003) Gating of GIRK channels: Details of an intricate, membrane-delimited signaling complex, *Neuron* **39**, 9–12.

Sarndahl, E., Bokoch, G.M., Stendahl, O., Andersson, T. (1993) Stimulus-induced dissociation of α subunits of heterotrimeric GTP-binding proteins from the cytoskeleton of human neutrophils, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6552–6556.

Sato, M., Ribas, C., Hildebrandt, J.D., Lanier, S.M. (1996) Characterization of a G-protein activator in the neuroblastoma-glioma cell hybrid NG108–15, *J. Biol. Chem.* **271**, 30052–30060.

Schaefer, M., Petronczki, M., Dorner, D., Forte, M., Knoblich, J.A. (2001) Heterotrimeric G proteins direct two modes of asymmetric cell division in the *Drosophila* nervous system, *Cell* **107**, 183–194.

Schiff, M.L., Siderovski, D.P., Jordan, J.D., Brothers, G., Snow, B., De Vries, L., Ortiz, D.F., Diverse-Pierluissi, M. (2000) Tyrosine-kinase-dependent recruitment of RGS12 to the N-type calcium channel, *Nature* **408**, 723–727.

Schramm, N.L., Limbird, L.E. (1999) Stimulation of mitogen-activated protein kinase by G protein-coupled α_2-adrenergic receptors does not require agonist-elicited endocytosis, *J. Biol. Chem.* **274**, 24935–24940.

Schulz, R. (2001) The pharmacology of phosducin, *Pharmacol. Res.* **43**, 1–10.

Scott, J.K., Huang, S.-F., Gangadhar, B.P., Samoriski, G.M., Clapp, P., Gross, R.A., Taussig, R., Smrcka, A.V. (2001) Evidence that a protein–protein interaction 'hot spot' on heterotrimeric G protein subunits is used for recognition of a subclass of effectors, *EMBO J.* **20**, 767–776.

Seachrist, J.L., Anborgh, P.H., Ferguson, S.S.G. (2000) β_2-Adrenergic receptor internalization, endosomal sorting, and plasma membrane recycling are regulated by Rab GTPases, *J. Biol. Chem.* **275**, 27221–27228.

Seamon, K.B., Daly, J.W. (1981) Forskolin: A unique diterpene activator of cyclic AMP-generating systems, *J. Cyclic Nucleotide Res.* **7**, 201–224.

Seno, K., Kishigami, A., Ihara, S., Maeda, T., Bondarenko, V.A., Nishizawa, Y., Usukura, J., Yamazaki, A., Hayashi, F. (1998) A possible role of RGS9 in phototransduction: A bridge between the cGMP-phosphodiesterase system and the guanylyl cyclase system, *J. Biol. Chem.* **273**, 22169–22172.

Shenker, A., Laue, L., Kosugi, S., Merendino, J.J., Minegishi, T., Cutler, G.B. (1993) A constitutively activating mutation of the luteinizing hormone receptor in familial male precocious puberty, *Nature* **365**, 652–654.

Shieh, B.H., Zhu, M.Y. (1996) Regulation of the TRP Ca^{2+} channel by INAD in *Drosophila* photoreceptors, *Neuron* **16**, 991–998.

Shih, M., Lin, F., Scott, J.D., Wang, H.Y., Malbon, C.C. (1999) Dynamic complexes of β_2-adrenergic receptors with protein kinases and phosphatases and the role of gravin, *J. Biol. Chem.* **274**, 1588–1595.

Siderovski, D.P., Strockbine, B., Behe, C.I. (1999) Whither goest the RGS proteins?, *Crit. Rev. Biochem. Mol. Biol.* **34**, 215–251.

Siffert, W. (2005) G protein polymorphisms in hypertension, atherosclerosis, and diabetes, *Annu. Rev. Med.* **56**, 17–28.

Sitaramayya, A., Harkness, J., Parkes, J.H., Gonzalez-Oliva, C., Liebman, P.A. (1986) Kinetic studies suggest that light-activated cyclic GMP phosphodiesterase is a complex with G-protein subunits, *J. Biochem.* **25**, 651–656.

Skiba, N.P., Yang, C.-S., Huang, T., Bae, H., Hamm, H.E. (1999) The α-helical domain of $G\alpha_t$ determines specific interaction with regulator of G protein signaling 9, *J. Biol. Chem.* **274**, 8770–8778.

Slesinger, P.A., Reuveny, E., Jan, Y.N., Jan, L.Y. (1995) Identification of structural elements involved in G protein gating of the GIRK1 potassium channel, *Neuron* **15**, 1145–1156.

Smine, A., Xu, X., Nishiyama, K., Katada, T., Gambetti, P., Yadav, S.P., Wu, X., Shi, Y.-C., Yasuhara, S., Homburger, V., Okamoto, T. (1998) Regulation of brain G-protein Go by Alzheimer's disease gene presenilin-1, *J. Biol. Chem.* **273**, 16281–16288.

Smotrys, J.E., Linder, M.E. (2004) Palmitoylation of intracellular signaling proteins: Regulation and function, *Annu. Rev. Biochem.* **73**, 559–587.

Solomon, K.R., Kurt-Jones, E.A., Saladino, R.A., Stack, A.M., Dunn, I.F., Ferretti, M., Golenbock, D., Fleisher, G.R., Finberg, R.W. (1998) Heterotrimeric G proteins physically associated with the lipopolysaccharide receptor CD14 modulate both in vivo and in vitro responses to lipopolysaccharide, *J. Clin. Invest.* **102**, 2019–2027.

Sondek, J., Siderovski, D.P. (2001) Gγ-like (ggl) domains: New frontiers in G-protein signaling and β-propeller scaffolding, *Biochem. Pharmacol.* **61**, 1329–1337.

Sorota, S., Rybina, I., Yamamoto, A., Du, X.-Y. (1999) Isoprenaline can activate the acetylcholine-induced K^+ current in canine atrial myocytes via G_s-derived $\beta\gamma$ subunits, *J. Physiol.* **514**, 413–423.

Spiegel, A.M., Weinstein, L.S. (2004) Inherited diseases involving G proteins and G protein-coupled receptors, *Annu. Rev. Med.* **55**, 27–39.

Steinberg, S.F., Brunton, L.L. (2001) Compartmentation of G protein-coupled signaling pathways in cardiac myocytes, *Annu. Rev. Pharmacol. Toxicol.* **41**, 751–773.

Stephens, L.R., Eguinoa, A., Erdjument-Bromage, H., Lui, M., Cooke, F., Coadwell, J., Smrcka, A.S., Thelen, M., Cadwallader, K., Tempst, P., Hawkins, P.T. (1997) The G$\beta\gamma$ sensitivity of a PI3K is dependent upon a tightly associated adaptor, p101, *Cell* **89**, 105–114.

Stoyanov, B., Volinia, S., Hanck, T., Rubio, I., Loubtchenkov, M., Malek, D., Stoyanova, S., Vanhaesebroeck, B., Dhand, R., Nurnberg, B., Gierschik, P., Seedorf, K., Hsuan, J.J., Waterfield, M.D., Wetzker, R. (1995) Cloning and characterization of a G protein-activated human phosphoinositide-3 kinase, *Science* **269**, 690–693.

Strittmatter, S., Valenzuela, D., Sudo, Y., Linder, M., Fishman, M. (1991) An intracellular guanine nucleotide release protein for G0. GAP-43 stimulates isolated alpha subunits by a novel mechanism, *J. Biol. Chem.* **266**, 22465–22471.

Sukumar, M., Ross, E.M., Higashijima, T. (1997) A G(s)-selective analog of the receptor-mimetic peptide mastoparan binds to G(s)α in a kinked helical conformation, *J. Biochem.* **36**, 3632–3639.

Svoboda, P., Milligan, G. (1994) Agonist-induced transfer of the alpha subunits of the guanine-nucleotide-binding regulatory proteins G_q and G_{11} and of muscarinic m1 acetylcholine receptors from plasma membranes to a light-vesicular membrane fraction, *Eur. J. Biochem.* **224**, 455–462.

Tolkovsky, A.M., Levitzki, A. (1981) Theories and predictions of models describing sequential interactions between the receptor, the GTP regulatory unit, and the catalytic unit of hormone dependent adenylate cyclases, *J. Cyclic Nucleotide Res.* **7**, 139–150.

Toya, Y., Schwencke, C., Couet, J., Lisanti, M.P., Ishikawa, Y. (1998) Inhibition of adenylyl cyclase by caveolin peptides, *Endocrinology* **139**, 2025–2031.

Toyoshige, M., Basi, N.S., Rebois, R.V. (1996) Chloride effects on Gs subunit dissociation: Fluoroaluminate binding to Gs does not cause subunit dissociation in the absence of chloride ion, *J. Biol. Chem.* **271**, 8791–8795.

Toyoshige, M., Okuya, S., Rebois, R.V. (1994) Choleragen catalyzes ADP-ribosylation of the stimulatory G protein heterotrimer but not is free α-subunit, *J. Biochem.* **33**, 4865–4871.

Tsao, P., Cao, T., von Zastrow, M. (2001) Role of endocytosis in mediating downregulation of G-protein-coupled receptors, *Trends Pharmacol. Sci.* **22**, 91–96.

Tsao, P.I., von Zastrow, M. (2001) Diversity and specificity in the regulated endocytic membrane trafficking of G-protein-coupled receptors, *Pharmacol. Ther.* **89**, 139–147.

Tsunoda, S., Zuker, C.S. (1999) The organization of INAD-signaling complexes by a multivalent PDZ domain protein in *Drosophila* photoreceptor cells ensures sensitivity and speed of signaling, *Cell Calcium* **26**, 165–171.

Tsunoda, S., Sierralta, J., Sun, Y., Bodner, R., Suzuki, E., Becker, A., Socolich, M., Zuker, C.S. (1997) A multivalent PDZ-domain protein assembles signalling complexes in a G-protein-coupled cascade, *Nature* **388**, 243–249.

van Biesen, T., Luttrell, L.M., Hawes, B.E., Lefkowitz, R.J. (1996) Mitogenic signaling via G protein-coupled receptors, *Endocr. Rev.* **17**, 698–714.

Vanhaesebroeck, B., Waterfield, M.D. (1999) Signaling by distinct classes of phosphoinositide 3-kinases, *Exp. Cell Res.* **253**, 239–254.

Vassilev, L.T., Vu, B.T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., Fotouhi, N.,

Liu, E.A. (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2, *Science* **303**, 844–848.

Vaziri, C., Downes, C.P. (1992) Association of a receptor and G-protein-regulated phospholipase C with the cytoskeleton, *J. Biol. Chem.* **267**, 22973–22981.

Viard, P., Exner, T., Maier, U., Mironneau, J., Nurnberg, B., Macrez, N. (1999) $G\beta\gamma$ dimers stimulate vascular L-type Ca2+ channels via phosphoinositide 3-kinase, *FASEB J.* **13**, 685–694.

Volpicelli, L.A., Lah, J.J., Levey, A.I. (2001) Rab5-dependent trafficking of the m4 muscarinic acetylcholine receptor to the plasma membrane, early endosomes, and multivesicular bodies, *J. Biol. Chem.* **276**, 47590–47598.

Wang, N., Rasenick, M.M. (1991) Tubulin-G protein interactions involve microtubule polymerization domains, *J. Biochem.* **30**, 10957–10965.

Wang, N., Yan, K., Rasenick, M. (1990) Tubulin binds specifically to the signal-transducing proteins, $G_s\alpha$ and $G_i\alpha_1$, *J. Biol. Chem.* **265**, 1239–1242.

Wedegaertner, P., Bourne, H., von Zastrow, M. (1996) Activation-induced subcellular redistribution of $G_s\alpha$, *Mol. Biol. Cell* **7**, 1225–1233.

Wetzker, R., Bohmer, F.-D. (2003) Transactivation joins multiple tracks to the ERK/MAPK cascade, *Nat. Rev. Mol. Cell Biol.* **4**, 651–657.

White, J.H., Wise, A., Main, M.J., Green, A., Fraser, N.J., Disney, G.H., Barnes, A.A., Emson, P., Foord, S.M., Marshall, F.H. (1998) Heterodimerization is required for the formation of a functional GABA(B) receptor, *Nature* **396**, 679–682.

Whitmarsh, A.J., Davis, R.J. (1998) Structural organization of MAP-kinase signaling modules by scaffold proteins in yeast and mammals, *Trends Biochem. Sci.* **23**, 481–485.

Wickman, K.D., Jiñiguez-Lluhi, o.A., Davenport, P.A., Taussig, R., Krapivinsky, G.B., Linder, M.E., Gilman, A.G., Clapham, D.E. (1994) Recombinant G-protein $\beta\gamma$-subunits activate the muscarinic-gated atrial potassium channel, *Nature* **368**, 255–257.

Willard, F.S., Crouch, M.F. (2000) Nuclear and cytoskeletal translocation and localization of heterotrimeric G-proteins, *Immunol. Cell Biol.* **78**, 387–394.

Witherow, D.S., Wang, Q., Levay, K., Cabrera, J.L., Chen, J., Willars, G.B., Slepak, V.Z. (2000) Complexes of the G protein subunit $G\beta_5$ with the regulators of G protein signaling RGS7 and RGS9: Characterization in native tissues and in transfected cells, *J. Biol. Chem.* **275**, 24872–24880.

Wu, H.-C., Chiu, C.-Y., Huang, P.-H., Lin, C.-T. (2001) The association of heterotrimeric GTP-binding protein (G_o) with microtubules, *J. Biomed. Sci.* **8**, 349–358.

Würtele, M., Jelich-Ottmann, C., Wittinghofer, A., Oecking, C. (2003) Structural view of a fungal toxin acting on a 14-3-3 regulatory complex, *EMBO J.* **22**, 987–994.

Xu, X.Z., Choudhury, A., Li, X., Montell, C. (1998) Coordination of an array of signaling proteins through homo- and heteromeric interactions between PDZ domains and target proteins, *J. Cell Biol.* **142**, 545–555.

Yamada, M., Inanobe, A., Kurachi, Y. (1998) G protein regulation of potassium ion channels, *Pharmacol. Rev.* **50**, 723–757.

Yamane, H.K., Fung, B.K.K. (1993) Covalent modifications of G-proteins, *Annu. Rev. Pharmacol. Toxicol.* **33**, 201–241.

Yan, K., Greene, E., Belga, F., Rasenick, M.M. (1996) Synaptic membrane G proteins are complexed with tubulin in situ, *J. Neurochem.* **66**, 1489–1495.

Yan, K., Popova, J.S., Moss, A., Shah, B., Rasenick, M.M. (2001) Tubulin stimulates adenylyl cyclase activity in C6 glioma cells by bypassing the β-adrenergic receptor: A potential mechanism of G protein activation, *J. Neurochem.* **76**, 182–190.

Yoshida, T., Willardson, B.M., Wilkins, J.F., Jensen, G.J., Thornton, B.D., Bitensky, M.W. (1994) The phosphorylation state of phosducin determines its ability to block transducin subunit interactions and inhibit transducin binding to activated rhodopsin, *J. Biol. Chem.* **269**, 24050–24057.

Zamponi, G.W. (2001) Determinants of G protein inhibition of presynaptic calcium channels, *Cell Biochem. Biophys.* **34**, 79–84.

Zeng, W., Xu, X., Popov, S., Mukhopadhyay, S., Chidiac, P., Swistok, J., Danho, W., Yagaloff, K.A., Fisher, S.L., Ross, E.M., Muallem, S., Wilkie, T.M. (1998) The N-terminal domain of RGS4 confers receptor-selective inhibition of G protein signaling, *J. Biol. Chem.* **273**, 34687–34690.

Zhang, C., Qiu, H.E., Krafft, G.A., Klein, W.L. (1996) Protein kinase C and F-actin are essential for stimulation of neuronal FAK tyrosine phosphorylation by G-proteins and amyloid beta protein, *FEBS Lett.* **386**, 185–188.

Zhang, F.L., Casey, P.J. (1996) Protein prenylation: Molecular mechanisms and functional consequences, *Annu. Rev. Biochem.* **65**, 241–269.

Single Biomolecule Detection, Characterization and Reactions: *see* Nucleic Acid and Protein Single Molecule Detection and Characterization

Single-Cell Molecular Neurobiology: *see* Molecular Neurobiology, Single-Cell

Single-Neuron Molecular and Cell Biology: *see* Molecular Neurobiology, Single-Cell

Signaling Through JAKs and STATs: Interferons Lead the Way

Christian Schindler and Jessica Melillo
Columbia University, New York, USA

1 Introduction 118

2 Discovery of the JAK-STAT Signaling Paradigm 120

3 The Janus Kinases (JAK) Family 123
3.1 JAK Structure 123
3.2 JAK Family Members 124
3.2.1 Jak1 124
3.2.2 Jak2 124
3.2.3 Jak3 124
3.2.4 Tyk2 125

4 The STAT Family of Transcription Factors 125
4.1 STAT Evolution 125
4.2 STAT Structure 126
4.2.1 NH_2 Domain 126
4.2.2 Coiled-coil Domain 127
4.2.3 DNA Binding Domain 127
4.2.4 Linker Domain 127
4.2.5 SH2 and Tyrosine Activation Domain 127
4.2.6 Transcription Activation Domain 128
4.3 STAT Family Members 128
4.3.1 Stat1 128
4.3.2 Stat2 129
4.3.3 Stat3 129
4.3.4 Stat4 129
4.3.5 Stat5a and Stat5b 129
4.3.6 Stat6 130

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

5 Hematopoietin Family of Cytokine Receptors 130
5.1 The Cytokine Type II or IFN Receptor Family 130
5.1.1 Type I IFN Receptor 130
5.1.2 Type II IFN Receptor 132
5.1.3 IFN-λ Receptor 132
5.1.4 IL-10 Receptor Family 132
5.2 Extended gp130 Receptor Family 133
5.2.1 IL6-gp130 Receptor Subfamily 133
5.2.2 gp130 Receptor-like Subfamily 134
5.2.3 IL-12 Family of gp130-Related Receptors 134
5.3 γC Receptor Family 135
5.3.1 IL-2 Receptor Family 135
5.3.2 IL-4 Receptor Family 135
5.4 IL-3 Receptor Family 136
5.5 Single-chain Receptor Family 136
5.6 Noncytokine Receptors 136

6 Regulation of JAK-STAT Signaling 136
6.1 Phosphatases 137
6.2 Nuclear Export 137
6.3 The SOCS Family 137

7 Concluding Comments 138

Bibliography 138
Books and Reviews 138
Primary Literature 139

Keywords

Cytokines
A large family of secreted peptide ligands that bind to specific receptors on target cells to mediate potent biological responses. This includes immunomodulatory peptides from a number of important families including the TGF-β, TNF, and hematopoietin families.

GAS (Gamma Activation Site)
The enhancer element recognized by most activated STAT homodimers. It was first identified as an IFN-γ response element.

Hematopoietins
These four-helix-bundle cytokines mediate potent biological responses on target cells (see Table 1). They are predominantly involved in regulating inflammation and immunity.

IFN (Interferon)
A family of cytokines initially identified for their antiviral activity. This family includes the type I, type II, and λ IFNs.

IRFs (Interferon Response Factors)
A family of transcription factors that is important in immune response. They bind to the family of IFN response elements (IREs).

ISGF-3
An IFN-α-activated transcription factor consisting of Stat1, Stat2, and IRF-9.

ISRE (IFN-stimulated Response Element)
A specific enhancer that is recognized by ISGF-3. They constitute a subset of the IRE enhancer family.

JAK (Janus kinase)
Tyrosine kinase that associates with cytokine receptors and provides catalytic activity, which is essential to initiate signal transduction. Members include Jak1, Jak2, Jak3, and Tyk2.

Ligand
The specific peptide that binds to a receptor.

NK Cells (Natural Killer Cells)
A family of lymphocytes that are important in innate immunity. They are potently activated by IFN-I and secrete IFN-γ.

Protein Tyrosine Kinases
Enzymes that catalyze phosphorylation of protein tyrosine residues, which serve an important regulatory role in higher eukaryotes.

Receptors
Transmembrane-spanning proteins that specifically bind extracellular ligands, like hematopoietins, to mediate a biological response.

SH2 (Src Homology 2)
The domain that mediates protein–protein interactions through its capacity to recognize and bind to phosphotyrosines in the context of a specific sequence of amino acids, referred to as a *tyrosine motif* (see final entry). SH2-phosphotyrosine interactions play an important role in signal transduction.

SOCS (Suppressors of Cytokine Signaling)
A family of rapidly induced proteins that feed back inhibit JAK-STAT signaling.

STAT (Signal Transducers and Activators of Transcription)
A family of latent cytoplasmic transcription factors that transduces signals for cytokines. Members include Stat1, Stat2, Stat3, Stat4, Stat5a, Stat5b, and Stat6.

T cells
Thymic derived lymphocyte family whose members are activated by specific antigen-MHC complex. They constitute an important limb of adaptive immunity. They can be divided into several subsets including Th1 and Th2 cells.

Tyrosine Motif
A phosphotyrosine and the four-carboxy-terminal flanking amino acids that constitute an SH2 domain recognition/binding site.

Characterization of the "viral interfering activity" secreted by infected tissues led to the identification of interferons, the founding members of the cytokine family. Subsequent studies, directed at understanding how IFNs mediated their biological response then led to identification of the JAK-STAT signaling cascade. Today, the important role this pathway plays in transducing signals for the over 50 members of the hematopoietin family of cytokines is widely appreciated. After providing a brief historical perspective, this review will provide an overview of the JAK-STAT signaling cascade, with emphasis on some of the more recent developments in this field.

1 Introduction

In 1957, Isaacs and Lindenmann first described the potent antiviral properties of interferons (IFNs). It was not until 20 years later, when the IFNs had been purified to homogeneity, that the size and diversity of this family was recognized. They were segregated into two distinct families, now referred to as type I (i.e. IFN-α/β) and type II (i.e. IFN-γ) IFNs. The early generation IFN cDNA clones, circa 1980, served to propel them as one of the first products of a nascent biotechnology industry. Concomitantly, other secreted peptides, also with potent biological activity, were identified. These were initially referred to as *monokines* or *lymphokines*, reflecting their cellular source (i.e. monocytes and lymphocytes respectively). This confusing terminology was later replaced by more generic terms, interleukins (i.e. between leukocytes) and cytokines. The cytokine family of biological response modifiers has subsequently grown to number in the hundreds, including members of the four-helix bundle, TNF, TGF-β, growth factor, and chemokines families of ligands.

The early availability of relatively large quantities of purified recombinant IFNs provided an opportunity to determine

how these cytokines mediate their potent biological responses. Initial studies determined that the rapid induction of genes were an important feature of this response. Subsequent characterization of the ability of IFN-α to drive the expression of new genes led to the identification of Stat1, Stat2, and the JAK-STAT signaling cascade (STAT refers to signal transducers and activators of transcription and JAK refers to Janus kinase). Further studies determined that IFN-γ also signals through an analogous, but distinct JAK-STAT pathway.

Within three years, five additional members of the STAT family had been cloned and found to transduce signals for other members of the cytokine family of ligands. Subsequent gene targeting studies have confirmed the pivotal role the STATs play in transducing the biological response for each four-helix-bundle cytokines (see Table 1). More recent efforts have focused on understanding how the transient nature of STAT signaling is regulated. These studies have revealed an elaborate system of controls, including tight regulation

Tab. 1 Hematopoietin dependent JAK-STAT signaling. Hematopoietins transduce their signals through specific sets of JAKs and STATs as indicated. The assignments for which there is the most confidence (i.e. based on knockout and biochemical studies) are shown in bold. Those with less confidence are shown in plain lettering, and those with the least confidence are shown in brackets.

Ligands	*JAKs*	*STATs*
IFN family		
IFN-I (Type I)[a]	**Jak1**, Tyk2	**Stat1**, **Stat2**, Stats3, Stat4 (Stats5-6)
IFN-γ (Type II)	**Jak1**, **Jak2**	**Stat1**
IFN-λ (IL-28a,b,-29)	Jak1, Tyk2	Stat1, Stat2, Stat3
IL-10	**Jak1**, Tyk2	**Stat3**, Stat1
IL-19	?	?
IL-20	?	Stat3
IL-22	Jak1, Tyk2	Stat3, (Stat5)
IL-24	?	Stat3
IL-26 (AK155)	?	Stat1/Stat3
Extended gp130 family		
IL-6	**Jak1**, (Jak2)	**Stat3**, Stat1
IL-11	Jak1	**Stat3**, Stat1
LIF	**Jak1**, (Jak2)	**Stat3**, Stat1
CNTF	Jak1, (Jak2)	**Stat3**, Stat1
NNT-1/BSF-3	Jak1, (Jak2)	**Stat3**, Stat1
NP	Jak1, (Jak2)	Stat3
CT-1	Jak1, (Jak2)	**Stat3**
OSM	Jak1, (Jak2)	**Stat3**, Stat1
IL-31	Jak1, (Jak2)	Stat3, Stat5
G-CSF	Jak1, (Jak2)	**Stat3**
Leptin	Jak2	**Stat3**
IL-12	**Tyk2**, Jak2	**Stat4**
IL-23	?	**Stat4**
IL-27	?	(Stat1), Stat3, Stat4, (Stat5)
IL-30	?	(Stat1), Stat3, Stat4, (Stat5)

(*continued overleaf*)

Tab. 1 (*Continued*)

Ligands	*JAKs*	*STATs*
γC family		
IL-2	**Jak1**, **Jak3**	**Stat5**, (Stat3)
IL-7	**Jak1**, **Jak3**	**Stat5**, (Stat3)
TSLP[b]	?	Stat5
IL-9	Jak1, Jak3	**Stat5**, Stat3
IL-15	**Jak1**, Jak3	**Stat5**, (Stat3)
IL-21	Jak3, (Jak1),	**Stat3**, Stat5, (Stat1)
IL-4	**Jak1**, **Jak3**	**Stat6**
IL-13[b]	Jak1, Jak2	**Stat6**, (Stat3)
IL-3 family		
IL-3	**Jak2**	**Stat5**
IL-5	**Jak2**	**Stat5**
GM-CSF	**Jak2**	**Stat5**
Single-chain family		
Epo	**Jak2**	**Stat5**
GH	Jak2	**Stat5**, (Stat3)
Prl	Jak2	**Stat5**
Tpo	**Jak2**	**Stat5**

[a] In humans this family consists of 12 IFN-αs, IFN-β.ω and Limitin.
[b] Bind to a related, but γc independent receptor.

of nuclear import–export, phosphorylation, as well as novel regulatory proteins, most notably the SOCS (suppressors of cytokine signaling) family. As there are many excellent reviews on cytokines (i.e. hematopoietins) and JAK-STAT signaling, this review will focus on the more recent developments and the important role IFN studies have played in elucidating hematopoietin biology.

2 Discovery of the JAK-STAT Signaling Paradigm

Studies with early preparations of recombinant IFN-Is led to the identification of a set of rapidly induced genes. Classic promoter analysis of these IFN-stimulated genes (ISGs), identified a single conserved IFN-stimulated response element (ISRE), with a consensus of AGTTTN$_3$TTTCC. Three unique ISRE binding complexes were identified in nuclear extracts prepared from IFN-α–stimulated cells. They were named *IFN-stimulated gene factor 1* (ISGF-1; aka IRF-2), ISGF-2 (aka IRF-1), and ISGF-3. Kinetic and biochemical studies suggested that ISGF-3 was the critical factor directing the rapid expression of IFN-α stimulated genes. Purification of ISGF-3 led to the identification of four component proteins, with molecular weights of 113 kDa, 91 kDa, 84 kDa, and 48 kDa. The 48-kDa protein was recognized as the ISRE binding component of this complex. Analysis of the three larger proteins determined that they were members of a novel family of proteins, for which the term STAT was coined. The 84 and 91 kDa proteins turned out to

be alternative mRNA splice products of a single gene and were renamed Stat1α and Stat1β. The 113-kDa product became known as Stat2 and the 48-kDa protein, with its homology to interferon response factors (IRFs) was named IRF-9.

With the development of antibodies against Stat1 and Stat2, it was possible to demonstrate that both proteins resided in the cytoplasm of resting cells. Upon IFN-α stimulation, Stat1 and Stat2 (both isoforms) were rapidly phosphorylated on a single conserved tyrosine, leading to the formation of stable Stat1–Stat2 heterodimer. Subsequent biochemical and structural studies determined that STAT dimerization is driven by the stable interaction between the phosphotyrosine residue of one STAT with the SH2 (Src homology 2) domain of the corresponding STAT. In contrast to the inactive STATs, Stat1–Stat2 dimers are competent for translocation into the nucleus, where they associate with IRF-9 to form a stable ISRE DNA binding complex, and direct the expression of target genes (see Fig. 1). Numerous ISRE-driven genes have been characterized, including PKR (the dsRNA dependent protein kinase), OAS (2′-5′ oligoadenylate synthetase), Mx, PML, the

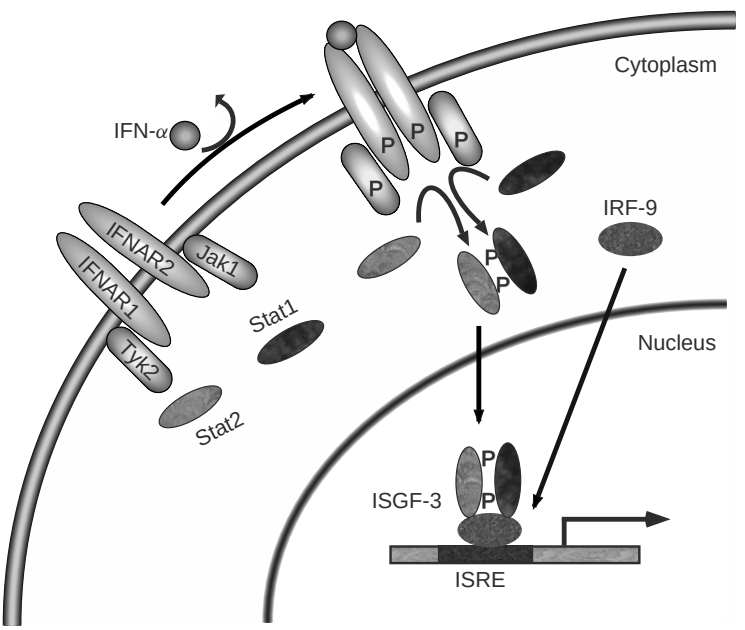

Fig. 1 The IFN-I signaling paradigm. Upon binding to its dimeric receptor, type I IFN promotes the apposition of two receptor associated JAKs, which then activate each other by transphosphorylation. The activated JAKs in turn phosphorylate receptor tyrosine(s), directing the SH2 domain dependent recruitment of Stat1 and Stat2. At the receptor, Stat1 and Stat2 are activated by phosphorylation, whereupon they heterodimerize, translocate into the nucleus, and associate with IRF-9 to bind to the ISRE enhancers. This DNA binding complex, referred to as ISGF-3, directs the expression of ISRE target genes. IFN-I also promotes the formation of Stat1 homodimers, which directs the expression of GAS-driven genes as outlined in Fig. 2.

Ifi 200 cluster, GBP, iNOS (inducible nitric oxide synthetase, aka NOS2), IP-10, IRF-7, 9–27, and ISG-15.

Contemporaneous with the discovery of STATs, was the recognition that members of the JAK family of tyrosine kinases were important in IFN-I signaling. This provided the missing link as to how STATs were activated and the term JAK-STAT Pathway was coined. Parallel efforts to understand how IFN-γ mediated potent but distinct biological responses led to the discovery of an analogous yet different JAK-STAT signaling pathway (see Fig. 2). Notable similarities with the IFN-α signaling cascade included requirement of a unique receptor, JAKs and Stat1. As was the case with IFN-α, elucidation of this pathway began with the identification of an IFN-γ responsive element (i.e. the gamma activation site; GAS) in the promoters of GBP and FcγRI. Subsequently, a unique IFN-γ-induced GAS binding activity was recognized and determined to consist of Stat1. Moreover, IFN-γ was found to stimulate the phosphorylation of the same conserved tyrosine residue on Stat1 as IFN-α. However, in the absence of Stat2 activation, a Stat1 homodimer was formed that translocated into the nucleus and directly bound DNA, culminating in the expression of GAS-driven genes. Subsequent studies determined that all STATs, with the exception of Stat2, form homodimers upon activation and then directly bind members of the GAS family of enhancers. This family of conserved enhancers features a palindrome, where the

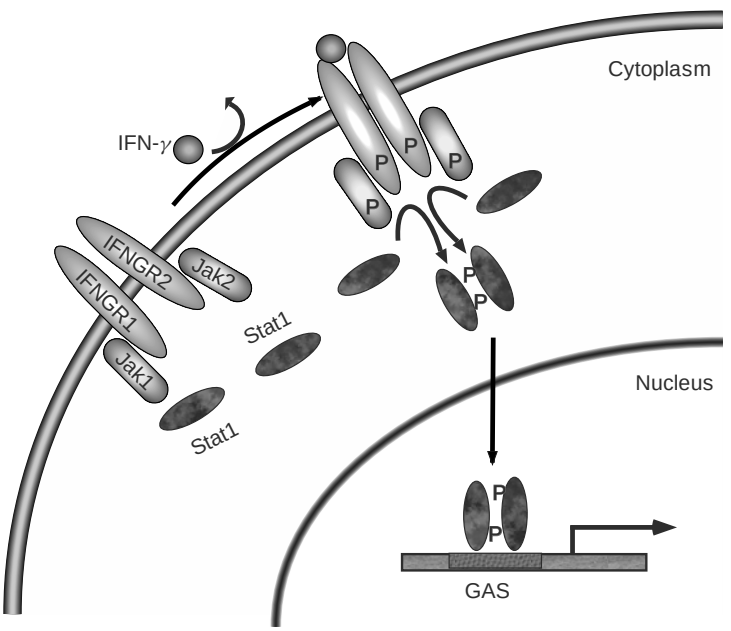

Fig. 2 The IFN-II signaling paradigm. Upon binding to its dimeric receptor, IFN-γ promotes the activation of two receptor associated JAKs. The activated JAKs in turn activate the receptor. Stat1 is recruited to the receptor, whereupon it is activated and forms homodimers. Stat1 homodimers translocate to the nucleus and bind to the GAS family of enhancers, culminating in the expression of a distinct set of genes.

optimal sequence is $TTCN_{2-4}GAA$ (Stat1 homodimers bind to an element where $N = 3$; Stat3 homodimers favor $N = 2$; and Stat6 homodimers favor $N = 4$). IFN-γ stimulated GAS-driven genes include a number of important immunoregulatory transcription factors (e.g. IRF-1, ICSBP and CIITA), chemokines (e.g. Mig), and cell surface molecules (e.g. ICAM-1, Ly6E and FcγRI). CIITA is critical for adaptive immunity, driving the expression of major histocompatibility complex (MHC) class II and the Invariant chain, whereas IRF-1 has more broad effects, directing the expression of immune response genes like *iNOS, MHC I, LMP2, TAP2,* and *IL-12*. The ability of IRF-1 also to drive ISRE gene expression is likely to account for some of the overlap in the biological responses to IFN-α and IFN-γ.

The rapid identification of additional members of the STAT family provided an opportunity to determine that the other four-helix-bundle cytokines (i.e. hematopoietins) also transduce signals through this pathway. As the number of four-helix-bundle cytokines has grown, so has the recognition that they all transduce pivotal signals through JAK-STAT signaling cascade. The critical role hematopoietins play in directing immune response has directed further attention to this important pathway.

3 The Janus Kinases (JAK) Family

Studies directed at understanding how IFNs mediated biological response directly implicated three known members of the JAK family, Jak1, Jak2, and Tyk2 in signal transduction. Subsequent homology screens identified a fourth member, Jak3, whose expression was restricted to hematopoietic tissues. However, each of these JAKs shared important structural and functional features that enable them to serve as an essential kinase for cytokine receptors. This is also the case for a JAK homolog, Hop, which has been identified in *Drosophila melanogaster* (see Sect. 4.1).

3.1 JAK Structure

JAKs range in size from 120 to 130 kDa. Comparison of these ~1000 residue kinases reveals seven high homology regions, JH1–JH7 (see Fig. 3a). JH1 encodes the tyrosine kinase, whose activity is regulated by an activation loop. Thus, in resting cells when the activation loop sits in the catalytic pocket, activity is limited to a low or basal state. However, cytokine dependent ligation brings the two receptor associated JAKs into close apposition, enabling them to activate each other through a transphosphorylation event. This entails exploiting basal activity to phosphorylate the opposing kinases activation loop. The loops move out of the catalytic site, rendering these kinases fully active, whereupon they sequentially phosphorylate critical receptor and STAT tyrosine residues (see Figs. 1 and 2). The second JAK homology domain, JH2, represents a pseudokinase domain (i.e. catalytically inactive) that is both a signature feature of this kinase family and appears to be required for full catalytic activity. The five amino-terminal JAK homology domains, JH3–JH7, constitute a FERM (four point one, ezrin, radixin, moesin) domain, which is responsible for association with hematopoietin family of cytokine receptors. Specifically, JAKs associate with proline rich, membrane proximal box1/box2 domain found in these receptors. As outlined in Sect. 5,

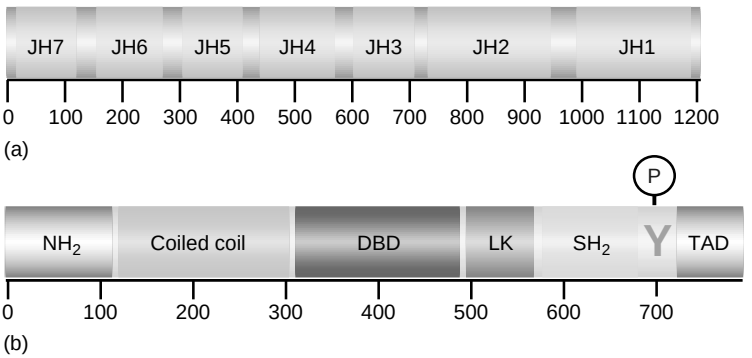

Fig. 3 Structural models of the JAK and STAT family of signaling proteins. (a) The four members of the JAK family share seven homology domains regions, JH1–JH7. JH1 serves as the catalytic domain, whereas JH2 represents a pseudokinase domain, which is a characteristic feature of this family of tyrosine kinases. JH3–JH7 comprise a FERM domain that is responsible for association with cytokine receptors. (b) The seven members of the STAT family of transcription factors share seven functionally conserved domains. This includes the amino-terminal domain (NH_2), the coiled-coiled domain (coiled coil), the DNA binding domain (DBD), the Linker domain (LK), the SH2 domain, the tyrosine activation domain, and the transcriptional activation domain (TAD), which is conserved in function but not in sequence. Domain colors correspond with those in Stat1 Crystal structure in Fig. 4.

the cytokine receptors depend on JAKs to transduce signals that can be functionally divided into several groups, each exhibiting a unique pattern of JAK association and STAT activation (see Table 1).

3.2 JAK Family Members

3.2.1 Jak1

This kinase was initially recognized for its association with the second chain of the IFN-α receptor (IFNAR2) and then the first chain of the IFN-γ receptor (IFNGR1). Subsequent biochemical studies demonstrated association with receptors for the gp130 and the common γ (γC) families. Jak1 knockout mice die perinatally, secondary to neurological defects attributed to a loss in gp130 signaling. Jak1$^{-/-}$ mice also exhibit reduced lymphocyte counts due to a lack of T cells and NK (natural killer) cells. Biochemical studies on Jak1$^{-/-}$ cells have underscored a failure to respond to cytokines from the gp130, γC, and IL-10 families.

3.2.2 Jak2

This kinase was initially determined to associate with IFNGR2 and the gp130 component of the IL-6 receptor. Targeted deletion of Jak2 yielded embryonic lethal phenotype, at day 12.5, due to failure in definitive erythropoiesis. Tissues collected from Jak2 knockout mice fail to respond to IFN-γ and most of the cytokines from the IL-3 and single-chain cytokine receptor families (see Table 1).

3.2.3 Jak3

The expression in this JAK is limited to hematopoietic cells, where it exclusively associates with the γC chain. Hence, the Jak3 knockout mice exhibit a severe combined immunodeficiency disease (SCID)

like syndrome, analogous to what has been reported for γC deficient mice.

3.2.4 Tyk2

Although the first member of the family to be associated with cytokine signaling, it was the last to be knocked. Earlier biochemical studies implicated Tyk2 in IFN-I, IL-10, IL-12, and IL-13 signaling. Yet, Tyk2 knockout mice exhibit surprisingly modest defects in their response to IFN-I and IL-10. They do, however, exhibit significant defects in their response to IL-12, and unexpectedly profound defects in their responses to LPS. While the mechanism behind this latter defect is not fully elucidated, it underscores the intimate relationship between toll-like receptor (TLR) dependent signaling (LPS signals through TLR-4), IFN-I production, and the IFN-I autocrine loop (see also Fig. 5).

4 The STAT Family of Transcription Factors

The rapid identification of seven members of the STAT family facilitated a functional classification of the growing family of cytokine receptors (see the following and Table 1). It also provided an opportunity to explore STAT function. The seven mammalian STATs (Stats 1–6 with two Stat5 genes), range in size from 750 to 900 amino acids. They consist of six distinct domains (Fig. 3b), and except for Stat2, they all transduce signals through the paradigm illustrated in Fig. 2. STATs have also been identified in numerous more primitive model eukaryotes, including *D. melanogaster*, *Caenorhabditis elegans*, and *Dictyostelium*. Growth in the number of STATs during evolution that appears in these lower eukaryotes reflect an increasing need for cell-to-cell communication as eukaryotes have become more complex.

4.1 STAT Evolution

STATs have been remarkably well conserved during eukaryotic evolution and may encode the primordial SH2 domain (see Sect. 4.2.5). Homologs have been identified in *Dictyostelium, C. elegans, Drosophila, Anopheles, Xenopus*, and zebrafish. The JAK-STAT pathway has been genetically characterized in *Drosophila*, where it plays important roles in developmental segmentation, larval hematopoiesis, sexual identity, and the development of eye polarity. The single *Drosophila* JAK exhibits 27% identity to Jak2 and the single STAT exhibits 37% identity to Stat5. Also present in *Drosophila* are a distant homolog of a cytokine receptor (Upd) and ligand (Dom). The three STAT homologs in *Dictyostelium* are distantly related to mammalian Stat5 and Stat3, but the signaling pathways they are involved in are unrelated to the one found in *Drosophila* and vertebrates. STAT homologs in *C. elegans* are also more distantly related. Zebrafish encode two STATs, homologous to Stat1 and Stat3. The Stat3 homolog is expressed in early development and is syntenic with both murine and human Stat3. The zebrafish Stat1 homolog is more divergent and expressed later in life.

The seven murine and human STATs are segregated into three syntenic genetic clusters, each consisting of two tandem STATs and suggesting evolution from a single tandem cluster (see the following). Specifically, in mice, Stat1 and Stat4 map to chromosome 2; Stat2 and Stat6 map to chromosome 12; and Stat3 and Stat5 map to chromosome 17. The Stat5 gene,

however, has undergone an additional evolutionarily more recent duplication, yielding tandem Stat5a and Stat5b genes. Analysis of these clusters, the recent duplication of Stat5, and the sequence conservation of Stat3 and Stat5 amongst lower eukaryotes, all serve to suggest that Stat3–Stat5 gene cluster represents the primordial tandem array mammalian STAT genes. This may also explain the relative functional pleiotropy of these two STATs (see the following).

4.2 STAT Structure

Biochemical, genetic, and structural studies have provided clear evidence that each of the seven mammalian STATs consist of six domains, as well as a tyrosine near residue 700 that serves as the activation site (see Figs. 3b and 4). These functionally conserved domains include: the amino-terminal domain (NH_2), the coiled-coiled domain (CCD), the DNA binding domain (DBD), the linker domain, the SH2 domain and the carboxy-terminal transcriptional activation domain (TAD; conserved in function but not sequence).

4.2.1 NH_2 Domain

The amino-terminal domain, about 130 residues in length, forms a unique dimeric structure when crystallized as an independent domain. Likewise, biochemical and genetic studies of full-length inactive Stat1, Stat3, Stat4, and Stat5 indicate that the amino-terminal domain drives homotypic

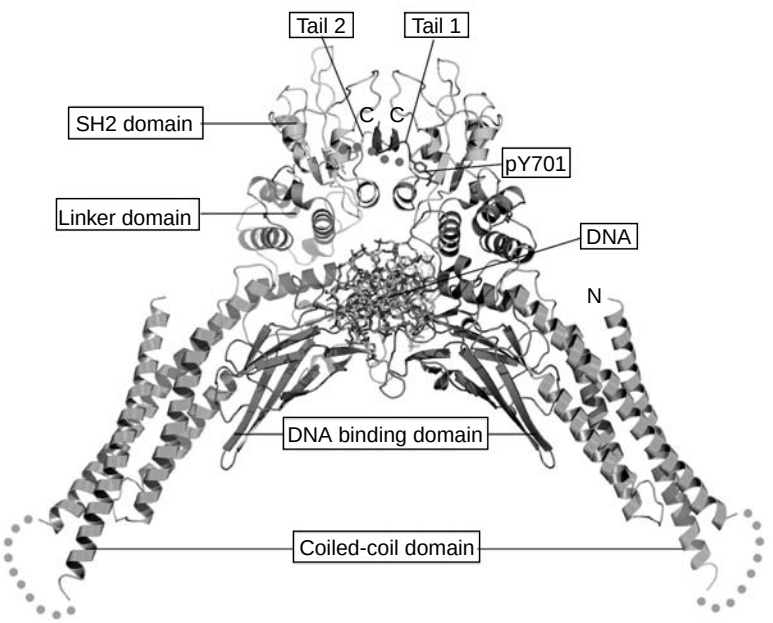

Fig. 4 Stat1 crystal structure. The structure represents two activated Stat1 fragments (amino acids 135–710) that are dimerized and bound to a GAS palindrome. Each domain is shown in a distinct color and labeled (see Fig. 3). Note that the CCD forms a four-α-helix bundle and the DBD forms a β-barrel with an immunoglobulin fold (see color plate p. xlii).

(i.e. to self) dimerization. This may assure delivery of "packets" of STAT dimers to the receptor in a conformation that renders the tyrosine, near residue 700, exposed. Upon activation, STATs undergo a conformation change, in which this phosphotyrosine becomes buried in an activated structure. Earlier studies also implicated the ability of the amino termini to dimerize, thereby promoting cooperativity of DNA binding to tandem GAS elements. Finally, the amino terminus may also serve to regulate STAT nuclear import and export.

4.2.2 Coiled-coil Domain

This amino proximal domain, ~150 residues in length (amino acids 135 to 315 for Stat1), folds up into a four-α-helix bundle (see Fig. 4). This structural feature is likely to afford the ability to dynamically regulate interactions with regulatory proteins (e.g. karyophilins) and a large surface area for protein–protein interactions. Consistent with this, the coiled-coil domain protrudes 80 Å laterally from the core structure of both Stat1 and Stat3, providing a hydrophilic surface that is known to associate with other proteins including IRF9, c-Jun, N-myc interacting protein (Nmi), and StIP.

4.2.3 DNA Binding Domain

This highly conserved ~200 amino acid domain (residues 320–480 in Stat1) forms a β-barrel with an immunoglobulin fold, reminiscent of the NF-κB and p53 DBDs. This affords each STAT component in the dimer the ability to recognize bases in the proximal half of the GAS palindrome. The number of direct contact sites between STAT residues and DNA are modest, yielding a dissociation constant in the nanomolar range, and suggesting that cooperative DNA binding activity may be an important aspect for STAT function. Recall, that Stat2 does not appear to form active homodimers or bind DNA directly.

4.2.4 Linker Domain

As the name suggests, this short domain (amino acids 488–576 for Stat1), links the DBD with the SH2 domain. Detailed analysis of the linker structure suggests that this domain serves to appropriately "translate" the SH2 dependent dimerization to the DBD.

4.2.5 SH2 and Tyrosine Activation Domain

The STAT SH2 domain is the most highly conserved and arguably the most important domain (residues 580–680 in Stat1). It plays critical roles in signal transduction through its capacity to recognize specific receptor phosphotyrosine motifs (i.e. affecting specific receptor recruitment). After activation it specifically recognizes the phosphotyrosine motif on the partner STAT (i.e. directing active STAT dimerization). Moreover, the SH2 domain in *Dictyostelium* STAT appears to represent the "earliest" identified SH2 domain, raising the possibility that it is the primordial SH2 domain. Although the STAT SH2 domain is well conserved evolutionarily, it is surprisingly divergent from other SH2 domains in sequence, but not structure. The structure consists of an antiparallel β-sheet, flanked by two α-helices, forming a pocket. An absolutely conserved arginine (Arg-602 for Stat1) resides at the bottom of this pocket and mediates the interaction with phosphotyrosine. As is the case with all SH2 domains, flanking residues recognize the other amino acids that are characteristic of a tyrosine motif (i.e. phosphotyrosine + four carboxy-terminal amino acids). Of note, the residues most important in defining

the specificity of binding to the STAT tyrosine motif (i.e. the tyrosine activation domain) during dimerization are located at positions +1 and +3 in this motif. In Stat1 residues, +5, +6, and +7 also contribute to the specificity of dimerization by associating with Ala-641 and Val-642 of the SH2 domain. The proximity of the SH2 to the corresponding STAT tyrosine activation domain precludes an intramolecular association between these domains. Thus, the STAT SH2 domain is only able to bind to tyrosine motifs on other proteins.

4.2.6 Transcription Activation Domain

The critical TAD resides at the STAT carboxy terminus. Consistent with the ability of this domain to regulate unique transcriptional responses for each STAT, it is the most poorly conserved STAT domain. The function of this domain can be regulated by either deletion or serine phosphorylation. The former feature was recognized with the initial isolation of Stat1β (i.e. the 84 kDa ISGF-3 component), which is an alternate splice isoform. The *Stat3* gene gives rise to a similar carboxy terminally deleted Stat3 isoform, Stat3β. Intriguingly, recent studies suggest that truncated isoforms are not necessarily transcriptionally inert, but rather can also direct the expression of a unique gene program through the association with distinct sets of transcriptional regulators. Regulation of STAT transcriptional activity by serine kinases has also received considerable attention. These studies have identified numerous candidate serine kinases, potentially reflecting the ability of multiple "stress" pathways to modulate STAT activity. Elegant studies, in which the critical serine residue in Stat1 and Stat3 have been mutated by a knockin approach, have demonstrated that serine phosphorylation is important for the expression of a subset of STAT target genes.

4.3 STAT Family Members

As indicated above, seven STAT genes have been identified in both mouse and man. They include *Stats1–6* and two Stat5 genes, *Stat5a* and *Stat5b*.

4.3.1 Stat1

Both biochemical and gene targeting studies have determined that Stat1 is critical for both type I and type II IFN signaling (see Figs. 1 and 2). Stat1 forms transcriptionally active homodimers in response to stimulation with both classes of IFNs (studies with λ-IFNs are under way; see Sect. 5.1.3). IFN-Is and IFN-λs also direct the formation of active Stat1–Stat2 heterodimers. Intriguingly, the relative role that Stat1 homodimers versus Stat1–Stat2 heterodimers play in the biological response to IFN-Is is tissue specific, with Stat1–Stat2 heterodimers playing a considerably more prominent role in fibroblasts and epithelial cells. Finally, gene targeting studies have also underscored a modest role for Stat1 in the biological response to IL-6 (see Table 1). Specifically, Stat1 appears to alter the nature of the response to Stat3 (the major gp130 activated STAT; see Sects. 4.3.3 and 5.2.2), highlighting the seemingly antagonistic roles Stat1 and Stat3 play in cancer and immune response. That is to say, numerous studies have implicated Stat3 as an oncogene, whereas Stat1 appears to antagonize cell proliferation. Likewise, Stat1 drives the expression of proinflammatory genes (aka "danger signals"), whereas Stat3 activation

can be associated with the suppression of inflammation.

4.3.2 Stat2

Stat2 is a most unique member of the STAT family by virtue of its large size, inability to homodimerize, and inability to directly bind DNA. Sequence comparison of the murine and human homologs also reveals an uncharacteristic level of carboxy-terminal sequence divergence. Both biochemical and gene targeting studies have determined that Stat2 is critical for the biological response to IFN-Is. Upon activation, Stat2 heterodimerizes with its exclusive partner Stat1. This heterodimer is competent for nuclear translocation, but must associate with IRF-9 to form a stable complex on ISRE elements, a signaling paradigm that is unique to Stat2 (see Fig. 1). Biochemical studies have suggested a similar role in the response to λ-IFNs. Although gene targeting studies have highlighted an essential role for Stat2 in IFN-I signaling, there is growing evidence for a regulatory role in some IFN-γ responses.

4.3.3 Stat3

Biochemical and gene targeting studies have demonstrated that Stat3 plays a critical role in mediating the biological response to members of the IL-6/gp130 and IL-10 families of cytokines. Underscoring an important developmental role for gp130 dependent signals, deletion of the *Stat3* gene yields an early embryonic lethal phenotype (i.e. day 6.5–7.5). However, tissue-specific deletions of Stat3 are largely associated with abnormal inflammation and/or decreased susceptibility to cancer. As suggested by the latter phenotype, there is a burgeoning interest in understanding the role Stat3 plays in tumorigenesis. This contrasts the biological activities associated with Stat1, pointing to the seemingly antagonistic relationship between Stat1 and Stat3 with respect to the control of growth and inflammation alluded to in Sect. 4.3.1.

4.3.4 Stat4

Stat4 is predominately involved in signaling for the IL-12 family (see Table 1 and Sect. 5.2.3), which plays a critical role in activating the effector functions of both Th1 and NK cells (e.g. enhanced IFN-γ production). Studies on Stat4 and Stat6 (see Sect. 4.3.6) knockout mice have underscored the important role these two STATs play in maintaining a healthy balance between Th1 and Th2 dependent immune responses. More recent studies have also implicated Stat4 in the ability of IFN-Is to synergize with IL-18 in the activation of Th1 cells. For these responses, IFN-I has been shown to drive Stat4 tyrosine phosphorylation and IL-18 to stimulate Stat4 serine phosphorylation, both of which are required for full activity. Although Stat4 is also important in the biological response to IL-23, a potential role in the biological response to the two newest members of the family, IL-27 and IL-30, remains more controversial.

4.3.5 Stat5a and Stat5b

Consistent with their location in the primordial STAT locus, Stat5a/Stat5b, like Stat3, are relatively pleiotropic in their effects. They transduce important signals for all members of the IL-2 family (all γC receptors except for IL-4 and IL-13), the IL-3 family, and the single-chain receptor family (see Table 1). Underscoring their homology, Stat5a and Stat5b play overlapping roles in mediating the responses to IL-2 and IL-3 ligand families.

Surprisingly however, the responses to members of the single-chain family are far more specific. Stat5a knockout mice exhibit a profound defect in the response to prolactin (PRL), and Stat5b knockout mice exhibit a profound defect in the response to growth hormone (GH). Reproductive defects and an autoimmune phenotype, attributed to a deficient IL-2 response, have made Stat5a/5b double knockout mice difficult to propagate and study.

4.3.6 Stat6

Biochemical and gene targeting studies have determined that Stat6 transduces signals for two closely related and distinct members of the γC family of receptors, the IL-4 and IL-13 receptors. Both receptors, and their corresponding ligands, play an essential role in the development of an effective Th2 dependent response. Thus, as outlined above, Stat4 and Stat6 serve as protagonists in maintaining a critical balance between the Th1 and Th2 subsets of T-helper cells.

5 Hematopoietin Family of Cytokine Receptors

To date, over 50 members of the four-helix-bundle family of cytokines, also referred to as *hematopoietins*, have been identified. They each bind to a conserved family of receptors, referred to as *hematopoietin receptors*, or more commonly known as *type I* and *type II cytokine receptors*, based on the presence or absence, respectively, of a conserved WSXWS amino acid sequence motif in their extracellular domain. Members of the cytokine receptor family type I can be further subdivided into four functionally related groups that transduce signals through analogous JAK-STAT signaling pathways (see Sects. 5.2–5.5; see also Table 1). Members of the cytokine receptor family type II (CRF2) include receptors for the many members of the IFN and IL-10 ligand families. Consistent with this, they also signal through a characteristic, albeit more pleiotropic set of JAK-STAT signaling pathways (see Sect. 5.1; see also Table 1).

5.1 The Cytokine Type II or IFN Receptor Family

IFNs represent the largest single family of cytokines that bind to and signal through receptors composed of CRF2 chains. Until recently, IFNs were divided into two families, type I and type II, on the basis of distinct receptors they bind. However, recently a third family of "λ-IFNs" has been identified, which signal through a unique receptor, consisting of an IL-10 receptor chain and an "orphan" CRF2 chain. The related IL-10 family of ligands also binds to a partially overlapping set of receptors consisting of CRF2 chains.

5.1.1 Type I IFN Receptor

Ligands binding to the type I IFN receptor represent the largest single family of four-helix-bundle cytokines, numbering close to 20 in man and mouse. Referred to as type I IFNs (IFN-Is), they are encoded in a large locus of single exon genes. IFN-Is include numerous IFN-αs, IFN-β, IFN-δ, IFN-ε, IFN-κ, IFN-τ, IFN-ω, and Limitin. Their receptor, which is usually referred to as the IFN-α receptor (IFNAR), is expressed in all cells. It consists of two CRF2 chains, IFNAR1 and IFNAR2. The cytoplasmic domain of IFNAR1 is physically associated with Tyk2,

whereas the cytoplasmic domain of IFNAR2 binds Jak1. As outlined in Fig. 1, upon binding IFN-I, these receptor associated tyrosine kinases become activated by transphosphorylation, whereupon they phosphorylated receptor tyrosine(s) to affect STAT recruitment (see Sect. 3.1). At the receptor, Stat1 and Stat2 become phosphorylated (i.e. activated) and form both Stat1–Stat2 heterodimers and Stat1 homodimers, which in turn direct the expression of ISRE- and GAS-driven genes. The relative balance of these two pathways differs between tissues, with Stat1–Stat2 dimers playing a more important role in fibroblasts and macrophages relying more heavily on Stat1 homodimers. Although IFN-Is were initially recognized for their important role in the innate immune response to viral infection, they are now also recognized for their capacity to regulate the interface between innate and adaptive immunity. Consistent with their important role in immune response, the regulation of IFN-I expression is quite complex (see Fig. 5). During infection with either a viral or bacterial pathogen, pattern recognition receptors, often from the TLR family, become activated. These receptors trigger activation of IKK-ε (aka IKK-i) and TBK-1, culminating in the activation of IRF-3. Activated IRF-3 drives the expression of IFN-β, the prototypical immediate early IFN-I. IFN-β is secreted, and binds to the IFN-α receptor directing the activation of ISGF-3 and expression of ISRE-driven genes. One such target gene, *IRF-7*, either independently or along with *IRF-3* (and potentially other IRFs) directs the robust secretions of the many delayed IFN-Is (e.g. IFN-αs, IFN-δ, IFN-ω, etc.). Intriguingly, the IFN-I autocrine/paracrine loop appears to be active at low levels in resting cells (i.e. basal activity), potentially underscoring its role as a sentry for infection. More recently however, studies have implicated this autocrine loop in the regulation of noninfectious cellular stresses, including cancer and autoimmune disease. Finally, while it is clear that both Stat1 and Stat2 play a critical role in mediating IFN-I response, the role other STATs play in this response has remained more controversial. Recent studies suggest that Stat3, which is activated by IFN-Is in most tissues, may serve to antagonize Stat1 activity, a theme that was outlined in the preceding sections. (Sects.4.3.1 and 4.3.3). In contrast, the IFN-I dependent activation

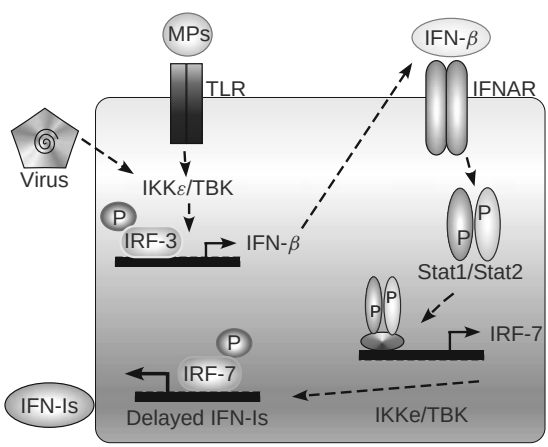

Fig. 5 IFN-I autocrine/paracrine loop. Viral and bacterial pathogens include molecular patterns (MPs) that are recognized by toll-like receptors (TLRs). Once bound to these receptors, they promote the activation of IKKε/TBK1 kinases, which in turn phosphorylate IRF-3. Activated IRF-3 then drives the expression and secretion of IFN-β, an immediate early IFN-I. IFN-β binds the IFN-α receptor (IFNAR) initiating an important autocrine loop. This leads to the expression of IRF-7, an important ISGF-3 target gene, which is responsible for driving the expression/secretion of the delayed IFN-Is.

of Stat4 is limited to leukocytes, where it appears to play an important role in promoting the expression of IFN-γ.

5.1.2 Type II IFN Receptor

There is only one ligand that binds to the type II IFN receptor, IFN-γ. In contrast to the type I IFNs, the gene encoding IFN-γ consists of three exons and resides in a distinct locus. The type II IFN receptor, which is usually referred to as the IFN-γ *receptor*, consists of two CRF2 chains, IFNGR1 and IFNGR2. They associate with Jak1 and Jak2 respectively. In contrast to type I IFNs, IFN-γ secretion is largely restricted to T cells and NK cells, underscoring the need to carefully regulate the activity of this potent proinflammatory cytokine. In T cells, IFN-γ is the signature cytokine of effector Th1 cells, which play a critical role in adaptive immunity. However, NK cell dependent IFN-γ production assures an important role in innate immunity as well. As outlined in Fig. 2, upon binding, ligand two receptor associated JAKs are activated by transphosphorylation, whereupon IFNGR1 becomes phosphorylated on a single tyrosine residue that specifically serves to recruit Stat1 to the receptor complex. Stat1 is activated at the receptor, dimerizes, translocates to the nucleus, and directly binds GAS elements to drive the expression of target genes. Notable GAS-driven target genes include chemokines and transcription regulators. One such transcription regulator is IRF-1, which then directs the expression of ISRE-driven genes, accounting for some of the functional overlap between type I and type II IFNs. Another important transcriptional regulator is CIITA, which coordinates the expression of MHC class II, a signature response associated with IFN-γ stimulation, and is essential for adaptive immunity.

5.1.3 IFN-λ Receptor

The ligands that bind to the IFN-λ receptor are either referred to as λ IFNs or IL-28a, IL28b, and IL-29. Both receptor and ligands were recently identified through genome mining. Although the IFN-λs are functionally quite similar to IFN-Is, they reside in a distinct genomic cluster and are encoded for by multiexon genes, more reminiscent of the gene for IFN-γ. In contrast, the genes encoding IFN-λ receptor chains reside in IFN-I/IL-10 locus (aka CRF2 locus), consisting of CRF2–12 (provisionally named IFN-λR1 or IL-28Rα) and CRF2–4 (a component of the IL-10 receptor; see the following). Although many functional and mechanistic details of this family have not yet been fully elucidated, λ-IFNs do appear to be induced in response to viral infection and predominately signal (at least in fibroblasts) through Stat1–Stat2 heterodimers. The relative role of this unique family during viral or bacterial infections is an area of current investigation.

5.1.4 IL-10 Receptor Family

This family of receptors and their corresponding ligands has grown considerably in size over the last few years. There are five distinct receptor chains that alternatively pair to form at least 5 unique receptors (not including the IFN-λ receptor). They bind and transduce signals for six ligands, including IL-10, IL-19, IL-20, IL-22 (IL-TIF), IL-24 (MDA-7; FISP), and IL-26 (AK-155). IL-10, the founding member of this family, is widely expressed and recognized for its potent and clinically important anti-inflammatory activity. The first evidence that IL-10 might be related to the IFNs came when the sequence of the first IL-10 receptor chain (IL-10R1) was found to be homologous to the IFN-γ receptor. Subsequently, several orphan *CRF2* genes were found to serve as receptors

for IL-10 family members. First was CRF-2–4, identified as the second chain of the IL-10 receptor (IL-10R2). More recently, three additional receptor chains, IL-20R1, IL-20R2, and IL-22R, have been identified. IL-19, IL-20, and IL-24 have been shown to signal through a receptor consisting of IL-20R1 and IL-20R2. However, both IL-20 and IL-24 can also signal through a second receptor consisting of IL-22R and IL-20R2. In contrast, IL-22 appears to signal exclusively through a receptor consisting of IL-22R and IL-10R2, whereas IL-26 appears to signal through a receptor consisting of IL-20R1 and IL-10R2. Of these ligands and receptors, IL-10 and its receptor remain the best characterized. The IL-10 receptor is expressed in most tissues. Its cytoplasmic domains associate with Jak1 and Tyk2 respectively, and serve to recruit and activate Stat3, the same STAT that transduces signals for the gp130 family (Sect. 5.2). Recent studies have provided important insight as to how IL-6 and IL-10 can both transduce signals through Stat3, and yet achieve very different biological responses. It has been determined that because the IL-10 receptor fails to recruit SOCS-3, a potent Stat3 antagonist (see Sect. 6.3), IL-10 directs a significantly prolonged pattern of Stat3 activation, culminating in a distinct pattern of gene expression. In addition, in contrast to gp130 responses (Sects. 4.3.3 and 5.2.1), Stat1 does not appear to play a significant role in directing the biological response to IL-10. With the recent growth in IL-10 like ligands and receptors, much remains to be learned about their biological responses. Initial studies have suggested that IL-19 and IL-20 are preferentially expressed in monocytes, and that IL-20 may play an important role in dermal inflammation. IL-22 has been reported to induce the expression of acute phase response genes in the liver, and IL-24 appears to be a Th2 secreted effector cytokine. Intriguingly, IL-24 appears to selectively antagonize the growth of some tumors. Likewise, IL-26 may also function as a T-cell effector cytokine.

5.2 Extended gp130 Receptor Family

The extended family of gp130 receptors includes a large group of receptors that either employs gp130, or a gp130 homolog as one of their receptor chains. The first, and therefore prototypical member of this family is the receptor for IL-6, consisting of a unique IL-6 ligand-binding chain (IL-6Rα) and gp130. The extended gp130 receptor family can be subdivided into those receptors that are functionally distinct from the IL-6 receptor (i.e. IL-12 family of gp130-related receptors; Sect. 5.2.3) and those that are functionally similar. The latter group can also be further subdivided into two groups, those receptors in which gp130 participates (i.e. IL6-gp130 receptor subfamily; Sect. 5.2.1), and those in which the receptor solely employs gp130 homologs (i.e. gp130-like receptors; Sect. 5.2.2). The discussion of this extended family will start with the prototypical IL6-gp130 receptor subfamily.

5.2.1 IL6-gp130 Receptor Subfamily

gp130 was first identified as the signaling component of the IL-6 receptor. It associates predominantly with Jak2 and encodes four STAT recruitment motifs (referred to as the YXXQ amino acid motifs), which exhibit a differential affinity for Stat1 and Stat3. gp130 also encodes fifth and distinct tyrosine motif (Y759 in mice) that exhibits a unique affinity for SHP2 and SOCS-3. Gene targeting studies creating gp130 chains that either only express a single YXXQ motif

or Y759, have elegantly served to elucidate their roles in mediating the biological response to IL-6. Specifically, the Y759 motif has been shown to direct the activation of an SHP2-MAP kinase signaling cascade that serves to promote cytoprotective (i.e. anti-inflammatory) responses, as well as antagonizing Stat3-dependent signals through the recruitment of SOCS-3. In contrast, YXXQ motifs direct recruitment/activation of Stat3 (and Stat1) to drive the expression of a number of proinflammatory genes, as well as SOCS-3, a Stat3-specific antagonist (Sect. 6.3). These genes are associated with both inflammation and cancer progression. Other IL-6 related ligands that transduce signals through gp130 include, IL-11, CNTF, LIF, OSM, CT-1, NP, and NNT-1/BSF-3. In each case, the corresponding receptors employ a unique chain that either only serves as a ligand-binding chain (like IL-6Rα), or both as a ligand-binding and signal transducing chain. The latter group includes specific receptor chains for OSM (i.e. OSM-Rα) and LIF (i.e. LIF-Rβ), which both exhibit significant homology to gp130. Of note, OSM appears to straddle this and the following group (Sect. 5.2.2), because it binds to two receptors, one that consists of gp130 and OSM-Rα, and the other consisting of OSM-Rα and OSM-Rβ (also a gp130 homolog).

5.2.2 gp130 Receptor-like Subfamily

Although most members of the extended gp130 receptor family include a gp130 chain, several appear to function independent of gp130. As alluded to above, this includes one of the OSM receptors, as well as receptors for IL-31, G-CSF, and Leptin. Curiously, the IL-31 receptor consists of both a unique chain (IL-31Rα) and OSM-Rβ. The receptors for both G-CSF and Leptin function as homodimers (see also section on single-chain receptors). The G-CSF receptor is important for neutrophil maturation, whereas the Leptin receptor is important in weight homeostasis. Importantly however, like in the IL6-gp130 subfamily, each of these ligand–receptor pairs transduces pivotal signals through Stat3.

5.2.3 IL-12 Family of gp130-Related Receptors

This group includes three receptors that are distantly related to, but functionally distinct from the IL-6-gp130 family. They bind to and transduce signals for a unique family of related ligands, which include IL-12, IL-23, IL-27. IL-12, the founding and most fully characterized member of this family is composed of two disulfide-linked peptides of 35 kDa (i.e. p35) and 40 kDa (p40). p35 has homology to IL-6, whereas p40 has homology to the IL-6Rα chain (also IL-11Rα and CNTFRα). p40 can also associate with a distinct 19-kDa subunit (i.e. p19) to form a novel cytokine, IL-23. Another recently identified and more poorly characterized member of this family is IL-27, which consists of EBI3, an IL-12 p40 homolog (recently renamed IL-30), and p28, an IL-12 p35 homolog. Not unexpectedly, this family of related ligands binds to and transduces signals through a family of related receptors. IL-12 binds to a receptor consisting of IL-12Rβ1 and IL-12Rβ2. This receptor is known to direct the sequential activation of Tyk2 and Stat4, both of which have been shown to be important in the biological response to IL-12. The IL-23 receptor consists of the p40 binding IL-12β1 chain and a unique p19 binding chain, now known as IL-23R. Although, one chain of the IL-27 receptor has recently been identified as the orphan WSX-1 receptor chain, the

nature of the second chain and the STAT-dependent signal downstream from this receptor remains controversial. However, the important role these ligands play in regulating the maturation of naïve T cells indicates that their signaling pathways will be similar.

5.3 γC Receptor Family

This family of receptors exploits the common hematopoietic receptor gamma chain (i.e. γC) and plays an important role in directing development and activity of lymphocytes. These receptors can be divided into two functionally distinct subgroups, the IL-2 family and the IL-4 family.

5.3.1 IL-2 Receptor Family

The receptor for IL-2 is a founding member of the cytokine receptor family. The role this receptor and its ligand play in lymphocyte development has been studied extensively. Additional family members that also play an important role in lymphocyte regulation, and their corresponding ligands have been identified over the ensuing years. They include the receptors for IL-7, TSLP, IL-9, IL-15, and IL-21. Each receptor consists of a ligand specific chain (the α-chain) and the common gamma chain (γC). In most cases, α-chains are associated with Jak1 and serve to recruit and mediate the activation of Stat5. The γC, whose expression is limited to hematopoietic tissues, serves to bring Jak3 to the receptor complex. The critical role these latter two components play in lymphocyte function is underscored by the severe combined immunodeficiency (SCID) phenotype that develops when γC or Jak3 are defective. The IL-2 family exhibits a number of additional noteworthy features. First, two closely related members of this receptor family (i.e. the IL-2 and IL-15 receptors) employ a third receptor chain, the β-chain, which serves as the signal transducing chain. In the absence of the α-chain, this β-chain directs the activity of a low affinity IL-2 receptor. Also, the receptor for TSLP does not use γC, rather, it consists of the IL-7R α-chain and a unique TSLP receptor chain. Finally, as mentioned to earlier, each of these receptors, except perhaps the IL-21 receptor, rely on Stat5 to transduce important ligand specific signals.

5.3.2 IL-4 Receptor Family

The receptors for IL-4 and IL-13 consist of an overlapping set of chains and are functionally distinct in the IL-2 family. These receptors and their corresponding ligands have received considerable attention for the role they play in regulating humeral immunity and allergic response. IL-4 binds two distinct receptors. One is composed of the IL-4 receptor α-chain (IL-4Rα), which associates with Jak1, and γC chain, which is associated with Jak3. The second receptor is a heterodimer between IL-4Rα (associated with Jak1) and the IL-13 receptor α-chain (IL-13Rα), which associates with Tyk2. IL-13 binds and signals through the second of these two receptors (i.e. IL-4Rα and IL-13Rα). For both receptors, the IL-4Rα chain serves as the signaling chain, directing the recruitment and activation of Stat6. This, along with the ability to signal through IRS adaptor proteins, functionally distinguishes this family from the IL-2 family. While this small family has overlapping function, gene targeting studies have identified important differences. IL-4 plays a more important role in lymphocyte maturation, whereas IL-13 is more important in directing lymphocyte effector functions.

5.4
IL-3 Receptor Family

The IL-3 family includes receptors for IL-3, IL-5, and GM-CSF (granulocyte macrophage-colony-stimulating factor), all of which play important roles in the development of myeloid lineages. The IL-3 and GM-CSF receptors, and their corresponding ligands, play a more important general role, directing development of granulocytes, macrophages, and dendritic cells (DCs). IL-5 and its receptor appear more specialized, promoting the development of mast cells and eosinophils. Each receptor in this family is composed of two chains: a ligand specific α-chain (i.e. IL-3Rα, IL-5Rα, and GM-CSFRα); and a shared signal transducing β-chain (i.e. β_{common}). The α-chains associate with Jak1 and the β-chain associates with Jak2. The β-chain also serves to recruit and activate Stat5. In mice, however, the receptor for IL-3 is a bit more complex. There are two functionally homologous IL-3 receptor β-chains, β_{IL3} and β_{common}. Analogous to human system, β_{common} serves as the β-chain for GM-CSF and IL-5 receptors, as well as some IL-3 receptors. But the remaining IL-3 receptors employ β_{IL3}, providing redundancy in the ability to respond to IL-3.

5.5
Single-chain Receptor Family

The single-chain receptors are the final family of structurally and functionally related hematopoietin receptors. As their name suggests, these receptors consist of a single chain that both binds ligand and transduces signals. Members of this family include receptors for traditional hematopoietins, like Epo (erythropoietin) and Tpo (thrombopoietin), as well as the receptors for GH and PRL (prolactin). In each case, the ligand binds to and activates a receptor homodimer, culminating in the sequential activation of receptor associated Jak2 and then Stat5. It is worth noting that two receptors from the gp130-related family, the G-CSF and Leptin receptors also function as homodimers.

5.6
Noncytokine Receptors

A number of receptors that are not from the cytokine family also appear to direct a subset of signals through STAT activation. Well-known examples include several receptor tyrosine kinases (RTks), like the receptors for EGF (epidermal growth factor), PDGF (platelet-derived growth factor), FGF (fibroblast growth factor) and potentially Flt3. There are also numerous reports implicating STAT signaling in the response to G proteins and G-protein-coupled receptors. The best-characterized example is the receptor for Angiotensin II (i.e. AT_1). This should not be surprising, since some of the receptors implicated in STAT activation in *Dictyostelium* are G-protein-coupled receptors.

6
Regulation of JAK-STAT Signaling

Two characteristic features of the JAK-STAT signals are their rapid onset and transient nature. With respect to rapid onset, STATs are briskly activated at the receptor, whereupon they quickly translocate to the nucleus and bind DNA. Like for many other nuclear proteins, the process of nuclear import is dependent on a well-characterized set of proteins. Within a period of hours, however, the cytokine stimulated signals decay and the STATs are reexported back to the cytoplasm, resetting the cell for its next round of

stimulation. This signal decay entails both downregulation of the upstream signaling components (i.e. the receptors and JAKs), as well as terminating STAT transcription activity. This section will quickly review three mechanisms that are known to play an important role in STAT signal decay: dephosphorylation, active nuclear export, and inhibition by members of the SOCS family of regulators.

6.1 Phosphatases

It is not surprising that several phosphatases, which regulate all kinase-based signals, have been implicated in the control of STAT signal transduction. Genetic and biochemical approaches have been exploited to identify several cytoplasmic phosphatases that promote the decay of activated cytokine receptors and potentially JAKs. They include SHP-1, SHP-2 and potentially CD45. In contrast, studies directed at characterizing the decay of activated STATs have suggested that STATs are dephosphorylated in the nucleus prior to nuclear export. Although TC-PTP has been implicated as a STAT nuclear phosphatase, it is likely that other STAT phosphatases remain to be identified.

6.2 Nuclear Export

Within hours of having accumulated in the nucleus, STATs are dephosphorylated and reexported back to the cytoplasm, resetting the cell for the next round of stimulation. The process of nuclear export, just like that of nuclear import, is dependent on Ran-GTPases, but the sensitivity of export to leptomycin B has greatly facilitated the analysis of the export process. As expected, leptomycin B has been shown to effectively block poststimulation nuclear export. Unexpectedly however, treatment of resting cells with leptomycin B also promotes STAT nuclear accumulation. This observation highlights an intriguing aspect of STAT biology, not only are STATs rapidly translocated into and out of the nucleus in response to activation, but they also shuttle into and out of the nucleus in resting cells. Moreover, this process is independent of tyrosine phosphorylation! Thus, the predominately cytoplasmic localization of STATs in resting cells represents a steady state balance between nuclear import and export. While the reason for this continuous flux has not been elucidated, distinct nuclear export sequence (NES) elements, which direct either poststimulation or basal nuclear export, have been identified. The distribution of NES elements suggests that their activity (i.e. access to nuclear export machinery) is regulated by conformational changes STATs undergo as they become activated, bind DNA, and then become dephosphorylated.

6.3 The SOCS Family

The SOCS proteins were identified as a family of STAT target genes that directly antagonize STAT activation, establishing a classic "feedback loop." CIS-1 (cytokine-inducible SH2 containing protein), the founding member, was identified as an Epo induced, Stat5 dependent target gene that blocked subsequent Stat5 recruitment to the Epo receptor. SOCS-1 was identified next through several approaches. Comparison of the CIS-1 and SOCS-1 sequences identified a conserved SH2 domain and SOCS box, now recognized as characteristic features of this family. Subsequent gene targeting studies determined that SOCS-1 knockout mice die perinatally from a

profound IFN-γ – Stat1 dependent inflammatory process, underscoring a critical role for SOCS-1 regulating these potent cytokine. In contrast, SOCS-2 has been shown to antagonize the ability of GH to activate Stat5. Consistent with this, the SOCS-2 knockout mice develop gigantism. The final SOCS protein found to regulate cytokine activity is SOCS-3. (SOCS4-7 appear to regulate responses to different classes of stimuli.) Careful studies have determined that SOCS-3 is a Stat3 antagonist. Furthermore, its specificity for IL-6 dependent Stat3 activation appears to be the major determinant in distinguishing between the differing Stat3 dependent responses to IL-6 and IL-10. This level of specificity is achieved through a single SOCS-3 recruitment motif found in gp130 receptor chain (i.e. Y759; see Sect. 5.2.1), but not either of the IL-10 receptor chains. An analogous SOCS-2 recruitment motif has been identified in the GH receptor. Finally, the mechanism by which the SOCS proteins regulate STAT signaling appears to entail more than competition for receptor recruitment. A number of studies have determined that the conserved SOCS box promotes association with the elongin B/C complex, which in turn binds an E3-like ubiquitin ligase, cullin-2, to promote proteosome dependent degradation. Thus, SOCS proteins appear to direct the degradation of at least some of the components in the JAK-STAT signaling cascade. The existence of SOCS proteins underscores the need to counterregulate STAT based signals.

7 Concluding Comments

Studies over the last decade have provided an important insight into how members of the JAK and STAT family direct signal transduction for all four-helix-bundle cytokines (i.e. hematopoietins). While much remains to be learned, significant advances in the future are likely to entail the identification of additional molecules that serve to fine tune this signaling cascade, as well as provide an insight into how these signals integrate with other important signaling pathways. Consistent with this prediction, a number of regulatory molecules, whose role in STAT signaling are not yet fully elucidated, have been identified, including, members of the PIAS (protein inhibitors of activated STATs) family, StIP, Nmi, several phosphatases, and several serine kinases. Recent studies, which have highlighted the role covalent modifications play in the regulation of transcription factors, beget the question of whether STAT signaling will also be found to be regulation by acetylation, methylation, SUMOylation, or nitrosylation.

See also Cytokines: Interleukins; Molecular Mediators: Cytokines; Receptor Biochemistry; Viral Inhibitors and Immune Response Mediators: The Interferons.

Bibliography

Books and Reviews

Akira, S., Takeda, K. (2004) Toll-like receptor signalling, *Nat. Rev. Immunol.* **4**, 499–511.

Barton, G.M., Medzhitov, R. (2003) Linking Toll-like receptors to IFN-alpha/beta expression, *Nat. Immunol.* **4**, 432–433.

Dearolf, C.R. (1999) JAKs and STATs in invertebrate model organisms, *Cell. Mol. Life Sci.* **55**, 1578–1584.

Decker, T., Kovarik, P. (2000) Serine phosphorylation of STATs, *Oncogene* **19**, 2628–2637.

Decker, T., Kovarik, P., Meinke, A. (1997) GAS elements: a few nucleotides with a major impact on cytokine-induced gene expression, *J. Interferon Cytokine Res.* **17**, 121–134.

Ehret, G.B., Reichenbach, P., Schindler, U., Horvath, C.M., Fritz, S., Nabholz, M., Bucher, P. (2001) DNA binding specificity of different STAT proteins. Comparison of in vitro specificity with natural target sites, *J. Biol. Chem.* **276**, 6675–6688.

Fickenscher, H., Hor, S., Kupers, H., Knappe, A., Wittmann, S., Sticht, H. (2002) The interleukin-10 family of cytokines, *Trends Immunol.* **23**, 89–96.

Heinrich, P.C., Behrmann, I., Haan, S., Hermanns, H.M., Muller-Newen, G., Schaper, F. (2003) Principles of interleukin (IL)-6-type cytokine signalling and its regulation, *Biochem. J.* **374**, 1–20.

Horvath, C.M. (2000) STAT proteins and transcriptional responses to extracellular signals, *Trends Biochem. Sci.* **25**, 496–502.

Ihle, J.N. (2001) The Stat family in cytokine signaling, *Curr. Opin. Cell Biol.* **13**, 211–217.

Kile, B.T., Nicola, N.A., Alexander, W.S. (2001) Negative regulators of cytokine signaling, *Int. J. Hematol.* **73**, 292–298.

Kisseleva, T., Bhattacharya, S., Braunstein, J., Schindler, C.W. (2002) Signaling through the JAK/STAT pathway, recent advances and future challenges, *Gene* **285**, 1–24.

Kovanen, P.E., Leonard, W.J. (2004) Cytokines and immunodeficiency diseases: critical roles of the gamma(c)-dependent cytokines interleukins 2, 4, 7, 9, 15, and 21, and their signaling pathways, *Immunol. Rev.* **202**, 67–83.

Levy, D.E., Darnell, J.E. Jr. (2002) Stats: transcriptional control and biological impact, *Nat. Rev. Mol. Cell Biol.* **3**, 651–662.

Miyoshi, K., Cui, Y., Riedlinger, G., Lehoczky, J., Zon, L., Oka, T., Dewar, K., Hennighausen, L. (2001) Structure of the mouse stat 3/5 locus: evolution from drosophila to zebrafish to mouse, *Genomics* **71**, 150–155.

O'Shea, J.J., Gadina, M., Schreiber, R.D. (2002). Cytokine signaling in 2002: new surprises in the Jak/Stat pathway, *Cell* **109**, S123–S131.

Pestka, S., Krause, C.D., Sarkar, D., Walter, M.R., Shi, Y., Fisher, P.B. (2004a) Interleukin-10 and related cytokines and receptors, *Annu. Rev. Immunol.* **22**, 929–979.

Pestka, S., Krause, C.D., Walter, M.R. (2004b) Interferons, interferon-like cytokines, and their receptors, *Immunol. Rev.* **202**, 8–32.

Shuai, K. (2000) Modulation of STAT signaling by STAT-interacting proteins, *Oncogene* **19**, 2638–2644.

Zeidler, M.P., Bach, E.A., Perrimon, N. (2000) The roles of the Drosophila JAK/STAT pathway, *Oncogene* **19**, 2598–2606.

Primary Literature

Akira, S., Nishio, Y., Inoue, M., Wang, X.-J., Wei, S., Matsusaka, T., Yoshida, K., Sudo, T., Naruto, M., Kishimoto, T. (1994) Molecular cloning of APRF, a novel IFN-stimulated gene factor 3 p91-related transcription factor involved in the gp130-mediated signaling pathway, *Cell* **77**, 63–71.

Alexander, W.S., Starr, R., Fenner, J.E., Scott, C.L., Handman, E., Sprigg, N.S., Corbin, J.E., Cornish, A.L., Darwiche, R., Owczarek, C.M. et al. (1999) SOCS1 is a critical inhibitor of interferon γ signaling and prevents the potentially fatal neonatal actions of this cytokine, *Cell* **98**, 597–608.

Argetsinger, L.S., Carter-Su, C. (1996) Mechanism of signaling by growth hormone receptor, *Physiol. Rev.* **76**, 1089–1107.

Azam, M., Erdjument-Bromage, H., Kreider, B.L., Xia, M., Quelle, F., Basu, R., Saris, C., Tempst, P., Ihle, J.N., Schindler, C. (1995) Interleukin-3 signals through multiple isoforms of Stat5, *EMBO J.* **14**, 1402–1411.

Becker, S., Groner, B., Müller, C.W. (1998) Three-dimensional structure of the Stat3β homodimer bound to DNA, *Nature* **394**, 145–151.

Begitt, A., Meyer, T., van Rossum, M., Vinkemeier, U. (2000) Nucleocytoplasmic translocation of Stat1 is regulated by a leucine-rich export signal in the coiled-coil domain, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10418–10423.

Bhattacharya, S., Schindler, C. (2003) Regulation of Stat3 nuclear export, *J. Clin. Invest.* **111**, 553–559.

Braunstein, J., Brutsaert, S., Olson, R., Schindler, C. (2003) STATs dimerize in the absence of phosphorylation, *J. Biol. Chem.* **278**, 34133–34140.

Bromberg, J.F., Wrzeszczynska, M.H., Devgan, G., Zhao, Y., Pestell, R.G., Albanese, C., Darnell, J.E. (1999) Stat3 as an Oncogene, *Cell* **98**, 295–303.

Chen, X., Vinkemeier, U., Zhao, Y., Jeruzalmi, D., Darnell, J.E. Jr., Kuriyan, J. (1998) Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA, *Cell* **93**, 827–839.

Chen, X., Bhandari, R., Vinkemeier, U., Van Den Akker, F., Darnell, J.E. Jr., Kuriyan, J. (2003) A reinterpretation of the dimerization interface of the N-terminal domains of STATs, *Protein Sci.* **12**, 361–365.

Copeland, N.G., Gilbert, D.J., Schindler, C., Zhong, Z., Wen, Z., Darnell, J.E. Jr., Mui, A.L., Miyajima, A., Quelle, F.W., Ihle, J.N., et al. (1995) Distribution of the mammalian Stat gene family in mouse chromosomes, *Genomics* **29**, 225–228.

Croker, B.A., Krebs, D.L., Zhang, J.G., Wormald, S., Willson, T.A., Stanley, E.G., Robb, L., Greenhalgh, C.J., Forster, I., Clausen, B.E., et al. (2003) SOCS3 negatively regulates IL-6 signaling in vivo, *Nat. Immunol.* **4**, 540–545.

Durbin, J.E., Hackenmiller, R., Simon, M.C., Levy, D.E. (1996) Targeted disruption of the mouse Stat1 gene results in compromised innate immunity to viral disease, *Cell* **84**, 443–450.

Gupta, S., Yan, H., Wong, L.H., Ralph, S., Krolewski, J., Schindler, C. (1996) The SH2 domains of Stat1 and Stat2 mediate multiple interactions in the transduction of IFN-alpha signals, *EMBO J.* **15**, 1075–1084.

Heim, M.H., Kerr, I.M., Stark, G.R., Darnell, J.E. (1995) Contribution of STAT SH2 groups to specific interferon signaling by the JAK-STAT pathway, *Science* **267**, 1347–1349.

Hou, J., Schindler, U., Henzel, W.J., Ho, T.C., Brasseur, M., McKnight, S.L. (1994) An Interleukin-4 induced transcription factor: IL-4 Stat, *Science* **265**, 1701–1705.

Kaplan, M.H., Schindler, U., Smiley, S.T., Grusby, M.J. (1996a) Stat6 is required for mediating responses to IL-4 and for the development of Th2 cells, *Immunity* **4**, 313–319.

Kaplan, M.H., Sun, Y.-L., Hoey, T., Grusby, M.J. (1996b) Impaired IL-12 responses and enhanced development of Th2 cells in Stat4-deficient mice, *Nature* **382**, 174–177.

Karaghiosoff, M., Neubauer, H., Lassnig, C., Kovarik, P., Schindler, H., Pircher, H., McCoy, B., Bogdan, C., Decker, T., Brem, G., et al. (2000) Partial impairment of cytokine responses in Tyk2-deficient mice, *Immunity* **13**, 549–560.

Kawata, T., Shevchenko, A., Fukuzawa, M., Jermyn, K.A., Totty, N.F., Zhukovskaya, N.V., Sterling, A.E., Mann, M., Williams, J.G. (1997) SH2 signaling in a lower eukaryote: a STAT protein that regulates stalk cell differentiation in dictyostelium, *Cell* **89**, 909–916.

Kieslinger, M., Woldman, I., Moriggl, R., Hofmann, J., Marine, J.-C., Ihle, J.N., Beug, H., Decker, T. (2000) Anti-apoptotic activity of Stat5 required during terminal stages of myeloid differentiation, *Genes Dev.* **14**, 232–244.

Klingmuller, U., Lorenz, U., Cantley, L.C., Neel, B.G., Lodish, H.F. (1995) Specific recruitment of SH-PTP1 to the erythropoietin receptor causes inactivation of JAK2 and termination of proliferative signals, *Cell* **80**, 729–738.

Marie, I., Durbin, J.E., Levy, D.E. (1998) Differential viral induction of distinct alpha-interferon genes by positive feedback through interferon regulatory factor-7, *EMBO J.* **17**, 6660–6669.

Marine, J.-C., Topham, D.J., McKay, C., Wang, D., Parganas, E., Stravopodis, D., Yoshimura, A., Ihle, J.N. (1999a) SOCS1 deficiency causes a lymphocyte-dependent perinatal lethality, *Cell* **98**, 609–616.

Marine, J.C., McKay, C., Wang, D., Topham, D.J., Parganas, E., Nakajima, H., Pendeville, H., Yasukawa, H., Sasaki, A., Yoshimura, A., Ihle, J.N. (1999b) SOCS3 is essential in the regulation of fetal liver erythropoiesis, *Cell* **98**, 617–627.

Maritano, D., Sugrue, M.L., Tininini, S., Dewilde, S., Strobl, B., Fu, X., Murray-Tait, V., Chiarle, R., Poli, V. (2004) The STAT3 isoforms alpha and beta have unique and specific functions, *Nat. Immunol.* **5**, 401–409.

McBride, K.M., Banninger, G., McDonald, C., Reich, N.C. (2002) Regulated nuclear import of the STAT1 transcription factor by direct binding of importin-alpha, *EMBO J.* **21**, 1754–1763.

Meraz, M.A., White, J.M., Sheehan, K.C., Bach, E.A., Rodig, S.J., Dighe, A.S., Kaplan, D.H., Riley, J.K., Greenlund, A.C., Campbell, D., et al. (1996) Targeted disruption of the Stat1 gene in mice reveals unexpected physiological specificity in the JAK-STAT pathway, *Cell* **84**, 431–442.

Metcalf, D., Greenhalgh, C.J., Viney, E., Willson, T.A., Starr, R., Nicola, N.A., Hilton, D.J., Alexander, W.S. (2000) Gigantism in mice lacking suppressor of cytokine signalling-2, *Nature* **405**, 1069–1073.

Moriggl, R., Sexl, V., Piekorz, R., Topham, D., Ihle, J.N. (1999) Stat5 activation is uniquely associated with cytokine signaling in peripheral T cells, *Immunity* **11**, 225–230.

Neubauer, H., Cumano, A., Mueller, M., Wu, H., Huffstadt, U., Pfeffer, K. (1998) Jak2 deficiency defines an essential developmental checkpoint in definitive hematopoiesis, *Cell* **93**, 397–409.

Nosaka, T., van Deursen, J.M., Tripp, R.A., Thierfelder, W.E., Witthuhn, B.A., McMickle, A.P., Doherty, P.C., Grosveld, G.C., Ihle, J.N. (1995) Defective lymphoid development in mice lacking Jak3, *Science* **270**, 800–802.

Ohtani, T., Ishihara, K., Atsumi, T., Nishida, K., Kaneko, Y., Miyata, T., Itoh, S., Narimatsu, M., Maeda, H., Fukada, T., et al. (2000) Dissection of signaling cascades through gp130 in vivo: reciprocal roles for STAT3- and SHP2-mediated signals in immune responses, *Immunity* **12**, 95–105.

Parganas, E., Wang, D., Stravopodis, D., Topham, D., Marine, J.-C., Teglund, S., Vanin, E.F., Bodner, S., Colamonici, O.R., van Deursen, J.M., et al. (1998) Jak2 is essential for signaling through a variety of cytokine receptors, *Cell* **93**, 385–395.

Park, C., Li, S., Cha, E., Schindler, C. (2000) Immune response in Stat2 knockout mice, *Immunity* **13**, 795–804.

Park, S.Y., Saijo, K., Takahashi, T., Osawa, M., Arase, H., Hirayama, N., Miyake, K., Nakauchi, H., Shirasawa, T., Saito, T. (1995) Developmental defects of lymphoid cells in Jak3 kinase-deficient mice, *Immunity* **3**, 771–782.

Remy, I., Wilson, I.A., Michnick, S.W. (1999) Erythropoietin receptor activation by a ligand-induced conformation change, *Science* **283**, 990–993.

Rodig, S.J., Meraz, M.A., White, J.M., Lampe, P.A., Riley, J.K., Arthur, C.D., King, K.L., Sheehan, K.C., Yin, L., Pennica, D., et al. (1998) Disruption of the Jak1 gene demonstrates obligatory and nonredundant roles of the Jaks in cytokine-induced biologic responses, *Cell* **93**, 373–383.

Schindler, C., Fu, X.Y., Improta, T., Aebersold, R., Darnell, J.E. Jr. (1992a) Proteins of transcription factor ISGF-3: one gene encodes the 91-and 84-kDa ISGF-3 proteins that are activated by interferon alpha, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7836–7839.

Schindler, C., Shuai, K., Prezioso, V., Darnell, J.E. (1992b) Interferon-dependent tyrosine phosphorylation of a latent cytoplasmic transcription factor, *Science* **257**, 809–813.

Seidel, H.M., Milocco, L.H., Lamb, P., Darnell, J.E., Stein, R.B., Rosen, J.B. (1995) Spacing of palindromic half sites as a determinant of selective STAT (signal transducers and activators of transcription) DNA binding and transcriptional activity, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3041–3045.

Shimoda, K., van Deursen, J., Sangster, M.Y., Sarawar, S.R., Carson, R.T., Tripp, R.A., Chu, C., Quelle, F.W., Nosaka, T., Vignali, D.A., et al. (1996) Lack of IL-4 induce Th2 response in IgE class switching mice with disrupted Stat6 gene, *Nature* **380**, 630–633.

Shimoda, K., Kato, K., Aoki, K., Matsuda, T., Miyamoto, A., Shibamori, M., Yamashita, M., Numata, A., Takase, K., Kobayashi, S., et al. (2000) Tyk2 plays a restricted role in IFN alpha signaling, although it is required for IL-12-mediated T cell function, *Immunity* **13**, 561–571.

Shuai, K., Schindler, C., Prezioso, V., Darnell, J.E. (1992) Activation of transcription by IFN-γ: tyrosine phosphorylation of a 91-kDa DNA binding protein, *Science* **258**, 1808–1812.

Shuai, K., Stark, G.R., Kerr, I.M., Darnell, J.E. (1993) A single phosphotyrosine residue of Stat91 required for gene activation by interferon-γ, *Science* **261**, 1744–1743.

Silvennoinen, O., Schindler, C., Schlessinger, J., Levy, D.E. (1993) Ras-independent signal transduction in response to growth factors and cytokines by tyrosine phosphorylation of a common transcription factor, *Science* **261**, 1737–1739.

Stahl, N., Boulton, T.G., Farruggella, T., Ip, N.Y., Davis, S., Witthuhn, B.A., Quelle, F.W., Silvennoinen, O., Barbieri, G., Pellegrini, S., et al. (1994) Association and activation of Jak-Tyk kinases by CNTF-LIF-OSM-IL-6 beta receptor components, *Science* **263**, 92–95.

Starr, R., Willson, T.A., Viney, E.M., Murray, L.J., Rayner, J.R., Jenkins, B.J., Gonda, T.J., Alexander, W.S., Metcalf, D., Nicola, N.A., Hilton, D.J. (1997) A family of cytokine-inducible inhibitors of signaling, *Nature* **387**, 917–921.

Takeda, K., Clausen, B.E., Kaisho, T., Tsujimura, T., Terada, N., Forster, I., Akira, S. (1999) Enhanced Th1 activity and development of chronic enterocolitis in mice devoid of Stat3 in macrophages and neutrophils, *Immunity* **10**, 39–49.

Takeda, K., Noguchi, K., Shi, W., Tanaka, T., Matsumoto, M., Yoshida, N., Kishimoto, T., Akira, S. (1997) Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3801–3804.

Takeda, K., Tanaka, T., Shi, W., Matsumoto, M., Minami, M., Kashiwamura, S., Nakanishi, K., Yoshida, N., Kishimoto, T., Akira, S. (1996) Essential role of Stat6 in IL-4 signaling, *Nature* **380**, 627–630.

Teglund, S., McKay, C., Schuetz, E., Van Deursen, J.M., Stravopodis, D., Wang, D., Brown, M., Bodner, S., Grosveld, G., Ihle, J.N. (1998) Stat5a and Stat5b proteins have essential roles and nonessential, or redundant, roles in cytokine responses, *Cell* **93**, 841–850.

Thierfelder, W.E., van Deursen, J.M., Yamamoto, K., Tripp, R.A., Sarawar, S.R., Carson, R.T., Sangster, M.Y., Vignali, D.A., Doherty, P.C., Grosveld, G., Ihle, J.N. (1996) Requirement for Stat4 in interleukin-12 mediated response of natural killer and T cells, *Nature* **382**, 171–174.

Thomis, D.C., Gurniak, C.B., Tivol, E., Sharpe, A.H., Berg, L.J. (1995) Defects in B lymphocyte maturation and T lymphocyte activation in mice lacking Jak3, *Science* **270**, 794–797.

Varinou, L., Ramsauer, K., Karaghiosoff, M., Kolbe, T., Pfeffer, K., Muller, M., Decker, T. (2003) Phosphorylation of the Stat1 transactivation domain is required for full-fledged IFN-gamma-dependent innate immunity, *Immunity* **19**, 793–802.

Velazquez, L., Fellous, M., Stark, G.R., Pellegrini, S. (1992) A protein tyrosine kinase in the Interferon α/β signaling pathway, *Cell* **70**, 313–322.

Vinkemeier, U., Moarefi, I., Darnell, J.E. Jr., Kuriyan, J. (1998) Structure of the amino-terminal protein interaction domain of STAT-4, *Science* **279**, 1048–1052.

Wen, Z., Zhong, Z., Darnell, J.E. (1995) Maximal activation of transcription by Stat1 and Stat3 requires both tyrosine and serine phosphorylation, *Cell* **82**, 241–250.

Yasukawa, H., Ohishi, M., Mori, H., Murakami, M., Chinen, T., Aki, D., Hanada, T., Takeda, K., Akira, S., Hoshijima, M., et al. (2003) IL-6 induces an anti-inflammatory response in the absence of SOCS3 in macrophages, *Nat. Immunol.* **4**, 551–556.

Single Nucleotide Polymorphism Blocks and Haplotypes: Human MHC Block Diversity

E.J. Yunis[1,2], J. Zuñiga[1,2,3], C.E. Larsen[2,4], M. Fernández-Viña[5], J. Granados[6], Z.L. Awdeh[4], and C.A. Alper[2,4]
[1] *Dana-Farber Cancer Institute, Boston, MA, USA*
[2] *Harvard Medical School, Boston, MA, USA*
[3] *Instituto Nacional de Enfermedades Respiratorias, Mexico City, Mexico*
[4] *The CBR Institute for Biomedical Research, Boston, MA, USA*
[5] *University of Texas, Houston, TX, USA*
[6] *Instituto Nacional de Ciencias Médicas y de la Nutrición Salvador Zubirán, Mexico City, Mexico*

1 Introduction 194

2 The Major Histocompatibility Complex (MHC) 195
2.1 MHC Blocks 198
2.2 HLA Blocks 198
2.2.1 HLA-Cw/B Block 198
2.2.2 HLA-DR/DQ Block 199
2.3 Common HLA Haplotypes Defined Serologically and Deduced HLA-Cw/B, HLA-DRB1/DQB1 Common Haplotypes and their Frequencies 201
2.4 The Aggregate Block Frequency (ABF) Measures Genetic Diversity 201
2.5 Class III Region Blocks 202
2.5.1 TNF Block 202
2.5.2 Complotype Block 202

3 Conserved Extended Haplotypes (CEHs) 202
3.1 Extension of the HLA-Cw/B Block to HLA-A and of the HLA-DR/DQ Block to HLA-DPB1 203
3.1.1 Frequency Distribution of MHC Blocks, CEH Fragments, and CEHs 203
3.1.2 The Global LD Measurement Related to the Size of the Database 203
3.1.3 Association of CEHs with Disease 203

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

4 Mapping of MHC Disease Genes: New Model for Future Studies Combining Genotyped Individuals with the Use of High-throughput SNPs 206

5 Paralogous Chromosomal Regions 206

6 Discussion 207

7 Future Studies 210

8 Summary 211

Acknowledgment 212

Bibliography 212
Book and Reviews 212
Primary Literature 213

Keywords

Alleles
Variants of genes, alternative forms of a gene that occur at a given location of a chromosome.

Conserved Extended Haplotypes
Relatively large stretches (1.5 Mb or more) of conserved DNA sequence, in which essentially all allelic markers within the genetic region are identical and shared among a relatively large number of unrelated individuals (typically, although not definitively, greater than or equal to 0.5% of a given population). Operationally, human MHC CEHs are identified by identical markers on a single chromosome at *HLA-B*, complotype and *HLA-DRB1*.

Disease Association
Genetic markers significantly associated with diseases.

Fragments of Conserved Extended Haplotypes
Genetic regions with different sizes of DNA blocks.

genes
Self-reproducing hereditary characters involved in production of proteins.

Genetic Recombination and Hotspots
Points of crossing-over between chromosomes during meiosis determined by family studies. Meiosis is the creation of sex cells by replication of chromosomes followed by

cell division and production of gametes; sperms and eggs. Hotspots refer to hypothetical chromosomal regions with higher rates of recombination.

Genetic Stratification and Population Admixture
Studies of DNA, RNA, or proteins that characterize different nationalities or races.

HLA Polymorphism
Allelic or specificity variation within specific genes or loci of the human MHC.

Human Genetic Diversity
Genetic markers that differ among individuals.

Linkage
Alleles on the same chromosome close enough to be inherited together.

Linkage Disequilibrium
Nonrandom association of inherited alleles in a population.

Major Histocompatibility Complex (MHC)
Polymorphic genes located in the human chromosome 6 involved in immune responses and in transplantation.

Population Genetics
Genetic differences among individuals of different nationalities or racial groups.

We describe genetically fixed segments of DNA within the major histocompatibility complex (MHC) extending to 3.2 Mb of DNA from *HLA-A* to *HLA-DPB1*. These are variable sized DNA fragments that vary in frequency in different ethnicities or races. Within the region, one 1.5-Mb block is relatively frequent, the DNA conserved extended haplotypes (CEHs). These span from *HLA-Cw*, *B* to *HLA-DRB1*, *DQB1* and include the polymorphisms of complement genes (complotypes) and TNF (tumor necrosis factor genes). This segment is informative for mapping disease susceptibility, immune responses, and allotransplantation matching. Other chromosomal regions (paralogous) have conserved genes, including those genetically related to those of the MHC, which also form DNA blocks that could be involved in immune functions and show disease association. DNA blocks can be used to measure human diversity. For example, the aggregate frequency of DNA blocks (ABF), determine the degree of genetic diversity in different populations; Africans, Asians, and Hispanics have higher genetic diversity than Caucasian Americans. The use of single nucleotide polymorphisms (SNPs) to determine DNA blocks (by measurement of linkage disequilibrium (LD)) is limited when ignoring the well-documented variability of frequency and size of DNA blocks among different populations.

1 Introduction

Recent interest in mapping human disease genes using single nucleotide polymorphisms (SNPs) has prompted us to review the structure of the human major histocompatibility complex (MHC) in order to discuss the limitations of such methods. SNPs occur every 200 to 2000 bp of human genomic DNA, and groups of relatively nearby SNPs have been analyzed as haplotypes on the basis of the measurement of linkage disequilibrium (LD). Genetic variation is nonrandomly distributed in the genome, and SNP density and distribution varies. The haplotypes deduced are said to be islands of nucleotide segments that are stable and form blocks. Claims are also made that meiotic crossing-over occurred between these blocks at what are called "hotspots" with the assumption that historic recombinations between the blocks explain and define operationally such blocks. Thus defined, the sizes of such blocks, using LD measurements, are between 5 to approximately 200 kb in size or more than 800 kb. A major drawback of all these reports of SNP blocks is that they disregard the well-documented existence and analysis of DNA blocks in the MHC, which vary in size among individuals and vary in frequency in different ethnicities.

We summarized in two previous reports some of the structural elements and polymorphic character of the human MHC regions between *HLA-A* and *HLA-B*, and between *HLA-B* and *HLA-DRB1*. These two regions have comparable sizes: 1460 kb from *HLA-A* to *HLA-B* and 1375 kb from *HLA-B* to *HLA-DRB1*. We also described the existence of small blocks and other relatively fixed genetic fragments in the human MHC: A/Cw = 1365 kb, Cw/B = 95 kb, TNF region ≥ 7 kb, complotype = 120 kb, and DR/DQ = 150 kb. We used the known frequencies of the latter four different small blocks to deduce the frequencies of common MHC haplotypes with sizes up to 3.4 Mb. We described a model using those four blocks for studying the population-based structure of the MHC with 16 potential kinds of block combinations that are either fragments of or individual blocks of conserved extended haplotypes (CEHs) with a total length of 1470 kb. This segment of the MHC (from *HLA-Cw* to *HLA-DQB1*) has been studied more extensively than the segment between *HLA-A* and *HLA-B* despite the fact that the latter was discovered much earlier.

MHC haplotype blocks and the larger CEHs are usually inherited intact as a unit, and the allele frequency distribution of particular MHC locus combinations in individuals is nonrandom. Specific alleles at various MHC loci tend to occur together on the same chromosome much more or much less frequently than the product of their individual frequencies. The LD between *HLA-A* and *HLA-B* and between *HLA-B* and *HLA-DRB1* has been well determined in some populations. By contrast, recent studies describing "blocks" of conserved DNA sequence (15–150 kb) within the human genome separated by sites of recombination (deduced) based on LD analysis of SNPs, suggest the existence of uniform lengths of conserved DNA sequences. Some investigators believe that each of these short blocks can be studied with high-throughput screening without the need for doing pedigree analysis. This, however, has questionable utility in the MHC. Of interest is a related SNP approach, analyzing a larger stretch of DNA (3.5 Mb) within the MHC, which was

reported recently, with the finding of what were called *surrogate haplotypes* arbitrarily anchored at *HLA-B*. Unfortunately, the haplotypes found did not always correlate with well-defined common haplotypes and CEHs found by pedigree analysis.

The small blocks of the MHC have different sizes and their variants differ in frequency in populations where they behave as fixed genetic units. These genetic units have been used as markers of human diversity, of ethnicity and/or nationality, of immune response, and of disease risk. They can also be used for genetic stratification studies to confirm the similarity or degree of admixture of populations. The HLA-Cw/B, complotype, and DR/DQ small blocks and the larger CEHs or common MHC haplotypes are important, because some of their variants are ethnic specific, while others are shared between different ethnicities. By contrast, SNP analysis in its infancy has rarely been used to distinguish individual variations in and to identify the population-specific distribution of the sizes of conserved DNA sequence longer than that of the small blocks. Furthermore, little effort has been made to understand the discrepancies between measurements of SNP LD as compared with methods of direct counting of different sizes of blocks in the MHC. Our view is that a thorough understanding of the relatively well-characterized MHC blocks in different human populations should serve investigators to better understand the sizes of DNA blocks throughout the genome.

2 The Major Histocompatibility Complex (MHC)

The human MHC is located within chromosomal region 6p21.3, spans at least 3.4 Mb of DNA and contains many genes and pseudogenes. In addition to the classical HLA genes, the MHC contains more than 140 other functional genes, many of which have immune function. The MHC genes are divided into three regions (class I, II, and III). The class I region is located at the telomeric end of the MHC, while the class II is at the centromeric end, with the class III region located between the class I and class II regions (Fig. 1). HLA allelic variants were first studied using serological reagents detecting human HLA

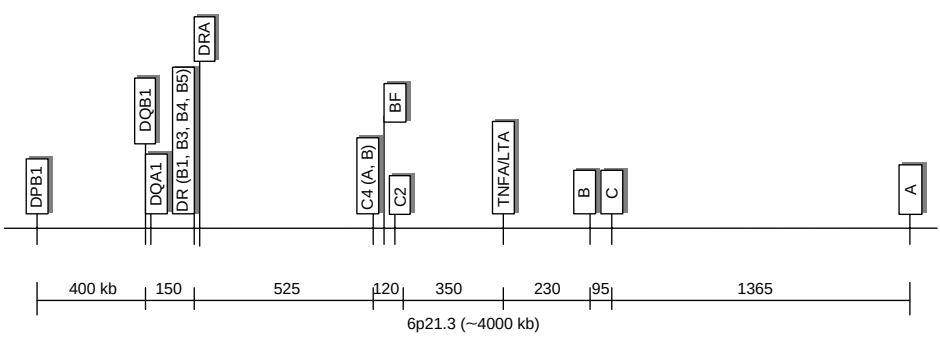

Fig. 1 Map of the human major histocompatibility complex (MHC) in 6p21.3. The physical distances between specific loci are shown. Distances are approximate and vary at many locations in different haplotypes as a result of various DNA insertions, deletions, and gene duplications. Modified from Yunis, E.J. et al. (2003) *Tissue Antigens*.

specificities, but HLA alleles can only be determined with precision using DNA methods. Genetic variants (alleles) at more than one locus inherited together on a parental chromosome are referred to as *haplotypes*.

MHC haplotypes (a term derived from haploid cells) found in an individual can only be defined by pedigree analysis in family studies, but several studies of large populations of unrelated subjects have been published that predicted frequencies of haplotypes calculated by statistical analysis. The original description of an HLA haplotype in 1967 using genotyped data that was obtained by pedigree analysis explained the coinheritance of alleles of two closely linked loci in several unrelated individuals. The concept of LD refers to the nonrandom association between alleles (usually at relatively nearby genetic loci). Therefore, this analytical method was quickly exploited to analyze preferential association between alleles at two or more loci based solely on phenotypic data (i.e. HLA typing of unrelated individuals, in which haplotype phase is not accurately known) has been described for the HLA region by several authors. LD has been an important tool since its early description for the genetic analysis of the human MHC, but the choice of methods for defining LD can be problematic. LD is defined as the difference between the predicted and observed haplotype frequency of specific alleles at two or more loci.

The alleles A and B at two randomly associated loci having frequencies $f(A)$ and $f(B)$ should occur together on the same chromosome with a haplotype frequency of $f(AB) = f(A) \times f(B)$. If this condition is not met, the alleles are said to be nonrandomly associated and the departure from prediction indicates the degree to which the alleles are in LD. The pattern of LD in the human genome varies between genomic regions and within a given genomic region between different ethnic groups. LD analysis has played an important role in both gene mapping and as one approach to identifying susceptibility markers in complex diseases. The extent of LD is given by delta $(D) = f(AB) - (f(A) \times f(B))$, where larger delta values indicate greater LD. Thus, the LD of a two-locus haplotype, $AiBj$ will be: $LD\ \ (AiBj) = HF(AiBj) - (f(Ai) \times f(Bj))$, where HF is the haplotype frequency and $f(Ai)$ and $f(Bj)$ are the frequencies of the Ai and Bj alleles. The standard delta value (D) is related to the frequency of alleles ($f(A)$ and $f(B)$). The allele frequency effect can be partially compensated by calculating a normalized LD value (D') that reflects the relative LD irrespective of allele frequencies. D' is calculated as: $D' = D/Dmax$, where $Dmax$ is the maximum D value possible. The definition of "strong" LD varies by the specific type of locus analysis under study. The highly polymorphic classical MHC haplotypes (and even their small blocks) containing alleles in strong LD generally have a $D' > 0.10$. For statistical tests of the significance, chi-square values are calculated using 2×2 contingency tables converted to p values.

LD determinations between *HLA-A* and *HLA-B* and between *HLA-B* and *HLA-DRB1* have been well described. Despite the opportunity for genetic recombination within the MHC seen in human sperm, the effect of meiotic recombination on the population distribution of MHC haplotypes is still uncertain. For example, an apparent "hotspot" of genetic recombination was found in human sperm within the class III region of the MHC near *LTA*, whereas a lower than expected recombination rate

was seen between *HLA-A* and *HLA-B*. This would suggest less relative fixity between *HLA-Cw* and *HLA-DRB1* as compared with the class I region. Nevertheless, this recombination site has not been confirmed in family studies. In the absence of both consistent family study confirmation and studies of the effect of sperm recombination on fertilization success and on pregnancy outcome, sperm recombination studies should not be generalized to describe the effect of recombination on haplotype frequencies in the population. Furthermore, the existence of CEHs, found in family studies, within the region between *HLA-Cw* and *HLA-DRB1* (often without a frequent *HLA-A* allele) and the comparable LD between *HLA-A* and *HLA-B* alleles as compared with that between *HLA-B* and *HLA-DRB1* alleles suggests that the genetic fixity of the two regions is similar.

We believe that it is necessary to study more intensively the genetic region between *HLA-Cw* and *HLA-A* to define new class I blocks and to analyze the possibility that the hotspots or regions of historical recombination represent small block boundaries. Recently, other recombination hotspots have been described. However, some of the described hotspots need to be confirmed with finer methods to establish their precise position. The physical distances for the two regions, HLA-A to HLA-Cw/B and HLA-Cw/B to HLA-DR/DQ are nearly identical. Pedigree analyses by family study have shown that both regions have approximately 1% recombination frequencies. If hotspots of genetic recombination between *HLA-B* and *HLA-DRB1* had a significant effect on the population frequencies of specific haplotypes, one would not have expected that the degree of nonrandom association would be higher in the latter. The full extent of LD within the MHC was further demonstrated with the elucidation of genetic polymorphism in the class III region intermediate between *HLA-B* and *HLA-DRB1*. Class III region genes include the complement genes *C2*, *C4A*, *C4B*, and *BF* as well as the cytokine genes *TNF*, *LTA*, and *LTB*. The haplotypes formed by complement genes in MHC class III region are called *complotypes* [K]. There are 14 relatively high-frequency complotypes in Caucasians that occur with a combined frequency of 0.96. MHC haplotypes defined by pedigree analysis containing essentially fixed DNA sequence across class I, II, and III loci found at relatively high frequency in specific populations are referred to as *CEHs* or *ancestral haplotypes* (AHs).

The presence in the population of CEHs, each defined by pedigree analysis, provides evidence for the DNA fixity in specific haplotypes of unrelated individuals. Two recent studies using LD analysis failed to find a significant amount of nonrandom association within the *HLA-Cw* to *HLA-DRB1* region other than in small blocks, although a third study found such associations. The former results are in apparent conflict with those reported by us and by others. We are not surprised that statistical analyses of a relatively few haplotypes (or unrelated individuals) fail to detect more than a few nonrandom associations of alleles of the MHC.

Reports describing the existence of "blocks" of conserved DNA sequence in the range of 5 to 150 kb within the human genome, separated by sites of high recombination activity, are based solely on LD analysis applied to SNP data. Most of these reports have concluded that such blocks represent relatively uniform lengths of conserved DNA sequences maintained throughout the human population as haplotypes. Many investigators now believe that such short blocks can be studied

with high-throughput screening without pedigree analysis. Although these techniques have been used recently to detect some of the previously demonstrated associations in the already well-characterized MHC genes and variants, their current utility, at least within the MHC, is uncertain. For example, LD analysis of SNP typing results, from unrelated Caucasian donors, was used recently across a much longer stretch of DNA (3.5 Mb) within the MHC to study "surrogate" *HLA-B* haplotypes. Genetic fixity within the MHC is so great that even LD measurements, taken over long distances of computer-constructed surrogate haplotypes (i.e. not defined by pedigree analysis) predicted to varying degrees the existence of specific common MHC haplotypes, although the haplotypes predicted often did not conform to high-frequency haplotypes seen in larger population studies based on direct haplotype counting through pedigree analysis. Another report, using pedigree-defined SNP-typed MHC haplotypes to construct an "integrated haplotype map" of the MHC, claimed that genetic fixity is higher within the MHC than throughout the rest of the human genome. However, those results did not achieve statistical significance, and the report only identified one common haplotype (although shorter) found in larger haplotype studies. SNP haplotype analysis, in its infancy, clearly still has much to prove before it might be considered a useful method for providing new insights into structure–function relationships, at least within the already well-characterized MHC.

2.1 MHC Blocks

A block size range of 5 to 150 kb in the human MHC have been well defined. These blocks are different from those recently reported, are never characterized solely by SNP typing, are highly polymorphic, and are diverse in different ethnic groups. Specific MHC blocks with specific alleles of one locus are often "haplospecific" for particular CEHs. Frequently, one allele at one MHC locus is highly correlated with an entire CEH. In other cases, a specific group of alleles within the same MHC block or in separate blocks is required to have a high correlation with a specific CEH. However, definitive characterization of haplotypes requires the analysis of families. Some haplotypes are not CEHs but could contain fragments or small blocks from specific CEHs. The best determination of haplotypes would require family studies.

2.2 HLA Blocks

In previous reports, we have summarized the two best characterized blocks of HLA loci, but we will emphasize in the following that the frequency distribution of the inheritance of the two blocks together has not been done in a critical manner. Instead, it has been possible to deduce the resulting haplotypes for the purpose of comparing frequencies of blocks and haplotypes in several ethnicities.

2.2.1 HLA-Cw/B Block

The HLA-Cw/B block is a relatively small segment of DNA (95 kb in length). Figure 2 shows the most frequent HLA-Cw/B block associations in four different ethnic groups of Americans, which, as reported before, are found in strong ($D' > 0.10$) and significant ($p > 0.0005$) LD. African-Americans, Asian-Americans, and Caucasian Americans share several high-frequency small blocks. Also shown

Fig. 2 Common HLA-B*, Cw* haplotypes in different populations.

*HLA-B**	*HLA-Cw**	*CA*	*AA*	*AsA*	*HA*
0702	*0702*	■	■		■
0801	*0701*	■	■		■
4402	*0501*	■			■
3501	*0401*	■	■		■
4001	*0304*	■		■	
5701	*0602*	■			
1501	*0303*	■			
1402	*0802*	■			■
3801	*1203*	■			
3701	*0602*	■			
4403	*0401*		■		
5301	*0401*		■		
1503	*0202*		■		
4201	*1701*		■		
4901	*0701*		■		
5802	*0602*		■		
5801	*0701*		■		
4501	*1601*		■		
4501	*0602*		■		
5801	*0302*		■	■	
5101	*1402*			■	
3802	*0702*			■	
4601	*0102*			■	
1502	*0801*			■	
4001	*0702*			■	
5401	*0102*			■	
1301	*0304*			■	
4403	*1403*			■	
3901	*0702*			■	
4403	*1601*				■
4002	*0304*				■
1801	*0501*				■
3512	*0401*				■
5101	*1502*				■
3905	*0702*				■
5001	*0602*				■

■ Frequency ≥2.0 %;
CA: Caucasian American, AA: African-American, AsA: Asian-American,
HA: Hispanic American.
HLA-B*, Cw haplotype frequency based on Cao K, Yunis E., et al.

are the frequent HLA-Cw/B blocks of Hispanic Americans to show haplotypes that indicate ethnic admixture. However, some blocks are much more frequently found in only one or two of the groups. Certain blocks are of higher frequency: (*HLA-Cw*0702, B*0702*) ($f = 0.11$), (*HLA-Cw*0701, B*0801*) ($f = 0.11$), and (*HLA-Cw*0502, B*4402*) ($f = 0.09$) in Caucasian Americans; (*HLA-Cw*0401, B*5301*) ($f = 0.09$), (*HLA-Cw*0202, B*1503*) ($f = 0.07$), and (*HLA-Cw*1701, B*4201*) ($f = 0.05$) in African-Americans; and (*HLA-Cw*0302, B*5801*) ($f = 0.07$), (*HLA-Cw*0702, B*3802*) ($f = 0.07$), and (*HLA-Cw*0102, B*4601*) ($f = 0.06$) in Asian-Americans.

2.2.2 HLA-DR/DQ Block

The HLA-DR/DQ region, spanning approximately 150 kb, is the longest of the four MHC blocks. The definition of this block is based on published frequencies of *HLA-DRB1*, *-DQB1* haplotypes in three different American populations. *HLA-DRB3/4/5* and *-DQA1* are other well-characterized polymorphic loci located within this DNA segment. Figure 3 summarizes the frequency data for the most frequent DR/DQ haplotypes in various ethnic groups. Some of these DR/DQ blocks are found much more frequently in certain ethnic groups. For example, (*HLA-DRB1*0401, DQB1*0301*) is primarily found in Caucasian Americans ($f = 0.03$); (*HLA DRB1*1503, DQB1*0602*) is primarily found in African-Americans ($f = 0.07$); and (*HLA-DRB1*0901, DQB1*0303*) is primarily found in Asian-Americans ($f = 0.10$). Other DR/DQ haplotypes are found frequently in more than one ethnic group. There are also a few haplotypes shared between Americans of all major ethnicities. For North American Hispanics, some of the haplotypes are ethnically specific and probably resulted from admixture with Amerindians. This is easily demonstrated when studying Mexicans, Mexican Americans, or other Hispanic Americans. In such cases the degree of admixture with African, Asian, and Amerindian haplotypes is variable (Fig. 3).

DRB1*, DQB1*	CA	AA	AsA	HA	P.R. A	MM	Mazatecan	Zapotec	Mixe	Mixtec
1501, 0602	■		■	■	■	■				
0101, 0501	■			■	■	■				
0301, 0201	■	■		■	■	■				
1101, 0301	■		■	■		■				
1301, 0603	■			■	■					
0701, 0201	■	■				■				
0401, 0301	■				■					
0401, 0302	■			■	■					
0404, 0302	■			■	■	■	■	■	■	
1302, 0604	■		■	■						
0701, 0202	■	■		■	■					
0701, 0303	■			■						
1001, 0501				■						
1101, 0302					■					
1102, 0302					■					
0302, 0402		■			■					
1102, 0301		■								
0804, 0301		■								
1503, 0602		■								
1301, 0303		■								
1101, 0602		■			■					
0102, 0501		■								
0901, 0303			■							
0405, 0401			■							
0803, 0601			■							
1202, 0301			■							
1401, 0503					■					■
0802, 0402						■	■	■	■	■
0407, 0302						■	■	■	■	■
1602, 0301							■	■	■	■
1406, 0301							■	■	■	■
0411, 0302							■			
1402, 0301							■	■	■	■
1502, 0601			■				■			
0403, 0302			■					■	■	■
1602, 0502										■

■ Frequency ≥ 2.0 %

Shared and ethnic-specific DRB1*, DQB1* haplotypes based on Yunis EJ et al. (B, C), Zuñiga J, Azocar J, Yunis EJ (unpublished results P.R.A.), Vargas-Alarcon G et al., Arnaiz-Villena A et al., Hollenbach JA., et al.

CA: Caucasian American, AA: African-American, AsA: Asian-American, HA: Hispanic-American, P.R.A: Puerto Rican American, MM: Mexican Mestizos.

Fig. 3 Common *HLA-DRB1*, DQB1** haplotypes in different populations.

2.3 Common HLA Haplotypes Defined Serologically and Deduced HLA-Cw/B, HLA-DRB1/DQB1 Common Haplotypes and their Frequencies

We previously summarized the frequency distribution of common HLA haplotypes defined by serological specificities for *HLA-A, B,* and *DR*. The allele-level typing of these haplotypes has not been determined systematically in a large database. Figure 4 shows the probable allele-level typing of the most common (serologically defined) HLA haplotypes in four American ethnic groups. The alleles assigned were determined by a combination of deduction and specific studies of individual haplotypes, often found in homozygous cell lines. Not shown in the figure is the fact that there is a large number of Hispanic common haplotypes with a frequency of less than 0.005, which is an indication of the higher genetic diversity of Hispanic Americans.

2.4 The Aggregate Block Frequency (ABF) Measures Genetic Diversity

The sum of frequencies of small blocks, larger blocks, or entire haplotypes within

HLA-A*	HLA-B*, HLA-Cw*	HLA-DRB1* HLA-DQB1*	CA	AA	AsA	HA
0201	0702 0702	1501 0602	■			
0301	0702 0702	1501 0602	* ■			■
0101	0801 0701	0301 0201	* ■	■		* ■
3301	1402 0802	0102 0501	■			* ■
0201	1501 0303	0401 0302	■			
0301	3501 0401	0101 0501	■			
0201	4402 0501	0401 0301	* ■			
2902	4403 1601	0701 0202	■			■
0101	5701 0602	0701 0303	■			
3001	4201 1701	0302 0402		■		
3001	1302 0602	0701 0202			■	
2402	0702 0702	0101 0501			■	
3303	4403 1403	1302 0604			■	
3303	4403 1403	0701 0202			■	
0207	4601 0102	0901 0303			■	
2402	5201 1202	1502 0601			■	
3303	5801 0302	0301 0201			■	
0201	3501 0401	0407 0302			* ■	
0201	3512 0401	0802 0402				■

■ Frequency ≥1.0% , *Haplotypes with frequency ≥2.0 %.

CA: Caucasian American, AA: African-American, AsA: Asian-American, HA: Hispanic-American.

Frequencies based on Yunis EJ et al., and Cao K et al.

Fig. 4 Common *HLA-A*, B*, Cw*, DRB1*, DQB1** haplotypes in different populations.

the MHC, defined by two or more loci sharing significant LD, is named ABF (*aggregate block frequency*). The frequencies higher than 1% (arbitrarily) of variants of particular blocks are added to compare the degree of diversity of populations studied. It is of interest that the ABFs of individual HLA blocks usually do not show much difference in all ethnicities studied. However, the ABFs of two or more combined blocks show that the Caucasian Americans have less diversity than African- or Asian-Americans. Of interest, the Mexican and Puerto Rican Americans appear to have a higher level of diversity by the use of available information of the DR/DQ block. In preliminary studies, we have determined that the Puerto Rican population of Massachusetts shows a high genetic diversity due to the admixture of Caucasian, African, and Asian haplotypes that are present in frequencies of less than 1%. The ABF for the common haplotypes from *HLA-A* to *HLA-DQB1* is 0.17 in Caucasian Americans, whereas it is 0.03 in African-Americans and 0.10 in Asian-Americans. The ABFs for common haplotypes of Hispanic Americans is approximately 0.06.

2.5 Class III Region Blocks

2.5.1 TNF Block

The shortest conserved MHC block studied is of 7 kb size and contains the genes for TNF-α and TNF-β (lymphotoxin-α). The *TNF* (tumor necrosis factor) microsatellites a, b, and c are upstream (a, b) or within (c) the *LTA* gene. Also, the TNF block includes the SNPs of the *TNF* promoter region. Specific variants of this block are present in all populations while others may be population-specific. Specific variants of this block are also haplospecific. For example, the nucleotide A (substituting for the common G nucleotide) at the −243 position is a marker for the African CEH [*HLA-A*3001, Cw*1701, B*4201, FC(1,90)0, DRB1*0302, DQB1*0402*]. Also, the *TNF* 237(G-A)* SNP is haplospecific for the *HLA-B*4001* CEH found often in Caucasians and for at least some of the *HLA-B48* haplotypes of East Asians found also in Colombian Amerindians. Likewise, a *TNF*-307(G-A), *862(C-A)*, and **856(C-T)* combination is found in both Caucasian and Asian haplotypes. It is not surprising that variants of this block are in LD with both complotypes and the HLA-Cw/B block variants.

2.5.2 Complotype Block

The four MHC-encoded complement genes (*C4B, C4A, BF*, and *C2*) are located between *HLA-DRB1* (525-kb centromeric to the complement genes) and *TNF* (350-kb telomeric to the complement genes). This block is variable in length from 75 to 120 kb. The gene order is Telomere-*C2-BF-C4A-CYP21A-C4B-CYP21B*-Centromere. Complotypes are inherited together as a unit, essentially without recombination. The most frequent haplotypes are *SC31* and *FC31* detected in all ethnic groups and *SC01* frequent among the British. Two complotypes are ethnic specific, the *SC21* of the Ashkenazi Jews and the *F1C30* of Basques and Sardinians.

3 Conserved Extended Haplotypes (CEHs)

Family studies have identified MHC haplotypes that show a significant degree of sequence conservation across a large stretch of the MHC in the region between *HLA-Cw, B, TNF, HSP-70*, complotype, and *DR,DQ*. Twelve of the CEHs found in

Caucasians are listed in Fig. 5. For detailed information, consult our earlier reviews.

3.1 Extension of the HLA-Cw/B Block to HLA-A and of the HLA-DR/DQ Block to HLA-DPB1

Figure 4 lists the high-frequency CEHs and common HLA haplotypes that extend to *HLA-A*. Inclusion of *HLA-A* alleles diminishes the ABF. At least four CEHs found in Caucasians extend to the HLA-DP region. Strong LD between the HLA-DR/DQ region and *HLA-DPB1* was shown for several ethnic groups during the 11th international histo compatibility workshop (IHWS), and more recently in Japanese. LD predictions, however, have reported a lack of nonrandom association between HLA-DR/DQ and DP, perhaps reflecting high levels of recombination within this region and/or the inability of statistical analysis to find genetic fixity or nonrandom associations in relatively small databases of haplotypes or unrelated individuals. In Caucasians, the CEH [*HLA-A*0101, Cw*0701, B*0801, SC01, DRB1*0301, DQB1*0201*] frequent in both the British and in Ashkenazi Jews, often extends to *HLA-DPB1*0401*. Also, the CEH [*HLA-A*3002, Cw*0501, B*1801, F1C30, DRB1*0301, DQB1*0201*] of the Basques and Sardinians often extends to *HLA-DPB1*0202*.

3.1.1 Frequency Distribution of MHC Blocks, CEH Fragments, and CEHs

Table 1 shows the frequency distribution of DNA blocks within a region of MHC of approximately 1.5 MB, between *HLA-Cw* and *HLA-DQB1*, in a study of 372 haplotypes determined in 92 pedigree-analyzed Caucasian families. Figure 5, showing the distribution of these haplotypes as analyzed by our previously published model, demonstrates that, of the 12 most frequent Caucasian CEHs, only 9 CEHs show nonrandom association with significant LD (<0.0005). Interestingly, approximately one-third of the haplotypes are composed of combinations of blocks in random association. Therefore, some CEHs cannot be found with statistical significance in this limited database. However, the small HLA-Cw/B and DR/DQ blocks were found to have significant LD corroborating those described previously. These data are the first demonstration of the frequency distribution of blocks in genotyped families ($n = 92$), although the data for complotypes were deduced from previous studies. The ABF measurement is comparable to those that had been deduced by us from separate reports of HLA-Cw/B and DR/DQ blocks. The high frequency of the complotype SC31 explains why CEHs carrying it require a larger database to produce LD to reach statistical significance.

3.1.2 The Global LD Measurement Related to the Size of the Database

As had been reported before, global LD measurements of large size DNA blocks or haplotypes are not informative when the size of the data studied is small. However, there were several small blocks and fragments of CEHs that can be found even with a limited number of haplotypes as demonstrated in the analyses of the data shown in Fig. 5 and Table 1.

3.1.3 Association of CEHs with Disease

Disease susceptibility or protective alleles of particular loci are included in several CEHs or their associated fragments. Important studies of pedigree-analyzed CEHs have been conducted to localize disease susceptibility genes, including type 1 diabetes, IgA deficiency,

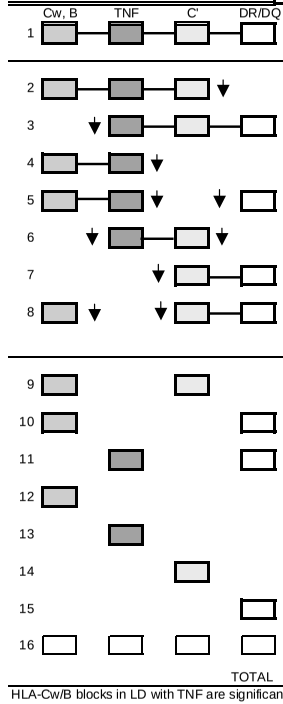

Block combinations	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII		
HLA-Cw*	0701	0702	0501	0602	0304	1601	0602	0802	0304	0401	0401	0401		
B*	0801	0702	4402	5701	4001	4403	1302	1402	1501	3501	3501	3501		
TNF microsatellites	2, 3, 1	11, 4, 1	6, 5, 1	2, 5, 2	4, 5, 1	7, 4.1	7, 4.1	2, 1, 2	2.1, 2	5,5, 2	10, 4,1	5,5, 2		
Complotype	SC01	SC31	SC30	SC61	SC02	FC31	SC31	SC2(1,2)	SC33	SC31	SC31	FC(3,2)0		
HLA-DRB1*	0301	1501	0401	0701	1302	0701	0701	0102	1401	1101	0401	0101		
DQB1*	0201	0602	0301	0303	0604	0202	0202	0501	0503	0301	0301	0501		
DQA1*	0501	0102	0302	0201	0102	0201	0201	0101	0104	05	0302	0101	Total	
1	30	28	14	7	5	6	3	3	3	5	3	2	109	109 Conserved extended haplotypes
2	3	9	2	1	2	1	2	2	1	2	1	0	26	
3	4	11	2	1	1	1	2	1	0	1	1	0	25	
4	1	3	2	1	1	1	1	1	1	1	2	1	16	
5	1	2	2	0	0	0	0	0	0	1	1	0	7	
6	3	8	1	0	2	0	0	0	0	1	1	0	16	
7	2	6	2	1	1	0	1	0	1	0	0	1	15	
8	1	2	1	1	0	0	0	0	0	1	0	0	6	111 Fragments
9	1	3	1	0	0	0	0	0	0	0	1	0	6	
10	1	3	1	0	1	1	0	1	1	1	1	1	12	
11	1	1	1	0	0	0	0	0	1	0	0	1	5	
12	2	2	2	2	1	1	1	1	1	2*	2*	2*	15	
13	1	8	3	0	3	1^^	1^^	0	0	0^	1	0^	17	
14	1	41*	2	0	0	5	41*	0	1	41*	41*	1	51	
15	2	4	3''	2	1	2'	2'	3	2	3	3''	5	27	133 Random
16													19	19 Rare blocks""
TOTAL														372 total

HLA-Cw/B blocks in LD with TNF are significant at < 0.001 for all CEHs (listed as Roman numerals at the top of the page), but all four blocks are in LD in only 9/12 haplotypes (excluding VII, X and XI).
*, ^, ^^, ', '' These marks are used to indicate numbers of haplotypes for variants 9–15 that cannot be assigned to a unique CEH. The values listed are counted only once for totals but also for each CEH.
Significance of LD between Cw, B and DR, DQ was also present in the same 9 common haplotypes.
The arrows represent regions of potential recombination.
"" Not included in the upper 15 kinds of haplotypes which are or could be part of one of the 12 CEHs listed.

Fig. 5 Model of four blocks or fragments of DNA within the MHC in a CEH. Using 16 kinds of haplotypes, the solid lines represent stretches of conserved DNA.

Tab. 1 Frequencies of MHC blocks (Cw/B, TNF, complotype, and class II genes) in a Caucasian population of 92 families.

Cw, B**			*TNF*		*Complotype*[d]		*DRB1*, DQB1*, DQA1**		
	HF			*HF*		*HF*		*HF*	
Blocks	*Published*[a]	*Observed*	*Blocks*	*Observed*	*Blocks*	*Observed*[b]	*Blocks*	*Published*[c]	*Observed*
0701,0801	0.111	0.108	*2,3,1*	0.118	*SC01*	0.120	*0301,0201,0501*	0.056	0.112
0702,0702	0.107	0.140	*11,4,1*	0.188	*SC31*	0.340	*1501,0602,0102*	0.085	0.139
0501,4402	0.091	0.070	*6,5,1*	0.075	*SC30*	0.067	*0401,0301,0302*	0.030	0.086
0602,5701	0.039	0.032	*2,5,2*	0.026	*SC61*	0.030	*0701,0303,0201*	0.024	0.032
0304,4001	0.064	0.026	*4,5,1*	0.038	*SC02*	0.030	*1302,0604,0102*	0.023	0.024
1601,4403	0.017	0.026	*7,4,1*	0.048	*FC31*	0.035	*0701,0201,0201*	0.031	0.043
0602,1302	0.013	0.019	*2,1,2*	0.035	*SC2(1,2)*	0.016	*0102,0501,0101*	0.010	0.021
0802,1402	0.032	0.022	*5,5,2*	0.040	*SC33*	0.016	*1401,0302,0501*	–	0.021
0304,1501	0.011	0.019	*10,4,1*	0.019	*FC(3,2)0*	0.010	*1101,0301,05*	0.037	0.032
0401,3501	0.064	0.070					*0101,0501,0101*	0.060	0.027

[a]Cao, K., Hollenbach, J., Shi, X., Shi, W., Chopek, M., Fernandez-Viña, M.A. (2001) Analysis of the frequencies of *HLA-A*, *B*, and *C* alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations, *Hum. Immunol.* **62**, 1009–1030.
[b]Modified from Alper, C.A. (1998) Extended MHC haplotypes and disease markers, *ISI Atlas Sci: Immunol.* **1**, 79–83.
[c]Yunis, E.J. et al. (2005).
[d]Complotypes were deduced on the basis of the association between Cw/B, TNF, and complotypes.
The term "blocks" was defined as published. LDs were significant between Cw/B and TNF in all 12 kinds of block combinations listed by Roman numerals in Fig. 5. However, only 9 were significant for CEHs (not VII, X, and XI).

dermatitis herpetiformis, and gluten-sensitive enteropathy. Likewise, fragments of the CEH [*HLA-Cw*0303, B*5502, SB45, DRB1*1401, DQB1*0503*] served to map pemphigus vulgaris (PV) genetic susceptibility in non-Jewish patients to the DR, DQ region. But, PV in Jewish patients is associated with the entire Ashkenazi CEH [*HLA-A*2601, Cw*1203, B*3801, SC21, DRB1*0402, DQB1*0302*]. One of the most frequent CEH of Caucasians, [*HLA-A*0301, Cw*0702, B*0702, SC31, DRB1*0501, DQB1*0602*], has been found to be associated with multiple sclerosis (MS) and with ragweed allergy. In another study, we had prior evidence that the CEH [*HLA-B8, SC01, DR3*] was increased among nonresponders to hepatitis B vaccine and that such subjects were more homozygous than expected. We prospectively immunized homozygotes and heterozygotes and none of the homozygotes responded, confirming the dominant role of the MHC in the immune response.

4 Mapping of MHC Disease Genes: New Model for Future Studies Combining Genotyped Individuals with the Use of High-throughput SNPs

Unfortunately, the great genetic fixity of the MHC has prevented formal identification of the true susceptibility gene for any of the diseases described earlier. A more practical approach than simply SNP "mapping" the entire MHC would be to genotype at least the four blocks we described and then, as a second step, to analyze more intensively the localized region (small block) most frequently associated with the disease. The "random" variants of that small block present in patients would perhaps be more informative than those variants found in the CEHs associated with the disease. In the example described in Fig. 5, one-third of the haplotypes contained only such random small blocks.

5 Paralogous Chromosomal Regions

Paralogous chromosomal regions are those genomic regions found in a single species, which are thought to have a common genetic origin, but which have been separated by mechanisms of duplication from an ancestral genetic unit. Other hypotheses suggest that perhaps independently duplicated genes were grouped during evolution by selective forces. Paralogous genes have been found in different chromosomal segments in humans. The early findings demonstrate the existence in chromosomes 11 and 12 of duplicated chromosomal segments in which two pairs of duplicate genes were contained. In general, the block duplication mechanism can better explain why genes paralogous to some found in the MHC are clustered in three other specific regions of the human genome.

Several published studies have described the existence of three chromosomal regions paralogous to the MHC: 1q21-25, 9q33-34, and 19p13.4–13.1. The possible origin of MHC paralogous regions on chromosomes 1, 9, and 19 could be from two rounds of duplication of the whole genome (2R hypothesis). These findings are supported by the identification of paralogous genes in *Drosophila melanogaster* and in other species in human chromosomes 1, 9, and 19. Phylogenetic analyses have revealed that the MHC and its paralogous regions have a common (ancestral) genomic region that was duplicated as

a block after the divergence of cephalochordates and vertebrates. Duplications (polyploidizations) could have occurred between 766 and 528 Myr ago. Interestingly, the MHC paralogous region in chromosome 9q32-34 retains an ancestral state in the context of organization and gene substitution patterns compared with both the human 1p31-p11, 19p13, and 6p21.3 regions and with the amphibious, chordate, and gnathostomata proto-MHC regions. The MHC paralogous genes are located in the classical MHC class I and class II regions, but also can be found in neighboring regions of the MHC. Gene families such as *NOTCH*, *BRD*, and *PBX* have copies in all of the MHC paralogous regions, but some genes that have paralogous copies in chromosomes 1, 9, and 19 are not found in the MHC.

Among the other previously mentioned paralogous regions, the 1q21-25 is unique because it contains histocompatibility-like loci in the CD1 and MR1 regions. CD1 molecules show structural and functional similarities in relation to HLA class I molecules. The CD1 molecules can present certain glycolipids to cytotoxic T lymphocytes. Genomic sequence of the human chromosome 9 reveals a highly polymorphic structure with a total of 31 genes paralogous to genes found in the human MHC. Interestingly, in chromosome 9, more than 90 genes associated with different human diseases have been mapped. These include *AIF1-L*, *BAT2-L*, *C5*, *PBX3*, *BRD3*, *COL1A5*, and *RXRA*. This fact is important in attempting to decipher the possible influence of several loci in polygenic diseases. Also of interest is that, in the extended MHC class I region of chromosome 6, families of olfactory receptor genes have been mapped also in the chromosome 9 paralogous region. This fact could be a determinant in the species preservation and reproduction, as recent reports suggested that mate selection could be influenced by HLA segregation of specific parental haplotypes that could be in LD with specific olfactory related genes in the proximity of MHC.

6 Discussion

High-resolution definition of the most frequent MHC haplotypes extending from at least *HLA-Cw* to *HLA-DQB1* is incomplete at the phenotype level, much less at the genotype level using pedigree analysis. We summarized the current knowledge of allelic variants of well-studied loci in common HLA haplotypes and, to a limited extent only, in MHC CEHs. Using the frequencies of nonrandomly associated alleles of two close loci, *HLA-Cw, -B* in one example and *HLA-DRB1, -DQB1* in another, we described blocks of DNA measuring approximately 100 to 150 kb each. We deduced that the existence of common HLA haplotypes, defined as the *nonrandom association* of these two blocks, is related to genetic fixity within the MHC spanning at least 1.4 Mb as exemplified by the existence of CEHs. When typing of the intermediate class III region of the MHC (complotypes, TNF, and *HSP-70*) and haplotype identification by pedigree analysis are conducted, approximately 85% of common HLA haplotypes are the same as genetically fixed CEHs. We showed in Fig. 5 that approximately 22% of common HLA haplotypes are not CEHs because the intermediate regions have different degrees of LDs (often related to separate CEH variants).

Although several of the MHC loci are among the most polymorphic known in the human genome, they represent only a small percentage of the total MHC DNA

sequence. Formal proof of the essential fixity of the complete DNA sequence for specific MHC CEHs shall require DNA sequencing of the entire region in several independent examples. This work has begun on consanguineous cell lines, with the MHC Haplotype Project. If one expands the CEH definition to include the loci from *HLA-A* to *HLA-DPB1*, the fixed DNA length would be more than 3 Mb of DNA.

The frequency of block combinations varies between major ethnic groups and/or in different nationalities. The preponderance of common haplotypes in Caucasians may be due to the fact that more Caucasians than Africans or Asians have been studied or may simply reflect a lower genetic diversity among Caucasians. Prediction of individual HLA blocks (e.g. HLA-Cw/B) by LD has been more complete than has identification of larger fragments or CEHs (which include TNF and/or complotype blocks) by family study. Identifying CEHs in a population depends on the size of the haplotype database. Lower frequency CEHs can only be identified in larger databases. It is important to keep in mind that the number of and the particular types of CEHs (or common haplotypes) detected will also depend on the type (e.g. ethnicity, disease status) of the population studied. Some CEHs, such as the first six listed in Table 1 and Fig. 5, are relatively frequent in healthy Caucasians. In a database of 500 Caucasian haplotypes, we would predict the aggregate frequency of all of the CEHs listed in Fig. 5 (not including *HLA-A*) to be at least 30%. In this review, we reported an analysis of 372 haplotypes derived from pedigree analysis of 92 parents, although the complotypes were deduced from probable nonrandom associations using our previous studies. Only nine CEHs showed significant nonrandom associations. The number of individuals studied would also affect the distribution of CEHs, their fragments, and small blocks. The arrows in Fig. 5 suggest sites of possible historical recombination that in many recent articles using SNPs are referred to as *hotspots*.

Structural analysis of MHC haplotypes is incomplete, but we believe that the available evidence demonstrates that there are fixed genetical segments of DNA within the MHC, the frequency of which vary in different human populations. Some CEHs have shown significant genetic fixity across blocks of DNA as large as 3.2 Mb from *HLA-A* to *-DPB1*. However, the most completely described segment spans *HLA-Cw* and *-DQB1*, and includes polymorphisms of complement genes (complotypes) and TNF. The boundaries of this 1.5-Mb block mark CEHs that are relatively frequent, which have proven informative for mapping disease susceptibility and immune response genes and they could be used for allotransplantation matching.

The concept of the aggregate frequency of individual common block variants spanning the MHC segment between HLA-Cw/B and HLA-DR/DQ, or ABF, may be used to determine the degree of relative diversity of Americans of Asian, Caucasian, African, or Hispanic ancestry. Likewise, we suggest that determining the ABF of a combination of small blocks can be used to estimate the degree of genetic fixity of even larger segments of DNA called *common HLA haplotypes* (which are usually representative of CEHs). Thus, ABF is a measurement of the frequency of DNA blocks of any particular size range, based on the aggregate frequency of nonrandomly associated alleles at two or more loci, which estimates the frequency of various sizes of genetically fixed DNA within the MHC of well-defined populations. The

result can be used to compare human diversity among well-defined populations.

Other DNA methods, such as RFLP (restriction fragment length polymorphism), SNP analysis, and microsatellite polymorphisms, have also been useful in defining extended genomic variation throughout the MHC at sites other than the classical HLA loci (e.g. complement and TNF locus typing). Allele-level typing of *HLA-A*, *-Cw*, *-B*, *-DRB1*, and/or *-DQB1* has been reported for a limited number of Americans of different ethnicities. At least three characteristics of these MHC haplotype blocks are quite different from those recently reported. MHC blocks described here are never defined solely by SNP typing and are highly polymorphic within well-characterized populations. In contrast to two to five common SNP-defined haplotype variants present in diverse ethnic groups, MHC blocks have many more common variants even within a single ethnic group. Finally, MHC haplotype variants exhibit a high degree of isolated population and/or ethnic specificity. Furthermore, despite both the likelihood of many recombination ''hotspots'' within the MHC and this great diversity within specific MHC blocks, many CEHs exist at relatively high frequency. Specific alleles or specific MHC blocks are often ''haplospecific'' for particular CEHs. In many cases, one particular allele at one locus is highly correlated with an entire MHC CEH. In other cases, a specific set of alleles within a single MHC block or in more than one block is required to achieve a high correlation with a particular CEH. However, haplotype determination by family study is required to rigorously define a complete haplotype.

Many MHC haplotypes are not CEHs but contain only fragments or even only a single small block from any specific CEH. Non-CEH haplotypes containing CEH fragments are represented in Fig. 5 by any haplotype except #1 (the full CEH) or #16 (containing no part of any CEH). For example, not all haplotypes containing the (*HLA-Cw*0701*, *B*0801*) block exist solely as the CEH [*HLA-Cw*0701*, *B*0801*, *TNF 2,3,1*, *SC01*, *DRB1*0301*, *DQB1*0201*] (although, we argue, the majority are contributed by this CEH). The distribution in frequencies between any given CEH and its specific blocks depends on several factors, including at least the degree to which a particular block exists in more than one extended haplotype (i.e. is not haplospecific) and the degree of fragmentation of the extended haplotype in the study population. If a disease gene is mapped to a CEH, the number of haplotypes bearing fragments of that CEH that is required to localize the gene will be greater than the number of ''random'' haplotypes required. If the gene is already mapped to a small block, variants of that block that are not associated with any CEH required to localize the gene should be fewer, because the complete set of these ''random'' variants should share little that is unique except for the susceptibility locus. For these reasons, it is probably more difficult to map susceptibility genes in American Caucasians than in other American ethnicities simply because the level of genetic diversity is lower in Caucasian Americans. We also believe that most of the diversity is possibly related to the degree of genetic admixture of individuals recruited, and this admixture could be within nationalities or between ethnicities.

An example of this complexity can be the use of the term Hispanic to describe individuals that originated from Spanish-speaking countries. Of course, such individuals probably resulted from different degrees of population admixture

producing stratification, and they represent many different ethnicities. The use of SNPs or short tandem repeats or variants of genes that could have been conserved in all humans or within a particular ethnicity will need to be used to rule out population genetic stratification in every study population. In this case, DNA blocks that are conserved in the genome need to be identified in order to distinguish them from DNA blocks that show variability caused by population admixture or by natural selection.

CEHs might result from many possible mechanisms, including recent population bottlenecks, recombination suppression, preferential transmission, migration and admixture, and/or genetic drift or natural selection. Duplication of genes involved in immune functions argues in favor of selection for blocks of DNA, which vary in different populations. Paralogous genes to some found in the human MHC with equal or related functions are located in other chromosomes (Chromosomes 1, 9, and 19). Such duplicated genes add another dimension of difficulty to disease gene localization because there may be several genes involved in a disease, which are molecularly related. For example, the *NOTCH* genes are found in the four paralogous chromosomal regions. One caveat, however, is that the MHC paralogous region of chromosome 9 contains many other genes that may be in LD with *NOTCH1*, but the equivalent genes to those are not present in chromosome 6. In the case of schizophrenia, *NOTCH4* could be in LD with another marker different from that of *NOTCH1* and it could suggest a difference in the populations studied. However, the *BRD2* and *BRD3* genes of chromosomes 6 and 9, respectively, are at similar distances to the *C4* genes of chromosome 6 and the *C5* of chromosome 9. The latter region is expected to have genes in LD that may be acting synergistically.

7 Future Studies

Complete human population studies based on the CEH and DNA blocks concept are needed. Requirements will include collecting data on the basis of geography (nationality), language, and ethnicity, as well as identifying other (non-MHC) genetic markers. The latter include the Y chromosome, mitochondrial DNA, blood groups, and short tandem repeats outside the MHC region. Such analysis could derive the age of haplotypes and population diversity relationships. Genetic admixtures need to be analyzed in order to know the markers that indicate genetic drifts or old admixtures. We believe that any unique marker of the polymorphic loci of a haplotype would be a marker of the age of the MHC haplotype. Two or more ancient mutations or gene conversions inherited together will have random frequencies, but during evolution they are fixed at the population level and may prove to have selective advantage. Natural selection is often mentioned as a major influence in MHC evolution and in the formation of CEH. However, nonrandom association between distant loci may be the result of directional selection (molecular cooperation during the immune response). In this context, the possible role of paralogous genes may be operating. Alternatively, balancing selection by a higher replacement rate of different alleles at certain loci could have been an influence, particularly a departure from neutral expectation of HLA frequencies.

8 Summary

We described an updated interpretation or definition of MHC or HLA haplotypes. Although the information is incomplete, we believe that there are genetically fixed MHC haplotypes of variable sizes in the human population and that the stretch of fixed DNA could be up to 3.2 Mb of DNA in several haplotypes that extend from *HLA-A* to *HLA-DPB1*. However, the fixity of the DNA between *HLA-Cw* and *HLA-DQB1* has been studied more extensively, using polymorphisms of the complement system (complotype) and the TNF region to define that stretch of fixed DNA of approximately 1.5 Mb. The boundaries of this large block, between *HLA-Cw* and *HLA-DQ*, mark MHC CEHs with fixed (or nearly so) DNA that are relatively frequent and have been used for mapping studies of disease and immune responses, and they could be used for matching for allotransplantation. The aggregate block frequencies (ABFs) of both the HLA-Cw/B and HLA-DRB1/DQB1 small blocks were investigated in three different ethnic groups living in the United States and in five African populations. The results suggest that the aggregate frequency of the most common small blocks in strong and significant LD may be used to determine the degree of relative diversity of Asian, Caucasian, or African-Americans and Hispanic populations. Likewise, we suggest that determining the ABF of a combination of specific small blocks can be used to determine the degree of genetic fixity of even larger segments of DNA. Thus, ABF is a measurement of the total frequency of fixed genetic (high population frequency) variants of a particular DNA block, to estimate the frequency of various sizes of DNA within the MHC of well-defined populations. This measurement adds a new dimension in the studies of human diversity. Essentially, we have developed a model of combinations of blocks of 16 kinds of MHC haplotypes, with one of them being the CEH, seven of them being CEH fragments containing combinations of the CEH blocks, seven more having only one CEH block or two separated blocks, and one of them containing only rare blocks that do not show any LD. We used as an example the higher degree of genetic diversity of the Mexican Mestizos as compared with Caucasian Americans or specific Amerindian subgroups to demonstrate that diversity as measured by ABF can vary related to the contribution of the genetic diversity of the populations that contributed to a particular population admixture. We anticipate that genetic diversity will also differ between supposedly ethnically similar individuals when they are recruited from different geographic regions. Although not defined in the same manner, as recently reported blocks of SNPs, previously identified MHC blocks are essentially genetic units. We discussed the problem of attempting to use SNP blocks to localize disease genes without previous assignment of MHC blocks determined by high-resolution typing and preferably by pedigree analysis. We suggest that after this is done, it is possible to map disease or protection genes using positional cloning and high-throughput technologies. Disease genes could be located in either large or small blocks of DNA, but we believe that it will be easier to look for them in random blocks not associated with high-frequency CEHs. We also discussed the possibility of applying the techniques and analytical methodologies developed for the MHC to other chromosomes of the human genome, with the thought that other regions paralogous to the MHC might be

best suited for initial work. Studies will be required to compare genes that are involved in the same pathways or networks, especially those that have evolved with high levels of polymorphism. Individual locus and haplotype variation in these paralogous regions characteristic of both specific ethnicities and specific diseases need to be studied to determine whether paralogous genes with similar or related functions might explain ethnic variation in the pathogenesis of certain diseases such as autoimmune diseases.

Acknowledgment

Yunis and Zuñiga were supported by Public Health Service grants HL29583 and HL59838 from the National Heart, Lung, and Blood Institute of the National Institutes of Health (NIH) and a gift from Granados. Larsen, Awdeh, and Alper were supported by Public Health Service grant HL29583 from the National Heart, Lung, and Blood Institute of the NIH.

See also Genetic Analysis of Populations; Genetic Variation and Molecular Evolution; Genetics, Molecular Basis of; Human Genetic Variation and Disease; Targeted Therapy: Genomic Approaches.

Bibliography

Book and Reviews

Allcock, R.J.N., Atrazhev, A.M., Beck, S., de Jong, P.J., Elliott, J.F., Forbes, S., Halls, K., Horton, R., Osoegawa, K., Rogers, J., Sawcer, S., Todd, J.A., Trowsdale, J., Wang, Y., Williams, S. (2002) The MHC haplotype project: a resource for HLA linked association studies, *Tissue Antigens* **59**, 520–521.

Alper, C.A. (1998) Extended MHC haplotypes and disease markers, *ISI Atlas Sci: Immunol.* **1**, 79–83.

Alper, C.A., Awdeh, Z.L., Yunis, E.J. (1989) Complotypes and extended haplotypes in laboratory medicine, *Complement Inflamm.* **6**, 8–18.

Alper, C.A., Awdeh, Z.L., Yunis, E.J. (1992) Conserved extended MHC haplotypes, *Exp. Clin. Immunogenet.* **9**, 58–71.

Alper, C.A., Raum, D., Karp, S., Awdeh, Z.L., Yunis, E.J. (1983) Serum complement "supergenes" of the major histocompatibility complex in man (complotypes), *Vox Sang.* **45**, 62–67.

Bodmer, W.F., Bodmer, J., Ihde, D., Adler, S. (1969) Genetic and Serological Association Analysis of the HLA Leukocyte System, in: Morton, N.E. (Ed.) *Computer Applications in Genetics*, University of Hawaii Press, Honolulu, HI, pp. 117–127.

Ceppellini, R., Curtoni, E.S., Mattiuz, P.L., Miggiano, V., Scudeller, G., Serra, A. (1967) Genetics of leukocyte antigens. A family study of segregation and linkage, in: Curton, E.S., Mattiuz, P.L., Tosi, R.M. (Eds.) *Histocompatibility Testing*, Munksgaard, Copenhagen, pp. 149–185.

Crawford, K., Alper, C.A. (2000) Genetics of the complement system, *Rev. Immunogenet.* **2**, 323–338.

Dawkins, R., Leelayuwat, C., Gaudieri, S., et al. (1999) Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease, *Immunol. Rev.* **167**, 275–304.

Degli-Esposti, M.A., Abraham, L.J., McCann, V., Spies, T., Christiansen, F.T., Dawkins, R.L. (1992) Ancestral haplotypes are conserved population haplotypes with unique MHC structure, in: Tsuji, K., Aizawa, M., Sasazuki, T. (Eds.) *HLA 1991*, Vol. 2, Oxford University Press, Oxford, pp. 169–172.

Gjertson, D.W., Lee, S.-H. (1998) HLA-A/B and DR/DQ allele level haplotype frequencies, in: Gjertson, D.W., Terasaki, P.I. (Eds.) *HLA 1998*, American Society for Histocompatibility and Immunogenetics, Lenexa, KS, pp. 365–450.

Hokamp, K., McLysaght, A., Wolfe, K.H. (2003) The 2R hypothesis and the human genome sequence, *J. Struct. Funct. Genomics* **3**, 95–110.

Horton, R., Wilming, L., Rand, V., Lovering, R.C., Bruford, E.A., Khodiyar, V.K., Lush, M.J., Povey, S., Talbot, C.C. Jr., Wright, M.W., Wain, H.M., Trowsdale, J., Ziegler, A., Beck, S. (2004) Gene map of the extended human MHC, *Nat. Rev. Genet.* **5**, 889–899.

Kasahara, M. (1999) Genome dynamics of the major histocompatibility complex: insights from genome paralogy, *Immunogenetics* **50**, 134–145.

Larsen, C.E., Alper, C.A. (2004) The genetics of HLA-associated disease, *Curr. Opin. Immunol.* **16**, 660–667.

Ohno, S. (1970) *Evolution by Gene Duplication*, Springer-Verlag, Berlin,.

Piazza, A. (1975) Haplotypes and linkage disequilibria from the three-locus phenotypes, in: Kissmeyer Nielsen, F. (Ed.) *Histocompatibility Testing 1975*, Munksgaard, Copenhagen, pp. 923–927.

Wall, J.D., Pritchard, J.K. (2003) Haplotype blocks and linkage disequilibrium in the human genome, *Nat. Rev. Genet.* **4**, 587–597.

Weiss, K.M., Terwilliger, J.D. (2000) How many diseases does it take to map a gene with SNPs? *Nat. Genet.* **26**, 151–157.

Yunis, E.J., Larsen, C.E., Fernandez-Viña, M., Awdeh, Z.L., Romero, T., Hansen, J.A., Alper, C.A. (2003) Inheritable variable sizes of DNA stretches in the human MHC: conserved extended haplotypes and their fragments or blocks, *Tissue Antigens* **62**, 1–20.

Yunis, E.J., Larsen, C.E., Fernandez-Viña, M., Awdeh, Z.L., Cao, K., Romero, T., Alper, C.A., Hansen, J.A. (2005) Human MHC haplotypes and their fragments or blocks, in: Hansen, J.A., Dupont, B. (Eds.) *Proceedings of the 13th International Histocompatibility Workshop*, Blackwell Munksgaard, Oxford.

Primary Literature

Abi-Rached, L., Gilles, A., Shiina, T., Pontarotti, P., Inoko, H. (2002) Evidence of en bloc duplication in vertebrate genomes, *Nat. Genet.* **31**, 100–105.

Ahmad, T., Neville, M., Marshall, S.E., Armuzzi, A., Mulcahy-Hawes, K., Crawshaw, J., Sato, H., Ling, K.L., Barnardo, M., Goldthorpe, S., Walton, R., Bunce, M., Jewell, D.P., Welsh, K.I. (2003) Haplotype specific linkage disequilibrium patterns define the genetic topography of the human MHC, *Hum. Mol. Genet.* **12**, 647–656.

Ahmed, A.R., Wagner, R., Khatri, K., et al. (1991) Major histocompatibility complex haplotypes and class II genes in non-Jewish patients with pemphigus vulgaris, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 5056–5060.

Ahmed, A.R., Yunis, J.J., Marcus-Bagley, D., et al. (1993) Major histocompatibility complex susceptibility genes for dermatitis herpetiformis compared with those for gluten-sensitive enteropathy, *J. Exp. Med.* **178**, 2067–2075.

Alper, C.A., Kruskall, M.S., Marcus-Bagley, D., et al. (1989) Genetic prediction of non response to hepatitis B vaccine, *N. Engl. J. Med.* **321**, 708–712.

Arnaiz-Villena, A., Vargas-Alarcón, G., Granados, J., et al. (2000) HLA genes in Mexican Mazatecans, the peopling of the Americas and the uniqueness of Amerindians, *Tissue Antigens* **56**, 405–416.

Awdeh, Z.L., Raum, D., Yunis, E.J., Alper, C.A. (1983) Extended HLA/complement allele haplotypes: evidence for T/t-like complex in man, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 259–263.

Awdeh, Z.L., Alper, C.A., Fici, D.A., Ronco, P. II, Yunis, E.J. (1995) Predictability of alloreactivity among unrelated individuals: role for HLA-DPB1, *Tissue Antigens* **46**, 180–186.

Baena, A., Leung, J.Y., Sullivan, A.D., et al. (2002) TNFα polymorphisms are markers of human ancestry, *Genes Immun.* **3**, 482–487.

Begovich, A.B., McClure, G.R., Suraj, V.C., et al. (1992) Polymorphism, recombination, and linkage disequilibrium within the HLA class II region, *J. Immunol.* **148**, 249–258.

Blumenthal, M., Marcus-Bagley, D., Awdeh, Z., Johnson, B., Yunis, E.J., Alper, C.A. (1992) HLA-DR2, [HLA-B7, SC31, DR2], and [HLA-B8, SC01, DR3] haplotypes distinguish subjects with asthma from those with rhinitis only in ragweed pollen allergy, *J. Immunol.* **148**, 411–416.

Cao, K., Hollenbach, J., Shi, X., Shi, W., Chopek, M., Fernandez-Viña, M.A. (2001) Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations, *Hum. Immunol.* **62**, 1009–1030.

Cullen, M., Noble, J., Erlich, H., et al. (1997) Characterization of recombination in the HLA

class II region, *Am. J. Hum. Genet.* **60**, 397–407.

Cullen, M., Perfetto, S.P., Klitz, W., Nelson, G., Carrington, M. (2002) High-resolution patterns of meiotic recombination across the human major histocompatibility complex, *Am. J. Hum. Genet.* **71**, 759–776.

Daly, M.J., Rioux, J.D., Schaffner, S.F., Hudson, T.J., Lander, E.S. (2001) High-resolution haplotype structure in the human genome, *Nat. Genet.* **29**, 229–232.

Danchin, E.G., Abi-Rached, L., Gilles, A., Pontarotti, P. (2003) Conservation of the MHC-like region throughout evolution, *Immunogenetics* **55**, 141–148.

Dausset, J., Legrand, L., Lepage, V., et al. (1978) A haplotype study of HLA complex with special reference to the HLA-DR series and to Bf. C2 and glyoxalase I polymorphisms, *Tissue Antigens* **12**, 297–307.

Dawson, E., Abecasis, G.R., Bumpstead, S., Chen, Y., Hunt, S., Beare, D.M., et al. (2002) A first-generation linkage disequilibrium map of human chromosome 22, *Nature* **418**, 544–548.

De la Concha, E.G., Fernandez-Arquero, M., Gual, L., et al. (2002) MHC susceptibility genes to IgA deficiency are located in different regions on different HLA haplotypes, *J. Immunol.* **169**, 4637–4643.

Fernandez-Viña, M.A., Gao, X.J., Moraes, M.E., et al. (1991) Alleles at four HLA class II loci determined by oligonucleotide hybridization and their associations in five ethnic groups, *Immunogenetics* **34**, 299–312.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., et al. (2002) The structure of haplotype blocks in the human genome, *Science* **296**, 2225–2229.

Garcia-Merino, A., Alper, C., Usuku, K., et al. (1996) Tumor necrosis factor (TNF) microsatellite haplotypes in relation to extended haplotypes, susceptibility to diseases associated with the major histocompatibility complex and TNF secretion, *Hum. Immunol.* **50**, 11–21.

Gaudieri, S., Dawkins, R.L., Habara, K., Kulski, J.K., Gojobori, T. (2000) SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity, *Genome Res.* **10**, 1579–1586.

Hauser, S.L., Fleischnick, E., Weiner, H.L., Marcus, D., Awdeh, Z., Alper, C.A. (1989) Extended major histocompatibility complex haplotypes in patients with multiple sclerosis, *Neurology* **39**, 275–277.

Hedrick, P.W. (1987) Gametic disequilibrium measures: proceed with caution, *Genetics* **117**, 331–341.

Hollenbach, J.A., Thomson, G., Cao, K., et al. (2001) HLA diversity, differentiation, and haplotype evolution in Mesoamerican natives, *Hum. Immunol.* **62**, 378–390.

Howell, W.M., Evans, P.R., Devereux, S.A., et al. (1993) Absence of strong HLA-DR/DQ-DP linkage disequilibrium in the British and French Canadian Caucasoid populations, *Eur. J. Immunogenet.* **20**, 363–371.

Humphray, S.J., Oliver, K., Hunt, A.R., et al. (2004) DNA sequence and analysis of human chromosome 9, *Nature* **429**, 369–374.

Jacob, S., McClintock, M.K., Zelano, B., Ober, C. (2002) Paternally inherited HLA alleles are associated with women's choice of male odor, *Nat. Genet.* **30**, 175–179.

Jeffreys, A.J. (2001) Intensively punctate meiotic recombination in the class II region of the MHC, *Nat. Genet.* **29**, 217–222.

Katsanis, N., Fitzgibbon, J., Fisher, E.M.C. (1996) Paralogy mapping: Identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel *PBX* and *NOTCH* loci, *Genomics* **35**, 101–108.

Kauppi, L., Sajantila, A., Jeffreys, A.J. (2003) Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region, *Hum. Mol. Genet.* **12**, 33–40.

Kirk, K.M., Cardon, L.R. (2002) The impact of genotyping error on haplotype reconstruction and frequency estimation, *Eur. J. Hum. Genet.* **10**, 616–622.

Lewontin, R.C. (1998) On measures of genetic disequilibrium, *Genetics* **120**, 840–852.

Mungall, A.J., Palmer, S.A., Sims, S.K. (2003) The DNA sequence and analysis of human chromosome 6, *Nature* **425**, 805–811.

Raum, D., Awdeh, Z., Yunis, E.J., Alper, C.A., Gabbay, K.H. (1984) Extended major histocompatibility complex haplotypes in type 1 diabetes mellitus, *J. Clin. Invest.* **74**, 449–454.

Saito, S., Ota, S., Yamada, E., Inoko, H., Ota, M. (2000) Allele frequencies and haplotypic associations defined by allelic DNA typing at HLA class I and class II loci in the Japanese population, *Tissue Antigens* **56**, 522–529.

Sanchez-Mazas, A., Djoulah, S., Busson, M., et al. (2000) A linkage disequilibrium map of the MHC region based on the analysis of 14 loci haplotypes in 50 French families, *Eur. J. Hum. Genet.* **8**, 33–41.

Sieling, P.A., Chatterjee, D., Porcelli, S.A., et al. (1995) CD1-restricted T cell recognition of microbial lipoglycan antigens, *Science* **269**, 227–230.

Simon, S., Awdeh, Z., Campbell, R.D., et al. (1991) A restriction fragment of the C2 gene is a unique marker for C2 deficiency and the uncommon C2 allele C2*B (a marker for type 1 diabetes), *J. Clin. Invest.* **88**, 2142–2145.

Stenzel, A., Lu, T., Koch, W.A., Hampe, J., Guenther, S.M., De La Vega, F.M., Krawczak, M., Schreiber, S. (2004) Patterns of linkage disequilibrium in the MHC region on human chromosome 6p, *Hum. Genet.* **114**, 377–385.

Stewart, C.A., Horton, R., Allcock, R.J.N., et al. (2004) Complete MHC haplotype sequencing for common disease gene mapping, *Genome Res.* **14**, 1176–1187.

Stumpf, M.P.H. (2002) Haplotype diversity and the block structure of linkage disequilibrium, *Trends Genet.* **18**, 226–228.

Tokunaga, K., Ohashi, J., Bannai, M., Juji, T. (2001) Genetic link between Asians and native Americans: evidence from HLA genes and haplotypes, *Hum. Immunol.* **62**, 1001–1008.

Tsunoda, T., Lathrop, G.M., Sekine, A., Yamada, R., Takahashi, A., Ohnishi, Y., Tanaka, T., Nakamura, Y. (2004) Variation of gene-based SNPs and linkage disequilibrium patterns in the human genome, *Hum. Mol. Genet.* **3**, 1623–1632.

Turbay, D., Lieberman, J., Alper, C.A., et al. (1997) Tumor necrosis factor constellation polymorphism and clozapine-induced agranulocytosis in two different ethnic groups, *Blood* **89**, 4167–4174.

Vargas-Alarcón, G., Salgado, N., Granados, J., Gómez-Casado, E., Martinez-Laso, J., Alcocer-Varela, J., Arnaiz-Villena, A., Alarcón-Segovia, D. (2001) Class II allele and haplotype frequencies in Mexican systemic lupus erythematosus patients: the relevance of considering homologous chromosomes in determining susceptibility, *Hum. Immunol.* **62**, 814–820.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Walsh, E.C., Mather, K.A., Schaffner, S.F., et al. (2003) An integrated map of the human major histocompatibility complex, *Am. J. Hum. Genet.* **73**, 580–590.

Wei, J., Hemmings, G.P. (2000) The NOTCH4 locus is associated with susceptibility to schizophrenia, *Nat. Genet.* **25**, 376–377.

Wilton, A.N., Cobain, T.J., Dawkins, R.L. (1985) Family studies of IgA deficiency, *Immunogenetics* **21**, 333–342.

Zavattari, P., Deidda, E., Whalen, M., Lampis, R., Mulargia, A., Loddo, M., Eaves, I., Mastio, G., Todd, J.A., Cucca, F. (2000) Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection, *Hum. Mol. Genet.* **9**, 2947–2957.

Skin, Molecular Cell Biology of

Annette Wysocki[1], *Thomas Mustoe*[2], *and Gregory Schultz*[3]
[1] *University of Mississippi Medical Center, Jackson, MS*
[2] *Northwestern University, Chicago, IL*
[3] *University of Florida, Gainesville, FL*

1 Embryology of Skin 219

2 Epidermis 221
2.1 Stratum Corneum 221
2.2 Stratum Lucidum 221
2.3 Stratum Granulosum 222
2.4 Stratum Spinosum 222
2.5 Stratum Basale, or Stratum Germinativum 222
2.6 Basement Membrane Zone 223

3 Dermis 224
3.1 Papillary Dermis 224
3.2 Reticular Dermis 224
3.3 Dermal Vasculature 224
3.4 Extracellular Matrix Proteins 225
3.4.1 Collagen 225
3.4.2 Elastin 226
3.4.3 Proteoglycans and Glycosaminoglycans 227
3.5 Hypodermis 227

4 Skin Functions 227
4.1 Protection Against Pathogens 228
4.2 Protection Against Ultraviolet Radiation 229
4.3 Skin Immune System 230
4.4 Thermoregulation 231
4.5 Sensation 232
4.6 Vitamin Metabolism 233
4.7 Social Communication 233

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

5 Factors Altering Skin Characteristics 233
5.1 Age 233
5.1.1 Neonatal Skin 234
5.1.2 Adolescent Skin 235
5.1.3 Elderly Skin 235
5.2 Sun 237
5.3 Hydration 237
5.4 Soaps 238
5.5 Nutrition 238
5.6 Medications 239

6 Wound Healing 239
6.1 Hemostasis 239
6.1.1 Provisional Wound Matrix 240
6.2 Inflammatory Phase 241
6.2.1 Neutrophils 241
6.2.2 Monocytes/Macrophages 241
6.3 Proliferative Phase 242
6.3.1 Granulation Tissue 242
6.3.2 Fibroblasts 242
6.3.3 Angiogenesis 243
6.3.4 Epithelialization 243
6.4 Remodeling 244
6.5 Chronic Wounds 244
6.6 Wound Bed Preparation 245

7 Summary 245

Bibliography 247
Books and Reviews 247
Primary Literature 248

Keywords

Dermis
The dermis, or corium, is the thickest tissue layer of the skin and underlies the epidermis. It contains very large amounts of extracellular matrix components that provide strength (collagen fibers), elasticity (elastin fibers), and compressibility (proteoglycans) to the skin. The dermis is extensively vascularized and innervated.

Epidermis
The outermost layer of the skin consisting predominately of epidermal cells arranged in six histologically distinct layers: stratum corneum, stratum lucidum, stratum

granulosum, stratum spinosum, stratum basale, and basement membrane zone. The epidermis provides the major barrier to water loss and microbial invasion of the skin.

Wound Healing
The process of repairing injuries to skin that normally progresses through sequential phases of hemostasis, inflammation, repair, and remodeling.

The skin is the one organ of the body that is constantly exposed to a changing environment. Maintaining its integrity is a complex process, and assaults from surgical incisions, injuries, or burns can lead to life-threatening consequences without appropriate treatment.

Human skin is divided into two primary layers – epidermis (outermost layer) and dermis (innermost layer) (Fig. 1). These two layers are separated by a structure called the *basement membrane*. Beneath the dermis is a layer of loose connective tissue called the *hypodermis* or *subcutis*, and beneath the subcutis is the fat layer. Major functions of the skin are protection, immunity, thermoregulation, sensation, metabolism, and communication. The skin of the average adult covers approximately 3000 square inches, or an area almost equivalent to 2 m^2. From birth to maturity the skin covering will undergo a sevenfold expansion. It weighs about 6 lbs (or up to 15% of total adult body weight), is the largest organ, and receives one-third of the body's circulating blood volume. The skin forms a protective barrier from the external environment while maintaining a homeostatic internal environment. Epidermal appendages – nails, hair follicles, sweat or sebaceous glands – which are lined with epidermal cells, are also present in the skin. During the healing of partial-thickness wounds, these epidermal cells migrate to resurface the wound. This organ is capable of self-regeneration and can withstand limited mechanical and chemical assaults. The skin varies in thickness from 0.5 mm in the tympanic membrane to 6 mm in the soles of the feet and the palms of the hands. Variations are attributable to differences in the thickness of the skin layers covering underlying organs, bones, muscle, and cartilage. Diseases of the skin can result from genetic causes, so-called genodermatoses, infection, immune dysfunction, and trauma. Skin diseases can involve some or all of the layers of the skin.

1 Embryology of Skin

Skin is derived from both ectoderm and mesoderm (see Fig. 1). The epidermis, pilosebaceous glands, apocrine and eccrine sweat glands, hair follicles, and nail units are derived from the surface ectoderm. The ectodermal skin appendages develop with the formation of epidermis at 11 weeks of gestation and complete their development in 5 months. Melanocytes, nerves, and specialized sensory units originate from the neuroectoderm. As early as 5

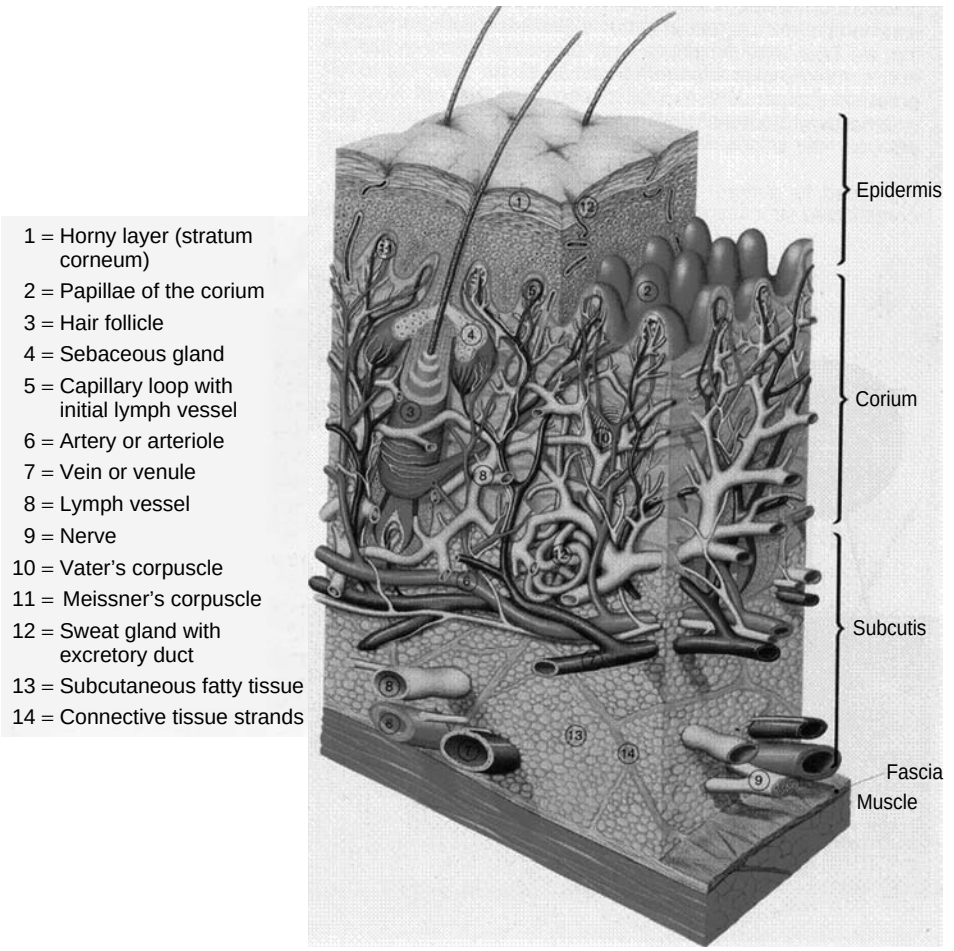

Fig. 1 Structure of skin. Major divisions of skin are the epidermis, dermis (corium), and subcutis (fat), which are separated from muscle by a layer of fascia.

weeks of human gestation, the neuroectodermal components from the neural crest cells can be detected.

The epidermis initially comes from a single layer of ectodermal cells. Soon they differentiate and form a second layer of squamous epithelium, called *periderm*, over the surface. Peridermal cells are eventually replaced by the cells arising from the basal layer.

The keratinized peridermal cells desquamate and form a part of the vernix caseosa, which covers fetal skin throughout pregnancy. It is the greasy white substance enveloping the baby at birth and functions to protect the fetal skin from uremic amniotic fluid and to act as a lubricant during birth. All components of epidermis are present at birth, and the peridermal cells are completely replaced by stratum corneum within a few months after birth. The mesoderm-derived skin elements form the structural components of dermis: the macrophages,

mast cells, Langerhan's cells, Merkel cells, fibroblasts, blood and lymphatic vessels, and fat cells. By 11 weeks of gestation the mesoderm-derived mesenchymal cells begin to produce collagen and elastin connective tissue fibers.

The dermis projects into the epidermal ridges and forms dermal papillae. The initial, simple, endothelium-lined capillaries in dermis acquire a muscular layer from the surrounding mesoderm-derived myofibroblasts, and as more simple single-layered vessels grow out from them, become arterioles and venules. The embryology of skin has profound implications for tumor genesis, prognosis, and treatment.

2 Epidermis

The epidermis, the outermost skin layer, is avascular and is derived from embryonic ectoderm. This layer is relatively uniform in thickness over the body and is between 75 and 150 μm, except for the soles and palms where thickness is between 0.4 and 0.6 mm. The epidermal layer is constantly being renewed with a turnover time ranging from 26 to 42 days. Complete epidermal renewal occurs between 45 and 75 days, or about every 2 months. The epidermal layer is composed of stratified squamous epithelial cells, or keratinocytes, and is divided into five layers. These layers, beginning from the outermost to the innermost, are the stratum corneum, stratum lucidum, stratum granulosum, stratum spinosum, and stratum basale (stratum germinativum, or simply, basal layer).

2.1 Stratum Corneum

The stratum corneum, or horny layer, is the top layer and is composed of dead keratinized cells. These squames, or corneocytes, are the cells that are abraded by the daily mechanical and chemical trauma of hand washing, scratching, bathing, exercising, and changing of clothes. The stratum corneum is composed of layers of thin, stacked, pancake-appearing, anucleate cells. These cells are filled with about 80% keratin, a tough, fibrous, insoluble protein; hence, they are called *keratinocytes*. Keratinocytes are initially formed in the basal layer and undergo the process of differentiation. The normal stratum corneum is composed of completely differentiated keratinocytes. Keratin is resistant to changes in temperature or pH and to chemical digestion by trypsin and pepsin. This same protein is found in hair and nails; in these structures, keratin is referred to as *hard* keratin compared with the *soft* keratin of the skin. Stratum corneum thickness varies with age, sex, and disease. The deeper layer of the stratum corneum is called the *stratum compactum* and the keratin in these cells is more densely packed. These cells also have a diminished capacity to bind water. Cells in the uppermost layer are partly shed because of the proteolytic degradation of the desmosomes and this layer is also called the *stratum dysjunctum*.

2.2 Stratum Lucidum

The stratum lucidum is the layer directly below the stratum corneum. This layer is found in areas where the epidermis is thicker, such as the palms of the hands or soles of the feet, and where it is prominent but absent from thinner skin, such as the eyelids. This layer can be one to five cells thick and is transparent. Cell boundaries are often difficult to identify in histological sections under a light microscope. The stratum lucidum is a

transitional layer where active lysosomal enzymes degrade the nucleus and cellular organelles before the cells move into the stratum corneum.

2.3 Stratum Granulosum

The stratum granulosum, or granular layer, is beneath the stratum lucidum when present; otherwise it lies beneath the stratum corneum. This layer is one to five cells thick and is so named because of the granules present in the keratinocytes of this layer. The cells of the stratum granulosum have not yet been compressed into a flattened layer and are diamond shaped. The structures contained in these cells are keratohyalin granules, which become intensely stained with the appropriate acid and basic dyes. The proteins contained in these granules, profilaggrin, intermediate keratin filaments, and loricrin, helps to organize the keratin filaments in the intracellular space. Cells in this layer still have active nuclei.

2.4 Stratum Spinosum

The stratum spinosum is the next layer below the stratum granulosum. This layer is often described as the prickly layer because cytoplasmic structures in these cells take on a spicule morphology. Generally, the cells of this layer are polyhedral in shape. A prominent feature of the prickle layer is desmosomes, a type of cell–cell junction. Cells in this layer begin to synthesize involucrin, a soluble precursor of the cornified envelopes. The spinous cells contain large bundles of newly synthesized keratin filaments, K1/K10 in addition to K5/K14 still present from the basal layer.

2.5 Stratum Basale, or Stratum Germinativum

The stratum basale, or stratum germinativum, is the innermost epidermal layer. It is often referred to simply as the *basal layer*. It is a single layer of mitotically active cells called *basal keratinocytes* or *basal cells*. These active cells respond to several factors, such as extracellular matrix (ECM), growth factors, hormones, and vitamins. Skin metabolism is mediated by glucose. Glucose utilization in the skin is comparable to muscle. Glucose leaving the circulation crosses the basement membrane and forms a concentration gradient that decreases as it moves to the upper layers of the epidermis.

Once cells leave the basal layer they begin an upward migration, which can take 2 to 3 weeks. It takes approximately 14 days for a cell to move to the stratum corneum and another 14 days to move through the stratum corneum and desquamate. After leaving the basal layer the cells begin the process of differentiation. All layers of the epidermis consist of peaks and valleys. This arrangement is more dramatic in the basal layer. In fact, the epidermal protrusions of the basal layer that point downward into the dermis are called *rete ridges* or *rete pegs*. Rete ridges are partly responsible for anchoring the epidermis, thus providing structural integrity. Basal cells in the base of the rete ridges have an increased proliferative capacity compared with cells at the top of the ridges.

Epidermal stem cells comprise about 10% of the basal cell population. These stem cells have a slow cell cycle and when their DNA is labeled with a radioactive nucleic acid, they retain the label for long periods. Hence, these stemlike cells are called *label retaining cells*. Once cell division occurs, the daughter cell that

will undergo differentiation as it moves toward the stratum corneum is called the *transient amplifying cell*. These transient amplifying cells make up about 50% of the basal keratinocyte population. A portion of the stem cell population is found in the epidermal crypts of the rete pegs and in the bulge region of hair follicles. Dividing cells go through the cell cycle and the G_1 phase is shortened in states such as wound healing. The normal keratinocyte cell cycle time is 300 h, but it may be as short as 36 h in psoriasis. These stem cell compartments and the transient amplifying cells are both capable of limited or continuing cell division, and thus, are the cells most likely to reside long enough in the skin to undergo genetic modifications leading to the development of skin cancers. The use of stem cells alone and in combination with tissue engineering approaches is an area of active research and development.

Also distributed in this layer are melanocytes, the cells responsible for skin pigmentation. These are dendritic cells arising from the neural crest that synthesize melanin. Melanocytes can be detected at 50 days gestation in fetal development. Under normal conditions melanocytes rarely divide. In normal skin, the number of melanocytes present is nearly the same regardless of skin color. There is approximately one melanocyte for every 36 basal cells. Dendritic melanocyte structures are responsible for the transfer of pigment to a large number of keratinocytes. The primary difference between light- and dark-skinned individuals is the size and distribution of the melanosomes, the structures containing the melanin pigment, and the activity of the melanocytes. Carotene or carotenoids are responsible for imparting the yellow hue to the skin of some individuals.

2.6 Basement Membrane Zone

The basement membrane zone (BMZ), or dermal–epidermal junction (DEJ), is the area that separates the epidermis from the dermis. The BMZ anchors the epidermis to the dermis and is the layer that is affected in blister formation. During wound healing the BMZ is disrupted and must be reformed. Closer examination of the BMZ in the past decade has revealed it to be more complex than previously believed. Basal keratinocytes use hemidesmosomes to structurally and functionally attach to the BMZ. The BMZ is subdivided into two distinct zones – lamina lucida and lamina densa. The lamina lucida is so named because it is an electron-translucent zone compared with the electron-dense zone of the lamina densa. The major proteins found in the BMZ are fibronectin, an adhesive glycoprotein; laminin, also a glycoprotein; type IV collagen, a nonfiber-forming collagen; and heparan sulfate proteoglycan, a glycosaminoglycan that probably acts as a type of ground substance. A lesser amount of type VII collagen is present, but it plays a very important role in attachment of cells to the lamina densa of the BMZ. Type VII collagen molecules form antiparallel dimers with the associated C termini linked by disulfide bonds. The N termini are linked into the basement membrane and anchoring plaque of epidermal cells in the stratum basale. Mutations of the type VII collagen gene that disrupt the proper association of the dimers leads to the dystrophic form of epidermolysis bullosa, a family of disorders characterized by separation of the epidermis from the dermis by relatively minor shear forces or trauma, resulting in blister-like injuries. This family of disorders, most of which are inherited, range in severity from mild to the severely

disabling and life-threatening diseases. In severe epidermolysis bullosa, blisters are not confined to the outer skin. They may develop inside the body, in such places as the linings of the mouth, esophagus, stomach, intestines, upper airway, bladder, corneal epithelial layer, and the genitals. This rare genetic disorder affects all ethnic and racial groups. Estimates indicate that as many as 100 000 Americans suffer from some form of epidermolysis bullosa, and depending on the severity, have significantly shortened life expectancy.

3 Dermis

The dermis, or corium, is the thickest tissue layer of the skin. Compared with the cellular epidermal layer, the dermis is sparsely populated primarily by fibroblast cells and is vascularized and innervated. The dermal layer is derived from the middle embryonic germ layer, the mesoderm. Dermal thickness ranges from 2 to 4 mm but on average is 2 mm. Variations in dermal thickness account for differences in total skin thickness that have been measured throughout the body. The dermis of the back is thicker than the dermis covering the scalp, forehead, abdomen, thigh, wrist, and palm.

3.1 Papillary Dermis

The papillary dermis lies immediately below the BMZ and forms interdigitating structures with the rete ridges of the epidermis called *dermal papillae*. The dermal papillae contain capillary loops, which supply the necessary oxygen and nutrients to the overlying epidermis via the BMZ. The collagen fibers contained in the papillary dermis are much smaller in diameter and form smaller, wavy, cable-like structures compared with the reticular dermis. This portion of the dermis also contains small elastic fibers and has a greater proportion of ground substance than the reticular dermis.

3.2 Reticular Dermis

The reticular dermis is the area below the papillary dermis and forms the base of the dermis. The collagen fibers in this layer are thicker in diameter and form larger cable-like structures. There is no clear separation of papillary and reticular dermis because the collagen fibers change in size gradually between the two layers. Thicker elastic fibers are found in the reticular dermis but substantially less ground substance is present. A complex of cutaneous blood vessels is also found in this part of the dermis.

3.3 Dermal Vasculature

The dermal vasculature consists of a network of papillary loops, supported by a deep horizontal plexus. The vasculature functions to provide nutritional support, immune surveillance, wound healing, thermal regulation, hemostasis, and inflammation. The complete formation of the vascular system involves both vasculogenesis and angiogenesis. Vasculogenesis is the process that occurs *de novo* during embryonic development, and this process is mediated by angioblasts that are derived from embryonic mesodermal cells. Angiogenesis is the process whereby new vessels are formed from the vascular endothelial cells of preexisting vessels. Angiogenesis is the process involved in wound

healing, tumor growth and metastasis, hemangiomas, telangectasia, psoriasis, and scleroderma. Until recently, angiogenesis during wound healing was thought to only involve proliferation of vascular endothelial cells from capillaries around the wound, but recent data indicate that some of the cells that form new capillaries in wounds are actually angioblasts that originate in bone marrow, circulate to the wound and become incorporated into the new capillaries. Vascular endothelial growth factor (VEGF), also known as *vascular permeability factor* (VPF) is secreted by keratinocytes in response to hypoxia and is a major factor responsible for the stimulation of angiogenesis. VEGF/VPF is known to increase during wound healing. Other factors known to stimulate angiogenesis include acidic and basic fibroblast growth factor (FGF), interleukin-8, platelet-derived endothelial growth factor, placental growth factor, transforming growth factor α and β, oncostatin M, angiogenin, and heparin binding epidermal growth factor. Inhibitors of angiogenesis include angiostatin, thrombospondin, endostatin, interferon-γ, interleukin-12, and platelet factor 4.

3.4 Extracellular Matrix Proteins

3.4.1 Collagen

Collagen is the major structural protein found in the dermis and is secreted by dermal fibroblasts as tropocollagen (Fig. 2). After additional extracellular processing, mature collagen fibers are formed. Normal human dermis is primarily composed of type I collagen, a fiber-forming collagen. Type I collagen represents about

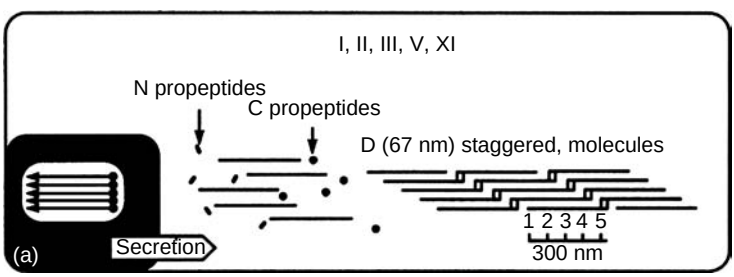

Fig. 2 Different types of collagen molecules. All collagen molecules contain some segments of triple stranded helix. (a) The fibrillar collagen molecules (types I, II, III, V, and XI) assemble in a head-to-tail and staggered, side-by-side alignment and are the most abundant types of collagens. They provide high tensile strength to the extracellular matrix. (b) The network-forming collagen molecules (types IV, VIII, and X) assemble in head-to-head and head-to-tail alignments, which generates a sheetlike structure with "pores" in basement membranes. (c) The fibril-associated collagen molecules with interrupted triple helices (FACITs) bind with collagen fibrils, but also bend at the interrupted helix hinge region, and associate with other molecules in the surrounding extracellular matrix. (d) Type VI collagen molecules retain globular regions at the N and C termini which form "beads on a chain" like appearance when they associate into tetramers. (e) Type VII collagen molecules also retain globular regions at the N and C termini, associate in tail-to-tail configuration, while the N terminus associates with basement membrane and anchoring plaque regions of epidermal cells.

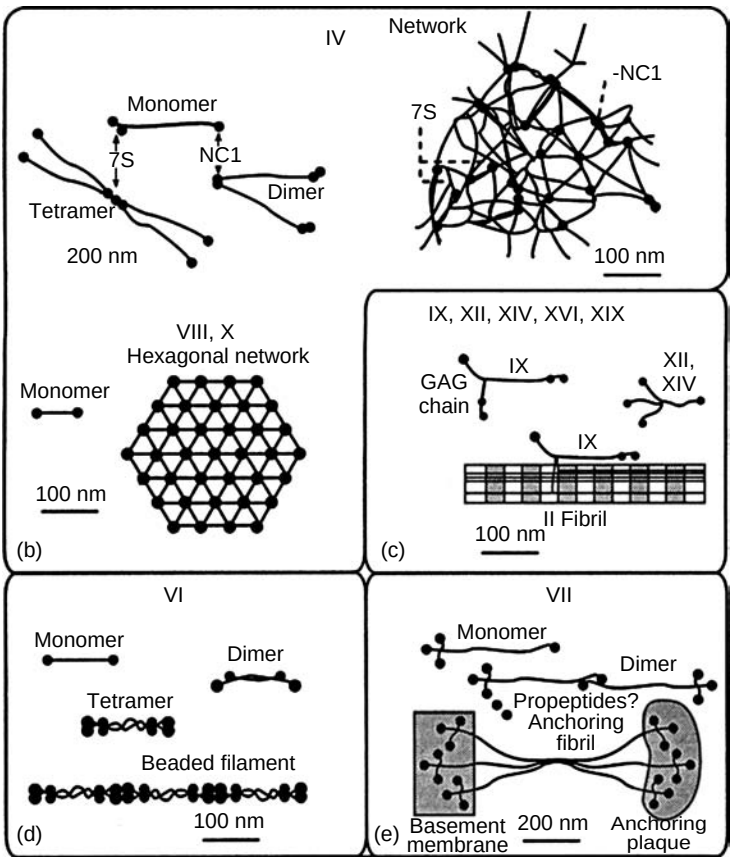

Fig. 2 (*Continued*)

77 to 85% of the collagen present, and type III collagen, also a fiber-forming collagen, represents the remaining 15 to 23%. Types V and VI collagen are also found in small amounts. Associated with the major fiber-forming collagens are types IX and XII collagens. These fibril-associated collagens do not self-associate into large self-polymers, but instead monomers bind to the sides of collagen fibers and extend a segment away from the fiber into the surrounding matrix that binds to proteoglycans (PGs). Thus, these fibril-associated collagen molecules help to bind collagen fibers with the other components of the ECM. Collagen is the protein that gives the skin its tensile strength. Chemically processed collagen from bovine sources results in leather handbags valued for their strength and long life. The primary amino acids that constitute collagen are proline, glycine, hydroxyproline, and hydroxylysine.

3.4.2 Elastin

Elastin, another protein found in the dermis, provides the skin with its elastic recoil. This prevents the skin from being permanently reshaped. Elastin is a fiber-forming

protein like collagen and has a high amount of proline and glycine. However, unlike collagen, elastin lacks large amounts of hydroxyproline. Elastin fibers form structures similar to a spring or coil that allow this protein to be stretched and, when released, to return to its inherent configuration. It accounts for less than 2% of the skin's dry weight. Other cells found in the dermis are mast cells, macrophages, and lymphocytes. All of these cells are involved with immune surveillance of the skin, often referred to as the *skin immune system* (SIS).

3.4.3 Proteoglycans and Glycosaminoglycans

The major proteins found in the dermis are collagen and elastin. The other category of proteins found occupying the space between collagen and elastin fibers is referred to as the *ground substance*. This category of proteins is largely composed of proteoglycans and glycosaminoglycans (GAGs). Included in this category of proteins are chondroitin sulfates and dermatan sulfate (versican, decorin, biglycan), heparan and heparan sulfate proteoglycans (syndecan, perlecan), and chondroitin-6 sulfate proteoglycans. Although these proteins only account for about 0.2% of the dry weight of the dermis, these large molecules are capable of binding up to 1000 times their volume. Thus, PGs and GAGs play a role in regulating the water-binding capacity of the dermis that can determine dermal volume and compressibility. Hyaluronan is also found in the dermis and its higher abundance in fetal skin where it forms a watery less stable matrix allows greater cell movement and is a factor in scarless healing. This material can bind growth factors and provide cellular linkages with other matrix materials. Fibroblasts are the cells distributed in this layer that synthesize and secrete these proteins. The dermis is a matrix supporting the epidermis and can be divided into two areas – papillary dermis and reticular dermis.

3.5 Hypodermis

Hypodermis, or superficial fascia, forms a subcutaneous layer below the dermis. This is an adipose layer containing a subdermal plexus of blood vessels giving rise to the cutaneous plexus in the dermis, which in turn gives rise to the papillary plexus and loops of the papillary dermis. The hypodermis attaches the dermis to underlying structures. This layer provides insulation for the body, a ready reserve of energy, and additional cushioning, and it also adds to the mobility of the skin over underlying structures. In certain pathologic disease states such as Werner's syndrome and scleroderma, this layer is largely absent.

4 Skin Functions

The skin provides protection against aqueous, chemical, and mechanical assaults; bacterial and viral pathogens; and ultraviolet radiation (UVR). It also prevents excessive loss of fluids and electrolytes to maintain the homeostatic environment. The effectiveness of the skin in preventing excessive fluid loss can be seen in burn patients; patients with burns involving 30% of their body can lose up to 4.1 L of fluid compared with 710 mL for a normal adult. The skin maintains the internal milieu and a progressive decrease in water content across the epidermal layers helps prevent excessive transepidermal water loss under basal conditions. The basal layer is

about 65 to 70% water decreasing to 40% in the granular layer and progresses to the stratum corneum that is 15% water. Skin barrier function is so effective that percutaneous drug delivery is limited and in general compounds are restricted to those with a molecular mass of 500 Da or less with a daily dose around 10 mg or less. Protection against mechanical assaults is mainly provided by the tough fibroelastic tissue of the dermis, collagen, and elastin. Collagen, the most abundant protein in mammals, represents 25% of total weight and provides tensile strength, which makes the skin resistant to tearing forces. Elastin is distributed with collagen but in smaller amounts. Large concentrations of elastin are present in blood vessels, especially the aortic arch near the heart.

4.1 Protection Against Pathogens

Protection against aqueous, chemical, bacterial, and viral pathogens is provided by the stratum corneum, secretions from the sebaceous glands, and the SIS. The primary line of defense against all of these agents is an intact stratum corneum. As mentioned previously, the insoluble protein keratin found in the horny layer provides good resistance. In addition, the constant shedding of squames from the stratum corneum prevents the entrenchment of microorganisms.

Sebum, a lipid-rich oily substance secreted by the sebaceous glands onto the skin surface, usually via hair follicles and shafts, provides an acidic coating with a pH ranging from 4 to 6.8 and a mean pH of 5.5. This acidity and natural antibacterial substances found in sebum retard the growth of microorganisms. These glands are stimulated by sex hormones (androgens) and become very active during adolescence. Sebum, along with keratin, provides resistance to aqueous and chemical solutions. When sebaceous glands occur in association with hair follicles, they are called a *pilosebaceous unit*. Sebaceous glands are not found on palms or soles and occur in areas that lack hair, such as the lips. These glands are largest on the face and when associated with hair follicles. Sebaceous glands may increase in size by 100 to 150 times as the sebum accumulates. Maximum secretion occurs in the late teens to early twenties. Rates of secretion are higher in men and decline 32% per decade in females compared to 23% per decade in males.

Resistance to pathogenic microorganisms is also provided by normal skin flora through bacterial interference. Conceptually, there are two categories of skin flora – resident (the bacteria normally found on a person) and transient (bacteria that are not normally found on a person and are usually shed by daily hygienic practices, such as bathing and hand washing). Resident bacteria are found on exposed skin; moist areas such as the axilla, perineum, and toe webs; and covered skin. Bacterial microcolonies are found in hair follicles and at the edges of squames as halos in the upper loose surface layers. The following species of bacteria are found in human skin: Staphylococcus, Micrococcus, Peptococcus, Corynebacterium, Brevibacterium, Proprionibacterium, Streptococcus, Neisseria, and Acinetobacter. The yeast Pityrosporum and the mite Demodex are also found. Not all species are found on any one individual, but most carry at least five of these genera. Normal viral floras are not known to exist. However, viruses have been detected in compromised skin and in individuals who are not immunocompetent.

There is an association between skin pH and bacteria, for example, *Propionibacterium acnes* grows well at pH values such as 6.0 and 6.5, but there is a marked decrease in growth at pH 5.5. However, it appears that this relationship can vary depending on pH and the specific bacterial species involved and body location. More recently the formation of bacterial biofilms is being scrutinized, and the role of normal skin flora on the formation of these biofilms is an area of active investigation. Biofilms are essentially an extracellular polysaccharide matrix or glycocalyx in which microorganisms are embedded. These biofilms are made up of mixed bacterial species living in their own microniche in a complex, metabolically cooperative microbial community that maintains its own form of homeostasis and rudimentary circulatory system. Biofilms are resistant to host immune responses and are markedly more resistant to antibiotic and topical bactericidal agents. Reports indicate that biofilm cells can be at least 500 times more resistant to antibacterial agents. The formation of these biofilms are familiar to clinicians where they can be found on endotracheal tubes, Hickman catheters, central venous catheters, contact lenses, and orthopedic devices. *Pseudomonas aeruginosa* is known to form biofilms in conjunction with other bacterial species. Quorum sensing is a feature of the bacteria in biofilm formation. These biofilms complicate the eradication of infections because they give rise to sessile and planktonic bacteria. Planktonic bacteria can be cleared by phagocytosis, antibodies, and antibiotics. The sessile bacteria in biofilms can evade antibiotics by giving rise to planktonic bacteria that responds to antibiotics and the host immune responses while the sessile bacteria remain. Many times cycles of antibiotic treatment are instituted to no avail because the symptoms of infection recur. In these situations, biofilm formation should be suspected and the surgical removal of the sessile population will most likely be required to eliminate the pathogenic bacteria. Moist skin areas such as the axilla, perineum, toe webs, hair follicles, nail beds, and sweat glands, are especially prone to an increased presence of bacteria. Protection from bacterial invasion is in part mediated by proteins called *defensins*. Findings show that β-defensin is abundant in skin and may be important for wound healing. Defensins act in conjunction with phagocytosing neutrophils and the membrane attack complex of complement. Thus, optimal skin and wound care is a cornerstone of clinical practice to prevent the progression from bacterial colonization to infection.

4.2 Protection Against Ultraviolet Radiation

Protection against UVR is provided by skin pigmentation, which results from synthesis of the pigment melanin. Harmful effects are primarily attributable to the longwave form of UVR, or UVA, which ranges spectrally from 320 to 400 nm and UVB or shortwave ultraviolet ranging from 290 to 320 nm. The shorter the waves, the more dangerous they become. UVC is effectively blocked by an intact ozone layer. With holes appearing in the ozone layer there is concern over the effects that this may have on skin diseases caused by UVR. Because of the increased synthesis, amount, and distribution of melanin in dark skin, these individuals are better protected against skin cancer. Melanin is distributed in all layers of the epidermis in dark skin in contrast to light skin. Exposure to UVR can lead to skin cancer, sunburn (first- or second-degree burns),

compromised immunity, and long-term skin damage. Sun exposed epidermal skin is thicker and rougher than unexposed skin. It is more prone to benign and malignant growths and to have a greater decrease in Langerhans cells. Sun exposed dermis has increased elastic properties probably due to increased synthesis of matrix metalloproteinases (MMPs), which cause an increase in fragmented collagen fibrils compared to unexposed dermis.

4.3 Skin Immune System

The SIS also provides protection against invading microorganisms and antigens. The cells of the skin that provide immune protection are the Langerhans' cells, an antigen-presenting cell found in the epidermis; tissue macrophages, which ingest and digest bacteria and other substances; mast cells, which contain histamine (released in inflammatory reactions), and dendrocytes. Both macrophages and mast cells are found in the dermis. Langerhans cells are responsible for the recognition, uptake, processing and presentation of soluble antigens, and haptens to sensitized T lymphocytes. This occurs through the binding of T cells to Langerhans cells. Exposure to UVB light decreases the functional capability of the Langerhans cells.

Tissue macrophages are derived from monocytes, which are derived from bone marrow precursor cells. Macrophages are among the most important cells of the SIS because they are versatile. Once these monocytes migrate into the tissue they differentiate and become macrophages. Cells in the dermis that are not completely differentiated are difficult to distinguish, and much effort has been made to recognize the various cells in the dermis in the last decade. Macrophages, in addition to their antibacterial activity, can process and present antigen to immunocompetent lymphoid cells; are tumoricidal; and can secrete growth factors, cytokines, and other immunomodulatory molecules. Macrophages are involved in coagulation, atherogenesis, wound healing, and tissue remodeling.

Mast cells are usually found distributed in the papillary dermis; around epidermal appendages, blood vessels, and nerves found in the subpapillary plexus; and in subcutaneous fat. These cells are distributed in connective tissue throughout the body in places where there is an interface of an organ with the environment. Mast cells contain or secrete on demand a host of proteins. Thus, mast cells are the primary effector cells in an allergic reaction. They are also involved in the presence of subacute and chronic inflammatory disease. Increased numbers of mast cells have been detected in chronic eczema, psoriasis, scleroderma, porphyria cutanea tarda, lichen simplex chronicus, lichen planus, and healing wounds. As a part of the SIS, mast cells play a role in protection against parasites; stimulate chemotaxis; promote phagocytosis; are involved in the activation and proliferation of eosinophils; are capable of altering vasotension and vascular permeability; and can promote connective tissue repair and angiogenesis. Their role in tumor surveillance awaits further study.

Dermal dendrocytes are highly phagocytic cells found in the papillary and upper reticular dermis. They are also distributed near vessels in the subpapillary plexus, reticular dermis, and subcutaneous fat. These are immunologically competent cells that are highly phagocytic and can be recognized as melanophages. In fetal, infant, photoaged, and pathological skin conditions, such as psoriasis or eczema

their number is increased. In malignant fibrotic tumors and fibroproliferative lesions such as keloids, scars, and scleroderma, their numbers are decreased.

4.4 Thermoregulation

Thermoregulation of the body is provided by the skin, which acts as a barrier between the outside and inside environments to maintain body temperature. The two primary thermoregulatory mechanisms are circulation and sweating. Blood vessels can either dilate to dissipate heat or constrict to shunt heat to underlying body organs. When dilated, these vessels have an increased blood flow and release heat by conduction, convection, radiation, and evaporation. Vasoconstriction is often accompanied by actions of the arrector pili muscle attached to hair follicles, which results in the hair standing vertically. In mammals that depend on hair for warmth, this action fluffs up the fur to increase thermal capacity. The bulge around the hair shaft that is visible when this occurs is commonly referred to as goose bumps. In humans, shivering is more important for maintaining body temperature when the outside environment is cold than the vertical orientation of hair. In cold weather, the "core" body temperature encompasses a smaller zone, whereas in warm weather the "core" is expanded. Sensations of cold and warm are generally detected below 30 °C and above 37 °C. Clinically it has been estimated that for each 1 °C (1.8 °F) increase in fever there is a 13% increase in a patient's fluid and calorie needs. At rest, the trunk, viscera, and brain account for 70% of heat production but are only 36% of body mass. However, during exercise, muscle and skin account for 90% of heat production while representing 56% of body mass.

Hair follicles, unlike other skin structures have a repeated cycle of growth and regression. Follicle development begins in the fourth month of gestation. The hair growth cycle consists of anagen (growth phase), catagen (follicle involution), and telogen phases (dormant or resting phase). A population of epidermal stem cells is found in the bulge region of the hair follicle, and this cell population contributes to epithelialization of partial-thickness wounds. Hair shape varies depending on ethnicity and location. Asians have the largest diameter scalp hair and Caucasians the smallest. In the scalp, about 85% of hair is in anagen phase and 15% in telogen phase with the anagen phase lasting from 2 to 5 years. On the extremities, anagen lasts from 22 to 28 days.

Sweating occurs when there is an increase in the activity of the sweat glands. It has been estimated that there are about two to five million sweat glands. Sweat glands are of two types – eccrine and apocrine. Eccrine glands arise from epidermal invagination and are found abundantly on the palms of the hand and soles of the feet. There are anywhere from 2 to 5 million glands, 0.05 to 0.1 mm in size. These glands are largely under the control of the nervous system, responding to temperature differences and emotional stimulation. Muscular activity also influences their secretory activity. The sweat glands, located in the dermis as a coil, secrete fluid that is 99.0 to 99.5% water with the remainder consisting of sodium chloride, urea, sulfates, and phosphates. The pH is slightly acidic. Thermoregulatory control occurs as a result of cooling when fluid is evaporated from the skin surface, since such evaporation requires heat. The odor associated with sweat is

largely a result of bacterial action. Eccrine sweat glands are capable of producing 1 to 4 L h^{-1} resulting in a 75 to 90% reduction in body heat. Man is capable of losing heat more rapidly and for longer periods than any other animal.

Apocrine sweat glands are usually found in association with hair follicles but do not play a significant role in thermoregulation. These coiled tubular glands are present in the axilla and anogenital area; modifications of these glands are found in the ear and it secretes ear wax, or cerumen. There are approximately 100 000 apocrine glands each 2 to 3 mm in diameter. Secretions from these glands are small in amount, turbid, and contain iron, carbohydrates, and lipids.

4.5 Sensation

Nerve receptors located in the skin are sensitive to pain, touch, temperature, and pressure. Nerve fibers are located in the dermis and throughout the epidermis. Nerve structures found in the skin originate from the neural crest and are detectable in the developing embryo around 5 weeks. When stimulated, these receptors transmit impulses to the cerebral cortex where they are interpreted. Combinations of these four basic types of sensations result in burning, tickling, and itching. These sensations are propagated by unmyelinated free nerve endings, Merkel's cells, Meissner's corpuscles, Krause's end bulbs, Ruffini's terminals, and Pacini's corpuscles. Identification of particular responses with specific nerve structures has not been successful. In part, the reason is that some receptors seem to respond to a variety of stimuli. However, it is known that Meissner's corpuscles are involved in touch reception; Pacini's corpuscles respond to pressure, coarse touch, vibration, and tension; and free nerve endings respond to touch, pain, and temperature. Skin sensation is a part of the body's integrative response to protect itself from the surrounding environment. Sensation assists with the skins regulatory function and can signal sweating, shivering, weight shifts, laughter, and scratching.

Merkel cells function as mechanoreceptors and in the epidermis these cells produce nerve growth factor, whereas in the dermis, Merkel cells express receptors for nerve growth factor. Merkel cells are found around the arrector pili muscle in the bulge region of hair follicles and contribute to the development of eccrine sweat glands, nails, and nerves of the skin.

Sensation also moderates psychobiologic phenomena made famous by Harlow, who demonstrated the preference of young animals for warm objects and those that provided better tactile sensitivity. In addition, early studies by Spitz in 1947 pointed to the importance of touch in mediating social interactions with children and infants. Deprivation of touch can lead to psychomotor retardation and an increased risk of death. Other studies on touch indicate that it may be a factor in the development of atherosclerotic lesions in animals. Studies examining the effect of stroking, handling, talking, and playing with rabbits reported a 60% reduction in aortic atherosclerotic lesions in treated animals compared with controls, even though both groups were fed the same cholesterol-containing diet and had similar blood pressures, heart rates, and serum cholesterol levels.

In healing wounds, sensory nerves sprout abundantly for about the first three weeks subsequently returning to their normal density. The role of neuropeptides to stimulate the growth of connective tissue

is increasingly being recognized. This includes a role for these neuropeptides in modulating matrix production by fibroblasts and in acting as growth factors for keratinocytes.

4.6 Vitamin Metabolism

Synthesis of vitamin D occurs in the skin in the presence of sunlight. UVR converts a sterol (7-dehydrocholesterol) to cholecalciferol (vitamin D). Vitamin D participates in calcium and phosphate metabolism and is important in the mineralization of bone. Because vitamin D is synthesized in the skin but then transmitted to other parts of the body, it is considered an active hormone when converted to calcitriol (1,25-dihydroxycholecalciferol).

4.7 Social Communication

In addition to its biologic, structural, functional, and physiologic functions, human skin also functions as an organ of communication and identification. The skin over the face is especially important for identification of a person and plays a role in internal and external assessments of beauty. Injury to the skin can result in not only functional and physiologic consequences but also changes in body image. Scarring from trauma, surgery, or incisions can lead to changes in clothing choices, avoidance of public exposure, and a decrease in self-esteem. Research indicates that with increased scarring from facial acne, the self-image is progressively reduced. Adolescents are especially sensitive to physical appearance. As an organ of communication, facial skin along with underlying muscles is capable of expressions such as smiling, frowning, and pouting. The sensation of touching can also convey feelings of comfort, concern, friendship, and love.

5 Factors Altering Skin Characteristics

5.1 Age

Age is an important factor in altering skin characteristics. More recently the scarless healing of fetal tissue has come under more intense investigation. It has been found that wounds heal without scarring in fetal lambs up until 120 days of gestation. Collagen deposition in fetal wounds occurs more rapidly and in a normal dermal pattern. The ratio of type I and type III collagen is different in fetal skin, and about 30 to 60% of the collagen is type III while in adult skin it is about 10 to 15%. In addition, an important difference between fetal and adult skin is the amount of hyaluronic acid, a glycosaminoglycan. In the laboratory setting, topical application of hyaluronic acid has been associated with a reduction in scar formation in postnatal wounds. This glycosaminoglycan is associated with collagen, and it has been proposed that a hyaluronic acid–collagen–protein complex plays a role in fetal scarless healing. Transforming growth factor β-1 (TGF-β1) is the other modulator of scarless healing in fetal wounds. Fetal wounds usually contain less TGF-β1 compared with adult wounds. In addition, differential patterns of expression of the various isoforms of TGF-β also affect the relative amount of scarring in fetal versus adult wounds. Wounding in fetal skin also reveals differences in the cell responses with a decrease in inflammatory cells, particularly polymorphonuclear leukocytes and macrophages. Platelets do not aggregate in response to collagen or release the amount

of TGF-β and PDGF-AB like that in adult cells. Fibroblasts are able to migrate at a faster rate and synthesize more total collagen. Other differences in protease and inhibitor activity, level of growth factors, and expression of homeobox genes has been detected.

5.1.1 Neonatal Skin

At birth the skin and nails are thinner than those in an adult but will gradually increase in thickness with aging. Formation of the epidermal and dermal layers occurs within the first 2 weeks of embryonic development. Epidermal development is complete by the end of the second trimester, and at birth epidermal thickness is almost that of adult skin, although newborn skin is not as effective as adult skin in providing an effective barrier to transcutaneous water loss. On the other hand, development of the dermis lags behind and does not take on the characteristics of adult dermis until after birth.

The ratio of type I to III collagen is similar to the fetal ratios. Soluble collagen is about 24%, compared with 1% in the adult. This persists until about 6 months of age. The newborn dermis is about 60% as thick as that found in an adult, and as expected, the dermal fibers are significantly finer. Newborn dermis contains a much higher cellular component compared with mature adult skin. The epidermal–dermal junction remains flat until the beginning of the third trimester, and at birth, the rete ridges are only weakly developed, thus making premature and newborn skin subject to tearing or blistering. At birth the capillary beds do not have a mature adult pattern and are still disorganized. An adult pattern of capillary loops occurs at about 14 to 17 weeks after skin growth slows.

Immature skin, or skin from premature infants between 23 and 24 weeks of gestation up to 32 weeks of gestation, requires special attention compared with that of infants beyond 32 weeks. In particular, before 28 weeks gestation the skin is thin and poorly keratinized and functions weakly as a barrier. The skin of infants born at the limits of viability has been characterized as more suitable to an "aquatic environment" than to atmospheric conditions. Transepidermal water loss is high, and application of tape strips the outer immature epidermal layer that can leave behind raw damaged skin prone to infection and occasional scarring. At 24 weeks' gestation, transepidermal water loss can be 10 times greater per unit area compared with an infant born at term. Infants born between 22 and 25 weeks may require up to 4 weeks to develop a functional stratum corneum. In addition, premature infants have high evaporative heat losses resulting in increased risk for hypothermia.

Because premature infants have a greater surface-area-to-volume ratio compared with full-term infants, they are at an increased risk for skin complications and systemic toxicity from topically applied agents. These infants may also have alterations in metabolism, excretion, distribution, and protein binding of chemical agents, placing them at increased risk for local or systemic toxicity from soaps, lotions, or other topical agents. Other dangers are percutaneous absorption of topical agents, including antiseptics. Hemorrhagic necrosis of the dermis from alcohol absorption has been reported if the alcohol does not quickly evaporate and is sometimes mistaken for bruising. The use of topical antibiotic sprays containing neomycin should be avoided since it is an ototoxic aminoglycoside.

Thus, water-based topical antiseptics are preferred but should be used sparingly. Cleaning should be done with care using normal saline or water. Chlorhexidine, a commonly used antiseptic, is not known to have any adverse effects, but it is probably absorbed from the skin and should be used judiciously. Likewise, iodine has been reported to be absorbed, leading to goiter and hypothyroidism. If required, moisturizing creams or ointments may be applied to dry, flaking, or fissured skin, and the best agents appear to be those with little or no preservatives, since these offer the greatest benefit with decreased risk.

5.1.2 Adolescent Skin

The next period of change occurs in adolescence when hormonal stimulation results in increased activity of sebaceous glands and hair follicles. Sebaceous glands increase their secretory rate, and hair follicles, giving rise to secondary sexual characteristics, become activated. From adolescence to adulthood there is a gradual change in skin characteristics. By the time the skin reaches mature adulthood, several changes become apparent. The dermis decreases in thickness by about 20%, whereas the epidermis remains relatively unchanged. Epidermal turnover time is increased; this means that wound healing may take longer. For instance, in young adults, epidermal turnover takes about 21 days, but by 35 years of age this turnover time is doubled. Barrier function is reduced, and such reduction may increase the risk of irritation. The number of active melanocytes per unit body surface area decreases with aging, which means that protection against UVR is diminished. Skin dryness is also associated with aging and an increase in wrinkles. Sensory receptors are diminished in capacity, meaning that the skin is more likely to be burned or traumatized without perception. Vitamin D production is decreased and may be a factor in osteomalacia.

5.1.3 Elderly Skin

With aging there is a decrease in the number of Langerhans' cells, which affects the immunocompetence of the skin and can lead to an increased risk of skin cancer and infection by invading microorganisms. The density of Langerhans' cells changes from 10.0 per 3 mm cross section of unexposed skin in 22–26-year-old individuals to 5.8 per 3 mm cross section in 62–68-year-old individuals. There is also a decrease in mast cells and melanocytes. The inflammatory response is also decreased, which may alter allergic reactions and healing. The thermoregulatory capacity of the skin is also compromised by decreased number of sweat glands, diminished vascularity, and reduced subcutaneous fat. Epidermal–dermal junction changes, such as the flattening of the prominent dermal papillae and of the rete ridges, alter junctional integrity. Consequently, the skin is more easily torn in response to mechanical trauma, especially shearing forces. Because the epidermal rete pegs flatten, the unique microenvironment of the basal keratinocytes changes; it is thought that this explains the decrease in epidermal proliferative capacity with aging.

Skin elasticity decreases with age and is related to a combination of aging and solar damage. Microscopic analysis of collagen and elastin fibers reveals that these are more compact with a loss of ground substance from the spaces between these cable-like structures. Collagen fibers appear to be unwinding, while elastin fibers appear to be undergoing digestion. The degradation of elastin can be detected at about 30 years of age but becomes

marked at 70 years of age. Changes in dermal proteoglycans usually occur after 40 years of age. By 70 years of age, most of the elastin network is affected. Changes in collagen content and structure are mediated by an underexpression of procollagen, overexpression of collagenase (MMP-1), stromelysin (MMP-3) and gelatinase A (MMP-2), and decreased expression of tissue inhibitors of matrix metalloproteinase-1 (TMP-1). There is also a marked reduction in vascular beds in the vertical capillary loops in the dermal papillae. There is an approximate 35% decrease in the cross-sectional area of these loops in aged skin. It is thought that this leads to atrophy of the hair bulbs, sweat glands, and sebaceous glands. Because the hypodermis also becomes thinner, mature individuals are more prone to pressure necrosis. With aging there is a progressive loss of mechanoreceptors to one-third of their average density that occurs from the second to ninth decade.

Clinical observations established that repair of injuries to skin of aged persons (over 50 years old) proceeded much more slowly than injuries in skin of young persons, but with improved quality of scarring both microscopically and macroscopically. However, the molecular changes that caused these effects were not known until a series of experiments by Ashcroft and colleagues revealed that the slower healing and reduced scarring in postmenopausal women were associated with reduced levels of TGF-β1 protein and mRNA in wounds. Furthermore, treatment of postmenopausal women (or aged men) either with oral or topical estrogen, accelerated healing of skin wounds and elevated levels of TGF-β1 while simultaneously reducing levels of MMPs and elevating levels of the tissue inhibitors of metalloproteinases (TIMPs). Estrogen treatment also reduced inflammation in test wounds and reduced levels of neutrophil elastase, which was associated with increased levels of collagen and fibronectin in wound biopsies, resulting in the enhanced strength of wounds 80 days after injury. These data demonstrated that delays in wound healing in the elderly could be significantly diminished by topical estrogen in both male and female subjects.

Changes in hair color and hair follicles also accompany aging. Age-related changes in active melanocytes result in gray hair. About 50% of the body's hair will be gray by the age of 50 in about 50% of the population. This change is accompanied by a reduction in the number of hair follicles and a decrease in the diameter of the hair. The rate of hair growth is also decreased. Nail growth rates also decrease by 40 to 50%.

The density of skin melanocytes is relatively constant until about 40 years of age. By about 45 years of age, skin melanocytes have decreased in density to approximately half of that seen between 30 and 39 years of age. Melanocytes decrease 6 to 8% each decade after age 30. It is thought that loss of skin melanocytes contributes to an increase in the formation of skin cancers. Other overt changes are wrinkling and sagging, which occur as a result of loss of underlying tissue in addition to changes seen in collagen and elastin.

Changes in thermoregulatory capacity occur with age, and older individuals are more prone to hypothermia and heat stroke. This has been attributed to changes in blood capillaries and eccrine sweat glands. In healthy older individuals, sweating may be decreased by up to 70%. Sebum secretion also declines with age. Barrier function decreases with aging owing to the decreased level of all the

major lipid species, especially ceramides. In addition, corneocytes are larger and less cohesive. In addition to these changes, pain perception is dulled, and there is reduced skin reactivity upon exposure to irritants. Cutaneous immune function also changes with aging as seen by a reduction in Langerhans' cell density. Skin damaged by sun exposure, or actinically damaged skin, has been found to have a 50% reduction of Langerhans' cell density compared with sun-protected skin. Reduction in immunocompetence of the skin is thought to contribute in part to skin cancer in the elderly.

Other factors that may contribute to the development of skin cancer in aged individuals are cumulative exposure to carcinogens, diminished DNA repair capacity, decreased melanocyte density, and alterations in dermal matrix. Not surprisingly, wound healing in older individuals is delayed compared with that of younger individuals.

5.2 Sun

Excessive exposure to UVR can have harmful effects that accelerate aging of the skin. For this reason, the condition associated with UVR-damaged skin is referred to as photoaging. Dermatologically, it is called *dermatoheliosis*. Obvious clinical signs of photodamaged skin are dryness, tough leathery texture, wrinkling (as a result of collagen and elastin degeneration), and irregular pigmentation (from changes in melanin distribution). Excessive exposure to UV radiation increases the risk of developing skin cancers, such as basal or squamous cell carcinoma and malignant melanoma. Damage to the DNA of skin cells leads to transformation of cells and cancer. Changes also occur in epidermal and dermal cells; epidermal cells become thickened, fibroblasts become more numerous, and dermal vessels become dilated and tortuous. Langerhans' cells are reduced in number by about 50%, thereby diminishing the immunocompetence of the skin. Immediate short-term exposure to UVR can lead to sunburn. This type of red sunburn is the result of a vasodilatory response that increases blood volume. Whether an individual will become sunburned depends on the extent of skin pigmentation. Naturally, those with the least pigmentation are more prone to sunburn and the harmful long-term effects of UVR. Severe short-term exposure of unprotected, lightly pigmented skin can lead to blistering (a second-degree burn).

There is an association between melanoma and sunburn: if a patient has had more than six serious sunburns, then he or she is at an increased risk for melanoma. Exposure to UVR and the rise of malignant melanomas has led to the development of more effective sunblocking agents. Over time, the incidence of malignant melanoma has increased from 1/1500 people in 1930 to 1/250 in 1980. By the year 2000 the incidence is expected to be 1/150. Sunscreens should be used on a regular basis, be applied at least 30 min before exposure, and have a sun protection factor (SPF) ranging from 15 to 30. Individuals with moderately pigmented skin require about 3 to 5 times more exposure to UVR to induce sunburn inflammation compared with Caucasians, and for individuals with darker skin, 10 times more exposure is required.

5.3 Hydration

Adequate skin hydration is normally provided by sebum secretion and an intact

stratum corneum with its keratinized cells. Several factors can affect skin hydration. Among these are relative humidity, removal of sebum, and age. Each of these factors increases water loss from the skin, leading to dryness and scaling. Application of emollients to the skin replaces the barrier function of lost sebum or decreased evaporative water loss when the relative humidity is low. Retention of water in the epidermal layers after application of a lotion leads to swelling of the skin, which is perceived as smoothness and softness. Often various products are promoted with claims of superiority over others without adequate *in vitro*, *in vivo*, or clinical data. The superiority of oil baths over water baths was found to be only marginal. Twenty minutes after both kinds of bath, skin hydration was increased when measured by water evaporation, electrical conductance, and capacitance. A small but significantly greater amount of water was bound in the skin after the oil bath, whereas no change was seen in evaporation, conductance, or capacitance. Thus, increases in water-holding capacity of the skin after an oil bath may not be of importance. On the other hand, a difference was found in skin-surface lipids, which lasted at least 3 h. This effect is comparable with application of a traditional moisturizing lotion. The authors of this study conclude that because daily use of bath oil is not practical, application of moisturizing lotions may be more advantageous and that the beneficial effects of bath oils are related to lipidization of the skin surface.

5.4 Soaps

Washing or bathing with an alkaline soap reduces the thickness and number of cell layers in the stratum corneum. Generally, soap emulsifies the lipid coating of the skin and removes it along with resident and transient bacteria. Excessive use of soap or detergents can interfere with the water-holding capacity of the skin and may impair bacterial resistance. Use of alkaline soaps increases skin pH, which may change bacterial resistance. The time for recovery to normal skin pH of 5.5 depends on the length of exposure. Ordinary washing requires 45 min to restore skin pH, whereas prolonged exposure can require 19 h. Other agents that can lead to delipidization or dehydration of skin are alcohol and acetone. Currently, acidic skin cleansers appear to be less irritating than a neutral or alkaline one and there is some evidence to suggest that acidic cleansers decrease the number of acne lesions on the face.

5.5 Nutrition

Normal healthy skin integrity can be maintained by an adequate dietary intake of protein, carbohydrate, fats, vitamins, and minerals. Under normal conditions in healthy individuals, increased nutrition is not beneficial if there is adequate dietary intake. If the skin is damaged, increased dietary intake of some substances such as vitamin C for collagen formation may be beneficial. A healthy diet of protein breaks down to supply the necessary amino acids for protein synthesis. Fats are broken down into essential fatty acids, which can then be used by cells to form their lipid bilayer. Carbohydrates are digested to supply energy for cell metabolism. Vitamins C, D, and A; the B vitamins pyridoxine and riboflavin; the mineral elements iron, zinc, and copper; and many others are needed to maintain a normal

healthy skin. Adequate dietary intake can be ensured by ingestion of amounts consistent with the recommended daily allowances (RDA).

5.6 Medications

Various medications affect the skin. Some of the best studied are the corticosteroids, which are known to interfere with epidermal regeneration and collagen synthesis. Photosensitive and phototoxic reactions are also known to occur from medications. Some categories of medications that can affect the skin are antibacterials, antihypertensives, analgesics, tricyclic antidepressants, antihistamines, antineoplastic agents, antipsychotic drugs, diuretics, hypoglycemic agents, sunscreens, and oral contraceptives. Skin flora can be changed by the use of antibacterials, orally administered steroids, and hormones. Analgesics, antihistamines, and nonsteroidal anti-inflammatory drugs (NSAIDs) can alter inflammatory reactions. Thus, whenever drugs are prescribed and skin reactions occur, medications should always be examined to check whether they are responsible.

6 Wound Healing

Healing of skin wounds normally occurs in a predictable sequence of phases including: (1) hemostasis; (2) inflammation; (3) repair (characterized by mitosis, angiogenesis, and synthesis of ECM); and (4) remodeling of the scar matrix (Fig. 3). These processes are regulated by numerous molecules including growth factors, cytokines, chemokines, receptors, proteinases and their inhibitors.

6.1 Hemostasis

The first major event following a skin injury is hemostasis, and the platelets play key roles in this process as they undergo activation, adhesion, and aggregation. These events are triggered by interactions of platelets with both soluble mediators and matrix components. Platelets become activated when exposed to extravascular collagen (such as type I collagen), which they detect via specific integrin receptors in their plasma membrane. Platelet adhesion to collagen activates them to release soluble mediators (growth factors and cyclic AMP) and adhesive glycoproteins from their α-granules, which are in turn, signals for subsequent platelets to change their morphology, become sticky and to aggregate. The aggregated platelets become trapped in the fibrin web and provide the bulk of the clot, and their membranes act as a surface on which inactive clotting enzyme proteases are bound, activated, and then released to accelerate the clotting cascade. The key glycoproteins released from the platelet granules include fibrinogen, fibronectin, thrombospondin, and von Willebrand factor. Soluble mediators released from the α-granules include platelet-derived growth factor (PDGF), transforming growth factor β (TGFβ), transforming growth factor α (TGFα), basic fibroblast growth factor (bFGF), and VEGF. PDGF and TGFβ are chemotactic for neutrophils and monocytes and recruit them from the vasculature to initiate the inflammatory response. VEGF, TGFα, and bFGF act on endothelial cells to initiate angiogenesis. PDGF is also chemotactic for fibroblasts, which will migrate to the area to begin the process of collagen deposition. Thus, through the interaction with the ECM and release of

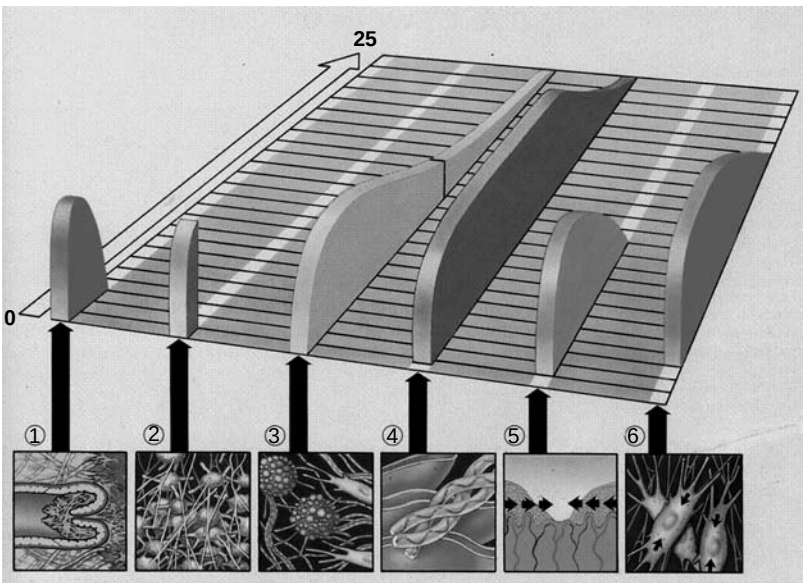

Fig. 3 Time sequences of normal wound healing. After an injury to skin, the vasculature responds (1) immediately by constriction of severed capillaries followed by increased permeability, which lasts for several days. The coagulation (2) cascade rapidly generates the fibrin clot, which forms the "provisional wound matrix" that is rich in growth factors released from granules of platelets that are trapped in the fibrin matrix. Inflammation (3) initially involves neutrophils that are draw to the injury by cytokines, followed a few days later by macrophages, which continue to destroy bacteria and release proteases to remove damaged components of the extracellular matrix. After one to two weeks, inflammation subsides unless bacterial contamination persists. Synthesis of new vasculature and extracellular matrix (granulation tissue) repairs the wounded dermal tissue (4), enabling epidermal cells (5) to proliferate, migrate, and resurface the injured area. Myofibroblasts (6) contract the new scar matrix, and eventually the scar remodels (late phase of 4).

soluble mediators, platelets not only contribute to hemostasis but their products also set the stage for the subsequent events of the healing process.

6.1.1 Provisional Wound Matrix

An injury that extends into the dermis of the skin causes disruption of the ECM and initiates the clotting cascade. Fibrin is generated at the site of injury when thrombin catalyzes conversion of the plasma-derived fibrinogen to fibrin. The fibrin clot acts to tamponade the bleeding and also to form a temporary scaffolding for the repair process. Serum-derived adhesion proteins, such as fibronectin and vitronectin, subsequently coat the fibrin network, forming a "provisional wound matrix" that plays a crucial role in cellular attachment through their interactions with integrin receptors. Fibronectin facilitates the movement of cells into the wound site from the vasculature (such as leukocytes migrating from the vasculature), as well as promoting migration of fibroblasts and

capillary cells through granulation tissue and epithelial cells moving over granulation tissue during resurfacing. The provisional wound matrix of fibrin/fibronectin also serves to bind and store growth factors, which are released by proteases (MMPs and serine proteases) as the provisional wound matrix is replaced by new collagen molecules in the proliferative/repair phase. Initially, type III collagen is synthesized by cells in the wound, which is slowly replaced over weeks and months by type I collagen molecules, which are found in normal skin. Furthermore, during the early phases of scar formation, the collagen molecules tend to be arranged in bundles that are orientated parallel to the edges of the wound rather than the more complex basket weave pattern of collagen bundles, which is characteristic of normal dermis. As the initial scar tissue remodels over many months, the architecture and composition of the scar matrix begins to resemble nonwounded skin, but never fully regenerates nonwounded ECM.

6.2 Inflammatory Phase

6.2.1 Neutrophils

Neutrophils are the first inflammatory cells to respond to the soluble mediators released by platelets and the coagulation cascade. They begin their journey from the vasculature by first adhering to the endothelial cell walls via their surface integrins and then releasing elastase and collagenase, which facilitate their migration through the basement membrane surrounding the endothelial cells and into the ECM at the wound site. Their primary role is to mount the first line of defense against infection by phagocytosing and killing bacteria, and by destroying foreign materials and devitalized tissue. They are able to interact with these extracellular targets via their integrin receptors. Neutrophils also produce and release inflammatory mediators such as TNFα (tumor necrosis factor) and IL-1 (Interleukin 1) that further recruit and activate fibroblasts and epithelial cells. Neutrophils produce and contain high levels of destructive proteases and oxygen free radicals, which they use to digest phagocytized materials. Upon their inevitable death, neutrophils release these substances into the local wound area where they can contribute to extensive tissue damage and prolong the inflammatory phase. The persistent presence of high levels of bacteria in a wound may contribute to chronicity through continued recruitment of neutrophils and their release of proteases, cytokines, and intracellular contents.

6.2.2 Monocytes/Macrophages

Usually, neutrophils are depleted in the wound after 2 to 3 days and are replaced by tissue macrophages. Macrophages begin as circulating monocytes that are attracted to the wound site beginning about 24 h after injury, by both soluble mediators and matrix degradation components such as fragments of collagen and fibronectin. Monocytes bind to the ECM via integrins and immediately differentiate into tissue macrophages. Serum factors and fibronectin mediate this differentiation. Tissue macrophages have a dual role in the healing process. They are voracious phagocytes and patrol the wound area ingesting bacteria, devitalized tissue, and depleted neutrophils. Macrophages also produce collagenases and elastase to assist them in breaking down devitalized tissues. They are able to regulate proteolytic destruction of tissue in the wound by producing and secreting inhibitors for these

enzymes. As important as their phagocytic role, macrophages also mediate the transition from the inflammatory phase to the proliferative phase of healing. They release a wide variety of growth factors and cytokines including TNFα, PDGF, TGFβ, TGFα, IL-1, IL-6, Insulin-like Growth Factor (IGF-1), and FGF. Some of these soluble mediators recruit and activate fibroblasts, which will synthesize, deposit, and organize the new tissue matrix, while others promote angiogenesis.

6.3 Proliferative Phase

Events of the proliferative or tissue building phase of healing include the synthesis, deposition, and organization of granulation tissue, ingrowth of new blood vessels and epithelial migration, proliferation, and stratification. In this phase of healing fibroblasts, macrophages, endothelial, and epithelial cells interact with matrix and with each other.

6.3.1 Granulation Tissue

Granulation tissue is a transitional replacement for normal dermis, which eventually matures into a scar during the remodeling phase of healing. It is characterized from unwounded dermis by its extremely dense network of blood vessels and capillaries, elevated cellular density of fibroblasts and macrophages, and randomly organized collagen fibers. It also has an elevated metabolic rate compared to normal dermis, which reflects the activity required for cellular migration, division, and protein synthesis.

6.3.2 Fibroblasts

The collagen, proteoglycans, and other components that comprise granulation tissue are synthesized and deposited primarily by fibroblasts. Fibroblasts migrate into the wound in response to multiple soluble mediators released by initially platelets and later by macrophages. Fibroblast migration in the ECM depends on precise recognition and interaction with specific components of the matrix. Fibroblasts in normal dermis are typically quiescent and sparsely distributed, whereas, in the provisional matrix of the wound site and in the granulation tissue, they are quite active and crowded together. Their migration and accumulation in the wound site requires them to change their morphology and to produce and secrete enzymes to clear a path for their movement from the ECM into the wound site.

Fibroblasts begin moving by first binding to matrix components such as fibronectin, vitronectin, and fibrin via their integrin receptors. Integrin receptors attach to specific amino acid sequences (such as R-G-D or arginine–glycine–aspartic acid) or binding sites in these matrix components. While one end of the fibroblast remains bound to the matrix component the cell extends a cytoplasmic projection to find another binding site, when the next site is found, the original site is released and the cell uses its cytoskeleton network of actin fibers to pull itself forward.

The direction of fibroblast movement is determined by the alignment of the fibrils in the ECM and provisional matrix and the gradient of chemotactic growth factors, cytokines, and chemokines. Fibroblasts tend to migrate along these fibrils as opposed to across them. Fibroblasts secrete proteolytic enzymes locally to facilitate their forward motion through the matrix. The enzymes secreted by the fibroblasts include three forms of the MMP family – collagenase (MMP-1) gelatinases (MMP-2 and MMP-9), which degrade gelatin substrates and

stromelysin (MMP-3), which has multiple protein substrates in the ECM.

Once the fibroblasts have migrated into the matrix, they again change their morphology, settle down and begin to proliferate and to synthesize granulation tissue components including collagen, elastin, and proteoglycans.

6.3.3 Angiogenesis

Damaged vasculature must be replaced to maintain tissue viability. The process of angiogenesis is stimulated by local factors of the microenvironment including low oxygen tension, low pH, and high lactate levels. Also, certain soluble mediators are potent angiogenic signals for endothelial cells. Many of these are produced by epidermal cells, fibroblasts, vascular endothelial cells and macrophages, and include bFGF, TGFα, and VEGF. It is now recognized that oxygen levels in tissues directly regulate angiogenesis by interacting with oxygen sensing proteins that regulate transcription of angiogenic and antiangiogenic genes. For example, synthesis of VEGF by capillary endothelial cells is directly increased by hypoxia through the activation of the recently identified transcription factor, hypoxia-inducible factor (HIF), which binds oxygen. When oxygen levels surrounding capillary endothelial cells drop, levels of HIF increase inside the cells, HIF-1 binds to specific DNA sequences and stimulates transcription of specific genes such as VEGF that promote angiogenesis. When oxygen levels in wound tissue increase, oxygen binds to HIF, leading to the destruction of HIF molecules in cells and decreased synthesis of angiogenic factors. Regulation of angiogenesis involves both stimulatory factors like VEGF and antiangiogenic factors like angiostatin, endostatin, and pigment epithelium-derived factor (PEDF).

Binding of angiogenic factors causes endothelial cells of the capillaries adjacent to the devascularized site to begin to migrate into the matrix and then proliferate to form buds or sprouts. Once again the migration of these cells into the matrix requires the local secretion of proteolytic enzymes, especially MMPs. As the tip of the sprout from one endothelial cell extends and encounters a sprout from another endothelial cell, the sprouts develop a cleft that subsequently becomes the lumen of the evolving vessel and complete a new vascular loop. This process continues until the capillary system is sufficiently repaired and the tissue oxygenation and metabolic needs are met. It is these new capillary tuffs that give granulation tissue its characteristic bumpy or granular appearance.

6.3.4 Epithelialization

Epithelialization is a multistep process that involves epithelial cell detachment and change in their internal structure, migration, proliferation, and differentiation. The intact mature epidermis consists of five layers of differentiated epithelial cells ranging from the cuboidal basal keratinocytes nearest the dermis up to the flattened, hexagonal, tough keratinocytes in the uppermost layer. Only the basal epithelial cells are capable of proliferation. These basal cells are normally attached to their neighboring cells by intercellular connectors called *desmosomes* and to the basement membrane by hemidesmosomes. When growth factors such as epidermal growth factor (EGF), keratinocyte growth factor (KGF) and TGFα are released during the healing process, they bind to receptors on these epithelial cells and stimulate migration and proliferation. The binding of the growth factors triggers the desmosomes and hemidesmosomes to dissolve so that the cells can detach in preparation for

migration. Integrin receptors are then expressed and the normally cuboidal basal epithelial cells flatten in shape and begin to migrate as a monolayer over the newly deposited granulation tissue, following along collagen fibers. Proliferation of the basal epithelial cells near the wound margin supply new cells to the advancing monolayer apron of cells (cells that are actively migrating are incapable of proliferation). Epithelial cells in the leading edge of the monolayer produce and secrete proteolytic enzymes (MMPs), which enable the cells to penetrate scab, surface necrosis, or eschar. Migration continues until the epithelial cells contact other advancing cells to form a confluent sheet. Once this contact has been made, the entire epithelial monolayer enters a proliferative mode and the stratified layers of the epidermis are reestablished and begin to mature to restore barrier function. TGFβ is one growth factor that can speed up the maturation (differentiation and keratinization) of the epidermal layers. The intercellular desmosomes and the hemidesmosome attachments to the newly formed basement membrane are also reestablished. Epithelialization is the clinical hallmark of healing but it is not the final event – remodeling of the granulation tissue is yet to occur.

6.4 Remodeling

Remodeling is the final phase of the healing process in which the granulation tissue matures into scar and tissue tensile strength increases. The maturation of granulation tissue involves a reduction in the number of capillaries via aggregation into larger vessels and a decrease in the amount of glycosaminoglycans and the water associated with the GAGs and proteoglycans. Cell density and metabolic activity in the granulation tissue drop during maturation. Changes are also made in collagen content and organization, which enhance tensile strength. The tensile strength of a newly epithelialized wound is only about 25% of normal tissue. Healed or repaired tissue is never as strong as normal tissues that have never been wounded. Tissue tensile strength is enhanced primarily by the reorganization of collagen fibers that were deposited randomly during granulation and increased covalent cross-linking of collagen molecules by the enzyme, lysyl oxidase, which is secreted into the ECM by fibroblasts. Over several months or more, changes in collagen organization in the repaired tissue will slowly increase the tensile strength to a maximum of about 80% of normal tissue.

Fibroblasts are the main cell type involved in remodeling. They synthesize new collagen, which is deposited along lines of tension. Initially, wound fibroblasts secrete type III collagen, which is then replaced by type I collagen during remodeling. Fibroblasts also secrete MMPs, which breakdown the type III collagen that was randomly deposited during the proliferative phase. A balance must be maintained between collagen removal and synthesis during remodeling, and this is achieved through the action of enzyme inhibitors like the tissue inhibitors of metalloproteinases that are also produced by fibroblasts.

6.5 Chronic Wounds

As outlined earlier, healing of skin wounds normally occurs in a predictable sequence of phases including: (1) hemostasis; (2) inflammation; (3) repair (characterized by mitosis, angiogenesis, and synthesis of ECM); and (4) remodeling of

the scar matrix. These processes are regulated by numerous molecules including growth factors, cytokines, chemokines, receptors, proteinases, and their inhibitors. It is reasonable to assume that the failure of an acute wound to progress through the phases of healing is due to conditions that interfere with the normal interactions of these molecules. Previous studies that analyzed fluids collected from chronic wounds found elevated levels of inflammatory cytokines, elevated levels of proteinases, and low levels of growth factor activity compared to acute, healing wounds. These observations led to the hypothesis that chronic wounds develop because of prolonged inflammation in acute wounds, which produces elevated levels of proteinases that destroy growth factors, receptors, and ECM proteins that are essential for healing. Furthermore, analysis of fibroblasts cultured from chronic wounds indicate that they frequently have reduced capacity to proliferate in response to specific growth factors. If this hypothesis is correct, it follows that elevated levels of inflammatory cytokines and proteinases should decrease and mitogenic activity should increase in fluids as chronic wounds begin to heal. Several studies have, indeed, reported reduced levels of inflammatory cytokines and proteases and increased mitogenic activity in wound fluids as chronic wounds begin to heal. These results support the concept that treatments that shift the molecular balance in chronic wounds to a pattern that resembles the molecular profile found in acute healing wounds will promote healing.

6.6 Wound Bed Preparation

Wound bed preparation is a comprehensive approach to wound management, which focuses on optimizing conditions in the wound to encourage normal endogenous processes of healing, by removing or correcting the barriers to healing. It is based on an understanding of the underlying molecular and cellular processes that occur in normal wound healing and the differences that have been identified in chronic wounds. The TIME acronym, developed by a group of wound care experts, was useful in focusing on four major attributes of chronic wounds that must be optimized for wounds to begin to heal. As shown in Table 1, the TIME acronym is a practical guide to wound management, which relates clinical observations and interventions to the underlying wound pathology in each of four areas: (1) T for tissue: nonviable or deficient; (2) I for infection/inflammation; (3) M for moisture balance; (4) E for edge of wound, nonadvancing or undermined. Application of the concepts of wound bed preparation has led to an improved dressing that contains collagen (gelatin) and is designed to reduce destruction of essential proteins in wound fluids by proteases in wound fluids by acting as a "substrate sink." A hydrogel containing doxycycline, which is an inhibitor of MMPs as well as an antibiotic, stimulated healing of chronic diabetic foot ulcers. Thus, wound bed preparation is emerging as a useful education aid as well as a practical guide for treatment strategies.

7 Summary

As the body's largest organ, the skin serves several complex functions: protection, thermoregulation, sensation, metabolism, and communication. Numerous factors influence the skin's ability to adequately

Tab. 1 TIME – Principles of wound bed preparation. Wound bed preparation is a comprehensive approach to wound management which focuses on optimizing conditions in the wound to encourage normal endogenous processes of healing by removing or correcting the barriers to healing. The TIME acronym is a practical guide to wound management which relates clinical observations and interventions to the underlying wound pathology in each of four areas: (1) T for tissue: nonviable or deficient; (2) I for infection/inflammation; (3) M for moisture balance; (4) E for edge of wound, nonadvancing or undermined.

Clinical observations	*Molecular & cellular problems*	*WBP clinical actions*	*Effects of clinical actions*	*Clinical outcome*
Tissue management Nonviable or deficient	Defective matrix and cell debris impair healing	Debridement (episodic or continuous) • Autolytic, sharp surgical, enzymatic, mechanical, or biological • Biological agents	Restoration of wound base and functional extracellular matrix proteins	Viable wound base
Infection and inflammation control	High bacteria counts or prolonged inflammation ↑ Inflammatory cytokines ↑ Proteases ↓ Growth factor activity	Remove infected foci – topical/systemic Antimicrobials Anti-inflammatories Protease inhibitors	Low bacteria counts or controlled inflammation ↓ Inflammatory cytokines ↓ Proteases ↑ Growth factor activity	Bacterial balance and reduced inflammation
Moisture balance	Desiccation slows epithelial cell migration Excessive fluid causes maceration of wound	Apply moisture balancing dressings Compression, negative pressure or other methods of removing fluid	Restore epithelial migration, desiccation avoided Edema, excessive fluid controlled, maceration avoided	Moisture balance
Epithelial edge advancement	Nonmigrating epidermal margin Nonresponsive wound cells and abnormalities in protease activities	Reassess cause, refer or consider corrective advanced therapies • Debridement • Bioengineered skin • Skin grafts • Adjunctive therapies	Migrating keratinocytes and responsive wound cells. Restoration of appropriate protease profile in wound	Advancing epithelial margin

provide these functions, such as age, UV exposure, hydration, medications, nutrition, and soaps. Knowledge of the structure and functions of normal skin are important for understanding the processes that occur during normal aging and healing of injuries to the skin. Extensive signaling through autocrine, paracrine, and endocrine pathways by growth factors, cytokines, and proteases occurs by multiple types of cells in wounds to regulate the rate and amount of scar tissue that forms. In addition, interaction of cells with ECM components via integrin receptors further regulates gene expression. When the sequential processes of normal wound healing are disrupted, wound healing is slowed or stopped resulting in a chronic wound. Identification of molecular and cellular abnormalities in chronic wounds provides a targeted approach to correct the abnormal imbalances and promote healing.

See also Immunology; Membrane Transport; Vascular Development and Angiogenesis.

Bibliography

Books and Reviews

Costerton, J.W., Lewandowski, Z., Caldwell, D.E., Korber, D.R., Lappin-Scott, H.M. (1995) Microbial biofilms, *Annu. Rev. Microbiol.* **49**, 711–745.

Council on Scientific Affairs. (1989) Harmful effects of ultraviolet radiation, *JAMA* **262**, 380–384.

Freinkel, R.K. (2001) Hair, in: Freinkel, R.K., Woodley, D.T. (Eds.) *The Biology of the Skin*, The Parthenon Publishing Group, New York.

Gilchrest, B.A. (1991) Physiology and Pathophysiology of Aging Skin, in: Goldsmith, L.A. (Ed.) *Physiology, Biochemistry and Molecular Biology of the Skin*, Oxford University Press, New York.

Haake, A.R., Holbrook, K.A. (1999) The Structure and Development of Skin, in: Freedberg, I.M., Eisen, A.Z., Wolff, K., Austen, K.F., Goldsmith, L.A., Katz, S. (Eds.) *Fitzpatrick's Dermatology in General Medicine*, McGraw-Hill, New York.

Holbrook, K.A. (2001) Structure and Function of the Developing Human Skin, in: Goldsmith, L.A. (Ed.) *Physiology, Biochemistry and Molecular Biology of the Skin*, Oxford University Press, New York.

Jacob, S.W., Francone, C.A., Lossow, W.J. (1982) *Structure and Function of Skin in Man*, WB Saunders, Philadelphia, PA.

McGregor, J.M., Hawk, J.L.M. (1999) Acute Effects of Ultraviolet Radiation on the Skin, in: Freedberg, I.M., Eisen, A.Z., Wolff, K., Austen, K.F., Goldsmith, L.A., Katz, S. (Eds.) *Fitzpatrick's Dermatology in General Medicine*, McGraw-Hill, New York.

Metze, D. and T. Lugen. 2001. Nervous System in the Skin, in: Freinkel, R.K., Woodley, D.T. (Eds.) *Biology of the skin*, Parthenon Publishing Group. New York.

Millington, P.F., Wilkinson, R. (1983) *Skin*, Cambridge University Press, Cambridge.

Odland, G.F. (1991) Structure of the Skin, in: Goldsmith, L.A. (Ed.) *Physiology, Biochemistry, and Molecular Biology of the Skin*, Oxford University Press, New York.

Quinton, P.M. (1983) Sweating and its Disorders, *Annu. Rev. Med.* **34**, 429–452.

Roberts, A.B., Sporn, M.B. (1996) Transforming Growth Factor-β, in: Clark, R.A.F. (Ed.) *The Molecular and Cellular Biology of Wound Repair*, Plenum Press, New York, pp. 275–308.

Roth, R.R., James, W.D. (1988) Microbial ecology of the skin, *Annu. Rev. Microbiol.* **42**, 441–464.

Sams, W.M. (1990) Structure and Function of the Skin, in: Sams, W.M., Lynch, P.J. (Eds.) *Principles and Practice of Dermatology*, Churchill Livingstone, New York.

Spince, A.P., Mason, E.B. (1987) *Human Anatomy and Physiology*, The Benjamin/Cummings Publishing Company, Menlow Park, CA.

Spitz, R. (1947) An Inquiry into the Genesis of Psychiatric Conditions in Early Childhood, in: Nagera, H. (Ed.) *Psychoanalytical Studies of the Child*, International, London.

Stryer, L. (1995) *Biochemistry*, W.H. Freeman and Company, New York.

Wenger, C.B. (1999) Thermoregulation, in: Fitzpatick, T.B., et al. (Eds.) *Fitzpatrick's Dermatology in General Medicine*, McGraw-Hill. New York.

Weston, W.L., Lane, A.T. (1999) Neonatal Dermatology, in Freeberg, I.M., et al. (Eds.) *Fitzpatrick's Dermatology in General Medicine*, McGraw-Hill, New York.

Wheater, P.R., Burkitt, H.G., Daniels, V.G. (1987) *Functional Histology*, Churchill Livingstone, Edinburgh.

Wysocki, A.B. (1995) A review of the skin and its appendages, *Adv. Wound Care* **8**, 53–62, 64.

Primary Literature

Agren, M.S., Steenfos, H.H., Dabelsteen, S., Hansen, J.B., Dabelsteen, E. (1999) Proliferation and mitogenic response to PDGF-BB of fibroblasts isolated from chronic venous leg ulcers is ulcer-age dependent, *J. Invest Dermatol.* **112**, 463–469.

Alterescu, V., Alterescu, K. (1988) Etiology and treatment of pressure ulcers, *Decubitus* **1**, 28–35.

Ashcroft, G.S., Greenwell-Wild, T., Horan, M.A., Wahl, S.M., Ferguson, M.W. (1999) Topical estrogen accelerates cutaneous wound healing in aged humans associated with an altered inflammatory response, *Am. J. Pathol.* **155**, 1137–1146.

Ashcroft, G.S., Herrick, S.E., Tarnuzzer, R.W., Horan, M.A., Schultz, G.S., Ferguson, M.W. (1997) Human ageing impairs injury-induced in vivo expression of tissue inhibitor of matrix metalloproteinases (TIMP)-1 and -2 proteins and mRNA, *J. Pathol.* **183**, 169–176.

Ashcroft, G.S., Horan, M.A., Herrick, S.E., Tarnuzzer, R.W., Schultz, G.S., Ferguson, M.W. (1997) Age-related differences in the temporal and spatial regulation of matrix metalloproteinases (MMPs) in normal skin and acute cutaneous wounds of healthy humans, *Cell Tissue Res.* **290**, 581–591.

Ashcroft, G.S., Dodsworth, J., van Boxtel, E., Tarnuzzer, R.W., Horan, M.A., Schultz, G.S., Ferguson, M.W. (1997) Estrogen accelerates cutaneous wound healing associated with an increase in TGF-beta1 levels, *Nat. Med.* **3**, 1209–1215.

Auger, M.J. (1989) Mononuclear phagocytes, *Br Med J* **298**, 546–548.

Benyon, R.C. (1989) The human skin mast cell, *Clin. Exp. Allergy* **19**, 375–387.

Bergstresser, P.R., Toews, G.B., Streilein, J.W. (1980) Natural and perturbed distributions of Langerhans cells: responses to ultraviolet light, heterotopic skin grafting, and dinitrofluorobenzene sensitization, *J. Invest. Dermatol.* **75**, 73–77.

Bettley, F.R. (1960) Some effects of soap on the skin, *Br. Med. J.* **5187**, 1675–1679.

Braverman, I.M. (1986) Elastic fiber and microvascular abnormalities in aging skin, *Dermatol. Clin.* **4**, 391–405.

Briggaman, R.A., Wheeler, C.E. Jr. (1968) Epidermal-dermal interactions in adult human skin: role of dermis in epidermal maintenance, *J. Invest. Dermatol.* **51**, 454–465.

Briggaman, R.A., Wheeler, C.E. Jr. (1971) Epidermal-dermal interactions in adult human skin. II. The nature of the dermal influence, *J. Invest. Dermatol.* **56**, 18–26.

Bullard, K.M., Longaker, M.T., Lorenz, H.P. (2003) Fetal wound healing: current biology, *World J. Surg.* **27**, 54–61.

Chin, G.A., Thigpin, T.G., Perrin, K.J., Moldawer, L.L., Schultz, G.S. (2003) Treatment of chronic ulcers in diabetic patients with a topical metalloproteinase inhibitor, doxycycline, *Wounds* **15**, 315–323.

Costerton, J.W., Stewart, P.S., Greenberg, E.P. (1999) Bacterial biofilms: a common cause of persistent infections, *Science* **284**, 1318–1322.

Cullen, B., Smith, R., McCulloch, E., Silcock, D., Morrison, L. (2002) Mechanism of action of PROMOGRAN, a protease modulating matrix, for the treatment of diabetic foot ulcers, *Wound Repair Regen.* **10**, 16–25.

Ehrlich, H.P., Hunt, T.K. (1968) Effects of cortisone and vitamin A on wound healing, *Ann. Surg.* **167**, 324–328.

Evans, N.J., Rutter, N. (1986) Development of the epidermis in the newborn, *Biol. Neonate* **49**, 74–80.

Gilchrest, B.A. (1989) Skin aging and photoaging: an overview, *J. Am. Acad. Dermatol.* **21**, 610–613.

Gilchrest, B.A., Murphy, G.F., Soter, N.A. (1982) Effect of chronologic aging and ultraviolet irradiation on Langerhans cells in human epidermis, *J. Invest. Dermatol.* **79**, 85–88.

Green, A., Siskind, V., Bain, C., Alexander, J. (1985) Sunburn and malignant melanoma, *Br. J. Cancer* **51**, 393–397.

Griffith, L.G., Naughton, G. (2002) Tissue engineering – current challenges and expanding opportunities, *Science* **295**, 1009–1014.

Harpin, V.A., Rutter, N. (1983) Barrier properties of the newborn infant's skin, *J. Pediatr.* **102**, 419–425.

Headington, J.T. (1986) The dermal dendrocyte, *Adv. Dermatol.* **1**, 159–171.

Headington, J.T., Cerio, R. (1990) Dendritic cells and the dermis: 1990, *Am. J. Dermatopathol.* **12**, 217–220.

Hoffmann, J.A., Kafatos, F.C., Janeway, C.A., Ezekowitz, R.A. (1999) Phylogenetic perspectives in innate immunity, *Science* **284**, 1313–1318.

Holbrook, K.A., Underwood, R.A., Vogel, A.M., Gown, A.M., Kimball, H. (1989) The appearance, density and distribution of melanocytes in human embryonic and fetal skin revealed by the anti-melanoma monoclonal antibody, HMB-45, *Anat. Embryol. (Berl)* **180**, 443–455.

Kalia, Y.N., Nonato, L.B., Lund, C.H., Guy, R.H. (1998) Development of skin barrier function in premature infants, *J. Invest. Dermatol.* **111**, 320–326.

Kim, D.K., Holbrook, K.A. (1995) The appearance, density, and distribution of Merkel cells in human embryonic and fetal skin: their relation to sweat gland and hair follicle development, *J. Invest. Dermatol.* **104**, 411–416.

Kiss, M., Wlaschek, M., Brenneisen, P., Michel, G., Hommel, C., Lange, T.S., Peus, D., Kemeny, L., Dobozy, A., Scharffetter-Kochanek, K. (1995) Alpha-melanocyte stimulating hormone induces collagenase/matrix metalloproteinase-1 in human dermal fibroblasts, *Biol. Chem. Hoppe Seyler* **376**, 425–430.

Korting, H.C., Braun-Falco, O. (1996) The effect of detergents on skin pH and its consequences, *Clin. Dermatol.* **14**, 23–27.

Ladwig, G.P., Robson, M.C., Liu, R., Kuhn, M.A., Muir, D.F., Schultz, G.S. (2002) Ratios of activated matrix metalloproteinase-9 to tissue inhibitor of matrix metalloproteinase-1 in wound fluids are inversely correlated with healing of pressure ulcers, *Wound Repair Regen.* **10**, 26–37.

Lavker, R.M., Zheng, P.S., Dong, G. (1986) Morphology of aged skin, *Dermatol. Clin.* **4**, 379–389.

Lin, A.N., Carter, D.M. (1986) Skin cancer in the elderly, *Dermatol. Clin.* **4**, 467–471.

Lober, C.W., Fenske, N.A. (1990) Photoaging and the skin: differentiation and clinical response *Geriatrics* **45**, 36–40, 42.

Longaker, M.T., Adzick, N.S., Hall, J.L., Stair, S.E., Crombleholme, T.M., Duncan, B.W., Bradley, S.M., Harrison, M.R., Stern, R. (1990) Studies in fetal wound healing, VII. Fetal wound healing may be modulated by hyaluronic acid stimulating activity in amniotic fluid, *J. Pediatr. Surg.* **25**, 430–433.

Longaker, M.T., Whitby, D.J., Adzick, N.S., Crombleholme, T.M., Langer, J.C., Duncan, B.W., Bradley, S.M., Stern, R., Ferguson, M.W., Harrison, M.R. (1990) Studies in fetal wound healing, VI. Second and early third trimester fetal wounds demonstrate rapid collagen deposition without scar formation, *J. Pediatr. Surg.* **25**, 63–68.

Mast, B.A., Schultz, G.S. (1996) Interactions of cytokines, growth factors, and proteases in acute and chronic wounds, *Wound Repair Regen.* **4**, 411–420.

Narisawa, Y., Hashimoto, K., Kohda, H. (1996) Merkel cells participate in the induction and alignment of epidermal ends of arrector pili muscles of human fetal skin, *Br. J. Dermatol.* **134**, 494–498.

Narisawa, Y., Hashimoto, K., Nihei, Y., Pietruk, T. (1992) Biological significance of dermal Merkel cells in development of cutaneous nerves in human fetal skin, *J. Histochem. Cytochem.* **40**, 65–71.

Nerem, R.M., Levesque, M.J., Cornhill, J.F. (1980) Social environment as a factor in diet-induced atherosclerosis, *Science* **208**, 1475–1476.

Nordlund, J.J. (1986) The lives of pigment cells, *Dermatol. Clin.* **4**, 407–418.

Nwomeh, B.C., Yager, D.R., Cohen, I.K. (1998) Physiology of the chronic wound, *Clin. Plast. Surg.* **25**, 341–356.

O'Toole, E.A. (2001) Extracellular matrix and keratinocyte migration, *Clin. Exp. Dermatol.* **26**, 525–530.

Ottenbacher, K.J., Muller, L., Brandt, D., Heintzelman, A., Hojem, P., Sharpe, P. (1987) The effectiveness of tactile stimulation as a form of early intervention: a quantitative evaluation, *J. Dev. Behav. Pediatr.* **8**, 68–76.

Pollack, S.V. (1982) Systemic medications and wound healing, *Int. J. Dermatol.* **21**, 489–496.

Potts, J.F. (1990) Sunlight, sunburn, and sunscreens. Preventing and remedying problems from 'too much fun in the sun', *Postgrad. Med.* **87**, 52–60, 63.
Rutter, N. (1988) The immature skin, *Br. Med. Bull.* **44**, 957–970.
Sauder, D.N. (1986) Effect of age on epidermal immune function, *Dermatol. Clin.* **4**, 447–454.
Schultz, G.S., Sibbald, R.G., Falanga, V., Ayello, E.A., Dowsett, C., Harding, K., Romanelli, M., Stacey, M.C., Teot, L., Vanscheidt, W. (2003) Wound bed preparation: a systematic approach to wound management, *Wound Repair Regen.* **11**(Suppl. 1), S1–S28.
Shuster, S., Fisher, G.H., Harris, E., Binnell, D. (1978) The effect of skin disease on self image, *Br. J. Dermatol.* **99**, 18–19.
Siebert, J.W., Burd, A.R., McCarthy, J.G., Weinzweig, J., Ehrlich, H.P. (1990) Fetal wound healing: a biochemical study of scarless healing, *Plast. Reconstr. Surg.* **85**, 495–502.
Silverberg, N., Silverberg, L. (1989) Aging and the skin, *Postgrad. Med.* **86**, 131–134.
Stender, I.M., Blichmann, C., Serup, J. (1990) Effects of oil and water baths on the hydration state of the epidermis, *Clin. Exp. Dermatol.* **15**, 206–209.
Trengove, N.J., Stacey, M.C., Macauley, S., Bennett, N., Gibson, J., Burslem, F., Murphy, G., Schultz, G. (1999) Analysis of the acute and chronic wound environments: the role of proteases and their inhibitors, *Wound Repair Regen.* **7**, 442–452.
Warner, R.R., Myers, M.C., Taylor, D.A. (1988) Electron probe analysis of human skin: determination of the water concentration profile, *J. Invest. Dermatol.* **90**, 218–224.
White, M.I., Jenkinson, D.M., Lloyd, D.H. (1987) The effect of washing on the thickness of the stratum corneum in normal and atopic individuals, *Br. J. Dermatol.* **116**, 525–530.
Wolff, K., Stingl, G. (1983) The Langerhans cell. *J. Invest. Dermatol.* **80**(Suppl.), 17s–21s.
Woodburne, R.T., Burkel, W.E. (1988) *Essentials of Human Anatomy*, Oxford University Press, New York.
Wysocki, A.B. (2002) Evaluating and managing open skin wounds: colonization versus infection, *AACN Clin. Issues* **13**, 382–397.
Xu, K.D., McFeters, G.A., Stewart, P.S. (2000) Biofilm resistance to antimicrobial agents, *Microbiology* **146**, 547–549.
Yager, D.R., Nwomeh, B.C. (1999) The proteolytic environment of chronic wounds, *Wound Repair Regen.* **7**, 433–441.
Yanagishita, M. (1994) A brief history of proteoglycans, *EXS* **70**, 3–7.

Slime Mold (Dictyostelium discoideum), Molecular Biology of: *see* Dictyostelium Discoideum: Cellular Slime Mold

Smell, sense of: *see* Sense of Smell

Somatic Cell Nuclear Transfer (Cloning): *see* Nuclear Transfer for Cloning Animals

Somatic Gene Therapy

M. Schweizer, E. Flory, C. Münk, Uwe Gottschalk, and K. Cichutek
Paul-Ehrlich-Institut, Langen, Germany

1 Introduction 252

2 Gene Transfer Methods 254
2.1 Nonviral Vectors and Naked Nucleic Acid 256
2.2 Viral Vectors 257
2.2.1 Retroviral Vectors 257
2.2.2 Lentiviral Vectors 258
2.2.3 Adenoviral Vectors 258
2.2.4 AAV (Adeno-associated Viral) Vectors 259
2.2.5 Poxvirus Vectors 259

3 Clinical Use 259
3.1 Overview on Clinical Gene Therapy Trials 259
3.2 Gene Therapy of Monogeneic Congenital Diseases 259
3.3 Tumor Gene Therapy 261
3.4 Gene Therapy of Cardiovascular Diseases 262
3.5 Preventive Vaccination and Gene Therapy of Infectious Diseases 262
3.6 Clinical Gene Therapy for the Treatment of Other Diseases 263

4 Manufacture and Regulatory Aspects 263

5 First Experience with the Clinical Use of Gene Transfer Medicinal Products 267

Bibliography 267
Books and Reviews 267
Primary Literature 268

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

1 Introduction

Innovative biotechnologicals of the future include gene transfer medicinal products. It can be assumed that by mid-2003, ~4000 patients or healthy individuals would have been treated within a clinical gene therapy trial, ~600 of those in Europe, and ~260 in Germany (The data originate from some statistics published in the Internet by Wiley Genetic Medicine Clinical Trials Online, http:///www.wiley.co.uk.genetherapy, and are reproduced by courtesy of the publisher.). Most of the clinical trials are currently in phase I or II due to a great diversity of ongoing developments, clinical experience must first be gained, before target-orientated product development and phase III clinical trials can be initiated. In this regard, investigator-driven gene therapy strategies developed by biomedical laboratories together with special clinic teams are very distinct from those developed by pharmaceutical industry. Investigator-driven gene therapy strategies are being invented by teams of biomedical researchers and physicians while developing a new approach for the treatment of a special disease in a defined stage. This is used, for the first time, on a selected group of patients in first clinical trials of phase I/II and aimed at proving the safety of the medicinal product. In clinical trials sponsored by pharmaceutical industry, this phase of orientation has often already been completed and further development in phase II or III is aimed at dose finding or proving efficacy. Concerning product development, there are no standard approaches because, at this stage of development, little experience has been gained and the types of gene transfer medicinal products are very diverse. Therefore, in the following sections, the main current clinical developments will be described while a brief outline of a single example of a manufacturing process is given, also due to manifold diversity.

Gene transfer medicinal products for human use are medicinal products used for *in vivo* diagnosis, prophylaxis, or therapy (Fig. 1). They contain or consist of:

1. genetically modified cells,
2. viral vectors, nonviral vectors or so-called naked nucleic acids, or
3. recombinant replication-competent microorganisms used for purposes other than the prevention or therapy of the infectious diseases that they cause.

The aim of the nucleic acid or gene transfer is the genetic modification of human somatic cells, either in the human body, that is, *in vivo*, or outside the human body, that is, *ex vivo*, in the latter case followed by transfer of the modified cells to the human body. The simplest case of genetic modification of a cell results from addition of a therapeutic gene encompassed by an expression vector. At least in theory, nucleic acid transfer may also be aimed at exchange of individual point mutations or other minimal genetic aberrations. Scientifically, this process is termed *homologous recombination* with the aim of repairing a defective endogenous gene at its locus. In principle, this can be achieved by so-called homologous recombination achieved by transferring oligonucleotides, where – thanks to 5′ and 3′ flanking homology regions – the new correct DNA sequence is replacing the existing defective one. In practice, homologous recombination is technically not yet achievable with the efficiency that will be required for clinical use.

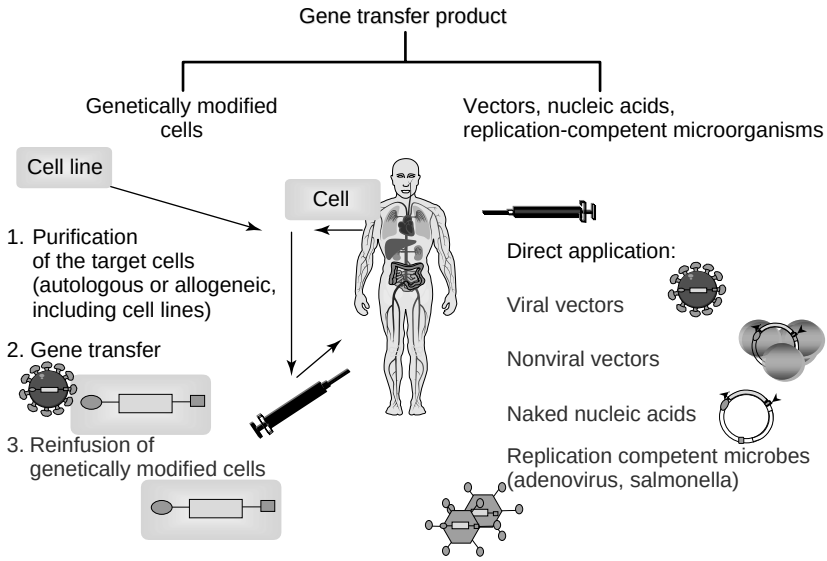

Fig. 1 Gene transfer medicinal products. The gene transfer medicinal products mentioned here are identical with those described in Table 1 of the European "Note for guidance on the quality, preclinical, and clinical aspects of gene transfer medicinal products (CPMP/BWP/3088/99)." The definition given is in compliance with the legally binding definition of gene therapeutics in Part IV, Annex I of Directive 2003/63/EC amending Directive 2001/83/EC.

Normally, genetic modification of cells is nowadays achieved by the transfer of an expression vector on which the therapeutic gene is located. The vector is transferred to cells via a delivery system (Fig. 2) such as a viral vector particle, a nonviral vector complex or a plasmid. In the latter case, the expression vector is inserted into and therefore part of a bacterial plasmid, which allows its manufacture and amplification in bacteria. Viral expression vectors contain the sequence signals (nucleic acid sequences) required for transfer by a particular viral vector particle. For retroviral vectors, for example, such signals are encompassed by the flanking "long terminal repeat" (LTR) sequences, the packaging signal psi (Ψ) required for incorporation of the expression vector by the retroviral vector particle, and other sequence signals. For nonviral vector complexes and naked nucleic acid, the expression vector is part of a bacterial carrier, the so-called plasmid DNA. Nonviral vectors are, for example, plasmid DNA mixed with a transfection reagent, whereas naked DNA does not contain a transfection reagent.

Another example of a gene transfer medicinal product is a recombinant microorganism such as conditionally replication-competent adenoviruses for tumor therapy. Here, neither is an endogenous cellular gene repaired by homologous recombination nor is a nonadenoviral therapeutic gene transferred. The transfer of conditional replicating adenoviruses to the malignant tumor cells induces cell lysis and local tumor ablation. The entire genome of the adenovirus is transferred without an additional therapeutic gene. The adenoviral genome may therefore be considered as the therapeutic gene.

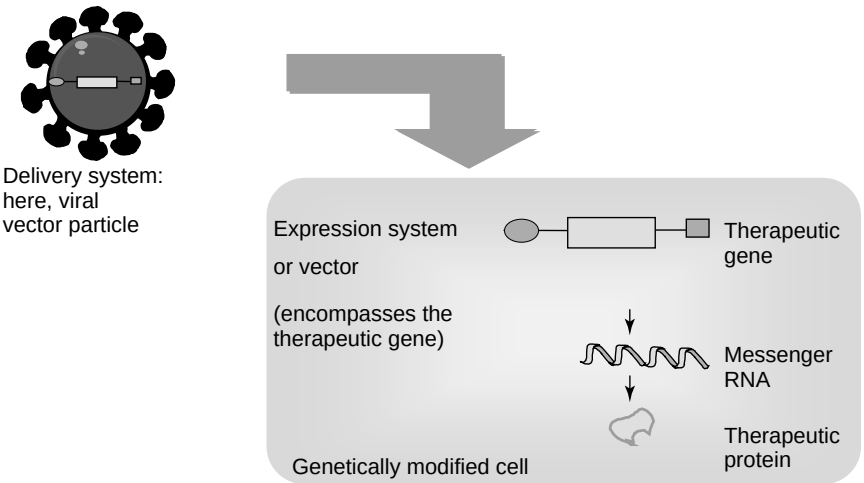

Fig. 2 Delivery system and expression vector used as gene transfer medicinal products. The terminology complies with the definition of gene therapeutics in Part IV, Annex I of Directive 2003/63/EC amending Directive 2001/83/EC.

Gene transfer efficiency plays a central role in gene transfer. It depends on a number of factors, for example, target cell, type of application (*ex vivo* or *in vivo* strategy), the tissue or organ containing the target cells, the physiological situation, and the disease and disease stage. Table 1 shows the most common viral vectors currently in clinical use. The vectors shown are replication incompetent and only transfer the expression vector void of any viral genes as much as possible. So-called integrating vectors mediate chromosomal integration of the expression vector (e.g. retroviral vectors), whereas nonintegrating vectors lead to an episomal status of the expression vector in the cell (e.g. adenoviral vectors), or to its cytoplasmatic replication (e.g. α-virus-derived vectors, vaccinia). Vectors derived from vaccinia, for example, used for tumor vaccination, may be replication incompetent such as modified vaccinia ancara (MVA) or ALVAC or replication competent, but attenuated like vaccinia.

After uptake by human somatic cells, the expression vector is transcribed like a normal cell gene. The resulting messenger RNA (mRNA) is translated and the therapeutic protein is synthesized by the cellular machinery. When so-called ribozyme genes are used, the mRNA acts like a catalytic enzyme and is itself the therapeutic gene product. As already mentioned, when a recombinant microorganism such as a conditionally replication-competent adenovirus (RCA) is used, the genome of the microorganism may be seen as the therapeutic gene.

2 Gene Transfer Methods

The objective of clinical gene transfer is the transfer of nucleic acids for the purpose of genetically modifying human cells (Fig. 3).

Whether a viral, a nonviral vector, or a naked plasmid DNA is used depends on

Tab. 1 Gene transfer methods (vectors/delivery systems).

Delivery system	*Description*	*Chromosomal integration*
Naked nucleic acid	Plasmid DNA, in absence of transfection reagents	No (after im inoculation)
Nonviral vector	Plasmid DNA/transfection reagent mixture	No (application dependent)
Viral vector		
Retroviral vector	Derived from murine leukemia virus (MLV)	Yes
Lentiviral vector	Derived from HIV-1	Yes
Adenoviral vector	Deletions in the virus genes E1, E3 or E4, E2ts, combinations thereof or "gutted" (gene-depleted)	No
Conditionally replication-competent adenovirus	No therapeutic gene except for the virus genome	No
Adeno-associated virus (AAV) vector	Wild-type AAV-derived	Yes /no (application dependent)
Smallpox virus vector	MVA ("Modified Vaccinia Ancara")	No
	ALVAC ("Avian Vaccinia")	No
	Vaccinia	No
Alphavirus vector	SFV	No
Herpes-viral vector	Herpes simplex virus	No

Note: SFV: Semliki Forest virus; MLV: murine leukemia virus.

the target cell of the genetic modification and whether an *in vivo* modification of the cell is at all possible. For a monogeneic disease affecting immune cells, it is for example, possible to purify CD34-positive cells or lymphocytes from the peripheral blood (e.g. by leukapharesis), to genetically modify the cells in culture, and to return the treated cells. Before reapplication, the treated cells may or may not be enriched. Currently, long-term correction of cells is only possible when integrating vectors such as retroviral or lentiviral vectors are used. Owing to chromosomal integration of the expression vector, the genetic modification is passed on to the daughter cells during cell division and persists. Only long-term expression may still be a problem. For therapy of a monogeneic disease such as cystic fibrosis, the target cells are primarily the endothelial cells of the bronchopulmonary tract, which can only be subjected to *in vivo* modification attempts. Although long-term correction would be desirable, *in vivo* modification using adenoviral vectors appeared to be more promising, because the target cells are largely in a resting state of the cell cycle amenable to adenoviral gene transfer due to expression of the cell surface receptors used by adenoviruses for cell entry. In addition, the amount and titers of adenoviral vectors seemed suitable. These examples illustrate that a number of factors contribute to the choice of the treatment strategy, the vector, and the route of administration. No single "ideal" vector is therefore suitable

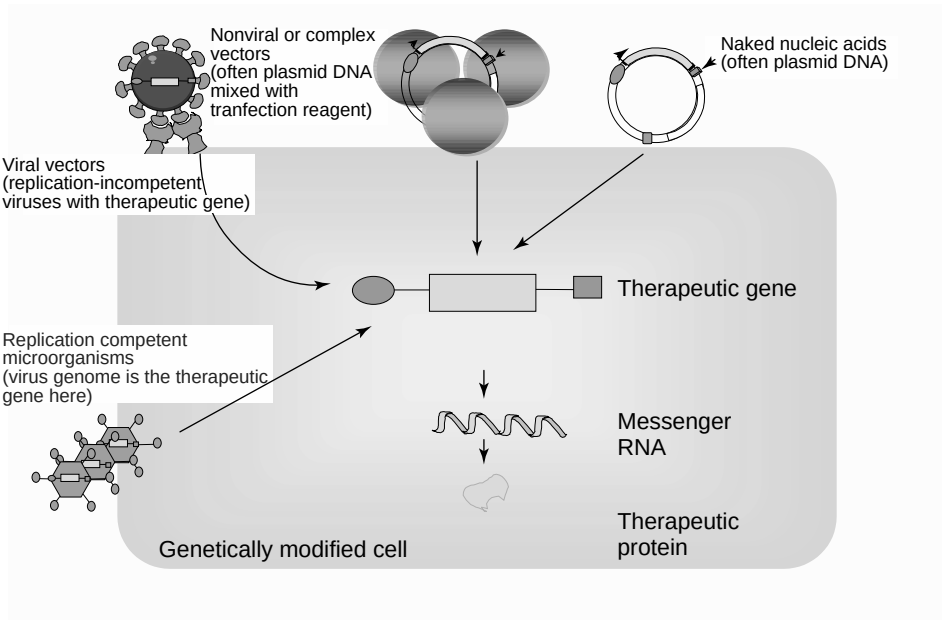

Fig. 3 Delivery systems used in clinical gene transfer. During gene therapy, an expression vector (therapeutic genes) is transferred to somatic cells via a delivery system, for example, a viral or nonviral vector (replication incompetent), a naked nucleic acid, or a recombinant, mostly conditionally replication-competent microorganism. The gene transfer, termed *transfection* when a viral vector is used or termed *transfection* when naked DNA or a nonviral vector is used, leads to genetic modification of the cell. The gene transfer can be carried out *in vivo*, that is, directly in or on the human body, or *ex vivo*, that is, in cell culture followed by the transfer of the modified cells to the human body.

for a large variety of gene therapies. In the past 15 years, many novel gene transfer techniques have been developed and used in clinical studies. In the following section and in Table 1, specific characteristics of the vectors most commonly used in the clinic are summarized.

2.1 Nonviral Vectors and Naked Nucleic Acid

The advantage of nonviral gene transfer systems compared with viral gene transfer systems is the smaller size limitations for the genes to be transferred. The expression vectors are nowadays usually part of a bacterial plasmid that can easily be amplified and grown in bacterial cultures. Plasmid DNA of up to 20 kb pairs encompassing an expression vector of up to 17 kb pairs can easily be manufactured. Promising methods for the *in vivo* administration of plasmid DNA include intradermal or intramuscular injection for the so-called naked nucleic acid transfer. Needle injection or application by medical devices such as gene guns can be used for this purpose. For so-called nonviral vectors, for example, synthetic liposomes or other transfection reagents mixed with plasmid DNA, the DNA-binding liposomes mediate contact with the cellular

plasma membrane thus releasing the DNA into the cytoplasma of the cell where uptake by the nucleus has to occur subsequently. During receptor-mediated uptake of nonviral vectors, cell surface proteins (receptors), for example, asialoglycoprotein or the transferrin receptor, mediate cellular uptake of the DNA complex containing a specific receptor ligand.

2.2 Viral Vectors

During evolution, viruses have been optimized to efficiently enter mammalian cells and replicate. Infected mammalian cells transcribe the viral genes and synthesize the viral gene products with high efficiency, sometimes to the disadvantage of endogenous protein production. Viral vectors are replication-incompetent particles derived from viruses by genetic engineering, which no longer transfer to cells the complete set or any viral genes. Instead, an expression vector with one or more therapeutic genes is transferred to cells. Since no complete viral genome is transferred, virus replication is impossible or, in some cases, it is impaired like with first- or second-generation adenoviral vectors. The following section briefly describes the properties of the current frequently used viral vectors.

2.2.1 Retroviral Vectors

The retroviral vectors in clinical use have mainly been derived from murine leukemia virus (MLV). MLV causes leukemia in mice and replication-competent retrovirus (RCR) in a contaminated vector preparation was shown to cause leukemia in severely immunosuppressed monkeys. RCR absence has therefore to be verified before human use of retrovirally modified cells. MLV vector use *in vivo* has been very rare. The genome of the retroviral vectors consists of two copies of single-stranded RNA, which contains one or more coding regions flanked by the viral control elements, the so-called "long-terminal repeat" regions. In the infected cells, the RNA is translated into double-stranded viral DNA and integrated into the cell. The integrated vector DNA is the expression vector. MLV vectors allow efficient genetic modification of proliferating cells by chromosomally integrating the expression vector.

Advantages of retroviral vectors include high gene transfer (transduction) efficiency, and long-term modification of cells due to stable integration of the expression vector into the chromosome of the cells. In addition, the MLV envelope proteins can be exchanged against those from other viruses (which is termed *vector pseudotyping*). This allows preparation of MLV vectors with improved transduction efficiency for certain cell types. Disadvantages of retroviral vectors include the small size of the coding region (~9 kb pairs or less), the restriction of transduction to proliferating cells only, insertional mutagenesis due to integration and the low titer of usually not more than 10^8 transducing units per milliliter of vector preparation. Although chromosomal integration occurs generally at random, it may lead to activation of cellular cancer genes, so-called proto-oncogenes, or, theoretically, to inactivation of tumor suppressor cells. In conjunction with additional genetic mutations, this may result in very low frequency in malignant cell transformation. Hundreds of patients that have been treated with retrovirally modified hematopoietic cells years ago have not shown any signs of cancer related to the gene transfer except for two patients

treated during an SCID-X1 (severe combined immunodeficiency disease) gene therapy trial in France. In the latter two leukemia cases, the vector-mediated overexpression of the proto-oncogene LMO2, possibly in conjunction with the therapeutic γc chain gene (which may influence cell proliferation and signal transduction) and the SCID-X1 disease, are the probable cause of the leukemia (see the following sections).

2.2.2 Lentiviral Vectors

Lentiviral vectors have been derived from human immunodeficiency virus type 1 (HIV-1), simian immunodeficiency virus (SIV) isolated from various old-world monkeys, feline immunodeficiency virus (FIV) and equine infectious anemia virus (EIAV) isolated from horses. Lentiviruses cause an acquired immunodeficiency syndrome and a replication-competent virus has therefore to be excluded before human use by batch-to-batch analysis and verification of the absence of replication-competent lentivirus (RCL). Lentiviral vectors may transfer coding regions of up to 9 kb pairs and allow pseudotyping just like MLV vectors. Their advantage is the dual capacity to transfer therapeutic genes into nonproliferating cells in conjunction with persistent genetic modification due to chromosomal integration. This could be useful for *ex vivo* modification of stem cells and *in vivo* modification of neuronal cells. Most lentiviral vectors have been pseudotyped with the G-protein of vesicular stomatitis virus (VSV-G) or the envelope proteins of Gibbon ape leukemia virus. The first clinical study using lentiviral vectors was started in 2003 and involves the *ex vivo* modification of autologous lymphocytes of HIV-infected patients with a therapeutic ribozyme gene shown *in vitro* to inhibit HIV-1 replication.

2.2.3 Adenoviral Vectors

The adenoviral genome consists of double-stranded DNA that persists episomally, that is, inside the nucleus, but not integrated into the chromosome of the cell. Therefore, the genetic modification may be lost during cell proliferation. Adenoviral vectors are the currently preferred vectors for the *in vivo* transduction of a variety of human somatic cells including nonproliferating cells. In contrast to lentiviral vectors, they allow insertion of larger coding regions of therapeutic genes above 10 kb pairs and are not associated with a detectable risk of insertional oncogenesis. In addition, vector titers above 10^{11} transducing units per milliliter can usually be achieved. The lack of long-term expression is in part due to the fact that certain adenovirus genes have been kept on first or second-generation adenoviral expression vectors, and because of the frequent generation of RCA during production. So-called gutless vectors are void of any adenoviral genes, but have to be purified from RCA after production.

Some wild-type (replicating) adenovirus strains cause inflammations of the airways and the conjunctivae. Adenoviral vectors may therefore also be transferred by inhalation of aerosols and inflammations observed following vector applications are mainly local, transient, and associated with very high titer applications. High-titer adenoviral vectors are no longer systemically administrated, because one patient died during systemic administration of a maximum dose of $\sim 10^{13}$ vector particles during gene therapy of the monogeneic disease OTCOTC (''ornithine transcarbamylase'') deficiency, a life-threatening metabolic disorder.

2.2.4 AAV (Adeno-associated Viral) Vectors

Adeno-associated viruses (AAVs) belong to the family of parvoviruses. Their genome consists of single-stranded DNA. Wild-type AVV can only replicate in the presence of helper viruses like adenovirus or herpesvirus and has not been associated with disease. AAV can infect hematopoietic cells including nonproliferating cells. Integration in infected human somatic cells is often confined to a distinct locus on human chromosome 19. AAV-derived vectors are usually classified as integrating vectors, although vector integration is unfortunately no longer confined to chromosome 19, but absence of integration may be observed, for example, following intramuscular administration The size of the coding region is very limited (~4 kb pairs).

2.2.5 Poxvirus Vectors

Poxvirus vectors encompass vaccinia derived from the smallpox vaccine and more attenuated variants like ALVAC or MVA. Their genome consists of single-stranded DNA of 130 to 300 kb pairs. Replication is restricted to the cytoplasm of cells and high amounts of protein are synthesized by the cell following transduction. Most applications, therefore, involve intramuscular vaccination.

3 Clinical Use

3.1 Overview on Clinical Gene Therapy Trials

A number of clinical trials show promising results (see Table 2). In the past few years, it has become increasingly clear that for each disease, the development of a particular and specific gene transfer method in connection with a particular treatment approach will probably be necessary. The first standard use of an approved gene transfer medicinal product is to be expected within the next seven years since ~1% of the clinical gene therapy studies are in an advanced stage of phase II or phase III clinical trial.

Clinical gene therapy studies have been performed initially in North America and Europe. About 50 clinical gene transfer studies have been registered in Germany, with slightly more than 250 patients that have been treated (http://www.pei.de) (http://www.zks.uni-freiburg.de/dereg.html). A general overview on registered studies is listed on the following Web sites: (http://www.wiley.co.uk/genetherapy, or www.pei.de). In Germany, a public registry has been available since 2004.

Target diseases in most clinical gene therapy trials have been cancer, cardiovascular diseases, infectious diseases such as AIDS, or monogeneic congenital disorders. The vectors most frequently used *ex vivo* are MLV vectors derived from murine leukaemia virus (MLV), whereas vectors derived from adenovirus, pox viruses, and AAV are usually used *in vivo*. A growing number of studies involves the use of nonviral vectors or naked DNA.

3.2 Gene Therapy of Monogeneic Congenital Diseases

The idea underlying gene therapy is the replacement of a defective gene by its normal, functional counterpart; for example, a mutation of the gene encoding the γc chain of the interleukin-2 and other receptors is the cause of the congenital immune disorder SCID-X1.

Tab. 2 Promising clinical gene therapy trials.

Disease	*Therapeutic gene*	*Pharmaceutical form/vector*	*Target cell*	*Remarks*
Human severe combined immunodeficiency SCID-X1	γc Chain gene (e.g. interleukin-2-receptor part)	MLV vector	Bone marrow stem cells *ex vivo*	4 of 1 patient cured, 2 leukemias
PAOD	VEGFgene	Plasmid DNA	Muscle/ endothelial cells *in vivo* (im)	Improved blood flow
Head and neck tumor	Adenovirus genome (cell lysis/apoptosis)	Tumor cell specific replicating adenovirus	p53-Negative tumor cells *in vivo*	Local tumor remission in combination with chemotherapy
Graft versus host disease in donor lymphocyte transfer for leukemia treatment	Thymidin kinase gene of the herpes simplex virus, followed by treatment with Ganciclovir	T cells, MLV vectors	T-lymphocytes *ex vivo*	Successful treatment of host-versus-graft disease
Hemophilia B	Coagulation factor IX gene	AAV vector	Muscle cells *in vivo* (im)	Improved coagulation factor concentration

Note: PAOD: peripheral artery occlusive disease; VEGF: vascular endothelial growth factor; AAV: Adeno-associated virus; MLV: murine leukemia virus.

Owing to this defect, immunologically relevant receptors are unable to mediate the normal differentiation and immune function of lymphoid cells such as T cells and natural killer lymphocytes (NK). Therefore, newborn babies suffering from SCID-X1 have a very limited immune system and must live in a germ-free environment. Their life expectancy is strongly reduced. Conventional treatment, that is, bone marrow transplantation, can provide a cure to a certain extent, but involves a high risk if no HLA haploidentical donor is available. For the latter situation, gene therapy within the framework of a clinical study was considered in France.

In this study, autologous CD34-positive bone marrow stem cells were retrovirally modified to express the functional γc chain gene. T cells and other hemaotpoietic cells derived from corrected stem cells were shown to repopulate the hematopoietic cell compartment, and over a period of up to three years, 11 treated patients, mostly newborns, displayed a functional and nearly normal immune system. This represents the first reproducible cure of a disease by gene therapy.

A leukemia-like lymphoproliferative disease was diagnosed roughly three years after treatment of two obviously cured patients. Treatment had been started at the age of a few months. Subsequent analysis revealed that the leukemia-like disease was indeed caused by the MLV vector; the disease mechanism is termed *insertional oncogenesis* resulting from insertional mutagenesis of the proto-oncogene LMO2 (mentioned earlier). According to current knowledge, up to 50 cells with an integration in LMO2 may have been administered together with the $\sim 10^8$ genetically modified CD34-positive bone marrow cells. Owing to the expression vector integration, the transcription of the LMO2 gene was deregulated and activated. Under normal circumstances, the body can cope with individual cells presenting preneoplastic changes like the one described. In the two treated children, however, further genetic changes must have accumulated to finally result in leukemia. Contributing factors discussed include the effect of the therapeutic γc chain gene, the product of which influences cell proliferation and differentiation, and other so far unknown genetic changes that may have occurred during the massive *in vivo* cell replication. In SCID-X1, the T-cell compartment is completely depleted, and is replenished after gene therapy by differentiation and replication of a few genetically corrected blood stem cells. During this process, genetic aberrations may occur with substantial frequency. However, further analysis will be required to understand the exact cause of leukemia development in SCID-X1 gene therapy. Since hundreds of patients treated with retrovirally modified cells in the past 10 years have not developed leukemia up to now, it is currently assumed that a practical risk of leukemia only exists in SCID-X1 gene therapy.

Gene therapy of hemophilia B also seems promising. Here, AAV vectors encoding a smaller, but functional version of the human coagulation factor IX gene were administered by intramuscular injection. A detectable increase in factor IX plasma concentration was observed. Even repeated AAV injections were well tolerated.

3.3 Tumor Gene Therapy

There are various gene therapy approaches that are being developed for the treatment

of cancer. They are aimed at inhibiting molecular pathways underlying malignant cell transformation. In other cases, tumor cell ablation by directly applying cell-killing mechanisms or, more indirectly, by improving immunological defense mechanisms directed against tumor cells are attempted.

A number of gene therapy studies involving the adenoviral transfer of tumor suppressor genes like p53 have already been performed. This is aimed at reverting malignant cell transformation or at inducing apoptosis. However, transduction following, for example, needle inoculation into tumors has been shown to be limited to a few cells close to the needle tracks. Direct tumor cell ablation by local injection of conditional RCAs in head and neck tumors led to detectable local tumor regression by direct virus-mediated cell lysis, especially when chemotherapy was used in parallel. Here, virus replication improved transduction efficiency *in vivo*. For the treatment of malignant brain tumors, variant herpesviruses have been inoculated into the tumor in order to lyse the tumor cells *in vivo*, especially if prodrugs have been administered that are converted by the viral thymidin kinase gene to a toxic drug.

In addition, a number of clinical approaches have already been tested that led to an improvement of immune recognition of tumors. They involved intratumoral injection of vectors, which transfer foreign MHC genes, such as B7.1 or B7.2, or cytokine genes, for example, interleukin-2 or granulocyte-macrophage colony-stimulating factor (GM-CSF). Here, vaccinia derived vectors such as MVA or ALVAC have often been used. Autologous or allogeneic tumor cells were also modified *ex vivo* by transfer of immunostimulating genes. Promising results have been reported from a phase I-study in which autologous tumor cells were adenovirally modified with the GM-CSF gene and rapidly reinoculated to stimulate antitumor immunity.

3.4 Gene Therapy of Cardiovascular Diseases

Local intramuscular injection of plasmid DNA or adenoviral vectors encoding vascular epithelial growth factor or fibroblast growth factor, both able to induce the formation of new blood vessels, has been used to improve microcirculation in ischemic tissue. Needle injection of plasmid DNA has been used in leg muscle, catheter application, or needle injection was also tried in ischemic heart muscle. The formation of new blood vessels and an improvement in the microcirculation has been observed.

A narrowing of the blood vessels (restenosis) often occurs after coronary blood vessel dilatation by stent implantation. This is probably caused by the proliferation of smooth muscle cells following injuring of the blood vessel endothelium by the stent. Here, the role of adenoviral or plasmid DNA mediated transfer of the gene encoding inducible nitroxide synthase (iNOS) is thought to result in reduced cell proliferation.

3.5 Preventive Vaccination and Gene Therapy of Infectious Diseases

During the past five to ten years, effective medicines have been developed for the treatment of AIDS. Combinations of effective chemotherapeutics are able to inhibit various steps of the replication cycle of HIV-1. This often results in reduction of the viral load in the peripheral blood, sometimes down to a level barely detectable with modern techniques. Because of the

requirement for long-term treatment and the massive adverse effects related to conventional treatment by chemotherapy, gene therapy of HIV infection could offer additional therapy options. *Ex vivo* retroviral transfer of HIV-inhibiting genes into peripheral blood lymphocytes or CD34-positive human cells has been attempted, so far with little success. The therapeutic molecules used include (1) decoy-RNA specifying multiple copies of the Rev- or the Tat-responsive element, so-called poly-TAR or poly-RRE sequences, (2) mini antibodies (single chain Fv; scFv), able to capture viral gene products within the cell, (3) trans-dominant negative mutants of viral proteins, for example, RevM10, or (4) ribozyme RNA, which enzymatically cleaves RNA. Other genes still under development are designed to prevent entry or chromosomal integration of HIV. It remains to be shown whether such gene therapy approaches present a suitable therapeutic option compared with existing chemotherapy.

The best prevention of infectious diseases is achieved by prophylactic vaccines. Clinical trials using vectored vaccines based on ALVAC or MVA have been initiated. Other clinical trials pursue the goal of developing vaccines against HIV-1, malaria, hepatitis B, tuberculosis, and influenza A virus infections. Vaccination regiments using poxvirus vectors such as ALVAC or MVA in combination with naked DNA as a prime vaccine, sometimes followed by further booster injections of recombinant viral antigens, are being tested in humans. Such regiments have been shown to prevent disease progression after lentivirus infection of monkeys. This illustrates the complexity of vaccination strategies that are currently pursued in vaccine research.

3.6
Clinical Gene Therapy for the Treatment of Other Diseases

Clinical gene therapy can also be used for the treatment of diseases not necessarily caused by single known gene defects, if promising therapeutic genes can be reasonably applied. Patients with chronic rheumatoid arthritis, for instance, should benefit from a reduction of the inflammations in joints. Such inflammations are caused or at least maintained by a cascade of events including the overexpression and increased release of a number of inflammatory cytokines. Monoclonal antibodies able to reduce the local concentration of the tumor necrosis factor (TNF) have already been successfully used to treat disease. Here, clinical gene transfer approaches involve the transfer of autologous synovial cells modified *ex vivo* by a therapeutic gene encoding interleukin-1 receptor antagonist. Alternatively, adenoviral vectors with the same gene have been directly injected into the affected joint.

4
Manufacture and Regulatory Aspects

The regulation of gene therapy is very complex and differs considerably in the European Union and the United States. In Part IV, Annex I of Directive 2003/63/EC (which replaces Annex I of Directive 2001/83/EC), a definition of so-called gene therapeutics is given. As gene therapy not only includes therapeutic but also preventive and diagnostic use of vectors, nucleic acids, certain microorganisms and genetically modified cells, the term ''gene transfer medicinal products'' as used in the relevant European guideline ''Note for guidance on the quality, preclinical, and

clinical aspects of gene transfer medicinal products (CPMP/BWP/3088/99)'' seems more exact. An accurate listing of the medicinal products that belong to the group of gene transfer medicinal products can be found in the table contained in the guideline. The definition given in Sect. 1 of this article is in accordance with this guideline and is in agreement with the definition of gene therapeutic products of Directive 2003/63/EC. The annex of the latter Directive contains legally binding requirements for quality and safety specifications of gene transfer products. Although targeted at product licensing, these requirements may have a bearing on their characterization before clinical use. Active ingredients of gene transfer as defined medicinal products may include, for example, vectors, naked plasmid DNA, or certain microorganisms such as conditionally replicated adenovirus. For the *ex vivo* strategy, the active ingredients are the genetically modified cells.

Written approval by a competent authority in conjunction with positive appraisal by an ethics committee will in the future be necessary for the initiation of clinical gene therapy trials. Respective regulatory processes are currently established in all EU member states during transformation of Directive 2001/20/EC. The manufacture of clinical samples in compliance with good manufacturing practice (GMP) will become compulsory. Germline therapy is illegal in the EU. The law relevant for clinical gene therapy trials and manufacture of gene transfer medicinal products in Germany is the German Drug Law (AMG) and respective decrees and operation ordinances. The law governing the physicians' profession stipulates in the ''Guidelines on gene transfer into human somatic cells'' (''Richtlinien zum Gentransfer in menschlichen Körperzellen'') that the competent ethics committee may seek advice from the central ''Commission of Somatic Gene Therapy'' of the Scientific Council of the German Medical Association before coming to its vote. The Paul-Ehrlich-Institut is the competent authority in Germany and offers information on current clinical trial regulations.

Gene transfer medicinal products will be licensed via the centralized procedure by the European Commission. The licensing process is coordinated by the EMEA (''European Agency for the Evaluation of Medicinal Products'') following submission of a licensing application. The marketing authorization is governed by Council Regulation (EC) No. 2309/93. The recommendation in favor or against marketing authorization is made on the basis of Directives 75/319/EEC and 91/507/EEC by experts of the national competent authorities which are members of the ''Committee for Proprietary Medicinal Products'' (CPMP).

In the United States, the Center for Biologics Evaluation and Research ''(CBER) of the ''Food and Drug Administration'' (FDA) is responsible for clinical trial approval and marketing authorization.

The assessment of the licensing application focuses on the quality, safety, efficacy, and environmental risk of a gene transfer medicinal product. The manufacturing process has to be designed and performed according to GMP regulations. Like other biologicals, gene therapy products have considerably larger size and complexity compared to chemicals, and analysis of the finished product is not sufficient to control their quality and safety. A suitable process management, in-process control of all critical parameters identified within process validation are decisive factors. Gene transfer medicinal products

containing or consisting of genetically modified organisms are also subject to contained use regulations before licensing and until these organisms are applied to humans.

From the economic point of view, procedures for the manufacture of therapeutic DNA must be scalable and efficient, and at the same time simple and robust. Manufacturing processes are as manifold as the gene transfer methods used in gene therapy. As an example, manufacture of plasmid DNA for naked nucleic acid transfer can be briefly described as follows. The methods available for plasmid production today largely originate from lab procedures for the production of DNA for analytical purposes (mini preparations) and have been adapted to fit process scale. Toxic substances and those that present a hazard to the environment, expensive ingredients, and nonscalable methods must be avoided. In this context, the experience gained from industrial manufacture of raw materials with the aid of bacterial cultures and virus production for the purpose of vaccine production are useful for fermentation. Suitable methods for downstream processing above all include chromatographic methods with high

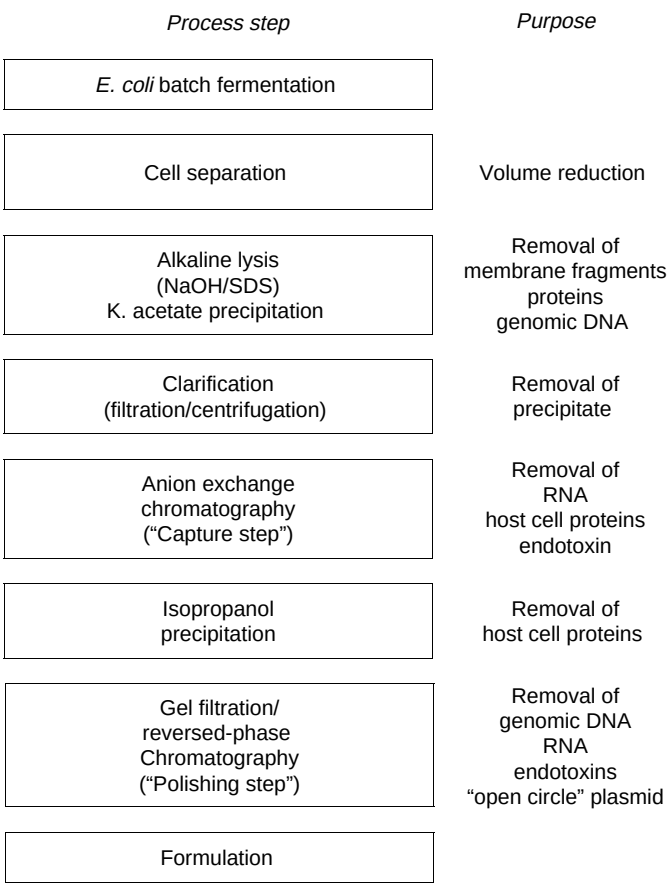

Fig. 4 Therapeutic plasmid DNA: typical manufacturing process.

Test	*Specification* (*method*)
Appearance	Clear colorless solution (visual inspection)
Size, restriction interfaces (identity)	Agreement with plasmid card (agarose gel electrophoresis, restriction enzyme assay)
Circular plasmid DNA (ccc)	>95% (Agarose gelelektrophorese, HPLC)
E. coli DNA	<0.02 µg/µg plasmid DNA (southern blot)
Protein	Not detectable (BCA protein assay)
RNA	Not (Agarose gel electrophoresis)
Endotoxin	<0.1 EU/µg plasmid DNA (LAL assay)
Sterility	No growth after 14 days (USP)
Specific activity	Conforms to reference standard (*in vitro* transfection)

Fig. 5 Therapeutic plasmid DNA: typical release specifications.

dynamic capacity and selectivity as well as high throughput.

In a typical procedure for the manufacture of therapeutic plasmid DNA (cf. Figs. 4 and 5), the first step is batch fermentation of *Escherichia coli* cells from a comprehensively characterized "Master Working Cell Bank" (MWCB). For this purpose, modern methods use high-density fermentation with optimized and safe *E. coli* K12 strains bearing a high number of copies of the required plasmid. The bacterial cells are harvested for further processing, resuspended in a small buffer volume, and lysed in an alkaline lysis procedure. By neutralization, the plasmid DNA is renatured while a large quantity of proteins, membrane components, and genomic DNA remain denatured. After separation of the precipitate by filtration, a chromatographic step can be performed as "capture step." Because of the anionic character of the nucleic acid, anion exchange chromatography (AEX) is the method of choice. In fractionated gradient elutions, differences in the charge enable the separation from contaminated RNA. Gel filtration (GF) or reversed phase (RP) steps can be used for fine purification. For final product analysis, evidence must be provided batch by batch that besides the correct identity and homogeneity, critical impurities like microorganisms, host cell proteins, genomic DNA, RNA, or endotoxins have been reduced below the specified limits. Removal of endotoxins is critical for *in vivo* gene transfer efficiency achieved with naked DNA.

Some established methods from protein chemistry can be used for processing therapeutic DNA. In parallel with the processing of proteins, however, the fact cannot be concealed that nucleic acids have some very specific properties. These include the extremely high viscosity of DNA solutions, the high sensitivity of nucleic acids to gravity, the low static and dynamic capacity of their chromatographic adsorption, and the ability to penetrate filtration media with porosities well below their molecular weight ("spaghetti effect").

After first experience, plasmid concentrations of ~200 mg L^{-1} fermentation broth can be obtained in high-density fermentation (optical density >50), corresponding to an yield of ~800 mg plasmid DNA per kg of dry biomass. Thus, from a fermenter of 1000 L usable volume, ~100 g plasmid DNA can be isolated per run in a batch fermentation at a purification yield of ~50%. Consequently, capacities for production of kilogram amounts can be built up with existing technologies.

5
First Experience with the Clinical Use of Gene Transfer Medicinal Products

The development of somatic gene therapy is still in its infancy. A number of theoretical risks of gene therapy have been listed, and numerous approaches and gene transfer methods are being developed in the clinic, even more in preclinical experiments.

In spite of this, a few SCID patients have been apparently cured by gene therapy using retrovirally modified bone marrow stem cells. At the same time, the occurrence of leukemia in two of the ~10 successfully treated children showed that theoretical risks cannot be clearly distinguished from clinically relevant risks due to the so far insufficient clinical experience. Trends, however, show that each pathological situation will require the development of a certain adapted gene therapy approach. Thus, in the long run, gene therapy will present real therapy or prevention options, especially for a number of up to now insufficiently treatable or untreatable diseases.

See also Gene Therapy and Cardiovascular Diseases; Liposome Gene Transfection; Medicinal Chemistry; Oncology, Molecular; Vaccination, Genetic; Vectors and Gene Therapy.

Bibliography

Books and Reviews

Anderson, F. (1995) Gene therapy, *Sci. Am.* **273**, 96B–98B.

Armentato, D., Sookdeo, C., White, G. (1994) Second generation adenovirus vectors for cystic fibrosis gene therapy, *J. Cell Biochem.* **18A**, 102–107.

Birnboim, H.C. (1983) A rapid alkaline extraction method for the isolation of plasmid DNA, *Methods Enzymol.* **100**, 243–249.

Buchholz, C.J., Stitz, J., Cichutek, K. (1999) Retroviral cell targeting vectors, *Curr. Opin. Mol. Ther.* **5**, 613–621.

Cichutek, K. (2000) DNA vaccines: development, standardization and regulation, *Intervirology* **43**, 331–338.

Cohen-Haguenauer, O., Rosenthal, F., Gansbacher, B., Cichutek, K., et al. (2002) Euregenethy network. Opinion paper on the current status of the regulation of gene therapy in Europe, *Hum. Gene Ther.* **13**, 2085–2110.

Culver, K.W. (1994).*Gene Therapy*, Mary Ann Liebert, New York

Green, A.P., Prior, G.M., Helveston, N.M., Taittinger, B.E., et al. (1997) Preparative purification of supercoiled plasmid DNA for therapeutic applications, *Biopharmacology* **10**, 52–62.

Gottschalk, U.. (1997) The Industrial Perspective of Somatic Gene Therapy, in: Müller, S., Simon, J.W., Vesting J.W. (Eds) *Interdisciplinary Approaches to Gene Therapy*, Springer.

Gottschalk, U., Chan, S. (1998) Somatic gene therapy. Present situation and future perspective, *Arzneimittelforschung/Drug Res.* **48**, 1111–1120.

Hodgson, C.P. (1995) The vector void in gene therapy, *Biotechnology* **13**, 222–229.

Jain, K.K. (1996) *Vectors for Gene Therapy. Scrip Reports*, PJB Publications, Surrey.

Kotin, R.M. (1994) Prospects for the use of adeno-associated virus as a vector for human gene therapy, *Hum. Gene Ther.* **5**, 793–797.

Marquet, M., Horn, N.A., Meek, J.A. (1995) Process development for the manufacture of plasmid DNA vectors for the use in gene therapy, *Biopharmacology* **10**, 26–37.

Marquet, M., Horn, N.A., Meek, J.A. (1997) Characterization of plasmid DNA vectors for use in human gene therapy. Part 1. *Biopharmacology* **10**, 42–50; Part 2. (1997) *Biopharmacology* **10**, 40–45.

Morgan, R.A., Anderson, W.F. (1993) Human gene therapy, *Annu. Rev. Biochem.* **62**, 191–217.

Müller, M. (2003) Considerations for the Scale-up of Plasmid DNA Purification, in: Bowlen,

B., Dürre, P. (Eds) *Nucleic Acid Isol. Meth.*, American Scientific Publishers, New York.
Mulligan, R.C. (1993) The basic science of somatic gene therapy, *Science* **260**, 926–932.
Nussenzweig, R.S., Long, C.A. (1994) Malaria vaccines: multiple targets, *Science* **265**, 1381–1384.
Prior, C., Bay, P., Ebert, B. (1995) Process development for the manufacture of inactivated HIV-1, *Pharm. Technol.* **19**, 30–52.
Randrianarison-Jewtoukoff, V., Perricaudet, M. (1995) Recombinant adenoviruses as vaccines, *Biologicals* **23**, 145–147.
Spooner, R.A., Deonarian, M.P., Epenetos, A.A. (1995) DNA vaccination for cancer treatment, *Gene Ther.* **2**, 1–11.
Tolstoshev, P., Anderson, W.F. (1993) Gene transfer techniques for use in human gene therapy, *Genome Res. Mol. Med. Virol.* **7**, 35–47.
Vile, R.G., Russell, S.J. (1995) Retroviruses as vectors, *Br. Med. Bull.* **51**, 12–15.

Primary Literature

Chandra, G., Patel, P., Kost, T.A., Gray, J.G. (1992) Large scale purification of plasmid DNA by fast protein liquid chromatography using a Hi-load Q sepharose column, *Anal. Biochem.* **203**, 169–177.
Gansbacher, B., Zier, K., Daniels, B. (1990) Interleukin-2 gene transfer into tumor cells abrogates tumorigenicity and induces protective immunity, *J. Exp. Med.* **172**, 1217–1222.
Horn, N.A., Meek, J.A., Budahazi, G., et al. (1995) Cancer gene therapy using plasmid DNA: purification of DNA for human clinical trials, *Hum. Gene Ther.* **6**, 565–573.
Mannino, R.J., Gould-Fogerite, S. (1988) Liposome-mediated gene transfer, *Biotechniques* **6**, 682–688.
Michael, S.I., Curiel, D.T. (1994) Strategies to achieve targeted gene delivery via the receptor-mediated endocytosis pathway, *Gene Ther.* **1**, 223–232.
Stitz, J., Buchholz, C.J., Cichutek, K., et al. (2000) Lentiviral vectors pseudotyped with envelope glycoproteins derived from gibbon ape leukemia virus and murine leukemia virus 10A1, *Virology* **273**, 16–20.
Stitz, J., Muhlebach, M.D., Cichutek, K., et al. (2001) A novel lentivirus vector derived from apathogenic simian immunodeficiency virus, *Virology* **291**, 191–197.
Wilson, J.W., Grossmann, M., Cabrera, J.A., et al. (1992) A novel mechanism for achieving transgene persistence in vivo after somatic gene transfer into hepatocytes, *J. Biol. Chem.* **267**, 11483–11489.

Spatial and Temporal Expression Patterns in Animals

Werner E.G. Müller
Abteilung Angewandte Molekularbiologie, Universität Mainz, Mainz, Germany

1 Introduction 272

2 Basis of Metazoan Pattern Formation 276
2.1 Stem Cells 276
2.2 Cell-to-cell Adhesion 279
2.2.1 Aggregation Factor 279
2.2.2 Aggregation Receptor 281
2.3 Cell-matrix Interaction: Tissue Formation 281
2.4 Cell Junction Molecules 282
2.5 Morphogens 283
2.6 Immunity, Individuality, and Apoptosis 285

3 Skeleton 286
3.1 Morphogenesis 286
3.2 Spicule Network in Sponges: A Unique Skeleton 289
3.3 Spicule Network in Sponges: A Unique Nervous System? 291

4 Axis-pattern Formation 293
4.1 Signaling Molecules: Wnt Pathway 294
4.2 Transcription Factors: T-box – Forkhead 295
4.3 Organizers 295
4.3.1 Spemann Organizer 296
4.3.2 Hensen's Node 296
4.3.3 Coelenterata 296
4.3.4 Porifera 296
4.4 Homeodomain Protein 298
4.4.1 Insects – Drosophila melanogaster 299
4.4.2 Vertebrates – Mouse 301

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

4.4.3 Coelenterates–Cnidaria 302
4.4.4 Porifera 303

5 Origin of Metazoan Pattern Formation 303
5.1 Serial Module Construction of the Body Plan in Diploblasts 304
5.2 Compartmental/Segmental Construction of the Body Plan in Triploblasts 305
5.2.1 Protostomia (Example, Insects) 305
5.2.2 Deuterostomia (e.g. Vertebrates) 306

6 General Principles in Metazoan Pattern Formation 307

Acknowledgment 309

Bibliography 309
Books and Reviews 309
Primary Literature 309

Keywords

Axis
It defines the polarity of an organism in various directions, for example, anteroposterior or apical–basal.

Bilateria
Multicellular animals comprising three germ layers (ectoderm, mesoderm, and endoderm). Their body plans are characterized by an anteroposterior and a dorsoventral axis, which is genetically controlled by gene clusters. Many structures occur in pairs.

Body Plan
Organization of an organism according to an axis.

Diploblasts
Animals comprising two germ layers (ectoderm and endoderm).

Metazoa
Multicellular animals composed of specialized cells that share labor.

Morphogenesis
A process that describes the formation of organs in a developing embryo.

Pattern
Entities of tissue units that are arranged during development in a well-ordered spatial pattern of cell activities.

Porifera (Sponge)
The phylogenetically oldest metazoan phylum, which evolved first from the common ancestor of all multicellular animals, the Urmetazoa.

Segments
Similar tissue units that are simultaneously formed during embryogenesis and arranged along an axis. These elements are remodeled during the development.

Serial Modules
Similar tissue units that are arranged along an axis and are formed in succession. These modules of the animals remain morphologically mostly unchanged during their development to adults.

Somites
Tissue units that are arranged along an axis and are formed consecutively. These elements are morphologically changed during development.

Triploblasts
Multicellular animals comprising three germ layers (ectoderm, mesoderm, and endoderm).

All metazoan animals comprise a body plan of different complexity. Since it is well established – especially on the basis of molecular and cell biological data – that all metazoan phyla, including the Porifera (sponges) evolved from a common ancestor, the fundamental common principles of pattern formation in Metazoa have been studied to a great extent. Common to all metazoan body plans is the formation of at least one axis that runs from the apical to the basal region; examples for this organization step are the Porifera and the Cnidaria (diploblastic animals). Animals belonging to the triploblasts, the Protostomia (model systems: insects and nematodes) and the Deuterostomia (mammals), comprise in addition to the anteroposterior axis also a dorsoventral axis that runs at a right angle to the first. Body plan formation starts from totipotent germ cells that lose their plastic differentiation propensity during their proliferation/differentiation. In the course of the developmental processes, a differential expression of cell–cell and cell–matrix adhesion molecules proceeds, which also involves the formation of cell junction molecules. After formation of tissue units, immune molecules are expressed, which allow the discrimination between self and nonself (establishment of the individuality), a process that is also controlled by apoptotic regulatory systems. Subsequently, extracellular solute factors (morphogens) and other secreted molecules in concert with transcription factors establish the axes. Organizer regions are formed which are localized in diploblasts (Porifera and Cnidaria) in the oscule or mouth region, respectively. In triploblasts, as it is well documented in the vertebrate

Spemann organizer or Hensen's node, cells are formed in a distinct region, which displays the property to induce a complete body axis. In parallel, skeleton formation (exo- or endoskeleton) supports the body construction.

The major difference in the genetic systems controlling body plan patterning between diploblasts and triploblasts lies in the fact that only the triploblasts (both Protostomia and Deuterostomia) comprise clusters of homeotic genes. In diploblasts, those clusters are missing and nonclustered paired-class genes and LIM/homeodomain genes control pattern formation. In metazoans, in diploblasts as well as in triploblasts, body plan formation starts by polarization of the embryo; the subsequent development is followed in diploblasts only to the stage of serial module patterning, while in triploblasts, the development to the adults includes compartmental and the subsequent segmental patterning.

1 Introduction

The multicellular animal kingdom of Metazoa comprises approximately 30 phyla that, as has been elucidated recently, share one common ancestor, the Urmetazoa; this term has now been widely adopted. Contrary to the metazoan organisms, in colonies of animal cells, for example, in choanoflagellates, there is no division of labor. In Metazoa, the cells have been differentiated to perform a series of functions, like digestion, sensation, contraction, and secretion. The determination of metazoan cells to a distinct fate occurs through an alteration in the pattern of gene activity and thus allows a specification to distinct roles. The result of such specialization is that nutrients are taken up by distinct feeding cells and from there are transported to nonfeeding cells, which accomplish different functions. Specialized cells are the basis for pattern formation, a process during which a spatial and temporal pattern of cell activities is organized within the well-ordered organism. Most animals are characterized by a bilateral symmetry and comprise regulatory genes, which are involved in head formation. Metazoa are grouped into the morphologically more diverse bilaterians, which themselves are further subdivided into Protostomia, including Ecdysozoa and Lophotrochozoa, and Deuterostomia and the nonbilaterian metazoans (phyla Porifera, Placozoa, Coelenterata, and Ctenophora).

During pattern formation a body plan is established, which also defines the main axes of the multicellular animal. From Porifera to Arthropoda (Protostomia) and Vertebrata (Deuterostomia) the level of complexity of the axes (anteroposterior ends (oral–aboral polarity), dorsal–ventral sides) increases. The two phyla Porifera and Coelenterata that possess only two epithelial layers, the ectoepithelium surrounding the body and the endoepithelium that encloses the digestive cavity(ies), are termed *diploblastic animals* and are defined by one apical–basal polarity. The triploblastic, bilaterian animals are composed, in addition to these external layers (the ectoderm and the endoderm) of a middle, mesodermal, cell layer that originates usually from the endoderm. Pattern formation can be studied during embryogenesis

or during differentiation of embryonic cells, for example, in three-dimensional cell cultures. These morphogenetic processes are based on and controlled by differential spatial and temporal expressions of genes that initiate or maintain a large number of signaling pathways. Activation of a specific pathway can determine the character of a specific region as in the embryonic organizer, like the "Spemann organizer" or the "Nieuwkoop center." In recent years, it has been established that the function of their key signaling pathways act by negative regulation. The tuned interactions result, for example, in the frog blastula/gastrula stages, in a sequential activation of the centers; like in the Nieuwkoop center in the ectoderm of late blastulae, which releases signals required for the activation of the Spemann organizer in the mesoderm of the gastrulae. This example shows that the timing and position of such centers is crucial and allows the synchronization of developmental processes, which drive the correct organization of the future tissue in an organism.

Segmentation (a process during which very similar functional units along a body axis are formed almost simultaneously (in insects)) or somite formation (the units along a body axis are formed sequentially (in vertebrates)) is a feature of triploblastic organisms. Each segment/somite is arranged along the anteroposterior axis. The orientation is controlled by characteristic sets of genes; for example, in insects the parasegments are delimited by the function of the *pair-rule* genes and the subsequent segments by the *segment polarity* genes. In these triploblasts, each segment/somite has its own identity, which is defined by master regulatory genes, termed *homeotic selector* genes. These homeotic genes are arranged into gene clusters. In diploblasts, Coelenterata and Porifera, ancestors of these genes exist as single molecules. It remains, however, enigmatic by which mechanism the organization of the gene clusters present in triploblasts evolved from the single gene of the diploblasts.

In the past few years, it has been elaborated that the Porifera (sponges) as the phylogenetically oldest metazoan phylum comprise already the basic structural and functional elements required for the construction of a body plan; they evolved approximately 800 million years ago. According to the inorganic composition of their skeleton (the elements of their skeleton are termed *spicules*), sponges are grouped into the classes of Demospongiae and Hexactinellida, which possess hydrated, amorphous, noncrystalline silica spicules and the class Calcarea, whose skeletal spicules are formed from calcium carbonate. The Demospongiae and Hexactinellida are the only two taxa within the kingdom of Metazoa which primarily utilize silica instead of calcium in their mineral skeleton; calcium is otherwise the dominant inorganic skeletal component in Metazoa.

Like in the second phylum of the diploblasts, the coelenterates, also the sponges are built from two major epithelial layers. In sponges, the exopinacoderm (formed by exopinacocytes) represents the external epithelium that delimits the organism toward the environment and the endopinacoderm (endopinacocytes), which lines the aquiferous canals. The water flow is directed from the lateral surface of the specimens, through the porocytes and the canals to the lacunae and the choanoderm (choanocyte chambers); Fig. 1. These chambers are composed of two kinds of epithelial cell layers, formed by choanocytes (flagellated cells) and by cone cells (cells with a double-conical shape which hang into the

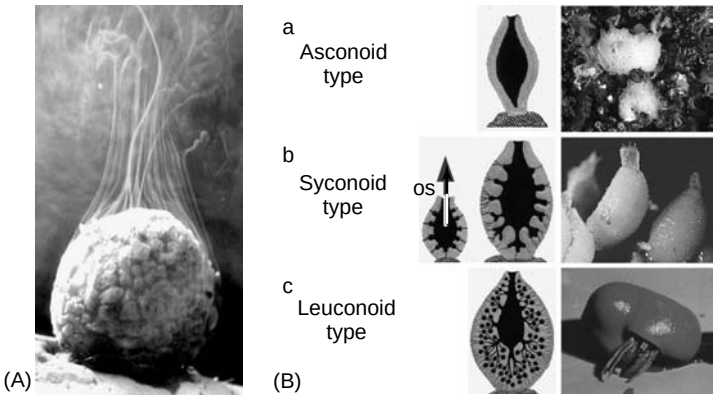

Fig. 1 Body plan organization in the phylum Porifera. (A) Sponges, here *Tethya lyncurium*, have the potency to conduct water from the external milieu through their inhalant openings at the surface to the exhalant apertures, the oscules. This highly complex aquiferous system is formed from differentiated somatic cells, which interact in a tuned manner through cell–cell/cell–matrix interaction. The water flow through a specimen has been visualized with fluorescein; the pulsatory extrusion of the water current can be seen. (B) The three different organization/body plan patterns of sponges are shown in the middle panel. The choanoderm layer is in red and the pinacoderm layer is in blue. (a) In the asconoid type, the sack-like organization shows an unfolded choanoderm. (b) In the syconoid type, the folding of the choanoderm layer results in the formation of choanocyte chambers, which are connected with the incurrent canals. After passing the chambers, the excurrent canals lead the water current to the oscule (os). (c) The organization of the leuconoid type. This complex canal network is composed of discrete choanocyte chambers, connected with a network of incurrent and excurrent canals. In all three types of body organization, a basal site (attachment to the substratum) and an oscular/apical pole (oscule; opening at which the central cavity which is connected with the excurrent canals extrudes the water) exists, which determines the axis (arrow). Representative species of the different organization levels are shown: (a) *Clathrina coriacea*, a calcareous sponge of the asconoid type. Magnification x0.2. (b) A syconoid type of organization as in *Sycon raphanus* (Calcarea); x3. (c) *Suberites domuncula* a siliceous sponge (Demospongiae) of the leuconoid type; x0.5. (See color plate p. xxvi).

chamber formed by choanocytes). Therefore, these chambers can be considered as organ-like assemblies. The function of the choanocyte chambers as organs or organ-like assemblies is to orient the water flow unidirectionally from the incurrent to the excurrent canal to allow the extrusion of the water from the body through the exhalant oscule(s). The two epithelial layers of the sponges surround the mesohyl, which comprises differentiated cells that interact in a coordinate way. In this central body of the sponges, the cells are loosely embedded in a ground matrix, composed primarily of collagen, galectin, and glycoconjugates. This matrix surrounds motile archaeocytes, which are pluripotent, and other differentiated cells, for example, the collencytes and lophocytes, which are involved in the formation of collagen as well

as the sclerocytes that form the spicules. In addition, myocytes exist which allow contraction of the sponge body upon mechanical irritation.

According to the folding pattern of the two epithelial layers of the pinacoderm, in correlation to the choanoderm, three types of sponge complexity are distinguished; the asconoid, the syconoid, and the leuconoid type. The folding of the choanoderm allows the formation of spherical chambers, the choanocyte chambers of different structural degrees (Fig. 1). In the asconoid type, the lateral opening(s) direct the water current to the continuous layer of choanocytes which face the single atrium and then to the oscule (Fig. 1B-a). This type of organization is seen in some Calcarea, for example, *Clathrina coriacea*. In the syconoid type, the folding pattern forms choanocyte chambers and the water flow enters the sponge through the porocytes into choanocyte chambers through which the water is pressed into the atrium and finally via the oscule again to the external milieu (Fig. 1B-b). This type of organization is seen in Calcarea and Demospongiae; as an example the calcareous sponge *Sycon raphanus* is shown here. The leuconoid type derives from the syconoid type of organization (Fig. 1B-c). Here, the porocyte openings lead from the vestibule to the connecting incurrent canals and into the choanocyte chambers. Subsequently, the water current is pressed into excurrent canals that open into the atrium and leaves the organism via the oscule (Fig. 1B-c). Most sponge species, like the marine (siliceous) demosponge *Suberites domuncula* (Fig. 1B-c) show the leuconoid type of organization. The Hexactinellida have an organization pattern that is reminiscent of the syconoid type and are composed of a choanoderm syncytium.

By molecular biological techniques – in the meantime over 15 000 ESTs (expressed-sequence tags; http://spongebase.genoserv.de) have been elaborated from the sponge *S. domuncula* – and subsequent identification of the functions of sponge genes, it has been disclosed that the basic strategies of body plan formation and of gene expression patterns found in Porifera are characteristic of Metazoa in general. Furthermore, an *in vitro* 3D-cell culture system, termed *primmorphs*, was established. Primmorph cultures are a newly developed technique to grow sponge cells *in vitro*; their cells have the potency to proliferate and to differentiate. This system allows an understanding of the roles of the morphogenetic and pattern-forming genes. The outcome was: first, Porifera are the phylum that is most closely related to the common ancestor of all metazoans, the Urmetazoa; and second, their body plan is controlled by pathways that remained conserved from Porifera to the crown taxa, human or *Drosophila melanogaster*/*Caenorhabditis elegans*. Amazing is the finding that the complexity of the sponge genome, which comprises over 100 000 genes, is high, and that sponge sequences share a high similarity to vertebrate genes. Therefore, it can be proposed that the developmental regulatory genes/proteins present in Porifera are representative of the related molecules found in higher metazoan taxa. The major distinction between Porifera and Cnidaria/Ctenophora on one side and the Bilateria on the other is the number of distinct tissue layers (diploblastic vs. triploblastic animals); Fig. 2.

In the following sections, the different aspects of pattern formation of Metazoa are highlighted, referring to the Porifera as the model organism for multicellular animals.

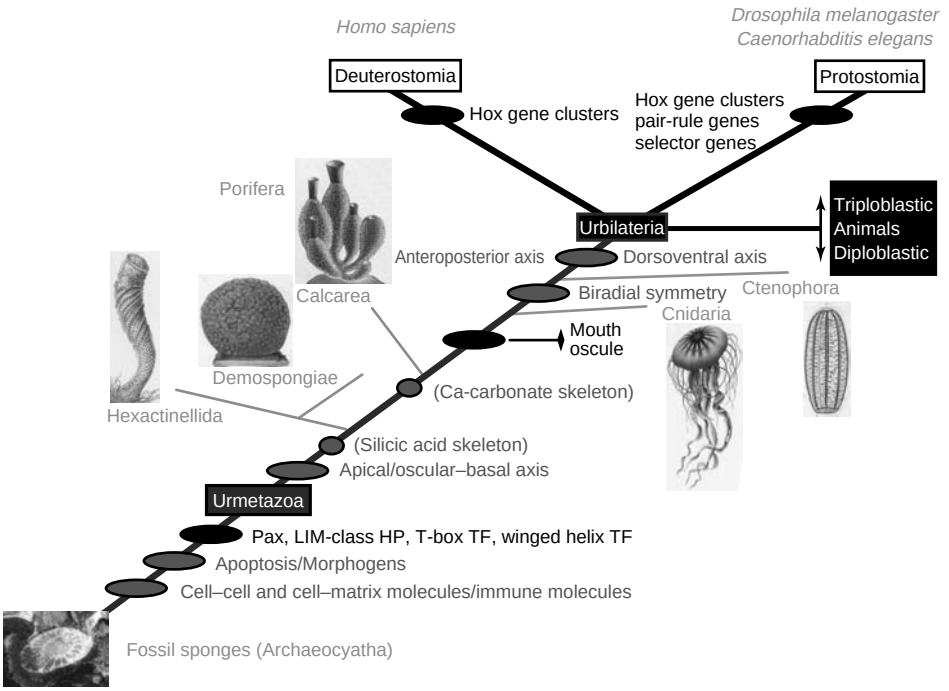

Fig. 2 Phylogenetic position of the Porifera within the metazoan kingdom. As first branches, the three poriferan classes (Hexactinellida and Demospongiae emerged and finally the Calcarea) evolve for the common ancestor of all metazoan phyla, the Urmetazoa. The major evolutionary novelties, which have to be attributed to the Urmetazoa, are those molecules that mediate apoptosis, control morphogenesis, the immune molecules, and cell adhesion molecules. The three classes of Porifera are the model systems that comprise a genetic reservoir of pattern formation characteristic of Metazoa; for example, the transcription factors, like paired box homeodomain molecules, LIM-class homeodomain, T-box (Brachyury) or winged helix (Forkhead). As a sister group to the Calcarea, the Cnidaria evolved. Subsequently, the Ctenophora emerged, which comprise not only an oral/aboral polarity but also a biradial symmetry. In the Cnidaria, the paired box transcription factors have been identified. Finally, the Urbilateria developed from the diploblasts (two epithelial layers), which are built from three germ layers (triploblasts). They diverged into the Protostomia, with the crown species *D. melanogaster* and *Caenorhabditis elegans*, and the Deuterostomia with the humans. Within the triploblastic animals, the homeobox genes are arranged in clusters. (See color plate p. xxvii).

2 Basis of Metazoan Pattern Formation

2.1 Stem Cells

Multicellular animals are characterized by the existence of a series of differentiated somatic cells, in addition to the totipotent germ cells. During evolution from the lowest metazoans, the sponges, to the crown taxa the number of distinct cell types increased steadily. The different somatic cells types derive from the zygote through the respective stem cell stage; the differentiation processes proceed in niches, which form a microenvironment in which

defined expression patterns of signaling molecules and local environmental factors direct the fate of the cells. These niches modify their regulatory properties in response to a changing environment to ensure that stem cell activities meet the needs of an organism for a given differentiated cell type. Two major types of stem cells derive from the zygote; the "germline" stem cells and the "somatic" stem cell. While the germline stem cells retain their total differentiation capacity, these properties are restricted in the somatic stem cells, which gradually lose their stem cell propensity. However, recent studies indicate that this traditional view of an irreversible loss of the stem cell ability during maturation of somatic cells might not completely reflect the physiological situation. It appears that such fixed "points of no return" do not always exist, but that at all levels of differentiation from the pluripotent progenitor cells to the committed progenitors, to the lineage progenitors, and finally to the terminally differentiated cells the propensity to act as a stem cell is retained, even though with a decreasing propensity. In consequence, the differentiation lineages of somatic cells are dynamic and plastic, and the strong distinction between embryonic stem cells and adult stem cells should be reconsidered.

With this new stem cell concept, one can ask for the simplest models of metazoan stem cells. The systems of choice are those organisms that already comprise a body plan organization and that reproduce sexually. These criteria are fulfilled in sponges. It was first discussed in sponges that some lineages of their somatic cells might not have lost the property of re- and transdifferentiation, a finding that became only increasingly obvious in higher metazoans approximately during the last five years.

It is generally agreed that the archaeocytes are the toti-/pluripotent cells in sponges from which the other cells originate; recently, evidence was presented indicating a localization of archaeocytes not only within the mesohyl but also in the endopinacoderm layer. Archaeocytes give rise to the major classes of differentiated somatic cells, (1) the epithelial cells, pinacocytes, and choanocytes; (2) the cells forming the skeleton, collencytes, and sclerocytes; and (3) the contractile cells, myocytes (Fig. 3). It had been operationally outlined that the sperm and egg cells derive from the archaeocytes. It appears that in Porifera the archaeocytes give rise not only to the different somatic cells (epithelial-, skeletal-, and contractile cells) but also to the germ cells from which the embryos originate. Another line of differentiation of the archaeocytes is to the thesocytes, the totipotent cells of gemmules, which are asexually propagative dissemination bodies. Pinacocytes, collencytes/sclerocytes and myocytes are cells with a low stem cell propensity, implying that these somatic cells are "terminally" differentiated.

In order to underline the view that the metazoan stem cell concept can also be applied for Porifera, characteristic marker genes have been cloned from *S. domuncula*. The first cDNAs identified whose deduced proteins share sequence similarity to mammalian stem cell markers, were the mesenchymal stem cell-like protein (MSCP) and noggin. MSCP is present in mesenchymal human stem cells; experimental evidence exists that MSCP is expressed in osteogenic mesenchymal stem cells. Functional studies revealed that in sponges the expression of this gene is under positive control of the morphogenetic inorganic elements, silicon, and ferric iron. In addition, two further potential genes

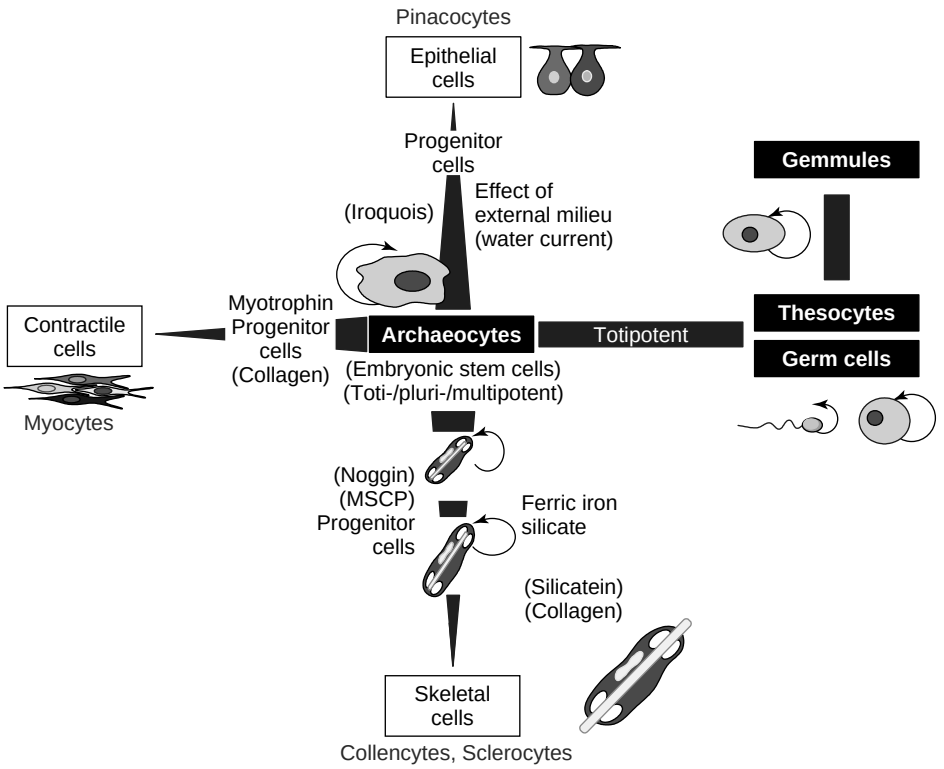

Fig. 3 Sponge stem cell system. Schematic outline of the development of the toti-/pluripotent sponge embryonic stem cells, the archaeocytes, to the germ cells on one side and to the three major differentiated cell types, the epithelial-, the contractile-, and the skeletal cells. It is indicated that during these transitions progenitor cells characteristic of these lineages have to be passed. The (potential) factors, for example, noggin and the mesenchymal stem cell–like protein (MSCP) on the path to the skeletal cells, which trigger the differentiation, are shown. In addition, it is outlined that committed progenitor cells are formed, which respond to the silicate/Fe(+ + +) stimulus through differentiation to skeletal cells, the sclerocytes (= skeletal cells). Without losing the high level of stem cell propensity, the archaeocytes change to germ cells and also to thesocytes, the dominant cells in gemmules (asexual propagation bodies). The triangles schematically indicate the decrease in stem cell propensity during the differentiation process.

involved in the differentiation of stem cells in sponges were isolated; noggin and the glia maturation factor. Noggin is a glycoprotein that binds bone morphogenetic proteins (BMP) selectively and antagonizes their effects. It was initially isolated from Xenopus and found to be expressed in the Spemann's organizer. In vertebrate development, noggin is involved in the formation of dorsal mesoderm derivatives, for example, the skeletal muscles.

In the initial phase of formation, the *in vitro* 3D-cell primmorph system contains predominantly the toti-/pluripotent archaeocytes, which can be stimulated to differentiate into four main tissue-specific directions. If they are induced by the

inorganic factors silicate or ferric iron, the archaeocytes give rise to the skeletal cells through an increased expression of genes encoding the structural proteins silicatein and collagen; a process that is mediated by noggin and MSCP (Fig. 3). Second, if archaeocytes are exposed to a morphogen (myotrophin), they are directed toward the contractile cell lineage. A third lineage, which gives rise to the aquiferous canal system, is induced by a physical factor; there *Iroquois* gene expression is induced as a result of an increased water current, which is paralleled with the formation of canal-like pores in the primmorphs. In higher metazoans, the expression of the *Iroquois* genes is thought to confer identity to a particular region, and hence it can be classified to the homeotic selector genes. The early function of *Iroquois* in insects includes the definition of the eye and notum territories, and in vertebrates, it includes the formation of the neural ectoderm. Late functions of *Iroquois* are subdividing the territories; the dorsal–lateral subdivision in *Drosophila* notum and the patterning of notum bristles and wing veins, and in vertebrates, the subdivision of neural tube and heart patterning. The sponge *S. domuncula* contains in its genome an *Iroquois* gene, which not only has remarkable sequence similarity with those sequences described from triploblastic animals but also expresses it during the formation of one major morphogenetic remodeling process, the construction of the aquiferous canal system (Fig. 3).

In general, toti-/pluripotent stem cells have the potential for extensive proliferation coupled with a high degree of differentiation capacity. The stem cells give rise to somatic cells that are – to a large extent – restricted to their differentiative and regenerative potential. The key role of the stem cells is to establish and maintain tissue homeostasis. The character of the stem cells and also of the differentiated somatic cells depends both on direct adhesion/contact to the adjacent cells and also on solute factors.

2.2 Cell-to-cell Adhesion

For more than 100 years, sponges have been a classical model for basic studies to understand metazoan cell–cell adhesion. Primarily, the marine demosponges *Microciona prolifera*, *Geodia cydonium*, and S. domuncula are the most thoroughly studied species.

2.2.1 Aggregation Factor

In 1973, it had been a success to isolate and purify both from *M. prolifera* and *G. cydonium*, the first extracellular particle termed *aggregation factor* (AF), which promotes the species-specific aggregation of sponge cells (Fig. 4). The AF was characterized as high molecular–weight complexes, which are assembled from a series of proteins that are bound to the core structure both covalently and noncovalently. The AF that mediates cell–cell aggregation depends on the presence of Ca^{2+} and binds to the membrane-associated aggregation receptor (AR).

The molecules in the very complex and dynamic cell–cell and cell–matrix recognition in sponges have been obtained by molecular cloning, for example, the galectin gene as the first cell–cell adhesion molecule, and integrin as the first cell–matrix adhesion receptor in *G. cydonium*. These sponge molecules are closely related to those proteins known to promote adhesion in Protostomia and Deuterostomia. Sequence analyses of the galectin from *G. cydonium* revealed that those amino acid (aa) residues that are involved

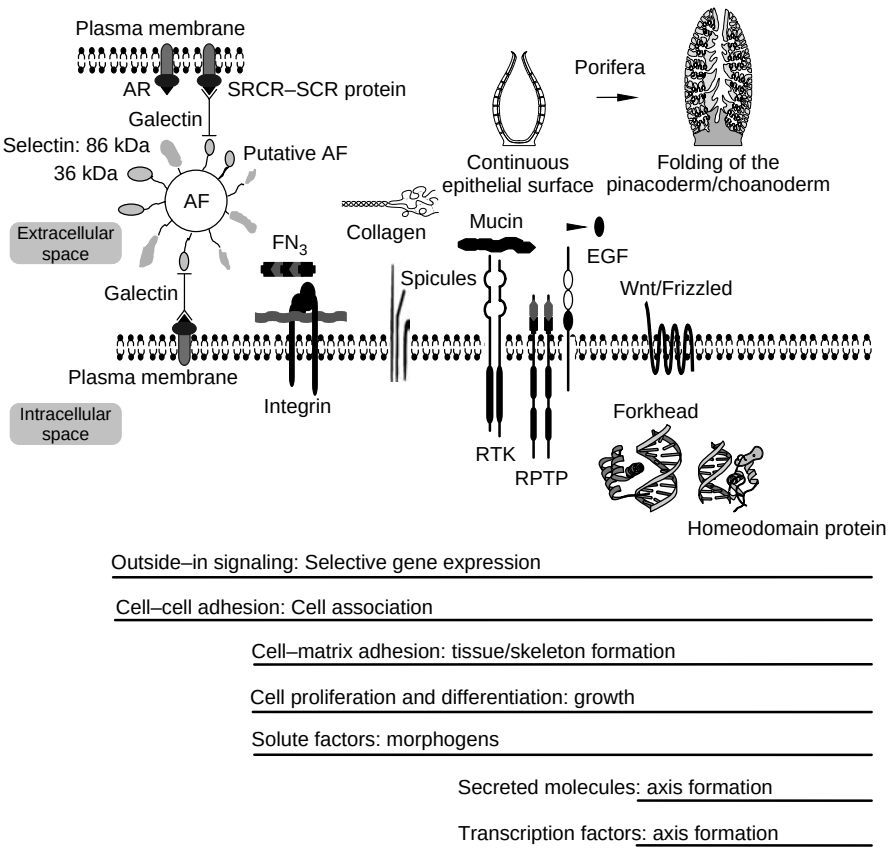

Fig. 4 Different phases required for the initiation of pattern formation (schematic model based on studies on the demosponges *Geodia cydonium* and *Suberites domuncula*). First, cell–cell and second cell–matrix adhesion molecules allow the mechanical interaction between adjacent cells or cells and extracellular matrix molecules. Aggregation factors (AF) in the extracellular space interact with aggregation receptors (AR) and mediate the first mechanical contact. The AF and the AR are complexes that are held together by protein–protein or protein–glycoprotein interactions. The AF-mediated cell–cell recognition is species-specific and very likely controlled by the complex SRCR/SCR-AR, which is embedded in the plasma membrane. Third, after this primary AF-AR-AF contact, intracellular signal transduction pathways are activated, resulting in a selective gene expression. New insertion of adhesion receptors into the plasma membrane follows; for example, integrin and other receptors, involved in tissue and skeleton formation. Like those, the receptor tyrosine kinases (RTK) and the receptor protein-tyrosine phosphatases (RPTP) also are uniquely found in Metazoa. Together with these receptors their ligands are synthesized, for example, collagen, molecules containing the fibronectin FN_3 modules, or mucinlike molecules, which establish the cell–matrix adhesion system. During this phase, growth factors are synthesized, for example, the precursor and the mature epidermal growth factor (EGF), which interact with the newly synthesized receptors. Fourth, solute molecules that initiate axis formation are released. Fifth, after completion of these phases pattern formation can start, a process that is controlled by ''morphogenetic'' cell surface receptors, like Frizzled, and by transcription factors, for example, Forkhead. Also, LIM-class homeodomain transcription factors are activated. It is suggested that in parallel with the expression of these gene cascades, the complexity of the sponge body plan, as an example, increases (upper right).

in binding of galectins from mammalians to galactose are also conserved in the sponge sequence. The sponge galectin is one polypeptide that is associated with the adhesion system in sponges; it links this AF complex to the membrane-associated aggregation receptor. These molecular biological and cell biological findings have been taken as one major clue for the now established view that Metazoa, including Porifera are of monophyletic origin with Urmetazoa as the hypothetical ancestor; Fig. 2.

2.2.2 Aggregation Receptor

The putative AR was cloned from *G. cydonium* and found to comprise fourteen SRCR domains, six SCR repeats, a C-terminal transmembrane domain and a cytoplasmic tail. More recently, one further molecule of the AF complex in the *G. cydonium* system was characterized, which showed a distant relationship to amphiphysin. Interestingly, the AR displays high sequence similarity to the macrophage type I scavenger receptor, macrophage M130 antigen and complement factor I, further supporting the fact that the adhesion molecules are highly conserved from Porifera to Chordata.

The AF and the corresponding AR are not merely adhesion molecules but also cause activation of signal transduction pathways resulting in an increased cell metabolism and also a higher cell proliferation and cell differentiation. The most obvious testimony for these properties can be deduced from the fact that single cells from sponges are not able to divide. Only if cell–cell contact is allowed do they form the 3D-cell cultures, the primmorphs, which contain both proliferating and differentiating cells.

2.3 Cell-matrix Interaction: Tissue Formation

Like in most metazoans, in sponges also the dominant molecule present in the extracellular matrix (ECM) that functions as a cell–matrix adhesion molecule is collagen. The genes were cloned both from the freshwater sponge *Ephydatia muelleri* as well as from *S. domuncula*; the sponge collagens belong to the short chain collagens. In protostomians as well as in deuterostomians, collagen binds to the cell surface receptor integrin. As a first nonbilaterian integrin, the sponge molecules have been identified in *G. cydonium* as well as in *S. domuncula* and the function was partially characterized. A second ligand for the integrin receptors is fibronectin, from which the FN_3 module has been identified (Fig. 4).

In addition, a series of further membrane receptors are required for tissue homeostasis and the subsequent pattern formation. Cell surface–spanning receptors, in particular, the single- (1-TMR) as well as the seven-pass transmembrane receptor proteins (7-TMR), serve as receivers for extracellular signaling molecules. Also, these molecules could be identified in sponges and represent autapomorphic characters for all Metazoa. The sponge1-TMR, the receptor tyrosine kinase (RTK), comprises the following segments; the extracellular domain (ligand-binding domain), the transmembrane domain, the juxtamembrane domain, and the catalytic domain. Its expression is upregulated during tissue fusion processes of autografts. The extracellular domain is composed of two immunoglobulin (Ig)-like segments, closely related to mammalian immunoglobulins, which are also polymorphic in the sponge molecule. The RTK interacts

with a mucin (Fig. 4). The corresponding dephosphorylating receptor, the receptor protein-tyrosine phosphatase (RPTP), was identified in *G. cydonium*. As a 7-TMR, the metabotropic glutamate/GABA-like receptor, a neuronal receptor, was characterized in *G. cydonium*; it responds to extracellular agonists as well as antagonists known from metabotropic glutamate/GABA-like receptors present on nerve cells from mammals. Another 7-TMR, the Frizzled receptor, has been found in *S. domuncula*. We will refer to it later, as a component of the Wnt signaling pathway.

Adhesion receptors, for example, integrins; or cell surface receptors, for example, RTKs, are evolutionary novelties and are found only in Metazoa. These molecules are more than merely mechanical linkages between structural molecules of the ECM and cells, collagen integrin, or other extracellular ligands and receptors, for example, mucin-RTK. During mechanical interaction between the extracellular molecule and the cell surface receptor, the receptor is activated; either, the receptor dimerizes or undergoes a conformational change and becomes enzymatically active, or both. It is amazing how manifold this interaction at the outside region of the cell surface receptor can be. The integrin receptor heterodimerizes with α- and β-subunits. Both subunits comprise a family of different molecules; hence, the number of possible combinations between the individual members is huge. Depending on the dimer formed the subsequent intracellular second messenger generation is different. This so termed *outside-in* signaling through cell surface receptors is also characteristic of 1-TMRs, for example, in RTKs. If an RTK interacts with an extracellular specific stimulus from outside, it undergoes dimerization resulting in a binding-induced conformational change; it then becomes enzymatically active and couples specifically with intracellular pathways. Determined by the members involved in the ligand–receptor complex, different outside-in signaling processes are activated that cause the initiation of a specific gene expression program.

2.4 Cell Junction Molecules

Formation of tight connections between cells is imperative to establish and maintain the tuned interaction between cells in a developing metazoan embryo or tissue. Such connections can be based on direct cell–cell interactions or on solute molecules (morphogens) and also on hormones. Cell junction molecules give the mechanical support for epithelial sheets, function as permeability barrier and, more importantly, provide communication pathways between adjacent cells. Several classes of junctions have been developed in Metazoa; "occluding junctions," which seal cells together in an epithelial cell sheet allowing the formation of a selective permeability barrier (e.g. tight junctions), "anchoring junctions," which mechanically attach cells through the actin- or intermediate filament network (focal contacts, desmosomes), and "communicating networks" (gap junctions, synapses). The tight junctions seal adjacent cells in a specific way, allowing the transport only of selected nutrients across a given cell sheet. Tight junctions in the epithelia allow the formation of an intraorganismic milieu, which is different from the external one and maintain this homeostasis. Especially for aquatic organisms, the selective-barrier function of cell sheets was one prerequisite for the formation of integrated multicellular animals. Hence, it was not surprising to find that this form

of junctions, that is, tight junctions, is evolutionarily conserved from sponges to the triploblastic crown species.

As in other metazoans, the tight junction scaffold proteins from sponges comprise the characteristic scaffold protein, a membrane-associated guanylate kinase with inverted arrangement (MAGI; Fig. 5). The sponge MAGI scaffold protein comprises six PDZ domains that are involved in protein–protein interaction, two WW domains that bind to proline-rich peptide motifs, and the conserved guanylate kinase motif. In addition, the existence of one tetraspan receptor, tetraspanin, in *S. domuncula* has been demonstrated. The tetraspanins belong to a group of hydrophobic proteins, comprising four transmembrane domains with a series of conserved aa residues in the extracellular loops. By *in situ* hybridization it had been shown that the *MAGI* gene is highly expressed in the epithelial cell layer and in the cells surrounding the canals; again, this result also shows that from sponges to higher metazoans, the basic requirements for tissue and subsequent pattern formation exists.

2.5 Morphogens

Pattern formation can be caused by a regional/segmental expression of a given gene; the protein formed acts in this region usually in a "yes" or "no" manner. In addition, morphogenesis can be driven by morphogens. These soluble factors are synthesized at very restricted sites and diffuse from there under formation of a smooth concentration gradient. Cells arranged along this morphogen gradient

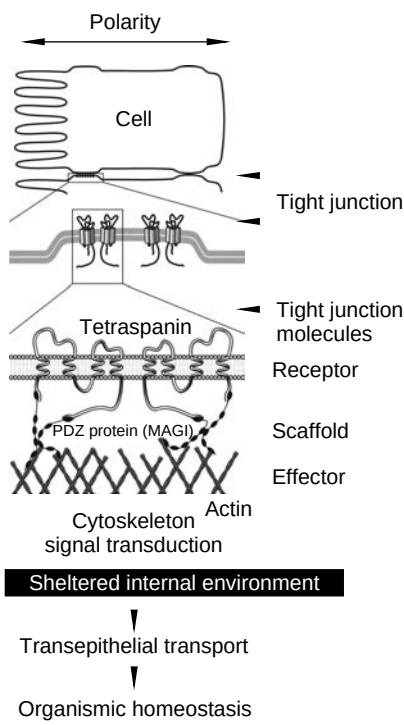

Fig. 5 Elements of tight junction proteins (here from *S. domuncula*). This scheme shows the molecules that seal the epithelial cell layer to control the lateral-extracellular transport in the aqueous milieu. By formation of tight junctions, cells undergo polarization. The tight junctional cell membrane–spanning receptors, here tetraspanin, associate with the scaffold protein, the PDZ protein MAGI. The scaffold protein MAGI plays a crucial role in the organization of the membrane receptor molecules and the effector molecules; the latter compose the cytoskeleton and the signal transduction molecules. Finally, tight junctions allow organismic homeostasis to occur.

react to it in a threshold-like manner. The decrease of the morphogen concentration provides the cells with positional information to which they can respond depending on their genetic program. The consequence of these specifying parameters, concentration, and genetic predisposition (activation of transcription factors), is that sharp distinctions arise in an initially uniform population of cells. Morphogens have been identified in all taxa studied and also in Porifera, for example, myotrophin (Fig. 4).

Myotrophin was first found in mammalian systems; in cardiac myocytes, it stimulates protein biosynthesis, suggesting a crucial role in the formation of cardiac hypertrophy. In *S. domuncula*, the gene encoding myotrophin has been found to multifold stimulate protein synthesis and induce growth of primmorphs to an elongated, oval shape. Furthermore, it could be demonstrated that in the presence of recombinant myotrophin, the cells upregulate the expression of the collagen gene. It is worth noting that the highest myotrophin gene expression is seen in the region of the oscule.

Recently the epidermal growth factor (EGF) was identified in *S. domuncula*. In chordates, EGF interacts with a series of different RTKs resulting in a regionally different cell response. In order to prove if EGF causes a differential gene expression in tissue depending on the regions from where the samples were taken, parabiosis studies were performed. Tissue samples were taken from the oscule region and the basal region of *S. domuncula* (Fig. 6a) The two tissue samples (oscule-basis) were placed together and attached with nylon fibers (Fig. 6b); between the two tissue samples, affi-gel blue beads were interspersed (Fig. 6c). The beads remained uncoated (controls) or were coated with recombinant EGF. After an

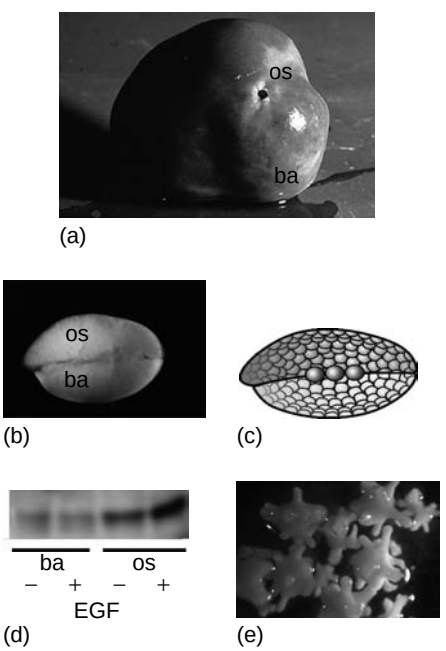

Fig. 6 Morphogenetic activity of the epidermal growth factor (EGF) in the homologous sponge system. (a) A specimen of *S. domuncula* with the pronounced oscule (os), the organizational center of sponge; below the oscule an atrium opens, a cavity where the excurrent canals meet; x2. (b) Tissue samples from the surface of the oscule region (os) and the basis (ba) of an animal tied together with nylon fibers; x2. (c) Between the two caps beads were interspersed, which remained either untreated or were coated with recombinant EGF. (d) After a 24-h incubation period, extracts from the oscule (os) and the basal region (ba) were prepared and equal amounts of protein were determined for the level of Brachyury by Western blotting. (e) Incubation of primmorph for three days in the presence of EGF; the shape of the primmorphs changed from round/oval to lobular; x2.5. (See color plate p. xxiii).

incubation period of 24 h, both the oscule and the basis tissue sample were taken, homogenized and the level of the morphogenetically active transcription factor Brachyury was determined by Western blotting (Fig. 6d). The results show that the expression level of Brachyury in untreated controls (without Brachyury) is low in the basal region of the sponge, in comparison to the level determined in the oscule region. After incubation with recombinant EGF, the level in the basis remained unchanged while it strongly increased in the sample region from the oscule. This finding indicates that predominantly cells from the oscule region have the ability to respond to EGF, while cells from the other region remain silent. As a direct proof that EGF causes oscule-like processes, primmorphs were incubated with low concentrations of recombinant protein for three days. During that period, the round/oval-shaped primmorphs changed to a lobular morphology (Fig. 6e). These data suggest that EGF comprises a morphogenetic potential that leads to the formation of oscules in the described homologous system.

Endogenous retinoids can be considered as morphogens that are involved in the suppression of head formation and activation of expression of posterior-specific genes in amphibians. A retinoid-responsive transcription factor, the retinoid X receptor (RXR), was identified in *S. domuncula*, as well as other elements of the retinoic acid pathway, for example, the ß,ß-carotene-15,15′-dioxygenase. The sponge RXR comprises two motifs of nuclear hormone receptors that exist exclusively in metazoan proteins: the DNA-binding (hormone receptor) domain including its zinc finger modules, and the ligand-binding domain. Functional studies revealed that the expression of the *S. domuncula* RXR undergoes strong upregulation in response to treatment with retinoic acid, while the expression of the sponge caspase is not increased. The gene encoding the morphogenetic LIM/homeobox protein was found to be strongly upregulated in response to retinoic acid treatment. These data indicate that the RXR and its ligand retinoic acid play a role in the control of morphogenetic events in sponges.

2.6 Immunity, Individuality, and Apoptosis

The early metazoans, the hypothetical Urmetazoa, lived in an aquatic environment and consequently were exposed to a massive load of both pro- and eukaryotic organisms trying to invade and destruct them. To cope with these threats, sponges have developed an efficient chemical defense system as well as humoral and cellular defense mechanisms.

The property to distinguish between self–self and self–nonself by fusion or rejection is common to all metazoans. The finding that the strategies used during these histo-(in)compatibility reactions are very similar from Porifera to Chordata was again very surprising. Key molecules involved in allo/autoimmunity in sponges that share high sequence and functional similarity with those molecules that had been found to control historecognition in deuterostomes, have been identified. Among those are the molecules comprising polymorphic Ig-like domains present in the sponge adhesion molecules, the allograft inflammatory factor (a sponge cytokine) as well as the (2–5)A system (control of infection).

Functionally connected with the immune system in Metazoa is the process

of apoptosis. In the last five years, it was again disclosed that the basic apoptotic response and effector systems are conserved among all metazoan phyla. On the basis of the view that developmental processes can only function if they are paralleled with apoptosis, a comprehensive search for molecules in sponges that were involved in apoptotic pathways was initiated. As the most promising segment to screen for a proapoptotic molecule, we selected the death domain, which is found in the mammalian apoptosis controlling proteins Fas, tumor necrosis factor-α or its receptor, and FADD. This approach was successful; the molecule isolated from *G. cydonium* even comprises two death domains. Functional assays were performed with allografts from *G. cydonium*, which revealed that in rejecting tissue a strong increase of the expression of the death domain–comprising gene occurs.

In vertebrates, the death domain containing receptors/adapter molecules interact intracellularly with the caspase-8 proenzyme through the death-effector domain with a similar region in the caspase. An adapter-mediated oligomerization causes an activation of the procaspase(s) that undergo cleavage and subsequently, dimerization. Upstream caspase(s) then process procaspase-3, which after dimerization activates a factor necessary for a proapoptotic DNase to degrade chromatin into the nucleosomal fragments, a hallmark of apoptosis. In Bilateria, a series of caspases are involved in the tuned control of apoptosis starting with caspase-8 in the cascade and ending with caspase-3. One gene has been identified in sponges (*G. cydonium*), which encodes two transcript forms, both for caspase-8 and for -3 equivalents. Functional studies indicate that the two forms of the sponge caspases act in *G. cydonium* in the apoptotic pathway.

In line with the biological evidence that in sponges apoptosis can be initiated by environmental stress factors, for example, bacterial load or cadmium, an intense screening for members of the Bcl-2 family was started. This effort resulted in the cloning and functional analysis of the antiapoptotic/cell survival proteins from the two sponge species. The proof that the sponge gene product acts as a cell survival protein was performed by transfection studies using mammalian cells. It could be shown that transfection of mammalian cells with the sponge *Bcl-2*-related gene confers resistance against heat shock and growth factor deprivation.

Taken together, the bulk of evidence confirms that from sponges to the higher metazoans a complex immune and apoptotic machinery exists, which allows the elimination of unwanted tissue (e.g. in allotransplantation) and the establishment of an organized Bauplan.

3 Skeleton

3.1 Morphogenesis

Morphogenesis is a process that is restricted to multicellular organisms. Like in plants, in animals also the cells are not merely a collection of randomly distributed cell types. During development, differentiated cells precisely associate to tissue, organs, and finally to the organism. In embryogenesis, first a proliferation phase takes place, which is paralleled by the differentiation and the morphogenesis phases (Fig. 4); these processes are driven by controlled gene expression patterns. Characteristic genes/proteins involved in proliferation are, for example,

DNA polymerase or thymidine kinase. Under the control of morphogenetic factors, specific cell surface receptors are expressed, which induce the organization of the intracellular signal transduction molecules, which are then used to allow the expression of extracellular matrices or other specific cell interacting molecules. These events establish differential cell affinities between differentiated cells and provide the basis for cell-specific adhesion and motility. The genes encoding morphogenetic switch proteins establish the formation of an apical–basal or anteroposterior polarity axis.

In all metazoans, a mouth region is formed, which allows the ingestion of food or elimination of metabolic end products. In Coelenterata and higher metazoans, the mouth region is characterized by the existence of an organized nerve net or a central nervous system; the mouth region forms a head and allows active feeding behavior. In Porifera, as the basal metazoan phylum, a mouth *sensu strictu* does not exist since they do not have a nerve system. However, these species also have one organizational center, which shows properties reminiscent of the mouth. There, the excurrent canals meet in a lacuna that opens to the external milieu through the oscule. Moreover, in the oscule region morphological singular structures exist, for example, a contractile rim, which allows the contraction of the oscule and hence influences feeding behavior. The oscule reacts to changes in the environment by contracting or opening and hence, it is provided with key control molecules for food uptake and respiration. Therefore, it can be proposed to consider the oscule as an apical center. This view is also supported by the fact that in the oscule region a high expression of morphogenetically active molecules (organizer molecules), for example, of characteristic homeobox genes, can be monitored.

The morphogenetic proteins establish a developmental gradient and act as transcription factors, morphogens, or signaling molecules. In all metazoans, at least one organizing gradient exists, which originates from a distinct region around the mouth, primarily during embryogenesis. This region was termed *organizer*. Organizer-specific genes include transcription factors and homeobox genes, and also soluble and secreted factors that are involved in cell-to-cell signaling in the extracellular space, and either inhibit or induce growth factor action.

In all metazoans, a dramatic and complex rearrangement of cell assemblies/tissue occurs, which is especially obvious during gastrulation. In a narrow sense, embryonal gastrulation is restricted to Coelenterata and triploblasts that have been conventionally attributed to be composed of the two/three germ layers. However, in view of recent molecular biological studies, it became overt that the Porifera are also provided with the basic elements for the formation of germ layers, which include the ECM (collagen and other adhesion molecules) or the cell–cell junctions (tight junction). Until now, the remaining controversial issue has been the lack of a conventional nerve system in Porifera; however, as outlined in the following, impulse conduction in Porifera might be mediated by their spicule/skeletal system. Genetic experiments revealed that Nodal, BMP, and the Wnt pathway, as well as transcription factors like Forkhead are key elements of germ layer formation and gastrulation. The latter two programming and pattern-forming factors act from Porifera to Bilateria.

The head/apical center regions are characterized by the expression of morphogenetically inductive molecules and hence by the ability to form axes. These centers display organizing functions whose effectiveness drops with distance from them. From Porifera to the triploblastic crown species, these centers display a distinguished regeneration capacity that cannot usually be reached by other regions in the body; removal of these centers can only poorly be replaced by cells from other parts of the body.

Finally, morphogenesis also implies that stage-specific parts of the body are eliminated in order to replace these unnecessary cells by others that are required for the formation of new structures and tissue during development. In recent years, it could be established that the genes and the expressed proteins, which control this apoptotic process, are conserved from sponges to human. In this respect, it is noteworthy to mention that the antiapoptotic protein Bcl-2 from the sponge *G. cydonium* confers resistance against apoptotic stimuli to human cells.

Skeletal elements are formed in all metazoan phyla to stabilize the body and to allow growth. In triploblasts, two different strategies have been followed. In Protostomia, an exoskeleton (integument), especially well developed in Arthropoda/Antennata, is formed which is made of organic molecules (chitin). Growth of the segmentally arranged integument is controlled by pair-rule, segment polarity, and selector genes. In Deuterostomia, and similarly in somites, arranged skeleton is formed during development. There, the skeletal structures (bones) derive from the somatic mesoderm and can be divided into three lineages that generate the skeleton. First, from the sclerotome the axial skeleton derives, then the lateral plate mesoderm generates the limb skeleton, and finally the cranial neural crest gives rise to the branchial arch and craniofacial bones. Osteogenesis involves two processes; the transformation of preexisting mesenchymal tissue into bone tissue (intramembranous ossification), and the differentiation of mesenchymal cells into cartilage, which is subsequently replaced by bone (endochondral ossification). Somite formation occurs along the anteroposterior axis and is controlled by Hox gene clusters and additionally by a series of soluble factors, for example, BMPs or insulin-like growth factors. The matrix macromolecules of the bones comprise the insoluble framework of collagen, which is associated with noncollagenous proteins. Around this matrix the inorganic calcium hydroxyapatite crystals are deposited. The growth of bones follows highly organized patterns; either the bone forming cells are arranged through the collagen matrix in concentric/cylindrical structures or in parallel arrays. Hence, the biomineralization process is guided by organic structural proteins. It should be noted here that biomineralization in various forms around an organic matrix is among the triploblasts not restricted to Deuterostomia, but also occurs in Protostomia, for example, in mussel shells as aragonite (calcium carbonate).

In diploblasts, in coelenterates, the basic inorganic biomineral is calcium carbonate. Especially in corals, the growth pattern has been studied in detail and revealed that around the axis a radiating accretive growth process occurs. This implies that layers of newly formed organic/inorganic material are added to the tip of the preceding growth step. By this process, the thickness of the layers is highest around the central axis and decreases

toward the side of the tip. In analogy to bone formation, the "apposition" of new layers of organic/inorganic material occurs through the guidance of organic matrices and the morphogenetically determined body or limb axis. It is outlined later that also in Porifera such an appositive growth pattern is seen.

Taken together the basic processes that allow the formation of a metazoan organism, are cell–cell adhesion, which is followed by cell–matrix adhesion. Then cell proliferation and differentiation occur, which allow growth of the embryo during development. Axis formation is established either by secreted molecules, for example, morphogens, or by transcription factors. This sequential process is schematically outlined in Fig. 4.

3.2 Spicule Network in Sponges: A Unique Skeleton

In living organisms, four major groups of biominerals exist; (1) iron compounds, which are restricted primarily to Prokaryota; (2) calcium phosphates, found in Metazoa; (3) calcium carbonates, used by Prokaryota, Protozoa, Plantae, Fungi, and Metazoa; and (4) silica (opal) present in sponges and diatoms. It is surprising that the occurrence of silica as a major skeletal element is restricted to some unicellular organisms and to sponges (Demospongiae and Hexactinellida). At present, it is not known why sponges have used silica as the biomineral to form one major innovation during the evolution from Protozoa to Metazoa, the hard skeleton. Since the transition from Protozoa to Metazoa is dated back 600 to 1000 million years ago, it has been proposed that oxygen level, temperature, and seawater chemistry played a major role in the evolution from Protozoa to Metazoa. In particular, it was assumed that during the period of the appearance of sponges, the ocean was richer in sodium carbonate than in sodium chloride; and that such a "soda ocean" had probably a pH above 9. Under such conditions, the concentration of silica, the dioxide form of silicon, was presumably higher in seawater than it is today.

The formation of the skeleton is a multifaceted process and will be explained in an exemplary way for Porifera. Even though these animals comprise the most simple body plan, their biomineral structure formation is already highly complex and by far not completely understood. Like in triploblasts, the diploblastic Porifera skeleton formation also has a pronounced effect on morphogenesis. As an example, if the animals are growing under unfavorable conditions that do not allow the formation of inorganic deposits (silica or calcium biominerals), the growth of the specimens is extremely suppressed.

Silica is the major constituent of sponge spicules in the classes Demospongiae and Hexactinellida. The spicules of these sponges are composed of hydrated, amorphous, noncrystalline silica. The secretion of spicules occurs in Demospongiae in specialized cells, the sclerocytes; there, silica is deposited around an organic filament. If the formation of siliceous spicules is inhibited, the sponge body collapses. The synthesis of spicules is a rapid process; the 100 μm long megasclere are formed within 40 h.

Inhibition studies revealed that skeletogenesis of siliceous spicules is enzyme-mediated, more particularly, by an Fe^{++}-dependent enzyme. Recent studies revealed that the dominant enzyme that catalyzes the formation of monomeric to polymeric silica is mediated by an enzyme

that belongs to the cathepsin subfamily and was termed *silicatein*.

The formation of siliceous spicules in sponges is genetically controlled; this process initiates the morphogenesis phase. Data demonstrated that at suitable concentrations, silicate induces genes, for example, those encoding collagen, silicatein, and myotrophin. A major step forward in elucidating the formation of the siliceous spicules on a molecular level was the finding that the "axial organic filament" of siliceous spicules is in reality an enzyme, silicatein, which mediates the apposition of amorphous silica and hence the formation of spicules.

The skeletal framework of the sponges is highly ordered. In two examples, the demosponge *Lubomirskia baicalensis* (Fig. 7a) and the hexactinellid *Euplectella aspergillum* (Fig. 7b), it can strikingly be seen that the growth of the sponges proceeds in a radiate accretive manner, meaning that growth zones, which are highly ordered are delimited by growth lines. These growth centers are constructed by an ordered arrangement of the spicules within the body. Most siliceous sponges are composed of larger megascleres, >10 μm long, and smaller microscleres (<10 μm). Some microscleres have a bizarre shape (Fig. 7c), while the megascleres usually have a long central rod with ends that in some cases are decorated with spines. Interestingly, the spicules (Fig. 7d) have the enzyme silicatein both in the central canal and also around the outer concentric silica layers. The central core rod of the spicule is synthesized first by the silicatein fiber, which exists in the central hole around which a 1 to 3 μm thick siliceous hollow fiber is formed. This layer becomes surrounded by an organic silicatein layer, which synthesizes the second siliceous layer and so on. This process of growth is very reminiscent of the process of bone formation.

By applying the *in vivo* cell culture technique, the primmorphs, (Fig. 7e) the dynamics and cell biology of spicule formation could be studied. These experiments revealed that the spicules are formed extracellularly in close association with fibrils (Fig. 7(f,g)). Cross sections through a young spicule within cells from primmorphs show that the amorphous silica of the spicule is initially not very compact (Fig. 7h).

At present, no experimental data are available which could explain the species-specific form and shape of the spicules; it had been suggested that their polarity is the result of homeobox genes.

Fig. 7 The skeletal framework of the sponges is highly ordered. Radiate accretive growth pattern seen in the skeletal arrangement of the spicules from the freshwater demosponge *Lubomirskia baicalensis* (a) [x1] and the marine hexactinellid *Euplectella aspergillum* [x0.5] (b); the zones that are uniformly constructed are dissected by growth lines. (c) A microsclere from *Geodia cydonium*; x10 000. (d) Scanning electron microscopic analysis of basal spicules from *Hyalonema sieboldi* showing the concentric layers of silica layers by which they are built. The center of the spicule is composed of an organic filament, which comprises of the silica-forming enzyme silicatein and forms the first hollow fiber. Around the first fiber structure, silicatein is deposited again and forms the second concentric layer; x10. (e) A primmorph (pr), the cell culture spheroid from *S. domuncula*, has the capacity to synthesize spicules (sp); x20. Formation of spicules in primmorphs: (f) The newly formed spicule (sp) is closely associated with fibrils (fi); x5000. (g) Synthesis of a spicule in a sclerocyte within a primmorph; x1000. (h) Section through a primmorph, showing a sectioned young spicule; the amorphous silica of the spicule is not very compact.

3.3
Spicule Network in Sponges: A Unique Nervous System?

Sponges are devoid of a nervous system. In a recent study, the (potential) biological relevance of the spicules as optical fibers in sponges had been discussed from the view that they might substitute for a nerve system. For these studies, the hexactinellid sponge species *Hyalonema sieboldi* was selected; it is

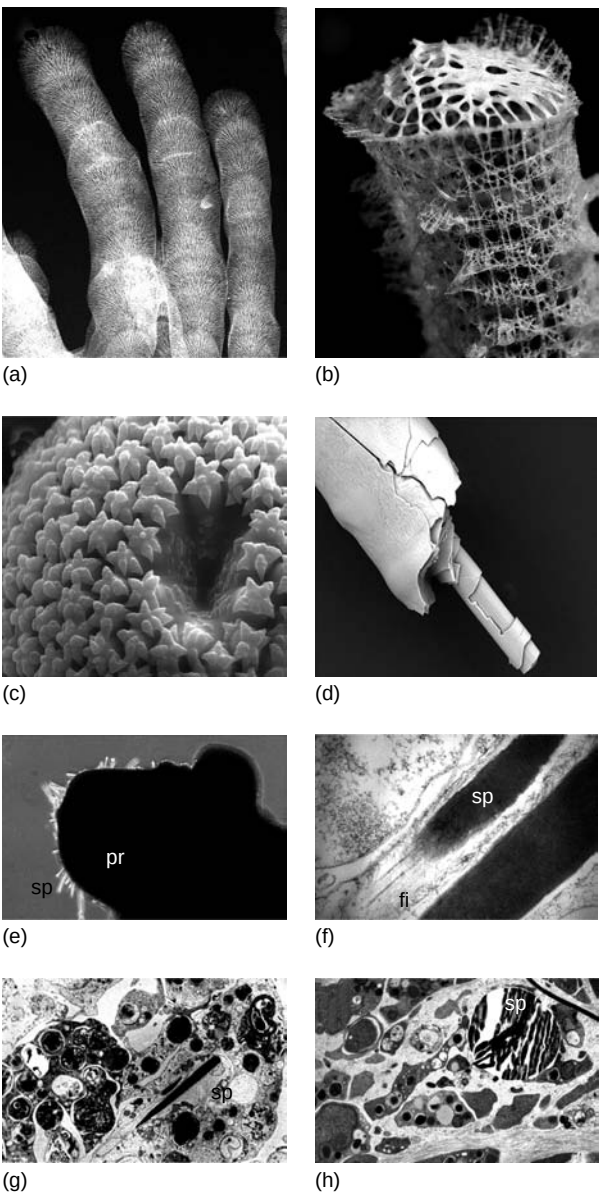

characterized by extremely long spicules that form the stalk of these animals. It is remarkable that the internal siliceous layer(s) that surround the axial filament within the spicules display a higher flexibility than the layers that are deposited more outward. In order to analyze the optical properties of the stalk spicules as optical fibers, they were exposed at one end to white light. The first surprising result was that light was cut off below 615 nm and above 1310 nm. Within the visible part of the spectrum transmitted by the fibers, three less pronounced absorption maxima were determined. The percentage yield of transmission of the different wavelengths was again astonishing. It was found that almost 40 to 60% was determined at wavelengths around 900 nm (length of the spicules: 10 cm). From comprehensive studies of luminous genera, in organisms that are able to emit light through their bioluminescent systems, it is known that most pelagic deep-sea organisms emit light with maxima of emission spectra around 480 nm, while those that are terrestrial or live in freshwater produce light with longer wavelengths. Therefore, sponges were screened for a bioluminescent system, which generates light especially within the wavelength of 400 to 600 nm. This approach was successful; a gene was identified coding for a luciferase.

On the basis of these experimental findings, it appears to be justified to propose that sponges are provided with a so-far unrecognized photoreception system (Fig. 8A). Therefore, our group postulated that sponges coordinate their sensory reception systems not on the basis of a protein-controlled nervous network, but by their inorganic siliceous spicule system. Since they are provided with the genetic machinery to express luciferase enzymes, the optical fibers (spicules) might guide and convert the light, via a photoelectric reaction, into electric signals. From model systems, using supported bilayer lipid membranes, it is well established that some of them show photoelectric properties; likewise, photoelectric effects in plants are well known. Focusing on the sponge, in earlier elegant studies it has been demonstrated that tissue from the hexactinellid *Rhabdocalyptus dawsoni* could transmit electrical impulses more than 1 cm away from the place where the electric impulse was sent. The subsequent amplification system that translates the electric signals into the (nervous) transmission system, both in sponges and in other metazoan phyla are likely mediated by the same biological amplifiers, by receptors mediating the metabolism of the cells. While sponges are provided only with cells containing neuronal receptors, the neuronal cells of higher metazoans are connected by a nervous system. However, on the basis of the experimental data presented here, and the data from the literature, the system of spicules in sponges might substitute in these animals for the nervous systems of higher metazoans. This view is supported by *in situ* hybridization studies, which revealed that the spicules are surrounded by cells that selectively express not only the silica-forming silicatein (Fig. 8B(a)) but also a photomorphotropic protein (Fig. 8B(b)). These cells are characterized by their large diameter (>15μm).

It is concluded that sponges are able to react to physical stimuli, especially to light, with a coordinate reaction based probably on a system of glass fibers (spicules). It seems evident that this novel view of the photoresponsive

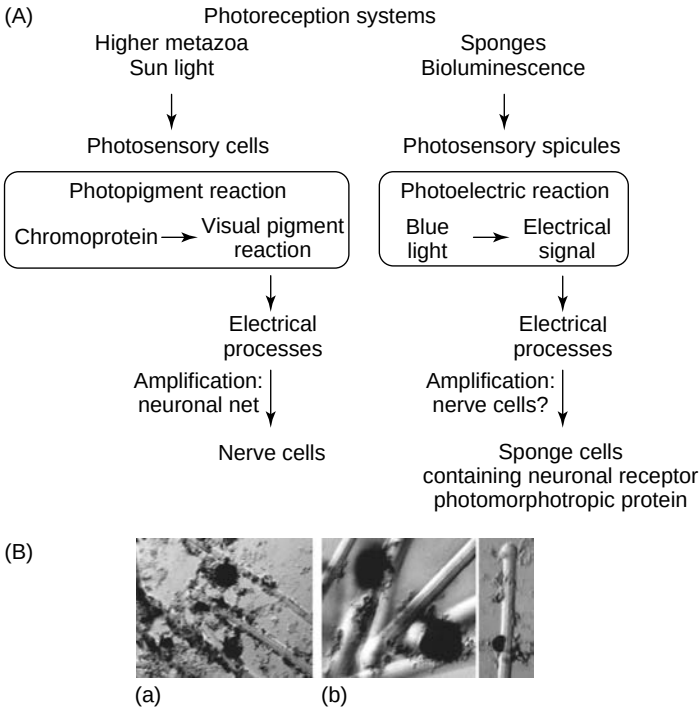

Fig. 8 Proposed photoreception system in sponges in comparison with that of higher metazoans. (A) In sponges, evidence is available that light is generated by bioluminescence (luciferase system), which is bundled by silica-based fibers. Within the blue light range, a photoelectric reaction takes place, converting the light into electric impulses. In contrast, the electrical impulses in higher Metazoa are generated by protein-based reactions. The amplification of the electric impulses in both systems occurs in nerve cells only, as in sponges, or in a neuronal net. (B) Large cells (>15 μm) surrounding the spicules which overexpress the *silicatein* gene [B(a)] and the gene expressing the photomorphotropic protein [B(b)].

properties of these animals, with the siliceous spicules as central elements provided with a photoelectric effect, will open new concepts in photoperception in animals.

4 Axis-pattern Formation

Pattern formation requires the defining of the main axes. As outlined, the diploblastic animals are characterized by one main body axis (apical/oscular–basal), the triploblasts have two axes (in addition to the anteroposterior one, the dorsal–ventral). Many signaling pathways are involved in those polarity-forming processes. Two large groups can be distinguished; first, pathways that originate from secreted signaling molecules, and second, those that are controlled by transcription factors. One example for each is given here.

4.1
Signaling Molecules: Wnt Pathway

The Wnt signaling pathway is a cell communication system that regulates cell-fate decisions, tissue polarity, and morphogenesis (Fig. 9). The Frizzled protein is the membrane receptor for the Wnt secreted glycoproteins. Through the canonical Wnt signaling pathway, the activated Frizzled binds to Disheveled (Dsh), which leads to the stabilization and accumulation of ß-catenin in the nucleus, where it activates the TCF/LEF transcription factor. Besides this canonical Wnt signaling pathway, two further related pathways have been identified: first, a signaling pathway downstream of Dsh, which includes the Rho and JNK cascade and was termed *noncanonical Wnt signaling pathway*; and second, the Wnt/calcium pathway that stimulates intracellular calcium release in a G protein–dependent manner. In vertebrates, the canonical Wnt signaling pathway is involved in axis specification, noncanonical Wnt signaling in the formation of cell polarity and convergent extension, and the Wnt/calcium pathway in tissue separation.

Recent studies demonstrated that genes encoding the Frizzled (Fz) receptor, as well as genes expressing proteins downstream of this receptor are present already in sponges. *In situ* hybridization analysis in adult *S. domuncula* specimens demonstrated Fz expression in the cortex region and in the epithelial layer of the aquiferous canals. Furthermore, Northern blot analysis revealed an upregulation of its levels of expression during the formation of sponge primmorphs. These findings suggest that a Wnt pathway involved in cell-fate determination and morphogenesis is already established in the phylogenetically oldest metazoans, the sponges.

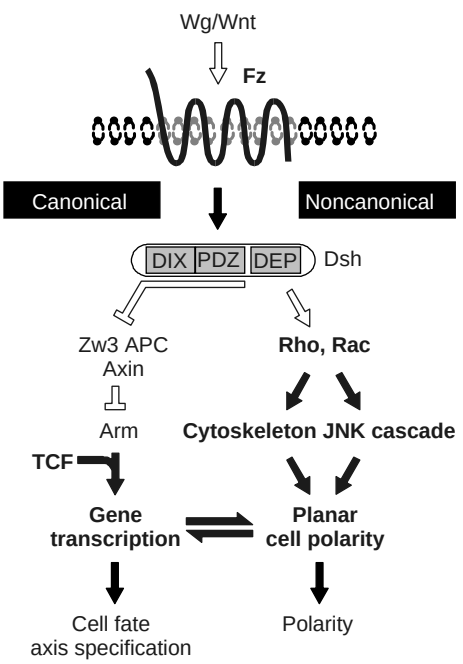

Fig. 9 Wnt signaling pathway. The extracellular Wingless (Wnt) ligand binds to the Frizzled receptor (Fz), which regulates cell-fate decisions through the Disheveled (Dsh) molecule that is composed of three functional domains (DIX, PDZ, DEP). From there two pathways lead either to the activation of the TCF/LEF transcription factor or to the JNK kinase cascade. The resulting selective gene expression causes a planar cell polarity. Those molecules that have already been identified in sponges are highlighted in bold.

4.2 Transcription Factors: T-box – Forkhead

During development, sets of genes, most of them transcription factors that are responsible for cell-fate and pattern determination, are expressed. Among them, T-box, Forkhead, and Homeobox gene families are mentioned, since they have been found to be extremely conserved on sequence and functional level in all metazoans.

Members of the T-box family, for example, Brachyury, are involved in the formation and differentiation of the third germ layer, the mesoderm, in triploblastic animals. They are expressed first at the end of the blastula stage in the region of the presumptive mesoderm, and later during gastrulation; the expression is restricted to the region of the notochord and the tail bud. In the diploblastic *Hydra, Brachyury* is expressed in the endothermal epithelial cells of the hypostome, which forms the apical part of the head of the polyp. On the basis of this observation, it has been proposed that *Brachyury* is involved in the formation of the first body axis, underscoring that *Brachyury* is crucial for the patterning of axes and germ layers in diploblasts as well as in triploblasts. Moreover, *Brachyury* is also important for the formation of the apical pole, the head. Recently, two T-box genes have been isolated from the sponge *S. domuncula*; a *Brachyury* gene and a homolog of the Tbx2-3-4-5 genes from chordates, which, in the latter taxa are involved in the formation of the limbs.

Since the larvae of *S. domuncula* cannot be routinely cultivated, it is – at present – not possible to conduct experiments with them. Therefore, studies had been restricted to adult animals and cultured sponge primmorphs, which comprise proliferating and differentiating cells. Interestingly, it could be demonstrated that the expression of the *Brachyury* gene is upregulated in differentiating sponge cells during formation of canal-like structures, suggesting that already in Porifera the primordial axis is genetically fixed.

This assumption was subsequently confirmed by the isolation and phylogenetic characterization of five members of the winged-helix/Forkhead gene family from the sponge *S. domuncula*. Forkhead proteins form a subfamily within the large group of helix–turn–helix proteins. They are responsible for a wide range of functions and key roles in early developmental processes, during organogenesis, and also for the function of the major organs and tissues in the adult animals. HNF3β, the founding member of this gene family, is responsible for the formation of terminal structures that develop into the gut.

The expression pattern of *Forkhead* and *T-box/Brachyury* genes during late blastulae and early gastrulae, supports the assumption that these two sets of genes are required for the morphogenetic movements occurring during processes identical or phylogenetically preceding gastrulation. Moreover, the overexpression of *Brachyury* is seen from Porifera and Coelenterata onwards to the triploblasts in distinct regions, usually adjacent to the organizers.

4.3 Organizers

Axis formation requires the concentration of cells at a distinct region, which displays the property to induce a complete body axis, even after transplantation; this region is termed *organizer*. In triploblastic animals, the organizer is involved both in dorsal-ventral and in anteroposterior patterning, for example, the Spemann organizer or the Hensen's node. In

diploblasts, like Coelenterata or Porifera, the organizer is closely connected to the circulation cavity.

4.3.1 Spemann Organizer

Spemann in 1936 was the first to describe the action of an organizer in amphibians by demonstrating that the dorsal lip of the amphibian blastopore to the ventral side of another embryo resulted in the development of a twinned embryo. The lip was able to induce the central nervous system, a dorsal mesoderm, and a secondary gut cavity. These results showed that in vertebrates the processes of cell differentiation progress through a series of sequential cell–cell interactions involving different cell affinities. Subsequently, organizer-specific genes were isolated, for example, *goosecoid*, *LIM homeobox protein*, and *Forkhead*. Microinjection studies revealed that mRNA, for example, of *goosecoid*, is able to induce a secondary axis. Furthermore, cells that had been injected with organizer-specific transcription factors are able to recruit adjacent cells for axis formation. In addition to these nuclear factors, a collection of genes encoding secreted factors were also identified, noggin, chordin, and follistatin being among these. Interestingly, these molecules act as antagonists for extracellularly present growth factors; hence, those organizer secreted molecules were termed according to the growth factors they antagonize. Among these is the Frzb-1 protein (Frizzled-related protein), which has the same binding affinity as the Frizzled receptor but lacks the transmembrane domain. Therefore, the soluble Frzb-1 protein binds to the Wnt signaling molecule, removes the ligand from the signaling pool and in consequence antagonizes Wnt signaling.

4.3.2 Hensen's Node

This node, found in birds, is a structure analogous to the Spemann organizer. During the progression of the avian embryo from the blastula to the gastrula, one major structure is formed, which is termed *primitive streak*. Within this streak, a depression or groove is formed that gives rise to the blastopore. At the anterior end of the primitive streak, a regional thickening of cells is found, which represents the Hensen's node. Cells from the node form the notochord. Transplantation studies also confirmed that the node could induce a new axis in avian embryos again supporting the high functional homology of the morphogenetic proteins; the same set of expressed genes as in the Spemann organizer was found in the Hensen's node.

4.3.3 Coelenterata

The hydrozoan polyps comprise a gastrovascular cavity that opens to the outside environment with a mouth/head. The mouth is set on an elevated hypostome, which comprises organizer activity. If this region is transplanted to the gastric region, a new head, together with tentacles and body axis, is formed. Molecular studies revealed that the organizer expresses the characteristic regulatory genes, coding for transcription factors, for example, LIM-class homeodomain protein or Brachyury.

4.3.4 Porifera

In sponges, the water enters the animals through many porocytes on their surface, the inhalant openings, into the canals formed by the endopinacoderm and is then passed to the choanocyte chambers from where the water is driven to the central atrium and finally pressed through the oscule out into the environment. Generally, the number of oscules is restricted and most species possess only one

major oscule. Several lines of morphological/cytological and molecular biological evidence indicate that the aquiferous canal system in sponges represents the organizational center of the sponges (Fig. 10a). During embryogenesis, amphiblastula or parenchymella larvae are formed. In their interior cavity, first choanocyte chambers are formed; these chambers are the central organ-like structure that directs and controls the water current. Subsequently, canals are formed which finally fuse with the outer pinacoderm layer. The 3D-cell culture experiments likewise revealed that an increase of the water current in the culture fluid results in canal formation, controlled by the homeodomain protein Iroquois.

In order to determine if the oscule, as the morphologically most prominent region in the sponge, also acts as an organizer-like region, ablation/transplantation studies were performed. In *S. domuncula*, the specimens comprise only one oscule (Fig. 10a); the histological section shows a large atrium in which the excurrent

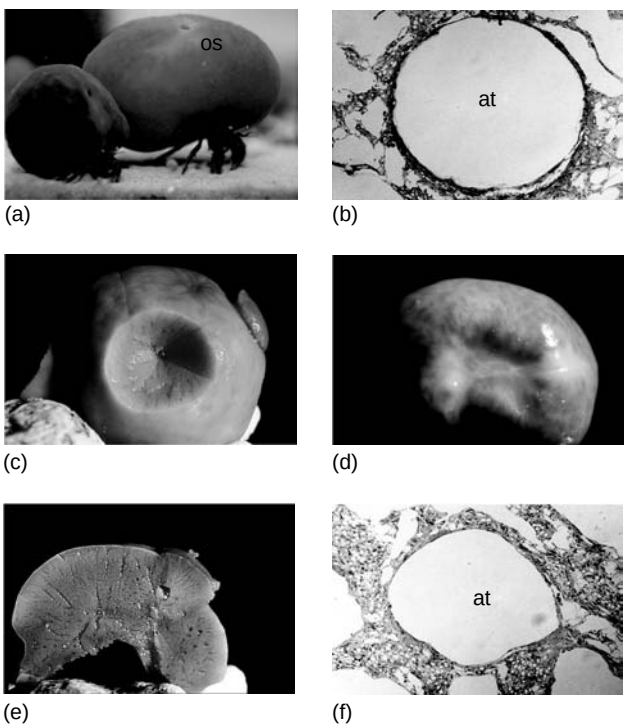

Fig. 10 Regeneration capacity of the oscule region in *S. domuncula*. (a) Each specimen contains one oscule (os); x1. (b) The histological section shows the atrium (at) below the oscule opening; x300. (c) If the oscule region is removed, the open surface (d) regenerates (for 14 days) by forming new epithelial layers, which, however, never show an oscule. (e) In contrast, if the oscule is removed from one place of the specimen and inserted into another region, the structure/atrium (at) of the oscule remains intact (14 days) (f); x300. (See color plate p. xxviii).

canals end and that opens into the oscule (Fig. 10b). The regeneration capacity of the oscule region is distinguished from the other regions of the body. After removal of the oscule (Fig. 10c), this area regenerates and is sealed only by an intact epithelium with a double pinacoderm layer, and not by the formation of a new oscule (Fig. 10d); also, at no other surface region is an oscule formed. However, if the oscule is excised and transplanted to another site (Fig. 10e), an intact oscule with an atrium is formed (Fig. 10f).

By applying a series of molecular/cell biological techniques, it could be demonstrated that the oscule region comprises the highest level of expressed genes, indicative of organizer regions. Among the genes that are overexpressed in the tissue surrounding the oscule are the LIM-homeodomain protein, Brachyury, Frizzled receptor, and also the secreted Frizzled-related proteins, noggin, or Iroquois. These data are strong arguments for the assumption that regionalized organizer centers are present from Porifera to the crown triploblastic species and are localized close to the openings of the body cavity.

4.4 Homeodomain Protein

In contrast to unicellular organisms, cells from metazoans also comprise genes for molecules that impart positional information in addition to genes encoding proteins for structural and narrow functional properties This property implies that cell movement within a specimen proceeds in a coordinate way to allow body plan formation. Hence, the cells have the ability to react to diverse temporally regulated interactions between shifting populations of cells and by that find their functional positions. These processes are driven by inductive proteins that are hierarchically expressed and that regulate cellular differentiation in order to connect the cells through complex form building cell movements.

In the two triploblastic model animals used for understanding the developmental gene network, *D. melanogaster* (Protostomia) and mouse (Deuterostomia) the anteroposterior/axial patterning of development is controlled by homeobox genes, coding for DNA-binding homeodomain transcription factors, which are even clustered in the genome. Most of the genes are expressed in a "collinear" way; the outstanding feature of this cluster is that genes at one end of the cluster are expressed and induce pattern formation at the anterior end of the embryo, while the genes at the other end of the cluster pattern the posterior of the embryo. Finally, the genes in the central region of the cluster pattern the middle of the embryo. This feature is termed *spatial colinearity*. In addition to the process of spatial colinearity, "temporal colinearity" exists, a process that implies that genes at one end of the cluster are activated first and genes expression is initiated progressively through the cluster until finally the genes at the opposite end are turned on. The building blocks of these clusters are represented by homeotic genes, comprising the homeobox. After the discovery that this homeobox is present in all of the clustered homeotic genes that code for sequence-specific DNA-binding domains, and the almost parallel discovery that the arrangement of the homologous genes is similar in insects and vertebrates, a comprehensive molecular comparison of the developmental processes in triploblasts became possible.

4.4.1 Insects – Drosophila melanogaster

The major difference between the mechanisms of expression of the homeobox genes in Protostomia and Deuterostomia is that the protostomian insects retain much of the segmental pattern of the body plan of a more simple organism. In *D. melanogaster*, the definite anteroposterior pattern is achieved in stages. First, genes of maternal origin are transcribed, followed by the sequential expression and function of the segmentation genes, which had been termed *gap* genes, *pair-rule* genes, and *segmental polarity* genes. On the basis of this sequential expression, a segmental periodicity is caused in the embryo along the anteroposterior axis; the identity of these individual segments is specified by homeotic selector genes (Fig. 11).

The basic modules of the segments in insect embryos are the "parasegments" from which the "segments" derive. The parasegments appear on the surface of the embryos after gastrulation. In *D. melanogaster*, 14 parasegments are delimited, which represent independent developmental units and acquire their own identity. The characteristics of the parasegments are that they are out of register with the final segments, which are seen in the late embryos and the adult specimens. Each segment is composed of a posterior region of one parasegment and an anterior region of the next one. In the head region of the adults, the segmental arrangement is lost. The gene clusters composed of homeobox genes, which delimit the parasegments are termed *pair-rule* genes. They are expressed in stripes; seven even-skipped pair-rule genes define the odd-numbered parasegments while seven others, for example, the *fushi tarazu*, define the even-numbered parasegments. The activity of the *pair-rule* genes is only temporary. During the process of cellularization of the blastoderm, the segment polarity genes are expressed. Among these is the transcription factor "engrailed," which is expressed in the anterior region of every parasegment. In contrast to the *pair-rule* genes, the *engrailed* gene is active during the entire life of the fly and displays two functions; first, it delimits the boundary of cell lineage restriction and also acts as a selector gene, which helps to identify a region through its control of the activity of other genes. The anterior border of each parasegment has the important property to prevent movement of the cells of this parasegment into the adjacent parasegment. Such a cell lineage restriction seen at the anterior margin of the parasegment in insects or at the boundaries between the rhombomeres of the vertebrate hindbrain implies that the same common genetic control mechanism(s) act within a stripe of the anteroposterior oriented embryo. An area that is delimited by the restriction of cell lineages has been termed *compartment*. Within a compartment, genes such as *engrailed*, which are expressed in the anterior margin of the parasegments, establish a boundary of lineage restriction and prevent primordial cells and their descendants from one side of the boundary from crossing.

The segments of the late embryos and the adults derive from the parasegments. Each larval segment is composed of the posterior region of one parasegment and the anterior region of the next parasegment. Hence, in the segments the *engrailed* gene is now expressed in the posterior region. Segmentation has a well-defined anteroposterior pattern and can be morphologically seen on the ventral site of the abdomen. The anterior region of each segment comprises the denticles (outgrowth of the cuticle), while the posterior region is naked. Besides *engrailed* which together with *hedgehog* is expressed at the anterior

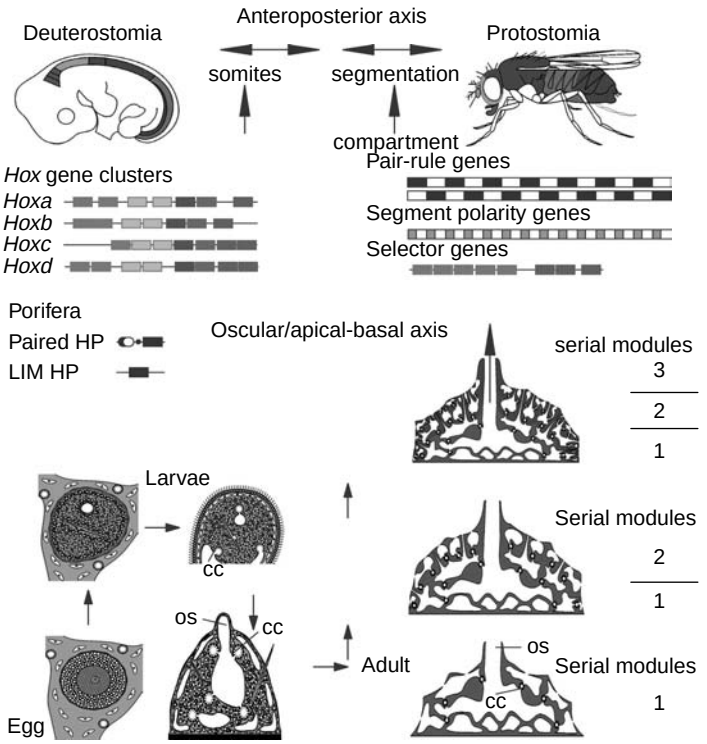

Fig. 11 Schematized view of the pattern formation in Metazoa. In the triploblasts, both the Deuterostomia, with the mouse as a model system, and the Protostomia, with the *Drosophila melanogaster* model, pattern formation is spatially and temporally controlled by homeobox gene clusters. In the mouse, several different clusters exist; their genes are positionally and combinatorially expressed along the anteroposterior axis. Colinearity exists between the order of *Hox* genes on the chromosome and the order by which the individual genes are expressed along this axis. In the insect, *Drosophila*, regular segmentation along the anteroposterior axis takes place. Prior to segment formation, parasegments develop. Sets of gene clusters (pair-rule genes) control parasegment formation; both genes defining odd-numbered parasegments and even-numbered parasegments are in action. Parasegment polarity is accomplished by the ''segment polarity'' genes. Subsequently, segments are formed from one anterior and one posterior region of a parasegment. Homeotic selector genes that help identify a particular region are present in a cluster and are grouped into an Antennapedia and a bithorax complex. In contrast to these triploblasts, the sessile diploblastic Porifera comprise an axis, which is usually defined by the location of the oscule (os). During embryogenesis, the fertilized egg develops into a larva, which comprises a cavity. During and after the settlement of the larvae, first choanocyte chambers (cc) are formed and the body cavity opens to the external milieu through an oscule. The growth of the adult proceeds along the oscule axis, apical/oscular–basal axis in a compartmental manner. It is proposed that the growth of sponges, occurs by serial modules, along the body axis; these processes are controlled by paired-class and LIM-class homeoproteins (HP), which are not arranged in a cluster on the chromosome. (See color plate p. xxx).

margin of each parasegment, additional genes are expressed at the posterior margin, for example, *wingless*. These polarity genes determine the segment patterning. It is interesting to note that *wingless* and *hedgehog* code for secreted signaling proteins that are not restricted to insects, but are also found in vertebrates where they are key players in pattern formation also.

Each segment has its own identity. As mentioned earlier, the same segment polarity gene is turned on in each segment. The specification of the identity of each segment is under the control of master regulator genes, termed *homeotic selector* genes. They control the expression of other genes throughout the development and are arranged into two gene complexes that share some homology with the *Hox* genes in vertebrates. In insects, two complexes are distinguished, the Antennapedia complex and the bithorax complex. Since all of these transcription factors contain a homeobox, they have also been collectively termed *HOM* genes. The genes of the Antennapedia complex encode transcription factors, which specify the anterior region, while the bithorax complex comprises the homeobox genes involved in the specification of the thorax. In the bithorax complex, three homeobox genes exist, *ultrabithorax* as well as *abdominal-A* and *-B*. The expression of these genes has already been seen in the parasegments; *ultrabithorax* is expressed in all parasegments from 5 to 14, while the expression of *abdominal-A* occurs in parasegments from 7 onwards, and *abdominal-B* from parasegments 10 onwards. Their combined/combinatorial activities define the character of each parasegment and are under the control of the pair-rule proteins, which are turned down in the initial phase of the development; subsequently, their correct expression is controlled by two groups of genes the *polycomb* and *trithorax* groups. The genes of the Antennapedia complex (*labial, proboscidea, deformed, Sex combs reduced, antennapedia*) are likewise spatially expressed and controlled in the anterior parasegments from anterior to posterior.

It is one characteristic of the bithorax and the Antennapedia complex that the order of the genes in the complex is the same as they are expressed along the anteroposterior axis during development. As an example, the *ultrabithorax*, which is located at the 3′ side of *abdominal-A* on the chromosome, is expressed more anteriorly and activated earlier than *abdominal-A*. This feature of temporal and spatial colinear expression of the genes within the insect *HOM* genes is highly conserved and also seen in vertebrates.

4.4.2 Vertebrates – Mouse

After gastrulation, the main axial structures of the vertebrate embryos (Deuterostomia) – somites, notochord, and neuronal tube – are formed. The regional organization and the anteroposterior axis formation start with the development of the somites from the most ventral region of the mesoderm at both sides of the embryo, while the notochord and the subsequent neurulation develop as a rod in the dorsal midline of the mesoderm. From the somites, the body and limb muscles derive and the cartilage is formed, which gives rise to the vertebrae, ribs, and dermis. The anteroposterior orientation of the mesoderm can be morphologically identified on the basis of the anatomical characteristics of the vertebrae, which depend on their locations along the axis.

The anteroposterior patterning is genetically controlled by clusters of homeobox genes (Fig. 11). Like in *Drosophila*, the

homeobox genes (*Hox*) specify the positional identity along the anteroposterior axis. *Hox* genes are known not only to control patterning of mesodermal structures along the anteroposterior axis but are also involved in the positional organization of other regions of the embryo, for example, the hindbrain. Vertebrates have four separate *Hox* gene clusters, which very likely arose by gene duplication. Again, like in insects, the vertebrate genes in the clusters are also expressed in a temporal and spatial order and reflect the arrangement on the chromosome. The best-studied model is the mouse. The *Hox* genes are first expressed in the anterior region in mesoderm cells at the early stage of gastrula. The pattern of expression of a given *Hox* gene is relatively sharp at the anterior border while it is less defined at the posterior borders. Most regions within the anteroposterior axis in the mouse embryo are characterized by a defined set of expressed *Hox* genes.

The positional identity of the somites is specified by the combinatorial expression of the genes of the Hox complexes along the anteroposterior axis. Again, the order of expression of the genes corresponds to their order on the chromosome. It should be stressed here that in vertebrates, retinoic acid can alter the positional "value" of a cell during development, especially during limb formation, and modulates the orientation of the anteroposterior axis.

A further class of Hox clusters has been identified in triploblasts, the paraHox cluster. It has been postulated that the paraHox cluster evolved from a putative Proto-Hox cluster. Also, these genes exhibit colinearity in neural tube and gut endoderm where they are associated with craniocaudal patterning. In mammals, the cluster consists of *Gsh1*, *Pdx1*, and *Cdx2*. The paraHox cluster also exhibits spatial and temporal colinearity; however, the temporal colinearity is inverted with respect to the pattern of the Hox cluster. In the paraHox cluster, the posterior gene *Cdx* is activated first, followed by *Xlox*, and then finally the anterior *Gsx* gene.

It should be stressed here that most of the homeobox genes present in Metazoa do not belong to the homeobox complex or are not involved in homeotic transformations.

4.4.3 Coelenterates–Cnidaria

In Cnidaria, both *Hox* and *paraHox* genes have been identified. In consequence, it was postulated that these two classes of clusters emerged by duplication from an ancestral minimal cluster of two to three *proto-Hox* genes within the Cnidaria. In *Hydra*, several *Hox/paraHox*-related genes, for example, *cnox-1*, *cnox-2*, or *cnox-3* have been identified and their expression was analyzed by *in situ* hybridization. In adult specimens, the three genes are expressed in the ectodermal layer, but are regionally different. These genes are not clustered; nevertheless, they display developmental functions, suggesting that clustering is not required for pattern formation.

More recently, a series of additional transcription factors have been identified in Cnidaria, which belong to the *Pax* gene family. These genes encode factors with multiple DNA-binding domains; they are defined by the presence of the paired box, which had been first identified in the Drosophila pair-rule gene *paired*, and encode a large DNA-binding domain. This domain is very similar to the homeodomain structure; some of these transcription factors contain a homeobox domain upstream of the paired box. Expression studies revealed that the cnidarian *Pax* gene is expressed in

transectodermal cells, very likely in areas where neurons differentiate.

Besides the *paired*-class homeobox genes, Antennapedia-class homeobox genes have also been identified, which are expressed in the head region and the apical pole. Until now, the LIM-class homeodomain proteins, known to be involved in the anterior organizer activity in chordates, have not been found in Cnidaria. These LIM-homeodomain proteins control cellular differentiation by activating transcription of cell type–specific genes.

4.4.4 Porifera

After the new view of the monophyletic origin of all metazoans, including Porifera, and the findings that both the immune system and the body plan comprise common basic elements in all animals, it was not too surprising that the ancestral homeobox genes are also present in sponges. The developmental processes resulting in the formation of a body axis require a head center; for example, in bilaterians, the Spemann's organizer. The genes that are involved in the establishment of the head organizer during embryogenesis have been grouped into three classes of *homeobox* genes, the *paired*-class, the *Antennapedia*-class, and the *LIM*-class genes.

In this area, rapid progress has been made in sponges in the last few years. A *paired*-class (*Pax-2/5/8*)-gene had been isolated from the freshwater sponge *Ephydatia fluviatilis*, which encodes a complete although substantially degenerated homeodomain. In *S. domuncula* a cDNA encoding an LIM/homeobox protein has been isolated, which comprises high sequence similarity to the related LIM-homeodomain proteins in other animals. The potential function of this protein in the sponge was elucidated in the primmorph model. If these 3D-cell aggregates were cultured on a homologous galectin matrix on which they form canal-like structures, morphogenetic processes are triggered, which also involve the LIM/homeobox protein/transcription factor.

A further putative homeobox gene, the *Iroquois*, was isolated from *S. domuncula*. The putative Iroquois transcription factor was found to be expressed in those cells that are adjacent to the canal system of this species. The finding that in sponges also the expression of the *Iroquois* gene is restricted to a specific tissue region, here the aquiferous system, adds a further piece to understanding the complexity of tissue organization in sponges. This finding underscores the fact that spatial and temporal expression of genes is one key process that establishes axes in Metazoa.

As mentioned earlier, the small hydrophobic molecule, retinoic acid, plays an important role in local signaling of homeodomain factor-mediated vertebrate development. As the first metazoan nuclear hormone receptor, the RXR had been identified in the phylogenetic oldest phylum Porifera. After identification, the role of the RXR was studied in the primmorph system. As in vertebrates or in Drosophila, retinoic acid is also formed in *S. domuncula* from ß-carotene via cleavage by the ß,ß-carotene-15,15′-dioxygenase to retinal and on further oxidization, to retinoic acid. Retinoic acid (9-*cis*-retinoic acid) binds to RXR resulting in a regulation of transcriptional activity of morphogenetic genes, including a homeobox gene also.

5 Origin of Metazoan Pattern Formation

The major difference in the genetic control of the body plan patterning between

diploblasts and triploblasts should be seen in the fact that only the triploblasts (both Protostomia and Deuterostomia) are composed of clusters of homeotic genes. In diploblasts, these clusters are missing and nonclustered *paired*-class (*Pax-2/5/8*)- and *LIM/homeodomain* genes control pattern formation. In triploblasts as well as in diploblasts, body plan formation starts by polarization of the embryo and subsequent development is followed in diploblasts only to the stage of serial module patterning, while in triploblasts the development to the adults includes both compartmental patterning and the subsequent segmental patterning (Fig. 11).

5.1 Serial Module Construction of the Body Plan in Diploblasts

Focusing on Porifera, after fertilization the zygote increases in size and develops flagellae. Depending on the taxa, morphologically slightly different types of larvae (amphiblastulae, parenchymellae, or coeloblastulae) are formed and released from the maternal body. During this phase the embryo polarizes, a process that is morphologically primarily obvious in the localization of the cilia and the formation of the body cavity (Fig. 11). Then gastrulation takes place driven by an asymmetric and tangentially arranged cleavage. The epithelia within the body cavity invaginate under formation of the choanocyte chambers. The embryos become sessile and the young sponge forms an oscule through which the body cavity opens to the external milieu; with this process, the young sponge forms an oscular/apical–basal axis.

The oscule region – as mentioned earlier – can be considered as an organizer, since there the characteristic vertebrate organizer genes are expressed. In addition, this region is provided with a regeneration capacity, which is distinguished from the rest of the animal. Especially the organizer-specific genes, *LIM-homeodomain, Brachyury, Frizzled receptor, noggin,* and *Iroquois,* are expressed along the aquiferous system and the oscule. This water canal system is the combined route for feeding, secretion, and gas exchange. Strong evidence has been collected in the last two years, which indicates that these "organizer" genes are expressed around the canals and the oscule and display morphogenetic activity as well.

Already morphological inspection shows that sponges grow in a radiate accretive pattern, regardless of whether they inhabit an environment with pronounced seasonal changes, for example, of the temperature (Lake Baikal; Fig. 7a), or a more stable milieu (Antarctic, or tropic regions; Fig. 7b). X-ray analysis of the skeleton of the Lake Baikal sponge *L. baicalensis* reveals that the architecture of the specimens comprises a highly ordered arrangement of the spicules within the body (Fig. 7a). *L. baicalensis* follows an organized growing pattern, reflecting a radiate accretive growth process. This pattern is maintained throughout the body of the animals and can be seen both in the growing parts of the specimens, the tips, and the basis. During growth, new layers/rings of new longitudinal spicule bundles are added at the tip of the branches. A similar highly regular arrangement of the spicule formation can also be seen in the hexactinellid *E. aspergillum* (Fig. 7b). The latter species shows a regular arrangement of the spicules, which form the lateral stabilization network that leaves holes for the water canal system. The apical end of the specimens, which harbors the oscule(s),

also shows a considerably ordered arrangement. By morphological inspection it has become obvious that specification of the different units of the body is absent; the growth zones look identical.

There is no doubt that this highly ordered growth process of the Porifera is genetically controlled. With the discovery that organizer-specific genes are present at the tips of the branches, around the oscules, together with the fact that growth along the oscular/apical–basal axis proceeds by a radiate accretive pattern, it is suggested that at the "growth lines," which are readily seen between the preceding and the succeeding growth layers, genes are expressed which form "serial modules," along the oscular/apical axis (Fig. 11). The serial modules should not be termed segments because this would imply that they are built by a more complex genetic network, as a system seen in insects.

The growth of the Coelenterata follows the pattern seen in Porifera. In *Hydra*, a continuous growth proceeds with organization center(s) localized around the mouth. In these animals, organizer-specific genes are highly expressed around the apical mouth region. Identical to sponges the coral growth pattern also follows the radiate accretive type. The apical/basal axis formation is modulated in these diploblasts by morphogens, which have two centers in Coelenterata, the foot and the mouth, while in sponges apparently only one center exists, which is found around the oscule.

5.2 Compartmental/Segmental Construction of the Body Plan in Triploblasts

Both in Deuterostomia (e.g. vertebrates) and in Protostomia (insects) development starts by polarization of the fertilized egg, a process that is driven by differential gene expression. While in vertebrates axis formation is determined by a signaling center, the Nieuwkoop center during the initial cleavages, and the Spemann organizer in the late blastula, in insects, body axis is set up by asymmetrical expression of the maternal genes.

5.2.1 Protostomia (Example, Insects)

A series of maternal genes are locally differently expressed in the egg and establish the anteroposterior axis. By spatial distribution of the mRNAs and, subsequently, of the resulting proteins, a regionalization within the egg is established. Among these genes is the *bicoid*, which is expressed at the anterior end of the egg. The bicoid protein acts as a transcription factor and as a morphogen and by doing so switches on zygotic genes at different threshold concentrations; thus, it initiates a new pattern within the zygote along the anteroposterior axis. Other genes, for example, *nanos*, polarizing the embryo from the posterior end act like a morphogen, while *hunchback* is more uniformly distributed within the egg. Hence, the anteroposterior axis of the egg is determined by several oppositely directed gradients. Besides the anteroposterior axis, the dorsal–ventral axis is determined in the egg through a later set of oocyte–follicle cell interactions. The oocyte-surrounding follicle cells specify the ventral side of the oocyte by deposition of maternal proteins.

After fertilization, four main classes of zygotic genes are activated along the anteroposterior axis; these are the *gap* genes, the *pair-rule* gene clusters, the *segment polarity* genes and the *selector homeotic* gene cluster. The *gap* genes, the first zygotic genes to be expressed, are also transcription factors. Their expression is initiated

by the bicoid gradient in the syncytial embryo; resulting in a coordinated expression of the *gap* genes along the anteroposterior axis in the following order: *hunchback*, *giant*, *Krüppel*, and *knirps*. The perfect regular segmentation of the insect larva seen at its cuticle starts with the expression of the *pair-rule* gene clusters, which delimit the parasegments. During this process, the blastoderm of the larva is cellularized. The boundaries of these periodic parasegment patterns are fixed by the segment polarity proteins, for example, engrail, which act on cellular levels and cause lineage restriction are also termed *Compartmentalization*. The "segment polarity" proteins stabilize the position not only of the parasegments but also of the segments. Then, the identity of the segments, based on the arrangement of the parasegments, is conferred by the selector/homeotic genes. The spatial expression pattern of the two clusters, the Antennapedia and the bithorax complex, is determined by the preceding expression of the *gap* genes and *pair-rule* genes.

5.2.2 Deuterostomia (e.g. Vertebrates)

In contrast to the initial phases of the development seen in insects, the patterning of the vertebrate body plan occurs always on the cellular level. Hence, intercellular signaling molecules are key players that interact with cell surface receptors and in turn induce intracellular signal transduction pathways; examples are the fibroblast growth factor, hedgehog, Wnt proteins, or ephrins. As in insects, in vertebrates also the eggs comprise a distinct polarity (animal pole – vegetal pole), which influences the cleavage pattern. The site at which the sperm interacts with the egg determines the dorsal–ventral polarity of the embryo. The following movement of the cytoplasm/yolk material is the first important developmental process, termed *cortical rotation*. In turn, a signaling center becomes localized in the embryo, which is termed *Nieuwkoop* center; it determines the dorsal side.

The Nieuwkoop center has the potential to induce a second embryo at another site. Its molecular nature is that it consists initially of a uniform low level of transforming growth factor-ß (TGF-ß) signal expressed all over the embryo. Acting alone, TGF-ß induces the formation of premesoderm and generates a BMP signal, which causes the transformation to the mesoderm. The major role of the Nieuwkoop center is the specification of a further dorsal signaling center, the Spemann organizer. The latter center is located in the immediate neighborhood of the Nieuwkoop center at the late blastula/early gastrula. The anteroposterior axis is the result of the action of this Spemann organizer (in vertebrates), or the Hensen's node in chick. Both are localized at the dorsal lip of the blastopore. The centers mediate their morphogenetic function through their transcription factors, for example, Brachyury and LIM protein, and through secreted proteins, for example, noggin, all of which are molecules that are highly conserved from Porifera to vertebrates.

As a result of the morphogenetic activity displayed by molecules originating from the Spemann organizer, the embryo finishes gastrulation and establishes the main axial structures, the somites (originating from the mesoderm), the notochord (mesoderm), and the neural tube (ectoderm). Interestingly enough, the formation of the somites and the notochord is under the control of the paired-type homeobox gene family *Pax*, again molecules conserved from Porifera to Chordata. The expression of *Pax* is subsequently modulated by BMP as well as by Wnt family

proteins so that the somites confine muscle precursors. This process could be termed *Compartmentalization*. Finally, the positional identity of the somites along the anteroposterior axis is specified by the sequential expression of the *Hox* gene cluster. This specification is reflected by the differences seen in the vertebrae along the anteroposterior axis. Hence, the vertebrate *Hox* gene clusters parallel the function of the insect homeotic selector gene clusters and follow a "somite" pattern. The transcription factors that are expressed in a temporal and spatial defined order are also arranged in the same order on the chromosome. Again – like in insects – the anterior border of the expression is more sharply defined than the posterior one. Several *Hox* gene clusters are expressed in vertebrates; all of them are involved in the specification of the regional identity.

6 General Principles in Metazoan Pattern Formation

The genetic programs that control development and body plan formation of multicellular organisms, plants, and animals, have been shaped during evolution. In consequence, such gene networks that control development also provide information about the evolution of these organisms. During development, distinct cell types are generated in distinct spatial and morphological domains, which establish organized structures and a functional organism. After fertilization of an egg, a genetic program controls cleavage divisions, resulting in a rapid increase of cells during the different stages of embryogenesis, a phase that is usually not characterized by an increase in cell mass. In these phases, pattern formation starts a process that organizes the cells of different affinities and activities in a spatial and temporal pattern/way.

Setting up the body axis during pattern formation is different between animals and plants; in plants, cell migration and tissue movement are absent and also distinct differences in embryogenesis exist with the lack of gastrulation as an example. In plants, the "set aside" of undifferentiated cells in the meristems in shoot and root tips is a dominant principle to allow and maintain continuous cell division at each axis. In animals, developmental mechanisms involve asymmetric cell division, pattern formation, establishment of the form and growth. Until sequence data became increasingly available, only a few model organisms had been studied to elucidate the basic principles in developmental biology of animals. These were primarily the protostomians *D. melanogaster* and *C. elegans* and the deuterostomians, for example, mouse and frog. At present, we have surpassed the threshold for a comprehensive view in body-plan formation in Metazoa, including axial pattern formation, which will also contribute to a solution in completing the hypothesis in developmental biology, for example, in segmentation/somite formation. Until 10 years ago, it was not clear whether all metazoan phyla, including the Porifera, originated from one common ancestor and hence are monophyletic, or, that multicellularity in animals occurred more than once. In 1994, sequence data clearly showed that the cell–cell and cell–matrix adhesion molecules of similar structure and function are present in all animal phyla from Porifera to the crown triploblastic phyla, a view that resulted in the formulation of the first hypothetical metazoan ancestor, the Urmetazoa.

In parallel with the elucidation that in all metazoans the characteristic adhesion molecules are functional, it became increasingly obvious that also individuality of a metazoan specimen, including its immune system, is based on conserved molecules and signal transduction pathways. Genes and their functional proteins, which control cell behavior and cell fate were identified, also including the process of apoptosis, which is required for the elimination of unwanted cells that remain at a later differentiation stage. Inductive interaction between cells is one major characteristic in development. The underlying signals are either secreted diffusible molecules, which interact with cell surface receptors and elicit there a new gene expression program, or signals that pass from one cell to the other via junctions. The response of a given cell to inductive signals is strongly dependent on its actual state.

During embryogenesis, pattern formation requires a spatial and temporal gene expression resulting in the formation of one or several axes. While diploblasts, for example, Porifera and Coelenterata, have only two epithelial layers, triploblasts have three germ layers; furthermore, diploblasts are provided only with one axis (apical/oscular–basal), while triploblasts/bilaterians are provided with at least two (anteroposterior and dorsal–ventral). Axis formation requires a polarized organization center. In Porifera, it is the oscular region that lays down the body axis, and in Coelenterata it is the mouth region. Common to these centers in the diploblasts, and the Spemann organizer and Nieuwkoop center, existing in the triploblasts/Deuterostomia, is their repertoire on expressed genes, for example, *T-box* family and *Brachyury* or *noggin*. It is noteworthy that these molecules are conserved from Porifera to the triploblastic crown species. In Protostomia (insects), the axis formation starts with the polarized expression of the maternal gene in the egg.

Pattern formation in diploblasts and triploblasts is also strikingly initiated by homeobox genes, for example, *paired*-class and *LIM*-class ones. Again, these genes have been isolated and (partially) functionally analyzed from Porifera to triploblasts. However, with respect to these genes, a characteristic difference exists between the arrangement of the sponge and *Cnidaria*/*Hydra* gene on the chromosome and those in triploblasts. In insects and in vertebrates, the morphogenetic homeobox genes are arranged in clusters. In insects, the *pair-rule* genes and the homeotic selector genes are clustered on the chromosomes, and likewise in vertebrates the four *Hox* gene complexes are arranged and expressed in a colinear way. Since the body plan/axis formation in diploblasts is established in the absence of homeobox gene clustering, it has been suggested that clustering of the homeobox genes is not a prerequisite for axis formation. As one consequence of the absence of homeobox gene clusters, the body axis in diploblasts is established in a compartmental construction, while in triploblasts, compartmental arrangement of the tissue stripes is the basis for a segmental/somital construction of the body plan.

In conclusion, this survey provides a basis for a unifying concept of pattern formation in all metazoans, which is based on genes expressing soluble morphogens that interact with their corresponding receptors and morphogenetically active transcription factors, including homeodomain proteins.

Acknowledgment

This work was supported by a grant from the Deutsche Forschungsgemeinschaft (Mü 348/12-4 and Mü/14-4), the Bundesministerium für Bildung und Forschung (project: Center of Excellence *BIOTEC*marin), and the International Human Frontier Science Program (RG-333/96-M).

See also Cellular Interactions; Matrix, Extracellular and Interstitial; Principles and Applications of Embryogenomics; Regulome; Stem Cells; Stem Cells, Embryonic; Variation and Conservation of the Development of Animal Bodies: Evolutionary Developmental Biology.

Bibliography

Books and Reviews

Dewel, R.A. (2000) Colonial origin of Eumetazoa: major morphological transitions and the origin of bilaterian complexity, *J. Morphol.* **243**, 35–74.

Galliot, B., Miller, D. (2000) Origin of anterior patterning – how old is our head? *Trends Genet.* **16**, 1–5.

Grunz, H. (Ed.) *The Vertebrate Organizer*, Springer-Verlag, Berlin, Heidelberg, New York.

Maldonado, M. (2004) Choanoflagellates, choanocytes, and animal multicellularity, *Invertebr. Biol.* **123**, 1–22.

Müller, W.E.G. (1995) Molecular phylogeny of Metazoa (animals): monophyletic origin, *Naturwissenschaften* **82**, 321–329.

Müller, W.E.G. (Ed.) (1998) *Molecular Evolution: Evidence for Monophyly of Metazoa*, Progr. Molec. Subcell. Biol. 19, Springer-Verlag, Berlin, Heidelberg, New York.

Müller, W.E.G. (2001) How was metazoan threshold crossed: the hypothetical Urmetazoa, *Comp. Biochem. Physiol. [A]* **129**, 433–460.

Müller, W.E.G., Wiens, M., Adell, T., Gamulin, V., Schröder, H.C., Müller, I.M. (2004) Bauplan of Urmetazoa: basis for genetic complexity of metazoa, *Int. Rev. Cytol.* **235**, 53–92.

Nüsslein-Volhard, C., Wieschaus, E. (1980) Mutations affecting segment number and polarity in *Drosophila*, *Nature* **287**, 795–801.

Simpson, T.L. (1984) *The Cell Biology of Sponges*, Springer-Verlag, New York.

Spemann, H. (1936) *Experimentelle Beiträge zu einer Theorie der Entwicklung*, Springer-Verlag, Berlin, Germany.

Wolpert, L. (1998) *Principles of Development*, Oxford University Press, Oxford, England.

Primary Literature

Adell, T., Grebenjuk, V.A., Wiens, M., Müller, W.E.G. (2003) Isolation and characterization of two T-box genes from sponges, the phylogenetically oldest metazoan taxon, *Dev. Genes Evol.* **213**, 421–434.

Chevreux, B., Pisterer, T., Drescher, B., Müller, W.E.G., Wetter, T., Suhai, S. (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs, *Genome Res.* **14**, 1147–1159.

Gamulin, V., Müller, I.M., Müller, W.E.G. (2000) Sponge proteins are more similar to those of *Homo sapiens* than to *Caenorhabditis elegans*, *Biol. J. Linnean Soc.* **71**, 821–828.

Kruse, M., Müller, I.M., Müller, W.E.G. (1997) Early evolution of metazoan serine/threonine- and tyrosine kinases: identification of selected kinases in marine sponges, *Mol. Biol. Evol.* **14**, 1326–1334.

Technau, U., Scholz, C.B. (2003) Origin and evolution of endoderm and mesoderm, *Int. J. Dev. Biol.* **47**, 531–539.

Wiens, M., Diehl-Seifert, B., Müller, W.E.G. (2001) Sponge Bcl-2 homologous protein (BHP2-GC) confers selected stress resistance to human HEK-293 cells, *Cell Death Differ.* **8**, 887–898.

Wimmer, W., Perovic, S., Kruse, M., Krasko, A., Batel, R., Müller, W.E.G. (1999) Origin of the integrin-mediated signal transduction: functional studies with cell cultures from the sponge *Suberites domuncula*, *Eur. J. Biochem.* **178**, 156–165.

Spermatogenesis, Molecular Biology of: *see* Male Reproductive System: Testis Development and Spermatogenesis

Spina Bifida, Molecular Biology of: *see* Neural Tube Defects, Molecular Biology of

Sphingolipid and Glycolipid Biology

Anthony H. Futerman
Weizmann Institute of Science, Rehovot, Israel

1 Introduction 312

2 Structures of Sphingolipids and Glycolipids 313

3 Metabolism and Intracellular Transport of Sphingolipids and Glycolipids 313

4 Functions of Sphingolipids and Glycolipids 317
4.1 Signaling 317
4.2 Functions of Complex Glycolipids 319

5 Sphingolipids and Glycolipids in Disease 320

6 Perspectives 322

Bibliography 322
Books and Reviews 322
Primary Literature 322

Keywords

Ceramide
The building block of sphingolipids and glycosphingolipids and an important second messenger that inhibits proliferation and promotes apoptosis.

Ganglioside
Glycosphingolipids containing one or more sialic acid residues in their glycan portion.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

Glycolipid
A sphingolipid that contains a head group consisting of simple or complex glycans.

Raft
A membrane microdomain enriched in sphingolipids and cholesterol.

Sphingolipid Storage Diseases
A class of inherited metabolic disorders caused by the defective lysosomal degradation of sphingolipids or glycosphingolipids.

Sphingosine-1-phosphate
A sphingolipid first and second messenger that stimulates growth and suppresses apoptosis.

Sphingolipids are ubiquitous components of cell membranes, and are found in plants, animals, and fungi, as well as in some unicellular organisms and prokaryotes. Sphingolipids consist of a sphingoid long-chain base to which a fatty acid is attached at carbon 2, and a variety of polar and nonpolar head groups at C-1; glycosphingolipids are sphingolipids containing one or more sugar moieties in their head groups. Metabolic pathways have been determined for most sphingolipids, but little is known about how these pathways are regulated. Recent interest in sphingolipids and glycosphingolipids has focused largely on their roles in signaling events, with ceramide and sphingosine-1-phosphate the focus of particular attention. Sphingolipids also play vital roles in the formation and stability of membrane microdomains (rafts), and complex glycosphingolipids play roles in recognition. These lipids are also involved in disease pathology, and the pathways of sphingolipid metabolism, as well as the lipids themselves, have been implicated as possible therapeutic targets and agents.

1 Introduction

Biological membranes contain three major classes of lipids, namely, glycerolipids such as phosphatidylcholine and phosphatidylinositol, sterols such as cholesterol and ergosterol, and sphingolipids (SLs) such as sphingomyelin and gangliosides. Until recently it was believed that SLs were no more than structural components of biological membranes having no regulatory roles, but studies initiated in the mid-1980s have demonstrated that SL biology is much more complicated than once thought. It is now known that SLs act as important signaling molecules in an analogous fashion to the better-described signaling functions of glycerolipids such as phosphatidylinositol and its derivates. These findings have generated tremendous excitement and renewed interest in SLs, and

in the complex glycosylated SLs, the glycosphingolipids (GSLs), which is reflected in the increased number of publications on this lipid class. For instance, sphingosine-1-phosphate was virtually unheard of until the early 1990s, and in 1996, the time of publication of the first edition of the "Encyclopedia of Molecular Cell Biology and Molecular Medicine," only 35 citations to sphingosine-1-phosphate can be found in Medline; in striking contrast, more than 230 citations to this important lipid second messenger were found in 2003 alone. This is typical of the renewed interest in all SLs, including ceramide, sphingomyelin, and of no less importance, the GSLs. Thus, the last 20 years have seen an explosion of interest in this field, and the goal of this chapter is to briefly summarize progress and our current state of knowledge of SL and GSL biology. As J.L.W. Thudichum, the discoverer of SLs, stated in his treatise on the chemical constitution of the brain, first published in 1884, SLs are indeed of fundamental importance. However, as reflected in the name chosen for these lipids (inspired by the Sphinx of Greek mythology, whose raison d'étre was enigmatic, and still is), many questions remain unresolved concerning SL and GSL biology. Accordingly, although perhaps not quite as enigmatic as once thought, challenges in this field still abound.

2 Structures of Sphingolipids and Glycolipids

As is the case for all membrane lipids, SLs are amphipathic molecules having both hydrophobic and hydrophilic regions. However, unlike the glycerolipids, in which two fatty acids are esterified to the primary and secondary hydroxyl groups of glycerol, SLs contain only one fatty acid (although up to 20 different species of fatty acid varying in chain length, degree of saturation, and degree of hydroxylation can be attached to the sphingoid long-chain base), which is attached via an amide bond to a sphingoid long-chain base, normally sphingosine (*trans*-D-erythro-2-amino-4-actadecene-1,3-diol), sphinganine, or phytosphingosine (Fig. 1). The simplest SL, ceramide, has a minimal hydrophilic region comprising two OH-groups, whereas complex SLs contain a hydrophilic moiety such as phosphate in the case of ceramide-1-phosphate, phosphorylcholine in the case of sphingomyelin (SM), and sugar residues in GSLs. At least 500 different carbohydrate structures have been described in GSLs. Both neutral and acidic GSLs exist, and particular variety is found in the acidic GSLs, the gangliosides, which are defined as GSLs containing one or more sialic acid residues, with sialic acid being a generic term for *N*-acetyl- or *N*-glycolyl-neuraminic acid. Gangliosides can be divided into at least eight different classes depending on the type of sugars linked to ceramide; a selection of typical GSL structures is shown in Table 1. The number of possible individual SL and GSL structures, varying in sphingoid base, fatty acid, and hydrophilic head group, is therefore huge, but little or nothing is known at present about how this enormous complexity is generated and maintained, or about the possible specific functions of individual SLs and GSLs.

3 Metabolism and Intracellular Transport of Sphingolipids and Glycolipids

The biochemical pathways of SL synthesis and degradation are well documented, and

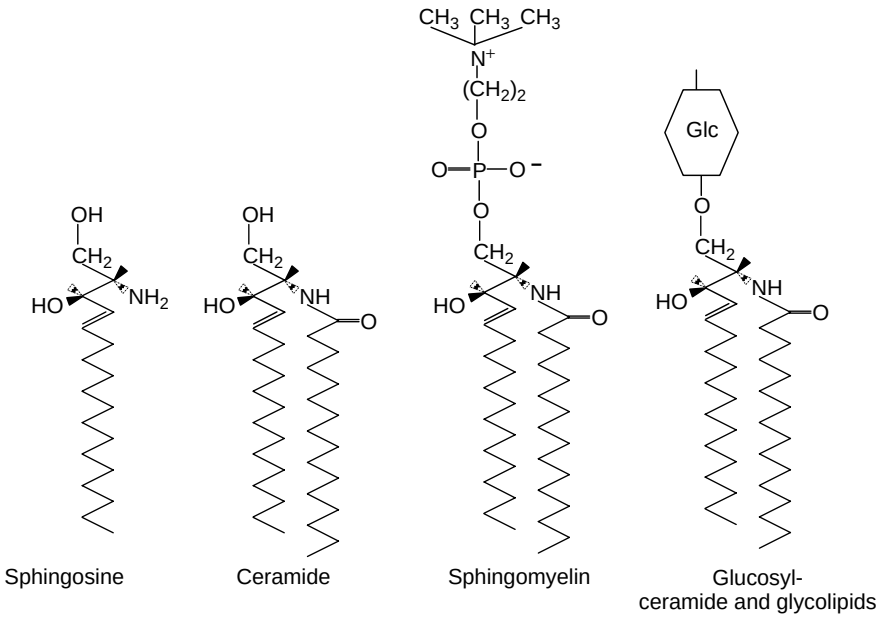

Fig. 1 The structure of simple SLs. For simplicity, only one kind of sphingoid base (sphingosine) is shown, to which one kind of fatty acid (palmitic acid) is attached. Sequential addition of other sugar residues to Glc(β1 $\rightarrow$ 1)Cer results in the formation of complex GSLs (Table 1 and Fig. 2).

most of the enzymes involved have been cloned, albeit some of them recently. SL synthesis begins with the condensation of palmitoyl CoA and serine-by-serine palmitoyl transferase to yield 3-ketosphinganine, from which sphinganine is produced by the action of 3-ketosphinganine reductase (Fig. 2). These two reactions, as well as the acylation of sphinganine to yield dihydroceramide, occur on the cytosolic leaflet of the endoplasmic reticulum, although the precise topology of each of these enzymes has not been fully elucidated. Sphinganine and sphingosine (the former produced mainly from the biosynthetic pathway and the latter via the degradative pathway) can be phosphorylated by sphingosine kinase to yield sphinganine-1-phosphate and sphingosine-1-phosphate, respectively.

The location of dihydroceramide reductase is not known, but most subsequent reactions, involving addition of head groups, occur in the Golgi apparatus. As the hydrophobic nature of ceramide prevents its spontaneous transfer through the cytosol, ceramide travels from the endoplasmic reticulum to the Golgi apparatus by facilitated vesicular and nonvesicular transport mechanisms. The latter involves a protein, CERT, which specifically extracts ceramide from membranes, is absolutely required for sphingomyelin synthesis, and contains a pleckstrin homology domain to allow its targeting to the Golgi apparatus. Interestingly, CERT does not appear to be involved in glucosylceramide formation, which occurs on the cytosolic leaflet of the Golgi apparatus, whereas addition of all other sugar moieties occurs on the

Tab. 1 The structure of a selection of GSLs and gangliosides.

GSLs	***Structure***
GalCer	Gal(β1 $\rightarrow$ 1)Cer
Sulfatide	HSO_3-3Gal(β1 $\rightarrow$ 1)Cer
GlcCer	Glc(β1 $\rightarrow$ 1)Cer
LacCer	Gal(β1 $\rightarrow$ 4)GlcCer
GM3	NeuAc(α2 $\rightarrow$ 3)Gal(β1 $\rightarrow$ 4)Glc(β1 $\rightarrow$ 1)Cer
GM2	GalNAc(β1 $\rightarrow$ 4)Gal[(3 $\leftarrow$ 2α) NeuAc](β1 $\rightarrow$ 4)Glc(β1 $\rightarrow$ 1)Cer
GM1	Gal(β1 $\rightarrow$ 3)GalNAc(β1 $\rightarrow$ 4)Gal[(3 $\leftarrow$ 2α)NeuAc](β1 $\rightarrow$ 4)Glc(β1 $\rightarrow$ 1)Cer
GD1a	NeuAc(α2 $\rightarrow$ 3)Gal(β1 $\rightarrow$ 3)GalNAc(β1 $\rightarrow$ 4)Gal[(3 $\leftarrow$ 2α) NeuAc](β1 $\rightarrow$ 4)Glc(β1 $\rightarrow$ 1)Cer
GT1a	NeuAc(8 $\rightarrow$ 2α) NeuAc(α2 $\rightarrow$ 3)Gal(β1 $\rightarrow$ 3)GalNAc(β1 $\rightarrow$ 4)Gal-[(3 $\leftarrow$ 2α)NeuAc](β1 $\rightarrow$ 4)Glc(β1 $\rightarrow$ 1)Cer
GD3	Gal[(3 $\leftarrow$ 2α) NeuAc(8 $\leftarrow$ 2α) NeuAc](β1 $\rightarrow$ 4)Glc(β1 $\rightarrow$ 1)Cer
GD2	GalNAc(β1 $\rightarrow$ 4)Gal[(3 $\leftarrow$ 2α)NeuAc(8 $\leftarrow$ 2α) NeuAc](β1 $\rightarrow$ 4)Glc(β1 $\rightarrow$ 1)Cer
GD1b	Gal(β1 $\rightarrow$ 3)GalNAc(β1 $\rightarrow$ 4)Gal[(3 $\leftarrow$ 2α)NeuAc(8 $\leftarrow$ 2α)NeuAc](β1 $\rightarrow$ 4)Glc(β1 $\rightarrow$ 1)Cer
GT1b	NeuAc(α2 $\rightarrow$ 3)Gal(β1 $\rightarrow$ 3)GalNAc(β1 $\rightarrow$ 4)Gal[(3 $\leftarrow$ 2α) NeuAc(8 $\leftarrow$ 2α)NeuAc](β1 $\rightarrow$ 4)Glc(β1 $\rightarrow$ 1)Cer
GQ1b	NeuAc(8 $\rightarrow$ 2α)NeuAc(α2 $\rightarrow$ 3)Gal(β1 $\rightarrow$ 3)GalNAc(β1 $\rightarrow$ 4)Gal-[(3 $\leftarrow$ 2α)NeuAc(8 $\leftarrow$ 2α)NeuAc](β1 $\rightarrow$ 4)Glc(β1 $\rightarrow$ 1)Cer
9-OAc-GD3	Gal[(3 $\leftarrow$ 2α)NeuAc(8 $\leftarrow$ 2α)(9-OAc)NeuAc](β1 $\rightarrow$ 4)Glc(β1 $\rightarrow$ 1)Cer
Asialo-GM2 (GA2)	GalNAc(β1 $\rightarrow$ 4)Gal(β1 $\rightarrow$ 4)Glc(β1 $\rightarrow$ 1)Cer
Asialo-GM1 (GA1)	Gal(β1 $\rightarrow$ 3)GalNAc(β1 $\rightarrow$ 4)Gal(β1 $\rightarrow$ 4)Glc(β1 $\rightarrow$ 1)Cer

Notes: Cer: ceramide; GalNAc: *N*-acetylgalactosamine; GlcNAc: *N*-acetylglucosamine: NeuAc: *N*-acetylneuraminic acid; 9-Oac: 9-O acetyl.

lumenal leaflet of the Golgi apparatus, as does the formation of sphingomyelin. Thus, CERT may be involved in regulating the synthesis of specific SLs by either its intrinsic specificity, or its interaction with dihydroceramide synthase, or with proteins involved in the transbilayer movement of SLs and GSLs in the Golgi apparatus. Subsequent formation of GSLs, including gangliosides, occurs by the stepwise addition of sugar residues by glycosyltransferases, which have defined substrate specificities but act on more than one GSL (Fig. 2). After their formation in the Golgi apparatus, GSLs and sphingomyelin move by vesicular transport to the plasma membrane, where they reside.

Although once believed to be restricted to the endomembrane system (i.e. those intracellular organelles connected by vesicular transport pathways), emerging evidence suggests that SLs and GSLs can also be found, albeit in small amounts, in other organelles such as the nucleus and the mitochondria. Thus, ganglioside GD3 relocates to mitochondria in response to some inducers of apoptosis, a neutral ceramidase has been found in mitochondria, and there is evidence for a sphingomyelin cycle in the nucleus. Little if anything is known about the dynamic interplay

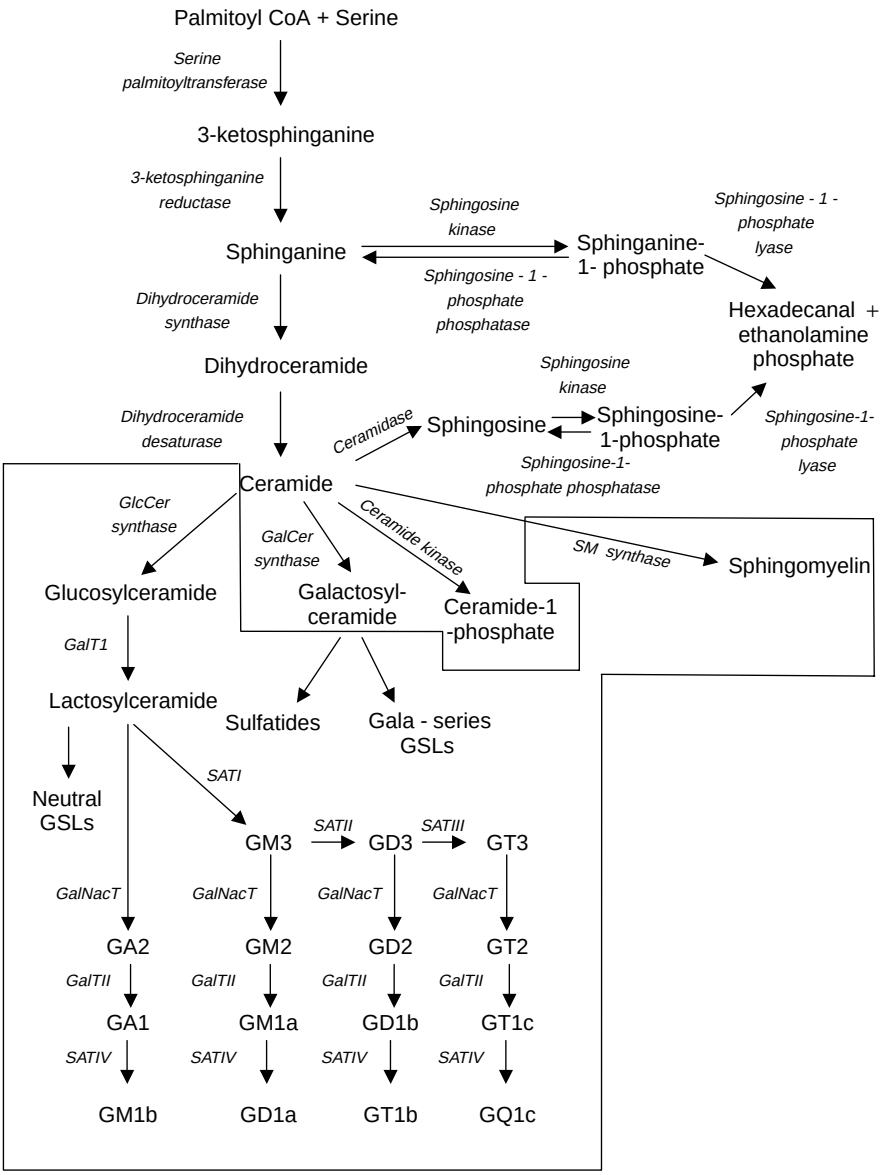

Fig. 2 The biosynthesis of SLs and GSLs. Not all metabolic pathways are shown. Enzymes are in italics. For GSL catabolism, see Table 2. Reactions within the box occur in the Golgi apparatus; all other reactions take place in the endoplasmic reticulum or cytosol.

between SL metabolism in the organelles of the endomembrane system and in other organelles not connected to them by classical vesicular transport pathways.

SLs and GSLs are internalized from the plasma membrane by vesicular transport involving both clathrin-mediated pathways and other pathways, including those

mediated by caveolin. Once internalized to endosomes, SLs and GSLs can either be recycled back to the plasma membrane or be further internalized to lysosomes, where their breakdown occurs by stepwise degradation, eventually yielding ceramide; SL and GSL degradation is defective in a number of inherited metabolic diseases, which will be discussed in more detail in Sect. 5. Ceramide is degraded by ceramidases to regenerate sphingoid bases. These bases act as substrates for the sphingosine kinases that form sphingosine-1-phosphate, which is subsequently degraded by sphingosine-1-phosphate lyase to yield hexadecanal and ethanolamine phosphate, providing a metabolic link between SLs and glycerolipids, and the only exit route of SL products from the SL metabolic pathway.

SL synthesis and degradation can be regulated at multiple levels, including through the control of enzyme expression, post-translational modification, and allosteric mechanisms. Some of these are cell-type specific, so as to control which SLs are synthesized during different stages of cell development, and some are in response to specific signals. Hence, although the basic blueprint of the biochemical pathways is well described, much still remains to be learned about modes of regulation of individual enzymes, and more importantly, about how flux through the metabolic pathways, and particularly at branch points where more than one enzyme can act on the same substrate, is regulated.

4 Functions of Sphingolipids and Glycolipids

Multiple functions have been ascribed to SLs and GSLs, and one of the challenges over the past couple of decades has been to make sense of the intracellular signaling functions newly ascribed to the simple SLs, mainly ceramide and sphingosine-1-phosphate, and of the functions ascribed to complex GSLs, mainly in events that occur at the cell surface. Elucidation of the exact function(s) that specific SLs and GSLs play in cell physiology, and how these functions are integrated and regulated, will represent a major achievement.

4.1 Signaling

Part of the reason for the recent explosion of interest in SL biology is the role that these lipids, particularly sphingosine, ceramide, sphingosine-1-phosphate, and ceramide-1-phosphate, have been shown to play in intracellular signaling. Ceramide and sphingosine-1-phosphate are the best-studied bioactive SLs and are produced for signaling pathways via the regulated activity of various enzymes. Thus, ceramide levels can be elevated in response to signals by activation of sphingomyelinases, or by *de novo* synthesis (Fig. 2). Sphingosine-1-phosphate is formed via the action of sphingosine kinase, but more than one sphingosine kinase exists and each enzyme is regulated by different means.

Ceramide and sphingosine-1-phosphate often exert opposite effects, with ceramide usually inhibiting proliferation and promoting apoptosis, and sphingosine-1-phosphate stimulating growth and suppressing apoptosis. Thus, many inducers of apoptosis or growth arrest activate one or more pathways of ceramide generation, and in many cases, there is evidence that endogenous ceramide mediates, at least in part, the apoptotic and growth arrest responses of these inducers (Fig. 3). Ceramide has been implicated in the actions of, for instance, tumor necrosis factor-α, and in the cytotoxic responses

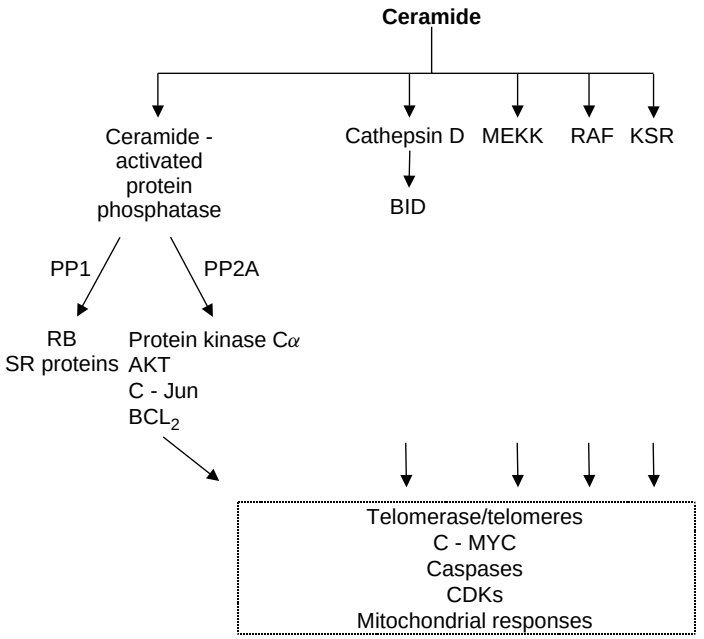

Fig. 3 Some proposed downstream targets of ceramide in signaling pathways. Ceramide generation, either via the *de novo* pathway or via the action of sphingomyelinases, results in a number of cellular responses, some of which are mediated by the signaling proteins and pathways shown in this scheme.

to amyloid ß-peptide, which suggests a role for this pathway in Alzheimer's disease and neurodegeneration. Conversely, many growth factors such as platelet-derived growth factor, as well as tumor necrosis factor-α, have been shown to induce sphingosine-1-phosphate formation, which in turn has been implicated in mediating viability and inflammatory responses. Much progress has been made in the past few years in determining the mechanisms of action of sphingosine (such as through the activation of pkh kinases in yeast), sphingosine-1-phosphate (primarily through the edg/sphingosine-1-phosphate receptors), and ceramide (through phosphatases and cathepsin D; Fig. 3), and such progress should allow the molecular determination of the specific mechanisms and roles of each individual bioactive SL.

Another huge impetus for study of SLs and GSLs was the observation that they are found in detergent-insoluble membrane lipid rafts. These microdomains have been implicated as both signaling platforms, and also as a means to traffic GSLs and certain proteins to specific membranes of polarized cells, such as epithelial cells and neurons. In terms of signaling platforms, it is now believed that acid-sphingomyelinase becomes exposed on the cell surface in response to stimulation via certain receptors, and the ceramide produced thereby modifies preexisting small rafts to fuse into larger signaling platforms and to trap activated receptor molecules in the platforms, thereby enhancing the

efficacy of signaling, and providing a functional and conceptual link between ceramide signaling and SL-enriched microdomains. This kind of mechanism may be particularly important in T-cell signaling. In terms of the role of rafts in sorting, considerable evidence exists that protein targeting can be mediated via rafts, but the fact that a protein is detergent-insoluble, together with SLs and GSLs, does not necessarily imply a functional interaction.

4.2 Functions of Complex Glycolipids

Early studies on complex GSLs were mainly devoted to attempting to understand their roles in the nervous system where GSLs, particularly gangliosides, are expressed at high levels. Gangliosides are expressed at different levels in different regions of the brain during development, suggesting functional roles for the gangliosides in brain development. For instance, during embryogenesis, and during the postnatal period, the hemato-series gangliosides, GM3, GD3, and 9-OAc-GD3, are highly expressed in the brain, but in adult tissue, these gangliosides are found at much lower amounts, so that the ganglioside composition of adult brain differs significantly from that of embryonic brain, inasmuch as adult brain expresses a much wider range of complex gangliosides, with GM1, GD1a, GD1b, and GT1b predominant. However, it was not until the cloning of the glycosyltransferases responsible for GSL synthesis, and subsequent availability of genetically engineered mice, that specific functions for specific gangliosides, or at least classes of gangliosides, began to emerge.

The essential nature of GSLs for sustaining life was shown in knock-out mice lacking glucosylceramide synthase, the first enzyme in the pathway of GSL biosynthesis (Fig. 2); these mice showed embryonic lethality, but whether this was caused by ceramide accumulation, or due to lack of complex GSLs, is unclear. Mice lacking the GM2/GD2 synthase gene (GalNacT, Fig. 2) contain high levels of GM3 and GD3, but no complex gangliosides. Surprisingly, these mice are viable and the nervous system is only mildly affected, although male fertility is severely disrupted due to aspermatogenesis. A number of other knock-out mice have been produced in the ganglioside biosynthetic pathway, and the overall consensus is that there is considerable functional redundancy between different ganglioside species, and that there may be considerable overlap in the roles of individual gangliosides.

One area in which individual gangliosides appear to have well-defined functions is their role as toxin receptors. In order for toxins to reach their intracellular targets, they bind to cell surface receptors, which are then internalized together with the toxins. Two toxins, namely, cholera toxin from *Vibrio cholera* and *Escherichia coli* heat-labile enterotoxin bind to ganglioside GM1, and Shiga toxin from *Shigella dysenteriae* binds to the neutral GSL, globotriaosylceramide (Gb3). The B-subunit of the toxins binds to their respective GSL receptors, and is internalized together with the A-subunit, which carries the catalytic domain. Lipid rafts have been implicated in the internalization mechanism, although there is some disagreement about the precise internalization itinerary undertaken by toxins attached to GSLs. What is clear is that the toxins are eventually targeted to the Golgi apparatus and then to the endoplasmic reticulum, where the A-subunit is released and initiates toxicity.

GSLs at the cell surface are also involved in recognition events that are

beneficial, unlike the toxic effects caused by toxin binding. Thus, a large body of data has shown roles for GSLs as antigens, as mediators of cell adhesion, and as modulators of signal transduction. The list of receptor functions of GSLs is extensive. For instance, GSLs mediate E-selectin-dependent rolling and tethering. α-Galactosylceramide acts as a ligand recognized by a special group of immune T cells, known as *invariant NKT cells*, and activates these cells to produce immunoregulatory cytokines, including interferon-γ and IL-4. 9-O-acetyl GD3, expressed in regions of cell migration and neurite outgrowth in the developing and adult nervous system, plays a role in neuronal motility. GSLs also act as T cell-stimulatory antigens by stimulating T cells expressing TCRαß or $\gamma\delta$ receptors. Together, these varied examples illustrate that the complex glycan structures of GSLs are involved in vital recognition events at the cell surface.

Finally, although this chapter focuses for the most part on SLs and GSLs in mammalian cells, these lipids are also important in lower organisms. Of particular note is the yeast, *Saccharomyces cerevisiae*, in which many of the enzymes of the initial steps in the biosynthetic pathway were first identified and cloned, and defined signaling functions of sphingosine-1-phosphate and ceramide, have been delineated. *Saccharomyces cerevisiae* does not contain complex GSLs, but rather contains inositolphosphorylceramides. Inositol-containing GSLs and glucosylceramide are found in plants and fungi, and galactosylceramide-based GSLs are found in fungi. As is the case in animal cells, SLs are involved in both intracellular processes and in cell-to-cell interactions in plants and fungi. These interactions may occur between cells of a multicellular organism or between cells of different species, as in host–pathogen interactions. In plants, recent studies have indicated that SLs are involved in signal transduction pathways, such as programmed cell death. Sphingenine-1-phosphate is involved in linking the plant hormone abscisic acid to stomatal pore regulation. In fungi, glucosylceramide has been linked to fruit-inducing activity. Inhibition of glucosylceramide synthesis can block morphological transitions such as budding and germ tube formation, and evidence is accumulating that GSLs play roles in fungal cell differentiation. One ongoing area of research is determining whether GSLs are involved in the interaction between mammals and pathogenic fungi.

5 Sphingolipids and Glycolipids in Disease

Much of the early work on SLs and GSLs was stimulated by their accumulation in a number of inherited metabolic diseases caused by defects in the lysosomal hydrolases responsible for their degradation (Table 2). For instance, Gaucher, Tay-Sachs, and Niemann Pick A and B diseases, in which glucosylceramide, ganglioside GM2, and sphingomyelin (SM) accumulate, respectively, are caused by the defective activity of the lysosomal enzymes glucocerebrosidase, β-hexosaminidase, and sphingomyelinase. A variety of symptoms are associated with GSL storage diseases, many of which involve central nervous system dysfunction, but the clinical presentation is very heterogeneous and progressive, with the rate of progression varying between disorders, and also within each disorder, and the diseases normally exist in infantile, juvenile, and adult forms. The most severe infantile forms present with acute brain

Tab. 2 SL and GSL storage diseases.

Disease	***Enzymatic defect***	***Major GSL storage material***
Gaucher	β-glucosidase, saposin C activator	Glucosylceramide
Sphingolipid activator deficiency	Sphingolipid activator protein	Glycolipids
GM1 gangliosidosis	β-Galactosidase	GM1
Tay Sachs	β-Hexosaminidase A	GM2
Sandhoff	β-Hexosaminidase A and B	GM2
GM2 activator deficiency	GM2 activator protein	GM2
Krabbe	β-Galactosidase	Galactosylceramide
Fabry	α-Galactosidase A	Globotriaosylceramide
Metachromatic leukodystrophy	Arylsulfatase A, saposin B activator	Sulfated glycolipids
Farber	Ceramidase	Ceramide
Niemann-Pick A & B	Sphingomyelinase	Sphingomyelin

involvement and patients die within the first years of life. In adult forms, symptoms develop more slowly and disability often arises mainly from peripheral symptoms. Juvenile forms are intermediate between infantile and adult forms. The reason for this heterogeneity is not known, but presumably depends on the level and nature of the accumulating GSL, and on the downstream biochemical and cellular pathways that are altered. However, despite years of study of the genetic and molecular basis of GSL storage disorders, little is known about the events that lead from lysosomal accumulation to the pathological manifestations described above. Presumably, study of these diseases would also yield important information about the normal physiological roles of SLs and GSLs.

In addition to the SL and GSL storage diseases, which result from defective degradation, SLs and GSLs have been implicated in many other diseases, notably cancer. In particular, crucial roles for ceramide and sphingosine-1-phosphate-mediated pathways have been identified in both cancer development and progression, with ceramide functioning as a tumor-suppressor lipid, inducing antiproliferative and apoptotic responses in various cancer cells, and sphingosine-1-phosphate acting as a tumor promotor. For example, an inverse correlation exists between the levels of ceramide and the degree of malignant progression and prognostic severity in patients with malignant astrocytomas. Similarly, the total content of ceramide is decreased in ovarian tumors compared with normal ovarian tissues. Ionizing radiation causes ceramide formation, and mice deficient in acid-sphingomyelinase lose the ability to accumulate ceramide and acquire resistance to radiation-induced apoptosis. In contrast, overexpression of sphingosine kinase results in cellular transformation in tissue culture and tumor formation. Sphingosine-1-phosphate has also been implicated in inflammatory pathways, which are themselves implicated in cancer pathogenesis. Complex GSLs might also be involved in cancer since they are shed from a number of

cancer cells and ganglioside depletion has been shown to reduce tumorigenicity. Together, these discoveries are paving the way for the advancement of anticancer therapies involving manipulation of SL signaling pathways and the formation of complex GSLs.

SLs and GSLs have also been implicated in the pathology of other diseases. Thus, anti-GSL antibodies are found in some peripheral neuropathies. Sulfatide has been implicated in thrombogenesis and hemostasis. Galactosylceramide and sulfatide are essential for node of Ranvier formation, involved in myelin formation. Other examples abound and based on all of these, SLs have also received attention as possible therapeutics. One of the most exciting is the possible use of ceramide-coated balloon catheters to target restenosis after angioplasty and/or intra-arterial stenting, which significantly reduces neointimal hyperplasia. Thus, SLs may eventually find their place as important therapeutic agents.

6 Perspectives

In this chapter, I have tried to share some of the recent excitement in SL and GSL research, and to show how these important lipids are impacting not only basic research but also modern medicine. Assuming that most structures of SLs and GSLs have now been discovered, and that the biochemical pathways have been delineated, the next generation of researchers in this field will no doubt devote most of its time to working out how these pathways are regulated, and whether the multiple different structures have defined functions in specific aspects of cell physiology and pathophysiology, or whether there is significant functional and structural redundancy.

See also Cellular Interactions; Glycobiology; Membrane Traffic: Vesicle Budding and Fusion; Membrane Transport.

Bibliography

Books and Reviews

Futerman, A.H. (Ed.) (2003) *Ceramide Signaling*, Kluwer Academic/Plenum Publishers, New York, USA.

Futerman, A.H., van Meer, G. (2004) The cell biology of lysosomal storage disease, *Nat. Rev. Mol. Cell Biol.* **5**, 554–565.

Gulbins, E., Kolesnick, R. (2003) Raft ceramide in molecular medicine, *Oncogene* **22**, 7070–7077.

Hakomori, S. (2003) Structure, organization, and function of glycosphingolipids in membrane, *Curr. Opin. Hematol.* **10**, 16–24.

Kester, M., Kolesnick, R. (2003) Sphingolipids as therapeutics, *Pharmacol. Res.* **47**, 365–371.

Kolter, T., Proia, R.L., Sandhoff, K. (2002) Combinatorial ganglioside biosynthesis, *J. Biol. Chem.* **277**, 25859–25862.

Marks, D.L., Pagano, R.E. (2002) Endocytosis and sorting of glycosphingolipids in sphingolipid storage disease, *Trends Cell. Biol.* **12**, 605–613.

Merrill, A.H. Jr. (2002) De novo sphingolipid biosynthesis. A necessary, but dangerous, pathway, *J. Biol. Chem.* **277**, 25843–25846.

Ogretmen, B., Hannun, Y.A. (2004) Biologically active sphingolipids in cancer pathogenesis and treatment, *Nat. Rev. Cancer* **4**, 604–616.

Spiegel, S., Milstien, S. (2003) Sphingosine-1-phosphate: an enigmatic signaling lipid, *Nat. Rev. Mol. Cell Biol.* **4**, 397–407.

Primary Literature

Alvarez-Vasquez, F., Sims, K.J., Hannun, Y.A., Voit, E.O. (2004) Integration of kinetic

information on yeast sphingolipid metabolism in dynamical pathway models, *J. Theor. Biol.* **226**, 265–291.

Bourbon, N.A., Yun, J., Kester, M. (2000) Ceramide directly activates protein kinase C zeta to regulate a stress- activated protein kinase signaling complex, *J. Biol. Chem.* **275**, 35617–35623.

Brown, D., London, E. (2000) Structure and function of sphingolipid- and cholesterol-rich membrane rafts, *J. Biol. Chem.* **275**, 17221–17224.

Brown, D.A., Rose, J.K. (1992) Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface, *Cell* **68**, 533–544.

Brugger, B., Graham, C., Leibrecht, I., Mombelli, E., Jen, A., Wieland, F., Morris, R. (2004) The membrane domains occupied by glycosylphosphatidylinositol-anchored prion protein and Thy-1 differ in lipid composition, *J. Biol. Chem.* **279**, 7530–7536.

Cremesti, A., Paris, F., Grassme, H., Holler, N., Tschopp, J., Fuks, Z., Gulbins, E., Kolesnick, R. (2001) Ceramide enables Fas to cap and kill, *J. Biol. Chem.* **276**, 23954–23961.

Cutler, R.G., Kelly, J., Storie, K., Pedersen, W.A., Tammara, A., Hatanpaa, K., Troncoso, J.C., Mattson, M.P. (2004) Involvement of oxidative stress-induced abnormalities in ceramide and cholesterol metabolism in brain aging and Alzheimer's disease, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2070–2075.

Dickson, R.C., Lester, R.L. (2002) Sphingolipid functions in Saccharomyces cerevisiae, *Biochim. Biophys. Acta* **1583**, 13–25.

Dupree, J.L., Coetzee, T., Blight, A., Suzuki, K., Popko, B. (1998) Myelin galactolipids are essential for proper node of Ranvier formation in the CNS, *J. Neurosci.* **18**, 1642–1649.

Fukasawa, M., Nishijima, M., Hanada, K. (1999) Genetic evidence for ATP-dependent endoplasmic reticulum-to-Golgi apparatus trafficking of ceramide for sphingomyelin synthesis in Chinese hamster ovary cells, *J. Cell. Biol.* **144**, 673–685.

Fukushima, N., Ishii, I., Contos, J.J., Weiner, J.A., Chun, J. (2001) Lysophospholipid receptors, *Annu. Rev. Pharmacol. Toxicol.* **41**, 507–534.

Furukawa, K. (2004) Glycosphingolipids in engineered mice: insights into function and generation of disease models, *Sem. Cell Dev. Biol.* **15**, 389–396.

Futerman, A.H., Hannun, Y.A. (2004) The complex life of simple sphingolipids, *EMBO Rep.* **5**, 777–782.

Garcia-Barros, M., Paris, F., Cordon-Cardo, C., Lyden, D., Rafii, S., Haimovitz-Friedman, A., Fuks, Z., Kolesnick, R. (2003) Tumor response to radiotherapy regulated by endothelial cell apoptosis, *Science* **300**, 1155–1159.

Garcia-Ruiz, C., Colell, A., Morales, A., Calvo, M., Enrich, C., Fernandez-Checa, J.C. (2002) Trafficking of ganglioside GD3 to mitochondria by tumor necrosis factor-alpha, *J. Biol. Chem.* **277**, 36443–36448.

Guerrera, M., Ladisch, S. (2003) N-butyldeoxynojirimycin inhibits murine melanoma cell ganglioside metabolism and delays tumor onset, *Cancer Lett.* **201**, 31–40.

Hakomori, S. (1984) Glycosphingolipids as differentiation-dependent, tumor-associated markers and as regulators of cell proliferation, *Trends Biochem.* **9**, 453–459.

Hakomori, S. (1990) Bifunctional role of glycosphingolipids. Modulators for transmembrane signaling and mediators for cellular interactions, *J. Biol. Chem.* **265**, 18713–18716.

Hanada, K., Kumagai, K., Yasuda, S., Miura, Y., Kawano, M., Fukasawa, M., Nishijima, M. (2003) Molecular machinery for nonvesicular trafficking of ceramide, *Nature* **426**, 803–809.

Hannun, Y.A., Obeid, L.M. (2002) The ceramide-centric universe of lipid-mediated cell regulation: Stress encounters of the lipid kind, *J. Biol. Chem.* **277**, 25847–25850.

Hannun, Y.A., Loomis, C.R., Merrill, A.H.J., Bell, R.M. (1986) Sphingosine inhibition of protein kinase C activity and of phorbol dibutyrate binding in vitro and in human platelets, *J. Biol. Chem.* **261**, 12604–12609.

Hayakawa, Y., Godfrey, D.I., Smyth, M.J. (2004) Alpha-galactosylceramide: potential immunomodulatory activity and future application, *Curr. Med. Chem.* **11**, 241–252.

Hopwood, J., Brooks, D.A. (1997) An Introduction to the Basic Science and Biology of the Lysosome and Storage Diseases. In Applegarth, D.A., Dimmick, J.E., Hall, J. (Eds.) *Organelle Disease*, Chapman and Hall Medical, London, pp. 7–36.

Huitema, K., Van Den Dikkenberg, J., Brouwers, J.F., Holthuis, J.C. (2004) Identification of a family of animal sphingomyelin synthases, *EMBO J.* **23**, 33–44.

Kolesnick, R. (2002) The therapeutic potential of modulating the ceramide/sphingomyelin pathway, *J. Clin. Invest.* **110**, 3–8.

Kolesnick, R.N., Clegg, S. (1988) 1,2-Diacylglycerols, but not phorbol esters, activate a potential inhibitory pathway for protein kinase C in GH3 pituitary cells, *J. Biol. Chem.* **263**, 6534–6537.

Kyogashima, M. (2004) The role of sulfatide in thrombogenesis and haemostasis, *Arch. Biochem. Biophys.* **426**, 157–162.

Lencer, W.I., Hirst, T.R., Holmes, R.K. (1999) Membrane traffic and the cellular uptake of cholera toxin, *Biochim. Biophys. Acta* **1450**, 177–190.

Lipsky, N.G., Pagano, R.E. (1985) A vital stain for the Golgi, *Science* **228**, 745–747.

Maceyka, M., Payne, S.G., Milstien, S., Spiegel, S. (2002) Sphingosine kinase, sphingosine-1-phosphate, and apoptosis, *Biochim. Biophys. Acta* **1585**, 193–201.

Malinina, L., Malakhova, M.L., Teplov, A., Brown, R.E., Patel, D.J. (2004) Structural basis for glycosphingolipid transfer specificity, *Nature* **430**, 1048–1053.

Malisan, F., Testi, R. (2002) GD3 ganglioside and apoptosis, *Biochim. Biophys. Acta* **1585**, 179–187.

Marchesini, N., Osta, W., Bielawski, J., Luberto, C., Obeid, L.M., Hannun, Y.A. (2004) Role for mammalian neutral sphingomyelinase 2 in confluence-induced growth arrest of MCF7 cells, *J. Biol. Chem.* **279**, 25101–25111.

Mizushina, Y., Hanashima, L., Yamaguchi, T., Takemura, M., Sugawara, F., Saneyoshi, M., Matsukage, A., Yoshida, S., Sakaguchi, K. (1998) A mushroom fruiting body-inducing substance inhibits activities of replicative DNA polymerases, *Biochem. Biophys. Res. Commun.* **249**, 17–22.

Mukherjee, S., Soe, T.T., Maxfield, F.R. (1999) Endocytic sorting of lipid analogues differing solely in the chemistry of their hydrophobic tails, *J. Cell Biol.* **144**, 1271–1284.

Munro, S. (2003) Lipid rafts: elusive or illusive? *Cell* **115**, 377–388.

Myerowitz, R., Lawson, D., Mizukami, H., Mi, Y., Tifft, C.J., Proia, R.L. (2002) Molecular pathophysiology in Tay-Sachs and Sandhoff diseases as revealed by gene expression profiling, *Hum. Mol. Genet.* **11**, 1343–1350.

Ng, C.K., Carr, K., McAinsh, M.R., Powell, B., Hetherington, A.M. (2001) Drought-induced guard cell signal transduction involves sphingosine-1-phosphate, *Nature* **410**, 596–599.

Paparounas, K. (2004) Anti-GQ1b ganglioside antibody in peripheral nervous system disorders: pathophysiologic role and clinical relevance, *Arch. Neurol.* **61**, 1013–1016.

Paris, F., Fuks, Z., Kang, A., Capodieci, P., Juan, G., Ehleiter, D., Haimovitz-Friedman, A., Cordon-Cardo, C., Kolesnick, R. (2001) Endothelial apoptosis as the primary lesion initiating intestinal radiation damage in mice, *Science* **293**, 293–297.

Raggers, R.J., Pomorski, T., Holthuis, J.C.M., Klin, N., van Meer, G. (2000) Lipid traffic: The ABC of transbilayer movement, *Traffic* **1**, 226–234.

Riboni, L., Campanella, R., Bassi, R., Villani, R., Gaini, S.M., Martinelli-Boneschi, F., Viani, P., Tettamanti, G. (2002) Ceramide levels are inversely associated with malignant progression of human glial tumors, *Glia* **39**, 105–113.

Riebeling, C., Allegood, J.C., Wang, E., Merrill, A.H. Jr., Futerman, A.H. (2003) Two mammalian longevity assurance gene (LAG1) family members, trh1 and trh4, regulate dihydroceramide synthesis using different fatty acyl-CoA donors, *J. Biol. Chem.* **278**, 43452–43459.

Rodrigues, M.L., Travassos, L.R., Miranda, K.R., Franzen, A.J., Rozental, S., de Souza, W., Alviano, C.S., Barreto-Bergter, E. (2000) Human antibodies against a purified glucosylceramide from Cryptococcus neoformans inhibit cell budding and fungal growth, *Infect. Immun.* **68**, 7049–7060.

Rohrbough, J., Rushton, E., Palanker, L., Woodruff, E., Matthies, H.J., Acharya, U., Acharya, J.K., Broadie, K. (2004) Ceramidase regulates synaptic vesicle exocytosis and trafficking, *J. Neurosci.* **24**, 7789–7803.

Ruvolo, P.P., Deng, X., Ito, T., Carr, B.K., May, W.S. (1999) Ceramide induces Bcl2 dephosphorylation via a mechanism involving mitochondrial PP2A, *J. Biol. Chem.* **274**, 20296–20300.

Sandhoff, K., Kolter, T. (1996) Topology of glycosphingolipid degradation, *Trends Cell. Biol.* **6**, 98–103.

Sango, K., Yamanaka, S., Hoffmann, A., Okuda, Y., Grinberg, A., Westphal, H., McDonald, M.P., Crawley, J.N., Sandhoff, K., Suzuki, K., Proia, R.L. (1995) Mouse models of Tay-Sachs and Sandhoff diseases differ in neurologic phenotype and ganglioside metabolism, *Nat. Genet.* **11**, 170–176.

Santana, P., Pena, L.A., Haimovitz-Friedman, A., Martin, S., Green, D., McLoughlin, M., Cordon-Cardo, C., Schuchman, E.H., Fuks, Z., Kolesnick, R. (1996) Acid sphingomyelinase-deficient human lymphoblasts and mice are defective in radiation-induced apoptosis, *Cell* **86**, 189–199.

Sheikh, K.A., Sun, J., Liu, Y., Kawai, H., Crawford, T.O., Proia, R.L., Griffin, J.W., Schnaar, R.L. (1999) Mice lacking complex gangliosides develop Wallerian degeneration and myelination defects, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 7532–7537.

Sidobre, S., Hammond, K.J., Benazet-Sidobre, L., Maltsev, S.D., Richardson, S.K., Ndonye, R.M., Howell, A.R., Sakai, T., Besra, G.S., Porcelli, S.A., Kronenberg, M. (2004) The T cell antigen receptor expressed by Valpha14i NKT cells has a unique mode of glycosphingolipid antigen recognition, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12254–12259.

Sillence, D.J., Platt, F.M. (2003) Storage diseases: new insights into sphingolipid functions, *Trends Cell. Biol.* **13**, 195–203.

Simon, C.G.J., Holloway, P.W., Gear, A.R.L. (1999) Exchange of C16-ceramide between phospholipid vesicles, *Biochemistry* **38**, 14676–14682.

Smith, D.C., Lord, J.M., Roberts, L.M., Johannes, L. (2004) Glycosphingolipids as toxin receptors, *Semin. Cell Dev. Biol.* **15**, 397–408.

Sprong, H., Degroote, S., Claessens, T., van Drunen, J., Oorschot, V., Westerink, B.H., Hirabayashi, Y., Klumperman, J., van der Sluijs, P., van Meer, G. (2001) Glycosphingolipids are required for sorting melanosomal proteins in the Golgi complex, *J. Cell Biol.* **155**, 369–380.

Suzuki, K. (1998) Twenty-five years of the 'psychosine hypothesis': A personal perspective of its history and present status, *Neurochem. Res.* **23**, 251–259.

Svennerholm, L. (1963) Chromatographic separation of human brain gangliosides, *J. Neurochem.* **10**, 613–623.

Thudicum, J. (1884) *A Treatise on the Chemical Constitution of the Brain*, Baillière, Tindall and Cox, London, UK.

Toledo, M.S., Suzuki, E., Levery, S.B., Straus, A.H., Takahashi, H.K. (2001) Characterization of monoclonal antibody MEST-2 specific to glucosylceramide of fungi and plants, *Glycobiology* **11**, 105–112.

Tomassini, B., Malisan, F., Franchi, L., Nicolo, C., Brea-Calvo, G., Saito, T., Testi, R. (2004) Calnexin suppresses GD3 synthase-induced apoptosis, *FASEB J.* **18**, 1553–1555.

van Meer, G., Lisman, Q. (2002) Sphingolipid transport: rafts and translocators, *J. Biol. Chem.* **277**, 25855–25858.

van Meer, G., Gumbiner, B., Simons, K. (1986) The tight junction does not allow lipid molecules to diffuse from one epithelial cell to the next, *Nature* **322**, 639–641.

van Meer, G., Stelzer, E.H.K., Wijnaendts-van-Resandt, R.W., Simons, K. (1987) Sorting of sphingolipids in epithelial (Madine-Darby Canine Kidney) Cells, *J. Cell Biol.* **105**, 1623–1635.

Venkataraman, K., Futerman, A. (2002) Do longevity assurance genes containing Hox domains regulate cell development via ceramide synthesis? *FEBS Lett.* **528**, 3–4.

Wada, R., Tifft, C.J., Proia, R.L. (2000) Microglial activation precedes acute neurodegeneration in Sandhoff disease and is suppressed by bone marrow transplantation, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 10954–10959.

Walkley, S.U. (2004) Secondary accumulation of gangliosides in lysosomal storage disorders, *Semin. Cell Dev. Biol.* **15**, 433–444.

Warnecke, D., Heinz, E. (2003) Recently discovered functions of glucosylceramides in plants and fungi, *Cell. Mol. Life Sci.* **60**, 919–941.

Winau, F., Schwierzeck, V., Hurwitz, R., Remmel, N., Sieling, P.A., Modlin, R.L., Porcelli, S.A., Brinkmann, V., Sugita, M., Sandhoff, K., Kaufmann, S.H., Schaible, U.E. (2004) Saposin C is required for lipid presentation by human CD1b, *Nat. Immunol.* **5**, 169–174.

Yamaoka, S., Miyaji, M., Kitano, T., Umehara, H., Okazaki, T. (2004) Expression cloning of a human cDNA restoring sphingomyelin synthesis and cell growth in sphingomyelin synthase-defective lymphoid cells, *J. Biol. Chem.* **279**, 18688–18693.

Yamashita, T., Wada, R., Sasaki, T., Deng, C., Bierfreund, U., Sandhoff, K., Proia, R.L. (1999) A vital role for glycosphingolipid synthesis during development and differentiation, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 9142–9147.

Yu, R.K., Saito, M. (1989) Structure and Localization of Gangliosides. In Margolis, R.U., Margolis, R.K. (Eds.) *Neurobiology of Glycoconjugates*, Plenum Press, New York, pp. 1–42.

Zervas, M., Somers, K.L., Thrall, M.A., Walkley, S.U. (2001) Critical role for glycosphingolipids in Niemann-Pick disease type C, *Curr. Biol.* **11**, 1283–1287.

Stem Cell Populations, Regulation of: *see* Regulation of Stem Cell Populations

Stem Cells (Human Embryonic) and Cancer

Maty Tzukerman and Karl Skorecki
Rambam Medical Center and Rappaport Faculty of Medicine and Research Institute, Technion-Israel Institute of Technology, Haifa, Israel

1 Introduction 403

2 Methods 403

3 Results 404

4 Discussion 404

5 Summary 406

Bibliography 406
Books and Reviews 406
Primary Literature 406

Keywords

Angiogenesis
The formation of new blood vessels from preexisting blood vessels.

Human embryonic stem cells
Pluripotent embryonic stem (ES) cell lines derived from the inner cell mass of human blastocysts that have the developmental potential to form derivatives of all three embryonic germ layers even after prolonged culture.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

Immunodeficient SCID/beige mouse
Severe combined immunodeficient (SCID) mice have a defect in the immune system that results in a lack of mature B and T cells, and thus are able to accept xenografts of human origin.

Teratoma
A nodule comprised of a wide variety of differentiated tissues and structures of normal human cells derived from injection of human embryonic stem cells into immunodeficient SCID/beige mouse.

Tumorigenesis
The process of tumor formation.

Tumor microenvironment
The complex network of intercellular interactions mediated by physical attachment, as in direct cell–cell or cell–stromal interactions, which has a significant effect on tumor physiologic characteristics such as growth pattern, vascular function, angiogenesis, and metastasis.

Vasculogenesis
The formation of new blood vessels that originate from endothelial cell precursors (angioblasts) during development.

There is currently no experimental system available wherein human cancer cells can be grown in the context of a mixed population of normal differentiated human cells for testing biological aspects of cancer cell growth (tumor cell invasion, angiogenesis), or response to anticancer therapies. Human embryonic stem cells when implanted into immunocompromised mice develop teratomas containing complex structures, comprising differentiated cell types representing the major germ-line derived lineages. We sought to determine whether human cancer cells would grow within such teratomas and display properties associated with malignancy, such as invasiveness and recruitment of blood vessels. Ovarian cancer cells (HEY), stably expressing an H2A-GFP fusion protein, which allows tracking of tumor cells, were injected into mature teratomas and developed into tumors. The growth, proliferative capacity, invasion, and induction of blood vessel formation were examined. We propose using this novel experimental platform to develop a preclinical model for investigating and manipulating the stromal response in tumor cell growth as an additional tool in cancer research.

1 Introduction

Human embryonic stem cells (hESC) have the potential to differentiate into complex structures, comprising differentiated cell types representing derivatives of all three major embryonic lineages. Embryoid bodies generated *in vitro* from human embryonic stem cells grown in suspension, and teratomas generated *in vivo* following injection into immunocompromised mice, contain a wide variety of differentiated tissues and structures comprised of normal human cells. Using these human differentiated cellular microenvironments, we have been able to demonstrate and investigate the growth and tumorigenic properties of a variety of human cancer cell types.

In recent years, there has been growing awareness that tumorigenesis properties are markedly affected by the surrounding tissue microenvironment at both primary and metastatic sites of tumor growth. Thus, it has been shown that tumor progression is associated with extensive remodeling of adjacent tissues to provide a supportive microenvironment for cancer cell proliferation, migration, invasion, and the formation of blood vessels required for supporting cancer growth. As an example, human prostate carcinoma-associated fibroblasts can promote tumorigenic transformation in initiated human prostate epithelial cells. In more recent studies, it was shown that the connective tissue stroma surrounding neoplastic breast tissue was induced into a state that promoted subsequent neoplastic growth in naïve breast epithelium. Also, in recent studies it was demonstrated that the genomic rearrangements characteristic of a given tumor type (e.g. lymphoma), could also be demonstrated in the adjacent endothelial cells of tumor-induced blood vessels. Appreciation of the important role stromal response plays in supporting tumor growth has led to the development of novel anticancer therapeutic agents targeted to disrupt stromal response factors, and thereby inhibit tumor growth. Particular attention has been directed at the development of anti-angiogenic molecules in particular, some of which have reached clinical trials.

Purely *in vitro* models using malignant cells in continuous culture, such as anchorage-independent growth and focus formation are not particularly amenable to the investigation of stromal interactions. It was the emergence of *in vivo* models using tumor xenograft growth in immunocompromised mice, which highlighted the importance of the stromal response in an experimental setting and have led to corresponding studies in the clinical setting. Accordingly, most preclinical studies of anticancer reagents have been conducted using injection of tumor cells as xenografts in immunocompromised mice. However, in the numerous published studies using this model, it has been the murine rather than the human stromal response, which has been the target of investigation or experimental therapeutic intervention.

Therefore, we utilized human embryonic stem cells in order to study the interactions of tumor cells with the surrounding heterogeneous microenvironment of differentiated human cell tissues and structures.

2 Methods

HEY ovarian cancer cells were injected into human embryonic stem cell–derived

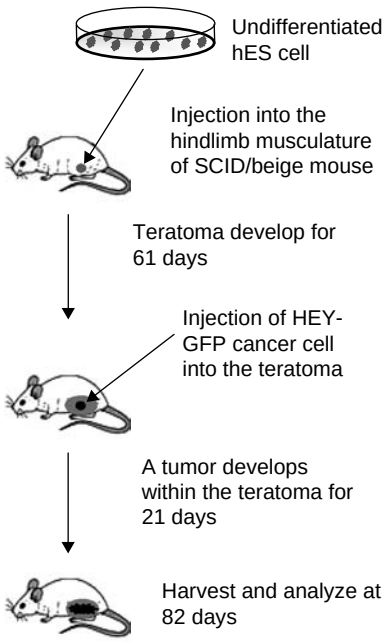

Fig. 1 Schematic representation of the experimental protocol. Copyright (2003) National Academy of Sciences, USA.

teratomas that had been allowed to mature in SCID/beige mice (Fig. 1). The HEY ovarian cancer cells had been transfected so as to stably express a histone H_2A-GFP fusion protein, in order to facilitate the tracking of tumor cell growth and invasion. Following various periods of growth, immunohistochemical analysis of the mixed teratoma-tumor structures confirmed the feasibility of using this approach to examine human tumorigenesis properties in a human cellular microenvironment.

3 Results

The important findings using this novel approach included: (1) the development of a homogeneous mass of tumor cells with high proliferative capacity within the differentiated microenvironment of the teratoma; (2) the identification of invasion by tumor cells into the surrounding differentiated teratoma structures; and (3) the existence of blood vessels of human origin derived from hES differentiation. In this regard, we have demonstrated staining of such blood vessels by using antibodies to human-specific endothelial cell surface antigenic epitopes such as CD31 or CD34 in teratomas bearing tumors, and demonstrated the presence of blood vessels of hES origin, of various calibers and levels of maturity, growing adjacent to and within the tumor. Therefore, it will be important to rigorously quantitate differences in blood vessel development between teratomas without and with tumor growth using these immunoreagents. This immunostaining approach can be supplemented with quantitative molecular analysis of the known neovasculogenesis gene expression profile.

4 Discussion

This novel experimental approach using human stem cell technology has been extended to generate a novel experimental platform, in which a wide variety of human tumor cells can be grown in a preclinical setting in a tissue microenvironment consisting of differentiated human cells, for investigating cancer growth and anticancer therapy. We have examined and compared different multiple human tumor cell types with varying tumorigenesis properties, and correspondingly found differences in growth, invasion, and angiogenic responses using this new model.

Most significantly, we find that the heterogeneous hES derived microenvironment provides a supportive niche for tumor growth, which is not available to human tumors growing xenografts following direct injection in murine tissues. We postulated that this reflects the availability to the tumor of the variety of potential human supporting cells that the corresponding tumor would actually encounter in the clinical setting. This postulate is supported and strengthened by the finding of significant differences in the response to anticancer therapy in tumors growing in the human cell microenvironment of a teratoma in immunocompromised mice in comparison with the same tumors growing following direct injection into the hind limb muscle tissue of these mice. It remains to be seen whether such differences may be predictive of expected responses in clinical trials. It will also be possible to extend the applications of this model in a number of basic and applied directions. Among others these extensions include:

1. pharmacologic and/or genetic manipulation of either the tumor cells or the hES derived stromal cells, which could include stable transfection of the hES cells with constructs targeted to modify the levels of expression of key enzymes and growth factors involved in the stromal response to tumor growth and progression;
2. molecular analysis of hES derived surrounding stromal cells, isolated from the tumor growths; isolation of stromal cells of different histologic specificity

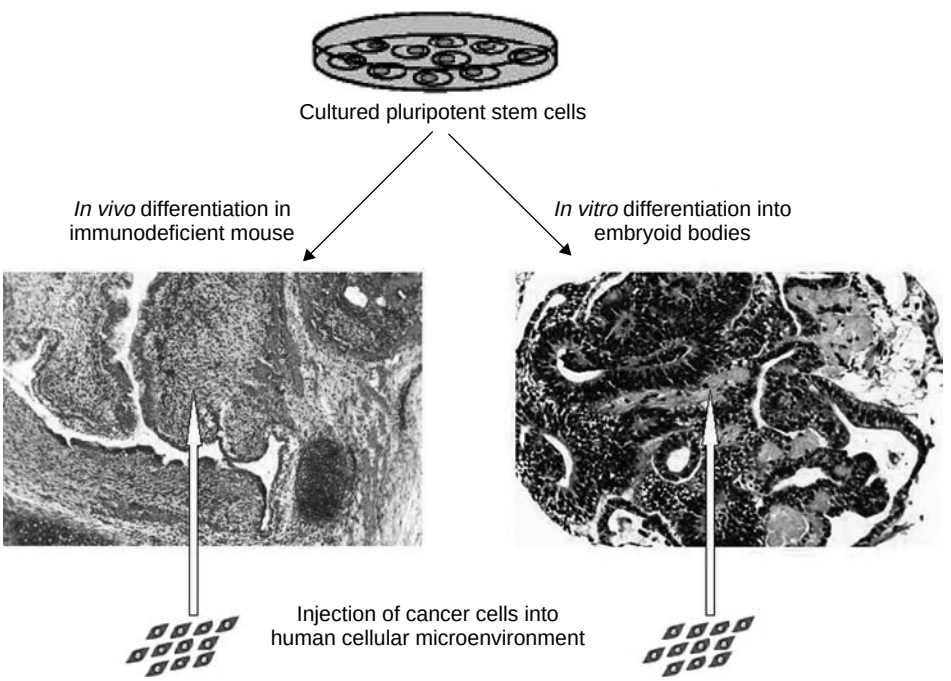

Fig. 2 An experimental platform for growing human tumor cells within a human cellular microenvironment derived from differentiation of human embryonic stem cells *in vitro* and *in vivo*. Copyright (2003) Discovery Medicine.

can be achieved by laser microdissection, and can be subjected to gene or protein expression analysis on relevant microchip arrays;
3. vital and functional monitoring of the tumor growth in response to therapeutic intervention, using modern imaging techniques applied to appropriately labeled or marked hES or tumor cells, including hybrid imaging with micro-PET/CT, or with luminescence or fluorescence detection.

It should be recognized that in the case of bulk teratomas and embryoid bodies, these structures are comprised of a mixture of cells, and lack the homogeneity or organization present in authentic human tissues. To this end, enrichment protocols and 3D polymer-based *in vitro* models have already been applied to human embryonic stem cell growth and will likely represent an important future step in optimizing the utility of the experimental platform that we have described. This model can be extended to the growth of tumor cells in a purely *in vitro* embryoid body configuration (Fig. 2).

5 Summary

In summary, we present and propose using a novel experimental platform consisting of human tumor cells growing within a human cellular microenvironment derived from human embryonic stem cells. The growth, invasion, and human neoangiogenic response observed indicate that these structures may serve as an important preclinical model for investigating and manipulating the stromal response in tumor cell growth, and thereby contribute to the progress of cancer research.

See also Brain Development; Cancer Stem Cells; Cellular Interactions; Mammalian Cell Culture Methods; Nuclear Transfer for Cloning Animals; Parthenogenetic Stem Cells; Principles and Applications of Embryogenomics; Stem Cells.

Bibliography

Books and Reviews

Alt-Holland, A., Zhang, W., Margulis, A., Garlick, J.A. (2005) Microenvironmental control of premalignant disease: the role of intercellular adhesion in the progression of squamous cell carcinoma, *Semin. Cancer Biol.* **15**, 84–96.

Bissell, M.J., Radisky, D. (2001) Putting tumours in context, *Nat. Rev. Cancer* **1**, 46–54.

Condon, M.S. (2005) The role of the stromal microenvironment in prostate cancer, *Semin. Cancer Biol.* **15**, 132–137.

Folkman, J. (2002) Role of angiogenesis in tumor growth and metastasis, *Semin. Oncol.* **6**(Suppl. 16), 15–18.

Pera, M.F., Reubinoff, B., Trounson, A. (2000) Human embryonic stem cells, *J. Cell. Sci.* **113**, 5–10.

Rangarajan, A., Weinberg, R.A. (2003) Opinion: comparative biology of mouse versus human cells: modelling human cancer in mice, *Nat. Rev. Cancer* **3**(12), 952–959.

Primary Literature

Folkman, J. (2004) A novel anti-vascular therapy for cancer, *Cancer Biol. Ther.* **3**, 338–339.

George, D.J., Kaelin, W.G. Jr. (2003) The von Hippel-Lindau protein, vascular endothelial growth factor, and kidney cancer, *N. Engl. J. Med.* **349**, 419–421.

Gerecht-Nir, S., Osenberg, S., Nevo, O., Ziskind, A., Coleman, R., Itskovitz-Eldor, J. (2004) Vascular development in early human embryos and in teratoma-derived from human embryonic stem cells, *Biol. Reprod.* **71**, 2029–2036.

Gertow, K., Wolbank, S., Rozell, B., Sugars, R., Andang, M., Parish, C.L., Imreh, M.P., Wendel, M., Ahrlund-Richter, L. (2004) Organized development from human embryonic stem cells after injection into immunodeficient mice, *Stem Cells Dev.* **13**, 421–435.

Goldshmidt, O., Zcharia, E., Abramovitch, R., Metzger, S., Aingorn, H., Friedmann, Y., Schirrmacher, V., Mitrani, E., Vlodavsky, I. (2002) Cell surface expression and secretion of heparanase markedly promote tumor angiogenesis and metastasis, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10031–10036.

Huang, J., Huang, J., Frischer, J.S., Serur, A., Kadenhe, A., Yokoi, A., McCrudden, K.W., New, T., O'Toole, K., Zabski, S., Rudge, J.S., Holash, J., Yancopoulos, G.D., Yamashiro, D.J., Kandel, J.J. (2003) Regression of established tumors and metastases by potent vascular endothelial growth factor blockade, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7785–7790.

Kuperwasser, C., Chavarria, T., Wu, M., Magrane, G., Gray, J.W., Carey, L., Richardson, A., Weinberg, R.A. (2004) Reconstruction of functionally normal and malignant human breast tissues in mice, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4966–4971.

Olumi, A.F., Grossfeld, G.D., Hayward, S.W., Carroll, P.R., Tlsty, T.D., Cunha, G.R. (1999) Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium, *Cancer Res.* **59**, 5002–5011.

Streubel, B., Chott, A., Huber, D., Exner, M., Jager, U., Wagner, O., Schwarzinger, I. (2004) Lymphoma-specific genetic aberrations in microvascular endothelial cells in B-cell lymphomas, *N. Engl. J. Med.* **351**, 250–259.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., Jones, J.M. (1998) Embryonic stem cell lines derived from human blastocysts, *Science* **282**, 1145–1147.

Tzukerman, M., Skorecki, K. (2003) A novel experimental platform for investigating tumorigenesis and anti-cancer therapy in a human microenvironment derived from embryonic stem cells, *Discov Med* **3**, 51–54.

Tzukerman, M., Rosenberg, T., Ravel, Y., Reiter, I., Coleman, R., Skorecki, K. (2003) An experimental platform for studying growth and invasiveness of tumor cells within teratomas derived from human embryonic stem cells, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13507–13512.

Wylie, S., MacDonald, I.C., Varghese, H.J., Schmidt, E.E., Morris, V.L., Groom, A.C., Chambers, A.F. (1999) The matrix metalloproteinase inhibitor batimastat inhibits angiogenesis in liver metastases of B16F1 melanoma cells, *Clin. Exp. Metastasis* **17**, 111–117.

Stem Cells in the Adult Brain: Their Identification and Role in Neurogenesis

Arturo Alvarez-Buylla and Erica L. Jackson
University of California, San Francisco, CA

1 Introduction 411

2 Adult Neurogenesis 411

3 Identification of the Primary Precursors in the Adult Brain 415
3.1 A Ventricular Zone (VZ) in Adult Birds: Radial Glia as Neuronal Progenitors 415

4 Progenitors of New Neurons in the Adult Mammalian Brain 416
4.1 Stem Cells Persist in the Adult Mammalian Brain 416

5 Organization and Primary Precursors of New Neurons in Germinal Zones of the Adult Mammalian Brain 417
5.1 The Subventricular Zone (SVZ) 418
5.2 The Subgranular Zone (SGZ) 420

6 Organization and Primary Precursors of New Neurons in the Embryonic and Postnatal Mammalian Brain 423
6.1 Radial Glia are the Neural Progenitors of the Developing Brain 424

7 The Lineage of Neural Stem Cells; Evolving Concepts 425

8 Concluding Remarks 429

Acknowledgments 429

Bibliography 429
Books and Reviews 429
Primary Literature 429

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

Keywords

Astrocytes
The most numerous glial cells in the CNS, astrocytes were traditionally defined by their star shaped morphology and believed to serve as support cells for neurons. It is now known that some astrocytes function as the neural stem cells in the adult brain.

Neural Stem Cells
Cells with the capacity for long-term proliferation that are capable of giving rise to progeny that differentiate into neurons, astrocytes, and oligodendrocytes. These cells are also able to self-renew, meaning they can divide to produce more stem cells.

Neurogenesis
The production of new neurons and the process by which they are born, migrate, and mature.

Radial Glia
Cells with a long basal process that extends from their cell body in the ventricular zone, through the parenchyma toward the brain surface. Classically, radial glia were viewed as immature glial cells that guide neuronal migration and function as scaffolding for brain development, but more recently have been determined to have neural stem cell properties.

Subgranular Zone
The region at the border between the granule cell layer and the hilus in the dentate gyrus of the hippocampus. This region is a germinal center and site of ongoing neurogenesis in the adult brain.

Subventricular Zone
The region of the brain lining the walls of the lateral ventricles. This region is a major germinal center and site of ongoing neurogenesis in the adult brain.

The adult brain continues generating new neurons. This discovery has dramatically changed our view of brain plasticity, and has also revealed previously unsuspected mechanisms of neuronal birth, migration, and replacement. In this chapter, we focus on the mechanism of neuronal birth and discuss recent data on how these new findings have changed concepts on the origin of neurons and glial cells in the adult, and developing central nervous system (CNS). Adult neurogenesis has been demonstrated in the forebrain of songbirds, and in the hippocampus and olfactory bulb (OB) of adult mammals, suggesting that neural stem cell are retained in the adult brain. These primary precursors have now been identified and shown to reside in specialized germinal niches. The composition and organization of these niches is discussed. Surprisingly, the primary precursor of new neurons in the adult brain has been found to be radial glial cells (birds) and astrocytes (mammals), glial cells

long considered to be part of a lineage separate from that of neurons. Work in the developing mammalian brain has shown that radial glia are not only precursors of astrocytes, but also of neurons, oligodendrocytes, and ependymal cells. Radial glia also serve as the precursors of adult neural stem cells. It is becoming increasingly clear that neural stem cells are contained within the lineage of neuroepithelial cells-radial glia-astrocytes. The identification of primary precursors in the developing and adult brain is fundamental to our understanding of neural development and adult neural replacement.

1 Introduction

For almost a century, the brain was conceived to differ significantly from other tissues that constantly self-renew. In these self-renewing tissues, primary progenitors or stem cells divide to regenerate different cell types, thus constantly replaying developmental programs in the adult for the purpose of tissue maintenance. For example, the bone marrow contains hematopoietic stem cells that divide to generate secondary progenitors that differentiate into different types of blood cells. The brain was considered different. It was thought that the important functions carried out by the brain required a stable network of neurons and that these fundamental cellular elements were not added or replaced in adult life, but only formed during a restricted period of embryonic and fetal life. It was believed that the mature brain was not able to support the birth, migration, and incorporation of new neurons, much less that these cells could then be integrated into functioning brain circuits. The progenitors and scaffolding required to support neuronal production and incorporation were thought to be present in the brain during just a limited period of development.

This traditional view has changed. Whereas many neurons in the brain are very long lived and indeed not replaced throughout life, new neurons are recruited into selective brain circuits probably in all vertebrate species, including humans. This conceptual shift raised fundamental questions about the mechanism by which new neurons are added to the adult brain. Investigations on the mechanism of adult neurogenesis revealed that neural stem cells were retained in the adult brain. The identification of these primary progenitors (stem cells) yielded an unexpected observation: these cells had the characteristics of astrocytes, a type of glial cell (the supportive cells found around neurons) previously thought to be fully differentiated and completely separate from the neuronal lineages. Here we will review how this field has evolved and how these discoveries, together with the recent identification of the stem cell in the developing brain, radically change the concepts in developmental neuroscience.

2 Adult Neurogenesis

Neurons do not divide. No one has ever observed a fully mature neuron undergo mitosis. Mitosis in the adult brain is rare and when observed it is associated with small cells and with

cells of glial character. It is for this reason that it was classically thought that while gliogenesis continues in the adult brain, neurogenesis stops soon after birth. However, determining what a cell that has undergone mitosis ultimately becomes is not a simple problem. [^{3}H]thymidine, an analog of thymidine that gets incorporated into DNA during S phase of the cell cycle, was instrumental to our understanding of the fate of dividing cells and to the discovery of adult neurogenesis. Early studies suggested that [^{3}H]thymidine was incorporated into neurons of adult brains after an electrolytic lesion. It was suggested that multiple neuronal types, including cortical pyramidal neurons, may be formed in the adult brain, and that some of these cells appear labeled with [^{3}H]thymidine as soon as one day after injection. Some of these early claims are not supported by more recent work, as we now know that it takes much longer for new neurons to form in the adult brain. More recent work, however, has confirmed that new neurons are indeed added to very specific circuits of the adult brain.

It was in an adult songbird, the canary, that neurogenesis was first demonstrated using, in addition to anatomical methods, ultrastructure and electrophysiology. Canaries display dramatic hormone-induced changes in brain regions associated with song production. It was this remarkable plasticity that led Goldman and Nottebohm to use [^{3}H]thymidine to determine whether new cells were added to song control nuclei. It was found that cells that incorporated [^{3}H]thymidine had the characteristic morphology of neurons at the light microscope. These cells also had the ultrastructure of neurons including synapses, they displayed action potentials and synaptic inputs typical of neurons, and some of the new neurons grew axons, which are only observed on nerve cells and not on glial cells. Neuronal recruitment in adult songbirds was first described in the song control nuclei, but was then shown to occur throughout most of the telencephalon. The new neurons are born in the ventricular zone (VZ) on the walls of the lateral ventricles. Just as in development, the precursors for the new neurons in the adult are not located at the site of differentiation, but in specialized germinal niches sometimes millimeters away from their site of final differentiation. Radial glia persist in the ventricular wall of adult birds, and the long processes of these cells guide the migration of young neurons in the brain from their site of birth to their ultimate destination where differentiation occurs. Little is known about the microenvironment in the avian VZ that supports the proliferation of neural precursors. The cell adhesion molecule N-Cadherin is concentrated in this region and has been suggested to play a role in maintaining the progenitor cells within the germinal niche. These studies firmly established that new neurons are added to adult birds.

New glial cells are also formed in the brain of adult canaries. However, in contrast to neurons, these cells seem to be derived from cells that divide locally within the brain parenchyma. This inference comes from the observation that at short survivals after [^{3}H]thymidine administration, cells of glial morphology appear labeled throughout the brain parenchyma. Furthermore, cells that migrate away from the VZ seem to give rise only to neurons. Evidence *in vitro*, however, suggests that both neurons and glia may be generated from dividing VZ cells in the brain of adult canaries. This may suggest that cells derived from the adult VZ in birds have the potential to form both neurons and glia,

but that the environment within the adult avian brain is only conducive to the differentiation of these new cells into neurons. Alternatively, separate precursors for neurons and glial cells may exist in the VZ of adults, but only the progeny of neuronal precursors migrates away from the VZ and differentiate.

Neurogenesis also continues in several regions of the central nervous system (CNS) of other vertebrate species including adult fish, lizards, and amphibians. The body and brain of most poikilotherm vertebrates continues to grow throughout life. Unlike mammals or birds in which neurogenesis is thought of as a process of cell replacement, in poikilotherms it has been suggested that neurogenesis is associated with growth of the body. For example, the retina of adult fish and frog keeps growing and this growth results, at least in part, from the addition of new neurons, including ganglion cells. Neurogenesis in adult lizards, however, appears to also have a role in regeneration. The medial cortex in lizards, a region equivalent to part of the hippocampal formation in mammals, thought to be important for spatial and olfactory learning, receives many new neurons in adults. Most interestingly, following a neurotoxic lesion that destroys most of the neurons of the medial cortex in adult lizards, rapid regeneration ensues, resulting in remarkable histological restoration. The regions in the VZ of adult lizards that give rise to the new neurons are well defined, and this should allow for the identification of the proliferative neuronal precursors that make these regenerative processes possible.

In mammals, the study of adult neurogenesis has been highly contentious. After years of debate over the identity of [^{3}H]thymidine labeled cells and whether it occurs at all in adult primates, it has now been clearly established that neurogenesis occurs in two regions of the adult mammalian forebrain: the olfactory bulb (OB) in the anterior telencephalon and the dentate gyrus (DG) of the hippocampus. Neurons destined for the OB are born in a specialized germinal region called the *subventricular zone* (SVZ). The SVZ is on the walls of the lateral ventricles caudal to the olfactory bulb. New neurons destined for the OB have to migrate a long distance (in mice 3–8 mm) to get to their final site of differentiation. In contrast, those destined for the dentate gyrus of the hippocampus are born locally in the subgranular zone (SGZ) (see Fig. 1).

Neuronal recruitment has also been demonstrated in adult primates, including humans. Similar to rodents, cells in the adult primate hippocampus are born in the SGZ. Also in a similar manner to what has been described in adult rodents, new neurons born on the lateral wall of the lateral ventricles of adult monkeys migrate into the olfactory bulb where they differentiate into local interneurons. Interestingly, the migration from the SVZ to the olfactory bulb does not occur in adult humans, but has been inferred to take place in human embryos. This suggests that this migratory process stops sometime during development in the human brain. Gould and colleagues in 1999 proposed that limited neuronal recruitment may also occur in neocortex. However, other laboratories did not observe incorporation of new neurons in adult cortex.

In the case of the olfactory bulb and dentate gyrus, the recruitment of the new neurons in the adult has been firmly established. Analysis by electron microscopy has shown that newly formed cells in the dentate gyrus and the olfactory bulb of adult rats have the ultrastructure of neurons and receive synaptic inputs.

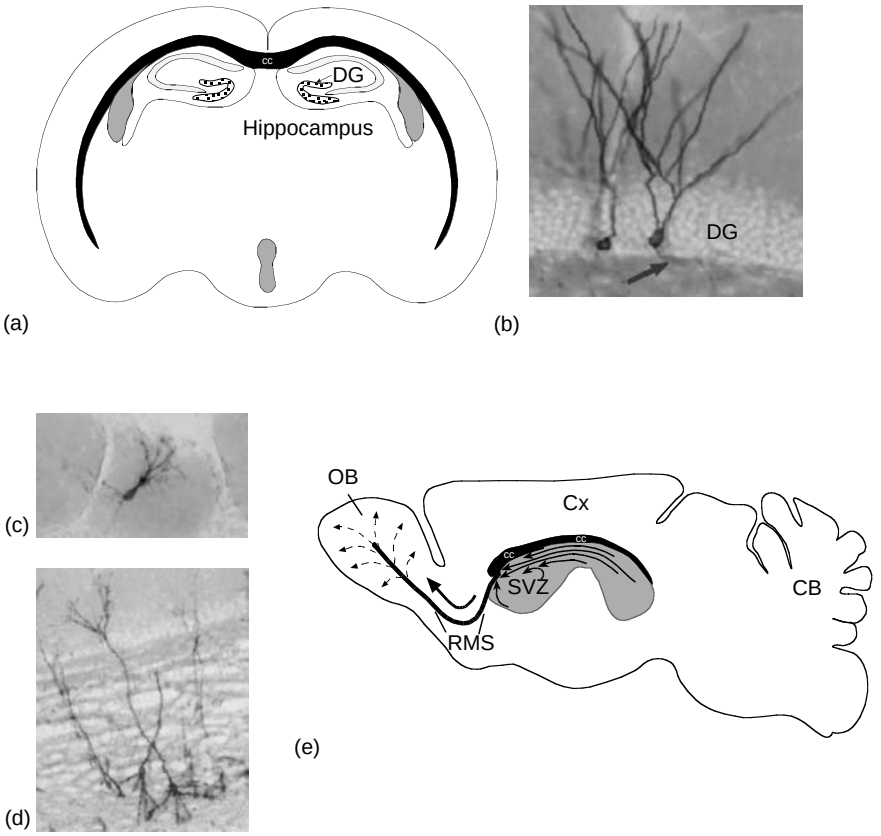

Fig. 1 Neurogenesis occurs in two regions of the adult brain. (a) Schematic cross-sectional view of the adult mouse brain at the level of the hippocampus. Retrovirally labeled progenitors in the subgranular layer of the dentate gyrus (DG) give rise to new neurons in the granule cell layer (b). Newly generated neurons, periglomerular cells (c) and granule cells (d) in the olfactory bulb are derived from labeled progenitors in the subventricular zone (SVZ). (e) The SVZ-OB system. Schematic sagittal view of the adult mouse brain with the OB to the left and cerebellum (CB) to the right. New neurons are produced throughout the SVZ (orange). The young neurons become aligned into long chains that form a series of interconnected paths throughout the SVZ. Once the chains reach the anterior SVZ, they join the rostral migratory stream (RMS), which leads the young neurons into the core of the olfactory bulb. Within the OB, cells disperse radially (dashed arrows) as individual cells and complete their differentiation into periglomerular cells (c) and granule cells (d). CX, Cortex; cc, corpus callosum. (See color plate p. xxxi).

In addition, newly formed cells in the dentate gyrus have axons that can be backfilled when a retrograde tracer is injected in CA3 of Ammon's horn. More recently, electrophysiological studies in slice cultures have demonstrated that new cells in the olfactory bulb and dentate gyrus display action potentials typical of neurons and are integrated into synaptic networks. These observations clearly establish the neuronal identity of new cells incorporated in the olfactory bulb and hippocampus.

The demonstration of adult neurogenesis in birds and mammals has changed the long-held dogma that no new neurons are formed in adults. These discoveries have raised the very basic question: which cells in the adult brain function as the primary progenitors or stem cells that make adult neurogenesis possible. The rest of this chapter will focus on the neural stem cells that are retained in the adult brain.

3 Identification of the Primary Precursors in the Adult Brain

During development neurons are born in specialized proliferative zones, generally in regions on the walls of the brain ventricle. As we have described earlier, this is also the case in adult birds where new neurons originate form progenitor cells in the wall of the ventricle. In adult mammals, cells born in the SVZ are also born close to the walls of the lateral ventricles, but neurons born in the SGZ are generated away from the ventricular system. In this section, we will discuss how these different germinal regions are organized and also discuss which cells are the stem cells for the continual generation of new neurons.

3.1 A Ventricular Zone (VZ) in Adult Birds: Radial Glia as Neuronal Progenitors

Classical studies have shown how radial glia serve as scaffolding for building the brain. As discussed below, until quite recently radial glia were considered committed progenitors of astrocytes, the most common type of glial cells in the brain; thus the name glia. Radial glia have their soma in the VZ, and possess a long process that extends toward the brain surface. These processes serve as guides for the migration of many young neurons from their site of birth in the VZ. Sometime after neuronal production ceases, radial glia in the mammalian brain migrate away from the VZ and transform into astrocytes. In contrast to mammals, radial glia persist in the adult brain of songbirds, as described earlier, and also in the adult brain of many species of cold-blooded vertebrates.

In adult canaries, new neurons are incorporated throughout most of the telencephalon, but not outside of it. This spatial restriction emerges during development around the time of hatching. As explained above, neurons are born in the walls of the lateral ventricles from which they migrate long distances to reach most areas of the telencephalon. The proliferative layer in the walls of the lateral ventricles in adult birds has many of the properties of the developing VZ. It is within this layer of cells that the primary precursor cells that support continuous adult neurogenesis reside. Proliferation occurs at higher rates in VZ "hot spots" where a large number of new neurons are born.

Young neurons born in these hot spots use the long processes of radial glia as guides to migrate away from the VZ. In this respect, the function of radial glia is similar to that described during the development of the cerebral cortex. However, a series of observations began to suggest that radial glia in adult birds may not only function as migratory guides but also as precursor cells. Neurogenic hot spots in birds are enriched for radial glia, and neurogenesis appears to be the primary function for these proliferative hot spots as no glial cells emerge from these regions. Surprisingly, radial glia divide, and their division correlates spatially and temporally with the appearance of new neurons in the proliferative hot spots.

This strongly suggests that radial cells in adult songbirds divide to generate the new neurons. Retroviral lineage studies in the adult avian telencephalon demonstrated radial glia in the same clones as that of new neurons, which was consistent with the model that radial glia gave rise to the new neurons. Radial glia in the adult avian brain also undergo interkinetic nuclear migration, a behavior typically associated with progenitor cells in the developing VZ. Adult avian radial glia also extend a single short cilium into the cerebrospinal fluid; primary cilia of similar structure and localization are found in neuroepithelial stem cells during embryonic development.

The above analyses strongly suggest that radial glial cells in the VZ of the adult avian brain function as the primary precursors of new neurons. The lineage and precise mode of division of radial glial in adult birds is not known. Nor is it known whether new neurons are produced directly from the division of a radial glia or whether the radial glia divide to generate intermediate progenitors that then generate the young neurons. The idea of radial glia as neuronal precursors was considered a peculiarity of adult birds, perhaps an adaptation for the continual generation of new neurons that is so widespread in the telencephalon of adult songbirds. In mammals, radial glia were considered to be in a lineage separate from that of neurons. However, as we will review below, this idea has begun to change.

4 Progenitors of New Neurons in the Adult Mammalian Brain

After more than a century of believing that the regenerative potential of the adult mammalian brain ended with development, the discovery of adult neurogenesis in restricted regions of the adult mammalian brain opened a whole new perspective. It suggested that neuronal progenitors and mechanisms to put these progenitors to work were in place at least in some regions of the adult brain. This raised hopes both for understanding a totally new mechanism of neural plasticity and for the possible use of this regenerative capacity for therapeutic purposes. One of the most basic questions was: which cells in the adult brain have this unique potential to give birth to the new neurons? It was well established that in mammals, unlike birds, radial glia are not retained in the adult brain, but transform into astrocytes at the end of fetal development as explained above. The origin of the new neurons in the SVZ and the SGZ, the two regions of the adult mammalian brain that continue generating new neurons in the adult, will be the theme discussed in this section.

4.1 Stem Cells Persist in the Adult Mammalian Brain

Before we understood the organization of adult germinal centers or the identity of the primary precursors of new neurons, studies of isolated cells *in vitro* made a very basic and important contribution. When cells from the adult brain were exposed to growth factors *in vitro* (in the initial studies on EGF or epidermal growth factor), a small subpopulation of the plated cells began to grow in suspension to form spherical clusters that were called *neurospheres*. Neurospheres have the interesting property of generating differentiated cells, including neurons, oligodendrocytes, and astrocytes when the growth factors are removed from the medium and the cells

are transferred onto a culture substratum that allows the cells to adhere to the tissue culture plate. Moreover, when primary neurospheres are dissociated and individual cells are placed again in nonadherent culture conditions with growth factors, a subpopulation of the cells in the neurosphere is able to generate secondary neurospheres that are also capable of generating differentiated progeny. Clonal cultures subsequently demonstrated that a single cell isolated from the adult brain can generate a neurosphere. Since self-renewal and multilineage differentiation are some of the basic defining properties of stem cells, these studies indicated that neural stem cells were retained in the adult brain. A similar conclusion was reached independently using adherent culture conditions with FGF (fibroblast growth factor) as the growth factor. FGF can also promote the growth of neural stem cells as neurospheres alone or in combination with EGF. Neurospheres have been isolated from other mammalian species including humans. It has been suggested that neurospheres could potentially provide a method for the amplification of neuronal precursors for therapeutic purposes. Neurospheres have become widely used in the neuroscience field to address questions on the regulation of self-renewal of neural stem cells, and to study factors that influence the differentiation of neural stem cells into specific types of neurons or glial cells.

Initial studies using EGF to make neurospheres used cells isolated from the entire striatum, which included parts of the walls of the lateral ventricle. Subsequent studies demonstrated that the self-renewing multipotent neurospheres were derived from a thin strip of the wall of the lateral ventricle. It is only on this wall that EGF-responsive cells capable of forming neurospheres are present. Interestingly, FGF-resposive neurospheres can be grown not only from the walls of the lateral ventricle but also from other ventricular walls, including the central canal in the spinal cord. FGF-responsive neural stem cells have also been isolated from regions of the adult brain, which apparently do not contain ventricular wall.

Interestingly, the populations of neural stem cells that respond to FGF or EGF are apparently different. There is some evidence suggesting that exposure to FGF induces the formation of EGF-responsive progenitors. This suggests that in the adult brain FGF may activate an earlier progenitor than EGF. The identification of the cells that respond to these growth factors *in vivo* is a topic of considerable interest and debate. We will next describe the organization of the SVZ and SGZ and describe their cellular composition. This will allow discussion of the possible cell types that give rise to neurospheres and provide an *in vivo* picture of how neurogenesis normally occurs in the adult brain.

5 Organization and Primary Precursors of New Neurons in Germinal Zones of the Adult Mammalian Brain

As explained above, neuronal birth in adult mammals has been demonstrated in two brain regions: (1) in the SVZ of the lateral ventricles and (2) in the subgranular layer (SGL) of the hippocampal dentate gyrus. In the adult mammalian brain, very few, if any, radial glial cells are retained into adult life, raising the question of the identity of the primary precursors for these new neurons.

5.1 The Subventricular Zone (SVZ)

The presence of large numbers of dividing cells in the SVZ has been known for more than a hundred years and is known to occur in many vertebrates including mice, rats, rabbits, hamster, dogs, monkeys, and humans. However, the fate of the large number of newly generated SVZ cells has been controversial. While pioneering work by Joseph Altman suggested that the SVZ could generate neurons, most other studies suggested that dividing SVZ cells serve as a reservoir of precursors for glial cells. Yet other investigators suggested that the dividing cells in the SVZ die soon after mitosis.

Direct demonstration that large numbers of neuronal precursors exist in the adult SVZ was obtained by culturing SVZ cells that had been labeled *in vivo* with [^{3}H]thymidine. *In vivo* adult SVZ cells generate new neurons that migrate to the olfactory bulb where they constantly replace interneurons, a process that also continues in adult primates. It has been suggested that SVZ cells may also give rise to neurons that migrate to other brain regions, but this process remains to be clearly demonstrated. In addition, the postnatal SVZ appears to be an important source of glial cells.

The cellular composition and organization of the SVZ has been described (Fig. 2). SVZ neuroblasts (type A cells), migrate in homotypic chains through a network of interconnecting pathways distributed throughout the wall of the lateral ventricle. These chains are ensheathed by slowly proliferating type B cells. B cells have the ultrastructural properties of astrocytes, including a light cytoplasm, thick bundles of glial fibrillary acidic protein (GFAP) positive intermediate filaments, gap junctions, glycogen granules, and dense bodies. Scattered along the chains of type A cells are clusters of rapidly dividing type C cells. Most of the SVZ is separated from the ventricle by a layer of multiciliated ependymal cells. Many of the chains of type A cells coalesce in the anterior and dorsal SVZ forming a restricted path of migration called the *rostral migratory stream* (RMS). The RMS is the pathway along which neuroblasts migrate from the SVZ to the olfactory bulb, where the neuroblasts differentiate into granule and periglomerular neurons.

Several lines of evidence indicate that a subset of SVZ astrocytes function as the primary neuronal precursors *in vivo*. Following infusion of the antimitotic drug cytosine-β-D-arabinofuranoside (Ara-C) into the adult mouse brain for six days, neuroblasts (A cells) and rapidly dividing C cells are eliminated. This treatment spares some SVZ astrocytes and ependymal cells. About 12 h after Ara-C removal, SVZ astrocytes begin dividing. Dividing SVZ astrocytes give rise to type C cells, which in turn generate neuroblasts. Within 10 days, the entire SVZ regenerates. Ependymal cells do not incorporate mitotic markers at any of the survivals studied, indicating that these cells are not the primary precursors of the new neurons.

GFAP-expressing astrocytes can be specifically labeled in transgenic mice expressing the receptor for an avian retrovirus under control of the GFAP promoter. SVZ astrocytes labeled with this method also generate type C cells, followed by neuroblasts that migrate to the olfactory bulb, indicating that under nonregenerating conditions SVZ astrocytes are also the primary precursors of the new neurons. As indicated above, a subpopulation of SVZ cells in the adult rodent brain can respond to EGF to generate neural

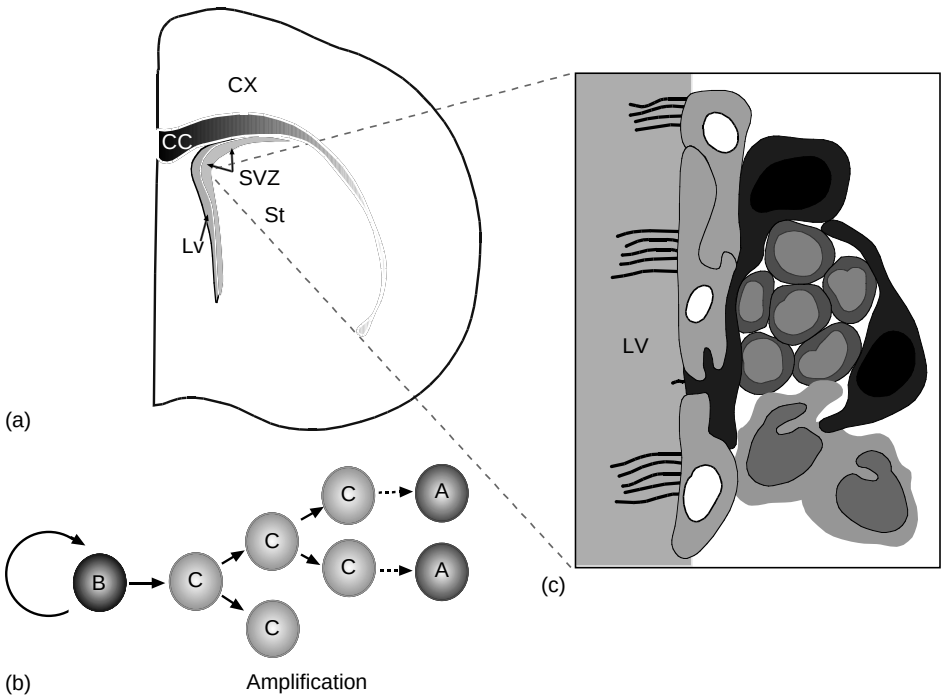

Fig. 2 Neurogenesis in the adult SVZ. (a) Cross-section of the adult mouse brain and (b) schematic illustration of SVZ composition and architecture; (c) expanded region. The SVZ lies adjacent to the lateral wall of the lateral ventricle (LV, tan). The chains of neuroblasts (A cells, red) destined for the olfactory bulb migrate as chains through glial tunnels formed by the processes of slowly dividing astrocytes (B cells, blue). Focal clusters of rapidly dividing precursors (C cells, green) are found scattered along the network of chains. Multiciliated ependymal cells line the wall of the ventricle (gray). Occasionally, an SVZ astrocyte extends a process with a single cilium to contact the ventricle. The lineage of olfactory bulb neurons has been described. SVZ astrocytes (B cells) act as neural stem cells in this region and divide to give rise to the rapidly dividing precursors (C cells) that in turn generate the neuroblasts that migrate to the olfactory bulb. (See color plate p. xxix).

stem cells *in vitro*. Cells derived from vitally labeled SVZ astrocytes respond to EGF signaling, both *in vivo* and *in vitro*. These cells can undergo self-renewal and can generate both neurons and glia. This suggests that SVZ astrocytes (B cells) correspond to the primary precursors of not only neurons, but also of neurospheres in the adult SVZ. Interestingly, astrocytes from other brain regions can also generate *in vitro* stem cells, but only if isolated from mice younger than 10 days. After this time, only astrocytes from the SVZ have this capacity. Importantly however, most of the EGF-responsive SVZ progenitors *in vivo* that give rise to neurospheres are not SVZ astrocytes, but rather the transit amplifying cells, the C cells. This suggests that as primary precursors (type B cells) are activated for proliferation and convert into type C cells, they retain self-renewal and multilineage potential if exposed *in vitro* to high concentrations of growth factors. Consistently, O2A progenitors, a transit

amplifying cell population from the optic nerve previously believed to be committed to the glial lineage, also generate cells with neural stem cell characteristics *in vitro* when exposed to growth factors.

Further verification that astrocytes are the neural stem cells comes from studies using transgenic mice expressing herpes simplex virus thymidine kinase (TK) from the GFAP promoter. Exposure to ganciclovir (GCV) leads to the production of a toxic metabolite, only in cells expressing TK, resulting in their selective killing. *In vivo* infusion of GCV into GFAP-TK mice for three days resulted in a nearly complete depletion (96%) of neurosphere forming cells from the SVZ. Together, the above results suggest that neural stem cells are contained within a population of cells classically considered to be differentiated macroglia.

Multiciliated ependymal cells have also been suggested to function as the neural stem cells of the SVZ *in vivo* and *in vitro*. However, several other studies do not support this interpretation. Most importantly, as mentioned above, there is no evidence that multiciliated ependymal cells divide *in vivo*. Ependymal cells do, however, play an important role in adult neurogenesis. Bone morphogenetic proteins (BMPs), which are powerful inhibitors of SVZ neurogenesis, are produced by SVZ astrocytes. This suggests that the neurogenic potential of many SVZ astrocytes is under constant inhibition by BMPs. Ependymal cells secrete noggin, which neutralizes the inhibitory effects of BMPs in the SVZ. Thus, the close interaction between the SVZ astrocytes and the ependymal layer is probably critical for the creation of a neurogenic niche.

Proliferation in the adult SVZ may be affected by lesions of the brain and of the corpus callosum, inhibition of nitric oxide, and stimulation of estrogen release in prairie voles. It will be interesting to determine precisely which cells are affected by the above manipulations. However, the problem is complex, as multiple cells divide in the SVZ, including the young neurons. Since these neuroblasts are migrating tangentially, apparent effects on BrdU labeling may not be due to changes in the rate of division of progenitors, but to the accumulation or depletion of neuroblasts resulting from changes in migratory rates.

The above observations demonstrate that SVZ astrocytes behave as neural stem cells *in vivo* and *in vitro*. This raises fundamental questions about the nature of SVZ astrocytes and how they differ from astrocytes in other brain regions. Astrocytes are very heterogeneous in structure, morphology, and function. Better markers are required to identify the different classes of astrocytes in the adult brain and, in particular, those that can function as stem cells. As mentioned earlier, neurogenesis also continues in the SGL layer of the hippocampus. In the next section, we will review recent work that suggests that astrocytes in the SGL also function as the primary precursors for new neurons.

5.2 The Subgranular Zone (SGZ)

The SGZ is the birth site of new neurons that become incorporated into the adult dentate gyrus of the hippocampus. In contrast to the SVZ or the adult avian VZ, the SGZ is not located close to the walls of the brain ventricles. It lies deep within the brain at the interface of the granule cell layer of the dentate gyrus and the hilus in the hippocampus. There is much interest in understanding the origin of these new neurons, as the hippocampus has such

an important function in learning and memory. Earlier work suggested that small darkly stained cells of "undifferentiated" appearance corresponded to the neural stem cells of the adult SGZ. *In vitro* studies also indicated that cells isolated from the adult hippocampus have neural stem cell properties *in vitro*. However, the identity of the cells *in vivo* that give rise to these stem cells in culture remains unknown.

In addition to the dark cells, astrocytes continue to divide in the SGZ of the adult hippocampus. Division of astrocytes was generally attributed to a process of local gliogenesis. This ongoing gliogenesis was thought to support neuronal function and neuronal turnover. However, more recent work suggests that SGZ astrocytes are the primary precursors of the new neurons of the adult hippocampus. We will next discuss how the SGZ is organized and the types of cells involved in adult hippocampal neurogenesis.

Two types of astrocytes have been identified in the SGZ, radial astrocytes, and horizontal astrocytes. Both types of astrocytes express GFAP, but only radial astrocytes express Nestin. Radial astrocytes have their cell bodies in the SGZ and have a prominent radial process that traverses the granular cell layer (Fig. 3). Radial SGZ astrocytes also have shorter tangential processes extending along the SGZ under the blades of the dentate gyrus. At the electron microscope, these cells have ultrastructural features of astrocytes similar to those of the SVZ, with a light cytoplasm containing intermediate filaments, and thin processes that intercalate between neighboring cells. Soon after labeling with a marker of DNA synthesis, such as BrdU

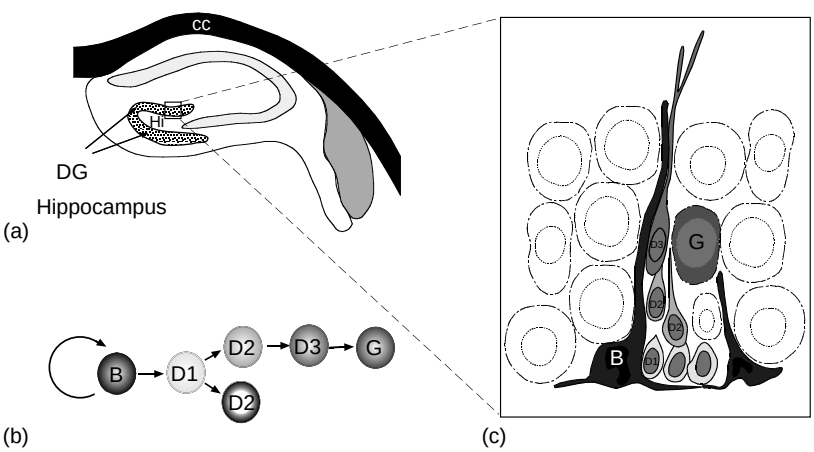

Fig. 3 Neurogenesis in the adult SGZ. (a) Cross-section of the adult mouse brain at the level of the hippocampus and (b) schematic illustration of composition and architecture of the SGZ and granule cell layer; (c) expanded region. SGZ astrocytes (B cells) have long radial processes that penetrate the granule cell layer and short tangential ones that run parallel to the SGZ. Dividing B cells generate immature, dark cells (D cells) that divide and mature into new granule neurons. D cells have been classified into three subtypes, D1, D2, and D3 that correspond to different stages in the maturation of the new neurons. The SGZ astrocytes (Blue cytoplasm, black nucleus); D cells (yellow or orange cytoplasm, gray nucleus), G: mature granule neuron (dotted, not filled). (See color plate p. xxiv).

or [^{3}H]thymidine, about half of the labeled cells in the SGZ correspond to astrocytes. Interestingly, with longer survivals after BrdU or [^{3}H]thymidine labeling, the number of labeled astrocytes rapidly decreases and the number of labeled GFAP-negative cells increases. A rapid decrease in the number of labeled GFAP-positive astrocytes with extended survival after BrdU injection has also been described in the dentate gyrus of adult rats. Close examination of this population of GFAP-negative cells that appears after increasing survival time revealed a small dark cell population (D cells). These cells probably correspond to the small dark cells previously suggested to function as the precursors of the new neurons (mentioned earlier).

Type D cells (or Type 2 progenitors as they were called in other studies), however, do not correspond to the primary precursors of the new neurons, but to an intermediate progenitor that is derived from the division of radial astrocytes. This was demonstrated by following the fate of dividing SGZ astrocytes after an antimitotic treatment and demonstrating that radial SGZ astrocytes cells give rise to D cells and granule neurons. Moreover, when transgenic mice that express the receptor for an avian retrovirus under control of the GFAP promoter are infected with an avian retrovirus expressing the reporter gene alkaline phosphatase, the dividing population of GFAP-expressing cells in the SGZ can be selectively labeled. One month after such selective labeling, labeled neurons are found in the granule cell layer. The neurons derived from the GFAP-expressing cells extended axonal projections into the CA3 region of Ammon's horn, indicating that they were incorporated into the normal circuitry of the hippocampus. Together, these experiments indicate that SGL astrocytes are the primary precursors that give rise to new granule neurons in the adult dentate gyrus through the intermediate D cells. As indicated above, neural stem cells can also be isolated and grown in culture from the adult hippocampus, but the *in vivo* cell type that gives rise to these stem cells *in vitro* has not been identified.

Detailed examination of the SGZ architecture has revealed a close interaction of radial astrocytes, the primary progenitors of the new neurons, with their progeny the D cells (Fig. 3). SGZ astrocytes and their processes form basketlike structures that nest tight clusters of D cells. Within these nests, D cells (Type 2 precursors) have cellular protrusions of different lengths and orientations, and begin expressing neuronal markers including Doublecortin, PSA-NCAM, Tuc4, NeuroD, Prox 1, and NeuN. Consistently, D cells have electrophysiological properties of young neurons. D cells have been classified into three subtypes, D1, D2, and D3, and shown to correspond to different stages in the maturation of the new neurons.

D1 cells are most immature, bearing no process or only a short one. Apparently, radial astrocytes divide asymmetrically to generate D1 cells that remain closely associated with the cell body of the progenitor astrocyte. Interestingly, D1 cells divide again to generate two D2 cells that bear a short process. Sometimes both daughter cells survive, but more frequently one of the daughters dies (Fig. 3). This is consistent with the finding that about half of all cells die during the maturation process. D2 cells then mature into D3 cells with a longer process that traverses the granule cell layer to become the primary dendrite of the newly formed granule neuron. D3 cells then mature into fully differentiated granule neurons.

Newly formed neurons then progressively lose PSA-NCAM from dendrites, cell body, axon and terminal synaptic arbor as they mature. During this maturation, the developing primary process of D3 cells, which gives rise to the dendritic arbor, remains closely associated with the radial processes of SGZ astrocytes. Thus D1, D2, and D3 cells remain closely associated with astrocytes throughout their maturation. Radial astrocytes appear to have a dual function; these cells give birth to the new neurons and support them throughout their maturation. They appear to provide scaffolding and probably support their development. The importance of astrocytes for supporting neurogenesis has also been demonstrated *in vitro* for SVZ and SGZ progenitors.

It is not known how much tangential dispersion of D cells occurs within the SGZ before these cells complete their differentiation. One possibility is that D cells do not migrate tangentially but remain largely confined to one radial domain, perhaps associated to their mother astrocytic cells. If this is the case, interesting hypotheses can be raised on the role that the primary precursor may play in the local control of neuronal addition.

The large radially oriented process of astrocytes that function as the primary precursors of new neurons may have important functions in relaying information from within the granule cell layer to the SGZ. This process may serve like an "antenna" to relay information from the granule cell layer and to the SGZ. This information may help regulate the proliferation of progenitor cells. Multiple factors can affect neurogenesis, and/or neuronal recruitment in the adult dentate gyrus, including adrenal steroids, the blockade of NMDA receptors, seizures, housing animals in enriched conditions, exercise, inflammation, and antidepressants. These different conditions may directly or indirectly affect the level of neuronal activity within the dentate gyrus and this may be detected by radial astrocytes inducing changes in the proliferation rate of these cells and in their progeny. It will be interesting to determine how these experimental paradigms affect the proliferation rates of astrocytes and D cells. Increased neurogenesis may require the induction of larger numbers of SGZ neural stem cells to enter the cell cycle. Interestingly, neuronal birth in the dentate gyrus increases following adrenalectomy, and this same condition results in a parallel increase in the proliferation of SGZ astrocytes. In this case, more astrocytes may become activated to accommodate the increased demand for new neurons following adrenalectomy. The identification of the primary precursors and of the intermediate cellular stages in the production of new hippocampal neurons will aid in the study of how neurogenesis is regulated in the adult mammalian hippocampus.

6 Organization and Primary Precursors of New Neurons in the Embryonic and Postnatal Mammalian Brain

In order to understand the developmental origin of adult neural stem cells and the derivation of adult germinal centers, we must be familiar with the identity and properties of embryonic neural stem cells. This section will discuss the organization of the early neuroepithelium and why these cells are believed to give rise to radial glia. Furthermore, we will discuss the structural and

stem cell functions of radial glia during development.

6.1 Radial Glia are the Neural Progenitors of the Developing Brain

As mentioned above, radial glia and their processes serve as scaffolding during brain development and, in particular, as guides for the migration of young neurons from their site of birth in the VZ to their final destination in different substrata of the developing brain. At the end of fetal development, radial glia in the mammalian brain retract from the ventricular and pial surfaces and transform into astrocytes that migrate into the underlying brain. As the name radial "glia" implies, these cells are thought to be glial cells. This is not only because they are immediate precursors of astrocytes but also because radial glial cells express markers and have morphological and ultrastructural characteristics of astrocytic cells.

Radial glia are thought to be derived from neuroepithelial cells, the earlier brain progenitors that form the neural plate and neural tube. In the embryonic neural tube, neuroepithelial cells contact both the ventricular and pial surfaces. As development proceeds and this neuroepithelium gives rise to the multilayered fetal brain, the distance from the ventricle to the brain surface greatly increases. Neuroepithelial cells transform into elongated radial glia by maintaining their ventricular and pial contacts. Radial glia and neuroepithelial cells have many things in common. Both cell types undergo mitosis with interkinetic nuclear migration. Radial glia contact the ventricular lumen into which they extend a single short cilium, which is very similar, if not identical, to the cilia found on the ventricular surface of neuroepithelial cells. Interestingly, this short cilium on radial glia and neuroepithelial cells is very similar to the one observed on adult neural stem cells (mentioned earlier). The function of this subcellular structure is not known, but it likely holds important secrets about neural stem cell regulation.

Both radial glia and neuroepithelial cells express the intermediate filament nestin, which is widely used to label neural stem cells in culture. This, together with the work in adult birds discussed earlier, began in the early 1990s and suggests that radial glial cells may be something more than just "glia." After more than a century since Magini, Ramón y Cajal and others began to describe radial glia, or *spongioblasts* as they were also called at that time, the true nature of these cells has emerged. Radial glia are the stem cells of the developing CNS. A recent series of elegant studies in the developing mammalian brain have shown that radial glia divide and generate a wide array of brain cells including neurons. Radial glia isolated from embryonic mouse and rat neocortex can produce neurons and glial cells *in vitro*. Radial glia isolated during the time when a large number of neurons are formed in the cerebral cortex, produced mostly colonies containing neurons. In contrast, when radial glial cells are isolated after this peak in neurogenesis, they generate mostly glial or mixed colonies. This suggests that radial glial cells can generate both neurons and glia and that their potential changes with age.

The most direct proof of the progenitor function of radial glia comes from direct visualization of the behavior of fluorescently labeled radial glia in slice cultures. With this approach, one can directly observe cortical radial glia dividing asymmetrically to

produce young neurons, which then migrate into the cortex. Noctor et al. went on to make electrophysiological recordings from clonally related cells to confirm the identity of radial glia and their neuronal progeny. Taken together, the above clearly demonstrate that radial glia can generate neurons. More recent work indicates that in some instances radial glia divide asymmetrically, but produce neurons indirectly. The daughter cell moves away from the VZ into the SVZ and divides again to generate two neurons. This last pattern of neuronal generation, also having as an origin a radial glial cell, is reminiscent of that observed in the adult SVZ and SGZ, where a primary precursor first gives rise to a transit amplifying cell, which then divides to generate the neurons.

Radial glial cells appear to be much more than just "glia" or guide posts for neuronal migration. These complex cells are the stem cells that through repeated division directly or indirectly generate the amazing array of neurons and glial cells that shape the brain. Radial glial cells are not simple undifferentiated cells, as would be expected for a stem cell, and this fact had likely tricked us into believing for over a century that they were differentiated glial cells. Radial glia are large and complex cells that maintain specialized contacts with all strata of the developing brain; including the pial surface, blood vessels, neurons, and other progenitors in the VZ. These cells also form organized palisades that likely play an important structural role for the developing brain as classically thought. In addition to providing spatial alignment between germinal centers close to the brain ventricles and the more superficial brain regions where neurons take up residence, the long processes of radial glia could carry signals from more superficial strata of the developing brain mantel into the germinal epithelium in the VZ. This radially organized alignment system may be particularly important for the development and evolution of the mammalian neocortex.

The notion that radial glia are just scaffolding and support for embryonic brain development must change, as these cells are primary precursors for an enormous number of neurons. This new concept has immense implications for our understanding of the molecular and cellular basis of brain development. It is within radial glia that much of the program for brain histogenesis must unfold. Therefore, we must look to these cells to find the molecular switches that control the well-orchestrated generation of the different cell types that form the brain.

Contrary to what was expected, the cell type in the embryonic brain with the most complex and "differentiated" characteristics (long radial processes, elaborate cytoskeleton, complex endings in the pial surface, and contacts with blood vessels), the radial glia, turn out to be a primary precursor of neurons. Recent evidence indicates that these cells are also the precursors of adult neural stem cells, thus, revealing the lineage tree of neural stem cells.

7
The Lineage of Neural Stem Cells; Evolving Concepts

The above indicates that cells that have been classically considered as differentiated glial cells in the adult and the embryo correspond to the primary precursors of new neurons. The lineages leading to the formation of neurons and

glial cells were classically considered to diverge early during embryonic development. In this traditional lineage tree, astrocytes and radial glia were drawn in one branch, while neurons were placed in a separate one. On the basis of the above findings, this concept has begun to change. In this last section, we will discuss how neural stem cells in the developing and adult brain are likely to be contained within what was classically considered a macroglia lineage – neuroepithelium-radial glia-astrocytic cells – (Fig. 4). It is from this central stem that the differentiated lineages in the brain emerge.

Since the discovery of glial cells in the nervous system in the middle of the nineteenth century, these supportive cells have been considered the end product of differentiation rather than the stem from which other cells were derived. Initially, it was believed that glial cells had a mesenchymal origin and corresponded to the connective tissue of the brain. This idea was overturned by Wilhelm His, who showed that glial cells were

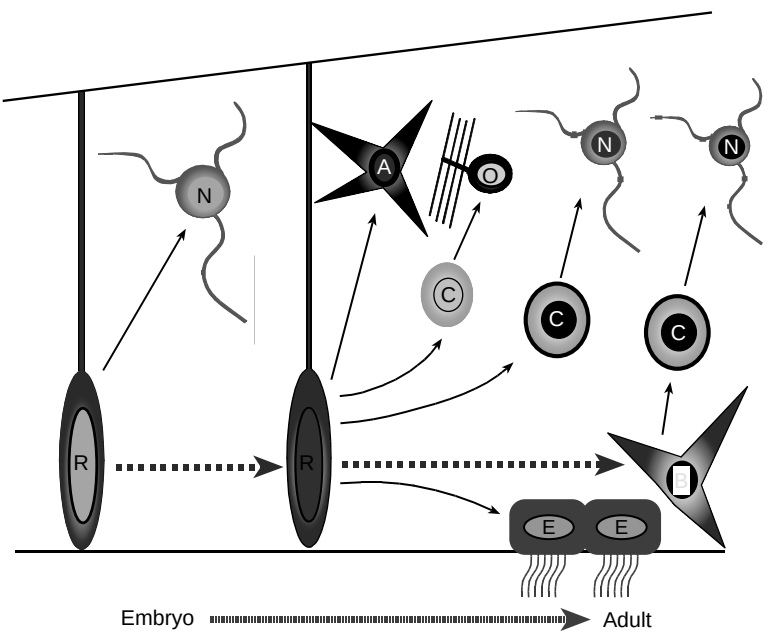

Fig. 4 Development of neural stem cells from the embryo to the adult. Primary progenitors in the embryo correspond to radial glia (R) in blue. These cells produce different generations of neurons (N, red) during development either directly, or indirectly through transit amplifying progenitors (C cells, green). Radial glia also generate oligodendrocytes (O, black) and ependymal cells (E, gray). It is not known whether single radial glia generate these different cell types or whether radial glia are lineage-restricted *in vivo*. Radial glia also turn into parenchymal astrocytes (A, black-blue) and germinal region astrocytes (B, blue-black). The latter continue to divide in the adult brain and generate neurons (N) and oligodendrocytes (not shown) through transit amplifying cells (C)]. The neural stem cell lineage is contained within what was classically considered an astrocytic lineage (Blue cells and arrows). (See color plate p. xxxiii).

derived from the so-called "spongioblasts" (what we now know as radial glia). From his work emerged the dominant hypothesis that in the early neural epithelium, predetermined precursors produced either neurons or glia. This concept of a separate origin for neurons and glia has remained very influential in neuroscience until fairly recently.

Neurons and glial cells are not as separate as once thought. Work in insects, for example, has shown that neurons and glial cells may be produced from a final division of a progenitor. Work in vertebrates has also suggested that both glial cells and neurons may be derived from a common progenitor. Other studies, however, continue to support the idea that glial-restricted and neuronal-restricted progenitors give rise to the corresponding differentiated cells in the brain. Both lineage-restricted precursors and multipotential stem cells may exist in the brain and as such these two camps may not necessarily subscribe to mutually exclusive ideas.

Two major findings suggest that stem cells are contained within what was earlier considered a glial-restricted lineage; the demonstration of radial glia as progenitor cells of not only astrocytes but also neurons, and the identification of multipotent precursors with astrocytic characteristics that support ongoing neurogenesis in the adult brain. This is not to say that all cells with astrocytic characteristics are stem cells. In fact, the name "astrocyte" has been used to describe a wide array of cells in the brain, including some (e.g. stellate astrocytes) that do not divide and appear to be terminally differentiated glial cells. Clearly, progenitor potential is inherited by some astrocytes, and this suggests that a continuum between embryonic and adult neural stem cells may exist to form the stem cell lineage in the brain. The question is: which cells in the developing brain are the progenitors of the adult neural stem cells?

One hypothesis is that it is the radial glial cells. These cells are known to give rise to astrocytes, some of which may be the ones that retain progenitor function in the adult germinal niches. Notch is believed to maintain cells in a proliferative progenitor state and inhibit neuronal differentiation. When Notch is activated in early neuroepithelial cells, it promotes radial glial identity, consistent with the idea that radial glia are derived from neuroepithelial cells and that they possess stem cell potential. This same study suggests that Notch activation in the early neuroepithelium promotes astrocytic identity in the adult brain including the SVZ. This observation lends support to the idea that radial glia in the lateral wall of the lateral ventricle may transform into the SVZ astrocytes that continue to function as stem cells in the adult. Direct evidence linking radial glia in development with adult stem cells has recently been obtained.

In mice, as in other mammals, most but not all radial glia have transformed into astrocytes by birth. A subpopulation of radial glia persists in the lateral wall of the lateral ventricle of newborn mice. Interestingly, these radial cells are precisely in the region where the adult germinal SVZ will develop. This offered an opportunity to directly test whether radial glial cells in this region serve as progenitors of adult neural stem cells. The cell bodies of neonatal radial glia line the lateral wall of the lateral ventricle and form a VZ. Their long basal processes radiate outward, curving through the striatum toward the surface of the brain. The disappearance, within the first week of life, of radial glia from

the lateral ventricular wall correlates with the appearance of GFAP+ astrocytes, suggesting a possible link between these two cell populations. Taking advantage of their unique long processes that curve around the striatum, a technique was developed to selectively and permanently label a small cohort of radial glia. When a virus expressing the DNA recombinase, Cre, is injected into the lateral striatum near the brain surface, it infects the radial glial cell processes and is transported to their cell bodies in the VZ. Specific labeling of these radial glial cells can be attained when the infection is performed in transgenic mice carrying a reporter gene behind a transcriptional stop element flanked by Cre-recombination sites. Examination of the fate of these labeled radial glial cells at varying time points after the initial injection, demonstrated that radial glia give rise to olfactory bulb granular and periglomerular interneurons, and striatal and SVZ astrocytes. This is consistent with the traditional concept of radial glial cells being precursors of astrocytes and with the more recent view that radial glia are also neuronal progenitors (mentioned earlier). Interestingly, however, these radial glial cells also give rise to two other cell populations that were not known to be derived from radial glia: oligodendrocytes and ependymal cells. This demonstrates that this population of neonatal radial glia act as multipotent progenitors, but it remains unknown whether individual radial glia are multipotent *in vivo*. When individual labeled radial glia were placed in culture, they generated neurospheres. The neurospheres grown from radial glia self-renewed and differentiated into neurons, astrocytes, and oligodendrocytes, indicating that striatal radial glia act as true neonatal stem cells *in vitro*. Most importantly, some of these neonatal stem cells transformed to generate the neurogenic stem cells of the adult SVZ (Fig. 4). These results strongly support the hypothesis that neural stem cells are contained within the radial glia–astrocyte lineage.

The evolving view suggests that differentiated cells, including neurons and glia, are derived from neural stem cells which include neuroepithelial cells, radial glia, and astrocytes. Progenitors in the VZ of the developing mammalian brain are programmed spatially and temporally to produce different types of neurons. It is, therefore, within neuroepithelial cells, radial glia, and astrocytes that these molecular programs unfold for the generation of different cell types. As developmental time progresses and these cells convert into adult progenitors, neural stem cells constantly change. The most rapid change probably occurs during embryonic development, as neuroepithelial cells and radial glia adjust their genetic program to produce the different types of neurons and glial cells required for assembly of the nervous system. The temporal control of neurogenesis is illustrated in Fig. 4, by showing neurons of different sizes and shades produced at the different developmental stages. This is an oversimplification of a process that includes multiple reiterations and many more cell types, affecting not only neurons but also glial cells. Some of the genetic control for the generation of different types of neurons has begun to be uncovered in the insect brain, and this is likely to become a dominant theme in vertebrate neural development and in trying to understand neural stem cell regulation. Thus, while we postulate the neuroepithelium-radial glia-astrocyte as the lineage within which neural stem cells are contained, this does not mean that all these cells are equivalent. The types of neurons and glial cells produced at each

developmental stage may be quite unique and dependent on an internal program that these cells carry. Therefore, in the strictest sense, neural stem cells *in vivo* are probably not self-renewing, especially during development.

8 Concluding Remarks

A profound change in concepts is taking place in our understanding of neural development and the nature of neural stem cells. The demonstration of adult neurogenesis has not only challenged long-held dogmas about brain development and plasticity, but has allowed us to study the nature and origin of neural stem cells. We have reviewed above a string of discoveries leading to the suggestion that neural stem cells are contained within a lineage of neuroepithelial cells, radial glia, and astrocytes, which was once considered a lineage leading to the generation of only differentiated glial (support) cells. These new concepts could have important implications for how we think about development and help us devise new strategies for brain repair. Perhaps most surprising is that neural stem cells are not simple ''undifferentiated'' progenitor cells. They are structurally quite complex, with long processes that interact closely with other cells in the brain and most prominently with their own progeny. Thus, in addition to being progenitors, they also appear to be true supportive cells as had been originally attributed to glia. This combined function – progenitor and support – may be inherited from a time in evolution when epithelial cells first ''invented'' how to make nerve cells. As suggested by Ramón y Cajal almost a century ago, astrocytes may be nothing more than modified epithelial cells. An important challenge for the future is to try to identify the different subtypes of astrocytes in the adult brain and learn to distinguish those that retain stem cell potential.

Acknowledgments

We thank Cynthia Yaschine for helpful assistance during the preparation of the manuscript. This work was supported by National Institutes of Health Grants NS28478 and HD32116 and by a gift from Frances and John Bowes. Erica Jackson is an American Cancer Society Postdoctoral Fellow.

See also Brain Development; Molecular Neurobiology, Single-Cell; Neuron Chemistry; Stem Cells; Stem Cells, Embryonic.

Bibliography

Books and Reviews

Jacobson, M. (1991) *Developmental Neurobiology*, 3rd edition, Plenum Press New York.

Primary Literature

Ahmed, S., Reynolds, B.A., Weiss, S. (1995) BDNF enhances the differentiation but not the survival of CNS stem cell-derived neuronal precursors, *J. Neurosci.* **15**, 5765–5778.

Allen, E. (1912) The cessation of mitosis in the central nervous system of the albino rat, *J. Comp. Neurol.* **22**, 547–568.

Altman, J. (1962) Are new neurons formed in the brains of adult mammals? *Science* **135**, 1127–1128.

Altman, J. (1969) Autoradiographic and histological studies of postnatal neurogenesis. IV. Cell proliferation and migration in the

anterior forebrain, with special reference to persisting neurogenesis in the olfactory bulb, *J. Comp. Neurol.* **137**, 433–458.

Altman, J. (1970) Postnatal Neurogenesis and the Problem of Neural Plasticity, in: Himwich, W.A. (Ed.) *Developmental Neurobiology*, C.C.Thomas, Springfield, IL, pp. 197–237.

Altman, J., Das, G.D. (1965a) Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats, *J. Comp. Neurol.* **124**, 319–336.

Altman, J., Gopal, D.D. (1965b) Post-natal origin of microneurones in the rat brain, *Nature* **207**, 953–956.

Alvarez-Buylla, A., Nottebohm, F. (1988) Migration of young neurons in adult avian brain, *Nature* **335**, 353–354.

Alvarez-Buylla, A., Theelen, M., Nottebohm, F. (1990a) Proliferation "hot spots" in adult avian ventricular zone reveal radial cell division, *Neuron* **5**, 101–109.

Alvarez-Buylla, A., Kirn, J.R., Nottebohm, F. (1990b) Birth of projection neurons in adult avian brain may be related to perceptual or motor learning, *Science* **249**, 1444–1446.

Alvarez-Buylla, A., Ling, C.-Y., Yu, W.S. (1994) Contribution of neurons born during embryonic, juvenile, and adult life to the brain of adult canaries: regional specificity and delayed birth of neurons in the song-control nuclei, *J. Comp. Neurol.* **347**, 233–248.

Alvarez-Buylla, A., Garcia-Verdugo, J.M., Tramontin, A.D. (2001) A unified hypothesis on the lineage of neural stem cells, *Nat. Rev. Neurosci.* **2**, 287–293.

Alvarez-Buylla, A., Garcia-Verdugo, J.M., Mateo, A., Merchant-Larios, H. (1998a) Primary neural precursors and intermitotic nuclear migration in the ventricular zone of adult canaries, *J. Neurosci.* **18**, 1020–1037.

Alvarez-Buylla, A., Garcia-Verdugo, J.M., Mateo, A.S., Merchant-Larios, H. (1998b) Primary neural precursors and intermitotic nuclear migration in the ventricular zone of adult canaries, *J. Neurosci.* **18**, 1020–1037.

Arsenijevic, Y., Weiss, S. (1998) Insulin-like growth factor-1 is a differentiation factor for postmitotic CNS stem cell-derived neuronal precursors: distinct actions from those of brain-derived neurotrophic factor, *J. Neurosci.* **18**, 2118–2128.

Barami, K., Kirschenbaum, B., Lemmon, V., Goldman, S.A. (1994) N-cadherin and Ng-CAM/8D9 are involved serially in the migration of newly generated neurons into the adult songbird brain, *Neuron* **13**, 567–582.

Barres, B.A. (2003) What is a glial cell? *Glia* **43**, 4–5.

Bayer, S.A., Yackel, J.W., Puri, P.S. (1982) Neurons in the rat dentate gyrus granular layer substantially increase during juvenile and adult life, *Science* **216**, 890–892.

Bentivoglio, M., Mazzarello, P. (1999) The history of radial glia, *Brain Res. Bull.* **49**, 305–315.

Blakemore, W.F. (1969) The ultrastructure of the subependymal plate in the rat, *J. Anat.* **104**, 423–433.

Blakemore, W.F., Jolly, D.R. (1972) The subependymal plate and associated ependyma in the dog. An ultrastructural study, *J. Neurocytol.* **1**, 69–84.

Burd, G.D., Nottebohm, F. (1985) Ultrastructural characterization of synaptic terminals formed on newly generated neurons in a song control nucleus of the adult canary forebrain, *J. Comp. Neurol.* **240**, 143–152.

Cameron, H.A., Tanapat, P., Gould, E. (1998) Adrenal steroids and N-methyl-D-aspartate receptor activation regulate neurogenesis in the dentate gyrus of adult rats through a common pathway, *Neuroscience* **82**, 349–354.

Cameron, H.A., Wooley, C.S., McEwen, B.S., Gould, E. (1993) Differentiation of newly born neuron and glia in the dentate gyrus of the adult rat, *Neuroscience* **56**, 337–344.

Carleton, A., Petreanu, L.T., Lansford, R., Alvarez-Buylla, A., Lledo, P.M. (2003) Becoming a new neuron in the adult olfactory bulb, *Nat. Neurosci.* **6**, 507–518.

Chetverukhin, V.K., Polenov, A.L. (1993) Ultrastructural radioautographic analysis of neurogenesis in the hypothalamus of the adult frog, *Rana temporaria*, with special reference to physiological regeneration of the preoptic nucleus. I. Ventricular zone cell proliferation, *Cell Tissue Res.* **271**, 341–350.

Chiasson, B.J., Tropepe, V., Morshead, C.M., Van der Kooy, D. (1999) Adult mammalian forebrain ependymal and subependymal cells demonstrate proliferative potential, but only subependymal cells have neural stem cell characteristics, *J. Neurosci.* **19**, 4462–4471.

Cohen, E., Meininger, V. (1987) Ultrastructural analysis of primary cilium in the embryonic

nervous tissue of mouse, *Int. J. Dev. Neurosci.* **5**, 43–51.

Connors, B.W., Ransom, B.R. (1987) Electrophysiological properties of ependymal cells (radial glia) in dorsal cortex of the turtle, *Pseudemys scripta*, *J. Physiol.* **385**, 287–306.

Doe, C.Q., Fuerstenberg, S., Peng, C.-Y. (1998) Neural stem cells: from fly to vertebrates, *J. Neurobiol.* **36**, 111–127.

Doetsch, F., Alvarez-Buylla, A. (1996) Network of tangential pathways for neuronal migration in adult mammalian brain, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14895–14900.

Doetsch, F., Garcia-Verdugo, J.M., Alvarez-Buylla, A. (1997) Cellular composition and three-dimensional organization of the subventricular germinal zone in the adult mammalian brain, *J. Neurosci.* **17**, 5046–5061.

Doetsch, F., Garcia-Verdugo, J.M., Alvarez-Buylla, A. (1999a) Regeneration of a germinal layer in the adult mammalian brain, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11619–11624.

Doetsch, F., Caille, I., Lim, D.A., Garcia-Verdugo, J.M., Alvarez-Buylla, A. (1999b) Subventricular zone astrocytes are neural stem cells in the adult mammalian brain, *Cell* **97**, 1–20.

Easter, S.S. (1983) Postnatal neurogenesis and changing connections, *Trends Neurosci.* **6**, 53–56.

Eriksson, P.S., Perfilieva, E., Bjork-Eriksson, T., Alborn, A., Nordborg, C., Peterson, D.A., Gage, F.H. (1998) Neurogenesis in the adult human hippocampus, *Nat. Med.* **4**, 1313–1317.

Filippov, V., Kronenberg, G., Pivneva, T., Reuter, K., Steiner, B., Wang, L.P., Yamaguchi, M., Kettenmann, H., Kempermann, G. (2003) Subpopulation of nestin-expressing progenitor cells in the adult murine hippocampus shows electrophysiological and morphological characteristics of astrocytes, *Mol. Cell. Neurosci.* **23**, 373–382.

Font, E., Garcia-Verdugo, J.M., Alcantara, S., Lopez-Garcia, C. (1991) Neuron regeneration reverses 3-acetylpyridine-induced cell loss in the cerebral cortex of adult lizards, *Brain Res.* **551**, 230–235.

Fukuda, S., Kato, F., Tozuka, Y., Yamaguchi, M., Miyamoto, Y., Hisatsune, T. (2003) Two distinct subpopulations of nestin-positive cells in adult mouse dentate gyrus, *J. Neurosci.* **23**, 9357–9366.

Gadisseux, J.F., Evrard, P., Misson, J.P., Caviness, V.S. (1989) Dynamic structure of the radial glial fiber system of the developing murine cerebral wall. An immunocytochemical analysis, *Dev. Brain Res.* **50**, 55–67.

Gaiano, N., Nye, J.S., Fishell, G. (2000) Radial glial identity is promoted by Notch1 signaling in the murine forebrain, *Neuron* **26**, 395–404.

Garcia-Verdugo, J.M., Ferron, S., Flames, N., Collado, L., Desfilis, E., Font, E. (2002) The proliferative ventricular zone in adult vertebrates: a comparative study using reptiles, birds, and mammals, *Brain Res. Bull.* **57**, 765–775.

Goldman, S.A. (1995) Neuronal Precursor cells and neurogenesis in the adult forebrain, *Neuroscientist* **1**, 338–350.

Goldman, S. (2003) Glia as neural progenitor cells, *Trends Neurosci.* **26**, 590–596.

Goldman, S.A., Nottebohm, F. (1983) Neuronal production, migration, and differentiation in a vocal control nucleus of the adult female canary brain, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2390–2394.

Goldman, S.A., Zaremba, A., Niedzwiecki, D. (1992) *In vitro* neurogenesis by neuronal precursor cells derived from the adult songbird brain, *J. Neurosci.* **12**, 2532–2541.

Gotz, M., Hartfuss, E., Malatesta, P. (2002) Radial glial cells as neuronal precursors: a new perspective on the correlation of morphology and lineage restriction in the developing cerebral cortex of mice, *Brain Res. Bull.* **57**, 777–788.

Gould, E., Cameron, H.A., McEwen, B.S. (1994) Blockade of NMDA receptors increases cell death and birth in the developing rat dentate gyrus, *J. Comp. Neurol.* **340**, 551–565.

Gould, E., Reeves, A.J., Graziano, M.S.A., Gross, C.G. (1999a) Neurogenesis in the neocortex of adult primates, *Science* **286**, 548–552.

Gould, E., Cameron, H.A., Daniels, D.C., Wooley, C.S., McEwen, B.S. (1992) Adrenal hormones suppress cell division in the adult rat dentate gyrus, *J. Neurosci.* **12**, 3642–3650.

Gould, E., Reeves, A.J., Fallah, M., Tanapat, P., Gross, C.G., Fuchs, E. (1999b) Hippocampal neurogenesis in adult old world primates, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5263–5267.

Gray, G.E., Sanes, J.R. (1992) Lineage of radial glia in the chicken optic tectum, *Development* **114**, 271.

Gritti, A., Cova, L., Parati, E.A., Galli, R., Vescovi, A.L. (1995) Basic fibroblast growth factor

supports the proliferation of epidermal growth factor-generated neuronal precursor cells of the adult mouse CNS, *Neurosci. Lett.* **185**, 151–154.

Grosskortenhaus, R., Pearson, B.J., Marusich, A., Doe, C.Q. (2005) Regulation of temporal identity transitions in Drosophila neuroblasts, *Dev. Cell.* **8**, 193–202.

Halliday, A.L., Cepko, C.L. (1992) Generation and migration of cells in the developing striatum, *Neuron* **9**, 15–26.

Hinds, J.W. (1968) Autoradiographic study of histogenesis in the mouse olfactory bulb. I. Time of origin of neurons and neuroglia, *J. Comp. Neurol.* **134**, 287–304.

Holland, E.C., Varmus, H.E. (1998) Basic fibroblast growth factor induces cell migration and proliferation after glia-specific gene transfer in mice, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1218–1223.

Horstmann, E. (1954) Die faserglia des selachiergehirns, *Z. Zellforsch.* **39**, 588–617.

Huang, L., DeVries, G.J., Bittman, E.L. (1998) Photoperiod regulates neuronal bromodeoxyuridine labeling in the brain of a seasonally breeding mammal, *J. Neurobiol.* **36**, 410–420.

Irvin, D.K., Zurcher, S.D., Nguyen, T., Weinmaster, G., Kornblum, H.I. (2001) Expression patterns of Notch1, Notch2, and Notch3 suggest multiple functional roles for the Notch-DSL signaling system during brain development, *J. Comp. Neurol.* **436**, 167–181.

Johansson, C.B., Momma, S., Clarke, D.L., Risling, M., Lendahl, U., Frisén, J. (1999) Identification of a neural stem cell in the adult mammalian central nervous system, *Cell* **96**, 25–34.

Kaplan, M.S., Hinds, J.W. (1977) Neurogenesis in the adult rat: electron microscopic analysis of light radioautographs, *Science* **197**, 1092–1094.

Kempermann, G., Kuhn, H.G., Gage, F.H. (1997) More hippocampal neurons in adult mice living in an enriched environment, *Nature* **386**, 493–495.

Kempermann, G., Kuhn, H.G., Gage, F.H. (1998) Experience-induced neurogenesis in the senescent dentate gyrus, *J. Neurosci.* **18**, 3206–3212.

Kirschenbaum, B., Doetsch, F., Lois, C., Alvarez-Buylla, A. (1999) Adult subventricular zone neuronal precursors continue to proliferate and migrate in the absence of the olfactory bulb, *J. Neurosci.* **19**, 2171–2180.

Koketsu, D., Mikami, A., Miyamoto, Y., Hisatsune, T. (2003) Nonrenewal of neurons in the cerebral neocortex of adult macaque monkeys, *J. Neurosci.* **23**, 937–942.

Kondo, T., Raff, M. (2000) Oligodendrocyte precursor cells reprogrammed to become multipotential CNS stem cells, *Science* **289**, 1754–1757.

Kornack, D.R., Rakic, P. (1999) Continuation of neurogenesis in the hippocampus of the adult macaque monkey, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5768–5773.

Kornack, D.R., Rakic, P. (2001a) The generation, migration, and differentiation of olfactory neurons in the adult primate brain, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4752–4757.

Kornack, D.R., Rakic, P. (2001b) Cell proliferation without neurogenesis in adult primate neocortex, *Science* **294**, 2127–2130.

Kosaka, T., Hama, K. (1986) Three-dimensional structure of astrocytes in the rat dentate gyrus, *J. Comp. Neurol.* **249**, 242–260.

Kriegstein, A.R., Gotz, M. (2003) Radial glia diversity: a matter of cell fate, *Glia* **43**, 37–43.

Laywell, E.D., Rakic, P., Kukekov, V.G., Holland, E.C., Steindler, D.A. (2000) Identification of a multipotent astrocytic stem cell in the immature and adult mouse brain, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13883–13888.

Lemullois, M., Klotz, C., Sandoz, D. (1991) Evolutionary conservation of an epitope associated with striated rootlets in different epithelial ciliated cells, *Biol. Cell.* **71**, 201–208.

Lendahl, U., Zimmerman, L.B., McKay, R.D.G. (1990) CNS stem cells express a new class of intermediate filament protein, *Cell* **60**, 585–595.

Levison, S.W., Goldman, J.E. (1993) Both oligodendrocytes and astrocytes develop from progenitors in the subventricular zone of postnatal rat forebrain, *Neuron* **10**, 201–212.

Levitt, P.R., Cooper, M.L., Rakic, P. (1981) Coexistence of neuronal and glial precursor cells in the cerebral ventricular zone of the fetal monkey: an ultrastructural immunoperoxidase analysis, *J. Neurosci.* **1**, 27–39.

Lewis, P.D. (1968) Mitotic activity in the primate subependymal layer and the genesis of gliomas, *Nature* **217**, 974–975.

Lewis, J. (1998) Notch signalling and the control of cell fate choices in vertebrates, *Semin. Cell. Dev. Biol.* **9**, 583–589.

Lim, D.A., Alvarez-Buylla, A. (1999) Interaction between astrocytes and adult subventricular zone precursors stimulates neurogenesis, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7526–7531.

Lim, D.A., Tramontin, A.D., Trevejo, J.M., Herrera, D.G., Garcia-Verdugo, J.M., Alvarez-Buylla, A. (2000) Noggin antagonizes BMP signaling to create a niche for adult neurogenesis, *Neuron* **28**, 713–726.

Lois, C., Alvarez-Buylla, A. (1993) Proliferating subventricular zone cells in the adult mammalian forebrain can differentiate into neurons and glia, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 2074–2077.

Lois, C., Alvarez-Buylla, A. (1994) Long-distance neuronal migration in the adult mammalian brain, *Science* **264**, 1145–1148.

Lopez-Garcia, C. (1993) Postnatal neurogenesis and regeneration in the lizard cerebral cortex, in: Cuello, C. (Ed.) *Neuronal Cell Death and Repair*, Elsevier, Amsterdam, pp. 237–246.

Lopez-Garcia, C., Tineo, P.L., Del Corral, J. (1984) Increase of the neuron number in some cerebral cortical areas of a lizard Podarcis hispanica, (Steind., 1870), during postnatal periods of life, *J. Hirnforsch.* **3**, 255–259.

Lopez-Garcia, C., Molowny, A., Garcia-Verdugo, J.M., Ferrer, I. (1988) Delayed postnatal neurogenesis in the cerebral cortex of lizards, *Dev. Brain Res.* **43**, 167–174.

Luskin, M.B. (1993) Restricted proliferation and migration of postnatally generated neurons derived from the forebrain subventricular zone, *Neuron* **11**, 173–189.

Magavi, S.S., Leavitt, B.R., Macklis, J.D. (2000) Induction of neurogenesis in the neocortex of adult mice, *Nature* **405**, 951–955.

Malatesta, P., Hartfuss, E., Gotz, M. (2000) Isolation of radial glial cells by fluorescent-activated cell sorting reveals a neuronal lineage, *Development* **127**, 5253–5263.

Mayer-Proschel, M., Kalyani, A.J., Mujtaba, T., Rao, M.S. (1997) Isolation of lineage-restricted neuronal precursors from multipotent neuroepithelial stem cells, *Neuron* **19**, 773–785.

McConnell, S.K. (1995) Strategies for the generation of neuronal diversity in the developing central nervous system, *J. Neurosci.* **15**, 6987–6998.

McDermott, K.W.G., Lantos, P.L. (1990) Cell proliferation in the subependymal layer of the postnatal marmoset, *Callithrix jacchus*, *Dev. Brain Res.* **57**, 269–277.

McKay, R. (1997) Stem cells in the central nervous system, *Science* **276**, 66–71.

Menezes, J.R.L., Smith, C.M., Nelson, K.C., Luskin, M.B. (1995) The division of neuronal progenitor cells during migration in the neonatal mammalian forebrain, *Mol. Cell. Neurosci.* **6**, 496–508.

Merkle, F.T., Tramontin, A.D., Garcia-Verdugo, J.M., Alvarez-Buylla, A. (2004) Radial glia give rise to adult neural stem cells in the subventricular zone, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17528–17532.

Misson, J.P., Edwards, M.A., Yamamoto, M., Caviness, V.S. Jr. (1988) Mitotic cycling of radial glial cells of the fetal murine cerebral wall: a combined autoradiographic and immunohistochemical study, *Dev. Brain Res.* **38**, 183–190.

Miyata, T., Kawaguchi, A., Okano, H., Ogawa, M. (2001) Asymmetric inheritance of radial glial fibers by cortical neurons, *Neuron* **31**, 727–741.

Molofsky, A.V., Pardal, R., Iwashita, T., Park, I.K., Clarke, M.F., Morrison, S.J. (2003) Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation, *Nature* **425**, 962–967.

Monje, M.L., Toda, H., Palmer, T.D. (2003) Inflammatory blockade restores adult hippocampal neurogenesis, *Science* **302**, 1760–1765.

Moreno-Lopez, B., Noval, J.A., Gonzalez-Bonet, L.G., Estrada, C. (2000) Morphological bases for a role of nitric oxide in adult neurogenesis, *Brain Res.* **869**, 244–250.

Morshead, C.M., Van der Kooy, D. (1992) Postmitotic death is the fate of constitutively proliferating cells in the subependymal layer of the adult mouse brain, *J. Neurosci.* **12**, 249–256.

Morshead, C.M., Garcia, A.D., Sofroniew, M.V., van Der Kooy, D. (2003) The ablation of glial fibrillary acidic protein-positive cells from the adult central nervous system results in the loss of forebrain neural stem cells but not retinal stem cells, *Eur. J. Neurosci.* **18**, 76–84.

Morshead, C.M., Reynolds, B.A., Craig, C.G., McBurney, M.W., Staines, W.A., Morassutti, D., Weiss, S., Van der Kooy, D. (1994) Neural stem cells in the adult mammalian

forebrain: a relatively quiescent subpopulation of subependymal cells, *Neuron* **13**, 1071–1082.

Mujtaba, T., Piper, D.R., Kalyani, A., Groves, A.K., Lucero, M.T., Rao, M.S. (1999) Lineage-restricted neural precursors can be isolated from both the mouse neural tube and cultured ES cells, *Dev. Biol.* **214**, 113–127.

Nait-Oumesmar, B., Decker, L., Lachapelle, F., Avellana-Adalid, V., Bachelin, C., Van Evercooren, A.B. (1999) Progenitor cells of the adult mouse subventricular zone proliferate, migrate and differentiate into oligodendrocytes after demyelination, *Eur. J. Neurosci.* **11**, 4357–4366.

Noctor, S.C., Martinez-Cerdeno, V., Ivic, L., Kriegstein, A.R. (2004) Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases, *Nat. Neurosci.* **7**, 136–144.

Noctor, S.C., Flint, A.C., Weissman, T.A., Dammerman, R.S., Kriegstein, A.R. (2001) Neurons derived from radial glial cells establish radial units in neocortex, *Nature* **409**, 714–720.

Nottebohm, F. (1985) Neuronal replacement in adulthood, in: Nottebohm, F. (Ed.) *Hope for a New Neurology*, 457 edition. New York Academy of Sciences, New York, pp. 143–161.

Palmer, T.D., Ray, J., Gage, F.H. (1995) FGF-2 responsive neuronal progenitors reside in proliferative and quiescent regions of the adult rodent brain, *Mol. Cell. Neurosci.* **6**, 474–486.

Palmer, T.D., Takahashi, J., Gage, F.H. (1997) The adult rat hippocampus contains primordial neural stem cells, *Mol. Cell. Neurosci.* **8**, 389–404.

Palmer, T.D., Willhoite, A.R., Gage, F.H. (2000) Vascular niche for adult hippocampal neurogenesis, *J. Comp. Neurol.* **425**, 479–494.

Palmer, T.D., Markakis, E.A., Willhoite, A.R., Safar, F., Gage, F.H. (1999) Fibroblast growth factor-2 activates a latent neurogenic program in neural stem cells from diverse regions of the adult CNS, *J. Neurosci.* **19**, 8487–8497.

Parent, J.M., Yu, T.W., Leibowitz, R.T., Geschwind, D.H., Sloviter, R.S., Lowenstein, D.H. (1997) Dentate granule cell neurogenesis is increased by seizures and contributes to aberrant network reorganization in the adult rat hippocampus, *J. Neurosci.* **17**, 3727–3738.

Paton, J.A., Nottebohm, F. (1984) Neurons generated in the adult brain are recruited into functional circuits, *Science* **225**, 1046–1048.

Privat, A. (1977) The ependyma and subependymal layer of the young rat: a new contribution with freeze-fracture, *Neuroscience* **2**, 447–457.

Privat, A., Leblond, C.P. (1972) The subependymal layer and neighboring region in the brain of the young rat, *J. Comp. Neurol.* **146**, 277–302.

Qian, X., Goderie, S.K., Shen, G., Stern, J.H., Temple, S. (1998) Intrinsic programs of patterned cell lineages in isolated vertebrate CNS ventricular zone cells, *Development* **125**, 3143–3152.

Rakic, P. (1971) Guidance of neurons migrating to the fetal monkey neocortex, *Brain Res.* **33**, 471–476.

Rakic, P. (1972) Mode of cell migration to the superficial layers of fetal monkey neocortex, *J. Comp. Neurol.* **145**, 61–84.

Rakic, P. (1981) Neuronal-glial interaction during development, *Trends Neurosci.* **4**, 184–187.

Rakic, P. (1985) Limits of neurogenesis in primates, *Science* **227**, 1054–1056.

Rakic, P. (1988) Specification of cerebral cortical areas, *Science* **241**, 170–176.

Ramón y Cajal, S. (1911) *Histologie du Système Nerveux de l'Homme et des Vertébrés*, Maloine, Paris.

Reynolds, B., Weiss, S. (1992) Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system, *Science* **255**, 1707–1710.

Reynolds, B.A., Weiss, S. (1996) Clonal and population analyses demonstrate that an EGF-responsive mammalian embryonic CNS precursor is a stem cell, *Dev. Biol.* **175**, 1–13.

Richards, L.J., Kilpatrick, T.J., Bartlett, P.F. (1992) De novo generation of neuronal cells from the adult mouse brain, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8591–8595.

Rubenstein, J.L., Anderson, S., Shi, L., Miyashita-Lin, E., Bulfone, A., Hevner, R. (1999) Genetic control of cortical regionalization and connectivity, *Cereb. Cortex* **9**(6), 524–532.

Sanai, N., Tramontin, A.D., Quinones-Hinojosa, A., Barbaro, N.M., Gupta, N., Kunwar, S., Lawton, M.T., McDermott, M.W., Parsa, A.T., Manuel-Garcia Verdugo, J., Berger, M.S., Alvarez-Buylla, A. (2004) Unique astrocyte ribbon in adult human brain contains neural stem cells but lacks chain migration, *Nature* **427**, 740–744.

Santa-Olalla, J., Covarrubias, L. (1999) Basic fibroblast growth factor promotes epidermal

growth factor responsiveness and survival of mesencephalic neural precursor cells, *J. Neurobiol.* **40**, 14–27.

Santarelli, L., Saxe, M., Gross, C., Surget, A., Battaglia, F., Dulawa, S., Weisstaub, N., Lee, J., Duman, R., Arancio, O., Belzung, C., Hen, R. (2003) Requirement of hippocampal neurogenesis for the behavioral effects of antidepressants, *Science* **301**, 805–809.

Schmechel, D.E., Rakic, P. (1979) A Golgi study of radial glia cells in developing monkey telencephalon: morphogenesis and transformation into astrocytes, *Anat. Embryol.* **156**, 115–152.

Seki, T., Arai, Y. (1993) Highly polysialylated neural cell adhesion molecule (NCAM-H) is expressed by newly generated granule cells in the dentate gyrus of the adult rat, *J. Neurosci.* **13**, 2351–2358.

Seri, B., Alvarez-Buylla, A. (2002) Neural stem cells and the regulation of neurogenesis in the adult hippocampus, *Clin. Neurosci. Res.* **2**, 11–16.

Seri, B., Garcia-Verdugo, J.M., McEwen, B.S., Alvarez-Buylla, A. (2001) Astrocytes give rise to new neurons in the adult mammalian hippocampus, *J. Neurosci.* **21**, 7153–7160.

Seri, B., Garcia-Verdugo, J.M., Collado-Morente, L., McEwen, B.S., Alvarez-Buylla, A. (2004) Cell types, lineage, and architecture of the germinal zone in the adult dentate gyrus, *J. Comp. Neurol.* **478**, 359–378.

Smart, I. (1961) The subependymal layer of the mouse brain and its cell production as shown by radioautography after thymidine-H3 injection, *J. Comp. Neurol.* **116**, 325–348.

Smart I.H.M. (1972) Cortical histogenesis and the 'glial coordinate system' of Nieuwenhuys. *J. Comp. Neurol.* **145**.

Smith, M.T., Pencea, V., Wang, Z., Luskin, M.B., Insel, T.R. (2001) Increased number of BrdU-labeled neurons in the rostral migratory stream of the estrous prairie vole, *Horm. Behav.* **39**, 11–21.

Song, H., Stevens, C.F., Gage, F.H. (2002) Astroglia induce neurogenesis from adult neural stem cells, *Nature* **417**, 39–44.

Sotelo, J.R., Trujillo-Cenóz, O. (1958) Electron microscope study on the development of ciliary components of the neural epithelium of the chick embryo, *Z. Zellforsch.* **49**, 1–12.

Spassky, N., Merkle, F.T., Flames, N., Tramontin, A.D., Garcia-Verdugo, J.M., Alvarez-Buylla, A. (2005) Adult ependymal cells are postmitotic and are derived from radial glial cells during embryogenesis, *J. Neurosci.* **25**, 10–18.

Steiner, B., Kronenberg, G., Jessberger, S., Brandt, M.D., Reuter, K., Kempermann, G. (2004) Differential regulation of gliogenesis in the context of adult hippocampal neurogenesis in mice, *Glia* **46**, 41–52.

Stensaas, L.J., Stensass, S.S. (1968) Light microscopy of glial cells in turtles and birds, *Z. Zellforsch.* **91**, 315–340.

Stevenson, J.A., Yoon, M.G. (1982) Morphology of radial glia, ependymal cells, and periventricular neurons in the optic tectum of goldfish (*Carassius auratus*), *J. Comp. Neurol.* **205**, 128–138.

Svendsen, C.N., Caldwell, M.A., Ostenfeld, T. (1999) Human neural stem cells: isolation, expansion and transplantation, *Brain Pathol.* **9**, 499–513.

Svendsen, C.N., Clarke, D.J., Rosser, A.E., Dunnett, S.B. (1996) Survival and differentiation of rat and human epidermal growth factor-responsive precursor cells following grafting into the lesioned adult central nervous system, *Exp. Neurol.* **137**, 376–388.

Szele, F.G., Chesselet, M.F. (1996) Cortical lesions induce an increase in cell number and PSA-NCAM expression in the subventricular zone of adult rats, *J. Comp. Neurol.* **368**, 439–454.

Tamamaki, N., Nakamura, K., Okamoto, K., Kaneko, T. (2001) Radial glia is a progenitor of neocortical neurons in the developing cerebral cortex, *Neurosci. Res.* **41**, 51–60.

Tramontin, A.D., Garcia-Verdugo, J.M., Lim, D.A., Alvarez-Buylla, A. (2003) Postnatal development of radial glia and the ventricular zone (VZ): a continuum of the neural stem cell compartment, *Cereb. Cortex* **13**, 580–587.

Tropepe, V., Sibilia, M., Ciruna, B.G., Rossant, J., Wagner, E.F., van der Kooy, D. (1999) Distinct neural stem cells proliferate in response to EGF and FGF in the developing mouse telencephalon, *Dev. Biol.* **208**, 166–188.

Vescovi, A.L., Reynolds, B.A., Fraser, D.D., Weiss, S. (1993) bFGF regulates the proliferative fate of unipotent (neuronal) and bipotent (neuronal-astroglial) EGF-generated CNS progenitor cells, *Neuron* **11**, 951–966.

Vescovi, A.L., Parati, E.A., Gritti, A., Poulin, P., Ferrario, M., Wanke, E., Frolichsthal-Schoeller, P., Cova, L., Arcellana-Pahlilio, M., Colombo, A., Galli, R. (1999) Isolation and cloning of multipotential stem cells from the embryonic human CNS and establishment of transplantable human neural stem cell lines by epigenetic stimulation, *Exp. Biol.* **156**, 71–83.

Voigt, T. (1989) Development of glial cells in the cerebral wall of ferrets: direct tracing of their transformation from radial glia into astrocytes, *J. Comp. Neurol.* **289**, 74–88.

Walsh, C., Cepko, C.L. (1988) Clonally related cortical cells show several migration patterns, *Science* **241**, 1342–1345.

Weiss, S., Dunne, C., Hewson, J., Wohl, C., Wheatley, M., Peterson, A.C., Reynolds, B.A. (1996) Multipotent CNS stem cells are present in the adult mammalian spinal cord and ventricular neuroaxis, *J. Neurosci.* **16**, 7599–7609.

Zamora, A.J., Mutin, M. (1988) Vimentin and glial fibrillary acidic protein filaments in radial glia of the adult urodele spinal cord, *Neuroscience* **27**, 279–288.

Zupanc, G.K. (2001) Adult neurogenesis and neuronal regeneration in the central nervous system of teleost fish, *Brain Behav. Evol.* **58**, 250–275.

Steroids, Antitumor: *see* Antitumor Steroids

Stem Cells, Adult

Jason H. Pomerantz[1,2] *and Helen M. Blau*[2]
[1] *University of California, San Francisco, CA, USA*
[2] *Stanford University School of Medicine, Stanford, CA, USA*

1 Introduction 364

2 Characteristics of Adult Stem Cells 367

3 Plasticity of Gene Expression 367

4 Conclusion 375

Bibliography 376
Books and Reviews 376
Primary Literature 376

Keywords

Bone Marrow–Derived Cell
Any cell that was present in the bone marrow at the time of extraction before transplant into a new host; includes hematopoietic stem cells, stromal cells, and other cell types.

Cloning
The introduction of a donor nucleus into an enucleated oocyte to produce an embryo.

Heterokaryon
A cell containing multiple distinct nuclei derived from different types of cells.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

Programming
A global change in gene expression profile resulting in a specialized, mature differentiated cell.

Reprogramming
Global change in gene expression profile resulting in a switch from one cell type to another.

Stem cells are responsible for providing a source of differentiated cells in adult tissues that undergo renewal or repair. Considerable progress has been made recently in the identification and characterization of stem cells in adults. They have been the focus of intense interest because of their potential therapeutic use in restoring form and function to damaged tissues. Stem cells in postnatal animals have been identified in specific locales, but another promising and less well-characterized population of stem cells are circulating cells derived from the bone marrow (BMDC). Advances in nuclear transplantation and the observed plasticity of gene expression and phenotype reported in multiple phyla, show that adult cells, including derivatives of HSCs, have a remarkable capacity to be reprogrammed. Irrespective of the mechanism by which reprogramming occurs, fusion or differentiation in response to the microenvironment, the potential for new gene expression patterns by BMDC (bone marrow–derived cell) progeny in recipient tissues holds promise for developing cellular therapies for both proliferative and postmitotic tissues.

1 Introduction

The concept of stem cells continues to evolve and the list of examples of stem cells is growing. Yet, the need for rigorous experimental verification coupled with problems of semantics and long-held dogma make identification of a cell as a stem cell difficult. The necessary functional characteristics of stem cells as well as the physical criteria that endow them with tangible qualities that can be used to detect, isolate, and manipulate them remain elusive. There is no universally agreed upon definition of a stem cell, yet most investigators claim to know one when they see one.

This chapter concerns "adult" stem cells, as distinct from "embryonic" (ES) stem cells. For the purposes of this review, all stem cells existing in postnatal animals are "adult." Like all somatic cells, adult stem cells are presumably derivatives of embryonic stem (ES) cells, although the ontogeny of diverse adult stem cells remains unclear, since to date lineage tracing has not been performed. ES cells studied in tissue culture generally are cell lines selected for their ability to proliferate indefinitely *in vitro*. There is no evidence that such a characteristic bears relevance to the *in vivo* functions of ES cells and it is likely that these cells have acquired properties that differ from their cell of origin in the blastula. Since ES cells can be

made to grow and differentiate *in vitro*, similar expectations are often made of adult stem cells. As is the case with ES cells, such manipulations are unnatural, and place a different set of demands on cells that may have no bearing on their function *in vivo*. A case in point is that although to date cartilage can be produced in a culture dish, it has never been possible to obtain function in joints after implantation in mammals. Therefore, it is necessary to distinguish the biology of ES and adult stem cells as they exist and function in an organism from what they or their progeny can do in a dish.

A conceptual paradox arises regarding one of the most commonly ascribed attributes of a stem cell: the potential for self-renewal. As defined, self-renewal is the ability to yield progeny that are identical to the parent cell. Self-renewal is known to occur at a functional level based on serially transplanted hematopoietic stem cells (HSCs) in mice that continue to repopulate the lineages of the blood. However, the property of self-renewal is more complex than it first appears, as HSCs cannot be serially transplanted indefinitely and HSC function decreases with age. Self-renewal carries the connotation of the ability to divide without limit (as ES cell lines appear to do). However, should self-renewal be a property of stem cells *in vivo*, it must be under strict control. Moreover, a common determinant of stem cells *in vivo* is that they *rarely divide*. In fact, some of the most elegant studies used to track stem cells *in vivo* exploit the property of label retention. A prerequisite for label retention is limited division to avoid dilution. Accordingly, the number of divisions that stem cells undergo in the adult (although more than other somatic cells) is quite limited and has been estimated for a number of tissues to be in the range of 100 to 1000 during the lifetime of the organism. These confounding findings raise the question as to whether self-renewal, or the ability to be cultured extensively in a clonal fashion should be considered a property essential to stem cells that exist naturally *in vivo*. We suggest here that the importance of clonality in stem cell biology is limited to providing experimental material for *in vitro* manipulation in order to work with homogeneous cell populations uncontaminated by other cell types. Its relevance *in vivo* remains unclear; certainly extensive self-renewal without specialization does not appear to be the norm in adult tissues.

Another attribute often ascribed to adult stem cells is their potency, or the number of different types of differentiated progeny that can be produced by a given stem cell. ES cells, by definition, give rise to all the cell types of the body. Adult stem cells differ in that respect. Most "stem cells" found in adult tissues are largely responsible for producing a few different specialized cell types, or possibly only one type (satellite cells in the case of muscle, for example). While this characteristic may reflect intrinsic differences in the ability of cells to express different patterns of genes, an equally plausible explanation is that it may reflect differences among cells in exposure to factors in their *in vivo* microenvironment. These factors influence experimental findings and are responsible for much of the debate surrounding the "nuclear plasticity" or differentiative potential, of cells found in postnatal animals.

All tissues that undergo renewal or repair are associated with relatively less differentiated cells that divide and differentiate to replenish cells lost in response to injury or in the course of aging. The number of different cell types that can arise from any given progenitor may vary, is poorly understood, and is a subject of

debate and interest. In addition, such considerations are essential for defining stem cells versus the more specialized progenitors and then precursors to which they give rise, for example, in hematopoiesis. Consequently, in this chapter we avoid a strict definition of "adult stem cells," as any definition is likely to exclude or underestimate the contribution of cells with a critical role in tissue repair, the essential property of a stem cell *in vivo*. Thus, terminology can hinder progress by providing constraints to the acceptance of novel findings that do not fit a definition intrinsic to a long-held dogma. Such semantic problems have contributed to the confusion that plagues the stem cell field, in general, especially with respect to experiments using bone marrow–derived cells (BMDC) or the subpopulation of BMDC that are HSCs.

Terms such as "plasticity," and "transdifferentiation" have a range of different connotations, and thus clarification based on historical definitions may be useful. The phenotypic conversion of one cell or tissue type into another has been termed "metaplasia," and the literature is replete with examples in humans. "Transdifferentiation," on the other hand, is a specific type of metaplasia involving a change in differentiated phenotype. As classically defined, the criteria for transdifferentiation of a cell are very specific: initially the cell must exhibit a mature, specialized stable phenotype, followed by a change to another specialized stable phenotype, not necessarily involving an intervening cell division. As used here, a "mature" cell refers to a cell type that performs a function, with stable, identifiable characteristics of the specific tissue or organ system to which it contributes and is often postmitotic, for example, neurons, keratinocytes, and cardiomyocytes. On the other hand, "plasticity" is characteristic of all examples of nuclear reprogramming, including cloning, heterokaryons, or metaplasia. Simply put, plasticity means "the ability to change or adapt." It is a descriptive term that relies on context. To minimize confusion, we propose that the term "plasticity" always be qualified and used in a specified context; for example, plasticity of nuclear gene expression, plasticity of cell morphology, plasticity of the differentiated state.

A clear distinction between that which occurs in life and that which we can induce is essential, yet often confused. With increased technological advances, these differences may become more obscure, affecting results and their interpretation. Until recently, frequent and easily discernible fates of cells were reported. Now it is possible to detect extremely rare events. Technological advances have revealed unexpected properties of cells, such as the ability of HSCs or their progeny in adults to incorporate into muscle fibers. The repertoire of individual HSCs (and their progeny) was not supposed to include participation in skeletal muscle and expression of muscle-specific genes. Is this heresy, a rare fluke of nature, or due to the restrictions scientists have imposed on the HSC inherent in its name and definition? Below, we summarize the current state of knowledge regarding stem cells that function in postnatal animals. We then discuss our hypothesis that what will ultimately dictate the medical value of any particular postnatal cell as a stem cell is its potential to be "programmed" or "reprogrammed, which is a property intrinsic to its nuclear chromatin structure. Furthermore, we speculate that the ability to alter this property and remodel chromatin at a molecular level can enhance the plasticity of any given cell and consequently its utility.

2
Characteristics of Adult Stem Cells

Stem cells have been characterized in a number of different adult somatic tissues in which turnover is the norm. Stem cells are not known for cells that are not generated postnatally, such as Purkinje neurons. Moreover, cells of most somatic tissues, other than the blood, do not circulate and remain relatively localized within one region of the tissue throughout the lifetime of the cell. Accordingly, somatic stem cells have been found within the tissues to which they are known to contribute in adulthood (see Table 1). The concept that the circulation can be an alternate route for somatic stem cells to reach and contribute to injured target tissues has arisen as a possibility in the past few years as markers of blood (GFP (green fluorescent protein), Y chromosome, and β-galactosidase) and methods for definitively resolving single cells expressing more than one marker (laser confocal microscopy) have come available. Transit of non-HSCs via the circulation would require functions in addition to those known to be performed by localized stem cells (extravasation and homing) heretofore thought to be limited to hematopoietic cells. To be convincing, physical incorporation into tissues must be coupled with novel patterns of gene expression. This process may be termed *reprogramming*, or more aptly in the case of stem cells, "programming" since a stem cell should not need to be *re*programmed to do something that is part of its natural function. One concern is whether the programming or reprogramming that occurs attains the frequency required to affect the function of the recipient tissue. This second requirement is key for any natural property of stem cells in living animals. However, rare events if real do not preclude use of a particular cell therapeutically for tissue regeneration. A clear understanding of the biology underlying such naturally occurring events may allow manipulation of cells and augmentation of their ability to perform new functions.

The designation of a cell as a "stem cell" has been limited to cells residing within proliferation-competent tissues that give rise to progeny that exhibit tissue-specific morphology and function. By marked contrast to ES cells, cultured cells from adults display a much lower propensity to differentiate into a broad range of cell types, and primary cultures are often difficult to establish. In most cases, adult stem cells have exhibited oligo or unipotent abilities (see Table 1 for known examples and their unique characteristics). The characterization of stem cells in adults is very much a work in progress. A detailed account of each of the cell types currently designated as "stem cell" in adult tissues is beyond the scope of this text and the reader is referred to a number of excellent reviews and primary articles. Below we focus on one central principle of stem cell function: the ability to induce the nucleus to express a specific repertoire of genes, thereby modifying cell identity in order to maintain, replace, or rescue a particular tissue.

3
Plasticity of Gene Expression

The recently discovered capacity of adult BMDCs to contribute to nonhematopoietic tissues suggests that they may reprogram their nuclei for functions other than blood. Nuclear transplantation (cloning) and cell fusion in nondividing heterokaryons and subsequently cloning, provide historical evidence for plasticity of mammalian gene expression. These

Tab. 1 Best-characterized stem cells found in adult tissues. For each tissue listed, stem cells have been located in specific anatomic sites with reasonable certainty and stem cell function has been demonstrated *in vivo* without intervening cell culture. To this date, only the hematopoietic stem cell can be prospectively isolated at the single-cell level based on phenotypic characteristics, and shown to function *in vivo*.

Tissue	*Location/niche*	*Defining characteristics in vivo*	*Surface markers defined for prospective isolation?*	*Differentiated progeny*
Skin	Follicle bulge, basal epidermis?	Label retaining, repopulate skin lineages after transplant	No	Hair follicle, sebaceous gland, epidermis
Blood	Bone marrow and peripheral blood	Repopulate blood lineages after transplant	Yes	Lymphoid, myeloid, erythroid cells
Skeletal muscle	Muscle fiber periphery between sarcolemma and basal lamina	Distinct anatomic location, fuse with and generate new myofibers after transplant	No	Skeletal myofibers
CNS	Periventricular zone	Label retaining, migrate to olfactory bulb or hippocampus	No	Neurons, astrocytes, oligodendrocytes
Gut epithelia	Basal region of intestinal crypts	Single cells regenerate entire crypts after irradiation	No	Enterocytes, goblet cells, paneth cells, enteroendocrine cells
Melanocyte	Lower permanent portion of the hair follicle	Label retaining, distinct anatomic location, regenerate pigmentation after transplant	No	Melanocyte
Skeletal	Marrow	Formation of a complete bone organ after transplant	No	Chondrocytes, osteocytes, adipocytes
Testis	Seminiferous tubule basement membrane associated with Sertoli cells	Restore sperm production after transplant	No	Spermatogonia

fundamental observations provide clear evidence of nuclear plasticity and serve as a starting point for understanding this developing field and its potential applications to regenerative medicine. Examples from numerous publications reveal that the potential exists at the genetic, cellular, and organismal levels for cells to assume new functions: gene expression patterns can change, differentiation can be manipulated, and cells can incorporate into tissues from which they did not originate. Consideration of the evidence leading to this conclusion describes the current state of the field. A consideration of the progress in the understanding of molecular mechanisms for remodeling chromatin may determine whether there are insurmountable blocks to using endogenous genetic or cellular complements (directly or after *ex vivo* manipulation) to repair the body. The goal of this chapter is to draw attention to the fundamental concept that plasticity of gene expression (required of stem cells) is widespread in mammalian cells, and provide the reader with evidence that this idea is supported by decades of experimentation.

The most direct way to investigate the capacity of a nucleus to express a new set of genes is to expose it to new intracellular factors and observe how it responds. Nuclear cloning, the introduction of a donor cell nucleus into an enucleated oocyte to produce an embryo, was first studied in amphibians. The results showed for the first time that differentiation and development do not require loss or irreversible inactivation of genes. The disparate interpretations of the early results are instructive *vis-a-vis* the current controversy surrounding the behavior of BMDC, HSC, and their derivatives in different settings. Briggs and King first reported that transplantation of nuclei from frog blastulae into enucleated oocytes yielded swimming tadpoles. They observed that irrespective of developmental stage, donor nuclei from a given lineage typically gave rise to normal tissues of the same lineage, but to abnormal tissues from other lineages. These experiments led them to conclude that differentiated donor nuclei are restricted in their developmental potential. Diberardino demonstrated that differences in cloning efficiency were, in part, related to the cell cycle of the donor nucleus, which had to switch to the rapid cycle typical of the egg. Nuclei derived from cells that were further along in development lost totipotency, presumably because they were more differentiated, had slower cell cycles and were therefore more difficult to activate. In other nuclear transplant experiments, Gurdon showed that when tadpoles were used as the source, some normal adult frogs could be produced using nuclei derived from differentiated intestinal endoderm. By contrast with Briggs and King, Gurdon heralded these findings as strong evidence for the plasticity of gene expression, and concluded that genes are not permanently inactivated, but instead can be reawakened, even in differentiated cells. Some caveats remained. Attempts to generate adult frogs from nuclei derived from adult animals were unsuccessful. King favored the interpretation that reprogramming of differentiated nuclei was limited, and that Gurdon's results likely represented contaminated transplants that included less differentiated cells. Nonetheless, Gurdon's bold interpretation stimulated thinking in the field of differentiation. Remarkably, confirmation of these results in mammals was not forthcoming for decades.

Despite the experiments in amphibians, 20 years later their relevance to mechanisms of differentiation in mammalian

cells had still not gained acceptance. Experimentally induced fusion of different cell types using viruses or the chemical polyethylene glycol became a powerful method to study gene expression changes after exposure of nuclei to new cytoplasm. Studies using synkaryons (cell hybrids that proliferate and exhibit nuclear fusion) first showed the existence of trans-acting repressors and activators, but gene activation was only transient because of cell division and the consequent loss of chromosomes. These studies were extended by fusion of cultured cells to form multinucleate, stable heterokaryons, in which each nucleus remained distinct and intact. Such nondividing cell fusion products allowed changes in gene expression to be monitored over time. Gene expression patterns depended on the nuclear ratio or relative gene dosage, and the consequent balance of factors present in the cytoplasm at any given time, showing that differentiation is dynamic and continuously regulated.

Clear evidence of reprogramming adult somatic nuclei evolved as investigators learned to manipulate heterokaryons. In early work with heterokaryons, Ringertz showed that fusion of rat myoblasts and chick erythrocytes caused the nuclei of the erythrocytes to swell and chromatin to become diffuse, but reprogramming was not observed. When transformed lines of mouse muscle cells were fused in tissue culture with rat cells in high serum medium that did not favor differentiation, gene repression, not activation was observed. However, when stable heterokaryons were produced in low serum medium in the presence of inhibitors of cell division, fused diploid mouse muscle cells and nontransformed human diploid cells derived from all three embryonic lineages, endoderm (hepatocytes), ectoderm (keratinocytes), and mesoderm (fibroblasts), were shown to activate a number of previously silent genes (muscle) for the first time. These results indicated that the differentiated state could be altered, even in specialized cells derived from human adults. Critical to these findings were the cell type used (nontransformed cells), the medium (which favored differentiation not proliferation), and a means of preventing mitosis and unequal chromosome segregation. Similar results were also obtained using similar methods with other differentiated cell types. Together, these *in vitro* heterokaryon studies established that the expression of genes that had been previously "silenced" during development and differentiation could be reactivated by cytoplasmic factors present in somatic cells, without the passage of nuclei through the oocyte or exposure to sequential factors characteristic of embryogenesis that was necessary for amphibian cloning. The heterokaryon studies first conclusively demonstrated that the differentiated state in mammals was not terminal, but could be altered.

The cloning of sheep nearly 15 years later confirmed the findings in heterokaryons and cloned amphibians. Micromanipulation of the much smaller, more delicate mammalian oocyte was pivotal to the success of nuclear transplantation in mammals. Techniques for injecting oocytes, rather than zygotes, and other conditions for cloning mammals were progressively improved in the 1980s and 1990s culminating in 1997 with the successful cloning of Dolly, a sheep derived from the transplantation of an adult mammary epithelial cell nucleus into a mammalian oocyte. Mammalian cloning has now been replicated in a number of species, including humans, due to a progression of technological advances.

The low efficiency of mammalian cloning (less than 1% of all nuclear transfers produce live offspring) left open the question of the cell of origin of successful clones, with the possibility that tissue-specific stem cells were the source of the effective nuclei. That concern was addressed first in cloning experiments using mature B and T lymphocytes, and has recently been laid to rest with the generation of live mice from unambiguous, fully differentiated, postmitotic olfactory neurons. These experiments leave no doubt that even "terminally" differentiated cells in mammals have nuclei that retain the remarkable capacity to be reprogrammed to totipotency and generate all the cell types necessary to form an intact animal.

From a practical standpoint, the efficiency of mammalian cloning has not improved since the first cloned mammals were reported. In addition, surviving cloned animals often continue to have a high incidence of defects, for example, large offspring, placental problems, and arthritis. These abnormal phenotypes are likely to be the result of epigenetic rather than genetic abnormalities, because they are not passed on to the offspring of affected cloned animals. Accordingly, a major effort is being directed at identifying the factors that are responsible for promoting the orderly silencing and reactivation of genes.

In summary, we now know that the cytoplasm has the capacity to direct reprogramming, and that at least in oocytes, reprogramming to a totipotent state can occur. Therapeutic cloning via the derivation of ES cells derived from somatic nuclear transfer into oocytes has already proven to be a viable approach to cell therapy with the cure of RAG 2 deficient mice. Oocytes and sperm have also been produced from existing ES cell lines, potentially bypassing the need to repeatedly obtain oocytes from live humans for the purpose of cloning, and associated ethical problems. Approximately half a century has passed since the pioneering work of Briggs and King, and Gurdon, and although the practical limitations of low efficiency, high cost, and an incomplete understanding of reprogramming mechanisms must be overcome, it seems likely that therapeutic cloning will play a role in regenerative medicine.

Evidence that plasticity of gene expression (nuclear reprogramming) occurs naturally *in vivo* is important, since both cloning and PEG-induced heterokaryon formation are examples of extreme, experimental manipulations. Examples of *tissue* metaplasia in humans include bone formation in scars and muscle, as well as conversion of one type of epithelium to another in the respiratory tract, the urinary bladder, and in the esophagus (Barrett's metaplasia). Metaplasia can also be induced experimentally. For example, pancreas can be converted to liver *in vitro* and *in vivo* in response to diverse factors In Drosophila, conversion of one imaginal disc to another in response to changes in homeotic gene expression is another example of metaplasia. In urodeles, a differentiated limb muscle cell can dedifferentiate and yield muscle, cartilage, and connective tissue progeny, and seminal studies of newt lens regeneration provide the most well-known examples of transdifferentiation. Like urodeles, zebrafish also have a marked capacity for regeneration, but as in humans, rigorous documentation of cellular metaplasia in zebrafish by tracking labeled cells, has yet to be reported. Finally, in axolotls, individually labeled spinal cord glial cells can dedifferentiate after tail amputation, and give rise to mature mesodermal (and neural) progeny, a definitive germ layer switch. Thus, in

response to injury, cells in differentiated tissues exist that can change and exhibit different phenotypes in lower animals as well as mammals, and the mechanisms may be evolutionarily conserved.

Notably, in humans metaplasia usually occurs in response to pathological conditions (e.g. gastroesophageal reflux in Barrett's esophagus). Indeed, it is unclear whether there are examples, in humans, of metaplasia as a component of normal development, and the cells responsible for tissue metaplasia in mammals have not been defined. It is possible that reserve cells for each different cell type are present as distinct entities in metaplastic tissues. By contrast, in mammalian lung, electron microscopy has revealed that morphological characteristics and organelles of two different cell types can appear in one cell during metaplastic conversion suggesting that in some cases multiple gene expression patterns are available to a particular cell in response to environmental changes. Taken together, the many examples of metaplasia are a testament to the intrinsic ability of cells to alter significant patterns of the genes they express in response to their *in vivo* environment.

The mammalian examples of metaplasia described above do not involve transcendence of embryonic lineages or apparent transit of cells via the circulation. Intense investigation in mammals has recently been directed at seeking evidence of metaplasia across embryonic lineages both for fundamental interest and potential therapeutic benefit. This research has primarily focused on BMDC in adults, which have efficient access to all tissues of the body via the circulation.

When the first report appeared demonstrating that bone marrow cells could contribute to muscle in mice, it was unclear whether this was a sporadic event or of fundamental physiologic significance. Using bone marrow from a transgenic animal expressing the myosin light-chain enhancer driving β-galactosidase (B-gal), muscle fibers expressing the reporter gene were found in transplant recipients. There followed a number of reports of similar findings, using marrow cells marked genetically with B-gal, Y chromosome, and GFP. Such cells were reported to be present in the brain, liver, heart, skeletal muscle, and epithelia of the kidney, lung, and skin. The incidence was generally low (0.01 to 0.1% of total cells) and many of the initial observations were not reproduced. Although semantics seems to confound results and their interpretation, there is justifiable concern about the validity of reported findings and each should be met with appropriate skepticism, as well as rigorous examination of the stringency of the methodologies used and accuracy of the conclusions drawn. To be conclusive, experimental findings must be replicated by others.

The emerging picture is that the contribution of BMDC to nonhematopoietic tissues is a rare but real event under normal physiological conditions. This contribution can be significantly increased in response to tissue stress or damage, for example, in liver In another case, damage to muscle either by local toxin injection or exercise resulted in a markedly higher BMDC contribution, and it may be possible to increase this frequency further. While recognizing that other excellent examples exist, we focus below on two that are illustrative: skeletal muscle and Purkinje neurons, which have been extensively studied by our group.

The identification of one cell type within bone marrow with the capacity to contribute to muscle was a critical finding. Since the bone marrow contains "stromal"

cells with the ability to differentiate into various tissues, it was not clear whether stromal cells were responsible for the BMDC contributions to liver and skeletal muscle following bone marrow transplant. Single-cell transplants have now shown definitively that ckit+, sca-1+, lin-HSC progeny not only reconstitute all lineages of the blood, but also incorporate into muscle in lethally irradiated mice. Whether additional cells in marrow share this capacity and can transit naturally to other tissues via the circulation remains to be determined.

BMDC also contribute to a population of cells that are unique to the cerebellum. In both mice and humans, BMT-derived nuclei can be found in Purkinje neurons. In mice, this is a low frequency event that increases over time after BMT, and most such Purkinje neurons exist as binucleate heterokaryons. The use of transgenic mice harboring the Purkinje specific promoter, L7, showed that BMDC nuclei that are present in heterokaryons exhibit dispersed chromatin and express the L7 promoter. Heterokaryon formation also occurs in damaged multinucleated muscle fibers. These findings suggest an intriguing mechanism by which complex, postmitotic structures may incorporate genomic material from endogenous cells in adulthood.

The contribution of bone marrow to other tissues appears to represent an additional example of metaplasia in mammals. However, available data do not address the question of whether BMDC contribution to other tissues is functionally significant in normal life. From a therapeutic standpoint, proof of principle has already been achieved with the survival of tyrosinemic fumarylacetoacetate hydrolase (FAH) mice following HSC transplantation. Thus, the degree to which this phenomenon can be manipulated for therapeutic goals is an important line of investigation.

Two different mechanisms have been proposed whereby BMDC contribute to nonhematopoietic tissues (see Fig. 1). Either mechanism ultimately involves nuclear reprogramming. Initial reports of "plasticity" in adult stem cells suggested that a developmentally immature BMDC could alter its typical course of differentiation to that of nonhematopoietic tissues. Therefore, BMDC could function as traditional stem cells for other tissues. This first mechanism implies a response to extracellular signals that are detected by the BMDC and result in an alteration of gene expression and differentiation along an alternative developmental path. Such a process could, in theory, result in mononuclear cells, either directly or via a tissue stem cell intermediate. Examples that have been reported include BMDC-derived kidney epithelium, pulmonary epithelium, and pancreatic islets. Three groups have also reported BMDC in the satellite cell position, underneath the basal lamina, expressing muscle stem cell markers Although skeletal muscle may be the most amenable system for discerning stem cell intermediates, it also has inherent complexities, as the end state via stem cell or direct fusion is the same: a multinucleated fiber. The strongest evidence to date that BMDC can be reprogrammed as mononuclear cells has recently been shown in pulmonary epithelium using the Cre/Lox system to exclude fusion.

As evidence for the second mechanism, fusion, the heterokaryons formed between BMDC and muscle or BMDC and Purkinje neurons are perhaps the clearest examples. In the case of fusion, reprogramming is achieved by cytoplasmic factors that may

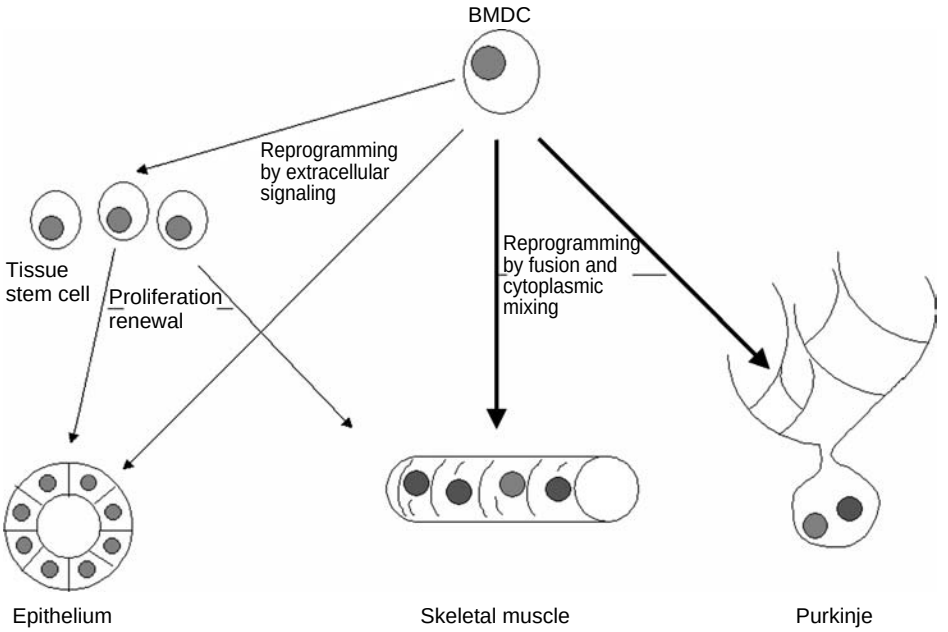

Fig. 1 Schematic representation of the possible mechanisms of BMDC incorporation into nonhematopoietic tissues. As discussed in the text, not all mechanisms are well established *in vivo*, and each may occur to a different extent or not at all in different tissues. Thick arrows indicate mechanisms that are better supported in the current literature. First, nuclear reprogramming can occur in cells in response to extracellular signals to yield a mature differentiated cell; second, nuclear reprogramming may first result in differentiation into a tissue-specific stem cell that has the ability to generate mature, fully differentiated cells; third, nuclear reprogramming is achieved when the nucleus of the donor cell is exposed to intracellular signals within a mixed cytoplasm following cell fusion. Fusion of BMDC directly with recipient cells is most clearly shown in Purkinje neuron x BMDC heterokaryons, but most probably also occurs in skeletal muscle that exists naturally as a syncytium.

be similar to those associated with reprogramming during cloning, or in the nuclei of heterokaryons formed *in vitro*. Initial reports of contribution of BMDC to nonhematopoietic tissues by fusion described the phenomenon as "illegitimate," "mere," or "random." However, time has granted a new respect for fusion. It is becoming clear that fusion of BMDC with other tissues is a specific process limited to particular donor and recipient cells. Moreover, it occurs naturally *in vivo*. A number of molecules have been implicated in cell–cell fusion in eukaryotic species, some of which may play a role in BMDC fusion. It will be interesting to determine whether fusion mechanisms are evolutionarily conserved and if cells that are involved in BMDC contribution to other tissues fuse by similar mechanisms.

In mammals, why fusion has remained a property of some cells except in the case of muscle or sperm and egg, is unclear. We propose that the observed fusion of BMDC with cells of other tissues may represent a way for cells from different parts of the body with diverse histories to contribute to one another.

A plausible significant impact from such a cellular interaction is demonstrated by rescue of the FAH mouse. And while other models are currently lacking, one could envision situations in which a cell suffers a deleterious loss-of-function mutation and is rescued by expression of a reprogrammed, wild-type copy present in a donated nucleus. If such scenarios occur, a provocative and as yet unanswered question arises: if a BMDC exists that can self-renew and yield progeny such as myeloid cells amenable to fusion, reprogramming, and rescue of a defunct Purkinje nucleus, is it functioning as a Purkinje stem cell?

As the factors involved in the control of gene expression at the molecular level are identified, approaches may emerge for increasing the efficiency of programming and reprogramming. Chromatin remodeling by histone acetylation or deacetylation, methylation of histones or DNA, and other mechanisms such as SWI/SNF comprise the basic mechanisms of gene activation and repression. Pharmacological or genetic interventions that target such remodeling factors may prove useful for relaxing chromatin in order to direct gene expression toward desired end points.

4 Conclusion

The previous half-century has provided the foundation and paved the way for stem cell biology and therapy. We have learned that (1) genetic material is generally not lost during development and differentiation, (2) the cytoplasm of the oocyte as well as somatic cells has the ability to reprogram gene expression, (3) most genes, even in terminally differentiated cells, can be reactivated, (4) a single somatic nucleus has the replicative capacity to yield enough tissue for a whole new organism, and (5) in life, cells change their phenotypes, and in some cases, nuclear gene expression is reprogrammed across lineages. These findings indicate that the theoretical roadblocks based on long-held dogma and old paradigms that might have precluded the use of cells from one source to contribute to another have been overcome.

Nonetheless, much of what we have observed and learned comes from unnatural biological perturbations. Even the incorporation of BMDC into nonhematopoietic tissues is not only extremely rare under most conditions, but experiments necessarily depart from natural circumstances. Furthermore, the efficiency with which incorporation occurs remains low and will not be of clinical interest until the incidence can be increased. Thus, the practical utility of BMDC and their advantages and disadvantages will become clear relative to ES and other stem cells with further experimentation. Another unresolved question regarding the current understanding of *in vivo* contribution of bone marrow to other tissues involves the potential for proliferation of reprogrammed cells. The different mechanisms by which BMDC may contribute to nonhematopoietic tissues are each potentially useful, but have different implications. Differentiation via a stem cell intermediate implies the ability to expand reprogrammed cells, as does fusion followed by reduction division. By contrast, direct fusion with postmitotic cells that cannot proliferate but form stable heterokaryons could serve as an alternative means of tissue repair. Fusion to form heterokaryons presents the intriguing possibility that a structurally complex, nondividing cell could

be rescued by a cell from another tissue. Nuclear donation and subsequent reprogramming may already constitute an elegant solution used by nature in humans and mice, and one that could be capitalized upon for therapeutic purposes. In one potential application relevant to either mechanism, BMDC could be engineered to serve as cellular delivery vehicles for therapeutic genes to nonhematopoietic tissues.

A distinction between the potential mechanisms required for nuclear reprogramming is important to consider when drawing parallels between cloning and heterokaryons formed *in vivo*. Nondividing heterokaryons observed *in vivo* are not subject to many of the demands that cloned nuclei are. In addition, the more limited gene expression patterns required by *in vivo* reprogramming (BMDC only need to express genes of one different cell type) may have different mechanistic requirements. Whereas the gold standard test for reprogramming in nuclear transplantation is the generation of an animal, the requirements for reprogramming BMDC nuclei are less extensive if not less stringent.

If the existence of technological hurdles and questions of degree or "low frequency" had daunted researchers of the last half-century, then cloning would not have persisted beyond a dream or the subject of science fiction. Perhaps the progress of cloning research, leading to therapeutic (not reproductive) cloning may predict the future with respect to the field of nuclear reprogramming and its applications to regenerative medicine. Like cloning, we should approach the biology and technology of adult stem cells with vigor, rigor, and persistence.

Bibliography

Books and Reviews

Alonso, L., Fuchs, E. (2003) Stem cells of the skin epithelium, *Proc. Natl. Acad. Sci. U.S.A.* **100**(Suppl. 1), 11830–11835.

Blau, H.M. (1992) Differentiation requires continuous active control, *Annu. Rev. Biochem.* **61**, 1213–1230.

Blau, H.M., Brazelton, T.R., Weimann, J.M. (2001) The evolving concept of a stem cell: entity or function? *Cell* **105**(7), 829–841.

Gage, F.H. (2000) Mammalian neural stem cells, *Science* **287**(5457), 1433–1438.

Goldring, K., Partridge, T., Watt, D. (2002) Muscle stem cells, *J. Pathol.* **197**(4), 457–467.

Gurdon, J.B., Byrne, J.A. (2003) The first half-century of nuclear transplantation, *Proc. Natl. Acad. Sci. U.S.A.* **100**(14), 8048–8052.

Hadorn, E. (1976) in: Ashburner, M. (Ed.) *The Genetics and Biology of Drosophila*, Academic Press, San Diego, pp. 556–617.

Lanza, R., Blau, H., Melton, D., Moore, M., Thomas, E.D., Verfaille, C., Weissman, I., West, M. (2004) *Handbook of stem cells*, Vol. 2, 1st edition, Elsevier, San Diego, p. 820.

Li, E. (2002) Chromatin modification and epigenetic reprogramming in mammalian development, *Nat. Rev. Genet.* **3**(9), 662–673.

Ringertz, Na.S., R., (1976) Heterokaryons. *Cell Hybrids*, Academic Press, New York, pp. 87–118.

Primary Literature

Alvarez-Dolado, M., et al. (2003) Fusion of bone-marrow-derived cells with Purkinje neurons, cardiomyocytes and hepatocytes, *Nature* **425**(6961), 968–973.

Anderson, D.J., Gage, F.H., Weissman, I.L. (2001) Can stem cells cross lineage boundaries? *Nat. Med.* **7**(4), 393–395.

Baron, M.H., Maniatis, T. (1986) Rapid reprogramming of globin gene expression in transient heterokaryons, *Cell* **46**(4), 591–602.

Beresford, W.A. (1990) Direct transdifferentiation: can cells change their phenotype without dividing? *Cell Differ. Dev.* **29**(2), 81–93.

Blau, H.M., et al. (1985) Muscle gene expression in heterokaryons, *Adv. Exp. Med. Biol.* **182**, 231–247.

Blau, H.M. (1989) How fixed is the differentiated state? Lessons from heterokaryons, *Trends Genet.* **5**(8), 268–272.

Blau, H.M. (2002) A twist of fate, *Nature* **419**(6906), 437.

Blau, H.M., Baltimore, D. (1991) Differentiation requires continuous regulation, *J. Cell Biol.* **112**(5), 781–783.

Blau, H.M., Chiu, C.P., Webster, C. (1983) Cytoplasmic activation of human nuclear genes in stable heterocaryons, *Cell* **32**(4), 1171–1180.

Boecker, W., Buerger, H. (2003) Evidence of progenitor cells of glandular and myoepithelial cell lineages in the human adult female breast epithelium: a new progenitor (adult stem) cell concept, *Cell Prolif.* **36**(Suppl. 1), 73–84.

Braun, K.M., Watt, F.M. (2004) Epidermal label-retaining cells: background and recent applications, *J. Investig. Dermatol. Symp. Proc.* **9**(3), 196–201.

Brazelton, T.R., et al. (2000) From marrow to brain: expression of neuronal phenotypes in adult mice, *Science* **290**(5497), 1775–1779.

Briggs, R., King, T. (1952) Transplantation of living nuclei from Blastula cells into enucleated frogs' eggs, *Proc. Natl. Acad. Sci. U.S.A.* **38**, 455–463.

Briggs, R., King, T.J. (1959) Nucleocytoplasmic interactions in eggs and embryos, in: Mirsky, A.E., Brachet J., (Ed.) *The Cell*, Academic Press, New York, pp. 537–617.

Brinster, R.L. (2002) Germline stem cell transplantation and transgenesis, *Science* **296**(5576), 2174–2176.

Brockes, J.P., Kumar, A. (2002) Plasticity and reprogramming of differentiated cells in amphibian regeneration, *Nat. Rev. Mol. Cell. Biol.* **3**(8), 566–574.

Camargo, F.D., et al. (2003) Single hematopoietic stem cells generate skeletal muscle through myeloid intermediates, *Nat. Med.* **9**(12), 1520–1527.

Camargo, F.D., Finegold, M., Goodell, M.A. (2004) Hematopoietic myelomonocytic cells are the major source of hepatocyte fusion partners, *J. Clin. Invest.* **113**(9), 1266–1270.

Charge, S., Rudnicki, M.A. (2003) Fusion with the fused: a new role for interleukin-4 in the building of muscle, *Cell* **113**(4), 422–423.

Cheong, H.T., Takahashi, Y., Kanagawa, H. (1993) Birth of mice after transplantation of early cell-cycle-stage embryonic nuclei into enucleated oocytes, *Biol. Reprod.* **48**(5), 958–963.

Chiu, C.P., Blau, H.M. (1984) Reprogramming cell differentiation in the absence of DNA synthesis, *Cell* **37**(3), 879–887.

Chiu, C.P., Blau, H.M. (1985) 5-Azacytidine permits gene activation in a previously noninducible cell type, *Cell* **40**(2), 417–424.

Corbel, S.Y., et al. (2003) Contribution of hematopoietic stem cells to skeletal muscle, *Nat. Med.* **9**(12), 1528–1532.

Cotran, C., Kumar, V., Tucker, C. (1999) *Pathologic Basis of Disease*, 6th edition, Philadelphia: WB Saunders, p. 1425.

Dabeva, M.D., et al. (1997) Differentiation of pancreatic epithelial progenitor cells into hepatocytes following transplantation into rat liver, *Proc. Natl. Acad. Sci. U.S.A.* **94**(14), 7356–7361.

Davidson, R.L. (1974) Gene expression in somatic cell hybrids, *Annu. Rev. Genet.* **8**, 195–218.

Di Berardino, M.A., King, T.J. (1967) Development and cellular differentiation of neural nuclear-transplants of known karyotype, *Dev. Biol.* **15**(2), 102–128.

Dontu, G., et al. (2003) Stem cells in normal breast development and breast cancer, *Cell Prolif.* **36**(Suppl. 1), 59–72.

Doyonnas, R., LaBarge, M.A., Sacco, A., Charlton, C., Blau, H.M. (2004) Hematopoietic contribution to skeletal muscle by myelomonocytic precursors, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 13507–13512.

Dreyfus, P.A., et al. (2004) Adult bone marrow-derived stem cells in muscle connective tissue and satellite cell niches, *Am. J. Pathol.* **164**(3), 773–779.

Echeverri, K., Tanaka, E.M. (2002) Ectoderm to mesoderm lineage switching during axolotl tail regeneration, *Science* **298**(5600), 1993–1996.

Eggan, K., et al. (2004) Mice cloned from olfactory sensory neurons, *Nature* **428**(6978), 44–49.

Eguchi, G., Kodama, R. (1993) Transdifferentiation, *Curr. Opin. Cell Biol.* **5**(6), 1023–1028.

Ephrussi, B., et al. (1969) Malignancy of somatic cell hybrids, *Nature* **224**(226), 1314–1316.

Ferrari, G., et al. (1998) Muscle regeneration by bone marrow-derived myogenic progenitors, *Science* **279**(5356), 1528–1530.

Fukada, S., et al. (2002) Muscle regeneration by reconstitution with bone marrow or fetal

liver cells from green fluorescent protein-gene transgenic mice, *J. Cell Sci.* **115**(Pt 6), 1285–1293.

Grove, J.E., et al. (2002) Marrow-derived cells as vehicles for delivery of gene therapy to pulmonary epithelium, *Am. J. Respir. Cell Mol. Biol.* **27**(6), 645–651.

Gurdon, J.B. (1962) Adult frogs derived from the nuclei of single somatic cells, *Dev. Biol.* **4**, 256–273.

Gussoni, E., et al. (1999) Dystrophin expression in the mdx mouse restored by stem cell transplantation, *Nature* **401**(6751), 390–394.

Gurdon, J.B., Uehlinger, V. (1966) "Fertile" intestine nuclei, *Nature* **210**(42), 1240–1241.

Harris, R.G., et al. (2004) Lack of a fusion requirement for development of bone marrow-derived epithelia, *Science* **305**(5680), 90–93.

Harris, H., Wiener, F., Klein, G. (1971) The analysis of malignancy by cell fusion 3. Hybrids between diploid fibroblasts and other tumour cells, *J. Cell Sci.* **8**(3), 681–692.

Hochedlinger, K., Jaenisch, R. (2002a) Monoclonal mice generated by nuclear transfer from mature B and T donor cells, *Nature* **415**(6875), 1035–1038.

Hochedlinger, K., Jaenisch, R. (2002b) Nuclear transplantation: lessons from frogs and mice, *Curr. Opin. Cell Biol.* **14**(6), 741–748.

Hubner, K., et al. (2003) Derivation of oocytes from mouse embryonic stem cells, *Science* **300**(5623), 1251–1256.

Hwang, W.S., et al. (2004) Evidence of a pluripotent human embryonic stem cell line derived from a cloned blastocyst, *Science* **303**(5664), 1669–1674.

Ianus, A., et al. (2003) In vivo derivation of glucose-competent pancreatic endocrine cells from bone marrow without evidence of cell fusion, *J. Clin. Invest.* **111**(6), 843–850.

Jackson, K.A., Mi, T., Goodell, M.A. (1999) Hematopoietic potential of stem cells isolated from murine skeletal muscle, *Proc. Natl. Acad. Sci. U.S.A.* **96**(25), 14482–14486.

Johansson, C.B. (2003) Mechanism of stem cells in the central nervous system, *J. Cell. Physiol.* **196**(3), 409–418.

Kale, S., et al. (2003) Bone marrow stem cells contribute to repair of the ischemically injured renal tubule, *J. Clin. Invest.* **112**(1), 42–49.

Kondo, M., et al. (2003) Biology of hematopoietic stem cells and progenitors: implications for clinical application, *Annu. Rev. Immunol.* **21**, 759–806.

Korbling, M., Estrov, Z. (2003) Adult stem cells for tissue repair – a new therapeutic concept? *N. Engl. J. Med.* **349**(6), 570–582.

Krakowski, M.L., et al. (1999) Pancreatic expression of keratinocyte growth factor leads to differentiation of islet hepatocytes and proliferation of duct cells, *Am. J. Pathol.* **154**(3), 683–691.

Krause, D.S., et al. (2001) Multi-organ, multi-lineage engraftment by a single bone marrow-derived stem cell, *Cell* **105**(3), 369–377.

LaBarge, M.A., Blau, H.M. (2002) Biological progression from adult bone marrow to mononucleate muscle stem cell to multinucleate muscle fiber in response to injury, *Cell* **111**(4), 589–601.

Lagasse, E., et al. (2000) Purified hematopoietic stem cells can differentiate into hepatocytes in vivo, *Nat. Med.* **6**(11), 1229–1234.

Lavker, R.M., et al. (2003) Hair follicle stem cells, *J. Investig. Dermatol. Symp. Proc.* **8**(1), 28–38.

Lawrence, P.A., Morata, G. (1983) The elements of the bithorax complex, *Cell* **35**(3 Pt 2), 595–601.

Lo, D.C., Allen, F., Brockes, J.P. (1993) Reversal of muscle differentiation during urodele limb regeneration, *Proc. Natl. Acad. Sci. U.S.A.* **90**(15), 7230–7234.

Marshman, E., Booth, C., Potten, C.S. (2002) The intestinal epithelial stem cell, *BioEssays* **24**(1), 91–98.

Mauro, A. (1961) Satellite cell of skeletal muscle fibers, *J. Biophys. Biochem. Cytol.* **9**, 493–495.

Meng, L., et al. (1997) Rhesus monkeys produced by nuclear transfer, *Biol. Reprod.* **57**(2), 454–459.

Mezey, E., et al. (2000) Turning blood into brain: cells bearing neuronal antigens generated in vivo from bone marrow, *Science* **290**(5497), 1779–1782.

Mombaerts, P. (2003) Therapeutic cloning in the mouse, *Proc. Natl. Acad. Sci. U.S.A.* **100**(Suppl. 1), 11924–11925.

Morrison, S.J., Weissman, I.L. (1994) The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype, *Immunity* **1**(8), 661–673.

Murtaugh, L.C., Melton, D.A. (2003) Genes, signals, and lineages in pancreas development, *Annu. Rev. Cell Dev. Biol.* **19**, 71–89.

Nishimura, E.K., et al. (2002) Dominant role of the niche in melanocyte stem-cell fate determination, *Nature* **416**(6883), 854–860.

Okada, T.S. (1991) *Transdifferentiation Flexibility in cell differentiation*, Clarendon Press, Oxford.

Osawa, M., et al. (1996) *Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell*, *Science* **273**(5272), 242–245.

Oshima, H., et al. (2001) Morphogenesis and renewal of hair follicles from adult multipotent stem cells, *Cell* **104**(2), 233–245.

Owen, M. (1988) Marrow stromal stem cells, *J. Cell Sci. Suppl.* **10**, 63–76.

Pereira, R.F., et al. (1995) Cultured adherent cells from marrow can serve as long-lasting precursor cells for bone, cartilage, and lung in irradiated mice, *Proc. Natl. Acad. Sci. U.S.A.* **92**(11), 4857–4861.

Petersen, B.E., et al. (1999) Bone marrow as a potential source of hepatic oval cells, *Science* **284**(5417), 1168–1170.

Pomerantz, J., Blau, H.M. (2004) Nuclear reprogramming: a key to stem cell function in regenerative medicine, *Nat. Cell Biol.* **6**(9), 810–816.

Poss, K.D., Keating, M.T., Nechiporuk, A. (2003) Tales of regeneration in zebrafish, *Dev. Dyn.* **226**(2), 202–210.

Potten, C.S., Morris, R.J. (1988) Epithelial stem cells in vivo, *J. Cell Sci. Suppl.* **10**, 45–62.

Prather, R.S., Sims, M.M., First, N.L. (1989) Nuclear transplantation in early pig embryos, *Biol. Reprod.* **41**(3), 414–418.

Priller, J., et al. (2001) Neogenesis of cerebellar Purkinje neurons from gene-marked bone marrow cells in vivo, *J. Cell Biol.* **155**(5), 733–738.

Raff, M. (2003) Adult stem cell plasticity: fact or artifact? *Annu. Rev. Cell. Dev. Biol.* **19**, 1–22.

Rao, M.S., et al. (1988) Almost total conversion of pancreas to liver in the adult rat: a reliable model to study transdifferentiation, *Biochem. Biophys. Res. Commun.* **156**(1), 131–136.

Reader, J.R., et al. (2003) Pathogenesis of mucous cell metaplasia in a murine asthma model, *Am. J. Pathol.* **162**(6), 2069–2078.

Reddy, J.K., et al. (1991) Pancreatic hepatocytes. An in vivo model for cell lineage in pancreas of adult rat, *Dig. Dis. Sci.* **36**(4), 502–509.

Rhind, S.M., et al. (2003) Human cloning: can it be made safe? *Nat. Rev. Genet.* **4**(11), 855–864.

Rideout, W.M. III, Eggan, K., Jaenisch, R. (2001) Nuclear cloning and epigenetic reprogramming of the genome, *Science* **293**(5532), 1093–1098.

Rubin, H. (2002) The disparity between human cell senescence in vitro and lifelong replication in vivo, *Nat. Biotechnol.* **20**(7), 675–681.

Sims, M., First, N.L. (1994) Production of calves by transfer of nuclei from cultured inner cell mass cells, *Proc. Natl. Acad. Sci. U.S.A.* **91**(13), 6143–6147.

Smith, C., et al. (1991) Purification and partial characterization of a human hematopoietic precursor population, *Blood* **77**(10), 2122–2128.

Spangrude, G.J., Heimfeld, S., Weissman, I.L. (1988) Purification and characterization of mouse hematopoietic stem cells, *Science* **241**(4861), 58–62.

Spear, B.T., Tilghman, S.M. (1990) Role of alpha-fetoprotein regulatory elements in transcriptional activation in transient heterokaryons, *Mol. Cell. Biol.* **10**(10), 5047–5054.

Spradling, A., Drummond-Barbosa, D., Kai, T. (2001) Stem cells find their niche, *Nature* **414**(6859), 98–104.

Stice, S.L., Robl, J.M. (1988) Nuclear reprogramming in nuclear transplant rabbit embryos, *Biol. Reprod.* **39**(3), 657–664.

Suzuki, A., Nakauchi, H. (2002) Identification and propagation of liver stem cells, *Semin. Cell Dev. Biol.* **13**(6), 455–461.

Tosh, D., Slack, J.M. (2002) How cells change their phenotype, *Nat. Rev. Mol. Cell Biol.* **3**(3), 187–194.

Uchida, N., Weissman, I.L. (1992) Searching for hematopoietic stem cells: evidence that Thy-1.1lo Lin- Sca-1+ cells are the only stem cells in C57BL/Ka-Thy-1.1 bone marrow, *J. Exp. Med.* **175**(1), 175–184.

Vassilopoulos, G., Wang, P.R., Russell, D.W. (2003) Transplanted bone marrow regenerates liver by cell fusion, *Nature* **422**(6934), 901–904.

Wagers, A.J., et al. (2002) Little evidence for developmental plasticity of adult hematopoietic stem cells, *Science* **297**(5590), 2256–2259.

Wagers, A.J., Weissman, I.L. (2004) Plasticity of adult stem cells, *Cell* **116**(5), 639–648.

Wang, X., et al. (2002) Kinetics of liver repopulation after bone marrow transplantation, *Am. J. Pathol.* **161**(2), 565–574.

Wang, X., et al. (2003a) Cell fusion is the principal source of bone-marrow-derived hepatocytes, *Nature* **422**(6934), 897–901.

Wang, X., et al. (2003b) The origin and liver repopulating capacity of murine oval cells,

Proc. Natl. Acad. Sci. U.S.A. **100**(Suppl. 1), 11881–11888.

Watt, F.M. (2001) Stem cell fate and patterning in mammalian epidermis, *Curr. Opin. Genet. Dev.* **11**(4), 410–417.

Weimann, J.M., et al. (2003a) Contribution of transplanted bone marrow cells to Purkinje neurons in human adult brains, *Proc. Natl. Acad. Sci. U.S.A.* **100**(4), 2088–2093.

Weimann, J.M., et al. (2003b) Stable reprogrammed heterokaryons form spontaneously in Purkinje neurons after bone marrow transplant, *Nat. Cell Biol.* **5**(11), 959–966.

Weissman, I.L. (2002) The road ended up at stem cells, *Immunol. Rev.* **185**, 159–174.

Willenbring, H., et al. (2004) Myelomonocytic cells are sufficient for therapeutic cell fusion in liver, *Nat. Med.* **10**(7), 744–748.

Wilmut, I., et al. (1997) Viable offspring derived from fetal and adult mammalian cells, *Nature* **385**(6619), 810–813.

Wright, W.E. (1984a) Expression of differentiated functions in heterokaryons between skeletal myocytes, adrenal cells, fibroblasts and glial cells, *Exp. Cell Res.* **151**(1), 55–69.

Wright, W.E. (1984b) Induction of muscle genes in neural cells, *J. Cell Biol.* **98**(2), 427–435.

Wright, N.A. (2000) Epithelial stem cell repertoire in the gut: clues to the origin of cell lineages, proliferative units and cancer, *Int. J. Exp. Pathol.* **81**(2), 117–143.

Stem Cells, Cancer: *see* Cancer Stem Cells

Stem Cells, Embryonic

Mahendra Rao
Stem Cell Unit, Laboratory of Neuroscience, National Institute on Aging, NIH, USA

1 Embryonic Development 383

2 Derivation of ES Cells 385

3 Basic Properties of ES Cells 388

4 Similarities and Differences among ES Cells Derived from Different Species 389

5 Differentiation of ES Cells 390

6 Other Stem Cell Populations, ES-like Cells, Primordial Germ Cell, and Epiblast Cells 392

7 Uses of Embryonic Stem Cells 395

Bibliography 396
Books and Reviews 396
Primary Literature 397

Keywords

Blastocyst
Stage in development of the preimplantation embryo that follows the morula stage where cells are segregated into the ICM and trophoectoderm.

Centrosome
The organizing center for chromosome replication.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

Chimera
An organism that contains cells from more than one set of gametes.

Cloning
The generation of a genetically identical copy. Therapeutic cloning is the term used when a cell line, rather than a unique individual, is generated.

ES-like
Cells that appear to be pluripotent and share some similarity with ESC lines.

Gastrulation
A series of morphogenetic movements that reorganize the developing embryo into three germ layers.

Homologous Recombination
The process of replacing a specific sequence in the genome by a modified sequence such that the inserted sequence is placed in a defined location.

ICM
Inner cell mass or a group of cells that is destined to contribute to the embryo proper.

IVF
In vitro fertilization. A term encompassing the process of generating fertilized preembryos using artificial reproductive technology.

MAPC
Multipotent adult progenitor cells.

Morula
An early stage of preembryo development.

Preembryo
A fertilized embryo prior to implantation in the uterus.

SCNT
Somatic cell nuclear transfer. A process by which somatic cell nuclei are inserted into an enucleated fertilized egg.

Self-renewal
The process by which stem cells generate daughter cells whose properties are identical to the parent cell.

Stem Cells
A population of cells with specialized properties that distinguish them from other cells. Stem cells are classified on the basis of the tissue from which they were isolated and their differentiation ability.

Teratocarcinomas
Tumors derived from transformed germ cells that often include differentiated cells from all germ layers. Teratomas are similar nonmalignant tumors derived from implanted ES cells.

Trophoectoderm
The outer layer of cells after embryo compaction that is destined to contribute to the developing placenta.

Embryonic development prior to implantation of the embryo in the uterus includes a stage of blastocyst formation, where the initial mass of cells is segregated into an inner cell mass destined to contribute to the developing embryo and an outer layer of cells that will contribute to the fetal portion of the placenta. Embryonic stem cells (ESCs) are cells derived from the inner cell mass of the developing preembryo that can be propagated indefinitely in culture. ESCs are characterized by the expression of a characteristic set of markers, the ability to self-renew indefinitely a lack of contact inhibition, atypical cell cycle regulation, the ability to form teratocarcinomas in nude mice, and the ability to differentiate into ectoderm, endoderm, and mesoderm either *in vitro* or *in vivo* after injection into blastocysts. This constellation of abilities distinguishes ESCs from all other cell populations, including adult stem cells.

The unique ability of ESCs to be propagated in culture, undergo genetic manipulation, and, nevertheless, retain the ability to contribute to all portions of the developing embryos, including germ cells, has been utilized to generate transgenic animals in which specific genes are introduced or removed. Such transgenic and knockouts have been utilized effectively to generate animal models of human disease, to develop improved strains, and to study critical aspects of development.

The ability to manipulate ESCs such that subsequent generations are derived solely from cultured manipulated cells, coupled with recent advances in deriving primate (including human) ESC lines has generated much excitement and hope as well as profound ethical concerns and considerable debate on how best to proceed.

1 Embryonic Development

Prior to the union of human gametes, the oocyte first requires timed completion of meiosis. This vital step does not occur throughout a woman's life, rather oocytes are arrested at the first meiotic division until puberty when small numbers mature competitively during the reproductive years. Maturation is complete at the "one day" event of ovulation that occurs in a regular, approximately, monthly cycle. Oocytes that are not fertilized are not retained, and, in humans, oocytes can be successfully fertilized only during a short

period after ovulation. Sperm cells mature from spermatic stem cells through a sequential process that has been well characterized. Spermatic stem cells are, in turn, generated from primordial germ cells that are "set aside" during early embryonic development. Generally, tens of millions of sperm are present in an ejaculate of which only one will successfully fertilize the mature oocyte. Sperms that fail to fertilize are discarded.

Fertilization initiates the process of cell differentiation. Maternal inherited mRNA regulates the initial developmental events, and embryonic transcription is not initiated till later. Once the sperm enters the egg, its DNA-associated proteins are replaced by oocyte histones. The two pronuclei become enveloped with oocyte-derived membranes, which fuse and begin the zygote's mitotic cell cycle. The sperm centrosome controls the first mitotic divisions and zygotic genomic activation occurs around day 4 at the morula stage. Embryonic development initiates with a series of cleavages to produce eight undetermined and essentially equivalent blastomeres. The pattern of cleavage is well coordinated by cytoplasmic factors and, in mammalian eggs, is holoblastic and rotational. The individual blastomeres are initially totipotent until the morula begins compaction and when cells initiate polarization. During compaction, cell boundaries become tightly opposed and cells are no longer equivalent. Cells in the inner cell mass contribute to the embryo proper, while cells on the outside contribute to the trophoectoderm. The blastocyst forms approximately 24 h after the morula stage by the development of an inner fluid-filled cavity, the blastocele. While genomic DNA is inherited from both parents, mitochondrial DNA is inherited from only the mother. Paternal mitochondria transferred at the time of sperm entry are discarded by a process that is little understood.

Human eggs are generally smaller than other eggs (about 100 μm), are usually fertilized within the fallopian tubes, and undergo cellular division in a defined mileu as they migrate toward the uterus. Cellular division follows a predictable 12- to 18-h cycle, resulting in 2 to 16 cell preembryos over the first several days.

Implantation is the process by which the compact zonula is thinned and the blastocyst released to implant. The blastocyst must first hatch from the thinning *zona pellucida* by alternating expansion and contraction, and this process of hatching is critical to further development. Implantation of the hatched blastocyst requires several steps, including apposition, attachment, penetration, and trophoblast invasion, and cannot occur till the first cell specification into trophoectoderm has occurred. As the trophoblast is developing to form the fetal component of the placenta, the endometrial lining of the maternal uterus is undergoing a decidual reaction to generate the maternal component of the placenta.

Simultaneously, the inner cell mass undergoes a process of gastrulation. Gastrulation is defined as a process of complex, orchestrated cell movements that vary widely among species, but include the same basic movements. These include epiboly, invagination, involution, ingression, and delamination. Thus, several major developmental events have taken place as the fertilized egg is migrating from the site of fertilization (fallopian tubes) to the body of the uterus over a period of four days. At all stages, the egg is shielded from the external environment, initially by the zonula and subsequently by the trophoectoderm, but is accessible to *ex utero* manipulation (see Fig. 1). It is important to emphasize

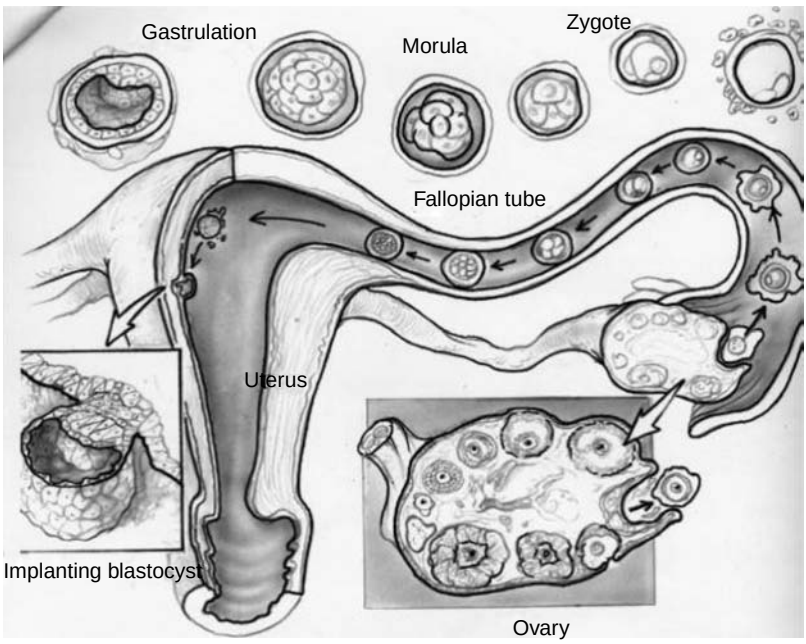

Fig. 1 Normal development of the preimplantation embryo – The mature follicle is released into the fallopian tubes, where it is fertilized by the sperm. The zygote travels down the fallopian tube, undergoing maturation to implant into the uterine wall four to – five days after fertilization. Figure by Tony Med Art.

that these important developmental events have occurred prior to implantation and well before blood vessel growth and heart development. The early-stage fertilized egg that has not yet been implanted has been termed a *preembryo* to distinguish it from the implanted embryo.

As our understanding of early developmental events has increased, our ability to safely manipulate the reproductive process has also increased, and various strategies to assist the reproductive process have been devised. Manipulation of the preimplanted embryo has been feasible for the past three decades, and *in vitro* fertilization, which was first performed 26 years ago, is now a common procedure. Currently, well over a thousand established clinics worldwide provide reproductive services, and approximately 1 to 3% of all live births are achieved with the utilization of some IVF services. The widespread use of IVF technology has led to the availability of large number of embryos that are in excess of reproductive requirement or have been stored for years against future need, and detailed rules governing the procurement, storage, use, gift/donation, destruction of blastocysts have been developed with perhaps the most regulated being the process of derivation of embryonic stem cell lines from human blastocysts.

2 Derivation of ES Cells

Embryonic stem cells (ESCs) are cells harvested from the inner cell mass (ICM) of

the developing preembryo that were destined to develop into all major derivatives of the developing embryo including germ cells that would ultimately develop into the next generation. Remarkably, cells harvested from the ICM can be maintained in culture over prolonged time periods (years) such that they retain this pluripotency in culture. Cultured ESCs differentiate into multiple phenotypes, both in a culture dish and after implantation into immunocompromised mice and contribute to an entire embryo after being injected back into a blastocyst that is subsequently implanted and allowed to develop to term.

ESCs can only be derived from a particular stage of embryonic development prior to implantation, at the time when the preembryo (preimplantation embryo) has separated into an inner cell mass and an outer trophoectoderm. The derivation process involves separating the inner cell mass from other portions of the developing preembryo by mechanical, enzymatic, and immunological means, or by laser surgery. The intact inner cell mass is placed in a dish with feeder cells and complex media, and colonies that propagate and have the appropriate characteristics are selected for additional growth to develop an ES cell line (see Fig. 2). The efficiency of the derivation process is variable, ranging from 30% to less than 1%, and varies among species and even among different strains of the same species. For example, it has been very difficult to obtain ES lines from rats, while it has been comparatively easy to obtain them from mice. Among mice as well, ESC lines can be produced with high efficiency only from

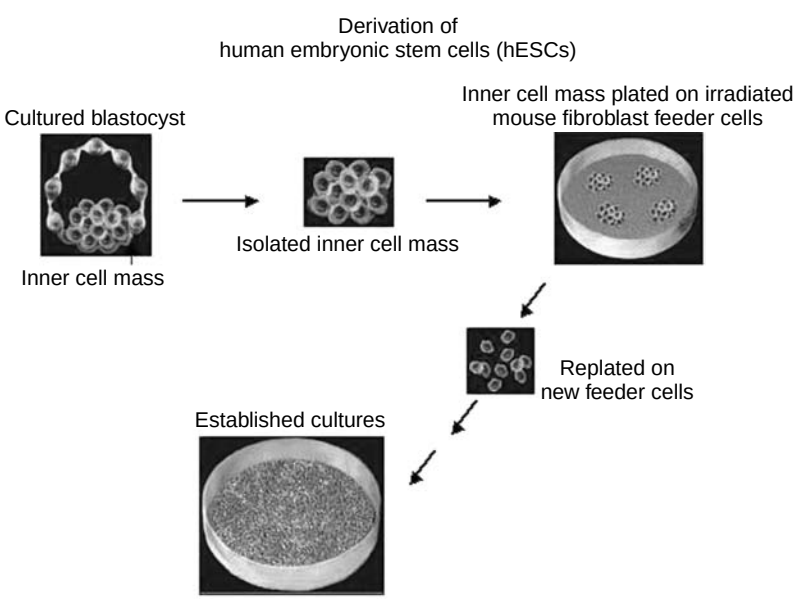

Fig. 2 Derivation of ES cell lines – A diagrammatic representation of the process of ES cell derivation. Cells of the ICM are removed and placed on support cells and propagated in ES cell medium. Cells are periodically tested for marker expression, and stability and lines of pluripotent cells are derived. A single blastocyst may generate one to three lines. Figure courtesy Dr J. Lebkowski (Geron).

some strains. It is unclear as to what dictates success, though the consensus suggests that subtle differences in the timing of rapid developmental change that characterize this stage underlie the differences observed.

ESCs were first described in mice, and were first established in the 1980s by Gail Martin, D. Solter, and others. In most cases, ESC isolation was performed by either mechanical dissociation of the embryo or immune ablation of the cells surrounding the ICM. Subsequently, ES cell lines have been derived from multiple other species, including fish, equines, bovines, and primates. Variations in success have been reported with some difficulty in isolation of ES lines from rats and some strains of mice.

Jamie Thomson and colleagues utilized excess human preimplantation embryos to harvest the inner cell mass and propagate it in culture to derive several lines from human blastocysts. These human cell lines have been maintained successfully in culture for over three years and retain their ability to differentiate into all major components of the embryos (see Fig. 3). Subsequent successful derivations have been performed in multiple laboratories. Prior to August 9, 2001, over 70 independent derivations had been reported. These derivations and cell lines obtained from them have been termed *presidential lines* under the Presidential (George W. Bush – USA) decree that only these lines will be eligible for federal funding. Additional lines have been subsequently derived and, currently, over hundred human embryonic cell lines have been established, and their ability to differentiate into exoderm, endoderm, and mesoderm

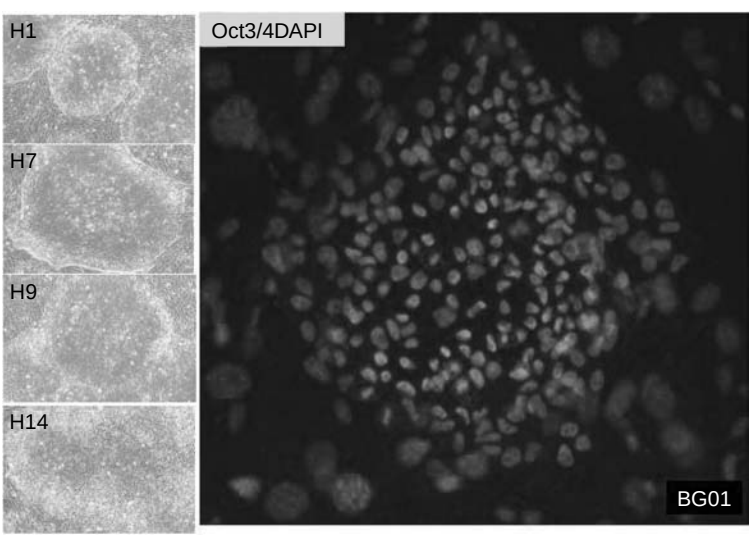

Fig. 3 ES cells – Typical colonies derived from four human ES cell lines are shown. On the right, a colony from a fifth line has been stained with an antibody to Oct3/4 (ES cell marker) and has been counterstained with a nuclear dye to show that the colony consists of a uniform population of undifferentiated cells. Photographs courtesy Dr M. Carpenter (Robarts Research Institute) and Dr X. Zheng (NIDA).

repeatedly demonstrated. Progress in the field has been rapid, and the efficiency of isolation has improved from 1% to about 30%. Variant lines derived from abnormal embryos diagnosed by isolation of single cells from suspect genetically abnormal embryos have been established and, more recently, ES lines from compound blastocysts generated from enucleated fertilized eggs, and somatic nuclei from donor cells have been established.

Ongoing research in the field of ESC derivation has been modified by a variety of factors that include the relatively broad patent issued to Wicell, covering all pluripotent ES lines such as the US president's decision to allow federal funds to be used only for lines derived prior to August 9, the differences in regulation across countries, the longer time period required to propagate hES cells, the proprietary interests of ES cell companies that invested in deriving cell lines, and the broad ethical concerns raised by the ability to manipulate the human genome. Apart from the ethical issues related to consent for donation, destruction of potential human life and other fundamental issues and important concerns have been raised about cloning. The ability to generate human embryonic stem cells and the ability to perform somatic nuclear transfer and successfully clone mammalian species raises fears, often fueled by limited information about cloning human beings. This relatively recent technological breakthrough has created maximum controversy, primarily because of its perceived potential to be used or misused on a large scale. It is useful to remember, however, that at each stage of technological sophistication, profound ethical issues have always been raised and publicly debated. Consensus will no doubt be reached, and one can look forward to accelerated progress once controversies are resolved.

3
Basic Properties of ES Cells

ES cells appear to be differ from all other cell populations in the long-term self-renewal ability, lack of contact inhibition in culture, regulation of the cell division cycle, their ability to form teratomas in immune compromised mice, and their ability to contribute to the germline after injection into blastocysts. These functional differences reflect in differences in gene expression, presence of cell surface epitopes, growth factor, and cell culture requirements, and serve to distinguish ESC from other stem cell populations derived from different developmental stages or different tissues. A few other embryonic cell types have been shown to retain some of the properties of ICM-derived ES cells, but so far no adult stem cell or even a stem cell derived from late fetal stages has shown a similar profile.

Much of what we know about ESCs comes from experiments performed on mouse ES cells. ES cells, irrespective of which strain of mouse they are derived from, grow as compact, well-defined colonies of tightly packed cells that are attached to each other by gap junctions. Cells have a high nuclear-to-cytoplasm ratio, with each cell typically having one or more prominent nucleoli. ESCs require growth factors to permit their growth, and these factors appear to be present in fetal bovine serum and in feeder cells prepared from mouse embryonic fibroblasts. In addition, supplementation with leukemia inhibitory factor (LIF) will enhance growth of mESC and may substitute for requirement of coculture with fibroblasts. Other genes

identified that are required for the maintenance of the ESC or required for proper maturation of the blastocyst include a variety of growth factors, transcription factors, and other regulatory genes, a few of which are discussed briefly below. Oct4-/- embryos die before the egg cylinder stage, and ES cells cannot be established from Oct4-/- cells. Levels of Oct4 are critical to the fate of the cells. Low cells lead to differentiation into trophoblast giant cells, while high levels cause differentiation into primitive endoderm and mesoderm. Sox-2 is another transcription factor that is thought to cooperate with Oct4 to maintain undifferentiated cells and activate key ES cell–enriched genes. Fibroblast growth factor (FGF4) is required for multiple aspects of early development and FGF4-/- embryos fail to develop after implantation, and ICM-derived cells do not proliferate *in vitro*. Foxd3/Genesis is another transcription factor that may be required for early embryonic development. Nanog is a recently identified homeodomain containing a transcription factor that appears critical for the proper self-renewal of ES cells. BMPs (bone morphogenetic proteins) and other members of the TGFß family appear to be important at multiple stages of early embryo development. BMP acting via BMPR1 can enhance self-renewal of ESC and substitute for the serum in ESC medium. FGFs and wnts are thought to play an important role in the process of gastrulation.

Examination of ESCs has made clear that multiple pathways that interact with each other at multiple levels regulate their proliferation and self-renewal. These pathways appear similar to the pathways activated as the ICM matures *in vivo* to generate embryonic tissue. ESC lines therefore serve as a useful and possibly unique model of early stages of embryonic development. These early stages have been very difficult to study in humans.

4 Similarities and Differences among ES Cells Derived from Different Species

The ability to compare ES cells from human and rodents has allowed one to assess conserved pathways likely to be important in ES cell self-renewal. Nanog, Oct3/4, sox, rex, utf, tert, connexin, Glut-1, Dicer, DNMT, and so on all appear to be conserved, and are likely to be important in ES cell self-renewal. This conservation of key pathways suggests a core ES cell identity that is conserved across species. This core identity that would include Oct, Sox, REX, UTF, TERT, and so on will be clearly distinct from genes expressed in other "adult/non ES" stem cell populations. Many, but not all, of these pathways are likely conserved between species (see Fig. 4).

Differences in gene expression and regulation have been reported as well, and, surprisingly, these differences have included genes important for self-renewal and survival of mouse ES cells. For example, Oct3/4 homologs likely do not exist in chicken embryos, while LIF signaling, which is critical for ES cell self-renewal in rodents, does not appear to be critical or even required for human ES cells. No paralogs of *E-Hox*, a gene critical for mouse ES cell self-renewal and differentiation, have been identified in humans. Eras, which is required for the growth of mouse ES cells, appears to be a pseudogene in humans. Human ES cells appear capable of differentiating into trophoectoderm, while mouse cells lack this capability. Species differences in equivalent populations are to be expected; however, the number of

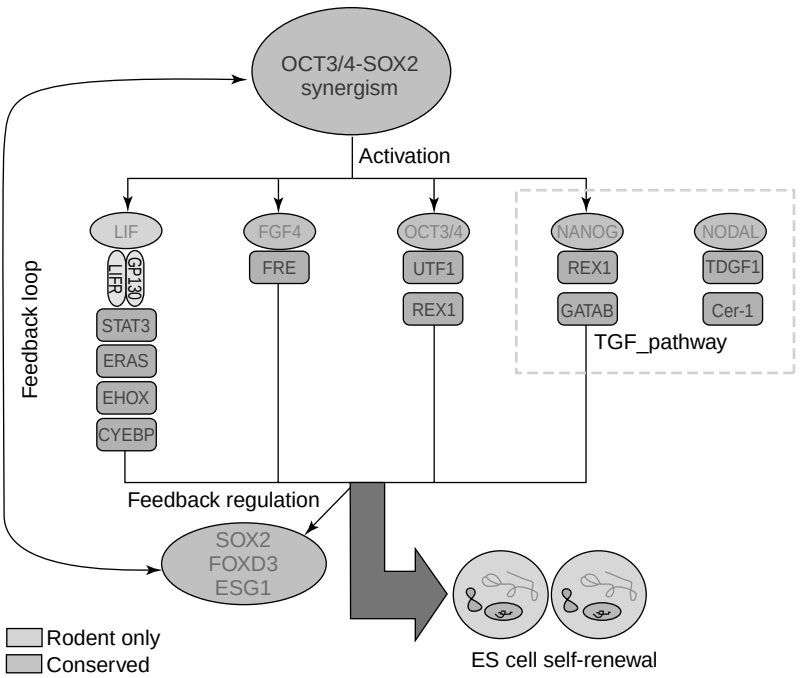

Fig. 4 Pathways to self-renewal in ESC – A model of the multiple pathways whose activation is required to maintain ES cells in an undifferentiated state. Model courtesy Dr M. Zhang (NIA).

differences reported from limited comparisons is quite surprising and may indicate evolutionary pressure for speciation.

5 Differentiation of ES Cells

ES cells can be induced to differentiate by altering culture conditions, injecting them into mice, or developing blastocysts or generating embryoid bodies by a process of aggregation, and many reports have provided evidence of the ability of direct ES cell differentiation toward particular lineages. Some generalizations of the process of differentiation can be made (see Fig. 5).

In no case has it been possible to direct 100% of the cells toward particular phenotype, and, in general, the efficiency of differentiation has been in the 10 to 20% range. Not enough markers (see Fig. 6) have been identified to allow generalized sorting strategies that have

Fig. 6 Markers to define ESC – The table summarizes some of the common markers used to examine ES cell colonies. Note that absence of markers of differentiated cells is an important criterion for determining if ES cells are undifferentiated. Panel on the right shows double labeling with two ES cell markers, while the lower panel shows the results of a PCR (polymerase chain reaction) amplification of ES cell markers in undifferentiated ES cells and differentiated cells. 1 = ES cells, 2 = embryoid body cells.

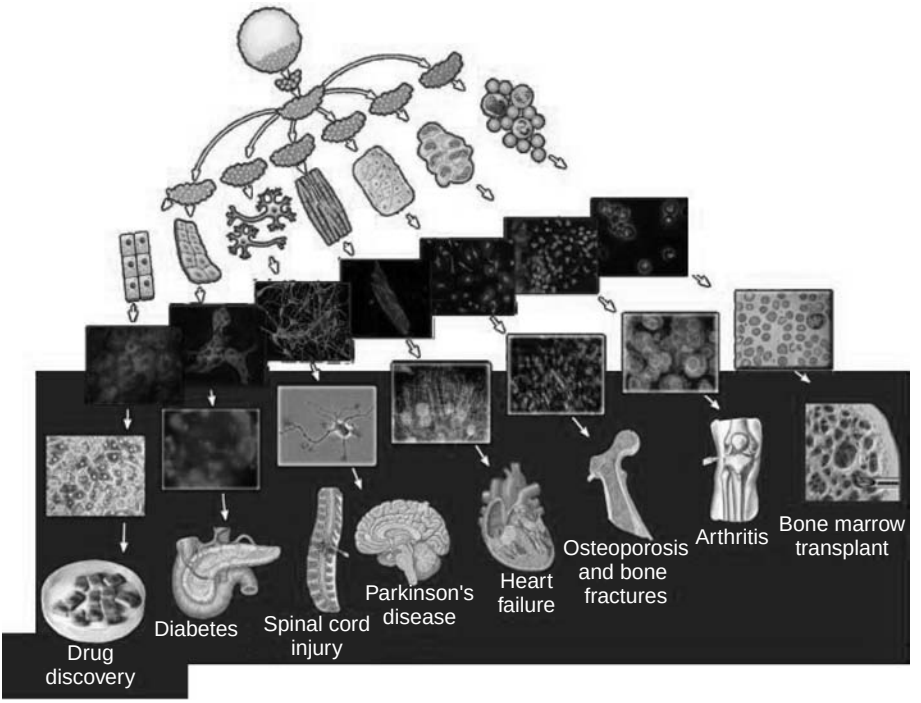

Fig. 5 Differentiation of ES cells – ES cells are capable of differentiating into an astounding variety of cells under the appropriate culture conditions or after transplantation in animal models. The figure summarizes some of the cells types that have been generated from hESC cultures. Figure courtesy Dr J. Lebkowski (Geron).

Markers
Alkaline phosphatase SSEA1, SSEA4, TRA160, TRA181/CD9/E-cadherin
Sox, oct, nanog, Rex, utf-1,dppa5
Absence of trophoectoderm markers (BEX)
Absence of neuroectoderm (SOX-1)
Absence of endoderm (GATA genes, HNF3β, PDX-1)
Absence of mesoderm (brachyury, MSX-1)
Cell cycle/TERT/karyotype/gap junctions

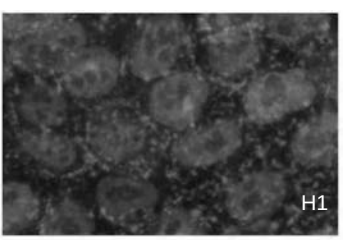

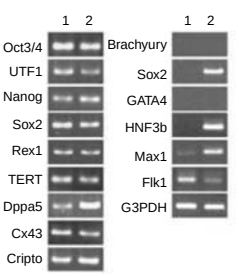

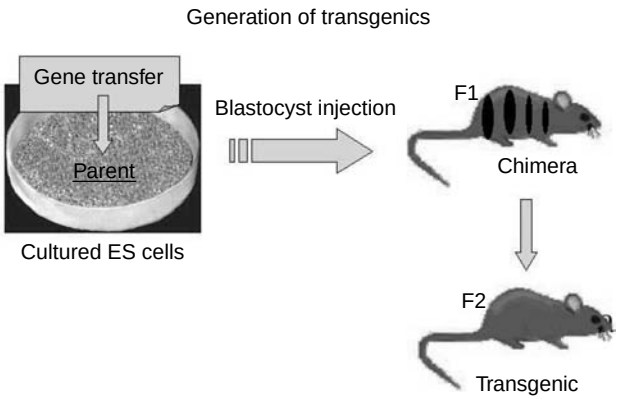

Fig. 7 Generation of transgenics – The figure summarizes the process of making transgenics. This particular use of ES cells has revolutionized the study of early developmental events and has allowed biologists to generate precise models of human disease in mice. The fear that similar strategies may be employed using hESC has raised concerns and much political debate.

been possible in other systems to be used with equal efficiency. In general, tissues or cell types that are generated early in development are easier to produce from ES cell cultures with current technology. Cells that do differentiate generally have fetal or embryonic characteristics, are not identical to adult cells, and, in many cases, are functionally immature. The differentiation process, while more rapid than that seen during embryonic development, is nevertheless slow, and it is unclear as to how much time in culture will be required to obtain differentiated human cells.

Progress in identifying and initiating studies to solve these constraints on the use of ES cell–derived differentiated cells has been rapid, though much remains to be done. Already, preliminary experiments have been performed that show that dopaminergic neurons, which are lost in Parkinsons Disease, can be generated from hESC cultures. Transplant experiments in mouse models have shown functional integration and prolonged survival with modest improvements in behavioral readouts. Similar promising results have been obtained with transplants in a spinal cord injury model and in a cardiac injury model. Pancreatic islet differentiation has been demonstrated from mouse ES cells, and preliminary reports suggest that a similar strategy can be used to generate islet cells from hESC as well (see Fig. 7).

The field is still in its infancy, and much needs to be understood about the process of differentiation. However, the initial results with biasing differentiation, isolating specific types of cells, and transplanting in animal models of disease have shown enough promise that scientists have been encouraged to proceed further.

6 Other Stem Cell Populations, ES-like Cells, Primordial Germ Cell, and Epiblast Cells

As the blastocyst matures, it undergoes a process of gastrulation to generate cells that will contribute to the different germ layers. The trilaminar embryo, formed as

gastrulation proceeds, will undergo additional morphogenetic movements and further progressive differentiation to generate more differentiated cells, tissues, and organs. What has become clear is that the process of differentiation involves the sequential generation of cells, which can still divide and self-renew, and thus fulfill the criteria of stem cells.

These more tissue-specific stem cells have been classified on the basis of their differentiation ability (hepatic, muscle, neural, hematopoietic etc.), and differ from embryonic stem cells (ESCs) by numerous criteria (see above). These tissue-specific stem cells are present throughout development, and may be present in the adult organism as well. Adult tissue-specific stem cells have been harvested from a variety of tissues, including neural and hematopoietic tissue, and adult HSCs are used for a variety of therapeutic purposes.

It is unclear if ICM cells are set aside during development or persist in special niches in the adult. The general consensus has been that they do not. However, two reports have suggested otherwise. Perhaps the most widely studied is the multipotent adult progenitor cell (MAPC) derived from mesenchymal cell cultures and first identified by Catherin Verfaille and colleagues. More recently, Young and colleagues have suggested that pluripotent cells persist as a rare population in multiple tissues (see Fig. 8). These cells that no doubt differ from ICM-derived ES cells by multiple criteria, however, may be functionally equivalent for therapeutic applications and may have the potential advantage of being derived from the individual requiring cell

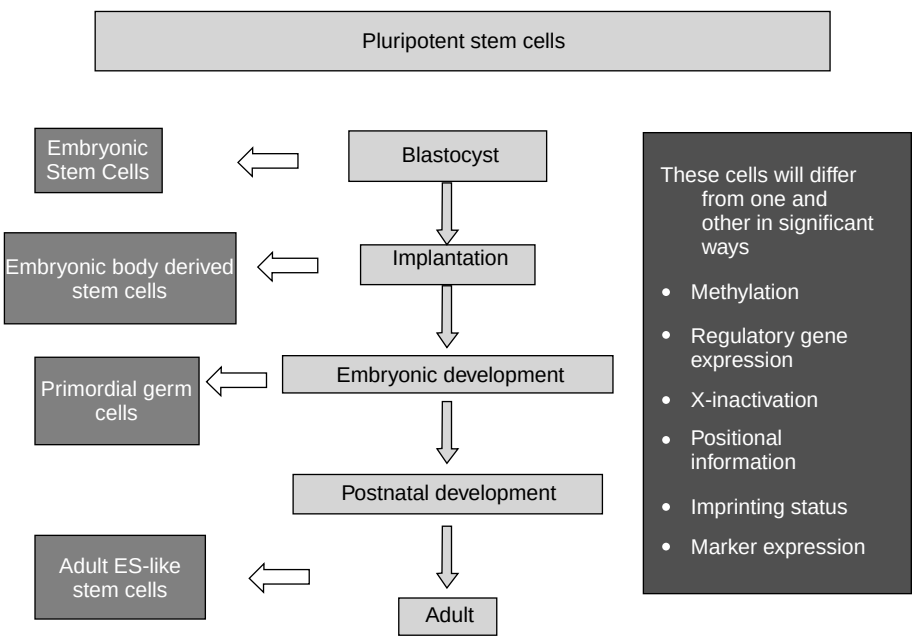

Fig. 8 Pluripotent cells – ES cells as well as cells derived from later stages of embryonic development may be pluripotent and capable of generating ectoderm, endoderm, and mesoderm derivatives. However, these cells will differ from each other in marker expression and other changes that occur as development proceeds.

therapy. In addition to these two populations, other investigators have suggested that stem cells (non-ESC), though tissue specific in their contribution during development, can be induced to reacquire additional capabilities by manipulating cell culture conditions or exposing them to drugs that alter nuclear structure or exposing them to cytoplasm from ES cells, unfertilized eggs, or germ cell tumors. These transdifferentiated or dedifferentiated cells, while differing from ES cells, may be functionally equivalent for therapeutic applications. The possible existence of such cells has fanned an ethical debate on determining whether deriving new ES lines is more appropriate or whether one should focus on dedifferentiating or transdifferentiating adult stem cells. In any case, it has fueled a drive to assess the properties of pluripotent cells and understand the underlying differences between stem cell populations (see Fig. 9).

In addition to ES-like and transdifferentiated cell populations, two additional populations of cells merit mention. These are primordial germ cells that represent the precursor of the sperm and oocyte and epiblast cells, which are present at a stage of development soon after gastrulation. Both of these populations appear similar to ESC in that they are pluripotent, express many of the markers characteristic of ESC including ES cell enriched genes, and

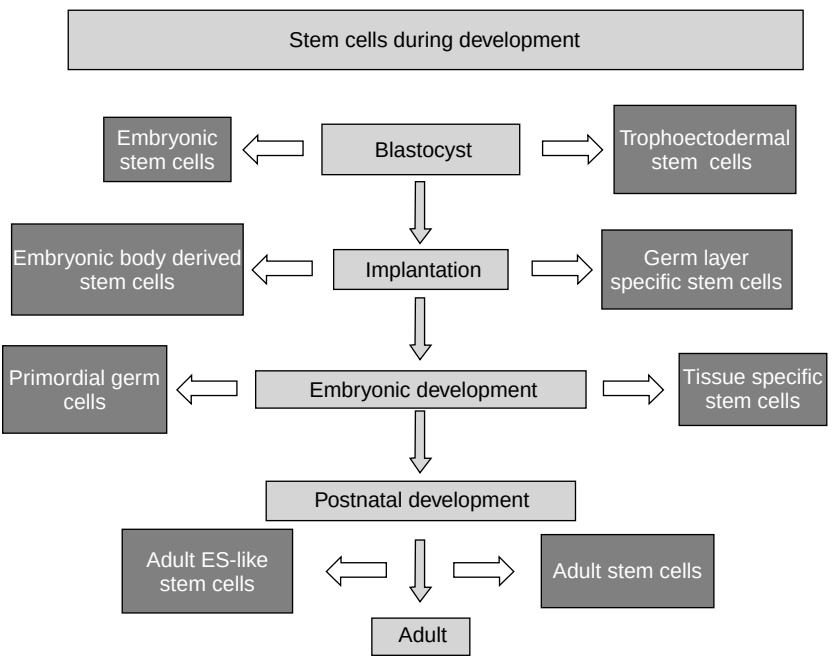

Fig. 9 Adult stem cells. ES cells and other pluripotent cells need to be distinguished from cells derived from more mature stages of embryonic, fetal, or postnatal development. These cells differ from ES cells, but may offer benefits distinct from those of ES cells. It is important to note that adult cells can be derived from ES cell cultures and, as such, the field of regenerative medicine does not consider choices to be either one type of cell or another. Rather, each cell type is uniquely suited for particular therapeutic indications.

can contribute to the germ line after blastocyst injection. These pluripotent stem cell populations do differ from ES cells in patterns of methylation, X chromosome inactivation, growth conditions, and growth factor requirements. However, their functional similarity suggests that these cells may be potential alternates for therapeutic use, more so when compared with transdifferentiated cells or ES-like cells. Work on germ cells and embryoid body derived cells was pioneered by Brigid Hogan and John Gearhart, and these lines offer a valuable alternative to ICM-derived pluripotent ESC lines.

7
Uses of Embryonic Stem Cells

Perhaps the most common use of mouse ES cells has been to generate transgenic mice. In the early 1980s, soon after the successful isolation of ES cell lines, Mario Capecchi and colleagues showed that it was possible to insert foreign DNA into a specific genomic locus via a process termed *homologous recombination*. This has allowed replacement of defective genes with a corrected copy, insertion of reporter molecules to trace the development of cells, or alteration of the levels of a particular product to assess the effect of changing a single protein. This remarkable ability to study directed changes throughout development has offered an unprecedented insight into the biology of complex organisms and has transformed developmental biology.

ES cells, by recapitulating development *in vitro*, have also provided a window into early developmental processes such as X chromosome inactivation, regulation of methylation, segregation of maternal and paternal mitochondria, evolutionary differences, speciation, and the differentiation of germ cells. Somatic nuclear transfer has allowed cloning of mice, sheep, cows, and pigs, and has been widely heralded in the animal husbandry field as a technique to perhaps revive extinct species or develop better strains of animals. These techniques are routinely used and some have been commercialized.

Human ESCs are currently not used in any clinical or commercial application, although there is much excitement about their potential use, and a wide variety of strategies have been proposed. An obvious use of hES cells is to use them as a source of cells for transplantation, and indeed much of the current focus is on isolating cells of the appropriate phenotype for cell replacement. Isolation and transplantation of dopaminergic neurons to treat Parkinson's disease, islet cells to treat diabetes, and cardiac muscle to treat heart failure are strategies under evaluation.

Many other uses of hESCs can be envisaged. These include modifying transplant strategies to take advantage of the unique properties of hESC behavior *in vivo*, using hESC in gene and drug discovery strategies, and using hESC to understand the basic biology of early developmental events such as mitochondrial segregation, regulation of telomerase expression, or maintenance of the pluripotent state. These additional or alternate uses of hESC do not require any fundamental breakthroughs; rather, these are an adaptation of the existing technology to hESC cultures. These specific and unique uses of hESC represent an important and less-discussed aspect of the advantages of hES cells over other cell types.

Additional techniques and manipulation strategies have provoked controversy even while promising possible novel therapies.

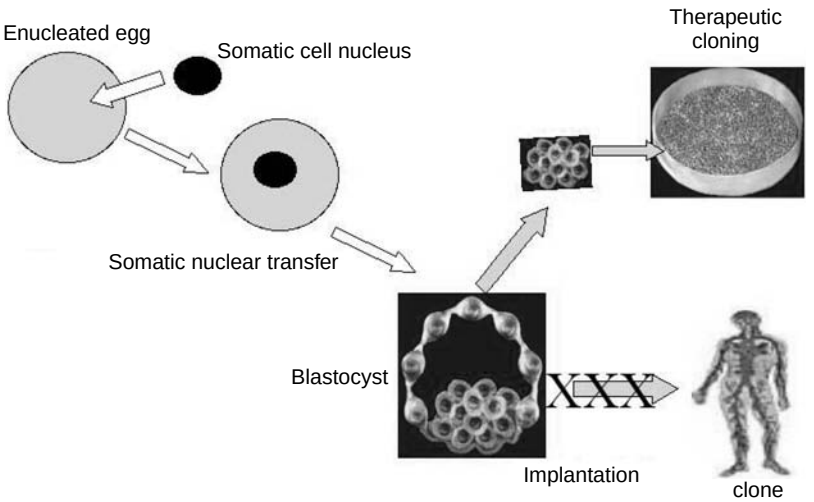

Fig. 10 Therapeutic cloning – The process of enucleating a fertilized egg and replacing it with a nucleus from any cell is termed *nuclear transfer* and such a compound egg can be implanted and, in a small number of cases, can lead to the development of viable offspring. Such cloned offspring are genetically identical to the donor nucleus, and the process of cloning is in use in animal husbandry. In humans, compound eggs can be used to generate ES cell lines that are genetically identical to the donor. These lines could then be used as a source for immunologically matched cells.

The ability to transfer a nucleus from any cell in the body to a fertilized egg that has been enucleated to generate novel ES cell lines (therapeutic cloning) or an embryo that can survive to term (cloning) in animals has raised fears that this theoretical possibility may be actualized in humans as well. The ability to generate ES lines from embryos that carry a genetic defect diagnosed at the time of blastocyst implantation by PGD (preimplantation genetic diagnosis) has important potential benefits in understanding the etiology and progression of congenital diseases. Nevertheless, enthusiasm has been tempered by concerns that embryos may be specifically created to generate designer ES cell lines (see Fig. 10).

The rich and unique promise of ES cell biology coupled with real and perceived risks of misuse of this promise has provoked healthy debate, has raised fears of disruptive social change, and guarantees interesting times ahead.

See also Brain Development; Cancer Stem Cells; Cellular Interactions; Mammalian Cell Culture Methods; Nuclear Transfer for Cloning Animals; Parthenogenetic Stem Cells; Principles and Applications of Embryogenomics; Stem Cells.

Bibliography

Books and Reviews

Chiu, A., Rao, M.S. (Eds.) (2003) *Human Embryonic Stem Cells*, Humana Press, Totowa. NJ, pp. 461.

Marshak, D.R. et al. (Eds.) (2001) *Stem Cell Biology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, ME, pp. 500.

Robertson, E.J., Kaufman, M.H., Bradley, A., Evans, M. (1983) Isolation, Properties, and Karyotype Analysis of Pluripotent (EK) Cell Lines from Normal and Parthenogenetic Embryos, in: Silver, L.M., Martin, G.R., Strickland, S. (Eds.) *Teratocarcinoma Stem Cells*, Cold Spring Harbor Press, Cold Spring Harbor, New York, pp. 647–663.

Smith, A. (2001) Embryonic Stem Cells, in: Marshak, D.R. (Ed.) *Stem Cell Biology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, ME, pp. 205–230.

Thomas, K.R., Folger, K.R., Capecchi, M.R. (1986) High frequency targeting of genes to specific sites in the mammalian genome, *Cell* **44**(3), 419–428.

Wilmut, I. (2003) Dolly-her life and legacy, *Cloning Stem Cells* **5**(2), 99–100.

Primary Literature

Allegrucci, C., Denning, C., Priddle, H., Young, L. (2004) Stem-cell consequences of embryo epigenetic defects, *Lancet* **364**, 206–208.

Andrews, P.W. (1998) Teratocarcinomas and human embryology: pluripotent human EC cell lines. Review article, *Apmis* **106**, 158–167. discussion 167–8.

Andrews, P.W., Meyer, L.J., Bednarz, K.L., Harris, H. (1984) Two monoclonal antibodies recognizing determinants on human embryonal carcinoma cells react specifically with the liver isozyme of human alkaline phosphatase, *Hybridoma* **3**, 33–39.

Assady, S., Maor, G., Amit, M., Itskovitz-Eldor, J., Skorecki, K.L., Tzukerman, M. (2001) Insulin production by human embryonic stem cells, *Diabetes* **50**, 1691–1697.

Beddington, R.S., Robertson, E.J. (1989) An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo, *Development* **105**, 733–737.

Besser, D. (2004) Expression of Nodal, Lefty-A, and Lefty-B in undifferentiated human embryonic stem cells requires activation of Smad2/3, *J. Biol. Chem.* **279**, 45076–45084.

Bhattacharya, B., Miura, T., Brandenberger, R., Mejido, J., Luo, Y., Yang, A.X., Joshi, B.H., Ginis, I., Thies, R.S., Amit, M., Lyons, I., Condie, B.G., Itskovitz-Eldor, J., Rao, M.S., Puri, R.K. (2004) Gene expression in human embryonic stem cell lines: unique molecular signature, *Blood* **103**, 2956–2964.

Bradley, A., Hasty, P., Davis, A., Ramirez-Solis, R. (1992) Modifying the mouse: design and desire, *Biotechnology (N Y)* **10**, 534–539.

Brandenberger, R., Khrebtukova, I., Thies, S.R., Cai, J., Miura, T., Puri, R., Vasicek, T., Lebkowski, J., Rao, M.S. (2004) MPSS profiling of human embryonic stem cells, *BMC Dev. Biol.* **4**, 10.

Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J., Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R., Lebkowski, J., Stanton, L.W. (2004) Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation, *Nat. Biotechnol.* **22**, 707–716.

Burdon, T., Stracey, C., Chambers, I., Nichols, J., Smith, A. (1999) Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells, *Dev. Biol.* **210**, 30–43.

Cai, J., Weiss, M.L., Rao, M.S. (2004) In search of "stemness", *Exp. Hematol.* **32**, 585–598.

Carpenter, M.K., Inokuma, M.S., Denham, J., Mujtaba, T., Chiu, C.P., Rao, M.S. (2001) Enrichment of neurons and neural precursors from human embryonic stem cells, *Exp. Neurol.* **172**, 383–397.

Carpenter, M.K., Rosler, E.S., Fisk, G.J., Brandenberger, R., Ares, X., Miura, T., Lucero, M., Rao, M.S. (2004) Properties of four human embryonic stem cell lines maintained in a feeder-free culture system, *Dev. Dyn.* **229**, 243–258.

Caulfield, T., Knowles, L., Meslin, E.M. (2004) Law and policy in the era of reproductive genetics, *J. Med. Ethics* **30**, 414–417.

Chalmers, D., Nicol, D. (2003) Embryonic stem cell research: can the law balance ethical, scientific and economic values? (Part II), *Law. Hum. Genome Rev.* **19**, 91–108.

Chiu, A., Rao, M.Se. (2003) *Human Embryonic Stem Cells*, Humana Press, Totowa, NJ.

Conley, B.J., Trounson, A.O., Mollard, R. (2004) Human embryonic stem cells form embryoid bodies containing visceral endoderm-like derivatives, *Fetal Diagn. Ther.* **19**, 218–223.

Doetschman, T.C., Eistetter, H., Katz, M., Schmidt, W., Kemler, R. (1985) The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac,

blood islands and myocardium, *J. Embryol. Exp. Morphol.* **87**, 27–45.

Draper, J.S., Smith, K., Gokhale, P., Moore, H.D., Maltby, E., Johnson, J., Meisner, L., Zwaka, T.P., Thomson, J.A., Andrews, P.W. (2004) Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells, *Nat. Biotechnol.* **22**, 53–54.

Drukker, M. (2004) Immunogenicity of human embryonic stem cells: can we achieve tolerance? *Springer Semin. Immunopathol.* **26**, 201–213.

Evans, M.J., Kaufman, M.H. (1981) Establishment in culture of pluripotential cells from mouse embryos, *Nature* **292**, 154–156.

Galli, C., Lagutina, I., Lazzari, G. (2003) Introduction to cloning by nuclear transplantation, *Cloning Stem Cells* **5**, 223–232.

Gardner, R.L., Brook, F.A. (1997) Reflections on the biology of embryonic stem (ES) cells, *Int. J. Dev. Biol.* **41**, 235–243.

Ginis, I., Luo, Y., Miura, T., Thies, S., Brandenberger, R., Gerecht-Nir, S., Amit, M., Hoke, A., Carpenter, M.K., Itskovitz-Eldor, J., Rao, M.S. (2004) Differences between human and mouse embryonic stem cells, *Dev. Biol.* **269**, 360–380.

Hansen, B. (2004) Embryonic stem cell research: terminological ambiguity may lead to legal obscurity, *Med. Law* **23**, 19–28.

Hart, A.H., Hartley, L., Ibrahim, M., Robb, L. (2004) Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human, *Dev. Dyn.* **230**, 187–198.

Heins, N., Englund, M.C., Sjoblom, C., Dahl, U., Tonning, A., Bergh, C., Lindahl, A., Hanson, C., Semb, H. (2004) Derivation, characterization, and differentiation of human embryonic stem cells, *Stem Cells* **22**, 367–376.

Hosler, B.A., LaRosa, G.J., Grippo, J.F., Gudas, L.J. (1989) Expression of REX-1, a gene containing zinc finger motifs, is rapidly reduced by retinoic acid in F9 teratocarcinoma cells, *Mol. Cell. Biol.* **9**, 5623–5629.

Humphrey, R.K., Beattie, G.M., Lopez, A.D., Bucay, N., King, C.C., Firpo, M.T., Rose-John, S., Hayek, A. (2004) Maintenance of pluripotency in human embryonic stem cells is STAT3 independent, *Stem Cells* **22**, 522–530.

Hwang, W.S., Ryu, Y.J., Park, J.H., Park, E.S., Lee, E.G., Koo, J.M., Jeon, H.Y., Lee, B.C., Kang, S.K., Kim, S.J., Ahn, C., Hwang, J.H., Park, K.Y., Cibelli, J.B., Moon, S.Y. (2004) Evidence of a pluripotent human embryonic stem cell line derived from a cloned blastocyst, *Science* **303**, 1669–1674.

Ida, R. (2004) [Ethical issues in the development of regenerative medicine: research on human embryonic and adult stem cells], *Rinsho Ketsueki* **45**, 454–460.

Kannagi, R., Cochran, N.A., Ishigami, F., Hakomori, S., Andrews, P.W., Knowles, B.B., Solter, D. (1983) Stage-specific embryonic antigens (SSEA-3 and -4) are epitopes of a unique globo-series ganglioside isolated from human teratocarcinoma cells, *EMBO J.* **2**, 2355–2361.

Kaufman, D.S., Hanson, E.T., Lewis, R.L., Auerbach, R., Thomson, J.A. (2001) Hematopoietic colony-forming cells derived from human embryonic stem cells, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10716–10721.

Kumar, R.A., Chan, K.L., Wong, A.H., Little, K.Q., Rajcan-Separovic, E., Abrahams, B.S., Simpson, E.M. (2004) Unexpected embryonic stem (ES) cell mutations represent a concern in gene targeting: lessons from "fierce" mice, *Genesis* **38**, 51–57.

Lavon, N., Yanuka, O., Benvenisty, N. (2004) Differentiation and isolation of hepatic-like cells from human embryonic stem cells, *Differentiation* **72**, 230–238.

Levenberg, S., Golub, J.S., Amit, M., Itskovitz-Eldor, J., Langer, R. (2002) Endothelial cells derived from human embryonic stem cells, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 4391–4396.

Li, L., Baroja, M.L., Majumdar, A., Chadwick, K., Rouleau, A., Gallacher, L., Ferber, I., Lebkowski, J., Martin, T., Madrenas, J., Bhatia, M. (2004) Human embryonic stem cells possess immune-privileged properties, *Stem Cells* **22**, 448–456.

Lisker, R. (2003) Ethical and legal issues in therapeutic cloning and the study of stem cells, *Arch. Med. Res.* **34**, 607–611.

Lorincz, M.T., Detloff, P.J., Albin, R.L., O'Shea, K.S. (2004) Embryonic stem cells expressing expanded CAG repeats undergo aberrant neuronal differentiation and have persistent Oct-4 and REST/NRSF expression, *Mol. Cell. Neurosci.* **26**, 135–143.

Marshak, D.Rea. (2001) *Stem Cell Biology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, ME.

Martin, G.R. (1980) Teratocarcinomas and mammalian embryogenesis, *Science* **209**, 768–776.

Martin, G.R. (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells, *Proc. Natl. Acad. Sci. U. S. A.* **78**, 7634–7638.

Matsui, Y., Zsebo, K., Hogan, B.L. (1992) Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture, *Cell* **70**, 841–847.

McHugh, P.R. (2004) Zygote and "clonote"–the ethical use of embryonic stem cells, *N. Engl. J. Med.* **351**, 209–211.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., Yamanaka, S. (2003) The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells, *Cell* **113**, 631–642.

Nichols, J. (2001) Introducing embryonic stem cells, *Curr. Biol.* **11**, R503–R505.

Niswander, L., Martin, G.R. (1992) Fgf-4 expression during gastrulation, myogenesis, limb and tooth development in the mouse, *Development* **114**, 755–768.

Niwa, H., Miyazaki, J., Smith, A.G. (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells, *Nat. Genet.* **24**, 372–376.

Niwa, H., Burdon, T., Chambers, I., Smith, A. (1998) Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3, *Genes Dev.* **12**, 2048–2060.

Odorico, J.S., Kaufman, D.S., Thomson, J.A. (2001) Multilineage differentiation from human embryonic stem cell lines, *Stem Cells* **19**, 193–204.

Park, J.H., Kim, S.J., Lee, J.B., Song, J.M., Kim, C.G., Roh, S. 2nd, Yoon, H.S. (2004) Establishment of a human embryonic germ cell line and comparison with mouse and human embryonic stem cells, *Mol. Cells* **17**, 309–315.

Park, S.P., Lee, Y.J., Lee, K.S., Ah Shin, H., Cho, H.Y., Chung, K.S., Kim, E.Y., Lim, J.H. (2004) Establishment of human embryonic stem cell lines from frozen-thawed blastocysts using STO cell feeder layers, *Hum. Reprod.* **19**, 676–684.

Pau, K.Y., Wolf, D.P. (2004) Derivation and characterization of monkey embryonic stem cells, *Reprod. Biol. Endocrinol.* **2**, 41.

Reubinoff, B.E., Pera, M.F., Fong, C.Y., Trounson, A., Bongso, A. (2000) Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro, *Nat. Biotechnol.* **18**, 399–404.

Robertson, E.J., Kaufman, M.H., Bradley, A., Evans, M. (1983) Isolation, Properties, and Karyotype Analysis of Pluripotent (EK) Cell Lines from Normal and Parthenogenetic Embryos, in: Silver, L.M., Martin, G.R., Strickland, S. (Eds.) *Teratocarcinoma Stem Cells*, Cold Spring Harbor Press, Cold Spring Harbor, New York, pp. 647–663.

Rosler, E.S., Fisk, G.J., Ares, X., Irving, J., Miura, T., Rao, M.S., Carpenter, M.K. (2004) Long-term culture of human embryonic stem cells in feeder-free conditions, *Dev. Dyn.* **229**, 259–274.

Rosner, M.H., Vigano, M.A., Ozato, K., Timmons, P.M., Poirier, F., Rigby, P.W., Staudt, L.M. (1990) A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo, *Nature* **345**, 686–692.

Rossant, J., Cross, J.C. (2001) Placental development: lessons from mouse mutants, *Nat. Rev. Genet.* **2**, 538–548.

Sandel, M.J. (2004) Embryo ethics–the moral logic of stem-cell research, *N. Engl. J. Med.* **351**, 207–209.

Shamblott, M.J., Axelman, J., Wang, S., Bugg, E.M., Littlefield, J.W., Donovan, P.J., Blumenthal, P.D., Huggins, G.R., Gearhart, J.D. (1998) Derivation of pluripotent stem cells from cultured human primordial germ cells, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 13726–13731.

Sherley, J.L. (2004) Human Embryonic Stem Cell Research: No Way Around a Scientific Bottleneck, *J. Biomed. Biotechnol.* **2004**, 71–72.

Shuster, E. (2003) Human cloning: category, dignity, and the role of bioethics, *Bioethics* **17**, 517–525.

Smith, A.G. (2001) Embryo-derived stem cells: of mice and men, *Annu. Rev. Cell Dev. Biol.* **17**, 435–462.

Solter, D., Damjanov, I. (1979) Teratocarcinomas rarely develop from embryos transplanted into athymic mice, *Nature* **278**, 554–555.

Solter, D., Knowles, B.B. (1978) Monoclonal antibody defining a stage-specific mouse embryonic antigen (SSEA-1), *Proc. Natl. Acad. Sci. U. S. A.* **75**, 5565–5569.

Suda, Y., Suzuki, M., Ikawa, Y., Aizawa, S. (1987) Mouse embryonic stem cells exhibit indefinite proliferative potential, *J. Cell. Physiol.* **133**, 197–201.

Surani, M.A. (2004) Stem cells: how to make eggs and sperm, *Nature* **427**, 106–107.

Suss-Toby, E., Gerecht-Nir, S., Amit, M., Manor, D., Itskovitz-Eldor, J. (2004) Derivation of a diploid human embryonic stem cell line from a mononuclear zygote, *Hum. Reprod.* **19**, 670–675.

Tamkins, T. (2004) South Koreans create human stem cell line using nuclear transfer, *Lancet* **363**, 623.

Tarin, J.J., Cano, A. (1999) Do human concept have the potential to enter into diapause?, *Hum. Reprod.* **14**, 2434–2436.

Thomas, K.R., Folger, K.R., Capecchi, M.R. (1986) High frequency targeting of genes to specific sites in the mammalian genome, *Cell* **44**, 419–428.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., Jones, J.M. (1998) Embryonic stem cell lines derived from human blastocysts, *Science* **282**, 1145–1147.

Thomson, J.A., Kalishman, J., Golos, T.G., Durning, M., Harris, C.P., Becker, R.A., Hearn, J.P. (1995) Isolation of a primate embryonic stem cell line, *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7844–7848.

Towns, C.R., Jones, D.G. (2004) Stem cells, embryos, and the environment: a context for both science and ethics, *J. Med. Ethics.* **30**, 410–413.

Walters, L. (2004) Human embryonic stem cell research: an intercultural perspective, *Kennedy Inst. Ethics J.* **14**, 3–38.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., Gough, N.M. (1988) Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells, *Nature* **336**, 684–687.

Wilmut, I. (2003) Dolly-her life and legacy, *Cloning Stem Cells* **5**, 99–100.

Wolf, E. (2002) Cloning of animals–perspectives for biomedical research, *Exp. Toxicol. Pathol.* **54**, 161–162.

Xu, C., Police, S., Rao, N., Carpenter, M.K. (2002) Characterization and enrichment of cardiomyocytes derived from human embryonic stem cells, *Circ. Res.* **91**, 501–508.

Xu, R.H., Chen, X., Li, D.S., Li, R., Addicks, G.C., Glennon, C., Zwaka, T.P., Thomson, J.A. (2002) BMP4 initiates human embryonic stem cell differentiation to trophoblast, *Nat. Biotechnol.* **20**, 1261–1264.

Yamashita, J.K. (2004) Differentiation and diversification of vascular cells from embryonic stem cells, *Int. J. Hematol.* **80**, 1–6.

Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., Scholer, H.R. (1996) Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells, *Development* **122**, 881–894.

Zeng, X., Miura, T., Luo, Y., Bhattacharya, B., Condie, B., Chen, J., Ginis, I., Lyons, I., Mejido, J., Puri, R.K., Rao, M.S., Freed, W.J. (2004) Properties of pluripotent human embryonic stem cells BG01 and BG02, *Stem Cells* **22**, 292–312.

Zhang, S.C., Wernig, M., Duncan, I.D., Brustle, O., Thomson, J.A. (2001) In vitro differentiation of transplantable neural precursors from human embryonic stem cells, *Nat. Biotechnol.* **19**, 1129–1133.

Zhou, C.Q., Mai, Q.Y., Li, T., Zhuang, G.L. (2004) Cryopreservation of human embryonic stem cells by vitrification, *Chin. Med. J. (Engl).* **117**, 1050–1055.

Stem Cells

Stewart Sell
New York State Health Department, Albany, NY

1 Properties of Stem Cells 329
1.1 The Fertilized Egg and Embryonic Development 329
1.2 Symmetric and Asymmetric Division 331
1.3 Stem Cells and the Origin of Life 331
1.4 The Cycle of Life and Various Kinds of Stem Cells 333

2 Determination and Potential of Stem Cells 334
2.1 Determination/Differentiation 334
2.2 Terminology of Potential 334
2.3 Dedifferentiation 335

3 Embryonic Stem Cells 335
3.1 Isolation and Culture of Embryonic Stem Cells 335
3.2 Embryonic Stem Cell Therapy 337

4 Nuclear Transplantation and Reproductive Cloning 337

5 Germ Stem Cells 339
5.1 Germ Cell Niches 339
5.2 Teratocarcinoma 340

6 Adult (Somatic) Stem Cells 340
6.1 Normal Tissue Renewal 340
6.2 Plasticity of Adult Stem Cells 341
6.3 Hematopoietic Stem Cells 342
6.3.1 Hematopoietic Cell Lineage 342
6.3.2 Pleoplasticity of Hematopoietic Stem Cells 343
6.3.3 Pleopotency Controversies 344
6.4 Glandular Stem Cells 345

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

6.5 Gastrointestinal (gi) Stem Cells 345
6.6 Skin Stem Cells 345
6.7 Neuronal Stem Cells (NSC) 345

7 Tumor Stem Cells 346
7.1 Embryonal Cells and Cancer 346
7.2 Maturation Arrest 347
7.3 Childhood Tumors 348
7.4 Leukemias as Models of Maturation Arrest 349
7.4.1 Gene Rearrangements and Molecular Lesions 349
7.4.2 Stem Cell Origin of Leukemia 349
7.5 Epidermal Carcinogenesis and Stem Cells 350
7.6 Stem cells and Adenocarcinomas 352

8 Therapeutic Applications of Stem Cells: Promises and Problems 352
8.1 Cellular Therapy 352
8.1.1 Central Nervous System 352
8.1.2 Type I Diabetes 354
8.1.3 Heart 355
8.2 Pleopotency or Fusion 355
8.3 Immune Rejection of Stem Cells 355
8.4 Seed and Soil Hypothesis 356
8.5 Designer Therapeutic Stem Cells 356

Bibliography 357
Books and Reviews 357
Primary Literature 357

Keywords

Asymmetric Division
Cellular division of stem cells wherein one daughter cell remains a stem cell and retains stem cell properties, while the other daughter cell begins the process of determination.

Blastocyst
A stage in the development of the embryo when the daughter cells derived from the fertilized egg (blastomeres) form a cyst. The outer layers of the cyst form the fetal membranes; the inner cell mass develops into the fetus.

Determination
The process by which cells lose potential and gain specialized function. This is also known as *differentiation*.

Embryonic Stem Cells
Cells from the first six to seven divisions of the fertilized egg found in the inner cell mass of the embryo. Embryonic stem cells are isolated from blastocysts before implantation into the uterus.

Potential
The property of a stem cell that allows it to give rise to progeny that are able to differentiate into various cell types. A totipotent cell can give rise to progeny that differentiate into all of the various mature cell types of the organism. A multipotent or pluripotent cell can differentiate into many of the cell types of the organism, but not into all cell types.

Stem Cell
A cell that renews itself after division so that it is never lost by differentiation.

Symmetric Division
Cellular division of stem cells in which each daughter cell remains a stem cell and retains the properties and potential of the mother cell.

A stem cell is a self-renewing cell that is both the first cell of a new organism and the cell that is responsible for the renewal of adult tissues. The fertilized egg is the ultimate stem cell. Following conception, the fertilized egg divides to produce identical daughter cells, which later give rise to progeny that provide all of the cells of the developing organism. These are called *embryonic stem cells* (ESCs). In the adult organism, individual stem cells are responsible for not only normal tissue renewal (*adult stem cells*) but also for sexual reproduction (*germ stem cells*). The adult stem cell divides to give rise to two daughter cells, one of which remains a stem cell; the other begins the process of differentiation. Stem cells in the adult not only replace the mature dying cells of specific tissues but may also have the potential to contribute to broader repair of injured tissues. Stem cells of the adult are also the cells that give rise to cancers. It is anticipated that treatment of genetic diseases and repair of damaged organs will be accomplished by exploitation of the properties of either embryonic or adult stem cells.

1 Properties of Stem Cells

1.1 The Fertilized Egg and Embryonic Development

The fertilized egg is a single stem cell, which divides to give rise to progeny that can differentiate into any of the specialized cells of adult tissues, as well as the membranes and placenta of the fetus (Fig. 1). During early embryonic development, the daughter cells of the fertilized egg, called *blastomeres*, clump together to form a cluster of cells known as a *morula*. At each division, the blastomeres are reduced in

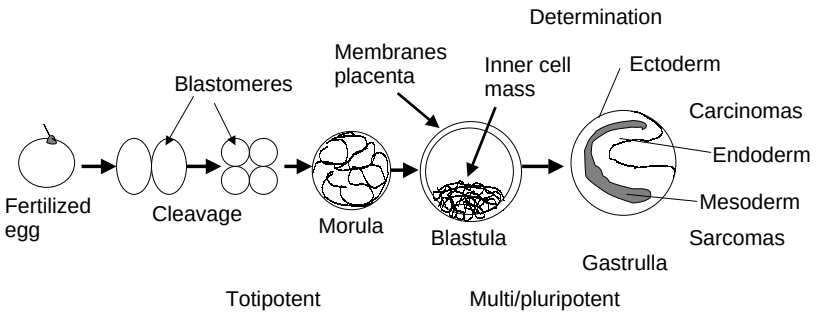

Fig. 1 Early development of the embryo. Division of the fertilized egg results in the formation of daughter cells called *blastomeres*, which form a ball of cells called a *morula*. After six or so divisions, a cavity appears at one end of the morula to form the blastocyst. Until this stage, each of the cells divides evenly to produce two daughter cells, each of which retains the same characteristics, that is, they are totipotent. The inner cell mass of the blastula is the source of ESCs (see below). Invagination of one pole of the blastula leads to formation of the primitive germ cell layers, with formation of the gastrula. During gastrulation and later formation of the fetus, the daughter cells lose potential as they gain specialized function. This process is known as *determination* (see below).

size, but each embryonic blastomere is able to produce all differentiated cell types, that is, it is totipotent. Eventually, as the number of blastomeres approaches 64 cells (six divisions), a cell-free center appears in the expanding cluster of blastomeres, and a hollow sphere of cells is formed, containing a clump of cells at one pole, the inner cell mass. This is the blastula. In humans, this process takes five to six days, as the morula moves down the fallopian tube of the uterus to the uterine cavity. The blastula then implants in the lining of the uterus (endometrium), where the outer layers of the blastula form the fetal membranes and placenta, and the inner cell mass begins the process of determination or differentiation (see below), to produce branches (lines) of cells that form the differentiated organs of the adult. At this stage, the outer cells of the blastocyst are believed to be no longer totipotent as they become committed to the fetal membranes and placenta. However, the inner cell mass remains multipotent until it becomes committed to the cells that make up the germ layers of the developing fetus. As the inner cell mass develops, the daughter cells begin to acquire properties different from one another, so that specific regions are formed that are destined to become different components of the developing embryo, a process known as *gastrulation*. During gastrulation, the blastula is rearranged to form layered "germ" zones known as *ectoderm* (outer skin), *mesoderm* (middle skin), and *endoderm* (inner skin), which are destined to form the adult organs. Skin, dermal appendages (including breast), and neural tissue are derived from ectoderm; connective tissue, muscle, bone, and blood vessels from mesoderm; and the gastrointestinal (gi) tract and glandular organs from endoderm. The relationship of cancer to the development of the gastrula is exemplified by the fact that cancers that arise from endoderm- and ectoderm-derived cells are

called *carcinomas*, whereas cancer that arise from mesoderm-derived cells are called *sarcomas*. For practical purposes, embryonic stem cells (ESCs) are the cells of the inner cell mass that are obtained from the blastocyst prior to implantation.

1.2 Symmetric and Asymmetric Division

ESCs divide to produce daughter cells with the same properties (symmetric division), whereas adult stem cells divide to produce daughter cells with different properties (asymmetric division, Fig. 2). During early embryonic development, symmetric division leads to a rapid exponential increase in the number of ESCs. After the first six or eight divisions, there is a gradual switch from symmetric division to asymmetric division, as differentiated organs are formed. Organogenesis involves both a change in the character of the cells and a loss of cells (programmed cell death, or apoptosis). Finally, in the adult during normal tissue renewal, the number of newly generated cells equals the number of terminally differentiating cells, so that the cell number in the organ remains at a constant level. This is accomplished by asymmetric division, in which one daughter cell retains the properties of the stem cell while the other daughter cell begins the process of determination (differentiation). However, the adult organs can respond to stress or injury by a shift to symmetric division when an increased number of cells is beneficial, such as an increase in the number of blood cells in response to reduced availability of oxygen at high altitudes, or the replacement of lost liver cells after toxic or viral liver injury. How asynchronous division is controlled is not well understood. In germ cells, one pole of a dividing cell contains protein complexes that are different from those present at the other pole of the cell, or one of the poles is anchored to a supporting cell or membrane, whereas the other pole is not, so that one daughter cell remains attached to a surrounding "niche" cell and remains a stem cell, whereas the other does not and begins the process of determination.

1.3 Stem Cells and the Origin of Life

In the fourth century B.C., Aristotle studied the early stages of human development in the uterus and concluded that the embryo was derived from the mother's menstrual blood. This conclusion was consistent with the hypothesis of spontaneous generation that living animals arise *de novo* from slime or decaying matter. This and similar concepts of the origin of life were generally accepted for over 2000 years until the hypothesis that life did not arise spontaneously, but only from preexisting life (*omne vivum ex vivo*), was pronounced by Leydig in 1855. Rudolph Virchow, the father of pathology, extended this to postulate that all cells in an organism are derived from preexisting cells (*onmis cellula e cellula*). Thus, all the cells of the human body arise from a preexisting stem cell, the fertilized egg. In 1864, Louis Pasteur formally disproved the spontaneous generation hypothesis through carefully controlled experiments in which he demonstrated the failure of microorganisms to grow (corruption) in sterilized broth-containing vessels with long necks that prevented ambient organisms from entering.

The dominant hypothesis of the 19th century was that a preformed individual

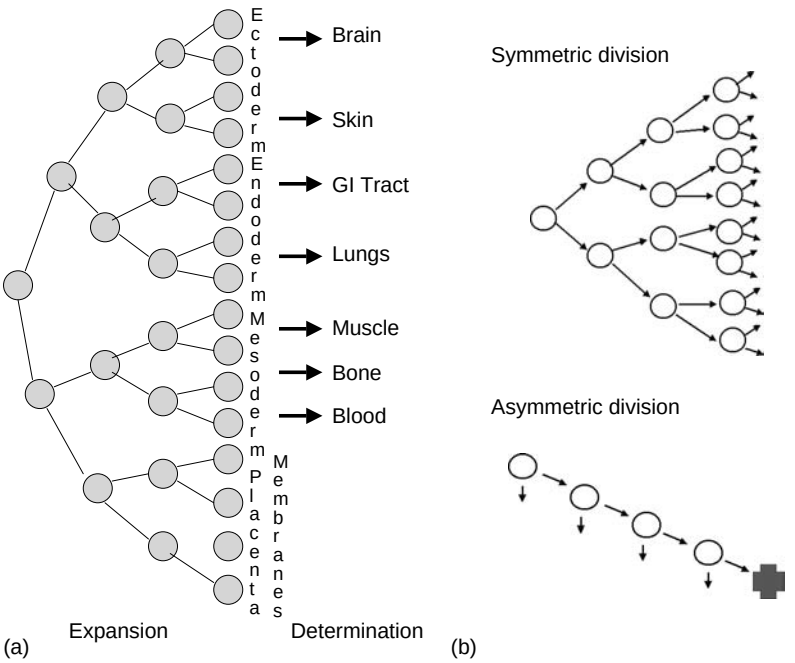

Fig. 2 Properties of stem cells. (a) Stem cells and organogenesis. During embryonic development, the ultimate stem cell, the fertilized egg, gives rise to progeny that remain totipotential as the population expands. Then, determination occurs and the cells begin to lose potential and begin to gain the specialized functions required for the formation of mature organs. As the adult organs form, the stem cells alter the nature of the division, so that one daughter cell begins the process of determination, while the other daughter cell remains a tissue stem cell (asymmetric division). The first determination step results in differentiation of cells that form the membranes and placenta of the developing embryo. Then, the cells remaining in the inner cell mass form the primitive germ layers called *ectoderm*, *endoderm*, and *mesoderm*. The cells in these germ layers then differentiate further into adult organs. (b) Symmetric and asymmetric division. During symmetric division, each cell divides and gives rise to two daughter cells with the same potential. This is a property of ESCs. During asymmetric division, a property of normal tissue renewal in the adult, each progenitor cell gives rise to one daughter cell that remains a progenitor cell plus one daughter cell that begins the process of determination to a terminally differentiated cell (i.e. asymmetric division). The number of cells increases exponentially during early embryogenesis, but the cell number remains constant during normal tissue renewal, as the number of new progenitor cells equals the number of cells destined to die.

resided in the egg or the sperm. This preformed individual was called a homunculus. The homunculus in the egg was activated to develop after stimulation with sperm, or the homunculus in the sperm was activated to develop when provided an appropriate environment in the uterus. By the early 1900s, it had been definitively shown that this concept was incorrect, and that the embryo formed by

the fertilization of an egg that developed in the ovary of the female by fusion with a sperm provided by the male. The product of the union of a sperm with an egg is the primordial totipotent stem cell.

1.4 The Cycle of Life and Various Kinds of Stem Cells

According to the principles derived from Leydig, Virchow, and Pasteur, life, as we know it, neither ends nor begins, but is continuous. The adult human, for example, is only one stage in the cycle of life (Fig. 3). The life of an individual begins with fertilization of the egg through formation of a fetus. Totipotent cells in the developing fetus migrate to the genital ridge. In adults, they produce germ stem cells in the gonads. Germ cells give rise to gametes (egg and sperm) by reduction division (meiosis), resulting in half the genetic information of the adult cells. Genetic reconstitution occurs when the sperm fertilizes the egg. In this process, life is continuous; it is neither created nor destroyed. As the cycle of life is manifested, successive types of stem cells are produced: embryonal, fetal, germ, and adult. The properties and promise of each of these types of stem cells will be presented below.

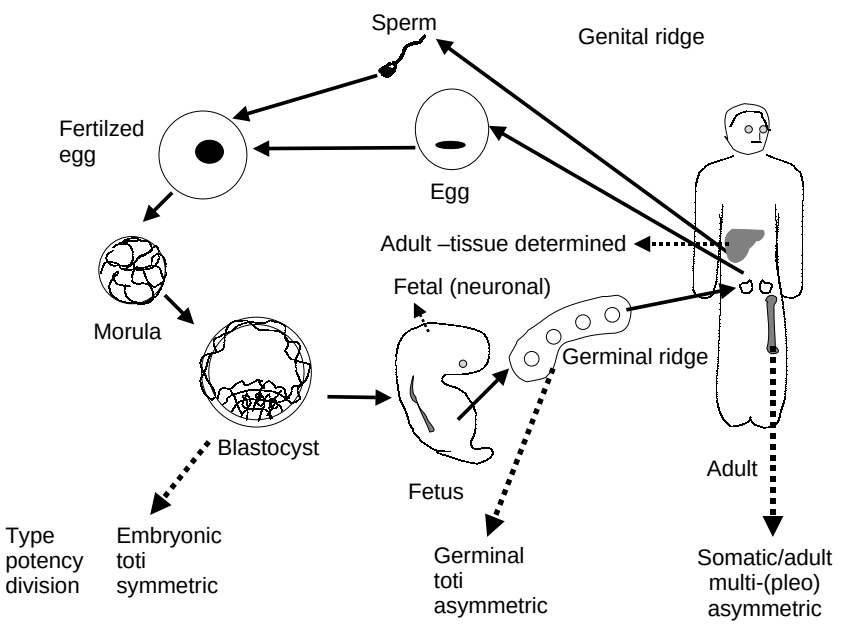

Fig. 3 Stem cells and the cycle of life. Stem cells of various kinds form a cycle of life. The fertilized egg divides to produce ESCs. When the fetus is formed, a set of germ cells develops in the genital ridge that gives rise to the germ cells in the gonads in the adult. These, in turn, produce egg or sperm that unites to form the fertilized egg, thus completing the cycle of life. In addition, there are in the adult various types of stem cells with differing potentials. The various types of stem cells, embryonic, germ, fetal, and adult (somatic), are each the subject of intense investigation for determination of their potential to restore or replace damaged cells (stem cell therapy).

2
Determination and Potential of Stem Cells

2.1
Determination/Differentiation

The process of the loss of potential and the gain of specialized function is known as determination or differentiation. In this process, the *totipotent stem cells* of the blastomere give rise to *multi/pluripotent cells* of the germ layers. These, in turn, give rise to tissue-determined stem cells of the developing organs. A change from symmetric to asymmetric division is a consequence of determination or differentiation. Tissue determination is accomplished in progenitor cells through interaction with other cell types. For example, determination of primitive gut cells into cells of the liver or pancreas takes place in association with developing endothelium and mesenchyme. As the cells become specialized into tissue, their potential is increasingly limited, and they finally differentiate to terminally differentiated cells. The molecular signals controlling determination are not well understood. Cellular signaling pathways, such as Oct-4, Wnt/β-catenin, Notch, BMP (bone morphogenic protein), Janus family kinase, and sonic hedgehog signaling pathways are under intense study for elucidation of their possible roles in determining which cells differentiate and which remain stem cells.

2.2
Terminology of Potential

The terminology used regarding cellular potential is listed in Table 1. The term totipotent should be reserved for those stem cells that can give rise to all of the differentiated tissues of the body as well as the fetal membranes and placenta, that is, the ESCs prior to blastulation. The terms multipotential and pluripotential are often used for the same attribute, the ability of a given stem cell to form many or all of the cells of the body, but not the cells of the fetal membranes, that is, the cells of the inner cell mass. In an official NIH primer released in 2000, pluripotent was defined as "capable of giving rise to most tissues of an organism," and totipotent as "having unlimited capacity" (www.nih.gov/news/stemcell/primer.htm). The potential decreases with progression of development of the embryo. However, as will be described in more detail below, some adult tissue stem cells appear to retain the potential to differentiate into cells of other lineages. Thus, these adult cells express unexpected or excess potential and are termed *pleopotent stem cells* (Table 1, and see below).

Tab. 1 Terminology of potential (plasticity).

Prefix	*Meaning*	*Example*
Toti	All	Embryonal
Multi	Many	Hematopoietic
Pluri	Some	Hematopoietic
Pleo[a]	More, excess	Hematopoietic
Oligo	Few	GI stem cell
Tri	Three	Bronchial lining
Bi	Two	Bile duct
Uni	One	Prostate

Source: The Random House Dictionary of the English Language.
[a] Recent evidence indicates that some cells in the tissues used as examples of determined potential may actually have more potential than was previously appreciated. For example, the skin or bone marrow may contain stem cells with the potential to form tissues of other organs. The term pleopotential is suggested for this property.

2.3 Dedifferentiation

Dedifferentiation, the reversion to a less-differentiated state of a cell after commitment to a more mature functional state, is difficult, if not impossible, to demonstrate in mammals, but it occurs during regeneration of tissues in fish and amphibians, such as Zebrafish, salamanders, and *Xenopus laevis* larvae. During limb regeneration in salamanders, the amputation site is covered by a thin layer of epidermis. Then, the major regeneration of the tissues of the lost limb is accomplished by a group of about 10 000 cells arising by dedifferentiation from the remaining muscle, cartilage, dermis, and glial (Schwann) cells, which form a cluster of undifferentiated cells called a *blastema*. These dedifferentiated cells then redifferentiate to form the mature cartilage, muscle, nerve, and so on, of the regenerated limb. This process appears to be limited to cells of mesenchymal origin as the overlying epidermis is restored by the epidermal layer and not the blastema. Thus, the differentiated muscle, nerve, and connective tissue cells of the salamander limb appear to retain the multipotency of more primitive mesodermal cells, even after differentiation into cells with mature function. It is not known why mammalian limb cells appear not to retain this capacity. One hypothesis for the cellular origin of cancer is that mature cells dedifferentiate to less-mature stages of cellular development, at which growth cannot be controlled. However, as will be discussed in more detail below, this appears less likely than the alternative hypothesis, namely, that cancer represents cells arrested in a stage of maturation at which they continue to proliferate without differentiating into mature nondividing cells (terminal differentiation).

3 Embryonic Stem Cells

3.1 Isolation and Culture of Embryonic Stem Cells

Under optimal conditions, cells from the inner cell mass of the preimplantation embryo (ESCs) are able to proliferate indefinitely and, under inducing conditions, to undergo determination and then to differentiate into the cells of any of the tissue types of the adult. In contrast, after formation of germ layers, most somatic progenitor cells have limited life spans and they exhibit decreasing differentiation potential as mature organs are formed. In practice, ESCs are obtained from preimplantation blastocysts created for purposes of *in vitro* fertilization (Fig. 4). Human oocytes fertilized *in vitro* will grow to the blastocyst stage in culture. Subsequent transfer of the *in vitro* fertilized embryos to the uteri of infertile patients, supplemented with luteal support, eventually led to the successful clinical application of *in vitro* fertilization. During *in vitro* fertilization, several hundred human oocytes are fertilized, and many blastocysts are produced. These are usually frozen, so that they can be used if the first uterine transplantation of four to six blastocysts is not successful. At present, there are hundreds of thousands of frozen human blastocysts available for culture. To obtain ESCs, the outer layer of the cells of the blastocyst is removed, destroying the embryo, and the inner cell mass cells are dispersed and placed into tissue culture. It is a common misconception that ESCs

are obtained from aborted fetuses. This is simply not correct. At the time of induced abortion, the embryo or fetus is far past the blastocyst stage of development where ECSs could be obtained.

Through adjustment of the conditions of culture, addition of different growth factors, or transfection of the cells with genes for transcription factors, the ESCs can be directed to differentiate into various cell types. For example, high glucose in the medium or transcription with pancreatic transcription factors can induce the cells to differentiate into pancreatic cells; addition of fibroblast growth factor (FGF) will induce the cells to become neuronal cells; addition of epidermal growth factor (EGF) will induce formation of glial cells. Through manipulation of the culture conditions, ESCs can be induced to differentiate into blood islands, muscle, connective tissue, neurons, macrophages, liver, and other cell types. ESCs injected into a normal blastocyst will incorporate into the developing embryo to form a chimeric individual; the organs of the

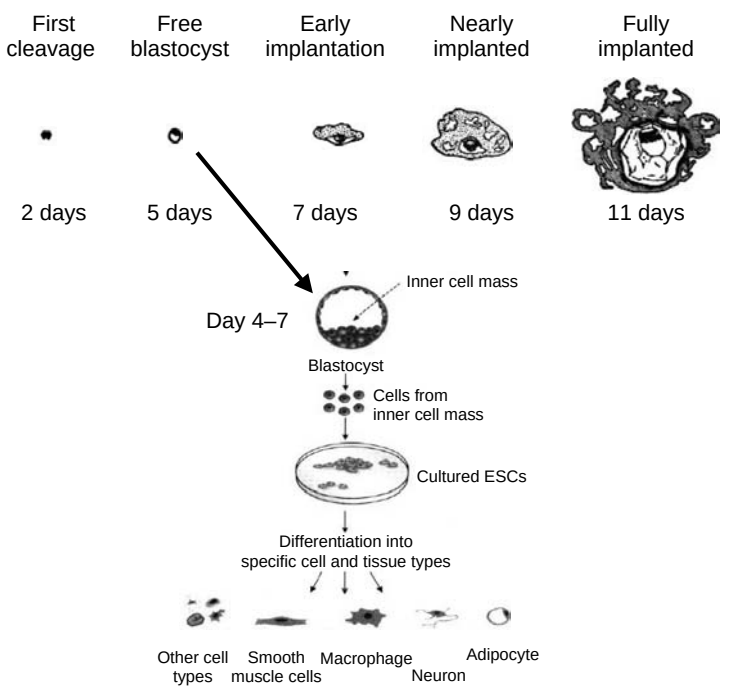

Fig. 4 Isolation and culture of human embryonic stem cells. ESCs are isolated from the preimplantation blastocyst cultured after *in vitro* fertilization. The inner cell mass cells from the blastocyst are cultured in special medium, and single cells are selected to produce cell lines. The cell lines are then cultured with selective growth factors, or transfected with tissue-specific promoters, to induce differentiation to more mature cell types. Modified from: Arey, L.B. (1974) *Developmental Anatomy*, W.B. Saunders Co., Philadelphia, PA and National Research Council – Institute of Medicine (2002) *Stem Cells and the Future of Regenerative Medicine*, Washington, DC, http://stemcells.nih.gov.

individual that develops will contain a mixture of cells derived from the ESCs and the blastocyst. The tremendous potential for the therapeutic use of ESCs is now only limited by our knowledge of how to induce these cells to differentiate into the cell type that we want, when we want them to, and how to manipulate them so that they will incorporate into adult tissues after transplantation.

3.2 Embryonic Stem Cell Therapy

The potential of the *in vivo* use of ESCs for therapy was demonstrated when it was shown that injection of ESCs into lethally irradiated mice could restore the lost bone marrow stem cells. Difficulties in using ESCs for replacement of adult tissues are the low efficiency and long time required by ESC for the process of differentiation into functional adult cells. This may be addressed either by using adult stem cells (see below) or by directing ESCs to a specialized tissue pathway *in vitro* prior to transplantation. ESCs may require a series of signals in order to produce progeny of a more differentiated type. For example, specialized culture conditions, including exposure to and withdrawal from FGF or culture with FGF and other growth factors, allow generation of neural stem cells, which will incorporate into developing mouse brain, and will migrate into damaged zones in the mouse brain and reconstitute the damaged tissues.

Since the ESCs will express histocompatibility markers when they differentiate, ESC therapy is only possible when the donor of the ESCs and the recipient are histocompatible. In preclinical models, cells are transplanted between inbred mouse strains or human cells are transplanted into immunodeficient mice, which will not reject the human cells. In order to carry out successful ESC therapy between histoincompatible individuals, it is necessary to obtain ESCs that bear the same histocompatibility genes as the recipient. The method proposed to meet this requirement is nuclear transplantation.

4 Nuclear Transplantation and Reproductive Cloning

ESCs with genetic characteristics of an adult individual may be obtained by transplantation of an adult nucleus into an enucleated oocyte from a different individual (Fig. 5). The microenvironment of the egg allows expression of the genes of the transplanted adult nucleus and results in development of a "cloned" embryo, which carries the genotype of the nuclear donor. The egg with the transplanted nucleus is cultured *in vitro* to the blastocyst stage. At this point, the cells of the developing blastocyst may be cultured as ESC lines, which can then be transplanted back into the adult nuclear donor without immune rejection. If the donor is lacking a critical gene, it can be transfected into the ESCs and then transplanted into the adult, thus effecting nuclear transplantation. The blastocysts may also be transplanted into the prepared uterus of a surrogate foster mother, in whom they may develop into a mature individual that is a clone of the nuclear donor (reproductive cloning). In experimental animals, this procedure has a very low success rate (generally less than 1%). Although reproductive cloning has been accomplished in several mammalian species, including primates, there are inherent problems in the procedure as the cloned individual exhibits accelerated

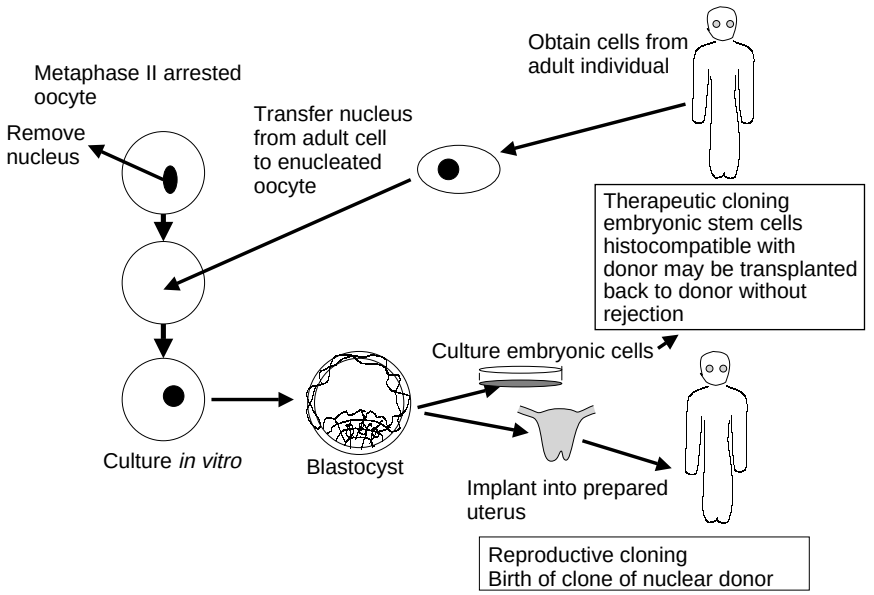

Fig. 5 Nuclear transplantation and cloning. Nuclear transplantation and cloning are accomplished by transferring the nucleus of an adult cell to a fertilized egg from which the nucleus has been removed. The developing egg is able to reactivate expression of the genes of the donated nucleus and to proliferate so as to form a blastocyst, which will have the same nuclear genes as the adult donor. The ESCs of this blastocyst can now be injected back into the donor of the nuclei and the ESCs will not be rejected, because they will express the same major histocompatibility antigens (MHCs) as the recipient. These ESCs could be used to replace a damaged organ, or they could be transfected with a replacement gene that is malfunctioning in the recipient (therapeutic cloning). The blastocyst may also be transplanted into a prepared uterus to develop into a clone of the nuclear donor (reproductive cloning is given in Table 2).

aging. Nuclear transplantation is generally viewed as ethically acceptable, since the blastocyst has been created for the purposes of *in vitro* fertilization and is destroyed before implantation. Thus, a truly viable embryo is not produced. However, reproductive cloning is not considered acceptable for application to humans.

The critical role of the interaction of the cytoplasmic environment of the oocyte and the transplanted nucleus of the donor cell is exemplified by experiments in which the nucleus of a mature cell is transplanted into the cytoplasm of a less-mature cell. In classic embryology, determination was understood as a one-way process: once a cell is fully differentiated in the adult, it is stable and not able to dedifferentiate. Thus, it was thought that the only nuclei from adult cells that could be programmed to produce a new organism would be those of a tissue stem cell. In experiments using nuclei from tadpoles, the potential of the nucleus was progressively lost during gastrulation. In the case of the cloning of sheep, for example, Dolly, nuclei from a day-nine embryo, a 26-day fetus, and the mammary gland of a six-year old ewe in the last trimester of pregnancy

were used; the use of the latter donor implies that nuclei from adult mammals retain all of the potential of the germ cells. However, the exact nature of the cells in the mixture used for cloning that contribute the nuclei, in the case of the transplant from mammary gland of the pregnant ewe, was not identified. The mammary epithelium during the third trimester is actively proliferating, and the proliferating gland contains large numbers of active progenitor cells. Since the number of nuclei in the mature adult tissue that can do this appears to be vanishing small, it is likely that those nuclei from an adult tissue that are capable of "reprogramming" in an oocyte may actually be the tissue progenitor cells or even the putative circulating totipotent stem cells of the adult.

On the other hand, under some circumstances, the differentiated state appears to be reversible. Transfer of nuclei from differentiated adult cells into an oocyte can result in the restoration of the totipotency of the nucleus in the oocyte. This appears to be true even for mature T and B cells, which have a rearranged T-cell receptor and immunoglobulin genes respectively. Every tissue in mice cloned from a B cell had the same immunoglobulin gene rearrangement as the original donor nucleus.

5 Germ Stem Cells

Germ stem cells are those cells in the reproductive organs that give rise to sperm or eggs. In practice, germ cells from the adult are not usually used for therapy; germ cells from the genital ridge of fetal animals are more readily obtained. Even so, it is more difficult to obtain and culture these cells than ESCs, and much less work has been done on them.

5.1 Germ Cell Niches

Early in embryogenesis, a few cells are designated to become germ cells; the cells that differentiate into egg or sperm in the adult individual. These cells migrate into the primitive gonad (genital ridge) and differentiate into female or male germ cell precursors, depending on the presence of two X chromosomes (female) or one X and one Y chromosome (male). In *Drosophila*, germ cells, both in the female's ovary and in the male's testes, reside in specialized tissue (niches) that control the division of the stem cells; thus, after mitosis, one daughter cell remains a stem cell and the other becomes a transit-amplifying cell that enters the process of producing sperm or egg cells (asymmetric division). During this process, the germ cells undergo reduction division or meiosis, whereby the number of chromosomes in each cell is reduced by half, so that each gamete contains a single set of chromosomes. Other cells in the organism contain two sets of chromosomes. When the egg is fertilized by sperm, the nuclei of the egg and sperm combine to reform a complete set of two chromosomes. In the testes, the germ cells surround a central core of supporting somatic cells known as the *hub*; in the ovary, the germ cells are wrapped inside somatic cells in a structure known as an *ovariole*. Upon division, the daughter cell next to the hub or at the end of the ovariole remains a stem cell, whereas the other distal daughter cell begins the process of determination to sperm or egg. Germ stem cell niches are critical for normal maturation and asymmetric division of germ cells. When germ cells are

displaced from their niches, teratocarcinomas may develop, indicating a role for the tissue microenvironment in controlling cell proliferation and differentiation.

5.2 Teratocarcinoma

Germ stem cells will express uncontrolled proliferation, haphazard differentiation, and tumor formation when growth controls normally placed on the developing germ cells are missing. For example, transplantation of germ cells from adult testes of mice into the genital ridges of 21-day-old fetal mice results in uncontrolled differentiation and tumor formation (teratocarcinoma). Teratocarcinomas are cancers of mixed cell types and may contain essentially all cellular elements of adult tissues, including brain, bone, teeth, bone marrow, eyes, endocrine glands, muscle, skin, and intestine, arising from all three germ layers as well as placenta and yolk-sac tissue.

The environment of the normal blastocyst can control the malignant potential of teratocarcinoma tumor cells. After transplantation into the inner cell mass of normal developing murine blastocysts, malignant teratocarcinoma cells will incorporate into the developing embryo with production of mosaic or chimeric mice, which contain normally functioning tissues derived from the tumor cells. These studies not only demonstrate the totipotential of germ cells and the teratocarcinomas derived from them but also the principle that the differentiation potential of totipotent cells – and in fact their ability to produce cancers – is controlled by appropriate environmental signals. These findings raise the possibility that other types of stem cells, including ESCs, will also undergo loss of growth control and will produce cancer if placed in the "wrong" environment.

6 Adult (Somatic) Stem Cells

The cells in normal adult organs are continuously replaced through a series of asymmetric cell divisions of adult stem cells. The rate of renewal varies in different organs. For example, skin, gastrointestinal epithelial cells, and some blood cells (polymorphonuclear cells) will be completely replaced in two weeks or less, whereas replacement of other tissues such as liver (approximately two years) or brain (a lifetime) are much more slowly replaced, if at all.

6.1 Normal Tissue Renewal

Normal replacement of adult tissue cells is accomplished by proliferation of tissue-determined stem cells, which differentiate into more mature cells that can no longer proliferate, and eventually die (Fig. 6). In the process of normal tissue renewal, in contrast to the symmetric division that occurs during embryologic development, division is asymmetric, with one daughter remaining a stem cell and the other beginning the process of determination. Normal renewal is accomplished by cohorts of transit-amplifying cells or progenitor cells, which are the daughter cells of the more primitive stem cell. The most primitive tissue stem cell normally does not divide, but it can be stimulated to proliferate if there is an increased demand for more cells (see hematopoietic progenitor cells, Sect. 6.3 below). When a tissue progenitor cell divides, one daughter cell remains a progenitor cell, whereas the other begins the process of determination, eventually producing a terminally differentiated cell, which in time dies. To maintain the total number of cells in an adult organ in equilibrium, the number of progenitor cells

that divide must equal the number of cells that differentiate and die. A comparison of the rates of normal tissue turnover in various organs is given in Table 2.

6.2 Plasticity of Adult Stem Cells

In classic embryology, determination is considered a one-way process. For example, once primitive endoderm becomes committed to forming liver, it is not able to dedifferentiate or "transdifferentiate" into another tissue type. However, there is increasing evidence for an unexpected plasticity of tissue stem cells. Adult stem cells of a given tissue, for example, blood, are able to differentiate into cells of other tissues, including liver, brain, and skin, of other germ layers. This is unexpected plasticity (*pleoplasticity*). Of particular importance is the presence of pleoplastic stem cells in the bone marrow that could be used for replacement of damaged cells in other organs of the same individual (autologous). Thus, it might be possible to clone cells from the bone marrow that could be exposed to selective determination signals *in vitro* and then be transplanted back into the same individual to replace lost cells in a specific organ. However, it has not been possible to obtain clones of bone marrow–derived stem cells with this potential. On the other hand, intravenous injection of

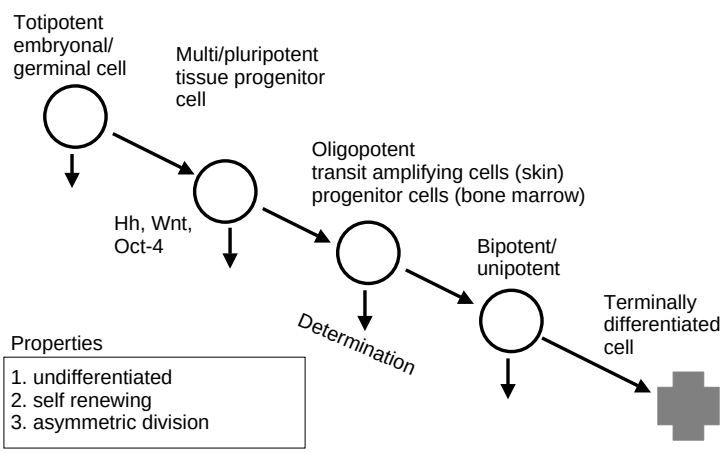

Fig. 6 Stem cell model of tissue renewal. Stem cells are part of a continuum of cell lineages with decreasing potential and increasing differentiated function. Depicted is a postulated lineage of cells including ESCs, tissue stem cells, transit-amplifying cells (progenitor cells), and terminally differentiated cells. ESCs give rise to tissue-determined stem cells during fetal development. In adults, the tissue-determined stem cell is usually quiescent and does not proliferate unless there is an increased need to produce new tissue cells. Normal tissue renewal is accomplished by transit-amplifying cells, which are the daughter cells of the tissue stem cells. In rapidly renewing tissues such as the bone marrow or gi tract, the population of transit-amplifying cells is relative large, whereas in slowly renewing organs, such as the liver, the number of these cells is very low. In normal tissue renewal, the number of new tissue cells produced essentially equals the number of cells that are lost by apoptosis or sloughing, so that the total number of cells in the tissue remains the same.

Tab. 2 Turnover and potential of progenitor cells in some adult organs.

Organ	*Replacement time*	*Potential*
Bone marrow	2–3 days (polys)	Multi/pluripotential (?pleo)
GI lining	2–4 days	Quadripotential (4 types)
Skin	2 weeks	Bipotential
Liver	1–2 years	Bipotential
Brain (neurons)	Lifetime	Unipotential[a]

[a]Mature neurons cannot divide, but neuronal progenitor cells are present in some parts of the adult brain. A generalization that can be made from these data is that the faster the cell turnover in a given organ, the more likely it is that that organ contains a stem or progenitor cell with the potential to form other cell types.

large numbers of autologous bone marrow cells allows the organ to select a few cells among the millions of transplanted bone marrow hematopoietic cells that might be able to differentiate into the cells of that organ's tissue.

6.3 Hematopoietic Stem Cells

6.3.1 Hematopoietic Cell Lineage

One of the most rapidly replaced tissues in the body is the blood. The hematopoietic or blood-forming cells are located in the bone marrow. The existence of an undifferentiated stem cell for blood cells ("gemeinsame Stammzelle") was hypothesized as early as 1917. This hematopoietic stem cell's lineage extends from the resting stem cells to transit-amplifying precursor cells to mature circulating blood cells (Fig. 7). The resting stem cell normally does not proliferate; production of new blood cells is accomplished by proliferation of the precursor or transit amplifying cells. The ability of the bone marrow stem cells to form colonies of various types of blood cells was first recognized when cells from the bone marrow would form colonies of cells in the spleens of histocompatible recipients. The number of stem cells was determined by counting the number of colonies formed in the spleen (colony-forming units). Through the use of such techniques, it was possible to show that a very small number of stem cells could not only reproduce themselves, but also give rise to the complete spectrum of phenotypic colonies of blood cells. Less than 0.05% of the cells in the bone marrow are pluripotent hematopoietic stem cells, most of the proliferating cells are transit-amplifying cells.

The number of transit-amplifying cells in the bone marrow varies, depending on the hematopoietic cell lineage. The survival time of circulating blood cells varies from polymorphonuclear cells, which only live for a few days – to red blood cells, which may survive for many months – to selected populations of lymphocytes, which may survive for years. The majority of circulating blood cells cannot proliferate. For example, red blood cells exclude their nuclei and polymorphonuclear cells inactivate their nuclei and condense them in clumps, so that neither of these cells can be stimulated to proliferate. Because of the rapid loss of polymorphonuclear cells, the number of transit-amplifying cells belonging to the polymorphonuclear

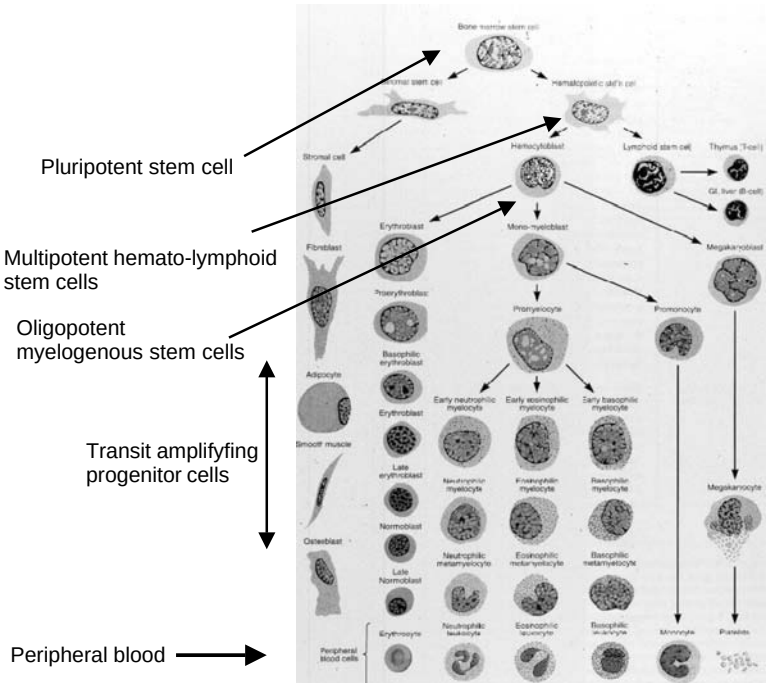

Fig. 7 Diagram of hematopoietic cell lineage. The bone marrow contains a pluripotent hematopoietic stem cell that gives rise to multipotent mesenchymal and hematopoietic lineages. The hematopoietic progenitor cell is believed to produce progeny that enter the blood cell pathway (hemocytoblast) or the lymphocytic pathway (lymphoid stem cell). The hemocytoblast is the common progenitor cell for red blood cells, white blood cells, and megakaryocytes (platelets). The white blood cell precursor (monomyeloblast) gives rise to transit-amplifying cells for monocytes (promonocytes) and polymorphonuclear cells (promyelocytes). The myelocytic series of transit-amplifying cells produces progenitors for neutrophils, eosinophils, and basophils. During this process, there is an exponential increase in numbers of cells at each stage. Thus, the totipotent stem cell is extremely rare, whereas there are enormous numbers of precursors for red blood cells and polymorphonuclear neutrophils in the bone marrow.

lineage (myelocytic series) in the bone marrow is very large (billions). Also, even though the survival time of red blood cells is much longer, there are approximately 1000 red blood cells for each white blood cell. Thus, the number of precursor cells for red blood cells is also very large. These are present in the bone marrow as transit-amplifying blast cells and their progeny at various stages of differentiation into mature polymorphonuclear cells or red blood cells.

6.3.2 Pleoplasticity of Hematopoietic Stem Cells

Until recently, the most primitive bone marrow progenitor cell was believed to be pluripotent, giving rise to all types of blood cells, including lymphocytes, red blood cells, monocytes, and megakaryocytes, as

well as a mesenchymal progenitor cell that can give rise to many other cell types, such as hematopoietic supporting stromal cells, osteocytes, adipocytes, and muscle cells. Now, it has been found that after bone marrow transplantation, the donor bone marrow cells may give rise to circulating cells that have the potential to differentiate into endothelium, muscle, liver, heart, brain (astrocytes and neurons), type I and type II pneumocytes, kidney, retina of the eye, and other organs. Serial transplantation indicates that a single bone marrow cell may be able to give rise to many different tissue types, and suggests that there must be a common precursor not only for the stromal and hematopoietic lineages but also for other germ layer-derived cell types. It is this putative pleopotent bone marrow cell that has stimulated the great revival of interest in adult stem cells in the last few years. The pleopotent stem cells arising from the bone marrow of adults may have just as much promise as do ESCs for stem cell therapy. However, the pleoplasticity of bone marrow stem cells remains controversial.

6.3.3 Pleopotency Controversies

There are conflicting data regarding the pleopotential of tissue stem cells, including bone marrow stem cells. For example, although it was reported that transplanted bone marrow stem cells provided new endothelium in transplant models of atherosclerosis, it now appears, after careful analysis, that the new endothelial cells were actually derived from preexisting endothelium. In a study of muscle-derived cells contributing to bone marrow, it was found that the muscle-derived bone marrow stem cells were derived from hematopoietic stem cells present in the muscle, and not from bone marrow–derived cells. In addition, *in vitro* culture of adult mouse brain or bone marrow stem cells with ESCs generates hybrid cells containing chromosomes of both adult and embryonic cells (cell fusion), and studies using a mouse injury model indicate that bone marrow–derived cells can fuse with liver cells during regeneration. On the other hand, differentiation of liver cells from bone marrow cells without fusion has been reported in other models. Injection of autologous bone marrow cells into the artery supplying a zone of infarcted heart muscle appears to increase healing by supplying nuclei through fusion of bone marrow–derived mononuclear cells *in situ* with precursors of heart muscle cells (satellite cells). However, others are unable to reproduce these findings and found no contribution of bone marrow–derived cells to heart muscle after injury.

Assuming that pleopotent bone marrow stem cells actually exist, a number of basic questions remain unanswered. Is there a single cell type in the bone marrow that is pleopotent, or are there multiple types of progenitor cells that can give rise to different cell types in other organs? There is some evidence that there is a single cell that can do it all. Does the progenitor cell normally circulate, or is it liberated from the bone marrow by humoral signals after injury? The evidence thus far is that this cell circulates normally, but its numbers may increase after injury. Are there normally progenitor cells in mature organs, such as liver or kidney, which can recirculate back to the bone marrow? Are these cells derived from the bone marrow, or can pleopotent progenitor cells be generated in differentiated organs? The ability of brain cells to become blood cells suggests that the answer to this question is yes, but recirculation of blood-derived cells to the brain and back to the bone marrow

cannot be ruled out. Can progenitor cell lines suitable for transplantation to repair damaged organs or for delivery of gene therapy be derived from bone marrow or other organs? Extensive culture of neural stem cells (NSCs) may be required before such cells can differentiate into blood cells, suggesting that rare transformation events may occur when cultured *in vitro* and may account for neural to blood differentiation. Answers to these questions will undoubtedly be forthcoming in the next few years.

6.4 Glandular Stem Cells

Glands consist of ducts draining collections of cells known as *acini*. During development, glands such as salivary glands, breast, pancreas, liver, and prostate arise from endodermal outpouchings that bud from the endodermal layer into the adjacent mesoderm and interact with the mesoderm to form a gland. In the adult, the epithelial cells of the gland are continually renewed from progenitor cells. For the most part, these progenitor cells are located in the basal layer of the duct and at the junction of the duct and the gland. An example of the ability of a glandular stem cell to proliferate and differentiate is the response of the breast to the stimulus of pregnancy. In the normal breast, the glandular tissue is only a duct, which terminates in a pouch of duct cells (terminal duct cells). During pregnancy, the terminal duct cells first proliferate and then differentiate into milk secreting cells. After weaning of the offspring, the glandular cells die, and the breast tissue reverts to the ductal form.

6.5 Gastrointestinal (gi) Stem Cells

The glands that line the gi tract are derived from the endodermal germ layer. They also consist of ductlike structures: the base of the gland, which is a tubular structure called the *crypt*, and the top, which forms fingerlike structures (villi) that extend up into the lumen of intestine. The progenitor cells are located near the base of the crypt (Fig. 8). These cells give rise to cohorts of approximately 40 to 60 new cells per day, which then move up the walls of the crypt to be sloughed off into the intestinal lumen in four to five days. Similar to the hematopoietic progenitor cells, the proliferating cells in the GI tract are transit-amplifying cells. The true GI stem cell most likely resides near the base of the crypt and is not actively proliferating.

6.6 Skin Stem Cells

The lineage of epidermal cells includes the most primitive skin progenitor located in the bulge of the hair follicle, bipolar stem cells in the basal layer of the stratified epithelium (basal cells), and committed transit-amplifying cells in the basal layer and in a few layers of cells above the basal layer. The basal layer is also called the *stratum germinativum* (where the dividing cells are located). The layer above is termed the *stratum spinosum*, because the intercellular bridges joining these squamous cells are visualized as spines microscopically (Fig. 9). Above this layer, there is increasing keratinization of the cells and loss of the ability to proliferate.

6.7 Neuronal Stem Cells (NSC)

Until recently, the adult brain was not believed to contain stem cells. Now, NSCs have been isolated from many parts of the central nervous system (CNS), including the forebrain subventricular

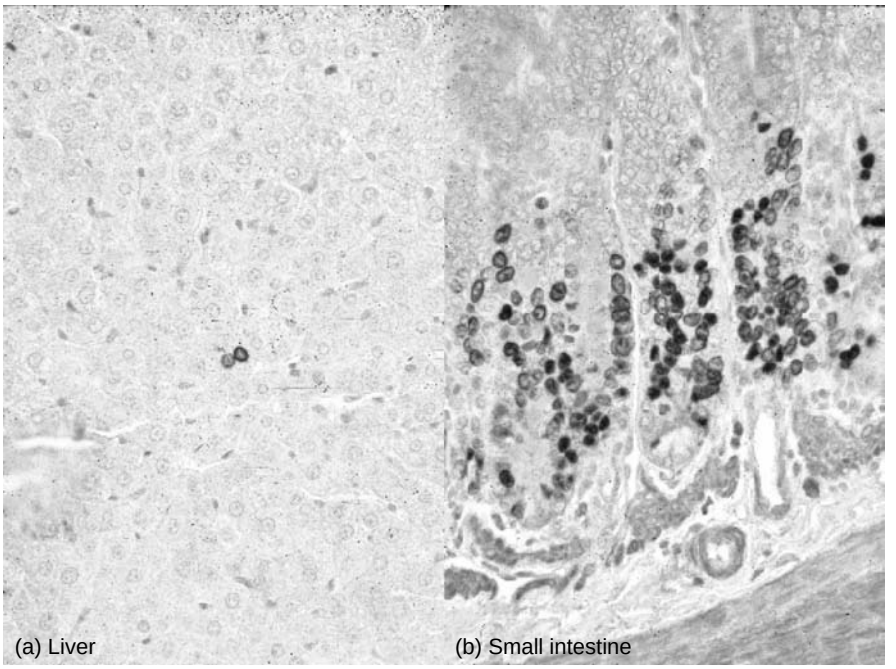

Fig. 8 Comparison of proliferating cells in liver and small intestine The dark-labeled nuclei indicate the cells that are proliferating in the liver (a) and small intestine (b) during a 24-h period. It is estimated that there are about 1 : 40 000 cells dividing in the liver at any given time, whereas millions of cells are dividing in the gi lining. The labeled cohort of cells, in panel B, will migrate up the intestinal crypt and will slough off into the intestine in four to five days.

zone, hippocampus, spinal cord, corpus striatum, septum, cortex, and optic nerve. These cells can be cultured as monolayers or as floating aggregates called neurospheres (see below). It is now clear that self-renewing NSCs not only are present in the adult brain but also can be derived from ESCs as well as other types of adult stem cells, such as hematopoietic stem cells. Multipotent neuronal stem cells are self-renewing, are able to differentiate into cells of all glial and neuronal linages, and populate developing or degenerating CNS. *In vitro* NSCs may be directed to differentiate into various cells types by addition of specific growth factors and can then be used for treatment of CNS lesions (see below).

7 Tumor Stem Cells

7.1 Embryonal Cells and Cancer

The idea that an adult human might harbor primitive ESCs was first conceived from studies on the pathology of cancer. With the invention of the microscope, pathologists by the mid-nineteenth century began to recognize that the cells that make up cancer tissues resembled embryonic tissue; Durante in Italy and Cohnheim in Austria hypothesized that tumors arose from embryonic cells present in the adult, a theory known as the embryonal rest theory of cancer. Rudolph

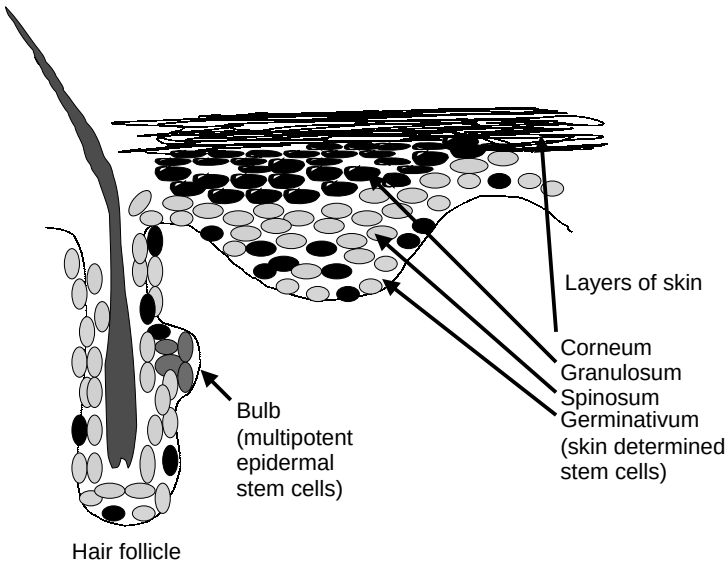

Fig. 9 Cellular layers of the skin (epidermis) and tissue renewal. Illustrated are a hair follicle and the layers of the epidermis, designated by their Latin names. The most primitive stem cells are located in the bulb of the hair follicle. When cultured *in vitro*, these cells can be passaged for many more generations than can any of the other cells derived from the epidermis. These cells give rise to various specialized appendages in the dermis, such as sweat and sebaceous glands, as well as hair follicles. The proliferating transit-amplifying cells of the skin are located in the basal layer (stratum germinativum) or in the lower layers of the stratum spinosum (indicated in black). There is then progressive accumulation of cytokeratin in the stratum granulosum associated with terminal differentiation and loss of cell structure, as the outer keratinized layer (stratum corneum) is formed. All of the cells of the skin, except for the nonproliferating resting stem cells in the basal layer, are replaced every two weeks.

Virchow observed that teratocarcinomas were made up of tissues, also found in developing embryos. The term ''dedifferentiation'' was coined by pathologists to describe this resemblance, and pathologists today still refer to the ''degree of dedifferentiation'' to grade tumors as fast growing (poorly differentiated) or slow growing (well differentiated). However, it is unfortunate that this term implies that tumors arise from dedifferentiation of mature tissues, which is unlikely.

7.2 Maturation Arrest

A cancer may arise from any cell in the body that can divide from a totipotent germ cell (teratocarcinoma) to a monopotent, tissue-determined cell. G. Barry Pierce postulated that the degree of differentiation depends on the stage of differentiation at which the majority of the cancer cells become arrested (Fig. 10). One interesting theory was that primitive sex cells (germ cells) could wander through the tissue

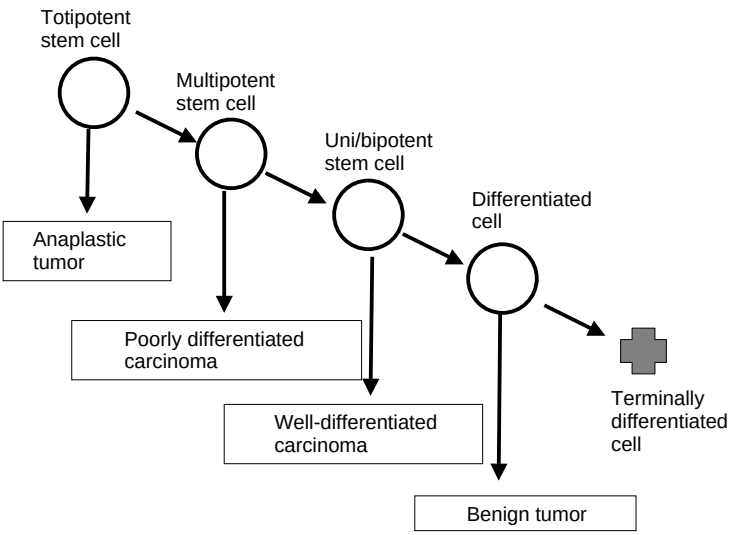

Fig. 10 Hypothesis that characteristic of a cancer is related to the stage of maturation arrest. The type of cancer in a given organ is related to the stage of determination at which the cancer arises. If it is a totipotent undifferented cell, the cancer will be anaplastic (without form). If it is a pluripotent cell, it is a poorly differentiated tumor. If it is a well-differentiated cell, it is a well-differentiated tumor. If a cell is late in differentiation, it is a benign tumor.

of the developing embryo and randomly lodge in organs other than the sex glands. In such locations, they could give rise to tumors that would express the phenotype of the tissue in which they were located, due to loss of the controlling effects of the gonadal microenvironment.

7.3 Childhood Tumors

Tumors arising in children provide clues to the relationship between "residual" embryonic cells and cancer. In the late 1800s, a tumor of the kidney in children was found to contain a mixture of embryonal cell types. In 1899, Wilms wrote a definitive monograph on this tumor and concluded that it arose from a fragment of the primitive undifferentiated mesodermal tissue, which normally is the analage of the kidney. Some of the cells in the tumor look like embryonal kidney and others like connective tissue, with a mixture of undifferentiated spindle cells and immature epithelial tubules. Because of this, it is also called a *mixed tumor* or *nephroblastoma*. Wilms' tumors appear to arise from embryonal cells in infants; these "embryonic rests" differentiate and involute with aging so that they are no longer a source of cancers after seven years of age.

A similar embryonic tumor of sympathetic ganglia, called *neuroblastoma*, originates in fetal neural crest cells of the sympathetic nervous system. These cancers have a peak age of occurrence at two years and are composed of NSCs that become more differentiated with age. In young infants, these tumors are undifferentiated and aggressive, whereas, in older children, they appear as more differentiated tumors or even as benign, well-differentiated

ganglion cells. It appears that the undifferentiated NSCs giving rise to these tumors can be differentiated into benign mature cells by the environment of the growing infant.

7.4 Leukemias as Models of Maturation Arrest

The term "leukemia" means white blood. It is used for cancers of the hematopoietic system, in which the number of white blood cells becomes so large that the blood appears creamy or white.

7.4.1 Gene Rearrangements and Molecular Lesions

The growth characteristics, morphology, and clinical behavior of human leukemias are related to molecular lesions caused by gene rearrangements that are found in all of the cells of the tumor, suggesting an origin in a single cell that underwent the original rearrangement. The genetic change most likely occurs in the primitive myeloid stem cells, but it is expressed when the maturing cells reach the level of differentiation at which point the translocation is activated (see Table 3 and Fig. 11). In chronic myelogenous leukemia, there is a translocation of bcr (breakpoint cluster region) to abl (Abelson tyrosine kinase), resulting in maturation arrest at the neutrophilic myelocytic series of cells. In acute myelogenous leukemia, there is constitutive activation of the IL-3 receptor, leading to continued proliferation of cells throughout the myeloid series, beginning at the level of the hemocytoblast and also including the transit-amplifying cells in the myeloid lineage. In acute promyelocytic leukemia, there is formation of a fusion protein that includes the retinoic acid receptor, which blocks differentiation at the promyelocytic stage. Rearrangement of the immunoglobulin G (IgG) promoter to the *bcl-2* or *c-myc* gene leads to expression of the product of these genes in B cells which then develop into B-cell tumors. Although the mutation may be present in all of the cells of the body, activation of the IgG promoter only occurs in B cells. Thus, manifestations of the IgG promoter rearrangements are seen in B-cell tumors.

7.4.2 Stem Cell Origin of Leukemia

The response to chemotherapy of acute leukemias also implies the involvement of a primitive stem cell, since repeated cycles of antiproliferative treatment are required. Since this therapy acts on actively proliferating cells, it would be predicted that one cycle of drug treatment would suffice to eliminate the actively

Tab. 3 Some gene rearrangements in leukemias and lymphomas.

Leukemia type	*Chromosomal rearrangement*	*Translocation*	*Function*	*Cell type*
Chronic myelogenous	t9:22	bcr/abl	Tyrosine kinase	Neutrophilic myelocyte
Acute myelogenous	t8:21	IL-3R	Tyrosine kinase	Hemocytoblast
Acute promyelocytic	t15:17	PML/RARα	Retinoic acid receptor	Promyelocyte
Burkitt's (B cell)	t8:14	IgG/myc	Transcription	B cell
B-cell lymphoma	t14:	IgG/bcl2	Blocks apoptosis	B cell

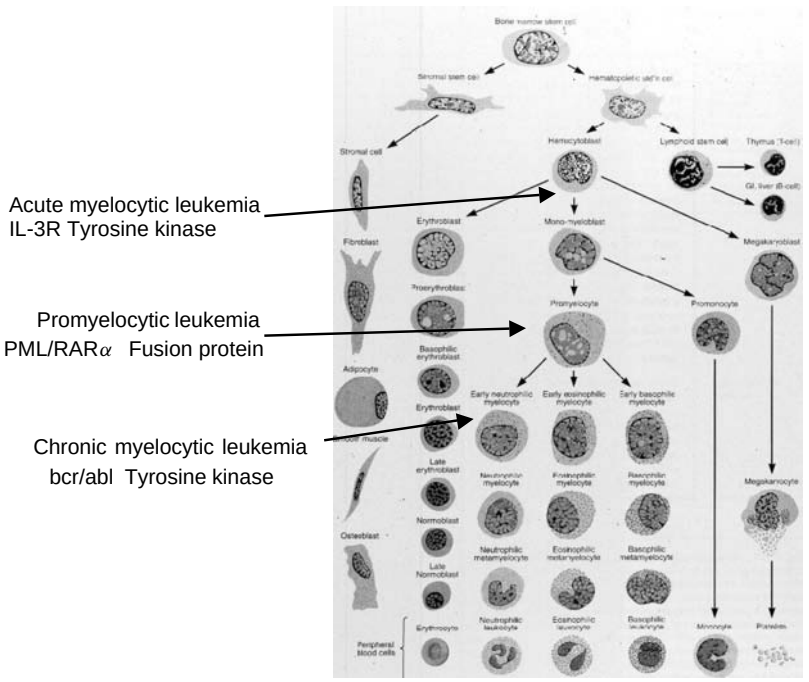

Fig. 11 Molecular defects and maturation arrest in leukemias. Gene rearrangements represent the major molecular lesions of leukemia. The examples shown here illustrate the stage of maturation at which the molecular lesions are expressed. For acute myelogenous leukemia, there is constitutive activation of the IL-3 receptor (tyrosine kinase). This causes proliferation of cells in the monomyelocytic series, as IL-3 is activated during activation of cells from the monomyeloblast through the monocyte, promyelocyte, and proerythrocyte series. For promyelocytic leukemia, there is formation of a fusion protein, which results in a block in differentiation at the promyelocytic cell. In chronic myelogenous leukemia, there is activation of abl kinase in cells of the neutrophilic series (neutrophilic myelocytes).

proliferating cells. However, the genetic damage is also present in the resting stem cells that are not proliferating at the time when the drug is given. When one cycle of treatment is completed, these resting cells will be stimulated to proliferate. When in a resting state, they will not be susceptible to treatment. Successful treatment may thus require four to five cycles of antiproliferative therapy to catch all of the leukemic cells. Even so, up to 40% of patients may require total bone marrow ablation and stem cell transplantation from a normal donor. This is most likely because the genetic lesion is present in the primitive resting stem cell in the bone marrow, which is not susceptible to antiproliferative treatment. The malignant stem cells must be destroyed by irradiation, or else the leukemia will reappear.

7.5 Epidermal Carcinogenesis and Stem Cells

Perhaps the best evidence that tumors arise from stem or progenitor cells is

provided by the two-step model of skin cancer: initiation and promotion (Fig. 12). In the classic model, benz(o)pyrene, the initiator, is painted onto the skin. This chemical binds to DNA in the skin cells, causing a permanent genetic alteration. However, cancers will not arise unless a proliferative stimulus is administered. This is provided by treating the skin with phorbol ester. Thus, the initiation induces genetic damage, and the promoter stimulates the damaged cells to proliferate, leading to cancer. Initiation must be done before promotion. If promotion is done prior to initiation, cancer will not develop. The elapse of time between initiation and promotion clearly indicates that the self-renewing stem cell is the target of initiation. Months or even years can separate initiation and promotion. Thus, in order for the tumors to grow out in this model, the initiated cells must survive from the time of initiation to the time of promotion. Given the well-established fact that all cells in the skin, except for the long-term, self-renewing stem cells, turn over completely every two to three weeks, the only way that the initiated cells could still be present if months or years elapsed between initiation and promotion would be for initiation to have occurred in the self-renewing progenitor-cell population.

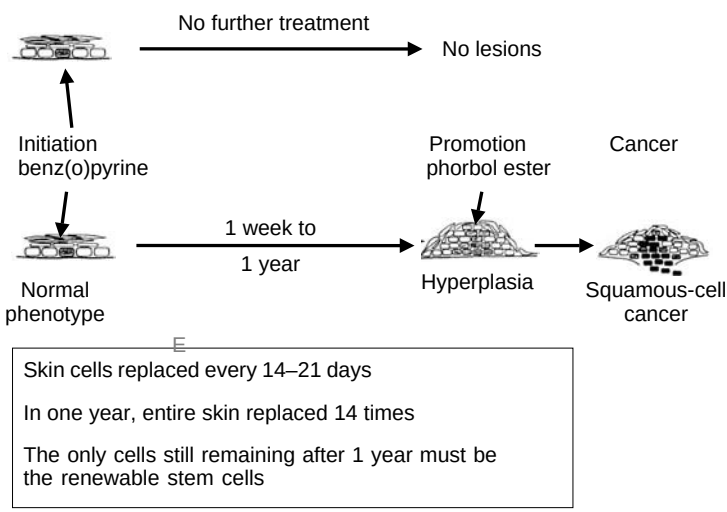

Fig. 12 Chemical carcinogenesis and the stem cell origin of skin cancer. The induction of skin cancer in animals using the initiation–promotion procedure demonstrates that skin cancers arise from stem cells. The two-step procedure of initiation and promotion employs application of a carcinogen, such as benz(o)pyrine, to the skin. Following one application, the skin appears normal and no tumors develop. Promotion is then accomplished by application of an agent that stimulates proliferation, such as phorbol ester, at the site of initiation. After promotion, tumors will appear in a short time. The critical finding is that promotion may be separated from initiation for up to a year. However, all of the transit-amplifying cells in the skin turn over within two weeks. Thus, the only cells that could remain extant from the initiation step are the self-renewing skin stem cells. Drawings modified from DiGiovanni, J. (1992) Multistage carcinogenesis in mouse skin, *Pharmacol. Ther.* **54**, 63–128.

Thus, in the initiation–promotion model for skin carcinogenesis, the initiated cell must be a self-renewing progenitor cell. If this is true, then why do the initiated cells not grow out and produce cancer during normal tissue renewal? Why is promotion required for initiated cells to grow into a cancer? It is likely that normal tissue renewal does not involve proliferation of the long-term self-renewing cells. Most normal cell renewal involves not the most primitive tissue stem cell, but rather the transit-amplifying cells. Thus, it is likely that the initiated transit-amplifying cells are not self-renewing over long periods of time, and that they eventually differentiate. The initiated cell that gives rise to cancer must be a resting stem cell that is only called upon to proliferate under the stress of promotion.

7.6 Stem cells and Adenocarcinomas

Extrapolation of the observations on the stem cell origin of carcinomas may be made to glandular organs. Exocrine glands consist of acini of glandular cells drained by a collecting duct. The key progenitor cells are located between the gland and the duct. For example, both from histologic examination of human lesions and from examination of "preneoplastic" lesions of experimental mouse models, breast cancers are believed to arise from basal progenitor cells located in the terminal ductular lobular unit from which the gland forms during pregnancy. Prostate cancer also arises from basal stem cells. Intestinal carcinogenesis is directed to functionally anchored stem cells in the crypt, and mutated stem cells appear to be able to spread through the mucosa by fusion of crypts, before becoming invasive. Mutation in the *adenomatous polyposis coli* gene (*APC*) is associated with increased expression of survivin, which allows expansion of crypt stem cells and development of colon cancer. Finally, mixed tumors of epithelial and mesenchymal components (carcinosarcomas) appear to arise from a common monoclonal origin. In each instance, carcinomas arise from tissue progenitor cells. There is also evidence that the cancer cells can arise from circulating stem cells that have the ability to cross basement membranes and then localize in various glandular tissues.

8 Therapeutic Applications of Stem Cells: Promises and Problems

8.1 Cellular Therapy

The results of cellular therapy of preclinical animal models of injury and disease point out both the promise and problems of using stem cells for cellular therapy. There are many models that could be presented to illustrate the need for "Designer Therapy", that is, the careful selection of particular cell types for specific therapies. A simplified comparison of the various types of stem cells available for cellular therapy is given in Fig. 13. As illustrations, the treatments of CNS lesions, myocardial infarcts, and diabetes will be considered.

8.1.1 Central Nervous System

Both infarcts in the brain and damage to the spinal cord have been partially corrected in mice by injection of NSCs. NSCs have the ability to migrate within the brain, site themselves in areas of brain or spinal cord damage, and restore cells in the damaged tissue with reconstitution of function – a truly remarkable capability.

	Embryonal	fetal	somatic	Tissue-determined
	Blastocyst	Neural crest	Cord blood Adult Bone marrow	Liver Pancreas etc.
No of cells	Tens	Hundreds	Millions	Hundreds
Form	Cell lines	Cell cultures	Cell isolates	Cell isolates
Potential	Totipotent	Multipotent	Pleopotent	Oligopotent
Differentiation	In vitro	In vitro/In vivo	In vivo	Differentiated
Clinical use	Experimental	Brain	Bone marrow	Experimental
Projected use for	Brain Heart	Brain Islet cells	Heart Liver	Cell replacement
Advantages	Totipotent Unlimited	Tissue directed proliferation	High numbers In vivo Selection Autotransplantation	Tissue directed
Disadvantages	Ethical issues Small number Graft rejection	Ethical issues Small number Graft rejection	Potential questioned culture limited	Small numbers difficult to get culture limited

Fig. 13 Comparison of stem cells used for cellular therapy. Each of the major sources of stem cells, embryonal, fetal (germ), and adult, has advantages and disadvantages for cell-mediated therapy. Among these, only bone marrow or cord-blood stem cells have been tested clinically. Fetal neural crest cells have been used, with some indication of success, for treatment of Parkinson's disease, but conclusive clinical trials have not been published.

The donor cells may be acquired from inner cell masses of blastocysts, from the neural crest cells of developing fetuses, or from certain areas of the adult brain (Fig. 14). The cells are cultured in special "neural" medium that stimulates their differentiation into neural progenitor cells. In addition, growth factors may further the differentiation of the cells into more mature types. For example, epidermal growth factor stimulates the glial differentiation required for restoration of the myelin sheath of nerves, whereas fibroblast growth factor (FGF) stimulates differentiation of nerve cells. Of these cell sources, the neural crest cells from 10.5-day gestation in the mouse have thus far been found to be the most efficient at differentiating into nerve cells; cells from 14.5-day-old fetuses are the best at differentiating into glial cells. Similar cells obtained from aborted 16- to 20-week human fetuses have been used with some success for treatment of human diseases, such as Parkinson's disease. It is very difficult to obtain enough differentiating cells from ESC cultures to use for this purpose; obtaining cells from the adult brain is not practical at this time. The use of fetal neural crest cells is also limited. The hope for the future is to obtain human ESC lines that will have the property of differentiation into neuronal or glial cells in high numbers. This will require the selection of cell lines, not available now, or alteration of cell lines through transfection with neural-specific transcription factors that will drive the cells to the desired fate.

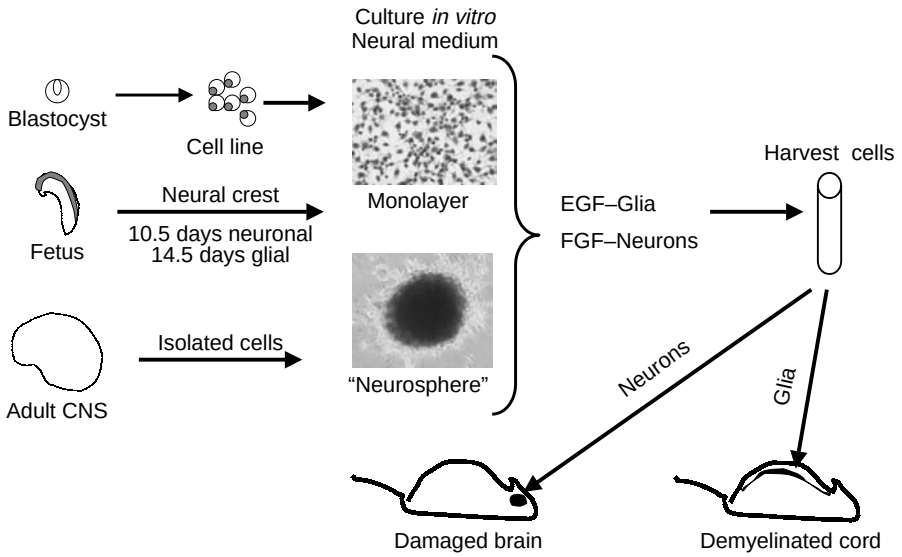

Fig. 14 Stem cell therapy of neurological deficits. Areas of necrosis in the brain, or demyelination in the spinal cord, may be repaired in mouse models by injection of stem cells that have been cultured *in vitro* and stimulated to differentiate into glial cells or neurons. This has been accomplished using ESCs, fetal neural crest stem cells, or stem cells from adult brains. Fetal neural crest cells appear to be the most effective.

8.1.2 Type I Diabetes

Type I diabetes is caused by the loss of insulin-secreting cells, with a consequent loss of regulation of the glucose levels in the blood and tissues, resulting in high blood glucose levels (hyperglycemia). Cellular therapy is directed toward restoring the level of insulin by transplanting insulin-secreting β-cells of the islets of Langerhans in the pancreas. Unfortunately, not only are β-cells difficult to obtain but also the desired effects of transplantation of β-cells are only achieved in less than 20% of attempts. It is predicted that stem cell precursors to β-cells contain the self-renewing cells required to maintain the number of β-cells at the level needed to maintain normoglycemia. Again, various sources of cells have been tested in preclinical animal models of diabetes, either genetically or chemically induced. These have included cells derived from adult pancreatic ducts (bipotent ductal progenitor cells), adult liver stem cells, and neonatal islet cells, as well as bone marrow cells and ESCs. Clearly, the cells more committed to the pancreatic lineage, such as neonatal islet cells and pancreatic duct-derived cells, give better results than do bone marrow stem cells or ESCs. The cells are cultured *in vitro* in the presence of high glucose, which stimulates β-cell differentiation, and are then injected into selected sites in the diabetic animal (for example, under the kidney capsule). This results in restoration of normoglycemia and prevents death due to

complications of diabetes. The take-home lesson is that cells differentiated to the pancreatic lineage are better than less-differentiated cells.

8.1.3 Heart

The severity of damage after a myocardial infarct may be lessened by injection of autologous mononuclear bone marrow stem cells into the infarct-related artery within a week after infarction. This procedure may be done in humans during coronary angioplasty after acute myocardial infarction. In a recent clinical study, the 10 patients treated with stem cells had smaller infarct areas and improved stroke volume index, left ventricular end-systolic volume, contractility of the infarct area, and myocardial perfusion of the infarct region, as compared to 10 patients in a control group not treated with stem cells.

8.2 Pleopotency or Fusion

The tantalizing findings that stem cells at various stages along the determination pathway can contribute to repair of damaged tissue have raised questions on exactly how this happens. The simplest explanation is that self-renewing stem cells are able to proliferate and differentiate into the cell types needed for repair. For example, bone marrow stem cells have been shown to contribute to many mature tissue types, including heart, liver, endothelium, lung, GI tract, and brain. Are there stem cells in the bone marrow that achieve this or are there other mechanisms? For example, the data are conflicting on the ability of bone marrow stem cells to become the smooth muscle cells of the heart. In fact, either the bone marrow cells may act through indirect mechanisms, such as by stimulation of the formation of new vessels (angiogenesis) or else mononuclear cells derived from the bone marrow fuse with regenerating heart muscle cells and contribute to there survival by adding nuclei to the muscle cells. Heart muscle cells are multinuclear, and thus it is normal for the progenitor cells of the heart (satellite cells) to fuse with other cells to form mature muscle. Cells from the bone marrow appear to do this and may increase the ability of infracted heart muscle to regenerate. The situation in the liver is also controversial. In animal models of chronic liver injury due to genetic abnormalities in the hepatocytes, transplanted bone marrow cells may fuse with residual hepatocytes and form polyploid hepatocytes that are resistant to the original injury. However, there are also convincing studies that show bone marrow cells becoming hepatocytes and gi cells without fusion. It may be that ongoing injury, with resultant inflammation and infiltration of bone marrow–derived mononuclear cells combined with damaged hepatocytes, leads to fusion of the monocytes with the hepatocytes. The contribution of the monocyte nucleus may overcome the genetic defect in the hepatocytes to allow survival.

8.3 Immune Rejection of Stem Cells

A major problem in the use of stem cells derived from embryonic or fetal tissues is the potential for such cells to be rejected when they differentiate in a histoincompatible recipient and express MHC antigens. This problem may be solved by nuclear transplantation and cloning, or by genetic manipulation of an embryonal cell line so

that it expresses the same MHC type as the recipient (see Fig. 5). However, cloning is an inefficient process and presents ethical problems, in that a human embryo must be produced to obtain the desired cells. Genetic engineering of ESC lines would be a time-consuming and expensive process. Either the cells would have to be designed specifically for the MHC of the individual recipient or else multiple cell lines expressing various MHC types would have to be produced, in order to have the cells available for any given recipient.

Histocompatibility problems are avoided through the use of autologous adult stem cells. Bone marrow transplantations are also possible across allogenic MHC barriers because the immune cells of the host are eliminated or greatly reduced by irradiation or chemical treatment, allowing the donor cells to survive. In addition, autologous progenitor cells from other tissues could be obtained from injured organs having a higher level of potential than the cells from normal organs. For example, it is possible to culture and then clone progenitor cells from animals with acute liver injury that have the potential to differentiate into other cell types, such as pancreatic cells, in addition to the expected differentiation into liver cells. It is not yet clear whether these injury-induced progenitor cells are derived from the very few progenitor cells in the mature liver, or from the circulating blood stem cells that enter the liver in larger numbers in response to the injury. In any case, there is every reason to believe that stem cell therapy is not only not limited to ESCs but is also feasible using adult stem cells. There is a lot of work to be done to find out which cells to use, how to manipulate them, and how to transplant them so as to get an optimal therapeutic result.

8.4 Seed and Soil Hypothesis

The ability of stem cells to replace tissue cells damaged by either genetic lesions or epigenetic injury is determined both by the innate potential of the stem cells and by the environment in which they are placed (seed and soil hypothesis). The ability of a stem cell to function in a given location is dependent on its potential (seed) as well as on microenvironmental signals (soil) such as cell–matrix interactions, cell–cell interactions, growth factors, hormones, and circulation gradients. Stem cells become restricted in the variety of differentiated progeny that they can produce during the determination process. For pleopotent stem cells to exist in the adult, preservation of the ESC potential must occur in some cells, the tissue progenitor cells. NSCs from the developing brain "remember" where they came from. On the other hand, ESCs and fetal stem cells are able to repopulate damaged neural tissue and to provide cells that will remyelinate damaged spinal cord in experimental models; bone marrow stem cells may also repair brain lesions.

8.5 Designer Therapeutic Stem Cells

Most public attention has focused on the use of ESCs for possible replacement of damaged tissue, or for use in gene therapy. This is because ESCs are capable of giving rise to any other cell type and have unlimited potential for tissue replacement. At the time of this writing, the critical and extensive work that must be done to realize the potential of ESCs is mired in ethical arguments and legal impediments. Until additional cell lines are made available, and until the studies

are accomplished, we cannot know how these cells respond to signals derived from the microenvironment of tissue culture conditions, or from the stroma of adult organs, to differentiate into the mature cell types desired. If they do respond, how many cells will differentiate, and how long will this take? In addition, the possibility that ESCs will give rise to cancer when sited out of their normal environment must be tested critically in preclinical models.

On the other hand, some form of tissue stem cell from fetal or adult individuals may be just as good as, or even better than, ESCs for some purposes. Thus, the cell of choice for treatment of a given disease may need to be selected from a specific stage along the determination pathway (see Fig. 6). In some cases, totipotent ESCs may be the optimal choice; in other cases, the tissue-determined stem cells (see Fig. 13).

There is no clear-cut answer to the question of which stem cells are optimal for any particular type of therapeutic approach. However, if we want cells to replace injured liver, or to grow up cell lines to provide liver function *in vitro* (liver bioreactors), it is logical that cells already some distance along the liver lineage pathway will be better to use than would be totipotent ESCs. Cells committed to neuronal lineages are probably better than ESCs for replacement of damaged neuronal cells. It may be possible to obtain cell lines representing cells at a particular stage of tissue determination for specific cellular therapy. In addition, these cells could be transfected with various genes under the control of promoters that are only activated by differentiated cells, so that expression of the gene is limited to the cells at the appropriate stage of differentiation. It is timely for us to develop and test the potential of designer stem cells for specific therapeutic challenges.

See also Brain Development; Cancer Stem Cells; Cellular Interactions; Mammalian Cell Culture Methods; Nuclear Transfer for Cloning Animals; Parthenogenetic Stem Cells; Principles and Applications of Embryogenomics; Spatial and Temporal Expression Patterns in Animals.

Bibliography

Books and Reviews

Blau, H.M., Brazelton, T.R., Weimann, J.M. (2001) The evolving concept of a stem cell: entity or function? *Cell* **105**, 829–841.

Korbling, M., Estrov, Z. (2003) Adult stem cells for tissue repair – a new therapeutic concept, *N. Engl. J. Med.* **349**, 570–582.

Nakano, T. (2003) Hematopoietic stem cells; generation and manipulation, *Trends Immunol.* **24**, 589–594.

Sell, S. (Ed.) (2004) *Stem Cell Handbook*, Humana Press, Totowa, NJ.

Sell, S. (2004) Stem cell origin of cancer and differentiation therapy, *Crit. Rev. Oncol. Hematol.* **51**, 1–28.

Primary Literature

Alcalay, M., Meani, N., Gelmetti, V. et al. (2003) Acute myeloid leukemia fusion proteins deregulate genes involved in stem cell maintenance and DNA repair, *J. Clin. Invest.* **112**, 1751–1761.

Alvarez-Dolando, N.A., Pardal, R., Garcia-Verdungo, J.M. et al. (2003) Fusion of bone marrow-derived cells with Purkinje neurons, cardiomyocytes and hepatocytes, *Nature* **425**, 968–973.

Arey, L.B. (1974) *Developmental Anatomy: A Textbook and Laboratory Manual of Embryology*, 7th edition, WB Saunders Co, Philadelphia, PA and London.

Asahara, T., Masuda, H., Takahashi, T., Kalka, C., Pastore, C., Silver, M., Dearne, M., Magner, M., Isner, J.M. (1999) Bone marrow origin of endothelial progenitor cells responsible for postnatal vasculogenesis in physiological and pathological neovascularization, *Circ. Res.* **85**, 221–228.

Bartlett, P.F. (1982) Pluripotential hemopoietic stem cells in the adult mouse brain, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2722–2725.

Beckwith, J.B., Palmer, N.F. (1978) Histopathology and prognosis of Wilm's tumor: results from the first national Wilm's tumor study, *Cancer* **41**, 1937–1948.

Beremblum, I. (1954) Carcinogenesis and tumor pathogenesis, *Adv. Cancer Res.* **2**, 129–175.

Bonnet, D., Dick, J.E. (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic, *Cell. Nat. Med.* **3**, 730–737.

Bookmaaker, M.B., Smits, A.M., Tolboon, H. et al. (2003) Bone marrow-derived cells contribute to glomerular endothelial repair in experimental glomerulonephritis, *Am. J. Pathol.* **163**, 553–562.

Bradley, A., Evans, M., Kaufman, M.H., Robertson., E. (1984) Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines, *Nature* **309**, 255–256.

Brinster, R.E. (1974) The effect of cells transferred into the mouse blastocyst on subsequent development, *J. Exp. Med.* **140**, 1049–1056.

Choi, K. (1998) Hemangioblast development and regulation, *Biochem. Cell Biol.* **76**, 947–956.

Cohnheim, J. (1867) Ueber entzundung und eiterung, *Path. Anat. Physiol. Klin. Med.* **40**, 1–79.

Cohnheim, J. (1889) *Lectures in General Pathology, Translated from the second German edition by McKee AB*, The New Sydenham Society, London, UK.

Cory, S., Vaux, D.L., Strasser, A., Harris, A.W., Adams, J.M. (1999) Insights from Bcl-2 and Myc: malignancy involves abrogation of apoptosis as well as sustained proliferation, *Cancer Res.* **59**, 1686s–1692s.

Dontu, G., Al-Hajj, M., Abdallah, W.M., Clarke, M.F., Wicha, M.S. (2003) Stem cells in normal breast development and breast cancer, *Cell Prolif.* **36**(Suppl. 1), 59–72.

Durante, F. (1874) Nesso fisio-pathologico tra la struttura dei nei materni e la genesi di alcuni tumori maligni, *Arch. Memori ed Osservazioni di Chirugia Practica* **11**, 217–226.

Eglitis, M.A., Mezey, E. (1999) Hematopoietic cells differentiate into both microglia and macroglia in the brains of adult mice, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4080–4085.

Evans, M.J., Kaufman, M.Y. (1978) Establishment in culture of pluripotential cells from mouse embryos, *Nature* **292**, 154–156.

Forbes, S.J., Poulsom, R., Wright, N.A. (2002) Hepatic and renal differentiation from blood-borne stem cells, *Gene Ther.* **9**, 625–630.

Friedwald, W.F., Rous, P. (1944) The initiating and promoting elements in tumor production. An analysis of the effects of tar, benzpyrene and methylcholanthrere on rabbit skin, *J. Exp. Med.* **80**, 101–126.

Furth, J., Kahn, M.C. (1937) The transmission of leukemia of mice with a single cell, *Am. J. Cancer* **31**, 276–282.

Garcia, S.B., Park, H.Y., Novelli, M., Wright, N.A. (1999) Field cancerization, clonality, and epithelial stem cells: the spread of mutated clones in epithelial sheets, *J. Pathol.* **187**, 61–81.

Gardner, R.L. (1969) Mouse chimaeras obtained by the injection of cells into the blastocyst, *Nature* **220**, 596–597.

Gilbert, D.M. (2004) The future of human embryonic stem cell research: addressing ethical conflict with responsible scientific research, *Med. Sci. Monit.* **10**, RA99–R103.

Goodell, M.A., Brose, K., Paradis, G., Conner, A.S., Mulligan, R.C. (1996) Isolation and function properties of murine hematopoietic stem cells that are replicating *in vivo*, *J. Exp. Med.* **183**, 1797–1806.

Gootwine, E., Webb, C.G., Sachs, L. (1982) Participation of myeloid leukaemia cells injected into embryos in hematopoietic differentiation in adult mice, *Nature* **299**, 63–65.

Graf, T. (2002) Differentiation plasticity of hematopoietic cells, *Blood* **99**, 3089–3101.

Haslan, S.Z., Bern, H.A. (1977) Histopathogenesis of 7,12-dimethylbenz[a]anathracene-induced rat mammary tumors, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4020–4024.

Herzog, E.L., Chai., L., Krause., D.S. (2003) Plasticity of marrow-derived stem cells, *Blood* **102**, 3483–3493.

Hillebrands, J.L., Klatter, F.A., van Dijk, W.D., Rozing, J. (2002) Bone marrow does not contribute substantially to endothelial-cell

replacement in transplant arteriosclerosis, *Nat. Med.* **8**, 194–195.

Hubner, K., Fuhrmann, G., Christenson, L.K., Kehler, J., Reinbold, R., De La Fuente, R., Wood, G., Strauss, J.F., III., Boiani, M., Scholer, J.R. (2003) Derivation of oocytes from mouse embryonic stem cells, *Science* **300**, 1251–1256.

Ianus, A., Holz, G.G., Theise, N.D., Hussain, M.A. (2003) In vivo derivation of glucose-competent pancreatic endocrine cells from bone marrow with no evidence of cell fusion, *J. Clin. Invest.* **111**, 843–850.

Ivanova, N.B., Dimos, J.T., Schaniel, C., Hackney, J.A., Moore, K.A., Lemischka, I.R. (2002) A stem cell molecular signature, *Science* **298**, 601–604.

Jackson, K.A., Majka, S.M., Wang, H. et al. (2001) Regeneration of ischemic cardiac muscle and vascular endothelium by adult stem cells, *J. Clin. Investig.* **107**, 1395–1402.

Jang, Y-Y., Bollector, M.I., Baylin, S.B., Diehl, A.M., Sharkis, S.J. (2004) Hematopoietic stem cells convert into liver cells within days without fusion, *Nat. Cell Biol.* **6**, 532–539.

Jiang, Y., Jahagirdar, B.N., Reinhardt, R.L., Schwartz, R.E., Keene, C.D. et al. (2002) Pluripotency of mesenchymal stem cells derived from adult marrow, *Nature* **418**, 41–49.

Kataoka, K., Medina, R.J., Kageyama, T., Miyazaki, M., Yoshino, T., Makino, T., Huh, N. (2003) Participation of adult mouse bone marrow cells in reconstitution of skin, *Am. J. Pathol.* **163**, 1227–1231.

Kleinsmith, L.J., Pierce, G.B. (1964) Multipotentiality of single embryonal carcinoma cells, *Cancer Res.* **24**, 1544–1551.

Kotten, D.N., Ma, B.Y., Cardoso, W.V. et al. (2001) Bone marrow-derived cells as progenitors of lung alveolar epithelium, *Development* **128**, 5181–5188.

Krause, D.S., Theise, N.D., Collector, M.I. et al. (2001) Multi-organ, multi-lineage engraftment by a single bone marrow-derived stem cell, *Cell* **105**, 369–377.

Labat, M.L. (2001) Stem cell and the promise of eternal youth: embryonic versus adult stem cells, *Biomed. Pharmacother.* **55**, 179–185.

Lagasse, E., Conners, H., Al-Shalimy, M. et al. (2000) Purified hematopoietic stem cells can differentiate into hepatocytes in vivo, *Nat. Med.* **6**, 1229–1234.

Lee, V.M., Stoffel, M. (2003) Bone marrow: an extra-pancreatic hideout for the elusive pancreatic stem cell? *J. Clin. Investig.* **111**, 799–801.

Li, Y., Chen, J., Wang, L., Lu, M., Chopp, M. (2001) Treatment of stroke in rat with intracarotid administration of marrow stromal cells, *Neurology* **56**, 1666–1672.

Liang, L., Bickenbach, J.R. (2002) Somatic epidermal stem cells can produce multiple cell lineages during development, *Stem Cells* **20**, 21–31.

Lin, Y., Weisdorf, D.J., Solovey, A., Hebbel, R.P. (2000) Origins of circulating endothelial cells and endothelial outgrowth from blood, *J. Clin. Invest.* **05**, 71–77.

Mansergh, F.C., Wride, M.A., Rancourt, D.E. (2000) Neurons from stem cells; implications for understanding nervous system development and repair, *Biochem. Cell Biol.* **78**, 613–628.

Martin, G. (1980) Teratocarcinomas and mammalian embryogenesis, *Science* **209**, 768–776.

Masuya, M., Drake, C.J., Fleming, P.A., Zeng, H., Hill, W.D., Martin-Studdard, A., Hess, D.C., Ogawa, M. (2003) Hematopoietic origin of glomerular mesangial cells, *Blood* **101**, 2215–2218.

McKinney-Freeman, S.L., Jackson, K.A., Camargo, F.D., Ferrari, G., Mavilio, F., Goodell, M.A. (2003) Muscle-derived hematopoietic stem cells are hematopoietic in origin, *Proc. Natl. Acad. Sci. U.S.A.* **299**, 1341–1346.

Medina, D. (1988) The preneoplastic state in mouse mammary tumorigenesis, *Carcinogenesis* **9**, 113–119.

Merock, J.R., Lansita, J.A., James, R., Tunstead, J.R., Sherley, J.L. (2002) Cosegregation of chromosomes containing immortal DNA strands in cells that cycle with asymmetric stem cell kinetics, *Cancer Res.* **62**, 6791–6795.

Mezey, E., Key, S., Vogelsang, G., Szlayova, I., Lang, G.G., Crain, B. (2003) Transplanted bone marrow generates new neurons in human brains, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1364–1369.

Minguell, J.J., Erices, A., Conget, P. (2001) Mesenchymal stem cells, *Exp. Biol. Med.* **226**, 506–520.

Mintz, B., Illmensee, K. (1975) Normal genetically mosaic mice produced from malignant teratocarcinoma cells, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3583–3589.

Morris, R.J., Coulter, K., Tryson, K., Steinberg, S.R. (1997) Evidence that cutaneous

carcinogen-initiated epithelial cells from mice are quiescent rather than actively cycling, *Cancer Res.* **57**, 3436–3443.

Morshead, C.M., Benveniste, P., Iscove, N.N., van der Kooy, D. (2002) Hematopoietic competence is a rare property of neural stem cells that may depend on genetic and epigenetic alterations, *Nat. Med.* **2**, 268–273.

National Research Council – Institute of Medicine (2002) *Stem Cells and the Future of Regenerative Medicine*, Washington, DC, http://stemcells.nih.gov

Nelson, W.J., Nusse, R. (2004) Convergence of Wnt, β-catentin, and cadherin pathways, *Science* **303**, 1483–1487.

Newsome, P.N., Johannessen, I., Boyle, J. et al. (2003) Human cord blood-derived cells can differentiate into hepatocytes in the mouse liver with no evidence of fusion, *Gastroenterology* **124**, 1891–1900.

Oberling, C. (1944) *The Riddle of Cancer*, Yale University Press, New Haven, CT.

Owens, D.M., Watt, F.M. (2003) Contribution of stem cells and differentiated cells to epidermal tumors, *Nat. Rev. Cancer* **3**, 444–451.

Papaioannou, V.E., McBurney, M.W., Gardner, R.L., Evans, R.L. (1975) Fate of teratocarcinoma cells injected into early mouse embryos, *Nature* **258**, 70–73.

Papaioannou, V.E., McBurney, M.W., Gardner, R.L., Evans, R.L., Rossant, J. (1983) Effects of the embryonic environment on proliferation and differentiation of embryonal carcinoma cells, *Cancer Suv.* **2**, 165–183.

Pappenheim, A. (1917) Prinzipien der neueren morphologischen Haematozytologie nach zytogenetischer Grundlage, *Folia Haematologica* **21**, 91–101.

Parchment, R.E., Pierce, G.B. (1989) Polyamine oxidation, programmed cell death, and regulation of melanoma in the murine embryonic limb, *Cancer Res.* **49**, 6680–6686.

Pardal, R., Clarke, M.F., Morrison, S.J. (2003) Applying the principles of stem-cell biology to cancer, *Nature Rev. Cancer* **3**, 895–902.

Parmar, S., Tallman, M.S. (2003) Acute promyelocytic leukaemia: a review, *Expert Opin. Pharmacother.* **4**, 1379–1392.

Peifer, M., Plakis, P. (2000) Wnt signaling in oncogenesis and embryogenesis-a look outside the nucleus, *Science* **287**, 1606–1609.

Pera, M.F., Herszfeld, D. (1998) Differentiation of human pluripotent teratocarcinoma stem cells induced by bone morphogenic protein-2, *Reprod. Fertil. Dev.* **10**, 551–555.

Perez-Losada, J., Balmain, A. (2003) Stem cell hierarchy in skin cancer, *Nature Rev. Cancer* **3**, 434–443.

Petersen, B.E., Bowen, W.C., Patrene, K.D., Mars, W.M. et al. (1999) Bone marrow as a potential source of hepatic of hepatic oval cells, *Science* **284**, 1168–1170.

Peyron, A. (1938) Sur la presence des cellules genitales primordiales dans les boutons embryonnaires des embryomes parthenogenetiques chez l'homme, *C.R. Acad. Sci. (Paris)* **206**, 1680–16831.

Pierce, G.B. (1983) The cancer cell and its control by the embryo, *Am. J. Pathol.* **113**, 117–124.

Pierce, G.B., Shikes, R., Fink, L.M. (1978) *Cancer: a problem of developmental biology*, Prentice-Hall, Inc., Englewood Cliffs, NJ.

Pierce, G.B., Wallace, C. (1971) Differentiation of malignant to benign cells, *Cancer Res.* **31**, 127–134.

Pittenger, M.F., McKay, A.M., Beck, S.C. et al. (1999) Multilineage potential of adult human mesenchymal stem cells, *Science* **284**, 143–147.

Potten, C.S., Morris, R.J. (1988) Epithelial stem cells in vivo., *J. Cell Sci. (Suppl.)* **10**, 45–62.

Potten, C.S., Morris, R.J., Owen, G., Booth, D. (2002) Intestinal stem cells protect their genome by selective segregation of template DNA strands, *J. Cell Sci.* **115**, 2381–2388.

Potter, V.R. (1978) Phenotypic diversity in experimental hepatomas: the concept of partially blocked ontogeny, *Br. J. Cancer* **38**, 1–23.

Poulsom, R., Forbes, S.J., Hodivala-Dilke, K. et al. (2001) Bone marrow contributes to renal parenchymal turnover and regeneration, *J. Pathol.* **195**, 229–235.

Priller, J., Persons, D.A., Klett, F.F., Kempermann, G., Kreutzberg, G.W., Dirnagl, U. (2001) Neogenesis of cerebellar Purkinje neurons form gene-marked bone marrow cells in vivo, *J. Cell Biol.* **155**, 733–738.

Prockop, D.J., Gregory, C.A., Spees, J.L. (2003) One strategy for cell and gene therapy: Harnessing the power of adult stem cells to repair tissues, *Proc. Natl. Acad. Sci. U.S.A.* **100**(Suppl. 1), 11917–11923.

Quaini, F., Urbanek, K., Beltrami, A.P. et al. (2002) Chimerisms of the transplanted heart, *N. Engl. J. Med.* **346**, 5–15.

Reya, T., Morrison, S.J., Clark, M.F., Weissman, I.L. (2001) Stem cells, cancer and cancer stem cells, *Nature* **414**, 105–111.

Rifai, S., Lyden, D. (2003) Therapeutic stem and progenitor cell transplantation for organ vascularization and regeneration, *Nat. Med.* **9**, 702–712.

Rippert, V. (1904) *Ueber ein myosarcoma striocellulare des nierenbeckens und des ureters*, Geschwulstlehre, Bonn, Germany.

Robertson, E.J. (1986) Pluripotential stem cells lines as a route into the mouse germ line, *Trends Genet.* **2**, 9–14.

Rotter, W. (1921) Histogenese der malignen Geschwulste, *Ztschr. Krebsforschung* **18**, 171–208.

Rowley, J.D. (1999) The role of chromosomal translocations in leukemogenesis, *Semin. Hematol.* **36**, 59–72.

Sherley, J.L. (2002) Asymmetric cell kinetics genes: the key to expansion of adult stem cells in culture, *Stem Cells* **20**, 561–572.

Shimada, H., Chatten, J., Newton, W.A. et al. (1984) Histopathologic prognostic factors in neuroblastic tumors: Definition of subtypes of ganglioneuroblastoma and an age-linked classification of neuroblastomas, *J. Natl. Cancer Inst.* **73**, 405–416.

Spradling, A., Drummond-Barbosa, D., Kai, T. (2001) *Stem cells find their niche*, *Nature* **414**, 98–104.

Stevens, L.C. (1967) Experimental production of testicular teratoma in mice, *J. Natl. Cancer Inst.* **52**, 549–552.

Stevens, L.C. (1970) The development of transplantable teratocarcinomas from intratesticular grafts of pre- and post-implantation of mouse embryos. *Dev. Biol.* **21**, 364–382.

Stocum, D.L. (2004) Amphibian regeneration and stem cells, *Curr. Top. Microbiol. Immunol.* **280**, 1–70.

Taipale, J., Beachy, P.A. (2002) The hedgehog and Wnt signaling pathways in cancer, *Nature* **411**, 349–354.

Terada, N., Hamazaki, T., Oka, M. et al. (2002) Bone marrow cells adopt the phenotype of other cells by spontaneous cell fusion, *Nature* **416**, 542–545.

Theise, N.D., Badve, S., Saxena, R. et al. (2000) Derivation of hepatocytes from bone marrow cells in mice after radiation-induced myeloablation, *Hepatology* **31**, 235–240.

Till, J.E., McCulloch, E.A. (1961) A direct measurement of the radiation sensitivity of normal mouse bone marrow cells, *Radiat. Res.* **14**, 213–222.

Tomita, M., Adachi, Y., Yamada, H. et al. (2002) Bone marrow-derived stem cells can differentiate into retinal cells in injured rat retina, *Stem Cells* **20**, 279–283.

Tran, S.D., Pillemer, S.R., Dutra, A. et al. (2003) Differentiation of human bone marrow-derived cells into buccal epithelial cell in vivo; a molecular analytical study, *The Lancet* **361**, 1084–1088.

Van Duuren, B.L., Sivak, A., Katz, C., Seidman, I., Melchionne, S. (1975) The effect of ageing and interval between primary and secondary treatment in two-stage carcinogenesis on mouse skin, *Cancer Res.* **35**, 502–505.

Vassilopoulos, G., Wang, P-R., Russell, D.W. (2003) Transplanted bone marrow regenerates liver by cell fusion, *Nature* **422**, 901–904.

Virchow, R. (1863) *Die krankhafter Geschwulste*, Bd 1. Berlin, Hirschwald.

Wagers, A.J., Sherwood, R.I., Christensen, J.L., Weissman, I.L. (2002) Little evidence for developmental plasticity of adult hematopoietic stem cells, *Science* **297**, 2256–2262.

Wang, X., Willenbring, H., Akkari, Y. et al. (2003) Cell fusion is the principal source of bone-marrow-derived hepatocytes, *Nature* **422**, 897–901.

Wang, Z., Lin, J. (2004) Nano maintains germline stem cell self-renewal by preventing differentiation, *Science* **303**, 2016–2019.

Watt, F.M. (1989) Terminal differentiation of epidermal keratinocytes, *Curr. Opin. Cell Biol.* **1**, 1107–1115.

Wilms, M. (1899) *Die Mischgeschwuelste*, Arthur Georgi, Leipzig, Germany.

Wodarz, A., Nusse, E. (1998) Mechanisms of Wnt signaling in development, *Annu. Rev. Cell Dev. Biol.* **14**, 59–88.

Yamashita, Y.M., Jones, D.L., Fuller, M.T. (2003) Orientation of asymmetric stem cell division by the AAPC tumor suppressor and centrosome, *Science* **301**, 1547–1550.

Ying, Q-L., Nichols, J., Evans, E.P., Smith, A.G. (2002) Changing potency by spontaneous fusion, *Nature* **416**, 541–548.

Young, H.E. (2004) Existence of reserve quiescent stem cells in adults, from

amphibians to humans, *Curr. Top. Microbiol. Immunol.* **280**, 71–109.
Zhao, L.R., Duan, W.M., Reyes, M., Keene, C.D., Verfaillie, C.M., Low, W.C. (2002) Human bone marrow stem cells exhibit neural phenotypes and ameliorate neurological deficits after grafting into the ischemic brain of rats, *Exp. Neurol.* **74**, 11–20.

Streptomyces, Molecular Biology of

Paul Herron, Emma Tilley, and Iain Hunter
University of Strathclyde, Glasgow, UK

1 Introduction 440

2 The *Streptomyces* Chromosome 443

3 *Streptomyces* Genome Sequences 444

4 *Streptomyces* Genomics 447
4.1 Microarrays 447
4.2 Proteomics 447

5 Genetic Tools for *In Vivo* Genetic Manipulation of *Streptomyces* 448
5.1 Plasmids 449
5.2 Bacteriophages 450
5.3 Transposons 451

6 Genetic Tools for *In Vitro* Genetic Manipulation of *Streptomyces* 452
6.1 *In Vitro* Shuttle Transposon Mutagenesis 453
6.2 PCR-targeted Mutagenesis 455

7 Antibiotic Selection 455

8 Morphological Development 456

9 Antibiotic Production 458

10 Optimization of Gene Expression in *Streptomyces* 458

11 Conclusions 459

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

Bibliography 460
Books and Reviews 460
Primary Literature 460

Keywords

Auxotroph
A mutant that is unable to grow on a simple "minimal" nutrient, due to its inability to make a key cellular building block, such as an amino acid or nucleic acid base, from the simple carbon source provided, for example, glucose. Such "markers," for example, a requirement for histidine, or proline, or uracil are used in genetic mapping experiments. See "complementation."

Complementation
A mutant that makes a defective (or no) gene product can often be recognized by a requirement for a particular nutritional supplement (its auxotrophic marker). For example, a strain with a mutation in the biosynthetic pathway for histidine will not grow unless histidine is added to the simple medium. In classical genetic mapping experiments, two strains are mated and the proficient gene from one partner is passed to the recipient that has the defective gene. In this case, the nutritional requirement for histidine will disappear. This process is called "complementation." In recombinant DNA experiments, when the gene that corresponds to the mutation is cloned, the event can be recognized by complementation

Cosmid
A specialized cloning vector that depends on the property of bacteriophage lambda to package selectively around 50 kb of DNA within its virus particle. The cosmid vector (typically around 10 kb is, therefore, a device that enables large (~40 kb) fragments of genomic DNA to be cloned efficiently, and is useful while preparing libraries as part of a genome sequencing project. Once the "phage" has been used to package the DNA, and the recombinant cosmid is introduced into *Escherichia coli*, it behaves as a plasmid.

Generalized Transduction
A change in phenotype of a recipient bacterium brought about by infection of a defective phage particle containing only donor bacterial DNA, followed by recombinational integration into the recipient's genome.

Lysogeny
The state in which a temperate phage has infected a host, which results in a stable association of the phage with the bacterium. The prophage is replicated with the bacterial genome, and is inherited by each daughter cell.

Open Reading Frame (ORF)
A nucleic acid sequence containing a series of codons without any stop codons, which may encode a protein.

Proteome
The collection of proteins made by a strain at a particular point in time. The proteome is displayed by two-dimensional electrophoresis and appears as individual spots. The technique is at its most powerful when different proteomes of the same strain are compared, to identify which proteins are upregulated or downregulated under given conditions.

Protoplast
The term given to a microbial cell after its cell wall has been removed substantially. This enables two different strains to be genetically recombined by fusion of their protoplasts in the presence of polyethylene glycol. Protoplasts are also used as recipients for naked DNA.

Secondary Metabolite
A metabolite that is made by the microorganism, often after growth has ceased – this metabolite is not essential for growth, viability, or survival of the microbe.

Transcriptome
The collection of RNA molecules present at any point in time. The transcriptome is displayed by hybridizing dye-labeled RNA to arrays of oligonucleotides, or short PCR products, which represent each gene of the organism. Each oligonucleotide or PCR product is immobilized in a fixed position on the array. A comparison of different transcriptomes indicates which genes are upregulated or downregulated under various conditions.

Sigma (σ) Factor
A protein that associates with core RNA polymerase and determines protein specificity.

Transformation/Transformant
The process by which naked DNA is introduced into a cell (or protoplast). The resulting cell is the transformant.

Transposition
The process by which a DNA element can insert itself, independent of host-cell recombination.

Transposon
A discrete DNA element capable of transposition and carrying recognizable genes, such as antibiotic-resistance markers.

Streptomyces are mycelial, soil-dwelling, differentiating bacteria that are renowned for the production of many clinically-useful antibiotics (e.g. streptomycin, erythromycin, tetracyclines, and a number of the third-generation penicillin-like antibiotics). They also make other useful metabolites (e.g. anticancer agents, immunosuppressants, antihelmintics, growth promoters, herbicides, and insecticides). Their status as simple microbial models of differentiation is well established and, because they are taxonomically related to pathogens such as mycobacteria that cause tuberculosis and leprosy (but are themselves "safe" organisms), they are being cited increasingly as informative models for these life-threatening pathogens. Much effort has gone into investigating the fundamental biology of *Streptomyces*. In recent years, the dissection of the architecture and regulation of antibiotic production gene clusters has been undertaken, together with studies on the regulatory circuits that control production and integrate with the signaling pathways of the differentiation processes. As clinical infections invariably develop resistance to antibiotics, there is an ongoing need to discover new drugs to combat this problem. The *Streptomyces* are viewed as a rich source of such drugs and contemporary effort, using recombinant technologies, is aimed at altering the genes for antibiotic production pathways so that they produce new, potentially useful metabolites that may keep them one step ahead of the resistance problem. Recent advances in the molecular biological techniques applicable to streptomycetes have assisted greatly in this task. The genome sequence of the model streptomycete, *Streptomyces coelicolor*, was published in 2002, followed recently by that of *Streptomyces avermitilis*. In 2004, a number of other sequencing projects were in progress, so that these resources will grow significantly in the near future. The *Streptomyces* are remarkable in that their genomes are among the largest (8–9 Mb, compared with *E. coli* at 4.6 Mb) reported so far for any bacteria. Bioinformatics, combined with postgenomic techniques such as microarrays and proteome gels, are now being used to define and characterize global gene expression under a variety of conditions. Rapid methods have been developed to create mutants for any gene, using its genome-derived DNA sequence to guide the strategy. Mutant libraries generated by random transposon insertion, are also becoming available. These aim to provide a collection of mutants in which each of the almost 8000 genes in *Streptomyces* is mutated individually. With these recent postgenomic resources building on the robust genetic systems that have been set up for *Streptomyces*, the next few years promise to be an exciting phase in our understanding of these unique bacteria.

1 Introduction

Streptomyces are members of the Gram-positive order *Actinomycetales* that includes:

- mycobacteria (the causative agents of tuberculosis and leprosy),
- corynebacteria (the causative agent of diphtheria, while other corynebacterial species are now used commercially in amino acid production),

- arthrobacters (often associated with plant pathology, but also able to degrade toxic chemicals), and
- nocardias (some of which are pathogenic).

The order *Actinomycetales* is very diverse. However, there is a convenient split, in that some members of the order are filamentous, while the others are not. The *Streptomyces* are filamentous bacteria (a feature that they share with *Nocardia*), found predominantly in the soil. They have a complex life cycle involving a sporulation phase (see later) and are remarkable for the vast range of secondary metabolites that they produce. Streptomycetes make more than 60% of known natural antibiotics, many of which are of commercial importance. They also produce secondary metabolites, which are used diversely as anticancer agents, immunosuppressants, antihelmintics, growth promoters, herbicides, and insecticides.

Thus, the streptomycetes have generated considerable interest – as model organisms to investigate microbial differentiation, as safe laboratory models (few are pathogenic) of their medically-important relations, but most importantly because of their industrial significance.

The *Streptomyces* were originally misclassified taxonomically as fungi, due to their appearance (mycelial, pigmented, and sporulating) that was reminiscent of moulds. However, they are true bacteria, displaying typical features such as bacterial cell walls, sensitivity to antibiotics targeted against bacteria, prokaryotic ribosomal structure (including 16S and 23S rRNA), and a single chromosome (albeit a linear chromosome, rather than the single circular chromosome of most other bacteria – see later). A notable feature of *Streptomyces* is the high (73%) G + C content in their DNA. Most of the other organisms have a G + C content of around 50%, and the range is 40 to 60%. It is not clear why the G + C content is so high. One theory states that because they are soil-dwelling and often found in arid desert conditions, there has been selective pressure to move away from DNA with a reasonable A + T content – which would make the organism more susceptible to damage by the ultraviolet rays of the sun.

The life cycle of *Streptomyces* involves extensive differentiation. Starting from a dormant spore, germination is followed by radial outgrowth which, on an agar plate, results in the formation of a mycelial colony. In their natural habitat, growth will often colonize soil particles. For each species, the radial growth involves branching of the hyphae in a particular pattern – the branch-to-branch distance of which, on a specific agar, is characteristic of the species and must be somehow programmed into its biology. Radial growth proceeds by apical extension of the mycelia, until some environmental cue triggers the first step in differentiation. Cues may be nutrient limitation, lack of water, or other factors that result in switching on of stress responses. Many *Streptomyces* undergo most elaborate forms of differentiation that involve lysis of the mycelia formed during radial growth and use the metabolic intermediates released from these lysed cells as building blocks to construct an aerial form of the mycelium, which reaches up and away from the radial mycelial mat. This aerial mycelium has a complex morphology that is again quite species-specific, often with helical coils that help to give the structure some rigidity. In the final phase of differentiation, the aerial mycelia septate and round off to form fresh spores. These can remain dormant and desiccated until the environmental conditions improve and a

fresh life cycle can begin again through germination.

Streptomyces coelicolor A3(2), the "model" streptomycete, has been studied for over 50 years – particularly through the lifelong research in the United Kingdom by Sir David Hopwood and his colleagues. The *in vivo* genetics of *S. coelicolor* were unraveled by Hopwood in the 1950s and the 1960s at the same time as the seminal discoveries on the genetics of other bacteria. Stable auxotrophs were formed easily, and they were used to construct a rudimentary linkage map. Over the years, many auxotrophic and other nutritional markers were added. Many mutants that were defective in the differentiation processes were isolated, characterized, and the positions of the genes on the linkage map were mapped carefully.

The original isolate of A3(2) produced a blue pigment, now characterized as the antibiotic actinorhodin. It was much later, in the late 1970s that attention was turned to this antibiotic, for which mutants could be recognized easily by the loss of blue pigment. In other mutant classes, there was a change in color associated with the mutation that now caused the buildup of an actinorhodin precursor or shunt metabolite. Thus, it was possible to quickly build up a collection of mutants altered in antibiotic production. It then became clear that *S. coelicolor* made a number of antibiotics: the red pigment, later identified as undecylprodigiosin; a colorless antibiotic, methylenomycin, whose production gene cluster is located on a plasmid; and a "calcium-dependent" antibiotic (CDA), which was only effective in the presence of the cation and later shown to have a lipopolypeptide structure that is made by a nonribosomal peptide synthase.

In 2001, when the genome sequence of *S. coelicolor* was completed, it became apparent that this model organism has the genetic potential to make over 20 different secondary metabolites, indicating the speed, power, and potential of the postgenomic approaches (see Sects. 4 and 6), compared to the classical approach. The "classical genetic" phase of *S. coelicolor* provided the bedrock of mutants, genetic loci, and fundamental procedures for the next phase – the "molecular biology" stage, involving *in vitro* techniques. The enabling methodology that resulted in the first success in cloning with *Streptomyces* was undoubtedly the ability to form and regenerate protoplasts from specially-treated mycelium. Originally developed as a tool for classical genetic analysis, whereby two strains could be genetically recombined without recourse to spore-based mating, protoplasts were fundamental to the introduction of naked plasmid DNA – enabling the first cloning experiments to be undertaken in 1980.

For the ensuing 15 years these cloning techniques, using the transformation of protoplasts, were refined substantially. Entire antibiotic pathways were cloned, dissected, and altered by mutation *in vitro*. Many of the genes identified during the classical phase of *Streptomyces* genetics were cloned by complementation and analyzed. New genes were identified and mutants of them were created by gene disruption. This was the era of laborious and meticulous characterization of the genes on an individual basis – analysis of their expression, and understanding of the complex interrelationships between them and their gene products. The construction of the ordered cosmid library of *S. coelicolor*, by Redenbach and coworkers in 1996, heralded the era of genome sequencing of *Streptomyces* (Sect. 3) and a number of other genome sequences are now completed, or projects are in progress.

These genome sequences are being annotated and bioinformatic tools are readily available for their further analysis. Gene-specific microarrays have been developed, that are now available to allow inspection of the transcriptome at any point in time or nutritional/environmental perturbation (Sect. 4). Rapid methods have now been developed to create mutants for any gene, using its genome-derived DNA sequence to guide the strategy (Sect. 6). Mutant libraries, generated by random transposon insertion, are also becoming available (Sect. 6). These aim to provide a collection of mutants in which each of the 7825 genes in *S. coelicolor* is mutated individually.

The resources discussed above can now be deployed in a number of ways. Display of the proteome of an organism at any point in time and nutritional/environmental regime (Sect. 4) provides valuable insight into the proteins being used at any instance (around 1000 of the possible 7825 will be expressed). Comparison with proteomes at other times and regimes gives further insight and, when combined with transcriptome data, provides a detailed analysis of the expression profiles of the genes. Combining this data-intensive information with parallel analyses of transcriptomes and proteomes of mutants that lack a particular gene then allows the relationship between expression of that gene with the others to be defined, and so an integrative view of the physiology of the organism results.

The next few years promise to be most exciting. We now have the genome sequences and postgenomic resources to conduct a thorough investigation of these unique bacteria. Many of the remaining questions about the integration of the morphological development with antibiotic production are sure to be answered in the near future.

Whereas these resources have been developed primarily for *S. coelicolor*, many have broad application in the industrial species, such as *Streptomyces clavuligerus* (clavulanic acid), *Streptomyces fradiae* (tylosin), *S. avermitilis* (avermectin), and *Streptomyces rimosus* (oxytetracycline). In the near future, we are also likely to see advances in our knowledge of how to exploit these industrial processes better and make them more productive.

2 The *Streptomyces* Chromosome

The precedent for chromosomal circularity established for *E. coli* during the 1960s meant that it was assumed that this was also the case for streptomycetes. In addition, early genetic studies also supported the circularity of the streptomycete chromosome. However, a circular genetic map is also consistent with having a linear chromosome. As a result of groundbreaking work by Carton Chen and Matthias Redenbach, this subsequently turned out to be the case for *Streptomyces*. Despite the inherent difficulty in determining the physical topology of this very large molecule directly, linearity was established by pulsed field gel electrophoresis of *Streptomyces lividans*, and by the ordering of the *S. coelicolor* cosmid library. Further analysis has revealed that chromosome linearity is a universal phenomenon in *Streptomyces*, as it is in related actinobacteria with similar sophisticated lifestyles. Some aspects of chromosome replication in *Streptomyces* are similar to other bacteria; for example, replication is initiated at a single origin near the chromosome center and proceeds bidirectionally. However, there is an inherent problem in replicating the ends of a linear DNA molecule, as DNA

replication can only progress in the 5′ to 3′ direction. This is because all known DNA polymerases require a 3′-OH from the adjacent "upstream" base to begin adding nucleotides, one at a time. In a linear chromosome, there is no free 3′-OH on one of the strands at each end. A similar problem arises in eukaryotic chromosomes, which are also linear, but here the highly-specialized telomerase complex makes a short RNA to provide that 3′-OH. Clearly, chromosomal circularity in most bacteria avoids this difficulty, as the DNA molecule does not have any "ends." Streptomycetes solve the problem of the free 3′-OH by having terminal proteins (Tpg) that are covalently attached to the 5′ ends of the chromosome. They are believed to act as the primer containing the 3′-OH and this provides a solution to this "end replication" problem. Genes encoding Tpg have now been identified in *S. coelicolor*, *S. lividans*, *S. avermitilis*, and *Streptomyces rochei*. Recently, Tap, a telomere-binding protein, has been described; *tap* is co-transcribed with *tpg* and it interacts with both Tpg and the telomeres; both are essential for maintenance of chromosomal linearity and mutants lacking these genes invariably possess circular chromosomes. Such strains often show increased genetic instability but generally show no impairment in growth or sporulation. The implications of chromosome linearity on chromosome segregation and decatanation following replication are therefore unclear. Both ends are physically associated through an unknown interaction. The ends of the chromosome are characterized by long terminal inverted repeats (TIRs), which vary in length between 30 and 600 kb depending on the species. A comparison of TIRs from different species has revealed conserved sequences within the first 170 bp arranged as seven palindromes, and may cause the formation of DNA secondary structure important for stability of the terminal sequences. The significance of the TIRs is unclear and the overall structure of the chromosome is thought to resemble a tennis racket with the TIRs forming the handle through an association between proteins bound to each of the inverted repeats.

3 *Streptomyces* Genome Sequences

The genome sequence of the model streptomycete, *S. coelicolor* was completed in July 2001 at the Sanger Centre, Cambridge, UK. The sequenced strain, M145, is plasmid-free, and although originally derived from a prototrophic revertant of a threonine auxotroph of the wild-type strain, A3(2), M145 behaves substantially like its parent *in vivo*. The genome of *S. coelicolor* was sequenced as 325 overlapping clones. Of these, 305 were cosmids, one was an existing plasmid carrying one of the chromosome termini and 19 were selected from a set of bacterial artificial chromosomes mapped to the sequences of finished cosmid contigs by end sequencing. Each clone was fractionated into 1.4 to 2 kb fragments, prior to sequencing. The finished 325 clones formed a contiguous sequence extending from within the left TIR to the right-hand end of the genome. The genome sequence was completed by extending the incomplete left TIR with a 7-kb consensus sequence copied from the right TIR.

S. avermitilis produces the antiparasitic agent avermectin, used in veterinary medicine and for the treatment of river blindness in Africa. In contrast to

S. coelicolor, its genome was sequenced using a whole genome shotgun approach combined with cosmid end sequencing to facilitate the final assembly, and with some sequencing of entire cosmids to sequence repeated regions of DNA accurately. Several other streptomycete genome sequencing projects (Table 1) are either under way, or have been completed but are currently publicly unavailable for commercial reasons.

The localization of open reading frame (orfs) within streptomycete genome sequences is facilitated by highly-biased codon usage at the third position (mol% $G + C \approx 90$) and the two published genome sequences, for *S. coelicolor* and *S. avermitilis*, confirm this predicted consequence of the high $G + C$ content of streptomycete DNA on codon usage. The 8 667 507 bp *S. coelicolor* genome encodes 7825 genes, while the 9 025 608 bp *S. avermitilis* genome encodes 7574. The publication of a second genome sequence has meant that it is now possible to assess the characteristic features of the streptomycete genome; it seems that the chromosomes of members of this genus have an unusual biphasic structure. *S. coelicolor* has a central core to its chromosome covering around 5 Mb of DNA and containing all genes that are unconditionally essential, such as DNA replication, transcription, and translation. In addition, it carries two arms of about 1.5 and 2.3 Mb containing conditionally-adaptive genes, such as those encoding the enzymes necessary to degrade cellulose and chitin likely to be present in the organism's natural habitat, which is soil. The *S. avermitilis* genome shows a similar division between core and arm regions.

Strong synteny exists between the core regions of the two *Streptomyces* genomes and the genomes of the two related pathogens, *Mycobacterium tuberculosis* and *Corynebacterium diptheriae*. This suggests that the streptomycete chromosome core is derived from a common actinobacterial ancestor with the arms being subsequently added by horizontal gene transfer. These organisms are superbly adapted to life in highly-variable, stressful environments. Soil is a heterogeneous environment, both temporally and spatially, which is constantly changing with respect to nutrient status, temperature, pH, and salinity. Unlike higher organisms bacteria are unable to control, to any great extent, the environment in which their cells live. Therefore, it is not surprising that this is reflected in the complexity of streptomycete genome sequences, reflecting the diversity of gene products that they have to make to cope with these changes.

The genomes of both organisms encode many new classes of proteins that are able to respond to this changing environment, such as those involved in transport and degradation of extracellular nutrients with a particular emphasis on regulation. Over 12% of proteins encoded by *S. coelicolor* are predicted to have a regulatory function (965 proteins), with a clear preference for certain regulatory groups. *S. coelicolor* encodes 65 different sigma factors responsible for controlling the promoter specificity of the RNA polymerase core enzyme. Two-component regulatory systems are found in high abundance in streptomycetes; *S. coelicolor* encodes 85 sensory kinases and 79 response regulators. The elucidation of the complex signal transduction pathways involved in the regulation of antibiotic production and morphological differentiation may now be possible through the study of previously unknown regulatory pathways in an ecologically-relevant context, rather than in the nutrient-rich environment of the laboratory.

Tab. 1 Current *Streptomyces* genome-sequencing projects. The information in this table is accurate till January 2005. For up-to-date information on the status of these and new streptomycete genome sequencing projects see www.genomesonline.org/.

Strain	*Genome size*	*Organization(s)*	*Status*
Streptomyces avermitilis MA-4680 (ATCC 31267)	9025 kb (7573 orfs)	Kitasato University, University of Tokyo, NITE, NIID, MEXT	Complete (http://avermitilis.ls.kitasato-u.ac.jp/)
Streptomyces coelicolor M145	8667 kb (7825 orfs)	Sanger Institute, John Innes Centre, BBSRC	Complete (www.sanger.ac.uk/Projects/S_coelicolor/)
Streptomyces ambofaciens ATCC 23877	8000 kb	Genoscope	Incomplete (unavailable) (more information at: www.genoscope.cns.fr/externe/English/Projets/Projet_DL/DL.html)
Streptomyces diversa		Diversa, Celera	Complete (unavailable)
Streptomyces noursei		QIAGEN	Complete (unavailable)
Streptomyces peucetius		IBR, Sunmoon University, Gene Chem Inc, MIC.	Incomplete (unavailable)
Streptomyces scabies 87–22	8100 kb	Cornell University, Sanger Institute, USDA, NSF	Incomplete (partially available at www.sanger.ac.uk/Projects/S_scabies/)

4 *Streptomyces* Genomics

4.1 Microarrays

Despite the large number of genes and problems of paralogous cross-hybridization compounded by the high streptomycete G + C content, the ability to monitor global gene expression in *S. coelicolor* is now achievable by microarray analysis. Two sources of microarrays are available for the *Streptomyces* research community: firstly, those produced by Stanford University (http://sncohenlab.stanford.edu/streptomyces/) in the United States and secondly, those produced by UMIST and University of Surrey, in the United Kingdom (www.surrey.ac.uk/SBMS/Fgenomics/Microarrays/index.html). The Stanford arrays were constructed by spotting between 5000 and 7000 orf-specific PCR products on glass slides. PCR products were generated from a single PCR amplification using 20-mer gene-specific primers that resulted in around 10% of the probes in the Stanford arrays being duplicates or overlapping sequences. The PCR products used to make the Surrey microarrays were originally generated by a versatile two-step procedure. The probe size designed for each orf was deliberately small in order to generate a unique product and so minimize cross-hybridization to other sequences in the genome. The first-round PCR was carried out using long PCR primers; the 3′ region of which was specific for the target gene while the 5′ region contained a universal tag sequence that allowed reamplification of the entire library with the same primer pair. In addition to the simple reamplification of the PCR products, there are other advantages with this approach. It minimizes the amount of genomic template carried onto the array, and as one of the primers can be derivatized (e.g. 5′-amidated), attachment of the PCR products to glass slides is facilitated. At present, the microarrays contain just over 7000 unique probes as well as landing lights, *Pseudomonas aeruginosa* PAO1 DNA (negative control) and rRNA (positive control). These arrays will be superseded by long oligonucleotide arrays in the near future. Both sets of microarrays have been successfully employed to investigate aspects of *Streptomyces* biology: the Stanford arrays were used to analyze dynamic changes in transcript profiles associated with developmental transition, while the Surrey arrays were used to demonstrate negative-feedback regulation of the *dnaK* operon by the DnaK chaperone machine.

4.2 Proteomics

Protein analysis by two-dimensional gel electrophoresis (2DGE) has been used with *S. coelicolor* for many years in both liquid and surface grown cultures in order to define the four stages in the life cycle of this species. Large changes in the radiolabeled protein profile obtained by 2DGE were observed when *S. coelicolor* was subjected to a variety of stresses including heat, cold, osmotic stress, and pH (http://proteom.biomed.cas.cz/). Since the completion of the *S. coelicolor* genome sequence, it has been possible to use matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) to identify many of the protein spots that show altered patterns of expression in response to stress. More extensive spot identification has led to over 10% of the theoretical proteome being mapped experimentally using 2DGE and MALDI-TOF MS

(http://dbkweb.ch.umist.ac.uk/StreptoBASE/s_coeli/referencegel/). In this study, extensive posttranslational modification was indicated as more than one protein spot per gene was observed. Subsequently, extensive posttranslational modification by N-acetylation, adenylation, and proteolytic processing was demonstrated by quadruple time-of-flight mass spectrometry (Q-TOF MS MS).

5
Genetic Tools for *In Vivo* Genetic Manipulation of *Streptomyces*

Many research groups investigating the molecular genetics of *Streptomyces* confine their activities to *S. coelicolor*, or the closely related laboratory "workhorse," *S. lividans*. This is understandable, given that the genome sequence is available and postgenomic resources have been designed and developed for this application. Furthermore, there is an excellent Laboratory Manual, written by Kieser et al. in 2000, that provides comprehensive protocols and information resources for these species. However, neither strain makes a commercially-important product, so applied researchers often have to adapt and modify the existing protocols for the strains that they study. Many protocols are transportable, but some are not.

In conducting recombinant DNA experiments, the transformation efficiency (transformants per μg of DNA) of the strain under study is very important. Construction of a representative gene bank, which will give a statistically-significant number of recombinants to screen, requires a transformation efficiency of at least 10^5/μg DNA and preferentially 10^6. This is easily achievable with *S. coelicolor* and *S. lividans*, but industrial strains are often recalcitrant to formation of protoplasts and, if this can be achieved, are often fastidious in their requirements of the agar medium that they will regenerate viable mycelia from these protoplasts. The protocols need extensive modification and optimization and, for some species, an acceptable efficiency is never achieved.

Two other methods may (or may not) prove useful in introducing DNA to the strain of interest. For some streptomycetes, DNA can be introduced by electroporation. This circumvents the need to develop protocols to make protoplasts and regenerate them. Experiments can be conducted more quickly, as electrotransformed cells grow rapidly on conventional agar (often in two days), whereas protoplasts can take seven to ten days to regenerate. The original protocol, for *S. rimosus*, was published by Pigac and Schrempf in 1995. Very few examples of successful adaptation to other species have been reported subsequently, and the technique is certainly not applicable to *S. coelicolor*, or *S. lividans*. The procedure seems to work well with species that grow with short fragmented filaments, for example, *Streptomyces albus*, so that characteristic may be a benchmark to indicate the likelihood of success. Certainly, the approach should be attempted, as success carries the benefits detailed earlier.

The second method is to introduce DNA by conjugation from *E. coli*. This unlikely strategy actually works very well. The vector DNA obviously has to contain an origin of replication for *E. coli*, but it may contain either a replication origin for *Streptomyces* (so that the delivered vector replicates as an autonomous plasmid in *Streptomyces*), or no streptomycete replication origin at all (so that the delivered DNA integrates into the *Streptomyces* chromosome). The other key component of the system is that the

vector has to contain a "transfer origin," *oriT*, to enable the DNA to be mobilized from *E. coli* to *Streptomyces*. The *oriT* component is a standard feature of *E. coli* plasmids that conjugate among *E. coli*. The last requirement is that the *E. coli* donor strain contains the genes necessary for promoting conjugal transfer – these genes may be resident on the chromosome, or supplied on a second plasmid. The system works extremely well for *S. coelicolor* and *S. lividans*, and has been adapted for industrial strains, such as *S. clavuligerus* and *S. rimosus*. It is now the method of choice for introducing DNA to *Streptomyces* and has superseded the protoplast route. The advantages of making all of the DNA constructs in *E. coli* and then mating them directly into *Streptomyces*, thus saving time, have been very compelling.

The final consideration while introducing DNA to *Streptomyces* is that many species have extremely active restriction systems. This is particularly true of industrial strains. Thus, DNA prepared in one host will not transform into another, as the newly-introduced DNA is recognized as "foreign" by the recipient and destroyed by the endogenous restriction enzymes. It is therefore crucial to test the prospective host strain to ensure that it does not restrict the DNA from the donor strain. This is most easily achieved by preparation of vector DNA from the donor strain and then transforming it into the recipient host. Vector DNA prepared from the recipient host acts as the control for this experiment. If there is no (or minimal) restriction barrier, the transformation efficiency of DNA prepared from both sources will be equivalent. A reduction in transformation efficiency when DNA from the donor is used (compared to the control when the DNA has been prepared in the recipient) indicates that the recipient has restriction enzyme activity for the DNA prepared from the donor strain. The DNA prepared in *E. coli*, or conjugated from *E. coli*, is heavily restricted by most *Streptomyces*. However, the problem may be circumvented by using a modification-deficient (*dam*, *dcm*) *E. coli* strain. Indeed, such a strain is now the preferred donor for conjugation, as detailed in the paragraph earlier.

5.1 Plasmids

Plasmids are essential components of the "tool kit" to develop the molecular genetics of any system. *Streptomyces* are endowed with many indigenous plasmids, so the initial cloning vectors were derived from them. Natural plasmids of *Streptomyces* fall into two categories: linear and circular.

The linear plasmids have often been termed *giant linear plasmids* (GLP's) because of their size. The best characterized example is SCP1 (*Streptomyces coelicolor* plasmid 1), which is of the order of 350 kb, but small linear plasmids have also been characterized (e.g. the 11.7-kb plasmid, pSCL1, from *S. clavuligerus*). Linear plasmids share a common feature with the chromosome in their possession of terminal proteins. As most protocols for plasmid isolation are for circular plasmids, it may well be that there are many more small linear plasmids to be discovered – simply by using the relevant protocol. Many of the GLPs are fertile and can conjugate between strains. Some integrate into the chromosome, for example, SCP1 has a preferred site of integration near the center of the linear chromosome, and then promote mobilization of chromosomal markers into the recipients by conjugation. This fertility scheme has outline similarity to the formation and mating of *Hfr* strains in *E. coli*.

Circular plasmids range from a few kilobases to many tens of kilobases in size. In the natural state, the smaller plasmids often have very high copy numbers (around 50 per cell), whereas, the larger versions are closer to unit copy number. The best characterized low-copy vector is SCP2, a 31-kb fertility plasmid of *S. coelicolor*. This natural plasmid has no selectable marker but its presence within a cell can be recognized by "lethal zygosis" – a phenomenon whereby a colony containing the plasmid causes a translucent zone of inhibition of growth in the lawn of surrounding cells, which do not contain the plasmid. Lethal zygosis is still a useful attribute when screening novel streptomycetes for plasmids, as both linear and circular plasmids display the phenomenon.

More advanced versions of SCP2 have been developed (e.g. pIJ922, pIJ916) by reducing the plasmid's size and introducing a selectable marker. These engineered derivatives still maintain their low copy number, and versions are available that retain the fertility/transfer genes, or have had them deleted. The utility of low-copy vectors in cloning entire antibiotic pathways has been ably demonstrated over the years. For example, the first demonstration was the cloning of the actinorhodin pathway by Malpartida and coworkers in 1984. Entire antibiotic pathways do not express in higher-copy vectors, probably because chromosomally-derived transactivating factors are then titrated out.

High-copy vectors for *Streptomyces* are almost all derived from the natural plasmid, pIJ101. Engineered versions have had extraneous DNA removed, together with the transfer functions to derive small (~4 kb) minimal replicons. Thereafter, antibiotic-resistance genes to enable selection (see Sect. 7) have been added (e.g. pIJ350) and other marker genes, such as tyrosinase (which confers a pigmented phenotype, therefore insertions within the gene are colorless, e.g. pIJ702). Transcriptional fusion versions (pIJ486 and pIJ487) have cloning sites just upstream of a kanamycin-resistance reporter gene, so that the level of resistance reflects the transcriptional potential of the cloned DNA. Lastly, there is a positive selection vector (pIJ699), in which a "stuffer" fragment is removed and replaced by cloned DNA. Only recombinants containing cloned DNA propagate and so cloning experiments conducted with this vector are assured 100% success of containing insert DNA.

High-copy vectors carrying highly expressed genes can often be unstable, or impair the performance of the streptomycete host. A growing realization of this problem stimulated the development of a class of integrating vectors, which use site-specific recombination to insert into the chromosome at a fixed location. Components of the actinophage ΦC31 have been adapted for this purpose. They all contain the attachment site for the phage, together with the integrase that undertakes the recombination event. The vectors may be propagated in *E. coli* and the various constructions made in this host. Then, through interspecific conjugation, the plasmid is transferred to the streptomycete, where it integrates at the attachment site. Plasmid pSET152 is the most commonly-used version of this type of vector.

5.2 Bacteriophages

Soil provides a rich source for the isolation of *Streptomyces* phages and most soils yield phages after enrichment, some

of which are specific for different species while others are polyvalent. Several phages have been developed as genetic tools for the transfer of DNA encoding biosynthetic enzymes between different streptomycete species. By far the best characterized and most useful actinophage tool is the temperate phage ΦC31 that infects both *S. coelicolor* and *S. lividans*. The frequency of lysogenization of ΦC31 is 20 to 40% and 1% of germinated lysogenic spores spontaneously release phage. The phage ΦC31 has a wide host range and infects about half of the 137 strains that it has been tested against. It is naked, showing a polyhedral head (53 nM wide), a long noncontractile tail (100 × 5 mm), a basal plate (15 nm in diameter) with at least one pin and a prominent knot between head and tail. ΦC31 is unable to transduce host DNA and adsorbs to a glycoprotein of susceptible cells by its tail. Newly germinated mycelia are especially susceptible to ΦC31, although it has been suggested that older mycelia are more efficient at adsorbing actinophages. The DNA enters the cell by an unknown mechanism in a nonpermuted linear form. The molecule then circularizes at the cohesive ends, which are then ligated to give a covalently-closed circular molecule. In the case of a lysogenic infection, the prophage is inserted in the *S. coelicolor* chromosome at gene *SCO3798*, which encodes a possible chromosome condensation protein. The establishment and maintenance of lysogeny involve the action of a single repressor protein, the reduction in concentration of which, such as by hyphal fusion, can result in zygotic induction. Induced lysogens show a latent period of 45 min, followed by a rise period pf 20 to 30 min giving a burst size of 100 to 200 virus particles. DNA replication involves a ''rolling circle'' mechanism generating concatameric molecules; the packaging of these into mature phages involves the site-specific cleavage of the concatamers into monomers. The genome of ΦC31 is 41 491 bp in length and 63.6% G + C. The arrangement of genes indicates that transcription of all genes, except one, is in the left to right organization. As in other phage genomes, early and late genes are clustered, with a further subclustering of genes according to function.

Few transducing phages have been characterized to date. SV1 and related phages, infecting the chloramphenicol-producing strains *Streptomyces venezuelae* and *Streptomyces phaeochromogenes*, were the first examples of generalized transducing phages. The prophage of ΦSF1 exists as two plasmids and is capable of transducing *S. fradiae*. Recently, a phage infecting *S. coelicolor* has been isolated, which is reported to be capable of transducing this species.

5.3 Transposons

A possible contribution to the inherent genetic instability in streptomycetes is the large number of transposons found within the *S. coelicolor* genome, as deletion of intervening sequences is a consequence of an intramolecular transposition event. There is some circumstantial evidence to support this as the terminal DNA of the *S. coelicolor* and *S. avermitilis* chromosomes contain large numbers of transposons that may generate the genetic instability associated with the chromosome ends. It is thought that these regions act as ''transposon traps'' due to the relatively low numbers of essential genes in the chromosome arms. Several streptomycete transposable elements have now been discovered by screening disrupted genes or by direct inspection of the

genomic sequence. The only compound transposon found to date in *Streptomyces* is within an unstable region of *S. lividans* and consists of two copies of IS*1373* that flank mercury resistance genes, although transposition of only one of the insertion sequences has been proven experimentally. The use of transposon mutagenesis in streptomycetes has been limited, in part because of the high target-site specificity exhibited by many indigenous streptomycete transposable elements. In addition, many transposons are delivered to *Streptomyces* on a temperature-sensitive delivery plasmid by protoplast transformation. At least in *S. coelicolor*, both protoplasting and incubation at high temperature can be mutagenic and can lead to unlinked mutations. Two indigenous transposons with a relatively low degree of target-site specificity are the phenotypically cryptic Tn*4556* from *S. fradiae* and IS*493* from *S. lividans*, and a consensus target sequence for the latter has been determined to be NgNCaNTgNNgN. More recently, two nonstreptomycete transposons have been put to use in the genus: first, the mycobacterial insertion sequence, IS*6100*, and more recently Tn*5493*, based on the Gram-negative, "cut and paste" transposon, Tn*5*, that also forms the basis of the *in vitro* transposition described elsewhere in this review. *In vitro* transposition has also been reported using Tn*5* and *mariner* with *S. coelicolor* DNA shotgun cloned in an *E. coli* plasmid. After localization of insertions by sequencing of the mutated plasmids, the disruptions were transferred into the *S. coelicolor* chromosome through marker replacement either by protoplast transformation or conjugation; the latter relying on an origin of transfer from *E. coli* conjugative plasmids in the cloning vector. IS*6100* was isolated as part of the compound transposon Tn*610* from *Mycobacterium fortuitum* and belongs to the IS*6* family of transposable elements, which form cointegrates as an end point of transposition. It encodes a transposase gene flanked by two 14-bp perfect inverted repeats. On the basis of the Southern hybridization studies, there is no evidence for preferred target site selection by IS*6100* in either *S. lividans* or *S. avermitilis*. A minitransposon, Tn*1792*, based on IS*6100* has also been developed and successfully used in *S. lividans* and *S. avermitilis*. The first large-scale global library of *in vivo* transposon mutants to be produced for *S. coelicolor* was generated using Tn*4560*, originally derived from Tn*4556*. This transposon had previously been used on a small scale, but it was only when it was delivered by conjugation from *E. coli* on a suicide vector that its full value became apparent. The result was a library of over 10^5 independent insertions that could easily be characterized by sequencing of ligation-mediated PCR products.

6 Genetic Tools for *In Vitro* Genetic Manipulation of *Streptomyces*

The predicted 7825 *S. coelicolor* orfs encode many proteins that represent new families as well as an expansion of existing groups. An essential prerequisite for understanding the biological function of this collection of relatively uncharacterized genes is the availability of a defined, comprehensive mutant library. In this way, the phenotypic effect of inactivating individual genes can be investigated systematically. In *Streptomyces*, the process of systematic mutagenesis has proceeded relatively slowly, due to the complexity of the protocols available to generate and identify

the location of a given mutation. While these classical approaches to the isolation and study of streptomycete mutants will continue to be used, two techniques have appeared recently that have greatly facilitated the isolation and characterization of mutants in *S. coelicolor*. Both procedures make use of the *S. coelicolor* cosmid library that provides a resource of cloned DNA covering the entire genome.

6.1 *In Vitro* Shuttle Transposon Mutagenesis

Gene disruption by *in vitro* shuttle transposon mutagenesis is achieved by adapting *in vitro* transposon mutagenesis, using a derivative of Tn*5*, to obtain insertions in representatives of the ordered cosmid library that was previously used to sequence the *S. coelicolor* genome. A transposon insertion library of a cosmid is recovered initially in *E. coli* and the location of each insertion then determined by DNA sequencing. The inclusion of an origin of transfer (*oriT*) in the transposon allows each cosmid to be subsequently mobilized into *S. coelicolor* by conjugal transfer, yielding transconjugants at high efficiency with the desired gene disruption. The minitransposon, Tn*5062* (Fig. 1), was constructed to mutagenize the *S. coelicolor* cosmid library is flanked by two Tn*5* mosaic ends and allows for *in vitro* transposition catalyzed by hyperactive Tn*5* transposase. Tn*5062* also contains a three-frame translational stop and *Streptomyces* consensus ribosome-binding site upstream of a promoterless *egfp* reporter gene to generate transcriptional fusions with the mutated gene. It also carries an apramycin-resistance gene for selection in *Streptomyces* and *E. coli* and the *oriT* origin of transfer; the latter enabling transfer of a resulting transposed cosmid by conjugation from a suitable *E. coli* donor strain to *S. coelicolor*. As the transposed cosmid does not contain a *Streptomyces* origin of replication, recovery of the transposon-borne marker depends on replacement of the wild-type gene with the disrupted cosmid-borne copy by allelic replacement and can easily be obtained by selecting for apramycin-resistant, kanamycin-sensitive transconjugants (Fig. 2). Double crossovers are obtained from over 10% of

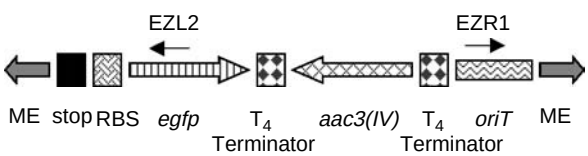

Fig. 1 Organization of Tn*5062*. The transposon consists of two inverted repeats (ME; Mosaic Ends) flanking an internal region containing the following elements: (1) a sequence consisting of translational stop codons in all three reading frames (stop); (2) a consensus streptomycete ribosome site(RBS); (3) a promoterless copy of the green fluorescent protein gene (*egfp*); (4) an apramycin-resistance gene (*aac3(IV)*) flanked by two T4 transcriptional terminators (T_4); and (5) the RK2 origin of transfer (*oriT*). Also indicated are positions of two sequencing primers, EZL2 and EZR1, used to precisely determine the location of each insertion. The sequence of Tn*5062* is deposited at EMBL (Accession number AJ566337).

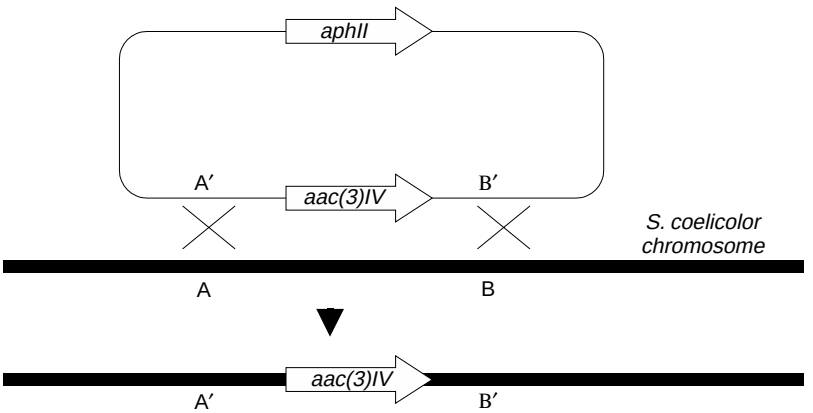

Fig. 2 Mutant generation in *S. coelicolor* after Tn*5062*- or PCR-targeted cosmid mutagenesis. Following introduction of the PCR targeting cassette or Tn*5062* (both apramycin resistant) into a cosmid that is marked with the kanamycin-resistance gene by PCR-targeted mutagenesis or *in vitro* transposition. A conjugative strain of *E. coli*, such as ET12567(pUZ8002), is transformed with the mutant cosmid. This strain is then mated with the recipient streptomycete, making use of the RK2 *oriT* region present in both cassettes, and following a double crossover between the flanking arms (A′ × A and B′ × B) of the knockout/disruption cassette, mutants in the desired gene can be isolated by selecting for apramycin-resistance and kanamycin-sensitive transconjugants.

primary transconjugant colonies due to the long flanking regions of the cosmid insertion. An appealing feature of this approach is that mutants can readily be moved into alternative genetic backgrounds and related species. Similarly, as transposition is carried out *in vitro*, an inability to recover viable transconjugants is indicative of the essential nature of that gene. Mutants are generated through gene disruption rather than deletion and as a result polar effects in many insertions are to be expected, although this may give information on domain functionality in some circumstances. *In vitro* transposon shuttle mutagenesis is, by its nature, a high-throughput technique and is designed to identify the site of the transposon insertion prior to generation of the mutant in *Streptomyces*. Eventually, this transposon library will consist of around 60 000 mutants. Many transposon insertions are located at intergenic sites or are in the wrong orientation to make use of the promoterless *egfp* reporter gene. However, sequencing DNA flanking the transposon of many thousands of transposed cosmids generated a massive amount of data, and, for this reason, a software application, Transposon Express (www.swan.ac.uk/genetics/dyson/InstallTE/), was developed to swiftly analyze a large data set of independent insertions created using Tn*5062*. This revealed a transposition target site consensus sequence of GCCCNGGGC in *Streptomyces* DNA and indicated an insertional preference for GC-rich streptomycete DNA compared to *E. coli* vector DNA. Full details of Tn*5062* insertions in the *S. coelicolor* genome are available at the *S. coelicolor* genome database, ScoDB (http://streptomyces.org.uk/). This provides access to a wide range of information relevant to *Streptomyces* research, such as

a full description of each cosmid, the annotation and gene ontologies of each gene, as well as information on primer design for PCR-targeted mutagenesis.

6.2 PCR-targeted Mutagenesis

A complementary technique based on PCR-targeted mutagenesis, known as ReDirect, was developed in parallel with *in vitro* transposon shuttle mutagenesis for use in *Streptomyces*. This method is based on generalized PCR targeting used in yeast to obtain homologous recombination via short stretches of sequence homology. In bacteria, homologous recombination at frequencies that make the technique practicable generally requires more sequence homology than can be supplied by PCR primers. However, by exploiting the recombination potential of phage λ, *E. coli* cells expressing λ red proteins can be used to enhance recombination frequencies via short homologous sequences and so generate detectable numbers of recombinants. As a result, this approach was applied to generate gene knockouts in the *S. coelicolor* cosmid library. Initially, an *E. coli* strain carrying the λ *red* genes under the control of the P_{araBAD} promoter is transformed with one of the *S. coelicolor* cosmids. In the meantime, an antibiotic-resistance gene coupled to *oriT* and selectable in both *E. coli* and *Streptomyces*, is amplified using primers whose 39 bp 5′ ends are homologous to DNA flanking a gene selected for deletion in the cosmid. This amplicon is used to transform the cosmid-carrying *E. coli* strain, where the λ *red* genes are induced with arabinose and transformants selected with the amplicon-borne resistance marker. As a result of homologous recombination between the 5′ ends of the amplicon and the *S. coelicolor* cosmid, the selected gene is deleted and the antibiotic-resistance marker is introduced into the cosmid. The equivalent region in the *S. coelicolor* chromosome can now be replaced in an analogous manner to introduction of Tn*5062* insertion (Fig. 2). This technique is less efficient than Tn*5062* mutagenesis in terms of throughput, but offers a greater degree of flexibility through the use of alternative resistance markers and reporter genes, as well as the generation of true null mutants. PCR-targeted mutagenesis has now been used to generate many hundreds of mutants in *S. coelicolor* and other strains.

7 Antibiotic Selection

More advanced vectors have been developed from naturally-occurring plasmids. They contain drug-resistance markers, which (of course) facilitates the selection of transformants. As streptomycetes are naturally-resistant to many antibiotics (e.g. the ß-lactams), a very limited number of quite unusual antibiotic-resistance genes have been used. Among these, the gene that confers resistance to thiostrepton is most popular – because most streptomycetes are sensitive to the drug. Thiostrepton acts as an antibiotic by binding to the ribosome and arresting translation. This resistance factor, cloned from thiostrepton-producer *Streptomyces azureus*, methylates host ribosomal RNA and prevents binding of the drug to the ribosome. Perhaps the most useful drug-resistance markers are those that encode resistance to antibiotics that are selective in both *E. coli* and *Streptomyces* and include *aac(3)IV* (apramycin), *aphII* (neomycin/kanamycin), *hph* (hygromycin),

aadA (streptomycin/spectinomycin), *aacI* (gentamicin), and *ermE* (erythromycin) in a decreasing order of efficacy. The first two of these selective markers are used in Tn*5062* mutagenesis and PCR-targeted mutagenesis, and as such are used for the selection of allelic replacements by mutated cosmids. pSupercos1, the vector used for generating the *S. coelicolor* cosmid library, encodes resistance to kanamycin (*aphII*) and ampicillin (*β-la*). As a result of this, and the *aac(3)IV* (apramycin) gene employed in the two *in vitro* protocols, the further use of these resistance genes is restricted in subsequent manipulations. Similarly, the most common strain used for the conjugative delivery of *oriT*-marked vectors, such as *E. coli* ET12567(pUZ8002), is already marked with chloramphenicol, tetracycline, and kanamycin resistance and may affect the choice of vectors delivered by conjugation.

8 Morphological Development

The complex life cycle of *Streptomyces* initiates with the germination of a dormant spore with subsequent hyphal emergence. Unlike unicellular bacteria, in which daughter cells separate, *Streptomyces* cells remain associated, forming long hyphal filaments that grow by apical extension. There are two kinds of cell division in *Streptomyces*: (1) vegetative septation leading to crosswall formation in substrate hyphae; and (2) sporulation septation leading to the formation of ovoid spores. Both involve the same division machinery, involving FtsQ and FtsZ, which are dispensable for vegetative growth. Differentiation involves a cascade of signaling events with nucleoid migration and septation, creating a subapical daughter cell. Septation arises through the action of the tubulin homolog FtsZ and its polymerization into the cytokinetic Z ring, and is thought to involve several membrane proteins. Exponential growth is achieved by lateral branching, allowing the *Streptomyces* colony to eventually yield radial outgrowth with formation of a mycelial mat. During this first growth phase, *Streptomyces* generate a dense multicellular network within the nutrient substrate. This apical extension and branching of vegetative hyphae enables the colonization of the media until nutrients are exhausted.

Streptomyces then form elaborately coiled, hydrophobic aerial hyphae that rise away from the plane of radial growth. A variety of extracellular factors can trigger aerial hyphae formation. Genetic dissection of the developmental programme of *S. coelicolor* reveals that at least two classes of genes are required: the *ram* genes and the bald (*bld*) genes, mutants of which are defective in the formation of an aerial mycelium and grow as smooth hairless colonies. The *ram* genes were discovered by their ability to induce rapid aerial mycelium (hence *ram*) when introduced on a low-copy plasmid in *S. lividans*. There are four *ram* genes found as an operon, *ramCSAB*, where the binding of RamR, a response regulator, to the *ramC* promoter, activates transcription. Interestingly, a *ramR* mutation prevents production of SapB and formation of aerial hyphae. SapB is a small, secreted lantibiotic-like peptide that acts as a biological surfactant, releasing the surface tension at the air–water interface and allowing the aerial hyphae to grow upright. Until recently, the product of *ramS* was unknown, with the product of *ramC* showing similarity with serine/threonine protein kinases and *ramAB* encoding components of an ATP-binding cassette transporter. However, recent studies have revealed

that SapB is derived from the product of *ramS*, after posttranslational modification by *ramC*.

Analyses of *bld* mutants allowed the detection of various proteins that have important roles during differentiation, cell–cell signaling and carbon–catabolite repression. Evidence suggests that certain bald genes are part of an organized developmental system – a signaling network in which the restoration of aerial hyphal growth is achieved through extracellular complementation. This signaling cascade involves one *bld* mutant that acts as a donor and rescues morphogenesis in the recipient to the left of a defined hierarchy, leading to erection of aerial hyphae:

$$[bldJ] < [bldK] < [bldA, bldH] < [bldG] < [bldC] < [bldD]$$

In these studies, all *bld* mutants rescue development in *bldJ* mutants, while *bldD* mutants can only be rescued by the wild-type strain. Further complementation assays indicated that once morphogenesis is restored in the acceptor strain, this backtracks and restores aerial hyphae formation in the donor, suggesting that the molecules that result from expression of these genes work in a signaling cascade for the generation of aerial hyphae. Each signal molecule then stimulates the next, so that there is finally *bldD*-induced production of SapB. The stimulation of the *bld*-signaling cascade is dependent on environmental cues such as nutrient and cell density. Catabolite repression and the accumulation of organic acids may cause many of the bald mutants to induce an acidogenic environment on glucose-based media and so prevent differentiation. Acidogenesis is the effect rather than the cause of the phenotypic expression of some *bld* mutants, particularly with the *bldA*, *bldB*, *bldG*, and *bldH* strains. It is suggested that the accumulation of organic acids by these mutants may be a consequence of limited oxygen availability in older, densely-packed substrate mycelium, thus causing a metabolic imbalance of the TCA cycle. The projection of aerial hyphae, and therefore the increase of surface area, induces oxygen availability that allows renewed metabolic activity and consequently the consumption of organic acids. Observations on the developmental signaling of *Bacillus subtilis* support this hypothesis. The *bldA* gene encodes the leu-tRNA that reads the rare UUA codon, and controls the initiation of the reproductive phase and secondary metabolism in *S. coelicolor*. Meanwhile, *bldB* and *bldD* genes encode putative transcription factors that have potential roles in phosphorylation and ADP-ribosylation, leading to the next stage of morphological development.

Development of aerial hyphae and tip extension is accompanied by a number of other physiological changes including the accumulation of polymeric storage compounds such as glycogen and the production of secondary metabolites. Stationary phase is characterized by septation and the formation of compartments that are destined to become spores. Each compartment contains at least one copy of the chromosome to ensure spore viability. Subsequently, the walls thicken and round up to form ovoid spores in a chain that is confined within a fibrous sheath that ruptures after the spore matures. Coordination of the metabolic processes with the morphological events of sporulation, especially with sporulation-associated cell division, appears to be provided by a second hierarchy of gene expression, governed initially by two σ factors. The first, σ^{WhiG}, influences sporulation in aerial mycelium. Excess σ^{WhiG} can actually

change the developmental fate of hyphae, by causing ectopic sporulation in the substrate hyphae, which normally lyse during formation of aerial hyphae. The second sigma factor, σ^F, triggers the maturation of the spore compartments, allowing rounding and thickening of the cell wall to form the characteristic gray-pigmented spores. The activity of σ^F is blocked in *whiG* mutants, thus defining part of the cascade that controls development. This highlights only 2 of the 65 sigma factors that have been identified in the *S. coelicolor* genome sequencing project. It therefore allows only a glimpse into the complex nature of initiation of aerial mycelium and spore formation.

9
Antibiotic Production

The production of antibiotics and secondary metabolites in *Streptomyces* is integrated into the complex pathway of differentiation and development, which these microorganisms pursue. Understanding how the regulatory circuits are "wired up" is a topic of academic interest, and also one that has substantial commercial application. A large number of mutants have now been isolated in the model genetic organism, *S. coelicolor*. Some of these mutants fail to make only one of the antibiotics within the repertoire of this organism. When such lesions do not map to the structural genes of the biosynthetic pathway, the implication is that they must be mutated in an arm of the regulatory cascade, which is specific for that pathway. Other mutants fail to make all the antibiotics, but have normal morphologies – they must have a lesion "further up" in the network. Yet others are blocked in some part of the differentiation process and fail to make antibiotics – they must be mutated in a central part of the regulatory cascade. Sufficient mutants have now been analyzed by molecular genetics to begin to construct testable models of the regulatory cascades. It is now well-established that cells within the mycelial mat communicate with one other using specialized small molecular-weight signaling molecules. Changes in the environment are sensed and converted to a message, which can be interpreted by the biology of the organism through cyclic phosphorylation and dephosphorylation of key proteins. *Streptomyces*, which provide an important model for such universally used "signal transduction" systems, are likely to be a rich source of experimental data to increase our understanding of how such systems work.

10
Optimization of Gene Expression in *Streptomyces*

The high G + C content of their DNA places some constraints on *Streptomyces* on their use of synonymous codons within protein-coding regions. The overall composition of DNA within coding regions of streptomycete DNA is 73%. However, analysis of a large number of protein-coding regions by M.J. Bibb and coworkers showed that, within the three positions of bases within each triplet, the distribution was not equal. On an average, the first base of a triplet contained DNA of 73% composition, the second base of around 50% and the third of around 95% (making the average of 73%). In pragmatic terms, when there is a choice of codons for a particular amino acid (e.g. there are six codons for leucine), those that contain a G or C in the third base will be used most frequently. Thus, for leucine, TTA

will be used very infrequently because it has no G or C, TTG will be less preferred as two of the three bases are Ts, CTA and CTT will also be less preferred because of their third bases, but CTG and CTC will be used most frequently. When it comes to the design of oligonucleotide primers based on known amino acid sequences, this feature becomes very important. For example, if a leucine is present, there is a high probability that it will be encoded by CTG or CTC rather than by the other four synonymous codons. In the design of oligonucleotide probes, this reduces the degeneracy (i.e. the number of different bases required for each amino acid) and so these much less complex oligonucleotide mixtures are more specific for streptomycete genes – giving a much better signal on hybridization. This feature may be used also in analysis of DNA sequences for protein-coding regions. If the computer is asked to scan a DNA sequence and tabulate the G + C content in each of the three positions, it is relatively easy to deduce where the likely orfs may lie.

These considerations also need to be taken into account in the design of reporter genes with which to monitor gene expression and cellular location of proteins. The array of reporter genes available for use in *Streptomyces* is more limited than those that are available in many other bacteria. ß-galactosidase is produced by all streptomycetes, and, as a result, the use of the chromogenic substrate X-GAL, is restricted to those strains in which a mutant in the gene encoding this enzyme has been created. Perhaps the most commonly used reporter gene that makes use of a chromogenic substrate in *Streptomyces* is the *xylE* gene from *Pseudomonas*, which has been used to monitor expression of many genes as a transcriptional fusion. More recently, a version of the *luxAB* cassette has been created in which all TTA codons have been replaced and the proximal few codons of both genes altered to reflect the high G + C content of streptomycete DNA. Luciferase, encoded by *luxAB*, is a very sensitive system for monitoring gene expression and, in conjunction with luminometry, provides a very sensitive and accurate system for quantifying gene expression. One drawback with this system is that the substrate, *n*-decylaldehyde, must be added exogenously and is toxic. However, when light emission is monitored by means of a 96-well luminometer, many samples can be exposed to the substrate simultaneously and a large amount of quantitative data, suitable for statistical analysis, can be generated in a short time. The best reporter system for the spatial monitoring of gene expression in *Streptomyces* is Green Fluorescent Protein. Unfortunately, the codon usage of wild-type GFP is not optimal for use in *Streptomyces*, and as a result, the enhanced fluorescence version (eGFP) is routinely used and has been used successfully to demonstrate the tissue-specific expression of various genes when used as a transcriptional fusion. Similarly, it has also been used to demonstrate the exact cellular location of proteins involved in cell division, such as FtsZ.

11 Conclusions

In the 50 years that *Streptomyces* have been studied, classical genetic systems have been developed and mutants isolated to provide the framework for the current campaign of postgenomic investigation. With the completion of the genome sequences, there has been development

in postgenomic techniques to interrogate transcriptomes and proteomes, and there has also been development of mutant libraries and rapid methods to make specific mutation and introduce them rapidly back into the chromosome; the next few years will see a quantum leap in our understanding of these important bacteria.

See also Bacterial Cell Culture Methods; Bacterial Growth and Division; Bacterial Pathogenesis, Molecular Basis of; Protein Microarrays; Proteomics; Total Analysis Systems, Micro; Vector System: Plasmid DNA.

Bibliography

Books and Reviews

Chater, K.F. (2001) Regulation of sporulation in *Streptomyces coelicolor* A3(2), *Curr. Opin. Microbiol.* **4**, 667–673.

Chater, K.F., Horinuchi, S. (2003) Signalling and early developmental events in two highly diverged *Streptomyces* species, *Mol. Microbiol.* **48**, 9–15.

Chater, K.F., Bucca, G., Dyson, P., Fowler, K., Gust, B., Herron, P.R., Hesketh, A., Hotchkiss, G., Kieser, T., Mersinias, V., Smith, C.P. (2002) *Streptomyces coelicolor* A3(2): from genome sequence to function, *Funct. Microb. Genomics Methods Microbiol.* **33**, 321–336.

Chen, C.W., Huang, C.-H., Lee, H.-H., Tsai, H.-H., Kirby, R. (2002) Once the circle has been broken: dynamics and evolution of Streptomyces chromosomes, *Trends Genet.* **18**, 522–528.

Dyson, P.J. (2000) *Streptomyces Genetics in Encyclopaedia of Microbiology*, Alexander, M., Bloom, B.R., Laskin, A.I., Iglewski, B.H., Schaechter, M. (Eds.) Academic Press, New York.

Flärdh, K. (2003) Growth polarity and cell division in *Streptomyces*, *Curr. Opin. Microbiol.* **6**, 564–571.

Hopwood, D.A. (1999) Forty years of genetics with *Streptomyces*: from *in vivo* through *in vitro* to *in silico*, *Microbiok.* **145**, 2183–2202.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., Hopwood, D.A. (2000) *Practical Streptomyces Genetics*, The John Innes Foundation, Norwich. ISBN 0-7084-0623-8,

Paradakar, A., Trefzer, A., Chakraburtty, R., Stassi, D. (2003) *Streptomyces* genetics a genomic perspective, *Crit. Rev. Biotechnol.* **23**, 1–27.

Primary Literature

Baltz, R.H., McHenney, M.A., Cantwell, C.A., Queener, S.W., Solenberg, P.J. (1997) Application of transposon mutagenesis in antibiotic producing streptomycetes, *Antonie Van Leeuwenhoek* **71**, 179–187.

Bentley, S.D., Chater, K.F., Cerdeno-Tirraga, A.-M., Challis, G.L., Thomson, N.R., James, K.D., Hams, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandrat, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Gable, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.-H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbanowitsch, E., Rajandream, M.-A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Parkhill, J., Hopwood, D.A. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2), *Nature* **417**, 141–147.

Bishop, A., Fielding, S., Dyson, P.J., Herron, P.R. (2004) Concerted mutagenesis of a streptomycete genome: a link between osmoadaptation, development and antibiotic production, *Gen. Res.* **14**, 893–900.

Bucca, G., Brassington, A.M.E., Hotchkiss, G., Mersinias, V., Smith, C.P. (2003) Negative feedback regulation of dnaK, clpB and lon expression by the DnaK chaperone machine in *Streptomyces coelicolor* identified by transcriptome and in vivo DnaK-depletion analysis, *Mol. Microbiol.* **50**, 153–166.

Chung, S.-T., Crose, L.L. (1989) *Streptomyces transposon Tn4556 and its Applications in Genetics and Molecular Biology of Industrial Microorganisms*, Hershberger, C.L.,

Queener, S.W., Hegman, G. (Eds.) American Society for Microbiology, Washington, DC.

Gehring, A.M., Yoo, N.J., Losick, R. (2001) RNA polymerase sigma factor that blocks morphological differentiation by *Streptomyces coelicolor*, *J. Bacteriol.* **183**, 5991–5996.

Gehring, A.M., Nodwell, J.R., Beverley, S.M., Losick, R. (2000) Genome wide insertional mutagenesis in Streptomyces coelicolor reveals additional genes involved in morphological differentiation, *Proc. Natl. Acad Sci. U.S.A.* **97**, 9642–9647.

Güst, B., Challis, G.L., Fowler, K., Chater, K.F. (2003) PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odour geosmin, *Proc. Nat. Acad. Sci. U.S.A.* **100**, 1541–1546.

Herron, P.R., Evans, M.C., Dyson, P.J. (1999) Low target site specificity of an IS*6100*-based mini-transposon, Tn*1792*, developed for transposon mutagenesis of antibiotic-producing *Streptomyces*, *FEMS Microbiol. Lett.* **171**, 215–221.

Herron, P.R., Hughes, G., Govind, C., Fielding, S., Dyson, P.J. (2004) TRANSPOSON EXPRESS, a software application to report the identity of insertions obtained by comprehensive transposon mutagenesis of sequenced genomes: analysis of the preference for *in vitro* Tn*5* transposition into GC-rich DNA, *Nucleic Acids Res.* **32**, e113.

Hesketh, A.R., Chandra, G., Shaw, A.D., Rowland, J.J., Kell, D.B., Chater, K.F. (2002) Primary and secondary metabolism, and post-translational protein modifications, as portrayed by proteomic analysis of *Streptomyces coelicolor*, *Mol. Microbiol.* **46**, 917–932.

Huang, J., Lih, C.-J., Pan, K.-H., Cohen, S.N. (2001) Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays, *Genes Dev.* **15**, 3183–3192.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kicuchi, H., Shiba, T., Sakaki, Y., Hattori, M., Omura, S. (2003) Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis, *Nat. Biotechnol.* **21**, 505–506.

Malpartida, F., Hopwood, D.A. (1984) Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host, *Nature* **309**, 462–464.

Pitman, A., Herron, P.R., Dyson, P.J. (2002) Cointegrate resolution following transposition of Tn*1792* in *Streptomyces avermitilis* facilitates analysis of transposon-tagged genes, *J. Micro. Methods* **49**, 89–96.

Redenbach, M., Kieser, H.M., Denapaite, D., Eichner, A., Cullum, J., Kinashi, H., Hopwood, D.A. (1996) A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome, *Mol. Microbiol.* **21**, 77–96.

Vohradsky, J., Li, X.M., Dale, G., Folcher, M., Nguyen, L., Viollier, P.H., Thompson, C.J. (2000) Developmental control of stress stimulons in *Streptomyces coelicolor* revealed by statistical analyses of global gene expression patterns, *J. Bacteriol.* **182**, 4979–4986.

Volff, J.-N., Altenbuchner, J. (1997) High frequency transposition of the Tn*5* derivative Tn*5493* in *Streptomyces lividans*, *Gene* **194**, 81–86.

Yang, M.C., Losick, R. (2001) Cytological evidence for association of the ends of the linear chromosome in *Streptomyces coelicolor*, *J. Bacteriol.* **183**, 5180–5186.

Stress-Induced Expression of Heat Shock Proteins and Action of the Heat Shock Protein Effectors

Bernd Bukau[1], *Elizabeth A. Craig*[2], *Richard I. Morimoto*[3], *and Arthur L. Horwich*[4,5]
[1] *University of Heidelberg, Heidelberg, Germany*
[2] *University of Wisconsin, Madison, WI, USA*
[3] *Northwestern University, Evanston, IL, USA*
[4] *Yale School of Medicine, New Haven, CT and The Scripps Research Institute, La Jolla, CA, USA*
[5] *HHMI, Chevy Chase, MD, USA*

1 Sensing of Protein Misfolding in Bacteria: Heat Shock and Envelope Stress Responses of *E. Coli* 466
1.1 The Cytosolic Heat Shock Response 466
1.1.1 Inducers, Sensors, and Signals 466
1.1.2 The Regulator 466
1.1.3 The Modulators 469
1.2 THE σ^E-dependent Envelope Stress Response 471
1.2.1 Inducers, Sensors, and Signals 472
1.2.2 The Regulator 473

2 Eukaryotic Heat Shock Response – Heat Shock Factors 476
2.1 Activation of Heat Shock Factor 1 479
2.2 Pharmacological Regulation of Eukaryotic Heat Shock Response 481

3 Unfolded Protein Response (UPR) of the Eukaryotic Endoplasmic Reticulum 483

4 Chaperone Effectors of the Heat Shock Response 485
4.1 Chaperonin (Hsp60) System 485
4.2 Hsp70 System 492
4.3 Small Heat Shock Proteins 500

5 Other Heat-inducible Components 503
5.1 Hsp90 503

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

5.2 Hsp33 and Oxidative Stress 505
5.3 Hsp15 and Free 50S Ribosomal Subunits 506
5.4 FtsJ 506

6 Chaperone-protease Pairs – ClpP, HslV, and Lon, and Associated ATP-dependent Chaperone Unfoldases, ClpA, ClpX, and HslU 506
6.1 Substrate Recognition and Translocation Mediated by the Chaperone Moiety 506
6.2 Proteolytic Components 508
6.3 HslU–HslV 508
6.4 Lon 509

7 Disaggregases – Prokaryotic ClpB and Eukaryotic Hsp104 509

8 Periplasmic Proteolytic System – HtrA Proteases DegP and DegS 509
8.1 DegP 510
8.2 DegS 512
8.3 Omi 512

Acknowledgment 512

Bibliography 513
Books and Reviews 513
Primary Literature 513

Keywords

Chaperonin
Ring structure that provides assisted folding to the native state, typically binding a nonnative protein in an open ring and then folding it after encapsulation, switching between the two states governed by ATP binding and hydrolysis.

Heat Shock Factor – HSF
A transcriptional regulatory factor that induces the expression of specific genes encoding the heat shock proteins and other components involved in the stress response, by binding to characteristic upstream elements.

Heat Shock Proteins
Specialized class of proteins whose abundance is thermally induced, many of which are molecular chaperones; the Hsps mediate the effector limb of the stress response, protecting proteins from misfolding and aggregation.

Molecular Chaperone
A specialized cellular protein that binds to nonnative proteins, stabilizing them or assisting their proper folding, but which is not associated with them in their final form.

Protein Aggregation
Multimolecular assembly of nonnative proteins; included are such structures as inclusion bodies and amyloid fibrils.

Stress Response
The cellular response to abnormal conditions, including abnormal temperature, oxidants, heavy metals, and organic solvents.

Unfolded Protein Response
The response of the eukaryotic secretory endoplasmic reticulum to the accumulation of unfolded proteins.

The heat shock response is a universal cellular response to stressful conditions, including not only heat exposure but also exposure to a variety of other noxious agents and conditions including oxidants, heavy metals, pH change, and nutrient starvation. In all cases, cells sense the stressful condition and respond by inducing the expression of a set of genes encoding so-called heat shock proteins (Hsps). Sensing, as reviewed below, appears in general to involve the recognition of misfolding proteins, as an immediate by-product of the stress conditions. Thus, there is a fairly uniform response mediated to the variety of different stresses. The induced heat shock proteins, whose structure and action is also reviewed here, include molecular chaperones and proteases, which act as agents that serve to protect the cell against the imposed stress. The chaperones act to stabilize misfolding proteins and assist them back to native form while in the case of proteases the action is to remove misfolded proteins by degradation. Notably, whereas such components were identified originally by their induction during stress, many of them have been subsequently observed to mediate the same actions, in many cases essential, under normal conditions. Included also in the stress-responding armamentarium are a number of components that bind to or modify nucleic acids and, while such components are at an earlier stage of study, these also may confer protection on nucleic acid containing machines such as chromatin and the ribosome.

The heat shock response can thus be conveniently considered as a ''sensing'' circuitry that mounts a transcriptional response to imposed stress, and an ''effector'' arm consisting of a variety of actions mediated by the heat shock proteins. We consider first the sensing system in bacteria and then eukaryotes, and then review the structure and action of the effectors.

1 Sensing of Protein Misfolding in Bacteria: Heat Shock and Envelope Stress Responses of *E. Coli*

Bacterial cells react to the accumulation of misfolded proteins by the induction of several different stress responses, which activate expression of heat shock genes encoding chaperones and proteases. These responses are compartment specific and therefore require the sensing of misfolded proteins in both cytosol and periplasmic space of gram-negative bacteria. Paradigms of such signaling pathways are the cytosolic heat shock response and the σ^E-dependent envelope stress response of *Escherichia coli*, both of which are described here.

1.1 The Cytosolic Heat Shock Response

The bacterial heat shock response was first described independently by Neidhardt and Yura in 1978 as reaction of *E. coli* cells to a rapid temperature upshift that allows them to survive thermal stress. Within seconds after the temperature upshift, the synthesis of at least 30 heat shock proteins (Hsps) are transiently induced, including the major chaperones (GroEL, GroES, DnaK, DnaJ, GrpE, HspG, ClpB, IbpA, and IbpB) and ATP-dependent proteases (Lon, ClpX, ClpP, HslU, and HslV) of the cytosol. The higher the temperature the longer the heat shock response and the higher the resulting Hsp levels. In case of a mild heat shock from 30 to 42 °C, the synthesis of Hsps peaks after 5 to 10 min at maximal rates of 10- to 15-fold above basal rates. The induction period is followed by an adaptation period in which the rate of Hsp synthesis decreases to a new steady state level (two to threefold above the pre-heat shock rate). The net result of this response is that the cellular levels of Hsps such as GroEL and DnaK increase by two- to fourfold upon shift to 42 °C, concomitant with an increased capacity for repair and degradation of misfolded proteins in the cytosol.

1.1.1 Inducers, Sensors, and Signals

The heat shock response is induced by a large variety of stress regimens including physicochemical factors and metabolically harmful substances (e.g. temperature upshift, pH and osmolarity changes, carbon source or amino acid starvation, oxidative stress, UV, ethanol, antibiotics, heavy metals, phage infection). Inducers have in common the potential to generate nonnative proteins, which are folding intermediates on the way to the native state, or misfolded proteins. Proteins are therefore the primary sensors of stress, and accumulation of aberrant proteins is an important signal that induces the heat shock response (Fig. 1). This scenario is strongly supported by the demonstration that the production of heterologous proteins, fusion proteins, mutant proteins, or protein fragments induces heat shock response in *E. coli*. However, the existence of additional sensors and signals such as membrane components cannot be excluded.

1.1.2 The Regulator

The heat shock regulon is positively regulated by the product of the *rpoH* gene, that is, the sigma 32 factor (σ^{32}) subunit of RNA polymerase (RNAP) which directs RNAP to promoters of heat shock genes, resulting in their transcription (Fig. 1). Heat shock promoters differ from regular promoters with respect to the – 35 region (CTTGAA, consensus), the – 10 region (CCCCATNT, consensus),

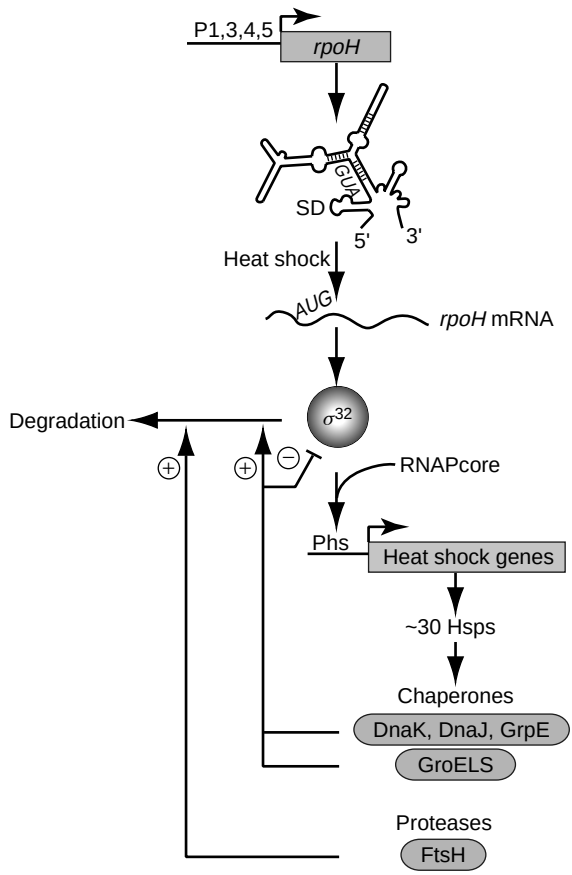

Fig. 1 Regulation of the heat shock response of *E. coli*. The positive regulator is the product of the *rpoH* gene, σ^{32}, which targets RNA polymerase to the promoters of heat shock genes. Induction of the heat shock response by temperature upshift is achieved by (1) heat-induced melting of the *rpoH* mRNA leading to its increased translation; (2) transient stabilization of the normally highly unstable σ^{32}; and (3) activation of σ^{32} allowing it to bind to RNA polymerase. The DnaK and GroEL chaperones with their cochaperones and the FtsH protease act as negative modulators of σ^{32}.

and the length of the spacer that separates these two regions (13–17 nucleotides). Some heat shock genes including *groEL*, *groES*, and *grpE* have additional promoter elements transcribed by the housekeeping sigma factor, which contribute to the basal levels of their expression. However, despite these additional promoters, σ^{32} is essential for growth of *E. coli* above 20 °C. Elegant genetic studies by the Yura laboratory revealed that the overproduction of a small set of chaperones with ATP-dependent protein folding activity, can rescue viability and protein aggregation defects of Δ*rpoH* cells up to 40 °C (GroELS overproduction) or even 42 °C (GroELS and DnaK, DnaJ, GrpE). These intriguing results demonstrate the essentiality of the heat shock regulon for viability of *E. coli* at most growth conditions, and the central importance of the GroEL and DnaK chaperone systems in protein quality control.

The cellular concentration of σ^{32} is very low, with 10 to 30 copies per cell at 30 °C, and is limiting for heat shock gene transcription. The temperature-induced increase in heat shock gene transcription is due to a rapid increase in the level of σ^{32} (15- to 20-fold at 5 min after shift from 30 to 42 °C), which results from increases in both synthesis and stability

of σ^{32}. Adaptation is achieved through a rapid repression of σ^{32} activity and, on a longer timescale, a decline in the levels of σ^{32}. Central to these controls of σ^{32} is the negative modulator function of chaperones and proteases.

1.1.2.1 **Control of σ^{32} synthesis.** The transcriptional regulation of *rpoH* is complex. Four promoters have been identified, three of them (P1, P4, P5) are transcribed by the housekeeping RNAPσ^{70} holoenzyme, while P3 is transcribed by RNAP containing the alternative σ^{E} (see Sect. 1.1.2), which ensures transcription of *rpoH* at extreme high temperatures at which σ^{70} is rendered inactive. Several DNA binding proteins regulate these promoters: the catabolite repression protein (CRP) activates P5; the initiator protein for DNA replication, DnaA, negatively modulates P3 and P4; the CRP–CytR protein complex negatively controls P4 and P5. This complex transcriptional regulation ensures that the levels of *rpoH* mRNA are sufficient to produce the basal levels of Hsps required for growth in a variety of environmental and metabolic conditions.

The temperature regulation of σ^{32} synthesis occurs primarily at the level of translational initiation, and relies on the formation of secondary and tertiary structures in the *rpoH* mRNA which sequester the ribosome binding site. An additional effect is that, upon increased translation during heat induction, the *rpoH* mRNA becomes stabilized and thereby accumulates. It should be emphasized that the translation regulation of *rpoH*, although contributing to the induction of heat shock gene expression after temperature upshift, does not contribute to the induction of a heat shock response at isothermic conditions, for example, following the accumulation of misfolded proteins.

The temperature-dependent structural changes in the *rpoH* mRNA require two cis acting elements in the coding sequence. A positively acting sequence element (region A) contains both the initiation site and a region of 15 nucleotides located immediately downstream of it (nucleotides +6 to +20 downstream of the AUG initiation codon). This "downstream box" is also found in unrelated mRNAs and acts by providing additional basepairing to the 16S rRNA, which increases translation efficiency. A negatively acting element (region B) required for temperature-dependent repression of *rpoH* translation is located between approximately +112 to +208 nucleotides downstream of the initiation codon. The repression of *rpoH* translation is mediated through the formation of secondary structures between the two elements, which occlude the translation initiation site. Consistent with this model, several mutational alterations of the *rpoH* mRNA, predicted to destabilize the base pairing between both elements, abolish the repression of *rpoH* translation. A clear correlation exists between the stability of the *rpoH* mRNA structure and the basal and heat induction of its translation, showing that intrinsic mRNA stability controls the synthesis of σ^{32}. The secondary structure formed between the two regulatory elements melts upon temperature upshift to 42 °C in the absence of any additional factor, demonstrating that the *rpoH* mRNA itself senses temperature and acts as a cellular thermometer.

1.1.2.2 **Control of σ^{32} degradation.** A second mechanism leading to the accumulation of σ^{32} during stress induction involves control of σ^{32} stability. At steady state temperature, either low or high, σ^{32} is extremely short lived *in vivo*, with a half-life of less than 1 min. Upon heat shock,

σ^{32} becomes rapidly stabilized (by eightfold upon shift from 30 to 42 °C), until during the adaptation period the stability of σ^{32} decreases again to a new steady state half-lifetime that is even shorter than at low temperature (10–15 s at 42 °C).

Degradation of σ^{32} *in vivo* is achieved mainly by the membrane associated AAA+ protease FtsH, although additional proteases such as HslUV, Lon, and ClpAP contribute to degradation as well. FtsH forms a ring-shaped oligomer with its catalytic center facing the cytosol. Degradation requires the energy of ATP, presumably for translocating σ^{32} to the catalytic center where cleavage takes place.

The molecular mechanism of the degradation process is poorly understood. A segment within the N-terminal part of σ^{32}, located within the conserved region 2, has been identified genetically as important for efficient degradation, but it is not clear whether this region constitutes a recognition site for the protease. FtsH differs from other AAA+ proteases of the *E. coli* cytosol in that it lacks a robust unfolding activity. This feature necessitates the substrate being rather unfolded before encountering FtsH. σ^{32} *per se* is indeed structurally unstable, which may facilitate its degradation. *In vitro* experiments showed that a shift of σ^{32} to heat shock temperatures within the physiological range leads to its partial unfolding, including region 2, and higher susceptibility to degradation. This suggests that the regulator of the heat shock response itself senses temperature changes. The physiological role of the structural changes within σ^{32}, leading to its increased instability at heat shock temperatures, may be to avoid unwanted overexpression of heat shock genes that are known to be inhibitory for growth.

However, the degradation process is more complex, as genetic evidence indicates a critical role for the DnaK chaperone system in efficient degradation (see the following). Whether this involves direct interaction between σ^{32} and the chaperone, perhaps by inducing a structural remodeling of σ^{32} which targets it for degradation, is unclear.

1.1.2.3 **Control of σ^{32} activity.** The control of activity of σ^{32} provides yet another layer of regulation of the heat shock response. This regulation is most apparent during a downshift of *E. coli* cells from heat shock to regular temperature, which leads to a strong and rapid repression of transcription of heat shock genes. By 15 min after a shift from 42 to 30 °C, the synthesis of GroEL decreases by 20-fold, although the level of σ^{32} decreases by only twofold. A similar repression of σ^{32} activity contributes to the regulation of σ^{32} activity at steady state conditions. This mode of regulation allows for the rapid adjustment of Hsp levels to altered requirements for chaperones and proteases, which seems crucial for cell growth.

The repression of σ^{32} activity involves the DnaK as well as the GroEL chaperone systems, both of which associate with σ^{32}. How this is achieved will be discussed in the next section.

1.1.3 The Modulators

The expression of heat shock genes in *E. coli* is regulated by the antagonistic action of the regulator σ^{32} and negative modulators (Fig. 2). Modulators are the DnaK and GroEL chaperone systems, which inactivate and destabilize σ^{32}, and to a lesser extent the FtsH protease, which is largely responsible for σ^{32} degradation. The generally accepted

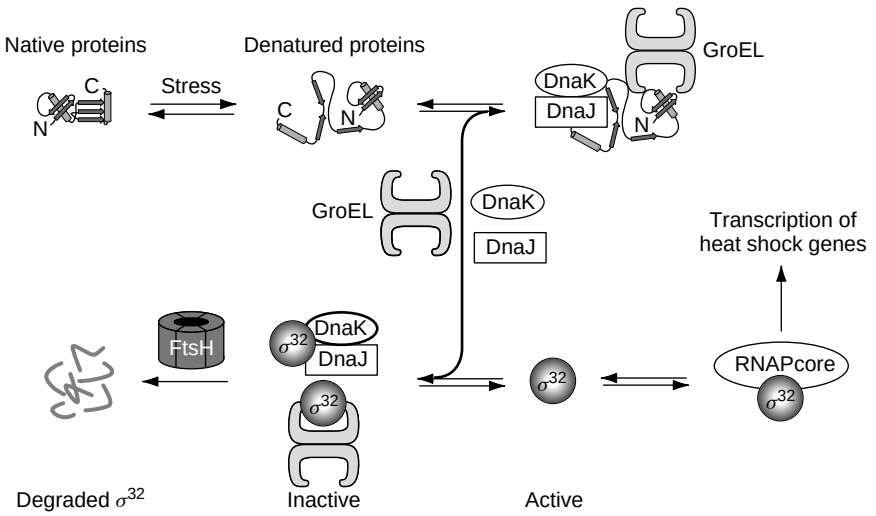

Fig. 2 Titration model of the heat shock response. DnaK and DnaJ negatively modulate σ^{32} by promoting its degradation by the FtsH protease and, through direct association, by preventing its association with the RNA polymerase core enzyme. GroEL also associates with σ^{32} resulting in negative modulation as well. Stress treatment of the cells causes accumulation of misfolded proteins in the cytosol and the sequestration of DnaK, DnaJ, and GroEL away from σ^{32}. This leads to stabilization and activation of σ^{32} and in consequence to induction of the heat shock response. Chaperone mediated repair of damaged proteins frees the chaperones to again repress and destabilize σ^{32} which reestablishes homeostasis with respect to heat shock gene expression.

model for the signal transduction pathway is that stress sequesters chaperones and proteases through binding to damaged proteins, which derepresses and stabilizes σ^{32}. Damage repair ameliorates the inducing signal and frees chaperones and proteases to shut off the heat shock response. This signal transduction is quite sensitive as it allows responses to relatively small changes in the amounts of both, chaperones and misfolded proteins. Fewer than 1500 molecules of a structurally unstable protein substrate of the DnaK system, firefly luciferase, produced per cell suffice to elicit a heat shock response. Furthermore, the artificial increase in the levels of DnaK and DnaJ, to less than 1.5-fold above wild-type levels, results in decreased level and activity of σ^{32} at intermediate temperature and faster shut off of the heat shock response; the artificial decrease in DnaK and DnaJ levels leads to opposite effects. In contrast, a decrease in FtsH levels increased the σ^{32} levels, but the accumulated σ^{32} was inactive, indicating that sequestration of FtsH alone cannot induce the heat shock response efficiently. The chaperones thus constitute the primary stress-sensing and transducing system of the *E. coli* heat shock response, while the protease may play a more indirect role.

Involvement of the DnaK system in the regulation of the heat shock response has been indicated first by genetic analysis. *dnaK, dnaJ,* and *grpE* mutants show an increased basal level of heat shock gene expression at 30 °C, and a delay in

the adaptation phase of the heat shock response. These alterations result from multiple defects in regulation. First, the DnaK system is crucial for degradation of σ^{32} at all temperatures, as indicated by the 10-fold stabilization of σ^{32} at 30 and 42 °C in the chaperone gene mutants. During the adaptation phase of the heat shock response, the stability of σ^{32} is also increased in *dnaK* (10-fold) and *grpE* (fivefold) mutants, but strikingly, not in *dnaJ* mutants. The weak phenotype of *dnaJ* mutants is explained by the demonstration that another DnaJ homolog existing in the *E. coli* cytosol, CbpA, can substitute for DnaJ in promoting σ^{32} degradation (and inactivation).

The second regulatory defect of *dnaK*, *dnaJ*, and *grpE* mutants is the lack of repression of σ^{32} activity following a temperature downshift (approx. 10-fold repression in wild-type cells). This defect relies on the role for the DnaK system to physically interact with free σ^{32} and to thereby inhibit its activity. DnaJ acts to target DnaK to substrates including σ^{32}, by associating with substrates and DnaK, thereby stimulating the ATP hydrolysis-driven locking in of substrates into the binding pocket of DnaK. The location of the DnaK and DnaJ binding site(s) within σ^{32} are still unclear. GrpE acts as the nucleotide exchange factor for DnaK, allowing for ADP dissociation and rebinding of ATP, which consequently leads to substrate dissociation from DnaK. σ^{32} is thereby subjected to a dynamic chaperone binding-and-release cycle, which prevents it from association with the RNAP enzyme and, consequently, from transcribing heat shock genes. It is tempting to speculate that during this cycle, the conformation of σ^{32} is altered such that it becomes more accessible for proteolytic degradation by FtsH.

Several findings indicate a role for the GroEL chaperone and its cochaperone GroES (GroELS) in the regulation of heat shock gene expression. First, the overproduction of GroELS leads to repression of σ^{32} activity, to an extent similar to the repression of σ^{32} resulting from overproduction of the DnaK system. This inhibition is alleviated by overproduction of GroELS substrates, suggesting that it is the pool of free GroELS that inhibits σ^{32}. Furthermore, the artificial depletion of GroEL results in an increased and prolonged heat shock response, presumably due to increased stabilization of σ^{32}. Finally, *in vitro*, GroEL associates with σ^{32} and inhibits its activity in transcription in synergistic action with the DnaK system.

Together, these findings indicate that modulation of the heat shock response is mediated by a network of cooperating chaperones, comprised of the two major chaperone systems of the cytosol with "foldase" activity, the DnaK and GroEL systems. These chaperones provide the direct link between the sensing of protein folding stress and the expression of heat shock genes, and act in concert with the direct temperature sensing by the *rpoH* mRNA, which does not involve the action of proteinaceous factors.

1.2 THE σ^{E}-dependent Envelope Stress Response

Gram-negative bacteria have a complex cell envelope composed of an inner membrane, an outer membrane, and a peptidoglycan-containing periplasmic space located in between. The periplasmic space constitutes a unique cellular compartment with respect to protein folding. First, it is exposed to strong variations in environmental conditions, in particular,

through the existence of porins in the outer membrane (OMPs), which allow the influx of ions, nutrients, osmolytes, as well as of harmful substances. Second, it is an oxidizing milieu, permitting the formation of disulfide bonds, both within and between proteins. Third, it lacks ATP and, consequently, the classical ATP-dependent chaperones of the cytosol. These features necessitate the existence of a unique set of periplasmic chaperones, folding catalysts and proteases, and a system that allows to adjust this set of folding proteins to the requirements of this compartment.

E. coli has two partially overlapping stress responses reacting to protein misfolding in the envelope, one directed by the alternative σ factor σ^E, and a second response regulated by the CpxAR two-component system. The primary function of the Cpx pathway is to control the biogenesis and structural integrity of surface structures such as bacterial pili, and to detect the attachment of the bacterial cell to surfaces. The primary function of the σ^E pathway is to counteract folding defects of outer-membrane proteins and presumably other proteins of the periplasm. The σ^E pathway is presented here in more detail as a paradigm for the sensing and transmembrane signaling of protein folding stress.

The σ^E-dependent envelope stress response (also referred to as *extracytoplasmic stress response*) was first discovered by C. Gross and coworkers in 1993 by the demonstration that excessive amounts of unfolded OMPs induce activation of the alternative σ^E subunit of RNAP. This sigma factor controls the expression of at least 11 genes, including genes encoding periplasmic folding catalysts, chaperones, and proteases, and also genes involved in biogenesis of the lipopolysaccharide (LPS) constituent lipid A and a number of lipoproteins. Among the first σ^E-controlled genes identified is *htrA* (*degP*), which encodes a periplasmic stress protein that undergoes a temperature-controlled switch from chaperone to protease activity (see Fig. 18 and text). In addition, σ^E mediates the expression of *rpoH* at very high temperatures (45–50 °C), thereby providing a regulatory link between the heat shock response and the cell envelope response. Historically, this role in the transcription of *rpoH* led to the discovery of σ^E by the C. Gross laboratory in 1989, a novel sigma factor without assigned regulon.

The σ^E pathway is induced rapidly upon shift of the cells from 30 to 43 °C, leading to 2- to 2.5-fold increased rates of synthesis of the gene products of the regulon. The rates of synthesis stay at these elevated levels, in contrast to the cytosolic heat shock response that readily enters an adaptation phase during which Hsp synthesis declines. Only when the cells are returned to pre-heat shock temperature do the rates of synthesis of the stress proteins drop rapidly to levels below those observed during steady state growth (e.g. 20-fold decline of DegP synthesis at 15 min after temperature downshift), finally reaching steady state levels.

1.2.1 Inducers, Sensors, and Signals

The σ^E-regulated envelope stress response is induced by several different stress regimen including temperature upshift, hyperosmotic stress, and ethanol treatment. The accumulation of unassembled OMPs (outer-membrane porin proteins) such as OmpF, OmpC, OmpT, and OmpX generates a particularly strong signal. Notably, OMPs only induce the response if they accumulate as unfolded species in the periplasmic space on their route, productive or abortive, to the outer membrane.

They do not generate a signal once they are properly assembled into their native oligomeric states in the outer membrane, or when they fail to be translocated across the inner membrane, remaining cytosolic. The σ^E pathway is induced, for example, by mutational alterations in OmpC, which prevent its insertion into the outer membrane and lead to its periplasmic accumulation. A complementary finding is that the reduced production of OmpC and OmpF leads to decreased expression of the genes controlled by the pathway.

The σ^E pathway primarily monitors assembly defects of outer-membrane proteins. However, misfolding of other proteins is also presumably sensed by this pathway as it is induced in mutants lacking periplasmic folding catalysts and disulfide bond oxidoreductases, in mutants producing a defective LPS, and in mutants producing unfolded variants of the periplasmic maltose binding protein MalE.

1.2.2 The Regulator

σ^E belongs to the large extracytoplasmic functions (ECF) subfamily of sigma factors, which are defined according to their ability to activate the expression of extracytoplasmic functions. σ^E confers to RNAP a promoter specificity – GAACTT, – 35 consensus; TCTGA, – 10 consensus; 16-bp spacer between both promoter regions – that is unique for promoters driving the expression of stress envelope genes. The *rpoE* gene encoding σ^E is organized in an operon together with the *rseA, rseB*, and *rseC* genes, at least two of which (*rseA, rseB*) are involved in the regulation of σ^E activity (see the following). The expression of *rpoE* and the other genes of this operon is autoregulated; a temperature upshift causes a slight (1.2- to 1.5-fold) increase in transcription of the *rpoE* gene.

It appeared at first that σ^E is nonessential for growth at temperatures below 40 °C. However, recent work established that σ^E is an essential sigma factor at all growth temperatures, but that genetically unlinked suppressor mutations arise with high frequency, which allow cell survival at lower temperatures. The σ^E regulon therefore provides essential functions related to protein folding and turnover in the cell envelope.

1.2.2.1 **Signal transduction across the membrane – control of σ^Eactivity by RseA, RseB, and RseC.** The sensing of misfolded proteins accumulating in the periplasm and outer membrane of gram-negative bacteria requires the transduction of these signals to the cytosol (Fig. 3). The key protein that relays the stress signals across the inner membrane and ultimately leads to activation of periplasmic chaperone and protease genes is RseA, a protein located in the inner membrane that acts as antisigma factor for σ^E. RseA is composed of a single membrane spanning segment, an N-terminal cytoplasmic domain, and a C-terminal periplasmic domain. With its cytoplasmic domain, RseA associates with σ^E thereby sequestering it away from RNAP. The determination of the atomic structure of the inhibitory complex between σ^E and the cytoplasmic domain of RseA elucidated the molecular basis for the sequestration. RseA is sandwiched between the two major domains of σ^E and acts by sterically occluding the two sites in σ^E, which are most important for association with RNAP. With its periplasmic domain, RseA binds the periplasmic protein RseB, a minor negative regulator of σ^E. Intriguingly, all three components of this protein complex (σ^E, RseA, RseB) are coregulated and induced during the stress

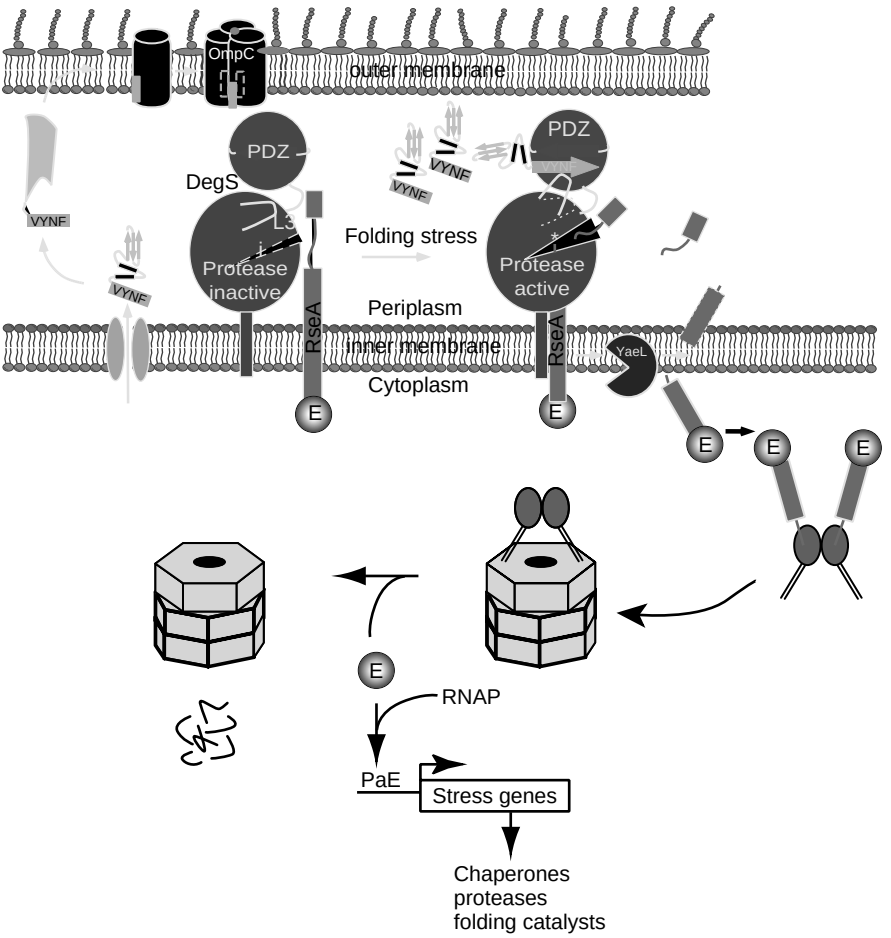

Fig. 3 σ^E-regulated envelope stress response. At noninducing conditions, the periplasmic DegS protease exists in an inactive conformation. The overproduction or mutational alteration of trimeric outer-membrane porins (Omp) lead to accumulation of unassembled folding intermediates, which expose usually buried C-terminal motifs (in red). The motif residues bind to the PDZ domain of DegS by formation of an additional ß-strand (red arrow), leading to a flipping of the flexible loop L3 toward the PDZ-protease domain interface. L3 repositioning activates the catalytic center (from i to *), making it accessible for a short unstructured segment of the DegS substrate, the anti-σ^E factor RseA. Initial cleavage of RseA by activated DegS initiates a subsequent cleavage of RseA by YaeL, which releases a fragment of RseA in complex with σ^E into the cytoplasm. Subsequently the SspB adaptor of the ClpXP protease binds to the RseA fragment and targets it to degradation, ultimately releasing σ^E, which then drives expression of genes encoding periplasmic chaperones and proteases. For clarity only one subunit of the DegS trimer is depicted and the RseB modulator is omitted. This figure has been modified from Schlieker, C., Mogk, A., Bukau, B. (2004) *Cell* **117**, 417–420. (See color plate p. xxxvi.)

response. This upregulation provides an autoregulatory loop, which ensures that the response can be rapidly shutoff after damage repair.

1.2.2.2 **A proteolytic cascade derepresses σ^E.** The affinity of RseA for σ^E is very high. The K_d of the complex formed between the cytosolic fragment of RseA and σ^E is below 1 nM. This high affinity implies that the activation of the cell envelope response strictly requires the active dissociation of σ^E from RseA. The dissociation of RseA from σ^E is achieved by a proteolytic cascade involving three proteases, whose action ultimately leads to liberation of σ^E and activation of the σ^E pathway. Indeed, during all stages of the envelope stress response induced by heat shock, the rate of RseA degradation correlates with the activity of σ^E, suggesting that the degradation of RseA is the key step in the regulation of this response. The RseA half-life is 8 min at nonstress conditions, it decreases to 2 min during the induction of the stress response, and increases to >50 min during temperature downshift.

The proteolytic cascade is triggered by the accumulation of misfolded porins in the periplasm (Fig. 3). It is initiated by activation of DegS, a member of the family of HtrA (DegP) proteases. (This is an interesting example of evolutionary diversification, since the regulatory DegS protease controls expression of the homologous DegP protease of the periplasm, which, in contrast to DegS, has a broad substrate range and lacks regulatory function.) Activated DegS initially cleaves RseA at its periplasmic site, which then allows a second cleavage by the membrane-embedded metalloprotease YaeL (also termed *RseP*) at the cytosolic site of RseA. The modulator of RseA stability, RseB, may act to allow the protease YaeL to cleave RseA more efficiently. YaeL is an ortholog of the mammalian Site-2 protease, which is involved in the proteolytic regulation of the unfolded protein response (UPR) in the endoplasmic reticulum (ER) and of the expression of cholesterol biosynthesis genes.

The successive cleavage of RseA by DegS and YaeL releases the σ^E–RseA fragment complex from its association with the inner membrane. Then, a final proteolytic cleavage occurs in the cytosol, mediated by the ClpXP member of the AAA+ protease family in conjunction with the adaptor protein SspB. The adaptor SspB binds to a C-terminal stretch of residues within the RseA fragment and, through additional binding to the N-terminal domain of ClpX, brings the σ^E–RseA fragment complex to the protease. Importantly, only the RseA fragment is cleaved by ClpXP while σ^E remains intact and becomes liberated to associate with the core RNAP, enabling transcription of σ^E-controlled stress genes.

1.2.2.3 **Recognition of the stress signal by DegS.** The activation of DegS is the first decisive step in the proteolytic cascade. This step is tightly controlled by the association of unassembled OMPs via a conserved sequence motif (YXF) located at their C termini. Crystal structures of different OMPs show that in their fully assembled state, this motif is buried and therefore cannot bind to DegS. This strongly suggests that the exposure of the YXF motif within the OMPs signals folding stress. Strong support for this hypothesis comes from the finding that short peptides carrying the YXF motif bind to the PDZ domain of DegS, a widespread structural motif involved in protein interactions, and stimulate RseA degradation *in vitro*.

How might binding of the YXF peptide to the PDZ domain control the activity of the protease domain of DegS? The atomic structure of DegS revealed that it is a homotrimer, where the protease domains form the core and the PDZ domains protrude outwards, creating a funnel-like arrangement (see Fig. 18 for homologous DegP structure). In this structure, the serine protease catalytic site of DegS is in an inactive state. An active conformation results from the addition of an activating peptide resembling the C terminus of OmpC. The peptide becomes bound in a pocket of the PDZ domain close to the interface with the protease domain. The PDZ domain provides a first docking site for the activating signal peptide. Association of the activator peptide results in several important conformational changes in the protease domain. A flexible loop (L3), protruding from the protease domain is displaced by 15 Å from its previous location in the inactive form, and is stabilized by a contact to the OmpC peptide. Upon peptide association, several residues in the active site involved in substrate binding and catalysis, previously blocked or distorted in inactive DegS, become accessible for substrate binding and correctly aligned for catalysis. In addition, the activator positioning allows a flexible loop to transduce the structural information induced by ligand binding to the catalytic center, resulting in highly specific cleavage of the downstream target, RseA.

The activation of DegS through ligand binding to the PDZ domain is reversible, making it well suited to trigger a transient response that can be turned off upon alleviation of the stress signal. The reversible activation of DegS allows a flexible and fast adaptation to the stress signal amplitude, a prerequisite for a fine-tuned cellular response.

2
Eukaryotic Heat Shock Response – Heat Shock Factors

In multicellular eukaryotes, the stress-induced regulation of the heat shock response occurs, as with prokaryotes, principally by activation of a family of heat shock transcription factors (HSFs), resulting in elevated expression of the same family of heat shock effector proteins that are involved in protein homeostasis in bacterial systems (Fig. 4). The complexity of HSF-dependent regulation varies among species, with yeast, *Caenorhabditis elegans* and Drosophila expressing a single HSF gene, vertebrates expressing four HSF genes (HSF1 to 4), and plants expressing more than a dozen HSF genes.

Comparison of domain organization and sequence identity among HSF genes reveals that the overall structure is highly conserved among eukaryotes (Fig. 5a). Located at the amino terminus of vertebrate HSFs is a helix–loop–helix DNA binding domain, followed by an 80 amino acid leucine zipper heptad repeat (HR-A/B) necessary for homotrimer formation, a centrally located negative regulatory domain that serves to maintain HSFs in an inert monomeric state, another leucine zipper heptad repeat, HR-C, and a C-terminal transactivation domain (Fig. 5a). HR-C is thought to interact with HR-A/B in the inert state and becomes dissociated from it upon heat shock, allowing the formation of homotrimers. Notably, HR-C is absent from budding yeast HSF and human HSF4 and correspondingly, the HSF of budding yeast exhibits constitutive trimerization and DNA binding, unlike other HSFs, and for at least some genes it is not negatively regulated as in the other eukaryotes, apparently supporting basal transcription.

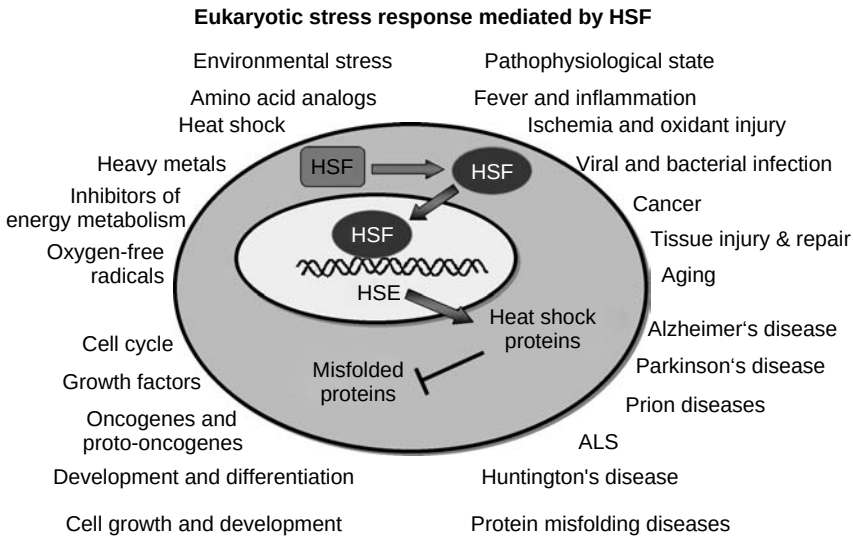

Fig. 4 Stress conditions that activate the heat shock response. Heat shock gene expression is induced by environmental and physiological stress (top left and right, respectively), by nonstressful conditions including cell growth and development (lower left), by pathophysiological states (upper right), and by protein aggregation diseases associated with the appearance of misfolded proteins (lower right). Activation of the heat shock transcription factor (HSF) leads to the transcriptional activation of heat shock genes, and the induced heat shock proteins function to prevent and repair protein damage through interaction with misfolded proteins.

HSF1 cycles between the cytoplasm and the nucleus of human cells prior to heat shock (Fig. 5b, top), and upon stress induction (Fig. 5b, right), its localization becomes nuclear and it binds to promoter elements, so-called heat shock elements (HSEs) upstream of heat-inducible genes, containing typically three contiguous inverted repeats (e.g. nTTCnnGAAnnTTCn), activating transcription (Fig. 5b, lower panels). In its nuclear location, it also rapidly forms distinctive heat shock "stress granules" that are associated with satellite repeats at chromosome 9q12, which become actively transcribed, the product message remaining associated with this locus. The nature of such heat shock transcriptional activation remains unknown.

Human HSF1 is constitutively phosphorylated in the negative regulatory domain at multiple serine residues (303, 307, and 363; Fig. 5a), but upon heat shock becomes hyperphosphorylated at these and other serine residues, for example, Ser230, associated with transcriptional activation. It is likely that additional stress-inducible phosphorylation sites will have regulatory roles for HSF1 function as suggested by mathematical modeling and sensitivity analysis of HSF1 in the human heat shock response. In addition to phosphorylation, HSF1 is also sumoylated on Lys298 in a manner that requires phosphorylation on serines 303 and 307. The role of sumoylation in HSF1 activity has not been fully established and may not be required for transcriptional activity.

Of the HSFs expressed in vertebrates, HSF1 exhibits properties that are most closely related to yeast and Drosophila

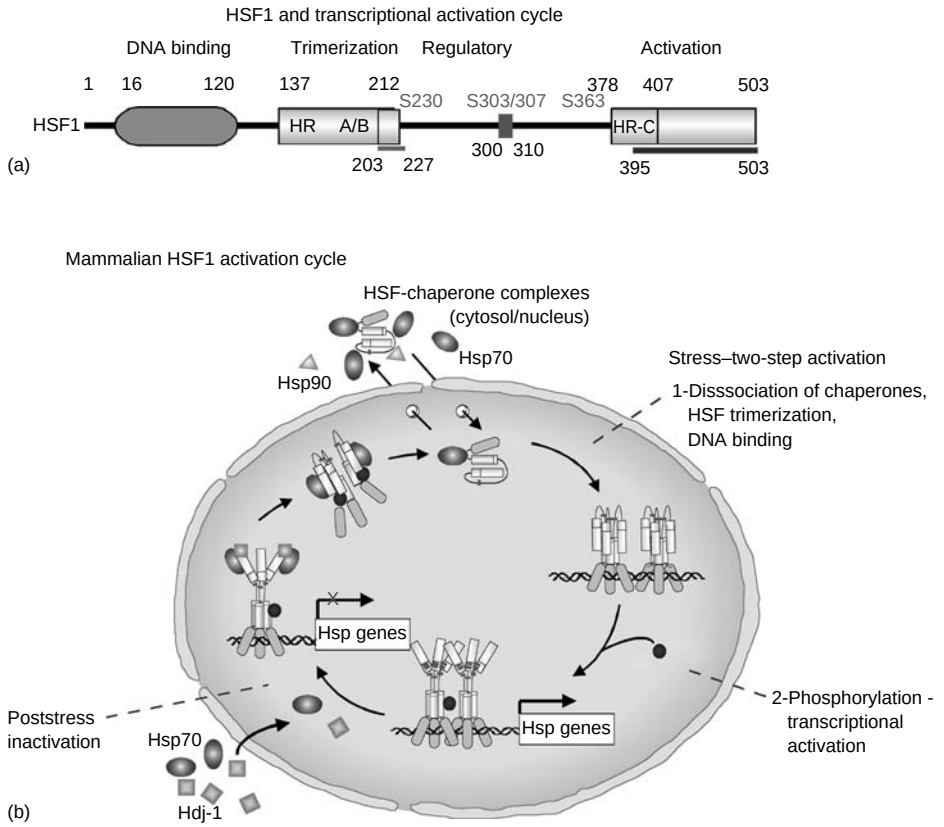

Fig. 5 Regulation of the heat shock response by HSF. (a) Domain representation of human HSF1 corresponding to the amino-terminal DNA binding domain, the heptad repeat-containing HR-A/B, the negative regulatory domain indicating sites of serine phosphorylation, the carboxyl-terminal heptad repeat, HR-C, and the transactivation domain. (b) Activation of heat shock factor (HSF1) is linked to the appearance of nonnative proteins and the requirement for molecular chaperones (e.g. hsp90, hsp70, and Hdj1) to prevent the appearance of misfolded proteins. Mammalian, Drosophila, and plant HSF1 exists in the basal state as an inert monomer (shown as intramolecularly negatively regulated for DNA binding and transcriptional activity via HR AB/C interaction). It undergoes stepwise activation, first to a DNA binding competent homotrimeric state which is transcriptionally inert, then, upon acquisition of inducible phosphorylation (green dot), to a state of activation and inducible transcription of heat shock genes. HSF1 activity is subsequently downregulated by the chaperones hsp70, hsp90, and Hdj1 which bind to HSF1 and repress its activity, and by a heat shock factor binding protein (HSBP) which binds to the region of HSF1 corresponding to the heptad repeat (not shown). (See color plate p. xxxvii.)

HSF, becoming activated as a transcription factor by various forms of stress. Notably, yeast HSF is essential for cell growth and viability, indicating that it has roles in supporting basal transcription as well as its activating role under stress. Drosophila HSF, by contrast, is dispensable to growth and viability of adult flies under nonstress conditions, but it is required during development. In mouse, deletion of HSF1

leads to placental defects, with significant prenatal mortality. Yet viable deleted animals can be obtained, but they exhibit growth retardation, female sterility, and, as expected, inability to mount a heat shock response with attendant lack of thermotolerance. Increased mortality in the face of endotoxic and inflammatory stimuli is also observed.

Recent studies in yeast employing microarray and CHIP analyses, as well as ts versions of HSF, have sought to identify the range of promoters bound by HSF and genes activated by it under heat shock conditions. More than 100 genes were both bound in their promoter regions by HSF and transcriptionally activated upon exposure to heat by at least 1.5-fold. Included were chaperones from all cellular compartments and also several components involved in ubiquitination and proteolysis, a number of components in energy and carbohydrate metabolism, as well as several to do with cell wall and cytoskeleton metabolism, and small-molecule transport. An even broader set of components was identified in microarray studies examining the entire spectrum of transcripts induced by a variety of environmental stresses, beyond regulation by HSF1, with many hundreds of transcripts upregulated. Thus, there appears to be a set of non-HSF-regulated components that can also participate in stress responses in eukaryotes. Adding to the complexity, it seems clear that there are also genes activated by HSF1, independent of its binding to classical HSE elements. A number of non-canonical elements have been identified, but for other activated transcripts there is no recognizable element.

Studies have begun to address the roles of individual HSFs where more than one is present. In avian cells, for example, the heat shock response involves two stress-activated factors, HSF1 and HSF3, which exhibit functional codependence; HSF3 null cells that express HSF1 are deficient for the heat shock response. HSF3 also interacts with other transcription factors including the oncogene Myb, via direct protein–protein interaction, through the HSF3 DNA binding domain. Myb is a growth regulated transcription factor – consequently, the interaction between Myb and HSF3 suggests genetic crosstalk between cell growth and the stress response.

One of the most provocative members of the HSF family is HSF2, originally shown to be activated in human erythroleukemia (K562) cells exposed to hemin, and subsequently detected during murine spermatocyte differentiation and embryogenesis. The stress-sensing pathway for HSF2 activation is associated with changes in the protein degradation machinery, with HSF2 DNA binding activity induced either by chemical inhibition of the ubiquitin-dependent proteasome pathway or by conditional mutations affecting the proteasome. The relationship between HSF2 activity and the protein degradative machinery reveals a requirement for Hsps, perhaps to prevent misfolding and aggregation of nonnative proteins targeted for degradation. HSF2 is also modified by sumoylation, and recently HSF2 has been shown to have a role in cell cycle control via binding to a subunit of condensin, influencing DNA compaction.

2.1 Activation of Heat Shock Factor 1

Stress-induced activation of HSF1 involves a multistep process (Fig. 5b) involving its relocalization within the nucleus, transition from an inert monomer to a trimeric DNA binding species, binding to heat

shock promoter elements, hyperphosphorylation at serine residues, and the rapid elevation in transcription of heat shock genes. Activation of HSF1 DNA binding activity is induced within minutes and attains maximal DNA binding and transcriptional activity by 20 to 30 min of continued heat shock. Thereafter, both activities attenuate to control levels within 60 to 120 min. The attenuation phase is autoregulated and requires the synthesis of high levels of chaperones that accumulate initially in the cytoplasm and relocalize to the nucleus and bind to the active trimeric state of HSF1 (Fig. 5b). The appearance of complexes between HSF1 and molecular chaperones correlates with the dissociation of HSF1 trimers from DNA and refolding of the trimer to the inert monomer. Consequently, maintenance of HSF1 in a repressed state is delicately balanced and easily disrupted. For example, transient overexpression of HSF1 is sufficient to override the negative regulators resulting in constitutive HSF1 activity. Likewise, overexpression of chaperones suppresses HSF1 activity. Specific sequences in the central negative regulatory domain of HSF1 maintain HSF1 in a repressed state; mutations cause derepression and constitutive HSF1 activity. Acquisition of HSF1 DNA binding activity alone, however, is insufficient to achieve transcription of target heat shock genes as demonstrated by certain small-molecule regulators of HSF1, such as the anti-inflammatory drug sodium salicylate, which induces DNA binding competent trimers that lack some of the posttranslational modifications associated with transcriptional activity.

The initial observations to suggest that the heat shock response is autoregulated arose from the use of amino acid analogs as inducers of the heat shock response. Treatment of Drosophila tissue culture cells with azetidine, a proline analog, led to chronic activation of heat shock gene expression. Incorporation of azetidine, or other amino acid analogs, into nascent polypeptides results in the chronic expression of misfolded proteins, led to activation of HSF1 and induction of heat shock genes. However, unlike the heat shock response itself, where attenuation is linked to the *de novo* synthesis of Hsps, the Hsp70 synthesized in amino acid analog-treated cells is likewise misfolded and nonfunctional. Consequently, in the presence of amino acid analogs, the heat shock response is chronically upregulated and does not attenuate. These observations support the idea that Hsps themselves are involved in autoregulation of the heat shock response.

Genetic evidence further supports Hsp-mediated autoregulation of the heat shock response. Overexpression of the yeast Ssa1p (cytosolic Hsp70) dampens the heat shock response, and deletion of the yeast Hsp70 genes, ssa1 and ssa2, results in an unusually high level of expression of another Hsp70 gene (*Ssa3p*) in an HSF-dependent manner. In addition, HSF itself was isolated as an extragenic suppressor of the temperature-sensitive phenotype of an Hsp70 mutant (ssa1ssa2 strain).

Biochemical evidence also supports that molecular chaperones autoregulate the heat shock response. HSF1 trimers associate with Hsp70, Hdj1 (Hsp40), and Hsp90 in ATP-sensitive complexes that can be reconstituted *in vitro*. Conditional overexpression of Hsp70 and Hdj1 results in repression of heat shock gene transcription with little or no effect on HSF1 trimer formation or DNA binding activity. Similar observations have been made for members of the DnaJ family in regulation of the heat shock response. For example, the *Saccharomyces cerevisiae*

DnaJ homolog, SIS1, negatively regulates its own expression in an HSE-dependent manner, and Drosophila Ddj1 associates with HSF1. Other chaperones including Hsp90 have been shown to interact with HSF1 or the HSF1 complex. Collectively, these results reveal that multiple chaperones are involved in the regulation of HSF1 and, moreover, that differences in specificity and kinetics may influence different aspects of HSF regulation and the heat shock response.

2.2 Pharmacological Regulation of Eukaryotic Heat Shock Response

Among the small molecules with heat shock regulatory properties are non-steroidal anti-inflammatory drugs (NSAIDs), cyclopentenone prostaglandins, serine protease inhibitors, proteasome inhibitors, the triterpene celastrol, and ansamycin benzoquinones. Salicylates and other NSAIDs have been shown to result in activation of HSF1. Moreover, treatment with sodium salicylate or indomethacin (at subthreshold concentrations) decreases the temperature threshold of the heat shock response. Exposure to salicylate or indomethacin at concentrations comparable to clinically achieved levels results in the priming of human cells for subsequent exposure to heat shock and other stresses by the enhanced transcription of heat shock genes and cytoprotection from stress-induced cell death.

Many molecules that regulate inflammation also activate HSF1, suggesting a link between inflammatory pathways and the heat shock response. The inflammatory response is induced by a cascade involving various signaling molecules including phospholipases, arachidonic acid, and prostaglandins. Some of these molecules have also been shown to induce HSF1 activity and Hsp expression. Of the prostaglandins, only the cyclopentenone prostaglandins, including PGA1, PGA2, and PGJ2, are functional in HSF1 induction. This class of prostaglandin is produced at a late stage in the inflammatory reaction, and is thought to be involved in the resolution of the inflammatory process, so cyclopentenone prostaglandins may actually be considered to be anti-inflammatory mediators.

Inhibition of protein degradation by small molecules that block proteasome or serine protease activities results in elevated levels of modified and misfolded proteins, leading to the chronic activation of a heat shock response. For example, exposure of eukaryotic cells to the proteasome inhibitors, MG132 and lactacystin, activate both HSF1 and HSF2. Likewise, the serine protease inhibitors DCIC, TPCK, and TLCK also induce these factors and elevate transcription of Hsp genes.

The aberrant expression of Hsps observed in many cancers has led to various hypotheses that the chronic upregulation of chaperones could afford the cancer cell with the capacity to suppress multiple independent mutations that affect protein function. Equally compelling are arguments that the transient activation of the heat shock response and elevated levels of chaperones establish a cytoprotective state that is resistant to cell death signals and with the capacity to endure stressful conditions. For these reasons, there has been much interest in the therapeutics of the heat shock response and chaperones. The discovery that inhibitors of Hsp90 function could block cell proliferative pathways important in cancer has led to the development of a family of structurally related benzoquinone ansamycins including geldanamycin and its less toxic derivative

17-AAG and the fungal antibiotic radicicol. Although structurally distinct, radicicol and geldanamycin both bind to the ATP binding domain of Hsp90 and inactivate Hsp90 function, resulting in decreased activity and levels of essential key signal transduction proteins including steroid receptors, cell cycle kinases, transcription factors, and p53, with immediate consequences on cell proliferation. Indeed, a recent report suggested that Hsp90 in tumor cells may have selectively greater affinity for 17-AAG, associated with presence of Hsp90 in such cells in multichaperone complexes as opposed to a predominantly noncomplexed state in normal cells. Regardless, a consequence of treatment with these Hsp90 inhibitors is that they produce constitutive activation of the heat shock response, presumably by derepression of HSF1.

A speculation and systems considerations

It is intriguing to consider that among the Hsp90 targets are the hormone regulated intracellular receptors that share many features in common with the regulation of the heat shock response. Both provide inducible responses regulated at the level of transcriptional control, and involve families of transcription factors that cycle between inert and activated states by adopting multiple conformations and rapid reversible kinetics of activation. Chaperone networks are employed in both systems to maintain the activators in an inert but responsive state, and later in the regulatory cycle, to repress and disassemble the activated factor. Perhaps transcription factors such as hormone receptors and HSF1, and other signaling molecules including kinases and caspases, have complex folding requirements that

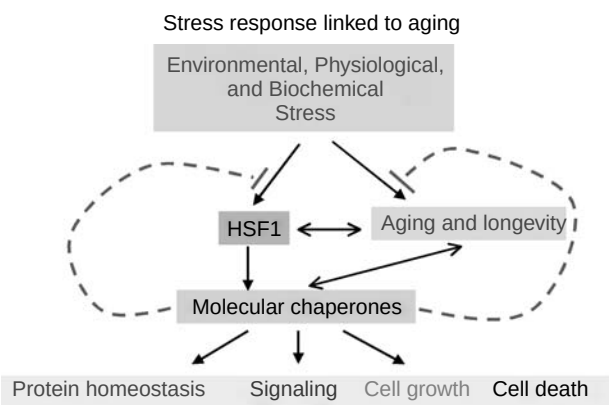

Fig. 6 Diverse roles of HSF1 and the heat shock response, including an effect on aging observed in *C. elegans*. In the presence of proteotoxic stress inducers, HSF1 is activated, leading to the elevated expression of molecular chaperones and other proteins that regulate protein homeostasis. These events influence stress signaling pathways, cell growth, and cell death. Recent results also indicate that HSF1 and certain molecular chaperones influence aging. See for example, the report of Hsu et al, 2003, where increased HSF-1 activity extended *C. elegans* lifespan and reduced HSF-1 activity shortened it. Such effects were linked to the DAF-16 insulin/IGF1 pathway and to small heat shock protein expression.

may have arisen from exon shuffling and other evolutionary pressures, for which chaperones have become essential to regulate alternate conformational states. Or did this arise because of the unique properties of chaperones and cochaperones to generate stable folded intermediates that have become prime targets in the design of regulatory complexes?

From a systems perspective, the linkage of signaling pathways that feed back through a shared pool of chaperones posits the presence of hierarchical genetic networks through which changes in the cellular environment can be transduced (Fig. 6). In support of this speculation, aberrant expression of individual chaperones or cochaperones influences proliferation and development by preventing or enhancing cell growth and cell death. For example, decreasing the levels of Hsp90, whether by mutation or by the Hsp90 inhibitor geldanamycin, uncovers a Pandora's box of developmental abnormalities in Drosophila and Arabidopsis. Likewise, heat shock and elevated levels of Hsp70 have growth inhibitory effects on signaling pathways.

3
Unfolded Protein Response (UPR) of the Eukaryotic Endoplasmic Reticulum

A further stress response system has been recognized in the secretory endoplasmic reticulum of eukaryotes, signaling the presence of unfolded proteins and producing transcriptional and translational responses. In some respects the UPR parallels the σ^E-mediated stress response of the bacterial periplasm, an analogous secretory compartment, where, likewise, signals must transmit across a membrane to reach the site of chromosomal transcription. The eukaryotic UPR induces the transcription of several hundred genes whose products increase both the folding capacity of the ER and also its ability to rid itself of misfolded proteins by retrotranslocation to the cytosol followed by proteasomal degradation. A further arm of the UPR, however, involves attenuation of general protein translation, which decreases the "load" of unfolded proteins arriving in the ER.

The UPR system is regulated by three ER transmembrane proteins, IRE1, ATF6, and PERK (Fig. 7a). These components appear to bind the Hsp70 chaperone protein in the ER, known as BiP, under normal conditions, but in the setting of unfolded proteins, BiP is recruited to these substrates, freeing these transmembrane components.

The first signaling mechanism to be understood is that mediated by IRE1, articulated from elegant experiments in a yeast system by Peter Walter and his colleagues in the mid-1990s. IRE1 responds to accumulation of unfolded proteins by homodimerization and activation of both a kinase activity and a site-specific endonuclease activity contained in a domain on the cytosolic side of the ER membrane, which enables splicing and efficient translation of the message for a transcriptional activator (HAC1 in yeast; Xbp1 in mammals; Fig. 7a). This activator binds and activates specific upstream sequence elements in many UPR-inducible genes. ATF6 (activating transcription factor 6) can also bind regulatory elements. This component is synthesized as a larger ER transmembrane precursor protein that is retained in the ER until unfolded proteins trigger its translocation to the Golgi, where it is cleaved in its transmembrane region by

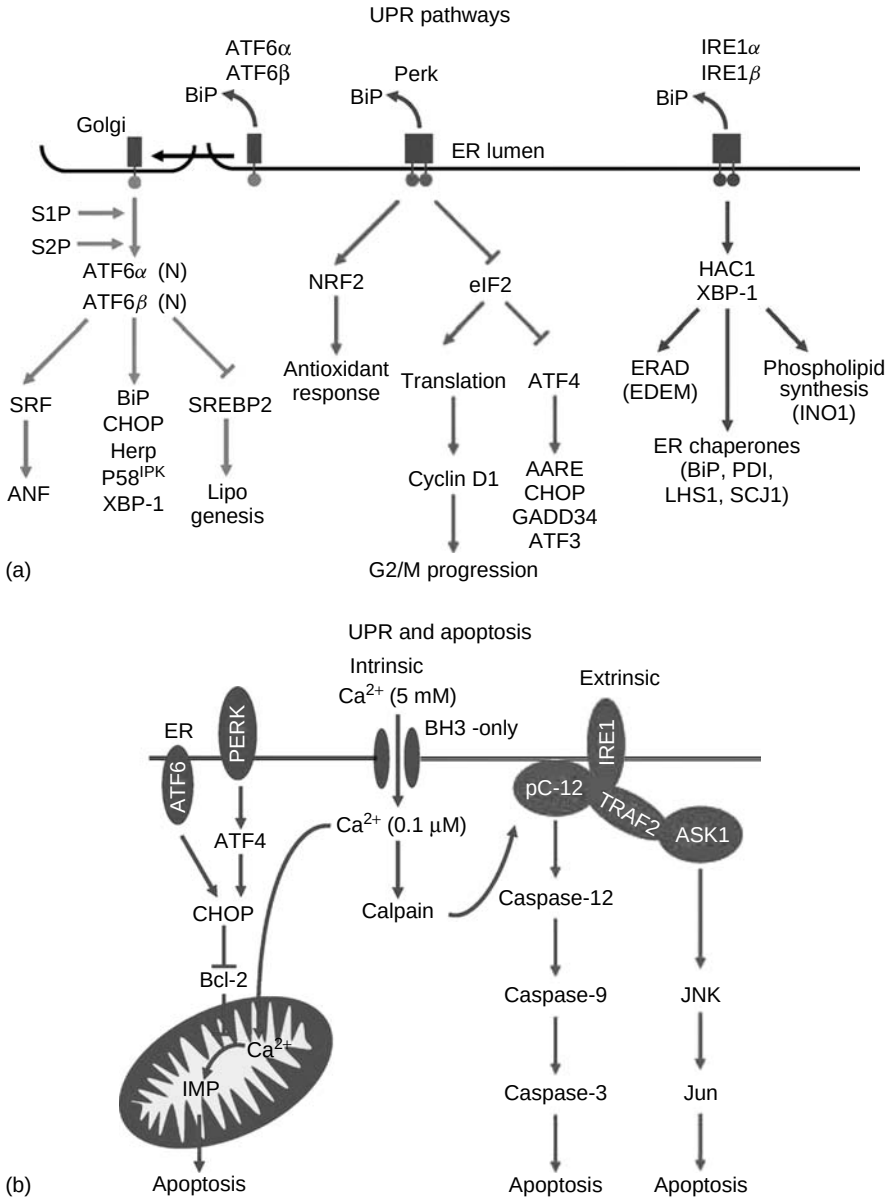

Fig. 7 Pathways of the unfolded protein response (UPR) by which the eukaryotic endoplasmic reticulum recognizes accumulation of unfolded proteins and signals to the nucleus and cytosol. (a) IRE1 and ATF6 trigger induction of transcription of components that enhance the folding and secretory capacity of the ER, while PERK inhibits general translation in the cytosol, diminishing the load of unfolded protein reaching the ER. (b) Prolonged activation of the UPR can lead to apoptotic cell death via the same proximal pathways, involving IRE1, ATF6, and PERK, but with distal components such as ATF4, CHOP, and JNK activating apoptosis. Figure taken from Schröder, M., Kaufman, R.J. (2005) ER stress and the unfolded protein response, *Mutat. Res.* **569**, 29–63.

two proteases known as S1P and S2P, which are also involved in the processing of the sterol response factor SREBP. Cleavage releases a cytosolic fragment of ATF6 that enters the nucleus and binds to ER stress response DNA elements including that for the IRE1 splicing target, Xbp1. Thus, ATF6 and IRE1 cooperate to activate the stress response. In a third arm of the UPR, the transmembrane kinase, PERK, first identified by David Ron and his colleagues, becomes activated, phosphorylating eIF2α and attenuating general translation, while at the same time enhancing translation of UPR-induced messages.

At the effector level, the UPR induces a host of components. Included are ER lumenal components involved in protein folding, such as the chaperones BiP (an Hsp70) and the DnaJ protein SCJ1, but also induced are components involved in disulfide bond formation, such as Ero1 and PDI (protein disulfide isomerase). Also induced are a host of glycosylation enzymes. At the level of the ER membrane, components of the translocon as well as signal peptidase are induced, and also induced is the class of components involved with ERAD (ER-associated degradation), including derlin, a putative protein retrotranslocase; Ubc7, a ubiquitinating component, and several components of the proteasome pathway. Additionally induced are components of vesicle trafficking and of lipid metabolism, reflecting a general induction of secretory functions by the UPR.

Notably, prolonged UPR activation can lead to apoptotic cell death, involving several pathways (Fig. 7b). One involves IRE and a Jun inhibitory kinase, and another involves ATF6 and activation of the bZIP transcription factor CHOP.

The UPR is implicated in a range of normal physiological states such as B cell maturation, where the ER must massively expand its capacity to import, fold, assemble, and export antibodies. More generally, the UPR appears to balance the translational burden relative to the maturational capacity of the ER. This may reflect the cell's overall metabolic state. Thus, as has been suggested, the UPR may play a contributing role in disease states. For example, obese mice were shown to exhibit increased levels of PERK and eIF2α phosphorylation, as well as elevation of the ER stress-responsive protein GRP78, in liver and fat. This ER stress-responsive state was associated with diminished insulin receptor signaling, most immediately mediated by Jun kinase-dependent phosphorylation of IRS-1 (insulin receptor substrate-1). This was likewise observed *in vivo*, in Xbp-1 +/− mice, lacking compensatory capacity for ER stress – these mice developed peripheral insulin resistance and type II diabetes. Consistently, and independently, mutations in PERK also lead to diabetes in mice and man. Thus, a link between chronic ER stress and peripheral insulin resistance seems to be indicated. The nature of how obesity translates into ER stress remains unclear, but it seems inescapable that stress response is integrated with metabolic regulation in the ER.

4 Chaperone Effectors of the Heat Shock Response

4.1 Chaperonin (Hsp60) System

A major sigma32-responsive heat shock operon in bacteria is that encoding the double ring chaperonin, GroEL, and its cooperating single ring lid structure,

GroES. These two ring structures mediate ATP-dependent polypeptide chain folding, with nonnative protein binding to the hydrophobic wall of the central cavity of an open GroEL ring via exposed hydrophobic surfaces, followed by folding to the native state occurring inside the encapsulated and hydrophilic chamber formed upon binding of the GroES lid and ATP to the same GroEL ring. Chaperonins resembling GroEL/GroES are also found in mitochondria (Hsp60/Hsp10) and chloroplasts (Rubisco binding protein/cpn10). A second family of chaperonins, functioning with a built-in lid structure, is found in archaebacteria (thermosome) and in the eukaryotic cytosol (CCT or TRiC). In all of these contexts, the chaperonins are essential for cell growth, indicating that the kinetic assistance they provide to polypeptide chain folding is essential under all conditions.

More explicitly, the role of these assemblies, like other chaperones, is to prevent or reverse off-pathway steps of folding that can occur under *in vivo* conditions. While Anfinsen and his coworkers showed in the late 1950s that the primary amino acid sequence of a polypeptide chain contains all of the information required to direct folding to the native state typically lying at the energetic minimum, under conditions in the cell, including high temperature and high solute concentration, polypeptide chains, particularly ones larger than 150 amino acids and with multiple domains, often cannot efficiently reach their native state. Under such conditions, in seeking to reach the energetic minimum, they face a more rugged folding "landscape," with the possibility of landing in local energetic minima, kinetically trapped states. In structural terms, such states, usually misfolded but in some cases on-pathway intermediates, expose hydrophobic surfaces that will become buried in the native state. These can recruit one molecule to another, leading to irreversible multimolecular aggregation, a functional dead end for the proteins involved, and a state that can be damaging to cells, as for example, in the case of the amyloid diseases. The action of molecular chaperones in general is to bind such states incipiently, via the exposed hydrophobic surfaces, masking such surfaces and preventing aggregation. Chaperones then release the bound protein, often through the action of binding of ATP, allowing the substrate to attempt to pursue a correct pathway of folding or biogenesis. The chaperonin class of molecular chaperones acts uniquely to support folding to native form by releasing bound proteins into an encapsulated hydrophilic cavity, enforcing folding in isolation where aggregation cannot occur. The hydrophilic environment likely contributes to productive folding as it energetically favors burial of hydrophobic surfaces to the interior of the folding protein and exposure of hydrophilic surfaces to the outside, properties of the native state.

History and physiology GroEL was originally uncovered in the early 1970s by Georgopoulos and coworkers and independently by investigators in Japan as a host gene required for phage assembly. Cells bearing mutations in this operon could survive infection with λ or T4 phage, and when lysates were examined by EM, they contained aggregated head or tail structures. GroEL mutant cells, in the absence of phage infection, were also observed to be defective in growth, particularly at high temperature. A role in macromolecular assembly was originally suggested, supported by experiments by Ellis and coworkers in 1980 with a chloroplast homolog that suggested a role for it in

assembly of the abundant hexadecameric CO_2-fixing enzyme Rubisco. Ultimately, however, a role in monomeric polypeptide chain folding was elucidated by studies of an yeast mutant affecting an additional homolog in mitochondria, Hsp60, which was observed to be required for the folding of imported proteins that pass through the mitochondrial membranes as unfolded monomers. A role of these double ring "chaperonins" in polypeptide chain folding was confirmed by reconstitution experiments, first with isolated mitochondria, then *in vitro* with the purified bacterial chaperonin, GroEL and its cochaperonin GroES, demonstrating assisted folding of imported or chemically denatured polypeptides, respectively, in the presence of ATP.

GroEL–GroES-mediated protein folding *in vitro* was first accomplished by Lorimer and coworkers as a two-step reaction. In the first step, a nonnative protein unfolded in denaturant was diluted into a buffer containing the GroEL double ring, and formed a stoichiometric complex with it. Such a binary complex formation protected the substrate protein from complete and irreversible aggregation that supervened in the absence of chaperonin. Later EM studies showed that in such complexes the bound substrate protein localizes inside the central cavity of a chaperonin ring. In the second step, upon addition of ATP and the GroES single ring cochaperonin, production of the native state of the bound protein occurred, associated with its release from GroEL, over a period of several minutes. The mechanism of GroEL/GroES-mediated folding is further detailed below.

The second family of chaperonins was identified via studies of a temperature-induced protein in thermophilic archaebacteria that had the same size as the subunit of the Hsp60 family under both native and denaturing conditions. When its primary sequence was determined, it was found to be homologous to an essential protein of the eukaryotic cytosol called TCP1. Cold-sensitive conditional mutation of tcp1 in yeast led to impaired tubulin biogenesis, but the role of this protein had remained unclear. This protein was recognized to be a subunit of the TCP1 or CCT (chaperonin containing TCP1) complex. This complex, composed of 8 related but nonidentical subunits per ring, was found to play an essential role in mediating folding of actin and tubulin but also, more recently, of G_α transducin and β-propeller proteins.

The chaperonins are abundant, reflecting their involvement in assisting the folding of many proteins. In bacteria, for example, their abundance amounts to ~1% of soluble protein under normal conditions, and they mainly provide assistance to folding of newly translated polypeptides. A number of studies estimate that ~10 to 20% of newly made species are assisted by this system in reaching the native state. Under stress conditions such as heat shock, however, the synthesis of the bacterial components rises to a level of 10 to 15% or more of total soluble protein, an amount of these ring complexes equivalent to or greater than the number of ribosomes. Under these conditions, the chaperonin system acts particularly to protect preexistent proteins from irreversible misfolding and aggregation, and there may be a large percentage of proteins requiring such action.

In mitochondria, a significant fraction of the imported polypeptides, approximating 50% from a proteomic study, require the action of Hsp60 and its cooperating component Hsp10 in order to be properly folded. In the eukaryotic cytosol, the CCT chaperonin is approximately a tenth

as abundant as GroEL is in bacteria, but it is essential, as mentioned, to mediating folding of actin, tubulin, transducin, β-propeller proteins, and likely others as well. The range of substrate proteins recognized by CCT remains to be fully resolved. In archaebacteria, by contrast, the thermosome has a high basal abundance, and under heat shock conditions it becomes induced as virtually the only protein synthesized. The lack of facile genetic manipulation in the archaeal system has made difficult a full evaluation of substrates and action. It is notable that both CCT and the thermosome employ another chaperone known as *prefoldin*, a jellyfish-shaped component that binds nonnative proteins via hydrophobic tentacle tips, as an agent that delivers substrate proteins to them. In the bacterial cytoplasm, it appears that nascent chain chaperones such as trigger factor, which binds nascent chains at the ribosome, and DnaK, may interact with substrates occupying more extended conformations, before they arrive at the chaperonins in collapsed globular states.

Chaperonin architecture

GroEL The bacterial chaperonin assembly is composed of identical 58-kDa subunits arranged in two back-to-back seven-membered rings (Fig. 8). Each subunit is folded into three domains, an equatorial domain housing a nucleotide pocket, the collective of which form the waistline of the cylinder; an apical domain at the terminal end of the cylinder, the collective of which form a hydrophobic polypeptide binding surface at the cavity-facing aspect of a ring; and a hingelike intermediate domain between the equatorial and apical domains, which enables nucleotide-directed opening of the apical domains. The equatorial domains form tight contacts of the subunits with each other, both around a ring and between rings, providing a stable base from which movements of the intermediate and apical domains are directed by nucleotide and GroES binding. ATP is bound and hydrolyzed cooperatively within a ring but anti-cooperatively between them. As indicated, the apical domains positioned at the terminal ends of the GroEL cylinder present a hydrophobic surface at their cavity-facing aspect that is involved with binding nonnative polypeptides in the central cavity (see Fig. 8 rightmost panels, yellow). Such binding involves hydrophobic contacts between the apical cavity lining of the chaperonin and hydrophobic surfaces exposed specifically in nonnative polypeptides, surfaces that will be buried to the interior in the native state. It is such exposed hydrophobic surfaces in substrate proteins, as mentioned above, that are responsible for the

Fig. 8 Chaperonin structures. Crystallographic models of the bacterial chaperonin, GroEL (top row; pdb 1GRL), GroES (middle row), and GroEL–GroES complex (pdb 1AON) are shown, as ribbons traces (left panels) and space filling (right-hand panels), illustrating the arrangement of subunits (middle columns) and domains (left column). a, apical; i, intermediate; e, equatorial domains. Note that association of GroES with a GroEL ring leads to large-scale elevation and twisting of its apical domains (compare top and bottom rows). This is attended by a change of cavity wall character from hydrophobic (yellow in right panels) to hydrophilic (blue in top GroES-bound ring of lower right panel). Productive folding proceeds inside of the GroES-encapsulated cis ring, whereas the initial binding of nonnative polypeptide occurs to an open ring of an asymmetric GroEL–GroES–ADP complex. (See color plate p. xliii.)

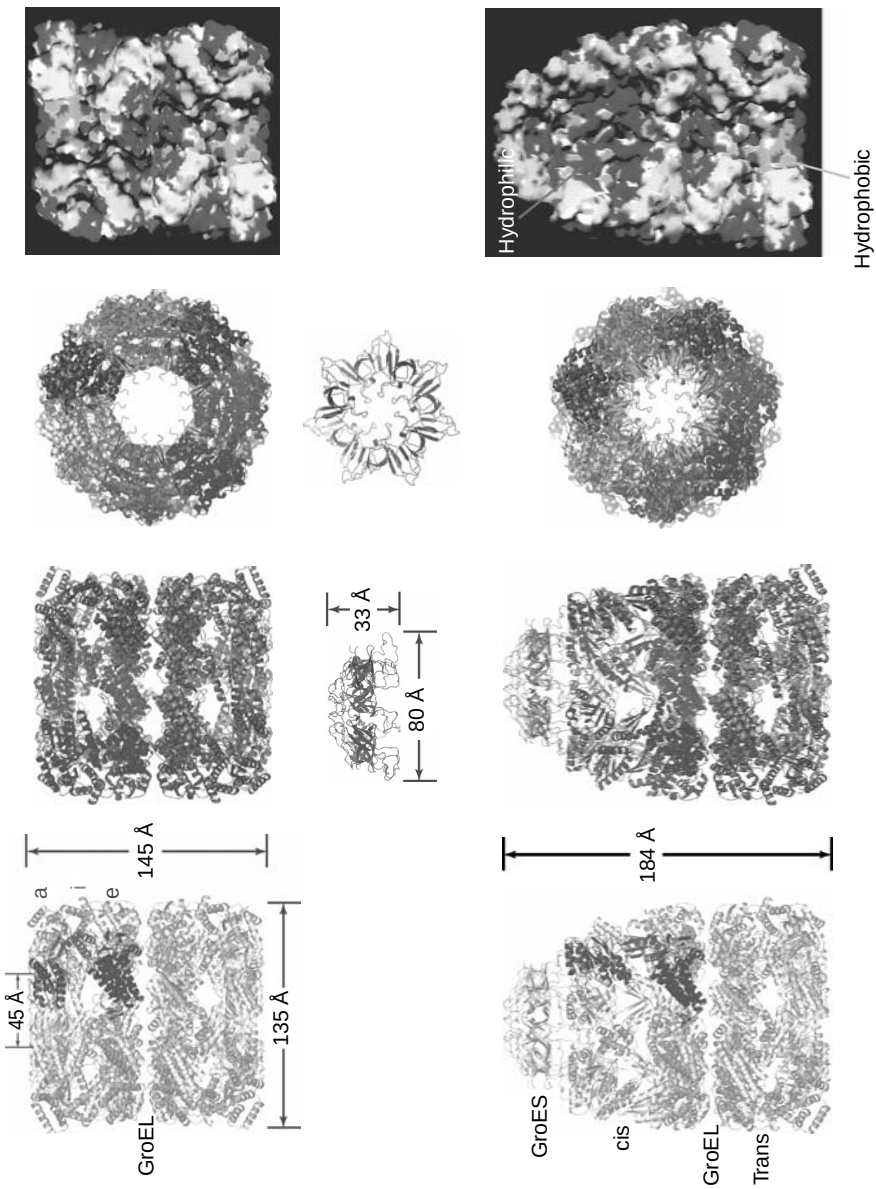
45 Å
a
i
e
GroEL
145 Å
135 Å
33 Å
80 Å
GroES
cis
GroEL
Trans
184 Å
Hydrophilic
Hydrophobic

undesirable behavior of multimolecular aggregation, which is associated with loss of substrate function, and with the harmful effects to the cell of protein aggregation. Thus, binding by the chaperonin, mediated through multiple surrounding apical domains simultaneously, effectively forestalls multimolecular aggregation. Upon binding ATP and then GroES to a GroEL ring, this same hydrophobic surface is elevated and twisted away from its original position, effectively removed from facing the central cavity, to form new contacts with the cochaperonin GroES through hydrophobic segments in each of its mobile loops. Such movement away from bound substrate polypeptide serves to release the polypeptide into what is now a hydrophilic chamber in which it can productively fold (Fig. 8, lower right panel, top ring).

GroES This single ring structure is composed of seven identical subunits, each comprised of a β-barrel from which extends a mobile loop segment of 17 to 18 amino acids that is freely mobile in the stand-alone state but which become immobilized in GroEL-GroES complexes (Fig. 8 middle and lower rows). Each GroES mobile loop contacts a hydrophobic aspect of a corresponding GroEL apical domain via a hydrophobic edge containing the sequence IVL. The inside aspect of the GroES cavity itself in GroEL–GroES complexes forms a smooth continuation of the GroEL cavity (Fig. 8).

Action in assisting protein folding The action of chaperonins in assisting folding is provided at two different steps of the reaction cycle (Fig. 9). During the step of binding nonnative protein in an open ring (left panel), the off-pathway events of misfolding and aggregation are prevented. Such binding may act, as discussed below, to effectively unfold nonnative proteins, giving them a fresh trial at productive folding at each round of binding to the chaperonin, occurring upon subsequent ATP/GroES-driven release into the chaperonin cavity (panel 2). During the step of such release, productive folding is supported by encapsulation in solitary confinement, by

GroEL – GroES reaction cycle

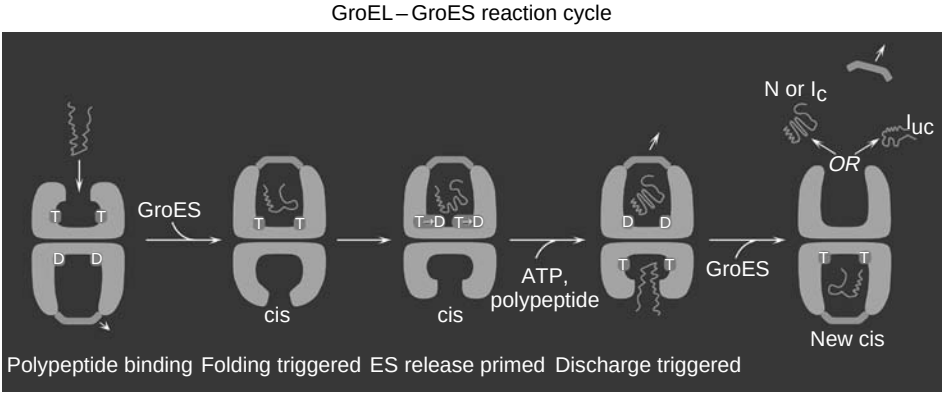

Fig. 9 GroEL–GroES–ATP-mediated folding cycle. Polypeptide is designated as a yellow line, GroES is designated in orange. T, ATP; D, ADP. I_c, intermediate folded form committed to reaching the native state in the bulk solution, that is, a state for which GroEL has no affinity. I_{uc}, nonnative form that is uncommitted, that is, that will rebind to a GroEL ring. See text for discussion of the steps of the reaction cycle. (See color plate p. xxvi.)

hydrophilic wall character and also by the narrow confinement of the walls, which may limit the off-pathway states that can be populated, excluding, for example, extended states.

An action of unfolding nonnative proteins in association with binding them in an open chaperonin ring has been supported by findings that chaperonin-bound nonnative proteins are very susceptible to proteases and also very susceptible to hydrogen-deuterium exchange, indicating little or no stable structure. Such a relatively unfolded state could be reached by two general mechanisms. One, called *thermodynamic partitioning*, recognizes that there is an ensemble of unfolded states for any given polypeptide species in the bulk solution and postulates that the chaperonin has greater affinity for the less folded of these states. Binding them will thus shift the equilibrium of states by mass action toward the less folded, effectively acting as an unfolding action. The second mechanism of unfolding would involve catalyzed unfolding, in which a polypeptide coming in contact with the chaperonin cavity wall becomes bound by multiple surrounding apical domains, which could act to unfold a protein. In one experiment, a catalytic effect of the chaperonin GroEL, supplied in substoichiometric amounts, on a small RNAse has been observed by hydrogen-deuterium exchange. Thus, either or both of these mechanisms could effectively reverse misfolding in the act of binding to chaperonin.

The nucleotide cycle The chaperonin reaction cycle is illustrated in Fig. 9. The two rings of GroEL behave in an asymmetric manner as a function of the behavior of ATP binding and hydrolysis, occurring cooperatively in the seven subunits of a ring but anti-cooperatively between rings. As a consequence of this, since GroES binding to a ring requires occupancy of the ring with ATP, which produces a degree of apical domain elevation and twist that enables initial interaction, GroES, like ATP, is asymmetrically bound to GroEL. The normal acceptor state for a nonnative polypeptide is thus the open ring of a GroEL–GroES–ADP asymmetric complex (panel 1). Polypeptide entering the open ring associates through its own exposed hydrophobic surface with the surrounding hydrophobic apical domains. The arrival of ATP at the same ring as polypeptide is an allosteric trigger to release the ligands from the opposite ring, a step that is accelerated by the coincident binding of nonnative polypeptide. In the presence of bound ATP, the polypeptide-bound GroEL ring binds GroES, and the further large apical domain movements attendant to this step effectively remove the hydrophobic binding surface from the central cavity, and trigger ejection of nonnative polypeptide off the cavity wall and into the now hydrophilic folding chamber (Fig. 9, panel 2). This folding-active state is the longest step of the GroEL–GroES reaction, ~8–10 s, following which ATP hydrolysis in the GroES-bound (cis) ring weakens the affinity of GroEL for GroES, "priming" the release of GroES (panel 3), which is triggered by the allosteric signal subsequently produced by ATP arriving in the opposite ring (panel 4). All of the cis ligands are ejected (panel 5): GroES, polypeptide whether it is native or not, and ADP. In the case of polypeptide, only a fraction of the molecules reach native form in any given cycle. For example, only ~5% of cis ternary complexes containing the monomeric enzyme, rhodanese, produce the native enzyme in a single round, so many additional cycles of apparently all-or-none folding are required in order for all of the molecules

to reach native form. Nonnative proteins released into the bulk solution have been shown to be unable to attain the native state in that location in an experiment preventing their rebinding to GroEL. Thus, in the cell, they must be rebound either by GroEL or by other chaperones, attempting once again to reach native form. Alternatively, if unable to be correctly folded, they are likely to be recognized and degraded by the proteolytic machinery.

Behavior during heat shock Early studies of GroEL/GroES showed that the abundance of these components was increased by more than 10-fold during heat shock, with levels accreting to 10 to 15% of total soluble protein. This is the result of an elevation of synthesis of these components directed via σ^{32} and the promoter element in the *GroESL* operon, as discussed earlier. Additional studies done later *in vitro* indicate that the chaperonins themselves are stable up to 60 °C, and the ATPase activity of GroEL is likewise accelerated. Yet under these conditions, efficiency of refolding is reduced, potentially because newly folded proteins become rapidly unfolded. Following relief of thermal exposure, however, productive folding promptly resumes.

4.2 Hsp70 System

Another bacterial operon under the control of σ^{32} is that encoding the chaperones DnaK (Hsp70) and DnaJ. These two proteins function together in the bacterial cytoplasm and in most other cellular compartments as an ATP-dependent chaperone machine. Unlike the GroEL/ES ring structures that bind nonnative polypeptide multivalently in a ring and carry out refolding in an encapsulated space, these components act as monomers, binding short hydrophobic stretches of amino acids present in extended segments of polypeptide chain, then releasing the chain into the bulk solution for further biogenesis.

These widely distributed chaperones have diversified through evolution in structure, localization, and function; and individual Hsp70s generally function with specific J-proteins. Bacteria have three or more related Hsp70s in their cytosol, but the major one, with general action, is DnaK. It makes up 1% of bacterial protein under optimal growth conditions and rises to 3% under stress conditions. In eukaryotes, there are multiple Hsp70s and J-proteins, localizing both in the cytosol and in major organelles, including

Fig. 10 Structure of the bacterial hsp70 protein, DnaK. Models of the crystal structures of the two major domains, the nucleotide binding domain (left, in green, pdb 1BUP) with complexed nucleotide, and the substrate binding domain with a complexed heptapeptide (right, in blue, pdb 1DKX). The two domains are joined together by a dotted line that signifies a linker region between the two domains. The structure and position of this linker and the orientation of the two domains relative to each other remain unknown. The peptide binding site is formed by loops emanating from a β-sheet, and an overlying α-helical lid structure contributes to opening and closing of access to the site. Lower panel shows the contacts formed between the heptapeptide and the binding site from a vantage point of the helical lid. Both hydrogen bonds (left, dashed lines) and hydrophobic interactions (right, DnaK side chains in space filling and those of the peptide as stick representation) contribute to binding of the central hydrophobic portion of the heptapeptide (Figure panels taken from Mayer, M.P., Bukau, B. (2004) Regulation of Hsp70 Chaperones by Co-chaperones, in: Buchner, J., Kiefhaber, T. (Eds.) *Protein Folding Handbook*, Part II, Chap. 15, Wiley-VCH, Weinheim, pp. 508–548). (See color plate p. xxxviii.)

the nucleus, endoplasmic reticulum, mitochondrion, and chloroplast.

While these different machines have diverged in structure and function, a fundamental mechanism of action is shared. The Hsp70s bind short extended hydrophobic stretches of polypeptide through a hydrophobic archway in the 15-kDa COOH-terminal β-sandwich domain (Fig. 10, top right), an interaction regulated by the binding and hydrolysis of ATP in the cleft of the bilobed N-terminal domain (Fig. 10, top left). Substrate protein can be directly bound by Hsp70 in

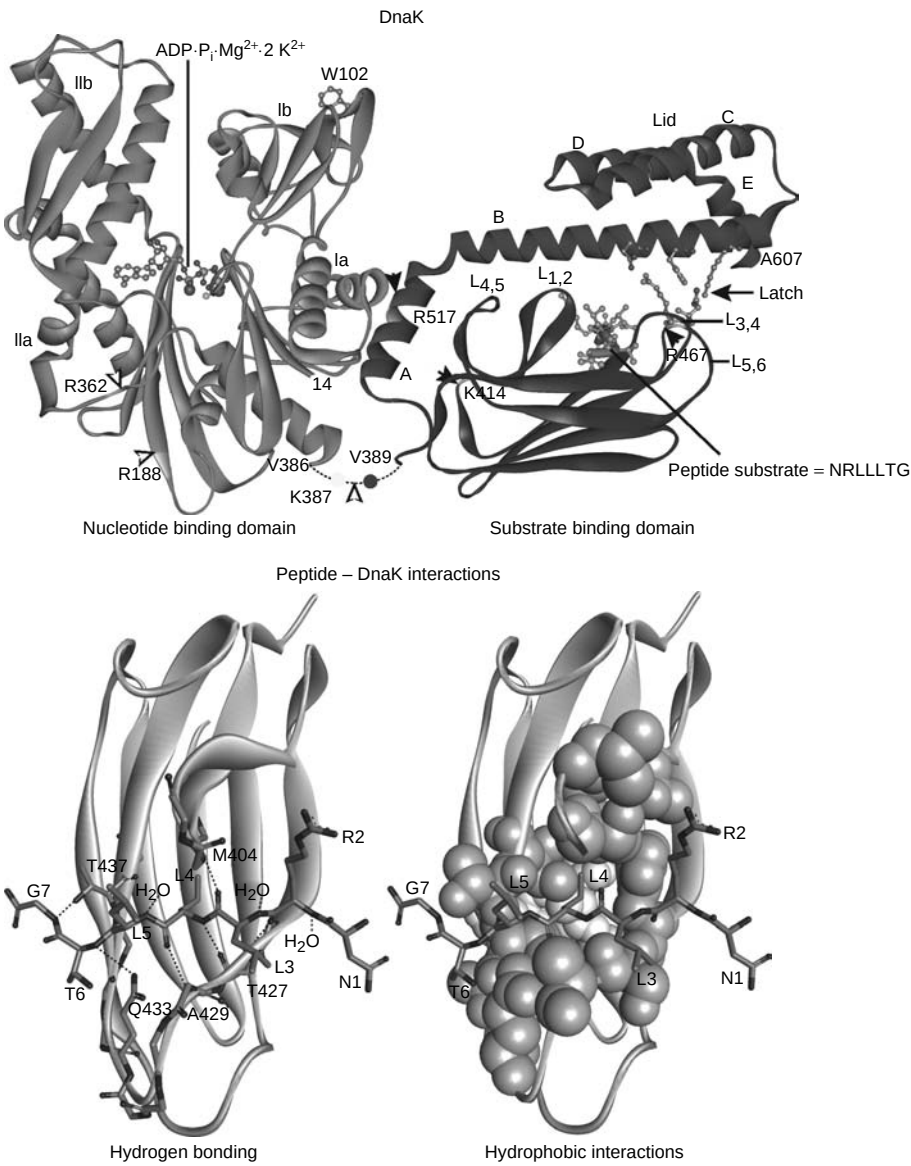

its ATP-bound state, but in many cases it seems likely that substrate is "delivered" to an Hsp70 peptide binding site by a J-protein partner chaperone that forms the initial interaction with substrate via its own C-terminal binding site for hydrophobic amino acid sequences (Figs. 11, 12c). Subsequent hydrolysis of ATP by the Hsp70 is strongly stimulated by interaction with the J-protein, an interaction remote from the peptide binding sites between the so-called "J domain" of the partner J-protein and the nucleotide binding domain of the Hsp70 protein (Figs. 10, 11, 12b). ATP hydrolysis leads to an allosterically directed large increase in affinity of the Hsp70 for its substrate protein, characteristic of the ADP-bound state of Hsp70s. The substrate thus becomes tightly bound in the Hsp70 peptide binding pocket until a nucleotide

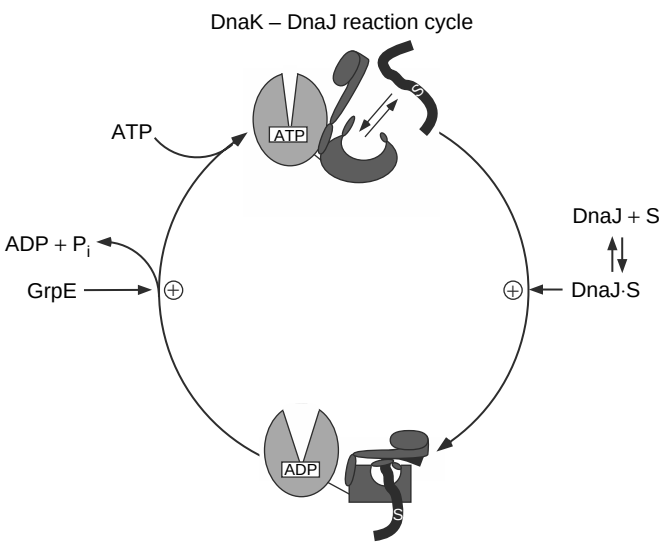

Fig. 11 DnaK–DnaJ reaction cycle. DnaK is represented schematically, with nucleotide binding domain colored light gray and peptide binding domain a darker gray. Nonnative substrate protein, S, is shown in black. The ability of DnaK to interact with substrates in its ATP-bound form is shown at the top, with rapid on and off-rates signified by the arrows. DnaJ's ability to also recognize substrate and to deliver substrate to DnaK is shown at the right. Such delivery of substrate by DnaJ, associated with binding to DnaK, is accompanied by accelerated ATP hydrolysis by DnaK (see text), associated with a locking in of bound substrate protein by the ADP-bound state of DnaK. In bacterial and mitochondrial systems, a nucleotide exchange factor, GrpE, binds to the ADP-bound nucleotide pocket of DnaK (mhsp70) (see Fig.13), releasing the nucleotide and allowing entry of ATP. In other hsp70 systems, such an exchanger is not operative, but other factors that modify nucleotide binding or residence can be involved (see Mayer, M.P., Bukau, B. (2004) Regulation of Hsp70 Chaperones by Co-chaperones, in: Buchner, J., Kiefhaber, T. (Eds.) *Protein Folding Handbook*, Part II, Chap. 15, Wiley-VCH, Weinheim, pp. 508–548; figure modified from their review).

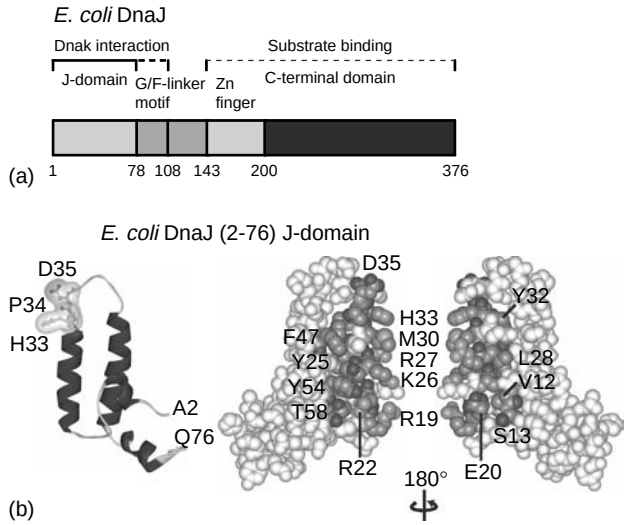

Fig. 12 DnaJ organization and structure. (a) Domain layout of *E. coli* DnaJ. Taken from Bukau, B., Horwich, A.L. (1998) The Hsp70 and Hsp60 chaperone machines (review), *Cell* **92**, 351–366. (b) Structure of the J-domain region (1XBL); at left, a ribbons diagram of the J-domain shows the arrangement of α-helices and a conserved HPD motif involved with interaction of DnaJ with DnaK's ATP binding domain; at right, a space-filling view shows in color the region of the J-domain that undergoes NMR chemical shift changes upon interaction with DnaK, from the work of Greene, M.K., Maskos, K., Landry, S.J. (1998) Role of the J-domain in the cooperation of Hsp40 with Hsp70, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6108–6113. (c) Crystallographic model shown in ribbons (left) and space-filling representations (right) of the C-terminal substrate binding domain of DnaJ homolog, Sis1, from *S. cerevisiae* (pdb 1C3G). Hydrophobic depression comprising the substrate binding site is designated in yellow. This assignment has been supported by mutational studies. Panels B and C taken from Mayer, M.P., Bukau, B. (2004) Regulation of Hsp70 Chaperones by Co-chaperones, in: Buchner, J., Kiefhaber, T. (Eds.) *Protein Folding Handbook*, Part II, Chap. 15, Wiley-VCH, Weinheim, pp. 508–548. (See color plate p. xl.)

exchange event opens the cleft by replacing bound ADP with ATP, associated with reduced affinity for substrate and its dissociation, freeing it for subsequent steps of folding or targeting (Fig. 11). Nucleotide exchange in some cases is catalyzed by an exchange factor, for example, the protein in the bacterial cytoplasm known as *GrpE*.

Using the foregoing general mechanism, Hsp70 proteins carry out a variety of functions in the cell, including the following:

1. Assisting in the folding of newly translated proteins emerging from or recently departed from the ribosome. In the former case, binding to nascent

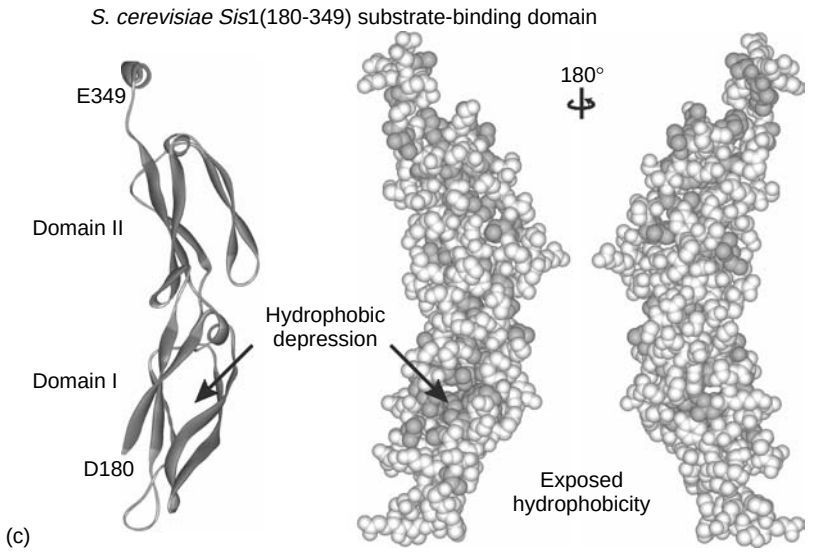

Fig. 12 (*Continued*)

chains may serve to prevent premature folding of a polypeptide. For example, where there are interactions in the native state between N and C-terminal regions, such a "holding" action of the N-terminal region could prevent premature folding until the C terminus has been translated and is available. In the case of proteins already released from the ribosome, binding by Hsp70 may prevent misfolding and aggregation during trials at correct folding to the native state.

2. Assisting posttranslational protein translocation across membranes, for example, translocation into the ER or mitochondria. On the cytosolic (cis) side of such membranes, Hsp70s act to maintain polypeptides in "loose" conformations, allowing them to engage and pass through translocons. Inside the organelles, at the trans side of membranes, Hsp70s act to facilitate translocation by binding unfolded stretches of polypeptide, biasing their progressive forward movement through the translocon and in some cases potentially providing a forceful pulling action.
3. Reversing incipient aggregation, under normal conditions or under stress conditions. Here Hsp70s appear to be able to pry apart low order aggregates of two to three monomers, potentially by shifting an equilibrium between such states toward that of bound monomer. An action on higher-order aggregates can also be mediated, but here the action of specialized Hsp100 hexameric AAA+ ring assemblies such as Hsp104 or ClpB, discussed in the following sections, is required. Recently it has been demonstrated that such disaggregating action by ClpB involves passage of monomers through its central channel, assisted by the action of Hsp70.
4. Mediating dissociation of specific oligomeric proteins. In bacterial cells, for example, such dissociation is

observed in the λ phage biogenesis system, involving dissociation of λP from λO-λP-DnaB ternary complexes, mediated by the Hsp70 system. In eukaryotic cells, an example is the dissociation of clathrin triskelions during clathrin coat disassembly, with Hsc70 recruited by the DnaJ domain of a clathrin-associating component, auxilin. Recent cryoEM studies of clathrin and clathrin–auxilin complexes suggest that Hsp70 may bind to a hydrophobic unstructured C-terminal region of the clathrin heavy chain subunit that localizes inside the clathrin lattice.

5. Facilitating biogenesis of steroid receptors and signal transducing kinases in the eukaryotic cytosol by participation in biogenesis of Hsp90 complexes. Here, cytosolic Hsp70 proteins function in cooperation with J-proteins to mediate an early step in the biogenesis of these molecules on their pathway of assembly into Hsp90 multichaperone complexes.
6. Regulating the heat shock response via binding of the bacterial heat shock regulatory transcription factor, σ^{32}, as discussed above, sequestering it from binding to RNAP. A similar action appears to occur in eukaryotes, with Hsc70 binding to the heat shock factor HSF.

History and general physiology Like GroEL/GroES, DnaK and DnaJ were first identified in the 1970s as genes required for growth of phage-λ. While GroEL/GroES are necessary for phage head and tail morphogenesis, DnaK and DnaJ are critical for the replication of λ DNA. DnaK/J were found to be important for remodeling the protein complex required for initiation of λ DNA replication, mediating an action of dissociation of a component from a ternary complex. During these studies, an Hsp70 of the lumen of the endoplasmic reticulum, BiP, was observed to complex with immunoglobulin light chains in the absence of partner heavy chains, a clue that led to appreciation of the involvement of Hsp70 proteins in protein biogenesis. However, a role in binding monomeric polypeptide chains was not clear until *in vivo* experiments with yeast reported in 1988 by a collaboration of Craig and Schekman groups, and *in vitro* ones reported by Blobel and coworkers, demonstrated that lack of Hsp70s had deleterious effects on protein conformation, preventing precursor proteins in the cytosol from occupying "loose" conformations required for entering either the ER or mitochondria. Similarly, a blocking of the function of Hsp70s at the other side of these membranes inside the respective organelles led to a blocking of translocation into them, as observed by the combined work of the groups of Craig and Pfanner/Neupert, suggesting that the role of binding the entering extended chains is required for translocation. Later experiments *in vitro* corroborated an action of Hsp70 on monomeric polypeptides, showing that purified DnaK, DnaJ, and GrpE were able to facilitate the refolding of denatured firefly luciferase. Recent experiments indicate that DnaK interacts with up to 18% of newly synthesized proteins prior to their folding, and that the action of the DnaK machinery is essential for survival upon exposure to stress such as heat shock.

Structure of Hsp70s, DnaJ proteins, and GrpE

Hsp70 Hsp70s are highly conserved monomeric proteins of approximately

70-kDa molecular mass. An amino terminal 44-kDa adenine nucleotide binding domain has a low intrinsic ATPase activity (Fig. 10). The carboxy terminus contains two regions. The more amino-terminal subdomain is the substrate binding region consisting of eight antiparallel β-strands that form a hydrophobic cleft that, in both peptide binding experiments and in a cocrystal, interacted with five consecutive hydrophobic residues in a bound peptide. Binding was favored by the presence of basic residues flanking the hydrophobic stretch in the substrate peptides. A survey by Bukau and coworkers of a number of proteins suggested the presence of a recognizable hydrophobic stretch that was on average every 36 amino acids across primary structures. Where correspondent three-dimensional native structures were available, in many cases these regions localized to a buried hydrophobic core of the native state.

Concerning the mechanism of Hsp70 binding, an α-helical region in the substrate binding domain of DnaK, following the sequences forming the β-sheet sandwich of the peptide binding cleft, appears to be able to form a lid over the bound peptide (Fig. 10). Deletion of this region affects affinity for nonnative substrates. The function of the C-terminal-most 10-kDa segment of Hsp70s, the most divergent region among the family, is less well understood, but it may play a role in modulating the interaction between the ATPase and substrate binding domains, or it may mediate interaction with partner J-proteins.

The two nucleotide states of Hsp70 have profoundly different characteristics in binding polypeptide substrates. In the ATP-bound state, the affinity for substrate is very low, with the on-rates and off-rates being very rapid (on the order of milliseconds) (Fig. 11). Interaction with substrate in the ADP-bound state, by contrast, is quite stable, with the off-rates and on-rates both being relatively slow (on the order of minutes) (Fig. 11). It is thought that allosterically mediated conformational changes of the COOH-terminal peptide binding domain are brought about by nucleotide binding in the NH_2-terminal domain. Such binding affects, for example, positioning of the α-helical lid, affecting affinity for nonnative proteins. In the ATP-bound state, the lid is disengaged allowing rapid entry and exit of substrate protein; in the ADP-bound state, the lid is proposed to be clamped tightly over the peptide binding pocket preventing release.

J-proteins J-proteins as a class are a very divergent group of proteins, having in common a J-domain that interacts with the ATPase domain of Hsp70 (Fig. 12a,b). J-domains are composed of four α-helices and have a highly conserved histidine, proline, aspartic acid (HPD) tripeptide found to be critical for J-domain function. While all J-proteins interact via HPD segments with Hsp70 ATPase domains, this alone is not sufficient to determine the partnering relationship, and other regions of J-proteins are divergent both in structure and function. A common class of J-proteins, represented by *E. coli* DnaJ, has three domains in addition to the J-domain: a glycine/phenylalanine-rich region, a zinc-binding region, and a C-terminal region. The two latter carboxy-terminal regions are involved in binding of nonnative protein substrates (Fig. 12c). Other J-proteins have domains that tether them to particular cellular locations, such as ribosomes (i.e. Zuo1 of yeast) or particular membranes (i.e. Pam18 of the

mitochondrial inner membrane or Sec63 of the endoplasmic reticulum membrane).

Nucleotide release factors Factors that destabilize the interaction of nucleotides with Hsp70 are important in the function of Hsp70/J-protein machinery. GrpE of the DnaK/DnaJ system is the best understood. It is a dimer that interacts with both the ATPase and substrate binding domain (Fig. 13). The interaction, which is stable in the presence of ADP, facilitates the release of this nucleotide. Orthologs of GrpE are found in mitochondria. However, structurally unrelated factors, such as Bag1 in the cytosol of mammals, serve as nucleotide release factors for other Hsp70/J-protein machineries.

The substrate polypeptide interaction cycle Interaction of Hsp70 with substrate polypeptides must be transient for them to carry out their biological function. This is accomplished by utilizing the great disparity in the interactions between Hsp70s and substrate polypeptides in the ATP- and ADP-bound states. Hsp70s bind substrate when in the ATP-bound state, as the on-rate is very rapid. The interaction is stabilized through hydrolysis of ATP. Hydrolysis is stimulated by at least two mechanisms. First, the interaction of polypeptide in the substrate binding site stimulates ATPase activity. Second, and more importantly, the J-protein partner stimulates ATP hydrolysis. The release of the substrate is modulated by the activity of nucleotide release factors (Fig. 11). Upon release of ADP, ATP rapidly binds, as it is typically present in much higher concentrations *in vivo*. This binding converts Hsp70 to a form having a rapid off-rate for substrate. Thus, substrate is released and Hsp70 is primed for another cycle of substrate interaction.

Multiple physiological roles of a single Hsp70 – example of mitochondria Hsp70s can function with more than one

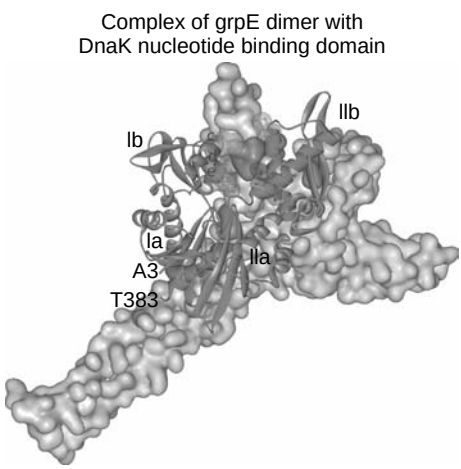

Fig. 13 Nucleotide exchange factor, GrpE, in space-filling representation (yellow and gold), complexed as a homodimer with the ATP binding domain of DnaK shown in ribbons (pdb 1DKG). The GrpE dimer interface is formed by two long α-helices (lower aspect) and, at the opposite end, a four-helix bundle. Interaction with the ATPase domain of DnaK occurs principally via a β-sheet of one of the monomers (protruding up from the long axis of GrpE, with sites of interaction colored orange) interacting mainly with DnaK's Ib and IIb subdomains. This interaction leads to a rotational displacement of IIb relative to the nucleotide-bound state, displacing bonds that would be formed with the base and ribose of ADP. The nucleotide binding site is thus disrupted by GrpE binding. Notably, in biochemical studies, GrpE binds to the ADP-bound and nucleotide-free states of DnaK with nanomolar affinity, but it is instantly dissociated from DnaK upon addition of ATP. Figure taken from Mayer, M.P., Bukau, B. (2004) Regulation of Hsp70 Chaperones by Co-chaperones, in: Buchner, J., Kiefhaber, T. (Eds.) *Protein Folding Handbook*, Part II, Chap. 15, Wiley-VCH, Weinheim, pp. 508–548. (See color plate p. xxviii.)

J-protein partner, which in part explains the functional diversity amongst these Hsp70-based machineries. For example, the soluble Hsp70 of the mitochondrial matrix, Ssc1, not only works with the J-protein of the mitochondrial inner membrane, Pam18, in the translocation of proteins from the cytosol into the matrix space of mitochondria but also functions with the soluble J-protein, Mdj1, in the folding of the newly imported proteins and refolding of partially denatured proteins. The specialized role of Pam18 in import is in part a function of being precisely tethered to a particular cellular location. Pam18 and Ssc1 are themselves independently targeted to the translocon of the inner mitochondrial membrane through which polypeptides are imported. This targeting ensures a high concentration of Hsp70/J-protein ready to interact with the unfolded polypeptide as it enters the matrix. (Similarly the cytosolic Hsp70/J-protein pair of yeast, Ssb and Zuo, localize near the exit of the tunnel of the ribosome, positioned to interact with emerging nascent polypeptides.) Second, the Ssc1/Pam18 machinery serves as an example of the adaptation of the ability of Hsp70s to bind short hydrophobic segments of polypeptide for biological roles other than protein folding. The interaction with translocating polypeptides serves as the driving force for the translocation of the bulk of cytosolically synthesized proteins across the inner membrane. This process utilizes the regulation of transient interactions by both a J-protein partner and a nucleotide release factor for cycles of interaction with the polypeptide, preventing backward slipping toward the cytosol.

In addition to gaining specialization through tethering to specific cellular targets, there are examples of specialization of substrate specificity. The best studied example is that of the specialized Hsp70/J-protein pair involved in the biogenesis of Fe−S cluster proteins: HscA(Ssq1)/HscB(Jac1) in bacteria (yeast). Apparently only one substrate exists for this machinery. The folded scaffold protein on which an Fe−S cluster is built, IscU (Isu), binds both the J-protein and the Hsp70 of this machinery. How these interactions promote Fe−S cluster assembly has yet to be resolved. Such interaction of the Hsp70 system with a native structured component also occurs in the case of λ DNA replication where DnaK pries apart a protein–protein interaction, and in regulation of the heat shock response where DnaK binds sigma 32 itself regulating its lifetime, as discussed earlier.

4.3 Small Heat Shock Proteins

Additional components that are σ^{32} inducible are the small Hsps, members of a ubiquitous family of proteins whose small subunits assemble into large oligomeric assemblies, containing 9 to more than 30 subunits, able to bind nonnative proteins at their surface via hydrophobic contacts, protecting them from aggregation. The subunits range in size from 16 to 40 kDa and invariably contain a C-terminal 90 amino acid domain, the so-called α-crystallin domain. Included in the sHsp family are the lens α-crystallins, which play a key role in preventing protein aggregation that would result in cataract formation. Multiple sHsp genes are found in most species, the products localizing typically to the cytosol, but under some conditions to the nucleus, and in some species also to organellar compartments. The nature of the heterogeneity of sHsp structure and function within any given

cellular compartment remains to be understood. In many cases, coassembly of different sHsp subunits occurs into any given sHsp complex within a given cellular compartment, further complicating functional evaluation.

Two different sHsps that form regular assemblies have been resolved crystallographically (Fig. 14a), one from an archaeon, *Methanococcus jannaschii*, and the other from wheat, showing two distinct types of assembly, the former a 24mer, soccer-ball like structure with a hollow core and side windows, with an outer diameter of 120 Å, 432 symmetry, and monomers all in equivalent structural environments, effectively four trimers of dimers. By contrast, the wheat structure is a 12-subunit assembly composed of two stacked six-membered discs, each a trimer of dimers, 95 Å in diameter with a hollow cavity. For both of these structures, there are nonresolving sequences from the nonshared N-terminal regions that may localize within the hollow cavities. Yet both reveal the conserved C-terminal α-crystallin domain, a β-sheet sandwich. A basic dimer building block of these sHsps is formed from longitudinal stacking of the α-crystallin domains of two subunits. The nonconserved N-terminal regions then contribute to the unique assemblies that are formed. By contrast with these two crystallized sHsps, many other sHsps form assemblies that are heterogeneous in subunit number and organization as judged from EM studies.

The nature of substrate binding by sHsps remains to be fully understood, but properties of subunit dissociation

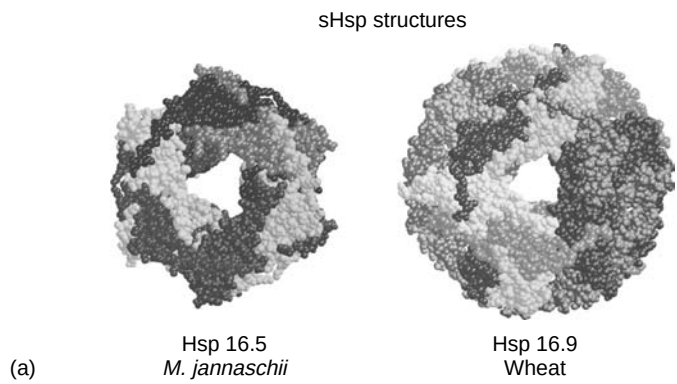

Fig. 14 Small heat shock protein structures and mechanism of action. (a) Crystallographic models of small hsps from archaeon *M. jannaschii* and wheat (pdb 1SHS, Kim et al, 1998 and 1GME, van Montfort et al, 2001, respectively) are shown, with coloring of individual subunits. (b) Mechanism of action of small hsps. Under heat stress, shsps undergo accelerated subunit exchange and structural rearrangement, associated with presumed exposure of hydrophobic surfaces that recruit nonnative polypeptides, forming substrate-chaperone complexes. Such action prevents aggregation of the nonnative polypeptides that occurs otherwise. Upon relief of heat stress, bound proteins are released and can be assisted to refold by the hsp70 system. Figure taken from van Montfort, R., Slingsby, C., Vierling, E. (2002) Structure and function of the small heat shock protein/α-crystallin family of molecular chaperones, *Adv. Protein Chem.* **59**, 105–156.

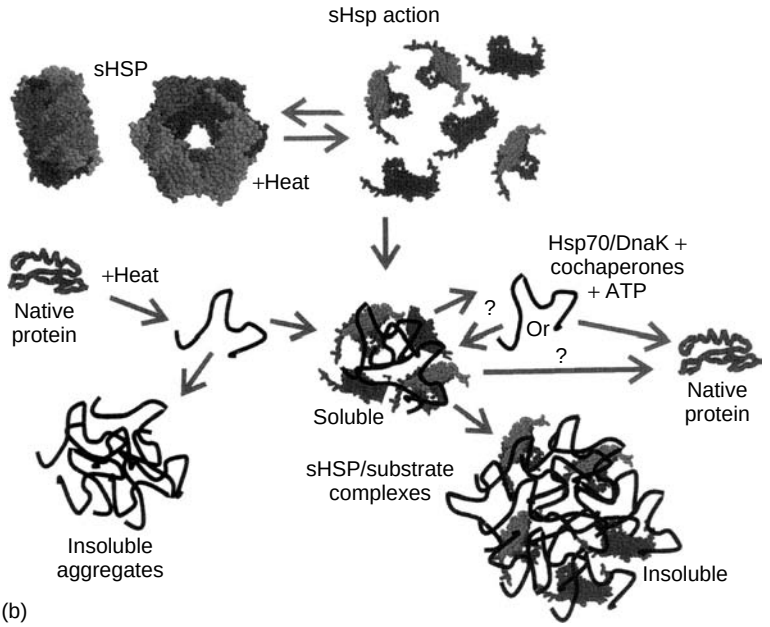

(b)

Fig. 14 (*Continued*)

and exchange, and/or structural rearrangement of subunits under stress conditions appear to figure in exposing hydrophobic binding sites that are otherwise buried as viewed in the X-ray structures (Fig. 14b). Both N-terminal sequences and a conserved motif in the α-crystallin domain of the wheat sHsp have been implicated by Bis-ANS binding experiments, the latter sequence noted to be buried in the structure as viewed by X ray, necessitating subunit dissociation or rearrangement to enable surface exposure.

Experiments *in vitro* reveal that complexes between sHsps and substrate proteins form larger-order heterogeneous structures, with multiple substrate proteins able to be bound in nonnative protease-susceptible form by an assembly of sHsps. Such binding of substrate serves to prevent stress-exposed proteins from aggregation. Recovery of the native form of bound substrate proteins ultimately requires release from the sHsps followed by refolding, assisted in many cases by other molecular chaperones, for example, the Hsp70 and chaperonin systems.

In addition to a protective role under stress and a role in lens maintenance, sHsps have been implicated in maintenance of the actin and intermediate filament cytoskeletons. Consistent with such a role, mutation of αB-crystallin has been found in a desmin-related myopathy. Accumulation of αB-crystallin in aggresome-like structures with staining properties of amyloids has also been observed in heart muscle of mice carrying the same mutation. An involvement of sHsps in blocking apoptosis has also been indicated by observation of binding to holocytochrome *c* released from mitochondria, blocking its activation of caspases.

5 Other Heat-inducible Components

The early studies of Neidhardt and VanBogelen and more recent ones have identified a number of other components that are induced by heat shock.

5.1 Hsp90

This chaperone is present in high abundance in the eukaryotic cytosol (~2–5% soluble protein), in the ER as the homolog GRP94, and in the bacterial cytoplasm (HtpG); in all three locations it is stress inducible. This suggests that it plays a general role in protecting misfolding proteins, but, despite the availability of both conditional mutants and pharmacological inhibitors, such action has not been demonstrated to date in any cellular context. Rather, a more specialized role has been identified in a late stage of the folding of several classes of substrate protein. This was first uncovered for Hsp90 in the eukaryotic cytosol, where, functioning in the context of multiprotein complexes containing a variety of "co-chaperone" components, Hsp90 binds near-native conformations and assists the final folding of a variety of "client" proteins including nuclear receptors, such as the glucocorticoid, progesterone, estrogen, and androgen receptors, and protein kinases such as src and PKB/Akt. In the case of the receptors, action by Hsp90 complexes is required in order to attain a ligand-activatable state, and the step of ligand binding releases such receptors from Hsp90 complexes in active forms that carry out transcriptional regulation. In the case of kinases it appears that binding of the catalytic domain is involved, potentially of the activation loop segment, enabling these molecules to assume an active state. A similar maturational action in the ER has been resolved for GRP94, assisting the biogenesis of, for example, ErbB2, immunoglobulins, and MHC class II molecules, but GRP94 also participates in binding peptides that will be presented by Class I molecules. Notably, in contrast with Hsp90 in the cytosol, the action of Grp94 in the ER appears to occur in the absence of associating cochaperones.

The multiprotein complexes involved in cytosolic Hsp90 action include such cochaperone components as Hsp70 or Hsp40 which likely mediate initial binding of client protein substrates prior to binding by Hsp90; Hip which modifies Hsp70 ATP binding activity; immunophilins bearing peptidylprolyl isomerase activity, such as Cyp40 and FKBP51/52; Hop/Sti1 a TPR (tetratricopeptide) repeat-containing "adaptor"/scaffold protein able to bind EEVD motifs at the C termini of Hsp70 and Hsp90; Cdc37 implicated in recruiting client kinases via its N terminus to Hsp90, which it binds through its C terminus; and p23/Sba1 which binds Hsp90 in the ATP-bound dimeric state and which appears to play a role in enabling productive release of client proteins. The formation and maturation of these multiprotein complexes is associated with an ordered binding and release of the specific cochaperone components, and this appears to regulate ATP binding and hydrolysis by Hsp90, which is in turn coupled with the conformational activation and release of client proteins.

For example, "early" steroid receptor complexes contain Hsp70 and Hop/Sti1 but lack Hsp90, whereas "later" ones contain Hsp90, immunophilins, and Sba1 but lack Hsp70 and Hop/Sti1. It is "intermediate" complexes that are involved with client protein recruitment to Hsp90, inhibited in ATP hydrolysis in such complexes

by the presence of Hop/Sti1. Departure of Hop/Sti1 and binding of immunophilins, however, then allows Hsp90 in a "late" complex to bind ATP, maintaining the client protein in an activatable state. This most likely involves ATP-bound Hsp90, producing or maintaining an open hydrophobic ligand binding pocket in the receptor molecule, producing a conformation that allows steroid ligand binding. The site(s) on Hsp90 that is involved in forming such physical association with client proteins remains to be defined, but presumably hydrophobic surfaces in the chaperone itself are involved, with candidate surfaces identified in crystallographic studies in both the middle and C-terminal domains. An open cavity formed between two outwardly curved regions of the molecule, dimerized through the C terminus (see Fig. 15), could be a site for localization of the bound substrate. Once ATP is turned over by the very slow N-terminal ATPase domain of hsp90 ($\sim$0.2 min^{-1}), the molecule may assume a more closed state, and these hydrophobic regions may become buried, driving release of the client receptor molecule, regardless of whether steroid hormone ligand has been bound.

Given the nature of its client proteins, involved with cellular signaling and growth control, hsp90 has been suggested to potentially function as an evolutionary "capacitor" that allows mutant versions of signaling kinases or receptors to remain functionally stabilized under normal conditions such that under unfavorable conditions, where hsp90 may be recruited to misfolded states, these mutant components may then become phenotypically expressed and may specifically allow survival of an "evolved" organism. Evidence in support of such a model comes from stress treatment of flies and plants, as well as specific inhibition of hsp90 by ansamycin drugs or RNAi – such treatments prompt exhibition of phenotypes that reflect uncovering of mutations lying in these pathways. In sum, the idea is that if hsp90 becomes tied up with stress response it is no longer able to suppress the misbehavior of mutants in critical signaling pathways, leading to a phenotype.

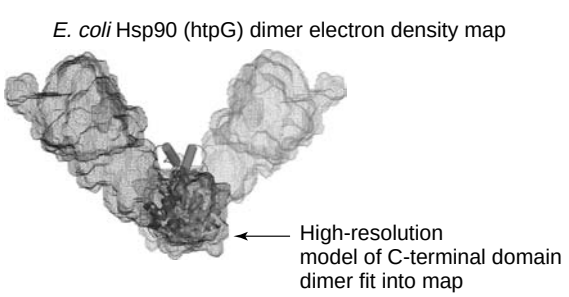

Fig. 15 Electron density map of intact *E. coli* HtpG (hsp90) dimer, illustrating the potential canyon between monomers in which nonnative substrates may become bound. Contact between the HtpG monomers (in the absence of nucleotide) occurs through their C termini, for which a high-resolution crystal structure has recently been solved (pdb 1SF8). Notably, structures have also been solved for the N-terminal ATP binding domain (also the site of ansamycin drug binding) as well as middle region of eukaryotic hsp90. Here, in the high-resolution structure of the dimer of HtpG C termini, the dimer interface involves a four-helix bundle. Two additional helices point out of density and their hydrophobic character and position within the "jaws" of the dimer, as well as functional studies, suggest potential involvement in substrate binding. Additional hydrophobic sequences, however, from the middle region have also been implicated (see text). Figure taken from Harris, S.F., Shiau, A.K., Agard, D.A. (2004) The crystal structure of the carboxy-terminal dimerization domain of htpG, the *Escherichia coli* Hsp90, reveals a potential substrate binding site, *Structure* **12**, 1087–1097.

5.2
Hsp33 and Oxidative Stress

Another heat responsive component that also counters oxidative stress, manifesting general chaperone activity as reflected by binding of a host of nonnative proteins under stress conditions, is the protein known as hsp33. Under normal cellular conditions hsp33 is abundant but occupies a reduced and inactive state, coordinating zinc through its four conserved cysteines (see Fig. 16), but under conditions of oxidative stress the zinc is released, allowing the cysteines to form disulfide bonds, associated with a conformational switch of the protein that allows its dimerization, exposing surfaces that bind nonnative polypeptides. Such action appears to be further enhanced by exposure to the added stress of heat, which strongly induces the abundance of hsp33. Relief of stress and return to reducing conditions allow release of substrate and dissociation of the hsp33 dimer (see Fig. 16). Released substrate is assisted with reaching native form by action of the hsp70 system. A recent study suggests that inactivation of DnaK occurs

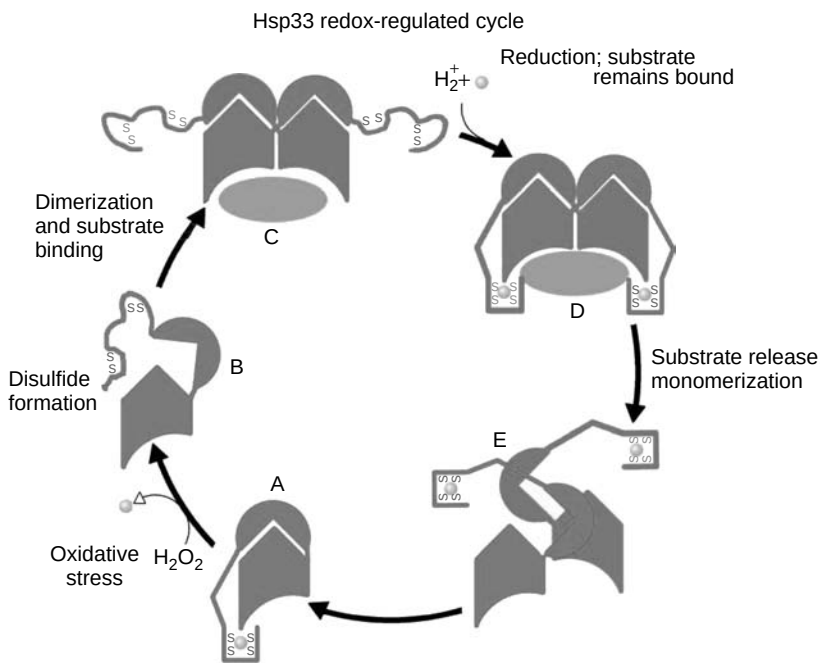

Fig. 16 hsp33 chaperone cycle. Under oxidative stress conditions, four cysteine residues in the C-terminal region of hsp33 switch from a zinc-coordinated state to forming a pair of disulfide bonds, associated with hsp33 dimerization and with exposure of a binding surface for nonnative polypeptides. Upon return to normal conditions, these steps are reversed and substrate proteins are released. They are subsequently acted on by the hsp70 system to regain the native state. Figure taken from Janda, I., Devedjiev, Y., Derewenda, U., Dauter, Z., Bielnicki, J., Cooper, D.R., Graf, P.C.F., Joachimiak, A., Jakob, U., Derewenda, Z.S. (2004) The crystal structure of the reduced, Zn^{+2}-bound form of the B subtilis Hsp33 chaperone and its implications for the activation mechanism, *Structure* **12**, 1901–1907.

under conditions of both heat plus oxidative exposure, a condition where hsp33 becomes the active chaperone; the roles are then reversed as normal conditions are resumed.

5.3
Hsp15 and Free 50S Ribosomal Subunits

This abundant protein is highly heat inducible and binds with high affinity to free 50S ribosomal subunits bearing nascent chains, as distinct from intact 70S ribosomes. A role in recycling 50S subunits bearing a nascent chain has been proposed. hsp15 appears to most immediately recognize an RNA moiety.

5.4
FtsJ

FtsJ also appears to bind nucleic acids. It also carries a methyltransferase fold, and its ability to methylate 23S rRNA within 50S ribosomal subunits of *E. coli* has been demonstrated both *in vivo* and *in vitro*. Null mutants exhibit an altered ribosome profile and growth defects. A condition of human X-linked mental retardation has recently been associated with defects in an FtsJ homolog whose yeast counterpart is involved in methylation of tRNA anticodon loops. Thus, the human condition likely results from defective translation.

6
Chaperone-protease Pairs – ClpP, HslV, and Lon, and Associated ATP-dependent Chaperone Unfoldases, ClpA, ClpX, and HslU

A set of proteases is also heat inducible, enabling the efficient degradation under stress conditions of damaged proteins. In the bacterial cytoplasm and mitochondrial compartment, these include the cylindrical proteasome-like proteases ClpP, HslV, and Lon. These components strongly resemble the eukaryotic proteasome in architecture and overall action, and HslV, in particular, resembles it in the mechanism of proteolysis. In all cases, access to these proteases occurs via narrow openings in their terminal ends, allowing only those substrates that have been selected and unfolded to enter, effectively producing a compartment within a compartment, that is, a proteolytic chamber within a cellular compartment (Fig. 17). The selection, unfolding, and translocation of substrates in all of these cases is governed by a cylindrical chaperone/unfoldase moiety, also housing a narrow central passage, either in a separate molecule that coaxially associates with the protease (ClpA and ClpX with ClpP, and HslU with HslV), or, in the case of Lon, the unfoldase moiety resides within the same protease cylinder. The chaperone moiety recognizes the specific proteins to be targeted for degradation via specific elements, and, through the action of ATP binding and hydrolysis, couples unfolding of the substrate protein with translocation down the axial passageway into the protease (Fig. 17).

6.1
Substrate Recognition and Translocation Mediated by the Chaperone Moiety

It is the step of recognition that confers specificity, triggering the subsequent steps of unfolding and degradation, and such recognition involves somewhat different strategies between prokaryotes and eukaryotes. In prokaryotes, such hsp100 unfoldases as ClpA and ClpX, both hexameric ring assemblies, recognize short-terminal

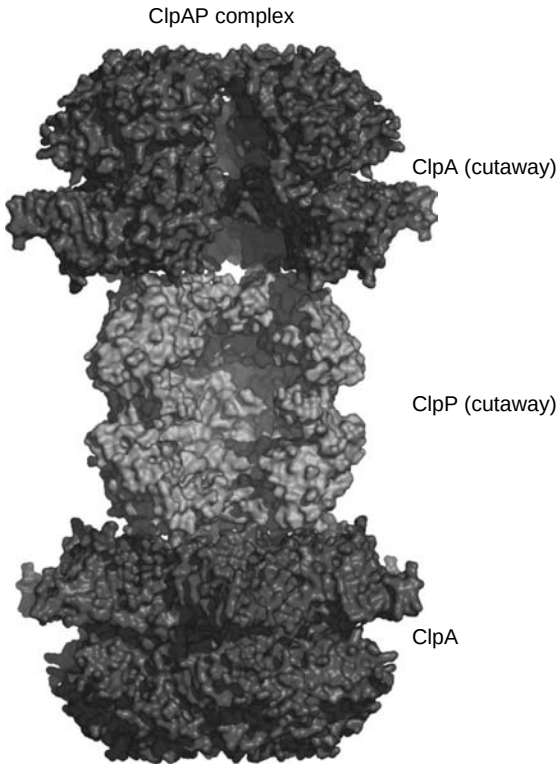

Fig. 17 ClpA–ClpP complex, illustrating the central channel pathway taken by substrate proteins as they are unfolded by the hsp100 chaperone ClpA (green) in association with ATP hydrolysis and translocated into the protease ClpP (blue), where they are degraded. Red regions inside ClpA designate loop segments at the level of the D2 ATPase domain of each identical subunit that has recently been shown to be involved with both binding and ATP-driven translocation of substrate proteins. Yellow inside ClpP designates the serine proteolytic active sites. Model generated by docking of a hexameric model of ClpA derived from the crystal structure of a monomer (1KSF) (employing the structure of the intact p97 hexamer as a guide – 1OZ4), with the crystallographic model of ClpP (1TYF). (See color plate p. xxxix.)

amino acid sequences of proteins to be unfolded and degraded. For example, an 11 amino acid sequence, ssrA, is attached to the C terminus of nascent chains that become arrested in translation at the ribosome, and this sequence is specifically recognized by both ClpA and ClpX. In the latter case, a "delivery" agent known as SspB can bind the ssrA-tagged substrate at the ribosome, lodging it in a relatively extended shallow groove, and acts to deliver the substrate to ClpX. Terminal recognition elements like ssrA appear to be recognized by the hsp100 cylinders themselves, either by flexible (usually N terminal) domains at the top of the cylinders or by loop elements inside their central channels. Delivery agents like SspB are recognized by the flexible N-domains, and in binding them, may serve to deliver the tag elements directly to the central channel recognition elements. The action of ATP hydrolysis then appears to drive pulling movements associated with translocation of substrates, tagged end first, down the axial channels, associated with unraveling and unfolding of the substrate protein's tertiary structure.

In eukaryotes, the system of recognition of proteins to be unfolded principally involves the attachment of the 79 amino acid moiety known as *ubiquitin* to either the N terminus or to lysine side chains of the target protein, mediated by a system of conjugating enzymes (E2s) and ligases (E3s). Once a protein is multiubiquitinated, it is efficiently bound by the 19S "cap" structure of the

proteasome, which subsequently employs a six-membered ring of ATPases at its base, contiguous with the 20S core protease particle, to mediate unfolding of the recognized protein, most likely using the same mechanism as that being defined for the prokaryotic AAA+ unfoldases just described.

6.2 Proteolytic Components

On the part of the proteolytic cylinders, unfolded polypeptides arriving from the coaxially associated unfoldases see a local concentration of protease active sites that is molar in concentration, assuring efficient cleavage by active sites lining the protease cavity walls. A variety of chemical strategies for peptide bond cleavage are utilized by the proteasome, but one that is novel, employed by the class of Ntn hydrolases including HslV, involves N-terminal threonine, acting as nucleophile via its side-chain hydroxyl group, attacking the carbonyl of the scissile peptide bond. This mechanism acts both to remove the propeptide from the proteasome subunit precursor and also to cleave bonds within substrates. In the latter case, the free amino group of the subunit strips the proton from the hydroxyl group of the active threonine side chain, activating the nucleophile. By contrast, in the case of ClpP, a serine proteolytic site is employed, with a classic Ser-His-Asp triad, but one positioned in a cleft formed between two β-strands, one in a "head" structure and the other in a "handle" of each subunit, the site facing the central cavity of the protease.

It is of interest that many of the prokaryotic unfoldases are composed of six-membered rings, as is the base of the proteasome, while the proteolytic cylinders with which these components associate are composed of seven-membered rings. Such symmetry mismatch is intriguing, potentially allowing for sliding or rotational movements of one component relative to the other, or allowing for modulation of the affinity of the components for each other. Notably, in the case of the ClpA and ClpX proteins, it is loop elements at the bottom aspect of their subunits that make contacts with a continuous ring of hydrophobic surface in the end surface of ClpP to mediate interaction between these components. Yet symmetry mismatch does not seem to be universal amongst these complexes as, for example, HslU–HslV are hexamer–hexamer assemblies, and other proteases, like Lon, already contain a chaperone moiety built within each of the proteolytically active subunits (see the following section).

6.3 HslU–HslV

HslV is also a double ring protease, composed of six-membered rings, whose active sites, containing a catalytic N-terminal threonine, most closely resemble those of the eukaryotic proteasome. Here the associating chaperone "unfoldase," HslU, also an AAA+ hexameric ring, aligns subunit to subunit with HslV. Each subunit of HslU exposes at its open end a flexible antiparallel segment known as an *I domain*. These domains, emanating from the ATPase domains of the subunits, collectively appear as tentacles that seem likely to be involved with capturing substrate proteins and potentially transferring them into the axial passageway of HslUV. Crystallographic studies show that association of HslU with HslV leads to allosteric adjustments at the

HslV active sites that serve to activate the protease activity of HslV.

6.4 Lon

Lon protease appears to function similar to the ClpP and HslV systems, except that each subunit of this hexameric ring assembly contains built into it an N-terminal domain, an AAA+ unfoldase domain, and a protease domain. While these elements are covalently linked, the sequence of events of recognition of substrate, translocation and unfolding, and degradation, seem likely to be the same. The protease domain has recently been crystallized, revealing an axial pore of 18-Å diameter giving rise to a distal concave surface in which is situated a serine-lysine active site dyad flanked by residues including a threonine which may functionally resemble the Aspartate in Ser-His-Asp triads and NHs that may stabilize the oxyanion. Interestingly, a similar X-ray structural analysis of an archaeal lon protease domain (*M. janaschii*) revealed a more classic Ser-Lys-Asp triad.

7 Disaggregases – Prokaryotic ClpB and Eukaryotic Hsp104

The ClpX and ClpA proteins have relatives, ClpB and hsp104, which are strongly heat-induced components of the bacterial cytoplasm, and yeast and plant cytosols, respectively, with architecture resembling that of ClpA, a ring of six identical subunits with a stacked domain structure of N-terminal flexible domains at the top of the cylinder and two AAA+ ATPase domains. Here, however, there is no protease partner, and there is also an additional striking structural feature; long coiled-coil "propeller" domains emanating from the first AAA+ domain at the outside of each of the six subunits at the midsection of the cylinder. These proteins have a unique ability to dissociate protein aggregates in the presence of the hsp70 (DnaK) system and ATP. The nature of such cooperation and role of the coiled-coil elements remain unclear, but it seems that hsp70 may be involved with enabling aggregated subunits to be bound or threaded through the central channel of ClpB, which thus unfolds them and allows them a chance to properly refold in the presence of the hsp70 and hsp60 systems.

8 Periplasmic Proteolytic System – HtrA Proteases DegP and DegS

Regulated proteolytic steps like those carried out by the Clp system, the proteasomal system, HslUV, and Lon occur in cellular locations that contain millimolar concentrations of ATP, needed to drive the unfolding and translocation of substrate proteins mediated by the AAA+ moieties of these proteases. Yet the cell has apparent need for the same strategy of regulated proteolysis in providing protein quality control to compartments that do not have access to ATP, for example, the bacterial periplasm and mitochondrial intermembrane space. In these compartments, it appears that a similar protease system is employed, preserving a strategy of caging the active sites inside an oligomeric assembly, but with substrate access not dependent on ATP. This family of proteases, called the *HtrA proteases* (high temperature requirement), is comprised of subunits, each containing a serine proteolytic domain and one or two C-terminal PDZ domains.

8.1 DegP

The notion that chaperone and protease activities are brought together in HtrA components, as with the ATP-dependent ones, is supported by structure and functional studies of DegP and of the more specialized membrane-anchored DegS. In functional studies of DegP, it was observed that at low temperatures this component has little or no proteolytic activity, yet it exhibited ability both *in vivo* and *in vitro* to assist the folding to native form of such periplasmic enzymes as malS. At heat shock temperatures, however, DegP protease activity became activated as observed *in vitro*, and recovery of native malS was reduced *in vivo*, as the apparent result of proteolytic turnover of molecules whose proper folding did not occur at the higher temperature. X-ray structural study of DegP followed, providing indications of how the machine might be functioning (Fig. 18). DegP is a hexamer composed of two oppositely positioned and loosely connected tight trimers of protease domains, with the PDZ domains packed loosely at the outside aspect forming mobile side-walls that can serve as gates for substrate entry into a central cavity. The oppositely positioned protease trimers are connected to each other by three pillars of polypeptide loop segment extending from each protease domain into the active site of a protease domain in the opposite ring. Two molecular forms were observed in the crystals, an open one with lateral openings between the pillars into the proteolytic cavity, wherein the PDZ domains lie at the lateral aspects of the trimers, and a closed form where the PDZ domains lie at the equatorial aspect blocking access to the cavity. These geometries further support that it is lateral gating that is involved with entry into the DegP cavity. The two PDZ domains themselves differed in character, one with an open groove, available to accept a segment of polypeptide as an added strand between a resident one and α-helix as in canonical PDZ motifs, and the other with an already occupying resident segment that may be mobilized under conditions of substrate binding but which could then conceivably displace substrate in the context of entry into the protease. The proteolytic cavity itself, ~15 Å in height and ~45 Å in diameter contains patches of hydrophobic surface at its ceiling (and floor), which may be involved with binding nonnative substrate entering the central cavity.

Concerning the protease sites, in the crystals obtained, transferred to 4°, the protease would be expected to be inactive and, consistently, the serine proteolytic sites, as classic trypsin-like sites formed between two orthogonal β-barrels, were indeed found to be distorted, with twists of the active site loops distorting the catalytic triad, oxyanion hole, and S1 specificity pocket. The catalytic sites were also blocked. The molecular nature of the thermal switch in DegP thus remains to be resolved, but there must be significant readjustment of the loops both extending from one trimeric ring into the other, and also locally of the active site loops. Concerning the nature of recruitment to the protease sites, it seems that either already unfolded substrates are recruited by the PDZ domains through the cavity wall into the proteolytic central cavity or that PDZ action and threading into the cavity mediates an unfolding action that favors proteolysis. Perhaps the PDZ domains can exert a pulling action akin to that suggested for the loops in the central channel of ClpA.

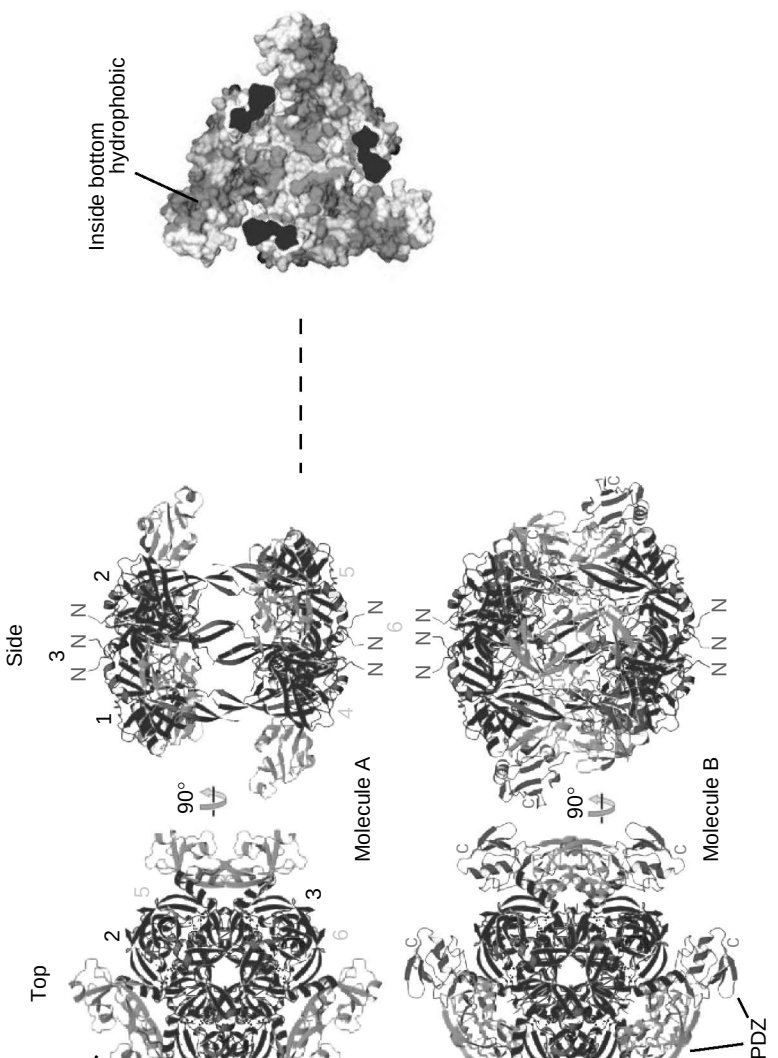

Fig. 18 Structure of the *E. coli* periplasmic protease DegP. The protease is composed of two trimeric protease rings facing each other (green) interconnected by struts of β-strand that enter the protease active sites and regulate activity (see text). The protease domains are flanked by PDZ domains (yellow and orange) that are involved with capture and perhaps unfolding of substrate proteins. The substrates are then transferred in an ATP-independent fashion through the sidehole aspects of the assembly into the proteolytic cavity, the inside of which contains hydrophobic surface at its terminal inside aspects, shown in yellow, likely involved with "guiding" substrate into the serine proteolytic sites. Figure taken from Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M., Clausen, T. (2002) Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine, *Nature* **416**, 455–459, pdb 1KY9. (See color plate p. xliv.)

8.2 DegS

In the case of the related DegS, additional glimpses of regulatory behavior of this class of component have been possible. Here, the serine proteolytic active sites were also found to be structurally nonorganized in the absence of substrate protein, but when a C-terminal peptide presented by unfolded Omp proteins became bound to the PDZ domains, adding a β-strand to one edge of the β-sheet of the PDZ domain, then the DegS active sites became reorganized into the active Ser-His-Asp triad. Thus, based on the behavior of DegS, it seems even more apparent that the PDZ domains are involved with substrate recognition, and that such recognition can directly influence the conformation of the proteolytic sites at least in DegS.

While DegP and DegS have very similar structural arrangements, their physiological roles differ considerably. DegP is a soluble periplasmic protease that appears to have broad substrate reactivity while DegS has a dedicated role in sensing specifically the free C termini of Omp proteins in the periplasmic space, termini that are normally buried but made specifically available by protein misfolding. Such binding, as described in Fig. 3, activates the protease of DegS, enabling cleavage of the transmembrane protein, RseA, followed by a second cleavage event mediated by YaeL, which leads to release of σE, the periplasmic stress-responsive sigma factor, from RseA, enabling transcriptional activation of periplasmic heat shock components such as DegP.

8.3 Omi

Finally, a mammalian homolog of DegP (Omi), residing in the mitochondrial intermembrane space, has been observed to be affected in a recessively inherited mouse motor neurodegeneration mutant known as *mnd2*. Purified mutant protein was devoid of proteolytic activity against casein. Omi appears to have a role in apoptosis, as it is released along with cytochrome *c* during this process, and it can degrade the class of proteins that are inhibitors of apoptosis (IAPs). This degradation activity was also defective. Interestingly, however, deletion of the PDZ domain from the mutant protein partially restored activity to the mutant, suggesting that the mutation, involving one of the large active site loop structures, affects substrate access to the active site. *In vivo*, loss of the protease activity of Omi increased the susceptibility of mitochondria in affected fibroblasts to a permeability transition of the membranes triggered by Ca^{+2}, associated with loss of membrane potential and with release of proapoptotic proteins. Thus, loss of a quality control function of this protease, particularly in striatal neurons, may lead to induction of the permeability transition, both reducing mitochondrial respiratory function and triggering release of apoptotic proteins leading to cell death. Consistently, mitochondrial disintegration is observable early during neuronal degeneration.

Acknowledgment

We thank Wolfgang Rist, Christian Schlieker, and George Farr for figure production. Richard Morimoto acknowledges support from NIH, HD Society, ALS Association, and Daniel F and Ada L Rice Foundation.

See also Proteasomes; Protein and Nucleic Acid Enzymes; Ubiquitin-Proteasome System for Controlling Cellular Protein Levels.

Bibliography

Books and Reviews

Alba, B., Gross, C. (2004) Regulation of the *Escherichia coli* σ^E–dependent envelope stress response, *Mol. Microbiol.* **52**, 613–619.

Caplan, A.J., Jackson, S., Smith, D. (2003) Hsp90 reaches new heights, *EMBO Rep.* **4**, 126–130.

Clausen, T., Southan, C., Ehrmann, M. (2002) The HtrA family of proteases: Implications for protein composition and cell fate, *Mol. Cell* **10**, 443–455.

Duguay, A., Silhavy, T. (2004) Quality control in the bacterial periplasm, *Biochim. Biophys. Acta* **1694**, 121–134.

Fenton, W.A., Horwich, A.L. (2003) Chaperonin-mediated protein folding: fate of substrate polypeptide, *Q. Rev. Biophys.* **36**, 229–256.

Gross, C. (1996) Function and Regulation of the Heat Shock Proteins, in: Neidhardt, F.C. (Ed.) *E. coli and Salmonella*, 2nd edition, ASM Press, pp. 1382–1399.

Hartl, F.U., Hayer-Hartl, M. (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein, *Science* **295**, 1852–1858.

Horwich, A.L., Fenton, W.A. (2004) Role of ATP in GroEL-GroES-assisted polypeptide folding, in: Hackney, D.D., Tamanoi, F. (Eds.) *The Enzymes*, Vol. XXIII, Academic Press, New York, pp. 399–433.

Kaufman, R.J. (2002) Orchestrating the unfolded protein response in health and disease, *J. Clin. Invest.* **110**, 1389–1398.

Mayer, M.P. (2004) Timing the catch (N&V), *Nat. Struct. Biol.* **11**, 6–8.

Mayer, M.P., Bukau, B. (2004) Regulation of Hsp70 Chaperones by Co-chaperones, in: Buchner, J., Kiefhaber, T. (Eds.) *Protein Folding Handbook*, Part II, Chap. 15, Wiley-VCH, Weinheim, pp. 508–548.

Morimoto, R.I. (1998) Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators, *Genes Dev.* **12**, 3788–3796.

Pearl, L.H., Prodromou, C. (2002) Structure, function, and mechanism of the Hsp90 molecular chaperone, *Adv. Protein Chem.* **59**, 157–186.

Pirkkala, L., Nykanen, P., Sistonen, L. (2001) Roles of the heat shock transcription factors in regulation of the heat shock response and beyond, *FASEB J.* **15**(7), 1118–1131.

Ryan, M.T., Pfanner, N. (2002) Hsp70 proteins in protein translocation, *Adv. Protein Chem.* **59**, 223–242.

Sauer, R.T., et al. (2004) Sculpting the proteome with AAA+ proteases and disassembly machines, *Cell* **119**, 9–18.

Schröder, M., Kaufman, R.J. (2005) ER stress and the unfolded protein response, *Mutat. Res.* **569**, 29–63.

Shi, Y., Mosser, D.D., Morimoto, R.I. (1998) Molecular chaperones as HSF1-specific transcriptional repressors, *Genes Dev.* **12**, 654–666.

van Montfort, R., Slingsby, C., Vierling, E. (2002) Structure and function of the small heat shock protein/α-crystallin family of molecular chaperones, *Adv. Protein Chem.* **59**, 105–156.

Wu, C. (1995) Heat shock transcription factors: structure and regulation, *Annu. Rev. Cell Dev. Biol.* **11**, 441–469.

Yura, T., Kanemori, M., Morita, M. (2000) The Heat Shock Response: Regulation and Function, in: Storz, G., Hengge-Aronis, R. (Eds.) *Bacterial Stress Responses*, ASM Press, Washington, DC, pp. 3–18.

Primary Literature

Abravaya, K., Myers, M.P., Murphy, S.P., Morimoto, R.I. (1992) The human heat shock protein hsp70 interacts with HSF, the transcription factor that regulates heat shock gene expression, *Genes & Dev.* **6**, 1153–1164.

Bochtler, M., Ditzel, L., Groll, M., Huber, R. (1997) Crystal structure of heat shock locus V (HslV) from Escherichia coli, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6070–6074.

Boorstein, W.R., Craig, E.A. (1990) Transcriptional regulation of SSA3, an HS70 gene from *Saccharomyces cerevisiae*, *Mol. Cell. Biol.* **10**, 3262–3267.

Botos, I., et al. (2004) The catalytic domain of *Escherichia coli* Lon protease has a unique fold

and a ser-lys dyad in the active site, *J. Biol. Chem.* **279**, 8140–8148.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D.C., Joachimiak, A., Horwich, A.L., Sigler, P.B. (1994) The crystal structure of the bacterial chaperonin GroEL at 2.8 Å, *Nature* **371**, 578–586.

Bügl, H., Fauman, E.B., Staker, B.L., Zheng, F., Kushner, S.R., Saper, M.A., Bardwell, J.C.A., Jakob, U. (2000) RNA methylation under heat shock control, *Mol. Cell* **6**, 349–360.

Bukau, B., Horwich, A.L. (1998) The Hsp70 and Hsp60 chaperone machines (review), *Cell* **92**, 351–366.

DeLaBarre, B., Brunger, A.T. (2003) Complete structure of p97/valosin-containing protein reveals communication between nucleotide domains, *Nat. Struct. Biol.* **10**, 856–863.

Flaherty, K.M., Deluca-Flaherty, C., McKay, D.B. (1990) Three dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein, *Nature* **346**, 623–628.

Greene, M.K., Maskos, K., Landry, S.J. (1998) Role of the J-domain in the cooperation of Hsp40 with Hsp70, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6108–6113.

Guo, F., Maurizi, M.R., Esser, L., Xia, D. (2002) Crystal structure of ClpA, an Hsp100 chaperone and regulator of ClpAP protease, *J. Biol. Chem.* **277**, 46743–46752.

Harris, S.F., Shiau, A.K., Agard, D.A. (2004) The crystal structure of the carboxy-terminal dimerization domain of htpG, the Escherichia coli Hsp90, reveals a potential substrate binding site, *Structure* **12**, 1087–1097.

Harrison, C.J., Hayer-Hartl, M., Di Liberto, M., Hartl, F.U., Kuriyan, J. (1997) Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK, *Science* **276**, 431–435.

Hsu, A.-L., Murphy, C.T., Kenyon, C. (2003) Regulation of aging and age-related disease by DAF-16 and heat-shock factor, *Science* **300**, 1142–1145.

Hunt, J.F., Weaver, A.J., Landry, S.J., Gierasch, L., Deisenhofer, J. (1996) The crystal structure of the GroES co-chaperonin at 2.8 Å resolution, *Nature* **379**, 37–45.

Jakob, U., Muse, W., Eser, M., Bardwell, J.C.A. (1999) Chaperone activity with a redox switch, *Cell* **96**, 341–352.

Janda, I., Devedjiev, Y., Derewenda, U., Dauter, Z., Bielnicki, J., Cooper, D.R., Graf, P.C.F., Joachimiak, A., Jakob, U., Derewenda, Z.S. (2004) The crystal structure of the reduced, Zn^{+2}-bound form of the B subtilis Hsp33 chaperone and its implications for the activation mechanism, *Structure* **12**, 1901–1907.

Jones, J.M., et al. (2003) Loss of Omi mitochondrial protease activity causes the neuromuscular disorder of *mnd2* mutant mice, *Nature* **425**, 721–727.

Jurivich, D.A., Sistonen, L., Kroes, R.A., Morimoto, R.I. (1992) Effect of sodium salicylate on the human heat shock response, *Science* **255**, 1243–1245.

Kim, K.K., Kim, R., Kim, S.H. (1998) Crystal structure of a small heat shock protein, *Nature* **394**, 595–599.

Korber, P., Stahl, J.M., Nierhaus, K.H., Bardwell, J.C.A. (2000) Hsp15: a ribosome-associated heat shock protein, *EMBO J.* **19**, 741–748.

Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M., Clausen, T. (2002) Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine, *Nature* **416**, 455–459.

Lee, S., Sowa, M.E., Watanabe, Y.-H., Sigler, P.B., Chiu, W., Yoshida, M., Tsai, F.T.F. (2003) The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state, *Cell* **115**, 229–240.

Ozcan, U., Cao, Q., Yilmaz, E., Lee, A.-H., Iwakoshi, N.N., Ozdelen, E., Tuncman, G., Görgün, C., Glimcher, L.H., Hotamisligil, G.S. (2004) Endoplasmic reticulum stress links obesity, insulin action, and type 2 diabetes, *Science* **306**, 457–461.

Rutherford, S.L., Lindquist, S. (1998) Hsp90 as a capacitor for morphological evolution, *Nature* **396**, 336–342.

Rye, H.S., Roseman, A.M., Furtak, K., Fenton, W.A., Saibil, H.R., Horwich, A.L. (1999) GroEL-GroES cycling: ATP and non-native polypeptide direct alternation of folding-active rings, *Cell* **97**, 325–338.

Sha, B., Lee, S., Cyr, D.M. (2000) The crystal structure of the peptide-binding fragment from the yeast Hsp40 protein Sis1, *Structure* **8**, 799–807.

Szyperski, T., Pellecchia, M., Wall, D., Georgopoulos, C., Wüthrich, K. (1994) NMR structure determination of the *Escherichia coli* DnaJ molecular chaperone: secondary structure and backbone fold of the N-terminal region (residues 2–108) containing the highly

conserved J domain, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11343–11347.

Sousa, M.C., Trame, C.B., Tsuruta, H., Wilbanks, S.M., Reddy, V.S., McKay, D.B. (2000) Crystal and solution structures of an HslUV protease-chaperone complex, *Cell* **103**, 633–643.

Spiess, C., Beil, A., Ehrmann, M. (1999) A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein, *Cell* **97**, 339–347.

Travers, K.J., Patil, C.K., Wodicka, L., Lockhart, D.J., Weissman, J.S., Walter, P. (2000) Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation, *Cell* **101**, 249–258.

van Montfort, R.L.M., Basha, E., Friedrich, K.L., Slingsby, C., Vierling, E. (2001) Crystal structure and assembly of a eukaryotic small heat shock protein, *Nat. Struct. Biol.* **8**, 1025–1030.

Walsh, N.P., Alba, B.M., Bose, B., Gross, C.A., Sauer, R.T. (2003) OMP peptide signals initiate the envelope stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain, *Cell* **113**, 61–71.

Wang, J., Hartling, J.A., Flanagan, J.M. (1997) The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis, *Cell* **91**, 447–456.

Weibezahn, J., et al. (2004) Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB, *Cell* **119**, 653–665.

Wilken, C., Kitzing, K., Kurzbauer, R., Ehrmann, M., Clausen, T. (2004) Crystal structure of the DegS stress sensor: how a PDZ domain recognizes misfolded protein and activates a protease, *Cell* **117**, 483–494.

Winter, J., Linke, K., Jatzek, A., Jakob, U. (2005) Severe oxidative stress causes inactivation of DnaK and activation of the redox-regulated chaperone Hsp33, *Mol. Cell* **17**, 381–392.

Xu, Z., Horwich, A.L., Sigler, P.B. (1997) The crystal structure of the asymmetric GroEL-GroES-$(ADP)_7$ chaperonin complex, *Nature* **388**, 741–749.

Zhu, X., Zhao, X., Burkholder, W.F., Gragerov, A., Ogata, C.M., Gottesman, M., Hendrickson, W.A. (1996) Structural analysis of substrate binding by the molecular chaperone DnaK, *Science* **272**, 1606–1614.

Zou, J., Guo, Y., Guettouche, T., Smith, D.F., Voellmy, R. (1998) Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1, *Cell* **94**, 471–480.

Stroke – Epidemiology, clinical overview and model systems: *see* Pathophysiology of Stroke

Structure, Function and Biogenesis of Peroxisomes

Joel M. Goodman
University of Texas Southwestern Medical Center at Dallas, Dallas, TX

1 Introduction 609

2 Common Structural and Functional Motifs 610

3 Specific Functions of Peroxisomes among Organisms 614
3.1 Peroxisomes in Fungi 614
3.2 Peroxisomes in Trypanosomatids 618
3.3 Peroxisomes in Plants 619
3.4 Peroxisomes in Mammals 620

4 Biogenesis of Peroxisomes 622
4.1 Origin of the Peroxisome 622
4.2 Peroxisome Fission 622
4.3 The Origin and Assembly of the Peroxisomal Membrane 623
4.4 Import of Proteins into Peroxisomes 624

5 Peroxisomal Motility 627

6 Peroxisomal Diseases 628
6.1 Single Enzyme Disorders 628
6.2 Peroxisome Biogenesis Disorders 629
6.3 Mouse Models for PBDs 630

Bibliography 631
Books and Reviews 631
Primary Literature 631

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

Keywords

Extended Shuttling Model
The hypothesis for protein import into peroxisomes whereby the PTS receptor enters the matrix where dissociation from cargo occurs.

Glycosome
The peroxisome in trypanosomatids, which houses several enzymes of the glycolytic pathway.

Glyoxysome
A subclass of peroxisomes in which the conversion of fatty acids to Krebs cycle intermediates is a key function.

Microbody
The original term for the peroxisome; still used as the morphological equivalent.

mPTS
The targeting information in peroxisome integral membrane proteins.

PBD
Peroxisomal biogenesis disorder. A patient afflicted by a PBD has a defective PEX gene.

Peroxin
A product of a PEX gene.

Peroxisome
An organelle, 0.2 to 0.5 μm in diameter, distinguished by a single membrane, a dense matrix, and common biosynthetic features; most are ovoid or spherical, contain flavin oxidases and catalase, and participate in lipid metabolism; many contain crystalline cores.

PEX Gene
A gene encoding a protein that functions in peroxisomal biogenesis. Before 1996, when nomenclature was unified, such genes were also known as *PAS*, *PEB*, or *PER*.

PTS
Peroxisomal targeting signal. Most matrix proteins contain a PTS1, a tripeptidyl sequence of which SKL is the prototype, at their extreme carboxy termini. A few others contain a PTS2 at or near their amino terminus.

Zellweger Syndrome
The most serious of the PBDs, characterized by facial dysmorphism, calcific stippling of cartilage, liver and renal disease, and nerve demyelination.

Peroxisomes comprise a family of organelles that are found in virtually every eukaryotic cell; they contain a single membrane enclosing a dense matrix. Crystalloid cores may also be present. While the functions of peroxisomes in some organisms and tissues can be highly specialized, most peroxisomes perform crucial reactions in the biogenesis and degradation of lipids. Many metabolic reactions within peroxisomes are catalyzed by flavin oxidases that generate hydrogen peroxide (thus the name "peroxisome"); catalase and other antioxidants exist in the matrix to protect the cell from oxidative damage. While isolated cells can survive without peroxisomes, animals and plants require them. Peroxisomes provide substrates for several developmental programs. Defective peroxisomes cause serious disease and early death in humans. Peroxisomes can arise from fission of preexisting organelles, as well as from a less understood *de novo* pathway that originates in the endoplasmic reticulum. Most matrix proteins fold and assemble in the cytosol and then are targeted to peroxisomes via chaperones and shuttling receptors. Great progress has been made in understanding protein import into peroxisomes, but important mysteries remain.

1 Introduction

A novel organelle was identified by Rhodin in electron micrographs of mouse kidney brush border in 1954. *Microbodies*, as they were termed, consisted of a single membrane surrounding a finely granular matrix and were 0.5 μm or smaller in diameter. Similar organelles in liver parenchymal cells, except those also containing crystalloid cores, were reported two years later by Rouiller and hypothesized to be mitochondrial precursors. Concurrent with these observations, the laboratories of de Duve and Baudhuin began the biochemical purification of lysosomes, particles that contained hydrolytic enzymes. Other enzymes, namely, catalase, uric acid oxidase (uricase), and a D-amino acid oxidase were initially fractionated with the hydrolytic enzymes but later could be purified away from them. In 1960, these nonlysosomal enzymes were ascribed to microbodies and shortly thereafter, in 1965, de Duve proposed a new name for the particle, "peroxisome," in recognition of hydrogen peroxide, the metabolic intermediate between the oxidases and catalase.

In the next decade, particles of similar morphology were found in many diverse cell types. In trypanosomes, an organelle was identified that contained, unexpectedly, several enzymes of the glycolytic pathway and thus were named *glycosomes*. In plant seedlings and in filamentous fungi, a different organelle, termed the *glyoxysome*, was described that contained enzymes of the glyoxylate cycle, in which acetyl-CoA could be converted to succinate, thereby linking the product of fatty acid oxidation with carbohydrate metabolism. In methylotrophic yeasts, microbodies were described that could proliferate to a huge extent, virtually hijacking the cytoplasm; these organelles performed the initial steps of methanol assimilation. Since these diverse organelles performed such different functions, it was not at all clear that they were related. This issue was finally

resolved after Gould, Subramani, and colleagues reported the first peroxisomal targeting signal, and it became clear that this motif was shared among matrix proteins of microbodies from all of these sources. The close relationship among these structures was confirmed as similar proteins involved with peroxisomal biogenesis were identified in all of these cell types.

The origin of peroxisomes has been a fascinating area of research since their identification. The first peroxisome may have been a prokaryotic endosymbiont that gave the host a selective advantage in an oxygen-dominated environment, but the lack of a genome and the ancient nature of this event, if it occurred at all, may make this hypothesis impossible to validate or disprove. Early electron microscopic analysis suggested that peroxisomes arose as buds from the endoplasmic reticulum. By 1980, however, it was clear that peroxisomes can arise by fission, and it was thought that a peroxisome was necessary for the generation of new organelles. The pendulum has now settled midway between these theories: the consensus at present is that peroxisomes can develop both from preexisting peroxisomes and from elements (of mysterious identity at present) of the endoplasmic reticulum.

2 Common Structural and Functional Motifs

Typical peroxisomes in electron micrographs are spherical or ovoid with a diameter of 0.2 to 0.5 μm. As a general rule, peroxisomes distribute evenly throughout the cytosol and appear as uniform puncta by optical microscopy using fluorescent or antibody tags. Yeast contain a dozen or fewer puncta, mammalian cells contain several hundred. They are sometimes clustered, perhaps as a result of recent fission events, and in yeast they may concentrate at the periphery. Young peroxisomes may have a more complex reticulated structure with internal membranes. Serial sectioning of mammalian liver or sebaceous glands, sources that have very active peroxisomal lipid metabolism, reveal that peroxisomes can be connected to each other by tubules, forming an extensive peroxisomal reticulum. Whether such tubules could also represent endoplasmic reticulum has been a contentious issue since the 1960s, although recent strong evidence in favor of this link has been provided. Besides a relationship to the endoplasmic reticulum, peroxisomes have been noted to be in close apposition to mitochondria and lipid bodies, and evidence has been put forward to support hypotheses of peroxisomal origins from each of these compartments.

The matrix is usually more electron dense than the surrounding cytoplasm. Peroxisomes of many organisms, from fungi to mammals (but not humans), contain a core of particularly dense protein within the matrix that is organized in a paracrystalline matrix. These can allow easy discrimination of peroxisomes from other structures. Cores can be cuboidal, plate-like, or more amorphous. The core can govern organellar shape. In the methylotroph *Hansenula polymorpha*, for example, peroxisomes are cuboidal, the membrane wrapping around the large core of alcohol oxidase (Fig. 1). The regular pattern of crystalloids represents a single protein, although others may penetrate the core in a more soluble state. The ground substance of the matrix usually appears uniform and granular, although areas of fibrillar structure have been described for mammalian tissue. The matrix, even disregarding the core, is not a uniform

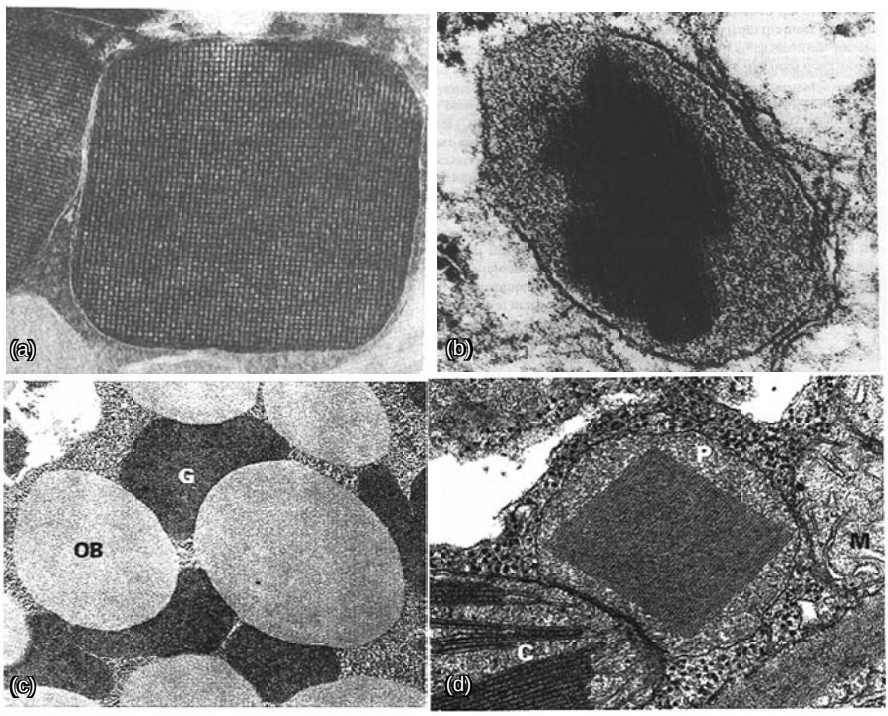

Fig. 1 Morphology of peroxisomes from four sources (transmission electron microscopy. (a) Hansenula polymorpha grown on methanol. The cuboidal shape is caused by the alcohol oxidase crystalloid. Source: van Dijken et al. (1982) *Ann. N. Y. Acad. Sci.* **386**, 200–216. (b) Rat. Crystalloid is composed of urate oxidase. Source: Masters, C., Crane, D. (1995) *The Peroxisome: A Vital Organelle*, Cambridge University Press, Cambridge, p. 9. (c) Glyoxysome (G) next to oil body (OB) from germinating tomato seed. Source: Buchanan, B.B., Gruissem, W., Jones, R. (Eds.) (2000) Biochemistry & Molecular Biology of Plants, American Society of Plant Physiologists, Rockville, MD, p. 35. (d) Tobacco leaf. Peroxisome (P) next to chloroplast (C). Buchanan, B.B., Gruissem, W., Jones, R. (Eds.) (2000) Biochemistry & Molecular Biology of Plants, American Society of Plant Physiologists, Rockville, MD, p. 33.

mixture of proteins; as peroxisomes are gently lysed, some matrix proteins leak much more readily than others; the less soluble proteins are presumed to be bound to the membrane, to the core, or to other less apparent structural components.

As noted above, a hallmark of peroxisomes is the presence of oxidases (Fig. 2). Typical substrates include lipids, alcohols, purines, and amino acids. These enzymes are responsible not only for the disposition of the end products of metabolism but also to generate important molecules for signal transduction, intermediates in assimilatory pathways, or substrates for specialized functions in the same cells or for remote tissues of the organism. Besides oxidative reactions, peroxisomes catalyze steps in many other pathways. The packaging of enzymes within peroxisomes may serve two important functions. Segregation of peroxisomal oxidative enzymes with catalase, other peroxidases, or superoxide dismutase certainly protects the cytoplasm

from oxidative stress. (Interestingly, reactive oxygen species that escape from peroxisomes can perform useful functions under certain conditions.) The other result of sequestering peroxisomal enzymes may be to prevent "metabolic interference"; their presence in the cytoplasm may initiate energetically wasteful futile cycles of metabolites. Thus, while the initial segregation of peroxisomal proteins may have been the result of an ancient microbial symbiosis, the host cells seem to have parlayed this relationship to optimal advantage.

Many peroxisomal enzymatic reactions are performed in other compartments as well. This is the case for β-oxidation of fatty acids (shared by the mitochondria in mammals), the glyoxylate cycle (shared by the cytoplasm in yeast), or sterol biosynthesis (shared by the endoplasmic reticulum in animals). In some cases, this is explained by a requirement of an enzymatic step in two distinct pathways that must be segregated from each other in two compartments to maximize flux. Often, specialization has occurred in the two compartments. For example, the oxidation of very long versus medium-length fatty acids is accommodated by peroxisomes and mitochondria, respectively, in mammalian tissues. In this case, two distinct sets of enzymes perform largely similar reactions in the two organelles. Often, however, gene products are dually localized (Table 1). Proteins that are colocalized to mitochondria and peroxisomes have the potential for an amino terminal mitochondrial-targeting signal, as well as a peroxisomal signal at the carboxy terminus. The mitochondrial signal may be absent in some molecules because of alternative splicing or initiation at an alternative translational start site. In several cases, however, the identical translation product is dually localized. A weak peroxisomal targeting signal (i.e. straying from the consensus sequence) is often but not always

Tab. 1 Some dually localized peroxisomal enzymes.[a]

Enzyme	*Species*	*Other location*	*Comments*
Epoxide hydrolase	Rat	Cytosol	PTS1 (SKI) inefficiently targeted?
Vanillyl-alcohol oxidase	*Penicillium simplicissimum*	Cytosol	PTS1 (WKL) inefficiently targeted?
HMG-CoA ligase	Human/mouse	Mitochondria	Cryptic mitochondrial-targeting sequence
2-methylacyl-CoA racemase	Human/rodent	Mitochondria	Poor PTS1 (KASL or KANL)
Carnitine acetyltransferase	*Candida tropicalis*	Mitochondria	Different transcription start sites
$\Delta^{3,5};\Delta^{2,4}$-dienoyl-CoA isomerase	Rat	Mitochondria	Peroxisomal protein has truncated amino terminus
Cu, Zn SOD	Mammals	Mitochondria /cytosol/lysosome?	Cryptic mitochondrial-targeting sequence

[a] Does not include peroxins known to shuttle from cytosol to peroxisomes.

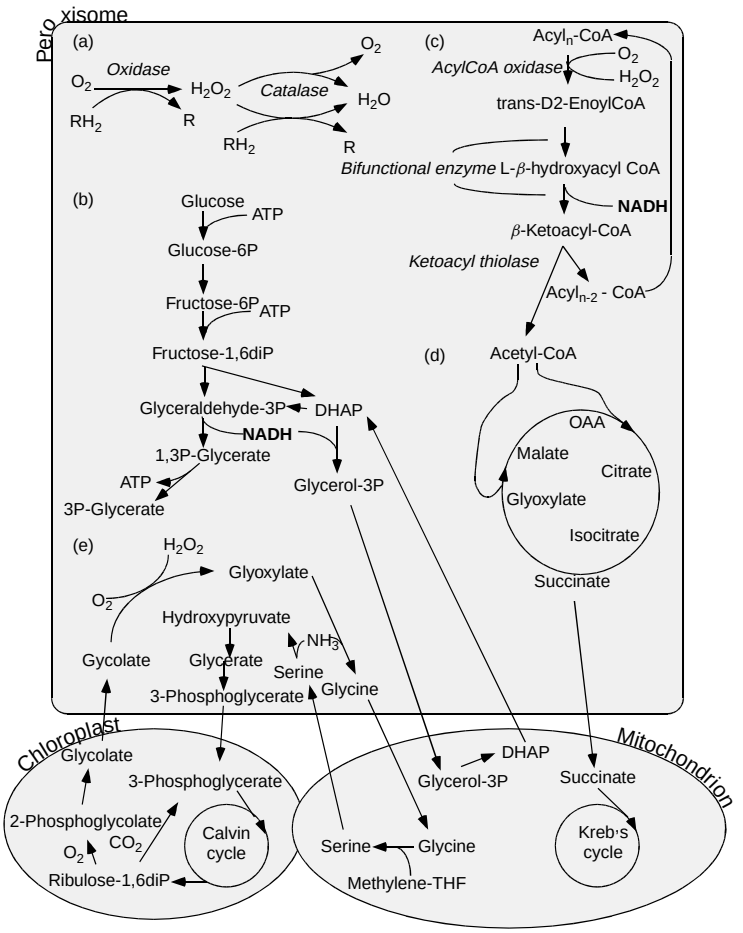

Fig. 2 Examples of peroxisome metabolic pathways. (a) The prototypic peroxisomal oxidase reaction and detoxification of hydrogen peroxide by catalase. (b) Glycolytic pathway in glycosomes. DHAP (dihydroxyacetone phosphate) is replenished from glycerol-3P in the mitochondria. 3P-glycerate exits the peroxisomes for further metabolism to pyruvate in the cytosol. (c) β-oxidation of fatty acids. (d) The glyoxylate cycle in plants, which allows the conversion of acetate to succinate, which can diffuse to mitochondria for gluconeogenesis ultimately to sucrose. (e) Contribution of peroxisomes in leaves to photorespiration, where 2-phosphoglycolate, the product of the rubisco oxygenase reaction in chloroplasts, will get converted, with the help of mitochondria, to 3-phosphoglycerate, the product of the rubisco carboxylase reaction.

found in these proteins, leading to ''mis-localization'' in the cytosol or preferential targeting to mitochondria. The balance of targeting can be shifted under different metabolic conditions, which can be caused by the induction or repression of receptors to the different targeting elements. It is a reasonable assumption that targeting of

these proteins is sometimes inefficient for important physiological reasons.

Peroxisomes have been most thoroughly studied in four systems: fungi (mainly yeasts), trypanosomatids, plants, and mammals. (Their functions and biogenesis in other vertebrates and in lower animals are relatively obscure.) Since the functions of peroxisomes have important and specific functions in these organisms, they will be discussed separately.

3 Specific Functions of Peroxisomes among Organisms

3.1 Peroxisomes in Fungi

Yeasts have played a major role in our understanding of peroxisomal function and biogenesis. It may be interesting for the reader to first consider a few major discoveries in the peroxisomal field using yeast systems in the context of other concurrent work.

The study of peroxisomes in fungi began as a search in the 1960s for a cheap source of food in animal husbandry and possibly for human use. Yeasts were identified by groups in Holland, Germany, and Japan that were able to grow on inexpensive and readily available carbon sources such as methanol or *n*-alkanes. When these strains were inspected by electron microscopy, investigators reported that microbodies were a prominent feature of the cytoplasm. Subsequent histocytochemistry and analysis of organellar fractions revealed that these structures contained oxidases and other enzymes that allowed the cells to assimilate the carbon source; they also contained catalase. Auspiciously, these observations were made shortly after the pioneering cell fractionation work of de Duve and Bauduin. The particles were therefore classified as peroxisomes. It soon became apparent that the size and number of these particles were controlled very tightly by growth substrate. Methylotrophs, for example, grown in glucose have only a few tiny peroxisomes, compared to the massive organelles in methanol medium. The plasticity of peroxisomes was further demonstrated in these yeasts when it was shown that the content of peroxisomes within a given species varied widely depending on the growth substrate; in contrast, a similar protein composition of the peroxisomal membrane demonstrated that these organelles were indeed related.

By this time, *Saccharomyces cerevisiae* had become a powerful model organism for genetic and biochemical approaches to biological problems. Although peroxisomes were discovered by Avers and Federman in *S. cerevisiae* in 1968, their existence in this model system was debated for many years. In normal glucose medium, they were difficult to detect by electron microscopy, no tools were available at this time for optical microscopy, and particulate catalase was attributed to vacuoles (lysosomes). Finally, in 1987 it was shown by Kunau, Veenhuis, and colleagues that growth on fatty acids caused peroxisomal proliferation such that peroxisomes were easily visible in the electron microscope. This observation, made at a time when the tools for molecular genetics were becoming very powerful in this system and the role of peroxisomes in human disease was apparent (see Sect. 6), generated much interest in the use of yeast for understanding peroxisomal assembly and function. Isolation of the organelles revealed that the β-oxidation pathway and several enzymes of the glyoxylate pathway

were contained within the peroxisome, allowing them to convert fatty acids to substrates for gluconeogensis and mitochondrial respiration. Not long after this, the genome sequence of *S. cerevisiae* was sequenced. Probing the genome for peroxisomal targeting sequences in the Gould laboratory revealed several proteins to be peroxisomal, including two components of the lysine biosynthetic pathway. Proteomic and genomic approaches have been since used for several other systems.

To complement ongoing work in *S. cerevisiae,* genetics were developed in methylotrophs (mainly *Pichia pastoris* and *H. polymorpha* and oleaginous yeasts (*Yarrowia lipolytica* and *Candida tropicalis*) as model systems to understand peroxisomal biogenesis. These systems have been crucial for the identification

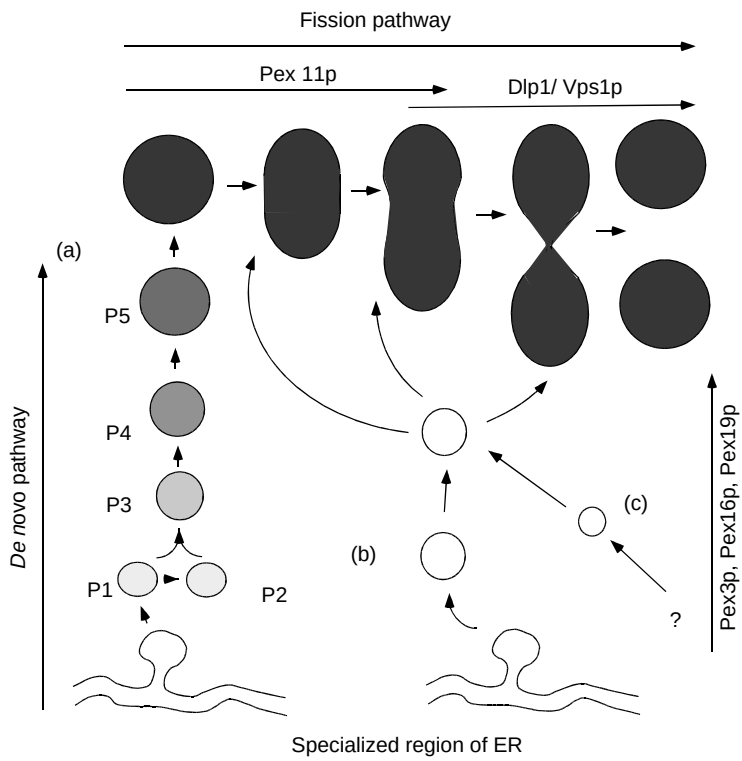

Fig. 3 Biogenesis of peroxisomes. Peroxisomes can divide by fission, top. Pex11p elongates the organelle and recruits the dynamin homolog Dlp1 (Vps1p in yeast) for organelle fission. (a) The *de novo* pathway in *Yarrowia lipolytica*, in which a series of discrete vesicular intermediates produce peroxisomes and also import subsets of matrix proteins (shading). (b) A hypothetical *de novo* pathway where vesicles bud from the endoplasmic reticulum (ER) and incorporate peroxins (Pex3p, which binds Pex19p and Pex16p) to target and assemble peroxisomal membrane proteins, which are required for matrix protein import. They fuse with mature peroxisomes. (c) Another hypothetical pathway in which vesicles are formed, but their source is unknown. This is considered as a possibility since the factors involved in classical budding from the ER seem not to be involved with peroxisomal assembly.

of genes and processes in the peroxisomal biosynthetic pathway. Genes so identified are termed *PEX*, and the corresponding proteins peroxins (Table 2). A discussion of their functions is provided in Sect. 4 below.

Yeasts, particularly *S. cerevisiae* and methylotrophs, have also been very important model systems for the study of peroxisomal degradation, since the switch from fatty acid or methanol to glucose normally results in rapid disappearance of peroxisomes. This process, termed *pexophagy*, takes two forms, micropexophagy and macropexophagy. Both are related to general autophagy, the process by which cells, when starved, digest portions of their cytoplasm to attempt to maintain viability. In macropexophagy, single peroxisomes are enwrapped in several layers of membrane (the source of which is unclear), which later fuse with elements of the vacuole, leading to digestion. Micropexophagy is analogous to general autophagy, where extensions of the vacuole surround groups of peroxisomes, directly leading to their digestion. Many genes have recently been identified that are required for pexophagy. Many of them overlap with autophagic genes. Interestingly, some of them also function in the cytosol-to-vacuole (cvt) pathway whereby precursors of at least two vacuolar proteins are normally synthesized in the cytosol and are delivered to the vacuole by a process resembling macropexophagy.

Molds and filamentous fungi also have peroxisomes that serve critical functions. In the filamentous fungus *Neurospora crassa*, peroxisomes contain the glyoxylate cycle and permit the cells to grow on acetate. More importantly, peroxisomes are essential for growth during hyphal growth of *N. crassa*. Mycelia formation is accompanied by the induction of Hex1p, which crystallizes in peroxisomes and buds from the organelle forming a structure termed a *Woronin body*. These particles are able to repair hyphae that become damaged by physically plugging holes in the cell membrane, thereby preventing the loss of cytoplasm and allowing other cell factors to seal the breach. New hyphae then can form adjacent to the initial event, expanding the organism.

Tab. 2 The localization and function of peroxins.

Peroxin	*Localization*	*Function*	*Comments*
Pex1p	Mainly cytosolic	Late in import; vesicle fusion?	AAA ATPase; associates with Pex6p
Pex2p	Peroxisomal membrane	Downstream of receptor docking	Zinc RING; associates with Pex10p
Pex3p	Peroxisomal membrane	Peroxisomal membrane biogenesis	Dock for Pex19p;
Pex4p	Cytosolic, peroxisomal membrane	Downstream of receptor docking	Binds to membrane via Pex22p; ubiquitin E2 ligase
Pex5p	Mainly cytosolic	PTS1 receptor	Binds cargo via TPR domains
Pex6p	Mainly cytosolic	Late in import; vesicle fusion?	Binds to membrane via Pex15p
Pex7p	Mainly cytosolic	PTS2 receptor	In mammals, can deliver cargo via Pex5p

Tab. 2 (*Continued*)

Peroxin	***Localization***	***Function***	***Comments***
Pex8p	Peroxisomal membrane	Links dock to RING proteins	Regulator of protein import?
Pex9p	Peroxisomal membrane	Protein import	Unique to *Y. lipolytica*
Pex10p	Peroxisomal membrane	Downstream of receptor docking	Zinc RING; binds to Pex2p, Pex10p, Pex5p
Pex11p	Peroxisomal membrane	Peroxisomal fission	Recruits dynamin-family members for fission
Pex12p	Peroxisomal membrane	Downstream of receptor docking	Zinc RING; binds to Pex10p and pex5p
Pex13p	Peroxisomal membrane	Docking	Binds to Pex14p and both PTS receptors
Pex14p	Peroxisomal membrane	Docking	Initial binding of Pex5p and Pex7p
Pex15p	Peroxisomal membrane	Protein import	Unique to *S. cerevisiae*; dock for Pex6p
Pex16p	Peroxisomal membrane	Peroxisomal membrane biogenesis	Binds to Pex19; absent in *S. cerevisiae*
Pex17p	Peroxisomal membrane	Docking	Binds to Pex14p; not in humans
Pex18p	Mainly cytosolic	PTS2 import	Binds to Pex7p
Pex19p	Mainly cytosolic	Peroxisomal membrane biogenesis	Docks on Pex3p; the likely PMP receptor
Pex20p	Mainly cytosolic	Thiolase import	Thiolase oligomerization; only in *Y. lipolytica*
Pex21p	Mainly cytosolic	PTS2 import	Only in *S. cerevisiae*
Pex22p	Peroxisomal membrane	Downstream of receptor docking	Docking of Pex4p
Pex23p	Peroxisomal membrane	Regulates peroxisomal size and number	
Pex24p	Peroxisomal membrane	Defective peroxisomal assembly	Mutants contain structures with a subset of peroxisomal proteins
Pex25p	Peroxisomal membrane	Regulates peroxisomal size and number	Binds to itself and Pex27p
Pex26p	Peroxisomal membrane	Protein import	Dock for Pex6p in humans
Pex27p	Peroxisomal membrane	Regulates peroxisomal size and number	Binds to itself and Pex25p
Pex28p	Peroxisomal membrane	Regulates peroxisomal size and number	Homologous to Pex24p
Pex29p	Peroxisomal membrane	Regulates peroxisomal size and number	Homologous to Pex24p
Pex30p	Peroxisomal membrane	Regulates peroxisomal size and number	Homologous to Pex23p; functions downstream of Pex28p and Pex29p
Pex31p	Peroxisomal membrane	Regulates peroxisomal size and number	Homologous to Pex23p; functions downstream of Pex28p and Pex29p
Pex32p	Peroxisomal membrane	Regulates peroxisomal size and number	Homologous to Pex23p; functions downstream of Pex28p and Pex29p

In penicillium molds, peroxisomes catalyze two critical steps in the biosynthesis of penicillin. Their importance is underscored by the finding that the amount of penicillin produced among species correlates with the number and size of their peroxisomes.

3.2 Peroxisomes in Trypanosomatids

Trypanosomatids are primitive flagellated protozoa that are responsible for widespread human suffering. The group contains trypanosomes and leishmania, the agents of African sleeping sickness, Chagas disease, and leishmaniasis.

These organisms contain glycosomes, organelles that contain enzymes of the Embden–Meyerhof pathway. The organelles are clearly in the peroxisome family since they contain catalase (in some species), their proteins usually terminate in peroxisomal targeting signals, and their assembly depends on peroxins, proteins that catalyze peroxisomal assembly.

Glycosomes contain the enzymes to catalyze the metabolism of glucose to 3-phosphoglycerate. This pathway does not result in net formation of ATP (this occurs at the pyruvate kinase step in the cytosol) but NADH is generated. NAD+ is reformed by the reduction of dihydroacetone (DHAP) phosphate to glycerol-3-phosphate. In itself this would result in loss of half the ATP from glucose (due to the loss of DHAP to glycerol), but the DHAP is regenerated from glycerol-3P in the mitochondria and circulated back to the glycosome, where it reenters the second half of the glycolytic pathway.

Trypanosomatids cycle between the mammalian bloodstream, insect gut, and salivary glands. When in the blood, trypanosomes are bathed in millimolar glucose. Mitochondrial respiration in the parasite is minimal and glycolytic enzymes are the major constituents of the glycosome. Once in the insect, however, amino acids are the major nutrients for the parasite (in its procyclic form). Subsequently, the composition of the glycosome changes; other pathways become more important for the organelle, and mitochondrial respiration and ATP generation are essential for survival.

Unlike yeast or isolated mammalian cells in culture, where peroxisomes are dispensable for maintenance and replication, trypanosomes die without glycosomes; organisms cannot be recovered from experiments in which PEX genes (i.e. those required for peroxisomal formation, see Sect. 6) are disrupted. It is thought that this reflects toxic effects of enzymes now in the cytosol, rather than a lack of glycosomal metabolism per se. Thus, when phosphoglycerate kinase, normally in glycosomes, is expressed in the cytosol, cells die. Interestingly, if the expression of PEX14 (required for peroxisomal protein import) is knocked down by RNAi technology, cells survived in the procyclic form, but only if glucose were not in the culture medium. Thus, the cytosolic mislocalization of glycolytic enzymes probably results in altered fluxes of intermediary metabolites in this compartment. While it was originally believed that the segregation of glycolytic enzymes within glycosomes was important for efficient channeling of pathway intermediates, the current thinking is that the compartmentation is essential to avoid interference of cytosolic metabolism by phosphorylated sugars from the glycosome. The precise mechanism of such toxicity remains to be elucidated.

Similar to mammalian peroxisomes (see below), glycosomes catalyze the biosynthesis of ether-linked phospholipids. The major surface glycosylphosphatidylinositol-linked protein of Leishmania, lipophosphoglycan, is bonded to the membrane via an ether- (instead of the normal ester-) phospholipid, a product of the glycosome. This surface molecule is not essential for life but is necessary for survival of the parasite within macrophages, which must occur for infection of mammals.

Finally, while peroxisomes in other cell types are well known to catabolize purines, glycosomes contain enzymes to salvage purines from the environment. This is an essential function since trypanosomes cannot synthesize them *de novo*. Both hypoxanthine guanine phosphoribosyl transferase and xanthine phosphoribosyl transferase contain peroxisomal targeting signals and are localized in glycosomes.

Recent analysis suggests that some glycosomal enzymes are of plant origin; similarities exist between algal and glycosomal forms. These include enzymes of glycolysis as well as the pentose-phosphate shunt and even the Calvin cycle! While this could indicate that glycosomes (and perhaps peroxisomes in other organisms) originated from an algal endosymbiont, it is also possible that the plant genes were acquired through lateral transfer after a phagocytic event.

3.3 Peroxisomes in Plants

Peroxisomes in higher plants are essential for seed germination and important in several metabolic pathways in the mature plant, including the synthesis of plant hormones and elucidation of signaling molecules. The development of molecular genetics in *Arabidopsis thaliana* has been the key to identifying many of these functions.

In oilseeds, energy is stored principally as triglycerides. Germination requires the conversion of these lipids to sugar, a process that requires fatty acid oxidation and the glyoxylate pathway. Fatty acids are metabolized to acetyl-CoA by β-oxidation, and the acetate is subsequently converted to succinate through the glyoxylate cycle. Both of these processes occur in an organelle termed the *glyoxysome*, a specialized type of peroxisome. During germination and leaf development, these organelles convert into leaf peroxisomes, as a specific subset of proteins is induced and imported into the organelles. That leaf peroxisomes develop from glyoxysomes and not from a separate source was shown by double immunogold electron microscopy using probes against enzymes in both pathways. In senescent plant tissues, backwards conversion into glyoxysomes occurs.

Leaf peroxisomes are central in photorespiration. In this process, the CO_2-fixing enzyme within plastids, ribulose-5-phosphate carboxylase/oxidase (Rubisco), accepts O_2 instead of CO_2, resulting in the formation of phosphoglycolate and phosphoglycerate (instead of two molecules of phosphoglycerate). Phosphoglycolate is dephosphorylated and shuttled into peroxisomes. Peroxisomes convert glycolate into glycine, which then can be transformed into serine within mitochondria, with a loss of CO_2. Serine shuttles back into peroxisomes and is converted there into phosphoglycerate, which can enter the plastid to complete the cycle. On balance, photorespiration appears wasteful for plants since it results in a loss of about 25% of fixed carbon with no gain in energy. However, it appears useful under conditions of low tissue CO_2 and

may be important for efficient utilization of nitrogen.

Plant peroxisomes are also important in hormone production and signal transduction. The hormone indole-3-acetic acid is generated in peroxisomes from indole-3-butyric acid. Similarly, the final step in the biosynthesis of jasmonate occurs in peroxisomes. Nitric oxide synthase, responsible for production of NO, has also been shown recently to be peroxisomal. Both hydrogen peroxide and superoxide, which can act as second messengers relaying information about the redox state of the cytoplasm, are generated within peroxisomes. Conversely, peroxisomal catalase has been shown to be essential for protection of the plant from oxidizing toxins such as paraquat and ozone.

Finally, there is evidence that peroxisomes play key roles in morphogenesis beyond their requirement for germination. The development of normal floral morphology can be severely altered and inhibited in an *aim1* mutant strain; *AIM1* appears to encode a peroxisomal fatty acid oxidation enzyme. Light-stimulated morphogenesis is also affected by peroxisomes. This pathway is rendered constitutive in a *det1* mutant; *DET1* encodes a DNA repressor. However, the effects of *det1* can be suppressed by a dominant mutant in *TED3*, which itself encodes Pex2p, a peroxisomal integral membrane protein that is a member of the protein import complex. While the details of the *DET1-TED3* interaction remain to be worked out, it is another example of a role of peroxisomes in development and signaling.

3.4 Peroxisomes in Mammals

A principal role of mammalian peroxisomes is the β-oxidation of long-chain ($C_{12}-C_{18}$) and very long chain fatty acids ($C_{20}-C_{26}$). In yeast, all fatty acid β-oxidation occurs in peroxisomes and it is the major locus for this activity in plants. In mammals, the peroxisomal pathway prefers the long and very long chains. In the organelle, they are metabolized to medium chain fatty acids, which are then shuttled as carnitine conjugates to the mitochondria for further oxidation. Peroxisomal enzymes of this pathway are distinct from the mitochondrial forms and contain peroxisomal targeting signals. The reactions in the pathway are similar in the two organelles except for the initial step. In the mitochondria, acyl-CoAs are converted to enoyl-CoAs by an NADH dehydrogenase, thereby generating reducing equivalents. In peroxisomes, the step is catalyzed by an oxidase, which generates hydrogen peroxide from molecular oxygen. The peroxide is then destroyed by catalase. The enzymology of subsequent steps is similar in the two organelles.

Peroxisomes are also critical in metabolizing branched chain fatty acids. Two-branched fatty acids, such as phytanic acid, a product of chlorophyll breakdown and an abundant product in injested plants, cannot utilize the β-oxidation pathway directly and must be first metabolized by a single round of α-oxidation, using peroxisomal enzymes to convert phytanic to pristanic acid, which can then be metabolized by the β-oxidation pathway.

Induction of the peroxisomal fatty acid oxidation pathway is catalyzed by the α-isoform of peroxisome proliferator–activator receptor (PPARα). The PPARs are a family of nuclear receptors that form heterodimers with retinoid X receptor (RXR) and control transcription of several metabolic pathways. PPARα is expressed mainly to liver and to a lesser

extent to heart and skeletal muscle. After fasting, fatty acids released by adipose tissue bind to the receptor, inducing peroxisome and mitochondrial proliferation and fatty acid oxidation, thus providing ketone bodies to the periphery as an energy source. While the induction of metabolic enzymes by PPARα is understood in some detail, the way in which organelle proliferation per se is induced is less clear.

Reactions similar to those of peroxisomal β-oxidation catalyze chain shortening of C_{27}-sterols to form bile acids within peroxisomes.

Peroxisomes also catalyze the initial steps in the synthesis of ether-linked phospholipids. Compared to the better-known ester-linked glycerolipids, there is an ether (in plasmanyl phospholipids such as platelet activating factor) or ether-vinyl linkage (in plasmalogens) at the sn-1 position of the glycerol backbone. Platelet activating factor is a potent cytokine that promotes inflammation through PAF receptors on the cell surface. Plasmalogens are at high levels in nervous system tissue and comprise most of the phospholipids in myelin. Although the precise physiological roles of plasmalogens are not fully understood, they are probably important for maintaining lipid fluidity and contributing to efficient endocytosis and other events of membrane dynamics. Since the sn-2 position of plasmalogens is often polyunsaturated fatty acids, they are targets of PLA_2 enzymes and may generate the bulk of arachidonic and docosahexaenoic acids, sources for ecosinoids such as prostaglandins, extremely potent second messengers.

Peroxisomes contain several enzymes in the cholesterol biosynthetic pathway. However, none of these enzymes are uniquely found in this compartment, and the amount of cholesterol biosynthesis in fibroblasts is not much decreased in the absence of the organelle (i.e. from patients with peroxisomal biogenesis disorders). Thus, the physiological significance of peroxisomal cholesterol biosynthesis is not clear.

There are several known flavin oxidases in mammalian peroxisomes besides the acyl-CoA oxidase that functions in the β-oxidation pathway. L-α-hydroxyacid oxidase is active against small molecules such as glycolate, lactate, and a few L-amino acids. Distinct from this enzyme is D-amino acid oxidase, which has a broad specificity for D-amino acids. Since intermediary mammalian metabolism utilizes L-amino acids, the function of this oxidase remains a mystery. One hypothesis is that it is there to metabolize bacterial products that find their way to the bloodstream. Another idea is that it metabolizes glycine or D-serine in the nervous system, thus possibly regulating the action of these important neurotransmitters.

Uric acid oxidase (uricase) is the initial enzyme in the breakdown of uric acid in many mammalian species. Rodent peroxisomes are distinguished by a crystalloid core in which uricase and other proteins are localized. The enzyme is absent in humans and anthropoid apes, whose peroxisomes lack cores. Finally, peroxisomal oxidases metabolize polyamines, which are important for stability of nucleic acids and for a variety of growth processes, and pipecolic acid, a degradation product of lysine.

Mammalian peroxisomes contain alanine: glyoxylate aminotransferase, which catalyzes the conversion of glyoxylate to glycine, thereby detoxifying cells of glyoxylate. Otherwise, glyoxylate is oxidized to oxalic acid, which is insoluble and crystallizes in the kidney leading to serious disease, further discussed in Sect. 6.

4
Biogenesis of Peroxisomes

4.1
Origin of the Peroxisome

It is not clear whether the first eukaryotic peroxisome was the result of internalization of a prokaryotic ancestor, as is the case for mitochondria, or whether eukaryotes were entirely responsible for biosynthesis for these organelles. Mitochondria contain a genome and machinery for DNA replication and protein synthesis that can be traced back to prokaryotic ancestors. Peroxisomes have no genome, and there is no apparent homology between groups of structural or enzymatic components of these organelles and prokaryotic counterparts, except for the homology between glycosomal and algal proteins noted earlier. There is no evidence, such as skewed codon usage, for ancient transfer of hypothetical genetic elements between peroxisome and nucleus. However, such differences might have existed long ago and have been erased over hundreds of millions of years of evolution. Thus far, there is no compelling evidence to support or refute the endosymbionic origin of peroxisomes.

4.2
Peroxisome Fission

Similar to mitochondria, peroxisomes can divide by fission (Fig. 3). Pex11p is considered to be the major fission factor that acts from the peroxisomal membrane, since its absence results in fewer peroxisomes, some of which are much larger than in wild type; expression of Pex11p in a null background causes the fission of the large organelles. Overexpression of Pex11p leads to an increased number of small organelles. Pex11p can self-associate in oligomers, and there is evidence that this behavior regulates its action. The carboxy terminus of Pex11p can bind coatomer, which is involved in vesicle formation and thus might promote fission, but the physiological significance of this observation has been questioned. Recently several other peroxins, Pex23p, Pex25p, and Pex27p through Pex32p have been identified that also regulate peroxisomal number and size. Many of these interact with each other, and epistasis analysis has demonstrated some order of events; for example, Pex31p and Pex32p function after Pex28p and Pex29p. Some of these factors may be working in a *de novo* pathway rather than in peroxisomal fission (see below). The number of proteins that regulate peroxisomal size and number suggest a complex mechanism for these processes.

Pex11p recruits Dlp1 (Vps1p in yeast), a member of the dynamin family, to peroxisomes. Dynamin has been shown to promote scission of vesicles from the plasma membrane during endocytosis. Whether dynamins themselves induce membrane fission or act as switches for other motor proteins is under active investigation. Regardless, this family of proteins promotes both mitochondrial and peroxisomal division. If expression of Dlp1 is silenced or dominant-negative alleles are used, long tubulated peroxisomes are formed, and these structures are even more pronounced if Pex11p is overexpressed. A fluorescently tagged Dlp1 catalytically inactive mutant shows a striped pattern on these tubules, which alternates with the Pex11p pattern. Although Pex11p is necessary for recruitment of Dlp1p, direct association of these two proteins cannot yet be demonstrated, suggesting that the interaction is indirect.

While yeast has 1 Pex11p, humans have three isoforms. Pex11α is induced by PPAR while Pex11β and Pex11γ are constitutive. In mice, 11β is essential for peroxisomal function (see Sect. 6).

4.3 The Origin and Assembly of the Peroxisomal Membrane

Besides peroxisomal fission, *de novo* biogenesis of peroxisomes also occurs. One line of evidence for this alternative pathway is the formation of peroxisomes in cells apparently lacking any. For example, a pex16 mutant fibroblast line has no visible peroxisomes or remnant peroxisomal membranes. If wild-type PEX16 is then introduced into these cells, peroxisomes appear. A second line of evidence is that continuities have been shown between membranes of the endoplasmic reticulum (ER) and peroxisome-related structures. Finally, mutants in the secretory pathway mislocalize peroxisomal proteins. While all these data point to the ER as the source of peroxisomes, such a pathway would have novel elements: peroxisomal matrix proteins do not enter the lumen of the ER *en route* to peroxisomes. The Sec61 complex in the ER membrane, through which secreted proteins and most membrane proteins destined for the exocytic pathway pass, is not involved with peroxisomal biogenesis. Neither COP1 nor COPII is required for peroxisomal assembly. With two probable exceptions in the yeast *Y. lipolytica* (Pex2p and Pex16p), no peroxisomal proteins are glycosylated. Experiments in *Y. lipolytica* have demonstrated that a discrete series of vesicular intermediates derived from the ER comprises the *de novo* pathway in that organism. Such a pathway has not yet been shown in other organisms, however. Regardless of the details, it is clear that the ER plays a role in *de novo* peroxisomal assembly in some way.

Three proteins have been shown to be important in the delivery of peroxisomal membrane proteins (PMPs) to nascent organelles: Pex19p, Pex3p, and in some species, Pex16p. Pex19p is a mainly cytosolic protein that binds to most PMPs. While there was much debate on whether Pex19p served as the PMP receptor to guide newly synthesized PMPs to the membrane, or instead served as a chaperone for integration once the protein was targeted, the consensus is now that it indeed binds to targeting elements on PMPs. A critical experiment to this end demonstrated that if Pex19p was intentionally mistargeted to the nucleus, many PMPs arrived there as well. Pex19p can bind to Pex3p, an integral membrane protein, which probably serves as a docking element for Pex19p-PMP complexes. In mammalian cells, the membrane protein Pex16p is also involved with PMP targeting and may stabilize Pex3p-Pex19p interactions. In the absence of any one of these three peroxins, peroxisomes are absent, and membranous remnants are difficult to find.

Unlike PTS1, which is very easy to recognize (see below), there is no motif (termed *mPTS*) easily recognizable on PMPs. Many PMPs have been analyzed for targeting information. In general, the mPTS consists of a transmembrane element and sequences adjacent. The matrix-facing sequences contain net basic charge and the cytosol-facing adjacent sequence often predicts an amphipathic helix. Several PMPs contain two mPTSs that can function independently of each other. Their role may be to increase the efficiency of targeting (for example, by binding to multiple Pex19p molecules) or to target the cargo to different subpopulations of peroxisomes. The extent to which the Pex(3,16,19)p system

acts to target PMPs to a preperoxisomal compartment (to establish the machinery for protein import into the matrix) versus to mature peroxisomes is not known but may depend on the fraction of organelles being generated by the fission versus *de novo* pathway.

To summarize, the weight of the evidence supports the existence of an ER-based *de novo* pathway for peroxisomal generation. Many details of this pathway, however, are not clear. The ER does not appear to function as it does in the classical exocytic pathway. There is no consensus on whether peroxisomes, near maturity, bud from the ER or whether a series of small vesicles are involved. Different mechanisms may work in different cell types. If a specialized region of the ER is involved, the nature of this compartment has not been elucidated. The next few years are likely to see resolution to this fascinating problem.

4.4 Import of Proteins into Peroxisomes

The last 15 years has seen remarkable progress toward a molecular understanding of peroxisomal import of proteins. However, some aspects of the basic mechanism remain unresolved.

Peroxisomal matrix proteins are synthesized in the cytoplasm. Nearly all of them contain a short targeting motif at their extreme carboxy terminus termed *PTS1* (*peroxisomal targeting signal 1*). The prototypic sequence of the PTS1 is SKL (serine-lysine-leucine), although many variations of this can be found, some of which are favored in various cell types. Several matrix proteins are targeted instead by a non-apeptidyl PTS2 sequence at or near their amino termini. This targeting signal is less conserved than PTS1 and consists of –(R/K)(L/V/I/Q)XX(L/V/I/H/Q)(L/S/G/A/K)X(H/Q)(L/A/F)–. In some systems, the amino terminus containing PTS2 is proteolytically removed during import. A few peroxisomal proteins have both PTS1 and PTS2 signals, and a few others have neither (see below).

Unlike classical protein import pathways across mitochondrial or ER membranes where unfolded proteins enter the import apparatus, peroxisomal cargo folds and assembles into oligomers before translocation. This has been demonstrated in several systems by coexpressing protein subunits that are lacking or containing PTS sequences (in which subunits without the signal are imported by a "piggyback" mechanism), as well as in pulse-chase experiments. The capacity for peroxisomes to import large molecules is also demonstrated by their ability to import colloidal gold particles conjugated with a peroxisomal targeting signal. Proteins, with few exceptions, enter peroxisomes as holoenzymes. The sole exception so far to this rule is alcohol oxidase, a 600-kDa homooctamer in methylotrophic yeasts, which is imported as monomers and assembles within the matrix of yeast peroxisomes. Whether this protein is simply too large to cross the peroxisomal membrane, or must assemble internally for other reasons, is not known.

Cytosolic chaperones play important roles in various protein import pathways. In many systems, proteins in the dnaK (Hsp70) and dnaJ families (initially identified in *Escherichia coli* in DNA replication and later as heat-shock proteins, Hsp) interact to maintain cargo in an unfolded state and protect them from aggregation before they encounter an import apparatus. Hsp70 and Djp1 (a dnaJ protein) have been shown to be important for peroxisomal import as well. However, instead of

keeping peroxisomal cargo unfolded, they act either in the folding of cargo or in keeping the PTS accessible to receptor.

The receptors to which the PTS1 and PTS2 bind are Pex5p and Pex7p, respectively (Fig. 4). The structure of Pex5p in complex with PTS1 has been solved. The carboxyl half of the molecule consists of six tetratricopeptide repeats, which form a sandwich around the PTS1 peptide. As mentioned earlier, a few matrix proteins contain no recognizable PTSs. However, their import depends on Pex5p, and they probably bind to this molecule outside of the normal PTS1-binding area.

In yeast, import of PTS2 cargo by Pex7p occurs independently of Pex5p. However, in mammals import requires Pex5p, an unexpected observation. Two forms of Pex5p are synthesized, one with a 37-amino acid extension produced by alternate splicing of the RNA. The Pex5L (long) form can bind Pex7p through the extension, and the extension is required for Pex7p-mediated PTS2 import. In yeast systems, there are accessory proteins in PTS2 import that may replace the function of the Pex5L extension in PTS2 import.

Since most peroxisomal cargos are oligomers, and Pex5p itself form tetramers, it is easy to imagine that large complexes may develop between receptor and cargo. According to the preimplex model of Gould, the formation of these complexes on the peroxisomal surface may be an important intermediate in import.

Once receptor binds to cargo, the complex docks on the peroxisomal surface. The minimal docking structure consists of Pex14p, Pex13p, and Pex17p, all membrane proteins. Binding studies indicate that Pex14p can self-associate and probably dimerizes in the membrane. In most systems, it behaves as a transmembrane protein and probably provides the first contact with the cargo-bound receptor. Cargo-bound Pex5p preferentially associates with Pex14p, probably because the cargo unmasks the Pex14p binding site on the amino terminal segment of Pex5p. Pex14p in turn associates with Pex13p and Pex17p. Pex17p is a peripheral element (facing the cytoplasm) and is nonessential to import in some systems. Pex13p, an integral membrane protein, binds to Pex7p in yeast through its amino terminus, and binds to Pex5p and Pex14p at its carboxy terminus. Both termini face the cytoplasm.

What happens after binding of receptor-cargo to the docking complex is not clearly understood yet. Several proteins act downstream of docking, shown by epistasis experiments; docking is not affected if they are eliminated. The docking complex can associate with another complex of three proteins, Pex2p, Pex10p, and Pex12p. All of these are essential for import. Each member of this trio contains a zinc RING motif facing the cytoplasm. Recent data indicates that the link between the docking and RING complexes is Pex8p, a peripheral membrane element that faces the matrix. However, Pex8p is present at lower concentrations than the other proteins of the two complexes, so that the linkage may be regulated rather than constitutive.

There has been much debate over whether the cytosolic receptors enter the matrix to drop off cargo (the simple vs extended shuttle model). The weight of the evidence favors the extended model whereby Pex5p and Pex7p release their cargo in the matrix. A fraction of these molecules are sometimes found in the matrix, and defective alleles in downstream peroxins often lead to the accumulation of receptors in that compartment. Pex5p has been shown to become protease-protected (i.e. protected by the membrane) during its

import cycle. Furthermore, Pex8p, which is not exposed to the cytoplasm face, can bind Pex5p. Finally, Pex5p with an added PTS2 extension that is normally cleaved in the matrix is similarly cleaved during import. Data from a mammalian *in vitro* import model suggests that Pex5p remains associated with the membrane at every step of its cycling reaction, so the protein may not necessarily entirely enter the matrix, but only drop off its cargo at the internal surface.

Two other membrane protein complexes function downstream of import: one containing two ATPases, and the other an E2 ubiquitin ligase.

Pex1p (altered *PEX1* alleles commonly cause human peroxisomal diseases, see Sect. 6) and Pex6p contain Walker ATPase domains and associate with each other. Although largely cytosolic, a fraction of each associates with the peroxisomal membrane by binding to integral membrane proteins: Pex15p in yeast and Pex26p in mammals. The ATPase function on both is critical to the import cycle, but its precise role is controversial. According to the preimplex model, Pex1p and Pex6p may serve to disassemble large receptor–cargo complexes once they have bound to the docking apparatus, and allow the import of single units. Another idea is that they are essential for the export of receptor after release of cargo in the matrix. In support of this hypothesis, it has been shown that export, but not import, of proteins is an ATPase-requiring step. Finally, in the yeast *Y. lipolytica*, Pex1p and Pex6p appear to be involved with fusion of preperoxisomal vesicles. It is very difficult to reconcile a role for these proteins in vesicle fusion versus protein import or export unless the import cycle is not strictly the movement of proteins through a proteinaceous gate in the membrane but instead involves membrane dynamics to an extent not appreciated or understood at present.

The last complex that participates in the import cycle contains Pex4p and Pex22p. Pex4p is a peripheral protein that is anchored to the bilayer by Pex22p. The dimer can bind to the RING complex through Pex4p–Pex10p interactions, linking them indirectly to the docking complex. Epistasis experiments place them after the Pex1p–Pex6p complex, however, and they are assumed to participate in Pex5p recycling to the cytosol.

Pex4p is an E2 ubiquitin ligase. Protein ubiquitination (ubiquitin is a small peptide) is widely known to be a tagging mechanism for proteins destined for degradation in the proteosome. It is also known, however, to be involved in other protein trafficking pathways, for example, recycling of plasma membrane receptors. How ubiquitination relates to the importance of Pex4p for protein import into peroxisomes has been a riddle since its sequence was determined. Very recently it was shown that Pex5p is ubiquitinated by Pex4p, and that this reaction prevents Pex5p from polyubiquitination and degradation. E2 ligases usually transfer ubiquitin moieties through E3 ligases. Pex10p is required for Pex5p-mediated monoubiquitylation, and thus Pex10p is a good candidate for the E3 enzyme. These data explain older experiments in which peroxin mutants led to instability of Pex5p. Pex4p may serve to regulate the concentration of Pex5p to peroxisomal need: The receptor is kept at low levels with relatively inactive peroxisomes, and increases in concentration (due to stabilization by monoubiquitylation) when peroxisomal import is upregulated.

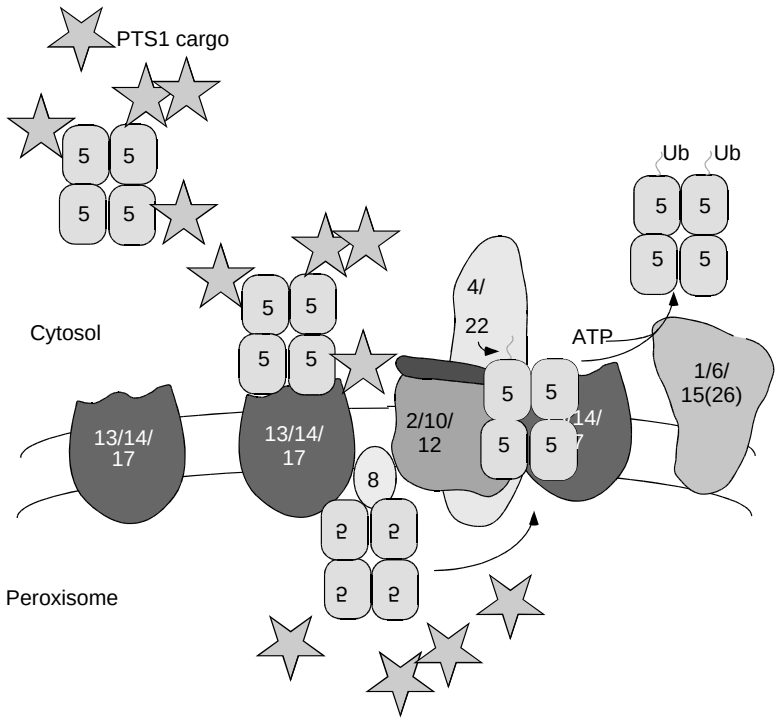

Fig. 4 Protein translocation across the peroxisomal membrane. Pex5p ("5" in the figure) with PTS1 cargo (stars) docks to the Pex13p–Pex14p–Pex17p complex. Subsequent steps are less clear, but require the other peroxins shown. Cargo is most likely released in the matrix. The Pex2–Pex10p–Pex12p complex contain interacting RING motifs, shown in dark red; it is not clear whether they are required for import or recycling of receptor. Pex8p links these two complexes as shown. Pex4p is an E2 ubiquitin ligase that monoubiquitinates Pex5p, preventing polyubiquitination and degradation by the proteosome. The ATPases Pex1p and Pex6p are shown promoting the export of receptor, although it is possible that they function earlier in the pathway, such as in the dissociation of receptor and cargo. (See color plate p. xxxiv).

5 Peroxisomal Motility

Observations of peroxisomes in living cells was made possible by the use of fluorescent tags, such as green fluorescent protein (GFP)-PTS1. Experiments in mammalian cultured cells revealed that peroxisomes fell into two classes; most moved randomly with no net spacial translation, while 10 to 15% move rapidly and directionally. Since inhibitors of microtubule assembly, but not inhibitors of actin assembly, inhibited this movement, it was concluded that peroxisomes can move on microtubules.

Particles that move on microtubules can utilize either dynein or kinesin-based motors. Both systems utilize ATP hydrolysis for movement, and peroxisomes have been shown *in vitro* to move on microtubules in an ATP-dependent fashion. While neither dynein nor kinesin has been shown to bind to peroxisomes, a dominant-negative fragment of dynactin, a dynein-accessory

protein, inhibits peroxisomal movement. Interestingly, purified peroxisomes can bind directly to purified microtubules, suggesting that the peroxisomal membrane contains specific microtubule-binding elements. This is supported by the finding that protease-treatment of intact peroxisomes results in the loss of microtubule binding. The factor responsible, however, has not yet been identified.

In contrast to their behavior in animal cells, peroxisomes in plant and yeast cells utilize actin-based motors for movement. In yeast, Myo2p is necessary for movement on actin cables, and this movement is essential for segregation of peroxisomes into buds during division. Actin is also important for the movement of plant peroxisomes; drugs that block actin also inhibit movement of peroxisomes.

Why move peroxisomes along microtubules? Peroxisomes display a uniform distribution in animal cells. In contrast, they can be in close proximity to the endoplasmic reticulum and to lipid bodies, which may reflect their origin or need for lipid substrates, respectively. Microtubules may be responsible for transporting peroxisomes to or from these sites. In some tissues, peroxisomes show a polar distribution, such as their appearance at the basolateral pole of proximal tubular cells of the kidney, and microtubules may be required to maintain that polarity. These filaments appear not to be important for segregation in animal cells during cytokinesis, since this process seems to be random.

6 Peroxisomal Diseases

The vital role of peroxisomes in early plant development has been described earlier (Sect. 3). In contrast, mammals can give birth to offspring that lack peroxisomes, but they are severely deficient in many organ systems and do not survive long afterwards. Human peroxisomal diseases are subdivided into those with a single enzyme deficiency, and the peroxisome biogenesis disorders (PBDs) with defects in *PEX* genes.

6.1 Single Enzyme Disorders

There are several disorders caused by defects in single peroxisomal matrix proteins. These range from the relatively benign (catalasemia) to the more serious (X-linked adrenoleukodystrophy (XALD)). XALD is characterized by an accumulation of very long chain fatty acids caused by defects in a peroxisomal membrane protein that probably encodes a subunit of the fatty acid transporter. Depending on the allele, patients may be stricken in adulthood with a disease that is degenerative over several generations, or present with the disease in childhood and show rapid disease progression. Symptoms are neurological (including severe mental retardation in severe cases) as well as endocrine, particularly involving adrenal insufficiency.

A peroxisomal single enzyme deficiency with an interesting mechanism is hyperoxaluria type I; afflicted patients generate extensive renal oxalate stones resulting in extensive kidney damage. The disease is caused by point mutations in the peroxisomal enzyme alanine glyoxylate aminotransferase (AGT), which normally converts glyoxylate to glycine in peroxisomes. Wild-type AGT forms homodimers in the cytosol and is targeted to peroxisomes by a PTS1. In some patients, the enzyme is inactive or forms

aggregates. In other patients, however, two mutations in AGT prevent cytosolic dimerization and concomitantly convert a cryptic mitochondrial-targeting signal to an active one. These mutations in combination cause AGT to mistarget to mitochondria. In this compartment, glyoxylate is not metabolized. Instead, it becomes oxidized to the more insoluble oxalate, causing hyperoxaluria and calcium oxalate stones.

6.2 Peroxisome Biogenesis Disorders

In 1964, a hereditary metabolic disease in infants was described by Zellweger and colleagues, characterized by severe hypotonia, craniofacial abnormalities, hepatomegaly, renal cysts, and mental retardation. Children severely affected lived only weeks or a few months. In 1973, Goldfischer and coworkers made the observation that cells from these patients lacked peroxisomes. Since then it has become clear that peroxisomal biogenesis disorders comprise a spectrum of diseases that range from Zellweger syndrome as the most severe, through neonatal adrenoleukodystrophy in intermediate severity, to infantile Refsum's disease as the mildest form. As it became appreciated that all of these diseases shared a lack of peroxisomes, complementation studies were performed by cell fusion assays to determine the number of genes involved. Fibroblasts from patients fell into 12 (now 14) complementation groups. Identification of the genes responsible was greatly facilitated by sequence information of yeast *PEX* genes. The severity of symptoms (and the assignment to the particular PBD) was found to be more a function of the degree of inactivation of the PEX gene product rather than of specific *PEX* genes. Mutations in PEX6 are responsible for nearly 70% of all PBDs.

Severely affected patients have large increases in concentrations of very long chain fatty acids and of bile acid intermediates, consistent with a decrease in peroxisomal fatty acid β-oxidation pathways (Table 3). They show deficiencies in plasmalogens and the enzymes of ether lipid biogenesis associated with peroxisomes. Defective ether lipid synthesis is thought to be responsible for defects in neuronal migration and the establishment

Tab. 3 Some biochemical abnormalities in Zellweger spectrum PBDs.[a]

Substrate or Enzyme	*Change*	*Comment*
Very long chain fatty acids	Greatly elevated	Defect in β-oxidation
Bile acid intermediates	Greatly elevated	Defect in β-oxidation
Pristanic acid	Elevated	Defect in β-oxidation
Phytanic acid	Elevated	Defect in α-oxidation
Plasmalogens (products of peroxisomal metabolism)	Greatly decreased	Defect in ether lipid biogenesis
DHAP aminotransferase	Greatly decreased	Mislocalized
Acyl-CoA oxidase	Greatly decreased	Mislocalized and degraded
Catalase	Mislocalized	Active in cytosol

[a]Modified from Weller, S., Gould, S.J., Valle, D. (2003) Peroxisome biogenesis disorders, *Annu. Rev. Genomics Hum. Genet.* **4**, 165–211.

of white matter during the development of the central nervous system.

Another disease, rhizomelic chondrodysplasia punctata (RCDP) has several common features to Zellweger syndrome, particularly altered bone growth (punctate calcifications of cartilage and shortening of the humerus and femur) as well as hypotonia and neurological symptoms. Cataracts are common. Depending on allele and with proper support, patients with RCDP can live for several decades. The main cause of the disease appears to be poor synthesis of ether lipids, either from a defect in *PEX7* leading to a specific PTS2 import defect, or the loss of dihydroxyacetonephosphate acyltransferase, a key peroxisomal enzyme in the ether lipid biosynthetic and a PTS2 substrate. RCDP illustrates the importance of ether lipids for normal physiology.

6.3 Mouse Models for PBDs

Gene disruptions in mice have been extraordinarily helpful in understanding human disease. Thus far knockouts have been generated in PEX2, PEX5, PEX13, PEX11α, and PEX11β. Mice homozygous for disruptions in PEX2, PEX5, and PEX13 die shortly after birth and exhibit many symptoms similar to Zellweger syndrome: low birth weight, lack of neuronal migration, poor feeding, and hypotonia. Similar to humans they lack viable peroxisomes but contain peroxisomal ghosts. Thus, they appear to be good models for human PBDs. Interestingly, in PEX5 −/−cells, the defect in neuronal migration (certainly involved with the mental retardation in children) is not corrected by reintroducing brain-specific PEX5 expression; the defect is still seen in animals not expressing the gene product in liver. Only by expressing PEX5 in liver and in brain is this neurological defect corrected.

As indicated earlier, PEX11 has three isoforms in humans, α, β and γ; knockouts have been made in the α- and β-isoforms. Unexpectedly, there was no phenotype in the mice lacking PEX11α, the form that is induced by PPARα. Even more surprising was the finding that peroxisomal number increased in response to PPARα agents, demonstrating that Pex11α is not necessary for peroxisomal induction by this transcription factor. In contrast, peroxisomes were less abundant in mice in which PEX11β, a constitutive and widely expressed form, was disrupted, and the animals displayed a Zellweger-like phenotype. The concentration of peroxisomal products and intermediates were only mildly altered; the reason for the severe phenotype in these animals is not yet clear.

The mouse disruption has also helped settle a controversy on the importance of peroxisomal cholesterol biosynthesis. Many enzymes in isoprenoid biosynthesis (shared in the cholesterol pathway) are localized in peroxisomes and their physiological significance has been uncertain. Mice lacking peroxisomes have only minor (if any) alterations in overall sterol biosynthesis; there may be a minor difference in liver biosynthesis. Thus, the role of peroxisomes in this pathway remains uncertain.

See also Intracellular Fatty Acid Binding Proteins in Metabolic Regulation; Lipid and Lipoprotein Metabolism; Metabolic Basis of Cellular Energy; Phospholipids.

Bibliography

Books and Reviews

Bellu, A.R., Kiel, J.A. (2003) Selective degradation of peroxisomes in yeasts, *Microsc. Res. Tech* **61**, 161–170.

Brites, P., Waterham, H.R., Wanders, R.J.A. (2004) Functions and biosynthesis of plasmalogens in health and disease, *Biochim. Biophys. Acta* **1636**, 219–231.

Brown, L.-A., Baker, A. (2003) Peroxisome biogenesis and the role of protein import, *J. Cell. Mol. Med.* **7**, 388–400.

Eckert, J.H., Erdmann, R. (2003) Peroxisome biogenesis, *Rev. Physiol. Biochem. Pharmacol.* **147**, 75–121.

Gould, S.J., Collins, C.S. (2002) Opinion: peroxisomal-protein import: is it really that complex? *Nat. Rev. Mol. Cell. Biol.* **3**, 382–389.

Hayashi, M., Toriyama, K., Kondo, M., Kato, A., Mano, S., de Bellis, L., Hayashi-Ishimaru, Y., Yamaguchi, K., Hayashi, H., Mishimura, M. (2000) Functional transformation of plant peroxisomes, *Cell. Biochem. Biophys.* **32**, 295–304.

Masters, C., Crane, D. (1995) *The Peroxisome: A Vital Organelle*, Cambridge University Press, Cambridge, UK.

Parsons, M. (2004) Glycosomes: parasites and the divergence of peroxisomal purpose, *Mol. Microbiol.* **53**, 717–724.

Titorenko, V.I., Rachubinski, R.A. (2004) The peroxisome: orchestrating important developmental decisions from inside the cell, *J. Cell Biol* **164**, 641–645.

Titorenko, V.I., Rachubinski, R.A. (2001) The life cycle of the peroxisome, *Nat. Rev. Mol. Cell Biol.* **2**, 357–368.

Tolbert, N.E. (1981) Metabolic pathways in peroxisomes and glyoxysomes, *Annu. Rev. Biochem* **50**, 133–157.

Weller, S., Gould, S.J., Valle, D. (2003) Peroxisome biogenesis disorders, *Annu. Rev. Genomics Hum. Genet.* **4**, 165–211.

Primary Literature

Aboushadi, N., Engfelt, W.H., Paton, V.G., Krisans, S.K. (1999) Role of peroxisomes in isoprenoid biosynthesis, *J. Histochem. Cytochem.* **47**, 1127–1132.

Agne, B., Meindl, N.M., Niederhoff, K., Einwachter, H., Rehling, P., Sickmann, A., Meyer, H.E., Girzalsky, W., Kunau, W.H. (2003) Pex8p: an intraperoxisomal organizer of the peroxisomal import machinery, *Mol. Cell* **11**, 635–646.

Avers, C.J., Federman, M. (1968) The occurrence in yeast of cytoplasmic granules which resemble microbodies, *J. Cell Biol.* **37**, 555–559.

Baudhuin, P., Beaufay, H., De Duve, C. (1965) Combined biochemical and morphological study of particulate fractions from rat liver. Analysis of preparations enriched in lysosomes or in particles containing urate oxidase, D-amino acid oxidase, and catalase, *J. Cell Biol.* **26**, 219–243.

Blattner, J., Helfert, S., Michels, P., Clayton, C. (1998) Compartmentation of phosphoglycerate kinase in Trypanosoma brucei plays a critical role in parasite energy metabolism, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 11596–11600.

Bowen, P., Lee, C.S., Zellweger, H., Lindenberg, R. (1964) A familial syndrome of multiple congenital defects, *Bull Johns Hopkins Hosp* **114**, 402–414.

Braverman, N., Dodt, G., Gould, S.J., Valle, D. (1998) An isoform of pex5p, the human PTS1 receptor, is required for the import of PTS2 proteins into peroxisomes, *Hum. Mol. Genet.* **7**, 1195–1205.

Collings, D.A., Harper, J.D., Vaughn, K.C. (2003) The association of peroxisomes with the developing cell plate in dividing onion root cells depends on actin microfilaments and myosin, *Planta* **218**, 204–216.

Collins, C.S., Kalish, J.E., Morrell, J.C., McCaffery, J.M., Gould, S.J. (2000) The peroxisome biogenesis factors pex4p, pex22p, pex1p, and pex6p act in the terminal steps of peroxisomal matrix protein import, *Mol. Cell. Biol.* **20**, 7516–7526.

Dammai, V., Subramani, S. (2001) The human peroxisomal targeting signal receptor, Pex5p, is translocated into the peroxisomal matrix and recycled to the cytosol, *Cell* **105**, 187–196.

Danpure, C.J. (1993) Primary hyperoxaluria type 1 and peroxisome-to-mitochondrion mistargeting of alanine:glyoxylate aminotransferase, *Biochimie* **75**, 309–315.

De Duve, C. (1965) Functions of microbodies (peroxisomes), *J. Cell Biol.* **27**, 25A.

Distel, B., Erdmann, R., Gould, S.J., Blobel, G., Crane, D.I., Cregg, J.M., Dodt, G., Fujiki, Y.,

Goodman, J.M., Just, W.W., Kiel, J.A., Kunau, W.H., Lazarow, P.B., Mannaerts, G.P., Moser, H.W., Osumi, T., Rachubinski, R.A., Roscher, A., Subramani, S., Tabak, H.F., Tsukamoto, T., Valle, D., van der Klei, I., van Veldhoven, P.P., Veenhuis, M. (1996) A unified nomenclature for peroxisome biogenesis factors, *J. Cell Biol.* **135**, 1–3.

Dyer, J.M., McNew, J.A., Goodman, J.M. (1996) The sorting sequence of the peroxisomal integral membrane protein PMP47 is contained within a short hydrophilic loop, *J. Cell Biol.* **133**, 269–280.

Erdmann, R., Blobel, G. (1995) Giant peroxisomes in oleic acid-induced Saccharomyces cerevisiae lacking the peroxisomal membrane protein Pmp27p, *J. Cell Biol.* **128**, 509–523.

Erdmann, R., Veenhuis, M., Mertens, D., Kunau, W.H. (1989) Isolation of peroxisome-deficient mutants of Saccharomyces cerevisiae, *Proc. Natl. Acad. Sci. U. S. A.* **86**, 5419–5423.

Faber, K.N., Heyman, J.A., Subramani, S. (1998) Two AAA family peroxins, PpPex1p and PpPex6p, interact with each other in an ATP-dependent manner and are associated with different subcellular membranous structures distinct from peroxisomes, *Mol. Cell. Biol.* **18**, 936–943.

Fang, Y., Morrell, J.C., Jones, J.M., Gould, S.J. (2004) PEX3 functions as a PEX19 docking factor in the import of class I peroxisomal membrane proteins, *J. Cell Biol.* **164**, 863–875.

Furuya, T., Kessler, P., Jardim, A., Schnaufer, A., Crudder, C., Parsons, M. (2002) Glucose is toxic to glycosome-deficient trypanosomes, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14177–14182.

Geraghty, M.T., Bassett, D., Morrell, J.C., Gatto, G.J., Bai, J., Geisbrecht, B.V., Hieter, P., Gould, S.J. Jr. (1999) Detecting patterns of protein distribution and gene expression in silico, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 2937–2942.

Goldfischer, S., Moore, C.L., Johnson, A.B., Spiro, A.J., Valsamis, M.P., Wisniewski, H.K., Ritch, R.H., Norton, W.T., Rapin, I., Gartner, L.M. (1973) Peroxisomal and mitochondrial defects in the cerebro-hepato-renal syndrome, *Science* **182**, 62–64.

Goodman, J.M., Scott, C.W., Donahue, P.N., Atherton, J.P. (1984) Alcohol oxidase assembles post-translationally into the peroxisome of Candida boidinii, *J. Biol. Chem.* **259**, 8485–8493.

Gorgas, K. (1984) Peroxisomes in sebaceous glands. V. Complex peroxisomes in the mouse preputial gland: serial sectioning and three-dimensional reconstruction studies, *Anat. Embryol. (Berl)* **169**, 261–270.

Gotte, K., Girzalsky, W., Linkert, M., Baumgart, E., Kammerer, S., Kunau, W.H., Erdmann, R. (1998) Pex19p, a farnesylated protein essential for peroxisome biogenesis, *Mol. Cell Biol.* **18**, 616–628.

Gould, S.J., Keller, G.A., Hosken, N., Wilkinson, J., Subramani, S. (1989) A conserved tripeptide sorts proteins to peroxisomes, *J. Cell Biol.* **108**, 1657–1664.

Gould, S.J., Keller, G.A., Schneider, M., Howell, S.H., Garrard, L.J., Goodman, J.M., Distel, B., Tabak, H., Subramani, S. (1990) Peroxisomal protein import is conserved between yeast, plants, insects and mammals, *EMBO J.* **9**, 85–90.

Gouveia, A.M., Guimaraes, C.P., Oliveira, M.E., Reguenga, C., Sa-Miranda, C., Azevedo, J.E. (2003) Characterization of the peroxisomal cycling receptor, Pex5p, using a cell-free in vitro import system, *J. Biol. Chem.* **278**, 226–232.

Hannaert, V., Bringaud, F., Opperdoes, F.R., Michels, P.A. (2003) Evolution of energy metabolism and its compartmentation in Kinetoplastida, *Kinetoplastid Biol Dis* **2**, 11.

Harano, T., Nose, S., Uezu, R., Shimizu, N., Fujiki, Y. (2001) Hsp70 regulates the interaction between the peroxisome targeting signal type 1 (PTS1)-receptor Pex5p and PTS1, *Biochem. J.* **357**, 157–165.

Harper, C.C., Berg, J.M., Gould, S.J. (2003) PEX5 binds the PTS1 independently of Hsp70 and the peroxin PEX12, *J. Biol. Chem.* **278**, 7897–7901.

Hettema, E.H., Girzalsky, W., van Den Berg, M., Erdmann, R., Distel, B. (2000) Saccharomyces cerevisiae pex3p and pex19p are required for proper localization and stability of peroxisomal membrane proteins, *EMBO J.* **19**, 223–233.

Hettema, E.H., Ruigrok, C.C., Koerkamp, M.G., van den Berg, M., Tabak, H.F., Distel, B., Braakman, I. (1998) The cytosolic DnaJ-like protein djp1p is involved specifically in peroxisomal protein import, *J. Cell. Biol.* **142**, 421–434.

Hoepfner, D., van den Berg, M., Philippsen, P., Tabak, H.F., Hettema, E.H. (2001) A role for Vps1p, actin, and the Myo2p motor in peroxisome abundance and inheritance in

Saccharomyces cerevisiae, *J. Cell. Biol.* **155**, 979–990.

Hu, J., Aguirre, M., Peto, C., Alonso, J., Ecker, J., Chory, J. (2002) A role for peroxisomes in photomorphogenesis and development of Arabidopsis, *Science* **297**, 405–409.

Issemann, I., Green, S. (1990) Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators, *Nature* **347**, 645–650.

Jedd, G., Chua, N.H. (2000) A new self-assembled peroxisomal vesicle required for efficient resealing of the plasma membrane, *Nat. Cell. Biol.* **2**, 226–231.

Jedd, G., Chua, N.H. (2002) Visualization of peroxisomes in living plant cells reveals acto-myosin-dependent cytoplasmic streaming and peroxisome budding, *Plant Cell Physiol.* **43**, 384–392.

Kovacs, W.J., Shackelford, J.E., Tape, K.N., Richards, M.J., Faust, P.L., Fliesler, S.J., Krisans, S.K. (2004) Disturbed cholesterol homeostasis in a peroxisome-deficient PEX2 knockout mouse model, *Mol. Cell. Biol.* **24**, 1–13.

Lazarow, P.B. (1978) Rat liver peroxisomes catalyze the beta oxidation of fatty acids, *J. Biol. Chem.* **253**, 1522–1528.

Lazarow, P.B., de Duve, C. (1973) The synthesis and turnover of rat liver peroxisomes. V. Intracellular pathway of catalase synthesis, *J. Cell. Biol.* **59**, 507–524.

Li, X., Gould, S.J. (2003) The dynamin-like GTPase DLP1 is essential for peroxisome division and is recruited to peroxisomes in part by PEX11, *J. Biol. Chem.* **278**, 17012–17020.

Mano, S., Nakamori, C., Hayashi, M., Kato, A., Kondo, M., Nishimura, M. (2002) Distribution and characterization of peroxisomes in Arabidopsis by visualization with GFP: dynamic morphology and actin-dependent movement, *Plant Cell Physiol.* **43**, 331–341.

Marshall, P.A., Krimkevich, Y.I., Lark, R.H., Dyer, J.M., Veenhuis, M., Goodman, J.M. (1995) Pmp27 promotes peroxisomal proliferation, *J. Cell. Biol.* **129**, 345–355.

Marzioch, M., Erdmann, R., Veenhuis, M., Kunau, W.H. (1994) PAS7 encodes a novel yeast member of the WD-40 protein family essential for import of 3-oxoacyl-CoA thiolase, a PTS2-containing protein, into peroxisomes, *EMBO J.* **13**, 4908–4918.

Mathur, J., Mathur, N., Hulskamp, M. (2002) Simultaneous visualization of peroxisomes and cytoskeletal elements reveals actin and not microtubule-based peroxisome motility in plants, *Plant Physiol.* **128**, 1031–1045.

McCammon, M.T., Veenhuis, M., Trapp, S.B., Goodman, J.M. (1990) Association of glyoxylate and beta-oxidation enzymes with peroxisomes of Saccharomyces cerevisiae, *J. Bacteriol.* **172**, 5816–5827.

McNew, J.A., Goodman, J.M. (1994) An oligomeric protein is imported into peroxisomes in vivo, *J. Cell. Biol.* **127**, 1245–1257.

Muller, W.H., Bovenberg, R.A., Groothuis, M.H., Kattevilder, F., Smaal, E.B., Van der Voort, L.H., Verkleij, A.J. (1992) Involvement of microbodies in penicillin biosynthesis, *Biochim. Biophys. Acta* **1116**, 210–213.

Nishimura, M., Hayashi, M., Kato, A., Yamaguchi, K., Mano, S. (1996) Functional transformation of microbodies in higher plant cells, *Cell. Struct. Funct.* **21**, 387–393.

Osumi, M., Miwa, N., Teranishi, Y., Tanaka, A., Fukui, S. (1974) Ultrastructure of Candida yeasts grown on n-alkanes. Appearance of microbodies and its relationship to high catalase activity, *Arch. Microbiol.* **99**, 181–201.

Parsons, M. (2004) Glycosomes: parasites and the divergence of peroxisomal purpose, *Mol. Microbiol.* **53**, 717–724.

Petriv, O.I., Tang, L., Titorenko, V.I., Rachubinski, R.A. (2004) A new definition for the consensus sequence of the peroxisome targeting signal type 2, *J. Mol. Biol.* **341**, 119–134.

Pires, J.R., Hong, X., Brockmann, C., Volkmer-Engert, R., Schneider-Mergener, J., Oschkinat, H., Erdmann, R. (2003) The ScPex13p SH3 domain exposes two distinct binding sites for Pex5p and Pex14p, *J. Mol. Biol.* **326**, 1427–1435.

Platta, H.W., Girzalsky, W., Erdmann, R. (2004) Ubiquitination of the peroxisomal import receptor Pex5p, *Biochem. J.* **384**, 37–45.

Reuber, B.E., Germain-Lee, E., Collins, C.S., Morrell, J.C., Ameritunga, R., Moser, H.W., Valle, D., Gould, S.J. (1997) Mutations in PEX1 are the most common cause of peroxisome biogenesis disorders, *Nat. Genet.* **17**, 445–448.

Rhodin, J. (1954) *Ultrastructural Organization and Function in Normal and Experimentally Changed Proximal Convoluted Tubule Cells of the Mouse*, Karolinska Institutet, Stockholm, Sweden.

Rouiller, C., Bernhard, W. (1956) "Microbodies" and the problem of mitochondrial regeneration in liver cells, *J. Biophys. Biochem. Cytol.* **2**(4 suppl), 355.

Sacksteder, K.A., Jones, J.M., South, S.T., Li, X., Liu, Y., Gould, S.J. (2000) PEX19 binds multiple peroxisomal membrane proteins, is predominantly cytoplasmic, and is required for peroxisome membrane synthesis, *J. Cell. Biol.* **148**, 931–944.

Sahm, H., Roggenkamp, R., Wagner, F., Hinkelmann, W. (1975) Microbiodies in methanol-grown Candida boidinii, *J. Gen. Microbiol.* **88**, 218–222.

Santos, M.J., Imanaka, T., Shio, H., Small, G.M., Lazarow, P.B. (1988) Peroxisomal membrane ghosts in Zellweger syndrome–aberrant organelle assembly, *Science* **239**, 1536–1538.

Schafer, H., Nau, K., Sickmann, A., Erdmann, R., Meyer, H.E. (2001) Identification of peroxisomal membrane proteins of Saccharomyces cerevisiae by mass spectrometry, *Electrophoresis* **22**, 2955–2968.

Schrader, M., Reuber, B.E., Morrell, J.C., Jimenez-Sanchez, G., Obie, C., Stroh, T.A., Valle, D., Schroer, T.A., Gould, S.J. (1998) Expression of PEX11beta mediates peroxisome proliferation in the absence of extracellular stimuli, *J. Biol. Chem.* **273**, 29607–29614.

Schrader, M., Thiemann, M., Fahimi, H.D. (2003) Peroxisomal motility and interaction with microtubules, *Microsc. Res. Tech.* **61**, 171–178.

Snyder, W.B., Koller, A., Choy, A.J., Subramani, S. (2000) The peroxin Pex19p interacts with multiple, integral membrane proteins at the peroxisomal membrane, *J. Cell. Biol.* **149**, 1171–1178.

South, S.T., Gould, S.J. (1999) Peroxisome synthesis in the absence of preexisting peroxisomes, *J. Cell. Biol.* **144**, 255–266.

South, S.T., Baumgart, E., Gould, S.J. (2001) Inactivation of the endoplasmic reticulum protein translocation factor, Sec61p, or its homolog, Ssh1p, does not affect peroxisome biogenesis, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 12027–12031.

South, S.T., Sacksteder, K.A., Li, X., Liu, Y., Gould, S.J. (2000) Inhibitors of COPI and COPII do not block PEX3-mediated peroxisome synthesis, *J. Cell. Biol.* **149**, 1345–1360.

Swinkels, B.W., Gould, S.J., Bodnar, A.G., Rachubinski, R.A., Subramani, S. (1991) A novel, cleavable peroxisomal targeting signal at the amino-terminus of the rat 3-ketoacyl-CoA thiolase, *EMBO J.* **10**, 3255–3262.

Titorenko, V.I., Rachubinski, R.A. (1998) Mutants of the yeast Yarrowia lipolytica defective in protein exit from the endoplasmic reticulum are also defective in peroxisome biogenesis, *Mol. Cell. Biol.* **18**, 2789–2803.

Titorenko, V.I., Chan, H., Rachubinski, R.A. (2000) Fusion of small peroxisomal vesicles in vitro reconstructs an early step in the in vivo multistep peroxisome assembly pathway of Yarrowia lipolytica, *J. Cell. Biol.* **148**, 29–44.

Titorenko, V.I., Ogrydziak, D.M., Rachubinski, R.A. (1997) Four distinct secretory pathways serve protein secretion, cell surface growth, and peroxisome biogenesis in the yeast Yarrowia lipolytica, *Mol. Cell. Biol.* **17**, 5210–5226.

Titus, D.E., Becker, W.M. (1985) Investigation of the glyoxysome-peroxisome transition in germinating cucumber cotyledons using double-label immunoelectron microscopy, *J. Cell. Biol.* **101**, 1288–1299.

van Dijken, J.P., Veenhuis, M., Kreger-van Rij, N.J., Harder, W. (1975) Microbodies in methanol-assimilating yeasts, *Arch. Microbiol.* **102**, 41–44.

Veenhuis, M., Goodman, J.M. (1990) Peroxisomal assembly: membrane proliferation precedes the induction of the abundant matrix proteins in the methylotrophic yeast Candida boidinii, *J. Cell. Sci.* **96**(4), 583–590.

Veenhuis, M., Mateblowski, M., Kunau, W.H., Harder, W. (1987) Proliferation of microbodies in Saccharomyces cerevisiae, *Yeast* **3**, 77–84.

Vizeacoumar, F.J., Torres-Guzman, J.C., Bouard, D., Aitchison, J.D., Rachubinski, R.A. (2004) Pex30p, Pex31p, and Pex32p form a family of peroxisomal integral membrane proteins regulating peroxisome size and number in Saccharomyces cerevisiae, *Mol. Biol. Cell.* **15**, 665–677.

Walton, P.A., Hill, P.E., Subramani, S. (1995) Import of stably folded proteins into peroxisomes, *Mol. Biol. Cell.* **6**, 675–683.

Wang, X., McMahon, M.A., Shelton, S.N., Nampaisansuk, M., Ballard, J.L., Goodman, J.M. (2004) Multiple targeting modules on peroxisomal proteins are not redundant: discrete functions of targeting signals within

Pmp47 and Pex8p, *Mol. Biol. Cell.* **15**, 1702–1710.

Waterham, H.R., Titorenko, V.I., Swaving, G.J., Harder, W., Veenhuis, M. (1993) Peroxisomes in the methylotrophic yeast Hansenula polymorpha do not necessarily derive from preexisting organelles, *EMBO J.* **12**, 4785–4794.

Zolman, B.K., Monroe-Augustus, M., Thompson, B., Hawes, J.W., Krukenberg, K.A., Matsuda, S.P., Bartel, B. (2001) chy1, an Arabidopsis mutant with impaired beta-oxidation, is defective in a peroxisomal beta-hydroxyisobutyryl-CoA hydrolase, *J. Biol. Chem.* **276**, 31037–31046.

Structure-based Drug Design and NMR-based Screening

David J. Craik and Richard J. Clark
University of Queensland, Brisbane, Australia

1 Introduction 520
1.1 Overview of Drug Development 520
1.2 Scope of Chapter 522
1.3 Principles of NMR Spectroscopy 523
1.4 Instrumentation 526
1.5 Applications of NMR in Drug Design and Discovery 528

2 Ligand-based Design 529
2.1 Structure Elucidation 530
2.1.1 Structure Elucidation of Natural Products 530
2.1.2 Structure Determination of Bioactive Peptides 531
2.1.3 Instrumental Advances and their Impact on Structure Elucidation 538
2.2 Conformational Analysis 539
2.3 Charge State 540
2.4 Tautomeric Equilibria 541
2.5 Ligand Dynamics: Line-shape and Relaxation Data 542
2.6 Pharmocophore Modeling: Conformations of a Set of Ligands 546
2.7 Limitations of Analog-based Design 547
2.8 Conformation of Bound Ligands: Transferred NOEs 547

3 Receptor-based Design 548
3.1 Macromolecular Structure Determination 548
3.1.1 Overview of Approach 549
3.1.2 Sample Requirements and Assignment Protocols 549
3.1.3 Recent Developments 550
3.1.4 Dynamics 550
3.1.5 Nucleic Acid Structures 550
3.1.6 Challenges for the Future: Membrane-bound Proteins 551
3.2 Macromolecule–Ligand Interactions 551

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

3.2.1 Overview 551
3.2.2 Influence of Kinetics and NMR Timescales 552
3.2.3 NMR Techniques 561
3.2.4 Selected Examples 565

4 NMR Screening 582
4.1 Methods 583
4.1.1 Chemical-shift Perturbation 583
4.1.2 Magnetization Transfer Experiments 589
4.1.3 Molecular Diffusion 591
4.1.4 Relaxation 592
4.1.5 NOE 593
4.1.6 Spin Labels 594
4.2 Practical Considerations 595
4.2.1 Screening Approach 595
4.2.2 Library Design 596
4.2.3 Hardware and Automation 598

5 Conclusions 599

Acknowledgment 599

Bibliography 599
Books and Reviews 599
Primary Literature 600

Keywords

Chemical Shift
A very important NMR parameter that provides information on the local electronic environment of nuclei in molecules. Chemical shifts are measured on the frequency axes of NMR spectra.

Chemical Shift Mapping
An approach to characterizing interactions between molecules by determining which sites have their chemical shifts perturbed when a ligand binds to a macromolecule. Since chemical shifts reflect the local environment, only signals close to the binding site are perturbed following ligand binding or conformational changes.

HSQC Spectroscopy
Short for heteronuclear single quantum coherence spectroscopy. This is a very powerful 2D NMR method that provides a useful ‘‘fingerprint’’ spectrum that can be used in binding studies.

Isotope Labeling
The use of stable isotopes (e.g. ^{13}C, ^{15}N) to enhance the prevalence of NMR-active nuclei in a protein to allow a sophisticated range of multinuclear NMR experiments to be carried out. Labeled proteins are easily produced using recombinant methods.

Multidimensional NMR
NMR experiments that result in spectra having multiple frequency axes so that spectral overlap is reduced. One-dimensional NMR spectra have one frequency axis (e.g. ^{1}H chemical shift) and one intensity axis, while in 2D NMR there are two frequency axes, with peak intensities normally represented in contour form. For example, in 2D HSQC spectra, the frequency axes often correspond to ^{1}H and ^{15}N chemical shifts.

NMR Screening
A generic term for a range of NMR techniques that identify ligands (usually as part of a library of compounds) that bind to a target macromolecule.

NOE (Nuclear Overhauser Effect)
An NMR parameter that provides information on distances between nuclei. This distance information in turn can be used to determine the three-dimensional structures of molecules, particularly peptides and proteins. NOEs are derived from intensities in NOESY spectra.

Pulse Sequence
An array of radio-frequency pulses that is used to perturb the nuclei in a sample in a defined way so as to produce a particular type of NMR spectrum. For example, different pulse sequences are used to generate NOESY or HSQC spectra.

Structure-based Design
An approach to designing drugs on the basis of knowledge of the structure of a target macromolecule and/or macromolecule–ligand complexes.

NMR spectroscopy is a powerful technique for determining the structures of molecules in solution. It involves placing a sample in a powerful magnetic field and irradiating it with radio waves of defined frequency. The emitted signals provide information about the local molecular environments of nuclei in the sample, from which structures can be derived. NMR can also be used for determining interactions between molecules and is particularly useful for determining the nature of binding interactions between ligands and macromolecules. This information is very important in drug design: by understanding how bioactive molecules interact with a target protein or nucleic acid it is, in principle, possible to design ligands with improved affinity and specificity, which may make useful drug leads. In addition to this structure-based approach to drug design, NMR is also useful as a screening tool in drug discovery programs to identify ligands that bind to target macromolecules.

1 Introduction

NMR spectroscopy has been widely used as a frontline tool in the pharmaceutical industry for several decades. In the past, the main use of NMR was in the structural characterization of organic molecules synthesized in the course of medicinal chemistry programs. Indeed, medicinal chemists have long regarded NMR as the premier tool for structure characterization, to confirm the identity of intermediates, or to determine the conformation of lead molecules. Over the last decade, major developments in both instrumentation and methods have resulted in this traditional use of NMR in the pharmaceutical industry being augmented by a range of exciting new applications. Two of the most important of these are the use of NMR in structure-based drug design and in screening for drug discovery. Both applications differ from the traditional use of NMR in that the macromolecular binding partner of the medicinal compound is now included in the mixture to be analyzed; that is, current applications of NMR in drug discovery are predominantly focused on the *interaction* between drug molecules and their macromolecular targets.

The aim of this chapter is to describe how NMR spectroscopy is used in modern drug discovery. The term *discovery* is used generically throughout to include processes that involve rational drug design as well as those that involve discovery through NMR screening. The latter is a relatively recent development and refers to the use of NMR as a tool to screen a compound library to identify molecules that bind to a chosen macromolecular target. Of course, the distinction between "design" and "discovery" is often quite blurred. This is nowhere more evident than in the recently developed SAR-by-NMR approach, in which the discovery of several weakly bound ligands from a screening program is intimately linked to a design process to chemically join them. SAR-by-NMR represents an exciting new technique for lead generation and is described in more detail later in this chapter.

Drug design/discovery represents only the first stage in the drug development process. There are many other steps that need to be taken once a lead molecule has been designed or discovered. Although other stages of the process, including lead optimization, toxicity studies, preclinical investigations, and clinical monitoring, do not fall within the scope of this chapter, it is worth mentioning that NMR spectroscopy contributes significantly across the whole spectrum of drug development, right through into the clinical domain. For example, NMR spectroscopy has been applied for the detection of drug metabolites in biological fluids and magnetic resonance imaging, which is based on the fundamental principles of NMR, plays an important role in clinical investigations. It is increasingly being used to monitor the functional outcomes of drug therapy. We will briefly address these broader applications of NMR before returning to the main topic of NMR in drug discovery.

1.1 Overview of Drug Development

To give an overview of the breadth of applications of NMR, Fig. 1 summarizes the drug development process and indicates the role of NMR at various stages. Drug development is an iterative process and can be simplified by representing it with

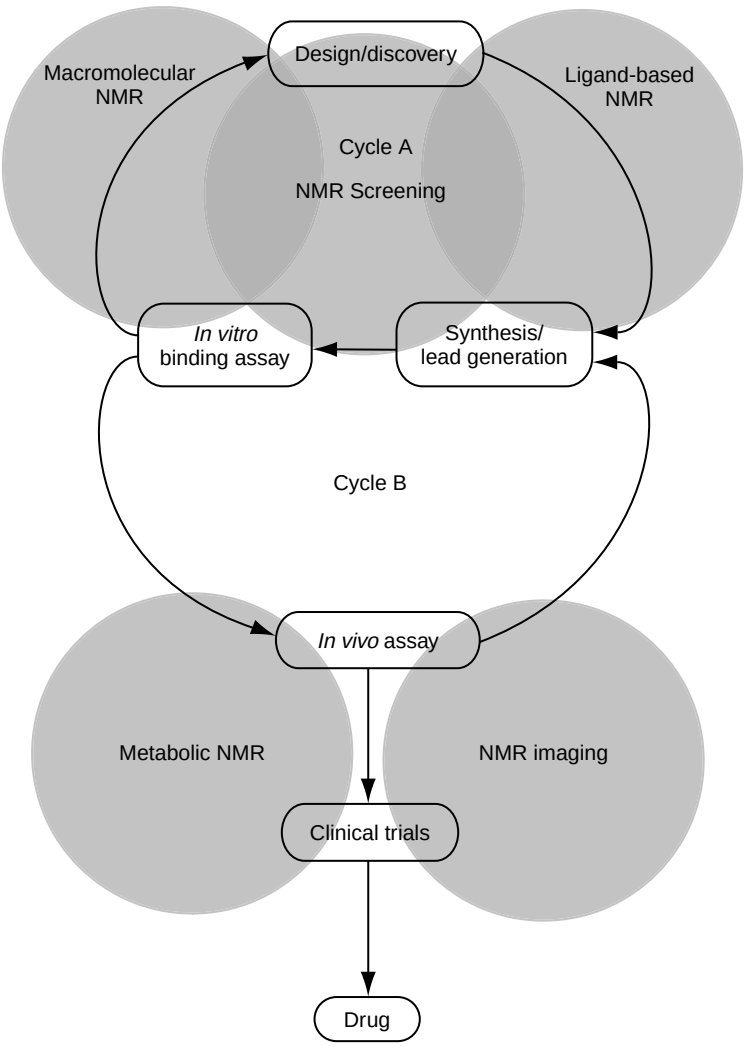

Fig. 1 Overview of the drug development process and summary of various types of NMR experiments that contribute at different stages.

two interconnected cycles of activity. Cycle A involves the design or discovery of an initial lead followed by its synthesis and bioassay. On the basis of the initial assay results, there may be several loops around this cycle before commencing the *in vivo* studies represented in Cycle B. At this stage, consideration of bioavailability, metabolism, and pharmacokinetic profiles must be made and this may involve synthetic modifications of the lead molecules to improve their druglike properties. Again, several loops around Cycle B may be necessary before one or more development candidates are identified. Ultimately one or more of these development candidates are identified for progression through clinical trials.

As indicated in Fig. 1, it is convenient to envisage five broad categories of NMR experiments that may contribute to this overall drug development process.

1. *Small molecule, or ligand-based, NMR.* This involves studies of drugs and drug leads, typically organic molecules with a molecular weight <500 Da, but also including small proteins of up to a few kDa. These studies may be used to characterize natural products or synthetic drug leads, or to determine their conformation.
2. *Macromolecular NMR.* This involves studies of the macromolecular targets of drugs, typically to determine their three-dimensional structure and/or the nature of their complexes with ligands.
3. *NMR screening.* This involves the use of NMR to identify lead molecules that bind to macromolecular targets. These studies typically involve both small molecules and macromolecules and seek to detect the presence of binding interactions between them.
4. *Metabolic NMR.* This involves studies of endogenous molecules whose concentrations may be modified by drug treatment, or studies of the metabolites of drugs themselves.
5. *NMR imaging.* Such studies provide anatomical information in an animal model or human patient. This includes, for example, monitoring the size of plaques or tumors in the brains of Alzheimer's or cancer patients, respectively, during drug therapy.

It is clear from these descriptions that NMR covers a wide range of applications in the pharmaceutical industry, although for the remainder of this chapter we will focus on NMR in the drug design/discovery phase of drug development, that is, on categories 1 to 3 of the preceding list. Together, the studies in categories 1 and 2 may be classified as structure-based design, whereas category 3 relates to drug discovery.

1.2 Scope of Chapter

Our aim is to give a broad overview on the use of NMR as a tool in structure-based design and in screening approaches to drug discovery. This chapter also contains a description of the relevant NMR methods, which are highlighted by illustrative examples. We briefly describe the instrumentation required for such studies and discuss emerging trends in the field. This includes developments in the field of drug discovery in the postgenomic era, which are likely to have an impact on the way in which NMR is used, as seen, for example, by the recent interest in structural genomics programs. NMR instrument developments are also described. For example, recent advances in cryoprobe technology promise to dramatically increase the sensitivity of NMR spectroscopy and also increase its application across the pharmaceutical industry. Finally, a section outlining some of the practical considerations in structure-based design and screening is included. Future directions in this field are mentioned throughout the discussion.

There are a number of reviews that describe applications of NMR in drug discovery or screening, and the reader is referred to these for additional information. Recent books covering aspects of NMR in drug design are also available.

It is assumed that most readers would be familiar with the basic principles of NMR. However, for completeness and to define some of the terms that will be used in this

chapter it is useful to give a brief overview of the principles.

1.3 Principles of NMR Spectroscopy

The underlying basis of NMR is that when nuclei with a nonzero spin quantum number are placed in a magnetic field, they take up one of a discrete number of quantized states. The application of radio-frequency (rf) energy produces transitions between these states. The energy changes associated with these transitions are detected as small voltages induced in a receiver coil, which are subsequently amplified, digitized, and processed to yield spectra, as illustrated in Fig. 2. The most commonly studied NMR-active nucleus is the proton, ^{1}H, but in modern NMR experiments ^{2}H, ^{13}C, and ^{15}N nuclei are also very important. For these heteronuclei, it is common to isotopically enrich the sample because of their low natural abundance. This is particularly important for studies of proteins, as will become apparent later in this chapter. Occasionally, other nuclei find specialist applications. For example, in fluorine-containing drugs it is possible to use sensitive ^{19}F-NMR signals to monitor interaction with target proteins, as described later in this chapter.

In modern spectrometers, the rf energy is supplied in the form of short pulses (typically, ~10 μs) that simultaneously excite all nuclei of a given isotope type (e.g. all protons or all ^{13}C nuclei). Nuclei of a given isotope, which are in different chemical environments by virtue of their atomic locations in the molecule, have slightly different resonance frequencies and lead to different oscillating voltages in the receiver coil. The resultant combined signal, termed a *free induction decay* (FID), is Fourier transformed to give a spectrum that is basically a plot of peak intensity versus frequency, with one peak for each chemically distinct nucleus. These features are schematically illustrated in Fig. 2. The frequency axis is termed the *chemical shift* because it reflects the local chemical environment of each nucleus. The range of chemical environments of nuclei in a molecule is such that chemical shifts range up to only a few hundred parts per million (ppm) of the base resonance frequency for ^{13}C and ^{15}N. For ^{1}H the range is smaller, covering only about 10 ppm. Despite this small range, chemical shifts provide valuable diagnostic information on

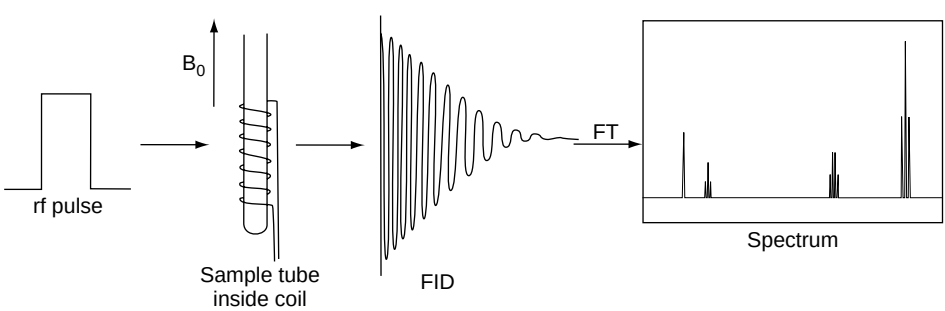

Fig. 2 Overview of the principles of NMR spectroscopy. Polarization of nuclear spins by a magnetic field is perturbed by application of a radio-frequency (rf) pulse. The resultant signal is Fourier transformed, to yield a spectrum reflecting the number and environments of nuclei in the sample.

the environment of the nucleus that is giving rise to the signal.

The chemical shift is an extremely important NMR parameter, but there are many other parameters that can be discerned from NMR spectra. Indeed, NMR is unique among many forms of spectroscopy in that there are so many parameters associated with a spectrum other than just peak intensity and frequency. These include coupling constants, which provide information on local conformations and also on molecular connectivities; nuclear Overhauser effects (NOEs), which provide information on internuclear distances; and relaxation parameters, which provide information on molecular dynamics. Table 1 summarizes the main NMR parameters that may be measured and highlights their applications in the drug discovery process.

The following sections of this chapter provide specific examples of how these various parameters are useful in the drug discovery process. Before doing this, though, it is useful to consider some of the limitations of one-dimensional (1D) NMR spectroscopy, particularly when the detected nucleus is ^{1}H, as is most commonly the case. With one signal coming from each chemically distinct proton and with those signals spread only over 10 ppm, it is clear that spectral overlap can potentially be a major problem for anything but the simplest of molecules. The development of higher field NMR spectrometers, which effectively provide greater dispersion in the frequency dimension, has contributed significantly to overcoming this limitation and increasing the application of NMR for studying pharmaceutically relevant molecules. In addition to such instrumental developments, methodological advances have also played a key role in extending the use of NMR. Multidimensional NMR methods have revolutionized biomolecular NMR spectroscopy by removing the limitations of a single frequency dimension, leading to the development of 2D, 3D, and 4D spectra.

A simple way of illustrating multidimensional NMR is through reference to heteronuclear correlation spectroscopy, in which two or more separate frequency dimensions are correlated with one another. For example, a particularly valuable 2D experiment is $^{1}H-^{15}N$ heteronuclear single quantum correlation (HSQC)

Tab. 1 NMR parameters and their applications in drug design/discovery.

Parameter	*Information provided relevant to drug design*
Chemical shift	Reflects local chemical environment; provides a fingerprint marker of structure (particularly in HSQC spectra)
Coupling constants	Conformational analysis, establishing molecular connectivity
Nuclear Overhauser effect	Determining interproton distances, three-dimensional structures
Relaxation times	Molecular dynamics
Line shape	Detecting and quantifying chemical-exchange processes
Peak intensities	Reflect relative number of nuclei, molecular symmetry
Amide-exchange rates/temperature coefficients	Hydrogen bonding or solvent exposure of amide protons

spectroscopy, in which the resultant spectrum has two frequency axes, corresponding to ^{1}H and ^{15}N frequency dimensions, and one intensity axis. Analogous $^{1}H-^{13}C$ HSQC spectra are also widely used. Such spectra are normally represented with the intensity axis in contour form so that they may be drawn in two dimensions as a set of contour peaks. Spectral peaks occur for pairs of $^{15}N/^{1}H$ or $^{13}C/^{1}H$ nuclei that are directly bonded to one another, and with each frequency being characteristic for the local chemical environment, HSQC spectra represent a relatively simple but highly characteristic fingerprint of the sample. Figure 3 shows the relationship between 1D and 2D spectra for the immunosuppressive drug cyclosporin, and includes a region of both the $^{1}H/^{15}N$ and $^{1}H/^{13}C$ HSQC spectra. In HSQC spectra, overlap problems are alleviated because, even if two protons have the same chemical shift and would hence be overlapped in a 1D spectrum, chances are that the respective heteronuclear signals will not be overlapped, allowing the signals to be resolved in the 2D spectrum. HSQC spectra are widely used in NMR-based drug screening and we will return to them later.

Multidimensional NMR spectra are not restricted to cases where the separate frequency axes encode signals from different nuclear types. Indeed, much of the early work on the development of 2D NMR was performed on cases where both axes involved ^{1}H chemical shifts. The main value in such spectra comes from the information content in cross peaks between pairs of protons. In COSY-type spectra (COSY = COrrelation SpectroscopY), cross peaks occur only between protons

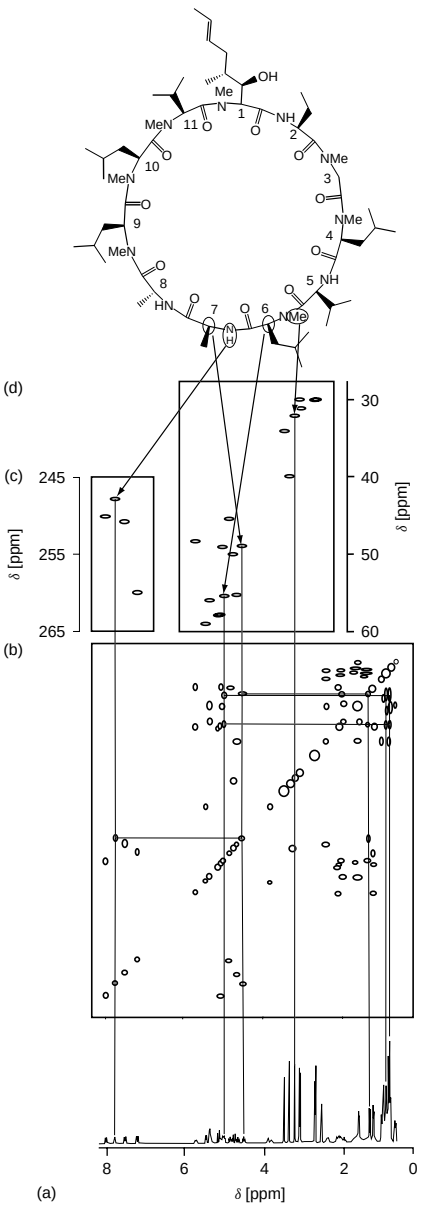

Fig. 3 A schematic representation of the (a) 1D ^{1}H; (b) 2D DQF-COSY; (c) $^{15}N/^{1}H$-HSQC; and (d) $^{13}C/^{1}H$-HSQC spectra of the immunosuppressive agent cyclosporin. Example resonances/correlations from residues 6 and 7 have been highlighted to illustrate the assignment process.

that are scalar coupled (i.e. within 2 or 3 bonds) to each other, whereas in NOESY (Nuclear Overhauser Effect SpectroscopY), spectra cross peaks occur for protons that are physically close in space (<5 Å apart). A combination of these two types of 2D spectra may be used to assign the NMR signals of small proteins and provides sufficient information on internuclear distances to calculate three-dimensional structures. Figure 3 includes a panel showing the COSY spectrum of cyclosporin and highlights the relationships between 1D ^{1}H-NMR spectra and corresponding 2D homonuclear (COSY) and heteronuclear (HSQC) spectra.

Homonuclear 2D spectra are generally applicable for the study of proteins up to approximately 80 amino acids in size. For larger proteins, the increased number of signals leads to overlap problems and, in addition, COSY-type spectra suffer from poor sensitivity when the signal linewidths are of the same order as or larger than ^{1}H, ^{1}H scalar coupling constants. Such limitations are reduced by the use of spectra of higher dimensionality (i.e. 3D or 4D spectra) that are based on correlations involving heteronuclear rather than homonuclear coupling constants. Such spectra are important in the structure determination process for larger proteins and are typically recorded for samples that incorporate uniform labeling with ^{15}N, or both ^{13}C and ^{15}N nuclei. Multidimensional spectra that involve irradiation of ^{1}H, ^{13}C, and ^{15}N nuclei are referred to as *triple resonance spectra*.

The details of how multidimensional spectra are obtained is beyond the scope of this chapter, but it suffices to say that, like most other modern NMR experiments, they involve irradiation of the sample with a set of rf pulses of defined length, frequency, and phase, with specific interpulse delays. The pulse programs for such experiments are commonly provided with the spectrometer as part of a standard library of experiments and may easily be run by novice users after input of an appropriate set of parameters to define the relevant spectral widths and type of experiment required.

The above discussion provides a basic overview of some of the methods important in modern NMR spectroscopy. Before examining specific applications in drug discovery, it is useful to describe the instrumental requirements for such studies.

1.4 Instrumentation

NMR spectrometers constitute a powerful and homogeneous magnet, a radio-frequency console for generating appropriate rf pulses, a probe for applying this rf energy to the sample and receiving the resultant signals, and a computer console for controlling the experiments and acquiring the resultant data. These features are summarized in Fig. 4. Spectrometers are normally specified in terms of the resonant frequency of protons at the given magnetic field (e.g. 500 MHz corresponds to a magnetic field of 11.7 T). Both sensitivity and dispersion of signals increase with increasing magnetic field.

There have been some major breakthroughs in both NMR instrumentation and methodology over the last decade, which has greatly increased the utility of NMR for drug discovery applications. These are summarized in Table 2, which also includes some of the earlier milestones in the development of NMR. Most notable among recent innovations are the use of pulsed-field gradient methods for improving spectral quality and allowing new types of experiments

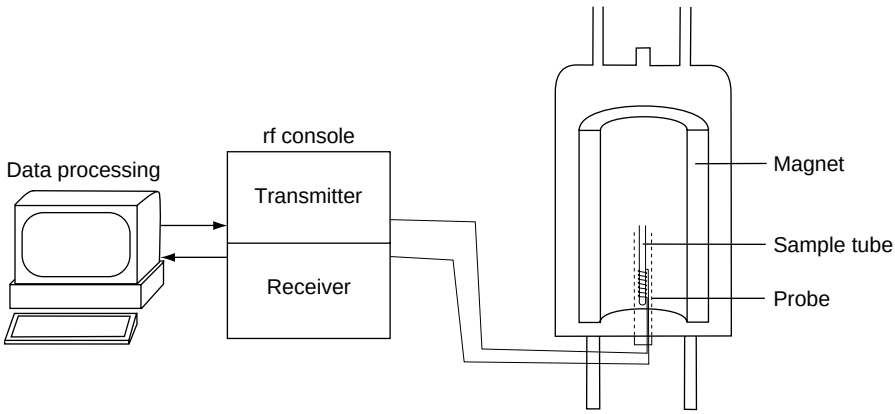

Fig. 4 Block diagram of a modern NMR spectrometer. These systems use superconducting magnets that are based on a solenoid of a suitable alloy (e.g. niobium/titanium or niobium/tin) immersed in a dewar of liquid helium. The extremely low temperature of the magnet itself (4.2 K) is well insulated from the sample chamber in the center of the magnet bore. The probe in which the sample is housed usually incorporates accurate temperature control over the range typically of 4 to 40 °C for biological samples. The rf coil in the probe is connected in turn to a preamplifier, receiver circuitry, analog-to-digital converter (ADC), and a computer for data collection.

Tab. 2 Milestones in the development of NMR spectroscopy.

Year	*Development*	*Nature*
1970	FT NMR	Instrumental
1975	Superconducting magnets	Instrumental
1980	2D NMR	Methodological
1985	Protein structure determination	Methodological
1990	Isotope labeling/multidimensional NMR	Methodological
1990	Pulsed-field gradients	Instrumental/methodological
1995	NMR screening	Methodological
1997	TROSY	Methodological
1998	LC-NMR/ LCMS-NMR	Instrumental
2000	Cryoprobes	Instrumental

to be performed, transverse relaxation-optimized spectroscopy (TROSY) methods for increasing the size of macromolecules that can be examined, and cryoprobes for enhancing sensitivity. The development of cryoprobes has resulted in the biggest single gain in sensitivity over recent years, effectively giving 500-MHz spectrometers the sensitivity of 800-MHz spectrometers (although without the gain in resolution!). The enhanced sensitivity is obtained by cooling the receiver coil and associated circuitry to near liquid helium temperatures, thereby reducing thermal noise. There were considerable technical barriers to be overcome in developing such probes because of the large difference in temperature between the receiver coils and the sample, which are only a few millimeters apart. These barriers have now

been overcome and cryoprobes are being installed in a large number of laboratories. They are also becoming available for higher field systems (900 MHz), thus providing further sensitivity gains.

Although the basic configurations of instruments tailored for structure-based design or for NMR drug screening are similar, there are some minor differences. For structure-based design applications, a relatively high field spectrometer is required (>500 MHz), usually equipped with three or four radio-frequency channels for the simultaneous irradiation of ^{1}H, ^{13}C, ^{15}N, and in some cases ^{2}H nuclei. The greatest sensitivity and dispersion are obtained with the highest possible magnetic field. Instruments of up to 900 MHz are currently available but, at the time of writing, only a few have been installed. Numerous 800-MHz systems dedicated to structure-based design have been installed in pharmaceutical laboratories. The high field instruments provide another advantage in that TROSY experiments can be used to produce a marked improvement in spectral quality for larger proteins. Such developments promise to push higher the size of proteins whose structure can be determined by NMR.

For NMR drug screening programs, the basic requirement of a spectrometer of 500 MHz or greater remains, but in addition, an interface that allows the spectrometer to sample a library of compounds of potential binding ligands needs to be present. This may be done either by use of a discrete sample changer or a flow-type system. Flow systems have the potential advantage of increased throughput and also have the potential disadvantage of precipitation of protein samples. In practice, this appears not to have been a major problem and both types of systems are in use in the pharmaceutical industry. Cryoprobes allow dramatically enhanced sensitivity gains, which bring particular advantages to the study of macromolecule–ligand interactions used in screening programs.

Pulsed-field gradients have become integral to most modern NMR spectrometers and are routinely used both for structure determination and screening experiments. Another recent development has been the interface of NMR spectrometers with other instrumentation such as liquid chromatography (LC) and/or mass spectrometry (MS).

1.5 Applications of NMR in Drug Design and Discovery

Our focus here is on the use of NMR in the discovery and design phase of drug development. The major role of NMR in the *design* process comes about by its exquisite ability to provide structural information, whereas the major role of NMR in *discovery* comes through its use as a screening tool to detect the binding of novel ligands to macromolecular targets. As already noted, the latter application is a relatively recent development but has created much interest in the pharmaceutical industry and promises to significantly enhance applications of NMR in this industry. The impact of the methodology is already becoming evident even at this early stage, with several SAR-by-NMR-derived leads currently in clinical development. As already noted, though, the discovery and design phases are often intimately connected, with lead molecules discovered in screening programs routinely being optimized by use of structure-based design approaches (Fig. 5).

In the context of this chapter, structure-based design refers to the process of

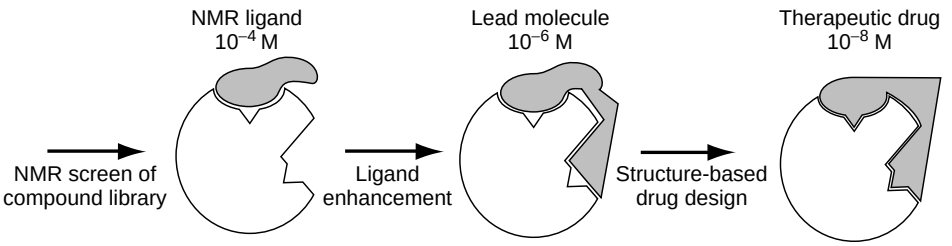

Fig. 5 A summary of the relationship between NMR screening and structure-based design. (Adapted from Peng, J.W, Lepre, C.A., Fejzo, J., Abdul-Manan, N., Moore, J.M. (2001) *Methods Enzymol.* **338**, 202.)

determining the three-dimensional structure of a lead molecule or macromolecular target, or determining the structure of the macromolecule–ligand complex, and using this information to design new drugs. The questions that may be asked when embarking on structure-based design projects are:

- What are the solution and bound conformations of the ligand?
- What is its charge/tautomeric state?
- Which functional groups bind to the receptor and what charge state are they in?
- What is the structure of the receptor?
- Which parts interact with the ligand?
- What is the geometry of the ligand–receptor complex?
- What are the kinetics of binding and are there dynamic motions of ligand, receptor, or the complex?

Table 3 summarizes these and other questions and indicates the type of NMR approaches that can provide answers. Remaining sections of this chapter are organized around the headings identified in Table 3.

In considering these questions, it is convenient to distinguish between ligand-based design, where the structural focus is on the small lead molecule, and receptor-based design, where the aim is to determine the structure of the macromolecular target. The NMR methods used in ligand-based design have been well established for many years, based on the use of NMR by organic and natural product chemists for more than four decades. However, there have been some important recent advances in NMR methods such as the use of pulsed-field gradients, and in the combination of NMR with other technologies such as LC and MS, which promise to enhance applications in this field. The use of NMR to determine the three-dimensional structures of macromolecules is a newer field, which commenced only in 1985 and is still rapidly evolving. NMR screening is a still newer approach, developed around 1996. Ligand-based and receptor-based design are examined in Sects. 2 and 3, respectively, and screening-based approaches are examined in Sect. 4.

2 Ligand-based Design

Many naturally occurring molecules have potent bioactivity that renders them useful as leads in the drug design process. These may be naturally occurring hormones, neurotransmitters, or other endogenous

Tab. 3 Information on ligands, macromolecules and their complexes sought in structure-based design and relevant NMR technologies used to derive this information.

Target	*Information*	*NMR Technology*
Ligand	Solution conformation	1D/2D NMR
	Charge/tautomeric state	Chemical shift/titrations
	Solution dynamics	Line-shape/relaxation analysis
	Pharmacophore models	All of the above, and TrNOE, of multiple ligands
	Bound ligand conformation	TrNOE
Macromolecule	3D structure	2D/3D/4D NMR
	Macromolecular dynamics	Relaxation time measurements
	Structure of articulated macromolecules (e.g. multimeric or membrane-bound receptors)	TROSY
Ligand–macromolecular complex	Stoichiometry of complex	Chemical-shift titration
	Kinetics of binding	Line width, titration analysis
	Location of interacting sites	HSQC, isotope editing
	Orientation of bound ligand	NOE docking
	Bound ligand conformation	TrNOE
	Structure of complex	3D/4D NMR
	Dynamics of complex	Relaxation time measurements

molecules, or they may be bioactive molecules from plants or microorganisms. Furthermore, screening programs on synthetic compound libraries frequently result in the discovery of bioactive molecules, which then become starting points in drug design. The general aim of ligand- or analog-based design is to determine the structure and conformation of a known bioactive molecule and then mimic this conformation in a designed lead compound, with the aim of improving its activity or druglike properties. The following sections examine various aspects of ligand-based design and illustrate them with examples.

2.1 Structure Elucidation

If the bioactive molecule is a synthetic product, its structure may be rapidly deduced by a simple comparison of NMR parameters (often combined with MS) of the product relative to those of the known precursor, to see whether the desired chemical transformation has taken place. If the bioactive compound is an unknown molecule discovered in an active fraction in bioassay-guided screening, then the first step is to elucidate its structure. Typical molecules that form the basis of such natural products-based drug discovery studies include "organic" natural products as well as small peptides and proteins. The approaches to structure elucidation for natural products and peptides/proteins are a little different from each other and are described in turn.

2.1.1 Structure Elucidation of Natural Products

In the case of nonpeptidic natural products, the main structural focus initially is

to elucidate the carbon framework. This normally involves a combination of 1D ^{1}H and ^{13}C-NMR, followed by homonuclear (DQF-COSY, TOCSY, ROESY, or NOESY) and heteronuclear (HSQC, HMBC) 2D experiments. Heteronuclear multiple bond correlation (HMBC) spectra are particularly valuable because they assist in tracing the backbone of the molecule. Such spectra display cross peaks between a ^{13}C nucleus and protons connected within two or three bonds and, in doing so, provide valuable information on molecular connectivity. Figure 6 shows typical HMBC correlations seen for selected regions of taxol, a plant-derived natural product that is currently a leading treatment for breast and ovarian cancers. Although the structure of taxol itself was originally deduced from a combination of X-ray crystallography on a degradation product and a range of ^{1}H and ^{13}C spectra in the 1970s, before HMBC spectra had been invented, HMBC spectra have been widely used for studies of the many taxol derivatives that have been examined in the last decade.

Elucidation of the carbon framework of natural products often yields substantial information about the three-dimensional structure at the same time, but if there are remaining questions on the stereochemistry of chiral centers or other factors affecting the three-dimensional structure, these can usually be resolved from NOESY spectra and/or an analysis of coupling constants. We will return to the taxol example later in Sect. 2.2 when describing conformational analysis.

2.1.2 Structure Determination of Bioactive Peptides

In contrast to the process described for organic molecules, the structure elucidation of peptide-based natural products involves two distinct steps: (1) the elucidation of the primary structure (amino acid sequence) followed by (2) the determination of secondary/tertiary structure. The primary structure determination is

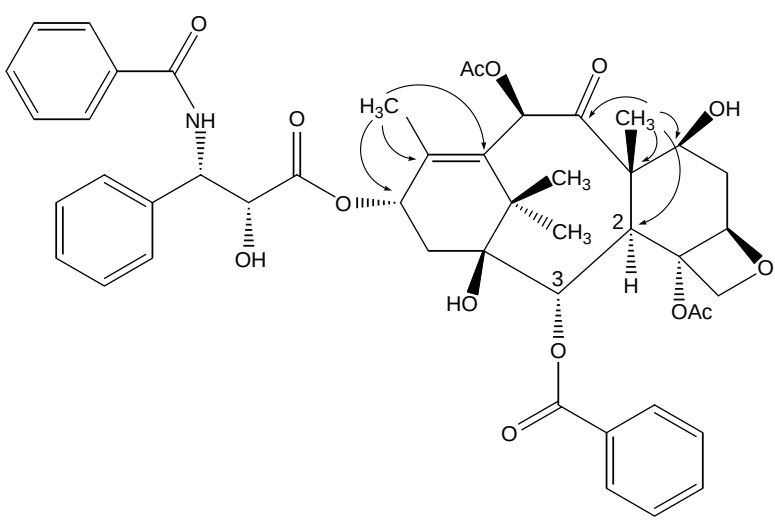

Fig. 6 Illustration of the HMBC correlations (arrows) used to assign the positions of two of the methyl quaternary methyl groups in taxol.

routinely done through Edman sequencing, or more recently by MS−MS methods. NMR plays a key role in the elucidation of the secondary and tertiary structure of peptides, mainly based on 2D homonuclear NMR spectroscopy. A combination of DQF-COSY and TOCSY (TOtal Correlation SpectroscopY) spectra are used to assign spin systems to amino acid types and then NOESY spectra are used to sequentially assign the resonances to individual protons in the peptide. The three-dimensional structure is then determined by deriving a series of internuclear distance restraints from the NOESY spectrum and using them in a simulated annealing algorithm to calculate a family of structures consistent with them.

Because the structure determination of peptides and proteins represents a very important contribution of NMR to the drug development process, it is informative to describe the process in more detail. To do this, we will use the recently developed peptide-based drug MVIIA as an example.

2.1.2.1 **NMR structure of Prialt® (MVIIA): a novel treatment for pain.** MVIIA, now known as Prialt®, is a 25-amino acid peptide originally discovered from the venom of the marine cone snail, *Conus magus*. Like other ω-conotoxins, it is a potent blocker of N-type calcium channels, giving it a wide range of potential therapeutic applications. When delivered intrathecally (i.e. through spinal infusion), it is approximately 1000 times more potent than morphine as an analgesic and has great potential for the treatment of intractable cancer pain. Figure 7 shows the peptide sequence and illustrates selected regions of the TOCSY and NOESY spectra.

As seen in Fig. 7, the TOCSY experiment is useful for classifying spin systems to amino acid type, with typically the most useful region being the "skewers" emanating from individual NH shifts (~7–10 ppm). For each NH proton in the peptide, a series of cross peaks to the α, β, and other side-chain protons is observed and these patterns define the spin system as belonging to a particular type of amino acid. Note, however, that there is some degeneracy in the resultant patterns. The NH side-chain pattern is truncated if there is a break of more than three bonds between protons within the spin system. This means, for example, that the skewers for aromatic residues extend only as far as the β-protons and they therefore appear similar to other "AMX" residues such as Cys, Ser, Asp, or Asn. Nevertheless, the ability to assign signals to either individual amino acid types or to the AMX group is a useful starting point in the assignment. However, such spectra provide no information about the sequential location of an amino acid if it is not unique in the sequence. These sequential assignments are obtained from the NOESY spectrum, as illustrated in the sequential walk shown in the middle panel of Fig. 7. The aim of the sequential assignment process is to locate adjacent amino acid spin systems, principally through a cross peak between the αH proton of one residue (i) and the NH of the following residue ($i+1$), often denoted as dαN(i, $i+1$). Additional support for the assignment is usually also sought in dβN(i, $i+1$) and dNN(i, $i+1$) correlations. At the early stages of an assignment, it is impossible to be certain whether a particular cross peak is a sequential or longer range cross peak; however, as the assignment procedure progresses, ambiguities become resolved. The assignment process is generally highly convergent, in that once a series of correct assignments is

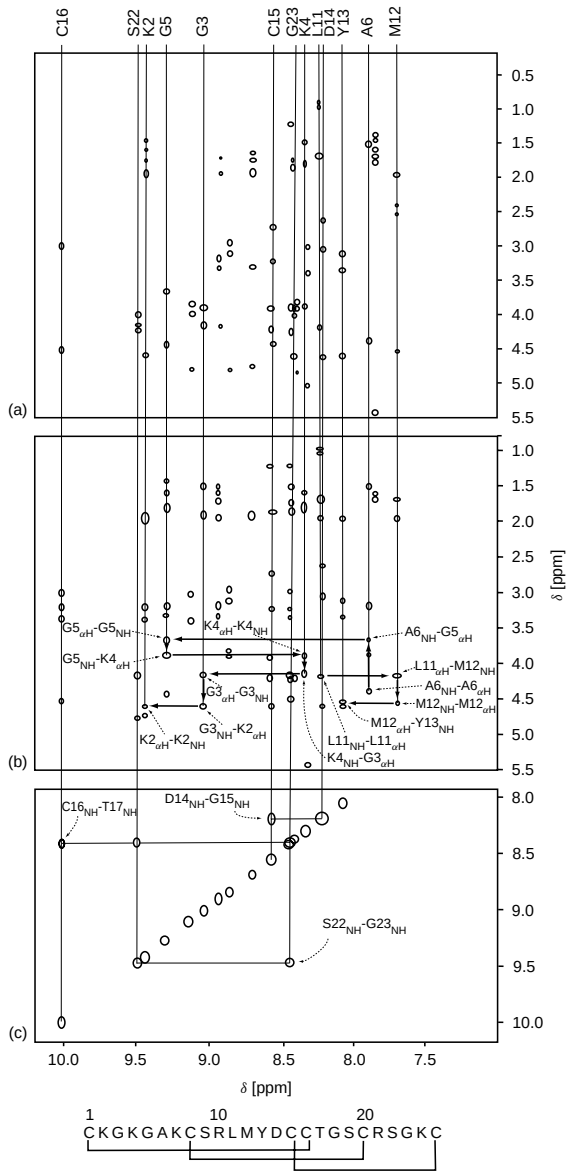

Fig. 7 Schematic representations of 2D-NMR spectra of the conotoxin MVIIA. (a) The fingerprint region of the TOCSY spectrum with selected spin systems marked. (b) Fingerprint region of the NOESY spectrum showing two (K2-A6 and L11-Y13) sequential walks. (c) NH—NH region of the NOESY spectrum showing correlations between the NH protons of D14 and G15; C16 and T17; and S22 and G23.

made, the number of choices for remaining cross peaks diminishes, in principle making their assignment easier.

Because peptides are polymers of amino acid units, the repeated NH, Hα, and side-chain protons tend to fall in characteristic chemical-shift ranges that can be useful in looking for patterns to identify amino acid types. Table 4 shows the typical chemical shifts for each of the 20 common amino acids when located in a "random-coil" environment. It is important to stress that

Tab. 4 ^{1}H chemical shifts for the 20 common amino acid residues in random-coil peptides.

Residue	*NH*	*αH*	*βH*	*Others*
Ala	8.24	4.32	1.39	
Arg	8.23	4.34	1.89, 1.79	γCH_2 1.70, 1.70
				δCH_2 3.32, 3.32
				NH 7.17, 6.62
Asn	8.40	4.74	2.83, 2.75	γNH_2 7.59, 6.91
Asp	8.34	4.64	2.84, 2.75	
Cys	8.43	4.71	3.28, 2.96	
Gln	8.32	4.34	2.13, 2.01	γCH_2 2.38, 2.38
				δNH_2 6.87, 7.59
Glu	8.42	4.35	2.09, 1.97	γCH_2 2.31, 2.28
Gly	8.33	3.96		
His	8.42	4.73	3.26, 3.20	2H 8.12
				4H 7.14
Ile	8.00	4.17	1.90	γCH_2 1.48, 1.19
				γCH_3 0.95
				δCH_3 0.89
Leu	8.16	4.34	1.65, 1.65	γH 1.64
				δCH_3 0.94, 0.90
Lys	8.29	4.32	1.85, 1.76	γCH_2 1.45, 1.45
				$\delta CH_2$1.70, 1.70
				εCH_2 3.02, 3.02
				εNH_3^+ 7.52
Met	8.28	4.48	2.15, 2.01	γCH_2 2.64, 2.64
				εCH_3 2.13
Phe	8.30	4.62	3.22, 2.99	2,6H 7.30
				3,5H 7.39
				4H 7.34
Pro		4.42	2.28, 2.02	γCH_2 2.03, 2.03
				δCH_2 3.68, 3.65
Ser	8.31	4.47	3.88, 3.88	
Thr	8.15	4.35	4.22	γCH3 1.23
Trp	8.25	4.66	3.32, 3.19	2H 7.24
				4H 7.65
				5H 7.17
				6H 7.24
				7H 7.50
				NH 10.22
Tyr	8.12	4.55	3.13, 2.92	2,6H 7.15
				3,5H 6.86
Val	8.03	4.12	2.13	γCH3 0.97, 0.94

The backbone shifts (αH and NH, ppm) are from Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., Sykes, B.D. (1995) ^{1}H, ^{13}C and ^{15}N random-coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, *J Biomol NMR* **5**, 67–81. The remaining shifts are from Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

these shifts can vary quite considerably in structured proteins (by up to several ppm) and are more useful for pattern recognition purposes than for exact identification of a particular residue. In the case of the Hα protons, the differences between the actual shifts in a structured protein and these random-coil values have an additional important use, in that they provide an indication of the local secondary structure. Intuitively, the further a chemical shift is from a random-coil value, the more likely it is attributed to the corresponding atom being in a structured environment.

After the assignment is complete, it is possible to derive substantial information about the secondary structure from an analysis of chemical shifts, coupling constants, and NOEs, even before the three-dimensional structure calculations are commenced. Figure 8 shows a typical summary of the relevant NMR information, again using the data for MVIIA as an example. Trends in these data provide a general indication of major elements of secondary structure. For example, a series of strong dαN(i, $i+1$), relative to dNN(i, $i+1$) NOEs often indicates an extended or β-type structure, whereas strong dNN(i, $i+1$) NOEs indicate local helical structure or turns. Large JαN coupling constants (>8.5 Hz) are associated with extended structure and small ones (<5 Hz) are associated with helical structure. Similarly, deviations of chemical shifts from random-coil values, often represented in terms of "chemical shift indices," indicate extended (positive values) or helical structure (negative values).

An additional useful parameter is the exchange rate of amide protons after dissolution of the sample in D_2O. Slowly exchanging amide protons indicate protection from solvent and possible involvement in intramolecular hydrogen bonds associated with elements of secondary structure. All of the NMR and slow exchange data can be consolidated to give an accurate representation of secondary structure, as indicated in the lower panel of Fig. 8. In the case of MVIIA, a triple-stranded β-sheet may be deduced on the basis of the local NOE, coupling, chemical shift, and amide-exchange NMR data.

Once all peaks in the 2D spectra have been assigned, cross peaks in the NOESY spectrum are used to derive a series of interproton distance restraints. These are then used in a simulated annealing algorithm to calculate a family of 3D structures consistent with the input restraints. Figure 9 shows two commonly used methods of representing such NMR-derived structures, either as a stereoview of the superimposed family of structures or as a ribbon diagram, in which elements of secondary structure are highlighted. For the latter representation, the lowest energy or average member of the ensemble is often chosen as representative of the structure. It is important, however, to examine the full ensemble to gain a complete understanding of the structure. Regions of disorder in the ensemble can be indicative of a lack of sufficient distance restraints, perhaps attributable to overlap or assignment errors, or may be related to local flexibility.

In the case of MVIIA, the peptide itself is being clinically developed as the active drug for administration through the intrathecal (spinal infusion) route. However, in general, peptides have a range of potential disadvantages as drugs, including poor bioavailability and susceptibility to proteolytic breakdown. Thus, for many cases involving peptide-based leads, the structural information of the type described above might be used as a starting point

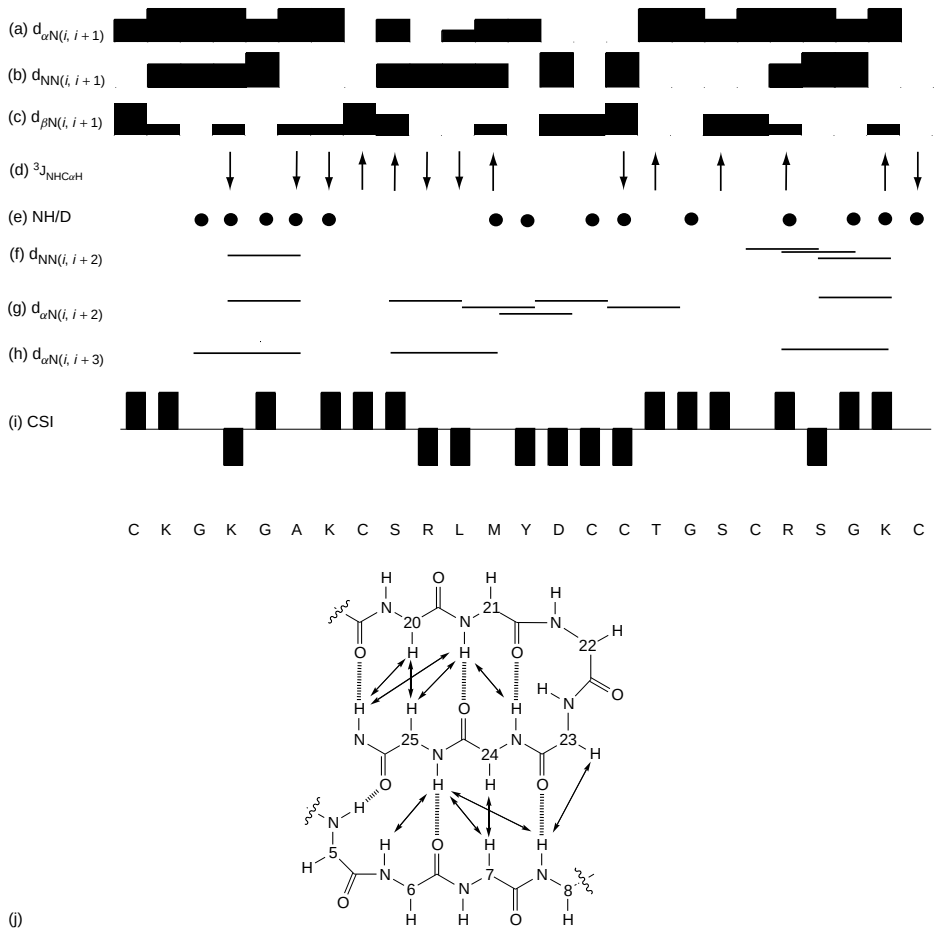

Fig. 8 A summary of the NMR data observed for MVIIA. (a) Hα−NH sequential NOEs. (b) NH−NH sequential NOEs. (c) Hβ−NH sequential NOEs. (f−h) Other short-range NOEs. The thickness of the bar indicates the strength of the observed NOE (weak, medium, or strong). (d) Three-bond NH−Hα coupling data, where upward-pointing arrows indicate a large coupling (>8 Hz) and downward-pointing arrows indicate a small coupling (<5 Hz). (e) H/D exchange data, where a filled circle represents a slow exchanging NH. (i) Chemical-shift index (CSI) data. The CSI uses a scoring system that compares Hα shifts to random-coil chemical shifts. A sequence of consecutive +1 scores is indicative of β-structure, whereas a sequence of consecutive −1 scores suggests helical structure. (j) The β-sheet of MVIIA. Double-headed arrows indicate observed NOEs and broken lines indicate proposed H bonds.

to design smaller constrained peptides or nonpeptidic mimics. This is the case, for example, in the development of endothelin antagonists described below.

2.1.2.2 **Endothelin as a lead in ligand-based design.** Endothelin (ET), shown in Fig. 10, is a 21-amino acid endothelial-derived constricting factor that has gained

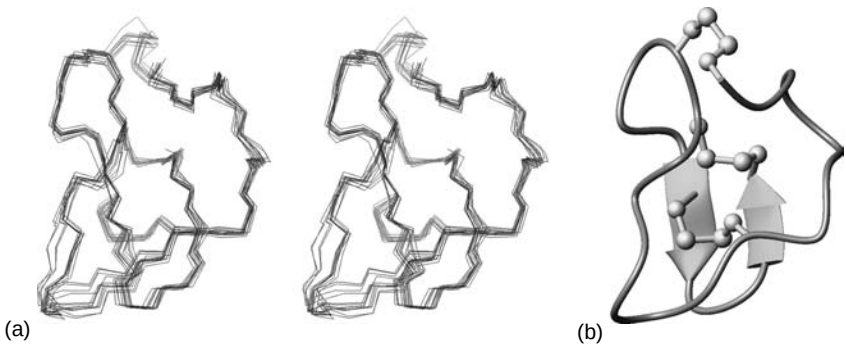

Fig. 9 (a) A stereoview of the superimposed backbone structures of the 20 lowest energy conformations for MVIIA (reproduced from Nielsen, K.J., Thomas, L., Lewis, R.J., Alewood, P.F., Craik, D.J. (1996) *J. Mol. Biol.* **263**, 297). (b) Ribbon diagram of MVIIA.

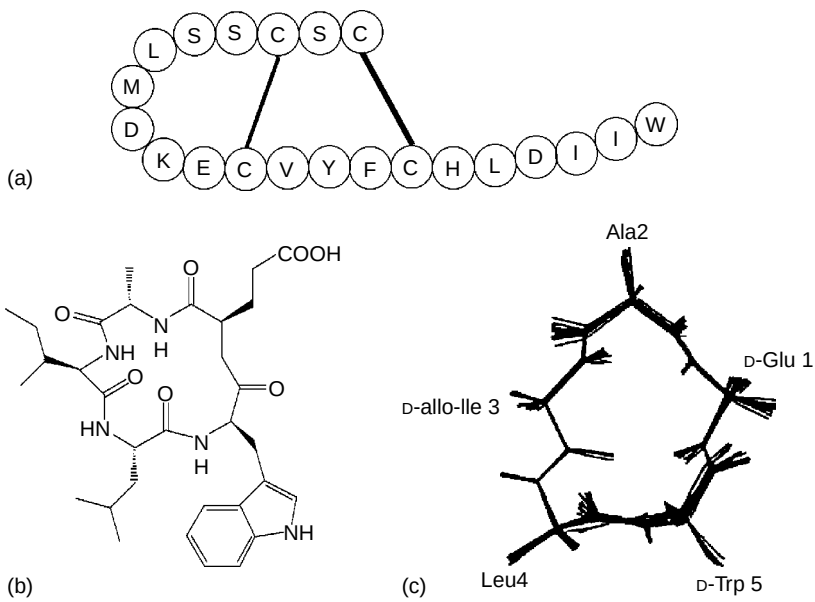

Fig. 10 (a) Primary sequence and disulfide connectivities of endothelin-1 (ET-1). (b) Primary structure of the cyclic endothelin antagonist BE18257B and (c) a family of 36 NMR structures, which demonstrate the well-defined nature of the cyclic peptide backbone.

prominence as a pharmacological lead molecule. Interest in the peptide arose because of its potent renal, pulmonary, and neuroendocrine activities. Endothelin and its isoforms have been implicated in a wide variety of disease states including ischemia, cerebral vasospasm, stroke, renal failure, hypertension, and heart failure. It exerts its pharmacological effect by acting on specific G-protein-coupled receptors. In mammalian species, two receptors, ET_A and ET_B, have been cloned; both are widely distributed in human tissue and are distinguished

by different responses to various ET isoforms.

The NMR-derived three-dimensional structure of ET-1 consists of several distinct regions, including a random-coil N terminus, a β-turn involving residues 5–8, followed by a short helical region and a flexible C-terminal tail. The presence of the flexible tail in solution is not surprising, as may be imagined from the primary sequence shown in Fig. 10. Although solution structures of ET and its analogs determined by NMR have been valuable in defining the gross conformation of these molecules, the flexibility of the tail in solution makes it difficult to extrapolate to the bound state. Indeed, an X-ray structure of ET-1 has quite a different structure for the C-terminal tail than for the random-coil arrangement in solution. The bound conformation may be different again.

There is clearly an advantage in having lead molecules with reduced flexibility, given that their solution conformation will intrinsically provide a better reflection of the bound conformation. In addition, the development of a more rigid drug will reduce unfavorable entropic contributions to binding energy. Indeed, a range of small cyclic peptides that are ET_A- or ET_B-selective antagonists have been discovered and provide valuable leads to the development of potential therapeutics. NMR studies have been instrumental in determining their solution conformations. For example, the rather well-defined solution conformation of the ET_A-selective antagonist BE18257B (shown in Fig. 10) contrasts with the flexibility of the tail region of ET that this peptide is thought to mimic. The discovery and development of these molecules illustrate the principle that cyclic peptides are often more suitable than linear peptides as lead ligands in drug design. In addition to their better defined and less-flexible conformations than those of their linear counterparts, they generally have improved bioavailability and resistance to protease attack.

We shall return to endothelin as a lead in drug design, in relation to a nonpeptidic antagonist. The underlying theme illustrated by the endothelin example is that ligand-based design often proceeds from initial studies of flexible endogenous molecules (particularly peptides) to constrained mimics (e.g. cyclic peptides) and often culminates in the development of nonpeptidic drug leads. NMR assists by defining the structures of the lead and subsequent molecules.

2.1.3 Instrumental Advances and their Impact on Structure Elucidation

Over the last few years, there have been several exciting instrumental developments that promise to dramatically expand the role NMR will play in the drug discovery process. These relate to the combination of NMR with other technologies such as LC and/or MS and the use of NMR to directly monitor reactions carried out on solid-phase resins. The latter promises to indirectly enhance drug discovery programs by improving the monitoring and hence efficiency of solid-phase combinatorial synthesis. Effectively, resin-based syntheses can be monitored at successive stages without the need to cleave intermediate products from the resin.

As already mentioned, the additional sensitivity brought about by cryoprobe technology promises to enhance a wide range of NMR applications, but will be particularly important in natural products-based drug discovery. In many cases, only limited amounts of pure compounds are isolated from natural product extracts and sensitivity has been a major limiting factor on structure elucidation. LC/MS/NMR

systems will greatly improve the efficiency of such analyses by minimizing the need for separate sample-handling steps for the different analytical technologies.

2.2 Conformational Analysis

Usually only 1D or 2D NMR methods are required to determine the solution conformation of bioactive ligands. Useful tools include analysis of chemical shifts, coupling constants, and NOEs. An assumption inherent in the application of such studies to drug design is that the solution conformation will be maintained on binding to the receptor. This is justified in the case of relatively rigid ligands. However, for potentially flexible ligands, the possibility of changes in conformation on binding must be considered, as noted above for the case of endothelin.

Coupling constants and NOEs are the main NMR parameters used in determining the solution conformations of drug leads. NOEs provide information about through-space proximity. Three-bond vicinal-coupling constants are particularly valuable because their dependency on the intervening dihedral angle through the Karplus relationship allows local geometry to be determined. This is illustrated in Fig. 11 for taxol. Although there are several vicinal-coupling constants in this molecule (Fig. 6), only one $^3J_{H2H3}$ occurs in a region of the molecule that is expected to be conformationally rigid and thus suitable for conformational determination by use of coupling constants. In taxol and a range of analogs, this coupling is in the range 4–7 Hz, consistent with partially eclipsed dihedral angles of approximately 120–140° for this ring-constrained structure. This is in good agreement with the X-ray structure of a taxol analog, where the angle is 120°. Note that, in general, such a Karplus analysis does not give a unique solution unless several coupling constants sampling the same dihedral angle are present and is reliant on the

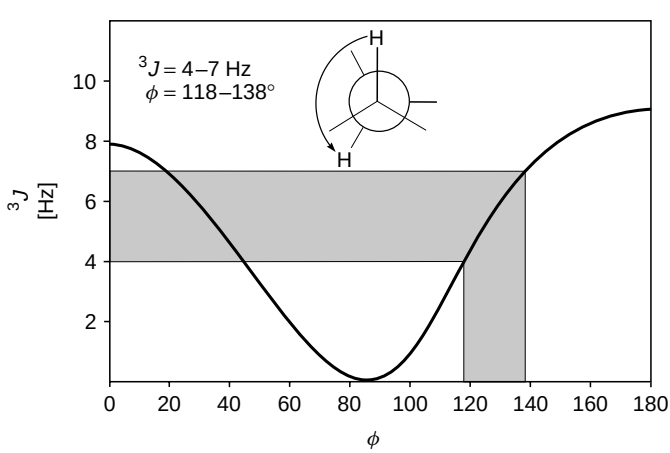

Fig. 11 Illustration of the Karplus relationship between three-bond scalar coupling constants and the dihedral angle of the intervening bond. The relationship is indicated for the phi (Φ) torsion angle of the H2 and H3 protons within the rigid core of taxol and related derivatives. See Fig. 6 for the structure of taxol.

assumption that the molecule exists only in a single conformation in solution. Although it is generally believed that this is the case for the core of taxol, recent relaxation data described in Sect. 2.5 suggest that this conclusion may need to be reexamined.

In addition to studies of the taxol core, there have been a large number of studies of the conformations of the side chains of taxol, and it appears that these are certainly flexible and that the molecule may adopt both extended and folded conformations of the side chains. In a case like this, the observed vicinal-coupling constants are a weighted average of those from the participating conformers.

2.3 Charge State

An advantage of NMR over other structural techniques such as X-ray crystallography is that it has the potential to provide information not only on structure but also on the electronic properties of molecules. Many drug leads contain ionizable groups and a determination of their charge state in solution and/or at the bound site is important in the design of analogs. Simple plots of chemical shifts as a function of pH for nuclei near these ionizable groups provide a convenient way of determining the pK_a value and hence the charge state. This is illustrated for Prialt® in Fig. 12, where it was suspected that one of the ionizable groups in the molecule, Asp^{14}, may be involved in a stabilizing salt–bridge interaction. This was confirmed by noting that the pK_a value for this residue is lowered considerably relative to the usual value for Asp. The β-proton chemical shifts were essentially independent of pH over the range 3–7 (indicating a p$K_a < 3$), whereas those of a control, random-coil peptide, titrated as expected over this range.

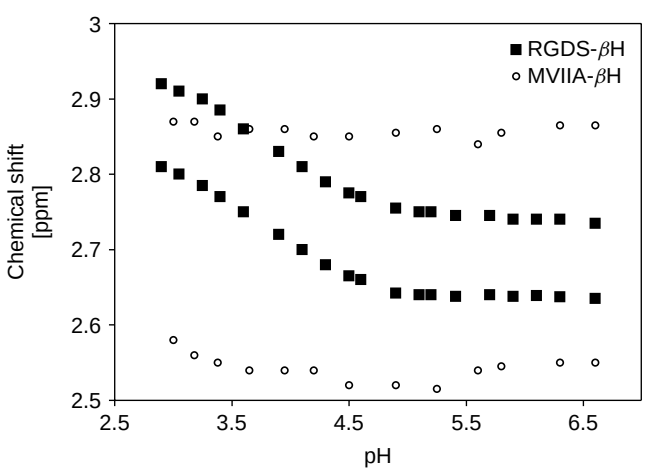

Fig. 12 Chemical-shift changes of the β-protons of Asp14 in MVIIA illustrating the lack of titration of the adjacent carboxyl group, indicating its involvement in salt bridge. By contrast, the shift of a control random-coil peptide varies with an apparent pK_a value of 3.7, as expected for an uncomplexed carboxyl moiety in peptides.

2.4 Tautomeric Equilibria

Tautomerization is a relatively common feature of drug molecules that is amenable to analysis through the use of chemical shifts or coupling constants as probes. This was recently demonstrated in a study of some nonpeptide endothelin analogs. Starting from the modestly active compound (**1**) (Table 5), derived by screening a compound library for ET_A antagonists, the nanomolar inhibitor (**2**) was developed. Further optimization through examination of electronic and structural requirements led to the subnanomolar inhibitor (**3**), which was subsequently put forward for evaluation in a number of preclinical disease models for stroke.

These molecules display keto-enol tautomerization, as illustrated in the keto-enol structures. The open form keto-acid salts and the closed-form butenolides exist in a pH-dependent equilibrium in solution, and at physiological pH, both forms exist. In principle, the biological activity could reside in either or both forms.

The extent of tautomerization was established by evaluation of NMR spectra as a function of pH, from 2.65–9.05. At acidic pH, compound (**2**) exists essentially in the closed butenolide form. Because the pH is slowly raised by addition of NaOD, the spectrum begins to exhibit properties associated with the open form keto acid,

Tab. 5 Substitution pattern and receptor-binding affinity of nonpeptidic endothelin antagonists.

Compound	R_1	R_2	IC_{50} *[nM]*	
			ET_A	ET_B
(**1**) PD012527	Cl	H	430	27000
(**2**) PD 155080	OCH_3	H	>0.4	4550
(**3**) PD156707	OCH_3	3,4,5-OCH_3	0.3	780
(**4**)	OCH_3	3,5-OCH_3,4-$O(CH_2)_3SO_3Na$	0.38	1600

From Patt, W.C., Edmunds, J.J., Repine, J.T., Berryman, K.A., Reisdorph, B.R., Lee, C., Plummer, M.S., Shahripour, A., Haleen, S.J., Keiser, J.A., Flynn, M.A., Welch, K.M., Reynolds, E.E., Rubin, R., Tobias, B., Hallak, H., Doherty, A.M. (1997) Structure-activity relationships in a series of orally active gamma-hydroxy butenolide endothelin antagonists, *J. Med. Chem.* **40**, 1063–1074 and Patt, W.C., Cheng, X.M., Repine, J.T., Lee, C., Reisdorph, B.R., Massa, M.A., Doherty, A.M., Welch, K.M., Bryant, J.W., Flynn, M.A., Walker, D.M., Schroeder, R.L., Haleen, S.J., Keiser, J.A. (1999) Butenolide endothelin antagonists with improved aqueous solubility, *J. Med. Chem.* **42**, 2162–2168.

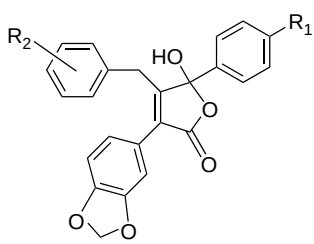

(**1–4**)

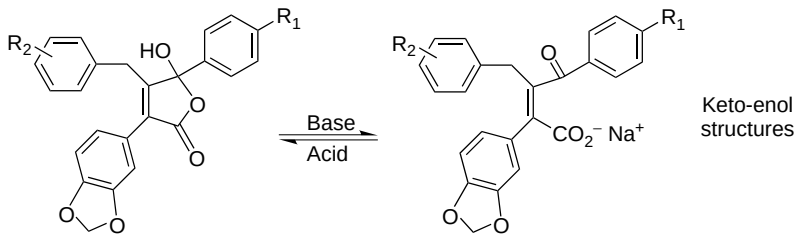

and at basic pH the compound is essentially all in the open form. The coupling pattern shown by the benzylic protons is a particularly characteristic marker of the tautomeric process. At acidic pH, the benzylic protons exhibit an AB quartet pattern consistent with the ring-closed structure. As the pH is raised, this pattern coalesces to a singlet, broad at neutral pH and sharp at basic pH, as would be expected with the open form keto-acid structure. After the pH was basic, addition of DCl to acidify the solution caused the spectrum to return to its original appearance, consistent with a reversible tautomerization process.

Identical biological results were obtained with the salt and closed butenolide form in all pharmacological assays, reflecting equilibration at physiological pH. This made it difficult to identify the biologically active form from these experiments alone, although methylation of the OH group in compounds (**1**–**3**) resulted in a loss of activity. Because these analogs cannot tautomerize to form open keto acids, it seems likely that the open form is responsible for activity.

In addition to its impact on the biologically active form, the tautomeric process has profound implications for formulation of drug candidates, as illustrated in some recent development work on compound (**3**). Although it is easy to synthesize and isolate water-soluble salts of the keto acids, once they are placed in aqueous solution the tautomeric equilibrium determines how much of each form is present. Indeed, if the closed butenolide tautomer is sufficiently water insoluble, it can precipitate out of solution and the equilibrium can drive the complete precipitation of the compound. Although (**3**) has good oral activity, its intravenous use is limited by the insolubility of the closed-form butenolide tautomer without the use of a specific and complex buffered formulation. Thus, in recent work, a series of water-soluble butenolides was developed to overcome this limitation for parenteral uses. This culminated in the development of (**4**) (Table 5), currently in preclinical evaluation.

This description of the development of (**4**) provides a good illustration of the fact that the synthesis of an active molecule is not the end of the drug development pathway, and that formulation considerations can be critical. In this case, NMR played a significant role in understanding tautomeric processes that had a direct bearing on solubility and hence formulation.

2.5 Ligand Dynamics: Line-shape and Relaxation Data

It is increasingly being recognized that the solution molecular dynamics of drugs may have an important role in modulating biological activity. For example, dynamics may influence entropic contributions to the free energy of binding. In general,

the more flexible a ligand is, the more unfavorable will be the loss in entropy on binding, assuming a relatively rigid bound state of the ligand. However, in some cases flexibility of a ligand may be a positive factor. This applies, for example, if a degree of flexibility is required to allow a ligand access to a buried active site, or if activation of a receptor requires a conformational change mediated by ligand binding. Therefore, knowledge of the flexibility of lead molecules is an important supplement to the structural and electronic information available from NMR.

The two major NMR methods for obtaining information on ligand flexibility are line-shape analysis and relaxation measurements (usually ^{13}C or ^{15}N T_1, T_2, or heteronuclear NOE measurements). In general terms, the former is sensitive to motions in the milli- to microsecond timescale and the latter in the nanosecond timescales. To some extent, structure calculations on peptide-based lead molecules can also give an indication of regions of flexibility from an examination of local regions of disorder among a family of calculated structures. Caution must be exercised because other factors can contribute to disorder, although in many cases there is a connection between disorder in a structural ensemble and molecular flexibility.

In an example that illustrates the application of NMR relaxation measurements for studying ligand flexibility, Kessler and colleagues investigated the role of disulfide bonds in the α-amylase inhibitor, tendamistat. This small protein contains two disulfide bonds (C11–C27 and C45–C73) and opening of the latter is known to reduce the melting temperature of the protein (i.e. reduce its stability), but not affect its α-amylase inhibitor function. The latter observation suggests that this disulfide bond may not affect either the structure or the dynamics of the molecule. To investigate this possibility, heteronuclear NOE measurements were used to determine the effect of the selective removal of the single disulfide bond on the dynamics of the molecule. To assess structural changes, chemical-shift differences, intrastrand NOE effects, and protected amide protons were measured for the disulfide-deficient variant C45A/C73A and wild-type tendamistat. Removal of the C45–C73 bond by the C45A/C73A mutation was found to have no influence on the β-barrel structure, apart from very local changes at the mutation sites. ^{13}C and ^{15}N relaxation data showed that the only region of significant internal mobility in either the wild-type or variant protein was at the N terminus. There was little difference in internal flexibility between the two proteins, consistent with their similar α-amylase inhibitory activity. In this case, it appears that the role of the C45/C73 disulfide bond is more related to thermodynamic stability and the secretion efficiency of the protein rather than to limiting its flexibility.

The thyroid hormones, exemplified by thyroxine (**5**), provide an example of the use of both line-shape analysis and NMR relaxation time measurements to give an insight into the internal flexibility, and perhaps the mode of action, of pharmaceutically important molecules. The thyroid hormones act by binding to a nuclear receptor and appear to control receptor function by inducing a conformational change that directs the alignment of functionally critical secondary-structure elements of the receptor. Synthetic thyroxine is widely used for the treatment of thyroid disorders and indeed has been the second most widely prescribed drug in the United States over recent years.

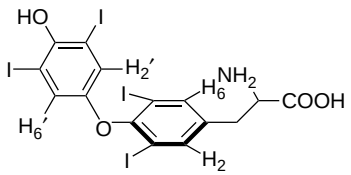

(5)

Table 6 shows experimental ^{13}C T_1 and heteronuclear NOE data for thyroxine together with best-fit theoretical values for these parameters based on two different models of the motion of thyroxine in solution. A simple isotropic model, in which the drug is regarded as a rigid body tumbling in solution, is clearly unable to simultaneously fit the T_1 and NOE data, whereas the two-state jump model, which incorporates a degree of internal mobility based on rapid flipping between conformational states, is better able to account for the experimental data. Although it is difficult to precisely define the nature of the internal motion, it is clear that rapid (nanosecond) internal motion is present, and this appears to involve small-amplitude torsions about the aromatic ring axes. The correlation time for overall tumbling of thyroxine was deduced to be approximately 0.35 ns, with the internal motion approximately 30-fold faster.

Studies of the ^{1}H-NMR spectrum of thyroxine as a function of temperature had earlier demonstrated the presence of additional larger amplitude, but slower ring flips. At low temperature, two signals were seen for the H2′ and H6′ protons. These signals broadened with increasing temperature, then coalesced and sharpened as the temperature was further increased. This was attributed to exchange of the environments of the two protons brought about by 180° rotation of the "outer" ring of thyroxine. Substitution of the observed coalescence temperature (T_c) and the chemical-shift difference of the two signals at low temperature ($\delta\nu$) allowed the free energy of activation for this slow ring flip process to be established from Eq. 1

$$\Delta G^{\neq} = 19.14\ T_c \left[9.97 + \log\left(\frac{T_c}{\delta\nu}\right) \right] \quad (1)$$

The derived barriers for several thyroid hormones are in the range 36–38 kJ mol^{-1}, which corresponds to large-amplitude ring flips on the milli- to microsecond timescale. From a combination of the relaxation data and the

Tab. 6 ^{13}C NMR Chemical shift and relaxation data for thyroxine.

Position	*Chemical shift [ppm]*	*Experimental*[a]		*Theoretical*[b]			
				Isotropic motion		*Two-state internal motion*	
		T_1[s]	*NOE*	*T_1[s]*	*NOE*	*T_1[s]*	*NOE*
C2′,6′	127.3	0.63	2.53	0.63	2.96	0.63	2.53
C2,6	142.6	0.63	2.63	0.63	2.96	0.63	2.58
C-α	57.1	0.51	2.37	0.51	2.94	0.51	2.57
C-β	36.4	0.64	2.29	0.64	2.96	0.64	2.51

[a]Measured relaxation data at 75 MHz and 305 K.
[b]Theoretical best-fit values based on the indicated models for molecular motion.

dynamic line-shape analysis data, it was possible to propose a unified model that accounts for both the fast and slow internal motions, as summarized in Fig. 13.

In this model, both aromatic rings of the thyroid hormones jump rapidly between two energetically equivalent conformations on a nanosecond timescale (a ↔ b and c ↔ d in Fig. 13). The half-angle of the jump varies, depending on the solvent, corresponding to an average displacement of about 90° between the two extreme jump positions. These separate states are not detectable on the chemical-shift timescale but lead to an average proximal environment for Ha and an average distal environment for Hb (attributed to rapid interchange between a and b in Fig. 13), which are seen in the low temperature spectra. However, these fast motions are detected by relaxation studies. Although the rate of this motion is rapid, its amplitude is not sufficient to average the environment of proximal and distal protons. Occasionally (about once every 1000 jumps), the outer ring jumps further

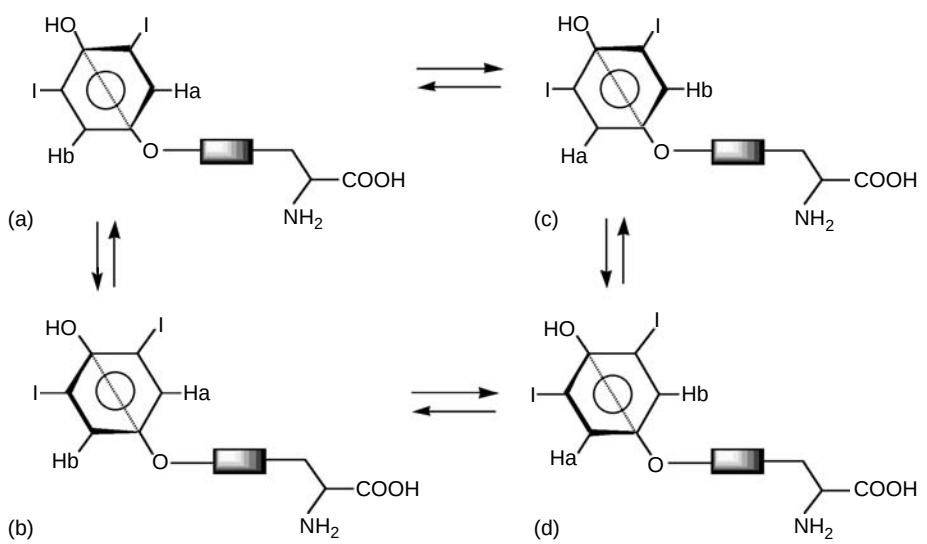

Fig. 13 Schematic illustrations of motions of the outer ring of thyroxine. The dotted line through the outer ring shows the jump axis about which the ring rotates. (a) Ha is shown in the proximal position and is closer to the viewer than Hb because the torsion angle phi $(\Phi)'$ is greater than 0°. This conformation corresponds to one of the two states of the two-state jump model and agrees with the "twist" of the outer ring observed in the crystal structure.
(b) Rotation about the dotted line through the center of the outer ring moves Ha away from the viewer and brings Hb toward the viewer. This corresponds to the second state of the outer ring in the two-state jump model. (c) Hb is now in the proximal position and closer to the viewer than Ha. (d) Hb is in the proximal position and is now further from the viewer than Ha. Transition from a to b and from c to d involves small-amplitude jumps on the nanosecond timescale and is detected by NMR relaxation measurements. Although not illustrated in the figure, the inner ring also exhibits this type of motion. Transitions a to c and b to d result in 180° flips of the outer ring and exchange of the environments of Ha and Hb. This ring flip occurs on a microsecond timescale and is detected by variable-temperature line-shape studies.
(Reprinted with permission from the American Chemical Society from Duggan, B.M., Craik, D.J. (1996) *J. Med. Chem.* **39**, 4007.)

than the nominal 90° range, exchanging the environments of the proximal and distal protons (a ↔ c and b ↔ d in Fig. 13). Although the actual rate of an individual ring flip is rapid, the effective rate of the process is on the microsecond timescale, because on average a large number of small-amplitude jumps occur for every large-amplitude ring flip. It is the exchanging of proximal and distal protons on the microsecond timescale that is detected by the variable-temperature line-shape studies.

The fact that thyroxine is apparently able to so freely move over a moderately large region of conformational space has implications for receptor binding. The crystal structure of the thyroid receptor ligand-binding domain complexed with the thyroid agonist 3,5-dimethyl-3′-isopropylthyronine shows that the thyroid hormones bind at the center of the hydrophobic core of the ligand-binding domain and may play a structural role in the conformational changes that activate the receptor. The structures of the retinoid-X receptor ligand-binding domain and the retinoic acid–retinoic acid receptor ligand-binding domain complex indicate that significant conformational changes accompany ligand binding in those cases. The conformational flexibility exhibited by the thyroid hormones may also be required for binding. It has been suggested that the rapid "wiggling" of the aromatic rings could enable the hormone to work its way to the center of the ligand-binding domain as the protein reorders itself about the ligand and may in fact trigger receptor conformational changes.

As briefly mentioned earlier, taxol provides another example where relaxation time measurements provide an insight into dynamics processes. Although it is generally thought that the taxane core is rigid, ^{13}C relaxation data suggest that a degree of flexibility (on the nanosecond timescale) may be present and may vary for different taxol analogs. In particular, it appears that the removal of certain side chains may introduce additional flexibility into the core region that would not easily be predicted on the basis of a simple inspection of the structure.

Another example of the application of line-shape analysis to ligand dynamics is described in Sect. 3.2 for the drug trimetrexate when bound to dihydrofolate reductase (DHFR). From that example and earlier studies on DHFR, it is clear that the techniques described earlier can equally be applied to ligands when bound to their receptor. In some cases, significant but highly specific mobility appears to be present at the bound site.

2.6 Pharmocophore Modeling: Conformations of a Set of Ligands

Determination of the conformations of a range of ligands all of which act at the same receptor site can provide significantly more information than just a single ligand structure. With a sufficiently broad range of ligands, it is often possible to generate a pharmocophore model of the receptor site, deduced on the basis of the conserved structural features and the conformations of the ligands. This has been done recently, for example, for the ω-conotoxins, the broad class of conotoxins to which Prialt® belongs. From structural studies of a range of ω-conotoxins and from literature data on various mutants with altered binding affinities, it was determined that only a localized region of the surface of these molecules is involved in receptor binding. This allowed a pharmocophore model of

putative receptor-binding pockets to be developed.

The advantage of such a pharmocophore model is that smaller, nonpeptide molecules that might have improved stability and bioavailability over their peptidic counterparts can, in principle, be designed. The NMR approach used in such pharmocophore modeling often involves a combination of many of the techniques already described. By determining information about structure and electronic properties for a range of different ligands, all acting at the same receptor site, it is often possible to infer information about the binding site, even if direct structural studies of this site are not possible.

2.7 Limitations of Analog-based Design

Although a determination of the structure of bioactive molecules is of key importance, there are distinct limitations on the use of solution structures for drug design. In particular, unless the molecule is rigid there is no certainty that the solution conformation is the same as the bioactive bound conformation. For this reason, there has been a shift over recent years to approaches in which information about the bound state is obtained. The other approach has been to probe the bound conformation by making a range of constrained analogs of a flexible lead molecule, as illustrated earlier for endothelin.

The most direct way of determining the conformation of a drug lead is to determine the full three-dimensional structure of its receptor complex. This has now been achieved in a significant number of cases but represents a substantial undertaking, as described in later sections of this chapter. A simpler approach that has also been applied is to use transferred nuclear Overhauser effect (TrNOE) methods, as described in the following sections. This approach fits at the interface of ligand-based design and receptor-based design. It fits with the former because no knowledge of the receptor structure is required, but it also fits with the latter because it requires the macromolecule of interest to be included in the mixture to be analyzed. It is appropriate therefore to introduce the topic here but also to discuss it further in Sect. 3.

2.8 Conformation of Bound Ligands: Transferred NOEs

In ligand-based drug design, it is not necessary to know the structure of the receptor, or even the location of the binding site, although the conformation of the ligand bound to the receptor is crucial. It is clearly better if this can be measured directly rather than be inferred from the conformation of the free ligand. In certain circumstances this information on the bound conformation can be obtained from the TrNOE technique. This method takes advantage of the fact that NOEs build up more rapidly in a ligand–macromolecule complex than they do in free ligand, and given appropriate exchange conditions for a mixture of ligand and macromolecule (typically satisfied for $K_D \geq 10^{-7}\ M^{-1}$), then signals from a free ligand may be used to determine the bound conformation.

Because it is not necessary to monitor signals from the macromolecule in this technique, as it is usually present in substoichiometric amounts, thus requiring only minimal amounts of what is sometimes the more expensive component of ligand–macromolecule complexes. In addition, the molecular weight restrictions inherent in full 3D-structure

determinations of complexes are ameliorated and the conformations of ligands bound to very large macromolecules may be determined. For example, the technique was recently used to determine the structure of an antibiotic bound to the ribosome and a range of other applications including enzyme–substrate, protein–carbohydrate, and protein–peptide interactions have been reported.

In addition to its application as a tool for determining bound conformations of ligands, the TrNOE method has also been used recently as a screening aid for the identification of ligands from mixtures that bind to a protein of interest. This application is addressed in more detail later in this chapter.

3 Receptor-based Design

Receptor-based design refers to the process of determining the three-dimensional structure of a macromolecular target and using this information to design ligands to interact with it. In general, there have been few cases where the structure of a macromolecule or receptor alone has been successfully used to design, *de novo*, a ligand to interact with that receptor. However, such an approach is likely to become more common with improved computer-based approaches to molecular design in the future. Currently, the most common approach is to study a ligand–macromolecule complex and to initiate the design process based on the interaction of the lead ligand with the macromolecule.

Although the structure of the macromolecule alone is of less interest than that of the complex, in many cases a determination of the structure of the complex follows from earlier studies on the unbound macromolecule. It is thus useful to describe the approaches to structure determination of macromolecular targets. This is followed by a discussion of the dynamic aspects of protein structures in Sect. 3.1, before addressing the main topic of macromolecule–ligand interactions in Sect. 3.2.

3.1 Macromolecular Structure Determination

The two major techniques for determining three-dimensional structures of proteins or nucleic acids are X-ray crystallography and NMR spectroscopy. NMR has been used to determine the structures of proteins for only 20 years, with the first NMR structure determination being made in 1985. NMR has a number of advantages over X-ray crystallography, including the fact that the requirement that the protein needs to be crystallized is avoided, and that the dynamic information available from NMR studies complements the structural information. A major disadvantage of NMR spectroscopy, though, is that it is currently limited to the determination of structures of <35 kDa. With the development of new NMR techniques, such as TROSY, this seems certain to increase significantly over the coming years, although the fact remains that among all structures currently deposited in the protein database the average size of NMR structures is about 8 kDa, substantially smaller than the average size of protein structures determined by X-ray crystallography. Despite this limitation, NMR has made major inroads into the macromolecular structure determination process, and currently approximately one-fifth of all new structures deposited in the protein database have been determined by NMR spectroscopy.

3.1.1 Overview of Approach

The basis for structure determination by NMR is that, by determining a large number of distance restraints between pairs of protons, it is possible to reconstruct a three-dimensional image of the molecule. These distance restraints are derived primarily from NOE measurements, which detect distances up to about 5 ´Å. Over recent years, such distance restraints have been supplemented by a range of other restraints, including dihedral angle restraints derived from coupling constant measurements and orientation restraints derived from residual dipolar couplings. These restraints are input into a simulated annealing algorithm, which is used to calculate a family of structures consistent with the restraints.

NMR is unique in that it can provide detailed and specific information on molecular dynamics in addition to structural information. The use of relaxation time measurements allows the relative mobility of individual atomic positions within a macromolecule to be determined. The dynamic information obtained includes not only the rates or frequencies of internal motions but also their amplitudes. Such amplitudes are often expressed by order parameters. Not surprisingly, it is observed in many cases that the termini of proteins are more flexible than internal regions. More interestingly, NMR has provided a number of examples where internal loops in proteins have been shown to have dynamics that may be associated with their function. A good example of this is HIV protease, where NMR studies have identified lowered order parameters in the flap region of the molecule, which may reflect flexibility to allow entry of substrates or inhibitors into the active site.

In summary, a major strength of NMR is that a global picture not only of the structure but also of the dynamics of the macromolecular target is obtained. Further, NMR provides information on ionization states of titratable groups and other electronic features within macromolecules, which may have an impact on ligand binding and function.

3.1.2 Sample Requirements and Assignment Protocols

Structure determination by NMR typically requires 500 μL of a 1–2 mM solution of the protein of interest. It is important that the macromolecule does not aggregate because this causes spectral broadening and may preclude assignment. The sample should preferably be stable in solution over the extended period of time required to collect the range of NMR experiments needed for assignment and structure determination. Individual experiments may last from a few hours to several days, with several weeks of data acquisition required for studies of larger proteins.

The particular set of NMR experiments required for NMR structure determination depends on the size of the protein. For smaller proteins ($\leq$7 kDa) it is usually possible to determine the structures, mainly using 2D NMR, without the need for isotopic labeling, by use of procedures described in detail earlier for Prialt®. For proteins in the range 7–14 kDa, ^{15}N-labeling and a combination of 2D/3D NMR experiments is usually sufficient, whereas for larger proteins ^{13}C/^{15}N labeling and 3D or 4D NMR is more or less mandatory. For proteins at the top end of the currently accessible range (25–35 kDa), there are additional advantages associated with partial deuteration of the protein.

3.1.3 Recent Developments

A number of recently developed methods offer the potential for improving the quality of NMR structures and for increasing the size of proteins that can be examined. In particular, the use of residual dipolar couplings and of anisotropic contributions to relaxation provide new kinds of restraints that promise to lead to more accurate NMR structures. As already mentioned, the TROSY method exploits relaxation phenomena to produce spectra with narrow lines and promises to significantly expand the size of protein targets that can be examined by NMR, from the current limit of about 35 kDa to perhaps >100 kDa.

Another development that is likely to have a significant impact is the increasing number of structural genomics programs being developed. The demands arising from such programs will no doubt stimulate new methods for the large-scale production of labeled proteins, and for speeding up the rate of structure determination by both NMR and crystallography.

3.1.4 Dynamics

Proteins exhibit a range of internal motions, from the millisecond to nanosecond timescale, and a full understanding of how small drugs might interact with such a "moving target" requires more than just the time-averaged macromolecular structure. Thus, over recent years much effort has been directed toward defining motions within proteins.

The most commonly applied approach has been to use ^{13}C or ^{15}N relaxation parameters such as T_1, T_2, and the heteronuclear NOE to derive correlation times for overall motion, together with rates and amplitudes of internal motions. Although the precise interpretation of the NMR relaxation data in terms of motional parameters remains dependent on the appropriateness of the motional model chosen, the results from many studies on the dynamics of proteins are sufficiently clear to confirm that nanosecond timescale motions in proteins are common. The functional significance of motions on the nanosecond timescale remains unclear and so far there have been few cases where significant differences in motions on this timescale between ligand-free and ligand-bound forms of proteins have been measured. It will be interesting to assess the functional significance of such motions as more data become available. However, slower motions have been correlated with function in a number of proteins, with a good example being HIV protease, described in more detail in Sect. 3.2.4.5

3.1.5 Nucleic Acid Structures

Most of the discussion on macromolecular targets so far has focused on proteins. DNA represents another valuable target in drug design. Most studies in which DNA is the target are done using short model oligonucleotides to mimic the binding region of DNA. The regular repeating nature of DNA structures makes this a more successful approach than similar attempts to dissect out binding regions of receptor proteins, where often the whole protein must be present to maintain a viable binding site. Similar comments apply for RNA, where improvements in synthetic methods have led to an increasing number of structure determinations over recent years. The principles involved in structure determination of nucleic acid targets are similar to those of proteins, but in practice nucleic acid structures are somewhat more difficult to solve.

3.1.6 Challenges for the Future: Membrane-bound Proteins

The majority of targets for currently known drugs are membrane-bound receptors, yet this represents the class of proteins for which least structural information is known. Membrane proteins are notoriously difficult to characterize at a structural level because they are difficult to crystallize, thus inhibiting X-ray crystallographic studies, and are both too large and too difficult to reconstitute in suitable media for NMR studies. Nevertheless, solid-state NMR methods are beginning to show promise that eventually such targets may be structurally characterized.

3.2 Macromolecule–Ligand Interactions

Macromolecule–ligand interactions are integral to a wide range of biological processes, including hormone, neurotransmitter or drug binding, antigen recognition, and enzyme–substrate interactions. Fundamental to each of these interactions is the recognition by a ligand of a unique binding site on the macromolecule. Through an understanding of the specific interactions involved it may be possible to design or discover analogous ligands with altered binding properties that might inhibit the biochemical function of the macromolecule in a highly specific manner. The study of macromolecule–ligand interactions thus forms the cornerstone of most structure-based drug design applications. The macromolecule of interest may be a protein or a nucleic acid, although the majority of drug design applications have focused on protein–ligand interactions. For this reason, we will refer mainly to protein–ligand interactions in the following discussion, but will also include some examples of drug–DNA interactions.

3.2.1 Overview

There are several important aspects of macromolecule–ligand interactions, which have a bearing on structure-based design. The simplest question that might be asked is "what is the strength of the binding interaction?" whereas the most detailed task would be to precisely define the atomic coordinates of the complete protein–ligand complex. In between these extremes there are many other questions that are important to the drug design process; these include questions about the binding stoichiometry and kinetics, the conformation of the bound ligand, and about the nature of functional group interactions between the protein and bound ligand. These and other important questions were introduced briefly in Table 3 and are examined in more detail later in this section. Before doing this it is first necessary to consider NMR timescales, because the ability of NMR methods to answer questions about macromolecule–ligand complexes depend critically on the kinetics of the binding interaction. Section 3.2.2 thus describes how various NMR parameters depend on binding kinetics and, in particular, how fast- and slow-exchange conditions affect the interpretation of NMR data.

Having identified the exchange regime, the task then becomes to decide which NMR parameters can be used to answer the questions posed above about the complex. Many of the NMR parameters that were described earlier for deriving information about ligands are also applicable to studies of complexes. These include chemical shifts, NOEs, and relaxation parameters. However, the presence of two interacting partners means that there are some differences in the way such parameters are measured, and this has led to the development of several techniques that are particularly important for the

study of macromolecule–ligand interactions, including chemical-shift mapping, isotope editing, and various NMR titrations. Section 3.2.3 describes these techniques. Finally, illustrative examples of the application of these techniques to specific drug design problems are given in Sect. 3.2.4.

3.2.2 Influence of Kinetics and NMR Timescales

Macromolecule–ligand interactions are characterized by an equilibrium reaction that potentially has a wide range of affinities and rates:

$$\mathrm{M} + \mathrm{L} \longleftrightarrow \mathrm{ML}$$

The rate constant for the forward reaction is referred to as the *on rate* (k_{on}), whereas dissociation of the complex is characterized by the reverse rate constant, k_{off}. The equilibrium constant for this interaction, represented in terms of the dissociation constant of the complex K_{D}, reflects a balance of the on and off rates, as shown in Eq. 2:

$$K_{\mathrm{D}} = \frac{[\mathrm{M}][\mathrm{L}]}{[\mathrm{ML}]} = \frac{k_{\mathrm{off}}}{k_{\mathrm{on}}} \qquad (2)$$

For many protein–ligand interactions, k_{on} is of the order of $10^8\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, and is typically quite similar for different ligands. The observation that K_{D} values may vary over a wide range, typically from millimolar to nanomolar (i.e. $K_{\mathrm{D}} = 10^{-3} - 10^{-9}$ M) for cases of interest, is a reflection of a variation in k_{off} for different ligands. Consideration of the k_{on} value above and the range of K_{D} values noted suggests a range in k_{off} from 10^{-1} to $10^5\ \mathrm{s}^{-1}$. The lifetime of the bound complex ($\tau_{\mathrm{ML}} = 1/k_{\mathrm{off}}$) may thus vary from much less than a millisecond to tens of seconds (10^{-5} to 10 s based on the above off rates). The exchange rate for the second-order binding process is given by:

$$k = \frac{1}{\tau} = \frac{1}{\tau_{\mathrm{ML}}} + \frac{1}{\tau_{\mathrm{L}}} = k_{\mathrm{off}}\left(1 + \frac{p_{\mathrm{ML}}}{p_{\mathrm{L}}}\right) \qquad (3)$$

where p_{ML} and p_{L} are the mole fractions of bound and free ligand, respectively.

The appearance of an NMR spectrum of a protein–ligand complex is dependent on the rate of chemical interchange between free and bound states. In particular, the effects of exchange on an individual NMR parameter (e.g. chemical shift, coupling constant, or relaxation rate) depend on the relative magnitude of the exchange rate and the difference in the NMR parameter between the two states. The cases where the rate of interchange is greater than, about equal to, or less than, the parameter difference are respectively referred to as *fast*, *intermediate*, and *slow* exchange, as indicated in Table 7.

Table 7 shows that the changes in chemical shifts on ligand binding (for signals either from the ligand or from the macromolecule) are in general greater than those for coupling constants or relaxation rates. Given that 100 s^{-1} might represent a typical exchange rate between free and bound states, it is clear that individual NMR signals may be found in either slow, fast, or intermediate exchange on the chemical-shift timescale, but it is more likely that couplings or relaxation parameters will be in fast exchange. Thus, in most cases where the term "NMR timescale" is used in the literature, it refers to the chemical-shift timescale. The table also emphasizes that there are two types of signals that can be monitored, those from the ligand and those from the macromolecule. In general, the typical magnitude of changes to chemical shifts or couplings of either

Tab. 7 NMR parameters and their changes on binding.

Parameter	*Difference*	*Typical magnitude*[a]
Chemical shift	$\nu_L - \nu_{ML}$	0–1000 s^{-1}
	$\nu_M - \nu_{ML}$	0–500 s^{-1}
Coupling constant	$J_L - J_{ML}$	0–12 s^{-1}
	$J_M - J_{ML}$	0–12 s^{-1}
Relaxation rate[b]	$1/T_L - 1/T_{ML}$	0–50 s^{-1} (for T_1, larger for T_2)
	$1/T_M - 1/T_{ML}$	0–10 s^{-1}

[a] Ranges are approximate only and larger effects may be seen in some cases.
[b] $1/T$ refers to either $1/T_1$ or $1/T_2$.

type of signal on binding are similar, although the changes to ligand signals may be larger than those from the macromolecule. However, changes to relaxation parameters for signals from ligands are much more likely to be greater than those for protein signals. This reflects the sensitivity of relaxation parameters to molecular mobility: a ligand undergoes a greater relative change in mobility on binding than does a protein, given that the relative increase in molecular weight in the complex is much greater for the ligand than for the protein.

The exchange regime (slow, intermediate, or fast) determines how a spectrum of a protein–ligand mixture changes during a titration, or as a function of temperature. Figure 14 schematically illustrates the various exchange regimes for macromolecule–ligand binding interactions. Slow exchange, corresponding to tight binding, is potentially the most useful regime, given that much detailed information on the nature of a complex can be deduced in this case. Nevertheless, fast exchange also allows valuable kinetic and thermodynamic parameters to be derived. The analysis is more complex for intermediate exchange and few quantitative studies are attempted for this situation.

3.2.2.1 **Slow exchange.** This situation applies when the rate of exchange is much slower than the difference in chemical shifts between the two states (i.e. $k \ll \nu B - \nu F$), where we now change to a nomenclature using subscripts to refer to the bound (B) and free (F) states. It should be understood that the signals may be derived from either ligand or macromolecule, so although B is always related to the ML complex, F might refer to either free ligand (L) or free macromolecule (M). In this situation, separate peaks are potentially observable for both free and bound states at their respective chemical shifts. Whether such signals are *actually* observed depends on the mole ratio of the individual species in a titration, and on the signals not being obscured by overlap or broadening.

Addition of a ligand to a solution of a protein results in the appearance of new signals attributed to bound protein resonances, with a concurrent decrease in the intensity of the free protein resonances, reflecting the decreased proportion of free protein during the titration. Once a stoichiometric mole ratio is achieved (usually 1 : 1, but sometimes 2 : 1 or higher if multiple binding sites are present on the protein), peaks from free ligand appear

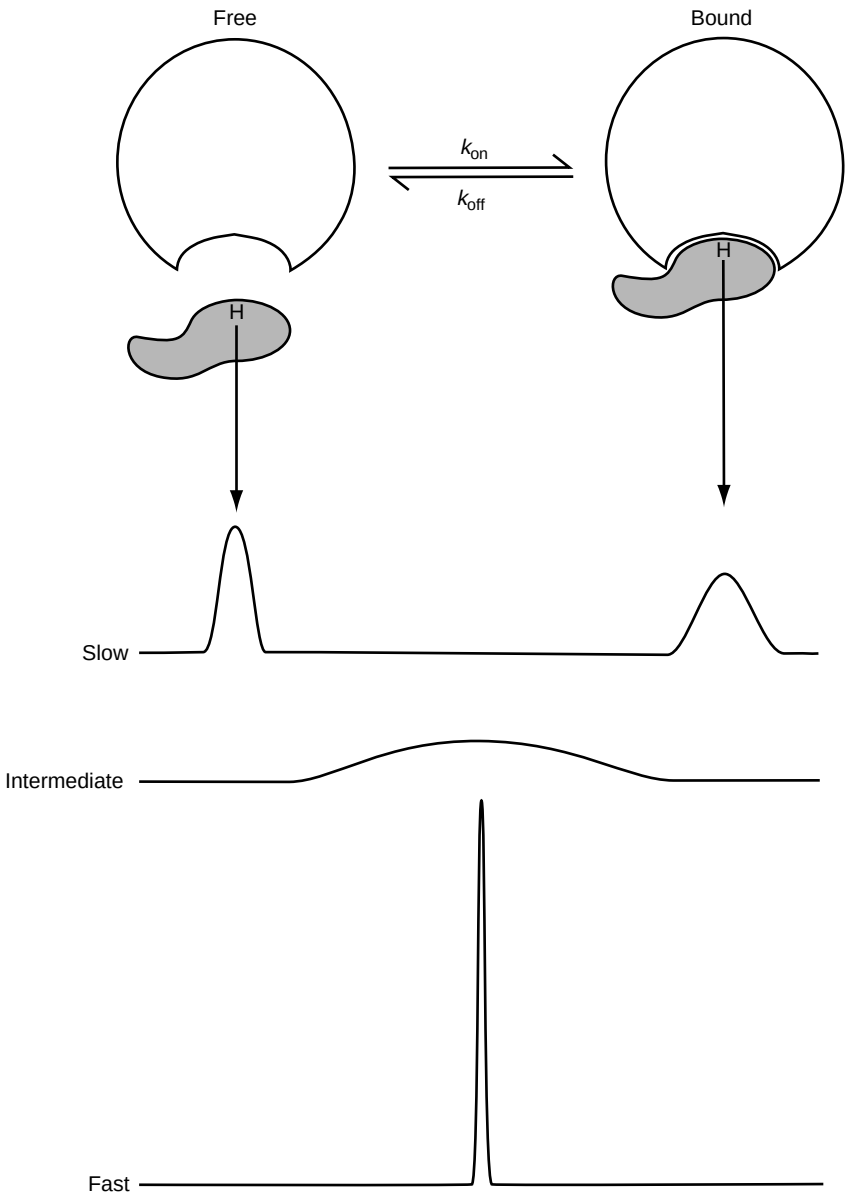

Fig. 14 Schematic illustration of the effects of slow, intermediate, and fast exchange on the appearance of peaks in NMR spectra of macromolecule–ligand complexes. In the slow-exchange case, separate peaks are seen for free and bound forms. Note the broader peak for the bound ligand because it now adopts the correlation time of the macromolecule. In the fast-exchange case, only an averaged peak is observed.

with increasing intensity as the excess of free ligand increases.

From such a titration it is possible to determine the stoichiometry of the complex, together with the chemical shifts of the bound states of the ligand and protein. In 1D NMR spectra, overlap of peaks makes it difficult to monitor more than a few resonances from either species, and such studies are most readily done when there is a well-resolved signal on one of the interacting species. Selective isotope labels have been used in the past for such studies, but it is now more common to use uniform ^{15}N- or ^{13}C-labeling of the protein and detect the chemical shifts in 2D HSQC spectra. It is often more difficult to label the ligand, but in some cases, the presence of rare nuclei such as ^{19}F can be used to advantage. A good example is the binding of the inhibitor 4-fluorobenzenesulfonamide to the enzyme carbonic anhydrase. Figure 15 shows ^{19}F spectra of the enzyme inhibitor complex at various mole ratios. The broadened peak for the bound ligand has a chemical shift of approximately 6 ppm and is in slow exchange with the peak from free ligand at 0 ppm. The stoichiometry of the complex in this case is 2 : 1, so that no signal from free ligand is visible until more than 2 mol of inhibitor are present. Addition of increasing amounts of ligand results in an increase in the free ligand signal, but no change in the bound ligand signal.

Determination of the binding constant from slow exchange spectra is not usually attempted. Generally, for slow-exchange conditions to exist in the first place, the binding is required to be submicromolar in affinity and non-NMR methods are more suitable for determining affinities in these cases. NMR studies are done at millimolar concentrations, making it difficult to determine K_D with any accuracy for tight binding systems.

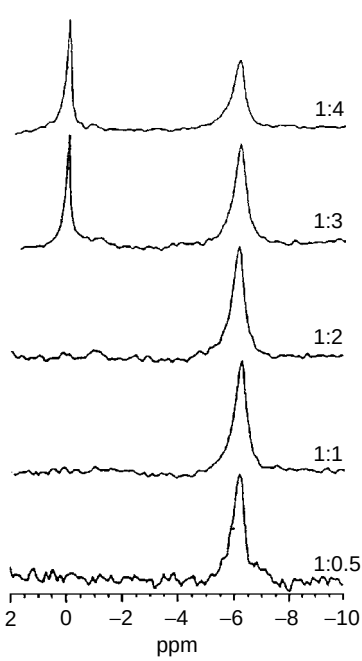

Fig. 15 ^{19}F NMR spectra at 282 MHz of the 4-fluorobenzenesulfonamide-carbonic-anhydrase-1 system at various ratios of enzyme to inhibitor, as indicated on the traces. The peak at ~6 ppm is caused by bound inhibitor. The enzyme concentration was 1 mM at pH 7.2 in D_2O at 25 °C. (Reprinted with permission from the American Chemical Society from Dugad, L., Gerig, J.T. (1988) *Biochemistry*, **27**, 4310.)

In principle, kinetic information on the complex can be obtained from slow-exchange spectra, as seen from the expressions for T_2 for free and bound ligand signals:

$$\frac{1}{T_{2\mathrm{F,obs}}} = \frac{1}{T_{2\mathrm{F}}} + k_{\mathrm{off}}\frac{p_{\mathrm{B}}}{p_{\mathrm{F}}} \quad (4)$$

$$\frac{1}{T_{2\mathrm{B,obs}}} = \frac{1}{T_{2\mathrm{B}}} + k_{\mathrm{off}} \quad (5)$$

Because the linewidth of a peak is related to T_2 by

$$\mathrm{LW} = \frac{1}{\pi T_2} \quad (6)$$

then measurements of linewidth during a titration can be used to derive k_{off}. Equations 4 and 5 show that, although the signal from bound ligand is independent of concentration, that of the free ligand decreases in linewidth as more ligand is added. A plot of linewidth versus ligand/macromolecule mole ratio allows k_{off} to be determined.

Although the determination of off rates is of significance in assessing the stability of the complex, the major interest in more recent studies of complexes in the slow-exchange limit has centered on determining the complete geometry of the complex through the use of intra- and intermolecular NOEs. A recent, but already classic, example of this approach is illustrated by the binding of immunosuppressant peptides such as cyclosporin and FK506 to their receptors. These types of examples are discussed in more detail in Sect. 3.2.4.2.

3.2.2.2 **Fast exchange.** When exchange between free and bound states is very fast, observed NMR parameters are a simple weighted average of those from the two contributing states, illustrated by Eq. 7 for chemical shifts and Eq. 8 for linewidths

$$\nu_{\mathrm{obs}} = p_{\mathrm{F}}\nu_{\mathrm{F}} + p_{\mathrm{B}}\nu_{\mathrm{B}} \quad (7)$$

$$\frac{1}{T_{2,\mathrm{obs}}} = \frac{p_{\mathrm{F}}}{T_{2,\mathrm{F}}} + \frac{p_{\mathrm{B}}}{T_{2,\mathrm{B}}} \quad (8)$$

These equations show that, in the fast-exchange limit, addition of a ligand to a protein solution will cause a progressive change in chemical shift. Signals from the protein initially reflect the free state, but as ligand is added, the population of bound protein increases, and the observed signals move toward those of the bound state. Similarly, when ligand signals are first detected, they predominantly reflect the bound state, but with increasing amounts of ligand they move toward the chemical shift of the free state. By regression analysis to Eq. 7, taking into account the dependency of the mole fractions on K_{D} by the standard quadratic binding equation, it is possible to obtain estimates of both K_{D} and the bound shift. The procedure works best for rather weakly binding ligands (e.g. millimolar dissociation constants).

When exchange is somewhat slower, but still within the fast-exchange limit, there is an exchange contribution to linewidth, as shown in Eq. 9:

$$\frac{1}{T_{2,\mathrm{obs}}} = \frac{p_{\mathrm{F}}}{T_{2,\mathrm{F}}} + \frac{p_{\mathrm{B}}}{T_{2,\mathrm{B}}} + \left\{\frac{p_{\mathrm{B}}p_{\mathrm{F}}^2\, 4\pi^2(\nu_{\mathrm{B}} - \nu_{\mathrm{F}})^2}{k_{\mathrm{off}}}\right\} \quad (9)$$

In this case, a maximum in the broadening of ligand or protein peaks occurs during the titration at a mole ratio of approximately 0.3, as illustrated below.

The spectral changes that occur in the fast-exchange regime can conveniently be illustrated by studies on the binding of a series of terephthalamide ligands to an

oligonucleotide model of DNA. The ligands, referred to as L(NO_2), L(NH_2), and L(Gly) (**6**–**8**), were synthesized as precursors for potential anticancer agents. To establish whether they bind in the minor groove of AT-rich DNA, a series of NMR titration experiments was undertaken

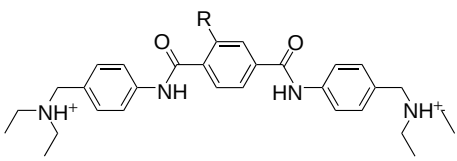

(**6**) L(NH_2) R = NH_2
(**7**) L(NO_2) R = NO_2
(**8**) L(Gly) R = $NHCOCH_2NH_2$

Figure 16 shows an expansion of the aliphatic region of a series of ^{1}H-NMR spectra of 0.5 mM of the oligonucleotide d(GGTAATTACC)$_2$, to which increasing amounts of L(NH_2) were added, of which the spectra cover mole ratios of ligand to DNA duplex ranging from 0 : 1 to 2.6 : 1. Although the spectra are complicated by overlap in some regions, it is clear that addition of the ligand causes significant changes to the DNA peaks. A typical example is seen for the T6 methyl peak, for which addition of ligand causes both an upfield shift and broadening of the peak at certain stages of the titration. The chemical shift moves monotonically with ligand concentration up to a mole ratio of 1 : 1 and then reaches a plateau, remaining constant as larger amounts of ligand are added. Broadening of the peak reaches a maximum at a ligand:DNA mole ratio of approximately 0.3. Both observations are consistent with there being moderately fast exchange on the chemical-shift timescale between the free and ligand-bound forms of the DNA in solution. In this case, the observed spectral peaks reflect neither the free nor the bound form of DNA, but are averaged signals.

Ligand peaks are also in fast exchange, as seen with the L(NH_2) methyl peak, which first appears at a ligand:DNA ratio of 1.36 : 1 as a shoulder on the overlapped T3 and T7 methyl peaks at approximately 1.27 ppm. This peak is not initially visible in spectra at low ligand:DNA mole ratios because of the small population of bound species and the overlapping DNA peaks. It moves upfield with increasing ligand concentration and, again, represents an averaged peak intermediate in chemical shift between free and bound forms, reflecting fast-exchange kinetics. Eventually, the chemical shift of this signal approaches that of the free ligand at 1.1 ppm, as measured in a separate experiment with a solution of ligand alone.

In fast-exchange cases such as this, it is possible to obtain an estimate of the dissociation constant for the complex (K_D) and the bound chemical shift (ν_B) of DNA resonances by fitting the observed chemical-shift changes as a function of ligand concentration to Eq. 7. The parameters that best fit the experimental data for the T6 methyl peak were $K_D = 1.2 \times 10^{-6}$ M and $(\nu_B - \nu_F) = 46$ Hz. To further define the thermodynamic constants associated with binding, the linewidth data were also quantitatively examined by use of Eqs. 6 and 9. In the case of moderately fast exchange, a maximum linewidth is predicted at a ligand:DNA mole ratio of 0.33, and this was indeed observed. Derived binding parameters were $K_D \leq 1.0 \times 10^{-6}$ M, $k_{off} \approx 250\ s^{-1}$, $(\nu_B - \nu_F) = 49$ Hz, and $LW_B = 12$ Hz, consistent with the values derived from the analysis of chemical shifts. Subsequent studies with the related ligand L(NO_2) showed similar binding to L(NH_2). However, a third ligand, L(Gly), was found to bind somewhat more tightly,

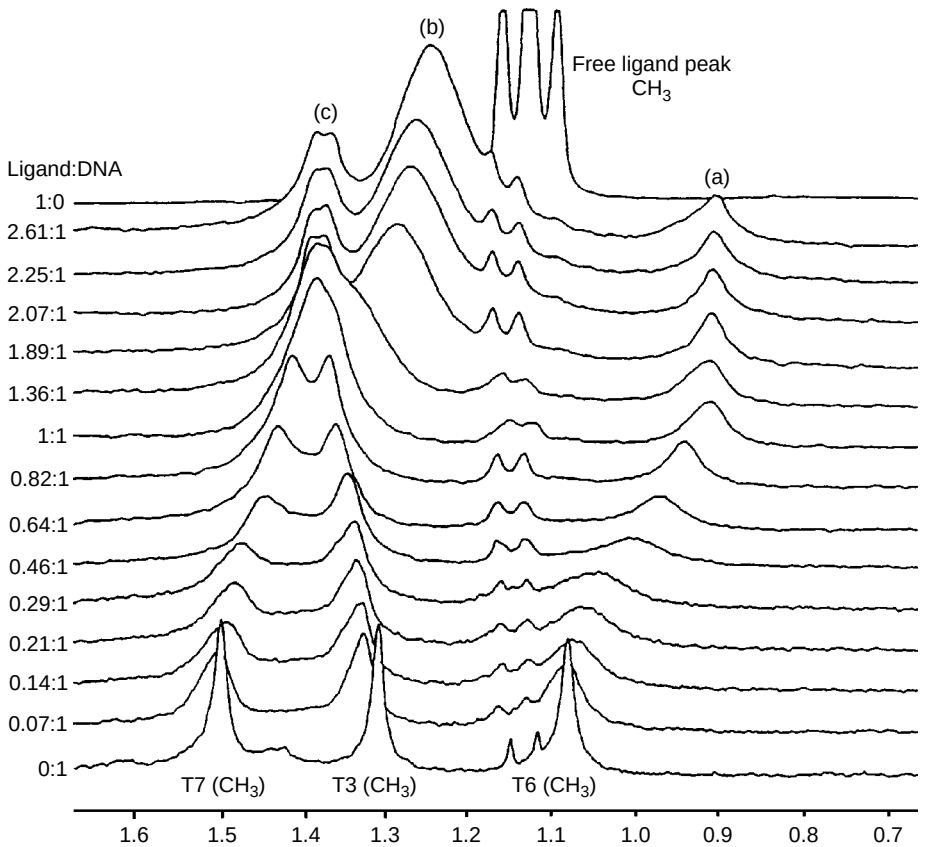

Fig. 16 Expanded regions from 300-MHz ^{1}H-NMR spectra for complexes between L(NH_2) and d(GGTAATTACC)$_2$ recorded at 10 °C. The two small peaks at 1.12 and 1.14 ppm arise from an impurity. Increasing ligand concentration causes an upfield shift of the T6 methyl resonance (a), and causes the T7 and T3 resonances to become overlapped at later stages of the titration (c). Peak (b) is an averaged resonance from the ligand methyl groups intermediate in shift between the bound and free forms of the ligand. (Reprinted with permission from Pavlopoulos, S., Rose, M., Wickham, G., Craik, D.J. (1995) *Anticancer Drug Des.* **10**, 623.)

with some signals in the intermediate exchange regime.

3.2.2.3 **Intermediate exchange.** In this regime, the rate of exchange between bound and free states is comparable to the differences in NMR parameters associated with the exchange. In general, the spectral peaks often become very broad and analysis is difficult. This is the case, for example, for L(Gly). In the methyl region of the spectra shown in Fig. 17, the T7 CH_3 signal moves upfield and the T3 CH_3 signal moves slightly downfield with increasing ligand concentration, as seen previously for L(NH_2) and L(NO_2). However, in contrast to the case for the other ligands, the characteristic broadening of peaks at intermediate ratios is non-Lorentzian, suggesting kinetics in the

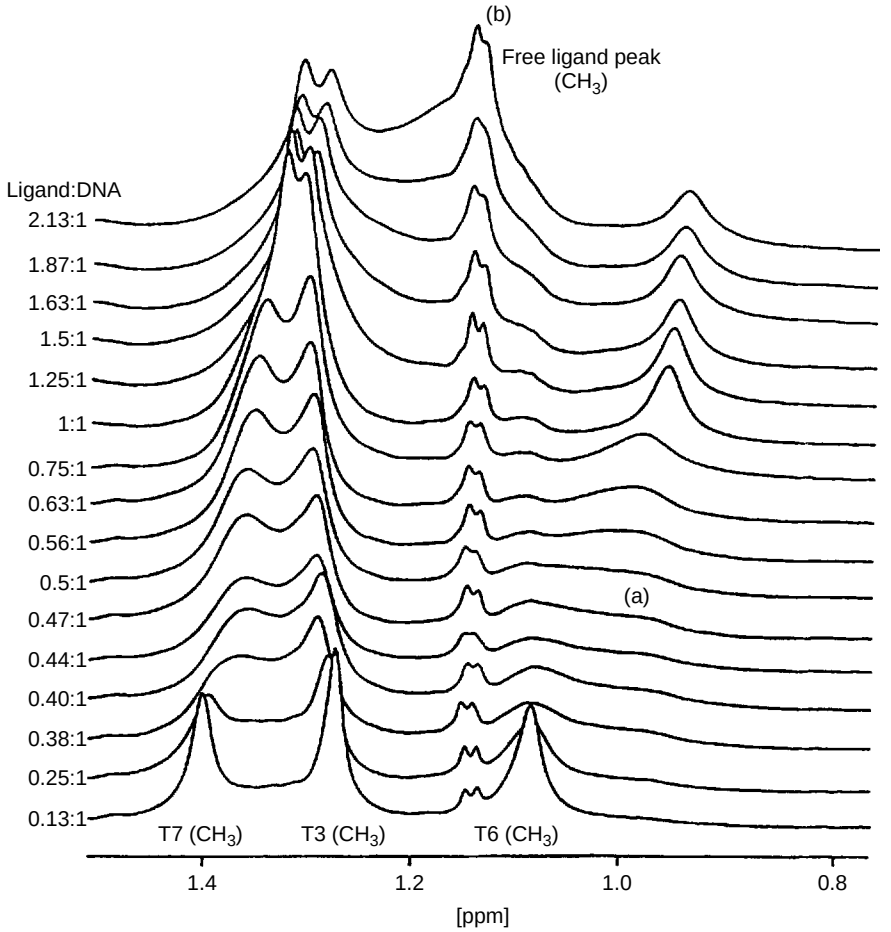

Fig. 17 Expansions from the 600-MHz ^{1}H-NMR spectra for complexes formed between L(Gly) and d(GGTAATTACC)$_2$ showing the methyl resonances. The two small peaks at 1.12 and 1.14 ppm are attributed to an impurity. The complex nature of the T6 methyl resonance at ligand:DNA ratios less than 1 : 1 (a), and the manner in which signal intensity increases at about 1.15 ppm at DNA:ligand ratios greater than 1 : 1 (b), are indicative of intermediate exchange. (Reprinted with permission from Pavlopoulos, S., Rose, M., Wickham, G., Craik, D.J. (1995) *Anticancer Drug Des.*, **10**, 623.)

intermediate exchange regime. The T6 CH_3 peak does not shift in the characteristic fast-exchange manner, but instead, a new broad resonance appears close to the expected position of the bound T6 CH_3 chemical shift on the first addition of ligand, and increases in intensity with increasing ligand concentration. This observation is consistent with the ligand being in slow to intermediate exchange between the free and bound forms, with $k_{off} \approx (\nu_B - \nu_F)$. On the basis of the magnitude of $\nu_B - \nu_F$ for this resonance, k_{off} for L(Gly) is estimated to be 50 s^{-1}, which is significantly slower than that for L(NO_2) and L(NH_2).

At a ligand:DNA ratio of approximately 1 : 1, the ratio of the integrals of the T6 methyl peak and the overlapped T3 and T7 methyl peaks is about 1 : 6. The expected value is 1 : 2, which indicates that the bound ligand methyl peak ($4 \times CH_3$) is overlapped with the T7 and T3 methyl peaks, as observed with $L(NH_2)$ and $L(NO_2)$. When the ligand:DNA ratio is increased beyond a 1 : 1 ratio, a new peak appears at about 1.15 ppm and increases in intensity as the ligand concentration is increased. This new peak corresponds to the methyl peak of the free ligand and its appearance in this manner is consistent with slow exchange on the chemical-shift timescale. To confirm this, spectra of a 2 : 1 mixture of L(Gly) and $d(GGTAATTACC)_2$ were acquired at different temperatures, as illustrated in Fig. 18.

At low temperatures, signals at 1.15 and 1.30 ppm (overlapped with the T7 CH_3 and T3 CH_3 peaks) attributable to the methyl groups from the free ligand and bound ligand, respectively, are distinguishable. As the temperature is increased, a broad peak appears between these two signals (at ~1.22 ppm). At the lower temperatures $k_{off} \leq (\nu_B - \nu_F)$, so that methyl resonances of the ligand have complex characteristics reflecting slow-intermediate exchange. At higher temperatures, $k_{off} \geq (\nu_B - \nu_F)$, so the signal appears as a fast-exchanged average between the free and bound resonances. From a qualitative analysis of the spectra, k_{off} for L(Gly) was estimated to be 50–60 s^{-1} at 283 K.

The fact that some peaks (e.g. the oligonucleotide T7 and T3 methyl signals)

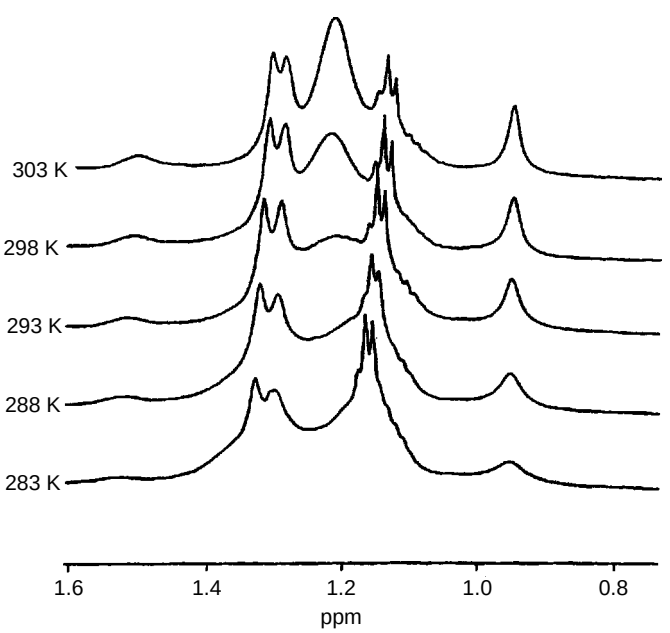

Fig. 18 Expansions of ^{1}H-NMR spectra of a 2 : 1 mixture of L(Gly) and $d(GGTAATTACC)_2$ acquired at different temperatures. (Reprinted with permission from Pavlopoulos, S., Rose, M., Wickham, G., Craik, D.J. (1995) *Anticancer Drug Des.*, **10**, 623.)

exhibit fast exchange, whereas others in the same spectrum of the same complex exhibit slow-intermediate characteristics, is a reflection of the different ($\nu_B - \nu_F$) values for different peaks. This emphasizes the point made earlier that the "exchange regime" is a relative expression and depends not only on the rate of exchange but also on the size of the chemical-shift differences involved. In summary, the observations suggest that k_{off} for binding of L(Gly) to the oligonucleotide duplex is much slower than that for the other two derivatives. This provides an illustration of the value of NMR as a quick method for comparing the binding of different ligands and for confirming ligand-binding hypotheses.

The change in binding kinetics may be rationalized by considering the different structure of the L(Gly) ligand relative to the other ligands. It was anticipated that, upon binding to the minor groove, the terephthalamides would adopt a conformation in which the substituent on the central ring would form part of the convex edge of the ligands and therefore be directed toward the "mouth" of the groove. Given this binding arrangement, the ligand L(Gly) would have a positively charged alkylamine group positioned to interact with the negatively charged phosphate groups of the DNA backbone. The L(Gly) derivative also has a bulkier substituent than that of the other ligands, and this is also consistent with some differences in its binding.

3.2.3 NMR Techniques

NMR is a particularly versatile tool for the analysis of protein–ligand interactions. As well as being able to observe different nuclei, measurements may be made of a range of different NMR parameters, including chemical shifts, linewidths, coupling constants, and relaxation parameters. In addition, there are several specific NMR techniques that have been applied for the measurement of these parameters. The techniques that are particularly valuable for the study of macromolecule–ligand interactions are described in the following sections.

3.2.3.1 **Chemical-shift mapping.** Chemical shifts are exquisitely sensitive markers of the local charge state and environment. Although it is not possible to construct an accurate model of a binding site from knowledge of the chemical shifts of a bound ligand, a qualitative interpretation of *changes* in chemical shifts of the macromolecule on binding provides significant insight into the location of the binding site. Traditionally, such studies were done using 1D NMR, but are now increasingly done by 2D HSQC spectra. By simultaneously obtaining information on chemical shifts for a large number of sites in a macromolecule and seeing which ones change when a ligand binds, and which ones do not, it is possible to deduce the location of the binding site. This procedure is referred to as *chemical-shift mapping*. A prerequisite of the approach is that the chemical shifts have to be assigned. Chemical-shift mapping by use of HSQC spectra is widely used in NMR screening approaches and we will defer a more detailed discussion on it until Sect. 4.

The relative simplicity of how chemical-shift information localizes binding sites may be illustrated by continuing with the example introduced above of terephthalamide binding to DNA. Figure 19 shows that, upon binding of the terephthalamides $L(NH_2)$ and $L(NO_2)$ to $d(GGTAATTACC)_2$, the DNA protons on the four base pairs between A5 and A8 are perturbed to a much larger degree than protons in the rest of

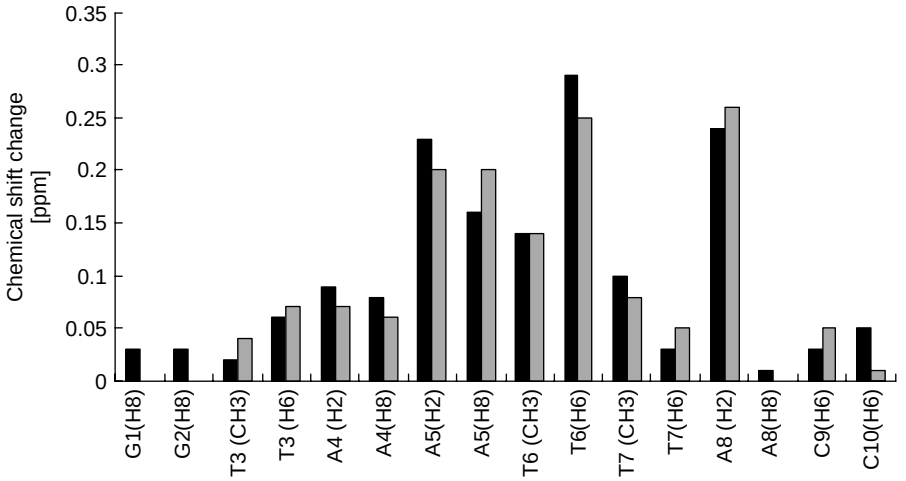

Fig. 19 Chemical-shift perturbations of DNA protons upon ligand binding. The lighter and darker columns represent shifts attributed to $L(NO_2)$ and $L(NH_2)$ derivatives, respectively.

the sequence. It is thus likely that these four residues form the binding site.

A more detailed analysis allows the binding site to be further localized to the minor, rather than to the major, groove in the region of these bases. A4, A5, and A8 are the only residues containing easily detectable minor groove protons (H2). These resonances, which originate from the floor of the minor groove, are shifted downfield with ligand binding, whereas most other resonances are shifted upfield. This observation is consistent with the ligands binding in the minor groove and has been reported for other minor groove binders such as Hoechst 33258 and SN-6999, where adenine H2 protons on the floor of the groove experience deshielding ring-current effects. However, significant chemical-shift changes were also observed for some major groove protons. This illustrates the general point that sometimes, allosteric effects can cause changes at sites not directly involved in binding. In the case of DNA, binding perturbations in the major groove have also been observed for other established minor groove binders such as distamycin, netropsin, and Hoechst 33258. On the basis of NOE and crystallographic data, it was concluded that the effects were caused by distortions of the B-DNA duplex, including changes in the ''base roll'' of residues within the binding site, upon complexation. Electronic effects arising from the close proximity of charged groups on the ligand to neighboring nucleotides were also found to perturb major groove resonances.

In the case of the terephthalamides, a comparison of the minor and major groove perturbations for a particular residue shows that the minor groove protons are affected to a much greater extent. This is particularly evident for A8, where the H2 proton shifts by approximately 0.25 ppm and the H8 proton is not affected (Fig. 19). It is difficult to conceive of a binding mode in the major groove that would account for such a large effect on the minor groove A8 H2 resonance without a simultaneous effect on the major groove protons of T7

and A8. The observed 1 : 1 stoichiometry of the complex excludes the possibility that the ligand binds to the major and minor groove at the same time. It is more likely, therefore, that binding in the minor groove causes distortion of the DNA structure so that perturbations are observed for the major groove protons of A5 and T6, but not neighboring nucleotides.

Other examples of the use of chemical-shift mapping to locate binding sites have been made for ligands binding to a range of drug targets, including immunophilins, matrix metalloproteases, and DHFR. Some of these examples are described in more detail in Sect. 3.2.4

3.2.3.2 **NMR titrations.** There are a number of advantages in undertaking a titration of ligand against macromolecule or vice versa rather than just examining the final complex. These include introducing the possibility of distinguishing signals from the individual components on the basis of intensities at intermediate stages of the titration in the slow-exchange case and obtaining kinetic and thermodynamic parameters associated with the interaction in the fast-exchange case. Such titrations may be done using either 1D or 2D spectra and are very useful for establishing the exchange regime of the complex, as described in Sect. 3.2.2. A variety of parameters may be monitored in the titration, although the two most common are chemical shifts and linewidths. Examples of such titrations are given in Figs. 16 and 17.

3.2.3.3 **Isotope editing and filtering.** Isotope editing provides a powerful way of distinguishing between the components in a complex without the need for a titration. It is one of the most useful tools for the study of macromolecule–ligand complexes, and indeed the background NMR technology that underpins isotope editing was developed specifically for the study of complexes. The principle of the approach is illustrated in Fig. 20 and is based on the use of isotopes to select for signals from either the ligand or macromolecule, or signals exclusively linking both of them.

The conceptually simplest approach is to uniformly deuterate the macromolecule, thereby removing its signals from ^{1}H-detected NMR spectra, and allowing signals from only the ligand to be observed. This substantially simplifies the spectrum and allows, for example, the bound conformation of a ligand to be determined from NOESY data recorded in D_2O. By rerunning the spectrum in H_2O, additional NOEs to exchangeable amide protons on the protein may be detected, thereby providing information on contacts between ligand and protein. Alternatively, ^{15}N or ^{13}C signals may be introduced selectively into either the ligand or protein, and editing techniques used to select only signals attached to these labels and their proximate protons.

Potentially, the most useful approach involves uniform labeling of one of the components with either ^{15}N or ^{13}C and leaving the other component unlabeled. It is then possible to edit the spectrum by selecting for interactions (either through bond or through space) that connect protons that are both one-bond coupled to ^{15}N or ^{13}C. Alternatively, the spectrum may be filtered to specifically remove such signals, thereby selecting only signals involving protons coupled to ^{14}N or ^{12}C (i.e. on the unlabeled component). It is generally easier to uniformly label the protein rather than the ligand, and editing methods are highly efficient, thus making it easy to visualize just the protein. However, because ligand signals

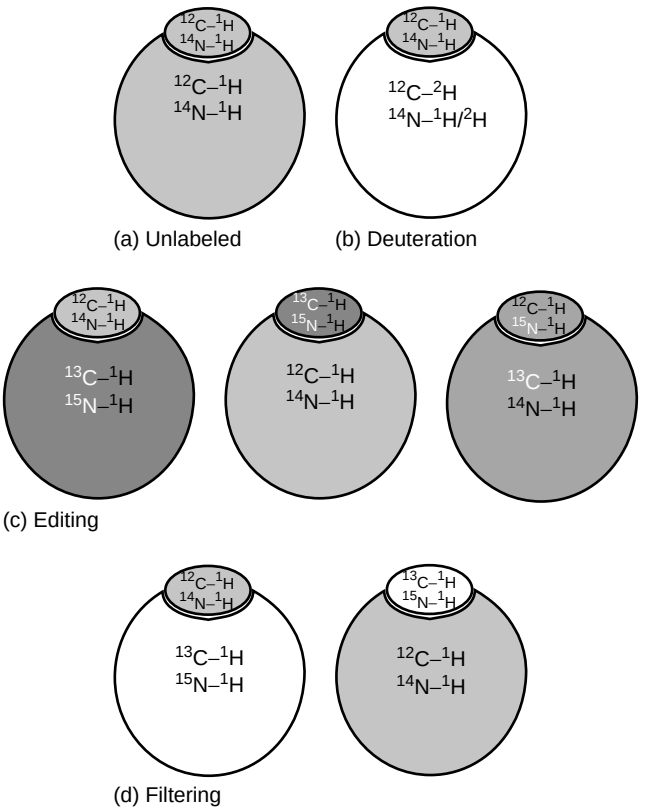

Fig. 20 Isotope editing and filtering can be used to select signals from either the ligand or the protein. (a) Normal protein and ligand with no filtering or editing. (b) Selection of the ligand signals by ^{2}H labeling of the protein. (c) Selection of protein or ligand signals by ^{13}C and/or ^{15}N labeling/editing. (d) Removal of protein or ligand signals by ^{13}C or ^{15}N filtering.

are often of interest, filtering experiments play a valuable role in visualizing them. Unfortunately, filtering experiments are more susceptible to artifacts than are editing experiments.

Another possibility is to use half-edited/half-filtered 2D experiments to detect NOEs that specifically involve interactions between protons attached to ^{15}N or ^{13}C and those that are not. This approach is used, for example, to detect intermolecular NOEs between a labeled protein and an unlabeled ligand. Examples of isotope editing/filtering are given in Sect. 3.2.4.

3.2.3.4 **NOE docking.** In many cases, the study of a complex may follow a previous structure determination of the isolated macromolecule and in that case it may be possible to determine much information about a complex by obtaining a relatively small number of NOEs linking the ligand and macromolecule. Gradwell and Feeney recently analyzed factors important

in such NOE docking experiments. In their analysis, a high-resolution X-ray structure of a protein–ligand complex was used to simulate loose distance restraints of varying degrees of quality that might typically be estimated from experimental NOE intensities. These simulated data were used to examine the effect of the number, distribution, and representation of the experimental constraints on the precision and accuracy of the calculated structures. A standard simulated annealing protocol was used, as well as a more novel method based on rigid-body dynamics. The results showed some parallels with those from similar studies on complete protein NMR structure determinations, but it was found that more constraints per torsion angle are required to define docked structures of similar quality. This is because the conformation and orientation of the ligand are defined only by NOEs and not by covalent attachment, as is the case for amino acid side chains in a protein structure. The effectiveness of different NOE-constraint averaging methods was explored and the benefits of using "R^{-6} averaging" rather than "center averaging" with small sets of NOE constraints were demonstrated. With these considerations in mind, it appears that NOE docking can be a very cost-efficient procedure for defining the environment, orientation, and conformation of ligands.

3.2.4 Selected Examples

Applications of the various NMR techniques described are now illustrated with selected examples. The examples have been chosen to give a broad perspective on the types of NMR experiments that can be done and the types of information they provide. Specifically, the first example covers the case of drug-nucleic acid binding and focuses on more traditional NMR experiments, involving relatively standard homonuclear methods. The second example covers binding of moderately large ligands to immunophilins and highlights modern isotope editing techniques. The third example, covering ligand binding to a matrix metalloproteinase (MMP), also highlights the importance of these techniques and shows how relatively simple spectra involving ^{19}F-containing ligands can be very informative. The fourth example describes ligand binding to DHFR, one of the most extensively studied systems by NMR, and illustrates the derivation of a broad range of kinetic and geometric information on intermolecular complexes. The final example, on HIV protease, describes how NMR complements X-ray studies and provides information on dynamic motions within complexes.

3.2.4.1 **DNA-binding drugs.** The NMR approaches that have been used to examine the interactions of minor groove binding drugs with DNA can be illustrated with studies on the bisbenzimidazole-based compound, Hoechst 33258, (**9**). It has been used widely as a fluorescent cytological DNA stain and is also active as an anthelmintic agent. It has activity against intraperitoneally implanted L1210 and P388 leukemias in mice.

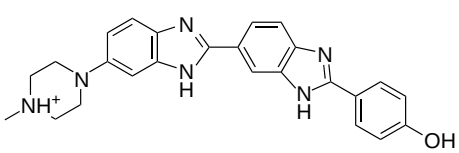

(**9**)

Footprinting studies have shown that sequences of four AT base pairs are a prerequisite for strong binding to DNA, consistent with similar observations for other structurally related molecules such

as distamycin and netropsin. The first structural studies of Hoechst 33258 complexed to short sequences of synthetic oligonucleotides were done using X-ray crystallographic methods, but NMR and further X-ray studies followed. Three of the X-ray studies used the *Eco*RI sequence d(CGCGAATTCGCG)$_2$, and another used the sequence d(CGCGATATCGCG)$_2$. Both sequences fulfil the requirement of at least four consecutive AT base pairs, and the resulting complexes showed similar modes of binding. In all of the X-ray studies, the Hoechst ligand was found to bind to the minor groove.

The NMR studies of complexes between Hoechst 33258 and oligonucleotide sequences provided complementary information to the crystal structure data. Because the binding is reversible, the NMR data offer the opportunity to derive information about the kinetics of the interaction. As with the crystallographic studies, the oligonucleotide sequences were designed to contain runs of AT base pairs. Some NMR studies were performed with dodecanucleotide sequences used in crystallographic studies, including d(CGCGAATTCGCG)$_2$, which allowed a direct comparison with the crystallographic data. Experiments were also performed with sequences specifically designed to investigate different aspects of the interaction. The sequence d(CTTTTGCAAAAG)$_2$ was designed to offer two binding sites, and it was shown that two Hoechst molecules interacted with the DNA duplex in symmetry-related orientations at the 5′-TTTT-3′ and 5′-AAAA-3′ sites.

Stoichiometry and kinetics The starting point in studies of ligand–DNA complexes is usually a titration experiment to establish the nature and stoichiometry of the complex. Complexes between the ligand and DNA duplex are obtained by adding small aliquots of ligand solution to a sample of the DNA duplex with one-dimensional ^{1}H NMR spectra acquired after each addition. The effects observed on the NMR spectrum after each addition reveal whether an interaction is taking place and allow the interaction to be characterized as fast or slow exchange on the NMR timescale. The stoichiometry of the interaction can also be determined from the titration.

In general, the addition of Hoechst 33258 to the oligonucleotide duplexes causes a decrease in the intensity of free DNA resonances and a concomitant increase in the intensity of new resonances, which appear in previously unoccupied spectral regions. This is consistent with the free and bound forms of the DNA duplex being in slow exchange with each other. For example, when Hoechst 33258 is added to d(GGTAATTACC)$_2$, the free DNA signals completely disappear at a DNA:drug ratio of 1 : 1, and the number of new resonances is seen to be twice the number of previously observed free DNA resonances (Fig. 21). This is a common feature of complexes with 1 : 1 stoichiometry and reflects a loss of the dyad symmetry of the duplex attributed to ligand binding.

Upon addition of Hoechst 33258 to d(CTTTTCGAAAAG)$_2$, the free DNA signals completely disappeared at a ratio of 2 : 1 drug:DNA and there was no doubling of the number of DNA resonances in the spectrum. From this, it could be concluded that two molecules were bound per duplex in a manner that retained the dyad symmetry of the DNA duplex. The binding was also determined to be cooperative, in that no intermediate 1 : 1 complex was detected. The formation of a 1 : 1 complex

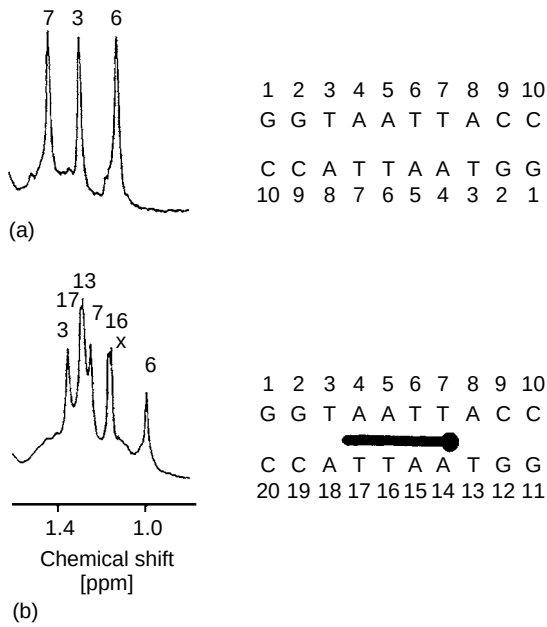

Fig. 21 1D ^{1}H-NMR spectra (recorded at 20 °C) illustrating the thymine methyl region for the symmetrical ligand-free duplex (a) and for the 1 : 1 Hoechst:d(GGTAATTACC)$_2$ complex (b), which is no longer symmetrical because of the ligand binding. x corresponds to a small impurity peak. The DNA strands are numbered to the right of the spectra and the approximate location of the ligand is indicated by a black bar. (Reprinted with permission from Blackwell Publishing Science from Embrey, K.J., Searle, M.S., Craik, D.J. (1993) *Eur. J. Biochem.* **211**, 437.)

would have resulted in a very complicated spectrum at intermediate ligand:DNA ratios, given that resonances arising from the free DNA, the 1 : 1 complex, and the 2 : 1 complex, would have produced four times as many observable peaks relative to the free DNA species. At intermediate ligand concentration, however, only two sets of peaks arising from DNA molecules were detected. In the 2 : 1 complex, only four thymine methyl resonances were detected (1.0–1.5 ppm), as expected for a symmetrical DNA duplex. These are all overlapped in the free DNA spectrum. In the 1 : 1 mixture, only signals from free DNA and from the 2 : 1 complex were detected.

The reversible nature of the Hoechst:DNA interaction is illustrated by the observation of chemical-exchange cross peaks in NOESY spectra of mixtures of free and complexed oligonucleotides. This may be seen in the NOESY spectrum of a mixture of free and complexed d(CTTTTCGAAAAG)$_2$, shown in Fig. 22, in which many chemical-exchange cross peaks are observed between resonances arising from the free and bound oligonucleotide. In a NOESY spectrum acquired at lower temperature, the intensity of these chemical-exchange cross peaks is significantly reduced, indicating that the exchange is slowed at lower temperatures. The exchange rate was estimated to be $<10\ s^{-1}$ at 10 °C.

The ability to observe such dynamic exchange phenomena is one of the strengths of NMR relative to X-ray crystallography, and several examples of these phenomena are described later in the chapter.

Binding site A combination of chemical shift and NOE information can be used to locate and characterize binding sites. Chemical-shift differences between resonances arising from free and bound forms of DNA are indicative of the nature of the interaction. In all studies of

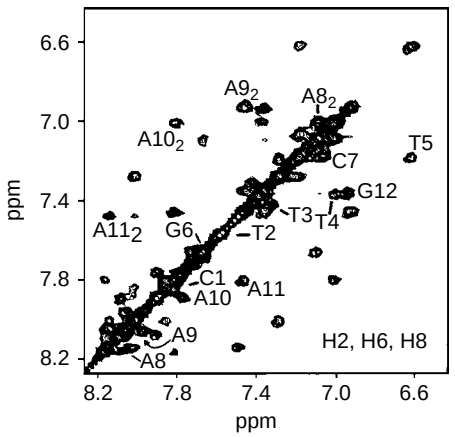

Fig. 22 Aromatic region of NOESY spectrum of a 1 : 1 mixture of Hoechst versus d(CTTTTCGAAAAG)$_2$ recorded with a 200-ms mixing time. Chemical-exchange cross peaks between protons of the free DNA and the 2 : 1 Hoechst:DNA complex are labeled with their identifying base pair. Below the diagonal the H6 and H8 cross peaks are shown, whereas those of the adenine H2 resonances are highlighted in the upper portion of the figure and labeled with a subscript 2. (Reprinted with permission from Oxford University Press from Searle, M.S., Embrey, K.J. (1990) *Nucleic Acids Res.* **18**, 3753.)

the Hoechst complexes described above, significant changes to the chemical shifts of thymine H1′ protons and adenine H2 protons were observed, in contrast to the generally small perturbations observed for the base H8/H6 and CH_3 resonances located in the major groove. Perturbations of this nature are consistent with binding to the minor groove. In some instances, significant perturbations were observed to major groove protons located well within the binding site, reflecting changes in the conformation of the DNA duplex (e.g. base roll, propeller twisting).

Further evidence of minor groove binding is provided by the fact that resonances arising from protons on the floor of the groove, such as the adenine H2 and imino resonances, are shifted downfield, whereas resonances from protons on the minor groove walls, such as the H1′ protons, are shifted upfield. This is a consequence of the ligand's being inserted edge-on into the minor groove. The deoxyribose protons that form the walls of the minor groove are positioned above the π-plane of the aromatic rings and consequently receive upfield perturbations to their chemical shifts. Protons positioned on the floor of the groove, however, generally lie in the plane of the aromatic rings and experience downfield perturbations to their chemical shifts, as illustrated in Fig. 23.

The magnitude of chemical-shift changes is a strong indicator of the location of the binding site. In the case of the 2 : 1 complex with d(CTTTTCGAAAAG)$_2$, the largest chemical-shift changes occur over the 5′-TTTT-3′ and 5′-AAAA-3′ regions of the duplex. In the case of 1 : 1 complexes, where the DNA duplex contains an AT base-pair segment located at the center of the sequence, greater chemical-shift perturbations are observed for resonances in that region, consistent with the binding site being located there.

Assignment of the bound ligand and DNA resonances enables the identification of intermolecular NOEs, which are required for a precise determination of the binding site. The interaction of Hoechst 33258 with the oligonucleotides produced a large number of intermolecular NOEs (~25–30), placing considerable constraints on the structure of the complex and enabling the orientation of the ligand within the binding site to be determined. The NOE contacts observed for different complexes have a few features in common. The contacts generally involve

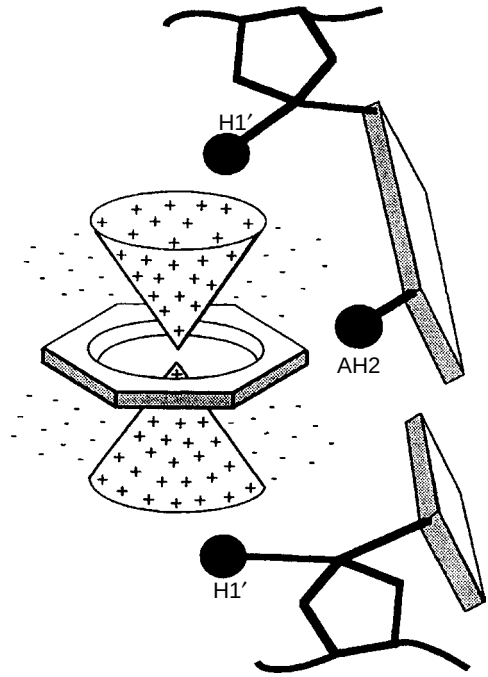

Fig. 23 Schematic representation of ligand-induced ring-current effects on nucleotide protons that form the walls (deoxyribose H1′) and floor (adenine H2) of the minor groove. (+), shielding effects; (−), deshielding effects. (Reprinted with permission from Blackwell Publishing Science from Embrey, K.J., Searle, M.S., Craik, D.J. (1993) *Eur. J. Biochem.* **211**, 437.)

DNA protons associated with the minor groove, such as ribose H1′ and adenine H2, clearly locating Hoechst in the minor groove. Protons of all four spin systems of the ligand show NOEs to protons of the DNA, demonstrating that the interaction occurs along the entire length of the drug. Typically, protons along one edge of the ligand (e.g. NH and H4′/H4′′′′) exhibit close contacts to protons on the floor of the minor groove, showing that the bound drug is crescent shaped and isohelical with the DNA.

Models of the interaction of Hoechst 33258 with the oligonucleotides studied were generated on the basis of the intermolecular NOEs. The models of the 1 : 1 complexes indicated that the ligand interacted with the four AT base pairs located at the center of the sequence. Interestingly, there was no evidence for interactions with GC base pairs on the periphery of the binding sites. In the 2 : 1 complex, the array of contacts observed located the ligand in the minor groove at the center of the 5′-TTTT-3′ and 5′-AAAA-3′ sites, as illustrated in Fig. 24.

As well as defining the *location* of the binding site, intermolecular NOEs can be used to determine the *orientation* of the ligand at that site. In the case of the 2 : 1 complex, the *N*-methylpiperizine moieties were found to point toward the center of the duplex, as indicated by NOEs between the protons from the piperizine ring and the 5′-terminus of the adenine tract (Fig. 24). Corresponding NOEs were also observed between the drug phenolic protons and the 5′-terminus of the thymine tract, as well as the 3′-terminus of the adenine tract of the complementary strand. This model did not indicate any interaction with the central GC base pairs.

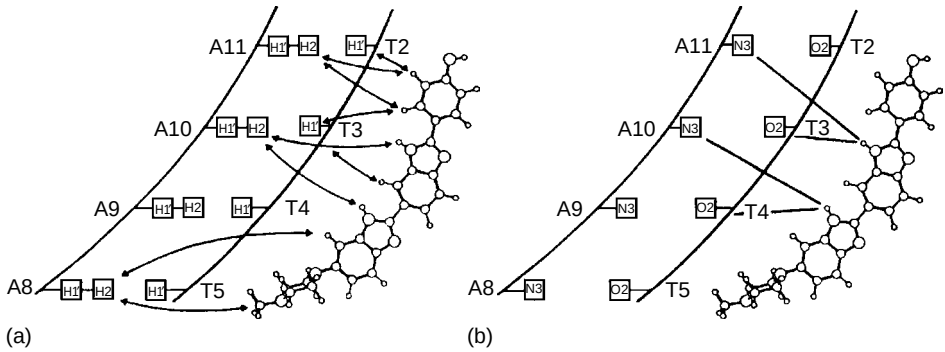

Fig. 24 Schematic representation of Hoechst 33258 bound to the minor groove of the 5′-TTTT sequence. (a) Highlights of some of the NOEs that determine the position and orientation of the Hoechst molecule within the minor groove. (b) Intermolecular hydrogen-bonding scheme. Molecular-modeling studies with an idealized B-DNA helical structure indicate that the benzimidazole H3′ is capable of forming bifurcated interactions with A11N3 and T302, whereas the benzimidazole H3″ hydrogen bonds in a similar manner, but with A10N3 and T402. In the proposed model of the complex, these distances fall within 3.5 Å and are thus within acceptable hydrogen-bonding limits. (Reprinted with permission from Oxford University Press from Searle, M.S., Embrey, K.J. (1990) *Nucleic Acids Res.* **18**, 3753.)

The orientation of the ligand was similarly determined in the 1 : 1 complexes on the basis of the intermolecular NOEs between protons located at the extremities of the Hoechst molecule and protons of the binding site. For example, in the interaction with $d(GTGGAATTCCAC)_2$, NOEs between protons from the piperizine moiety and the H2 and H1′ protons of the dinucleotide fragment d(A5T5)·d(A6T6) were reported.

Dynamic processes The binding of the Hoechst molecule to the self-complementary oligonucleotide duplexes in a 1 : 1 ratio lifts the dyad symmetry of the duplexes so that two sets of DNA resonances are observed. This indicates that the drug is in slow exchange between the free and the bound forms. Close examination of the 2D NOE data, however, reveals the presence of chemical-exchange cross peaks between symmetry-related protons on opposite sides of the dyad axis of the DNA duplex. The mechanism by which this occurs has been described as dissociation of the Hoechst molecule from the duplex, followed by a 180°-reorientation and rebinding. The self-complementary nature of the sequences ensures that the same complex is formed for either ligand orientation but with the net effect of interchanging the two strands with respect to the orientation of the Hoechst molecule. The rate at which this process occurs was estimated using cross-peak intensities in the NOESY spectrum. When interacting with $d(GGTAATTACC)_2$ and $d(GTGGAATTCCAC)_2$, the lifetime of the complex in each state ($1/k_{ex}$) was reported to be approximately 0.8 and 0.45 s, respectively. These values indicate a small but significant difference in the affinity of Hoechst for TAATTA and GAATTC sites.

Intramolecular dynamic processes that are fast on the NMR timescale are also observable in the ^{1}H-NMR spectrum of the bound Hoechst molecule. Resonance

averaging is observed for the H2/H6 and H3/H5 protons of the phenol group, which is consistent with the environments on either side of the ring being averaged by rapid ring-flipping motions about the C4–C2′ axis. This occurs despite the apparent tight fit between the phenyl ring and the walls of the minor groove, which, in a static model of the complex, must present a large barrier for rotation. The rate for this process is as high as 1000 s^{-1}, which is much higher than the rate of interconversion between free and bound forms of the duplex; thus, dissociation of the drug from the complex cannot be the rate-limiting factor for phenol ring flipping. Dynamic fluctuations of the DNA conformation are more likely to provide the rate-limiting step.

Summary of solution studies The data obtained from these NMR studies are consistent with the bound ligand fitting tightly within the minor groove of AT tetramers, with the aromatic rings of the ligand being roughly coplanar. The AT tract provides the key recognition features required for binding, including the narrowness of the minor groove. The importance of van der Waals interactions is evident, given the large number of NOE contacts between the ligand and the walls and floor of the groove. Hydrogen bonding also plays a significant role in stabilizing the interaction, as do electrostatic interactions between the positively charged piperizine ring and the minor groove. Electrostatic interactions are also likely to play a significant role in orienting the ligand within the binding site, as shown in the 2 : 1 complex, in which the piperazine rings point toward the center of the duplex where the positive charge is best stabilized. The information derived from these studies, as well as from NMR studies of the interactions of other minor groove binders with DNA, is useful for the design of ligands with altered specificity or increased binding affinity, with the overall goal being the development of novel drugs.

3.2.4.2 **Immunophilins: Studies of FK506 analog binding to FKBP.** Some of the most detailed investigations of the interaction between ligands and their target proteins have been made for the immunophilin class of proteins. The major FK506 binding protein (FKBP) has a molecular mass of about 11.8 kDa, whereas cyclophilin (Cp) has a mass of about 17 kDa. These proteins are unrelated in amino acid sequence, but both have peptidyl-prolyl cis–trans activities that are inhibited by immunosuppressants that block signal transduction pathways leading to T-lymphocyte activation. FK506 (**10**) and rapamycin (**11**) inhibit the cis–trans isomerase activity of FKBP, whereas cyclosporin A (structure shown in Fig. 3) inhibits that of Cp. NMR has contributed significantly to the understanding of binding interactions to both these proteins.

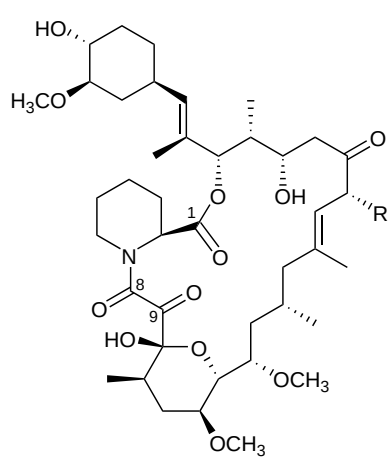

(**10**) FK506 R = CH_2CHCH_2
(**12**) Ascomycin R = CH_2CH_3

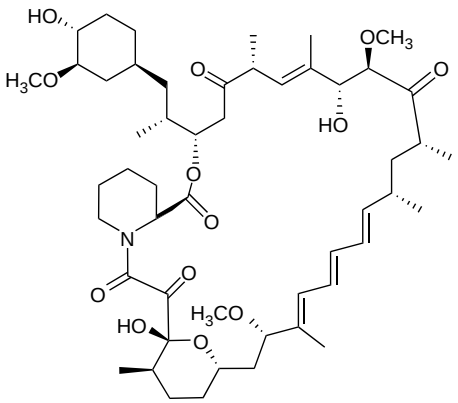

(**11**) Rapamycin

Initial studies on FK506 focused on the structure of the free ligand to aid in the design of further analogs. However, it was established from studies of the cyclosporin A–cyclophilin complex that the conformation of a molecule bound to its target site might be very different from that in the free state. In addition, analog design is assisted by knowing the location of the binding region of the ligand. Studies were therefore undertaken to determine the bound state of the ligand as well as to identify those portions of the drug interacting with the binding protein.

The first investigations involved the analysis of ^{13}C carbonyl chemical shifts of C8 and C9 and the ^{1}H chemical shifts of the piperidine ring of FK506 bound to FKBP. The upfield shifts of the piperidine ring protons, as well as NOEs observed between these protons and aromatic protons of FKBP, suggested that the bound site on FKBP resided in an aromatic-rich domain, and allowed a putative binding site on FKBP to be proposed. It was also evident that the pipecolinyl functionality of FK506 and analogs was involved in the binding face of the ligand.

In another study, a uniformly ^{13}C-labeled ascomycin, (**12**), was prepared, allowing the bound conformation of ascomycin to be determined in the presence of FKBP. The enhanced ^{13}C signals were used to edit the ^{1}H NOESY spectra used for the structural analysis. Not only were the assignments of side-chain methyls made possible by the ^{13}C enrichment but ligand resonances also could be distinguished readily from those of the protein. The conformation of the ligand was determined from NOEs observed in a 3D HMQC-NOESY spectrum. The resulting ascomycin structure (Fig. 25) differed considerably from that of the uncomplexed FK506 obtained by X-ray crystallography, but was similar to that of rapamycin. In particular, the bound ascomycin displayed a trans orientation of the 7,8-amide bond, whereas this bond is cis in free FK506 and trans in rapamycin. The backbone structure of the macrocyclic ring differed from that of uncomplexed FK506, but showed a similarity in the piperidine ring region to that of rapamycin. This study also showed that both the piperidine ring and the pyranoyl moiety of ascomycin are involved in the binding interface in the complex with FKBP. Ligand protons that show NOEs to the protein are in bold in Fig. 25. X-ray studies have since confirmed these results for both the FK506-FKBP and rapamycin–FKBP complexes, showing the trans orientation of the ligand amide bond in the bound conformation, and verifying the involvement of the piperidine and pyranoyl regions of the ligand in the binding interface.

The binding site of the FKBP complex has also been investigated through use of NMR spectroscopy. In early studies, the structure of uncomplexed FKBP was solved using ^{1}H-NMR methods. Although spectral overlap did not allow every

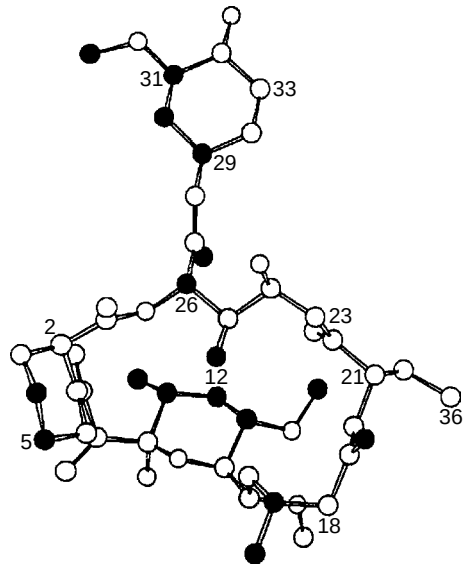

Fig. 25 Three-dimensional structure of ascomycin bound to FKBP. Protons on the ligand that showed NOEs to the protein are denoted by a black shading of the carbons to which they are attached. Although no NOEs were observed from the protons at position 3 to the protein, the upfield shift of their resonances, −1.09 and 0.25 ppm, suggests that they are in close proximity to an aromatic region of FKBP. (Reprinted with permission from American Chemical Society from Petros, A.M., Gampe Jr., R.T., Gemmecker, G., Neri, P., Holzman, T.F., Edalji, R., Hochlowski, J., Jackson, M., McAlpine, J., Luly, J.R. et al. (1991) *J. Med. Chem.*, **34**, 2925.)

structural constraint present to be identified unambiguously, convergent structures defining the global fold of the 107-residue FKBP protein were obtained. Previous biochemical data allowed the extensive aromatic cluster within the core of the structure to be identified as the ligand-binding pocket. The loop regions of the protein between residues 37–43 and 83–90, situated at the open end of the binding pocket, were also of interest. The loops were the least well-defined regions of FKBP and were thought to be flexible, and perhaps involved in the binding interaction. Examination of ^{1}H and ^{15}N chemical-shift changes on addition of ligand supported this notion and suggested that significant structural changes in these loop regions occurred upon ligand binding.

In a later study, a high-resolution structure of the complete ascomycin-FKBP complex was calculated by heteronuclear 3D and 4D NMR. The extra detail afforded by the multidimensional NMR approach allowed the ligand–protein contact area to be located unambiguously and even specific intermolecular hydrogen bonds could be identified. The structure of the complexed FKBP was essentially similar to that of the uncomplexed structure, except that the "ill-defined" loop regions between residues 36–45 and 78–92 were found to adopt well-defined conformations in the complexed proteins, as preempted by previous studies. Although this difference may partially be a result of the differences in resolution achieved in the complexed and uncomplexed FKBP NMR studies, generally it was thought that binding involved some rearrangement of the 36–45 and 78–92 loops. This provides a good example of the dynamic nature of protein binding as revealed by NMR spectroscopy.

The dynamic aspects of the ligand–FKBP complex formation have been pursued through analysis of ^{15}N-NMR relaxation data. In particular, the increased backbone mobility for several residues within the 36–45 and 78–95 loops compared with that of the rest of the protein has been noted. From analysis of the ^{15}N

relaxation rates of FKBP complexed with FK506, it was found that flexibility was restricted along the entire polypeptide chain. This confirmed the proposition that the binding interaction of FKBP with ligand involves stabilization and structuring of the protein loops adjacent to the binding site.

In summary, it was possible not only to define the free and bound conformations of the ligand but also to identify the two binding interfaces involved in the interaction and demonstrate a reduction in protein mobility in a defined region of the protein upon binding. This level of analysis was possible because of the tight binding of the FKBP–ligand complex, its small size, and the availability of labeled species. The information proved to be complementary to X-ray crystallographic studies and will help clarify the role of FKBP complex formation in immunoregulation.

3.2.4.3 **Matrix metalloproteinases.** Matrix metalloproteinases or MMPs, including stromelysin, collagenase, and gelatinase, are involved in tissue remodeling associated with embryonic development, growth, and wound healing. Unregulated or overexpressed MMPs have been implicated in several pathological conditions, including arthritis and cancer, and inhibitors of stromelysin and other MMPs have attracted much interest because of their potential for the treatment of these diseases.

Several NMR structural studies of stromelysin and collagenase complexes have been reported. The secondary structure and global fold have been found to be quite similar for the catalytic domains of both enzymes and their various complexes with ligands. The active site in each enzyme is a cleft spanning the width of the enzyme, with a catalytic zinc atom coordinated by three histidine residues located in the center. Different dynamic properties of active-site residues in stromelysin/ligand complexes and of collagenase with and without bound inhibitor have been reported. It has been proposed that structural and dynamic differences can be exploited in structure-based drug design to achieve broad inhibitor activity against several MMPs or to obtain more selective inhibition.

Of recent interest have been structural data on a novel class of MMP-binding inhibitors, represented by PNU-107859 (**13**) and PNU-142372 (**14**), which contain a thiadiazole moiety that coordinates the catalytic zinc atom through its exocyclic sulfur atom.

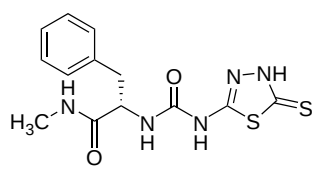

(**13**)

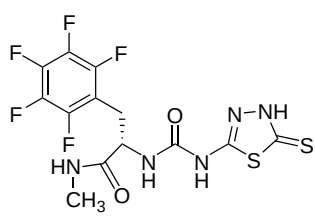

(**14**)

Isotope editing/filtering studies played an important role in defining interactions between the ligands and stromelysin. For example, for the stromelysin/PNU-107859 complex, a 3D ^{12}C-filtered, ^{13}C-edited NOESY spectrum recorded on the [^{12}C,^{14}N]PNU-107859/[^{13}C,^{15}N]-stromelysin complex was used to assign protein/ligand NOEs. Of the 11 observed

NOEs between the ligand and protein aliphatic protons, nine involved the aromatic ring of (**13**) and one involved the terminal methyl group. NOEs were observed between (**13**) and protons of Tyr[155], His[166], Try[168], and Ala[169]. All four of these residues are located in the S_1-S_3 binding sites on one side of the active site. Comparison of 2D $^1H-^{15}N$ HSQC spectra showed that differences between the ^1H and ^{15}N chemical shifts for the stromelysin/**13** and stromelysin/**14** complexes are concentrated in the active site, indicating that no gross conformational differences in protein structure exist. The aromatic rings of (**13**) and (**14**) bind in the same region of the protein.

A region of the 1D ^{19}F spectrum of the stromelysin/**14** complex is shown in Fig. 26. Two separate resonances were observed for the two ortho fluorine atoms of the bound ligand in contrast to the single resonance observed for both ortho protons of stromelysin-bound (**13**), indicating that the ring flip rate (rotation about the $C^{\beta}-C^{\gamma}$ bond) is reduced for stromelysin-bound (**14**) compared to stromelysin-bound (**13**). A ring flip rate of approximately 100 s^{-1} was estimated from the difference in linewidths for the bound ortho and para fluorine atom resonances of (**14**), more than 2 orders of magnitude slower than the ring flip rate for (**13**). The ^{19}F spectrum in Fig. 26 illustrates several general principles that are useful in NMR studies of ligand–macromolecule complexes. First, note that the use of a rare probe nucleus such as ^{19}F produces spectra of elegant simplicity. Because there is no naturally occurring ^{19}F in the

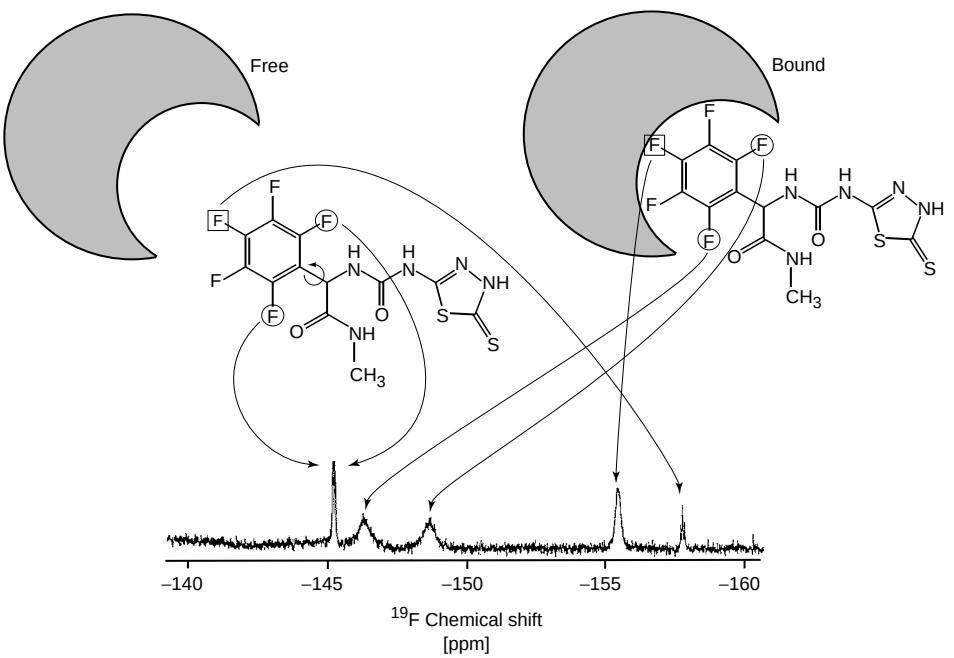

Fig. 26 Region of the 1D ^{19}F spectrum of the stromelysin/PNU-142372 complex. Signals from free (sharp) and bound (broad) PNU-142372 are observed. (Adapted with permission from Elsevier Science from Stockman, B. (1998) *Prog. Nucl. Magn. Reson. Spectrosc.* **33**, 109.)

macromolecule, it generates no interfering signals. Second, the offset in chemical shift between bound and free signals reflects the different environment of the bound and free states. Third, signals from the bound ligand are broader than those from the free ligand because of the higher molecular weight of the complex but are still clearly visible for a complex of this size.

NMR studies have also been reported for ligands bound to collagenase. Interest so far has focused on hydroxamate-containing ligands, where it has been shown that binding causes a decrease in mobility of some but not all active-site residues. Interestingly, some active-site residues adjacent to residues that interact directly with inhibitor were found to have high mobility both in the presence and the absence of inhibitor. This contrasts with what is observed for stromelysin complexed to hydroxamate ligands and a more complete understanding of the dynamics of the respective interactions may provide critical information for drug design.

Hydroxamate-containing ligands have also featured in other NMR studies, this time using TrNOEs to determine their bioactive conformations. TrNOE data were used to determine the conformation of the inhibitors when bound to stromelysin. The NOE-derived structures of the bound inhibitors were used as templates to screen a database of 260 000 compounds. Eighteen of the 23 compounds identified for which stromelysin binding data were available had affinities less than 200 nM, demonstrating the value of deriving a conformationally restricted template for structure-based drug design. This study also demonstrates the close synergy that exists between structure-based design and screening approaches, either *in silico* or experimental.

3.2.4.4 **Dihydrofolate reductase.** Dihydrofolate reductase or DHFR is an important intracellular enzyme that is the target of several clinically used drugs, including methotrexate (**15**), an anticancer compound, and trimethoprim (**16**), an antibacterial. These act by inhibiting the enzyme in malignant cells and parasites, respectively. The small size of DHFR (18–20 kDa) makes it amenable to structural studies and there have been numerous complexes determined using both X-ray and NMR methods. The focus here will be on a recent illustrative example of the structure of a new complex of DHFR with trimetrexate (**17**). Trimetrexate was initially investigated as an antimalarial agent but has subsequently been found to have antineoplastic activity against breast, neck, and head cancers. It has also been used as an antibacterial for the treatment of *Pneumocystis carinii* pneumonia in AIDS patients. As seen from the following structures, trimetrexate combines some of the features of trimethoprim and methotrexate:

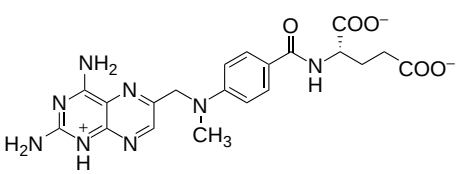

(**15**)

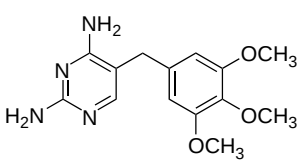

(**16**)

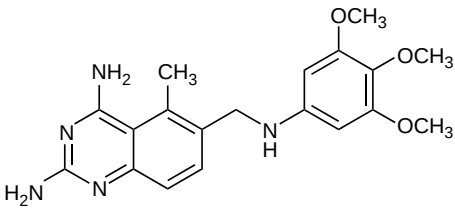

(**17**)

The three-dimensional structure of the complex of DHFR with trimetrexate was determined using about 2000 distance restraints, 300 angle restraints, and 100 hydrogen-bonding restraints. Simulated annealing calculations produced a family of 22 structures consistent with the constraints. Several intermolecular protein–ligand NOEs were obtained by using a novel approach that monitored temperature effects of NOE signals resulting from dynamic processes in the bound ligand. At a low temperature (5 °C), the trimethoxy ring of bound trimetrexate flips sufficiently slowly to give narrow signals in slow exchange, which give good NOE cross peaks. At higher temperature, these broaden and their NOE cross peaks disappear, thus allowing the signals in the lower temperature spectrum to be identified as NOEs involving ligand protons. Figure 27 shows the structure of the complex, including the orientation of the ligand in the binding site.

The binding site for trimetrexate is well defined and was compared with the binding sites in related complexes formed with methotrexate and trimethoprim. No major conformational differences were detected between the different complexes. The 2,4-diaminopyrimidine-containing moieties in the three drugs bind essentially in the same binding pocket and the remaining parts of their molecules adapt their conformations such that they can make effective van der Waals interactions with essentially the same set of hydrophobic amino acids. The side-chain orientations and local conformations are not greatly changed in the different complexes.

The ring flipping of the trimethoxy aromatic ring mentioned earlier was detected by variable-temperature studies of the spectral line shape. The presence of such dynamic processes involving the ligand appear to be not uncommon in macromolecule–ligand complexes and the ability of NMR methods to detect such phenomena represents one distinct advantage of NMR over X-ray methods of structure determination. Relaxation measurements were also used to probe dynamics of the protein and no large-amplitude motions were found, apart from that at the C terminus. The power of NMR methods for studying dynamics of complexes is further illustrated by an earlier study of the complex of DHFR with methotrexate. In this case, a correlated dynamic rotation of a carboxylate group on the ligand and Arg^{57} of the protein was detected, as illustrated in Fig. 28.

3.2.4.5 **HIV protease.** Because of its essential role in the HIV life cycle, HIV protease is a major target for structure-based design of anti-AIDS drugs. There are now more than 100 structures of HIV protease and protease-inhibitor complexes in the HIV-protease structure database, and the availability of this wealth of high-resolution structural information has been the driving force behind numerous structure-based design programs. Most of the high-resolution structural information on HIV protease has been obtained from X-ray crystallography data. Although there are relatively few examples of HIV protease/inhibitor

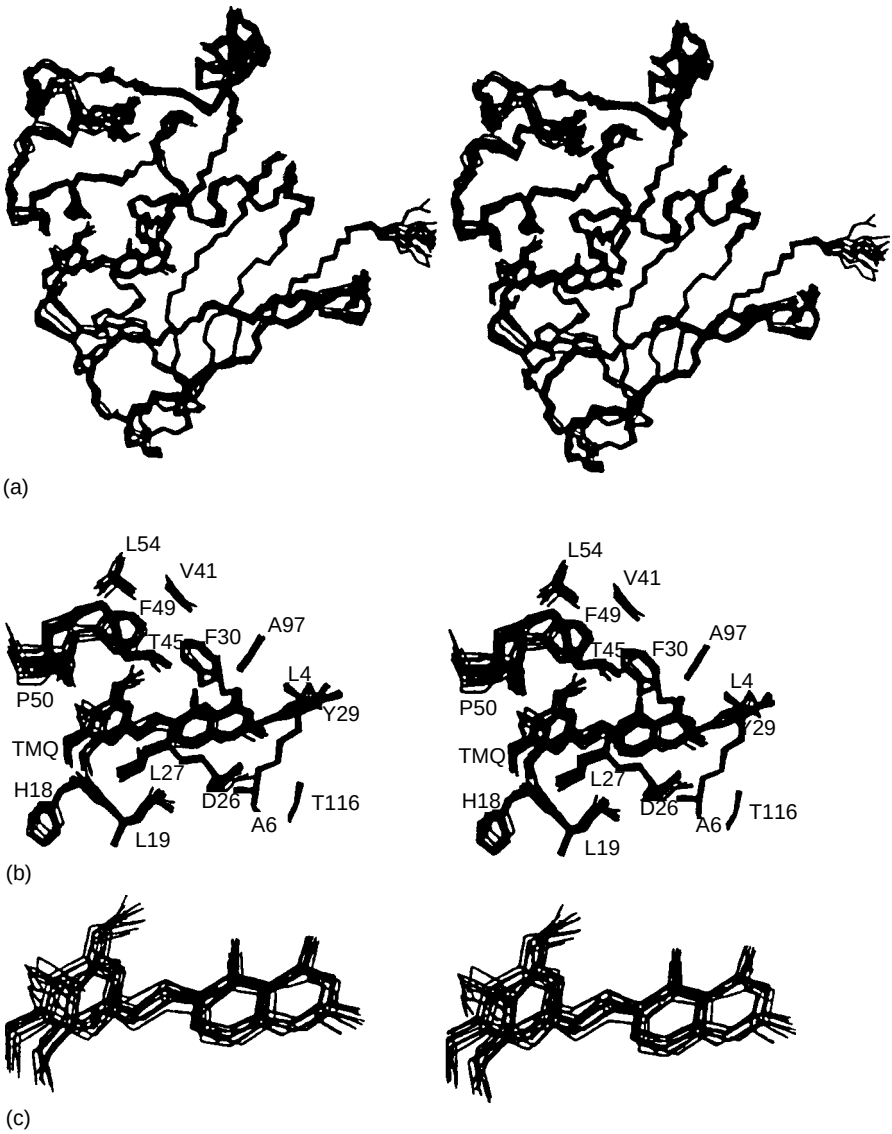

Fig. 27 Stereoview of a superposition over the backbone atoms (N, Cα, and C) of residues 1–162 of the final 22 structures of the DHFR–trimetrexate complex. (a) View of the protein backbone and the trimetrexate heavy atoms. (b) View of trimetrexate in the binding site of enzyme. (c) Conformation of trimetrexate in the binding site of enzyme. The orientation of trimetrexate is identical for (a–c) and only its heavy atoms are shown. (Reprinted with permission from The Protein Society from Polshakov, V.I., Birdsall, B., Frenkiel, T.A., Gargaro, A.R., Feeney, J. (1999) *Protein Sci.* **8**, 467.)

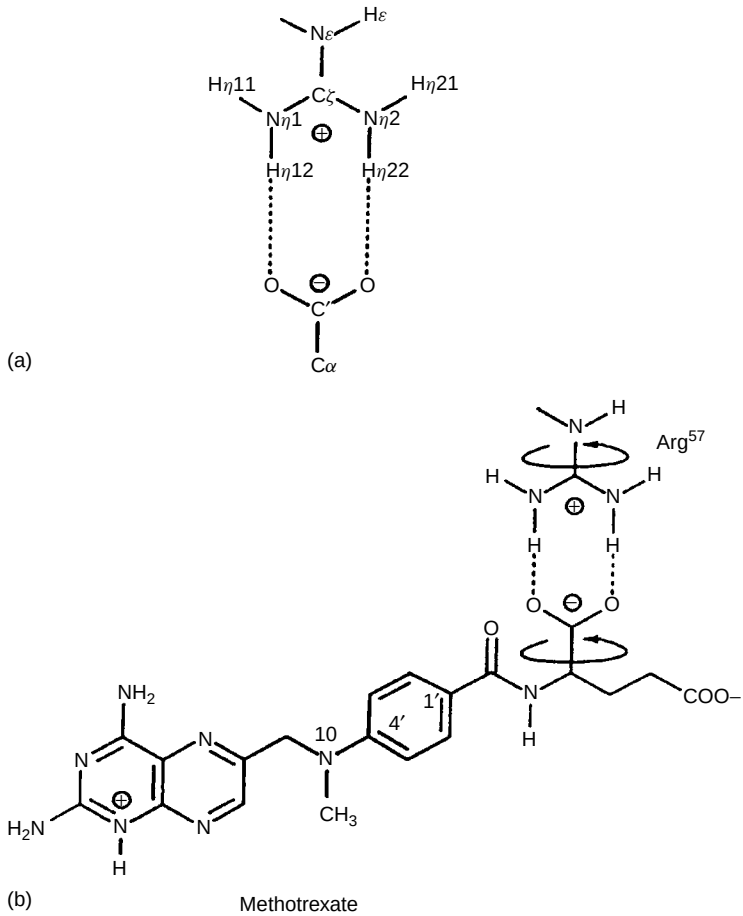

Fig. 28 Correlated motions of a carboxylate group from methotrexate and Arg^{57} of DHFR detected by NMR. (a) Structure of an arginine–carboxylate complex formed with symmetrical end-on interactions and (b) structure of methotrexate showing its interaction with the guanidine group of Arg^{57} of DHFR. (Reprinted with permission from Nieto, P.M., Birdsall, B., Morgan, W.D., Frenkiel, T.A., Gargaro, A.R., Feeney, J. (1997) *FEBS Lett.* **405**, 16.)

complexes that have been determined by use of NMR spectroscopy, the NMR data, taken together with the structural data from X-ray experiments, have contributed to an understanding of protease-inhibitor recognition and dynamics. Indeed, studies of HIV protease/inhibitor complexes are a powerful example of the way in which complementary information obtained from X-ray crystallography and NMR spectroscopy can be used to facilitate structure-based drug design.

HIV protease/inhibitor complexes have a molecular weight of approximately 22 kDa. Although NMR spectroscopy is well suited for the determination of the structure of molecules in this size range, efforts to determine the solution structure

of the complex were hampered by the fact that the protease undergoes rapid autocatalysis in solution. It required the development of potent inhibitors before NMR studies of the complex became feasible. The first solution structure (Fig. 29) of HIV protease bound to the cyclic urea inhibitor DMP-323 (**18**) was reported in 1996.

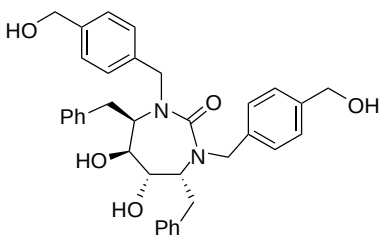

(**18**)

The protease exists as a homodimer. Each 99-residue monomer contains 10 β-strands and the dimer is stabilized by a four-stranded antiparallel β-sheet formed by the N- and C-terminal strands of each monomer. The active site of the enzyme is formed at the interface, where each monomer contributes a catalytic triad (Asp^{25}-Thr^{26}-Gly^{27}) that is responsible for cleavage of the protease substrates. The "flap region" is located above the reactive site and is formed by a hairpin from each monomer of two antiparallel β-strands joined by a β-turn. There is little difference between the solution and crystal structures of protease–inhibitor complexes, except in those regions where the polypeptide chain is disordered. However, experiments in solution have allowed access to parameters that are not directly accessible from crystal data. These parameters, such as the amplitude and frequency of backbone dynamics, the protonation states of the catalytic aspartate residues, and the rate of monomer interchange, are essential

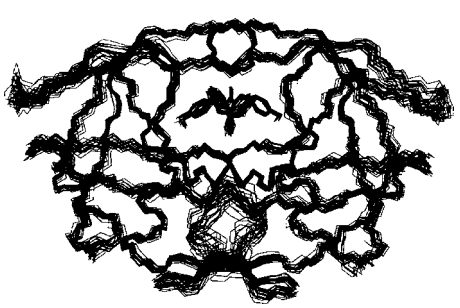

(a)

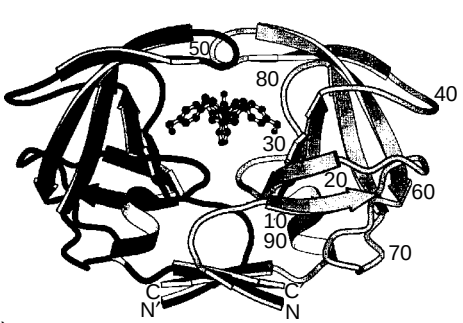

(b)

Fig. 29 (a) View of the superimposed heavy atom (N, Ca, C) of the ensemble of structures of the HIV-1 proteases/DMP323 complex. (b) Ribbon diagram of the minimized average structure of the complex. (Reprinted with permission from The Protein Society from Yamazaki, T., Hinck, A.P., Wang, Y.X., Nicholson, L.K., Torchia, D.A., Wingfield, P., Stahl, S.J., Kaufman, J.D., Chang, C.H., Domaille, P.J., Lam, P.Y. (1996), *Protein Sci.* **5**, 495.)

in understanding the interaction of HIV protease with potent inhibitors.

The cyclic urea inhibitor DMP-323 was designed by analysis of crystal structures of HIV protease/inhibitor complexes. A feature common to many of the complexes of HIV protease is a buried water molecule that bridges the inhibitor and Ile^{50} in the flaps. Interactions with this water molecule are thought to induce the fit of the flaps over the inhibitor. In contrast, mammalian aspartic-protease/inhibitor complexes are unable to accommodate an equivalent water molecule. This observation led to the design of a series of cyclic urea-based inhibitors that are capable of displacing the buried water molecule. As well as improving the specificity of inhibitors to the viral protease, displacement of the water molecule was expected to increase the entropic contribution to inhibitor binding and thus enhance the affinity of complex formation. The cyclic urea inhibitors are highly potent and specific inhibitors of HIV protease and for DMP-323, it has been shown in both the crystal structure and in solution that the urea moiety does indeed replace the buried water molecule.

Although DMP-323 replaces one buried water molecule, several others are observed in the crystal structure of the complex. In favorable cases, NMR can be used to estimate the residence times of hydration water molecules, thus providing information about the timescale of the interaction of buried water with the bulk solvent. This analysis led to the identification of a symmetry-related pair of water molecules that may have a structural role in formation of the complex. Such information may prove useful in the design of future cyclic urea inhibitors. An interesting finding was the fact that each of the hydroxyl protons of DMP-323 is in rapid exchange with solvent. This is a surprising result, given that two of these hydroxyl protons are completely buried and form a network of hydrogen bonds with the catalytic Asp^{25}/Asp^{125} side chains. Furthermore, the dissociation rate of DMP-323 is less than $1\ s^{-1}$ under the conditions of the experiment, which is too small to average the chemical shifts of the hydroxyl protons and the bulk water. The observation is ascribed to local fluctuations in the complex that allow solvent molecules to penetrate into the binding site. This conclusion is supported by the observation that the catalytic protons of the Asp^{25}/Asp^{125} side chains in the protease/DMP-323 complex undergo H-D exchange with solvent, even though they are buried and hydrogen bonded to the inhibitor. These studies highlight that even well-ordered structures such as the protease/DMP-323 complex may be flexible on the millisecond to microsecond timescale.

Interestingly, in the DMP-323 complex, both of the catalytic Asp^{25}/Asp^{125} side chains are protonated over the pH range 2–7. The protonated Asp^{25}/Asp^{125} residues form a network of hydrogen bonds with the hydroxyl groups of DMP-323. In contrast, it has been shown that in the complex with the asymmetric inhibitor KNI-272, the side chain of Asp^{25} is protonated, whereas that of Asp^{125} is not. A suggested explanation for this is that both oxygens of the Asp^{125} side chain are deprotonated to accept two hydrogen bonds, one from a bound water molecule and one from the inhibitor. In contrast, the side chain of Asp^{25} is protonated so that it can donate a hydrogen bond to the inhibitor. Consequently, the protonation state of the enzyme is influenced strongly by interaction with specific inhibitors and this

knowledge is essential for a detailed understanding of the protease/drug interactions.

NMR has also been used to study the relationship between flexibility and enzymatic function for HIV protease. For the protease/DMP-323 complex, ^{15}N spin-relaxation studies determined that residues that are flexible correlate well with residues that are disordered in the NMR structure of the complex. For example, residues in poorly defined loops were found to undergo large-amplitude internal motions on the nanosecond–picosecond timescale. In contrast, two regions of the molecule were found to exhibit motions on the millisecond–microsecond timescale. The first of these is at the N terminus of the protein around Thr^4-Leu^5. This is adjacent to the major site of autolysis of the protease and it has been suggested that the rate of cleavage may regulate HIV protease activity *in vivo*. Consequently, the observed flexibility may be important for regulation of protein function. The second region found to be undergoing millisecond–microsecond motion was the tips of the flaps around Ile^{50}-Gly^{51}. In crystal structures, this region of the protease is well ordered and not involved in crystal contacts, although its conformation varies from structure to structure. This motion is interpreted as a dynamic conformational exchange process, which is fast, relative to the chemical-shift timescale. Thus, when the protease is bound to a symmetric inhibitor in solution, this conformational exchange results in the chemical shifts of the flap residues in the two monomers being identical. In contrast, when the protease is bound to an asymmetric inhibitor, such as KNI-272, crystal structures show that each monomer interacts with the inhibitor in a different way. This is reflected in the fact that the chemical shifts of the monomers are different when asymmetric inhibitors are bound. Analysis of spectra from such an asymmetric complex has revealed that the inhibitor is capable of "flipping" its orientation with respect to the two monomers without dissociating from the complex. These data again highlight the importance of defining both the structural and dynamic aspects of binding to understand the requirements for potent interactions between HIV protease and its inhibitors.

The development of inhibitors of HIV protease represents a major success for structure-based drug design. When HIV was first identified in the early 1980s, there were no known drugs effective for treatment of infection. A combination of X-ray crystallography, NMR spectroscopy, computer modeling, and chemical synthesis has resulted in the development of several effective HIV protease inhibitors. However, in common with other retroviruses, HIV has a high transcription error rate that results in a rapid mutational rate. One of the results of this is the production of a divergent population of viruses in which the sequence of the HIV protease produced may differ substantially. As a consequence, drug-resistant strains of the virus emerge. Clearly, knowledge of the structural principles that govern inhibition of the protease and the mechanism by which the virus develops resistance will continue to be important in the development of effective new drugs.

4 NMR Screening

In the past, NMR was predominantly used in the design stage of drug discovery rather than the screening stages. Recently, new methods that make use of NMR to screen ligands for binding to a protein

target have been developed and are proving to be a powerful tool in the discovery of new drug leads. This section gives an overview of the various experimental methods, summarized in Table 8, which can be used to screen mixtures of ligands for binding to a drug target. There will also be a brief discussion on the practical considerations that need to be made when designing an NMR screening program.

4.1 Methods

4.1.1 Chemical-shift Perturbation

Chemical shift is a function of the chemical (and hence magnetic) environment that individual nuclei experience. Perturbations of chemical shifts can be used to detect binding of a ligand to a protein target. When a ligand binds to a protein, the local chemical environment is changed, and this is reflected by a change in the chemical shifts of nuclei in close proximity to the ligand-binding site. The most common experiment used in this screening methodology is the $^{1}H/^{15}N$ HSQC that generates a discrete signal for each amide group within the protein. A reference $^{1}H/^{15}N$ HSQC spectrum, which is acquired in the absence of potential ligands, is compared to a spectrum recorded in the presence of ligands and any changes in the amide chemical shifts are indicative of a ligand binding to a location close to the corresponding amide groups. The major advantage of this technique is that, if the NMR assignment of the amide resonances is known, then the site of binding for each ligand can be determined. This is a valuable piece of information in the development of more potent second-generation drug leads. Binding affinities can also be determined by measuring the change in chemical shift as a function of ligand concentration.

One technique that utilizes this screening method for drug design is "SAR-by-NMR," developed by Fesik and coworkers. SAR-by-NMR is a fragment-based drug design approach in which a potent drug candidate is derived by chemically linking two or more small low affinity ligands for a target. In theory, the binding energy of the linked compounds will be the sum of the binding energies of the two individual compounds plus contributions to binding energy attributed to linkage. Therefore, it is possible to generate a drug lead with a nanomolar dissociation constant (K_D) from two milli- to micromolar fragments.

The first step in this process (Fig. 30) involves screening a library of ligands (typically with a MW <400) in mixtures of up to 10 for binding to a protein target by comparing the $^{1}H/^{15}N$ HSQC spectrum of an ^{15}N-enriched protein in both the presence and the absence of ligands. Any ligand-induced changes in the chemical shift of the nitrogen and amide proton signals indicate binding of one or more ligands in the mixture to the protein target. The mixture containing the binding ligand(s) is deconvoluted and each individual compound screened to identify the individual ligand(s) responsible for the observed chemical-shift perturbations. Once a binding ligand is identified, analogs can be screened to optimize binding.

A second ligand, which binds at a proximal site, is then identified either from the original screen or by repeating the library screening with the first ligand site bound to the protein. This ligand is then optimized and the structure of the ternary complex determined by use of either NMR or X-ray crystallography. The ternary complex structure provides information on the conformation and orientation of

Tab. 8 A summary of the methods available for NMR screening and their respective characteristics.

Screening methodology	*Signals observed*	*Protein size limit*	*Labelling*	*Binding information obtained*	*K_D limit*	*K_D determined*	*Suitable for HTS (High throughput screening)*	*Mixture deconvolution required?*
Chemical-shift perturbation, e.g. SAR-by-NMR	Protein	<30 kDa	^{15}N protein	Location	10^{-3}–10^{-7} M	Yes	Yes	Yes
STD	Ligand	None	None	Orientation	10^{-3}–10^{-7} M	No	Yes	No
Diffusion-based, e.g. affinity NMR	Ligand	None	^{2}H protein for isotope editing	None	10^{-3}–10^{-7} M	Yes	Yes	No
Relaxation based	Ligand	None	None	None	10^{-3}–10^{-7} M	Yes	Yes	No
TrNOE	Ligand	None	None	Bound conformation	10^{-3}–10^{-7} M	No	Yes	No
NOE pumping	Ligand or protein (reverse)	None	None	Bound conformation	10^{-3}–10^{-7} M	No	Yes	No Yes[a]
Spin labeling	Ligand or protein	None	Spin label for either ligand or protein	Orientation, simultaneous binding	10^{-3}–10^{-6} M	No	Yes[b]	Yes[b]

[a]For reverse NOE pumping.
[b]For primary screening if the protein is spin labeled or for second-site screening if the first-site ligand is spin labeled.

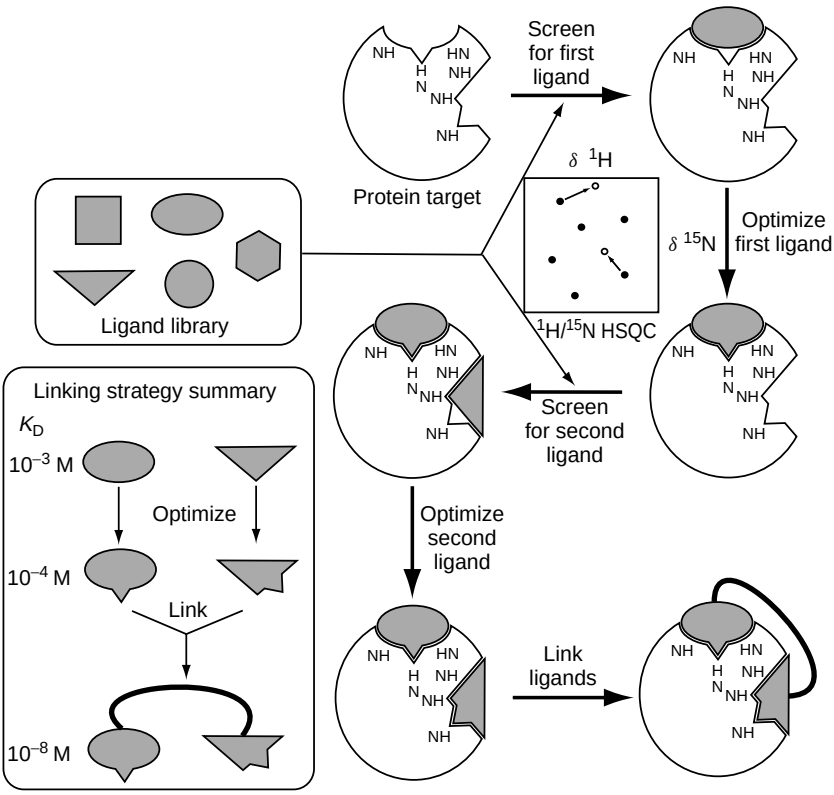

Fig. 30 Summary of the SAR-by-NMR drug discovery methodology. A protein target is screened against a library consisting of small organic molecules by use of the $^{1}H/^{15}N$ HSQC experiment. When two ligands that bind in close proximity are identified, they are linked to form a composite ligand with an increased affinity for the target.

the bound ligands, which facilitates the synthesis of hybrid molecules where the two ligands are joined by a suitable linking moiety.

There are several examples that illustrate the potential of SAR-by-NMR. As noted earlier, FKBP inhibits calcineurin and blocks T-cell activation when complexed to the immunosuppressant FK506. This protein was used as a target for SAR-by-NMR screening and subsequently two ligands, (**19**) and (**20**), were identified with K_D values of 2 μM and 0.1 mM, respectively. A model of the ternary complex between the protein and both

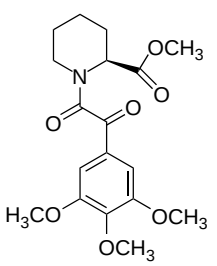

(**19**)

ligands was produced, which indicated that the methyl ester of (**19**) was close to the benzoyl hydroxyl group in (**20**). These two groups were linked with alkyl

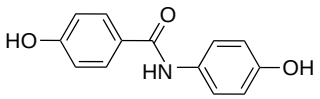

(**20**)

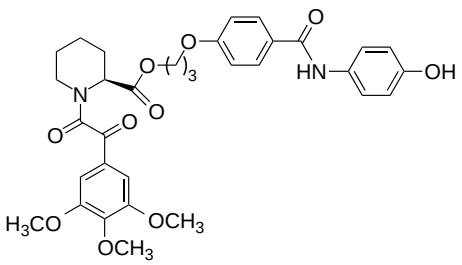

(**21**)

chains of various lengths, with the most active compound (**21**) having a three-carbon linker and a K_D value of 19 nM.

Inhibitors of the MMP stromelysin have also been designed through the use of the SAR-by-NMR screening methodology. As mentioned in Sect. 3.2.4.3, MMPs are involved in matrix degradation and tissue remodeling, with overexpression of these enzymes being associated with arthritis and tumor metastasis. Acetoxyhydroxamate (**22**) was used as one ligand because it was known previously that MMP inhibitors contain a hydroxamate moiety. The K_D value of (**22**) was determined to be 17 mM. To identify a second ligand the protein was screened against a ligand library in the presence of saturating amounts of (**22**). The library was biased for hydrophobic compounds, given that stromelysin demonstrates a substrate preference for a hydrophobic amino acid, and structural studies had identified a hydrophobic binding pocket supporting this observation. From the library screen, a series of biphenyl compounds were identified and analogs of these compounds were synthesized. A biphenyl derivative (**23**) was produced with a K_D value of 0.02 mM. The NMR structure of a ternary complex, consisting of stromelysin (**22**) and the biaryl derivative (**24**) (chosen for its superior aqueous solubility), was determined and indicated that the methyl group of (**22**) was in close proximity to the pyrimidine ring of (**24**). With this information, (**22**) and (**23**) were subsequently linked by different length linkers and the most active compound produced, (**25**), had a K_D value of 15 nM.

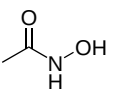

(**22**)

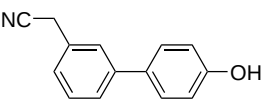

(**23**)

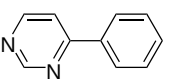

(**24**)

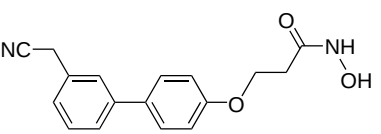

(**25**)

A variation of SAR-by-NMR is to optimize binding or improve the pharmacological properties of known drug leads generated by other methods (e.g. natural products isolation or combinatorial chemistry). A compound can be fragmented into

individual subunits and then alternative fragments can be identified through use of $^{1}H/^{15}N$ HSQC screening. These fragments can then be incorporated into the molecular structure in the hope of improving the binding and/or pharmacological properties of the parent compound (Fig. 31). The alternative fragment must bind in the same location as the corresponding section of the original molecule, making $^{1}H/^{15}N$ HSQC screening method ideal as it provides information on the binding site of ligands.

In a demonstration of this fragmentation method, an antagonist of the interaction between leukocyte function-associated antigen 1 (LFA-1) and intracellular adhesion molecule 1 (ICAM-1) was used as a starting molecule. This interaction plays a role in the inflammatory response and specific T-cell immune responses, and inhibitors have applications in the treatment of inflammation and organ transplant rejection. The *p*-arylthio cinnamide antagonist (**26**) had an IC50 value of 44 nM; however, it was envisaged that the molecule's activity and physical properties could be improved by replacing the isopropyl phenyl group with a more hydrophilic moiety. Screening of a 2500-compound library provided several hits, and analogs of (**26**) were made, which incorporated these ligands in place of the isopropyl phenyl group. Compounds (**27**) and (**28**) had both improved aqueous solubility and pharmacokinetic profiles, with similar or improved activity (IC50 values of 20 and 40 nM, respectively) when compared to that of the parent compound (**26**).

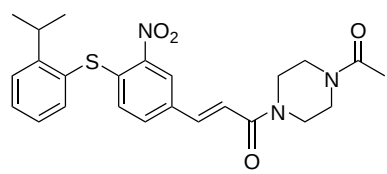

(**26**)

Many compounds bind to human serum albumin (HSA), which significantly reduces their *in vivo* activity and hence their

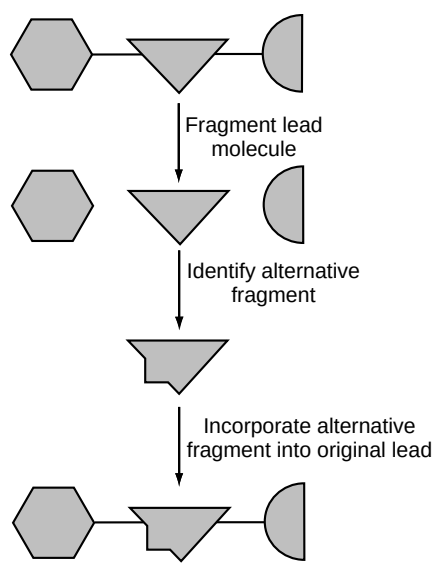

Fig. 31 The fragment optimization approach developed from SAR-by-NMR. A known ligand of a protein is broken into fragments, and small molecules based on the fragments are screened for binding. Any molecules that are found to bind can then be incorporated into the original lead compound with the hope of improving its binding and/or physicochemical properties.

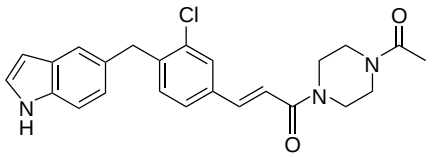

(**27**)

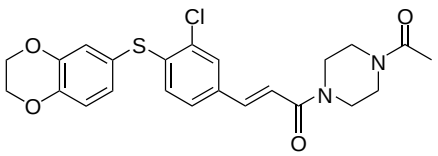

(**28**)

potential as a drug lead. The fragmentation method has recently been used to find analogs of diflusinal (**29**) that have a reduced affinity toward HSA. Diflusinal is a nonsteroidal anti-inflammatory that is more potent, longer acting, and is more tolerated *in vivo* than aspirin. However, 99% of diflusinal is bound to albumin in plasma and, as a result, high doses are required for it to be effective. Structural studies of the diflusinal/HSA-III complex indicated that, by introducing polar functionality to the difluorophenyl moiety, binding affinity to HSA may be reduced without affecting activity. A series of organic compounds analogous to the difluorophenyl fragment were screened using the $^{1}H/^{15}N$ HSQC chemical-shift perturbation method and several alternative fragments were identified. These were incorporated into the diflusinal structure, resulting in a number of analogs (e.g. **30** and **31**) with reduced affinity for HSA but still maintaining some activity. It was predicted that the next generation of compounds would incorporate the fluorine atoms present in (**29**) into leads such as (**30**) and (**31**) because the presence of fluorine increased activity without increasing the affinity for HSA.

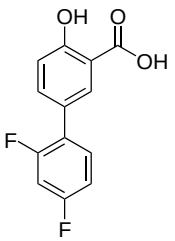

(**29**)

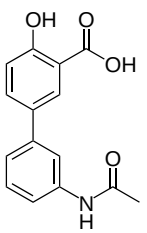

(**30**)

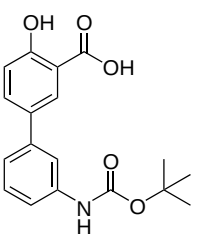

(**31**)

Overall, SAR-by-NMR is already showing great promise as a major new tool in drug design, but it has a few limitations. First, the screening of large ligand libraries and any subsequent deconvolution steps can be expensive, in that a large amount of ^{15}N-labeled protein is needed. However, as already noted, recent advances in cryoprobe technology has reduced the amount of protein required. The use of smaller, more druglike libraries such as that described in the SHAPES ideology could also be used to reduce the amount of protein required for screening. The second limitation is that the NMR assignments for

the protein must be known or determined. This limits the size of the protein target to 30 kDa or less, although this value will presumably increase as TROSY-based NMR technologies for structure determination advance. The method also requires the comparison of $^{1}H/^{15}N$ HSQC of the protein in both the absence and the presence of ligands. Changes in solvent conditions such as pH, polarity, salt concentration, and viscosity may cause shifts in amide resonances, leading to false positives.

Recently, a protocol was described that screens ligands using the $^{1}H/^{15}N$ HSQC experiment but includes a mass spectrometry–prescreening step. Ligand mixtures are added to a protein target and then subjected to size-exclusion chromatography, which separates ligand/protein complexes from free ligands. The identity of the bound ligands can then be determined using MS data and, once identified, screened by $^{1}H/^{15}N$ HSQC to determine the location and specificity of binding. This MS/NMR methodology reduces the amount of ^{15}N-labeled protein required because only a fraction of the library is screened by NMR and there is no deconvolution step.

4.1.2 Magnetization Transfer Experiments

Proteins are composed of a large network of dipole–dipole interactions, resulting in the efficient transfer of magnetization throughout the molecule. The saturation transfer difference (STD) experiment (Fig. 32) uses this phenomenon to detect the binding of ligands to a protein. It relies on the fact that saturation of a single protein resonance results in saturation of all protein resonances and any ligands that bind to the protein, provided they are not affected directly by the selective saturation pulse. The STD experiment is able to detect the binding of ligands with a K_D between 10^{-3} and 10^{-8} M.

An STD experiment consists of irradiating an isolated protein resonance (either at low or high field) with a series of pulses that saturate the entire protein and any binding ligands. This results in a spectrum

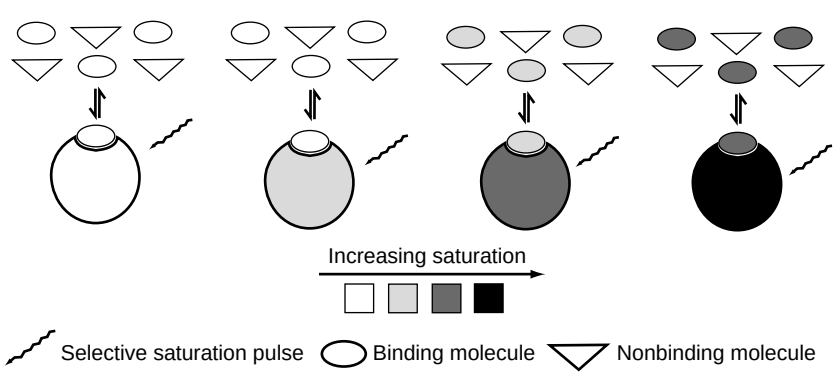

Fig. 32 A schematic representation of the saturation transfer effect. The protein resonances are saturated (indicated by shading) by a selective pulse by spin diffusion. Resonances of nonbinding ligands (triangles) are not affected by this pulse but ligands that are interacting with the protein (ellipses) will also become saturated. These interacting ligands are transferred to solution through chemical exchange where they are detected. (Reprinted with permission from Wiley-VCH from Craik, D.J. (Ed.) (1996) *NMR in Drug Design*, CRC Press, Boca Raton, FL, pp. 1–476.)

containing reduced signal intensities from both the protein and the ligands that bind to it. A second spectrum of the protein and ligand library is then recorded with the saturation pulses off-resonance. Subtraction of these two spectra results in the STD spectrum that shows only those ligands that bind to the protein (residual protein resonances are removed using a T_2 relaxation filter). This subtraction occurs internally through phase cycling after every scan, to reduce artifacts attributed to temperature or magnetic field variations.

An STD can be added to most forms of NMR experiments including COSY, TOCSY, NOESY, and inversely detected ^{13}C or ^{15}N spectra. A high-resolution magic angle spinning (HR-MAS) STD experiment has been developed to study the binding of ligands to a protein immobilized on a solid support. HR-MAS STD NMR provides a way of obtaining ligand-binding information for proteins that are difficult to work with in solution attributed to either poor solubility or conformational changes. In addition to screening ligands for binding to a protein, the binding epitope of a molecule can also be determined by examining the intensities of ligand resonances. The proton signals that have the strongest signals will correspond to those that are part of the ligand's binding epitope. For example, it was shown that when methyl β-D-galactoside (**32**) bound to *Ricinus communis* agglutin I the H2, H3, and H4 were saturated to the highest degree (values on structure indicate relative signal strengths) and hence were in close proximity to the protein protons. This analysis was subsequently extended to the decasaccharide NA_2 (**33**) and demonstrated that the Gal-6′ and GlcNAc-5′ residues bind edge-on to the protein, with the binding contribution of the terminal galactose residue being the greater.

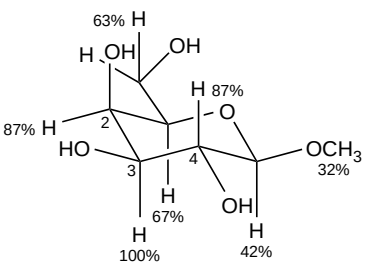

(**32**)

There are a number of advantages in using the STD experiment for the detection

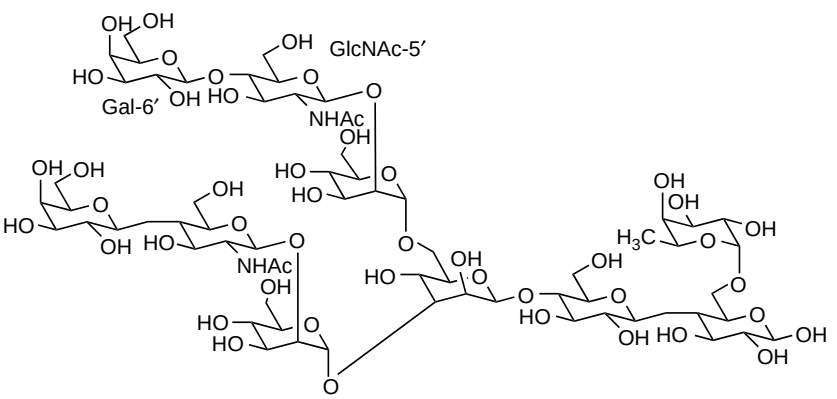

(**33**)

of ligand binding. The saturation transfer effect is an efficient process, which results in high sensitivity, and hence only small quantities of protein are required (nanomolar concentrations of a protein with MW >10 kDa). In addition, protein size is noncritical; in fact, as the protein becomes larger, the saturation transfer effect becomes more efficient. The acquisition time for each experiment is also quite short and, because the experiment is ligand observed, no deconvolution of mixtures is required, making this a good technique for high-throughput screening of large ligand libraries. Unlike the chemical-shift perturbation techniques, STD experiments provide no information on the site of ligand binding.

A second variation of saturation transfer experiment has been devised, which uses the transfer of magnetization from the water. Water is intimately associated with proteins being bound either within or on the surface of the macromolecular structure. Saturation of the water resonance will lead to protein saturation through a variety of mechanisms, including saturation of the αH resonances, saturation of exchanging protein resonances, and NOE interactions between water and the protein. If a compound is bound to the protein, it will also become saturated, and this effect can be used as an indication of ligand binding.

4.1.3 Molecular Diffusion

Molecules can be distinguished on the basis of their diffusion coefficients, which are related to molecular size. Large macromolecules, such as proteins, diffuse more slowly than small molecules and it is this size difference that can be exploited to screen for ligand binding. If a small molecule binds to a protein target, its diffusion coefficient is altered to a value more like that of the protein. Therefore, by utilizing a diffusion filter, resonances generated by small molecules that do not bind to the protein can be removed from the spectrum.

Diffusion editing is achieved with the use of a pair of gradient pulses. If field homogeneity is ignored, then all spins experience an identical magnetic field despite having different positions throughout the sample. The application of a field gradient has the effect of making field strength dependent on position. Under the influence of the gradient pulse, the phase of individual spins become dependent on their position within the sample and hence the spins are spatially "encoded." If diffusion does not occur, this spatial encoding is fully reversible by a second gradient of inverse polarity and no loss of NMR signal will occur. However, the second gradient pulse will be unable to "decode" the spins that have undergone diffusion and the resulting NMR signal will be reduced. Acquiring spectra of a sample with and without the diffusion filter and then subtracting them allows the ligands binding to the protein to be identified. This filtering method can be used for both 1D and 2D experiments and can be "tuned" by altering the strength and duration of the gradients.

Because the ligand signals are being observed in this screening method, no convolution of the ligand mixture is required, given that any signals can be assigned directly to individual compounds within the mixture. However, signals from the protein are always present, which can pose a problem in interpreting spectra. An isotope-edited version of the diffusion experiment has been designed to avoid this problem, although labeled protein is required. Generally, there is no requirement for labeling of the protein

target or for the protein resonances to be assigned and thus, in theory, there is no size limit on the proteins that can be screened by use of this method, although no information is obtained on the location of ligand binding. However, if the protein is large, then the transverse relaxation time may be too short to observe the bound ligands in the diffusion-edited spectrum. Only one sample, containing protein and ligands, is used to obtain both reference and screening data and therefore differences between the sample and reference spectra caused by addition of the ligands (pH, salt concentration, etc.) are avoided.

Diffusion-filtered NMR screening requires that there is a significant difference in observed translational diffusion between the free and bound states. The ligands are in fast exchange on the diffusion timescale and as a consequence the observed diffusion coefficient for binding ligands is an average between the free and bound diffusion values. Free ligands diffuse at a much faster rate than those in the bound state, and thus just a small amount of free ligand has a considerable effect on the observed average diffusion coefficient. This effect may be significant enough to reduce the difference between binding and nonbinding ligands, making it more difficult to interpret results.

A methodology based on diffusion filtering named "affinity NMR," has also been used to screen for binding. Diffusion-edited NMR experiments were able to identify two known binding tetrapeptide ligands of vancomycin from a mixture of 10 peptides. Similarly, the application of diffusion-editing experiments has been demonstrated by differentiating ligands of stromelysin from a mixture containing nonbinding compounds.

4.1.4 Relaxation

Like diffusion, the transverse relaxation time (T_2) of molecules is also dependent on molecular size. Large molecules, such as proteins, have a short T_2 and hence exhibit broad NMR signals, whereas small molecules have a longer T_2 and hence narrower linewidths. Therefore, if a small molecule ligand binds to a protein, its T_2 value will decrease and a line-broadening effect of bound ligand signals can be observed. Alternatively, a relaxation filter can be used to remove signals from molecules with a short T_2 value. Subtraction from a reference spectrum will result in a spectrum containing only those ligands that bind to the protein. The ability to identify binding ligands using relaxation filters has been demonstrated using FKBP. A mixture of nine compounds consisting of one known ligand of FKBP, 2-phenylimidazole (**34**), and eight nonbinding compounds (e.g. **35**–**37**) were screened and only signals from (**34**) were observed.

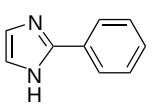

(**34**)

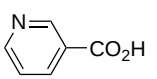

(**35**)

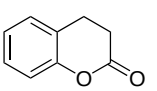

(**36**)

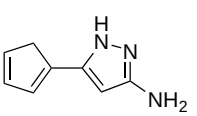

(**37**)

4.1.5 NOE

NOE experiments can also be used to identify ligands that bind to protein targets. Small molecules have a fast tumbling rate and, as a consequence, generally exhibit small positive NOEs. In contrast, large molecules such as proteins generate strong negative NOEs because of their slow tumbling time. When a small molecule binds to a protein, its tumbling rate is slowed to that of the protein and it exhibits strong negative NOEs. On dissociation, these are transiently retained and are known as *transfer NOEs* or *TrNOEs* (Fig. 33). TrNOEs and those arising directly from the free ligand can be distinguished by the rate of signal build up. TrNOEs accumulate significantly faster and therefore can be selected by using shorter mixing times in the NOE experiment.

In practice, a 2D NOESY spectrum of the mixture of potential ligands in the absence of protein is recorded and all molecules exhibit small positive NOEs. The experiment is then repeated in the presence of protein and molecules that bind display negative TrNOEs. Subtraction of the two spectra provides signals arising from only those compounds that bind. These TrNOEs can be interpreted to provide information regarding the bound conformation of the active ligands. However, when analyzing the conformational data, care must be taken to ensure that the ligands are in fast exchange and that the observed TrNOEs are not affected by contributions from spin diffusion. Relative binding affinities between ligands can also be determined by comparison of TrNOE signal strength but, again, the fast-exchange regime and spin-diffusion effects need to be taken into account. If all ligands are in fast exchange, the stronger binding ligands occupy more binding sites and thus give larger TrNOE intensities. Because of the need for an averaging effect, brought about by fast chemical exchange, TrNOE experiments are limited to those ligands with a K_D value from 10^{-3} to 10^{-7} M. The spectral properties of excess ligand in solution are evoked by small fractions of bound molecules, greatly enhancing sensitivity.

TrNOE experiments have been used to identify a bioactive disaccharide from a library of 15 mono- and disaccharides that bound to *Aleuria aurantia* agglutinin. Another study has described the identification of a silalyl Lewis mimetic (**38**) that binds to E-selectin from a library of 10 compounds. As well as being used to detect binding, TrNOEs may also be used to determine bound ligand conformations, as described earlier in this chapter.

A second technique that uses NOEs to detect binding is NOE pumping. This

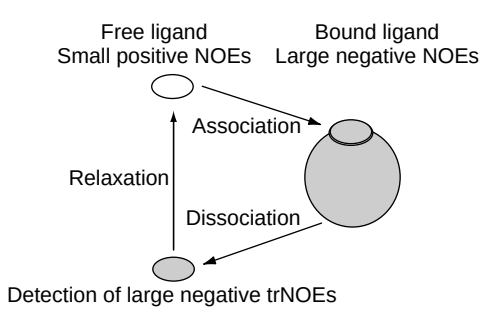

Fig. 33 A schematic representation of the TrNOE experiment used to detect ligand binding (as described by Mayer, M., Meyer, B. (2000) *J. Med. Chem.* **43**, 2093). The free ligand (white ellipse) exhibits only small positive NOEs, although binding to the large protein target results in the generation of large negative TrNOEs. The appearance of these large negative TrNOE signals can be used to identify ligands within a mixture that are binding to the protein and also provide some information on the bound conformation of the ligand.

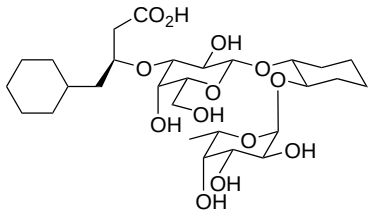

(**38**)

method was designed to alleviate some of the problems associated with the diffusion-edited screening methods. Signals from ligand molecules are removed using a diffusion filter and then transfer of signal from the protein to bound ligands by NOE occurs. The inverse of this is possible (known as *reverse NOE pumping*), which uses a relaxation filter to attenuate the protein resonances, after which the signal is transferred to the protein by NOE. Ligands may lose signal either by relaxation (for a free ligand) or through relaxation and NOE transfer (for a bound ligand). Therefore, by subtracting spectra (which is done internally to reduce subtraction artifacts) from experiments with and without NOE pumping to the protein, the binding ligands can be detected.

4.1.6 Spin Labels

Spin–spin relaxation rates are proportional to the product of the squares of the gyromagnetic ratios of the involved spins. The gyromagnetic ratio of an unpaired electron is significantly larger than that of a proton and therefore any spins influenced by this electron will have substantially shortened relaxation times. The resonances of protons that are within 15–20 Å from the unpaired electron will experience this effect and be significantly broadened. The introduction of a short spin-lock period will significantly reduce the intensity or quench these signals.

The spin-label method can be used as either a primary screening method or to identify a second ligand-binding site. The primary screening method requires residues around the binding pocket of the target to be spin labeled. Residues suitable for this labeling include lysine, cysteine, histidine, glutamate, aspartate, tyrosine, and methionine. Any ligands that bind to the protein in close proximity to the spin-labeled residues will be able to be identified. To screen for second-site ligand binding, the known first-site binding ligand is spin labeled. A reduced signal will be observed for any ligands that bind simultaneously and in close proximity to the first ligand-binding site. In addition, the degree of reduction in signal intensity gives an indication of the orientation of the second ligand in relation to the first, given that the effect of the spin label is inversely proportional to the distance separating the electron and proton. This information is valuable in the design of linkers to join the two ligands.

There are several advantages to using the spin-label screening method. Currently, it is the only method that can detect ligands that bind to the protein simultaneously, unlike other methods that can produce false positives if the first ligand-binding site is not fully saturated. The concentration of protein required for screening is relatively small (~10 μM) because of the substantial enhancement of the relaxation rate by the spin label. The protein can also be unlabeled and partially purified and there is no molecular weight limit. The spin labels also quench protein signals, making interpretation of spectra easier. The experiment is easy to set up and analyze, making

it amenable to automation. It is also insensitive to small changes in solvent conditions, which can generate false positives in other methods. The information obtained on the orientation of ligands is also valuable and makes it an alternative to the chemical-shift perturbation methods when the proteins are large and NMR assignments have not been made.

A disadvantage of the method is the requirement for spin-labeled proteins and ligands. In addition, any ligands with slow dissociation rates will show no averaging of relaxation rates and therefore tightly binding compounds ($K_D < 10^{-6}$ M) will produce false negatives. Protein spin labeling must occur adjacent but not within the binding site to minimize alteration of its binding properties.

The antiapoptopic protein Bcl-xL is responsible for the reduced susceptibility of cancer cells to undergo apoptosis and is therefore a target for the development of new anticancer agents. The structure of a previously identified ligand for Bcl-xL (**39**) was modified to incorporate a TEMPO spin label (**40**). By use of spin labeled (**40**), an eight-compound library was screened for simultaneous binding to Bcl-xL. From this library, an aromatic ketoxime (**41**) was identified as binding simultaneously with and in the vicinity of (**40**). Analysis of relaxation enhancements revealed that the protons around the indole ring were closest to the spin label.

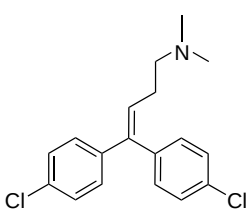

(**39**)

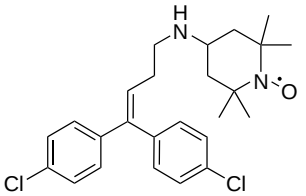

(**40**)

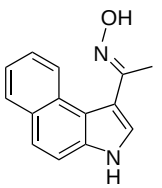

(**41**)

4.2 Practical Considerations

4.2.1 Screening Approach

The first step in using NMR screening is to select a suitable screening method for the target protein being used. Table 8 lists the characteristics of each NMR screening method. The choice of experiment will be determined by the characteristics of the protein target and the information that is desired from the screen. For example, the SAR-by-NMR method is suitable only for small, easily expressed proteins because the NMR assignments for the target need to be known so the location of binding can be determined, and a large amount of ^{15}N-enriched protein is required. If a simple "yes/no" answer on ligand binding is wanted, then the shorter, less resource intensive ligand-observed experiments (e.g. STD, diffusion-edited, or TrNOE) may suffice.

It is also important to determine the correct NMR solvent conditions for the screening procedure. These should facilitate good solubility, with little precipitation or aggregation, and acquisition of good

quality data; maintain protein structure and activity; and provide a sufficient buffering effect to allow for ligands to be added. Two methods that permit the screening of a range of solvent conditions to determine without the need for a large amount of protein are the microdialysis button test and the microdrop-screening method.

4.2.2 Library Design

Effective design and management of the ligand library to be used for screening is essential if successful results are to be obtained.

4.2.2.1 **Ligand properties.** Diversity of ligands is an important factor to consider in the design of a library for NMR screening and there are a number of factors to take into consideration. Although it would seem logical to maximize diversity, this may not always be the most efficient approach. If the system being studied exhibits neighborhood behavior, then maximizing diversity is a good option. Neighborhoods are regions of multidimensional molecular space defined by a set of molecular descriptors. By the choice of a molecule that is in the center of a neighborhood, it is possible, in theory, to represent all molecules within that molecular space. By spreading out the molecules that are selected for the library so that each neighborhood does not overlap, diversity is maximized.

However, if neighborhoods are only small, then compound libraries must be very large so that the neighborhoods overlap and hence all molecular space is covered. In addition, some systems do not exhibit neighborhood behavior and relatively small changes to the structure of a compound may lead to large changes in its binding affinity for the target. Maximizing diversity may also be inefficient because many molecules do not possess physicochemical characteristics that are suitable as the basis for a drug. In practice, the more that is known about the drug target, the less diverse and more focused the library can be. However, if the library is too focused, then some outlying "new" ligand type for the target being screened may be missed.

One strategy for library design is to select compounds that have druglike characteristics. A simple set of rules, determined by Lipinski and coworkers, for determining whether a compound is druglike is known as *the rule of 5*. According to this set of criteria, the majority of orally available drugs have five or fewer hydrogen bond donors, 10 or fewer hydrogen bond acceptors, a log P of less than 5, and a molecular weight less than 500. Additional factors that can be taken into consideration include the number of heavy atoms, rotatable bonds, and ring systems. Another study has revealed that there are a number of frameworks and side chains that commonly occur in many drugs. Drug molecules, from the comprehensive medicinal database, were broken down into systems consisting of frameworks (Fig. 34) and side chains. Analysis of these two structural features revealed that approximately 50% of all known drugs could be represented by only 32 different frameworks. When atom type and bond order were included in the analysis, 41 frameworks were found to describe 24% of all drugs. A similar analysis of side-chain frequency indicated that approximately 70% of all side chains present in the compound database analyzed were from the top 20 occurring side chains.

The presence of these common frameworks and side chains has been exploited in the SHAPES methodology for NMR

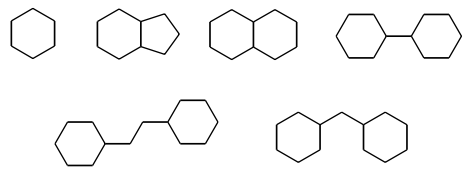

Fig. 34 Examples of molecular frameworks from the SHAPES library.

screening. This strategy employs a small focused library based on these common frameworks and side chains to screen against protein targets through the use of relaxation and NOE experiments. The advantages of this approach are that the library is small and hence only relatively small amounts of protein are required and any hits from the library will possess druglike characteristics. However, a disadvantage of the method is that it is unlikely to yield new drug types, given that the library is based on known drug frameworks.

Diversity of molecular type is not the only factor that must be taken into account when designing a library to be used in an NMR screening program. Because the screening occurs in an aqueous solution, the organic compounds chosen for the library must demonstrate reasonable solubility in the aqueous conditions used. In general, compounds are dissolved in dimethyl sulfoxide (DMSO) and then added to the protein solution at the appropriate concentration. Currently, there are no good methods for determining the solubility of a wide range of compounds before screening commences. A simple method is to dilute the DMSO solution in buffer and observe whether any precipitation or aggregation occurs. However, this method will not be suitable for compounds that precipitate or aggregate over several hours, and solutions that appear clear may still contain high MW aggregates, which will cause false positives in experiments such as the diffusion, relaxation, and TrNOE methods.

It is also preferable to choose ligands that are synthetically accessible and/or possess suitable moieties to build upon or link to other fragments. This is especially important in the SAR-by-NMR screening methodology because this relies on the ability to link individual fragments to form a more potent drug lead. If the ligands to be linked are not synthetically accessible or do not possess suitable linking functional groups, then this process is severely hindered.

4.2.2.2 **Mixture design.** The optimal number of compounds per mixture is dependent on the screening method. For ligand-observed experiments, the limiting factor for the number of compounds in a mixture is spectral overlap. Ligands need to be chosen so that spectral overlap is minimized, making interpretation of the resulting data far simpler. In theory, protein-observed experiments could have a large number of compounds per mixture, which would both minimize screening time and the requirement for large amounts of protein. However, because the experiments are protein observed, deconvolution of the mixtures and rescreening of each individual compound are required to identify any hits. Therefore, the number of compounds per mixture is dependent on the hit rate in the screening procedure, given that the greater the hit rate, the more deconvolution steps that are required and consequently more protein and spectrometer time are needed. The number of experiments required is at a minimum

when the number of compounds is equal to $1/(\text{hit rate})^{1/2}$; thus, with a hit rate of 10%, the optimal number of compounds per mixture is three. In addition to these factors, if the hit rate is high, then it is likely that several compounds within a mixture containing a large number of compounds may compete for the same binding pocket, which may lead to false negatives.

In mixtures of organic compounds, the possibility of interactions between compounds, such as reactions or ion pairing, is also possible and should be taken into consideration, especially when one uses large numbers of compounds per mixture. It has been demonstrated that in random mixtures of 10 compounds in DMSO, the probability of a reaction occurring between two of the mixture's components is 26%. This value can be reduced by careful selection of mixture components (e.g. separating acids from bases) to approximately 9%.

4.2.3 Hardware and Automation

Automation is a requirement if libraries containing a large number of compounds are to be screened. Technology has been developed that allows the automation of almost all steps of the NMR screening process from sample preparation through to data analysis. The general setup for NMR screening consists of a robot for just-in-time preparation of each sample, which is then transferred to the magnet either through a flow system or as discrete samples on a rail system. There are several disadvantages in using a flow system, including the possibility of contamination of samples by previously screened compounds, the capillary line can be blocked if the protein or ligands precipitate or form aggregates, recovery of the sample is more laborious because it has been diluted, and cryoprobe technology (discussed later in this section) is not yet available in the flow system. Many of these problems can be overcome by using the discrete samples with the rail system.

Data acquisition is easily automated and there are several software packages that will automate data processing for 2D spectra. The processing of 1D spectra automatically is reported to be less reliable because of the large solvent signal, and usually requires manual adjustment of phasing. One of the most laborious tasks in NMR screening is the analysis and comparison of the resulting spectra. For 1D ligand-observed experiments, difference methods (e.g. STD) provide the most reliable method for interpretation of results, in that the presence of signals in the spectra will correspond to the ligands that are binding. In 2D protein-observed experiments (e.g. $^{1}H/^{15}N$ HSQC), a more statistically rigorous analysis of changes in chemical shift is required and a discussion of this is beyond the scope of this chapter.

Currently, approximately 50 to 100 samples can be screened per day and if mixtures contain 10 compounds each, this provides a substantial throughput. This throughput rate will increase as technology improves, as has been demonstrated by the use of cryoprobes. Cryogenic NMR probes, where the preamplifier and radiofrequency coils are cooled to low temperatures, can significantly increase the signal-to-noise ratio of an NMR spectrum. By use of these probes NMR data can be obtained in much faster times and by use of lower protein concentrations, which subsequently increases throughput, the total amount of protein needed to screen a library is reduced. For example, Hajduk and coworkers demonstrated the substantial improvements made through the use of a cryoprobe instead of a

conventional probe in $^{1}H/^{15}N$ chemical-shift perturbation screening. Stromelysin (50 μM) was screened against mixtures of 100 compounds (50 μM each), facilitating the screening of more than 10 000 compounds in one day. The use of lower concentrations of both protein and ligands increases the stringency levels for the binding strength of ligands. At a protein/ligand concentration of 0.5 mM, ligands with dissociation constants in the millimolar range can be detected, although at a protein/ligand concentration of 50 μM, this dissociation constant limit is reduced to approximately 0.15 mM. Although using higher protein/ligand concentrations can be advantageous when screening libraries containing small low affinity ligands, a higher stringency is required when screening large libraries, to reduce the number of hits obtained to a manageable number.

5 Conclusions

In this chapter, we have given an overview of the two major approaches used in NMR and drug discovery, structure-based design and NMR-based screening. Both areas are flourishing and, together with more traditional uses of NMR, they demonstrate the versatility of NMR as a tool in medicinal chemistry. The power of NMR has been dramatically enhanced over the last decade by developments in both instruments and methodology. On the instrumental side, increases in magnetic field strengths and the development of cryoprobes have greatly increased sensitivity. Linkages of NMR to LC and MS have increased versatility. On the methods front, there have been a range of new approaches discovered, which will enhance the study of larger molecular complexes. Advances in protein expression and labeling have played a major role in stimulating the development of new NMR pulse sequences to extract information from such complexes.

Acknowledgment

Work in our laboratory on NMR in drug design and development is supported by the Australian Research Council. David J. Craik is an ARC Professorial fellow.

The present text is a modified and updated version of the chapter "NMR and Drug Discovery" published in "Burger's Medicinal Chemistry and Drug Discovery" (ISBN 0-471-37032-0) edited by D.J. Abraham in 2003.

See also Medicinal Chemistry; Oligonucleotides; Pharmacokinetics and Pharmacodynamics of Biotech Drugs; Protein and Nucleic Acid Enzymes; Protein NMR Spectroscopy; Receptor Targets in Drug Discovery; Synthetic Peptides: Chemistry, Biology, and Drug Design; Targeting and Intracellular Delivery of Drugs.

Bibliography

Books and Reviews

Craik, D.J. (Ed.) (1996) *NMR in Drug Design*, CRC Press, Boca Raton, FL.

Hajduk, P.J., Meadows, R.P., Fesik, S.W. (1999) NMR-based screening in drug discovery, *Q. Rev. Biophys.* **32**, 211–240.

Holzgrabe, U., Wawer, I., Diehl, B. (1999) *NMR Spectroscopy in Drug Development and Analysis*, Wiley-VCH, Weinheim.

Keifer, P.A. (1999) NMR tools for biotechnology, *Curr. Opin. Biotechnol.* **10**, 34–41.

Lepre, C.A., Moore, J.M., Peng, J.W. (2004) Theory and applications of NMR-based screening in pharmaceutical research, *Chem. Rev.* **104**, 3641–3675.

Roberts, G.C.K. (Ed.) (1993) *NMR of Macromolecules: A Practical Approach*, Oxford University Press, Oxford.

Roberts, G.C. (1999) NMR spectroscopy in structure-based drug design, *Curr. Opin. Biotechnol.* **10**, 42–47.

Stockman, B. (1998) NMR Spectroscopy as a tool for structure-based drug design, *Prog. Nucl. Magn. Reson. Spectrosc.* **33**, 109–151.

Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

Primary Literature

Abseher, R., Horstink, L., Hilbers, C.W., Nilges, M. (1998) Essential spaces defined by NMR structure ensembles and molecular dynamics simulation show significant overlap, *Proteins: Struct. Func. Genet.* **31**, 370–382.

Andersen, N.H., Chen, C.P., Marschner, T.M., Krystek, S.R. Jr., Bassolino, D.A. (1992) Conformational isomerism of endothelin in acidic aqueous media: a quantitative NOESY analysis, *Biochemistry* **31**, 1280–1295.

Aumelas, A., Chiche, L., Kubo, S., Chino, N., Tamaoki, H., Kobayashi, Y. (1995) [Lys(-2)-Arg(-1)]endothelin-1 solution structure by two-dimensional 1H-NMR: possible involvement of electrostatic interactions in native disulfide bridge formation and in biological activity decrease, *Biochemistry* **34**, 4546–4561.

Bagby, S., Tong, K.I., Liu, D., Alattia, J.R., Ikura, M. (1997) The button test: a small scale method using microdialysis cells for assessing protein solubility at concentrations suitable for NMR, *J. Biomol. NMR* **10**, 279–282.

Balbach, J., Seip, S., Kessler, H., Scharf, M., Kashani-Poor, N., Engels, J.W. (1998) Structure and dynamic properties of the single disulfide-deficient alpha-amylase inhibitor [C45A/C73A]tendamistat: an NMR study, *Proteins* **33**, 285–294.

Bemis, G., Murcko, M. (1996) The properties of known drugs. 1. Molecular frameworks, *J. Med. Chem.* **39**, 2887–2893.

Bemis, G., Murcko, M. (1999) Properties of known drugs. 2. Side chains, *J. Med. Chem.* **42**, 5095–5099.

Bierer, B.E., Somers, P.K., Wandless, T.J., Burakoff, S.J., Schreiber, S.L. (1990) Probing immunosuppressant action with a non-natural immunophilin ligand, *Science* **250**, 556–559.

Bleicher, K., Lin, M., Shapiro, M.J., Wareing, J.R. (1998) Diffusion edited NMR: screening compound mixtures by affinity NMR to detect binding ligands to Vancomycin, *J. Org. Chem.* **63**, 8486–8490.

Bourguet, W., Ruff, M., Chambon, P., Gronemeyer, H., Moras, D. (1995) Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha, *Nature* **375**, 377–382.

Campbell, A.P., Sykes, B.D. (1993) The two-dimensional transferred nuclear Overhauser effect: theory and practice, *Annu. Rev. Biophys. Biomol. Struct.* **22**, 99–122.

Carrondo, M.A., Coll, M., Aymami, J., Wang, A.H., van der Marel, G.A., van Boom, J.H., Rich, A. (1989) Binding of a Hoechst dye to d(CGCGATATCGCG) and its influence on the conformation of the DNA fragment, *Biochemistry* **28**, 7849–7859.

Casarotto, M.G., Craik, D.J. (2001) Ring flexibility within tricyclic antidepressant drugs, *J. Pharm. Sci.* **90**, 713–721.

Chen, A., Shapiro, M.J. (1998) NOE Pumping: a novel NMR technique for identification of compounds with binding affinity to macromolecules, *J. Am. Chem. Soc.* **120**, 10258–10259.

Chen, A., Shapiro, M.J. (1999) Affinity NMR, *Anal. Chem.* **71**, 669A–675A.

Cheng, J.W., Lepre, C.A., Moore, J.M. (1994) 15N NMR relaxation studies of the FK506 binding protein: dynamic effects of ligand binding and implications for calcineurin recognition, *Biochemistry* **33**, 4093–4100.

Cheng, J.W., Lepre, C.A., Chambers, S.P., Fulghum, J.R., Thomson, J.A., Moore, J.M. (1993) 15N NMR relaxation studies of the FK506 binding protein: backbone dynamics of the uncomplexed receptor, *Biochemistry* **32**, 9000–9010.

Coles, M., Sowemimo, V., Scanlon, D., Munro, S.L., Craik, D.J. (1993) A conformational study by 1H NMR of a cyclic pentapeptide antagonist of endothelin, *J. Med. Chem.* **36**, 2658–2665.

Dalvit, C., Pevarello, P., Tato, M., Veronesi, M., Vulpetti, A., Sundstrom, M. (2000)

Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water, *J. Biomol. NMR* **18**, 65–68.

Dasgupta, F., Mukherjee, A.K., Gangadhar, N. (2002) Endothelin receptor antagonists – an overview, *Curr. Med. Chem.* **9**, 549–575.

Detlefsen, D.J., Hill, S.E., Day, S.H., Lee, M.S. (1999) Molecular flexibility profiling using NMR spectroscopy, *Curr. Med. Chem.* **6**, 353–358.

Dugad, L., Gerig, J.T. (1988) NMR studies of carbonic anhydrase-4-fluorobenzenesulfonamide complexes, *Biochemistry* **27**, 4310–4316.

Duggan, B.M., Craik, D.J. (1996) 1H and 13C NMR relaxation studies of molecular dynamics of the thyroid hormones thyroxine, 3,5,3′-triiodothyronine, and 3,5-diiodothyronine, *J. Med. Chem.* **39**, 4007–4016.

Duggan, B.M., Craik, D.J. (1997) Conformational dynamics of thyroid hormones by variable temperature nuclear magnetic resonance: the role of side chain rotations and cisoid/transoid interconversions, *J. Med. Chem.* **40**, 2259–2265.

Embrey, K.J., Searle, M.S., Craik, D.J. (1993) Interaction of Hoechst 33258 with the minor groove of the A + T-rich DNA duplex d(GGTAATTACC)2 studied in solution by NMR spectroscopy, *Eur. J. Biochem.* **211**, 437–447.

Fede, A., Billeter, M., Leupin, W., Wuthrich, K. (1993) Determination of the NMR solution structure of the Hoechst 33258-d(GTGGAATTCCAC)2 complex and comparison with the X-ray crystal structure, *Structure* **1**, 177–186.

Feeney, J., Batchelor, J.G., Albrand, J.P., Roberts, G.C.K. (1979) The effects of intermediate exchange processes on the estimation of equilibrium constants by NMR, *J. Magn. Reson.* **33**, 519–529.

Fejzo, J., Lepre, C.A., Peng, J.W., Bemis, G.W., Ajay, Murcko, M.A., Moore, J.M. (1999) The SHAPES strategy: an NMR-based approach for lead generation in drug discovery, *Chem. Biol.* **6**, 755–769.

Fesik, S.W. (1993) NMR structure-based drug design, *J. Biomol. NMR* **3**, 261–269.

Fesik, S.W., Gampe, R.T. Jr., Holzman, T.F., Egan, D.A., Edalji, R., Luly, J.R., Simmer, R., Helfrich, R., Klahore, V., Rich, D.H. (1990) Isotope-edited NMR of cyclosporin A bound to cyclophilin: evidence for a trans 9, 10-amide bond, *Science* **250**, 1406–1409.

Fesik, S.W., Gampe, R.T. Jr., Eaton, H.L., Gemmecker, G., Olejniczak, E.T., Neri, P., Holzman, T.F., Egan, D.A., Edalji, R., Simmer, R., Helfrich, R., Hochlowski, J., Jackson, M. (1991) NMR studies of [U-13C]cyclosporin A bound to cyclophilin: bound conformation and portions of cyclosporin involved in binding, *Biochemistry* **31**, 6574–6583.

Freedberg, D.I., Wang, Y.X., Stahl, S.J., Kaufman, J.D., Wingfield, P., Kiso, Y., Torchia, D.A. (1998) Flexibility and function in HIV protease: dynamics of the HIV-1 protease bound to the asymmetric inhibitor kynostatin 272 (KNI-272), *J. Am. Chem. Soc.* **120**, 7916.

Gale, D.J., Craik, D.J., Brownlee, R.T.C. (1988) Variable-temperature NMR studies of thyroid hormone conformations, *Magn. Reson. Chem.* **26**, 275–280.

Gehrmann, J., Alewood, P.F., Craik, D.J. (1998) Structure determination of the three disulfide bond isomers of alpha-conotoxin GI: a model for the role of disulfide bonds in structural stability, *J. Mol. Biol.* **278**, 401–415.

Ghose, A.K., Viswanadhan, V.N., Wendoloski, J.J. (1999) A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases, *J. Comb. Chem.* **1**, 55–68.

Gonnella, N., Lin, M., Shapiro, M.J., Wareing, J.R., Zhang, X. (1998) Isotope-filtered affinity NMR, *J. Magn. Reson.* **131**, 336–338.

Gonnella, N.C., Bohacek, R., Zhang, X., Kolossvary, I., Paris, C.G., Melton, R., Winter, C., Hu, S.I., Ganu, V. (1995) Bioactive conformation of stromelysin inhibitors determined by transferred nuclear Overhauser effects, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 462–466.

Gooley, P.R., O'Connell, J.F., Marcy, A.I., Cuca, G.C., Axel, M.G., Caldwell, C.G., Hagmann, W.K., Becker, J.W. (1996) Comparison of the structure of human recombinant short form stromelysin by multidimensional heteronuclear NMR and X-ray crystallography, *J. Biomol. NMR* **7**, 8–28.

Gradwell, M.J., Feeney, J. (1996) Validation of the use of intermolecular NOE constraints for obtaining docked structures of protein-ligand complexes, *J. Biomol. NMR* **7**, 48–58.

Grzesiek, S., Bax, A., Nicholson, L.K., Yamazaki, T., Wingfield, P.T., Stahl, S.J., Eyermann, C.J., Torchia, D.A., Hodge, C.N., Lam, P.Y., Jadhav,

P.K., Chang, C.H. (1994) NMR evidence for the displacement of a conserved interior water molecule in HIV protease by a nonpeptide cyclic urea-based inhibitor, *J. Am. Chem. Soc.* **116**, 1581–1583.

Hajduk, P.J., Meadows, R.P., Fesik, S.W. (1997) Discovering high-affinity ligands for proteins, *Science* **278**, 497–499.

Hajduk, P.J., Olejniczak, E.T., Fesik, S.W. (1997) One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules, *J. Am. Chem. Soc.* **119**, 12257–12261.

Hajduk, P., Gerfin, T., Boehlen, J., Haberli, M., Marek, D., Fesik, S. (1999) High-throughput nuclear magnetic resonance-based screening, *J. Med. Chem.* **42**, 2315–2317.

Hajduk, P., Dinges, J., Schkeryantz, J., Janowick, D., Kaminski, M., Tufano, M., Augeri, D., Petros, A., Nienbar, V., Zhong, P., Hammond, R., Coen, M., Beutel, B., Katz, L., Fesik, S. (1999) Novel inhibitors of Erm methyltransferases from NMR and parallel synthesis, *J. Med. Chem.* **42**, 3852–3859.

Hajduk, P., Sheppard, G., Nettesheim, D., Olejniczak, E., Shuker, S., Meadows, R., Steinman, D., Carrera, G., Marcotte, P., Severin, J., Walter, K., Smith, H., Gubbins, E., Simmer, R., Holzman, T., Morgan, D., Davidsen, S., Summers, J., Fesik, S. (1997) Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR, *J. Am. Chem. Soc.* **119**, 5818–5827.

Hann, M., Hudson, B., Lewell, X., Lifely, R., Miller, L., Ramsden, N. (1999) Strategic pooling of compounds for high-throughput screening, *J. Chem. Inf. Comput. Sci.* **39**, 897–902.

Henrichson, D., Ernst, B., Magnani, J.L., Wang, W., Meyer, B., Peters, T. (1999) Bioaffinity NMR spectroscopy: identification of an E-Sepectin antagonist in a substance mixture by transfer NOE, *Angew. Chem. Int. Ed.* **38**, 98–102.

Jacobsen, E.J., Mitchell, M.A., Hendges, S.K., Belonga, K.L., Skaletzky, L.L., Stelzer, L.S., Lindberg, T.J., Fritzen, E.L., Schostarez, H.J., O'Sullivan, T.J., Maggiora, L.L., Stuchly, C.W., Laborde, A.L., Kubicek, M.F., Poorman, R.A., Beck, J.M., Miller, H.R., Petzold, G.L., Scott, P.S., Truesdell, S.E., Wallace, T.L., Wilks, J.W., Fisher, C., Goodman, L.V., Kaytes, P.S., Ledbetter, S.R., Powers, E.A., Vogeli, G., Mott, J.E., Trepod, C.M., Staples, D.J., Baldwin, E.T., Finzel, B.C. (1999) Synthesis of a series of stromelysin-selective thiadiazole urea matrix metalloproteinase inhibitors, *J. Med. Chem.* **42**, 1525–1536.

Katoh, E., Yamazaki, T., Kiso, Y., Wingfield, P., Stahl, S.J., Kaufman, J.D., Torchia, D.A. (1999) Determination of the rate of monomer interchange in a ligand-bound homodimeric protein from NOESY cross peaks: application to the HIV protease/KNI-529 complex, *J. Am. Chem. Soc.* **120**, 7916.

Kay, L.E. (1998) Protein dynamics from NMR, *Nat. Struct. Biol.* **5**(Suppl. S), 513–517.

Keifer, P.A. (2000) NMR spectroscopy in drug discovery: tools for combinatorial chemistry, natural products, and metabolism research, *Prog. Drug Res.* **55**, 137–211.

Lam, P.Y., Jadhav, P.K., Eyermann, C.J., Hodge, C.N., Ru, Y., Bacheler, L.T., Meek, J.L., Otto, M.J., Rayner, M.M., Wong, Y.N., et al. (1994) Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors, *Science* **263**, 380–384.

Lepre, C.A. (2001) Library design for NMR-based screening, *Drug Discov. Today* **6**, 133–140.

Lepre, C., Moore, J. (1998) Microdrop screening: a rapid method to optimize solvent conditions for NMR spectroscopy of proteins, *J. Biomol. NMR* **12**, 493–499.

Lin, M., Shapiro, M.J., Wareing, J.R. (1997) Screening mixtures by affinity NMR, *J. Org. Chem.* **62**, 8930–8931.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Deliv. Rev.* **46**, 3–26.

Liu, G., Huth, J.R., Olejniczak, E.T., Mendoza, R., DeVries, P., Leitza, S., Reilly, E.B., Okasinski, G.F., Fesik, S.W., von Geldern, T.W. (2001) Novel p-arylthio cinnamides as antagonists of leukocyte function-associated antigen-1/intracellular adhesion molecule-1 interaction. 2. Mechanism of inhibition and structure-based improvement of pharmaceutical properties, *J. Med. Chem.* **44**, 1202–1210.

MacLachlan, L.K., Middleton, D.A., Edwards, A.J., Reid, D.G. (1997) A Case History – NMR Studies of the Structure of a Small Protein, w-Conotoxin MVIIA, in: Reid, D.G. (Ed.) *Protein NMR Techniques*, Humana Press, Totowa, NJ, pp. 337–362.

Marassi, F.M., Opella, S.J. (1998) NMR structural studies of membrane proteins, *Curr. Opin. Struct. Biol.* **8**, 640–648.

Mayer, M., Meyer, B. (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor, *J. Am. Chem. Soc.* **123**, 6108–6117.

McCoy, M.A., Dellwo, M.J., Schneider, D.M., Banks, T.M., Falvo, J., Vavra, K.J., Mathiowetz, A.M., Qoronfleh, M.W., Ciccarelli, R., Cook, E.R., Pulvino, T.A., Wahl, R.C., Wang, H. (1997) Assignments and structure determination of the catalytic domain of human fibroblast collagenase using 3D double and triple resonance NMR spectroscopy, *J. Biomol. NMR* **9**, 11–24.

Meadows, R.P., Nettesheim, D.G., Xu, R.X., Olejniczak, E.T., Petros, A.M., Holzman, T.F., Severin, J., Gubbins, E., Smith, H., Fesik, S.W. (1993) Three dimensional structure of the FK506 binding protein/ascomycin complex in solution by heteronuclear three- and four-dimensional NMR, *Biochemistry* **32**, 754–765.

Michnick, S.W., Rosen, M.K., Wandless, T.J., Karplus, M., Schreiber, S.L. (1991) Solution structure of FKBP, a rotamase enzyme and receptor for FK506 and rapamycin., *Science* **252**, 836–839.

Miljanich, G.P. (2004) Ziconotide: neuronal calcium channel blocker for treating severe chronic pain, *Curr. Med. Chem.* **11**, 3029–3040.

Moore, J.M., Peattie, D.A., Fitzgibbon, M.J., Thomson, J.A. (1991) Solution structure of the major binding protein for the immunosuppressant FK506, *Nature* **351**, 248–250.

Moy, F.J., Pisano, M.R., Chanda, P.K., Urbano, C., Killar, L.M., Sung, M.L., Powers, R. (1997) Assignments, secondary structure and dynamics of the inhibitor-free catalytic fragment of human fibroblast collagenase, *J. Biomol. NMR* **10**, 9–19.

Moy, F.J., Haraki, K., Mobilio, D., Walker, G., Powers, R., Tabei, K., Tong, H., Siegel, M.M. (2001) MS/NMR: a structure-based approach for discovering protein ligands and for drug design by coupling size exclusion chromatography, mass spectrometry, and nuclear magnetic resonance spectroscopy, *Anal. Chem.* **73**, 571–581.

Munro, S.L., Andrews, P.R., Craik, D.J., Gale, D.J. (1986) 13C NMR studies of the molecular flexibility of antidepressants, *J. Pharm. Sci.* **75**, 133–141.

Munro, S., Craik, D., McConville, C., Hall, J., Searle, M., Bicknell, W., Scanlon, D., Chandler, C. (1991) Solution conformation of endothelin, a potent vaso-constricting bicyclic peptide. A combined use of 1H NMR spectroscopy and distance geometry calculations, *FEBS Lett.* **278**, 9–13.

Nicholson, L.K., Yamazaki, T., Torchia, D.A., Grzesiek, S., Bax, A., Stahl, S.J., Kaufman, J.D., Wingfield, P.T., Lam, P.Y., Jadhav, P.K., et al. (1995) Flexibility and function in HIV-1 protease, *Nat. Struct. Biol.* **2**, 274–280.

Nielsen, K.J., Thomas, L., Lewis, R.J., Alewood, P.F., Craik, D.J. (1996) A consensus structure for omega-conotoxins with different selectivities for voltage-sensitive calcium channel subtypes: comparison of MVIIA, SVIB and SNX-202, *J. Mol. Biol.* **263**, 297–310.

Nielsen, K.J., Adams, D., Thomas, L., Bond, T., Alewood, P.F., Craik, D.J., Lewis, R.J. (1999) Structure-activity relationships of omega-conotoxins MVIIA, MVIIC and 14 loop splice hybrids at N and P/Q-type calcium channels, *J. Mol. Biol.* **289**, 1405–1421.

Nieto, P.M., Birdsall, B., Morgan, W.D., Frenkiel, T.A., Gargaro, A.R., Feeney, J. (1997) Correlated bond rotations in interactions of arginine residues with ligand carboxylate groups in protein ligand complexes, *FEBS Lett.* **405**, 16–20.

Oprea, T.I., Gottfries, J., Sherbukhin, V., Svensson, P., Kuhler, T.C. (2000) Chemical information management in drug discovery: optimizing the computational and combinatorial chemistry interfaces, *J. Mol. Graph. Model.* **18**, 512–524, 541.

Otting, G., Liepinsh, E., Wuthrich, K. (1991) Protein hydration in aqueous solution, *Science* **254**, 974–980.

Patt, W.C., Edmunds, J.J., Repine, J.T., Berryman, K.A., Reisdorph, B.R., Lee, C., Plummer, M.S., Shahripour, A., Haleen, S.J., Keiser, J.A., Flynn, M.A., Welch, K.M., Reynolds, E.E., Rubin, R., Tobias, B., Hallak, H., Doherty, A.M. (1997) Structure-activity relationships in a series of orally active gamma-hydroxy butenolide endothelin antagonists, *J. Med. Chem.* **40**, 1063–1074.

Patt, W.C., Cheng, X.M., Repine, J.T., Lee, C., Reisdorph, B.R., Massa, M.A., Doherty, A.M., Welch, K.M., Bryant, J.W., Flynn, M.A., Walker, D.M., Schroeder, R.L., Haleen, S.J.,

Keiser, J.A. (1999) Butenolide endothelin antagonists with improved aqueous solubility, *J. Med. Chem.* **42**, 2162–2168.

Pavlopoulos, S., Rose, M., Wickham, G., Craik, D.J. (1995) A 1H NMR analysis of the interaction between terephthalamide derivatives and the oligonucleotide duplex d(GGTAATTACC)2, *Anticancer Drug Des.* **10**, 623–639.

Pelton, J.G., Wemmer, D.E. (1989) Structural characterization of a 2:1 distamycin A.d(CGCAAATTGGC) complex by two-dimensional NMR, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5723–5727.

Pelton, G.J., Wemmer, D.E. (1990) Structural characterization of a 2:1 distamycin A d(CGCAAATTGGC)2 determined by two-dimensional NMR, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5723–5727.

Peng, J.W., Lepre, C.A., Fejzo, J., Abdul-Manan, N., Moore, J.M. (2001) Nuclear magnetic resonance-based approaches for lead generation in drug discovery, *Meth. Enzymol.* **338**, 202–230.

Pervushin, K., Riek, R., Wider, G., Wuthrich, K. (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12366–12371.

Petros, A.M., Gampe, R.T., Jr., Gemmecker, G., Neri, P., Holzman, T.F., Edalji, R., Hochlowski, J., Jackson, M., McAlpine, J., Luly, J.R., et al. (1991) NMR studies of an FK-506 analogue, [U-13C]ascomycin, bound to FKBP: conformation and regions of ascomycin involved in binding, *J. Med. Chem.* **34**, 2925–2928.

Pjura, P.E., Grzeskowiak, K., Dickerson, R.E. (1987) Binding of Hoechst 33258 to the minor groove of B-DNA, *J. Mol. Biol.* **197**, 257–271.

Polshakov, V.I., Birdsall, B., Frenkiel, T.A., Gargaro, A.R., Feeney, J. (1999) Structure and dynamics in solution of the complex of *Lactobacillus casei* dihydrofolate reductase with the new lipophilic antifolate drug trimetrexate, *Protein Sci.* **8**, 467–481.

Renaud, J.P., Rochel, N., Ruff, M., Vivat, V., Chambon, P., Gronemeyer, H., Moras, D. (1995) Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid, *Nature* **378**, 681–689.

Rose, J.R., Salto, R., Craik, C.S. (1993) Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitutions, *J. Biol. Chem.* **268**, 11939–11945.

Rosen, M.K., Standaert, R.F., Galat, A., Nakatsuka, M., Schreiber, S.L. (1990) Inhibition of FKBP rotamase activity by immunosuppressant FK506: twisted amide surrogate, *Science* **248**, 863–866.

Ross, A., Schlotterbeck, G., Klaus, W., Senn, H. (2000) Automation of NMR measurements and data evaluation for systematically screening interactions of small molecules with target proteins, *J. Biomol. NMR* **16**, 139–146.

Searle, M.S., Embrey, K.J. (1990) Sequence-specific interaction of Hoechst 33258 with the minor groove of an adenine-tract DNA duplex studied in solution by 1H NMR spectroscopy, *Nucleic Acids Res.* **18**, 3753–3762.

Shapiro, M. (2001) Applications of NMR screening in the pharmaceutical industry, *Farmaco* **56**, 141–143.

Shuker, S.B., Hajduk, P.J., Meadows, R.P., Fesik, S.W. (1996) Discovering high-affinity ligands for proteins: SAR by NMR, *Science* **274**, 1531–1534.

Teng, M.K., Usman, N., Frederick, C.A., Wang, A.H. (1988) The molecular structure of the complex of Hoechst 33258 and the DNA dodecamer d(CGCGAATTCGCG), *Nucleic Acids Res.* **16**, 2671–2690.

Tjandra, N., Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, *Science* **278**, 1111–1114.

Tjandra, N., Garrett, A.M., Gronenborn, A.M., Bax, A., Clore, G.M. (1997) Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy, *Nat. Struct. Biol.* **4**, 443–449.

Van Doren, S.R., Kurochkin, A.V., Hu, W., Ye, Q.-Z., Johnson, L.L., Hupe, D.J., Zuiderweg, E.R. (1995) Solution structure of the catalytic domain of human stromelysin complexed with a hydrophobic inhibitor, *Protein Sci.* **4**, 2487–2498.

van Duyne, G.D., Standaert, R.F., Karplus, M., Schreiber, S.L., Clardy, J. (1991) Atomic structure of FKBP-FK506, an immunophilin-immunosuppressant complex., *Science* **252**, 839–842.

Wagner, R.L., Apriletti, J.W., McGrath, M.E., West, B.L., Baxter, J.D., Fletterick, R.J. (1995)

A structural role for hormone in the thyroid hormone receptor, *Nature* **378**, 690–697.

Wallace, B.A., Janes, R.W., Bassolino, D.A., Krystek, S.R. Jr. (1995) A comparison of X-ray and NMR structures for human endothelin-1, *Protein Sci.* **4**, 75–83.

Wandless, T.J., Michnick, S.W., Rosen, M.K., Karplus, M., Schreiber, S.L. (1991) FK506 and rapamycin binding to FKBP: common elements in immunophilin-ligand complexation, *J. Am. Chem. Soc.* **113**, 2339–2341.

Wang, Y.X., Freedberg, D.I., Grzesiek, S., Torchia, D.A., Wingfield, P.T., Kaufman, J.D., Stahl, S.J., Chang, C.H., Hodge, C.N. (1996) Mapping hydration water molecules in the HIV-1 protease/DMP323 complex in solution by NMR spectroscopy, *Biochemistry* **35**, 12694–12704.

Wishart, D.S., Sykes, B.D., Richards, F.M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy, *Biochemistry* **31**, 1647–1651.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., Sykes, B.D. (1995) 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, *J. Biomol. NMR* **5**, 67–81.

Wuthrich, K. (2003) NMR studies of structure and function of biological macromolecules (Nobel Lecture), *Angew. Chem.-Int. Ed.* **42**, 3340–3363.

Xu, J., Stevenson, J. (2000) Drug-like index: a new approach to measure drug-like compounds and their diversity, *J. Chem. Inf. Comput. Sci.* **40**, 1177–1187.

Yamazaki, T., Nicholson, L.K., Wingfield, P., Stahl, S.J., Kaufman, J.D., Domaille, P.J., Torchia, D.A. (1994) NMR and X-ray evidence that the HIV protease catalytic aspartyl groups are protonated in the complex formed by the protease and a nonpeptide cyclic urea-based inhibitor, *J. Am. Chem. Soc.* **116**, 10791–10792.

Yamazaki, T., Hinck, A.P., Wang, Y.X., Nicholson, L.K., Torchia, D.A., Wingfield, P., Stahl, S.J., Kaufman, J.D., Chang, C.H., Domaille, P.J., Lam, P.Y. (1996) Three-dimensional solution structure of the HIV-1 protease complexed with DMP323, a novel cyclic urea-type inhibitor, determined by nuclear magnetic resonance spectroscopy, *Protein Sci.* **5**, 495–506.

Superantigens

Matthew D. Baker and K. Ravi Acharya
University of Bath, Claverton Down, Bath, UK

1 Introduction 639

2 The Superantigen Family 639
2.1 Molecular Architecture 640
2.2 Nonclassical Superantigens 641
2.3 The Immune Response to Bacterial Superantigens 641

3 Binding to MHC Class II Molecules 641
3.1 The Generic Binding Site and the MHC Class II α-chain 642
3.2 Zinc-dependent Binding to MHC Class II β-chain 643

4 Binding to the T-Cell Receptor 645
4.1 General Mechanism of Binding for Classical Superantigens to T-Cell Receptor V_β Elements 645
4.2 Binding to the T-Cell Receptor Vα Element 646

5 Formation of the Trimeric Complex for Signal Transduction 647

6 Roles of Superantigens in Disease 647
6.1 Toxic Shock Syndrome 648
6.2 Staphylococcal Food Poisoning 648
6.3 Autoimmunity 649

Bibliography 649
Books and Reviews 649
Primary Literature 649

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

Keywords

Antigen Presenting Cell
Cells, including macrophages, B cells, and dendritic cells that take up and process an antigen into a peptide fragment capable of being presented by an MHC class II molecule on the cell surface.

Major Histocompatibility Complex Class II
Present peptide antigen to T cells.

Major Histocompatibility Complex Class II Binding Site
The region of a superantigen that recognizes and binds to MHC class II molecules.

Pyrogenic Exotoxins
A group of superantigens produced by the gram-negative bacteria *Streptococcus pyogenes*.

Staphylococcal Enterotoxins
A group of superantigens produced by the gram-negative bacteria *Staphylococcus aureus*.

Superantigen
A substance (as an enterotoxin) that acts as an antigen capable of stimulating much larger numbers of T cells than ordinary antigen.

T Cell
T lymphocyte; a cell bearing a T-cell receptor, which can control and mediate humoral immunity.

T-Cell Receptor
The receptor that enables the T cell to recognize and react to specific antigens.

T-Cell Receptor Binding Site
The region of a superantigen that recognizes and binds to the V_β elements of T-cell receptors.

Superantigens (SAgs) are a family of proteins produced primarily by *Staphylococcus aureus* and *Streptococcus pyogenes*. These toxins acquired the name *Superantigen* due to their ability to hyperstimulate the immune system. SAgs are not recognized by the immune system in the same way as conventional foreign proteins, and interact in a unique fashion with major histocompatibility complex (MHC) class II molecules and T-cell receptors (TCRs). Once formed, the trimolecular complex induces vast

T-cell proliferation resulting in massive cytokine release, epithelial damage, capillary leak, and hypotension. Recent additions to the family include toxins produced by *Mycoplasma arthritidis* and *Yersinia Pestis*. SAgs are implicated in several diseases including toxic shock syndrome (TSS), scarlet fever, and food poisoning; their function appears to be primarily to weaken the host sufficiently to allow disease to take hold.

1 Introduction

Conventional antigens: peptide fragments derived from exogenous proteins are displayed on the surface of antigen presenting cells (APC) by major histocompatibility complex II (MHC) molecules. These peptides are then recognized by circulating T cells and an immune response targeted toward that particular peptide is initiated. In contrast, superantigens (SAgs) are not processed by APC and remain as fully intact proteins. Structural evidence shows SAgs to bind to APCs on the external face of the MHC class II molecule and concomitantly to T cells via the external face of the T-cell receptor (TCR) V_β element. In contrast with the specificity that results from TCR binding to one particular type of conventional peptide antigen, interaction between SAg and the V_β element of the TCR supplies a much broader range of specificity, as whole subsets of T cells displaying the same V_β types are able to recognize the SAg.

Superantigenic toxins are very stable proteins. They are highly resistant to proteases, denaturation from heat and extreme pH, and are often found in food that are high in protein, salt, and sugar. In order to cause food poisoning, they must be able to survive not only cooking and preserving procedures, but also the digestive process.

2 The Superantigen Family

In order to be classified as an SAg, a toxin must demonstrate the ability to stimulate T cells in a V_β specific manner. The number of bacterial SAgs or proteins highly homologous to known SAgs has grown considerably over the last decade so much so that an international nomenclature committee for staphylococcal SAgs have described an international procedure for the naming of newly described SAgs and putative SAgs.

The staphylococcal enterotoxins (SEs) A, B, C1–3, D, E, H, I, and J toxic shock syndrome toxin-1 (TSST-1) were originally designated enterotoxins because of their emetic property when ingested. Since 2000, several new "SEs" have been discovered on the basis of sequence alignments, though it is proposed that none should be classified as SEs until it is demonstrated that they induce emesis after oral administration.

Along with the staphylococcal enterotoxins, the streptococcal pyrogenic exotoxins (SPEs) A, C, and H; streptococcal mitogenic exotoxin (SME)-Z_2, and streptococcal superantigen (SSA) produced by *Streptococcus pyogenes* represent the most well-characterized bacterial SAgs. Other pathogens, such as *Mycoplasma arthritidis* and *Yersinia pseudotuberculosis* have also been shown to secrete superantigenic

proteins. For the most part, this review will focus on the staphylococcal and streptococcal SAgs, with some discussion on the emerging SAgs from *M. arthritidis* and *Y. pseudotuberculosis* (*Mycoplasma arthritidis* mitogen – MAM and *Yersinia pseudotuberculosis* mitogen – YPM), as well as the wealth of putative SAgs uncovered by genomic initiatives. Broadly speaking, staphylococcal and streptococcal SAgs can be grouped into four subfamilies on the basis of amino acid sequence and three-dimensional structure: The first group is made up of the staphylococcal enterotoxins SEA, SED, SEE, SEH, SEI, and SEJ; both staphylococcal and streptococcal toxins form the second group, which consists of SEB, SEC1-3, SpeA1-3, SSA, and SEG; the third group contains streptococcal pyrogenic- and mitogenic toxins SpeC, SpeJ, SpeG, SpeH, SME-Z, SME-Z_2, and streptococcus dysgalactiae-derived mitogen (SDM); the SAg-like proteins SSL and TSST-1 form the fourth group. MAM and YPM have no sequence or structural homology to the rest of the bacterial SAgs and cannot be grouped with any of these subfamilies.

2.1 Molecular Architecture

Comparison of the three-dimensional structures of staphylococcal and streptococcal SAgs reveal a conserved two-domain architecture (Fig. 1). The two domains, known as the N- and C-terminal domains, are separated by a long solvent-exposed α-helix, spanning the center of the molecule. Typically, the N-terminal domain contains many hydrophobic residues in its solvent-exposed regions and structurally it has a significant similarity to the oligosaccharide/oligonucleotide binding fold (OB-fold) found in other proteins. In spite of this similarity, SAgs have not been seen to bind to either DNA or carbohydrates. The C-terminal domain resembles the β-grasp motif and is composed of a four-stranded β-sheet that is centrally capped by a long α-helix. Several SAgs also contain a highly flexible disulfide loop. Located in the N-terminal domain (Fig. 1), this flexible loop is implicated in the emetic properties of the staphylococcal and streptococcal toxins, as mutations of the two cysteine residues that form at the base of this disulfide loop abolishes the ability of these toxins

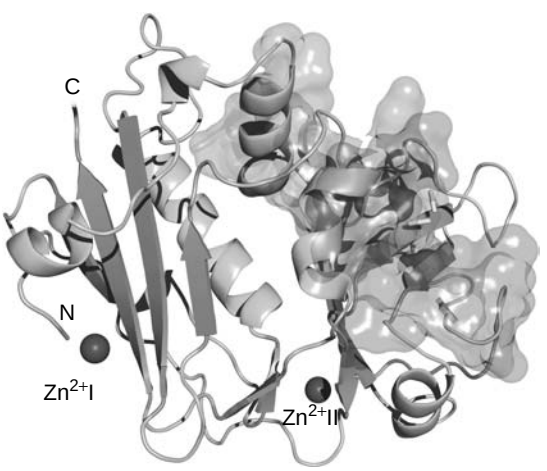

Fig. 1 Ribbon diagram of staphylococcal enterotoxin A (SEA), representative of the common structural features of the staphylococcal and streptococcal SAg family. Blue spheres represent the positions of the two zinc sites. The cysteine residues that form the disulfide loop are shown in stick representation (yellow), the surfaces defining the TCR binding region (green), and MHC class II binding region (red) are shown (see color plate p. xlii).

to induce vomiting in primates. Additionally, this loop has also been implicated in T-cell stimulation for several members of the staphylococcal and streptococcal SAg family.

2.2 Nonclassical Superantigens

The staphylococcal SAg-like proteins SSL5 and SSL7 also show the common architecture preserved among these related groups of proteins. This group of proteins is most closely related to TSST-1, although none of the SSLs have demonstratable superantigenic activity, and as such are linked to the SAg family purely due to sequence identity and structural homology. Differences between the SSLs and classical SAgs include extended secondary structure elements, extended loop regions, and differences in charge distribution over the protein surface. At present, it is not known whether these features have any biological significance.

The discovery and characterization of SAgs from bacterial sources such as *M. arthritidis* and *Y. pseudotuberculosis* have extended the topology of the family even further. MAM and YPMa show no structural homology to the rest of the SAg family whatsoever. MAM is composed entirely of α-helices, arranged in two domains to form an L shape. The N-terminal domain consists of four α-helices and a 25-residue N-terminal loop that wraps around the four-helix bundle, whereas The C-terminal domain contains six α-helices. YPMa is much smaller than the classical staphylococcal and streptococcal toxins (14 kDa), and consists of a jelly-roll fold comprising two β-sheets, each containing four antiparallel strands. Similarly, the structure of YPMa is, unlike any other known SAgs, structurally similar to viral capsid proteins.

2.3 The Immune Response to Bacterial Superantigens

Superantigens primarily target $CD4^+$ T cells, leading to a T-helper type 1 (Th1) response and no Th2 response. A dominant Th1 response causes the suppression of antibody production, which in turn reduces the clearance of the invading microbe, allowing disease progression. The resultant release of a variety of cytokines in high quantities from macrophages and T cells cause many of the symptoms associated with Sag-related disease, and eventually leads to hypotension through capillary leak. The array of cytokines produced include IL-1 and TNF-α from macrophages and TNF-β, IL-2, and IFN-γ from T cells. Most cytokine expression is transient and occurs within 12 h of SAg exposure, whereas signals involved in proliferation such as DNA synthesis and IL-2 receptor expression do not occur until over 24 h after toxin exposure. Once cells have undergone DNA synthesis, they are capable of re-expressing cytokines prior to a second round of proliferation.

3 Binding to MHC Class II Molecules

MHC class II molecules possess two separate binding sites for the classical staphylococcal and streptococcal SAgs. The first, a low-affinity binding site that is also known as the *generic site* is located on the α-chain of the MHC class II molecule (Fig. 2), whereas the second, a high-affinity zinc-dependent site is located

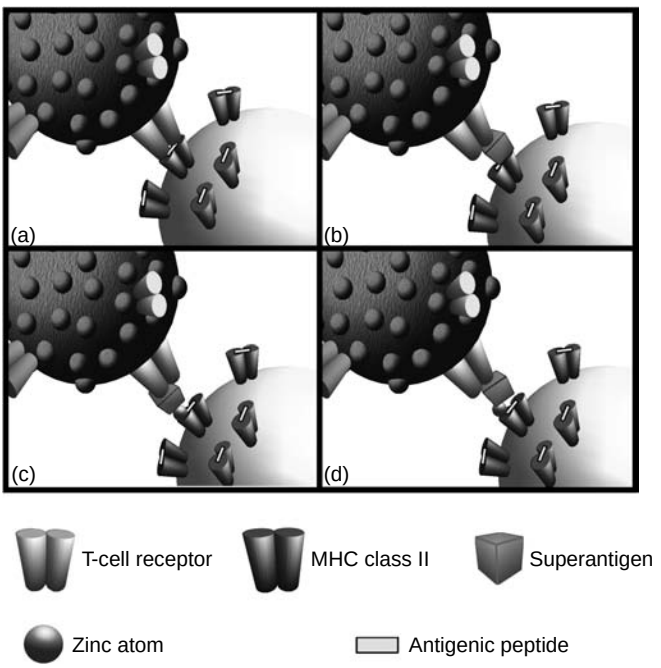

Fig. 2 Schematic representation illustrating the differences between conventional peptide antigen presentation and SAg presentation to MHC class II and T-cell receptors. (a) Conventional antigen is processed by the antigen presenting cell and displayed as discrete peptide fragments within the peptide binding groove of MHC class II molecules, direct interaction occurs between TCR and MHC II molecule; (b) SAgs bind to the solvent-exposed face of the MHC class II molecule (α-1) via its generic site, forming a bridge between TCR (Vβ) and MHC class II; (c) SAg binds to MHC class II (β1) via a bridging zinc atom in a high affinity, zinc-dependent manner; and interacts with TCR Vα as predicted for SEH; (d) SAg binds to MHC class II (β1) via a high-affinity zinc site, and interacts with TCR Vβ as is postulated for the remainder of the zinc containing SAgs. In both cases, the MHC class II molecule associated antigenic peptide has been shown to influence T-cell recognition of SAg/MHC class II molecule.

on the β-chain. SAgs bind via this high-affinity site approximately 100 times more strongly than to the generic site. A great deal of structural and mutational data has revealed that a range of diversity exists within the SAg family for binding to MHC class II molecules. Each SAg binds to different alleles of class II molecules to varying extent. While most of the SAgs, including TSST-1, SEB, and MAM bind preferentially to HLA-DR alleles, SAgs, such as SEC, SpeA, and SSA bind predominantly to HLA-DQ alleles.

3.1 The Generic Binding Site and the MHC Class II α-chain

X-ray crystallographic studies on SEB and TSST-1 in complex with the MHC class

II molecule HLA-DR1 via the low-affinity generic site, SpeC in complex with HLA-DR2, and SEH in complex with HLA-DR1 via the high-affinity zinc site have yielded a great deal of structural information about the binding of SAgs to MHC class II molecules. The SEB– and TSST-1–HLA-DR1 complexes demonstrate that both SEB and TSST-1 have similar binding modes to HLA-DR1. Interaction between HLA-DR1 and these SAgs mainly involves the α-chain of the DR1 molecule and the solvent-exposed, hydrophobic core at the N-terminal domain of the toxin. Similar hydrophobic ridge regions exist in several other SAgs and form the generic MHC class II binding site. One feature seen in the TSST-1–HLA-DR1 complex, (but not in the SEB–HLA-DR1 complex) is the number of additional contacts between TSST-1 and the peptide antigen displayed by the HLA-DR1. Indeed, truncating the C-terminal end of the antigenic peptide significantly affects TSST-1 binding to MHC class II molecules in mice.

3.2
Zinc-dependent Binding to MHC Class II β-chain

Several of the bacterial SAgs, with the exception of SEB, TSST-1, SSA, SSL1-, and YPM, have one or more zinc binding sites (Fig. 1). Zinc ions have been shown to play two distinct roles in SAg function. First, they act as dimerization sites, second, they have been shown to be important for the recognition of the MHC class II β-chain. SEA possesses a high-affinity zinc binding site in its C-terminal domain; this site has a substantially higher affinity for DR1 ($K_d = 100$ nM) compared with the generic binding site ($K_d = 10$ μM), and if the two binding sites coexist, a K_d of 13 nM is observed. The importance of both of these sites is demonstrated by the fact that mutation of residues in either of these sites abolishes the toxin's ability to induce cytokine expression. The coexistence of these two distinct MHC class II binding sites may enable it to form a trimeric SEA–MHC–SEA complex, a phenomenon that has been observed in solution. Several other SAgs including SEE possess identical zinc ligands, and a similar case for SAg–MHC–SAg complex formation can be argued. Both SED and SpeC possess analogous zinc binding sites to SEA. In SED, the zinc ion acts as a means for toxin dimerization; SED can form zinc-dependent homodimers, whereas SpeC forms zinc-independent homodimers. Both SED and SpeC bind to the β-chain of MHC class II molecule by a zinc-mediated mechanism similar to that of SEA, and because of the ability of these toxins to dimerize, formation of trimers and/ or tetramers is a possibility. A similar binding mechanism has been proposed for SEH, SME-Z, SME-Z_2, SpeG, and SpeH, all of which lack a generic MHC class II binding site. This high-affinity zinc site, present in SEA is not seen in SEC or SpeA. Instead, a separate zinc binding site (with an estimated K_d of approximately 1 μM) is seen at the N-terminal domain of these toxins. In a similar fashion to the SEC-like zinc binding site, this secondary zinc binding site is also important for MHC class II binding (Figs. 1 and 2). Direct observation of zinc-mediated SAg–MHC class II complexes such as those seen in the crystal structures of SEH–HLA-DR1 and SpeC–HLA-DR2, have enabled a detailed analysis of these interactions. A great deal of similarity exists between the two complexes. The interactions between both SAgs and their MHC class II molecules are governed by a tetrahedrally coordinated zinc ion. For SpeC, three of the

coordinating ligands are from the toxin, and for SEH, two are from the toxin and one is from a water molecule. In both complexes, the fourth ligand is provided by residue His81 of the β1 chain of the class II molecule. In a similar fashion to the TSST-1-MHC class II complex, interaction between the SAg and the antigenic peptide is observed. Up to one-third of the contact area between SAg and MHC class II is taken up by antigenic peptide. In both cases, a majority of the interactions occur with the backbone atoms of the peptide, enabling both SpeC and SEH to have similar interactions with the antigenic peptide even though the composition of the peptide differs in both complexes. This would suggest that although the antigenic peptide plays an important role in the complex interaction, MHC class II binding is not entirely peptide specific.

To confuse matters further, another zinc binding site has recently been discovered in SEC2. This zinc ion is located near the generic MHC class II binding site and is thought to act as a site of dimerization for SEC2. If this zinc ion does indeed allow dimerization of SEC2, the generic MHC class II site would be blocked by the dimer interface; thus, MHC class II binding would be governed by the SEC2 primary zinc binding site. As a result, SEC2 would be able to bind to MHC class II molecules as a zinc-mediated dimer. SpeA1 and SpeC are also capable of forming homodimers, although in this case the mechanism is zinc independent. SpeA1 has recently been shown to exist in a disulfide linked dimeric form via cysteine residues located within the flexible disulfide loop, whereas the SpeC dimer is formed by direct interaction of the surfaces that usually form the generic MHC class II binding site.

Zinc was also thought to play a similar role in MAM, namely, binding to MHC class II molecules and toxin dimerization. However, the absence of zinc in the crystal structure of dimeric MAM in complex with HLA-DR1 suggests that zinc fulfills neither role. Clearly, the exact role of zinc in the action of this novel SAg requires further investigation.

Despite the MAM_2–HLA-DR1 complex bearing no immediate likeness to that of other bacterial SAgs, there are some similarities in the overall mechanism of interaction. Complex formation does not induce any major structural changes at the binding interface, which is consistent with other SAg–MHC class II complexes. The MAM_2–HLA-DR1 complex is formed through interaction of the N-terminal domain of MAM and a number of regions of HLA-DR1. Contact with the antigenic peptide (hemagglutinin peptide) in the peptide binding groove of the HLA-DR1 molecule is again observed. Interestingly, the MAM binding site on HLA-DR1 also overlaps the generic binding site for staphylococcal and streptococcal SAgs. In support of this, both SEB and TSST-1 have been shown to block MAM binding to MHC class II molecules. Moreover, MAM binding is also disrupted by those SAgs that bind to the high-affinity zinc site, as the MAM binding site on MHC class II also includes the region on the MHC β-chain, which contains His81. Although the crystal structure reveals MAM to bind to MHC class II molecules as a dimer, both monomer and homodimers are shown to be present in solution, indicating that MAM may be able to act on MHC class II molecules as both monomer and dimer.

Current understanding of SAg structure and function suggests that there is an increasing range of mechanisms by which SAgs can interact with MHC class II molecules. Although differences exist

between SAgs, three main binding modes are observed: (1) through zinc-mediated interaction; (2) via the generic site; or (3) as a homodimer, which can be either zinc or nonzinc mediated. In addition, it is also possible that a combination of any of these three mechanisms may be employed.

4 Binding to the T-Cell Receptor

In comparison to their interaction with MHC class II molecules, the binding of SAgs to T-cell receptors is at first sight a much more localized interaction. SAgs were thought to bind exclusively TCR via the TCR V_β element through a universally similar mechanism, where the binding is mediated by interactions between the side chains of the SAg and the backbone atoms from the V_β region. This results in each SAg selectively expanding T-cell subsets bearing certain V_β elements, while excluding others.

4.1 General Mechanism of Binding for Classical Superantigens to T-Cell Receptor V_β Elements

The TCR binding sites of the staphylococcal and streptococcal SAgs consist of comparable scaffold regions within which specific amino acids affect the V_β specificity. The binding site is formed by a shallow groove between the two domains of the molecule (Fig. 1). The structures of SEB, SEC2, and SEC3 in complex with TCR V_β seem to support this simple binding mechanism, a majority of the contacts between SAg and TCR are between the amino acid side chains within this groove and backbone atoms of the TCR V_β elements. The main interactions are between the side-chain atoms of the SAg and complementarity determining regions 1 and 2 (CDRs 1 and 2), and hypervariable region 4 (HV4) of the V_β chain. Close examination of the TCR binding sites of SEC2/3, SEA, and SEB indicate that an invariant asparagine residue (Asn23 in SEB/SECs; Asn25 in SEA) is essential for direct interactions with the TCR, and if this residue is mutated, T-cell stimulation is lost. This residue is located on the surface of the toxin, and is thought to have similar interactions in all the SEs. Mutagenesis studies have revealed that further interaction exists between SAg and TCR, which are conserved among many of the staphylococcal and streptococcal SAgs, including SEC1-3, SEB, SpeA, and SSA. In addition to this "common core" of residues, the differences in TCR affinity among the SAgs can be accounted for by several residues unique to each toxin. For example, the type of residue at position 26 of SEC2 confers specificity between SEC1, SEC2, SEA, and SEB via its interaction with Gly53 from the V_β chain. Similarly, the residue at position 91 of SEC2 is also implicated in TCR binding and is not conserved in SEA or SEB. By switching several residues for a dissimilar group of amino acids, a decreased affinity for certain V_β chains and an increase in affinity for others may occur. This is likely to be as a direct result of an increase in the amount of unfavorable interactions for one V_β element followed by an increase of favorable interactions for another V_β element. Experiments with SEA support this idea. SEA specificity for TCR is thought to be governed by three residues at position 21, 206, and 207. Exchange of these three residues for the homologous residues in SEE causes a change in the profile of the responding

V_β elements to that of the profile normally seen for T cells stimulated by SEE. It has also been shown that those residues that define the specificity of an SAg for particular V_β elements make the greatest energetic contribution to the overall stability of the SAg–TCR complex, and as the above evidence shows, these residues are comparatively few in number.

More recently, the two SAg–TCR V_β complexes SpeA1–V_β 8.2 and SpeC–V_β 2.1 have indicated that TCR binding, for at least some SAgs is not as simple as first thought. The interaction of SpeA1 with the TCR β-chain is similar to that of SEB, with several of the important contacts preserved. In addition, SpeA1 has an extra five hydrogen bonds with the TCR V_β compared to SEB. Thus, the mechanism of TCR binding employed by SpeA1 is slightly more intricate than those employed by SEB and SEC2/3. The complexity increases still further with SpeC; this toxin was found to bind to significantly more V_β residues than SEC2/3, SEB, or SpeA1, including numerous interactions with both main-chain and side-chain atoms of the TCR β-chain. The increased amount of contacts are most likely due to the deeper and broader cleft between the two domains of SpeC compared with other SAgs, as this is where the TCR binding site is located. Structural similarity between SpeC and TSST-1 mean that although the TCR binding site of TSST-1 is not yet fully characterized, it is likely that it shares similar binding characteristics. The available mutagenesis data suggests that the TCR binding site of TSST-1 is located in a similar position to SpeC, between the two domains of the toxin.

Both MAM and YPMa have been shown to bind to the TCR β-chain, although in the absence of a crystal structure of their complexes, the exact location and composition of their TCR binding sites are speculative. Clearly, more work needs to be carried out to determine the binding mechanisms of these novel SAgs before a direct comparison can be made to the staphylococcal and streptococcal toxins.

The characterization of SAg–TCR complexes have demonstrated that multiple modes of TCR interaction exist. The first, a high specificity binding mode, involves many contacts by both backbone atoms and side-chain atoms over an increased area as demonstrated by SpeC. This mode has high affinity for relatively few TCR V_β elements. The second, a moderate specificity mode, has fewer interactions over a reduced contact area. This mode has a modest affinity for a larger group of TCR V_β elements, such as seen in SpeA1. The final mode, a promiscuous binding mode, such as is seen in SEB and SEC2/3, suggests that binding to TCR V_β chains for these SAgs is a matter of simple conformational dependence.

4.2
Binding to the T-Cell Receptor Vα Element

The discovery that SEH, in contrast to all other SAgs, stimulates T cells in a V_α specific fashion, completely lacking any V_β expansion, adds further complication to the scheme. It is suggested that SEH may bind to TCR in a V_α specific fashion due to its presentation by MHC class II molecules via a zinc atom. Whether other SAgs that are presented to TCR by MHC class II in the same way can bind to V_α in the same way is yet to be established.

5
Formation of the Trimeric Complex for Signal Transduction

The complex course of events at the cell surface that lead to the formation of the MHC II–SAg–TCR complex is still ambiguous. It is possible for SAg to exist as monomer and dimer, and many SAgs have multiple MHC class II binding sites enabling them to form complexes with more than one MHC class II molecule or TCR. In terms of trimeric complex formation, this means that even further diversity exists in the ability of these toxins to interact with the immune system.

When considering complex formation, it is important to first consider the nature of membrane-bound receptors. In order for a complex to form, cell membranes need to be in close proximity to each other, and the receptors must diffuse to the site of interaction. It has been shown that in order for superantigenic T-cell activation to occur, a comparatively low percentage (<0.3%) of the MHC class II molecules must be occupied by the SAg. Higher concentrations of bound toxin result in an aborted T-cell response after a few cell divisions due to apoptosis. Thus, a low local concentration of MHC class II molecules on the cell interface is preferable for optimum superantigenicity. It is thought that the binding of a SAg to TCR induces clustering of the TCRs on the cell surface and the acquirement of the intracellular components required to transduce a signal. This is thought to occur in a manner that mimics peptide antigens with regard to receptor clustering, either through direct clustering events as proposed by the TCR oligomerization model, or by the binding of SAg homodimers to multiple MHC class II molecules, which in turn would promote T-cell clustering. Molecular modeling suggests that signal transduction stimulated by SEA through large-scale assembly is limited to four or five TCR–(DR1β-SEA-DR1α) tetramers, and requires the dimerization of MHC class II molecules. While TCRs would be clustered together in this model, TCR dimerization is thought unlikely. SEA is not unique in its ability to form zinc-mediated dimers. SED and SEC2 can form zinc-dependent homodimers, whereas SpeA1 forms a nonzinc mediated disulfide linked dimer; a SpeC dimer is also formed in the absence of zinc. MAM has also been shown to be able to form zinc-dependent and zinc-independent dimers, and YPM is thought to be able to form zinc-independent trimers. Clearly, SAgs have evolved slightly different ways of inducing receptor clustering. SAgs that act as monomers and possess only a single MHC class II binding site appear to rely on the interactions of the TCR V_α and MHC class II-β1, which increases the stability of the ternary complex to within the range seen for conventional antigen. A stable MHC–SAg–TCR complex with an extended half-life may therefore assist receptor clustering.

6
Roles of Superantigens in Disease

It should be noted that the association of an SAg with a certain disease does not mean that it necessarily causes the disease. In some cases, the presence of a microbe known to produce an SAg has been correlated with the disease. In other cases, expansion of a certain V_β subset is interpreted as evidence of the involvement of SAg in a disease. A definitive confirmation of the involvement of

an SAg requires the cloning of the putative SAg and its subsequent testing in an animal model. Thus, the involvement of SAgs in disease is established for relatively few conditions: these include toxic shock syndrome (TSS), food poisoning, Kawasaki disease, and scarlet fever.

6.1 Toxic Shock Syndrome

TSST-1 is the main causative agent for TSS and induces most of the symptoms associated with the disease in experimental animals. TSST-1 is responsible for nearly all cases of menstrual TSS, and approximately 60% of nonmenstrual staphylococcal TSS. The remainder of the cases can be attributed primarily to other SEs including SEB, SEC, and SEA, whereas SpeA is associated with most cases of streptococcal TSS. Although the exact nature of the disease is not fully understood, it is clear that toxic shock is in part a result of the superantigenic nature of these toxins. Superantigenicity and massive T-cell proliferation affect the cardiovascular system by causing extensive epithelial damage, capillary leak, and a decrease in peripheral vascular resistance resulting in shock as well as affecting kidney and liver function. The ability of TSST-1 to cause both localized or systemic symptoms in this manner dictates that it must be able to transverse the epithelial barrier. At present, the mechanism by which it is able to do this is unknown, but could include either passive diffusion, or cellular receptors. The ability of TSST-1 to cross epithelial barriers, if it exists, would make it unique among the SAgs. The use of a cellular receptor would require some kind of specific interaction with TSST-1, and therefore a receptor binding site. Structural evidence for such a binding site on TSST-1 is limited, and is confined to the observable structural differences between TSST-1 and the other SAgs. These differences include the lack of an α-helix in the C-terminal domain, the long N-terminal extension, and the absence of a disulfide loop. TSST-1 also has unique patches of hydrophobic and neutral residues on the front and rear of the β-barrel at the N-terminal domain. It is possible that a combination of these features could produce a receptor specific binding site in order for TSST-1 to traverse epithelial cells and allow systemic shock.

6.2 Staphylococcal Food Poisoning

Staphylococcus aureus is the leading cause of microbial food-borne disease in the world. Symptoms of staphylococcal food poisoning (SFP) include abdominal pain, nausea, vomiting, and diarrhea and are usually experienced within 2 to 6 h of ingestion. Little is known about how the physical properties of SAgs relate to the symptoms of emesis and diarrhea. It has been suggested that the symptoms of food poisoning are a result of high cytokine levels. Indeed, the side effects of IL-2 therapy often mimic those of SFP. More recent work indicates that the emetic properties of these toxins are not completely correlated with their superantigenicity. Several areas in the N-terminal region of SEA have been identified that are important for both emetic and superantigenic function. Carboxymethylation of histidine residues in SEB was found to abrogate emetic activity but still induce T-cell proliferation, indicating that the two activities are separate. The areas of the toxin that induce emesis and diarrhea are also thought to be

separate, because intravenous administration of a C-terminal fragment of SEC1 was found to induce diarrhea but not emesis in primates. The disulfide loop is considered to be one of the main regions responsible for emesis. Investigation has shown that the disulfide bond itself is not an absolute requirement for emetic activity, but the conformation within or adjacent to the loop is important for emesis. Further analysis of these toxins is required in order to assess the contribution of certain amino acids to the overall pathology of SAgs in the context of the diseases they cause.

6.3 Autoimmunity

The action of SAgs as immunomodulators has identified them as possible candidates for autoimmune disease. Since they do not discriminate between autoimmune and normal lymphocytes, SAgs would be able to stimulate autoreactive T cells, both locally and systemically. Although SAgs have not yet been directly implicated in human autoimmune disease, there is evidence to suggest their involvement. For example, SED-reactive T-cell lines have been shown to stimulate autoantibody production by B cells. Superantigens have been shown to be associated with a growing list of autoimmune conditions, including multiple sclerosis, arthritis, diabetes, and Kawasaki disease. Whether SAgs act as the initial trigger for autoimmunity or as external factors changing an ordinary controllable illness into an autoimmune condition for susceptible individuals, both remain possible, and clearly further work is needed to assess the roles of these toxins in this type of disease.

See also Autoantibodies and Autoimmunity; Immune Defence, Cell Mediated; Immunoassays; Immunologic Memory; Immunology; Innate Immunity; Transcription Factor NF-κB: Function, Structure, Regulation, Pathways, and Applications.

Bibliography

Books and Reviews

Baker, M., Acharya, K.R. (2002) Superantigens: Structure Function and Diversity, in: Krakauer, T. (Ed.) *Superantigen Protocols*, Humana Press Totowa, New Jersey.

Krakauer, T. (Ed.) (2002) *Superantigen Protocols*, Humana press, Totowa, New Jersey.

Papageorgiou, A.C., Acharya, K.R. (2000) Microbial superantigens: from structure to function, *Trends Microbiol.* **8**, 369–375.

Petersson, K., Forsberg, G., Walse, B. (2004) Interplay between superantigens and immunoreceptors, *Scand. J. Immunol.* **59**, 345–355.

Thibodeau, J., Sekaly, R.P. (Eds.) (1995) *Superantigens: Structure Function and Therapeutic Potential*, R.G. Landes company, Austin, Texas.

Primary Literature

Arcus, V.L., Langley, R., Proft, T., Fraser, J.D., Baker, E.N. (2002) The three-dimensional structure of a superantigen-like protein, SET3, from a pathogenicity island of the *Staphylococcus aureus* genome, *J. Biol. Chem.* **277**, 32274–32281.

Cole, B.C. (1991) The immunobiology of *Mycoplasma arthritidis* and its superantigen MAM, *Curr. Top. Microbiol. Immunol.* **174**, 107–119.

Deringer, J.R., Ely, R.J., Stauffacher, C.V., Bohach, G.A. (1996) Subtype-specific interactions of type C staphylococcal enterotoxins

with the T-cell receptor, *Mol. Microbiol.* **22**, 523–534.

Donadini, R., Liew, C.W., Kwan, A.H., Mackay, J.P., Fields, B.A. (2004) Crystal and solution structures of a superantigen from *Yersinia pseudotuberculosis* reveal a jelly-roll fold, *Structure (Camb.)* **12**, 145–156.

Fields, B.A., Malchiodi, E.L., Li, H., Ysern, X., Stauffacher, C.V., Schlievert, P.M., Karjalainen, K., Mariuzza, R.A. (1996) Crystal structure of a T-cell receptor β-chain complexed with a superantigen, *Nature* **384**, 188–192.

Germain, R.N. (1997) T-cell signaling: the importance of receptor clustering, *Curr. Biol.* **7**, 640–644.

Jardetzky, T.S., Brown, J.H., Gorga, J.C., Stern, L.J., Urban, R.G., Chi, Y.I., Stauffacher, C., Strominger, J.L., Wiley, D.C. (1994) Three-dimensional structure of a human class II histocompatibility molecule complexed with superantigen, *Nature* **368**, 711–718.

Kim, J., Urban, R.G., Strominger, J.L., Wiley, D.C. (1994) Toxic shock syndrome toxin-1 complexed with a class II major histocompatibility molecule HLA-DR1, *Science* **266**, 1870–1874.

Kline, J.B., Collins, C.M. (1997) Analysis of the interaction between the bacterial superantigen streptococcal pyrogenic exotoxin A (SpeA) and the human T-cell receptor, *Mol. Microbiol.* **24**, 191–202.

Lavoie, P.M., McGrath, H., Shoukry, N.H., Cazenave, P.A., Sekaly, R.P., Thibodeau, J. (2001) Quantitative relationship between MHC class II-superantigen complexes and the balance of T cell activation versus death, *J. Immunol.* **166**, 7229–7237.

Leder, L., Llera, A., Lavoie, P.M., Lebedeva, M.I., Li, H., Sekaly, R.P., Bohach, G.A., Gahr, P.J., Schlievert, P.M., Karjalainen, K., Mariuzza, R.A. (1998) A mutational analysis of the binding of staphylococcal enterotoxins B and C3 to the T cell receptor β chain and major histocompatibility complex class II, *J. Exp. Med.* **187**, 823–833.

Li, H., Llera, A., Mariuzza, R.A. (1998a) Structure-function studies of T-cell receptor-superantigen interactions, *Immunol. Rev.* **163**, 177–186.

Li, H., Llera, A., Tsuchiya, D., Leder, L., Ysern, X., Schlievert, P.M., Karjalainen, K., Mariuzza, R.A. (1998b) Three-dimensional structure of the complex between a T cell receptor β chain and the superantigen staphylococcal enterotoxin B, *Immunity* **9**, 807–816.

Li, Y., Li, H., Dimasi, N., McCormick, J.K., Martin, R., Schuck, P., Schlievert, P.M., Mariuzza, R.A. (2001) Crystal structure of a superantigen bound to the high-affinity, zinc-dependent site on MHC class II, *Immunity* **14**, 93–104.

Lina, G., Bohach, G.A., Nair, S.P., Hiramatsu, K., Jouvin-Marche, E., Mariuzza, R. (2004) Standard nomenclature for the superantigens expressed by Staphylococcus: international nomenclature committee for staphylococcal superantigens, *J. Infect. Dis.* **189**, 2334–2336.

Marrack, P., Kappler, J. (1990) The staphylococcal enterotoxins and their relatives, *Science* **248**, 705–711.

Petersson, K., Hakansson, M., Nilsson, H., Forsberg, G., Svensson, L.A., Liljas, A., Walse, B. (2001) Crystal structure of a superantigen bound to MHC class II displays zinc and peptide dependence, *EMBO. J.* **20**, 3306–3312.

Petersson, K., Pettersson, H., Skartved, N.J., Walse, B., Forsberg, G. (2003) Staphylococcal enterotoxin H induces V α-specific expansion of T cells, *J Immunol.* **170**, 4148–4154.

Roussel, A., Anderson, B.F., Baker, H.M., Fraser, J.D., Baker, E.N. (1997) Crystal structure of the streptococcal superantigen SPE-C: dimerization and zinc binding suggest a novel mode of interaction with MHC class II molecules, *Nat. Struct. Biol.* **4**, 635–643.

Schlievert, P.M., Jablonski, L.M., Roggiani, M., Sadler, I., Callantine, S., Mitchell, D.T., Ohlendorf, D.H., Bohach, G.A. (2000) Pyrogenic toxin superantigen site specificity in toxic shock syndrome and food poisoning in animals, *Infect. Immun.* **68**, 3630–3634.

Sundberg, E., Jardetzky, T.S. (1999) Structural basis for HLA-DQ binding by the streptococcal superantigen SSA., *Nat. Struct. Biol.* **6**, 123–129.

Sundberg, E.J., Li, H., Llera, A.S., McCormick, J.K., Tormo, J., Schlievert, P.M., Karjalainen, K., Mariuzza, R.A. (2002) Structures of two streptococcal superantigens bound to TCR β chains reveal diversity in the architecture of T cell signaling complexes, *Structure (Camb.)* **10**, 687–699.

Swaminathan, S., Furey, W., Pletcher, J., Sax, M. (1992) Crystal structure of staphylococcal

enterotoxin B, a superantigen., *Nature* **359**, 801–806.

Swaminathan, S., Furey, W., Pletcher, J., Sax, M. (1995) Residues defining Vβ specificity in staphylococcal enterotoxins, *Nat. Struct. Biol.* **2**, 680–686.

Woodland, D.L., Wen, R., Blackman, M.A. (1997) Why do superantigens care about peptides?, *Immunol. Today* **18**, 18–22.

Zhao, Y., Li, Z., Drozd, S.J., Guo, Y., Mourad, W., Li, H. (2004) Crystal structure of Mycoplasma arthritidis mitogen complexed with HLA-DR1 reveals a novel superantigen fold and a dimerized superantigen-MHC complex, *Structure (Camb.)* **12**, 277–288.

Supermolecular Protein Complexes

Jianpeng Ma
Baylor College of Medicine and Rice University, Houston, TX, USA

1 Architectural and Motional Characteristics of Supramolecular Complexes 655

2 Simulating Supramolecular Complexes at Multiresolution and Multilength Scales 656
2.1 State-of-the-art Simulation Methods 656
2.2 Biological Applications 658
2.2.1 Molecular Chaperonin GroEL 658
2.2.2 Fatty Acid Synthase (FAS) 660
2.2.3 F-actin Filaments 663

3 Refining Supramolecular Complexes at Multiresolution and Multilength Scales 665
3.1 Normal-mode-assisted Cryo-EM Structural Refinement 665
3.2 Fiber Diffraction Refinement Based on Long-range Normal Modes 666

4 Discussions and Future Perspectives 667

Acknowledgment 668

Bibliography 668
Books and Reviews 668
Primary Literature 669

Keywords

Coarse-grained Simulation
A particular type of simulation strategy that is designed to reduce the complexity of calculation so as to allow researchers to capture essential properties of dynamic

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

systems, such as low-frequency motions, by filtering out the nonessential components such as high-frequency motions. This type of study is very useful when coping with very large systems, especially when the computers are not powerful enough to sample all the biological timescales.

Cryogenic Electron Microscopy
A low-resolution experimental technique that measures structures of macromolecules. The resolution of the technique is usually not high enough for establishing atomic coordinates, but it can provide rough shapes of molecules. It is an effective way of getting the structures of supramolecular complexes that are usually hard to form diffracting crystals for X-ray crystallography.

Fiber Diffraction
A low-resolution experimental technique that measures structures of long filamentous fibers. The technique made a critical contribution in elucidating double-helical structure of DNA in the early stage of the field of molecular biology. In recent years, aided by the modern synchrotron radiation, the method is gaining momentum in studying many biologically important fibers such as F-actin filaments.

Large Conformational Change
Biological macromolecules fulfill their functions in a dynamic fashion. Motion is an integral part of function. Large conformational changes are large structural rearrangements that alter physical properties of protein complexes such as their affinity to their ligands or their ability to catalyze biochemical reactions. These changes can take place at different levels of structural components, which can be a localized loop, a single domain, or entire quaternary arrangement of the complex.

Molecular Dynamics Simulation
A computer simulation technique that maps the spatial and temporal trajectories of protein movement by numerically propagating Newton's equations of motion for all atoms in biologically relevant environment. It is widely used for studying dynamics of biological molecules.

Normal Mode Analysis
A simulation algorithm that decomposes motions of proteins into a set of orthogonal harmonic modes, which serves as a complete basis set from which any arbitrary vibration can be expressed as a linear combination. Mathematically, it is an eigenvalue problem. Usually, low-frequency modes are more biologically relevant. In recent years, normal mode analysis and other related algorithms have been shown to be very effective in modeling large-scale conformational changes, especially those of supramolecular protein complexes.

Structural Refinement
A computational procedure that optimizes the structural model against experimentally measured data by, for example, minimizing certain target functions that define the difference between the two.

Supramolecular Protein Complexes
Large protein assemblies that are formed by multiple subunits, or polypeptides, usually arranged in a unique quaternary architectural style to carry out specific biological functions. Such complexes are frequently seen in higher species due to higher level of complexity in regulatory mechanisms. The particular architecture is usually designed to accommodate the required structural motions involved such as those in allosteric proteins. Sometimes, supramolecular complexes are also referred to as *molecular machines*.

Many important biological functions are carried out by supramolecular protein complexes, which are characteristically formed in a hierarchical arrangement. For those complexes, just like large man-made structures, flexible motions are essential for maintaining robustness and for fulfilling their functions. To study the motion, various computer simulation techniques have played an important role in elucidating spatial and temporal trajectories of proteins. Furthermore, in recent years, computational modeling has also dramatically contributed to the improvement of structural models in refinement. In this chapter, to give a representative illustration of the subject, the author outlines several examples taken from recent studies of supramolecular protein complexes.

1
Architectural and Motional Characteristics of Supramolecular Complexes

Just like all kinds of natural or man-made structures, proteins are built in a modular fashion. The smallest structural modules for proteins are likely the individual globular domains that may fold relatively independently. Many proteins have more than one domain. In large assemblies, multiple subunits are organized in a hierarchical way to form the complexes. The size of complexes varies dramatically ranging from relatively simple dimers to huge multicomponent complexes such as viral particles and cytoskeleton filaments. Functionally, structural complexity is often related to the level of sophistication of regulation mechanisms in different organisms. In lower species, a particular biological function is carried out by smaller proteins, while, in higher species, it may be carried out by larger complexes with increased complexity in regulation.

There are evolutionary advantages for building large complexes in a modular fashion. By doing so, nature only needs to ensure the folding ability of relatively small sets of modules (often small domains) that act like building blocks, and then to optimize the outer shape of those modules to fit into the assemblies. In fact, using shape as an additional parameter for diversifying complex structures is a way by which nature achieves its vast architectural complexity via only 20 types of amino acids. Moreover, nature follows only a few design principles that are also commonly employed in man-made structures. For

example, shape complementarity among different protein subunits is common for ensuring the overall integrity of complexes similar to what is observed in the domes of ancient bridges and cathedrals.

Another important issue regarding complex structures is flexibility. Similar to large man-made structures, flexibility plays an important role in maintaining robustness of complexes and in fulfilling their functions. Many large molecular complexes behave just like man-made machines, for which motion is an integral part of functions. In general, a particular type of architecture is designed not only to ensure structural integrity, but also to optimize the coupling of motions among various interacting parts so as to transmit the actions from one part to another, for example, allosteric mechanisms.

As a consequence of the hierarchical nature of complex structures, typical molecular motions in supramolecular complexes are hinged domain movements. This is likely because these kinds of movements are energetically easier to move a whole chunk of structure around some flexible hinges so as to avoid substantial disturbance of the inner folding cores of each building block, which might have been evolved separately. In that sense, the conformational dynamics of the supramolecular complexes is somehow simpler in terms of their topological changes (although very large in scale) in comparison with folding in which all intricate side-chain placement needs to be satisfied. For the same reason, the motions of supramolecular complexes often follow some universally similar rules. Understanding a few representatives will enormously enlighten the understanding of many of them.

The best way to understand the role of protein conformational motions is to study the pathways of these motions and systematically map out the intermediate states they experience along the pathways. However, protein conformational motions are relatively fast, and the intermediates along the pathways are transient. They are usually very difficult to be captured by experimental methods. To overcome such a difficulty, state-of-the-art computer simulation techniques have made it possible to study the transient motional features in a reasonably accurate fashion. In particular, the power of computational investigation has been repeatedly demonstrated in numerous cases when the methods are combined with the experimental data. Outstanding contributions have been made for understanding dynamic functions of supramolecular complexes, and particularly to assist structural refinement for experimental measurement.

In the following sections, we will outline several examples taken from recent computational studies of supramolecular complexes. The purpose is not a comprehensive review; rather, it is a representative illustration to give readers a general overview of the subject.

2 Simulating Supramolecular Complexes at Multiresolution and Multilength Scales

2.1 State-of-the-art Simulation Methods

A large portion of biomolecular functions involves structural dynamics, which occurs in a wide range of length scales, for example, from vibrations of small chemical bonds to global conformational changes of supramolecular complexes. Meanwhile, experimental structural data are only available in varying resolution scales. In some

cases, one has high-resolution atomic coordinates by methods such as X-ray crystallography, but in others, one only has low- to intermediate-resolution density maps by, for example, cryoelectron microscopy (cryo-EM). To cope with such a wide-range scale distribution of biological dynamics, computer simulations have been playing an increasingly important role in providing a spatial and temporal framework for interpreting experimental observations.

Molecular dynamics (MD) simulation is a widely adopted method for studying dynamics-related properties of proteins. The essence of the method is to numerically propagate Newtonian equations of motion for all atoms in time so as to generate a trajectory in space from which all the dynamic properties can be computed. For studying supramolecular complexes and their associated large-scale conformational changes, certain types of modifications to standard MD simulation have been developed. One of the typical methods is the targeted molecular dynamics (TMD) method. Together with a few related versions such as steered molecular dynamics (SMD), they are widely used in modeling large-scale conformational changes. The basic idea for all these methods is the constrained simulation. Usually, the input of those simulations would be the two static end-point structures of a conformational change determined by X-ray crystallography. Then the dynamic pathways between the two end-point structures can be simulated by slowly "guiding" the simulation from one state to the other. The methods are approximate, but they have been demonstrated to be very effective in delivering a trajectory from which biochemical data can be rationalized and, more importantly, further experimental probes can be designed to verify the suggested mechanisms.

Another effective way of analyzing molecular motions is harmonic modal analyses. The most commonly used ones are normal mode analysis (NMA), quasi-harmonic analysis and its related version of essential dynamics. Mathematically, modal analyses are eigenvalue problems. The modes form a complete basis set, by which any arbitrary molecular deformational motions can be expressed as a linear combination. In general, modal analyses are particularly effective where elastic (harmonic) properties of molecules are concerned. The underlying principle for modal analysis is to study biomolecular functions by filtering out the less important high-frequency motions and focusing on the dominating low-frequency components.

Numerous examples have demonstrated that functionally important transition pathways of biomolecules, especially those of supramolecular complexes, often follow the trajectories of one or a few low-frequency normal modes. An important fact that emerged from those studies is that the protein structures have evolved in such a way that their intrinsic structural flexibility, as manifested in normal modes, facilitates the functionally important conformational variations. Moreover, in recent years, several important new algorithms have been developed. Among them, particularly important ones are the coarse-grained NMA methods. They have enabled one to simulate deformational motions of supramolecular complexes at dramatically extended resolution and length scales.

Most importantly, NMA methods have enabled structural refinement in several structural biology fields including fiber diffraction, cryo-EM (cryo-electron microscopy), and X-ray crystallography.

This is because the biological specimens intrinsically contain the information of molecular motions. Even in the most ordered crystal lattices, certain mobile components of molecules still present certain degrees of structural and motional anisotropy, which are often functionally related. Powerful computational methods are necessary to take those motional features into account in structural refinement. For supramolecular complexes, another difficulty comes from the lack of enough independent experimental observations that could lead to severe overfitting if too many adjustable parameters are used for refinement. Using harmonic modal analysis to aid refinement can significantly ease such a problem.

2.2 Biological Applications

In this section, we will briefly outline several typical examples for modeling large-scale motions of supramolecular complexes. The first one is molecular chaperonin GroEL with available atomic coordinates. Another one is human fatty acid synthase (FAS) based on low-resolution electron density maps from cryo-EM measurements. And the final one is F-actin filament based on a model established from fiber diffraction. Those examples cover a wide range of length scale and resolution scale. Other examples in literature include ribosome, viruses, and so on.

2.2.1 Molecular Chaperonin GroEL

Chaperonin GroEL is a huge molecular complex that facilitates protein folding in bacterial cells. Its architecture consists of two cylindrical rings with seven identical subunits arranged in a back-to-back fashion with a dyad symmetry (see Fig. 1). In the middle of the complex, there is a central channel that can drastically alter its size as a consequence of large allosteric conformational changes of the complex driven by the binding of ATPs, the heptameric cochaperonin GroES, and nonnative polypeptides. Essentially, the central cavity plays a role as a cage in one of the two rings, the so-called "Anfinsen Cage," for facilitating polypeptide folding in a protected environment.

The static X-ray structures of different conformational states of the GroEL complex elucidate the end points of conformational changes. But they do not provide information of pathways for the transition. This information has to be supplemented by computational simulations. Two major types of simulation methods have been applied. One is NMA and the other is TMD simulation.

In NMA of GroEL, the study revealed intrinsic flexibility of each individual subunit that plays a vital role in its structural response to ATP binding. It was shown that two dominating low-frequency modes could capture the essential features of ATP-driven conformational changes, which include the downward motion of the intermediate domain and upward motion of the apical domain. NMA also captured the twisting motion of the apical domain that was later shown to be essential for facilitating protein folding.

The results of NMA were later corroborated by a more detailed TMD simulation performed from the close state to the open state. The TMD trajectory revealed the sequence of events in time along the transition pathway (which was missing in NMA) (see Fig. 2). More detailed atomic interactions between key residues were also observed. The results suggested that the intermediate domain, especially the M helix, has to come down first to close the binding site of ATP, which is important for hydrolysis. Then the large-scale upward motion of

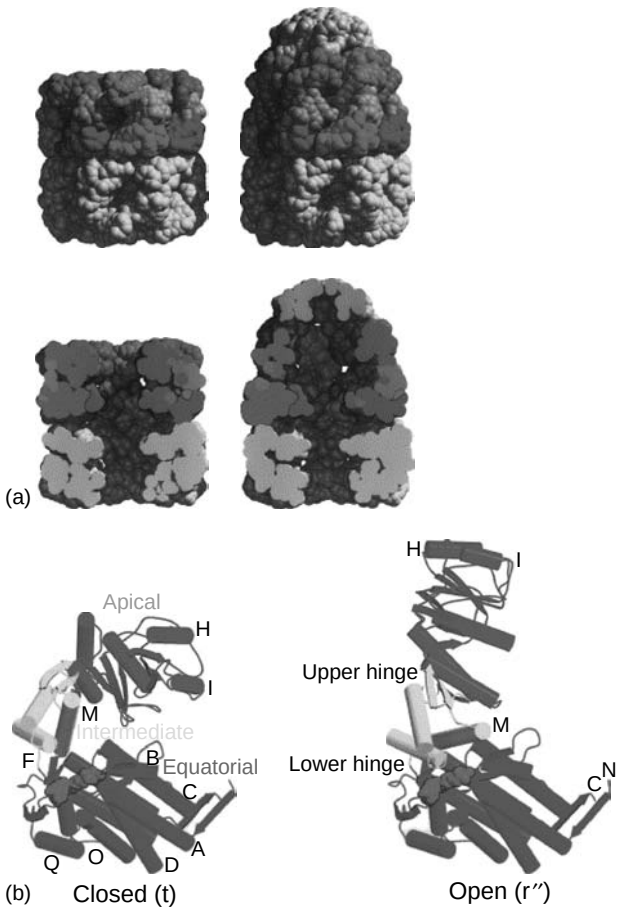

Fig. 1 The overall architecture and conformational change of GroEL. The closed ATPγS crystal structure, and the GroEL–GroES–$(ADP)_7$ structure are shown from left to right. (a) van der Waals space-filling models (6 Å spheres around C atoms). Upper panels show the overall dimension; lower panels show the interior of the complex by a cut-through view. Different colors are used to indicate the subunits in the upper ring. The domains are distinguished by shading: apical, medium hue; intermediate, light hue; equatorial, dark hue. The lower ring is uniformly in yellow and GroES is uniformly in gray.
(b) Representations of a subunit of GroEL corresponding to the structures is shown in (a). The color code is: apical, green; intermediate, yellow; equatorial, red. The nucleotide is shown as a blue space-filling model. Some of the key secondary structural elements are marked. The view of the subunits is roughly from inside out of the GroEL central cavity. This figure was modified from Ma, J., Sigler, P.B., Xu, Z., Karplus, M. (2000) A dynamic model for the allosteric mechanism of GroEL, *J. Mol. Biol.* **302**, 303. (See color plate p. xli.)

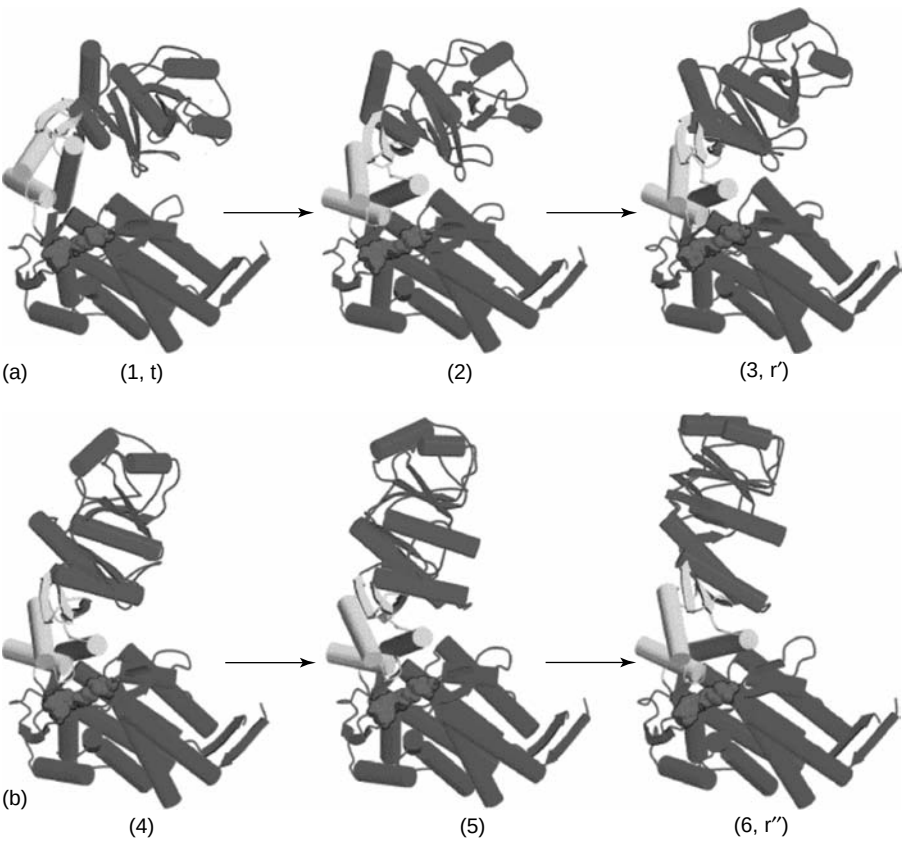

Fig. 2 Calculated transition path. (a) Schematic representation of the two-stage transition. (b) A set of structures on the path: (1) through (3) correspond to the first stages associated with nucleotide binding, that is, the t to r′ transition; (4) through (6) correspond to the second stage involving GroES binding, that is, the r′ to r″ transition. Structures (1) and (6) labeled t and r″, respectively, are the X-ray structures. The other structures are from the TMD simulation. The equatorial domain is superimposed as a reference in the various structures and so it has an essentially fixed orientation. The view of the subunit is the same as that of Fig. 1(b) and the twist of the apical domains, especially in the second stage, is toward the reader. This figure was adapted from Ma, J., Sigler, P.B., Xu, Z., Karplus, M. (2000) A dynamic model for the allosteric mechanism of GroEL, *J. Mol. Biol.* **302**, 303.

the apical domain follows. The transition is completed by a nearly 90° twist of the apical domain so that the chemical nature of the inner wall of GroEL central cage changes from hydrophobic in the close state to hydrophilic in the open state, which plays a determining role for GroEL-assisted protein folding. In addition to the transition pathways, the computational results also provided rational evidence for the allosteric mechanisms of GroEL, which has positive cooperativity within a ring and negative cooperativity between two rings.

2.2.2 Fatty Acid Synthase (FAS)

FAS catalyzes the synthesis of long-chain saturated fatty acids – palmitate, from acetyl-CoA, malonyl-CoA, and NADPH. In

animals, including humans, the catalytic centers of FAS involve an acyl-carrier protein (ACP) and seven catalytic enzymes. All of them are covalently connected together in a single polypeptide chain encoded by a single gene. The native FAS is a dimer of two identical subunits and each monomer has a molecular weight of 272 000 Da. The homodimer, essential to the functions, is arranged in a head-to-tail fashion forming two active catalytic centers. The 3D structure of FAS was determined by cryo-EM technique to about 19-Å resolution (see Fig. 3).

FAS is a multidomain complex with a very unique nonglobular quaternary arrangement. Much like many other multidomain supramolecular complexes, structural flexibility is one of the central issues in the function of FAS. In particular, domain movements of FAS are likely to be essential for facilitating the multistep catalytic reactions of the synthesis of palmitate. Although the traditional textbook view of FAS mechanism is such that the long and flexible (20-Å maximal length) phosphopantetheinyl moiety of ACP carries the substrate from one active site to the other, there are growing lines of biochemical evidence showing that the flexibility of the enzyme may also be involved in shuffling the substrates.

To computationally investigate the flexibility, the NMA method based on cryo-EM density map was applied to the 19-Å density maps of FAS to compute the patterns of thermal deformational motions. The most significant simulation results were that there are two types of major flexible hinges in the dimeric structure of FAS. One is an inter-subunit hinge located at the small inter-subunit connection (Fig. 4a,b). The other is an intra-subunit hinge located at the relatively thin neck region (Fig. 4c,d). Thus, a FAS dimer has one inter-subunit hinge between two subunits and two intra-subunit hinges from two subunits.

The patterns of the low-frequency deformational modes demonstrated that the domain movements of FAS around these hinges are primarily rigid-body movements. In the very lowest-frequency mode (mode 7), the motion is primarily around the inter-subunit hinge (Fig. 4a). Each subunit moves as a rigid body and the dimer undergoes a see-saw-like bending around the twofold symmetry axis in the pseudoplane of the dimer. Biologically, one may regard such a motion to be the one that can transiently alter the width of the cleft by, for example, bringing domains in the opposite subunit closer, which would be advantageous in facilitating the elongation of fatty acyl chain between the cysteine sulfhydryl

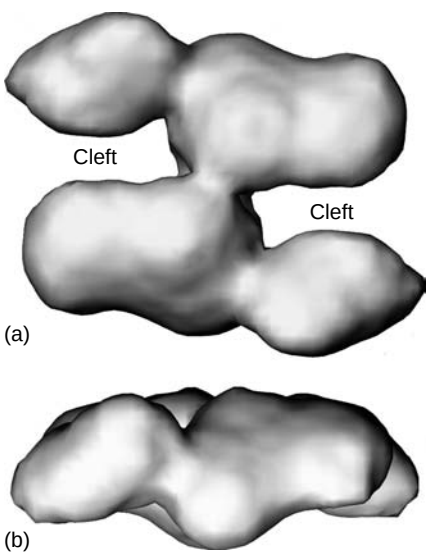

Fig. 3 The electron density maps of FAS determined by cryo-EM at 19-Å resolution. Only the one with a twofold symmetry imposed in the reconstruction is shown. The top one is a top view and the bottom one is a side view. The clefts between subunits are labeled.

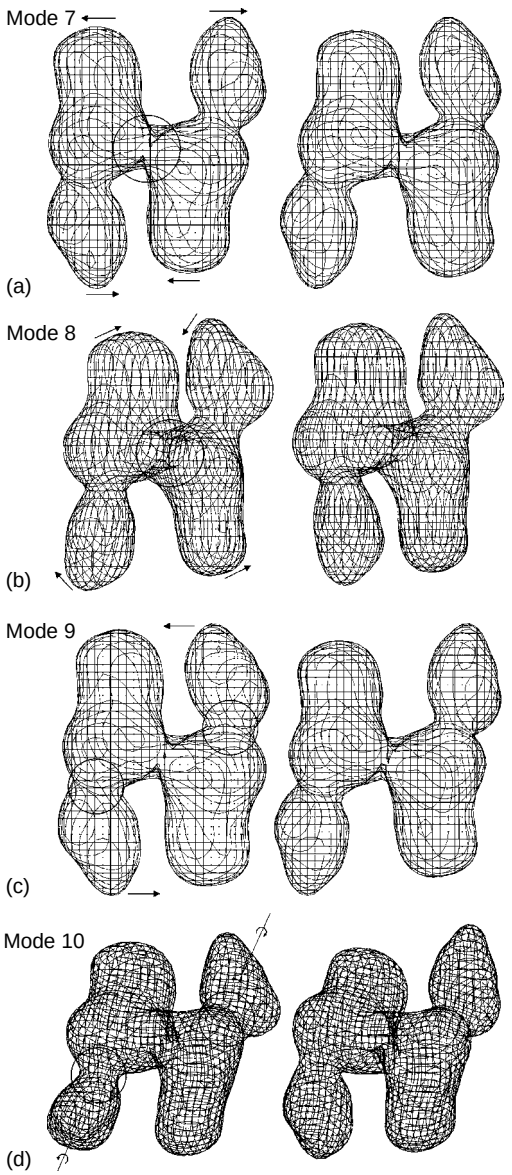

Fig. 4 The motional patterns of the first four lowest-frequency deformational modes: mode 7(a), 8(b), 9(c), and 10(d). Note, the first six modes are zero modes corresponding to global translation and rotation of the entire molecule. For each mode, two opposite conformations (left and right) during harmonic vibration are shown to illustrate the direction of the motion. The amplitude of the motion was arbitrarily chosen for visual clarity. The arrows are used to indicate the directions of the motions. The larger circles in (a) and (b) indicate the inter-subunit hinge and the smaller circles in (c) and (d) indicate the intra-subunit hinges. The dotted lines in (d) indicate the longest axes of the subunits. This figure was adapted from Ming, D., Kong, Y., Wakil, S.J., Brink, J., Ma, J. (2002) Domain movements in human fatty acid synthase by quantized elastic deformational model, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7895.

group of condensing enzyme and the phosphopantetheine sulfhydryl group of ACP. Early biochemical study showed that the two sulfhydryl groups from two opposite sides of the catalytic reaction centers could be cross-linked by a disulfide bond quickly. This fact indicates that the two sides are in direct proximity – an event that can presumably take place via large domain movements similar to what is described in normal modes.

Besides the very lowest-frequency mode, other low-frequency modes also revealed important domain movements.

For example, the second lowest-frequency mode (mode 8) describes an out-of-plane twisting between two rigid subunits around the inter-subunit hinge (nearly perpendicular to mode 7) (Fig. 4b). It is possible that this type of motion could bring together the subunits in two opposite catalytic reaction centers. The linear combination of modes 7 and 8 would facilitate the two subunits large-scale motions around the inter-subunit hinge. Similarly, the third (mode 9) (Fig. 4c) and fourth (mode 10) (Fig. 4d) lowest-frequency modes present the motions around the two intra-subunit hinges.

All together, the conformational motion of FAS can be regarded as the combination of these fundamental deformational modes. Although the dimeric FAS molecule has a chemically symmetric structure and the calculation was performed on the structure with a twofold symmetry imposed, the NMA clearly demonstrate that FAS is a highly flexible and heterogeneous molecule. Large-scale domain movements can take place in various directions, which results in a wide range of conformation of FAS. Biologically, those domain movements could be important for facilitating and regulating the entire synthesis by coordinating the communication between components of the structure, for instance, altering the distance between various active sites inside the catalytic reaction center. Finally, another important usage of the heterogeneous conformational distribution revealed in NMA is to improve the refinement in the cryo-EM reconstruction (see Sect. 3.1).

2.2.3 F-actin Filaments

Actins are filamentous protein complexes that are important for cell transport, cytoskeletal support, and contractile events in almost all eukaryotic cells. They exist in two forms. In nonionic solutions, the dominant form is a monomeric G-actin of 41 kD. In the presence of salt, G-actin monomers polymerize into a double-stranded helical polymer called *F-actin*, which is the functional form of actin. Structural information on several forms of actin includes the atomic coordinates of G-actin (Fig. 5a) and the Holmes model of F-actin filaments established from fiber diffraction. In the Holmes model, F-actin filaments appear to be formed by two right-handed long-pitch helical strands that twist around each other with a rise of 27.5 Å and

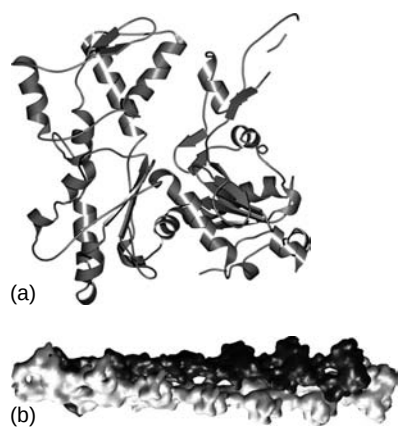

Fig. 5 Structures of actins. (a) Structure of G-actin monomer that contains 375 amino acids. (b) The Holmes model for a 13-subunit repeat of F-actin filament established from fiber diffraction with two helical strands differently colored. To illustrate the overall shape, the atomic coordinates were blurred to 8-Å resolution. This figure was adapted from Ming, D., Kong, Y., Wu, Y., Ma, J. (2003) Substructure synthesis method for simulating large molecular complexes, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 104.

a rotation angle of -166.15° per monomer around the filament axis (Fig. 5b). The minimum repeat of the double-stranded helix contains 13 monomers (the 13-subunit repeat) and the total length of F-actin could be several microns or longer.

In recent years, a growing body of experimental evidence indicates that a large portion of dynamic properties of F-actin filaments, such as those modulating operations of molecular motors, is directly related to their intrinsic elasticity. A deep understanding of the mechanical elastic properties of F-actin filaments is, therefore, vital to elucidating their roles in motile and cytoskeletal systems.

Since the size of F-actin filament is way beyond conventional computational methods, a new method, substructure synthesis method (SSM), was developed to extend the computational capacity in simulating vibrational motions of very long filamentous systems. In SSM, a given molecular complex structure is regarded as an assemblage of substructures that can be chosen arbitrarily, and sometimes quite naturally, for example, as domains, subunits, or large segments of biomolecular complexes. One first computes a set of substructure modes by methods such as NMA to represent the motions of substructure, then, the motions of the entire structure are generated by linking the substructures together using a set of constraints to enforce geometric compatibility at the interfaces of adjacent substructures. For periodically repeating systems, a hierarchical synthesis scheme (HSS) can be employed in combination with SSM to rapidly achieve the desired complexity. Computationally, the efficiency of SSM comes from avoiding the full eigenvalue problem for the entire structure.

The procedure of SSM–HSS was applied to the synthesis of the vibrational modes of F-actin filaments of 4.6 μm from the modes of a single 13-subunit repeat. Figure 6 shows a few typical lowest-frequency bending modes. All of these modes move in a plane, which is a typical feature for a homogeneous elastic rod. Similar behavior was also observed for twisting and stretching modes. These results clearly suggest that one can use SSM–HSS to scale-up the microscopic dynamic information, obtained from atomistic simulations, to a wide range of macroscopic length scale.

In summary, the computationally synthesized vibrational modes for F-actin filaments by SSM–HSS provide an orthonormalized basis set by which any arbitrary elastic deformation of the filaments

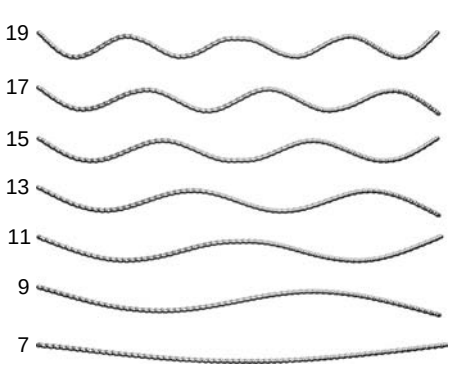

Fig. 6 The motional patterns of several lowest-frequency bending modes for the 4.6 – μm F-actin filament calculated by SSM–HSS. The indices of the modes are marked. The even indices are the ones that degenerate to the displayed one. The 7th mode is the lowest-frequency mode. This figure was adopted from Ming, D., Kong, Y., Wu, Y., Ma, J. (2003) Simulation of F-actin filaments of several microns, *Biophys. J.* **85**, 27.

can then be expressed as a linear combination. These modes can also be used to characterize the motions of F-actin filaments at a continuously varying length scale, that is, from the extremely localized high-frequency motions all the way to global elastic deformational motions. The bending motions are likely to dominate the motions of long filaments, while other types of deformations such as twisting and stretching are likely to be important to short-range local deformations of the filaments. Finally, the methods are equally applicable to other filamentous systems in cells, for example, tubulin, collagen, and DNA. A major application of these modes is in assisting structural refinement for fiber diffraction (see Sect. 3.2).

3 Refining Supramolecular Complexes at Multiresolution and Multilength Scales

One of the most important applications of NMA is in assisting structural refinement for large flexible systems. Since the low-frequency normal modes are collective parameters that describe the motions in a global fashion, the biggest advantage of using normal modes as refinement parameters is that only a very small set of modes are often enough to represent the motional features during refinement. Such an approach drastically cuts down the number of adjustable parameters needed for the refinement. Particularly in cases such as cryo-EM and fiber diffraction, there are not many independent measures due to low resolution, and a smaller number of parameters is always good for avoiding overfitting in refinement.

3.1 Normal-mode-assisted Cryo-EM Structural Refinement

Similar to X-ray crystallography and many other structural determination methods, the structural dynamics within the particle population due to the inherent flexibility of biomolecules can lead to smearing out of the cryo-EM averaging. At the current stage, only extraordinarily flexible systems have a realistic issue with conformational heterogeneity. But, in the near future, once the resolution of cryo-EM reaches a certain level, the conformational heterogeneity will become a major limiting factor for resolution of many systems. FAS is one of the systems that was currently influenced by the conformational heterogeneity in measurement. Although the actual data reached a much higher resolution, the final cryo-EM reconstruction for FAS was only at 19 Å. A computational strategy in conjunction with NMA was developed to separate the heterogeneous data such that structurally homogeneous particle images can be grouped to achieve higher resolution reconstructions within each group.

In the procedure of standard cryo-EM single model refinement, one first generates a set of projections of an initial 3-D model. After that, one aligns each raw particle image against each projection in 2D and groups the image with the projection it is most similar to. Once the alignment is done, average images are taken for the particles grouped with each projection. Such a procedure produces a class average from the raw particle data corresponding to each projection image. Finally, the class averages can be used to construct a new 3-D model, which is used to seed the next iteration of refinement.

In the normal-mode-based refinement of FAS, the initial models were established from the original structure and the two end-point conformers of the first two vibrational modes (modes 7 and 8 shown in Fig. 4a,b) were with an arbitrary amplitude. Totally, there were $n = 5$ initial models for the multirefinement algorithm (Fig. 7). The particles were classified, and in this case, one produced n class averages for each orientation, which could be used to produce n new 3D models to seed the next round of iterative refinement. The procedure also led to n subsets of the raw particle data associated with each of the 3D models. The results clearly showed that the conformational heterogeneity predicted by NMA were real in measured cryo-EM particle population.

Although with some degree of success, the normal-mode-assisted cryo-EM refinement has not become a general method for every system. One must overcome several technical difficulties in order for the method to become a routine structural refinement procedure. But the concept involved in normal-mode-assisted cryo-EM refinement harbors the hope for facilitating high-resolution structural determination by cryo-EM in the future. Twenty or thirty years ago, X-ray crystallography went through the same stage; sooner or later, cryo-EM will also reach a resolution range where molecular motions become a limiting factor for resolution.

3.2 Fiber Diffraction Refinement Based on Long-range Normal Modes

A very important application of long-range normal modes for biological fibers is in assisting refinement of structural models against data from fiber diffraction, which is an effective technique for studying filamentous systems that are ubiquitous in cell. Typical examples range from simple polypeptides, polynucleotides, and polysaccharides, to cytoskeletal filaments and filamentous viruses.

In the field of fiber diffraction, it has been known for a long time that long-range filamentous deformations may have significant contributions to the errors of the refined models. To cope with this problem, the atomic model of F-actin was refined against fiber diffraction data using long-range normal modes as adjustable

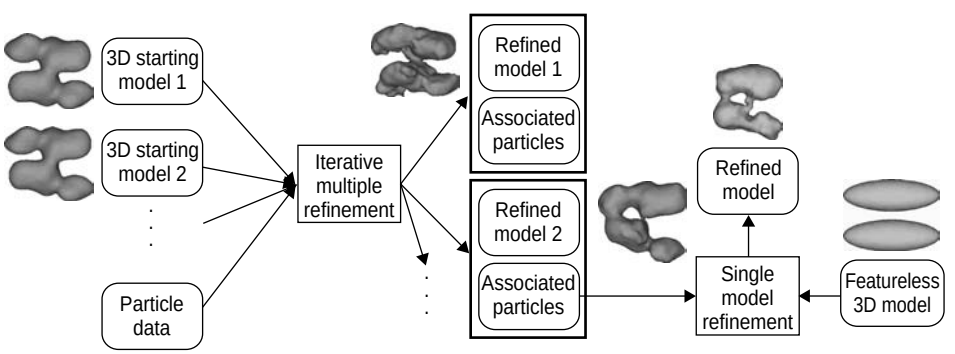

Fig. 7 A flow diagram of the second test used to verify that structurally consistent subpopulations were determined from the raw heterogeneous particle data set. A featureless model (right) was refined against the particle data used to generate refined model 2. The result of this refinement appears to be in the same structural substrate as refined model 2.

parameters to account for the collective long-range filament deformations. Each of the four domains of monomeric G-actin was treated as a rigid body to isolate the contributions of long-range deformations. The improvement in refinement was monitored by the drop in R- and R_{free}-factors using individual or combinations of modes calculated by SSM. It was found that the bending modes make the most significant contributions to the refinement. By including only ~7 to 9 bending modes as adjustable parameters, one was able to decrease R-factor for 2.4% (Fig. 8). This means that, for any fiber diffraction refinement, a substantial portion of errors is due to long-range deformations, in particular, because of the bending of the filaments. Therefore, the effects of these long-range deformations must be included in the refinement in structure determinations by fiber diffraction. Moreover, the advantage of SSM-based normal mode refinement is that it uses a small set of long-range modes as adjustable parameters to achieve a good fit, thus preventing the potential overfitting.

4 Discussions and Future Perspectives

For most of biochemical reactions in living organisms, what makes them different from the same reactions in test tubes is that biochemical reactions in cells are strictly regulated. They take place only when needed, to the required extent, and at the desired rate. Any malfunctioning in regulation of the reactions could result in

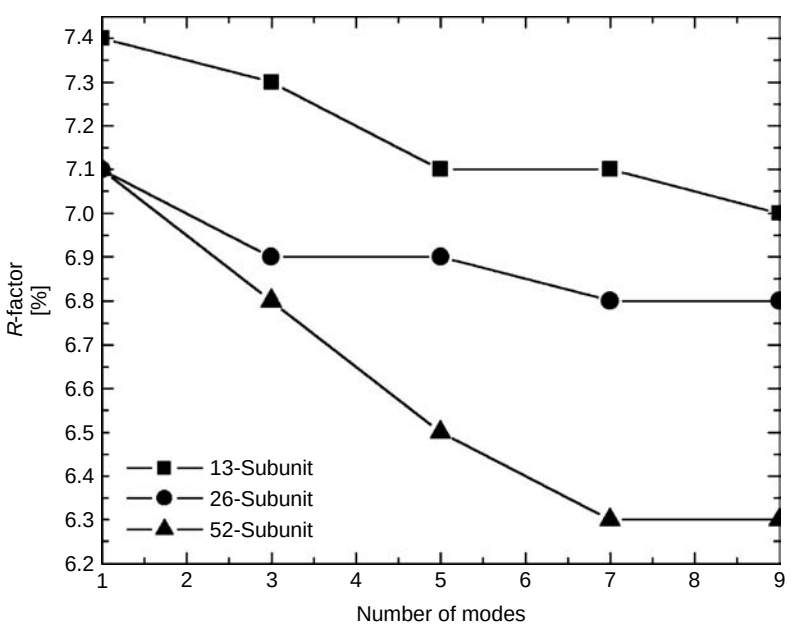

Fig. 8 Refinement results using various lengths of repeats as helical unit cells with a combination of multiple low-frequency normal modes. This figure was adopted from Wu, Y., Ma, J. (2004) Refinement of F-actin model against fiber diffraction data by long-range normal modes, *Biophys. J.* **86**, 116.

damage to the organisms by causes such as diseases. In most systems, especially supramolecular complexes, the regulatory mechanisms almost always involve large-scale conformational changes as the cases in allosteric proteins. Thus, understanding the structural and energetic nature of those conformational changes are vital for understanding biological functions.

As described earlier, the TMD and its related methods have played an important role in simulating large conformational motions. In recent years, the most dramatic success came from the widespread applications of coarse-grained NMA, which lies in the realization that, in modeling dynamics of large complexes, it is feasible to abandon the traditional concepts of protein structure (bonds, angles, dihedrals, etc.) and to treat the structure as an elastic object, whose shape and mass distribution can be obtained from electron density maps of the complexes, at various resolutions, from methods such as X-ray diffraction or cryo-EM. The directionality of the elastic deformational modes of the complexes can then be determined solely on the basis of low-resolution electron density maps. Numerous studies have shown that the patterns of those elastic modes match the biologically relevant conformational changes of proteins.

In the future, various biophysical methods will continue to be developed for studying dynamics of biological complexes. Of particular importance is to combine the computational methods with the experimental methods to iteratively elucidate the structures and mechanisms. This is exemplified by a recent ongoing study of the huge structure of nuclear pore complex by Sali's group at the University of California, San Francisco. A wide range of experimental data derived from all kinds of experimental methods, each providing partial and useful information regarding structure and dynamics, are gathered as constraints for computational modeling. As a consequence, a fairly accurate structural model can be established on the basis of experimental data from which no such model was possible by any conventional means. Here, the author would like to emphasize that the concept of utilizing limited, but diverse, experimental data as constraints to guide the computational modeling will play an increasingly important role in the future study of supramolecular complexes. This is because, for those large systems, no single technique can deliver a complete description of mechanisms.

Acknowledgment

The author acknowledges support from the National Institutes of Health (R01-GM067801) and the National Science Foundation Career Award (MCB-0237796).

See also Informatics (Computational Biology); Intracellular Fatty Acid Binding Proteins and Fatty Acid Transport; Intracellular Fatty Acid Binding Proteins in Metabolic Regulation.

Bibliography

Books and Reviews

Brooks, C.L. III, Karplus, M., Pettitt, B.M. (1988) Proteins: a theoretical perspective of dynamics, structure, and thermodynamics, *Adv. Chem. Phys.* **71**, 1.

Ma, J. (2004) New advances in normal mode analysis of supermolecular complexes and

applications to structural refinement, *Curr. Protein Pept. Sci.* **5**, 119.
Ma, J. (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes, *Structure* **13**, 373.

Primary Literature

Brink, J., Ludtke, S.J., Yang, C.Y., Gu, Z.W., Wakil, S.J., Chiu, W. (2002) Quaternary structure of human fatty acid synthase by electron cryomicroscopy, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 138.
Brink, J., Ludtke, S.J., Kong, Y., Wakil, S.J., Ma, J., Chiu, W. (2004) Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis, *Structure (Camb.)* **12**, 185.
Brooks, B.R., Janezic, D., Karplus, M. (1995) Harmonic analysis of large systems. I. Methodology, *J. Comput. Chem.* **16**, 1522.
Ma, J., Karplus, M. (1998) The allosteric mechanism of the chaperonin GroEL: a dynamic analysis, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8502.
Ma, J., Sigler, P.B., Xu, Z., Karplus, M. (2000) A dynamic model for the allosteric mechanism of GroEL, *J. Mol. Biol.* **302**, 303.
Ming, D., Kong, Y., Wu, Y., Ma, J. (2003) Substructure synthesis method for simulating large molecular complexes, *Proc. Nalt. Acad. Sci. U.S.A.* **100**, 104.
Ming, D., Kong, Y., Wu, Y., Ma, J. (2003) Simulation of F-actin filaments of several microns, *Biophys. J.* **85**, 27.
Ming, D., Kong, Y., Lambert, M., Huang, Z., Ma, J. (2002) How to describe protein motion without amino-acid sequence and atomic coordinates, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8620.
Ming, D., Kong, Y., Wakil, S.J., Brink, J., Ma, J. (2002) Domain movements in human fatty acid synthase by quantized elastic deformational model, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7895.
Schlitter, J., Engels, M., Kruger, P., Jacoby, E., Wollmer, A. (1993) Targeted molecular dynamics simulation of conformational change - application to the T-R transition in insulin, *Mol. Sim.* **10**, 291.
Tirion, M.M. (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis, *Phys. Rev. Lett.* **77**, 1905.

Surface Polysaccharides Antigens: *see* Carbohydrate Antigens

Synchrotron Infrared Microspectroscopy

Nebojsa S. Marinkovic and Mark R. Chance
Department of Physiology and Biophysics and Center for Synchrotron Biosciences, Albert Einstein College of Medicine, Bronx, NY 10461

1 Introduction 673
1.1 Development of IR Spectrometers 674
1.2 FTIR Microscopes 676

2 Synchrotron Radiation 676
2.1 Synchrotron Infrared Facilities Worldwide 676
2.2 Production of Synchrotron IR Radiation 678
2.3 Advantage of the Synchrotron Source 678
2.4 U2B Beamline Microspectroscopic Instrumentation and Capabilities 679

3 Experimental Aspects 680
3.1 Microscopic Sampling for Transmission Spectroscopy 680
3.2 Microscopic Sampling for Reflection Spectroscopy 682
3.3 Microscopic Sampling for ATR Spectroscopy 682
3.4 Microscopic Sampling for Grazing Angle Spectroscopy 683

4 Sample Preparation 683

5 Band Assignments 684
5.1 Region 3600–3100 cm^{-1} 685
5.2 Region 3100–2800 cm^{-1} 685
5.3 Region 2800–1800 cm^{-1} 685
5.4 Region 1800–1600 cm^{-1} 686
5.5 Region 1600–1000 cm^{-1} 686

6 Applications: Analysis of Infrared Microscopic Maps and Images 686
6.1 Analysis of Bone Chemistry in Osteoporosis 687
6.2 IR Microscopy Studies of Epilepsy and Neurodegeneration in Animals 690

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 13
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30650-1

6.3 Microscopic Structure and Properties of Seeds 691
6.4 Biotransformation of Pollutants in Roots 692
6.5 IR Microscopy of Hair Sections 693
6.6 Other Biological Applications 695

7 Conclusion 695

Bibliography 696
Books and Reviews 696
Web Pages 696
Primary Literature 696

Keywords

Fourier Transform (FT)
A mathematical technique capable of converting a time domain signal to a frequency domain signal.

Infrared (IR)Light
Electromagnetic radiation at wavelengths longer than the red end of visible light and shorter than microwaves (roughly between 1 and 100 microns).

Infrared Microspectroscope
An instrument that combines features of a spectrometer and microscope apparatus, providing an infrared spectrum for a small sample area of the specimen.

Microscope
An instrument that provides a greatly enlarged image of an object, allowing the viewing of features that cannot be seen with the naked eye.

Spectrometer
An instrument for analyzing a spectrum, that is, the intensity of radiation as a function of wavelength.

Synchrotron
A circular accelerator of charged particles, such as electrons. As the electrons accelerate around each bend in the ring, they are guided by powerful magnets and give off energy in the form of light.

Synchrotron Infrared Microspectroscopy is a powerful method for obtaining infrared spectrum of a small area of the specimen, using synchrotron light as the source. The brightness of the broadband synchrotron beam is 1000 times higher than that of thermal infrared sources, and allows analysis of sample areas down to the diffraction limit (e.g. 5 μm at 2000 cm^{-1}) at high signal-to-noise ratios in a short time. The technique is most widely used in biosciences, providing information on chemistry on microscales of virtually all biological specimens, like bone, plants, and other biological tissues. Some applications of the technique include infrared imaging of neurons subjected to various brain diseases, bone osteoporosis, biodegradation of man-made pollutants in plants, nutrition value of seeds, and protein folding in microsecond time regimes.

1 Introduction

Infrared (IR) spectroscopy measures intensity of the IR light at a particular frequency, providing information on the vibration of covalent bonds within molecules. The wavelengths of the IR light absorbed by a vibrating bond depend on the nature and type of the bond, the applicable selection rules, and the type of the vibration excited, and are also influenced by the inter- and intramolecular interactions of the molecules. The intensity of the light depends (as defined by the Beer–Lambert law) on the concentration of the vibrating molecules. Therefore, the IR spectrum furnishes information related to the quantity, composition, structure, and environment of the sample. Since the spectrum is additive, that is, all IR-active bonds present in the sample contribute to it simultaneously, a complete picture of the absorptions of functional groups is obtained, offering detailed information on sample chemistry.

IR spectroscopy has a number of advantageous features for the analysis of biological samples. Data acquisition can be quite rapid and dilute species can be detected. The method is also flexible, as samples can be examined in liquid or solid states without destruction of the sample, while instruments are widely available, reliable, and easy to use. Infrared studies of purified materials are quite straightforward. However, biological samples from tissues or cell preparations are both spatially and chemically heterogeneous. Probing samples with a beam size larger than the discrete spatial elements produces an average spectrum of all tissues probed by the beam, as the IR spectrum is additive. Without a sufficiently small beam, the discrete chemical species within discrete features of the tissue cannot be identified.

The device that focuses the infrared light onto a small spot is an IR microscope. The importance of the IR microscopy was realized early in the last century, but only during the last two decades has it experienced rapid development. Spatially resolved IR microspectroscopy can produce a two-dimensional array of spectra from which a chemical image (often called "map") is generated. Infrared microscopy has been utilized in various branches of sciences, including forensic approaches, materials science, art restoration, cell biology, and disease. Its advantage over a number of other spectroscopic techniques

is that it provides chemical information for multiple species simultaneously and directly on a micron scale without stains or contrast agents. Light microscopy can provide chemical information using specific chemical imaging agents, while X-ray techniques can be tuned to an absorption edge to probe specific elements. Magnetic resonance spectroscopy can also provide chemical information and has sensitivity similar to infrared, but submillimeter resolution is difficult. The high brightness of the synchrotron as an infrared source provides enhanced lateral resolution, faster acquisition times, and superior spectral quality. As a result, synchrotron infrared spectroscopy has flourished at the turn of the century, and a number of book chapters and review articles have been published. This chapter illustrates the potential offered by infrared microspectroscopy coupled with synchrotron radiation in biochemistry and biology.

1.1 Development of IR Spectrometers

The first IR spectrometers were equipped with monochromators, which select single wavelengths (or frequencies) of IR light and deliver them to the sample. The monochromator could be based on prisms or diffraction grating. These dispersive spectroscopes were slow and the observed source size limited the IR light that could be focused onto the sample such that spatial resolutions of less than 1 mm were impractical, even with elaborate beam condensers. The development of technology, especially in computers, brought the first interferometric spectrometers to the market. These developments changed not only the IR spectrometry in general but also made IR microspectroscopy possible.

Interferometric spectrometers are based on the Michelson interferometer schematically shown in Fig. 1. The IR radiation from the source is divided in half by a beamsplitter, usually a half-silverized mirror placed at 45° to the incident beam. The two halves of the IR beam further assume separate paths, each going toward one arm of the interferometric stage. One half travels toward a stationary mirror, and the other is directed toward a moving mirror. After the reflection, the two halves of the incident radiation recombine at the beamsplitter and again divide into two parts, traveling either toward the source or the sample (see Fig. 1). Only the half traveling toward the sample is used. If the mirror is moved at constant velocity, the two halves of the incident beam travel different pathlengths and the light reaching the sample, and ultimately the detector, is modulated in intensity due to constructive and destructive interference. The intensity of the light at the detector will vary sinusoidally, a maximum being registered each time that the retardation (i.e. difference of the pathlengths) is an integral multiple of the light wavelength λ. The plot of the intensity of the light versus retardation is called interferogram. Since retardation is also time dependent as the mirror moves at constant velocity, one obtains a single-beam IR spectrum by computing cosine Fourier transform of the beam intensity at the detector. The single-beam IR spectrum is a variation of the light intensity as a function of wavelength (or frequency) of the light. The abscissa of the IR spectrum is therefore inverse time, as the frequency is, by definition, the number of oscillations in unit time. A transmittance spectrum is obtained by dividing two single-beam spectra, one from an empty cell (background) and one with the sample. For traditional reasons, the abscissa of an IR spectrum is

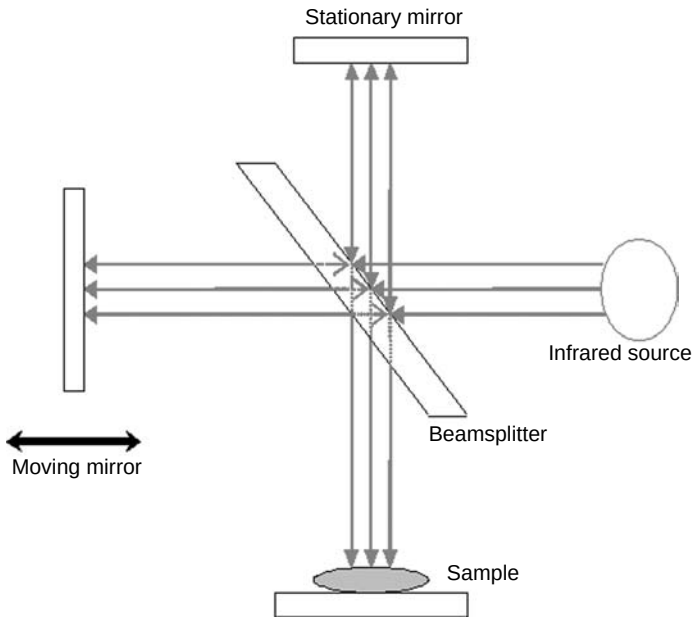

Fig. 1 Michelson interferometer, showing incident radiation coming from the source on the right onto the beamsplitter that delivers one half of the beam to the stationary mirror (a) and the other half to the moving mirror (b). After recombining the two half-beams, the modulated light is reflected toward the sample held by a substrate on the bottom.

usually plotted with increasing wavelength (in micrometers), or with decreasing wave numbers (reciprocal wavelength, in cm^{-1}), while the ordinate is given in absorbance (i.e. negative logarithm of transmittance) because the latter is directly proportional to the concentration of the sample by the Beer–Lambert law.

It should be realized that all frequencies of the incident light are going toward the two arms of the detector simultaneously, so that the intensity of the signal registered at the detector is the sum of all IR frequencies. The single-beam spectrum is produced by complex Fourier transform of the light intensity for all frequencies and all retardations. Fourier Transform Infrared (FTIR) spectrometers are able to compute the IR spectrum by the use of an algorithm for fast-Fourier transform (FFT), developed in mid-1960s. Along with the development of fast, low-cost computers required for the spectrum computation, other technical advances brought FTIR spectrometers into routine usage, ultimately replacing dispersive spectrometers. The ability to shine all wavelengths onto the sample simultaneously and deconvolve the discrete frequencies and the ability of the beamsplitter to accept greater input of radiation compared to a grating or monochromator are the major advantages of modern instruments (termed *Fellgett's and Jacquinot's advantages*, respectively). Additional advantages include optical verification of mechanical motion by interference fringes of a laser positioned parallel to the IR beam, and development

of fast and sensitive detectors that greatly enhanced sensitivity such as mercury cadmium telluride (MCT) detectors.

1.2 FTIR Microscopes

Infrared microspectroscopy provides information relating the spatial distribution of materials within IR suitable samples (described below) with the goal in biology to obtain spectra of individual tissue components. An infrared microscope is a beam condenser that reduces the diameter of the IR beam by focusing the light onto a small area. Even though the first commercial IR microscopy attachment was designed in the 1950s, it was not until the 1980s that it became a practical analytical tool, coinciding with the wider use of FTIR spectrometers worldwide. During the 1980s, IR microscopes became available for nearly all FTIR spectrometers. Stringent requirements are needed in the design of a microscope to achieve high signal-to-noise ratio with high spatial resolution. IR microscopes are available from a number of companies worldwide and their Web pages are given in the Bibliography section.

In design and operation, infrared microscopes are similar to the visible microscopes, except that their optics contain only mirrors rather than lenses, so that the IR light can pass through with minimal absorption. The IR light follows the same path as the visible light. Furthermore, modern IR microscopes can be equipped with a number of methods for better visualization of the sample, including polarized light, fluorescence illumination, or differential interference contrast imaging.

Although novel FTIR spectrometers and microscopes deliver high performance, often the practical limit of the spot size at the sample is tens of micrometers. This limits the applicability of the IR microspectroscopy when imaging single cells or tissues, as the subcellular structures relevant to biology or disease rely on imaging at the diffraction limit of light (5 μm at 2000 cm^{-1}). A synchrotron source allows the spot size to be reduced, because the synchrotron IR light is essentially diffraction limited and orders of magnitude brighter than thermal sources, as shown over twenty years ago. Most commercially available IR spectrometers and microscopes accept input from an external source, so that very little modification of the instruments is needed to adapt it to accept synchrotron light.

2 Synchrotron Radiation

2.1 Synchrotron Infrared Facilities Worldwide

Synchrotrons are facilities that produce high-flux electromagnetic radiation in a continuous spectrum from infrared through the visible and UV to the X-ray region onto a small spot. Generally, their operation relies on emission of radiation of electrons orbiting at a speed close to the speed of light (see Sect. 2.2). Because of the advantages of the synchrotron over thermal IR source (see Sect. 2.3), many synchrotron facilities around the world have at least one infrared beamline either in operation or planned. The list of the current operational and planned synchrotron infrared facilities is given in Table 1. Design and current research on infrared beamlines are reviewed in recent publications.

Tab. 1 Operational and planned synchrotron infrared facilities worldwide.

Operational	
National Synchrotron Light Source-Upton, USA	http://infrared.nsls.bnl.gov/
Advanced Light Source-Berkeley, USA	http://infrared.als.lbl.gov/
Synchrotron Radiation Center-Madison, USA	http://www.src.wisc.edu/
Synchrotron Radiation Source-Daresbury, UK	http://www.srs.dl.ac.uk/
Laboratoire Utilisation du RayonnementElectro-Magnetique-Orsay, France	http://www.lure.u-psud.fr/Experiences/SACO/SA5/intro.htm
Synchrotron Radiation Source ANKA-Karlsruhe, Germany	http://www.fzk.de/anka/beamlines_details/infrared.htm
Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung (BESSY)-Berlin, Germany	http://www.bessy.de/lab_profile/en.06.iris.php?language=en
Spring-8-Hyogo, Japan	http://www.spring8.or.jp/ENGLISH/facility/bl/
PublicBeamline/BL43IR/micro.html Synchrotron Radiation Research Center-Hsinchu (Taiwan, ROC)	http://user.srrc.gov.tw/BeamLine/beamline.html
Canadian Light Source-Saskatoon, SK, Canada	http://cls-ccrs.usask.ca/
Planned	
Centre for Advanced Microstructures and Devices (CAMD)-Baton Rouge, USA	http://www.camd.lsu.edu/
Elettra-Trieste-Italy	http://www.elettra.trieste.it/
Laboratorio Nacional de Luz Sincrotron (LNLS)-Campinas SP, Brazil	http://www.lnls.br/
MAXLAB-Lund, Sweden	http://www.maxlab.lu.se/
Source Optimise'e de Lumie're d'Energie Interme'diaire de LURE (SOLEIL)-LURE, France	http://www.synchrotron-soleil.fr/anglais/index.html
Swiss Light Source (SLS)-Paul Scherer Institute In Switzerland	http://sls.web.psi.ch/
Duke Free Electron Laser Laboratory (DFELL)-Duke University, NC, USA	http://www.fel.duke.edu/
Thomas Jefferson Laboratory-Newport News, USA	http://www.jlab.org/FEL/
SINBAD-Daphne Light-Frascati, Italy	http://www.lnf.infn.it/esperimenti/sr_dafne_light/
UVSOR Facility, Okazaki, Japan	http://www.uvsor.ims.ac.jp/defaultE.htm
National Synchrotron Radiation Laboratory-Hefei, China	http://www.nsrl.ustc.edu.cn/en/enhome.html
SURF III	http://physics.nist.gov/SURF
ISMI at SSLS-Singapore	http://ssls.nus.edu.sg

2.2 Production of Synchrotron IR Radiation

At the National Synchrotron Light Source (NSLS), electrons from an electron source are accelerated with a linear accelerator. Subsequently, they enter into a booster ring where they are further accelerated before entering a storage ring (see Fig. 2). Infrared radiation is delivered by the smaller of the two storage rings of NSLS, the VUV-IR ring. In this ring, the electrons are accelerated to energy of 800 MeV. The storage ring uses magnetic field to bend electron trajectory onto a closed orbit. Because the velocity of electrons is close to the speed of light, they give off the radiation in a relativistic manner at the end of each of these ''bending'' magnets. The radiation thus produced is directional, concentrated, free of thermal noise, and predominantly polarized. The highest current (1 A) is obtained immediately after the injection of electrons in the storage ring, and the current is allowed to decay over time to about 400 mA after several hours. In order to keep the average current around 750 to 800 mA, the storage ring is filled seven times in 24 h, with the fill interval of 4.5 to 5 h during the normal working hours, and 3 to 3.5 h during the remainder of the day.

2.3 Advantage of the Synchrotron Source

To obtain a good-quality spectrum, the intensity of an IR band must be at least 100 times greater than noise. Signal-to-noise

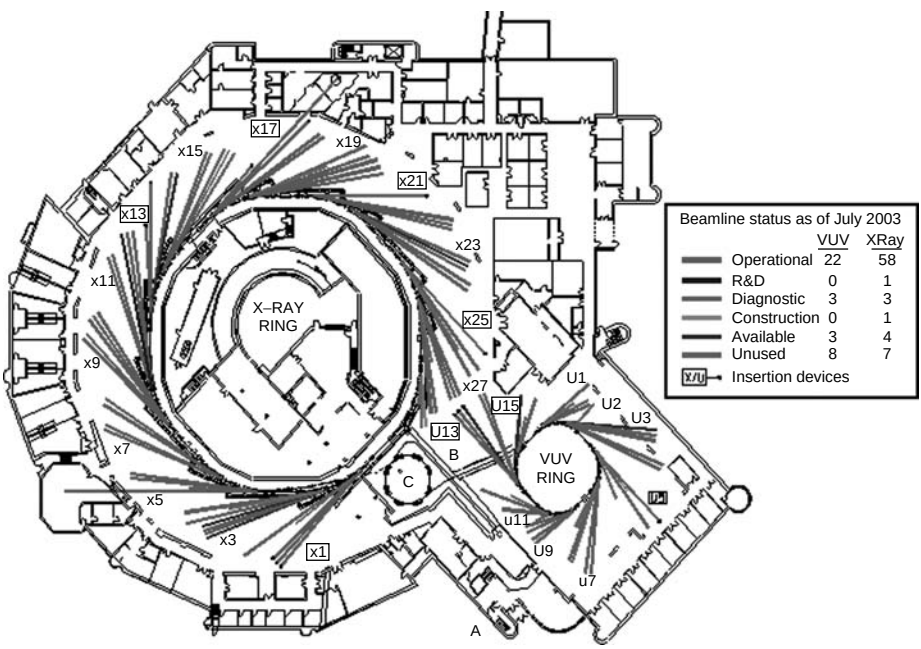

Fig. 2 Schematic diagram of the synchrotron facility at the National Synchrotron Light Source, Brookhaven National Laboratory, Upton, NY, showing the electron source (A), linear accelerator (B), and booster ring (C), as well as the VUV/IR storage ring and beamlines. The location of U2 port on which the Albert Einstein Center for Synchrotron Biosciences' U2B beamline is constructed is also shown.

ratio (S/N) is related to the ratio of source power (SP) that can be calculated for each source, and noise power (NP) that is characteristic of the instrument. Source power at a wavelength λ is directly proportional to the brightness $B(\lambda)$ of the light source. The source comparison of the synchrotron and globar sources was recently calculated. For a standard FTIR spectrometer with a globar source, SP and NP can be calculated to be about 4.5 mW and 0.1 nW respectively, for IR radiation at $\lambda = 2000\ cm^{-1}$ (5 μm) and one-second-long measuring time, so that S/N is 45 000. However, such a high S/N is obtained with the beam size at the sample no smaller than 0.6 mm. SP degrades rapidly when the spot size is reduced. For IR microscopy where spatial resolutions in micron scales are needed, $SP(\lambda)$ at a 5-μm spot size can be calculated to be only about 0.05 nW, so that S/N = 0.5. To obtain S/N of 100, the scan time would be about 11 h. Unrealistically long scanning time (1.5 months) is needed to achieve S/N of 1000.

A brighter source like synchrotron easily overcomes this difficulty. The advantage of the synchrotron lies in its brightness that is 2 to 3 orders of magnitude higher than that of globar sources. Brightness of a storage ring is given by the relation: $B(\lambda) = 75\ I\ BW/\lambda^3$, where I is the ring current and BW is the bandwidth. Using the parameters for the NSLS VUV-IR ring, one obtains $SP(\lambda)$ of 0.1 mW, so that S/N is about 1000 for one-second measuring time. By comparison, the synchrotron is an excellent source for infrared microscopy studies at spatial resolutions on the order of the wavelength λ. It should be noted that the source power of a synchrotron is directly related to the ring current (not the ring power), which explains why the NSLS VUV-IR ring, with a maximum current of 1 A, is the world's premiere facility for the infrared microscopy. The synchrotron infrared beam is nondestructive to biological samples, producing an average temperature rise of only 0.5 K to a test sample closely resembling biological materials.

2.4 U2B Beamline Microspectroscopic Instrumentation and Capabilities

The Albert Einstein Center for Synchrotron Biosciences' U2B beamline is located on the U2 synchrotron IR port of the VUV-IR ring of the National Synchrotron Light Source (see Fig. 2). The port delivers light in solid angle of about 80 mrad horizontally by 40 mrad vertically. The U2 extraction port chamber was installed in the VUV ring in December 1996. The mirror box that divides the light for the use of two equal-end stations (U2A and U2B) was installed in October 1997, and the remainder of the beamline in March 1998. After conditioning, it took full beam in May 1998. U2A and U2B end stations were then constructed, and each took light in November 1998. The U2B beamline has been in full operation and has been accepting outside users since the spring of 2000, after a commissioning period of one year.

The synchrotron beam is steered and focused by a combination of ellipsoidal and planar mirrors to provide a 1 : 1 image of the source. The beam passes through a series of valves and masks, then through a wedged diamond window into a mirror box where it is divided into two equal lines (U2A and U2B). The beam that is steered toward U2B beamline continues through a rough vacuum-pumped tube and then through a window (KBr or polyethylene, depending on the desired frequency range) into the nitrogen-purged instrumentation listed below. The total angular acceptance is 45 H X 40 V mrad.

The beamline is equipped with a dry nitrogen-purged FTIR spectrometer (Magna 860, Thermo Nicolet Instruments, Madison, WI, USA), a Nicolet NicPlan IR microscope, and a custom-made external cryostat chamber. The spectrometer is equipped with a Michelson interferometer having capabilities such as rapid scan, step scan, and the ability to accept external beams and utilize external detectors. For measurements in its internal chamber, a narrowband mercury–cadmium telluride (MCT/A) and a deuterated triglycine sulfate (DTGS) detector with polyethylene window are available, covering a range from 5000 to 100 cm^{-1}. The doubly confocal NicPlan IR microscope is outfitted with a full range of objectives, such as 10X and 20X visible objectives, a 32X IR objective (Thermo Spectra-Tech) for both transmission and reflection IR measurements, and an attenuated total reflection (ATR) ZnSe objective (Thermo Spectra-Tech) for reflection measurements of highly absorbing and/or thick samples. The microscope is capable of diffraction-limited spatial resolution. It is further equipped with a computer-controlled stage that brings the sample under the objective, and fully automated IR microscopic experiments are performed by Atlμs software (Thermo Nicolet) with the possibility to view the visual image of the sample together with its three-dimensional IR representation (IR map).

In order to broaden the range of vibrational spectra collected by the microscope, a flat pick-off mirror has been introduced in the optical path to allow the use of external detectors. The mirror directs the light either to the internal narrowband MCT/A detector operating in the 7000 to 650 cm^{-1} range, or to the external detectors. With the use of external detectors, the lower frequency limit of the spectra obtained through the microscope is extended, at the expense of the signal-to-noise ratio. Three detectors are placed behind the microscope: a liquid nitrogen–cooled broadband MCT/B detector (Belov Technology, New Brunswick, NJ, USA) operating in a 5000 to 450 cm^{-1} spectral range with a sensor size of 0.25×0.25 mm, a liquid helium–cooled dual-detector system (Infrared Laboratories, Tucson, AZ, USA) combining a silicon bolometer with a wedged white polyethylene window that covers the far-spectral range (650–10 cm^{-1}), and a Si:B photodetector with a KRS-5 window for measurements in the mid-IR (3500–350 cm^{-1}). The signal-to-noise ratio is the highest for MCT/A (3–8 times higher than MCT/B, depending on the frequency, while that of the Si:B photodetector is comparable with MCT/B around 600 cm^{-1}, though improved by about twofold to around 450 cm^{-1}.

With the use of the bolometer, the lower frequency limit is extended to about 30 cm^{-1}. A set of motorized and manual translation stages and collimating optics are placed behind the microscope to precisely align the external detectors. The complete optical path, from the spectrometer through the microscope and to the detectors, is continuously purged with dry nitrogen to insure a significant reduction in water vapor absorption that would otherwise render portions of both the mid-IR and far-IR unusable.

3 Experimental Aspects

3.1 Microscopic Sampling for Transmission Spectroscopy

A collimated beam from the synchrotron is directed onto the interferometer, which

modulates the wavelength of the light, and then is passed on into the IR microscope. Depending on the design of the microscope, the light can be delivered to the microscope's first objective either from the top or the bottom. In the case of the NicPlan microscope, at the U2B beamline, the light enters from the top. Before entering the first condenser, the light passes through a variable aperture that limits the illumination of the sample to the selected area. The schematic representation of the microscope in transmission and reflection modes is shown in Fig. 3.

In transmission mode, the IR light enters the first condenser that focuses the beam onto the sample and passes through the sample that partially absorbs the light. The nonabsorbed beam then passes through an IR window that holds the sample and is collected by a second condenser that recollimates the beam. To compensate for spherical aberration caused by the IR window (or windows), condensers are equipped by a variable compensation ring. If there is only one window beneath the sample, the compensation ring at the first condenser is set to zero, while that at the secondary condenser should be set to the thickness of the window. In the case when a sample is squeezed between two windows, compensation rings on both condensers should be set to the thickness of the window. To minimize light scattered from the sample, the IR light recollimated by the second condenser usually passes through another aperture on its path to the detector.

Windows for transmission spectroscopy should be less than 3 mm thick. Choosing a window depends on the frequency range desired, physical properties (e.g hardness, if sample needs to be squeezed between two windows), and chemical properties (e.g. solubility in water). Care should be taken if the sample was kept in an ammonium-based solution prior to sectioning and placing on the window, as many of these window materials can make complexes. The list of the common

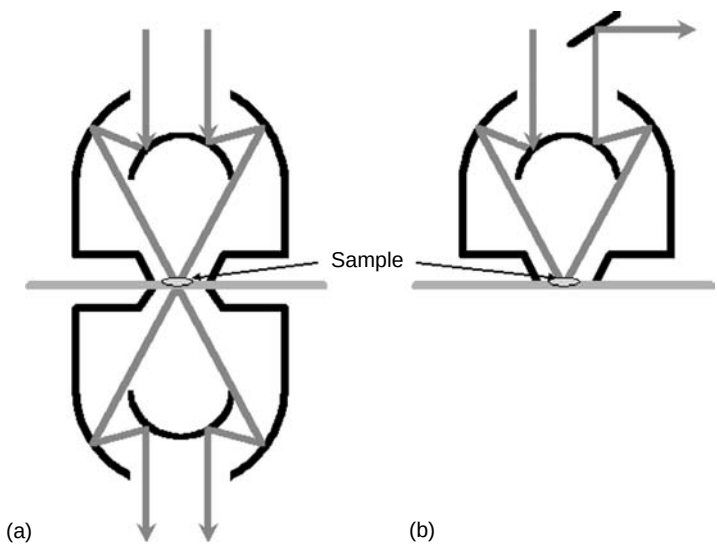

Fig. 3 Optical path of the infrared beam through Schwarzchild IR objective in (a) transmission and (b) reflection modes.

Tab. 2 Physical and chemical properties of some commonly used window materials for IR spectroscopy.

Window Material	*Transmission Range [cm^{-1}]*	*Properties: solubility in H_2O [g 100 cm^{-3}]*
CaF_2	4000–1100	Clear; 0.002
BaF_2	4000–800	Clear; 0.17
ZnS	4000–600	Clear; insoluble
ZnSe	4000–650	Orange; insoluble
KBr	4000–400	Clear, very soluble
KRS-5 (TlBr-TlI)	4000–200	Red; 0.05
Diamond	4000–50	Clear; insoluble; expensive

transmission window materials is given in Table 2. Most of these windows are also transparent to visible light.

3.2 Microscopic Sampling for Reflection Spectroscopy

The second most common way of collecting IR microspectra is in reflection mode. In this case, only about one half of the beam reaches the sample (see Fig. 3). The modulated IR light is sent through the top aperture and focused onto a small area of the sample by the first condenser. The portion of the light not absorbed by the sample reflects off the flat surface beneath the sample and passes through the same portion of the sample once again. The first condenser recollimates the light and sends it to the detector by a mirror. Since this mirror blocks 40 to 50% of the incident light, transmission mode is a preferred method for collecting microspectra because of the higher flux of the incident light on the sample. Furthermore, inhomogeneities in the sample may cause oscillations in the background of the reflection spectra. Because the beam passes through the sample twice, the thickness of the sample should be approximately half of that for the transmission spectroscopy.

Traditional substrate material on which the sample is placed is a highly polished metal slide that efficiently reflects IR radiation, for instance, gold. Recently, IR-reflective slides became commercially available at a fraction of the cost of the windows for transmission spectroscopy. They are made of a common glass slide on which an IR-reflective coating is deposited on one side. The reflective surface consists of several layers, comprising two silver coatings that reflect IR light, and a protective coating of SnO_2. One of the first vendors of these slides, Kevley Technologies named them low-e MirrIR. The slides are durable, insoluble in water and common organic solvents, and transparent for the visible light while reflecting about 95% of the IR radiation in the range 4000 to 400 cm^{-1}. Because of these properties, reflection spectroscopy recently became a very popular method for microspectroscopy.

3.3 Microscopic Sampling for ATR Spectroscopy

Attenuated total reflection spectroscopy is done on samples that cannot be cut into thin sections suitable for either transmission or reflection methods. An ATR

objective is a Schwarzchild IR objective that contains a small hemisphere made of an IR-transparent material, like ZnSe, Ge, or diamond. Since Ge is opaque for visible light, and diamond can be cost prohibitive, ZnSe is usually preferred. The objective has two masks. For viewing of the sample, the mask lets in radiation that is nearly perpendicular to the sample. During data collection, mask permits only radiation that enters the ATR hemisphere at near-grazing angles (65–80°), so that the rays are totally internally reflected back into the hemisphere. However, because of the nature of light propagation at the interface between two media having different refractive indices, the light penetrates into the sample via evanescent field. Penetration depth of the light is usually of the order of the wavelength (i.e. several micrometers), which is suitable for the IR measurements. The totally reflected light is then recollimated by the objective as in the reflection mode, and sent to the detector. Because the mask only allows beams that are close to grazing angle, a severe reduction in flux, up to 70%, is expected in the ATR spectroscopy in comparison to the normal reflectance mode.

In contrast to the microsampling discussed above, the thickness of the sample is not a factor in ATR spectroscopy. Although this technique involves very little sample preparation, it is not applicable in all cases. Since the sample needs to be in close contact to the ATR element to ensure a good-quality spectrum, the use of automated mapping methods cannot be done because ATR element and/or the sample may be damaged while the sample stage is moved. Samples must be either firm (like bone) or mounted on a stiff substrate. Care should be taken with soft samples as they may be damaged, and could also leave a residue on the ATR element, so that it would need cleaning before each data collection. Furthermore, the spectral region is obviously limited to the transmission range of the ATR element (see Table 2).

3.4 Microscopic Sampling for Grazing Angle Spectroscopy

Grazing angle spectroscopy is commonly used for probing very thin samples placed on a highly reflective surface. The surface-sensitive technique uses similar objective and mask as in ATR spectroscopy, and the IR light path is also similar. No ATR element is needed for grazing angle mode. The substrate is a highly polished, reflective surface and the sample is deposited directly on it. The thickness of the sample must be less than the wavelength of the IR radiation, to get a good signal with this technique, and because of the difficulty to cut such thin biological samples, this technique is rarely used.

4 Sample Preparation

The sample preparation of biological samples is possibly the most critical part of a successful IR microscopy experiment. Questions regarding stability or degradation of the sample, especially right after the removal from the body, fixation, imbedding into materials for sectioning, thickness of the samples for different IR microscopy experiments, and so on, need to be addressed before the actual sample preparation. Also, for the IR data to be biochemically relevant, the cells must be preserved in a lifelike state without introducing artifacts. A few general principles are outlined below.

If fixation of the sample tissue is desired to prevent possible degradation, formalin should be used rather than water/ethanol mixture, as ethanol causes proteins to aggregate. While formalin has less of a tendency to induce unfolding of proteins, the fixation should be done with low concentrations and washed with copious amount of saline fluid or deionized water before drying to prevent the formalin IR signatures in the spectra.

Tissue samples for IR microspectroscopy should be prepared in a similar fashion as for ordinary light microscopy. Although the optimal thickness varies greatly from one sample to another, generally it should be kept below ~50 μ or ~20 μm for transmission and reflection measurements, respectively. Sectioning of a sample is usually done at low temperatures (between −10 and −20 °C), that is, by cryotoming. If possible, the sample should be cut without any support. Otherwise, the sample can be imbedded in a medium that maintains the physical integrity of the sample during sectioning. The embedding compound should be carefully chosen to match the hardness of the sample. Furthermore, its IR absorption bands should not overlap with those of interest in the sample, as the embedding compound may penetrate into the sample. In most cases, a liquid hydrocarbon such as paraffin wax is a good choice, as its IR absorption is only seen in the CH region. Penetration of the embedding material into the sample can be avoided in some cases if the embedding compound is placed only at one side of the sample block before sectioning. However, physical integrity of the biological materials may be lost.

Once a thin section is cut, it is placed onto an IR substrate. For transmission microscopy, the substrate should be one of the IR windows described above (see Sect. 3.1). Regular glass windows are only partially transparent in IR region (4000–2500 cm^{-1}) so that only CH stretching region can be probed with a sample on a glass substrate. For reflection microscopy, the media are polished reflective surfaces, like low-e slides.

Biological materials often strongly adhere to IR windows and/or low-e reflective slides, and washing and staining can be done with the sample placed on the substrate. While staining allows better visualization of the sample components, care should be taken as the stains may produce artifacts, either by exhibiting their own spectral features or by changing the IR absorption of the sample by interaction.

5 Band Assignments

Since the IR spectrum is additive (see Sect. 1), most tissue samples can be approximated as a sum of three spectral types, namely, lipids, proteins, and polynucleides. Each of these building blocks contributes to the overall spectrum with its relative concentration. Relevant examples of them are given in Fig. 4.

The spectrum of a protein (Fig. 4a) is characterized by strong amide bands, arising from the protein backbone. Lipids (Fig. 4b) are characterized by strong absorption methyl and methylene bands (CH region around 2900 cm^{-1}). Nucleic acids (Fig. 4c) show strong characteristic absorptions of P=O moieties in 1000–1200 cm^{-1} region, while carbohydrates (Fig. 4d) contain broad O−H stretching bands above 3000 cm^{-1}. For easier viewing, we shall divide the spectrum into several parts and discuss each of them on the basis of the building blocks mentioned.

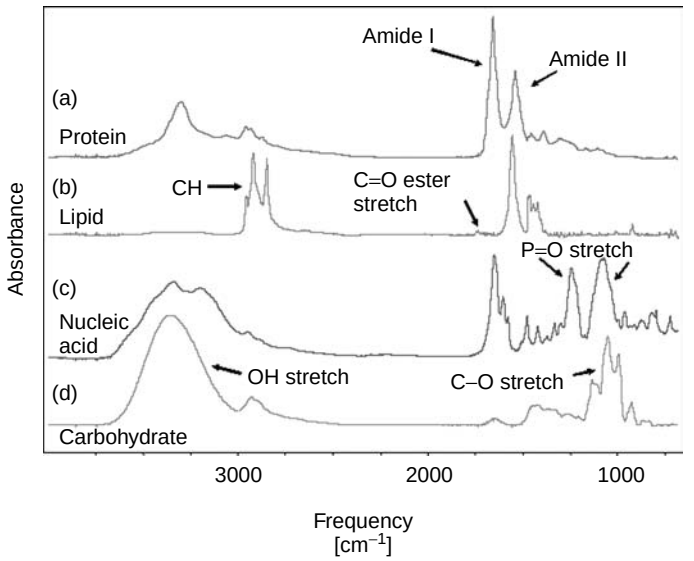

Fig. 4 Fundamental spectral features of biological components.

5.1
Region 3600–3100 cm^{-1}

This region consists largely of O−H and N−H stretching vibrations. Unlike all other bands that are usually sharp with frequencies sensitive to the electron affinities of the groups attached to them, this region is mostly devoid of sharp bands. Instead, a very broad band is observed, due to hydrogen bonding. Its intensity varies with the concentration of the O−H and N−H functional groups. All building blocks show bands in this region, but their shapes and intensities vary. Lipids, having only −OH groups at the end of the molecules, produce only a small hump. Proteins show larger peak due to −NH groups present in every amino acid, while nucleic acids and carbohydrates show distinctive features in this range. Unfortunately, both the intensity and peak position vary significantly with the concentration of water (external or internal) and the interpretation of bands in this region should be taken with caution.

5.2
Region 3100–2800 cm^{-1}

This region of the spectra where −CH stretches are observed is distinctive in lipids. By comparison to alkanes, they are assigned to asymmetric and symmetric CH_3 and CH_2 stretches. Since lipids are long-chain acids, the intensity of the CH_2 absorptions (2920 and 2850 cm^{-1}) is 10 to 20 times that of CH_3 bands (2960 and 2875 cm^{-1}). Proteins show only weak features in this region, as the methyl and methylene groups are present only in the amino acid side chains, with almost equal proportions. Nucleic acids and carbohydrates practically have no CH_3 groups and only weak CH_2 and CH bands are seen in the spectrum.

5.3
Region 2800–1800 cm^{-1}

This region is generally free of bands associated with biological samples, except

for combination bands of water in wet tissues.

5.4 Region 1800–1600 cm^{-1}

The absorption in this spectral region is due mostly to the ester C=O stretching vibrations. Lipids show a distinctive band around 1750 cm^{-1} in this region. Its frequency is strongly affected by hydrogen bonding to water in biological tissues. Nucleic acids show two distinctive features in 1720–1670 cm^{-1}, arising from purine and pyrimidine bases, respectively. Proteins show one major absorption in this region, called *amide I band*, arising predominantly from the C=O group of amide. It is sensitive to the conformation of the protein, as shown below. The frequency of the C=O vibration is sensitive to the hydrogen bonding, and the dielectric constant of the surrounding microenvironment.

The 1700–1610 cm^{-1} amide I envelope consists of several bands, depending on the protein structure. Antiparallel β-sheets (1685–1680 cm^{-1}), turns and coils (1680–1660 cm^{-1}), α-helices (1660–1648 cm^{-1}), unordered structures (1660–1648 cm^{-1}), β-sheets (1640–1625 cm^{-1}), 3_{10} helices (1637 cm^{-1}), and aggregated strands (1625–1610 cm^{-1}) all contribute to this complex region. Fitting of this envelope with series of bands with above frequencies can be used not only to predict protein secondary structure but also to track the conformational changes and assess the kinetics of proteins folding processes.

5.5 Region 1600–1000 cm^{-1}

This region, often-called the fingerprint region, is characterized by a number of different bands. Proteins show two bands (Fig. 4a) in this region. Both bands, usually called amide II and amide III are sensitive to protein conformation, as is amide I. But, because of their complex nature (amide II is predominantly N−H bending coupled with C=N stretching vibration, occurring between 1560 and 1500 cm^{-1}, and amide III involves C−N stretching, N−H bending and CH_2 wagging modes, seen within 1350–1220 cm^{-1}) they are not as useful as amide I in determining structure. Proteins containing acidic amino acid residues (aspartate and glutamate) show COO^- vibrations around 1580 and 1400 cm^{-1}.

Nucleic acids show a series of sharp but weak bands. The most prominent features are the two bands at 1225 and 1090 cm^{-1}, arising from phosphate PO_2^- asymmetric and symmetric vibrations, respectively. Major bands of lipids in this region arise from CH_2 scissoring vibration at 1470 cm^{-1} and C−O−C stretching vibration, around 1060 cm^{-1}. The C−O stretching vibration of carbohydrates occurs at 1200–1000 cm^{-1} and can be strong in liver tissues.

6 Applications: Analysis of Infrared Microscopic Maps and Images

Once the sample is prepared, placed on an IR substrate, and onto the microscope stage, obtaining IR spectral map is rather straightforward. The map is a spatially resolved array of spectra from which a chemical image is generated. The usual scenario is that an area map is set to collect sample scan through an aperture that is the same, or somewhat larger than the wavelength of the lowest frequency band of interest. For instance, if one is interested only in C−H stretching, an aperture of about 4 μm will be appropriate. Larger aperture (~7 μm) settings are needed if the investigator needs to collect amide I band

in addition to the CH stretching, and even larger (12 μm) if the whole mid-infrared spectrum is desired.

When the aperture is set, one chooses the step size, that is, the distance between two adjacent sample points. The step size is usually set to be equal to, or slightly smaller than the aperture size. The latter choice usually produces smoother transition in band intensities of a map. Finally, one chooses a number of scans per data point. For that, the experimenter should do a set of measurements with a different number of scans at the same data point inside the map. Since the counting statistics component (white noise) of the signal-to-noise ratio is proportional to the square root of the number of scans, one chooses the smallest number of scans that produces satisfactory S/N. Although this is very much sample dependent, 32 or 64 scans are usually good choices.

Mapping a sample is done by collecting spectra in either one-dimensional rows, two-dimensional grids, or in a set of predetermined points. In all cases, spectral collections are completely computer-controlled.

While the interpretation of the data from the map can be done by analyzing a full spectrum of the bands as described earlier for each observed point, this approach is impractical when analyzing a map that may contain hundreds or thousands of spectra. Many FTIR microscopy vendors have developed programs that can present the data in a form of image for easy visualization and comparison to the optical image. Functional group mapping is the most straightforward method, in which a parameter is presented as a function of position in the map. The parameter can be absorption maximum, peak intensity, integrated intensity of a band, or ratios of various parameters. For example, plotting the ratio of band intensity versus that of a baseline produces a reasonable measure of the concentration of a particular species in the map. Each method has advantages and disadvantages. While the intensity of a band (either peak intensity or integrated intensity) is the most straightforward method to obtain the distribution of amount of material represented by a functional group within the map, it may suffer from the baseline fluctuations and changes in thickness of the specimen. Structural properties of a sample can be better visualized by plotting the frequency of absorption maximum as a function of the position in the map, but possible changes in adjacent and/or overlapping bands must be taken into consideration. The ratio of two peak intensities or integrated peak intensities is the method that is often used to avoid the changes of the sample thickness across the area mapped, as well as the synchrotron beam fluctuations over time, but the representing bands must be carefully chosen to avoid problems arising from the baseline changes, artifacts, and so on. Finally, Principle Component Analysis (PCA) is an effective method to display chemical information in a large set of data, as in an IR map, revealing differences between spectra that are not easily seen by inspection and/or spectral subtraction. The major variability between spectra is calculated and concentrated into small sets of values, called *principle components*. Plotting principle components against one another shows clustering of the data.

6.1 Analysis of Bone Chemistry in Osteoporosis

Osteoporosis, characterized by decreased bone density, altered bone architecture,

and increased susceptibility to fracture, is a common and costly disease afflicting millions of women. Although the role of bone density in osteoporosis is well established, little is known about the changes in bone chemistry as the disease progresses. Alterations of several chemical components of bone, including carbonate, phosphate, and collagen, may affect bone strength. Furthermore, in the study of heterogeneous bone tissues, chemical variations at different locations of bone, that is, cortical versus trabecular, may be important in determining the overall mechanical properties of bone.

The application of synchrotron infrared microspectroscopy to examine changes in bone chemistry in a monkey model of osteoporosis has been described. The microscopic matrix and mineral chemistry of osteopenic bone can be monitored by studying the peaks of infrared spectra at amide I (1700–1600 cm^{-1}), carbonate ($\nu_2 CO_3^{2-}$, 825–905 cm^{-1}), and phosphate ($\nu_4 PO_4^{3-}$, 500–650 cm^{-1}). It has been demonstrated that the ratios between the integrated area of phosphate (or carbonate) and amide I bands represent the amount of mineral relative to matrix in bone. Curve fitting the three regions yields information about the compositions of mineral and matrix components that may be altered during disease process. To further analyze these changes, the overall and localized bone quality using two spectral data collection approaches can be used: Mapping broad, randomly selected bone regions to obtain a large number

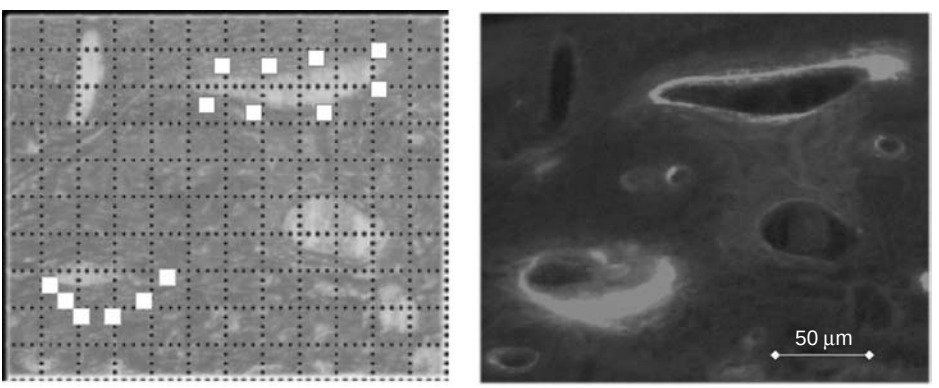

(a) (b)

Fig. 5 (a) 100× light image from a subchondral region of proximal tibia from one of the ovariectomized animals. The squares within the grid indicate the fields of data collection limited by 20 × 20 μm aperture during mapping. White-colored squares correspond to data collection at the fluorescent labels shown in (b). The aperture size in this case was limited to 20 × 20 μm. For each data point, a spectrum was acquired using 64 coadditions and Happ–Genzel apodization at 4 cm^{-1} resolution. A background spectrum through air (no sample) was collected for absorbance reference. (b) Fluorescent image of the same field. Calcin (green) and alizarin (red) indicate bone formed one and two years, respectively, after ovariectomy. (Reproduced by permission from International Union of Crystallography, from Marinkovic, N.S., Huang, R., Bromberg, P., Sullivan, M., Toomey, J., Miller, L.M., Sperber, E., Moshe, S., Jones, K.W., Chouparova, E., Lappi, S., Franzen, S., Chance, M.R. (2002) Center for Synchrotron Biosciences' U2B Beamline: an internal resource for biological infrared spectroscopy, *J. Synchrotron Radiat.* **9**, 189–197) (see color plate p. xxxiv).

of spectra; and, individual point-and-shoot at regions labeled by fluorochromes to be of a given age, Fig. 5. In both cases, synchrotron light source is necessary for data collection at spatial resolutions near the diffraction limit (3–20 μm). The proximal tibiae of animals at three distinct anatomical regions are of interest: cortical, subchondral, and trabecular bone. A bone density enhancing drug, nandrolone decanoate, was administered to a subset of ovariectomized monkeys as a comparison to the nontreated subjects to demonstrate the effect of the drug on bone chemistry during disease progression.

For visual comparisons between different ratios, contour maps of ratio intensities can be generated using Microcal Origin 5.0 (Northampton, MA) with the same coordinate as in the data collection field. For example, the contour map of carbonate to matrix ratios (Fig. 6) from the same field shown in Fig. 5 demonstrates that areas with low intensities correspond to alizarin labeled region (red label), indicating less carbonate mineralization in the newly formed bone. By comparing the results from diseased monkeys with those from intact as well as treated subjects, we are able to detect specific changes in bone chemistry due to disease and therapy. To date, the results of macroscopic chemical changes examined by infrared mapping has been reported, where a significant decrease in carbonate, and increased acid phosphate content were observed in ovariectomized animals. The microscopic results combining FTIR and fluorescence microscopy further demonstrate localized changes in the amount of

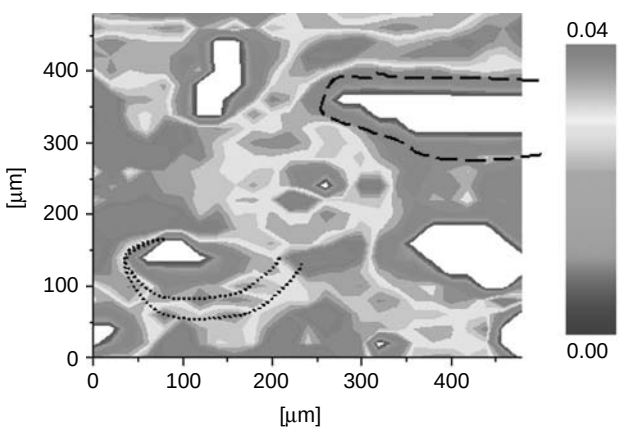

Fig. 6 Contour maps of $\nu_2 CO_3^{2-}$/amide I ratio from the same field shown in Fig. 5. The map was generated by tabulating the ratios obtained from curve-fitting analysis into the same $x - y$ coordinate in Fig. 5(a). The intensity is represented by a color gradient indicated by the column to the right of the figure. (Reproduced by permission from International Union of Crystallography, from Marinkovic, N.S., Huang, R., Bromberg, P., Sullivan, M., Toomey, J., Miller, L.M., Sperber, E., Moshe, S., Jones, K.W., Chouparova, E., Lappi, S., Franzen, S., Chance, M.R. (2002) Center for Synchrotron Biosciences' U2B Beamline: an internal resource for biological infrared spectroscopy, *J. Synchrotron Radiat.* **9**, 189–197) (see color plate p. xxiv).

phosphate and carbonate as well as the degree of collagen crosslinking. These results show that synchrotron FTIR microspectroscopy can study bone tissue chemistry with a level of detail that is not possible with other techniques. In particular, it has the ability to monitor both the mineral and the matrix components of bone simultaneously. In addition, the automated fitting procedures allow the chemical information of the infrared spectra to be rapidly extracted such that quantitative mapping is possible.

6.2 IR Microscopy Studies of Epilepsy and Neurodegeneration in Animals

Hippocampal tissue from humans with chronic temporal lobe epilepsy reveals a specific pattern of neuronal degeneration referred to as *mesial temporal sclerosis*. The predominant feature of this disease includes a selective loss of pyramidal neurons in the hippocampal endblades and frequently in the Sommer sector, that is, CA3 and CA1 areas, respectively. In contrast, pyramidal cells in the transitional zone between the two areas (CA2 area), as well as granular cell layer, subiculum, and fibers *en passage* remain intact. Kainic acid is often used to induce seizures in animals. Neuronal loss is apparent at 24 h, and is maximal at 48 h after the onset of kainic acid status epilepticus. A wide variety of morphological changes associated with dying during development or following a toxic insult is observed in cells. A binary classification scheme suggests that physiologically appropriate death is due to apoptosis and that pathological mechanisms involve necrosis. Necrosis, or accidental or pathological cell death is characterized by rupture of the membrane due to swelling of organelles, so that cytoplasmic contents are released into the extracellular space. In contrast, apoptosis, or physiological cell death is characterized by nuclear condensation and cellular and membrane shrinkage. Apoptosis has been reported to increase following a wide range of pathological states, such as neurodegenerative diseases, adrenalectomy, and seizure.

Visible and infrared images of healthy and diseased neural tissue have been used to observe changes in the pyramidal neurons in the CA3 and CA1 regions. Samples were prepared as follows: An adult rat was exposed to severe status epilepticus and sacrificed by transcardiac perfusion 48 h after the onset of seizure. The normal rat had a sham injection. The brain was fixed overnight in the same perfusate and left in 30% sucrose solution. When the brain sunk, it was rapidly frozen in methylbutane (−40 °C) and cut through the dorsal hippocampus in a cryostat (−18 °C) into 12-μm thick sections. The tissue was mounted onto BaF_2 windows and placed under the IR microscope. The IR maps of CA1, CA2, and CA3 pyramidal cells and surrounding tissue were collected using 64 scans with 4 cm^{-1} resolution. The aperture size was 10 μm, which is roughly the size of a pyramidal cell (see Fig. 7) and the stage was moved in 10-μm increments. The protein, lipid, disulfide, ketone, and phosphate mid-infrared regions were examined to assess biochemical differences in the tissue samples.

We found that examination of the peak intensities of the amide I band to lipid bands (1654 and 2922 cm^{-1}, respectively) provided an indicator of the disease. As seen in Fig. 7, the IR image of bands closely follows the optical image, indicating that the highest amide to lipid peak

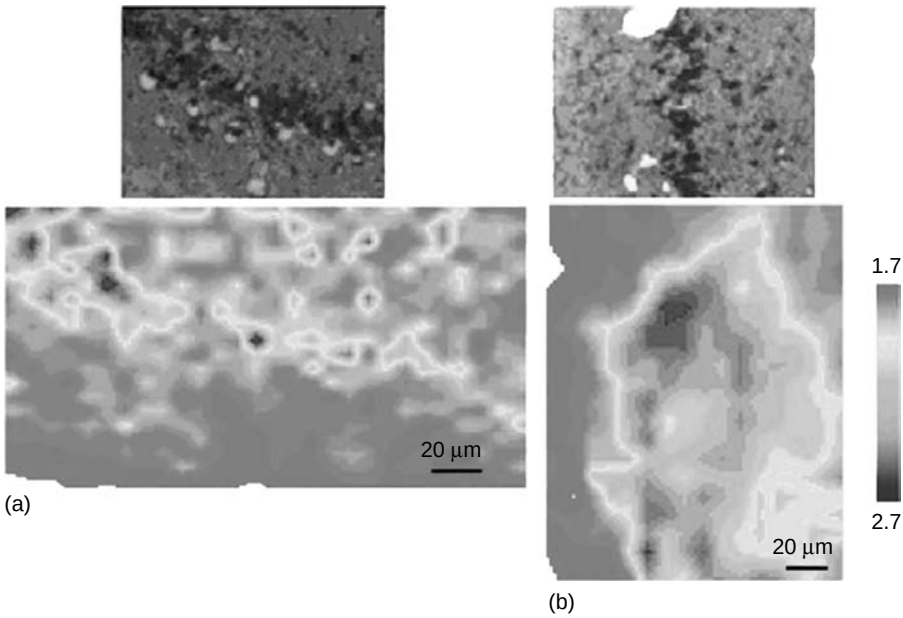

Fig. 7 Optical (top) and false-colored IR images of brain tissue from (a) a normal rat and (b) a rat with severe seizure. Tissues were mounted on BaF_2 windows. The aperture size was 10×10 μm (see the green square at the center of the optical image of normal rat tissue). For each data point in the IR map, a spectrum was acquired using 64 coadditions and Happ–Genzel apodization at 4 cm^{-1} resolution. A background spectrum through a spot without sample was collected for absorbance reference. The map is generated using an amide I to lipid peak height ratio of bands at 1654 and 2922 cm^{-1}, respectively, and the color represents the intensity of the ratio. (Reproduced by permission from International Union of Crystallography, from Marinkovic, N.S., Huang, R., Bromberg, P., Sullivan, M., Toomey, J., Miller, L.M., Sperber, E., Moshe, S., Jones, K.W., Chouparova, E., Lappi, S., Franzen, S., Chance, M.R. (2002) Center for Synchrotron Biosciences' U2B Beamline: an internal resource for biological infrared spectroscopy, *J. Synchrotron Radiat.* **9**, 189–197) (see color plate p. xxxii).

ratio is found within the CA3 cells in the normal tissue. In contrast, the IR image of the tissue sample of diseased animal shows that the amide I to lipid ratio is high not only within the cells, but spreads several tens of micrometers further away from the cell wall. It appears that the diseased neural cells lose their integrity, spreading their biomaterial toward the surrounding tissue. This example indicates the power of IR microspectroscopy to investigate changes in tissues during disease progression.

6.3 Microscopic Structure and Properties of Seeds

Various grain cultures are commonly grown in the world for human and animal diets. High quality grain silage requires attention to details such as plant maturity, speed of harvest, chop length, and silo packing, as they influence the fermentation process, storage losses, palatability, and nutritional value of the resulting feed. The selection of seed varieties has been undertaken throughout the world to

improve the nutritive value and nutrient utilization. To assess total seed chemical composition on protein and carbohydrates, usual procedure involves standard "wet" chemistry, which relies on homogenization of seed and use of harsh chemicals that alter the plant structure and possibly generate artifacts. Probing of the localized chemical microstructures to assess the spatial origin and distribution of components has been conducted by FTIR microspectroscopy, first with a globar source, and later with synchrotron source on varieties of corn, wheat, and barley. Feathers are also considered a potential protein supplement to animal feed as they are widely available as a by-product in poultry and are exceptionally high in protein. A comparison of secondary structure of feather to other protein sources from grain using FTIR microspectroscopy revealed high contribution of β-sheet in feather, up to 88%, whereas α-helix is the main protein component in barley (71%), oat (92%), and wheat (50%) grain. The low α-helix content, as well as other factors including the high keratin content, strong disulfide bonding of amino acids, and the relationship of protein secondary structure of feather to solubility, is believed to be in relationship to the low feather digestibility. Wheat specimen from different wheat cultivars showed that the percentage of α-helix is much higher in hard winter wheats, known for their bread-making quality. The α-helix to β-sheet peak area ratios in hard wheat is between 1.6 and 2.1, while that of soft wheat is about 1.

6.4 Biotransformation of Pollutants in Roots

Root and root exudates were studied by synchrotron FTIR microspectroscopy, in order to obtain information on interaction of plant with the environment. The level of phosphorus in the soil affected lignin concentration in epidermis, and increased level of sugars and lignin precursors in rhizosphere surrounding the root of mung bean. Stem, root, and leaf tissues of a model genetic organism and mutant plants were studied to understand defensive mechanism of the plant toward a predator, Erysiphe spp. fungus.

The plant defense mechanisms toward toxic compounds were the topic of investigation of recent studies. The fate of organic contaminants within the plants used for waste cleanup (phytoremediation), as well as bioavailability of transformed contaminants and their toxicity relative to the parent compound was investigated using benzotriazole as the model compound. This compound is commonly used as a corrosion inhibitor in antifreeze, airplane deicing fluids, gasoline, oil, lubricants, and heat exchanger fluids. The changes in root constituents including carbohydrate, cellulose, lignin, and protein, due to uptake, incorporation, and/or transformation of benzotriazole (BT) were monitored. The distribution of the characteristic CH out-of-plane bending mode of BT at 745 cm^{-1} (Fig. 8) across the IR microscopically scanned region demonstrated a high concentration of BT across the vascular region of BT-treated roots, and a low concentration in the epidermis. This provided evidence that the aromatic ring of BT remained unchanged during the incorporation within sunflower secondary root.

While changes in intensity of the characteristic BT peak were obvious, differences in other portions of spectra are not readily observed (Fig. 9A). This is particularly true for the protein amide I and II regions. To visualize the changes in treated versus untreated sunflower roots better, principal component analysis has been conducted.

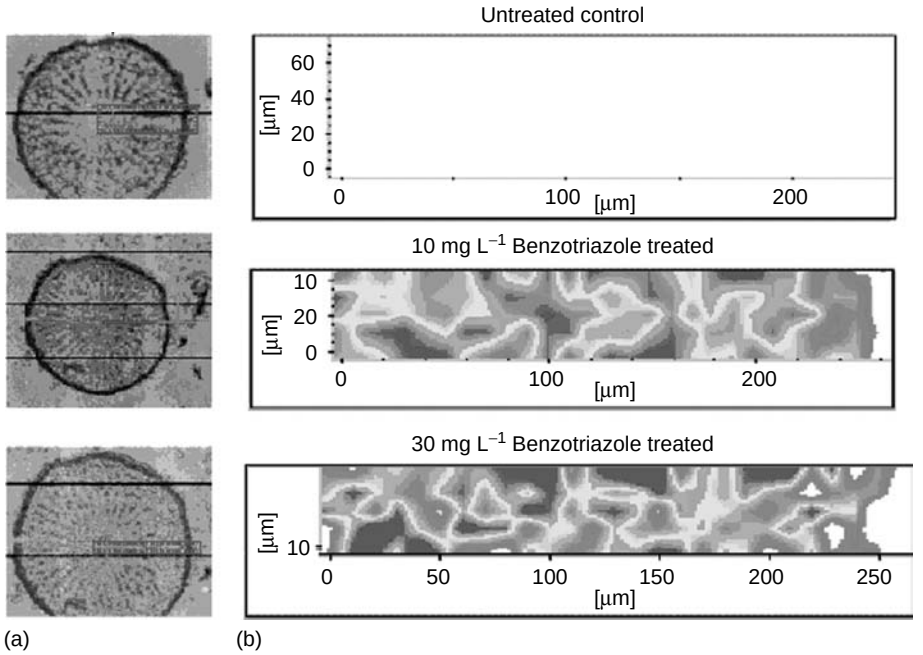

Fig. 8 (a) Visible images of root sections of a control plant and plants treated with 10 and 30 mg L^{-1} benzotriazole. Blue rectangles represent the area mapped by synchrotron IR microspectroscopy. (b) False-color intensity maps of untreated (control) and treated sunflower root sections, processed for the aromatic C—H out-of-plane bending mode of benzotriazole at 745 cm^{-1}. The intensity of the band in the inner (cortex) area of the secondary root is proportional to the initial concentration of the benzotriazole in which the plant was grown. (Reproduced by permission from International Scientific Communication from Marinkovic, N.S., Dokken, K.M., Davis, L.C., Linkovs, D.H., Flinn, J.M., Chance, M.R. (2005) Synchrotron infrared spectroscopy at beamline U2B, *Am. Biotechnology Lab.* **23**, 12–13) (see color plate p. xxxv).

Major spectral variability was observed in 180 data points, chosen from the infrared spectra of each 10 mg L^{-1}, 30 mg L^{-1} BT-treated and untreated roots (Fig. 9B and C). 89% of spectral variability accounted for changes in protein (1650 and 1553 cm^{-1}), benzotriazole peak at 745 cm^{-1}, carbohydrate (1102, 1039 cm^{-1}) and lignin peak (1250 cm^{-1}), revealing that the BT treatment causes damage to the proteins and/or alters their production in the plant tissue, while facilitating production of carbohydrate and possibly altering lignin structure.

6.5 IR Microscopy of Hair Sections

Analysis of human hair is done extensively. In cross-section, human hair is 50 to 100 μm in diameter, and consists of three distinguished regions, that is, the outer region (cuticle), inner region (cortex), and central part (medulla). The cuticle is a 5 μm-thick dense layer that protects the hair fiber. The cortex that is 45 to 90 μm thick determines the strength of the fiber and contains hair pigment (melanin), while loosely packed keratinized cells in 5

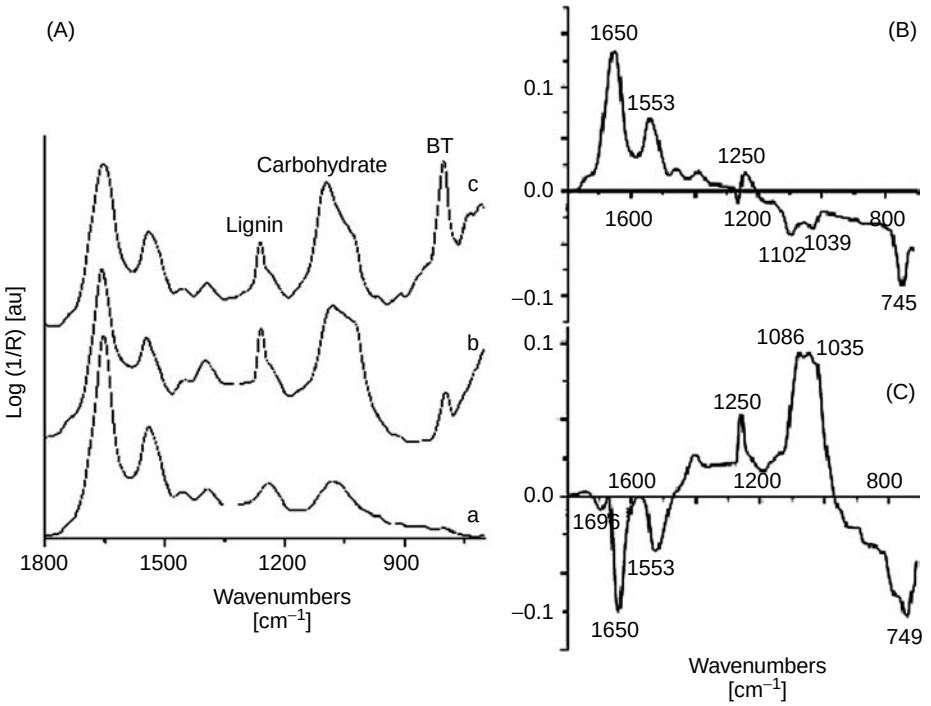

Fig. 9 (A) FTIR spectra of root tips from untreated control, 10 mg L^{-1}, and 30 mg L^{-1} benzotriazole-treated sunflower plants (a, b, and c, respectively). Each spectrum is an average of 10 spectra taken from arbitrarily selected locations from their respective IR maps. The spectra of treated plants roots show the appearance of the characteristic aromatic C—H peak of BT at 750 cm^{-1}, and differences in carbohydrate region at 1000–1100 cm^{-1} and lignin peak at 1250 cm^{-1}, with respect to the control. (B) The first principle component loading shows major variability in the protein amide I and II bands (1650 and 1553 cm^{-1}) and the characteristic BT peak at 745 cm^{-1}. (C) The second PC loading shows major variability in the protein and carbohydrate regions, as well as in lignin and BT peaks.

to 10 μm-thick medulla distributes moisture and nutrients to the hair strand. Even though infrared microspectroscopy of hair is now routinely used in forensic science and was applied for analysis of drug abuse in earlier studies with globar source, only with the high-brightness synchrotron FTIR microspectroscopy was it revealed that drugs are mainly localized in medulla. The latter technique revealed that Caucasian human hair contains fatty acids, mostly present in medulla and to a lesser extent in cuticle, whereas Afro-American hair showed no presence of these long-chain acids, in contrast to previous beliefs. First derivative spectra of hair showed a slight downward frequency shift for the symmetric CH_2 stretch, in agreement with the hypothesis of organized lipids inside the outer layer of hair, such as the skin stratum corneum. Amide I band is different in the three hair parts, suggesting differences in secondary structure composition of the proteins in each region. Chemical images of peptide and lipid bands in transverse hair cross-sections

before and after bleaching showed that keratin, the major protein component in hair containing high concentration of cysteine residues, becomes oxidized after the treatment, producing a sharp sulfonate stretching band around 1040 cm^{-1} in cortex, while medulla stays unaffected.

6.6 Other Biological Applications

Synchrotron IR microspectroscopy was used to show distribution of proteins, lipids, and nucleic acids in a single living cell through 3 × 3-μm aperture, and in a cell undergoing mitosis. Live cyanobacterial cells response to silicification was monitored to understand how microorganisms react to sudden and extreme changes in chemical and physical properties of their environment. Environmentally induced changes in the overall molecular framework of cells were studied during biogeochemical reduction of Cr(VI) on basalt surfaces on a bacterial colony on a magnetite surface during biodegradation of toxic polycyclic aromatic hydrocarbons, and on a micropopulation isolated from a eutrophic lake. Response of a single living human HepG2 cell derived from hepatocellular carcinoma in various concentrations of a model environmental contaminant for the aryl hydrocarbon receptor revealed changes in the phosphate backbone of nucleic acids and intracellular changes due to induction of the *CYP1A1* gene. Human lung fibroblast cells at different points in their life cycle revealed changes in secondary structure of proteins, variation of DNA/RNA content, and packing.

Hypercholesterolomic rabbit liver was probed by functional group mapping and nonsubjective cluster analysis, identifying regions of tissue with similar properties. Human skin on human explants were analyzed to show distribution of lipids inside stratum corneum, epidermis, and dermis, and to assess drug localization. The model drug, 4-cyanophenol with a distinct signature of $\nu_{C\equiv N}$ at 2222 cm^{-1} showed that it penetrates mostly into the stratum corneum, and extends into epidermis, and that the agent does not produce secondary structural modifications of the tissue during penetration.

Synchrotron FTIR spectroscopy was used to characterize organization of proteins in amyloid plaques from individuals with Alzheimer's disease (AD). The affected plaques showed appearance of β-sheet structure at 1630–1634 cm^{-1} inside amide I envelope, while the proteins in surrounding tissue contained mostly α-helical conformation. Classification of the IR spectroscopic data nonsubjectively by multivariate methods showed that IR spectroscopy could potentially be used in the diagnosis of AD from autopsy tissue. Combining two techniques in a single microscope, it was shown that β-amyloid plaques labeled by Thioflavin S could be identified by fluorescence microscopy, and analyzed by synchrotron IR microspectroscopy. Dietary enhancement of iron in drinking water was shown to significantly alter disease progression in the transgenic APP2576 murine model of AD in mice, altering the protein conformation in hippocampal neurons. It is likely that the change of protein secondary structure leads to aggregation of the misfolded protein.

7 Conclusion

Synchrotron infrared microspectroscopy is a powerful method for obtaining

high signal-to-noise spectra at diffraction-limited spot sizes and providing diffraction-limited spatial resolution. The high-brightness advantage of synchrotron source that is typically of 3 orders of magnitude higher than the thermal sources has had a significant impact in biological and biomedical research. In particular, mid-infrared vibrational modes of tissue chemistry can reveal small but significant chemical changes associated with various diseases and altered environmental conditions. The applications of synchrotron IR microspectroscopy are exemplified in neuroscience, bone research, and plant science, showing a variety of possible uses of this rapidly growing technique.

See also Alzheimer's Disease; Molecular Neurobiology, Single-Cell; Nucleic Acid and Protein Single Molecule Detection and Characterization.

Bibliography

Books and Reviews

Dumas, P., Miller, L.M. (2003) Biological and biomedical applications of synchrotron infrared microspectroscopy, *J. Biol. Phys.* **29**, 201–218.

Jackson, M. Mantsch, H.H. (1996) Medical Challenge to Infrared Spectroscopy in: Mantsch, H.H., Chapman, D. (Eds.) *Infrared Spectroscopy of Biomolecules*, Wiley-Liss, Toronto, Canada, pp. 311–340.

Jackson, M., Mantsch, H.H., (2000) Ex-Vivo Analysis by Infrared Spectroscopy in: Myers, R.A. (Ed.) *Encyclopedia of Analytical Chemistry*, Wiley, Chichester, England, pp. 131–151.

Marinkovic, N.S., Huang, R., Bromberg, P., Sullivan, M., Toomey, J., Miller, L.M., Sperber, E., Moshe, S., Jones, K.W., Chouparova, E., Lappi, S., Franzen, S., Chance, M.R. (2002) Center for Synchrotron Biosciences' U2B Beamline: an internal resource for biological infrared spectroscopy, *J. Synchrotron Radiat.* **9**, 189–197.

Messerschmidt, R.G. (1995) in: Huecki, H.J. (Ed.) *Practical Guide to Infrared Microspectroscopy*, Marcel Dekker, New York, pp. 3–39.

Messerschmidt, R.G., Harthcock, M.A. (Eds.) (1988) *Infrared Microscopy: Theory and Applications*, Marcel Dekker, New York.

Wetzel, D.L. (1993) in: Charalambous, G. (Ed.) *Food Flavors, Ingredients and Composition*, Elsevier Science, Amsterdam, NY, pp. 679–728.

Wetzel, D.L., Fulcher, R.G. (1990) in: Charalambous, G. (Ed.) *Food Flavors and Off Flavors*, Elsevier Science, Amsterdam, NY, pp. 485–510.

Wetzel, D.L., Levine, S.M. (2001) Biological Applications of Infrared Microspectroscopy, in: Gremlich, H.-U., Yan, B. (Eds.) *Infrared and Raman Spectroscopy of Biological Materials*, Marcel Dekker, New York, pp. 101–142.

Web Pages

Corn guide to hybrid selection for the Maritimes, (2004), available at:
http://www.gnb.ca/0316/03160004-e.pdf.

IR microscope vendors:
1. Bruker – http://www.brukeroptics.com;
2. Thermo-Nicolet – http://www.nicolet.com;
3. Biorad – http://www.bio-rad.com;
4. Perkin Elmer – http://www.instruments.perkinelmer.com

Kevley Technologies web page:
http://www.kevley.com
Variable Compensation Condensers:
http://www.thermo.com/com/cda/product/detail/1,1055,17977,00.html

Primary Literature

Bantignies, J.L., Fuchs, G., Carr, G.L., Williams, G.P., Lutz, D., Marul, S. (1998) Organic reagent interaction with hair spatially characterized by infrared microspectroscopy using synchrotron radiation, *Int. J. Cosmet. Sci.* **20**, 381–394.

Benning, L.G., Phoenix, V.R., Yee, N., Tobin, M.J. (2004) Molecular characterization of cyanobacterial silicification using synchrotron

infrared microspectroscopy, *Geochim. Cosmochim. Acta* **68**, 729–741.

Benning, L.G., Phoenix, V., Yee, N., Tobin, M.J., Konhauser, K.O., Mountain, B.W. (2002) Molecular characterization of cyanobacterial cells during silification: A synchrotron-based infrared study, *Geochem. Earth Surf.* **6**, 259–263.

Budevska, B.O. (2001) Vibrational Spectroscopy Imaging of Agricultural Products, in: Chalmers, J.M., Griffiths, P.R. (Eds.) *Handbook of Vibrational Spectroscopy*, Vol .5, John Wiley and Sons, Inc. New York, pp. 3720–3732.

Camacho, N.P., Rinnerthaler, S., Paschalis, E.P., Mendelsohn, R., Boskey, A.L., Fratzl, P. (1999) Complementary information on bone ultrastructure from scanning small angle X-ray scattering and Fourier transform infrared microspectroscopy, *Bone* **25**, 287–293.

Charriaut-Marlangue, C., Aggoun-Zouaoui, D., Represa, A., Ben-Ari, Y. (1996) Apoptotic features of selective neuronal death in ischemia epilepsy and gp 120 toxicity, *Trends Neurosci.* **19**, 109–114.

Choo, L.-P., Wetzel, D.L., Halliday, W.C., Jackson, M., Levine, S.M., Mantsch, H.H. (1996) *In situ* characterization of b-amyloid in Alzheimer's diseased tissue by synchrotron Fourier transform infrared microspectroscopy, *Biophys. J.* **71**, 1672–1679.

Choo, L.-P., Mansfield, J.R., Pizzi, M., Somorjai, R.I., Jackson, M., Halliday, W.C., Mantsch, H.H. (1995) Infrared spectra if human central nervous system tissue: Diagnosis of Alzheimer's disease by multivariate analysis, *Biospectroscopy* **1**, 141–148.

Coates, V.J., Offner, A., Siegler, E.H. Jr. (1953) Design and performance of an infrared microscope attachment, *J. Opt. Soc. Am.* **43**, 984.

Cooley, J.W., Tukey, J.W. (1965) An algorithm for machine calculation of complex Fourier series, *Math.Comput* **19**, 297.

Dokken, K.M., Davis, L.C., Erickson, L.E., Castro-Diaz, S., Marinkovic, N.S. (2005) Synchrotron Fourier transform infrared microspectroscopy: a novel tool to monitor the fate of organic contaminants in plants, *Microchem. J.* in press.

Dumas, P., Miller, L.M. (2003b) The use of synchrotron infrared microspectroscopy in biological and biomedical investigations, *Vib. Spectrosc.* **32**, 3–21.

Dumas, P., Jamin, N., Teillaud, J.L., Miller, L.M., Beccard, B. (2004) Imaging capabilities of synchrotron infrared spectroscopy, *Faraday Discuss.* **126**, 289–302.

Duncan, W.D., Williams, G.P. (1983) Infrared synchrotron radiation from electron storage rings, *Appl. Opt.* **22**, 2914.

Gazi, E., Dwyer, J., Lockyer, N.P., Miyan, J., Gardner, P., Hart, C., Brown, B., Clarke, N.W. (2004) Fixation protocols for subcellular imaging by synchrotron-based Fourier transform infrared microspectroscopy, *Biopolymers* **77**, 18–30.

Griffiths, P.R., de Haseth, J.A. (1986) *Fourier Transform Infrared Spectrometry*, John Wiley and Sons, New York.

Hansen, W.N. (1973) in: Muller, H. (Ed.) *Advances in Electrochemistry and Electrochemical Engineering*, Vol. 9, John Wiley and Sons, New York.

Holman, H.-Y.N., Goth-Goldstein, R., Martin, M.C., Russell, M.L., McKinney, W.R. (2000) Low-dose responses to 2.3.7.8-tetrachlorodibenzo-p-dioxin in single living human cells measured by synchrotron infrared spectromicroscopy, *Environ. Sci. Technol.* **34**, 2513–2517.

Holman, H.-Y.N., Martin, M.C., Blakely, E.A., Bjornstad, K., McKinney, W.R. (2000) IR spectroscopic characteristics of cell cycle and cell death probed by synchrotron radiation based Fourier transform IR spectromicroscopy, *Biopolymers* **57**, 329–335.

Holman, H.-Y.N., Perry, D.L., Martin, M.C., Lamble, G.M., McKinney, W.R., Hunter-Cevera, J.C. (1999) Real-time characterization of biogeochemical reduction of Cr(VI) on basalt surfaces by SR-FTIR imaging, *Geomicrobiol. J.* **16**, 307–324.

Holman, H.-Y.N., Nieman, K., Sorensen, D.L., Miller, C.D., Martin, M.C., Borch, T., McKinney, W.R., Sims, R.S. (2002) Catalysis of PAH Biodegradation by hymic acid shown in synchrotron infrared studies, *Environ. Sci. Technol.* **36**, 1276–1280.

Huang, R.Y., Miller, L.M., Carlson, C.S., Chance, M.R. (2002) Characterization of bone mineral composition in the Proximal Tibia of Cynomolgus monkeys: The effect of ovariectomy and nandrolone decanoate treatment, *Bone* **30**, 492–497.

Huang, R.Y., Miller, L.M., Carlson, C.S., Chance, M.R. (2003) Alterations of bone chemistry

in ovariectomized monkeys revealed by synchrotron infrared microSpectroscopy, *Bone* **33**, 514–521.

Jackson, M., Mantsch, H.H. (1995) The use and misuse of FTIR spectrometry in the determination of protein structure, *Crit. Rev. Biochem. Mol. Biol.* **30**, 95–120.

Jackson, M., Ranmjiawan, B., Mantsch, H.H. (1998) Functional group imaging of the jaundiced rabbit liver, *Cell. Mol. Biol.* **44**, 89–98.

Jamin, N., Dumas, P., Moncuit, J., Fridman, W.H., Teillaud, J.L., Carr, G.L., Williams, G.P. (1998a) Chemical imaging of nucleic acids, proteins, and lipids of a single living cell. Application of synchrotron infrared microspectroscopy in cell biology, *Cell. Mol. Biol.* **44**, 9–13.

Jamin, N., Dumas, P., Moncuit, J., Fridman, W.H., Teillaud, J.L., Carr, G.L., Williams, G.P. (1998b) Highly resolved chemical imaging of living cells by using synchrotron infrared microspectroscopy, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4837–4840.

Kalasinsky, K.S. (1998) Drug distribution of human hair by infrared microspectroscopy, *Cell. Mol. Biol.* **44**, 81–87.

Kalasinsky, K.S., Magluilo, J., Shaefer, T. (1993) Hair analysis for drugs of abuse, *Forensic Sci. Int.* **63**, 253–260.

Kemsley, E.K. (1998) *Discriminant Analysis and Class Modeling of Spectroscopic Data*, John Wiley and Sons Ltd., Chichester, England.

Kreplak, I., Briki, F., Doucet, J., Merigoux, C., Leroy, F., Leveque, J.L., Miller, L.M., Carr, G.L., Williams, G.P., Dumas, P. (2001) Profiling lipids across Caucasian and Afro-American hair transverse cuts, using synchrotron infrared microspectroscopy, *Int. J. Cosmet. Sci.* **23**, 369–374.

Lin-Vien, D., Colthup, N.B., Fateley, W.G., Grasselli, J.G. (1991) *Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*, Academic Press, Boston, MA.

Marinkovic, N.S., Adzic, A.R., Sullivan., M., Kovacs, K., Rousseau, D.L., Yeh, S.-R., Chance, M.R. (2000) Design and implementation of a rapid mixer flow cell for time-resolved infrared microspectroscopy, *Rev. Sci. Instrum.* **71**, 405–4060.

Marinkovic, N.S., Dokken, K.M., Davis, L.C., Linkous, D.H., Flinn, J.M., Chance, M.R. (2005) Synchrotron infrared spectroscopy at beamline U2B, *Am. Biotechnol. Lab.* **23**, 12–13.

Martin, M.C., Tsvetkova, N.M., Crowe, J.H., McKinney, W.R. (2001) Negligible sample heating from synchrotron infrared beam, *Appl. Spectrosc.* **55**, 111–113.

Michelson, A.A. (1902) *Light Waves and their Uses*, University of Chicago Press, Chicago.

Miller, L.M., Huang, R., Chance, M.R. (1999) Applications of fluorescence-assisted infrared microspectroscopy to the study of osteoporosis, *Synth. Rad. News* **12**, 21–27.

Miller, L.M., Smith, G.D., Carr, G.L. (2003) Synchrotron-based biological microspectroscopy: from the mid-infrared through the far-infrared regimes, *J. Biol. Phys.* **29**, 219–230.

Miller, L.M., Dumas, P., Jamin, N., Teillaud, J.-L., Miklossy, J., Forro, L. (2002) Combining IR spectroscopy and fluorescence imaging in a single microscope: biomedical applications using a synchrotron infrared source, *Rev. Sci. Instrum.* **73**, 1357–1360.

Miller, L.M., Vairavamurthy, V., Chance, M.R., Mendelsohn, R., Paschalis, E.P., Betts, F., Boskey, A.L. (2001) *In Situ* analysis of mineral content and crystallinity in bone using Infrared Microspectroscopy of the ν_4 PO_4^{3-} vibration, *Biochem. Biophys. Acta* **1527**, 11–19.

Nucara, A., Lupi, S., Calvani, P. (2004) The synchrotron infrared beamline SISSI at ELETTRA, *Infrared Phys. Technol.* **45**, 375–381.

Paschalis, E.P., DiCarlo, E., Betts, F., Sherman, P., Mendelsohn, R., Boskey, A.L. (1996) FTIR microscopic analysis of human osteonal bone, *Calcif. Tissue Int.* **59**, 480–487.

Piekut, D., Phipps, B., Pretel, S., Applegate, C. (1997) Effects of generalized convulsive seizures on corticotropin-releasing factor neuronal systems, *Brain Res.* **743**, 63–69.

Raab, T.K., Martin, M.C. (2001) Visualyzing rhizosphere chemistry of legumes with mid-infrared synchrotron radiation, *Planta* **213**, 881–887.

Raab, T.K., Vogel, J.P. (2004) Ecological and agricultural applications of synchrotron IR microscopy, *Infrared Phys. Technol.* **45**, 393–402.

Reffner, J.A., Hornlein, R.W. (1998) Experimental confirmation of sample area definition in infrared microspectrometry, *AIP Conference Proceedings* **430**, 425–427.

Roy, P., Brubach, J.-B., Calvani, P., deMarzi, G., Filabozzi, A., Gerschel, A., Giura, P., Lupi, S., Marcouille, O., Mermet, A., Nucara, A., Orphal, J., Paolone, A., Vervloet, M. (2001) Infrared synchrotron radiation: from the

production to the spectroscopic and microscopic applications, *Nucl. Instrum. Methods. Phys. Res., Sect. A* **467–468**, 426–436.

Schade, U., Peatman, W.B. (2003) The synchrotron activities at BESSY, *J. Biol. Phys.* **29**, 309–312.

Siege, D.C., Dean, A., Evado, E.L., Tobin, M.J. (2002) Fourier transform infrared spectroscopy of Pediastrum duplex: characterization of micropopulation isolated from a eutrophic lake, *Eur. J. Phycol.* **37**, 19–26.

Sloviter, R.S., Dean, E., Sollas, A.L., Goodman, J.H. (1996) Apoptosis and necrosis induced in different hippocampal neuron populations by repetitive perforant path stimulation in the rat, *J. Comp. Neurol.* **366**, 516–533.

Smith, T.I. (2002) The source issue in infrared microspectroscopy, *Nucl. Instrum. Methods. Phys. Res., Sect. A* **483**, 565–570.

Sperk, G. (1994) Kainic acid seizures in the rat, *Prog. Neurobiol.* **42**, 1–32.

Sperk, G., Gruber, B., Roder, C., Schwarzer, C. (1994) Neuropeptide Y inhibits glutamate release from rat hippocampal slices through the Ysub2 receptor, *Neuropeptides* **26**, 44.

Wetzel, D.L., Levine, S.M. (1999) Imaging molecular chemistry with infrared microscopy, *Science* **285**, 1224–1225.

Wetzel, D.L., Srivarin, P., Finney, J.R. (2003) Revealing protein infrared spectral detail in a heterogeneous matrix dominated by starch, *Vib. Specrosc.* **31**, 109–114.

Williams, G.P. (1982) The National Synchrotron Light Source in the infrared region, *Nucl. Instrum. Methods. Phys. Res., Sect. A* **195**, 383.

Williams, G.P., Dumas, P. (2001) Profiling lipids across Caucasian and Afro-American hair transverse cuts, using synchrotron infrared microspectroscopy, *Int. J. Cosmet. Sci.* **23**, 369–374.

Yu, P., McKinnon, J.J., Christensen, C.R., Christensen, D.A. (2004a) Using synchrotron transmission FTIR microspectroscopy as a rapid, direct, and nondestructive analytical technique to reveal molecular microstructural-chemical features within tissue in grain barley, *J. Agric. Food Chem.* **52**, 1484–1494.

Yu, P., McKinnon, J.J., Christensen, C.R., Christensen, D.A. (2004b) Imaging molecular chemistry of pioneer corn, *J. Agric. Food Chem.* **52**, 7345–7352.

Yu, P., McKinnon, J.J., Christensen, C.R., Christensen, D.A. (2004c) Using synchrotron-base FTIR microspectroscopy to reveal chemical features of feather protein secondary structure: comparison with other protein sources, *J. Agric. Food Chem.* **52**, 7353–7361.

Yu, P., McKinnon, J.J., Christensen, C.R., Christensen, D.A., Marinkovic, N.S., Miller, L.M. (2003) Chemical Imaging of microstructures within cellular dimension using synchrotron infrared microspectroscopy, *J. Agric. Food Chem.* **51**, 6062–6067.

Synthetic Peptide and Non-Peptide Combinatorial Libraries: *see* Peptide and Non-Peptide Combinatorial Libraries

Systematics and Evolution: *see* Molecular Systematics and Evolution

Syngamy and Cell Cycle Control

Michael Whitaker
University of Newcastle upon Tyne, Newcastle upon Tyne, UK

1 Formation of Gametes 6
1.1 Meiosis and Genetic Recombination 6
1.2 Male Gametes 7
1.3 Female Gametes 8

2 Interaction of Sperm and Egg 9
2.1 Sperm Activation and Chemotaxis 9
2.2 The Acrosome Reaction 10
2.3 Sperm–Egg Interaction and Speciation 11
2.4 Sperm–Egg Fusion and Sperm Incorporation 11
2.5 Syngamy 13

3 Fertilization Calcium Signals 14
3.1 Calcium Ions are a Universal Egg Activator at Fertilization 14
3.2 Calcium Signaling 14
3.3 Fertilization Calcium Waves 15
3.4 Repetitive Fertilization Calcium Transients 16
3.5 Signal Transduction during the Fertilization Calcium Wave 17
3.5.1 Reaction Diffusion Waves 17
3.5.2 Initiation of the Calcium Wave by the Sperm 18

4 Fertilization and the Cell Division Cycle 19
4.1 Fertilization Occurs at Different Stages of the Cell Division Cycle in Different Species 19
4.2 Maturation Promoting Factor Plays a Part in Meiosis and in Meiotic Arrest 19
4.3 Calcium Signals and Cell Cycle Progression at Fertilization 20
4.3.1 Metaphase Arrest 20
4.3.2 Interphase Arrest 21
4.4 The Cell Cycle and Syngamy 22

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

5 Calcium and Cell Cycle Control 23
5.1 Cell Cycle Checkpoints 23
5.2 The DNA Damage Checkpoint 24
5.3 The Mitosis Exit Checkpoint 24
5.4 The Checkpoint at Cell Cycle Onset 25
5.5 Calcium and Cell Cycle Checkpoints 25

Bibliography 25
Books and reviews 25
Primary Literature 26

Keywords

Actin
A small protein monomer that concatenates to produce thin filaments when it polymerizes.

Allosteric
Modifications to protein structure occur when the binding of a small molecule or another protein to one site in a protein leads to a conformational change in another part of the protein distant from the site of interaction.

Apoptosis
A process of programmed cell death in which cellular components are dismantled in orderly fashion without leakage of cell contents.

Centromeres
Stretches of DNA sequence generally toward the middle of a chromosome that attract accessory proteins that allow mechanical attachment of microtubules in the mitotic spindle and permit the movement of chromosomes into the daughter cells.

Chaperones
Proteins that bind to other proteins to enable them to fold correctly or to remain soluble.

Chromatin
DNA and its accessory proteins, found *condensed* in chromosomes during mitosis and *uncondensed* in the interphase nucleus, where it is surrounded by a double nuclear membrane.

Cyclic AMP
A messenger molecule within cells that activates a protein kinase (protein kinase A); it is made in a single step from ATP by adenylyl cyclase under the control of transmembrane receptors of hormones or growth factors.

Cytokinesis
The process by which the daughter cells separate following mitosis; it involves the actin-based contraction of a purse stringlike contractile ring that generates the cleavage furrow between the daughter cells.

Depolarization
The usually transient loss of the electrical potential that cells maintain across their plasma membranes by setting up asymmetric ionic concentration gradients; it acts as a signal, above all in nerve and muscle.

Disjunction
The separation of the chromosomes at the start of anaphase in mitosis and meiosis; disjunction allows the chromosomes to be pulled into the daughter cells.

Echinoderms
A phylum that sits at the base of the vertebrate lineage to which sea urchins and sea stars belong.

Enzymes
Proteins that act as catalysts in chemical reactions.

Glycoprotein
Protein that has been modified after translation by the addition of sugars, often branching chains of mannose and glucosamine derivatives.

Histones
Very basic (positively charged) proteins that bind to and order negatively charged DNA.

Homologs
The name given to the complementary chromosome pairs that result from recombination.

Hydrophobic
Molecules prefer nonaqueous environments.

Imprinting
An epigenetic modification of key genes that occurs in placental mammals; it regulates embryonic growth.

Interphase
Comprises all stages of the cell division cycle except mitosis and cytokinesis.

Introns
Stretches of DNA within a gene that do not encode amino acids and are cut out before protein sequence is decoded from messenger RNA.

Kinases
Enzymes that attach phosphate groups to other molecules. Protein kinases turn other proteins on and off in this way.

Lipid bilayers
Separate the cell from the extracellular milieu and also separate the cell's internal compartments; they consist of cylindrical lipid molecules with a long hydrophobic tail and a water-loving head group that assemble into planar sheets, lipid tails inwards.

Microtubules
Long cylindrical tubes that assemble from tubulin protein dimers; the dimers pack in a helical array.

Mitosis
The process that precisely segregates the genes between daughter cells after gene duplication; DNA becomes very tightly coiled and packed into chromosomes that are assembled into a disc and then pulled apart into the daughter cells by microtubules.

NADPH
Nicotine-adenine dinucleotide phosphate in its reduced form; it is used by cells as a cofactor to build molecules as this requires a reducing agent.

Phospholipases
Enzymes that cleave phospholipids and remove some or all of the headgroup.

Polyamines
Large linear molecules with many positively charged amine groups.

Polymorphism
The term given to small sequence differences between the same gene in different individuals within a species; many polymorphisms show no obvious phenotype, while at the other end of the spectrum some polymorphisms give rise to debilitating diseases such as cystic fibrosis and muscular dystrophy; the sum of all polymorphisms account for the genetic differences between individuals.

Proteoglycans
Very large proteins found in the extracellular matrix that have an amino acid backbone and very long carbohydrate side chains.

Proteolysis
The process of protein destruction; all proteins eventually undergo proteolysis, but in some cases, proteins are rapidly destroyed as part of a regulatory mechanism rapidly inactivate their activity; proteins can be cleaved into large fragments by proteases and into very short peptides by a cytoplasmic machine known as the *proteasome*.

S-phase
The phase in the cell division cycle during which DNA synthesis occurs and the genome is duplicated.

Spindle
The mitotic *spindle* is the array of microtubules that attaches to the chromosomes during mitosis and drags the chromosomes into their respective daughter cells; in its shape, this structure resembles spun thread wrapped around its spindle.

Sympatric
Species are found in the same or overlapping geographical areas.

Vesicles
Spherical compartments bounded by a lipid bilayer membrane; their contents are often released into the extracellular milieu; this is achieved by fusion of the limiting bilayer membrane with the plasma membrane.

The key aspect of sexual reproduction is the generation of genetic variation through the swapping of variable lengths of chromosome arms between pairs of homologous chromosomes during meiotic recombination in the germline. Following recombination, it is the general rule that a single (haploid) set of chromosomes is packaged into each male and female gamete (sperm and egg). These two sets of chromosomes, one from each parent, then come together to form a single nucleus: the process of syngamy. This simple outcome is achieved in a complex way through the process of fertilization. Fertilization comprises the delivery of the male set of chromosomes, tightly packaged in a small and motile sperm, to the female set of chromosomes, contained within a much larger nutrient laden and relatively sessile egg or oocyte; at fertilization, fusion of the two gametes occurs as a prelude to syngamy. In addition, fertilization triggers the onset of the very rapid rounds of cell division that quickly produce a very large number of embryonic cells. The cell division cycle must be coordinated with the events of syngamy. Large intracellular calcium signals occur at fertilization and ensure that the events of the cell cycle are controlled to permit successful melding of the parents' genomes.

1
Formation of Gametes

1.1
Meiosis and Genetic Recombination

Each round of sexual reproduction shuffles the genetic pack giving rise to a unique combination of genetic alleles and thus to a unique individual. The shuffling is brought about by recombination, a form of genetic cutting and pasting. Human genetic information is packaged into 22 pairs of fully recombining chromosomes (the autosomes); the X and Y sex chromosomes recombine only in a limited (pseudoautosomal) region and are ignored in this account. Each of the 22 autosomes differs in the genes it carries and in its shape because chromosomes have two arms, each of which varies in length between autosome pairs. The autosome pairs look identical to one another, like twins. But unlike identical (monozygotic) twins, though they carry the same genes, they are not genetically identical: each member of the pair carries different alleles of the same genes. Recombination cuts one or more of the autosome arms in each member of the autosome pair at the same place (in fact almost always at exactly the same nucleotide position in the DNA on the two chromosomes) and then swaps the fragments, pasting them back on the opposite chromosome (Fig. 1a).

Recombination occurs during meiosis, the specialized form of chromosome separation that is believed to have evolved from the more normal mitosis seen each time a somatic cell divides (It has recently been suggested by Krylov, Nasmyth, and Koonin that meiosis may be the phylogenetically primitive form of chromosome separation.). Meiosis is at the center of the mechanism of sexual reproduction and occurs only once as each gamete is formed. A germ cell destined to become a sperm or egg contains 22 autosomes from the father and 22 from the mother (the 22 autosome pairs). Just before meiosis, the germ cell undergoes a round of DNA synthesis. This generates 44 chromosome pairs, consisting of 44 autosomes with the father's alleles (the paternal homologs) and 44 autosomes with the mother's alleles (the maternal homologs). Each homolog pair is joined at the centromere. The outcome of meiosis is a haploid gamete with 22 nonidentical autosomes, so the 88 autosomes at the outset of meiosis must be packaged into four cells. This requires two rounds of cell division (Fig. 1b). The first meiotic division (meiosis I) is very unusual. As a consequence of recombination, pairs of recombined maternal/paternal homologs are formed, attached at the points on the arms at which recombination has occurred (the chiasmata). Chromosomes are pulled apart during meiosis I by microtubule-based forces, just as in mitosis. As the chromosomes line up in the central spindle under the tension generated by the microtubules of the meiotic spindle attached to the centromeres, they are seen to be arranged as pairs of pairs (tetrads). The first pairing reflects the attachments between sister chromatids of both maternal and paternal homologs established during DNA synthesis at the centromere and along the chromosomal arms; the second pairing is due to the chiasmata that are a consequence of recombination. It is a unique feature of meiosis I that when the spindle microtubules pull the chromosomes apart during anaphase, the chromosomes separate at the arms (synaptonemes), not at the centromeres. This results in detachment of homologs at the chiasmata (a process known as resolution) and disjunction of the homolog pairs, leaving the sister chomatid pairs

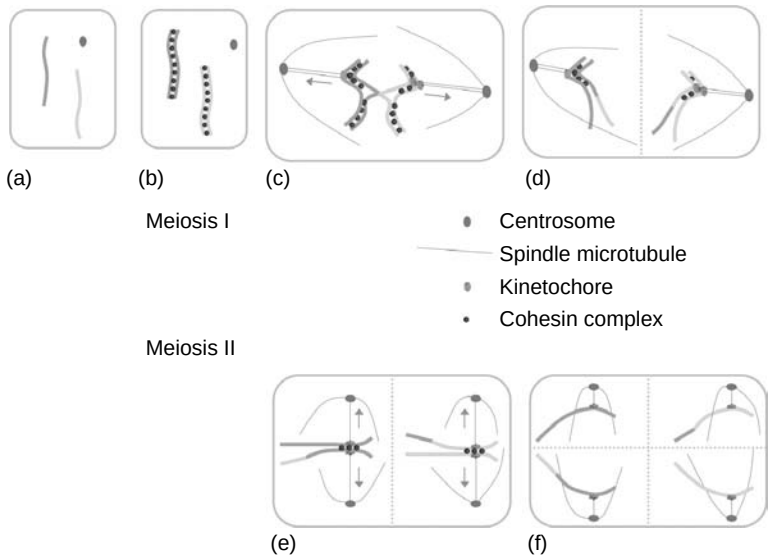

Fig. 1 Recombination and chromatid separation during meiosis. Before meiosis begins, the diploid germ cell contains pairs of homologous chromosomes; for simplicity, only one chromosome pair is illustrated (a). Both chromosomes in each pair undergo DNA synthesis (b); the newly replicated chromosomes are held together along their length by the protein cohesin. Then a complex series of events occurs during recombination that results in the swapping of chromosome arms between the replicated pairs (not shown). The outcome of these recombination events is revealed during the first meiotic division (c): chromosomes aligned on the meiotic spindle can be seen as tetrads, linked by the chiasm that is formed at the site of recombination. At anaphase cohesin falls off the chromosome arms as the chromosomes are pulled toward the centrosomes that organize the meiotic spindle (d); cohesin persists around the kinetochore, maintaining the association between sister chromatids. The meiotic spindle reforms during second meiosis (e). This time cohesin is lost at the kinetochore during anaphase (f). The result is four haploid cells, two of which have swapped chromosome arms in this illustration. With thanks to Neil Hunter for the building blocks of this illustration.

attached at the centrosome. The second meiotic division takes place without any further genome duplication through DNA synthesis. It involves a conventional chromosomal separation at the centromeres, the outcome being four haploid cells.

1.2 Male Gametes

In males, each of the four haploid outcomes of meiosis becomes a gamete. Driven presumably by a strong selection pressure (an unsuccessful gamete has a wasted life and an individual who produces unsuccessful gametes will have no offspring), male gametes have evolved as small, higher energetic cells. As they mature, they lose most of their cytoplasm, develop very highly condensed chromatin due to the replacement of histones with polyamines and grow a very motile tail (flagellum). They also retain a secretory vesicle (the acrosome) whose purpose is

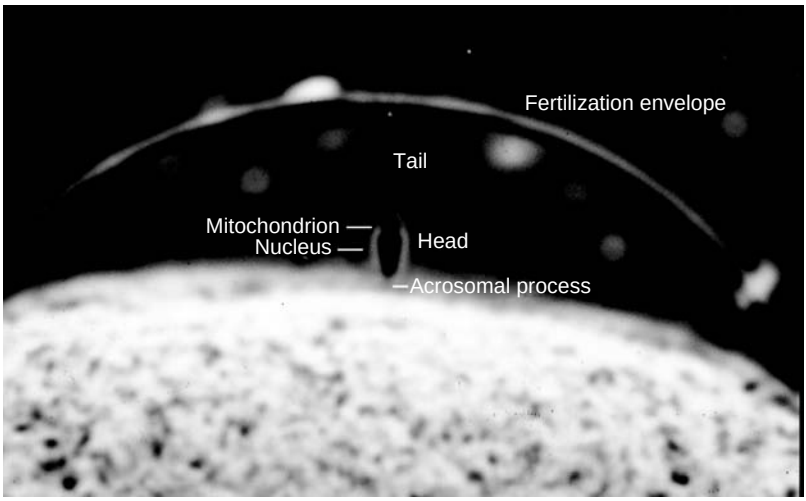

Fig. 2 The spermatozoon as it fertilizes the egg. Light microscope image of a sperm as it fertilizes the sea urchin egg. The sperm has fused with the egg through its acrosomal process and started the calcium wave. The local increase in calcium has caused the local elevation of the fertilization envelope that will exclude other sperm. The sperm head contains the DNA-containing nucleus and the mitochondrion that provides energy to the tail. Once the sperm has fused with the egg, the tail stops beating and stands straight as the sperm enters the egg. The sperm is only 2–3 microns across; the egg is 100 microns in diameter. The disparity in the sizes of the gametes is very obvious in this image.

to secrete digestive enzymes at the time that the sperm is passing through the extracellular matrix surrounding the egg and a mitochondrion or mitochondria to provide the ATP required for motility (Fig. 2).

1.3 Female Gametes

All embryos undergo a period of autonomous development and during this period they require a source of metabolites. As we have seen, sperm offer slim pickings; it is the egg that has specialized to contain a nutrient store (yolk) on which the embryo will live until it can itself feed or undergo uterine implantation. As a consequence, the embryo can grow only a little bigger than the egg itself (by, for example, developing fluid-filled internal cavities). The size of the egg thus determines the size of the autonomous embryo. Chicks undergo autonomous development within an eggshell to a very large size and thus have a very large yolk store. Frog eggs are around 1 mm in diameter and give rise to a macroscopic tadpole. Many eggs, including mammalian eggs are in the 0.1 mm range, with correspondingly smaller microscopic autonomous embryos, but still weigh in at ten times the size and one thousand times the volume of most somatic cells.

Size matters in two ways. First, it sets a constraint when the male and female gametes meet, as the much smaller sperm has had to evolve alongside mechanisms that signal its presence to the relatively enormous egg. We shall return to this.

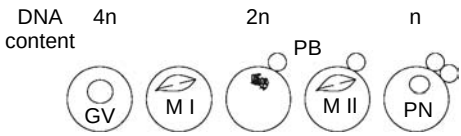

Fig. 3 The female oocyte during meiotic maturation. The immature oocyte is readily identified by its large nucleus (germinal vesicle, GV); by this point, DNA synthesis and recombination have taken place and the oocyte contains four copies of the genome (4n). The first meiotic division (meiosis I, MI) results in the extrusion of a polar body (PB); the oocyte now contains a 2n genomic complement. After second meiosis, a second polar body is formed and the oocyte reforms a haploid (n) pronucleus (PN).

Second, it has made its mark on the way the egg undergoes meiosis.

Strictly speaking, it is an oocyte that undergoes meiosis, the oocyte becoming an egg only once meiosis is complete. In most cases, the oocyte is already a very large cell by the time meiosis occurs, pumped up with yolk by the surrounding nurse cells in the ovary. In the female germline, both meiotic divisions are very asymmetric, both generating a very small cell, budded as it were on the surface of the large oocyte. After meiosis I, the small cell (polar body) contains 46 chromosomes and does not divide further. The polar body extruded during meiosis II contains 23 chromosomes and again has an inconsequential fate (Fig. 3). Thus, of the four possible haploid cells that can be generated by meiosis, only the large oocyte survives in the female germline.

2 Interaction of Sperm and Egg

2.1 Sperm Activation and Chemotaxis

Active sperm have a very limited shelf life. Sperm of external spawners shed into the water column have no ready access to respiratory substrates beyond their own limited stores and sperm of this sort cease to swim within minutes. Mammalian sperm within the uterus are supplied with external substrates; they also have many more mitochondria and can survive as motile and fecund sperm for several days. Mechanisms therefore exist to suppress sperm motility until sperm are delivered to the vicinity of the egg.

A widespread mechanism for suppression of sperm activation appears to be the acidic environment in testis and seminal fluid, combined with a high partial pressure of carbon dioxide. Once the seminal fluid is diluted, pH rapidly increases and activates motility.

Sperm activation may also be mediated or enhanced by signaling molecules released by the egg or present in the uterus. In echinoderm sperm, small peptides released from the egg coat interact with their receptor, a transmembrane guanylate cyclase located in the sperm tail. Stimulation of the receptor increases both pH and calcium concentrations within the sperm, leading to enhanced motility (Fig. 4). Similarly, progesterone stimulates calcium signals in mammalian sperm and enhances motility.

These agents may also be inducing chemotaxis, the directed movement of sperm toward the egg along a chemical gradient, though this is not yet proven. In other animals, for example, ascidians and siphonophores, sperm chemotaxis has been well demonstrated. Chemotaxis between gametes of the fern can be demonstrated using simple kits designed for high school students. It has recently

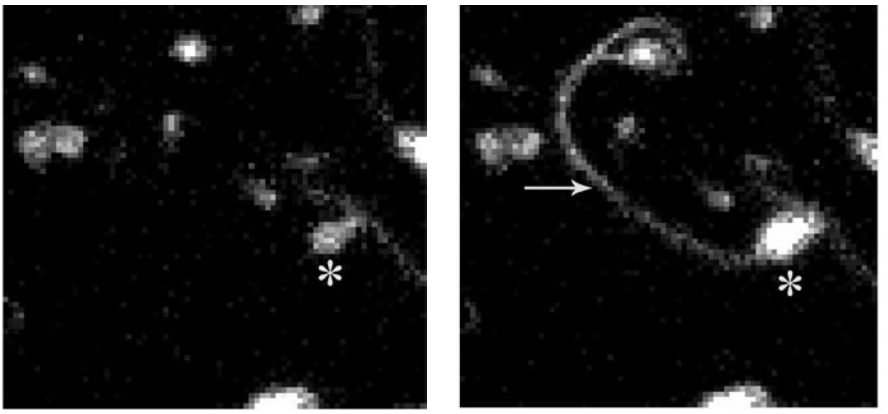

Fig. 4 Calcium signaling and sperm chemotaxis. Calcium ion concentrations can be detected inside cells by using fluorescent dyes whose intensity depends on calcium concentration; sperm head is marked with an asterisk. Sperm have low resting calcium concentrations (left). When the chemotactic peptide speract is added, calcium concentration increases in the tail (arrow); this alters swimming behavior.

been found that human sperm possess receptors equivalent to those found in the olfactory epithelium and that appropriate odorants can induce changes in swimming behavior consistent with chemotaxis.

2.2 The Acrosome Reaction

The naked surface of the egg or oocyte is always well protected, sometimes by thick proteoglycan coats and often in addition with a cloud of accessory cells embedded in this extracellular matrix. Sperm must traverse these barriers in order to reach the plasma membrane and enter the egg and to do so they must dissolve the extracellular matrix. The release of the contents of the acrosomal vesicle is a process akin to the release of neurotransmitters and circulating hormones: a large increase in intracellular calcium concentration triggers fusion of the acrosomal vesicle with the sperm plasma membrane. The acrosomal vesicle contains glycanases and proteases in many cases, capable of hydrolyzing the extracellular matrix.

Fittingly, the release of acrosomal contents is triggered by the proximity of the extracellular coats themselves. In echinoderms, the acrosome reaction can be experimentally triggered by a component of egg jelly in the complete absence of the egg itself. In mammalian sperm, the trigger to the acrosome reaction is a glycoprotein component of the extracellular matrix immediately surrounding the egg, the zona pellucida.

In abalone, the contents of the acrosomal vesicle consist of a protein, lysin, which acts not as an enzyme, but stoichiometrically as a structurally specific chaotropic agent (chaperone) that dissolves the egg vitelline envelope.

One consequence of the acrosome reaction in many species is the formation of the acrosomal process, a fine protrusion at the tip of the sperm head that is the locus of interaction with the egg membrane. This fact is exemplified by the behavior of *Thyone* sperm: they undergo

an acrosome reaction at the very periphery of the egg jelly, sending a 100-μm process through the jelly coat to reach the egg plasma membrane. In *Thyone*, as in other species, the process grows due to actin polymerization. The growing plus end of the actin filaments is at the tip of the acrosomal process while the reservoir of monomeric actin lies at the base of the filament; the rate of growth decreases exponentially, limited by the increasing diffusion time for actin monomers along the length of the growing process. Actin polymerization is triggered by the increase in pH that accompanies the calcium rise that triggers the acrosome reaction.

2.3 Sperm–Egg Interaction and Speciation

Reproductive isolation is the major mechanism driving the formation of new species. This often occurs as a result of geographic isolation of a population, but can also occur sympatrically. Closely related sympatric species may remain speciated by separation of habitat or by asynchronous spawning, but reproductive isolation is very often due to rapid evolution of the molecules that are the gatekeepers for syngamy.

A striking example is the abalone lysin released during the acrosome reaction. There are five species of abalone living sympatrically on the Pacific coast of North America. Their reproductive isolation is due in large part to the species specificity of the lysin reaction: lysins from other related abalone species are much less potent at dissolving the vitelline envelope. Lysins show very little polymorphism within species. Nonconservative substitutions within the protein are very high relative both to housekeeping proteins and to lysin introns, implying very strong positive selection. Lysin interacts with a protein designated VERL in the vitelline envelope. VERL is a very large (>1 m kDa) protein that consists of well conserved ~100 amino acid repeats. Analysis indicates that it is not under strong positive selection. The hypothesis put forward is that VERL undergoes a process of concerted evolution that puts very strong evolutionary pressure on lysin that can lead to speciation. Concerted evolution is a phenomenon found in proteins with multiple repeats. A mutation in one repeat can propagate through all repeats as a result of preferential recombination events. The finding that lysin shows a very low polymorphism within a population indicates that positive selection produced an optimal lysin in response to concerted evolution in VERL that then swept through a population, creating a new species.

Lysin/VERL is the only known example of how sperm/egg receptor interaction may govern speciation. Very little is known about other sperm/egg receptors. It has been suggested that an 18-kDa protein that coats the acrosomal process in abalone (a close relative of lysin, but with little sequence similarity) may be the sperm receptor for the egg plasma membrane. Similarly, the very hydrophobic protein bindin, which coats the echinoderm acrosomal process, may interact with a receptor on the egg plasma membrane. In mammalian eggs, it has been suggested that integrin/disintegrin ligand receptor pairs that are known to be involved in cell–matrix interaction and signaling may be the sperm/egg ligand/receptor at fertilization, but the evidence is contradictory.

2.4 Sperm–Egg Fusion and Sperm Incorporation

Once the sperm and egg plasma membranes are in close proximity, they fuse

to establish continuity of egg and sperm cytoplasm. Cell–cell fusion in general is not well understood and gamete fusion is no exception. The protein bindin that coats the acrosomal process in echinoderms is very hydrophobic, as are the cores of lipid bilayers: it is thought that bindin may promote membrane fusion, not least because it has been found to promote fusion of lipid vesicles *in vitro*. The 18-kDa protein of abalone has been ascribed a similar function. In mammalian fertilization, the integrin/disintegrin interaction may be fusogenic, but there is no sound evidence for this.

It is humbling that we do not yet understand the mechanism of sperm–egg fusion: it represents the moment of fertilization. A variety of observations isolate the event of sperm–egg fusion as the key event at fertilization:

- In echinoderms, sperm–egg fusion causes depolarization of the egg, the first response observed at fertilization that, significantly, much reduces the probability of other sperm fusing with the egg (an electrical block to polyspermy).
- In echinoderms, under certain conditions, sperm–egg fusion is reversible; when it is reversed, eggs may or may not undergo activation and the longer the sperm cytoplasm is in continuity with the egg, the greater the chances of egg activation.
- Sperm–egg fusion is the earliest detectable event at fertilization, as measured by dye transfer between egg and sperm.
- Mammalian sperm carry an enzyme in their cytoplasm that can activate the egg and is delivered to the egg when the sperm fuses with it.

Once the sperm fuses with the egg, it has fulfilled its purpose of finding the egg and fusing with in local competition with tens (mammals) or millions (external spawners) of other sperm. Incorporation

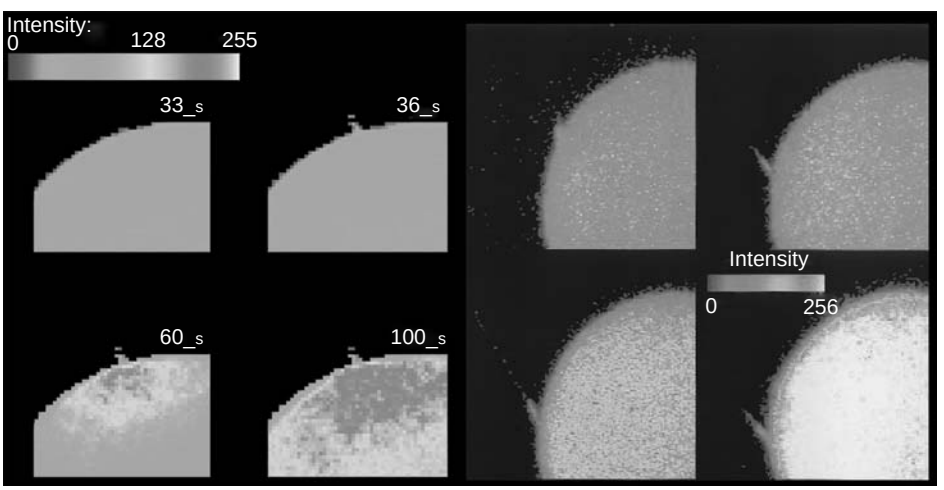

Fig. 5 Sperm–egg fusion in sea urchin and mouse eggs. The eggs contain a fluorescent dye sensitive to calcium concentration. When the sperm fuses with the egg, dye enters the sperm. Sperm–egg fusion triggers an increase of calcium within the egg. The calcium increase occurs many seconds after sperm–egg fusion. Left: sea urchin. Right: mouse.

of the sperm into the egg relies on the egg itself (Fig. 5).

Immediately after sperm–egg fusion, the connection between egg and sperm is tenuous. In many species, fusion occurs at the tip of an acrosomal process little more than 100 nm in diameter. Even during mammalian sperm–egg fusion where it is the side of the sperm head that fuses with the egg plasma membrane, the area of contact is no more than 1 μm^2. Movement of the sperm head (and eventually sperm tail) into the egg is due to actin-based traction. Actin is a force generating protein found in all cells whose acme is rapid and coordinated force generation in skeletal muscle. Actin filaments surround the sperm head, forming an extrusion from the egg surface known as a fertilization cone.

Actin filaments drag the sperm nucleus into the egg. Mitochondria also enter the egg cytoplasm. These degenerate; it is clear by observation and from the genetics of mitochondrial inheritance that all mitochondria in offspring are inherited from the egg and so from the mother. The sperm tail is incorporated. Its components dissolve in the egg cytoplasm, save for the centriole, an organelle that in the sperm organizes the tail microtubules but that, incorporated into the egg, takes over the role of organizing the mitotic spindle microtubules.

Once inside the egg, the behavior of the sperm nucleus depends upon the cell cycle stage of the egg that the sperm has fertilized.

2.5 Syngamy

The blind aim of sexual reproduction is to ensure that one sperm pronucleus coalesces with one female pronucleus. Coalescence of more than the complementary pair of pronuclei leads to genetic chaos. The penetration of more than one sperm nucleus into an egg is known as polyspermy and its consequences are usually fatal to the embryo.

Mechanisms exist to exclude supernumerary sperm. The electrical block to polyspermy (depolarization), found in echinoderms and amphibians, has already been mentioned: the fertilizing sperm depolarizes the egg, making it (for unknown reasons) much harder for subsequently arriving sperm to fuse with the egg. The electrical block is very fast and can come into play within 10 ms of sperm–egg fusion. Other polyspermy blocks are effective, but slower, requiring tens of seconds to take effect. A common mechanism comprises the secretion of the contents of secretory vesicles from the egg after fertilization. In many cases, this leads to the elaboration of a shell (vitelline envelope) around the fertilized egg that sperm cannot penetrate. In mammalian eggs, the secretions lead to inactivation of the glycoprotein that triggers the acrosome reaction in incoming sperm; sperm that have not undergone the acrosome reaction cannot fuse with the egg. In fish eggs, sperm must enter in single file down a narrow canal in the egg investments (the micropyle). Once the egg is activated by the fertilizing sperm, the secretions plug the canal, obstructing the progress of further sperm.

Not all strategies to ensure one-to-one fusion of the sperm and egg nucleus rely on sperm exclusion. Bird eggs are usually polyspermic. Once a lucky sperm nucleus has fused with the egg nucleus, all the other sperm nuclei present in the egg degenerate. In ctenophores, more than luck seems to be in play. The ctenophore egg is naturally polyspermic. The egg nucleus roams from one sperm nucleus

to the next, eventually fusing with one of them; the basis on which this choice is made is unknown.

Once the sperm nucleus has entered the egg it becomes surrounded by a nuclear envelope membrane and is known as the male pronucleus; its counterpart in the egg is the female pronucleus.

As the male pronucleus forms, it swells. The increase in pronuclear volume is due to chromatin decondensation consequent on the replacement of the polyamines surrounding sperm DNA with egg histones.

Ultimately, male and female pronuclei congress and fuse to realize syngamy. Congression of the pronuclei occurs because the male pronucleus is closely associated with the sperm centrosome: it organizes a microtubule array (the aster) that spreads through the egg cytoplasm and captures the female pronucleus, drawing it toward the male pronucleus, while the growing aster pushes against the egg cortex to move both male and female pronucleus toward the center of the egg.

3
Fertilization Calcium Signals

3.1
Calcium Ions are a Universal Egg Activator at Fertilization

In the early part of the twentieth century, Jacques Loeb experimented with various ways of stimulating the egg to undergo development in the absence of sperm – parthenogenesis. Treatment of echinoderm eggs with butyric acid followed by hypertonic seawater was effective, proving Loeb's antivitalist point that life was chemistry but offering no insight into the underlying mechanisms that governed the onset of development.

In the 1970s, a sharp tool became available that enabled these mechanisms to be defined: a calcium ionophore with the designation A23187. Calcium ionophores are small molecules that can carry calcium ions across lipid membranes. It was shown that treatment of unfertilized eggs of both vertebrates and invertebrates with A23187 led to many of the sequelae of fertilization: secretion of cortical secretory granules, DNA synthesis, and migration of the female pronucleus. The eggs had been activated, but did not divide. What was missing? It turned out to be the centrosomes that are required to form the mitotic spindle and that are donated to the egg by the sperm at fertilization. When centrosomes were induced to form in the echinoderm egg using hypertonic seawater, full development ensued in many cases and these parthenotes could be raised to adulthood.

These investigations set calcium ions center stage at fertilization.

3.2
Calcium Signaling

Calcium ions are the bearers of messages delivered by universal signaling systems found in all higher eukaryotes. Calcium ion concentrations are very low in the cytoplasm of all cells and are below micromolar concentrations at rest. It is postulated that calcium ion concentrations must of necessity be very low in the cytoplasm because of the existence of high concentrations of inorganic phosphate, itself a key component of intermediary metabolism. High concentrations of calcium and phosphate cannot coexist because of the low solubility product of calcium phosphate salt. On the other hand, calcium concentrations are in the millimolar range in the seas and in the body fluids of metazoans,

leading to very large gradients of calcium concentration across cell membranes. The existence of these gradients has two consequences. One is that calcium ion pumps driven by the cell's metabolism must constantly eject calcium from the cell since no lipid membrane is completely impermeant to calcium. The second is that relatively small calcium fluxes across the plasma membrane can give rise to large and rapid changes in the intracellular calcium concentrations. The latter consequence makes the cytoplasmic calcium concentration very suitable as an energy-efficient cell signal.

Calcium fluxes across plasma membranes control neurotransmitter release and are the basis of all neuronal activity. In skeletal muscle, calcium causes contraction but is released into the cytoplasm not from the extracellular fluid but from a pervasive network of reticular internal membranes that form a compartment known as the sarcoplasmic reticulum. This arrangement allows calcium ion concentrations to increase rapidly throughout the cytoplasm without the diffusion delays that arise when calcium enters through the plasma membrane. The reticular network is present in other cell types, where it is known as endoplasmic reticulum (ER). It exists not solely as a source of calcium signals; in fact, its primary purpose is to permit the folding and export of proteins destined for the plasma membrane either as integral membrane proteins or as secreted products. The concentration of calcium ions in ER is closer to the concentrations in extracellular fluid than to that of the cytoplasm and is maintained by a calcium pump. We know that folding and export of ER proteins is compromised when the calcium concentration in the ER falls, so it is reasonable to suppose that the compartment's high calcium concentration exists to provide an environment for proteins comparable to that they will experience once they leave the cell. But, as with the plasma membrane calcium gradient, the ER calcium gradient has been exploited in evolution as a calcium signaling system.

3.3 Fertilization Calcium Waves

When the calcium ionophore is used to activate an unfertilized egg, it is causing the release of calcium from the internal calcium stores of the ER. This is easily demonstrated by removal of all calcium external to the egg: this has no effect on egg activation. It is also possible, though rather more complicated, to demonstrate that calcium ion concentrations increase at fertilization.

The first demonstrations of fertilization calcium increases employed a protein (aequorin) that emits light in response to calcium in a graded way. The increase in calcium at fertilization could be seen as a ring of light travelling through the peripheral cytoplasm of a large medaka fish egg. This established the basic features of the fertilization calcium signal as a large increase in cytoplasmic calcium that arises at the point of sperm–egg fusion and traverses the egg as a constant velocity reaction-diffusion wave. Similar waves were observed in frog, echinoderm, nematode, and mammalian eggs using both aequorin and more convenient fluorescent calcium indicators developed by Dr Roger Tsien (Fig. 6).

Fertilization calcium waves are the largest and longest calcium signals measured. They rise to around 5–10 μM and last for many minutes. Preparation for these large explosive signals takes place

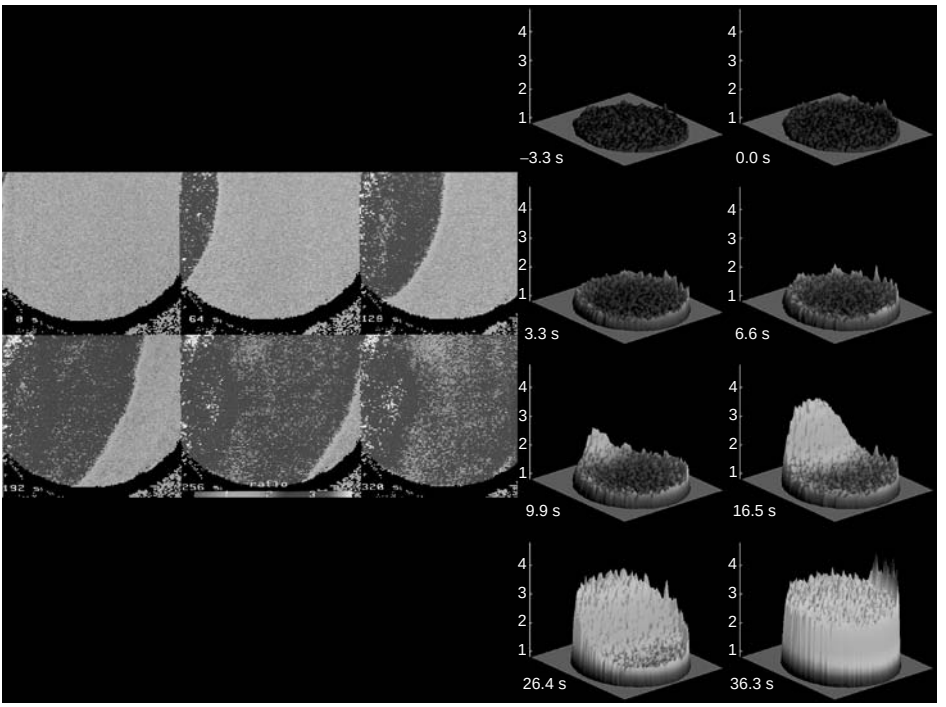

Fig. 6 Fertilization calcium waves in sea urchin and frog eggs. Calcium sensitive dyes are used to detect calcium waves. Warm colors reflect high calcium concentrations. The calcium wave crosses the sea urchin egg in around 15 s with a velocity of around 5 μm s^{-1}. The wave has a similar velocity in the frog egg, but here it takes 5 min to traverse the egg as it is over 1 mm in diameter. Left: frog egg. Right: sea urchin egg. A small early increase in calcium concentration can be seen at the egg periphery, a consequence of a calcium action potential that is triggered by sperm–egg fusion.

during meiosis. As meiosis progresses, the oocyte becomes competent to generate a large calcium wave; after fertilization, this competence disappears. The calcium ionophore experiments tell us that the waves' purpose is to initiate development and it is reasonable to suppose that their extent, magnitude, and duration are necessary to ensure that each part of the egg cytoplasm receives the message that fertilization has occurred. They ensure that the small and local interaction that occurs between egg and sperm is amplified and transmitted throughout the egg.

3.4 Repetitive Fertilization Calcium Transients

Both ascidian and mammalian oocytes manifest multiple large calcium transients at fertilization. In ascidian oocytes, the locus of initiation of the calcium wave migrates with the male pronucleus as it moves toward the oocyte's vegetal pole. In mammalian oocytes, the first few calcium transients originate at the point of sperm – egg fusion, but thereafter, each calcium transient appears to rise homogeneously within the oocyte cytoplasm. This behavior can be explained by proposing

that each transient is generated by a factor associated with the sperm and male pronucleus. In ascidian oocytes, it is supposed that this factor remains closely associated with the male pronucleus, while in mammalian oocytes, the factor slowly diffuses away from the male pronucleus to become uniformly distributed in the cytoplasm. We shall consider the likely purpose of the repetitive transients when we come to discuss the link between fertilization and the cell division cycle.

3.5 Signal Transduction during the Fertilization Calcium Wave

The question before us is how sperm–egg fusion, an event that initially involves a cytoplasmic bridge as little as 100 nm wide, is able to trigger the large fertilization calcium wave. An analogy is that the fertilization calcium wave is an explosion and that the sperm lights a fuse to set it off. This leads to two further questions: what gives the calcium wave its explosive properties and how is the fuse lit? We shall answer them in turn.

3.5.1 Reaction Diffusion Waves

Explosions occur because the products of a chemical reaction further increase the rate of the reaction itself in a process of positive feedback. The fuel for the calcium explosion at fertilization is the calcium stored within the ER. Positive feedback occurs because of the particular properties of calcium channels that exist in the ER membrane: the probability of their opening increases when cytoplasmic calcium increases. Once calcium rises at a point within the cytoplasm, a wave of calcium release will propagate from that point (Fig. 7). There is good evidence for this mechanism in heart cells made to propagate artificial calcium waves.

The identity of the calcium channel in the ER responsible for the fertilization calcium wave was determined by

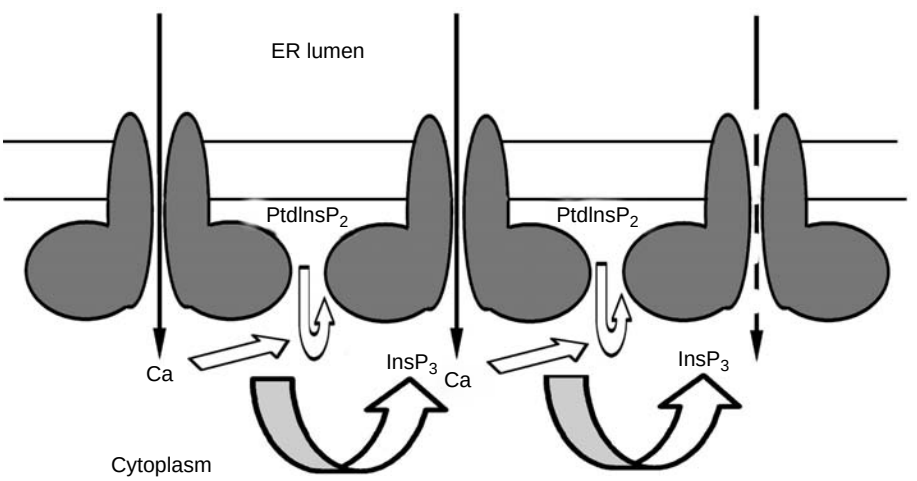

Fig. 7 Mechanism of propagation of the fertilization calcium wave. The calcium wave propagates via calcium-dependent production of $InsP_3$ and diffusion of calcium from one release channel to the next. Both $InsP_3$ and calcium are required in the vicinity of the channel to trigger channel opening and movement of calcium from the ER into the cytoplasm.

investigation of the phosphoinositide messenger pathway, a ubiquitous signaling pathway in cells of higher eukaryotes. The substrate of this pathway is phosphatidylinositol bisphosphate ($PtdInsP_2$). A signaling event, for example, the binding of a hormone or neurotransmitter to its receptor, leads to activation of a specific phospholipase C that hydrolyzes $PtdInsP_2$ to produce two products, inositol trisphosphate ($InsP_3$) and diacylglycerol. $InsP_3$ is a small sugar that causes calcium release from ER. It achieves this by binding to and allosterically activating an ER calcium channel known as the $InsP_3$ receptor ($InsP_3R$), causing it to open. Calcium in the cytoplasm is a coactivator of the $InsP_3R$. The sequence in which $InsP_3$ and calcium are presented to the channel is crucial: if calcium binds after $InsP_3$, then the channel is activated, but if it binds before, $InsP_3$ is unable to open the channel.

Calcium waves can be triggered in eggs by injecting very small amounts of $InsP_3$ into unfertilized eggs and can be blocked by agents that prevent $InsP_3$ from interacting fruitfully with the $InsP_3R$. These observations demonstrate that the $InsP_3R$ is the calcium channel responsible for propagation of the fertilization calcium wave, but they also suggest that the mechanism by which the wave propagates is more complex than that established for calcium waves in heart cells.

The properties of the $InsP_3R$ are not immediately consistent with the observation that very small quantities of $InsP_3$ are sufficient to trigger a calcium wave since once the locally injected $InsP_3$ is diluted by diffusion into the large volume of the egg, its concentration is insufficient to activate the $InsP_3R$ which, as we have seen, requires coactivation by both $InsP_3$ and calcium. There is evidence, in fact, that increases in calcium induce further hydrolysis of $PtdInsP_2$, generating $InsP_3$ locally as the wave propagates. The mechanism of wave propagation thus relies on two intertwined positive feedback pathways, calcium-induced production of $InsP_3$ and coactivation of the $InsP_3R$ by $InsP_3$ and calcium (Fig. 7).

3.5.2 Initiation of the Calcium Wave by the Sperm

The simplest idea is that the sperm generates a small amount of $InsP_3$, so setting off the calcium wave in much the same way as happens when $InsP_3$ is microinjected. This is almost certainly the case, though there is no direct evidence for an increase in $InsP_3$ in the tiny volume close to the point of sperm–egg fusion.

The evidence that an increase in $InsP_3$ triggers the fertilization calcium wave is more subtle. A significant amount of time elapses after sperm–egg fusion before the increase in cytoplasmic calcium that marks the onset of the wave is seen. In echinoderms, around 10–20 s elapses, comparable to the time it then takes for the wave to cross the egg; in mammalian eggs, several minutes pass. What is found is that an agent that interferes with the activation of the $InsP_3R$ increases this latency.

$InsP_3$ is generated by activation of $PtdInsP_2$-specific phospholipases. There has been a debate about which of these phospholipases is responsible for triggering the calcium wave. It was initially thought that an interaction of the sperm with an egg receptor might activate the β-form of phospholipase C (PLCβ); PLCβ is known to be coupled to transmembrane receptors in all cell types examined. However, as it became clearer that sperm–egg fusion was the key activating event at fertilization, attention shifted to another form of phospholipase, PLCγ, that is activated

by phosphorylation on tyrosine. Protein fragments that were able to compete with and block either PLCγ or tyrosine (Src family) kinases were found to increase the latency of the fertilization response and at higher concentrations, to block it completely. Mammalian eggs were, however, an exception, being completely insensitive to these inhibitors. On the assumption that the activating phospholipase would be present in sperm, the mammalian testis was screened for the existence of appropriate phospholipases. A novel phospholipase found only in testis (PLCζ) was shown to induce repetitive calcium transients when expressed in unfertilized mouse eggs.

The picture we have of signal transduction at fertilization is that a protein (Src kinase, PLCγ, PLCζ, depending on species) passes into the egg when sperm and egg fuse. The presence of this protein in the egg leads to the very local production of $InsP_3$. As the latent period progresses, a concentration of $InsP_3$ builds up sufficiently to trigger the onset of the explosive fertilization calcium wave (Fig. 7). Observation of the repetitive fertilization calcium transients in ascidians and mammals suggests that in these species, the activating protein is bound to a component of the sperm head. In ascidian oocytes, the association persists as the male pronucleus forms, while in mammalian eggs, the protein (PLCζ) is thought to dissociate from the male pronucleus (see Sects. 3.4 and 4.3.1).

4 Fertilization and the Cell Division Cycle

4.1 Fertilization Occurs at Different Stages of the Cell Division Cycle in Different Species

It is striking given the ubiquity of calcium signals at fertilization that there are significant differences between eggs of different species in the cell cycle stage in meiosis at which fertilization occurs. Meiosis comprises two cell divisions without an intervening S-phase. Fertilization can occur before the first meiotic division begins, at metaphase of first meiosis, at metaphase of second meiosis, and in interphase once meiosis is complete (Fig. 8). With very few exceptions, this is because at the time that the egg is most receptive to sperm, its cell cycle has arrested at one of the above stages, a stage specific to the species in question.

4.2 Maturation Promoting Factor Plays a Part in Meiosis and in Meiotic Arrest

Classic experiments in the field of cell cycle control showed that cytoplasm from oocytes arrested and awaiting fertilization in meiosis I or meiosis II would induce the maturation of immature oocytes. Similarly, classical experiments in the genetics of yeast cell cycle control genes determined that two key components of

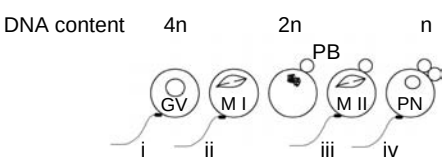

Fig. 8 Cell cycle arrest before fertilization. Oocyte of different species arrest before fertilization at different points of the meiotic division cycle. Some are fertilized as immature oocytes (i: clam), some during first meiotic metaphase (ii: other molluscs, ascidian), some during second meiotic metaphase (iii: frog, starfish, mammals) and some after meiosis at the start of the first mitotic cell division cycle (iv: sea urchin).

maturation promoting factor activity were cdk1, a protein kinase, and its activating partner, cyclin B.

Cdk1/cyclin B activation linked to accumulation of cyclin B drives the cell into meiosis or mitosis, causing condensation of chromatin and formation of the mitotic or meiotic spindle. Oocytes (and other cells) arrested at metaphase show high levels of cdk1/cyclin B activity, but what marks them out is the persistence of this normally transient activity: the persistence leads to cell cycle arrest.

Stabilization of cdk1/cyclin B activity during metaphase cell cycle arrest is due to what appear to be the two other components of maturation promoting factor, a protein termed mos and its target, MAP kinase. Mos is a protein unique to meiotic maturation. Its transcription early during maturation leads to increasing levels of MAP kinase activity. MAP kinase activity remains high throughout meiosis until fertilization occurs. It is responsible for the persistence of cdk1/cyclin B activity (Fig. 9) and, incidentally, for the suppression of S-phase between meiosis I and meiosis II.

Persistent activation of cdk1/cyclin B accounts for the two metaphase cell cycle arrest points during meiosis observed in various species, but other explanations are necessary to explain cell cycle arrest before meiosis begins or after it is completed.

Cell cycle arrest before the onset of meiosis is seen in most organisms. In most species, premeiotic arrest is a distinct stage preceding the prefertilization arrest that occurs later in meiosis and that has already been discussed earlier. Ovaries contain many immature oocytes, a proportion of which are recruited by hormonal stimulation to begin maturation. It is thought that hormonal stimulation leads to a fall in the second messenger cyclic adenosine monophosphate (cAMP). In species whose oocytes are fertilized when immature, it is the fertilization calcium signal that triggers the onset of maturation and reinitiation of meiosis (In some molluscs fertilized at this stage, calcium influx across the plasma membrane takes the place of a calcium wave.).

Cell cycle arrest in interphase of the first mitotic cell cycle occurs in some echinoderms. This arrest can be explained by the suppression of protein synthesis by a relatively acidic cytoplasmic pH and the maintenance of a redox state that lead to low levels of the synthetic nicotine-adenine dinucleotide phosphate (NADPH).

4.3 Calcium Signals and Cell Cycle Progression at Fertilization

4.3.1 Metaphase Arrest

Persistent cdk1/cyclin B activity driven by mos/MAP kinase is the cause of the cell cycle arrest at metaphase in oocytes. Inactivation of these kinases by removal of both cyclin B and mos is how the

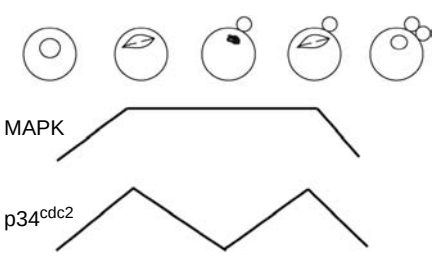

Fig. 9 MAP kinase and cdk1 during meiotic maturation. MAP kinase activity remains high throughout meiosis, while cdk1 activity dips between meiosis I and meiosis II.

fertilization calcium signal causes meiosis to resume.

Frog oocytes are arrested during meiosis II. The fertilization calcium wave triggers cyclin proteolysis by the proteasome through activating calmodulin and so calmodulin-dependent protein kinase II (CamKII). CamKII in turn activates another molecular machine known as the *anaphase promoting complex* (APC) that targets cyclin for destruction by the proteasome. A parallel mechanism, activation of the calcium-dependent protease calpain, leads to proteolysis of mos. In frog oocytes, the proximal cause of anaphase onset is the destruction of cyclin; destruction of mos occurs later than cyclin destruction. Mouse oocytes are also arrested at meiosis II awaiting fertilization. In mammals, 10 to 12 h elapses between fertilization and formation of the nuclear membrane around male and female pronuclei, the lengthy period coinciding with the process of imprinting, found only in mammals. During this period, calcium transients occur every few minutes and cyclin B is slowly degraded, the rate extent of cyclin degradation determined by the frequency of calcium signals. Formation of the pronuclear membranes marks the end of meiosis and coincides with the final disappearance of cyclin B. Mos, on the other hand, is degraded within a few hours. Ascidian oocytes are arrested before fertilization at metaphase of meiosis I. A first set of calcium transients occurs immediately after fertilization, leading to degradation of cyclin B and exit from meiosis I. Mos/MAP kinase activity is not affected by this first set of transients (remember that a role of mos/MAP kinase is to suppress S-phase in the interval between the two meiotic divisions). The repetitive transients begin again as the oocyte enters meiosis II, again leading to degradation of cyclin B and, on this occasion, degradation of mos too. The events described in this paragraph are summarized in Fig. 10.

The reader will have noted a correlation between the state of the nucleus and the occurrence of repetitive calcium transients: the transients are present during meiotic divisions when chromatin is condensed and the nuclear envelope is absent, and are absent during interphase when chromatin in decondensed and a nuclear envelope surrounds the membrane. The transients may indeed be triggered by a factor that is sequestered in the nucleus and released when the nuclear envelope breaks down as the oocyte enters meiotic division. The evidence for this is that nuclei isolated from first cell cycle interphase in mouse embryos will break down under the influence of cdk1/cyclin B when transferred to unfertilized mouse oocytes and induce repetitive calcium transients. The simplest explanation is that PLCζ, the egg activator from the sperm, is sequestered into the nucleus as the nuclear envelope reforms. This conjecture is consistent with observations that repetitive calcium transients continue indefinitely when import of proteins into the nucleus though the nuclear pores are blocked.

4.3.2 Interphase Arrest

Little more is known about how calcium lifts the cell cycle blockade after fertilization of immature oocytes than that calcium signals are known to be capable of antagonizing signaling by cyclic AMP.

Sea urchin eggs are arrested after completion of meiosis, in interphase of the first cell cycle. The fertilization calcium wave stimulates a calcium-dependent NAD^+ kinase to generate NADP, which is rapidly reduced in the cell to NADPH, providing fuel for protein synthesis. Activation of phospholipase C is known to generate

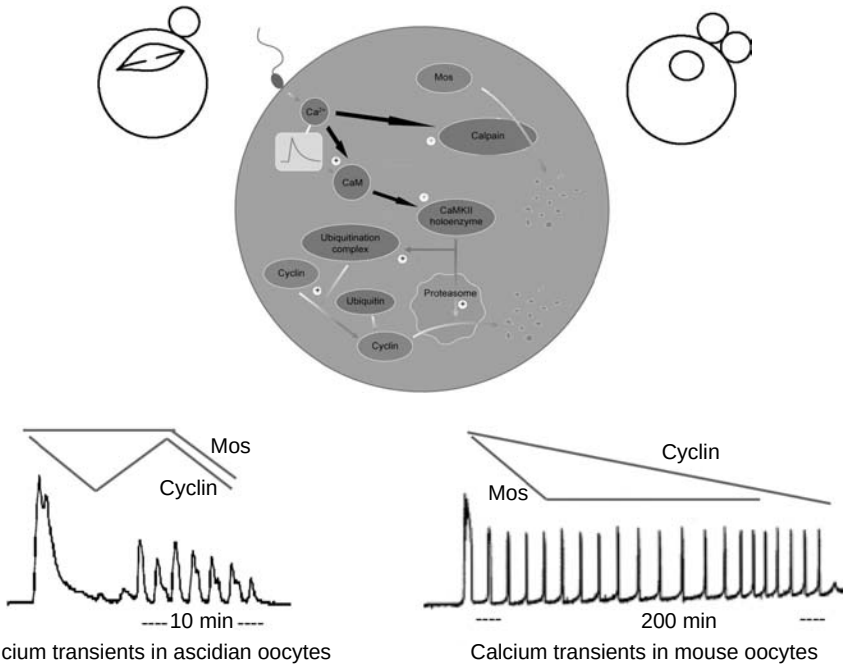

Fig. 10 Destruction of mos and cyclin B at fertilization. Two calcium-dependent pathways for proteolysis at fertilization. One degrades mos by activating calpain. The other activates CaMKinase II to stimulate the APC and cyclin degradation by the proteasome. Calcium dependent proteolysis is rapid in ascidians; in mouse and other mammals mos is rapidly degraded, but cyclin degradation takes several hours.

a diacylglycerol from $PtdInsP_2$ in addition to $InsP_3$ (Fig. 7). At fertilization in the sea urchin egg, the diacylglycerol so generated activates a protein kinase C (no relation of phospholipase C) in the egg plasma membrane in turn stimulating a sodium–hydrogen antiporter that ejects hydrogen ions from the egg, returning the cytoplasm to its normal pH of 7.2. The suppression of the protein synthetic machinery is lifted, cyclin synthesis begins, and the cell cycle proceeds.

4.4 The Cell Cycle and Syngamy

The behavior of the male pronucleus once inside the oocyte or egg after fertilization depends on the point at which fertilization occurs: during or after meiosis. Oocytes fertilized during meiosis must complete meiosis to generate a haploid female pronucleus with which the male pronucleus can undergo syngamy. In this case, the male pronucleus remains separate with its chromatin in a condensed state. Once meiosis II is completed, cyclin B/cdk1 activity falls and the male pronucleus develops a nuclear envelope, as does the female pronucleus. Syngamy comprises the mixing of male and female genomes and occurs through fusion of the pronuclei's nuclear envelopes. S-phase of the first mitotic cell cycle often begins before syngamy. In an extreme case, in mouse

zygotes, S-phase is completed and the nuclear envelope breaks down as the zygote enters the first mitotic division before syngamy occurs. Syngamy occurs on the metaphase plate as male and female chromosomes are captured by spindle microtubules.

5 Calcium and Cell Cycle Control

5.1 Cell Cycle Checkpoints

It is noteworthy that the same cytoplasmic calcium signal, the wave generated by the sperm, operates to trigger resumption of the cell cycle at three distinct points in the cell cycle in different organisms: G2/M, metaphase of MI and MII (which are formally equivalent) and G1. The question arises as to whether these fertilization mechanisms are specific examples of a more general requirement for calcium during cell cycle progression of all cell types.

More general study of the regulation of the cell cycle in yeast and in mammalian somatic cells as well as in embryos has led to the concept of cell cycle checkpoints. The concept is straightforward: before a cell makes a crucial step through the cell cycle, a checkpoint mechanism is invoked that is designed to ensure that everything is correctly in place, a preflight checklist as it were. Two examples of this are the DNA damage checkpoint and the mitosis exit checkpoint.

DNA replication is not a faultless process. Errors are introduced during copying of the DNA strands that are then repaired during late S-phase. The point here is that if mitosis were to occur before repair had taken place, then one or both daughter cells would inherit mutations that resulted from errors introduced at S-phase. The DNA damage checkpoint mechanism is invoked by damaged DNA; while it is active, the cell cannot proceed into mitosis.

The mitosis exit checkpoint is invoked during mitosis to ensure that all chromosome pairs are captured and properly aligned on the metaphase plate before they are separated by the processes of anaphase. A single misaligned chromosome will prevent mitosis. The checkpoint machinery monitors attachment of two sets of microtubules to the chromosome and also verifies by measuring tension in the microtubules that they are correctly attached to the spindle poles. In this instance, the checkpoint mechanism aims to avoid an unequal distribution of chromosomes into the daughter cells. This condition known as *aneuploidy* leads to cell death or on occasion to uncontrolled cell proliferation. A third checkpoint exists at the beginning of the cell cycle in G1. Here, a cell decides, on the basis of signals from its environment, whether it will differentiate, commit suicide (apoptosis), or progress into the cell cycle and divide. The cell cycle stage at which oocytes are found to arrest in the cell cycle correspond to the stages at which these checkpoints act.

Oocytes arrested in G2 are stopped at the stage at which the DNA damage checkpoint acts, those arrested at metaphase are stopped at the mitosis exit checkpoint stage, and those arrested in G1 resemble those cells making choices about whether to divide, die, or differentiate. It is thus very likely that cell cycle arrest before fertilization makes use of underlying cell cycle checkpoint mechanisms.

5.2 The DNA Damage Checkpoint

The DNA damage checkpoint operates by preventing activation of the mitotic kinase cdk1/cyclin B. It does this by preventing a phosphatase (cdc25) from removing a phosphate group on cdk1 that inhibits kinase function. It has been shown in mouse oocytes that activation of cdc25 is necessary and sufficient to bypass the G2/M block and trigger resumption of meiosis. In most oocytes, as mentioned earlier, resumption of meiosis is triggered by a fall in cyclic AMP (Fig. 11). In mouse oocytes, it has been shown that cdc25 is downstream of the reduction in cyclic AMP levels. The calcium transient in oocytes fertilized at G2/M may activate cdc25 via calmodulin-dependent protein kinase II, as it has been shown that progression through the G2/M checkpoint in mammalian somatic cells involves activation of this kinase.

5.3 The Mitosis Exit Checkpoint

The mitosis exit checkpoint operates by preventing activation of the APC that targets key proteins for destruction including securin and cyclin B. Loss of securin leads to loss of adhesion between paired chromosomes on the metaphase plate, allowing them to separate and spring to the poles; loss of cyclin B leads to inactivation of the cdk1/cyclin B kinase, permitting decondensation of chromatin and reformation of the nuclear envelope. These events are triggered by the fertilization calcium transient in oocytes arrested at metaphase meiosis I and meiosis II before fertilization.

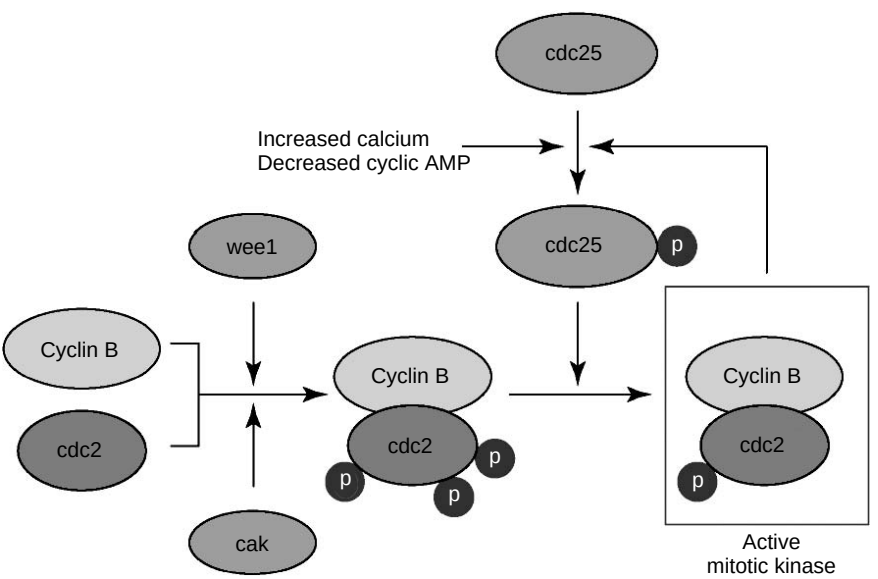

Fig. 11 Activation of mitotic kinase by the cdc25 phosphatase. Cdk activating kinase (cak) and wee1 kinase phosphorylate the cdk1/cyclin complex. Cak phosphorylation activates the mitotic kinase and wee1 phosphorylation is inhibitory. Cdc25 phosphatase removes the inhibitory phosphates. Cdc25 is controlled by cAMP and calcium.

5.4 The Checkpoint at Cell Cycle Onset

There are analogies between fertilization in species such as echinoderms that are fertilized in G1 and the G1 decision point in mammalian somatic cells. One striking example is starfish oocytes: if they do not sense a fertilization calcium transient, they proceed to apoptosis. Other similarities are the activation of an MAP kinase signaling cascade, stimulation of the sodium-hydrogen antiporter by protein kinase C, and stimulation of the protein synthetic machinery. Nonetheless, the comparison is less precise at the molecular level.

5.5 Calcium and Cell Cycle Checkpoints

The existence of identical molecular mechanisms in arrested oocytes and in the checkpoint mechanisms of somatic cells strongly suggests that the arrest before fertilization uses the checkpoint machinery and raises a question: is calcium signaling a component of general cell cycle regulation or is control of the checkpoint machinery by calcium unique to fertilization?

That calcium is a cell cycle regulatory signal is clear in some cases, and less clear in others. In some embryos, calcium signals can be identified at the times at which the checkpoint mechanisms operate; the amplitude of the signals is small, less than one-tenth of the size of the fertilization calcium signal and signals are often local to the nucleus. The calcium-calmodulin-calmodulin kinase II pathway has been shown to control entry into mitosis and calcium and calmodulin control separation of the chromosomes at mitosis. Embryos do not possess a G1 decision point mechanism.

In somatic cells, it has been shown, for example, that calmodulin-dependent protein kinase II must be activated at mitosis entry and that it phosphorylates cdc25, the activator of cdk1/cyclin B. It has also been shown that rapidly raising intracellular calcium can induce anaphase. Checkpoint-associated calcium signals can be detected, but they are not present in all circumstances. The jury is out on the universality of cell cycle calcium signaling, but if the hypothesis turns out to be true, then the calcium signals central to fertilization will have pointed the way.

See also Calcium Biochemistry; Female Reproduction System, Molecular Biology of; *In vitro* Fertilization; Male Reproductive System: Testis Development and Spermatogenesis.

Bibliography

Books and reviews

Berridge, M.J., Lipp, P., Bootman, M.D. (2000) The versatility and universality of calcium signalling, *Nat. Rev. Mol. Cell Biol.* **1**, 11–21.

Darszon, A., Labarca, P., Nishigaki, T., Espinosa, F. (1999) Ion channels in sperm physiology, *Physiol. Rev.* **79**, 481–510.

Eisenbach, M. (1999) Sperm chemotaxis, *Rev. Reprod.* **4**, 56–66.

Heilbrunn, L.V. (1928) *The Colloid Chemistry of Protoplasm*, Verlag von Gebrueder Borntraeger, Berlin, Germany.

Hepler, P.K. (1992) Calcium and mitosis, *Int. Rev. Cytol.* **138**, 239–268.

Jaffe, L.F. (1980) Calcium explosions as triggers of development, *Ann. N.Y. Acad. Sci.* **339**, 86–101.

Jaffe, L.A., Cross, N.L. (1986) Electrical regulation of sperm-egg fusion, *Annu. Rev. Physiol.* **48**, 191–200.

Jaffe, L.F., Creton, R. (1998) On the conservation of calcium wave speeds, *Cell Calcium* **24**, 1–8.
Jaffe, L.A., Giusti, A.F., Carroll, D.J., Foltz, K.R. (2001) Ca^{2+} signalling during fertilization of echinoderm eggs, *Semin. Cell Dev. Biol.* **12**, 45–51.
Kishimoto, T. (2003) Cell-cycle control during meiotic maturation, *Curr. Opin. Cell Biol.* **15**, 654–663.
Lillie, F.R. (1919) *Problems of Fertilization*, University of Chicago Press, Chicago, IL.
Loeb, J. (1913) *Artificial Parthenogenesis and Fertilization*, University of Chicago Press, Chicago, IL.
Maller, J.L. (1998) Recurring themes in oocyte maturation, *Biol. Cell* **90**, 453–460.
Rieder, C. (1998) *Mitosis and Meiosis*, Academic Press, New York, p. 504.
Rothschild, L. (1956) *Fertilization*, Methuen, London, England.
Schatten, G. (1994) The centrosome and its mode of inheritance: the reduction of the centrosome during gametogenesis and its restoration during fertilization, *Dev. Biol.* **165**, 299–335.
Stricker, S.A. (1999) Comparative biology of calcium signaling during fertilization and egg activation in animals, *Dev. Biol.* **211**, 157–176.
Wassarman, P.M. (1999) Fertilization in animals, *Dev. Genet.* **25**, 83–86.
Whitaker, M. (1996) Control of meiotic arrest, *Rev. Reprod.* **1**, 127–135.
Whitaker, M.J. (2005) Calcium signalling in eggs and embryos, *Physiol. Rev.* in press.
Whitaker, M.J., Steinhardt, R.A. (1982) Ionic regulation of egg activation, *Q. Rev. Biophys.* **15**, 593–666.
Whitaker, M.J., Patel, R. (1990) Calcium and cell cycle control, *Development* **108**, 525–542.
Whitaker, M., Swann, K. (1993) Lighting the fuse at fertilization, *Development* **117**, 1–12.

Primary Literature

Abassi, Y.A., Carroll, D.J., Giusti, A.F., Belton, R.J., Foltz, K.R. (2000) Evidence that Src-type tyrosine kinase activity is necessary for initiation of calcium release at fertilization in sea urchin eggs, *Dev. Biol.* **218**, 206–219.
Adkins, C.E., Taylor, C.W. (1999) Lateral inhibition of inositol 1,4,5-trisphosphate receptors by cytosolic Ca^{2+}, *Curr. Biol.: CB* **9**, 1115–1118.
Aizawa, H., Kawahara, H., Tanaka, K., Yokosawa, H. (1996) Activation of the proteasome during Xenopus egg activation implies a link between proteasome activation and intracellular calcium release, *Biochem. Biophys. Res. Commun.* **218**, 224–228.
Albrieux, M., Sardet, C., Villaz, M. (1997) The two intracellular Ca^{2+} release channels, ryanodine receptor and inositol 1,4,5-trisphosphate receptor, play different roles during fertilization in ascidians, *Dev. Biol.* **189**, 174–185.
Albrieux, M., Lee, H.C., Villaz, M. (1998) Calcium signaling by cyclic ADP-ribose, NAADP, and inositol trisphosphate are involved in distinct functions in ascidian oocytes, *J. Biol. Chem.* **273**, 14566–14574.
Allen, R.D., Griffin, J.L. (1958) The time sequence of early events in the fertilization of sea urchin eggs, *Exp. Cell Res.* **15**, 163–173.
Ayabe, T., Kopf, G.S., Schultz, R.M. (1995) Regulation of mouse egg activation: presence of ryanodine receptors and effects of microinjected ryanodine and cyclic ADP ribose on uninseminated and inseminated eggs, *Development (Cambridge, England)* **121**, 2233–2244.
Baitinger, C., Alderton, J., Poenie, M., Schulman, H., Steinhardt, R.A. (1990) Multifunctional Ca^{2+} /calmodulin-dependent protein kinase is necessary for nuclear-envelope breakdown, *J. Cell Biol.* **111**, 1763–1773.
Bandyopadhyay, J., Lee, J., Lee, J., Lee Jin, I., Yu, J.-R., Jee, C., Cho, J.-H., Jung, S., Lee Myon, H., Zannoni, S., Singson, A., Kim Do, H., Koo, H.-S., Ahnn, J. (2002) Calcineurin, a calcium/calmodulin-dependent protein phosphatase, is involved in movement, fertility, egg laying, and growth in *Caenorhabditis elegans*, *Mol. Biol. Cell* **13**, 3281–3293.
Begg, D.A., Rebhun, L.I. (1979) pH regulates the polymerization of actin in the sea urchin egg cortex, *J. Cell Biol.* **83**, 241–248.
Bement, W.M., Capco, D.G. (1989) Activators of protein kinase C trigger cortical granule exocytosis, cortical contraction, and cleavage furrow formation in Xenopus laevis oocytes and eggs, *J. Cell Biol.* **108**, 885–892.
Berridge, G., Dickinson, G., Parrington, J., Galione, A., Patel, S. (2002) Solubilization of receptors for the novel Ca2+-mobilizing messenger, nicotinic acid adenine

dinucleotide phosphate, *J. Biol. Chem.* **277**, 43717–43723.

Bloom, T.L., Szuts, E.Z., Eckberg, W.R. (1988) Inositol trisphosphate, inositol phospholipid metabolism, and germinal vesicle breakdown in surf clam oocytes, *Dev. Biol.* **129**, 532–540.

Brandriff, B., Hinegardner, R.I., Steinhardt, R. (1975) Development and life cycle of the parthenogenetically activated sea urchin embryo, *J. Exp. Zool.* **192**, 13–24.

Brind, S., Swann, K., Carroll, J. (2000) Inositol 1,4,5-trisphosphate receptors are downregulated in mouse oocytes in response to sperm or adenophostin A but not to increases in intracellular Ca^{2+} or egg activation, *Dev. Biol.* **223**, 251–265.

Busa, W.B., Nuccitelli, R. (1985) An elevated free cytosolic Ca^{2+} wave follows fertilization in eggs of the frog, *Xenopus laevis*, *J. Cell Biol.* **100**, 1325–1329.

Busa, W.B., Ferguson, J.E., Joseph, S.K., Williamson, J.R., Nuccitelli, R. (1985) Activation of frog (*Xenopus laevis*) eggs by inositol trisphosphate. I. Characterization of Ca^{2+} release from intracellular stores, *J. Cell Biol.* **101**, 677–682.

Campbell, K.D., Reed, W.A., White, K.L. (2000) Ability of integrins to mediate fertilization, intracellular calcium release, and parthenogenetic development in bovine oocytes, *Biol. Reprod.* **62**, 1702–1709.

Carrasco, D., Allende, C.C., Allende, J.E. (1990) The incorporation of myo-inositol into phosphatidylinositol derivatives is stimulated during hormone-induced meiotic maturation, *Exp. Cell Res.* **191**, 313–318.

Carroll, J., Swann, K. (1992) Spontaneous cytosolic calcium oscillations driven by inositol trisphosphate occur during in vitro maturation of mouse oocytes, *J. Biol. Chem.* **267**, 11196–11201.

Carroll, J., Swann, K., Whittingham, D., Whitaker, M. (1994) Spatiotemporal dynamics of intracellular $[Ca^{2+}]_i$ oscillations during the growth and meiotic maturation of mouse oocytes, *Development (Cambridge, England)* **120**, 3507–3517.

Carroll, D.J., Albay, D.T., Terasaki, M., Jaffe, L.A., Foltz, K.R. (1999) Identification of PLC gamma-dependent and -independent events during fertilization of sea urchin eggs, *Dev. Biol.* **206**, 232–247.

Carroll, D.J., Ramarao, C.S., Mehlmann, L.M., Roche, S., Terasaki, M., Jaffe, L.A. (1997) Calcium release at fertilization in starfish eggs is mediated by phospholipase Cgamma, *J. Cell Biol.* **138**, 1303–1311.

Carroll, M., Levasseur, M., Wood, C., Whitaker, M.J., Jones, K.T., McDougall, A. (2003) Exploring the mechanism of action of the sperm-triggered calcium-wave pacemaker in ascidian zygotes, *J. Cell. Sci.* **116**, 4997–5004.

Cavalier-Smith, T. (2002) Origins of the machinery of recombination and sex, *Heredity* **88**, 125–141.

Centonze, V.E., White, J.G. (1995) Calcium oscillations during early *C. elegans* development. http://elegans.swmed.edu/wli/[wm95e148]. (cited with permission).

Chambers, E.L. (1980) Fertilization and cleavage of the eggs of the sea urchin *Lytechinus variegatus* in Ca^{2+}-free sea water, *Eur. J. Cell Biol.* **22**, 476.

Chambers, E.L., de Armendi, J. (1979) Membrane potential, action potential and activation potential of eggs of the sea urchin, *Lytechinus variegatus*, *Exp. Cell Res.* **122**, 203–218.

Chambers, E.L., Pressman, B.C., Rose, B. (1974) The activation of sea urchin eggs by the divalent ionophores A23187 and X-537A, *Biochem. Biophys. Res. Commun.* **60**, 126–132.

Charbonneau, M., Moreau, M., Picheral, B., Vilain, J.P., Guerrier, P. (1983) Fertilization of amphibian eggs: a comparison of electrical responses between anurans and urodeles, *Dev. Biol.* **98**, 304–318.

Cheung, A., Swann, K., Carroll, J. (2000) The ability to generate normal Ca^{2+} transients in response to spermatozoa develops during the final stages of oocyte growth and maturation, *Hum. Reprod. (Oxford, England)* **15**, 1389–1395.

Chiba, K., Kado, R.T., Jaffe, L.A. (1990) Development of calcium release mechanisms during starfish oocyte maturation, *Dev. Biol.* **140**, 300–306.

Chini, E.N., Beers, K.W., Chini, C.C., Dousa, T.P. (1995) Specific modulation of cyclic ADP-ribose-induced Ca^{2+} release by polyamines, *Am. J. Physiol.* **269**, C1042–C1047.

Cho, C., Bunch, D.O., Faure, J.E., Goulding, E.H., Eddy, E.M., Primakoff, P., Myles, D.G. (1998) Fertilization defects in sperm from mice lacking fertilin beta, *Science* **281**, 1857–1859.

Churchill Grant, C., Okada, Y., Thomas Justyn, M., Genazzani Armando, A., Patel, S., Galione,

A. (2002) NAADP mobilizes Ca(2+) from reserve granules, lysosome-related organelles, in sea urchin eggs, *Cell* **111**, 703–708.

Ciapa, B., Whitaker, M. (1986) Two phases of inositol polyphosphate and diacylglycerol production at fertilisation, *FEBS Lett.* **195**, 347–351.

Ciapa, B., Epel, D. (1991) A rapid change in phosphorylation on tyrosine accompanies fertilization of sea urchin eggs, *FEBS Lett.* **295**, 167–170.

Ciapa, B., Borg, B., Whitaker, M. (1992) Polyphosphoinositide metabolism during the fertilization wave in sea urchin eggs, *Development (Cambridge, England)* **115**, 187–195.

Ciapa, B., Pesando, D., Wilding, M., Whitaker, M. (1994) Cell-cycle calcium transients driven by cyclic changes in inositol trisphosphate levels, *Nature* **368**, 875–878.

Cosson, M.P., Carre, D., Cosson, J. (1984) Sperm chemotaxis in siphonophores. II. Calcium-dependent asymmetrical movement of spermatozoa induced by the attractant, *J. Cell Sci.* **68**, 163–181.

Coward, K., Campos-Mendoza, A., Larman, M., Hibbitt, O., McAndrew, B., Bromage, N., Parrington, J. (2003) Teleost fish spermatozoa contain a cytosolic protein factor that induces calcium release in sea urchin egg homogenates and triggers calcium oscillations when injected into mouse oocytes, *Biochem. Biophys. Res. Commun.* **305**, 299–304.

Cox, L.J., Larman, M.G., Saunders, C.M., Hashimoto, K., Swann, K., Lai, F.A. (2002) Sperm phospholipase Czeta from humans and cynomolgus monkeys triggers Ca2+ oscillations, activation and development of mouse oocytes, *Reproduction (Cambridge, England)* **124**, 611–623.

Crossley, I., Swann, K., Chambers, E., Whitaker, M. (1988) Activation of sea urchin eggs by inositol phosphates is independent of external calcium, *Biochem. J.* **252**, 257–262.

Dale, B., De Belice, L.J., Taglietti, V. (1978) Membrane noise and conductance increase during single spermatozoon-egg interactions, *Nature* **275**, 217–219.

Day, M.L., McGuinness, O.M., Berridge, M.J., Johnson, M.H. (2000) Regulation of fertilization-induced Ca^{2+} spiking in the mouse zygote, *Cell Calcium* **28**, 47–54.

Deguchi, R., Osanai, K. (1994) Repetitive intracellular Ca^{2+} increases at fertilization and the role of Ca^{2+} in meiosis reinitiation from the first metaphase in oocytes of marine bivalves, *Dev. Biol.* **163**, 162–174.

Deng, M.Q., Huang, X.Y., Tang, T.S., Sun, F.Z. (1998) Spontaneous and fertilization-induced Ca^{2+} oscillations in mouse immature germinal vesicle-stage oocytes, *Biol. Reprod.* **58**, 807–813.

Dong, J.B., Tang, T.S., Sun, F.Z. (2000) Xenopus and chicken sperm contain a cytosolic soluble protein factor which can trigger calcium oscillations in mouse eggs, *Biochem. Biophys. Res. Commun.* **268**, 947–951.

Dube, F. (1988) The relationships between early ionic events, the pattern of protein synthesis, and oocyte activation in the surf clam, *Spisula solidissima*, *Dev. Biol.* **126**, 233–241.

Dube, F., Guerrier, P. (1982) Activation of Barnea candida (Mollusca, Pelecypoda) oocytes by sperm of KCl, but not by NH4Cl, requires a calcium influx, *Dev. Biol.* **92**, 408–417.

Ducibella, T., Huneau, D., Angelichio, E., Xu, Z., Schultz Richard, M., Kopf Gregory, S., Fissore, R., Madoux, S., Ozil, J.-P. (2002) Egg-to-embryo transition is driven by differential responses to Ca(2+) oscillation number, *Dev. Biol.* **250**, 280–291.

Dumollard, R., Sardet, C. (2001) Three different calcium wave pacemakers in ascidian eggs, *J. Cell. Sci.* **114**, 2471–2481.

Dumollard, R., Marangos, P., Fitzharris, G., Swann, K., Duchen, M., Carroll, J. (2004) Sperm-triggered [Ca2+] oscillations and Ca2+ homeostasis in the mouse egg have an absolute requirement for mitochondrial ATP production, *Development (Cambridge, England)* **131**, 3057–3067.

Dumollard, R., Hammar, K., Porterfield, M., Smith Peter, J., Cibert, C., Rouviere, C., Sardet, C. (2003) Mitochondrial respiration and Ca2+ waves are linked during fertilization and meiosis completion, *Development (Cambridge, England)* **130**, 683–692.

Eisen, A., Reynolds, G.T. (1984) Calcium transients during early development in single starfish (*Asterias forbesi*) oocytes, *J. Cell Biol.* **99**, 1878–1882.

Eisen, A., Reynolds, G.T. (1985) Source and sinks for the calcium released during fertilization of single sea urchin eggs, *J. Cell Biol.* **100**, 1522–1527.

Eisen, A., Kiehart, D.P., Wieland, S.J., Reynolds, G.T. (1984) Temporal sequence and spatial distribution of early events of fertilization

in single sea urchin eggs, *J. Cell Biol.* **99**, 1647–1654.

Evans, J.P., Kopf, G.S., Schultz, R.M. (1997) Characterization of the binding of recombinant mouse sperm fertilin beta subunit to mouse eggs: evidence for adhesive activity via an egg beta1 integrin-mediated interaction, *Dev. Biol.* **187**, 79–93.

Evans, T., Rosenthal, E.T., Youngblom, J., Distel, D., Hunt, T. (1983) Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division, *Cell* **33**, 389–396.

Fechter, J., Schoneberg, A., Schatten, G. (1996) Excision and disassembly of sperm tail microtubules during sea urchin fertilization: requirements for microtubule dynamics, *Cell Motil. Cytoskeleton* **35**, 281–288.

FitzHarris, G., Marangos, P., Carroll, J. (2003) Cell cycle-dependent regulation of the structure of the endoplasmic reticulum and inositol 1,4,5-trisphosphate-induced Ca^{2+} release in mouse oocytes and embryos, *Mol. Biol. Cell* **14**, 288–301.

Florman, H.M., Wassarman, P.M. (1985) O-linked oligosaccharides of mouse egg ZP3 account for its sperm receptor activity, *Cell* **41**, 313–324.

Fluck, R., Abraham, V., Miller, A., Galione, A. (1999) Microinjection of cyclic ADP-ribose triggers a regenerative wave of Ca2+ release and exocytosis of cortical alveoli in medaka eggs, *Zygote* **7**, 285–292.

Fontanilla, R.A., Nuccitelli, R. (1998) Characterization of the sperm-induced calcium wave in *Xenopus* eggs using confocal microscopy, *Biophys. J.* **75**, 2079–2087.

Fujiwara, T., Nakada, K., Shirakawa, H., Miyazaki, S. (1993) Development of inositol trisphosphate-induced calcium release mechanism during maturation of hamster oocytes, *Dev. Biol.* **156**, 69–79.

Galione, A., McDougall, A., Busa, W.B., Willmott, N., Gillot, I., Whitaker, M. (1993) Redundant mechanisms of calcium-induced calcium release underlying calcium waves during fertilization of sea urchin eggs, *Science* **261**, 348–352.

Gilkey, J.C., Jaffe, L.F., Ridgway, E.B., Reynolds, G.T. (1978) A free calcium wave traverses the activating egg of the medaka, *Oryzias latipes*, *J. Cell Biol.* **76**, 448–466.

Gillot, I., Whitaker, M. (1994) Calcium signals in and around the nucleus in sea urchin eggs, *Cell Calcium* **16**, 269–278.

Giusti, A.F., Xu, W., Hinkle, B., Terasaki, M., Jaffe, L.A. (2000) Evidence that fertilization activates starfish eggs by sequential activation of a Src-like kinase and phospholipase cgamma, *J. Biol. Chem.* **275**, 16788–16794.

Giusti, A.F., O'Neill, F.J., Yamasu, K., Foltz, K.R., Jaffe, L.A. (2003) Function of a sea urchin egg Src family kinase in initiating Ca2+ release at fertilization, *Dev. Biol.* **256**, 367–378.

Giusti, A.F., Carroll, D.J., Abassi, Y.A., Terasaki, M., Foltz, K.R., Jaffe, L.A. (1999) Requirement of a Src family kinase for initiating calcium release at fertilization in starfish eggs, *J. Biol. Chem.* **274**, 29318–29322.

Glabe, C.G. (1985) Interaction of the sperm adhesive protein, bindin, with phospholipid vesicles. II. Bindin induces the fusion of mixed-phase vesicles that contain phosphatidylcholine and phosphatidylserine in vitro, *J. Cell Biol.* **100**, 800–806.

Glahn, D., Mark, S.D., Behr, R.K., Nuccitelli, R. (1999) Tyrosine kinase inhibitors block sperm-induced egg activation in *Xenopus laevis*, *Dev. Biol.* **205**, 171–180.

Gordo, A.C., Rodrigues, P., Kurokawa, M., Jellerette, T., Exley, G.E., Warner, C., Fissore, R. (2002) Intracellular calcium oscillations signal apoptosis rather than activation in in vitro aged mouse eggs, *Biol. Reprod.* **66**, 1828–1837.

Goudeau, H., Depresle, Y., Rosa, A., Goudeau, M. (1994) Evidence by a voltage clamp study of an electrically mediated block to polyspermy in the egg of the ascidian Phallusia mammillata, *Dev. Biol.* **166**, 489–501.

Grainger, J.L., Winkler, M.M., Shen, S.S., Steinhardt, R.A. (1979) Intracellular pH controls protein synthesis rate in the sea urchin egg and early embryo, *Dev. Biol.* **68**, 396–406.

Guerrier, P., Leclerc-David, C., Moreau, M. (1993) Evidence for the involvement of internal calcium stores during serotonin-induced meiosis reinitiation in oocytes of the bivalve mollusc Ruditapes philippinarum, *Dev. Biol.* **159**, 474–484.

Halet, G., Tunwell, R., Balla, T., Swann, K., Carroll, J. (2002) The dynamics of plasma membrane PtdIns(4,5)P(2) at fertilization of mouse eggs, *J. Cell. Sci.* **115**, 2139–2149.

Hamaguchi, Y., Hiramoto, Y. (1981) Activation of sea urchin eggs by microinjection of calcium buffers, *Exp. Cell Res.* **134**, 171–179.

Han, J.K., Fukami, K., Nuccitelli, R. (1992) Reducing inositol lipid hydrolysis, Ins(1,4,5)P_3 receptor availability, or Ca^{2+} gradients lengthens the duration of the cell-cycle in *Xenopus laevis* blastomeres, *J. Cell Biol.* **116**, 147–156.

Heinecke, J.W., Meier, K.E., Lorenzen, J.A., Shapiro, B.M. (1990) A specific requirement for protein kinase C in activation of the respiratory burst oxidase of fertilization, *J. Biol. Chem.* **265**, 7717–7720.

Hesketh, T.R., Moore, J.P., Morris, J.D., Taylor, M.V., Rogers, J., Smith, G.A., Metcalfe, J.C. (1985) A common sequence of calcium and pH signals in the mitogenic stimulation of eukaryotic cells, *Nature* **313**, 481–484.

Hirose, K., Kadowaki, S., Tanabe, M., Takeshima, H., Lino, M. (1999) Spatiotemporal dynamics of inositol 1,4,5-triphosphate that underlies complex Ca^{2+} metabolism patterns, *Science* **284**, 1527–1530.

Ho, H.C., Suarez, S.S. (2001) Hyperactivation of mammalian spermatozoa: function and regulation, *Reproduction (Cambridge, England)* **122**, 519–526.

Homa, S.T., Carroll, J., Swann, K. (1993) The role of calcium in mammalian oocyte maturation and egg activation, *Hum. Reprod. (Oxford, England)* **8**, 1274–1281.

Hyslop, L.A., Carroll, M., Nixon, V.L., McDougall, A., Jones, K.T. (2001) Simultaneous measurement of intracellular nitric oxide and free calcium levels in chordate eggs demonstrates that nitric oxide has no role at fertilization, *Dev. Biol.* **234**, 216–230.

Inoue, S., Tilney, L.G. (1982) Acrosomal reaction of thyone sperm. I. Changes in the sperm head visualized by high resolution video microscopy, *J. Cell Biol.* **93**, 812–819.

Jaffe, L.A. (1976) Fast block to polyspermy in sea urchin eggs is electrically mediated, *Nature* **261**, 68–71.

Jaffe, L.A., Sharp, A.P., Wolf, D.P. (1983) Absence of an electrical polyspermy block in the mouse, *Dev. Biol.* **96**, 317–323.

Jellerette, T., He, C.L., Wu, H., Parys, J.B., Fissore, R.A. (2000) Down-regulation of the inositol 1,4,5-trisphosphate receptor in mouse eggs following fertilization or parthenogenetic activation, *Dev. Biol.* **223**, 238–250.

Johnson, J.D., Epel, D. (1976) Intracellular pH and activation of sea urchin eggs after fertilisation, *Nature* **262**, 661–664.

Jones, K.T., Nixon, V.L. (2000) Sperm-induced Ca^{2+} oscillations in mouse oocytes and eggs can be mimicked by photolysis of caged inositol 1,4,5-trisphosphate: evidence to support a continuous low level production of inositol 1, 4,5-trisphosphate during mammalian fertilization, *Dev. Biol.* **225**, 1–12.

Jones, K.T., Soeller, C., Cannell, M.B. (1998) The passage of Ca^{2+} and fluorescent markers between the sperm and egg after fusion in the mouse, *Development (Cambridge, England)* **125**, 4627–4635.

Kaji, K., Oda, S., Shikano, T., Ohnuki, T., Uematsu, Y., Sakagami, J., Tada, N., Miyazaki, S., Kudo, A. (2000) The gamete fusion process is defective in eggs of Cd9-deficient mice, *Nat. Genet.* **24**, 279–282.

Kaltschmidt, J.A., Brand, A.H. (2002) Asymmetric cell division: microtubule dynamics and spindle asymmetry, *J. Cell. Sci.* **115**, 2257–2264.

Kawahara, H., Yokosawa, H. (1994) Intracellular calcium mobilization regulates the activity of 26 S proteasome during the metaphase-anaphase transition in the ascidian meiotic cell cycle, *Dev. Biol.* **166**, 623–633.

Keizer, J., Li, Y.X., Stojilkovic, S., Rinzel, J. (1995) $InsP_3$-induced Ca^{2+} excitability of the endoplasmic reticulum, *Mol. Biol. Cell* **6**, 945–951.

Kishimoto, T., Usui, N., Kanatani, H. (1984) Breakdown of starfish ovarian follicle induced by maturation-promoting factor, *Dev. Biol.* **101**, 28–34.

Kline, D. (1988) Calcium-dependent events at fertilization of the frog egg: injection of a calcium buffer blocks ion channel opening, exocytosis, and formation of pronuclei, *Dev. Biol.* **126**, 346–361.

Kono, T., Carroll, J., Swann, K., Whittingham, D.G. (1995) Nuclei from fertilized mouse embryos have calcium-releasing activity, *Development (Cambridge, England)* **121**, 1123–1128.

Kono, T., Jones, K.T., Bos-Mikich, A., Whittingham, D.G., Carroll, J. (1996) A cell cycle-associated change in Ca^{2+} releasing activity leads to the generation of Ca^{2+} transients in mouse embryos during the first mitotic division, *J. Cell Biol.* **132**, 915–923.

Kranz, E., Lorz, H. (1994) In vitro fertilisation of maize by single egg and sperm cell protoplast fusion mediated by high calcium and high pH, *Zygote (Cambridge, England)* **2**, 125–128.

Krylov, D.M., Nasmyth, K., Koonin, E.V. (2003) Evolution of eukaryotic cell cycle regulation: stepwise addition of regulatory kinases and late advent of the CDKs, *Curr. Biol.* **13**, 173–177.

Kuo, R.C., Baxter, G.T., Thompson, S.H., Stricker, S.A., Patton, C., Bonaventura, J., Epel, D. (2000) NO is necessary and sufficient for egg activation at fertilization, *Nature* **406**, 633–636.

Kuroda, R., Kontani, K., Kanda, Y., Katada, T., Nakano, T., Satoh, Y., Suzuki, N., Kuroda, H. (2001) Increase of cGMP, cADP-ribose and inositol 1,4,5-trisphosphate preceding Ca^{2+} transients in fertilization of sea urchin eggs, *Development* **128**, 4405–4414.

Kyozuka, K., Deguchi, R., Mohri, T., Miyazaki, S. (1998) Injection of sperm extract mimics spatiotemporal dynamics of Ca^{2+} responses and progression of meiosis at fertilization of ascidian oocytes, *Development* **125**, 4099–4105.

Labbe, J.C., Lee, M.G., Nurse, P., Picard, A., Doree, M. (1988) Activation at M-phase of a protein kinase encoded by a starfish homologue of the cell cycle control gene cdc2+, *Nature* **335**, 251–254.

Lambert, C., Goudeau, H., Franchet, C., Lambert, G., Goudeau, M. (1997) Ascidian eggs block polyspermy by two independent mechanisms: one at the egg plasma membrane, the other involving the follicle cells, *Mol. Reprod. Dev.* **48**, 137–143.

Lawrence, Y., Whitaker, M., Swann, K. (1997) Sperm-egg fusion is the prelude to the initial Ca^{2+} increase at fertilization in the mouse, *Development (Cambridge, England)* **124**, 233–241.

Lawrence, Y., Ozil, J.P., Swann, K. (1998) The effects of a Ca^{2+} chelator and heavy-metal-ion chelators upon Ca^{2+} oscillations and activation at fertilization in mouse eggs suggest a role for repetitive Ca^{2+} increases, *Biochem. J.* **335**, 335–342.

Leckie, M.P., Empson, R.M., Bechetti, A., Thomas, J., Galione, A., Whitaker, M. (2003) The NO pathway acts late during the fertilization response in sea urchin eggs, *J. Biol. Chem.* **278**, 12247–12254.

Lee, M.G., Nurse, P. (1987) Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2, *Nature* **327**, 31–35.

Levasseur, M., McDougall, A. (2000) Sperm-induced calcium oscillations at fertilisation in ascidians are controlled by cyclin B1-dependent kinase activity, *Development (Cambridge, England)* **127**, 631–641.

Li, S.T., Huang, X.Y., Sun, F.Z. (2001) Flowering plant sperm contains a cytosolic soluble protein factor which can trigger calcium oscillations in mouse eggs, *Biochem. Biophys. Res. Commun.* **287**, 56–59.

Li, W., Llopis, J., Whitney, M., Zlokarnik, G., Tsien, R.Y. (1998) Cell-permeant caged InsP3 ester shows that Ca^{2+} spike frequency can optimize gene expression, *Nature* **392**, 936–941.

Longo, F.J., Lynn, J.W., McCulloh, D.H., Chambers, E.L. (1986) Correlative ultrastructural and electrophysiological studies of sperm egg interactions of the sea-urchin, *Lytechinus variegatus*, *Dev. Biol.* **118**, 155–166.

Lorca, T., Abrieu, A., Means, A., Doree, M. (1994) Ca^{2+} is involved through type II calmodulin-dependent protein kinase in cyclin degradation and exit from metaphase, *Biochim. Biophys. Acta* **1223**, 325–332.

Lorca, T., Galas, S., Fesquet, D., Devault, A., Cavadore, J.C., Doree, M. (1991) Degradation of the proto-oncogene product p39mos is not necessary for cyclin proteolysis and exit from meiotic metaphase: requirement for a Ca^{2+}-calmodulin dependent event, *EMBO J.* **10**, 2087–2093.

Lorca, T., Cruzalegui, F.H., Fesquet, D., Cavadore, J.C., Mery, J., Means, A., Doree, M. (1993) Calmodulin-dependent protein kinase II mediates inactivation of MPF and CSF upon fertilization of Xenopus eggs, *Nature* **366**, 270–273.

Machaca, K., Haun, S. (2002) Induction of maturation-promoting factor during Xenopus oocyte maturation uncouples Ca(2+) store depletion from store-operated Ca(2+) entry, *J. Cell Biol.* **156**, 75–85.

Marangos, P., FitzHarris, G., Carroll, J. (2003) Ca2+ oscillations at fertilization in mammals are regulated by the formation of pronuclei, *Development (Cambridge, England)* **130**, 1461–1472.

Markoulaki, S., Matson, S., Abbott Allison, L., Ducibella, T. (2003) Oscillatory CaMKII activity in mouse egg activation, *Dev. Biol.* **258**, 464–474.

McCulloh, D.H., Chambers, E.L. (1992) Fusion of membranes during fertilization - increases

of the sea-urchin eggs membrane capacitance and membrane conductance at the site of contact with the sperm, *J. Gen. Physiol.* **99**, 137–175.

McDougall, A., Levasseur, M. (1998) Sperm-triggered calcium oscillations during meiosis in ascidian oocytes first pause, restart, then stop: correlations with cell cycle kinase activity, *Development (Cambridge, England)* **125**, 4451–4459.

McDougall, A., Levasseur, M., O'Sullivan, A.J., Jones, K.T. (2000) Cell cycle-dependent repetitive Ca^{2+} waves induced by a cytosolic sperm extract in mature ascidian eggs mimic those observed at fertilization, *J. Cell. Sci.* **113**, 3453–3462.

Mehlmann, L.M., Terasaki, M., Jaffe, L.A., Kline, D. (1995) Reorganization of the endoplasmic reticulum during meiotic maturation of the mouse oocyte, *Dev. Biol.* **170**, 607–615.

Mehlmann, L.M., Carpenter, G., Rhee, S.G., Jaffe, L.A. (1998) SH2 domain-mediated activation of phospholipase C gamma is not required to initiate Ca^{2+} release at fertilization of mouse eggs, *Dev. Biol.* **203**, 221–232.

Miller, B.J., Georges-Labouesse, E., Primakoff, P., Myles, D.G. (2000) Normal fertilization occurs with eggs lacking the integrin alpha6beta1 and is CD9-dependent, *J. Cell Biol.* **149**, 1289–1296.

Miyazaki, S., Igusa, Y. (1982) Ca-mediated activation of a K current at fertilization of golden hamster eggs, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 931–935.

Miyazaki, S., Yuzaki, M., Nakada, K., Shirakawa, H., Nakanishi, S., Nakade, S., Mikoshiba, K. (1992) Block of Ca^{2+} wave and Ca^{2+} oscillation by antibody to the inositol 1,4,5-trisphosphate receptor in fertilized hamster eggs, *Science* **257**, 251–255.

Moolenaar, W.H., Tertoolen, L.G., de Laat, S.W. (1984) Phorbol ester and diacylglycerol mimic growth factors in raising cytoplasmic pH, *Nature* **312**, 371–374.

Nakano, Y., Shirakawa, H., Mitsuhashi, N., Kuwabara, Y., Miyazaki, S. (1997) Spatiotemporal dynamics of intracellular calcium in the mouse egg injected with a spermatozoon, *Mol. Hum. Reprod.* **3**, 1087–1093.

Nixon, V.L., Levasseur, M., McDougall, A., Jones, K.T. (2002) Ca^{2+} oscillations promote APC/C-dependent cyclin B1 degradation during metaphase arrest and completion of meiosis in fertilizing mouse eggs, *Curr. Biol.: CB* **12**, 746–750.

Nuccitelli, R. (1980) The fertilization potential is not necessary for the block to polyspermy or the activation of development in the medaka egg, *Dev. Biol.* **76**, 499–504.

Ozil, J.P., Swann, K. (1995) Stimulation of repetitive calcium transients in mouse eggs, *J. Physiol.* **483**, 331–346.

Ozil, J.P., Huneau, D. (2001) Activation of rabbit oocytes: the impact of the Ca^{2+} signal regime on development, *Development (Cambridge, England)* **128**, 917–928.

Parrington, J., Brind, S., De Smedt, H., Gangeswaran, R., Lai, F.A., Wojcikiewicz, R., Carroll, J. (1998) Expression of inositol 1,4,5-trisphosphate receptors in mouse oocytes and early embryos: the type I isoform is upregulated in oocytes and downregulated after fertilization, *Dev. Biol.* **203**, 451–461.

Patel, R., Holt, M., Philipova, R., Moss, S., Schulman, H., Hidaka, H., Whitaker, M. (1999) Calcium/calmodulin-dependent phosphorylation and activation of human Cdc25-C at the G2/M phase transition in HeLa cells, *J. Biol. Chem.* **274**, 7958–7968.

Poenie, M., Alderton, J., Tsien, R.Y., Steinhardt, R.A. (1985) Changes of free calcium levels with stages of the cell division cycle, *Nature* **315**, 147–149.

Poenie, M., Alderton, J., Steinhardt, R., Tsien, R. (1986) Calcium rises abruptly and briefly throughout the cell at the onset of anaphase, *Science* **233**, 886–889.

Reimann, J.D.R., Jackson, P.K. (2002) Emi1 is required for cytostatic factor arrest in vertebrate eggs, *Nature* **416**, 850–854.

Rice, A., Parrington, J., Jones, K.T., Swann, K. (2000) Mammalian sperm contain a Ca^{2+}-sensitive phospholipase C activity that can generate $InsP_3$ from PIP_2 associated with intracellular organelles, *Dev. Biol.* **228**, 125–135.

Rongish, B.J., Wu, W., Kinsey, W.H. (1999) Fertilization-induced activation of phospholipase C in the sea urchin egg, *Dev. Biol.* **215**, 147–154.

Runft, L.L., Jaffe, L.A. (2000) Sperm extract injection into ascidian eggs signals Ca^{2+} release by the same pathway as fertilization, *Development (Cambridge, England)* **127**, 3227–3236.

Runft, L.L., Watras, J., Jaffe, L.A. (1999) Calcium release at fertilization of Xenopus eggs

requires type I IP(3) receptors, but not SH2 domain-mediated activation of PLCgamma or G(q)-mediated activation of PLCbeta, *Dev. Biol.* **214**, 399–411.

Sato, K., Tokmakov, A.A., Iwasaki, T., Fukami, Y. (2000) Tyrosine kinase-dependent activation of phospholipase Cgamma is required for calcium transient in Xenopus egg fertilization, *Dev. Biol.* **224**, 453–469.

Sato, Y., Miyazaki, S., Shikano, T., Mitsuhashi, N., Takeuchi, H., Mikoshiba, K., Kuwabara, Y. (1998) Adenophostin, a potent agonist of the inositol 1,4,5-trisphosphate receptor, is useful for fertilization of mouse oocytes injected with round spermatids leading to normal offspring, *Biol. Reprod.* **58**, 867–873.

Saunders, C.M., Larman, M.G., Parrington, J., Cox, L.J., Royse, J., Blayney, L.M., Swann, K., Lai, F.A. (2002) PLC zeta: a sperm-specific trigger of Ca^{2+} oscillations in eggs and embryo development, *Development (Cambridge, England)* **129**, 3533–3544.

Schmidt, T., Patton, C., Epel, D. (1982) Is there a role for the Ca2+ influx during fertilization of the sea urchin egg? *Dev. Biol.* **90**, 284–290.

SeGall, G.K., Lennarz, W.J. (1981) Jelly coat and induction of the acrosome reaction in echinoid sperm, *Dev. Biol.* **86**, 87–93.

Shapiro, B.M. (1984) Molecular aspects of sperm-egg fusion, *Ciba Found. Symp.* **103**, 86–99.

Shearer, J., De Nadai, C., Emily-Fenouil, F., Gache, C., Whitaker, M., Ciapa, B. (1999) Role of phospholipase Cgamma at fertilization and during mitosis in sea urchin eggs and embryos, *Development (Cambridge, England)* **126**, 2273–2284.

Shen, S.S., Steinhardt, R.A. (1979) Intracellular pH and the sodium requirement at fertilisation, *Nature* **282**, 87–89.

Shilling, F.M., Carroll, D.J., Muslin, A.J., Escobedo, J.A., Williams, L.T., Jaffe, L.A. (1994) Evidence for both tyrosine kinase and G-protein-coupled pathways leading to starfish egg activation, *Dev. Biol.* **162**, 590–599.

Shiraishi, K., Okada, A., Shirakawa, H., Nakanishi, S., Mikoshiba, K., Miyazaki, S. (1995) Developmental changes in the distribution of the endoplasmic reticulum and inositol 1,4,5-trisphosphate receptors and the spatial pattern of Ca^{2+} release during maturation of hamster oocytes, *Dev. Biol.* **170**, 594–606.

Shiwa, M., Murayama, T., Ogawa, Y. (2002) Molecular cloning and characterization of ryanodine receptor from unfertilized sea urchin eggs, *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **282**, R727–R737.

Snow, P., Yim, D.L., Leibow, J.D., Saini, S., Nuccitelli, R. (1996) Fertilization stimulates an increase in inositol trisphosphate and inositol lipid levels in Xenopus eggs, *Dev. Biol.* **180**, 108–118.

Speksnijder, J.E. (1992) The repetitive calcium waves in the fertilized ascidian egg are initiated near the vegetal pole by a cortical pacemaker, *Dev. Biol.* **153**, 259–271.

Speksnijder, J.E., Sardet, C., Jaffe, L.F. (1990a) The activation wave of calcium in the ascidian egg and its role in ooplasmic segregation, *J. Cell Biol.* **110**, 1589–1598.

Speksnijder, J.E., Sardet, C., Jaffe, L.F. (1990b) Periodic calcium waves cross ascidian eggs after fertilization, *Dev. Biol.* **142**, 246–249.

Speksnijder, J.E., Corson, D.W., Sardet, C., Jaffe, L.F. (1989) Free calcium pulses following fertilization in the ascidian egg, *Dev. Biol.* **135**, 182–190.

Steinhardt, R.A., Epel, D. (1974) Activation of sea-urchin eggs by a calcium ionophore, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1915–1919.

Steinhardt, R.A., Alderton, J. (1988) Intracellular free calcium rise triggers nuclear envelope breakdown in the sea urchin embryo, *Nature* **332**, 364–366.

Steinhardt, R.A., Lundin, L., Mazia, D. (1971) Bioelectric responses of the echinoderm egg to fertilization, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 2426–2430.

Steinhardt, R., Zucker, R., Schatten, G. (1977) Intracellular calcium release at fertilization in the sea urchin egg, *Dev. Biol.* **58**, 185–196.

Steinhardt, R.A., Epel, D., Carroll, E.J., Yanagimachi, R. (1974) Is calcium ionophore a universal activator for unfertilised eggs? *Nature* **252**, 41–43.

Stephano, J.L., Gould, M.C. (1997) The intracellular calcium increase at fertilization in Urechis caupo oocytes: activation without waves, *Dev. Biol.* **191**, 53–68.

Stith, B.J., Espinoza, R., Roberts, D., Smart, T. (1994) Sperm increase inositol 1,4,5-trisphosphate mass in *Xenopus laevis* eggs preinjected with calcium buffers or heparin, *Dev. Biol.* **165**, 206–215.

Stricker, S.A. (1996) Repetitive calcium waves induced by fertilization in the nemertean worm *Cerebratulus lacteus*, *Dev. Biol.* **176**, 243–263.

Stricker, S.A., Smythe, T.L. (2003) Endoplasmic reticulum reorganizations and Ca^{2+} signaling in maturing and fertilized oocytes of marine protostome worms: the roles of MAPKs and MPF, *Development (Cambridge, England)* **130**, 2867–2879.

Stricker, S.A., Silva, R., Smythe, T. (1998) Calcium and endoplasmic reticulum dynamics during oocyte maturation and fertilization in the marine worm *Cerebratulus lacteus*, *Dev. Biol.* **203**, 305–322.

Stricker, S.A., Swann, K., Jones, K.T., Fissore, R.A. (2000) Injections of porcine sperm extracts trigger fertilization-like calcium oscillations in oocytes of a marine worm, *Exp. Cell Res.* **257**, 341–347.

Swann, K. (1990) A cytosolic sperm factor stimulates repetitive calcium increases and mimics fertilization in hamster eggs, *Development (Cambridge, England)* **110**, 1295–1302.

Swann, K., Whitaker, M. (1985) Stimulation of the Na/H exchanger of sea urchin eggs by phorbol ester, *Nature* **314**, 274–277.

Swann, K., Whitaker, M. (1986) The part played by inositol trisphosphate and calcium in the propagation of the fertilization wave in sea urchin eggs, *J. Cell Biol.* **103**, 2333–2342.

Swann, K., McDougall, A., Whitaker, M. (1994) Calcium signalling at fertilization, *J. Mar. Biolog. Assoc. U.K.* **74**, 3–16.

Swanson, W.J., Vacquier, V.D. (1998) Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein, *Science* **281**, 710–712.

Tachibana, K., Tanaka, D., Isobe, T., Kishimoto, T. (2000) c-Mos forces the mitotic cell cycle to undergo meiosis II to produce haploid gametes, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14301–14306.

Tang, T.S., Dong, J.B., Huang, X.Y., Sun, F.Z. (2000) Ca^{2+} oscillations induced by a cytosolic sperm protein factor are mediated by a maternal machinery that functions only once in mammalian eggs, *Development (Cambridge, England)* **127**, 1141–1150.

Terasaki, M., Jaffe, L.A. (1991) Organization of the sea urchin egg endoplasmic reticulum and its reorganization at fertilization, *J. Cell Biol.* **114**, 929–940.

Thomas, T.W., Eckberg, W.R., Dube, F., Galione, A. (1998) Mechanisms of calcium release and sequestration in eggs of *Chaetopterus pergamentaceus*, *Cell Calcium* **24**, 285–292.

Tilney, L.G., Jaffe, L.A. (1980) Actin, microvilli and the fertilization cone of sea urchin eggs, *J. Cell Biol.* **87**, 771–782.

Tilney, L.G., Inoue, S. (1982) Acrosomal reaction of Thyone sperm. II. The kinetics and possible mechanism of acrosomal process elongation, *J. Cell Biol.* **93**, 820–827.

Tokmakov, A.A., Sato, K.I., Iwasaki, T., Fukami, Y. (2002) Src kinase induces calcium release in Xenopus egg extracts via PLCgamma and IP(3)-dependent mechanism, *Cell Calcium* **32**, 11–20.

Tombes, R.M., Simerly, C., Borisy, G.G., Schatten, G. (1992) Meiosis, egg activation, and nuclear envelope breakdown are differentially reliant on Ca^{2+}, whereas germinal vesicle breakdown is Ca^{2+} independent in the mouse oocyte, *J. Cell Biol.* **117**, 799–811.

Török, K., Wilding, M., Groigno, L., Patel, R., Whitaker, M. (1998) Imaging the spatial dynamics of calmodulin activation during mitosis, *Curr. Biol.* **8**, 692–699.

Tsien, R.Y. (1986) New tetracarboxylate chelators for fluorescence measurement and photochemical manipulation of cytosolic free calcium concentrations, *Soc. Gen. Physiol. Ser.* **40**, 327–345.

Tunquist, B.J., Schwab, M.S., Chen, L.G., Maller, J.L. (2002) The spindle checkpoint kinase bub1 and cyclin e/cdk2 both contribute to the establishment of meiotic metaphase arrest by cytostatic factor, *Curr. Biol.: CB* **12**, 1027–1033.

Turner, P.R., Sheetz, M.P., Jaffe, L.A. (1984) Fertilization increases the polyphosphoinositide content of sea urchin eggs, *Nature* **310**, 414–415.

Twigg, J., Patel, R., Whitaker, M. (1988) Translational control of InsP3-induced chromatin condensation during the early cell cycles of sea urchin embryos, *Nature* **332**, 366–369.

Vacquier, V.D. (1998) Evolution of gamete recognition proteins, *Science* **281**, 1995–1998.

Verlhac, M.H., Kubiak, J.Z., Weber, M., Geraud, G., Colledge, W.H., Evans, M.J., Maro, B. (1996) Mos is required for MAP kinase activation and is involved in microtubule organization during meiotic maturation in the mouse, *Development (Cambridge, England)* **122**, 815–822.

Viets, L.N., Campbell, K.D., White, K.L. (2001) Pathways involved in RGD-mediated calcium

transients in mature bovine oocytes, *Cloning Stem Cells* **3**, 105–113.

Walker, D.S., Gower, N.J.D., Ly, S., Bradley, G.L., Baylis, H.A. (2002) Regulated disruption of inositol 1,4,5-trisphosphate signalling in Caenorhabditis elegans reveals new functions in feeding and embryogenesis, *Mol. Biol. Cell* **13**, 1329–1337.

Webb, D.J., Nuccitelli, R. (1985) Fertilization potential and electrical properties of the *Xenopus laevis* egg, *Dev. Biol.* **107**, 395–406.

Whitaker, M., Irvine, R.F. (1984) Inositol 1,4,5-trisphosphate microinjection activates sea urchin eggs, *Nature* **312**, 636–639.

Whitaker, M., Aitchison, M. (1985) Calcium-dependent polyphosphoinositide hydrolysis is associated with exocytosis in vitro, *FEBS Lett.* **182**, 119–124.

Wilding, M., Dale, B. (1998) Soluble extracts from ascidian spermatozoa trigger intracellular calcium release independently of the activation of the ADP ribose channel, *Zygote (Cambridge, England)* **6**, 149–154.

Wilding, M., Torok, K., Whitaker, M. (1995) Activation-dependent and activation-independent localisation of calmodulin to the mitotic apparatus during the first cell cycle of the Lytechinus pictus embryo, *Zygote (Cambridge, England)* **3**, 219–224.

Wilding, M., Russo, G.L., Galione, A., Marino, M., Dale, B. (1998) ADP-ribose gates the fertilization channel in ascidian oocytes, *Am. J. Physiol.* **275**, C1277–C1283.

Williams, C.J., Mehlmann, L.M., Jaffe, L.A., Kopf, G.S., Schultz, R.M. (1998) Evidence that Gq family G proteins do not function in mouse egg activation at fertilization, *Dev. Biol.* **198**, 116–127.

Winkler, M.M., Steinhardt, R.A., Grainger, J.L., Minning, L. (1980) Dual ionic controls for the activation of protein synthesis at fertilization, *Nature* **287**, 558–560.

Witchel, H.J., Steinhardt, R.A. (1990) 1-Methyladenine can consistently induce a fura-detectable transient calcium increase which is neither necessary nor sufficient for maturation in oocytes of the starfish *Asterina miniata*, *Dev. Biol.* **141**, 393–398.

Wood, C.D., Nisihigaki, T., Furuta, T., Baba, S.A., Darszon, A. (2005) Real-time analysis of the role of Ca^{2+} in flagellar movement and motility in single sea urchin sperm, *J. Cell. Biol.* **169**, 725–731.

Wu, G.J., Simerly, C., Zoran, S.S., Funte, L.R., Schatten, G. (1996) Microtubule and chromatin dynamics during fertilization and early development in rhesus monkeys, and regulation by intracellular calcium ions, *Biol. Reprod.* **55**, 260–270.

Wu, H., Smyth, J., Luzzi, V., Fukami, K., Takenawa, T., Black, S.L., Allbritton, N.L., Fissore, R.A. (2001) Sperm factor induces intracellular free calcium oscillations by stimulating the phosphoinositide pathway, *Biol. Reprod.* **64**, 1338–1349.

Wyrick, R.E., Nishihara, T., Hedrick, J.L. (1974) Agglutination of jelly coat and cortical granule components and the block to polyspermy in the amphibian *Xenopus laevis*, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2067–2071.

Xu, X.Z.S., Sternberg Paul, W. (2003) A C. elegans sperm TRP protein required for sperm-egg interactions during fertilization, *Cell* **114**, 285–297.

Yamamoto, S., Kubota, H.Y., Yoshimoto, Y., Iwao, Y. (2001) Injection of a sperm extract triggers egg activation in the newt *Cynops pyrrhogaster*, *Dev. Biol.* **230**, 89–99.

Yoshida, M., Murata, M., Inaba, K., Morisawa, M. (2002) A chemoattractant for ascidian spermatozoa is a sulfated steroid, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14831–14836.

Yoshida, M., Sensui, N., Inoue, T., Morisawa, M., Mikoshiba, K. (1998) Role of two series of Ca^{2+} oscillations in activation of ascidian eggs, *Dev. Biol.* **203**, 122–133.

Yoshida, M., Ishikawa, M., Izumi, H., De Santis, R., Morisawa, M. (2003) Store-operated calcium channel regulates the chemotactic behavior of ascidian sperm, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 149–154.

Zhu, C.C., Furuichi, T., Mikoshiba, K., Wojcikiewicz, R.J. (1999) Inositol 1,4,5-trisphosphate receptor downregulation is activated directly by inositol 1.4.5-trisphosphate binding, *J. Biol. Chem.* **274**, 3476–3484.

Zucker, R.S., Steinhardt, R.A. (1978) Prevention of the cortical reaction in fertilized sea urchin eggs by injection of calcium-chelating ligands, *Biochim. Biophys. Acta* **541**, 459–466.

Synovial Mast Cells in Inflammatory Arthritis

Theoharis C. Theoharides
Tufts University School of Medicine, Boston, MA, USA

1 Osteoarthritis Versus Rheumatoid Arthritis 39
1.1 Specific Triggers Versus Lack of Inhibitory Molecules 39

2 Mast Cells in the Joints 40
2.1 Inflammatory Mediators 40

3 Chemoattractants 41
3.1 Animal Models 41

4 Pathophysiology of Mast Cells 42
4.1 Triggers 42
4.2 Mediators 43
4.3 Selective Secretion 45

5 Neurohormonal Stimulation of Mast Cells and Inflammatory Disorders 47
5.1 Anatomical Observations 47
5.2 Stress and the HPA Axis 47
5.3 Proinflammatory Effects of Stress 47
5.4 CRH Effects on Mast Cells 48

6 Therapy 48
6.1 Proteoglycans 48
6.2 Flavonoids 49
6.3 Synergistic Effects 50

7 Perspectives 50

Acknowledgment 51

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

Bibliography 52
Books and Reviews 52
Primary Literature 53

Keywords

Activation
Stimulation of cells to undergo partial, selective, or complete synthesis and secretion of mediators, or generation of other molecules or responses in response to a variety of triggers; such processes may be associated with subtle morphologic changes discernible only by appropriate ultrastructural studies.

Corticotropin-releasing Hormone
A peptide originally shown to be secreted from the hypothalamus in response to stress and primarily stimulating the anterior pituitary cells to secrete ACTH. It has since been shown to affect numerous systems including the immune system, and we proposed that it be renamed "stress response hormone (SRH)" instead of CRH.

Degranulation
The release of the contents of secretory granules, from mast cells or other secretory cells such as polymorphonuclear leukocytes, platelets, pituitary, or β-pancreatic cells in response to specific receptor-mediated triggers through a process dependent on energy and extracellular calcium ions.

Inflammation
An orchestrated, often excessive response of the organism, involving the immune, endocrine, and neural systems against real or perceived triggers.

Protease-activated Receptors (PARs)
Unique receptors on cell surfaces, stimulation of which leads to widespread inflammation.

Stress
The response of cells and tissues to real or perceived noxious stimuli with anti-inflammatory effects through brain secretion of CRH, but proinflammatory effects through peripheral secretion of CRH and related peptides.

Tryptase
One of many unique proteases secreted from mast cells that have multiple actions, such as microvascular leakage, stimulation of protease-activated receptors (PARs) and induction of inflammation, generation of active peptides, degradation of cytokines.

Inflammatory arthritis is a highly prevalent condition with substantial morbidity and degree of disability. The pathogenesis is still poorly understood limiting effective therapy. Increasing evidence has implicated synovial mast cells in the initiation of inflammatory arthritis. Mast cells are commonly known for their involvement in allergic reactions through degranulation and release of histamine. However, these cells are capable of releasing over 50 potent vasoactive, proinflammatory and neurosensitizing molecules in response to nonallergic triggers without degranulation. Such selective release of cytokines, especially in response to stress hormones, provides a unique explanation and possible therapeutic targets for inflammatory arthritis.

1 Osteoarthritis Versus Rheumatoid Arthritis

A 2002 report from the US Center for Disease Control (CDC) indicated that one-third of all adults in the United States (almost 70 million) suffer from arthritis or chronic joint pain, up from one-fifth in 1993, with an estimated annual cost of $82 billion. The impact of arthritis is comparable in Australia, Canada, Europe, and the United Kingdom, with an estimated total of another 60 million affected individuals.

Osteoarthritis (OA) is a degenerative joint disease and is characterized by the breakdown of the joint's cartilage and loss of joint space especially in hands and weight-bearing joints, such as knees, hips, feet, and the back. Cartilage breakdown allows bones to rub against each other, causing inflammation, pain, and loss of movement. Although age is a risk factor, obesity, and joint injuries due to sports, work-related activity, or accidents increase the risk of developing OA.

Rheumatoid arthritis (RA) involves chronic inflammation and destruction of the joints. RA is a systemic disease that affects the entire body and is characterized by inflammation and reactive hyperplasia of the synovium, which causes redness, swelling, stiffness, and pain. RA typically affects many different joints and is noted for remissions and flares, often precipitated or exacerbated by stress. Articular damage in RA involves cartilage erosion, inflammatory cell accumulation and finally to bone destruction and reactive bone spur formation. The involved joints lose their shape and alignment, resulting in pain and loss of movement.

1.1 Specific Triggers Versus Lack of Inhibitory Molecules

As in common practice in medicine, there have been numerous investigations in the potential trigger(s) implicated in the pathophysiology of inflammatory and especially of rheumatoid arthritis. In certain cases, as in crystal (gout) arthritis or Lyme arthritis, the etiologic agents are known. However, such triggers are not known for RA. There is some evidence from animal models that inflammatory arthritis may involve an inappropriate autoimmune response against structural (colagen) or infection (bacterial, viral) antigens. However, little attention has been paid to the potential lack of innate inhibitory molecules, such

as a_1-antitrypsin or a_2-macroglobulin in other tissues.

2 Mast Cells in the Joints

The inflammatory and subsequent lytic processes in inflammatory arthritis are not well understood, especially the initial triggering events. Until recently, discussion of the cellular elements involved did not even mention mast cells, even though a number of papers had reported the presence of mast cells in joints. Over the last 10 years, histochemical observations were correlated to clinical findings. Increased numbers of mast cells have been routinely noted and their mediators could initiate and promote inflammation while mast cells constitute a small fraction of the cells in a normal human synovium, they increase to almost 10% in RA. Mast cells, cytokines, and metalloproteinases were colocalized in the RA lesion and activation of synovial mast cells was shown to be autoregulated by their own mediators' histamine and tryptase (Fig. 1).

It was recently convincingly argued that mast cells may be involved in inflammatory arthritis; however, this presentation was still limited primarily to the role of histamine and tumor necrosis factor, TNF-α, and did not discuss potential triggers such as C3a and C5a. One of the reasons why mast cells have been neglected comes from the persistence of associating them only with allergies in spite of mounting evidence to the contrary. Another reason is the fact that mast cells are not commonly seen to degranulate in tissues such as joints that do not get allergic reactions; they were, therefore, considered an "innocent bystander's" renewed evidence that mast cells can release mediators selectively without degranulation, especially in response to nonallergic triggers (discussed in the following) has expanded the potential importance of mast cells.

2.1 Inflammatory Mediators

Mast cell-derived TNF-α mediates inflammation and induces hypertrophy of draining lymph nodes during infection.

Mast cells are known to secrete IL-6 and recent studies showed that mast cell–deficient mice could not increase their serum IL-6 in response to acute stress. This is of interest in view of the fact that IL-6 is elevated, especially in juvenile RA. IL-6 is

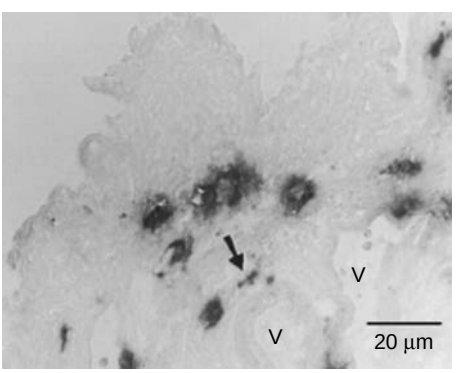

Fig. 1 Tryptase-positive mast cells (brown color) adjacent to blood vessels (V) from the joint of a patient with rheumatoid arthritis (See color plate p. xxxviii).

also important in collagen-induced and antigen-induced arthritis.

A number of chemokine receptors have also been identified on mast cells, especially CXCR3 protein, which was preferentially expressed on mast cells in the synovium of RA patients.

3 Chemoattractants

Tryptase can cause microvascular leakage and stimulate protease-activated receptors (PAR) leading to widespread inflammation, including autocrine stimulation of mast cells. Protease also acts as a tissue metalloproteinase, degrading matrix, as well as gelatinase, thus contributing to tissue remodeling. Tryptase also induces hyperexcitability of submucosal neurons, while histamine directly stimulates SP and calcitonin gene-related peptide (CGRP)-containing neurons of the trigeminal system.

Cells obtained from fluid aspirated from joints of patients with arthrosynovitis express RANTES and MCP-1, both of which are potent mast cell chemoattractants. In addition, stem cell factor (SCF) and nerve growth factor (NGF) are also known to induce mast cell migration, proliferation, and secretion. Other mast cell chemoattractants include vascular endothelial growth factor (VEGF), epidermal growth factor (EGF), and platelet-derived growth factor (PDGF).

3.1 Animal Models

In spite of the fact that TNF blockers are used clinically with considerable success in adult RA, increasing evidence indicates that IL-6 may be equally or more important. In fact, we recently showed that experimental inflammatory arthritis in mice was unaffected in TNF −/− mice. Instead, the IL-6 −/− mice were resistant to inflammatory arthritis, and humanized antibody to human IL-6 receptor inhibited collagen-induced arthritis. These findings are of particular significance as we recently showed that mast cells are the sole source of serum IL-6 elevations induced by acute stress in mice. Most recently, mast cell deficient mice were shown to be resistant to autoimmune and inflammatory arthritis.

In particular, inflammatory arthritis induced by the injection of carrageenan in the hind knee joints, increased the joint size by 1.94 ± 0.41 mm in C57BL mice in 4 days, by which time the mice were obviously limping (Fig. 2). These effects were inhibited in W/W^v mast cell deficient and corticotropin-releasing hormone (CRH) −/− mice, which were clinically indistinguishable from mice injected with normal saline (Fig. 2). These results are of interest because CRH is increased in the joints of RA patients and CRH receptors (CRHR) are present on articular mast cells, implying that CRH could trigger or regulate mast cell activation. In fact, we showed that CRH could activate mast cells in the skin and increase vascular permeability in rodents. CRH was recently shown to also induce mast cell–dependent vascular permeability in humans. The pathophysiological implication of such finding is that CRH could be released locally under stress and could exacerbate inflammatory diseases (Fig. 3).

It is quite interesting that stress increased serum IL-6 levels in care givers of chronically ill patients, but it was decreased in those who went to church regularly, indicating that reduction of stress could lead to decrease in a key proinflammatory cytokine producer.

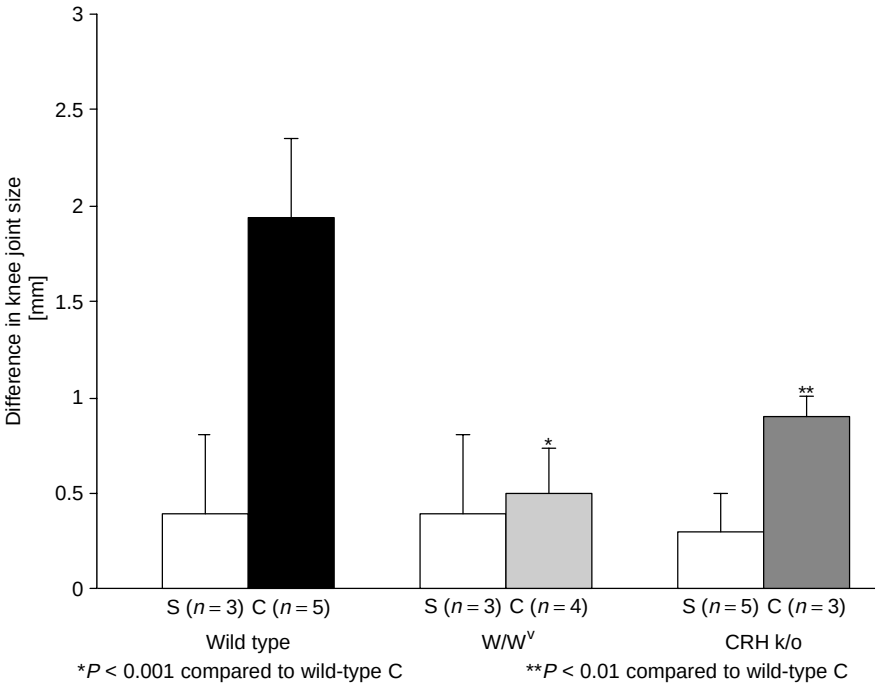

Fig. 2 Lack of inflammatory arthritis in W/W^v mast cell deficient and CRH −/− mice. W/W^v = mast cell deficient mice; k/o = knockout mice; S = normal saline; C = carrageenan.

The K/BxN mouse has been used as a model for human inflammatory arthritis and it was recently shown that the presence of autoantibodies against glucose-G phosphate isomerase is dependent on mast cells.

4 Pathophysiology of Mast Cells

Mast cells derive from a bone marrow progenitor and mature in tissues depending on microenvironmental conditions; they are characterized by the surface expression of the stem cell factor (CSF) receptor c-kit and the high affinity receptor for immunoglobulin E (FcεRI). Mast cells are important not only for allergic reactions but also in inflammation, autoimmunity, and inflammatory arthritis. However, though mast cells do not always express their c-kit receptor or FceRJ, they do so during inflammation. They are ubiquitous in the body, including the brain and joints that are not known to develop allergic reactions. Mature mast cells vary considerably in their cytokine and proteolytic enzyme content. However, the phenotypic expression of mast cells does not appear to be fixed.

In fact, mast cells cultured together with pheochromocytoma cells were shown to develop contacts with the latter and acquire the ability to respond to neuropeptides.

4.1 Triggers

In addition to IgE and antigen, anaphylatoxins, C3a, C5a antibody light chains, bacterial or viral superantigens aggregated

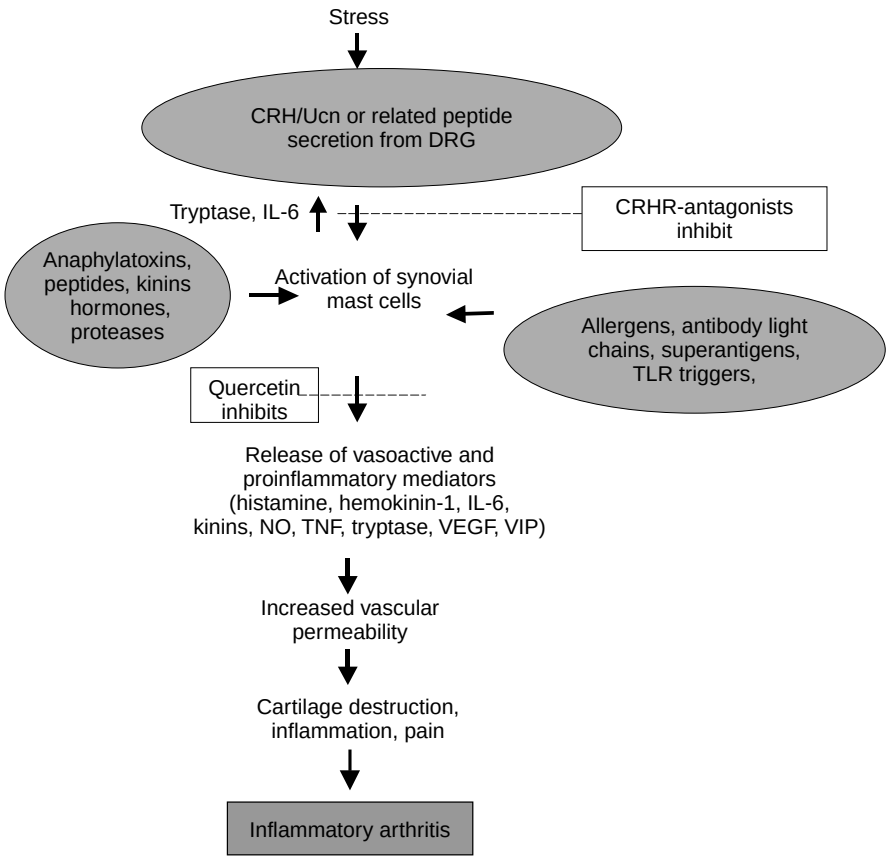

Fig. 3 Schematic representation of the possible role of mast cells in inflammatory arthritis. DRG = dorsal root ganglia; NO = nitric oxide; VIP = vasoactive intestinal peptide; VEGF = vascular endothelial growth factor.

IgG cytokines, histamine-releasing factors, hormones, and neuropeptides can trigger mast cell secretion (Table 1). The latter include SP, somatostatin, neurotensin (NT), parathyroid hormone, and NGF, which is also released under stress.

Mast cells can also be activated by endothelin, and blockade of the endothelin receptor was recently shown to block *in vivo* effects attributed to mast cells in mice. Proteases released from mast cells or other inflammatory cells can generate biologically active peptides, like histamine-releasing peptide (HRP) from plasma albumin, and HRP is found in synovial fluids obtained from patients with RA. Adherent *Escherichia coli* can activate mast cells that express Toll-like (Toll) receptors 2, activated by *E. coli* lipopolysaccharide (LPS); such receptors are critical in innate and acquired immunity.

4.2 Mediators

Mast cells synthesize and secrete over 50 biologically potent molecules that include most of the known cytokines. They

Tab. 1 Triggers of mast cell activation.

Antigens	***Hormones***
Allergens/IgE	ACTH
Aggregated IgE	CRH
Antibody light chains	Estradiol
Chemokine receptor ligands (CXCR3)	PTH
Superantigens (bacterial, viral)	Urocortin
Toll-like receptor (TLR 2, 4, 9) triggers (LPS, PDG, dd DNA)	*Peptides*
	Albumin-derived histamine-releasing molecules
Anaphylatoxins	Bradykinin
C3a, C5a	CGRP
	Endorphins
Cytokines	Endothelin
IL-1	Histamine-releasing factor
TNF-α	NT
	SRIF
Drugs	SP
Contrast media	VIP
Curare	
Morphine	*Neurotransmitters*
	Acetylcholine
Enzymes	Purines
Chymase	
PARs	*Free Radicals*
Tryptase	Oxygen
	Hydroxy
Growth Factors	
NGF	
SCF	

Notes: dd: double stranded; LPS: lipopolysaccharides; PDG: peptidoglycan; TLR: Toll-like receptor.

can, therefore, easily induce vasodilation, vascular endothelial molecule expression, and chemoattraction, which is necessary for the initiation of inflammation.

These vasodilatory, angiogenic, proinflammatory, and neurosensitizing mediators (Table 2) include histamine, heparin, kinins, neuropeptides such as vasoactive intestinal peptide (VIP) and proteases, such as chymase and tryptase (preformed) as well as chemoattractants, cytokines, growth factors, leukotrienes, prostaglandins, nitric oxide (NO), SCF, and VEGF (newly synthesized) (Table 2).

The most abundant mediators are proteases that can cause microvascular leakage, activation of PAR receptors and inflammation, as well as activation of neuropeptide precursors and direct proteolytic damage to synovium. Mast cells are also particularly rich in TNF-α and IL-6 both of which are implicated in arthritis, as discussed earlier. Mast cells are also the richest source of CRH outside the brain and could help propagate local inflammatory processes in response to chemical, mechanical, or oxidative stress.

Tab. 2 Mast cell mediators and their biologic effects.

Molecules	***Major actions***
Prestored	
Enzymes	
Arylsulfatases	Lipid/proteoglycan hydrolysis
Carboxypeptidase A	Peptide processing
Cholinesterase	Angiotensin II synthesis
Chymase	Kinins synthesis vasodilation, *pain*
Kinogenases	Arachidonic acid generation
Phospholipases	Angiotensin I generation
Renin	Tissue damage, microvascular leakage, gelatin and matrix degradation inflammation, *pain*
Tryptase (7 such proteases)	
Biogenic Amines	
Histamine	Vasodilation, angiogenesis, mitogenesis, *pain*
5-HT (serotonin)	Vasoconstriction, *pain*
Peptides	
Bradykinin	Vasodilation, *pain*
CRH	Vasodilation, inflammation
Endorphins	Analgesia
SP	Anti-inflammatory (?)
SRIF	Inflammation, *pain*
Ucn	Vasodilation, inflammation
VIP	Vasodilation
Proteoglycans	
Chondroitin sulfate	Cartilage synthesis, anti-inflammatory
Heparin	Angiogenesis, NGF stabilization
Hyaluronic acid	Cell surface recognition
De novo synthesized	
LTB_4	Leukocyte chemotaxis
LTC_4	Vasoconstriction, *pain*
NO	Vasodilation
PAF	Platelet activation and serotonin release
PGD2	Vasodilation, *pain*
Chemokines	
IL-8, MCP-1, MCP-3, MCP-4, RANTES	Chemoattraction, leukocyte infiltration
Cytokines	
IL-1,2,3,4,5,6,9,10,13,16	Inflammation, leukocyte migration, *pain*
INF-α; MIF; TNF-α	Inflammation, leukocyte proliferation/activation
Growth Factors	
CSF, GM-CSF, β-FGF, NGF, VEGF	Endothelial, immune and neuronal cell growth

4.3 Selective Secretion

One of the reasons mast cells have remained fairly obscure in inflammatory arthritis has been the fact that they appear "intact," giving the impression that they are quiescent.

Unlike degranulation, where most of the 500 secretory granules, each 1000 nm in diameter release their contents in an explosive manner, mast cells can undergo

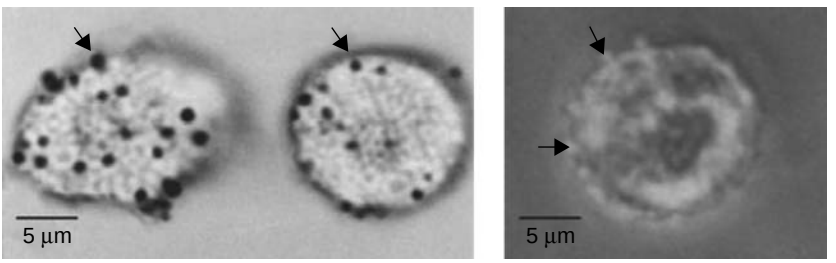

(a) (b)

Fig. 4 Punctate release of contents of individual secretory granules from purified rat peritoneal mast cells incubated with calcium containing liposomes. (a) Content of individual granules (arrow) visualized by ruthenium red staining (with a complementary filter to show black) using Nomarski optics; (b) individual granules in the process of exocytosis from a mast cell visualized by ruthenium red staining.

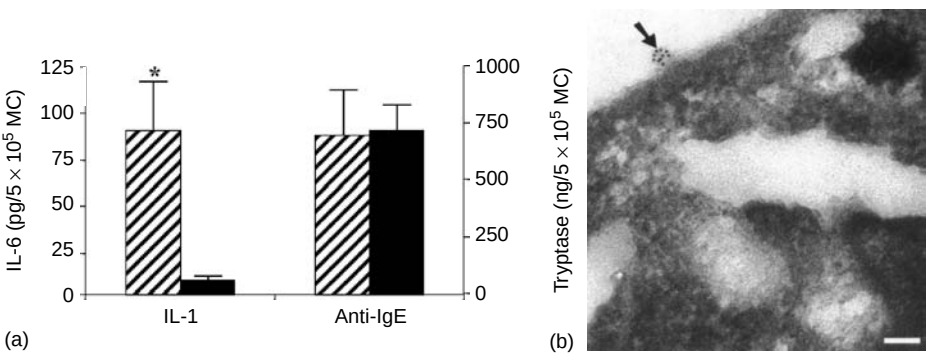

(a) (b)

Fig. 5 Selective vesicular IL-6 release induced by IL-1 from normal human cultured mast cells. (a) IL-6 and tryptase release in response to either IL-1 or anti-IgE ($n = 8$ $^*p < 0.05$); (b) a vesicle (arrow) budding from a mast cell observed by cryo-immuno-ultrastructural microscopy "white" areas represent secretory granules. Scale bar = 70 nm.

secretion of the content of individual granules (Fig. 4).

Mast cells also have the ability to release some mediators selectively without degranulation. This unique phenomenon has been shown for biogenic amines, eicosanoids, and IL-6. In fact, IL-1 was shown to induce selective release of IL-6 without histamine or tryptase, through a process that involves small vesicles (50–70 nm each in diameter) (Fig. 5). Mast cells *in vivo* also often have ultrastructural alterations of their electron dense granular core indicative of secretion, but without degranulation, a process termed *activation*, or *intragranular activation*, or *piecemeal* degranulation. This appearance was prominent in brain mast cells of nonhuman primates with experimental allergic encephalomyelitis (EAE), and in human inflammatory conditions such as inflammatory bowel disease (IBD) or interstitial cystitis (IC) of the urinary bladder. Moreover, in certain diseases such as

scleroderma and IC, mast cells appear totally depleted of their granule content and they could not be recognized by light microscopy (phantom mast cells).

Unpublished evidence indicates that synovial mast cells in RA undergo mostly selective release of mediators that is not obvious by routine histopathology.

5 Neurohormonal Stimulation of Mast Cells and Inflammatory Disorders

Mast cells could also be activated by antidromic nerve stimulation as shown for trigeminal or cervical ganglion stimulation that may result in neuropeptide secretion of reactive fibers when localized close to mast cells, release of SP from sensory afferents could stimulate mast cell secretion *in vivo*. NT also could be released from dorsal root ganglia (DRG) alone or together with CRH and further stimulate mast cells.

5.1 Anatomical Observations

Mast cells are located perivascularly in close proximity to neurons. A functional association between mast cells and neurons has been reported in 1999, and the potential pathophysiological role of brain mast cells has been reviewed by Theoharides in 1996.

5.2 Stress and the HPA Axis

Chronic stress affects illness and suppresses the immune system. Acute stress, however, can exacerbate inflammatory syndromes. In fact, there is evidence that acute stress can stimulate the immune system. Stress has been shown to worsen arthritis and activate mast cells.

Corticotropin-releasing hormone or CRH is a 41-amino acid peptide that regulates the hypothalamic-pituitary adrenal (HPA) axis and coordinates the stress response through activation of the sympathetic nervous system. CRH acts through specific receptors, which include CRHR-1 and CRHR-2. Both receptor types are located on brain neurons, but CRHR-2 has also been identified on cerebral arterioles that could be stimulated by CRH and urocortin (Ucn) directly. CRHR-2 has been further subdivided to CRHR-2α and CRHR-2β, which are best activated by Ucn, a peptide with about 50% structural similarity to CRH; Ucn II and Ucn III have also been identified and are potent CRHR-2 agonists.

5.3 Proinflammatory Effects of Stress

Both CRH and CRH mRNA have been demonstrated in rodent spleen and thymus, while human peripheral blood leukocytes and enterochromaffin cells express mRNA for Ucn. CRH and CRHR mRNA is expressed in rodent and human skin, CRH-like immunoreactivity is present in the dorsal horn of the spinal cord and in DRG, as well as in sympathetic ganglia. We recently showed that acute stress could increase the skin content of CRH, most likely from DRG. We also showed that human mast cells contain both CRH and Ucn that could be released in response to immunologic stimulation.

CRH can stimulate leukocytes to produce β-endorphin, adrenocorticotropic hormone (ACTH), and α-melanocyte stimulating hormone (α-MSH), as well as monocytes to secrete interleukin-1 (IL-1), and lymphocytes to produce IL-2.

CRH also stimulated lymphocyte proliferation, increased IL-2 receptor expression on T lymphocytes, was chemotactic for mononuclear leukocytes and activated CRHR-1 on spleen cells.

Relevant examples of possible proinflammatory actions of CRH include carrageenan-induced aseptic inflammation and arthritis, where both CRH and Ucn have been identified in the joints. Moreover, inflammatory arthritis shown to be absent in W/W^v mast cell deficient mice, was blocked by the preferential CRHR-1 antagonist Antalarmin and was greatly attenuated in CRH knockout mice. In fact, mast cells in the joints of RA patients were reported to express CRH receptors.

5.4 CRH Effects on Mast Cells

Mast cells are localized close to CRH positive neurons in the rat median eminence. CRH administration in humans causes peripheral vasodilation and flushing reminiscent of mast cell activation. Iontophoresis of CRH increased human skin vasodilation, detected by laser Doppler, which was dependent on CRHR-1 and mast cells. Intradermal CRH administration leads to histamine-dependent swelling; in rodents, intradermal CRH leads to activation of mast cells and Evans blue extravasation was also induced. Moreover, CRHR-1 is involved in stress-induced exacerbation of chronic contact dermatitis in rats.

In addition to CRH stimulating mast cells, mast cell mediators could influence CRH release. For instance, human mast cells synthesize and secrete large amounts of CRH. Histamine can increase CRH mRNA expression in the hypothalamus, and mast cells could stimulate the HPA axis. This could occur possibly through IL-6 and IL-1, both of which are released from mast cells; conversely, CRH stimulates IL-6 release. Acute stress also increased BBB permeability in rats and mice only in brain areas containing mast cells.

CRH Ucn can directly activate mast cells that express CRHRs; in fact, mast cells were recently shown to express five CRHR-1 isoforms that could have different functions.

6 Therapy

Typical therapy for osteoarthritis involves weight loss, exercise, and use of nonsteroidal anti-inflammatory drugs (NSAIDs). In RA, active articular inflammation often requires immunosuppressant and immune modifying drugs. Severe RA requires multiple agents that include steroids, methotrexate, penicillamine, gold salts, and most recently TNF blockers, such as etanercept or infliximab. Unfortunately, recent evidence that cyclooxygenase 2 (COX-2) inhibitors increase cardiovascular disease has forced the withdrawal or limited use of such drugs.

Except for glucocorticoids, there are no clinically effective mast cell inhibitors.

6.1 Proteoglycans

Glucosamine sulfate presumably acts as a building block for new cartilage, while chondroitin sulfate acts as a "ready-made" component.

D-Glucosamine, *N*-acetylglucosamine, glucosamine sulfate, and chondroitin are commonly used for arthritis. A meta analysis of clinical trials using glucosamine

and/or chondroitin originally had indicated potential usefulness in OA. The validation of WOMAC, a self-administered health status instrument for OA has helped with the design of double-blind studies; similarly the Cedars–Sinai Health Related Quality of Life Instrument (CSHQ-RA) can be used for RA research. One of the first double-blind studies compared 500-mg glucosamine sulfate tid to 400-mg ibuprofen tid for 4 weeks. From the second week onwards, the clinical improvement was the same (~50%) in both groups, but 35% of those on ibuprofen reported adverse effects as compared to 6% on glucosamine. At least two studies have shown that 1500 mg of glucosamine per day for 3 years can delay progression of OA and knee space reduction. In another randomized, double-blind placebo-controlled trial, 212 patients with OA of the knee were assigned either 1500 mg of oral glucosamine sulfate or a placebo for 3 years. Patients on active treatment had clinical improvement of symptoms and no significant joint space loss, compared to progressive narrowing of those on placebo. In a subsequent study, 202 patients with OA of the knee were randomized to either 1500 mg of glucosamine sulfate or placebo for 3 years; symptoms on the active group improved by about 25% and joint space narrowing (>0.5 mm) occurred in 5% of those on the active arm as compared to 15% on the placebo. However, more recent studies with glucosamine, especially when administered for shorter duration, have failed to show consistent benefit. Moreover, use of 1500 mg of glucosamine per day could interfere with insulin sensitivity.

The proteoglycan chondroitin sulfate has also been added to glucosamine, but there are a number of problems with such preparations. Very little oral chondroitin sulfate is absorbed in powder form due to its high molecular weight (150 000–1 000 000 Da) and degree of sulfation. In particular, less than 5% of chondroitin sulfate is absorbed intact when administered orally. In another study, after oral intravenous administration of chondroitin sulfate (16 000 Da), the absolute bioavailability was 13%, but when chondroitin sulfate of about double the size (26 000 Da) was used, less than 4% of the oral dose administered in rats was intact chondroitin sulfate in the blood. However, the smaller size (16 000 Da) has almost no protective value in laboratory experiments. Other papers found no oral absorption of chondroitin sulfate and concluded that any protective effect in the joints after oral administration was unfounded.

There are no good studies on the use of proteoglycans in RA, presumably because the aim has traditionally been the synovial inflammatory response. However, chondroitin sulfate was shown to inhibit activation of connective tissue mast cells. There is also some evidence that bacterial adhesion and invasion depends on sulfated surface polysaccharides and use of chondroitin sulfate as "decoy" may prevent bacteria from adhering to the cell surface and causing infection. Aloe vera has been reported to reduce mast cell secretion and mast cell infiltration in an inflamed synovial pouch model.

6.2 Flavonoids

Few clinically available drugs can effectively inhibit human mast cell activation. For instance, even though disodium cromoglycate (cromolyn) had been known to inhibit activation of rodent mast cells, it was unable to inhibit human mast cell activation. The most well documented evidence published to date on the inhibitory

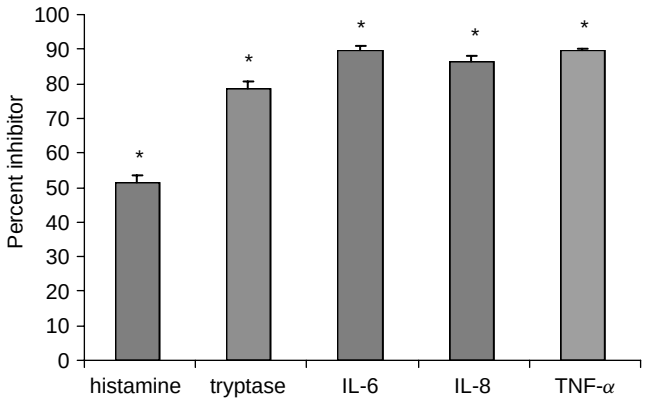

Fig. 6 Inhibitory effect of quercetin on normal human mast cell proinflammatory mediator secretion. Mast cells were sensitized with human-IgE and was pretreated for 15 min prior to stimulation with anti-IgE (10 μg mL^{-1}) for 30 min (histamine, $n = 6–31$; and tryptase, $n = 6–16$) and for 6 h (IL-6, IL-8, and TNF-α, $n = 4–7$) at 37 °C. Quercetin significantly inhibited all these mediators when compared to anti-IgE alone treated mast cells. $* = p < 0.05$.

action of mast cells has focused on the naturally occurring flavonoids. For instance, quercetin can inhibit not only the prestored mediators histamine and tryptase from normal human mast cells but also the synthesis of the cytokines IL-6, IL-8, and TNF-α (Fig. 6).

Some flavonoids like morin have weak inhibitory activity, even though it differs from quercetin on the addition of one hydroxyl group in the β-phenolic ring.

6.3 Synergistic Effects

However, chondroitin sulfate appears to block mast cell activation and histamine release; quercetin blocks mast cell secretion and particularly cytokine secretion. The two molecules together could have synergistic actions. Patients do not have an allergic reaction to glucosamine sulfate or chondroitin sulfate if they are allergic to sulfonamide antibiotics. Rutin, the glycoside form of quercetin, also has antiarthritis properties.

The possibility is rather unlikely, because there was absence of cross-reactivity between sulfonamide antibiotics and sulfonamide nonantibiotics.

7 Perspectives

Evidence dating back over 20 years that mast cells may participate in inflammation through selective release of proinflammatory molecules had been ignored until recently (see Figure 7). It now appears that mast cells are critical for the initiation of inflammation, especially in arthritis. Unfortunately, the triggers of synovial mast cell activation are not known and there are no clinically available mast cell inhibitors.

A combination of proteoglycans and select flavonoids, such as quercetin (e.g. ArthroSoft®) together with low dose buffered salicylates, (e.g. Trilisate) may

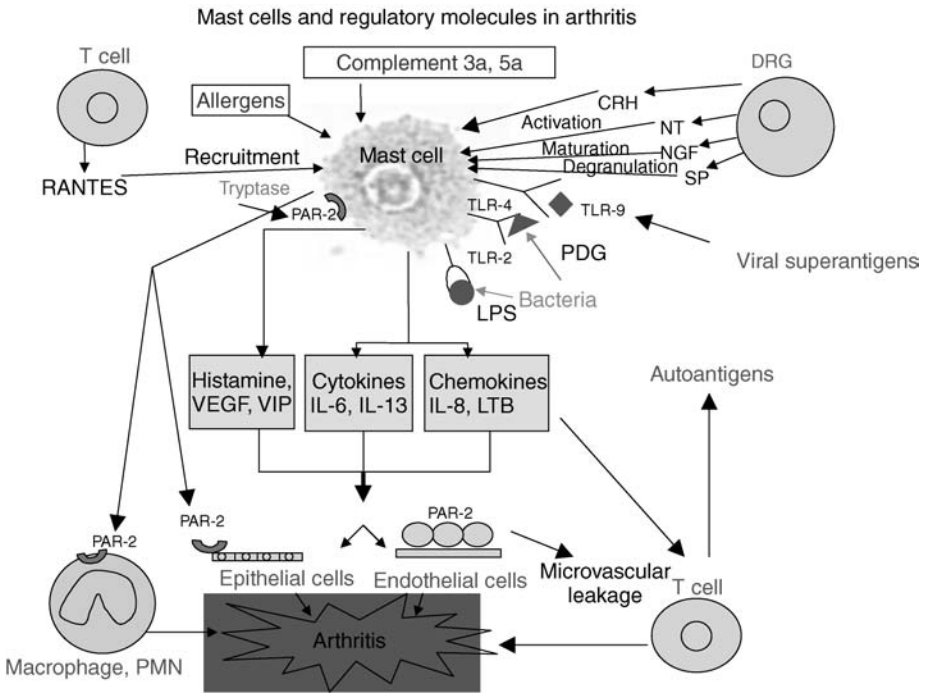

Fig. 7 Diagrammatic representation of synovial mast cells interaction with other cells in the initiation of synovial inflammation and arthritis.

be the best initial approach for RA. The additional benefit of quercetin is its unique ability to inhibit secretion of many proinflammatory molecules, especially IL-6 that has been increasingly implicated in juvenile RA. Such molecules could be combined with CRH receptor antagonists when they become clinically available.

Acknowledgment

Aspects of our own research have been funded by NIH grants No. DK 42409, DK 44816, DK 62861, NS38326, and AR47652. Use of CRH receptor antagonists, as well as proteoglycans and flavonoids, for inflammatory conditions is covered by US patents # 6,020,305; 6,624,148; # 6,635,625; 6,641,806; 6,645,482; 6,689,748 and 09/771,669 (allowed Aug 10, 2004) awarded to TCT and assigned to Theta Biomedical Consulting and Development, Co., Inc. (Brookline, MA). Thanks are due to Dr D. Kempuraj and Dr N. Papadopoulou, as well as to Mrs. Jill Donelan and Mrs. Jessica Christian for some of the graphics, and to the latter also for her patience and word processing skills.

See also Autoantibodies and Autoimmunity; Growth Factors; Immunology; Innate Immunity; Molecular Mediators: Cytokines; Transcription Factor NF-κB: Function,

Structure, Regulation, Pathways, and Applications.

Bibliography

Books and Reviews

Akira, S., Takeda, K., Kaisho, T. (2001) Toll-like receptors: critical proteins linking innate and acquired immunity, *Nat. Immunol.* **2**, 675–680.

Chalmers, D.T., Lovenberg, T.W., Grigoriadis, D.E., Behan, D.P., DeSouza, E.B. (1996) Corticotropin-releasing factor receptors: from molecular biology to drug design, *Trends Pharmacol. Sci.* **17**, 166–172.

Chrousos, G.P. (1995) The hypothalamic-pituitary-adrenal axis and immune-mediated inflammation, *N. Engl. J. Med.* **332**, 1351–1362.

Dautzenberg, F.M., Hauger, R.L. (2002) The CRF peptide family and their receptors: yet more partners discovered, *Trends Pharmacol. Sci.* **23**, 71–77.

Dieterich, K.D., Lehnert, H., De Souza, E.B. (1997) Corticotropin-releasing factor receptors: an overview, *Exp. Clin. Endocrinol. Diabetes* **105**, 65–82.

Fitzgerald, G.A. (2005) Coxibs and cardiovascular disease, *N. Engl. J. Med.* **351**, 1709–1711.

Foreman, J.C. (1987) Peptides and neurogenic inflammation, *Brain Res. Bull.* **43**, 386–398.

Galli, S.J. (1993) New concepts about the mast cell, *N. Engl. J. Med.* **328**, 257–265.

Goetzl, E.J., Cheng, P.P.J., Hassner, A., Adelman, D.C., Frick, O.L., Speedharan, S.P. (1990) Neuropeptides, mast cells and allergy: novel mechanisms and therapeutic possibilities, *Clin. Exp. Allergy* **20**, 3–7.

Grammatopoulos, D.K., Chrousos, G.P. (2002) Functional characteristics of CRH receptors and potential clinical applications of CRH-receptor antagonists, *Trends Endocrinol. Metab.* **13**, 436–444.

Karalis, K., Louis, J.M., Bae, D., Hilderbrand, H., Majzoub, J.A. (1997) CRH and the immune system, *J. Neuroimmunol.* **72**, 131–136.

Khansari, D.N., Murgo, A.J., Faith, R.E. (1990) Effects of stress on the immune system, *Immunol. Today* **11**, 170–175.

Lindahl, U., Hook, M. (1978) Glycosaminoglycans and their binding to biological macromolecules, *Annu. Rev. Biochem.* **47**, 385–417.

Manek, N.J. (2001) Medical management of osteoarthritis, *Mayo Clin. Proc.* **76**, 533–539.

McAlindon, T. (2003) Why are clinical trials of glucosamine no longer uniformly positive? *Rheum. Dis. Clin. North Am.* **29**, 789–801.

Rabkin, J.G., Struening, E.L. (1976) Life events, stress and illness, *Science* **194**, 1013–1020.

Redegeld, F.A., Nijkamp, F.P. (2005) Immunoglobulin free light chains and mast cells: pivotal role in T-cell-mediated immune reactions? *Trends Immunol.* **24**, 181–185.

Rosenwasser, L.J., Boyce, J.A. (2003) Mast cells: beyond IgE, *J. Allergy Clin. Immunol.* **111**, 24–32.

Rostand, K.S., Esko, J.D. (1997) Minireview: microbial adherence to and invasion through proteoglycans, *Infect. Immunol.* **65**, 1–8.

Schmidlin, F., Bunnett, N.W. (2001) Protease-activated receptors: how proteases signal to cells, *Curr. Opin. Pharmacol.* **1**, 575–582.

Schwartz, L.B. (1987) Mediators of human mast cells and human mast cell subsets, *Ann. Allergy* **58**, 226–235.

Serafin, W.E., Austen, K.F. (1987) Mediators of immediate hypersensitivity reactions, *N. Engl. J. Med.* **317**, 30–34.

Stead, R.H., Bienenstock, J. (1990) Cellular Interactions Between the Immune and Peripheral Nervous System. A Normal Role for Mast Cells? in: Burger, M.M., Sordat, B., Zinkernagel, R.M. (Eds.) *Cell to Cell Interaction*, Karger, Basel, pp. 170–187.

Taylor, P.C. (2003) Anti-TNFalpha therapy for rheumatoid arthritis: an update, *Intern. Med.* **42**, 15–20.

Theoharides, T.C. (1996) Mast cell: a neuroimmunoendocrine master player, *Int. J. Tissue React.* **18**, 1–21.

Theoharides, T.C. (2002) Mast cells and stress – a psychoneuroimmunological perspective,
J. Clin. Psychopharmacol. **22**, 103–108.

Theoharides, T.C. (2003) Dietary supplements for arthritis and other inflammatory conditions: key role of mast cells and benefit of combining anti-inflammatory and proteoglycan products, *Eur. J. Inflamm.* **1**, 1–8.

Theoharides, T.C., Bielory, L. (2004) Mast cells and mast cell mediators as targets of dietary

supplements, *Ann. Allergy Asthma Immunol.* **93**, S24–S34.

Theoharides, T.C., Cochrane, D.E. (2004) Critical role of mast cells in inflammatory diseases and the effect of acute stress, *J. Neuroimmunol.* **146**, 1–12.

Theoharides, T.C., Sant, G.R. (2003) Neuroimmune Connections and Regulation of Function in the Urinary Bladder, in: Bienenstock, J., Goetzl, E., Blennerhassett, M. (Eds.) *Autonomic Neuroimmunology*, Hardwood Academic Publishers, Lausanne, Switzerland, pp. 345–369.

Theoharides, T.C., Alexandrakis, M., Kempuraj, D., Lytinas, M. (2001) Anti-inflammatory actions of flavonoids and structural requirements for new design, *Int. J. Immunopathol. Pharmacol.* **14**, 119–127.

Theoharides, T.C., Donelan, J.M., Papadopoulou, N., Cao, J., Kempuraj, D., Conti, P. (2004) Mast cells as targets of corticotropin-releasing factor and related peptides, *Trends Pharmacol. Sci.* **25**, 563–568.

Topol, E.J. (2005) Failing the public health – rofecoxib, Merck, and the FDA, *N. Engl. J. Med.* **351**, 1707–1709.

Walker-Bone, K., Javaid, K., Arden, N., Cooper, C. (2000) Regular review: medical management of osteoarthritis, *BMJ* **321**, 936–940.

Wasserman, S.I. (1984) The mast cell and synovial inflammation, *Arthritis Rheum.* **27**, 841–844.

Woolley, D.E. (2003) The mast cell in inflammatory arthritis, *N. Engl. J. Med.* **348**, 1709–1711.

Primary Literature

Akira, S., Takeda, K., Kaisho, T. (2001). Toll-like receptors: critical proteins linking innate and acquired immunity, *Nat. Immunol.* **2**, 675–680.

Alkaabi, J.K., Ho, M., Levison, R., Pullar, T., Belch, J.J. (2003) Rheumatoid arthritis and macrovascular disease, *Rheumatology (Oxford)* **42**, 292–297.

Aloe, L., Levi-Montalcini, R. (1977) Mast cells increase in tissues of neonatal rats injected with the nerve growth factor, *Brain Res.* **133**, 358–366.

Alonzi, T., Fattori, E., Lazzaro, D., Costa, P., Probert, L., Kollias, G., De Benedetti, F., Poli, V., Ciliberto, G. (1998) Interleukin 6 is required for the development of collagen-induced arthritis, *J. Exp. Med.* **187**, 461–468.

Ando, T., Rivier, J., Yanaihara, H., Arimura, A. (1998) Peripheral corticotropin-releasing factor mediates the elevation of plasma IL-6 by immobilization stress in rats, *Am. J. Physiol.* **275**, R1461–R1467.

Angioni, S., Petraglia, F., Gallinelli, A., Cossarizza, A., Franceschi, C., Muscettola, M., Genazzani, A.D., Surico, N., Genazzani, A.R. (1993) Corticotropin-releasing hormone modulates cytokines release in cultured human peripheral blood mononuclear cells, *Life Sci.* **53**, 1735–1742.

Askenase, P.W. (2005) Mast cells and the mediation of T-cell recruitment in arthritis. *N. Engl. J. Med.* **349**, 1294.

Baici, A., Horler, D., Moser, B., Hofer, H.O., Fehr, K., Wagenhauser, F.J. (1992) Analysis of glycosaminoglycans in human serum after oral administration of chondroitin sulfate, *Rheumatol. Int.* **12**, 81–88.

Baigent, S.M. (2001) Peripheral corticotropin-releasing hormone and urocortin in the control of the immune response, *Peptides* **22**, 809–820.

Bamberger, C.M., Wald, M., Bamberger, A.-M., Ergun, S., Beil, F.U., Schulte, H.M. (1998) Human lymphocytes produce urocortin, but not corticotropin-releasing hormone, *J. Clin. Endocrinol. Metab.* **83**, 708–711.

Batler, R.A., Sengupta, S., Forrestal, S.G., Schaeffer, A.J., Klumpp, D.J. (2002) Mast cell activation triggers a urothelial inflammatory response mediated by tumor necrosis factor-alpha, *J. Urol.* **168**, 819–825.

Benyon, R., Robinson, C., Church, M.K. (1989) Differential release of histamine and eicosanoids from human skin mast cells activated by IgE-dependent and non-immunological stimuli, *Br. J. Pharmacol.* **97**, 898–904.

Bethin, K.E., Vogt, S.K., Muglia, L.J. (2000) Interleukin-6 is an essential, corticotropin-releasing hormone-independent stimulator of the adrenal axis during immune system activation, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9317–9322.

Bienenstock, J., Tomioka, M., Matsuda, H., Stead, R.H., Quinonez, G., Simon, G.T., Coughlin, M.D., Denburg, J.A. (1987) The role of mast cells in inflammatory processes: evidence for nerve mast cell interactions, *Int. Arch. Allergy Appl. Immunol.* **82**, 238–243.

Bischoff, S.C., Sellge, G., Lorentz, A., Sebald, W., Raab, R., Manns, M.P. (1999) IL-4 enhances proliferation and mediator release in mature human mast cells, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8080–8085.

Boe, A., Baiocchi, M., Carbonatto, M., Papoian, R., Serlupi-Crescenzi, O. (1999) Interleukin 6 knock-out mice are resistant to antigen-induced experimental arthritis, *Cytokine* **11**, 1057–1064.

Bradding, P., Okayama, Y., Howarth, P.H., Church, M.K., Holgate, S.T. (1995) Heterogeneity of human mast cells based on cytokine content, *J. Immunol.* **155**, 297–307.

Bridges, A.J., Malone, D.G., Jicinsky, J., Chen, M., Ory, P., Engber, W., Graziano, F.M. (1991) Human synovial mast cell involvement in rheumatoid arthritis and osteoarthritis. Relationship to disease type, clinical activity, and antirheumatic therapy, *Arthritis Rheum.* **34**, 1116–1124.

Buchanan, W.W., Goldsmith, C.H., Campbell, J., Stitt, L.W., Bellamy, N. (1988) Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee, *J. Rheumatol.* **15**, 1833–1840.

Bugajski, A.J., Chlap, Z., Gadek-Michalska, A., Borycz, J., Bugajski, J. (1995) Degranulation and decrease in histamine levels of thalamic mast cells coincides with corticosterone secretion induced by compound 48/80, *Inflamm. Res.* **44**(Supp.1), S50–S51.

Cao, J., Papadopoulou, N., Kempuraj, D., Sugimoto, K., Cetrulo, C.L., Theoharides, T.C. (2005) Human mast cells express corticotropin-releasing hormone (CRH) receptors and CRH leads to selective secretion of vascular endothelial growth factor (VEGF), *J. Immunol.*, in press.

Carraway, R., Cochrane, D.E., Lansman, J.B., Leeman, S.E., Paterson, B.M., Welch, H.J. (1982) Neurotensin stimulates exocytotic histamine secretion from rat mast cells and elevates plasma histamine levels, *J. Physiol.* **323**, 403–414.

Ceponis, A., Konttinen, Y.T., Takagi, M., Xu, J.W., Sorsa, T., Matucci-Cerinic, M., Santavirta, S., Bankl, H.C., Valent, P. (1998) Expression of stem cell factor (SCF) and SCF receptor (c-kit) in synovial membrane in arthritis: correlation with synovial mast cell hyperplasia and inflammation, *J. Rheumatol.* **25**, 2304–2314.

Chalmers, D.T., Lovenberg, T.W., DeSouza, E.B. (1995) Localization of novel corticotropin-releasing factor receptor (CRF2) mRNA expression to specific subcortical nuclei in rat brain: comparison with CRF1 receptor mRNA expression, *J. Neurosci.* **10**, 6340–6350.

Chalmers, D.T., Lovenberg, T.W., Grigoriadis, D.E., Behan, D.P., DeSouza, E.B. (1996) Corticotropin-releasing factor receptors: from molecular biology to drug design, *Trends Pharmacol. Sci.* **17**, 166–172.

Chen, R., Lewis, K.A., Perrin, M.H., Vale, W.W. (1993) Expression cloning of a human corticotropin-releasing factor receptor, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8967–8971.

Chrousos, G.P. (1995) The hypothalamic-pituitary-adrenal axis and immune-mediated inflammation, *N. Engl. J. Med.* **332**, 1351–1362.

Claman, H.N., Choi, K.L., Sujansky, W., Vatter, A.E. (1986) Mast cell "disappearance" in chronic murine graft-vs-host disease (GVHD)-ultrastructural demonstration of "phantom mast cells", *J. Immunol.* **137**, 2009–2013.

Clifton, V.L., Crompton, R., Smith, R., Wright, I.M. (2002) Microvascular effects of CRH in human skin vary in relation to gender, *J. Clin. Endocrinol. Metab.* **87**, 267–270.

Cochrane, D.E., Carrawan, R.E., Miller, L.A., Feldberg, R.S., Bernheim, H. (2003) Histamine releasing peptide (HRP) has proinflammatory effects and is present at sites of inflammation, *Biochem. Pharmacol.* **66**, 331–342.

Conte, A., de Bernardi, M., Palmieri, L., Lualdi, P., Mautone, G., Ronca, G. (1991) Metabolic fate of exogenous chondroitin sulfate in man, *Arzneimittelforschung* **41**, 768–772.

Conte, A., Volpi, N., Palmieri, L., Bahous, I., Ronca, G. (1995) Biochemical and pharmacokinetic aspects of oral treatment with chondroitin sulfate, *Arzneimittelforschung* **45**, 918–925.

Conti, P., Pang, X., Boucher, W., Letourneau, R., Reale, M., Barbacane, R.C., Thibault, J., Theoharides, T.C. (1997) Impact of Rantes and MCP-1 chemokines on *in vivo* basophilic mast cell recruitment in rat skin injection model and their role in modifying the protein and mRNA levels for histidine decarboxylase, *Blood* **89**, 4120–4127.

Conti, P., Reale, M., Barbacane, R.C., Castellani, M.L., Orso, C. (2002) Differential production of RANTES and MCP-1 in synovial fluid from the inflamed human knee, *Immunol. Lett.* **80**, 105–111.

Conti, P., Reale, M., Barbacane, R.C., Letourneau, R., Theoharides, T.C. (1998) Intramuscular injection of hrRANTES causes mast cell recruitment and increased transcription of histidine decarboxylase: lack of effects in genetically mast cell-deficient W/W^v mice, *FASEB J.* **12**, 1693–1700.

Correa, S.G., Riera, C.M., Spiess, J., Bianco, I.D. (1997) Modulation of the inflammatory response by corticotropin-releasing factor, *Eur. J. Pharmacol.* **319**, 85–90.

Crisp, A.J., Champan, C.M., Kirkham, S.E., Schiller, A.L., Keane, S.M. (1984) Articular mastocytosis in rheumatoid arthritis, *Arthritis Rheum.* **27**, 845–851.

Crofford, L.J., Sano, H., Karalis, K., Webster, E.A., Friedman, T.C., Chrousos, G.P., Wilder, R.L. (1995) Local expression of corticotropin-releasing hormone in inflammatory arthritis, *Ann. N.Y. Acad. Sci.* **771**, 459–471.

Crofford, L.J., Sano, H., Karalis, K., Friedman, T.C., Epps, H.R., Remmers, E.F., Mathern, P., Chrousos, G.P., Wilder, R.L. (1993) Corticotropin-releasing hormone in synovial fluids and tissues of patients with rheumatoid arthritis and osteoarthritis, *J. Immunol.* **151**, 1–10.

Crompton, R., Clifton, V.L., Bisits, A.T., Read, M.A., Smith, R., Wright, I.M. (2003) Corticotropin-releasing hormone causes vasodilation in human skin via mast cell-dependent pathways, *J. Clin. Endocrinol. Metab.* **88**, 5427–5432.

Dautzenberg, F.M., Hauger, R.L. (2002) The CRF peptide family and their receptors: yet more partners discovered, *Trends Pharmacol. Sci.* **23**, 71–77.

Davis, R.H., Stewart, G.J., Bregman, P.J. (1992) Aloe vera and the inflamed synovial pouch model, *J. Am. Podiatr. Med. Assoc.* **82**, 140–148.

De Benedetti, F., Martini, A. (1998) Is systemic juvenile rheumatoid arthritis an interleukin 6 mediated disease? *J. Rheumatol.* **25**, 203–207.

De Benedetti, F., Massa, M., Pignatti, P., Albani, S., Novick, D., Martini, A. (1994) Serum soluble interleukin 6 (IL-6) receptor and IL-6/soluble IL-6 receptor complex in systemic juvenile rheumatoid arthritis, *J. Clin. Invest.* **93**, 2114–2119.

de Hooge, A.S., van De Loo, F.A., Arntz, O.J., van den Berg, W.B. (2000) Involvement of IL-6, apart from its role in immunity, in mediating a chronic response during experimental arthritis, *Am. J. Pathol.* **157**, 2081–2091.

de Paulis, A., Ciccarelli, A., Marinò, I., de Crescenzo, G., Marinò, D., Marone, G. (1997) Human synovial mast cells. II. Heterogeneity of the pharmacologic effects of antiinflammatory and immunosuppressive drugs, *Arthritis Rheum.* **40**, 469–478.

de Paulis, A., Marino, I., Ciccarelli, A., de Crescenzo, G., Concardi, M., Verga, L., Arbustini, E., Marone, G. (1996) Human synovial mast cells. I. Ultrastructural in situ and in vitro immunologic characterization, *Arthritis Rheum.* **39**, 1222–1233.

De Simone, R., Alleva, E., Tirassa, P., Aloe, L. (1990) Nerve growth factor released into the bloodstream following intraspecific fighting induces mast cell degranulation in adult male mice, *Brain Behav. Immun.* **4**, 74–81.

Dhabhar, F., McEwen, B.S. (1996) Stress-induced enhancement of antigen-specific cell-mediated immunity, *J. Immunol.* **156**, 2608–2615.

Dhabhar, F.S., McEwen, B.S. (1999) Enhancing versus suppressive effects of stress hormones on skin immune function, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1059–1064.

Dhabhar, F.S., Miller, A.H., McEwen, B.S., Spencer, R.L. (1995) Effects of stress on immune cell distribution: dynamics and hormonal mechanisms, *J. Immunol.* **154**, 5511–5527.

Dieterich, K.D., Lehnert, H., De Souza, E.B. (1997) Corticotropin-releasing factor receptors: an overview, *Exp. Clin. Endocrinol. Diabetes* **105**, 65–82.

Dimitriadou, V., Pang, X., Theoharides, T.C. (2000) Hydroxyzine inhibits experimental allergic encephalomyelitis (EAE) and associated brain mast cell activation, *Int. J. Immunopharmacol.* **22**, 673–684.

Dimitriadou, V., Aubineau, P., Taxi, J., Seylaz, J. (1987) Ultrastructural evidence for a functional unit between nerve fibers and type II cerebral mast cells in the cerebral vascular wall, *Neuroscience* **22**, 621–630.

Dimitriadou, V., Lambracht-Hall, M., Reichler, J., Theoharides, T.C. (1990) Histochemical and ultrastructural characteristics of rat brain

perivascular mast cells stimulated with compound 48/80 and carbachol, *Neuroscience* **39**, 209–224.

Dimitriadou, V., Buzzi, M.G., Moskowitz, M.A., Theoharides, T.C. (1991) Trigeminal sensory fiber stimulation induces morphologic changes reflecting secretion in rat dura mast cells, *Neuroscience* **44**, 97–112.

Dimitriadou, V., Rouleau, A., Trung Tuong, M.D., Newlands, G.J.F., Miller, H.R.P., Luffau, G., Schwartz, J.-C., Garbarg, M. (1997) Functional relationships between sensory nerve fibers and mast cells of dura mater in normal and inflammatory conditions, *Neuroscience* **77**, 829–839.

Ditzel, H.J. (2005) The K/BxN mouse: a model of human inflammatory arthritis, *Trends Mol. Med.* **10**, 40–45.

Douglas, W.W. (1974) Involvement of calcium in exocytosis and the exocytosis-vesiculation sequence, *Biochem. Soc. Symp.* **39**, 1–28.

Duensing, T.D., Wing, J.S., vanPutten, J.P.M. (1999) Sulfated polysaccharide-directed recruitment of mammalian host proteins: a novel strategy in microbial pathogenesis, *Infect. Immunol.* **67**, 4463–4468.

Dunlop, D.D., Manheim, L.M., Yelin, E.H., Song, J., Chang, R.W. (2003) The costs of arthritis, *Arthritis Rheum.* **49**, 101–113.

Dvorak, A.M., Macglashan, D.W., Morgan, E.S., Lichtenstein, L.M. Jr. (1996) Vesicular transport of histamine in stimulated human basophils, *Blood* **88**, 4090–4101.

Dvorak, A.M., McLeod, R.S., Onderdonk, A.B., Monahan-Earley, R.A., Cullen, J.B., Antonioli, D.A., Morgan, E., Blair, J.E., Estrella, P., Cisneros, R.L., Cohen, Z., Silen, W. (1992) Human gut mucosal mast cells: ultrastructural observations and anatomic variation in mast cell-nerve associations *in vivo*, *Int. Arch. Allergy Immunol.* **98**, 158–168.

Esposito, P., Basu, S., Letourneau, R., Jacobson, S., Theoharides, T.C. (2003) Corticotropin-releasing factor (CRF) can directly affect brain microvessel endothelial cells, *Brain Res.* **968**, 192–198.

Esposito, P., Gheorghe, D., Kandere, K., Pang, X., Conally, R., Jacobson, S., Theoharides, T.C. (2001) Acute stress increases permeability of the blood-brain-barrier through activation of brain mast cells, *Brain Res.* **888**, 117–127.

Fajardo, I., Pejler, G. (2003) Human mast cell beta-tryptase is a gelatinase, *J. Immunol.* **171**, 1493–1499.

Fitzgerald, G.A. (2005) Coxibs and cardiovascular disease, *N. Engl. J. Med.* **351**, 1709–1711.

Foreman, J.C. (1987) Peptides and neurogenic inflammation, *Brain Res. Bull.* **43**, 386–398.

Gadek-Michalska, A., Chlap, Z., Turon, M., Bugajski, J., Fogel, W.A. (1991) The intracerebroventicularly administered mast cells degranulator compound 48/80 increases the pituitary-adrenocortical activity in rats, *Agents Actions* **32**, 203–208.

Gagari, E., Tsai, M., Lantz, C.S., Fox, L.G., Galli, S.J. (1997) Differential release of mast cell interleukin-6 via c-kit, *Blood* **89**, 2654–2663.

Galli, S.J. (1993) New concepts about the mast cell, *N. Engl. J. Med.* **328**, 257–265.

Genedani, S., Bernardi, M., Baldini, M.G., Bertolini, A. (1992) Influence of CRH and α-MSH on the migration of human monocytes *in vitro*, *Neuropeptides* **23**, 99–102.

Genovese, A., Borgia, G., Bouvet, J.P., Detoraki, A., de Paulis, A., Piazza, M., Marone, G. (2005) Protein Fv produced during viral hepatitis is an endogenous immunoglobulin superantigen activating human heart mast cells, *Int. Arch. Allergy Immunol.* **132**, 336–345.

Goetzl, E.J., Cheng, P.P.J., Hassner, A., Adelman, D.C., Frick, O.L., Speedharan, S.P. (1990) Neuropeptides, mast cells and allergy: novel mechanisms and therapeutic possibilities, *Clin. Exp. Allergy* **20**, 3–7.

Gotis-Graham, I., Smith, M.D., Parker, A., McNeil, H.P. (1998) Synovial mast cell responses during clinical improvement in early rheumatoid arthritis, *Ann. Rheum. Dis.* **57**, 664–671.

Grabbe, J., Welker, P., Möller, A., Dippel, E., Ashman, L.K., Czarnetzki, B.M. (1994) Comparative cytokine release from human monocytes, monocyte-derived immature mast cells and a human mast cell line (HMC-1), *J. Invest. Dermatol.* **103**, 504–508.

Grammatopoulos, D.K., Chrousos, G.P. (2002) Functional characteristics of CRH receptors and potential clinical applications of CRH-receptor antagonists, *Trends Endocrinol. Metab.* **13**, 436–444.

Hagen, L.E., Schneider, R., Stephens, D., Modrusan, D., Feldman, B.M. (2003) Use of complementary and alternative medicine by pediatric rheumatology patients, *Arthritis Rheum.* **49**, 3–6.

Hardingham, T.E. (1986) Structure and biosynthesis of proteoglycans, *J. Rheumatol.* **10**, 143–183.

He, S., Walls, A.F. (1997) Human mast cell tryptase: a stimulus of microvascular leakage and mast cell activation, *Eur. J. Pharmacol.* **328**, 89–97.

He, S., Gaca, M.D., Walls, A.F. (1998) A role for tryptase in the activation of human mast cells: modulation of histamine release by tryptase and inhibitors of tryptase, *J. Pharmacol. Exp. Ther.* **286**, 289–298.

He, S., Gaca, M.D., Walls, A.F. (2001) The activation of synovial mast cells: modulation of histamine release by tryptase and chymase and their inhibitors, *Eur. J. Pharmacol.* **412**, 223–229.

Herrmann, M., Scholmerich, J., Straub, R.H. (2000) Stress and rheumatic diseases, *Rheum. Dis. Clin. North Am.* **26**, 737–763.

Hiromatsu, Y., Toda, S. (2003) Mast cells and angiogenesis, *Microsc. Res. Tech.* **60**, 64–69.

Huang, M., Berry, J., Kandere, K., Lytinas, M., Karalis, K., Theoharides, T.C. (2002) Mast cell deficient W/W^v mice lack stress-induced increase in serum IL-6 levels, as well as in peripheral CRH and vascular permeability, a model of rheumatoid arthritis, *Int. J. Immunopathol. Pharmacol.* **15**, 249–254.

Hughes, R., Carr, A. (2002) A randomized, double-blind, placebo-controlled trial of glucosamine sulphate as an analgesic in osteoarthritis on the knee, *Rheumatology (Oxford)* **41**, 279–284.

Jain, S., Stevenson, J.R. (1991) Enhancement by restraint stress of natural killer cell activity and splenocyte responsiveness to concanavalin A in Fischer 344 rats, *Immunol. Invest.* **20**, 365–376.

Janiszewski, J., Bienenstock, J., Blennerhassett, M.G. (1994) Picomolar doses of substance P trigger electrical responses in mast cells without degranulation, *Am. J. Physiol.* **267**, C138–C145.

Johnson, E.O., Moutsopoulos, M. (1992) Neuroimmunological axis and rheumatic diseases, *Eur. J. Clin. Invest.* **22**, S2–S5.

Kandere-Grzybowska, K., Letourneau, R., Boucher, W., Bery, J., Kempuraj, D., Poplawski, S., Athanassiou, A., Theoharides, T.C. (2003) IL-1 induces vesicular secretion of IL-6 without degranulation from human mast cells, *J. Immunol.* **171**, 4830–4836.

Kaneko, K., Kawana, S., Arai, K., Shibasaki, T. (2003) Corticotropin-releasing factor receptor type 1 is involved in the stress-induced exacerbation of chronic contact dermatitis in rats, *Exp. Dermatol.* **12**, 47–52.

Kaneko, S., Satoh, T., Chiba, J., Ju, C., Inoue, K., Kagawa, J. (2000) Interleukin-6 and interleukin-8 levels in serum and synovial fluid of patients with osteoarthritis, *Cytokines Cell. Mol. Ther.* **6**, 71–79.

Karalis, K., Louis, J.M., Bae, D., Hilderbrand, H., Majzoub, J.A. (1997) CRH and the immune system, *J. Neuroimmunol.* **72**, 131–136.

Karalis, K., Sano, H., Redwine, J., Listwak, S., Wilder, R.L., Chrousos, G.P. (1991) Autocrine or paracrine inflammatory actions of corticotropin-releasing hormone *in vivo*, *Science* **254**, 421–423.

Kempuraj, D., Huang, M., Kandere-Grzybowska, K., Basu, S., Boucher, W., Letourneau, R., Athanasiou, A., Theoharides, T.C. (2003) Azelastine inhibits secretion of IL-6, TNF-α and IL-8 as well as NF-κB activation and intracellular calcium ion levels in normal human mast cells, *Int. Arch. Allergy Immunol.* **132**, 231–239.

Kempuraj, D., Madhappan, B., Christodoulou, S., Boucher, W., Cao, J., Papadopoulou, N., Cetrulo, C.L., Theoharides, T.C. (2005) Flavonols inhibit pro-inflammatory mediator release, intracellular calcium ion levels and protein kinase C theta phosphorylation in human mast cells, *Br. J. Pharmacol.*, in press.

Kempuraj, D., Papadopoulou, N.G., Lytinas, M., Huang, M., Kandere-Grzybowska, K., Madhappan, B., Boucher, W., Christodoulou, S., Athanasiou, A., Theoharides, T.C. (2004) Corticotropin-releasing hormone and its structurally related urocortin are synthesized and secreted by human mast cells, *Endocrinology* **145**, 43–48.

Keul, R., Heinrich, P.C., Muller-Newen, G., Muller, K., Woo, P. (1998) A possible role for soluble IL-6 receptor in the pathogenesis of systemic onset juvenile chronic arthritis, *Cytokine* **10**, 729–734.

Khansari, D.N., Murgo, A.J., Faith, R.E. (1990) Effects of stress on the immune system, *Immunol. Today* **11**, 170–175.

Kiecolt-Glaser, J.K., Preacher, K.J., MacCallum, R.C., Atkinson, C., Malarkey, W.B., Glaser, R. (2003) Chronic stress and age-related increases in the proinflammatory cytokine IL-6, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9090–9095.

Kiener, H.P., Baghestanian, M., Dominkus, M., Walchshofer, S., Ghannadan, M., Willheim, M., Sillaber, C., Graninger, W.B., Smolen, J.S., Valent, P. (1998) Expression of the C5a receptor (CD88) on synovial mast cells in patients with rheumatoid arthritis, *Arthritis Rheum.* **41**, 233–245.

Kobayashi, H., Ishizuka, T., Okayama, Y. (2000) Human mast cells and basophils as sources of cytokines, *Clin. Exp. Allergy* **30**, 1205–1212.

Kohno, M., Kawahito, Y., Tsubouchi, Y., Hashiramoto, A., Yamada, R., Inoue, K.I., Kusaka, Y., Kubo, T., Elenkov, I.J., Chrousos, G.P., Kondo, M., Sano, H. (2001) Urocortin expression in synovium of patients with rheumatoid arthritis and osteoarthritis: relation to inflammatory activity, *J. Clin. Endocrinol. Metab.* **86**, 4344–4352.

Kops, S.K., Theoharides, T.C., Cronin, C.T., Kashgarian, M.G., Askenase, P.W. (1990) Ultrastructural characteristics of rat peritoneal mast cells undergoing differential release of serotonin without histamine and without degranulation, *Cell Tissue Res.* **262**, 415–424.

Kops, S.K., Van Loveren, H., Rosenstein, R.W., Ptak, W., Askenase, P.W. (1984) Mast cell activation and vascular alterations in immediate hypersensitivity-like reactions induced by a T cell derived antigen-binding factor, *Lab. Invest.* **50**, 421–434.

Krüger-Krasagakes, S., Möller, A.M., Kolde, G., Lippert, U., Weber, M., Henz, B.M. (1996) Production of inteuleukin-6 by human mast cells and basophilic cells, *J. Invest. Dermatol.* **106**, 75–79.

Kubler, A., Rothacher, G., Knappertz, V.A., Kramer, G., Nink, M., Beyer, J., Lehnert, H. (1994) Intra and extracerebral blood flow changes and flushing after intravenous injection of human corticotropin-releasing hormone, *Clin. Invest.* **72**, 331–336.

Leal-Berumen, I., Conlon, P., Marshall, J.S. (1994) IL-6 production by rat peritoneal mast cells is not necessarily preceded by histamine release and can be induced by bacterial lipopolysaccharide, *J. Immunol.* **152**, 5468–5476.

Lee, D.M., Friend, D.S., Gurish, M.F., Benoist, C., Mathis, D., Brenner, M.B. (2002) Mast cells: a cellular link between autoantibodies and inflammatory arthritis, *Science* **297**, 1689–1692.

Letourneau, R., Pang, X., Sant, G.R., Theoharides, T.C. (1996) Intragranular activation of bladder mast cells and their association with nerve processes in interstitial cystitis, *Br. J. Urol.* **77**, 41–54.

Letourneau, R., Rozniecki, J.J., Dimitriadou, V., Theoharides, T.C. (2003) Ultrastructural evidence of brain mast cell activation without degranulation in monkey experimental allergic encephalomyelitis, *J. Neuroimmunol.* **145**, 18–26.

Leu, S.J., Singh, V.K. (1992) Stimulation of interleukin-6 production by corticotropin-releasing factor, *Cell. Immunol.* **143**, 220–227.

Levi-Schaffer, F., Shalit, M. (1989) Differential release of histamine and prostaglandin D2 in rat peritoneal mast cells activated with peptides, *Int. Arch. Allergy Appl. Immunol.* **90**, 352–357.

Levi-Schaffer, F., Austen, K.F., Gravallese, P.M., Stevens, R.L. (1986) Co-culture of interleukin 3-dependent mouse mast cells with fibroblasts results in a phenotypic change of the mast cells, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6485–6488.

Lewis, K., Li, C., Perrin, M.H., Blount, A., Kunitake, K., Donaldson, C., Vaughan, J., Reyes, T.M., Gulyas, J., Fischer, W., Bilezikjian, L., Rivier, J., Sawchenko, P.E., Vale, W.W. (2001) Identification of urocortin III, an additional member of the corticotropin-releasing factor (CRF) family with high affinity for the CRF 2 receptor, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7570–7575.

Lindahl, U., Hook, M. (1978) Glycosaminoglycans and their binding to biological macromolecules, *Annu. Rev. Biochem.* **47**, 385–417.

Lovenberg, T.W., Chalmers, D.T., Liu, C., DeSouza, E.B. (1995a) CRF2α and CRF2β receptor mRNAs are differentially distributed between the rat central nervous system and peripheral tissues, *Endocrinology* **136**, 4139–4142.

Lovenberg, T.W., Liaw, C.W., Grigoriadis, D.E., Clevenger, W., Charmers, D.T., DeSouza, E.B., Oltersdorf, T. (1995b) Cloning and characterization of a functionally distinct corticotropin-releasing factor receptor subtype from rat brain, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 836–840.

Lowry, P.J., Woods, R.J., Baigent, S. (1996) Corticotropin releasing factor and its binding protein, *Pharmacol. Biochem. Behav.* **54**, 305–308.

Lutgendorf, S.K., Russell, D., Ullrich, P., Harris, T.B., Wallace, R. (2004) Religious participation, interleukin-6, and mortality in older adults, *Health Psychol.* **23**, 465–475.

Lytinas, M., Kempuraj, D., Huang, M., Boucher, W., Esposito, P., Theoharides, T.C. (2003) Acute stress results in skin corticotropin-releasing hormone secretion, mast cell activation and vascular permeability, an effect mimicked by intradermal corticotropin-releasing hormone and inhibited by histamine-1 receptor antagonists, *Int. Arch. Allergy Immunol.* **130**, 224–231.

Malaviya, R., Ross, E., Jakschik, B.A., Abraham, S.N. (1994) Mast cell degranulation induced by type 1 fimbriated *Escherichia coli* in mice, *J. Clin. Invest.* **93**, 1645–1653.

Malone, E.D. (2002) Managing chronic arthritis, *Vet. Clin. North Am. Equine Pract.* **18**, 411–437.

Manek, N.J. (2001) Medical management of osteoarthritis, *Mayo Clin. Proc.* **76**, 533–539.

Mastorakos, G., Chrousos, G.P., Weber, J.S. (1993) Recombinant interleukin-6 activates the hypothalamic-pituitary-adrenal axis in humans, *J. Clin. Endocrinol. Metab.* **77**, 1690–1694.

Matsuda, H., Kawakita, K., Kiso, Y., Nakano, T., Kitamura, Y. (1989) Substance P induces granulocyte infiltration through degranulation of mast cells, *J. Immunol.* **142**, 927–931.

Matsuda, H., Kannan, Y., Ushio, H., Kiso, Y., Kanemoto, T., Suzuki, H., Kitamura, Y. (1991) Nerve growth factor induces development of connective tissue-type mast cells *in vitro* from murine bone marrow cells, *J. Exp. Med.* **174**, 7–14.

Matsumoto, I., Inoue, Y., Shimada, T., Aikawa, T. (2001) Brain mast cells act as an immune gate to the hypothalamic-pituitary-adrenal axis in dogs, *J. Exp. Med.* **194**, 71–78.

Matsushima, H., Yamada, N., Matsue, H., Shimada, S. (2004) The effects of endothelin-1 on degranulation, cytokine, and growth factor production by skin-derived mast cells, *Eur. J. Immunol.* **34**, 1910–1019.

Mattheos, S., Christodoulou, S., Kempuraj, D., Kempuraj, B., Karalis, K., Theoharides, T.C. (2003) Mast cells and corticotropin-releasing hormone (CRH) are required for experimental inflammatory arthritis, *FASEB J.* **17**, C44.

Maurer, M., Wedemeyer, J., Metz, M., Piliponsky, A.M., Weller, K., Chatterjea, D., Clouthier, D.E., Yanagisawa, M.M., Tsai, M., Galli, S.J. (2004) Mast cells promote homeostasis by limiting endothelin-1-induced toxicity, *Nature* **432**, 512–516.

McAlindon, T. (2003) Why are clinical trials of glucosamine no longer uniformly positive? *Rheum. Dis. Clin. North Am.* **29**, 789–801.

McAlindon, T.E., LaValley, M.P., Gulin, J.P., Felson, D.T. (2000) Glucosamine and chondroitin for treatment of osteoarthritis, *JAMA* **283**, 1469–1475.

McCurdy, J.D., Olynych, T.J., Maher, L.H., Marshall, J.S. (2003) Cutting edge: distinct Toll-like receptor 2 activators selectively induce different classes of mediator production from human mast cells, *J. Immunol.* **170**, 1625–1629.

McEvoy, A.N., Bresnihan, B., FitzGerald, O., Murphy, E.P. (2001) Corticotropin-releasing hormone signaling in synovial tissue from patients with early inflammatory arthritis is mediated by the type 1α corticotropin-releasing hormone receptor, *Arthritis Rheum.* **44**, 1761–1767.

McLachlan, J.B., Hart, J.P., Pizzo, S.V., Shelburne, C.P., Staats, H.F., Gunn, M.D., Abraham, S.N. (2003) Mast cell-derived tumor necrosis factor induces hypertrophy of draining lymph nodes during infection, *Nat. Immunol.* **4**, 1199–1205.

Merchenthaler, I., Hynes, M.A., Vingh, S., Schally, A.V., Petrusz, P. (1983) Immunocytochemical localization of corticotropin-releasing factor (CRF) in the rat spinal cord, *Brain Res.* **275**, 373–377.

Middleton, E., Kandaswami, C., Theoharides, T.C. Jr. (2000) The effects of plant flavonoids on mammalian cells:implications for inflammation, heart disease and cancer, *Pharmacol. Rev.* **52**, 673–751.

Mihara, M., Kotoh, M., Nishimoto, N., Oda, Y., Kumagai, E., Takagi, N., Tsunemi, K., Ohsugi, Y., Kishimoto, T., Yoshizaki, K., Takeda, Y. (2001) Humanized antibody to human interleukin-6 receptor inhibits the development of collagen arthritis in cynomolgus monkeys, *Clin. Immunol.* **98**, 319–326.

Million, M., Maillot, C., Saunders, P., Rivier, J., Vale, W., Taché, Y. (2002) Human urocortin II, a new CRF related peptide, displays selective CRF2-mediated action on gastric transit in

rats, *Am. J. Physiol. Gastrointest. Liver Physiol.* **282**, G34–G40.

Muller-Fassbender, H., Bach, G.L., Haase, W., Rovati, L.C., Setnikar, I. (1994) Glucosamine sulfate compared to ibuprofen in osteoarthritis of the knee, *Osteoarthritis Cartil.* **2**, 61–69.

Navarra, P., Tsagarakis, S., Faria, M.S., Rees, L.H., Besser, G.M., Grossman, A.B. (1991) Interleukins-1 and -6 stimulate the release of corticotropin-releasing hormone-41 from rat hypothalamus *in vitro* via the eicosanoid cyclooxygenase pathway, *Endocrinology* **128**, 37–44.

Nigrovic, P.A., Lee, D.M. (2005) Mast cells in inflammatory arthritis, *Arthritis Res. Ther.* **7**, 1–11.

Nishioka, T., Kurokawa, H., Takao, R., Kumon, Y., Nishiya, K., Hashimoto, K. (1996) Differential changes of corticotropin releasing hormone (CRH) concentrations in plasma and synovial fluids of patients with rheumatoid arthritis (RA), *Endocr. J.* **43**, 241–247.

Ohshima, S., Saeki, Y., Mima, T., Sasai, M., Nishioka, K., Nomura, S., Kopf, M., Katada, Y., Tanaka, T., Suemura, M., Kishimoto, T. (1998) Interleukin 6 plays a key role in the development of antigen-induced arthritis, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8222–8226.

Olsson, N., Ulfgre, A.K., Nilsson, G. (2001) Demonstration of mast cell chemotactic activity in synovial fluid from rheumatoid patients, *Ann. Rheum. Dis.* **60**, 187–193.

Ostrakhovitch, E.A., Afanas'ev, I.B. (2001) Oxidative stress in rheumatoid arthritis leukocytes: suppression by rutin and other antioxidants and chelators, *Biochem. Pharmacol.* **62**, 743–746.

Palmieri, L., Conte, A., Giovannini, L., Lualdi, P., Ronca, G. (1990) Metabolic fate of exogenous chondroitin sulfate in the experimental animal, *Arzneimittelforschung* **40**, 319–323.

Pavelka, K., Gatterova, J., Olejarova, M., Machacek, S., Giacovelli, G., Rovati, L.C. (2002) Glucosamine sulfate use and delay of progression of knee osteoarthritis: a 3-year, randomized, placebo-controlled, double-blind study, *Arch. Intern. Med.* **162**, 2113–2123.

Pendleton, A., Arden, N., Dougados, M., Doherty, M., Bannwarth, B., Bijlsma, J.W., Cluzeau, F., Cooper, C., Dieppe, P.A., Gunther, K.P., Hauselmann, H.J., Herrero-Beaumont, G., Kaklamanis, P.M., Lequesne, M., Lohmander, S., Mazieres, B., Mola, E.M., Pavelka, K., Serni, U., Swoboda, B., Verbruggen, A.A., Weseloh, G., Zimmermann-Gorska, I. (2000) EULAR recommendations for the management of knee osteoarthritis: report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT), *Ann. Rheum. Dis.* **59**, 936–944.

Pignatti, P., Vivarelli, M., Meazza, C., Rizzolo, M.G., Martini, A., De Benedetti, F. (2001) Abnormal regulation of interleukin 6 in systemic juvenile idiopathic arthritis, *J. Rheumatol.* **28**, 1670–1676.

Rabkin, J.G., Struening, E.L. (1976) Life events, stress and illness, *Science* **194**, 1013–1020.

Radulovic, M., Dautzenberg, F.M., Sydow, S., Radulovic, J., Spiess, J. (1999) Corticotropin-releasing factor receptor 1 in mouse spleen: expression after immune stimulation and identification of receptor-bearing cells, *J. Immunol.* **162**, 3013–3021.

Redegeld, F.A., Nijkamp, F.P. (2005) Immunoglobulin free light chains and mast cells: pivotal role in T-cell-mediated immune reactions? *Trends Immunol.* **24**, 181–185.

Reginster, J.Y., Deroisy, R., Rovati, L.C., Lee, R.L., Lejeune, E., Bruyere, O., Giacovelli, G., Henrotin, Y., Dacre, J.E., Gossett, C. (2001) Long-term effects of glucosamine sulphate on osteoarthritis progression: a randomized, placebo-controlled clinical trial, *Lancet* **357**, 251–256.

Rinner, I., Schauenstein, K., Mangge, H., Porta, S., Kvetnansky, R. (1992) Opposite effects of mild and severe stress on in vitro activation of rat peripheral blood lymphocytes, *Brain. Behav. Immunol.* **6**, 130–140.

Rodewald, H.-R., Dessing, M., Dvorak, A.M., Galli, S.J. (1996) Identification of a committed precursor for the mast cell lineage, *Science* **271**, 818–822.

Rosenwasser, L.J., Boyce, J.A. (2003) Mast cells: beyond IgE, *J. Allergy Clin. Immunol.* **111**, 24–32.

Rostand, K.S., Esko, J.D. (1997) Minireview: microbial adherence to and invasion through proteoglycans, *Infect. Immunol.* **65**, 1–8.

Roupe van der Voort, C., Heijnen, C.J., Wulffraat, N., Kuis, W., Kavelaars, A. (2000) Stress induces increases in IL-6 production by leucocytes of patients with the chronic inflammatory disease juvenile rheumatoid arthritis: a putative role for alpha(1)-adrenergic receptors, *J. Neuroimmunol.* **110**, 223–229.

Roznieckі, J.J., Dimitriadou, V., Lambracht-Hall, M., Pang, X., Theoharides, T.C. (1999) Morphological and functional demonstration of rat dura mast cell-neuron interactions *in vitro* and *in vivo*, *Brain Res.* **849**, 1–15.

Ruschpler, P., Lorenz, P., Eichler, W., Koczan, D., Hanel, C., Scholz, R., Melzer, C., Thiesen, H.J., Stiehl, P. (2005) High CXCR3 expression in synovial mast cells associated with CXCL9 and CXCL10 expression in inflammatory synovial tissues of patients with rheumatoid arthritis, *Arthritis Res. Ther.* **5**, R241–R252.

Sakai, K., Clemmons, D.R. (2005) Glucosamine induces resistance to insulin-like growth factor I (IGF-I) and insulin in Hep G2 cell cultures: biological significance of IGF-I/insulin hybrid receptors, *Endocrinology* **144**, 2388–2395.

Schmidlin, F., Bunnett, N.W. (2001) Protease-activated receptors: how proteases signal to cells, *Curr. Opin. Pharmacol.* **1**, 575–582.

Schwartz, L.B. (1987) Mediators of human mast cells and human mast cell subsets, *Ann. Allergy* **58**, 226–235.

Serafin, W.E., Austen, K.F. (1987) Mediators of immediate hypersensitivity reactions, *N. Engl. J. Med.* **317**, 30–34.

Shanas, U., Bhasin, R., Sutherland, A.K., Silverman, A.-J., Silver, R. (1998) Brain mast cells lack the c-kit receptor: immunocytochemical evidence, *J. Neuroimmunol.* **90**, 207–211.

Singh, L.K., Boucher, W., Pang, X., Letourneau, R., Seretakis, D., Green, M., Theoharides, T.C. (1999) Potent mast cell degranulation and vascular permeability triggered by urocortin through activation of CRH receptors, *J. Pharmacol. Exp. Ther.* **288**, 1349–1356.

Singh, V.K. (1989) Stimulatory effect of corticotropin-releasing neurohormone on human lymphocyte proliferation and interleukin-2 receptor expression, *J. Neuroimmunol.* **23**, 256–262.

Singh, V.K., Leu, C.S.J. (1990) Enhancing effect of corticotropin-releasing neurohormone on the production of interleukin-1 and interleukin-2, *Neurosci. Lett.* **120**, 151–154.

Skofitsch, G., Savitt, J.M., Jacobowitz, D.M. (1985a) Suggestive evidence for a functional unit between mast cells and substance P fibers in the rat diaphragm and mesentery, *Histochemistry* **82**, 5–8.

Skofitsch, G., Zamir, N., Helke, C.J., Savitt, J.M., Jacobowitz, D.M. (1985b) Corticotropin-releasing factor-like immunoreactivity in sensory ganglia and capsaicin sensitive neurons of the rat central nervous system: colocalization with other neuropeptides, *Peptides* **6**, 307–318.

Slominski, A., Pisarchik, A., Tobin, D.J., Mazurkiewicz, J., Wortsman, J. (2004) Differential expression of a cutaneous corticotropin-releasing hormone system, *Endocrinology* **145**, 941–950.

Slominski, A., Wortsman, J., Luger, T., Paus, R., Solomon, S. (2000) Corticotropin releasing hormone and proopiomelanocortin involvement in the cutaneous response to stress, *Physiol. Rev.* **80**, 979–1020.

Spinedi, E., Hadid, R., Daneva, T., Gaillard, R.C. (1992) Cytokines stimulate the CRH but not the vasopressin neuronal system: evidence for a median eminence site of interleukin-6 action, *Neuroendocrinology* **56**, 46–53.

Stead, R.H., Bienenstock, J. (1990) Cellular Interactions Between the Immune and Peripheral Nervous System. A Normal Role for Mast Cells? in: Burger, M.M., Sordat, B., Zinkernagel, R.M. (Eds.) *Cell to Cell Interaction*, Karger, Basel, pp. 170–187.

Stead, R.H., Tomioka, M., Quinonez, G., Simon, G.T., Felten, S.Y., Bienenstock, J. (1987) Intestinal mucosal mast cells in normal and nematode-infected rat intestines are in intimate contact with peptidergic nerves, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2975–2979.

Strom, B.L., Schinzel, R., Apter, A.J., Margolis, D.J., Lautenbach, E., Hennessy, S., Bilker, W.B., Pettitt, D. (2003) Absence of cross-reactivity between sulfonamide antibiotics and sulfonamide nonantibiotics, *N. Engl. J. Med.* **349**, 1628–1635.

Suda, T., Tomori, N., Tozawa, F., Mouri, T., Demura, H., Shizume, K. (1984) Distribution and characterization of immunoreactive corticotropin-releasing factor in human tissues, *J. Clin. Endocrinol. Metab.* **59**, 861–866.

Takagi, M., Nakahata, T., Kubo, T., Shiohara, M., Koike, K., Miyajima, A., Arai, K.-I., Nishikawa, S.-I., Zsebo, K.M., Komiyama, A. (1992) Stimulation of mouse connective tissue-type mast cells by hemopoietic stem cell factor, a ligand for the *c-kit* receptor, *J. Immunol.* **148**, 3446–3453.

Tal, M., Liberman, R. (1997) Local injection of nerve growth factor (NGF) triggers degranulation of mast cells in rat paw, *Neurosci. Lett.* **221**, 129–132.

Taylor, P.C. (2003) Anti-TNFalpha therapy for rheumatoid arthritis: an update, *Intern. Med.* **42**, 15–20.

Taylor, A.M., Galli, S.J., Coleman, J.W. (1995) Stem-cell factor, the kit ligand, induces direct degranulation of rat peritoneal mast cells in vitro and in vivo: dependence of the in vitro effect on period of culture and comparisons of stem- cell factor with other mast cell-activating agents, *Immunology* **86**, 427–433.

Tetlow, L.C., Woolley, D.E. (1995b) Distribution, activation and tryptase/chymase phenotype of mast cells in the rheumatoid lesion, *Ann. Rheum. Dis.* **54**, 549–555.

Tetlow, L.C., Woolley, D.E. (1995a) Mast cells, cytokines, and metalloproteinases at the rheumatoid lesion: dual immunolocalization studies, *Ann. Rheum. Dis.* **54**, 896–903.

Theoharides, T.C. (1996) Mast cell: a neuroimmunoendocrine master player, *Int. J. Tissue React.* **18**, 1–21.

Theoharides, T.C. (2002) Mast cells and stress – a psychoneuroimmunological perspective, *J. Clin. Psychopharmacol.* **22**, 103–108.

Theoharides, T.C. (2003) Dietary supplements for arthritis and other inflammatory conditions: key role of mast cells and benefit of combining anti-inflammatory and proteoglycan products, *Eur. J. Inflamm.* **1**, 1–8.

Theoharides, T.C., Bielory, L. (2004) Mast cells and mast cell mediators as targets of dietary supplements, *Ann. Allergy Asthma Immunol.* **93**, S24–S34.

Theoharides, T.C., Cochrane, D.E. (2004) Critical role of mast cells in inflammatory diseases and the effect of acute stress, *J. Neuroimmunol.* **146**, 1–12.

Theoharides, T.C., Douglas, W.W. (1978a) Secretion in mast cells induced by calcium entrapped within phospholipid vesicles, *Science* **201**, 1143–1145.

Theoharides, T.C., Douglas, W.W. (1978b) Somatostatin induces histamine secretion from rat peritoneal mast cells, *Endocrinology* **102**, 1637–1640.

Theoharides, T.C., Sant, G.R. (2003) Neuroimmune Connections and Regulation of Function in the Urinary Bladder, in: Bienenstock, J., Goetzl, E., Blennerhassett, M. (Eds.) *Autonomic Neuroimmunology*, Hardwood Academic Publishers, Lausanne, Switzerland, pp. 345–369.

Theoharides, T.C., Alexandrakis, M., Kempuraj, D., Lytinas, M. (2001) Anti-inflammatory actions of flavonoids and structural requirements for new design, *Int. J. Immunopathol. Pharmacol.* **14**, 119–127.

Theoharides, T.C., Bondy, P.K., Tsakalos, N.D., Askenase, P.W. (1982) Differential release of serotonin and histamine from mast cells, *Nature* **297**, 229–231.

Theoharides, T.C., Sieghart, W., Greengard, P., Douglas, W.W. (1980) Antiallergic drug cromolyn may inhibit histamine secretion by regulating phosphorylation of a mast cell protein, *Science* **207**, 80–82.

Theoharides, T.C., Donelan, J.M., Papadopoulou, N., Cao, J., Kempuraj, D., Conti, P. (2004) Mast cells as targets of corticotropin-releasing factor and related peptides, *Trends Pharmacol. Sci.* **25**, 563–568.

Theoharides, T.C., Sant, G.R., El-Mansoury, M., Letourneau, R.J., Ucci, A.A., Jr., Meares, E.M., Jr. (1995a) Activation of bladder mast cells in interstitial cystitis: a light and electron microscopic study, *J. Urol.*, **153**, 629–636.

Theoharides, T.C., Singh, L.K., Boucher, W., Pang, X., Letourneau, R., Webster, E., Chrousos, G. (1998) Corticotropin-releasing hormone induces skin mast cell degranulation and increased vascular permeability, a possible explanation for its pro-inflammatory effects, *Endocrinology* **139**, 403–413.

Theoharides, T.C., Patra, P., Boucher, W., Letourneau, R., Kempuraj, D., Chiang, G., Jeudy, S., Hesse, L., Athanasiou, A. (2000) Chondroitin sulfate inhibits connective tissue mast cells, *Br. J. Pharmacol.* **131**, 1039–1049.

Theoharides, T.C., Spanos, C.P., Pang, X., Alferes, L., Ligris, K., Letourneau, R., Rozniecki, J.J., Webster, E., Chrousos, G. (1995b) Stress-induced intracranial mast cell degranulation. A corticotropin releasing hormone-mediated effect, *Endocrinology* **136**, 5745–5750.

Thomason, B.T., Brantley, P.J., Jones, G.N., Dyer, H.R., Morris, J.L. (1992) The relation between stress and disease activity in rheumatoid arthritis, *J. Behav. Med.* **15**, 215–220.

Topol, E.J. (2005) Failing the public health – rofecoxib, Merck, and the FDA, *N. Engl. J. Med.* **351**, 1707–1709.

Tsai, M., Takeishi, T., Thompson, H., Langley, K.E., Zsebo, K.M., Metcalfe, D.D., Geissler, E.N., Galli, S.J. (1991) Induction of mast cell proliferation, maturation and heparin synthesis by the rat c-*kit* ligand, stem cell factor, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6382–6386.

Tsakalos, N.D., Theoharides, T.C., Kops, S.K., Askenase, P.W. (1983) Induction of mast cell secretion by parathormone, *Biochem. Pharmacol.* **32**, 355–360.

Uzuki, M., Sasano, H., Muramatsu, Y., Totsune, K., Takahashi, K., Oki, Y., Iino, K., Sawai, T. (2001) Urocortin in the synovial tissue of patients with rheumatoid arthritis, *Clin. Sci.* **100**, 577–589.

Vaughan, J., Donaldson, C., Bittencourt, J., Perrin, M.H., Lewis, K., Sutton, S., Chan, R., Turnbull, A.V., Lovejoy, D., Rivier, C., Rivier, J., Sawchenko, P.E., Vale, W. (1995) Urocortin, a mammalian neuropeptide related to fish urotensin I and to corticotropin-releasing factor, *Nature* **378**, 287–292.

Venihaki, M., Dikkes, P., Carrigan, A., Karalis, K.P. (2001) Corticotropin-releasing hormone regulates IL-6 expression during inflammation, *J. Clin. Invest.* **108**, 1159–1166.

Walker-Bone, K., Javaid, K., Arden, N., Cooper, C. (2000) Regular review: medical management of osteoarthritis, *BMJ* **321**, 936–940.

Wasserman, S.I. (1984) The mast cell and synovial inflammation, *Arthritis Rheum.* **27**, 841–844.

Webster, E.L., Barrientos, R.M., Contoreggi, C., Isaac, M.G., Ligier, S., Gabry, K.E., Chrousos, G.P., McCarthy, E.F., Rice, K.C., Gold, P.W., Sternberg, E.M. (2002) Corticotropin releasing hormone (CRH) antagonist attenuates adjuvant induced arthritis: role of CRH in peripheral inflammation, *J. Rheumatol.* **29**, 1252–1261.

Weisman, M.H., Paulus, H.D., Russak, S.M., Lubeck, D.P., Chiou, C.F., Sengupta, N., Ofman, J.J., Borenstein, J., Moadel, A.B., Sherbourne, C.D. (2003) Development of a new instrument for rheumatoid arthritis: the Cedars-Sinai Health-Related Quality of Life instrument (CSHQ-RA), *Arthritis Rheum.* **49**, 78–84.

Wipke, B.T., Wang, Z., Nagengast, W., Reichert, D.E., Allen, P.M. (2005) Staging the initiation of autoantibody-induced arthritis: a critical role for immune complexes, *J. Immunol.* **172**, 7694–7702.

Wood, P.G., Karol, M.H., Kusnecov, A.W., Rabin, B.S. (1993) Enhancement of antigen-specific humoral and cell-mediated immunity by electric footshock stress in rats, *Brain Behav. Immun.* **7**, 121–134.

Woolley, D.E. (2003) The mast cell in inflammatory arthritis, *N. Engl. J. Med.* **348**, 1709–1711.

Woolley, D.E., Tetlow, L.C. (2000) Mast cell activation and its relation to proinflammatory cytokine production in the rheumatoid lesion, *Arthritis Res.* **2**, 65–74.

Yu, J.G., Boies, S.M., Olefsky, J.M. (2005) The effect of oral glucosamine sulfate on insulin sensitivity in human subjects, *Diabetes Care* **26**, 1941–1942.

Synthesis of Peptide Mimetics and their Building Blocks

Bruce K. Cassels[1] *and Patricio Sáez*[2]
[1] *Department of Chemistry, Faculty of Sciences, University of Chile, and*
[2] *Faculty of Medical Sciences, University of Santiago, Chile*

1 **Introduction 67**

2 **Azapeptides 68**

3 **β-, γ-, δ-, ε-, and Related Peptides 71**

4 **Aminooxy Peptides 72**

5 **Oligocarbamate and Oligourea Peptidomimetics 74**

6 **Peptoids 75**

7 **Variations on the Peptoid Theme 77**

8 **Proline Substitutes 80**

9 **Biological Activity and Therapeutic Hints 80**

Bibliography 86
Books and Reviews 86
Primary Literature 87

Keywords

Azapeptides
Peptide mimetics in which one or more amino acid residues is (or are) replaced by azaamino acid(s) (i.e. amino acid(s) where the carbon atom bearing the amino group is replaced by nitrogen).

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

Aminooxy Peptides
Analogs of β-, γ- or δ-peptides in which the β-, γ- or δ-carbon is replaced by an oxygen atom.

Oligocarbamate
A compound containing repeating O—CO—NH units separated by two carbon atoms, one of which may bear a side chain.

Oligourea
A compound containing repeating NH—CO—NH units separated by two carbon atoms, one of which may bear a side chain.

***β*-,*γ*-, *δ*-, *ε*-Peptides**
Peptide mimetics based on β-, γ-, δ- or ε-amino acids.

Peptide Mimetics (or Peptidomimetics)
Compounds imitating the structure of bioactive peptides, preserving overall conformational features and retaining some of the chemical groups that interact with biological targets, but incorporating nonamino acid units.

Peptoids
Peptide mimetics in which the amino acid side chains are shifted to the amide nitrogen atoms.

Very many peptides have the potential of being used as drugs. Nevertheless, they usually suffer from serious limitations stemming from their metabolic lability and unfavorable pharmacokinetics. A promising approach to circumvent these problems is the use of conformationally similar molecules, able to interact with the same biological targets, but less susceptible to metabolic degradation and with better absorption and distribution properties. This may be achieved through the synthesis of peptide mimetics where some or all of the amino acid residues have been replaced by different chemical moieties. This article gives a brief overview of peptide mimetic structures on which medicinal chemists have focused in the last few years, together with the chemical approaches used for their synthesis and of the amino acid replacements that serve as their building blocks. These methods range from classical organic synthetic methods to quite sophisticated catalytic and solid-phase protocols.

In many cases, the target compounds have been tested in biological systems with varying degrees of success. These results should not be construed as an indication of unreliability of the peptidomimetic approach, but rather as a reflection of the limited sample examined to date. Because of the solid foundations on which this therapeutic strategy is based, we think that a growing number of peptide mimetics are likely to enter the clinic in the coming years.

1 Introduction

The large number of endogenous peptides that act as cytokines, hormones, or neurotransmitters, in the regulation of neural function and blood pressure, and in inflammation, continues to increase. Not only are primary gene products important, but in many cases their breakdown also affords smaller peptides with increased potency or with qualitatively different properties. Nevertheless, the development of clinical therapies based on endogenous peptides has been rather intuitive and not actually realized because of pharmacokinetic as well as pharmacodynamic issues, which seem to be difficult to solve properly. In order to overcome these drawbacks, the discovery by screening, the design, synthesis, and pharmacological testing of nonpeptide ligands of peptide receptors – either as agonists or antagonists – has become a very active field, particularly during the last decade.

Peptide mimics, peptide mimetics, or peptidomimetics may be defined as compounds that can substitute for peptides in their interactions with receptors or enzymes. Early examples such as morphine – a mimic of opioid peptides – did not result from any awareness of their mechanism of action. More recently, screening efforts have led to "hits" such as the natural, moderately potent cholecystokinin receptor antagonist asperlicin, described by Chang et al. in 1985, which has been modified to afford a plethora of synthetic analogs based on either the 1,4-benzodiazepine or the quinazolidinone moieties, reviewed by Herranz in 2003, in both of which series peptide analog portions may be discerned.

The design, synthesis, and evaluation of peptide mimics dates back at least to the late 1960s, but the vast majority of papers in the field are less than 20 years old, and the number of publications in the last decade is close to three-quarters of the total. The subject of peptide mimics can be divided into three major areas: modifications of the amino acid side chains, modifications of the peptide backbone, and dipeptide mimics with constrained conformations. The first aspect has been addressed by Williams in the previous edition of this encyclopedia (1996), and will not be considered here. Isosteric replacements of the peptide bond (CONH) by groups such as thioamido (CSNH), ester (COO), ketomethylene ($COCH_2$), methyleneoxy-, methylenethio-, or methyleneamino (CH_2O, CH_2S, CH_2NH), retro-amido (NHCO), or trans C=C double bonds (CH=CH, CF=CH) were reviewed by Goodman and Ro in 1995 in a chapter of Burger's Medicinal Chemistry that is more concerned with conformationally constrained dipeptide mimics. These are a newer development that relies heavily on a firm understanding of protein conformation, which has been reviewed by Hruby and Balse in 2000. Another review volume is that of Abell in 1999. An extensive, five-volume treatise on the subject of peptide and peptidomimetic synthesis was published by Goodman et al. in 2002, and again, as a "workbench edition," in 2004. We will consequently concentrate on advances made in the last few years.

The synthesis of these compounds relies in part on classical peptide chemistry, but almost always incorporating specific modifications to accommodate the different building blocks required. An important general methodology for the construction of peptoids, for example, is the approach of Zuckermann, described below, which continues to be used in most cases.

2
Azapeptides

Azapeptides are peptide mimics in which one or more backbone α-carbon atoms have been replaced by nitrogen. From the structural standpoint, this means introducing one or more urea moieties in the peptide chain. Occasional syntheses of monoazapeptides have been described at least for the last 40 years and were reviewed as far back as 1970. However, this type of peptide mimic was not included in the Goodman and Ro chapter published in 1995, although it has been reviewed in 2002 by Zega and Urleb, so we will discuss them in some detail.

Azapeptide synthesis has generally been achieved by a combination of hydrazine chemistry and peptide chemistry. For example, the reaction of a 1-alkyl-2-Boc-protected hydrazine or an unprotected monoalkylhydrazine with a carboxylic acid ester α-isocyanate affords an N-protected or unprotected monoazadipeptide derivative (Scheme 1). Even in unprotected monoalkylhydrazines, reaction at thesubstituted nitrogen atom is favored. If deprotected, the resulting semicarbazide can be coupled to the carboxyl group of an N-protected amino acid or peptide using, for example, conventional dicyclohexylcarbodiimide (DCC) coupling, to provide the elongated product (Scheme 2).

A stepwise example of this, requiring purification at each step, but which resulted in potent, selective cysteine protease inhibitors, is that of Wieczerzak et al. in 2002. In this case, peptides were prepared in solution and then an azaglycine residue was appended to the free amino group by coupling *t*-butyl carbazate (Boc-hydrazine) and an amino acid or peptide ester or amide with carbonyldiimidazole, where the key reaction is shown in Scheme 3. Although experimental details are lacking, the Boc-azapeptide could presumably be deprotected and coupled to a Boc-protected amino acid or peptide using conventional methodology (the use of 2-(1-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate – TBTU – or DCC, and 1-hydroxybenzotriazole – HOBt – are mentioned).

The synthesis of a series of Boc-protected alkylhydrazines was described by Dutta and Morley in 1975, and still constitutes a generally useful approach. This is

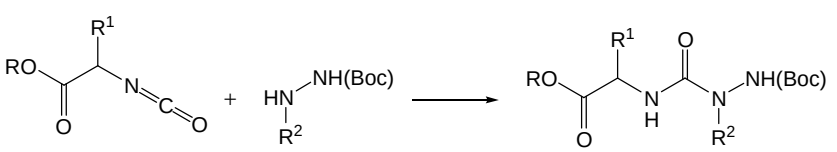

Scheme 1

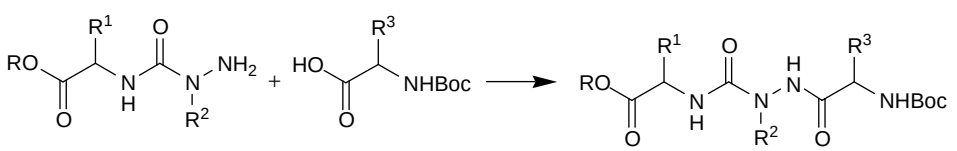

Scheme 2

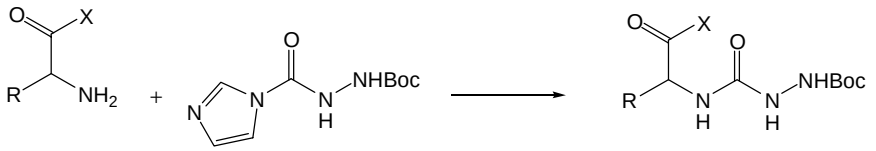

Scheme 3

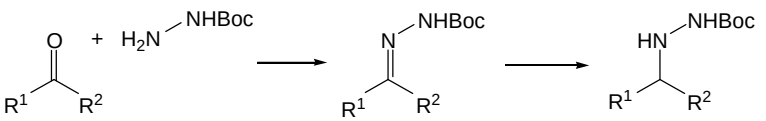

Scheme 4

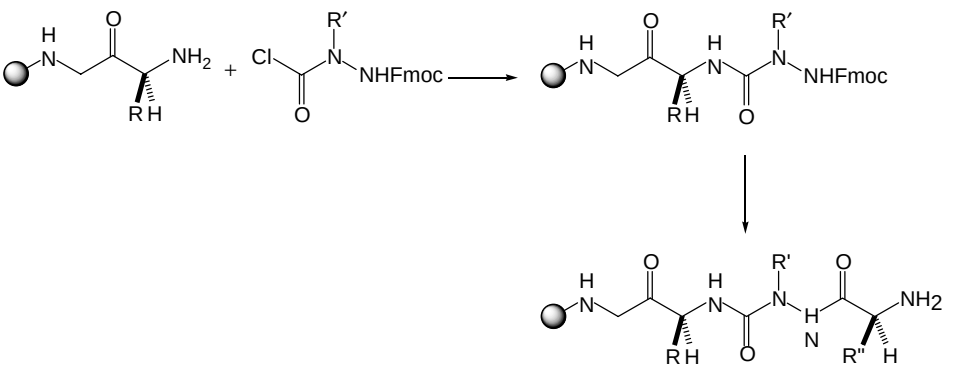

Scheme 5

done by preparing the Boc-hydrazides of suitable aldehydes or ketones and then reducing them by catalytic hydrogenation (Scheme 4).

In 2001, Hart and Beeson in a newer example of compounds introduced important modifications allowing the use of solid-phase Fmoc methodology on Rink amide resin (Scheme 5).

A large series of compounds of this class, based on the sequence of the antigenic hen ovalbumin pentadecapeptide Gln-Ala-Val-His-Ala-Ala-His-Ala-Glu-Ile-Asn-Glu-Ala-Gly-Arg by replacing one amino acid at a time by azaglycine, azaalanine, or azaglutamic acid, bound detectably to the major histocompatibility complex. Some of them, for example, Gln-Ala-Val-His-Ala-Ala-His-Ala-Glu-Ile-Gly[a]-Glu-Ala-Gly-Arg (where the superscript[a] indicates an azaamino acid), were as potent as activators of T cells as the native peptide.

The azaalanine monomer was built starting from methylhydrazine (Scheme 6). Synthesis of the azaglutamic acid started from *tert*-butyl acrylate (Scheme 7).

A recent patent describes an apparently general method for the preparation of azapeptides or azatides that, like Hart and Beeson's procedure in 2001 but unlike earlier methods, is amenable in theory to automated synthesis by modification of existing protocols. An important advantage

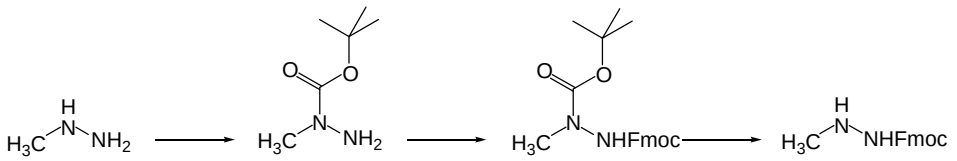

Scheme 6

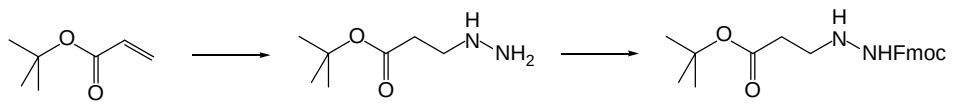

Scheme 7

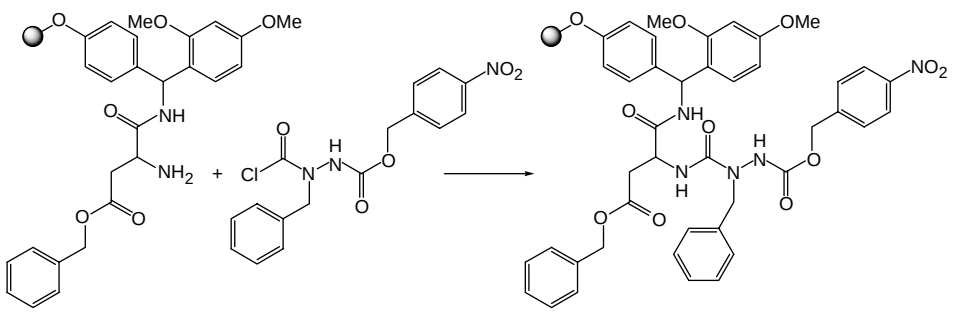

Scheme 8

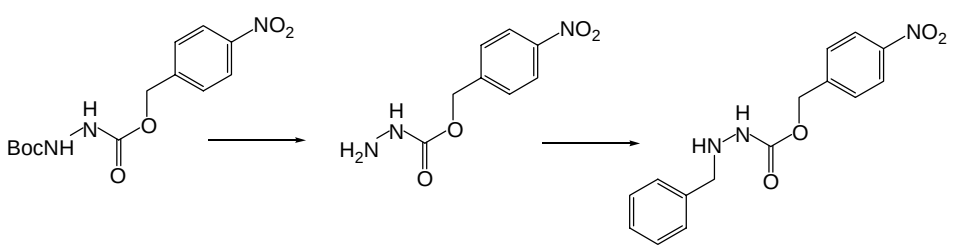

Scheme 9

of this over earlier methods is that it uses the same basic reaction sequence as the Merrifield synthetic strategy. This involves the reaction of an azaamino acid building block, protected with a group such as *p*-nitrobenzyl, with a carbonyl-releasing agent such as phosgene in the presence of base to afford a protected, activated carbonyl azaamino acid. This crude product is subsequently allowed to react with a resin-bound amino acid, peptide, or amino acid or peptide azaanalog (Scheme 8). Subsequent cleavage of the protecting group, for example, with $SnCl_2$, releases the freshly synthesized, resin-bound azapeptide. This methodology requires the synthesis of azaamino acid building blocks, which can usually be achieved in three steps by reaction

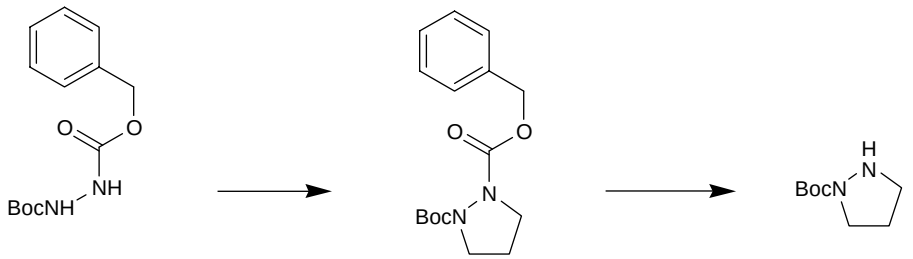

Scheme 10

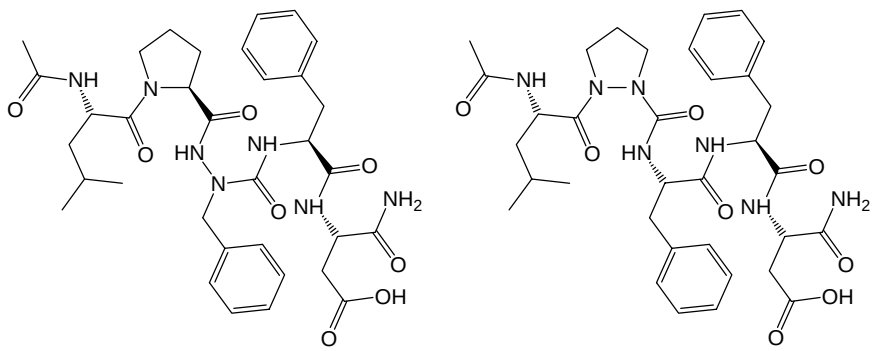

Fig. 1 Two azapeptide analogs of Ac-Leu-Pro-Phe-Phe-AspNH_2.

of *t*-butyl carbazate with *p*-nitrobenzyl chloroformate, selective removal of the *t*-butyloxycarbonyl group, and alkylation of the primary amino group (Scheme 9).

A slightly different challenge is posed by the azaproline moiety, which was synthesized according to the sequence shown in Scheme 10. In this particular case, the t-butyloxycarbonyl derivative was used directly in the chlorocarbonylation and coupling steps. This methodology was used to synthesize azapeptides such as Ac-Leu-Pro-Phe^a-Phe-AspNH_2 and Ac-Leu-Pro^a-Phe-Phe-AspNH_2 (Fig. 1).

3
β-, γ-, δ-, ε-, and Related Peptides

β-, γ-, δ- and ε-peptides are oligomers of β-, γ-, δ- or ε-amino acids (Fig. 2). Depending upon their backbone substitution, they may be classified as β^2-, $\beta^{2,2}$-, β^3-, and so on (depicted are only β^2-, γ^2-, δ^2- and ε^2- peptides). Although sugar amino acids are not covered in this chapter, they may also be viewed as building blocks for conformationally restricted β-, γ-, δ-, and ε-peptides, for example, Fig. 3. These compounds have been reviewed as recently as 2002 by Gruner et al.

Another recent review, concentrating on the structure of β-peptides, but also briefly reviewing their biological properties, is that of Cheng et al. in 2001. Another review by Lelais and Seebach in 2004 with an update of the subject almost to the present day would render any other attempt repetitious at this time.

On several occasions, γ-peptides have been reported to adopt hydrogen bond–stabilized helical conformations, but in

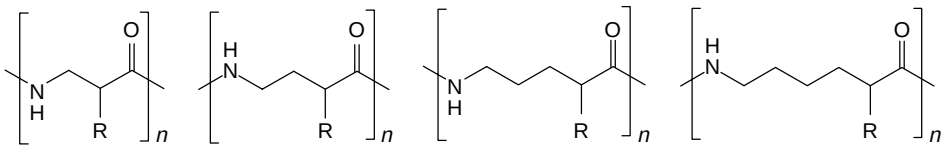

Fig. 2 General structures of β^2-, γ^2-, δ^2- and ε^2- peptides.

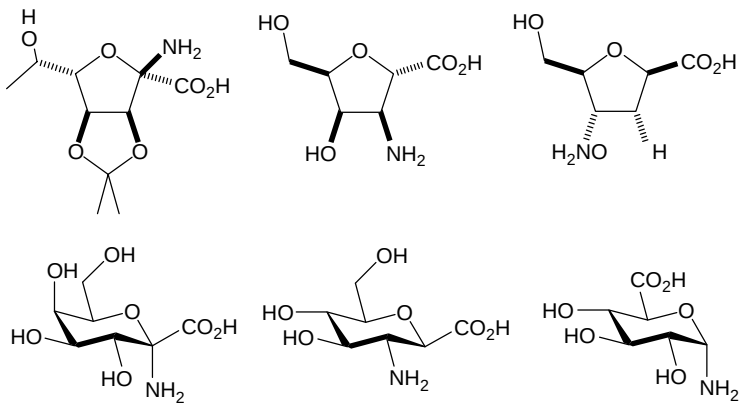

Fig. 3 Examples of sugar amino acid building blocks.

oligomers of γ-amino acid residues, a sheet secondary structure is preferred. Nowick and Brower in 2003 and Zhao et al. in 2004 have constructed bulky N^{α}-acylornithine-derived δ-peptides that adopt well-defined folded conformations, unlike earlier, presumably more flexible examples. The synthesis of these compounds usually adheres to standard peptide synthetic methodology.

4 Aminooxy Peptides

α-,β- and γ-aminooxypeptides are analogs of β-, γ- or δ-peptides in which the β-, γ- or δ-carbon is replaced by an oxygen atom (Fig. 4). The unusual conformational properties of peptides containing a single α-aminooxy acid residue were first predicted and observed by Yang et al. in 1996. A recent important development in this field was the first solid-phase synthesis of oligomeric α-aminooxy peptides. This was based on the stepwise extension of a resin-bound α-phthalimidoxy acid (Scheme 11). The same group has introduced 2-oxanipecotic acid as a cyclic α-aminooxy acid for conformational studies, for example, Fig. 5. The synthesis of phthaloyl δ-aminooxy acids, the building blocks of chiral δ-aminooxy peptides, has been reviewed and improved upon by Shin et al. in 2000.

Chiral β^3-aminooxy peptides are a recent development. They have been obtained in a straightforward manner by coupling of phthalimidoxy acids with the free amino groups of their partners, under similar conditions to those used previously by the same authors. The chiral β^3-aminooxy amino acid building blocks were synthesized from chiral α-aminooxy

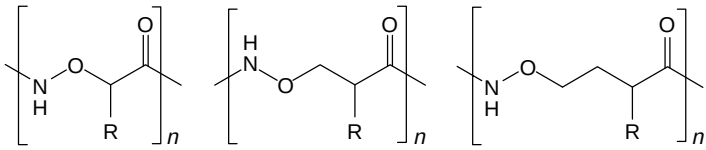

Fig. 4 General structures of α-, β- and γ- aminooxypeptides.

Fig. 5 A 2-oxanipecotic acid dimer moiety.

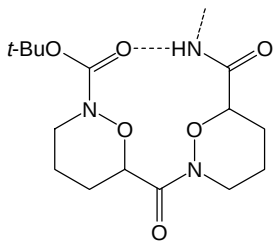

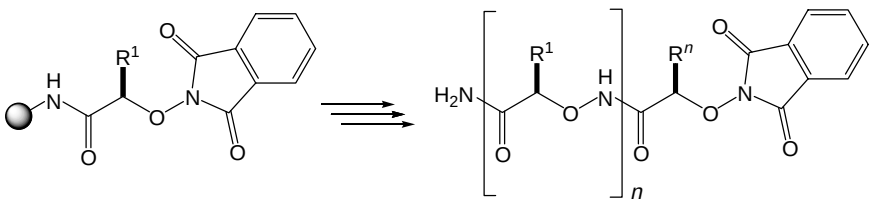

Scheme 11

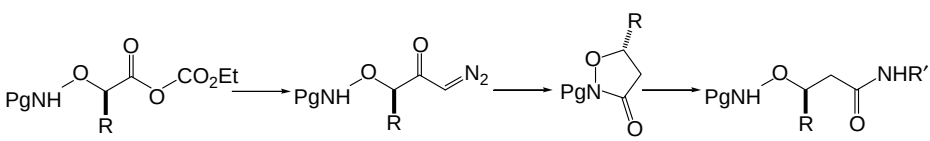

Scheme 12

Fig. 6 A γ^4-aminooxydipeptide.

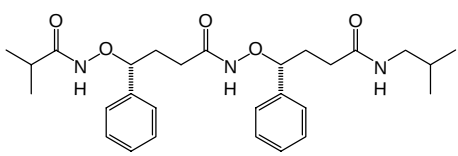

acids by Arndt-Eistert homologation (Pg = protecting group) (Scheme 12) or, preferably, by enantioselective reduction of β-keto esters using either baker's yeast or chiral Ru(II) complexes, followed by reduction of the ester group and amination of the secondary alcohol (Scheme 13).

γ^4-Aminooxy peptides have also been synthesized very recently and have been shown to adopt unique helical conformations, different from those of homochiral oligomers of α- and β-aminooxy acids (Fig. 6). Experimental details of their syntheses have not yet been published.

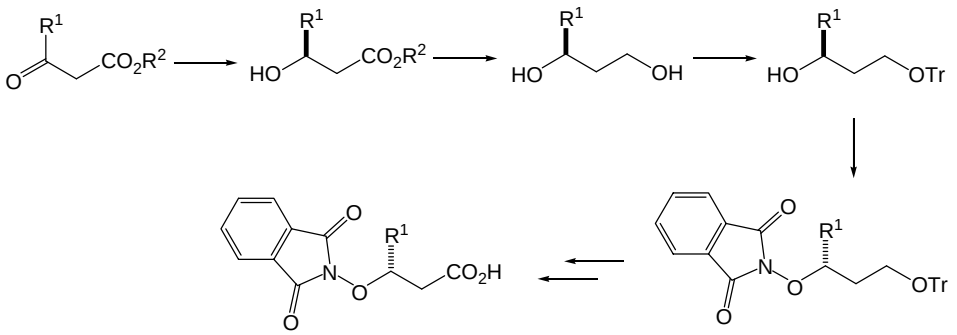

Scheme 13

5 Oligocarbamate and Oligourea Peptidomimetics

Oligocarbamate peptide mimetics were first synthesized as a library by Cho et al. in 1993, using methodology described for the preparation of peptide libraries. Their general structure may be depicted as shown in Fig. 7. The first successful attempts to prepare oligourea peptidomimetics (Fig. 8) using solid-phase technology were published in detail by Burgess et al. in 1997 and Kim et al. in 1996. The carbamate or urea amino acid analogs are generally represented by the usual three- or one-letter amino acid notations with the superscripts c or u, respectively.

The earlier methods were based on phthalimido-protected isocyanates and azido 4-nitrophenyl carbamates as building blocks, which diverge considerably from the standard methodologies using, for example, Fmoc-protected amino acids or amino acid analogs, such as described by Chan and White in 2000. More recent detailed protocols for the solid-phase synthesis of oligourea peptidomimetics and their building blocks using Fmoc protection under conditions mimicking those used in automated peptide synthesis are given in Boeijen et al. (2001). The general procedure is based on the extension of a Rink amide resin-linked starter by successive addition of Fmoc and p-nitrophenyl-protected urea monomers with intermediate capping and deprotection steps (Scheme 14). Most of the monomers were synthesized from Fmoc-protected amino acids by successive conversion to the corresponding alcohols, azides,

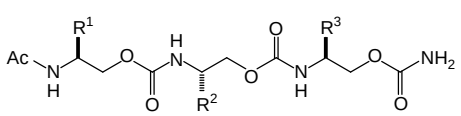

Fig. 7 An oligocarbamate peptide mimetic.

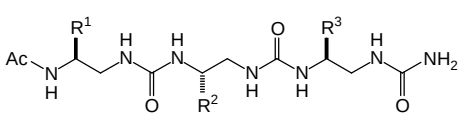

Fig. 8 An oligourea peptide mimetic.

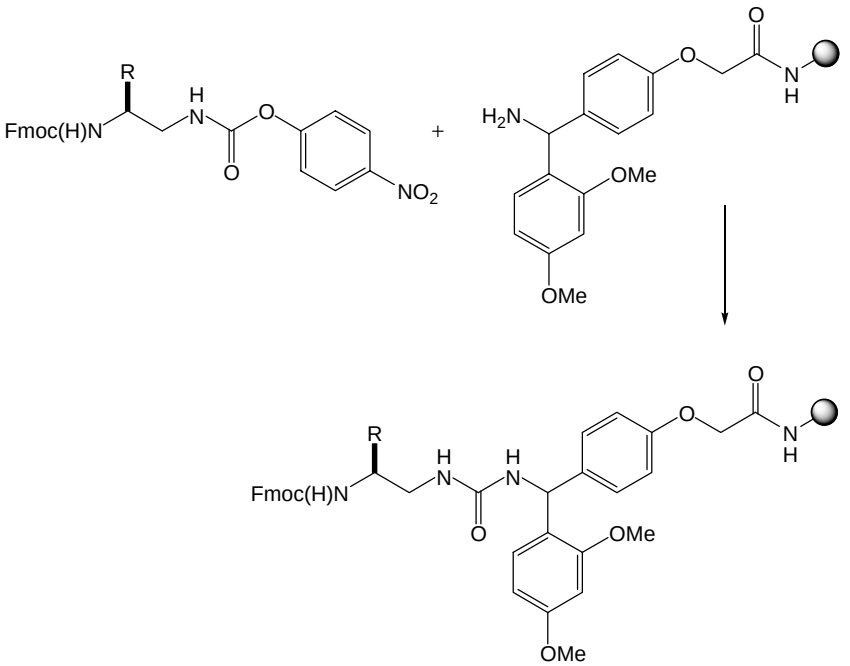

Scheme 14

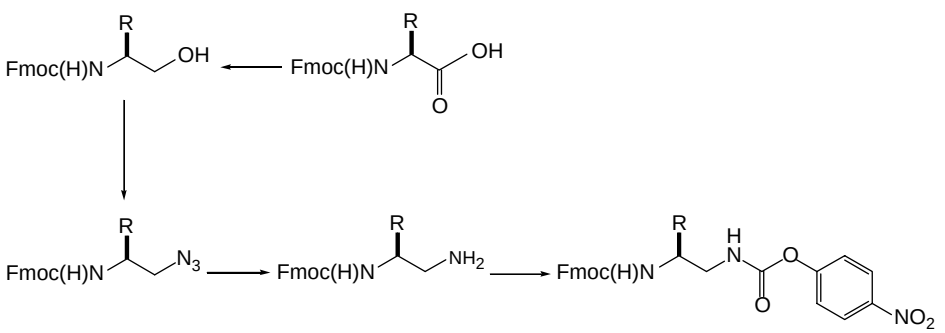

Scheme 15

and amines, which were finally derivatized to the p-nitrophenyl carbamates (Scheme 15). The Fmoc-protected glycine urea monomer, however, was prepared from Boc-ethylenediamine. This methodology was used to synthesize Ac-Gly^{u}-Gly^{u}-Phe^{u}-Leu^{u}-NH_2 and a couple of neurotensin analogs.

6 Peptoids

Peptoids are N-alkylglycine oligomers, that is, peptide mimetics in which side chains are bound to the amide nitrogen atom rather than to the α-carbon atom (Fig. 9). Peptoid synthesis based on commercially

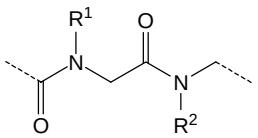

Fig. 9 A dipeptoid moiety.

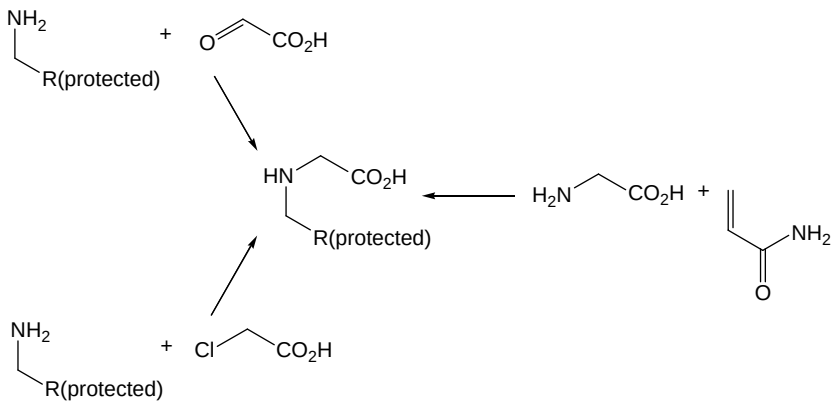

Scheme 16

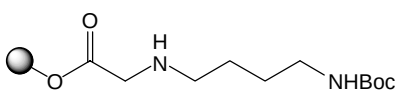

Fig. 10 A resin-bound peptoid residue.

available primary amines is eminently amenable to the preparation of very large libraries using combinatorial approaches. Earlier work in this direction, pioneered by Zuckermann's group has been reviewed by Figliozzi et al. in 1996. A limitation of such processes has been the long reaction time (of up to several hours) required for the coupling of each residue. However, it has been shown in 2002, by Olivos et al. that the solid-phase synthesis of peptoids can be successfully accelerated under microwave irradiation, achieving reaction times of less than one minute. The basic strategy is the approach described by Simon et al. in 1992), which uses standard solid-phase synthesis techniques. The monomers are prepared by different methods from appropriate primary amines (Scheme 16).

Shankaramma et al. in 2003 synthesized tetradecapeptide-peptoid macrocycles containing a single N-substituted glycine residue, allowing a β-hairpin conformation induced by two successive proline residues to be maintained, stabilized here by the array of intramolecular hydrogen bonds between carbonyl and NH groups. The synthesis of the linear precursor was carried out on 2-chlorotrityl chloride-polystyrene resin using Fmoc chemistry, initially binding bromoacetic acid to the resin and then replacing the bromine with a 4-Boc-aminobutylamino group to create the single peptoid residue as a starting point (Fig. 10). This precursor was then

released from the resin, cyclized and deprotected to afford the final product.

7 Variations on the Peptoid Theme

Several new variations on the peptoid theme have been synthesized. One of the earliest is the ureapeptoid structure built by Kruijtzer et al. in 1997 and by Wilson and Nowick in 1998 (Fig. 11). Ureapeptoids can be prepared from Boc-protected *N*-1-substituted ethylenediamines, by conversion to the corresponding isocyanates and condensation of the latter with sarcosine methyl ester (Scheme 17). Subsequent removal of the Boc protective group and condensation with an appropriate *p*-nitrophenyl carbamate allowed extension of the chain to a ureapeptoid trimer.

A more efficient method starts from Boc-ethylenediamine or oligomers containing this moiety, by chain extension with *N*-(2-nitrobenzenesulfonyl)-2-imidazolidinone containing the Ns protective group of Fukuyama et al., 1995 followed by N-alkylation and deprotection (Scheme 18).

Retropeptoids, for example, Fig. 12, are peptidomimetics in which the substituted nitrogen atom of the peptoid N-substituted glycine is moved up one position to replace the α-carbon atom as described

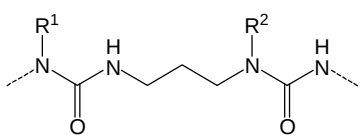

Fig. 11 A ureapeptoid moiety.

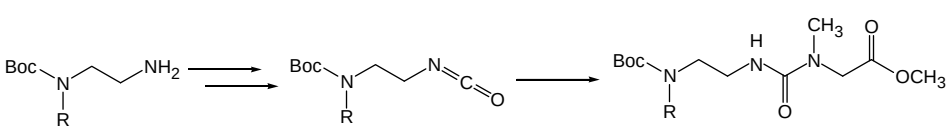

Scheme 17

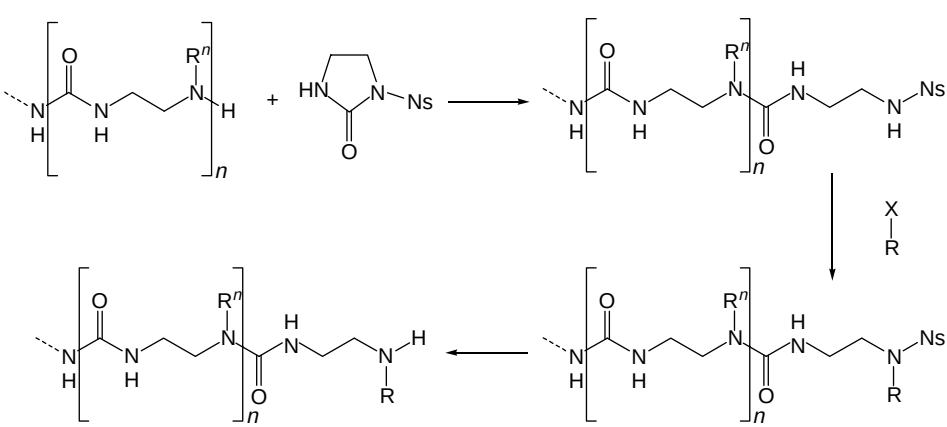

Scheme 18

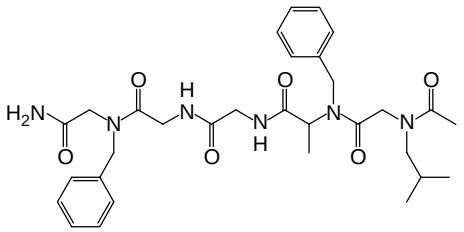

Fig. 12 A retropeptide.

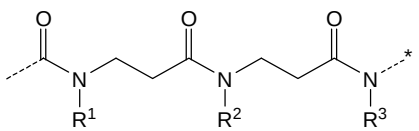

Fig. 13 A β-peptoid moiety.

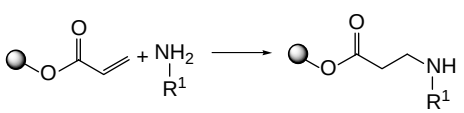

Scheme 19

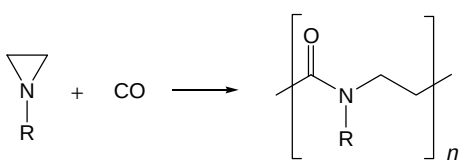

Scheme 20

by Kruijtzer et al. in 1998. β-Peptoids (Fig. 13) are N-substituted oligomers of β-aminopropionic acid. This type of oligomer had been prepared previously by coupling of N-substituted β-amino acids as explained by Seebach et al. in 1996 using standard peptide methodology, and shown to be stable toward pepsin. The Seebach approach suffers from the limitation that it requires a collection of β-amino acid monomers as building blocks. Hamper et al. in 1998 devised a solid-phase strategy, based on the Michael addition of primary amines to an acrylic acid residue bound to Wang resin by treatment with acryloyl chloride (Scheme 19). Successive couplings with acryloyl chloride followed by the addition of primary amines and finally cleavage from the resin led to the synthesis of several di-β-peptoids, as well as trimeric *N*-benzyl-β-aminopropionic acid, the latter in 67% overall yield.

Poly-β-peptoids are oligomers of a single N-substituted β-aminopropionic acid residue. Approaches to these structures by copolymerization of CO and imines were explored initially by Kacker et al. in 1998 and by Dghaym et al. also in 1998, and have been developed into a successful catalytic method by the group of Jia in 2002 and Darensbourg in 2004. This involves the alternating copolymerization of carbon monoxide and an N-alkylaziridine using a cobalt carbonyl catalyst (Scheme 20).

Hydrazinoazapeptoids are based on the hydrazinopeptide concept (Fig. 14). The synthesis of hydrazinoazapeptoids (as hybrids with a hydrazinopeptide moiety)

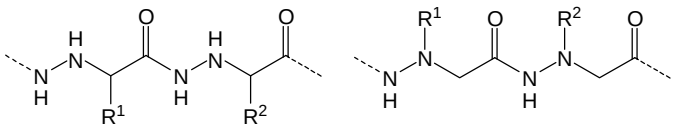

Fig. 14 A hydrazinopeptide and a hydrazinopeptoid moieties.

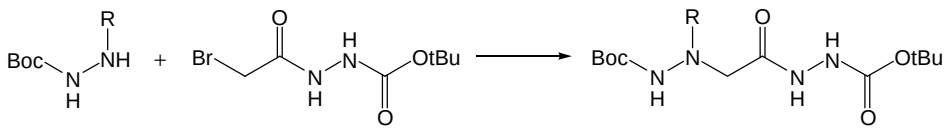

Scheme 21

Scheme 22

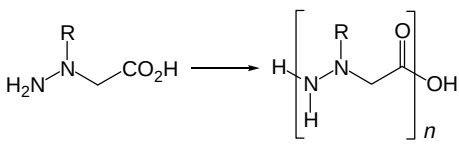

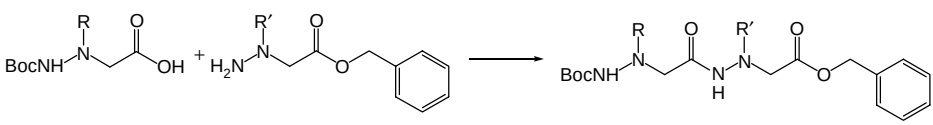

Scheme 23

Fig. 15 An aminooxypeptide and an aminooxypeptoid moieties.

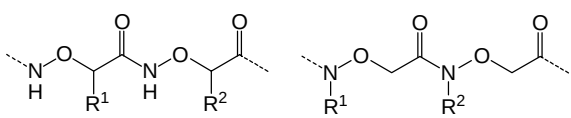

has been carried out in solution, using the submonomer approach, which implies alkylation of one of the nitrogen atoms of a monosubstituted alkyl or arylalkylhydrazine with a bromoacetyl residue on the growing oligomer, requiring the use of protected hydrazine derivatives to insure alkylation at the desired position (Scheme 21).

Aza-β^3-peptides are oligomers of N^{α}-substituted hydrazinoacetic acid, and may also be regarded as aza analogs of β^3-amino acids (Scheme 22). The monomers can be synthesized as N^{β}-Boc-protected benzyl or methyl esters, and monomers, dimers, and oligomers all coupled very efficiently under DCC/DMAP (dicyclohexylcarbodiimide/dimethylaminopyridine) activation, followed by deprotection and elongation steps as required (Scheme 23).

Aminooxypeptoids, based on the aminooxypeptide concept, can be envisioned as *N*-alkyl- or arylalkylhydroxylamine derivatives (Fig. 15). The synthesis of aminooxypeptoids has also been accomplished using the submonomer approach

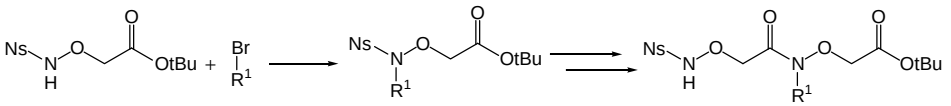

Scheme 24

(Ns represents *o*-nitrobenzenesulfonyl), but unacceptably low yields were obtained (Scheme 24). Stepwise assembly of the N-alkylated monomers in either the N to C or the C to N direction proved to be more satisfactory, although yields were still low when bulky side chains were present at the terminal positions of the deprotected monomers.

8 Proline Substitutes

The cyclic structure of proline and the similar stabilities of cis- and trans-peptide bonds involving its secondary amine nitrogen, separated by a relatively low potential energy barrier, are features that explain the unique roles of this amino acid. As similar properties are found in other $N^{\alpha}-C^{\alpha}$-cyclized amino acids, a number of proline mimics with different ring sizes have been synthesized and incorporated into peptides. These modifications are discussed in the Goodman and Ro review (1995), as well as others in which an additional heteroatom and/or a conformationally stabilizing feature such as a double bond, a bicyclic system, or a *gem*-dimethyl substitution is included in the proline ring. A number of proline substitutes have been biologically characterized (see Sect. 9).

9 Biological Activity and Therapeutic Hints

Although theoretically considered as promising fields to develop innovative therapies for a number of relevant human diseases, the biological characterization of most of the peptide mimetics described here has not been the subject of too much research, as compared with the abundant literature regarding synthetic aspects. Regarding endogenous peptides, these were used rather intuitively. Some classical examples are insulin, human growth hormone, interferons, or erythropoietin. Peptidase inhibition, as illustrated by angiotensin-converting enzyme (ACE) inhibitors, also offers enormous therapeutic opportunities, and the many successful ACE inhibitors currently available may be viewed broadly as mimics of the peptide substrate. Moreover, peptide processing is a crucial step in viral replication, and protease inhibition has become a fundamental aspect of HIV antiretroviral therapy. Increasing knowledge of the roles of peptidases is continually suggesting novel approaches to the treatment of disease. Therefore, the design and synthesis of substrate or transition state analogs to inhibit these enzymes is another fertile field of research.

While these molecules have found some clinical use, the drug potential of many others remains unrealized. One reason for this is that peptides are not bioavailable orally, their transport through biological membranes is usually unsatisfactory, and they are rapidly metabolized. Another is that the inherent conformational flexibility of peptides signifies an entropic penalty for them to be able to adopt appropriate conformations to interact with their targets.

Azapeptides such as Ac-Leu-Pro-Phe^a-Phe-AspNH_2 and Ac-Leu-Pro^a-Phe-Phe-AspNH_2 are effective, proteolytically stable inhibitors of the cytotoxicity of β-amyloid peptide, and are therefore of interest as β-sheet breaking compounds of potential utility in the treatment of diseases such as Alzheimer's, chronic progressive traumatic encephalopathy, and vascular dementia with amyloid angiopathy.

Pure or all-azaaminoacid oligomers or "azatides" are azapeptides in which all the α-carbon atoms have been replaced by nitrogen. Although early attempts to synthesize this kind of analog date back as far as the history of (mono)azapeptides, the first successful azatide synthesis appears to be a report by Gante et al. in 1995 on the preparation of an analog of a peptidic renin inhibitor, which, however, does not seem to have been tested. A year later, Han and Janda described the development of general solution-phase procedures for the coupling of Boc-protected azaamino acid monomers and the synthesis of these building blocks. This methodology was exemplified by synthesizing the azatide mimic of Leu-enkephalin. Unfortunately, this product did not compete with Leu-enkephalin for an antibody raised against β-endorphin, the only assay carried out as a test of biological activity. This disappointing result, together with the subsequent discovery that newer azapeptide analogs of the original compound, but with one or more natural amino acid residues, also lacked this activity, may explain why this avenue of research does not seem to have been followed more extensively.

On the basis of the sequence of the RNA-binding domain of HIV-1 Tat protein, ^{48}Gly-Arg-Lys-Lys-Arg-Arg-Gln-Arg-Arg-Arg^{57}, Tamilarasu et al. in 2001 synthesized an oligocarbamate (Phe-Gly^c-Arg^c-Lys^c-Lys^c-Arg^c-Arg^c-Gln^c-Arg^c-Arg^c-Arg^c) and an oligourea (Gly^u-Arg^u-Lys^u-Lys^u-Arg^u-Arg^u-Gln^u-Arg^u-Arg^u-Arg^u-Tyr) analog, which inhibited transcriptional activation by Tat protein in human cells with IC_{50} values of 1 and 0.5 μM, respectively.

The pharmacological activity of peptoids designed as analogs of peptides with recognized bioactivity has often been disappointing. Nevertheless, some peptoid structures lacking any obvious fit to a target exhibit interesting activities. Thus, a combinatorial library of peptoid trimers yielded several new compounds (with a large, lipophilic, nonpeptidomimetic, also antibacterial dehydroabietylamine moiety and unnatural amino acid–like residues) exhibiting activity against both gram-negative and gram-positive bacteria, including some resistant strains to all known antibiotics. The MIC (minimal inhibitory concentration) values were as low as 5 μM in some cases. One of the analogs fully protected *Staphylococcus aureus*-infected mice at 10 and 30 mg kg^{-1} (Fig. 16).

It should be pointed out that dehydroabietylamine itself is no more than an order of magnitude less potent *in vitro* than the most effective peptoid hybrid of this series. Nevertheless, it is inactive in the infection

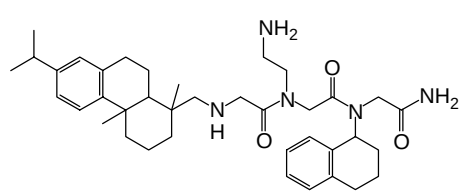

Fig. 16 A dehydroabietylaminopeptoid.

model, suggesting that the pharmacokinetics for some peptoids is more favorable. The authors suggest that, because of their broad spectrum of antibacterial activity, their lack of structural specificity, and β-galactosidase and propidium iodide leakage, these substances act on the bacterial cell membrane. Although the more thoroughly studied compound depicted above had some hemolytic activity, this appeared at concentrations greater than the MIC for staphylococci. The screen that led to these compounds showed that dehydroabietylamine, at any of the three positions, although preferably at the N-terminal location, was a frequent (but neither necessary nor sufficient) component of the active analogs. The extension of the antibacterial spectrum to gram-positive organisms seems to be related to the presence of a basic residue (such as the aminoethyl group mentioned earlier) at the intermediate position.

Quite recently, a conjugate of ethylenediaminetetraacetic acid (EDTA) with two 5-aminosalicylic acid (5-ASA) methyl ester moieties (EBAME), was described (Fig. 17). This compound can be viewed as a peptoid in which two glycine conjugates are joined through their α-amino acid nitrogen atoms, combining the anti-inflammatory activities of 5-ASA, with the redox-active metal chelating activity of acid (EDTA), suppressing the gastric irritation mediated by 5-ASA, removing potentially harmful transition metal ions and mimicking superoxide dismutase (SOD) when complexed with Cu^{2+} ions. EBAME was shown to bind to Cu^{2+} in a 1 : 1 ratio, but the superoxide-destroying activity of this complex was rather low. With Mn^{2+}, however, a very satisfactory SOD-like activity was demonstrated. The authors claim that EBAME has potential as a dual-function anti-inflammatory agent with reduced gastric irritant potential.

A couple of macrocyclic peptide–peptoid hybrids, mentioned in Sect. 6, has also quite recently been shown to have antibiotic activity against gram-positive and – negative bacteria, with low hemolytic activity against erythrocytes, one of the major drawbacks of earlier members of this class (Fig. 18). The replacement of an arginine group N-substituted glycine residue present in a previously synthesized purely peptidic prototype by an *N*-4-aminobutyl-substituted glycine residue led to a slightly

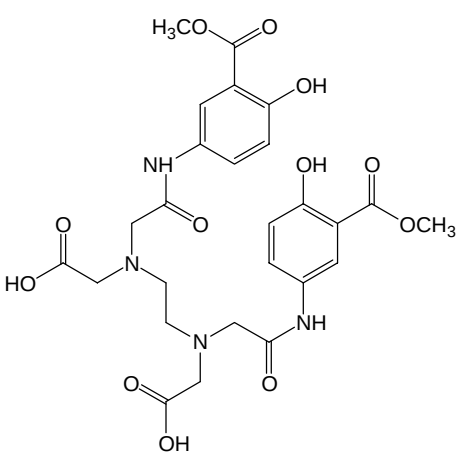

Fig. 17
Ethylenediaminetetraaceptic acid bis-(5-aminosalicylic acid methyl ester) (EBAME).

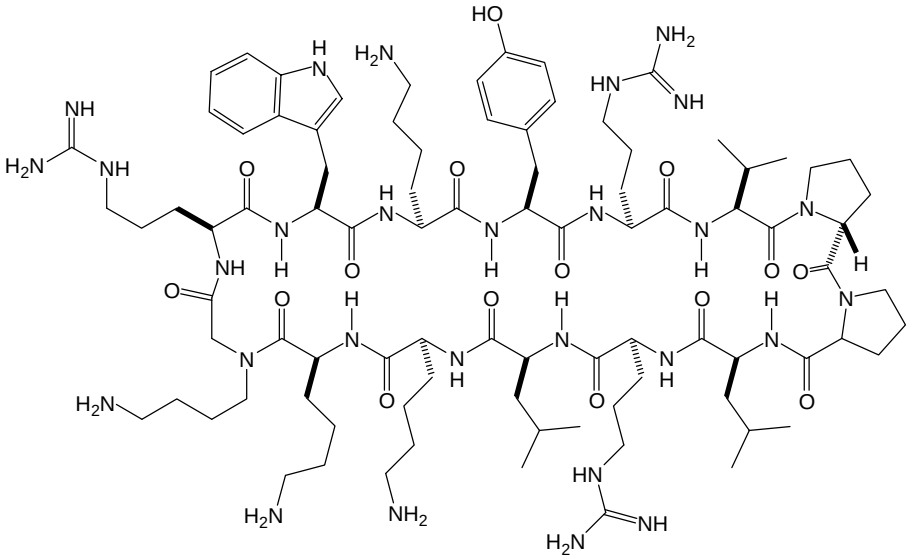

Fig. 18 A macrocyclic peptide–peptoid hybrid.

improved antibiotic potency (MIC between 4 and 64 μg mL^{-1}) and reduced hemolytic activity (0.5% at 100 μg mL^{-1}), while preserving the β-hairpin conformation believed to be important for this biological activity.

When α-chiral, bulky side chains are repeated along a peptoid backbone, stable helical structures arise in spite of the lack of potentially hydrogen-bonding NH groups. The same structural concept has been applied to the design and synthesis of a mimic of lung surfactant protein C, containing 22 N-substituted glycine residues. This product successfully reproduces several desirable characteristics such as the ability to adsorb rapidly to an air–water interface, to reduce and control surface tension and to respread quickly upon surface expansion (Fig. 19).

Several helical, cationic, amphipathic peptoids designed as mimics of the natural magainins have also shown antibacterial activity, for example, Fig. 20.

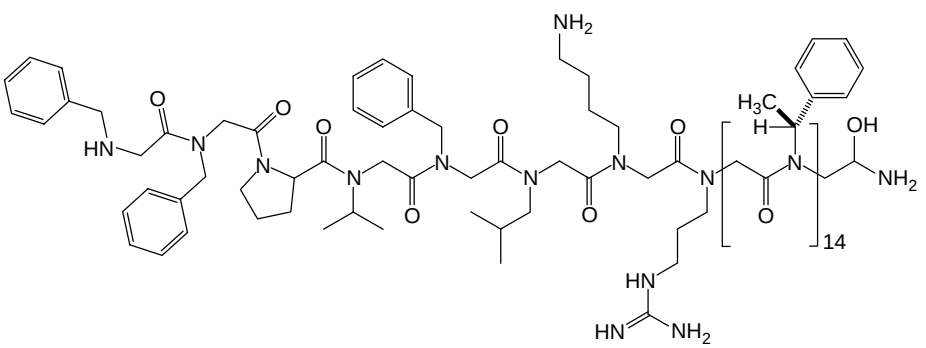

Fig. 19 A peptoid mimic of lung surfactant protein C.

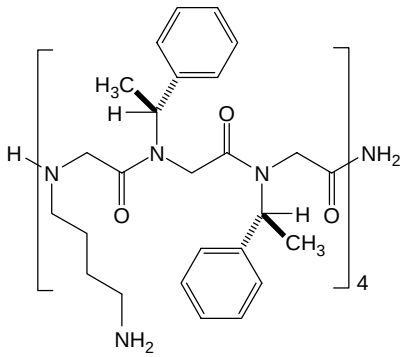

Fig. 20 A peptoid magainin mimic.

These oligomers were prepared from peptoid polyamines, synthesized sequentially on Rink amide resin, which were then guanidinylated using pyrazole 1-carboxamidine (Scheme 25).

The retropeptide Leu-enkephalin amide mimic depicted earlier (Fig. 12) appears not to have been subjected to biological testing. The more complex substance P peptoid and retropeptoid analogs, however, are reported to be *in vitro* agonists in the above-mentioned article, although this observation does not seem to have been followed up by a more thorough pharmacological study. Rather surprisingly, the high resistance of the peptoid analog to degradation by pepsin is mentioned as evidence that the retropeptoid is similarly stable.

Some hydrazinoazapeptoids designed as possible proteasome inhibitors, based on the Ac-Leu-Leu-Norleucinal template, exhibit antiproliferative activity in L1210 cells, for example, Fig. 21, but their potency is weak compared to that of the structurally unrelated, proven proteasome inhibitor bortemozib.

A recent successful example is the synthesis of a potent ($K_i = 1.5$ nM), specific human proline endopeptidase inhibitor

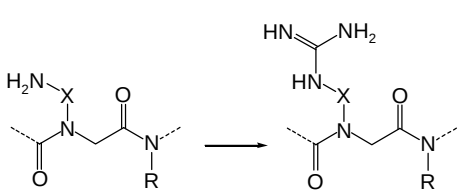

Scheme 25

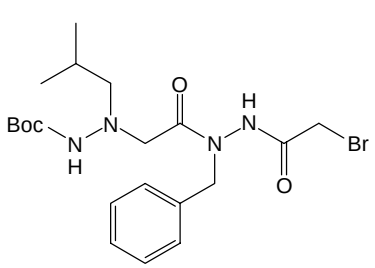

Fig. 21

that was viewed as a probe of the possible role of this enzyme in the production of β-amyloid peptide, a hallmark of Alzheimer's disease (Fig. 22).

A newer development is the replacement of a proline residue by a cyclopent-2-ene-1,2-dicarboxylic acid moiety. Applying this concept, these authors synthesized and evaluated a series of thrombin inhibitors related to inogatran and melagatran (in which the pyrrolidine ring of a proline residue is replaced by a piperidine or an azetidine ring, respectively), some of them with submicromolar IC_{50} values, which, however, are still somewhat unsatisfactory when compared with the 22.4 and 4.7 nM IC_{50} values of the reference compounds. Their most potent example ($IC_{50} = 0.87\ \mu M$) is shown in Fig. 23. For the sake of comparison, the analog lacking the N-ethoxy group has $IC_{50} = 1.51\ \mu M$. Although the (2R)-cyclopent-2-ene-1,2-dicarboxylic acid moiety generally afforded more potent compounds, its replacement by trans-(1S,2S)-cyclopentane-1,2-dicarboxylic acid in the latter example results in slightly enhanced activity ($IC_{50} = 1.32\ \mu M$). This structural modification has been incorporated much more successfully into inhibitors of prolyl oligopeptidase, a serine peptidase that preferentially catalyzes the hydrolysis of peptide bonds at the carboxyl side of proline residues. Thus, the following analog has a 0.3 nM IC_{50}, while the corresponding value for the reference compound with a proline residue in the central position is 0.2 nM (Fig. 24).

Recents efforts in this direction are those of Han et al. in 1999, Avenoza et al. in 2002, and Jenkins et al. in 2004. The former authors used 7-azabicyclo[2.2.1]heptane-1-carboxylic acid as a proline replacement in a boroarginine thrombin inhibitor, and as an inducer of a β-turn in the dipeptides Ser-Pro and Pro-Ser for conformational studies (Fig. 25).

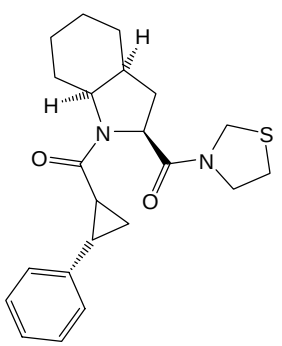

Fig. 22

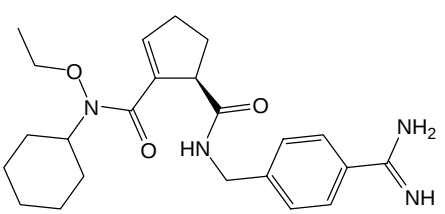

Fig. 23

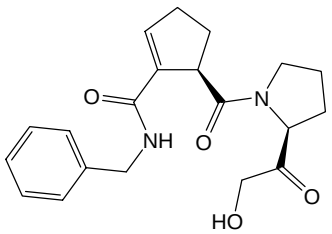

Fig. 24

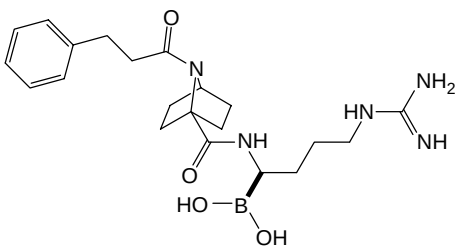

Fig. 25

This substance, synthesized from 7-Boc-1-carboxy-7-azabicyclo[2.2.1]heptane t-butyl ester, inhibited thrombin with K_i = 2.9 nM, which compares reasonably well with the parent proline compound (K_i = 0.10 nM) suggesting that this replacement mimics the active conformation in Han et al. in 1999.

The more recent paper analyzed the implications of this replacement, with electronegative (HO or F) substituents at C-4, for the conformation of collagen. Jenkins et al. in 2004 concluded that the bridge abolishes the effect of the electronegative substituents on the proline ring on the trans/cis ratio of the peptide bonds, previously observed by the same group with unbridged, electronegatively substituted proline residues.

The unique features of peptide-inspired derivatives pose many synthetic challenges that must still be met. Certain structural aspects, such as the tendency of different peptidomimetic scaffolds to adopt stable conformations in solution, are currently a very active field of research that is also a stimulus for the development of novel synthetic methodologies. In view of the extremely varied therapeutic potential of peptide mimetics, the development of reliable therapies based on these compounds will require further expansion of the broad spectrum of synthetic methods available in order to overcome the current practical limitations of these very versatile moieties, without altering their essential chemical nature. This frontier in the field of peptide mimetics will foreseeably be a major area of research in the coming years.

Bibliography

Books and Reviews

Abell, A. (Ed.) (1999) *Advances in Amino Acid Mimetics and Peptidomimetics*, JAI Press, Greenwich.

Chan, W.C., White, P.D. (Eds.) (2000) *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*, Oxford University Press, Oxford.

Cheng, R.P., Gellman, S.H., DeGrado, W.F. (2001) β-Peptides: From structure to function, *Chem. Rev.* **101**, 3219–3232.

Goodman, M., Ro, S. (1995) in: Wolff, M. (Ed.) *Burger's Medicinal Chemistry and Drug*

Discovery, Vol. 1, 5th edition, John Wiley & Sons, New York, p. 803.

Goodman, M., Moroder, L., Toniolo, C., Felix, A. (Eds.) (2002) *Houben-Weyl, Methods in Organic Chemistry, Vol. E22. Synthesis of Peptides and Peptidomimetics*, Thieme Medical Publishers, Stuttgart and New York.

Hruby, V.J., Balse, P.M. (2000) Conformational and topographical considerations in designing agonist peptidomimetics from peptide leads, *Curr. Med. Chem.* **7**, 945–970.

Simon, R.J., Kania, R.S., Zuckermann, R.N., Huebner, V.D., Jewell, D.A., Banville, S., Ng, S., Wang, L., Rosenberg, S., Marlowe, C.K., Spellmeyer, D.C., Tan, R., Frankel, A.D., Santi, D.V., Cohen, F.E., Bartlett, P.A. (1992) Peptoids: a modular approach to drug discovery, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9367–9371.

Williams, R.M. (1996) Amino Acid Synthesis, in: Meyers, R.B. (Ed.) *Encyclopedia of Molecular Biology and Molecular Medicine*, Vol. 1, Wiley-VCH Verlag GmbH, Weinheim, p. 52.

Zega, A., Urleb, U. (2002) Azapeptides, *Acta Chim. Slov.* **49**, 649–662.

Primary Literature

Avenoza, A., Busto, J.H., Peregrina, J.M., Rodríguez, F. (2002) Incorporation of Ahc into model dipeptides as an inducer of a β-turn with a distorted amide bond. Conformational analysis, *J. Org. Chem.* **67**, 4241–4249.

Baek, B.-H., Lee, M.-R., Kim, K.-Y., Cho, U.-I., Boo, D.W., Shin, I. (2003) Novel consecutive β- and γ-turn mimetics composed of a-aminooxy acid tripeptides, *Org. Lett.* **5**, 971–974.

Bailey, M.A., Ingram, M.J., Naughton, D.P. (2004) A novel anti-oxidant and anti-cancer strategy: a peptoid anti-inflammatory drug conjugate with SOD mimic activity, *Biochem. Biophys. Res. Commun.* **317**, 1155–1158.

Barelli, H., Petit, A., Hirsch, E., Wilk, S., De Nanteuil, G., Morain, P., Checler, F. (1999) S 17092-1, a highly potent, specific and cell permeant inhibitor of human proline endopeptidase, *Biochem. Biophys. Res. Commun.* **257**, 657–661.

Barré, C., Le Grel, P., Robert, A., Baudy-Floc'h, M. (1994) Design and synthesis of a new series of peptide analogs: the hydrazinopeptides, *J. Chem. Soc. Chem. Commun.* **1994**, 607–608.

Boeijen, A., van Ameijde, J., Liskamp, R.M.J. (2001) Solid-phase synthesis of oligourea peptidomimetics employing the Fmoc protection strategy, *J. Org. Chem.* **66**, 8454–8462.

Bouget, K., Aubin, S., Delcros, J.-G., Arlot-Bonnemains, Y., Baudy-Floc'h, M. (2003) Hydrazino-aza and N-azapeptoids with therapeutic potential as anticancer agents, *Bioorg. Med. Chem.* **11**, 4881–4889.

Bretscher, L.E., Jenkins, C.L., Taylor, K.M., DeRider, M.L., Raines, R.T. (2001) Conformational stability of collagen relies on a stereoelectronic effect, *J. Am. Chem. Soc.* **123**, 777–778.

Burgess, K., Linthicum, D.S., Shin, H. (1995) Solid-phase syntheses of unnatural biopolymers containing repeating urea units, *Angew. Chem., Int. Ed. Engl.* **34**, 907–909.

Burgess, K., Ibarzo, J., Linthicum, D.S., Russell, D.H., Shin, H., Shitangkoon, A., Totani, R., Zhang, A.J. (1997) Solid phase synthesis of oligoureas, *J. Am. Chem. Soc.* **119**, 1556–1564.

Chang, R.S., Lotti, V.J., Monaghan, R.L., Birnbaum, J., Stapley, E.O., Goetz, M.A., Albers-Schonberg, G., Patchett, A.A., Liesch, J.M., Hensens, O.D., Springer, J.P. (1985) A potent nonpeptide cholecystokinin antagonist selective for peripheral tissues isolated from Aspergillus alliaceus, *Science* **230**, 177–179.

Cheguillaume, A., Lehardy, F., Bouget, K., Baudy-Floc'h, M., Le Grel, P.M. (1999) Submonomer solution synthesis of hydrazinoazapeptoids, a new class of pseudopeptides, *J. Org. Chem.* **64**, 2924–2927.

Cheguillaume, A., Salaün, A., Sinbandhit, S., Potel, M., Gall, P., Baudy-Floc'h, M., Le Grel, P.M. (2001) Solution synthesis and characterization of aza-β^3-peptides (N^α-substituted hydrazino acetic acid oligomers), *J. Org. Chem.* **66**, 4923–4929.

Chen, F., Zhu, N.-Y., Yang, D. (2004) γ^4-Aminoxy peptides as new peptidomimetic foldamers, *J. Am. Chem. Soc.* **126**, 15980–15981.

Cho, C.Y., Moran, E.J., Cherry, S.R., Stephans, J.C., Fodor, S.P.A., Adams, C.L., Sundaram, A., Jacobs, J.W., Schultz, P.G. (1993) An unnatural biopolymer, *Science* **261**, 1303–1305.

Darensbourg, D.J., Phelps, A.L., Le Gall, N., Jia, L. (2004) Mechanistic studies on the copolymerization reaction of aziridines and carbon monoxide to produce poly-β-peptoids, *J. Am. Chem. Soc.* **126**, 13808–13815.

Dghaym, R.D., Yaccato, K.J., Arndtsen, B.A. (1998) The novel insertion of imines into a late metal-carbon σ bond: developing a palladium-mediated route to polypeptides, *Organometallics* **17**, 4–6.

Dutta, A., Morley, J. (1975) Polypeptides. Part XIII. Preparation of α-aza-amino-acid (carbazic acid) derivatives and intermediates for the preparation of α-aza-peptides, *J. Chem. Soc., Perkin Trans. 1* 1712–1720.

Figliozzi, G.M., Goldsmith, R., Ng, S.C., Banville, S.C., Zuckermann, R.N. (1996) Synthesis of N-substituted glycine peptoid libraries, *Methods Enzymol.* **267**, 437–447.

Fukuyama, T., Jow, C.-K., Cheung, M. (1995) 2- and 4-nitrobenzenesulfonamides: exceptionally versatile means for preparation of secondary amines and protection of amines, *Tetrahedron Lett.* **36**, 6373–6374.

Gante, J. (1970) Azapeptides, a novel class of peptide analogs, *Angew. Chem., Int. Ed. Engl.* **9**, 813.

Gante, J., Krug, M., Lauterbach, G., Weitzel, R., Hiller, W. (1995) Synthesis and properties of the first all-aza analog of a biologically active peptide, *J. Pept. Sci.* **1**, 201–206.

Gibson, C., Goodman, S.L., Hahn, D., Hoelzemann, G., Kessler, H. (1999) Novel solid-phase synthesis of azapeptides and azapeptoids via Fmoc-strategy and its application in the synthesis of RGD-mimetics, *J. Org. Chem.* **64**, 7388–7394.

Goodson, B., Ehrhardt, A., Ng, S., Nuss, J., Johnson, K., Giedlin, M., Yamamoto, R., Moos, W.H., Krebber, A., Ladner, M., Giacona, M.B., Vitt, C., Winter, J. (1999) Characterization of novel antimicrobial peptoids, *Antimicrob. Agents Chemother.* **43**, 1429–1434.

Gruner, S.A.W., Locardi, E., Lohof, E., Kessler, H. (2002) Carbohydrate-based mimetics in drug design: sugar amino acids and carbohydrate scaffolds, *Chem. Rev.* **102**, 491–514.

Hamper, B.C., Kolodziej, S.A., Scates, A.M., Smith, R.G., Cortez, E. (1998) Solid phase synthesis of β-peptoids: N-substituted β-aminopropionic acid oligomers, *J. Org. Chem.* **63**, 708–718.

Han, H., Janda, K.D. (1996) Azatides: solution and liquid phase syntheses of a new peptidomimetic, *J. Am. Chem. Soc.* **118**, 2539–2544.

Han, H., Yoon, J., Janda, K.D. (1998) Investigations of azapeptides as mimetics of Leu-enkephalin, *Bioorg. Med. Chem. Lett.* **8**, 117–120.

Han, W., Pelletier, J.C., Mersinger, L.J., Kettner, C.A., Hodge, C.N. (1999) 7-Azabicycloheptane carboxylic acid: a proline replacement in a boroarginine thrombin inhibitor, *Org. Lett.* **1**, 1875–1877.

Hart, M., Beeson, C. (2001) Utility of azapeptides as major histocompatibility complex class II protein ligands for T-cell activation, *J. Med. Chem.* **44**, 3700–3709.

Herranz, R. (2003) Cholecystokinin antagonists: pharmacological and therapeutic potential, *Med. Res. Rev.* **23**, 559–605.

Hess, H.-J., Moreland, W.T., Laubach, G.D. (1963) N-[2-Isopropyl-3-(L-aspartyl-L-arginyl)-carbazoyl]-L-tyrosyl-L-valyl-L-histidyl-L-prolyl-L-phenylalanine, an isostere of bovine angiotensin II, *J. Am. Chem. Soc.* **85**, 4041–4042.

Jarho, E.M., Venäläinen, J.I., Huuskonen, J., Christiaans, J.A.M., García-Horsman, J.A., Forsberg, M.A., Järvinen, T., Gynther, J., Männistö, P.T., Wallén, E.A.A. (2004) A cyclopent-2-enecarbonyl group mimics proline at the P2 position of prolyl oligopeptidase inhibitors, *J. Med. Chem.* **47**, 5605–5607.

Jenkins, C.L., Lin, G., Duo, J., Rapolu, D., Guzei, I.A., Raines, R.T., Krow, G.R. (2004) Substituted 2-azabicyclo[2.1.1]hexanes as constrained proline analogs: Implications for collagen stability, *J. Org. Chem.* **69**, 8565–8573.

Jia, L., Sun, H., Shay, J.T., Allgeier, A.M., Hanton, S.D. (2002) Living alternate copolymerization of N-alkylaziridines and carbon monoxide as a route for synthesis of poly-β-peptoids, *J. Am. Chem. Soc.* **124**, 7282–7283.

Kacker, S., Kim, S.J., Sen, A. (1998) Insertion of imines into palladium-acyl bonds: Towards metal-catalyzed alternating copolymerization of imines with carbon monoxide to form polypeptides, *Angew. Chem., Int. Ed. Engl.* **37**, 1251–1254.

Kirshenbaum, K., Barron, A.E., Armand, P., Goldsmith, R., Bradley, E., Cohen, F.E., Dill, K.A., Zuckermann, R.N. (1998) Sequence-specific polypeptoids: a diverse family of heteropolymers with stable secondary structure, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4303–4308.

Kruijtzer, J.A.W., Lefeber, D.J., Liskamp, R.M.J. (1997) Approaches to the synthesis

of ureapeptoid peptidomimetics, *Tetrahedron Lett.* **38**, 5335–5338.

Kruijtzer, J.A.W., Hofmeyer, L.J.F., Heerma, W., Versluis, C., Liskamp, R.M.J. (1998) Solid-phase syntheses of peptoids using Fmoc-protected N-substituted glycines: the synthesis of (retro)peptoids of Leu-enkephalin and substance P, *Chem. Eur. J.* **4**, 1570–1580.

Lee, M.-R., Lee, J., Baek, B.-H., Shin, I. (2003) The first solid-phase synthesis of oligomeric alpha-aminooxy peptides, *Synlett* 325–328.

Lelais, G., Seebach, D. (2004) β^2-Amino acids – Syntheses, occurrence in natural products, and components of β-peptides, *Biopolymers* **76**, 206–243.

López-Areiza, J.J., Rueckle, T., Soto-Jara, C. (2004) Aza-peptides, WO 2004/050689 A2, June 17, 2004.

Murphy, J.E., Uno, T., Hamer, J.D., Dwarki, V., Zuckermann, R.N. (1998) A combinatorial approach to the discovery of efficient cationic peptoid reagents for gene delivery, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1517–1522.

Ng, S., Goodson, B., Ehrhardt, A., Moos, W.H., Siani, M., Winter, J. (1999) Combinatorial discovery process yields antimicrobial peptoids, *Bioorg. Med. Chem.* **7**, 1781–1785.

Nöteberg, D., Brånalt, J., Kvanrnström, I., Linschoten, M., Musil, D., Nyström, J.-E., Zuccarello, G., Samuelsson, B. (2000) New proline mimetics: synthesis of thrombin inhibitors incorporating cyclopentane- and cyclopentenedicarboxylic acid templates in the P2 position. Binding conformation investigated by X-ray crystallography, *J. Med. Chem.* **43**, 1705–1713.

Nowick, J.S., Brower, J.O. (2003) A new turn structure for the formation of beta-hairpins in peptides, *J. Am. Chem. Soc.* **125**, 876–877.

Olivos, H.J., Alluri, P.G., Reddy, M.M., Salony, D., Kodadek, T. (2002) Microwave-assisted solid-phase synthesis of peptoids, *Org. Lett.* **4**, 4057–4059.

Patch, J.A., Barron, A.E. (2003) Helical peptoid mimics of magainin-2 amide, *J. Am. Chem. Soc.* **125**, 12092–12093.

Seebach, D., Brenner, M., Rueping, M., Schweizer, B., Jaun, B. (2001) Preparation and determination of X-ray-crystal and NMR-solution structures of $\gamma^{2,3,4}$ -peptides, *Chem. Commun.* 207–208.

Seebach, D., Overhand, M., Kühnle, F.N.M., Martinoni, B., Oberer, L., Hommel, U., Widmer, H. (1996) β-Peptides: synthesis by Arndt-Eistert homologation with concomitant peptide coupling. Structure determination by NMR and CD spectroscopy and by X-ray crystallography. Helical secondary structure of a β-hexapeptide in solution and its stability towards pepsin, *Helv. Chim. Acta* **79**, 913–941.

Shankaramma, S.C., Moehle, K., James, S., Vrijbloed, J.W., Obrecht, D., Robinson, J.A. (2003) A family of macrocyclic antibiotics with a mixed peptide-peptoid beta-hairpin backbone conformation, *Chem. Commun.* 1842–1843.

Shankaramma, S.C., Athanassiou, Z., Zerbe, O., Moehle, K., Mouton, C., Bernardini, F., Vrijbloed, J., Obrecht, D., Robinson, J. (2002) Macrocyclic hairpin mimetics of the cationic antimicrobial peptide protegrin I: a new family of broad-spectrum antibiotics, *ChemBioChem.* **3**, 1126–1123.

Shin, I., Park, K. (2002) Solution-phase synthesis of aminooxy peptoids in the C to N and N to C directions, *Org. Lett.* **4**, 869–872.

Shin, I., Lee, M.R., Lee, J., Jung, M., Lee, W., Yoon, J. (2000) Synthesis of optically active D-aminooxy acids from L-amino acids or L-hydroxy acids as building blocks for the preparation of aminooxy peptides, *J. Org. Chem.* **65**, 7667–7675.

Tamilarasu, N., Huq, I., Rana, T.M. (1999) High affinity and specific binding of HIV-1 TAR RNA by a Tat-derived oligourea, *J. Am. Chem. Soc.* **121**, 1597–1598.

Tamilarasu, N., Huq, I., Rana, T.M. (2001) Targeting RNA with peptidomimetic oligomers in human cells, *Bioorg. Med. Chem. Lett.* **11**, 505–507.

Wang, X., Huq, I., Rana, T.M. (1997) HIV-1 TAR RNA recognition by an unnatural biopolymer, *J. Am. Chem. Soc.* **119**, 6444–6445.

Wender, P.A., Mitchell, D.J., Pattabiraman, K., Pelkey, E.T., Steinman, L., Rothbard, J.B. (2000) The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: peptoid molecular transporters, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13003–13008.

Wieczerzak, E., Drabik, P., Lankiewicz, L., Ołdziej, S., Grzonka, Z., Abrahamson, M., Grubb, A., Brömme, D. (2002) Azapeptides structurally based upon inhibitory sites of cystatins as potent and selective inhibitors of cysteine proteases, *J. Med. Chem.* **45**, 4202–4211.

Wilson, M.E., Nowick, J.S. (1998) An efficient synthesis of N,N′-linked oligoureas, *Tetrahedron Lett.* **39**, 6613–6616.

Woll, M.G., Lai, J.R., Guzei, I.A., Taylor, S.J.C., Smith, M.E.B., Gellman, S.H. (2001) Parallel sheet secondary structure in γ-peptides, *J. Am. Chem. Soc.* **123**, 11077–11078.

Wu, C.W., Sanborn, T.J., Zuckermann, R.N., Barron, A.E. (2001) Peptoid oligomers with α-chiral, aromatic side chains: effects of chain length on secondary structure, *J. Am. Chem. Soc.* **123**, 2958–2963.

Wu, C.W., Seurynck, S.L., Lee, K.Y.C., Barron, A.E. (2003) Helical peptoid mimics of lung surfactant protein C, *Chem. Biol.* **10**, 1057–1063.

Yang, D., Zhang, Y.-H., Zhu, N.Y. (2002) $\beta^{2,2}$-Aminoxy acids: A new building block for turns and helices, *J. Am. Chem. Soc.* **124**, 9966–9967.

Yang, D., Ng, F.-F., Li, Z.-J., Wu, Y.-D., Chan, K.W.K., Wang, D.-P. (1996) An unusual turn structure in peptides containing α-aminoxy acids, *J. Am. Chem. Soc.* **118**, 9794–9795.

Yang, D., Li, B., Ng, F.F., Yan, Y.L., Qu, J., Wu, Y.D. (2001) Synthesis and characterization of chiral N-O turns induced by α-aminoxy acids, *J. Org. Chem.* **66**, 7303–7312.

Yang, D., Qu, J., Li, B., Ng, F.-F., Wang, X.-C., Cheung, K.-K., Wang, D.-P., Wu, Y.-D. (1999) Novel turns and helices in peptides of chiral α-aminoxy acids, *J. Am. Chem. Soc.* **121**, 589–590.

Yang, D., Zhang, Y.-H., Li, B., Zhang, D.-W., Chan, J.C.Y., Zhu, N.Y., Luo, S.W., Wu, Y.D. (2004a) Effect of side chains on turns and helices in peptides of β^3-aminoxy acids, *J. Am. Chem. Soc.* **126**, 6956–6966.

Yang, D., Zhang, Y.-H., Li, B., Zhang, D.-W. (2004b) Synthesis of chiral β^3-aminoxy peptides, *J. Org. Chem.* **69**, 7577–7581.

Zhao, X., Jua, M.-X., Jiang, X.-K., Wu, L.-Z., Li, Z.-T., Chen, G.-J. (2004) Zipper-featured δ-peptide foldamers driven by donor-acceptor interaction. Design, synthesis and characterization, *J. Org. Chem.* **69**, 270–279.

Synthetic Peptides: Chemistry, Biology, and Drug Design

Tomi K. Sawyer
ARIAD Pharmaceuticals, Cambridge, MA, USA

1 Peptide Chemistry 94
1.1 Peptide Structure and Three-dimensional Properties 94
1.2 Peptide Chemical Synthesis and Molecular Diversity 102

2 Peptide Biology 104
2.1 Peptide Biosynthesis and Metabolism 104
2.2 Peptide Biological Functions and Mechanisms of Action 106

3 Peptide Drug Design 110
3.1 Peptide Pathophysiology and Therapeutic Targets 110
3.2 Peptide, Peptidomimetic, and Nonpeptide Drug Discovery 112

Bibliography 119
Books and Reviews 119
Primary Literature 120

Keywords

Amino Acids
The building blocks of peptides (as well as polypeptides and proteins) are amino acids. Chemically, an amino acid consists of a central carbon (termed *α-carbon* or C-α) to which are attached an amino (NH_2) group, a carboxyl (CO_2H) group, a side chain (generically referred to as an *R group*), and a hydrogen (H) atom. Amino acids differ in terms of their R group, and the three-dimensional chemical structure (or configuration) of amino acids also may be differentiated at the central α-carbon, except where R is hydrogen. With the exception of glycine, the configuration of the 20 naturally occurring amino acids is designated as L (unless specified otherwise).

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

First and Second Messengers
Peptide hormones are considered to be first messengers because they initiate a cellular/tissue response by interacting with cell surface receptors (ligand-specific binding molecules). Peptide (exemplifying a class of ligand) binding then leads to an amplified cascade of signal transduction events that increase or decrease production of specific intracellular (second) messengers. Second messengers include cyclic 3′,5′-adenosine monophosphate (cAMP), cyclic 3′,5′-guanosine monophosphate (cGMP), diacylglycerol (DAG), inositol triphosphate (IP3), and calcium ion (Ca^{2+}). Receptor-coupled effector enzymes or transport proteins (e.g. adenylate or guanylate cyclase, phospholipase, calcium channels) regulate intracellular levels of such second messengers.

Peptidomimetics and Nonpeptides
More complete synthetic tailoring of a peptide to yield analogs having essentially no naturally occurring amino acids or dipeptide substructure became popularized during the past two decades on the basis of the premise that such compounds might be better suited for drug development, being low in molecular mass, metabolically stable, and bioavailable by oral administration. Such radically modified peptides are generically referred to as *peptidomimetics* (including peptoids and pseudopeptides), and such compounds vary significantly in terms of the nature of the chemical template and design strategy used. ''Nonpeptide'' is used here to generically refer to molecules originally derived from screening of natural product sources or chemical files that competitively displace the natural peptide (or synthetic peptide analog) from the native receptor/enzyme target. In contrast to peptidomimetics, which are designed on the basis of chemical information of a peptide lead or its target receptor/enzyme active site, the discovery of nonpeptides from screening strategies may yield compounds that have no obvious chemical similarity to the native peptide (or synthetic peptide analog).

Primary Structure
The primary structure of a peptide is defined as the linear ordered linkage of constituent amino acids from the N terminus to the C terminus. ''Primary sequence,'' although often used, represents incorrect terminology.

Recombinant Peptides
These include naturally occurring peptides as well as analogs containing naturally occurring amino acids. Such peptides may be prepared by biosynthetic methods based on overexpression of a cloned gene to yield high levels of the desired peptide (or precursor) in a host organism or cell line from which the peptide can subsequently be processed and/or purified to homogeneity.

Secondary Structure (Conformation)
The three-dimensional properties of peptides (as well as polypeptides and proteins) are determined by the individual compound's amino acid sequence, the spatial orientation of both its backbone and side chains, and its characteristic specific intramolecular hydrogen-bonding interactions. Peptide conformations include α-helix, β-sheet, and

β- and γ-turns. Secondary structures may be experimentally determined by spectroscopic and/or crystallographic techniques.

Synthetic Peptides

These peptides, prepared by total chemical synthesis or, in some cases, semisynthesis, provide the opportunity for sophisticated structural modification by incorporation of unusual amino acids, dipeptide replacements or surrogates, and cyclization (involving backbone and/or side-chain linkage), among other types of chemical modifications. Synthetic peptide analogs may be chemically modified in a modest or radical manner with respect to backbone and/or side-chain functionalization. Synthetic alterations of a peptide are exemplified by amide substitution with alkylamide, aminomethylene, ketomethylene, or other nonhydrolyzable isosteres. With the exception of the alkylamide modification, such partial transformation of the backbone amide substructure of a peptide results in analogs generally referred to as *pseudopeptides*. Other typical chemical modifications of synthetic peptides are alterations of the backbone hydrocarbon (CH) substructure by alkylation or stereoinversion.

Peptides are molecules composed of amino acids that are chemically bonded in a specific manner by backbone amide linkage (peptide bond) and, in several instances, by side-chain linkage (e.g. disulfide bond). The discovery and chemical determination of biologically active peptides have become well established over the past few decades. Such work originally needed to overcome two particular challenges: (1) peptides generally exist in very low concentrations, making isolation extremely difficult for early scientists who lacked sophisticated technologies for purification and/or characterization of such molecules as have been more recently developed; and (2) since peptides often consist of a complex molecular framework, early scientists faced major difficulty in accurately determining the structure as well as in proving it by independent chemical synthesis (the latter work, especially, requires highly specific and reversible masking of chemically reactive amine, carboxylate, and other side-chain functionalities of each constituent amino acid).

The rapid development of basic and biomedical research on peptides over the past half-century has been predicated on the premise that peptides are promising molecules for the discovery of new drugs that mimic or block their biological properties. Peptides play key roles in the physiology of all living organisms, as exemplified in humans, where peptides regulate growth, metabolism and development, reproduction, cardiovascular homeostasis, central nervous system (CNS) function, gastrointestinal (GI) function, immune system function, and a plethora of other biological activities.

Over recent years, worldwide attention on peptide chemistry and biological research has been the general theme of international scientific conferences (e.g. biennial symposia of the American, Australian, Chinese, European, and Japanese peptide societies; the Peptide Gordon Conference) and scientific journals dedicated

to such research (e.g. *Journal of Peptide Research, Journal of Peptide Science, Peptide Science (Biopolymers), Peptides, Regulatory Peptides, Neuropeptides*, and *Letters in Peptide Science*). The relative importance of peptide research is evident by the number of Nobel prizes awarded to investigators for their contributions to the science, technology, and/or medicine of these important biomolecules.

Some peptide hormones (e.g. insulin, growth hormone, growth hormone–releasing factor) are being produced on a large scale by recombinant biosynthesis technology to treat specific human afflictions. Human diseases or disorders for which peptide-based drug therapy exists include osteoporosis (calcitonin), diabetes (insulin), prostate cancer and endometriosis (gonadotropin-releasing hormone), acromegaly and ulcers (somatostatin), diuresis and hypertension (vasopressin), hypoglycemia (glucagon), and hypothyroidism (thyrotropin-releasing hormone).

On the basis of the development of the aforementioned naturally occurring or synthetically modified peptide hormones, which are currently marketed as prescription medications, there is hope that such chemically altered peptides (hereinafter referred to as *pseudopeptides, peptidomimetics*, or *peptoids*, according to the type of chemical modification) may also prove useful in therapeutic intervention for still other life-threatening or disabling afflictions, such as immunological dysfunction (e.g. AIDS), cancer, neural dysfunction (e.g. Alzheimer's disease), and cardiovascular dysfunction (e.g. hypertension). Furthermore, the past success of peptide sweeteners (e.g. the dipeptide aspartame) may provide impetus to research and development strategies aimed at the discovery of new applications of peptides.

1 Peptide Chemistry

1.1 Peptide Structure and Three-dimensional Properties

Peptides are composed of two or more amino acids linked together by a peptide bond (Fig. 1). The number of amino acids determine whether the peptide is referred to as a *dipeptide, tripeptide, tetrapeptide*, and so forth. Of the 20 or so genetically coded amino acids, naturally occurring peptides are generally composed of L-stereoisomers (except for Gly). Peptides, polypeptides, or proteins may be structurally differentiated on the basis of the number of constitutive amino acids (see Table 1). For example, the term "peptide" is frequently used to categorize such compounds ranging from 2 to 20 amino acids. Similarly, "polypeptides" often designate compounds ranging from 20 to 50 amino acids, and "proteins" is used for such molecules of more than 50 amino acids (see Table 2). Although a more precise definition may take into consideration other factors (e.g. three-dimensional chemical complexity or biological function), there still exist a substantial amount of personal preference among experts in this field of research. In this article, the generic term "peptide" is used to designate all such chemical agents or messengers.

The precise linear arrangement, amino → carboxy (N → C) directionality, of amino acids in a peptide is referred to as

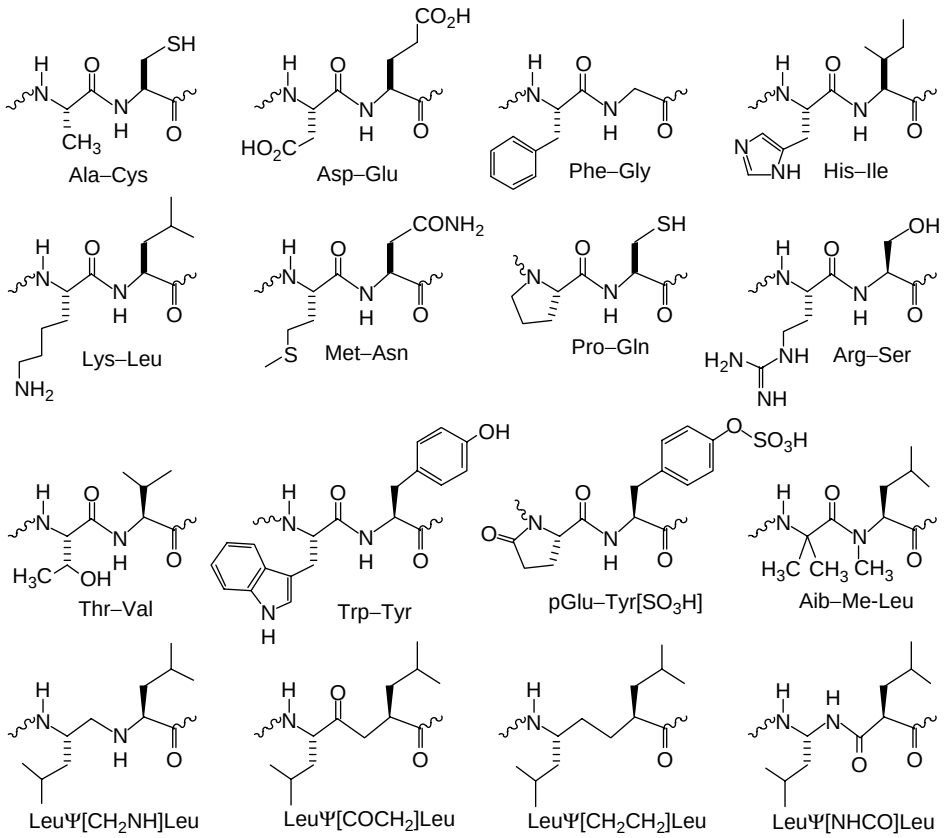

Fig. 1 Chemical structures of dipeptides composed of 20 genetically coded amino acids and some synthetically modified analogs.

its primary structure (Table 3). The three-dimensional folding of the peptide onto itself through covalent (e.g. S−S, disulfide) or noncovalent (e.g. hydrogen and/or ionic) bonding determines the secondary structure of the peptide. Furthermore, the covalent linkage between two or more cysteine residues through disulfide bond formation dictates the transformation of linear peptides to cyclic (or multicyclic) peptides. Such chemically defined restriction of the molecular flexibility of linear peptides does frequently lead toward higher ordered three-dimensional structural complexity and defined spatial arrangement of constitutive amino acid side chains (i.e. tertiary structure). For example, the secondary structure of neuropeptide-Y has been proposed to include a poly-Pro-type helix at the N terminus and an α-helix within its central fragment sequence and a conformationally flexible C terminus. Contributing to such a secondary structure is the tendency of peptides to adopt spatial orientations of both the backbone and side-chain functionalities to bring hydrophobic amino acids (e.g. Trp, Tyr, Phe, Leu, Ile, Val, Met) into proximity to form a hydrophobic surface. Similarly, residues of hydrophilic amino acids (e.g. Lys, Arg,

Tab. 1 Structure determinations of the first 100 peptide hormone, neurotransmitters, growth factors, and cytokines.

Year	*Peptide*
1951	Oxytocin
1953	Insulin
1954	Adrenocorticotropin, vasopressin
1956	β-Melanotropin, angiotensin
1957	α-Melanotropin, glucagon
1960	Bradykinin
1962	Eledoisin
1964	Gastrin, physalaemin
1965	β-Lipotropin
1966	Secretin, somatotropin
1968	Calcitonin, cholecystokinin, C-peptide
1969	Thyrotropin-releasing hormone, nerve growth factor, prolactin
1971	Chorionic gonadotropin, vasoactive intestinal peptide, gastric inhibitory peptide, lutropin, lutropin-releasing hormone, substance P, bombesin, neurophysin
1972	Follitropin, epidermal growth factor
1973	Motilin, neurotensin, somatostatin, tuftsin
1975	Enkephalin, pancreatic polypeptide, thymopoietin
1976	Delta sleep-inducing peptide, β-endorphin, insulin-like growth factor
1977	Relaxin, granuloliberin
1978	Fibroblast growth factor
1979	Gastrin-releasing peptide, glicentin, α/β-neoendorphin, kyotorphin, thymosin
1980	Interferon-α/β
1981	Corticotropin-releasing factor, dynorphin, oxytomodulin, peptide-HI, calcitonin-gene-related peptide, erythropoietin, sauvagine
1982	Growth hormone-releasing factor, neuropeptide-Y, peptide YY, rimorphin, katacalcin, urotensin, interferon-γ
1983	Neuromedin B, neuromedin K, galanin, melanin-concentrating hormone, atriopeptide, platelet-derived growth factor, transforming growth factor-α/β, interleukin 2
1984	Tumor necrosis factor-α/β, cerebellin, interleukin-3
1985	Valosin, granulocyte/macrophage colony-stimulating factor, interleukin-1α/β, neuropeptide K, neutrophil peptide
1986	Cyclinopeptin, leukopyrokinin, leucokinin, leukomyosuppressin, pancreastatin, galanin-associated peptide, interleukin-6
1987	Follicular gonadotropin-releasing peptide, cardioactive peptide, interleukin-4, interleukin-5, interleukin-8
1988	Corticostatin, interleukin 7, neuropeptide-Y, γ-endothelin

Source: Adapted from Eberle, A.N. (1991) *Chimia* **45**, 145.

Glu, Asp, Thr, Ser, Gln, Asn, His) may be distributed within such three-dimensional structures to form a hydrophilic surface that may provide a high degree of solvation by water. In the case of peptides such as growth hormone–releasing factor (GRF) and glucagon, an α-helical-type secondary structure gives rise to amphiphilicity (one side-chain surface being hydrophobic and the other hydrophilic). Furthermore, peptides may be composed of two amino acid chains, or subunits, forming a so-called

Tab. 2 Physiological actions of some regulatory peptides.

Peptide (abbreviation)	*Amino acids*[a]	*Principal physiological actions*[b]
Angiotensin II (AII)	8	↑Vasoconstriction, ↑aldosterone secretion, ↑dipsogenesis (i.e. thirst)
Bradykinin	9	↑Smooth muscle contraction, ↑vasodilation, ↑inflammatory algesia
Calcitonin	32	↓Blood Ca^{2+}
Cholecystokinin (CCK)	8	↑Pancreatic enzyme and electrolyte secretion, ↑gallbladder contraction, ↑satiety, ↑secretion of insulin and glucagons
Adrenocorticotropin (ACTH)	39	↑Adrenal steroidogenesis
Corticotropin-releasing hormone (CRF)	41	↑Corticotropin secretion
Dynorphin	17	↑Sedative analgesia, ↑feeding behavior
β-Endorphin	31	↑Opiate-like activity, ↑central analgesia, ↑respiratory depression, ↑euphoria
Leu-enkephalin	5	↑Spinal analgesia, ↑emotional effects
Epidermal growth factor (EGF)	53	↑Epidermal growth and keratinization,
Follicle-stimulating hormone (FSH)	230	↑Spermatogenesis in males, ↑ovarian follicle growth and estradiol synthesis in females
Gastrin-17	17	↑Pepsin, gastric acid–HCl and pancreatic enzyme secretion, ↑intestine smooth muscle contraction
Glucagon	29	↑Blood glucose, ↑gluconeogenesis, ↑Glycogenolysis
Gonadotropin-releasing hormone (GnRH)	10	↑Luteinizing hormone and follicle-stimulating hormone secretion
Insulin	51	↓Blood glucose, ↑gluconeogenesis, ↑glycogenesis
Luteinizing hormone (LH)	210	↑Testicular androgen synthesis in males, ↑Ovarian estradiol and progesterone synthesis in females
α-Melanotropin (MSH)	13	↑Melanin biosynthesis
Motilin	22	↑Gastrointestinal villous motility

(*continued overleaf*)

Tab. 2 (*Continued*)

Peptide (abbreviation)	*Amino acids*[a]	*Principal physiological actions*[b]
Nerve growth factor (NGF)	118	↑Sympathetic neurite development
Oxytocin	9	↑Milk ejection, ↑uterine contraction
Parathyroid hormone (PTH)	84	↑Blood calcium
Prolactin	199	↑Milk synthesis, ↑mammary gland development
Secretin	27	↑Pancreatic secretion of H_2O, bicarbonate, electrolyte; ↑stomach secretion of gastrin
Somatomedins	50–80	↑Peripheral nervous system growth and development
Somatostatin-14 (SRIF)	14	↓Somatotropin, thyrotropin, parathyroid hormone, calcitonin, renin, and gastric acid secretion
Somatotropin (growth hormone, GH)	191	↑Hepatic somatomedin synthesis
Growth hormone-releasing factor (GRF)	44	↑Somatotropin secretion
Substance P (NK1)	11	↑Peripheral nervous system pain transmission, ↓central nervous system pain transmission
α_1-Thymosin	28	↑Lymphocyte proliferation and differentiation
Thyrotropin (TSH)	201	↑Thyroid hormone T3 and T4 synthesis and secretion
Thyrotropin-releasing hormone (TRH)	3	↑Thyrotropin and prolactin secretion
Vasoactive intestinal peptide (VIP)	28	↑Vasodilation, ↑bronchodilation, ↑glucagon and insulin secretion, ↑glycogenolysis, ↑lipolysis
Vasopressin	9	↑Renal H_2O adsorption, ↑vasoconstriction

[a]Prevalent form of peptide in humans.
[b]Other biological activities have been described (only principal actions are summarized here).
Source: Adapted from Sawyer, T.K., Smith, C.W. (1990) in: Dulbecco, R. (Ed.), *Encyclopedia of Human Biology*, Vol. 5, Academic Press, Harcourt Brace Jovanovich Publishers, San Diego, CA, pp. 725–735.

Tab. 3 Primary structures of some naturally occurring peptides and synthetic analogs thereof.

Peptide	*Primary structure*[a]
Aspartame	Asp[1]-Phe[2]-OMe
Thyrotropin-releasing hormone	<Glu[1]-His-Pro[3]-NH_2
FMRF-amide	Phe[1]-Met-Arg-Phe[4]-NH_2
Enkephalin (Met)	Tyr [1]-Gly-Gly-Phe-Met[5]
DPDPE	Tyr1-D -Pen-Gly-Phe-D-Pen[5] *Pen, penicillamine
BQ-123	cyclo(Leu D-Trp-D-Asp-Pro-D-Val)
L-365,209	cyclo(Pro-D-Phe-Ile-D-Dhp-Dhp- D-MePhe) *Dhp, dehydropiperazyl
GHRP	His[1]-D-Trp-Ala-Trp-D-Phe-Lys[6]-NH_2
Ebiratide	Met(O_2)[1]-Glu-His-Phe-D-Lys-Trp[6]-NH-$(CH_2)_8$-NH_2 *Met(O_2), methionine sulfone
Cholecystokinin 8	Asp-Tyr[SO3H]-Met-Gly-Trp-Met-Asp-Phe[8]-NH_2
Angiotensin II	Asp1-Arg-Val-Tyr-Ile-His-Pro-Phe8
Sandostatin™	D-Phe[1]-Cys-Phe-D-Trp-Lys-Thr-Cys-Thr[8]-ol *Thr-ol, threoninol
Oxytocin	Cys[1]-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly[9]-NH_2
Vasopressin	Cys[1]-Tyr-Phe-Gln-Asn-Cys-Pro-Arg-Gly[9]-NH_2
Desmopressin™	Mpa[1]-Tyr-Phe-Gln-Asn-Cys-Pro-D-Arg-Gly[9]-NH_2 *Mpa, mercaptoproprionic acid
Bradykinin	Arg[1]-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg[9]

(*continued overleaf*)

Tab. 3 (*Continued*)

Peptide	***Primary structure***[a]
HOE-140	Arg^{0}-D-Arg^{1}-Lys-Arg-Hyp-Thi-Ser-D-Tic-Oic-Arg^{9} *Hyp, 4-hydroxyproline; Thi, thienylalanine; Oic, octahydroindole-2-carboxylic acid
Lutropin-releasing hormone	<Glu^{1}-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly^{10}-NH_2
Buserelin	<Glu^{1}-His-Trp-Ser-Tyr-D-Ser(tBu)-Leu-Arg-Pro^{9}-NH-Et
MDL-28050	Suc-Tyr^{1}-Glu-Pro-Pro-Glu-Glu-Tyr-Ala-Cha-Gln^{10} *Suc, succinyl; Cha, cyclohexyalanine
Substance P	Arg^{1}-Pro-Lys-Pro-Gln-Gln-Phe-Phe-Gly-Leu-Met^{11}-NH_2
Cyclosporin A	cyclo(MeLeu-MeLeu-MeThr[4 *R*-4(*E*-2-butenyl)-4-methyl]-Abu-Sar-MeLeu-Leu-MeLeu-Ala-D-Ala-MeLeu)
α-Melanotropin	Ac-Ser^{1}-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly-Lys-Pro-Val^{13}-NH_2
Melanotan-I (NDP-MSH)	Ac-Ser^{1}-Tyr-Ser-Nle-Glu-His-D-Phe-Arg-Trp-Gly-Lys-Pro-Val^{13}-NH_2 *Nle, norleucine
Neurotensin	Glu^{1}-Leu-Tyr-Glu-Asn-Lys-Pro-Arg-Arg-Pro-Tyr-Ile-Leu^{13}
Somatostatin	Ala^{1}-Gly-Cys-Lys-Asn-Phe-Phe-Trp-Lys-Thr-Phe-Thr-Ser-Cys^{14}
Endothelin	Cys^{1}-Ser-Cys-Ser-Ser-Leu-Met-Asp-Lys-Glu-Cys-Val-Tyr-Phe-Cys-His-Leu-Asp-Ile-Ile-Trp^{21}
Glucagon	His^{1}-Ser-Gln-Gly-Thr-Phe-Thr-Ser-Asp-Tyr-Ser-Lys-Tyr-Leu-Asp-Ser-Arg-Arg-Ala-Gln-Asp-Phe-Val-Gln-Trp-Leu-Met-Asp-Thr^{29}
Galanin	Gly^{1}-Trp-Thr-Leu-Asn-Ser-Ala-Gly-Tyr-Leu-Leu-Gly-Pro-His-Ala-Ile-Asp-Asn-His-Arg-Ser-Phe-His-Asp-Lys-Tyr-Gly-Leu-Ala^{29}-NH_2

Calcitonin	Cys[1]-Gly-Asn-Leu-Ser-Thr-Cys-Met-Leu-Gly-Thr-Tyr-Thr-Gln-Asp-Phe-Asn-Lys-Phe-His-Thr-Phe-Pro-Gln-Thr-Ala-Ile-Gly-Val-Gly-Ala-Pro[33]-NH_2
Neuropeptide-Y	Tyr[1]-Pro-Ser-Lys-Pro-Asp-Asn-Pro-Gly-Glu-Asp-Ala-Pro-Ala-Glu-Asp-Leu-Ala-Arg-Tyr-Tyr-Ser-Ala-Leu-Arg-His-Tyr-Ile-Asn-Leu-Met-Thr-Arg-Gln-Arg-Tyr[36]-NH_2
Adrenocorticotropin	Ser[1]-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly-Lys-Pro-Val-Gly-Lys-Lys-Arg-Arg-Pro-Val-Lys-Val-Tyr-Pro-Asn-Gly-Ala-Glu-Asp-Glu-Ser-Ala-Glu-Ala-Phe-Pro-Leu-Glu-Phe[39]
Corticotropin-releasing factor	Ser[1]-Glu-Glu-Pro-Pro-Ile-Ser-Leu-Asp-Leu-Thr-Phe-His-Leu-Leu-Arg-Glu-Val-Leu-Glu-Met-Ala-Arg-Ala-Glu-Gln-Leu-Ala-Gln-Gln-Ala-His-Ser-Asn-Arg-Lys-Leu-Leu-Asp-Ile-Ile[41]-NH_2
Growth hormone releasing factor	Tyr[1]-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Met-Ser-Arg-Gln-Gln-Gly-Glu-Ser-Asn-Gln-Glu-Arg-Gly-Ala-Arg-Ala-Arg-Leu[44]-NH_2
Insulin	Gly[1]-Ile-Val-Glu-Gln-Cys-Cys-Thr-Ser-Ile-Cys-Ser-Leu-Tyr-Gln-Leu-Glu-Asn-Tyr-Cys-Asn- *(A-chain above; B-chain below)* Phe[1]-Val-Asn-Gln-His-Leu-Cys-Gly-Ser-His-Leu-Val-Glu-Ala-Leu-Tyr-Leu-Val-Cys-Gly-Glu-Arg-Gly-Phe-Phe-Tyr-Thr-Pro-Lys-Thr[31]
Parathyroid hormone	Ser[1]-Val-Ser-Glu-Ile-Gln-Leu-Met-His-Asn-Leu-Gly-Lys-His-Leu-Asn-Ser-Met-Glu-Arg-Val-Glu-Trp-Leu-Arg-Lys-Lys-Leu-Gln-Asp-Val-His-Asn-Phe-Val-Ala-Leu-Gly-Ala-Pro-Leu-Ala-Pro-Arg-Asp-Ala-Gly-Ser-Gln-Arg-Pro-Arg-Lys-Lys-Glu-Asp-Asn-Val-Leu-Val-Glu-Ser-His-Glu-Lys-Ser-Leu-Gly-Glu-Ala-Asp-Lys-Ala-Asp-Val-Asp-Val-Leu-Thr-Lys-Ala-Lys-Ser-Gln[84]

[a]Asterisks indicate abbreviations commonly found in the literature.

dimeric structure composed of either homologous or heterologous subunits. The individual subunits may be either covalently (e.g. disulfide) or noncovalently (e.g. ionic or hydrophobic) bonded together. Both interchain and intrachain disulfide bonds may be present within a peptide (e.g. insulin, Table 3). In some cases, such a complex chemical species may be required to establish the biologically active form of a peptide (e.g. nerve growth factor dimer).

The three-dimensional substructure of peptides may be further described in terms of the torsion angles between the backbone amine nitrogen (N-α), backbone carbonyl carbon (C′), backbone hydrocarbon (C-α), and side-chain hydrocarbon functionalization (e.g. C-β, C-γ, C-δ, C-ε of Lys) as depending on the amino acid sequence (Fig. 2). The torsion angle nomenclature is exemplified by the following cases: Ψ, N-α−C-α−C′−N-α; ω, C-α−C′−N-α−C-α; Φ, C′−N-α−C-α−C′; χ_1 N-α−C-α−C-β−X (where X is not hydrogen); χ_2, C-α−C-β−C-γ−X. A Ramachandran plot of Ψ versus Φ for peptides possessing intrinsic secondary structure has further indicated that particular combinations of torsion angles for a helical, reverse-turn, or extended conformation do exist in a predominant fashion. For amide bond torsion angle (ω) the trans geometry is preferred for most dipeptide substructures; when Pro is the C-terminal partner, however, the cis geometry is possible. It is noted that the three-dimensional structural flexibility is directly related to covalent and/or noncovalent bonding interactions within the amino acid sequence of a particular peptide, and synthetic modifications such as N-α alkylation and C-α or C-β alkylation effect significant conformational constraints locally as related to backbone and/or side-chain torsion angles (e.g. conformationally modified Phe analogs, Fig. 2). Finally, it is noted that backbone amide replacements may be defined in terms of a nomenclature system, in which the functional group substitution of the dipeptide substructure is identified as a Ψ[surrogate] (e.g. substitution of the aminomethylene surrogate in Leu-Val would be described as LeuΨ[CH_2NH]Val).

1.2 Peptide Chemical Synthesis and Molecular Diversity

The chemical synthesis of peptides and peptide libraries has advanced tremendously over the past century relative to the pioneering efforts of many scientists (including, in some cases, Nobel laureates) such as George Barany, Max Bergmann, Miklos Bodanszky, Louis Carpino, Theodor Curtius, Emil Fischer, Murray Goodman, Mario Geysen, Richard Houghten, Victor Hruby, Stephen Kent, R. Bruce Merrifield, Stanford Moore, Shumpei Sakakibara, William Stein, James Tam, Claudio Toniolo, Vincent du Vigneaud, and Theodor Wieland. In retrospect, the field of peptide chemistry is enriched with a phenomenal record of achievements involving the solution- and solid-phase synthesis of complex peptides, purification and analytical characterization of peptides, and more recently, the development and implementation of specialized technologies that expedite peptide-based drug discovery in diverse areas of research (e.g. immunology; cardiovascular and neurological physiology). The scope of peptide synthesis pervades both naturally occurring peptides and their analogs, as exemplified by a variety of possible side-chain and/or backbone substitutions that have been developed to exploit conformational modeling hypotheses (local or global

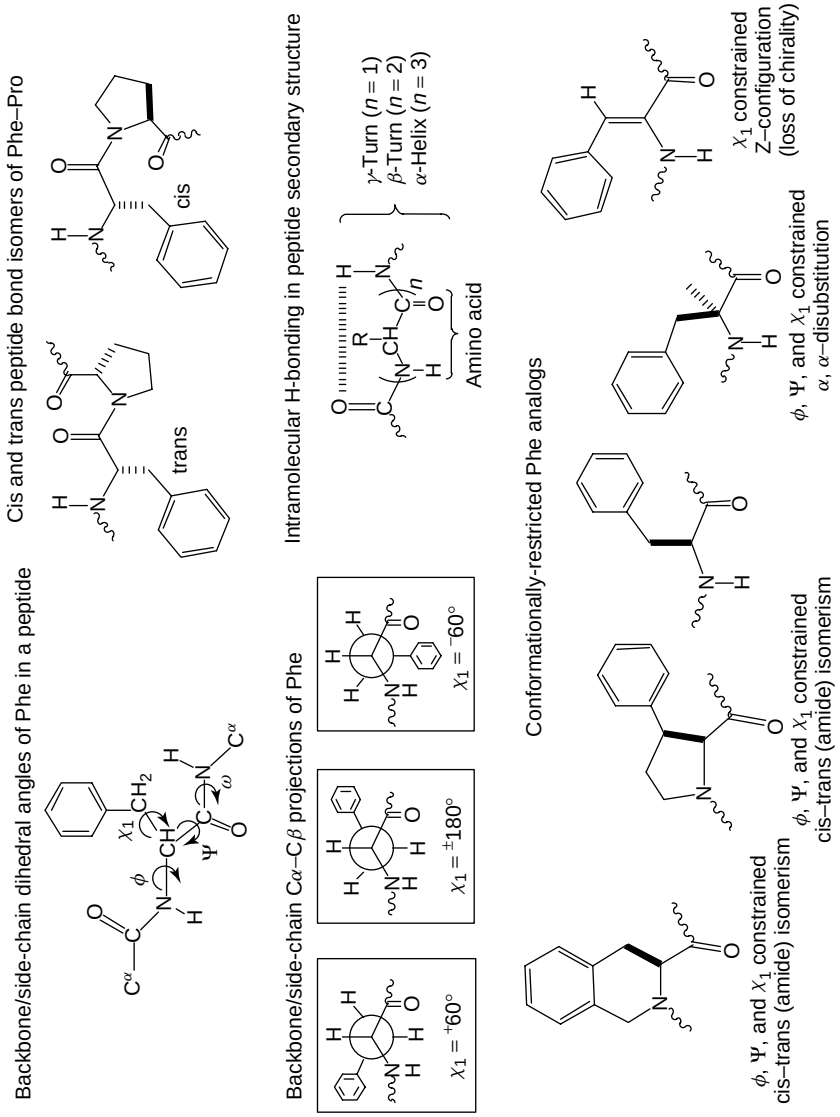

Fig. 2 Three-dimensional structural properties of peptides.

three-dimensional structural constraints), molecular recognition and mechanistic properties at biological targets (receptors, enzymes, antibodies, nucleic acids), and/or metabolism by processing or degrading peptidases. The chemical synthesis of peptides has become a rapidly developed scientific art owing to such factors as automated solid-phase methods, orthogonal protection–cleavage strategies, and a sophisticated database of literature that details possible reagents and potential side reactions to further optimize the task of preparing simple and/or complex peptides. The synthesis of large-sized and/or multicyclic peptides such as interleukin-3, conotoxin, endothelin, transforming growth factor, corticotropin-releasing factor, and calcitonin has been achieved in many research laboratories, and contemporary peptide chemistry methods have been successfully applied to the synthesis of yet more complex macromolecular targets (e.g. HIV protease dimer). Most notably, the chemical synthesis of peptides has expanded to the preparation of combinatorial peptide libraries for which high volume target screening (e.g. receptor binding or enzyme inhibition) is performed to identify compound(s) that possess biological activity. This generally requires the interface of two specialized technologies, namely, synthesis of combinatorial peptide libraries, and their target screening. As opposed to focused analysis of the native peptide with respect to biological activity, such techniques provide tremendous molecular diversity of potential peptide lead compounds to study molecular recognition and mechanistic aspects of peptide interaction with protein (receptors, enzymes), nucleic acid, or carbohydrate targets. Similarly, phage-based peptide libraries have been developed for such applications. The principal advantage of the synthetic peptide or peptidomimetic libraries is the extraordinary molecular diversity of amino acid building blocks that may include unnatural amino acids, pseudodipeptides, and nonpeptide template intermediates.

2 Peptide Biology

2.1 Peptide Biosynthesis and Metabolism

Peptides are biosynthesized on ribosomes, where their specific amino acid sequence is determined (translated) by a specific messenger RNA sequence (codon triplet codes for an amino acid). The nucleotide sequences of the RNA are dictated (transcribed) from specific chromosomal deoxyribonucleotide sequences (DNA genes). Cellular biosynthesis of peptides is believed to proceed very specifically, with error rates of less than 1 in 104 amino acids at a rate of 20 amino acids/ second. The nascent peptides are subsequently released and transported into the cisternae of the rough endoplasmic reticulum and then to the Golgi elements, where they may be posttranslationally modified (e.g. sulfated, glycosylated). Vesicles containing the peptide are pinched off the terminal cisternae of the Golgi apparatus, and then they are targeted to intracellular organelles, the plasma membrane, or are secreted into the extracellular space. In some peptides, the tyrosine residues may be posttranslationally sulfated (Tyr[SO_3H], Fig. 1) as exemplified by the GI peptide hormones cholecystokinin-8 and gastrin. If glutamic acid is present, it may be cyclized into a pyroglutamic acid derivative (<Glu, Fig. 1) as exemplified by thyrotropin-releasing

factor, a hypothalamic peptide. Peptides vary with regard to whether their C terminus is a free carboxylic acid (i.e. $-CO_2H$) or carboxamidated (i.e. $-CONH_2$) and whether the N terminus is a free amine (H_2N-), acetylated (CH_3CONH-), formylated ($HCONH-$), or bearing some other modified amine group (Fig. 1). Some peptides are glycosylated; that is, they are conjugated to one or more carbohydrate groups (e.g. sialic acid). These various examples of posttranslational functionalization are often directly linked to the biological activities of the parent peptides as related to the "bioactive" conformation and/or "message" sequence of the peptide. Such posttranslational "tag"î also serve as biomolecular zip codes to target the peptide to a particular cellular compartment. Finally, it is important to note that the amino acid sequence itself may differ to varying degrees among species. This information is useful in determining the evolutionary relatedness between species. Two or more isoforms of a peptide may also exist within an individual species of animal. These isoforms, although chemically similar, may differ either in primary structure length (e.g. cholecystokinin family of CCK-51, CCK-39, CCK-33, CCK-12, and CCK-8 peptides), or in other chemical aspects (e.g. site-specific amino acid substitutions and/or side-chain modification by glycosylation, sulfation, etc.). Gene duplication followed by single or multiple nucleic acid mutations most likely account for such variations in peptide chemical structure and subsequent evolution of various families of peptides (e.g. the neurohypophyseal hormones such as oxytocin and vasopressin).

Some peptides (e.g. oxytocin, vasopressin) require extensive posttranslational processing because they are not coded directly by DNA. Frequently, N- and/or C-terminally extended amino acid sequences are initially biosynthesized. These propeptides may then be packaged within secretory vesicles along with proteolytic enzymes. Both endopeptidases and exopeptidases (Table 4) can contribute to appropriate processing of the inactive precursor peptide to yield the active peptide. Therefore, peptides may be derived indirectly by way of a propeptide, which itself may be derived from a prepropeptide. For example, proopiomelanocortin is a large inactive peptide (MW 28 500) that is cleaved by specific peptidases into several active peptides, including adrenocorticotropin

Tab. 4 Some examples of endopeptidases and exopeptidases.

Aspartyl class	*Cysteinyl class*	*Metallo class*	*Serinyl class*
Pepsin	Cathepsin B	Peptidyl dipeptidase A	Thrombin
Renin	Cathepsin H	Collagenase	Trypsin
Cathepsin D	Cathepsin L	Endopeptidase 24.11	Chymotrypsin A
Cathepsin E	Cathepsin S	Aminopeptidase M	Elastase
HIV protease	Cathepsin M	Carboxypeptidase A	Kallikrein
	Cathepsin N	Stromolysin	Cathepsin A
	Cathepsin T	Gelatinase A	Cathepsin G
	Calpains	Gelatinase B	Cathepsin R
	Papain		Tissue plasminogen activator (TPA)
	Proline endopeptidase		

(ACTH), β-endorphin, α-melanotropin, and β-melanotropin. Some large plasma proteins also serve as propeptides for peptide production. Renin, an enzyme released from the juxtaglomerular cells of the kidneys, acts on a liver-borne substrate protein (angiotensinogen) to convert it to the decapeptide angiotensin I, which is then processed by another enzyme to yield the active peptide hormone angiotensin II. Other precursor plasma proteins, the kininogens, are similarly converted by specific serine proteases (kallikreins) to kinins, such as bradykinin, which is an important peptide hormone in regulating blood flow in certain vascular beds. Bradykinin and angiotensin II are examples of peptide hormones that are released from liver cells as larger propeptides to be converted into active hormonal peptides within the blood. In contrast, peptidases may also inactivate peptides by splitting the molecules at specific internal peptide bonds. Exopeptidases, both carboxypeptidases and aminopeptidases, cleave off the C-terminal or N-terminal amino acids, respectively. Some peptides may be inactivated by simple deamidation at the C-terminal (if amidated) end of the molecule. Endopeptidase cleavage is generally quite specific, and both exopeptidases and endopeptidases (Table 4) are well characterized in terms of mechanistic properties and substrate specificity. Most importantly, it is well known that the clinical use of some peptides has been limited significantly by their short plasma half-life, which may in part be related to their lability to peptidases. Therefore, knowledge of biodegradation mechanisms has been considered to be an important research objective to facilitate the design of synthetic peptide analogs that may exhibit sustained biological activities (see Sect. 3.2). Some peptides may also undergo inactivation by other chemical transformations such as that exemplified by insulin, in which its cystine-bridged heterodimeric structure is liable to cleavage by means of reduction of the interchain disulfide bonds by the enzyme insulinase. Considerable effort has been devoted to the objective of designing synthetic peptide analogs (including pseudopeptides) and peptidomimetics, which are structurally altered to compromise or eliminate biological cleavage–inactivation by peptidases. Alternatively, there has also been considerable effort devoted to directly inhibiting either processing (cleavage activation) or degrading peptidases to modify the generation or the lifetime of a particular endogenous (or exogenous) peptide, respectively. In particular, the historically important discoveries of peptide and, subsequently, peptidomimetic inhibitors of angiotensin-converting enzyme (ACE) (i.e. teprotide and captopril: see Sect. 3.2, Fig. 4) deserve particular mention because they provided tremendous incentive to rational drug design and natural product lead-finding technologies. Such peptidase targets exemplify processing enzymes, which are required for the generation of endogenous peptides, and inhibition of such peptidases compromises the agonist activity of a particular peptide.

2.2 Peptide Biological Functions and Mechanisms of Action

Peptides serve diverse physiological roles, and their biological functions and mechanisms of action have been the subject of intensive molecular and cell biology, biochemistry and pharmacology research, as exemplified by the pioneering efforts of many scientists (including, in some cases, Nobel laureates) such as Frederick

Banting, Stanley Cohen, Pedro Cuatrecasas, Charles Deber, Lila Gierasch, Alfred Gilman, Roger Guillemin, Mac Hadley, Rita Levi-Montalcini, Frank Porreca, Jean Rivier, Martin Rodbell, Alan Saltiel, Andrew Schally, Solomon Snyder, Donald Steiner, Wylie Vale, and Henry Yamamura. Glutathione is a tripeptide that serves as a cofactor for one or more enzyme systems as related to its oxidation–reduction chemical properties. Ion-binding peptides also constitute a category of transporting molecules (ionophores) as exemplified by valinomycin (Table 3). Ion channel–binding peptides have been identified from natural sources such as venoms, and such peptides may selectively regulate sodium (Na^+), potassium (K^+), or calcium ion (Ca^{2+}) transport. Furthermore, the relationship between calcium ion and peptides extends to the finding that some peptide hormones (α-melanotropin and adrenocorticotropin) apparently require this divalent metal ion to bind and/or effect receptor-mediated signal transduction. Antibiotic peptides include gramicidin (Table 3), actinomycin D, bacitracin, penicillin, and defensins. Peptide toxins include agatoxin, α-bungarotoxin, and ricin. Finally, it is well recognized that a vast number of peptide chemical messengers exist, which manifest their biological activities as regulatory peptide hormones, neurotransmitters, growth factors, and cytokines.

Hormones, which include peptides, steroids, catecholamines, prostaglandins, and related naturally occurring compounds, are chemical messengers released by one cell to act on one or more other cell types to elicit a physiological response. The largest category of hormones consists of peptides, and the biological properties of such peptide hormones is extensive, with well-established examples related to endocrine, cardiovascular, neural, immune, and gastrointestinal system regulation. In terms of functional properties, peptide hormones regulate differentiation, growth, reproduction, blood pressure, glucose homeostasis, and behavior, in addition to performing many other regulatory activities. Peptide hormones normally exist at very low concentrations (10^{-12}–10^{-9} M) in bodily fluids or tissues, in which they persist for short periods of time (1–30 min), since they are rapidly inactivated by peptidases and/or excreted (i.e. cleared) from the body. The transitory actions of peptide hormones are well suited for continuous regulation (i.e. homeostatis) of various physiological systems.

Some of the examples of peptide hormones and their particular regulatory functions summarized earlier (Table 2) are amplified briefly here. Thyroid-stimulating hormone (TSH), which acts on the thyroid gland, is a glycoprotein, as are lutropin and follitropin, which act on the gonads (testes and ovaries). Somatotropin (growth hormone, GH) is necessary for normal growth, and defective or excessive secretion can lead to dwarfism or acromegaly, respectively. Prolactin also possesses growth-promoting activity, most specifically, mammary growth and milk production. All the pituitary peptide hormones are individually controlled by peptide factors (hormones) of hypothalamic origin. These hypophysiotropic peptides either stimulate or inhibit pituitary hormone secretion. The peptide hormones glucagon and insulin, both of pancreatic origin, are responsible for elevating or depressing glucose levels through their actions on the liver and other organs. Glucose homeostasis is therefore very tightly regulated by these two counter-regulatory peptides. Similarly, parathyroid

hormone (PTH) and calcitonin, derived from the parathyroid and thyroid glands respectively, act in a counterregulatory manner to maintain serum Ca^{2+} levels. Specifically, PTH elevates and calcitonin lowers blood Ca^{2+} levels. The GI peptide hormones regulate the movement, metabolism, digestion, and intestinal absorption of metabolic substrates that are essential for life. Peptide hormones regulate GI smooth muscle motility (and therefore substrate transport) and the pancreatic enzyme secretions that degrade metabolic substrates into absorptive substrates. Gastrin, cholecystokinin (CCK), and vasoactive intestinal peptide (VIP) are examples of a dozen or so recognized peptide hormones that regulate GI, pancreatic, and other processes. A large number of peptide growth factors have been discovered, which play a role in normal as well as neoplastic (cancerous) growth. Factors such as epidermal growth factor (EGF), nerve growth factor (NGF), transforming growth factors, somatomedins, and erythropoietin each affect certain cell types to regulate their normal growth and proliferation. Excess secretion of certain of these peptide growth factors may correlate to neoplastic transformation, tumor growth, and/or metastasis. A most important discovery was that peptides produced by the brain exhibit analgesic activity, as do the opiate drugs (e.g. morphine). This observation led to the discovery of opiate receptors and the demonstration that endogenous morphine-like substances (e.g. Met-enkephalin, β-endorphin, dynorphin) interact with the same receptors to affect analgesia.

Peptide hormones, probably without exception, mediate their initial actions at the level of the cell plasma membrane (phospholipid bilayer). Peptide hormones ("first messengers") interact at the cell surface with protein (or glycoprotein) macromolecular receptors, which bind to the peptide ligand in a very specific manner. This results in activation (signal transduction) of proximally located enzymes or transport proteins, which may be modified in a positive or negative manner to produce within the cell, a "second messenger" (e.g. cAMP, cGMP, IP_3, DAG, Ca^{2+}/K^{+}; Table 5). Alternatively, the peptide–receptor complex may result in kinase-like activities to phosphorylate other cellular signaling proteins such as those exemplified by growth factor receptor kinases. Many peptide hormones and neurotransmitters effect receptor-mediated stimulation or inhibition of adenylate cyclase via intermediary signaling heterotrimeric proteins known as *G proteins*. Recognition domains of such G protein–coupled peptide receptors (typically, a seven transmembrane–spanning helical protein superfamily) exist for both the peptide ligand and the specific G protein. Three-dimensional models are being developed to provide further insight into the structure–activity relationships of these G protein–coupled receptors as related to protein-coupled receptors as related to probing the molecular basis of peptidergic receptor-mediated biological signaling by site-directed mutagenesis. In cases of oncogene activation, the uncontrolled production or mutation of cellular proteins can mimic, for example, a signaling pathway mediated by a growth factor peptide–receptor complex, which can lead to a pathophysiological (malignant carcinogenic) transformation of that particular cell. Examples of oncogenes that can lead to disregulated signal transduction pathways include peptide growth factors (e.g. *sis*), receptors (e.g. *erb B*), transduction proteins (e.g. *ras*), and transcription factors (e.g. *fos, myc*). Binding

Tab. 5 Peptidergic receptor subtypes and signal transduction pathways.

Peptide	***Receptor subtype***	***Signal transduction second messenger***	***Peptide//Peptidomimetic/Nonpeptide (Agonists or antagonists)***
Angiotensin II	AT_1	↓cAMP, ↑IP_3/DAG	DuP753 (nonpeptide antagonist)
	AT_2	Not determined	PD123177 (nonpeptide antagonist)
Bradykinin	B_1	Not determined	BK_{1-8} (peptide agonist) [Leu^8]BK_{1-8} (peptide antagonist)
	B_2	↑IP_3/DAG	[Phe^8 (CH_2NH)Arg^9]BK (pseudopeptide agonist) d-Arg[Hyp^3, Thi^5, D-Tic^7, Oic^8]BK (peptide antagonist)
Cholecystokinin	CCK_A	↑IP_3/DAG	A-71623 (peptide agonist) Devazepide (nonpeptide antagonist)
	CCK_B	Not determined	Gastrin (peptide agonist) CI-988 (peptidomimetic antagonist)
Endothelin	ET_A	↑IP3/DAG	ET-1 (peptide agonist; nonselective) BQ-123 (peptide antagonist)
	ET_B	↑IP_3/DAG	Sarafotoxin S6c (peptide agonist)
Neuropeptide-Y	Y_1	↓cAMP	[Leu^{31}, Pro^{34}]NPY (peptide agonist)
	Y_2	Not determined	NPY_{13-36} (peptide agonist)
Met-enkephalin	δ-Opioid	↓cAMP, ↑K^+ channel	DAMGO (peptide agonist) CTOP (peptide antagonist)
β-Endorphin	μ-Opioid	↓cAMP, ↑K^+ channel	DPDPE (peptide agonist) ICI-174864 (peptide antagonist)
Dynorphin	κ-Opioid	↓Ca^{2+} channel	U-69593 (nonpeptide agonist) Norbinaltorphimine (nonpeptide antagonist)
Substance P	NK_1	↑IP_3/DAG	[Pro^9]SP (peptide agonist) CP-96345 (nonpeptide antagonist)
Substance K (NK_A)	NK_2	↑IP_3/DAG	[Lys^5, $MeLeu^9$, Nle^{10}]NKA (peptide agonist) L-659877 (peptide antagonist)
Neurokinin-B (NK_B)	NK_3	↑IP_3/DAG	Senktide (peptide agonist) [Trp^7, β-Ala^8]NKB_{4-10} (peptide antagonist)
Vasopressin	V_{1A}	↑IP_3/DAG	d(CH_2)$_5$[Tyr(Me)$_2$]AVP (peptide antagonist)
	V_{1B}	↑IP_3/DAG	–
	V_2	↑cAMP	Desamino[D-Arg^8]AVP (peptide agonist) d(CH_2)$_5$[D-Ile^2, Ile^4]AVP (peptide antagonist)

Source: Adapted from Watson and Abbott, *Trends in Pharmacological Sciences* (TiPS Receptor Nomenclature Supplement – January 1992).

of a peptide to a receptor (or enzyme) results from a multiplicity of noncovalent intermolecular interactions, and these events are fundamentally related to structure–activity studies focused on a plethora of biologically active peptides. Binding is considered to precede transformation of the receptor (or enzyme) from its inactive to active state by molecular mechanisms often not understood. In many cases, it has been determined which substructural features of a peptide are required for binding (i.e. the "address" sequence) versus signaling/activation (i.e. the "message" sequence). In the case of adrenocorticotropin, for example, the central 5–24 amino acid sequence of the native peptide (Table 3) is capable of effecting full steroidogenic agonism, whereas further N-terminal truncation (i.e. $ACTH_{8-24}$ or $ACTH_{11-24}$) gives rise to partial agonism or antagonism. Similarly, corticotropin-releasing factor (CRF) appears to require a significant portion of its C-terminal sequence for binding (i.e. residues 12–41), whereas biological activity requires N-terminal extension (i.e. residues 6–41). The three-dimensional structural features of a peptide that provide the essential functional group ensemble for molecular recognition (binding) at a target receptor or enzyme may be composed of backbone and/or side-chain components, and such a three-dimensional substructure of the peptide has been referred to as the *pharmacophore*. Development of both intuitive peptide pharmacophore models and models derived experimentally (biophysical and computational chemistry) is providing an opportunity to explore the molecular recognition and mechanistic properties of peptide interactions with macromolecular targets. Such studies are of particular significance to peptidomimetic drug design.

3 Peptide Drug Design

A large number of pharmaceutical companies and biotechnology firms have successfully advanced the discovery and development of synthetic or recombinant peptides as well as synthetic peptidomimetic or nonpeptide drugs for numerous medical applications. Relative to synthetic peptide, peptidomimetic, and nonpeptide drug discovery, the integration of drug design using NMR spectroscopy, X-ray crystallography, computational chemistry, and other biophysical and *in silico* methods has been critical to provide insight to the intimate structure–activity relationships and related biological properties necessary to create potent, selective, metabolically stable, safe-acting and, in many cases, orally effective molecules. Of course, knowledge of the pathophysiology of biologically active peptides and their therapeutic target (e.g. receptor, peptidase or signal transduction protein) is essential to identify opportunities for peptide (synthetic or recombinant), peptidomimetic and nonpeptide drug discovery.

3.1 Peptide Pathophysiology and Therapeutic Targets

The pathophysiology related to biologically active peptides may be due to defects in the biosynthesis of the peptide or its target receptor (or enzyme), among a number of other factors. Table 6 lists a variety of endocrine disorders that result from excessive or deficient peptide hormone levels. Amino acid mutations of regulatory peptides are known to exist and to provide the molecular basis for pathophysiological conditions requiring replacement therapy. Causative factors in some cases of diabetes

Tab. 6 Pathophysiological states of known relationship to peptides.

Disease (symptom)	*Relationship to peptide*
Addison's disease (abnormal carbohydrate metabolism)	Deficiency in adrenocorticotropin
Cushing's disease (increased protein catabolism)	Excess adrenocorticotropin
Secondary hypogonadism	Deficiency in gonadotropin
Secondary hypergonadism	Excess gonadotropin
Hyperglycemia and glucosuria (diabetes mellitus or insulin-dependent diabetes, if resulting from cytotoxic autoantibodies to β-cells)	Deficiency in insulin
Abnormal blood Ca_2 (increased), hyperparathyroidism	Excess parathyroid hormone
Abnormal growth (decreased), Laron-type dwarfism	Deficiency in somatomedin
Abnormal growth (decreased), hypopituitary dwarfism	Deficiency in somatotropin
Abnormal growth (increased) in children, giantism	Excess somatotropin
Abnormal growth (increased) in adults, acromegaly	Excess somatotropin
Abnormal metabolism (increased), Graves' disease antibodies	Excess thyrotropin-mimicking, anti-TSH receptor
Hypovolemia–dehydration (increased), pituitary-type diabetes insipidus	Deficiency in vasopressin

Source: Adapted from Hadley, M.E. (2000) *Endocrinology*, 5th ed., Prentice Hall, Englewood Cliffs, NJ.

mellitus are defects in the primary structure of insulin, either in the active site of the peptide or at sites of cleavage, where the insulin chains are proteolytically cleaved from the prohormone structure. Some, albeit rare, types of diabetes are also due to defects in the amino acid sequence of the insulin receptor that compromise binding of the peptide hormone to the receptor or receptor-mediated signal transduction. In one form of familial (genetic) hypothyroidism, a mutation in one codon of the gene that transcribes for the β-subunit of the hormone thyrotropin (TSH) results in a single amino acid substitution to yield a conformationally abnormal β-subunit that cannot associate with the α-subunit to form a functionally active heterodimeric peptide hormone. Furthermore, since this particular defect in producing endogenous active TSH is familial, and therefore present at birth (congenital), the consequences (e.g. cretinism) are devastating. Within the scope of mimicking or blocking regulatory peptide effects at the level of target receptors, there exists a tremendous research effort in identifying peptidomimetic or nonpeptide derivatives (agonists or antagonists), and such studies have enabled both pharmacological analysis and, in some cases, the discovery of therapeutic agents of synthetic or recombinant origin (see Sect. 3.2). Furthermore, the design of peptide receptor-targeted agonists or antagonists that may be radiolabeled or conjugated to anticancer drugs has made it possible to use such agents in the diagnosis, localization, or chemotherapy of tumors. For example, radioactive indium attached to a somatostatin analog has been developed to localize certain tumors by external scintography.

Enzyme targets that have been determined to be of particular clinical importance as related to the pathophysiology of regulatory peptides include a number of peptidases (Table 4). Representative peptidase targets may be classified as

follows: serine peptidases, such as thrombin, tissue plasminogen activator (TPA), and elastase; metallopeptidases, such as ACE, collagenase, and enkephalinase; aspartic peptidases, such as renin, cathepsin D, and HIV protease; and cysteine peptidases, such as cathepsin B, and calpains. As described earlier, peptidases may function primarily as processing or degrading enzymes of regulatory peptides. Within the scope of blood pressure regulation, inhibitors of either ACE or renin may serve to reversibly terminate the stepwise processing of the α_2-macroglobulin angiotensinogen to give rise to the vasoconstrictor angiotensin II (AII). In the case of ACE inhibitors, the clinical efficacy of such drugs as antihypertensives has been well established. In the latter case, the discovery of potent peptidomimetic inhibitors of renin has not readily translated into the development of orally active therapeutic candidates to date. However, such research has catalyzed the identification and design of peptidomimetics that inhibit HIV protease, an aspartyl peptidase having structural and mechanistic resemblance to renin, and such potential drugs may enable a novel chemotherapeutic intervention strategy for the threatening spread and fatal pathogenesis related to AIDS. As a separate case, but still related to processing peptidase targets, peptidomimetic inhibitors of thrombin are being designed in the hope that they might be used to prevent thrombus (clot) formation following surgery. Within the realm of degrading peptidases, the discovery of peptidomimetic inhibitors of neutral endopeptidase (NEP) has been a strategy of interest to advance novel analgesic agents that function via inhibiting the cleavage inactivation of endogenous opioid peptides (e.g. enkephalins). Similarly, peptidomimetic inhibitors of collagenase, a collagen-degrading peptidase, are thought to be of potential utility in the discovery of therapeutic agents effective in pathological conditions such as rheumatoid arthritis.

3.2 Peptide, Peptidomimetic, and Nonpeptide Drug Discovery

Peptide, peptidomimetic, and nonpeptide drug discovery is an intriguing area of research efforts in numerous pharmaceutical and biotechnology companies, and interdisciplinary strategies have emerged over the past two decades that have provided the framework to advance lead compounds and, in some cases, drugs (Fig. 3). Noteworthy to such peptide, peptidomimetic, and nonpeptide drug discovery has been the pioneering work of many scientists from both academia and industry, including Richard DiMarchi, Roger Friedinger, Ralph Hirschmann, Victor Hruby, Isabella Karle, Horst Kessler, Garland Marshall, Henry Mosberg, John Nestor, Daniel Rich, Joseph Rudinger, Tomi Sawyer, Peter Schiller, Robert Schwyzer, and Daniel Veber.

The first recombinant peptide drug commercialized was insulin in 1982. Currently, a number of peptide drugs (or preclinical leads) have been advanced using genetic engineering technologies. Examples are somatotropin, tumor necrosis factor, EGF, erythropoietin, hepatitis B vaccine, interferons, various cytokines, and TPA (Table 7). On the basis of the success of insulin replacement therapy for the treatment of some types of diabetes mellitus, the potential for therapeutic application of other structurally complex peptides has been advanced by sophisticated molecular biology methodologies, which permit biosynthesis of peptides not readily obtained by chemical synthesis.

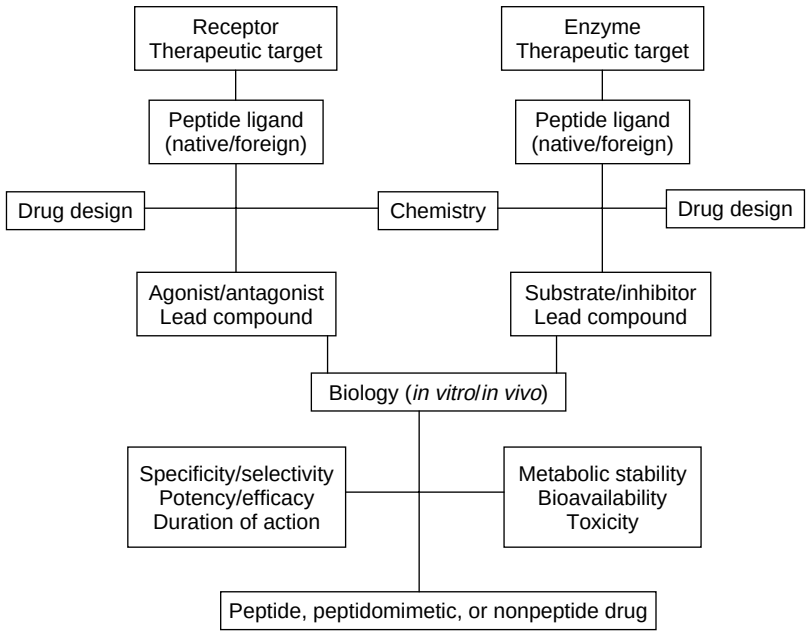

Fig. 3 Interdisciplinary strategies in peptide, peptidomimetic, and nonpeptide drug discovery.

The known chemical diversity of peptide, pseudopeptide, peptidomimetic, and nonpeptide structures that have been advanced as lead compounds or marketed drugs is extensive. Synthetic peptides include native molecules and analogs that incorporate side-chain and/or backbone modifications. In many cases, the resultant synthetic derivatives are more potent, target tissue (receptor) selective, and/or stable to peptidases relative to the native peptide. Such examples include Sandostatin (SRIF analog), Desmopressin (vasopressin analog), Goserelin (GnRH analog), Tetracosactide (corticotropin analog), and Melanotan-I (MSH analog). The discovery of "first-generation" peptide antagonists has, in a majority of cases, been based on chemically modified analogs. Peptide-based competitive antagonists of bradykinin, oxytocin, PTH, enkephalin, vasopressin, glucagon, CCK, gastrin, angiotensin II, GnRH, substance P, and CRF have been reported during the past decade, and such compounds have typically incorporated unusual amino acids (e.g. D-amino acids, cyclic amino acids, α,β-unsaturated amino acids; Fig. 2) and backbone CONH replacements.

The chemical transformation or replacement of peptide agonists or antagonists by simpler (molecular weight-wise), metabolically stable, and orally bioavailable "second-generation" peptidomimetics or nonpeptides has been accomplished (Table 7). For example, research focused on peptide receptor targets has led to the identification of effective antagonists. To some extent, molecular recognition properties required to mimic the agonist properties of peptides may be more complex and, perhaps, uncompromising relative to those effecting antagonist properties. In fact, the identification of many nonpeptide

Tab. 7 Examples of peptide, peptidomimetic or nonpeptide drugs, clinical candidates or preclinical lead compounds.

Recombinant peptides	***Proposed/Known therapeutic application***
Atrial natriuretic factor	Potential use in prophylaxis and/or treatment of acute renal failure
Epidermal growth factor	Potential use in skin grafting, eye surgery, and/or treatment of burns or ulcers
Somatotropin	Treatment of growth defects
Insulin	Treatment of type-I diabetes
Glucagon-like insulinotropic peptide	Potential treatment of insulin-insensitive diabetes
Teriparatide	Treatment of osteoporosis
Transforming growth factor	Potential treatment of wound healing and burns
Nerve growth factor	Potential treatment of neural plasticity defects as possibly related to Alzheimer's disease
Interleukin-1α/β	Potential use in cancer therapy and inflammation
Hirudin	Potential use in prevention or treatment of venous blood clots
Tissue plasminogen activator	Potential use as an anticoagulant in heart attacks
Factor VII	Potential use as a blood clotting factor in major forms of hemophiliacs
Erythropoietin	Potential uses in treatment of anemia in kidney dialysis patients, AIDS, and cancer
Platelet-derived growth factor	Potential uses in promoting growth of fibroblasts, keratinocytes, and formation of new blood vessels
Interferon-α	Potential use as an immune stimulant for cancer therapy
Interferon-β	Potential use as an immune stimulant for treatment of viral diseases and multiple sclerosis
Interferon-γ	Potential use as an immune stimulant for treatment of infectious diseases, cancer, and rheumatoid arthritis
Synthetic peptides (relationship to native peptide)	***Proposed/known therapeutic applications***
Aspartame (dipeptide mimic of glucose)	Artificial sweetener
Cyclosporin A (peptide natural product)	Immunosuppressive drug

Sandostatin™ (hexapeptide agonist analog of SRIF)	Symptomatic treatment of acromegaly and carcinoid syndrome
Techtide P829 (radiolabeled SRIF analog)	Detection of tumors
Desmopressin (nonapeptide agonist analog of vasopressin)	Treatment of severe diabetes insipidus
Buserelin (nonapeptide agonist analog of GnRH)	Treatment of prostate cancer
Goserelin (GnRH agonist)	Treatment of cancer and hormonal menstruation disorder
HOE-140 (decapeptide antagonist analog of BK)	Potential treatment of pain, inflammation, rhinitis, and/or asthma
Melanotan-I (tridecapeptide agonist analog of MSH)	Potential use as a stimulant of skin pigmentation
Tetracosactide (peptide agonist analog of ACTH)	Diagnostic agent of adrenal function
BQ-123 (cyclic pentapeptide, natural product antagonist of endothelin)	Potential treatment of hypertension, restenosis, and related disorders
Calcitonin (identical to native peptide)	Treatment of hypercalcemia, Paget's disease, osteoporosis, and pain affiliated with bone cancer
Ebiratide (hexapeptide agonist analog of ACTH)	Potential application for CNS disorders; cognition treatment
Pentigetide (pentapeptide antagonist analog of IgE)	Potential antiallergic agent
MDL28050 (decapeptide antagonist analog of Hirudin)	Potential use as thrombin inhibitor based, anticoagulant agent
Eptifibatide (cyclic heptapeptide)	Potential antithrombotic drug
L365209 (pseudohexapeptide, natural product antagonist of oxytocin)	Potential use as uterine relaxant for prevention of premature labor
Atosiban (oxytocin antagonist)	Potential use for preterm labor
DPDPE (pentapeptide agonist analog of enkephalin)	Potential analgesic lead; δ-selective opioid agonist
GHRP-6 (hexapeptide agonist at gherelin receptor)	Potential GH secretagogue lead
Groliberin (growth hormone-releasing factor analog)	Potential use for treatment of growth hormone deficiency
Enfuvirtide (inhibitor of viral fusion)	Potential antiviral and anti-HIV drug

(*continued overleaf*)

Tab. 7 (*Continued*)

Peptidomimetics or nonpeptides (relationship to native peptide)	***Proposed/known therapeutic applications***
Morphine (peptidoligand natural product agonist at μ-type enkephalin receptor)	Analgesic drug
Captopril, Enalapril (peptidomimetic inhibitors of ACE)	Antihypertensive drugs
DuP-753 (nonpeptide antagonist at AT_1-type AII receptor)	Potential antihypertensive drug
A-72517 (peptidomimetic inhibitor of renin)	Potential antihypertensive drug
Ro-318539, A-75925 (peptidomimetic inhibitors of HIV protease)	Potential anti-HIV/AIDS drug
CP-96345 (nonpeptide antagonist at NK_1 receptor)	Potential use as an antiallergic or analgesic drug
SC-47643 (peptidomimetic antagonist at fibrinogen receptor)	Potential use for prevention of thrombus formation
Integrilin (peptidomimetic antagonist at fibrinogen receptors)	Potential use for myocardial infarction
OPC-21268 (nonpeptide antagonist at V_1 vasopressin receptor)	Potential use as a diuretic agent
Ro-24 9975 (peptidomimetic agonist at TRH receptor)	Potential use as cognitive enhancer agent
CI-988 (peptidomimetic antagonist at CCK_B receptor)	Potential anxiolytic drug
MK-329	Potential use for treatment of pancreatitis (nonpeptide antagonist at CCKA receptor)
L-692429 (nonpeptide agonist at gherelin receptor)	Potential use as a GH secretagogue
SCH-34826 (peptidomimetic inhibitor of metallo-endopeptidase EC 3.4.24.11)	Potential antihypertensive drug; inhibitor of atrial naturetic peptide (ANP) degradation
CT-0543 (peptidomimetic inhibitor of gelatinase)	Potential antimetastatic agent
MD-805 (peptidomimetic inhibitor of thrombin)	Potential antithrombolytic drug
Bivalirudin (thrombin inhibitor)	Potential anticoagulant and antianginal drug
Ampicillin, amoxycillin (nonpeptide inhibitors of bacterial cell wall peptidoglycan synthesis)	Antibacterial penicillin-related drugs

Note: See Table 3 and Fig. 4 for some chemical structures.
Source: Adapted from Sawyer, T.K. (1995) in: Amidon, G., Taylor, M. (Eds.), *Peptide-Based Drug Design: Controlling Transport and Metabolism*, ACS Books, Washington, DC.

antagonists at peptide receptors has been significantly advanced by mass screening assays using chemical collections as well as various natural product sources, thereby illustrating the fact that the peptide ligand is not exclusive toward providing the molecular framework for peptide antagonist drug discovery. Interestingly, the recent development of peptide library technologies (including emerging pseudopeptide and peptidomimetic libraries) has accelerated opportunities to identify novel leads as well as the ability to evaluate analogs in a structure–activity sense to expedite the overall process of drug candidate selection. The application of biophysical chemistry (e.g. NMR spectroscopy and X-ray crystallography) and computational chemistry methods have collectively established powerful drug design technologies for the synthetic transformation of native peptides to chemically modified pseudopeptides or peptidomimetics. In the case of cyclic, conformationally constrained peptides, the synergism of NMR spectroscopy and computational chemistry-based analysis of three-dimensional structural properties and model building points to an extremely promising approach to advancing peptide-based drug design in a rational manner. In the case of X-ray crystallography and computational chemistry, the study of peptide or peptidomimetic interaction with proteins (peptidases, kinases, Src homology proteins, antibodies) exploits high-resolution (1.8–3.0 Å) molecular maps of the complex to provide insight into further design strategies. In this regard, recent work on HIV protease, an aspartyl peptidase, is particularly outstanding in that more than 120 inhibitor–enzyme complexes have been determined by X-ray crystallography. Furthermore, this work provides an excellent example of peptide and peptidomimetic drug design and development strategies that integrate molecular biology, biochemistry, synthetic chemistry, biophysical and computational chemistry, pharmacology, and drug delivery research (Fig. 4).

Among the major challenges of developing peptide drugs has been the mode of administration of the particular compound. Peptide drug delivery exists in many forms, including parenteral (intravenous or intramuscular injection), interstitial (subcutaneous implant), oral, nasal, and percutaneous (transdermal) administration. Perhaps, the most well-known injectable peptide is that of insulin, and with respect to implants the recent success of the gonadotropin-releasing hormone analog Zoladex deserves mention. In the case of oral administration, one of the key reasons for pursuing the chemical transformation of peptides into peptidomimetic derivatives has been to identify prototypic lead compounds that may exhibit oral bioavailability. Such efforts have shown success as exemplified by the development of orally effective ACE inhibitors, renin inhibitors, HIV protease inhibitors, thyrotropin-releasing hormone (TRH) agonists, fibrinogen antagonists, and so forth. Nevertheless, the possibility of alternative modes of administration of peptide (or peptidomimetic) drugs remain an intriguing area of research, and some success has been achieved with respect to nasal (oxytocin, desmopressin, buserelin, and calcitonin), pulmonary, buccal, rectal, and transdermal routes of peptide drug administration. Of particular impact on the oral delivery of peptide drugs is the limitation on passive or active transport of molecules of high molecular weight (typically including tetrapeptides and larger peptides) posed by the intestinal epithelia. Similarly, stability toward gastric acid and degradative enzymes, including

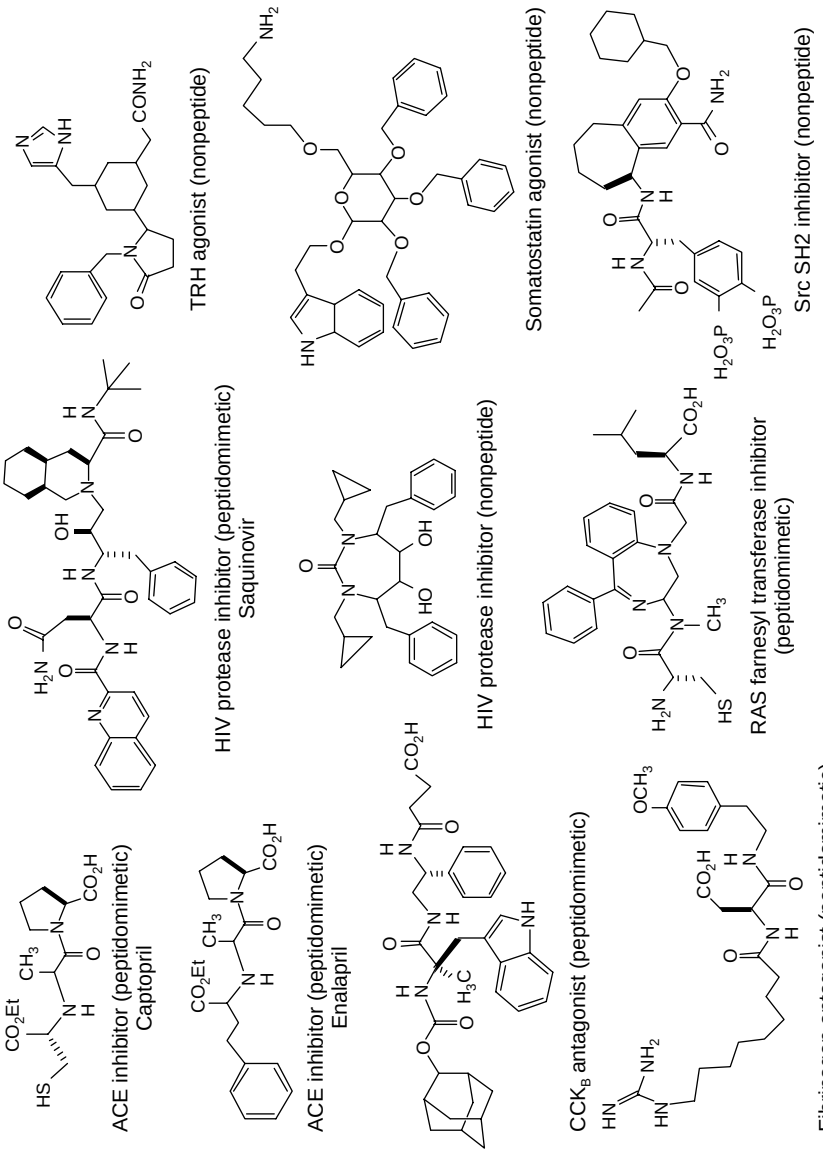

Fig. 4 Chemical structures of some peptidomimetic or nonpeptide drugs, clinical candidates, or preclinical lead compounds.

peptidases associated with the epithelial membrane (luminal side) or within such cells, provides another challenge a peptide-based drug candidate must surpass to eventually access systemic circulation after oral administration.

The apparent medical potential for peptide, peptidomimetic, or nonpeptide drugs is already tremendous, and future opportunities will continue to make this area of pharmaceutical research extremely competitive. There are approximately 100 synthetic and recombinant peptides that are marketed worldwide, and an estimated 200 molecules that are in clinical testing. Nine marketed peptides are responsible for annual sales surpassing the $4 billion mark. Some past and current examples of peptide, peptidomimetic, and nonpeptide preclinical lead compounds, clinical candidates, and marketed drugs are exemplified above in Table 7. Most recently, synthetic and recombinant peptides that have been approved by the FDA include Teriparatide (PTH agonist analog for osteoporosis), Goserelin (GnRH agonist analog for cancer and hormonal menstruation disorder), Integrilin (integrin gpIIb/IIIa antagonist for myocardial infarction) Enfuvirtide (inhibitor of viral fusion to effect antiviral and anti-HIV activity), Eptifibatide (cardiovascular, neuroprotective), Bivalirudin (thrombin inhibitor for anticoagulant and antianginal), Atosiban (oxytocin antagonist for preterm labor), Techtide P829 (radiolabeled somatostatin analog for tumor detection), and Groliberin (GRF analog for treatment of growth hormone deficiency). Past and still existing major peptide and peptidomimetic drugs include insulin (for treatment of diabetes) and dipeptide sweeteners such as aspartame. Without question, the field of peptide, peptidomimetic, and nonpeptide drug discovery has emerged as a superclass of molecules for a plethora of disease as exemplified by the therapeutic scope and molecular diversity as described in this chapter. Future research and drug development holds further promise to combat many existing diseases and to deal with many unmet needs throughout the world.

See also Biotransformations of Drugs and Chemicals; Design and Application of Synthetic Peptides; HPLC of Peptides and Proteins; Medicinal Chemistry; Peptide and Non-Peptide Combinatorial Libraries; Protein Expression by Expansion of the Genetic Code; Receptor Targets in Drug Discovery; Targeting and Intracellular Delivery of Drugs.

Bibliography

Books and Reviews

Amidon, G., Taylor, M. (Eds.) (1995) *Peptide-Based Drug Design: Controlling Transport and Metabolism*, ACS Books, Washington, DC.

Chorev, M., Sawyer, T.K. (Eds.) (2004) Peptide Revolution: Genomics, Proteomics and Therapeutics. *Proceedings of the Eighteenth Peptide Symposium*. American Peptide Society, San Diego, CA.

Gross, E., Meienhofer, J., Udenfriend, S. (Eds.) (1979–1987) *The Peptides: Analysis, Synthesis, Biology*, Vols. 1-IX, Academic Press, New York.

Hadley, M.E. (1992) *Endocrinology*, 3rd edition, Prentice Hall, Englewood Cliffs, NJ.

Hider, R.C., Barlow, D. (Eds.) (1991) *Polypeptide and Protein Drugs: Production, Characterization and Formulation*, Ellis Horwood, Chichester, UK.

Nego-Vilar, A., Conn, P.M. (Eds.) (1988) *Peptide Hormones: Effects and Mechanisms of Action*, Vols. 1–III, CRC Press, Boca Raton, FL.

Ward, D.J. (1991) *Peptide Pharmaceuticals: Approaches to the Design of Novel Drugs*, Open University Press, Buckingham, UK.

Wieland, T., Bodanszky, M. (1991) *The World of Peptides: A Brief History of Peptide Chemistry*, Springer-Verlag, Berlin, Germany.

Williams, W.V., Weiner, D.B. (Eds.) (1993) *Biologically Active Peptides: Design, Synthesis and Utilization*, Technomic, Lancaster, PA.

Primary Literature

Ahn, J.A., Boyle, N.A., MacDonald, M.T., Janda, K.D. (2002) Peptidomimetics and peptide backbone modifications, *Mini Rev. Med. Chem.* **2**, 463–473.

Al-Obeidi, F.A., Hruby, V.J., Sawyer, T.K. (1998) Peptide and peptidomimetic libraries: molecular diversity and drug design, *Mol. Biotechnol.* **9**, 205–223.

Annis, I., Hargittal, B., Barany, G. (1997) Disulfide bone formation in peptides, *Methods Enzymol.* **289**, 198–221.

Burke, T.R. Jr., Yao, Z.J., Liu, D.G., Voigt, J., Gao, Y. (2001) Phosphoryltyrosyl mimetics in the design of peptide-based signal transduction inhibitors, *Biopolymers* **60**, 32–44.

Bursavich, M.G., Rich, D.H. (2002) Designing non-peptide peptidomimetics in the 21st century: inhibitors targeting conformational ensembles, *J. Med. Chem.* **31**, 541–558.

Cheng, R.P., Gellman, S.H., DeGrado, W.F. (2001) Beta-peptides: from structure to function, *Chem. Rev.* **101**, 3219–3232.

Chorev, M., Goodman, M. (1995) Recent developments in retro peptides and proteins–an ongoing topochemical exploration, *Trends Biotechnol.* **13**, 438–445.

Craik, D.J., Simonsen, S., Daly, N.L. (2002) The cyclotides: novel macrocyclic peptides as scaffolds in drug design, *Curr. Opin. Drug Discov. Dev.* **5**, 251–260.

Dawson, P.E., Kent, S.B. (2000) Synthesis of native proteins by chemical ligation, *Annu. Rev. Biochem.* **69**, 923–960.

Eguchi, M., Kahn, M. (2002) Design, synthesis, and application of peptide secondary structure mimetics, *Mini Rev. Med. Chem.* **2**, 447–462.

Fairlie, D.P., Abbenante, G., March, D.R. (1995) Macrocyclic peptidomimetics: forcing peptides into bioactive conformations, *Curr. Med. Chem.* **2**, 654–686.

Fauchere, J.-L. (1986) Elements for the Rational Design of Peptide Drugs, in: Testa, B. (Ed.) *Advances in Drug Research*, Vol. 15, Academic Press, London, UK, pp. 29–69.

Figliozzi, G.M., Goldsmith, R., Ng, S.C., Banville, S.C., Zuckermann, R.N. (1996) Synthesis of N-substituted glycine peptoid libraries, *Methods Enzymol.* **267**, 437–447.

Freidinger, R.M. (1999) Nonpeptidic ligands for peptide and protein receptors, *Curr. Opin. Chem. Biol.* **3**, 395–406.

Goodman, M., Ro, S. (1994) Peptidomimetics for Drug Design, in: Wolff, M.E. (Ed.) *Medicinal Chemistry and Drug Design, Vol. I. Principles of Drug Discovery*, 5th edition, John Wiley & Sons, New York, pp. 803–861.

Holder, J.R., Haskell-Luevano, C. (2004) Melanocortin ligands: 30 years of structure activity relationship (SAR) studies, *Med. Res. Rev.* **24**, 325–356.

Houghten, R.A. (2000) Parallel array and mixture-based synthetic combinatorial chemistry: tools for the next millennium, *Annu. Rev. Pharmacol. Toxicol.* **40**, 273–282.

Hruby, V.J. (2002) Designing peptide receptor agonists and antagonists, *Nat. Rev. Drug Discov.* **1**, 847–858.

Hruby, V.J., Al-Obeidi, F.A., Kazmierski, W. (1990) Emerging approaches in the molecular design of receptor-selective peptide ligands, *Biochem. J.* **268**, 249–262.

Jackson, D.C., Purcell, A.W., Fitzmaurice, C.J., Zeng, W., Hart, D.N. (2002) The central role played by peptides in the immune response and the design of peptide-based vaccines against infectious diseases and cancer, *Curr. Drug Targets* **3**, 175–196.

Kempf, D., Sham, H.L. (1996) HIV protease inhibitors, *Curr. Pharm. Des.* **2**, 225–246.

Lam, K.S., Renil, M. (2002) From combinatorial chemistry to chemical microarray, *Curr. Opin. Chem. Biol.* **6**, 353–358.

Lauer-Fields, J.L., Juska, D., Fields, G.B. (2002) Matrix metalloproteinases and collagen catabolism, *Biopolymers* **66**, 19–32.

Liu, L.P., Deber, C.M. (1998) Guidelines for membrane proteins engineering derived from de novo designed model peptides, *Biopolymers* **47**, 41–62.

Marshall, G.R. (1993) A hierarchical approach to peptidomimetic design, *Tetrahedron* **49**, 3547–3558.

Marshall, G.R. (2001) Peptide interactions with G-protein coupled receptors, *Biopolymers* **60**, 246–277.

Micklatcher, C., Chmielewski, J. (1999) Helical peptide and protein design, *Curr. Opin. Chem. Biol.* **3**, 724–729.

Olson, G.L., Bolin, D.R., Bonner, N.P., Bos, M., Cook, C.M., Fry, D.C., Graves, B.J., Hatada, M., Hill, D.E., Kahn, M., Madison, V.S., Rasiecki, V.K., Sarabu, R., Sepinwall, J., Vincent, G.P., Voss, M.E. (1993) Concepts and progress in the development of peptide mimetics, *J. Med. Chem.* **36**, 3039–3049.

Pavia, M.R., Sawyer, T.K., Moos, W.H., Eds. (1993) The generation of molecular diversity. *Bioorg. Med. Chem. Lett.* **3**, 387–396, Symposium-in-print.

Qian, Y., Sebti, S.M., Hamilton, A.D. (1997) Farnesyltransferase as a target for anticancer drug design, *Biopolymers (Peptide Sci.)* **43**, 25–41.

Quan, M.L., Wexler, R.R. (2001) The design and synthesis of noncovalent factor Xa inhibitors, *Curr. Top. Med. Chem.* **1**, 137–149.

Rich, D.H. (2002) Discovery of nonpeptide, peptidomimetic peptidase inhibitors that target alternate enzyme active site conformations, *Biopolymers* **66**, 115–125.

Ripka, A.S., Rich, D.H. (1998) Peptidomimetic design, *Curr. Opin. Chem. Biol.* **2**, 441–452.

Sadler, K., Tam, J.P. (2002) Peptide dendrimers: applications and synthesis, *J. Biotechnol.* **90**, 195–229.

Sakakibara, S. (1995) Synthesis of large peptides in solution, *Biopolymers* **37**, 17–28.

Sawyer, T.K. (1997) Peptidomimetic and nonpeptide drug discovery: chemical nature and biological targets, in: Reid, R. (Ed.) *Drugs and the Pharmaceutical Sciences*, Vol. 101, Marcel Dekker, New York, pp. 81–114.

Sawyer, T.K. (1999) Peptidomimetic and Nonpeptide Drug Discovery: Impact of Structure-Based Drug Design, in: Veerapandian, P. (Ed.) *Structure-Based Drug Design: Diseases, Targets, Techniques and Developments*, Marcel Dekker, New York, pp. 559–634.

Sawyer, T.K., Bohacek, R.S., Dalgarno, D.C., Eyermann, C.J., Kawahata, N., Metcalf, C.A., Shakespeare, W.C., Sundaramoorthi, R., Wang, Y., Yang, M.G. (2002) Src homology-2 inhibitors: peptidomimetic and nonpeptide, *Mini Rev. Med. Chem.* **2**, 475–488.

Singh, S.B., Lingham, R.B. (2002) Current progress on farnesyl protein transferase inhibitors, *Curr. Opin. Drug Discov. Dev.* **5**, 225–244.

Skiles, J.W., Gonnella, N.C., Jeng, A.Y. (2001) The design, structure and therapeutic application of matrix metalloproteinase inhibitors, *Curr. Med. Chem.* **8**, 425–474.

Songyang, Z., Cantley, L.C. (1995) Recognition and specificity in protein tyrosine kinase-mediated signalling, *Trends Biochem. Sci.* **20**, 471–475.

Sundaram, R., Dakappagari, N.K., Kaumaya, P.T. (2002) Synthetic peptides as cancer vaccines, *Biopolymers* **66**, 200–216.

Yamashita, D.S., Dodds, R.A. (2000) Cathepsin K and the design of inhibitors of cathepsin K, *Curr. Pharm. Des.* **6**, 1–24.

Zhang, Y.Z. (2001) Protein tyrosine phosphatases: prospects for therapeutics, *Curr. Opin. Chem. Biol.* **5**, 416–423.

Systems Biology

H. Busch and R. Eils
German Cancer Research Center, Heidelberg, Germany

1 Introduction 125

2 What is System-level Understanding? 126

3 What Makes Biological Systems so Special? 128
3.1 Biological Complexity 129
3.2 Global Properties of Biological Systems 130
3.2.1 Robustness of Biological Systems 130
3.2.2 System Adaptation and Control 131
3.2.3 Modules and Protocols 132

4 Systems Biology Modeling 133
4.1 Network Biology 136
4.1.1 Dynamic Network Models 138
4.2 Reaction-diffusion Models 138
4.3 Intracellular Noise Models 139
4.4 Modeling Resources 141

5 Future Prospects of Systems Biology 142
5.1 Synthetic Cell Biology 143
5.2 Conclusions – Where are We? 144

Bibliography 146
Books and Reviews 146
Primary Literature 146

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

Keywords

System
A set of interacting parts functioning as a whole and distinguishable from its surroundings by identifiable boundaries.

Systems Theory
Systems theory denotes the cross-disciplinary investigation of the abstract organization of systems, irrespective of their substance, type, or spatiotemporal scale of existence. The goal is the study of emerging properties that arise from the interconnectedness of the individual parts making up the system.

Robustness
Robustness of biological systems denote the maintenance of specific system functionalities in the presence of fluctuations or change in environmental parameters.

Control
Control is defined as the response action taken by a system to counteract parameter changes to maintain system functions at a certain, predefined level.

Modularity
A design concept of complex systems to integrate simpler, self-contained functional building blocks into the framework of one larger system.

Model
The concept of representing causal relationships from real systems in the language of mathematics.

Systems biology is a new field of biology, which puts the theoretical foundations of system-level analysis of living matter into the context of modern high-throughput quantitative experimental data, mathematics, and *in silico* simulations. It aims at analyzing the organization and gaining engineering-control of metabolic and genetic pathways. The ultimate goal will be to gain a 'holistic' view on the complex workings of life. This article reviews the need for a system-level understanding of biology, comments on the current scientific progress in this field, and points out future directions of experimental design strategies and theoretical approaches.

1
Introduction

Systems biology is a newly established field in life sciences that aims at promoting a global system-level understanding of living matter through the integration of various scientific domains.

The considerate attention Systems biology receives nowadays is due to the fact that it will most likely cause a paradigmatic shift in biological research. Modern molecular cell biology so far has been a descriptive science, devoting insight mainly to isolated small compartments of a system as a whole, for example, by investigating the influence of individual molecules types within the whole cell cycle. The study of the interconnected nature of cellular processes has long been avoided in favor of a reductionist approach. This is on one hand due to the sheer amount of new challenges that come about when tackling complex systems. On the other hand, it has been the common leitmotif in other natural sciences, such as physics, to shed light mostly on well-controlled systems that are either small and isolated, or large and homogeneous so that they can be tackled with the laws of statistics. It is dawning on us only now that the true challenge lies in the description of dynamical, mesoscopic, open, spatiotemporally extended, nonlinear systems that operate in the absence of thermodynamic equilibrium, and these systems are the most important to understand as they are the ones that support life

Although the reductionist approach has been successful in identifying key factors for many fundamentally important biological processes, contemporary science has to recognize the importance of wholeness, which is defined as problems of organization instead. Emergent phenomena arise from the interaction of various units or modules, which are neither resolvable nor understandable through the study of local events or the respective parts in isolation. Hence, traditional reductionist models and methods of cell and molecular biology are not very well suited and can be incomplete, misleading, or even completely wrong.

Historically, Jan Christiaan Smuts was among the first to formulate a theory of the whole that was hoped to fill the gap between science and philosophy. In his book "Holism and Evolution," published in 1926, Smuts argued that nature consisted of discrete objects, or 'wholes,' that are not entirely resolvable into their respective parts. The wholes and parts mutually depend on each other in their functionality, thus forming one organic, unified web of relations, which comprises matter, life, and mind that cannot be accounted for by a reductionistic analysis. Smuts saw his idea confirmed in evolution, regarding Holism as the active driving force toward more perfect wholes or species.

The theoretical foundations of systems engineering have already been laid 50 years ago. The concept of systems theory in biology (cf. Sec. 2) was proposed in the 1940s by the biologist Ludwig von Bertalanffy and further developed in the 1950s by Ross Ashby as a countermovement against reductionism in science. In the sense of holism, von Bertalanffy emphasized the need for a study of the informational organization within real, open systems. The assembly of such interrelated elements then comprise a unified whole, which in turn can show new emergent properties.

In 1948, the mathematician Norbert Wiener established the field of cybernetics as the science of communication and control of systems in regard to their environment. Cybernetics is closely related

to systems theory, using the same concepts of information, control, or feedback. However, the former focuses on systems function for providing regular and reproducible behavior, whereas the latter deals more with system structure. Even so, both terms are often used in conjunction, for both structure and function cannot be understood as separate entities.

Although the mathematical tools required for Systems biology had already been refined decades ago, the approach of systems theory in biology has gained momentum only now in the era of genomics and proteomics as a consequence of high-throughput measurements, which provide the necessary amount of quantitative data for the establishment of appropriate holistic models of cellular processes.

Today, biology embarks on systems thinking in two different ways. One way is to regard Systems biology as a new way toward integrating information from different organizational levels, starting from DNA to proteins via metabolic pathways to functional modules, into the context of a holistic organizational view. The primary goal of the second view on Systems biology is to establish a conceptual framework and working methodologies for the augmentation of knowledge on the workings of biological phenomena, thus combining systems theory and molecular biology: "Systems biology is not a collection of facts but a way of thinking". This view has already been shared by Mesarović in the late 1960s. He predicted that Systems biology would be an established field of science as soon as "biologist start asking the right questions". Put differently, biologists need not recast facts already known from molecular biology in a different language, but they need to ask questions based on system-theoretic concepts.

Both ways share the extensive need for high-quality, quantitative biological data through extensive experimental endeavors and new mathematical analysis tools as a basis of developing detailed models that are essential for the study of systemic properties and behavior. Currently, it is the hope in the postgenomic era that new breakthroughs in experimental techniques lay the foundation for the integration of mathematics, engineering, physics, and computer science into biology to understand the range of complex biological regulatory systems at multiple hierarchical and spatiotemporal levels of cellular organization.

Section 2 defines the concept of system-level understanding in detail. The need for Systems biology will become clear from the discussion on the outstanding properties of biological systems, such as complexity, modularity, stochasticity, and hierarchical organization in Sect. 3. These insights lead to guidelines and strategies for the design of experiments and models with the need of a unified computational infrastructure as explained in Sect. 4. The last section (Sect. 5) concludes with an outlook on the future prospects of Systems biology and the hope for the initiation of a Systeome Project.

2
What is System-level Understanding?

The word "system" derives from the Greek and is composed of the prefix *syn*, meaning together, and the root of *histanai*, meaning "cause to stand." A system is defined as the assembly or set of interrelated elements comprising of a unified whole that is distinct from its environment. It can be hierarchically organized and made up of

other subsystems or modules, which allows for constructing a complex entity out of simpler units. For example, organelles like mitochondria constitute distinct subsystems within the organization of a cell. The subdivision of natural entities into systems is an abstract construct. Systems *per se* do not really exist in reality. They are rather defined as a set of elements interacting over time.

Systems theory denotes the transdisciplinary investigation of the abstract organization of phenomena, independent of their substance, type, or spatiotemporal scale of existence. Its goal is to study emerging properties arising from the interconnectedness and complexity of relationships between parts. It argues that however complex or diverse a system is, there are always different types of organizational structures present, which can be represented as a network of information flow. Because these concepts and principles remain the same across different scientific disciplines like biology, physics or engineering, systems theory provides a basis for their unification. The systems ansatz distinguishes itself from the more traditional analytic approach by emphasizing concepts of system-environment boundaries, signal input–output relationships, signal and information processing, system states, and hierarchies. Albeit systems theory is valid for all system types, it usually focuses on complex, adaptive, self-regulating systems that are called *cybernetic*.

Elegant, simple, and globally valid models are rare in biology as compared to other fields of science. Few examples exist where some function can be attributed to the workings of a single small molecule or protein, as is the case with hemoglobin during gas transport in the blood stream. In most cases, several genes are coexpressed and a multitude of proteins is involved for each cellular signaling pathway. Nevertheless, modern biology so far mostly followed a reductionist approach, typically tackling small isolated parts of an organism, trying to deduce biological phenomena from molecular behavior, which oftentimes results in a simplistic "one gene for one function"-approach. However, genetic analysis has shown that the genotype of different species is mostly identical. It must be thus the signal processing stages on the way from the genome to the phenotype, in other words, an ever more elaborate regulation of gene expression that amplify the subtle particularities in the respective genetic codes.

Therefore, it becomes clear that future explanations of biological phenomena ought to be explained in the vocabulary of system theory, such as amplification, control, adaptation, sensitivity, autoregulation and error correction, taking a holistic view of the system under consideration. In short, Systems biology is needed to uncover the laws of the whole that cannot be deduced by delving deeper into the details.

Accordingly, system-level investigation of biological matter comprises the understanding of system structure and function in the sense of systems theory and cybernetics, respectively.

System structure: This denotes the identification of the static connection topology and regulatory relationships within the network of genes, proteins, and other small molecules that constitute the signal transduction and metabolic pathways as well as the physical structure of organisms or cells. Some of the experimental techniques currently available to deduce the cells' global

system structure are DNA microarray or protein complex identification using TAP tagging and mass spectrometry analysis.

System dynamics: This refers to the qualitative and quantitative evolution of the above network over time. Dynamics include the temporal variation of the metabolites' concentrations as well as the structure of the network itself. Different methods are applicable for the proper establishment of the model, such as metabolic, sensitivity, or dynamic analysis.
The system dynamics of gene regulation can be determined to a certain degree by quantitative Westernblots and RT-PCR. Spatiotemporal dynamics of individual molecules has become available with the help of fluorescent marker proteins. In particular FRAP (fluorescence recovery after photobleaching), FRET (fluorescence resonance energy transfer and FCS (fluorescence correlation spectroscopy) allow the determination of molecular diffusion constants, molecular interaction, and the localization of individual proteins or protein complexes in time and space.

Once the structure and its spatiotemporal dynamics have been (partly) understood, it is a future vision to develop methods controlling the state of the respective biological systems. The control schemes could be utilized to minimize malfunctions or treatment against disease within the particular system. The ultimate goal will be to design new multitarget methods for cellular control schemes as well as the forward engineering of living matter from scratch on the basis of the knowledge from model predictions and understanding. The area of systems control and design in biology is usually referred to as *Synthetic biology* (See Sec. 5.1).

3 What Makes Biological Systems so Special?

While biological systems still need to be based on laws of (quantum-)mechanics, chemistry, and thermodynamics, biology additionally incorporates the notion of organismal function, which represents the need for survival and reproduction, as well as the possibility to evolve in and adapt to changing environments. This inherent purpose differentiates biology significantly from all other natural sciences. Moreover, there is no distinct separation of information storage and regulatory units. Both genes and proteins are used for long- and short-term storage of interaction rules as well as being active regulatory elements. Genes, for example, can regulate their own expression. Classical physics views the emergence of every effect to be determined by a cause residing in the past. Biological systems on the contrary are teleonomic, that is, they are oriented toward a state in the future. It was the insight of Cybernetics that purposeful activity can be described through the use of circular mechanisms, where the effect equals the cause. The simplest example of a circular logic is a feedback loop, where the output of the system is fed into its input again (See Sec. 3.2).

Having these prerequisites in mind, the concepts of systems theory and cybernetics should be utilized, if one wants to establish successful formal mathematical models in biology. The introduction of circular causalities has far reaching consequences on the general design and global

properties of biological systems, such as robustness, complexity and control, which are discussed in detail in this section.

3.1 Biological Complexity

In order to comprehend the challenges in Systems biology, it is important to understand the origin of the complexity encountered in nature. First, it is important to note that complexity within a system does not necessarily come about as a consequence of the number of component parts. In physics, the macroscopic state or the dynamics of objects are oftentimes well described by a few parameters or simple mathematical equations. For example, the temperature of an object comes about due to the thermal agitation of its individual atoms. However, their average energy can be ascribed to a single, macroscopically measurable quantity.

A complex system is one in which the laws that describe its behavior are qualitatively different from those that govern its units, such that new features emerge as one moves from one temporal, spatial, or organizational scale to another. The science of complexity is about revealing the principles that govern the ways in which these new properties appear. Thus, it is rather the organization of the system into irreducible, heterogeneous parts that are highly structured and hierarchically organized on various spatiotemporal scales, which makes a system to become complex, or in other words, "complicated."

The notion of complexity is not well defined amongst the various science disciplines. Information theory usually characterizes complexity as the amount of information needed to optimally predict the behavior (or state) of the system on the basis of entropy measures. However, this definition refers to sequence complexity and is thus unsuitable for biology, where, for example, the sequential makeup of proteins is largely uncorrelated with their respective function. One viable approach toward defining the complexity of biological systems may be the identification of the topological structure at a higher level of large-scale organization in terms of hierarchically organized networks.

Complex systems in physics are usually defined to consist of many individual, yet simple, and identical or similar elements, whose complex dynamics arise from the interaction of its elements alone. Physicists try to deduce simple and few interaction laws, which are then sufficient to describe the complexity emerging at the macroscopic scale of a system, for example, by a power-law behavior with respect to the size of events and their probability to occur.

Different approaches toward describing this type of complex systems exist. One is the so-called "slaving-principle". It makes use of the fact that under certain conditions the global, macroscopic behavior of a system is governed by a few, slowly evolving state variables, which "slave" all other dynamics. This way, the relevant degrees of freedom of such a complex system are largely reduced, allowing the system to find its own structure, that is, to self-organize. One use of the slaving principle is the analytical description of lasers, explaining the spontaneous, synchronous light emission by the atoms in the lasing medium.

Another popular framework from statistical physics for the explanation of emergent phenomena in multibody systems is the notion of self-organized criticality (SOC), which was acclaimed to explain the frequent occurrence of long-tail distributions in many natural phenomena. SOC denotes the ability of open and dissipative

systems to display critical, that is, scale-invariant, behavior even in the absence of external pressures. Unfortunately, this theory did not live up to its promise of being universally applicable, but it has its applications in, for example, earthquake, forest fire, or avalanche models.

The limited applicability of SOC or the slaving principle on biological phenomena is mainly due to the fact that biological systems need a redefinition of complexity that is quite different from that in physics. Biological systems are heterogeneous, modular, highly structured on multiple irreducible spatiotemporal scales, and self-dissimilar with each entity usually having several functional and regulatory properties. Nevertheless, the individual components "interact selectively and nonlinearly to produce coherent rather than complex behavior."

Carlson and Doyle therefore introduced the theory of highly optimized tolerance (HOT). It accounts for the intrinsic design of biological and engineered systems. Their theory reflects the behavior of such high-throughput, high-density systems, which are faced with limited resources. These systems show a high tolerance, that is, robustness, against environmental parameter fluctuations. This robustness is achieved at the cost of a high degree of complexity through the addition of control units to the system. Resource constraints then call for an optimized tradeoff between fault tolerance toward frequently experienced perturbations and fragility to rare, yet possible events. Thus, there exists a "conservation law of robustness."

The complex design of even the simplest eukaryotic cells is often compared to the makeup of computers or today's commercial aircraft. Indeed, those man-made machines have a bewildering complexity such that no man alone understands all the parts and their interplay in complete detail. Many of the strongly interacting and irreducible complex functions are there to automate, control, or backup operations. This elucidates the main difference between systems describable by SOC or HOT. The former theory relates to scale-invariant and self-similar features, while the latter deals with self-dissimilar structures and sensitivities (dis-)appearing on each scale of observation such that the degrees of freedom in these systems are irreducibly many.

3.2 Global Properties of Biological Systems

Systems thinking provides universal properties of biological systems that link the emergence of complexity to other general design features. These can be used as a guideline for abstract mathematical modeling, looking for principles that are commonly shared between diverse species. Moreover, these properties show how biology and engineering converge on a systems level view.

3.2.1 Robustness of Biological Systems

Fault tolerance or robustness of biological systems denotes the maintenance of specific system functionalities even in the presence of fluctuations or a change in environmental parameters. In biology, it refers to the concept of homeostasis and the stability of developmental control. Robustness is achieved through the incorporation of regulatory control loops into a system, thus shielding or buffering the desired system functionality from environmental influence. Hence, it is important to note that robustness is a relative system property. For the maintenance of a certain equilibrium, other properties must evolve

and adapt. It is therefore required to define which cellular functions change and which resist as a reaction toward a disturbance.

Most of the "complicatedness" encountered in biological systems is a direct consequence of the implementation of control schemes rather than the core function itself. In particular, if robustness is to be achieved for a wide range of disturbances, the control must have an equally increasing variety as stated in the "Law of Requisite Variety." Consequently, complexity serves simplicity in the sense that the complicated control schemes are hidden underneath the simple, yet reliable output.

Robustness in biological systems is usually achieved by redundant or functionally overlapping regulatory pathways, signal control with the help of feedback loops and certain checkpoints within the cell, all of which add to the total number of components regulating the cell, while keeping the number of distinct phenotypic characteristics low. Cellular checkpoints play an important role, for example, in mitosis, the cell cycle, and in the early stages of embryo development. Bacterial chemotaxis is an intensively studied example of a robust behavior due to multiple feedback control. An example for the use of genetic buffering for robustness is shown by the fact that only about 20% of genes in budding yeast are essential for viability.

Most biological systems employ complexity to buffer against environmental changes or genetic defects. For example, the simplest bacteria such as Mycoplasma with a couple of hundred genes survive only within a narrow band of environmental parameters. *Escherichia coli* on the other hand, possessing about 3000 genes, survive in various environmental conditions. Hence, most of the additional control schemes are inactive under normal laboratory conditions and are possibly activated under severe environmental conditions only, which usually hides the complexity of regulation.

Another interesting consequence of robustness comes from the assumption that dynamical behavior in biological systems is coupled less to parameters themselves, but more to the overall system structure itself. Von Dassow showed that robustness was the simplifying criterion for the determination of the correct topology of the segment polarity network, producing highly robust patterning over a large range of parameter variation.

Robustness in complex systems comes at the price of some fragility due to tiny, yet rare events. Indeed, a small rearrangement of some cellular signal pathways can oftentimes lead to a spectacular failure, that is, impaired cell death. For example, cancer can be caused through the accumulation of tiny mutations in the genetic code over the human life span. Nevertheless, the outbreak of cancer during the reproductive years of a human are rather rare and nature seems to cope well with this tradeoff for the benefit of robust and reliable "normal" functioning.

3.2.2 System Adaptation and Control

Adaptation is the ability of a system to accommodate for varying external input stimuli or disturbances to gain and maintain the correct and optimized output. This is usually achieved with the help of feedback control. Popular examples of adaptation using integral feedback control is chemotaxis, in which bacteria adapt their movement not to the absolute level of pheromones but to the chemical gradient only, or the activation of the heat shock protein in *E. coli* under temperature stress.

Control is a recurrent design encountered in natural systems, which effectively increases robustness with the goal to keep certain values within predefined physiological limits. One distinguishes between feedforward and feedback system control. The former is an open-loop sequence of predefined actions triggered by a certain stimulus, working reliably only within strict ranges of input stimuli, yet simple in design. The latter employs a closed-loop design, which feeds part of the output signal back into the system input. Depending on the sign of the feedback, positive feedback (or autocatalysis) will amplify the output signal, thus enhancing reliable sensitivity toward cellular decision derived from noisy input signals. Negative feedback, homeostasis, stabilizes the output around some desired value by opposing any changes caused by disturbances.

Although both types of control are successfully employed in intracellular signaling, both bear fragility as well. Autocatalysis can lead to self-sustained, unrestricted signal amplification, as observed in uncontrolled tumor growth, while ultrastable homeostasis regulation can cause large, possibly disturbing, transient signals in response to external fluctuations.

Most often, biological systems use combinations of open and closed-loop designs, basing their control action upon the absolute, accumulative, or differential value of the input stimulus, thus balancing both sensitivity and stability.

3.2.3 Modules and Protocols

Modules are characterized by being mostly self-contained, having fixed interfaces for external communication and are evolutionarily detached, possessing their own identity, having many internal, but only few external links through which information and matter is exchanged. Systems are then thought of as assemblies of modules, like entire living organisms or an individual.

Modular design is a commonly found principle in modern engineering. It enables the independent development and testing of units before integration into a common system. Further benefits are the savings in developmental and maintenance cost and the prospect of graceful degradation instead of catastrophic failure, as errors are usually restricted to a single module.

Similarly, modularity seems to be an important concept in biology, probably stemming from the evolutionary pressure for optimal flexibility and the low chance of damage spread. Spatially distinct entities, like organelles in the cell or metabolic units seem to be a recurrent scheme found throughout various organisms. Genomic research on the gene regulatory network of yeast led to the discovery of a set of regulatory motifs, which can be seen as conceptual modular entities. Such motifs have also been discovered in *E. coli* and are not unique to cellular regulation, emerging in food webs, the Internet and neural networks as well.

If modular design of biological systems is a governing principle, it opens up new possibilities for simplification of *in vivo* and *in silico* experiments. Modules would provide fixed levels of detail and size, which are easily abstracted once their functions are known. With the help of these core modules, it would be possible to build ever more complex models without the need for segmentation of every level of detail.

Modules communicate using protocols. Protocols are fixed, commonly agreed on rules that standardize communication with their respective surroundings. They ensure error correction, cellular coordination, as well as evolvability through the

possibility of adding new functions to a particular module. It has been shown that protocols are an efficient means for the hierarchical organization of complex systems through their integration of different layers, thus reducing the costs of information transmission. For example, negative feedback is a powerful module for the establishment of homeostasis. Similarly, gene regulation, membrane potentials, or signal transduction pathways can all be regarded as protocols being utilized by the different biological modules, such as the DNA, ion channels or kinases, and phosphatases. Elaborate feedback control protocols are robustly established for the spatiotemporal development of various species. An extensive genome-wide analysis of prokaryotic cell-cycle progression revealed a hierarchical control structure with three to four master regulatory proteins acting in a coordinated fashion.

Interestingly, the use of protocols on the cellular level has striking similarities with modern communication of computers via the Internet. The Internet employs many protocols for persistent communication, which can be ordered into a general hierarchy starting from the data link via the network and transport to the application layer. Popular examples of each hierarchy are the reverse address resolution protocol (RARP) for the IP lookup and search of the communication partner, IP for routing for the appropriate subnet of the data, TCP for the secure, error-free data exchange, and finally the application layer like HTTP or FTP for retrieving a Web page or downloading files. The fact that these protocols have been used for decades, irrespective of the fast and ongoing evolution of computer hard- and software, highlights their importance. The Internet moved into the focus of intensive research, beginning to have a rich theory. Its close resemblance with other complex systems, while at the same time having available enormous data sets, make it an attractive example to compare with biological networks.

4
Systems Biology Modeling

The word model derives from the Latin language "modus," that is, "manner, measure," and refers to the concept of representing causal relationships from real systems in the language of mathematics. This mapping oftentimes involves simplifications of the original systems with the hope to gain predictive powers on experimental results and to explain functional design principles.

Modeling and simulation have become indispensable tools for gaining insights into natural systems. In biology, they help to bridge the gap between theory and experiment. Oftentimes, the biologists are faced with the dilemma that experiments do not provide sufficient data for theoretical interpretation, while at the same time, clues for new experiments are missing as a consequence of lacking hypotheses.

Experimental results need proper mathematical interpretation and model hypotheses require experimental proofs. Thus, the process of knowledge generation is an iterative process consisting of two feedback loops (Fig. 1). The experimental part employs experimental techniques to obtain quantitative data of the system dynamics. These are then compared with predictions by the mathematical model. In the case of failure, that is, in the case of diverging results from model and experiment, this process must be repeated by

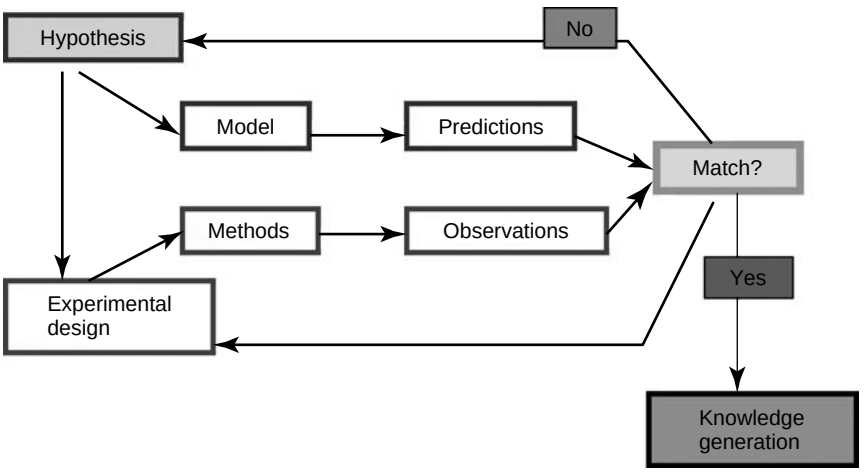

Fig. 1 Schematic to the experiment-modeling loop for knowledge generation in Systems biology.

adjusting the old hypothesis and thus reconciling the model with the new results from the experiment. As an example of this cycle of model building, experiment, and model refinement, Ideker et al. recently developed an integrated approach to construct, test, confirm, and refine the simulation of the yeast galactose utilization network through the combined numerical simulation and analysis of networks with systematic experimental perturbation and global measurements.

The major guideline for optimal modeling should be the concept of Occam's razor, which states that models should be void of any redundant information. In the times of modern biology, this paradigm also needs to be viewed from the other side, that is, that models must not be oversimplified, and thus missing the essential clues of functionality of real-life systems. In particular, it is important to note that modeling in biology must establish different concepts than those in physics. The paradigm of nonlinear science is that even simple systems can lead to dynamically rich behavior. Despite the beauty of the simplicity of this idea, one must realize, however, that living systems regulate their dynamics differently, that is, through a complex makeup of interconnected regulatory functions, a fact that is often neglected or ignored.

Models in biology are usually heuristic. They arise embedded in the process of biological experiments and are coupled tightly to them. Information is, for example, deduced from perturbation analysis of the experimental system and thus contains assumptions about the causality and the passage of time. The construction of a valid working model of a biological system from experimental data can be approached from two sides.

The bottom-up modeling of cellular functions is based on the integration of established biological knowledge on the dynamics of the relevant biological components of the system or regulatory network under consideration. This attempt is useful, when most of the reaction partners are known and their interaction dynamics are understood. The research goal is the establishment of an accurate computer simulation that allows for (1) the analysis

of the system dynamics (2) the scan of parameter ranges that are unattainable in experiments (3) the prediction of unknown functionality or interactions. Attempts of a bottom-up modeling are, for example, the λ-phage decision circuit or the data-driven simulation of a cancer cell.

The top-down modeling ansatz tries to apply statistical analysis to data from high-throughput experiments coming, for example, from DNA microarrays. Data mining techniques search for clusters of coexpressed genes as a consequence of an external perturbation, like the knockout or overexpression of certain genes. The working assumption for these methods is that coexpressed genes also share common relationships with respect to other biological processes, for example, metabolic pathways. The result of a cluster analysis is the construction of an interaction network of genes or proteins, whose topology can hint at biologically reasonable organization principles.

The knowledge-based bottom-up approaches toward modeling possess a certain appeal to biologists. However, successful modeling needs to overcome the gaps in understanding. Most of protein–protein interactions are still unknown and the underlying physics of interacting molecules needs better attention. The influence of high molecule concentration in subcompartments of a cell environment on reaction rates is as yet widely unknown. The spatial distribution of reactants and molecular crowding may well affect reactions and their rates, for example, explaining the symmetry breaking process of mitosis. A particular challenge arises from identifying the relevant components, which is most difficult due to the vast number of combinations of active molecules. These obstacles presently impair mathematical modeling in biology.

Besides the ever growing need for more experimental biological data, the success of quantitative modeling will especially depend on the quality and diversity of such data. Experimental data needs to be comprehensive with respect to four aspects:

Factor: the need to capture the behavior of all important target factors, such as the genes and proteins that play decisive roles in the experimental system under consideration.

Item: this refers to the simultaneous measuring of necessary sets of variables, such as transcription level, molecule concentration or spatial information that are required for reliable hypothesis building from experiments.

Time and space: this refers to the need for a sufficiently high spatiotemporal sampling rate of the experimental data to obtain a reasonable resolution of the spatiotemporal dynamics.

Repetition: experiments need to be repeated in a reliable fashion to obtain a statistically reliable estimate for the biological variability of the system and other sources of error induced by the experimental setup.

In particular, the last aspect is often neglected in biology both in terms of the necessary number of experiment repetitions, as well as the false integration of data from inconsistent samples. This can lead, for example, to many false-positive or false-negative results in public genomic databases.

Regardless of the approach, one needs to define the level of model abstraction, complexity, and spatiotemporal scale of the system under investigation. The level of

abstraction leads to certain dynamics being modeled and simulated in detail, while for other parts, the details can be neglected, for example, due to low sensitivity toward specific parameter values, or considered as black boxes characterized by their input/output behavior.

The major goal of theory and model simulations is reverse engineering of biological systems, which is "the process of analyzing a system to identify its components and their interrelationships and create representations of the system in another form or at a higher level of abstraction." However, biological systems pose a major challenge toward reverse engineering due to the hidden complexities, inherent robustness, and possibly suboptimal design of functional units. Circular causality makes it hard to distinguish between cause and effect from biological data. All of this imposes ambiguities so as to deduce the correct biological "wiring diagram" from the experimental data. It is the hope that systems thinking in biology with its concepts of robustness, hierarchy and modularity, and protocols will provide detailed bottom-up models with testable hypotheses for model discrimination.

Different tools are available for the translation of experimental data into the language of mathematics, whose use depends on the degree of model abstraction and the spatiotemporal resolution of the system under consideration. In the following, we will focus on the motivation of network biology, as an example of system-level dissection of biological systems from an organizational point of view.

4.1 Network Biology

Network biology attempts to deduce the internal organization of a biological system from its organizational network topology that can be understood easily with relatively few and simple mathematical rules. It is hoped that fundamental relationships will emerge from essentially statistical analyses of large genomic and proteomics databases through this system-level approach.

Graph theoretic concepts have long since been used by biologists to visualize the interactions of genes, proteins, and metabolites. Such graphs consist of a set of nodes forming an interaction network. Usually, the nodes stand for the state variables of the system, like molecule concentrations, while the connections define the dynamics, that is, the type of interaction between them. Emergent phenomena then arise from the interplay of local and global dynamics due to the nodes and connections, respectively.

Networks are characterized by the connectivity, path length, and clustering distributions. Connectivity tells how many neighbors each node has on average. The path length denotes the average separation between arbitrarily chosen nodes, and the clustering coefficient is a measure of the grouping tendency of the nodes. Apart from these network measures, it is important to investigate the network architecture. Depending on the way the various nodes organize themselves into a network, different final network topologies appear.

Random networks appear by randomly connecting pairs out of a fixed set of nodes. It can be shown that the connectivity distribution follows a Poisson statistic, while the mean path length increases proportional to $\log N$, where N is the number of nodes. This type of network shows a small-world effect: any two nodes can be connected with just a few links.

Scale-free networks possess few highly connected nodes, called *hubs*, while many

nodes share only few connections. Hence, the connectivity distribution follows a power-law, that is, $P(k) \propto k^{-\gamma}$, where k and $P(k)$ denote the connectivity and distribution, respectively, and γ is the connectivity exponent. Such networks appear, if new nodes have a preferential attachment to already highly connected hubs. They possess ultrashort path lengths proportional to $\log \log N$.

Scale-free hierarchical networks are constructed out of a growth process, which iteratively replicates and loosely connects clusters or highly connected nodes, while also establishing links to clusters of previous grow steps. These rules assure a hierarchical structure of ever-smaller clusters down to an original seeding cluster.

High-throughput data collection has empowered the biological community to draw ever more detailed interaction webs of protein–protein, metabolite and gene transcription networks. Network biology relates to the field that attempts to discover universal design and organization principles, which govern the functioning and evolution of inner- and intracell networks. It has been discovered recently that all biological networks share certain topologies best described by scale-free networks, usually having exponents of $2 < \gamma < 3$. Interestingly, this type of complex network also emerges in other aspects of life and society, such as the World Wide Web and social interaction webs. Moreover, metabolic networks in the cell seem to share the property of ultrasmall world effects.

First experimental results on yeast and *E. coli* seem to support the view that metabolic as well as genomic regulatory networks show a hierarchical organization with few recurrent subnetworks called *motifs* that hint at the existence of elementary regulatory units. While modules are discrete functional units that are semi-detached from the whole system, motifs comprise a set of genes or metabolites that form recurring, significant patterns of interconnections, inseparable from the rest of the system. Ravasz et al. showed that the metabolic networks of 43 distinct organisms have a modular organization interconnected in a hierarchical manner – a system-level cellular organization that might be generic in nature.

The true reason for the preferred occurrence of scale-free networks in nature still remains elusive. Yet, there are a few possible explanations. Scale-free networks come about due to growth with preferred attachment. Considering the evolutionary history of metabolic networks, they might have grown through the addition of ever more metabolites to highly connected nodes already present in the system. Scale-free networks are also preferable with respect to functional robustness. Randomly deleting nodes from the network will most likely hit a loosely connected node, while there is a low probability that one deletes a hub.

The need to investigate the interconnections of genes–proteins and proteins–proteins stems from the fact that the genome of various species is quite similar, whereas it has been argued that the evolution of ever more complex species is correlated with an increased functional connectivity between a constant number of genes. Motifs might provide a way to increase interconnectivity of existing proteins, this way creating new functionalities. "This is likely one of several reasons that the apparent complexity of organisms can increase so markedly without a corresponding increase in gene number. An attribute of proteins encoded by the human genome is that they have a richer assembly of domains than do their counterparts

in invertebrates or yeast, and indeed the assortment of domains into novel combinations is likely an important aspect of genome divergence.''

Usually, only an incomplete knowledge about the system under investigation is available. In these cases, Bayesian networks can ameliorate the insufficient knowledge by modeling the known and unknown interactions in a directed, probabilistic cause–effect graph. The Bayes' rules then allow the researcher to combine new hypotheses with the existing knowledge from experiment. Likewise, the Bayesian networks allow to induce those network structures that best match the known expression data of a given regulatory network.

From the success of network biology, Oltvai and Barabasi propose to view the organization of cellular function in terms of a complexity pyramid. Accordingly, cellular functions are separated into groups of heterogeneous components and function, all interacting within large networks. Each higher functional cellular level, starting from molecular components like DNA to metabolic pathways, functional modules to large-scale network organization, is ever more universally applicable to a multitude of species. This way, network biology provides a new view on the organizational structure, behavior, and function of the cell. It similarly bestows new approaches of bottom-up and top-down modeling, that is, moving from molecules to motifs and modules, or starting from the other end with the construction of scale-free, hierarchical networks and then including organism-specific modules and motifs. Altogether, this global organization seems to be a hint for an evolutionary family of circuits with recurring dynamic organizational principles, which opens up the possibility of establishing ''a periodic table for functional regulatory circuits.''

4.1.1 Dynamic Network Models

Despite the intuitive appeal of network biology toward the structure identification of biological systems, it is limited in the sense that it does not include the temporal dynamics of the system. Boolean networks are an abstract, yet viable, approach toward the inclusion of a temporal evolution of large-scale networks, while at the same time being easy to handle both analytically and numerically. Each node can assume an ON/OFF value. Then, the state of each node at the next time step $(t+1)$ is deduced deterministically from a logical Boolean function, that is, *AND*, *OR*, *NAND*..., based on its current state and external input. While Boolean networks have been successfully used for the analysis of gene regulatory networks, they completely neglect the continuous nature of concentration gradients or stochastic effects. The inclusion of these into model simulations is the topic of the following sections.

4.2 Reaction-diffusion Models

Differential equations provide a mathematical description of system dynamics with continuous system states in time and space. They are the most widespread formalism to model systems throughout the various scientific domains, having a long history of more than 300 years. In biology, they are widely used to describe the time and space course of molecular concentrations with the interaction between the various reactants being modeled through functional and differential relations. There is a vast literature available on modeling

biochemical reactions based on differential equations, especially in the context of metabolic processes.

In general, one distinguishes between purely time-dependent ordinary differential equations (ODE), and more general partial differential equations (PDE), which additionally include spatial dimensions. A PDE describing the spatiotemporal evolution of a system has the general form

$$\frac{\partial \psi_i(\mathbf{r}, t)}{\partial t} = \underbrace{f[\psi_i(\mathbf{r}, t)]}_{\text{temporal evolution}} + \underbrace{\nabla \mathbf{D_i}(\mathbf{r}, t) \nabla \psi_i(\mathbf{r}, t)}_{\text{spatial Diffusion}}, \quad (1)$$

where $\psi_i(\mathbf{r}, t)$ denotes the respective system state variables of the various molecules labeled by the subscript i. The above equation reduces to an ODE in the absence of spatial diffusion, which is $\mathbf{D}(\mathbf{r}, t) = 0$. The term $f[\psi_i(\mathbf{r}, t)]$ denotes the respective synthesis rates, depending on the various concentrations of $\psi_i(\mathbf{r}, t)$ and possibly external signals. It is often composed of the various, mostly nonlinear, interaction functions between the system metabolites based on the "law of mass action" of metabolic processes, for example, Michaelis–Menten-like enzymatic degradation, inhibition of the Hill type.

The spatial aspect of the equation enters the description due to the inclusion of a diffusion term, where the diffusion term $\mathbf{D_i}(\mathbf{r}, t)$ depends on space and time in general. Alan Turing was the first to point out such reaction-diffusion systems as a possible explanation for morphogenesis in natural systems. He showed theoretically that a system of reacting and diffusing chemicals can evolve spontaneously to a spatially heterogeneous state as a response to an infinitesimally small forcing. This pioneering work led to a new branch of research and many so-called Turing systems have been proposed (however, not finally been proven) to account for pattern formation in developmental biology, for example, explaining the pattern formation on snail houses, modeling Drosophila embryogenesis, or explaining the coating patterns of animals.

The importance of spatiotemporal inhomogeneities in molecular concentrations and molecular crowding in signal transduction pathways has recently been acknowledged. As a consequence, cellular functions also employ on active transport mechanisms to sustain reliable cellular processes.

Owing to the nonlinear interaction terms $f[\psi_i(\mathbf{r}, t)]$, an analytical solution of the differential equation is possible in rare cases, only. However, a stability analysis is always feasible and gives a good insight into the qualitative behavior of the local stability of the steady state solutions and the occurrence of periodic solutions in space and time.

In these cases, a quantitative insight into the respective equations can be obtained from numerical simulations. A variety of numerical integration algorithms can be found in standard literature; moreover, several computer software packages have been developed specifically for the simulation and bifurcation analysis of nonlinear systems, such as GEPASI or DBSolve.

4.3 Intracellular Noise Models

The above approach of differential equations for the mechanistic description of biological phenomena implicitly relies on the "law of mass action." This law relates the molecular component concentrations to reaction rates. It is valid in homogeneous, well-mixed environment, assuming the underlying reactions to be continuous

and deterministic. Under these prerequisites, the complete time-course of the reactions is determined by the initial conditions.

Deducing effective reaction rates from molecular interactions fails in the case of low particle numbers. There, the discrete and random nature of the individual, elementary reactions between molecules inevitably comes forth, showing its impact on a macroscopic scale as fluctuations in the molecule concentration over time. In molecular biology, this intrinsic noise becomes most significant in the regulation of gene expression, as it is usually accompanied with low copy numbers of mRNA transcripts and the genes themselves. Stochastic effects are attributed to be the cause of phenotypic diversity in isogenetic populations of cells. They play a major role in the lysis-lysogeny decision circle of the λ-phage and are shown to be a decisive element in the threshold behavior of CD95-induced apoptosis.

Nature both exploits and copes with noise. Regulatory determinism of gene expression and thus reduction of noise is achieved by redundant gene configurations, highly improving the regulation reliability. Another strategy is the embedding of the respective gene function into a pool of many "background" reaction partners.

At the same time, there exists a "positive" effect of noise in biological systems. Noise-induced phenotypic diversity increases the likelihood of species survival within a wider range of environmental parameters and allows for a more rapid evolutionary adaptation to environmental pressure. The constructive and ordering role of noise in terms of information transfer is long since known from the phenomenon of Stochastic Resonance (SR). Experiments have shown that neurons increase their sensitivity toward weak input signal through the addition of ambient noise. Likewise, Paulsson et al. theoretically deduced an effect called *stochastic focusing*. Cellular mechanisms might show an increased sensitivity toward input signals as a result of additional noise-induced degrees of freedom.

The analytical treatment of stochasticity in chemical reactions starts from defining a joint probability distribution $p(\mathbf{X}, t)$ of the system, where $\mathbf{X} = [X_1, \ldots, X_n]$ denotes a discrete number of n different molecule types at time t. The temporal evolution of p is described by a master equation, which is a gain–loss equation for the probabilities of the separate states $\mathbf{X}$

$$\frac{\partial p(\mathbf{X}, t)}{\partial t} = \sum_{j=1}^{m} \left(\beta_j - \alpha_j p(\mathbf{X}, t)\right), \qquad (2)$$

where m different elementary reactions are considered with $\alpha_j \Delta t$ and $\beta_j \Delta t$ denoting the loss and gain terms, respectively. The former are the probabilities that some reaction j occurs within the time interval $[t, t + \Delta t]$, while the latter is the gain due to transitions from other states to $p(\mathbf{X}, t)$. Albeit the intuitive simplicity of the master equation approach, its analytical solution is mostly impossible and even its numerical manageability is limited.

An alternative to the evaluation of the master equation is the stochastic simulation algorithm (SSA), which is the numerical approximation to the microscopic behavior of reacting chemicals in a "well-mixed" environment. The algorithm randomly chooses the reaction type and the integration step size on the basis of the current rate constants and population size of the chemical reactants. The relative probabilities for each elementary reaction

are derived from the respective propensity functions, which relate the reaction probability to the possible molecule combinations and their effective reaction speeds. One simulation run then consists of iteratively calculating the probability density according to the propensity functions, updating the integration time and chemical populations with respect to the randomly chosen step size and reaction stoichiometry, respectively. It can be shown that the SSA leads to the correct master equation representation, such that the average of many SSA realizations converge to the correct probability density $p(\mathbf{X}, t)$.

No matter how useful the SSA algorithm is, it has a main limitation stemming from the need for detailed knowledge about every single reaction mechanism as well as the requirement for only one reaction to take place within a given time interval, which can be very small indeed. Furthermore, several simulation runs need to be performed in order to obtain valid averages of the probability density $p(\mathbf{X}, t)$. These conditions render the algorithm computationally demanding, especially in the case of large numbers of molecule species. Therefore, various refinements have been proposed for the above algorithm that either aim at reducing the number of calculations at each integration step or increasing the time step size by allowing several reactions to take place at once.

Stochastic modeling converges to the "Law of Mass Action" in the case of large molecule numbers, where the fluctuation reaction probabilities can be replaced by effective kinetic parameters. In between, there exists an intermediate range, however, in which the discrete structure of reaction events becomes insignificant, yet the stochastic effects still prevail on the macroscopic scale. Then, the dynamics become quasi time-continuous and can be described by a stochastic differential equation (SDE), called *chemical Langevin equation*, that is, an ordinary differential equation with an additional noise term. While the numerical simulation of SDEs is mathematically well established, uncertainties persist concerning the validity range of this approximation as well as statistical representation of the ad hoc noise term added to the equations.

To reduce the computational complexity of stochastic simulation, hybrid simulation systems have been proposed recently. Noting the fact that molecule numbers in the cell can vary substantially over time, it is possible to simply apply efficient ODEs or SDEs for system integration at times of large and intermediate molecule numbers, while switching to discrete stochastic simulation, for molecules with a number below a certain threshold. This way, the crucial effects of stochasticity at low particle numbers are conserved while maintaining acceptable losses in accuracy.

4.4 Modeling Resources

The development of a standardized computer infrastructure is of utmost importance to manage the increasing amount of knowledge on biological systems. This helps the effective utilization of resources and the exchange of models, ideas, and data. An integrative software tool for Systems biology should comprise the following features for proper systems understanding. From a mathematical point of view, such tools need support for different simulation algorithms, such as deterministic and stochastic ordinary and partial differential equation solver and should include analysis algorithms for parameter estimation and model discrimination. From an experimental point of view, it

should have support for a standardized modeling language. In terms of software, the simulation package should run independently from the computer platform and preferably on a computer grid or cluster environment.

A number of simulation packages are currently under active development, most of them freely available. As a consequence of the complexity encountered in biological systems, there is no integrative software package, capable of handling all phenomena in signal transduction, metabolic pathways, or spatiotemporal simulations. Some of the currently most versatile packages among many are ECell, Virtual Cell, or Cellware. Simulation packages for truly spatiotemporal simulation of biological systems are currently under active development.

Model building is a complex enterprise and is usually accomplished in collaboration with scientists from different research institutions. Serving the need for seamless information exchange of computer-based models, common standard markup languages based on the XML format are currently developed. The most popular ones are CellML, BioPAX, and SBML. The Systems biology workbench extends the above approach by providing a standardized application interface for researchers to not only exchange model data and results but also the simulation tools themselves.

Another great challenge lies in the inherent modeling of stochasticity of inner cell processes. The SSA algorithm from Gillespie is computationally infeasible with an increasing number of molecules, and the computational power for exact stochastic simulations will be immense. Therefore, different algorithms have been developed, like reducing the number of random variables for simulation or allowing to take larger integration time steps at a justifiable error in the respective reaction probabilities. Several program packages for the simulation of stochastic molecular dynamics exist, for example, StochSim or Stocks. Programs like SmartCell tackle the simulation of spatially extended stochastic systems.

Regardless of the modeling approach of biological systems there is always the need for parameter estimates of at least a few constituents of the model. Because of the sheer amount of parameters required for successful modeling on one hand and the huge number of experiments already performed, the need to organize experimental data has brought about several publicly available databases of molecular properties, interactions, and pathways. They provide an invaluable infrastructure for future modeling efforts, enabling the modeler to start simulations from a certain degree of abstraction.

Spatiotemporal modeling will additionally require information about the physical structure of its model constituents as obtained from microscopy. For these needs, the development of an Open Microscope Environment, which is currently being built as a joint effort amongst various research institutions, will provide a unified data format and database environment for consistently annotating, storing, and retrieving five dimensional microscope images (four dimensions in space and time with additional color information) and exchanging them between research institutions.

5
Future Prospects of Systems Biology

The systematic synthesis of experimental data into model descriptions nourishes the

hope that biology will one day become a branch of science close to engineering. Once the biologists have learned the "language of the genes," that is, their syntax and their semantics, they will have the theoretical background knowledge and the experimental expertise to draft and create synthetic model cells or even organisms from scratch, on which new drugs and cures can be tested *in silico* before a possible assembly *in vivo*.

The impact of Systems biology will change medical practice. With the advent of functional understanding of living systems, it will be possible to devise new anticancer therapies that first identify the particularities of heterogeneous tumor cell populations and then try to predict and control cell activity to detect and use specific fragile points of each tumor cell type. The current problem with chemotherapy is the multidrug resistance of most tumor cells as well as the drugs' mutagenic property. They kill part of the cancer population, while fostering the diversification of other cancer cell strains. Thus, viewing cancer from a systemic level as a robust system instead might provide physicians with a framework for future anticancer strategies.

5.1 Synthetic Cell Biology

Currently, genetic engineering of organisms is mostly done by inserting genes into the DNA of plants and animals, yet the ultimate goal will be the design of living matter from scratch. Arkin claims that molecular biology has entered a stage of maturity that requires its transformation into an engineering discipline. Synthetic biology is a future vision for coming decades based on the current progress in cellular, molecular biology as well as genetics and associated fields of engineering and computer science. With synthetic biology being the technological counterpart of Systems biology, it is the hope that their combined knowledge will enable the creation of essentially artificial systems using biological design principles with new combinations of building modules from existing (sub)cellular systems.

Synthetic biology departs from a component-based approach. It views a living system as being composed of interacting modules, each having particular functions that exchange their information via protocols. Current research suggests that these modules are limited in number, albeit their specific tasks being diverse. Synthetic biology implies a scientific agenda on a higher level of abstraction toward identifying and categorizing the various module types within the cell and across organisms, investigating their interactions on the basis of the module, rather than molecule-interaction. From an engineering point of view, this means that modules can be taken out of their current evolutionary context and assembled differently in the sense of versatile building blocks.

As a first step, there are efforts to characterize the workings of switches, bi-stable systems or stabilizing control loops from simple chemical reaction schemes, in a systematic way. Different authors describe the fundamental principles of building logical circuits in the language of gene regulatory networks. Tyson et al. present a systematic overview of designs for biological control systems such as switches, sniffers, or buzzers, and combine the mathematical description with experimental findings. Hasty et al. review the possibilities for constructing gene circuits serving various functions such as autoregulation, repression or logical gates. In a first *de novo* design study of a gene regulatory network,

Guet et al. systematically analyzed the phenotypic behavior of different parameter and topology combinations in a genetic regulatory network. They constructed various logical gates, such as NAND, NOR, and NOT IF, through the combination of three nodes together with five promoters. As an important result, they found that both parameter values as well as network topology are important to unambiguously determine the systems' computational function.

A Biological Information System (BIS) is to be established in the long run, extending genomic databases with quantitative mechanistic knowledge. Using all data in an integrated form, it will be possible to build refined theories by checking on key molecules, replacing complex pathways with effective reaction parameters, and thus defining biologically meaningful reaction subsets from the large amount of possible reactions.

Kitano calls for the establishment of an international Systeome project (alluding to the joint efforts of the Human Genome Project), which aims at promoting Systems biology by creating such a comprehensive data resource of biological systems features. He defines this grand challenge as "the assembly of system profiles for all generic variations and environmental stimuli responses." Several model organisms, like yeast, *Caenorhabditis elegans*, *Drosophila C.*, the mouse and finally the human shall serve as the "rosetta stones" for the establishment of a working *in silico* simulation model of life, whose establishment is hoped to be finished by the year 2030.

One further important aspect of Synthetic biology is the possible risk analysis of current and future artificial genetic mutations in plants and animals. At this point, nobody can assess the long-term risks of releasing new artificial gene pools into the wild. Erwin Chargaff, who pioneered our understanding of the DNA, took a rather pessimistic view on the disproportionate ratio between technological possibilities of genetic engineering and the knowledge about its implications:

> "Bacteria and Viruses have always formed a most effective biological underground. The Guerilla warfare through which they act on higher forms of life is only imperfectly understood. By adding to this arsenal freakish forms of life – procaryotes propagating eukaryotic genes – we shall be throwing a veil of uncertainties over the life of coming generations. Have we the right to counteract, irreversibly, the evolutionary wisdom of millions of years, in order to satisfy the ambition and curiosity of a few scientists? [. . .] My generation, or perhaps the one preceding mine, has been the first to engage, under the leadership of the exact sciences, in a destructive colonial warfare against nature. The future will curse us for it."

Albeit the above quote dates from 1976, it is still relevant, for example, with respect to the ongoing debate on the commercial use of genetically mutated crop. Chargaff's pessimism is not without foundation. Genetics has established the techniques for altering systematically the genetic makeup; however, a deeper understanding on the effects of manipulating so intricate a design as the cell cycle is still lacking. A Systeome project is thus the logical and necessary sequitur of the era of genomics. It will help clarify the consequences of gene transfer and inheritance in order to avoid potentially dangerous mutations.

5.2 Conclusions – Where are We?

The headline question of this section is easily answered: we are at the very beginning of a new branch of science. The road ahead is clearly marked and commonly agreed on

by many people. Yet, the trouble in all processes lies with the detail of things. The authors are not saying that the goals of Systems biology cannot be achieved, though its requirements are so outstanding that they call for drastic measures. This starts with the need for the scientists of different disciplines to learn a common scientific language as well as to have a definition of common goals.

To date, most experimental biosciences are method-driven, paying a lot of attention to details and produce, in extreme cases, loads of data out of context. Theoretical sciences on the other hand are principle-driven, neglecting important details, and thus coming up with theorems and models out of this world. As a consequence, all researchers will have to learn to focus on the important parts of their respective systems that are under investigation, yet varying which part to measure in detail and which part to neglect.

The biologists must adopt abstract thinking, trusting in the language of mathematics and necessary simplification. The physicists must learn to abolish oversimplified thinking and to get used to the analysis of strongly interacting systems with many degrees of freedom. Engineers must gain a deeper understanding of their systems under investigation beyond numerical solutions, for example, in terms of process optimization. The necessary unification of the various branches of science for the sake of the advancement of Systems biology will make communication skills among scientists ever more important. Years of ''isolated'' research have diversified the scientific vocabulary and methods, making it sometimes hard to mediate and discuss interdisciplinary goals and methods. Furthermore, scientists must learn to put their own research into the bigger frame of interdisciplinary research and research teams.

Despite this increased complexity of systems, there is still the need for an intuitive understanding of the relevant elements of the system under study. Complex computer models of biological systems lacking intuitive understanding can be regarded as an isomorphic mapping from one unknown system *in vivo* to another unknown system *in silico*. Unfortunately, despite a certain reliability factor in the agreement of experiment and simulation, it is impossible to finally prove the validity of such a model. For as long as there is a lack of complete understanding, there is always the possibility of new insights, which might render the current concept false. Vilar et al. even argue, using the example of the lac operon that ''even in the 'postgenomic era' [modeling] will still rely more on good intuition and skills of quantitative biologists than on the sheer power of computers.''

It is the strong belief that Systems biology will promote a unification of sciences, if the community is working ''holistically'' in a joint effort of biology, mathematics, physics, and engineering. Then, a new generation of scientists with an interdisciplinary training will emerge, whose everyday business will comprise working in the lab as well as performing data analysis and model simulations, exchanging data and results freely through publicly available databases. With the establishment of new experimental techniques and mathematical tools, asking the right questions, we can be sure that the systems approach to biology will be a great success. It will not only change modern life sciences but it will also change the view on life itself and finally we could prove the late Cassandra, Mr. Chargaff, wrong.

See also Genetics, Molecular Basis of; Informatics (Computational Biology); Metabonomics and Metabolomics; Molecular Systematics and Evolution; Nucleic Acid and Protein Sequence Analysis and Bioinformatics; Pragmatic Computational Biology: Sequence Analysis and Biological Systems Description; Protein Repertoire, Evolution of.

Bibliography

Books and Reviews

Barabàsi, A.L., Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization, *Nat. Genet.* **5**, 101–114.

Bertalanffy, L. (1973) *General Systems Theory*, Penguin, Harmondsworth.

Ideker, T., Galitski, T., Hood, L. (2001) A new approach to decoding life: systems biology, *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372.

Kitano, H. (Ed.) (2001) *Foundations of Systems Biology*, MIT Press, Cambridge.

Nature Insight. (2002) Computational biology, *Nature* **420**, 205–251.

Ross Ashby, W. (1956) *Introduction to Cybernetics*, Chapman & Hall, London, Available electronically http://pespmcl.vub.ac.be/books/ IntroCyb.pdf.

Science special issue. (2002) Systems Biology, *Science* **295**, 1661–1682.

Wiener, N. (1948) *Cybernetics: Control and Communications in the Animal and the Machines*, MIT Press, Cambridge.

Primary Literature

Adami, C. (2002) What is complexity? *BioEssays* **24**, 1085–1094.

Alm, E., Arkin, A.P. (2003) Biological networks, *Curr. Opin. Struct. Biol.* **13**, 193–202.

Almaas, E., Kovács, B., Vicsek, T., Oltvai, Z.N., Barabási, A.-L. (2004) Global organization of metabolic fluxes in the bacterium Escherichia coli, *Nature* **427**, 839–843.

Alon, U., Surette, M.G., Barkai, N., Leibler, S. (1999) Robustness in bacterial chemotaxis, *Nature* **397**, 168–171.

Ander, M., Beltrao, P., Diventura, B., Foglierini, M., Lemerle, C., Oliveira, I., Kaplan, A., Serrano, L. (2004) Smartcell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks, *Systems Biology* **1**, 129–138.

Angeli, D., Ferrell, J.E., Sontag, E.D. (2004) Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1822–1827.

Arkin, A.P. (2001) Synthetic cell biology, *Curr. Opin. Biotechnol.* **12**(6), 638–644.

Arkin, A., Ross, J., McAdams, H.H. (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells, *Genetics* **149**(4), 1633–1648.

Bak, P., Tang, C., Wiesenfeld, K. (1987) Self-organized criticality: an explanation of the 1/f noise, *Phys. Rev. Lett.* **59**, 381–384.

Barabàsi, A.-L., Albert, R. (1999) Emergence of scaling in random networks, *Science* **286**, 509–512.

Barkai, N., Leibler, S. (1997) Robustness in simple biochemical networks, *Nature* **387**, 913–917.

Bentele, M., Eils, R. (2004) General stochastic hybrid method for the simulation of chemical reaction processes in cells. *CMSB04, Lecture Notes in Computer Science*, Springer, Heidelberg.

Bentele, M., Lavrik, I., Ulrich, M., Stosser, S., Heermann, D.W., Kalthoff, H., Krammer, P.H., Eils, R. (2004) Mathematical modeling reveals threshold mechanism in cd95-induced apoptosis, *J. Cell. Biol.* **166**, 839–851.

BioPAX Homepage. Internet 2004. http://www.biopax.org/.

Carlson, J.M., Doyle, J. (1999) Highly optimized tolerance: a mechanism for power laws in designed systems, *Phys. Rev. E.* **60**(2), 1412–1427.

Carlson, J.M., Doyle, J. (2002) Complexity and robustness, *Proc. Natl. Acad. Sci. U.S.A.* **99**(Suppl. 1), 2538–2545.

Chargaff, E. (1976) The dangers of genetic meddling, *Science* **192**, 938–940.

Chikofsky, E.J., Cross, J.H. (1990) Reverse engineering and design recovery: a taxonomy, *IEEE Software* **7**, 13–17.

Christopher, R., Dhiman, A., Fox, J., Gendelman, R., Haberichter, T., Kagle, D., Spizz, G., Khalil, I.G., Hill, C. (2004) Data-driven computer simulation of human cancer cell, *Ann. N. Y. Acad. Sci.* **1020**, 132–153.

Clarke, D.J., Giménez-Abián, J.F. (2000) Checkpoints controlling mitosis, *BioEssays* **22**(4), 351–363.

Claverie, J.-M. (2000) Do we need a huge new centre to annotate the human genome? *Nature* **403**, 12.

Crampin, E.J., Schnell, S., McSharry, P.E. (2004) Mathematical and computational techniques to deduce complex biochemical reaction mechanisms, *Prog. Biophys. Mol. Biol.* **86**, 77–112.

Cross, M.C., Hohenberg, P.C. (1993) Pattern formation outside of equilibrium, *Rev. Mod. Phys.* **65**(3), 851–1112.

Csete, M.E., Doyle, J.C. (2002) Reverse engineering of biological complexity, *Science* **295**, 1664–1669.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z.j., Schilstra, M.J., Clarke, P.J.C., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H. (2002) A genomic regulatory network for development, *Science* **295**, 1669–1678.

de Jong, H. (2002) Modeling and simulation of generic regulatory systems: a literature review, *J. Comput. Biol.* **9**(1), 67–103.

Demirel, Y., Sandler, S.I. (2002) Thermodynamics and bioenergetics, *Biophys. Chem.* **97**, 87–111.

Dhar, P.K., Zhu, H., Tan, C.M., Somani, S., Li, Y., Mishra, S.K. (2005) Modeling and simulation in systems biology: the genesis and evolution of challenges, strategies and methods, Internet http://web.bii.astar.edu.sg/~somani/documents/ieee_cw1.pdf.

Dhar, P., Meng, T.C., Somani, S., Ye, L., Sairam, A., Chitre, M., Hao, Z., Sakharkar, K. (2004) Cellware – multi-algorithmic software for computational systems biology, *Bioinformatics* **20**, 1319–1321.

Douglass, J.K., Wilkins, L., Pierson, D., Pantazelou, E., Moss, F. (1993) Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance, *Nature* **365**, 337–340.

Elowitz, M., Levine, A., Siggla, W., Swain, P. (2002) Stochastic gene expression in a single cell, *Science* **297**, 1183–1186.

Elowitz, M.B., Surette, M.G., Wolf, P.-E., Stock, J.B., Leibler, S.J. (1999) Protein mobility in the cytoplasm of Escherichia coli, *J. Bacteriol.* **181**, 197–203.

Endy, D., Brent, R. (2001) Modelling cellular behavior, *Nature* **409**, 391–395.

François, P., Hakim, V. (2004) Design of genetic networks with specified functions by evolution in silico, *Proc. Natl. Acad. Sci. U.S.A.* **101**(2), 580–585.

Freeman, M. (2000) Feedback control of intercellular signalling in development, *Nature* **408**, 313–319.

Fulka, J., Tesařík, J., Loi, P., Moor, R.M. (2000) Manipulating the human embryo: cell cycle checkpoint controls, *Cloning* **2**(1), 1–7.

Gammaitoni, L., Hänggi, P., Jung, P., Marchesoni, F. (1998) Stochastic resonance, *Rev. Mod. Phys.* **70**(1), 223–287.

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., Superti-Furga, G. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes, *Nature* **10**, 123,124.

Gibson, M.A., Bruck, J. (2000) Efficient exact stochastic simulation of chemical systems with many species and many channels, *J. Phys. Chem.* **104**, 1876–1889.

Gierer, A., Meinhardt, H. (1972) A theory of biological pattern formation, *Kybernetik* **12**, 30–39.

Gillespie, D.T. (1977) Exact stochastic simulation of coupled chemical reactions, *J. Phys. Chem.* **81**, 2340–2361.

Gillespie, D.T. (2001) Approximate accelerated stochastic simulation of chemically reacting systems, *J. Chem. Phys.* **115**, 1716–1733.

Goryanin, I., Hodgman, T.C., Selkov, E. (1999) Mathematical simulation and analysis

of cellular metabolism and regulation, *Bioinformatics* **15**(9), 749–758.

Grassberger, P. (1986) Toward a quantitative theory of self-generated complexity, *Int. J. Theor. Phys.* **25**, 907–928.

Gu, X. (2003) Evolution of duplicate genes versus genetic robustness against null mutations, *Trends Genet.* **19**, 354–356.

Guet, C.C., Elowitz, M.B., Hsing, W., Leibler, S. (2002) Combinatorial synthesis of genetic networks, *Science* **296**, 1466–1470.

Guimerà, R., Arenas, A., Díaz-Guilera, A. (2001) Communication and optimal hierarchical networks, *Physica A* **299**, 247–252.

Haken, H. (1987) *Advanced Synergetics*, Springer, Berlin.

Hartman, J.L., Garvik, B., Hartwell, L. (2001) Principles for the buffering of genetic variation, *Science* **291**, 1001–1004.

Hartwell, L.H., Hopfield, J.J., Leibler, S., Murray, A.W. (2002) From molecular to modular cell biology, *Nature* **420**, C47–C52.

Hasty, J., McMillen, D., Collins, J.J. (2002) Engineered gene circuits, *Nature* **420**, 224–230.

Haubelt, A., Bullinger, E., Sauter, T., Allgöwer, F., Gilles, E.D. (2005) *Systems Biology – A Glossary from two Perspectives*, Internet http://www.sysbio.de/.

Heinrich, R., Schuster, S. (1996) *The Regulation of Cellular Systems*, Chapman & Hall, New York.

Heylighen, F., Joslyn, C., Turchin, V. (2005) Principia cybernetica web. Internet http://pespmc1.vub.ac.be/.

Ho, M.-W. (1999) Seeking clarity in the debate over the safety of gm foods, *Nature* **402**, 575–576.

Howard, M., Rutenberg, A.D. (2003) Pattern formation inside bacteria: fluctuations due to the low copy number of proteins, *Phys. Rev. Lett.* **90**(1–4), 128102.

Huang, S. (1999) Gene expression profiling, genetic networks, and cellular states: an integrating concept for tumorigenesis and drug discovery, *J. Mol. Med.* **77**, 469–480.

Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J., Kitano, H. (2002) The ERATO Systems Biology Workbench: enabling interaction and exchange between software tools for computational biology, *Pac. Symp. Biocomput.* 450–461.

Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., Cuellar, A.A., Dronov, S., Gilles, E.D., Ginkel, M., Gor, V., Goryanin, I.I., Hedley, W.J., Hodgman, T.C., Hofmeyr, J.-H., Hunter, P.J., Juty, N.S., Kasberger, J.L., Kremling, A., Kummer, U., Le Novre, N., Loew, L.M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E.D., Nakayama, Y., Nelson, M.R., Nielsen, P.F., Sakurada, T., Schaff, J.C., Shapiro, B.E., Shimizu, T.S., Spence, H.D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., Wang, J. (2003) The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models, *Bioinformatics* **19**, 524–531.

Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R., Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network, *Science* **292**, 929–934.

Jeong, H., Mason, S.P., Barabási, A.-L., Oltvai, Z.N. (2001) Lethality and centrality in protein networks, *Nature* **411**, 41–42.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabási, A.-L. (2000) The large-scale organization of metabolic networks, *Nature* **407**, 651–654.

Kauffman, S.A. (1993) *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, New York.

Kearns, W.G., Liu, J.M. (2001) Cell cycle checkpoint genes and aneuploidy: a short review, *Curr. Genomics* **2**(2), 171–180.

Kholodenko, B.N. (2003) Four-dimensional organization of protein kinase signaling cascades: the roles of diffusion, endocytosis and molecular motors, *J. Exp. Biol.* **206**, 2073–2082.

Kierzek, A.M. (2002) Stocks: stochastic kinetic simulations of biochemical systems with Gillespie algorithm, *Bioinformatics* **18**(3), 470–481.

Kitano, H. (2002a) Computational systems biology, *Nature* **420**, 206–210.

Kitano, H. (2002b) Systems biology: a brief overview, *Science* **295**, 1662–1664.

Kitano, H. (2004) Cancer as a robust system: implications for anticancer therapy, *Nat. Rev. Cancer* **4**, 227–235.

Kloeden, P.E., Platen, E. (1999) *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin.

König, R., Eils, R. (2004) Gene expression analysis on biochemical networks using

the Potts spin model, *Bioinformatics* **20**, 1500–1505.

Laub, M.T., McAdams, H.H., Feldblum, T., Fraser, C.M., Shapiro, L. (2000) Global analysis of the genetic network controlling a bacterial cell cycle, *Science* **290**, 2144–2148.

Lauffenburger, D.A. (2000) Cell signaling pathways as control modules: complexity for simplicity? *Proc. Natl. Acad. Sci. U.S.A.* **97**(10), 5031–5033.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.L., Gordon, D.B., Ren, B., Wyrick, J.J., Tagne, J.-B., Volkert, T.L., Fraenkel, E., Gifford, D.K., Young, R.A. (2002) Transcriptional regulatory networks in Saccharomyces cerevisiae, *Science* **298**, 799–804.

Levine, M., Tjian, R. (2003) Transcription regulation and animal diversity, *Nature* **424**, 147–151.

Lindvall, J.M., Emelie, K., Blomberg, M., Edvard Smith, C.I. (2003) In silico tools for signal transduction research, *Brief. Bioinform.* **4**, 315–324.

Lippincott-Schwartz, J., Snapp, E., Kenworthy, A. (2001) Studying protein dynamics in living cells, *Nat. Rev. Mol. Cell Biol.* **2**, 444–456.

Lippincott-Schwartz, J., Altan-Bonnet, N., Patterson, G.H. (2003) Photo-bleaching and photoactivation: following protein dynamics in living cells, *Nature Cell Biol.* **5**, S7–S14.

Lloyd, C.M., Halstead, M.D.B., Nielsen, P.F. (2004) CellML: its future, present and past, *Prog. Biophys. Mol. Biol.* **85**, 433–450.

May, R.M. (2004) Uses and abuses of mathematics in biology, *Science* **303**, 790–793.

McAdams, H., Shapiro, L. (1995) Circuit simulations of genetic networks, *Science* **269**, 650–656.

McAdams, H.H., Arkin, A. (1999) It's a noisy business! – genetic regulation at the nanomolar scale, *Trends Genet.* **15**(2), 65–69.

Mendes, P. (1993) Gepasi: a software package for modelling the dynamics, steady states and control of biochemical and other systems, *Comput. Appl. Biosci.* **9**(5), 563–571.

Mesarovic, M.D. (1968) System theory and biology – view of a theoretician, in: Mesarovic, M.D. (Ed.) *Systems Theory and Biology*, Springer, New York, pp. 59–87.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U. (2002) Network motifs: simple building blocks of complex networks, *Science* **298**, 824–827.

Morishita, Y., Aihara, K. (2004) Noise-reduction through interaction in gene expression and biochemical reaction processes, *J. Theor. Biol.* **228**, 315–325.

Morton-Firth, C.J. (1998) *Stochastic Simulation of Cell Signalling Pathways*, PhD thesis, University of Cambridge, Cambridge.

Murray, J.D. (1993) *Mathematical Biology*, Springer, Heidelberg.

Oltvai, Z.N., Barabási, A.-L. (2002) Life's complexity pyramid, *Science* **298**, 763–764.

Swedlow, J., Sorger, P., Goldbert, I. (2005) *Open Microscopy Environment*. Internet http://www.openmicroscopy.org/.

Paulsson, J., Berg, O., Ehrenberg, M. (2000) Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation, *Proc. Natl. Acad. Sci. U.S.A.* **97**(13), 7148–7153.

Pawson, T., Nash, P. (2003) Assembly of cell regulatory systems through protein interaction domains, *Science* **300**, 445–452.

Press, W.H., Teukolsky, S.A. Vetterling, W.T., Flannery, B.P. (1993) *Numerical Recipes in C*, 2nd edition, Cambridge University Press, New York.

Rao, C.V., Wolf, D.M., Arkin, A.P. (2002) Control, exploitation and tolerance of intracellular noise, *Nature* **420**, 231–237.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, N., Barabási, A.-L. (2002) Hierarchical organization of modularity in metabolic networks, *Science* **297**, 1551–1555.

Schnell, S., Turner, T.E. (2004) Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws, *Prog. Biophys. Mol. Biol.* **85**, 235–260.

Searls, D.B. (2002) The language of the genes, *Nature* **420**, 211–217.

Sharom, J.R., Bellows, D.S., Tyers, M. (2004) From large networks to small molecules, *Curr. Opin. Chem. Biol.* **8**, 81–90.

Slepchenko, B.M., Schaff, J.C., Carson, J.H., Loew, L.M. (2002) Computational cell biology: spatiotemporal simulation of cellular events, *Annu. Rev. Biophys. Biomol. Struct.* **31**, 423–441.

Slepchenko, B.M., Schaff, J.C., Macara, I.G., Loew, L.M. (2003) Quantitative cell biology with the virtual cell, *Trends Cell Biol.* **13**, 570–576.

Smuts, J.C. (1926) *Holism and Evolution*, Macmillan, London.

Stelling, J., Sauer, U., Szallasi, Z., Doyle, F.J., Doyle, J. (2004) Robustness of cellular functions, *Cell* **118**, 675–685.

Strogatz, S.H. (2001) Exploring complex networks, *Nature* **410**, 268–276.

Swedlow, Jason.R., Goldberg, Ilya., Brauner, Erik., Sorger, Peter.K. (2003) Informatics and quantitative analysis in biological imaging, *Science* **300**, 100–102.

Thomann, D., Rines, D.R., Sorger, P.K., Danuser, G. (2002) Automatic fluorescent tag detection in 3d with super-resolution: application to the analysis of chromosome movement, *J. Microsc.* **208**, 49–64. PK, and

Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C., Hutchison, C.M. (1999) Ecell: software environment for whole cell simulation, *Bioinformatics* **15**(5), 72–74.

Turing, A.M. (1952) The chemical basis of morphogenesis, *Philos. Trans. R. Soc. London B*. **237**, 37–72.

Turner, T.E., Schnell, S., Burrage, K. (2004) Stochastic approaches for modelling in vivo reactions, *Comput. Biol. Chem.* **28**, 165–178.

Tyson, J.J., Chen, K.C., Novak, B. (2003) Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell, *Curr. Opin. Cell Biol.* **15**, 221–231.

van Kampen, N.G. (1992) *Stochastic Processes in Physics and Chemistry*, Elsevier Science Publishers, B. V., Amsterdam.

Vicsek, T. (2002) Complexity: the bigger picture, *Nature* **418**, 131.

Vilar, J.M.G., Kueh, H.Y., Barkai, N., Leibler, S. (2002) Mechanisms of noise-resistance in genetic oscillators, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5988–5992.

Volz, D., Eigel, M., Athale, C., Bastian, P., Hermann, H., Kappel, C., Eils, R. (2004) Spatial modeling and simulation of diffusion in nuclei of living cells. *CMSB04, Lecture Notes in Computer Science*, Springer, Heidelberg.

von Dassow, G., Meir, E., Munro, E.M., Odell, G. (2000) The segment polarity network is a robust developmental module, *Nature* **406**, 188–192.

Wolkenhauer, O. (2001) Systems biology: the reincarnation of systems theory applied in biology, *Brief. Bioinf.* **2**(3), 258–270.

Yi, T.-M., Huang, Y., Simon, M.I., Doyle, J.C. (2000) Robust perfect adaptation in bacterial chemotaxis through integral feedback control, *Proc. Natl. Acad. Sci. U.S.A.* **97**(9), 4649–4653.

Takifugu and Tetradon Genomes: *see* Pufferfish Genomes: *Takifugu* and *Tetraodon*

Targeted Therapy: Genomic Approaches

Anna C. Maroney and Sergey E. Ilyin
Johnson & Johnson Pharmaceutical Research & Development, Spring House, PA

1 Principals of Targeted Therapy 152
1.1 Molecular Basis 152
1.2 Chromosomal Translocation 152
1.3 Gene Amplification 153
1.4 Tumor-specific Receptor Expression 153
1.5 Targeting Enzymatic Cascades 154

2 Therapeutic Agents Used in Targeted Approaches 154
2.1 Small Molecules 154
2.2 Antibody-based Therapeutics 155
2.3 Emerging Therapeutic Agents 156

3 Technological Approaches 156
3.1 Integrated Clinical Strategies for Diagnosis, Treatment, and Prognosis 156
3.2 High-throughput Genomics Approaches 156
3.3 Integrative Approaches 157

4 Concluding Remarks 158

Bibliography 158
Books and Reviews 158
Primary Literature 159

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

Keywords

Biomarker
A characteristic that is objectively measured and evaluated as an indicator or prognostic of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

Functional Informatics
Experimental paradigm based on the convergence and integration of bioinformatics and broad-based automation.

Targeted Therapy
Treatments directed toward inhibition or modulation of specific signaling pathways, receptors, or enzymatic cascades.

The concept of Targeted Therapy was developed over the past 10 years and it relies on inhibition of specific signaling pathways, as opposed to nonspecific chemotherapy. Consequently, if targeted pathways are differentially activated between normal and malignant cells, significantly fewer side effects would be expected in clinical settings. Translocations and other genomic changes that occur in malignancies lead to activation of specific signaling pathways and/or enzymatic cascades, which can be specifically and selectively addressed with therapeutic agents.

1 Principals of Targeted Therapy

1.1 Molecular Basis

Genetic instability is the hallmark of cancer as a disease. Gene alterations range from amplification, insertions, deletions, chromosomal translocations, or point mutations in the tumor cell's DNA. These genetic modifications lead to changes in tumor cell physiology, which can be specifically and selectively targeted by therapeutic agents. Inhibitors of cell cycle, growth factor signaling cascades, protein folding, angiogenesis, and apoptosis are among the main areas of focus for targeted therapy in oncology. The Cancer Genome Anatomy Project (CGAP) was created to provide a database of these genetic alterations in cancer genomes and is currently the most comprehensive freely accessible database to date (http://cgap.nci.nih.gov). Further, analysis that link these genetic changes with poorer prognosis afford the opportunity to create molecular road maps that may lead to a greater understanding of the processes that lead to uncontrolled cell growth.

1.2 Chromosomal Translocation

Oncogenes (genes that cause cellular transformation) can be activated as a result

of chromosomal translocation. The fusion protein, Bcr-Abl, occurs as a result of a chromosomal translocation between chromosome 9 and 22 (referred to as the *Philadelphia chromosome*) in white blood cells of about 95% of patients with chronic myelogenous leukemia (CML). The FDA approved Gleevec/Glivec (imatinib mesylate), an inhibitor of the Bcr-Abl tyrosine protein kinase, in 2001, as first-line treatment for newly diagnosed CML patients. A side-by-side comparison of Gleevec with interferon-α plus cytarabine revealed remarkable complete cytogenetic response rates of 76% in the group treated with Gleevec as compared to 15% in the group treated with interferon-α plus cytarabine. Whereas resistance to Gleevec therapy is rare (5–10%) in newly diagnosed patients in the early chronic phase of CML, it can occur in up to 60 to 70% of patients in the late blast phase, a stage at which the number of immature abnormal white blood cells in the marrow and blood are extremely high. This resistance, in part, occurs as a result of acquired mutations in Bcr-Abl that make inhibitor binding less effective. Such mutations that can occur over treatment time and the acquisition of other genetic changes support the continual development of therapeutics for this disease including less mutation-sensitive inhibitors of Bcr-Abl. Gleevec also inhibits other related kinases, such as the platelet derived growth factor receptor (PDGFR) and c-kit. As a result of positive clinical outcome in patients with c-kit positive gastrointestinal stromal tumors, it was also approved for this indication in 2002. Gleevec is currently being tested in over 40 clinical trials in combination with a wide variety of agents including other targeted therapies as well as standard chemotherapy.

1.3 Gene Amplification

Gene amplification resulting in the overexpression of proteins that are critical in growth factor stimulated signaling pathways are also targets of potential therapy. Examples of this are inhibitors that block the epidermal growth factor receptor (EGFR). EGFR is overexpressed and associated with poor prognosis in a wide variety of tumors such as lung, gastrointestinal, pancreatic, and colorectal. There are currently three approved therapies that directly inhibit EGFR: Iressa (Gefitnib) for the treatment of non-small cell lung carcinoma (NSCLC), Erbitux (Cetuximab) for treating colorectal cancer and Tarceva (Irlotinib) for NSCLC. Clinical trials with Iressa revealed a relatively low 10% response rate and the response did not appear to correlate with EGFR overexpression. Retrospective analysis of responders revealed that the response was selective for a patient population with activating mutations in the receptor. These results underscore the need to understand how the targets are driving tumor growth and highlight the need for prescreening patients to identify those patient populations who are most likely to respond to the therapy. Another member of the tyrosine kinase EGFR family, HER2 or EGFR2, is overexpressed in breast tumors and is the target of an antibody therapeutic called *Herceptin* (Trastuzumab) that was approved in 1998 for the treatment of metastatic breast cancer.

1.4 Tumor-specific Receptor Expression

Arguably, the first targeted therapy approved by the FDA occurred over 20 years

ago and acts as a weakly binding estrogen mimetic that competes with endogenous estrogen and effectively blocks the activity of the estrogen receptor found on breast tumors. Nolvadex (Tamoxifen) was first approved for use in the treatment of estrogen receptor positive breast cancer. In 1998, Nolvadex gained a new indication when it was demonstrated in a National Cancer Institute funded study, that it reduced the incidence of developing breast cancer by 44% in women considered at high risk for developing the disease. However, a small risk factor of endometrial cancer has been associated with long-term use of Nolvadex in postmenopausal women. This risk factor potentially may be circumvented by a novel antiestrogen, Faslodex (Fulvestrant), that is a more potent estrogen receptor antagonist than Nolvadex and has been shown to be devoid of agonist effects on the endometrium in cell culture studies. Faslodex has been approved for second-line treatment in breast cancer patients and is currently in clinical trials for first-line therapy. Postmenopausal breast cancer patients who have completed antiestrogen therapy but have a recurrence now have another option. Aromatase is an enzyme involved in the production of estrogen. Recently, Femara (Letrozole), an aromatase inhibitor, which was previously approved for treating breast metastasis patients, has won FDA approval for treating postmenopausal women who have taken Nolvadex for five years, after which Nolvadex's efficacy declines. In a clinical study of over 5100 patients randomly assigned to take Femara or a placebo after Nolvadex therapy, it was found that Femara cut the risk of death by 39% for certain breast cancer patients.

1.5 Targeting Enzymatic Cascades

Beyond targeting specific proteins in growth factor signaling cascades, another targeted approach relies on inhibition of enzymatic cascades involved in protein degradation. Velcade (Bortezomib) blocks a key process that regulates protein turnover by inhibiting the activity of the 26S proteasome complex. The proteasome complex rids the cell of short-lived and regulatory proteins important in basic cellular processes such as regulation of cell cycle and modulation of growth factor receptors. Proteins are identified for destruction by an enzymatic cascade that tags the protein with multiple small molecules of ubiquitin. The polyubiquitinated proteins are directed to the 26S proteasome complex where they are shredded by protease activity. Velcade has been approved for the treatment of multiple myeloma and is currently being tested in multiple Phase II clinical trials including NSCLC, colorectal, and non-Hodgkin lymphoma.

2 Therapeutic Agents Used in Targeted Approaches

2.1 Small Molecules

Small molecules (usually possessing molecular weights of less then 500 Da) are attractive therapeutics for a number of reasons: if cell permeability is not limiting, then these molecules have the greatest potential for reaching an intracellular target; greater choices in delivery methods including the ease of oral administration; and offer more options for chemical modification than other approaches like antibody

or gene therapies. The major targets of small molecule inhibitors in current clinical oncology trials affect growth factor receptor signaling pathways. Several examples of small molecule inhibitors that are showing early promise in current clinical trials include: Zarnestra (tipifarnib), a protein farnesylation inhibitor in Phase II for acute myelogenous leukemia; Midostaurin (PKC-412) and MLN-518, Flt-3 inhibitors in Phase II for acute myelogenous leukemia; ZD6474, an inhibitor of the (vascular endothelial growth factor receptor) VEGFR/EGFR family in Phase II for NSCLC and multiple myeloma; BAY43-9006, a raf/PDGFR/VEGFR inhibitor in Phase III for renal cell carcinoma; and CCI-779, an analog of rapamycin and an inhibitor of mTor in Phase III for renal cell carcinoma.

2.2 Antibody-based Therapeutics

Therapeutic antibodies have proven an effective strategy for oncology. Early attempts at antibody therapy were met with limited success. Naked rodent antibodies lacked overall potency and induced a human anti-rodent antibody response preventing the use of repetitive dosing. Further, poor tumor cell penetrance and poor selection of early targets added to lack of efficacious antibody therapeutics. In the early 1980s, human/mouse chimeric antibodies were engineered through recombinant technology and reduced the generation of human anti-therapeutic antibody responses, but did not eliminate them. Advances in molecular immunology have resulted in fully human antibodies as therapeutics that may further reduce antigenicity. Toxicity with this approach is usually attributed to cross reactivity with normal tissue, for example, heart muscle damage associated with the use of Herceptin. Early successes with antibody therapeutics were directed to cell surface antigens in the myeloid lineage that when bound mediated cytolysis by activating complement and antibody-dependent cellular toxicity. Another application of these high specificity and high affinity binding antibodies was to conjugate them to radioisotopes or cytotoxins. Moreover, with the advent of better target selection, antibodies, such as Avastin (bevacizumab) and Erbitux (cetuximab), are having noteworthy effects in solid tumors. There are eight antibody therapeutics that are currently approved by the FDA for the treatment of cancer (Table 1).

Tab. 1 FDA approved antibody-based therapeutics.

Antibody	*Target*	*FDA approval for use [year]*
Rituxan	CD20	Non-Hodgkin's lymphoma (1997)
Herceptin	HER2/EGFR2	Breast cancer (1998)
Mylotarg	CD33/cali-cheamicin	Acute myeloid leukemia (2000)
Campath	CD52	Chronic lymphocytic leukemia (2001)
Zevalin	CD20/90Y	Non-Hodgkin's lymphoma (2002)
Bexxar	CD20	Non-Hodgkin's lymphoma (2003)
Erbitux	EGFR	Colorectal cancer (2004)
Avastin	VEGFR	Colorectal cancer (2004)

2.3
Emerging Therapeutic Agents

Advances in molecular technologies have opened up new areas for unique therapeutic approaches based on functional genomics tools. Therapeutic approaches that are in development stages at various academic and commercial institutions include vaccines, antisense oligonucleotides, and recombinant vaccines. For example, G3139, antisense oligonucleotide against Bcl2, a pro-survival mitochondrial protein, is in preregistration for melanoma. Agents based on functional genomics tools, such as antisense, ribozymes, and RNAi may potentially become stand-alone therapies. They also play an additional role by enabling functional profiling to identify targets and biomarkers. For example, RNAi technology enables any protein target to be inhibited by sequence-specific, double-stranded RNA molecules. siRNA-based approaches are effective and allow for functional characterization of new potential targets in preclinical studies. Other functional genomics tools, such as antisense ribozymes may also prove useful in the clinic as they already have in the research environment.

3
Technological Approaches

3.1
Integrated Clinical Strategies for Diagnosis, Treatment, and Prognosis

Investigation of tumor-specific changes that differentiate cancer cells from other cells in the body is important for identification of novel receptors and pathways, which may be targeted. Greater understanding of cancer biology is also important for discovery and validation of biomarkers. Targeted therapies and predictive biomarkers can potentially guide physicians to eventually treat each patient individually with medication that is personalized and potentially more effective. With the advent of the sequencing of the human genome and high-throughput functional genomics, the identification of possible targets has increased exponentially. Of the 30 000 to 40 000 human genes that have been identified, preliminary analysis indicates that about 60% of these are of known function: 21% are involved in signal transduction, 10% are transcription factors, 5% are protooncogenes and 3% are cell adhesion molecules. This information along with functional analysis of the remaining genes should lead to the identification of new targets in cancer therapeutics.

3.2
High-throughput Genomics Approaches

Expressional analysis is greatly enhanced by the recent advances in high-throughput technologies and in related bioinformatics tools to analyze, interpret, and manage increasing amount of genomic data. Microarray analysis significantly expanded the throughput of genomic studies. For example, hierarchical clustering of expression data revealed distinct groups of epithelial ovarian malignancies. On the basis of microarray data, tumors could be further subdivided into major groupings that correlated both to histological and clinical observations. Microarray-based comparison of gene expression profiles of paired diagnosis-relapse samples from patients with precursor-B acute lymphoblastic leukemia showed that only six genes differed significantly in expression. Microarray technology enables high-throughput

testing of expression and SNP genotyping. Microarrays could be manufactured by direct synthesis of DNA on the chip surface (for example, Affymetrix chips (http://www.affymetrix.com, and NimbleGen (http://www.nimblegen.com/). Microarray could also be produced by spotting materials on the chip surface.

Spotted microarrays may incorporate cDNA, oligonucleotides, and also proteins and functional genomics tools, such as siRNA. Minimum Information about Microarray Experiment (MIAME) is a set of rules aimed at making the microarray results reproducible (http://www.mged.org/Workgroups/MIAME/miame.html.) Microarray data undergoes statistical analysis to identify genes of interest. Statistical microarray analysis (SMA) package is available at http://stat-www.berkeley.edu/users/terry/zarray/Software/smacode.html. SMA is running in an R statistical environment; R (www.r-project.org) is a widely used open-source object-oriented statistical software package. Fixed content arrays (which are mass-produced) can provide many data points; however, sequences of interest may be missing. Microarray experiments could also miss changes in the expression of low-level abundance genes. A significant opportunity exists when combining the throughput of a microarray with the sensitivity of PCR. For example, OpenArray™ Transcript Analysis (BioTrove) uses parallel quantitative PCR to determine transcript copy numbers in samples of interest. OpenArray™ Transcript Analysis platform consumes nanoliter-scale quantities of reagents and samples. Microscope slide-sized OpenArray™ plates contain over 3000 through-holes, which accept 33-nL reactions. By harnessing surface energy, Thru-Hole™ technology creates a precise dispensing action that does not require complex equipment.

3.3 Integrative Approaches

Microarray analyses could be used to identify genes that are induced or suppressed in certain types of tumors and these can be studied as potential targets and biomarkers. Functional characterization of these novel factors is a rate-limiting step. Functional genomics approaches coupled with bioinformatics have the potential to accelerate target and biomarker identification and validation and subsequent development of novel therapeutic approaches and strategies. siRNA technology could be used to selectively probe involvement of genes in the control of proliferation and other cellular processes. Additional studies have demonstrated that siRNA transfection and subsequent processing and analysis could be automated. Automated application of gene specific siRNA produces modified cell populations, which can be examined via batteries of functional tests. Genes with impact on biological processes of interest are selected and investigated as potential targets and biomarkers. Bio-pharmaceutical discovery and development create a demand for conducting a significant number of assays for target validation, lead discovery, and optimization, biomarker support, and other relevant assessments. For example, individual patient's tumor cells may be tested against a panel of siRNA reagents to identify pathways driving tumor growth and progression. This information may lead to formulation of personalized strategies of treatment. To minimize cost, automation becomes very important. Robotics may be used at different stages of the discovery and

development process. A fairly complex experiment could be split into a series of relatively simple steps, similar to Henry Ford's approach to used automobile production. Robots may be positioned to accommodate a very complex experimental workflow, creating an assembly line approach to molecular biology. This experimental approach is termed *Functional Informatics*. (Healthtech Institute Bioinformatics Glossary: http://www.genomicglossaries.com/content/Bioinformatics_gloss.asp).

To make this transformational change happen, several elements are of critical importance: (1) assays have to be reformulated to automation-friendly formats, (2) data acquisition and data analysis have to be automated, and (3) each test system has to be independently validated. Automation drives increases in data volumes, and this in turn necessitates automated data processing and reporting. As data are generated on the same platform, significant efficiency is achieved via consistent data processing, data integration, data storage, and data reporting. As data generation and analysis become integrated on the same platform, an automated analysis via artificial intelligence allows the automated hypothesis generation and testing.

4 Concluding Remarks

Targeted Therapy relies on inhibition of specific signaling pathways and results in significantly fewer adverse effects in clinical settings. Genomic technologies and integrative paradigms such as Functional Informatics have significant potential to accelerate drug discovery and optimize the development of novel targeted therapies.

See also Antitumor Agents: Taxol and Taxanes – Production by Yew Cell Culture; Antitumor Steroids; Epigenetic Mechanisms in Tumorigenesis; Genetics, Molecular Basis of; Growth Factors and Oncogenes in Gastrointestinal Cancers; Informatics (Computational Biology); Oligonucleotides; Oncogenes; Oncology, Molecular; Targeting and Intracellular Delivery of Drugs.

Bibliography

Books and Reviews

Amador, M.L., Hidalgo, M. (2004) Epidermal growth factor receptor as a therapeutic target for the treatment of colorectal cancer, *Clin. Colorectal Cancer* **4**(1), 51–62.

Balmain, A., Gray, J., Ponder, B. (2003) The genetics and genomics of cancer, *Nat. Genet.* **33**, 238–244.

Boon, K., Osorio, E.C., Greenhut, S.F., Schaefer, C.F., Shoemaker, J., Polyak, K., Morin, P.J., Buetow, K.H., Strausberg, R.L., De Souza, S.J., Riggins, G.J. (2002) An anatomy of normal and malignant gene expression, *Proc. Natl. Acad. Sci. U.S.A.* **99**(17), 11287–11292.

Botstein, D. (2003) Genomic perspective and cancer, *Cold Spring Harbor Symp. Quant. Biol.* **68**, 417–424.

Campbell, S.L., Khosravi-Far, R., Rossman, K.L., Clark, G.J., Der, C.J. (1998) Increasing complexity of Ras signaling, *Oncogene* **17**(11), 1395–1413.

Choueiri, T.K., Alemany, C.A., Abou-Jawde, R.M., Budd, G.T. (2004) Role of aromatase inhibitors in the treatment of breast cancer, *Clin. Ther.* **26**(8), 1199–1214.

Cohen, I. (2004) Endometrial pathologies associated with postmenopausal tamoxifen treatment, *Gynecol. Oncol.* **94**(2), 256–266.

Demetri, G.D. (2002) Targeting the molecular pathophysiology of gastrointestinal stromal tumors with imatinib. Mechanisms, successes, and challenges to rational drug development, *Hematol. Oncol. Clin. North Am.* **16**(5), 1115–1124.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., Stratton, M.R. (2004) A census of human cancer genes, *Nat. Rev. Cancer* **4**(3), 177–183.

Gatto, B. (2004) Monoclonal antibodies in cancer therapy, *Curr. Med. Chem. Anti-Canc Agents* **4**(5), 411–414.

Heel, R.C., Brogden, R.N., Speight, T.M., Avery, G.S. (1978) Tamoxifen: a review of its pharmacological properties and therapeutic use in the treatment of breast cancer, *Drugs* **16**(1), 1–24.

Ilyin, S.E., Belkowski, S.M., Plata-Salaman, C.R. (2004) Biomarker discovery and validation: technologies and integrative approaches, *Trends Biotechnol.* **22**(8), 411–416.

Ilyin, S.E., Bernal, A., Horowitz, D., Derian, C.K., Xin, H. (2004) Functional informatics: convergence and integration of automation and bioinformatics, *Pharmacogenomics* **5**(6), 721–730.

Jain, R.K. (2002) Tumor angiogenesis and accessibility: role of vascular endothelial growth factor, *Semin. Oncol.* **29**(6 Suppl. 16), 3–9.

Marshall, M. (1995) Interactions between Ras and Raf: key regulatory proteins in cellular transformation, *Mol. Reprod. Dev.* **42**(4), 493–499.

Miles, D.W. (2001) Update on HER-2 as a target for cancer therapy: herceptin in the clinical setting, *Breast Cancer Res.* **3**(6), 380–384.

O'Reilly, M.S., Boehm, T., Shing, Y., Fukai, N., Vasios, G., Lane, W.S., Flynn, E., Birkhead, J.R., Olsen, B.R., Folkman, J. (1997) Endostatin: an endogenous inhibitor of angiogenesis and tumor growth, *Cell* **88**(2), 277–285.

Paez, J.G., Janne, P.A., Lee, J.C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F.J., Lindeman, N., Boggon, T.J., Naoki, K., Sasaki, H., Fujii, Y., Eck, M.J., Sellers, W.R., Johnson, B.E., Meyerson, M. (2004) EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy, *Science* **304**(5676), 1497–1500.

Sommer, T., Jarosch, E., Lenk, U. (2001) Compartment-specific functions of the ubiquitin-proteasome pathway, *Rev. Physiol. Biochem. Pharmacol.* **142**, 97–160.

Strausberg, R.L. (2001) The cancer genome anatomy project: new resources for reading the molecular signatures of cancer, *J. Pathol.* **195**(1), 31–40.

Strausberg, R.L., Buetow, K.H., Emmert-Buck, M.R., Klausner, R.D. (2000) The cancer genome anatomy project: building an annotated gene index, *Trends Genet.* **16**(3), 103–106.

Suter, T.M., Cook-Bruns, N., Barton, C. (2004) Cardiotoxicity associated with trastuzumab (Herceptin) therapy in the treatment of metastatic breast cancer, *Breast* **13**(3), 173–183.

Tauchi, T., Ohyashiki, K. (2004) Molecular mechanisms of resistance of leukemia to imatinib mesylate, *Leuk. Res.* **28**(Suppl. 1), S39–S45.

Tipping, A.J., Melo, J.V. (2003) Imatinib mesylate in combination with other chemotherapeutic drugs: in vitro studies, *Semin. Hematol.* **40**(2 Suppl. 2), 83–91.

Vogelstein, B., Kinzler, K.W. (2002) *The Genetic Basis of Human Cancer*, McGraw-Hill, New York.

Wannesson, L., Ghielmini, M. (2003) Overview of antibody therapy in B-cell non-Hodgkin's lymphoma, *Clin. Lymphoma* **4**(Suppl. 1), S5–12.

Wong, S. Witte, O.N. (2004) The BCR-ABL story: bench to bedside and back, *Annu. Rev. Immunol.* **22**, 247–306.

Primary Literature

Adams, J., Palombella, V.J. Sausville, E.A. Johnson, J. Destree, A. Lazarus, D.D. Maas, J. Pien, C.S. Prakash, S. Elliott, P.J. (1999) Proteasome inhibitors: a novel class of potent and effective antitumor agents, *Cancer Res.* **59**(11), 2615–2622.

Aiello, L.P., Pierce, E.A., Foley, E.D., Takagi, H., Chen, H., Riddle, L., Ferrara, N., King, G.L., Smith, L.E. (1995) Suppression of retinal neovascularization in vivo by inhibition of vascular endothelial growth factor (VEGF) using soluble VEGF-receptor chimeric proteins, *Proc. Natl. Acad. Sci. U.S.A.* **92**(23), 10457–10461.

Brooks, P., Fuertes, G., Murray, R.Z., Bose, S., Knecht, E., Rechsteiner, M.C., Hendil, K.B., Tanaka, K., Dyson, J., Rivett, J. (2000)

Subcellular localization of proteasomes and their regulatory complexes in mammalian cells, *Biochem. J.* **346**(Pt 1), 155–161.

Chan, T.L., Zhao, W., Leung, S.Y., Yuen, S.T. (2003) BRAF and KRAS mutations in colorectal hyperplastic polyps and serrated adenomas, *Cancer Res.* **63**(16), 4878–4881.

Ciardiello, F., Caputo, R., Damiano, V., Troiani, T., Vitagliano, D., Carlomagno, F., Veneziani, B.M., Fontanini, G., Bianco, A.R., Tortora, G. (2003) Antitumor effects of ZD6474, a small molecule vascular endothelial growth factor receptor tyrosine kinase inhibitor, with additional activity against epidermal growth factor receptor tyrosine kinase, *Clin. Cancer Res.* **9**(4), 1546–1556.

Cohen, M.H., Williams, G.A., Sridhara, R., Chen, G., Pazdur, R. (2003) FDA drug approval summary: gefitinib (ZD1839) (Iressa) tablets, *Oncologist* **8**(4), 303–306.

Fisher, B., Costantino, J.P., Wickerham, D.L., Redmond, C.K., Kavanah, M., Cronin, W.M., Vogel, V., Robidoux, A., Dimitrov, N., Atkins, J., Daly, M., Wieand, S., Tan-Chiu, E., Ford, L., Wolmark, N. (1998) Tamoxifen for prevention of breast cancer: report of the national surgical adjuvant breast and bowel project P-1 study, *J. Natl. Cancer Inst.* **90**(18), 1371–1388.

Goss, P.E., Ingle, J.N., Martino, S., Robert, N.J., Muss, H.B., Piccart, M.J., Castiglione, M., Tu, D., Shepherd, L.E., Pritchard, K.I., Livingston, R.B., Davidson, N.E., Norton, L., Perez, E.A., Abrams, J.S., et al. (2003) A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer, *N. Engl. J. Med.* **349**(19), 1793–1802.

Griswold, I.J., Shen, L.J., La Rosee, P., Demehri, S., Heinrich, M.C., Braziel, R.M., McGreevey, L., Haley, A.D., Giese, N., Druker, B.J., Deininger, M.W. (2004) Effects of MLN518, a dual FLT3 and KIT inhibitor, on normal and malignant hematopoiesis, *Blood* **104**(9), 2912–2918.

Groffen, J., Stephenson, J.R., Heisterkamp, N., de Klein, A., Bartram, C.R., Grosveld, G. (1984) Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22, *Cell* **36**(1), 93–99.

Houben, R., Becker, J.C., Kappel, A., Terheyden, P., Brocker, E.B., Goetz, R., Rapp, U.R. (2004) Constitutive activation of the Ras-Raf signaling pathway in metastatic melanoma is associated with poor prognosis, *J. Carcinog.* **3**(1), 6.

Howell, A., Osborne, C.K., Morris, C., Wakeling, A.E. (2000) ICI 182,780 (Faslodex): development of a novel, "pure" antiestrogen, *Cancer* **89**(4), 817–825.

Hurwitz, H., Fehrenbacher, L., Novotny, W., Cartwright, T., Hainsworth, J., Heim, W., Berlin, J., Baron, A., Griffing, S., Holmgren, E., Ferrara, N., Fyfe, G., Rogers, B., Ross, R., Kabbinavar, F. (2004) Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer, *N. Engl. J. Med.* **350**(23), 2335–2342.

Kimura, E.T., Nikiforova, M.N., Zhu, Z., Knauf, J.A., Nikiforov, Y.E., Fagin, J.A. (2003) High prevalence of BRAF mutations in thyroid cancer: genetic evidence for constitutive activation of the RET/PTC-RAS-BRAF signaling pathway in papillary thyroid carcinoma, *Cancer Res.* **63**(7), 1454–1457.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**(6822), 860–921.

Lemoine, N.R., Pignatelli, M. (1995) Type-1 growth factor receptors in pancreatic and gastrointestinal neoplasia, *Eur. J. Gastroenterol. Hepatol.* **7**(10), 939–944.

Lynch, T.J., Bell, D.W., Sordella, R., Gurubhagavatula, S., Okimoto, R.A., Brannigan, B.W., Harris, P.L., Haserlat, S.M., Supko, J.G., Haluska, F.G., Louis, D.N., Christiani, D.C., Settleman, J., Haber, D.A. (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib, *N. Engl. J. Med.* **350**(21), 2129–2139.

Margolin, K., Gordon, M.S., Holmgren, E., Gaudreault, J., Novotny, W., Fyfe, G., Adelman, D., Stalter, S., Breed, J. (2001) Phase Ib trial of intravenous recombinant humanized monoclonal antibody to vascular endothelial growth factor in combination with chemotherapy in patients with advanced cancer: pharmacologic and long-term safety data, *J. Clin. Oncol.* **19**(3), 851–856.

Mills, G.B., Lu, Y., Kohn, E.C. (2001) Linking molecular therapeutics to molecular diagnostics: inhibition of the FRAP/RAFT/TOR

component of the PI3K pathway preferentially blocks PTEN mutant cells in vitro and in vivo, *Proc. Natl. Acad. Sci. U.S.A.* **98**(18), 10031–10033.

O'Brien, S.G., Guilhot, F., Larson, R.A., Gathmann, I., Baccarani, M., Cervantes, F., Cornelissen, J.J., Fischer, T., Hochhaus, A., Hughes, T., Lechner, K., Nielsen, J.L., Rousselot, P., Reiffers, J., Saglio, G., et al. (2003) Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia, *N. Engl. J. Med.* **348**(11), 994–1004.

O'Hare, T., Pollock, R., Stoffregen, E.P., Keats, J.A., Abdullah, O.M., Moseson, E.M., Rivera, V.M., Tang, H., Metcalf, C.A. 3rd, Bohacek, R.S., Wang, Y., Sundaramoorthi, R., Shakespeare, W.C., Dalgarno, D., Clackson, T., et al. (2004) Inhibition of wild-type and mutant Bcr-Abl by AP23464, a potent ATP-based oncogenic protein kinase inhibitor: implications for CML, *Blood* **104**(8), 2532–2539.

O'Reilly, M.S., Holmgren, L., Chen, C., Folkman, J. (1996) Angiostatin induces and sustains dormancy of human primary tumors in mice, *Nat. Med.* **2**(6), 689–692.

Pandey, A., Volkots, D.L., Seroogy, J.M., Rose, J.W., Yu, J.C., Lambing, J.L., Hutchaleelaha, A., Hollenbach, S.J., Abe, K., Giese, N.A., Scarborough, N.A. (2002) Identification of orally active, potent, and selective 4-piperazinylquinazolines as antagonists of the platelet-derived growth factor receptor tyrosine kinase family, *J. Med. Chem.* **45**(17), 3772–3793.

Perez-Soler, R. (2004) The role of erlotinib (Tarceva, OSI 774) in the treatment of non-small cell lung cancer, *Clin. Cancer Res.* **10**(12 Pt 2), 4238s–4240s.

Richardson, P.G., Barlogie, B., Berenson, J., Singhal, S., Jagannath, S., Irwin, D., Rajkumar, S.V., Srkalovic, G., Alsina, M., Alexanian, R., Siegel, D., Orlowski, R.Z., Kuter, D., Limentani, S.A., Lee, S., et al. (2003) A phase 2 study of bortezomib in relapsed, refractory myeloma, *N. Engl. J. Med.* **348**(26), 2609–2617.

Shah, N.P., Tran, C., Lee, F.Y., Chen, P., Norris, D., Sawyers, C.L. (2004) Overriding imatinib resistance with a novel ABL kinase inhibitor, *Science* **305**(5682), 399–401.

Staal, F.J., van der Burg, M., Wessels, L.F., Barendregt, B.H., Baert, M.R., van den Burg, C.M., van Huffel, C., Langerak, A.W., van der Velden, V.H., Reinders, M.J., van Dongen, J.J. (2003) DNA microarrays for comparison of gene expression profiles between diagnosis and relapse in precursor-B acute lymphoblastic leukemia: choice of technique and purification influence the identification of potential diagnostic markers, *Leukemia* **17**(7), 1324–1332.

Udvardy, A. (1993) Purification and characterization of a multiprotein component of the Drosophila 26S (1500 kDa) proteolytic complex, *J. Biol. Chem.* **268**(12), 9055–9062.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., et al. (2001) The sequence of the human genome, *Science* **291**(5507), 1304–1351.

Weisberg, E., Boulton, C., Kelly, L.M., Manley, P., Fabbro, D., Meyer, T., Gilliland, D.G., Griffin, J.D. (2002) Inhibition of mutant FLT3 receptors in leukemia cells by the small molecule tyrosine kinase inhibitor PKC412, *Cancer Cell.* **1**(5), 433–443.

Welsh, J.B., Zarrinkar, P.P., Sapinoso, L.M., Kern, S.G., Behling, C.A., Monk, B.J., Lockhart, D.J., Burger, R.A., Hampton, G.M. (2001) Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer, *Proc. Natl. Acad. Sci. U.S.A.* **98**(3), 1176–1181.

Wilhelm, S., Chien, D.S. (2002) BAY 43–9006: preclinical data, *Curr. Pharm. Des.* **8**(25), 2255–2257.

Xin, H., Bernal, A., Amato, F.A., Pinhasov, A., Kauffman, J., Brenneman, D.E., Derian, C.K., Andrade-Gordon, P., Plata-Salaman, C.R., Ilyin, S.E. (2004) High-throughput siRNA-based functional target validation, *J. Biomol. Screen.* **9**(4), 286–293.

Zujewski, J., Horak, I.D., Bol, C.J., Woestenborghs, R., Bowden, C., End, D.W., Piotrovsky, V.K., Chiao, J., Belly, R.T., Todd, A., Kopp, W.C., Kohler, D.R., Chow, C., Noone, M., Hakim, F.T., et al. (2000) Phase I and pharmacokinetic study of farnesyl protein transferase inhibitor R115777 in advanced cancer, *J. Clin. Oncol.* **18**(4), 927–941.

Targeting and Intracellular Delivery of Drugs

Ruth Duncan
Welsh School of Pharmacy, Cardiff University, Cardiff, UK

1 Introduction 166
1.1 Challenges for Targeting and Intracellular Delivery 167
1.2 Principles of Drug Targeting and Controlled Release 167
1.3 Biological Barriers 170
1.3.1 Whole Organism Level 170
1.3.2 In the Circulation 170
1.3.3 Cellular Level 175
1.4 Opportunities for Targeting 177
1.4.1 Passive Targeting to Diseased Tissue 177
1.4.2 Active Targeting to Diseased Cells 178
1.4.3 How to Realize Intracellular Targeting? 181
1.5 Devices, Technologies, and Nanomedicines 185

2 Vectors: Design, Applications, and Clinical Status 186
2.1 Liposomes 187
2.2 Antibodies, Proteins, and Fusion Proteins 190
2.3 Nanoparticles 192
2.4 Polymer conjugates 193
2.4.1 Polymer–protein Conjugates 195
2.4.2 Polymer–drug Conjugates 196
2.5 Polymeric Micelles 197
2.6 Viral and Nonviral Vectors 198

3 To the Future – The Importance of "Nanomedicines"? 198

Bibliography 200
Books and Reviews 200
Primary Literature 200

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

Keywords

Angiogenesis
The process by which small new blood vessels are formed by budding from existing vessels in both normal and diseased (e.g. tumor) tissue.

Biomimetic Polymer
A polymer whose structure is designed to mimic the properties of a natural macromolecule, for example, a synthetic multivalent ligand designed for receptor interaction.

Controlled Release
Controlled release dosage forms maintain a constant plasma concentration of drug over a prolonged period of time.

Drug Targeting
Guidance of a drug to a diseased cell, or an intracellular compartment within that cell, where it can exert its therapeutic effect. Guidance of drug away from sites of potential toxicity is also a benefit.

Dendrimer
A macromolecule containing symmetrically arranged branches arising from a multifunctional core. Repeated reaction sequences add a precise number of terminal groups at each step or generation.

Endocytosis
Internalization of the cell's plasma membrane to form vesicles that capture macromolecules and particles present in the extracellular fluid and/or bound to membrane-associated receptors. These vesicles then undergo a complex series of fusion events directing the internalized material to an appropriate intracellular compartment.

Endosomotropic
Delivery of molecules to the cell via the endosomal compartment of the cell. Term applied to polymer constructs designed as viral mimetics that can permeabilize the endosomal membrane and thus deliver proteins and oligonucleotides into the cytoplasm of the cell.

Immunoconjugate
An antibody-based conjugate usually containing a covalently bound drug.

Liposome
Lipid-based vesicle used to entrap a drug payload and promote disease-specific targeting.

Lysosomotropic
Delivery of molecules to the cell via the lysosomal compartment of the cell. A term applied to polymer-drug constructs taken into the cell by endocytosis, which contains a linker that is cleaved by lysosomal hydrolases to liberate the drug molecule that diffuses across the lysosomal membrane into the cell.

Micelle
A colloidal aggregate (usually of spherical dimensions) formed in aqueous solution forms molecules that have both hydrophilic (water-loving) and hydrophobic domains (water-hating). They form when the concentration reaches the critical micelle concentration. Both surfactant and block copolymer micelles are being used as drug delivery systems.

Nanomedicines
The umbrella term used to describe nanosized (1–1000 nm) drugs and drug delivery systems used as medicines. This can include molecular medicines (e.g. proteins, antibodies, recombinant fusion proteins and hybrid polymer-protein constructs), but is more generally used to describe multicomponent drug delivery systems such as liposomes (which may also be polymer coated and contain antibody targeting elements), nanoparticles, and polymer therapeutics including polyplexes.

Nanoparticle
Tiny particles, usually of dimension 20 to 500 nm and formed from natural or synthetic polymers that are used to entrap drug for improved drug targeting and controlled release.

Nanomedicine, the Discipline
The newly emerging discipline called *nanomedicine* describes the application of nanotechnology (usually viewed as 1–1000 nm) to the design of systems and devices that can be (1) used to facilitate better understanding of disease pathophysiology and thus enables new target identification for therapeutic intervention; (2) nanoimaging at the cellular and patient level; and (3) design of nanomedicines and nanodiagnostics. Underpinning fields are nanorelated materials, nanorelated engineering and nanorelated toxicology.

Polymer Therapeutics
This is an umbrella term used to describe polymeric drugs, polymer–drug conjugates, polymer–protein conjugates, polymeric micelles to which drug is covalently bound, and multicomponent polyplexes being developed as nonviral vectors. From the industrial standpoint, these nanosized medicines are more like new chemical entities than conventional ''drug delivery systems or formulations'' which simply entrap, solubilize, or control drug release without resorting to chemical conjugation.

Polyplex
The term is usually used to describe the complex formed by a polycation and an anionic oligonucleotide or plasmid. The term interpolyelectrolyte complex (IPEC) is also used.

Effective targeting of drugs (particularly macromolecular therapeutics such as proteins, peptides, and oligonucleotides) to diseased cells, and moreover to the precise intracellular compartment where they are required has proved difficult to achieve. While the last decade has seen a number of drug delivery systems incorporating low molecular weight drugs and proteins enter routine clinical use, inadequate delivery is the single most important factor limiting our ability to utilize molecular medicines to their full potential. The body has evolved a number of very efficient biological barriers to protect itself from the harmful effects of pathogens (bacteria, viruses) and foreign materials such as proteins. These include environmental interfaces such as the gastrointestinal tract and the skin, internal cellular barriers such as the endothelial lining of blood vessels (particularly the tight blood–brain barrier), immune surveillance, and not least the cell membrane and inherent compartmentation of intracellular organelles. The current challenge is to design innovative devices and technologies (including nanomedicines) that will help guide the therapeutic to its correct location of action (drug targeting) and ensure that pharmacological activity is maintained for an adequate duration (controlled release).

1 Introduction

Two distinct approaches are customarily taken in the search for improved medicines. The first area is the identification of a distinctive structure that will make a perfect fit with a unique target in the diseased cell. Theoretically, this should generate the "perfect" drug molecule with exquisite therapeutic activity and no side effects. In reality, it is usually difficult to find or synthesize such "perfect" drugs. Efforts continue; these include screening of natural product molecules to identify candidates with pharmacological activity, preparation of synthetic low molecular weight drugs via traditional medicinal or combinatorial chemistry, and identification of natural macromolecules, including antibodies, proteins, and oligonucleotides that have inherent biological activity.

The second, and complementary approach is to create a drug delivery system that will act as a vehicle to carry and guide more precisely the above-mentioned agents to their desired site of action. For the last three decades, pharmaceutical scientists, with expertise in cell and molecular biology, chemistry, pharmaceutical technology, and medicine have been working in multidisciplinary teams to better understand the biological rationale for design of effective drug delivery systems. Once gained, this knowledge is used in combination with sophisticated synthetic chemistry, recombinant

molecular biology, and sometimes also materials science and engineering, to construct the delivery system appropriate to the particular drug payload. This article provides an introduction to the field of targeting and intracellular delivery (biological barriers and technologies used for targeting), describes the first-generation technologies carrying low molecular weight drugs and proteins that have already entered routine clinical use over the last decade (i.e. the first "nanomedicines"), and also reflects upon the challenges that still must be overcome if we are to efficiently deliver macromolecular medicines such as antibodies, proteins, and oligonucleotides into specific intracellular compartments.

The principal challenge remains the identification of safe and effective delivery systems able to localize macromolecular therapeutics to the diseased cells or tissues and once there, to promote their efficient delivery to the required intracellular compartment (*drug targeting*) and ensure that they are made available for the appropriate duration of time (*controlled release*).

1.1 Challenges for Targeting and Intracellular Delivery

Why is this field so important? Advances in the understanding of the molecular basis of diseases has brought expectations that novel medicines, able to dramatically improve treatment of life-threatening, and chronic debilitating diseases, are just around the corner. Despite progress in genomics and proteomics research, it has proved much more difficult – than originally envisaged – to realize effective and practical therapies based on protein-, peptide-, and oligonucleotide-based therapeutics into routine clinical practice. Conversion of such innovative macromolecular therapeutics into medicines has been delayed by the lack of parallel, equivalent investment in enabling "drug delivery" technologies. For example, from 1989 to January 2004, there have been more than 900 trials mostly focusing on cancer as a target. There is still no gene therapy medicine approved for routine clinical use. Similarly, while the biotechnology revolution is bringing increasingly widespread use of protein- (particularly antibody) and peptide-based therapeutics, these macromolecular drugs are still administered to the patient parenterally, for example, by injection. There is a pressing need for more convenient to use medicines, perhaps manufactured as a tablet, capsule, transdermal (across the skin) delivery system, or inhalation device, as these formulations are much more convenient for the patient to use at home. Most importantly, there is a pressing need to discover improved therapies (new drugs or drug delivery systems) that can cure life-threatening diseases (e.g. cancer and drug-resistant infectious diseases) and palliate the effects of chronic debilitating diseases (e.g. arthritis, multiple sclerosis, and senile dementia).

1.2 Principles of Drug Targeting and Controlled Release

Since the pioneering work of Paul Ehrlich at the beginning of the last century, the potential benefits of disease-specific targeting, using both chemotherapy and immunotherapy, have been well appreciated. Although progress in recent decades has been remarkable, these concepts have never been fully realized to practice.

Ehrlich coined the phrase "magic bullet" to describe targeted therapy. In the modern era, phrases such as "guided missiles" or "stealth delivery systems" are also often used. Most low molecular weight pharmaceuticals distribute randomly throughout the body after administration and few are perfectly selective in terms of action. Therefore, the desire to identify strategies for better drug targeting has been growing exponentially. Increasingly sophisticated systems are being explored, including engineered devices, technologies of nanoscale size (1–1000 nm) and recombinant or synthetic hybrid molecular conjugates.

When trying to design an ideal "drug delivery system," it is often necessary to consider two distinct factors. These work in concert to achieve maximum therapeutic benefit. The first is the aforementioned need to deliver the drug to its target site of action that is, *drug targeting*. The second is the need to deliver pharmacologically active drug at an appropriate rate, giving a therapeutic concentration and for an appropriate period of time, the exact duration of which depends on the mechanism of action of the particular compound (Fig. 1a). Minimum toxicity can be achieved by ensuring that drug concentration does not rise above the maximum tolerated concentration. Over the last three decades, many innovative drug delivery systems have been designed to sustain drug levels over a prolonged period; this can be over a day, a month, or even a year. These delivery systems are called *controlled release systems*, and they can have various forms including drug-containing tablets and capsules, polymer-based depots or devices.

Most modern drug delivery systems administered parenterally incorporate both the features of *drug targeting* and *controlled release*. They aim at targeting the drug to the diseased cell, and once there, are programmed to release the drug at an appropriate rate (geared to its mechanism of action) for a suitable duration of time.

It is important to note that drug delivery systems designed for *drug targeting* accomplish this goal at three different levels (Fig. 1b). Many diseases are resident in a single organ; therefore, drug delivery systems can be devised to localize high concentrations of drug locally within that particular organ. This phenomenon has been called *1st order targeting*. Lung targeting is used as an example in Fig. 1b. In this case, the drug may be delivered from the exterior using an inhaler to distribute drug-containing particles onto the lung surface, that is, topically, or alternatively by intravenous administration of a delivery system able to localize to the lung as a result of lung vasculature-specific targeting. While bringing the benefit of elevated local drug concentrations, 1st order targeting has the disadvantage that it is does not focus the agent specifically into diseased cells.

If cell-specific targeting can be achieved, for example, targeting to cancer cells within the lung, the process is referred to as *2nd order targeting*.

The ultimate challenge is often to localize the drug (especially if it is a macromolecular therapeutic) to a specific intracellular compartment within the target cell. This might be an intracellular vacuole if the target is a pathogen (e.g. leishmania, mycobacteria), the cytosol in the case of an antisense oligonucleotide or, for gene therapy, it may be necessary for a plasmid to be delivered to the nucleus. The ability of a drug delivery system to localize to a compartment at the subcellular level has been defined as *3rd order targeting*.

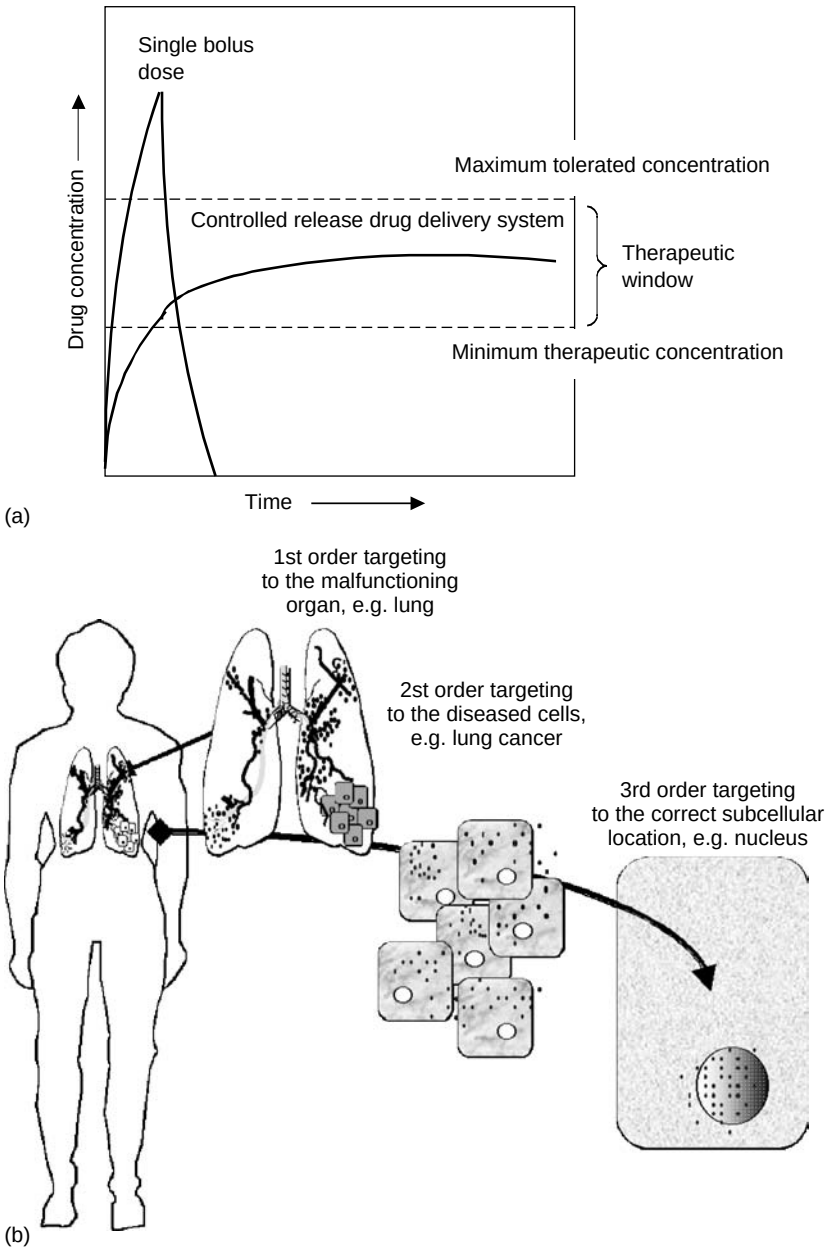

Fig. 1 Rationale for design of targeted drug delivery systems. Panel (a) illustrates the need to release drug at the desired concentration (active but not toxic) over the required period of time. Panel (b) different levels of lung targeting as an illustration of the concepts of first order (to the organ), second order (to the diseased cell within the organ), and third order targeting (to a subcellular compartment within the diseased cell).

1.3 Biological Barriers

1.3.1 Whole Organism Level

Even getting drugs into the body can be a formidable challenge. The body is well protected via the epithelial cell barriers that act as a coat of armor. For example, a layer of columnar epithelial cells lines the gastrointestinal tract, nose, and lung, and the differentiated squamous epithelium that forms skin covers external body surfaces. Depending on their physicochemical nature, low molecular weight drugs will often pass across the gastrointestinal tract, skin, or the pulmonary, nasal, vaginal mucosa. Even so, there are still a number of physicochemical and biological barriers to circumvent to ensure adequate bioavailability. Particular hazards include premature metabolism, for example, by the cytochrome P450 enzymes, and/or ejection of drugs from the very cells they are trying to cross by membrane-localized efflux pumps such as P-glycoprotein. The cytochrome P450 enzyme CYP3A4 is expressed at high levels in the intestinal mucosa and it is known to recognize as substrates more than 50% of drugs administered orally to humans. The complexity of each barrier depends on the particular route of administration and further details can be found in more specialized texts. Table 1 lists the principal routes used for drug (and delivery system) administration. Many specialized *in vitro* models of biological barriers have been established to enable the study of transport processes and to facilitate drug delivery system optimization.

Small, relatively lipophilic molecules have the advantage that they can usually traverse biomembranes. Macromolecular drugs (including proteins, peptides, oligonucleotides) and the newly emerging drug delivery technologies including protein carriers, liposomes, nanoparticles, immunoconjugates, and polymer therapeutics (see below) in general cannot. To gain access to the body, and subsequently travel to those tissues within which the target cells lie, they must either pass across the cells (*transcellular passage*) that comprise the epithelial barrier by hijacking physiological transport processes such as transcytosis or alternatively pass between these cells (*intercellular passage*) as shown schematically in Fig. 2. In the latter case, they must be small enough to pass through the intercellular junction complex or be able to transiently open the tight junctions. Intercellular junctions can be of three types: a tight junction, a gap junction, and a desmosome. Passage of molecules is limited by their size. Whereas the gap junction has a diameter of 2 to 3 nm, the tight junction can be as small as 0.8 to 2.0 nm, so transport across is usually restricted to molecules of <200 to 500 Da molecular weight. It should be noted that microparticles and nanoparticles are taken up and transported across the gastrointestinal tract by the immunosurveillance cells called M cells. This physiological pathway has been explored as a route for delivery of vaccines and drugs.

1.3.2 In the Circulation

To bypass these external barriers, many drug delivery systems are injected parenterally (e.g. by intravenous, intraperitoneal, subcutaneous administration etc.). However, irrespective of the route of administration, travel through the bloodstream is usually required to reach the target cells. This presents an additional series of biological hurdles. The reticuloendothelial system (RES), that is, the macrophages in the liver and spleen, is primed to eliminate circulating "foreign"

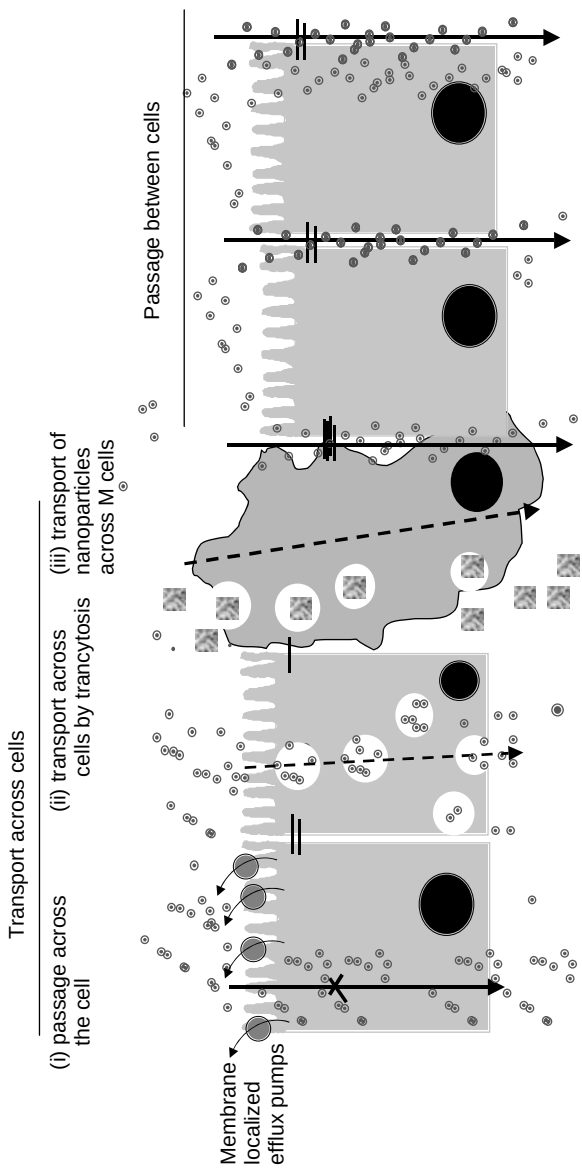

Fig. 2 Possible mechanisms of transport across an epithelial cell barrier. The gastrointestinal tract is given as an example, and in this case, the specialized M cells are also able to take up and transport nanoparticles and microparticles. (i) Transport of low molecular weight compounds (particularly if they are hydrophobic) can occur across the cell by transcellular diffusion or by active transport. (ii and iii) Alternatively, low molecular weight compounds and macromolecules and/or small particles can be transferred by transcytosis. Finally, very small molecules (ions, solutes) can potentially pass through the junctional complex and this movement between cells is termed the *paracellular route*.

Tab. 1 Routes of administration.

Route	***Use***	***Comments***
Noninvasive		
Oral	A wide variety of low molecular weight drugs are administered via this route.	Convenient for the patient. However, poor and irreproducible bioavailability can be a problem. This route is not well suited for delivery of macromolecular drugs that are poorly absorbed and/or degraded in the gastrointestinal tract.
Buccal	An alternative to oral delivery well suited to specific drugs.	Rapid access to brain avoids first pass metabolism in the liver.
Colonic	Can be achieved by use of suppositories or oral formulations designed for time-release or site-specific release in the colon.	Suppositories are a popular alternative to oral delivery in some countries; they can be useful in the event of the patient suffering from nausea or vomiting.
Vaginal	Being used as a means to block HIV virus entry.	Convenient for localized applications such as treatment of sexually transmitted diseases.
Ocular	Topical application to the eye or local injection of drugs expected to act locally.	Eye drops are convenient for the patient to use but often the drug is rapidly removed and it is difficult to maintain therapeutic concentrations.
Pulmonary	Delivery of drugs with target action in the lung or as a means of gaining access to the bloodstream.	The large absorptive surface of the extremely thin lung mucosal membrane and its good blood supply has given hope that this route can be used for systemic delivery of peptides and proteins.

Intranasal	Delivery of drugs that would be inactivated in the GI tract.	Although a convenient route for administration, reproducible delivery can be compromised if the patient has a cold or hay fever.
Topical and transdermal	Topically administered drugs for treatment of skin-related diseases and also the use of transdermal patches to deliver drugs across the skin.	Drugs with suitable correct physicochemical properties pass through the skin into the bloodstream. This has the advantage of avoiding first pass metabolism. Recently, delivery systems that promote transport using iontophoresis, ultrasound, or microneedles have also emerged. These painless techniques transiently disrupt the skin's barrier but permeability can be dependent on age and site of application.
Parenteral		
Local implantation	Use for delivery at the site of disease and to implant controlled release systems.	Ability to place the delivery device at the disease site, for example, brain cancer, blood vessel. Also can be tailored to control release over short or long periods of time. Not suitable for targeted delivery to disseminated disease.
Intravenous, subcutaneous, and intramuscular	Delivery of drugs that would be poorly absorbed using other routes and in circumstances where an immediate effect is required.	The only reliable route of administration for macromolecular drugs and compounds inactivated in the GI tract or poorly absorbed there.
Intraperitoneal	Delivery of drugs expected to act locally, for example, ovarian cancer.	Potential difficulties for repeated administration and rapid transfer to the circulation diminishes regional specific targeting.

particulate matter and "foreign" macromolecules, and can accomplish this task within minutes. The tendency of antibodies and blood proteins to adhere to circulating drug delivery systems (opsonization) potentiates RES-mediated clearance. Even if a drug delivery system is designed to target directly accessible cells (e.g. blood cells, the vascular endothelium, or liver cells), efficient targeting can take time and thus can be compromised by rapid RES clearance. The aim is often targeted delivery to cells in the extravascular space. To accomplish this goal, RES-avoiding, long-circulating, delivery systems are used. In this case, the vasculature itself may present a daunting physical barrier due to the simple squamous epithelium (made up of the endothelial cells) that lines the vessels. Depending on overall size and physicochemical properties, exit of the drug delivery system from the bloodstream (extravasation) can be difficult. These features also control subsequent, diffusion-mediated, penetration into a target tissue. The tight endothelial barrier in the brain, the so-called "blood–brain barrier" (BBB), is particularly difficult to penetrate.

The blood vessel endothelial cell wall is specifically designed to retain circulating proteins and cells. Smaller macromolecular-based drug carriers, and nanoparticles, usually 5 to 20 nm in size, can escape into extravascular compartments using the routes taken by proteins and antibodies that constantly circulate via more permeable, fenestrated blood vessels, into the lymphatic system and back into the circulation. For instance, they can escape more easily through vessels in the peripheral capillary beds, the peritoneal cavity, liver, and in the alveoli of the lung. In some cases, these smaller carriers also have the ability to utilize transcytosis pathways that facilitate transport of natural macromolecules across the endothelial wall. Once in the interstitial space the extracellular matrix can be an additional barrier. Charged or hydrophobic carriers can bind to it or become entangled there. Moreover, deeper penetration of antibody-targeted systems into a tissue can be compromised by efficient binding to antigen near the blood vessels; the so-called "antigen-binding barrier". Premature binding of an antibody to antigen shed into the bloodstream or extracellular fluid also neutralizes targeting potential.

Larger carriers and complexes, such as liposomes and polymer/DNA complexes (polyplexes) usually have a diameter of 30 to 150 nm and consequently they have greater difficulty in escaping normal blood vessels and subsequently penetrating far into complex tissues. Nonetheless, all long-circulating carriers share the ability to selectively extravasate into sites of enhanced vascular permeability, by a process known as *passive targeting* (see below). In disease, enhanced permeability typically arises at sites of local inflammation (e.g. during infection, arthritis) and at sites of new blood vessel formation, for example, the angiogenesis that occurs to sustain tumor development. Even liposomes of diameter up to 800 nm have been found to escape angiogenic vessels.

While small size is advantageous in terms of extravasation and avoidance of the RES surveillance, if the carrier is too small (e.g. a protein of $<60\,000$ Da or a polymer/particle $<40\,000$ Da or $<\sim 5-6$ nm in diameter) there is a high probability of rapid renal elimination via the pores in the kidney glomerulus. This is disadvantageous if a delivery system is removed prematurely before it can access the desired target. However, the

kidney receives ~25% of the cardiac output, so conversely renal filtration has potential for exploitation as a route for drug targeting to renal tubule cells. These cells are physiologically programmed to reabsorb certain small molecules (e.g. folate, saccharides, and amino acids) and some low molecular weight charged or hydrophobic macromolecules. Strategies under investigation have used low molecular weight proteins or synthetic polymers as renal tubule-targeted drug carriers (e.g. lysozyme-naproxen and lysozyme-captopril). The renal tubule cell tropism of antisense oligonucleotides has also been explored as a means to localize their pharmacological activity.

In addition to the above-mentioned physical barriers, in transit, the drug delivery system and its payload encounter a large number of biochemical barriers. Premature metabolism of drug or carrier can occur at every step. Proteases and nucleases are a particular hazard when trying to deliver proteins and oligonucleotides.

1.3.3 Cellular Level

Once a drug delivery system arrives safely at the target cell, yet another series of biological barriers present themselves. These include the physical barrier of the plasma and intracellular vesicle membranes, problems of intracellular organelle compartmentation (e.g. penetration into the nucleus), and a series of biochemical and metabolic barriers; for example, endocytic vesicle pH, the catabolic/metabolic enzymes present lysosomes, cytosol, and other intracellular organelles.

Low molecular weight, lipophilic drugs often readily traverse the membrane by diffusion, facilitated-, or active transport. However, they still have to bypass membrane-localized efflux pumps, which are a particular problem in drug-resistant tumor cells. Generally, natural macromolecules, macromolecular drug delivery systems, liposomes, nanoparticles, and polymer–DNA complexes (polyplexes) cannot easily traverse the plasma membrane. Their cellular uptake is almost exclusively limited to the endocytic route. Figure 3 shows the main intracellular trafficking pathways and illustrates the dynamics of the transport processes, the rapid rates of transfer between compartments, and the local pH encountered. Escaping the endocytic vesicular compartments presents a major challenge for a drug delivery system. This hostile, intravesicle environment can inactivate many low molecular weight drugs, and all proteins, peptides, and oligonucleotides that arrive there unless they rapidly escape into the cytosol or can be protected in transit by the drug delivery vehicle. The membrane-localized ATP driven H^+ pump is responsible for vesicle acidification and pH values of 4.5 to 6.5 are typically seen. This may change in disease. Additionally the lysosomal compartment contains a battery of hydrolytic enzymes (including peptidases and proteases, nucleases, lipases, and glycosidases) well able to degrade all natural macromolecules that arrive there.

As the endocytic pathway is the principal route of entry for targeted delivery systems, its mechanism(s) are of fundamental importance. Endocytosis involves the internalization of plasma membrane in the form of membrane-bounded vesicles. This is a constitutive process in all eukaryotic cell types, although each cell type may have a specialized endocytic machinery to allow it to fulfill particular physiological functions. Endocytosis fulfills a multitude of functions including removal of aging proteins from the extracellular fluids leading to their catabolism;

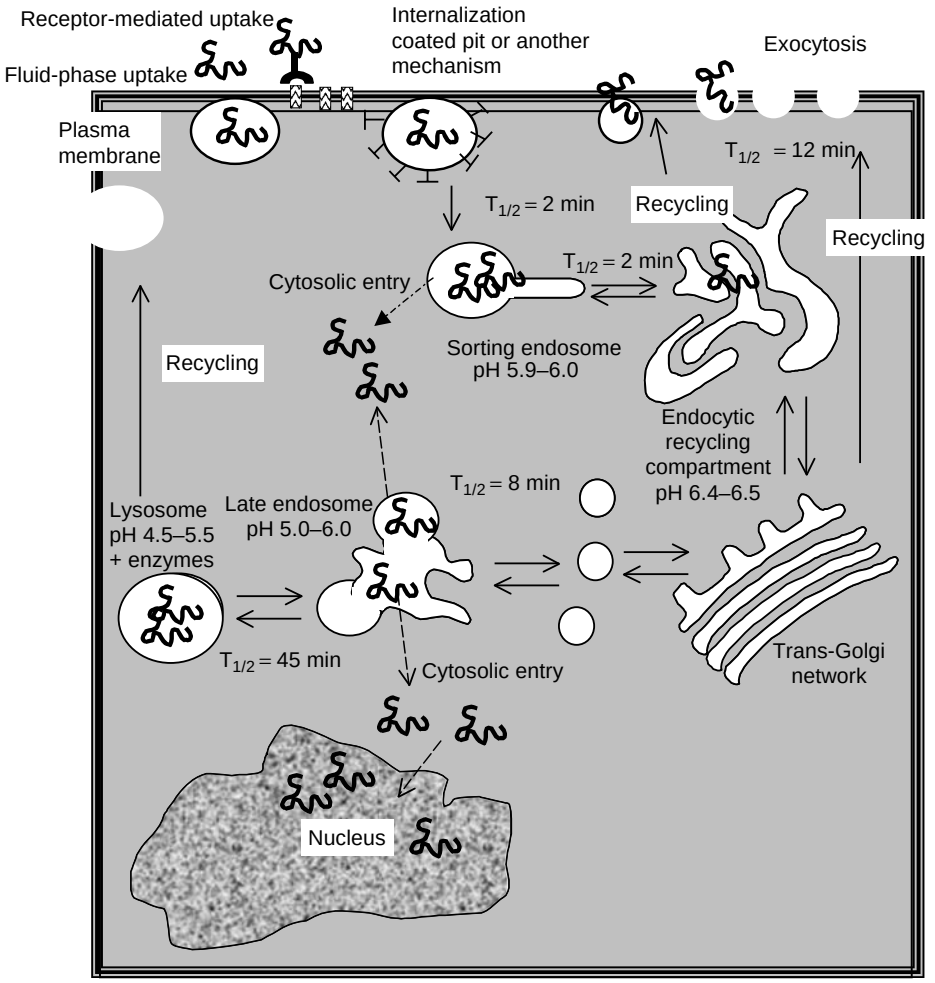

Fig. 3 Schematic diagram showing the main pathways for endocytic uptake and intracellular trafficking of macromolecular and particulate drug carriers. Of particular relevance is the decreasing pH encountered within intracellular vesicles and the potentially rapid exchange between one compartment and another. Uptake can occur simply due to the presence of the delivery system as a solute or colloidal suspension in the extracellular fluid (fluid-phase endocytosis) or as a result of receptor-mediated uptake. The cargo can potentially be directed into endosomal and lysosomal compartments, may be rapidly recycled and eliminated from the cell by exocytosis, or might potentially escape the endosomal vesicular network and localize to other intracellular organelles such as the nucleus.

internalization of nutrients required by the cell for anabolic synthesis (cholesterol, amino acids (via protein degradation), iron etc.), and regulation of plasma membrane receptor number and consequently signaling pathways.

It is noteworthy that there are many potential doorways for entry, and the

mechanisms of vesicle formation are ligand- and cell-type dependent. Internalization can occur via phagocytosis (uptake of larger particles usually by macrophages), clathrin-coated pits (receptor-mediated endocytosis), caveolae (cholesterol enriched domains particularly implicated in endothelial cell transcytosis), and clathrin-independent vesicles (those responsible for micropinocytosis). Uptake of natural molecules is usually triggered by ligand–receptor interaction, and assembly of membrane-associated, multiprotein complexes with signaling cytosolic domains carefully regulating the process of internalization. The membrane composition of vesicles and tubules (lipids and proteins) that comprise the endocytic network is constantly changing. Transfer of the contained cargo is regulated by vesicle–vesicle fusion events controlled by the drop in pH, vesicle membrane curvature, multiprotein complexes on the cytoplasmic face of the vesicle, and also the membrane lipid and protein composition of each compartment. Specific receptors are sorted into pathways that direct them to (1) the lysosomes (e.g. the asialoglycoprotein receptor of liver hepatocytes which traffics to the lysosome); (2) recycle them back to the plasma membrane (e.g. the transferrin/transferrin receptor complex; or (3) the Golgi network (e.g. protein toxins such as ricin).

With the biological rationale for design very much to the fore, current efforts are trying to achieve biomimetic, bioresponsive, macromolecular, and nanosized particulate delivery systems that are able to mimic natural transport vectors such as proteins (including protein toxins) or viruses. The overall goals are (1) design of constructs able to escape the immunosurveillance and metabolic pathways that lead to premature elimination; (2) cell-specific targeting; and (3) cellular entry by translocation across the plasma membrane or, using specific endocytic transport pathways, compartment-specific trafficking.

1.4 Opportunities for Targeting

1.4.1 Passive Targeting to Diseased Tissue

The physicochemical features of many drug delivery systems disadvantage them in terms of targeting disseminated peripheral diseases (e.g. metastatic cancer). However, sometimes these inherent properties can be usefully harnessed in the context of the phenomenon called *passive targeting*. The fact that the RES and/or liver rapidly accumulate >80% of the dose (within 5–10 min) after administration of many nanoparticles, liposomes, and charged (+ve or −ve charge) and/or hydrophobic macromolecular carriers can provide an opportunity for 1st order passive targeting to the liver cells or spleen. In the form of the liposomally entrapped (or lipid-associated) antifungal/ anti-Leishmaniasis drug amphotericin B, cell-specific passive targeting to macrophages harboring pathogens or parasites has been exploited effectively clinically. Another example of passive targeting is the propensity of circulating microparticles (or aggregates formed by interaction of the drug delivery system with plasma proteins or blood cells) of diameter >7 μM to become rapidly trapped in the narrow capillary beds of the lung, again >80% dose within minutes. This also provides an option for 1st order lung-specific passive targeting. However, the approach requires caution due to the risk of pulmonary embolism.

If the carrier can evade RES capture and displays prolonged plasma half-life, it is termed *long circulating*. In this case, it can exhibit *passive targeting* via the "enhanced permeability and retention effect" (EPR effect). First described by Matsumura and Maeda in 1986, EPR-mediated targeting of tumors has been attributed to two principal factors (Fig. 4). First, the disorganized pathology of angiogenic tumor vasculature with its discontinuous endothelium leads to hyperpermeability toward circulating macromolecules and colloidal carriers. Second, the absence of lymphatic drainage in the tumor encourages subsequent tumor retention and progressive accumulation (Fig. 4). The EPR effect is being exploited by a number of drug carriers (including polymer-coated liposomes, polymeric micelles, and polymer–drug conjugates) to promote tumor targeting.

It is now well established that long-circulating macromolecules including albumin, protein-, and polymer–conjugates, polymeric micelles and liposomes accumulate passively in solid tumor tissue by the EPR effect. Intravenously administered polymer–drug conjugates can increase tumor concentration of antitumor drugs up to 70-fold and tumor biopsy following administration of the liposomal anthracyclines Doxil® and DaunoXome™ confirmed EPR-mediated tumor targeting in the clinical setting. The magnitude of targeting achieved by the EPR effect can be surprisingly high, up to ~20% dose/g tumor in murine models, and interestingly the extent of targeting seen (expressed as % dose/g tumor) is often highest in smaller tumors probably because they are more actively involved in angiogenesis. These values are as high as reported for tumor-specific antibodies in rodent tumor models.

1.4.2 Active Targeting to Diseased Cells

Most urgently required are drug delivery systems that do not show inherent liver, spleen, or lung tropism. If the carrier is long circulating, it can be decorated with ligands to promote receptor-mediated uptake, that is, exhibit *active targeting*. Attempts to promote cell-specific delivery have explored a vast array of targeting ligands (see Table 2 for examples), particularly in an endeavor to realize tumor-specific delivery. Most widely explored have been saccharides, peptides, proteins, and antibodies and their fragments. Ideally, the target antigen or receptor must be present at a high density on the cells of interest. It should exhibit homogeneity of expression (preferably with no or minimal expression on normal cells), and the receptor should neither be shed into the bloodstream nor be downregulated upon ligand binding. In addition, receptor fate (whether it is internalized or not) must match the needs of the particular targeted therapeutic being delivered. Some delivery systems are active at the level of the plasma membrane; others have a requirement for internalization. Endogenous ligands present in extracellular fluids (e.g. transferrin or folate) may compete with targeted delivery systems so ideally their concentrations should be low.

The development of monoclonal antibody technology by Kohler and Milstein in the 1970s gave hope that the exquisite specificity of antibody–antigen interaction would provide perfect cell-specific targeting. However, antibody-targeted drug delivery systems have been relatively slow to come into practice. Challenges have included problems in identifying disease-cell specific antigenic determinants; antibody immunogenicity, nonspecific capture of antibodies by cells of the RES due to

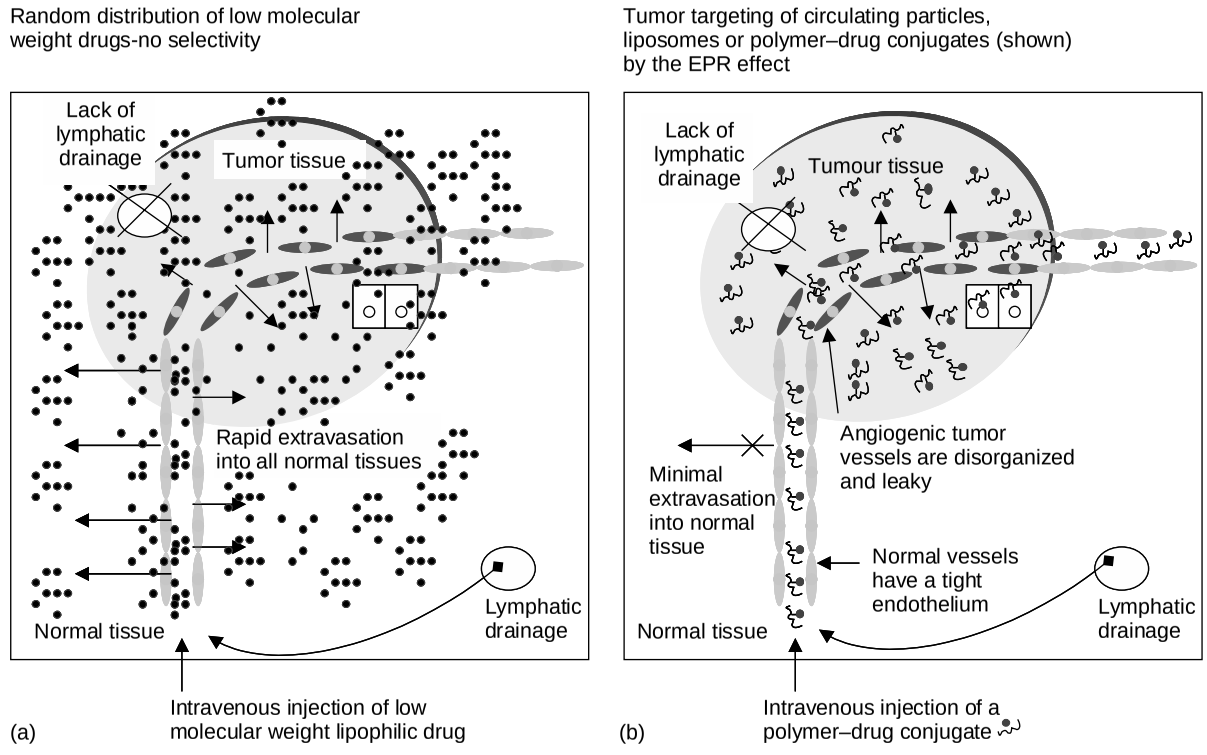

Fig. 4 Tumor targeting by the enhanced permeability and retention (EPR) effect. Panel (a) shows the random distribution of low molecular weight chemotherapy after intravenous injection. Panel (b) shows the vascular retention of a macromolecular drug conjugate (the same would be true for a colloidal carrier) which leads to enhanced extravasation into tumor tissue (via hyperpermeable blood vessels) and the drug delivery system progressively accumulates there due to the absence of lymphatic drainage.

Tab. 2 Ligands used to promote cell-specific targeting.

Ligand	*Target cell*	*Receptor/Antigen*	*Carriers*
Saccharides			
Galactose	Hepatocytes	Asialoglycoprotein receptor	Polymer and protein conjugates
Mannose	Macrophages	Mannose receptor	Polymer and protein conjugates
Sialyl Lewis-X	Endothelial cells	E-selectin	Liposomes, polymers
Peptides and proteins			
MSH	Melanoma	MSH receptor	Polymers
EBV peptide	Lymphocytes	CD21	Polymers
RDG peptides	Endothelial cells	Integrins $\alpha_v\beta_3$	Polymers, peptide-DNA complexes, lipoplexes
Transferrin	Various tumors	Transferrin receptor	Polymers, liposomes
FGF	Various tumors	FGF receptor	Polymers
LDL	Various tumors	LDL receptors	Protein conjugates
Modified albumin			Protein conjugates
Antibodies or fragments			
Anti-Thy 1.2	T lymphocytes	Thy 1.2 antigen	Polymers
Anti-CD19	Multiple myeloma	CD19	Liposomes
Antitransferrin receptor mAb	Various tumors	CD71	Polymers, liposomes, immunoconjugates
B1mAb	Leukemia and lymphoma	BCL1	Polymers
OV-TL	Ovarian carcinoma		Polymers
Anti-ELAM-1, VCAM-1 and ICAM-1 mAbs	Vascular endothelial cells	ELAM-1, VCAM-1 or ICAM-1	Liposomes, immunoconjugates
Anti-HER2-Fab'	Breast cancer	HER2	Liposomes
Anti-VEGF receptor mAb	Vasculature of solid tumors	VEGF receptor	Liposomes
Anti-CD20 mAb	B-cell lymphoma	CD20	Immunoconjugate, liposomes
Anti-TAG72 mAb	Colorectal and other solid tumors	TAG72	Immunoconjugate, liposomes
Anti-EGFR	Glioma	EGFR	Recombinant immunotoxin containing Pseudomonas exotoxin
Other			
Folate	Various tumors, particularly ovarian	Folate receptor	Polymer conjugates and liposomes
Hyaluronic acid	Murine melanoma	CD44	Liposomes

Notes: MSH: melanocyte stimulating hormone; FGF: fibroblast growth factor; EBV peptide: EDPGFFNVE, an Epstein–Barr virus epitope; LDL: low-density lipoprotein.

antibody Fc recognition; generation of antiideotypic antibodies that inactivate the antibody specificity or lead to the formation of immunocomplexes that are rapidly cleared. Immunoconjugates can have large size (Mw (weight average molecular weight) >150 000 Da) and this fact, together with the antigen-binding barrier, can lead to poor tissue penetration. However, exciting progress in the field of therapeutic antibodies (e.g. the breast cancer therapy Herceptin®, the recently approved antiangiogenic antibody Avastin®, and commercialization of the first antibody–drug conjugate Mylotarg®) has revitalized the field of antibody-targeted drug delivery systems. Recombinant techniques are being used to humanize antibodies and the versatility of chemical conjugation is being used to create antibody–(or Fab' fragments) polymer conjugates, and nanoparticle- and liposome-targeted systems.

Other ligands have been widely used to promote active targeting, and a list of examples are given in Table 2. Carbohydrates have been widely explored. For example, mannose can be used to target polymeric and protein drug carriers to macrophages. Galactose promotes targeted delivery to the asialoglycoprotein receptor present on normal hepatocytes and hepatocellular carcinoma. Using polymer–drug conjugates, galactose targeting has been explored clinically with the hope of improving treatment of primary and secondary liver cancer. Liver targeting was verified clinically using clinical gamma camera imaging. It is noteworthy that galactose targeting localizes drug within normal hepatocytes as well as hepatoma, so this approach relies on diffusion of the overall high local drug concentration into neighboring cells to create, what has been termed, a *bystander effect*. Phage display techniques have become popular as a means to identify peptide-mediated disease-specific targeting opportunities. For example, peptide hormones such as melanocyte stimulating hormone (MSH) and somatostatin have been employed as tumor targeting ligands, and RDG-containing peptides have been widely used to localize to β-integrins present on endothelial cells. In this case, RGD-containing cyclic peptides having a disulphide bond between two cysteine residues to constrain structure into a preferable motif have the highest receptor affinity. The fact that some cells, including some tumor cells, have elevated levels of LDL receptors, growth factor receptors (e.g. FGFR and EGFR), and folate receptors has encouraged their exploration as potential targets. For example, folate-targeted liposomes have been used to improve delivery of oligonucleotides and anticancer agents.

1.4.3 How to Realize Intracellular Targeting?

Even on arrival at the target cell, direction of macromolecular medicines and nano-sized drug delivery systems to the desired intracellular target remains a significant challenge. If the agent can act at the level of the plasma membrane, the journey is over. If the target lies within the endosome, lysosome, or Golgi, internalization and delivery via the endocytic trafficking pathways is relatively simple. For example, many transferrin-bearing drug delivery systems (including liposomes, nanoparticles, polymer–drug conjugates, and polyplexes) and antitransferrin receptor antibody–drug conjugates have been described. Receptor-mediated internalization leads to rapid and efficient uptake and this has been used, for example, to enhance gene delivery by nonviral vectors. Transfer of internalized macromolecular drugs (e.g.

proteins, oligonucleotides) and drug delivery systems (liposomes, nanoparticles, antibody, and polymer conjugates) to the lysosomal compartment is often unavoidable. Addition of ligands such as galactose, and LDL to various drug delivery systems is used to ensure lysosomotropic delivery. In this context, many polymer-, protein-, or antibody–drug conjugates have been successfully designed to use decreasing pH or lysosomal enzymes as the trigger to liberate low molecular weight conjugates drug. For example, pH-sensitive *cis*-aconityl and hydrazone linkers liberate drugs in acidic environments. In addition, the lysosomal thiol-dependent proteases (such as cathepsin B) have been widely used to activate polymer–drug conjugates synthesized to contain peptidyl polymer-drug linkers that are enzyme substrates, or designed using a polymeric carrier (e.g. polyglutamic acid) whose backbone is degraded by these enzymes.

The main challenges that remain are targeted delivery to other intracellular compartments and transfer of macromolecular medicines into the cytosol and/or the nucleus, and the barriers to gene delivery have been well documented. Several strategies are being explored to facilitate cytosolic delivery (Fig. 5). These include (1) improved transfer across the plasma membrane using fusogenic liposomes, membrane translocating peptides, pore-forming proteins, and membrane-active peptides; (2) following endocytic uptake, transfer out of the endosomal compartment into the cytosol with the aid of pH-sensitive endosomolytic polymers or peptides, pH-sensitive fusogenic liposomes and recombinant proteins; and (3) tailoring bioresponsive linkers (usually pH-sensitive or protease sensitive) to liberate a drug payload from polymeric and protein carriers when exposed to the appropriate trigger in the endosomal pathway.

Membrane translocation is a key step. It is becoming increasingly evident that many proteins and receptors can bypass the classical endocytic pathways to enter the cell, as they are equipped with signal peptide motifs that promote translocation across the plasma membrane (Table 3). Similarly, viruses and toxins have peptide sequences that change conformation on entry into the acidic endosome and permeabilize the endosomal membrane facilitating entry. As these peptide motifs become characterized, they are increasingly incorporated into drug delivery systems to facilitate cytosolic access, either using hybrid, recombinant proteins or with the aid of synthetic chemistry to incorporate them into a polymeric, liposomal, or nanoparticulate carriers.

As well as the peptidyl membrane translocating sequences, peptides that are membrane active at pH 7.4, such as the bee venom toxin melittin, are also being explored as components of drug targeting systems. However, in most cases, the pH gradient between extracellular fluid and the endosomal/lysosomal compartments is the important biotrigger used to promote delivery. As an alternative to peptides, a growing number of pH-responsive endosomolytic polymers have been synthesized as vectors for plasmid and protein delivery. Potentially, their reduced or lack of immunogenicity provides greater suitability for *in vivo* administration. Polymers used include polycations such as polyethyleneimine and polylysine, polyanions such as polypropylacrylic acid derivatives, amphoteric polymers such as linear polyamidoamines, and the hyperbranched polyamidoamine dendrimers. Their exact mechanism of endosomal permeabilization is the subject of some

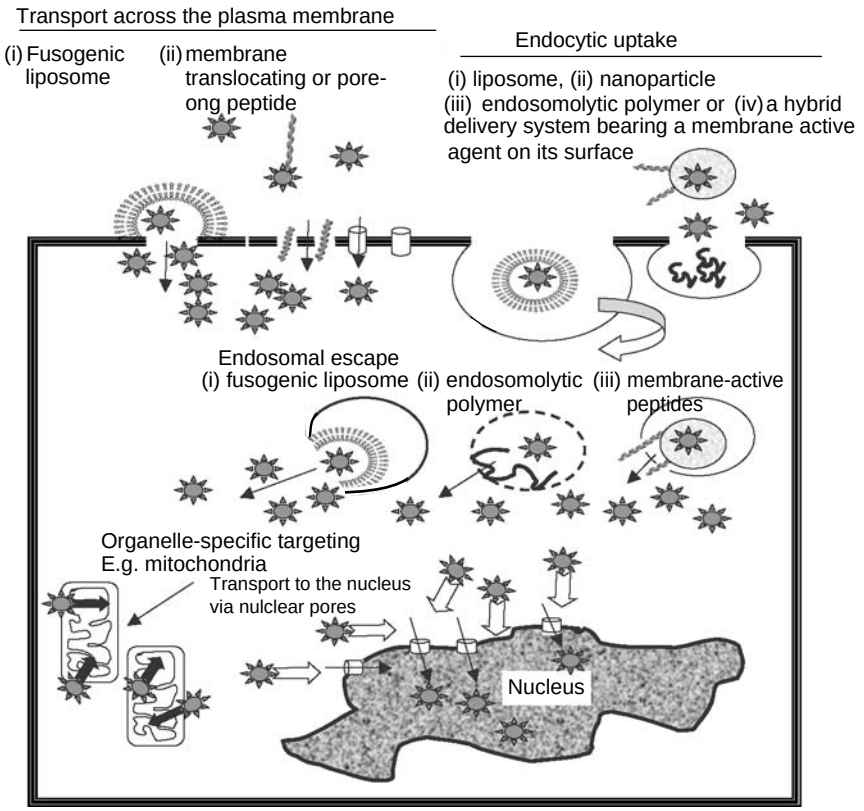

Fig. 5 Vectors used to promote cytosolic access and intracellular targeting. Three main approaches are used; vectors designed to fuse with the plasma membrane; vectors that can pass across the plasma membrane passage, and vectors that promote cytosolic access and/or organelle-specific targeting following endocytic uptake and entry into the endosomal compartments.

debate. It has been suggested that protonation of polycations results in swelling, and this, together with the influx of Cl^- ions and consequent osmotic effects, cause membrane destabilization. This is called the *proton sponge hypothesis*. However, it is clear that not all polycations act in this way. Linear polyamidoamines and polyamidoamine dendrimers seem to interact directly with the membrane to cause perturbation and perhaps pore formation. In the context of liposomal delivery, cationic helper lipids, for example, dioleoylphosphatidylethanolamine (DOPE), have been used to promote fusion in acidic pH environments.

Efficient delivery into the cytosol can be sufficient to ensure pharmacological activity of many macromolecular medicines, for example, antisense oligonucleotides and protein toxins. However, many oligonucleotides, plasmids, and protein drugs must subsequently traffic to the nucleus to achieve the desired effect. Stability of vector (and payload) in the cytosol and an appropriate rate of dissociation to deliver the payload can be problematic. Movement in the cytosol by diffusion can

Tab. 3 Motifs used to promote intracellular delivery targeting.

Ligand	*Function*	*Carrier/Use*
Membrane translocating peptides		
TAT protein transduction domain of HIV	Enhances endosomal escape after endocytic uptake	Fusion proteins or polymer conjugates
Herpes simplex virus type-1 22 protein – VP22	Facilitates an export-neighboring cell protein uptake pathway	Fusion proteins, plasmid delivery
Hydrophobic membrane permeable sequences (MPS) for example, AAVALLPAVLLALLAP	Delivers proteins and peptides across the plasma membrane Enhances endosomal escape after endocytic uptake	Recombinant fusion proteins or chemically synthesized peptide conjugates
Rhinovirus N-terminal peptide VP-1	pH-dependent, enhances endosomal escape after endocytic uptake	Polymers and polyplexes
Influenza virus N-terminal peptide HA-2	pH-dependent, enhances endosomal escape after endocytic uptake	Polymers and polyplexes
GALA	pH-dependent, enhances endosomal escape after endocytic uptake	Polymers and polyplexes
Membrane-active peptides		
Melittin	Pore-forming peptide from bee venom	Polymers and polyplexes
KALA		Polymers and polyplexes
Gramicidin S	Cationic peptide	Liposomes
Polyarginine ($n = 6 - 8$)	Cationic peptide	Polymers
Compartment-specific targeting motifs		
Lipophilic triphenylphosphonium cation	Localizes to mitochondria	Targeting drugs for example, antioxidants to mitochondria
KDEL motif	Promotes endoplasmic reticulum localization	Targeting oligonucleotides, antibodies, and toxins
A cluster of basic amino acids for example, PKKKRKV	Nuclear localization signals that promote nuclear pore binding before transfer to nucleus	Polyplexes designed for gene delivery, fusion proteins

be haphazard and slow, although it has been suggested that some polyplexes are more rapidly translocated to the perinuclear region of the cell in association with microtubules. Incorporation of peptidyl nuclear localization sequences (NLS) that have clusters of four or more cationic amino acids have become popular to promote nuclear trafficking and entry. Even so, nuclear entry of larger macromolecular and particulate carriers can be difficult if the cells are not dividing. Physiologically, macromolecules are imported through the nuclear pores, and there are approximately 3000 to 4000 pores per nucleus. However, in the closed state they will only allow passage of molecules <9 nm in diameter and even when open their diameter is only 26 nm.

With the realization that many diseases can be pinpointed to dysfunction of a specific intracellular compartment (or protein localized therein), many are now trying to design drug targeting systems aimed at specific intracellular organelles. For example, mitochondria dysfunction contributes to a wide number of diseases including, obesity, neurodegenerative diseases, ischemia-reperfusion injury, and cancer. Additionally, there are some 40 different lysosomal storage diseases that have a genetic basis, and it is known that lysosomal/endosomal malfunction can lead to accumulation of products such as lipofuscin, which is implicated in the progression of age-related neurodegeneration, for example, in Alzheimer's disease. A number of strategies are being explored as a means of targeting mitochondria. Interesting concepts at the early stage of development include "mitochondriotropic" vesicles prepared from dequalinium (diameter 70–700 nm) and use of mitochondrial signal peptides (MLS-peptide), both of which seem to promote mitochondrial targeting. Peptide localization signals are also being explored as tools for reaching other intracellular compartments, for example, to carry peptides and fluorophores to peroxisomes and for noninvasive cytosolic delivery of peptides and proteins.

1.5 Devices, Technologies, and Nanomedicines

Given the complex journey required to circumvent all these biological barriers, the ideal targeted delivery system will inevitably be multicomponent. It must be able to gain access to the body, and subsequently move through the body without recognition as "foreign" by the immunosurveillance systems. It must be biomimetic in the sense that it will be able to utilize natural transport pathways (such as transcytosis, endocytosis) and natural receptor-mediated targeting strategies at both the cellular and subcellular level to achieve the end goal. Ever more sophisticated devices are increasingly used to gain access to the body. These devices can be as simple as the syringe, or as complex as pulmonary inhalers and transdermal devices (including microneedle injectors, and the gene gun which is used to transiently physically damage the skin and thus allow particulate access). Sophisticated multicomponent oral delivery systems have also been designed for local and/or timed release. A detailed description of these devices is beyond the scope of this article.

While natural and synthetic polymers have traditionally found use as pharmaceutical excipients and biomedical polymers, it is important to note their expanding role as components of a wide variety of drug delivery systems. They are utilized to fabricate

devices, rate-controlling coatings, hydrogels for topical and parenteral controlled release, and as biodegradable polymeric implanted matrices for local controlled release of drugs, peptides, and proteins. Not least, the field of polymer therapeutics is now becoming well established and polymers are often components of other types of delivery vectors (see below).

At one time, antibody conjugates, liposomes, nanoparticles, and polymer conjugates were viewed as competing approaches. Naively it was thought that one technology would emerge as a universal platform for all targeting applications. It soon became evident that each has its individual advantages and disadvantages. Antibodies have potential for selective targeting, but even as humanized proteins they can be immunogenic, their pharmacokinetics is governed by molecular weight, and they have a limited drug carrying capacity. Liposomes have a high drug carrying capacity, but stability can be an issue (either releasing drug too quickly or entrapping drug too strongly) and they are prone to RES capture. Similarly, it has proven difficult to steer nanoparticles away from the RES and they can have relatively lower (than liposomes) drug carrying capacity. The versatility of synthetic polymer chemistry provides a unique opportunity to tailor synthetic, biomimetic, molecules having specific molecular weight and particular architecture. For linear polymers, this can provide multivalency needed for receptor recognition, and using dendrimer chemistry controlled 3D architecture can be achieved. In all cases, the polymer-drug linker can be tailored to facilitate site-specific drug release. However, the drug carrying capacity of polymer–drug conjugates is relatively low, conjugate structure can be complex and heterogeneous presenting challenges for characterization. Particular care must be taken to ensure that potential toxicity and/or immunogenicity of parenterally administered synthetic macromolecules is carefully addressed, particularly if the polymer is not biodegradable. There has been growing concern that polymeric nanoparticles, which have a particularly high surface area to volume ratio, could be exceptionally dangerous. However, even in this case any toxicity observed will clearly depend on route of administration, frequency of administration and the polymer (and chemistry) used to prepare the nanoparticle.

With the realization that an ideal platform for drug delivery would marry the benefits of these approaches, over the last decade families of hybrid, nanosized technologies have emerged. In general, they have multiple components optimized for each specific application considering location of the target site and nature of the drug payload to be delivered. There has been exponential growth in the transfer of such technologies into clinical development, and the number of Regulatory Authority approved products is growing year on year (see below). These factors underline the importance of these technologies as the first generation of "nanomedicines" (Fig. 6).

2 Vectors: Design, Applications, and Clinical Status

The literature relating to potential drug targeting systems increases year on year. Many claim design and/or synthesis of new drug delivery technologies. Where supporting biological data are provided, the vast majority of experiments are conducted *in vitro*. While these studies provide

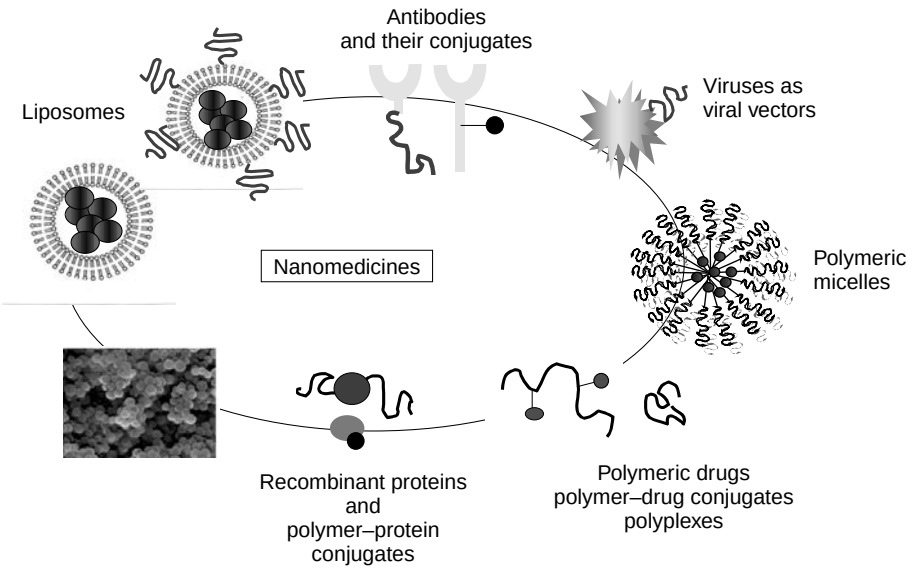

Fig. 6 Vectors that can be described as the first "nanomedicines."

the opportunity to quantitate receptor-mediated uptake by the target cells, we get an early impression of pharmacological potential, and it helps define intracellular localization; experience has shown that such *in vitro* assays rarely have value in predicting *in vivo* targeting potential. Nearly all technologies tested *in vivo* fail due to (1) inability to reach the target cell (often due to RES clearance); (2) inability to reach the correct intracellular compartment on arrival at the target cell; (3) inability to release/deliver drug at the appropriate rate on arrival at the target cell; and (4) not least the fact that many vectors display unacceptable toxicity and/or immunogenicity prohibiting safe human use.

In this section, only those vectors having well established *in vivo* and clinical targeting ability are described. In most cases, new technologies have been first explored in the context of cancer therapy. Globally over 6 million people die of cancers each year. Cure is only possible in ~35% of cases diagnosed, and commonly used cytotoxic chemotherapy is relatively ineffective due to lack of drug selectivity, nonspecific damage to healthy tissue, and multiple mechanisms of drug resistance (<5% of cures). The pressing medical need makes it ethically justifiable to administer novel drug delivery systems to cancer patients where the potential benefits far outweigh any risks associated with these new technologies. Once therapeutic benefit, and clinical safety of other vectors is established in cancer patients, exploitation in therapeutic indications has followed.

2.1 Liposomes

Liposomes were discovered in 1965 by Bangham and at that time used as model membranes. Since the pioneering work of Gregoriadis and Ryman in the 1970s,

liposomes have become widely used for targeted delivery of drugs, proteins, vaccines, and oligonucleotides. Several liposomal products have been approved as novel anticancer therapies, and for delivery of amphotericin B to treat opportunistic infections (Table 4). These vesicular carriers are prepared from phospholipids and cholesterol to form lipid bilayer-bounded vesicles. A single bilayer (unilamellar) or multiple bilayers (multilamellar) are used to entrap an aqueous interior (Fig. 7). Typically, those liposomes used clinically have a diameter of 80 to 150 nm. Hydrophilic drugs, including proteins, are entrapped in the aqueous interior and hydrophobic drugs carried within the lipid bilayer. Amphiphilic drugs that are weak bases or weak acids can be loaded into the liposomal interior using remote loading methods such as ammonium sulfate for doxorubicin and a pH-gradient method for vincristine to give high intravesicle concentrations. Alteration of their lipid composition (sometimes using cationic or anionic lipids) and/or the cholesterol:lipid ratio has been used to modulate properties. For example, cationic liposomes have been used to complex DNA and oligonucleotides on their surface.

The first liposomal formulations had major disadvantages including poor plasma stability, rapid blood clearance by the RES, and difficulty of manufacture on a commercial scale. However, these issues have been largely overcome using second-generation liposomes containing cholesterol to improve stability and also a hydrophilic polymer coating (particularly using polyethyleneglycol (PEG), so-called PEGylation) to minimize RES uptake and thus prolong plasma circulation. The PEGylated or ''Stealth'' liposomes are also able to capitalize on the EPR effect to promote passive tumor targeting. Although liposomes can be internalized by endocytosis, it is thought that delivery of entrapped antitumor agents is largely due to diffusion of drug out of the liposomes and into tumor cells. To promote receptor-mediated targeting of liposomes (active targeting) and cellular internalization, the field of ligand-targeted liposomes (LTLs) has been growing. For example, liposomes have been prepared with antibodies (or their fragments) on their surface. For example, anti-HER2 to target HER2-overexpressing breast cancer; anti-β–integrin Fab to target nonsmall cell lung cancer; antihuman E-selectin to localize to activated endothelial cells; folate and

Tab. 4 Examples of liposomal formulations in clinical use and clinical development.

Product	*Status*	*Payload*	*Indication*
DaunoXome®	Market	Daunorubicin	Cancer
Doxil®/Caelyx®	Market	Doxorubicin	Cancer
Myocet®	Market	Doxorubicin	Cancer
Ambisome®	Market	Amphotericin B	Fungal infections
Amphotech®	Market	Amphotericin B	Fungal infections
Liposomal-MTO	Phase I	Mitoxantrone (MTO)	Cancer
Aroplatin	Phase I/II	Platinate	Cancer
ATRA-IV	Phase I/II	Retinoic acid	Cancer
NX211	Phase I/II	A camptothecin	Cancer

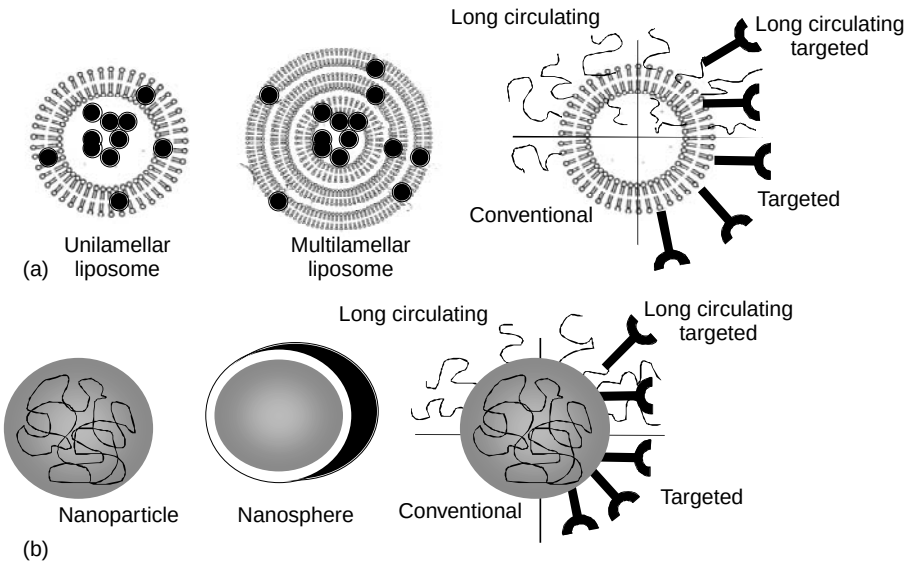

Fig. 7 Colloidal carriers used for drug targeting. A variety of liposomal (panel a) and nanoparticle (panel b) drug delivery systems have been developed for targeted drug delivery. The liposomal carriers can be unilamellar or multilamellar, drug can be entrapped in the aqueous core or within the lipid bilayer depending on its physicochemical nature. Nanoparticles can be made as a polymer matrix with drug entrapped in the polymer matrix, or as a nanosphere entrapping drug in the core. In both cases, ligands can be added to the surface to promote receptor-mediated targeting and a polymer coating applied to the surface to prolong plasma circulation times (usually PEGylation).

hyaluronic acid to target melanoma. PEG-coated liposomes containing cationic lipid have been shown to target chondroitin sulfate proteoglycans overexpressed on the surface of highly metastatic tumor cells. In addition, other experiments have reported neovascular targeting properties of cationic liposomes labeled with the fluorescent dye rhodamine in samples derived from patients with transitional cell cancer of the urinary bladder.

Clinical success of the anthracycline-containing liposomal formulations DaunoXome®, Myocet®, and Doxil® (called Caelyx® in Europe), both as single agents, and more recently as components of combination chemotherapy, really established liposomes as important drug carriers. The DaunoXome® aqueous formulation is composed of distearoylphosphatidylcholine (DSPC), cholesterol, and daunorubicin in a weight ratio of 18 : 7 : 1; that is, it has a daunorubicin concentration of 2 mg mL^{-1}. The liposomes are unilamellar vesicles of diameter ~45 nm. In contrast, Doxil® was the first polymer-coated liposome to come on the market. In this case, the vesicles are composed of hydrogenated soy phosphatidylcholine (HSPC); cholesterol; and diphosphatidylethanolamine (DSPE) conjugated to PEG in the ratio 56 : 38 : 5. The formulation consists of unilamellar vesicles of diameter ~80 to 90 nm and they contain doxorubicin (2 mg mL^{-1}). Doxil® was first approved as a treatment for AIDS-related Kaposi's sarcoma in patients with disease resistant to other chemotherapy or

in patients who are intolerant to such therapy, but now it is also used as a treatment for metastatic ovarian carcinoma (when disease is refractory to both paclitaxel- and platinum-based chemotherapy regimens) and breast cancer. Proof of concept that liposomes could increase anticancer agent activity and reduce toxicity has led to their further development as carriers of other anticancer agents including platinates, camptothecins, mitoxantrone, and vincristine.

Use of liposomal formulations is not limited to tumor targeting. Other uses include delivery of amphotericin B (AmBisome®) to treat various fungal infections and visceral leishmaniasis, antibiotics (Mikasome®), and for targeting radiotherapy. They are also currently under intensive development as vectors for vaccine delivery, antisense delivery, and for use as gene delivery vectors.

2.2 Antibodies, Proteins, and Fusion Proteins

Antibodies and proteins have long been considered as therapeutic agents in their own right and as components of targetable drug carriers. It is only through the last decade, however, that they have really begun to realize this potential. In the context of drug targeting, conjugates can be prepared either by chemical conjugation of drug or, in the case of protein therapeutics, often the conjugate is prepared by recombinant technology. Not only are antibodies themselves used as vehicles for targeted delivery, as mentioned above, they (or antibody fragments) are often used as the site-directing components of other delivery systems including liposomes, polymer conjugates, and nanoparticles.

Early murine antibodies were beset with problems of poor specificity and immunogenicity. However, the ability to humanize, and most recently prepare chimeric antibodies is rapidly generating a successful pipeline of novel targeted pharmaceuticals (Table 5). Seventeen therapeutic antibodies have entered the market, and there is an enormous pipeline of antibody-based products (>350) coming through clinical development.

Therapeutic antibodies can be divided into several categories according to their mechanism of action.

1. They can block the action of another molecule (i.e. act as an antagonist) and examples in this class include antibodies that bind to soluble ligands such as antivascular endothelial growth factor (VEGF), and antitumor necrosis factor (TNF) antibodies.
2. They can stimulate a biological response (i.e. act as an agonist).
3. They kill target cells through immune-mediated mechanisms, for example, the antitumor antibodies Rituxan® and Herceptin®.
4. They can be used to direct radiotherapy, toxins or other drug carrier systems (liposomes, nanoparticles, and polymer conjugates) to a target tissue.

Ideally, target antigens will have 1000-fold greater expression on the target cell than seen on normal tissues and the monoclonal antibody derived against the target should have high affinity. Often in the complex *in vivo* environment, this degree of selectivity has been difficult to achieve. With the aim of minimizing the biological barriers to be encountered, the most successful early products were designed to localize to cells circulating in the bloodstream. For example, the anti-CD20 antibody Rituxan® developed as a treatment for relapsed or refractory non-Hodgkin's lymphoma (NHL). Similarly,

Tab. 5 Monoclonal antibody-based products.

Antibody	*Target*	*Payload*	*Use*
Therapeutic antibodies			
Rituxan®:	CD20	Inherent activity	CD20 +ve non-Hodgkin's lymphoma – mediates antibody-dependent cytolysis
Herceptin®	HER2	Inherent activity	HER2 +ve breast cancer – mediates antibody-dependent cytolysis
Antibody-drug conjugates			
Mylotarg®	CD33	Calicheamicin	Acute myeloid leukemia – taken into the cell by endocytosis
Radioimmunotherapeutics			
Tositumomab®	CD20	[^{131}I]iodide	Non-Hodgkin's lymphoma-targeted radiotherapy
Zevalin®	CD20	^{90}Yttrium	Non-Hodgkin's lymphoma-targeted radiotherapy
Immunotoxins			
Anti-B4-blocked ricin	CD19	Blocked ricin	Targeted immunotoxin non-Hodgkin's lymphoma
Anti-Tac(Fv)-PE38 (LMB2)	CD25	*Pseudomonas* exotoxin fusion protein	Hematological malignancies

the first immunoconjugate to enter routine clinical use was an anti-CD33 antibody conjugated to the natural product anticancer agent calicheamicin (Mylotarg®). In this case, it is used as a treatment for acute myeloid leukemia (AML). Antibodies have not only been used to carry chemotherapy, but they also offer a unique opportunity to target radiotherapy and highly potent protein toxins (e.g. ricin, diphtheria toxin, and pseudomonas exotoxin) to tumor cells. Two radioimmunoconjugates have been approved for clinical use (Zevalin® and Bexxar®) and both utilize an anti-CD20 antibody linked to a chelating agent carrying a radionuclide as treatments for refractory NHL patients.

This year the recombinant humanized monoclonal antibody Avastin™ was approved as the first anticancer agent that works by preventing the formation of new blood vessels (angiogenesis). Avastin™ acts by binding vascular endothelial growth factor (VEGF) preventing its interaction with VEGF receptors (Flt-1 and KDR) on the surface of endothelial cells. The endothelial cell is also a relatively easily accessible target to circulating antibody. In this case, the mechanism of action is inhibition of new blood vessel formation (angiogenesis) and when administered intravenously together with a combination of anticancer drugs (including ironotecan, 5-fluorouracil (5FU), and leucovorin), Avastin™ extends colon cancer patient survival time.

In the future it will be essential to identify a library of antibodies (acting by themselves as therapeutics in their own right or as components of drug delivery systems) that will effectively target those common solid tumors that are the major cause of mortality (e.g. breast, prostate, lung, and gastrointestinal cancers). These tumors are inherently more difficult to

access using macromolecular and vesicular/particulate drug carriers due to relatively poor tumor vascularization and the difficulty of deep penetration into the tumor tissue following extravasation. The recent commercialization of Herceptin®, a new treatment for breast cancer, has been an important step forward. This monoclonal antibody recognizes the HER2 receptor on breast cancer cells and acts by slowing or stopping cell growth. It is only effective in treating those 15 to 25% of breast cancer patients who are receptor +ve and is given to patients in combination with chemotherapy to treat advanced metastatic disease.

Use of proteins (other than antibodies) and peptides to promote targeted drug delivery has also been widely explored, but mostly in preclinical studies. Early studies focused on natural carrier proteins such as albumin, modified albumin, and low-density lipoprotein (LDL). Many peptides have been explored as potential tumor targeting ligands. In this case, an ^{111}indium-labeled analog of the peptide somatostin (Octreoscan®) is used as a diagnostic imaging agent. The ligand localizes to primary and metastatic neuroendocrine tumors bearing somatostatin receptors, and clinical studies are also ongoing to see whether similar analogs can also target therapeutic levels of radioactivity.

2.3 Nanoparticles

Although nanotechnology is becoming very fashionable, Speiser and colleagues had already described the first nanoparticles for drug delivery in the 1970s. Since then there have been numerous studies examining their potential as targetable drug carriers in the context of parenteral administration, in particular for targeting anticancer agents, for oral administration, and most recently for promoting transport across the blood–brain barrier. Nanoparticles have been defined as colloidal carriers of diameter 1 to 1000 nm. They are usually made from natural or synthetic polymers, but recent technology has been exploring the possibility of making nanoparticles from the active drug substance itself. The polymers used can be preformed, for example, poly(lactic acid), and manufacture achieved by emulsion evaporation, precipitation, or salting out procedures. Alternatively, they are prepared by polymerization of monomers, for example, poly(alkylcyanoacrylate) nanoparticles. In some cases, for example, albumin nanoparticles, manufacture is achieved by addition of crosslinking agents. Structures are prepared as nanoparticles composed of a polymer matrix, or nanocapsules with a polymer shell surrounding an open or filled core (Fig. 7). The latter can be achieved by interfacial polymerization. Nanoparticles administered intravenously are particularly prone to opsonization and rapid elimination by RES clearance, so surface modification by addition of polymers (e.g. poloxamers and PEG) has become popular as an attempt to overcome this problem.

The technology is so versatile that nanoparticles can be prepared to contain low molecular weight chemotherapy, peptides, and proteins, and oligonucleotides. However, despite the vast number of preclinical studies, so far it has proved relatively difficult to transfer nanoparticulate drug carriers into clinical practice. Certain types of nanoparticles have shown the ability to reverse multidrug resistance (MDR), which is a major problem in cancer chemotherapy. Nanoparticles are being explored for the selective delivery of oligonucleotides to tumor cells.

An early problem was identification of long-circulating nanoparticles made from biodegradable, biocompatible polymers that would be safely eliminated after achieving the targeting objective. Although progress has been made, the most advanced systems at present are the albumin nanoparticles used as diagnostics and targetable carriers for the anticancer agent paclitaxel. In fact, albumin nanoparticles containing paclitaxel (Abraxane™) that have recently been undergoing extensive clinical studies in breast cancer patients with metastatic disease have shown promising early results. A nanoparticle-based MRI contrast agent, Endorem®, is already on the market, and this superparamagnetic, ferumoxide nanoparticle stabilized on the surface using dextran is used to diagnose liver lesions, particularly cancer metastases in the liver.

2.4 Polymer conjugates

The term *polymer therapeutics* has been used as an umbrella term to describe rationally designed polymeric drugs, polymer–drug and polymer–protein conjugates, polymeric micelles containing covalently bound drug and polyplexes for nucleic acid delivery (Fig. 8). Successful clinical application of polymer–protein conjugates (Table 6) and promising clinical

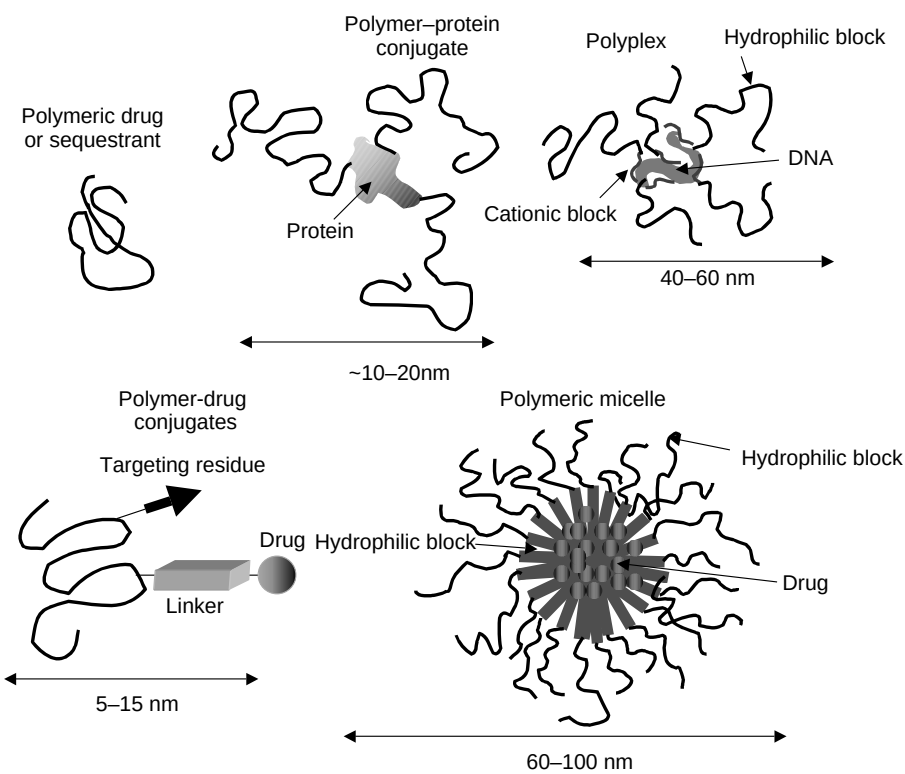

Fig. 8 A schematic diagram showing the range of polymer-based drugs and delivery systems that fall under the umbrella term *polymer therapeutics*.

Tab. 6 Polymer therapeutics in the market or transferred into clinical development.

Compound	*Name*	*Status*	*Indication*
Polymeric drugs			
Poly(alanine, lysine, glutamic acid, tyrosine)	Copaxone®	Market	Multiple sclerosis
Poly(allylamine)	Renagel®	Market	End-stage renal failure
Dextrin-2-sulfate	Emmelle® gel	Market	HIV/AIDS – a vaginal virucide formulated as a gel
Dextrin-2-sulfate		Phase III	HIV/AIDS – polymer administered intraperitoneally
Poly(I):Poly(C)	Ampligen®	Phase III	Chronic fatigue immune dysfunction (myalgic encephalomyelitis; ME)
Polyvalent, polylysine dendrimer containing SPL7013	VivaGel™	Phase I	Viral sexually transmitted diseases, formulated as a vaginal gel
Polymer–oligonucleotide conjugates			
PEG-aptamer	Macugen™	NDA filed	Age-related macular degeneration
Polymer–protein conjugates			
PEG-adenosine deaminase	Adagen®	Market	Severe combined immunodeficiency syndrome
SMANCS	Zinostatin Stimalmer®	Market	Cancer – hepatocellular carcinoma
PEG-L-asparaginase	Oncaspar®	Market	Acute lymphoblastic leukemia
PEG-α-interferon 2b	PEG-intron™	Market	Hepatitis C, also in clinical development in cancer, multiple sclerosis, HIV/AIDS
PEG-α-interferon 2a	PEGASYS®	Market	Hepatitis C
PEG-human growth hormone	Pegvisomant®	Market	Acromegaly
PEG-GCSF	Neulasta™	Market	Prevention of neutropenia associated with cancer chemotherapy
PEG-antiTNF Fab	CDP870	Phase II	Rheumatoid arthritis
Polymer–drug conjugates			
Polyglutamate-paclitaxel	CT-2103, XYOTAX™	Phase II/III	Cancer – particularly lung cancer
HPMA copolymer-doxorubicin	PK1; FCE28068	Phase II	Cancer – particularly lung and breast cancer
HPMA copolymer-doxorubicin-galactosamine	PK2; FCE28069	Phase I/II	Cancer – particularly hepatocellular carcinoma
HPMA copolymer-paclitaxel	PNU166945	Phase I	Cancer
HPMA copolymer camptothecin	MAG-CPT /PNU166148	Phase I	Cancer
HPMA copolymer platinate	AP5280	Phase II	Cancer
HPMA copolymer platinate	AP5346	Phase I	Cancer
Polyglutamate-camptothecin	CT-2106	Phase I	Cancer
PEG-camptothecin	PROTHECAN™	Phase II	Cancer
Polymeric micelles			
PEG-aspartic acid-doxorubicin micelle	NK911	Phase I	Cancer

results arising from trials with polymer–anticancer drug conjugates bodes well for future design and development of the ever more sophisticated bionanotechnologies needed to realize the full potential of the postgenomics age. Everyone uses water-soluble polymers, as an inert functional part of a multifaceted construct for improved drug, protein, or gene delivery.

Conceptually, polymer therapeutics share many features with other macromolecular drugs (proteins, antibodies, oligonucleotides) and macromolecular prodrugs including immunoconjugates. A bonus, however, is the versatility of synthetic chemistry, which allows tailoring of molecular weight, addition of biomimetic features to the man-made construct, and even the possibility to include bioresponsive elements.

2.4.1 Polymer–protein Conjugates

Recombinant DNA and monoclonal antibody technology has created a biotech revolution providing a growing number of peptide, protein, and antibody-based drugs. However, their limitations often include a short plasma half-life, poor stability and, for proteins, immunogenicity. Therefore, there has been a continuing search for improved protein-derived alternatives. In the 1970s, pioneering research of Davis, Abuchowski, and colleagues foresaw the potential of PEG conjugation to proteins. This technique is now well established and is called *PEGylation*. PEGylation is designed to increase protein solubility and stability, and reduce protein immunogenicity. Moreover, by preventing rapid renal clearance of small proteins and receptor-mediated protein uptake by cells of the RES, PEGylation can be used to prolong plasma half-life. The resultant need for less frequent dosing is of great benefit to the patient and encourages compliance.

PEG is a particularly attractive polymer for conjugation. It is widely used as a pharmaceutical excipient and the flexible highly water-soluble polymer chain extends to give a hydrodynamic radius that is some 5 to 10 times greater than that of a globular protein of equivalent molecular weight. There are three requirements for optimized synthesis of a polymer–protein conjugate: (1) a semitelechelic polymer, that is, with a single reactive group at one terminal end to avoid protein crosslinking during conjugation; (2) the ability to introduce a linker that will not itself generate toxic or immunogenic by-products and that will provide appropriate stability characteristics (dependent on the protein being bound); (3) and an approach that will provide reproducible site-specific protein modification. Linear and branched PEGs of Mw 5000 to 40 000 Da have been used to create protein conjugates, sometimes with multiple PEGs attached per protein, or alternatively a 1 : 1 ratio.

First-generation protein conjugates used linear monomethoxyPEGs (mPEG) and a variety of conjugation chemistries, but these early strategies had significant disadvantages including protein crosslinking (due to contaminating PEG-diol), modification of protein charge due to consumption of protein $-NH_2$ or $-COOH$ groups during conjugation, unstable PEG-protein linkages, and sometimes the need for reaction conditions that led to protein denaturation. More recently, improved conjugation techniques have been developed including site-specific modification following protein mutagenesis, the use of the enzyme transglutaminase to PEGylate site-specifically in the protein, and additionally, design of degradable PEG-protein

linkages to maximize return of protein bioactivity.

The clinical value of PEGylation is now well established (Table 6). PEG-L-asparaginase (ONCASPAR®) is used as treatment for acute lymphoblastic leukemia (ALL). Compared to the native enzyme, PEG-L-asparaginase brings advantages of reduced hypersensitivity, longer plasma half-life, and a slower total clearance. Consequently, it can be administered every two weeks instead of two to three times per week as required for the native enzyme. Most importantly, PEGylation of L-asparaginase decreases hypersensitivity reactions (only seen in ≤8% patients after conjugate administration). PEGylated-recombinant methionyl human granulocyte colony stimulating factor (G-CSF) is used to prevent severe cancer chemotherapy-induced neutropenia. Again, the PEG-GCSF (Neulasta™) has the benefit of less frequent administration being given by a single subcutaneous injection on day 2 of each chemotherapy cycle. The native G-CSF must be given daily for two weeks to achieve the same protection. Two PEG-interferon-α (IFNα) (2a or 2b) conjugates, PEGASYS® and PEG-INTRON®, have been approved as treatments for hepatitis C. Whereas IFNα-2a and IFNα-2b display similar biological activity and only differ in respect of a single amino acid, the molecular weight of PEG used for conjugation and the linker employed is very different in each product. PEGASYS® has a higher specific activity *in vitro* and a longer plasma half-life than PEG-INTRON®. Both conjugates have shown clinically superior antiviral activity compared to IFNα. PEG-IFNα is also under clinical evaluation for treatment of other diseases including, cancer, multiple sclerosis, and HIV/AIDS.

2.4.2 Polymer–drug Conjugates

In the 1970s, Ringsdorf first described the concept of targetable polymer–drug conjugates. These drug carriers have a tripartite structure comprising a water-soluble (hydrophilic) polymer backbone chosen to aid drug solubilization, a biodegradable polymer-drug linker designed to ensure stability in the circulation and subsequently facilitate drug release at the target site and, thirdly the inclusion of a targeting ligand to promote cell-specific delivery. A wide variety of natural, synthetic, and pseudosynthetic polymers have been explored as the platform, but so far those conjugates used clinically have only been based on *N*-(2-hydroxypropyl)methacrylamide (HPMA) copolymers, polyglutamatic acid (PGA), polyethyleneglycol (PEG) and polysaccharides derived from dextran (Table 6).

When selecting a polymeric carrier, a number of polymer features must be considered to ensure suitability for human use. The polymer should be nontoxic and nonimmunogenic. Preferably, the polymer backbone should be biodegradable, and if not (e.g. HPMA copolymers and PEG), the polymer molecular weight must be limited to <40 000 Da to ensure eventual renal elimination. To guarantee efficacy, the conjugate must have sufficient carrying capacity in relation to the potency of the anticancer drug to be carried. The first clinically tested polymer–drug conjugates all contained commonly used, and relatively potent, antitumour agents. The fundamental aim of polymer–drug conjugation is to modify pharmacokinetics at the cellular level and thus promote increased tumor targeting of drug while minimizing exposure of sensitive normal tissues. The polymer–drug linker must be stable in the circulation. At the cellular level, once internalized the conjugate is trafficked via

the endosomal compartments (acid pH 6.5–5.5) to lysosomes where it is exposed to both acid pH (~pH 5.0) and an array of lysosomal hydrolases.

The first polymer–drug conjugate to enter clinical trial did so in 1994; it was the HPMA copolymer-Gly-Phe-Leu-Gly-doxorubicin conjugate (PK1, FCE28068). It has an Mw ~30 000 Da and doxorubicin content of ~8.5 wt%. The peptidyl linker was designed to be cleared following lysosomotropic delivery by lysosomal cathepsin B. In Phase I trials, PK1 was administered as a short infusion every three weeks and the maximum tolerated dose was 320 mg m^{-2} (doxorubicin-equivalent). This is a ~fourfold increase compared to the normal safe dose of free drug and a higher dose than can be safely given in liposomal form. The dose-limiting toxicities seen were typical of the anthracyclines, including febrile neutropenia and mucositis. Despite cumulative doses up to 1680 mg m^{-2} (doxorubicin-equivalent), no cardiotoxicity, typical of anthracyclines, was observed. Antitumor activity seen in patients considered chemotherapy resistant/refractory and at lower doxorubicin doses (80–180 mg m^{-2}) was consistent with EPR-mediated targeting. Subsequently HPMA copolymer conjugates of paclitaxel, camptothecin, and platinates have also progressed into clinical trial. Despite a large number of studies exploring ligand-targeted polymer conjugates, an HPMA copolymer-Gly-Phe-Leu-Gly-doxorubicin containing additionally galactosamine (PK2, FCE28069) is still the only targeted conjugate to be tested clinically. It was designed to target the hepatocyte asialoglycoprotein receptor for treatment of liver cancer. Phase I/II trials were conducted in patients with primary hepatocellular carcinoma and the maximum tolerated dose was 160 mg m^{-2} Dox-equivalent. Although most of the conjugate localized in normal liver hepatocytes, it was estimated that the doxorubicin concentration in hepatoma tissue would still be 12 to 50-fold higher than could be achieved by administration of free drug.

Currently, a polyglutamic acid-paclitaxel conjugate (CT-2103, XYOTAX™) is showing considerable promise. This conjugate contains 37 wt% paclitaxel linked through the 2′ position of the ester bond to the γ-carboxylic acid of PGA (Mw ~40 000 Da). This is the first biodegradable polymer to be used for conjugate synthesis, and the polymer backbone is cleaved by cathepsin B to liberate diglutamyl-paclitaxel. When administered as a single agent intravenously over 30 min every 3 weeks, the maximum tolerated dose of CT-2103 was 266 mg m^{-2} paclitaxel equivalent. In these early trials, a significant number of patients have displayed partial responses or stable disease (mesothelioma, renal cell carcinoma, nonsmall cell lung cancer and a paclitaxel-resistant ovarian cancer patient) and the extensive Phase II and Phase III program now also includes combinations of CT-2103 with cisplatin and carboplatin.

2.5 Polymeric Micelles

Self-assembling block copolymer micelles provide an alternative targetable drug delivery vehicle. The most advanced example is the pluronic block copolymer micelle designed by Kataoka and colleagues incorporating doxorubicin (both covalently bound and noncovalently entrapped) that is able to circumvent P-glycoprotein-mediated resistance is currently in early clinical evaluation. The micelle is composed of PEG (Mw ~5000 Da)-poly(aspartic acid) block copolymers and has a diameter of 42 nm. Preferential accumulation in tumor tissue

by the EPR effect leads to a three- to fourfold improvement in drug targeting. In this case, the covalently bound drug is inactive, and it is the free drug that slowly escapes over 8 to 24 h that is responsible for tumor cell destruction.

2.6 Viral and Nonviral Vectors

The challenges of gene therapy are described in full elsewhere in this volume. However, as the natural (retroviral and adenoviral vectors are the favored tools for gene therapy) and synthetic nonviral vectors are being used as the new generation of targetable drug carriers, they are worthy of brief mention here. Viral vectors undeniably can circumvent all of the biological barriers to promote highly efficient gene delivery, and in certain cases show tissue tropism. However, viral vectors also have recognized disadvantages, in particular, their mixed safety record. The death of an 18-year-old patient in 1999 due to an acute inflammatory reaction after gene therapy, and the more recent observation that 2 of 10 children treated with gene therapy to correct severe combined immunodeficiency disease (SCID) developed a type of leukemia that was clearly caused by the viral vector, has led many to question the risk/benefit of this approach. While it is widely acknowledged that nonviral vectors offer potential advantages in terms of safety and ease of manufacture (and the lipidic-DNA complexes (lipoplexes) have been used in clinical trials), both lipoplexes and polymer-based delivery systems (polyplexes) are still far from ideal and there are many barriers to delivery. Their transfection efficiency is often extremely low and the most popular lipidic and polymeric vectors used are very toxic. Even with more than a decade of systematic study, so far all have failed to identify the "ideal" polymer construct that is able to promote efficient intracytosolic delivery of proteins and genes after repeated intravenous administration *in vivo*. Nonetheless, improving design of pH-responsive endosomolytic polymers gives hope that a synthetic viral mimetic will eventually become reality.

3 To the Future – The Importance of "Nanomedicines"?

The challenges and opportunities of drug targeting have long been appreciated, and now, the first applications are at last coming to fruition, principally, but not exclusively, in the context of tumor targeting. This has spurred on the search for other targeting systems to localize drugs to sites of inflammation, to infectious diseases, to the brain, and to atherosclerotic plaques. There is increasing anticipation that nanotechnology, as applied to medicine, will bring still further advances in the diagnosis and treatment of disease. Within the confines of the nanoscale size range (1–1000 nm), the discipline of "nanomedicine" can be best defined as the science and technology of diagnosing, treating, and preventing disease and traumatic injury, of relieving pain, and of preserving and improving human health, using molecular tools and molecular knowledge of human body. It has five distinct, but overlapping subthemes. They include (1) analytical techniques and diagnostic tools; (2) nanoimaging and manipulations; (3) nanomaterials and nanodevices; (4) the *nanomedicines* designed either as biologically active therapeutics or drug delivery systems; and (5) all issues

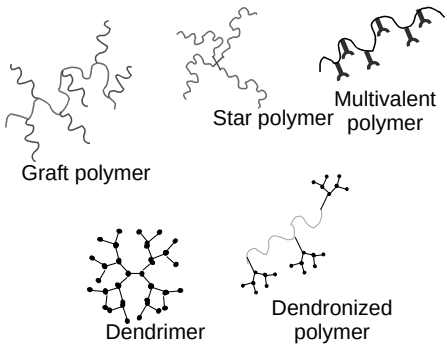

Fig. 9 The versatility of polymer chemistry can produce a range of polymer architectures now being used to develop second-generation drug targeting vectors.

relating to their pharmaceutical development; and clinical use with particular regard to potential toxicity.

The search for safe and effective targeted therapy continues in the postgenomics era, and as we understand more of the mechanisms of the disease progression and aging, ever more sophisticated diagnostics and therapeutics will undoubtedly emerge. In the long term, it is predicted that these activities will be combined, and the word ''theragnostic'' has been used to describe a system that is able to diagnose a problem and then deliver the therapy. Device miniaturization, with parallel efforts to bring together Mother Nature's molecules (as therapeutics and targeting ligands) and synthetic chemistry gives many exciting possibilities for designing novel therapeutics and targeting systems.

In the short term, research efforts are focusing on three main areas: (1) identification of new targets; (2) two-step approaches and combination therapy to increase specificity and efficacy; and (3) new vectors for improved targeting. Water-soluble synthetic polymers and polymer-based materials are playing an increasing role in drug delivery systems and biomaterial design. The search is on for new, safe, biodegradable polymers that are better suited to repeated parenteral administration, and also toward more elaborate 3D, bioresponsive architecture that bring the capacity to mimic the tertiary structure of proteins. A variety of architectures can now be synthesized and these are being used to create nanosized molecules with multifunctionality (Fig. 9).

Dendrimers and dendronized polymers are particularly interesting as they provide the multivalency needed for immobilization of drugs, imaging agents, peptidyl epitopes, and so on; they seem to offer unique opportunities for exploiting intracellular and transcellular trafficking pathways. Engineering shape memory into polymers has recently been described for polymer materials, and transfer of these concepts into water-soluble polymer architecture, perhaps using molecular imprinting techniques, remains an exciting challenge for the future.

See also Biotransformations of Drugs and Chemicals; Medicinal Chemistry; Oligonucleotides; Pharmacokinetics and Pharmacodynamics of Biotech Drugs; Receptor Targets in Drug Discovery; Targeted Therapy: Genomic Approaches; Therapeutic Compounds

in Nature as Leads for New Pharmaceuticals; Vector System: Plasmid DNA; Vector Targeting in Gene Therapy.

Bibliography

Books and Reviews

Advanced Drug Delivery Reviews: http://www.authors.elsevier.com/JournalDetail.html?PubID=505508&Precis=DESC

Allen, T.M. (2002) Ligand-targeted therapeutics in anticancer therapy, *Nat. Rev. Drug Discov.* **2**, 750–763.

Brekke, O.H., Sandlie, I. (2003) Therapeutic antibodies for human diseases at the dawn of the twenty-first century, *Nat. Rev. Drug Discov.* **2**, 52–62.

Duncan, R. (2003a) The dawning Era of polymer therapeutics, *Nat. Rev. Drug Discov.* **2**, 347–360.

Harris, J.M., Chess, R.B. (2003) Effect of pegylation on pharmaceuticals, *Nat. Rev. Drug Discov.* **2**, 214–221.

Jain, R.K. (1998) The next frontier of molecular medicine: delivery of therapeutics, *Nat. Med.* **4**, 655–657.

Journal of Controlled Release: http://www.authors.elsevier.com/JournalDetail.html?PubID=502690&Precis=DESC

Journal of Drug Targeting: http://www.tandf.co.uk/journals/titles/1061186x.asp

Langer, R. (2001) Perspectives: drug delivery – drugs on target, *Science* **293**, 58–509.

Molema, G., Meijer, D.K.F., Mannhold, R., Kubinyi, H., Timmerman, H. (Eds.) (2001) *Drug Targeting: Organ-specific Strategies, In Series Methods and Principles in Medicinal Chemistry*, Vol 12, Wiley-VCH Verlag GMbH, Weinheim, pp 1–381.

Muzykantov, V., Torchilin, V. (Eds.) (2002) *Biomedical Aspects of Drug Targeting*, Kluwer Academic Publishers, Dordrecht, pp. 1–506.

Primary Literature

Asokan, A., Cho, M.J. (2002) Exploitation of intracellular pH gradients in cellular delivery of macromolecules, *J. Pharm. Sci.* **91**, 903–912.

Atkins, J.H., Gershell, L.J. (2002) Selective anticancer drugs, *Nat. Rev. Cancer* **1**, 645–646.

Barik, S., Bitko, V. (2005) RNA Clinical Applications, Emerging, *reference to article in this Encyclopedia*.

Borm, P.J.A. (2002) Particle toxicology: from coal mining to nanotechnology, *Inhal. Toxicol.* **14**, 311–324.

Boussif, O., Lezoualch, F., Zanta, M.A., Mergny, M.D., Scherman, D., Demeneix, B., Behr, J.-P. (1995) A versatile vector for gene and oligonucleotide transfer into cells in culture and *in vivo*: polyethylenimine, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7297–7301.

Brigger, I., Dubernet, C., Couvreur, P. (2002) Nanoparticles in cancer therapy and diagnosis, *Adv. Drug Deliv. Rev.* **54**, 631–651.

Brocchini, S., Duncan, R. (1999) Pendent Drugs, Release from Polymers, in: Mathiowitz, E. (Ed.) *Encyclopedia of Controlled Drug Delivery*, John Wiley & Sons, New York, pp. 786–816.

Burton, E.A., Fink, D.J., Glorioso, J.C. (2005) Gene therapy, *reference to article in this Encyclopedia*.

Cavagnaro, J.A. (2002) Preclinical safety evaluation of biotechnology-derived pharmaceuticals, *Nat. Rev. Drug Discov.* **1**, 469–476.

Chanda, S.K., Caldwell, J.S. (2003) Fulfilling the promise: drug discovery in the postgenomics era, *Drug Discov. Today* **8**, 168–174.

Check, E. (2003) Harmful potential of viral vectors fuels doubts over gene therapy, *Nature* **423**, 573–574.

Couvreur, P., Kante, B., Roland, M., Speiser, P. (1979) Adsorption of antineoplastic drugs to polyalkylcyanoacrylate nanoparticles and their release in calf serum, *J. Pharm. Sci.* **68**, 1521–1524.

Couvreur, P., Grislain, L., Lenaert, V., Brasseur, F., Guiot, P., Biernacki, A. (1986) Biodegradable Polymeric Nanoparticles as Drug Carrier for Antitumor Agents, in: Guiot, P. Couvreur, P. (Eds.) *Polymeric Nanoparticles and Microspheres*, CRC Press, Boca Raton, Fla, pp. 27–93.

Crommelin, D.J., Storm, G.J. (2003) Liposomes: from the bench to the bed, *Liposome Res.* **13**, 33–36.

Damascelli, B., Cantu, G., Mattavelli, F., Tamplenizza, P., Bidoli, P., Leo, E., Dosio, F., Cerrotta, A.M., Di Tolla, G., Frigerio, L.F., Garbagnati, F., Lanocita, R., Marchiano, A., Patelli, G., Spreafico, C., Ticha, V., Vespro, V., Zunino, F. (2001) Intra-arterial chemotherapy with polyoxyethylated castor oil free paclitaxel,

incorporated in albumin nanoparticles (ABI-007) – phase i study of patients with squamous cell carcinoma of the head and neck and anal canal: preliminary evidence of clinic, al activity, *Cancer* **92**(10), 2592–2602.

Davis, F.F. (2002) The origin of pegnology, *Adv. Drug Deliv. Rev.* **54**, 457–458.

Duncan, R. (2003b) Polymer-anticancer Drug Conjugates, in: Budman, D., Calvert, H., Rowinsky, E. (Eds.) *Handbook of Anticancer Drug Development*, Lippincott Williams & Wilkins, Baltimore, Md, pp. 239–260.

Duncan, R. (2005) *N*-(2-Hydroxypropyl)methacrylamide Copolymer Conjugates, in: Kwon, G.S. (Ed.) *Polymeric Drug Delivery Systems*, Marcel Dekker, New York, pp. 1–92.

Duncan, R. (2004b) Nanomedicines in action, *Pharm. J.* **273**, 485–488.

Edelstein, M.L., Abedi, M.R., Wixon, J., Edelstein, R.M. (2004) Gene therapy clinical trials worldwide 1989–2004 – an overview, *J. Gene Med.* **6**, 597–602.

Editorial (2003) Nanomedicine: grounds for optimism, *Lancet* **362**, 673.

Ferruti, P., Marchisio, M.A., Duncan, R. (2002) Poly(amido-amine)s: biomedical applications, *Macromol. Rapid Comm.* **23**, 332–355.

Florence, A.T. (2004) Issues in oral nanoparticle drug carrier uptake and targeting, *J. Drug Target.* **12**(2), 65–70.

Funhoff, A.M., van Nostrum, C.F., Koning, G.A., Schuurmans-Nieuwenbroek, N.M.E., Crommelin, D.J.A., Hennink, W.E. (2004) Endosomal escape of polymeric gene delivery complexes is not always enhanced by polymers buffering at low pH, *Biomacromolecules* **5**, 32–39.

Frechet, J.M.J., Tomalia, D.A. (Eds.) (2001) *Dendrimers and Other Dendritic Polymers*, John Wiley & Sons, New York.

Gabizon, A., Shmeeda, H., Barenholz, Y. (2003) Pharmacokinetics of PEgylated liposomal doxorubicin: review of animal and human studies, *Clin. Pharmacokinet.* **42**, 419–436.

Goldberg, M., Gomez-Orellana, I. (2003) Challenges for the oral delivery of macromolecules, *Nat. Rev. Drug Discov.* **2**, 289.

Greenwald, R.B., Choe, Y.H., McGuire, J., Conover, C.D. (2003) Effective drug delivery by PEGylated drug conjugates, *Adv. Drug Deliv. Rev.* **55**, 217–250.

Gref, R., Minamitake, Y., Peracchia, M.T., Trubetskoy, V., Torchilin, V., Langer, R. (1994) Biodegradable long-circulating polymeric nanospheres, *Science* **263**, 1600–1603.

Gregoriadis, G., Ryman, B.E. (1972) Fate of protein-containing liposomes injected into rats. An approach to the treatment of storage diseases, *Eur. J. Biochem.* **24**, 485.

Greish, K., Fang, J., Inutsuka, T., Nagamitsu, A., Maeda, H. (2003) Macromolecular therapeutics: advantages and prospects with special emphasis on solid tumour targeting, *Clin. Pharmacokinet.* **42**, 1089–1105.

Gruenberg, J. (2001) The endocytic pathway: a mosaic of domains, *Nat. Rev. Mol. Cell Biol.* **2**, 721–730.

Haas, M., Kato, Y., Haverdings, F.G., Moolemaar, F., Suzuki, K., de Zeeuw, D., Sugiyama, Y., Meijer, D.K.F. (2001) Delivery of Drugs and Antisense Oligonucleotides to Proximal Tubular Cells of the Kidney Using Macromolecular and Pro-drug Approaches, in: Molema, G., Meijer, D.K.F., Mannhold, R., Kubinyi, H., Timmerman, H. (Eds.) *Drug Targeting: Organ-specific Strategies, Series Methods and Principles in Medicinal Chemistry,* Vol 12, Wiley-VCH Verlag GMbH, Weinheim, pp. 121–155.

Hacein-Bey-Abina, S., von Kalle, C., Schmidt, M., Le Deist, F., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J.-L., Fraser, C.C., Cavazzana-Calvo, M., Fischer, A. (2003) A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency, *N. Engl. J. Med.* **348**, 255–256.

Haensler, J., Szoka, F.C. (1993) Polyamidoamine cascade polymers mediate efficient transfection of cells in culture, *Bioconjug. Chem.* **4**, 372–379.

Hafez, I.M., Cullis, P.R. (2004) Tunable pH-sensitive liposomes, *Meth. Enzymol.* **387**, 113–134.

Hamm-Alvarez, S.F., Okamoto, C.T. (Eds.) (2003) Endocytosis and sorting of macromolecules, *Adv. Drug Deliv. Rev.* **55**(11), 1351–1524.

Harrington, K.J., Mohammadtaghi, S., Uster, P.S., Glass, D., Peters, A.M., Vile, R.G., Stewart, J.S.W. (2001) Effective targeting of solid tumors in patients with locally advanced encapsulated in liposomes by radiolabeled pegylated liposomes, *Clin. Cancer Res.* **7**, 243–254.

Hawiger, J. (1999) Noninvasive intracellular delivery of functional peptides and proteins, *Curr. Opin. Chem. Biol.* **3**, 89–94.

Huang, P.S., Oliff, A. (2001) Drug-targeting strategies in cancer therapy, *Curr. Opin. Genet. Dev.* **11**, 104–110.

Iwai, K., Maeda, H., Konno, T. (1984) Use of oily contrast medium for selective drug targeting to tumour: enhanced therapeutic effect and X-ray image, *Cancer Res.* **44**, 2114–2121.

Jain, R.K. (2001) Delivery of molecular and cellular medicine to solid tumours, *Adv. Drug Deliv. Rev.* **46**, 149–168.

Kabanov, A.V., Felgner, P.L., Seymour, L.W. (Eds.) (1998) *Self-assembling Complexes for Gene Delivery. From Laboratory to Clinical Trial*, John Wiley & Sons Ltd, Chichester.

Kreuter, J., Speiser, P. (1976) In vitro studies of poly(methyl methacrylate) adjuvants, *J. Pharm. Sci.* **65**, 1624–1627.

Kreuter, J., Shamenkov, D., Petrov, V., Ramge, P., Cychutek, K., Koch-Brandt, C., Alyautdin, R. (2002) Apolipoprotein-mediated transport of nanoparticle-bound drugs across the blood-brain barrier, *J. Drug Target.* **10**(4), 317–325.

Kueltzo, L.A., Middaugh, C.R. (2003) Non classical transport proteins and peptides: an alternative to classical macromolecule delivery systems, *J. Pharm. Sci.* **92**, 1754–1772.

Kukowska-Latallo, J.F., Bielinska, A.U., Johnson, J., Spindler, R., Tomalia, D.A., Baker, J.R.J. (1996) Efficient transfer of genetic material into mammalian cells using starburst polyamidoamine dendrimers, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 4897–4902.

Kyriakides, T.R., Cheung, C.Y., Murthy, N., Bornstein, P., Stayton, P.S., Hoffman, A.S. (2002) pH-Sensitive polymers that enhance intracellular drug delivery in vivo, *J. Cont. Rel.* **78**, 295–303.

Langer, R. (1998) Drug delivery and targeting, *Nature* **392**, 5–10.

LaVan, D., Lynn, D., Langer, R. (2002) Moving smaller in drug delivery and discovery, *Nat. Rev. Drug Discov.* **1**, 77–84.

Lehr, C.-M. (Ed.) (2002) *Cell Culture Models of Biological Barriers*, Taylor & Francis, London, pp. 1–430.

Lu, Y., Low, P.S. (2002) Folate-mediated delivery of macromolecular anticancer therapeutic agents, *Adv. Drug Deliv. Rev.* **54**, 675–693.

Maeda, H., Konno, T. (1997) Metamorphosis of Neocarzinostatin to SMANCS: Chemistry, Biology, Pharmacology, and Clinical Effect of the First Prototype Anticancer Polymer Therapeutic, in: Maeda, H., Edo, K., Ishida, N. (Eds.) *Neocarzinostatin: The Past, Present, and Future of an Anticancer Drug*, Springer Verlag, Berlin, pp. 227–267.

Martin, C.R., Kohli, P. (2003) The emerging field of nanotube biotechnology, *Nat. Rev. Drug Discov.* **2**, 29–37.

Mathiowitz, E. (Ed.) (1999) *Encyclopedia of Controlled Drug Delivery*, Vols 1 and 2, John Wiley & Sons, New York.

Matsumura, Y., Maeda, H. (1986) A new concept for macromolecular therapies in cancer chemotherapy: mechanism of tumouritropic accumulation of proteins and the antitumour agent SMANCS, *Cancer Res.* **6**, 6387–6392.

Maxfield, F.R., McGraw, T.E. (2004) Endocytic recycling, *Nat. Rev. Mol. Cell Biol.* **5**, 121–132.

Milenic, D.E., Brady, E.D., Brechbiel, M.W. (2004) Antibody-targeted radiation cancer therapy, *Nat. Rev. Drug Disc.* **3**, 488–498.

Nagle, T., Berg, C., Nassr, R., Pang, K. (2003) The further evolution of biotech, *Nat. Rev. Drug Discov.* **2**, 75–79.

Nakanishi, T., Fukushima, S., Okamoto, K., Suzuki, M., Matsumura, Y., Yokoyama, M., Okano, T., Sakurai, Y., Kataoka, K. (2001) Development of the polymer micelle carrier system for doxorubicin, *J. Cont. Rel.* **74**, 295–302.

Novina, C.D., Sharp, P.A. (2004) The RNAi revolution, *Nature* **430**, 161–164.

Pap, E.H.W., Dansena, T.B., Wirtza, K.W.A. (2001) Peptide-based targeting of fluorophores to peroxisomes in living cells, *Trends Cell Biol.* **11**, 10–12.

Parkes, R.J., Hart, S.L. (2000) Adhesion molecules and gene transfer, *Adv. Drug Deliv. Rev.* **44**, 135–152.

Pouten, C.W. (1998) Nuclear import of polypeptides, polynucleotides and supramolecular complexes, *Adv. Drug Deliv. Rev.* **34**, 51–64.

Prausnitz, M., Mitragotri, S., Langer, R. (2004) Current status and future potential of transdermal drug delivery, *Nat. Rev. Drug Discov.* **3**, 115–124.

Rabinow, B.E. (2004) Nanosuspensions in drug delivery, *Nat. Rev. Drug Disc.* **3**, 785–796.

Raper, S.E., Chirmule, N., Lee, F.S., Wivel, N.A., Bagg, A., Gao, G.P., Wilson, J.M., Batshaw, M.L. (2003) Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer, *Mol. Genet. Metab.* **80**, 148–158.

Reichert, J., Pavlou, A. (2004) Monoclonal antibodies market, *Nat. Rev. Drug Discov.* **3**, 383–385.

Reubi, J.C. (2003) Peptide receptors as molecular targets for cancer diagnosis and therapy, *Endocr. Rev.* **24**, 389–427.

Ringsdorf, H. (1975) Structure and properties of pharmacologically active polymers, *J. Polym. Sci. Polym. Symp.* **51**, 135–153.

Sapra, P., Allen, T.M. (2003) Ligand-targeted liposomal anticancer drugs, *Prog. Lipid Res.* **42**, 439–462.

Seymour, L.W., Ulbrich, K., Styger, P.S., Brereton, M., Subr, V., Strohalm, J., Duncan, R. (1994) Tumouritropism and anticancer efficacy of polymer-based doxorubicin prodrugs in the treatment of subcutaneous murine B16F10 melanoma, *Br. J. Cancer* **70**, 636–641.

Seymour, L.W., Ferry, D.R., Anderson, D., Hesslewood, S., Julyan, P.J., Payner, R., Doran, J., Young, A.M., Burtles, S., Kerr, D.J. (2002) Hepatic drug targeting: phase i evaluation of polymer bound doxorubicin, *J. Clin. Oncol.* **20**, 1668–1676.

Singer, J.W., Baker, B., de Vries, P., Kumar, A., Shaffer, S., Vawter, E., Bolton, M., Garzone, P. (2003) Poly-(L)-glutamic acid-paclitaxel (CT-2103) [XYOTAX], a biodegradable polymeric drug conjugate: characterization, preclinical pharmacology, and preliminary clinical data, *Adv. Exp. Med. Biol.* **519**, 81–99.

Triggle, D.J. (2004) Medicinal chemistry, *reference to article in this Encyclopedia*. Ulbrich K, Subr V 2004 polymeric anticancer drugs with pH-controlled activation, *Adv. Drug Deliv. Rev.* **23**, 1023–1050.

Vasey, P.A., Kaye, S.B., Morrison, R., Twelves, C., Wilson, P., Duncan, R., Thomson, A.H., Murray, L.S., Hilditch, T.E., Murray, T., Burtles, S., Fraier, D., Frigerio, E., Cassidy, J. (1999) Phase i clinical and pharmacokinetic study of PKI (HPMA copolymer doxorubicin) first member of a new class of chemotherapeutics agents: drug-polymer conjugates, *Clin. Cancer Res.* **5**, 83–94.

Veronese, F.M., Harris, J.M. (Eds.) (2002) Peptide and protein PEGylation, *Adv. Drug Deliv. Rev.* **54**(4).

Wadia, Y.S., Stan, R.V., Dowdy, S.F. (2004) Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis, *Nat. Med.* **10**, 1–6.

Wagner, E. (1999) Application of membrane-active peptides for nonviral gene delivery, *Adv. Drug Deliv. Rev.* **38**, 279–289.

Wagner, E. (2004) Strategies to improve DNA polyplexes for in vivo gene transfer: will artificial viruses be the answer?, *Pharm. Res.* **21**, 8–14.

Weisz, O.A. (2003) Organelle acidification and disease, *Traffic* **4**, 57–64.

Weissig, V. (2003) Mitochondrial-targeted drug and DNA delivery, *Crit. Rev. Ther. Drug Carrier Syst.* **20**, 1–62.

Weissig, V., De Souza, G., Torchilin, V.P. (2002) Targeting Mitochondria, in: Muzykantov, V. Torchilin, V. (Eds.) *Biomedical Aspects of Drug Targeting*, Kluwer Academic Publishers, Dordrecht, pp. 473–495.

Widera, A., Norouziyan, F., Shen, W.-C. (2003) Mechanisms of TfR-mediated transcytosis and sorting in epithelial cells and applications towards drug delivery, *Adv. Drug Deliv. Rev.* **55**, 1439–1466.

Wiethoff, C.M., Middaugh, C.R. (2003) Barriers to nonviral gene delivery, *J. Pharm. Sci.* **92**, 203–217.

Wu, G.Y., Wu, C.H. (1998) Receptor mediated delivery of foreign genes to hepatocytes, *Adv. Drug Deliv. Rev.* **29**, 243–248.

Yokoyama, M., Miyauchi, M., Yamada, N., Okano, T., Sakurai, Y., Kataoka, K., Inoue, S. (1990) Polymer micelles as novel drug carrier: adriamycin-conjugated poly(ethylene glycol)-poly(aspartic acid) block copolymer, *J. Cont. Rel.* **11**, 269–278.

Taxol and Taxanes, Production by Cell Culture: *see* Antitumor Agents: Taxol and Taxanes – Production by Yew Cell Culture

Telomerase

Maria Antonietta Cerone[1] *and Silvia Bacchetti*[2]
[1] *McGill University, Montreal, Canada*
[2] *Regina Elena Cancer Institute, Rome, Italy*

1 Telomerase and Telomere Maintenance 207

2 Telomerase Biochemistry 207
2.1 Telomerase: A Novel RNA-dependent DNA Polymerase 207
2.2 Functional Sites and Catalytic Activity 208
2.3 Telomerase Components 210
2.3.1 Telomerase RNA Component 210
2.3.2 Telomerase Protein Component 213
2.3.3 Telomerase-associated Proteins 215

3 Telomerase Assays 217
3.1 Biochemical Assays 217
3.1.1 The Standard Assay 217
3.1.2 TRAP: Telomere Repeat Amplification Protocol 217
3.2 Expression Assays 219

4 Telomeres, Telomerase, and Cell Life Span 219

5 Telomerase and Human Disease 219
5.1 Cancer 219
5.2 Age-related Diseases 222
5.3 Dyskeratosis Congenita 223

Bibliography 223
Books and Reviews 223
Primary Literature 223

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

Keywords

dNTP
Deoxyribonucleoside triphosphate.

rNTP
Ribonucleoside triphosphate.

RT–PCR
Reverse transcription–polymerase chain reaction.

Telomerase
The ribonucleoprotein enzyme that synthesizes telomeric DNA.

Telomeres
Essential genetic elements at the termini of linear chromosomes.

Telomerase Primer
An oligonucleotide, typically a telomeric DNA sequence, capable of being extended by telomerase.

TERT
Telomerase catalytic protein.

TR
Telomerase RNA subunit.

Immortal cells transmit essentially a complete genome from one division to the next. To ensure the full replication of their DNA, most immortal eukaryotic cells possess a unique enzyme, telomerase, which apparently evolved with the ends of linear chromosomes, specifically for the *de novo* synthesis of telomeric DNA. Telomere maintenance is the only known biochemical function of telomerase, and in the absence of telomerase, telomeric DNA is gradually lost with each round of cell division. Normal human somatic cells are considered mortal since they have a finite capacity for division. The observation that normal cells lack telomerase, while most tumor cell populations express telomerase, suggests an important biological function for this enzyme in controlling a key difference between mortal and immortal cells.

1
Telomerase and Telomere Maintenance

Telomeres are DNA-protein structures that cap the ends of eukaryotic chromosomes and have a distinctive chromatin organization. Telomeric DNA is noncoding and consists of tandem reiterations of simple G-rich repeats oriented 5′ to 3′ toward the chromosome termini and overhanging the C-rich strand (Fig. 1). The second component of telomeres are proteins that bind to telomeric DNA in a sequence-specific manner or through protein–protein interactions. In some species, the DNA terminus is folded back into a loop structure (t-loop) stabilized by telomeric proteins. Both DNA and protein components are required for telomere function. Telomeres are essential for chromosome stability and function because they prevent nucleolytic degradation and fusion of chromosome ends, contribute to chromosome movement and localization within the nucleus through their attachment to the nuclear matrix, and can affect gene expression through position effects. Moreover, telomeric DNA contributes to the complete replication of the chromosome. Since eukaryotic DNA polymerases catalyze synthesis only in the 5′ to 3′ direction and utilize an RNA primer for initiation, linear DNA molecules cannot be replicated in their entirety (Fig. 2). Telomeric DNA shoulders the loss of terminal sequences masked by the 5′-most primer, thus acting as a buffer against loss of internal and essential genes. In addition, being subject to elongation by telomerase, telomeric DNA provides a means for compensating for the loss of sequences stemming from the end replication problem.

Telomere length is extremely variable among species, possibly as a result of differences in minimal functional sizes that may reflect the number of telomeric proteins or protein-binding sites. In addition, there is marked heterogeneity in telomere length among chromosomes of individual cells, probably due to unequal loss and/or synthesis of telomeric DNA.

2
Telomerase Biochemistry

2.1
Telomerase: A Novel RNA-dependent DNA Polymerase

Telomerase activity was first discovered in the ciliate *Tetrahymena thermophila* on the basis of its novel property of extending a telomere-like oligonucleotide with *de novo* synthesized telomeric DNA. The enzyme is a ribonucleoprotein comprising a structural RNA molecule and a catalytic

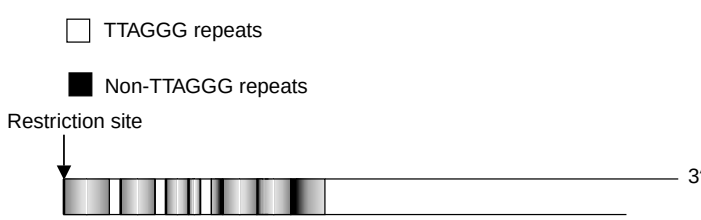

Fig. 1 Structure of terminal restriction fragments: digestion of genomic DNA with restriction enzymes releases terminal restriction fragments (TRFs) comprising a telomeric region of perfect TTAGGG repeats, and a subtelomeric region containing degenerate TTAGGG repeats, possibly interspersed with perfect repeats and unrelated simple sequences.

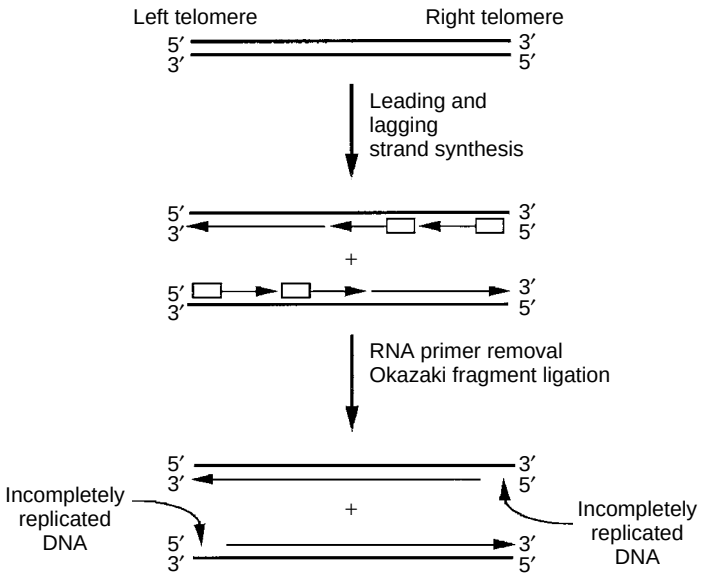

Fig. 2 The end replication problem: DNA polymerases synthesize DNA in the 5′ to 3′ direction and require an RNA primer (open boxes) for initiation of synthesis. Synthesis of the lagging strand is discontinuous and generates Okazaki fragments that are ligated to one another following removal of primers and new DNA synthesis. Removal of the 5′-most primer results in an irreparable gap at the 5′ end of the molecule. Thus, the daughter strand is shorter than the parental strand at this point. The extent of sequence loss reflects the size and position of the 5′-most primer; since the latter may vary, daughter molecules of different length may be generated.

protein. In yeast and human, the enzyme is a dimer of the two core components. A short region of the RNA subunit, called the *template domain*, is complementary to the species-specific telomere sequence and directs the synthesis of the proper telomeric DNA (Fig. 3). Thus, telomerase functionally is a reverse transcriptase: an RNA-dependent DNA polymerase. However, unlike viral reverse transcriptases that can fully copy varied "exogenous" RNAs, telomerase utilizes only a small template domain of its specific integral RNA. Most telomerases are processive and capable of reiteratively copying the template region without dissociating from the extended DNA chain. Thus, after binding to a DNA primer (typically a telomere sequence), telomerase cycles between an elongation phase in which deoxyribonucleotides (dNTPs) are sequentially added to the 3′ end of the DNA, and a "translocation" event, in which telomerase advances one repeat relative to the telomere (Fig. 3). The reiterative copying of the template domain accounts for the simple repeated structure of telomeric DNA at chromosome ends of most eukaryotic species.

2.2 Functional Sites and Catalytic Activity

The template region of the integral RNA component of telomerase typically

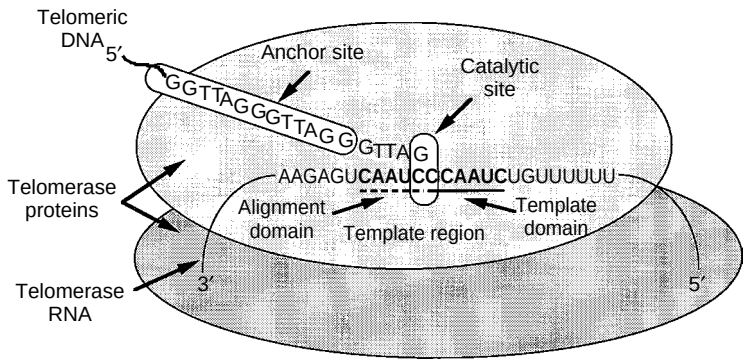

Fig. 3 The template region and primer-binding site of human telomerase: the presumed primer template configuration at the catalytic site of human telomerase is schematically illustrated, along with functional sites within the protein components (shaded ovals) and the RNA template region. This picture captures the relative position of telomeric DNA and the template region just after a translocation event: the 3′ end of the DNA is positioned opposite the alignment domain of the RNA template region, with the catalytic site ready to add dNTPs according to the sequence dictated by the template domain.

contains more than one repeat of the complementary telomeric sequence. The additional bases at the 3′ side of the template domain are utilized in part as an "alignment domain" to facilitate the binding of DNA to telomerase. This binding and alignment is especially important following the translocation event when the 3′ end of the DNA slips back such that it is entirely outside the template domain (Fig. 4). However, RNA:DNA hybridization is not sufficient to account for binding of the DNA primer to telomerase. An "anchor site" involving protein–DNA interactions also functions in initial primer binding as well as in maintaining the correct position of the DNA during the elongation/translocation events.

The anchor site of telomerase is somewhat promiscuous in that it is capable of utilizing various DNA sequences as primers for telomere extension. Telomerase can also utilize an RNA primer under certain circumstances and even incorporate ribonucleoside triphosphates (rNTPs) when they are present at high concentrations in the absence of the correct deoxyribonucleoside triphosphates (dNTPs). However, only dNTPs complementary to the template bases are efficiently added to the growing DNA primer. Nontelomeric primers that function in a telomerase assay are generally relatively G-rich single-stranded DNAs and typically have a 3′ end with some similarity to the natural telomere sequence. This suggests that the anchor site has some affinity for exposed guanine bases and that hybridization of the 3′ end of the DNA primer to the alignment region is important in forming a proper primer:template configuration. With the 3′-OH of the primer positioned in the catalytic site of the enzyme, formation of the phosphodiester bond catalyzed by all nucleic acid polymerases can occur between the terminal 3′-OH and the α-phosphate of the incoming dNTP, which is selected on the basis of its complementarity to the template ribonucleotide in the catalytic site at that time. No additional high-energy bond cleavage other than that of the dNTPs appears to be necessary for

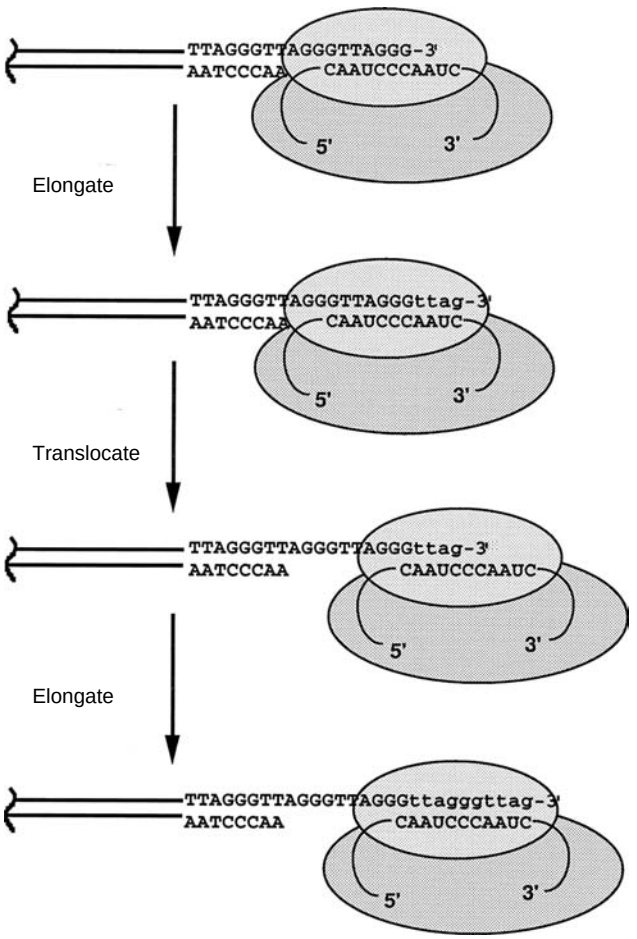

Fig. 4 Telomerase mechanism of action: the spatial relationships between telomerase protein subunits (shaded ovals), telomerase RNA (curved strand showing the human template region, CAAUCCCAAUC), and the end of a human chromosome during the cyclical process of elongation and translocation.

elongation and translocation. It is possible that translocation is facilitated by two additional properties of telomerase: a helicase-like activity that maintains a small number of base pairs between the growing DNA chain and the template, and a potential strain, which builds during the elongation steps when the relative position of the RNA and the catalytic site of the proteins must gradually change.

2.3 Telomerase Components

2.3.1 Telomerase RNA Component

The RNA subunit of telomerase (TR) has been characterized in numerous organisms including humans (Table 1). All TRs contain a region comprising the alignment and template domains whose functional role has been demonstrated by

Tab. 1 Some known telomerase RNAs.

Species[a]	*Telomere repeat*	*RNA size (nucleotide)*	*Template region (3′ to 5′)*[b]
Human	GGTTAG	~450	CAAUCCCAAUC
Mouse	GGTTAG	~450	AUCCCAAUC
T. thermophila	GGGTTG	159	AACCCCAAC
O. nova	GGGTTTTG	190	CCAAAACCCCAAAAC
E. crassus	GGGTTTTG	189	CCAAAACCCCAAAAC
S. cerevisiae	$(TG)_{1-3}$ TG_{2-3}	~1300	CACCACACCCACACAC

[a]Complete cDNAs for the RNA components have been published for human, mouse, *Tetrahymena thermophila*, *Oxytricha nova*, *Euplotes crassus*, *Saccharomyces cerevisiae*, and a number of other ciliates and yeast species.
[b]The putative template domain of the template region is underlined. The degenerate telomere sequence of *S. cerevisiae* can be explained by variable alignment of the telomeric DNA 3′ end on the template region and/or variable sites of termination during elongation.

site-directed mutagenesis. However, TRs from different species are highly divergent both in size (150–1300 nucleotides) and primary sequence. In addition, there are significant differences in their biogenesis: ciliate TRs have a polyU, are transcribed by DNA polymerase III, and matured through pathways specific for polIII transcripts, whereas vertebrate and yeast TRs are transcribed by DNA polymerase II and have a TMG cap. Mammalian TRs belong to the family of small Cajal bodies RNA (scaRNAs) and may be matured and/or assembled into these structures.

On the other hand, eukaryote TRs show a high degree of similarity in the secondary structure with a core that includes the template region and a pseudoknot motif (Fig. 5). The latter is important for the interaction of TR with the catalytic protein TERT, for the assembly of an active telomerase *in vivo* and for enzyme processivity and dimerization. An additional structural element upstream of the template in all TRs is the 5′ template boundary, which is important to prevent telomerase from copying beyond the template sequences. The mechanism that defines the template boundary varies among different species being based on a specific sequence in ciliates, a helix (helix 1) in yeast, or a helix (P1b) or sequence in vertebrates. The boundary element in most TRs represents also a site of high affinity interaction with the telomerase catalytic subunit.

Outside the template/pseudoknot core element, the secondary structure of TRs is less conserved. In ciliate TRs, a stem-loop element, called helix IV, binds TERT, is important for telomerase processivity, and for the formation of the pseudoknot. Vertebrate TRs contain a highly conserved region, CR4/CR5, which is crucial for telomerase function. Within this region are the P6.1 stem, which has been implicated in hTR binding to hTERT, and the P6.1 loop, which is important for catalytic activity. The main distinctive characteristic of vertebrate TRs is the catalytic subunit binding site, which includes two elements, the pseudoknot and the P6.1 stem, which are far away

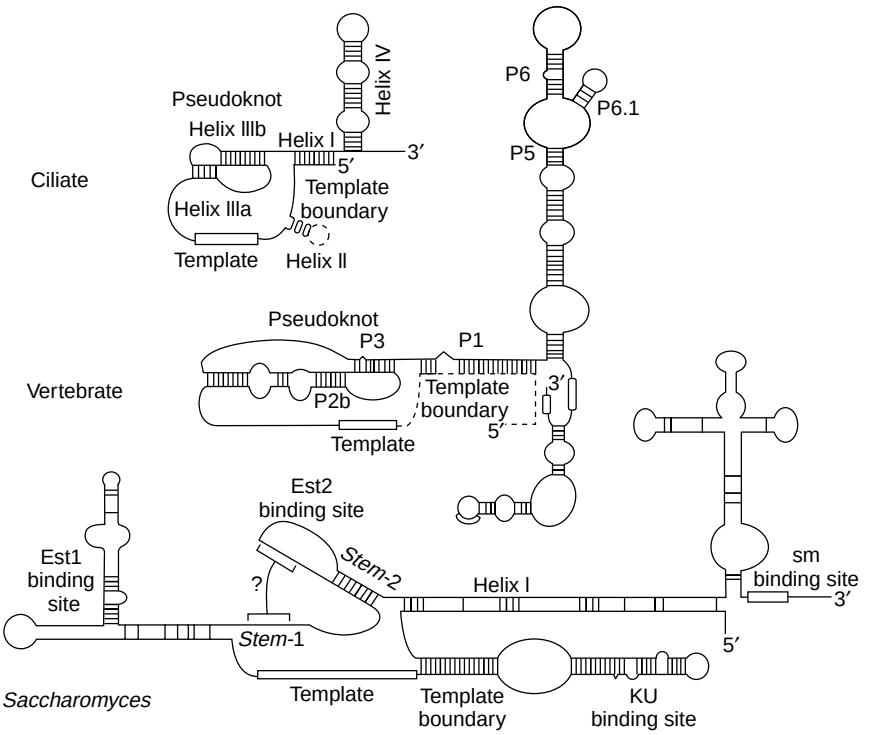

Fig. 5 Secondary structures of ciliate, vertebrate, and yeast (*Saccharomyces*) telomerase RNAs. The structures that are present in some but not all species within a group are represented by dashed lines. Conserved regions that bind to TERT are in green; structures defining the template region are in red, and those defining the template boundary are in blue. Reprinted from Chen, J.-L., Greider, C.W. (2004a) An emerging consensus for telomerase RNA structure, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14683–14684. With permission. (See color plate p. xxiii.)

from each other in the primary sequence, whereas ciliate and yeast TRs bind the catalytic subunit through the pseudoknot and the immediately adjacent stem-loop element.

The most divergent region among TRs is the 3′ region, which is essential for TR maturation and stability. Yeast TR contains a structural element that acts as a binding site for Sm proteins, which are involved in the maturation of snRNA and are required for efficient TR biogenesis. In vertebrate TRs, the 3′ portion contains two elements that are essential for stability and maturation, the H/ACA box and the CR7 domain. The H/ACA motif is usually found in small nucleolar RNAs (snoRNAs) involved in the posttranscriptional modification of rRNA. Among the proteins that bind to this motif is dyskerin/Cbf5p (see the following). The CR7 motif is a telomerase-specific stem-loop motif, which contains a localization signal specific for Cajal bodies (CAB motif), intranuclear structures where snRNAs and snoRNAs are matured and modified. Yeast TRs contain two additional structural elements not conserved in other TRs: the Ku-interacting stem-loop and the Est1p-interacting motif.

2.3.2 Telomerase Protein Component

The telomerase protein (TERT) is the rate-limiting determinant of enzymatic activity. The protein has at least three functions: binding the template RNA, binding DNA, and catalyzing the elongation/translocation events. Telomerases from ciliates, yeast, and humans have also a cleavage activity capable of removing the terminal nucleotide of the DNA extension product at the pause site (i.e. when the extension product is at the 5′ end of the template domain). The TERT gene has been cloned from several species from ciliates to mammals, including humans (Table 2). In essentially all cases there is a single gene encoding a protein of 100–130 kDa that is phylogenetically conserved. The TERT proteins belong to the family of reverse transcriptase with which they share a set of motifs (RT). In addition, all TERTs contain a unique T motif (T) and ciliate TERTs also contain a unique CP motif (Fig. 6). Mutational analysis has mapped the catalytic core of the enzyme within the RT domain; mutations of the T motif also abolish or reduce activity. In addition, amino acid residues at the N and C terminus are important for TR binding, nuclear or nucleolar localization, multimerization and activity. Among the latter there are residues that dissociate catalytic activity, as assayed *in vitro*, from telomere maintenance *in vivo*.

Regulation of telomerase has been extensively characterized in humans. The hTERT gene maps on the short arm of chromosome 5, is over 40 kb in size and comprises 16 exons and 15 introns (Fig. 7). The gene is alternatively spliced and several different transcripts are differentially expressed during embryonic

Tab. 2 Some cloned TERTs.

Species	*Gene # introns*	*Protein*	
		Size [kDa]	*PI*
Human	15	127	11.3
Mouse	>11	127	10.3
Tetrahymena thermophila	18	133	10.0
Schizosaccharomyces pombe	15	116	10.8
Caenorhabditis elegans	10	66	10.0
Oxytricha trifallax	Unspliced	134	10.3
Euplotes aediculatus	Unspliced	123	10.1
Saccharomyces cerevisiae	Unspliced	103	10.0

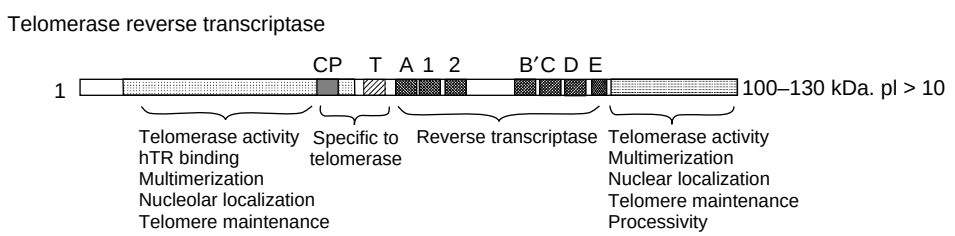

Fig. 6 Schematic structure of the telomerase reverse transcriptase. Reprinted from Harrington, L. (2003) Biochemical aspects of telomerase function, *Cancer Lett.* **194**, 139–154. With permission.

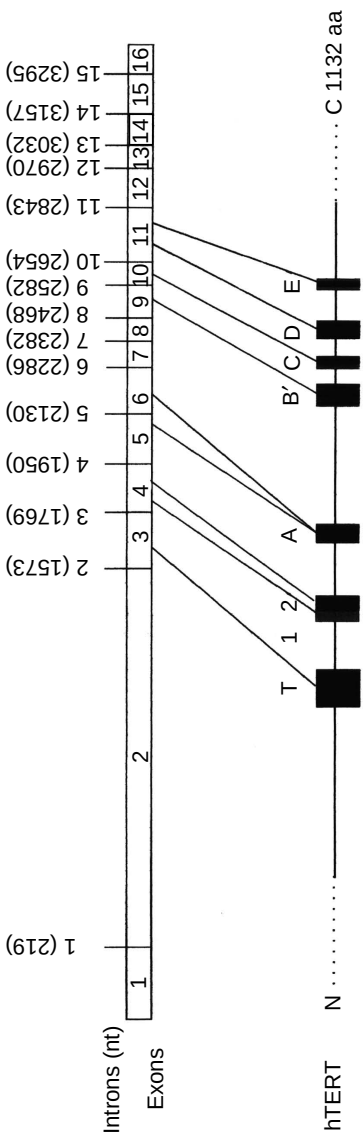

Fig. 7 Organization of the hTERT gene. Introns and exons are indicated. Numbers in brackets refer to the nucleotide position of the intron with nt 1 being the A of the ATG. The location of the telomerase-specific T motif and of the RT motifs and their assignment to exons encoding them is indicated. Reprinted from Cong, Y.S., Wen, J., Bacchetti, S. (1999) The human telomerase catalytic subunit htert: organization of the gene and characterisation of the promoter, *Hum. Mol. Genet.* **8**, 137–142. With permission.

development and are present in somatic cells. Spliced transcripts are not associated with enzymatic activity, but the product of the α splice variant has been reported to inhibit the enzyme when exogenously expressed in cells. It is, however, unclear whether alternative splicing contributes to regulation of hTERT expression.

Telomerase activity and expression of hTERT full-length mRNA disappears during embryonic development and remains absent in most adult somatic tissues. Transcriptional regulation appears to be the process primarily responsible for this phenomenon. The hTERT promoter has been cloned and characterized. It has no TATA or CAAT boxes but contains a large CpG island around the ATG. Promoter methylation, however, does not appear to play a major role in the control of hTERT expression. The minimal promoter, as determined by deletion analysis, spans 330 bp upstream of the ATG but regulatory sequences that may be essential *in vivo* extend further upstream for up to at least 7 kb. The hTERT promoter contains binding sites for many transcription factors several of which have been shown capable of activating (e.g. c-Myc, Sp1) or repressing (e.g. Mad1, p53) hTERT gene expression. The abundance of these sites has suggested that regulation of hTERT expression may operate at multiple levels and/or involve different factors in different tissues. This hypothesis is supported by the observation that transcriptional repressors of telomerase, mapping on different human chromosomes, are differentially active in different cell contexts. Activation of telomerase can also been induced by steroid hormones and the HPV16 E6 protein, while pRB, E2F, and the Wilms' tumor 1 tumor suppressor can repress hTERT transcription. Regulation of telomerase expression may also involve modification of the local chromatin structure through the recruitment by transcription factors of histone acetyltransferases or deacetylases on the hTERT promoter.

Posttranslational modifications, primarily phosphorylation, of the hTERT protein may represent an additional level of control for expression of the enzyme. Phosphorylation by PKC-α and ζ and by Akt has been reported to enhance enzyme activity, while phosphorylation by the tyrosine kinase c-Abl represses it. In addition, phosphorylation has been implicated in the translocation of hTERT from the cytoplasm to the nucleus where it can catalyze telomere elongation.

Lastly, in tumors cells, which are usually telomerase positive (see the following), supernumerary hTERT genes due to chromosome 5 polysomy may contribute to the upregulation of hTERT expression.

2.3.3 Telomerase-associated Proteins

The telomerase complex is larger than the sum of the two core components, TR and TERT, and varies in size in different organisms. Several proteins have been identified as components of the telomerase holoenzyme, and many have been implicated in regulating the function and/or the assembly of the complex. For some, however, the biological function has not been identified.

In the ciliate *T. thermophila* two proteins, p80 and p95, copurify with telomerase activity. Although not essential for catalysis, they act as negative regulators of telomere length. Two additional telomerase-associated proteins, p45 and p65, have been characterized in *Tetrahymena*. Both proteins appear involved in telomere length regulation since their absence results in telomere shortening. p65 interacts with TR and shares a La motif with the *Euplotes aediculatus* protein p43

and the mammalian La antigen, which are also associated with telomerase and are likely to be involved in telomere length regulation.

In *Saccharomyces cerevisiae*, two proteins, Est1p and Est3p, are important for telomerase activity *in vivo*: deletion of either of them causes telomere shortening without affecting catalytic activity *in vitro*. The specific role of these proteins is still under investigation, but the available data are consistent with a role of Est1p in the recruitment of telomerase to the telomere through its interaction with TR and the complex Cdc13p/Ten1/Stn1, which binds to the single-stranded telomeric end. Some data also suggest that Est1p may participate in the activation of the telomerase complex at the telomeres. Est3p is known to associate with the telomerase subunits, but its function is still unknown. The telomerase complex in yeast also includes the heterodimer yKu70/80, which binds TR. Ku, a major component of the nonhomologous end joining (NHEJ) pathway activated in response to double-strand DNA breaks, binds chromosome ends and contributes to telomere length regulation and end protection. In addition, Ku, together with other telomere-binding factors, such as Cdc13p, plays a role in facilitating telomerase action at telomeres, suggesting that its functions in telomere biology are complex.

Cdc13p, a single-strand DNA binding protein with high specificity for the telomeric repeats, plays a major role in telomere maintenance in yeast. Its function is regulated by two binding partners, Stn1p and Ten1p. The complex Cdc13/Ten1/Stn1 is required for telomere end protection, as well as for telomere replication through the recruitment of telomerase. Data from several laboratories support a dual role for Cdc13p: (1) recruitment of telomerase to the telomeres, through Est1p, thereby allowing elongation of chromosome ends, and (2) involvement in C-strand synthesis together with Stn1p, thereby limiting further elongation.

The vertebrate protein TEP1 (telomerase associated protein 1), is the homolog of *Tetrahymena* p80. The N terminus of TEP1 interacts with TR; in addition, the protein contains a C-terminal protein–protein interaction motif (WD40 repeats) that may mediate its association with TERT. The function of TEP1 is still unclear: removal of the protein from the telomerase complex in mice has no effect on telomerase activity or telomere length.

The organization of the telomerase complex in vertebrates shares some similarities with that of yeast, although not all proteins are common to the two organisms. Among hTERT-interacting proteins are the molecular chaperones p23 and hsp90, essential for assembly and activity *in vivo*, and the SMN (survival of motor neuron) protein, which may be involved in localization of telomerase to nucleoli and/or Cajal bodies. Nuclear retention of telomerase appears to require the 14-3-3 protein to which hTERT binds through its C terminus. More recent data have shown that an RNA chaperone, nucleolin, interacts with telomerase and regulates its subcellular localization, although it is not clear whether the protein participates in telomerase assembly or acts as a reservoir of active enzyme in the cell. Two other proteins interacting with hTERT *in vitro* and *in vivo* are EST1A and EST1B, putative human homologs of yeast Est1p. For EST1A, there is experimental evidence for its involvement in telomere capping and telomere length regulation. Both hTERT and hTR interact *in vitro* with PINX1, a TRF1 binding protein, which acts as a negative regulator of the activity. *In vivo* the interaction of PINX1

with telomerase depends on hTERT and involves the hTR-binding domain of this subunit. Among proteins that interact with hTR, dyskerin and its family members Nhp2, Nop10, and Gar1 are important for hTR accumulation, stability, and maturation; hStau and L22 are involved in telomerase localization, and the La antigen plays a yet undefined role in telomere length determination. Lastly, hnRNP proteins A1, C1/C2, and UP1 have been implicated in the regulation of telomerase access to telomeres through their interaction with both hTR and telomeric DNA.

In human cells, a similar role to that of Cdc13p has been proposed for POT1 (protection of telomeres 1). Like Cdc13p, POT1 is a single-strand DNA binding protein with high affinity for telomeric DNA and belongs to the family of DNA binding proteins containing an OB fold (oligonucleotide binding) DNA binding domain. In human cells, POT1 is recruited to the telomeres through the TRF1 complex and has been involved in telomere length regulation. Moreover, the data available so far suggests a possible involvement of POT1 in end protection and in regulating telomerase recruitment and/or activation at the telomeres. Access of telomerase to its telomeric DNA substrate is also dependent on telomeric proteins (e.g. TRF1 and TRF2 in mammals) that stabilize the t-loop structure of telomeres.

3 Telomerase Assays

3.1 Biochemical Assays

3.1.1 The Standard Assay

The standard telomerase reaction uses a telomeric oligonucleotide, such as $d(TTAGGG)_3$ for the human enzyme, as the primer for extension by telomerase in a cell-free extract mixed with appropriate buffers and deoxyribonucleotides. Most of the telomerase activity from mammalian cells resides in the supernatant (i.e. "cytoplasmic fraction") from a 100 000 xg centrifugation (the "S100" fraction) of either mechanically ruptured or detergent-treated human cells. Since telomeres reside in the nucleus of a cell, it is assumed that most telomerase *in vivo* would be found in the nucleus and that the enzyme may leak from this compartment during preparation of the extract.

The products of the telomerase reaction in a cell-free extract can be labeled by the incorporation of $[\alpha^{32}P]$ dGTP during elongation of the primer and then detected by autoradiography after resolution of the products on a polyacrylamide gel. In the case of human telomerase, which is processive and reiteratively copies the six-nucleotide template in a cyclical fashion, the products of the many simultaneous extension reactions occurring in the test tube generate a "ladder" of bands spaced at a six-nucleotide distance (Fig. 8). Moreover, the generation of this characteristic ladder is sensitive to pretreatment of the extract with RNase, since telomerase utilizes its integral RNA as the template for telomere synthesis. Thus, a "+RNase" reaction is often included as a negative control in telomerase assays.

3.1.2 TRAP: Telomere Repeat Amplification Protocol

A highly sensitive and reproducible assay for telomerase utilizes a polymerase chain reaction (PCR) step to amplify the telomerase products prior to detection. Modifications to the reaction conditions for the telomerase portion of this "telomere repeat

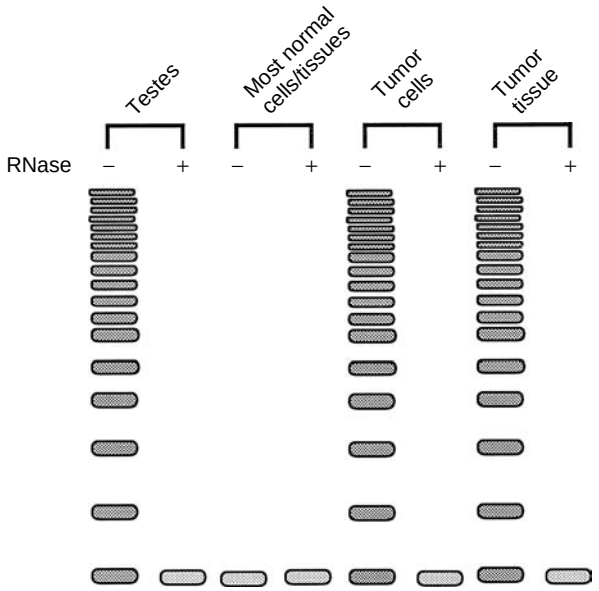

Fig. 8 A standard telomerase "ladder": gel electrophoresis of telomerase products resolves the DNA oligonucleotides according to size. Since the translocation event is slow relative to elongation, there is a greater frequency of partial products at six-nucleotide intervals corresponding to the extension products that reach the 5' end of the template domain. In telomerase-positive tissues such as testes, tumor cells in culture, or tumor biopsies, a distinct ladder of products is seen. This pattern is not seen when the extract is pretreated with RNase, since RNase destroys telomerase activity.

amplification protocol," or TRAP, streamline the overall assay and reduce the chance of PCR artifacts. A critical modification is the use of a non-TTAGGG oligonucleotide (TS), which acts as a good primer for telomerase but reduces the probability of primer/dimer artifacts during the PCR amplification step. The C-rich complementary primer for the PCR reaction is also changed from the perfect CCCTAA repeat to reduce the possibility of false-positive products in the PCR phase of the assay, which may be generated by primer/dimer formation. TRAP is more reliable and sensitive than the standard assay and is capable of detecting telomerase activity in as few as 10–100 telomerase-positive cells among 10^5–10^6 telomerase-negative cells. Thus, very small tissue samples can be analyzed for the presence of positive cells. Detection systems can use either radioactivity or stains incorporated or intercalated into the double-stranded DNA created by the PCR reaction. Fresh or frozen biopsy, autopsy, and surgical specimens, including frozen histological sections, can be analyzed for telomerase activity with TRAP. TRAP, however, is an activity assay and thus cannot be used on fixed material or on poorly stored or necrotic material in which enzyme activity is compromised. To avoid false negatives, tissues should also

be tested for the presence of inhibitors of telomerase or Taq polymerase by mixing or titration experiments.

3.2 Expression Assays

Unlike the RNA component, which is expressed ubiquitously, the telomerase catalytic protein is rate limiting for enzymatic activity and is expressed only in telomerase-positive cells. Detection of the enzyme, therefore, can be achieved by reverse transcription–polymerase chain reaction (RT–PCR) with primers specific for the telomerase mRNA, or by immune reactions using antibodies against the protein component.

4 Telomeres, Telomerase, and Cell Life Span

In multicellular eukaryotes, reproductive cells have unlimited proliferative potential (immortality), whereas somatic cells are capable of only a limited number of divisions. The finite replicative capacity of normal somatic cells has been termed *cellular senescence* and is believed to play a significant role in age-related disease.

It has been proposed that the proliferative capacity of somatic cells is at least partly controlled by the status of their telomeres and telomerase (Fig. 9). In the absence of the enzyme or of any mechanism for telomere maintenance, the loss of telomeric DNA that occurs with each round of replication is thought to act as a ''mitotic clock'' that registers the number of times a cell has divided and signals cessation of division (senescence), presumably when one or more telomeres have attained a functionally suboptimal size or structure. Somatic telomere length, therefore, may be both an indicator of the cell proliferative history and a predictor of its replicative potential. In accord with this model, sperm and ova express telomerase activity and maintain telomere length from one generation to the next. At least in humans, somatic cells have little or no telomerase activity and contain telomeres that are progressively shorter with donor age or cell division in culture. This pattern of telomerase expression is consistent with somatic repression of the enzyme. Telomerase repression in humans appears to be part of an early developmental program, since activity has been detected in certain fetal tissues, whereas, no or little activity is detectable in the corresponding tissue from young postnatal or adult donors. The stringency of telomerase repression is greater in humans than in rodents since telomerase can be easily detected in several adult mouse tissues.

5 Telomerase and Human Disease

5.1 Cancer

Escape from cell senescence and ultimately the immortalization of somatic cells are believed to be important aspects of tumorigenesis. Thus, repression of telomerase and cell senescence can be viewed as tumor suppressor mechanisms. The difference in telomerase regulation between humans and rodents may in part account for the much higher frequency of spontaneous cell immortalization and cancer in rodents compared to humans.

The role of telomerase in human cancer is supported by numerous data indicating that telomerase is reactivated upon cell immortalization *in vitro* (Fig. 9) and is present in the vast majority of tumor-derived cell lines and of tumor tissues

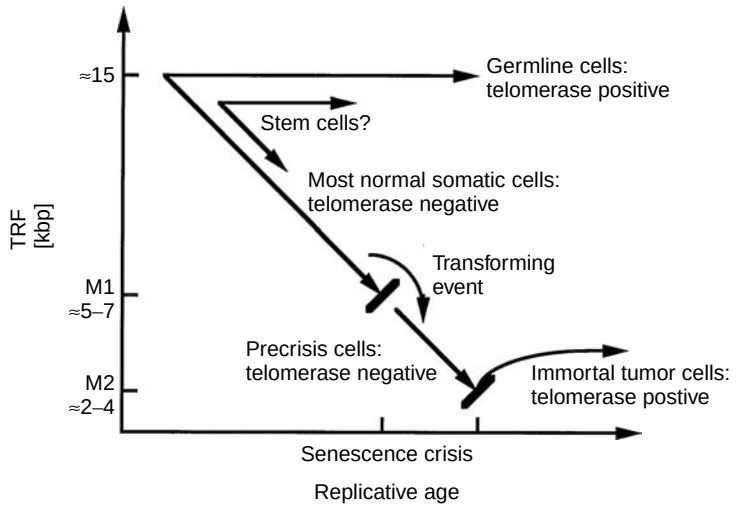

Fig. 9 The telomere hypothesis: reproductive cells, and perhaps stem cells, express telomerase activity and maintain stable telomeres. During development, telomerase is repressed in most somatic cells and, as a consequence, telomeres shorten with cell division. Shortening of telomeres is thought to monitor the proliferative history of somatic cells and to signal growth arrest (senescence) when telomeres have reached a functionally suboptimal size. Transformation extends the proliferative capacity of cells but does not reactivate telomerase, and transformed cells continue to lose telomeric DNA till the end of their extended life span (crisis). At crisis, telomeres on average are extremely short and some may be absent, and chromosomes become unstable. The loss of cell viability occurring at this stage may be due to nonfunctional telomeres and/or lethal chromosome aberrations resulting from terminal fusion of unprotected chromosomes. The rare cells that reacquire telomerase activity have stable telomeres and are immortal.

(Table 3). Indeed, telomerase rates as the most prevalent tumor marker described to date, and it is assumed that enzymatic activity signifies the existence of at least a subset of immortal cells in the malignant tissues. The apparent lack of telomerase in most somatic tissues and its widespread presence in malignant tissues has raised the hope that inhibition of enzymatic activity may be exploited in the management of cancer. Such inhibition should lead to loss of cell viability through resumption of telomere shortening. *In vitro* experimental data from lower eukaryotes and from human cells support this expectation: deleting the telomerase gene or inhibiting telomerase biogenesis by the antisense transcript of the template RNA or of the TERT mRNA can restore a mortal phenotype to otherwise immortal cells by causing telomere loss, and ultimately senescence or crisis of the cell population (Fig. 9). There are, however, several issues to be addressed before it may be possible to validate this approach for clinical treatment. Reagents that inhibit telomerase efficiently and specifically are not yet available. In addition, the biological outcome of telomerase inhibition on normal and malignant tissues *in vivo*

Tab. 3 Telomerase activity in human tumors.

Organ/Tissue	***Cancer*** [a]	***Telomerase*** [b] ***[% positive]***	
Bone marrow/blood	ALL	4/5	(80)
	AML	47/64	(73)
	APL	1/1	(100)
	CLL		
	Early	2/14	(14)
	Late	4/7	(57)
	CML		
	Chronic	30/42	(71)
	Blast	21/21	(100)
	MDS	4/6	(67)
	Myeloma	1/1	(100)
Bladder	Carcinoma	172/185	(92)
Breast	Carcinoma *in situ*	9/12	(75)
	Ductal and lobular	300/339	(88)
Cervix	Carcinoma	16/16	(100)
Colon/rectum	Carcinoma	123/138	(89)
Head/Neck	Squamous cell carcinoma	112/130	(86)
Kidney	Carcinoma	95/115	(83)
	Wilms' tumor	6/6	(100)
Liver	Carcinoma	149/173	(86)
Lung	NSCLC	98/125	(78)
	SCLC	15/15	(100)
Lymph nodes	Lymphoma		
	Low grade	12/14	(86)
	High grade	16/16	(100)
Muscle	Leiomyosarcoma	5/5	(100)
Neural	Anaplastic astrocytoma	2/20	(10)
	Ganglioneuroma	0/4	(0)
	Glioblastoma multiforme	45/60	(75)
	Meningioma		
	Ordinary	5/30	(17)
	Atypical	12/13	(92)
	Malignant	9/9	(100)
	Neuroblastoma	99/105	(94)
	Oligodendroglioma	19/19	(100)
	Retinoblastoma	17/34	(59)
Ovary	Carcinoma	21/23	(91)
Pancreas	Carcinoma	41/43	(95)
Prostate	High grade PIN	3/5	(60)
	Carcinoma	52/58	(90)

(*continued overleaf*)

Tab. 3 (*Continued*)

Organ/Tissue	*Cancer* [a]	*Telomerase* [b] *[% positive]*	
Skin	Basal cell carcinoma	73/77	(95)
	Squamous cell carcinoma	15/18	(83)
	Melanoma	6/7	(86)
Stomach	Adenocarcinoma	56/66	(85)
Testis	Carcinoma	3/3	(100)

[a] ALL: acute myelocytic leukemia; AML: acute myeloid leukemia; APL: acute promyelocytic leukemia; CLL: chronic lymphocytic leukemia; CML: chronic myelocytic leukemia; MDS: myelodysplastic syndrome; NSCLC: non–small cell lung cancer; SCLC: small-cell lung cancer; PIN: prostate intraepithelial neoplasia.
[b] Number of telomerase-positive specimens over total tested.

has yet to be investigated. Relevant to this issue is the existence of alternative mechanisms of telomere maintenance that may become active in transformed cells upon telomerase inhibition.

Last, the nearly ubiquitous presence of telomerase in tumors suggests that the enzyme should also serve as a diagnostic or prognostic marker for cancer. Telomerase activation in cancer does not appear to be associated directly with deregulation of growth or metastatic capacity. Thus, in the multihit process of tumorigenesis, there should be no direct selective advantage for telomerase activation until the tumor cells have reached crisis with critically short telomeres. The early events in tumorigenesis likely involve activation of proto-oncogenes and inactivation of tumor suppressor genes that confer various forms of growth advantage during successive rounds of clonal expansion and selection in the tumor population. Telomerase activation is, therefore, expected to be a late event in most cancers and would correlate with relatively aggressive forms of the disease.

5.2 Age-related Diseases

Genetic ablation of the TR gene in mice, which retain telomerase expression in adult somatic tissues, leads to age-related pathologies. Conversely, exogenous expression of hTERT in telomerase-negative normal human cells in culture extends their life span without acquisition of a transformed phenotype. Reactivation of the enzyme in a transient or regulated manner could theoretically be used to extend the replicative life span of somatic cells, particularly in the case of premature aging syndromes or in aging tissues of otherwise healthy individuals. There are a number of tissues in which replicative cell senescence may play a role in the pathology of age-related diseases. Examples include senescence of dermal fibroblasts in skin aging (decreased wound repair and susceptibility to ulcers), senescence of endothelial cells in vascular aging (cardiovascular diseases), senescence of osteoblasts in bones (osteoporosis), and senescence of lymphocytes in peripheral blood (susceptibility

to infection). An alternative to telomerase reactivation for increasing cell life span is intervention through any mechanism that slows the rate of telomere loss in telomerase-negative tissues, for example, mechanisms that counteract the enhanced rate of loss caused by oxidative stress.

5.3 Dyskeratosis Congenita

In addition to cancer, a human genetic disease, Dyskeratosis congenita (DC), has been linked to telomerase. DC is a rare inherited multisystem disease that targets highly regenerative tissues and results in bone marrow failure. The autosomal dominant form of DC is characterized by mutations in hTR, suggesting that defects in telomerase function may influence the DC phenotype. Indeed, in affected families the disease shows anticipation and correlates with the presence of shorter telomeres. Most mutations found in patients affected by autosomal dominant DC occur in regions of hTR that are important for either catalytic activity or RNA stability. Recent data both *in vitro* and *in vivo* suggests that the clinical manifestations of autosomal dominant DC may be caused by telomere shortening due to lack of activity or haploinsufficiency of telomerase.

See also Aging and Sex, DNA Repair in; Epigenetic Mechanisms in Tumorigenesis; Oncogenes; Oncology, Molecular; Protein and Nucleic Acid Enzymes.

Bibliography

Books and Reviews

Bacchetti, S., Counter, C.M. (1995) Telomeres and telomerase in human cancer, *Int. J. Oncol.* **7**, 423–432.

Chen, J.-L., Greider, C.W. (2004a) An emerging consensus for telomerase RNA structure, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14683–14684.

Chen, J.-L., Greider, C.W. (2004b) Telomerase RNA structure and function: implications for dyskeratosis congenita, *Trends Biochem. Sci.* **29**, 183–192.

Collins, K. (1999) Ciliate telomerase biochemistry, *Annu. Rev. Biochem.* **68**, 178–218.

Cong, Y.-S., Wright, W.E., Shay, J.W. (2002) Human telomerase and its regulation, *Microbiol. Mol. Biol. Rev.* **66**, 407–425.

Greider, C.W., Blackburn, E.H. (2004) Tracking telomerase, *Cell* **S116**, S83–S86.

Harrington, L. (2003) Biochemical aspects of telomerase function, *Cancer Lett.* **194**, 139–154.

Nugent, C.I., Lundblad, V. (1998) The telomerase reverse transcriptase: components and regulation, *Genes Dev.* **12**, 1073–1085.

Shay, J.W., Bacchetti, S. (1997) A survey of telomerase activity in human cancer, *Eur. J. Cancer* **33**, 787–791.

Smogorzewska, A., de Lange, T. (2004) Regulation of telomerase by telomeric proteins, *Annu. Rev. Biochem.* **73**, 177–208.

Primary Literature

Aigner, S., Postberg, J., Lipps, H.J., Cech, T.R. (2003) The euplotes La motif protein p43 has properties of a telomerase specific subunit, *Biochemistry* **42**, 5736–5747.

Armbruster, B.N., Banik, S.S., Gun, C., Smith, A.C., Counter, C.M. (2001) N-terminal domains of the human telomerase catalytic subunit required for enzyme activity in vivo, *Mol. Cell. Biol.* **21**, 7775–7786.

Autexier, C., Greider, C.W. (1998) Mutational analysis of the *tetrahymena* telomerase RNA: identification of residues affecting telomerase activity *in vitro*, *Nucleic Acids Res.* **26**, 787–795.

Bachand, F., Autexier, C. (2001a) Functional regions of human telomerase reverse

transcriptase and human telomerase RNA required for telomerase activity and RNA-protein interactions, *Mol. Cell. Biol.* **21**, 1888–1897.

Bachand, F., Autexier, C. (2001b) Human telomerase RNA-protein interactions, *Nucleic Acids Res.* **29**, 3385–3393.

Backsch, C., Wagenbach, N., Nonn, M., Leistritz, S., Stanbridge, E., Schneider, A., Durst, M. (2001) Microcell-mediated transfer of chromosome 4 into HeLa cells suppresses telomerase activity, *Genes Chromosomes Cancer* **31**, 196–198.

Banik, S.S., Counter, C.M. (2004) Characterisation of interaction between PinX1 and human telomerase subunits hTERT and hTR, *J. Biol. Chem.* **279**, 51745–51748.

Banik, S.S., Guo, C., Smith, A.C., Margolis, S.S., Richardson, D.A., Tirado, C.A., Counter, C.M. (2002) C-terminal regions of the human telomerase catalytic subunit essential for in vivo enzyme activity, *Mol. Cell. Biol.* **22**, 6234–6246.

Baumann, P., Cech, T.R. (2001) Pot1, the putative telomere end-binding protein in fission yeast and humans, *Science* **292**, 1171–1175.

Beattie, T.L., Zhou, W., Robinson, M.O., Harrington, L. (1998) Reconstitution of human telomerase activity in vitro, *Curr. Biol.* **8**, 177–180.

Beattie, T.L., Zhou, W., Robinson, M.O., Harrington, L. (2001) Functional multimerization of the human telomerase reverse transcriptase, *Mol. Cell. Biol.* **21**, 6151–6160.

Blasco, M.A., Funk, W., Villeponteau, B., Greider, C.W. (1995) Functional characterization and developmental regulation of mouse telomerase RNA, *Science* **269**, 1267–1270.

Blasco, M.A., Lee, H.W., Hande, M.P., Samper, E., Lansdorp, P.M., DePinho, R.A., Greider, C.W. (1997) Telomere shortening and tumor formation by mouse cells lacking telomerase rna, *Cell* **91**, 25–34.

Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C.P., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S., Wright, W.E. (1998) Extension of life-span by introduction of telomerase into normal human cells, *Science* **279**, 349–352.

Bryan, T.M., Sperger, J.M., Chapman, K.B., Cech, T.R. (1998) Telomerase reverse transcriptase genes identified in *Tetrahymena termophila* and *Oxytricha trifallax*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8479–8484.

Cohn, M., Blackburn, E.H. (1995) Telomerase in yeast, *Science* **269**, 396–400.

Colgin, L.M., Wilkinson, C., Englezou, A., Kilian, A., Robinson, M.O., Reddel, R.R. (2000) The hTERT alpha splice variant is a dominant negative inhibitor of telomerase activity, *Neoplasia* **2**, 426–432.

Collins, K., Gandhi, L. (1998) The reverse transcriptase component of the tetrahymena telomerase ribonucleoprotein complex, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8485–8490.

Cong, Y.-S., Bacchetti, S. (2000) Histone deacetylation is involved in the transcriptional repression of htert in normal human cells, *J. Biol. Chem.* **275**, 35665–35668.

Cong, Y.S., Wen, J., Bacchetti, S. (1999) The human telomerase catalytic subunit htert: organization of the gene and characterisation of the promoter, *Hum. Mol. Genet.* **8**, 137–142.

Counter, C.M., Meyerson, M., Eaton, E.N., Weinberg, R.A. (1997) The catalytic subunit of yeast telomerase, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9202–9207.

Counter, C.M., Avilion, A.A., LeFeuvre, C.M., Stewart, N.G., Greider, C.W., Harley, C.B., Bacchetti, S. (1992) Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity, *EMBO J.* **11**, 1921–1929.

Counter, C.M., Hahn, W.C., Wei, W., Caddle, S.D., Beijersbergen, R.L., Lansdorp, P.M., Sedivy, J.M., Weinberg, R.A. (1998) Dissociation among *in vitro* telomerase activity, telomere maintenance and cellular immortalization, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14723–14728.

Devereux, T.R., Horikawa, I., Anna, C.H., Annab, L.A., Afshari, C.A., Barrett, J.C. (1999) DNA methylation analysis of the promoter region of the human telomerase reverse transcriptase (hTERT) gene, *Cancer Res.* **59**, 6087–6090.

Etheridge, K.T., Banik, S.S., Armbruster, B.N., Zhu, Y., Terns, R.M., Terns, M.P., Counter, C.M. (2002) The nucleolar localization domain of the catalytic subunit of human telomerase, *J. Biol. Chem.* **15**, 15.

Evans, S.K., Lundblad, V. (2000) Positive and negative regulation of telomerase access to the telomere, *J. Cell. Sci.* **113**, 3357–3364.

Feng, J., Funk, W.D., Wang, S.-S., Weinrich, S.L.O., Avilion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le,

S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W., Villeponteau, B. (1995) The RNA component of human telomerase, *Science* **269**, 1236–1241.

Fitzgerald, M.S., McKnight, T.D., Shippen, D.E. (1996) Characterisation and developmental patterns of telomerase expression in plants, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14422–14427.

Friedman, K.L., Cech, T.R. (1999) Essential functions of aminoterminal domains in the yeast telomerase catalytic subunit revealed by selection of viable mutants, *Genes Dev.* **13**, 2863–2874.

Gewin, L., Galloway, D.A. (2001) E box-dependent activation of telomerase by human papillomavirus type 16 E6 does not require induction of c-myc, *J. Virol.* **75**, 7198–7201.

Greenberg, R.A., Allsopp, R.C., Chin, L., Morin, G.B. (1998) Expression of mouse telomerase reverse transcriptase during development, differentiation and proliferation, *Oncogene* **16**, 1723–1730.

Greider, C.W. (1991) Telomerase is processive, *Mol. Cell. Biol.* **11**, 4572–4580.

Greider, C.W., Blackburn, E.H. (1985) Identification of a specific terminal transferase activity in tetrahymena extracts, *Cell* **43**, 405–413.

Greider, C.W., Blackburn, E.H. (1987) The telomere terminal transferase of *tetrahymena* is a ribonucleoprotein enzyme with two kinds of primer specificity, *Cell* **51**, 887–898.

Greider, C.W., Blackburn, E.H. (1989) A telomeric sequence in the rna of *tetrahymena* telomerase required for telomere repeat synthesis, *Nature* **337**, 331–337.

Gunes, C., Lichtsteiner, S., Vasserot, A.P., Englert, C. (2000) Expression of the hTERT gene is regulated at the level of transcriptional initiation and repressed by Mad1, *Cancer Res.* **60**, 2116–2121.

Guo, C., Armbruster, B.N., Price, D.T., Counter, C.M. (2003) In vivo regulation of hTERT expression and telomerase activity by androgen, *J. Urol.* **170**, 615–618.

Hahn, W.C., Stewart, S.A., Brooks, M.W., York, S.G., Eaton, E., Kurachi, A., Beijersbergen, R.L., Knoll, J.H., Meyerson, M., Weinberg, R.A. (1999) Inhibition of telomerase limits the growth of human cancer cells, *Nat. Med.* **5**, 1164–1170.

Harley, C.B., Futcher, A.B., Greider, C.W. (1990) Telomere shorten during ageing of human fibroblasts, *Nature* **345**, 458–460.

Harrington, L., McPhail, T., Mar, V., Zhou, W., Oulton, R., Bass, M.B., Arruda, L., Robinson, M.O. (1997a) A mammalian telomerase associated protein, *Science* **275**, 973–977.

Harrington, L., Zhou, W., McPhail, T., Oulton, R., Yeung, D.S., Mar, V., Bass, M.B., Robinson, M.O. (1997b) Human telomerase contains evolutionary conserved catalytic and structural subunits, *Genes Dev.* **11**, 3109–3115.

Hayflick, L. (1965) The limited *in vitro* lifetime of human diploid cell strains, *Exp. Cell Res.* **37**, 614–636.

Herbert, B., Pitts, A.E., Baker, S.L., Hamilton, S.E., Wright, W.E., Shay, J.W., Corey, D.R. (1999) Inhibition of human telomerase in immortal human cells leads to progressive telomere shortening and cell death, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14276–14281.

Horikawa, I., Cable, P.L., Afshari, C., Barrett, J.C. (1999) Cloning and characterization of the promoter region of human telomerase reverse transcriptase gene, *Cancer Res.* **59**, 826–830.

Iishi, Y., Tsuyama, N., Maeda, S., Tahara, H., Ide, T. (1999) Telomerase activity in hybrids between telomerase-negative and telomerase-positive immortal human cells is repressed in the different complementation groups but not in the same complementation groups of immortality, *Mech. Ageing Dev.* **110**, 175–193.

Jiang, X.R., Jimenez, G., Chang, E., Frolkis, M., Kusler, B., Sage, M., Beeche, M., Bodnar, A.G., Wahl, G.M., Tlsty, T.D., Chiu, C.P. (1999) Telomerase expression in human somatic cells does not induce changes associated with a transformed phenotype, *Nat. Genet.* **21**, 111–114.

Kang, S.S., Kwon, T., Kwon, D.Y., Ido, S.I. (1999) Akt protein kinase enhances human telomerase activity through phosphorylation of telomerase reverse transcriptase subunit, *J. Biol. Chem.* **274**, 13085–13090.

Kharbanda, S., Kumar, V., Dhar, S., Pandey, P., Chen, C., Majumder, P., Yuan, Z.M., Whang, Y., Strauss, W., Pandita, T.K., Weaver, D., Kufe, D. (2000) Regulation of the hTERT telomerase catalytic subunit by c-Abl tyrosine kinase, *Curr. Biol.* **10**, 568–575.

Khurts, S., Masutomi, K., Delgermaa, L., Arai, K., Oishi, N., Mizuno, H., Hayashi, N., Hahn, W.C., Murakami, S. (2004) Nucleolin interacts with telomerase, *J. Biol. Chem.* **279**, 51508–51515.

Kilian, A., Bowtell, D.D., Abud, H.E., Hime, G.R., Venter, D.J., Keese, P.K., Duncan, E.L.,

Reddel, R.R., Jefferson, R.A. (1997) Isolation of a candidate human telomerase catalytic subunit gene, which reveals complex splicing patterns in different cell types, *Hum. Mol. Genet.* **6**, 2011–2019.

Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L., Shay, J.W. (1994) Specific association of human telomerase activity with immortal cells and cancer, *Science* **266**, 2011–2015.

Kyo, S., Takakura, M., Kanaya, T., Zhuo, W., Fujimoto, K., Nishio, Y., Orimo, A., Inoue, M. (1999) Estrogen activates telomerase, *Cancer Res.* **59**, 5917–5921.

Kyo, S., Takakura, M., Taira, T., Kanaya, T., Itoh, H., Yutsudo, M., Ariga, H., Inoue, M. (2000) Sp1 cooperates with c-Myc to activate transcription of the human telomerase reverse transcriptase gene (hTERT), *Nucleic Acids Res.* **28**, 669–677.

Li, H., Zhan, L.L., Funder, J.W., Liu, J.P. (1997) Protein phosphatase 2A inhibits nuclear telomerase activity in human breast cancer cells, *J. Biol. Chem.* **272**, 16729–16732.

Lin, J.-J. (1996) The Saccharomyces CDC13 protein is a single-strand TG_{1-3} telomeric DNA-binding protein in vitro that affects telomere behaviour in vivo, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13760–13765.

Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V., Cech, T.R. (1997) Reverse transcriptase motifs in the catalytic subunit of telomerase, *Science* **276**, 561–567.

Liu, D., Safari, A., O'Connor, M.S., Chan, D.W., Laegeler, A., Qin, J., Songyang, Z. (2004) PTOP interacts with POT1 and regulates its localization to telomeres, *Nat. Cell Biol.* **6**, 673–680.

Malik, H.S., Burke, W.D., Eickbush, T.H. (2000) Putative telomerase catalytic subunits from *Giardia lamblia* and *Caenorhabditis elegans*, *Gene* **251**, 101–108.

Mantell, L.L., Greider, C.W. (1994) Telomerase activity in germline and embryonic cells of xenopus, *EMBO J.* **13**, 3211–3217.

Marrone, A., Mason, P.J. (2003) Dyskeratosis congenita, *Cell. Mol. Life Sci.* **60**, 507–517.

McElligot, R., Wellinger, R.J. (1997) The terminal DNA structure of mammalian chromosomes, *EMBO J.* **16**, 3705–3714.

Metz, A.M., Love, R.A., Strobel, G.A., Long, D.M. (2001) Two telomerase reverse transcriptases (terts) expressed in *Candida albicans*, *Biotechnol. Appl. Biochem.* **34**, 47–54.

Meyerson, M., Counter, C.M., Eaton, E.N., Ellison, L.W., Steiner, P., Caddle, S.D., Ziaugra, L., Beijersbergen, R.L., Davidoff, M.J., Liu, Q., Bacchetti, S., Haber, D.A., Weinberg, R.A. (1997) Hest2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization, *Cell* **90**, 785–795.

Misiti, S., Nanni, S., Fontemaggi, G., Cong, Y.-S., Wen, J., Hirte, H.W., Piaggio, G., Sacchi, A., Pontecorvi, A., Bacchetti, S., Farsetti, A. (2000) Induction of htert expression and telomerase activity by estrogens in human ovary epithelium cells, *Mol. Cell. Biol.* **20**, 3764–3771.

Morales, C.P., Holt, S.E., Ouellette, M., Kaur, K.J., Yan, Y., Wilson, K.S., White, M.A., Wright, W.E., Shay, J.W. (1999) Absence of cancer-associated changes in human fibroblasts immortalized with telomerase, *Nat. Genet.* **21**, 115–118.

Moriarty, T.J., Huard, S., Dupuis, S., Autexier, C. (2002) Functional multimerization of human telomerase requires an RNA interaction domain in the N terminus of the catalytic subunit, *Mol. Cell. Biol.* **22**, 1253–1265.

Morin, G.B. (1989) The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats, *Cell* **59**, 1493–1497.

Muslin, A.J., Xing, H. (2000) 14-3-3 proteins: regulation of subcellular localization by molecular interference, *Cell. Signal.* **12**, 703–709.

Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., Harley, C.B., Cech, T.R. (1997) Telomerase catalytic subunit homologs from fission yeast and human, *Science* **277**, 955–959.

Nishimoto, A., Miura, N., Horikawa, I., Kugoh, H., Murakami, Y., Hirohashi, S., Kawasaki, H., Gazdar, A.F., Shay, J.W., Barrett, J.C., Oshimura, M. (2001) Functional evidence for a telomerase repressor gene on human chromosome 10p51.1, *Oncogene* **20**, 828–835.

Nugent, C.I., Hughes, T.R., Lue, N.F., Lundblad, V. (1996) Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance, *Science* **274**, 249–252.

Oguchi, K., Liu, H., Tamura, K., Takahaschi, H. (1999) Molecular cloning and characterization of AtTERT, a telomerase reverse transcriptase

homolog in *Arabidopsis thaliana*, *FEBS Lett.* **457**, 465–469.

Oh, S., Song, Y.H., Yim, J., Kim, T.K. (2000) Identification of Mad as a repressor of the human telomerase (hTERT) gene, *Oncogene* **19**, 1485–1490.

Oh, S., Song, Y.H., Kim, U.J., Yim, J., Kim, T.K. (1999) The Wilms' tumor 1 tumor suppressor gene represses transcription of the human telomerase reverse transcriptase gene, *J. Biol. Chem.* **274**, 37473–37478.

Poole, J.C., Andrews, L.G., Tollefsbol, T.O. (2001) Activity, function and gene regulation of the catalytic subunit of telomerase (hTERT), *Gene* **269**, 1–12.

Prowse, K.R., Avilion, A.A., Greider, C.W. (1993) Identification of a nonprocessive telomerase activity from mouse cells, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1493–1497.

Reichenbach, P., Hoss, M., Azzalin, C.M., Nabholz, M., Bucher, P., Lingner, J. (2003) A human homolog of yeast Est1 associates with telomerase and uncaps chromosome ends when overexpressed, *Curr. Biol.* **13**, 568–574.

Rudolph, K.L., Chang, S., Lee, H.W., Blasco, M., Gottlieb, C.J., Greider, C., DePinho, R.A. (1999) Longevity, stress response and cancer in aging telomerase-deficient mice, *Cell* **96**, 701–712.

Schramke, V., Luciano, P., Brevet, V., Guillot, S., Corda, Y., Longhese, M.P., Gilson, E., Geli, V. (2004) RPA regulates telomerase action by providing Est1p access to chromosome ends, *Nat. Genet.* **36**, 46–54.

Seto, A.G., Livengood, A.J., Tzfati, Y., Blackburn, E.H., Cech, T.R. (2002) A bulged stem tethers Est1p to telomerase RNA in budding yeast, *Genes Dev.* **16**, 2800–2812.

Singer, M.S., Gottschling, D.E. (1994) Tlc1: template rna component of *Saccharomyces cerevisiae* telomerase, *Science* **266**, 404–409.

Snow, B.E., Erdmann, N., Cruickshank, J., Goldman, H., Gill, R.M., Robinson, M.O., Harrington, L. (2003) Functional conservation of the telomerase protein Est1p in humans, *Curr. Biol.* **13**, 698–704.

Steenbergen, R.D., Kramer, D., Mijer, C.J., Walboomers, J.M., Trott, D.A., Cuthbert, A.P., Newbold, R.F., Overkamp, W.J., Zdzienicka, M.Z., Snijders, P.J. (2001) Telomerase suppression by chromosome 6 in a human papillomavirus type 16-immortalized keratinocyte cell line and in a cervical cancer cell line, *J. Natl. Cancer Inst.* **93**, 865–872.

Tahara, H., Yasui, W., Tahara, E., Fujimoto, J., Ito, K., Tamai, K., Nakayama, J., Ishikawa, F., Ide, T. (1999) Immuno-histochemical detection of human telomerase catalytic component, hTERT, in human colorectal tumor and nontumor tissue sections, *Oncogene* **18**, 1561–1567.

Takakura, M., Kyo, S., Kanaya, T., Hirano, H., Takeda, J., Yutsudo, M., Inoue, M. (1999) Cloning of human telomerase catalytic subunit (hTERT) gene promoter and identification of proximal core promoter sequences essential for transcriptional activation in immortalized and cancer cells, *Cancer Res.* **59**, 551–557.

Takakura, M., Kyo, S., Sowa, Y., Wang, Z., Yatabe, N., Maida, Y., Tanaka, M., Inoue, M. (2001) Telomerase activation by histone deacetylase inhibitor in normal cells, *Nucleic Acids Res.* **29**, 3006–3011.

Tanaka, H., Shimizo, M., Horikawa, I., Kugoh, H., Yokota, J., Barrett, J.C., Oshimura, M. (1998) Evidence for a putative telomerase repressor gene in the 3p14.2p21.1 region, *Genes Chromosomes Cancer* **23**, 123–133.

Ulaner, G.A., Hu, J.F., Vu, T.H., Giudice, L.C., Hoffman, A.R. (1998) Telomerase activity in human development is regulated by human telomerase reverse transcriptase (htert) transcription and by alternate splicing of htert transcripts, *Cancer Res.* **58**, 4168–4172.

Vaziri, H., Benchimol, S. (1998) Reconstitution of telomerase activity in normal human cells leads to elongation of telomeres and extended replicative life span, *Curr. Biol.* **8**, 279–282.

Veldman, T., Etheridge, K.T., Counter, C.M. (2004) Loss of hPot1 function leads to telomere instability and a cut-like phenotype, *Curr. Biol.* **14**, 2264–2270.

Wang, J., Xie, L.H., Allan, S., Beach, D., Hannon, G.J. (1998) Myc activates telomerase, *Genes Dev.* **12**, 1769–1774.

Weinrich, S.L., Pruzan, R., Ma, L., Tesmer, V.M., Holt, S.E., Bodnar, A.G., Lichtsteiner, S., Kim, N.W., Trager, J.B., Taylor, R.D., Carlos, R., Andrews, W.H., Wright, W.E., Shay, J.W., Harley, C.B., Morin, G.B. (1997) Reconstitution of human telomerase with the template rna component htr and the catalytic protein htrt, *Nat. Genet.* **17**, 498–502.

Wen, J., Cong, Y.-S., Bacchetti, S. (1998) Reconstitution of wild-type or mutant telomerase activity in telomerase-negative immortal human cells, *Hum. Mol. Genet.* **7**, 1137–1141.

Witkin, K.L., Collins, K. (2004) Holoenzyme proteins required for the physiological assembly and activity of telomerase, *Genes Dev.* **18**, 1107–1118.

Wu, K.J., Grandori, C., Amacker, M., Simon-Vermot, N., Polack, A., Lingner, J., Dalla Favera, R. (1999) Direct activation of tert transcription by c-myc, *Nat. Genet.* **21**, 220–224.

Xia, J., Peng, Y., Mian, I.S., Lue, N.F. (2000) Identification of functionally important domains in the N-terminal region of telomerase reverse transcriptase, *Mol. Cell. Biol.* **20**, 5196–5207.

Xu, H.J., Zhou, Y., Ji, W., Perug, G.S., Kruzelock, R., Kong, C.T., Bast, R.C., Mills, G.B., Li, J., Hu, S.X. (1997) Reexpression of the retinoblastoma protein in tumor cells induces senescence and telomerase inhibition, *Oncogene* **15**, 2589–2596.

Xu, D., Wang, Q., Gruber, A., Bjorkholm, M., Chen, Z., Zaid, A., Selivanova, G., Peterson, C., Wiman, K.G., Pisa, P. (2000) Downregulation of telomerase reverse transcriptase mRNA expression by wild type p53 in human tumor cells, *Oncogene* **19**, 5123–5133.

Zhang, A., Zheng, C., Lindvall, C., Hou, M., Ekedahl, J., Lewensohn, R., Yan, Z., Yang, X., Henriksson, M., Blennow, E., Nordenskjold, M., Zetterberg, A., Bjorkholm, M. (2000) Frequent amplification of the telomerase reverse transcriptase gene in human tumors, *Cancer Res.* **60**, 6230–6235.

Telomere-binding Proteins in Fission Yeast

Junko Kanoh
Graduate School of Biostudies, Kyoto University, Kyoto, Japan

1 Telomere-specific Complex 230
1.1 Taz1 230
1.2 Rap1 and Rif1 232
1.3 End-binding Proteins 234

2 Telomere Heterochromatin 237
2.1 Gene Silencing 237
2.2 Swi6 237
2.3 RNAi and RITS 238

Bibliography 239
Books and Reviews 239
Primary Literature 239

Keywords

Chromatin Immunoprecipitation (ChIP)
Living cells are fixed with formaldehyde, and cell extracts are prepared. Specific protein–DNA complexes are precipitated by immunoprecipitation and the associated DNA is purified and amplified by PCR (polymerase chain reaction) with specific primers.

Fission Yeast Telomere and Telomeric G-rich Overhang
Telomere DNA comprises specific G-rich tandem repeat sequences. The consensus sequence of fission yeast telomeric repeat is TTACAGG. The average length of fission yeast telomeric repeats is 300 bp. Although telomeric repeat sequence and telomere

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

length vary, the structure of the very end of the telomere is conserved among species. The G-rich strand protrudes as a 3′ single-strand overhang.

Life Cycle of Fission Yeast
The fission yeast, *Schizosaccharomyces pombe*, basically grows in the haploid state. When the medium is deprived of nitrogen, the h^+ and h^- cells mate, undergo meiosis and sporulation.

Telomerase
Telomerase is a reverse transcriptase containing the RNA template for telomeric repeats. It compensates for the reduction of telomere length caused by the end replication problem of linear DNA. Fission yeast telomerase is encoded by the *trt1*$^+$ gene.

Telomere Shortening in the Absence of Telomerase
In humans, although telomerase activity is relatively high in germ cells, it is significantly low in most somatic cells, resulting in the gradual reduction of telomere length. Fission yeast telomerase (Trt1) is active in both the mitotic cell cycle and the meiotic cell cycle, thereby maintaining telomere length. In the absence of Trt1, however, the telomere length is gradually decreased, leading to cell growth arrest.

The telomere is a specialized heterochromatin localized at the end of the linear chromosome, and is essential for the multiple aspects of genome integrity. Telomere functions are performed mostly by telomere-binding proteins. Although many of the telomere-binding proteins are conserved among species, some variations exist. Recent studies of the fission yeast telomere-binding proteins have revealed that the components of the fission yeast telomere are more similar to those of humans than to those of budding yeast, indicating that fission yeast is evolutionarily more closely related to humans than to budding yeast in terms of the telomere.

1 Telomere-specific Complex

1.1 Taz1

In budding yeast, the Rap1 (scRap1, repressor/activator protein1) protein binds to the telomeric repeats as well as to the mating-type silencer elements and the UAS (upstream activation sequence) of the genes for ribosomal proteins, translational components, and glycolytic enzymes. scRap1 regulates telomere length and telomere position effect by recruiting two groups of proteins (see below). Thus, scRap1 is the most fundamental protein that performs telomere functions.

The fission yeast has three chromosomes, and each chromosome end has approximately 300 bp of telomeric repeats.

Cooper and coworkers attempted to identify a protein that directly binds to the telomeric repeats in fission yeast by a one-hybrid screen. The Taz1 protein was identified as a protein that specifically binds to the fission yeast telomeric repeats. The amino acid sequence of Taz1 is highly homologous with that of hTRF1 and hTRF2 in humans, and these proteins constitute the TTAGGG repeat factor (TRF) family (see Fig. 1). There is, however, no TRF homolog in budding yeast. They contain a TRFH (TRF homology) domain and a single Myb domain in common. Their carboxyl terminal Myb domains form the helix-turn-helix structure. hTRF1 and hTRF2 also directly bind to the human telomeric repeat DNA. It has been shown that the TRF family proteins form a homodimer via their TRFH domains to bind to the duplex telomeric repeats by means of their Myb domains, because a single Myb domain is not sufficient for association with DNA.

The localization of Taz1 is not limited to the telomeric repeats alone; Taz1 is also localized at the subtelomeric TAS (telomere-associated sequence) spanning approximately 10 kb adjacent to the telomeric repeats. Furthermore, Taz1 can be newly recruited to the TAS in the absence of the telomeric repeats, indicating that Taz1 recognizes not only the telomeric repeats, but also some elements (the DNA sequence or the associating proteins) in the TAS. In budding yeast, scRap1 is also localized at the subtelomeric region, suggesting that there is a common mechanism that induces the spreading of telomeric repeat-binding proteins into the subtelomeric region.

The disruption of the *taz1*$^+$ gene causes severe defects in telomere functions. First, telomeres are markedly elongated to ~2.5 to 4.5 kb in the *taz1* mutant. Taz1 is surmised to act as a negative regulator of telomerase, limiting the accessibility of telomerase to telomere DNA. Second, telomere silencing is alleviated in the *taz1* mutant. Genes that are transcriptionally active at the original locus are silenced when they are located near the telomeric repeats. This telomere silencing is caused by the specific chromatin structure. The deletion of *taz1*$^+$ probably results in a dramatic change of the higher-order chromatin structure at the telomeric repeats. Third, the deletion of *taz1*$^+$ causes a defect in telomere clustering toward the spindle pole body (SPB) in meiotic prophase. In vegetative cells, all the centromeres associate near the SPB, whereas telomeres cluster partially at the nuclear periphery apart from the SPB and the centromeres.

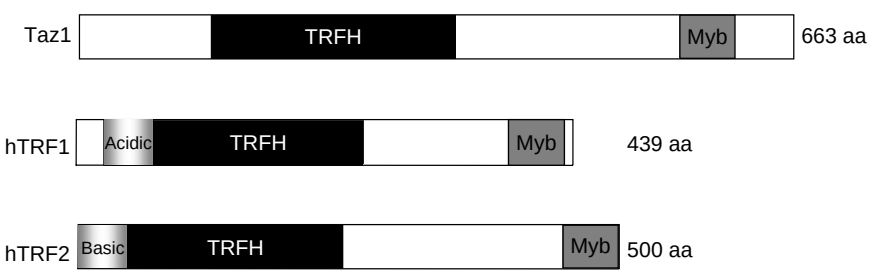

Fig. 1 Taz1 belongs to the TRF family. "TRFH" indicates a TRF homology domain, which mediates formation of a homodimer. "Myb" indicates a DNA-binding domain. Amino-terminal region of hTRF1 is very acidic, whereas that of hTRF2 is very basic. Amino-terminal region of Taz1 does not have either of the features.

Upon the induction of mating by nitrogen starvation, all the telomeres cluster toward the SPB, and the centromeres are displaced from the SPB. Then, the chromosomes (nucleus) move extensively in the cell with the SPB as the leading edge. In this horsetail stage, homologous chromosomes are aligned, inducing efficient homologous recombination. Thus, Taz1 is required for the stable association of the telomeres with the SPB during meiotic prophase and for the efficient meiotic recombination. Furthermore, *taz1* mutant cells produce abnormal spores, indicating that the precise telomere-SPB association is required for the completion of meiosis. Fourth, the *taz1* mutant exhibits lethal telomere fusions in the G1 phase after nitrogen starvation. This telomere fusion does not occur in the absence of pKu70 or Lig4, which mediates Ku-dependent nonhomologous end joining (NHEJ). Therefore, Taz1 protects chromosome ends from the DNA repair mediated by the Ku-dependent NHEJ. Fifth, the *taz1* mutant is defective in growth at 20 °C, although it grows normally at higher temperatures. At 20 °C, the *taz1* mutant shows DNA double-strand breaks (DSBs), G2/M cell cycle delay caused by DNA damage checkpoint, and chromosome missegregation. In addition, the mutant is hypersensitive to treatments that induce DSBs. Thus, Taz1 is required for preventing DSBs and for DNA repair throughout the genome.

1.2 Rap1 and Rif1

It is thought that scRap1 binds to the DNA as a monomer because scRap1 apparently has two Myb domains that are sufficient for the DNA-binding activity. The deletion of scRap1 is lethal, whereas some *rap1* mutants are viable. The *rap1-17* mutant has a strikingly elongated telomere DNA, suggesting that scRap1 is a negative regulator of telomerase, similar to the fission yeast Taz1. scRap1 recruits the two groups of proteins to the telomeres by interacting with its carboxyl terminal (RCT: Rap1 C terminus) domain. The proteins of the first group are scRif1 (Rap1-interacting factor1) and scRif2, which are mainly involved in telomere length control. The deletion of either protein causes moderate elongation of the telomere DNA, and the deletion of both proteins results in dramatic telomere elongation, similar to the phenotype of the *rap1-17* mutant. Proteins of the second group include Sir3 (silent information regulator 3) and Sir4, which are mainly involved in the formation of heterochromatin. Sir3 and Sir4 form a complex with Sir2, a histone deacetylase, and spread into the neighboring subtelomeric region from the telomeric repeats, resulting in a higher-order heterochromatin structure.

Fission yeast Rap1 (spRap1) was identified by a BLAST search in the genome database and by screening for nuclear proteins from the GFP-fusion genomic DNA library. spRap1, scRap1, and human Rap1 (hRap1) share homology in the BRCT (BRCA1 or breast cancer gene 1 C-terminal) domain and the Myb domain (see Fig. 2). spRap1 has an additional Myb-like domain that is similar to the second Myb domain of scRap1, although its physiological significance is not known. In contrast to spRap1, scRap1 and hRap1 have an RCT domain at their carboxyl termini. This suggests that protein–protein interaction is mediated in a different manner in fission yeast. Fission yeast Rif1 (spRif1) was identified as a homolog of scRif1 in the genome database. The amino acid sequence of spRif1 is homologous with that of scRif1 throughout the entire sequence.

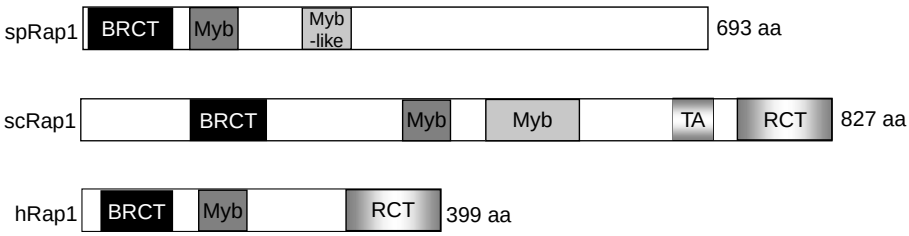

Fig. 2 Structures of the Rap1 family proteins. "BRCT" indicates BRCA1 C terminus. "Myb" indicates an Myb-related helix-turn-helix motif. "RCT" indicates Rap1 C terminus, which mediates interaction with the scRif proteins and with the Sir proteins in budding yeast. "TA" indicates a transactivation domain.

However, it is not known whether there is a Rif2 homolog in fission yeast.

Taz1 does not require spRap1 or spRif1 for telomere binding. On the other hand, both spRap1 and spRif1 interact with Taz1 and are localized at the telomeres in a Taz1-dependent manner. Thus, different from the case of scRap1, fission yeast Rap1 does not directly bind to the telomere DNA at least in mitotically growing cells. hRap1 also does not bind directly to the telomere DNA, and is localized at the telomeres via interaction with TRF2. It is speculated that the structure of the Rap1 protein was changed during the course of evolution. One possibility is that budding yeast scRap1 has gained two Myb domains for direct binding to the telomere DNA as a monomer. The second possibility is that fission yeast and human Rap1 proteins have lost one Myb domain and gained TRF proteins instead.

Both spRap1 and spRif1 are involved in telomere length control (see Fig. 3). The deletion of spRap1 causes a marked elongation of the telomere DNA to the extent similar to that of the *taz1* mutant. The *rif1* mutant has telomere DNA that is approximately 200 bp longer than that of the wild type. The *taz1rap1* and *taz1rif1* mutants have telomere lengths similar to that of the *taz1* mutant, indicating that spRap1 and spRif1 regulate telomere length in a Taz1-dependent manner, which is consistent with the fact that spRap1 and spRif1 are recruited to the telomeres by interacting with Taz1. The *rap1* mutant has additional telomere phenotypes. First, telomere silencing is strikingly alleviated. Second, it is defective in telomere clustering toward the SPB in meiotic prophase, resulting in the production of abnormal spores (see Fig. 4). Collectively, three telomere functions, namely, telomere length control, telomere silencing, and telomere clustering toward the SPB in the meiotic prophase, are impaired in the *rap1* mutant cells, in which Taz1 is still localized at the telomeres. Therefore, Taz1 itself cannot fulfill those telomere functions, and its primary role is to recruit such telomere regulators as spRap1 and spRif1 to the telomeres. However, not only does Taz1 recruit spRap1 and spRif1, but it is also required for the stability of the whole genome, whereas spRap1 and spRif1 are not. There may also be other regulators that associate with Taz1.

Indirect immunofluorescence microscopy has shown that spRap1 is sometimes colocalized with the telomeres and the SPB in the *taz1* mutant at the horsetail stage. This suggests two possibilities. The first is that spRap1 has an intrinsic tendency to be localized near the SPB.

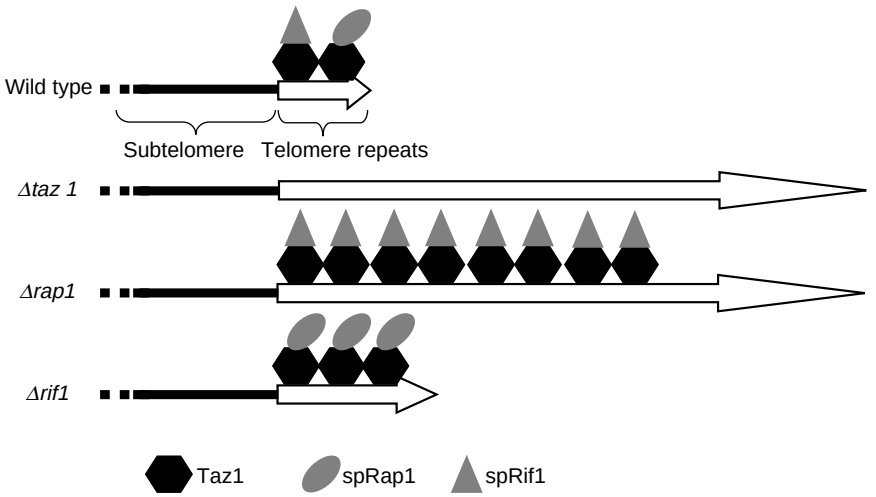

Fig. 3 Telomere-binding proteins regulate telomere length. In the wild-type strain, both spRap1 and spRif1 are associated with Taz1 at the telomeres. In the *taz1* mutant, none of them are localized at the markedly elongated telomeres (exception: spRap1 in meiotic prophase). In the *rap1* mutant, Taz1 and spRif1 are localized at the telomeres, but telomere DNA is markedly elongated. In the *rif1* mutant, Taz1 and spRap1 are localized at the telomeres, and telomere DNA is moderately elongated.

The protein complex that specifically functions for telomere clustering toward the SPB may recruit spRap1 to the SPB in a Taz1-independent manner. The second possibility is that spRap1 has an activity to bind to telomere DNA inefficiently. In fact, spRap1 contains an additional Myb-like domain besides a Myb domain that is conserved in yeasts and humans (described above). These two Myb domains, although incomplete, may serve as DNA-binding domains for spRap1 in the meiotic prophase. Localization of spRif1 is different from that of Taz1 and spRap1. spRif1 is localized at the telomeres, although mainly in the nucleolus in the wild-type cells. This suggests that spRif1 has functions that are not related to the telomere maintenance.

hRap1 is also involved in telomere length control. The overexpression of hRap1 results in telomere elongation. On the other hand, the primary function of hRif1 seems to be different from those of spRif1 and scRif1. hRif1 is localized at the DSBs, not at the functional telomeres. hRif1 is regulated by ataxia telangiectasia-mutated (ATM) and 53BP1 and functions in the S-phase checkpoint pathway. It is possible that the yeast Rif1 proteins also have similar functions.

1.3 End-binding Proteins

The Ku complex functions as a heterodimer consisting of Ku70 and Ku80, and basically binds to the DSBs to mediate DNA repair by NHEJ. Fission yeast Ku70 (pKu70) and Ku80 (pKu80) were identified by a BLAST search. pKu is localized near the chromosome ends in a Taz1-independent manner. The deletion of pKu70/pKu80 results in stable but shorter telomeric repeats, elongation of G-rich overhang, and TAS rearrangement. The rearrangement of TAS in the *pku* mutant

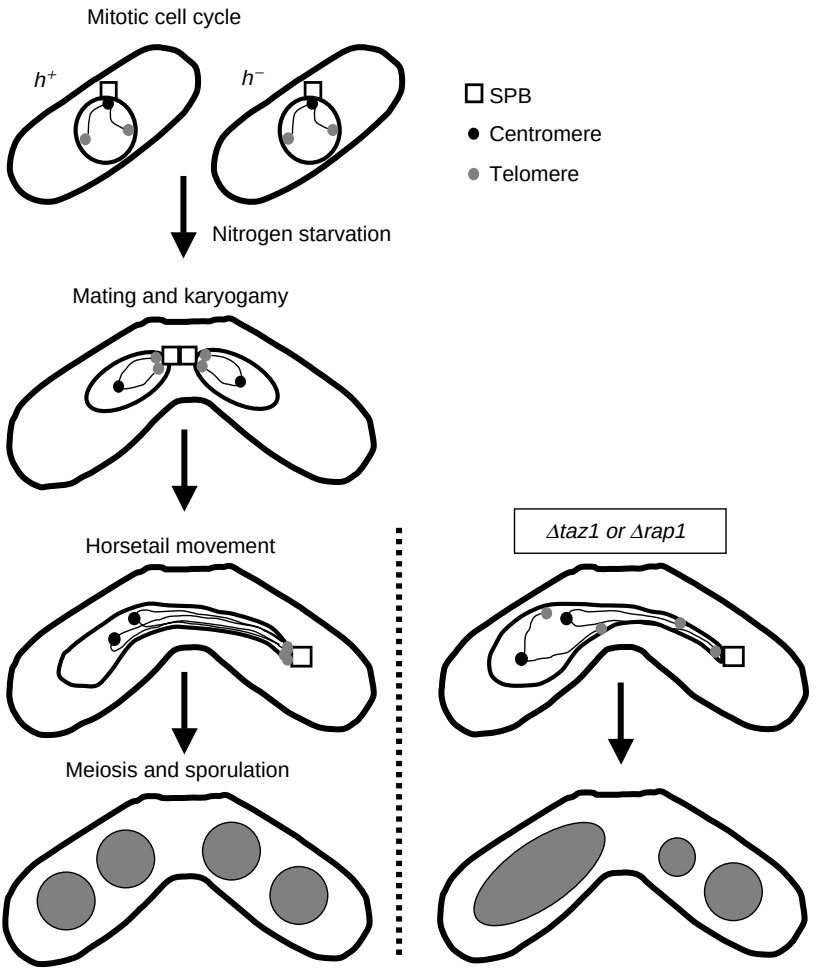

Fig. 4 The Taz1–spRap1 complex is required for normal meiosis. In vegetative cells, all the centromeres cluster near the SPB, whereas the telomeres cluster partially near the nuclear periphery (a single chromosome is shown in this figure). Upon nitrogen starvation, the h^+ and h^- cells mate and start karyogamy. At the same time, the telomeres cluster toward the SPB, whereas the centromeres are displaced from the SPB. After karyogamy, the nucleus moves extensively in the cell with the SPB as the leading edge (the horsetail movement). After meiosis is completed, four spores (shown as the gray circles) are produced in a zygote. In the *taz1* or *rap1* mutant, telomere clustering at the meiotic prophase is defective, and spores of abnormal number and size are produced.

is Rhp51 (fission yeast Rad51 homolog)-dependent, and Rhp51 is localized at the telomere especially in the absence of pKu, indicating that pKu sequesters Rhp51 from the telomere ends to inhibit homologous recombination. Furthermore, pKu protects the telomere end, and the lack of pKu leads to the acceleration of senescence in the absence of telomerase. Although the primary function of the pKu is to

mediate NHEJ, pKu at the telomeres is regulated not to fuse chromosome ends. Taz1 directly or indirectly prevents pKu from treating the telomeres as DSBs. In budding yeast, yKu70/yKu80 is involved in the telomere silencing and in the telomere localization to the nuclear periphery, as well as in the telomere length control. However, pKu is not involved in the telomere silencing, indicating that pKu and yKu function in differently at the telomeres.

Fission yeast Pot1 protein was identified by a BLAST search for a protein containing limited homology to the α-subunits of ciliate telomere proteins. The Pot1 protein specifically binds to the telomeric G-rich single-strand DNA. The *pot1* mutant loses the telomere DNA very quickly after germination, and most of the cells become elongated after ~10 generations. Most of the elongated cells show chromosome missegregation (see Fig. 5). However, some survivors emerge after ~75 generations, and all the three chromosomes are circularized in most of those cells. Thus, Pot1 directly protects the telomere at the end. Human Pot1 also specifically binds to the telomeric G-rich single-strand DNA, indicating that the function of Pot1 is conserved in fission yeast and humans. On the other hand, there is no apparent ortholog in budding yeast. However, the budding yeast Cdc13 protein binds to the telomeric G-rich single-strand DNA and regulates the accessibility of telomerase to the telomeres via interaction with the Est1 (ever shorter telomeres1) protein, which is associated with telomerase. In humans, hPot1 is recruited to the telomeres by interacting with PTOP/PIP1, which interacts with the TIN2-TRF1 complex at the telomeres. Thus, it is speculated that

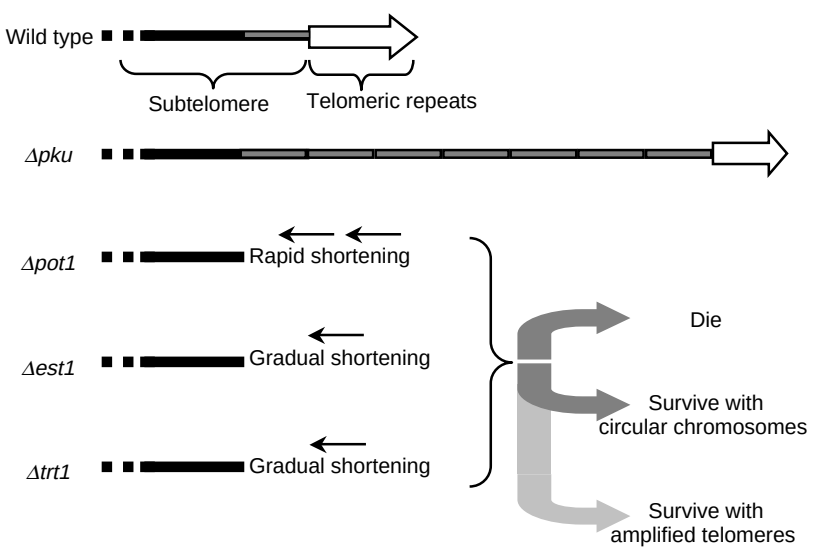

Fig. 5 Telomere length regulation by the telomere end–binding proteins. In the *pku* mutant, length of the telomeric repeats is slightly reduced, but telomere-proximal region of the TAS is amplified (gray boxes). In the *pot1*, *est1*, and *trt1* mutants, telomeres are shortened until cells die or survivors with circular chromosomes emerge. A very small population of the survivors has chromosomes with amplified telomere DNA.

the localization of hPot1 at the G-rich overhang is regulated by TRF1, which controls the accessibility of telomerase and the telomere length. The regulatory mechanism of fission yeast Pot1 is not known.

In budding yeast, scEst1 recruits or activates telomerase at the end of telomeres. In fission yeast, spEst1 was identified by a BLAST search as a protein that contains significant homology only at a small region near its amino terminus. Similar to scEst1, spEst1 physically interacts with telomerase, although it is not required for the catalytic activity of telomerase. The phenotype of the *spest1* mutant is very similar to that of the *trt1* mutant, that is, the telomere DNA is gradually shortened after germination (slower than the *pot1* mutant), and some survivors with circular chromosomes emerge. Thus, spEst1 may be required for the recruitment of telomerase to the telomere ends. Budding yeast Cdc13 associates with scEst1, so fission yeast Pot1 may interact with spEst1 to recruit telomerase.

2 Telomere Heterochromatin

2.1 Gene Silencing

Telomeres are specialized heterochromatin. Telomere silencing is alleviated in the mutants of many proteins. The first group contains Taz1 and spRap1, which are telomere-specific proteins. The second group contains Swi6, Clr4, and Rik1, which are required for heterochromatin formation at all the heterochromatin loci. Swi6 is the homolog of HP1s (heterochromatin protein1) in humans and fly. Clr4 is a methyltransferase specific for Lys9 of histone H3. The methylation of histone H3 at Lys9 is required for the binding of Swi6 to the histones. Rik1 interacts with Clr4 to act in concert to methylate histone H3 at Lys9 at the heterochromatin loci. Mutations in these *swi6*, *clr4*, and *rik1* genes do not cause any defects in telomere length.

2.2 Swi6

Indirect immunofluorescence microscopy and chromatin immunoprecipitation (ChIP) analyses have shown that Swi6 is localized at the heterochromatin loci, such as the centromeres, the telomeres, and the silent mating-type locus. Swi6 recruits cohesin molecules to the heterochromatin regions. Enriched cohesins at the pericentric region of the centromere are required for the precise chromosome segregation at mitosis. In fact, the *swi6* mutant shows the lagging chromosomes at a high frequency. At the silent mating-type locus, Swi6 is localized at the K-region, a ~15 kb chromosomal domain including the *mat2* and *mat3* loci. Swi6 heterochromatin is required for the maintenance of the silent epigenetic state of the K-region. Collectively, it seems that Swi6 has local functions at the heterochromatin regions. However, the function of Swi6 at the telomeres is not known. In budding yeast, Sir3, a component of the silencing complex, is localized at the telomeres and represses the expression of the stress-responsive genes of the *PAU* family located at the subtelomeres. When cells are in a stress condition, the Sir3 proteins are hyper-phosphorylated by the Mpk1 kinase and are removed from the telomeres, resulting in the derepression of the *PAU* family genes. No such stress-responsive genes at the fission yeast subtelomeres have been reported so far. However, there is an open reading frame

(ORF) of a putative RecQ helicase homolog at least on the four chromosome ends (Chr. I and II). The expression of the ORF is induced when the telomere DNA is shortened by the deletion of telomerase. The overexpression of the RecQ helicase homolog in the telomerase-deficient cells results in the earlier recovery from the senescence. Although the overexpression of the RecQ helicase homolog itself is apparently not toxic to the cells, Swi6 at the telomeres may be required for the repression of the RecQ helicase homolog.

2.3 RNAi and RITS

The pericentric outer (*otr*) region of the centromere consists of two types of repeat sequences, the *dg* and *dh* repeats. There exists a DNA sequence called *cenH* (centromere-homologous repeat) in the K-region at the silent mating-type locus. The *cenH* DNA sequence is highly homologous with those of the partial *dg* and *dh* repeats. Interestingly, the components of the fission yeast RNAi machinery, such as Argonaut (Ago1), Dicer (Dcr1), and RNA-dependent RNA polymerase (Rdp1), are involved in the establishment of Swi6 heterochromatin at the centromeres and at the silent mating-type locus. siRNAs are produced by Rdp1 and Dcr1 using the *dg* and/or *dh* repeats as a template, and then an RNAi effector complex RNAi-induced initiation of transcriptional gene silencing (RITS) containing Chp1, Tas3 and Ago1, which is associated with the

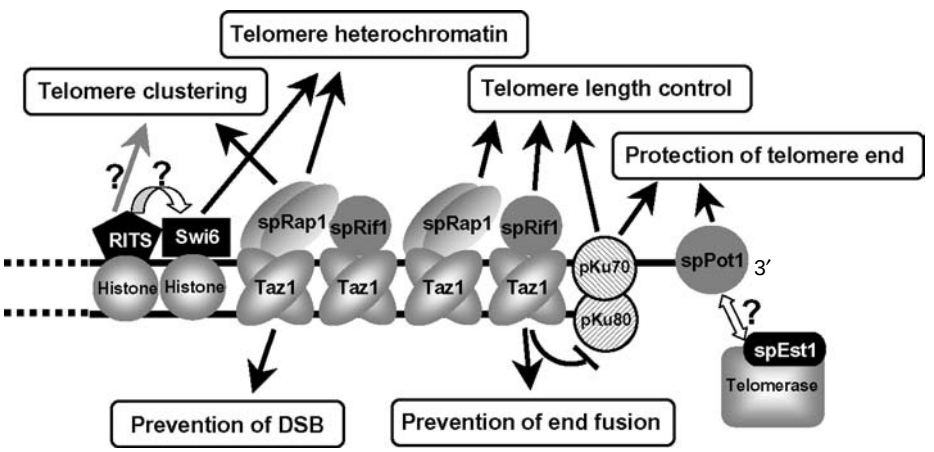

Fig. 6 Telomere-binding proteins in fission yeast. spRap1 and spRif1 are recruited to the telomeres by interacting with Taz1. spRap1 is involved in telomere length control, maintenance of the telomere heterochromatin and telomere clustering toward the SPB in meiotic prophase. spRif1 is involved in telomere length control. Taz1 prevents the Ku-dependent chromosome end fusion and DSBs throughout the genome. On the other hand, the pKu70/pKu80 heterodimer is involved in telomere length control and in the stability of TASs. Pot1 associates with the G-rich overhang and protects chromosome ends. Pot1 regulates the accessibility of telomerase possibly by recruiting spEst1 that associates with telomerase to the telomere ends. Swi6 (a HP1 homolog) and the RITS complex are localized at the telomeres in a methylation of histone H3 at Lys9-dependent manner. The RITS complex is possibly involved in the establishment of Swi6 heterochromatin at the telomeres and may be required for the telomere clustering.

siRNA, is recruited to the original target DNA sequence. It is thought that Clr4 is also recruited to the target sequence with the RITS complex to methylate histone H3, and then Swi6 is recruited to the methylated histone H3. Similar to Swi6, the RITS complex is also stably localized at the heterochromatin loci including the telomeres in a Clr4-dependent manner. Although how the RITS is recruited to the telomeres remains unknown, it is possibly recruited to the telomeres via the RNAi pathway. The RNAi mutant shows a defect in the telomere clustering both in the mitotic cell cycle and at the meiotic prophase, resulting in abnormal meiosis. Furthermore, the *clr4* and *rik1* mutant, in which the RITS complex is not localized at the heterochromatin loci, also show the defect in telomere clustering at the meiotic prophase. Therefore, some components of the RITS complex may be required for sticking the telomeres together and the telomere dynamics in the nucleus (see Fig. 6).

See also DNA Repair in Yeast; Heterochromatin and Eurochromatin – Organization, Packaging, and Gene Regulation; Telomerase.

Bibliography

Books and Reviews

Cooper, J.P. (2000) Telomere transitions in yeast: the end of the chromosome as we know it, *Curr. Opin. Genet. Dev.* **10**, 169–177.

Ferreira, M.G., Miller, K.M., Cooper, J.P. (2004) Indecent exposure: when telomeres become uncapped, *Mol. Cell* **13**, 7–18.

Grewal, S.I.S. (2000) Transcriptional silencing in fission yeast, *J. Cell. Phys.* **184**, 311–318.

Hiraoka, Y. (1998) Meiotic telomeres: a matchmaker for homologous chromosomes, *Genes Cells* **3**, 405–413.

Kanoh, J., Ishikawa, F. (2003) Composition and conservation of the telomeric complex, *Cell. Mol. Life Sci.* **60**, 2295–2302.

Smogorzewska, A., de Lange, T. (2004) Regulation of telomerase by telomeric proteins, *Annu. Rev. Biochem.* **73**, 177–208.

Primary Literature

Ai, W., Bertram, P.G., Tsang, C.K., Chan, T.-F., Zheng, X.F.S. (2002) Regulation of subtelomeric silencing during stress response, *Mol. Cell* **10**, 1295–1305.

Allshire, R.C., Nimmo, E.R., Ekwall, K., Javerzat, J.P., Cranston, G. (1995) Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation, *Genes Dev.* **9**, 218–233.

Baumann, P., Cech, T.R. (2000) Protection of telomeres by the Ku protein in fission yeast, *Mol. Biol. Cell* **11**, 3265–3275.

Baumann, P., Cech, T.R. (2001) Pot1, the putative telomere end-binding protein in fission yeast and humans, *Science* **292**, 1171–1175.

Beernink, H.T., Miller, K., Deshpande, A., Bucher, P., Cooper, J.P. (2003) Telomere maintenance in fission yeast requires an Est1 ortholog, *Curr. Biol.* **13**, 575–580.

Bernard, P., Maure, J.F., Partridge, J.F., Genier, S., Javerzat, J.P., Allshire, R.C. (2001) Requirement of heterochromatin for cohesion at centromeres, *Science* **294**, 2539–2542.

Bianchi, A., Smith, S., Chong, L., Elias, P., de Lange, T. (1997) TRF1 is a dimer and bends telomeric DNA, *EMBO J.* **16**, 1785–1794.

Bilaud, T., Brun, C., Ancelin, K., Koering, C.E., Laroche, T., Gilson, E. (1997) Telomeric localization of TRF2, a novel human telobox protein, *Nat. Genet.* **17**, 236–239.

Boulton, S.J., Jackson, S.P. (1998) Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing, *EMBO J.* **17**, 1819–1828.

Broccoli, D., Smogorzewska, A., Chong, L., de Lange, T. (1997) Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2, *Nat. Genet.* **17**, 231–235.

Buchman, A.R., Lue, N.F., Kornberg, R.D. (1988) Connections between transcriptional

activators, silencers, and telomeres as revealed by functional analysis of a yeast DNA-binding protein, *Mol. Cell. Biol.* **8**, 5086–5099.

Chikashige, Y., Hiraoka, Y. (2001) Telomere binding of the Rap1 protein is required for meiosis in fission yeast, *Curr. Biol.* **11**, 1618–1623.

Chikashige, Y., Ding, D., Funabiki, H., Haraguchi, T., Mashiko, S., Yanagida, M., Hiraoka, Y. (1994) Telomere-led premeiotic chromosome movement in fission yeast, *Science* **264**, 270–273.

Chikashige, Y., Kinoshita, N., Nakaseko, Y., Matsumoto, T., Murakami, S., Niwa, O., Yanagida, M. (1989) Composite motifs and repeat symmetry in S. pombe centromeres: direct analysis by integration of Not1 restriction sites, *Cell* **57**, 739–751.

Chong, L., van Steensel, B., Broccoli, D., Erdjument-Bromage, H., Hanish, J., Tempst, P., de Lange, T. (1995) A human telomeric protein, *Science* **270**, 1663–1667.

Cockell, M., Palladino, F., Laroche, T., Kyrion, G., Liu, C., Lustig, A.J., Gasser, S.M. (1995) The carboxy termini of Sir4 and Rap1 affect Sir3 localization: evidence for a multicomponent complex required for yeast telomeric silencing, *J. Cell Biol.* **129**, 909–924.

Conrad, M.N., Wright, J.H., Wolf, A.J., Zakian, V.A. (1990) RAP1 protein interacts with yeast telomeres in vivo: overproduction alters telomere structure and decreases chromosome stability, *Cell* **63**, 739–750.

Cooper, J.P., Watanabe, Y., Nurse, P. (1998) Fission yeast Taz1 protein is required for meiotic telomere clustering and recombination, *Nature* **392**, 828–831.

Cooper, J.P., Nimmo, E.R., Allshire, R.C., Cech, T.R. (1997) Regulation of telomere length and function by a Myb-domain protein in fission yeast, *Nature* **385**, 744–747.

Ekwall, K., Javerzat, J.P., Lorentz, A., Schmidt, H., Cranston, G., Allshire, R. (1995) The chromodomain protein Swi6: a key component at fission yeast centromeres, *Science* **269**, 1429–1431.

Ekwall, K., Nimmo, E.R., Javerzat, J.P., Borgstrom, B., Egel, R., Cranston, G., Allshire, R. (1996) Mutations in the fission yeast silencing factors *clr4*$^+$ and *rik1*$^+$ disrupt the localisation of the chromo domain protein Swi6p and impair centromere function, *J. Cell Sci.* **109**, 2637–2648.

Evans, S.K., Lundblad, V. (1999) Est1 and Cdc13 as comediators of telomerase access, *Science* **286**, 117–120.

Fairall, L., Chapman, L., Moss, H., de Lange, T., Rhodes, D. (2001) Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2, *Mol. Cell* **8**, 351–361.

Ferreira, M.G., Cooper, J.P. (2001) The fission yeast Taz1 protein protects chromosomes from Ku-dependent end-to-end fusions, *Mol. Cell* **7**, 55–63.

Funabiki, H., Hagan, I., Uzawa, S., Yanagida, M. (1993) Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast, *J. Cell Biol.* **121**, 961–976.

Galy, V., Olivo-Marin, J.C., Scherthan, H., Doye, V., Rascalou, N., Nehrbass, U. (2000) Nuclear pore complexes in the organization of silent telomeric chromatin, *Nature* **403**, 108–112.

Gravel, S., Larrivee, M., Labrecque, P., Wellinger, R.J. (1998) Yeast Ku as a regulator of chromosomal DNA end structure, *Science* **280**, 741–744.

Grewal, S.I.S., Klar, A.J.S. (1997) A recombinationally repressed region between *mat2* and *mat3* loci shares homology to centromeric repeats and regulates directionality of mating-type switching in fission yeast, *Genetics* **146**, 1221–1238.

Hall, I.M., Noma, K., Grewal, S.I.S. (2003) RNAi machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 193–198.

Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A., Grewal, S.I.S. (2002) Establishment and maintenance of a heterochromatin domain, *Science* **297**, 2232–2237.

Hardy, C.F.J., Sussel, L., Shore, D. (1992) A RAP1-interacting protein involved in transcriptional silencing and telomere length regulation, *Genes Dev.* **6**, 801–814.

Hecht, A., Strahl-Bolsinger, S., Grunstein, M. (1996) Spreading of transcriptional repressor SIR3 from telomeric heterochromatin, *Nature* **383**, 92–96.

Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S.M., Grunstein, M. (1995) Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast, *Cell* **80**, 583–592.

Hiraoka, Y., Henderson, E., Blackburn, E.H. (1998) Not so peculiar: fission yeast telomere repeats, *Trends Biochem. Sci.* **23**, 126.

Hoppe, G.J., Tanny, J.C., Rudner, A.D., Gerber, S.A., Danaie, S., Gygi, S., Moazed, D. (2002) Steps in assembly of silent chromatin in yeast: Sir3-independent binding of a Sir2/Sir4 complex to silencers and role for Sir2-dependent deacetylation, *Mol. Cell. Biol.* **22**, 4167–4180.

Huet, J., Cottrell, P., Cool, M., Vignais, M.L., Thiele, D., Marck, C., Buhler, J.M., Sentenac, A., Fromageot, P. (1985) A general upstream-binding factor for genes of the yeast translation apparatus, *EMBO J.* **4**, 3539–3547.

Kanoh, J., Ishikawa, F. (2001) spRap1 and spRif1, recruited to telomeres by Taz1, are essential for telomere function in fission yeast, *Curr. Biol.* **11**, 1624–1630.

Kibe, T., Tomita, K., Matsuura, A., Izawa, D., Kodaira, T., Ushimaru, T., Uritani, M., Ueno, M. (2003) Fission yeast Rhp51 is required for the maintenance of telomere structure in the absence of the Ku heterodimer, *Nucleic Acids Res.* **31**, 5054–5063.

Kyrion, G., Boakye, K.A., Lustig, A.J. (1992) C-terminal truncation of RAP1 results in the deregulation of telomere size, stability, and function in *Saccharomyces cerevisiae*, *Mol. Cell. Biol.* **12**, 5159–5173.

Laroche, T., Martin, S.G., Gotta, M., Gorham, H.C., Pryde, F.E., Louis, E.J., Gasser, S.M. (1998) Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres, *Curr. Biol.* **8**, 653–656.

Li, B., Oestreich, S., de Lange, T. (2000) Identification of human Rap1: Implications for telomere evolution, *Cell* **101**, 471–483.

Lingner, J., Cooper, J.P., Cech, T.R. (1995) Telomerase and DNA end replication: No longer a lagging strand problem?, *Science* **269**, 1533–1534.

Liu, D., Safari, A., O'Connor, M.S., Chan, D.W., Laegeler, A., Qin, J., Songyang, Z. (2004) PTOP interacts with POT1 and regulates its localization to telomeres, *Nat. Cell Biol.* **6**, 673–680.

Loayza, D., de Lange, T. (2003) POT1 as a terminal transducer of TRF1 telomere length control, *Nature* **423**, 1013–1018.

Lundblad, V., Szostak, J.W. (1989) A mutant with a defect in telomere elongation leads to senescence in yeast, *Cell* **57**, 633–643.

Lustig, A.J., Kurtz, S., Shore, D. (1990) Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length, *Science* **250**, 549–553.

Lustig, A.J., Liu, C., Zhang, C., Hanish, J.P. (1996) Tethered Sir3p nucleates silencing at telomeres and internal loci in *Saccharomyces cerevisiae*, *Mol. Cell. Biol.* **16**, 2483–2495.

Mandell, J.G., Goodrich, K.J., Bahler, J., Cech, T.R. (2005) Expression of a RecQ helicase homolog affects progression through crisis in fission yeast lacking telomerase, *J. Biol. Chem.* **280**, 5249–5257.

Miller, K.M., Cooper, J.P. (2003) The telomere protein Taz1 is required to prevent and repair genomic DNA breaks, *Mol. Cell* **11**, 303–313.

Miyoshi, T., Sadaie, M., Kanoh, J., Ishikawa, F. (2003) Telomeric DNA ends are essential for the localization of Ku at telomeres in fission yeast, *J. Biol. Chem.* **278**, 1924–1931.

Moretti, P., Freeman, K., Coodly, L., Shore, D. (1994) Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1, *Genes Dev.* **8**, 2257–2269.

Nakamura, T.M., Moser, B.A., Russell, P. (2002) Telomere binding of checkpoint sensor and DNA repair proteins contributes to maintenance of functional fission yeast telomeres, *Genetics* **161**, 1437–1452.

Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., Harley, C.B., Cech, T.R. (1997) Telomerase catalytic subunit homologs from fission yeast and human, *Science* **277**, 955–959.

Nakayama, J., Klar, A.J., Grewal, S.I.S. (2000) A chromodomain protein, Swi6, performs imprinting functions in fission yeast during mitosis and meiosis, *Cell* **101**, 307–317.

Nakayama, J., Allshire, R.C., Klar, A.J.S., Grewal, S.I.S. (2001a) A role for DNA polymerase α in epigenetic control of transcriptional silencing in fission yeast, *EMBO J.* **20**, 2857–2866.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., Grewal, S.I.S. (2001b) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly, *Science* **292**, 110–113.

Nimmo, E.R., Pidoux, A.L., Perry, P.E., Allshire, R.C. (1998) Defective meiosis in telomere-silencing mutants of *Schizosaccharomyces pombe*, *Nature* **392**, 825–828.

Niwa, O., Shimanuki, M., Miki, F. (2000) Telomere-led bouquet formation facilitates homologous chromosome pairing and restricts ectopic interaction in fission yeast, *EMBO J.* **19**, 3831–3840.

Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D., Grewal, S.I.S. (2004) RITS acts in *cis* to promote RNA-interference-mediated transcriptional and post-transcriptional silencing, *Nat. Genet.* **36**, 1174–1180.

Nonaka, N., Kitajima, T., Yokobayashi, S., Xiao, G., Yamamoto, M., Grewal, S.I., Watanabe, Y. (2002) Recruitment of cohesin to heterochromatic region by Swi6/HP1 in fission yeast, *Nat. Cell Biol.* **4**, 89–93.

Nugent, C.I., Hughes, T.R., Lue, N.F., Lundblad, V. (1996) Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance, *Science* **274**, 249–252.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., Nishimura, Y. (1994) Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices, *Cell* **79**, 639–648.

Polotnianka, R.M., Li, J., Lustig, A.J. (1998) The yeast Ku heterodimer is essential for protection of the telomere against nucleolytic and recombinational activities, *Curr. Biol.* **8**, 831–834.

Sadaie, M., Naito, T., Ishikawa, F. (2003) Stable inheritance of telomere chromatin structure and function in the absence of telomeric repeats, *Genes Dev.* **17**, 2271–2282.

Sadaie, M., Iida, T., Urano, T., Nakayama, J. (2004) A chromodomain protein, Chp1, is required for the establishment of heterochromatin in fission yeast, *EMBO J.* **23**, 3825–3835.

Scherthan, H., Bahler, J., Kohli, J. (1994) Dynamics of chromosome organization and pairing during meiotic prophase in fission yeast, *J. Cell Biol.* **127**, 273–285.

Shore, D., Nasmyth, K. (1987) Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements, *Cell* **51**, 721–732.

Silverman, J., Takai, H., Buonomo, S.B.C., Eisenhaber, F., de Lange, T. (2004) Human Rif1, ortholog of a yeast telomeric protein, is regulated by ATM and 53BP1 and functions in the S-phase checkpoint, *Genes Dev.* **18**, 2108–2119.

Smith, G.C., Jackson, S.P. (1999) The DNA-dependent protein kinase, *Genes Dev.* **13**, 916–934.

Smith, C.D., Smith, D.L., DeRisi, J.L., Blackburn, E.H. (2003) Telomeric protein distributions and remodeling through the cell cycle in *Saccharomyces cerevisiae*, *Mol. Biol. Cell* **14**, 556–570.

Spink, K.G., Evans, R.J., Chambers, A. (2000) Sequence-specific binding of Taz1 dimers to fission yeast telomeric DNA, *Nucleic Acids Res.* **28**, 527–533.

Strahl-Bolsinger, S., Hecht, A., Luo, K., Grunstein, M. (1997) SIR2 and SIR4 interactions differ in core and extended telomeric heterochromatin in yeast, *Genes Dev.* **11**, 83–93.

Sussel, L., Shore, D. (1991) Separation of transcriptional activation and silencing functions of the RAP1-encoded repressor/activator protein 1: isolation of viable mutants affecting both silencing and telomere length, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7749–7753.

Taggart, A.K.P., Teng, S.-C., Zakian, V.A. (2002) Est1p as a cell cycle-regulated activator of telomere-bound telomerase, *Science* **297**, 1023–1026.

Tuzon, C.T., Borgstrom, B., Weilguny, D., Egel, R., Cooper, J.P., Nielsen, O. (2004) The fission yeast heterochromatin protein Rik1 is required for telomere clustering during meiosis, *J. Cell Biol.* **165**, 759–765.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I.S., Moazed, D. (2004) RNAi-mediated targeting of heterochromatin by the RITS complex, *Science* **303**, 672–676.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., Martienssen, R.A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi, *Science* **297**, 1833–1837.

Wotton, D., Shore, D. (1997) A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in *Saccharomyces cerevisiae*, *Genes Dev.* **11**, 748–760.

Ye, J.Z.-S., Hockemeyer, D., Krutchinsky, A.N., Loayza, D., Hooper, S.M., Chait, B.T., de Lange, T. (2004) POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex, *Genes Dev.* **18**, 1649–1654.

Temporal Adjustment of Physiology: *see* Rhythmic and Temporal Processes in Biology

Testis Development: *see* Male Reproductive System: Testis Development and Spermatogenesis

Testosterone Analogs and Antagonists: *see* Sex Hormones (Male): Analogs and Antagonists

Thale Cress (*Arabidopsis thaliana*) Genome

Mark D. Curtis and Ueli Grossniklaus
University of Zürich, Zürich, Switzerland

1 Introduction 251
1.1 Model Organisms 252

2 Genome Structure 253
2.1 Linkage Groups 253
2.2 Genome Sequence 253
2.3 Chromsome Organization 254
2.3.1 Telomeres 254
2.3.2 Centromeres 254
2.3.3 Nucleolar Organizers (NORs) 255
2.4 Synteny 256
2.5 Gene Families 257
2.6 Genome Plasticity 257

3 Genome Function 258
3.1 Loss-of-function 258
3.2 Classical Mutants 259
3.3 Insertional Mutants 259
3.3.1 T-DNA 259
3.3.2 Transposons 260
3.3.3 Enhancer Detection and Gene Trapping 260
3.4 Targeted Gene Disruptions 261
3.5 Microarrays 262
3.6 Gain-of-function Mutations 262
3.6.1 Constitutive Ectopic Expression 263
3.6.2 Inducible Ectopic Expression 265

4 Gene Regulation 266
4.1 Transcription Factors 266

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

4.2 Epigenetics 267
4.3 Cellular Communication 269

5 Conclusions and Future Prospects 271

Acknowledgments 272

Bibliography 272
Books and Reviews 272
Primary Literature 273

Keywords

Abaxial
Dorsal, that is, facing away from the axis of an organ or organism; the lower side of a leaf is termed *abaxial*.

Accession
Plant genotype of a species collected at a specific location.

Acetylation
Introduction of an acetyl group (CH_3CO) into a compound.

Adaxial
Ventral, that is, nearest to or facing toward the axis of an organ or organism; the upper side of a leaf is termed *adaxial*.

ADP-ribosylation
Transfer of adenosine diphosphate ribose to proteins.

Angiosperms
Plants having seeds in a closed ovary.

Apomict
A plant that reproduces asexually through seed.

Bacterial Artificial Chromosomes (BACs)
A DNA-cloning vector capable of incorporating fragments of 100 to 300 kb (average of 150 kb) in *Escherichia coli* cells.

Centromere
The region of a chromosome required for chromosome segregation, usually repetitive sequences, to which the spindle attaches during mitosis and meiosis.

Cleaved Amplified Polymorphic Sequences (CAPS)
Sequences that show polymorphism between individuals when a particular locus is amplified, and the product digested with restriction enzymes, resulting in different sized restriction fragments.

Colinearity
Conservation of gene order in chromosomes over distantly related species.

Dicots
Short for dicotyledonae-an angiosperm with two embryonic leaves (cotyledons).

Dioecious Plants
Having male and female reproductive organs on separate plants.

Diploid
Containing paired chromosomes (2 × the haploid number).

DNA Polymorphism
A difference in the DNA sequence at a particular locus between individuals.

Ectopic Expression
Overexpression or expression in addition to endogenous gene expression.

Epigenetics
The study of heritable changes in gene expression that occur without a change in DNA sequence.

Euchromatin
Uncondensed, transcriptionally active chromatin.

Exon
Protein coding part of a gene.

Expressed Sequenced Tags
A sequence, derived from a cDNA library, that has been expressed at some stage of development, and that is used to identify the gene locus.

Forward Genetics
A classical genetic approach, where a gene is defined by a mutant phenotype.

Gametophyte
The gamete-producing organism or phase in the life cycle of a plant: the megagametophyte is the female gametophyte (embryo sac), containing the female gametes (egg and central cell); the microgametophyte is the male gametophyte (pollen), containing two sperm cells.

Genetic Map
The linear positioning of genetic markers (providing a visible phenotype) within a chromosome, or genome, as determined by their recombination frequency.

Haploid
Containing one set of unpaired chromosomes.

Hermaphroditic Plants
Having both male and female reproductive organs in the same flower.

Heterochromatin
Condensed, transcriptionally inactive, chromatin.

Hexaploid
Containing six sets of paired chromosomes.

Histone
Main protein of chromatin.

Homeotic Genes
A class of genes that when mutated lead to homeotic transformations. Homeotic genes play a central role in animals and plants in pattern formation and the specification of organ identity.

Intron
Nonprotein coding part of a gene.

Linkage Groups
Genes that tend to be transmitted together.

Locus
The position of a gene on the chromosome.

Methylation
Addition of a methyl group (CH_3) to a substrate.

Monocots
Short for monocotyledonae-an angiosperm with one embryonic leaf (cotyledon).

Monoecious Plants
Having separate male and female flowers on the same plant.

Morphological
Concerned with the visible structure of an organism.

Nucleosome
A unit made up of a histone (H) protein complex, wrapped in two turns of DNA.

Orthologous Genes
Genes that have diverged after a speciation event.

ParalogousGenes
Genes that have diverged after a duplication event.

Pericentromeric
Pertaining to the region of repetitive DNA at the periphery of the centromere.

Phosphorylation
The addition of a phosphate (PO_4) group to a protein.

Phylogenetic Relationship
The relationship of species or a taxonomic group of organisms, as derived from the comparison of characteristics such as DNA sequence or other characters.

Physical Map
The linear positioning of physical markers, or identifiable landmarks on DNA, such as restriction enzyme cleavage sites within a chromosome or genome, as determined by their recombination frequency. (The highest resolution physical map is the nucleotide sequence of the genome.)

Polyploid
A nucleus having more than one set of paired chromosomes.

Rapid Amplified Polymorphic DNA
Variation in the size of DNA fragments that arise from the amplification of DNA, from different individuals, using short nucleotide primers. This variation arises from DNA sequence polymorphisms that affect the efficiency with which the primer binds to the DNA sites.

Recombination
Combining genes or characters from parents so that the progeny contains a combination that is not the same as in either parent.

Retroelements
Genetic elements that transpose replicatively, leaving a copy of itself at the launch site and reinserting a new copy in the genome.

Reverse Genetics
A strategy that is used to identify the function of a gene, the sequence of which has been identified, by screening for mutations in that gene (using gene targeting, or the recovery of EMS, or insertional mutants).

Satellite Repeat
Highly repetitive eukaryotic DNA, often located around centromeres.

Simple Sequence Length Polymorphisms (SSLP)
Simple nucleotide repeat sequences that are highly variable in length between individuals.

Sumoylation
A protein modification, biochemically analogous to, but functionally distinct from, ubiquitination. Sumoylation involves the covalent attachment of a SUMO (small ubiquitin-related modifier) to substrate proteins.

Synteny
Colinearity between species of wide evolutionary distance.

Telomeres
Highly repetitive DNA, capped by specific proteins, that protect the ends of chromosomes.

Tiling Path
A minimum set of overlapping clones, usually BACs or YACs containing genomic DNA that make a contiguous sequence of genomic DNA that is anchored to the combined physical and genetic map.

Transformation
A permanent heritable modification of a cell by the uptake and incorporation of exogenous DNA.

Transposons
Sequences of DNA that can move, or transpose, around the genome.

Ubiquitination
The modification of a protein by addition of a ubiquitin molecule to lysine residues that targets the protein for destruction. Ubiquitination of histones leads to a conformational change in chromatin structure.

Vernalization
Prolonged cold treatment of plants that induces flowering.

Yeast Artificial Chromosomes (YACs)
A DNA cloning vector based on a bacterial plasmid that contains a yeast centromeric region (CEN), a yeast origin of replication (ARS), and a telomere region at the end of each arm. YACs are capable of containing more than 1 megabases of cloned DNA.

The genome of the Thale Cress, *Arabidopsis thaliana*, was the first plant genome to be fully sequenced, providing a tool for the rapid characterization of biological processes. This versatile model organism has had a major impact on our understanding of the physiology, growth, and development of flowering plants, and their response to environmental stress. An in-depth understanding of model organisms, coupled with knowledge of crop systems, will provide rapid advancements in technology, including genetic engineering, that are required to overcome future predicted global climate changes and short falls in food production.

1 Introduction

The importance and impact of plants on our planet is irrefutable, they are essential for its well-being. Our planet's biomass is mostly made up of plants, a characteristic illustrated by images of earth from space.

Plants are highly adaptive organisms, surviving in diverse and often severe environments, from deserts to marshlands. They combat disease and predation. They adjust their behavior to seasonal and diurnal changes, moving to avoid shade, or in anticipation of the rising sun. Some can even count; the Venus flytrap ignores a single stimulus, but snaps shut in response to a second one. Plants have evolved a wide variety of reproductive strategies. They reproduce sexually as hermaphroditic, monoecious, or dioecious species, or asexually by apomixis or through tubers or stolons. They include some of the largest and oldest living organisms on the planet. These features alone provide compelling reasons for plant study; however, coupled with our dependence on these organisms for food, medicines, clothing, and building materials, a deeper understanding of how these organisms function is essential to our future well-being.

The wide variety of exploited plant products has largely been dependent on the existence of a large reservoir of diverse plant species. This diversity has, in many ways, hampered progress in the functional analysis of plants, since many plant species of economic importance are also diverse. In animals, it has long been accepted that model organisms would provide acceptable alternatives to human experimentation. In plants, however, no widely accepted model plant species received early adoption. Historically, owing to plant diversity, it was felt that no single plant organism could provide universally applicable information of use to other plant species. This led to a preference for studies in economically important and frequently diverse plant species, which resulted in slow progress. Major drawbacks of many of the economically important crop species, which have made genetic and molecular studies difficult, are the large size and highly repetitive nature of their polyploid genomes. Bread wheat, for example, is a hexaploid with a genome of 1.6×10^{10} bp, 80% of which consists of repetitive DNA sequences. Although repetitive DNA is also a feature of animal genomes, including the human genome, polyploidy is rarely seen in animals; it was thought to occur only in a few insects, amphibians, and reptiles until 1999, when it was discovered in one mammal; a desert rodent in Argentina.

1.1 Model Organisms

Model diploid organisms in animal studies revealed that, despite the divergence of invertebrates and vertebrates more than 600 million years ago, many genes, including the *HOX* homeotic gene family, are functionally interchangeable between diverse organisms, for example, flies and vertebrates. This has allowed remarkable parallels to be drawn between the fruit fly model and human genetics. Since angiosperms are thought to have evolved from a common ancestor within the last 150 million years, genes identified in a model angiosperm could be expected, with confidence, to functionally replace an orthologous gene in many of the 250 000 angiosperm species.

In 1943, before these discoveries, Friedrich Laibach, realizing the benefit of adopting a model plant species, suggested the Thale Cress, *Arabidopsis thaliana*, as a model plant system (Fig. 1). He recognized that *Arabidopsis* had many features that would make it an ideal model organism for plant studies. It has a short life cycle, of between 6 to 8 weeks, it is physically small (20–30 cm tall at maturity), it can be self-fertilized and outcrossed, producing large quantities of seed (up to 6000 seeds per plant), and it is a genetically tractable diploid. Despite these many

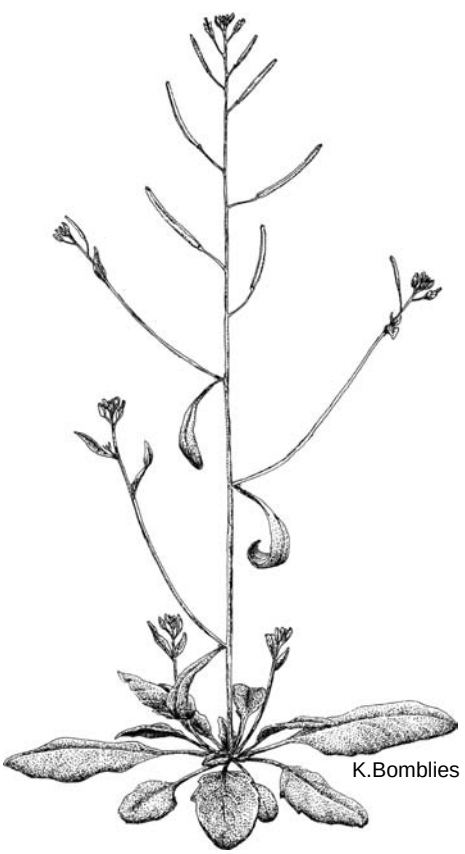

Fig. 1 A drawing of a mature flowering wild-type *Arabidopsis thaliana* plant (courtesy of Kirsten Bomblies, University of Wisconsin, Madison).

advantages for classical genetic studies, it was not until the 1960s that enthusiasm for *Arabidopsis* really developed. In addition to these earlier identified features, it was soon discovered that *Arabidopsis* could be mutated easily, and contained a small number of chromosomes (a haploid number of five chromosomes), as revealed by microscopic observations, and later by linkage analysis. In fact, it is this low number of chromosomes that has allowed the efficient identification of linkage groups in *Arabidopsis*.

2 Genome Structure

When confronted with an uncharacterized genome, isolates or accessions with different morphological traits can be used in classical genetic studies to determine the physical relationship (linkage) of the genes responsible for these traits.

2.1 Linkage Groups

Linkage analysis is dependent on recombination between different loci. These different loci (morphological or classical markers) are generated most easily by mutagenesis, and their position with respect to one another in the genome can be identified by their frequency of recombination. This type of analysis led to the production of an extensive genetic map. The interest in *Arabidopsis* as a model organism really accelerated when, in the early 1980s, the small size and simple structure of the genome was determined and high-efficiency transformation protocols were established.

Together with classical morphological markers, linkage analysis was performed with molecular markers, taking advantage of another important feature of *Arabidopsis*: the high degree of polymorphism at the DNA level between accessions. Restriction fragment length polymorphisms (RFLPs) and later, to a lesser extent, PCR-based markers, such as rapid amplified polymorphic DNAs (RAPDs), have been used to construct linkage maps in *Arabidopsis*. This type of analysis led to the production of a physical map, which was later integrated with the genetic map. These markers, together with the development of more reliable PCR-based markers, CAPS (cleaved amplified polymorphic sequences) and SSLP (simple sequence length polymorphisms), were used to construct a high-resolution combined physical and genetic map. The small size of the genome and this extensive map made *Arabidopsis* an attractive proposition for whole genome sequencing. The sequencing program began with a collaborative international initiative in 1996. To facilitate this sequencing effort, a minimal tiling path of large-insert yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs) was assembled, spanning the entire genome. These BAC and YAC contigs, anchored to the combined physical and genetic map, provided a framework that allowed the *Arabidopsis* genome to be sequenced. As a consequence of this enormous international effort, the *Arabidopsis* genome was the first reference plant genome to be sequenced.

2.2 Genome Sequence

The original analysis of the sequence revealed 25 498 genes in about 125 Mb of DNA (a figure, which is slowly increasing as methods of gene prediction improve, currently >28 000). In *Arabidopsis*, the

average gene length is around 2000 bp with an average exon length of 250 bp (~5 per gene) and an average intron of 180 bp in length. This means that *Arabidopsis* genes are mainly exonic. The rice genome is larger, containing 430 Mb of DNA packaged into 12 chromosomes. The rice genome sequence analysis has revealed a total of approximately 50 000 genes. These genes are similar in length to *Arabidopsis* genes at around 2350 bp, with an average exon length of 260 bp (~4 per gene) and an average intron length of 330 bp. Here, the genes are nearly equally exonic and intronic. This is in stark contrast to the human genome, which has a similar number of genes as found in *Arabidopsis* (Celera suggests 26 383 genes), but, here, the genes are mainly intronic and are contained in some 3000 Mb of DNA. (The average length of a human gene is 27 000 bp [ranging between 20 – 50 kb], the average exon length is 236 bp [~5 per gene], and the average intron is 5500 bp.)

2.3 Chromsome Organization

The *Arabidopsis* genome sequence provides information on chromosomal organization, including telomeres, centromeres, and nucleolar organizers that maintain genome stability even during evolutionary changes in genome structure.

2.3.1 Telomeres

Arabidopsis chromosomes, like all eukaryotic chromosomes, are capped by complex nucleoprotein structures known as *telomeres*, which stabilize the genome (Fig. 2b). The length of telomeres varies among eukaryotes, for example, in tobacco, telomeres are 150 kb, while in humans, telomeres are less than 20 kb. *Arabidopsis* telomeres were the first higher eukaryote telomeric DNA sequences to be determined. They are composed of TTTAGGG repeats, averaging 2 to 3 kb in length with G-overhangs of 20 to 30 nucleotides. Telomeric DNA is usually maintained by telomerase enzymes. Mutations in telomerase cause telomere shortening and ultimately cell death. Telomerase-deficient *Arabidopsis* plants show a remarkable tolerance to genome stress, showing no visible abnormalities even after five generations of inbreeding. After further inbreeding, telomerase-deficient mutants, however, show end-to-end chromosome fusions, which result from uncapped chromosome ends being recognized as damaged DNA. Varying degrees of developmental abnormalities follow, ultimately resulting in lethality by the 10th generation. Other mutations in *Arabidopsis* also produce dysfunctional telomeres; however, the slow onset of lethality makes *Arabidopsis* an excellent model for telomere study. Interestingly, although telomeres play a role in genome stability, the *Arabidopsis* genome sequence has revealed an unexpected degree of large segmental duplications and rearrangements. More than 60% of this small genome consists of duplicated DNA, suggesting that plant genomes are in a constant state of flux.

2.3.2 Centromeres

One of the most important domains of eukaryotic chromosomal DNA is the centromere (Fig. 2b). The centromeric regions are required for normal chromosome segregation during meiosis and mitosis. They usually consist of highly repetitive DNA that mediates the attachment of the spindle to the chromosomes during cell division. Despite their common function, centromeric DNA sequences are not conserved between organisms. *Arabidopsis* centromeres contain repetitive DNA

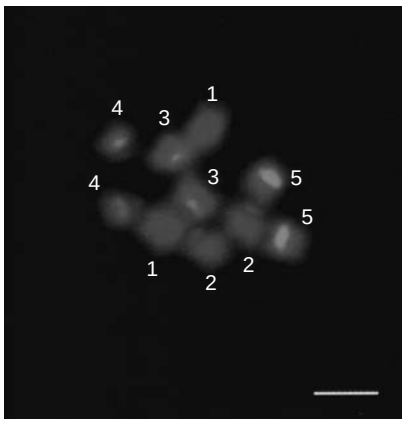

(a)

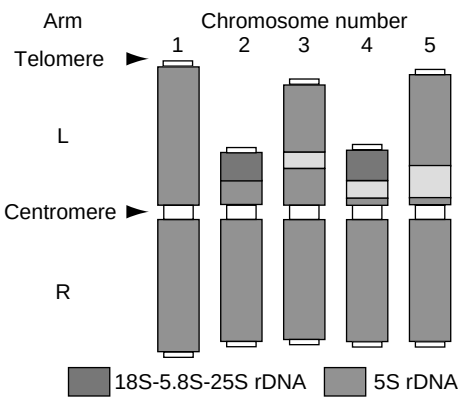

(b)

Fig. 2 (a) DAPI-stained metaphase chromosomes of *Arabidopsis thaliana* Landsberg *erecta*, showing green fluorescein signals from the 5S rDNA and red Cy-3 signals from the 18S-5.8S-25S rDNA. Chromosomes were identified by the FISH pattern and numbered according to the linkage group. Bar = 2 µm. (b) A karyotype showing the location of the 5S rDNA18S-5.8S-25S rDNA, centromeric 180 bp repeats and the telomeric repeats (Murata, M., Heslop-Harrison, J.S., Motoyoshi, F. (1997) Physical mapping of the 5S ribosomal RNA genes in *Arabidopsis thaliana* by multi-color fluorescence *in situ* hybridization with cosmid clones, *Plant J.* **12**, 31–37.) This figure is copyrighted by Blackwell Publishing and is reprinted with permission. (See color plate p. xxiv.)

elements, including telomere-like repeats, transposons, and retroelements; however, the most abundant sequence (several megabases) is a 180-bp satellite repeat, which forms long tandem arrays. The *Arabidopsis* genome contains five centromere domains, one for each chromosome. The sequence data corresponding to these regions are incomplete because of the inherent difficulties in sequencing large regions of repetitive DNA. It is apparent, however, that in *Arabidopsis,* as in maize and rice, the centromeric satellite repeat is a key functional element of centromere DNA. These satellite sequences associate with centromere-specific histone H3 variants (CenH3s). CenH3s are rapidly evolving proteins, suggesting that they respond to changes in rapidly evolving centromeric repeat sequences.

2.3.3 Nucleolar Organizers (NORs)

Like all organisms, *Arabidopsis* contains ribosomal DNA (rDNA), encoding clustered arrays of the 18S, 5.8S, and 25S ribosomal RNA (*rRNA*) genes, which, in concert with 5S RNA, form the structural core of cytoplasmic ribosomes (the sites of protein manufacture). These clusters of active *rRNA* genes are known as *nucleolar*

organizers (NORs) since, during interphase, the nucleolus forms at the site of rRNA transcription and ribosome assembly. In *Arabidopsis,* these NORs are roughly 3.6 to 3.7 Mbp, comprising 360 to 370 *rRNA* genes, and are situated adjacent to the telomeres on the short arm of chromosomes 2 and 4 (Fig. 2a and b). In contrast, the 5S *rRNA* genes are localized in clusters on chromosomes 3, 4, and 5, toward the centromeres (Fig. 2a and b). Coding regions of *rRNA* genes are highly conserved, which allows the phylogeny of diverse organisms to be established by comparative DNA sequence analysis. The intergenic spacer regions containing control elements, however, are more variable. The variants differ in the number of repetitive sequences located in the intergenic spacer region, suggesting that the *rRNA* genes evolve together as a group (concerted evolution). Comparing intergenic space sequences allows the phylogeny of very closely related species, or even isolates of the same species, to be established.

A powerful approach to look at the structure of chromosomes is to use fluorescence *in situ* hybridization (FISH) with chromosome-specific DNA probes. Using this chromosome painting technique, the physical position of the 5S ribosomal RNA genes, described earlier, was identified (Fig. 2a and b). Using a set of FISH probes, inter- and intrachromosomal rearrangements of mitotic or meiotic chromosomes could be determined, allowing the identification of break points that result in natural or mutagen-induced rearrangements. This approach has been used in large-scale physical mapping to compare homologous loci in *Arabidopsis* and *Brassica rapa* (Cabbage and Mustard), a close relative, providing evolutionary data on genome structure.

2.4 Synteny

One of the incentives for the genome sequencing initiative was an expectation of significant colinearity between *Arabidopsis* and more economically important genomes. The hope was that important genes, such as those encoding disease resistance, found in the model organism, would help locate similar genes in crops. Many studies have shown, however, that genome structure is constantly changing, causing differing degrees of complexity, limiting the usefulness of *Arabidopsis* as a reference organism for gene order. The ancestral lineages of *Arabidopsis* and *Brassica* diverged 12.2 to 19.2 million years ago and typically show more than 85% nucleotide conservation in coding regions. Despite these similarities, there is extensive divergence between the structures of these genomes, making it difficult to identify even the most closely related segments of the *Brassica oleracea* and *Arabidopsis thaliana* genomes. This is, in part, because of extensive amplifications and rearrangements in plant genomes. Despite these rearrangements, 34 significant regions of colinearity have been identified between *B. oleracea* and *A. thaliana,* which represent well over a quarter of the *B. oleracea* genome. Comparisons made between the sequence of the rice genome and *Arabidopsis* have revealed very low levels of colinearity. These lineages diverged approximately 200 million years ago. This low level of synteny suggests that further reference plants should be sequenced so that evolutionary intermediates can be analyzed. This, however, presents a dilemma. Should phylogenetically important species be sequenced, or should efforts be focused directly on economically important

species? Although synteny is limited between rice and *Arabidopsis*, comparisons between rice and other cereals show high degrees of synteny at the genetic map level. This has allowed the integration of their respective genetic maps, aiding fine mapping and candidate gene identification. Marker colinearity of rice and barley (*Hordeum vulgare*) chromosomes, for example, has provided a candidate gene for the *uzu* mutant that causes semidwarfism in barley. The position of *uzu* corresponds to a rice gene, *D61*, which is similar to the *Arabidopsis* brassinosteroid-insensitive1 (*BRI1*) gene (mutations in *BRI1* cause dwarfism in *Arabidopsis*). A comparison of the *HvBRI1* sequences in *uzu* and normal barley identified a single nucleotide substitution in the mutant gene, demonstrating the usefulness of this type of analysis. Comparative analysis can, however, be frustrated by the surprising number of rearrangements and deletions that have occurred at the microlevel between rice and closely related species such as sorghum and maize. Reduced levels of microcolinearity between these species, in particular, have frustrated the use of rice as an intergenomic cloning vehicle to isolate disease resistance genes.

2.5 Gene Families

Analysis of the complete *Arabidopsis* genome sequence has made it possible to identify the genomic position of entire gene families and to reconstruct their evolution. For example, putative disease resistance genes are unevenly distributed across the 5 chromosomes of *Arabidopsis*. There are 50 on chromosome 1; 7 on chromosome 2; 17 on chromosome 3; 32 on chromosome 4; and 60 on chromosome 5. The genes are arranged as single gene loci, clusters, and superclusters. The large number of genes in this family resulted from an ancient duplication of the genome and subsequent rearrangements. (This is likely to be the case for all large gene families in *Arabidopsis*.) Subsequently, a few ancestral genes underwent local amplifications, leading to tandem gene pairs, some of which were broken up by mechanisms such as chromosomal translocations, while others produced clusters or superclusters as a result of further amplifications. Understanding the evolution of gene families helps account for the low levels of colinearity observed between plant lineages such as *Arabidopsis* and rice. Despite these differences, however, it is interesting to note that the rice genome contains putative orthologs of nearly all genes known to control disease resistance responses in *Arabidopsis*. This suggests that, despite high levels of genomic rearrangement since their divergence some 200 million years ago, there remains extensive conservation of disease resistance signaling between the dicots and monocots.

2.6 Genome Plasticity

Comparative genome studies have demonstrated that plant genomes are highly dynamic. Amplifications, such as those described for disease resistance gene clusters, are thought to result from intra- and intergenic recombination and gene conversion. Chromosomal duplications, deletions, inversions, and translocations also occur frequently, and transposons may be major mediators of genome expansion and contraction. For instance, a large inversion of about 700 kb relocated pericentromeric regions to an interstitial

position on the short arm of chromosome 4, where it forms a heterochromatic knob.

In addition to rearrangements within the nuclear genome, lateral gene transfer from endosymbionts has played a significant role in the architecture of the plant nuclear genome. It has been estimated that as much as 18% of the *Arabidopsis* genome is derived from cyanobacteria, and that even recent integrations of chloroplast (derived from cyanobacteria) and mitochondrial DNA (derived from α-proteobacteria) have occurred. The *Arabidopsis* genome revealed 17 different transfer RNA- and intron-containing insertions of chloroplast DNA as well as a complete copy of the 367-kb mitochondrial genome near the centromere of chromosome 2. Evidence of these events have also been observed in rice and humans with 59 fragments of mitochondrial DNA >2 kb inserted in the human genome. In plants, these lateral gene transfers have recently been demonstrated in the laboratory, suggesting that this type of phenomenon occurs relatively frequently. It is interesting to note that chloroplast genomes encode only 50 to 200 proteins, while cyanobacterial genomes, from which the chloroplast originated, encode several thousand proteins. This reduction has been made possible by the nuclear acquisition of genes from the organelle. Subsequent evolution has provided the necessary targeting signals to allow these nuclear-encoded proteins to be synthesized on cytoplasmic ribosomes and reimported into the organelles.

3 Genome Function

Although the *Arabidopsis* genome has not demonstrated a high level of synteny with other plant genomes, it has been remarkably successful as a tool for identifying gene function. Despite the potential for functional redundancy, *Arabidopsis* has provided (and will continue to provide) an excellent reference organism for functional genomics. The function of 69% of genes can be classified according to sequences similar to proteins of known function in other organisms. However, the function of only 9% of genes has been determined experimentally (The *Arabidopsis* Genome Initiative).

The current challenge is to assign a function to the approximately 8000 genes with completely unknown function, and to verify the true functions of the more than 15 000 genes that have been assigned functions by analogy with other organisms. Even the approximately 2000 genes, whose functions have been identified experimentally, may have additional, for example, stage-specific, functions of which we are unaware. The challenge is to discover novel methods by which to unravel the secret life of plants.

3.1 Loss-of-function

Classically, genetic pathways have been dissected using screens for loss-of-function mutations that disrupt a single gene, thus revealing its function. This type of analysis is collectively known as *forward genetics*. Loss-of-function mutations can be created in a number of ways, including chemical mutagens, physical mutagens, or biological mutagens, such as T-DNA (transfer DNA) or transposable elements. The role of genes, for example, in disease resistance, development, signal transduction, membrane transport, and light perception has all

been inferred from the analysis of mutant phenotypes.

3.2 Classical Mutants

Classical loss-of-function mutations are frequently induced by chemical (most commonly ethyl methanesulfonate [EMS]) or physical mutagens such as radiation (X ray, γ ray, or fast neutron). EMS mutants often produce G to A transitions, while fast neutrons usually produce small deletions. Mutagenesis using EMS is the most commonly used method since it is efficient and generates a wide range of mutant phenotypes. These phenotypes identify varying degrees of disrupted gene function, providing complete loss of function, weak, dominant, or conditional alleles. These mutant alleles can be used in a variety of ways to further characterize the function of a gene, for example, weak alleles of mutant genes can be used to identify second site modifiers, which either enhance or suppress the primary phenotype. The disadvantage of EMS, however, is that it generates many mutations within the genome and extraneous mutations must be removed by repeated backcrossing to wild type. Additionally, genes that act redundantly cannot be identified in loss-of-function screens. Furthermore, the only way to isolate the mutant gene is by the tiresome task of positional cloning. Positional cloning, however, is no longer the technically demanding task it was in the 1990s. As a result of the complete genome sequence and the extensive number of molecular markers that are available, the procedure is now routine. Ultimately, a mutant gene is identified by the complementation of the mutant phenotype with the wild-type gene, or the molecular analysis of several independent mutant alleles. Fast neutron mutagenesis often results in small deletions, which make positional cloning even easier, since DNA blot hybridization, rather than DNA sequencing, can be used to identify the mutated gene.

3.3 Insertional Mutants

In contrast to EMS or fast neutron mutagens, insertional mutagens, such as defined T-DNAs or transposable elements can be used as a tag to identify the *Arabidopsis* genomic DNA that flanks them (gene tagging).

3.3.1 T-DNA

The T-DNA of *Agrobacterium*-mediated plant transformation vectors is used as a powerful insertional mutagen. T-DNA insertions occur randomly in the *Arabidopsis* genome, usually generating loss-of-function mutations. The relationship between the mutation and the T-DNA insertion can be verified simply by screening for co-segregation of the mutant phenotype with the marker gene on the T-DNA. As many mutants generated by this strategy are not tagged, the disrupted gene must be used to complement the mutant phenotype to confirm its function. The disadvantage of T-DNA insertions is that the number of mutations induced in each plant line is low (1 or 2), as compared to EMS (where 100 or more lesions are induced per genome). Thus, a much larger T-DNA-mutagenized population must be screened to reach levels of saturation similar to those obtained using EMS. High-efficiency transformation procedures have allowed the generation of large populations of T-DNA insertion lines over recent years.

3.3.2 Transposons

Transposons are an alternative to T-DNAs for the production of insertional mutations. Here, a tagged mutation can be verified by the remobilization of the transposable element, which restores the mutant phenotype to wild type, thus avoiding the need for complementation. Transposon insertion screens can make use of endogenous or heterologous transposable elements.

The most popular endogenous transposable element, *Tag1*, was first identified as an insertion in the nitrate transporter gene *CHL1* of *Arabidopsis*. This element was found in Landsberg *erecta*, but not in the Columbia or Wassilewskija accessions. *Tag1* is an autonomous element capable of independent excision. Its activity is restricted to developing male and female gametophytes, which means that excision events are predominantly germinal. Germinal excision is important for gene tagging since this determines the number of potentially useful heritable mutagenic events. Endogenous elements also account for at least 10% of the *Arabidopsis* genome (one-fifth of the intergenic DNA), but these are generally silenced (The *Arabidopsis* genome initiative, 2000). Heterochromatic regions are rich in transposable elements and evidence is accumulating that heterochromatin was derived from euchromatin by the insertion of transposable elements. Heterochromatic regions contain methylated DNA and other forms of chromatin modification that are thought to regulate the mobility of transposons. Although transposons are frequently silenced by methylation, active transposons have also been found in plants, most famously in maize.

When transposon tagging was under development in *Arabidopsis*, very few endogenous elements had been characterized, and this led to the introduction of heterologous transposable elements from other species. Maize elements, *Activator/Dissociation* (*Ac/Ds*), and *Enhancer/Inhibitor* (*En/I*), also known as *Suppressor-mutator* (*Spm*), are among the most popular of these heterologous elements. *Ac/Ds* has been widely used to great effect in a number of plant species, including tomato, tobacco, and *Arabidopsis*. In *Arabidopsis*, effective levels of transposition of *Ds* (a nonautonomous element) were achieved using a stabilized *Ac*-expressing transposase under the 35S promoter. Using this system, the *Ds* element is transposed when, by crossing plants, the *Ac* transposase is supplied *in trans*. Studies have shown that somatic excision, as a result of the constitutive expression of the 35S *Ac* transposase, occurs frequently, and that *Ds* transposition is often linked to the donor site (68% of transpositions). This linked transposition provides a clear advantage over T-DNA insertions in targeted strategies; it allows the transposition of *Ds* from launch sites close to the target locus. To prevent excision of the *Ds* element after transposition, *Ac* is usually segregated away, thereby generating stable insertions. Nevertheless, some insertional mutants may lose their *Ds* tag prior to stabilization. The *Ac/Ds* system has been used on many occasions to identify gene function in *Arabidopsis*, as has the *En/Spm* system.

3.3.3 Enhancer Detection and Gene Trapping

The spatial and temporal control of gene expression is mediated by *cis*-acting elements and *trans*-acting transcriptional regulators that bind in a sequence-specific manner. Such elements were first revealed through analyses in which promoter fragments were fused to a reporter gene.

Enhancer detection is a method that was first developed in *Drosophila*, allowing the identification of such *cis*-regulatory elements (enhancers) in the genome. The method, which relies on a mobile genetic element carrying a reporter gene under a weak promoter that can be influenced by nearby regulatory sequences, allows the identification of genes based on their pattern of expression. A modification of this approach, gene trapping, which relies on the generation of transcriptional fusions between the endogenous gene and the introduce reporter gene, was then quickly developed for mice. Both approaches were adapted to *Arabidopsis*, either using T-DNA or *Ds* elements that contain a reporter gene, encoding either beta-glucuronidase (GUS) or the green fluorescent protein (GFP). The expression of the reporter gene depends on the flanking *cis*-acting signal sequences. These produce cell- and tissue-type specific expression patterns that can be used not only as molecular markers but also to provide clues as to the function of the associated gene.

3.4 Targeted Gene Disruptions

A number of multicellular model organisms have benefited from routine methods of targeted gene disruption (TGD), for example, mouse, the moss *Physcomitrella patens*, and *Drosophila*. Despite much effort, TGD approaches in *Arabidopsis* have proved inefficient. Even the targeted disruption of transgenes has been difficult. However, a method for assessing the efficiency of TGD in endogenous *Arabidopsis* genes has recently been developed. Results have shown that, while the frequency of disruption is indeed low, the actual number of events is sufficient for the isolation of a targeted mutant from the large number of transgenic seeds that can be obtained. Unfortunately, these methods are not yet routine, so the best current options are either to screen for local insertions into the gene of interest by a nearby transposable element or to take advantage of the expanding number of insertion mutants now available (http://enhancertraps.bio.upenn.edu/; http://genetrap.cshl.org/; http://www.biotech.wisc.edu/Arabidopsis/; http://www.mpiz-koeln.mpg.de/GABI-Kat/; http://flagdb-genoplante-info.infobiogen.fr/projects/fst/DocsIntro/introCollection.html; http://ukcrop.net/agr/insert.html; or http://nasc.nott.ac.uk/info/slat_info1.html). These approaches, together with genome sequence data, have provided an invaluable resource for reverse genetics. Insertional mutants can be screened by PCR, or by BLAST-searching a gene of interest against the DNA sequences that are identified adjacent to the insertion site. Several collections of insertion lines can be screened for known insertions in a gene of interest. Knowing the complete sequence of the genome has made an alternative reverse genetics approach very powerful. This approach, known as *TILLING* (Targeted Induced Local Lesions IN Genomes), combines high-density point mutations, provided by traditional chemical mutagenesis, with sensitive detection methods. This allows mutations in a particular gene of interest to be identified, so that genes of unknown function can be targeted. The real advantage here is that a series of mutant alleles can be produced, including weak or possibly conditional alleles; whereas insertional mutants tend to be knockouts. This means that the function of essential genes can be studied as a result of identifying nonlethal mutant alleles.

3.5 Microarrays

A powerful experimental approach to help elucidate the function of the large numbers of genes that have been identified from the genome sequence is to use gene chips or microarrays. These tools allow the systemic analysis of expression profiles of large numbers of genes, including stress-induced, tissue-specific, or developmentally regulated genes. DNA microarrays are constructed either from cDNA expressed sequenced tags (ESTs), gene sequences, or gene-specific oligonucleotides. These systemic experimental approaches have generated large data sets that can be used to identify genes that have similar expression profiles under specific conditions or identify genes that are expressed at similar developmental stages. They can even be used to identify the expression of genes that are restricted to a small subset of cells. The advantage of this type of analysis is that candidate genes with parallel regulation to known genes that are involved in specific responses or developmental processes can be identified. These candidate genes must be further analyzed to determine their specific role in the plant function of interest. This type of analysis can also be used to identify common *cis*-acting regulatory elements in genes that are regulated in parallel. For example, 450 circadian-regulated genes were identified in *Arabidopsis* using microarrays. Comparative promoter sequence analysis of these genes revealed a nine-nucleotide "evening element", which was shown to be required for circadian control of gene expression. Similarly, genes may be identified that are involved in several responses, for example, in the stress responses to drought, cold, and high-salinity, highlighting cross talk between regulatory networks.

The combination of chip analysis with mutational analysis to identify downstream target genes of regulatory complexes is another powerful approach. Using this technology, differences in the expression profiles of the mutants *mea, fie* (components of the MEA/FIE Polycomb Group [PcG] complex) and wild-type plants were identified during early embryogenesis, revealing that two genes, *PHERES1* and *MEIDOS*, were significantly deregulated in these mutants. One of these candidate genes, *PHERES1*, was analyzed further, and, using chromatin immunoprecipitation, it was shown that its promoter region was bound directly by the MEA/FIE PcG complex.

3.6 Gain-of-function Mutations

Identifying the function of genes in large gene families often presents problems, especially if these genes act redundantly. Loss-of-function mutations in redundantly acting genes are frequently overlooked. For instance, the *kanadi1* (*kan1*) and *kanadi2* (*kan2*) single mutants show no visible phenotype in *Arabidopsis*, but together lead to loss of polarity in lateral organs. In the double mutants, abaxial cell types are replaced by adaxial cells in most lateral organs, most strikingly in reproductive tissue, leading to radial development of ovules (Fig. 3). Similarly, a triple mutant of the TGA family of transcription factors, *tga6, tga2 tga5*, blocks salicylic acid-induced pathogen-related gene expression and disease resistance, but not single or double mutant combinations.

Another way to identify the function of redundantly acting genes is to express them ectopically. In this way, a mutant phenotype can be produced if the ectopic

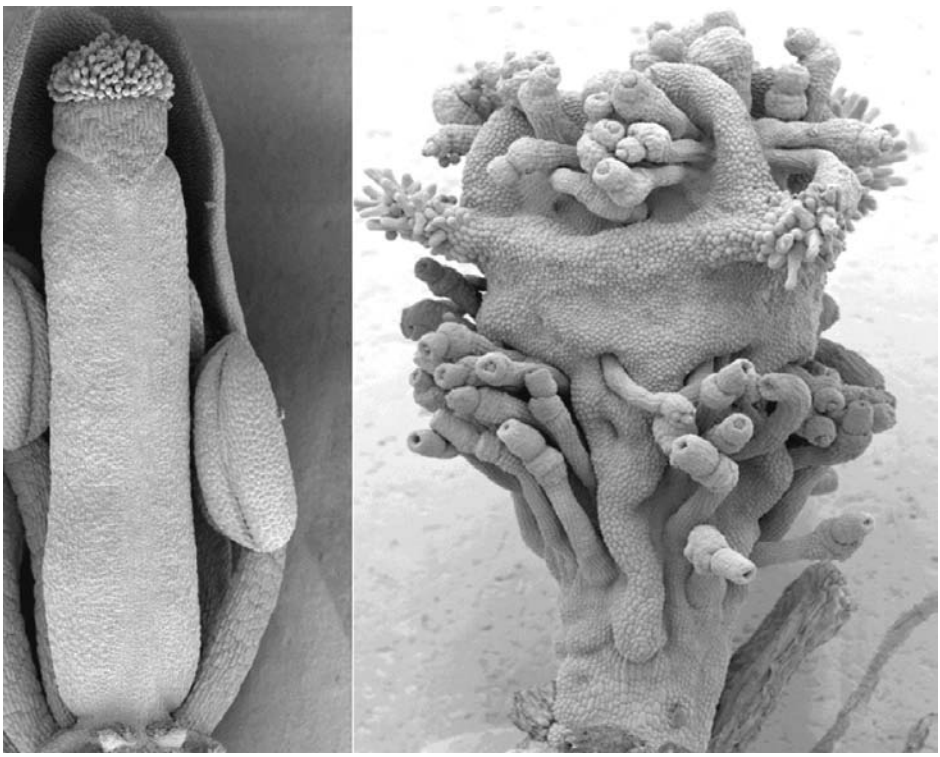

(a) (b)

Fig. 3 *kanadi* loss-of-function phenotypes. (a) In the wild-type *Arabidopsis* gynoecium, ovules are restricted to the internal (adaxial) side. (b) In *kan1 kan2* plants polarity is lost so that the placenta develops externally and ovules develop radially around the entire abaxial circumference of the ovary and on its distal end. (Eshed, Y., Baum, S.F., Perea, J.V., Bowman, J.L. (2001) Establishment of polarity in lateral organs of plants, *Curr. Biol.* **11**, 1251–1260.) Reproduced with permission from Elsevier.

expression of a gene deregulates a genetic pathway. Again, with the *KANADI* genes, single gene knockouts show no phenotype; however, the ectopic expression of either *KAN1* or *KAN2*, throughout leaf primordia, results in the dramatic transformation of adaxial to abaxial cell types. Several strategies to achieve a gain-of-function phenotype in plants have been devised.

3.6.1 Constitutive Ectopic Expression

The simplest approach is to place a gene under the transcriptional control of a constitutive promoter such as the CaMV 35S promoter. This type of analysis has been the basis of many gene function studies. For instance, overexpression of the floral homeotic gene *APETELA3* (*AP3*), under the control of the 35S promoter, showed the positive regulatory interaction between *AP3* and *PISTILLATA* (*PI*) in floral development (Fig. 4).

The ectopic expression of genes in one species can result in a phenotype reminiscent of another species, leading to

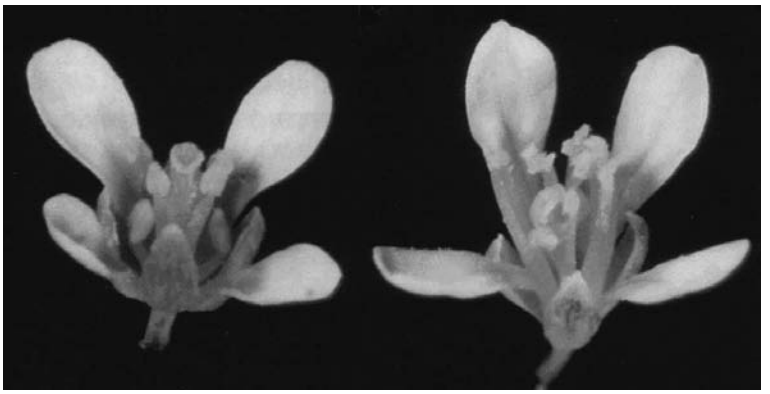

(a) (b)

Fig. 4 Phenotype of *Arabidopsis thaliana* flowers. (a) Wild type, showing sepals (whorl 1), petals (whorl 2), stamen (whorl 3), carpels (whorl 4). (b) Overexpression of *AP3* (*35S::AP3*) showing sepals (whorl 1), petals (whorl 2), stamen (whorl 3), stamen (whorl 4). *PI* is expressed in whorls 2, 3, and 4, early in floral development, but not in whorl 1. Ectopic expression of *AP3* in all whorls causes additional *PI* expression to be maintained in whorl 4, showing a positive regulatory reaction between *AP3* and *PI* in floral development. In wild-type flowers, *AP3* is only expressed in whorls 2 and 3, so that *PI* expression in late floral development is maintained only in whorls 2 and 3. (Jack, T., Fox, G.L., Meyerowitz, E.M. (1994) *Arabidopsis* homeotic gene APETALA3 ectopic expression: transcriptional and post-transcriptional regulation determine floral organ identity, *Cell* **76**, 703–716.) Reproduced with permission from Elsevier.

the identification of genes that play a role in the evolution of developmental pathways. Ubiquitous overexpression of the *Knotted*-class homeobox gene *KNAT1* (usually restricted to the shoot meristem), resulted in highly abnormal leaf morphology that included severely lobed leaves. The morphology of leaves is a prominent characteristic of different plant species. In plants with simple leaves, the *KNOTTED1*-like genes are expressed in the shoot apical meristem (SAM) and are downregulated in leaves. In many plants with complex leaves, it was found that *KNOTTED*-like genes are expressed late in leaf development. The gene in *Arabidopsis* responsible for downregulating *Knotted*-class genes is *ASYMMETRIC LEAVES1* (*AS1*), encoding a MYB-domain protein closely related to *PHANTASTICA* in *Antirrhinum* and *ROUGH SHEATH2* in maize. Downregulation of the *PHANTASTICA*-like MYB transcription factor in leaves seems to have been an important factor in the evolution of different leaf forms in plants. These discoveries have demonstrated that altered gene expression patterns contribute to the diversity of plant morphology. Another well-documented example is found in flowers. Here, the expression domains of the MADS-box proteins have contributed to the diversity of floral morphology found among the angiosperms. The mis-expression of genes is, therefore, a powerful tool for determining gene function and can identify candidate genes that may be responsible for the development of diverse morphologies over evolutionary time.

Several alternative approaches to misexpress genes have been developed that allow the activation of randomly tagged genes. These include specialized transposable elements that place a constitutive promoter upstream of genes as a result of random integration of transposons. This type of gain-of-function approach was pioneered by Coupland and colleagues. The technique allows ubiquitous ectopic expression of activation-tagged genes. Using this technique, however, genes essential for regulating early embryonic and gametophytic pathways may be lost, since the ubiquitous ectopic expression of these genes may produce a lethal effect. In addition, the ubiquitous ectopic expression may mask any tissue-specific effects that may be identified when deregulated gene expression is restricted to certain cell types. An alternative T-DNA-based system that contains multimerized transcriptional enhancers from the *CaMV 35S* gene has been developed. Using this approach, it appears that the expression pattern of endogenous genes is maintained; however, the level of expression is enhanced. This approach has not only been useful in identifying the function of genes but has also helped identify mechanisms that regulate the accumulation of specific mRNAs that are required for normal development. MicroRNAs downregulate their target genes either by cleaving the mRNA or by blocking their translation. Using this activation-tagging technique, specific expression of microRNA from the *JAW* locus was enhanced, revealing its role in targeting the mRNAs of a family of transcription factors for cleavage. These mRNAs encode the TCP domain transcription factors involved in leaf development, providing tangible evidence that microRNAs provide an endogenous mechanism of gene silencing that specifically regulates aspects of plant morphogenesis.

3.6.2 Inducible Ectopic Expression

The generation of mutants in critically important genes inevitably results in lethality. To overcome early lethality problems, inducible tagging systems have been devised. These rely on heat shock or chemical induction, so that ectopic gene expression can be controlled temporally. Using a chemically inducible activation-tagging system, the *WUSCHEL* gene, a homeodomain transcription factor normally restricted to the central zone of the SAM and nucellus tissue of the developing ovule primordium, was misexpressed, allowing ectopic embryo development in roots (Fig. 5). Here, the effect of gene expression is tissue-specific and demonstrates the importance of gene regulation to restrict expression to a particular subset of cells. An example demonstrating the value of restricted ectopic gene expression to very small regions of tissue (fractions of a meristem) is the induced ectopic expression of expansin in the meristem, leading to the formation of leaves. This will be of particular importance when attempting to deregulate gene expression in studies on the evolutionary development of reproductive strategies in plants. *Arabidopsis* reproduces sexually; however, it is generally accepted that apomictic reproduction (clonal reproduction through seed) results from the deregulation of sexual processes, since apomicts are derived from sexual ancestors. These rapid alterations in gene expression are thought to result from epigenetic effects, frequently produced by ploidy changes that result from interspecific hybridization. Further development of tissue-specific inducible gene activation may reveal candidate genes, whose altered gene expression patterns can account

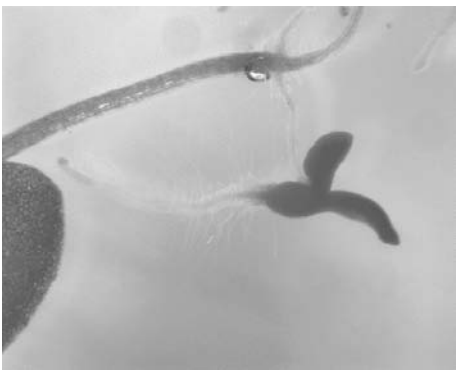

Fig. 5 Misexpression of a gene that controls stem cell identity (*WUSCHEL*) can be used to redirect root cells of an intact plant to any of the other major sporophytic developmental pathways, depending on additional cues, for example, leaf development (without additional cues), floral development (together with LEAFY), or embryogenesis (in response to increased auxin) (Gallois, J.L., Nora, F.R., Mizukami, Y., Sablowski, R. (2004) WUSCHEL induces shoot stem cell activity and developmental plasticity in the root meristem, *Genes Dev.* **18**, 375–380.) Image from Curtis and Grossniklaus, 2003, is copyrighted by the American Society of Plant Biologists and is reprinted with permission.

for changes in the reproductive pathway of plants.

4 Gene Regulation

The regulated expression of genes allows an organism to respond to environmental cues, to develop, and to reproduce. These mechanisms of gene regulation involve more than 12% of *Arabidopsis* genes. Gene expression is controlled, both spatially and temporally, through specific interactions of transcription factors with their DNA recognition sequences, and through the dynamic restructuring of chromatin, which alters the accessibility of the DNA to these factors.

4.1 Transcription Factors

There are more than 1500 transcription factors encoded by the *Arabidopsis* genome, which account for about 6% of the predicted total number of genes. The *Arabidopsis* genome encodes diverse families of transcription factors, many of which are unique to plants, for example, the WRKY, NAC, and AP2/EREBP families. Others factors are present in families that are common to plants, fungi, and animals. However, in plants, these families have grown to include large numbers of genes. Two such super families are those of the MYB and MADS transcription factors. The MYB factors comprise one of the largest families with approx. 190 MYB and MYB-related genes identified in *Arabidopsis*. A similar number, 156, have been identified in rice. In contrast, the corresponding family in yeast has only 10 members, *Drosophila* has 6, and *Caenorhabditis elegans* only 3. A similar pattern emerges for the MADS-box genes; *Arabidopsis* has 107, rice has 71, yeast has 4, and *Drosophila* and *C. elegans* have only 2 each. This type of comparative genome analysis, between *Arabidopsis*, rice, yeast, *C. elegans*, and *Drosophila*, has identified differences in the relative size of many other gene families, revealing that plants and animals have developed different regulatory mechanisms since their divergence to achieve similar processes. In both animals and plants,

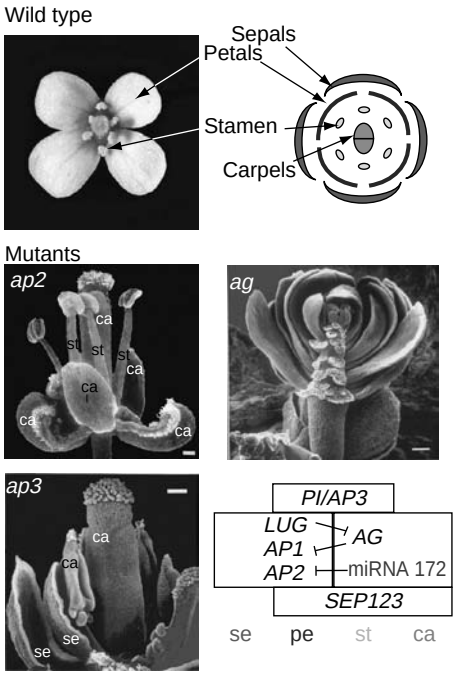

Fig. 6 The ABC model was proposed after the analysis of phenotypes of homeotic floral mutants. Floral organ identity is hypothesized to result from the combined activity of 3 classes of homeotic gene products. A function (*AP1* and *AP2*) = sepals, A + B (*AP3* and *PI*) function = petals. B + C (*AGAMOUS*) = stamens, and C function alone = carpels. Loss-of-function alleles of these genes result in homeotic transformations; for example, in *ap2* mutants, carpels develop in the position of sepals and stamen in the position of petals. In *ag* mutants, petals develop in the position occupied by stamen, and another flower meristem replaces the carpels. In *ap3* mutants, sepals replace petals and carpels replace stamen. In triple mutants, the floral organs form leaf-like structures; however, leaves are not converted to floral organs unless a *SEP* gene is expressed in addition. Other factors restrict the spatial activity of the A, B, and C class genes. The spatial activity of the A, B, and C class genes are controlled so that *AP2* and *LEUNIG* restrict the activity of *AG* to whorls 3 and 4. *AG* restricts the activity of *AP1* to whorls 1 and 2. *SUPERMAN* (not shown) prevents the activity of B class genes in whorl 4 and a microRNA, *miR172*, prevents the translation of *AP2* mRNA in whorls 3 and 4. The *ap2* SEM reproduced from Bowman et al. (1991) with permission from Development and the company of Biologists; the *ag* SEM reproduced from Bowman and Meyerowitz (1991) with permission from the company of Biologists; and the *ap3* SEM reproduced from Bowman, Smyth and Meyerowitz (1989) (*Plant Cell* **1**, 37–52, with permission from the American Society of Plant Biologists).

for example, the position and specification of organs is an essential prerequisite for normal development. In animals, this patterning is achieved by the localized activity of members of the homeobox gene family, which determines the spatial arrangement and proper positioning of organs, such as limbs. In plants, however, the specification of floral organ positioning and identity (sepals, petals, stamens, carpels) is reliant on the domain-specific expression of members of the MADS-box family (Fig. 6). Although both plants and animals follow the same basic concept (combinatorial gene regulation) to specify and position organs, they use distinct families of transcription factors, despite the fact that they both have homeo- and MADS-box genes. This divergence is further emphasized by the absence of certain transcription factor families from various lineages, as described earlier.

4.2 Epigenetics

Frequently, the development of technology to identify gene function has led to new discoveries in gene regulation. Inactivation of gene expression was an unexpected result of the insertion of multiple copies of transgenes for ectopic expression. This led to the first investigations into gene silencing in multicellular organisms using sense or antisense transgenes that were shown to downregulate target gene expression. Manipulations with antisense

genes have served as tools to determine the mechanisms of gene function in many aspects of the plant life cycle, including floral development and disease resistance. Although gene silencing has served as an alternative means by which to phenocopy the effect of mutants, it has also allowed the genetic analysis of silencing mechanisms.

Using plant lines with silent transgenes, it has been possible to show that specific gene inactivation can occur, either at the transcriptional level (transcriptional gene silencing, TGS), or posttranscriptionally (posttranscriptional gene silencing, PTGS), as a result of specific mRNA degradation. Mutant screens that identified plant lines with reactivated transgene expression produced lesions in genes that are involved in silencing. Using this type of approach, it was discovered that mutations in *ddm1* (decrease in DNA methylation) cause the activation of silenced transgenes. The *DDM1* gene encodes a SW12/SNF2 chromatin-remodeling factor. When this gene is mutated, cytosine methylation of DNA is reduced by 70%, with most immediate effects in repeated sequences, and also affecting dispersed sites upon inbreeding, causing progressive heritable developmental abnormalities. Reductions in the methylation of repetitive sequences lead to the reactivation of transcriptionally silent transposable elements and silenced endogenous genes.

Cytosine methylation at symmetric CG nucleotides is an evolutionarily conserved DNA modification that has been found in vertebrates, plants, and some fungi. In addition, plants have significant levels of cytosine methylation at non CG-sequences, including symmetrical CNG and asymmetrical CNN sequences. DNA methylation is established and maintained by a variety of DNA methyltransferase genes (at least 10 encoded by the *Arabidopsis* genome). In animals, the methylation signal is recognized by methyl-CpG-binding domain (MBD) proteins (there are at least 12 putative MBD proteins in *Arabidopsis*). In animals, these proteins recruit a variety of histone deacetylase-(HDAC) and chromatin-remodeling complexes, which lead to heterochromatin formation and transcriptional silencing. In the *Arabidopsis* mutant *flowering locus d* (*fld*), the expression of the repressor *FLOWERING LOCUS C* (*FLC*) (a MADS-box transcription factor that blocks the transition from vegetative to reproductive development) is elevated, resulting in late flowering. FLD is a plant homolog of the human protein KIAA0601, a component of the human histone deacetaylase 1,2 (HDAC1/2) corepressor complex. In the *fld* mutant, the *FLC* locus contains nucleosomes in which histone H4 is hyperacetylated. FLD therefore regulates flowering time in *Arabidopsis* by altering chromatin architecture; deacetylation of histone H4 causes condensation of chromatin to produce transcriptionally inactive heterochromatin. The N-terminal tails of histones can be subjected to a variety of posttranslational modifications, including phosphorylation, ADP-ribosylation, sumoylation, methylation, as well as acetylation. Interestingly, the *FLC* gene also undergoes dimethylation at lysine 9 and lysine 27 on histone H3 during vernalization (an extended period of cold treatment required for floral induction). These methylation marks are applied by a Polycomb complex that has histone methyltransferase activity toward histone H3. Methylated H3 histones at the *FLC* locus cause stable repression of *FLC* after vernalization, thus allowing the induction of flowering. It is known that maintenance DNA-methyltransferases perpetuate existing DNA methylation in CpG

sequences after DNA replication by methylating hemimethylated sites. Less is known about the way in which unmethylated DNA loci are initially targeted for methylation and gene silencing. This type of gene silencing is required by plants to defend against invading viruses and transposable elements, and is directed at repetitive sequences, suggesting a homology-based trigger for DNA methylation. In plants, the first clues that this phenomenon may be caused by RNA came from the observation that RNA viroids cause *de novo* cytosine methylation of homologous genomic DNA sequences. Among the candidate methyltransferase genes encoded by *Arabidopsis* for *de novo* DNA methylation are the domain rearranged methyltransferases (DRMs), which share homology with the mammalian *Dnmt3* genes, which encode *de novo* methyltransferases. Recently, DRMs have been shown to be involved in RNA-directed DNA methylation (RdDM). Significantly, mutations in *dicer-like3* (*dcl3*), *RNA-dependent-RNA-polymerase2* (*rdr2*), *silencing defective4* (*sde4*), and *argonaute4* (*ago4*) have been shown to affect DRM-dependent *de novo* methylation of a *FWA* transgene and endogenous *FWA* locus. Interestingly, *DCL3*, *RDR2*, and *AGO4* are homologs of proteins involved in RNA silencing. RNA silencing uses 21- to 25-nucleotide small interfering RNAs (siRNA) to direct mRNA cleavage or induce transcriptional gene silencing. This type of RNA-induced gene silencing has been exploited in plants to silence target genes. Single-stranded hairpin RNAs (hpRNAs), or dsRNAs, are used in an approach known as *RNAi* (RNA interference). Vectors producing dsRNA are introduced into a plant, where an RNase III–like helicase (Dicer) specifically cleaves the dsRNA into siRNA. The siRNAs are incorporated into a nuclease-containing complex called *RISC* (RNAi silencing complex – *ARGONAUTE* is a component of RISC), which degrades any mRNA that is complementary to the single-strand siRNA associated with the complex. The siRNA also targets homologous DNA sequences for *de novo* methylation by *DRM1* and *DRM2*.

RNAi constructs designed to silence endogenous genes can be made simply, using a variety of vectors, and can be stably expressed in plants following *Agrobacterium*-mediated transformation or transiently expressed following bombardment. The advantage of a homology-based system of gene silencing is that, using this technique, gene families can be silenced so that a mutant phenotype can be obtained, even when genes act redundantly. Interestingly, PTGS or RNAi silencing in plants can propagate, after the initiation of local silencing, from cell to cell through plasmodesmata (continuous cytoplasmic bridges between cells) and over long distances through the vascular system, causing sequence-specific degradation of transcripts systemically (Fig. 7). Thus, RNAi is an extremely powerful method for analyzing gene function. However, because of its systemic nature, tissue-specific knockdown of genes tend to be unsuccessful. A further disadvantage of gene knockdowns using RNAi is the potential instability of the RNAi effect. This is most likely caused by transcriptional silencing of the RNAi construct itself.

4.3 Cellular Communication

The function of many genes in plants is to communicate between cells. Cell

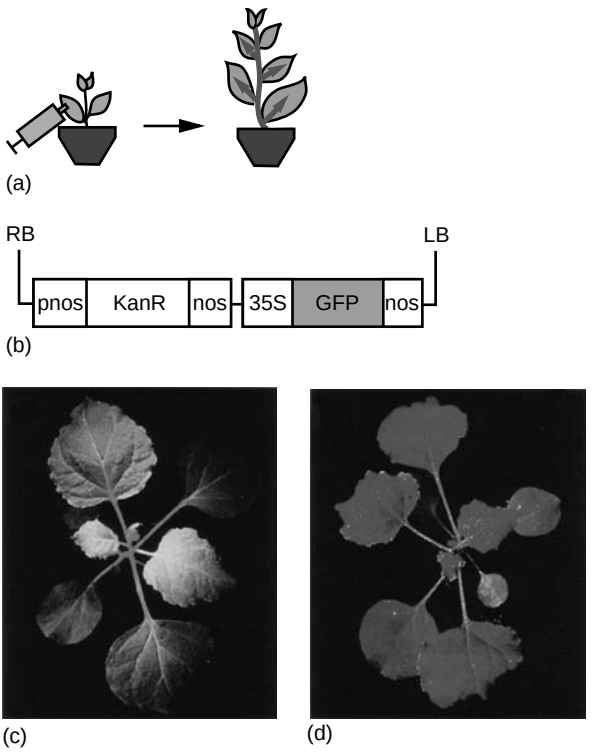

Fig. 7 Posttranslational gene silencing of a GFP transgene, induced by infiltration with *Agrobacterium tumefaciens*. Schematic representation of the experiment (a). Transgenic *Nicotiana benthamiana*, containing a T-DNA (b), and expressing GFP, was infiltrated with a hypervirulent strain of *A. tumefaciens* carrying the same T-DNA. Prior to infiltration, the plants show high levels of GFP expression under UV illumination (c). Ten days after infiltration complete GFP silencing was achieved (d). Plants appear uniformly red as a result of chlorophyll fluorescence under UV illumination. (Brigneti, G., Voinnet, O., Li, W.-X., Ji, L.-H., Ding, S.-W., Baulcombe, D.C. (1998) Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*, *EMBO J.* **17**, 6739–6746.) Reproduced with permission from Nature Publishing Group. (See color plate p. xxv.)

communication can involve short- or long-range signals that usually operate via ligand–receptor interactions that initiate intracellular signal transduction cascades, much as they occur in animal systems. There is, however, an intercellular cell-to-cell signaling system unique to plants that is based on plasmodesmata, which are thought to be involved in the trafficking of endogenous proteins, such as transcription factors, and are used by viruses for cell-to-cell spread. KNOTTED in maize, for example, has been shown to move to cells in which its RNA is not present. Mosaic

analysis is frequently used to determine whether the function of a protein is cell autonomous, or whether it has a role in cells where it is not expressed (noncell autonomous). This can be achieved in a number of ways using methods that rely either on transposons or on recombinases. The induced excision of a *Ds* element by *Ac*, or an intervening sequence by recombinases, can restore the wild-type function of a gene in a few cells of a plant. Therefore, the plant will be a chimera, containing cells of different genotypes. Using these techniques, it has been shown that APETALA3 (AP3) can regulate cell fate autonomously, while acting nonautonomously to control organ shape and size. Similarly, LEAFY (LFY) can move to adjacent cells, activating target genes, while APETALA1 activates early target genes cell autonomously. These powerful tools can be used to determine a gene's mode of action, which would not be possible by simply examining the genomic sequence. It may be possible to make predictions as to whether a protein encoded by a gene is involved in cell–cell communication, but the precise mechanism or functional significance can be determined only through experimentation.

5 Conclusions and Future Prospects

Plants are a key element of sustainable environments, they provide the oxygen we breathe and are the producers at the source of the food chain. They have played this important role for millions of years and have a tremendous influence on our weather and climate. Plant resources, and the way in which they are managed, have enormous implications for our future well-being. Careful crop management can provide a solution to famine and disease, but must also play a role in averting the current global warming crisis. In the next few decades, we must reverse our dependency on fossil fuels and grow plants for the sustainable production of oils and plastics. Plant biotechnology involves the modification of plant performance. In relation to crops, this includes increasing yield, enhancing tolerance to stress, and increasing disease resistance. Breeding high-yield, dwarf, disease-resistant wheat and rice varieties has significantly increased world grain production. However, these yields must continue to climb if food production is to keep pace with increases in world population. One way in which yields can be increased to keep pace with rising populations is to identify and manipulate the activity of genes that affect economically important traits. These modifications may be achieved by increasing or decreasing the activity of genes that are naturally present in an organism. The genes, and the techniques by which these genes can be alternately regulated, are frequently identified in model organisms like *Arabidopsis* and then transferred to crop species. A major advantage of genetically modifying organisms is that plants can be exploited, not only as sources of food but also as sources of industrial raw materials, such as oils, starches, and biodegradable plastics. They can even be engineered as bioreactors for the production of vaccines and therapeutic antibodies, or as tools for bioremediation and mopping up pollutants such as heavy metals from contaminated soils. In order to achieve these aims, a thorough and broad understanding of plant biology, in all its fields, is indispensable. This understanding will be most efficiently achieved through the continued use of model organisms.

The Thale Cress *Arabidopsis thaliana* has proved an excellent model organism for the study of many aspects of plant biology. Its small genome, and the abundance of polymorphisms between the two most commonly used accessions has made it ideal for positionally cloning mutant genes. *Arabidopsis* was initially selected as a model organism because of its small size, its ease of growth, short life cycle, and fecundity. Over the last few decades, it has emerged as an invaluable and versatile tool for the plant biologist. It possesses many features that could not have been predicted by its early proponents, including its small genome and the ease with which it can be transformed. Today *Arabidopsis* is proving an excellent model for telomere and centromere studies, and has played a prominent role in unraveling the mechanisms of epigenetic gene regulation. *Arabidopsis* is not only the model organism at the forefront of plant biology but it also contributes to our wider understanding of the genomic structures and mechanisms in eukaryotic organisms as a whole. Even before the discovery of the structure of DNA, Laibach championed a model organism that has inspired a generation of scientists and has changed the face of biology.

Acknowledgments

Several thousand people work on furthering our understanding of how the information in the *Arabidopsis* genome is converted into growth and development, physiology and environmental interactions. Naturally, we could only give a glimpse of the wealth of information available for this reference plant and apologize to all colleagues whose work we did not cover in this chapter. We thank Claudia Köhler for helpful comments and critical reading of the entire manuscript and to Catherine Feuillet for helpful comments on the synteny section.

See also Genetics, Molecular Basis of; Immunoassays; Plant Gene Expression, Regulation of; Plant Pathology, Molecular; Rhythmic and Temporal Processes in Biology.

Bibliography

Books and Reviews

Ashburner, M. (1990) *Drosophila, A Laboratory Handbook*, Cold Spring Harbor Press, Cold Spring Harbor, NY.

Cao, R., Zhang, Y. (2004) The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3, *Curr. Opin. Genet. Dev.* **14**, 155–164.

Gale, M.D., Devos, K.M. (1998) Comparative genetics in the grasses, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1971–1974.

Grossniklaus, U. (2001) From Sexuality to Apomixis: Molecular and Genetic Approaches, in: Savidan, Y., Carman, J.G., Dresselhaus, T. (Eds.) *The Flowering of Apomixis: From Mechanisms to Genetic Engineering*, CIMMYT, IRD, European Commission DG VI, Mexico, DF, pp. 168–211.

Irish, V.F. (2003) The evolution of floral homeotic gene function, *Bioessays* **25**, 637–646.

Keller, B., Feuillet, C. (2000) Colinearity and gene density in grass genomes, *Trends Plant Sci.* **5**, 246–251.

Koornneef, M., Stam, P. (1987b) Genetic Analysis, in: Koncz, C., et al. (Eds.) *Methods in Arabidopsis research*, World Scientific, Singapore, pp. 83–99.

Leister, D. (2003) Chloroplast research in the genomic age, *Trends Genet.* **19**, 47–56.

Lucas, W.J. (1995) Plasmodesmata: intercellular channels for macromolecular transport in plants, *Curr. Opin. Cell. Biol.* **7**, 673–680.

Meyerowitz, E.M., Pruitt, R.E. (1985) *Arabidopsis thaliana* and plant molecular genetics, *Science* **229**, 1214–1218.

Nowak, S.J., Corces, V.G. (2004) Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation, *Trends Genet.* **20**, 214–220.

Pruitt, R.E., Bowman, J.L., Grossniklaus, U. (2003) Plant genetics: a decade of integration, *Nat. Genet.* **33**(Suppl.), 294–304.

Reiter, R.S., Young, R.M., Scolink, P.A. (1992) Genetic Linkage of the *Arabidopsis* Genome: Methods for Mapping with Recombinant Inbreds and Random Amplified Polymorphic DNAs (RAPDs), in: Koncz, C., et al. (Eds.) *Methods in Arabidopsis Research*, World Scientific, Singapore, pp. 170–190.

Richmond, T., Somerville, S. (2000) Chasing the dream: plant EST microarrays, *Curr. Opin. Plant Biol.* **3**, 108–116.

Riha, K., Shippen, D.E. (2003) Telomere structure, function and maintenance in *Arabidopsis*, *Chromosome Res.* **11**, 263–275.

Tariq, M., Paskowski, J. (2004) DNA and histone methylation in plants, *Trends Genet.* **20**, 244–251.

Zhang, Y. (2003) Transcriptional regulation by histone ubiquitination and deubiquitination, *Genes Dev.* **17**, 2733–2740.

Primary Literature

Aarts, M.G., Corzaan, P., Stiekema, W.J., Pereira, A. (1995) A two-element *Enhancer-Inhibitor* transposon system in *Arabidopsis thaliana*, *Mol. Gen. Genet.* **247**, 555–564.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., Ecker, J.R. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*, *Science* **301**, 653–657.

Alonso-Blanco, C., Peeters, A.J., Koornneef, M., Lister, C., Dean, C., van den Bosch, N., Pot, J., Kuiper, M.T. (1998) Development of an AFLP based linkage map of Ler, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a Ler/Cvi recombinant inbred line population, *Plant J.* **14**, 259–271.

Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant, *Arabidopsis thaliana*, *Nature* **408**, 796–815.

Azpiroz-Leehan, R., Feldmann, K.A. (1997) T-DNA insertion mutagenesis in *Arabidopsis*: going back and forth, *Trends Genet.* **13**, 152–156.

Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function, *Cell* **116**, 281–297.

Bastow, R., Mylne, J.S., Lister, C., Lippman, Z., Martienssen, R.A., Dean, C. (2004) Vernalization requires epigenetic silencing of *FLC* by histone methylation, *Nature* **427**, 164–167.

Bechtold, N., Pelletier, G. (1989) In planta *Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration, *Methods Mol. Biol.* **82**, 259–266.

Bell, C.J., Ecker, J.R. (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*, *Genomics* **19**, 137–144.

Bellen, H.J., O'Kane, C.J., Wilson, C., Grossniklaus, U., Pearson, R.K., Gehring, W.J. (1989) P-element-mediated enhancer detection: a versatile method to study development in *Drosophila*, *Genes Dev.* **3**, 1288–2300.

Bender, J., Fink, G.R. (1995) Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of *Arabidopsis*, *Cell* **83**, 725–734.

Benolken, R.M., Jacobson, S.L. (1970) Response properties of a sensory hair excised from Venus's flytrap, *J. Gen. Physiol.* **56**, 64–82.

Bergelson, J., Stahl, E., Dudek, S., Kreitman, M. (1998) Genetic variation within and among populations of *Arabidopsis thaliana*, *Genetics* **148**, 1311–1323.

Berger, F., Linstead, P., Dolan, L., Haseloff, J. (1998) Stomata patterning on the hypocotyl of *Arabidopsis thaliana* is controlled by genes involved in the control of root epidermis patterning, *Dev. Biol.* **194**, 226–234.

Bharthan, G., Goliber, T.E., Moore, C., Kessler, S., Pham, T., Neelima, S.R. (2002) Homologies in leaf form inferred from KNOX1 gene expression during development, *Science* **296**, 1858–1860.

Bier, E., Vaessin, H., Shepherd, S., Lee, K., McCall, K., Barbel, S., Ackerman, L., Carretto, R., Uemera, T., Grell, F., Jan, I.Y., Jan, Y.N. (1989) Searching for pattern and mutation

in the *Drosophila* genome with P-lacz vector, *Genes Dev.* **3**, 1273–1287.

Blanc, G., Barakat, A., Guyot, R., Cooke, R., Delsey, M. (2000) Extensive duplication and reshuffling in the *Arabidopsis* genome, *Plant Cell* **12**, 1093–1101.

Bowman, J.L., Meyerowitz, E.M. (1991) Genetic control of pattern formation during flower development in *Arabidopsis*, *Symp. Soc. Exp. Biol.* **45**, 89–115.

Bowman, J.L., Smyth, D.R., Meyerowitz, E.M. (1989) Genes directing flower development in *Arabidopsis*, *Plant Cell.* **1**, 37–52.

Bowman, J.L., Smyth, D.R., Meyerowitz, E.M. (1991) Genetic interactions among floral homeotic genes in *Arabidopsis*, *Development* **112**, 1–20.

Breyne, P., Rombaut, D., Van Gysel, A., Van Montagu, M., Gerats, T. (1999) AFLP analysis of genetic diversity within and between *Arabidopsis thaliana* ecotypes, *Mol. Gen. Genet.* **261**, 627–634.

Brigneti, G., Voinnet, O., Li, W.-X., Ji, L.-H., Ding, S.-W., Baulcombe, D.C. (1998) Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*, *EMBO J.* **17**, 6739–6746.

Briza, J., Carroll, B.J., Klimyuk, V.I., Thomas, C.M., Jones, D.A., Jones, J.D. (1995) Distribution of unlinked transpositions of a Ds element from a T-DNA locus on tomato chromosome 4, *Genetics* **141**, 383–390.

Brzeski, J., Jerzmanowski, A. (2003) Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors, *J. Biol. Chem.* **278**, 823–828.

Byrne, M.E., Barley, R., Curtis, M., Arroyo, J.M., Dunham, M., Hudson, A., Martienssen, R.A. (2000) *Asymmetric leaves1* mediates leaf patterning and stem cell function in *Arabidopsis*, *Nature* **408**, 967–971.

Campisi, L., Yang, Y., Yi, Y., Heilig, E., Herman, B., Cassista, A.J., Allen, D.W., Xiang, H., Jack, T. (1999) Generation of enhancer trap lines in Arabidopsis and characterization of expression patterns in the inflorescence, *Plant J.* **17**, 699–707.

Canilleri, C., Lafleuriel, J., Macadre, C., Varoquaux, F., Parmentier, Y., Picard, G., Caboche, M., Bouchez, D. (1998) A YAC contig map of *Arabidopsis thaliana* chromosome 3, *Plant J.* **15**, 633–642.

Capecchi, M.R. (1989) Altering the genome by homologous recombination, *Science* **244**, 1288–1292.

Carpenter, R., Coen, E.S. (1995) Transposon-induced chimeras show that *floricaula*, a meristem identity gene, acts non-autonomously between cell layers, *Development* **121**, 19–26.

Cavell, A., Lydiate, D., Parkin, I., Dean, C., Trick, M.A. (1998) A 30 centimorgan segment of *Arabidopsis thaliana* chromosome 4 has six collinear homologues within the *Brassica napus* genome, *Genome* **41**, 62–69.

Chan, S.W.L., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., Jacobsen, S.E. (2004) RNA silencing genes control *de novo* DNA methylation, *Science* **303**, 1336.

Chang, C., Bowman, J.L., DeJohn, A.W., Lander, E.S., Meyerowitz, E.M. (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6856–6860.

Chono, M., Honda, I., Zeniya, H., Yoneyama, K., Saisho, D., Takeda, K., Takatsuto, S., Hoshino, T., Watanabe, Y. (2003) A semidwarf phenotype of barley uzu results from a nucleotide substitution in the gene encoding a putative brassinosteroid receptor, *Plant Physiol.* **133**, 1209–1219.

Clough, S.J., Bent, A. (1998) A floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*, *Plant J.* **16**, 735–743.

Coa, X., Aufsatz, W., Zilberman, D., Mette, M.F., Huang, M.S., Matzke, M., Jacobsen, S.E. (2003) Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation, *Curr. Biol.* **13**, 2212–2217.

Colbert, T., et al. (2001) High-throughput screening of induced point mutations, *Plant Physiol.* **126**, 480–484.

Copenhaver, G.P., Pikaard, C.S. (1996) Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana*, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution, *Plant J.* **9**, 273–282.

Curtis, M.D., Grossniklaus, U. (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta, *Plant Physiol.* **133**, 462–469.

Dean, C., Sjodin, C., Bancroft, I., Lawson, E., Lister, C., Scofield, S., Jones, J. (1991) Development of an efficient transposon

tagging system in *Arabidopsis thaliana*, *Symp. Soc. Exp. Biol.* **45**, 63–75.

Devos, K.M., Beales, J., Nagamura, Y., Sasaki, T. (1999) *Arabidopsis*-rice: will colinearity allow gene prediction across the eudicot-monocot divide? *Genome Res.* **9**, 825–829.

Donson, J., Fang, Y., Espiritu-Santo, G., Xing, W., Salazar, A., Miyamoto, S., Armendarez, V., Volkmuth, W. (2002) Comprehensive gene expression analysis by transcript profiling, *Plant Mol. Biol.* **48**, 75–97.

Eshed, Y., Baum, S.F., Perea, J.V., Bowman, J.L. (2001) Establishment of polarity in lateral organs of plants, *Curr. Biol.* **11**, 1251–1260.

Fabri, C.O., Schaeffner, A.R. (1994) An *Arabidopsis thaliana* RFLP mapping set to localize mutations to chromosomal regions, *Plant J.* **5**, 149–156.

Feldmann, K.A., Marks, M.D. (1987) *Agrobacterium*-mediated transformation of germinating seeds of *Arabidopsis thaliana*: a non-tissue culture approach, *Mol. Gen. Genet.* **208**, 1–9.

Feuillet, C., Travella, S., Stein, N., Albar, L., Nublat, A., Keller, B. (2003) Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15253–15258.

Fransz, P.F., Armstrong, S., de Jong, J.H., Parnell, L.D., van Drunen, C., Dean, C., Zabel, P., Bisseling, T., Jones, G.H. (2000) Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region, *Cell* **100**, 367–376.

Gallardo, M.H., Beckham, J.W., Honeycutt, R.L., Ojeda, R.A., Kohler, N. (1999) Discovery of tetraploidy in a mammal, *Nature* **401**, 341.

Gallois, J.L., Nora, F.R., Mizukami, Y., Sablowski, R. (2004) WUSCHEL induces shoot stem cell activity and developmental plasticity in the root meristem, *Genes Dev.* **18**, 375–380.

Gehring, W.J. (1987) Homeo boxes in the study of development, *Science* **236**, 1245–1252.

Glick, B.R. (2003) Phytoremediation: synergistic use of plants and bacteria to clean up the environment, *J. Environ. Qual.* **32**, 1583–1590.

Gossler, A., Joyner, A.L., Rossant, J., Skarnes, W.C. (1989) Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes, *Science* **244**, 463–465.

Greider, C.W. (1996) Telomere length regulation, *Annu. Rev. Biochem.* **65**, 337–365.

Gross-Hardt, R., Lenhard, M., Laux, T. (2002) WUSCHEL signaling functions in interregional communication during *Arabidopsis* ovule development, *Genes Dev.* **16**, 1129–1138.

Grossniklaus, U., Bellen, H.J., Wilson, C., Gehring, W.J. (1989) P-element-mediated enhancer detection applied to the study of oogenesis in *Drosophila*, *Development* **107**, 189–200.

Gruenbaum, Y., Naveh-Many, T., Cedar, H., Razin, I. (1981) Sequence specificity of methylation in higher plant DNA, *Nature* **292**, 860–862.

Gu, Q., Ferrandiz, C., Yanofsky, M.F., Martienssen, R. (1998) The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development, *Development* **125**, 1509–1517.

Hake, S., Frelling, M. (1986) Analysis of genetic mosaics shows that the extra epidermal cell divisions in *Knotted* mutant maize plants are induced by adjacent mesophyll cells, *Nature* **320**, 621–623.

Halfter, U., Morris, P.C., Willmitzer, L. (1992) Gene targeting in *Arabidopsis thaliana*, *Mol. Gen. Genet.* **231**, 186–193.

Hamilton, A.J., Baulcombe, D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants, *Science* **286**, 950–952.

Hanin, M., Volrath, S., Bogucki, A., Briker, M., Ward, E., Paszkowski, J. (2001) Gene targeting in *Arabidopsis*, *Plant J.* **28**, 671–677.

Harmer, S.L., Hogenesch, J.B., Straume, M., Chang, H.-S., Han, B., Zhu, T., Wang, X., Kreps, J.A., Kay, S.A. (2000) Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock, *Science* **290**, 2110–2113.

Hauge, B.M., Hanley, S.M., Cartinhour, S., Cherry, J.M., Goodman, H.M., Koornneef, M., Stam, P., Chang, C., Kempin, S., Medrano, L., Meyerowitz, E.M. (1993) An integrated genetic/RFLP map of the *Arabidopsis thaliana* genome, *Plant J.* **3**, 745–754.

Hauser, M.T., Adhami, F., Dorner, M., Fuchs, E., Glossl, J. (1998) Generation of co-dominant PCR-based markers by duplex analysis on high resolution gels, *Plant J.* **16**, 117–125.

He, Y., Michaels, S.D., Amasino, R.M. (2003) Regulation of flowering time by histone acetylation in *Arabidopsis*, *Science* **302**, 1751–1754.

Hehl, R., Baker, B. (1989) Induced transposition of *Ds* by a stable *Ac* in crosses of transgenic tobacco plants, *Mol. Gen. Genet.* **217**, 53–59.

Heinlein, M., Epel, B.L. (2004) Macromolecular transport and signaling through plasmodesmata, *Int. Rev. Cytol.* **235**, 93–164.

Huang, C.Y., Ayliffe, M.A., Timmis, J.N. (2003) Direct measurement of the transfer rate of chloroplast DNA into the nucleus, *Nature* **422**, 72–76.

Hwang, I., Kohchi, T., Hauge, B.M., Goodman, H.M., Schmidt, R., Cnops, G., Dean, C., Gibson, S., Iba, K., Lemieux, B., Arondel, V., Danhoff, L., Somerville, C. (1991) Identification and map position of YAC clones comprising one-third of the *Arabidopsis* genome, *Plant J.* **1**, 367–374.

Ilic, K., SanMiguel, P.J., Bennetzen, J.L. (2003) A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12265–12270.

Innan, H., Terauchi, R., Miyashita, N.T. (1997) Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*, *Genetics* **146**, 1441–1452.

Izawa, T., Ohnishi, T., Nakano, T., Ishida, N., Enoki, H., Hashimoto, H., Itoh, K., Terada, R., Wu, C., Miyazaki, C., Endo, T., Iida, S., Shimamoto, K. (1997) Transposon tagging in rice, *Plant Mol. Biol.* **35**, 219–229.

Jack, T., Fox, G.L., Meyerowitz, E.M. (1994) *Arabidopsis* homeotic gene APETALA3 ectopic expression: transcriptional and post-transcriptional regulation determine floral organ identity, *Cell* **76**, 703–716.

Jackson, D., Veit, B., Hake, S. (1994) Expression of maize *KNOTTED1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot, *Development* **120**, 405–413.

Jackson, S.A., Cheng, Z., Wang, M.L., Goodman, H.M., Jiang, J. (2000) Comparative fluorescence in situ hybridization mapping of a 431-kb *Arabidopsis thaliana* bacterial artificial chromosome contig reveals the role of chromosomal duplications in the expansion of the *Brassica rapa* genome, *Genetics* **156**, 833–838.

Jander, G., Norris, S.R., Rounsley, S.D., Bush, D.F., Levin, I.M., Last, R.L. (2002) *Arabidopsis* map-based cloning in the post-genome era, *Plant Physiol.* **129**, 440–450.

Jeddeloh, J.A., Bender, J., Richards, E. (1998) The DNA methylation locus *DDM1* is required for maintenance of gene silencing in *Arabidopsis*, *Gene Dev.* **12**, 1714–1725.

Jeddeloh, J.A., Stokes, T.L., Richards, E. (1999) Maintenance of genomic methylation requires a SWI2/SNF2-like protein, *Nat. Genet.* **22**, 94–97.

Jefferson, R.A., Kavanagh, T.A., Bevan, M.W. (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants, *EMBO J.* **6**, 3901–3907.

Jenik, P.D., Irish, V. (2001) The *Arabidopsis* floral homeotic gene APETALA3 differentially regulates intercellular signaling required for petal and stamen development, *Development* **128**, 13–23.

Kakutani, T., Jeddeloh, J.A., Flowers, S.K., Munakata, K., Richards, E.J. (1996) Developmental abnormalities and epimutations associated with DNA hypomethylation mutations, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12406–12411.

Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J., Weigel, D. (1999) Activation tagging of the floral inducer FT, *Science* **286**, 1962–1965.

Keller, J., Lim, E., James, D.W., Dooner, H.K. Jr. (1992) Germinal and somatic activity of the maize element *Activator* (*Ac*) in *Arabidopsis*, *Genetics* **131**, 449–459.

Kempin, S.A., Liljegren, S.J., Block, L.M., Rounsley, S.D., Yanofsky, M.F., Lam, E. (1997) Targeted disruption in *Arabidopsis*, *Nature* **389**, 802–803.

Kerk, N.M., Ceserani, T., Tausta, S.L., Sussex, I.M., Nelson, T.M. (2003) Laser capture microdissection of cells from plant tissues, *Plant Physiol.* **132**, 27–35.

Kilian, A., Chen, J., Han, F., Steffenson, B., Kleinhofs, A. (1997) Towards map-based cloning of the barley stem rust resistance genes *Rpg1* and *rpg4* using rice as an intergenomic cloning vehicle, *Plant Mol. Biol.* **35**, 187–195.

Kitajima, E.W., Lauritis, J.A. (1969) Plant virions in plasmodesmata, *Virology* **37**, 681–685.

Klimyuk, V.I., Nussaume, L., Harrison, K., Jones, J.D. (1995) Novel GUS expression patterns following transposition of an enhancer trap *Ds* element in *Arabidopsis*, *Mol. Gen. Genet.* **249**, 357–365.

Koch, M., Bishop, J., Mitchell-Olds, T. (1999) Molecular systematics of *Arabidopsis* and *Arabis*, *Plant Biology.* **1**, 529–537.

Köhler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W., Grossniklaus, U. (2003) The

Polycomb-group protein *MEDEA* regulates seed development by controlling expression of the MADS-box gene *PHERES1*, *Genes Dev.* **17**, 1540–1553.

Koltunow, A.M., Grossniklaus, U. (2003) Apomixis: a developmental perspective, *Ann. Rev. Plant Dev.* **54**, 547–574.

Koncz, C., Martini, N., Mayerhofer, R., Koncz-Kalman, Z., Korber, H., Redei, G.P., Schell, J. (1989) High-frequency T-DNA-mediated gene tagging in plants, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8467–8471.

Konieczny, A., Ausubel, F.M. (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers, *Plant J.* **4**, 403–410.

Koornneef, M., Stam, P. (1987a) Procedures for mapping by using F2 and F3 populations, *Arabidopsis Inf. Serv.* **26**, 73–78.

Koornneef, M., van der Veen, J.H. (1983) Trisomics in *Arabidopsis thaliana* and the location of linkage groups, *Genetica* **61**, 41–46.

Koornneef, M., van Eden, J., Hanhart, C.J., Stam, P., Braaksma, F.J., Feenstra, W.J. (1983) Linkage map of *Arabidopsis thaliana*, *J. Hered.* **74**, 265–272; Kuromori, T., Hirayama, T., Kiyosue, Y., Takabe, H., Mizukado, S., Sakurai, T., Akiyama, K., Kamiya, A., Ito, T., Shinozaki, K. (2004) A collection of 11 800 single-copy *Ds* transposon insertion lines in *Arabidopsis*, *Plant J.* **37**, 897–905.

Krysan, P.J., Young, J.C., Sussman, M.R. (1999) T-DNA as an insertional mutagen in *Arabidopsis*, *Plant Cell* **11**, 2283–2290.

Krysan, P.J., Young, J.C., Jester, P.J., Monson, S., Copenhaver, G., Preuss, D., Sussman, M.R. (2002) Characterization of T-DNA insertion sites in *Arabidopsis thaliana* and the implications for saturation mutagenesis, *OMICS* **6**, 163–174.

Laibach, F. (1943) *Arabidopsis thaliana* (L) Heynh. Als objekt für genetische und entwicklungsphysioligische Untersuchungen, *Bot. Archiv.* **44**, 265–455.

Langridge, J. (1955) Biochemical mutations in the crucifer *Arabidopsis thaliana* (L.) Heynh, *Nature* **176**, 260–261.

Langridge, J., Griffing, B. (1959) A study of high temperature lesions in *Arabidopsis thaliana*, *Aust. J. Biol. Sci.* **12**, 117–135.

Laux, T., Mayer, K.F., Berger, J., Jurgens, G. (1996) The *WUSCHEL* gene is required for shoot and floral meristem integrity in *Arabidopsis*, *Development* **122**, 87–96.

Lawrence, C.W. (1968) Radiation induced polygenic mutation in *Arabidopsis thaliana* (L) Heynh: selection for flowering time, *Heredity* **23**, 321–337.

Lebel, E.G. (1994) *Gene Targeting in Plants: Induction of Cellular Recombination and Transposons as Targeting Vectors*, PhD Thesis, DISS, ETH no. 10696, Federal Institute of Technology, Zurich.

Lee, K.Y., Lund, P., Lowe, K., Dunsmuir, P. (1990) Homologous recombination in plant cells after *Agrobacterium*-mediated transformation, *Plant Cell* **2**, 415–425.

Leutwiler, L.S., Hough-Evans, B.R., Meyerowitz, E.M. (1984) The DNA of *Arabidopsis thaliana*, *Mol. Gen. Genet.* **194**, 15–23.

Lin, X., Kaul, S., Rounsley, S., Shea, T.P., Benito, M.I., Town, C.D., Fujii, C.Y., Mason, T., Bowman, C.L., Barnstead, M., Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J., Ronning, C.M., Koo, H.L., Moffat, K.S., Cronin, L.A., Shen, M., Pai, G., Aken, S.V., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D., Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M., Venter, J.C. (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*, *Nature* **402**, 761–768.

Lippman, Z., Gendrel, A.-V., Blanck, M., Vaughn, M.W., Dehia, N., McCombie, R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., Carrington, J.C., Doerge, R.W., Colot, V., Martienssen, R. Role of transposable elements in heterochromatin and epigenetic control, *Nature* **430**, 471–476.

Lister, C., Dean, C. (1993) recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*, *Plant J.* **4**, 745–750.

Lucas, W.J., Bouche-Pillon, S., Jackson, D.P., Nguyen, L., Baker, L., Ding, B., Hake, S. (1995) Selective trafficking of KNOTTED1 homeodomain protein and its mRNA through plasmodesmata, *Science* **270**, 1980–1983.

Lukens, L., Zou, F., Lydiate, D., Parkin, I., Osborn, T. (2003) Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*, *Genetics* **164**, 359–372.

Lukowitz, W., Gillmor, C.S., Scheible, W.-R. (2000) Positional cloning in *Arabidopsis*. Why it feels good to have a genome initiative working for you, *Plant Physiol.* **123**, 795–805.

Lysak, M.A., Pecinka, A., Schubert, I. (2003) Recent progress in chromosome painting of Arabidopsis and related species, *Chromosome Res.* **11**, 195–204.

Ma, J., Bennetzen, J.L. (2004) Rapid recent growth and divergence of rice nuclear genomes, *Proc. Natl. Acad. Sci. U.S.A.* **10**, 12404–12410.

Malicki, J., Schugart, K., McGinnis, W. (1990) Mouse Hox-2.2 specifies thoracis segmental identity in *Drosophila* embryos and larvae, *Cell* **63**, 961–967.

Marra, M., Kucaba, T., Sekhon, M., Hillier, L., Martienssen, R., Chinwalla, A., Crockett, J., Fedele, J., Grover, H., Gund, C., McCombie, W.R., McDonald, K., McPherson, J., Mudd, N., Parnell, L., Schein, J., Seim, R., Shelby, P., Waterston, R., Wilson, R. (1999) zA map for sequence analysis of the *Arabidopsis thaliana* genome, *Nat. Genet.* **22**, 265–270.

Matsuhara, S., Jingu, F., Takahashi, T., Komeda, Y. (2000) Heat-shock tagging: a simple method for the expression and isolation of plant genome DNA flanked by T-DNA insertions, *Plant J.* **22**, 79–86.

McCallum, C.M., Comai, L., Greene, E.A., Henikoff, S. (2000) Targeted screening for induced mutations, *Nat. Biotechnol.* **18**, 455–457.

McGinnis, N., Kuziora, M.A., McGinnis, W. (1990) Human Hox-4.2 and *Drosophila* deformed encode similar regulatory specificities in *Drosophila* embryonic larvae, *Cell* **63**, 969–976.

Messenguy, F., Dubois, E. (2003) Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development, *Gene* **316**, 1–21.

Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A., Matzke, A.J. (2000) Transcriptional silencing and promoter methylation triggered by double-stranded RNA, *EMBO J.* **19**, 5194–5201.

Meyerowitz, E.M. (2002) Plants compared to animals: the broadest comparative study of development, *Science* **295**, 1482–1485.

Miao, Z.-H., Lam, E. (1995) Targeted disruption of the TGA3 locus in *Arabidopsis thaliana*, *Plant J.* **7**, 359–365.

Minsung, K., McCormick, S., Timmermans, M., Neelima, S. (2003) The expression domain of *PHANTASTICA* determines leaflet placement in compound leaves, *Nature* **424**, 438–443.

Mittelsten-Scheid, O., Afsar, K., Paszkowski, J. (1998) Release of epigenetic gene silencing by trans-acting mutations in *Arabidopsis*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 632–637.

Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., Kakutani, T. (2001) Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*, *Nature* **411**, 212–214.

Miyashita, N.T., Kawabe, A., Innan, H. (1999) DNA variation in the wild plant *Arabidopsis thaliana* revealed by amplified fragment length polymorphism analysis, *Genetics* **152**, 1723–1731.

Mizukami, Y., Ma, H. (1992) Ectopic expression of the floral homeotic gene *AGAMOUS* in transgenic *Arabidopsis* plants alters floral organ identity, *Cell* **71**, 119–131.

Mizukami, Y., Ma, H. (1995) Separation of AG function in floral meristem determinacy from that in reproductive organ identity by expressing antisense *AG* RNA, *Plant Mol. Biol.* **28**, 767–784.

Mourier, T., Hansen, A.J., Willerslev, E., Arctander, P. (2001) The human genome project reveals a continuous transfer of large mitochondrial fragments to the nucleus, *Mol. Biol. Evol.* **18**, 1833–1837.

Mozo, T., Dewar, K., Dunn, P., Ecker, J.R., Fischer, S., Kloska, S., Lehrach, H., Marra, M., Martienssen, R., Meier-Ewert, S., Altmann, T. (1999) A complete BAC-based physical map of the *Arabidopsis thaliana* genome, *Nat. Genet.* **22**, 271–275.

Murata, M., Heslop-Harrison, J.S., Motoyoshi, F. (1997) Physical mapping of the 5S ribosomal RNA genes in *Arabidopsis thaliana* by multicolor fluorescence in situ hybridization with cosmid clones, *Plant J.* **12**, 31–37.

Muskett, P.R., Clissold, L., Marocco, A., Springer, P.S., Martienssen, R., Dean, C. (2003) A resource of mapped dissociation launch pads for targeted insertional mutagenesis in the *Arabidopsis* genome, *Plant Physiol.* **132**, 506–516.

Nagaki, K., Talbert, P.B., Zhong, C.X., Dawe, R.K., Henikoff, S., Jiang, J. (2003) Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres, *Genetics* **163**, 1221–1225.

Nakazono, M., Qiu, F., Borsuk, L.A., Schnable, P.S. (2003) Laser-capture microdissection, a tool for the global analysis of gene

expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize, *Plant Cell* **15**, 583–596.

Napoli, C., Lemieux, C., Jorgensen, R. (1990) Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous gene *in trans*, *Plant Cell* **2**, 279–289.

Nathan, D., Sterner, D.E., Berger, S.L. (2003) Histone modifications: now summoning sumoylation, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13118–13120.

Odell, J.T., Nagy, F., Chua, N.H. (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter, *Nature* **313**, 810–812.

Offringa, R., de Groot, M.J., Haagsman, H.J., Does, M.P., van den Elzen, P.J., Hooykaas, P.J. (1990) Extrachromosomal homologous recombination and gene targeting in plant cells after *Agrobacterium*-mediated transformation, *EMBO J.* **9**, 3077–3084.

Page, D.R., Köhler, C., da Costa-Nunes, J.A., Baroux, C., Moore, J.M., Grossniklaus, U. (2004) Intrachromosomal excision of a hybrid *Ds* element induces large genomic deletions in *Arabidopsis*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2969–2974.

Parenicova, L., de Folter, S., Kieffer, M., Horner, D.S., Favalli, C., Busscher, J., Cook, H.E., Ingram, R.M., Kater, M.M., Davies, B., Angenent, G.C., Colombo, L. (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world, *Plant Cell* **15**, 1538–1551.

Parinov, S., Sundaresan, V. (2000) Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome sequencing project, *Curr. Opin. Biotechnol.* **11**, 157–161.

Parinov, S., Sevugan, M., Ye, D., Yang, W.C., Kumaran, M., Sundaresan, V. (1999) Analysis of flanking sequences from dissociation insertion lines: a database for reverse genetics in *Arabidopsis*, *Plant Cell* **11**, 2263–2270.

Paszkowski, J., Baur, M., Bogucki, A., Potrykus, I. (1988) Gene targeting in plants, *EMBO J.* **7**, 4021–4026.

Pien, S., Wyrzykowska, J., McQueen-Mason, S., Smart, C., Fleming, A. (2001) Local expression of expansin induces the entire process of leaf development and modifies leaf shape, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10981–10982.

Poirier, Y., Nawrath, C., Somerville, C. (1995) Production of polyhydroxyalkanoates, a family of biodegradable plastics and elastomers, in bacteria and plants, *Biotechnology* **13**, 142–150.

Prins, M., Kikkert, M., Ismayadi, C., de Graauw, W., de Haan, P., Goldbach, R. (1997) Characterization of RNA-mediated resistance to tomato spotted wilt virus in transgenic tobacco plants expressing NS(M) gene sequences, *Plant Mol. Biol.* **33**, 235–243.

Pruitt, R.E., Meyerowitz, E.M. (1986) Characterization of the genome of *Arabidopsis thaliana*, *J. Mol. Biol.* **187**, 169–183.

Quatrano, R.S., Assmann, S.A., Sanders, D., Eckhardt, N.A. (2002) Signal transduction, *Plant Cell* **14**(Suppl.), S1–S417.

Redei, G.P. (1962) Genetic block of 'vitamin thiazole' synthesis in *Arabidopsis*, *Genetics* **47**, 979.

Reinhardt, D., Pesce, E.R., Stieger, P., Mandel, T., Baltensperger, K., Bennett, M., Traas, J., Friml, J., Kuhlemeier, C. (2003) Regulation of phyllotaxis by polar auxin transport, *Nature* **426**, 255–260.

Reiss, B., Schubert, I., Kopchen, K., Wendeler, E., Schell, J., Puchta, H. (2000) RecA stimulates sister chromatid exchange and the fidelity of double-strand break repair, but not gene targeting, in plants transformed by *Agrobacterium*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3358–3363.

Rice Chromosome 10 Sequencing Consortium. (2003) In-depth view of structure, activity, and evolution of rice chromosome 10, *Science* **300**, 1566–1569.

Richly, E., Kurth, J., Leister, D. (2002) mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution, *Mol. Biol. Evol.* **19**, 76–84.

Richter, T.E., Ronald, P.C. (2000) The evolution of disease resistance genes, *Plant Mol. Biol.* **42**, 195–204.

Risseeuw, E., Offringa, R., Franke-van Dijk, M.E.I., Hooykaas, P.J.J. (1995) Targeted recombination in plants using *Agrobacterium* coincides with additional rearrangements at the target locus, *Plant J.* **7**, 109–119.

Rong, Y.S., Golic, K.G. (2000) Gene targeting by homologous recombination in *Drosophila*, *Science* **288**, 2013–2018.

Samson, F., Brunaud, V., Balzergue, S., Dubreucq, B., Lepiniec, L., Pelletier, G., Caboche, M., Lecharny, A. (2002) FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of *Arabidopsis thaliana* T-DNA transformants, *Nucleic Acids Res.* **30**, 94–97.

Schaefer, D.G., Zryd, J.P. (1997) Efficient gene targeting in the moss *Physcomitrella patens*, *Plant J.* **11**, 1195–1206.

Schaffer, R., Ramsay, N., Samach, A., Corden, S., Putterill, J., Carre, I.A., Coupland, G. (1999) The *late elongated hypocotyl* mutation of *Arabidopsis* disrupts circadian rhythms and the photoperiodic control of flowering, *Cell* **93**, 1219–1229.

Schmidt, R., West, J., Love, K., Lenehan, Z., Lister, C., Thompson, H., Bouchez, D., Dean, C. (1995) Physical map and organization of *Arabidopsis thaliana* chromosome 4, *Science* **270**, 480–483.

Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., taji, T., Yamaguchi-Shinozaki, K., Carnini, P., Kawai, J., Hayashizaki, Y., Shinozaki, K. (2002) Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold, and high-salinity stresses using a full-length cDNA microarray, *Plant J.* **31**, 279–292.

Sessions, A., Yanofsky, M.F., Weigel, D. (2000) Cell-cell signaling and movement of the floral transcription factors LEAFY and APETALA1, *Science* **289**, 779–782.

Sieburth, L.E., Drews, G.N., Meyerowitz, E.M. (1998) Non-autonomy of *AGAMOUS* function in flower development: use of *Cre/loxP* method for mosaic analysis in *Arabidopsis*, *Development* **125**, 4303–4312.

Singer, T., Yordan, C., Martienssen, R.A. (2001) Robertson's *Mutator* transposons in *A. thaliana* are regulated by the chromatin-remodeling gene *Decrease in DNA Methylation* (*DDM1*), *Genes Dev.* **15**, 591–602.

Skarnes, W.C., von Melchner, H., Wurst, W., Hicks, G., Nord, A.S., Cox, T., Young, S.G., Ruiz, P., Soriano, P., Tessier-Lavigne, M., Conklin, B.R., Stanford, W.L., Rossant, J. (2004) A public gene trap resource for mouse functional genomics, *Nat. Genet.* **36**, 543–544.

Smith, D.L., Fedoroff, N.V. (1995) LRP1, a gene expressed in lateral and adventitious root primordia of *Arabidopsis*, *Plant Cell* **7**, 735–745.

Smith, C.J., Watson, C.F., Bird, C.R., Ray, J., Schuch, W., Grierson, D. (1990) Expression of a truncated tomato polygalacturonase gene inhibits expression of the endogenous gene in transgenic plants, *Mol. Gen. Genet.* **224**, 477–481.

Smith, D., Yanai, Y., Liu, Y.-G., Ishiguro, S., Okada, K., Shibata, D., Whittier, R.F., Federoff, N.V. (1996) Characterization and mapping of *Ds*-GUS-T-DNA lines for targeted insertional mutagenesis, *Plant J.* **10**, 721–732.

Speulman, E., Metz, P.L., van Arkel, G., te Lintel Hekkert, B., Stiekema, W.J., Pereira, A. (1999) A two-component enhancer-inhibitor transposon mutagenesis system for functional analysis of the *Arabidopsis* genome, *Plant Cell* **11**, 1853–1866.

Spillane, C., Steimer, A., Grossniklaus, U. (2001) Apomixis in agriculture: the quest for clonal seeds, *Sex Plant Reprod.* **14**, 179–187.

Springer, P.S., McCombie, W.R., Sundaresan, V., Martienssen, R.A. (1995) Gene trap tagging of PROLIFERA, an essential MCM2-3-5-like gene *in Arabidopsis*, *Science* **268**, 877–880.

Stegemann, S., Hartmann, S., Ruf, S., Bock, R. (2003) High-frequency gene transfer from the chloroplast genome to the nucleus, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8828–8833.

Steinitz-Sears, L.M., Lee-Chen, S. (1970) Cytogenetic studies in *Arabidopsis thaliana*, *Can. J. Genet. Cytol.* **12**, 217–223.

Stupar, R.M., Lilly, J.W., Town, C.D., Cheng, Z., Kaul, S., Buell, C.R., Jiang, J. (2001) Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: implication of potential sequencing errors caused by large-unit repeats, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5099–5103.

Swinburne, J., Balcells, L., Scofield, S.R., Jones, J.D., Coupland, G. (1992) Elevated levels of activator transposase mRNA are associated with high frequencies of dissociation excision in *Arabidopsis*, *Plant Cell* **4**, 583–595.

Talbert, P.B., Masuelli, R., Tyagi, A.P., Comai, L., Henikoff, S. (2002) Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant, *Plant Cell* **14**, 1053–1066.

Tantikanjana, T., Mikkelsen, M.D., Hussain, M., Halkier, B.A., Sundaresan, V. (2004) Functional analysis of the tandem-duplicated P450 genes SPS/BUS/CYP79F1 and CYP79F2 in glucosinolate biosynthesis and plant development by *Ds* transposition-generated double mutants, *Plant Physiol.* **135**, 840–848.

Timmerman, M.C.P., Hudson, A., Becraft, P.W., Nelson, T. (1999) *ROUGH SHEATH2*: a *Myb* protein that represses *knox* homeobox genes in maize lateral organ primordia, *Science* **284**, 151–153.

Tissier, A.F., Marillonnet, S., Klimyuk, V., Patel, K., Torres, M.A., Murphy, G., Jones, J.D. (1999) Multiple independent defective suppressor-mutator transposon insertions in *Arabidopsis*: a tool for functional genomics, *Plant Cell* **11**, 1841–1852.

Topping, J.F., Wei, W., Lindsey, K. (1991) Functional tagging of regulatory elements in the plant genome, *Development* **112**, 1009–1019.

van der Krol, A.R., Mur, L.A., Beld, M., Mol, J.N.M., Stuitje, A.R. (1990) Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression, *Plant Cell* **2**, 291–299.

Vielle-Calzada, J.-P., Moore, J.M., Gagliano, W.B., Grossniklaus, U. (1998) Altering sexual development in *Arabidopsis*, *J. Plant Biol.* **41**, 71–83.

Vongs, A., Kakutani, T., Martienssen, R.A., Richards, E.J. (1993) *Arabidopsis thaliana* DNA methylation mutants, *Science* **260**, 1926–1928.

Waites, R., Selvandurai, H.R., Oliver, I.R., Hudson, A. (1998) The *PHANTASTICA* gene encodes a *MYB* transcription factor involved in growth and dorsoventrality in lateral organs in *Antirrhinum*, *Cell* **93**, 779–789.

Warzecha, H., Mason, H.S. (2003) Benefits and risks of antibody and vaccine production in transgenic plants, *J. Plant. Physiol.* **160**, 755–764.

Wassenegger, M., Heimes, S., Riedel, L., Sanger, H.L. (1994) RNA-directed *de novo* methylation of genomic sequences in plants, *Cell* **76**, 567–576.

Waterhouse, P.M., Helliwell, C.A. (2003) Exploring plant genomes by RNA-induced gene silencing, *Nat. Rev. Genet.* **4**, 29–38.

Weigel, D., Ahn, J.H., Blazquez, M.A., Borevitz, J.O., Christensen, S.K., Fankhauser, C., Ferrandiz, C., Kardailsky, I., Malancharuvil, E.J., Neff, M.M., Nguyen, J.T., Sato, S., Wang, Z.Y., Xia, Y., Dixon, R.A., Harrison, M.J., Lamb, C.J., Yanofsky, M.F., Chory, J. (2000) Activation tagging in Arabidopsis, *Plant Physiol.* **122**, 1003–1013.

Wesley, S.V., Helliwell, C.A., Smith, N.A., Wang, M.B., Rouse, D.T., Liu, Q., Gooding, P.S., Singh, S.P., Abbott, D., Stoutjesdijk, P.A., Robinson, S.P., Gleave, A.P., Green, A.G., Waterhouse, P.M. (2001) Construct design for efficient, effective and high-throughput gene silencing in plants, *Plant J.* **27**, 581–590.

Wicker, T., Yahiaoui, N., Guyot, R., Schlagenhauf, E., Liu, Z.D., Dubcovsky, J., Keller, B. (2003) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and Am genomes of wheat, *Plant Cell* **15**, 1186–1197.

Wilson, K., Long, D., Swinburne, J., Coupland, G. (1996) A *Dissociation* insertion causes a semidominant mutation that increases expression of *TINY*, an *Arabidopsis* gene related to *APETALA2*, *Plant Cell* **8**, 659–671.

Wilson, C., Pearson, R.K., Bellen, H.J., O'Kane, C.J., Grossniklaus, U., Gehring, W. (1989) P-element mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in *Drosophila*, *Genes Dev.* **3**, 1301–1313.

Wisman, E., Cardon, G.H., Fransz, P., Saedler, H. (1998) The behaviour of the autonomous maize transposable element *En/Spm* in *Arabidopsis thaliana* allows efficient mutagenesis, *Plant Mol. Biol.* **37**, 989–999.

Wolfe, K.H., Gouy, M., Yang, Y.-W., Sharp, P.M., Li, W.-H. (1989) Date of the monocot-dicot divergence estimated from the chloroplast DNA sequence data, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6201–6205.

Wong, G.K., Passey, D.A., Huang, Y., Yang, Z., Yu, J. (2000) Is "junk" DNA mostly intron DNA? *Genome Res.* **10**, 1672–1678.

Wilke, D. (1999) Chemicals from biotechnology: molecular plant genetics will challenge the chemical and fermentation industry, *Appl. Microbiol. Biotechnol.* **52**, 135–145.

Wyrzykowska, J., Pien, S., Shen, W.H., Fleming, A.J. (2002) Manipulation of leaf shape by modulation of cell division, *Development* **129**, 957–964.

Zachgo, E.A., Wang, M.L., Dewdney, J., Bouchez, D., Camilleri, C., Belmonte, S., Huang, L., Dolan, M., Goodman, H.M. (1996) A physical map of chromosome 2 of *Arabidopsis thaliana*, *Genome Res.* **6**, 19–25.

Zhang, Y., Tessaro, M.J., Lassner, M., Li, X. (2003) Knockout analysis of *Arabidopsis*

transcription factors *TGA2, TGA5* and *TGA6* reveal their redundant and essential roles in systemic acquired resistance, *Plant Cell* **15**, 2647–2653.

Zuo, J., Niu, Q.-W., Frugis, G., Chau, N.-H. (2002) The WUSCHEL gene promotes vegetative-to-embryonic transition in *Arabidopsis, Plant J.* **30**, 349–359.

The Evolution of the Protein Repertoire: *see* Protein Repertoire, Evolution of

Theoretical Bioinformatics: *see* Systems Biology

Theoretical Molecular Biology: *see* Pragmatic Computational Biology: Sequence Analysis and Biological Systems Description

Therapeutic Angiogenesis

Mary E. Gerritsen
Exelixis, South San Francisco, CA

1 Revascularization 286

2 The Principal Players 287
2.1 Cells of the Vessel Wall 287
2.2 Growth Factors and Their Receptors 289
2.2.1 VEGF Family 289
2.2.2 FGF Family 293
2.2.3 PDGF Family 294
2.2.4 TGF-β Family 296
2.2.5 HGF 297
2.2.6 Angiopoietins 298
2.3 Inflammation 299
2.4 Circulating Endothelial Precursors/Bone Marrow-derived Cells 300
2.5 Mechanical Factors 301
2.6 Hypoxia 301
2.7 Vascular Remodeling 302

3 Animal Models of Therapeutic Angiogenesis 302

4 Clinical Trials 303
4.1 Protein-based Trials 303
4.2 Gene Therapy-based Clinical Trials 305

5 Going Forward 306
5.1 Appropriate Animal Model with Accurate Readout of Arteriogenesis 306
5.2 The Right Growth Factor 307
5.3 The Role/Utility of Bone Marrow-derived Cells 307
5.4 Drug Delivery 308
5.5 New Approaches 309

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

5.6 Patient Selection 309
5.7 Clinical Endpoints 310

6 Conclusion 310

Bibliography 311
Books and Reviews 311
Primary Literature 311

Keywords

Angiogenesis
The formation of new blood vessels.

Arteriogenesis
The growth of arterioles (preformed collaterals) into mature arteries.

Ischemia
A reduction in the blood supply to a bodily organ, tissue, or part; usually caused by constriction or obstruction of the blood vessels.

Myocardial Infarction
Destruction of heart tissue resulting from obstruction of the blood supply to the heart muscle.

Peripheral Vascular Disease (PVD)
A disease of the large blood vessels of the extremities. PVD may occur when major blood vessels in these areas are blocked and do not receive enough blood.

Reperfusion
The restoration of blood flow to an organ or tissue that has had its blood supply cut off, as after a heart attack.

Revascularization
A broad term that is usually used to describe surgical and catheter procedures that are used to restore blood flow to the heart or another ischemic organ. In the context of therapeutic angiogenesis, this term also encompasses the use of drugs, gene therapy, and stem cells.

Therapeutic Angiogenesis
The use of drugs or other therapeutic modalities to stimulate the revascularization of ischemic tissue.

Vasculogenesis
The *de novo* formation of blood vessels from angiopoeitic precursor cells.

Therapeutic angiogenesis refers to the use of drugs or other therapeutic modalities to stimulate the revascularization of ischemic tissue. The concept was first introduced by the German gynecologist Michael Höckel in 1989 as a therapeutic approach to improve regional tissue perfusion and viability following surgery. In the intervening two decades, there has been a virtual explosion of information relating to the growth factors and receptors involved, the molecular mechanisms of new vessel formation, the development of numerous animal models, and a number of clinical trials. The two major clinical indications are the coronary artery (ischemic heart disease) and peripheral vascular disease, but therapeutic angiogenesis could also be applied to wound healing.

Ischemic heart disease is the leading cause of death and the major contributor to morbidity in the Western world. This problem will continue to grow as the population ages and will be even more severe if the ''epidemics'' of obesity, metabolic syndrome, and diabetes continue to develop unchecked. In the United States alone in 1999 there were 1.1 million percutaneous coronary revascularizing interventions, 355 000 coronary bypass surgeries, and 131 000 endarterectomies. Annually, about 1.2 million Americans will have a first or recurrent coronary attack and roughly 502 000 of these people will die −340 000 of them in an ER or before they reach a hospital. It is estimated that 7.8 million Americans aged 20 years and older have survived a heart attack and about 6.8 million Americans have angina pectoris (chest pain or discomfort due to reduced blood supply to the heart). Moreover, within six years of a recognized heart attack, 18% of men and 35% of women will have a second heart attack. A significant number of these patients will die or be disabled with heart failure or the aftermath of a stroke. In the United States, coronary heart disease is the number one cause of premature permanent disability (19% of total) in the labor force (American Heart Association Heart Disease and Stroke Statistics – 2004 Update). Peripheral ischemic vascular disease is also a significant cause of major morbidity, affecting up to 14 to 18 million patients in the United States. In these patients, intermittent claudication (pain while walking that abates during rest) is the most common symptom. Other symptoms include numbness or weakness in the legs, aching pain in the feet or toes while at rest, nonhealing ulcers on the leg or foot, cold legs or feet, and skin color changes of the legs or feet. However, less than 20% of the patients with peripheral vascular disease have been diagnosed, and of those that are, many are often undertreated. Of these patients, more than 70% experience no change or any improvement in symptoms after 5 to 10 years of conservative management. However, 20 to 30% of patients develop more severe symptoms that require intervention, and a small fraction (less than 10%) will require amputation. Moreover, the 5 and 10-year rates of mortality can reach 30 and 50% respectively.

1
Revascularization

There are several interrelated processes that contribute to revascularization (Table 1). *Angiogenesis* refers to the development of new vessels; in the context of "therapeutic angiogenesis," angiogenesis would be used to describe the proliferation of new capillaries in ischemic vascular beds. The primary driver of this angiogenesis is thought to be tissue hypoxia, which occurs as a consequence of poor tissue perfusion. The hypoxia sensor is believed to be a complex containing the oxygen-sensitive hypoxia inducible factor-1α (HIF-1α) subunit, which is rapidly destroyed in the presence of oxygen. In a reaction requiring oxygen and iron, a proline hydroxylase hydroxylates HIF-1α and makes it susceptible to ubiquitination by the von Hippel–Lindau protein and to subsequent degradation in the proteasome (Fig. 1). Genes of relevance to angiogenesis regulated by HIF-1α include vascular endothelial growth factor (VEGF), the VEGF receptor *flt*-1 (fms like tyrosine kinase-1; VEGFR-1), and angiopoietin-2 (Ang2). The products of these genes predominantly activate VEGF-dependent endothelial cell proliferation; interestingly the signals by other growth factors (e.g. fibroblast growth factors (FGFs), platelet-derived growth factors (PDGFs)) that are required for smooth muscle cell stimulation (necessary for support and stabilization of larger vessels) are not driven by hypoxia. However, the local growth of capillaries cannot compensate for ischemia because of the proximal occlusion of a major arterial vessel by an atheroma. Moreover, it is unclear whether tissue ischemia is really the primary driver of spontaneous revascularization.

Vasculogenesis is defined as the *de novo* formation of blood vessels from angiopoeitic precursor cells. This process is better known in embryonic development, where it describes the *de novo* formation

Tab. 1 Processes involved in new vessel formation and remodeling.

	Vasculogenesis	***Arteriogenesis***	***Angiogenesis***
Cell types involved	Endothelial precursor cells	Endothelial cells, smooth muscle cells, pericytes, monocytes	Endothelial cells pericytes
Primary stimulus	Developmental process; in adult not known	Inflammation shear stress	Inflammation ischemia
Product	Fully formed vessels (in embryo)	Arterioles/conduit arteries	Capillaries
Occurs in adult tissue	Minimal; unclear how important	Yes	Yes
Contribution to effective perfusion	Not clear; may contribute	Major contributor	Minor; important for local perfusion
Growth factors involved	VEGF, Ang-1, Ang-2, SDF1	PDGF, Ang-1, Ang-2, FGFs, MCP1	FGF1, FGF2, FGF4, FGF5, VEGF1, VEGF2, VEGF3, HGF, TGFβ

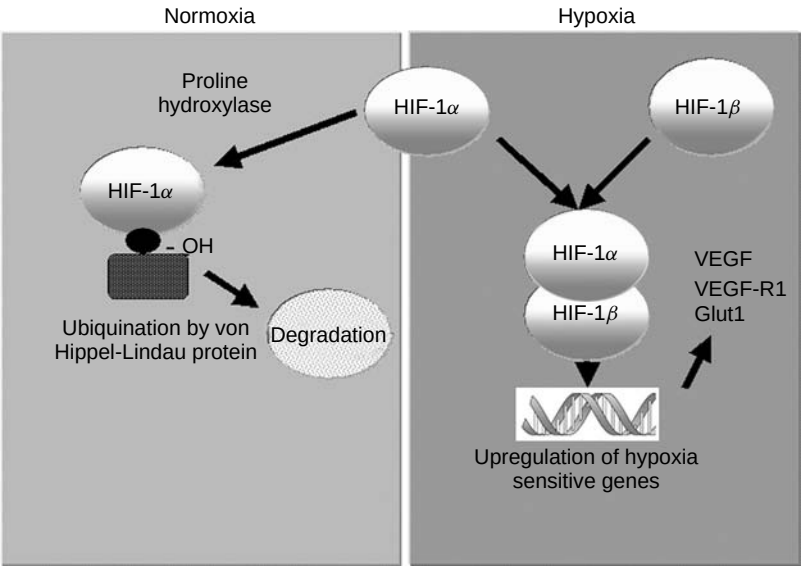

Fig. 1 The hypoxia sensor. The α-subunit of the transcription factor hypoxia-inducible factor 1 (HIF-1α) is hydroxylated by proline hydroxylase in the presence of oxygen and iron. It then undergoes ubiquitination by the von Hippel–Lindau protein and is then degraded by the proteasome. In contrast, under conditions of hypoxia, HIF-1α-levels increase and can interact with HIF-1β. The HIF-1 heterodimer can then bind to specific response elements on hypoxia-regulatable genes (e.g. VEGF, VEGFR-1, Glut1), leading to the upregulation expression.

of the vascular system from mesenchymal stem cells that form blood islands. However, there are a number of recent studies that suggest that bone marrow-derived cells could contribute to new vessel formation in the context of therapeutic angiogenesis.

Arteriogenesis by definition describes the growth of arterioles (preformed collaterals) into mature arteries. It may also be more broadly interpreted as the growth of new, larger vessels. Interestingly, the history of the study of collaterals and their functional importance predates early studies of angiogenesis and vasculogenesis by two to three centuries! Collaterals were first described by the English physician Richard Lower who pioneered transfusions from dog to dog and from lamb to man. He described coronary artery anastomoses on the basis of the observation that fluid injected into one artery flows into the other and concluded that "there is a need of vital heat and nourishment, so deficiency of these is very fully guarded against by such anastomoses".

2 The Principal Players

2.1 Cells of the Vessel Wall

Newly formed vessels as well as those that remodel are comprised of several basic cell types, each of which has specific roles and functions in the development of the vasculature. Endothelial cells, which

line the inner surface of all blood vessels, have received the most attention with respect to their role in angiogenesis. These cells, which are normally quiescent and nondividing, can undergo a rapid phenotypic switch in response to growth factors and inflammatory stimuli. When activated, the endothelial cells upregulate adhesion molecules for inflammatory cells, secrete chemokines that recruit inflammatory cells and possibly stem cells to sites of activation, and secrete metalloproteinases and other proteases that play a role in the breakdown of the extracellular matrix and liberation of additional growth factors either bound to the matrix or present in latent forms. Following the degradation of the basement membrane, endothelial cells then develop sprouts from preexisting microvessels, invade the extracellular matrix, form tubes, and connect the tips of these tubes to create loops capable of handling blood.

Periendothelial cells are required for the final stages of vessel maturation. At the level of the capillary, this role is filled by pericytes, a perivascular cell with multifunctional activities. The pericytes extend long cytoplasmic processes over the surface of the endothelial cells. Endothelial cells and pericytes make interdigitating contacts and can also communicate through the formation of gap junctions. There are differences in pericyte morphology and the degree of coverage in different vascular beds. However, the pericytes appear to play an important role in protecting new endothelium-lined vessels against rupture or regression. In the developing retinal vasculature, for example, vessels regress more easily as long as they are not covered by pericytes, and in diabetes, the loss of pericytes around the retinal vasculature is associated with aneurysms, vasodilatation, and bleeding.

Vascular smooth muscle cells are highly specialized cells, which in mature arteries have the principal function of contraction and regulation of the blood vessel tone. Normally, these cells proliferate at a very low rate and exhibit a low synthetic activity. These cells express unique contractile proteins and ion channels. Highly specific signaling molecules coordinate the contractile functions of these cells. As new vessels form, vascular smooth muscle cells play a key role in the overall morphology. During this process, or when vessels remodel, vascular smooth muscle cells exhibit a high rate of proliferation and synthesize extracellular matrix molecules such as collagen, elastin, and proteoglycans.

The formation of larger vessels can occur by the enlargement of preexisting collaterals (discussed below), or possibly through the *de novo* growth of collateral vessels. There is some argument that so-called "*de novo*" collaterals may be an artifact of the resolution limits of angiography. Collateral vessels on angiograms are thought to be new vessels simply because they were not visible previously. However, there may not have been enough flow through collateral vessels for them to be filled with the contrast reagent, and thus the abundance of preexisting collateral vessels has likely been seriously underestimated. It is possible that newly formed capillaries could remodel into arterioles by recruiting smooth muscle cells. The most convincing evidence for the formation of *de novo* collateral vessels comes from observations in pigs and dogs where relatively large, angiographically visible collateral vessels develop in the connective tissue that forms around ameroid occluders (although these could still be enlarged *vasa vasorum* or small blood vessels ramifying on the outside of a major artery or vein).

2.2 Growth Factors and Their Receptors

2.2.1 VEGF Family

Vascular endothelial growth factor (VEGF; also called VEGF-A to distinguish it from other VEGF family members) is probably the best characterized and certainly the most "popular" (in terms of number of publications on the subject!) of the angiogenic growth factors.

The VEGF family members and their receptors are depicted in Fig. 2. There are three VEGF receptors (VEGFR-1 to VEGFR-3), which interact with the VEGF family members (VEGF-A through -E, as well as the related placental growth factor, PlGF). Each of the VEGFs has a characteristic receptor-binding pattern (Fig. 2).

Placental growth factor was originally discovered, as its name implies, in the placenta. However, PlGF is found in many tissues and is produced by many cell types including vascular smooth muscle cells, inflammatory cells, bone marrow cells, neurons, and tumor cells. Although quiescent endothelial cells release minimal amounts of PlGF, activation by angiogenic stimuli such as VEGF or HGF markedly upregulates PlGF. PlGF can be alternatively spliced, and to date, four different isoforms have been described which differ in size and their binding properties (PlGF-1 ($PlGF_{131}$), PlGF-2 ($PlGF_{153}$), PlGF-3 ($PlGF_{203}$), and PlGF-4 ($PlGF_{224}$)). PlGF-2 and -4 both contain an insertion of 21 amino acids at the carboxy terminus,

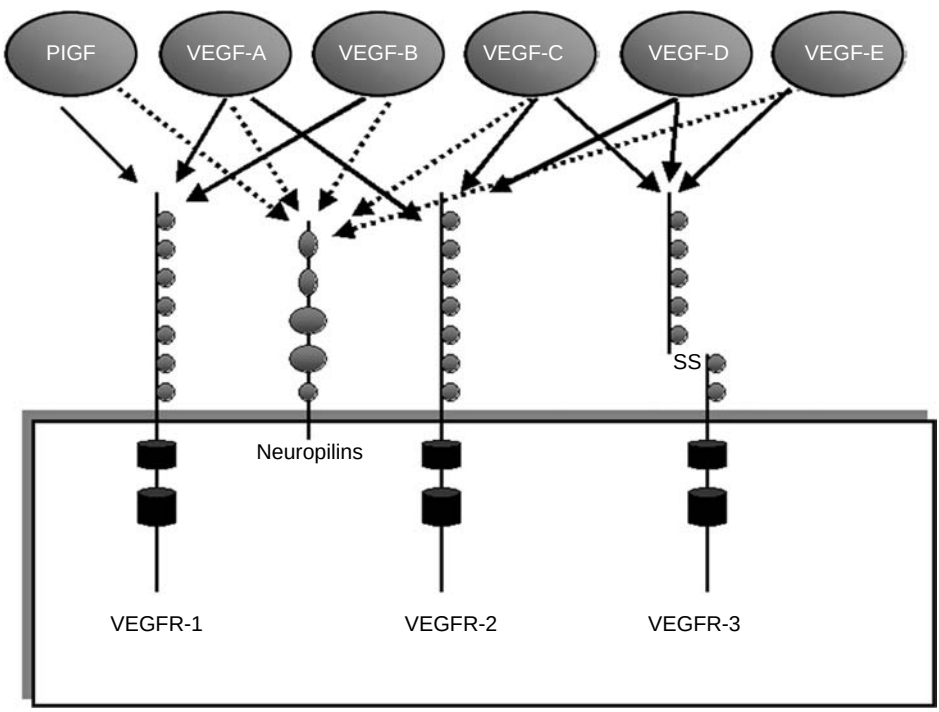

Fig. 2 Binding of the various VEGF family members to the VEGF tyrosine kinase receptors (VEGFR-1, VEGFR-2, and VEGFR-3) and the neuropilins.

which confers the ability to bind polyanionic substances such as heparan sulfate proteoglycans (HSPG) and also to the coreceptors neuropilin 1 and 2. PlGF-1 shares 53% amino acid identity with the PDGF-like region in VEGF, and an overall identity of 42%. However, the recent elucidation of the crystal structure of PlGF showed that the overall structure of human PlGF-1 is remarkably similar to VEGF-A. The physiological roles of PlGF are still unclear, but it may synergize with VEGF-A by forming heterodimers or by potentiating VEGF signaling. PlGF likely elicits responses in cells and tissues that express VEGF-R1 (e.g. smooth muscle cells, monocytes, possibly vascular, and hematopoietic stem cells/progenitors from the bone marrow). PlGF null mice appear to have normal embryonic development but aberrant responses in models of pathological angiogenesis.

VEGF-B shares about 43% amino acid sequence identity with $VEGF_{164}$. VEGF-B binds VEGFR-1 and neuropilins. VEGF-B is also alternatively spliced ($VEGF\text{-}B_{167}$ and $VEGF\text{-}B_{186}$) resulting in a frame shift and two nonhomologous C-termini. In contrast to other VEGF family members, VEGF-B does not contain any N-linked carbohydrates. The shorter isoform of VEGF-B, $VEGF\text{-}B_{167}$, is not glycosylated at all whereas the longer form, $VEGF\text{-}B_{186}$, is O-glycosylated. The function of VEGF-B is really not well characterized. Mice deficient in VEGF-B have a mild cardiac phenotype (defects in atrial conduction, smaller hearts, dysfunctional coronary vasculature, and impaired recovery from cardiac ischemia). VEGF-B has a wide tissue distribution abundant in heart, skeletal muscle, and brown fat. The predominant isoform seems to be $VEGF\text{-}B_{167}$.

VEGF-C and VEGF-D have less homology to VEGF-A, but are mitogenic for vascular endothelial cells *in vitro* (although less potent) and stimulate angiogenesis *in vivo*. VEGF-C and VEGF-D bind VEGFR-2 and VEGFR-3 as well as neuropilins. Interestingly, the expression of VEGFR-3 becomes highly restricted to lymphatic endothelial cells during embryogenesis. VEGF-C and VEGF-D specifically stimulate the proliferation of lymphatic endothelial cells, whereas VEGF-A does not induce any changes in the lymphatic vasculature. Mice deficient in VEGFR-3 die at E9.5 of cardiovascular failure as a result of an impairment in the remodeling and maturation of large vessels. VEGF-C and VEGF-D have long N- and C-terminal extensions that flank the VEGF-homology domain. VEGF-C is synthesized as a precursor protein that undergoes proteolytic processing. The C-terminal domain is cleaved upon secretion, but remains associated with the N-terminal domain by disulfide bonds, thus giving rise to a disulfide-linked tetramer composed of 29-kDa and 31-kDa polypeptides. Upon processing of the N-terminal propeptide, the "mature" form is released, which consists of two 21-kDa polypeptide chains. These polypeptide chains correspond to the VEGF homology domain. Incomplete or additional proteolytic processing can also lead to two other fragments that migrate on reducing SDS-PAGE gels with apparent MW of 15 and 43 kDa. The full-length and so-called "mature" form of VEGF-C will bind to VEGFR-3 with high affinity but high-affinity interactions with VEGFR-2 require proteolytic processing.

The functions of VEGF-A have been extensively characterized and new roles continue to be unveiled. VEGF-A binds to VEGFR-1 and VEGFR-2, as well as to neuropilins 1 and 2. It appears that VEGFR-2 mediates the biological effects of VEGF-A on endothelial cells, whereas VEGFR-1

probably functions as a decoy receptor (discussed below). VEGF-A has the distinction of being one of the few molecules where even the deletion of one allele (e.g. VEGF-A −/+ mice) is embryonically lethal, underscoring the critical importance of this growth factor to vascular development. The biological activities of VEGF-A are primarily restricted to the vascular endothelium, which selectively expresses VEGFR-2. Biological effects of VEGF-A on endothelial cells include stimulation of proliferation, migration, increased vascular permeability, and promotion of survival. However, VEGF-A also can elicit responses in VEGFR-1, expressing cells such as monocytes, smooth muscle cells, dendritic cells, and osteoclasts.

VEGF-A is differentially spliced from eight exons, resulting in a number of different isoforms that are named by their amino acid number: $VEGF_{206}$, $VEGF_{189}$, $VEGF_{183}$, $VEGF_{165}$, $VEGF_{145}$, and $VEGF_{121}$ (Fig. 3). These isoforms vary in their patterns of expression and their binding properties for heparan sulfate proteoglycans, and the coreceptors, neuropilin 1 and 2. For example, $VEGF_{121}$ binds neither HSPG nor the neuropilins unlike $VEGF_{145}$ and $VEGF_{165}$. It is thought that the differences in the interactions with neuropilins and heparan sulfate proteoglycans contribute to the differences in potency of the different isoforms.

Recently, a new series of VEGF isoforms that contain an alternative COOH terminus has been described. In these alternative isoforms, instead of the first 18 nucleotides of exon 8, a more distal splice site is used, resulting in a novel

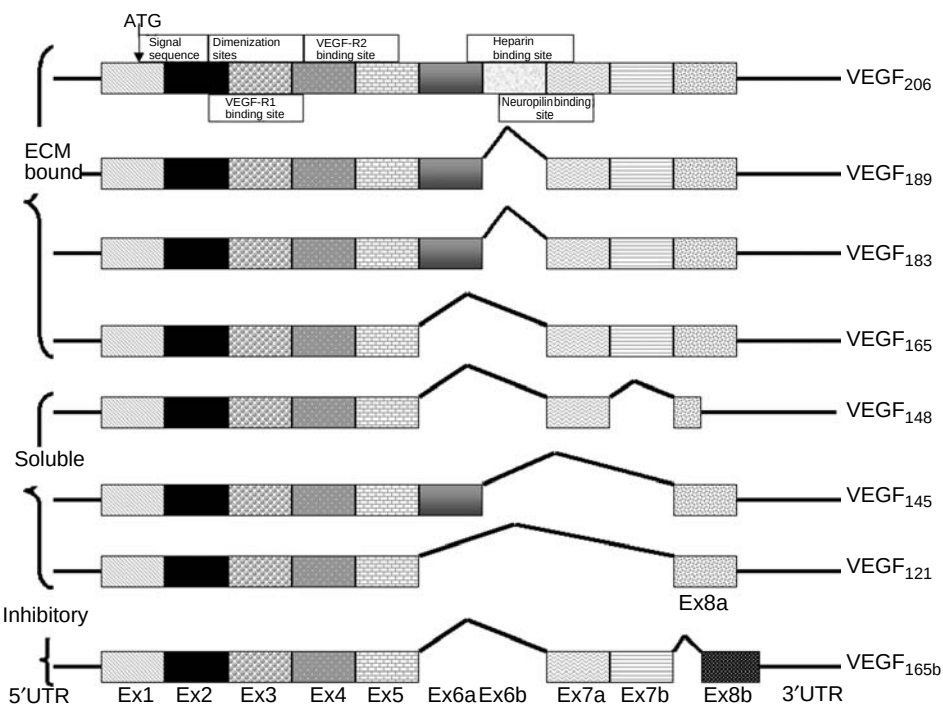

Fig. 3 Alternative splice variants of VEGF-A generating different VEGF-A isoforms.

COOH terminus. These new splice variants have a number of interesting features. Since they contain normal exons 1–5, the isoforms are likely to bind receptors, be secreted and dimerize. However, Keyt and coworkers demonstrated that it was the COOH-terminal of VEGF that conferred biological activity. Thus far, the alternative (so-called "b") isoforms of VEGF have none of the inherent biological activities of VEGF-A (e.g. proliferation, migration) and, moreover, appear to have potent inhibitory actions on VEGF-A mediated activities, presumably by either competing with the receptors or by forming inactive heterodimers. Since the commercial antibodies, cDNA probes, and RT-PCR reagents available against VEGF-A do not differentiate the active from the "b" isoforms, it raises some interesting issues with regard to the apparent levels of VEGF-A in normal and pathological specimens.

The VEGFR are glycosylated, type-I integral membrane proteins with seven immunoglobulin-like domains forming the extracellular portion and a split tyrosine kinase domain forming the interior portion. The VEGFR are closely related to the PDGF receptors (see below). VEGFR-1 (also known as *flt*-1) has the highest affinity for VEGF, but as discussed above, in endothelial cells at least, appears to function as a decoy receptor. VEGFR-2 (also known as *kdr*) is selectively expressed on endothelial and hematopoietic precursor cells, and mediates all of the known activities of VEGF-A on vascular endothelium.

The neuropilins (NRP) were first described as neuronal receptors for axon guidance factors for members of the semaphorin family. It was later found, however, that the neuropilins were expressed on endothelial cells and were also able to function as receptors for some forms of VEGF-A ($VEGF_{165}$ binds both NRP1 and NRP2; $VEGF_{145}$ binds NRP2 but not NRP1; $VEGF_{121}$ does not bind either). Other VEGF family members such as PlGF2 and VEGF-B bind to NRP1 and PlGF-2 and VEGF-C bind to NRP2. The neuropilins have short intracellular domains that apparently do not transduce signals subsequent to either semaphorin or VEGF binding. It is thought that the neuropilins form complexes with the VEGFR, facilitating signal transduction.

Potential beneficial effects of VEGF-A in therapeutic angiogenesis were first suggested by experiments in which supplemental VEGF, administered either as a recombinant protein or as naked DNA, augmented collateral blood flow in a number of species with experimentally induced hind limb or myocardial ischemia. A recent study by Rissanen and coworkers compared the activities of different VEGF family members in a rabbit hind-limb model. In these studies, the effects of administering adenoviral vectors expressing VEGF-A, VEGF-B, VEGF-C, and VEGF-D, as well as a VEGFR-3 specific mutant ($VEGF\text{-}C^{156S}$) were compared. The authors also explored the effects of adenoviruses expressing either full-length or proteolytically processed VEGF-D ($VEGF\text{-}D^{\Delta N \Delta C}$). They found that VEGF-A and $VEGF\text{-}D^{\Delta N \Delta C}$ induced the strongest angiogenic and vascular permeability effects. Full-length VEGF-C and VEGF-D induced predominantly lymphangiogenesis and the VEGFR-3 specific ligand $VEGF\text{-}C^{156S}$ exclusively promoted lymphangiogenesis. Interestingly, VEGF-B did not promote either angiogenesis or lymphangiogenesis. The lack of effect of VEGF-B suggested that VEGFR-1 signaling alone might not be capable of initiating angiogenesis. However, collateral formation/arteriogenesis is impaired in PlGF−/− mice and this phenotype is rescued when the PlGF−/−

mice are treated with a local administration of PlGF. PlGF and VEGFR-1 are upregulated in postischemic collateral vessels. Moreover, local infusion of PlGF in the rabbit hind-limb model of arteriogenesis potently induces collateral growth, an effect that is abolished by monocyte depletion. It is currently thought that the role of PlGF/VEGFR-1 may be to recruit monocytes (which express VEGFR-1). The apparent discrepancies between the effects of PlGF and VEGF-B remain confusing, although they might result from differences in the methodology as well as species.

2.2.2 FGF Family

The FGFs comprise a large family of 23 related peptide growth factors (FGF-1 to FGF-23), although four of them (FGF-11–14) are probably not canonical FGFs. FGFs modulate the proliferation and differentiation of many different cell types of both mesenchymal and neuroectoderm origins. Increased FGF signaling, due to mutations in either the FGFs or their receptors, leads to a variety of human skeletal disorders, including dwarfism and craniosynostosis syndromes. Most of the FGFs (FGF-3–8, -10, -15, -17–19 and -21–23) have N-terminal signal peptides and are readily secreted. In spite of the lack of an obvious signal sequence, FGFs-9, -16, and -20 are also secreted. FGF-1 and FGF-2 also lack a signal sequence and are apparently not secreted, but can be found on the cell surface and deposited within the extracellular matrix. FGF-9 has a noncleaved amino-terminal hydrophobic sequence that is required for secretion; a third set of the FGFs (FGF-11-14) lack signal sequences and are believed to remain inside the cell. In the adult, FGFs are thought to be involved in the processes of both physiological (e.g. in response to exercise, wound healing) and pathological (e.g. tumors, retinopathies) angiogenesis. With respect to therapeutic angiogenesis, the role/effects of two of the FGF family members (FGF-2 and FGF-4) have been evaluated.

FGF-1 and FGF-2 were the first FGF family members to be isolated and were originally named acidic and basic FGF respectively on the basis of their isoelectric points. However, FGF-1 and FGF-2 differ from most other FGFs in several important ways. FGF-1 is unique among FGFs in that it binds with high affinity to all of the known receptor isoforms. Both FGF-1 and FGF-2 are found in the cytosol and lack a signal peptide at their 5′ ends. Despite the lack of a signal peptide, both factors seem to be released from cells through a nonclassical secretory pathway. Both FGF-1 and FGF-2 have also been found in the cell nucleus and a putative nuclear localization signal has been identified at the 5′ end of the FGF-1 protein. There are also higher molecular weight forms of FGF-2, which derive from alternative translation initiation sites in the 5′ region of the *FGF2* gene. These isoforms localize to the nucleus. The precise function(s) of these nuclear forms of FGF-1 and FGF-2 remains unclear.

FGF-2 is a potent regulator of many cellular functions and phenomena, including cell proliferation, differentiation, survival, adhesion, migration, motility and apoptosis, and processes such as limb formation, wound healing, tumorigenesis, angiogenesis, vasculogenesis, and blood vessel remodeling. FGF-2 is a potent mitogen and chemotactic factor for endothelial cells and also stimulates smooth muscle cells and pericytes to proliferate. Other processes relevant to angiogenesis that are regulated by FGF include stimulation of plasminogen activator production and matrix metalloproteinase (MMPs) expression.

The latter two responses play an important role in vessel destabilization and breakdown of the extracellular matrix, a process required for endothelial cell sprouting and invasion. However, despite these activities, the deletion of the FGF-2 gene resulted in a phenotype with only mild cardiovascular, skeletal, and neuronal abnormalities.

FGF-4 has the distinction of being one of the minorities of the FGF family members whose deletion results in embryonic lethality (FGF-8, -9, -10, -15, and -18 are also embryonic lethal or lethal at P0). FGF-4 is absolutely required for preimplantation mouse development apparently due to a requirement for FGF-4 to maintain trophoectoderm and primitive endoderm identity at embryonic day 4.5. With respect to angiogenesis, it is not really clear whether FGF-4 plays a role in endogenous angiogenic responses. *In vitro* studies suggest that it acts in a manner similar to FGF-2, although somewhat weaker in terms of endothelial cell proliferation. Nonetheless, adenoviral gene delivery studies have now shown that delivery of FGF-4 by this means promotes neovascularization in animal models and has promising effects in the clinic.

A family of four receptor tyrosine kinases mediates the diverse effects of the FGFs. These receptors are characterized by an extracellular ligand-binding region consisting of three immunoglobulin-like (Ig) domains (D1 to D3), a single transmembrane helix, and a cytoplasmic portion with the protein tyrosine kinase activity. Alternative mRNA splicing leads to additional isoforms of these receptors and thereby alters ligand binding. One such splicing event results in the skipping of exons encoding D1, resulting in a shorter, two Ig-like domain form of the receptor. Binding specificity for the different FGF family members resides in the D2 and D3 Ig domains as well as a short linker between D2 and D3. It has been suggested that the short form of the receptor may have a higher affinity for some FGFs than the long form. Another alternative splicing site specifies the sequence of the carboxy-terminal half of D3, resulting in either the 3b (IIIb) or 3c (IIIc) isoform of the FGF-R. These splicing events are thought to be regulated in a tissue-specific manner, with exon IIIB more prevalent in cells of epithelial lineage, and IIIc in cells of mesenchymal origin.

As is the case for many of the receptor tyrosine kinases, receptor dimerization is required for FGF signaling. FGF signaling also requires heparin or HSPG. Interaction with HSPG or heparin stabilizes FGFs to thermal denaturation and proteolysis and, moreover, can restrict the diffusion distance from the site of FGF production.

There are now many published studies with different animal models that support the concept that administration of FGF-2 as protein, naked DNA, or as an adenoviral gene-therapy vector can promote angiogenesis, improve myocardial perfusion, improve endothelial vasodilatory function, and promote the development of collateral vessels. Gene therapy using adenovirus serotype 5, in which the E1A and E1B genes have been replaced by the human FGF-4 gene driven by a constitutive cytomegalovirus promoter (Ad5FGF-4), was effective in stimulating collateral vessels in a porcine model of stress-induced myocardial ischemia and in a rabbit hind-limb ischemia model.

2.2.3 PDGF Family

The platelet-derived growth factor family is comprised of four polypeptides, PDGFA, B, C, and D, which occur in the form of homodimers (PDGF-AA, PDGF-BB, PDGF-CC, PDGF-DD) or heterodimers

(PDGF-AB). To date, there are two known PDGF receptors, an α-receptor (PDGFR-α) and a β-receptor (PDGFR-β). The PDGFR-α receptor will bind all forms of PDGFs except PDGF-DD, while the PDGFR-β receptor binds all forms of PDGFs except PDGF-AA. When a dimerized PDGF molecule binds one of the above receptor molecules, receptor dimerization occurs, followed by activation of the receptor tyrosine kinase activity, leading to the autophosphorylation of the cytoplasmic domain of the receptor.

PDGF was originally purified from platelets (as implied by its name), but it was subsequently found to be made by many other cell types including fibroblasts, keratinocytes, myoblasts, epithelial cells, and macrophages. Endothelial cells derived from the microvasculature express PDGFR-β and the addition of PDGF-BB to microvascular endothelial cells stimulates DNA synthesis and tube formation in *in vitro* assays. Importantly, PDGF also stimulates the proliferation of smooth muscle cells and pericytes. The critical role for PDGF in the maturation of new blood vessels was demonstrated in genetic mouse models. Mice deficient in either PDGF-B or PDGFR-β develop grossly normal blood vessels, but die *in utero* from hemorrhage and edema. The cerebral microvasculature of these mice was the major site of hemorrhage and leakiness; in these vessels pericytes were absent and the vessel lumens were dilated. In other organs, pericytes may have been present (it is difficult to distinguish pericytes from smooth muscles because of a paucity of specific markers) but dysfunctional. It is thought that PDGF therefore plays a role in the recruitment and the maturation of pericytes and thus plays an important role in the stabilization of newly formed blood vessels.

Cao and coworkers studied the angiogenic synergy, vascular stability, and improvement of hind-limb ischemia using combinations of PDGF-BB, VEGF-A, and FGF-2. In this study, the authors employed *in vivo* angiogenesis assays (mouse corneal micropocket assay and mouse matrigel assay). PDGF-AB and PDGF-BB induced robust angiogenesis in the cornea, whereas the vessel areas induced by PDGF-AA were significantly smaller. To test the possibility that PDGF-BB might modulate FGF-2 or VEGF-A-induced angiogenesis, these growth factors were implanted alone or in combination into mouse corneas. FGF-2 and VEGF-A alone induced intense corneal neovascularization and marked synergy was observed in corneas coimplanted with PDGF-BB and FGF-2. The vessel areas were significantly greater and the newly formed blood vessels appeared better organized than those elicited by FGF-2 alone. In contrast, PDGF-BB did not improve the quality of the VEGF-induced vascular networks. Similar results were observed in the matrigel model. To evaluate vessel stability, these authors monitored the mouse corneal networks for more than 210 days. The newly formed vessels in corneas treated with PDGF-BB alone regressed after 24 days. FGF-2-induced vessels remained stable for up to 40 days, but had almost all regressed by 70 days. In contrast, corneas implanted with the combination of PDGF-BB and FGF-2 developed vascular networks that remained stable for over 210 days. PDGF-BB did not improve the longevity of VEGF-A induced vessels, which were barely detectable after ~70 days. The combination of VEGF-A and FGF-2 produced an additive effect on vascular network development and the vessels developed also remained for a somewhat longer period (~100 days). To evaluate the possible role of mural cells in

the longevity of vessels induced by PDGF-BB and FGF-2, the authors performed a histological analysis of the vessels, staining for endothelial markers (CD31) and pericyte/smooth muscle markers (desmin). The vessels induced by the combination of PDGF-BB and FGF-2 had a significantly greater smooth muscle/pericyte coating than the FGF-2 alone group. These authors also examined the synergistic interactions of FGF-2 and PDGF-BB in the rat hind-limb ischemia model. Angiographic analysis was performed 23 days after ligation of the femoral artery. In agreement with earlier studies, FGF-2 alone increased the number of collateral vessels. Delivery of PDGF-BB alone was only moderately effective in improving collateral vessel formation. The combination of FGF-2 and PDGF-BB potently stimulated collateral growth with a higher density of collateral vessels. The high numbers of collaterals induced by the combination of PDGF-BB and FGF-2 were maintained for more than 60 days without regression. The effects on collaterals also correlated well with hind-limb perfusion scores. These observations clearly suggest that the combination of FGF-2 and PDGF-BB can interact synergistically to induce angiogenesis, arteriogenesis, and long-lasting functional vessels. Recent studies by Hao and coworkers also suggest that the combination of FGF-2 and PDGF-BB (administered by dual gene transfer) also resulted in an increased number of capillaries and arterioles and more stable vessels in a rat model of myocardial infarction.

2.2.4 TGF-β Family

Transforming growth factor-βs are actually a superfamily of more than 35 members, including the TGF-βs, BMPs (bone morphogenic proteins), GDFs (growth differentiation factors), activins, inhibins, MIS (Müllerian inhibiting substance), nodal, and leftys. The best-known member is TGF-β1. TGF-β1 is secreted as a latent form where the activity of the mature domain is masked by the propeptide LAP (latency associated peptide). This biologically inactive precursor must be activated by proteases such as plasmin or cathepsin D. It can also be activated by low pH, chaotropic agents such as urea, and by heating. Many different cells, both normal and transformed, express TGF-β1 and like FGF2, TGF-β1 binds to and is found in the extracellular matrix of many tissues. Endothelial cells, pericytes, and smooth muscle cells all express TGF-β receptors, and thus TGF-β can exert many effects on the vasculature. TGF-β significantly inhibits the proliferation and migration of endothelial cells (although low doses may stimulate growth), and also may stimulate (low dose) or inhibit (higher doses) endothelial tube formation. The effects of TGF-β on tube formation may be mediated by the ability of TGF-β to modulate urokinase and plasminogen activator inhibitor levels and to stimulate the production of protease inhibitors such as tissue inhibitor of metalloproteinase. TGF-β is also a potent stimulant of extracellular matrix production.

The ligands of the TGF-β superfamily signal through a family of transmembrane serine/threonine receptor kinases. There are two subfamilies of TGF-β receptors known as the Type I and Type II receptors. Upon ligand binding, these receptors interact. The type I receptors have a somewhat confusing nomenclature due to their simultaneous discovery by different groups. However, a common name applied to many of them is ALK receptors (activin receptor-like kinases). TGF-β1 binds directly to the type II receptors, then

recruits type I receptors, and once an active ligand-type I/type II receptor complex is formed, the type II receptors activate type I receptors by phosphorylation. The type I receptors subsequently phosphorylate downstream SMAD proteins that propagate the signal from the cytoplasm to the nucleus.

A third distinct type of cellular receptors for the TGF-β ligands are the type III receptors, exemplified by betaglycan and endoglin. Endoglin is expressed at high levels in endothelial cells, and at lower levels in monocytes, erythroid precursors, and other cell types.

TGF-β1 can stimulate angiogenesis *in vivo*, an activity that may result from the proinflammatory effects of the protein TGF-β1, which is chemotactic for monocytes and fibroblasts.

Genetic studies have strongly suggested important roles for TGF-β1 in the development of the vasculature. Mice lacking TGF-β1 exhibit two phenotypes. About 50% of the mice die *in utero* as a result of defects in the yolk sac vasculature and hematopoietic system. The other embryos survive (possibly through maternal transfer of TGF-β1), but die at 3 to 4 weeks of age as a result of severe multifocal inflammation. Mice deficient in TGF-β receptor 1 (also known as ALK-5 or TβRI) also die midgestation as a result of severe defects in the vascular development of the yolk sac and placenta and an absence of circulating red cells. In the vascular system, there is an additional type I receptor for TGF-β1, ALK-1. ALK-1 deficient mice die around midgestation, also exhibiting severe vascular abnormalities. Targeted disruption of the murine endoglin (Eng) gene is also embryonically lethal at E11.5 due to defects in vascular development. Eng-null mutants exhibit both poor vascular smooth muscle cell development and arrested endothelial remodeling. Endoglin thus appears to be essential for the differential growth and sprouting of endothelial tubes and the recruitment and differentiation of mesenchymal cells into vascular smooth muscle cells and pericytes.

Van Royen and coworkers evaluated the activity of exogenous application of TGF-β1 in a model of arteriogenesis in the rabbit hind limb. TGF-β1 was delivered locally (0.48 $\mu g\ kg^{-1}\ d^{-1}$) into the collateral circulation via an osmotic minipump. Compared to Phosphate Buffered Saline (PBS)-treated controls, TGF-β1-treated animals had an increased number of visible collateral arteries as well as the conductance of the collateral circulation. The authors attributed at least part of the arteriogenic response to monocyte activation by TGF-β1. Isolated human monocytes treated with TGF-β1 exhibit an upregulation of the MAC-1 receptor as well as increased monocyte adhesion and transmigration across endothelial monolayers.

2.2.5 HGF

Hepatocyte growth factor (HGF; also known as scatter factor) is a mesenchyme-derived factor with multiple activities including the regulation of cell growth, motility, and morphogenesis of various cell types including epithelial and endothelial cells. HGF is a large, multidomain protein that is structurally similar to plasminogen. HGF has six domains – an amino-terminal domain, four Kringle domains, and a serine proteinase homology domain. The serine proteinase homology domain lacks enzymatic activity. HGF is synthesized as a large, single-chain inactive precursor (pro-HGF), which is converted proteolytically into a two-chain, active heterodimer. HGF also binds to heparan sulfate proteoglycans with high affinity, although HSPG binding is not essential for receptor activation. The

receptor for HGF was identified as c-*met*, a proto-oncogene.

Met is a disulfide-linked heterodimer, consisting of an extracellular α-chain and a longer β-chain. The β-chain consists of an ectodomain, the transmembrane helix, and the cytoplasmic portion, which contains kinase domains. Met has a unique bidentate docking site, which comprises two tyrosine residues (Y1349 and Y1356) and the surrounding amino acid residues. When phosphorylated, this docking site binds substrates such as Gab1, Grb2, and phosphatidylinositol 3-kinase. Although deletion of HGF or met (both of which were embryonically lethal) demonstrated an essential role of this signaling system in development, the phenotype was not overtly vascular. The livers of both the HGF−/− and met−/− mice were considerably reduced in size because of the decreased proliferation and increased apoptosis of the hepatocytes. The mice also demonstrated a placental defect and a reduced survival of sensory neurons and a complete absence of all muscle groups, which derive from migrating precursor cells.

In the adult, HGF and met are broadly expressed and have been implicated in responses to injury, liver regeneration, tumor growth and metastasis, and angiogenesis. HGF is a potent endothelial mitogen, motogen, and morphogen. Moreover, several *in vivo* studies have shown that this growth factor is a potent inducer of angiogenesis.

The role of HGF in arteriogenesis has not been extensively studied. However, HGF is induced in the skeletal muscle of both rats and mice following ischemic injury. HGF has also been implicated in the regeneration of capillary endothelial cells in ischemically injured myocardium. Van Belle and coworkers examined the effects of VEGF-A and HGF, alone and in combination, on endothelial proliferation and migration *in vitro*. They also compared the effects of VEGF-A and HGF on capillary density and collateral development in the rabbit hind-limb ischemia model. These authors found that the combination of VEGF-A and HGF synergistically induced endothelial proliferation and migration *in vitro*. In the rabbit hind-limb ischemia model, HGF appeared to produce a superior response in terms of capillary density and collateral formation compared to either VEGF or vehicle. The combination of HGF and VEGF also acts synergistically in *in vitro* assays of endothelial tube formation, and a recent gene expression profiling study demonstrated that these two growth factors elicit different profiles of gene expression and when combined, exhibit considerable synergy in both the magnitude and numbers of new genes upregulated. Several additional studies have provided substantial evidence that HGF, administered as a recombinant protein or as naked DNA or as an adenoviral vector, is a potent inducer of arteriogenesis. However, combination studies of the HGF with VEGF, TGF-β1, or PDGF-BB in arteriogenesis models have not been reported to date.

2.2.6 Angiopoietins

The angiopoietins (Ang) are a family of secreted proteins, four of which have been identified to date, that bind to Tie family receptor tyrosine kinases. Two Tie receptors have been identified thus far, Tie1 and Tie2 (Tek); interestingly the distribution of these receptors appears to be selective for the vascular endothelium and some cells of the hematopoietic cell lineage. An important role for these molecules was first revealed by gene knock-out studies. Ang1

and Tie2 deficient mice have similar embryonic lethal phenotypes, characterized by severe defects in vascular remodeling, reduced numbers of endothelial cells, insufficient vessel stabilization, and deferred vessel maturation. Application of the Ang1 protein or expression in transgenic mice can also counteract some of the side effects of VEGF such as increased vessel leakiness, while having an additive effect on the angiogenic effects of VEGF. Tie1-deficient mice develop extensive edema and hemorrhage and die either perinatally or at embryonic day 14.5. A ligand for Tie1 has not been identified. Ang-2 was first characterized as a related protein to Ang1 that could bind Tie2 and antagonize Ang1. Overexpression of Ang2 during development leads to phenotypes very similar to those observed in Ang1-deficient animals. It is thought that Ang2 may induce vessel destabilization and allow angiogenic sprouts to form and respond to other growth factors. Ang1 mRNA is predominantly expressed in cells surrounding the endothelium such as the smooth muscle cells, whereas in the adult, Ang2 is selectively expressed in endothelial cells at sites of vascular remodeling (e.g. ovary, uterus, placenta). Ang2 expression also appears to be regulated by hypoxia.

Siddiqui and coworkers examined the effects of Ang1 and VEGF-A_{165} gene therapy (alone and in combination) in a mouse model of myocardial ischemia. In the ischemic myocardium, administration of either plasmid resulted in an increase in the ratio of arterioles to capillaries; dual administration seemed to further increase the number of arterioles. These observations suggest that evaluation of the combinations of Ang1 with VEGF or other growth factors might be a fruitful area for further investigation.

2.3 Inflammation

Inflammation promotes angiogenesis in several different ways. Sites of inflammation are often also hypoxic, and as discussed above, hypoxia upregulates VEGF-A and Ang1. The plasma proteins that extravasate at sites of inflammation contain fibrinogen products and other protein fragments that can stimulate new vessel formation. Importantly, many of the cells recruited to the site of inflammation (macrophages, lymphocytes, mast cells, platelets, neutrophils) release angiogenic factors and enzymes that can further amplify an angiogenic response. These angiogenic factors include growth factors, cytokines, and proteinases like metalloproteinases.

There is increasing evidence that recruitment of monocytes may be a functionally important aspect of arteriogenesis. Monocytes have been identified on the endothelial lining of growing canine coronary collaterals; these cells are probably attracted by the endothelial expression of MCP-1, a monocyte chemoattractant. Local infusion of MCP-1 has been shown to improve collateral vessel formation in several studies. Fluid shear stress is a potent inducer of MCP-1 as well as of ICAM-1, an adhesion molecule to which the Mac-1 receptor of monocytes binds. Targeted disruption of the MCP1 receptor (CCR-2) in mice decreases collateral growth after femoral artery occlusion. Additionally, the weak arteriogenic effects of VEGF infusion may be due to the monocyte chemoattractant activity of VEGF for monocytes, which express VEGFR-1. This may also account for the reduced collateral artery growth observed in PlGF−/− mice (PlGF is selective for VEGFR-1).

An overlooked cell type in arteriogenesis may be the mast cell. Mast cells aggregate in the adventitia of growing collaterals and remain there for up to six weeks. Since these cells can secrete bFGF, VEGF, other vasoactive autocoids, platelet activating factors, and a variety of other cytokines, they may well play a paracrine role in the development or remodeling of collateral vessels.

2.4 Circulating Endothelial Precursors/Bone Marrow-derived Cells

Bone marrow provides a rich reservoir of stem and progenitor cells. Recently, a rare population of cells called "endothelial progenitor cells" (EPCs) has been described. The "EPCs" can be mobilized to the circulation, and are thought to contribute to the formation of new vessels. There is also a population of circulating endothelial precursors (CEPs). CEPs are thought to originate from EPCs in the bone marrow, and may also be recruited from the parenchyma of the systemic vasculature. These cells are defined by a number of phenotypic markers (see Table 2).

Angiogenic factors and chemokines are thought to recruit subsets of proangiogenic hematopoietic cells, which along with EPCs and CEPs may contribute to new blood vessel formation. There are numerous studies which now suggest that bone marrow-derived cells are recruited to areas of neoangiogenesis following wound healing, limb ischemia, post myocardial infarction, endothelialization of vascular grafts, retinal and lymphoid organ neovascularization, and angiogenesis associated with neonatal development and tumor growth. Vigorous exercise has also been shown to increase the number of CEPs in peripheral blood.

Several studies have shown that transplantation of a bone marrow-derived stem or EPCs seems to improve recovery of perfusion and function in animal models of myocardial and peripheral ischemia. The relative contribution of these transplanted cells to the endothelium of growing vessels varies considerably from one study to another, ranging from 0% to ≥50%, although

Tab. 2 Markers for endothelial precursor cells (EPCs), circulating endothelial precursors (CEPs), and circulating endothelial cells (CECs).

Cell type	*Origin*	*Specific phenotypic markers*
EPCs	Bone marrow, vascular parenchyma, umbilical cord blood	CD133(AC133)+, VE-cadherin, VEGFR-2+, CD34+, CD31+, VEGFR-1+, CD146+, CXCR4+, Di-Ac-LDL uptake, vWF+, E-selectin, Ulex europeas binding
CEPs	Peripheral blood	CD133+, VEGFR2+, VE-cadherin, CD34+, CD31+, VEGFR-1+, CD146+, CXCR4+, Di-Ac-LDL uptake, vWF+, E-selectin, Ulex europeas binding
CECs	Endothelial lining of vessel wall	CD133NEG, VEGFR-2+, CD146+, VE-cadherin+, CD34+, CD31+, Di-Ac-LDL uptake, vWF+, E-selectin, Ulex europeas binding

in most studies this figure denotes that 50% or more of the capillaries examined have at least one single transplanted cell. It is important to emphasize the results of recent studies employing laser confocal image analysis, which suggest that the so-called EPCs do not really incorporate into the vessel wall *per se*, but are closely associated with the vessel wall. The bone marrow-derived stem cells may thus play an important paracrine role, releasing cytokines or other growth factors.

2.5 Mechanical Factors

At present, relatively little is known about the endogenous signals for collateral artery growth. However, many investigators believe that the mechanical effects of increased blood flow play an important role. Prior to the occlusion of an artery or the formation of a hemodynamically significant stenosis, there is likely to be little flow across preexisting collateral arterial connections. However, with the narrowing or occlusion of a major artery, the distal arterial pressure falls, increasing the pressure gradient along preexisting collaterals, resulting in the establishment of unidirectional blood flow and an increase in flow in the midsection of the collateral vessels. In contrast to the distal site (which may be hypoxic), the region of the collateral artery is exposed to fully oxygenated blood. Physiologists have studied the effects of blood flow on arterial diameters using models of arteriovenous fistulas. Normally, blood flows from arteries into capillaries and then into the veins. When an arteriovenous fistula is present, blood flows directly from an artery into a vein, bypassing the capillaries. The increased flow in the artery results in arterial expansion. Wall shear stress (τ) can be calculated using the equation:

$$\tau = \frac{4\eta Q}{\pi a^3} \qquad (1)$$

where η = viscosity, Q = flow rate, and a = internal radius of the vessel. Therefore, an increase in flow will directly translate into an increase in shear stress and the expansion of the internal diameter (either mechanically, by relaxation of the vessel, or by active remodeling of the vessel) subsequently antagonizes this effect. There are other stresses on the vessel wall that may also be triggers for arterial remodeling, including the intraluminal pressure, which in turn creates circumferential, longitudinal, and radial stresses on the wall.

The vascular endothelium plays a critical role in the responses of the vessel wall to biomechanical stresses. Changes in wall shear stress can induce the expression of adhesion molecules such as ICAM-1 and VCAM-1, as well as the expression of chemokines (MCP-1) and matrix metalloproteinases. Mechanical strain may also induce MMP expression by vascular smooth muscle cells and monocyte/macrophages.

2.6 Hypoxia

It is important to emphasize that with regard to oxygen tension, arteriogenesis and angiogenesis differ fundamentally. Angiogenesis occurs in regions of hypoxia, which are usually quite distant from the site of collaterals that form to bridge a major arterial occlusion. Arteriogenesis, on the other hand, may occur in a normoxic environment. In the heart, this distinction may not be as clear; the distance from ischemic tissue and collateral vessel growth is small. As discussed above, hypoxia upregulates the expression of the

transcription factor, HIF-1α, which in turn regulates the expression of VEGF, Ang1, and VEGFR-1.

2.7 Vascular Remodeling

Following the acute phase of arteriogenesis that is dominated by inflammatory events, a second phase ensues, which is the slow remodeling of the arterial structure after the final diameters are attained. This requires the formation of a new elastic lamina by smooth muscle cells, the rebuilding of the media, and formation of an intima. Associated with these events is the down-regulation of tissue inhibitor of matrix metalloproteinases (TIMPs and MMPs) and upregulation and expression of new MMPs that remodel the matrix. Some of the smooth muscle cells of the old media appear to undergo apoptosis, and are replaced by new smooth muscle cells with a more synthetic embryonic-like phenotype. With the thickening of the vessel wall, the intercellular connections and communications between the cells change, an event paralleled by an increase in the expression of connexin 37 (an early marker of arteriogenesis) in the smooth muscle cells. The final stages of collateral remodeling are often accompanied by an increase in vessel length and thereby vessel-tortuosity.

3 Animal Models of Therapeutic Angiogenesis

A number of different species have been used to study arteriogenesis, the most common being the dog, pig, rabbit, rat, and mouse. These species vary significantly in their vascular anatomy, and these differences should be considered in evaluating the data. Hind-limb ischemia models usually involve ligation or resection of a femoral artery. It should be noted that in some species, for example, rats, this surgery may cause little or no ischemia (probably because of the presence of collaterals which provide sufficient blood flow). In rabbits, after femoral artery ligation, no increase in the expression of VEGF or hypoxia-inducible genes (HIF-1α, lactate dehydrogenase A) was found. For the heart, a common model is to ligate a major coronary artery or to place an ameroid constrictor around a major artery (which slowly constricts, allowing development of collaterals without necessarily causing a myocardial infarction). In recent years, mice have become very attractive models owing to the availability of different strains and genetically modified (e.g. transgenic and knock-out) animals.

Investigators use a variety of techniques to evaluate vascular structures in these models, and each method has its advantages and disadvantages. Histology, a commonly used method to analyze capillary and arteriole density and dimensions, is often a relatively subjective two-dimensional approach to a three-dimensional problem. Moreover, the sections examined may or may not provide an accurate representation of vascularity through the entire sample. Laser Doppler perfusion imaging is widely used to assess hind-limb perfusion in mice and rats; however, this method provides no anatomic information and only measures superficial perfusion. Other methods include the injection of fluorescent or radiolabeled microspheres, which offers the ability to study, in a longitudinal manner, perfusion to tissues. However, in a small animal such as a mouse, the injection of microspheres into the left ventricle can be

challenging because of their small size. Estimates of vascular surface area can be obtained using a radiolabeled endothelial specific antibody technique. Other recent imaging approaches include magnetic resonance angiography, positron emission tomography (PET), and micro-computed tomography (micro-CT). Recent studies with micro-CT suggest that this method may provide the required resolution and sensitivity for accurate assessment of mouse models.

Evidence for differences in blood flow recovery (as measured by Laser Doppler imaging) and hemoglobin oxygen saturation (as assessed by noninvasive tissue oximetry) in the distal hind limbs after femoral artery occlusion in different inbred strains of mice has been reported, with C57BL/6 mice recovering faster and more completely than BALB/c mice. At least, part of this difference may be due to the presence of more or larger preexistent collaterals in C57BL/6 than in BALB/c mice. Murine, rabbit, rat, and porcine models have been used to evaluate the potential roles of growth factors, gene therapy, stem cells, and other treatment modalities for therapeutic "angiogenesis." An important caveat for many of these studies is that they are generally performed in young, healthy animals. The clinical situation is much more complicated with the additional issues of age, obesity, atherosclerosis, and diabetes. Several studies have now shown that arteriogenesis/angiogenesis is often impaired in aged, atherosclerotic, hyperlipidemic, or diabetic animals and humans. However, relatively few studies have evaluated the efficacy of different protein, plasmid, or gene-based therapies in these models, which may more closely reflect the clinical situation.

4 Clinical Trials

As discussed above, there are now numerous preclinical studies in mice, rats, rabbits, dogs, and pigs that have shown the potential of various angiogenic growth factors to restore blood flow and function to ischemic tissues. Quite reasonably, therefore, a number of these approaches have been evaluated clinically. To date, however, the translation of this experience into clinical success has been disappointing and it is clear that this approach is "not ready for prescription" just yet. Early, open-label uncontrolled studies seemed to predict spectacular success, but the later, well-controlled studies were less encouraging. There have been several recent reviews of the clinical data from these studies.

4.1 Protein-based Trials

One of the first trials to test a recombinant protein employed a combination of intracoronary and intravenous infusions of VEGF-A_{165} (the VEGF in Ischemia for Vascular Angiogenesis (VIVA) trial) (see Table 3). The clinical endpoints were exercise tolerance, myocardial perfusion, angina symptoms, or functional class. Although there was no significant benefit at 60 days, there was a trend towards a benefit in the high-dose VEGF165 group at 120 days.

The FIRST trial (FGF Initiating Revascularization Trial) evaluated a single intracoronary FGF2 infusion. In this study, there were significant improvements in functional status and angina symptoms in the FGF2 group; however, there was no benefit compared with placebo in terms of exercise capacity or myocardial perfusion at 90 days. A more promising result was

Tab. 3 Clinical trials in angiogenesis. Protein-based therapies.

Trial	*Name*	*Growth factor and route of administration*	*Patient population [N]*	*Clinical endpoints*	*Results*
VIVA	VEGF in ischemia for vascular angiogenesis (Phase II)	Intracoronary and intravenous VEGF165 protein infusions	178	At 60 days Exercise tolerance Myocardial perfusion Angina symptoms Functional class	No significant benefit Trend in favor of high dose VEGF165 seen at 120 days
FIRST	FGF initiating revascularization Trial(Phase II)	Single intracoronary FGF2 infusion	337	At 90 days Functional status Symptom class Exercise tolerance Myocardial perfusion	Significant improvement in functional status and symptom class No benefit compared with placebo for exercise or myocardial perfusion
Hep-Alg	Phase I/II	Polymer released FGF2	Given at time of coronary bypass surgery24	At 90 days Myocardial perfusion Symptom status	Significant improvement Improvement still seen at three-year follow-up
TRAFFIC	Phase II	Double intra-arterial FGF2 infusion	Patients with peripheral vascular disease190	At 90 days Walking time	Improvement seen at 90 days. However, no differences among treated groups after 6 months

obtained in a very small ($n = 24$), but randomized trial of a slow release formulation of FGF2 implanted (at the time of coronary artery bypass surgery) into unrevascularized myocardium. In these patients, there was significant improvement in perfusion and symptom status after 90 days in the high FGF2 group that was still seen at the 3-year follow-up.

The (Therapeutic angiogenesis with recombinant fibroblast growth factor-2 for intermittent claudication) TRAFFIC trial was the first randomized trial in patients with peripheral vascular disease. Patients in the treatment arm of this trial received two (30 days apart) intra-arterial infusions of FGF2. Interestingly, at 90 days, the FGF2 patients had improved peak walking times, but this difference was not maintained after six months.

4.2 Gene Therapy-based Clinical Trials

Gene therapy has several potential advantages over protein-based therapy. Probably the most important is that gene transfer can allow prolonged, local exposure. This advantage may be offset if there is low efficiency of gene expression and also if there is an immune response to the protein or virally infected cells. Uncontrolled regulation of gene expression may also pose risks; for example, several animal model studies suggest that overexpression of VEGF can lead to the formation of hemangioma-like structures and severe edema.

Studies exploring the efficacy and safety of gene therapy for therapeutic arteriogenesis are appearing with greater frequency in the literature. Some of these approaches have entered clinical trials, and this information is summarized in Table 4. While these early studies look promising, the majority of studies have been open-label without placebo controls, and past experience with the protein-based trials has clearly shown that there is a significant placebo effect.

The angiogenic gene-therapy (AGENT) studies were the first randomized, double-blind, placebo-controlled trials of therapeutic angiogenesis by gene transfer for myocardial ischemia. In the AGENT1 trial, Ad5-FGF4 enrolled 79 patients between 1998 and 2000. The principal objectives of the trial were safety evaluation and dose identification for larger clinical trails. Efficacy was assessed by determining the change in total exercise time and the time of onset of angina from baseline to follow-up (at 4 and 12 weeks) using a symptom-limited exercise treadmill test. Ad5-FGF4 was administered via a cardiac catheter. Overall, the FGF-4 therapy using the adenoviral vector was well tolerated without significant safety concerns. Importantly, there was no detectable FGF4 protein outside of the heart, and the calculated first pass extraction across the coronary circulation was 87% of the injected dose. The exercise treadmill testing suggested a trend towards improvement in the FGF-4 treated group, although it was not statistically significant.

The AGENT-2 trial was a randomized, double-blind placebo-controlled trial. This study showed that Ad5-FGF4 was more effective than the placebo in reducing the reversible perfusion defect size in patients with stable angina. The Ad5-FGF4 group also had a higher percentage of symptom-free patients at 8 weeks. However, the pivotal AGENT-3/-4 trials were stopped early because an interim analysis showed that the efficacy endpoints were unlikely to be met. The reasons

Tab. 4 Clinical trials in angiogenesis: gene-based therapies.

Title	*Phase*	*Reagent*	*Method of delivery and [n]*
VEGF2	Phase I/II	VEGF2 plasmid	Intramyocardial injection(19)
KAT	Phase II	VEGF165 protein, plasmid or adenovirus	Injection with PTCA or stenting(103)
Euroinject one	Phase II	Plasmid-VEGFA165	Percutaneous intracardiac injection(74)
REVASC	Phase II	Adenovirus VEGFA121(CAD)	Intracardiac injection via thoracotomy67
VEGF-POAD	Phase II	Adenovirus-VEGFA165(PAD)	54
RAVE	Phase II	Adenovirus-VEGFA121	105
AGENT-1	Phase I	Adenovirus FGF4(CAD)	79 (19 placebo, 60 receiving AdFGF4)
AGENT-2	Phase II	Adenovirus FGF4(CAD)	52 A = 35, P = 17
AGENT-3	Phase III (US)	Adenovirus FGF4(CAD)	Single intracoronary administration 450 A = 300 P = 150
AGENT-4	Phase III (Multinational)	AdFGF4(CAD)	Single intracoronary administration 450 A = 300 P = 150
AMG0001	Phase II	HGF plasmid(PAD)	
Del 1	PhaseI/II	Del-1 (PAD)	Intramuscular injection
NV1FGF	Phase II	FGF1 plasmid (PAD)	
Biobypass	Phase II	CAD and PAD	
HIF-1a	Phase I	CAD and PAD	

for the failure are unclear since the results of the trial have not been published to date. Overall, however, the results from these and other gene-based therapy trials have demonstrated that these approaches, if well designed, are safe and may still hold the promise of future success.

At this juncture, scientists and clinicians need to reevaluate the critical issues revealed by these studies in order to progress. Clearly, there is a need to better understand the biology of compensatory vascular growth before rushing into the next clinical trial.

5 Going Forward

Going forward, there are several basic issues that need to be addressed.

5.1 Appropriate Animal Model with Accurate Readout of Arteriogenesis

Preclinical models that rely on young healthy animals may represent an entirely different biology than that of older human patients with advanced atherosclerosis. Therapeutic modalities

should therefore be evaluated in older, atherosclerotic animals.

The methodology used to assess revascularization should accurately assess collateral development, tissue blood flow, and muscle function. The more commonly used methods of histology and Doppler flow are not sufficiently sensitive or accurate. Newer methods using micro-CT and MRI imaging show considerable promise and should become more widely used. Currently, most methods to assess muscle function in animal models following revascularization are quite crude, relying on so-called semiquantitative "ambulatory scores" (e.g. an assessment of toe and foot flexing, dragging, inability to use limb) and ischemic sores (e.g. discoloration, necrosis, amputation). Some reports have used muscle stimulation studies; whether these accurately reflect recovery of muscle function still need further study.

The duration of the animal model should be more representative of the clinical situation. Typically, an animal model is conducted over a 3 to 4 week time frame; in contrast, the clinical trials assess revascularization at 90 and 180 days.

5.2 The Right Growth Factor

The focus, to date, has primarily been on growth factors that stimulate angiogenesis, not arteriogenesis. Thus, although a factor may successfully elicit the local growth of capillaries (which is certainly useful in wound healing), it is unlikely to compensate for the ischemia resulting from a proximal occlusion of a major arterial trunk.

Is VEGF the right growth factor? Are combinations of different growth factors required? As discussed above, there have been promising results from animal studies using combinations of growth factors (e.g. FGF-2 and PDGF-BB, Ang-1 and VEGF, VEGF and HGF). Alternatively, is there an arteriogenesis specific growth factor or stimulus (e.g. a master switch gene)? For a single growth factor to be effective, it would have to initiate a self-propagating cascade of the appropriate proliferative, migratory, and inflammatory responses and orchestrate the responses of several different types of cells involved in the response. Some investigators are evaluating the possible utility of gene therapy using the hypoxia "master switch" gene, HIF-1α. However, since arteriogenesis is not triggered by hypoxia, it is unclear whether this strategy will be successful.

Alternative candidates might be found by a more intensive evaluation of the role of monocytes in arteriogenesis. That these cells play a crucial role is underscored by the observations of depressed arteriogenic responses in animals following monocyte depletion, as well as in the CCR-2 null mice (which lack the receptor for the monocyte specific chemotactic factor, MCP-1). Macrophages are known to produce a variety of growth factors and cytokines including bFGF, TGF-α, GM-CSF, M-CSF, VEGF, IL-8, substance P, IGF-1, and PDGF. Nerve growth factor (NGF) has also been described to be proangiogenic and to improve collateral development in animal models of ischemia. Del-1, an integrin receptor ligand, has also been reported to be proangiogenic in animal models.

5.3 The Role/Utility of Bone Marrow-derived Cells

Vasculogenesis was a term that originally referred to a process restricted to embryonic development; however, more recently

it has been used to describe the formation of a primary vasculature from precursor cells in the adult. The contribution of such a process to ongoing angiogenesis and arteriogenesis is quite controversial. The controversies arise from a number of issues, (1) difficulties in defining what precisely is an endothelial precursor cell; (2) the possible role of cell fusion, in which the nuclear contents of two cells can fuse to form a cell with a twofold increase in chromosome content (for some reason circulating bone marrow cells seem to be particularly susceptible to this phenomenon); and (3) sampling or technical issues arising from the assessment of EPC contribution to new vessels. For example, two recent high-power laser confocal studies clearly demonstrated that cells which previously would have been considered as incorporated into newly forming vasculature were actually incorrectly identified because of the overlap of different fluorescent signals from adjacent cells. However, whether bone marrow-derived precursor cells are incorporated into collaterals or contribute to collateral development by some paracrine mechanism, data from several groups now clearly demonstrate that these cells are recruited to regions of collateral development. Thus, the role and importance of these cells to the process is an area in need of intensive investigation.

5.4 Drug Delivery

Translation of "therapeutic angiogenesis" from the lab bench to the clinic requires an assessment of what are truly clinically viable delivery strategies. With respect to coronary revascularization, prolonged or frequent repetitive intracoronary infusions are probably not acceptable. Moreover, intravascular delivery is presented with the problems of the endothelial barrier and the rapid washout of the drug. Local perivascular delivery may be more feasible if it could be accomplished noninvasively (e.g. thoracoscopically). Another approach is to instill a growth factor intrapericardially, which can be accomplished with a catheter and it obviates the need for open-chest surgery. However, the patient population that this would be applicable to is limited, since many have had prior coronary artery bypass graft surgery and no longer have a pericardial cavity suitable for such local delivery. Another approach might be intramyocardial delivery. While this has the appeal of delivering the drug to the site where required, it is an invasive approach requiring highly specialized equipment and a highly skilled operator. Sustained release formulations, where the growth factor of interest is incorporated into a polymer or gel, can allow prolonged, local release.

While the use of gene therapy has been touted as a means to get around the problems of protein delivery, gene transfer has its own problems. For example, the efficiency of gene transfer seems to be reduced in ischemic tissues. Aging myocytes demonstrate a decreased efficiency of adenoviral mediated gene expression. The proteins, whether administered directly or produced via gene therapy may also have a reduced half-life in ischemic tissue due to ongoing processes of tissue resorption or repair. A general problem with adenoviral vectors is their immunoreactivity. Expression is lost in most tissues within the first few weeks after gene transfer, although some of the more recent "second generation" viral vectors have been reported to last as long as 20 weeks. Further modifications in these vectors may lead to improvements in adenoviral gene transfer.

Improvements in the delivery strategy require an accurate assessment of the amount and residence time of the drug (or other therapeutic agent) delivered to the site. This may be improved by the use of sustained release formulations (for proteins or small molecules) and is clearly necessary when using gene therapy or stem cell therapy. Coadministration of agents such as sodium nitroprusside, serotonin, or acetylcholine should also be considered since these agents appear to increase the efficiency of transfection.

5.5 New Approaches

The use of bone marrow-derived stem cells for therapeutic applications has now been extended to a number of small clinical trials, and the initial data are promising. For example, Miyamoto and coworkers implanted bone marrow–mononuclear cells plus platelets in 12 patients with peripheral arterial disease. The initial results suggested improved blood flow, reduction in pain, and an improvement in walking time. A group in Korea recently published their findings from the Myoblast Autologous Graft in Ischemic Cardiomyopathy (MAGIC) cell-randomized clinical trial. Twenty-seven patients with myocardial infarction who underwent coronary stenting received peripheral blood stem cell infusion ($n = 10$), G-CSF alone ($n = 10$), or control (7). At follow-up, the patients who had received cell infusion improved significantly (indices monitored included exercise capacity, myocardial perfusion, and systolic function) compared to controls. The patients treated with G-CSF also showed an improved cardiac function but had an unexpectedly high rate of in-stent restenosis.

Cell-based gene therapy is a relatively new approach that utilizes autologous cells as vectors following *in vitro* transfection with the transgene of interest. This approach has the advantages of circumventing the inflammatory response through the use of autologous cells. Additionally, stable transfection of the cells helps to achieve a more prolonged expression of the transgene. This approach is also amenable to the use of more complex constructs that could allow a stable, regulatable expression, and also the possibility of expressing more than one transgene. To date, the major approaches for ischemic tissues have used gene transfer by transfected skeletal myoblasts, smooth muscle cells, and angioblasts.

5.6 Patient Selection

The majority of the clinical trials of therapeutic angiogenesis have selected those patients that have undergone multiple failed revascularization attempts including coronary artery bypass graft (CABG) surgery and percutaneous coronary interventions. Is this the population most likely to respond? One might disagree that this may represent a population that is particularly resistant to a new collateral formation or may have utilized (and occluded) preexisting collaterals. Moreover, the time period (probably 2–3 weeks) after the initial therapy that may be required to revascularize may not be applicable to a patient with a life-threatening occlusion of a proximal vessel. A recent consensus panel of experts agreed that the following patient groups may represent better candidates for induction of therapeutic angiogenesis :

1. Patients with single, long-standing occlusion(s) of proximal coronary arteries subtending viable myocardium

2. Patients with multivessel, diffuse disease with evidence of inducible ischemia and myocardial viability
3. Presence of adequate feeder vessels
4. Presence of adequate distal runoff

Therapeutic angiogenesis may also be a preferred therapy in patients with a coronary anatomy that is not suited to angioplasty or surgery. Additionally, there may be genetic factors, as yet not defined, that might play a role in the response to angiogenic stimulation. Clinical studies are always complicated by the complexity of the patients, and one of the background issues in these studies is the possibility that some common medications, such as captopril, aspirin, cyclooxygenase-2 inhibitors, lovastatin, isosorbide dinitrate, and antibiotics might have antiangiogenic effects. A potential risk with proangiogenesis therapy is the possibility that stimulation of new blood vessel growth might also result in undesirable outcomes such as in-plaque angiogenesis, acceleration of proliferative retinopathy or age-related macular degeneration, or cancer. Smokers are commonly known to have a diminished wound healing response, which may make them poor candidates for a therapeutic angiogenesis therapy. The effect of smoking on arteriogenesis has not been studied. All of the clinical studies conducted to date have demonstrated a high variability in response among patients, indicating that there may be some as yet to be determined biological factor(s), which might predict responders. Are there differences, for example, in the local or circulating inhibitors of angiogenesis (e.g. endostatin, angiostatin, platelet factor 4, tumistatin, thrombospondins, $VEGF165_b$)? The search for biomarkers and their validation as predictions of therapeutic response is one of the major issues that need to be addressed to design successful clinical trials.

5.7 Clinical Endpoints

The time-honored endpoint for many cardiovascular trials is "major adverse cardiovascular events," which is a composite endpoint that includes indices including death, myocardial infarctions, recurrent ischemia, and the necessity for intervention. Unfortunately, the complexity of the patient population is such that extremely large and expensive clinical trials would be required to achieve statistical significance. Consequently, clinical endpoints to date have relied on measurements of myocardial ischemia (during exercise testing) and other noninvasive measures of ischemia. These measurements revealed a confounding, potent placebo effect. Potential reasons for a genuine biological improvement in the placebo groups could be due to improved patient care or to the effects of promoting physical activity and exercise (a well-known physiological stimulus for "therapeutic" angiogenesis). The development of a surrogate marker that would identify drug effects is a clear challenge. Most clinical investigators believe that a method to image newly formed vessels or developing collaterals directly would provide such a surrogate. Emerging technologies for the noninvasive assessment of vascularity and perfusion, including magnetic resonance imaging, and positron-emission tomography may provide some of these necessary new tools.

6 Conclusion

Therapeutic arteriogenesis thus remains an unproven clinical treatment for both coronary artery disease and peripheral vascular disease. However, there is certainly

a reason to be optimistic. As with any new therapy, scientists and clinicians must learn from the failures and incorporate the lessons learned into new approaches. The questions have now been asked, and only the future will provide the answer.

See also Gene Therapy and Cardiovascular Diseases; Growth Factors; Vascular Development and Angiogenesis.

Bibliography

Books and Reviews

Hammond, H.K., McKirnan, D. (2001) Angiogenic gene therapy for heart disease: a review of animal studies and clinical trials, *Cardiovasc. Res.* **49**, 561–567.

Helisch, A., Schaper, W. (2003) Arteriogenesis: the development and growth of collateral arteries, *Microcirculation* **10**(1), 83–97.

Schaper, W., Scholz, D. (2003) Factors regulating arteriogenesis, *Arterioscler. Thromb. Vasc. Biol.* **23**, 1143–1151.

Simons, M., Ware, J.A. (2003) Therapeutic angiogenesis in cardiovascular disease, *Nat. Rev. Drug. Discov.* **2**, 863–871.

Simons, M., Bonow, R.O., Chronos, N.A., Cohen, D.J., Giordano, F.J., Hammond, H.K., Laham, R.J., Li, W., Pike, M., Sellke, F.W., Stegmann, T.J., Udelson, J.E., Rosengart, T.K. (2000) Clinical trials in coronary angiogenesis: issues, problems, consensus: an expert panel summary, *Circulation* **102**, E73–E86.

Primary Literature

Aase, K., von Euler, G., Li, X., Ponten, A., Thoren, P., Cao, R., Cao, Y., Olofsson, B., Gebre-Medhin, S., Pekny, M., Alitalo, K., Betsholtz, C., Eriksson, U. (2001) Vascular endothelial growth factor-B-deficient mice display an atrial conduction defect, *Circulation* **104**, 358–364.

Abaci, A., Oguzhan, A., Kahraman, S., Eryol, N.K., Unal, S., Arinc, H., Ergin, A. (1999) Effect of diabetes mellitus on formation of coronary collateral vessels, *Circulation* **99**, 2239–2242.

Achen, M.G., Jeltsch, M., Kukk, E., Makinen, T., Vitali, A., Wilks, A.F., Alitalo, K., Stacker, S.A. (1998) Vascular endothelial growth factor D (VEGF-D) is a ligand for the tyrosine kinases VEGF receptor 2 (Flk1) and VEGF receptor 3 (Flt4), *Proc. Natl. Acad. Sci. U.S.A.* **95**, 548–553.

Ahmet, I., Sawa, Y., Yamaguchi, T., Matsuda, H. (2003) Gene transfer of hepatocyte growth factor improves angiogenesis and function of chronic ischemic myocardium in canine heart, *Ann. Thorac. Surg.* **75**, 1283–1287.

Allt, G., Lawrenson, J. (2001) Pericytes: cell biology and pathology, *Cells Tissues Organs* **169**, 1–11.

Aoki, M., Morishita, R., Taniyama, Y., Kaneda, Y., Ogihara, T. (2000) Therapeutic angiogenesis induced by hepatocyte growth factor: potential gene therapy for ischemic diseases, *J. Atheroscler. Thromb.* **7**, 71–76.

Aoki, M., Morishita, R., Taniyama, Y., Kida, I., Moriguchi, A., Matsumoto, K., Nakamura, T., Kaneda, Y., Higaki, J., Ogihara, T. (2000) Angiogenesis induced by hepatocyte growth factor in non-infarcted myocardium and infarcted myocardium: up-regulation of essential transcription factor for angiogenesis, ets, *Gene. Ther.* **7**, 417–427.

Asahara, T., Isner, J.M. (2002) Endothelial progenitor cells for vascular regeneration, *J. Hematother. Stem Cell Res.* **11**, 171–178.

Asahara, T., Takahashi, T., Matsuda, H., Kalka, C., Chen, D., Iwaguro, H., Inai, Y., Silver, M., Isner, J.M. (1999) VEGF contributes to postnatal neovascularization by mobilizing bone marrow-derived endothelial progenitor cells, *EMBO J.* **18**, 3964–3972.

Autiero, M., Luttun, A., Tjwa, M., Carmeliet, P. (2003) Placental growth factor and its receptor, vascular endothelial growth factor receptor-1: novel targets for stimulation of ischemic tissue revascularization and inhibition of angiogenic and inflammatory disorders, *J. Thromb. Haemost.* **1**, 1356–1370.

Bates, D.O., Cui, T.G., Doughty, J.M., Winkler, M., Sugiono, M., Shields, J.D., Peat, D., Gillatt, D., Harper, S.J. (2002) VEGF165b, an inhibitory splice variant of vascular endothelial growth factor, is down-regulated in renal cell carcinoma, *Cancer Res.* **62**, 4123–4131.

Battegay, E.J., Rupp, J., Iruela-Arispe, L., Sage, E.H., Pech, M. (1994) PDGF-BB modulates endothelial proliferation and angiogenesis in vitro via PDGF beta-receptors, *J. Cell Biol.* **125**, 917–928.

Bell, S.E., Mavila, A., Salazar, R., Bayless, K.J., Kanagala, S., Maxwell, S.A., Davis, G.E. (2001) Differential gene expression during capillary morphogenesis in 3D collagen matrices: regulated expression of genes involved in basement membrane matrix assembly, cell cycle progression, cellular differentiation and G-protein signaling, *J. Cell Sci.* **114**, 2755–2773.

Bikfalvi, A., Klein, S., Pintucci, G., Rifkin, D. (1997) Biological roles of fibroblast growth factor-2, *Endocr. Rev.* **18**, 26–45.

Bikfalvi, A., Savona, C., Perollet, C., Javerzat, S. (1998) New insights in the biology of fibroblast growth factor-2, *Angiogenesis* **1**, 155–173.

Bladt, F., Riethmacher, D., Isenmann, S., Aguzzi, A., Birchmeier, C. (1995) Essential role for the c-met receptor in the migration of myogenic precursor cells into the limb bud, *Nature* **376**, 768–771.

Boros, P., Miller, M. (1995) Hepatocyte growth factor: a multifunctional cytokine, *Lancet* **345**, 293–295.

Bottaro, D.P., Rubin, J.S., Faletto, D., Chan, A., Kmiecik, T., Van de Woude, G., Aaronson, S.A. (1991) Identification of the hepatocyte growth factor receptor as the c-met proto-oncogene product, *Science* **251**, 802–804.

Bourdeau, A., Dumont, D.J., Letarte, M. (1999) A murine model of hereditary hemorrhagic telangiectasia, *J. Clin. Invest.* **104**, 1343–1351.

Bugler, B., Amalric, F., Prats, H. (1991) Alternative initiation of translation determines cytoplasmic or nuclear localization of basic fibroblast growth factor, *Mol. Cell. Biol.* **11**, 573–577.

Burgess, W., Maciag, T. (1989) The heparin-binding (fibroblast) growth factor family of proteins, *Annu. Rev. Biochem.* **58**, 575–606.

Bussolino, F., Di Renzo, M.F., Ziche, M., Bocchietto, E., Olivero, M., Naldini, L., Gaudino, G., Tamagnone, L., Coffer, A., Comoglio, P.M. (1992) Hepatocyte growth factor is a potent angiogenic factor which stimulates endothelial cell motility and growth, *J. Cell Biol.* **119**, 629–641.

Cai, W.-J., Koltai, S., Scholz, D., Schaper, W., Schaper, J. (2001) Connexin 37 and desmin, early markers for arteriogenesis in canine heart, *J. Mol. Cell. Cardiol.* **33**, A17.

Cai, W.J., Koltai, S., Kocsis, E., Scholz, D., Schaper, W., Schaper, J. (2001) Connexin37, not Cx40 and Cx43, is induced in vascular smooth muscle cells during coronary arteriogenesis, *J. Mol. Cell. Cardiol.* **33**, 957–967.

Calza, L., Giardino, L., Giuliani, A., Aloe, L., Levi-Montalcini, R. (2001) Nerve growth factor control of neuronal expression of angiogenetic and vasoactive factors, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4160–4165.

Cantarella, G., Lempereur, L., Presta, M., Ribatti, D., Lombardo, G., Lazarovici, P., Zappala, G., Pafumi, C., Bernardini, R. (2002) Nerve growth factor-endothelial cell interaction leads to angiogenesis in vitro and in vivo, *FASEB J.* **16**, 1307–1309.

Cao, R., Brakenhielm, E., Pawliuk, R., Wariaro, D., Post, M.J., Wahlberg, E., Leboulch, P., Cao, Y., Angiogenic, synergism. (2003) Vascular stability and improvement of hind-limb ischemia by a combination of PDGF-BB and FGF-2, *Nat. Med.* **9**, 604–613.

Carmeliet, P., Luttun, A. (2001) The emerging role of the bone marrow-derived stem cells in (therapeutic) angiogenesis, *Thromb. Haemost.* **86**, 289–297.

Carmeliet, P., Moons, L., Luttun, A., Vincenti, V., Compernolle, V., De Mol, M., Wu, Y., Bono, F., Devy, L., Beck, H., Scholz, D., Acker, T., DiPalma, T., Dewerchin, M., Noel, A., Stalmans, I., Barra, A., Blacher, S., Vandendriessche, T., Ponten, A., Eriksson, U., Plate, K.H., Foidart, J.M., Schaper, W., Charnock-Jones, D.S., Hicklin, D.J., Herbert, J.M., Collen, D., Persico, M.G. (2001) Synergism between vascular endothelial growth factor and placental growth factor contributes to angiogenesis and plasma extravasation in pathological conditions, *Nat. Med.* **7**, 575–583.

Cohen, T., Gitay-Goren, H., Sharon, R., Shibuya, M., Halaban, R., Levi, B.Z., Neufeld, G. (1995) VEGF121, a vascular endothelial growth factor (VEGF) isoform lacking heparin binding ability, requires cell-surface heparan sulfates for efficient binding to the VEGF receptors of human melanoma cells, *J. Biol. Chem.* **270**, 11322–11326.

Cui, T.G., Foster, R.R., Saleem, M., Mathieson, P.W., Gillatt, D.A., Bates, D.O., Harper, S.J. (2004) Differentiated human podocytes

endogenously express an inhibitory isoform of vascular endothelial growth factor (VEGF165b) mRNA and protein, *Am. J. Physiol. Renal. Physiol.* **286**, F767–F773.

D'Amore, P.A., Smith, S.R. (1993) Growth factor effects on cells of the vascular wall: a survey, *Growth Factors* **8**, 61–75.

De corde, L.R. (1968) in: *Early Science at Oxford*, Gunther, R (Ed.) Dawsons of Pall Mall, London, UK.

Deindl, E., Buschmann, I., Hoefer, I.E., Podzuweit, T., Boengler, K., Vogel, S., van Royen, N., Fernandez, B., Schaper, W. (2001) Role of ischemia and of hypoxia-inducible genes in arteriogenesis after femoral artery occlusion in the rabbit, *Circ. Res.* **89**, 779–786.

Dell'Era, P., Belleri, M., Stabile, H., Massardi, M.L., Ribatti, D., Presta, M. (2001) Paracrine and autocrine effects of fibroblast growth factor-4 in endothelial cells, *Oncogene* **20**, 2655–2663.

Detmar, M., Brown, L.F., Schon, M.P., Elicker, B.M., Velasco, P., Richard, L., Fukumura, D., Monsky, W., Claffey, K.P., Jain, R.K. (1998) Increased microvascular density and enhanced leukocyte rolling and adhesion in the skin of VEGF transgenic mice, *J. Invest. Dermatol.* **111**, 1–6.

Dias-Flores, L., Gutierrez, R., Varela, H., Rancel, N., Valladares, F. (1991) Microvascular pericytes: a review of their morphological and functional characteristics, *Histol. Histopathol.* **6**, 269–286.

Dickson, M., Martin, J., Cousins, F., Kulkarni, A., Karlsson, S., Akhurst, R. (1995) Defective haematopoiesis and vasculogenesis in transforming growth factor-β1 knock out mice, *Development* **121**, 1845–1854.

Donate, L., Gherardi, E., Srinivasan, N., Sowdhamini, R., Aparicio, S., Blundell, T. (1994) Molecular evolution and domain structure of plasminogen-related growth factors (HGF-SF) and HGF1/MSP, *Protein. Sci.* **3**, 2378–2394.

Dono, R., Texido, G., Dussel, R., Ehmke, H., Zeller, R. (1998) Impaired cerebral cortex development and blood pressure regulation in FGF-2-deficient mice, *EMBO J.* **17**, 4213–4225.

Duff, S.E., Li, C., Garland, J.M., Kumar, S. (2003) CD105 is important for angiogenesis: evidence and potential applications, *FASEB J.* **17**, 984–992.

Dumont, D.J., Jussila, L., Taipale, J., Lymboussaki, A., Mustonen, T., Pajusola, K., Breitman, M., Alitalo, K. (1998) Cardiovascular failure in mouse embryos deficient in VEGF receptor-3, *Science* **282**, 946–949.

Duvall, C.L., Robert Taylor, W., Weiss, D., Guldberg, R.E. (2004) Quantitative microcomputed tomography analysis of collateral vessel development after ischemic injury, *Am. J. Physiol. Heart Circ. Physiol.* **287**, H302–H310.

Edelberg, J.M., Tang, L., Hattori, K., Lyden, D., Rafii, S. (2002) Young adult bone marrow-derived endothelial precursor cells restore aging-impaired cardiac angiogenic function, *Circ. Res.* **90**, E89–E93.

Feldman, B., Poueymirou, W., Papaioannou, V.E., DeChiara, T.M., Goldfarb, M. (1995) Requirement of FGF-4 for postimplantation mouse development, *Science* **267**, 246–249.

Ferrara, N., Gerber, H.P., LeCouter, J. (2003) The biology of VEGF and its receptors, *Nat. Med.* **9**, 669–676.

Ferrara, N., Carver-Moore, K., Chen, H., Dowd, M., Lu, L., O'Shea, K.S., Powell-Braxton, L., Hillan, K.J., Moore, M.W. (1996) Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene, *Nature* **380**, 439–442.

Filvaroff, E.H., Guillet, S., Zlot, C., Bao, M., Ingle, G., Steinmetz, H., Hoeffel, J., Bunting, S., Ross, J., Carano, R.A., Powell-Braxton, L., Wagner, G.F., Eckert, R., Gerritsen, M.E., French, D.M. (2002) Stanniocalcin 1 alters muscle and bone structure and function in transgenic mice, *Endocrinology* **143**, 3681–3690.

Frick, M., Dulak, J., Cisowski, J., Jozkowicz, A., Zwick, R., Alber, H., Dichtl, W., Schwarzacher, S.P., Pachinger, O., Weidinger, F. (2003) Statins differentially regulate vascular endothelial growth factor synthesis in endothelial and vascular smooth muscle cells, *Atherosclerosis* **170**, 229–236.

Gao, M., Lai, N., McKirnan, M., Roth, D.A., Rubanyi, G.M., Dalton, N., Roth, D., Hammond, H.K. (2004) Increased regional function and perfusion after intracoronary delivery of adenovirus encoding fibroblast gorwth factor 4: report of preclinical data, *Hum. Gene Ther.* **15**, 574–587.

Gerhardt, H., Betsholtz, C. (2003) Endothelial-pericyte interactions in angiogenesis, *Cell Tissue Res.* **314**, 15–23.

Gerritsen, M.E., Tomlinson, J.E., Zlot, C., Ziman, M., Hwang, S. (2003) Using gene expression profiling to identify the molecular basis of the synergistic actions of hepatocyte growth factor and vascular endothelial growth factor in human endothelial cells, *Br. J. Pharmacol.* **140**, 595–610.

Gerritsen, M.E., Soriano, R., Yang, S., Zlot, C., Ingle, G., Toy, K., Williams, P.M. (2003) Branching out: a molecular fingerprint of endothelial differentiation into tube-like structures generated by Affymetrix oligonucleotide arrays, *Microcirculation* **10**, 63–81.

Gimbrone, M.J., Resnick, N., Nagel, T., Khachigian, L., Collins, T., Topper, J. (1997) Hemodynamics, endothelial gene expression, and atherogenesis, *Ann. N. Y. Acad. Sci.* **801**, 1–10.

Gluzman-Poltorak, Z., Cohen, T., Herzog, Y., Neufeld, G. (2000) Neuropilin-2 is a receptor for the vascular endothelial growth factor (VEGF) forms VEGF-145 and VEGF-165 [corrected], *J. Biol. Chem.* **275**, 18040–18045.

Goncalves, L. (1998) Fibroblast growth factor-mediated angiogenesis for the treatment of ischemia. Lessons learned from experimental models and early human experience, *Rev. Port. Cardiol.* **17**, II1–I20.

Goumans, M.J., Mummery, C. (2000) Functional analysis of the TGFbeta receptor/Smad pathway through gene ablation in mice, *Int. J. Dev. Biol.* **44**, 253–265.

Grines, C.L., Watkins, M.W., Mahmarian, J.J., Iskandrian, A.E., Rade, J.J., Marrott, P., Pratt, C., Kleiman, N. (2003) A randomized, double-blind, placebo-controlled trial of Ad5FGF-4 gene therapy and its effect on myocardial perfusion in patients with stable angina, *J. Am. Coll. Cardiol.* **42**, 1339–1347.

Grines, C.L., Watkins, M.W., Helmer, G., Penny, W., Brinker, J., Marmur, J.D., West, A., Rade, J.J., Marrott, P., Hammond, H.K., Engler, R.L. (2002) Angiogenic gene therapy (AGENT) trial in patients with stable angina pectoris, *Circulation* **105**, 1291–1297.

Guo, J.S., Cho, C.H., Lam Liu, E.S., Choy, H.T., Wang, J.Y., Leung Koo, M.W. (2002) Antiangiogenic effect of a highly selective cyclooxygenase-2 inhibitor on gastric ulcer healing in rats, *Toxicol. Appl. Pharmacol.* **183**, 41–45.

Hao, X., Mansson-Broberg, A., Blomberg, P., Dellgren, G., Siddiqui, A.J., Grinnemo, K.H., Wardell, E., Sylven, C. (2004) Angiogenic and cardiac functional effects of dual gene transfer of VEGF-A165 and PDGF-BB after myocardial infarction, *Biochem. Biophys. Res. Commun.* **322**, 292–296.

Hao, X., Mansson-Broberg, A., Gustafsson, T., Grinnemo, K.H., Blomberg, P., Siddiqui, A.J., Wardell, E., Sylven, C. (2004) Angiogenic effects of dual gene transfer of bFGF and PDGF-BB after myocardial infarction, *Biochem. Biophys. Res. Commun.* **315**, 1058–1063.

Heil, M., Ziegelhoeffer, T., Wagner, S., Fernandez, B., Helisch, A., Martin, S., Tribulova, S., Kuziel, W.A., Bachmann, G., Schaper, W. (2004) Collateral artery growth (arteriogenesis) after experimental arterial occlusion is impaired in mice lacking CC-chemokine receptor-2, *Circ. Res.* **94**, 671–677.

Helisch, A., Wagner, S., Brandt, U., Heil, M., Ziegelhoeffer, T., Bachmann, G., Schaper, W. (2003) Genetic background is a major determinant of arteriogenesis after femoral artery ligation in inbred mice, *J. Am. Coll. Cardiol.* **41**, 279A.

Hellstrom, M., Gerhardt, H., Kalen, M., Li, X., Eriksson, U., Wolburg, H., Betsholtz, C. (2001) Lack of pericytes leads to endothelial hyperplasia and abnormal vascular morphogenesis, *J. Cell Biol.* **153**, 543–553.

Henry, P.D. (1993) Hypercholesterolemia and angiogenesis, *Am. J. Cardiol.* **72**, 61 C–64C.

Henry, T.D., Annex, B.H., McKendall, G.R., Azrin, M.A., Lopez, J.J., Giordano, F.J., Shah, P.K., Willerson, J.T., Benza, R.L., Berman, D.S., Gibson, C.M., Bajamonde, A., Rundle, A.C., Fine, J., McCluskey, E.R. (2003) The VIVA trial: vascular endothelial growth factor in ischemia for vascular angiogenesis, *Circulation* **107**, 1359–1365.

Ho, H.K., Jang, J.J., Kaji, S., Spektor, G., Fong, A., Yang, P., Hu, B.S., Schatzman, R., Quertermous, T., Cooke, J.P. (2004) Developmental endothelial locus-1 (Del-1), a novel angiogenic protein: its role in ischemia, *Circulation* **109**, 1314–1319.

Hockel, M., Burke, J.F. (1989) Angiotropin treatment prevents flap necrosis and enhances dermal regeneration in rabbits, *Arch. Surg.* **124**, 693–698.

Hoefer, I.E., van Royen, N., Rectenwald, J.E., Deindl, E., Hua, J., Jost, M., Grundmann, S., Voskuil, M., Ozaki, C.K., Piek, J.J., Buschmann, I.R. (2004) Arteriogenesis proceeds via

ICAM-1/Mac-1- mediated mechanisms, *Circ. Res.* **94**, 1179–1185.

Hu, G.F. (1998) Neomycin inhibits angiogenin-induced angiogenesis, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9791–9795.

Hu, G.F. (2001) Neomycin inhibits the angiogenic activity of fibroblast and epidermal growth factors, *Biochem. Biophys. Res. Commun.* **287**, 870–874.

Imamura, T., Engleka, K., Zhan, X., Tokita, Y., Forough, R., Roeder, D., Jackson, A., Maier, J., Hla, T., Maciag, T. (1990) Recovery of mitogenic activity of a growth factor mutant with a nuclear translocation sequence, *Science* **249**, 1567–1570.

Ito, W.D., Arras, M., Winkler, B., Scholz, D., Schaper, J., Schaper, W. (1997) Monocyte chemotactic protein-1 increases collateral and peripheral conductance after femoral artery occlusion, *Circ. Res.* **80**, 829–837.

Iyer, S., Leonidas, D.D., Swaminathan, G.J., Maglione, D., Battisti, M., Tucci, M., Persico, M.G., Acharya, K.R. (2001) The crystal structure of human placenta growth factor-1 (PlGF-1), an angiogenic protein, at 2.0 A resolution, *J. Biol. Chem.* **276**, 12153–12161.

Javerzat, S., Auguste, P., Bikfalvi, A. (2002) The role of fibroblast growth factors in vascular development, *Trends Mol. Med.* **8**, 483–489.

Jeffers, M., Rong, S., Woude, G.F. (1996) Hepatocyte growth factor/scatter factor-Met signaling in tumorigenicity and invasion/metastasis, *J. Mol. Med.* **74**, 505–513.

Jennische, E., Ekberg, S., Matejka, G.L. (1993) Expression of hepatocyte growth factor in growing and regenerating rat skeletal muscle, *Am. J. Physiol.* **265**, C122–C128.

Jin, H., Wyss, J.M., Yang, R., Schwall, R. (2004) The therapeutic potential of hepatocyte growth factor for myocardial infarction and heart failure, *Curr. Pharm. Des.* **10**, 2525–2533.

Joukov, V., Kaipainen, A., Jeltsch, M., Pajusola, K., Olofsson, B., Kumar, V., Eriksson, U., Alitalo, K. (1997) Vascular endothelial growth factors VEGF-B and VEGF-C, *J. Cell. Physiol.* **173**, 211–215.

Joukov, V., Sorsa, T., Kumar, V., Jeltsch, M., Claesson-Welsh, L., Cao, Y., Saksela, O., Kalkkinen, N., Alitalo, K. (1997) Proteolytic processing regulates receptor specificity and activity of VEGF-C, *EMBO J.* **16**, 3898–3911.

Kahn, J., Mehraban, F., Ingle, G., Xin, X., Bryant, J., Vehar, G., Schoenfeld, J., Grimaldi, C., Peale, F., Drakharapu, A., Lewin, D., Gerritsen, M. (2000) Gene expression profiling in an in vitro model of angiogenesis, *Am. J. Pathol.* **156**, 1887–1900.

Kaipainen, A., Korhonen, J., Mustonen, T., van Hinsbergh, V.W., Fang, G.H., Dumont, D., Breitman, M., Alitalo, K. (1995) Expression of the fms-like tyrosine kinase 4 gene becomes restricted to lymphatic endothelium during development, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3566–3570.

Kamihata, H., Matsubara, H., Nishiue, T., Fujiyama, S., Tsutsumi, Y., Ozono, R., Masaki, H., Mori, Y., Iba, O., Tateishi, E., Kosaki, A., Shintani, S., Murohara, T., Imaizumi, T., Iwasaka, T. (2001) Implantation of bone marrow mononuclear cells into ischemic myocardium enhances collateral perfusion and regional function via side supply of angioblasts, angiogenic ligands, and cytokines, *Circulation* **104**, 1046–1052.

Kang, H.J., Kim, H.S., Zhang, S.Y., Park, K.W., Cho, H.J., Koo, B.K., Kim, Y.J., Soo Lee, D., Sohn, D.W., Han, K.S., Oh, B.H., Lee, M.M., Park, Y.B. (2004) Effects of intracoronary infusion of peripheral blood stem-cells mobilised with granulocyte-colony stimulating factor on left ventricular systolic function and restenosis after coronary stenting in myocardial infarction: the MAGIC cell randomised clinical trial, *Lancet* **363**, 751–756.

Keyt, B.A., Berleau, L.T., Nguyen, H.V., Chen, H., Heinsohn, H., Vandlen, R., Ferrara, N. (1996) The carboxyl-terminal domain (111–165) of vascular endothelial growth factor is critical for its mitogenic potency, *J. Biol. Chem.* **271**, 7788–7795.

Keyt, B.A., Nguyen, H.V., Berleau, L.T., Duarte, C.M., Park, J., Chen, H., Ferrara, N. (1996) Identification of vascular endothelial growth factor determinants for binding KDR and FLT-1 receptors. Generation of receptor-selective VEGF variants by site-directed mutagenesis, *J. Biol. Chem.* **271**, 5638–5646.

Klagsbrun, M., Takashima, S., Mamluk, R. (2002) The role of neuropilin in vascular and tumor biology, *Adv. Exp. Med. Biol.* **515**, 33–48.

Kleiman, N.S., Patel, N.C., Allen, K.B., Simons, M., Yla-Herttuala, S., Griffin, E., Dzau, V.J. (2003) Evolving revascularization approaches for myocardial ischemia, *Am. J. Cardiol.* **92**, 9N–17N.

Kondo, I., Ohmori, K., Oshita, A., Takeuchi, H., Fuke, S., Shinomiya, K., Noma, T., Namba, T., Kohno, M. (2004) Treatment of acute

myocardial infarction by hepatocyte growth factor gene transfer: the first demonstration of myocardial transfer of a ''functional'' gene using ultrasonic microbubble destruction, *J. Am. Coll. Cardiol.* **44**, 644–653.

Kornowski, R. (2003) Collateral formation and clinical variables in obstructive coronary artery disease: the influence of hypercholesterolemia and diabetes mellitus, *Coron. Artery Dis.* **14**, 61–64.

Krussel, J.S., Behr, B., Milki, A.A., Hirchenhain, J., Wen, Y., Bielfeld, P., Lake Polan, M. (2001) Vascular endothelial growth factor (VEGF) mRNA splice variants are differentially expressed in human blastocysts, *Mol. Hum. Reprod.* **7**, 57–63.

Kukk, E., Lymboussaki, A., Taira, S., Kaipainen, A., Jeltsch, M., Joukov, V., Alitalo, K. (1996) VEGF-C receptor binding and pattern of expression with VEGFR-3 suggests a role in lymphatic vascular development, *Development* **122**, 3829–3837.

Kulkarni, A., Huh, C., Becker, D., Geiser, A., Lyght, M., Flanders, K., Roberts, A., Sporn, M., Ward, J., Karlsson, S. (1993) Transforming growth factor β1 null mutation in mice causes excessive inflammatory response and early death, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 770–774.

Kumar, R., Yoneda, J., Bucana, C.D., Fidler, I.J. (1998) Regulation of distinct steps of angiogenesis by different angiogenic molecules, *Int. J. Oncol.* **12**, 749–757.

Lagercrantz, J., Farnebo, F., Larsson, C., Tvrdik, T., Weber, G., Piehl, F. (1998) A comparative study of the expression patterns for vegf, vegf-b/vrf and vegf-c in the developing and adult mouse, *Biochim. Biophys. Acta* **1398**, 157–163.

Laham, R.J., Sellke, F.W., Edelman, E.R., Pearlman, J.D., Ware, J.A., Brown, D.L., Gold, J.P., Simons, M. (1999) Local perivascular delivery of basic fibroblast growth factor in patients undergoing coronary bypass surgery: results of a phase I randomized, double-blind, placebo-controlled trial, *Circulation* **100**, 1865–1871.

Larsson, J., Goumans, M., Sjostrand, L., van Rooijen, M., Ward, D., Leveen, P., Xu, X., ten Dijke, P., Mummery, C., Karlsson, S. (2001) Abnormal angiogenesis but intact hematopoietic potential in TGF-β type I receptor-deficient mice, *EMBO J.* **20**, 1663–1673.

Laufs, U., Werner, N., Link, A., Endres, M., Wassmann, S., Jurgens, K., Miche, E., Bohm, M., Nickenig, G. (2004) Physical training increases endothelial progenitor cells, inhibits neointima formation, and enhances angiogenesis, *Circulation* **109**, 220–226.

Lederman, R.J., Mendelsohn, F.O., Anderson, R.D., Saucedo, J.F., Tenaglia, A.N., Hermiller, J.B., Hillegass, W.B., Rocha-Singh, K., Moon, T.E., Whitehouse, M.J., Annex, B.H. (2002) Therapeutic angiogenesis with recombinant fibroblast growth factor-2 for intermittent claudication (the TRAFFIC study): a randomised trial, *Lancet* **359**, 2053–2058.

Lee, J., Gray, A., Yuan, J., Luoh, S.M., Avraham, H., Wood, W.I. (1996) Vascular endothelial growth factor-related protein: a ligand and specific activator of the tyrosine kinase receptor Flt4, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1988–1992.

Lee, R.J., Springer, M.L., Blanco-Bose, W.E., Shaw, R., Ursell, P.C., Blau, H.M. (2000) VEGF gene delivery to myocardium: deleterious effects of unregulated expression, *Circulation* **102**, 898–901.

Leveen, P., Pekny, M., Gebre-Medhin, S., Swolin, B., Larsson, E., Betsholtz, C. (1994) Mice deficient for PDGF B show renal, cardiovascular, and hematological abnormalities, *Genes Dev.* **8**, 1875–1887.

Lindahl, P., Johansson, B.R., Leveen, P., Betsholtz, C. (1997) Pericyte loss and microaneurysm formation in PDGF-B-deficient mice, *Science* **277**, 242–245.

Luttun, A., Carmeliet, G., Carmeliet, P. (2002) Vascular progenitors: from biology to treatment, *Trends Cardiovasc. Med.* **12**, 88–96.

Luttun, A., Tjwa, M., Moons, L., Wu, Y., Angelillo-Scherrer, A., Liao, F., Nagy, J.A., Hooper, A., Priller, J., De Klerck, B., Compernolle, V., Daci, E., Bohlen, P., Dewerchin, M., Herbert, J.M., Fava, R., Matthys, P., Carmeliet, G., Collen, D., Dvorak, H.F., Hicklin, D.J., Carmeliet, P. (2002) Revascularization of ischemic tissues by PlGF treatment, and inhibition of tumor angiogenesis, arthritis and atherosclerosis by anti-Flt1, *Nat. Med.* **8**, 831–840.

Ma, L., del Soldato, P., Wallace, J.L. (2002) Divergent effects of new cyclooxygenase inhibitors on gastric ulcer healing: Shifting the angiogenic balance, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13243–13247.

Ma, P., Maulik, G., Christensen, J., Salgia, R. (2003) c-Met: structure, functions and potential for therapeutic inhibition, *Cancer Metastasis Rev.* **22**, 309–325.

Maisonpierre, P.C., Suri, C., Jones, P.F., Bartunkova, S., Wiegand, S.J., Radziejewski, C., Compton, D., McClain, J., Aldrich, T.H., Papadopoulos, N., Daly, T.J., Davis, S., Sato, T.N., Yancopoulos, G.D. (1997) Angiopoietin-2, a natural antagonist for Tie2 that disrupts in vivo angiogenesis [see comments], *Science* **277**, 55–60.

Mandriota, S.J., Pepper, M.S. (1998) Regulation of angiopoietin-2 mRNA levels in bovine microvascular endothelial cells by cytokines and hypoxia, *Circ. Res.* **83**, 852–859.

Massague, J. (1998) TGF-beta signal transduction, *Annu. Rev. Biochem.* **67**, 753–791.

Matsumoto, K., Nakamura, T. (1996) Emerging multipotent aspects of hepatocyte growth factor, *J. Biochem.* **119**, 591–600.

McCarthy, M., Crowther, M., Bell, P., Brindle, N. (1998) The endothelial receptor tyrosine kinase tie-1 is upregulated by hypoxia and vascular endothelial growth factor, *FEBS Lett.* **423**, 334–338.

McCollum, G.W., Rajaratnam, V.S., Bullard, L.E., Yang, R., Penn, J.S. (2004) Herbimycin A inhibits angiogenic activity in endothelial cells and reduces neovascularization in a rat model of retinopathy of prematurity, *Exp. Eye Res.* **78**, 987–995.

McKirnan, M., Guo, X., Waldman, L., Dalton, N., Lai, N., Gao, M., Roth, D., Hammond, H. (2000) Intracoronary gene transfer of fibroblast growth factor-4 increases regional contractile function and responsiveness to adrenergic stimulation in heart failure, *Card. Vascular Regenerat.* **1**, 11–21.

Merchan, J.R., Chan, B., Kale, S., Schnipper, L.E., Sukhatme, V.P. (2003) In vitro and in vivo induction of antiangiogenic activity by plasminogen activators and captopril, *J. Natl. Cancer Inst.* **95**, 388–399.

Miyamoto, M., Yasutake, M., Takano, H., Takagi, H., Takagi, G., Mizuno, H., Kumita, S., Takano, T. (2004) Therapeutic angiogenesis by autologous bone marrow cell implantation for refractory chronic peripheral arterial disease using assessment of neovascularization by ^{99m}Tc-tetrofosmin (TF) perfusion scintigraphy, *Cell Transplant.* **13**, 429–437.

Molteni, A., Ward, W.F., Ts'ao, C.H., Taylor, J., Small, W. Jr., Brizio-Molteni, L., Veno, P.A. (2003) Cytostatic properties of some angiotensin I converting enzyme inhibitors and of angiotensin II type I receptor antagonists, *Curr. Pharm. Des.* **9**, 751–761.

Montesano, R., Soriano, J.V., Malinda, K.M., Ponce, M.L., Bafico, A., Kleinman, H.K., Bottaro, D.P., Aaronson, S.A. (1998) Differential effects of hepatocyte growth factor isoforms on epithelial and endothelial tubulogenesis, *Cell Growth Differ.* **9**, 355–365.

Morimoto, A., Okamura, K., Hamanaka, R., Sato, Y., Shima, N., Higashio, K., Kuwano, M. (1991) Hepatocyte growth factor modulates migration and proliferation of human microvascular endothelial cells in culture, *Biochem. Biophys. Res. Commun.* **179**, 1042–1049.

Morishita, R., Aoki, M., Hashiya, N., Yamasaki, K., Kurinami, H., Shimizu, S., Makino, H., Takesya, Y., Azuma, J., Ogihara, T. (2004) Therapeutic angiogenesis using hepatocyte growth factor (HGF), *Curr. Gene Ther.* **4**, 199–206.

Morishita, R., Nakamura, S., Hayashi, S., Taniyama, Y., Moriguchi, A., Nagano, T., Taiji, M., Noguchi, H., Takeshita, S., Matsumoto, K., Nakamura, T., Higaki, J., Ogihara, T. (1999) Therapeutic angiogenesis induced by human recombinant hepatocyte growth factor in rabbit hind limb ischemia model as cytokine supplement therapy, *Hypertension* **33**, 1379–1384.

Murayama, T., Asahara, T. (2002) Bone marrow-derived endothelial progenitor cells for vascular regeneration, *Curr. Opin. Mol. Ther.* **4**, 395–402.

Neufeld, G., Cohen, T., Gitay-Goren, H., Poltorak, Z., Tessler, S., Sharon, R., Gengrinovitch, S., Levi, B.Z. (1996) Similarities and differences between the vascular endothelial growth factor (VEGF) splice variants, *Cancer Metastasis Rev.* **15**, 153–158.

O'Connell, P., McKenzie, A., Fisicaro, N., Rockman, S., Pearse, M., d'Apice, A. (1992) Endoglin: a 180-kD endothelial cell and macrophage restricted differentiation molecule, *Clin. Exp. Immunol.* **90**, 154–159.

Olofsson, B., Korpelainen, E., Pepper, M.S., Mandriota, S.J., Aase, K., Kumar, V., Gunji, Y., Jeltsch, M.M., Shibuya, M., Alitalo, K., Eriksson, U. (1998) Vascular endothelial growth factor B (VEGF-B) binds to VEGF receptor-1 and regulates plasminogen activator activity in

endothelial cells, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11709–11714.

Ono, K., Matsumori, A., Shioi, T., Furukawa, Y., Sasayama, S. (1997) Enhanced expression of hepatocyte growth factor/c-Met by myocardial ischemia and reperfusion in a rat model, *Circulation* **95**, 2552–2558.

Ornitz, D.M., Xu, J., Colvin, J.S., McEwen, D.G., MacArthur, C.A., Coulier, F., Gao, G., Goldfarb, M. (1996) Receptor specificity of the fibroblast growth factor family, *J. Biol. Chem.* **271**, 15292–15297.

Ortega, S., Ittmann, M., Tsang, S., Ehrlich, M., Basilico, C. (1998) Neuronal defects and delayed wound healing in mice lacking fibroblast growth factor 2, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5672–5677.

Owens, G., Kumar, M., Wamhoff, B. (2003) Molecular regulation of vascular smooth muscle cell differentiation in development and disease, *Physiol. Rev.* **84**, 767–801.

Ozawa, C.R., Banfi, A., Glazer, N.L., Thurston, G., Springer, M.L., Kraft, P.E., McDonald, D.M., Blau, H.M. (2004) Microenvironmental VEGF concentration, not total dose, determines a threshold between normal and aberrant angiogenesis, *J. Clin. Invest.* **113**, 516–527.

Paul, S.A., Simons, J.W., Mabjeesh, N.J. (2004) HIF at the crossroads between ischemia and carcinogenesis, *J. Cell. Physiol.* **200**, 20–30.

Petersen, W., Pufe, T., Unterhauser, F., Zantop, T., Mentlein, R., Weiler, A. (2003) The splice variants 120 and 164 of the angiogenic peptide vascular endothelial cell growth factor (VEGF) are expressed during Achilles tendon healing, *Arch. Orthop. Trauma Surg.* **123**, 475–480.

Phillips, G.D., Whitehead, R.A., Knighton, D.R. (1992) Inhibition by methylprednisolone acetate suggests an indirect mechanism for TGF-B induced angiogenesis, *Growth Factors* **6**, 77–84.

Pierelli, L., Bonanno, G., Rutella, S., Marone, M., Scambia, G., Leone, G. (2001) CD105 (endoglin) expression on hematopoietic stem/progenitor cells, *Leuk. Lymphoma* **42**, 1195–1206.

Pipili-Synetos, E., Papageorgiou, A., Sakkoula, E., Sotiropoulou, G., Fotsis, T., Karakiulakis, G., Maragoudakis, M.E. (1995) Inhibition of angiogenesis, tumour growth and metastasis by the NO-releasing vasodilators, isosorbide mononitrate and dinitrate, *Br. J. Pharmacol.* **116**, 1829–1834.

Pipp, F., Heil, M., Issbrucker, K., Ziegelhoeffer, T., Martin, S., van den Heuvel, J., Weich, H., Fernandez, B., Golomb, G., Carmeliet, P., Schaper, W., Clauss, M. (2003) VEGFR-1-selective VEGF homologue PlGF is arteriogenic: evidence for a monocyte-mediated mechanism, *Circ. Res.* **92**, 378–385.

Poltorak, Z., Cohen, T., Neufeld, G. (2000) The VEGF splice variants: properties, receptors, and usage for the treatment of ischemic diseases, *Herz* **25**, 126–129.

Puri, M., Rossant, J., Alitalo, K., Bernstein, A., Partanen, J. (1995) The receptor tyrosine kinase TIE is required for integrity and survival of vascular endothelial cells, *EMBO J.* **14**, 5884–5891.

Rajantie, I., Ilmonen, M., Alminaite, A., Ozerdem, U., Alitalo, K., Salven, P. (2004) Adult bone marrow-derived cells recruited during angiogenesis comprise precursors for periendothelial vascular mural cells, *Blood* **104**, 2084–2086.

Raychaudhuri, S.K., Raychaudhuri, S.P., Weltman, H., Farber, E.M. (2001) Effect of nerve growth factor on endothelial cell biology: proliferation and adherence molecule expression on human dermal microvascular endothelial cells, *Arch. Dermatol. Res.* **293**, 291–295.

Rehman, J., Li, J., Parvathaneni, L., Karlsson, G., Panchal, V.R., Temm, C.J., Mahenthiran, J., March, K.L. (2004) Exercise acutely increases circulating endothelial progenitor cells and monocyte-/macrophage-derived angiogenic cells, *J. Am. Coll. Cardiol.* **43**, 2314–2318.

Rissanen, T.T., Markkanen, J.E., Gruchala, M., Heikura, T., Puranen, A., Kettunen, M.I., Kholova, I., Kauppinen, R.A., Achen, M.G., Stacker, S.A., Alitalo, K., Yla-Herttuala, S. (2003) VEGF-D is the strongest angiogenic and lymphangiogenic effector among VEGFs delivered into skeletal muscle via adenoviruses, *Circ. Res.* **92**, 1098–1106.

Rivard, A., Isner, J.M. (1998) Angiogenesis and vasculogenesis in treatment of cardiovascular disease, *Mol. Med.* **4**, 429–440.

Rivard, A., Berthou-Soulie, L., Principe, N., Kearney, M., Curry, C., Branellec, D., Semenza, G.L., Isner, J.M. (2000) Age-dependent defect in vascular endothelial growth factor expression is associated with reduced hypoxia-inducible factor 1 activity, *J. Biol. Chem.* **275**, 29643–29647.

Robinson, C.J., Stringer, S.E. (2001) The splice variants of vascular endothelial growth factor (VEGF) and their receptors, *J. Cell Sci.* **114**, 853–865.

Rosen, E.M., Lamszus, K., Laterra, J., Polverini, P.J., Rubin, J.S., Goldberg, I.D. (1997) HGF/SF in angiogenesis, *Ciba Found. Symp.* **212**, 215–226; discussion 227–219.

Roth, D.M., Lai, N.C., Gao, M.H., Drumm, J.D., Jimenez, J., Feramisco, J.R., Hammond, H.K. (2004) Indirect intracoronary delivery of adenovirus encoding adenylyl cyclase increases left ventricular contractile function in mice, *Am. J. Physiol. Heart Circ. Physiol.* **287**, H172–H177.

Santos, P.M., Winterowd, J.G., Allen, G.G., Bothwell, M.A., Rubel, E.W. (1991) Nerve growth factor: increased angiogenesis without improved nerve regeneration, *Otolaryngol. Head Neck Surg.* **105**, 12–25.

Sasaki, K., Duan, J., Murohara, T., Ikeda, H., Shintani, S., Shimada, T., Akita, T., Egami, K., Imaizumi, T. (2003) Rescue of hypercholesterolemia-related impairment of angiogenesis by oral folate supplementation, *J. Am. Coll. Cardiol.* **42**, 364–372.

Satchell, S., Harper, S., Mathieson, P. (2001) Angiopoietin-1 is normally expressed by periendothelial cells, *Thromb. Haemost.* **86**, 1597–1598.

Sato, T., Qin, Y., Kozak, C., Audus, K. (1993) Tie-1 and Tie-2 define another class of putative receptor tyrosine kinase genes expressed in early embryonic vascular system, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9355–9358.

Sato, T., Tani, Y., Murao, S., Fujieda, H., Sato, H., Matsumoto, M., Takeuchi, T., Ohtsuki, Y. (2001) Focal enhancement of expression of c-Met/hepatocyte growth factor receptor in the myocardium in human myocardial infarction, *Cardiovasc. Pathol.* **10**, 235–240.

Sato, T.N., Tozawa, Y., Deutsch, U., Wolburg-Buchholz, K., Fujiwara, Y., Gendron-Maguire, M., Gridley, T., Wolburg, H., Risau, W., Qin, Y. (1995) Distinct roles of the receptor tyrosine kinases Tie-1 and Tie-2 in blood vessel formation, *Nature* **376**, 70–74.

Schaper, J., Konig, R., Franz, D., Schaper, W. (1976) The endothelial surface of growing coronary collateral arteries: intimal margination and diapedesis of monocytes: a combined SEM and TEM study, *Virchows Arch. A Pathol. Anat. Histol.* **370**, 193–205.

Schmidt, C., Bladt, F., Goedecke, S., Brinkmann, V., Zschiesche, W., Sharpe, M., Gherardi, E., Birchmeier, C. (1995) Scatter factor/hepatocyte growth factor is essential for liver development, *Nature* **373**, 699–702.

Scholz, D., Ziegelhoeffer, T., Helisch, A., Wagner, S., Friedrich, C., Podzuweit, T., Schaper, W. (2002) Contribution of arteriogenesis and angiogenesis to postocclusive hindlimb perfusion in mice, *J. Mol. Cell. Cardiol.* **34**, 775–787.

Schweiki, D., Itin, A., Soffer, D., Keshet, E. (1992) Vascular endothelial growth factor induced by hypoxia may mediate hypoxia-induced angiogenesis, *Nature* **359**, 843–845.

Sengupta, S., Gherardi, E., Sellers, L.A., Wood, J.M., Sasisekharan, R., Fan, T.P. (2003) Hepatocyte growth factor/scatter factor can induce angiogenesis independently of vascular endothelial growth factor, *Arterioscler. Thromb. Vasc. Biol.* **23**, 69–75.

Shi, Y., Massague, J. (2003) Mechanisms of TGF-beta signaling from cell membrane to the nucleus, *Cell* **113**, 685–700.

Shintani, S., Murohara, T., Ikeda, H., Ueno, T., Sasaki, K., Duan, J., Imaizumi, T. (2001) Augmentation of postnatal neovascularization with autologous bone marrow transplantation, *Circulation* **103**, 897–903.

Shyu, K.G., Wang, M.T., Wang, B.W., Chang, C.C., Leu, J.G., Kuan, P., Chang, H. (2002) Intramyocardial injection of naked DNA encoding HIF-1alpha/VP16 hybrid to enhance angiogenesis in an acute myocardial infarction model in the rat, *Cardiovasc. Res.* **54**, 576–583.

Siddiqui, A.J., Blomberg, P., Wardell, E., Hellgren, I., Eskandarpour, M., Islam, K.B., Sylven, C. (2003) Combination of angiopoietin-1 and vascular endothelial growth factor gene therapy enhances arteriogenesis in the ischemic myocardium, *Biochem. Biophys. Res. Commun.* **310**, 1002–1009.

Silverstein, P. (1992) Smoking and wound healing, *Am. J. Med.* **93**, 22S–24S.

Simons, M., Annex, B.H., Laham, R.J., Kleiman, N., Henry, T., Dauerman, H., Udelson, J.E., Gervino, E.V., Pike, M., Whitehouse, M.J., Moon, T., Chronos, N.A. (2002) Pharmacological treatment of coronary artery disease with recombinant fibroblast growth factor-2: double-blind, randomized, controlled clinical trial, *Circulation* **105**, 788–793.

Skaletz-Rorowski, A., Walsh, K. (2003) Statin therapy and angiogenesis, *Curr. Opin. Lipidol.* **14**, 599–603.

Suri, C., Jones, P.F., Patan, S., Bartunkova, S., Maisonpierre, P.C., Davis, S., Sato, T.N., Yancopoulos, G.D. (1996) Requisite role of angiopoietin-1, a ligand for the TIE2 receptor, during embryonic angiogenesis, *Cell* **87**, 1171–1180.

Takeshita, S., Pu, L.Q., Stein, L.A., Sniderman, A.D., Bunting, S., Ferrara, N., Isner, J.M., Symes, J.F. (1994) Intramuscular administration of vascular endothelial growth factor induces dose-dependent collateral artery augmentation in a rabbit model of chronic limb ischemia, *Circulation* **90**, II228–II234.

Takeshita, S., Tsurumi, Y., Couffinahl, T., Asahara, T., Bauters, C., Symes, J., Ferrara, N., Isner, J.M. (1996) Gene transfer of naked DNA encoding for three isoforms of vascular endothelial growth factor stimulates collateral development in vivo, *Lab. Invest.* **75**, 487–501.

Takeshita, S., Weir, L., Chen, D., Zheng, L.P., Riessen, R., Bauters, C., Symes, J.F., Ferrara, N., Isner, J.M. (1996) Therapeutic angiogenesis following arterial gene transfer of vascular endothelial growth factor in a rabbit model of hindlimb ischemia, *Biochem. Biophys. Res. Commun.* **227**, 628–635.

Takeshita, S., Zheng, L.P., Brogi, E., Kearney, M., Pu, L.Q., Bunting, S., Ferrara, N., Symes, J.F., Isner, J.M. (1994) Therapeutic angiogenesis. A single intraarterial bolus of vascular endothelial growth factor augments revascularization in a rabbit ischemic hind limb model, *J. Clin. Invest.* **93**, 662–670.

Taniyama, Y., Morishita, R., Aoki, M., Nakagami, H., Yamamoto, K., Yamazaki, K., Matsumoto, K., Nakamura, T., Kaneda, Y., Ogihara, T. (2001) Therapeutic angiogenesis induced by human hepatocyte growth factor gene in rat and rabbit hindlimb ischemia models: preclinical study for treatment of peripheral arterial disease, *Gene Ther.* **8**, 181–189.

Taniyama, Y., Morishita, R., Hiraoka, K., Aoki, M., Nakagami, H., Yamasaki, K., Matsumoto, K., Nakamura, T., Kaneda, Y., Ogihara, T. (2001) Therapeutic angiogenesis induced by human hepatocyte growth factor gene in rat diabetic hind limb ischemia model: molecular mechanisms of delayed angiogenesis in diabetes, *Circulation* **104**, 2344–2350.

Thurston, G. (2002) Complementary actions of VEGF and angiopoietin-1 on blood vessel growth and leakage, *J. Anat.* **200**, 575–580.

Thurston, G., Suri, C., Smith, K., McClain, J., Sato, T., Yancopoulos, G., McDonald, D. (1999) Leakage resistant blood vessels in mice transgenically overexpressing angiopoietin-1, *Science* **286**, 2511–2514.

Thurston, G., Rudge, J.S., Ioffe, E., Zhou, H., Ross, L., Croll, S.D., Glazer, N., Holash, J., McDonald, D.M., Yancopoulos, G.D. (2000) Angiopoietin-1 protects the adult vasculature against plasma leakage, *Nat. Med.* **6**, 460–463.

Tsujii, M., Kawano, S., Tsuji, S., Sawaoka, H., Hori, M., DuBois, R.N. (1998) Cyclooxygenase regulates angiogenesis induced by colon cancer cells, *Cell* **93**, 705–716.

Ueda, H., Sawa, Y., Matsumoto, K., Kitagawa-Sakakida, S., Kawahira, Y., Nakamura, T., Kaneda, Y., Matsuda, H. (1999) Gene transfection of hepatocyte growth factor attenuates reperfusion injury in the heart, *Ann. Thorac. Surg.* **67**, 1726–1731.

Uehara, Y., Minowa, O., Mori, C., Shiota, K., Kuno, J., Noda, T., Kitamura, N. (1995) Placental defect and embryonic lethality in mice lacking hepatocyte growth factor/scatter factor, *Nature* **373**, 702–705.

Van Belle, E., Witzenbichler, B., Chen, D., Silver, M., Chang, L., Schwall, R., Isner, J.M. (1998) Potentiated angiogenic effect of scatter factor/hepatocyte growth factor via induction of vascular endothelial growth factor: the case for paracrine amplification of angiogenesis, *Circulation* **97**, 381–390.

Van Belle, E., Rivard, A., Chen, D., Silver, M., Bunting, S., Ferrara, N., Symes, J.F., Bauters, C., Isner, J.M. (1997) Hypercholesterolemia attenuates angiogenesis but does not preclude augmentation by angiogenic cytokines, *Circulation* **96**, 2667–2674.

van Royen, N., Hoefer, I., Buschmann, I., Heil, M., Kostin, S., Deindl, E., Vogel, S., Korff, T., Augustin, H., Bode, C., Piek, J.J., Schaper, W. (2002) Exogenous application of transforming growth factor beta 1 stimulates arteriogenesis in the peripheral circulation, *FASEB J.* **16**, 432–434.

Veikkola, T., Jussila, L., Makinen, T., Karpanen, T., Jeltsch, M., Petrova, T.V., Kubo, H., Thurston, G., McDonald, D.M., Achen, M.G., Stacker, S.A., Alitalo, K. (2001) Signalling via vascular endothelial growth factor receptor-3 is

sufficient for lymphangiogenesis in transgenic mice, *EMBO J.* **20**, 1223–1231.

Vincent, K.A., Feron, O., Kelly, R.A. (2002) Harnessing the response to tissue hypoxia: HIF-1 alpha and therapeutic angiogenesis, *Trends Cardiovasc. Med.* **12**, 362–367.

Vincent, K.A., Shyu, K.G., Luo, Y., Magner, M., Tio, R.A., Jiang, C., Goldberg, M.A., Akita, G.Y., Gregory, R.J., Isner, J.M. (2000) Angiogenesis is induced in a rabbit model of hindlimb ischemia by naked DNA encoding an HIF-1alpha/VP16 hybrid transcription factor, *Circulation* **102**, 2255–2261.

Volpert, O.V., Ward, W.F., Lingen, M.W., Chesler, L., Solt, D.B., Johnson, M.D., Molteni, A., Polverini, P.J., Bouck, N.P. (1996) Captopril inhibits angiogenesis and slows the growth of experimental tumors in rats, *J. Clin. Invest.* **98**, 671–679.

Voskuil, M., van Royen, N., Hoefer, I.E., Seidler, R., Guth, B.D., Bode, C., Schaper, W., Piek, J.J., Buschmann, I.R. (2003) Modulation of collateral artery growth in a porcine hindlimb ligation model using MCP-1, *Am. J. Physiol. Heart Circ. Physiol.* **284**, H1422–H1428.

Waltenberger, J. (2001) Impaired collateral vessel development in diabetes: potential cellular mechanisms and therapeutic implications, *Cardiovasc. Res.* **49**, 554–560.

Weis, M., Heeschen, C., Glassford, A.J., Cooke, J.P. (2002) Statins have biphasic effects on angiogenesis, *Circulation* **105**, 739–745.

Weitz, J.I., Byrne, J., Clagett, G.P., Farkouh, M.E., Porter, J.M., Sackett, D.L., Strandness, D.E. Jr., Taylor, L.M. (1996) Diagnosis and treatment of chronic arterial insufficiency of the lower extremities: a critical review, *Circulation* **94**, 3026–3049.

Wolf, C., Cai, W.J., Vosschulte, R., Koltai, S., Mousavipour, D., Scholz, D., Afsah-Hedjri, A., Schaper, W., Schaper, J. (1998) Vascular remodeling and altered protein expression during growth of coronary collateral arteries, *J. Mol. Cell. Cardiol.* **30**, 2291–2305.

Wong, S., Hamel, L., Chevalier, S., A, P. (2000) Endoglin expression on human microvascular endothelial cells association with betaglycan and formation of higher order complexes with TGF-beta signalling receptors, *Eur. J. Biochem.* **267**, 5550–5560.

Xin, X., Yang, S., Ingle, G., Zlot, C., Rangell, L., Kowalski, J., Schwall, R., Ferrara, N., Gerritsen, M.E. (2001) Hepatocyte growth factor enhances vascular endothelial growth factor-induced angiogenesis in vitro and in vivo, *Am. J. Pathol.* **158**, 1111–1120.

Yang, E.Y., Moses, H.L., cell, migration. (1990) Transforming growth factor beta 1-induced changes in proliferation, and angiogenesis in the chicken chorioallantoic membrane, *J. Cell Biol.* **111**, 731–741.

Yang, W., Ahn, H., Hinrichs, M., Torry, R.J., Torry, D.S. (2003) Evidence of a novel isoform of placenta growth factor (PlGF-4) expressed in human trophoblast and endothelial cells, *J. Reprod. Immunol.* **60**, 53–60.

Zhong, J., Eliceiri, B., Stupack, D., Penta, K., Sakamoto, G., Quertermous, T., Coleman, M., Boudreau, N., Varner, J.A. (2003) Neovascularization of ischemic tissues by gene delivery of the extracellular matrix protein Del-1, *J. Clin. Invest.* **112**, 30–41.

Zhou, M., Sutliff, R., Paul, R., Lorenz, J., Hoying, J., Haudenschild, C., Yin, M., Coffin, J., Kong, L., Kranias, E., Luo, W., Boivin, G., Duffy, J., Pawlowski, S., Doetschman, T. (1998) Fibroblast growth factor 2 control of vascular tone, *Nat. Med.* **4**, 201–207.

Ziegelhoeffer, T., Fernandez, B., Kostin, S., Heil, M., Voswinckel, R., Helisch, A., Schaper, W. (2004) Bone marrow-derived cells do not incorporate into the adult growing vasculature, *Circ. Res.* **94**, 230–238.

Zlot, C., Ingle, G., Hongo, J., Yang, S., Sheng, Z., Schwall, R., Paoni, N., Wang, F., Peale, F.V. Jr., Gerritsen, M.E. (2003) Stanniocalcin 1 is an autocrine modulator of endothelial angiogenic responses to hepatocyte growth factor, *J. Biol. Chem.* **278**, 47654–47659.

Zwaginga, J.J., Doevendans, P. (2003) Stem cell-derived angiogenic/vasculogenic cells: possible therapies for tissue repair and tissue engineering, *Clin Exp Pharmacol Physiol* **30**, 900–908.

Therapeutic Compounds in Nature as Leads for New Pharmaceuticals

Antony D. Buss[1], *Brian Cox*[2], *and Roger D. Waigh*[3]
[1] *MerLion Pharmaceuticals Pte. Ltd, Science Park II, Singapore*
[2] *Novartis Pharma Research Centre, Horsham, UK*
[3] *University of Strathclyde, Glasgow, Scotland*

1 Introduction 325

2 Drugs Affecting the Central Nervous System 327
2.1 Morphine Alkaloids 327
2.2 Conotoxins 329
2.3 Cannabinoids 330
2.4 Asperlicin 332

3 Neuromuscular Blocking Drugs 334
3.1 Curare, Decamethonium, and Atracurium 334

4 Anticancer Drugs 337
4.1 Catharanthus (Vinca) Alkaloids 337
4.2 Camptothecin 339
4.3 Paclitaxel and Docetaxel 340
4.4 Epothilones 343
4.5 Podophyllotoxin, Etoposide, and Teniposide 344
4.6 Marine Sources 347

5 Antibiotics 348
5.1 β-Lactams 348
5.2 Erythromycin Macrolides 354
5.3 Streptogramins 355
5.4 Echinocandins 356

6 Cardiovascular Drugs 357
6.1 Lovastatin, Simvastatin, and Pravastatin 357

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

6.2 Teprotide and Captopril 360
6.3 Adrenaline, Propranolol, and Atenolol 361
6.4 Dicoumarol and Warfarin 362

7 Antiasthma Drugs 363
7.1 Khellin and Sodium Cromoglycate 363
7.2 Ephedrine, Isoprenaline, and Salbutamol 365
7.3 Contignasterol 366

8 Antiparasitic Drugs 366
8.1 Artemisinin, Artemether, and Arteether 366
8.2 Quinine, Chloroquine, and Mefloquine 368
8.3 Avermectins and Milbemycins 372

9 Immunosuppressant Drugs 374
9.1 Cyclosporin, Tacrolimus, and Sirolimus 374
9.2 Mycophenolic Acid and Mycophenolate Mofetil 375

10 Conclusion 376

Acknowledgment 377

Bibliography 377
Books and Reviews 377
Primary Literature 378

Keywords

Antibiotics
Chemical substances that inhibit the growth of microorganisms such as bacteria and fungi.

Drug Discovery
The process of identifying molecules that may be used to treat diseases and ailments.

Natural Products
Chemical substances produced by living organisms

Pharmaceuticals
Drugs that are used to treat human diseases or substances having medicinal properties.

Screening
Testing substances for biological activity and their potential use as pharmaceuticals.

Secondary Metabolites
Chemical compounds (natural products) that are not essential for normal growth, development or reproduction of the producing organism.

For most of Nature's flora and fauna, the struggle for survival is often waged at the molecular level. Over time this intense, unending competition has resulted in an armamentarium of small organic molecules with a structural diversity and biological potency often unrivalled by the best efforts and imaginations of scientists. Many of these natural products have proved invaluable in providing templates for the construction of novel drugs with applications in a wide range of therapeutic areas. In fact, the pharmaceuticals derived from microbial and plant secondary metabolites have been directly responsible for doubling life expectancy in the 20th century and more than half of approved drugs are either natural products or related to them. They have revolutionised clinical practice and are still one of the most important sources of anticancer, antiinfective and immunosuppressant agents.

1 Introduction

Of the 520 new pharmaceuticals approved between 1983 and 1994, 39% were derived from natural products, the proportion of antibacterials and anticancer agents of which was over 60%. Between 1998 and 2004, a total of 21 drugs derived from natural products were launched on the market by major pharmaceutical and biotech companies (Table 1), including caspofungin, daptomycin, ertapenem, orlistat, pimecrolimus, rosuvastatin, sirolimus, and telithromycin. In 2000, one-half of the top-selling pharmaceuticals were derived from natural products, having combined sales of more than US $40 billion. These included the biggest selling anticancer drug paclitaxel, the "statin" family of hypolipidemics, and the immunosuppressant cyclosporin. Despite the evident success in drug discovery, the popularity of natural products waned significantly from the early 1990s. An increasing number of pharmaceutical companies cut back or terminated their natural product drug discovery programs in favor of screening very large libraries of synthetic compounds generated by combinatorial methodologies. However, quality lead molecules for many drug targets have remained elusive and more recently, much greater emphasis has been placed on accessing smaller numbers of chemical compounds (focused libraries), but which possess predetermined, druglike physicochemical properties. It remains to be seen whether the unique features of compounds that are themselves derived from living organisms will once again feature prominently in industrial drug discovery programs.

The abundance of microbial and plant-derived natural products and their value in medicine are undisputed, but the reason why these complex chemical substances, known as *secondary metabolites*,

Tab. 1 Drugs derived from natural products (1998–2004).

Year	*Generic name (trade name)*	*Disease area*
1998	Orlistat (Xenical®)	Antiobesity
1998	Cefoselis (Wincef®)	Antibacterial
1999	Dalfopristin and quinupristin (Synercid®)	Antibacterial
1999	Valrubicin (Valstar®)	Oncology
1999	Colforsin daropate (Adele, Adehl®)	Cardiotonic
2000	Arteether (Artemotil®)	Antimalarial
2001	Ertapenem (Invanz™)	Antibacterial
2001	Caspofungin (Cancidas®)	Antifungal
2001	Telithromycin (Ketek®)	Antibacterial
2001	Pimecrolimus (Elidel®)	Atopic dermatitis
2002	Galantamine (Reminyl®)	Alzheimer's disease
2002	Micafungin (Funguard®)	Antifungal
2002	Amrubicin hydrochloride (Calsed®)	Oncology
2002	Biapenem (Omegacin®)	Antibacterial
2002	Nitisinone (Orfadin®)	Antityrosinaemia
2003	Miglustat (Zavesca®)	Type 1 Gaucher disease
2003	Mycophenolate sodium (Myfortic®)	Immunosuppression
2003	Rosuvastatin (Crestor®)	Dyslipidemia
2003	Pitavastatin (Livalo®)	Dyslipidemia
2003	Daptomycin (Cubicin™)	Antibacterial
2004	Everolimus (Certican™)	Immunosuppression

are produced has been subject to much debate. One popular view is that these compounds have a specific role in protecting the otherwise defenseless producing organisms from attack by predators. Taking morphine as an example of a secondary metabolite whose value to the producing plant is not entirely obvious, 14 steps are required from available amino acids, including at least one step that is highly substrate specific. The presence of morphine in the tissues of *Papaver somniferum* must therefore, confer a selectional advantage on the plant: genetic code is required for each of the enzymes involved in the biosynthesis, valuable amino acids are utilized in forming the enzymes, and a relatively scarce nutrient (nitrogen) is locked up in the compounds produced. If morphine did not continue to have value for the plant, mutants would have arisen with the advantage of not having a drain on their metabolic resources.

We can only guess at the ecological functions of morphine. Perhaps a mammalian herbivore that consumed too many poppies would become drowsy and itself fall prey to a carnivore. It may be significant that the cannabinoids (CB), produced in greatest abundance in the nutritious growing tips of the plant, also induce mental effects that would compromise a herbivore's ability to escape a predator. Whatever their natural protective functions, natural products are a rich source of biologically active compounds that have arisen as the result of natural selection, over perhaps 300 million years. The challenge to the medicinal chemist is to exploit this unique chemical diversity. The following account illustrates

how natural products have been used as what are called *lead compounds*, or templates for the development of important medicines.

2 Drugs Affecting the Central Nervous System

2.1 Morphine Alkaloids

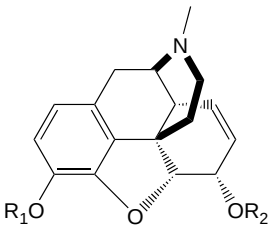

(**1**) Morphine $R_1 = R_2 = H$

(**2**) Codeine $R_1 = CH_3$, $R_2 = H$

(**3**) Heroin $R_1 = R_2 = COCH_3$

The history of the opium alkaloids is too well known to warrant repetition here, but the analgesics based on morphine (**1**) are too important to be left out of an account of natural products as leads. Thus, we shall summarize the clinically more important developments that have occurred since the isolation of morphine in 1803. Codeine (**2**) continues to be used widely for the treatment of moderate pain and, although present in the opium poppy (*P. somniferum*), it is normally synthesized in higher yield from morphine.

Other than codeine, the earliest significant semisynthetic derivative of morphine was the diacetate heroin (**3**), which is still used in terminal cancer where its addictiveness is irrelevant. Acetylation masks the polar hydroxy groups, so that penetration into the central nervous system (CNS) is enhanced; hydrolysis then occurs to liberate the phenolic hydroxyl, giving an active analgesic and ultimately regenerates morphine. Heroin was thus one of the first prodrugs.

Modifications to the C-ring of morphine are legion, but none of the derivatives is free from addictive liability, though many have been used clinically. N-Demethylation and realkylation yield more interesting analogs, notably *N*-allylnormorphine and nalorphine (**4**), which is a morphine antagonist. Further modification leads to naloxone (**5**), which unlike nalorphine has very little agonist activity and has retained a place in therapy for treatment of opiate-induced respiratory depression. Naloxone will also precipitate withdrawal symptoms in opiate addicts, thereby facilitating diagnosis.

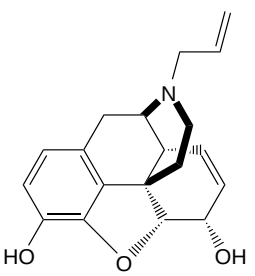

(**4**) Nalorphine

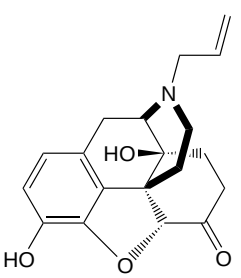

(**5**) Naloxone

Total synthesis of morphine is difficult, but analogs lacking the dihydrofuran ring are accessible from 1-benzylisoquinolines, in analogy with the biosynthesis of morphine, to give the morphinans (**6**). The system may be simplified even further to give the benzomorphans (**7**), although neither these nor the morphinans have provided the long-sought analgesic without addictive properties.

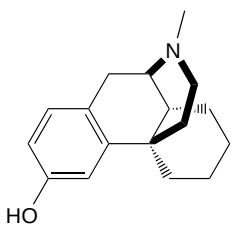

(**6**) Morphinan

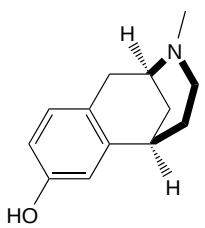

(**7**) Benzomorphan

A semisynthetic route to morphine analogs was found from thebaine (**8**) using Diels–Alder reactions in the C-ring. Adducts such as etorphine (**9**) have the distinction of very high potency, sufficient to immobilize rhinoceroses at moderate dose levels! Unfortunately, the addictive liability runs parallel to the increase in analgesic potency, a tendency that was partly overcome in the analog buprenorphine (**10**).

All this work was carried out in ignorance of the nature of the natural transmitter(s) which subsequently, proved to be the peptides known as *endorphins*

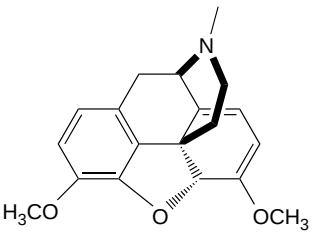

(**8**) Thebaine

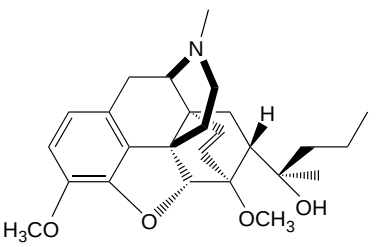

(**9**) Etorphine

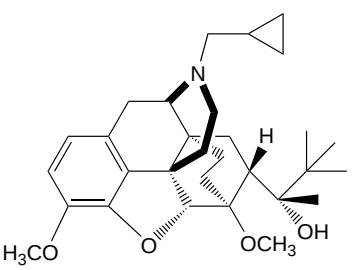

(**10**) Buprenorphine

and their pentapeptide fragments, as the *enkephalins*. It is perhaps significant that vastly improved understanding of the biochemical basis for analgesia and the characterization of a family of related receptors, known as δ, κ, and μ, have so far failed to yield any better drugs for the treatment of pain.

A series of analgesics that were discovered initially in an attempt to obtain smooth muscle relaxants based on another natural product, atropine (**11**),

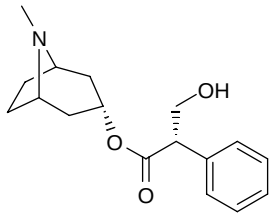

(**11**) Atropine

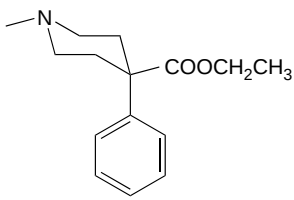

(**12**) Pethidine

started with the observation that meperidine (pethidine) (**12**) unexpectedly produced a reaction in mice known as *Straub tail*, normally characteristic of the morphine series. Meperidine itself is still used widely in childbirth in the belief that there is a lower incidence of respiratory depression in the fetus. The realization that 4-phenylpiperidines, which are not obvious structural analogs of morphine, could give rise to useful analgesic effects, led to the synthesis of many thousands of derivatives, several with far greater potency than that of meperidine. Unfortunately, as potency increases so does addiction liability, and respiratory depression.

2.2 Conotoxins

Elan Pharmaceuticals has developed the peptide SNX-111 (Ziconotide), the synthetic equivalent of ω-Conopeptide-MVIIA, found in the venom of the predatory marine snail *Conus magus*, for the treatment of severe pain and ischemia. Marketing approval in the United States and Europe has recently been granted for SNX-111 as an intrathecal infusion in the treatment of severe, chronic pain. The peptide has the structure H-^{1}Cys-Lys-Gly-Lys-Gly-Ala-Lys-^{8}Cys-Ser-Arg-Leu-Met-Try-Asp-^{15}Cys-^{16}Cys-Thr-Gly-Ser-^{20}Cys-Arg-Ser-Gly-Lys-^{25}Cys-NH_2 cyclic(1–16), (8–20), (15–25)-tris (disulfide), which does not make it an easy target for synthesis and gives it poor distribution properties *in vivo*.

SNX-111 blocks N-type calcium channels, which are located throughout the CNS on neuronal somata, dendrites, dendritic spines, and axon terminals, where they play a major role in the regulation of the *neurotransmitters* associated with pain transmission and stroke. The drive is to discover an orally active, selective, small-molecule modulator of N-type calcium channels to overcome the disadvantages of administration of SNX-111.

High-throughput *screening* campaigns have resulted in a number of leads being identified; whereas others have chosen to modify known drugs shown to block N-type channels. Workers at Parke-Davis, however, employed a ligand-based approach using the three-dimensional solution structure of the peptide. Compounds such as conotoxin analog (**13**) were designed where key binding motifs are attached to an alkylphenyl ether scaffold. The compound had an IC_{50} value of 3.3 μM in a human N-type channel assay, but showed no selectivity over the L-type channel. Structure-activity work on the conotoxins has shown that other regions of the peptide, absent in these synthetic ligands, are responsible for channel family selectivity.

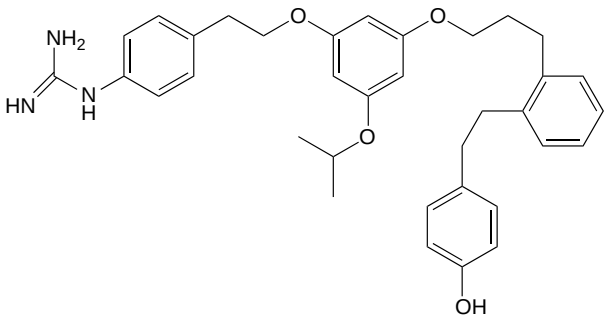

(**13**) Conotoxin analog

2.3 Cannabinoids

The plant *Cannabis sativa* has been used by humans for thousands of years, both for the effects when ingested and for making rope from the fibers in the stem. The major constituent of pharmacological interest is Δ_9-tetrahydrocannabinol (**14**) (THC), which has a multiplicity of actions. In animals, the effects include sedation and apparent hallucinations, which are similar to the major effects in the CNS in humans. There are also cardiovascular effects, notably tachycardia and postural hypotension, which can be separated from the CNS action, as in the synthetic analog $\Delta_{6a,10a}$-dimethylheptyl-THC (**15**), which has minimal CNS activity.

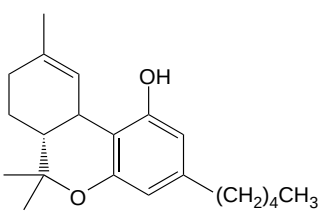

(**14**) THC

Given the widespread illicit use of *C. sativa*, it was perhaps inevitable that eventually one or two cancer patients receiving chemotherapy would dose themselves with their own sedative in the form of marijuana. An unexpected blessing from this uncontrolled combination was a reduction in the nausea experienced during chemotherapy. A variety of anticancer agents cause severe nausea and vomiting, including nitrogen mustard, adriamycin, 5-azacytidine, cyclophosphamide, and methotrexate: a unique situation arose in which the remedy was discovered by the patients themselves. Although smoking reefers gives rapid absorption and close control of the effects, smoking cannot be recommended and thus, when the physicians in charge were made aware of their patients' discovery, they devised a controlled clinical trial in which measured doses of THC were dissolved in sesame oil and administered in gelatin capsules. A placebo was similarly prepared for use in a randomized, double-blind, cross-over

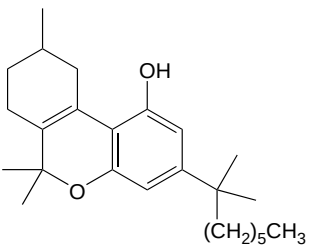

(**15**)

experiment. The results left no doubt that a majority of patients benefited from THC pretreatment, even those who had previously been refractory to the effects of the standard antiemetics such as prochlorperazine. There remained the problem of tachycardia associated with THC treatment. The multiplicity of effects of THC has led to the synthesis of large numbers of analogs, particularly in the hope of finding nonmorphine-like analgesics without addictiveness and without the other CNS effects of THC. The analog nabilone (**16**) had been shown to exert less effect than that of THC on the cardiovascular system, while retaining the mixture of CNS actions, including analgesic, antianxiety, and antipsychotic properties. When tested as an antiemetic, nabilone proved to be superior to THC and has been used for this purpose for more than 30 years.

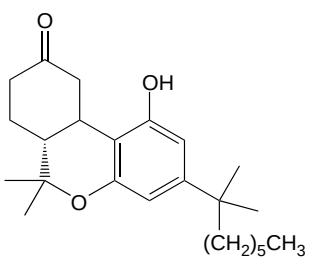

(**16**) Nabilone

After the demonstration of THC binding sites in the CNS, a search for an endogenous ligand produced the long-chain ethanolamine derivative (**17**) of arachidonic acid, known as *anandamide*. Subsequently, the glycerol ester of arachidonic acid (**18**), known as *2-AG*, was shown to be a more abundant endogenous ligand in the brain than anandamide. Further development has tended to concentrate on analogs of the natural ligands, notably the methyl derivative of anandamide (**19**), which is resistant to the amide hydrolase that terminates the action of anandamide itself and the dimethylheptyl analog (**20**) that is traceable to the earlier modifications to THC. Such analogs tend to have activity similar to that of THC.

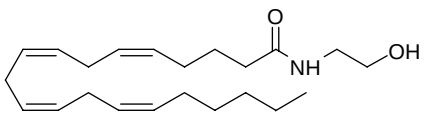

(**17**) Anandamide

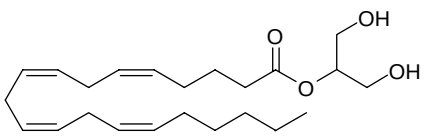

(**18**) 2-AG

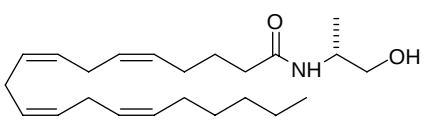

(**19**) *R*-methanandamide

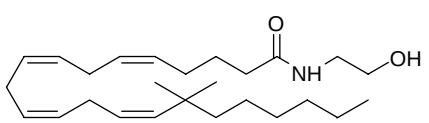

(**20**)

An interesting twist in the tail is provided by the observation that anandamide is also a ligand for the so-called *enigmatic vanilloid receptors*, previously characterized through their interactions with two other natural

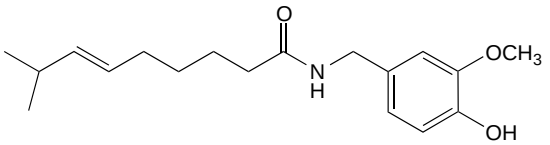

(**21**) Capsaicin

products, capsaicin (**21**) and resiniferatoxin (**22**) responsible for the "hot" sensation caused by compounds in, for example, chillies. A functional vanilloid receptor was cloned in 1997 and is activated by heat and acid as well as the chemical ligands. A combination of the anandamide structure with a vanilloid motif, as in AM404 (**23**), enhances the anandamide transport inhibitory properties. The situation is complex from the viewpoint of drug design, not least because there are two cannabinoid receptors, plus a hydrolase and a transport protein, interference with any or all of which might provide new drugs.

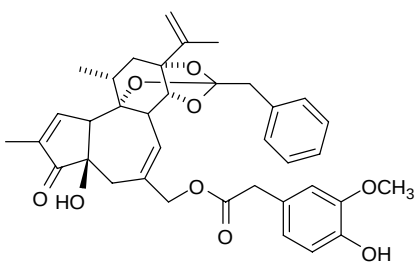

(**22**) Resiniferatoxin

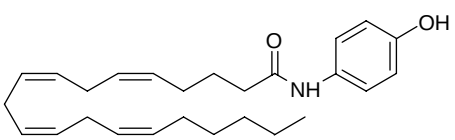

(**23**) AM404

The cannabinoid acids, which are devoid of psychotropic activity, are promising anti-inflammatory agents and it is possible that the next useful therapeutic agent will come from this direction, rather than the sought-after analgesic.

2.4 Asperlicin

Cholecystokinin (CCK) is a peptide hormone, present in the gut and CNS; it is one of the most abundant peptides in the brain. The whole peptide is composed of 33 amino acids, but the C-terminal octapeptide H-Asp-Tyr(SO_3H)-Met-Gly-Trp-Met-Asp-Phe-NH_2 possesses the full range of activities, sufficient for it to be classed as a *neurotransmitter*. Specific, high-affinity binding sites have been found on mammalian CNS cell membranes and in other organs such as pancreas, gall bladder, and colon. The latter have been classed as CCK-A receptors, but the majority of CNS receptors were classed as CCK-B, on the basis of affinity differences for various agonists and antagonists. To confuse the issue slightly, the gastrin receptor in the stomach is closely related to the CCK-B (now known as the CCK_2) receptor and is stimulated by the C-terminal tetrapeptide of CCK: in the periphery, gastrin receptors are the same as CCK_2 receptors.

The effects of CCK on intestinal smooth muscle and pancreas are easy to demonstrate pharmacologically, unlike the role in the CNS, which is a matter for conjecture. It was assumed that the CNS activity must be significant, given the abundance of the

peptide in the brain and that the discovery of antagonists might lead to new drug treatments, as yet unspecified.

Fishing in microbial broths, using radioreceptors as bait, produced asperlicin (**24**), the first potent, competitive, and selective CCK-A (CCK_1) antagonist, from a culture of *Aspergillus alliaceus*.

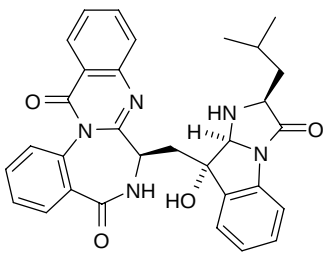

(**24**) Asperlicin

Asperlicin is moderately potent, poorly soluble in water, and not bioavailable by the oral route. When discovered it was also, with morphine, one of the very few nonpeptides with affinity for a peptide receptor (peptoids are discounted in this assessment). It was an interesting target for synthetic modification, particularly viewed as a benzodiazepine derivative with potential CNS activity.

On the basis of the benzodiazepine nucleus and an overt mimic of diazepam, one of the first successful synthetic analogs was L-364,286 (**25**), which had potency on CCK-A receptors similar to that of asperlicin. Better receptor affinity was achieved with 3-amide–substituted benzazepines: the 2-indolyl derivative L-364,718, also known as MK-329 (**26**), is 5 orders of magnitude more potent than asperlicin at CCK-A receptors and is a valuable pharmacological tool.

Modification of the 3-amide to give a urea linkage as in (**27**) led to a reduction in CCK-A receptor affinity. Importantly, discrimination between CCK-A

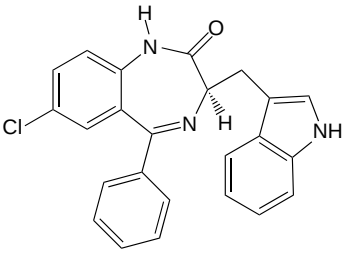

(**25**)

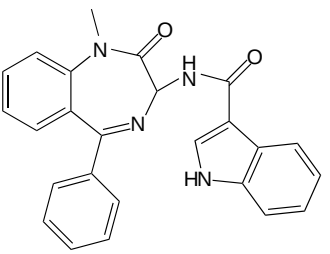

(**26**)

and CCK-B receptors by (**27**) is governed by the stereochemistry at C3, the (S)-enantiomer showing greater affinity for CCK-A receptors. The (R)-enantiomer, known as L-365,260, prefers CCK-B receptors, antagonizes gastrin-stimulated acid secretion in animal models and, among other CNS effects, induces analgesia in primates and displays anxiolytic properties.

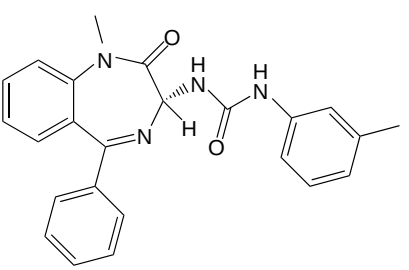

(**27**)

Further development in this series has substantially improved receptor affinity:

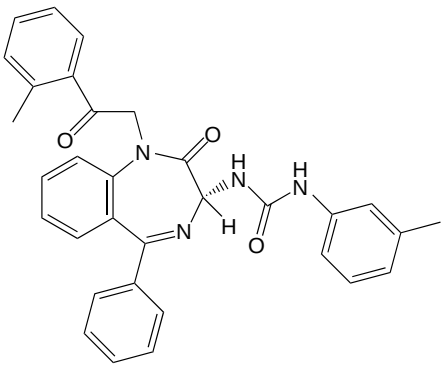

(**28**) YM-022

YM-022 (**28**) has IC_{50} 0.05 nM kg^{-1}. Clinical trials of compounds in this series have been disappointing because of poor bioavailability, but the general concept of finding a therapeutic agent through antagonism of CCK_2 receptors is still viable.

3
Neuromuscular Blocking Drugs

3.1
Curare, Decamethonium, and Atracurium

The development and use of muscle relaxants, to allow a reduction in the level of anesthesia during surgery, follows entirely from studies of South American arrow poisons and particularly from the isolation of pure D-tubocurarine (**29**) in the 1930s from tube curare. Another of the South American blowpipe poisons, *calabash curare*, was used for similar purposes and developed to give alcuronium (**30**) from the alkaloid C-toxiferine 1 (**31**). Both types of curare paralyze skeletal muscle by a similar mechanism, antagonizing the effect of acetylcholine at the neuromuscular junction.

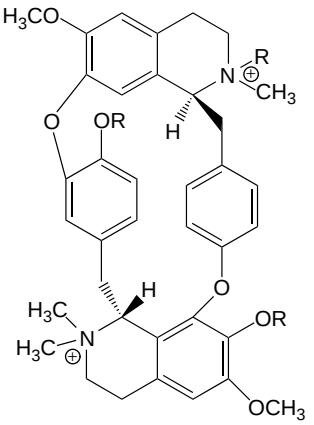

(**29**) Tubocurarine R = H
(**32**) Metocurine R = CH_3

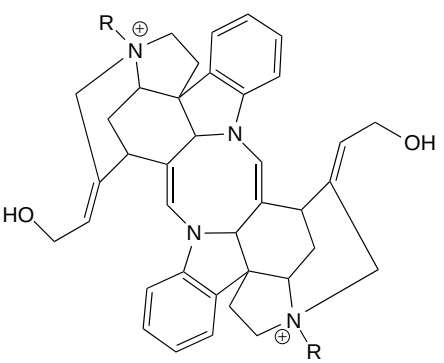

(**30**) Alcuronium R = $CH_2CH{=}CH_2$
(**31**) C-toxiferine 1 R = CH_3

The muscle-paralyzing curare alkaloids are quaternary salts that are not absorbed when taken orally. For surgical procedures, they must be administered by intravenous injection, which results in onset of paralysis in at most a few minutes: anesthesia is normally induced before administration of the muscle relaxant, which is followed by artificial respiration. Although the neuromuscular blocking agents are potentially lethal when administered alone, in the environment of an operating theater they are truly lifesaving drugs that have made

a major impact on survival rates during surgery.

In the 1930s, there were no spectroscopic aids to structure elucidation and it is not surprising that an error was made assigning two quaternary nitrogens to the structure of d-tubocurarine, a mistake that was not corrected until 1970. The methylation product of d-tubocurarine, known as *metocurine* (**32**), is a more potent muscle relaxant. It was known for a long time as dimethyltubocurarine because of the erroneous structure allocated to compound (**29**). The error, in assigning a bisquaternary structure to a molecule with one quaternary and one protonated tertiary nitrogen, led to a large number of highly active synthetic bisquaternaries. The simplest of these was decamethonium (**33**), which was nothing more than two trimethylammonium end groups connected with a decamethylene chain. As one of a series with different chain lengths, decamethonium became the prototype for many more complex structures with 10 atoms between the quaternary centers, which appeared to be optimal for binding to the acetylcholine receptor at the neuromuscular junction.

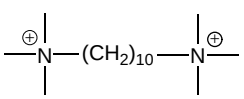

(**33**) Decamethonium

Unlike tubocurarine, decamethonium depolarizes the muscle endplate, rendering the membrane insensitive to acetylcholine. The action of tubocurarine is competitive and can be overcome with increased concentrations of acetylcholine, brought about by administration of an anticholinesterase: the latter is thus, an antidote to tubocurarine, but not to decamethonium. Despite the lack of an antidote, decamethonium was used very widely for over two decades. One of its disadvantages is an overlong duration of action, during which time the patient has to be maintained on artificial respiration, because the muscle of the diaphragm is also susceptible to the actions of the drug. An early and highly successful attempt to shorten the action of decamethonium gave suxamethonium (**34**), a diester formed between succinic acid and two molecules of choline, which hydrolyzes rapidly in the presence of pseudocholinesterase.

Tubocurarine causes cardiovascular side effects induced by direct interactions with ganglionic acetylcholine receptors and from stimulation of histamine release, so analogs have been well worth pursuing. The macrocyclic structure of tubocurarine is a difficult synthetic target, but fortunately, ring-opened analogs, such as laudexium (**35**), have high potency and relatively few side effects. The main problem with (**35**) is the duration of action, which

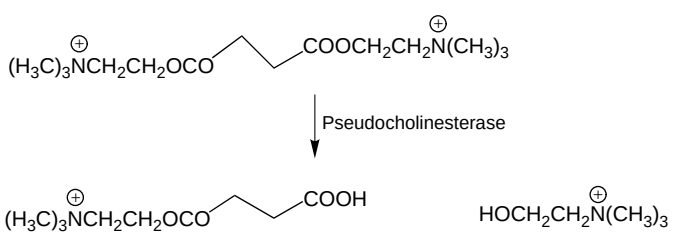

Decomposition of suxamethonium (**34**)

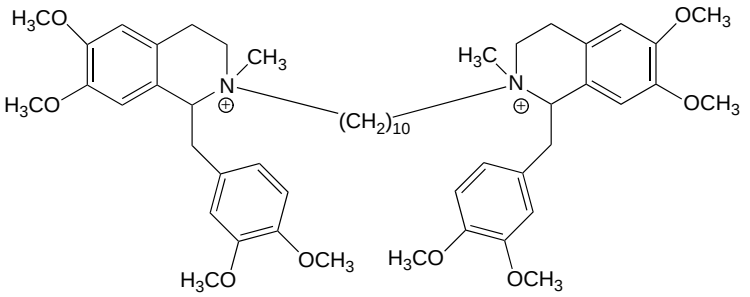

(**35**) Laudexium

at about 40 min is too long for many operations. Two approaches have been used to shorten the duration of action. The concept of pH-controlled Hofmann elimination was employed successfully in the design of atracurium (**36**), which in clinical use has the advantage that the drug disappears at a constant rate, irrespective of liver or kidney function. Some ester hydrolysis contributes to the degradation of atracurium *in vivo*, as might be expected. A slightly later development centered on an empirical search for structures that would undergo ester hydrolysis more rapidly, resulting in mivacurium (**37**), which has a slightly shorter duration of action than that

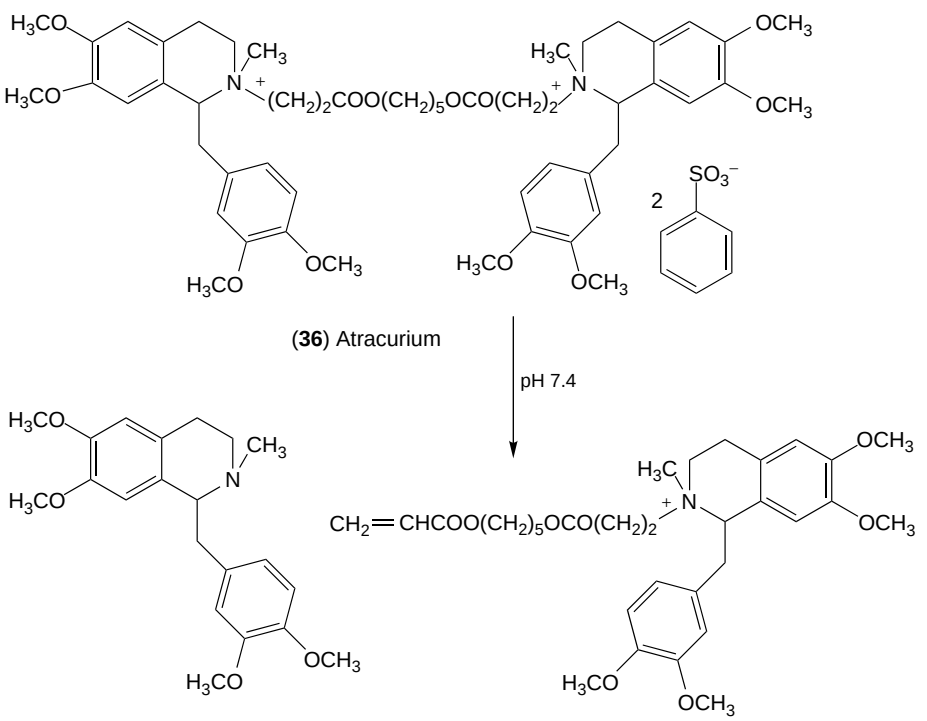

(**36**) Atracurium

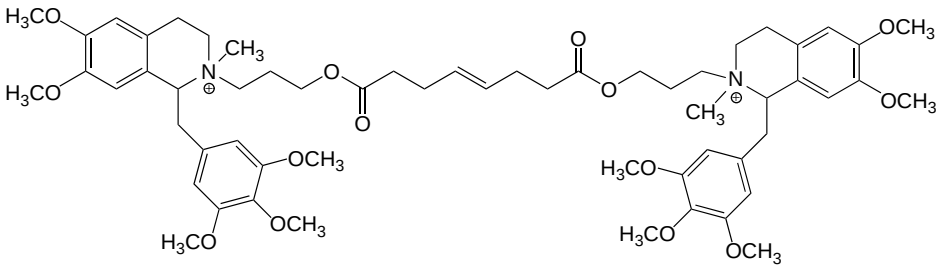

(**37**) Mivacurium

of atracurium, the latter being about 15 to 20 min.

4 Anticancer Drugs

4.1 Catharanthus (Vinca) Alkaloids

In 1949, Canadian researchers at the University of Western Ontario began investigating the medicinal properties of the rosy periwinkle (*Catharanthus roseus*), a plant that had been used for many years to treat diabetes mellitus in the West Indies. Despite finding that the plant extract when given orally had no effect on blood sugar levels in rats or rabbits, the researchers noted that when given intravenously, the extract caused the animals to succumb to bacterial infection and die. This curious observation prompted further studies, which showed that the plant extract reduced levels of white blood cells, causing granulocytopenia and bone marrow damage, toxic effects that are encountered with many antitumor drugs. These findings led the Canadian group to isolate an alkaloid fraction with potent cytotoxic activity. The active principle was eventually purified and became known as *vinblastine* (**38**), a dimeric indole-dihydroindole alkaloid.

Concurrently, researchers at the Lilly Research Laboratories had been investigating extracts of *C. roseus* and they too had detected cytotoxic activity, specifically against acute lymphocytic leukemia. The US group isolated several alkaloids, including vinblastine and another closely related alkaloid, vincristine (**39**).

Although many other alkaloids have been isolated from *C. roseus*, only vinblastine and vincristine have been developed for clinical use. The antiproliferative activity of the two compounds is related to their specific interaction with tubulin, thus preventing assembly of tubulin into microtubules and arresting cell division. However, despite this apparent identical mechanism of action and their clear chemical similarities, vinblastine and vincristine display very different clinical effects. Vinblastine, for example, is used to treat Hodgkin's disease and metastatic testicular tumors, whereas vincristine is used mainly in combination with other anticancer drugs for the treatment of acute lymphocytic leukemia in children. Toxicity profiles are also different, in that vinblastine causes bone marrow depression, whereas peripheral neuropathy often proves to be dose limiting in vincristine therapy.

Lilly introduced vinblastine and vincristine into the clinic in 1960 and 1963 respectively, but this did not preclude

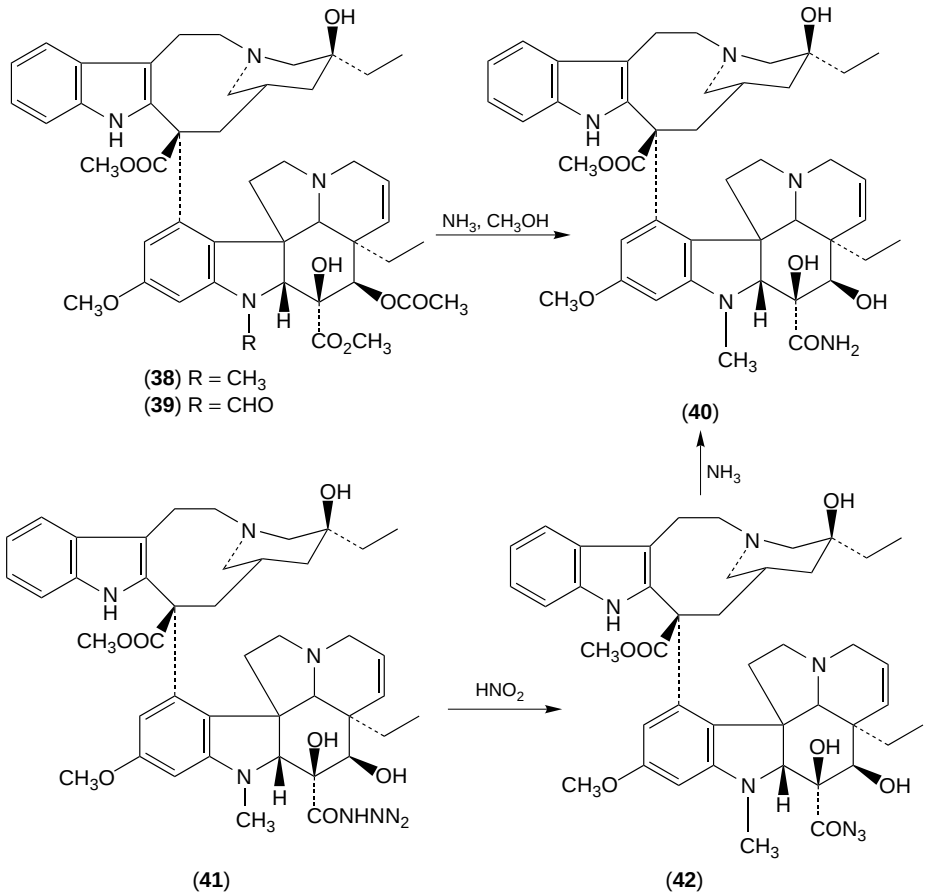

the search for improved derivatives. A chemical modification program aimed at improving antitumor activity and reducing toxicity was initiated in 1972. Concern about the neurotoxicity displayed by vincristine, its chemical instability, and low natural abundance (0.03 g kg^{-1} dried plant material) led to vinblastine being chosen as a template for semisynthetic modification. Selective ammonolysis of the ester function at C-3 and hydrolysis of the adjacent acetyl group yielded the desacetyl vinblastine amide, vindesine (**40**). Better yields of vindesine were obtained from the hydrazide (**41**) on treatment with nitrous acid and then reacting the resultant azide (**42**) with ammonia. The azide (**42**) proved to be a useful intermediate for the preparation of a range of substituted amides, although vindesine proved to be the derivative of choice, with significant differences in the spectrum of antitumor activity and toxicity compared to that of the naturally occurring alkaloids. Phase I clinical trials commenced in 1977 and vindesine has been used for the treatment of non–small cell lung cancer, lymphoblastic leukemia, and

non-Hodgkin's lymphomas. In combination with cisplatin, vindesine ranks among the foremost treatments for non–small cell lung cancer with respect to response rate and survival. Back in the 1950s, the US researchers could not have guessed that 30 years on, the demand for Catharanthus alkaloids would necessitate the processing of around 8000 kg of plant material per year!

4.2 Camptothecin

Camptothecin (**43**) was first isolated by Monroe Wall and Mansukh Wani in 1966, after ethanolic extracts of *Camptotheca acuminata*, a tree native to China, showing unusual and potent antitumor activity. Starting with 19 kg of dried wood and bark, Wall and Wani painstakingly purified the principal active component with a combination of hot solvent extraction, an 11-stage Craig countercurrent partition process, silica gel chromatography, and crystallization. Camptothecin was characterized as a novel pentacyclic alkaloid, present as just 0.01% w/w of the stem bark of *C. acuminata*. Of particular note was the unusual activity that camptothecin displayed in L1210 and P388 mouse leukemia life-prolongation assays. The compound also inhibited the growth of solid tumors *in vivo* and the water-soluble sodium salt was progressed to phase II clinical trials before being withdrawn because of severe bladder toxicity.

Interest in camptothecin gained new impetus in 1985, when it was discovered that the compound exerts its antitumor activity through a novel mechanism of action. Camptothecin binds to the covalent complex formed by topoisomerase I and DNA, which initiates DNA replication and thus, stabilizes the enzyme–DNA complex and prevents cell proliferation. The elucidation of the mechanism of action provided a means of evaluating camptothecin analogs as topoisomerase inhibitors *in vitro* and efforts then focused on synthesizing water-soluble analogs with broad-spectrum antitumor activities. The α-hydroxy lactone (ring E) and, in particular, the 20 (S)-form proved essential for maintaining biological activity, but the 10-hydroxy analog (**44**) showed greater activity than that of (**43**). Wall and Wani successfully deployed the Friedlander reaction between substituted 2-aminobenzaldehydes and the tricyclic intermediate (**45**), to synthesize a variety of ring-A–substituted analogs. These studies may have prompted SmithKline Beecham (now GlaxoSmithKline) to synthesize the water-soluble 10-hydroxycamptothecin analog topotecan (**46**) that was first approved in 1996 for the treatment of recurrent ovarian cancer and, two years later, for small cell lung cancer. Irinotecan (**47**), developed by Daiichi and Yakult Honsha in Japan and marketed by Pharmacia, was also approved in 1996 for the treatment of advanced colorectal cancer. Irinotecan is inactive as a topoisomerase I inhibitor, but acts as a prodrug of the active 7-ethyl-10-hydroxycamptothecin (**48**).

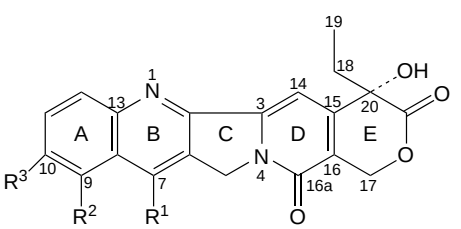

(**43**) Camptothecin: $R^1 = R^2 = R^3 = H$
(**44**) 10-Hydroxycamptothecin: $R^1 = R^2 = H$, $R^3 = OH$

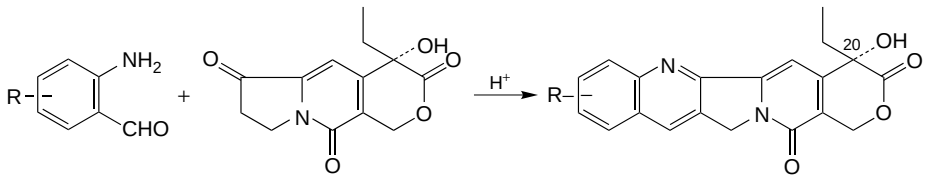

(**45**)

(**46**) Topotecan: R^1 = H, R^2 = CH_2—$N(CH_3)_2$, R^3 = OH

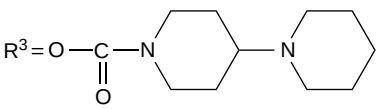

(**47**) Irinotecan: $R^1 = C_2H_5$, R^2 = H,

(**48**) 7-Ethyl-10-hydroxycamptothecin: $R^1 = C_2H_5$, R^2 = H, R^3 = OH

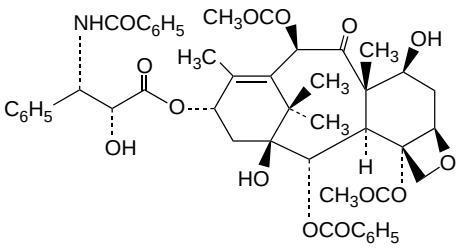

(**49**)

4.3 Paclitaxel and Docetaxel

Regarded as the tree of death by the Greeks and used to prepare arrow poison by the Celts, the yew tree has been associated with death and poisoning for centuries. The English yew, *Taxus baccata*, was used to make funeral wreaths and it was believed that one could die by merely standing beneath the boughs of the tree.

Yew certainly contains highly toxic metabolites and their potency and fast duration of action has often made extracts of yew the poison of choice for numerous murders and suicide attempts. It is thus, ironic that extracts from the Pacific yew, *Taxus brevifolia*, after being tested in the National Cancer Institute's (NCI) screening program during the 1960s, yielded one of the most exciting anticancer compounds discovered in recent years; that is, paclitaxel (**49**) (originally given the name taxol by Wall and Wani).

The initial isolation and characterization of paclitaxel proved particularly difficult because of (1) its very low natural abundance in *T. brevifolia* bark (although this was the best-known source, the isolated yield was only 0.02% w/w, equivalent to 650 mg per tree); (2) the poor analytical data obtained from the purified compound; and (3) the failure of paclitaxel to give crystals that were suitable for X-ray analysis. The structure of paclitaxel was published in 1971, but further biological testing continued to be troubled by difficulties. The compound showed only modest *in vivo* activity in various leukemia assays, which was no better than that displayed by a number of other new compounds at the time. In addition to the limited supplies of paclitaxel (the complexity of the molecule precluded chemical synthesis), the compound was very poorly soluble in water, which made formulation difficult. However, various new assays were developed in the 1970s, including the murine B16 melanoma model, in which paclitaxel showed very good activity and another boost came when Horwitz et al. discovered that the compound prevented cell division by a unique mode of action. In contrast to the antimitotic vinblastine

and podophyllotoxin analogs (qv), which prevent microtubule assembly, paclitaxel inhibits cell division by promoting assembly of stable microtubule bundles, which leads to cell death.

Phase I clinical trials were initiated in 1983, but these were to proceed at a slow and tortuous pace and proved all but disastrous when the high levels of oil-based adjuvant used to formulate paclitaxel caused severe allergic reactions in many volunteers. Undaunted by the formulation problem and spurred on by paclitaxel's novel mechanism of action, clinicians were eventually able to minimize the allergic events and demonstrate useful activity. Phase II clinical trials began in 1985 despite continuing supply problems and 4 years later the program received a significant boost when good responses were reported from patients suffering from refractory ovarian cancer, a disease that kills some 12 500 women a year in the United States alone.

In many ways, the development of paclitaxel mirrored that of the camptothecin analogs, both being dogged for many years by supply issues, poor pharmacokinetics, and toxicity, but the subsequent uncovering of novel mechanisms of action fueled renewed efforts to develop these leads into important new anticancer agents.

In 1991, Bristol-Myers Squibb in conjunction with the NCI agreed to manage the supplies of paclitaxel and they were granted a licence to further develop the compound. The following year the US Federal Drug Administration approved paclitaxel for the treatment of ovarian cancer in patients unresponsive to standard treatments and in December 1993 approval was given for the treatment of metastatic breast cancer.

The sourcing of paclitaxel from *T. brevifolia* was a major problem because to treat the groups of patients suffering ovarian cancer in the United States alone would require about 25 kg of compound per year, necessitating the felling of some 38 000 trees! Although the Pacific yew is not a rare tree, it is extremely slow growing and such harvesting could not be sustained indefinitely. It has been estimated that there were enough trees available to maintain a supply of paclitaxel for only 2 to 7 years. The isolation of paclitaxel from other Taxus species has been investigated at length and reasonable quantities have been obtained from the needles of several species including *T. baccata*. Using the needles has alleviated the supply problem because they can be harvested without damaging the tree. However, the needles contain much higher quantities of several biosynthetic precursors of paclitaxel and two of these, baccatin III (**50**) and 10-desacetylbaccatin III (**51**) have been used to prepare paclitaxel semisynthetically. One approach, developed by Potier et al., involved acylation of the sterically hindered C-13 position of baccatin III with cinnamic acid and subsequent double-bond functionalization through hydroxyamination, to give paclitaxel together with various regio- and stereoisomers. A better approach involved protection of 10-desacetylbaccatin III as the triethylsilyl ether, followed by direct acylation with the phenylisoserine derivative (**52**), giving paclitaxel in 38% overall

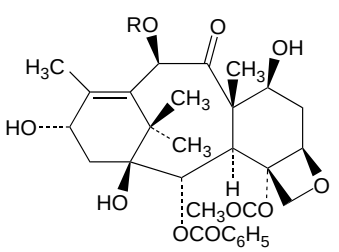

(**50**) R = $COCH_3$
(**51**) R = H

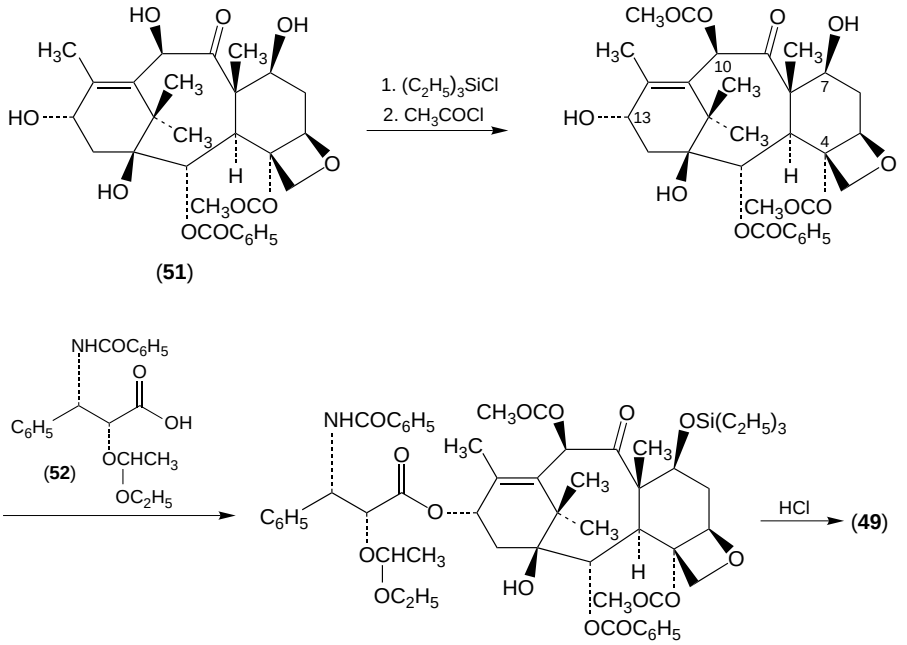

yield. Further improvements were made using less sterically demanding acylating reagents; for example, acylation with the β-lactam (**53**) gave paclitaxel in up to 90% yield and this may be the preferred method for commercial production.

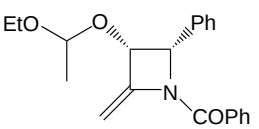

(**53**)

These semisynthetic approaches also provide access to analogs with potential advantages over paclitaxel itself. Structure-activity studies have shown that, although the oxetane ring appears to be essential for activity, wide variation in the nature and stereochemistry of the C-13 ester side chain can be tolerated. Thus, the *N-t*-(butoxycarbonyl) derivative, docetaxel (**54**), which appears to be more potent than paclitaxel and has better solubility characteristics, has been developed and launched by Sanofi-Aventis for the treatment of ovarian, breast, and lung cancers.

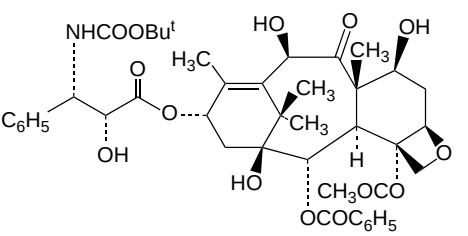

(**54**)

Various "protaxols," designed to release paclitaxel *in situ* under physiological conditions, have been prepared by acylating the C-2′ hydroxyl group. Nicolaou et al. reported the synthesis of the sulfone (**55**), which is soluble and stable in aqueous

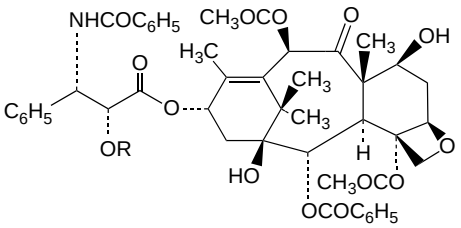

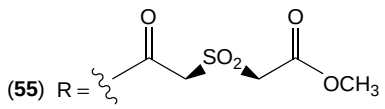

media, but is able to release paclitaxel rapidly in human blood plasma.

Plant tissue culture, microbial fermentation, and total synthesis provide other possibilities for the production of paclitaxel and its derivatives, although it is far from certain whether any of them will be commercially viable.

4.4 Epothilones

Epothilones A (**56**) and B (**57**), 16-membered macrocyclic polyketide lactones, were first isolated from the cellulose-degrading myxobacterium *Sorangium cellulosum* by Hoefle, Reichenbach, and coworkers as narrow-spectrum antifungal and cytotoxic metabolites. The compounds were then tested by the National Cancer Institute in the United States and found to be highly active against breast and colon cancer cell lines. Subsequently, Bollag et al. at the Merck Research Laboratories discovered that the epothilones stabilize microtubule assembly and thus, inhibit cell division by the same mechanism as that of paclitaxel (mentioned earlier). This observation, together with their less complex chemical structure, increased water solubility, more rapid action *in vitro*, and effectiveness against multidrug-resistant tumor cell lines, has prompted significant interest and a number of epothilones are currently undergoing clinical trials as anticancer agents.

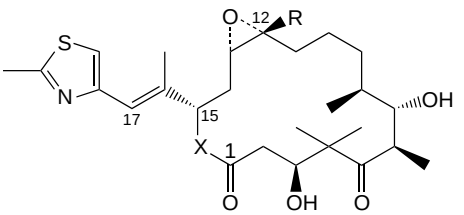

(**56**) Epothilone A: X = O, R = H
(**57**) Epothilone B: X = O, R = CH_3
(**59**) BMS-247550: X = NH, R = CH_3

On learning the absolute stereochemistry of (**56**) and (**57**), three academic research groups embarked on the total synthesis of the epothilones. Nicolaou, Danishefsky, and Schinzer independently adopted successful, elegant synthetic approaches involving olefin metathesis, macrolactonization, Suzuki coupling, or ester–enolate–aldehyde condensation. Within 3 years of the disclosure of their absolute stereochemistry, 17 different total syntheses of the natural products were reported. These syntheses paved the way for the generation of a large number of epothilone analogs for biological evaluation, including the use of solid-phase combinatorial approaches.

The academic groups focused on modifications around the core macrocyclic lactone, establishing important structure–activity relationships, but not improving on the *in vitro* biological activity of the most active natural product, epothilone B (**57**). *In vivo* biological data were comparatively scarce and, although one group reported that epothilones B (**57**) and D (**58**) showed activity in murine tumor models,

researchers at Bristol-Myers Squibb reported that epothilones D (**58**) lacks *in vivo* activity as a result of rapid metabolic inactivation. It was postulated that esterase-mediated hydrolysis of the macrocyclic lactone formed an inactive ring-opened species and, therefore, efforts were focused on replacing the lactone with a more stable macrocyclic lactam moiety. Several macrocyclic lactam derivatives were synthesized from (**57**) and (**58**). Of note was the preparation of BMS-247550 (**59**) in a three-step synthesis from epothilone B (**57**), utilizing a novel Pd(0)-catalyzed ring-opening reaction followed by reduction and macrolactamization. BMS-247550 (**59**), which is in phase III clinical trials, retains its activity against human cancer cells that are naturally insensitive to paclitaxel or that have developed resistance to paclitaxel, both *in vitro* and *in vivo*.

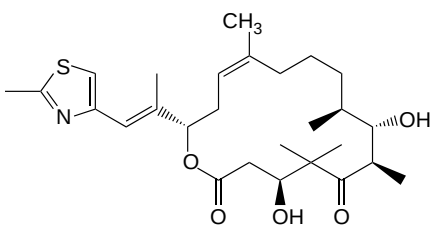

(**58**) Epothilone D

4.5 Podophyllotoxin, Etoposide, and Teniposide

The development of the natural constituents of podophyllum resin into effective semisynthetic and, ultimately, totally synthetic compounds for the treatment of various kinds of cancer provides one of the most sustained and intriguing stories of drug discovery. The story has all the classic ingredients, starting with observation and reasoning, extending through chance into new areas and characterized throughout by persistence and determination, particularly when biological activity had to be traced to very minor constituents in the crude plant extract.

Podophyllum peltatum (Mayapple, or American mandrake) and *Podophyllum emodi* are, respectively, American and Himalayan plants, widely separated geographically, but used in both places as cathartics in folk medicine. An alcoholic extract of the rhizome known as *podophyllin* was included in many pharmacopoeias for its gastrointestinal effects; it was included in the U.S.P., for example, from 1820 to 1942. At about this time the beneficial effect of podophyllin, applied topically to benign tumors known as *condylomata acuminata*, was demonstrated clinically. This usage was not inspirational, given that there are records of topical application in the treatment of cancer by the Penobscot Indians of Maine and, subsequently, by various medical practitioners in the United States from the nineteenth century. The crude resinous podophyllin is an irritant and unpleasant mixture unsuited to systemic administration.

The first chemical constituent was isolated from podophyllin in 1880 and named *podophyllotoxin*. A structure was proposed in 1932 and after some fine-tuning, was shown to be the lignan (**60**). As might be expected, the crude resin contains a variety of chemical types, including the flavonols, quercetin, and kaempferol. Although these other constituents undoubtedly have biological activity, it is the lignans that have received most attention and to which we shall devote the remainder of this section.

Chemists at Sandoz in the early 1950s reasoned that crude podophyllin might contain lignan glycosides with anticancer activity, which might be more water soluble and less toxic than podophyllotoxin.

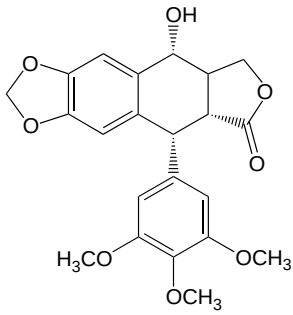

(**60**) Podophyllotoxin

The reasoning for the latter is not entirely clear, but in the event, they proved to be correct in both respects. Careful isolation gave podophyllotoxin β-D-glucopyranoside (**61**) its 4′-desmethyl analog (**62**), and some less important lignans lacking the B-ring hydroxy group. Unfortunately, the sugar derivatives were less active as inhibitors of cell proliferation than were the aglycones, as well as less toxic; however, as expected, they were much more water soluble. While continuing work to isolate more natural lignans, a substantial program of structural modification of the known compounds was undertaken with a view to protect the glucosides from hydrolytic enzymes and also to improve cellular uptake. Most of these changes were ineffective: the per-acylated derivatives, for example, were insoluble in water and had inferior cytostatic effects.

Condensation of the glucosides with a variety of aldehydes was more useful, in that not all the hydroxy groups were blocked. Despite this, water solubility was a problem with the podophyllotoxin derivatives (**63**). Gastrointestinal absorption was greatly improved, however, as was chemical stability and positive effects were observed in a few cancer patients with the benzylidene derivative (**64**). It was at this point that luck played a hand, backed up by a good deal of determination. A crude podophyllin fraction that was simpler and cheaper to prepare than pure podophyllin glucoside, was also treated with benzaldehyde to give a mixture of benzylidene derivatives, about 80% of which was compound (**64**). The crude product was found to be more potent than compound (**64**) alone and subsequently, to possess a different mode of action from that of the lead

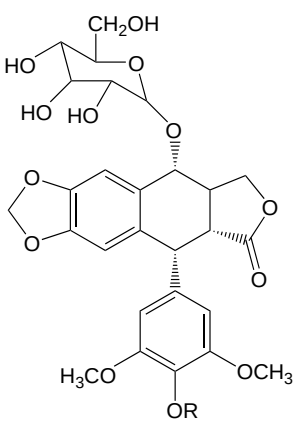

(**61**) R = CH_3
(**62**) R = H

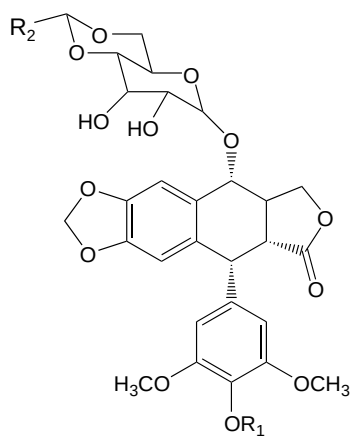

(**63**) R_1 = H, CH_3 R_2 = various alkyl, aryl
(**64**) R_1 = CH_3 R_2 = C_6H_5

compounds: rather than arresting cells in metaphase, cells were prevented from entering mitosis altogether. The crude mixture was marketed for cancer treatment as *Proresid*.

Improved biological assay methods indicated the presence of an unknown, highly active constituent of Proresid. For example, Proresid prolonged the life of mice inoculated with L1210 leukemia cells, an effect that was not observed with the known major constituent. In the early 1960s, chromatographic and spectroscopic techniques were not as highly developed as they are now and more than 2 years of work was required to isolate and identify the unknown component of the mixture, which proved to be the 4′-desmethoxy-1-epi analog (**65**) of the podophyllotoxin glucoside adduct. Present only in very small amounts in the derivatized extract, it was necessary to devise a synthesis from readily available materials. It was fortunate that the desired 1β configuration was readily secured from 1α-hydroxy-4′-desmethylpodophyllotoxin, itself obtained by selective demethylation of podophyllotoxin: the remainder of the synthesis would now be considered routine.

Given a large supply of the key intermediate (**66**), it was straightforward to prepare a number of aldehyde derivatives, resulting in analogs with up to a 1000-fold increase in potency. The selected adducts were those prepared from thiophen-2-aldehyde, giving teniposide (**67**) and from acetaldehyde, giving etoposide (**68**). Both drugs are of value, etoposide in the treatment of small cell lung cancer and testicular cancer, teniposide in the treatment of lymphomas and leukemias. The thiophene derivative is also of use in the treatment of brain tumors.

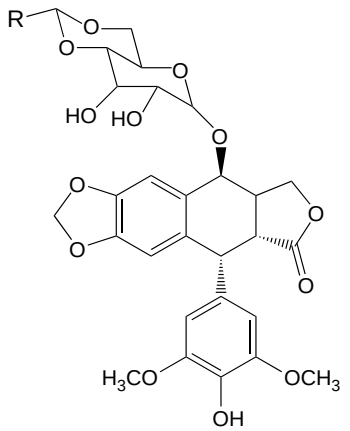

(**65**) R = C_6H_5
(**67**) R = 2-thienyl
(**68**) R = CH_3

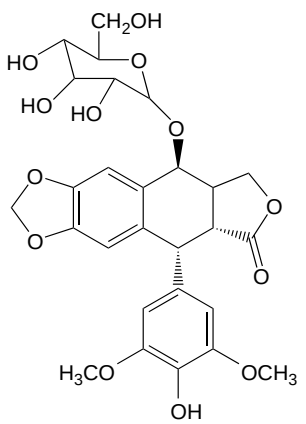

(**66**)

The natural products, podophyllotoxin and its congeners, are "spindle poisons" that inhibit cell proliferation by binding to tubulin and preventing formation of microtubules. Presumably, this effect is sufficient to account for the success of podophyllin in the treatment of condylomata acuminata, although the crude extract contains many other candidates for contribution to the biological activity. As

has been described, a very minor component of the natural mixture, missing the 4′-hydroxy group, having the 1β- instead of the 1α-hydroxy configuration and with this hydroxy group conjugated with β-D-glucose, must be treated with an aldehyde to produce the highly active and most important derivatives. These derivatives do not bind to tubulin, but have been shown to be inhibitors of topoisomerase II, which may account for most of the observed biological effects, including DNA strand breaks that lead to anticancer activity.

4.6 Marine Sources

Cytosine arabinoside (**69**), a synthetic analog of the C-nucleosides spongouridine (**70**) and spongothymidine (**71**) from the sea sponge *Cryptotheca cripta*, was the first and, so far, the only marine-derived compound used routinely as an anticancer agent. However, a number of chemically diverse natural products from marine sources have been progressed to clinical trials. Some of the most advanced compounds include trabectedin (ecteinascidin-743) (**72**), a tetrahydroisoquinoline alkaloid isolated from the mangrove ascidian *Ecteinascidia turbinata*, bryostatin-1 (**73**), a macrolide isolated from the bryozoan *Bugula neritina*, and dolastatin-10 (**74**), a linear peptide from the sea mollusk *Dolabella auricularia*. Trabectedin (**72**) was first isolated by Rinehart's group at the University of Illinois and has been progressed recently to phase III clinical trials by PharmaMar (Zeltia) for the treatment of ovarian cancer.

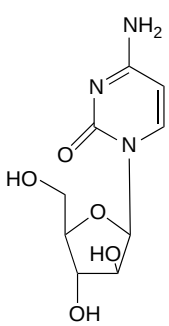

(**69**) Cytosine arabinoside

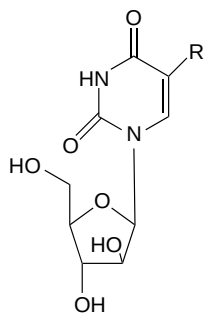

(**70**) Spongouridine: R = H
(**71**) Spongothymidine: R = CH_3

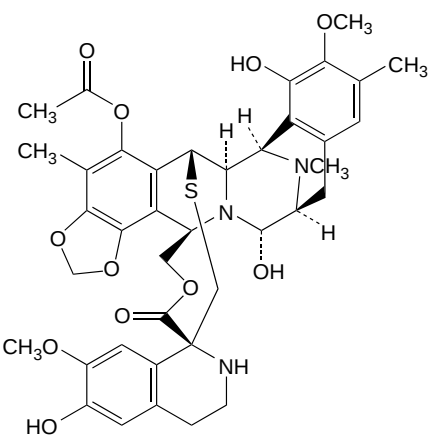

(**72**) Ecteinascidin-743

Yields of marine-derived natural products are invariably low and supply problems have delayed their development as useful pharmaceutical agents. For example, over 3000 kg of the sea squirt

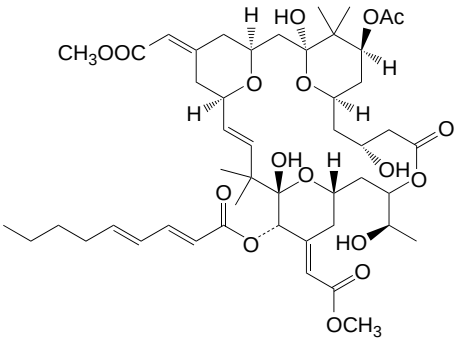

(**73**) Bryostatin-1

E. turbinata is required to produce 3 g of trabectedin (**72**), which is sufficient for just one cycle of treatment and 1000 kg of *B. neritina* yields 1.5 g of bryostatin-1 (**73**). Fortunately, supplies of trabectedin (**72**) have been met by semisynthesis and viable synthetic routes are now also available for bryostatin-1 (**73**) and dolastatin-10 (**74**).

5 Antibiotics

5.1 β-Lactams

In 1929, Alexander Fleming published the results of his chance finding that a penicillium mold caused lysis of staphylococcal colonies on an agar plate. He also showed that the culture filtrate, named *penicillin*, possessed activity against important pathogens including gram-positive bacteria and gram-negative cocci. However, it was not until 1940 that the true therapeutic efficacy of penicillin was revealed, when Chain et al. successfully tested the material in mice that had been previously infected with a lethal dose of streptococci. Several years later, the precise chemical structure of the main active component, benzylpenicillin (**75**), was determined and efforts to synthesize the compound were initiated. Benzylpenicillin proved to be an elusive target because of the instability of the β-lactam ring: it was unstable under acid conditions and was deactivated by β-lactamase enzymes produced by various gram-positive and gram-negative bacteria.

The discovery that the fused β-lactam nucleus, 6-aminopenicillanic acid (6-APA) (**76**), could be obtained from cultures of *Penicillium chrysogenum* led to the preparation of new, semisynthetic derivatives with improved stability to gastric acid and β-lactamases and with activity against a wider range of pathogenic organisms. Sheehan showed that compound (**76**) would react readily with acid chlorides to form new penicillin derivatives with novel substituents at the 6-position. Methicillin (**77**), with a sterically demanding 2,6-dimethoxybenzamide side chain, was the first semisynthetic penicillin to show resistance to staphylococcal β-lactamases, although the compound was

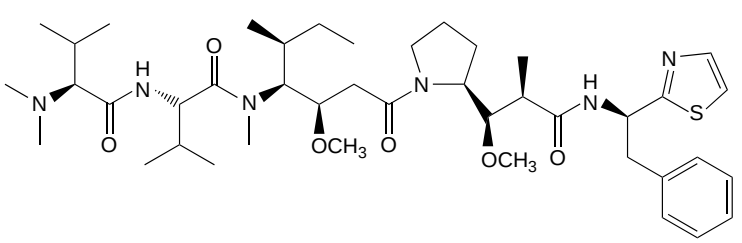

(**74**) Dolastatin-10

still acid labile. Ampicillin (**78**) has an α-aminophenylacetamido side chain and displays good activity against gram-negative organisms, it is stable to acid and thus can be administered orally, although it is susceptible to degradation by β-lactamases. Amoxycillin (**79**) differs from ampicillin by the addition of a single hydroxy group, but the compound is better absorbed by the gastrointestinal tract.

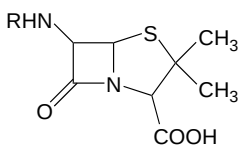

(**75**) R = $COCH_2Ph$
(**76**) R = H

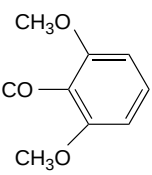

(**77**) R = CO

(**78**) R = COCHPh
NH_2

(**79**) R = COCH — OH
NH_2

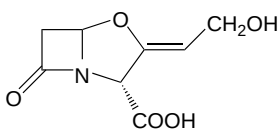

(**80**)

Clavulanic acid (**80**), isolated from *Streptomyces clavuligerus*, is similar in structure to the penicillins, except that oxygen replaces sulfur in the five-membered ring. Clavulanic acid has weak antibacterial activity, but is a potent inhibitor of β-lactamases. A mixture of clavulanic acid and the β-lactamase–sensitive amoxycillin was introduced in 1981 as Augmentin and has proved to be an effective combination to combat β-lactamase–producing bacteria. In 2001, 20 years after its launch, Augmentin was the best-selling antibacterial worldwide.

The clinical introduction of the penicillin group of antibiotics prompted an intensive search for novel antibiotic-producing organisms and Selman Waksman demonstrated the value of actinomycetes in this role, discovering the aminoglycoside streptomycin (**81**) from *Streptomyces griseus* in 1943. Pharmaceutical companies also embarked on large programs of screening soil samples for antibiotic-producing microorganisms. Chloramphenicol (**82**) was isolated from *Streptomyces venezuelae* in 1948 and other clinically important antibiotics followed: chlortetracycline (**83**), neomycin (**84**), oxytetracyclin (**85**), erythromycin (**86**), oleandomycin (**87**), kanamycin (**88**), and rifamycin (**89**).

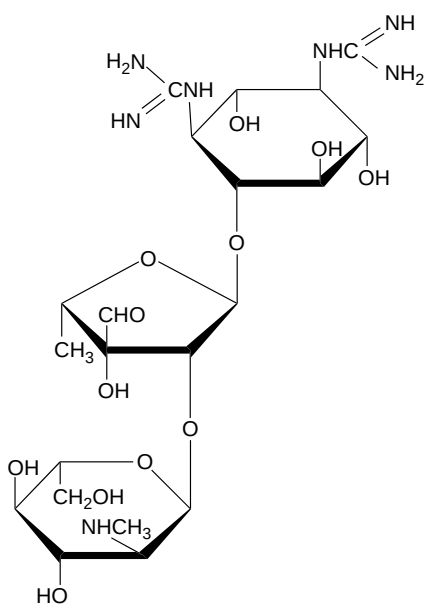

(**81**)

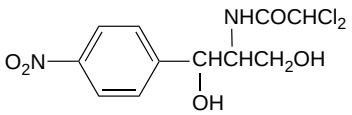

(**82**)

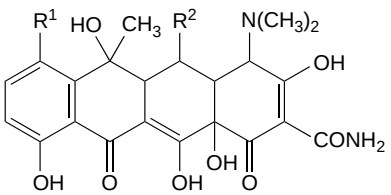

(**83**) R^1 = –Cl, R^2 = –H
(**85**) R^1 = –H, R^2 = –OH

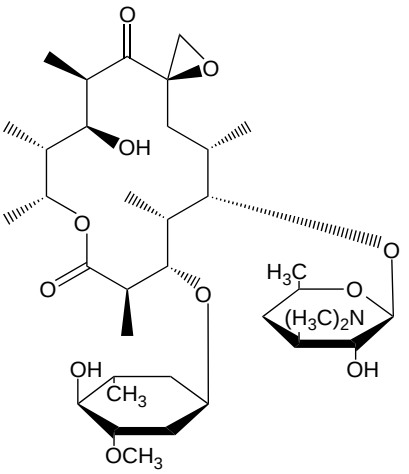

(**87**)

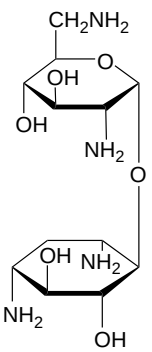

(**84**)

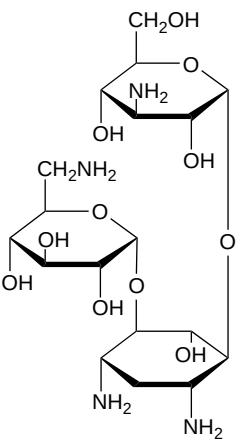

(**88**)

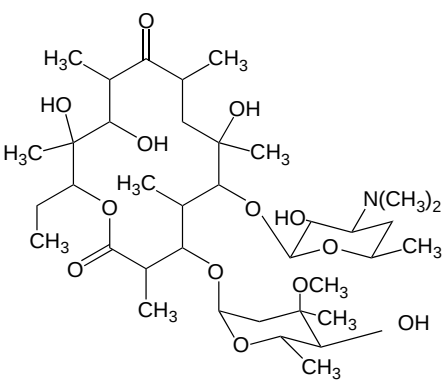

(**86**)

In 1948, Giuseppe Brotzu isolated the fungus *Cephalosporium acremonium* from a water sample collected off the coast of Sardinia. The culture showed significant antimicrobial activity, but Brotzu could not interest the Italian authorities in his discovery. He then turned to a friend in England for help, who arranged for Howard Florey at Oxford to receive a sample of the producing culture. Eventually,

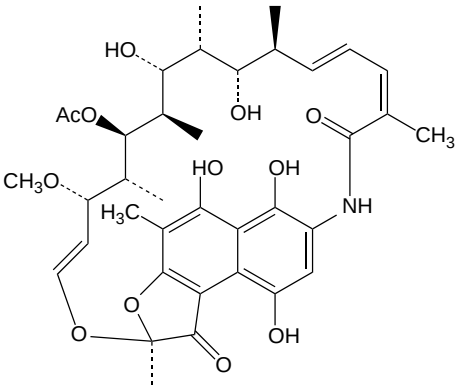

(**89**)

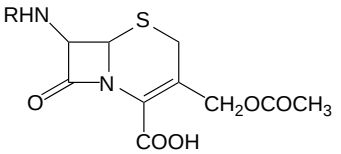

(**90**) R = $COCH_2CH_2CH_2CH(COOH)NH_2$

(**91**) R = H

(**92**) R = $COCH_2$ 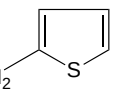

(**93**) R = $COCH_2CN$

(**94**) R = $COCH_2S$ 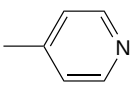

an antibacterial substance was isolated and named *cephalosporin C* (**90**). The compound, which had a structure similar to that of the penicillins, except that it had a dihydrothiazine ring fused to the β-lactam core, showed good resistance to β-lactamases and was less toxic than benzylpenicillin. However, plans to market the compound were terminated with the introduction of methicillin (mentioned earlier).

The discovery that the basic structural building block of cephalosporin C, that is, 7-aminocephalosporanic acid (7-ACA) (**91**), could be synthesized leading to the preparation of numerous cephalosporin derivatives in a similar way to the synthesis of penicillins from 6-aminopenicillanic acid. Modification of the substituent at the 7-position, while retaining the 3-acetoxymethyl group gave cephalothin (**92**), cephacetrile (**93**), and cephapirin (**94**), so-called first-generation cephalosporins with good activity against gram-positive bacteria, although the acetyl ester was susceptible to degradation by esterases and thus, limited the duration of action. Replacement of the acetoxy group by other substituents rendered the products less prone to esterase attack. For example, the pyridinium derivative, cephaloridine (**95**), has a longer duration of action than that of cephalothin.

The first orally active cephalosporin was cephaloglycin (**96**), which possessed a phenylglycine substituent in the C-7 side chain, although the labile 3-acetoxymethyl group was retained. Replacing the acetoxy group with a proton or chlorine, for example, cephalexin (**97**), cefadroxil (**98**), cephradine (**99**), and cefaclor (**100**), extended the duration of action of these orally active products. Cefaclor has been classified as a second-generation cephalosporin because it has a wider spectrum of activity, which includes gram-negative bacteria such as *Haemophilus influenzae*. Cephamandole (**101**) and cefuroxime (**102**) are parenterally administered cephalosporins with similar activities against clinically important gram-negative bacteria and are also resistant to many types of β-lactamases.

The newer third-generation cephalosporins, including ceftazidime (**103**), ceftizoxime (**104**), and ceftriaxone (**105**), which all contain an α-aminothiazolyl group in the C-7 side chain, have been developed for

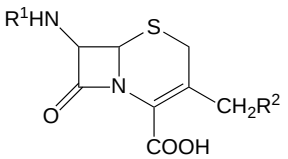

(**95**) R^1 = 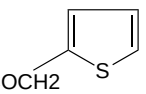

R^2 = 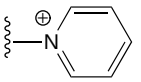

(**96**) NH_2 | COCHPh $R^2 = OCOCH_3$

(**97**) R^1 = NH_2 | COCHPh $R^2 = H$

(**98**) R^1 = NH_2 | COCH — 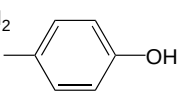 — OH $R^2 = H$

(**99**) R^1 = NH_2 | COCH — 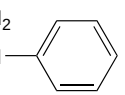 $R^2 = H$

treating specific pathogens such as *Pseudomonas aeruginosa*. Thienamycin (**106**), isolated from *Streptomyces cattleya* in 1976, represented a new class of β-lactam antibiotics produced by bacteria where the sulfur of the penicillin nucleus was replaced by a methylene group. An *N*-formylimidoyl derivative, imipenem (**107**),

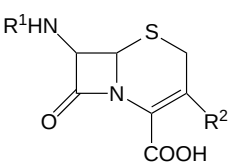

(**100**) R^1 = NH_2 | COCHPh $R^2 = Cl$

(**101**) R^1 = NH_2 | COCHPh R^2 = 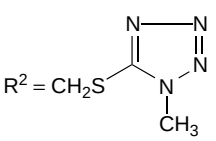

(**102**) R^1 = 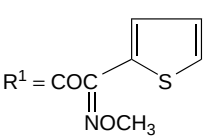

$R^2 = CH_2OCONH_2$

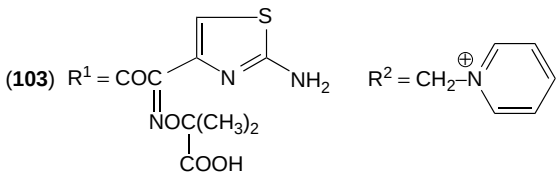

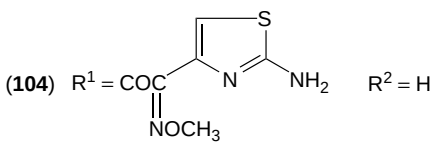

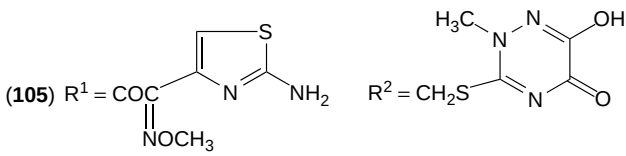

was the first example from this new class of carbapenem antibiotics to become available for clinical use. Imipenem has a very broad spectrum of activity against most gram-positive and gram-negative aerobic and anaerobic bacteria.

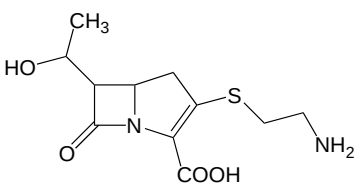

(**106**)

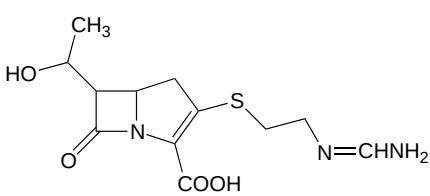

(**107**)

Screening bacteria such as *Pseudomonas acidophila* and *Chromobacterium violacium* for production of β-lactam antibiotics resulted in the discovery of naturally occurring monobactams, which had moderate antimicrobial activity. Side-chain variations, as developed for the penicillins and cephalosporins, led to compounds with improved activity against both gram-positive and gram-negative bacteria. A derivative containing the α-aminothiazolyl group, a side-chain component common to the third-generation cephalosporins (mentioned earlier), showed specific activity against gram-negative aerobic bacteria, including *Pseudomonas* spp. and was stable to most types of β-lactamases. The compound aztreonam (**108**) became the first

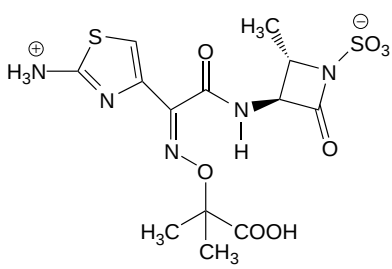

(**108**)

commercially available monobactam and showed a mode of action similar to that of the other β-lactam antibiotics by blocking bacterial cell wall synthesis.

5.2 Erythromycin Macrolides

Erythromycin (**109**) was isolated, in 1952, from a strain of *Saccharopolyspora erythraea* (formerly *Streptomyces erythreus*). As a broad-spectrum antibiotic, erythromycin has proved invaluable for the treatment of bacterial infections in patients with β-lactam hypersensitivity and is also the drug of choice in the treatment of infections caused by species of *Legionella, Mycoplasma, Campylobacter*, and *Bordetella*.

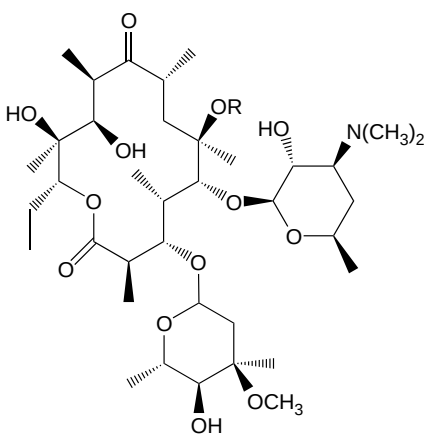

(**109**) Erythromycin A, R = H
(**114**) Clarithromycin, R = CH_3

Although safe and effective, erythromycin is not a perfect antibacterial. The presence of hydroxy groups suitably disposed with respect to the keto function at C-9 leads to the formation of a tautomeric mixture of hemiketals. The 6,9-hemiketal (**110**) may be dehydrated in stomach acid to give the inactive δ_8 analog (**111**), which can undergo further ring closure to give the 9,12-tetrahydrofuran (**112**) that is also inactive. The δ_8 derivative (**111**) may be responsible for some gastrointestinal disturbance. To avoid these problems by increasing the stability to acid, the 2′-stearate, estolate, and ethylsuccinate esters have been prepared, but even when the tablets are enteric-coated the bioavailability is erratic and relatively frequent dosing is required.

An understanding of the acid-catalyzed decomposition of erythromycin has led to a variety of semisynthetic derivatives with improved oral bioavailability. Reductive amination of the 9-keto function gives erythromycylamine, which reacts with (2-methoxyethoxy) acetaldehyde to give dithromycin. Beckmann rearrangement of the 9-oxime followed by reduction and methylation gives azithromycin (**113**), which shows good activity against gram-negative bacteria, including *H. influenzae*. An alternative for prevention of cyclization between the 9-keto and 6-hydroxy is to mask the 6-hydroxy group. If the 6-hydroxy is methylated, the result is clarithromycin (**114**), which like (**113**), has an improved pharmacokinetic profile compared with that of the parent molecule.

Both azithromycin and clarithromycin have been used for various bacterial infections for a number of years. Within the last decade, resistance has emerged to a range of antibacterials, including the macrolides, arising from methylation of an adenine in the 23S ribosomal RNA target site, which prevents binding. However, the ketolides (e.g. telithromycin (**115**)) overcome MLS_B resistance because they lack the L-cladinose moiety at position 3: the exposed hydroxyl is also oxidized to a ketone. The loss of potency, which would ensue, is compensated by two further modifications that improve binding; formation of

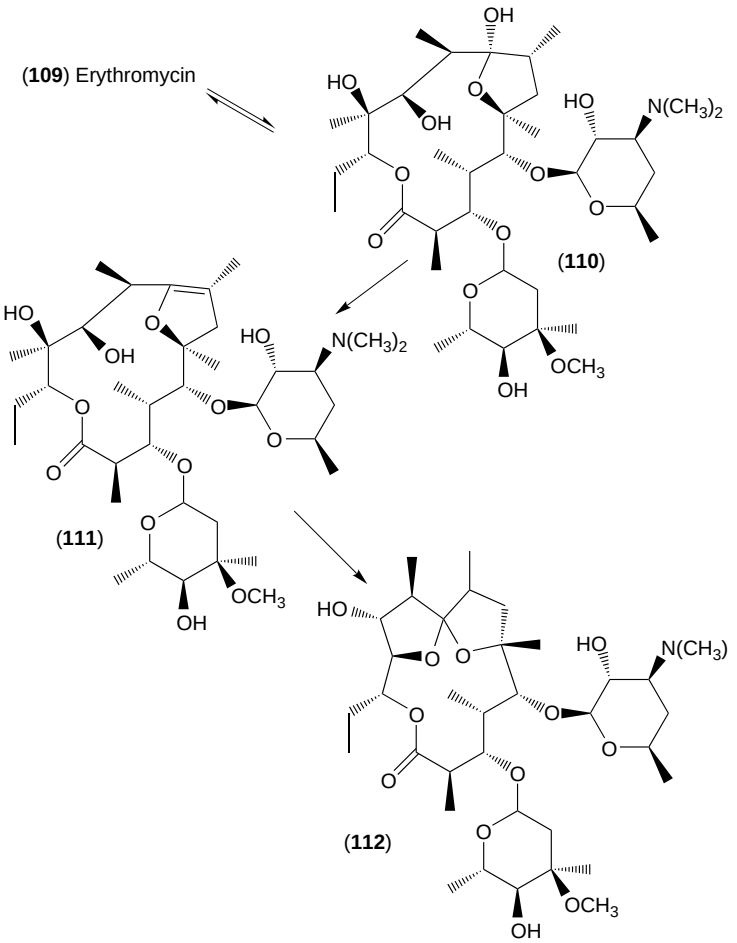

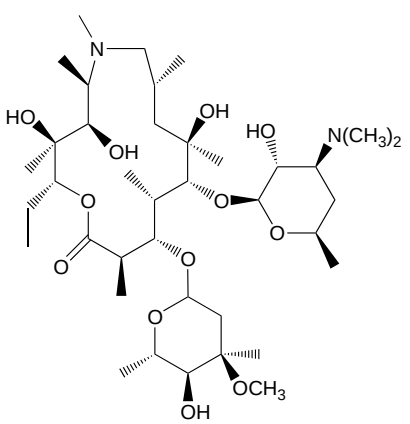

(113) Azithromycin

a carbamate at positions 11/12 and extension with a heterocycle-substituted side chain. In ABT 773 a similar side chain is placed at position 6 with comparable results.

5.3 Streptogramins

The streptogramins are produced by *Streptomyces* species and have been classified into two groups: Group A are polyunsaturated macrocyclic lactones and Group B are cyclic hexadepsipeptides. Both groups

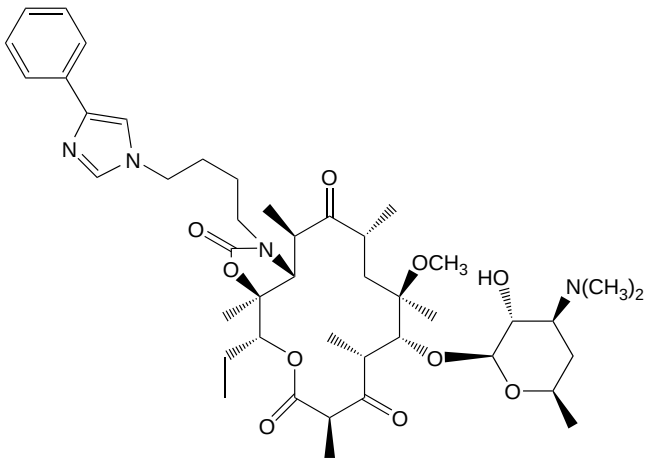

(**115**) Telithromycin

bind bacterial ribosomes, inhibit protein synthesis at the elongation step, and act synergistically against many gram-positive microorganisms. However, the naturally occurring streptogramins are poorly soluble in water and this, until recently, has limited their use for treating bacterial infections. New, water-soluble derivatives have been developed and the semisynthetic dalfopristin (**116**) and quinupristin (**117**) mixture (Synercid) has been approved for the treatment of gram-positive infections, including multidrug-resistant strains of *Enterococcus faecium*, *Staphylococcus aureus*, and *Streptococcus pneumoniae*.

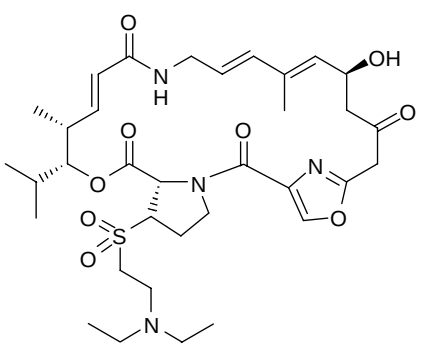

(**116**) Dalfopristin

5.4 Echinocandins

The fungal metabolite echinocandin B (**118**) is one of the lipopeptides in which a cyclic hexapeptide is combined with a long-chain fatty acid. Echinocandin B inhibits β-1,3-glucan synthesis and as a result has anti-Candida and anti-*Pneumocystis carinii* activity. As a group, the echinocandins are not orally bioavailable, are hemolytic, and are not very water soluble, despite the hydrogen-bonding ability of the polyhydroxylated hexapeptide.

Synthesis of the cyclic hexapeptide is unattractive for the purpose of securing analogs with improved biological activity because of the unusual nature of the amino acids used and the complex stereochemistry generated by the high degree of hydroxylation. However, echinocandin B can be produced efficiently by fermentation of *Aspergillus nidulans*

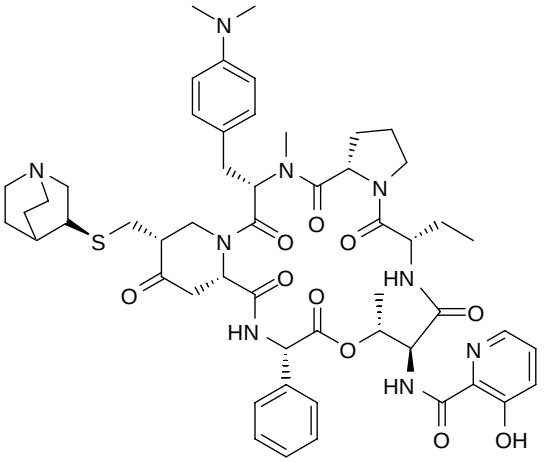

(**117**) Quinupristin

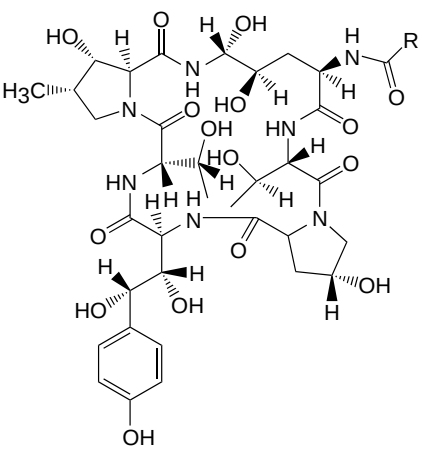

(**118**) Echinocandin B, R = linoleyl

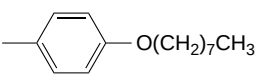

(**119**) Cilofungin, R = $O(CH_2)_7CH_3$

and then deacylated by fermentation with *Actinoplanes utahensis*. The free amino group thus exposed can be derivatized with a number of active esters. Synthesis of the amide from 4-octylbenzoic acid gives cilofungin (**119**), which has specifically high potency against *Candida albicans* and some other Candida species.

For systemic use, cilofungin had to be given intravenously and unfortunately, ran into problems associated with the cosolvent polyethylene glycol. A better derivative, anidulafungin (LY-303366) (**120**) has the major advantage of oral bioavailability and has been progressed to phase III clinical trials. Many other antifungal peptides are under investigation and two members of this series have now been marketed, caspofungin (MK-991, L-743,872) (**121**) and micafungin (FK463), for the treatment of aspergillosis and candidiasis.

6 Cardiovascular Drugs

6.1 Lovastatin, Simvastatin, and Pravastatin

One of the most significant natural product discoveries in the last 30 years has been a fungal secondary metabolite called *lovastatin* (**122**). Heralded as a major

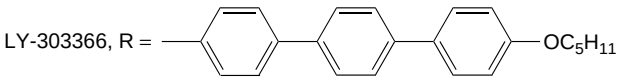

(**120**)

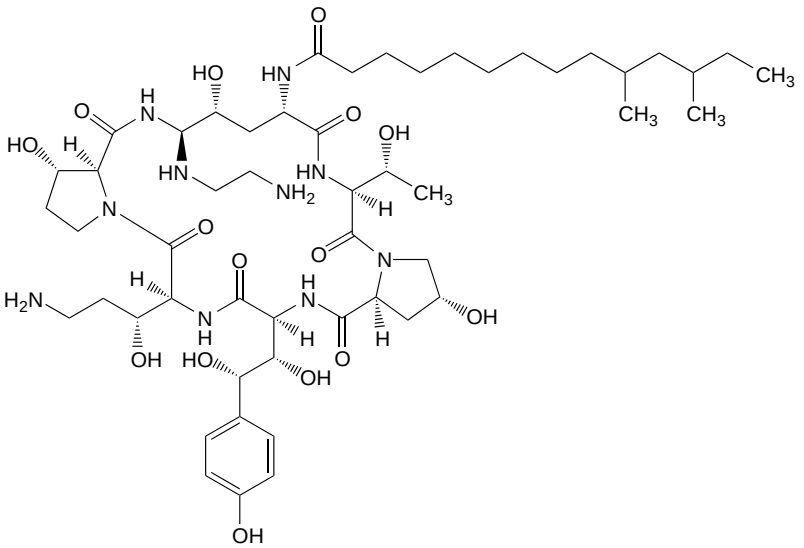

(**121**) Caspofungin

breakthrough in the treatment of coronary heart disease, lovastatin was introduced onto the market by Merck in 1987 for the treatment of hypercholesterolemia, a condition marked by elevated levels of cholesterol in the blood.

Lovastatin works by inhibiting 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, a key rate-limiting enzyme in the cholesterol biosynthetic pathway. However, the first specific inhibitors of this enzyme were discovered several years earlier by Endo et al. at Sankyo. The compounds, which are structurally related to lovastatin, were isolated from *Penicillium citrinum* and shown to block cholesterol synthesis in rats and lower cholesterol levels in the blood. Development of the most active compound, designated ML-236B (**123**), is believed to have been curtailed because of toxicity problems.

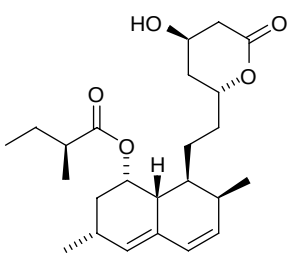

(**122**)

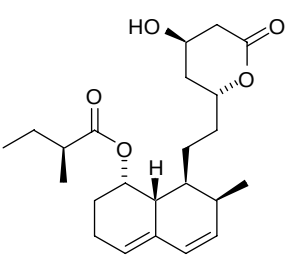

(**123**)

Brown et al. at Beechams also reported the isolation of (**123**), but as a metabolite from *Penicillium brevicompactum*. The group, naming the compound compactin, reported its antifungal activity but failed to reveal its mode of action as an inhibitor of HMG-CoA reductase. The search for naturally occurring inhibitors of HMG-CoA reductase gained pace and after spending several years developing appropriate screens, Merck found during only the second week of testing, a culture of *Aspergillus terreus* that displayed interesting inhibitory activity. In February 1979, the active component, lovastatin (mevinolin), was isolated and characterized, and in November the following year Merck was granted patent protection in the United States. Although lovastatin proved to be identical to monocolin K, a metabolite isolated earlier from *Monascus ruber*, the chemical structure of the latter compound had not been reported, whereas Merck filed for patent protection giving complete structural details for lovastatin.

The discovery of compactin and lovastatin prompted efforts to develop derivatives with improved biological properties. Modification of the methylbutyryl side chain of lovastatin led to a series of new ester derivatives with varying potency and, in particular, introduction of an additional methyl group α to the carbonyl gave a compound with 2.5 times the intrinsic enzyme activity of lovastatin. The new derivative, named *simvastatin* (**124**), was the second HMG-CoA reductase inhibitor to be marketed by Merck. Both lovastatin and simvastatin are prodrugs and are hydrolyzed to their active open-chain dihydroxy acid forms in the liver. A third compound, pravastatin (**125**), launched by Sankyo and Squibb in 1989, is the open hydroxyacid form of compactin that was first identified as a urinary metabolite in dogs. Pravastatin is produced by microbial biotransformation of compactin.

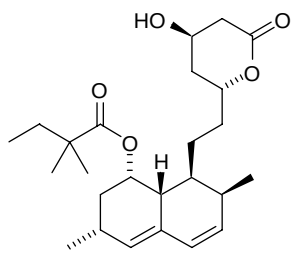

(**124**)

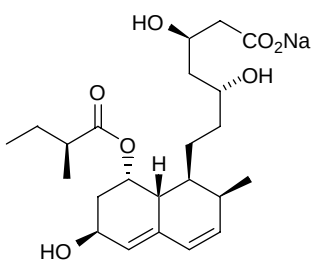

(**125**)

The HMG-CoA reductase inhibitors described earlier bind to two active sites on the enzyme: the hydroxymethylglutaryl binding domain and an adjacent hydrophobic pocket to which the decalin moiety binds. The recognition that the ring-opened hydroxy acids resemble mevalonic acid and that the decalin moiety could be replaced by 4-fluorophenyl–substituted heterocycles led to the launch of several new products including fluvastatin (**126**), the ill-fated cerivastatin (**127**) and the so-called turbostatin atorvastatin (**128**). Although cerivastatin was withdrawn from the market in 2001 because of fatal adverse drug–drug interactions, the "statins" remain one of the fastest growing

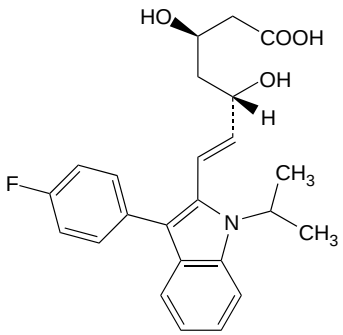

(**126**) Fluvastatin

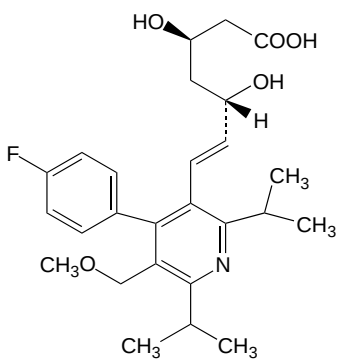

(**127**) Cerivastatin

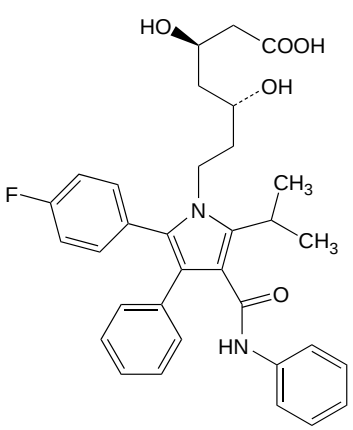

(**128**) Atorvastatin

segments of the pharmaceutical industry. One of the latest members of this group of cholesterol-lowering drugs to be marketed is AstraZeneca's rosuvastatin (**129**).

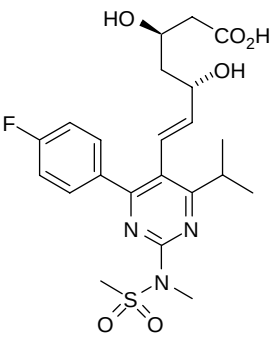

(**129**) Rosuvastatin

6.2 Teprotide and Captopril

While studying the physiological effects of snake poisoning, Ferreira discovered that specific components in the venom of the pit viper *Bothrops jararaca* inhibited degradation of the peptide bradykinin and potentiated its hypotensive action. The "potentiating factors" proved to be a family of peptides that worked by inhibiting the dipeptidyl carboxypeptidase, angiotensin-converting enzyme (ACE). In addition to catalyzing the degradation of bradykinin, ACE also catalyzes the conversion of human prohormone, angiotensin 1, to the potent vasoconstrictor octapeptide, angiotensin II. However, the significance of ACE in the pathogenesis of hypertension was not fully appreciated until the 1970s after Ondetti et al. had first isolated and then synthesized the naturally occurring nonapeptide, teprotide (**130**). The compound proved to be a specific potent inhibitor of ACE and showed excellent antihypertensive properties in clinical trials, although

its use was limited by the lack of oral activity.

Pyr–Trp–Pro–Arg–Pro–Gln–Ile–Pro–Pro
(**130**)

The discovery of teprotide led to a search for new, specific, orally active ACE inhibitors. Ondetti et al. proposed a hypothetical model of the active site of ACE, based on analogy with pancreatic carboxypeptidase A and used it to predict and design compounds that would occupy the carboxy-terminal binding site of the enzyme. Carboxyalkanoyl and mercaptoalkanoyl derivatives of proline were found to act as potent, specific inhibitors of ACE and 2-D-methyl-3-mercaptopropanoyl-L-proline (**131**) (captopril) was developed and launched in 1981 as an orally active treatment for patients with severe or advanced hypertension. Captopril, modeled on the biologically active peptides found in the venom of the pit viper, made an important contribution to the understanding of hypertension and paved the way for other ACE inhibitors, such as enalapril (**132**) and lisinopril, which have had a major impact on the treatment of cardiovascular disease.

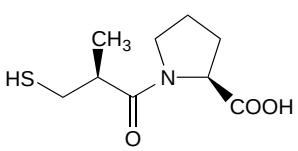

(**131**)

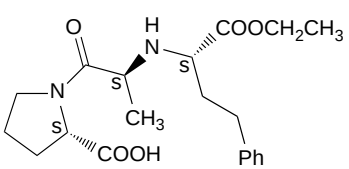

(**132**)

6.3 Adrenaline, Propranolol, and Atenolol

The true clinical potential of β-adrenoceptor blocking agents for treating angina, atrial fibrillation, and tachycardias was first recognized by James Black and colleagues at ICI. Black noted a report from Neil Moran of Emory University in 1958, showing that dichloroisoprenaline antagonized the effects of adrenaline on heart rate and muscle tension. The first effective β-adrenoceptor blocker, pronethalol (**133**), was synthesized 2 years later by the ICI group and marketed for limited use in 1963. Toxicity problems soon led pronethalol to be replaced by the 1-naphthyl analog, propranolol (**134**), which became the first β-adrenoceptor antagonist approved for general use, being more potent and yet devoid of the partial agonist or intrinsic sympathomimetic activity shown by many other analogs. Compounds with improved selectivity for the β-adrenoceptor of cardiac muscle (β-1-adrenoceptor blockers) were to follow, including atenolol (**135**), which became the most frequently prescribed beta-blocker and one of the best-selling drugs of the time.

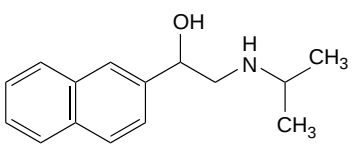

(**133**) Pronethalol

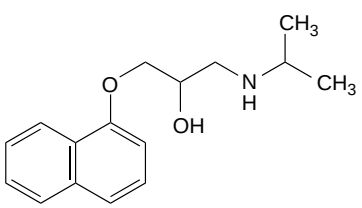

(**134**) Propranolol

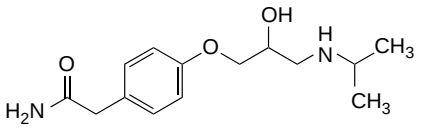

(**135**) Atenolol

6.4 Dicoumarol and Warfarin

Sweet clover has a long history of medicinal use, often as an antiflammatory or analgesic preparation in the form of ointments and poultices. *Melilotus officinalis* (yellow sweet clover, or ribbed melilot) was reputed to have been a favorite herbal treatment used by King Henry VIII of England and the plant is still referred to as *King's Clover* in some publications.

The plant flourishes in poor soil and was cultivated extensively in Europe for cattle fodder and for soil improvement. In the early 1920s, *M. officinalis* was planted on the prairies of North Dakota and Alberta, Canada, but with disastrous consequences. Soon cattle and sheep throughout these regions began literally bleeding to death. The mysterious hemorrhagic disease was traced to clover fodder that had not been stored properly and had become "spoiled," or moldy. However, the insolubility of the anticoagulant component and the difficulty of assaying extracts for biological activity made the task of isolating the active principal component intractable. It took almost 20 years before the compound was identified as 3,3′-methylenebis(4-hydroxycoumarin) (**136**), an oxidative degradation metabolite of coumarin (**137**), itself a common component of *Melilotus* sp. Soon after the compound had been identified, trials were initiated that confirmed the oral anticoagulant activity in humans and in 1942, it was marketed under the name dicoumarol. The compound had a slow, erratic onset of action and efforts were initiated to prepare synthetic analogs that acted faster and had longer duration of action. A 4-hydroxycoumarin residue, substituted at the 3-position, proved essential for biological activity and in 1948, after synthesizing over 150 compounds, a 4-hydroxycoumarin derivative that was longer acting and more potent than dicoumarol was selected not for clinical use, but as a rodenticide for development by the Wisconsin Alumni Research Foundation! The compound (**138**), named *warfarin* (an acronym derived from the name of the institute coupled with "arin" from coumarin), became a household name for rat poison. Concern over the use of oral anticoagulants and the inherent risk of hemorrhage inhibited the development of warfarin as a therapeutic agent. However, in 1951, a US Army cadet unsuccessfully attempted to commit suicide by taking massive doses of the compound. The incident prompted further clinical trials that resulted in warfarin being used as the anticoagulant of choice for prevention of thromboembolic disease.

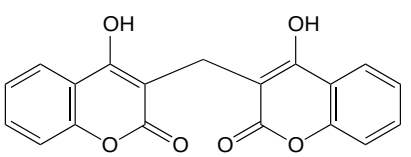

(**136**)

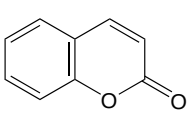

(**137**)

The mode of action of the coumarin anticoagulants involves blocking the regeneration of reduced vitamin K and induces a state of functional vitamin K

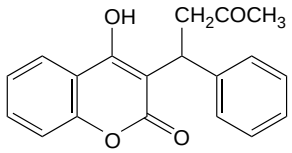

(**138**)

deficiency, thus interfering with the blood-clotting mechanism.

7 Antiasthma Drugs

7.1 Khellin and Sodium Cromoglycate

The toothpick plant, *Ammi visnaga*, had been used for centuries in Egypt as an antispasmodic agent to treat renal colic and ureteral spasm. In 1879, one of the plant's main constituents was isolated, crystallized, and named *khellin* (**139**). Subsequently, the pure compound was shown to relax smooth muscle and in 1938, the chemical structure was characterized as a chromone derivative. In 1945, a medical technician took khellin to treat renal colic and found instead that it acted as a potent coronary vasodilator and relieved his angina. This chance discovery, together with earlier observations, led to khellin being used as a coronary artery vasodilator and for treating bronchial asthma. However, its clinical use was severely limited by some unpleasant gastrointestinal side effects.

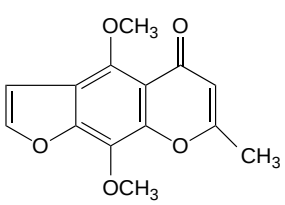

(**139**)

Five years later, a small British pharmaceutical company, called *Benger Laboratories*, initiated a program to synthesize khellin analogs as potential bronchodilators for treating asthma and had prepared a series of compounds that relaxed guinea pig bronchial smooth muscle and protected the animals against allergen-induced bronchospasm.

A clinical pharmacologist on Benger's staff, who suffered from chronic asthma, questioned the validity of the animal model and decided instead to test the compounds on himself. He then prepared a "soup" of guinea pig fur, inhaled the vapors to induce a reproducible asthma attack, and assessed the effects of the synthesized khellin derivatives. Many of the compounds first prepared were insoluble in water and caused nausea and other unpleasant side effects when taken orally. This led to the test compounds being formulated as aerosol sprays and in 1958, an aerosol preparation of a chromone-2-carboxylic acid derivative (**140**) was found to exert a protectant effect, albeit short lived, against bronchial allergen challenge without showing the bronchodilator activity seen with other compounds. The compound was completely inactive in the guinea pig asthma model and afforded its protectant effect in humans only when inhaled as an aerosol.

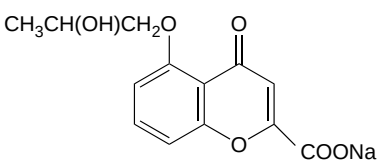

(**140**)

About two new compounds were tested each week and in 1965, after synthesizing

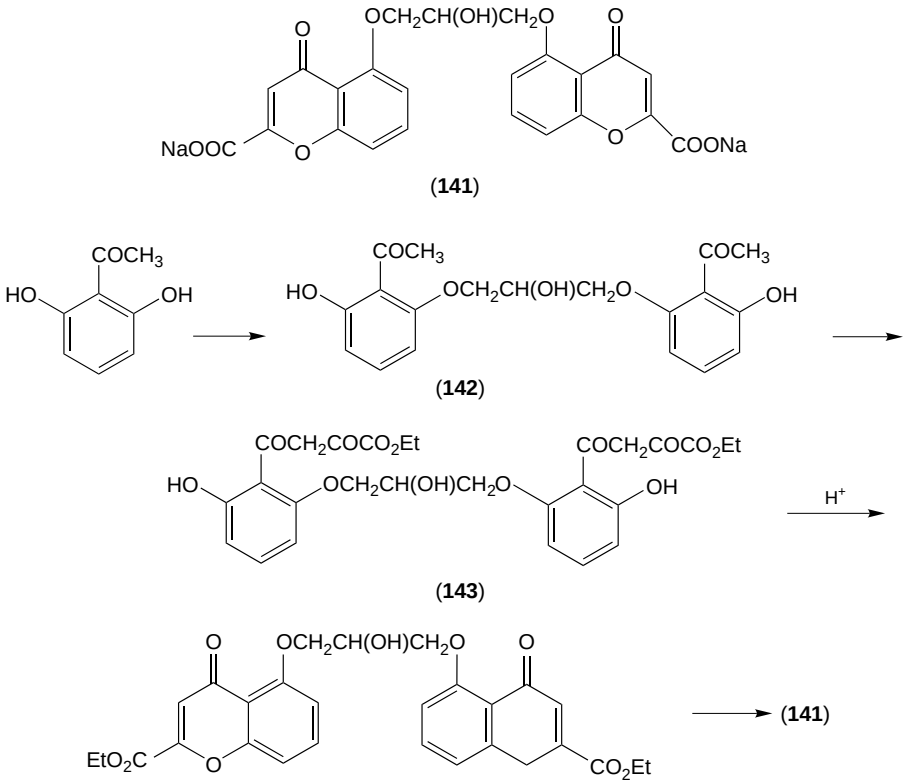

some 670 analogs, a bischromone was prepared that gave good protection, even when inhaled up to 6 h before bronchial allergen challenge. The compound sodium cromoglycate (**141**) was obtained by condensing diethyl oxalate with the bis(hydroxy acetophenone) (**142**) and cyclizing the resultant bis(2,4-dioxobutyric acid) ester (**143**) under acidic conditions. The essential chemical features required for activity appeared to be the coplanarity of the chromone nuclei, the flexible dioxyalkyl link and the carboxyl groups in the 2-positions. It is believed to act by stabilizing tissue mast cells against degranulation, thereby preventing release of inflammatory mediators.

Sodium cromoglycate entered clinical trials in 1967 and emerged to become a first-line prophylactic treatment for bronchial asthma.

The coronary dilator properties of khellin have not been ignored and at least one successful program was initiated to prepare analogs for testing as potential antiangina drugs. Benziodarone (**144**) was the first useful compound to emerge from the Labaz laboratories in Belgium based on the benzofuran ring system. However, the compound caused

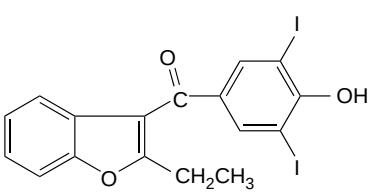

(**144**) Benziodarone

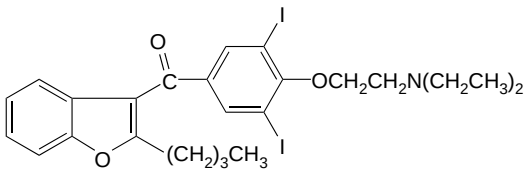

(**145**) Amiodarone

hepatotoxicity in man and was soon superseded by amiodarone (**145**), a more potent coronary dilator for treating angina. In 1970, the first report of antiarrhythmic activity in the clinic was published and amiodarone became established for prophylactic control of supraventricular and ventricular arrhythmias during the 1980s.

7.2 Ephedrine, Isoprenaline, and Salbutamol

The Chinese have been using a plant extract known as *ma huang* to treat asthma and hay fever for thousands of years. The extract is prepared from several species of *Ephedra*, a small leafless shrub found in China. Following experiments at the Peking Union Medical College and then at the University of Pennsylvania and the Mayo Clinic in the United States, the active ingredient, ephedrine (**146**), was introduced into Western medicine in 1926 as an orally active bronchodilator for the treatment of acute asthma.

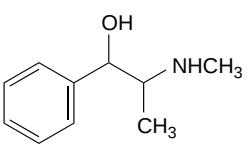

(**146**)

Ephedrine is related to another natural product that has been used to treat asthma, that is, the adrenal hormone adrenaline (**147**) (epinephrine). Adrenaline is a potent agonist of both α- and β-adrenoceptors and thus produces arterial hypertension as an undesirable side effect. In 1951, a synthetic alternative, isoprenaline (**148**), was

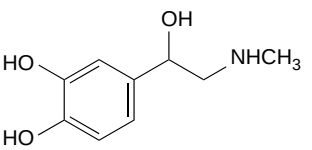

(**147**)

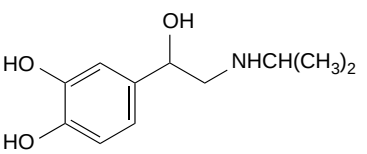

(**148**)

introduced and for almost 20 years it was considered the drug of choice for treating bronchospasm associated with acute asthmatic attack. Isoprenaline is a specific β-adrenoceptor agonist and, although it has no vasoconstrictor activity, the compound does have marked cardiac stimulant properties and a short duration of action. Ahlquist's concept of two types of adrenoceptor was developed further by Lands et al., who established the existence of β_1- and β_2-adrenoceptor subtypes. Clear structure–activity relationships emerged with the preparation of compounds related to adrenaline and ephedrine; the basic requirement for β-adrenoceptor agonist

activity was an aromatic ring substituted by an ethanolamine side chain. The branched methyl substituent on the side chain was associated with prolonged duration of action (i.e. ephedrine), whereas aromatic hydroxylation (in isoprenaline) prevented penetration across the blood–brain barrier and thus prevented stimulation of the CNS. However, 1,2-dihydroxy substituents were found to promote enzymic degradation, and replacement of the 3-hydroxy group by a hydroxymethyl substituent was required to extend the duration of action. In 1969, salbutamol (**149**) was launched by Glaxo as a longer-lasting, selective β_2-adrenoceptor agonist for the treatment of bronchial asthma and, subsequently, a lipophilic ether analog, salmeterol (**150**), was introduced with an even longer duration of action that has potential advantage in the prevention of nocturnal asthma.

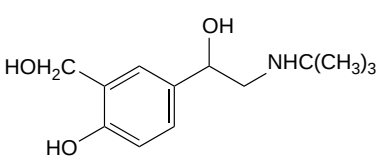

(**149**)

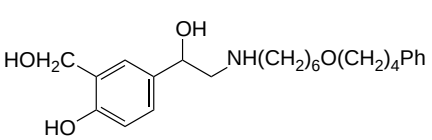

(**150**)

Despite the many chemical alterations that have been carried out on the phenylethanolamine "template," the key chemical features associated with modern β-agonists can be seen to have originated from the naturally occurring compounds, adrenaline and ephedrine.

7.3 Contignasterol

The use of inhaled corticosteroids such as fluticasone propionate to treat asthma and rhinitis has been well documented and will not be repeated here. Less well known is an unusual, highly oxygenated marine-derived steroid isolated from the sponge *Petrosia contignata* that possesses a unique cyclic hemiacyl side chain (**151**). The compound was isolated by Andersen and coworkers at the University of British Columbia and found to possess anti-inflammatory properties *in vivo*. Contignasterol was being developed by Inflazyme, in collaboration with Aventis, for the treatment of asthma and other inflammatory diseases and progressed to phase II clinical trials.

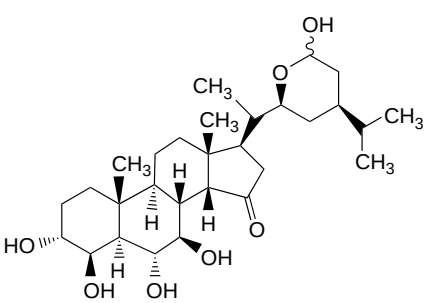

(**151**) Contignasterol

8 Antiparasitic Drugs

8.1 Artemisinin, Artemether, and Arteether

Artemisia annua (sweet wormwood, Qing hao) has been used in Chinese medicine for well over 1000 years. The earliest recommendation is for the treatment of hemorrhoids, but there is a written record of use in fevers dated 340 A.D. Modern

development dates from the isolation of a highly active antimalarial, artemisinin (qinghaosu), in 1972, and has been carried out almost entirely in China. Much of the original literature is therefore, in Chinese.

Artemisinin (**152**) is a sesquiterpene lactone with an unusual peroxide bridge. One of the earliest modifications involved catalytic reduction of the peroxide, resulting in loss of one oxygen and total loss of antimalarial activity in the adduct (**153**). The role of the peroxide bridge in producing antimalarial effects was not fully understood, but it appeared essential for activity, so much of the early work on analogs conserved this structural feature as an empirical finding. The mechanism of action of artemisinin has since been elucidated, although it is not without controversy. The drug has a high affinity for hemozoin, a storage form of hemin that is retained by the parasite after digestion of hemoglobin, leading to a highly selective accumulation of the drug by the parasite.

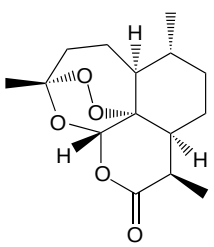

(**152**) Artemisinin

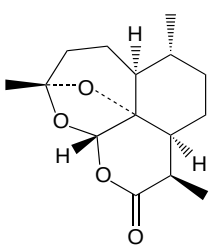

(**153**)

Artemisinin then decomposes in the presence of iron, probably from the hemozoin and releases free radicals, which kill the parasite. The peroxide bridge is, therefore, a crucial part of the drug molecule, as was suspected from structure–activity studies. Elucidation of the mechanism of action has led to the synthesis of a range of simple analogs capable of iron-catalyzed decomposition, some of which have good antimalarial activity.

In retrospect, it is not surprising that the peroxide-bridged compound (**154**), isolated from *Artabotrys uncinatus*, also has antimalarial activity. Because peroxides of this kind are likely to be formed from a variety of precursors in dried plant material (see the following), there may well be many more antimalarials of this kind to be found.

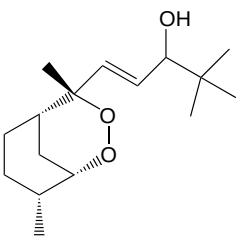

(**154**)

Artemisinin is an excellent antimalarial, approximately equal in potency to chloroquine, with a good therapeutic index except on the fetus. The preparation of semisynthetic derivatives has been stimulated primarily by a requirement for improved solubility because artemisinin is relatively insoluble in both water and oil.

Reduction of (**152**) with sodium borohydride occurs at the lactone carbonyl, leaving the peroxide intact. The resulting cyclic hemiacetal, dihydroartemisinin (**155**), which is a more potent antimalarial than the parent compound, shows typical acetal reactivity. In the presence of acid,

a highly reactive carbocation intermediate allows S_N1-type substitution with a variety of nucleophiles. For example, boron trifluoride catalyzes reactions with methanol and ethanol to give artemether (**156**) and arteether (**157**), respectively, two of the most important derivatives. Both are more potent than the parent compound and have improved solubility in oil. Artemether has been chosen for development under the name Paluther.

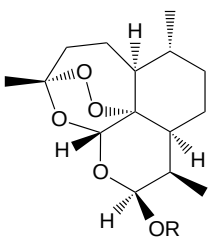

(**155**) R = H
(**156**) R = CH_3 artemether
(**157**) R = CH_2CH_3 arteether
(**158**) R = $COCH_2CH_2COONa$ sodium artesunate

Water solubility can be greatly improved by the standard ploy of esterification with succinic acid and conversion to the sodium salt. Applied to compound (**155**), this technique gives sodium artesunate (**158**), a water-soluble prodrug that may be given intravenously. It may be assumed that hydrolysis occurs *in vivo* to give back (**155**) as the active antimalarial because (**156**) has been shown to be unstable in aqueous solution and because analogous carboxylic acids with a nonhydrolyzable ether link are relatively inactive.

There are two reasons for the great interest being shown in artemisinin and its derivatives. First, there is little cross-resistance with *Plasmodium falciparum* between the members of this series and the quinoline-based antimalarials like chloroquine. On the contrary, significant potentiation of effect is observed in combination with chloroquine analogs such as mefloquine. Second, the high lipid solubility of, for example, artemether ensures rapid penetration into the CNS, so these sesquiterpene lactones are first-line drugs for the treatment of cerebral malaria caused by *P. falciparum*, which is otherwise fatal.

It seems likely that most of the artemisinin found in dried plant material is formed by autoxidation after the death of the plant. From the medicinal chemist's point of view, this is unimportant, but some plant biochemists might have doubts about the description of artemisinin as a "natural product." Air drying in sunlight may be considered a natural, although not a botanical, process. It is probable that many other plant-derived peroxides are formed in a similar way.

Whole plant extracts often show promising activity that may not be traceable to single components. This is obviously not true of *A. annua* extracts, but it is interesting to note that other constituents, notably methoxylated flavones, have potentiating effects on the antimalarial activity of artemisinin.

The reported effect of artemisinin on systemic lupus erythematosus is intriguing, given the history of use of quinine-type antimalarials in this disease.

8.2 Quinine, Chloroquine, and Mefloquine

The use of cinchona bark (e.g. *Cinchona succirubra*) by South American Indians to treat fevers and the subsequent importation of the bark into Europe by Jesuit priests in the seventeenth century is well known. At that time malaria was widespread, even as far north as eastern Scotland and there was no effective

treatment for "the ague." Although quinine (**159**) is not very potent or long acting, a good sample of cinchona bark contains about 5% of the alkaloid. This high concentration permitted genuinely therapeutic doses of bark to be given and allowed the pure alkaloid to be isolated as early as 1820. During the next 100 years, quinine was the only effective treatment for malaria known to Europeans. Without quinine, life in the tropics was impossible for those without natural immunity to malaria. Supplies of quinine to Europe were threatened during World War I, stimulating a major program of research into synthetic analogs.

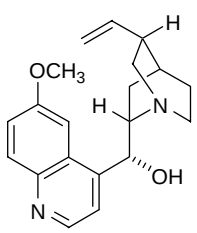

(**159**) Quinine

The chemical techniques available to chemists in the period 1820–1920, although improving rapidly, did not allow a structure to be proposed for quinine with any confidence: the first completely correct proposal came in 1922 and was finally confirmed by total synthesis as late as 1945. However, part structures were known, such as the 6-methoxyquinoline moiety and were sufficient to allow the synthesis of mimics. The first clinically successful mimics were the 8-aminoquinolines.

In the early years of the twentieth century, synthetic organic chemistry was a young discipline, largely governed by empirical rules. Progress toward synthetic analogs of complex natural structures was governed as much by synthetic feasibility as by a desire for close mimicry. The first quinine analogs were, therefore, a combination of the accessible 6-methoxyquinoline part of the quinine structure, with elements of the first successful antimicrobial agents, such as 9-aminoacridine. Nitration followed by reduction could be used to generate a number of new molecules from a variety of parent heterocycles. It is recorded that 4-, 6-, and 8-aminoquinolines have antimalarial properties and, quite extraordinarily, two of these chemical classes are still used today, have quite different uses as antimalarials, and quite possibly have different modes of action.

The first of the 8-aminoquinolines to be introduced into medicine was pamaquine (**160**), not long after World War I. Despite greater toxicity than that of quinine, this class of drugs was found to have radical curative ability against the relapsing malarias. Several hundred analogs were tested during World War II and of these, primaquine (**161**) survives to the present day for short-term use as a radical curative.

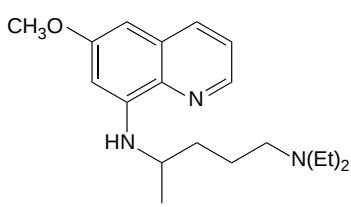

(**160**) Pamaquine

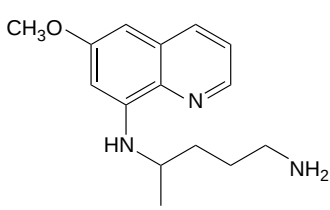

(**161**) Primaquine

Quinacrine (**162**) is an obvious embodiment of the principle outlined above; as a derivative of both quinine and 9-aminoacridine it combined a known antimalarial with a known antimicrobial. The result was a useful, relatively nontoxic antimalarial, although it stained the skin and eyeballs yellow. Despite this side effect and a high incidence of gastrointestinal disturbance, quinacrine was widely used during World War II by European troops in East Asia. The availability of the results of medicinal chemistry research to both sides in wartime is a curious feature of antimalarial development, highlighted below.

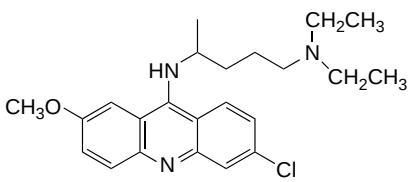

(**162**) Quinacrine (mepacrine)

As has been explained, the major stimulus for research into synthetic antimalarials was not so much the therapeutic inadequacy of quinine as the potential lack of availability in times of social upheaval. During World War II, the United States encouraged the planting of cinchona in Costa Rica, Peru, and Ecuador. The total synthesis of quinine was too difficult in the 1940s and is unlikely to become economically viable even in the new millennium. This problem was partly overcome with quinacrine, which was used widely in World War II, although quinacrine has the defects described earlier. The conceptual derivation of chloroquine (**163**) from quinacrine is obvious and apparently happened twice, in Germany and the United States, the latter about 10 years after the Germans had discarded the drug as being too toxic! The story of the rediscovery of chloroquine is fascinating, as an account of human muddle and misjudgment, finally leading to an extraordinarily valuable drug.

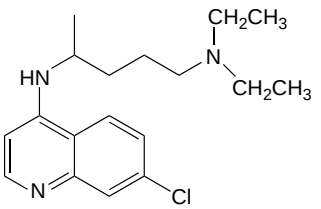

(**163**) Chloroquine

Over decades of sublethal exposure, the resistance of all types of malaria has increased to a point where chloroquine no longer offers certain protection. With the partial exception of quinine and dihydroquinine, resistance to antimalarials had reached the stage at the time of the Vietnam War where more research was required. The development of mefloquine (**164**) was a continuation of the World War II effort, with a gap of about 20 years. Resistance to chloroquine had developed widely during that period, but surprisingly less so to quinine, given the obvious similarities in structure. This observation stimulated a reappraisal of quinolines, known as *quinoline methanols*, which bear a hydroxy

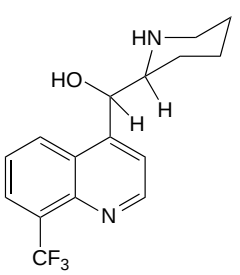

(**164**) Mefloquine

group on the α-carbon of a substituent attached to the 4-position. Up to 1944, a total of 177 quinoline methanols had been synthesized and tested, resulting in one compound (**165**) with activity superior to that of quinine. In human volunteers there was a high incidence of phototoxicity associated with (**165**), so research on quinoline methanols in 1944 had ceased in favor of the 4-amino series, which included chloroquine. Reappraisal of about 100 of the World War II compounds confirmed the high activity and phototoxicity of (**165**) and also showed the high potency of an analog (**166**), which had reduced phototoxicity. These data, together with results from about 200 newer compounds, fostered the belief that phototoxicity was separable from antimalarial activity. Extensive evaluation of (**166**) in humans with chloroquine-resistant *P. falciparum* infections showed promise, but with a significant incidence of toxic reactions; the dose required was also inconveniently large.

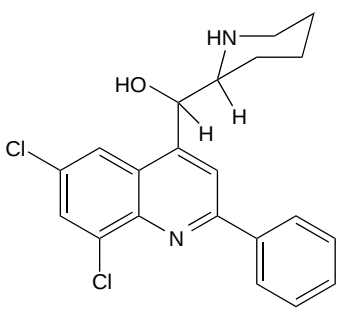

(**165**)

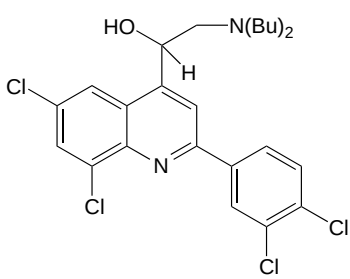

(**166**)

Two hypotheses concerning the effect of the 2-phenyl substituent were proposed. One was that metabolic oxidation was blocked at this position, so that duration of action was prolonged, which was considered desirable. Second, the UV chromophore was enlarged, which would increase the likelihood of drug-induced photosensitivity. The phenyl substituent was thus replaced by trifluoromethyl in the 2-position. Before the first such derivatives were tested, further analogs were prepared with an additional trifluoromethyl group on the benzene ring. This was serendipitous because the first series of 2-trifluoromethyl analogs had low potency and were also photosensitizing. The series with two trifluoromethyl groups, one at position 2 and another in the 6-, 7-, or 8-position were all potent and free from phototoxicity. The most potent was mefloquine (**164**), a very successful drug, but one that produces unacceptable CNS effects in a small proportion of users; parasite resistance has also been observed in parts of Southeast Asia. There is now a serious attempt by the World Health Organization and groups such as the Medicines for Malaria Venture Foundation to find new antimalarials.

Physicians are pragmatic when choosing therapy for patients whose suffering is not alleviated by accepted methods. A drug that has been shown to be toxicologically safe may be utilized in a new area for the flimsiest of reasons. Thus, Page described his use of quinacrine in two cases of lupus erythematosus as being based on "[a] chance observation. . .," although he did not describe the observation that led to his decision. He did, however, record

that quinine had been tried previously and "prevented extension of the lesions," so this may have been the basis for his rationale. In any event, the beneficial effects of quinacrine were remarkable and appeared to be related to the degree of yellowing of the skin that, as described earlier, is a common side effect of the use of quinacrine in malaria.

Among Page's group of patients with lupus erythematosus were two with rheumatoid arthritis, whose symptoms also responded to treatment with quinacrine. The following year, other physicians conducted a trial of quinacrine on a larger group of patients with rheumatoid arthritis; the results encouraged Haydu to test chloroquine on similar patients, again with positive results. A year later, two more physicians compared quinacrine with chloroquine and found the latter to be better tolerated, the majority of patients gaining some benefit. Both quinacrine and chloroquine caused gastrointestinal disturbances, which led to a trial of hydroxychloroquine (**167**), an unsuccessful antimalarial but with less effect on the gut, thus allowing larger doses to be given. Hydroxychloroquine has remained part of the standard drug therapy for rheumatoid arthritis ever since.

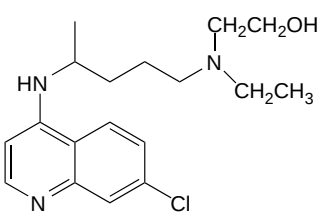

(**167**) Hydroxychloroquine

So far, the choice of quinine-like drugs to treat rheumatoid arthritis has been based on preliminary selection as antimalarials. Because the two types of action are presumably unconnected, there might be some value in a screening program aimed directly at rheumatoid disease.

8.3 Avermectins and Milbemycins

There is no major distinction between the avermectins and milbemycins, which are based on the same complex polyketide macrocycle (**168**): the avermectins are oxygenated at C-13 and bear a disaccharide on this oxygen. They have been isolated from cultures of a number of *Streptomyces* species, obtained from all over the world.

The avermectins, particularly, have been the subject of intense commercial interest because they possess potent activity against both nematode and arthropod parasites of livestock. A full discussion of structure–activity relationships would be out of place here, not least because the data are voluminous, so we shall concentrate on the development of ivermectin, which has been a major success.

Structural designation of avermectins is quaintly based on three series: A, B; a, b; and 1, 2. Greater activity resides in the B series, with a free OH at position 5. There is little difference in potency between the a and b series. In the more potent B series, there are important differences between the 1 series and the 2 series; B_1 is the more active orally, whereas B_2 is the more potent by injection. There are also differences in their spectrum of activity. The spectrum of activity was kept as broad as possible by hydrogenation of a mixture of avermectins B_1a and B_1b to give ivermectin (**169**), which contains at least 80% of 22,23-dihydroavermectin

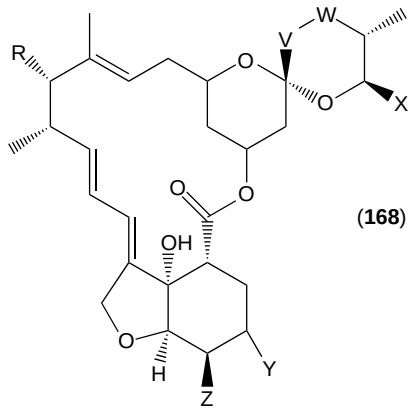

(**168**)

Avermectins R =

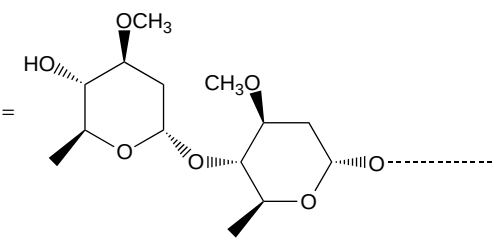

Milbemycins R = H

In the avermectins the series are designated as follows (Y = CH_3):

A, Z = CH_3
B, Z = H
a, X = $CH(CH_3)CH_2CH_3$
b, X = $CH(CH_3)_2$
1, V-W = CH═CH
2, V-W = $CH_2CH(OH)$

In (**169**), V-W = CH_2CH_2, X = $CH(CH_3)CH_2CH_3$ (major) or $CH(CH_3)_2$ (minor), Y = CH_3 and Z = H

B_1a and not more than 20% of 22,23-dihydroavermectin B_1b.

Ivermectin was developed for, and has been highly successful in, the treatment and control of parasites in cattle, horses, sheep, pigs, and dogs. Following studies in humans with river blindness (onchocerciasis), the developers of ivermectin have participated in a major program aimed at eradication of the disease. The sufferers inhabit some of the poorest parts of Africa and cannot pay for their treatment, so the drug has been donated by Merck and Co. Since 1996, more than 20 million treatments have been given. The drug does not kill the adult worms that cause onchocerciasis, but is useful in interrupting the life cycle. Ivermectin is also of value in treatment of scabies. A great deal of information on the biological

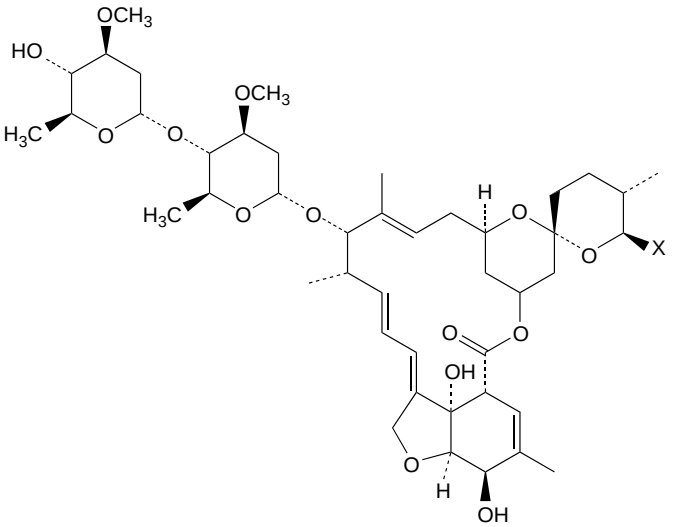

(**169**) X = $CH(CH_3)CH_2CH_3$ (major) or $CH(CH_3)_2$ (minor)

aspects of the use of ivermectin has been published.

9 Immunosuppressant Drugs

9.1 Cyclosporin, Tacrolimus, and Sirolimus

The discovery that the cyclic polypeptide, cyclosporin (**170**), acts as a T-cell selective immunosuppressant by inhibiting calcineurin has lead to a revolution in organ transplantation and related clinical practice. Cyclosporin is produced by fermentation of the fungus *Tolyplocladium inflatum* and was first approved in 1978 for prevention of kidney allograft rejection. The drug is now also used to prevent liver, heart, lung, and bone marrow transplant rejection and for the treatment of psoriasis, atopic dermatitis,

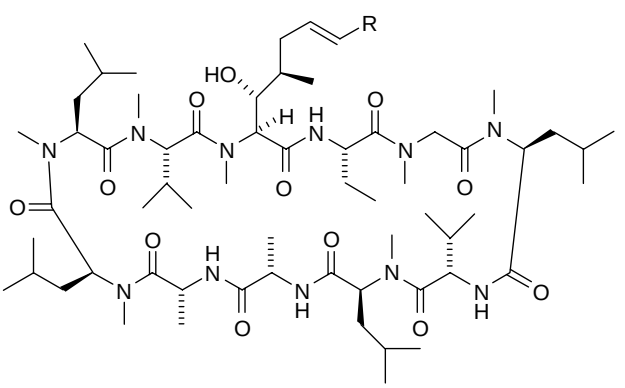

(**170**) R = CH_3
(**171**) R = $CH{=}CH_2$

rheumatoid arthritis, and nephrotic syndrome. A semisynthetic alkene analog of cyclosporin, ISA-247 (**171**), is undergoing clinical trials for prevention of kidney allograft rejection and for treating psoriasis. Two other fermentation products have also played a pivotal role in transplant surgery, that is, the macrolides, tacrolimus (FK506) (**172**) from Astellas and more recently, sirolimus (rapamycin) (**173**) from Wyeth. These drugs are produced by fermentation of *Streptomyces tsukubaenis* and *Streptomyces hygroscopicus* respectively. Various derivatives of sirolimus (**173**), which blocks activation of the lipid kinase mTOR (mammalian target of rapamycin), are being evaluated as anticancer agents.

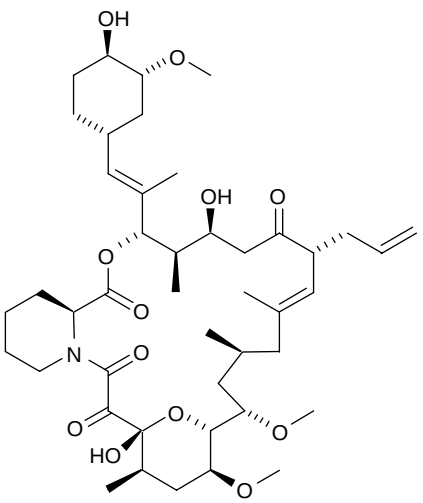

(**172**)

9.2 Mycophenolic Acid and Mycophenolate Mofetil

Finally, mycophenolic acid (**174**), first isolated in 1896 from various cultures of Penicillium was found to inhibit inosine monophosphate dehydrogenase and, as a result, selectively block DNA synthesis in lymphocytes. A prodrug of mycophenolic acid, mycophenolate mofetil (**175**), shows improved bioavailability and is now approved in many countries for use in

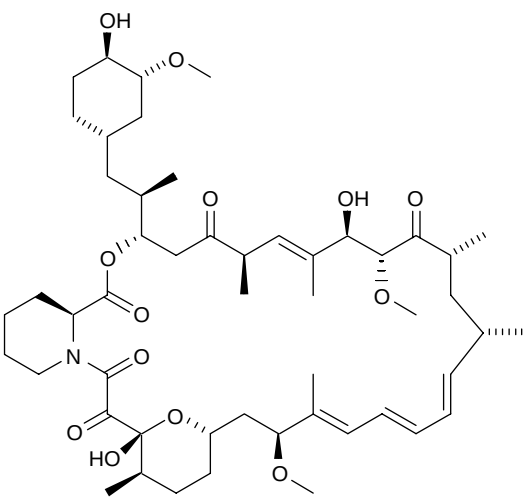

(**173**)

combination with cyclosporin to prevent kidney allograft rejection.

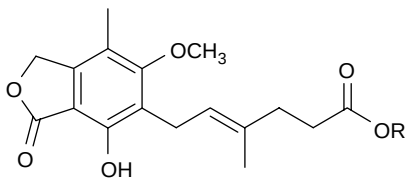

(**174**) R = H

(**175**) R = 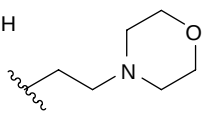

to see whether development leads to better drugs. There are as yet relatively few publications in this area.

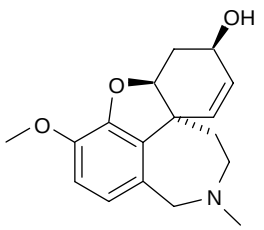

(**176**)

10 Conclusion

Natural product research has been the single most successful strategy for discovering new pharmaceuticals and has contributed dramatically to extending human life and improving clinical practice. As long as Nature continues to yield novel, diverse chemical entities possessing selective biological activities, natural products will play an important role as leads for new pharmaceuticals. An interesting example is the alkaloid galantamine (Nivalin, Reminyl) (**176**), originally isolated from the bulbs of the Amaryllidaceae family (snowdrops, daffodils, etc.), which has found use in the symptomatic treatment of Alzheimer's disease. It is a reversible and competitive inhibitor of acetylcholinesterase that also interacts allosterically with nicotinic acetylcholine receptors to potentiate the action of agonists. By acting to enhance the reduced central cholinergic function associated with this disease, significant improvements in cognition and behavioral symptoms have been observed in patients. In this case, it is the alkaloid itself that is used as the active compound and it will be interesting

Over 90% of bacterial, fungal, and plant species are still waiting to be investigated. High-throughput screening methods will allow large numbers of samples to be tested against more biological targets. An alternative view is that the elucidation of the biological effects of chosen compounds, in some detail, will yield insight into biological processes that may open avenues for medicinal chemistry research that is not based on pure chance. This view is based on the recognition that secondary metabolites have been produced and ruthlessly selected, by evolution, over a long period of time. Either way, the medicinal chemist has a wonderful opportunity to continue utilizing the rich chemical diversity offered by Nature.

The best approach for the identification of natural product leads is a matter of debate. Some very inventive techniques have been used for sample mixtures (extracts) by the bioassay-guided method; for example, by combining automated sample injection/fraction collection with the use of modern stationary phases for column chromatography and appropriate biological assays, active compounds can be readily isolated from crude extracts in just a few

days. An alternative is to use an ethnobotanical or ethnopharmacological technique, whereby the accumulated wisdom of many generations of native plant users may be harnessed in the search for better medicines for all. These two techniques may be combined, so that the native people describe the uses to which they put the plant and the researchers devise a bioassay that is used to find the active components. The problem with any bioassay-guided technique, however, is that the inactive constituents are not identified. This represents a potential waste, given that the microbial strain or plant specimen has had to be collected, preserved, and identified. An alternative approach is to use high-performance liquid chromatography coupled with modern detection technologies (UV, evaporative light scattering, mass spectrometry) to first reduce a microbial culture or plant extract to its secondary metabolites, either as semipurified mixtures or single compounds: the products are then able to be screened for biological activity in a high-throughput manner and the samples can be reevaluated when new screens become available. One thing is certain: the variety of natural product structures, after perhaps 300 million years of natural selection, far exceeds the bounds of human imagination and continues to offer exciting possibilities for novel drug discovery.

Acknowledgment

The present text is a modified and updated version of the chapter ''Natural Products as Leads for New Pharmaceuticals'' published in ''Burger's Medicinal Chemistry and Drug Discovery'' (ISBN 0-471-37032-0) edited by D.J. Abraham in 2003.

See also Biotransformations of Drugs and Chemicals; Medicinal Chemistry; Pharmacokinetics and Pharmacodynamics of Biotech Drugs; Structure-based Drug Design and NMR-based Screening; Targeting and Intracellular Delivery of Drugs.

Bibliography

Books and Reviews

Butler, M.S. (2004) The role of natural product chemistry in drug discovery, *J. Nat. Prod.* **67**, 2141–2153.

Cragg, G.M., Newman, D.J., Snader, K.M. (1997) Natural products in drug discovery and development, *J. Nat. Prod.* **60**, 52–60.

Flynn, E.H. (Ed.) (1972) *Cephalosporins and Penicillins Chemistry and Biology*, Academic Press, New York.

Kaufman, T.S., Ruveda, E.A. (2005) The quest for quinine: those who won the battles and those who won the war, *Angew. Chem., Int. Ed. Engl.* **44**, 854–885.

Koehn, F.K., Carter, G.T. (2005) The evolving role of natural products in drug discovery, *Nature Rev.* **4**, 206–220.

Livett, B.G., Gayler, K.R., Khalil, Z. (2004) Drugs from the sea: conopeptides as potential therapeutics, *Curr. Med. Chem.* **11**, 1715–1723.

Marshall, I.G., Waigh, R.D. (1993) A New Look at Muscle Relaxants – Alternatives to Curare, in: Harvey, A.L. (Ed.) *Drugs from Natural Products*, Ellis Horwood, Chichester, 1993, pp. 131–151.

Newman, D.J., Cragg, G.M., Snader, K.M. (2000) The influence of natural products upon drug discovery, *Nat. Prod. Rep.* **17**, 215–234.

Oberlies, N.H., Kroll, D.J. (2004) Camptothecin and taxol: historic achievements in natural products research, *J. Nat. Prod.* **67**, 129–135.

Shu, Y.-Z. (1998) Recent natural products based drug development: a pharmaceutical industry perspective, *J. Nat. Prod.* **61**, 1053–1071.

Sneader, W. (1985) *Drug Discovery: The Evolution of Modern Medicine*, John Wiley & Sons, New York, 1985, pp. 78–80.

Trigg, P.I. (1989) Qinghaosu (Artemisinin) as an Antimalarial Drug, in: Wagner, H., Hikino, H., Farnsworth, N.R. (Eds.) *Economic and Medicinal Plant Research*, Vol. 3, Academic Press, London, pp. 19–55.

Primary Literature

Ablordeppey, S.Y., Fan, P., Ablordeppey, J.H., Mardenborough, L. (1999) Systemic antifungal agents against AIDS-related opportunistic infections: current status and emerging drugs in development, *Curr. Med. Chem.* **6**, 1151–1195.

Ahlquist, R.P. (1948) A study of the adrenotropic receptors, *Am. J. Physiol.* **153**, 586–600.

Albers-Schönberg, G., Arison, B.H., Hensens, O.D., Hirshfield, J., Hoogsteen, K., Kaczka, E.A., Rhodes, R.E., Kahan, J.S., Kahan, F.M., Ratcliffe, R.W., Walton, E., Ruswinkle, L.J., Morin, R.B., Christensen, B.G. (1978) Structure and absolute configuration of thienamycin, *J. Am. Chem. Soc.* **100**, 6491–6499.

Alberts, A.W., Chen, J., Kuron, G., Hunt, V., Huff, J., Hoffman, C., Rothrock, J., Lopez, M., Joshua, H., Harris, E., Patchett, A., Monaghan, R., Currie, S., Stapley, E., Albers-Schonberg, G., Hensens, O., Hirshfield, J., Hoogsteen, K., Liesch, J., Springer, J. (1980) Mevinolin: a highly potent competitive inhibitor of hydroxymethylglutaryl-coenzyme A reductase and a cholesterol-lowering agent, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3957–3961.

Anrep, G.V., Barsourm, G.S., Kenawy, M.R., Misrahy, G. (1947) Therapeutic uses of khellin: method of standardisation, *Lancet* April 26, 557–558.

Bahal, N., Nahata, M.C. (1992) The new macrolide antibiotics: azithromycin, clarithromycin, dirithromycin, and roxithromycin, *Ann. Pharmacother.* **26**, 46–55.

Balkovec, J.M., Black, R.M., Hammond, M.L., Heck, J.V., Zambias, R.A., Abruzzo, G., Bartizal, K., Kropp, H., Trainor, C., Schwartz, R.E., McFadden, D.C., Nollstadt, K.H., Pittarelli, L.A., Powles, M.A., Schmatz, D.M. (1992) Synthesis, stability, and biological evaluation of water-soluble prodrugs of a new echinocandin lipopeptide. Discovery of a potential clinical agent for the treatment of systemic candidiasis and *Pneumocystis carinii* pneumonia (PCP), *J. Med. Chem.* **35**, 194–198.

Ball, A.P., Geddes, A.M., Davey, P.G., Farrell, I.D., Brookes, G.R. (1980) Clavulanic acid and amoxycillin: a clinical, bacteriological, and pharmacological study, *Lancet* **1**, 620–623.

Bock, M.G. (1991) Development of non-opioid cholecystokinin type B receptor antagonists, *Drug. Future* **16**, 631–640.

Bollag, D.M., McQueney, P.A., Zhu, J., Hensens, O., Koupal, L., Liesch, J., Goetz, M., Lazarides, E., Woods, C.M. (1995) Epothilones, a new class of microtubule-stabilizing agents with a taxol-like mechanism of action, *Cancer Res.* **55**, 2325–2333.

Bonfiglio, G., Furneri, P.M. (2001) Novel streptogramin antibiotics, *Expert Opin. Invest. Drugs* **10**, 185–198.

Bright, G.M., Nagel, A.A., Bordner, J., Desai, K.A., Dibrino, J.N., Nowakowska, J., Vincent, L., Watrous, R.M., Sciavolino, F.C., English, A.R., Retsema, J.A., Anderson, M.R., Brennan, L.A., Borovoy, R.J., Cimochowski, C.R., Faiella, J.A., Girard, A.E., Girard, D., Herbert, C., Manousos, M., Mason, R. (1988) Synthesis, in vitro and in vivo activity of novel 9-deoxo-9a-AZA-9a-homoerythromycin A derivatives; a new class of macrolide antibiotics, the azalides, *J. Antibiot.* **41**, 1029–1047.

Brown, A.G., Smale, T.C., King, T.J., Hasenkamp, R., Thompson, R.H. (1976) Crystal and molecular structure of compactin, a new antifungal metabolite from *Penicillium brevicompactum*, *J. Chem. Soc., Perkin Trans. 1* 1165–1170.

Burgoyne, D.L., Anderson, R.J., Allen, T.M. (1992) Contignasterol, a highly oxygenated steroid with the unnatural 14β configuration from the marine sponge *Petrosia contignata* Thiele, 1899, *J. Org. Chem.* **57**, 525–528.

Burkhart, C.N. (2000) Ivermectin: an assessment of its pharmacology, microbiology and safety, *Vet. Hum. Toxicol.* **42**, 30–35.

Burstein, S.H. (1999) The cannabinoid acids: nonpsychoactive derivatives with therapeutic potential, *Pharmacol. Ther.* **82**, 87–96.

Cairns, H., Fitzmaurice, C., Hunter, D., Johnson, P.B., King, J., Lee, T.B., Lord, G.H., Minshull, R., Cox, J.S.G. (1972) Synthesis and structure-activity relations of disodium cromoglycate and some related compounds, *J. Med. Chem.* **15**, 583–589.

Caldwell, J.E., Heier, T., Kitts, J.B., Lynam, D.P., Fahey, M.R., Miller, R.D. (1989) Comparison of the neuromuscular block induced by mivacurium, suxamethonium or atracurium

during nitrous oxide-fentanyl anaesthesia, *Br. J. Anaesth.* **63**, 393–399.

Casy, A.F., Parfitt, R.T. (1986) *Opioid Analgesics*, Plenum, New York, p. 407.

Chain, E., Florey, H.W., Gardner, A.D., Heatley, N.G., Jennings, M.A., Orr-Ewing, J., Sanders, A.G. (1940) Penicillin as a chemotherapeutic agent, *Lancet* **2**, 226–228.

Corey, E.J., Gin, D.Y., Kania, R. (1996) Enantioselective total synthesis of ecteinascidin 743, *J. Am. Chem. Soc.* **118**, 9202–9203.

Cox, B. (2000) Calcium channel blockers and pain therapy, *Curr. Rev. Pain* **4**, 488–498.

Cragg, G.M., Newman, D.J. (2004) A tale of two tumor targets: topoisomerase I and tubulin. The Wall and Wani contribution to cancer chemotherapy, *J. Nat. Prod.* **67**, 232–244.

Davies, H.G., Green, R.H. (1986) Avermectins and milbemycins, *Nat. Prod. Rep.* **3**, 87–121.

Davies, H.G., Green, R.H. (1991) Avermectins and milbemycins Part I, *Chem. Soc. Rev.* **20**, 211–269.

De Lucca, A.J. (2000) Antifungal peptides: potential candidates for the treatment of fungal infections, *Expert Opin. Invest. Drugs* **9**, 273–299.

Denis, J.N., Greene, A.E., Guénard, D., Guéritte-Voegelein, F., Mangatal, L., Potier, P. (1988) Highly efficient, practical approach to natural taxol, *J. Am. Chem. Soc.* **110**, 5917–5919.

Di Marzo, V., Bifulco, M., De Petrocellis, L. (2004) The endocannabinoid system and its therapeutic exploitation, *Nat. Rev. Drug Discov.* **3**, 771–784.

Douthwaite, S., Champney, W.S. (2001) Structures of ketolides and macrolides determine their mode of interaction with the ribosomal target site, *J. Antimicrob. Chemother.* **48**(Suppl. T1), 1–8.

Elford, B.C., Roberts, M.F., Phillipson, J.D., Wilson, R.J.M. (1987) Potentiation of the antimalarial activity of qinghaosu by methoxylated flavones, *Trans. R. Soc. Trop. Med. Hyg.* **81**, 434–436.

Emmenegger, H., Stahelin, H., Rutschmann, J., Rertz, J., von Wartburg, A. (1961a) On the chemistry and pharmacology of *Podophyllum* glucoside and its derivatives. 1, *Drug Res.* **11**, 327–333.

Emmenegger, H., Stahelin, H., Rutschmann, J., Rertz, J., von Wartburg, A. (1961b) On the chemistry and pharmacology of *Podophyllum* glucoside and its derivatives. 2, *Drug Res.* **11**, 459–469.

Endo, A. (1979) Monacolin K, a new hypocholesterolemic agent produced by a *Monascus* species, *J. Antibiot.* **32**, 852–854.

Endo, A., Kuroda, M., Tsujita, Y. (1976) ML-236A, ML-236B, and ML-236C, new inhibitors of cholesterogenesis produced by *Penicillium citrinium*, *J. Antibiot.* **29**, 1346–1348.

Evans, B.E., Rittle, K.E., Bock, M.G., Dipardo, R.M., Freidinger, R.M., Whitter, W.L., Lundell, G.F., Veber, D.F., Anderson, P.S. (1988) Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists, *J. Med. Chem.* **31**, 2235–2246.

Everett, A.J., Lowe, L.A., Wilkinson, S. (1970) Revision of the structures of (+)-tubocurarine chloride and (+)-chondrocurine, *J. Chem. Soc. D: Chem. Commun.* 1020–1021.

Ferreira, S.H. (1965) A Bradykinin-potentiating factor (BPF) present in the venom of *Bothrops jararca*, *Br. J. Pharmacol.* **24**, 163–169.

Ferreira, S.H., Greene, L.J., Alabaster, V.A., Bakhle, Y.S., Vane, J.R. (1970) Activity of various fractions of bradykinin potentiating factor against angiotensin I converting enzyme, *Nature* **225**, 379–380.

Fleming, A. (1929) On action of cultures of a *Penicillium* with special reference to their use in the isolation of *B. influenzae*, *Br. J. Exp. Med.* **10**, 226–236.

Gerardy, R., Zenk, M.H. (1993) Formation of salutaridine from (*R*)-reticuline by a membrane-bound cytochrome P-450 enzyme from *Papaver somniferum*, *Phytochemistry* **32**, 79–86.

Gerzon, K. (1980) Dimeric Catharanthus Alkaloids, in: Cassady, J.M., Douros, J.D. (Eds.) *Anticancer Agents Based on Natural Product Models*, Academic Press, New York, pp. 271–317.

Goetz, M.A., Lopez, M., Monaghan, R.L., Chang, R.S.L., Lotti, V.J., Chen, T.B. (1985) Asperlicin, a novel non-peptidal cholecystokinin antagonist from *Aspergillus alliaceus*. Fermentation, isolation and biological properties, *J. Antibiot.* **38**, 1633–1637.

Gordon, D.J., Riffind, B.M. (1987) 3-Hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors: a new class of cholesterol-lowering agents, *Ann. Intern. Med.* **107**, 759–761.

Guéritte-Voegelein, F., Guénard, D., Lavelle, F., Le Goff, M.-T., Mangatal, L., Potier, P. (1991)

Relationships between the structure of taxol analogs and their antimitotic activity, *J. Med. Chem.* **34**, 992–998.

Hamada, Y., Hayashi, K., Shiori, T. (1991) Efficient stereoselective synthesis of dolastatin 10, an antineoplastic peptide from a sea hare, *Tetrahedron Lett.* **32**, 931–934.

Haydu, G.G. (1953) Rheumatoid arthritis therapy; a rationale and the use of chloroquine diphosphate, *Am. J. Med. Sci.* **225**, 71–75.

Heinrich, M., Lee, T.H. (2004) Galanthamine from snowdrop – the development of a modern drug against Alzheimer's disease from local Caucasian knowledge, *J. Ethnopharmacol.* **92**, 147–162.

Hill, D.R., Campbell, N.J., Shaw, T.M., Woodruff, G.N. (1987) Autoradiographic localization and biochemical characterization of peripheral type CCK receptors in rat CNS using highly selective nonpeptide CCK antagonists, *J. Neurosci.* **7**, 2967–2976.

Hoffrnann, W.F., Alberts, A.W., Anderson, P.S., Chen, J.S., Smith, R.L., Willard, A.K. (1986) 3-Hydroxy-3-methylglutaryl-coenzyme A reductase inhibitors. 4. Side-chain ester derivatives of mevinolin, *J. Med. Chem.* **29**, 849–852.

Höfle, G., Bedorf, N., Gerth, K., Reichenbach, H., (1993) Epothilone, Deren herstellungsverfahren sowie sie enthaltende mittel, Ger. Pat. DE 91–4138042; *Chem Abstr.* **120**, 52841.

Hsiang, Y.-H., Hertzberg, R., Hecht, S., Lui, L. (1985) Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I, *J. Biol. Chem.* **260**, 14873–14878.

Keller-Juslen, C., Kuhn, M., von Wartburg, A., Stahelin, H. (1971) Mitosis-inhibiting natural products. 24. Synthesis and antimitotic activity of glycosidic lignan derivatives related to podophyllotoxin, *J. Med. Chem.* **14**, 936–940.

Kingsbury, W.D., Boehm, J.C., Jakas, D.R., Holden, K.G., Hecht, S.M., Gallagher, G., Caranfa, M.J., McCabe, F.L., Faucette, L.F., Johnson, R.K., Hertzberg, R.P. (1991) Synthesis of water-soluble (aminoalkyl)camptothecin analogs: inhibition of topoisomerase I and antitumor activity, *J. Med. Chem.* **34**, 98–107.

Kirst, H.A., Sides, G.D. (1989) New directions for macrolide antibiotics: structural modifications and in vitro activity, *Antimicrob. Agents Chemother.* **33**, 1413–1418.

Lands, A.M., Luduena, F.P., Buzzo, H.J. (1967) Differentiation of receptors responsive to isoproterenol, *Life Sci.* **6**, 2241–2249.

Leanza, W.J., Wildonger, K.J., Miller, T.W., Christensen, B.G. (1979) N-Acetimidoyl- and N-formimidoylthienamycin derivatives: antipseudomonal β-lactam antibiotics, *J. Med. Chem.* **22**, 1435–1436.

Lee, F.Y.F., Borzilleri, R., Fairchild, C.R., Kim, S.-H., Long, B.H., Reventos-Suarez, C., Vite, G.D., Rose, W.C., Kramer, R.A. (2001) BMS-247550: a novel epothilone analog with a mode of action similar to paclitaxel but possessing superior antitumor efficacy, *Clin. Cancer Res.* **7**, 1429–1437.

Link, K.P. (1959) The discovery of dicumarol and its sequels, *Circulation* **19**, 97–107.

Mangatal, L., Adeline, M.-T., Guénard, D., Guéritte-Voegelein, F., Potier, P. (1989) Application of the vicinal oxyamination reaction with asymmetric induction to the hemisynthesis of taxol and analogues, *Tetrahedron* **45**, 4177–4190.

Morimoto, S., Takahashi, Y., Watanabe, Y., Omura, S. (1984) Chemical modification of erythromycins. I. Synthesis and antibacterial activity of 6-O-methylerythromycins A, *J. Antibiot.* **37**, 187–189.

Nakamura, C.E., Abeles, R.H. (1985) Mode of interaction of β-hydroxy-β-methylglutaryl coenzyme A reductase with strong binding inhibitors: compactin and related compounds, *Biochemistry* **24**, 1364–1376.

Newton, G.G.F., Abraham, E.P. (1955) Cephalosporin C, a new antibiotic containing sulphur and D-α-aminoadipic acid, *Nature* **175**, 548.

Nicolaou, K.C., Roschangar, F., Dionisios, V. (1998) Chemical biology of epothilones, *Angew. Chem., Int. Ed. Eng1.* **37**, 2014–2045.

Nicolaou, K.C., Riemer, C., Kerr, M.A., Rideout, D., Wrasidlo, W. (1993) Design, synthesis and biological activity of protaxols, *Nature* **364**, 464–466.

Noble, R.L., Beer, C.T., Cutts, J.H. (1958) Role of chance observations in chemotherapy: *Vinca rosea*, *Ann. N. Y. Acad. Sci.* **76**, 882–894.

Ondetti, M.A., Rubin, B., Cushman, D.W. (1977) Design of specific inhibitors of angiotensin-converting enzyme: new class of orally active antihypertensive agents, *Science* **196**, 441–444.

Page, F. (1951) Treatment of lupus erythematosus with mepacrine, *Lancet* **2**, 755–758.

Palomo, C., Arrieta, A., Cossio, F., Aizpurua, J.M., Mielgo, A., Aurrekoetxea, N. (1990) Highly stereoselective synthesis of α-hydroxy β-amino acids through β-lactams: application to the synthesis of the taxol and bestatin side chains and related systems, *Tetrahedron Lett.* **31**, 6429–6432.

Pettit, G.R., Gao, F., Sengupta, D., Coll, J.C., Herald, C.L., Doubek, D.L., Schmidt, J.M., Camp, J.R., Rudloe, J.J., Nieman, R.A. (1991) Isolation and structure of bryostatins 14 and 15, *Tetrahedron* **47**, 3601–3610.

Reading, C., Hepburn, P. (1979) The inhibition of staphylococcal β-lactamase by clavulanic acid, *Biochem. J.* **179**, 67–76.

Rinehart, K.L., Holt, T.G., Fregeau, N.L., Stroh, J.G., Keifer, P.A., Sun, F., Li, L.H., Martin, D.G. (1990) Ecteinascidins 729, 743, 745, 759A, 759B, and 770: potent antitumor agents from the Caribbean tunicate *Ecteinascidia turbinata*, *J. Org. Chem.* **55**, 4512–4515.

Sartorelli, A.C., Creasey, W.A. (1969) Cancer chemotherapy, *Annu. Rev. Pharmacol.* **9**, 51–72.

Schiff, P.B., Fant, J., Horwitz, S.B. (1979) Promotion of microtubule assembly in vitro by taxol, *Nature* **277**, 665–667.

Sheehan, J.C. (1964) *Molecular Modification in Drug Design*, Advances in Chemistry Series, No. 45, American Chemical Society, Washington, DC, 1964, pp. 15–24.

Stahelin, H. (1972) Chemistry and mode of action of podophyllin derivatives, *Planta Med.* **22**, 336–347.

Stahelin, H., von Wartburg, A. (1991) The chemical and biological route from podophyllotoxin glucoside to etoposide: ninth Cain Memorial Award Lecture, *Cancer Res.* **51**, 5–15.

Stenlake, J.B., Waigh, R.D., Dewar, G.H., Hughes, R., Chapple, D.J., Coker, G.G. (1981) Biodegradable neuromuscular blocking agents. Part 4. Atracurium besylate and related polyalkylene diesters, *Eur. J. Med. Chem.* **16**, 515–524.

Sykes, R.B., Bonner, D.P. (1985) Monobactam Antibiotics: History and Development, in: Williams, J.D., Woods, P. (Eds.) *Aztreonam, The Antibiotic Discovery for Gram-negative Infections, Royal Society Medicine International Congress Symposium Series No. 89*, Royal Society Medicine, London, pp. 3–24.

Sykes, R.B., Cimarusti, C.M., Bonner, D.P., Bush, K., Floyd, D.M., Georgopapadakou, N.H., Koster, W.H., Liu, W.C., Parker, W.L., Principe, P.A., Rathnum, M.L., Slusarchyk, W.A., Trejo, W.H., Wells, J.S. (1981) Monocyclic β-lactam antibiotics produced by bacteria, *Nature* **291**, 489–491.

Szallasi, A., Di Marzo, V. (2000) New perspectives on enigmatic vanilloid receptors, *Trends Neurosci.* **23**, 491–497.

van Schepdael, J., Solvay, H. (1970) Clinical study of amiodarone in cardiac rhythmic disorders, *Presse Med.* **78**, 1849–1850.

Wall, M.E., Wani, M.C. (1995) Camptothecin and taxol: discovery to clinic – thirteenth Bruce F. Cain Memorial Award Lecture, *Cancer Res.* **55**, 753–760.

Wall, M.E., Wani, M.C., Cook, C.E., Palmer, K.H., McPhail, H.T., Sim, G.A. (1966) Plant antitumor agents. I. The isolation and structure of camptothecin, a novel alkaloidal leukemia and tumor inhibitor from *Camptotheca acuminata*, *J. Am. Chem. Soc.* **88**, 3888–3890.

Wani, M.C., Taylor, H.L., Wall, M.E., Coggon, P., McPhail, A.T. (1971) Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*, *J. Am. Chem. Soc.* **93**, 2325–2327.

Woodward, R.B., Doering, W.E. (1945) The total synthesis of quinine, *J. Am. Chem. Soc.* **67**, 860–874.

Yarbrough, G.G., Taylor, D.P., Rowlands, R.T., Crawford, M.S., Lasure, L. (1993) Screening microbial metabolites for new drugs – theoretical and practical issues, *J. Antibiot.* **46**, 535–544.

Therapy, Gene: *see* Somatic Gene Therapy

Therapy, Targeted: *see* Targeted Therapy: Genomic Approaches

Three Dimensional Cell Culture: The Importance of Microenvironment in Regulation of Function

Sophie A. Lelièvre[1] *and Mina J. Bissell*[2]
[1] *Purdue University, West Lafayette, IN, USA*
[2] *Lawrence Berkeley National Laboratory, Berkeley, CA, USA*

1 Three-dimensional (3-D) Cell Culture: Cell Shape, Tissue Structure, and Biochemical Signaling 385
1.1 The Discovery of Extracellular Matrix (ECM) Signaling 385
1.1.1 Plasticity of the ECM 386
1.1.2 ECM Signaling Encompasses Specific Cell Membrane Receptors 387
1.1.3 ECM-response Elements 388
1.2 Geometry Matters 388
1.2.1 The ECM Controls Cell Shape 388
1.2.2 Cell Adhesion to the ECM Triggers Intracellular Reorganization 389
1.3 Why 3D? 389

2 Setting up the Culture System 392
2.1 Choice of the Extracellular "Gel" and Other Substrata 392
2.1.1 The ECM 392
2.1.2 The Other Components of the 3D Culture Milieu 397
2.2 The Multiple Facets of Setting Culture Conditions 397
2.2.1 Making the Gel 398
2.2.2 Biochemical and Imaging Manipulations of 3D Cultures 398
2.3 Biochemical, Structural, and Functional Validations 400
2.3.1 Modeling Human Annulus in Culture 400
2.3.2 The Relation Between Functional Integrity Accuracy and Structural in Mammary Acini 400

3 Current Applications of 3D Cell Culture 401
3.1 Biochemical and Architectural Signaling in Physiologically Relevant Context 401
3.1.1 3-D Cell Culture Unravels Complex Interactions Among Signal Transduction Pathways 401

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

3.1.2 3-D Cell Culture Can Lead to the Discovery of Additional Functions for Known Proteins 403
3.2 Recapitulation of Stem Cell Growth, Developmental Fate and Pathological Conditions in 3D Cultures 403
3.2.1 Insights into Mechanisms that Control Normal Tissue Formation 403
3.2.2 3D Culture Systems that Mimic Neoplasias 404
3.3 Large-scale Cell Culture and Bioengineering 405
3.3.1 High-throughput Drug Testing 405
3.3.2 Bioreactors: The Power of Rotation and Hydrodynamics 406
3.3.3 Tissue Production in Bioengineering 406

4 A Glimpse into the Future 407
4.1 The Continuing Search for New Biological and Engineered Extracellular Matrices that Resemble the Physiological Status *In Vivo* 407
4.1.1 Biological Matrices 408
4.1.2 Future Directions in Engineered Matrices 408
4.2 Multiple Elements: Force, Tension and Compliance 409
4.3 The Need for Additional Complexity: 3D Cocultures 410

Acknowledgments 411

Bibliography 412
Books and Reviews 412
Primary Literature 413

Keywords

Biomaterial
Any material from either natural or synthetic sources used for biological applications, to replace or substitute for biological function of living cells or tissues.

Extracellular Matrix (ECM)
It is a meshwork of large fibrillar and globular glycoproteins and hydrophilic proteoglycans, secreted by cells. ECM plays an important role in cell and tissue structure and organ morphogenesis and function by providing biochemical and mechanical signaling to cells and, in combination with interstitial fluids, provides resistance to tensile and compressive stresses.

Hydrogel
It is a soft viscoelastic meshwork that consists of two or more components, one of which is water.

Mechanical Stress
Includes pressure (force acting perpendicular to a surface), shear (force acting parallel to a surface), or stretch, all of which influence the development, remodeling, and pathology of tissues.

Tissue Bioengineering
Formation of tissue *de novo* by promoting cell proliferation and differentiation *in vivo* or *ex vivo* using a physiologically relevant microenvironment and primary cells or appropriate cell lines.

The basic principle of three-dimensional (3D) cell cultures is to place cells in conditions that induce formation of multicellular, tissue-like structures developing in X, Y, and Z planes. The 3D structures are characterized by establishment of adhesion complexes and tissue polarity, and by changes in cytoskeletal structure and cell volume that are significantly different from those found in cells cultured as monolayers (2D). As a result, the phenotypes, function, and regulation of signaling pathways under 2D and 3D conditions are fundamentally different. The science, as well as the art, of 3D cultures center around the recapitulation of physiological behaviors found in the organism, where reciprocal and dynamic cell–cell and cell–ECM interactions affect intracellular biochemical and structural signals, and hence influence gene expression and homeostasis. The 3D culture systems provide unique opportunities to mimic many aspects of developmental and pathological phenomena, to elucidate molecular mechanisms, and to produce tissues for transplantation purposes. With apologies to many colleagues whose works are not cited, this chapter relies heavily on the examples from the authors' laboratories using mammary epithelial cells. This is partly because we are most familiar with this tissue and also, many of the details for mammary models have been worked out in nearly three decades. As such, the experience with mammary acinus is used as a possible road map for other tissues.

1 Three-dimensional (3-D) Cell Culture: Cell Shape, Tissue Structure, and Biochemical Signaling

1.1 The Discovery of Extracellular Matrix (ECM) Signaling

The existence of extracellular matrix (ECM) has been recognized since the origin of microscopy. The term extracellular matrix was proposed following the discovery of cells within connective tissues during the nineteenth century. It was defined as the material outside the cells. The importance of this structure for cell behavior was first suspected when quantitative changes in the density of the ECM were seen to affect cell adhesion properties. This phenomenon was well described

for the adhesion between lens rudiment and the optic vesicle, a process critical for lens cell spreading and during chronic inflammatory changes in the lungs.

During the 1970s, major observations established the importance of the ECM in developmental processes, and cell biologists showed that the different cell types were capable of responding to physical and biochemical influences of the ECM and also secreting ECM components.

These discoveries suggested the existence of a reciprocal relationship between the cells and their ECM (detailed in the following sections) and led to the postulate of the model of dynamic reciprocity between the ECM and the cell nucleus. The model integrated both the physical and biochemical connections between the putative "ECM receptors," the cytoskeleton, the nucleus, and chromatin. Thus, the combined work of cell and developmental biologists as well as biochemist in this area, set the stage for an astonishing scientific adventure aiming at deciphering the role of ECM in development, tissue specificity, and gene expression.

1.1.1 Plasticity of the ECM

The ECM represents a complex world of intricate molecular interactions and is characterized by an enormous diversity in morphology and composition. It is composed of fibrillar molecules such as collagens and elastin and glycoproteins such as fibronectin, laminins, and hydrophilic proteoglycans. In addition, soluble molecules such as growth factors or cytokines can be attached to, trapped, and released from the ECM. The ECM acts as a hydrogel in which fibrillar proteins interact within the hydrated network of the glycosaminoglycan chains of proteoglycans. The combination of the hydrogel properties with the diversity of ECM composition is critical for the function of each type of tissue.

The existence of a different ECM composition depending on the tissue type and/or the degree of tissue maturation was reported almost half a century ago. One of the striking examples of ECM plasticity occurs during skeletogenesis. At early stages, a hyaluronan-rich ECM is critical for mesenchymal cell migration and proliferation. Then, cartilage differentiation involves the removal of hyaluronan and the production of type II collagen and a specific chondroitin sulfate proteoglycan. Additionally, a more specialized ECM organization is seen in the epiphyseal growth plate, and also at the site of initiation of bone formation, which is characterized by a switch of collagen synthesis from type II to type I, and depletion of proteoglycans.

The diversity of ECM composition was soon linked to the fact that different cell types (e.g. mesenchymal, endothelial, and epithelial cells) within a tissue are capable of producing a wide range of ECM molecules. Smooth muscle cells were shown to synthesize connective tissue matrix components, including collagen, elastin, and glycosaminoglycans in the arterial system and other smooth muscle-rich tissues (e.g. uterine myometrium), indicating that the same cell type has the capacity to elaborate soluble, as well as insoluble, components of the ECM. Similarly, endothelial and epithelial cells were shown to synthesize the collagens and glycoproteins of basement membranes (BM, a specialized form of ECM) found around vessels and at the interface between epithelial and connective tissues. The demonstration that the different cell types were capable of synthesizing ECM proteins relied on the ability to culture cells

ex vivo. The ECM components released into the culture medium were found to be similar to that of molecules observed in human tissue, hence confirming that cells were at the origin of ECM production. For instance, fibronectin secreted by human endothelial cells in culture was identical to the fibronectin found in human plasma, and fetal epithelioid cells in culture synthesized type IV procollagen similar to that of collagenous proteins extracted from the tissue's BM.

1.1.2 ECM Signaling Encompasses Specific Cell Membrane Receptors

The observation of a meshwork of ECM molecules on the cell surface suggested a close interaction between cells and ECM molecules. An early milestone in the study of cell–ECM interactions came from work largely pioneered by Hay and colleagues. In an elegant series of experiments, they demonstrated that contacts between cells and the ECM were necessary to sustain the stimulatory effect of ECM on corneal epithelial collagen synthesis. Notably using filters of different pore sizes and thickness, they showed a linear correlation between the cell surface area in contact with the ECM and production of collagen. Many more reports followed describing both the positive and negative regulation of ECM components by cell–ECM interactions. For example, chondrocytes synthesize chondroitin sulfate in the presence of extracellular chondroitin sulfate, whereas they stop synthesis in the presence of hyaluronan.

Early indications that the ECM was influencing cell behavior came from observations that the morphology and biochemistry of normal parenchymal liver cells and normal mammary epithelial cells from prelactating mice were modified upon culture in the presence of collagen I gels compared to tissue culture plastic. Similarly the ECM was shown to influence metabolic pathways and cell proliferation. It was shown later that the ECM in general and the BM, in particular, also controlled cell survival.

Numerous specific ECM receptors have now been identified including prominent transmembrane proteins, the integrins. It is well established that the binding of ECM molecules to cell membrane receptors triggers intracellular signaling, which ultimately reaches the cell nucleus and influences gene expression. Although integrin signaling is quite complex, a major principle of this signaling clearly includes the existence of central signal transducers (e.g. FAK) that can bind to, and recruit other signaling molecules. This network of signaling molecules is connected to the cytoskeleton. Additionally, it has been shown that integrins and growth factor receptors cooperate to control major cellular events such as proliferation. This cooperation is critical to transmit signals via specific signaling cascades. Further transmission of ECM signaling to the cell nucleus partly relies on the nuclear translocation of signaling molecules that are first activated in the cytoplasm. As described in other sections of this chapter, ECM receptors participate also in mechanotransduction where forces affect cytoskeletal organization and signaling cascades, and ultimately influence the cell phenotype. Notably, the interaction between the cytoskeleton and signaling molecules is thought to be critical for changes in cell phenotype resulting from mechanical influence. Thus, both structural (e.g. via cytoskeletal rearrangement) and biochemical aspects of ECM signaling should be integrated to develop a complete view of the control of cell behavior.

1.1.3 ECM-response Elements

A direct connection between ECM signaling and gene regulation was described several years ago upon the discovery of ECM-response elements in several gene promoter sequences. The first ECM-response element was identified and characterized in the enhancer region of the β-casein gene, which is turned on in mammary epithelial cells upon contact between the cells and laminin. It was shown later that chromatin remodeling is involved in regulation of the ECM-response element of the β-casein gene. We identified also a negative ECM-response element and a few other positive elements in the promoters of the TGF-β1 (transforming growth factor) gene and tumor viruses, although these were not characterized in detail. An ECM-response element was identified also in the promoter of LpS1 genes in sea urchin, indicating that ECM-specific signaling to the genome may be a conserved phenomenon. In the latter study, it is proposed that ECM *disruption* triggers a signal that induces a remodeling of binding factors in the ECM-response element and turns off LpS1 gene activity. Thus, both cell adhesion and release of binding to ECM are capable of triggering local chromatin remodeling.

1.2 Geometry Matters

It was originally thought that the ECM had a purely physical role by providing a scaffold to maintain the shape of the organs. However, upon identification of ECM receptors and evidence of ECM signaling, intracellular biochemical pathways triggered by cell adhesion to ECM molecules became an important research focus. In addition, the role of cell shape in ECM-induced mechanical and biochemical signaling is beginning to be accepted, and is an active area of investigation.

1.2.1 The ECM Controls Cell Shape

The influence of the ECM on cell shape was already reported in the early days of ECM research. For instance, it was clear that axon extension and orientation was dependent on the contact with extracellular fibrils. Early on, Bornstein and colleagues suggested that the extracellular meshwork of ECM molecules was interacting directly or indirectly with the internal cytoskeleton, hence mediating changes in cell shape. Today, we can appreciate the importance of this hypothesis since ECM-mediated cell shape appears to play a critical role in the control of gene expression and cell phenotype, and the cytoskeleton appears to be essential for transmission of signals.

Using surfaces with dots of ECM regularly distributed and separated by non-adhesive regions, Singhvi and colleagues demonstrated that the number and distribution of available ECM binding sites dictate cell shape, spreading, and migration. More recently, it was shown that contacts with spacing between 5 and 25 μm trigger cells to adapt their shape, while the extent of cell spreading is correlated with the total surface covered by the ECM. Integrins were identified as critical receptors for the mechanical effect of the ECM. These transmembrane ECM receptors appear to act as anchor molecules that react to the tension exerted by binding to the ECM molecules. By resisting mechanical loads that are applied to ECM receptors, ECM molecules contribute not only to changes in cell shape but most probably also to the arrangement of cells into specific multicellular structures during tissue morphogenesis.

1.2.2 Cell Adhesion to the ECM Triggers Intracellular Reorganization

Whereas it was relatively straightforward to show the relation between the ECM molecules and the cytoskeleton, how ECM may influence structures beyond the cytoskeleton, such as the cell nucleus, has remained unclear. It has been shown that the mechanical effect induced by ECM on integrins is also accompanied by the deformation of the shape of the cell nucleus as well as the redistribution of nucleoli along the axis of the applied tension field, but much less is known about the subsequent alterations within the molecular organization of the nucleus. Nevertheless, there are reports suggesting an effect of cell shape change on gene expression. The simple fact is that certain cell types can adopt a rounded shape when cultured in the presence of certain ECM components instead of the flat shape seen in 2D cultures, and this is sufficient to induce or suppress the expression of certain genes. Such is the case for the expression of the milk protein, lactoferrin, which is turned on upon laminin-rich ECM-induced roundup of mammary epithelial cells. This effect of ECM is essentially mechanical since it can be reproduced without ECM when cells simply are cultured on the nonadhesive substratum, polyhema, which has been used for decades to show the effect of cell shape on growth regulation.

We have shown that ECM signaling is accompanied with the reorganization of certain nuclear structural proteins such as NuMA, and the cell proliferation regulator, retinoblastoma protein as well as changes in the expression and/or distribution of chromatin components. However, the role played by biochemical signaling versus mechanical signaling in nuclear organization is not clear yet. It will be difficult to sort out a direct ECM effect on nuclear components from an indirect effect due to alterations in gene expression. Indeed, changes in nuclear organization may be a consequence of alterations in gene expression that are associated with changes in chromatin structure and possibly modification of the RNP (ribonucleoprotein) network that has been proposed to connect structural nuclear proteins.

1.3 Why 3D?

The art and science of cell culture started more than a half century ago when Alexis Carrel grew a small piece of chick embryo on a fibrin clot, *ex vivo*. Since then, almost all cells have been cultured on tissue culture plastic. It is in fact remarkable that so many cells can survive when cultured in such extreme conditions, on a plastic surface without their regular microenvironment and in the presence of exogenous factors, which are often fluids that the cells do not encounter directly *in vivo*, as is the case of media containing serum (serum is clotted blood and contains excess growth factors and cytokines released from lysed platelets and other blood cells). Cells cultured on tissue culture plastic are referred to as *monolayer or two-dimensional (2D) cultures*. This is the standard method of culturing cells and has permitted numerous discoveries, including the identification of the different components of the cell machinery and regulatory principles of fundamental cellular events such as proliferation, adhesion, and cell death. However, almost all cells change function drastically in such cultures compared to *in vivo* situation, and differentiated normal cells essentially lose their ability to express tissue-specific genes. While the recognition for the need to "grow" cells

in more physiological settings started in the 1970s, it is only recently that this concept has gained a more universal appeal (Fig. 1). This is especially so because it is now appreciated that important signaling pathways are regulated differently in 3D compared to 2D. Moreover, the study of nuclear organization has revealed significant differences between cells cultured under 2D and 3D conditions; the latter recapitulates the organization observed in tissue sections.

"Do not ask what a cell can do, ask what it does do!" (Gloria Hepner, personal communication). What happens in 2-D cell culture is what a cell *can* do, so the

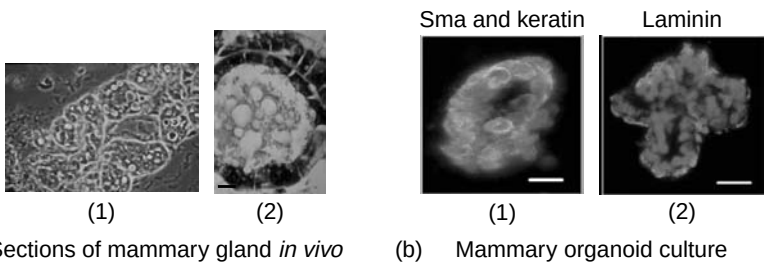

Fig. 1 (a) (1) Thick section through a midpregnant mammary gland shows the gross structure of developing alveolar buds. A dual exposure in which cells are viewed by Nomarski optics and the laminin visualized by rhodamine fluorescence. Bar 10 μm. (From Streuli and Bissell, 1991.) (2) Transverse section of an alveolus from a lactating mouse mammary gland (paraffin embedded, stained with hematoxylin and eosin). Bar, 5 μm (From Streuli and Bissell, 1991.) (b) Primary organoids from virgin mammary gland retain myoepithelial cells and basement membrane (BM). (1) Detection of keratin-positive and α-smooth muscle actin-positive cells in primary mammary organoids at the time of preparation. Frozen sections of mammary organoids embedded in OCT were immunostained simultaneously with mouse anti-α-smooth muscle actin to detect myoepithelial cells (red) and rabbit antikeratin antibodies (green). (2) Organoids were stained with rabbit antilaminin antibodies (green). Scale bars: 30 μm. DAPI nuclear counterstain is in blue. (From Simlan et al., 2001.) The Interplay of matrix metalloproteinases, morphogens and growth factors is necessary for branching of mammary epithellal cells. Development 128: 3117–3131.) (c) (1) Top: Transmission EM (TEM) of primary mammary epithelial cells cultured on plastic. Cells on plastic form a monolayer with the apical surface toward the medium [X 9000]. Bottom: schematic drawing of cells cultured on tissue culture plastic. (2) Left: Schematic drawing of cells cultured on floating collagen gels; right: TEM of primary mammary epithelial cells cultured for 5 days on floating collagen gel. The epithelial cells resemble the cells in the intact gland, and have microvilli and tight junctions at their apical surface. A basal lamina separates the epithelium from the gel and the cells below. The cytoplasm contains an extensive network of distended rough endoplasmic reticulum, Golgi apparatus, secretory vesicles containing dense granular material and fat droplets. Nuclei are central or basal [X 9000]. (3) Left: Transverse section of mammary cells cultured on Matrigel. TEM of a section through a polar and secretory cell structure (acinus) formed by aggregation of dissociated mammary cells from midpregnant mice grown on Matrigel. Right: Schematic presentation of these structures either on top of gel or inside the gel. (TEMs in (c) (1) and (c) (2) are from Emerman, J.T., Pitelka, D.R. (1977) Maintenance and induction of morphological differentiation in dissociated mammary epithelium on floating collagen membranes, *In Vitro* **13**, 316–328, with permission of the publisher; schemes are adapted from Streuli and Bissell, 1991.) (See color plate p. xl.)

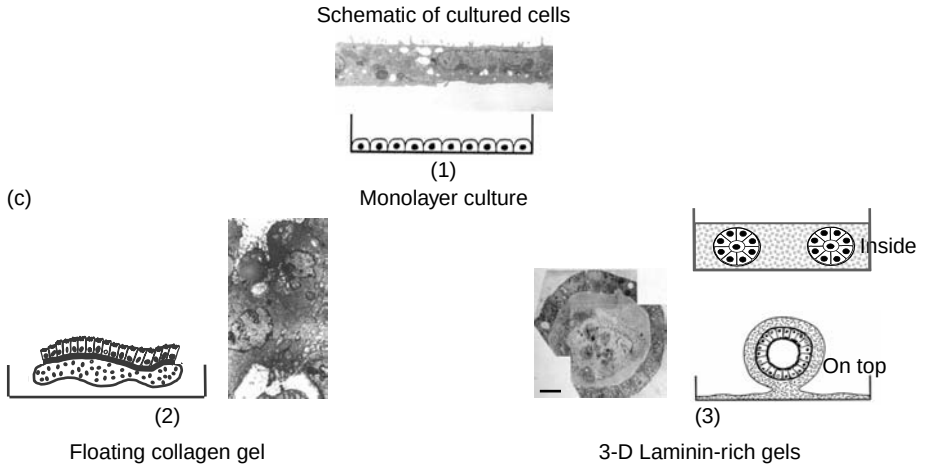

Fig. 1 (*Continued*)

challenge is to find out how one studies what the cell *does* do. Organ cultures are one approach, but unfortunately, this type of culture is usually short lived; the center of the explants becomes necrotic while fibrosis develops in the remaining live tissue. In contrast, 3D cultures within a malleable ECM permit the recapitulation of aspects of cell behavior that are physiologically relevant in a simplified environment as shown with mammary, liver, skin, pancreas, lung, prostate, salivary, thyroid, kidney, endothelial, intestinal, and bone cells.

Not surprisingly, as ECM molecules were being shown to signal to cells, alterations in cell–ECM interactions began to be viewed also as key events in the development of pathologies. Changes in the relationship between cells and the ECM include degradation of ECM molecules via the action of proteases. A striking example is that of the invasive phenotype of cancer cells. Alterations in the secretion of ECM components associated with pathologies were also described early on. Specifically, biochemical defects due to alterations in collagen synthesis were reported in humans as early as 1972. More recently, diseases that are emerging as the twenty-first century's new medical challenges have also been associated with alterations in the interaction between cells and their ECM. Indeed, the major pathological trait of muscular dystrophies is the inability to form correctly glycosylated dystroglycans, hence resulting in the lack of proper adhesion to laminin. Potential treatment strategies aim at promoting the reestablishment of cell–ECM interactions necessary to sustain muscle activity by targeting certain glycosyl transferases that stimulate dystroglycan glycosylation. Defect in dystroglycan has been proposed to play a role also in breast cancer.

3D cultures help greatly in unraveling mechanisms involved in normal and pathological developments because their experimental setup is based on established knowledge. We know that contact with appropriate ECM is critical for cell behavior and shape, and that specific soluble factors are required for the optimal function of tissues. Placing cells in the correct microenvironment promotes the formation of tissuelike 3D structures that develop in the X, Y, and Z planes. Thus, 3D cultures

provide a controlled microenvironment, the complexity of which may be increased to recapitulate additional features of the natural cellular environment. However, it is important to remember that ultimately the art and science of 3D cultures rely on the ability to test the system chosen for compliance with tissue or organ physiology *in vivo*. Thus, it is also crucial to know the limits of a particular system that is being used.

2 Setting up the Culture System

When thinking about setting up 3D cultures, it is important to take into account the physiological and histological studies performed on the tissue of interest. The culture system should allow the necessary readout for a particular question being asked without compromising the phenotype being studied. Once the necessary environment has been established, quality control has to be maintained throughout the experiment, whether the goal is to investigate a particular signal transduction pathway or to propagate cells for tissue replacement.

As described in the preceding sections, cell behavior is influenced by shape and chemistry. The 3D culture system has to address both of these aspects, but an ECM-like scaffold is usually not sufficient to provide the complete endocrine and paracrine milieu. Therefore, specific soluble factors, including growth factors and hormones should also be provided.

To illustrate the different aspects of 3D cultures, examples described in the following sections will be taken mainly from the monoculture of rodent and human mammary cells, a system we helped develop and thus are particularly familiar with. Nevertheless, the logic associated with setting up the correct 3D culture systems can be applied to any cell type. Cell coculture systems that combine several cell types will be discussed in Sect. 4.3 of this chapter. Monoculture, or monotypic cell models, refers to cell culture models involving one cell type, whereas coculture, or heterotypic cell models, refers to cell culture models involving two or more cells types.

2.1 Choice of the Extracellular "Gel" and Other Substrata

2.1.1 The ECM

Although most cell types studied so far have been shown to synthesize ECM components, setting up 3D cultures usually requires the use of either natural or synthetic exogenous ECM, in order to provide the scaffold necessary for initial cell shape change and to trigger the biochemical signaling necessary for differentiation-linked growth arrest and morphogenesis in non-neoplastic cells. Almost all cells make their own ECM environment and depending on factors such as cell type and tissue polarity, the ECM can be deposited as an organized structure that envelopes the multicellular structures (e.g. the BM in polarized epithelia) and is formed against the exogenous substrata.

The choice of the exogenous ECM depends on the tissue type; however, ECM components should be provided as a *hydrogel,* and not as a solid substratum, in order to comply with major physical characteristics of natural ECM environments. Normal epithelial and endothelial cells abut BM components, while normal mesenchymal cells are in contact with connective tissue ECM. In contrast, even when they originate from epithelial tissues, invasive

tumor cells both make, and are in contact with, connective ECM and often make aberrant amounts of BM components. The need for specific ECM components for structural and functional differentiation of mammary cells was demonstrated early on. We showed that mouse mammary epithelial cells made milk proteins if they were cultured on Matrigel™ or laminin-1, but would not do so in response to collagen I or fibronectin. Later, in collaboration with the Petersen laboratory we showed that the assay could be used also for distinguishing normal and malignant human breast epithelial cells, and that nonneoplastic mammary epithelial cells would not form polarized structures (acini) in 3D collagen I gels unless laminin-1 (but not laminins-5 or 10/11) was present. The characteristics that are used as a readout for proper acinar differentiation of breast luminal cells (Table 1) signify two important features of 3D cultures: First, given the appropriate milieu, the epithelial cells create their own endogenously produced ECM by secreting and organizing ECM components into a BM (e.g. human epithelial cells make the BM component, collagen IV, when cultured in the presence of exogenous laminin-rich Matrigel™ from murine origin). Second, vectorial secretion into the lumen indicates that this 3D culture setting recapitulates the formation of a polarized tissue structure signified by the presence of a polarity axis defined by an organized BM, lateral cell–cell junctions and apical polarity markers (Fig. 2). The formation of a polarity axis is characteristic of all epithelia and is indispensable for proper tissue function. Surprisingly, a 3D culture system made of a malleable, laminin-1 containing gel appears to be sufficient to induce luminal cell lines to form the structural and functional characteristics of mammary acini (Fig. 2).

With the exception of collagen I, it is difficult to use pure ECM components because they do not gel easily and are cost prohibitive. There are other alternatives: Matrigel™, an ECM mixture that is rich in laminin-1 (about 80%) can trigger the formation of both mouse and human mammary acini. Although different preparations of Matrigel™ have slightly different compositions (see Sect. 2.2.1), it is, nevertheless, an acceptable choice as long as Matrigel™ lots are tested for recapitulation of the expected phenotype in the particular cell type being cultured.

Since the first description of the Matrigel™ use in mammary cells, our laboratory has developed a number of variations outlined in Fig. 3. An earlier study had indicated that addition of laminin to a collagen gel would allow rodent mammary cells to express milk proteins. More

Tab. 1 Evaluation of breast acinar differentiation using immunofluorescence for cellular and extracellular components.

General characteristics	*Markers of baso-apical polarity*
• One layer cells around the lumen as shown by DAPI-staining of cell nuclei • Growth arrest as shown by absence of staining for BrdU and Ki-67 • Cortical localization of F-actin in the cytoplasm	• Basal location of basement membrane components (e.g., laminin, collagen IV) • Cell-cell junction location of beta-catenin and E-cadherin • Apical location of tight junction protein ZO-1 • Apical location of Mucin-1

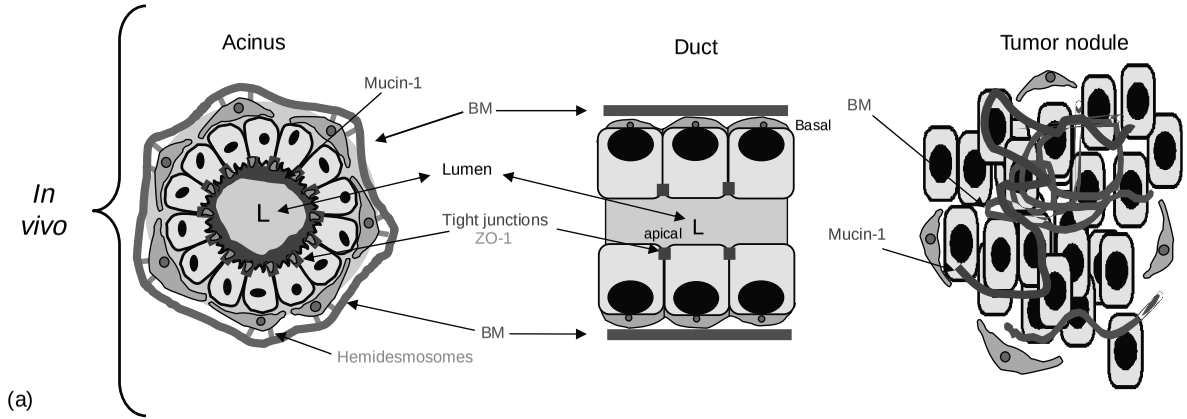

Fig. 2 Scheme of components of a polar acinus and alterations in tumors. (a) Schematic cross section through a mammary acinus (acinus) shows the typical polarized organization. From basal to apical poles: the BM (red), myoepithelial cells (myoeps) (green), luminal cells (eps) (yellow), apical mucin-1 (blue), and the lumen (L). Cell–BM junctions critical for basal polarity are represented in light blue (hemidesmosomes) and cell–cell junctions critical for apical polarity are represented in purple (tight junctions). The localization of a major organizer of tight junctions, ZO-1, is shown in pink. Similarly the polarity axis is shown on a longitudinal section of a normal tubular structure (duct). In contrast in tumors (tumor nodule), when BM components and mucin-1 are expressed and secreted, they are found throughout the multicellular structure and there is no distinguishable baso-apical polarity axis. Cell nuclei are shown in black. (b) In 3D monocellular culture, nonneoplastic epithelial mammary cells are plated as single cells in Matrigel and recapitulate the formation of acini-like structures with a polarity axis within a few days, whereas mammary tumors cells plated as single cells form tumorlike nodules in which mucin-1 and basement membrane components, when secreted, are observed throughout the multicellular structure and ZO-1 is either absent or observed in multiple locations including the cell nucleus and the cytoplasm (not drawn). Cell nuclei are shown in black. (Scheme for (b) is schematic drawing of results described in Petersen, O.W., Ronnov-Jessen, L., Howlett, A.R., Bissell, M.J., 1992.) (See color plate p. xliii.)

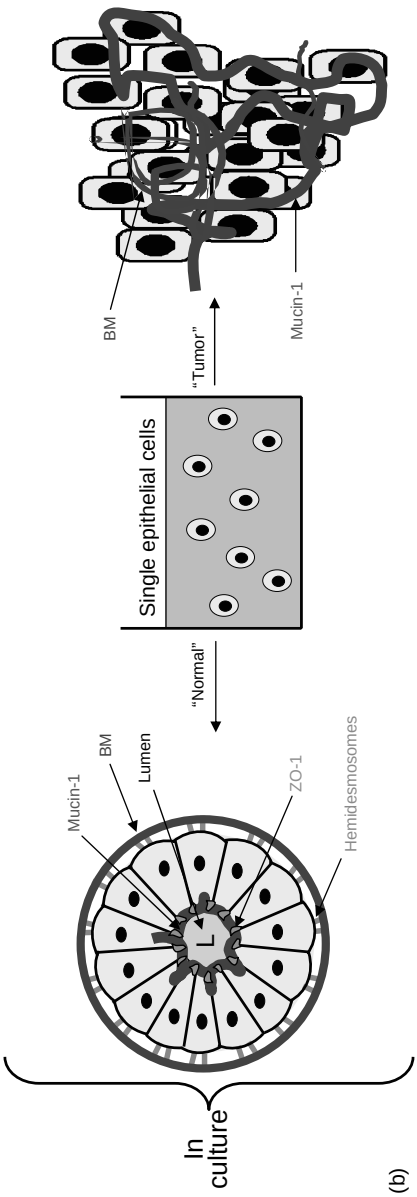

(b)

Fig. 2 (*Continued*)

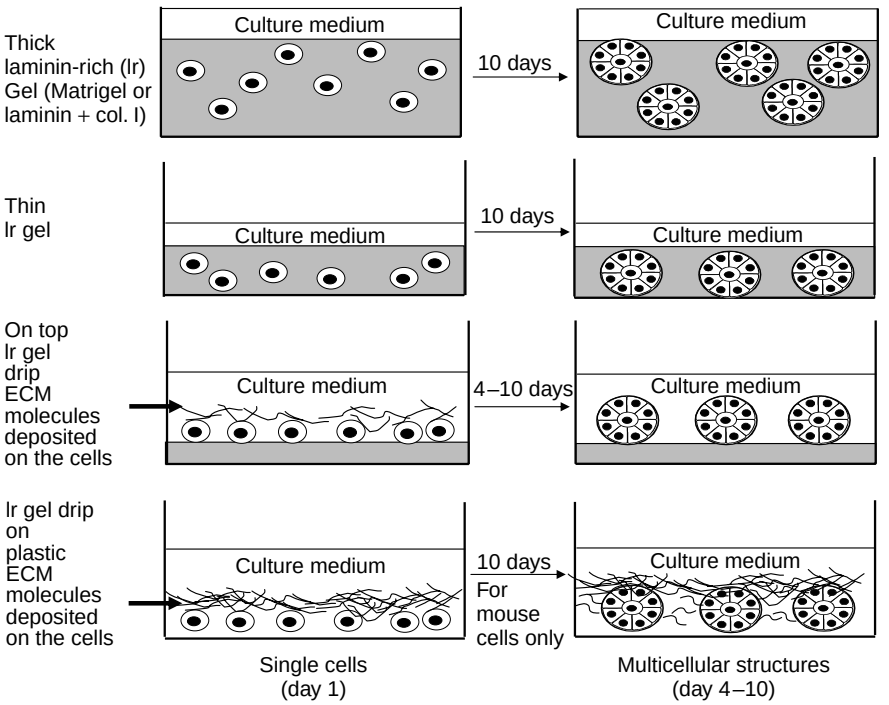

Fig. 3 Scheme of different types of 3D monocultures in the presence of Matrigel. Cells can be cultured as single cells in three different ways, including thick gel, thin gel, culture on top of thin Matrigel plus Matrigel drip, and culture on plastic surface with Matrigel drip (or Matrigel overlay). For primary epithelial breast cells and HMT3522-S1, the formation of multicellular structures that mimic acinar differentiation is observed by day 7, and is complete by day 10. Some mammary cell lines such as MCF10A take longer. We have recently developed techniques that allow formation of acini on top after 4 days (not shown). Although the three techniques using Matrigel generally lead to acini formation with human mammary epithelial cells, the Matrigel drip on tissue culture plastic gives rise to acini with certain cell lines only (e.g. murine cell lines). Matrigel is represented in light gray. Cell nuclei are shown in black.

recently, it was shown that in 3D collagen I gels, addition of laminin restores polarity to MDCK cells, and that addition of 10% laminin-1 to collagen I gels could mimic the formation of mammary acini. This method is particularly interesting because, in contrast to Matrigel™, it uses a controlled exogenous environment consisting of a well-defined scaffold, collagen I, and a relatively pure biochemical inducer, laminin-1.

Interestingly, when comparing the ability of laminin-1, Matrigel™, collagen IV, and fibronectin to induce the formation of acini using prostate epithelial cells, again only laminin-1 and Matrigel™ were efficient exogenous ECM environments confirming the importance of using appropriate BM components as initiators of differentiation. Laminin-1 was shown also to be critical for the formation of polarized acini by human submandibular epithelial cells as well as polarization of lung and kidney epithelial cells. Thus, a relatively similar gelatinous ECM scaffold may be used to initiate differentiation

of different types of epithelial cells with secretory functions.

2.1.2 The Other Components of the 3D Culture Milieu

Tissue differentiation would not be possible if the exogenous ECM and plain culture medium (i.e. medium lacking any additives) were the sole ingredients. Although serum has been of great help for culturing cells by providing necessary nutrients and proliferation and/or survival factors, its composition is poorly understood even now, and variations in its composition most certainly affect the phenotype of cells. It is imperative to culture cells in 3D using a defined culture medium specific for the cell type utilized. One of the earliest realizations of this was a demonstration using fetal human liver cells. Although different preparations of Matrigel™ have slightly different compositions (see Sect. 2.2.1), it is, nevertheless, an acceptable choice as long as lots are tested for the particular type of cells to be cultured, and for the recapitulation of the expected phenotype. As a rule of thumb, a defined medium contains generic survival factors as well as factors and hormones necessary for the differentiation of a particular cell type. For example, prolactin and insulin are important for acinar differentiation of mammary epithelial cells. Similarly, addition of testosterone is important for the differentiation of prostate cells. Growth factors are also chosen on the basis of the cell type. For instance, the epidermal growth factor (EGF) is critical for proliferation and differentiation of mammary epithelial cells. Thus, an understanding of cellular physiology will be particularly important to determine the exact composition of the medium.

The correct concentration of each factor will have to be determined in order to optimize the recapitulation of a particular phenotype. Once established, these concentrations should not be tampered with unless there is a scientific reason for doing so. For instance, addition of EGF to the culture medium was critical for the development of a nonneoplastic human cell line from reduction mammoplasty, and its removal led to selection and/or promotion of a cell variant that became malignant. Abolition of the EGF complement was based on the knowledge that a vast majority of breast tumor cells are EGF independent for their proliferation. Thus, by manipulating the defined culture medium, it was possible to achieve what was usually only achieved by introducing permanently activated oncogenes in cells.

2.2 The Multiple Facets of Setting Culture Conditions

There are at least five different ways to produce the ECM environment necessary for 3D cell culture within the usual laboratory settings: The ECM environment could be (1) from natural origin (e.g. purified denatured BM from tissues), (2) made from single ECM components such as hydrated collagen gels with or without other defined ECM components (can be purified from rat tails), (3) composed of BM mixtures such as Matrigel™ (purified from a mouse sarcoma), (4) made of cellulose sponges with mixtures of ECM components, or (5) composed of engineered ECM-like gels, which are becoming increasingly available but with mixed success. In addition, some of these exogenous ECMs can be used under different conditions so that cells and multicellular structures are either totally embedded or simply maintained on the surface of the exogenous ECM gels.

2.2.1 Making the Gel

3D cell culture is essentially based on the use of hydrated gels. Indeed culturing cells on a coat of dried ECM may be helpful to allow the expression of a few differentiated functions, but does not generally lead to full tissue-like differentiation. Most commonly, cells are embedded within the gel by mixing them in the exogenous ECM used in a liquid form just before plating so that the cells are equally distributed within the depth of the ECM scaffold. After the 40 to 60 min necessary for gel formation with most biologically derived ECMs, the desired medium is added to maintain hydration and to supply nutrients. This technique allows cells to have a fairly homogeneous gel environment and for most epithelial cells it creates an optimum condition for differentiation.

While so far the laminin-rich Matrigel™ has been the workhorse of those who use 3D cultures, the use of collagen I as the primary scaffold may confer additional advantages. Indeed, since purified ECM components can be mixed with collagen I gel to provide additional ECM specificity, the same basic system can be used to compare the effect of different ECM components on a given cell type. However, for each cell type for which conditions have been determined in Matrigel™, strict comparison with the Matrigel™ cultures are necessary to make sure that appropriate markers are being expressed and that the structures formed reflect the *in vivo* situation.

Examples of gel preparation:

1. *Collagen I gels*: Techniques for attached and floating collagen gels have been described in detail.
2. *Matrigel™*: Vials are thawed on ice several hours before use. 2×10^5 mammary epithelial cells are mixed with 0.3-mL chilled Matrigel™ per well of a 24-well dish. Gel formation occurs at 37 °C for about 40 min; then 1 mL of defined medium is added. Medium is changed according to the cell type used. For mammary acinar differentiation, medium is changed three times a week with addition of fresh additives.
3. *Variations on Matrigel technique*: These have been outlined in Fig. 3. Preparation of thinner 3D culture is possible depending on the cell type. Techniques for thin gel preparation are described in Sect. 2.2.2.
4. *Mixed gels*: Optimal gel conditions for mammary acinar differentiation are typically 2×10^5 epithelial cells per 0.5 mL of collagen gel containing 10% laminin or Matrigel™ per well of a 24-well dish. Gel formation occurs at 37 °C for a minimum of an hour, with subsequent addition of 1 mL of defined medium.
5. *Other 3D techniques*: There are a number of variations on the above techniques (e.g. the sponge technique described in Sect. 2.3.1 below). While these are not discussed in detail here, there is much room for innovations. There are already other advanced techniques described for other tissues such as skin (discussed later in section 4.3).

2.2.2 Biochemical and Imaging Manipulations of 3D Cultures

Almost all analytical techniques can be performed on structures produced on, or in, 3D gels. Isolation of multicellular structures requires treatments with ECM-specific proteases (collagenase for collagen I gels and dispase for Matrigel™). Gel solubilization typically takes 30 min at 37 °C with dispase grade II and 30 min at 37 °C with collagenase D. It is critical to wash out proteases: using three washes with

regular additive-free culture medium and a fourth wash in the buffer of choice (e.g. phosphate buffered saline (PBS) with protease and/or phosphatase inhibitors). Each wash is followed by 5-min centrifugations. Please note: it is important NOT to separate cells from each other using trypsin and/or ethylenediaminetetraacetic acid (EDTA) since these treatments will affect cell contacts, the internal cellular organization and gene expression, and consequently may alter the outcome of the experiments. Following washes, the pellet of multicellular structures can be treated for cell fractionation using classical protocols to investigate protein content in the different cellular compartments. We have obtained successful preparations of different types of cell extracts from 3D cultures including nuclear and cytoplasmic extracts, chromatin fractions, or crude membrane extracts by simply applying usual lysis and preparation buffers. The separation of nucleus from cytoplasm is achieved using a dounce homogenizer. We recommend carefully monitoring the separation by examining an aliquot of the cell sample under phase contrast microscope after sets of defined number of strokes, until separation is satisfactory. Indeed, depending on the multicellular structures used, separation may require a higher number of strokes compared to the protocol followed with cells obtained from 2D cultures.

RNA can be directly extracted from 3D cultures. The technique described by Gudjonsson and colleagues (2002a) is particularly simple yet clean enough to obtain RNA that can then be used for microarray analysis. This technique uses Trizol according to the manufacturer's instruction. If necessary more pipetting or even vortexing may help solubilize the gel following incubation with Trizol.

If immunostaining has to be performed, cultures can be embedded in sucrose and frozen in Tissue-Tek optimal cutting temperature (OCT). Fluorescence or horseradish peroxidase staining can be utilized on cryosections according to classical immunostaining protocols used for tissue biopsy.

Immunostaining may also be performed directly on fresh intact cultures. This option is quite useful whenever there is a risk that cryosectioning could damage the multicellular organization, or when the entire depth of the multicellular structure has to be imaged in order to get a global idea of tissue organization, or when freezing samples may interfere with the subtle organization of the subcellular domains to be imaged. It is possible to prepare thinner gels as long as the parameters measured are not impaired (i.e. the differentiation process should be compared to that obtained with the thick gels). For differentiation of mammary epithelial cells in acini, we found that 70 $\mu L\ cm^{-2}$ of an efficient Matrigel™ lot still produced well organized and functional tissue-like structures, permitting a reliable penetration of antibodies and the use of laser scanning microscopy. The thin gel method can also be used to introduce function-blocking antibodies into live cells. It is also possible to use a drip method, in which mammary epithelial cells are mixed with 4% of Matrigel™ before being plated on top of a coat of Matrigel™ (40 $\mu L\ cm^{-2}$) or even on plastic in case of mouse cells. In another drip method we have developed, cells are plated in half the usual volume of medium. The other half of the medium is then mixed with Matrigel™ to achieve 10% final concentration of the gel and dripped slowly over the entire culture surface. Others have described this technique with

slight variations. The Matrigel™ drip need not be repeated again when medium is changed. ECM molecules will rapidly bind to the cells, a process that usually initiates morphogenesis (Fig. 3). It is important to note, however, that the type of secondary fluorochrome-conjugated antibody and the Matrigel™ lots could influence the amount of background staining in the 3D cultures. Thus, immunostaining is performed as usual, except that when necessary, blocking should be performed for 2 h at room temperature, and each washing step should last for 20 min in order to reduce nonspecific background in the Matrigel™. Confocal microscopy and reconstruction of 3D images can be performed easily with these cultures, which are particularly appropriate for immunostaining done to visualize supramolecular patterns of protein distribution in the cell nucleus.

2.3 Biochemical, Structural, and Functional Validations

3D cell cultures could mimic physiologically relevant processes and thus offer a wide range of possible investigations into molecular mechanisms of regulation, as long as quality controls are applied frequently and the desired phenotypes are achieved reproducibly.

In this section, we will give a few examples of the type of controls that are important for determining how well the 3D conditions recapitulate the *in vivo* situation. Obviously, markers (quality controls) would differ with the type of cells and tissues used. This necessitates a thorough knowledge of the physiology and the anatomy of the tissue of interest, and as such will have to take the limits of the 3D culture systems into account.

Earlier we discussed conditions for recapitulation of the mammary acinus in 3D cultures. Whereas mammary acini require laminin-1 (see also the following), a tissue such as annulus would require a different microenvironment.

2.3.1 Modeling Human Annulus in Culture

In order to successfully mimic intervertebral disc formation, it is particularly important to assess the level and the nature of ECM molecules produced *in vivo*. These tissues make copious amounts of type II collagen, aggrecan, and chondroitin-6 sulfotransferase *in vivo*. In a series of tests performed by Gruber and colleagues, collagen sponge was determined to be the best carrier not only because of the correct profile of gene expression but also because sponge induced high levels of correct ECM molecules to be deposited compared to other carriers such as collagen gels.

2.3.2 The Relation Between Functional Integrity Accuracy and Structural in Mammary Acini

Three-dimensional cell cultures are powerful for analysis of signaling and functional regulation, and they also provide an interesting challenge: The acini formed by mammary epithelial cells in 3D culture, for example, are traditionally constructed with only one cell type: the luminal epithelial cells. However, acini formed *in vivo* in the mammary gland contain a dual "tube" consisting of luminal and myoepithelial cells (see Fig. 2). Although the cultured acini for all practical purposes resemble the acini formed *in vivo* (even the sizes are similar; see Fig. 1), the basal pole of luminal cells is touching only the BM whereas, it is in contact with the myoepithelial cells in the normal gland. Thus, in monoculture in 3D, basal polarity markers

hemidesmosomes are formed between the luminal cells and the BM, rather than between the myoepithelial cells and the BM as is the case *in vivo*. Interestingly, however, the hemidesmosomes appear to be normal and are made through connections with $\alpha 6\beta 4$ integrins, which is the correct integrin dimer. This means that luminal cells in culture in the absence of myoepithelial cells turn on the expression of myoepithelial markers, hence becoming a "hybrid" between the two cell types. Thus, quality control includes not only the assessment of the expression of differentiation markers but also the assessment of their localization within the luminal cells to verify the formation of a baso-apical polarity axis (see Fig. 2). Whereas the cultured acini appear to perform many of the functions of the acini in the gland (e.g. milk production in the case of the mouse mammary cells), it is unlikely that they would be able to perform every function the acini perform *in vivo*. Thus, it is desirable to add the other cell types to the mix to eventually make a complete gland. Whether or not the acini formed by the mixed cell population are equivalent to the double-layered acini *in vivo* remains to be determined. As described in Sect. 4.3 of this chapter, it has recently been possible to coculture mammary epithelial and myoepithelial cells, and induce the formation of double-layered structures. A comparison of gene expression between the acini *in vivo* and the epithelial versus epithelial plus myoepithelial cell cultures should determine the exact contribution of the myoepithelial cells to the ability to form acini; these experiments are in progress in a number of laboratories including ours.

We have reported often on the many differences in signaling that are observed between 2D and 3D cultures. In general, we find that the patterns of gene expression and localization of markers within multicellular structures obtained in 3D culture closely resemble the patterns *in vivo*. However, formation of multicellular structures in 3D takes only a few days and is in a different context than what occurs in natural development of the gland. Thus, it is possible that some of the mechanisms observed during the formation of tissuelike structures in 3D cell culture may not reflect what happens *in vivo*. Whenever possible, it is quite important to confirm the existence of novel genes or patterns of gene expression and localization observed in 3D cultures with sections of tissues *in vivo*.

3 Current Applications of 3D Cell Culture

The use of 3D cell culture has opened new directions for both research purposes and medical applications. In the next three sections of this chapter, we will briefly discuss some of the possible applications.

3.1 Biochemical and Architectural Signaling in Physiologically Relevant Context

3.1.1 3-D Cell Culture Unravels Complex Interactions Among Signal Transduction Pathways

An increasing number of studies have now revisited signal transduction in multicellular structures that recapitulate either phenotypically normal or pathological behaviors.

Both the epidermal growth factor receptor (EGFR) and the $\beta 1$-integrin signaling pathways have been widely studied on monolayer cultures, and were shown to be involved in the control of proliferation and survival. Using 3D cultures of neoplastic mammary epithelial cells

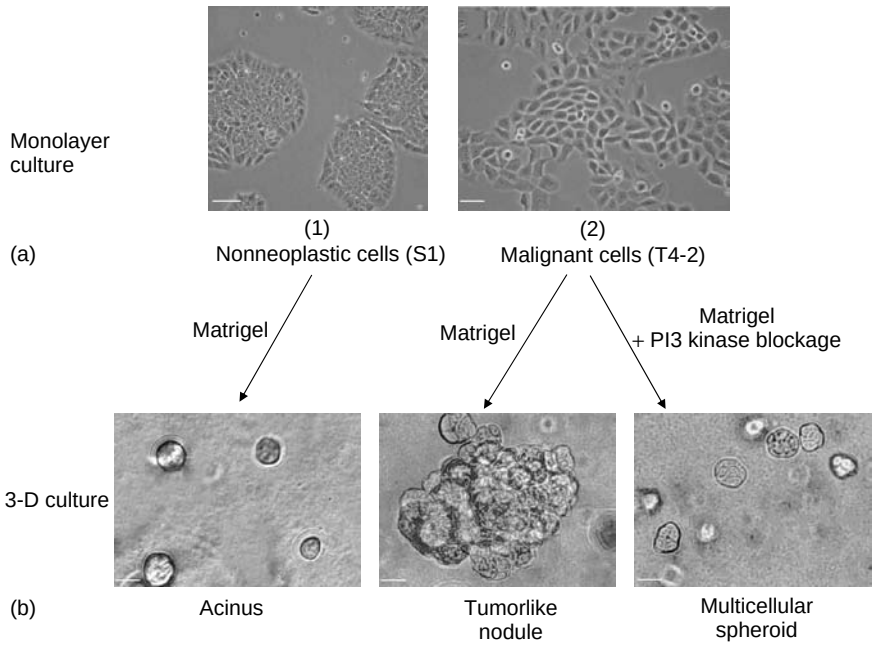

Fig. 4 Morphology of mammary cells in 2D and 3D. (a) Morphology of nonneoplastic S1 and malignant T4-2 HMT3522 cells cultured on plastic. (1) Phase contrast image of nonmalignant breast epithelial cells (HMT3522-S1) cultured for 10 days. Cells form circular, smooth colonies with the inner cells being somewhat compacted. (2) Phase contrast image of tumorigenic T4-2 cells, derived from S1 cell, cultured for 5 days. The cells are larger than S1 cells and do not compact in the center. The cells at the edges only rarely show any distinct organization. Size bars = 25 μm. (b) Phase contrast micrographs of S1, T4-2, and revertant T4-2 cell colonies following treatment with EGFR inhibitor (cells also revert using β-1 integrin inhibitory antibody or PI-3 Kinase inhibitors (see text). When plated inside Matrigel, S1 cells form growth-arrested multicellular structures resembling mammary acini within 10 days of culture, whereas untreated T4-2 cells form disorganized multicellular structures in which cells keep proliferating. (Size bar = 30 μm). (From Myers et al., in: Julio C. (Ed.) *Cell Biology: A Laboratory Handbook*, 3rd edition, 2005, in press.)

that can be "reverted" using inhibitors of either EGFR or β1-integrin pathways (Fig. 4), it was demonstrated that these two pathways were coupled, meaning that downmodulation of one pathway led to downmodulation of both signaling and the level of the relevant receptor in the other pathway. In contrast, no such coupling was observed in cells cultured on 2D substrata.

Beyond its importance for better understanding the differentiation of mammary epithelial cells, the discovery of coupling between a number of signaling pathways in 3D, demonstrated that there exists a higher order control of tissue phenotype, a process that is disrupted when cells are on 2D monolayers. Interestingly, other studies showed that uncoupling EGFR and integrin regulation by overexpressing EGFR in nonneoplastic mammary epithelial cells resulted in the maintenance of Erk activation and inhibition of apoptosis associated with downregulation of Bim. Another example of higher order signal

transduction came from studies of "premalignant" phenotypes: 3D cell cultures of ErbB2-MCF10A mammary epithelial cells were shown to recapitulate features of breast ductal carcinoma *in situ* (DCIS), a noninvasive stage of mammary neoplasia. Using this model, it was recently shown that cooperation between ErbB2 and TGF-β is critical for triggering the invasive behavior of ErB2-MCF10A cells, and that sustained activation of Erk is critical for ErbB2 and TGF-β cooperation.

3.1.2 3-D Cell Culture Can Lead to the Discovery of Additional Functions for Known Proteins

Overexpression of human carcinoembryonic antigen (CEA) and family member (CEACAM6) was shown to prevent differentiation of two colonic cell lines into polarized glandular structures in 3D cultures. Since the cell surface levels of these CEA proteins inversely correlate with the degree of cellular differentiation in freshly excised colonocytes, the 3D assay provides a plausible explanation for the differences in levels of CEA observed in colon carcinoma. In the long run, they also provide a good rationale for testing CEA as a potential target to fight colon cancer. Another example is how 3D cultures have enabled us to study the concept of "architectural" signaling in physiologically relevant assembly of cells. It is known that the internal organization of cells is dependent on the differentiation status. However, the use of 3D cell culture has permitted the discovery that the reverse is also true: the structural organization itself can regulate the degree of differentiation. For instance, the nuclear organization of the structural protein NuMA was significantly modified between early and late stages of breast acinar differentiation in 3D cultures. To demonstrate that this distribution pattern was physiologically relevant, we showed that the characteristic distribution of NuMA in full acinar differentiation in 3D culture was similar to that observed in tissue sections of normal human breast epithelium, but was not observed on monolayers in 2D cultures, even when cells were growth arrested. Conversely, we showed that altering NuMA organization in acinar cells impaired the proper formation of the polarity axis, thus suggesting a potentially important but unappreciated role for NuMA in acinar differentiation.

3.2 Recapitulation of Stem Cell Growth, Developmental Fate and Pathological Conditions in 3D Cultures

3.2.1 Insights into Mechanisms that Control Normal Tissue Formation

Formation of organs involves a combination of different cell types and tissues according to a particular spatial organization. Such complex structures are difficult to recapitulate even in 3D cell culture. However, while formation of complex organ structures in 3D culture is still in its infancy (see also Sect. 4.3), even monotypic 3D cultures have begun to enable us to unravel potential molecular key players for cell/tissue replacement therapies.

For example, myocardial repair is being actively investigated in 3D cultures. A recent study reported that culturing embryonic cardiac myocytes on a tubular scaffold engineered from aligned type I collagen efficiently induced a phenotype resembling neonatal ventricular myocytes *in vivo*.

Stem cells research is bringing additional exciting applications for 3D cell culture. Stem cells are attractive sources for the production of engineered tissues since identifying the conditions that permit the recapitulation of a particular

cell type will permit the replacement of different types of damaged tissues starting from very few progenitor cells. For instance, human mesenchymal stem cells have been induced to differentiate into osteoblasts and adipocytes when cultured on fibrous matrices. Their culture is associated with the secretion of an extensive ECM network composed of collagens I and IV, fibronectin, and laminin that align with fibrils of the exogenous artificial matrix. Differentiation of enteric neuronal precursor cells is being studied using enteric neurospheres in 3D culture in order to better apprehend the development of the enteric nervous system, and putative mammary progenitor cells have been shown to differentiate into terminal ductal lobulal units (TDLU). Many more studies of this kind are sure to follow.

3.2.2 3D Culture Systems that Mimic Neoplasias

The recapitulation of "tumor growth" in 3D cultures, and the ability to generate new information on tumor reversion, apoptosis, and resistance to chemotherapy have attracted much attention in recent years. Three-dimensional cell culture models that mimic malignant phenotypes should provide tools to better understand the mechanisms of malignancy and help develop novel anticancer strategies. Most tumor cell lines have been obtained from invasive tumors and they grow in both collagen I and Matrigel™, giving rise to characteristically disorganized colonies and masses. Moreover, tumor cell lines often produce copious amounts of cytokines, growth factors, and aberrant ECM components that allow them to be cultured in the 3D microenvironment (Fig. 4), despite the fact that cultures of primary cells from tumors are not often easy to grow – probably because they require additional factors from stromal cells; this means that epithelial cancer cell lines turn on stromal components to compensate. Indeed, it is now possible to recreate a reasonable replica of breast tumors by including stromal components in a collagen gel.

One of the most dramatic phenotypes of tumor cells that has been successfully recapitulated in culture is invasion through a thin coat of Matrigel™ deposited on a porous filter. If cells are found on the other side of the filter a few hours after plating, they are considered to possess invasive capabilities (Fig. 4a). This is not strictly a 3D culture technique because cells are cultured for short periods of time and do not form 3D structures. However, invasion can be visualized also in 3D in Matrigel™ after formation of tumorlike nodules. In this case, invasive potential can be visualized by the formation of invasive expansions emanating from the tumorlike nodules (Fig. 5). While the invasive phenotype of malignant cells is fairly easy to recapitulate, especially with cell lines, behavior of benign or premalignant cells has been a challenge to model. 3D cell culture models of premalignant and preinvasive lesions are particularly important to develop because the mechanisms that control formation and progression of these lesions are poorly understood. The HMT3522 progression series recapitulates the formation of nonmalignant breast tissue lesions where spontaneously derived precursors to T4-2 cells form disorganized masses in 3D culture but are not yet malignant in nude mice. Other models also have been developed by overexpression of various oncogenes that alter cellular structure but are not yet sufficient for malignancy. One such induced "pre-malignant" model was developed by activating the ErbB2 signaling pathway, using an inducible system, in nonneoplastic mammary epithelial

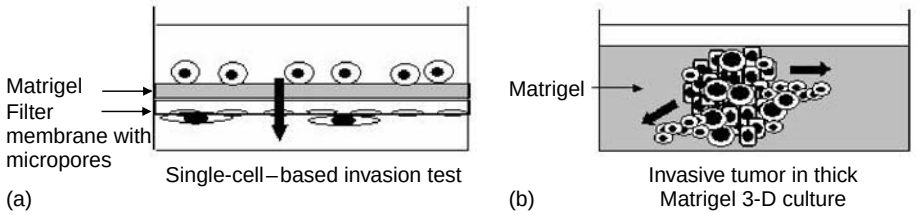

Fig. 5 Assays to assess the invasive phenotype of tumor cells. (a) Classical invasion test performed within 6 to 18 h of culture. The thick arrow shows the direction of invasion through the pores. (b) Invasive phenotype observed in thick Matrigel after 6 to 10 days of 3D culture. Thick arrows show the invasive expansions emanating from the tumorlike nodule. ECM (Matrigel) is shown in light gray. Cell nuclei are shown in black.

cells once they had formed acini. ErbB2 activation induced cell proliferation characterized by filling of the lumina, while the BM remained intact, hence mimicking carcinoma *in situ*, a preinvasive lesion that is associated with an increased risk of developing breast cancer.

Three-dimensional cell cultures have made it possible also to "revert" malignant cells into phenotypically normal acinus-like structures providing a venue to investigate the molecular mechanisms of this poorly studied area. The importance of targeting the microenvironment as well as the tumor itself for cancer therapy is gaining acceptance and modeling this process in 3D would allow increased momentum.

Three-dimensional culture systems are critical also for anticancer drug testing. It was shown almost 15 years ago that the sensitivity of tumor cells to drugs used for cancer treatment is radically different in cells cultured in monolayers compared to cells organized into multicellular spheroids. This has been dramatically illustrated further for drugs that cause apoptosis and chemotherapeutic resistance. Ideally therefore, drug testing should be performed on human cancer cells cultured in 3D and compared with studies in rodents *in vivo*.

3.3 Large-scale Cell Culture and Bioengineering

In the preceding sections, we have discussed techniques and experiments that can be easily performed in basic laboratory settings. There is, however, a need for high-throughput 3D cultures used for drug testing and to produce tissue replacement for medical use.

3.3.1 High-throughput Drug Testing

In order to standardize tumor spheroids for drug testing, techniques such as hanging drops have been suggested. This technique is based on the principle that most cells tend to attach to each other without the need for artificial scaffold. The use of this technique has been recommended because it induces the formation of homogeneously sized spheroids that could be used for automated drug screening. Although homogeneity is an important factor for reproducible drug testing, techniques that use an exogenous ECM scaffold may recapitulate certain alterations in phenotype and/or tumor structure more accurately. The technique we have developed referred to as "on top + drip" (Fig. 3) allows a rapid, 4-day

formation of structures that can be used for high-throughput drug testing as well as imaging.

3.3.2 Bioreactors: The Power of Rotation and Hydrodynamics

This type of 3D culture requires specific equipment, and often a complex exogenous ECM scaffold. The use of bioreactors for 3D culture is based on the idea that these systems may allow feeding of a large number of cells via a spatially homogeneous distribution of the fluid flow, and hence permit the development of highly complex and/or thick tissue structures. The rationale for using microfluidics technology with 3D cell culture systems is based on the hypothesis that it may further help generate an environment similar to the one that exists *in vivo*. Such complex systems have been successfully used to model microscale blood vessels using microfluidic channels. Bioreactors have been used also to recapitulate the differentiation of high-density tissues. For liver cell culture, for example, bioreactors were designed to provide a large surface area for the attachment of a high number of cells, and increased capacity for oxygen and nutrient supplies. One of the most recent culture systems is composed of 10-stacked layers of microfluidic channels made of poly(dimethylsiloxane) (PDMS); the system is linked also to an oxygen chamber. PDMS has high gas permeability and provides high oxygen supply to cells. As in small-scale 3D cultures, primary hepatocytes are plated as single-cell suspension in bioreactors. Upon rapid formation of spheroidal aggregates, differentiation proceeds as shown by the formation of tight junctions, glycogen storage, and canaliculi. 3D microcarrier scaffolds with fluid flow throughout the scaffold have been used also in rotating systems to enhance the elaboration of mineralized bone matrix by osteoblast-like cells.

It is important to emphasize that 3D cultures performed in bioreactors, as described above, are quite different from spinner flasks or rotary cell culture systems used to induce microgravity-like conditions. Bioreactors aim at promoting tissue-like differentiation on a large scale; in contrast, spinner flasks and more recently microgravity-inducing systems, are mostly used to produce large cell aggregates. These latter systems have been popular in the field of cancer research. However, we feel that these systems do not accurately imitate physiologically relevant conditions associated with tumor formation. This is especially so with microgravity where changes in proliferation properties, cytoskeleton, and nuclear organization have been recorded upon comparison with gravity controlled cultures. Thus, the use of purely microgravity-inducing devices should be intended mainly for experiments that investigate the effects of microgravity.

3.3.3 Tissue Production in Bioengineering

3D cultures have major applications in tissue engineering. Typically, cells taken from a donor are placed in 3D culture, on different exogenous substrata, in order to produce tissues that are readily transplantable. A few types of bioengineered tissues are already used for transplantation. However, for most tissue types, the use of 3D cultures is still at the exploratory stages.

Bioengineered tissue transplants have been successfully produced for skin replacement therapies. An example of bioengineered graft is composed of a bovine collagen lattice containing living human

fibroblasts overlaid with a fully differentiated layer of human keratinocytes. Studies performed with nude mice have shown that this system provides a normal healing process, and that the implanted collagen is progressively degraded and replaced with new endogenously produced ECM molecules, such as elastin, synthesized by the living human cells contained within the graft.

Engineering of bone tissue is being actively pursued. The goal is to culture osteoblasts or osteoprogenitor cells within a 3D scaffold, under conditions that favor tissue growth and mineralization. One of the major technical problems is to provide nutrients equally throughout the culture so that cell viability is maintained in all parts of the culture system. Passive nutrient diffusion is not appropriate for such a culture and different techniques of dynamic internal fluid perfusion are being tested. The topography of the culture surface may also be important for the optimal expression of the osteoblastic phenotype. Indeed, it was shown that micropore sizes of polycarbonate membrane surfaces influence protein synthesis and alkaline phosphatase activity. Notably, membranes with larger pore sizes seem to favor differentiation and matrix production.

For tissues with elastic properties such as cartilage, the challenge is to obtain a matrix architecture that has structural and mechanical equivalence to elastic cartilage *in vivo*. A solution to such a challenge may be to use biomimetic hydrogel in a mechanically stable biomimetic composite matrix. Another challenge in bioengineering consists of inducing selective gene expression, notably by triggering proper cell shape. It has been shown that the type of collagen scaffold plays a critical role in determining an optimal cell shape for the production of intervertebral disc tissue.

4 A Glimpse into the Future

In the last section, we presented areas of research that will bring 3D culture to the next level by permitting the recapitulation of complex tissue structures (i.e. tissues that contain several cell types, with each cell type displaying a specific differentiation status). To do so, research efforts focus on designing highly malleable and complex exogenous culture scaffolds and acquiring a better knowledge of the mechanical characteristics of tissues.

4.1 The Continuing Search for New Biological and Engineered Extracellular Matrices that Resemble the Physiological Status *In Vivo*

Exogenous matrices used for 3D cultures have to provide a malleable scaffold as well as the necessary chemical properties for the initiation of tissue differentiation. Natural ECM molecules have inherent properties that greatly influence cell adhesion and shape. For instance, in contrast to many other ECM proteins, tenascins only permit weak adhesion and do not promote cell spreading. Lack of tenascin has been associated with abnormalities in the nervous system, regenerative processes, and skin properties. These observations suggest that, in addition to classical adhesion-promoting ECM proteins, adhesion-modulating/repelling molecules such as tenascin also should be included in biomaterials used for the production of certain types of tissue. This simple example of a particular and critical property of an ECM molecule illustrates the challenges faced by bioengineers. Indeed, for successful tissue engineering, one needs to design complex exogenous scaffolds that

take into account the intricate relationship between different ECM molecules.

4.1.1 Biological Matrices

One of the problems with current biological matrices such as Matrigel™ that are abundantly used in the laboratory is the lack of reliable composition from one batch to another. The inconsistency in the composition of Matrigel™ is inherent to the fact that this material is produced from tumors induced in rodents. Although Matrigel™ has been a useful material for carefully monitored laboratory experiments, other physiological matrices obtained from normal tissues may be required for bioengineering purposes. These matrices have the advantage of providing the complexity of the ECM environment in a more reproducible manner.

Porcine small intestinal submucosa (SIS)-based ECM scaffolds come from ECM produced naturally from normal tissue and thus their composition is fairly constant from one sample to another. SIS-based ECM has been used for several years in tissue bioengineering. Implantation of SIS *in vivo* was shown to facilitate the repair of long-bone defects in rats. SIS has also been shown to be a promising biodegradable scaffold for bladder tissue regeneration *in vivo*. Physiological and complex basal lamina have been used also from living tissues attached to underlying stroma. BM was obtained also following separation from large preovulatory follicles in chicken. The ultrastructure of the isolated BM remains similar to that of the BM observed in the intact ovarian follicle. Although this BM has been successfully used for culturing granulosa cells, it has not been tested widely on other cell types.

4.1.2 Future Directions in Engineered Matrices

Biomaterials for tissue engineering are used for different purposes including facilitation of tissue regeneration, production of carriers to deliver cell transplants, or formation of matrices used to induce tissue morphogenesis in culture before tissue transplantation. Clinical perspectives for tissue engineering require the use of biomaterials that recapitulate physical and chemical properties of natural ECM, while ensuring biocompatibility. It is particularly important that synthetic scaffolds or matrices mimic the hydrogel characteristics of natural ECM in order to maintain tissue-like viscoelastic properties, diffusion, and interstitial flow.

The incorporation of biological characteristics into classical synthetic polymer hydrogels is particularly promising to improve the ECM scaffold required for the recapitulation of complex tissue structures. These gels form *in situ* in the presence of cells, thus facilitating 3D culture setup. Certain gels may even include enzymatic activity that favors cell migration and tissue remodeling. Other promising biomaterials consist of nanofibrillar components that mimic the fibrillar architecture of the ECM, and are formed by self-assembling processes *in situ*. Self-assembly occurs via noncovalent-intermolecular interactions (e.g. assembly of self-complementary amphiphilic peptides in physiological medium). Some of these novel scaffolds can form gels containing more than 99% water and have been shown to maintain and/or promote differentiation of different cell types including chondrocytes and liver progenitor cells. Furthermore, biomolecular signals can also be incorporated in these self-assembling scaffolds by inclusion of ECM-like peptides, for example,

the laminin-based peptide IKVAV. Another interesting approach consists of creating highly defined synthetic ECM analogs, the distribution and concentration of which can be set up as desired on a passive scaffold. These structures may facilitate cell migration or be sensitive to metalloproteinases. Other matrices are designed to permit growth factor concentration and cell-mediated release, hence mimicking natural ECM growth factor presentation and delivery. Finally, certain types of hydrogels formed from self-assembling artificial proteins can undergo controlled gelation and dissolution. These scaffolds are formed by gelation of proteins with terminal leucine zipper domains flanking a central water-soluble polyelectrolyte segment. Formation of the 3D polymer networks is induced in near neutral aqueous solutions. Temperature or pH elevation induces dissolution of the gel and a return to the viscous state. These types of gels are particularly interesting for tissue engineering as they may permit an easy recovery of multicellular structures for experimental analysis.

4.2 Multiple Elements: Force, Tension and Compliance

The importance of mechanical forces in the communication between cells and their ECM is now well established. Understanding how external directional forces affect cell behavior and patterns of gene expression will be of tremendous help for the design of appropriate 3D microenvironments.

Mechanical stress is the pressure, shear, or stretch that influences the development, remodeling, and pathology of tissues. The ECM transduces internal and external mechanical loads into changes in cellular structure and function. Thus, the ECM plays an important role in maintaining a mechanical environment compliant with proper tissue structure and function. The modification of the mechanical environment may affect the maintenance of tissue structure and differentiation by inducing changes in the ECM. These changes can include profound ECM remodeling, as shown by the increase in collagen synthesis and metalloproteinase activity in the presence of higher mechanical load. Mechanical stress is thought to regulate the production of ECM components by inducing the release of growth factors or by directly triggering intracellular signaling pathways. One important family of ECM receptors, integrins, has been shown to be critical for the transmission of mechanical cues to the cells. MAPK and NF-kB pathways are activated in response to variations in mechanical stress. In addition, tension within the cytoskeleton has been shown to be important for mechanotransduction, since ''relaxation'' induced by inhibiting Rho-dependent kinase prevents the induction of tenascin C in stretched fibroblasts. Interestingly, enhancer sequences that respond to static stretch have been identified in the tenascin and collagen XII promoters. Thus, in addition to ECM-response elements, the genome also possesses stretch-response elements within the sequence of specific genes.

The interaction between the ECM and cells plays a role in tissue homeostasis by controlling tissue remodeling, the migration of cells within the tissue, as well as the ECM integrity. An elegant design using different alignments of fibroblasts compared to the mechanical load (as defined by guidance cues) showed that the level of metalloproteinase expression remained high under conflicting guidance

cues, hence suggesting a potential mechanism for the control of ECM remodeling during wound healing. Since cell–ECM interactions are dependent on the mechanical load, the preparation of ECM or ECM-like scaffolds for tissue bioengineering will have to take into account the type of tensional forces that these scaffolds would create. Indeed, different preparations of collagen I (e.g. variations in concentration and pH of the polymerization reaction) were shown to influence fibril density, length, and diameter, and thus tensile mechanical properties of the ECM.

4.3 The Need for Additional Complexity: 3D Cocultures

This chapter has focused mainly on 3D culture settings that deal with "monotypic" cell models. However, tissues are made of complex interactions, occurring either by direct contact or via paracrine effects between multiple cell types. The next challenge in 3D culture is to develop systems that take into account an increasing number of tissue parameters. The knowledge obtained from monoculture of cells in 3D has been providing crucial information for this endeavor.

The influence of coculturing different cell types on phenotypes was observed early on by mixing epithelial and mesenchymal cells. The study of epithelial–mesenchymal interactions has gained increasing popularity for the understanding of tumor behavior. Indeed, peritumoral fibroblasts appear to promote epithelial tumor progression. In breast, carcinomas are typically associated with myofibroblasts. It has been shown that these myofibroblasts originate primarily from resident fibroblasts, although there are other precursor cells as well. Using a collagen matrix and breast tumor epithelial cells, a number of years ago, we developed a "tumor microenvironment" assay to recapitulate critical aspects of breast carcinoma in 3D cultures. These heterotypic cell culture models are likely to provide clues with regard to the influence of stromal cells on epithelial tumor progression. In addition, state-of-the-art techniques, such as microarray analysis, can be adapted for this kind of coculture setting and allows the identification of changes in gene expression profiles induced by heterotypic cell–cell interactions.

Cocultures are also being used to mimic normal tissue differentiation, and to maintain a specific cell population. Coculture of hematopoietic cells and stromal cells on a porous carrier has permitted a better expansion of hematopoietic progenitor cells, without the need for exogenous cytokines, compared to 2D coculture. A coculture system including Schwann cells and dorsal root ganglion neurons in the presence of laminin, is sufficient to provide a microenvironment that favors myelination. *In vivo*, breast luminal epithelial cells are surrounded by myoepithelial cells in the acinar structures. Recently, the use of separated luminal and myoepithelial cells has allowed recapitulation of the "double-layered" tube of breast terminal duct lobular units. The presence of myoepithelial cells was sufficient to induce luminal cell polarity in acini formed in 3D cultures in collagen I, whereas epithelial cells did not show correct polarity when cultured alone in collagen I (Fig. 6). Using this heterotypic cell culture model, it was demonstrated that normal myoepithelial cells dictate acini polarization through their ability to synthesize laminin-1. Thus, the development of new 3D coculture systems that include different cell types

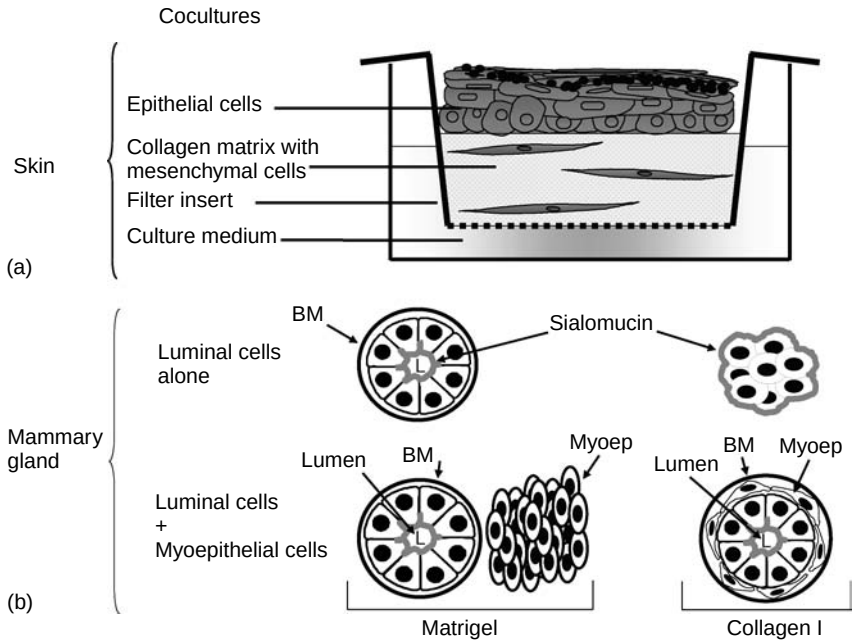

Fig. 6 Coculture models for skin and breast: (a) An experimental model to analyze differentiation of cultured keratinocytes *in vitro*. Keratinocytes have been cultured in organotypic cultures for many years. This particular system results in skinlike "tissues" that in the future could be used as grafts *in vivo*. To achieve this, keratinocytes are cultured on the top of a filter in the upper chamber, and the bottom chamber consists of collagen 1 gels mixed with mesenchymal cells. Medium is below the filter. (From Stark, H.J., Szabowski, A., Fusenig, N.E., Maas-Szabowski, N. (2004) Organotypic cocultures as skin equivalents: a complex and sophisticated in vitro system, *Biol. Proced. Online* **6**, 55–60 and by permission of the authors.) (b) Formation of complete breast epithelial structures (acini) in cocultures. Left panels: Matrigel culture of luminal epithelial cells alone (top) and a mixture (1 : 1) of luminal epithelial and myoepithelial cells (bottom). In Matrigel, luminal, and myoepithelial cells do not mix and each forms distinct colonies. Right panels: Collagen I culture of luminal epithelial cells (top) and a mixture (1 : 1) of luminal epithelial and myoepithelial cells (bottom). Mucin-1 (sialomucin) location is shown in light gray. Coculture in collagen I promotes formation of some dual layered and correctly polarized acinus. Symbols: BM = basement membrane; L = lumen; Myoep = myoepithelial cells. Cell nuclei are shown in black. (Schematic drawings of results from Gudjonsson et al., 2002a.)

found in normal and pathological tissues helps uncover specific interaction properties between cell types. Each new discovery brought by 3D cell culture models will further our knowledge in basic principles of tissue homeostasis and foster the development of optimal settings for tissue bioengineering.

Acknowledgments

We thank Penelope Siig for indispensable expert assistance with the preparation of this manuscript. We thank Drs. G. Steven Martin, Eva Turley, Celeste Nelson, Jordi Alcaraz, and Patricia Abad for critical reading of the manuscript. The work from

the authors' laboratories was supported by the US Department of Energy, Office of Biological and Environmental Research (DE-AC0376 SF00098 to Mina J Bissell), the National Cancer Institute (CA64786-02 to Mina J Bissell and CA57621 to Zena Werb and Mina J Bissell), US Department of Defense Breast Cancer Research Program (DAMD17-00-1-0226 and W81XWH-04-1-0670 to Sophie A Lelièvre), an Innovator Award from the US Department of Defense Breast Cancer Research Program (DAMD17-02-1-0438 to Mina J Bissell), and the Purdue Research Foundation and ''Friends You Can Count On'' Foundation (to Sophie A Lelièvre).

See also Cell Growth in Microgravity; Cell Junctions, Structure, Function, and Regulation; Cellular Interactions; E-Cell: Computer Simulation of the Cell; Mammalian Cell Culture Methods; Programmed Cell Death.

Bibliography

Books and Reviews

Bissell, M.J. (1981) The differentiated state of normal and malignant cells or how to define a ''normal'' cell in culture, *Int. Rev. Cytol.* **70**, 27–100.

Bissell, M.J., Radisky, D. (2001) Putting tumours in context, *Nat. Rev. Cancer* **1**, 46–54.

Bissell, M.J., Rizki, A., Mian, I.S. (2003) Tissue architecture: the ultimate regulator of breast epithelial function, *Curr. Opin. Cell Biol.* **15**, 753–762.

Bissell, M.J., Weaver, V.M., Lelièvre, S.A., Wang, F., Petersen, O.W., Schmeichel, K.L. (1999) Tissue structure, nuclear organization, and gene expression in normal and malignant breast, *Cancer Res.* **59**, 1757s–1763s; discussion 1763s–1764s.

Boudreau, N., Myers, C., Bissell, M.J. (1995) From laminin to lamin: regulation of tissue-specific gene expression by the ECM, *Trends Cell Biol.* **5**, 1–4.

Cairns, R.A., Khokha, R., Hill, R.P. (2003) Molecular mechanisms of tumor invasion and metastasis: an integrated view, *Curr. Mol. Med.* **3**, 659–671.

Chiquet-Ehrismann, R. (2004) Tenascins, *Int. J. Biochem. Cell Biol.* **36**, 986–990.

Cunha, G.R., Hayward, S.W., Wang, Y.Z., Ricke, W.A. (2003) Role of the stromal microenvironment in carcinogenesis of the prostate, *Int. J. Cancer* **107**, 1–10.

Davies, P., Allison, A.C., Cardella, C.J. (1975) The relation between connective tissue cells and intercellular substances, including basement membranes, *Philos. Trans. R. Soc. London, B: Biol. Sci.* **271**, 363–377.

Ingber, D.E. (2002) Mechanical signaling and the cellular response to extracellular matrix in angiogenesis and cardiovascular physiology, *Circ. Res.* **91**, 877–887.

Ingber, D.E. (2003a) Tensegrity II. How structural networks influence cellular information processing networks, *J. Cell Sci.* **116**, 1397–1408.

Ingber, D.E., Prusty, D., Sun, Z., Betensky, H., Wang, N. (1995) Cell shape, cytoskeletal mechanics, and cell cycle control in angiogenesis, *J. Biomech.* **28**, 1471–1484.

Jacks, T., Weinberg, R.A. (2002) Taking the study of cancer cell survival to a new dimension, *Cell* **111**, 923–925.

Kenny, P.A., Bissell, M.J. (2003) Tumor reversion: correction of malignant behavior by microenvironmental cues, *Int. J. Cancer* **107**, 688–695.

Kivirikko, K.I., Risteli, L. (1976) Biosynthesis of collagen and its alterations in pathological states, *Med. Biol.* **54**, 159–186.

Martin, K.H., Slack, J.K., Boerner, S.A., Martin, C.C., Parsons, J.T. (2002) Integrin connections map: to infinity and beyond, *Science* **296**, 1652–1653.

Miranti, C.K., Brugge, J.S. (2002) Sensing the environment: a historical perspective on integrin signal transduction, *Nat. Cell Biol.* **4**, E83–E90.

Piez, K.A. (1997) History of extracellular matrix: a personal view, *Matrix Biol.* **16**, 85–92.

Ross, R. (1975) Connective tissue cells, cell proliferation and synthesis of extracellular

matrix-a review, *Philos. Trans. R. Soc. London, B: Biol. Sci.* **271**, 247–259.

Rosso, F., Giordano, A., Barbarisi, M., Barbarisi, A. (2004) From cell-ECM interactions to tissue engineering, *J. Cell. Physiol.* **199**, 174–180.

Schmeichel, K.L., Bissell, M.J. (2003) Modeling tissue-specific signaling and organ function in three dimensions, *J. Cell Sci.* **116**, 2377–2388.

Silver, F.H., Siperko, L.M. (2003) Mechanosensing and mechanochemical transduction: how is mechanical energy sensed and converted into chemical energy in an extracellular matrix? *Crit. Rev. Biomed. Eng.* **31**, 255–331.

Streuli, C.H. and Bissell, M.J. (1991) Mammary epithellal cells, extracellular matrix and gene expression. In: Lippman, M., Dickson R. (eds.) *Regulatory Mechanisms in Breast Cancer*, Kluwer Academic Publishers, Norwell, MA, pp. 365–381.

Stupack, D.G., Cheresh, D.A. (2002) ECM remodeling regulates angiogenesis: endothelial integrins look for new ligands, *Sci. STKE* **2002**, PE7.

Tlsty, T.D., Hein, P.W. (2001) Know thy neighbor: stromal cells can contribute oncogenic signals, *Curr. Opin. Genet. Dev.* **11**, 54–59.

Toole, B.P., Linsenmayer, T.F. (1977) Newer knowledge of skeletogenesis: macromolecular transitions in the extracellular matrix, *Clin. Orthop.* **129**, 258–278.

Weaver, V.M., Howlett, A.R., Langton-Webster, B., Petersen, O.W., Bissell, M.J. (1995) The development of a functionally relevant cell culture model of progressive human breast cancer, *Semin. Cancer Biol.* **6**, 175–184.

Primary Literature

Abad, P.C., Mian, I.S., Plachot, C., Nelpurackal, A., Bator-Kelly, C., Lelièvre, S. (2004) The C-terminus of the nuclear protein NuMA: phylogenetic distribution and structure, *Protein Sci.* **13**, 2573–2577.

Abbott, A. (2003) Cell culture: biology's new dimension, *Nature* **424**, 870–872.

Aggeler, J., Ward, J., Blackie, L.M., Barcellos-Hoff, M.H., Streuli, C.H., Bissell, M.J. (1991) Cytodifferentiation of mouse mammary epithelial cells cultured on a reconstituted basement membrane reveals striking similarities to development in vivo, *J. Cell Sci.* **99**(Pt 2), 407–417.

Allinen, M., Beroukhim, R., Cai, L., Brennan, C., Lahti-Domenici, J., Huang, H., Porter, D., Hu, M., Chin, L., Richardson, A., Schmitt S., Sellers W.R., Polyak, K. (2004) Molecular characterization of the tumor microenvironment in breast cancer, *Cancer Cell* **6**, 17–32.

Anders, M., Hansen, R., Ding, R.X., Rauen, K.A., Bissell, M.J., Korn, W.M. (2003) Disruption of 3D tissue integrity facilitates adenovirus infection by deregulating the coxsackievirus and adenovirus receptor, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1943–1948.

Anderson, H.C. (1976) Osteogenetic epithelial-mesenchymal cell interactions, *Clin. Orthop.* **116**, 211–223.

Asem, E.K., Feng, S., Stingley-Salazar, S.R., Turek, J.J., Peter, A.T., Robinson, J.P. (2000) Basal lamina of avian ovarian follicle: influence on morphology of granulosa cells in-vitro, *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* **125**, 189–201.

Asselineau, D., Bernhard, B., Bailly, C., Darmon, M. (1985) Epidermal morphogenesis and induction of the 67 kD keratin polypeptide by culture of human keratinocytes at the liquid-air interface, *Exp. Cell Res.* **159**, 536–539.

Bader, A., Knop, E., Kern, A., Boker, K., Fruhauf, N., Crome, O., Esselmann, H., Pape, C., Kempka, G., Sewing, K.F. (1996) 3-D coculture of hepatic sinusoidal cells with primary hepatocytes-design of an organotypical model, *Exp. Cell Res.* **226**, 223–233.

Barboro, P., D'Arrigo, C., Diaspro, A., Mormino, M., Alberti, I., Parodi, S., Patrone, E., Balbi, C. (2002) Unraveling the organization of the internal nuclear matrix: RNA-dependent anchoring of NuMA to a lamin scaffold, *Exp. Cell Res.* **279**, 202–218.

Barcellos-Hoff, M.H., Aggeler, J., Ram, T.G., Bissell, M.J. (1989) Functional differentiation and alveolar morphogenesis of primary mammary cultures on reconstituted basement membrane, *Development* **105**, 223–235.

Barresi, R., Michele, D.E., Kanagawa, M., Harper, H.A., Dovico, S.A., Satz, J.S., Moore, S.A., Zhang, W., Schachter, H., Dumanski, J.P. et al. (2004) LARGE can functionally bypass alpha-dystroglycan glycosylation defects in distinct congenital muscular dystrophies, *Nat. Med.* **10**, 696–703.

Bello-DeOcampo, D., Kleinman, H.K., Deocampo, N.D., Webber, M.M. (2001) Laminin-1

and alpha6beta1 integrin regulate acinar morphogenesis of normal and malignant human prostate epithelial cells, *Prostate* **46**, 142–153.

Ben-Ze'ev, A., Robinson, G.S., Bucher, N.L., Farmer, S.R. (1988) Cell-cell and cell-matrix interactions differentially regulate the expression of hepatic and cytoskeletal genes in primary cultures of rat hepatocytes, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2161–2165.

Bissell, D.M., Tilles, J.G. (1971) Morphology and function of cells of human embryonic liver in monolayer culture, *J. Cell Biol.* **50**, 222–231.

Bissell, M.J., Hall, H.G., Parry, G. (1982) How does the extracellular matrix direct gene expression? *J. Theor. Biol.* **99**, 31–68.

Bissell, D.M., Arenson, D.M., Maher, J.J., Roll, F.J. (1987) Support of cultured hepatocytes by a laminin-rich gel. Evidence for a functionally significant subendothelial matrix in normal rat liver, *J. Clin. Invest.* **79**, 801–812.

Blaschke, R.J., Howlett, A.R., Desprez, P.Y., Petersen, O.W., Bissell, M.J. (1994) Cell differentiation by extracellular matrix components, *Methods Enzymol.* **245**, 535–556.

Bornstein, P., Ash, J.F. (1977) Cell surface-associated structural proteins in connective tissue cells, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2480–2484.

Bornstein, P., Duksin, D., Balian, G., Davidson, J.M., Crouch, E. (1978) Organization of extracellular proteins on the connective tissue cell surface: relevance to cell-matrix interactions in vitro and in vivo, *Ann. N. Y. Acad. Sci.* **312**, 93–105.

Botchwey, E.A., Dupree, M.A., Pollack, S.R., Levine, E.M., Laurencin, C.T. (2003a) Tissue engineered bone: measurement of nutrient transport in three-dimensional matrices, *J. Biomed. Mater. Res.* **67A**, 357–367.

Botchwey, E.A., Pollack, S.R., El-Amin, S., Levine, E.M., Tuan, R.S., Laurencin, C.T. (2003b) Human osteoblast-like cells in three-dimensional culture with fluid flow, *Biorheology* **40**, 299–306.

Briand, P., Nielsen, K.V., Madsen, M.W., Petersen, O.W. (1996) Trisomy 7p and malignant transformation of human breast epithelial cells following epidermal growth factor withdrawal, *Cancer Res.* **56**, 2039–2044.

Carrel, A. (1983a) Landmark article Nov 11, 1911: rejuvenation of cultures of tissues. By Alexis Carrel, *JAMA* **250**, 1085.

Carrel, A. (1983b) Landmark article, Nov 14, 1908: results of the transplantation of blood vessels, organs and limbs. By Alexis Carrel, *JAMA* **250**, 944–953.

Chen, H.M., Schmeichel, K.L., Mian, I.S., Lelièvre, S., Petersen, O.W., Bissell, M.J. (2000) AZU-1: a candidate breast tumor suppressor and biomarker for tumor progression, *Mol. Biol. Cell.* **11**, 1357–1367.

Chiquet, M., Renedo, A.S., Huber, F., Fluck, M. (2003) How do fibroblasts translate mechanical signals into changes in extracellular matrix production? *Matrix Biol.* **22**, 73–80.

Clejan, S., O'Connor, K., Rosensweig, N. (2001) Tri-dimensional prostate cell cultures in simulated microgravity and induced changes in lipid second messengers and signal transduction, *J. Cell. Mol. Med.* **5**, 60–73.

Crouch, E., Bornstein, P. (1979) Characterization of a type IV procollagen synthesized by human amniotic fluid cells in culture, *J. Biol. Chem.* **254**, 4197–4204.

Dairkee, S.H., Ji, Y., Ben, Y., Moore, D.H., Meng, Z., Jeffrey, S.S. (2004) A molecular 'signature' of primary breast cancer cultures; patterns resembling tumor tissue, *BMC Genomics* **5**, 47.

Debnath, J., Muthuswamy, S.K., Brugge, J.S. (2003a) Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures, *Methods* **30**, 256–268.

Debnath, J., Walker, S.J., Brugge, J.S. (2003b) Akt activation disrupts mammary acinar architecture and enhances proliferation in an mTOR-dependent manner, *J. Cell Biol.* **163**, 315–326.

Ebendal, T. (1977) Extracellular matrix fibrils and cell contacts in the chick embryo. Possible roles in orientation of cell migration and axon extension, *Cell Tissue Res.* **175**, 439–458.

Emerman, J.T., Pitelka, D.R. (1977) Maintenance and induction of morphological differentiation in dissociated mammary epithelium on floating collagen membranes, *In Vitro* **13**, 316–328.

Emerman, J.T., Bissell, M.J. (1988) Cultures of mammary epithelial cells: Extracellular matrix and functional differentiation, in: Maramorosch Kas, G.H. (Ed.) *Advances in Cell Culture*, Academic Press, San Diego, CA, pp. 137–155.

Emerman, J.T., Bartley, J.C., Bissell, M.J. (1981) Glucose metabolite patterns as markers of functional differentiation in freshly isolated

and cultured mouse mammary epithelial cells, *Exp. Cell Res.* **134**, 241–250.

Evans, H.J., Sweet, J.K., Price, R.L., Yost, M., Goodwin, R.L. (2003) Novel 3D culture system for study of cardiac myocyte development, *Am. J. Physiol. Heart Circ. Physiol.* **285**, H570–H578.

Folkman, J., Moscona, A. (1978) Role of cell shape in growth control, *Nature* **273**, 345–349.

Fragoso, G., Robertson, J., Athlan, E., Tam, E., Almazan, G., Mushynski, W.E. (2003) Inhibition of p38 mitogen-activated protein kinase interferes with cell shape changes and gene expression associated with Schwann cell myelination, *Exp. Neurol.* **183**, 34–46.

Friedrich, M.J. (2003) Studying cancer in 3 dimensions: 3-D models foster new insights into tumorigenesis, *JAMA* **290**, 1977–1979.

Gaboyard, S., Blanchard, M.P., Travo, C., Viso, M., Sans, A., Lehouelleur, J. (2002) Weightlessness affects cytoskeleton of rat utricular hair cells during maturation in vitro, *Neuroreport* **13**, 2139–2142.

Galbraith, C.G., Yamada, K.M., Sheetz, M.P. (2002) The relationship between force and focal complex development, *J. Cell Biol.* **159**, 695–705.

Geiger, B., Bershadsky, A., Pankov, R., Yamada, K.M. (2001) Transmembrane crosstalk between the extracellular matrix–cytoskeleton crosstalk, *Nat. Rev. Mol. Cell Biol.* **2**, 793–805.

Giguere, L., Cheng, J., Gospodarowicz, D. (1982) Factors involved in the control of proliferation of bovine corneal endothelial cells maintained in serum-free medium, *J. Cell. Physiol.* **110**, 72–80.

Gittes, G.K., Galante, P.E., Hanahan, D., Rutter, W.J., Debase, H.T. (1996) Lineage-specific morphogenesis in the developing pancreas: role of mesenchymal factors, *Development* **122**, 439–447.

Gobin, A.S., West, J.L. (2002) Cell migration through defined, synthetic ECM analogs, *FASEB J.* **16**, 751–753.

Gospodarowicz, D., Delgado, D., Vlodavsky, I. (1980) Permissive effect of the extracellular matrix on cell proliferation in vitro, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4094–4098.

Grant, M.E., Harwood, R., Williams, I.F. (1975) The biosynthesis of basement-membrane collagen by isolated rat glomeruli, *Eur. J. Biochem.* **54**, 531–540.

Grant, D.S., Tashiro, K., Segui-Real, B., Yamada, Y., Martin, G.R., Kleinman, H.K. (1989) Two different laminin domains mediate the differentiation of human endothelial cells into capillary-like structures in vitro, *Cell* **58**, 933–943.

Grayson, W.L., Ma, T., Bunnell, B. (2004) Human mesenchymal stem cells tissue development in 3D PET matrices, *Biotechnol. Prog.* **20**, 905–912.

Gruber, H.E., Ingram, J.A., Leslie, K., Norton, H.J., Hanley, E.N. Jr. (2003) Cell shape and gene expression in human intervertebral disc cells: in vitro tissue engineering studies, *Biotech. Histochem.* **78**, 109–117.

Gruber, H.E., Leslie, K., Ingram, J., Norton, H.J., Hanley, E.N. (2004) Cell-based tissue engineering for the intervertebral disc: in vitro studies of human disc cell gene expression and matrix production within selected cell carriers, *Spine J.* **4**, 44–55.

Gudjonsson, T., Ronnov-Jessen, L., Villadsen, R., Bissell, M.J., Petersen, O.W. (2003) To create the correct microenvironment: three-dimensional heterotypic collagen assays for human breast epithelial morphogenesis and neoplasia, *Methods* **30**, 247–255.

Gudjonsson, T., Ronnov-Jessen, L., Villadsen, R., Rank, F., Bissell, M.J., Petersen, O.W. (2002a) Normal and tumor-derived myoepithelial cells differ in their ability to interact with luminal breast epithelial cells for polarity and basement membrane deposition, *J. Cell Sci.* **115**, 39–50.

Gudjonsson, T., Villadsen, R., Nielsen, H.L., Ronnov-Jessen, L., Bissell, M.J., Petersen, O.W. (2002b) Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties, *Genes Dev.* **16**, 693–706.

Guerret, S., Govignon, E., Hartmann, D.J., Ronfard, V. (2003) Long-term remodeling of a bilayered living human skin equivalent (Apligraf) grafted onto nude mice: immunolocalization of human cells and characterization of extracellular matrix, *Wound Repair Regen.* **11**, 35–45.

Hampton, T. (2004) Studies show tissues surrounding tumors have a role in cancer progression, *JAMA* **291**, 2417–2419.

Hayashi, I., Sato, G.H. (1976) Replacement of serum by hormones permits growth of cells in a defined medium, *Nature* **259**, 132–134.

Hendrix, R.W., Zwaan, J. (1975) The matrix of the optic vesicle-presumptive lens interface during induction of the lens in the chicken embryo, *J. Embryol. Exp. Morphol.* **33**, 1023–1049.

Hennink, W.E., van Nostrum, C.F. (2002) Novel crosslinking methods to design hydrogels, *Adv. Drug Deliv. Rev.* **54**, 13–36.

Hoffman, M.P., Kibbey, M.C., Letterio, J.J., Kleinman, H.K. (1996) Role of laminin-1 and TGF-beta 3 in acinar differentiation of a human submandibular gland cell line (HSG), *J. Cell Sci.* **109**(Pt 8), 2013–2021.

Hoffman, M.P., Nomizu, M., Roque, E., Lee, S., Jung, D.W., Yamada, Y., Kleinman, H.K. (1998) Laminin-1 and laminin-2 G-domain synthetic peptides bind syndecan-1 and are involved in acinar formation of a human submandibular gland cell line, *J. Biol. Chem.* **273**, 28633–28641.

Huang, D. (1977) Extracellular matrix-cell interactions and chondrogenesis, *Clin. Orthop.* **123**, 169–176.

Hutmacher, D.W., Ng, K.W., Kaps, C., Sittinger, M., Klaring, S. (2003) Elastic cartilage engineering using novel scaffold architectures in combination with a biomimetic cell carrier, *Biomaterials* **24**, 4445–4458.

Ibraghimov-Beskrovnaya, O., Ervasti, J.M., Leveille, C.J., Slaughter, C.A., Sernett, S.W., Campbell, K.P. (1992) Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix, *Nature* **355**, 696–702.

Ilantzis, C., DeMarte, L., Screaton, R.A., Stanners, C.P. (2002) Deregulated expression of the human tumor marker CEA and CEA family member CEACAM6 disrupts tissue architecture and blocks colonocyte differentiation, *Neoplasia* **4**, 151–163.

Ingber, D.E. (2003b) Mechanobiology and diseases of mechanotransduction, *Ann. Med.* **35**, 564–577.

Ingber, D.E., Madri, J.A., Jamieson, J.D. (1981) Role of basal lamina in neoplastic disorganization of tissue architecture, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3901–3905.

Ingber, D.E., Madri, J.A., Jamieson, J.D. (1985) Neoplastic disorganization of pancreatic epithelial cell-cell relations. Role of basement membrane, *Am. J. Pathol.* **121**, 248–260.

Ingber, D.E., Madri, J.A., Jamieson, J.D. (1986) Basement membrane as a spatial organizer of polarized epithelia. Exogenous basement membrane reorients pancreatic epithelial tumor cells in vitro, *Am. J. Pathol.* **122**, 129–139.

Jaffe, E.A., Mosher, D.F. (1978) Synthesis of fibronectin by cultured human endothelial cells, *J. Exp. Med.* **147**, 1779–1791.

Kefalides, N.A., Cameron, J.D., Tomichek, E.A., Yanoff, M. (1976) Biosynthesis of basement membrane collagen by rabbit corneal endothelium in vitro, *J. Biol. Chem.* **251**, 730–733.

Kelm, J.M., Timmins, N.E., Brown, C.J., Fussenegger, M., Nielsen, L.K. (2003) Method for generation of homogeneous multicellular tumor spheroids applicable to a wide variety of cell types, *Biotechnol. Bioeng.* **83**, 173–180.

Kerr, D.J., Wheldon, T.E., Kerr, A.M., Kaye, S.B. (1987) In vitro chemosensitivity testing using the multicellular tumor spheroid model, *Cancer Drug Deliv.* **4**, 63–74.

Kirshner, J., Chen, C.J., Liu, P., Huang, J., Shively, J.E. (2003) CEACAM1-4S, a cell-cell adhesion molecule, mediates apoptosis and reverts mammary carcinoma cells to a normal morphogenic phenotype in a 3D culture, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 521–526.

Kisiday, J., Jin, M., Kurz, B., Hung, H., Semino, C., Zhang, S., Grodzinsky, A.J. (2002) Self-assembling peptide hydrogel fosters chondrocyte extracellular matrix production and cell division: implications for cartilage tissue repair, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9996–10001.

Klein, G., Langegger, M., Timpl, R., Ekblom, P. (1988) Role of laminin A chain in the development of epithelial cell polarity, *Cell* **55**, 331–341.

Kleinman, H.K., McGarvey, M.L., Liotta, L.A., Robey, P.G., Tryggvason, K., Martin, G.R. (1982) Isolation and characterization of type IV procollagen, laminin, and heparan sulfate proteoglycan from the EHS sarcoma, *Biochemistry* **21**, 6188–6193.

Kleinman, H.K., McGarvey, M.L., Hassell, J.R., Star, V.L., Cannon, F.B., Laurie, G.W., Martin, G.R. (1986) Basement membrane complexes with biological activity, *Biochemistry* **25**, 312–318.

Kopan, R., Traska, G., Fuchs, E. (1987) Retinoids as important regulators of terminal differentiation: examining keratin expression in individual epidermal cells at various stages of keratinization, *J. Cell Biol.* **105**, 427–440.

Kropp, B.P., Cheng, E.Y. (2000) Bioengineering organs using small intestinal submucosa scaffolds: in vivo tissue-engineering technology, *J. Endourol.* **14**, 59–62.

Kubota, Y., Kleinman, H.K., Martin, G.R., Lawley, T.J. (1988) Role of laminin and basement membrane in the morphological differentiation of human endothelial cells into capillary-like structures, *J. Cell Biol.* **107**, 1589–1598.

Kumar, A., Lnu, S., Malya, R., Barron, D., Moore, J., Corry, D.B., Boriek, A.M. (2003) Mechanical stretch activates nuclear factor-kappaB, activator protein-1, and mitogen-activated protein kinases in lung parenchyma: implications in asthma, *FASEB J.* **17**, 1800–1811.

Leclerc, E., Sakai, Y., Fujii, T. (2004) Microfluidic PDMS (polydimethylsiloxane) bioreactor for large-scale culture of hepatocytes, *Biotechnol. Prog.* **20**, 750–755.

Lee, E.Y., Parry, G., Bissell, M.J. (1984) Modulation of secreted proteins of mouse mammary epithelial cells by the collagenous substrata, *J. Cell Biol.* **98**, 146–155.

Lee, S.J., Choi, J.S., Park, K.S., Khang, G., Lee, Y.M., Lee, H.B. (2004) Response of MG63 osteoblast-like cells onto polycarbonate membrane surfaces with different micropore sizes, *Biomaterials* **25**, 4699–4707.

Lehnert, D., Wehrle-Haller, B., David, C., Weiland, U., Ballestrem, C., Imhof, B.A., Bastmeyer, M. (2004) Cell behaviour on micropatterned substrata: limits of extracellular matrix geometry for spreading and adhesion, *J. Cell Sci.* **117**, 41–52.

Lelièvre, S.A., Weaver, V.M., Nickerson, J.A., Larabell, C.A., Bhaumik, A., Petersen, O.W., Bissell, M.J. (1998) Tissue phenotype depends on reciprocal interactions between the extracellular matrix and the structural organization of the nucleus, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14711–14716.

Li, M.L., Aggeler, J., Farson, D.A., Hatier, C., Hassell, J., Bissell, M.J. (1987) Influence of a reconstituted basement membrane and its components on casein gene expression and secretion in mouse mammary epithelial cells, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 136–140.

Li, S., Edgar, D., Fassler, R., Wadsworth, W., Yurchenco, P.D. (2003) The role of laminin in embryonic cell polarization and tissue organization, *Dev. Cell* **4**, 613–624.

Liu, H., Radisky, D.C., Wang, F., Bissell, M.J. (2004) Polarity and proliferation are controlled by distinct signaling pathways downstream of PI3-kinase in breast epithelial tumor cells, *J. Cell Biol.* **164**, 603–612.

Lutolf, M.P., Lauer-Fields, J.L., Schmoekel, H.G., Metters, A.T., Weber, F.E., Fields, G.B., Hubbell, J.A. (2003) Synthetic matrix metalloproteinase-sensitive hydrogels for the conduction of tissue regeneration: engineering cell-invasion characteristics, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5413–5418.

Maniotis, A.J., Chen, C.S., Ingber, D.E. (1997) Demonstration of mechanical connections between integrins, cytoskeletal filaments, and nucleoplasm that stabilize nuclear structure, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 849–854.

Mauchamp, J., Mirrione, A., Alquier, C., Andre, F. (1998) Follicle-like structure and polarized monolayer: role of the extracellular matrix on thyroid cell organization in primary culture, *Biol. Cell.* **90**, 369–380.

Meier, L., Hay, E.D. (1975) Stimulation of corneal differentiation by interaction between cell surface and extracellular matrix. I. Morphometric analysis of transfilter "induction", *J. Cell Biol.* **66**, 275–291.

Michalopoulos, G., Pitot, H.C. (1975) Primary culture of parenchymal liver cells on collagen membranes. Morphological and biochemical observations, *Exp. Cell Res.* **94**, 70–78.

Mooney, D.J., Langer, R., Ingber, D.E. (1995) Cytoskeletal filament assembly and the control of cell spreading and function by extracellular matrix, *J. Cell Sci.* **108**(Pt 6), 2311–2320.

Mooney, D.J., Hansen, L.K., Langer, R., Vacanti, J.P., Ingber, D.E. (1994) Extracellular matrix controls tubulin monomer levels in hepatocytes by regulating protein turnover, *Mol. Biol. Cell* **5**, 1281–1288.

Mudera, V.C., Pleass, R., Eastwood, M., Tarnuzzer, R., Schultz, G., Khaw, P., McGrouther, D.A., Brown, R.A. (2000) Molecular responses of human dermal fibroblasts to dual cues: contact guidance and mechanical load, *Cell Motil. Cytoskeleton.* **45**, 1–9.

Muschler, J., Levy, D., Boudreau, R., Henry, M., Campbell, K., Bissell, M.J. (2002) A role for dystroglycan in epithelial polarization: loss of function in breast tumor cells, *Cancer Res.* **62**, 7102–7109.

Muthuswamy, S.K., Li, D., Lelièvre, S., Bissell, M.J., Brugge, J.S. (2001) ErbB2, but not ErbB1, reinitiates proliferation and induces luminal repopulation in epithelial acini, *Nat. Cell Biol.* **3**, 785–792.

Myers, C.A., Schmidhauser, C., Mellentin-Michelotti, J., Fragoso, G., Roskelley, C.D.,

Casperson, G., Mossi, R., Pujuguet, P., Hager, G., Bissell, M.J. (1998) Characterization of BCE-1, a transcriptional enhancer regulated by prolactin and extracellular matrix and modulated by the state of histone acetylation, *Mol. Cell. Biol.* **18**, 2184–2195.

Nakamura, K., Kuga, H., Morisaki, T., Baba, E., Sato, N., Mizumoto, K., Sueishi, K., Tanaka, M., Katano, M. (2002). Simulated microgravity culture system for a 3-D carcinoma tissue model. *Biotechniques* **33**, 1068–1070, 1072, 1074–1076.

Niece, K.L., Hartgerink, J.D., Donners, J.J., Stupp, S.I. (2003) Self-assembly combining two bioactive peptide-amphiphile molecules into nanofibers by electrostatic attraction, *J. Am. Chem. Soc.* **125**, 7146–7147.

Nielsen, H.L., Ronnov-Jessen, L., Villadsen, R., Petersen, O.W. (2002) Identification of EPSTI1, a novel gene induced by epithelial-stromal interaction in human breast cancer, *Genomics* **79**, 703–710.

O'Brien, L.E., Jou, T.S., Pollack, A.L., Zhang, Q., Hansen, S.H., Yurchenco, P., Mostov, K.E. (2001) Rac1 orientates epithelial apical polarity through effects on basolateral laminin assembly, *Nat. Cell Biol.* **3**, 831–838.

Oliver, C., Waters, J.F., Tolbert, C.L., Kleinman, H.K. (1987) Growth of exocrine acinar cells on a reconstituted basement membrane gel, *In Vitro Cell. Dev. Biol.* **23**, 465–473.

Petersen, O.W., Ronnov-Jessen, L., Howlett, A.R., Bissell, M.J. (1992) Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9064–9068.

Petka, W.A., Harden, J.L., McGrath, K.P., Wirtz, D., Tirrell, D.A. (1998) Reversible hydrogels from self-assembling artificial proteins, *Science* **281**, 389–392.

Plachot, C., Lelièvre, S.A. (2004) DNA methylation control of tissue polarity and cellular differentiation in the mammary epithelium, *Exp. Cell Res.* **298**, 122–132.

Powers, M.J., Janigian, D.M., Wack, K.E., Baker, C.S., Beer Stolz, D., Griffith, L.G. (2002) Functional behavior of primary rat liver cells in a three-dimensional perfused microarray bioreactor, *Tissue Eng.* **8**, 499–513.

Pujuguet, P., Radisky, D., Levy, D., Lacza, C., Bissell, M.J. (2001) Trichostatin A inhibits beta-casein expression in mammary epithelial cells, *J. Cell. Biochem.* **83**, 660–670.

Reginato, M.J., Mills, K.R., Paulus, J.K., Lynch, D.K., Sgroi, D.C., Debnath, J., Muthuswamy, S.K., Brugge, J.S. (2003) Integrins and EGFR coordinately regulate the pro-apoptotic protein Bim to prevent anoikis, *Nat. Cell Biol.* **5**, 733–740.

Reid, L.M., Rojkind, M. (1979) New techniques for culturing differentiated cells: reconstituted basement membrane rafts, *Meth. Enzymol.* **58**, 263–278.

Roeder, B.A., Kokini, K., Sturgis, J.E., Robinson, J.P., Voytik-Harbin, S.L. (2002) Tensile mechanical properties of three-dimensional type I collagen extracellular matrices with varied microstructure, *J. Biomech. Eng.* **124**, 214–222.

Ronnov-Jessen, L., Petersen, O.W., Koteliansky, V.E., Bissell, M.J. (1995) The origin of the myofibroblasts in breast cancer. Recapitulation of tumor environment in culture unravels diversity and implicates converted fibroblasts and recruited smooth muscle cells, *J. Clin. Invest.* **95**, 859–873.

Roskelley, C.D., Desprez, P.Y., Bissell, M.J. (1994) Extracellular matrix-dependent tissue-specific gene expression in mammary epithelial cells requires both physical and biochemical signal transduction, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12378–12382.

Runswick, S.K., O'Hare, M.J., Jones, L., Streuli, C.H., Garrod, D.R. (2001) Desmosomal adhesion regulates epithelial morphogenesis and cell positioning, *Nat. Cell Biol.* **3**, 823–830.

Sakamoto, T., Hirano, K., Morishima, Y., Masuyama, K., Ishii, Y., Nomura, A., Uchida, Y., Ohtsuka, M., Sekizawa, K. (2001) Maintenance of the differentiated type II cell characteristics by culture on an acellular human amnion membrane, *In Vitro Cell. Dev. Biol. Anim.* **37**, 471–479.

Sakiyama-Elbert, S.E., Panitch, A., Hubbell, J.A. (2001) Development of growth factor fusion proteins for cell-triggered drug delivery, *FASEB J.* **15**, 1300–1302.

Sakurai, H., Barros, E.J., Tsukamoto, T., Barasch, J., Nigam, S.K. (1997) An in vitro tubulogenesis system using cell lines derived from the embryonic kidney shows dependence on multiple soluble growth factors, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6279–6284.

Sanderson, I.R., Ezzell, R.M., Kedinger, M., Erlanger, M., Xu, Z.X., Pringault, E., Leon-Robine, S., Louvard, D., Walker, W.A. (1996) Human fetal enterocytes in vitro: modulation of the phenotype by extracellular matrix, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 7717–7722.

Sasaki, T., Takagi, M., Soma, T., Yoshida, T. (2002) 3D culture of murine hematopoietic cells with spatial development of stromal cells in nonwoven fabrics, *Cytotherapy* **4**, 285–291.

Schafer, K.H., Hagl, C.I., Rauch, U. (2003) Differentiation of neurospheres from the enteric nervous system, *Pediatr. Surg. Int.* **19**, 340–344.

Schmidhauser, C., Casperson, G.F., Bissell, M.J. (1994) Transcriptional activation by viral enhancers: critical dependence on extracellular matrix-cell interactions in mammary epithelial cells, *Mol. Carcinog.* **10**, 66–71.

Schmidhauser, C., Casperson, G.F., Myers, C.A., Sanzo, K.T., Bolten, S., Bissell, M.J. (1992) A novel transcriptional enhancer is involved in the prolactin- and extracellular matrix-dependent regulation of beta-casein gene expression, *Mol. Biol. Cell* **3**, 699–709.

Schuger, L., O'Shea, K.S., Nelson, B.B., Varani, J. (1990) Organotypic arrangement of mouse embryonic lung cells on a basement membrane extract: involvement of laminin, *Development* **110**, 1091–1099.

Schuger, L., Yurchenco, P., Relan, N.K., Yang, Y. (1998) Laminin fragment E4 inhibition studies: basement membrane assembly and embryonic lung epithelial cell polarization requires laminin polymerization, *Int. J. Dev. Biol.* **42**, 217–220.

Seid, C.A., Ramachandran, R.K., George, J.M., Govindarajan, V., Gonzalez-Rimbau, M.F., Flytzanis, C.N., Tomlinson, C.R. (1997) An extracellular matrix response element in the promoter of the LpS1 genes of the sea urchin *Lytechinus pictus*, *Nucleic Acids Res.* **25**, 3175–3182.

Semino, C.E., Merok, J.R., Crane, G.G., Panagiotakos, G., Zhang, S. (2003) Functional differentiation of hepatocyte-like spheroid structures from putative liver progenitor cells in three-dimensional peptide scaffolds, *Differentiation* **71**, 262–270.

Seton-Rogers, S.E., Lu, Y., Hines, L.M., Koundinya, M., LaBaer, J., Muthuswamy, S.K., Brugge, J.S. (2004) Cooperation of the ErbB2 receptor and transforming growth factor beta in induction of migration and invasion in mammary epithelial cells, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1257–1262.

Silva, G.A., Czeisler, C., Niece, K.L., Beniash, E., Harrington, D.A., Kessler, J.A., Stupp, S.I. (2004) Selective differentiation of neural progenitor cells by high-epitope density nanofibers, *Science* **303**, 1352–1355.

Simlan M., Hiral Y., Navre M., Werb Z., Lochter A., Bissell M.J. (2001) The Interplay of matrix metalloproteinases, morphogens and growth factors is necessary for branching of mammary epithellal cells, *Development* **128**, 3117–3131.

Singhvi, R., Kumar, A., Lopez, G.P., Stephanopoulos, G.N., Wang, D.I., Whitesides, G.M., Ingber, D.E. (1994) Engineering cell shape and function, *Science* **264**, 696–698.

Stark, H.J., Szabowski, A., Fusenig, N.E., Maas-Szabowski, N. (2004) Organotypic cocultures as skin equivalents: a complex and sophisticated in vitro system, *Biol. Proced. Online* **6**, 55–60.

Sternlicht, M.D., Kedeshian, P., Shao, Z.M., Safarians, S., Barsky, S.H. (1997) The human myoepithelial cell is a natural tumor suppressor, *Clin. Cancer Res.* **3**, 1949–1958.

Streuli, C.H., Bailey, N., Bissell, M.J. (1991) Control of mammary epithelial differentiation: basement membrane induces tissue-specific gene expression in the absence of cell-cell interaction and morphological polarity, *J. Cell Biol.* **115**, 1383–1395.

Streuli, C.H., Schmidhauser, C., Kobrin, M., Bissell, M.J., Derynck, R. (1993) Extracellular matrix regulates expression of the TGF-beta 1 gene, *J. Cell Biol.* **120**, 253–260.

Streuli, C.H., Schmidhauser, C., Bailey, N., Yurchenco, P., Skubitz, A.P., Roskelley, C., Bissell, M.J. (1995) Laminin mediates tissue-specific gene expression in mammary epithelia, *J. Cell Biol.* **129**, 591–603.

Suckow, M.A., Voytik-Harbin, S.L., Terril, L.A., Badylak, S.F. (1999) Enhanced bone regeneration using porcine small intestinal submucosa, *J. Invest. Surg.* **12**, 277–287.

Szabowski, A., Maas-Szabowski, N., Andrecht, S., Kolbus, A., Schorpp-Kistner, M., Fusenig, N.E., Angel, P. (2000) c-Jun and JunB antagonistically control cytokine-regulated mesenchymal-epidermal interaction in skin, *Cell* **103**, 745–755.

Takahashi, K., Kawahara, S., Ono, T. (1990) Effects of growth factors and hormones on

growth and morphological differentiation of human breast epithelial cells within collagen gel in serum-free medium, *Jpn. J. Cancer Res.* **81**, 52–57.

Tan, W., Desai, T.A. (2005) Microscale multilayer cocultures for biomimetic blood vessels, *J. Biomed. Mater. Res.* **72**, 146–160.

Vassy, J., Portet, S., Beil, M., Millot, G., Fauvel-Lafeve, F., Gasset, G., Schoevaert, D. (2003) Weightlessness acts on human breast cancer cell line MCF-7, *Adv. Space Res.* **32**, 1595–1603.

Vassy, J., Portet, S., Beil, M., Millot, G., Fauvel-Lafeve, F., Karniguian, A., Gasset, G., Irinopoulou, T., Calvo, F., Rigaut, J.P., et al. (2001) The effect of weightlessness on cytoskeleton architecture and proliferation of human breast cancer cell line MCF-7, *FASEB J.* **15**, 1104–1106.

Vukicevic, S., Luyten, F.P., Kleinman, H.K., Reddi, A.H. (1990) Differentiation of canalicular cell processes in bone cells by basement membrane matrix components: regulation by discrete domains of laminin, *Cell* **63**, 437–445.

Wang, N., Ingber, D.E. (1994) Control of cytoskeletal mechanics by extracellular matrix, cell shape, and mechanical tension, *Biophys. J.* **66**, 2181–2189.

Wang, F., Weaver, V.M., Petersen, O.W., Larabell, C.A., Dedhar, S., Briand, P., Lupu, R., Bissell, M.J. (1998) Reciprocal interactions between beta1-integrin and epidermal growth factor receptor in three-dimensional basement membrane breast cultures: a different perspective in epithelial biology, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14821–14826.

Weaver, V.M., Petersen, O.W., Wang, F., Larabell, C.A., Briand, P., Damsky, C., Bissell, M.J. (1997) Reversion of the malignant phenotype of human breast cells in three-dimensional culture and in vivo by integrin blocking antibodies, *J. Cell Biol.* **137**, 231–245.

Weaver, V.M., Lelièvre, S., Lakins, J.N., Chrenek, M.A., Jones, J.C., Giancotti, F., Werb, Z., Bissell, M.J. (2002) Beta4 integrin-dependent formation of polarized three-dimensional architecture confers resistance to apoptosis in normal and malignant mammary epithelium, *Cancer Cell* **2**, 205–216.

Yamada, K.M., Clark, K. (2002) Cell biology: survival in three dimensions, *Nature* **419**, 790–791.

Zhang, S. (2003) Fabrication of novel biomaterials through molecular self-assembly, *Nat. Biotechnol.* **21**, 1171–1178.

Tissue-Like Cell Culture: *see* Three Dimensional Cell Culture: The Importance of Microenvironment in Regulation of Function

Total Analysis Systems, Micro

Shigeori Takenaka
Kyushu Institute of Technology, Kitakyushu-shi, Japan

1 Introduction 422

2 Principle of the μ-TAS 423
2.1 Microfluidic Device System 423
2.2 DNA Chip 426

3 Applications 428
3.1 Combination with Polymerase Chain Reaction (PCR) 428
3.2 Integration of an Electrochemical Detector on the Device 429
3.3 Electrochemical DNA Chip 430
3.4 Combination of a DNA Chip with a Microfluidic Device 431
3.5 Application in Protein Analysis 432

4 Perspectives 433

Bibliography 434
Books and Reviews 434
Primary Literature 434

Keywords

Capillary Gel Electrophoresis
Electrophoretic separation of nucleic acids, proteins, and so on in the gel-packed capillary typically of 100 μm internal diameter and 50 to 100 cm in length under high voltage (1–2 kV).

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

DNA Probe
DNA fragment carrying a suitable labeling marker that can hybridize with the target DNA.

Hybridization
To form double-stranded DNA between respective complementary single-stranded DNA and/or RNA by heating and cooling.

Polymerase Chain Reaction (PCR)
Amplification method of DNA with two primer sets by repeating denaturation, annealing, and extension with a heat-stable polymerase.

Single Nucleotide Polymorphism (SNP)
One nucleotide variation located on a chromosomal DNA in individuals, which is observed in the population at a frequency greater than 1%.

The investigations concerned with downsizing and integration of chemical experiments and processes on a chip attract great attention. These researches have not only brought about integration or miniaturization of whole chemical processes to the scale of a semiconductor integrated circuit but also provided several advantages such as the reduction of chemicals, waste, reaction space, and reaction and measurement time. These devices generally exploit micro size fluidic lines on the flat surface made of glass or plastic. The device thus obtained is called a *microfluidic device* and the system is called μTAS (Micro Total Analysis Systems) or Lab-on-a-Chip. Such devices have been evolving in the biotechnology area, especially as a gel electrophoretic DNA separation module for DNA sequencing. On the other hand, the research fields connected with the integration of many different DNA sequences on a small space also have been actively explored to provide a tool for high-throughput analysis of events associated with genes. This system is called a *DNA chip* and helps to collect genetic information closely associated with many genetic diseases. A DNA chip is expected to provide a future gene diagnosis method, since this research combined with the microfluidic system has been studied to develop a rapid pretreatment method of the sample. An electrochemical detecting method is more suitable for such a miniature system than the conventional fluorescent one currently in use.

1 Introduction

The Lab-on-a-Chip analysis technology is based on an idea that allows one to duplicate many of the experimental procedures normally carried out in a flask or beaker in the chemical laboratory on a small substrate or chip. The advantage of doing everything in a tiny space lies

in the fact that the time and the quantity of reagents required for analysis can be reduced drastically. In addition, the whole system may be automated easily by integrating complicated analytical procedures in the form of a circuit. To date, significant progress has been seen and a practical Lab-on-a-Chip analysis technology is now available for biochemical analysis and medical diagnosis. One of the large tasks in biochemical research is how to separate, identify, and quantify biomacromolecules such as nucleic acids and proteins efficiently. Analysis of their interactions with each other and small molecules simultaneously is another task in the biochemical and medical areas. Microfluidic devices carrying microdrains or microchannel on the chip made of silicon, glass, or plastic have been developed for the former purpose of the Lab-on-a-Chip. The latter goal was attained by microfluidics and a unit was furnished in this device; and biomolecules were detected by this unit. The separation in this system is commonly conducted by electrophoresis in the microfluidic device. Many researches concerned with electrophoretic separation of DNA especially have been reported and one of them is used for DNA sequencing on the chip. The latter example is represented by a DNA chip or protein chip. This system enables simultaneous analysis of massive amounts of samples at a time or high-throughput analysis of biomolecules, where one can estimate the specific interaction of the target biomolecules through their amount held on the reactants arranged as many matrix-patterned dots on a small substrate. This technology has been used not only in basic studies such as DNA expression analysis but also in drug discovery. Although this technology is useful and important, it contains many manual procedures such as the introduction step of the sample. DNA chips or protein chips furnished with whole procedures are becoming more and more important in the diagnostic area and an assembly of total analysis of biochip has been tried. In the Lab-on-a-Chip analysis technology, highly sensitive detection is required because of limited availability of biomolecules. Fluorescence markers are often used to label biomolecules directly or indirectly to ease analysis. For the trial of a self-contained DNA chip or protein chip, an electrochemical detector was devised to integrate all the components in these chip substrates.

Here, recent development of microfluidic devices and DNA chips are summarized as a Lab-on-a-Chip analysis technology. The integration of PCR for DNA amplification on the chip or the integration of an electrochemical detector discusses the former topic as realization of a self-contained chip system. In the latter topic, the basic principle of the DNA chip is reviewed with an emphasis on recent development in related areas. Combination of microfluidic parts with a DNA chip should enable pretreatment on the chip resulting in a self-contained DNA analyzing system.

2 Principle of the μ-TAS

Here, the principle of an electrophoretic separation system of a microfluidic device is described as a typical example and an outline of the DNA chip is also described chronologically.

2.1 Microfluidic Device System

Separation and analysis with a microfluidic device were achieved on a drain or

fluidic channel of about 100 μm width on the substrate made of silicon, glass, or plastic. At first, silicon was mainly used as an extension of the photolithographic technique commonly used to prepare integrated circuits in electronics. Since then, glass or plastic have been used from a viewpoint of the construction of an inexpensive disposable system. Poly(dimethylsiloxane) (PDMS) initially used in this system was insufficient to dissipate the Joule heat and acrylic polymers overcame this problem. Figure 1 shows an example of a plastic microfluidic device prepared by injection molding. The drains or microchannels are filled with gel and electrolyte and DNA is separated according to its size by electrophoresis like in capillary gel electrophoresis. The linear polyacrylamide has been used as the filling gel in this system but nanosized particles in place of such a polymer enable better separation. Figure 2 shows the cross-type microfluidic channels carrying circular wells called a *reservoir* at each terminus. A typical separation procedure is explained with this system. The sieving material such as polyacrylamide and electrolyte are filled over the microchannel. The locations of the circular wells are shown in Fig. 2A(a–d). The DNA sample is injected into the well as shown in Fig. 2A and electrodes are dipped in (a) and (b), and a high voltage is applied between (a) and (b) to transfer the sample to the channel between (a) and (b) by an electroosmotic and electrochemically driven flow. When the sample reached and filled the position of the cross-type channel on the microfluidic device (Fig. 2B), the electrodes are switched to wells (c) and (d),

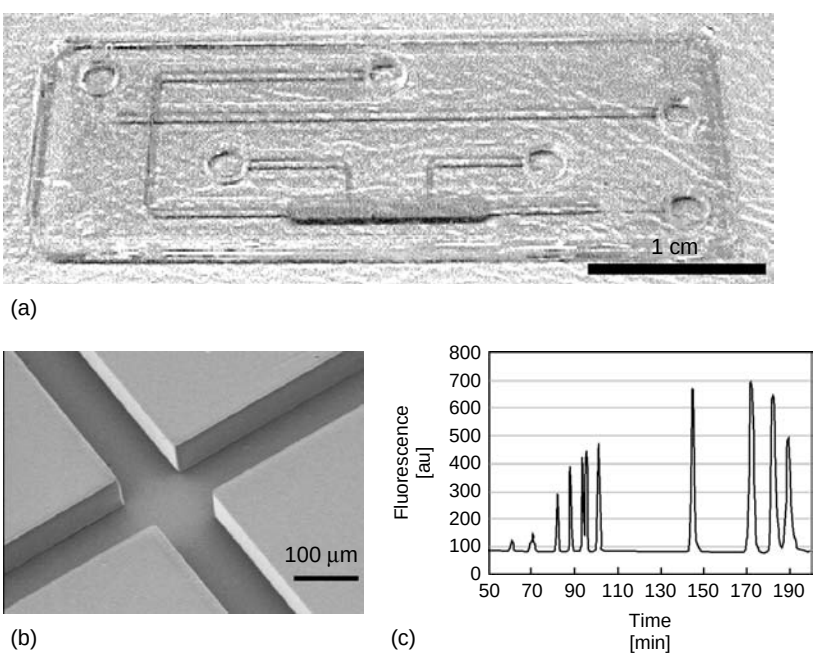

Fig. 1 (a) Example of a microfluidic device made of polymethylmetacrylate (PMMA) by injection molding (Enplas Laboratories, Inc., Saitama, Japan) 100 μm in width and 50 μm in depth (b) and separation of Hae III digestion fragments of. φX-174 DNA (c), which was stained with ethidium bromide and monitored by its fluorescence intensity.

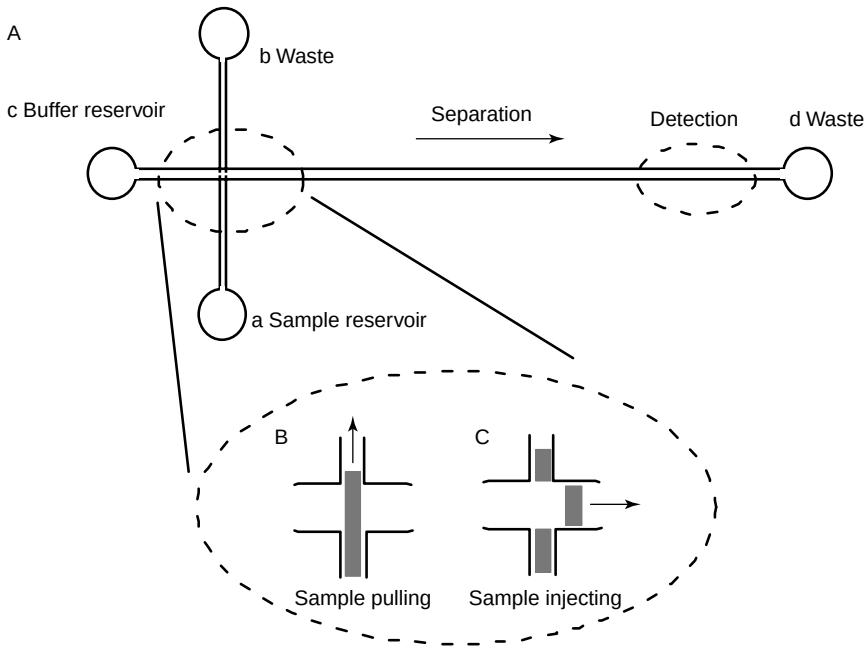

Fig. 2 Separation principle of a gel-filled microfluidic device for gel electrophoretic DNA separation.

and a high voltage is applied between (c) and (d) (Fig. 2C). The sample occupying the cross section is cut off and transferred to the fluidics line between (c) and (d). Continuous application of the voltage between (c) and (d) separates the sample in the separation channel, and the separated analyte is detected near site (d) usually by the fluorescence signal of labeled analyte emitted upon excitation with a laser. Such a type of microfluidic device allows introduction of an extremely small volume (100 pl in the case of 100-μm width and 20-μm depth) for each sample in the cross section of the fluidics line with high reproducibility. Theoretical investigation revealed that the diffusion rate constant of analytes in the microchannel increases with an increase in the strength of electric field but the separation efficiency is not improved. Nevertheless, this microfluidic device can effectively dissipate the heat (in Joule), produced by applying a high voltage during electrophoresis, and shorten the separation time without band broadening by the applied higher voltage from that in the ordinary capillary gel electrophoretic separation system. An optimized microfluidic system gave a 10 to 100 times shorter separation time than the conventional slab or capillary gel electrophoresis system. The microfluidic device was also applied to gene type testing and achieved ultrafast allelic profiling of short tandem repeats (STRs). This typing was successfully achieved by a microchannel that was 2.6 cm in length in 30 s, as compared with 2 to 3 h of slab gel electrophoresis and 10 to 30 min of capillary gel electrophoresis. Recently, some companies launched a microfluidic device for the separation of biomolecules. For example, a glass type microfluidic device is commercially available from Cariper Technologies Corporation

and Agilent Technologies, and a plastic one from Aclare Biosciences Inc.

DNA sequencing is another application of the microfluidic device. The advantages of this system are effective cooling of the Joule heat generated by applying a high voltage in DNA sequencing because of a large surface area compared with the inner solution of the channel and integration of many channels on the device. A DNA sequencing system furnished with 96 channels on a silicon wafer of 10 cm diameter was reported by Mathies's group. A rotary confocal scanner suitable for this device was also developed by the same group, which enabled analysis of one sample within 90 s. A raw DNA sequencing rate >150 kb h^{-1} per microchannel was achieved, which is much faster than that attained with the conventional DNA sequencing.

The electrophoretic separation in the microfluidic device was carried out by the molecular sieve mechanism of polymer molecules in the microchannel. This separation was tested by microfabrication of many nanosize pillars in the microchannel. Such a type of device was realized by constructing many pillars of 30 nm diameter 500 nm apart from each other on the microchannel. When DNA goes through the pillars in the microchannel, the pillars serve as obstacles to retard DNA migration. The retardation time depends on its molecular size. Nanopillars that were 2700 nm in length, 500 nm in width, and 500 nm in spacing (with a high aspect ratio of height/width = 5.4) in a microchannel on quartz have been fabricated by the electron beam lithography technique and used for DNA separation. DNA fragments carrying a range from 1 to 38 Kb were separated as clear bands at the detection window of 1450 μm from the entrance of the nanopillar channel (25 μm in width and 2.7 μm in height) in 170 s. The theoretical plate number of the channels (380–1450 μm in length) is as high as 1000 to 3000 μm ($0.7–2.1 \times 10^6$ plates m^{-1}).

2.2 DNA Chip

DNA chips have become an indispensable tool in genome exploration, disease diagnosis, drug discovery, and toxicological research. The DNA chip is a small substrate carrying different kinds of single-stranded DNA fragments immobilized on its specified position as a probe. Where DNA sequences complementary to the probe are present in the sample, a double-stranded DNA is formed on the substrate surface by a process called *hybridization*. The resulting hybrid DNA is detected somehow, and to facilitate analysis sample, DNA is labeled usually with a fluorescence dye prior to hybridization. A typical experiment using a DNA chip is shown in Fig. 3. Many different kinds of DNA fragments are immobilized in the specified position on the substrate as a capture probe. After labeling sample DNA with a fluorescence dye, it is allowed to hybridize with the capture probe on the chip. When target DNA exists in the sample, a DNA duplex is formed on the spot carrying the complementary DNA sequence. After washing the chip, the hybrid DNA is visualized by surveying by confocal microscopy. The intensity of labeling dye on the chip is proportional to the efficiency of hybridization or the amount of the specific DNA fragment. Comparative gene expression experiments are a typical example of the application of the DNA chip, in which two different kinds of cells, for example, normal and diseased, are studied.

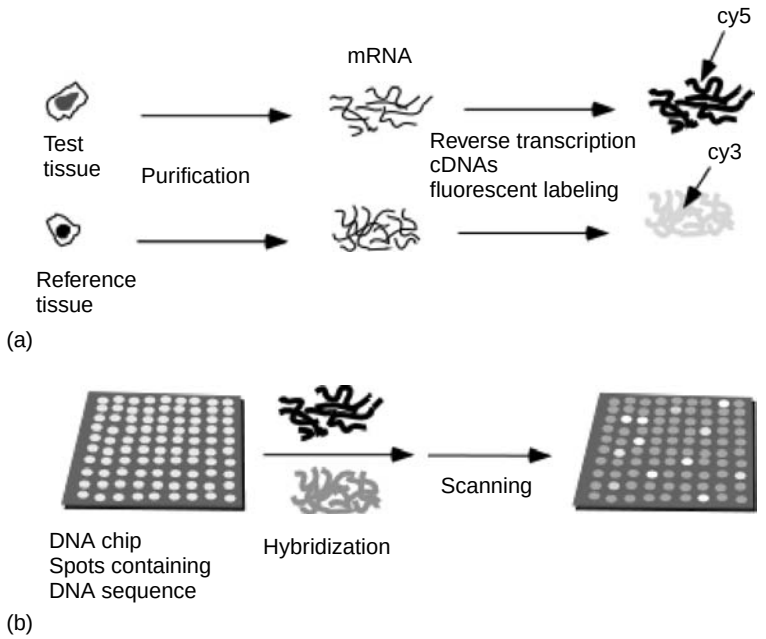

Fig. 3 Comparative gene expression analysis as an example of typical DNA chip experiments.

The purpose of this experiment is to quantify mRNA in the specific cell. In the first step, the extracted mRNAs are converted to cDNA by reverse transcriptase. cDNAs obtained from different cells are labeled with different colored fluorescence dyes, cyanine derivatives, Cy3 and Cy5, being used most often. cDNAs prepared from equal amounts of the cells are allowed to hybridize on the DNA chip. After hybridization and washing of the DNA chip, fluorescence images are taken by scanning. Information of gene expression in the cell can be obtained from the position and intensity of the fluorescence spot. Since the concept of DNA chip technology was conceived by Southern and coworkers in the early 1990s, many technologies have been devised with a focus on the immobilization method of capture DNA probe on the substrate. Two methods are mainly adopted to attach a capture probe on the substrate such as glass. The first approach of DNA chip preparation is direct DNA fragment (oligonucleotide) synthesis on the substrate by the computer-generated photolithographic mask, which defines the exposure site of the surface. In this approach, the light-directed synthesis of oligonucleotides is carried out combinatorially and all of the possible DNA sequences can be constructed on the surface with a limited number of synthesis. For example, the synthesis times of 4×7 give 4^7 kinds of DNA sequences. To reduce the cost associated with the photolithographic mask, *in situ* synthesis of oligonucleotides on the surface has been developed without relying on the photolithographic mask. The second approach to array fabrication is off-chip synthesis, which is more flexible and economical than the on-chip one.

The synthesized or extracted DNA fragment is spotted and covalently attached on the surface. The DNA chip produced by on-chip synthesis is commercially available from Affymetrix Inc. (Santa Clara, CA). Although the DNA chip technology is powerful, many manual procedures are involved such as the preparation of capture probes on the DNA chip surface and labeling of sample DNA with a fluorescent dye. To improve the performance of the DNA chip technology, especially for clinical disease diagnosis, it is important to upgrade the DNA chip technology with respect to cost, speed, and robustness, and also a totally self-contained gene analysis system needs to be assembled.

Recently, DNA chip technologies have been extended to proteome analysis, for example, in the form of a protein chip. Protein chips developed to date are categorized by antibody chip, peptide chip, protein chip, bead, or particle chip, which can be used as labeled particles in solution and specialized surface chemistry to perform combinatorial chemistry. The advantages conferred by its miniaturization might even be greater than those of the DNA chip systems.

3 Applications

3.1 Combination with Polymerase Chain Reaction (PCR)

It is necessary to concentrate the DNA as analyte on the microfluidic device to apply for DNA diagnosis. Polymerase chain reaction (PCR) is a useful technique suited for this purpose. A microfluidic device carrying a PCR unit has been developed from this point of view. PCR is a gene amplification reaction, where heat denaturation into the single-stranded form of DNA and enzymatic extension reaction to generate a double-stranded form are repeated several dozen times. A high-speed temperature changing device is necessary for PCR. Miniaturization of the reaction cell can improve the thermal exchange rate by increasing the ratio of the device surface to the reaction solution volume. When a heating and cooling system is introduced in a reservoir as shown in Fig. 2(a), the PCR product amplified there can be separated by the separation system on the microfluidic device described in Fig. 2. Ultrashort thermal cycling times of 17 s per cycle was achieved by infrared (IR) radiation-mediated heating and compressed air cooling of amplified DNA in capillary or on-chip microchambers was integrated onto microfluidic devices. Precise temperature control was achieved by controlling the light intensity by a system furnished with a tungsten lamp as an IR radiation source, thermocouple feedback computer interface coupled with an airstream.

A PCR system based on a microchannel was reported, where the microchannel crosses three distinct temperature zones of thermostatted copper blocks as shown in Fig. 4. As the PCR cocktail is pumped through the channel, it flows between the regions of different temperatures defined by three parts under the channel. The melting, annealing, and extension steps are executed by passing through the different temperature zones. In this system, the PCR cocktail is flowed to the different temperature channel instead of a temperature change in the same position of the microreactor. The channel incorporates 20 cycles giving a theoretical amplification of 2^{20}. A flow rate that provided a total cycling time of 10 min gave

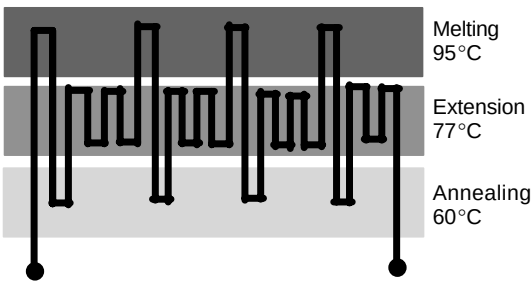

Fig. 4 PCR based on the microchannel. Three well-defined zones are maintained at 95, 77, and 60 °C by means of thermostatted copper blocks. The PCR cocktail is hydrostatically pumped through a single channel etched into a glass plate. The melting, extension, and annealing steps are defined by the channel passing through the three temperature zones.

70% of the yield of a fast commercial PCR themocycler in which the equivalent cycles took 50 min. A fivefold improvement in time was seen with little product degradation, free from the thermal inertia effect. Furthermore, the device allows multiple PCRs to run at a time on the same device, for example, by introducing a separate reaction loop.

3.2 Integration of an Electrochemical Detector on the Device

Fluorometric methods are commonly used for the detection of electrophoretically separated biomolecules such as DNA fragments on the microfluidic device. Following labeling of a sample directly (through a covalent bond) or indirectly (staining with a fluorescence dye), the fluorescence is detected by laser-induced fluorescence spectroscopy. However, it is difficult to incorporate a laser-induced fluorescence detector into this system to obtain a self-contained miniaturized and portable device. An electrochemical detection system would be an alternative to this and enables the development of a self-contained miniaturized analysis system with high sensitivity. Furthermore, both the detector and instrument may be miniaturized and the optical path can be omitted. In addition, even turbid sample solutions may be amenable to analysis. In summary, an inexpensive, low-power, and high-compatibility device can be constructed by advanced micromachining and microfabrication. To enable electrochemical detection of electrophoretically separated samples in the microfluidic device, the electrodes for detection (working, counter, and reference electrodes in the ordinary case) need to be incorporated into the outlet reservoir (Fig. 2A(d)). In this case, it is critical to position the detection electrode to avoid high potential at the separation microchannel (between c and d in Fig. 2A). Mathies and coworkers first reported an electrophoretic microfluidic device with an integrated amperometric detector where a dual-lead, platinum-working electrode 10 μm in width, is positioned just 30 μm beyond the end of the separation channel in the reservoir by the photolithographic technique. This is important, as this detection electrode finds itself just outside the high electric field applied for separation, and is effectively shielded from this field as a result. To place the working electrode just outside the exit of the separation channel, several types of electrochemical detectors were proposed by different types of arrangement between the microchannel and working electrode and the position of electrodes relative to flow direction. The distance between the separation channel outlet and working electrode affects the band broadening. When the working electrode is close to the separation channel, the

electrochemical response may be observed without band broadening, but may be less well shielded from the separation voltage. These detectors have mainly relied on microfabricated working electrodes as a thin or thick film, with the reference and counterelectrodes to complete the potentiostatic circuit, placed in the detector reservoir. Practical application of the electrophoretic separation based on a microfluidic device furnished with an electrochemical detector was reported not only for nucleic acids but also for neurotransmitters such as catecholamine, nitroaromatic explosives, organophosphate pesticides, or endocrine-disturbing chemicals such as chlorophenol derivatives. A complete self-contained microfluidic device will be realized through an on-device integration of the potentiostatic circuit and other functional elements such as preconcentrator.

3.3 Electrochemical DNA Chip

Incorporation of an electrochemical detector to the microfluidic device for the laser-induced fluorescence detector should yield a true Lab-on-a-Chip analysis device. The current DNA chips require a large-sized detecting system of a laser-based fluorescence scanner. Therefore, when an electrochemical detector is replaced for the fluorescence detector of a DNA chip, this also realizes a true DNA chip resulting in a portable gene diagnosis chip. The research along this line is still in its infancy. Nanogen Inc. (San Diego, CA) constructed a variety of microelectronic chips including 25–10 000 addressable electrodes and has developed an electronic DNA chip preparation process. Since oligonucleotides and DNA have negative charges at the phosphate backbone, their movement can be electrochemically controlled, leading to the electrochemically accelerated hybridization and electrochemical denaturation on the chip. Controlled electric fields have been used for discriminating oligonucleotide hybrids with varying binding strengths such as those between the completely matched and single-base mismatched and for removing unhybridized DNA. Such fine-tuned electronic stringency selection obviates the need for extensive washing. This approach is now successful in the analysis of single nucleotide polymorphisms (SNPs), STRs, insertions, deletions, and other genetic mutations. The electronically regulated sample preparation process was demonstrated for the dielectrophoretic separation of *Escherichia coli* from blood cells. After the isolation, the bacteria are lysed by a series of high-voltage pulses. Hybridization on this chip is detected fluorometrically. Electrochemical detection of DNA hybridization is a long-sought goal, as it will enable label-free detection of DNA or protein binding as well as curtailing the amount of reagents and processing costs, and amenability to portability and miniaturization.

As an approach to electrochemical detection of DNA hybridization, Motorola Clinical Micro Sensors Inc. (Pasadena, CA) and TUM-gene Inc. (Chiba, Japan) did pioneering work. They aim at a point-of-care diagnostic chip with a small battery-operated instrument. Motorola constructed a multi-electrode chip and introduced a ferrocene-labeled oligonucleotide as an electrochemically active DNA probe, instead of the fluorescence-labeled one. Two DNA probes carrying a complementary target DNA sequence were prepared as a DNA probe on the electrode and a ferrocene-labeled DNA probe. Since sample DNA is allowed to hybridize with these DNA probes, the ferrocene is fixed on the electrode only in the presence of target DNA

in sample DNA in the sandwich assay. The hybridized DNA is detected from the electrochemical response due to the ferrocene. TUM-gene has developed a simultaneous multiple mutation detection (SMMD) method, which enables simultaneous discrimination of several genetic mutations such as SNP, insertion, deletion, translocation, and STR, using a multielectrode DNA chip and a double-stranded nucleotide specific threading intercalator, ferrocenylnaphthalene diamide (FND). In this method, target genome is amplified by the first PCR and then the second asymmetric PCR with a specially designed primer set. The second PCR products are applied on two electrodes containing either a wild-type (Wt) or a mutant type (Mt) probe. After completion of the hybridization and ligation reactions of the PCR products with the probes, each electrode is washed and immersed in an electrolyte containing FND, and the electric current at specified potential for FND is measured. Genotype (Wt/Wt, Wt/Mt, Mt/Mt) were determined by calculating the ratio of current signals from the Wt probe and Mt probe. The validity of the SMMD method was examined for lipoprotein lipase gene mutation by blind testing 50 patients' samples including 2 Mt/Mt homozygotes and 6 Wt/Mt heterozygotes. As expected, all of samples of human lipoprotein lipase (LPL) mutations, which are connected with hypertriglyceridemia or high neutral lipid concentration in the blood, were identified by this method correctly, indicating that the SMMD method can be used reliably for screening of the genetic mutations for routine clinical patient samples. Both techniques succeeded in detecting hybridization electrochemically with a DNA chip, which will eventually turn to a self-contained, miniaturized, and portable device. However, it is noted that time-consuming manual procedures such as extraction of DNA from the sample and PCR are still involved.

3.4 Combination of a DNA Chip with a Microfluidic Device

Microfluidic devices capable of carrying out sample pretreatment, PCR, and detection on a DNA chip are needed for automated clinical diagnosis. Such a system has not yet been completed, but various attempts have been made. Fusion of microfluidics and DNA chips will be useful with the advantages for both sides. Microfluidic channels can be utilized as an autotransport system to deliver controlled volumes of sample and reagents in a DNA chip. Nanogen developed a way to regulate hybridization electrochemically as discussed earlier and DNA sample and reagents are introduced to the DNA chip through fluidic networks made by injection molding. The automatic, real-time monitoring of hybridization of target DNA with the capture DNA fixed on the DNA chip will be achieved by the combination of precise temperature and fluidic control. This will in turn enable automatic selection of assay parameters such as buffer species and the precise volume and temperature. Finally, this microfluidic device and DNA chip fusion system will automatically select optimal parameters for hybridization on the chip. Furthermore, this fusion system will allow kinetic monitoring of the hybridization reaction on the chip.

Affymetrix Inc. reported for the first time a highly integrated monolithic device made of polycarbonate for automated multistep genetic assays. The detector of this device is a DNA chip made by photolithography and the overall device dimensions are less than those of a credit card. The device

can accommodate more than 10 reagents for over 60 sequential operations and is capable of extracting and concentrating nucleic acids from milliliter aqueous samples and performing microliter chemical amplification, serial enzymatic reactions, melting, mixing, and nucleic acid hybridization. The performance of the device was tested for mutations in a 1.6-kb region of the HIV genome from serum samples, where starting concentrations were as few as 500 copies of RNA. This device was produced in plastic rather than glass or silicon for economic reasons, since a disposable device is required for these clinical applications. Fluids were moved pneumatically from one chamber to the next in this chip.

Motorola groups have been developing plastic microfluidic devices carrying an electrochemical DNA chip and a micropump. Hybridization kinetics was studied using this chip and they tried to improve the hybridization efficiency by oscillation of hybridization mixture using a micropump. A self-contained, fully integrated device was constructed consisting of microfluidic mixers, valves, pumps, channels, chambers, heater, and DNA chip for sample preparation, PCR amplification, and DNA chip analysis. Instead of incorporating a DNA chip into the microfluidic device, a DNA array was prepared in the microfluidic channel hot-embossed in poly(methyl methacrylate) to allow detection of low-abundant mutations of *K-ras* in gene fragments that carry point mutations of high diagnostic value for colorectal cancers.

The DNA probes carrying a complementary DNA sequence for both sides of the point mutation were prepared: one has a point mutation site and zip-code sequence for being caught on the capture probe on the DNA array and the other has a near IR-dye IRD-880 for highly sensitive detection. When sample DNA has target point mutation, two kinds of DNA probes are joined together enzymatically by a process called an *allele-specific ligase detection reaction (LDR)*. Near IR fluorescence imaging of the DNA microarray on the microfluidic device could reveal the existence of the linked DNA probe on the corresponding zip-code sequence. This system can reduce the hybridization time from 3 h with the conventional DNA chip to less than 1 min and detect a point mutation in the *K-ras* oncogene at a level of 1 mutant DNA in 10 000 wild-type sequences. A disposable polymer microfluidic device holding a 1000-spot DNA array has been manufactured by microinjection molding. This device provides a powerful tool for highly parallel studies of kinetics and thermodynamics of duplex formation on the DNA microarray in the microfluidic channel to show that the hybridization in microfluidic hybridization assays is diffusion-limited due to the small diffusion coefficients of the DNA and RNA molecules involved.

3.5 Application in Protein Analysis

Application of a microfluidic device in protein research is becoming more and more important. At first, the device was used for sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) separation and detection of proteins after staining with a fluorescence dye. This novel technology offers an important tool for immunoassay, clinical diagnostics, biomedical science, and industry. Electrospray mass spectroscopy (ESI-MS) coupled with a microfluidic device especially is useful in the analysis of protein and peptides obtained by protease treatment. In proteome

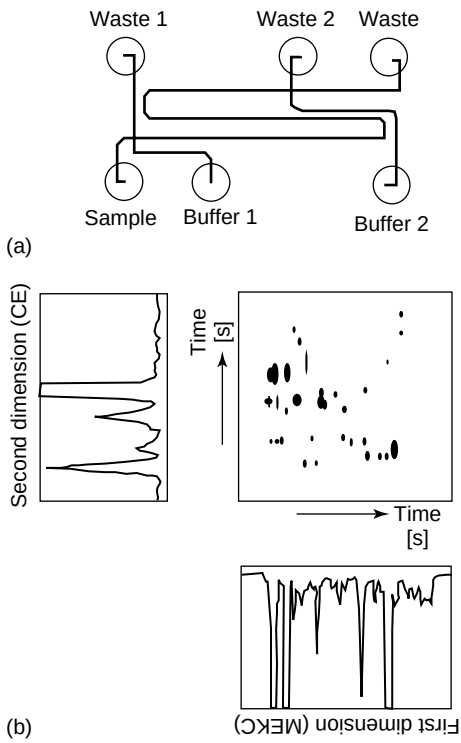

Fig. 5 (a) Illustration of a microfluidic device that enables 2-D electrophoretic separation of labeled peptides by MEKC and CE on a system. The 1-D separation channel extends from the first cross intersection to the second. The 2-D separation channel extends from the second cross to the point of detection. (b) Schematic cartoon of a 2-D separation map, which can be constructed from chromatograms of individual microfluidic separations.

analysis, two-dimensional electrophoresis is a powerful technology for separation of protein. Samples are separated in one dimension first by isoelectric focusing followed by SDS-PAGE separation in the second dimension. In this method, protein is first separated by the difference in isoelectric points and is further separated by the difference in molecular mass in the second dimension. Up to 10 000 different proteins can be separated on a gel, though a typical resolution is of the order of 5000. However, two-dimensional SDS-PAGE is a slow process taking up to two days for complete separation. This problem may be alleviated by a microfluidic system. The micellar electrokinetic chromatography (MEKC) with rapid channel electrophoresis (RCE) was used for the two-dimensional separation of peptides by using the microfluidic device carrying specially ordered microchannel construction as shown in Fig. 5(a). The peptides are separated by MEKC while at the same time the MEKC eluent is rapidly sampled for separation by RCE in this device and detected by laser-induced fluorescence. This device allows rapid sampling and separation, which are not easily achieved by conventional techniques. Figure 5(b) shows a two-dimensional map of separated protein, which can be constructed from the chromatograms from individual microfluidic separations.

4 Perspectives

Microfluidic devices, DNA chips, and the combination of both technologies were described here, focusing mainly on high-throughput analysis of genetic information. In this sense, these technologies are not simple extensions of the existing methods but a totally innovative analytical means suitable for research in the postgenome era. Powerful as they are, the current technologies suffer various weaknesses and they are being improved steadily. The information obtained by these technologies is useful in various areas, in pharmacogenomics, in particular, as it will help clarify the mechanism and side effects of drugs and boost molecular design of more effective

drugs. Such a device is the right step toward developing techniques that can replace two-dimensional electrophoresis. The application of a microfluidic device, integrating a flow cytometry unit, to the separation of cells was also reported. Like a microfluidic device, miniaturization of microplates has also been studied, which will be useful for high-throughput screening in drug discovery. Thus, the utilization of the microfluidic device has definitely spread very wide.

See also Immuno-PCR; Microarray-Based Technology: Basic Principles, Advantages and Limitations; Peptide and Non-Peptide Combinatorial Libraries; Protein Microarrays; Real-Time Quantitative PCR: Theory and Practice.

Bibliography

Books and Reviews

Auroux, P.-A., Iossifidis, D., Reyes, D.R., Manz, A. (2002) Micro total analysis systems. 2. Analytical standard operations and applications, *Anal. Chem.* **74**, 2637–2652.

Becker, H., Locascio, L.E. (2002) Polymer microfluidic devices, *Talanta* **56**, 267–287.

Jain, K.K. (2000) Application of biochip and microarray systems in pharmacogenomics, *Pharmacogenomics* **1**, 289–307.

Kan, C.W., Fredlake, C.P., Doherty, E.S., Barron, A.E. (2004) DNA sequencing and genotyping in miniaturized electrophoresis systems, *Electrophoresis* **25**, 3564–3588.

Lacher, N.A., Garrison, K.E., Martin, R.S., Lunte, S.M. (2001) Microchip capillary electrophoresis/electrochemistry, *Electrophoresis* **22**, 2526–2536.

Lin, Y.-W., Huang, M.-F., Chang, H.-T. (2005) Nanomaterials and chip-based nanostructures for capillary electrophoretic separations of DNA, *Electrophoresis* **26**, 320–330.

Lion, N., Rohner, T.C., Dayon, L., Arnaud, I.L., Damoc, E., Youhnovski, N., Wu, Z.-Y., Roussel, C., Josserand, J., Jensen, H., Rossier, J.S., Przybylski, M., Girault, H.H. (2003) Microfluidic systems in proteomics, *Electrophoresis* **24**, 3533–3562.

Oosterbroek, R.E., van den Berg, A. (2003) *Lab-on-a-Chip: Miniaturized Systems for (bio) Chemical Analysis and Synthesis*, Elsevier Biomedical, Amsterdam.

Reyes, D.R., Iossifidis, D., Auroux, P.-A., Manz, A. (2002) Micro total analysis systems. 1. Introduction, theory, and technology, *Anal. Chem.* **74**, 2623–2636.

Sanders, G.H.W., Manz, A. (2000) Chip-based microsystems for genomic and proteomic analysis, *Trends Analyt. Chem.* **19**, 364–378.

Schena, M. (1999) *DNA Microarray, A Practical Approach*, Oxford University Press, New York.

Takenaka, S. (2001) Highly sensitive probe for gene analysis by electrochemical approach, *Bull. Chem. Soc. Jpn.* **74**, 217–224.

Verpoorte, E. (2002) Microfluidic chips for clinical and forensic analysis, *Electrophoresis* **23**, 677–712.

Verpoorte, E., De Rooij, N.F. (2003) Microfluidics meets MEMS, *Proc. IEEE* **91**, 930–953.

Wang, J. (2002a) Electrochemical detection for microscale analytical systems: a review, *Talanta* **56**, 223–231.

Wang, J. (2002b) On-chip enzymatic assays, *Electrophoresis* **23**, 713–718.

Primary Literature

Ander, R.C., Su, X., Bogdan, G.J., Fenton, J. (2000) A miniature integrated device for automated multistep genetic assays, *Nucleic Acids Res.* **28**, e60.

Bakajin, O., Duke, T.A.J., Tegenfeldt, J., Chou, C.-F., Chan, S.S., Austin, R.H., Cox, E.C. (2001) Separation of 100-kilobase DNA molecules in 10 seconds, *Anal. Chem.* **73**, 6053–6056.

Baldwin, R.P., Roussel, T.J. Jr., Crain, M.M., Bathlagunda, V., Jackson, D.J., Gullapalli, J., Conklin, J.A., Pai, R., Naber, J.F., Walsh, K.M., Keynton, R.S. (2002) Fully integrated on-chip electrochemical detection for capillary electrophoresis in a microfabricated device, *Anal. Chem.* **74**, 3690–3697.

Bousse, L., Mouradian, S., Minalla, A., Yee, H., Williams, K., Dubrow, R. (2001) Protein sizing on a microchip, *Anal. Chem.* **73**, 1207–1212.

Fan, Z.H., Mangru, S., Granzow, R., Heaney, P., Ho, W., Dong, O., Kumar, R. (1999) Dynamic DNA hybridization on a chip using paramagnetic beads, *Anal. Chem.* **71**, 4851–4859.

Fu, A.Y., Spence, C., Scherer, A., Arnold, F.H., Quake, S.R. (1999) A microfabricated fluorescence-activated cell sorter, *Nat. Biotechnol.* **17**, 1109–1111.

Huhmer, A.F.R., Landers, J.P. (2000) Noncontact infrared-mediated thermocycling for effective polymerase chain reaction amplification of DNA in nanoliter volumes, *Anal. Chem.* **72**, 5507–5512.

Kaji, N., Tezuka, Y., Takamura, Y., Ueda, M., Nishimoto, T., Nakanishi, H., Horiike, Y., Baba, Y. (2004) Separation on long DNA molecules by quartz nanopillar chips under a direct current electric field, *Anal. Chem.* **76**, 15–22.

Kopp, M.U., DeMello, A.J., Manz, A. (1998) Chemical amplification: continuous-flow PCR on a chip, *Science* **280**, 1046–1048.

Koutuy, L.B., Schmalzing, D., Taylor, T.A., Fuchs, M. (1996) Microchip electrophoretic immunoassay for serum cortisol, *Anal. Chem.* **68**, 18–22.

Landers, J.P. (2003) Molecular diagnostics on electrophoretic microchips, *Anal. Chem.* **75**, 2919–2927.

Lenigk, R., Liu, R.H., Athavale, M., Chen, Z., Ganser, D., Yang, J., Rauch, C., Liu, Y., Chen, B., Yu, H., Ray, M., Marrero, R., Grodzinski, P. (2002) Plastic biochannel hybridization devices: a new concept for microfluidic DNA arrays, *Anal. Biochem.* **311**, 40–49.

Liu, Y., Rauch, C.B. (2003) DNA probe attachment on plastic surfaces and microfluidic hybridization array channel devices with sample oscillation, *Anal. Biochem.* **317**, 76–84.

Liu, R.H., Yang, J., Lenigk, R., Bonanno, J., Grodzinski, P. (2004) Self-contained, fully integrated biochip for sample preparation, polymerase chain reaction amplification, and DNA microarray detection, *Anal. Chem.* **76**, 1824–1831.

McClain, M.A., Culbertson, C.T., Jacobson, S.C., Ramsey, J.M. (2001) Flow cytometry of *Escherichia coli* on microfluidic devices, *Anal. Chem.* **73**, 5334–5338.

Medintz, I.L., Paegel, B.M., Blazej, R.G., Emrich, C.A., Berti, L., Scherer, J.R., Mathies, R.A. (2001) High-performance genetic analysis using microfabricated capillary array electrophoresis microplates, *Electrophoresis* **22**, 3845–3856.

Noerholm, M., Bruus, H., Jakobsen, M.H., Telleman, P., Ramsing, N.B. (2003) Polymer microfluidic chip for online monitoring of microarray hybridizations, *Lab Chip* **4**, 28–37.

Pease, A.C., Solas, D., Sullivan, E.J., Caronin, M.T., Holmes, C.P., Fodor, S.P.A. (1994) Light-generated oligonucleotide arrays for rapid DNA sequence analysis, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5022–5026.

Rocklin, R.D., Ramsey, R.S., Ramsey, J.M. (2000) A microfabricated fluidic device for performing two-dimensional liquid-phase separations, *Anal. Chem.* **72**, 5244–5249.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science* **270**, 467–470.

Schmalzing, D., Koutny, L., Adourian, A., Belgrader, P., Matsudaira, P., Ehrlich, D. (1997) DNA typing in thirty seconds with a microfabricated device, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10273–10278.

Schmalzing, D., Adourian, A., Koutny, L., Ziaugra, L., Matsudaira, P., Ehrlich, D. (1998) DNA sequencing on microfabricated electrophoretic devices, *Anal. Chem.* **70**, 2303–2310.

Schmalzing, D., Koutny, L., Chisholm, D., Adourian, A., Matsudaira, P., Ehrlich, D. (1999) Two-color multiplexed analysis of eight short tandem repeat loci with an electrophoretic microdevice, *Anal. Biochem.* **270**, 148–152.

Seong, G.H., Crooks, R.M. (2002) Efficient mixing and reactions within microfluidic channels using microbead-supported catalysts, *J. Am. Chem. Soc.* **124**, 13360–13361.

Southern, E.M., Maskos, U., Elder, J.K. (1992) Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models, *Genomics* **13**, 1008–1017.

Tabuchi, M., Ueda, M., Kaji, N., Yamasaki, Y., Nagasaki, Y., Yoshikawa, K., Kataoka,

K., Baba, Y. (2004) Nanospheres for DNA separation chips, *Nat. Biotechnol.* **22**, 337–340.

Wakai, J., Takagi, A., Nakayama, M., Miya, T., Miyahara, T., Iwanaga, T., Takenaka, S., Ikeda, Y., Amano, M. (2004) A novel method of identifying genetic mutations using an electrochemical DNA array, *Nucleic Acids Res.* **32**, e141.

Wang, J., Ibanez, A., Chatrathi, M.P., Escarpa, A. (2001) Electrochemical enzyme immunoassays on microchip platforms, *Anal. Chem.* **73**, 5323–5327.

Wang, Y., Vaidya, B., Farquar, H.D., Stryjewski, W., Hammer, R.P., McCarley, R.L., Soper, S.A. (2003) Microarrays assembled in microfluidic chips fabricated from poly(methyl methacrylate) for the detection of low-abundant DNA mutations, *Anal. Chem.* **75**, 1130–1140.

Waters, L.C., Jacobson, S.C., Kroutchinina, N., Khandurina, J., Foote, R.S., Ramsey, J.M. (1998a) Microchip device for cell lysis, multiplex PCR amplification, and electrophoretic sizing, *Anal. Chem.* **70**, 158–162.

Waters, L.C., Jacobson, S.C., Kroutchinina, N., Khandurina, J., Foote, R.S., Ramsey, J.M. (1998b) Multiple sample PCR amplification and electrophoretic analysis in a microchip, *Anal. Chem.* **70**, 5172–5176.

Woolley, A.T., Lao, K., Glazer, A.N., Mathies, R.A. (1998) Capillary electrophoresis chips with integrated electrochemical detection, *Anal. Chem.* **70**, 684–688.

Yuen, P.K., Kricka, L.J., Fortina, P., Panaro, N.J., Sakazume, T., Wilding, P. (2001) Microchip module for blood sample preparation and nucleic acid amplification reactions, *Genome Res.* **11**, 405–412.

Trace Elements Micronutrients: *see* Micronutrients, Trace Elements

Transcription Factor NF-κB: Function, Structure, Regulation, Pathways, and Applications

Raymond Cheong and Andre Levchenko
Johns Hopkins University, Baltimore, MD

1 Introduction 438

2 Target Genes 439

3 Protein Domains and Crystal Structure 440

4 Posttranslational Regulation of NF-κB 443

5 Canonical and Alternative Pathway: Roles of IκB and IKK 445
5.1 Canonical pathway 445
5.2 Alternative Pathway 448

6 NF-κB-regulating Signaling Pathways 449
6.1 TNFα 449
6.2 LPS 452
6.3 IL-1 453
6.4 Antigens 454
6.5 Stress 455

7 Other Roles of NF-κB in Normal Biology and Pathology 456
7.1 NF-κB in Cancer 456
7.2 Pharmacologic Inhibition of NF-κB 458
7.3 Atypical Roles of NF-κB: Example of Learning 459

Acknowledgment 460

Bibliography 460
Books and Reviews 460
Primary Literature 460

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

Keywords

Apoptosis
An organized form of cell death.

Canonical Pathway
NF-κB activation mediated by IKKβ phosphorylation of IκB.

IκB
Inhibitor of NF-κB.

IKK
IκB kinase, which contains at least three subunits named α, β, γ.

NF-κB
Nuclear factor that binds to κB enhancers.

Rel
The protein family that includes NF-κB.

NF-κB is a transcription factor found in most cells that plays a central role in the inflammatory, immunologic, apoptotic, and stress responses. NF-κB is a dimer composed of members of the Rel family of transcription factor subunits. It is normally sequestered in the cytoplasm by its inhibitor, IκB. Nearly all NF-κB-activating stimuli activate the IκB kinase, which phosphorylates IκB and thereby targets it for degradation. Newly freed NF-κB then enters the nucleus and mediates its various effects by activating its target genes. NF-κB activity is modulated by numerous mechanisms, including combinatorial specificity of its subunits and inhibitors, several forms of posttranslation modifications, and differences in signal transduction. NF-κB is of great therapeutic interest and we highlight its role in cancer and review modern drugs used to inhibit it. Finally, we discuss the role of NF-κB in learning, a hint that NF-κB has functions beyond those traditionally associated with it.

1 Introduction

In 1986, Sen and Baltimore isolated a small nuclear factor (protein) able to bind to the enhancer of and activate transcription of the immunoglobulin κ light chain gene in B cells, which they termed NF-κB. Originally this transcription factor was thought to be restricted to lymphoid cells, but it quickly became clear that NF-κB is present in most cell types

and that it regulates much more than immunoglobulins. Today, it is apparent that not only is NF-κB a transcription factor central to the regulation of the immune system, but it is also important in fundamental processes like apoptosis and cellular response to stress. For this reason, NF-κB is a topic of intense interest in a wide variety of laboratories.

NF-κB is an inducible transcription factor that responds to a bewildering array of stimuli including inflammatory molecules, viruses, bacteria, "stresses," and medications. The list of the target genes of NF-κB is equally extensive, and includes immune cytokines and receptors, antiapoptotic genes, and various stress response genes. Given the pleiotropic nature of the signals, functions, and targets of NF-κB, one expects NF-κB to be subjected to a great deal of regulation and indeed this is the case. This review takes a bottom-up approach to the many layers of NF-κB regulation. First, we discuss the nature of NF-κB's targets, and examine the NF-κB molecule itself to understand how its structure influences its function as a transcription factor. Then we consider modes of direct regulation in the form of posttranslational modifications and interactions with its inhibitor IκB. We continue to "zoom out" and consider the major NF-κB-regulating signaling pathways and finally end with discussion of the broader implications of NF-κB in biology, medicine, and disease progression.

By way of introduction, we briefly describe here the consensus "canonical pathway" so that molecular details of NF-κB in Sect. 1–3 can be placed in proper context (see Fig. 4 and discussed further in Sect. 4.1). In the absence of stimulation, NF-κB is sequestered in the cytoplasm, bound to its inhibitor IκB. Nearly all NF-κB-activating signals converge on the activation of the IκB kinase (IKK). IKK phosphorylates IκB and thus targets this NF-κB inhibitor for degradation by an ubiquitin-proteasome pathway. Newly freed NF-κB then translocates to the nucleus and activates transcription of its target genes. Among these targets is the α-isoform of IκB, and newly synthesized IκBα enters the nucleus, binds to NF-κB and exports it to the cytoplasm. This returns the system to its original state and thereby provides signal downregulation.

2 Target Genes

Originally NF-κB was isolated as a factor that bound to the enhancer of the κ immunoglobulin light chain, and it was thought to be a transcriptional regulator restricted to B cells. However, it soon became clear that NF-κB was not only found in nearly all cell types but that it also regulated a wide variety of genes. Comprehensive lists of *bona fide* NF-κB targets indicate NF-κB regulates on the order of 200 to 300 genes, plus dozens more putative targets. A sample of such targets is provided in Table 1. The consensus NF-κB binding sequence for these genes is 5′-GGGRNNYYCC-3′ (G = guanine, C = cytosine, R = adenine or guanine, Y = thymine or cytosine, N = any nucleotide).

Grouping the target genes into broad families indicates why NF-κB is considered a central mediator of the immune response. NF-κB induces expression of chemokines and cytokines that modulate the immune response, many immunologically important cell surface molecules, cell adhesion molecules involved in movement and spatial targeting of leukocytes, and components of

Tab. 1 Selected target genes activated by NF-κB. See http://people.bu.edu/gilmore/nf-kb/target/ for a comprehensive list.

Class	*Example target genes*
Cytokines	IFNβ/γ, IL−1α/1β/2/6/8/9/10/11/12/13/15, LTα/β, TNFα/β, TRAIL, RANTES
Immune-related	B7.1, CCR5/7, CD23/40/48/83/137/154, IL2Rα, IgCγ1/Cγ4/ε/κ, MHCI, β2m, TCR, TLR9, complement B/C4, CRP, LBP, β-defensin, tissue factor-1
Cell adhesion	ELAM-1, ICAM-1, P-Selectin, VCAM-1
Stress response	Angiotensin, COX-2, iNOS, SOD, Ferritin-H, CYP2E1, CYP2C11
Apoptosis	Bfl1, Bcl-xL, Bcl-2, c-FLIP, IAP, TRAF1, TRAF2, Fas, FasL
Growth factors	BMP2, G-CSF, GM-CSF, M-CSF, EPO, PDGFβ, VEGF
Transcription factors	c-Rel, RelB, p100, p105, IκBα, A20, c-myc, p53, junB, IRF-1/2/4/7, Stat5a
Viruses	Adenovirus, CMV, EBV, HBV, HIV-1, HSV, JCV, HPV16, SIV, SV40
Neuroreceptors	μ-opioid receptor, NPY receptor, NMDA (putative)

the complement system, among others. Interestingly, the expression of some viral proteins is also induced by NF-κB, which may allow some viruses to proliferate in immunostimulatory conditions and thereby avoid elimination by the immune system.

NF-κB target genes extend beyond the immune system. Of special importance is the fact that NF-κB upregulates various antiapoptotic genes, such as IAPs, c-FLIP, and Bcl-2, making it an important factor in cell survival. Also, since NF-κB is activated by reactive oxygen species, UV radiation, heavy metals, and similar noxious agents, and since NF-κB can upregulate antioxidant enzymes and factors like angiotensin II, iNOS, and COX-2, it has been proposed that NF-κB is more generally a central mediator of the cell stress response. However, "stress" is admittedly an amorphous concept, and the signaling pathways activated by stress are poorly characterized and not greatly specific to NF-κB activation as we shall see in Sect. 5.5.

It seems unreasonable to expect that any NF-κB-activating stimulus results in transcription of all its target genes. Indeed, there are several mechanisms by which specificity might be maintained. This includes differences in NF-κB promoter affinity, NF-κB/IκB/IKK composition and modification, activity kinetics, and signal-specific pathways. These forms of regulation are the subject of the next several sections.

3 Protein Domains and Crystal Structure

NF-κB is a dimer composed of five possible subunits, p65 (RelA), c-Rel, RelB, p50, and p52, (see Fig. 1) with the p65/p50 being the predominant heterodimer in most cells. (In this review, unless otherwise noted, we use NF-κB to denote p65/p50.) Thirteen of the 15 possible pairs of the 5 subunits can form *in vivo*, with the notable exception of RelB/c-Rel and RelB/RelB. These dimers not only differ in DNA-binding specificity, thereby contributing to specific gene activation and specific dimer functions, but

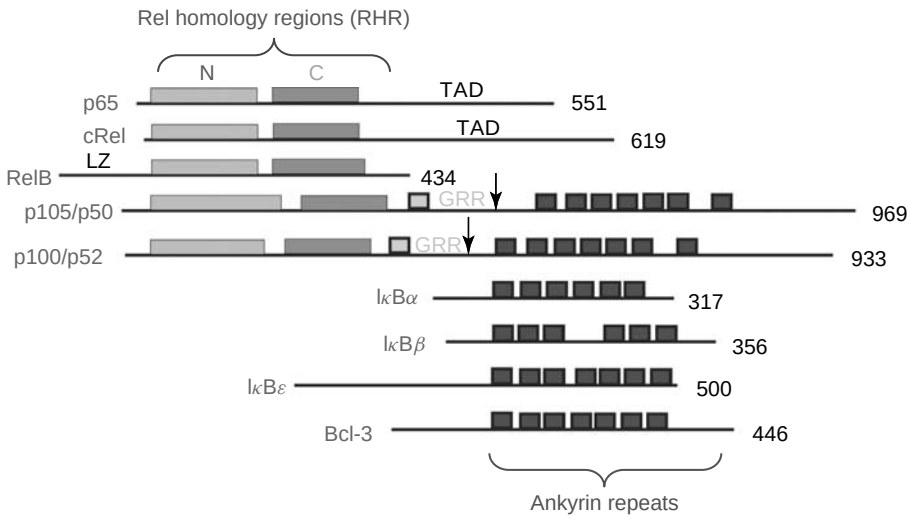

Fig. 1 Rel and IκB family members. p65, c-Rel, RelB, p105, and p100 share an N-terminal Rel-homology region (RHR) composed of two domains and a nuclear localization signal (NLS, see Fig. 2)). RelB also contains an N-terminal leucine zipper (LZ). p65, c-Rel, and RelB are transcriptionally active and contain C-terminal transcriptional activation domains (TAD). p105 and p100 are processed to p50 and p52 respectively by C-terminal proteolysis to the indicated point (arrow), which requires the glycine-rich region (GRR). p105, p100, IκBα, IκBβ, IκBε, and Bcl-3 contain 5–7 ankyrin repeats that mediate binding to Rel family members. The figure is based on ClustalW alignment of NCBI protein sequences NP_068810 (p65), NP_002899 (c-Rel), NP_006500 (RelB), NP_003989 (p105), NP_002493 (p100), NP_065390 (IκBα), NP_002494 (IκBβ), NP_004547 (IκBε), and NP_005169 (Bcl-3).

also differ in stability and affinity with the various IκB isoforms and therefore can be differentially regulated. In fact, stimulus-dependent alteration in dimer composition provides a direct mechanism by which NF-κB activity is modulated over time.

All five NF-κB subunits are related to v-rel, an oncogene encoded by the avian reticuloendotheliosis retrovirus that can induce acute leukemia in chickens. More generally, the Rel family is a group of evolutionarily conserved proteins and includes, for example, the *Drosophila* proteins Dorsal, Dif, and Relish, which are not only homologous in protein sequence but also have functions and participate in signaling pathways similar to that of NF-κB in humans. Interestingly, NFAT (nuclear factor of activated T cells) is a transcription factor that also has significant homology to Rel family members, but is primarily responsive to calcium/calcineurin signals.

All five subunits contain a ~300 amino acid N-terminal Rel-homology region (RHR), which is responsible for dimerization, DNA binding, and nuclear localization. Crystal structures of the p65, c-Rel, p50, and p52 homodimers as well as the p65/p50 heterodimer reveal that the RHR is composed of two domains each with immunoglobulin-like β-strand structure. When bound to DNA, NF-κB dimers adopt a unique butterfly-like structure (see Fig. 2). One set of "wings" is formed by the C-terminal domains of each RHR, which dimerize via interdigitating

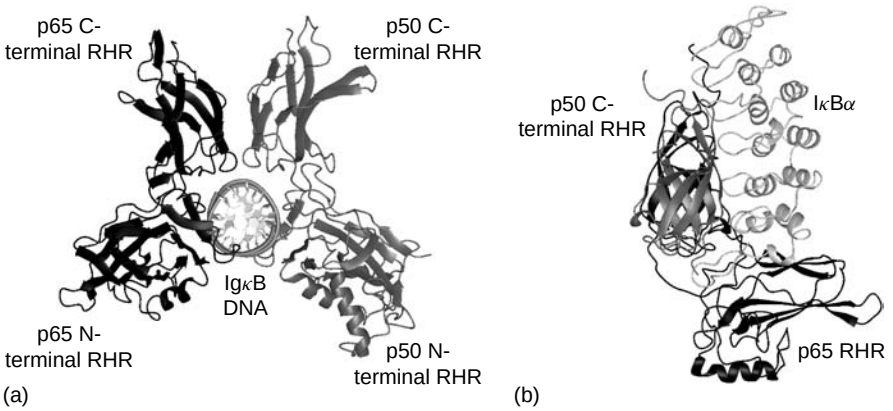

Fig. 2 Crystal structures of NF-κB. p65/p50 heterodimer bound to IgκB DNA (a) and bound to IκBα (b). p65 is shown in black, p50 in dark gray, IκBα in light gray (b), and IgκB in light gray (a). Images were created from protein databank accession numbers 1VKX (Chen, F.E., et al., (1998)) and 1IKN (Huxford, T., et al., (1998)) using Molw PDB Viewer 4.0 (Molecular Images) and Molmol 2.5.1.

hydrophobic side chains. These side chains as well as nearby polar/ionic amino acid interactions influence dimer stability. Another set of "wings" is formed by the N-terminal domains that form extensive contacts with the phosphate backbone of the DNA. The C-terminal domain also contributes to the dimer-DNA interface, although specific base pair recognition is mediated by the N-terminal domain. About a half-dozen key residues in the N-terminal domain mediate specific DNA binding, and combined with relative orientations of the subunits, largely accounts for the different binding affinities of the various NF-κB dimers to different DNA sequences. Additionally, only the major groove of DNA is bound by NF-κB dimers, and in some of them, no protein loops block the minor groove, which is thus open for potential simultaneous binding by other proteins such as HMG-I(Y). Finally, the C-terminal domain contains a nuclear localization signal (NLS), which allows free NF-κB to translocate to the nucleus.

Additionally, the p65, c-Rel, and RelB subunits contain a C-terminal transcriptional activation domain (TAD). However, p50 and p52 lack such a domain, which makes them transcriptionally inactive, with p50 and p52 homodimers serving to repress NF-κB target genes. Instead of a TAD, the C-terminus of p50 and p52 only contains a short glycine-rich region (GRR), a remnant of the C-terminus of their larger precursors p105 (NF-κB1) and p100 (NF-κB2). Interestingly, the C-terminal halves of these precursors, which are degraded, contain ankyrin repeats homologous to those of the IκB family. Thus, p105 and p100 can inhibit the other NF-κB subunits or the processed versions of themselves, a topic discussed in Sect. 4.2.

The NF-κB inhibitors, that is, the IκB family of proteins, have in common a series of 5–7 ankyrin repeats that mediate binding to and sequestration of the NF-κB proteins. In mammalian cells, the IκB family is generally considered to be limited to the three major isoforms IκBα, IκBβ, IκBε, as well as the p105 and

p100 NF-κB precursors, and the proto-oncogene Bcl-3. Discussion of the different functions of the IκB isoforms is postponed to Sect. 4.1, but here we note that there are also three lesser-studied members: IκBγ, which is identical to p105 C-terminus and is produced by alternative splicing; IκBζ, an apparently inducible IκB form; and IκBR, possibly a p50/p50-specific inhibitor. Additionally, Bcl-3 can also promote NF-κB activity by transcriptional derepression, that is, by inhibiting p50 homodimer binding to DNA, which has no transcriptional activity.

Crystal structures of IκBα bound to p65/p50 show that the ankyrin repeats of IκBα stack upon each other to form a slightly bent cylinder. Repeats 4–6 bind to the C-terminal domains of the RHR of p65 and p50, either causing the formation of or stabilizing the RHR N-terminal domains in a "closed" conformation incapable of DNA binding. Also, ankyrin repeats 1 and 2 interact with the NLS on p65, masking it from the nuclear import machinery. Together, these interactions help explain early observations that IκB not only sequesters NF-κB in the cytoplasm but also detaches NF-κB bound to DNA. IκBα (and IκBβ) also contains an N-terminal region that is phosphorylated in response to various signals and a C-terminal region that contains a PEST domain that regulates constitutive degradation, but neither of these regions were fully included in the crystal structure. Overall, the crystal structure of IκBβ bound to p65 homodimer is similar to the structure of the IκBα–p65/p50 complex, with the major differences relating to a longer loop between repeats 3 and 4 and an increased distance between repeat 6/C-terminus of IκBβ and NF-κB. It has been proposed but not shown conclusively that these differences could account for, respectively, the observations that IκBβ masks both NLSs of NF-κB and that it does not prevent DNA binding. In all, the reported crystal structures are immensely helpful in explaining a large volume of data regarding interactions between IκB, NF-κB, and DNA, and also provide an easy test for the plausibility of future proposed interactions.

4 Posttranslational Regulation of NF-κB

Posttranslational modifications provide an avenue of direct influence on the behavior of proteins and are a common mechanism of transcription factor regulation. NF-κB is no exception (see Fig. 3) and the major forms of posttranslational modifications known to affect its function are phosphorylation, acetylation, and ubiquitination. (Proteolytic processing may also be considered a form of posttranslational modification, but its discussion is postponed to Sect. 4.2, where it is considered in the context of the alternative pathway.) However, much of the work elucidating these mechanisms is relatively recent and, as a result, many questions remain unanswered.

The p65 subunit was early on recognized as a target for induced phosphorylation and it is evident today that p65 can be phosphorylated on numerous residues by a variety of kinases activated by different signals. The best-characterized phosphorylation site is serine 276. Initial evidence of its importance came as early as 1993, when it was discovered that mutation of the analogous site in c-Rel (S275) abrogated DNA binding. Phosphorylation of S276 in p65 by PKA or MSK1 increases the affinity of NF-κB for DNA and also promotes binding of coactivators like CBP/p300, which can then recruit the general transcriptional machinery. Other

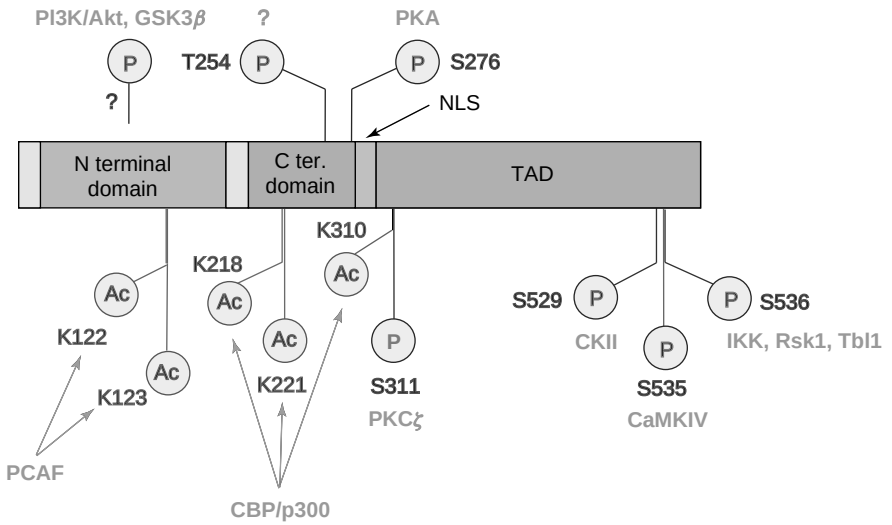

Fig. 3 Various phosphorylation and acetylation sites on p65 are indicated along with the corresponding enzymes. The sites are shown to scale and are explicitly identified. See text for more details.

C-terminal serines have been implicated as phosphorylation sites, including S311 phosphorylated by PKCζ, S529 phosphorylated by CKII, S535 phosphorylated by IKK/Rsk1/Tbk1, and S535 phosphorylated by CaMKIV. Finally, PI3K/Akt and GSK3β have been implicated in inducing p65 phosphorylation at unknown sites. Phosphorylation at any of these sites is associated with positive effects on NF-κB activity, but additional experiments are needed to establish the relative importance and biological role of these phosphorylation sites.

In recent years, acetylation of NF-κB has emerged as another important regulatory posttranslational modification. NF-κB is acetylated at several conserved lysines, although unlike phosphorylation, this modification may be activating or inhibitory depending on location. CBP/p300 can acetylate p65 at lysines 218, 221, and 310, which increases enhancer binding, decreases affinity for IκBα, and increases overall transcriptional activity. As CBP/p300 has also been implicated as a coactivator that binds to phosphorylated p65 (see above), this raises the possibility that NF-κB is brought to full transcriptional activity by phosphorylation followed by acetylation. NF-κB activity can then be downregulated by further acetylation of p65 at lysines 122 and 123 by p300 and PCAF (CBP-associated factor), reducing affinity of NF-κB for DNA. p50 also can be activated by acetylation at lysines 431, 440, and 441, hinting at a broad role of acetyltransferases in the regulation of different NF-κB dimers. Histone deacetylases like HDAC3 play a role in reversing these modifications and provide a potential mechanism whereby acetylation can reversibly regulate NF-κB activity.

A recent development has implicated ubiquitination in p65 regulation. The E3 ubiquitin ligase SOCS-1 mediates proteolysis of p65, and the peptidyl–prolyl isomerase Pin-1 protects p65 from this

degradation, provided p65 is phosphorylated at threonine 254 by a yet unidentified kinase. Pin-1 further promotes NF-κB activity by inhibiting association between p65 and IκBα. It is not clear whether these effects are limited to the specific tissues (breast epithelium and breast cancer cells) in which they were discovered, but this finding does demonstrate that there remains much to understand about the regulation of NF-κB at the level of the molecule itself.

5 Canonical and Alternative Pathway: Roles of IκB and IKK

5.1 Canonical pathway

Early experiments demonstrated that NF-κB and its inhibitor IκBα were locked in a negative feedback loop:

1. Nuclear and cytosolic protein extracts show no κ enhancer binding activity, but treatment of those fractions with detergents reveals as much NF-κB as found in the nucleus of stimulated cells, implying the presence of an inhibitor that keeps NF-κB inactive in the cytosol. This inhibitor, termed IκBα, was soon purified and later cloned.
2. Transient treatment by phorbol ester or TNFα causes degradation of IκB with concomitant NF-κB DNA-binding activity, followed by upregulation of IκBα mRNA, and replenishment of IκB with concomitant loss of NF-κB DNA-binding activity. This indicated that NF-κB regulates its own inhibitor.
3. The IκBα promoter contains at least one NF-κB binding site and failure to activate NF-κB leads to failure of IκBα transcription. Also, blockage of protein synthesis leads to persistent NF-κB activity and continued accumulation of IκBα mRNA. This demonstrated that NF-κB directly regulates its own inhibitor.

These observations formed the basis of the canonical pathway, which we discuss in detail in this section (see Fig. 4).

The canonical pathway hinges on the ability of NF-κB, IκBα, and their complex to move between the cytoplasm and nucleus, but attempts to understand the details of this mechanism have uncovered a more complex story. Nuclear import of free NF-κB is made possible by an NLS in the C-terminus of the RHR of p65, but p50 may contain a nuclear export signal (NES), which would antagonize NF-κB activity. Movement of free IκB and export of IκB–NF-κB complex is likewise believed to be mediated by an NES and NLS on IκB, but the exact sequences remain to be resolved. The consensus is that the ankyrin repeats of IκBα control nuclear localization, but it is unclear if this is due to a Ran-independent NLS near the second repeat, or if the ankyrin repeats recruit another protein, which then allows nuclear import via a Ran-dependent mechanism. IκBα contains a putative C-terminal NES, but multiple studies show that the true NES is actually in its N-terminus and its function is CRM-1 dependent. Nonetheless, it is clear that IκBα is capable of nuclear import and export on its own, which leads to the possibility that IκBα–NF-κB could move in and out of the nucleus independent of stimulus. Indeed, blocking nuclear export by pharmacologic inhibition of CRM-1 by leptomycin B leads to accumulation of IκBα and NF-κB in the nucleus. This demonstrates that despite the predominantly cytoplasmic localization of IκBα and NF-κB, they constitutively

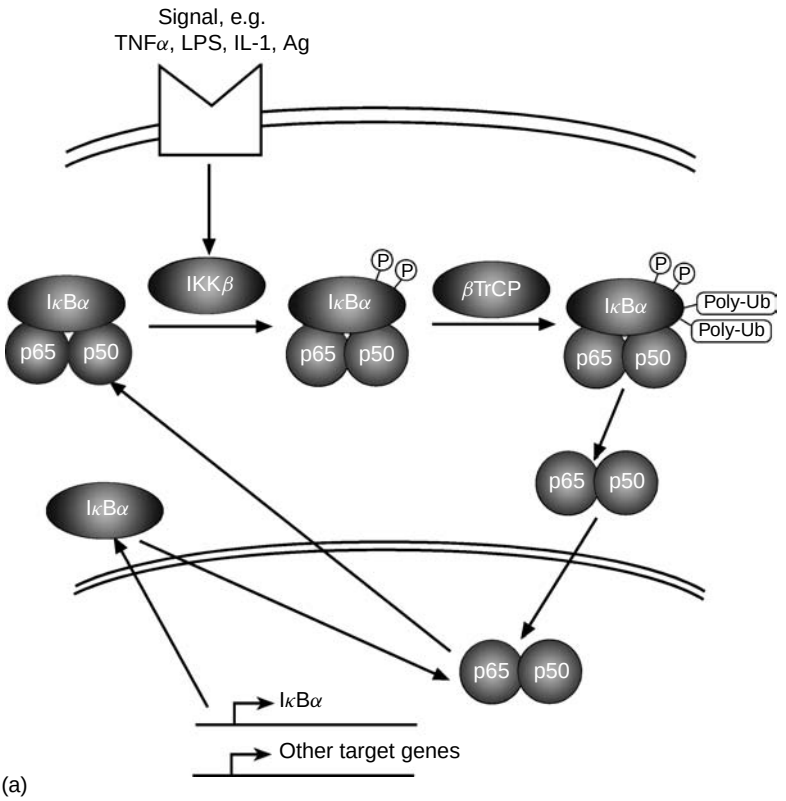

Fig. 4 Canonical and alternative pathways. (a) most stimuli activate NF-κB through the canonical pathway, which involves IKKβ-mediated degradation of IκB and involves a negative feedback loop through IκBα. (b) Some stimuli activate NF-κB through the alternative pathway, which involves IKKα-mediated processing of p100 to p52.

shuttle between the cytoplasm and nucleus. Consistent with this idea, shuttling is also observed in IκBε, which contains an NLS and NES similar to that proposed for IκBα, but not observed in the case of IκBβ, which has no identified NLS or NES. Shuttling of IκBα–NF-κB complexes may also be aided by the inability of IκBα to fully mask the NLSs of NF-κB. A quantitative model of shuttling suggests that it functions to inhibit the small amount of NF-κB not bound by IκB, due to a small but finite dissociation rate, but this model implies that the import and export rates of IκBα depend on whether or not it is bound to NF-κB, but how their interaction alters translocation rate is unclear. Needless to say, the mechanism and biological role of shuttling is still an open question. So, while it is clear that IκB and NF-κB are joined in a negative feedback loop, the dynamics of their interaction remain to be resolved.

More upstream in the canonical pathway, after discovery of the autoregulatory pathway, it soon became evident that a

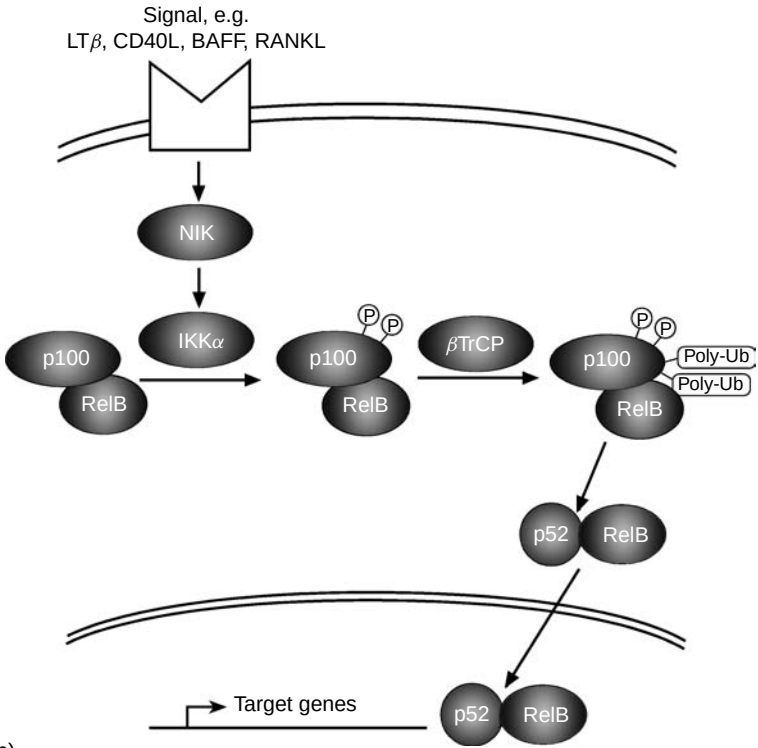

(b)

Fig. 4 (*Continued*)

multitude of NF-κB-activating signals all resulted in phosphorylation of IκB at two conserved C-terminal serines (S32 and S36 in IκBα). Phosphorylation at these sites allows IκBα to be recognized by β-TrCP, the recognition subunit of an SCF-type E3 ubiquitin ligase that polyubiquitinates IκBα at lysines 21 and 22. This finally targets IκB for degradation by the 26S proteasome. However, several authors have noted that the primary form of β-TrCP is localized in the nucleus, which would imply that separation of NF-κB from IκB in the cytoplasm is not a prerequisite for accumulation of nuclear NF-κB, which is directly at odds with the canonical model. It is not currently clear how to resolve this paradox, but it seems likely that any explanation would have to take into account the aforementioned shuttling property.

IκBα is essential for regulating NF-κB, as evidenced by the observation that IκBα$^{-/-}$ knockout mice die perinatally due to multiorgan inflammation. However, the role of the two other major isoforms IκBβ and IκBε is unclear. The phenotype of IκBβ$^{-/-}$ knockout mice is unpublished but reportedly mild. Also, despite their clear differences in structure and response kinetics, IκBβ can substitute for IκBα with no obvious deleterious effects. IκBε$^{-/-}$ mice likewise have no discernible phenotype except a decrease in a subpopulation of T cells. Additionally, IκBε appears to interact primarily with

p65/c-Rel or p65/p65 dimers. Two features that IκBβ and IκBε have in common that distinguish them from IκBα is that they respond much slower to the same stimuli and they are not transcriptionally activated by NF-κB. Consequently, IκBβ and IκBε might function to dampen oscillations that arise out of the NF-κB–IκBα negative feedback loop, thereby sustaining a prolonged secondary response to persistent stimuli.

Recognition of the need for IκB phosphorylation to initiate an NF-κB response led to intense interest in identification of the responsible kinase. The IκB kinase complex was first isolated in 1997 and is now known to be composed of at least three subunits termed IKKα, IKKβ, and NEMO (also called IKKγ). IKKα and IKKβ are highly homologous catalytic subunits with kinase activity, with the former chiefly involved in the alternative pathway (see below) and the latter critical to cytokine-mediated NF-κB activation. NEMO is considered a regulatory subunit with no enzymatic activity and genetic studies also confirm it is a necessary component for cytokine-mediated activity. A fourth subunit, IKKε, was proposed as a member of the IKK complex, however, its requirement is limited to a subset of NF-κB-activating stimuli. Likewise, the ELKS protein was recently coimmunopurified with the IKK complex and identified as another regulatory subunit that appears to recruit IκBα to the complex, but further characterization of ELKS is needed to understand its role. The IKK complex has an apparent molecular mass of 700 to 900 kDa as determined by gel filtration, although the individual components total about 300 kDa. As a result, the exact stoichiometric makeup is of some debate and may not even be constant, although some have proposed IKKα–IKKβ–$NEMO_3$ and $(IKK\alpha–IKK\beta–NEMO)_2$. The mechanism of IKK activation is even more unclear, though the lack of conclusive evidence of an IKK kinase has lead to the general belief that autophosphorylation or proximity-induced trans-phosphorylation between IKKα and IKKβ leads to activation of IKK. More detailed structural analysis of the components than currently available is probably required to resolve this question.

5.2 Alternative Pathway

Several studies revealed that aly/aly (alymphoplasia, which harbors a natural mutation in NF-κB-inducing kinase, NIK) and $nf\kappa b2^{-/-}$ knockout mice had an overlapping phenotype of disturbed development or loss of peripheral lymphoid organs. This seemed to suggest that NIK and p100 were linked together in a common pathway controlling lymphoid development, and, indeed, in 2001 evidence emerged that NIK induces p100 degradation resulting in p52 formation through a phosphorylation–ubiquitination mechanism. Additionally, it was found that NIK can phosphorylate and activate IKKα, and that $IKK\alpha^{-/-}$ and inactive mutant $IKK\alpha^{AA}$ show a similar phenotype. Consistent with these observations, it was soon shown that IKKα, but not IKKβ, is preferentially involved in p100 processing. Further work by several laboratories demonstrated that p100 processing is specifically activated by LTβ, CD40L, BAFF, and RANKL binding to their respective receptors but not by canonical pathway activators like TNFα. So, the signaling pathway consisting of NIK → IKKα → p100 processing is distinct from the canonical pathway, and thus is termed the alternative pathway.

Despite characterization as such, the alternative pathway is tied to the canonical pathway on several levels. Signals that activate the alternative pathway also activate the canonical pathway and the canonical pathway can initiate transcription of p100. Also, NIK is a binding partner of TRAF2 (TNF receptor associated factor 2), which is an essential factor for canonical signaling by TNFα and p100 processing depends on the same βTrCP as the canonical pathway. Finally, LTβR can activate an overlapping but not identical set of genes via the alternative versus the canonical pathway. This raises intriguing questions about how the overlaps/differences in the components of the canonical and alternative pathways translate into shared/specific gene activation, questions that remain to be answered.

In contrast to p100, p105 is constitutively processed and is completely degraded following appropriate stimulus. Processing of both p100 and p105 clearly requires the GRR, which may act as a stop signal for the proteasome. However, the GRR is not alone sufficient and the sequences/mechanisms that mediate partial versus complete proteolysis remain to be identified. Signal-induced proteolysis of p105 is mediated by a canonical-like pathway, requiring phosphorylation at serines 927 and 932 by the IKK complex (independent of IKKα) and βTrCP-mediated ubiquitination on multiple lysines. This mechanism is somehow also dependent on phosphorylation of p105 at series 903 and 907 by GSK3β; however, the role of this kinase in the pathway is unclear. Degradation of p105 allows three distinct pathways to be activated. First, p105 releases p50, which can enter the nucleus and stimulate NF-κB-responsive genes. Second, p105 releases TPL-2, an MAP 3-kinase that mediates MAP kinase activity stimulated by LPS. Thirdly, p105 releases ABIN-2 (A20-binding inhibitor of NF-κB) a protein of unknown function that appears to down-regulate the NF-κB response. Thus, while constitutive p105 processing probably provides a steady supply of p50 (for example, for association with p65), signal-induced p105 proteolysis is in part a *bona fide* arm of the canonical pathway.

6 NF-κB-regulating Signaling Pathways

Continuing to move upstream of NF-κB, here we further describe pathways that converge on NF-κB activation. There are more than 450 known NF-κB activators, including cytokines; bacterial, viral, or parasitic pathogens; and physical, chemical, or oxidative stress. It is beyond the scope of this chapter to review all of these stimuli in detail; however, we do turn our attention to those NF-κB-activating signals that have attracted the most attention (see Fig. 5).

6.1 TNFα

Tumor necrosis factor alpha (TNFα) is a soluble trimeric protein originally discovered as an antitumor factor produced during the immune response. Today, it continues to be recognized as a major proinflammatory cytokine, which is perhaps the best studied of the major activators of NF-κB. Furthermore, TNFα alone is a topic of great interest as it and its receptor are members of much larger families with functions perhaps as diverse as that of NF-κB itself.

TNFα activates the canonical pathway by first binding to and trimerizing its primary receptor, TNFR1. This causes SODD (suppressor of death domain), which is bound to TNFR keeping it

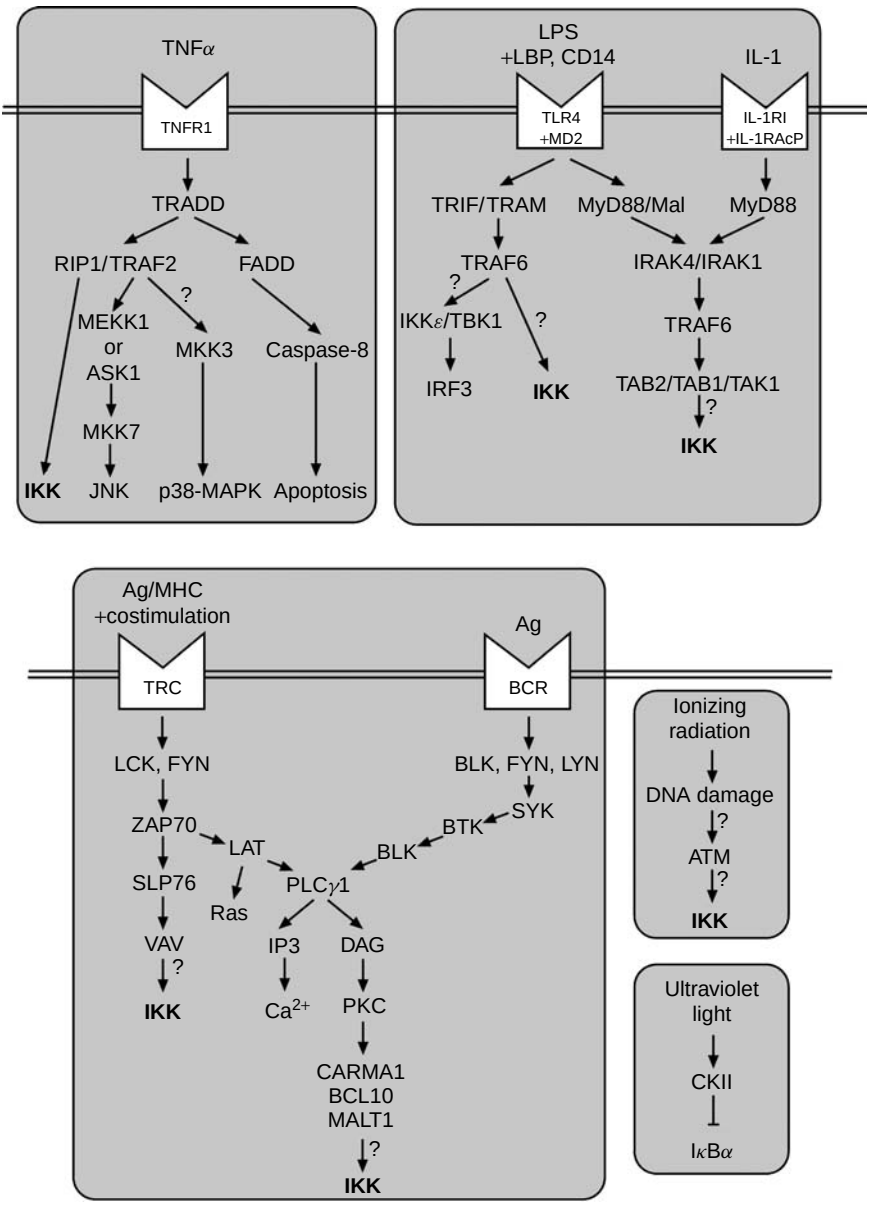

Fig. 5 NF-κB-dependent signaling pathways. The pathways from TNF, LPS, IL-1, antigens, ionizing radiation, and ultraviolet light to NF-κB are shown. Pathways that end in IKK continue via the canonical pathway. Indirect connections are indicated by question marks. See text for description.

in an inactive state, to be displaced by TRADD (TNFR-associated death domain). TRADD acts as a scaffold to recruit TRAF2 (TNFR-associated factor 2) and RIP (receptor interacting protein). In a simple model, TRAF2 further recruits IKK to the receptor with RIP required to activate IKK. However, the nature of interaction is more complex, as TRAF2 appears to destabilize IKK, whereas RIP may counter this action, and the kinase activity of RIP is not required for IKK activation. Nevertheless, somehow the RIP/TRAF2/IKK interaction appears to be sufficient for IKK activation. A few groups have proposed that RIP and TRAF2 recruit another kinase(s) capable of IKK activation, with the leading candidates being MEKK1 and MEKK3. However, neither has been shown to be absolutely essential in knockout studies. The lack of a genetically verified IKK kinase has lead to the hypothesis that proximity-induced autophosphorylation may instead activate IKK. Active IKK then phosphorylates IκB, stimulating the canonical pathway as previously discussed.

In addition to signaling via NF-κB, TNFR1 can activate at least three other pathways. One is the extrinsic apoptotic pathway, which for TNFR1 is widely believed to be mediated by TRADD. TRADD may bind FADD via their homologous death domains, and FADD can recruit the apoptosis initiator caspase-8 to the receptor to form the DISC (death inducing signaling complex). Now in close proximity on the DISC, the caspases molecules can cleave and activate each other. This initiates a cascade of proteolysis, as caspase-8 can proteolytically activate other caspase family members, which in turn cleave a wide assortment of target proteins, ultimately leading to apoptosis. It should be noted, however, that while the DISC has been isolated on *bona fide* death receptors like Fas, the same cannot be said about TNFR1, raising the question whether TNFR1-DISC is unstable or whether apoptosis is mediated through some alternative mechanism(s).

The second pathway clearly activated by TNFR1 is JNK (c-Jun N-terminal kinase). This may occur by either of two cascade mechanisms: TRAF2-dependent activation of the MAP 3-kinase MEKK1 phosphorylating MKK7, which in turn, phosphorylates JNK, or TRAF2-dependent activation of ASK1 phosphorylating MKK7, which phosphorylates JNK. JNK can then regulate other molecules by its kinase activity, most notably activating the c-Jun transcription factor. Though the evidence is somewhat equivocal, JNK appears to have a generally proapoptotic effect. For example, JNK has been implicated in the activation of proapoptotic molecules, such as Smac, Bim, and Bmf and the inhibition of antiapoptotic molecules such as cIAP, but the respective mechanisms are poorly characterized.

Finally, TNFR1 can robustly activate the p38 (a MAPK) pathway. This pathway appears to pass through TRAF2/RIP and the MAP 3-kinase MKK3. p38 has roles in mediating the inflammatory response, but may also antagonize the actions of NF-κB. This raises intriguing questions about how the NF-κB and p38 pathways interact, but additional work in this area is required.

One interesting aspect of TNFR1 signaling is that its stimulation can result in two entirely opposite effects. Under different circumstances, and depending on cell type, it can stimulate cell proliferation or induce apoptosis. Although normally TNFα binding does not lead to apoptosis, in situations in which NF-κB activity is diminished, this potential effect becomes apparent. This is most aptly demonstrated by $p65^{-/-}$ knockout mice

that suffer embryonic lethality due to massive liver apoptosis, a phenotype that can be rescued by also knocking out TNFR1 or TNF. These two different effects are probably a result of the ability of TNFR1 to activate both the antiapoptotic NF-κB pathways, the directly apoptotic DISC pathway, and the conditionally apoptotic JNK pathway, thereby raising important questions about how these pathways interact. NF-κB activates transcription of many survival factors including c-FLIP and IAPs, and these inhibit apoptosis at the level of the DISC. Likewise, other target genes of NF-κB, such as GADD45β and XIAP also inhibit JNK activation. On the other side, multiple components of the canonical pathway can be cleaved by effector caspases. So, while TNFα primarily elicits a proinflammatory response and not apoptosis, the cross talk between the various TNFR1 signaling pathways may function as a switch, which under some circumstances, commits the cell to death.

6.2 LPS

Lipopolysaccharide (LPS) is a component of the outer membrane of gram-negative bacteria. Systemic infections by such bacteria lead to accumulation and disseminated exposure to this potent toxin, which initiates an overexuberant inflammatory response ultimately leading to vasodilation, a severe drop in blood pressure, and septic shock. LPS is also a well-known activator of NF-κB, whose upregulation of many humoral mediators contributes significantly to this syndrome. Understandably, the signaling initiated by LPS is of major interest in infectious disease and immunology.

The receptor for LPS is Toll-like receptor-4 (TLR4), which was firmly established in 1998 by the discovery that C3H/HeJ mice that were resistant to LPS contained a single point mutation in the intracellular domain of TLR4. TLR4 is homologous to Toll, a *Drosophila* receptor tied to a very similar signaling pathway. More generally, the Toll-like receptor family (TLR) recognizes the so-called pathogen associated molecular patterns (PAMPs) and have an evolutionarily conserved function to combat infectious disease. TLRs are, in turn, part of the IL-1/TLR superfamily, which explains the great overlap between LPS and IL-1 signaling described later.

LPS does not bind its receptor by itself; rather several extracellular proteins are involved. In the simplest scheme, LPS is recognized in the bloodstream by LPS-binding protein (LBP). Soluble or membrane-bound CD14 then recognizes the LPS–LBP complex, which promotes loading of LPS onto the receptor. Furthermore, TLR-4 requires MD-2 to be associated with it to transduce LPS signals. Other additional molecules such as HSP70, HSP90, CD11, CD18, CD55, chemokine receptor 4, and growth differentiation factor-5 have also been implicated as components of the LPS receptor complex but their roles are unknown at this time. In this context, then, LPS induces TLR4 to dimerize and initiates signaling.

TLR4 can signal through at least two pathways, termed MyD88-dependent or MyD88-independent, both of which can activate NF-κB via IKK. MyD88 (myeloid differentiation factor 88) was the earliest known adaptor protein for TLR4 and is able to bind the receptor through homotypic interaction of the TIR (Toll/IL-1R) domains on each protein. In MyD88-dependent signaling, a second essential adaptor called Mal (MyD88 adaptor-like) is also required, but its exact role is unclear. Nevertheless, MyD88 binds IRAK-1 (IL-R1 associated

kinase-1) and promotes its activation by IRAK-4 mediated phosphorylation. Active IRAK-1 then leaves the receptor complex and binds to TRAF6, which in turn binds to TAB-2. The TRAF6/TAB-2 complex activates the adaptor-MAP kinase complex TAB-1/TAK-1. Finally, TAK-1 has not only been implicated in the activation of IKK, and thereby in NF-κB, but also in the JNK and p38 pathways. Alternatively, it has been proposed that TRAF6 activates IKK via other MAPK cascade members. However, the proposed links between TRAF6 and IKK have been criticized for lack of genetic experiments clearly demonstrating the role of these components in IKK activation.

TLR4 can also activate NF-κB via the MyD88-independent pathway, albeit in a delayed and much less potent fashion. This pathway relies on the TLR4 adaptors TRIF (TIR domain-containing adaptor inducing interferon β) and TRAM (TRIF-related adaptor molecule). The subsequent downstream events are unclear, but it has been suggested that IKK activation is again dependent on TRAF6. Also, as the name suggests, TRIF and TRAM can activate interferon-β expression. The mechanism is likewise unclear, but appears to involve IKKε and TBK-1 activation of IRF3, a transcription factor whose targets include interferon-β. Interestingly, p65 has also been implicated in the interferon arm of the MyD88-independent pathway. Thus, LPS is intimately linked to NF-κB by several mechanisms and therefore NF-κB plays a key role in the host response to pathogens.

6.3 IL-1

Interleukin-1 (IL-1) is the founding member of an ever-increasing family of cytokines that modulate inflammation and the immune response. IL-1 is primarily produced by macrophages and mediates a variety of physiologic effects, including stimulating fever and inducing synthesis of a variety of inflammatory mediators and adhesion molecules. It has a particularly prominent role in rheumatologic disease. The cloning of IL-1 cDNA in 1984 opened the door to understanding, on a molecular level, how IL-1 mediates so many different effects, and here we will briefly consider how IL-1 signals through NF-κB.

IL-1 actually represents a family of three proteins, IL-1α, IL-1β, and IL-1Ra. IL-1α and IL-1β both bind to the type I IL-1 receptor (IL-1RI) and have nearly identical effects, and indeed prior to their cloning the presence of two different molecules was a source of great confusion about whether one molecule could be responsible for such a broad range of effects. IL-1α is an acidic form not usually secreted into the extracellular environment, whereas IL-1β is a neutral form that is found in the bloodstream. For this reason, IL-1 generally refers to IL-1β unless otherwise noted. IL-1Ra also binds to IL-1R and acts as a receptor antagonist.

IL-1RI is the signaling receptor for IL-1 and is the founding member of the IL-1R/TLR superfamily. This superfamily is linked by a common intracellular TIR domain that mediates binding to TIR-containing adaptors. Members of the IL-1R subfamily contain three extracellular immunoglobulin domains, which differ from the extracellular leucine-rich region in the TLR subfamily. IL-1RI alone is not sufficient for IL-1 signaling, as IL-1RAcP (IL-1R accessory protein), also a member of the IL-1R/TLR superfamily, is required. However, IL-1 only binds IL-1RI, and it is not clear if IL-1RI is constitutively or inducibly associated with IL-1RAcP. In

any case, IL-1 causes trimerization of IL-1RI and initiates a signaling cascade to NF-κB identical to that of the MyD88-dependent pathway TLR4 (see above), with the important exception that IL-1 signaling does not require the Mal adaptor. IL-18, whose receptor is also a member of the IL-1R/TLR superfamily, likewise signals through the same MyD88 pathway. So, it is through these conserved pathways that NF-κB once again can be seen as a central mediator of the inflammatory response.

6.4 Antigens

The adaptive immune response is initiated when an antigen drives activation and proliferation of specific T and B lymphocytes. These cells cooperate to produce antibodies against the antigen and target and kill cells displaying such antigen, thereby mounting a specific host defense against the antigenic source. Antigens activate T and B cells through TCR (T-cell receptor) and BCR (B-cell receptor), respectively, and both of these receptors activate NF-κB, among other pathways. Thus, NF-κB is intimately linked with the adaptive immune response, and combined with the above descriptions of the role of NF-κB in inflammation and the innate immune response, one can see why NF-κB is often referred to as a central regulator of the immune system as a whole.

We first address connections between TCR and NF-κB, remembering that the pathways relaying BCR to NF-κB are quite similar. The TCR is activated when it binds to antigen held by an MHC (major histocompatibility complex) on the surface of an antigen-presenting cell (such as dendritic cells or macrophages) in the presence of costimulation, for example, B7 binding to its T-cell surface receptor CD28. TCR lacks intracellular enzymatic activity, and like TNFR1, IL-1R, and TLR4, requires adaptors to initiate signaling. Engagement of TCR causes tyrosine phosphorylation of its ITAM (immunoreceptor tyrosine-based activation motifs) by the Src-like kinases Lck and Fyn. This allows TCR to recruit ZAP-70 (TCR zeta-chain associated protein kinase, 70 kDa), another tyrosine kinase. At this point, there are two proposed routes for the signal to reach IKK and subsequently NF-κB. One route is that ZAP-70 phosphorylates the adaptor SLP-76 (SH2-domain containing leukocyte protein, 76 kDa), which in turn activates Vav, a GTP exchange factor. Vav then activates IKK through an unknown mechanism. The other pathway emanating from ZAP-70 goes through LAT (linker for activation of T cells), an adaptor that activates PLCγ1 (phospholipase C). PLCγ1 acts through its classical pathway, generating IP3 and DAG from the phospholipid PIP2. DAG can activate PKCθ (protein kinase C), an essential factor for NF-κB activation by the TCR. PKCθ then signals through CARMA1, Bcl10, and MALT1 in sequence, and MALT1 activates IKK through an unknown mechanism. IKK activates NF-κB through a canonical mechanism, which helps upregulate target genes such as IL-2 (a T-cell growth factor), antiapoptotic factors, and cell proliferation factors such as cyclin D, all of which promote T-cell activation.

With such a complex and indirect signaling pathway to NF-κB, it is not surprising that TCR can concomitantly robustly activate other major signal transduction systems. For example, IP3 produced by PLCγ1 causes release of calcium from intracellular stores, thereby activating the calcineurin-NFAT pathway. Likewise, LAT contains Grb-2 binding sites that may provide a link between the TCR and the Ras

pathway. Both the NFAT and Ras pathways cooperate with NF-κB to activate, among other targets, IL-2 and promote the growth of T cells specific to the encountered antigen.

The B-cell receptor is likewise important for B-cell development and activation, and NF-κB is one important pathway activated by BCR. As alluded to above, this activation occurs through a mechanism analogous to that of the TCR. An antigen molecule can bind to the BCR directly and induce its aggregation, which stimulates tyrosine phosphorylation of its ITAM motifs by Src-family kinases like Blk, Fyn, and Lyn. This recruits the tyrosine kinase Syk to BCR, and subsequently Btk via a PI3K-dependent mechanism. Btk activates the adaptor BLNK, which recruits PLCγ2 to the membrane. PLCγ2 mediates the production of DAG, which activates PKC. There are some questions about which PKC is activated, with possibilities including PKCβ or PKCλ, but these different observations may be accounted for by differences in B-cell maturation. PKC subsequently acts through CARMA1, Bcl10, and MALT1 to activate IKK, which finally activates NF-κB. This pathway, like that of the TCR, branches at several points to activate calcium, ERK, JNK, and p38 MAPK-dependent pathways, among others, hinting at the substantial complexity of the BCR signaling network.

6.5 Stress

Stress is an amorphous concept because stressors often lack tangible molecular receptors, in contrast to the pathways discussed above. Additionally, the stress itself may be intangible (e.g. radiation), come in multiple forms (e.g. various chemical oxidants), or is seemingly very nonspecific (e.g. osmotic stress). Nonetheless, as cells are routinely subject to stressful environments or stimuli, understanding the nature and response cell stress is of undeniable importance in cell physiology. It is additionally important from a clinical point of view in both pathophysiology and treatment, raising questions, for example, as to how oxidation contributes to major diseases like atherosclerosis, or how ionizing radiation (IR) therapy affects cancerous versus normal tissue. NF-κB is induced by a wide variety of apparently stressful stimuli and upregulates genes whose products might be useful in weathering the stress. In this section, we examine the most studied NF-κB-inducing stressors: oxidation, ionizing radiation, and ultraviolet light.

Oxidants like hydrogen peroxide were long known to activate NF-κB and early hypotheses even implicated oxygen radicals as second messengers in the TNF and IL-1 signaling pathways. However, increased understanding of the components of these pathways in the 1990s led to a consensus that oxygen radicals did not play a central role in NF-κB signaling. Furthermore, the importance of radical oxygen species (ROS) in activating NF-κB itself came under criticism, as effects of oxidant on NF-κB were found to be cell type- or stimulus-specific and the known components of NF-κB-regulating signaling pathways were found to be responsive to the redox state of the cell. Sadly, little progress has been made on these fronts and there is no clear mechanism by which the cell senses ROS and relays this information to NF-κB. Additionally, although many antioxidants are known to inhibit NF-κB, their effect may not be due to decreased upstream generation of ROS. Rather, they may instead target NF-κB signaling components. At this time, it is unclear what the relative importance of

ROS in stimulating NF-κB might be, although it is likely that ROS merely plays a facilitating role by modulating components of various signaling pathways.

Ionizing radiation has also long been known to activate NF-κB. Since NF-κB has antiapoptotic effects, the IR-NF-κB connection is of great interest to oncologists studying radiosensitivity, that is, the ability of IR to kill some cells and not others. One model of IR signaling to NF-κB that has been put forth involves ATM, a PI3-kinase mutated in the human disease ataxia-telangiectasia, in which afflicted patients are hypersensitive to radiation. ATM might sense IR by direct or indirect activation following DNA damage, and active ATM appears to be essential to NF-κB activity in response to IR. NF-κB activation by ATM clearly depends on IκBα phosphorylation, but the exact mechanism is unclear; some evidence for ATM directly phosphorylating IκBα and activation of IKK via an MAPK has been accumulated. Additionally, this may not be the only way NF-κB responds to genotoxic stress, as ATM-independent connections have been described.

Ultraviolet light is yet another atypical stimulus long known to activate NF-κB. This is of research interest for such disparate topics as the role of UV light in skin cancer and UV-induced transcription of HIV-1. Additionally, UV-induced DNA damage is often used as a model system for genotoxic stress. However, this is questionable as enucleation experiments have shown that DNA damage is not required for NF-κB activation for any UV wavelength. This has led to postulation that UV light affects NF-κB through the generation of ROS but this suggestion is difficult to evaluate given the existing problems in studying ROS (see above). Rather, it appears that UV light acts through a p38-MAPK dependent mechanism to stimulate CKII phosphorylation of the C-terminus of IκBα, a region known to regulate constitutive turnover of IκBα. This makes UV light particularly special, as it joins the very few NF-κB inducers that are known to act independently of IKK.

7 Other Roles of NF-κB in Normal Biology and Pathology

Given the pleiotropic functions of NF-κB, as well as its involvement in a number of key pathways, and detailed knowledge of many such pathways, it is not surprising that NF-κB is of interest to multiple fields of biomedical research. Indeed, a PubMed search (Nov 2004) reveals over 16 000 articles published about NF-κB. Clearly NF-κB is and has long been of great interest to immunologists, and the idea that it is a central regulator of the immune response has already been detailed in the previous sections and so will not be further discussed here. NF-κB is also of interest in virtually any disease with an inflammatory or cell death component. Here, we focus on the role of NF-κB in cancer, which is a good example of how the various properties of NF-κB have pathological importance. To get an idea of the wide variety of research now underway, we also examine the pharmacologic attempts to inhibit NF-κB and the more unusual role of NF-κB in the neurobiology of memory. Together, these examples hint at the broad scope of scientific research that NF-κB has ignited.

7.1 NF-κB in Cancer

The role of NF-κB in cancer–whether it is a tumor promoter or suppressor – is of

great controversy. Evidence of NF-κB as a tumor suppressor comes in two general categories. First, NF-κB is clearly required for the development and function of the adaptive immune response, one function of which is to survey the body for preneoplastic cells and eliminate them ("immune surveillance"). Second, although NF-κB is considered generally antiapoptotic, there are some circumstances in which activated NF-κB plays a proapoptotic role. In such circumstances, it would be disadvantageous to inhibit NF-κB as it could encourage tumor proliferation; for example, melanoma cells appear to be protected from apoptosis through downregulation of NF-κB. There has also been a proposal that the tumor suppressor/promoter status of NF-κB is regulated in part by the transcriptional coactivator/corepressor bound to it. However, there is no general consensus as to the mechanisms by which NF-κB can suppress tumor growth, and indeed, it may be cell-type or stimulus-dependent.

Evidence for NF-κB as a tumor promoter comes in many forms. The earliest hint came from classification of NF-κB as a member of the Rel family, as v-rel, another member of this family, is an avian leukemic oncogene. In humans, c-Rel amplifications are found in several types of lymphoma. However, genetic alterations of other subunits are generally not found in tumor tissues, which suggest that dysregulation of NF-κB activity plays a more important role in its oncogenic potential. Indeed, CYLD is a tumor suppressor that negatively regulates NF-κB at the level of TRAF2 and IKK. Its mutation in familial cylindromatosis, a disease of dramatic benign growth of skin appendages, results in upregulation of NF-κB.

Overactivation of NF-κB could promote tumor growth through several mechanisms. First, NF-κB can stimulate cell proliferation, for example, by upregulating cyclin D1 which mediates the G1 to S checkpoint. Likewise, by virtue of its antiapoptotic properties, NF-κB could contribute to neoplastic cell survival, for example, by resisting apoptosis normally induced by the immune surveillance provided by cytotoxic T lymphocytes and natural killer cells. Several oncogenic viruses, especially EBV (via LMP-1), HTLV-1 (via Tax), KSHV (via Pak-1), activate NF-κB and thereby implicate it in mediating oncogenic transformation. Additionally, NF-κB may promote tumor growth by upregulating VEGF needed for angiogenesis and promote metastasis by upregulating adhesion molecules like ICAM-1 needed for attachment to and subsequent penetration of the vasculature. With such a diverse array of potentially tumor supportive functions, it is not surprising that a large number of cancers display constitutive activation of NF-κB.

Exciting new genetic experiments from several laboratories indicate that NF-κB is a critical molecule in linking inflammation to cancer. One group used mice deficient in MyD88, TLR2, or TLR4 to demonstrate that under physiological conditions TLR signaling is required to protect the intestinal epithelium from injury. Since TLR2 and TLR4 signal primarily through NF-κB, this indicates a role for NF-κB in the survival of intestinal epithelium. An accompanying paper by another group demonstrated in a mouse model of ulcerative colitis–associated colon cancer that knocking out IKKβ in myeloid cells led to decreased tumor growth, whereas knocking out IKKβ in intestinal epithelium had no effect on tumor size but did lead to decreased tumor incidence due to increased apoptosis. A parallel study in a mouse model of

liver cancer showed that inhibition of NF-κB by either TNFα-neutralizing antibody treatment or induction of nondegradable "super-repressor" IκBα caused apoptosis of transformed hepatocytes. These studies together paint a picture in which NF-κB is involved in cancer progression and not initiation, in contrast to earlier ideas discussed previously. Specifically, this model of cancer progression involves NF-κB on two levels. First, NF-κB is required in immune cells (especially macrophage lineage) to produce cytokines in response to chronic inflammatory insult. Second, NF-κB is required in transformed epithelial cells to allow those cytokines to signal against apoptosis. While such antiapoptotic function has a normal physiologic role, it appears that it is errantly active during the period in which transformed cells develop into more aggressive forms. If true, this would seem to have clear implications for treatments that could halt the development of inflammation-associated cancers. If this model holds up to further scrutiny, it will be interesting to see what other cancers progress in a similar fashion.

7.2 Pharmacologic Inhibition of NF-κB

Keeping in mind potential side effects, there is ample evidence that downregulating NF-κB could be useful in treating numerous cancers and diseases with a chronic inflammatory component. Thus, finding specific and effective inhibitors of the NF-κB pathways is of intense interest, and indeed, the list of known inhibitory compounds/substances – perhaps even more impressive than the lists of known activators and target genes – numbers nearly 500. Here, we briefly review the major classes of pharmacologic inhibitors that have been applied against NF-κB.

Naturally, at least some anti-inflammatory compounds can be assumed to have negative effects on NF-κB activity. Indeed, glucocorticoids may exert their effects via several possible mechanisms, including upregulation of IκBα expression, competition for coactivators between NF-κB and glucocorticoid receptor, or by repressive crosstalk between the two pathways. The specific points of inhibition of the NF-κB pathway by several other immunosuppressive drugs, like cyclosporine A, FK506, thalidomide, and plant flavonoids are presently not precisely known. On the other hand, NSAIDs (nonsteroidal anti-inflammatory drugs) have more specific roles in the NF-κB pathway regulation, namely, by inhibiting IKK. Aspirin and sodium salicylate specifically inhibit IKKβ by antagonizing its ability to bind ATP; however, these molecules can exert anti-inflammatory effects via other mechanisms as well. Sulfasalazine, used as an NSAID to treat inflammatory bowel disease, inhibits both IKKα and IKKβ, also by reducing ATP binding. Finally, sulindac, an NSAID employed against rheumatic diseases, specifically inhibits IKKβ, but it is not known whether it exerts effects through ATP.

Since NSAIDs also have effects beyond IKK and because nearly all NF-κB-activating stimuli signal through IKK, IKK is an attractive target for further drug development. Specifically, IKKβ inhibitors are extremely sought after since it is the critical IKK complex component in the canonical pathway and much more is known about its role in diseases than IKKα. No fewer than 15 compounds have been isolated or synthesized with micromolar or better IKK binding constants. The very recent demonstration of successful small

interfering RNA (siRNA) treatment of hypercholesterolemia in mice also opens the door to siRNA-targeting of IKK proteins or other NF-κB targets. All of these compounds are in preclinical testing, and if any one of these are eventually approved for human use, the clinical impact could be enormous.

Another strategy for inhibiting NF-κB is by preventing IκB degradation. One successful strategy is inhibition of the 26S proteasome. Bortezomib (Velcade, Millennium Pharmaceutical, Inc.) is such a drug, and was approved in May 2003 for the treatment of refractory multiple myeloma. Work continues to see if bortezomib may have some other antitumor activity that would be useful in treating solid tumors. Four other 26S proteasome inhibitors are also in preclinical testing. It is a bit odd that inhibiting the proteasome, which degrades all polyubiquitinated proteins, would have specific activity against NF-κB. However, potentially more specific inhibitors of the $SCF^{\beta-TrCP}$ ubiquitin ligase complex have not been reported.

Last, but not least, receptor-level inhibitors are an important class of anti-NF-κB drugs. The anti-TNFR antibody etanercept (Enbrel, Immunex Corp.) and the anti-TNFα antibodies infliximab (Remicade, Centocor Inc.) and adulimumab (Humira, Abbott Laboratories) have been recently approved for treatment of rheumatoid arthritis. There is good potential for these drugs to have benefits in other rheumatologic diseases; however, these drugs, especially infliximab, are also associated with possibly fatal sepsis. Anakinra (Kineret, Amgen Inc.), a recombinant form of IL-R1a, was also approved in 2001 for treatment of refractory rheumatoid arthritis, and combination therapy with anti-TNF agents may have synergistic clinical effects. Finally, a rationally designed dominant-negative form of TNF is in preclinical testing, but it remains to be shown what advantages, if any, it has over the already approved drugs.

7.3 Atypical Roles of NF-κB: Example of Learning

In almost every context, NF-κB is thought to exert its effects in the context of inflammation, apoptosis, or stress response. This is also true in the nervous system, where NF-κB plays important roles in multiple sclerosis, excitotoxic neuronal death, and neuronal oxidative stress. Interestingly, however, NF-κB may have additional physiologic functions with implications for learning and memory as discussed below.

Although NF-κB was known to be involved in the functioning of neurons soon after the discovery of NF-κB itself, increased attention to its possible role in learning did not begin until much later, with the discovery that NF-κB is localized in the synapses of many neurons and that it could be induced at physiologic levels of synaptic stimulation. The significance of this discovery is that NF-κB could mediate memory (i.e. long-term potentiation, LTP) by acting as a bridge between short-term stimulation and long-term changes due to alterations in gene expression patterns. Later animal studies hinted that this was indeed the case, as fear memory formation in the amygdala was impaired by the NF-κB inhibitor thalidomide and by κB decoy DNA, whereas memory formation was enhanced by NF-κB-activating histone deacetylase (HDAC) inhibitors. LTP in the hippocampus is likewise dependent on NF-κB activity.

Some parts of the signaling pathway have been identified. The current model of synaptic plasticity (i.e. long-term changes

in a synapse) is that a prolonged glutaminergic stimulation leads to the opening of the synaptic calcium channel NMDA, followed by intracellular calcium initiating a signal transduction cascade that leads to gene expression or modulation. Consistent with this model, both the calcium/calmodulin-dependent kinase CaMKII and the calcium-dependent protease calpain have been implicated in the activation of synaptic NF-κB, which can then travel retrograde to the nucleus and modulate gene transcription. In either case, the mechanism of IκB degradation is unknown. It is also currently unclear which genes NF-κB upregulates to mediate synaptic plasticity, but NF-κB activity is associated with changes in glutaminergic synaptic currents. Additionally, NF-κB's role in LTP is probably limited to specific cell types in the brain since many neurons exhibit constitutive NF-κB activity. Nonetheless, evidence is strong that NF-κB mediates formation of some memories, and many open questions remain about this function. This example gives a flavor of various aspects of NF-κB function yet to be discovered despite nearly 20 years of intense investigation.

Acknowledgment

The authors wish to thank Alexander Hoffmann for useful discussions and NIH for support of this work.

See also Autoantibodies and Autoimmunity; Cytokines: Interleukins; Immune Defence, Cell Mediated; Immunoassays; Immunologic Memory; Immunology; Innate Immunity; Molecular Mediators: Cytokines.

Bibliography

Books and Reviews

Beinke, S., Ley, S.C. (2004) Functions of NF-kappaB1 and NF-kappaB2 in immune cell biology, *Biochem. J.* **382**(Pt 2), 393–409.

Caamano, J., Hunter, C.A. (2002) NF-kappaB family of transcription factors: central regulators of innate and adaptive immune functions, *Clin. Microbiol. Rev.* **15**(3), 414–429.

Chen, F.E., Ghosh, G. (1999) Regulation of DNA binding by Rel/NF-kappaB transcription factors: structural views, *Oncogene* **18**(49), 6845–6852.

Chen, L.F., Greene, W.C. (2004) Shaping the nuclear action of NF-kappaB, *Nat. Rev. Mol. Cell. Biol.* **5**(5), 392–401.

Ghosh, S., Karin, M. (2002) Missing pieces in the NF-kappaB puzzle, *Cell* **109**(Suppl), S81–S96.

Hayden, M.S., Ghosh, S. (2004) Signaling to NF-kappaB, *Genes Dev.* **18**(18), 2195–2224.

Karin, M., Yamamoto, Y., Wang, Q.M. (2004) The IKK NF-kappa B system: a treasure trove for drug development, *Nat. Rev. Drug Discov.* **3**(1), 17–26.

Kumar, A., Takada, Y., Boriek, A.M., Aggarwal, B.B. (2004) Nuclear factor-kappaB: its role in health and disease, *J. Mol. Med.* **82**(7), 434–448.

Mattson, M.P., Culmsee, C., Yu, Z., Camandola, S. (2000) Roles of nuclear factor kappaB in neuronal survival and plasticity, *J. Neurochem.* **74**(2), 443–456.

Pahl, H.L. (1999) Activators and target genes of Rel/NF-kappaB transcription factors, *Oncogene* **18**(49), 6853–6866.

Primary Literature

Aggarwal, B.B. (2003) Signalling pathways of the TNF superfamily: a double-edged sword, *Nat. Rev. Immunol.* **3**(9), 745–756.

Albensi, B.C., Mattson, M.P. (2000) Evidence for the involvement of TNF and NF-kappaB in hippocampal synaptic plasticity, *Synapse* **35**(2), 151–159.

Aleyasin, H., Cregan, S.P., Iyirhiaro, G., O'Hare, M.J., Callaghan, S.M., Slack, R.S., et al. (2004) Nuclear factor-(kappa)B modulates the p53 response in neurons exposed to DNA damage, *J. Neurosci.* **24**(12), 2963–2973.

Amir, R.E., Haecker, H., Karin, M., Ciechanover, A. (2004) Mechanism of processing of the NF-kappa B2 p100 precursor: identification of the specific polyubiquitin chain-anchoring lysine residue and analysis of the role of NEDD8-modification on the SCF(beta-TrCP) ubiquitin ligase, *Oncogene* **23**(14), 2540–2547.

Anrather, J., Csizmadia, V., Soares, M.P., Winkler, H. (1999) Regulation of NF-kappaB RelA phosphorylation and transcriptional activity by p21(ras) and protein kinase Czeta in primary endothelial cells, *J. Biol. Chem.* **274**(19), 13594–13603.

Auphan, N., DiDonato, J.A., Rosette, C., Helmberg, A., Karin, M. (1995) Immunosuppression by glucocorticoids: inhibition of NF-kappa B activity through induction of I kappa B synthesis, *Science* **270**(5234), 286–290.

Auron, P.E., Webb, A.C., Rosenwasser, L.J., Mucci, S.F., Rich, A., Wolff, S.M., et al. (1984) Nucleotide sequence of human monocyte interleukin 1 precursor cDNA, *Proc. Natl. Acad. Sci. U.S.A.* **81**(24), 7907–7911.

Baeuerle, P.A. (1998) IkappaB-NF-kappaB structures: at the interface of inflammation control, *Cell* **95**(6), 729–731.

Baeuerle, P.A., Baltimore, D. (1988) Activation of DNA-binding activity in an apparently cytoplasmic precursor of the NF-kappa B transcription factor, *Cell* **53**(2), 211–217.

Baeuerle, P.A., Baltimore, D. (1996) NF-kappa B: ten years after, *Cell* **87**(1), 13–20.

Ballard, D.W., Bohnlein, E., Lowenthal, J.W., Wano, Y., Franza, B.R., Greene, W.C. (1988) HTLV-I tax induces cellular proteins that activate the kappa B element in the IL-2 receptor alpha gene, *Science* **241**(4873), 1652–1655.

Baltimore, D., Beg, A.A. (1995) DNA-binding proteins. A butterfly flutters by, *Nature* **373**(6512), 287–288.

Barton, G.M., Medzhitov, R. (2003) Toll-like receptor signaling pathways, *Science* **300**(5625), 1524–1525.

Baud, V., Liu, Z.G., Bennett, B., Suzuki, N., Xia, Y., Karin, M. (1999) Signaling by proinflammatory cytokines: oligomerization of TRAF2 and TRAF6 is sufficient for JNK and IKK activation and target gene induction via an amino-terminal effector domain, *Genes Dev.* **13**(10), 1297–1308.

Beg, A.A., Baltimore, D. (1996) An essential role for NF-kappaB in preventing TNF-alpha-induced cell death, *Science* **274**(5288), 782–784.

Beg, A.A., Sha, W.C., Bronson, R.T., Baltimore, D. (1995) Constitutive NF-kappa B activation, enhanced granulopoiesis, and neonatal lethality in I kappa B alpha-deficient mice, *Genes Dev.* **9**(22), 2736–2746.

Beg, A.A., Sha, W.C., Bronson, R.T., Ghosh, S., Baltimore, D. (1995) Embryonic lethality and liver degeneration in mice lacking the RelA component of NF-kappa B, *Nature* **376**(6536), 167–170.

Beinke, S., Ley, S.C. (2004) Functions of NF-kappaB1 and NF-kappaB2 in immune cell biology, *Biochem. J.* **382**(Pt 2), 393–409.

Bender, K., Gottlicher, M., Whiteside, S., Rahmsdorf, H.J., Herrlich, P. (1998) Sequential DNA damage-independent and -dependent activation of NF-kappaB by UV, *EMBO J.* **17**(17), 5170–5181.

Ben-Neriah, Y. (2002) Regulatory functions of ubiquitination in the immune system, *Nat. Immunol.* **3**(1), 20–26.

Berkowitz, B., Huang, D.B., Chen-Park, F.E., Sigler, P.B., Ghosh, G. (2002) The x-ray crystal structure of the NF-kappa B p50.p65 heterodimer bound to the interferon beta -kappa B site, *J. Biol. Chem.* **277**(27), 24694–24700.

Bohuslav, J., Chen, L.F., Kwon, H., Mu, Y., Greene, W.C. (2004) p53 induces NF-kappaB activation by an IkappaB kinase-independent mechanism involving phosphorylation of p65 by ribosomal S6 kinase 1, *J. Biol. Chem.* **279**(25), 26115–26125.

Bowie, A., O'Neill, L.A. (2000) Oxidative stress and nuclear factor-kappaB activation: a reassessment of the evidence in the light of recent discoveries, *Biochem. Pharmacol.* **59**(1), 13–23.

Brach, M.A., Hass, R., Sherman, M.L., Gunji, H., Weichselbaum, R., Kufe, D. (1991) Ionizing radiation induces expression and binding activity of the nuclear factor kappa B, *J. Clin. Invest.* **88**(2), 691–695.

Braddock, M., Quinn, A., Canvin, J. (2004) Therapeutic potential of targeting IL-1 and IL-18 in inflammation, *Expert Opin. Biol. Ther.* **4**(6), 847–860.

Brown, K., Park, S., Kanno, T., Franzoso, G., Siebenlist, U. (1993) Mutual regulation of the transcriptional activator NF-kappa B and its inhibitor, I kappa B-alpha, *Proc. Natl. Acad. Sci. U.S.A.* **90**(6), 2532–2536.

Bulavin, D.V., Fornace, A.J. Jr. (2004) p38 MAP Kinase's emerging role as a tumor suppressor, *Adv. Cancer Res.* **92**, 95–118.

Caamano, J., Hunter, C.A. (2002) NF-kappaB family of transcription factors: central regulators of innate and adaptive immune functions, *Clin. Microbiol. Rev.* **15**(3), 414–429.

Caamano, J.H., Rizzo, C.A., Durham, S.K., Barton, D.S., Raventos-Suarez, C., Snapper, C.M., et al. (1998) Nuclear factor (NF)-kappa B2 (p100/p52) is required for normal splenic microarchitecture and B cell-mediated immune responses, *J. Exp. Med.* **187**(2), 185–196.

Cao, Z., Xiong, J., Takeuchi, M., Kurama, T., Goeddel, D.V. (1996) TRAF6 is a signal transducer for interleukin-1, *Nature* **383**(6599), 443–446.

Cao, Y., Bonizzi, G., Seagroves, T.N., Greten, F.R., Johnson, R., Schmidt, E.V., et al. (2001) IKKalpha provides an essential link between RANK signaling and cyclin D1 expression during mammary gland development, *Cell* **107**(6), 763–775.

Carlotti, F., Dower, S.K., Qwarnstrom, E.E. (2000) Dynamic shuttling of nuclear factor kappa B between the nucleus and cytoplasm as a consequence of inhibitor dissociation, *J. Biol. Chem.* **275**(52), 41028–41034.

Chen, F.E., Ghosh, G. (1999) Regulation of DNA binding by Rel/NF-kappaB transcription factors: structural views, *Oncogene* **18**(49), 6845–6852.

Chen, G., Goeddel, D.V. (2002) TNF-R1 signaling: a beautiful pathway, *Science* **296**(5573), 1634–1635.

Chen, L.F., Greene, W.C. (2004) Shaping the nuclear action of NF-kappaB, *Nat. Rev. Mol. Cell. Biol.* **5**(5), 392–401.

Chen, Y.Q., Ghosh, S., Ghosh, G. (1998) A novel DNA recognition mode by the NF-kappa B p65 homodimer, *Nat. Struct. Biol.* **5**(1), 67–73.

Chen, L.F., Mu, Y., Greene, W.C. (2002) Acetylation of RelA at discrete sites regulates distinct nuclear functions of NF-kappaB, *EMBO J.* **21**(23), 6539–6548.

Chen, F.E., Huang, D.B., Chen, Y.Q., Ghosh, G. (1998) Crystal structure of p50/p65 heterodimer of transcription factor NF-kappaB bound to DNA, *Nature* **391**(6665), 410–413.

Chen, L., Fischle, W., Verdin, E., Greene, W.C. (2001) Duration of nuclear NF-kappaB action regulated by reversible acetylation, *Science* **293**(5535), 1653–1657.

Cheng, J.D., Ryseck, R.P., Attar, R.M., Dambach, D., Bravo, R. (1998) Functional redundancy of the nuclear factor kappa B inhibitors I kappa B alpha and I kappa B beta, *J. Exp. Med.* **188**(6), 1055–1062.

Chiao, P.J., Miyamoto, S., Verma, I.M. (1994) Autoregulation of I kappa B alpha activity, *Proc. Natl. Acad. Sci. U.S.A.* **91**(1), 28–32.

Chilov, D., Kukk, E., Taira, S., Jeltsch, M., Kaukonen, J., Palotie, A., et al. (1997) Genomic organization of human and mouse genes for vascular endothelial growth factor C, *J. Biol. Chem.* **272**(40), 25176–25183.

Cramer, P., Larson, C.J., Verdine, G.L., Muller, C.W. (1997) Structure of the human NF-kappaB p52 homodimer-DNA complex at 2.1 A resolution, *EMBO J.* **16**(23), 7078–7090.

Cusack, J.C. (2003) Rationale for the treatment of solid tumors with the proteasome inhibitor bortezomib, *Cancer Treat. Rev.* **29**(Suppl. 1), 21–31.

Dadke, D., Fryer, B.H., Golemis, E.A., Field, J. (2003) Activation of p21-activated kinase 1-nuclear factor kappaB signaling by Kaposi's sarcoma-associated herpes virus G protein-coupled receptor during cellular transformation, *Cancer Res.* **63**(24), 8837–8847.

Dal Porto, J.M., Gauld, S.B., Merrell, K.T., Mills, D., Pugh-Bernard, A.E., Cambier, J. (2004) B cell antigen receptor signaling 101, *Mol. Immunol.* **41**(6–7), 599–613.

de Lumley, M., Hart, D.J., Cooper, M.A., Symeonides, S., Blackburn, J.M. (2004) A biophysical characterisation of factors controlling dimerization and selectivity in the NF-kappaB and NFAT families, *J. Mol. Biol.* **339**(5), 1059–1075.

Demarchi, F., Bertoli, C., Sandy, P., Schneider, C. (2003) Glycogen synthase kinase-3 beta regulates NF-kappa B1/p105 stability, *J. Biol. Chem.* **278**(41), 39583–39590.

Deng, L., Wang, C., Spencer, E., Yang, L., Braun, A., You, J., et al. (2000) Activation of the IkappaB kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain, *Cell* **103**(2), 351–361.

Devary, Y., Rosette, C., DiDonato, J.A., Karin, M. (1993) NF-kappa B activation by ultraviolet

light not dependent on a nuclear signal, *Science* **261**(5127), 1442–1445.

Devin, A., Cook, A., Lin, Y., Rodriguez, Y., Kelliher, M., Liu, Z. (2000) The distinct roles of TRAF2 and RIP in IKK activation by TNF-R1: TRAF2 recruits IKK to TNF-R1 while RIP mediates IKK activation, *Immunity* **12**(4), 419–429.

DiDonato, J.A., Hayakawa, M., Rothwarf, D.M., Zandi, E., Karin, M. (1997) A cytokine-responsive IkappaB kinase that activates the transcription factor NF-kappaB, *Nature* **388**(6642), 548–554.

Dinarello, C.A. (1994) The interleukin-1 family: 10 years of discovery, *FASEB J.* **8**(15), 1314–1325.

Doi, T.S., Marino, M.W., Takahashi, T., Yoshida, T., Sakakura, T., Old, L.J., et al. (1999) Absence of tumor necrosis factor rescues RelA-deficient mice from embryonic lethality, *Proc. Natl. Acad. Sci. U.S.A.* **96**(6), 2994–2999.

Ducut Sigala, J.L., Bottero, V., Young, D.B., Shevchenko, A., Mercurio, F., Verma, I.M. (2004) Activation of transcription factor NF-kappaB requires ELKS, an IkappaB kinase regulatory subunit, *Science* **304**(5679), 1963–1967.

Dunne, A., O'Neill, L.A. (2003) The interleukin-1 receptor/Toll-like receptor superfamily: signal transduction during inflammation and host defense, *Sci. STKE* **2003**(171), re3.

Eisenberg, S.P., Evans, R.J., Arend, W.P., Verderber, E., Brewer, M.T., Hannum, C.H., et al. (1990) Primary structure and functional expression from complementary DNA of a human interleukin-1 receptor antagonist, *Nature* **343**(6256), 341–346.

Fitzgerald, K.A., McWhirter, S.M., Faia, K.L., Rowe, D.C., Latz, E., Golenbock, D.T., et al. (2003) IKKepsilon and TBK1 are essential components of the IRF3 signaling pathway, *Nat. Immunol.* **4**(5), 491–496.

Fitzgerald, K.A., Rowe, D.C., Barnes, B.J., Caffrey, D.R., Visintin, A., Latz, E., et al. (2003) LPS-TLR4 signaling to IRF-3/7 and NF-kappaB involves the toll adapters TRAM and TRIF, *J. Exp. Med.* **198**(7), 1043–1055.

Fong, A., Sun, S.C. (2002) Genetic evidence for the essential role of beta-transducin repeat-containing protein in the inducible processing of NF-kappa B2/p100, *J. Biol. Chem.* **277**(25), 22111–22114.

Franzoso, G., Bours, V., Park, S., Tomita-Yamaguchi, M., Kelly, K., Siebenlist, U. (1992) The candidate oncoprotein Bcl-3 is an antagonist of p50/NF-kappa B-mediated inhibition, *Nature* **359**(6393), 339–342.

Franzoso, G., Carlson, L., Poljak, L., Shores, E.W., Epstein, S., Leonardi, A., et al. (1998) Mice deficient in nuclear factor (NF)-kappa B/p52 present with defects in humoral responses, germinal center reactions, and splenic microarchitecture, *J. Exp. Med.* **187**(2), 147–159.

Freudenthal, R., Romano, A., Routtenberg, A. (2004) Transcription factor NF-kappaB activation after in vivo perforant path LTP in mouse hippocampus, *Hippocampus* **14**(6), 677–683.

Furia, B., Deng, L., Wu, K., Baylor, S., Kehn, K., Li, H., et al. (2002) Enhancement of nuclear factor-kappa B acetylation by coactivator p300 and HIV-1 Tat proteins, *J. Biol. Chem.* **277**(7), 4973–4980.

Furukawa, K., Mattson, M.P. (1998) The transcription factor NF-kappaB mediates increases in calcium currents and decreases in NMDA-and AMPA/kainate-induced currents induced by tumor necrosis factor-alpha in hippocampal neurons, *J. Neurochem.* **70**(5), 1876–1886.

Garg, A., Aggarwal, B.B. (2002) Nuclear transcription factor-kappaB as a target for cancer drug development, *Leukemia* **16**(6), 1053–1068.

Geng, L., Rudd, C.E. (2002) Signalling scaffolds and adaptors in T-cell immunity, *Br. J. Haematol.* **116**(1), 19–27.

Gerritsen, M.E., Williams, A.J., Neish, A.S., Moore, S., Shi, Y., Collins, T. (1997) CREB-binding protein/p300 are transcriptional coactivators of p65, *Proc. Natl. Acad. Sci. U.S.A.* **94**(7), 2927–2932.

Ghosh, S., Karin, M. (2002) Missing pieces in the NF-kappaB puzzle, *Cell* **109**(Suppl), S81–S96.

Ghosh, S., May, M.J., Kopp, E.B. (1998) NF-kappa B and Rel proteins: evolutionarily conserved mediators of immune responses, *Annu. Rev. Immunol.* **16**, 225–260.

Ghosh, G., van Duyne, G., Ghosh, S., Sigler, P.B. (1995) Structure of NF-kappa B p50 homodimer bound to a kappa B site, *Nature* **373**(6512), 303–310.

Ghosh, S., Gifford, A.M., Riviere, L.R., Tempst, P., Nolan, G.P., Baltimore, D. (1990) Cloning of the p50 DNA binding subunit of NF-kappa B: homology to rel and dorsal, *Cell* **62**(5), 1019–1029.

Gilmore, T.D. *Target Genes of NF-kB [online]*. Available at: http://people.bu.edu/gilmore/nf-kb/target/index.html. Accessed Nov 14 2004.

Gilmore, T.D. *Inducers of NF-kB Activity [Online]*. Available at: http://people.bu.edu/gilmore/nf-kb/inducers/index.html. Accessed Nov 14 2004.

Gilmore, T.D. *Inhibitors of Rel/NF-kB [Online]*. Available at: http://people.bu.edu/gilmore/nf-kb/inhibitors/index.html. Accessed Nov 14 2004.

Gilmore, T.D., Kalaitzidis, D., Liang, M.C., Starczynowski, D.T. (2004) The c-Rel transcription factor and B-cell proliferation: a deal with the devil, *Oncogene* **23**(13), 2275–2286.

Goyarzu, P., Malin, D.H., Lau, F.C., Taglialatela, G., Moon, W.D., Jennings, R., et al. (2004) Blueberry supplemented diet: effects on object recognition memory and nuclear factor-kappa B levels in aged rats, *Nutr. Neurosci.* **7**(2), 75–83.

Greenfeder, S.A., Nunes, P., Kwee, L., Labow, M., Chizzonite, R.A., Ju, G. (1995) Molecular cloning and characterization of a second subunit of the interleukin 1 receptor complex, *J. Biol. Chem.* **270**(23), 13757–13765.

Greten, F.R., Eckmann, L., Greten, T.F., Park, J.M., Li, Z.W., Egan, L.J., et al. (2004) IKKbeta links inflammation and tumorigenesis in a mouse model of colitis-associated cancer, *Cell* **118**(3), 285–296.

Grumont, R.J., Gerondakis, S. (1994) Alternative splicing of RNA transcripts encoded by the murine p105 NF-kappa B gene generates I kappa B gamma isoforms with different inhibitory activities, *Proc. Natl. Acad. Sci. U.S.A.* **91**(10), 4367–4371.

Guttridge, D.C., Albanese, C., Reuther, J.Y., Pestell, R.G., Baldwin, A.S. Jr. (1999) NF-kappaB controls cell growth and differentiation through transcriptional regulation of cyclin D1, *Mol. Cell. Biol.* **19**(8), 5785–5799.

Hammarskjold, M.L., Simurda, M.C. (1992) Epstein-Barr virus latent membrane protein transactivates the human immunodeficiency virus type 1 long terminal repeat through induction of NF-kappa B activity, *J. Virol.* **66**(11), 6496–6501.

Hannum, C.H., Wilcox, C.J., Arend, W.P., Joslin, F.G., Dripps, D.J., Heimdal, P.L., et al. (1990) Interleukin-1 receptor antagonist activity of a human interleukin-1 inhibitor, *Nature* **343**(6256), 336–340.

Harhaj, E.W., Sun, S.C. (1999) Regulation of RelA subcellular localization by a putative nuclear export signal and p50, *Mol. Cell. Biol.* **19**(10), 7088–7095.

Harper, N., Hughes, M., MacFarlane, M., Cohen, G.M. (2003) Fas-associated death domain protein and caspase-8 are not recruited to the tumor necrosis factor receptor 1 signaling complex during tumor necrosis factor-induced apoptosis, *J. Biol. Chem.* **278**(28), 25534–25541.

Haskill, S., Beg, A.A., Tompkins, S.M., Morris, J.S., Yurochko, A.D., Sampson-Johannes, A., et al. (1991) Characterization of an immediate-early gene induced in adherent monocytes that encodes I kappa B-like activity, *Cell* **65**(7), 1281–1289.

Hayakawa, M., Miyashita, H., Sakamoto, I., Kitagawa, M., Tanaka, H., Yasuda, H., et al. (2003) Evidence that reactive oxygen species do not mediate NF-kappaB activation, *EMBO J.* **22**(13), 3356–3366.

Hayden, M.S., Ghosh, S. (2004) Signaling to NF-kappaB, *Genes Dev.* **18**(18), 2195–2224.

Heine, H., El-Samalouti, V.T., Notzel, C., Pfeiffer, A., Lentschat, A., Kusumoto, S., et al. (2003) CD55/decay accelerating factor is part of the lipopolysaccharide-induced receptor complex, *Eur. J. Immunol.* **33**(5), 1399–1408.

Hinz, M., Krappmann, D., Eichten, A., Heder, A., Scheidereit, C., Strauss, M. (1999) NF-kappaB function in growth control: regulation of cyclin D1 expression and G0/G1-to-S-phase transition, *Mol. Cell. Biol.* **19**(4), 2690–2698.

Hoeflich, K.P., Luo, J., Rubie, E.A., Tsao, M.S., Jin, O., Woodgett, J.R. (2000) Requirement for glycogen synthase kinase-3beta in cell survival and NF-kappaB activation, *Nature* **406**(6791), 86–90.

Hoffmann, A., Levchenko, A., Scott, M.L., Baltimore, D. (2002) The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation, *Science* **298**(5596), 1241–1245.

Hu, X. (2003) Proteolytic signaling by TNFalpha: caspase activation and IkappaB degradation, *Cytokine* **21**(6), 286–294.

Huang, T.T., Kudo, N., Yoshida, M., Miyamoto, S. (2000) A nuclear export signal in the N-terminal regulatory domain of IkappaBalpha controls cytoplasmic localization of inactive NF-kappaB/IkappaBalpha complexes, *Proc. Natl. Acad. Sci. U.S.A.* **97**(3), 1014–1019.

Huang, T.T., Wuerzberger-Davis, S.M., Wu, Z.H., Miyamoto, S. (2003) Sequential

modification of NEMO/IKKgamma by SUMO-1 and ubiquitin mediates NF-kappaB activation by genotoxic stress, *Cell* **115**(5), 565–576.

Huang, D.B., Chen, Y.Q., Ruetsche, M., Phelps, C.B., Ghosh, G. (2001) X-ray crystal structure of proto-oncogene product c-Rel bound to the CD28 response element of IL-2, *Structure (Cambridge)* **9**(8), 669–678.

Huang, Q., Yang, J., Lin, Y., Walker, C., Cheng, J., Liu, Z.G., et al. (2004) Differential regulation of interleukin 1 receptor and Toll-like receptor signaling by MEKK3, *Nat. Immunol.* **5**(1), 98–103.

Huxford, T., Huang, D.B., Malek, S., Ghosh, G. (1998) The crystal structure of the IkappaBalpha/NF-kappaB complex reveals mechanisms of NF-kappaB inactivation, *Cell* **95**(6), 759–770.

Imbert, V., Rupec, R.A., Livolsi, A., Pahl, H.L., Traenckner, E.B., Mueller-Dieckmann, C., et al. (1996) Tyrosine phosphorylation of I kappa B-alpha activates NF-kappa B without proteolytic degradation of I kappa B-alpha, *Cell* **86**(5), 787–798.

Inoue, J., Kerr, L.D., Kakizuka, A., Verma, I.M. (1992) I kappa B gamma, a 70 kd protein identical to the C-terminal half of p110 NF-kappa B: a new member of the I kappa B family, *Cell* **68**(6), 1109–1120.

Ivanov, V.N., Ronai, Z. (2000) p38 protects human melanoma cells from UV-induced apoptosis through down-regulation of NF-kappaB activity and Fas expression, *Oncogene* **19**(26), 3003–3012.

Jack, R.S., Fan, X., Bernheiden, M., Rune, G., Ehlers, M., Weber, A., et al. (1997) Lipopolysaccharide-binding protein is required to combat a murine gram-negative bacterial infection, *Nature* **389**(6652), 742–745.

Jacobs, M.D., Harrison, S.C. (1998) Structure of an IkappaBalpha/NF-kappaB complex, *Cell* **95**(6), 749–758.

Jang, M.K., Goo, Y.H., Sohn, Y.C., Kim, Y.S., Lee, S.K., Kang, H., et al. (2001) Ca2+/calmodulin-dependent protein kinase IV stimulates nuclear factor-kappa B transactivation via phosphorylation of the p65 subunit, *J. Biol. Chem.* **276**(23), 20005–20010.

Johnson, C., Van Antwerp, D., Hope, T.J. (1999) An N-terminal nuclear export signal is required for the nucleocytoplasmic shuttling of IkappaBalpha, *EMBO J.* **18**(23), 6682–6693.

Jung, M., Kondratyev, A., Lee, S.A., Dimtchev, A., Dritschilo, A. (1997) ATM gene product phosphorylates I kappa B-alpha, *Cancer Res.* **57**(1), 24–27.

Kaltschmidt, C., Kaltschmidt, B., Baeuerle, P.A. (1993) Brain synapses contain inducible forms of the transcription factor NF-kappa B, *Mech. Dev.* **43**(2–3), 135–147.

Kaltschmidt, C., Kaltschmidt, B., Neumann, H., Wekerle, H., Baeuerle, P.A. (1994) Constitutive NF-kappa B activity in neurons, *Mol. Cell. Biol.* **14**(6), 3981–3992.

Karin, M., Ben-Neriah, Y. (2000) Phosphorylation meets ubiquitination: the control of NF-[kappa]B activity, *Annu. Rev. Immunol.* **18**, 621–663.

Karin, M., Yamamoto, Y., Wang, Q.M. (2004) The IKK NF-kappa B system: a treasure trove for drug development, *Nat. Rev. Drug Discov.* **3**(1), 17–26.

Kato, T. Jr., Delhase, M., Hoffmann, A., Karin, M. (2003) CK2 is a C-terminal IkappaB Kinase responsible for NF-kappaB activation during the UV response, *Mol. Cell* **12**(4), 829–839.

Kiernan, R., Bres, V., Ng, R.W., Coudart, M.P., El Messaoudi, S., Sardet, C., et al. (2003) Post-activation turn-off of NF-kappa B-dependent transcription is regulated by acetylation of p65, *J. Biol. Chem.* **278**(4), 2758–2766.

Kishore, N., Huynh, Q.K., Mathialagan, S., Hall, T., Rouw, S., Creely, D., et al. (2002) IKK-i and TBK-1 are enzymatically distinct from the homologous enzyme IKK-2: comparative analysis of recombinant human IKK-i, TBK-1, and IKK-2, *J. Biol. Chem.* **277**(16), 13840–13847.

Klement, J.F., Rice, N.R., Car, B.D., Abbondanzo, S.J., Powers, G.D., Bhatt, P.H., et al. (1996) IkappaBalpha deficiency results in a sustained NF-kappaB response and severe widespread dermatitis in mice, *Mol. Cell. Biol.* **16**(5), 2341–2349.

Kulms, D., Poppelmann, B., Schwarz, T. (2000) Ultraviolet radiation-induced interleukin 6 release in HeLa cells is mediated via membrane events in a DNA damage-independent way, *J. Biol. Chem.* **275**(20), 15060–15066.

Kumar, A., Takada, Y., Boriek, A.M., Aggarwal, B.B. (2004) Nuclear factor-kappaB: its role in health and disease, *J. Mol. Med.* **82**(7), 434–448.

Lang, V., Janzen, J., Fischer, G.Z., Soneji, Y., Beinke, S., Salmeron, A., et al. (2003)

betaTrCP-mediated proteolysis of NF-kappaB1 p105 requires phosphorylation of p105 serines 927 and 932, *Mol. Cell. Biol.* **23**(1), 402–413.

Lee, S.H., Hannink, M. (2001) The N-terminal nuclear export sequence of IkappaBalpha is required for RanGTP-dependent binding to CRM1, *J. Biol. Chem.* **276**(26), 23599–23606.

Lee, S.H., Hannink, M. (2002) Characterization of the nuclear import and export functions of Ikappa B(epsilon), *J. Biol. Chem.* **277**(26), 23358–23366.

Lee, S.J., Dimtchev, A., Lavin, M.F., Dritschilo, A., Jung, M. (1998) A novel ionizing radiation-induced signaling pathway that activates the transcription factor NF-kappaB, *Oncogene* **17**(14), 1821–1826.

Legrand-Poels, S., Schoonbroodt, S., Matroule, J.Y., Piette, J. (1998) Nf-kappa B: an important transcription factor in photobiology, *J. Photochem. Photobiol. B* **45**(1), 1–8.

Leitges, M., Sanz, L., Martin, P., Duran, A., Braun, U., Garcia, J.F., et al. (2001) Targeted disruption of the zetaPKC gene results in the impairment of the NF-kappaB pathway, *Mol. Cell* **8**(4), 771–780.

Li, N., Karin, M. (1998) Ionizing radiation and short wavelength UV activate NF-kappaB through two distinct mechanisms, *Proc. Natl. Acad. Sci. U.S.A.* **95**(22), 13012–13017.

Li, N., Karin, M. (1999) Is NF-kappaB the sensor of oxidative stress? *FASEB J.* **13**(10), 1137–1143.

Li, Z., Nabel, G.J. (1997) A new member of the I kappaB protein family, I kappaB epsilon, inhibits RelA (p65)-mediated NF-kappaB transcription, *Mol. Cell. Biol.* **17**(10), 6184–6190.

Li, Z.W., Rickert, R.C., Karin, M. (2004) Genetic dissection of antigen receptor induced-NF-kappaB activation, *Mol. Immunol.* **41**(6–7), 701–714.

Li, S., Strelow, A., Fontana, E.J., Wesche, H. (2002) IRAK-4: a novel member of the IRAK family with the properties of an IRAK-kinase, *Proc. Natl. Acad. Sci. U.S.A.* **99**(8), 5567–5572.

Li, Q., Van Antwerp, D., Mercurio, F., Lee, K.F., Verma, I.M. (1999) Severe liver degeneration in mice lacking the IkappaB kinase 2 gene, *Science* **284**(5412), 321–325.

Li, N., Banin, S., Ouyang, H., Li, G.C., Courtois, G., Shiloh, Y., et al. (2001) ATM is required for IkappaB kinase (IKKk) activation in response to DNA double strand breaks, *J. Biol. Chem.* **276**(12), 8898–8903.

Lin, L., Ghosh, S. (1996) A glycine-rich region in NF-kappaB p105 functions as a processing signal for the generation of the p50 subunit, *Mol. Cell. Biol.* **16**(5), 2248–2254.

Luo, J.L., Maeda, S., Hsu, L.C., Yagita, H., Karin, M. (2004) Inhibition of NF-kappaB in cancer cells converts inflammation- induced tumor growth mediated by TNFalpha to TRAIL-mediated tumor regression, *Cancer Cell* **6**(3), 297–305.

Martin-Blanco, E. (2000) p38 MAPK signalling cascades: ancient roles and new functions, *Bioessays* **22**(7), 637–645.

Madrid, L.V., Mayo, M.W., Reuther, J.Y., Baldwin, A.S. Jr. (2001) Akt stimulates the transactivation potential of the RelA/p65 Subunit of NF-kappa B through utilization of the Ikappa B kinase and activation of the mitogen-activated protein kinase p38, *J. Biol. Chem.* **276**(22), 18934–18940.

Malek, S., Huang, D.B., Huxford, T., Ghosh, S., Ghosh, G. (2003) X-ray crystal structure of an IkappaBbeta x NF-kappaB p65 homodimer complex, *J. Biol. Chem.* **278**(25), 23094–23100.

Malinin, N.L., Boldin, M.P., Kovalenko, A.V., Wallach, D. (1997) MAP3K-related kinase involved in NF-kappaB induction by TNF, CD95 and IL-1, *Nature* **385**(6616), 540–544.

Martindale, J.L., Holbrook, N.J. (2002) Cellular response to oxidative stress: signaling for suicide and survival, *J. Cell. Physiol.* **192**(1), 1–15.

Mattson, M.P., Culmsee, C., Yu, Z., Camandola, S. (2000) Roles of nuclear factor kappaB in neuronal survival and plasticity, *J. Neurochem.* **74**(2), 443–456.

McKay, L.I., Cidlowski, J.A. (1998) Cross-talk between nuclear factor-kappa B and the steroid hormone receptors: mechanisms of mutual antagonism, *Mol. Endocrinol.* **12**(1), 45–56.

Meberg, P.J., Kinney, W.R., Valcourt, E.G., Routtenberg, A. (1996) Gene expression of the transcription factor NF-kappa B in hippocampus: regulation by synaptic activity, *Brain Res. Mol. Brain Res.* **38**(2), 179–190.

Meffert, M.K., Chang, J.M., Wiltgen, B.J., Fanselow, M.S., Baltimore, D. (2003) NF-kappa B functions in synaptic signaling and behavior, *Nat. Neurosci.* **6**(10), 1072–1078.

Memet, S., Laouini, D., Epinat, J.C., Whiteside, S.T., Goudeau, B., Philpott, D., et al. (1999) IkappaBepsilon-deficient mice: reduction of one T cell precursor subspecies and

enhanced Ig isotype switching and cytokine synthesis, *J. Immunol.* **163**(11), 5994–6005.

Miyamoto, S., Schmitt, M.J., Verma, I.M. (1994) Qualitative changes in the subunit composition of kappa B-binding complexes during murine B-cell differentiation, *Proc. Natl. Acad. Sci. U.S.A.* **91**(11), 5056–5060.

Miyawaki, S., Nakamura, Y., Suzuka, H., Koba, M., Yasumizu, R., Ikehara, S., et al. (1994) A new mutation, aly, that induces a generalized lack of lymph nodes accompanied by immunodeficiency in mice, *Eur. J. Immunol.* **24**(2), 429–434.

Monia, B.P., Cowsert, L.M. inventors; Isis Pharmaceuticals Inc., assignee. Antisense modulation of inhibitor-kappa B kinase-alpha expression. U.S. Pat. No. 5,962,673 patent 5,962,673. 1998 Oct 5 1999.

Monia, B.P., Cowsert, L.M. inventors; Isis Pharmaceuticals Inc., assignee. Antisense modulation of inhibitor-kappa B kinase-beta expression. U.S. Pat. No. 5,977,341 patent 5,977,341. 1998 Nov 2 1999.

Moore, B.E., Bose, H.R. Jr. (1988) Transformation of avian lymphoid cells by reticuloendotheliosis virus, *Mutat. Res.* **195**(1), 79–90.

Mosialos, G., Gilmore, T.D. (1993) v-Rel and c-Rel are differentially affected by mutations at a consensus protein kinase recognition sequence, *Oncogene* **8**(3), 721–730.

Muller, C.W., Rey, F.A., Sodeoka, M., Verdine, G.L., Harrison, S.C. (1995) Structure of the NF-kappa B p50 homodimer bound to DNA, *Nature* **373**(6512), 311–317.

Muta, T., Yamazaki, S., Eto, A., Motoyama, M., Takeshige, K. (2003) IkappaB-zeta, a new anti-inflammatory nuclear protein induced by lipopolysaccharide, is a negative regulator for nuclear factor-kappaB, *J. Endotoxin Res.* **9**(3), 187–191.

Nagai, Y., Akashi, S., Nagafuku, M., Ogata, M., Iwakura, Y., Akira, S., et al. (2002) Essential role of MD-2 in LPS responsiveness and TLR4 distribution, *Nat. Immunol.* **3**(7), 667–672.

Nelson, D.E., Ihekwaba, A.E., Elliott, M., Johnson, J.R., Gibney, C.A., Foreman, B.E., et al. (2004) Oscillations in NF-kappaB signaling control the dynamics of gene expression, *Science* **306**(5696), 704–708.

Naumann, M., Scheidereit, C. (1994) Activation of NF-kappa B in vivo is regulated by multiple phosphorylations, *EMBO J.* **13**(19), 4597–4607.

Neumann, M., Grieshammer, T., Chuvpilo, S., Kneitz, B., Lohoff, M., Schimpl, A., et al. (1995) RelA/p65 is a molecular target for the immunosuppressive action of protein kinase a. *EMBO J.* **14**(9), 1991–2004.

Nicholson, D.W. (1999) Caspase structure, proteolytic substrates, and function during apoptotic cell death, *Cell Death Differ.* **6**(11), 1028–1042.

Norman, D.A., Barton, P.J. (2000) Isolation, sequence, and chromosomal localisation of the human IkappaBR gene (NFKBIL2), *Ann. Hum. Genet.* **64**(Pt 1), 15–23.

O'Neill, E.A. (1998) A new target for aspirin, *Nature* **396**(6706), 15,17.

Oshiumi, H., Sasai, M., Shida, K., Fujita, T., Matsumoto, M., Seya, T. (2003) TIR-containing adapter molecule (TICAM)-2, a bridging adapter recruiting to toll-like receptor 4 TICAM-1 that induces interferon-beta, *J. Biol. Chem.* **278**(50), 49751–49762.

Pahl, H.L. (1999) Activators and target genes of Rel/NF-kappaB transcription factors, *Oncogene* **18**(49), 6853–6866.

Palsson-McDermott, E.M., O'Neill, L.A. (2004) Signal transduction by the lipopolysaccharide receptor, Toll-like receptor-4, *Immunology* **113**(2), 153–162.

Panta, G.R., Kaur, S., Cavin, L.G., Cortes, M.L., Mercurio, F., Lothstein, L., et al. (2004) ATM and the catalytic subunit of DNA-dependent protein kinase activate NF-kappaB through a common MEK/extracellular signal-regulated kinase/p90(rsk) signaling pathway in response to distinct forms of DNA damage, *Mol. Cell. Biol.* **24**(5), 1823–1835.

Parkin, J., Cohen, B. (2001) An overview of the immune system, *Lancet* **357**(9270), 1777–1789.

Perera, P.Y., Mayadas, T.N., Takeuchi, O., Akira, S., Zaks-Zilberman, M., Goyert, S.M., et al. (2001) CD11b/CD18 acts in concert with CD14 and Toll-like receptor (TLR) 4 to elicit full lipopolysaccharide and taxol-inducible gene expression, *J. Immunol.* **166**(1), 574–581.

Perkins, N.D. (2004) NF-kappaB: tumor promoter or suppressor? *Trends Cell Biol.* **14**(2), 64–69.

Peters, R.T., Liao, S.M., Maniatis, T. (2000) IKKepsilon is part of a novel PMA-inducible IkappaB kinase complex, *Mol. Cell* **5**(3), 513–522.

Piret, B., Schoonbroodt, S., Piette, J. (1999) The ATM protein is required for sustained

activation of NF-kappaB following DNA damage, *Oncogene* **18**(13), 2261–2271.

Poltorak, A., He, X., Smirnova, I., Liu, M.Y., Van Huffel, C., Du, X., et al. (1998) Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene, *Science* **282**(5396), 2085–2088.

Pomerantz, J.L., Baltimore, D. (2002) Two pathways to NF-kappaB, *Mol. Cell* **10**(4), 693–695.

Qing, G., Qu, Z., Xiao, G. (2005) Regulation of NF-kappaB2 p100 processing by its cis-acting domain, *J. Biol. Chem.* **280**(1), 18–27.

Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., Medzhitov, R. (2004) Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis, *Cell* **118**(2), 229–241.

Ravi, R., Bedi, A. (2004) NF-kappaB in cancer–a friend turned foe, *Drug Resist. Updat.* **7**(1), 53–67.

Ray, P., Zhang, D.H., Elias, J.A., Ray, A. (1995) Cloning of a differentially expressed I kappa B-related protein, *J. Biol. Chem.* **270**(18), 10680–10685.

Ray, P., Yang, L., Zhang, D.H., Ghosh, S.K., Ray, A. (1997) Selective up-regulation of cytokine-induced RANTES gene expression in lung epithelial cells by overexpression of IkappaBR, *J. Biol. Chem.* **272**(32), 20191–20197.

Redlich, K., Schett, G., Steiner, G., Hayer, S., Wagner, E.F., Smolen, J.S. (2003) Rheumatoid arthritis therapy after tumor necrosis factor and interleukin-1 blockade, *Arthritis Rheum.* **48**(12), 3308–3319.

Reed, J.C., Huang, Z. *Apoptosis Pathways and Drug Targets [Online]*. Available at: http://www.nature.com/reviews/poster/apoptosis/. Accessed Nov 18 2004.

Rosenfeld, M.E., Prichard, L., Shiojiri, N., Fausto, N. (2000) Prevention of hepatic apoptosis and embryonic lethality in RelA/TNFR-1 double knockout mice, *Am. J. Pathol.* **156**(3), 997–1007.

Rudolph, D., Yeh, W.C., Wakeham, A., Rudolph, B., Nallainathan, D., Potter, J., et al. (2000) Severe liver degeneration and lack of NF-kappaB activation in NEMO/IKKgamma-deficient mice, *Genes Dev.* **14**(7), 854–862.

Ruland, J., Mak, T.W. (2003) Transducing signals from antigen receptors to nuclear factor kappaB, *Immunol. Rev.* **193**, 93–100.

Ryo, A., Suizu, F., Yoshida, Y., Perrem, K., Liou, Y.C., Wulf, G., et al. (2003) Regulation of NF-kappaB signaling by Pin1-dependent prolyl isomerization and ubiquitin-mediated proteolysis of p65/RelA, *Mol. Cell* **12**(6), 1413–1426.

Saccani, S., Pantano, S., Natoli, G. (2003) Modulation of NF-kappaB activity by exchange of dimers, *Mol. Cell* **11**(6), 1563–1574.

Sachdev, S., Hoffmann, A., Hannink, M. (1998) Nuclear localization of IkappaB alpha is mediated by the second ankyrin repeat: the IkappaB alpha ankyrin repeats define a novel class of cis-acting nuclear import sequences, *Mol. Cell. Biol.* **18**(5), 2524–2534.

Saijo, K., Schmedt, C., Su, I.H., Karasuyama, H., Lowell, C.A., Reth, M., et al. (2003) Essential role of Src-family protein tyrosine kinases in NF-kappaB activation during B cell development, *Nat. Immunol.* **4**(3), 274–279.

Saklatvala, J. (2004) The p38 MAP kinase pathway as a therapeutic target in inflammatory disease, *Curr. Opin. Pharmacol.* **4**(4), 372–377.

Sakurai, H., Chiba, H., Miyoshi, H., Sugita, T., Toriumi, W. (1999) IkappaB kinases phosphorylate NF-kappaB p65 subunit on serine 536 in the transactivation domain, *J. Biol. Chem.* **274**(43), 30353–30356.

Sato, S., Sugiyama, M., Yamamoto, M., Watanabe, Y., Kawai, T., Takeda, K., et al. (2003) Toll/IL-1 receptor domain-containing adaptor inducing IFN-beta (TRIF) associates with TNF receptor-associated factor 6 and TANK-binding kinase 1, and activates two distinct transcription factors, NF-kappa B and IFN-regulatory factor-3, in the Toll-like receptor signaling, *J. Immunol.* **171**(8), 4304–4310.

Schmitz, M.L., Mattioli, I., Buss, H., Kracht, M. (2004) NF-kappaB: a multifaceted transcription factor regulated at several levels, *Chembiochem* **5**(10), 1348–1358.

Scholzke, M.N., Potrovita, I., Subramaniam, S., Prinz, S., Schwaninger, M. (2003) Glutamate activates NF-kappaB through calpain in neurons, *Eur. J. Neurosci.* **18**(12), 3305–3310.

Schreck, R., Baeuerle, P.A. (1991) A role for oxygen radicals as second messengers, *Trends Cell Biol.* **1**(2–3), 39–42.

Sen, R., Baltimore, D. (1986) Multiple nuclear factors interact with the immunoglobulin enhancer sequences, *Cell* **46**(5), 705–716.

Senftleben, U., Cao, Y., Xiao, G., Greten, F.R., Krahn, G., Bonizzi, G., et al. (2001) Activation

by IKKalpha of a second, evolutionary conserved, NF-kappa B signaling pathway, *Science* **293**(5534), 1495–1499.

Sheppard, K.A., Phelps, K.M., Williams, A.J., Thanos, D., Glass, C.K., Rosenfeld, M.G., et al. (1998) Nuclear integration of glucocorticoid receptor and nuclear factor-kappaB signaling by CREB-binding protein and steroid receptor coactivator-1, *J. Biol. Chem.* **273**(45), 29291–29294.

Simeonidis, S., Liang, S., Chen, G., Thanos, D. (1997) Cloning and functional characterization of mouse IkappaBepsilon, *Proc. Natl. Acad. Sci. U.S.A.* **94**(26), 14372–14377.

Simon, M.M., Aragane, Y., Schwarz, A., Luger, T.A., Schwarz, T. (1994) UVB light induces nuclear factor kappa B (NF kappa B) activity independently from chromosomal DNA damage in cell-free cytosolic extracts, *J. Invest. Dermatol.* **102**(4), 422–427.

Singh, S., Darnay, B.G., Aggarwal, B.B. (1996) Site-specific tyrosine phosphorylation of IkappaBalpha negatively regulates its inducible phosphorylation and degradation, *J. Biol. Chem.* **271**(49), 31049–31054.

Slifman, N.R., Gershon, S.K., Lee, J.H., Edwards, E.T., Braun, M.M. (2003) Listeria monocytogenes infection as a complication of treatment with tumor necrosis factor alpha-neutralizing agents, *Arthritis Rheum.* **48**(2), 319–324.

Soutschek, J., Akinc, A., Bramlage, B., Charisse, K., Constien, R., Donoghue, M., et al. (2004) Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs, *Nature* **432**(7014), 173–178.

Steed, P.M., Tansey, M.G., Zalevsky, J., Zhukovsky, E.A., Desjarlais, J.R., Szymkowski, D.E., et al. (2003) Inactivation of TNF signaling by rationally designed dominant-negative TNF variants, *Science* **301**(5641), 1895–1898.

Stein, B., Kramer, M., Rahmsdorf, H.J., Ponta, H., Herrlich, P. (1989) UV-induced transcription from the human immunodeficiency virus type 1 (HIV-1) long terminal repeat and UV-induced secretion of an extracellular factor that induces HIV-1 transcription in nonirradiated cells, *J. Virol.* **63**(11), 4540–4544.

Stone, J.D., Conroy, L.A., Byth, K.F., Hederer, R.A., Howlett, S., Takemoto, Y., et al. (1997) Aberrant TCR-mediated signaling in CD45-null thymocytes involves dysfunctional regulation of Lck, Fyn, TCR-zeta, and ZAP-70, *J. Immunol.* **158**(12), 5773–5782.

Sun, S.C., Ganchi, P.A., Ballard, D.W., Greene, W.C. (1993) NF-kappa B controls expression of inhibitor I kappa B alpha: evidence for an inducible autoregulatory pathway, *Science* **259**(5103), 1912–1915.

Suzuki, N., Suzuki, S., Duncan, G.S., Millar, D.G., Wada, T., Mirtsos, C., et al. (2002) Severe impairment of interleukin-1 and Toll-like receptor signalling in mice lacking IRAK-4, *Nature* **416**(6882), 750–756.

Takaesu, G., Surabhi, R.M., Park, K.J., Ninomiya-Tsuji, J., Matsumoto, K., Gaynor, R.B. (2003) TAK1 is critical for IkappaB kinase-mediated activation of the NF-kappaB pathway, *J. Mol. Biol.* **326**(1), 105–115.

Takaesu, G., Kishida, S., Hiyama, A., Yamaguchi, K., Shibuya, H., Irie, K., et al. (2000) TAB2, a novel adaptor protein, mediates activation of TAK1 MAPKKK by linking TAK1 to TRAF6 in the IL-1 signal transduction pathway, *Mol. Cell* **5**(4), 649–658.

Tam, W.F., Sen, R. (2001) IkappaB family members function by different mechanisms, *J. Biol. Chem.* **276**(11), 7701–7704.

Tergaonkar, V., Pando, M., Vafa, O., Wahl, G., Verma, I. (2002) p53 stabilization is decreased upon NFkappaB activation: a role for NFkappaB in acquisition of resistance to chemotherapy, *Cancer Cell* **1**(5), 493–503.

Tergaonkar, V., Bottero, V., Ikawa, M., Li, Q., Verma, I.M. (2003) IkappaB kinase-independent IkappaBalpha degradation pathway: functional NF-kappaB activity and implications for cancer therapy, *Mol. Cell. Biol.* **23**(22), 8070–8083.

Thanos, D., Maniatis, T. (1995) NF-kappa B: a lesson in family values, *Cell* **80**(4), 529–532.

Thompson, J.E., Phillips, R.J., Erdjument-Bromage, H., Tempst, P., Ghosh, S. (1995) I kappa B-beta regulates the persistent response in a biphasic activation of NF-kappa B, *Cell* **80**(4), 573–582.

Triantafilou, M., Triantafilou, K. (2004) Heat-shock protein 70 and heat-shock protein 90 associate with Toll-like receptor 4 in response to bacterial lipopolysaccharide, *Biochem. Soc. Trans.* **32**(Pt 4), 636–639.

Triantafilou, K., Triantafilou, M., Dedrick, R.L. (2001) A CD14-independent LPS receptor cluster, *Nat. Immunol.* **2**(4), 338–345.

Turpin, P., Hay, R.T., Dargemont, C. (1999) Characterization of IkappaBalpha nuclear import pathway, *J. Biol. Chem.* **274**(10), 6804–6812.

van de Stolpe, A., Caldenhoven, E., Stade, B.G., Koenderman, L., Raaijmakers, J.A., Johnson, J.P., et al. (1994) 12-O-tetradecanoylphorbol-13-acetate-and tumor necrosis factor alpha-mediated induction of intercellular adhesion molecule-1 is inhibited by dexamethasone. Functional analysis of the human intercellular adhesion molecular-1 promoter, *J. Biol. Chem.* **269**(8), 6185–6192.

Varfolomeev, E.E., Ashkenazi, A. (2004) Tumor necrosis factor: an apoptosis JuNKie? *Cell* **116**(4), 491–497.

Verma, I.M., Stevenson, J.K., Schwarz, E.M., Van Antwerp, D., Miyamoto, S. (1995) Rel/NF-kappa B/I kappa B family: intimate tales of association and dissociation, *Genes Dev.* **9**(22), 2723–2735.

Vermeulen, L., De Wilde, G., Van Damme, P., Vanden Berghe, W., Haegeman, G. (2003) Transcriptional activation of the NF-kappaB p65 subunit by mitogen- and stress-activated protein kinase-1 (MSK1), *EMBO J.* **22**(6), 1313–1324.

Vile, G.F., Tanew-Ilitschew, A., Tyrrell, R.M. (1995) Activation of NF-kappa B in human skin fibroblasts by the oxidative stress generated by UVA radiation, *Photochem. Photobiol.* **62**(3), 463–468.

Wajant, H., Pfizenmaier, K., Scheurich, P. (2003) Tumor necrosis factor signaling, *Cell Death Differ.* **10**(1), 45–65.

Wallis, R.S., Broder, M.S., Wong, J.Y., Hanson, M.E., Beenhouwer, D.O. (2004) Granulomatous infectious diseases associated with tumor necrosis factor antagonists, *Clin. Infect. Dis.* **38**(9), 1261–1265.

Wang, T., Zhang, X., Li, J.J. (2002) The role of NF-kappaB in the regulation of cell stress responses, *Int. Immunopharmacol.* **2**(11), 1509–1520.

Wang, D., Westerheide, S.D., Hanson, J.L., Baldwin, A.S. Jr. (2000) Tumor necrosis factor alpha-induced phosphorylation of RelA/p65 on Ser529 is controlled by casein kinase II, *J. Biol. Chem.* **275**(42), 32592–32597.

Wang, C., Deng, L., Hong, M., Akkaraju, G.R., Inoue, J., Chen, Z.J. (2001) TAK1 is a ubiquitin-dependent kinase of MKK and IKK, *Nature* **412**(6844), 346–351.

Weber, C.K., Liptay, S., Wirth, T., Adler, G., Schmid, R.M. (2000) Suppression of NF-kappaB activity by sulfasalazine is mediated by direct inhibition of IkappaB kinases alpha and beta, *Gastroenterology* **119**(5), 1209–1218.

Wesche, H., Henzel, W.J., Shillinglaw, W., Li, S., Cao, Z. (1997) MyD88: an adapter that recruits IRAK to the IL-1 receptor complex, *Immunity* **7**(6), 837–847.

Wietek, C., Miggin, S.M., Jefferies, C.A., O'Neill, L.A. (2003) Interferon regulatory factor-3-mediated activation of the interferon-sensitive response element by Toll-like receptor (TLR) 4 but not TLR3 requires the p65 subunit of NF-kappa, *J. Biol. Chem.* **278**(51), 50923–50931.

Wilkinson, K.D. (2003) Signal transduction: aspirin, ubiquitin and cancer, *Nature* **424**(6950), 738–739.

Wright, S.D., Ramos, R.A., Tobias, P.S., Ulevitch, R.J., Mathison, J.C. (1990) CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein, *Science* **249**(4975), 1431–1433.

Xiao, G., Harhaj, E.W., Sun, S.C. (2001) NF-kappaB-inducing kinase regulates the processing of NF-kappaB2 p100, *Mol. Cell* **7**(2), 401–409.

Yamamoto, Y., Yin, M.J., Lin, K.M., Gaynor, R.B. (1999) Sulindac inhibits activation of the NF-kappaB pathway, *J. Biol. Chem.* **274**(38), 27307–27314.

Yamamoto, M., Sato, S., Hemmi, H., Hoshino, K., Kaisho, T., Sanjo, H., et al. (2003) Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway, *Science* **301**(5633), 640–643.

Yamazaki, S., Muta, T., Takeshige, K. (2001) A novel IkappaB protein, IkappaB-zeta, induced by proinflammatory stimuli, negatively regulates nuclear factor-kappaB in the nuclei, *J. Biol. Chem.* **276**(29), 27657–27662.

Yamazaki, S., Muta, T., Matsuo, S., Takeshige, K. (2005) Stimulus-specific induction of a novel NF-kappa B regulator, Ikappa B-zeta, via toll/interleukin-1 receptor is mediated by mRNA stabilization, *J. Biol. Chem.* **280**(2), 1678–1687.

Yang, J., Xu, Z.P., Huang, Y., Hamrick, H.E., Duerksen-Hughes, P.J., Yu, Y.N. (2004) ATM and ATR: sensing DNA damage, *World J. Gastroenterol.* **10**(2), 155–160.

Yang, J., Lin, Y., Guo, Z., Cheng, J., Huang, J., Deng, L., et al. (2001) The essential role of MEKK3 in TNF-induced NF-kappaB activation, *Nat. Immunol.* **2**(7), 620–624.

Yeh, S.H., Lin, C.H., Gean, P.W. (2004) Acetylation of nuclear factor-kappaB in rat amygdala improves long-term but not

short-term retention of fear memory, *Mol. Pharmacol.* **65**(5), 1286–1292.

Yeh, S.H., Lin, C.H., Lee, C.F., Gean, P.W. (2002) A requirement of nuclear factor-kappaB activation in fear-potentiated startle, *J. Biol. Chem.* **277**(48), 46720–46729.

Yin, M.J., Yamamoto, Y., Gaynor, R.B. (1998) The anti-inflammatory agents aspirin and salicylate inhibit the activity of I(kappa)B kinase-beta, *Nature* **396**(6706), 77–80.

Yujiri, T., Ware, M., Widmann, C., Oyer, R., Russell, D., Chan, E., et al. (2000) MEK kinase 1 gene disruption alters cell migration and c-Jun NH2-terminal kinase regulation but does not cause a measurable defect in NF-kappa B activation, *Proc. Natl. Acad. Sci. U.S.A.* **97**(13), 7272–7277.

Zabel, U., Baeuerle, P.A. (1990) Purified human I kappa B can rapidly dissociate the complex of the NF-kappa B transcription factor with its cognate DNA, *Cell* **61**(2), 255–265.

Zhong, H., Voll, R.E., Ghosh, S. (1998) Phosphorylation of NF-kappa B p65 by PKA stimulates transcriptional activity by promoting a novel bivalent interaction with the coactivator CBP/p300, *Mol. Cell* **1**(5), 661–671.

Zhong, H., May, M.J., Jimi, E., Ghosh, S. (2002) The phosphorylation status of nuclear NF-kappa B determines its association with CBP/p300 or HDAC-1, *Mol. Cell* **9**(3), 625–636.

Zhong, H., SuYang, H., Erdjument-Bromage, H., Tempst, P., Ghosh, S. (1997) The transcriptional activity of NF-kappaB is regulated by the IkappaB-associated PKAc subunit through a cyclic AMP-independent mechanism, *Cell* **89**(3), 413–424.

Transcriptome and Interactome Correlation: *see* Mass Spectrometry of Proteins (Proteomics)

Transcriptome Characterization: *see* Serial Analysis of Gene Expression

Transgenic Animal Patents: *see* Living Organism (Animal) Patents

Transgenic Animal Technology: *see* Animal Biotechnology and Modeling

Transgenic Fish

Pinwen Peter Chiou, Jenny Khoo, Chung Zoon Chun, and Thomas T. Chen
University of Connecticut, Storrs, CT, USA

1 Introduction 475

2 Selection of Model Fish Species 475
2.1 Small-size Fish Model 476
2.2 Medium-size Fish Model 476
2.3 Large-size Fish Model 476

3 Types of Transgenes 477
3.1 Gain-of-function Transgenes 477
3.2 Reporter-function Transgenes 479
3.3 Loss-of-function Transgenes 480

4 Methods of Gene Transfer 480
4.1 Microinjection of DNA into Embryos or Unfertilized Eggs 480
4.2 Electroporation of DNA into Embryos or Sperm 481
4.3 Infection with Replication-defective Pantropic Retroviral Vectors 482
4.4 Emerging New Methodology 483

5 Characterization of Transgenic Fish 484
5.1 Identification of Transgenic Fish 484
5.2 Pattern of Transgene Integration 484
5.3 Inheritance of Transgenes 485
5.4 Expression of Transgenes 485

6 Application of Transgenic Fish 485
6.1 Transgenic Fish in Basic Research 486
6.1.1 Vertebrate Developmental Biology 487
6.1.2 Functional Analysis of Promoter/Enhancer Elements 488

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

6.1.3 Signaling Pathways 489
6.1.4 Human Disease Modeling 490
6.2 Growth Hormone Transgenic Fish 492
6.3 Transgenic Fish with Enhanced Resistance to Pathogen Infection 492
6.4 Transgenic Fish with Different Body Color 493
6.5 Transgenic Fish as Environmental Biomonitors 494
6.6 Other Examples of Transgenic Fish 496

7 Concerns and Future Perspectives 497

Bibliography 498
Books and Reviews 498
Primary Literature 499

Keywords

Electroporation
A process utilizing a series of short electrical pulses to permeate cell membranes, thereby permitting the entry of DNA molecules into animal or plant cells.

Polymerase Chain Reaction (PCR)
An *in vitro* method for the enzymatic synthesis of a specific DNA sequence, using two oligonucleotide primers hybridizing to opposite strands and flanking the region of interest in the target DNA.

Transgene
A piece of nonself origin gene or noncoding DNA fragment that is introduced into plants or animals for the production of transgenic species.

Transgenic Fish
Fish into which a foreign gene or noncoding DNA fragment has been introduced and stably integrated into its genome.

F_1 Transgenic Fish
Transgenic progeny derived from a cross between a P_1 transgenic individual and a nontransgenic individual or between two P_1 transgenic individuals.

P_1 Transgenic Fish
Transgenic fish developed from embryos that received foreign gene transfer.

Fish into which foreign DNA is artificially introduced and integrated in their genomes are called *transgenic fish*. Since the development of the first transgenic fish in 1985, techniques to produce transgenic fish have improved tremendously, resulting in the production of genetically modified (GM) fish of many species. In the past few years, the fast booming application of transgenic fish technology has led to many valuable discoveries in different disciplines of the biological sciences, many of which could potentially lead to the next breakthrough in new treatments for human diseases. Transgenic fish technology has also been applied to produce fish with beneficial traits, such as fast growth rate, enhanced disease resistance, and environmental biomonitors, in the hope of improving the quality of both human life and the earth environments.

1 Introduction

Fish into which a foreign gene or noncoding DNA fragment is artificially introduced and integrated in their genomes are called *transgenic fish*. Since 1985, a wide range of transgenic fish species have been produced mainly by microinjecting or electroporating homologous or heterologous transgenes into newly fertilized or unfertilized eggs and, sometimes, sperm. To produce a desired transgenic fish, several factors should be considered. First, an appropriate fish species must be chosen. This decision would depend on the nature of each study and the availability of the fish holding facility. Second, a specific gene construct must be designed on the basis of the special requirement of each study. For example, the gene construct may contain an open reading frame encoding a gene product of interest and regulatory elements that regulate the expression of the gene in a temporal, spatial, and/or developmental manner. Third, the gene construct has to be introduced into sperm or embryos in order for the transgene to be integrated stably into the genome of the embryonic cells. Fourth, since not all instances of gene transfer are efficient, a screening method must be adopted for identifying transgenic individuals.

Since the development of the first transgenic fish, techniques to produce transgenic fish have improved tremendously, resulting in the production of genetically modified (GM) fish of many species. In recent years, transgenic fish have been established as valuable models for different disciplines of biological research as well as for human disease modeling. The application of transgenic fish technology to produce fish with beneficial traits, such as environmental biomonitors and enhanced disease resistance, is also rising. In this chapter, we will review the progress of the transgenic fish technology and its application in basic research and biotechnological application.

2 Selection of Model Fish Species

Gene transfer studies have been conducted in several different fish species such as salmon, rainbow trout, common carp, goldfish, loach, channel catfish, tilapia, northern pike, walleye, zebrafish, and Japanese medaka. Depending on the

purpose of the transgenic fish studies, the embryos of some fish species prove to be more suitable than others. In selecting a fish species for gene transfer studies, a series of parameters are considered. These parameters are (1) length of the life cycle; (2) year round supply of eggs and sperm; (3) culture conditions; (4) size of the adult at maturity; and (5) availability of background information on genetics, physiology, and endocrinology of the fish species.

2.1 Small-size Fish Model

Japanese medaka (*Oryzias latipes*) and zebrafish (*Danio rerio*) are regarded as ideal fish species for conducting gene transfer studies. Both the fish species have short life cycles (3 months from eggs to mature adults), produce hundreds of eggs on a regular basis without exhibiting a seasonal breeding cycle, and can be maintained easily in the laboratory for two to three years. Their eggs are relatively large, 1.0 to 1.5 mm diameter, and possess very thin and semitransparent chorions. These features allow easy microinjection of DNA into the eggs if appropriate glass needles are used. Furthermore, inbred lines and various morphological mutants of both the fish species are available. These fish species are suitable candidates for producing transgenic fish for (1) studying developmental regulation of gene expression and gene action; (2) identifying regulatory elements that regulate the expression of a gene; (3) measuring the activities of promoters; and (4) producing transgenic models for human diseases and environmental toxicology. However, a major drawback of these two fish species is that their small body size makes them unsuitable for endocrinological or biochemical analysis.

2.2 Medium-size Fish Model

The handicap of using small body–size fish species such as medaka or zebrafish in gene transfer studies can be easily overcome by the use of fish species such as loach, killifish (*Fundulus*), goldfish, and tilapia. The body sizes of these fish species are big enough so that sufficient amounts of tissues can be isolated from individual fish for biochemical and endocrinological studies. Another important attribute of these fish species is their shorter maturation time compared to rainbow trout or salmon. In particular, tilapia is a very appropriate medium-size fish species for gene transfer studies since it requires only about three months to complete its entire life cycle. It is possible to produce three generations of transgenic tilapia in one year. Unfortunately, the lack of a well-defined genetic background and asynchronous reproductive behavior of these fish species present some problems for conducting gene transfer studies.

2.3 Large-size Fish Model

Rainbow trout, salmon, channel catfish, and common carp are commonly used large body–size model fish species for transgenic fish studies. Since the knowledge of endocrinology, reproductive biology, and physiology of these fish species have been well worked out, they are well suited for studies on comparative endocrinology as well as on aquaculture applications. However, the maturation time for each of these fish species is relatively long. For example, rainbow trout and salmon require two to three years to reach reproductive maturity, and common carp or channel catfish about one to two years.

Thus, it will prolong inheritance studies. In addition, since these fish species have a single spawning cycle per year, it restricts the number of gene transfer attempts that can be conducted per year.

3 Types of Transgenes

A transgene is a piece of nonself origin gene or DNA fragment that is introduced into fish for the production of transgenic fish. Transgenes are usually constructed in plasmids that contain appropriate promoter/enhancer elements for proper expression of the transgenes in fish. A variety of promoters from nonfish species have been employed to control the expression of transgenes in fish (Table 1). Additionally, increasing number of fish promoters (Table 1) are available for the purpose of generating "all-fish" expression cassettes, among which the carp β-actin promoter has been shown to drive strong expression of transgenes in various fish cell types. Tissue-specific promoters and inducible promoters such as zebrafish *heat shock protein 70* (*hsp70*) have also been used successfully to control the expression of transgene in the desired tissue(s) at the desired time(s). For example, activation of *hsp70*-controlled transgene can be achieved by heat shock and, intriguingly, by focusing a sublethal laser microbeam onto individual cells that carry the transgene. The targeted cells appear normal after the treatment.

Depending on the purpose of the gene transfer studies, transgenes can be grouped into three major types: (1) *gain-of-function*; (2) *reporter-function*; and (3) *loss-of-function*. The basic organization and the function of these transgenes will be discussed below.

3.1 Gain-of-function Transgenes

Gain-of-function transgenes are designed to add new functions to the transgenic

Tab. 1 Examples of promoters used in production of transgenic fish.

	Origin Species	***Promoter***
Nonfish origin:		
	Rat	GAP43[a]
	Chicken	β-actin
	Xenopus	elongation factor 1α
	Cytomegalovirus	CMV intermediate–early
	Moloney murine leukemia virus	MoMLV LTR
Fish origin:		
	Carp	β-actin
	Goldfish	α1tubulin[a]
	Medaka	elongation factor 1α, β-actin
	Zebrafish	H2A.F/Z, α-actin[a], heat shock protein 70[b]

[a]tissue-specific promoters- GAP43: nervous system; α1 tubulin: nervous system; α-actin: muscle (see able 3 for more examples).
[b]Inducible promoters; LTR: long terminal repeat; H2A.F/Z: histone H2A.F/Z.

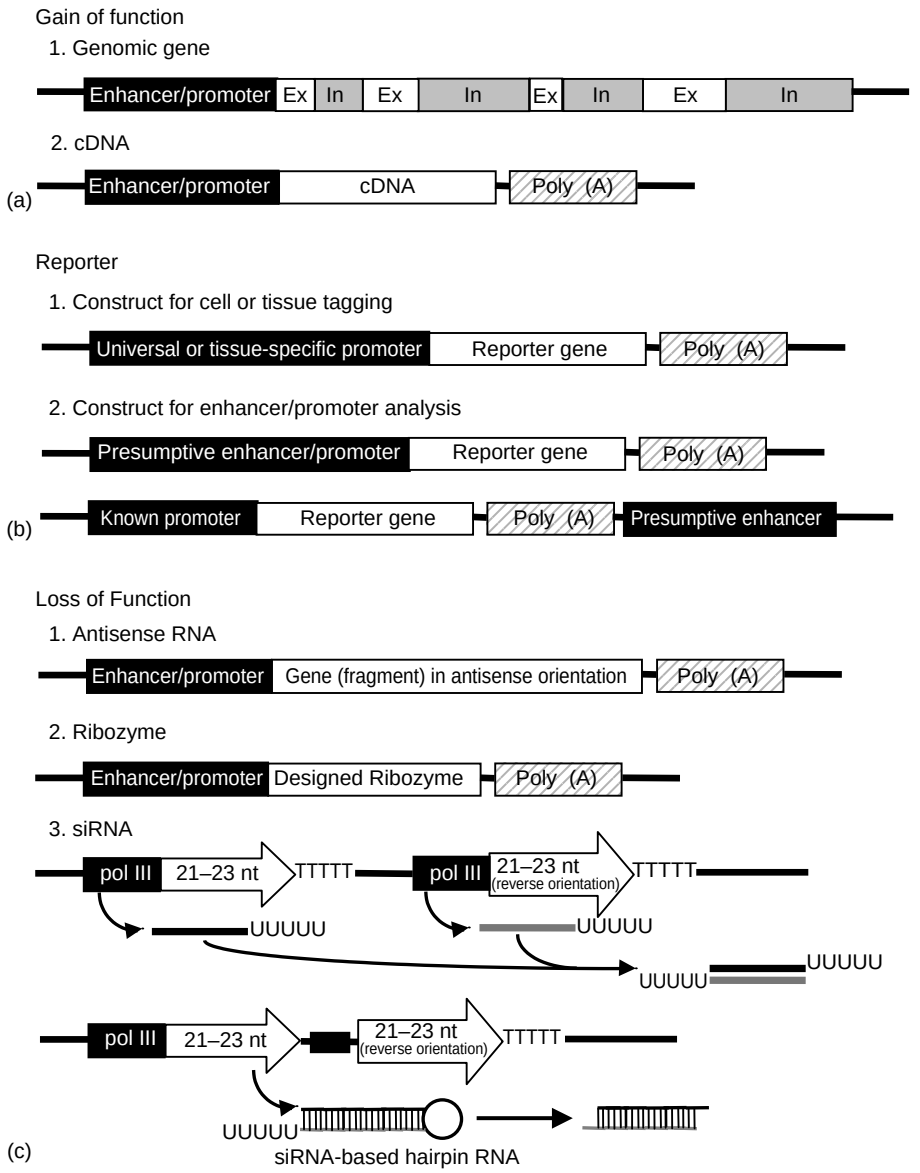

Fig. 1 Prototypes of transgenes for fish transgenesis. Depending on the experimental purpose, the promoter used in the transgene construct can be either constitutive or an inducible one. Ex: exon; in: intron; pol III: RNA polymerase III promoters.

individuals or to facilitate the identification of the transgenic individuals if the genes are expressed properly in the transgenic individuals. The coding regions of the transgenes are usually homologous or heterologous genomic or cDNA sequences, which encode polypeptide products (Fig. 1a). The expression

of the transgenes is usually driven by promoter/enhancer sequences of homologous or heterologous sources. If the transgenes contain functional signal peptide sequences, the gene products will be secreted out of cells once they are synthesized. Transgenes containing the structural gene of mammalian or fish growth hormones (GH) or the respective cDNA fused to a functional promoter/enhancer of chicken or fish β-actin gene are examples of the gain-of-function transgene constructs. Expression of such transgenes in transgenic individuals has been shown to result in growth enhancement. Winter flounder antifreeze protein gene and chicken δ-crystalline gene with their own promoter/enhancer sequences are other commonly used *gain-of-function* transgenes used in various gene transfer studies.

Another type of commonly used *gain-of-function* transgenes are the "reporter-function transgenes," which are discussed in the next section.

3.2 Reporter-function Transgenes

Genes such as *bacterial chloramphenicol acetyltransferase* (*CAT*), *β-galactosidase* (*β-gal*), *neomycin phosphotransferase* II (*neo*^{R}), *luciferase*, and *green fluorescent protein* (*GFP*), are frequently used as reporters. These reporter genes serve as a convenient marker for indicating the success of gene transfer into the target cells since the translational products of these genes can be easily monitored. Furthermore, because these gene products can be measured quantitatively, they are also routinely adopted for identifying the sequence of an undefined promoter/enhancer or for measuring the relative activity of a promoter/enhancer element in question. A less frequently used reporter type is the one with target-specificity. One such example is cameleon, a fluorescent calcium sensor, which was originally used to detect calcium transients in culture cells and pharyngeal muscle of *Caenorhabditis elegans*. Recently, *cameleon* and several other indicator genes have been used to establish transgenic zebrafish for monitoring neuronal activity. The prototypes of a *reporter-function* transgene are shown in Fig. 1b.

Among the commonly used reporters, the jellyfish (*Aequorea victoria*) GFP and its variants have become the choice markers for transgenic fish because the detection of GFP signal is direct and easy, requiring no exogenous substrate. There are numerous GFP variants available, including those producing green, blue, and yellow fluorescence light. Together with the red fluorescence protein (RFP), these reporters can be easily detected in live transgenic fish such as zebrafish and medaka whose embryos and body are transparent, without the need to sacrifice the fish. Additionally, these reporters can be used to study multiple transgenes simultaneously in the same fish. Since the first GFP transgenic fish in 1995, GFP-tagged transgenic embryos and fish have become the most convenient and widely used tools in basic biological research, particularly in the regulation of gene expression and morphogenetic movements during embryonic development (see Sect. 6.1). In many cases, similar applications can be achieved with the other reporter genes mentioned above. However, the detection of these reporters is indirect and involves additional steps such as enzymatic reaction with substrate and immunohistochemical staining with specific antibody to the reporter.

3.3 Loss-of-function Transgenes

The *loss-of-function* transgenes are constructed for interfering with the expression of host genes at the transcriptional or translational level. Mechanistically, there are at least three groups of "loss-of-function" transgenes. The transgenes might encode antisense RNAs or oligonucleotides that selectively hybridize to a target endogenous mRNA, resulting in degradation of the mRNA with RNase H or interference with translation of the mRNAs via preventing the binding of ribosomes. A variation of the antisense RNA approach is to introduce a group of chemically modified oligonucleotides known as morpholino phosphorodiamidate oligonucleotides (MOs) into fish embryos. In contrast to ordinary antisense molecules, MOs exhibit long-term stability with less toxicity to living cells and appear to function through translation blockage only. The second group of "loss-of-function" transgenes is catalytic RNA (ribozymes) that can cleave specific target mRNAs at a specific site and thereby "knock down" the normal gene expression. The most recent development of the *loss-of-function* transgenes is to introduce a group of small RNAs known as small (short) interfering RNAs (siRNAs) into cells. siRNA will guide a ribonuclease known as RISC (the RNA-induced silencing complex) to recognize mRNA containing a sequence homologous to the siRNA, resulting in degradation of the mRNA by the ribonuclease. The prototype of *loss-of-function* transgenes is shown in Fig. 1c. Although there is currently no report of stable lines of transgenic fish carrying this type of transgenes, the feasibility of this approach in silencing genes has been demonstrated in numerous fish cell lines *in vitro*. Moreover, many recent studies also prove that MOs and siRNA molecules can silence their targets specifically and effectively when introduced into fish embryos, such as zebrafish, medaka, and rainbow trout.

4 Methods of Gene Transfer

Techniques such as calcium phosphate precipitation, direct microinjection, retrovirus infection, electroporation, and particle gun bombardment have been widely used to introduce foreign DNA into animal cells, plant cells, and germ lines of mammals and other vertebrates. Among these methods, direct microinjection and electroporation of DNA into newly fertilized eggs have proven to be the most reliable methods of gene transfer in fish systems.

4.1 Microinjection of DNA into Embryos or Unfertilized Eggs

Microinjection of foreign DNA into newly fertilized eggs was first developed for the production of transgenic mice in the early 1980s. Since then, the technique of microinjection has been adopted for introducing transgenes into Atlantic salmon, common carp, catfish, goldfish, loach, medaka, rainbow trout, tilapia, and zebrafish. The gene constructs that were used in these studies include human or rat growth hormone gene, rainbow trout or salmon GH cDNA, chicken δ-crystalline protein gene, winter flounder antifreeze protein gene, *E. coli* β-galactosidase gene, and *E. coli* hygromycin resistance gene. In general, gene transfer in fish by direct microinjection is conducted as follows. The parameters for microinjection are summarized in Table 2. Eggs and sperm are collected in separate, dry containers.

Tab. 2 Parameters of fish transgenesis by microinjection and electroporation.

Parameters	*Gene Transfer Method*	
	Microinjection	*Electroporation*
Developmental stage	1 to 2 cells	1 to 2 cells
DNA size	<10 Kb	<10 Kb
DNA concentration	10^{6-7} molecules/embryo	100 μg ml^{-1}
DNA topology	Linear	Linear
Chorion barrier	Dechorionated/micropyle	Intact chorion
Electrical field strength	N/A[a]	500 to 3000 v
Pulse shape	N/A[a]	exponential/square
Pulse duration	N/A[a]	ms[c] to s
Temperature	RT[b]	RT[b]
Medium	PBS/saline[d]	PBS/saline

[a]N/A, not applicable.
[b]RT, room temperature (25 °C).
[c]ms, millisecond.
[d]PBS/saline: phosphate-buffered saline.

Fertilization is initiated by mixing sperm and eggs, then adding water, with gentle stirring to enhance fertilization. Eggs are microinjected within the first few hours after fertilization. The injection apparatus consists of a dissecting stereomicroscope and two micromanipulators, one with a micro-glass-needle for injection and the other with a micropipette for holding fish embryos in place. Routinely, about 10^6 to 10^8 molecules of a linearized transgene are injected into the egg cytoplasm. Following injection, the embryos are incubated in water until hatching. Since natural spawning in zebrafish or medaka can be induced by adjusting photoperiod and water temperature, precisely staged newly fertilized eggs can be collected from the aquaria for gene transfer. If the medaka eggs are maintained at 4 °C immediately after fertilization, the micropyle on the fertilized eggs will remain visible for two hours. The DNA solution can be delivered into the embryos by injecting through this opening during this time period.

Depending on the species, the survival rate of injected fish embryos ranges from 35 to 80% while the rate of DNA integration ranges from 10 to 70% in the survivors. The tough chorions of the fertilized eggs in some fish species, for example, rainbow trout and Atlantic salmon, can make insertion of glass needles difficult. This difficulty has been overcome by any one of the following methods: (1) inserting the injection needles through the micropyle; (2) making an opening on the egg chorions by microsurgery; (3) removing the chorion by mechanical or enzymatic means; (4) preventing chorion hardening by initiating fertilization in a solution containing 1 mM glutathione; or (5) injecting the unfertilized eggs directly.

4.2 Electroporation of DNA into Embryos or Sperm

Electroporation is a successful method for transferring foreign DNA into bacteria, yeast, and plant and animal cells in culture.

In the past years, this method has become popular for transferring foreign genes into fish embryos or sperm. Electroporation utilizes a series of short electrical pulses to permeate cell membranes, thereby permitting the entry of DNA molecules into embryos or sperm. The patterns of electrical pulses can be emitted in a single pulse of exponential decay form (i.e. exponential decay generator) or high-frequency multiple peaks of square waves (i.e. square-wave generator). The basic parameters are summarized in Table 2. Studies conducted in our laboratory and those of others have shown that the rate of DNA integration in electroporated embryos is on the order of 20% or higher in the survivors. Although the overall rate of DNA integration in transgenic fish produced by electroporation was equal to or slightly higher than that of microinjection, the actual amount of time required for handling a large number of embryos by electroporation is orders of magnitude less than that required for microinjection. Several reports have also appeared in the literature describing successful transfer of transgenes into fish by electroporating sperm instead of embryos. With the success seen in sperm electroporating, electroporation can therefore be considered as an efficient and versatile mass gene transfer technology.

4.3 Infection with Replication-defective Pantropic Retroviral Vectors

By microinjection or electroporation, plasmid-encoded transgenes can be introduced at a satisfactory efficiency into many fish species to produce transgenic offspring. However, the resulting P_1 transgenic individuals are almost always mosaics as a result of delayed transgene integration. Furthermore, both methods are not efficient or applicable in transferring foreign DNA into embryos of life-bearing fish, marine fish, and invertebrates because of the restrictions imposed by the unique spawning behavior or anatomical structure of these species. Modification of transgenes has been adopted to increase the efficiency of gene transfer by microinjection or electroporation. Hsiao et al. and Thermes et al. have taken the following approaches: (1) using the inverted terminal repeat DNA from adeno-associated virus to increase the stability of transgene integration; and (2) employing *I-SceI* meganuclease to mediate high transportation efficiency of the transgene into the nucleus. Despite the improvement, these modified approaches cannot be applied to the mentioned species, whose transgenic offspring can not be produced by either microinjection or electroporation. Instead, transformation of mature or immature germ cells by infection with transgene-encoded retroviral vectors may be an effective approach for transgenesis in these fish.

A replication-defective pantropic retroviral vector containing the long terminal repeat (LTR) sequence of Moloney murine leukemia virus (MoMLV) and transgenes in a viral envelop with the G-protein of vesicular stomatitis virus (VSV), was developed by Burns et al. Since entry of VSV into host cells is mediated by interaction of the VSV-G protein with a phospholipid component of the host cell membrane, this pseudotyped retroviral vector has a very broad host range and is able to transfer transgene into many different cell types. Using the pantropic pseudotyped defective retrovirus as a gene transfer vector, transgenes containing neo^R or β-galactosidase were transferred into zebrafish, medaka, and dwarf surf clams.

More recently, Sarmasik et al. and Chen et al. also used the pantropic retroviral vector to transfer genes into the immature gonads of crayfish, desert guppy, and shrimp, respectively. They found that, by using the pantropic retroviral vector as a gene transfer vector, the problem of transgene mosaicism in P_1 transgenic fish is eliminated as the transgenes are introduced into immature male gonads by the pantropic retroviral vector.

4.4 Emerging New Methodology

Nuclear transplantation is a key technique in animal cloning. This method involves the transfer of nuclei from donor cells to enucleated eggs, stimulating eggs into cleavage and early phase of embryonic development, and in the case of mammals, the developing embryos are transferred into pseudopregnant females to complete the final stages of development. The first successful cloning of animals by nuclear transplantation was established for Northern Leopard Frog (*Rana pipiens*) by Briggs and King in 1952. However, Dolly, the lamb, is the first mammal successfully cloned by the nuclear transplantation technique, and since then, many laboratories throughout the world have adopted this technique to clone other mammals and farm animals. To date, production of diploid and fertile fish by the nuclear transplantation technique have been achieved in two small laboratory fish species, medaka and zebrafish. Three types of donor cells have been used for nuclear transfer: sperm, long-term cultured somatic cells, and embryonic cells.

In the study reported by Wakamatsu et al., embryonic cells from two transgenic medaka lines were used as donor cells to generate transgenic medaka. Blastoderm cells from the donor embryos in the mid-blastula stage were collected and dissociated into single cells. These were subsequently transplanted into the cytoplasm of the recipient enucleated eggs at the animal pole through the micropyle. In this study, 1 out of 588 eggs transplanted with the first donor and 5 out of 298 eggs transplanted with the second donor developed successfully to adulthood. The transgene of the donor nuclei was also transmitted to the F_1 and F_2 offspring in a Mendelian fashion.

Two versions of nuclear transfer technique have been established for zebrafish recently. Jesuthasan and Subburaju reported the use of sperm nuclei as donor for nuclear transfer. Sperm collected from mature males were demembranated by treatment with lysolecithin, digitonin, or by freeze-thawing. The demembranated sperm were then incubated with transgene DNA at room temperature prior to microinjecting into the animal pole region of the unfertilized egg. This method could produce nonmosaic expression of transgene in all cells of the embryos if incubation of the demembranated sperm with DNA was prolonged to 20 min. Few of the treated eggs however developed into fertile adults and the transgene was shown to be transmitted through the germ line. In the report by Lee et al., embryonic fibroblast cells from disaggregated embryos were used as donor cells. These cells were first modified by retroviral insertions expressing GFP as a reporter and were cultured for at least 12 weeks before their nuclei were transplanted into enucleated, unfertilized eggs. The method resulted in 2% of the transplants developing into fertile, diploid offspring. This study clearly demonstrated that nuclei from slowly dividing cultured fish somatic cells could be reprogrammed to support rapid embryonic development

and used as donor cells for nuclear transfer in fish. Moreover, the promising use of modified somatic cells provides an alternative choice of vehicle for gene targeting and fish cloning since fish ES cells are still unavailable. Overall, these studies have demonstrated the feasibility of using germinal and somatic cells from fish as donors for conducting nuclear transfer studies in different important fish species for both basic research and application purposes. In addition, the problem of transgene mosaicism in P_1 transgenic fish can be overcome by the nuclear transplantation method.

5 Characterization of Transgenic Fish

5.1 Identification of Transgenic Fish

Identification of transgenic individuals is the most time-consuming step in the production of transgenic fish. Traditionally, dot blot and Southern blot hybridization of genomic DNA were common methods used to determine the presence of transgenes in the presumptive transgenic individuals. These methods involve isolation of genomic DNA from tissues of presumptive transgenic individuals, digestion of DNA samples with restriction enzymes, and Southern blot hybridization of the digested DNA products. Although this method is expensive, laborious, and insensitive, it offers a definitive answer as to whether a transgene has been integrated into the host genome. Furthermore, it also reveals the pattern of transgene integration if appropriate restriction enzymes are employed in the Southern blot hybridization analysis. In order to handle a large number of samples efficiently and economically, a polymerase chain reaction (PCR) based assay has been adopted. The strategy of the assay involves the isolation of genomic DNA from a very small piece of fin tissue, PCR amplification of the transgene sequence, and Southern blot analysis of the amplified products. Although this method does not differentiate whether the transgene is integrated in the host genome or exists as an extrachromosomal unit, it serves as a rapid and sensitive screening method for identifying individuals that contain the transgene at the time of analysis. This method has been used in our laboratory as a routine preliminary screen for the presence of transgenes in thousands of presumptive transgenic fish.

5.2 Pattern of Transgene Integration

Studies conducted in many fish species have shown that following injection of linear or circular transgene constructs into fish embryos, the transgenes are maintained as an extrachromosomal unit through many rounds of DNA replication in the early phase of the embryonic development. At a later stage of embryonic development, some of the transgenes are randomly integrated into the host genome while others are degraded, resulting in the production of mosaic transgenic fish. To determine the pattern of transgene integration, genomic DNA from PCR positive fish is digested with a series of restriction endonucleases and the resulting products resolved in agarose gels for Southern blot analysis. In many fish species studied to date, it is found that multiple copies of transgenes are integrated in a head-to-head, head-to-tail or tail-to-tail form, except in transgenic common carp and channel catfish where a single copy of transgene

was integrated at multiple sites on the host chromosomes.

5.3 Inheritance of Transgenes

Stable integration of transgenes is an absolute requirement for continuous vertical transmission to subsequent generations and establishment of a transgenic fish line. To determine whether the transgene is transmitted to subsequent generations, P_1 transgenic individuals are mated to nontransgenic individuals and the progeny are assayed for the presence of transgene by the PCR assay method described earlier. Although it has been shown that the transgene may persist in the F_1 generation of transgenic zebrafish as extrachromosomal DNA, detailed analysis of the rate of transmission of transgenes to F_1 and F_2 generations in many transgenic fish species indicates true and stable incorporation of gene constructs into the host genome. If the entire germ line of the P_1 transgenic fish is transformed with at least one copy of the transgene per haploid genome, at least 50% of the F_1 transgenic progeny will be expected in a backcross involving a P_1 transgenic with a nontransgenic control. In many such crosses, only about 20% of the progeny are transgenic. When the F_1 transgenic is backcrossed with a nontransgenic control, however, at least 50% of the F_2 progeny are transgenics. These results clearly suggest that the germ lines of the P_1 transgenic fish are mosaic as a result of delayed transgene integration during embryonic development.

5.4 Expression of Transgenes

An important aspect of gene transfer studies is the detection of transgene expression. Depending on the levels of transgene products in the transgenic individuals, the following listed methods are commonly used to detect transgene expression: (1) RNA northern or dot blot hybridization; (2) RNase protection assay; (3) reverse transcription-polymerase chain reaction (RT-PCR); (4) immunoblotting assay; and (5) other biochemical assays for determining the presence of the transgene protein products. Among these assays, RT-PCR is the most sensitive method and only requires a small amount of sample. The strategy of this assay is summarized in Fig. 2. Briefly, it involves the isolation of total RNA from a small piece of tissue, synthesis of single-stranded cDNA by reverse transcription, and PCR amplification of the transgene cDNA by employing a pair of oligonucleotide specific to the transgene product as amplification primers. The resulting products are resolved on agarose gels and analyzed by Southern blot hybridization using a radiolabeled transgene as a hybridization probe. Transgene expression can also be quantified by a quantitative RT-PCR method or a quantitative real-time RT-PCR.

6 Application of Transgenic Fish

As techniques of producing various species of transgenic fish have become available in the last two decades, there has been a rapid boom in applying transgenic fish strategy to different disciplines of basic research and biotechnological applications. In this section, we will discuss recent advances in applying the transgenic fish technology in basic research as well as in producing transgenic fish with beneficial traits such as enhanced growth and disease resistance for aquaculture.

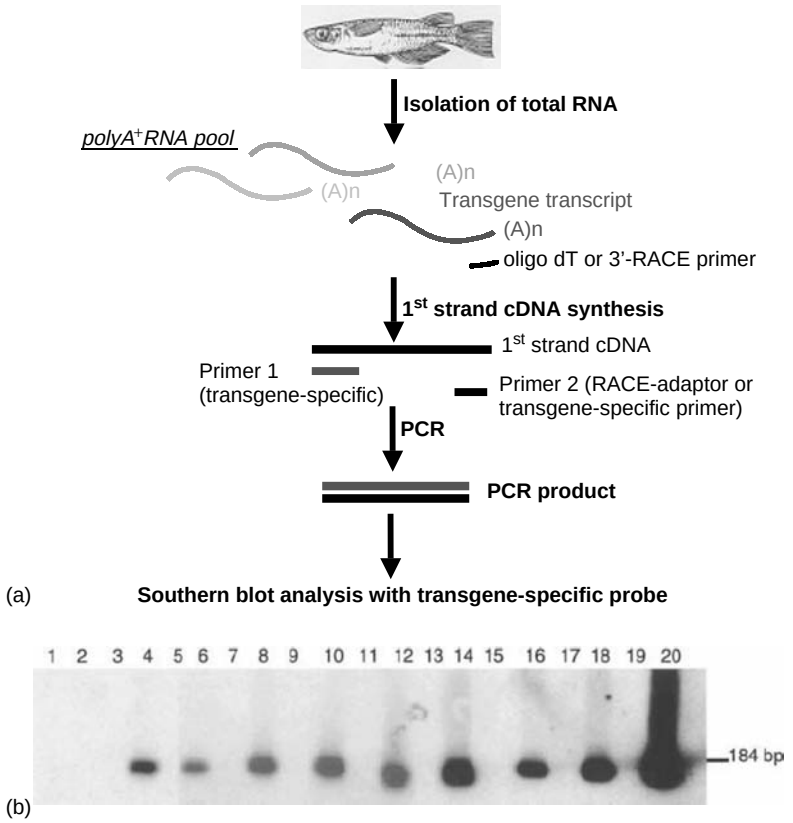

Fig. 2 Determination of transgene expression by RT-PCR or RACE-PCR. (a) Strategy of RT-PCR and RACE-PCR. (b) An example of detecting cecropin transgene in transgenic medaka. Total RNA was isolated from fin tissues of F_2 transgenic medaka. Expression of cecropin transgene was detected by reverse transcribing poly A^+ RNA into complementary cDNA and PCR amplification of the cecropin cDNA by using cecropin-specific primers. The amplified product was further confirmed by Southern hybridization using a cecropin-specific probe. RT-PCR was conducted with mRNA from transgenic fish without prior reverse transcription (lanes 5, 7, 9, 11, 13, 15, 17, 19) or with first cDNA transcribed from mRNA isolated from the transgenic fish (lanes 4, 6, 8, 10, 12, 14, 16, 18). Lane 1: RT-PCR with no mRNA input; lane 2: RT-PCR with mRNA from negative control fish (c-mRNA); lane 3: PCR with c-mRNA without prior reverse transcription; lane 20: PCR with control plasmid containing cecropin gene. From Sarmasik, A., Warr, G., Chen, T.T. (2002) Production of transgenic medaka with increased resistance to bacterial pathogens, *Mar. Biotechnol.* **4**, 310–322 with permission.

6.1 Transgenic Fish in Basic Research

Teleost fish serve as extremely valuable models for basic research in vertebrate biology as they are functionally similar to mammalian species, and because they are sufficiently evolutionarily distant to mammals, they are well suited for comparative genomics. Transgenic fish

can be easily and economically produced and reared in large quantities, granting them superiority over the transgenic mouse model, especially in studies requiring a large number of animals. In recent years, transgenic fish tagged with GFP have emerged as important tools in studies of vertebrate developmental biology, basic molecular and cell biology, and human disease modeling. Additionally, gene misexpression and knockdown are also widely used in these studies.

6.1.1 Vertebrate Developmental Biology

Transgenic zebrafish and medaka and their embryos are ideal models for studying the complex developmental mechanisms that transform a single-cell zygote into a multicellular organism with diverse functions. Both the fish species provide the following advantages: (1) short generation time (about 3 months); (2) constantly available and easily reared; and (3) most importantly, transparent embryos granting clear visualization of all stages of embryonic development. Additionally, there are mutant zebrafish and medaka whose transparent bodies consist of less interfering pigmentation. Although most of the developmental studies have so far been conducted in zebrafish, medaka embryos provide a larger window for observation at smaller developmental intervals because of their longer embryonic developmental time (2 days versus 10 days).

As shown in Table 3, researchers have established many lines of transgenic reporter zebrafish that express GFP in specific cells or tissues, including germ cells, blood, lymphoid cells, epithelia, muscle, heart (Fig. 4a), pancreas, liver, intestine (Fig. 4b), vasculature, and the nervous system. This is achieved by introducing a GFP, enhanced green fluorescence protein (EGFP), or RFP reporter that is placed under control of tissue- or cell-specific promoter/enhancer. Because the same fluorescence-tagged cells can be followed over time, these transgenic fish have been successfully used in the study of morphogenesis and gene regulation during embryonic development. For example, transgenic zebrafish with GFP-tagged blood vessels have been used to study the blood vessels assembly during development. Zebrafish *fli1* promoter is able to drive expression of GFP in all blood vessels throughout embryogenesis. By coupling this feature with high-resolution dynamic imaging techniques, confocal microangiography has clearly illustrated the anatomical architecture of the vasculature of zebrafish and show that blood vessels undergoing angiogenesis display pathfinding

Tab. 3 Examples of cell-type and tissue-type specific transgenic zebrafish lines with GFP or EGFP reporter.

Specificity	*Promoter*
Cell-specific	
Germ cells	*Xenopus efl* α (*-vasa* 3'UTR)
	Zebrafish *versa*
Blood	Zebrafish *gata 1*
Lymphoid cells	Zebrafish *rag1*
Epithelia	Zebrafish *krt8*
Tissue-specific	
Muscle	Zebrafish α-*actin*
Pancreas	Zebrafish *pdx-1*
	Zebrafish *insulin*
Liver	Zebrafish *l-fabp*
Intestine	Zebrafish *i-fabp*
Vasculature	Zebrafish *fli 1*
Heart	Zebrafish *cmlc2*
Nervous system	Zebrafish *islet1*
	Zebrafish *huC*
	Goldfish α1 *tubulin*

behavior similar to that of neuronal growth cones.

Application of fish transgenesis on developmental biology can also be achieved by gene misexpression or knockdown. So far, such an approach has been carried out exclusively in fish embryos. For example, zebrafish embryos have been used to study the role of Sema3D protein in guiding retinal ganglion cell (RGC) axons during embryonic development. Sema3D belongs to the semaphorin family, many members of which are known to repel or attract specific growth cones. Sema3D may participate in guiding RGC axons along the dorsoventral axis of the tectum, as indicated by the differentially higher expression of the protein in the ventral versus dorsal tectum. To understand the function of Sema3D, homozygous zebrafish embryos have been established from stable transgenic zebrafish fish carrying the *sema3D* gene under control of the zebrafish *heat shock protein 70* (*hsp70*) promoter. Ubiquitous misexpression (overexpression) of Sema3D in the embryos can be achieved by heat shock and has been found to inhibit ventral but not dorsal RGC axon growth. Additionally, mosaic Sema3D misexpression can be induced by heat shock in the embryos microinjected with the *hsp70*-controlled *sema3D* transgene. In these embryos, ventral RGC axons have been found to avoid or stop at individual cells misexpressing Sema3D along their pathway. Furthermore, knockdown of Sema3D with morpholino antisense causes errors in retinotectal mapping along the dorsoventral axis. These analyses suggest that Sema3D in the ventral tectum directs the correct innervation of ventral RGCs to dorsal tectum by inhibiting them from extending into ventral tectum. A similar approach has been employed to study the function of Sema4E protein in embryonic development.

6.1.2 Functional Analysis of Promoter/Enhancer Elements

Transgenic zebrafish and medaka have become popular tools in the identification and functional analysis of promoter/enhancer elements of vertebrates. The fast-growing genomic information of zebrafish and medaka facilitates the identification of evolutionarily conserved promoter/enhancer elements via cross-species comparison of genome sequences among fish, mouse, and human. Transgenic GFP-tagged zebrafish and medaka have been applied to define the essential regions of promoter/enhancer elements as well as to unmask the regulatory functions of these elements. For example, transgenic medaka and zebrafish have been used to characterize promoters of carp rhodopsin gene (cRh) and zebrafish *myf-5* gene (myf-5), respectively. In these studies, transgenic fish were designed to carry EGFP gene fused with different lengths of the upstream fragments of the presumed promoter/enhancer regulatory sequences. The role of motifs, enhancers, and minimal cis-regulatory elements involved in tissue-specific expression or the translocation of progenitor muscle cells have been identified by monitoring the expression of the EGFP reporter.

Another example of this approach is the investigation of the regulatory function of the LTRs of human ERV-9 endogenous retrovirus. The LTRs of endogenous ERV-9 retrovirus are conserved during primate evolution, but their function in the primate genomes is unknown. In human, the solitary LTRs of ERV-9 are identified as middle repetitive DNAs associated with up to 4000 human gene

loci including the β-globin gene locus where the ERV-9 LTR is juxtaposed to the locus control region (β-LCR) far upstream of the globin genes. The regulatory function of ERV-9 LTRs in the human genome has recently been demonstrated via using transgenic zebrafish carrying the β-globin ERV-9 LTR coupled to the GFP gene. The analysis suggests a regulatory role of ERV-9 LTR enhancers, during oogenesis, in the synthesis of maternal mRNAs that are required for early embryogenesis. Alternatively, the ERV-9 LTR enhancers could transcriptionally potentiate and preset chromatin structure of the cis-linked gene loci in oocytes and adult stem/progenitor cells.

6.1.3 Signaling Pathways

In addition to traditional cell-culture system, transgenic fish and their embryos have become an important tool for studying different signaling pathways in recent years. So far, most of the transgenic fish studies are conducted in zebrafish, thanks to the abundant genetic information of the species. Several achievements by the new approach are described here. Perz-Edwards et al. have identified specific regions within a whole developing zebrafish embryo where cellular transcription are responsive to retinoic acid signaling, which has been shown to be critical in normal vertebrate development. Lawson et al. have identified the downstream factors of receptor tyrosine kinase signaling that are necessary during *in vivo* blood vessel formation and arterial development. Additionally, Delta-Notch signaling in oligodendrocyte specification and the regulatory function of chemokine signaling in sensory cell migration have also been investigated by using transgenic zebrafish system. In a non-zebrafish system, Petrausch et al., by using transgenic goldfish, have identified the effects of several trophic factors converging on a purine-sensitive signaling mechanism that controls axonal outgrowth and the expression of multiple growth-associated proteins.

Despite the valuable discoveries achieved by using cell culture systems in studying signaling pathways in the last two decades, there are limitations to the use of cell culture, mainly being an artificial system that is isolated from the physiological environment of the animal. On the contrary, the most significant advantage of using transgenic fish for signaling pathway study is that transgenic fish provide an easy and representative system to dissect out signaling pathways and their biological implications with relevance to time and space in an animal. For example, while in many vertebrate species Wnt/b-catenin signaling is known to be involved in the development of neural crest cells and the specification of pigment cells derived from them at later stage, it is unclear which Wnts are involved, and when they are required. To address these issues, Lewis et al. established a transgenic zebrafish line that carried an inducible dominant-negative repressor of Wnt/β-catenin signaling, T-cell Factor 3 (Tcf-3a). The N-terminus of zebrafish Tcf-3a was conjugated with GFP as a reporter and was placed under the control of the zebrafish *hsp70* promoter. By activating the repressor to inhibit endogenous Wnt/b-catenin signaling at discrete times in development, a critical period for Wnt signaling in the initial induction of neural crest was defined. Wnt/b-catenin signaling is crucial for neural crest induction from the end of gastrulation through the 3-somite stage, and the cells lose sensitivity to transgene activation between the 3- and 6-somite stages despite the later requirement of the signaling for the pigment cell lineage. Researchers have also identified

a specific Wnt, Wnt8, and demonstrated its crucial role in initial neural crest induction, by blocking Wnt8 with antisense morpholino oligonucleotides. As demonstrated by these examples, transgenic fish is an excellent system to study a common scenario for many signaling pathways, such as BMP and Notch, in which the same signaling pathways are used in different places and at different times throughout development.

6.1.4 Human Disease Modeling

By providing a general proof of principle or reproducing specific aspects of the human diseases, animal models may be used to identify disease-associated factors or genes and to screen for new treatments for these diseases. In addition to murine models, fish have been used to study human diseases for decades, particularly in cancer research. In recent years, transgenic zebrafish and other species are emerging as valuable tools for modeling human diseases, especially in the domains of neurodegenerative diseases and cancers.

Among the neurodegenerative diseases, transgenic fish have been used for studies of Alzheimer's and Parkinson's diseases, two devastating diseases that affect tens of millions of people worldwide. The formation of neurofibrillary tangles (NFT), along with the β-amyloid plaques, are two central events in the development of Alzheimer's disease (AD). A microtubule-associated protein, tau, is critical in the process of NFT formation, in which tau protein is compromised in its normal association with microtubules, leading to the formation of paired helical filaments. Hyperphosphorylation of tau protein and mutations in the genes encoding the protein have been implicated in the pathogenesis of AD and a variety of hereditary dementias, collectively termed *frontotemporal dementia* with parkinsonism linked to chromosome 17 (FTDP-17).

The cellular consequence of NFT *in vivo* was first illustrated in a fish model by Hall et al. In this model, the giant neurons (anterior bulbar cells [ABCs]) of the central nervous system in the sea lamprey (*Petromyzon marinus*) were used to study the cytopathological changes caused by chronical overexpression of human tau protein. The unique advantages of this system include the already established morphological and cellular characterization of ABCs in detail, and most importantly, the ready accessibility of the neurons for manipulation by microinjection. This system demonstrates the association of filamentous tau deposits with stereotyped cytopathological changes, including the loss of dendritic microtubules and synapses, plasma membrane degeneration, and eventually the formation of extracellular tau deposits and cell death. Furthermore, this sequence of change is spatiotemporally correlated with the AD-related hyperphosphorylation of tau protein. By using this system, a lipid-soluble, low-molecular-weight proprietary compound has been shown to significantly retard the progressive degeneration of ABCs that express human tau protein, suggesting that development of tau-inhibitors may be a promising approach for AD treatment.

A transgenic zebrafish embryo model has also been established to study the functional consequence of FTDP-related mutations of tau protein in neurodegeneration. Zebrafish embryos at the 1–2 cell stage were microinjected with *tau* mutants under control of the neural specific GATA-2 promoter. An FTDP-17 mutant form of human tau has been shown to produce a cytoskeletal disruption that closely resembled the NFT in human disease.

This model system is therefore useful in the study of other mutant taus in vertebrate neurons *in vivo* and in the study of other molecules that are involved in the pathogenesis of human neurodegenerative diseases.

Parkinson's disease (PD) is characterized by the degeneration of mesencephalic dopaminergic neurons (MDNs) in the substantia nigra (SN). One of the most important functions of MDNs is the control of voluntary movement; loss of MDNs will eventually lead to a loss of dopamines, its metabolites, and the dopamine transporter. One of the most promising approaches to develop treatments for PD is to study the normal development of MDNs, in the hope that an understanding of their development will lead to means that can replenish damaged MDNs. Zebrafish have been used to screen for mutations that affect MDN development, and fish transgenesis can further assist in the understanding of the roles of these mutated genes in MDN development. Currently, a transgenic zebrafish model has been established to study the *ubiquitin C-terminal hydrolase L1* (*uch-L1*), a gene associated with the inherited form of PD. Zebrafish embryos were microinjected with *in vitro* synthesized zebrafish *uch-L1* mRNA or a plasmid construct containing GFP reporter under control of the *uch-L1* promoter. Together, these assays demonstrate the preferential activity of uch-L1 promoter in neurons and the spatiotemporal expression of the *uch-L1* in the ventral diencephalons, a functionally homologous region to the SN in human. The transgenic zebrafish system has thus provided further evidence for the connection of *uch-L1* in the development and degeneration of neural cells and can be further used to characterize other genes involved in neurodegenerative diseases.

In cancer research, stable lines of transgenic tumor-bearing zebrafish have been established to study the conserved cancer pathways involved in oncogene-induced lymphoblastic leukemia. In this model, zebrafish embryos at one-cell stage were injected with mouse oncogene *c-myc* (*mMyc*) and chimerical EGFP-*mMyc* transgenes under control of the zebrafish *Rag2* promoter, a lymphocyte specific promoter. In the resultant P_1 founder fish, the *myc*-induced clonally derived T-cell acute lymphoblastic leukemia was observed in almost 100% of the fish expressing green fluorescence. By tracing migration of the EGFP-tagged cells and confirming by RNA *in situ* hybridization, the leukemic T cells were clearly shown to arise in the thymus. The cells then spread to the gills and eyes and finally disseminate to the skeletal muscle and abdominal organs. The homing of leukemic T cells back to thymus was also illustrated in irradiated fish transplanted with GFP-tagged leukemic lymphoblasts. To establish stable lines of *mMyc*-transgenic fish, an *in vitro* fertilization technique was adopted because of the prompt onset of leukemia in the F_1 transgenic fish after germ-line transmission. By fertilizing the eggs collected from normal females with the sperm from 90-day-old leukemic males, the resultant transgene-positive progeny were shown to exhibit expansion of GFP-positive leukemic T cells from the thymus by one month of age. This transgenic zebrafish model can serve as a very powerful tool in the "forward-genetic" screening for modifier genes that are associated with the *myc*-induced leukemia or other types of cancers. The model can also be applied to large-scale screen for new drugs to treat or prevent human leukemia in the future.

In addition to the examples mentioned above, transgenic fish model can be applied in the studies of many other types of human diseases. We expect to see a plethora of such studies in the near future.

6.2 Growth Hormone Transgenic Fish

The initial drive of transgenic fish research came from attempts to increase production of economically important fish for human consumption. The worldwide supply of fishery products has traditionally depended upon commercial harvest of finfish, shellfish, and crustaceans from freshwater and marine-water sources. In recent years, while the worldwide commercial catch of fish experienced a sharp reduction, the worldwide demand of fish products has risen steeply. In order to cope with the demand of fish products, many countries have turned to aquaculture. Although aquaculture has the potential to meet the world demand for fish products, innovative strategies are required to improve its efficiency. What can transgenic fish technology offer in this regard?

There are three aspects of fish growth characteristics that could be improved for aquaculture: (1) increasing the initial growth rate of fry for an early head start; (2) enhancing somatic growth rate in adults to provide larger market body size; and (3) improving feed conversion efficiency of fish to achieve effective utilization of feeds. Among these three, enhanced somatic growth rates via manipulation of the GH gene shows considerable promise. Studies conducted by Agellon et al. and Paynter et al. showed that treatment of yearling rainbow trout and oysters with recombinant rainbow trout growth hormone resulted in significant growth enhancement. Similar results of growth enhancement in fish treated with recombinant GH have been reported by many investigators. These results point to the possibility of improving the somatic growth rate fish by manipulating fish GH or its gene.

Zhu et al. reported the first successful transfer of a human GH gene fused to a mouse metallothionein (MT) gene promoter into goldfish and loach. The F_1 offspring of these transgenic fish grew to be twice as large as their nontransgenic siblings. Since then, similar enhanced growth effect has been observed in the GH-transgenic fish of many other species, such as tilapia, carp, catfish, sea bream, and salmon. These studies have shown that the expression of a foreign human or fish GH gene could result in significant growth enhancement in the P_1, F_1, and F_2 transgenic fish. Additionally, Dunham et al. have demonstrated that transgenic common carp carrying rainbow trout GH transgene display, consistently in two consecutive generations, a favorable body shape, better dress-out yield, and better quality of flesh compared with controls. These studies have demonstrated that the application of GH-transgenic fish could be beneficial to the worldwide aquaculture, and hence could help to alleviate the starvation in many economically poor countries by providing more efficient, cheaper and yet high-quality protein sources.

6.3 Transgenic Fish with Enhanced Resistance to Pathogen Infection

Disease is one of the most severe bottlenecks in aquaculture. For the past few decades, efforts to control infectious diseases in commercially important teleosts have primarily focused on the development of suitable vaccines for fish and

selection of fish strains with robust resistance to infectious pathogens. Although effective vaccines have been developed for several important fish pathogens, the current vaccination practice is expensive, laborious, and time consuming. Genetic selection based on traditional crossbreeding techniques is time consuming and the outcome is frequently unpredictable, and sometimes unachievable because of the lack of desired genetic traits. More effective approaches for controlling fish disease in aquaculture are highly desirable.

Transgenic fish technology can facilitate the genetic selection process by directly modifying the undesirable genetic traits that confer vulnerability to pathogens or by introducing specific genes that are related to disease resistance into fish. The introduced transgenes can be fish-originated or characterized genes from other species. We have recently demonstrated the feasibility of introducing the antimicrobial peptide genes, cecropin B and cecropin P1, into fish to enhance disease resistance. Cecropins, first identified in *Cecropia* moth, are members of the antimicrobial peptide family that are evolutionarily conserved in species from insects to mammals. These peptides possess activities against a broad range of microorganisms, including bacteria, yeasts, and even viruses. *In vitro* studies conducted by Sarmasik et al. showed that recombinant cecropin B inhibited the propagation of common fish pathogens such as *Pseudomonas fluorescens, Aeromonas hydrophila, and Vibrio anguillarum*. Furthermore, Chiou et al. showed that cecropin and a designed analog, CF-17 peptide, effectively inhibited the replication of fish viruses such as infectious hematopoietic necrosis virus (IHNV), viral hemorrhagic septicemia virus (VHSV), snakehead rhabdovirus (SHRV), and infectious pancreatic necrosis virus (IPNV) (Fig. 3). More recently, our laboratory has shown that synthetic peptide CF-17 is also effective in inhibiting the propagation of insect baculovirus. Additionally, Jia et al. demonstrated the enhanced resistance to bacterial infection in fish which were continuously transfused with a cecropin–melittin hybrid peptide, CEME, and pleurocidin amide, a C-terminally amidated form of the natural flounder peptide. These results led to the hypothesis that production of disease-resistant fish strain may be achieved by introducing the known antimicrobial peptide genes into fish through the application of transgenic fish technology.

To test this hypothesis, we introduced gene constructs containing preprocecropin, procecropin, mature cecropin B, and cecropin P1 into medaka embryos by electroporation. The resulting F_2 transgenic medaka were subjected to bacterial challenges at a LD_{50} dose with *Pseudomonas fluorescens* and *Vibrio anguillarum*, respectively. The resulting relative percent survival (RPS) of the tested transgenic F_2 fish ranged from 72 to 100% against *P. fluorescens* and 25 to 75% against *V. anguillarum* (Table 4). These results clearly demonstrate the potential application of transgenic fish technology in producing fish with more robust resistance to infectious pathogens.

6.4 Transgenic Fish with Different Body Color

There are two main motives in producing fish with altered body color: (1) to generate novel varieties of ornamental fish with rare colors for the purpose of rearing as pets, and (2) to use the change of body color as indicators of environmental changes. GFP and several GFP variants are the most used genes for such purposes.

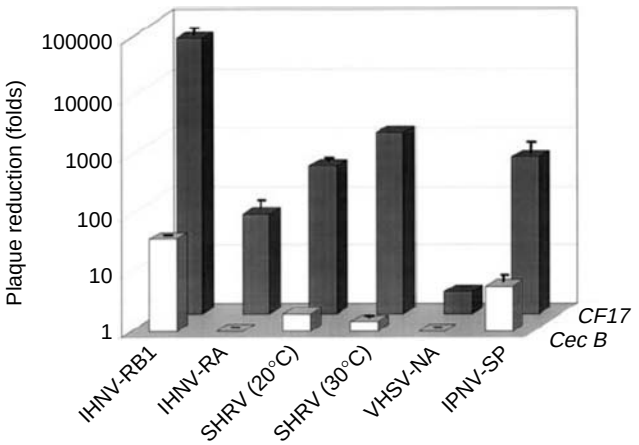

Fig. 3 Effect of peptides cecropin and CF-17 on inhibition of fish virus propagation *in vitro*. The antiviral activity of the peptides cecropin B (Cec B) and CF17 was evaluated against one nonenveloped birnavirus, the SP isolate of infectious pancreatic necrosis virus (IPNV), and three enveloped rhabdoviruses: the RB1 (type2) and RA (type 1) isolates of infectious hematopoietic necrosis virus (IHNV), snakehead rhabdovirus (SHRV), and an avirulent North America isolate (NA) of viral hemorrhagic septicemia virus (VHSV). Viral infection was carried out in EPC cells for SHRV, and CHSE-214 cells for the rest viruses. From Chiou, P.P., Lin, C.M., Perez, L., Chen, T.T. (2002) Effect of cecropin B and a synthetic analogue on propagation of fish viruses *in vitro, Mar. Biotechnol.* **4**, 294–302, with permission.

Additionally, melanin-concentrating hormone (MCH) isolated from chum salmon has also been exploited to generate transgenic medaka fish with altered body color. So far, researchers have successfully produced several colorful transgenic zebrafish displaying whole-body green, red, yellow, or orange fluorescent colors under daylight, dim light, or UV light. Some of these fish have been sterilized to avoid contaminating the wild population if they should accidentally be released from aquariums to the environments. For research purpose, mutant zebrafish and medaka with less interfering pigmentation for optical observation have also been developed. One such example is a line of "see-through" transgenic medaka which are transparent throughout their entire life, thus allowing clear visualization of GFP that is introduced into the fish as a reporter. The application of transgenic fish as biomonitors will be discussed in the next section.

6.5 Transgenic Fish as Environmental Biomonitors

Fish have long been used as models in environmental toxicology studies. Prompted by concerns of human health, tissues of wild-caught fish have been monitored as indicators for the presence of dangerous pollutants such as polycyclic hydrocarbons, oxidants, and heavy metals in the waters or in fish themselves. Different assays

Fig. 4 Examples of transgenic zebrafish that express GFP or RFP reporter in specific tissues. (a) A ''green-heart'' fish. Transgenic zebrafish carrying GFP reporter placed under control of the promoter of zebrafish cardiac myosin light chain 2 gene (*cmlc2*). Green fluorescence was intensively and specifically expressed in the myocardial cells located both around the heart chambers and the atrioventricular canal. The transgenic fish is kindly provided by Dr. Tsai (Huang, C.J., Tu, C.T., Hsiao, C.D., Hsieh, F.J., Tsai, H.J. (2003). Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish, *Dev. Dyn.* **228**, 30–40). (b) A fish with ''red guts.'' Transgenic zebrafish carrying RFP reporter placed under control of the promoter of zebrafish intestinal fatty acid binding protein gene (*i-fabp*). Homogenous red fluorescent signals are shown in the gut epithelial cells as demonstrated in this transverse section of a transgenic larva. n: notochord. From Her, G.M., Yeh, Y.H., Wu, J.L. (2003) 435-bp liver regulatory sequence in the liver fatty acid binding protein (L-FABP) gene is sufficient to modulate liver regional expression in transgenic zebrafish, *Dev. Dyn.* **227**, 347–356, with permission.

Tab. 4 Relative percent survival (RPS) of F_2 cecropin transgenic medaka challenged with *P. fluorescens* and *V. anguillarum*.

Transgenic lines	***P. fluorescene RPS***[a]	***V. anguillarum RPS***[a]
Preprocecropin B: CMV | pre | pro | cecropin B	100% (1)	50% (1)
Procecropin B: CMV | Ig | pro | cecropin B	72–100% (5)	25–50% (3)
Cecropin B: CMV | Ig | cecropin B	72–100% (2)	ND
Cecropin P1: CMV | Ig | cecropin P1	100% (2)	75% (1)

[a]RPS = [1 − %mortality of F_2 transgenic fish/%mortality of F_2 nontransgenic fish] × 100 CMV: cytomegalovirus promoter; Ig: catfish IgG signal peptide; ( ): number of F_2 families tested; ND: nondetermined.
Source: Adapted from Sarmasik, A., Warr, G., Chen, T.T. (2002) Production of transgenic medaka with increased resistance to bacterial pathogens, *Mar. Biotechnol.* **4**, 310–322 with permission.

have been employed to measure biological parameters that are impacted by toxic pollutants, including DNA damage, defense enzymes (e.g. glutathione peroxidase and superoxide dismutase), genes inducible by toxic chemicals (e.g. cytochrome P450 1A1 and 1A2), and factors that regulate redox potential (e.g. glutathione and ascorbic acid). Although these assays are very sensitive, they require specialized techniques and equipment, and thus cannot be performed in the field when the samples are collected. In addition, the obtained data from wild-caught fish samples give no indication of the exact time when those fish were exposed to the toxicants.

Transgenic fish can potentially serve as sensitive sentinels for aquatic pollution, with the benefit of avoiding the inconveniences mentioned above. The principle of such biomonitor fish is to introduce into the fish reporters that are easy to observe and are under the control of promoter/enhancer elements responsive to pollutants. GFP and its variants are thus the best reporters of choice for easy optical observation, whereas luciferase is for extreme sensitivity. So far, researchers have made significant progress in establishing transgenic zebrafish carrying luciferase or GFP reporter under control of the aromatic hydrocarbon response elements (AHREs), electrophile response elements (EPREs), or metal response elements (MREs). The specificity of each type of response elements is as follows: AHREs respond to numerous polycyclic hydrocarbons and halogenated coplanar molecules such as 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) and polychlorinated biphenyls; EPREs respond to quinones and numerous other potent electrophilic oxidants; MREs respond to heavy metal cations such as mercury, copper, nickel, cadmium, and zinc. In the zebrafish cell line ZEM2S, these three types of elements can drive the expression of luciferase reporter in a dose-dependent, chemical-class-specific manner in response to more than 20 environmental pollutants. The expression of GFP transgenes can be detected in the F_1 and, sometimes, in the F_2 generations; however, the transgenes are eventually lost following that. There are still technical obstacles in generating transgenic biomonitor fish, but the problems can be possibly overcome in the near future.

6.6 Other Examples of Transgenic Fish

Other applications include the use of transgenic fish as a model for studying mechanisms of seawater adaptation and generating cold-resistant fish carrying antifreeze proteins (AFPs). AFPs are found in many cold-water fish species, such as the winter flounder, ocean pout, and sea raven. Expression of *afp* genes permit these fish to survive in freezing seawater under as low as $-1.8\,^{\circ}C$. However, AFP transgenic Atlantic salmon does not demonstrate a measurable increase of cold tolerance.

Transgenic fish can potentially serve as bioreactors for producing therapeutic proteins beneficial to human health. The advantages of fish as bioreactors include (1) relatively low cost as compared to mammals such as cows; (2) easy maintenance; (3) producing large quantity of progeny; and (4) most importantly, no known viruses or prions have been reported to infect both human and fish. Successful expression of bioactive human proteins such as coagulation factor VII in fish embryos demonstrate this application of transgenic fish to produce therapeutic chemicals for human in the future.

7
Concerns and Future Perspectives

Transgenic fish serve as very valuable model systems for different disciplines of biological research as well as for human disease modeling. The application of transgenic fish technology to produce fish with beneficial traits, such as environmental biomonitors and GM food, is also on the rise. There has been groundbreaking discoveries made in the past few years based on the transgenic fish models, and such application is expected to continue to thrive in the future. The development of transgenic fish as environmental biomonitors is also encouraging and will continue to expand. The technique to produce transgenic fish has improved tremendously in the last two decades; nonetheless, there are still concerns about low efficiency, mosaic expression, nonfish components of the transgenes, and the long generation time for certain fish species. The following advances would thus be essential in developing the next generation transgenic fish: (1) developing more efficient mass gene transfer technologies; (2) developing targeted gene transfer technology such as embryonic stem cell gene transfer method; (3) identifying more suitable fish-origin promoters to direct the expression of transgenes at optimal levels and desired time; and (4) developing methods to shorten the time required to generate homozygous offspring.

Despite the promising future of transgenic fish as model systems in biological research, there are great concerns about the environmental impacts of such applications of transgenic fish technology to produce GM food or hobby pets. On the basis of mathematical modeling, a "Trojan Gene Effect" has been demonstrated in the male transgenic medaka carrying salmon growth hormone gene. Release of such fish into the environments could ultimately lead to the extinction of the wild type population owing to their advantage in mating with the wild type female. The study provides a sound analysis of the potential impact of the accidental release of transgenic fish into the environment, reminding us of the delicate interaction between a newly introduced species and the native population and environment. However, we would reason that the potential environmental impact by transgenic fish is controllable and even avoidable. First of all, the "Trojan Gene Effect" might not necessarily exist for other types of transgenes or in other fish species. In practice, transgenic fish can be maintained in confined areas to avoid the escape of fish into the nearby waters. Moreover, as the desirable genetic traits can be introduced into fish, the undesirable behavior traits could be possibly removed from the transgenic population by screening the behavior pattern of each transgenic individual. Ultimately, establishment of sterile transgenic population, such as a sterile triploid all-female strain, can be applied to avoid the spreading of transgenes into the wild population. Ideally, such sterility should be reversible; otherwise it would mean the loss of the goose that lays golden eggs. Finally, from a historical perspective, the concerns on the environmental impacts in the early days of the recombinant DNA technology have not been realized and many lessons can be learned from this experience.

Acceptance of GM fish as food by society is a key factor for the development of transgenic fish as a new source of cheaper and high-quality food. Largely, the concern is due to the public perception that GM food is unsafe to human health. Safety is the ultimate goal of GM fish development and relies on thorough characterization of

the transgenes and safety assessment of the GM fish, resembling the development of new drugs. Unfortunately, it is a fact that there is always a subpopulation of people who are allergic, to different degrees, to many of the existing natural food sources, such as peanuts and shellfish. Therefore, the challenge is great in developing GM fish with absolutely no adverse effect to every single individual in the human population. A thorough evaluation process to fully analyze any health risk of GM fish to human should be established and the information should be clearly delivered to the public and well marked on the food product. Development of "gene-inactivated" transgenic fish may be more acceptable by people who are concerned of the safety of the "new" genes that are introduced into fish. One candidate of such GM fish is transgenic fish strains with their myostatin gene being inactivated. Inactivation of myostatin gene causes double muscling in animals, resulting in significant increase in mass of skeletal muscle. Natural mutation of myostatin gene has been found in the meaty Belgian Blue and Piedmontes cattles, and the double muscling effect has been observed in mice whose myostatin gene is artificially inactivated. Such "gene-inactivation" approach would be very valuable in producing transgenic fish with more flesh and higher food conversion rate, particularly for the economically important aquaculture species.

In summary, the application of transgenic fish has resulted in many valuable discoveries in basic biological science, many of which could lead to the next breakthrough for new cures for human diseases. On the other hand, the potential of GM fish as bioreactors or as new food sources should not be deterred by the environmental and safety concerns but instead it requires more research efforts in the future to develop new GM species with beneficial traits that are also safe to human health and carry no harm to the environment.

See also Genetics, Molecular Basis of; Living Organism (Animal) Patents; Nuclear Transfer for Cloning Animals; Reporter Gene System: Green Fluorescent Protein; Zebrafish (*Danio rerio*) Genome and Genetics.

Bibliography

Books and Reviews

Boman, H.G. (1995) Peptide antibiotics and their role in innate immunity, *Annu. Rev. Immunol.* **13**, 61–92.

Chen, T.T., Powers, D.A. (1990) Transgenic fish, *Trends Biotechnol.* **8**, 209–215.

Chen, T.T., Lin, C-M., Zhu, Z., Gonzalez-Villasenor, L.I., Dunham, R.A., Powers, D.A. (1990) Gene transfer, expression and inheritance of rainbow trout growth hormone genes in carp and loach, in: Church, R. (Ed.) *Transgenic Models in Medicine and Agriculture*, Wiley, New York, pp. 127–139.

Di Berardino, M.A., McKinnell, R.G., Wolf, D.P. (2003) The golden anniversary of cloning: a celebratory essay, *Differentiation* **71**, 398–401.

Duxbury, M.S., Whang, E.E. (2004) RNA interference: a practical approach, *J. Surg. Res.* **117**, 339–344.

Götz, J., Streffer, J.R., David, D., Schild, A., Hoerndli, F., Pennanen, L., Kurosinski, P., Chen, F. (2004) Transgenic animal models of Alzheimer's disease and related disorders: histopathology, behavior and therapy, *Mol. Psychiatry* **9**, 1–20.

Hackett, P.B. (1993) The molecular biology of transgenic fish, in: Hochachka, P.W., Mommsen, T.P. (Eds.) *Biochemistry and*

Molecular Biology of Fish, 2, Elsevier, B.V., pp. 207–240.

Muller, F., Blader, P., Strahle, U. (2002) Search for enhancers: teleost models in comparative genomic and transgenic analysis of cis regulatory elements, *BioEssays* **24**, 564–572.

Stern, H.M., Zon, L.I. (2003) Cancer genetics and drug discovery in the zebrafish, *Nat. Rev. Cancer* **3**, 533–539.

Udvadia, A.J., Linney, E. (2003) Windows into development: historic, current, and future perspectives on transgenic zebrafish, *Dev. Biol.* **256**, 1–17.

Primary Literature

Agellon, L.B., Emery, C.J., Jones, J.M., Davies, S.L., Dingle, A.D., Chen, T.T. (1988) Growth hormone enhancement by genetically engineered rainbow trout growth hormone, *Can. J. Fish. Aquat. Sci.* **45**, 146–151.

Amsterdam, A., Lin, S., Hopkins, N. (1995) The *Aequorea victoria* green fluorescent protein can be used as a reporter in live zebrafish embryos, *Dev. Biol.* **171**, 123–129.

Ballagi-Pordany, A., Ballagi-Pordany, A., Funa, K. (1991) Quantative determination of mRNA phenotypes by polymerase chain reaction, *Anal. Biochem.* **196**, 88–94.

Boonanuntanasarn, S., Yoshizaki, G., Takeuchi, T. (2003) Specific gene silencing using small interfering RNAs in fish embryos, *Biochem. Biophys. Res. Commun.* **310**, 1089–1095.

Boonanuntanasarn, S., Yoshizaki, G., Takeuchi, Y., Morita, T., Takeuchi, T. (2002) Gene knock-down in rainbow trout embryos using antisense Morpholino phosphorodiamidate oligonucleotides, *Mar. Biotechnol.* **4**, 256–266.

Buono, R.J., Linser, P.J. (1992) Transient expression of RSVCAT in transgenic zebrafish made by electroporation, *Mol. Mar. Biol. Biotechnol.* **1**, 271–275.

Burns, J.C., Friedmann, T., Driever, W., Burrascano, M., Yee, J.K. (1993) VSV-G pseudotyped retroviral vector: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells, *Proc. Natl. Acad. Sci. USA* **90**, 8033–8037.

Carvan, M.J. III, Dalton, T.P., Stuart, G.W., Nebert, D.W. (2000) Transgenic zebrafish as sentinels for aquatic pollution, *Ann. N.Y. Acad. Sci.* **919**, 133–147.

Chen, T.T., Powers, D.A., Lin, C.M., Kight, K., Hayat, M., Chatakondi, N., Ramboux, A.C., Duncan, P.L., Dunham, R.A. (1993) Expression and inheritance of RSVLTR-rtGH1 cDNA in common carp, *Cyprinus carpio*, *Mol. Mar. Biol. Biotechnol.* **2**, 88–95.

Chiou, P.P., Lin, C.M., Perez, L., Chen, T.T. (2002) Effect of cecropin B and a synthetic analogue on propagation of fish viruses *in vitro*, *Mar. Biotechnol.* **4**, 294–302.

Cotten, M., Jennings, P. (1989) Ribozyme mediated destruction of RNA *in vivo*, *EMBO J.* **8**, 3861–3866.

Du, S.J., Gong, G.L., Fletcher, G.L., Shears, M.A., King, M.J., Idler, D.R., Hew, C.L. (1992) Growth enhancement in transgenic Atlantic salmon by the use of an "all fish" chimeric growth hormone gene construct, *Biotechnology* **10**, 176–181.

Dunham, R.A., Chatakondi, N., Nichols, A.J., Kucuktas, H., Chen, T.T., Powers, D.A., Weete, J.D., Cummins, K., Lovell, R.T. (2002) Effect of rainbow trout growth hormone complementary DNA on body shape, carcass yield, and carcass composition of F1 and F2 transgenic common carp (*Cyprinus carpio*), *Mar. Biotechnol.* **4**, 604–611.

Dunham, R.A., Ramboux, A.C., Duncan, P.L., Hayat, M., Chen, T.T., Lin, C-M., Kight, K., Gonzalez-Villasenor, L.I., Powers, D.A. (1992) Transfer, expression and inheritance of salmonid growth hormone genes in channel catfish, *Ictalurus punctatus*, and effects on performance traits, *Mol. Mar. Biol. Botechnol.* **1**, 380–389.

Fletcher, G.L., Davis, P.L. (1991) Transgenic fish for aquaculture, in: Setlow, J.K. (Ed.) *Genetic Engineering* Vol. 13, Plenum Publishing, New York, pp. 331–370.

Fletcher, G.L., Shears, M.A., King, J.J., Davis, P.L., Hew, C.L. (1988) Evidence of antifreeze protein gene transfer in Atlantic salmon (*Salmo salar*), *Can. J. Fish. Aquat. Sci.* **45**, 352–357.

Fronhoffs, S., Totzke, G., Stier, S., Wernert, N., Rothe, M., Bruning, T., Koch, B., Sachinidis, A., Vetter, H., Ko, Y. (2002) A method for the rapid construction of cRNA standard curves in quantitative real-time reverse transcription polymerase chain reaction, *Mol. Cell. Probes* **16**, 99–110.

Gill, J.A., Stumper, J.P., Donaldson, E.M., Dye, H.M. (1985) Recombinant chicken and bovine

growth hormone in cultured juvenile Pacific salmon, *Biotechnology* **3**, 4306–4310.

Goldman, D., Hankin, M., Li, Z., Dai, X., Ding, J. (2001) Transgenic zebrafish for studying nervous system development and regeneration, *Transgenic Res.* **10**, 21–33.

Gong, Z., Wan, H., Tay, T.L., Wang, H., Chen, M., Yan, T. (2003) Development of transgenic fish for ornamental and bioreactor by strong expression of fluorescent proteins in the skeletal muscle, *Biochem. Biophys. Res. Commun.* **308**, 58–63.

Gong, Z., Ju, B., Wang, X., He, J., Wan, H., Sudha, P.M., Yan, T. (2002) Green fluorescent protein expression in germ-line transmitted transgenic zebrafish under a stratified epithelial promoter from keratin8, *Dev. Dyn.* **223**, 204–215.

Hall, G.F., Lee, S., Yao, J. (2002) Neurofibrillary degeneration can be arrested in an in vivo cellular model of human tauopathy by application of a compound which inhibits tau filament formation *in vitro*, *J. Mol. Neurosci.* **19**, 253–260.

Hall, G.F., Lee, V.M., Lee, G., Yao, J. (2001) Staging of neurofibrillary degeneration caused by human tau overexpression in a unique cellular model of human tauopathy, *Am. J. Pathol.* **158**, 235–246.

Halloran, M.C., Sato-Maeda, M., Warren, J.T., Su, F., Lele, Z., Krone, P.H., Kuwada, J.Y., Shoji, W. (2000) Laser-induced gene expression in specific cells of transgenic zebrafish, *Development* **127**, 1953–1960.

Her, G.M., Chiang, C.C., Wu, J.L. (2004) Zebrafish intestinal fatty acid binding protein (I-FABP) gene promoter drives gut-specific expression in stable transgenic fish, *Genesis* **38**, 26–31.

Her, G.M., Yeh, Y.H., Wu, J.L. (2003) 435-bp liver regulatory sequence in the liver fatty acid binding protein (L-FABP) gene is sufficient to modulate liver regional expression in transgenic zebrafish, *Dev. Dyn.* **227**, 347–356.

Higashijima, S., Hotta, Y., Okamoto, H. (2000) Visualization of cranial motor neurons in live transgenic zebrafish expressing green fluorescent protein under the control of the islet-1 promoter/enhancer, *J. Neurosci.* **20**, 206–218.

Higashijima, S., Masino, M.A., Mandel, G., Fetcho, J.R. (2003) Imaging neuronal activity during zebrafish behavior with a genetically encoded calcium indicator, *J. Neurophysiol.* **90**, 3986–3997.

Higashijima, S., Okamoto, H., Ueno, N., Hotta, Y., Eguchi, G. (1997) High-frequency generation of transgenic zebrafish which reliably express GFP in whole muscles or the whole body by using promoters of zebrafish origin, *Dev. Biol.* **192**, 289–299.

Holden, C. (2003) "That special glow" (a news report in Random Samples). Holden, C. (Ed.), *Science* **300**, 1368.

Howard, R.D., DeWoody, J.A., Muir, W.M. (2004) Transgenic male mating advantage provides opportunity for Trojan gene effect in a fish, *Proc. Natl. Acad. Sci. USA* **101**, 2934–2938.

Hsiao, C.D., Hsieh, F.J., Tsai, H.J. (2001) Enhanced expression and stable transmission of transgenes flanked by inverted terminal repeats from adeno-associated virus in zebrafish, *Dev. Dyn.* **220**, 323–336.

Huang, C.J., Tu, C.T., Hsiao, C.D., Hsieh, F.J., Tsai, H.J. (2003) Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish, *Dev. Dyn.* **228**, 30–40.

Huang, H., Vogel, S.S., Liu, N., Melton, D.A., Lin, S. (2001) Analysis of pancreatic development in living transgenic zebrafish embryos, *Mol. Cell. Endocrinol.* **177**, 117–124.

Hwang, G., Muller, F., Rahman, M.A., Williams, D.W., Murdock, P.J., Pasi, K.J., Goldspink, G., Farahmand, H., Maclean, N. (2004) Fish as Bioreactors: Transgene Expression of Human Coagulation Factor VII in Fish Embryos, *Mar. Biotechnol.* (http://www.springerlink.com/media/3G267U4FJG5JWHE03N2G/contributions/C/X/A/N/CXAN5MRVF4817U62_html/fulltext.html. Published online Apr. 29, 2004)

Inoue, K., Takei, Y. (2003) Asian medaka fishes offer new models for studying mechanisms of seawater adaptation, *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **136**, 635–645.

Jessen, J.R., Willett, C.E., Lin, S. (1999) Artificial chromosome transgenesis reveals long-distance negative regulation of rag1 in zebrafish, *Nat. Genet.* **23**, 15–16.

Jesuthasan, S., Subburaju, S. (2002) Gene transfer into zebrafish by sperm nuclear transplantation, *Dev. Biol.* **242**, 88–95.

Jia, X., Patrzykat, A., Devlin, R.H., Ackerman, P.A., Iwama, G.K., Hancock, R.E. (2000)

Antimicrobial peptides protect coho salmon from Vibrio anguillarum infections, *Appl. Environ. Microbiol.* **66**, 1928–1932.

Kambadur, R., Sharma, M., Smith, T.P., Bass, J.J. (1997) Mutations in myostatin (GDF8) in double-muscled Belgian Blue and Piedmontese cattle, *Genome Res.* **7**, 910–916.

Kinoshita, M., Morita, T., Toyohara, H., Hirata, T., Sakaguchi, M., Ono, M., Inoue, K., Wakamatsu, Y., Ozato, K. (2001) Transgenic medaka overexpressing a melanin-concentrating hormone exhibit lightened body color but no remarkable abnormality, *Mar. Biotechnol.* **3**, 536–543.

Knaut, H., Steinbeisser, H., Schwarz, H., Nusslein-Volhard, C. (2002) An evolutionary conserved region in the vasa 3'UTR targets RNA translation to the germ cells in the zebrafish, *Curr. Biol.* **12**, 454–466.

Krøvel, A.V., Olsen, L.C. (2002) Expression of a vas::EGFP transgene in primordial germ cells of the zebrafish, *Mech. Dev.* **116**, 141–150.

Langenau, D.M., Traver, D., Ferrando, A.A., Kutok, J.L., Aster, J.C., Kanki, J.P., Lin, S., Prochownik, E., Trede, N.S., Zon, L.I., Look, A.T. (2003) *Myc*-induced T cell leukemia in transgenic zebrafish, *Science* **299**, 887–890.

Lawson, N.D., Mugford, J.W., Diamond, B.A., Weinstein, B.M. (2003) Phospholipase C gamma-1 is required downstream of vascular endothelial growth factor during arterial development, *Genes Dev.* **17**, 1346–1351.

Lawson, N.D., Weinstein, B.M. (2002) *In vivo* imaging of embryonic vascular development using transgenic zebrafish, *Dev. Biol.* **248**, 307–318.

Lee, K.Y., Huang, H., Ju, B., Yang, Z., Lin, S. (2002) Cloned zebrafish by nuclear transfer from long-term-cultured cells, *Nat. Biotechnol.* **20**, 795–799.

Lewis, J.L., Bonner, J., Modrell, M., Ragland, J.W., Moon, R.T., Dorsky, R.I., Raible, D.W. (2004) Reiterated Wnt signaling during zebrafish neural crest development, *Development* **131**, 1299–1308.

Li, Q., Shirabe, K., Kuwada, J.Y. (2004) Chemokine signaling regulates sensory cell migration in zebrafish, *Dev. Biol.* **269**, 123–136.

Lin, S., Gaiano, N., Culp, P., Burns, J.C., Friedmann, T., Yee, J-K., Hopkins, N. (1994) Integration and germ-line transmission of a pseudotyped retroviral vector in zebrafish, *Science* **265**, 666–668.

Liu, Y., Berndt, J., Su, F., Tawarayama, H., Shoji, W., Kuwada, J.Y., Halloran, M.C. (2004) Semaphorin3D guides retinal axons along the dorsoventral axis of the tectum, *J. Neurosci.* **24**, 310–318.

Long, Q., Meng, A., Wang, H., Jessen, J.R., Farrell, M.J., Lin, S. (1997) GATA-1 expression pattern can be recapitulated in living transgenic zebrafish using GFP reporter gene, *Development* **124**, 4105–4111.

Lu, J.K., Burns, J.C., Chen, T.T. (1994) Retrovirus-mediated transfer and expression of transgenes in medaka, in: *Proceedings of the Third International Marine Biotechnology Conference at Tromso, Norway*, pp. 72.

Lu, J.K., Fu, B.H., Wu, J.L., Chen, T.T. (2002) Production of transgenic silver sea bream (*Sparus sarba*) by different gene transfer methods, *Mar. Biotechnol.* **4**, 328–337.

Lu, J.K., Chen, T.T., Allen, S.K., Matsubara, T., Burns, J.C. (1996) Production of transgenic dwarf surfclams, *Mulinia lateralis*, with Pantropic retroviral vectors, *Proc. Natl. Acad. Sci. USA* **93**, 3482–3486.

Lu., J.K., Chrisman, C.L., Andrisani, O.M., Dixon, J.E., Chen, T.T. (1992) Integration expression and germ-line transmission of foreign growth hormone genes in medaka, *Oryzias latipes*, *Mol. Mar. Biol. Biotechnol.* **1**, 366–375.

Martinez, R., Estrada, M.P., Berlanga, J., Guillen, I., Hernandez, O., Cabrera, E., Pimentel, R., Morales, R., Herrera, F., Morales, A., Pina, J.C., Abad, Z., Sanchez, V., Melamed, P., Lleonart, R., de la Fuente, J. (1996) Growth enhancement in transgenic tilapia by ectopic expression of tilapia growth hormone, *Mol. Mar. Biol. Biotechnol.* **5**, 62–70.

McPherron, A.C., Lee, S.J. (1997) Double muscling in cattle due to mutations in the myostatin gene, *Proc. Natl. Acad. Sci. USA* **94**, 12457–12461.

Moav, Boaz, Liu, Z., Groll, Y., Hackett, P.R. (1992) Selection of promoters for gene transfer into fish, *Mol. Mar. Biol. Biotechnol.* **1**, 338–345.

Moriyama, S., Takahashi, A., Hirano, T., Kawauchi, H. (1990) Salmon growth hormone is transported into the circulation of rainbow trout (*Oncorhynchus mykiss*) after intestinal administration, *J. Comp. Physiol. B* **160**, 251–260.

Neumann, E., Schaefer-Ridder, M., Wang, Y., Hofschneider, P.H. (1982) Gene transfer into

mouse lyoma cells by electroporation in high electric fields, *EMBO J.* **1**, 841–845.

Ozato, K., Kondoh, H., Inohara, H., Iwanatsu, T., Wakamatsu, Y., Okada, T.S. (1986) Production of transgenic fish: introduction and expression of chicken δ-crystallin gene in medaka embryos, *Cell Differ.* **19**, 237–244.

Palmiter, R.D., Brinster, R.D. (1986) Germ-line transformation in mice, *Ann. Rev. Genet.* **20**, 465–499.

Park, H.C., Appel, B. (2003) Delta-Notch signaling regulates oligodendrocyte specification, *Development* **130**, 3747–3755.

Park, H.C., Kim, C.H., Bae, Y.K., Yeo, S.Y., Kim, S.H., Hong, S.K., Shin, J., Yoo, K.W., Hibi, M., Hirano, T., Miki, N., Chitnis, A.B., Huh, T.L. (2000) Analysis of upstream elements in the HuC promoter leads to the establishment of transgenic zebrafish with fluorescent neurons, *Dev. Biol.* **227**, 279–293.

Pauls, S., Geldmacher-Voss, B., Campos-Ortega, J.A. (2001) A zebrafish histone variant H2A.F/Z and a transgenic H2A.F/Z:GFP fusion protein for in vivo studies of embryonic development, *Dev. Genes Evol.* **211**, 603–610.

Paynter, K., Chen, T.T. (1991) Biological activity of biosynthetic rainbow trout growth hormone in the eastern oyster, (*Crassostrea virginica*), *Biol. Bull.* **181**, 459–462.

Perz-Edwards, A., Hardison, N.L., Linney, E. (2001) Retinoic acid-mediated gene expression in transgenic reporter zebrafish, *Dev. Biol.* **229**, 89–101.

Petrausch, B., Tabibiazar, R., Roser, T., Jing, Y., Goldman, D., Stuermer, C.A., Irwin, N., Benowitz, L.I. (2000) A purine-sensitive pathway regulates multiple genes involved in axon regeneration in goldfish retinal ganglion cells, *J. Neurosci.* **20**, 8031–8041.

Pi, W., Yang, Z., Wang, J., Ruan, L., Yu, X., Ling, J., Krantz, S., Isales, C., Conway, S.J., Lin, S., Tuan, D. (2004) The LTR enhancer of ERV-9 human endogenous retrovirus is active in oocytes and progenitor cells in transgenic zebrafish and humans, *Proc. Natl. Acad. Sci. USA* **101**, 805–810.

Potter, H., Weir, L., Leder, P. (1984) Enhancer-dependent expression of human *k* immunoglobulin genes introduced into mouse pre-B lymphocytes by electroporation, *Proc. Natl. Acad. Sci. USA* **81**, 7161–7165.

Powers, D.A., Hereford, L., Cole, T., Creech, K., Chen, T.T., Lin, C.M., Kight, K., Dunham, R.A. (1992) Electroporation: a method for transferring genes into gametes of zebrafish (*Brachydanio rerio*), channel catfish (*Ictalurus punctatus*), and common carp (*Cyprinus carpio*), *Mol. Mar. Biol. Biotechnol.* **1**, 301–308.

Rahman, M.A., Ronyai, A., Engidaw, B.Z., Jauncey, K., Hwang, G-L., Smith, A., Roderick, E., Penman, D., Varadi, L., Maclean, N. (2001) Growth and nutritional trials on transgenic Nile tilapia containing an exogenous fish growth hormone gene, *J. Fish Biol.* **59**, 62–78.

Reinhard, E., Nedivi, E., Wegner, J., Skene, J.H., Westerfield, M. (1994) Neural selective activation and temporal regulation of a mammalian GAP-43 promoter in zebrafish, *Development* **120**, 1767–1775.

Riddle, R., Pollock, J.D. (2003) Making connections: the development of mesencephalic dopaminergic neurons, *Brain Res. Dev. Brain Res.* **147**, 3–21.

Sarmasik, A., Warr, G., Chen, T.T. (2002) Production of transgenic medaka with increased resistance to bacterial pathogens, *Mar. Biotechnol.* **4**, 310–322.

Schreurs, R.H., Legler, J., Artola-Garicano, E., Sinnige, T.L., Lanser, P.H., Seinen, W., Van der Burg, B. (2004) *In vitro* and *in vivo* antiestrogenic effects of polycyclic musks in zebrafish, *Environ. Sci. Technol.* **38**, 997–1002.

Sekine, S., Miizukzmi, T., Nishi, T., Kuwana, Y., Saito, A., Sato, M., Itoh, H., Kawauchi, H. (1985) Cloning and expression of cDNA for salmon growth hormone in *E. coli*, *Proc. Natl. Acad. Sci. USA* **82**, 4306–4310.

Shears, M.A., Fletcher, G.L., Hew, C.L., Gauthier, S., Davies, P.L. (1991) Transfer, expression, and stable inheritance of antifreeze protein genes in Atlantic salmon (Salmo salar), *Mol. Mar. Biol. Biotechnol.* **1**, 58–63.

Shigekawa, K., Dower, W.J. (1988) Electroporation of eukaryotes and prokaryotes: a general approach to introduction of macromolecules into cells, *Biotechniques* **6**, 742–751.

Sin, F.Y., Walker, S.P., Symonds, J.E., Mukherjee, U.K., Khoo, J.G., Sin, I.L. (2000) Electroporation of salmon sperm for gene transfer: efficiency, reliability, and fate of transgene, *Mol. Reprod. Dev.* **56**(Suppl. 2), 285–288.

Son, O.L., Kim, H.T., Ji, M.H., Yoo, K.W., Rhee, M., Kim, C.H. (2003) Cloning and expression analysis of a Parkinson's disease gene, uch-L1, and its promoter in zebrafish, *Biochem. Biophys. Res. Commun.* **312**, 601–607.

Stuart, G.W., McMurray, J.V., Westerfield, M. (1988) Replication, integration, and stable germ-line transmission of foreign sequence injected into early zebrafish embryos, *Development* **109**, 403–412.

Stuart, G.W., Vielkind, J.V., McMurray, J.V., Westerfield, M. (1990) Stable lines of transgenic zebrafish exhibit reproduction patterns of transgene expression, *Development* **109**, 293–296.

Symonds, J.E., Walker, S.P., Sin, F.Y.T. (1994) Development of mass gene transfer method in Chinook salmon: optimization of gene transfer by electroporated sperm, *Mol. Mar. Biol. Biotechnol.* **3**, 104–111.

Thermes, V., Grabher, C., Ristoratore, F., Bourrat, F., Choulika, A., Wittbrodt, J., Joly, J.S. (2002) I-SceI meganuclease mediates highly efficient transgenesis in fish, *Mech. Dev.* **118**, 91–98.

Tomasiewicz, H.G., Flaherty, D.B., Soria, J.P., Wood, J.G. (2002) Transgenic zebrafish model of neurodegeneration, *J. Neurosci. Res.* **70**, 734–745.

Tseng, F.S., Lio, I.C., Tsai, H.J. (1994). Introducing the exogenous growth hormone cDNA into loach (*Misgurnus anguillicaudatus*) eggs via electroporated sperms as carrier, in: *3rd International Marine Biotechnology Conference*, Tromso, Norway. abstract pp. 71.

Wakamatsu, Y., Pristyazhnyuk, S., Kinoshita, M., Tanaka, M., Ozato, K. (2001b) The see-through medaka: a fish model that is transparent throughout life, *Proc. Natl. Acad. Sci. USA* **98**, 10046–10050.

Wakamatsu, Y., Ju, B., Pristyaznhyuk, I., Niwa, K., Ladygina, T., Kinoshita, M., Araki, K., Ozato, K. (2001a) Fertile and diploid nuclear transplants derived from embryonic cells of a small laboratory fish, medaka (Oryzias latipes), *Proc. Natl. Acad. Sci. USA* **98**, 1071–1076.

Wan, H., He, J., Ju, B., Yan, T., Lam, T.J., Gong, Z. (2002) Generation of two-color transgenic zebrafish using the green and red fluorescent protein reporter genes gfp and rfp, *Mar. Biotechnol.* **4**, 146–154.

Wang, T.M., Chen, Y.H., Liu, C.F., Tsai, H.J. (2002) Functional analysis of the proximal promoter regions of fish rhodopsin and myf-5 genes using transgenesis, *Mar. Biotechnol.* **4**, 247–255.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., Campbell, K.H. (1997) Viable offspring derived from fetal and adult mammalian cells, *Nature* **385**, 810–813.

Xiao, T., Shoji, W., Zhou, W., Su, F., Kuwada, J.Y. (2003) Transmembrane sema4E guides branchiomotor axons to their targets in zebrafish, *J. Neurosci.* **23**, 4190–4198.

Zhang, P., Hayat, M., Joyce, C., Gonzalez-Villasenor, L.I., Lin, C-M., Dunham, R., Chen, T.T., Powers, D.A. (1990) Gene transfer, expression and inheritance of pRSV-Rainbow Trout-GH-cDNA in the carp, *Cyprinus carpio* (Linnaeus), *Mol. Reprod. Dev.* **25**, 3–13.

Zhu, Z., Li, G., He, L., Chen, S.Z. (1985) Novel gene transfer into the goldfish (*Carassius auratus* L 1758), *Angew Ichthyol.* **1**, 31–34.

Transgenic Mice: *see* Transgenic Mice in Biomedical Research

Transgenic Mice in Biomedical Research

J. Willem Voncken and Marten Hofker
University of Maastricht, Maastricht, The Netherlands

1 A Brief History of Mice 508
1.1 Mouse Genetics 508
1.2 Molecular Mouse Genetics 510
1.2.1 Brief Overview of Developments in Transgenesis 511
1.2.2 Brief Overview of Developments in Gene Targeting 512

2 Transgenesis; Basics and Technology 513
2.1 Transgene Design 513
2.1.1 Origin and Structure of the Transgene 513
2.1.2 Regulatory Elements 515
2.2 Pronuclear Injection 518
2.3 Biotechnological Procedures 518

3 Gene Targeting; Basics and Technology 521
3.1 Targeting Vector Design 521
3.2 Embryonic Stem (ES) Cells 524
3.3 Genesis of Mosaic Embryos 524
3.4 Biotechnological Procedures 526

4 Refinements to the Models 527
4.1 Fully ES Cell–derived Embryos 527
4.2 Conditional Transgenesis 529
4.3 Conditional Gene Targeting 530
4.3.1 Controlled CRE Expression 532
4.4 Knockins: Humanized Mice 533

5 Random Mutagenesis in ES Cells 536
5.1 Gene Trap Mutagenesis 537

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

5.2 Retroviruses and Transposable Elements in Random Mutagenesis 538
5.3 Random Enu Mutation 539

6 Phenotype Analysis 539

7 Perspectives 540
7.1 Partial Transgenesis; Viral Shuttles 540
7.2 RNA Interference; Posttranscriptional Gene Silencing 541

Bibliography 542
Books and Reviews 542
Primary Literature 543
Websites (References Indicated as they Appear in the Original Text) 545

Keywords

Autologous
Derived from endogenous sequences.

Binary Model
Mouse model in which gene activity can be switched on and off.

Chimeric Mouse
Mouse derived from donor ES cells and recipient embryo, displaying coat color mosaicism.

Chromosome Substitution
A single full-length chromosome from one inbred strain has been transferred onto the genetic background of a second strain by repeated backcrossing.

Conditional Model
Mouse model in which a modification can be activated on command.

CRE-LoxP
Bacterio phage P1-derived recombination system: the recombinase CRE-mediated recombination between two identical recombinase recognition sites: LoxP.

CRE-MER
Fusion of a CRE recombinase with a mutated ligand-binding domain of the estrogen receptor.

Ectopic Expression
Expression out of place or out of time.

Embryonic Stem Cells
Pluripotent stem cells derived from preimplantation embryos.

Gene Targeting
Modification of endogenous genes through use of replacement vectors.

Gene Trapping
Process by which transcription of a marker gene is coupled to that of an endogenous locus in which the gene trap vector is inserted and transcripts are trapped.

Heterologous
Derived from exogenous sequences.

Inbred Strain
A mouse strain bred to homozygocity for all loci by 20 brother-to-sister back crosses.

Insertional Mutagenesis
Gene modification by insertion of foreign DNA vectors.

Isogenic DNA
DNA derived from the same genetic background.

Knockin
Targeted mutation of an existing locus, mostly replacing endogenous coding sequences with functional sequences

Pluripotent
Unlimited differentiation capacity.

Recombinant Inbred Strain
Recombinant inbred strains are derived from crosses of two inbred strains.

RNA interference
RNA-mediated posttranscriptional gene silencing.

Spatio-temporal Expression Pattern
Gene activity determined in time and location.

Tetraploid
Containing four haploid genome copies.

Transgene
A(n expression) construct that is inserted into the mouse genome.

Genetically modified mouse models have become a crucial tool in present day biomedical research. Modifying genes in mice provides an excellent approach to unravel gene function at the molecular and cellular level, as well as its role in the physiology and pathology of the intact organism. By doing so, numerous models have been developed that reliably replicate diseases in humans, thereby providing insight in disease mechanisms and paving the way toward prevention and therapy. Crucial for these advances has been the ability to modulate gene expression at will. It is possible to increase or decrease gene expression, or eliminate the expression of a gene completely. These alterations can be made cell type–specific and even inducible or reversible. Moreover, it is possible to replace mouse genes with the cognate human gene carrying a specific mutation. This chapter will deal with the essential approaches in transgenic mouse research and its impact on biomedical research.

1
A Brief History of Mice

1.1
Mouse Genetics

The resources directed toward studying the mouse have been exponentially growing. After the completion of the human genome sequence, the mouse was the second mammalian genome for which the complete sequence was determined. There is no mammalian system that comes close to rival the molecular and genetic possibilities that the mouse has to offer to researchers. Many physiological methods that worked fine in larger experimental animals have been miniaturized to make use of the genetic approaches uniquely available to mouse researchers. What makes the mouse so special?

The advantages that come to mind first are obvious: for biomedical researchers, mammals are highly preferred model systems, because the physiology of man and most mammals is very similar and the pathology of most diseases is reproduced well in animal models. The mouse stands out from the other mammals as being the smallest. Mouse husbandry is therefore easy and economical; its requirements for food are modest, and mice are excellent breeders with a short generation time, allowing around four successive generations each year. Upon completion of the human and mouse genome sequence, the usefulness of the mouse as a model animal for the human has been further reinforced. The number of genes in both species is around 30 000. With the exception of about 200 genes, which were shown unique to mouse or man, the vast majority of genes is conserved between mouse and man. These combined properties make the mouse an excellent animal model to study gene function and extrapolate the outcome to human disease.

Second, the mouse has advantages that may be less obvious, but which relate to the long tradition of mouse research starting in the early 1900s. Up until then, hobbyist mouse keepers had already been breeding mice for their coat color variation. By making use of these mice, it was possible to demonstrate that this variation followed the laws of Mendelian inheritance. Soon, it was also realized that there was a need for

genetically homogeneous mouse strains and the most widely used *inbred strains* such as BALB/c and C57BL6 originate from that period. An excellent historical overview on mouse genetics is provided by Dr. Silver.

At present, hundreds of inbred strains have been generated (see Sect. 6). This resource is the key to the success of mouse research. Inbred strains ensure that experiments can be carried out in the absence of inter-individual genetic variation, because mice within an inbred strain are all *genetically identical* to each other and *homozygous* throughout the genome. Moreover, each strain has one or more unique and, often, well-established properties that make the strain particularly useful for a specific area of research. For instance, some mouse strains appeared susceptible to cancer or cardiovascular diseases, while others are better suitable for behavioral research. The genetics of such complex diseases and biological processes is very difficult to study in man. This is mainly due to the fact that multiple genes are involved and that the genetic basis for susceptibility to, for instance, cancer will vary from person to person. In the mouse, it is possible to cross-breed a susceptible and a resistant strain. As a result, the susceptibility to cancer will vary in the progeny. Genetic variation in these mice is confined to the differences between the two parental strains, and every given locus will only have two different alleles. Hence, the genetics is confined to a maximum of two alleles per locus, which greatly simplifies genetic analysis. These properties allowed generating specific resources to study complex diseases, including *recombinant inbred* (RI) strains (http://jaxmice.jax.org/info/recombinbred.html). These RI strains combine a specific genetic contribution of one mouse strain with the genetic background of another strain. Along these lines, a very powerful resource has been developed recently, consisting of a panel of inbred mouse strains carrying only a single chromosome from the other strain (http://jaxmice.jax.org/library/notes/493c.html). These *chromosome substitution* strains are very similar to recombinant inbred strains, but offer a less costly strategy to move from a susceptibility locus toward the gene proper. The principles of mouse genetics are crucial when working with transgenic animals, because it is vital to control the genetic background in the experiments (see below): the interaction between the transgene and the genetic background will ultimately determine the phenotype; for example, changing the genetic background on a given genetic trait may prevent an early death of a particular mouse model.

While mouse genetics has a major impact on our understanding of the mechanisms underlying complex diseases, a constant influx of mice carrying single gene mutations greatly stimulates mouse research. These mutations have either occurred spontaneously in mouse breeding programs or originated from research on the gene toxicity of radiation and chemical compounds. The latter approach has been rejuvenated upon the availability of better technology to identify the causative mutations (see below). A good example of a spontaneous mutation is the leptin gene mutation that occurred in the C57BL6 mouse. The mutation was named *ob*, after obese, because the mice have no control over their feeding behavior and become overwhelmingly obese. Hence, the phenotype could be discovered easily. The subsequent discovery of the causative gene marks one of the most dramatic discoveries in this research field.

1.2 Molecular Mouse Genetics

Despite all the possibilities for the basic biomedical research offered by "traditional" mouse genetics and the discoveries of relevant mutations through random mutagenesis, the true breakthrough for the mouse as the animal of choice came with the accomplishment of (1) transgenesis through injecting DNA in fertilized oocytes and (2) gene targeting via homologous recombination in embryonic stem cells. The latter mice are commonly

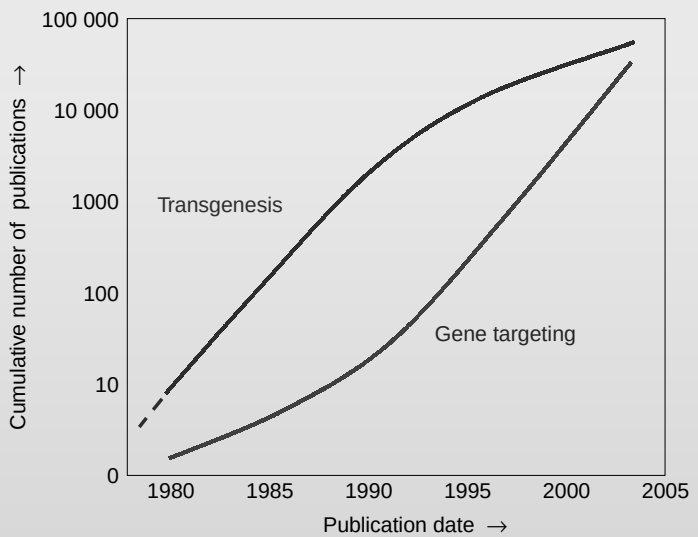

(a)

Milestones in mouse molecular genetics			
Transgenesis			
R.Jaenisch	Germ line transmission of retroviruses	Viral infection	1976
J.Gordon	Integration and germ line transmission of transgenes	Pronuclear injection	1981
R.Brinster F.Costantini E.Lacy	Expression of integrated transgene	Pronuclear injection	1982
R. Palmiter R.Brinster	First phenotype of a transgenic mouse (giant mouse; rat growth hormone)	Pronuclear injection	1982
Gene targeting			
M.Evans G.Martin	Establishment pluripotent ES cell cultures	*In vitro*	1981
M.R.Capecchi K.R.Thomas	Homologous recombination in mammalian cells Homologous recombination in ES cells	*In vitro*	1982 1987
D.W.Melton O.Smithies	Germ line transmission of targeted mutations in ES cells	*In vivo*	1989

(b)

known as *knockout* mouse models. This combination of technologies offers the possibility to design models with "gain of function" and "loss of function," which is an ideal situation for a (mouse) geneticist (Fig. 1a).

Notably, these technological breakthroughs were facilitated by the availability of inbred mouse strains. Transgenesis works most efficient when the donor eggs are produced by first generation (F1) hybrids between C57Bl6 and BALB/c. In the case of the gene-targeting approach using embryonic stem cells, it is critical to make use of the appropriate strain combinations when injecting the cells into blastocysts and propagating the embryos in foster strains. This dependence on the genetic background provides the ability to choose the optimal strain and may be one of the reasons why, mice are, as of yet, the only mammalian species suitable for gene targeting (see Sect. 3). The goal of altering the gene function through transgenesis is to generate animal models in which the role(s) of the gene of interest either in normal (e.g. development) or abnormal biological processes (e.g. disease) is revealed. Other important applications in fundamental science include the study of mammalian gene expression regulation. Currently, genetic modification is carried out in a wide range of invertebrate and vertebrate species, including mammals such as mice, rats, rabbits, and livestock such as sheep, pigs, cattle, and recently even primates. The choice of the models is primarily determined by its objective (e.g. improvement of economical value or development of pharmaceutical proteins). Although fundamental biological processes can be well studied in relatively simple organisms such as the nematode *Caenorhabditis elegans*, or the fruitfly *Drosophila melanogaster*, for experimental biomedical research the laboratory mouse is by far the most used animal model. Below, we will briefly outline the emergence of the two technologies and sketch the initial experimental strategies that caused great excitement within the scientific community.

1.2.1 Brief Overview of Developments in Transgenesis

By classical definition, a *transgenic* organism carries extra, often foreign (i.e. from a different organism) DNA in its genome, called a *transgene*. Transgenesis literally crosses (*trans*) species barriers. One of the most obvious conceptual characteristics is that a transgene integrates at random in the recipient host genome, whereas gene targeting, by homologous recombination, selectively alters an existent locus. A number of milestone achievements,

Fig. 1 Genetically modified mouse models in biomedical science. (a) A rough estimation of published scientific output using genetically modified mouse models. The blue line represents reports on transgenic models and the red line on gene targeting. Note: with the advent of conditional gene targeting, which utilizes transgenesis in conjunction with targeted mutagenesis, a strict separation of transgenesis and gene targeting into the twentieth century is difficult. (b) Milestones in mouse molecular genetics. The first report on germ-line transmission of integrated retroviruses (via infection) dates back to the mid-1970s. Subsequently, pronuclear microinjection of fertilized oocytes and subsequent proof of transgene integration, expression, and germ-line transmission, were pioneered in different laboratories. Similarly, breakthroughs in the use of embryonic stem (ES) cell technology are listed. Many other excellent laboratories have since contributed to refinement of technologies and strategies; these are listed throughout this chapter (see color plate p. xxvi).

which were pivotal for the development of modern-day mouse molecular genetics, are summarized in Fig. 1(b).

In the mid 1970, robust technology to clone genes and manipulate genes was still about five years away. Subsequently, as the structure of eukaryotic genes was revealed and recombinant DNA technology developed, gene constructs could be generated comprising an eukaryotic promoter fused to the full-length coding sequence of the gene of interest (consisting of a DNA copy (cDNA) of the mRNA molecule). Generating mice with a strong promoter driving an extra copy of the gene resulted in mouse models with high-level expression of the gene of interest. Interesting phenotypes did occur when the gene was markedly overexpressed. In addition, however, genes would become expressed in the wrong tissue or at the wrong place. This *ectopic* expression was also a powerful way to obtain phenotypic information. Initially, researchers were often confronted with a high failure rate: gene constructs would be integrated in high-copy head-to-tail arrays (see Sect. 2), but would not show a corresponding increase in expression level, or transgene expression would diminish in consecutive generations. A much-improved understanding of eukaryotic gene regulation in the mid-1980s provided the insight, leading to improved construct design. For instance, DNA elements located tens of kilobases outside of the β-globin gene were shown to control high-level tissue-specific gene expression through interaction with DNA sequences immediately 5′-prime of the gene. Also, sequence elements were discovered that isolate a specific gene from influences of flanking DNA. In addition, evidence was obtained for a role of splicing in mRNA stability. These advances led to the use of much larger constructs harboring the complete intron–exon structure of the genes, and including large up- and downstream DNA segments. Such fragments were obtained through cloning in cosmids, bacterial artificial chromosomes and yeast artificial chromosomes, allowing handling DNA fragments in the range from 40 up to 1000 kb.

1.2.2 Brief Overview of Developments in Gene Targeting

The development of gene targeting by homologous recombination in embryonic stem cells (ES) can be attributed to two crucial technological advances that ultimately compounded. One line of research involves ES research. Initially, ES cells were used to study cellular differentiation and cancer, because when injected subcutaneously into nude mice, teratomas would develop. Figure 1(b) itemizes a number of crucial advances in ES cell technology that helped develop ES cell technology to its current status. Improved culturing conditions ensured sustenance of pluripotence, a property of ES cells crucial for their application as vehicles to introduce germ-line mutations. Soon after the first mouse was made that showed germ-line transmission of genes introduced through genetic modification of ES cells *in vitro*, evidence was obtained that an endogenous locus could be exchanged for a mutated sequence by homologous recombination *in vitro* (see Sect. 3). These strategies targeted the murine *HPRT* gene and were based on the fact that HPRT is an excellent selectable marker expressed in the genome on the X chromosome of male cells. Spurred by success, the question arose whether gene targeting would also be feasible for nonselectable autosomal genes; this was proven possible shortly after.

Sections 2 and 3 will elaborate on the conventional transgenesis and gene-targeting technology. Noteworthy, with the introduction of novel concepts in molecular genetics, our ability to manipulate gene expression in the mouse currently knows few limitations. Sections 4, 5, and 7 of this chapter will elaborate on some of the more recent technological advances such as binary systems (see Sect. 4), random mutagenesis (see Sect. 5), and RNA interference (see Sect. 7).

2 Transgenesis; Basics and Technology

2.1 Transgene Design

Shortly after its introduction in the early 1980s, the main scientific applications of transgenesis in the mouse could be roughly divided in two areas: (1) studies of gene promoter activity (and/or other regulatory elements) and (2) studies of the effect(s) of gene overexpression (or a mutated form thereof; Fig. 2a). Gene activity is, simply put, controlled by a *promoter*. A promoter is a stretch of DNA sequence, which usually precedes the coding part of a gene. A promoters' task is to regulate gene expression: it controls the "when (at which time point) and where" (in what cells, tissues) of gene activity. Eukaryotic genes also carry a termination element. In order to function properly in a eukaryotic genome, a transgene minimally meets these requirements: it carries a promoter, a coding part, and a stop signal. The easiest way to read out promoter activity in cell or a developing animal is by placing a so-called *reporter gene* downstream of it. Reporter genes encode proteins, often found in lower organisms such as yeast, fireflies, or jellyfish, which reveal their presence by a microscopically or macroscopically clearly visible mark (Fig. 2a, b). This approach is standardly used to examine *spatio-temporal* gene expression patterns and/or tissue specificity of native promoter sequences *in vivo* (Fig. 2b).

By far, transgenesis is used to achieve and/or study the effects of overexpression (higher gene activity than normal) or *ectopic* expression (gene activity in different cell types than normal) of a gene of interest, or of a mutated form thereof. The common denominator of these studies is the analysis of the resulting altered phenotype. This can be abnormal development, altered physiology or behavior, development of disease, and so on. The choice of regulatory sequences (see below) determines whether the expression of a transgene follows the expression pattern of its endogenous counterpart or is limited to distinct cell types or particular developmental stages.

2.1.1 Origin and Structure of the Transgene

The origin of a transgene, or elements therein, may range from prokaryotes or from lower (e.g. reporter genes) to higher eukaryotes, like mice or men. In experimental biomedical research, human transgenes offer several advantages. Importantly, many known genetic disorders in humans have been extensively characterized at the molecular genetic level: mutations or deletions are charted and such "diseased alleles" are readily available. In addition, most genes and proteins are well conserved between mouse and man. This, of course, is important when the biological activity of a gene product depends on proper interaction with other cellular proteins. Finally, the sequence

divergence between mouse (transgene recipient) and human (transgene donor) of a given gene can be used to determine whether a mouse is transgenic or not, what its copy number is (number of transgene integrations), and whether a transgene is actively expressed, since it allows for discrimination between transgene and the endogenous gene. Figure 3 gives an example of such an analysis.

Most eukaryotic genes display a typical intron/exon structure, with only exon sequences represented in mRNA. Transgenes, in general, are more reliably expressed if the intron/exon boundaries are preserved in a transgene construct. Solely cDNA-based expression vectors (i.e. exons sequences only) frequently show low expression levels and are often silenced. Since many integrating DNA or RNA viruses do not carry the eukaryotic intron/exon structures, an expressed genetic sequence without such boundaries may be recognized as foreign and potentially harmful by the eukaryotic host cell and will be inactivated (silenced) by cellular defense mechanisms. The native intron/exon structure of a gene need not be preserved in its entirety: indeed, inclusion of only one intron in a transgene has been shown

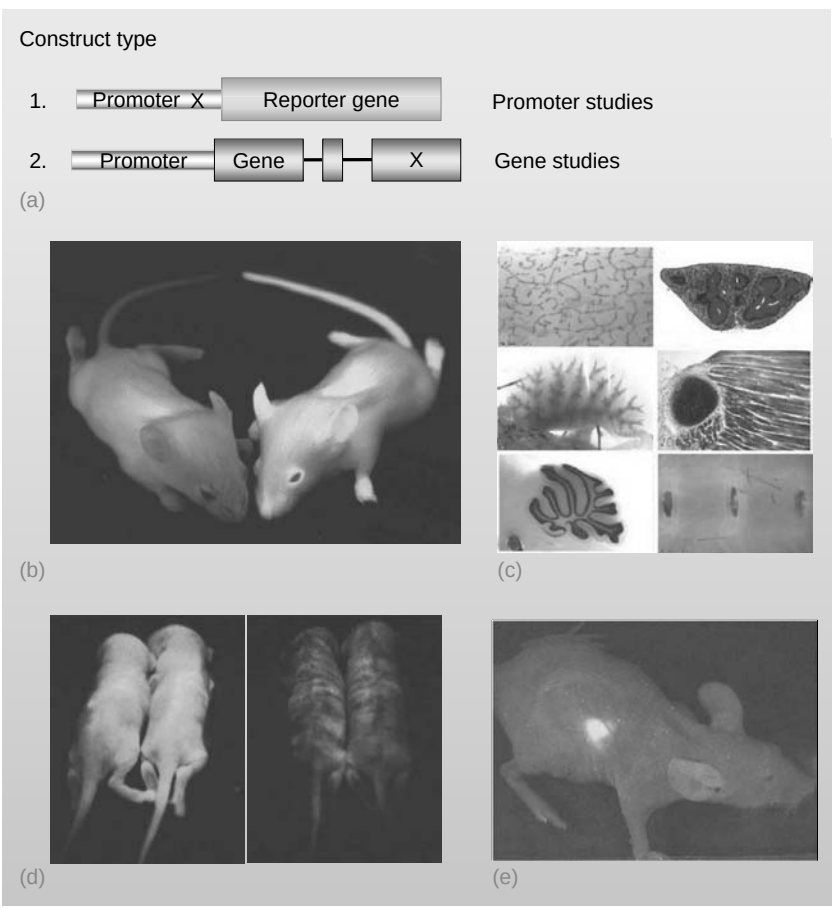

to augment transgene expression. Many successful transgene constructs include both genomic and cDNA sequences derived form the same gene (Fig. 4a).

2.1.2 Regulatory Elements

Any transgene comprises a number of essential elements that control gene expression (Fig. 4b). The most basic and essential elements are the promoter and termination signals for transcription (RNA synthesis). A promoter controls gene expression by providing binding sites for proteins (RNA polymerase transcriptional machinery, cell type–specific transcription factors) that regulate gene transcription (i.e. activation). The real-life situation is often much more complicated than this; many more regulatory elements exist that all play a role in establishing stable and faithful gene expression patterns. Among these are enhancers, which typically act in an orientation-independent manner, Matrix attachment regions (MARs), scaffold attachment regions (SARs), Locus control regions (LCR), chromosomal insulators, and antirepressor elements. Some of these elements are involved in subnuclear localization of genes to areas of active transcription and/or insulate gene expression from influences of surrounding chromatin. Some of these autologous (i.e. endogenous) regulatory elements may actually be many thousands of base pairs removed from the coding sequences of a gene. Indeed, if the purpose of a study is to examine faithful gene expression patterns, for instance throughout embryonic development, a transgene should include such endogenous elements. The exact location of such elements within a locus is often not known and hence there is an obvious advantage in using large DNA segments as transgenes. The development of molecular vectors, which can harbor large pieces of DNA, such as so-called artificial

Fig. 2 Transgenesis; concepts. (a) Original scientific applications of transgenesis in the mouse were roughly divided in two areas: (1) studies of gene promoter activity (and/or other regulatory elements) and (2) studies of the effect(s) of overexpression (or a mutated form) of gene X. Promoter studies typically make use of *reporter* genes (green box) that reveal their presence through specific catalytic activity. Often-used reporter genes include the green fluorescent protein jelly fish (b, d), bacterial beta-galactosidase (c), (also see Fig. 3), and fire fly luciferase (d). (b) Transgene expression is controlled by a keratinocyte-specific promoter (source image: http://www.mshri.on.ca/nagy/cre.htm; Maatman, R., Gertsenstein, M., de Meijer, E., Nagy, A. et al. (2003) Aggregation of embryos and embryonic stem cells, *Methods Mol. Biol.* **209**, 201–230). (c) Differential expression of beta-galactosidase (upper left to lower right): Lymphatic vessels (intestinal surface), marginal zones between red and white pulp (spleen), bronchial epithelium (lungs), Large skeletal muscles (limb), granule cell layer (cerebellum), Intervertebral discs (source images: Valenzuela, D.M., Murphy, A.J., Frendewey, D., Gale, N.W. et al. (2003) High-throughput engineering of the mouse genome coupled with high-resolution expression analysis, *Nat. Biotechnol.* **21**, 652–659). (d) Fluorescent images of an X-linked enhanced green fluorescent protein (EGFP) transgene; transgenic male offspring (left) of a mating between an X-linked EGFP transgenic female and a nontransgenic male have ubiquitous green fluorescence in the skin owing to the presence of a single active X chromosome. Transgenic female pups (right) have mosaic (tortoiseshell) green fluorescence in the skin owing to random inactivation of one X chromosome (source image: Hadjantonakis, A.K., Dickinson, M.E., Fraser, S.E., Papaioannou, V.E. (2003) Technicolour transgenics: imaging tools for functional genomics in the mouse, *Nat. Rev. Genet.* **4**, 613–625). (e) An external image of adenovirus-encoded GFP gene expression acquired from a nude mouse in the light box 72 h after portal vein injection (see color plate p. xxviii).

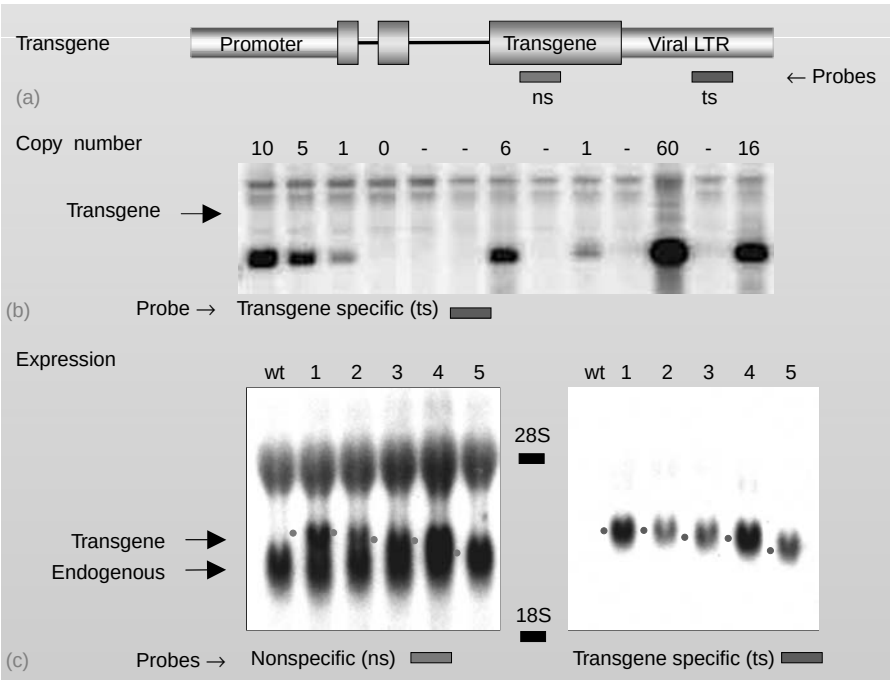

Fig. 3 Transgene construction and detection. (a) Schematic representation of a hypothetical transgene. The transgene is derived from mouse genomic sequences (brown). Transgene expression is under control of a general promoter and terminates in a retroviral long terminal repeat (LTR) element. Probes used for detection of transgene DNA or RNA are depicted as a brown box (nonspecific (ns) probe, which detects both the endogenous gene and the transgene) or a blue box (a transgene-specific (ts) probe, which only detects the transgene). (b) Copy number determination by Southern blot analysis with a ts probe. Bold print above image represent DNA samples with known copy numbers, resp. 10, 5, 1, and no transgene copies inserted; the remaining lanes show four founders with varying copy numbers (image adapted from:(http://www.healthsystem.virginia.edu/internet/transgenic-mouse/southerndesign.cfm)) (c) Detection of transgene expression by Northern blot hybridization clearly demonstrates the need for transgene specific probes (right panel). Left panel: expression signal of endogenous gene masks transgene expression (see color plate p. xxix).

chromosomes that can be propagated in yeast (YACs), phages P1 (bacterial viruses; PACs), or bacteria (BACs) has been imperative for the cloning and expression of such transgenic constructs. Exceedingly large constructs (>500 000 bp) are first transferred into embryonic stem cells (see Sect. 3), which are then used to generate transgenic mice.

Regulatory elements need not necessarily be autologous in nature, they may be *heterologous*. For instance, when overexpression or ectopic expression is required, general type heterologous promoters (e.g. viral promoters) and/or enhancers are widely used. The use of heterologous and autologous regulatory elements is often combined. The first published transgenic mouse model displaying a specific phenotype made use of the general-type metallothionein promoter (pMT). The MT promoter is inducible *in vitro* and *in*

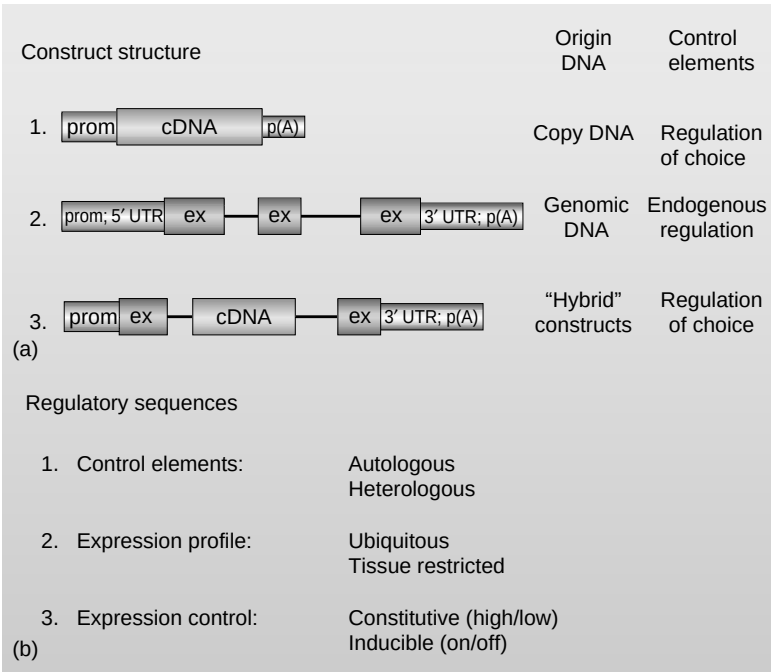

Fig. 4 Transgene structure and regulation. (a) Schematic overview of basic transgenic construct design. For all constructs (1,2,3), eukaryotic coding sequence is the starting material. Regulatory sequences are cloned separately into the construct, which is entirely cDNA based (1). Endogenous regulatory sequences may be included when genomic DNA is used (2). A transgene can be tailored to specific requirements. The hybrid genomic/cDNA construct depicted (3) combines intro-exon boundary inclusion with facile cloning options: the use of cDNA may reduce the size of a vector considerably. prom: promoter sequences, ex: exon, cDNA: copy DNA, 3′ UTR: 3-prime untranslated region, p(A): polyadenylation signal. (b) Regulatory sequences in transgene design. Depending on the specific purpose of the model, many different systems are available to control transgene expression. As indicated, regulatory element can be autologous (i.e. derived from an endogenous locus) or heterologous in nature (1). The required transgene expression profile may need to be ubiquitous, or cell type specific. This determines the choice of promoter (2). Finally, the model may require controlled expression; instead of constitutive expression, inducible expression would be preferred (3). (Figures adapted from: Voncken, J.W. (2003b) Transgene design, *Methods Mol. Biol.* **209**, 51–67).

vivo with glucocorticoids, heavy metals, or bacterial endotoxin (LPS). Heterologous promoters and other regulatory elements are applied widely, and as indicated before, their incorporation into constructs is determined primarily by the aim of the animal model (see Fig. 2). Not only general-type promoters, like the ones driving housekeeping genes (e.g. *phosphoglycerate kinase PGK*) or other genes that are widely expressed throughout eukaryotic cell types (e.g. *histones, ß-actin*), but also

viral promoters are often used to ensure high and ubiquitous transgene expression. LCRs are included in transgenic constructs to both enhance transgene expression and to achieve position-independent and copy number–dependent expression of a transgene, often with cell lineage–specific enhancer activity. Finally, promoter choice is often dictated by whether expression of a transgene is deleterious to the mouse or not. For example, many oncogenes play an important role in embryonic development, interfering with their normal expression (pattern) results in embryonic lethality. To bypass this problem, either tissue-specific or inducible promoters (or combinations of such regulatory systems are used (see Sect. 4)).

Taken together, the origin of DNA, structure and choice of regulatory elements for a transgene are largely determined by the scientific aim of the experimental model itself. For biomedical studies employing the laboratory mouse as a model system, the use of human transgenes is often preferred.

2.2 Pronuclear Injection

Transgenes can be introduced into living cells in many different ways. Among the most standard procedures *in vitro* ("in the test tube") are transfection methods either using calcium phosphate/DNA precipitates or liposomes (small DNA-containing vesicles surrounded by a lipid membrane), which are taken up by cells, electroporation (electroshock), viral vectors (e.g. adenovirus, retrovirus, which carry the transgene), particle bombardment, and microinjection. Some of these methods, however, cannot be used in combination with mouse oocytes. The very first proof-of-principle report on germ-line transmission of a foreign DNA sequence in the mouse was achieved through retroviral transduction. As indicated above, viruses of this class are often inactivated in eukaryotic cells. Although recent technological developments in the field of transgenesis again make use of retroviral constructs (see Sect. 7), pronuclear injection has been the golden standard for generating transgenic animals over the last two decades.

The principle of pronuclear injection is simple: a sterile DNA solution is prepared, which contains a very low concentration of the transgene (nanograms (10e-12 kg)/microliter (10e-15 l)). A very fine glass needle containing the transgene DNA solution is inserted into one of the pronuclei and DNA solution is expelled into the nucleoplasma (Fig. 5a, 6a). The amount of DNA injected is extremely low (femtograms (10e-18 kg)/picoliters (10e-12 l)). Injected transgene DNA in the nucleus tends to integrate at random into the genome. Often, this occurs in head-to-tail tandem arrays called *concatamers*. (Fig. 5b). The repetitive nature of such concatamers may lead to transgene inactivation, as does the random character of the insertion: it often causes transgenes to land in inactive heterochromatin, which constitutes most of a eukaryote cells' DNA. The use of insulator elements (see Sect. 2.1) or targeted transgenesis (see Sect. 4) provides solutions to these problems.

2.3 Biotechnological Procedures

Figure 6 gives an overview of the biotechnological procedures involved in the production of transgenic mice. Because generating transgenic mice by pronuclear injection is a rather inefficient process, a relatively large number of fertilized oocytes

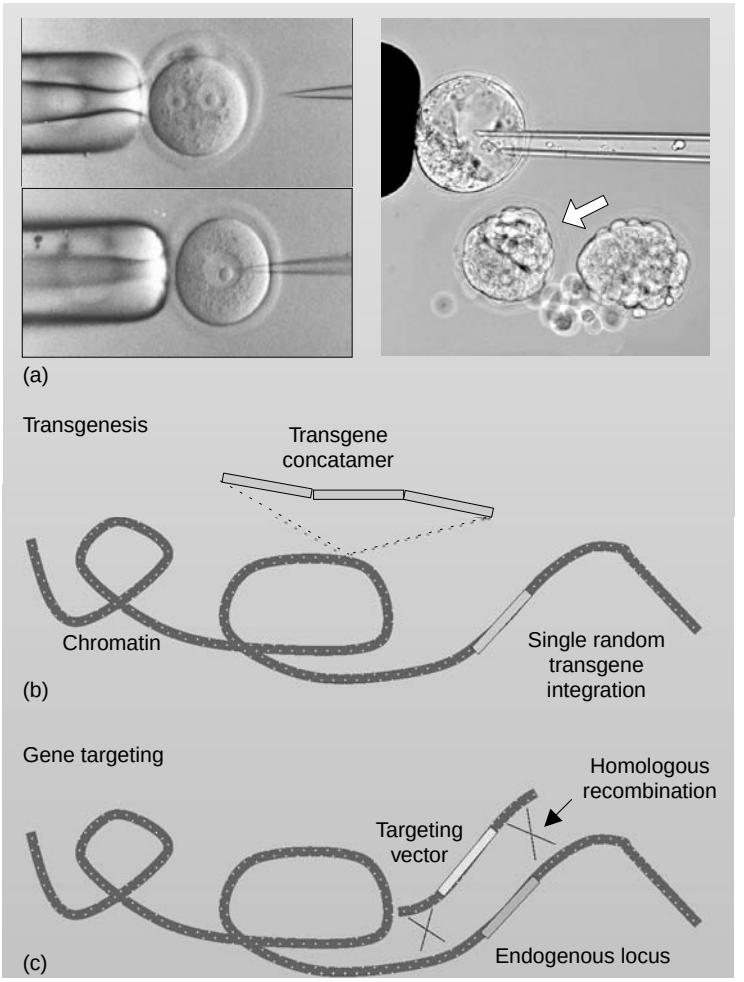

Fig. 5 Microinjection. (a) Pronuclear injection of DNA into a fertilized one-cell-stage embryo. In the early stages following fertilization, two pronuclei are visible in the zygote: one from the oocyte and one from the sperm cell; these will eventually fuse to form a diploid nucleus. Both prefusion pronuclei are clearly visible in the cytoplasm (the male pronucleus is mostly used for injection, since it is the largest) (left panel; source image: Hogan, B., Beddington, R., Costantini, F., Lacey, E. (1994) *Manipulating the Mouse Embryo. A Laboratory Manual*, 2nd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York). Microinjection of genetically modified ES cells into blastocysts; arrow indicates collapsed, injected blastocysts; ES cells are clearly visible in the blastocoel (right panel). (b) Schematic representation of random integration of transgenesis in the mouse genome: transgene either integrated single or as concatameric head-to-tail arrays. (c) Gene targeting is homology-driven: a targeting vector will exchange with an endogenous locus by homologous recombination.

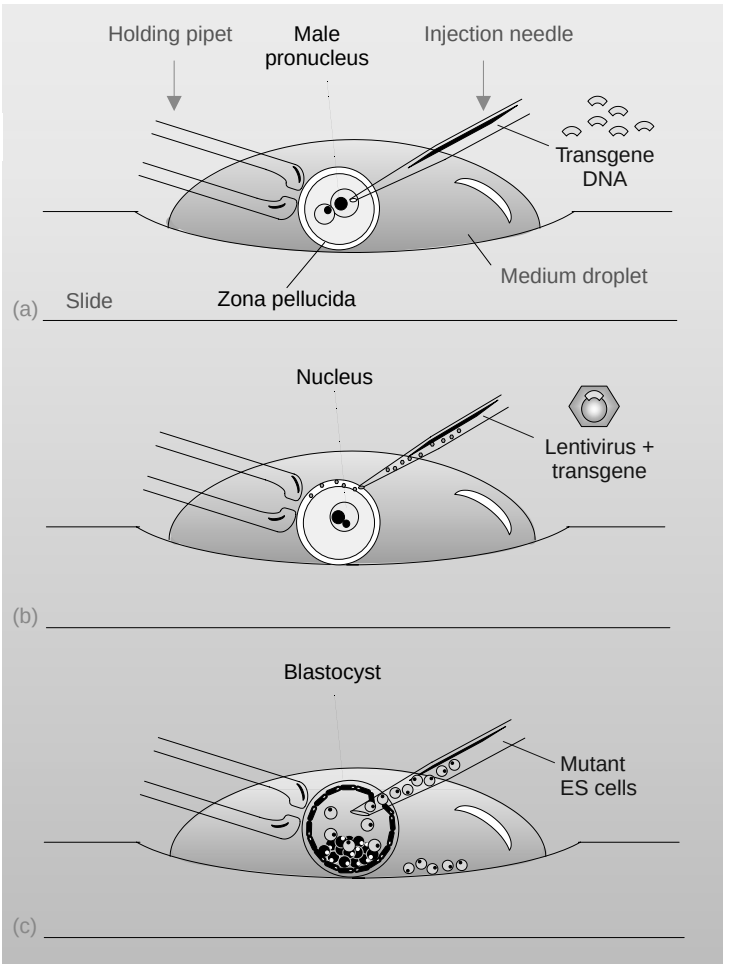

Fig. 6 Schematic representation of a number of techniques to generate genetically modified mouse models. (a) Pronuclear injection of transgene DNA. (b) Injection of infectious viral particle in the perivitellin space; the advantages of using lentiviral shuttle vector are significant: it obviates potentially harmful procedures otherwise used to introduce transgenes into a cell (pronuclear microinjection, electroporation) and offers some degree of control over transgene copy numbers. Lentiviral vectors are also used to generate transgenic lines through infection of ES cells. One of the few restrictions on recombinant lentiviral vector use is the size limitation on the transgene inserts. (c) Microinjection of pluripotent, genetically modified ES cells into blastocysts.

needs to be injected to obtain the so-called founder mice: transgenic founders developed *in utero* from successfully injected zygotes that carry an integrated transgene in their genome. A typical transgenic study uses at least three expressing transgenic founders to establish independent transgenic mouse lines from.

Convenient numbers of fertilized eggs are obtained by superovulation, much the same as way as in humans for *in vitro* fertilization purposes: exogenous hormones (follicle-stimulating hormone and luteinizing hormone analogs) are injected at a 48-h interval to achieve multiple ovulations per donor female, which are subsequently mated to fertile males. Simultaneously, recipient females, the so-called foster females, are endocrinologically primed for pregnancy by copulation to vasectomized males (see Fig. 7). Microsurgery is carried out under aseptic conditions and general anesthesia. A number of different mouse strains are useful for biotechnological procedures, many of which have their own advantages and specific traits (see Sect. 1). Detailed descriptions of the technology, useful strains, and husbandry are to be found in a number of media.

3 Gene Targeting; Basics and Technology

First established in the late 1980s and early 1990s of the last century, gene targeting is now a well-established technology for generating genetically modified mouse models. Gene targeting allows for germ-line manipulation at predetermined genomic loci by a process called *homologous recombination*: part of an endogenous allele is replaced or mutated by vector sequences that carry flanking sequences homologous to the endogenous gene (Sect. 3.1). It is in this aspect where gene targeting fundamentally differs from conventional transgenic technology: the possibility to change existing genes offers a very important and powerful extension of the molecular genetic tools to create experimental animal models in biomedical research. Whereas transgenesis asks for genes and their phenotypes to be dominant in nature (e.g. expression of (mutated) oncogenes), gene targeting now permits study of recessive gene function (e.g. tumor suppressor genes). The concept of gene targeting is simple: eradication of a gene may reveal its function by a resultant phenotype. Current applications range from conventional *null mutation* of a given gene, the so-called gene knockouts, to replacement of endogenous alleles with mutated alleles and even inducible (*conditional;* see Sect. 4.3) genetic modifications.

3.1 Targeting Vector Design

Generally speaking, two types of vectors exist by which genes can be mutated.

1. Replacement vectors
2. Random insertion vectors

All classical *replacement type vectors* have in common that they carry stretches of DNA, which are highly homologous or identical to the gene, which is to be targeted. The mechanism by which gene replacement is accomplished, *homologous recombination*, uses these identical or very similar sequences of DNA: homologous recombination is a sequence homology-initiated and driven process (Fig. 5c, 8a). It appears to be a highly selective process, which only takes place between relatively large (several kilobase pairs) DNA segments. In fact, it is important that the constructs used for homologous recombination are being derived from the exact same genetic background (*isogenic*) as the cell line used for gene targeting. This requirement implies that a sequence divergence of around 1 in 100 bp will cause a marked drop in

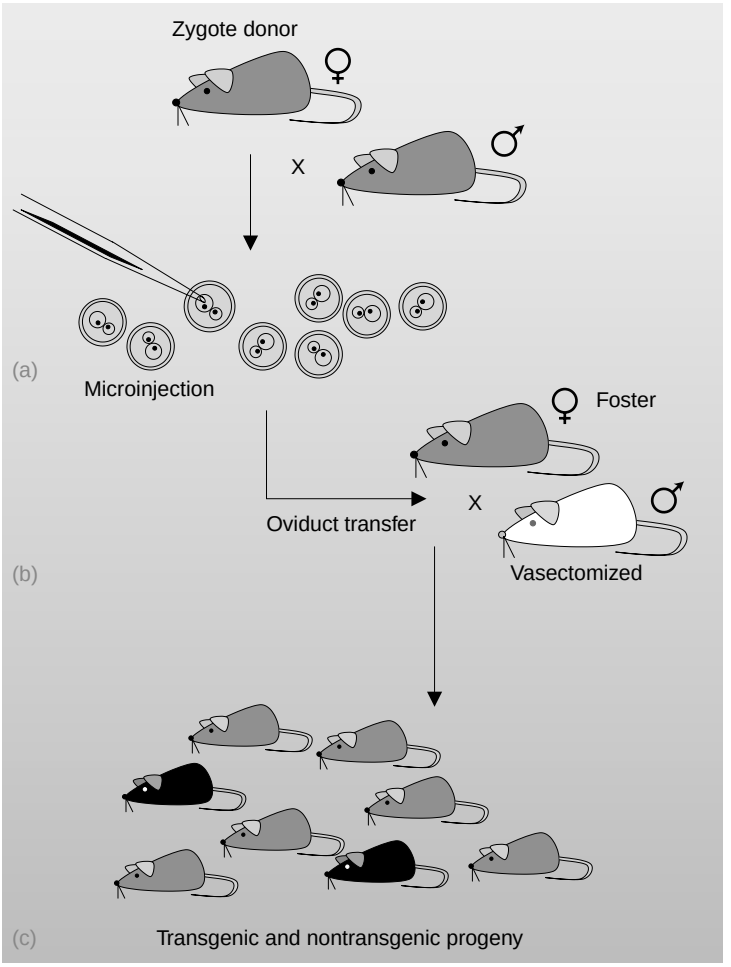

Fig. 7 Biotechnological procedures in conventional transgenesis. (a) Hybrid F1 animals are used for superovulation and mating to generate sufficient zygotes for injection. (b) Upon microinjection, the embryos are transferred into the oviduct of a pseudopregnant female; *pseudo-pregnancy*, is achieved by mating to a vasectomized male and exists for a maximum of 3 days; the surgical transfer of microinjected one- or two-cell-stage embryos directly into the fallopian tube at 0.5 day *post coitum* ensures sustenance of the pregnant state. (c) Offspring will manifest coat colors of both strains (e.g. C57Bl6 and CBA) used to generate the F1 hybrids (e.g. B6CBA/F1). Roughly, 10% of the offspring will carry an integrated transgene, and will be further examined for transgene expression and transmission. Such an animal is referred to as a *transgenic founder*. In addition, these media provide comprehensive information on historical and genetic backgrounds of in- and outbred strains on mouse embryology and dissection of specific developmental stages.

the frequency of homologous recombination. Conventional gene-targeting constructs mostly harbor a selection marker (neomycin, hygromycin, or puromycin resistance gene), which serves a number of purposes: it replaces endogenous gene sequences, thereby interrupting the gene and rendering it nonfunctional (Fig. 8a). In addition, the marker is used to positively select for embryonal stem cells that have integrated the targeting vector. A second feature often included in a replacement

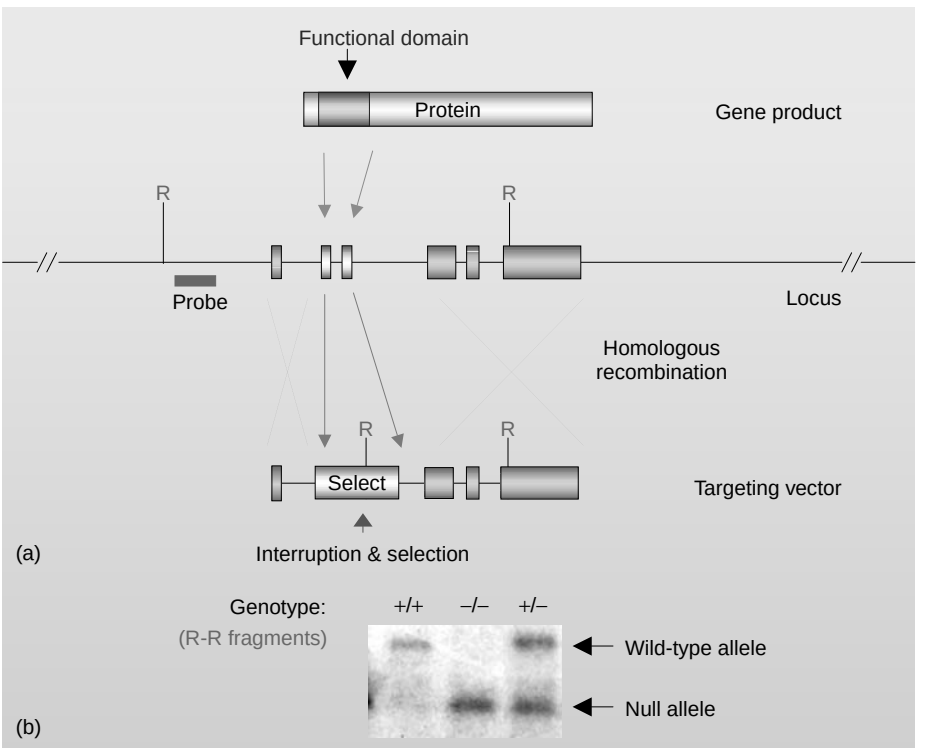

Fig. 8 Basic design of a conventional targeting vector. (a) The targeting vector duplicates part of the locus to be mutated (homologous sequences), but is interrupted by a positive selection marker (*select*), for instance, an antibiotic resistance gene. Many conventional targeting vectors also contain a negative selection marker. The herpes simplex virus Thymidine Kinase gene is often included: the TK gene product converts a harmless compound, for instance, Ganciclovir, into a cytotoxic substance. The positioning of the TK cassette at one of the ends of a targeting vector results in exclusion of TK from the genome upon homologous recombination, whereas TK is cointegrated in case of random integration of a targeting vector. (b) The vector is designed such that by means of an "external probe" (blue-grey box: *probe*; from within the to be targeted locus, but outside the homology regions used in the targeting construct) and strategically chosen restriction endonuclease sites (R) a shift is detected upon homologous recombination. The figure shows the three possible genotypes; two wild-type (+/+) alleles, two knockout (−/−) alleles (homozygous knockout), and a combination of both in the heterozygous (+/−) condition (see also Fig. 10). Mutagenesis vectors may be introduced into ES cells via several means; the most commonly used method is electroporation (electroshock) by which the cellular membrane is damaged, allowing uptake of DNA in cells (see color plate p. xxvii).

vector is a negative selection marker, for example, the *Herpes simplex virus* thymidine kinase (HSV-TK). Upon positive (negative) selection, a number of separately isolated ES cell clones are "genotyped" to ensure proper recombination on both ends of the targeting vector. Typically, these cells carry a monoallelic mutation, that is, only one of two alleles in the ES clone is replaced by the targeting vector (Fig. 8b).

Random insertion type vectors insert at will in the genome and, therefore, do not rely on sequence homology with target genes (see Sect. 5.1). Integration of a vector in a locus may, however, cause inactivation of a gene. The best-known application of insertion type vectors is *gene trapping* (see Sect. 4 and 5). Gene trap vectors are constructed such that they either trap promoter activity or trap a polyadenylation signal from an endogenous locus (Fig. 15). In this manner, not only actively transcribed but also inactive genes can be trapped (with promoter-trap or poly(A)-trap vector respectively). Variations on this concept are described in Sect. 5.

3.2 Embryonic Stem (ES) Cells

In contrast to classical transgenesis, where mutagenesis is done directly in the embryo, gene targeting is done in *embryonic stem cells* (ES cells) (Fig. 9a). These are very early stem cells isolated from preimplantation embryos (*blastocysts*). ES cells are *pluripotent*, that is, they are fully undifferentiated and, in essence, have the capacity to participate in mouse embryogenesis and become any cell type in the animal. The first steps in a gene targeting experiment are carried out *in vitro*. ES cell culturing conditions are aimed at maintaining their pluripotency. ES cells are either cultured on a layer of supporting feeder cells (mouse embryo fibroblasts (MEFs)) in the presence of factors that block differentiation of stem cells in general (like *Leukemia Inhibitory Factor*,) or are "feeder-independent." It is their pluripotency that is ultimately called upon: when generating targeted mutations in the mouse, male mutant ES cells are reintroduced into preimplantation embryos and made to participate in embryogenesis and spermatogenesis in particular. If the latter occurs, the targeted mutation can be stably transmitted via the germline to offspring and bred to homozygocity.

3.3 Genesis of Mosaic Embryos

ES cells may be used in a number of ways to generate genetically mosaic embryos. The most commonly used procedure is *blastocyst injection*: preimplantation blastocysts are used to introduce ES cell into (Fig. 5a). The donor mouse line (i.e. of which ES cells are derived) and the recipient line (i.e. of which blastocysts are procured) differ in coat color genetics. Typically, an inbred strain such as *C57blk* (black coat color) is used for blastocyst production to inject *129Sv* or a *129Ola* (sand-colored coat), some of the most commonly used ES cell lines, into. The procedure described above is schematically represented in Fig. 6c.

Alternatively, ES cells can be fused to morula stage embryos in a procedure termed *morula aggregation*. Embryos in the morula stage typically comprise 8 to 16 blastomeres (individual embryonic cells) surrounded by a protective layer called the *zona pellucida*. In order to achieve aggregation, this zone is first removed and the embryos together with a small number of ES cells are cocultured in a small depression, in which gravity forces them to

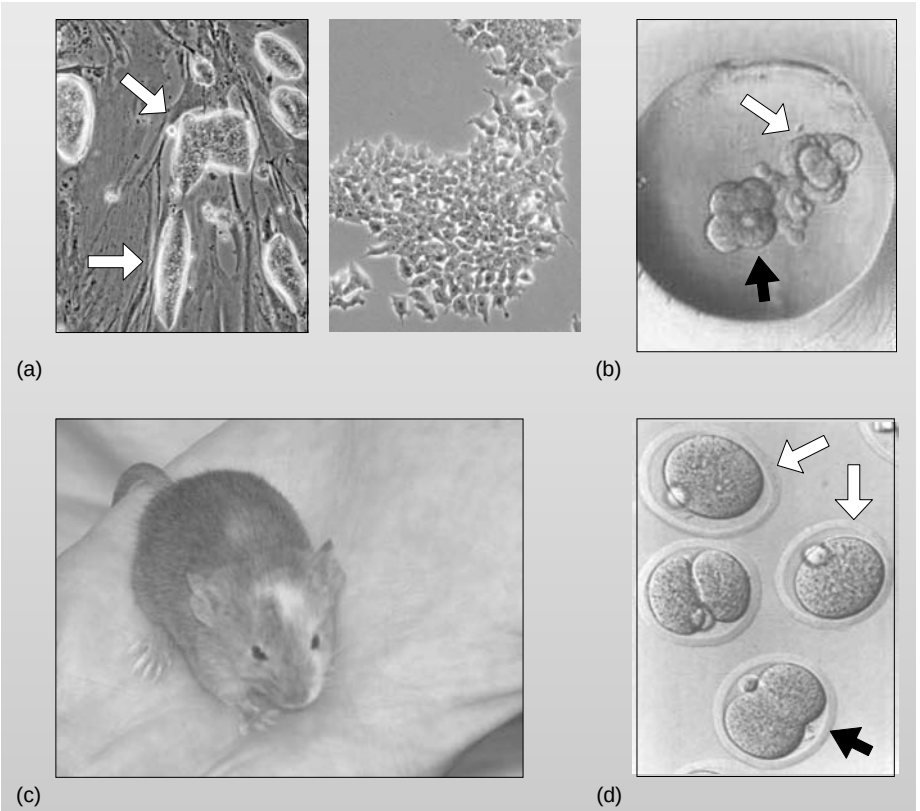

Fig. 9 Embryonic stem cells, embryos, and chimeric mice. (a) Left panel: ES cells grown on mouse embryonic fibroblasts (MEFs); white arrows point at colonies of tightly coherent ES cells; right panel: feeder-free ES cells. (b) Aggregation of 8-cell-stage morulas (black arrow) with genetically modified ES cells (white arrow). (c) Chimeric mice beautifully displaying the coat color mosaicism resulting from contribution of ES cells of two different genetic backgrounds to the developing embryo the mix of, in this case, C57Bl (recipient blastocyst – black) and 129Ola (donor ES cells – sand) becomes manifest as Agouti, patches of purely 129Ola-derived tissue is sand colored (e.g. over left eye). (d) Electrofusion of two-cell-stage embryos (middle left) results in fusion and formation of one-cell-stage tetraploid embryos (white arrows). Tetraploidy does not hamper placenta formation, but does not permit normal embryogenesis. Tetraploid blastocysts injected with diploid ES cells will support fully ES cell–derived embryonic development (source image: (http://www.mshri.on.ca/nagy/cre.htm)) (see color plate p. xxxi).

be in close contact with each other (Fig. 9b; http://www.mshri.on.ca/nagy/cre.htm).

Both methods yield the typical mosaic coat (e.g. a mixture of different coat colors, like black and sand). The mice that display the coat color mosaicism are referred to as *chimeric mice* (Fig. 9c). Chimeric males are bred to wild-type mice to generate a mouse that is heterozygous for the mutant allele in all its cells; the subsequent mating of two

heterozygous animals allows the mutation to be bred into a homozygous state.

3.4 Biotechnological Procedures

In contrast to transgenic technology, preimplantation embryos are isolated at the blastocyst stage. Microinjected embryos are transferred directly into the uterus of a 2.5-day pseudopregnant female, where they develop into chimeric embryos (Fig. 9b, 10). A typical targeting experiment uses at least two independent mutant embryonic stem cell lines for microinjection to verify and confirm the phenotype

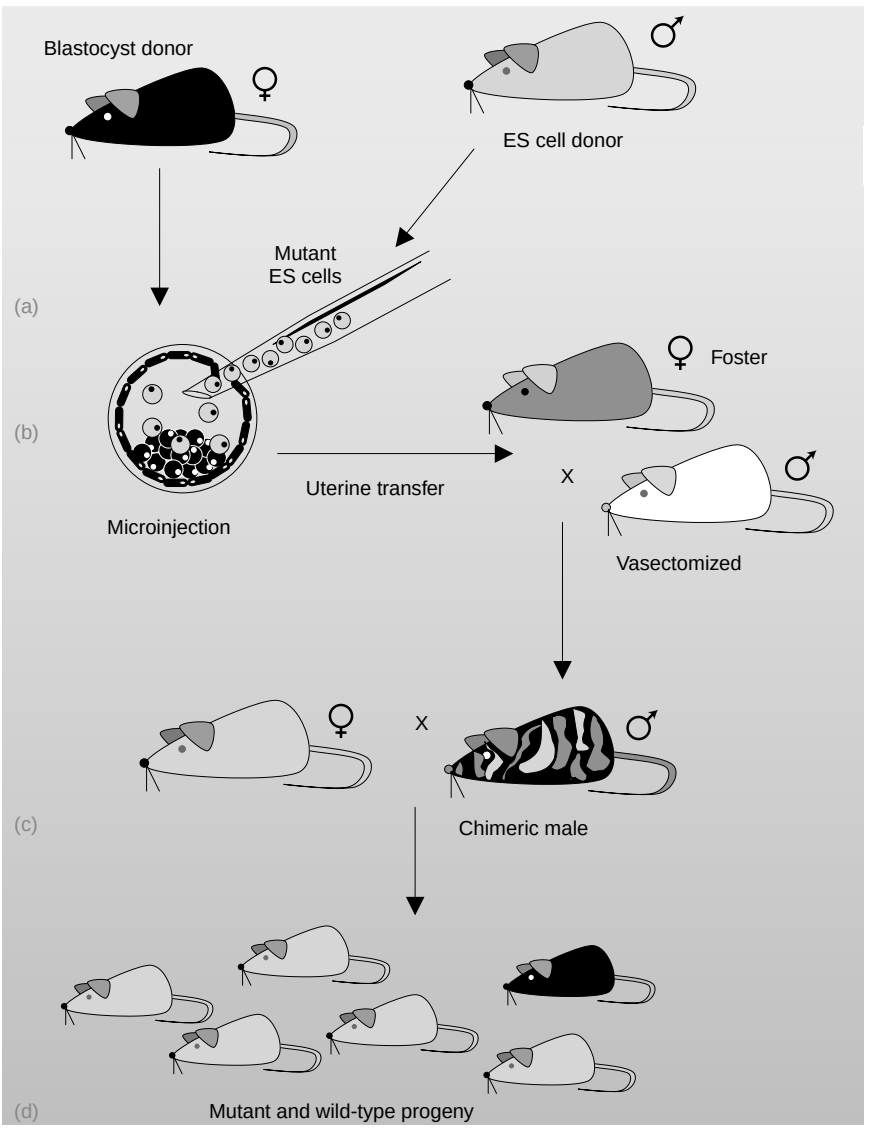

in two independently established homozygous knockout animals.

4 Refinements to the Models

Transgenes or gene knockouts may cause unexpected embryonic or neonatal lethality, either by interfering with embryo development or placentation. If the aim of an animal model is to, for example, study the effect of a transgene or a gene knockout on adult physiology or development of a disease, embryonic lethality limits the usefulness of such a model. To circumvent this problem, animal molecular geneticists have sought to refine animal models. The resulting mouse models often combine aspects of transgenesis and gene targeting and may include a molecular switch, by which null mutation of a gene now is rendered *conditional*. Such on/off models are referred to as *binary models*. The most pivotal advantages of such molecular switches are gain in accuracy and specificity of the model and an important concomitant reduction of disease burden for the animals. Some exciting models and applications that have been developed over the last decade are described below.

4.1 Fully ES Cell–derived Embryos

The placenta is one of the first organ systems to develop during early embryogenesis. It provides an essential means by which the developing embryo and the mother animal exchange essential factors such as nutrients and oxygen. A surprisingly large number of null mutations result in defective placentation. Defective placentation in a conventional genetic approach thwarts all possibilities to uncover a gene function during early embryogenesis. J. Rossant and coworkers developed an elegant solution to this problem during the early 1990s: tetraploid embryos were used to reintroduce diploid ES cells into. The concept is elegantly simple: tetraploid cells will contribute to placentation (polyploidy is a normal feature within the developing placenta); they will, however, not contribute to the development of the embryo itself. It was shown that in this approach, reintroduced normal diploid wild-type ES cells were fully capable of supporting normal embryonic development. The data demonstrated that with this technology

Fig. 10 Biotechnological procedures in conventional gene targeting. (a) Inbred strains are used for superovulation and mating to generate sufficient blastocysts for injection. The procedures in use to obtain sufficient blastocysts for microinjection are natural matings or superovulation. Breeding properties of some inbred strains are relatively poor, hence, natural mating often require large numbers of mating pairs. Super ovulation is often not very efficient in inbred strains either, but will usually yield adequate embryo numbers for injection. (b) Upon microinjection, the embryos are transferred into the oviduct of a pseudopregnant female (see also legend to Fig. 7). (c) Since male ES cells are used, researchers look for a gender bias among chimeric offspring: a potent male (XY) ES cell line forces even female recipient blastocysts (XX) into a male sex phenotype. As a general rule is considered: the higher the coat color mosaicism, the more likely it is that the ES cells participate in spermatogenesis, the higher the chance for germ-line transmission of the mutation; (d)–Germ-line transmission is revealed by the coat color of the donor (ES cell; sand-colored in the figure; several ES cell lines are in use) strain; 50% of the sand-colored offspring will be heterozygous for the germ-line mutation (see color plate p. xxx).

the study of genetically manipulated ES cells in all fetal lineages was possible, while, at the same time, potential negative effects of a null mutation on placental development were circumvented. Tetraploid recipient embryos were generated by electrofusion: normal diploid two-cell-stage embryos were subjected to a defined electroshock that results in fusion of the cells to form one tetraploid one-cell-stage embryo (Fig. 9d). This tetraploid early embryo is allowed to further develop *in vitro*, to be used at a later stage for aggregation or microinjection (see Sect. 3.3). It should be noted that although the role of a gene can now be studied independent of placentation, specific lethality as a result of gene mutation is not prevented. To bypass

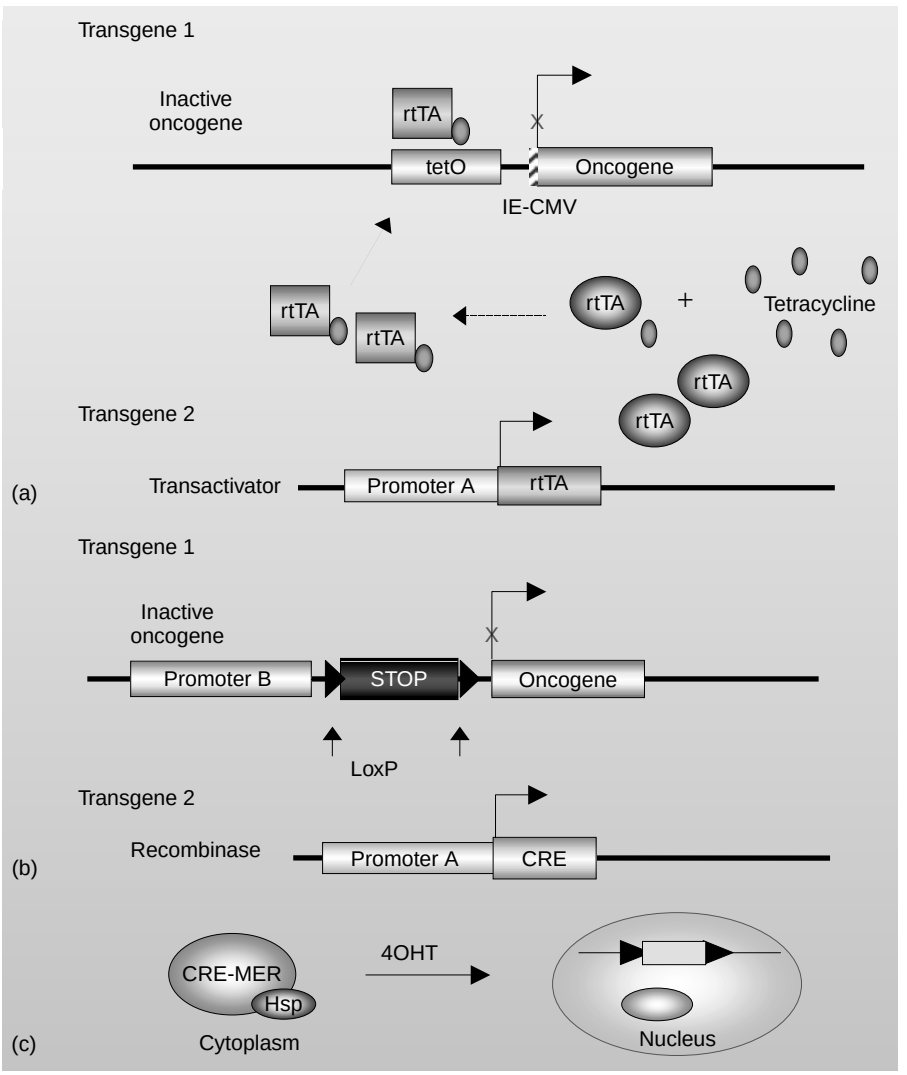

these problems, inducible or binary models are used.

4.2 Conditional Transgenesis

Conditional transgenesis is employed to have better control over (trans)gene expression in a given model. This may be advantageous for a number of reasons, for instance, to circumvent lethality during embryonic development caused by toxic or teratogenic properties of transgenes or to study the onset of disease from its earliest stages onward.

In its simplest form, conditionality is integrated in an animal model by means of tissue-specific promoters and/or control element usage (see Sect. 2.1). Transgene expression is confined to a defined target tissue or select cells within a tissue, for instance, epithelial lung cells (surfactant protein promoter) or anterior pituitary cells (POMC, pro-opiomelanocortin promoter). Several tens to hundreds of models have been created and used in this fashion to successfully study the specific effect(s) of transgene expression in development or disease.

Truly conditional transgenesis, in which the investigator not only determines where a transgene is turned on but also when and how strong, can be achieved with binary models. Binary models incorporate a switching mechanism. Several molecular switches exist to turn on or off transgenes. Figure 11 lists examples of these methods. Most binary switches are reversible. Two main concepts are

Fig. 11 Binary models. (a) The Tet-system requires two independently generated transgenic lines, which are combined by interbreeding. Transgene 1 carries a minimal ("enhancer-less") Cytomegalovirus immediate early (IE–CMV) promoter fused to heptamerized prokaryotic tetO regulatory sequences. Transgene 1 is transcriptionally silent until the transactivator line (transgene 2) provides transactivator protein. In the original Tet-system, transgene 2 expressed a tetracycline (Tc)-controlled transcriptional activator, tTA. tTA is a fusion between the Tet-repressor of the *Tn10* Tc resistance operon of *E. coli* and a C-terminal part of the herpes simplex transactivator protein VP16. tTA will induce transcription from the tetO-controlled transgene up to five orders of magnitude, but is inhibited by Tc. The example depicts rtTA (red ovals), a reverse Tc-controlled transactivator, which is a mutant form of tTA and needs Tc-derivatives to bind tetO (yellow squares). So rather than an inhibitable model, rtTA is an inducible model. Refinements on the Tet-system include choice of regulatory sequences (promoter A) in transgene 2 (cell type–specific, general promoter). In addition, several derivatives of Tc exist, among which Doxocyclin (Dox) has better tissue penetration and pharmacokinetic properties than Tc itself and very high and selective affinity for TetR. (b) The use of bacteriophage P1 CRE-LoxP recombination–dependent removal of a strong transcriptional block (STOP) in a transgene construct may be helpful in defining the molecular events at the onset of tumorigenesis and in addition circumvents possible embryonic lethal effects of the transgene. The choice of regulatory sequences (promoter A) in transgene 2 is determined by the model. Although the excision of a transcriptional block is, in principle, reversible, for practical reasons this system is rarely used as a binary switch, since it would entail obligatory screening for genetic modifications (insertions) in target cells: this obviously compromises the usefulness of the models in terms of precision and rapidity. (c) Fusion of CRE to the mutant forms of the ligand-binding domain of nuclear hormone receptors, such as the estrogen receptor (MER), has generated a recombinase CRE-MER that can be "induced" at the posttranslation level: CRE-MER is retained cytoplasmic through interaction with heatshock proteins (Hsp) such as HSsp90; upon addition of 4-hydroxytamoxifen (4OHT), a synthetic steroid, the Hsp-binding is released, and the recombinase shuttles to the nucleus to excise floxed sequences (also see Fig. 12) (see color plate p. xxxii).

used: (1) a combination of nonmammalian promoters and transcription factors and (2) integration of transcriptional interruptions that are removed by recombination. The use of inducible transcription by means of nonmammalian, for example, regulatory elements borrowed from bacteria or lower eukaryotes, like the fruit fly (*D. melanogaster*), allows transgenes to be activated and silenced "on command." The system that utilizes removal of a strong transcriptional block by means of recombination, although in principal reversible, is, for practical reasons, mostly used one-way.

Several binary switches are itemized below:

- *Tetracycline-controlled transgene expression: the Tet-system.* The transgene of interest is placed under the control of a Tetracycline resistance operon from *Escherichia coli*, the tet-operator (*tetO*). A second, independent transgenic line expressing a derivative of TetR, a transcriptional regulator of tetO, confers Tetracycline (Tc)-inducibility or repressiveness to the transgene of interest (see Fig. 11a). Transcription factors with regulatory activity toward the tetO are absent in eukaryotic cells, which means that theoretically the system cannot be leaky for transcription and pleiotropic effects are therefore not expected.
- *Insect hormone–controlled transgene expression.* The system uses the molting steroid hormone *ecdysone* for transgene induction. It requires a heterodimeric receptor and an ecdysone responsive element, which is incorporated as a regulatory element in the transgene of interest. Administration of *ecdysone* (or a derivative *muristerone A*) rapidly and selectively induces transgene expression. Upon withdrawal of the steroid hormone, expression is silenced. Again, since mammalian cells lack insect hormone receptors, the binary system theoretically works as an on/off system, much like the Tet-system. Ecdysone nor its derivatives are toxic or teratogenic. Owing to excellent pharmacokinetic properties, the method is ideally suited for very precise conditional control over gene expression.
- *Transgene induction by removal of a transcriptional block.* The system makes use of a recombination process derived from bacteriophage P1: phages (bacterial viruses) use a site-specific recombinase (CRE) to integrate their DNA into the genome of their bacterial host. An essential second component of this process is a recombinase recognition site, called an *LoxP site*: two LoxP sites are recognized and recombined by CRE (Fig. 12a). The two essential components, LoxP and CRE, also work in eukaryotic systems, and are commonly maintained on two separate transgenic lines. Flanking a sequence with two similarly oriented LoxP sites will lead to its excision in the presence of CRE. Hence, an on/off transgenic system employs a silenced transgene, in which a strong transcriptional block is bracketed by two similarly oriented LoxP sites; a second independent transgenic line, which expresses the CRE recombinase, may be driven off a general, tissue-specific or inducible promoter, depending on the objective of the study (Fig. 11b). More details on this recombination-based switching mechanism will be discussed in Sect. 4.3.

4.3 Conditional Gene Targeting

Gene targeting is widely used to study gene function in the mouse. Hundreds

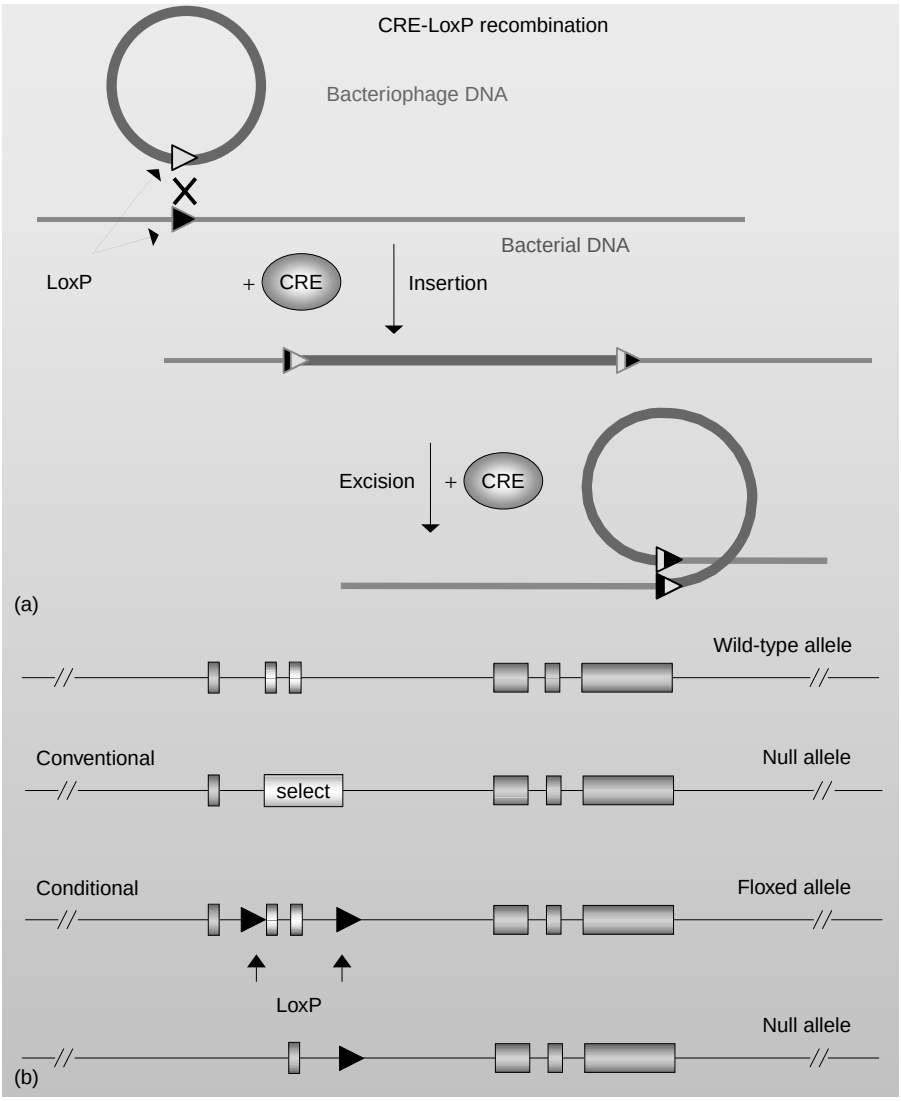

Fig. 12 (a) CRE-LoxP system: LoxP sites are short sequences harboring inverted 13 nucleotide repeats separated by an 8-nucleotide asymmetric spacer. The LoxP are recognition sites for the recombinase CRE. CRE drives site-specific recombination of two identical LoxP sites, for clarity depicted as a yellow and a black triangle. Two similarly oriented LoxP sites on one contiguous DNA strand will be recombined by CRE and the sequence in between the LoxP sites is excised. (b) This recombination principle is used in conditional gene-targeting vectors: a crucial coding region (two blue exons) is ''floxed'' (i.e. flanked by LoxP sites) in ES cells and excised *in vivo*. The LoxP sites in the ''floxed'' allele are assumed not to interfere with wild-type gene expression. The figure compares the structure of a conventional and a conditional knockout allele. A variation on this theme is the FLP-*frt* system, which is derived from yeast. The FLP gene product, like CRE, is a site-specific recombinase; *frt* (FLP recognition target) represents the small inverted repeats, analogous to LoxP, recognized by the recombinase (see color plate p. xxxiv).

of genes have been "knocked out" in the conventional approach, in which the null mutation is transmitted through the germline, as discussed in Sect. 3.1. An estimated one in three null mutations results in embryonic or early postpartum lethality. Therefore, germ-line mutations may be appropriate models to recapitulate human congenital genetic disorders. They may, however, not be the best tools to study gene function in adult mice, since this is hampered by often unexpected and unexplained lethality. K. Rajewsky and coworkers were among the first to develop a conditional gene-targeting strategy that allows inactivation of a gene in a defined *spatio-temporal* setting: the null mutation can, for instance, be confined to selected cell types or to a specific stage during development.

Conditional gene targeting makes use of the CRE-LoxP principle, as described in Sect. 4.2. Briefly, LoxP sites flank a genomic sequence in a targeting vector, which will be deleted *in vivo*; this allele is referred to as a *floxed* allele (Fig. 12b). This final "floxed" wild-type allele only carries recombinase recognition sites at positions within the locus where they are presumed inert, for example, introns: up until the moment of recombination-driven excision *in vivo*, the allele should function as a wild-type allele. Since the CRE-LoxP system is applied more widely than the FLP-*frt* system *in vivo*, further descriptions of conditional gene targeting will focus mainly on the CRE-LoxP system.

The current standard experimental conditional approach requires two separate mouse lines:

1. the "floxed" line (as described above)
2. the "deleter" line, expressing CRE.

The transcriptional control directing CRE expression determines the specificity and degree of regulation of the conditional knockout. CRE-transgene expression is controlled either by constitutively cell type–specific or inducible regulatory elements (also see Sect. 2.1 and 4.2). Deleter strains are most often generated by conventional transgenic technology. Alternatively, the CRE-cDNA can be inserted via targeted mutagenesis to a locus of interest, under control of the endogenous promoter. Although the latter method does not require a thorough knowledge of the regulatory elements controlling the locus at hand, the approach is not as straightforward as conventional transgenesis. Drawbacks of the former method comprise all known problems linked to conventional transgenesis: transgene expression is dependent on integration site and copy number. Consequently, a number of independent founders will have to be generated to select a useful strain. A substantial collection of tissue-specific CRE mouse lines is currently available.

Inducible CRE expression allows for a high level of control (see Sect. 4.2). Again, it should be noted that conditional gene targeting is mainly used unidirectionally, that is, once a gene of interest is inactivated through CRE-mediated recombination, reversion of the genotype is not easily accomplished.

4.3.1 Controlled CRE Expression

Inducibility can be conferred by rather simple measures such as CRE expression through infectious viral particles. Such viruses can be locally administered or systemically. In addition, their tropism (i.e. host-cell range) can be manipulated.

Inducible mammalian promoters have been used, as well as the Tet-system (see Sect. 4.2). Besides the induction of CRE

transcription and translation, CRE induction has also been achieved at the post-translational level. This was accomplished by fusing CRE-encoding sequences in frame to the ligand-binding domain (LBD) of nuclear steroid hormone receptors, such as the estrogen or progesterone receptor (ER, PR respectively). Ligand binding transfers the CRE-ER fusion product into the nucleus, where the recombinase exerts its catalytic activity toward LoxP sites. Mutation of the ER moiety, has generated *CRE-MER* (mutated estrogen receptors) fusions, which are only induced by synthetic hormones and not by their natural ligands (Fig. 11c). Ongoing improvements of these and alike inducible systems clearly generate a highly sensitive and specific means to control transgene expression and enable the study of gene ablation and overexpression that, without such tools, were impossible. An increasing number of published models employing CRE-MER variants attest to the feasibility of the approach *in vivo*.

4.4 Knockins: Humanized Mice

Many studies in human disease have uncovered mutations that are relevant to the etiology of the pathological condition. Examples of these are large congenital chromosomal duplications or deletions in inborn genetic disease, specific dominant mutations in oncogenes, recessive mutations in tumorsuppressor genes, or polymorphisms in genes involved in lipid metabolism linked to atherosclerosis. In order to understand the exact molecular mechanism underlying the development of disease, such mutations need to be exactly recreated in the mouse. Clearly, conventional and conditional genomic manipulations are not specific enough to address such issues. CRE-mediated recombination has provided possibilities and strategies to manipulate the genome beyond conditional gene knockouts or transgenesis. As illustrated in Fig. 13 and 14, the CRE-LoxP system can be used to manipulate the genome in various manners:

- *Gene knockin* or *gene exchange strategies* apply CRE-LoxP-mediated recombination to exchange endogenous coding regions for exogenous or mutated sequences in a locus (Fig. 13a). Especially, the latter application is of significant interest to the scientific community, since it permits the development of animal models, which reflect human genetic disorders more closely than models applying overexpression or ablation of a gene. Initial approaches inserted reporter genes into an endogenous locus; this simple variation on conventional gene targeting achieves the same purpose as a transgene harboring a reporter preceded by endogenous regulatory sequences (see Sect. 2.1). The main difference, of course, is that this strategy obviates the handling of large DNA fragments with potentially unidentified regulatory elements needed to achieve faithful expression patterns, since these elements are provided by the targeting into the endogenous locus. Recently developed strategies take knockin significantly further and allow for specific introduction of (point) mutations and repeated exchange of genetic sequences: recombination-mediated cassette exchange (RMCE) applies strategically positioned pairs of wt and mutated LoxP sites or LoxP and *frt* sites to achieve stepwise and directional replacement of genomic sequences with other of mutated sequences (Fig. 13b).

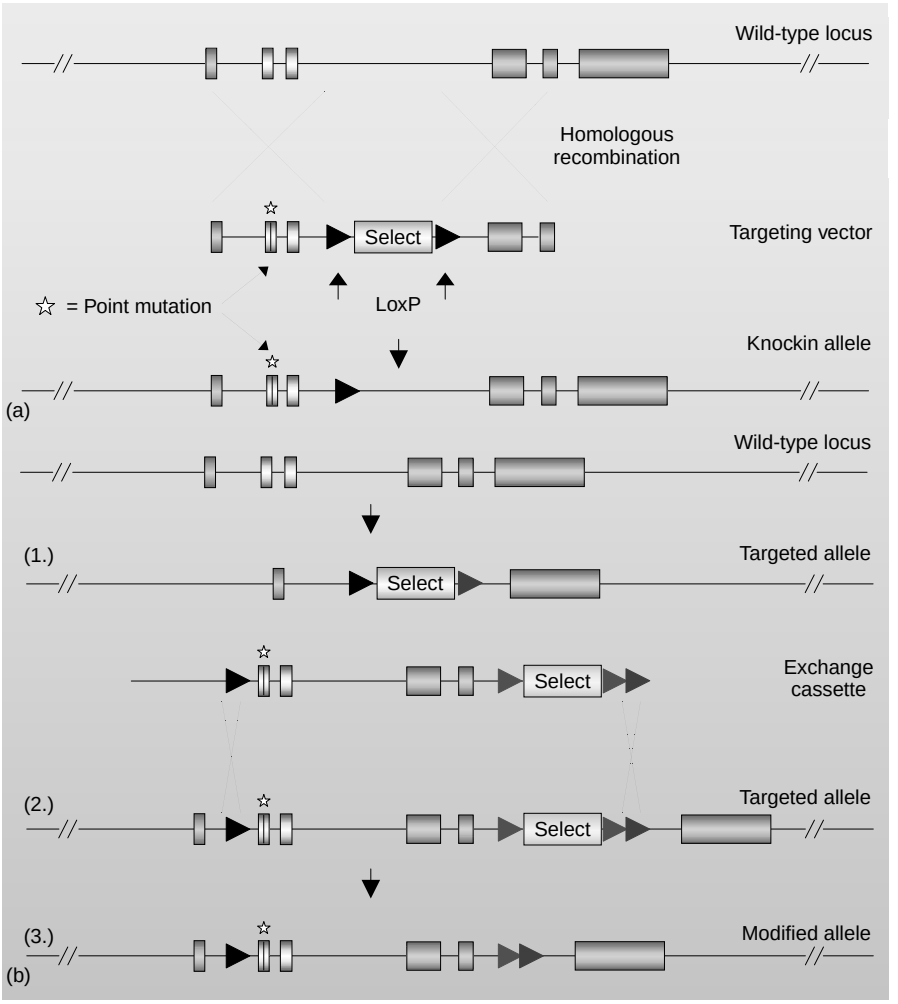

Fig. 13 Applications of site-directed recombination in molecular mouse genetics. (a) A conditional targeting construct usually harbors an antibiotic resistance gene (select) flanked by recombinase recognition sites (either LoxP or *frt*; see legend to Fig. 12). Prior to microinjection into blastocysts or aggregation with morulas, the selection marker is excised in ES cells by recombination (by transient expression of either CRE or FLP). The targeting construct depicted contains a specific point mutation (asterix) in one of its exons. Homologous recombination results in introduction of the mutation in the endogenous locus. (b) The use of multiple nonidentical LoxP and or *frt* sites (black, red, and blue triangles) creates the possibility to repeatedly exchange sequences between vectors and a locus via site-specific recombination. The procedure is referred to as *recombination-mediated cassette exchange (RCME)*. First, a locus is targeted and an asymmetrically "floxed" selection marker is introduced (1). Using the same LoxP variants, an exchange cassette is introduced (2) upon which the selection marker is removed by recombinase-mediated excision (3). The exchange procedure can be repeated multiple times, since the position of the most peripheral recombinase recognition sites (black and red triangles) is maintained (see color plate p. xxxv).

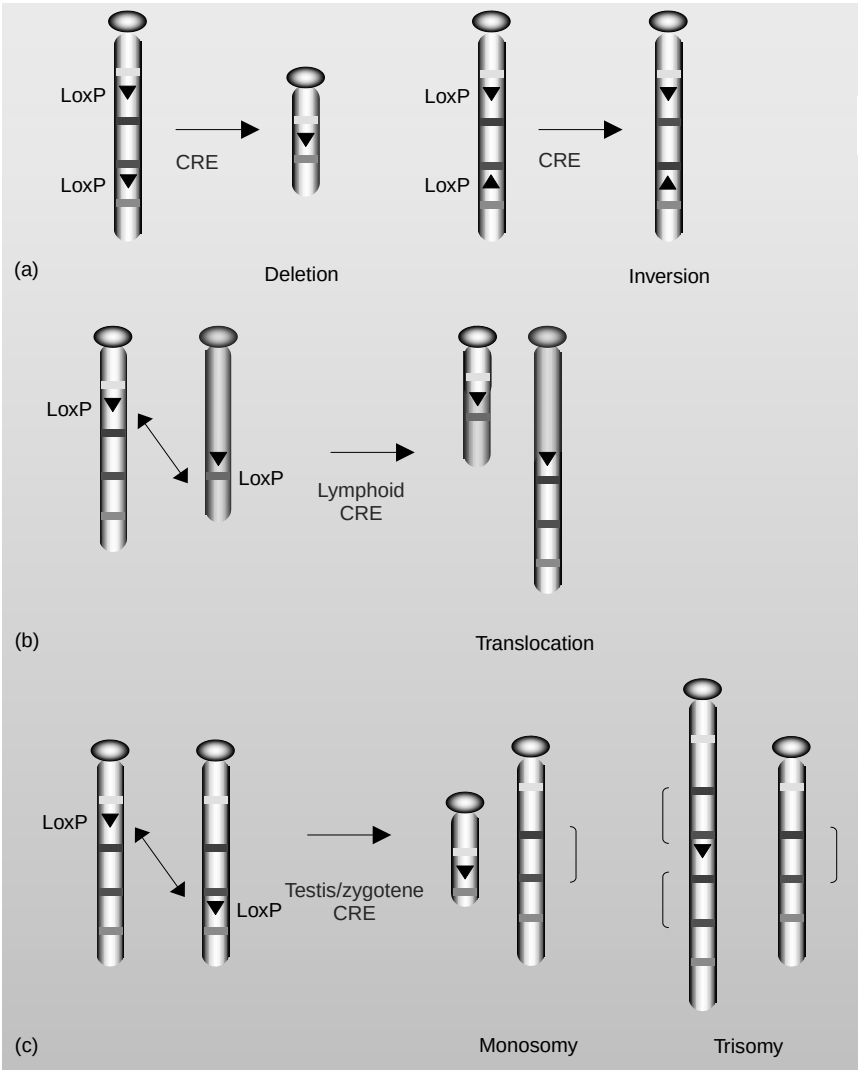

Fig. 14 CRE-LoxP applications in chromosome engineering. (a) Chromosome engineering is an important tool that facilitates functional analysis of large chromosomal regions: large segments of DNA can be either deleted or inversed, depending on the orientation of the LoxP sites in respect to each other. Series of overlapping deletions can be generated by combined use of one targeted LoxP site and retrovirus-mediated random insertion of additional LoxP sites. CRE-LoxP recombination may be used to generate balancer chromosomes to maintain lethal recessive mutations and has the potential to facilitate large-scale mutagenesis screens. Deletion or inversion carriers can be used to uncover and instantly map, for example, ENU-induced recessive phenotypes to deleted or inversed chromosomal regions. (b) Interchromosomal recombination by positioning LoxP sites on nonhomologous chromosomes will recreate translocations, which occur in human disease. (c) The use of CRE-mediated recombination to create defined autosomal trisomies in the mouse, for example, through meiosis-/zygotene-specific CRE-deleter strains that are solely active in testis.

- *Chromosomal deletions, inversions, translocations.* Simultaneous specific positioning of LoxP sites on one chromosome allows for gene inversions or complete deletions, depending on the orientation of the recombinase recognition sites relative to each other (Fig. 14). Positioning of such recombinase recognition sites on nonhomologous chromosomes has proven to be a feasible strategy to achieve chromosomal translocation, alike translocations that are implicated in some leukemias.
- *Targeted integration of transgenes* to loci, which are known to be "open," that is, not embedded in heterochromatin. The gain of this approach is a high probability of transgene expression. An often-used target locus is the *ROSA26* locus, an ubiquitously transcribed locus that does not encode a protein, but which is active in many different cell types throughout development and adult life. Transgenes are flanked by DNA sequences homologous to the *ROSA26* locus and targeted via conventional methods (see Sect. 3).
- A substantial number of other applications for site-directed recombination has been reported on in the scientific literature. Since the principal mechanism remains the same in all instances, a few of these applications are listed here, but not worked out in detail: *selection marker recycling, exchange of reporter genes, removal of reporter genes, or selection markers* (as in the generation of floxed alleles).

5 Random Mutagenesis in ES Cells

The completion of the mouse and human genome sequences has sparked renewed interest in developing tools for genome-wide mutagenesis in the mouse. Among the main molecular genetic technologies currently pursued are targeted mutagenesis (see Sect. 3 and 4), *insertional mutagenesis,* by insertion of gene trap vectors or viral or transposable sequences, and chemical mutagenesis. These technologies all represent random mutagenesis protocols, which, in principle, can be applied to mouse ES cells. The important advantage of random mutagenesis is that new mutations can be generated in mice at

Fig. 15 Random mutagenesis strategies. (a) Promoter-trap and poly(A)-trap vectors, respectively, catch or induce transcriptional activity and have the potential to identify most, if not all, of the active and inactive genes in the genome, including alternatively spliced forms and low-abundance transcripts, and is thus an important tool in genome annotation. (b) Gene mutation through random insertion of retroviral vectors (or transposable elements; see Sect. 5.2). Shown is an oncogene, which is activated by a nearby proviral insert, and a tumorsuppressor gene, which is inactivated proviral insertion into the gene. Inserted elements are engineered such that they simultaneously function as sequencing tags to identify the surrounding genomic DNA. (c) Stable or inducible RNA interference as an alternative to targeted mutagenesis to study gene function in mice: a short double-strand RNA (ds) stem-loop-stem structure (white arrows, blue loop) is produced from a RNA polymerase III-promoted vector. The loop is enzymatically removed inside the cell; the resulting short interfering RNAs will bind their complementary mRNA, upon which the target mRNA is degraded by a complex called *RISC (RNA-induced silencing complex)*. RNA interference–mediated gene silencing does not require targeted mutagenesis of a gene, and may therefore represent a more efficient and rapid method to assess gene function (see color plate p. xxxvi).

a rate far exceeding conventional gene targeting. In addition, it is possible to identify and maintain recessive lethal phenotypes at any developmental stage, which, for instance, is not possible with chemical mutagenesis (see Sect. 5.3).

5.1 Gene Trap Mutagenesis

Gene trapping provides a method to mutate genes in the genome through random insertion throughout the genome.

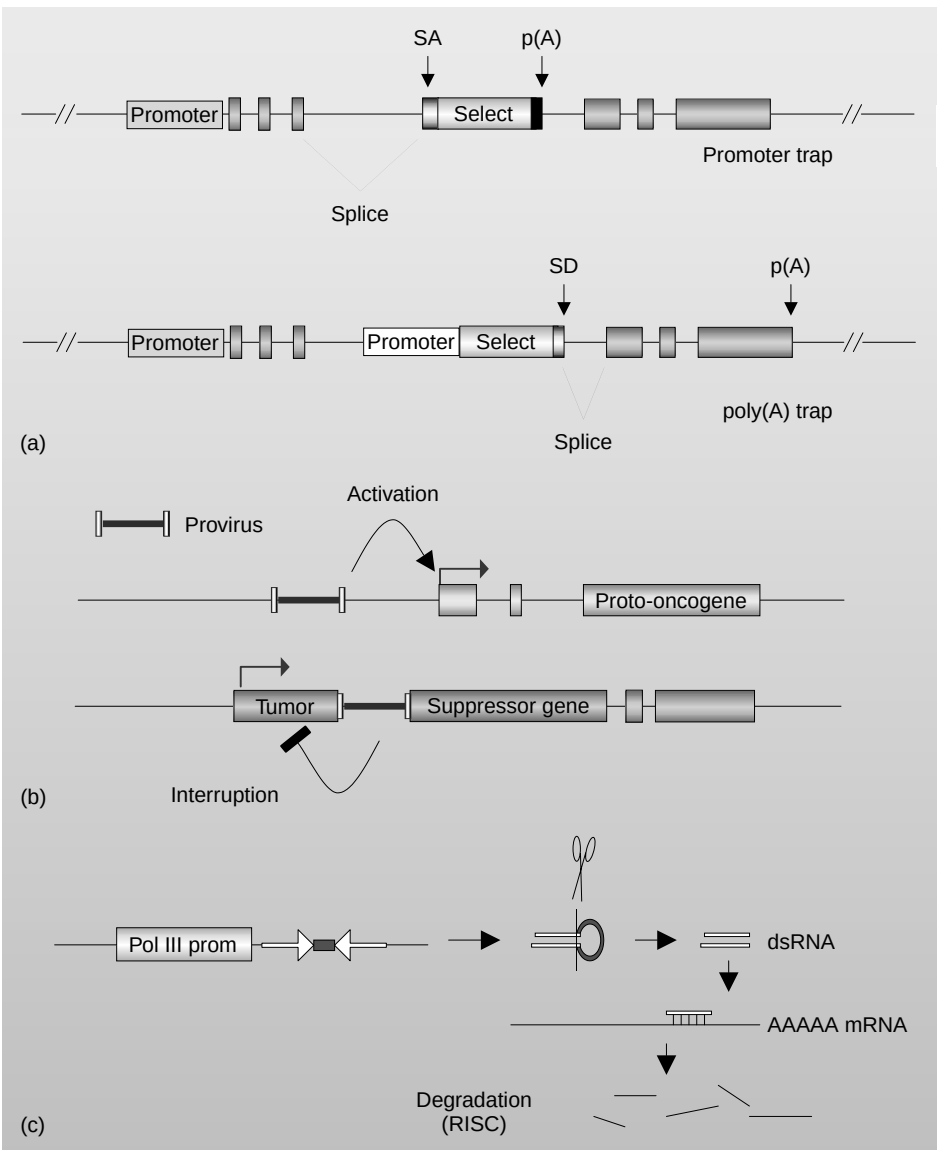

Essentially, all currently designed gene trap vectors insert a selectable marker or reporter gene lacking essential regulatory sequences into an endogenous gene. Through marker insertion the gene is interrupted, while marker expression is complemented by endogenous regulatory elements (Fig. 15a). In this manner, transcriptionally active as well as inactive loci can be trapped. Recent reports provided proof that gene trapping can also be used to selectively mutagenize genes that share certain properties, in this case, the fact that their gene products are all transmembrane proteins. A "secretory" trap vector was constructed such that it captured signal sequences in preceding exons by means of including a "capturing" transmembrane domain in the gene trap construct. These and other approaches collectively demonstrate the power and versatility of gene trapping in genome-wide functional analysis of molecular mechanisms important in development and physiology. A collective gene trap database has recently been initiated by a number of consortia (http://www.igtc.ca/index.html – http://www.escells.ca/ – http://www.mmrrc.org/ – http://tikus.gsf.de/project/web_new/index.html – http://www.sanger.ac.uk/PostGenomics/genetrap/ – http://baygenomics.ucsf.edu/overview/welcome.html).

5.2 Retroviruses and Transposable Elements in Random Mutagenesis

The use of *retroviruses* in random mutagenesis is adapted from applications in which proviruses are used to screen for collaborating oncogenes in leukemogenesis. Slow transforming retroviruses such as Moloney Murine Leukemia Virus (MoMuLV) cause transformation by integration into the genome. Proviral integration may activate oncogenes or interrupt tumor suppressor genes (Fig. 15b). This is an extremely useful asset in random mutagenesis, since the integrated proviruses now serve as a *molecular tag* for cancer genes: it can be used to identify the locus at which it is integrated by sequencing the flanking integration site. Retroviral vectors have been adapted for use in mouse embryos and ES cells as well, and now permits germ-line mutagenesis.

In analogy to random proviral insertion, also *transposable elements* from yeast, insects, fish, and mammals, among other organisms, can be used for functional genomic analyses. Transposable elements (TEs) are a heterogeneous class of genetic elements that have the capacity to move from one site on a chromosome to another: they "jump." TEs are very efficient at integrating into DNA, and some elements can carry extra DNA. TEs are therefore useful vectors for transferring new genetic material into genomes and for random insertional mutagenesis. By way of example, among the properties that make the Tc*1*/*mariner* superfamily of TEs useful for molecular genetic research are the facts that they are easy to work with, their efficient and stable integration, and persistent, long-term transgene expression following transposon-mediated gene transfer. Importantly, such TEs "jump" in species other than their original hosts, which underscores the need for further development and use of these and alike molecular tools for functional genomics in the mouse. Several reports on germline induced TE transposition and germline attest to the mutagenic potential of the approach.

5.3
Random Enu Mutation

For many years, mutant mouse models were obtained largely as a "spin-off" from research on the mutagenic properties of chemical compounds and radiation. With the recent advent of more genome resources, it has become possible to identify such induced mutations more rapidly. Initially, substantial efforts were undertaken, involving collaborations of many laboratories, each with their own phenotyping expertise. At present, robust technology is in place to mutagenize mice. One of the *caveats* remained however, because, although screens for recessive mutations are the most promising ones, such screens are exceedingly difficult and expensive. Recently, it was shown that a narrowly focused screening system for recessive mutations can be highly productive, while the expenses remained acceptable. Moreover, M. Justice, A. Bradley, and colleagues made use of a set of strains carrying engineered coat color markers and a chromosomal inversion (see Fig. 14a), enabling to further reduce costs by focusing the screen on a predefined chromosomal segment. Given the large number of loci controlling complex genetic disorders, this strategy permits an ordered approach to finding disease genes. The strategy is highly suitable to focus on some synthenic regions in the mouse genome, which are known to be associated with disease susceptibility in man.

6
Phenotype Analysis

Genetically engineered mouse models serve to study genotype–phenotype relationships and can show an impressive variety of phenotypes, ranging from embryonic lethality to increased susceptibility to chronic diseases such as cardiovascular disease and cancer. The models allow defining primary gene functions as well as the role of a particular gene mutation in disease.

Once mouse models have been generated, it is important to carry out an in-depth analysis of the pathology of the entire mouse and not restrict the analysis to the target organs that are the main subject of the intended study. A detailed strategy is outlined elsewhere. Often, one will find that the dramatic alterations in the gene expression pattern will show unexpected effects on the phenotype. For instance, in the case of a study on apolipoprotein C1, the expected perturbation of lipid metabolism was accompanied by a dramatic and progressive hair loss of the mice with the highest level of expression of the transgene. The most common problem encountered, however, in studies using genetically modified mouse models is a lack of phenotype, even when a unique and evolutionarily conserved gene has been altered for which there are no obvious functionally redundant homologs present in the mouse genome. The main cause of this is the large difference in environmental stress experienced by wild mice as opposed to inbred laboratory mice. Often, phenotypes need to be evoked by either increasing the environmental stress (i.e. high-fat diets, mutagens) or increasing the genetic stress using a second transgenic strain with a mutation in a related pathway. A good example is provided by the use of mouse strains in cardiovascular research. Wild-type laboratory mice are resistant to develop atherosclerosis because of its highly antiatherogenic lipoprotein profile. By knocking out genes controlling lipoprotein metabolism, such as the

apolipoprotein E gene, the mice become sensitized. Upon feeding a moderated high-fat diet, such mice will develop complex lesions in the aorta within 2 months. Many gene alterations will have no effect on atherogenesis when studied separately. However, when these mice are crossed with apolipoprotein E–deficient mice, it is now possible to study the role of such genes. Thus, the apolipoprotein E–deficient mouse serves as a sensitized mouse model that is highly susceptible to the effects of additional genetic changes. Likewise, changing the genetic background of the mice will affect susceptibility to disease; this approach has been used to define novel loci in human disease.

Recent technological advances permit analysis of transgene expression and the biological consequences thereof by non-invasive methods. In vital imaging, very sensitive cameras measure bioluminescence from, for example, tumors or embryos *in utero*, which express luciferase of GFP reporter transgenes in living animals (Fig. 2d). Clearly, the impact of technological developments like these is substantial not only in terms of increased specificity and resolution of the model but also in terms of reduced animal usage.

7 Perspectives

7.1 Partial Transgenesis; Viral Shuttles

Recent technological advances allow more rapid approaches to study gene function without the need for germ-line modification. Remarkably, partial transgenesis can be achieved through local electroporation of transgenic vectors into tissues. Using this approach, local muscle tissue in rats has been specifically modified to study its function. Another example is the study of cell migratory and (trans)differentiation processes during neural tube closure; in these studies, one side of the tube is selectively electroporated with an expression construct, while the other side functions as a control.

The use of viral vectors to deliver transgenes to mammalian cells is another method to achieve partial transgenesis. The use of several classes and combinations of viral vectors is being explored in gene therapeutic applications. Similar vectors are very useful in accomplishing partial transgenesis, restricted to certain tissues such as bone marrow (which is easily harvested and cultured *ex vivo* and reintroduced into a recipient mouse). The latter approach can also be employed to, for instance, study isolated effects of gene knockout in the hematopoietic compartment in combination with bone marrow transplantation. Local and transient transgenesis is currently used to activate conditional floxed transgenes (see Sect. 5), for instance by adenovirus-encoded CRE. This approach was proven useful in, for example, somatic activation of transforming oncogenes in the lung.

Lentiviral vectors, unlike the classical slow replicating retroviruses, harbor the advantage that they are not silenced in embryonic stem cells, which makes expression characteristics of such vectors good. Recombinant lentiviruses were recently used to introduce transgenes into fertilized oocytes and ES cells (Fig. 6b). The advantages of using lentiviral shuttle vector are significant: it obviates potentially harmful procedures otherwise used to introduce transgenes into a cell (pronuclear microinjection, electroporation) and offers some degree of control over transgene copy numbers.

7.2 RNA Interference; Posttranscriptional Gene Silencing

The abundance of biological information that has become available upon completion of the genome sequence of both mouse and human is exciting and overwhelming at the same time. An estimated 10 000 genes have been identified currently, which represents an estimated one-third of the total gene complement in a human or mouse. Of these genes, only a few thousand have been functionally studied. Clearly, one of the remaining challenges in molecular genetics is to analyze the function of novel genes. Gene targeting is beyond any doubt one of the most successful and informative technologies when it comes to elucidating gene function. Despite improvements in cloning technologies that employ recombination in prokaryotes instead of "standard" recombinant DNA technology and that significantly simplify otherwise laborious cloning strategies, gene targeting remains time-consuming and expensive.

Recent scientific and technological advancements may, however, provide faster means to address novel gene function. One of the most exciting developments finds its origin in the realization that gene expression is controlled at a posttranscriptional level by double-strand (ds) RNA-mediated degradation of cytoplasmic RNAs. This process, called *RNA silencing* – or *posttranscriptional gene silencing* – was first discovered in plants as an unexpected gene silencing in transgenic plants: silencing of an endogenous gene often occurred in concert with silencing of the transgene. Transgene-mediated RNA silencing is now recognized as a form of RNA silencing that uses small interfering dsRNAs (siRNA), which serve as a guide for specific target mRNA recognition and its subsequent enzymatic degradation. Investigations by many groups showed that RNA interference (RNAi) is a regulatory process used by many organisms to control gene expression. Many of the protein factors involved in RNAi are conserved, from protozoans to humans. RNAi has recently been developed into a technology by which endogenous gene function can be assessed in many vertebrates, including mammals. One obvious advantage of RNAi-mediated gene silencing is that it does not rely on genetic modification. This makes the technology very rapid compared to (conditional) gene targeting. Proof of principle has been achieved in many vertebrate systems including several human cells lines. RNAi also works in early mouse embryos, and was recently demonstrated to be transmittable via the germline in transgenic mice. Whereas initial strategies in cell lines made use of synthetic dsRNA sequences, a relative costly approach, RNAi, is currently achieved via stable expression. In addition, recent reports show that random siRNA libraries can be generated from cDNA libraries via ingenious cloning strategies. The potential held by this approach is that such libraries are self-selecting, thereby enabling investigators to set-up high-throughput genetic screens. RNAi can be made tissue specific, inducible, and can be delivered to mouse embryos using viral vectors. Such improvements will clearly enhance the application of RNAi in functional genomic research.

Although genetic manipulation is possible in tissue culture, the study of genetic interaction of transgenes with other genes and local or systemic effects of transgene expression in the intact organism provides a much more complete and physiologically relevant picture of a gene's function than can ever be achieved *in vitro*. Genetic modification therefore represents an

important and biologically relevant tool to complement *in vitro* gene expression studies aimed at, for example, delineation of signal transduction pathways or identification of tissue-specific regulatory elements. Refinements to genetically modified mouse models in biomedical research are of importance for a number of reasons. Conditional models recapitulate the pathophysiology as it occurs in humans and permit a high degree of selectivity and resolution, which not only renders studies more specific and focused but also significantly reduces the burden on the laboratory animal. The positive effects thereof are reduced stress on the system and, concurrently, improved specificity.

Bibliography

Books and Reviews

Bass, B.L. (2000) Double-stranded RNA as a template for gene silencing, *Cell* **101**, 235–238.

Beier, D.R. (2002) ENU mutagenesis: a work in progress, *Physiol. Genomics* **11**, 111–113.

Berns, A. (1988) Provirus tagging as an instrument to identify oncogenes and to establish synergism between oncogenes, *Arch. Virol.* **102**, 1–18.

Botteri, F.M., van der Putten, H., Miller, A.D., Fan, H. et al. (1986) Recombinant retroviruses in transgenic mice, *Ann. N. Y. Acad. Sci.* **478**, 255–268.

de Winther, M.P., Heeringa, P. (2003) Bone marrow transplantations to study gene function in hematopoietic cells, *Methods Mol. Biol.* **209**, 281–292.

Dudley, J.P. (2003) Tag, you're hit: retroviral insertions identify genes involved in cancer, *Trends Mol. Med.* **9**, 43–45.

Fazio, S., Linton, M.F. (2001) Mouse models of hyperlipidemia and atherosclerosis, *Front. Biosci.* **6**, D515–D525.

Gijbels, M.J., de Winther, M.P. (2003) Autopsy and histologic analysis of the transgenic mouse, *Methods Mol. Biol.* **209**, 267–279.

Hernandez, D., Fisher, E.M. (1999) Mouse autosomal trisomy: two's company, three's a crowd, *Trends Genet.* **15**, 241–247.

Hofker, M., van Deursen, J. (2002) *Transgenic Mouse: Methods and Protocols (Methods in molecular biology)*, Humana press, Totowa, NJ, U.S.A.

Hogan, B., Beddington, R., Costantini, F., Lacey, E. (1994) *Manipulating the Mouse Embryo. A Laboratory Manual*, 2nd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Ivics, Z., Izsvak, Z. (2004) Transposable elements for transgenesis and insertional mutagenesis in vertebrates: a contemporary review of experimental strategies, *Methods Mol. Biol.* **260**, 255–276.

Jackson, I.J., Catherine, M., Abbott, C.M. (2000) *Mouse Genetics and Transgenics: A Practical Approach*, Oxford University press, New York, U.S.A.

Justice, M.J., Noveroske, J.K., Weber, J.S., Zheng, B. et al. (1999) Mouse ENU mutagenesis, *Hum. Mol. Genet.* **8**, 1955–1963.

Kootstra, N.A., Verma, I.M. (2003) Gene therapy with viral vectors, *Annu. Rev. Pharmacol. Toxicol.* **43**, 413–439.

Kuhn, R., Schwenk, F. (2003) Conditional knockout mice, *Methods Mol. Biol.* **209**, 159–185.

Li, Q., Harju, S., Peterson, K.R. (1999) Locus control regions: coming of age at a decade plus, *Trends Genet.* **15**, 403–408.

Maatman, R., Gertsenstein, M., de Meijer, E., Nagy, A., Vintersten, K. (2003) Aggregation of embryos and embryonic stem cells, *Methods Mol. Biol.* **209**, 201–230.

Massoud, T.F., Gambhir, S.S. (2003) Molecular imaging in living subjects: seeing fundamental biological processes in a new light, *Genes Dev.* **17**, 545–580.

McCormick, S.P., Liu, C.Y., Young, S.G., Nielsen, L.B. (2003) Manipulating large insert clones for transgenesis, *Methods Mol. Biol.* **209**, 105–123.

Mikkers, H., Berns, A. (2003) Retroviral insertional mutagenesis: tagging cancer pathways, *Adv. Cancer Res.* **88**, 53–99.

Nagy, A. (2000) Cre recombinase: the universal reagent for genome tailoring, *Genesis* **26**, 99–109.

Nagy, A., Gertsenstein, M., Vintersten, K., Behringer, R. (2002) *Manipulating the Mouse Embryo: A Laboratory Manual*, 3rd edition,

Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
Osumi, N., Inoue, T. (2001) Gene transfer into cultured mammalian embryos by electroporation, *Methods* **24**, 35–42.
Pedersen, Ra.R.J. (1989) *Transgenic Techniques in Mice: A Video Guide*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
Ramirez-Solis, R., Liu, P., Bradley, A. (1995) Chromosome engineering in mice, *Nature* **378**, 720–724.
Roebroek, A.J., Wu, X., Bram, R.J. (2003) Knockin approaches, *Methods Mol. Biol.* **209**, 187–200.
Rossant, J., Cross, J.C. (2001) Placental development: lessons from mouse mutants, *Nat. Rev. Genet.* **2**, 538–548.
Saunders, T.L. (2003) Reporter molecules in genetically engineered mice, *Methods Mol. Biol.* **209**, 125–143.
Sharp, P.A. (1999) RNAi and double-strand RNA, *Genes Dev.* **13**, 139–141.
Silver, L. (1995) *Mouse Genetics: Concepts and Applications*, Oxford University press, New York, U.S.A.
Soewarto, D., Blanquet, V., Hrabe de Angelis, M. (2003) Random ENU mutagenesis, *Methods Mol. Biol.* **209**, 249–266.
Somia, N., Verma, I.M. (2000) Gene therapy: trials and tribulations, *Nat. Rev. Genet.* **1**, 91–99.
Soriano, P. (1999) Generalized lacZ expression with the ROSA26 Cre reporter strain, *Nat. Genet.* **21**, 70–71.
Voncken, J.W. (2003a) Genetic modification of the mouse. General technology; pronuclear and blastocyst injection, *Methods Mol. Biol.* **209**, 9–34.
Voncken, J.W. (2003b) Transgene design, *Methods Mol. Biol.* **209**, 51–67.
Yang, M., Baranov, E., Moossa, A.R., Penman, S. et al. (2000) Visualizing gene expression by whole-body fluorescence imaging, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 12278–12282.

Primary Literature

Baharvand, H., Matthaei, K.I. (2004) Culture condition difference for establishment of new embryonic stem cell lines from the C57BL/6 AND BALB/c mouse strains, *In Vitro Cell. Dev. Biol. Anim.* **40**, 76–81.
Brummelkamp, T.R., Bernards, R., Agami, R. (2002) A system for stable expression of short interfering RNAs in mammalian cells, *Science* **296**, 550–553.
Carmell, M.A., Zhang, L., Conklin, D.S., Hannon, G.J. et al. (2003) Germline transmission of RNAi in mice, *Nat. Struct. Biol.* **10**, 91–92.
Choi, T., Huang, M., Gorman, C., Jaenisch, R. (1991) A generic intron increases gene expression in transgenic mice, *Mol. Cell. Biol.* **11**, 3070–3074.
Christ, N., Droge, P. (2002) Genetic manipulation of mouse embryonic stem cells by mutant lambda integrase, *Genesis* **32**, 203–208.
Doerfler, W. (1991) Patterns of DNA methylation–evolutionary vestiges of foreign DNA inactivation as a host defense mechanism. A proposal, *Biol. Chem. Hoppe Seyler* **372**, 557–564.
Dupuy, A.J., Fritz, S., Largaespada, D.A. (2001) Transposition and gene disruption in the male germline of the mouse, *Genesis* **30**, 82–88.
Escary, J.L., Perreau, J., Dumenil, D., Ezine, S. et al. (1993) Leukaemia inhibitory factor is necessary for maintenance of haematopoietic stem cells and thymocyte stimulation, *Nature* **363**, 361–364.
Evans, M.J., Kaufman, M.H. (1981) Establishment in culture of pluripotential cells from mouse embryos, *Nature* **292**, 154–156.
Feil, R., Brocard, J., Mascrez, B., LeMeur, M. et al. (1996) Ligand-activated site-specific recombination in mice, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10887–10890.
Folger, K.R., Wong, E.A., Wahl, G., Capecchi, M.R. (1982) Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules, *Mol. Cell. Biol.* **2**, 1372–1387.
Forrester, L.M., Nagy, A., Sam, M., Watt, A. et al. (1996) An induction gene trap screen in embryonic stem cells: Identification of genes that respond to retinoic acid in vitro, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1677–1682.
Ghoshal, K., Jacob, S.T. (2000) Regulation of metallothionein gene expression, *Prog. Nucleic. Acid Res. Mol. Biol.* **66**, 357–384.
Gordon, J.W., Ruddle, F.H. (1981) Integration and stable germ line transmission of genes injected into mouse pronuclei, *Science* **214**, 1244–1246.
Gossen, M., Bujard, H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters, *Proc. Natl. Acad. Sci. U. S. A.* **89**, 5547–5551.

Gossler, A., Joyner, A.L., Rossant, J., Skarnes, W.C. (1989) Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes, *Science* **244**, 463–465.

Hadjantonakis, A.K., Dickinson, M.E., Fraser, S.E., Papaioannou, V.E. (2003) Technicolour transgenics: imaging tools for functional genomics in the mouse, *Nat. Rev. Genet.* **4**, 613–625.

Ikawa, M., Tanaka, N., Kao, W.W., Verma, I.M. (2003) Generation of transgenic mice using lentiviral vectors: a novel preclinical assessment of lentiviral vectors for gene therapy, *Mol. Ther.* **8**, 666–673.

Jaenisch, R. (1976) Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus, *Proc. Natl. Acad. Sci. U. S. A.* **73**, 1260–1264.

Jahner, D., Haase, K., Mulligan, R., Jaenisch, R. (1985) Insertion of the bacterial gpt gene into the germ line of mice by retroviral infection, *Proc. Natl. Acad. Sci. U. S. A.* **82**, 6927–6931.

Jong, M.C., Gijbels, M.J., Dahlmans, V.E., Gorp, P.J. et al. (1998) Hyperlipidemia and cutaneous abnormalities in transgenic mice overexpressing human apolipoprotein C1, *J. Clin. Invest.* **101**, 145–152.

Kisseberth, W.C., Brettingen, N.T., Lohse, J.K., Sandgren, E.P. (1999) Ubiquitous expression of marker transgenes in mice and rats, *Dev. Biol.* **214**, 128–138.

Koller, B.H., Hagemann, L.J., Doetschman, T., Hagaman, J.R. et al. (1989) Germ-line transmission of a planned alteration made in a hypoxanthine phosphoribosyltransferase gene by homologous recombination in embryonic stem cells, *Proc. Natl. Acad. Sci. U. S. A.* **86**, 8927–8931.

Lacy, E., Roberts, S., Evans, E.P., Burtenshaw, M.D. et al. (1983) A foreign beta-globin gene in transgenic mice: integration at abnormal chromosomal positions and expression in inappropriate tissues, *Cell* **34**, 343–358.

Lyons, S.K., Meuwissen, R., Krimpenfort, P., Berns, A. (2003) The generation of a conditional reporter that enables bioluminescence imaging of Cre/loxP-dependent tumorigenesis in mice, *Cancer Res.* **63**, 7042–7046.

Mansour, S.L., Thomas, K.R., Capecchi, M.R. (1988) Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to nonselectable genes, *Nature* **336**, 348–352.

Mao, X., Fujiwara, Y., Orkin, S.H. (1999) Improved reporter strain for monitoring Cre recombinase-mediated DNA excisions in mice, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 5037–5042.

Martin, G.R. (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells, *Proc. Natl. Acad. Sci. U. S. A.* **78**, 7634–7638.

Metzger, D., Clifford, J., Chiba, H., Chambon, P. (1995) Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase, *Proc. Natl. Acad. Sci. U. S. A.* **92**, 6991–6995.

Meuwissen, R., Linn, S.C., Linnoila, R.I., Zevenhoven, J. et al. (2003) Induction of small cell lung cancer by somatic inactivation of both Trp53 and Rb1 in a conditional mouse model, *Cancer Cell* **4**, 181–189.

Mural, R.J., Adams, M.D., Myers, E.W., Smith, H.O. et al. (2002) A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome, *Science* **296**, 1661–1671.

Nagy, A., Gocza, E., Diaz, E.M., Prideaux, V.R. et al. (1990) Embryonic stem cells alone are able to support fetal development in the mouse, *Development* **110**, 815–821.

Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. et al. (1993) Derivation of completely cell culture-derived mice from early-passage embryonic stem cells, *Proc. Natl. Acad. Sci. U. S. A.* **90**, 8424–8428.

No, D., Yao, T.P., Evans, R.M. (1996) Ecdysone-inducible gene expression in mammalian cells and transgenic mice, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 3346–3351.

Ostertag, E.M., DeBerardinis, R.J., Goodier, J.L., Zhang, Y. et al. (2002) A mouse model of human L1 retrotransposition, *Nat. Genet.* **32**, 655–660.

Paddison, P.J., Caudy, A.A., Hannon, G.J. (2002) Stable suppression of gene expression by RNAi in mammalian cells, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 1443–1448.

Palmiter, R.D., Chen, H.Y., Brinster, R.L. (1982) Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring, *Cell* **29**, 701–710.

Palmiter, R.D., Brinster, R.L., Hammer, R.E., Trumbauer, M.E. et al. (1982) Dramatic growth of mice that develop from eggs

microinjected with metallothionein-growth hormone fusion genes, *Nature* **300**, 611–615.
Rajewsky, K., Gu, H., Kuhn, R., Betz, U.A. et al. (1996) Conditional gene targeting, *J. Clin. Invest.* **98**, 600–603.
Robertson, E., Bradley, A., Kuehn, M., Evans, M. (1986) Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector, *Nature* **323**, 445–448.
Schonig, K., Bujard, H. (2003) Generating conditional mouse mutants via tetracycline-controlled gene expression, *Methods Mol. Biol.* **209**, 69–104.
Sen, G., Wehrman, T.S., Myers, J.W., Blau, H.M. (2004) Restriction enzyme-generated siRNA (REGS) vectors and libraries, *Nat. Genet.* **36**, 183–189.
Shirane, D., Sugao, K., Namiki, S., Tanabe, M. et al. (2004) Enzymatic production of RNAi libraries from cDNAs, *Nat. Genet.* **36**, 190–196.
Skarnes, W.C., Auerbach, B.A., Joyner, A.L. (1992) A gene trap approach in mouse embryonic stem cells: the lacZ reported is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice, *Genes Dev.* **6**, 903–918.
Su, H., Wang, X., Bradley, A. (2000) Nested chromosomal deletions induced with retroviral vectors in mice, *Nat. Genet.* **24**, 92–95.
te Riele, H., Maandag, E.R., Berns, A. (1992) Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs, *Proc. Natl. Acad. Sci. U. S. A.* **89**, 5128–5132.
Thomas, K.R., Capecchi, M.R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells, *Cell* **51**, 503–512.
Thompson, S., Clarke, A.R., Pow, A.M., Hooper, M.L. et al. (1989) Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells, *Cell* **56**, 313–321.
Valenzuela, D.M., Murphy, A.J., Frendewey, D., Gale, N.W. et al. (2003) High-throughput engineering of the mouse genome coupled with high-resolution expression analysis, *Nat. Biotechnol.* **21**, 652–659.
van Deursen, J., Wieringa, B. (1992) Targeting of the creatine kinase M gene in embryonic stem cells using isogenic and nonisogenic vectors, *Nucleic Acids Res.* **20**, 3815–3820.
Van Deursen, J., Fornerod, M., Van Rees, B., Grosveld, G. (1995) Cre-mediated site-specific translocation between nonhomologous mouse chromosomes, *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7376–7380.
van de Wetering, M., Oving, I., Muncan, V., Pon Fong, M.T. et al. (2003) Specific inhibition of gene expression using a stably integrated, inducible small-interfering-RNA vector, *EMBO Rep.* **4**, 609–615.
van der Putten, H., Botteri, F.M., Miller, A.D., Rosenfeld, M.G. et al. (1985) Efficient insertion of genes into the mouse germ line via retroviral vectors, *Proc. Natl. Acad. Sci. U. S. A.* **82**, 6148–6152.
Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W. et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.
Vicat, J.M., Boisseau, S., Jourdes, P., Laine, M. et al. (2000) Muscle transfection by electroporation with high-voltage and short-pulse currents provides high-level and long-lasting gene expression, *Hum. Gene Ther.* **11**, 909–916.
Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J. et al. (2002) Initial sequencing and comparative analysis of the mouse genome, *Nature* **420**, 520–562.
Zhang, Y., Buchholz, F., Muyrers, J.P., Stewart, A.F. (1998) A new logic for DNA engineering using recombination in Escherichia coli, *Nat. Genet.* **20**, 123–128.
Zhang, Y., Proenca, R., Maffei, M., Barone, M. et al. (1994) Positional cloning of the mouse obese gene and its human homolog, *Nature* **372**, 425–432.
Zheng, B., Sage, M., Cai, W.W., Thompson, D.M. et al. (1999) Engineering a mouse balancer chromosome, *Nat. Genet.* **22**, 375–378.

Websites (References Indicated as they Appear in the Original Text)

http://jaxmice.jax.org/info/recombinbred.html.
http://jaxmice.jax.org/library/notes/493c.html.
http://www.mshri.on.ca/nagy/cre.htm.
http://www.healthsystem.virginia.edu/internet/transgenic-mouse/southerndesign.cfm.

http://www.igtc.ca/index.html
http://www.escells.ca/
http://www.mmrrc.org/
http://tikus.gsf.de/project/web_new/index.html
http://www.sanger.ac.uk/PostGenomics/genetrap/
http://baygenomics.ucsf.edu/overview/welcome.html

Transgenic Plants for Food Use

Susanne Stirn and Horst Lörz
University of Hamburg, Hamburg, Germany

1 Methods to Establish Transgenic Plants 549
1.1 Transformation Methods 549
1.1.1 Agrobacterium Transformation 549
1.1.2 Direct Gene Transfer 550
1.2 Tissue Requirements 550
1.3 Molecular Requirements 551
1.3.1 Promoter 551
1.3.2 Codon Usage 551
1.3.3 Selectable Marker and Reporter Genes 551

2 GM Plants Already on the Market 554
2.1 Herbicide Resistance in Soybean, Maize, Oilseed Rape, Sugar Beet, Rice, and Cotton 554
2.2 Insect Resistance in Maize, Potatoes, Tomatoes, and Cotton 554
2.3 Virus Resistance, Male Sterility, Delayed Fruit Ripening, and Fatty Acid Contents in GMPs 554

3 GM Plants "In the Pipeline" 562
3.1 Input Traits 562
3.1.1 Insect Resistance 562
3.1.2 Virus, Fungal, Bacterial, and Nematode Resistance 563
3.1.3 Tolerance Against Abiotic Stress 566
3.1.4 Improved Agronomic Properties 567
3.2 Traits Affecting Food Quality 568
3.2.1 Increased Vitamin Content 568
3.2.2 Production of Very-long-chain Polyunsaturated Fatty Acids 569
3.2.3 Increased Iron Level 569
3.2.4 Improved Amino Acid Composition 569
3.2.5 Reduction in the Content of Antinutritive Factors 570

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

3.2.6 Production of "Low-calorie Sugar" 570
3.2.7 Seedless Fruits and Vegetables 570
3.3 Traits that Affect Processing 571
3.3.1 Altered Gluten Level in Wheat to Change Baking Quality 571
3.3.2 Altered Grain Composition in Barley to Improve Malting Quality 571
3.4 Traits of Pharmaceutical Interest 571
3.4.1 Production of Vaccines 571
3.4.2 Production of Pharmaceuticals 573

4 Outlook 574

Acknowledgment 574

Bibliography 574
Books and Reviews 574
Primary Literature 575

Keywords

B.t. Toxins
B.t. toxins are proteins produced by the soil bacterium *Bacillus thuringiensis*. The toxin recognizes certain receptors on the surface of insect midgut epithelial cells. A pore complex is formed through the membrane resulting in the loss of potassium ions (necessary for the osmotic balance). The insect will die due to massive water uptake.

Herbicide Resistance
By transferring the respective genes, plants can become tolerant to certain herbicides. Herbicide-resistant plants already on the market are mostly tolerant to the broad-spectrum herbicides glyphosate and glufosinate, which do not discriminate between certain plant species (as do selective herbicides).

Hypersensitive Reaction
Plants possess a natural defense mechanism against pathogens like fungi, bacteria, or viruses. During early infection, infected cells will be triggered to die (programmed cell death). Together with the plant cells, the pathogens die and their spread will be prevented. With the help of gene technology, the hypersensitive reaction shall be accelerated to make plants more resistant to pathogens.

Marker Gene
Marker genes are necessary to select early transformation events since plant transformation is of low efficiency (1–10% transformation success). Mostly, antibiotic resistance genes have been used as marker genes, since bacterial genes are readily available and antibiotics can be easily used as selective agents in culture media.

Transgenic or Genetically Modified Organism
A transgenic or genetically modified organism is an organism in which the genetic material has been modified in a way that does not occur naturally (e.g. by conventional breeding or natural recombination). Genetic modification involves moving, inserting, or deleting genes within or between species.

Here we describe the state of the art of genetically modified plants intended for food use. First, an introduction to the different methods of plant transformation as well as the molecular requirements for the stable introduction and expression of the gene(s) of interest is given. Then, we present an overview on the transgenic plants already on the market or approved for placing on the market worldwide. Finally, examples of new transgenic plants in different stages of development are presented. These comprise of well-established traits like herbicide or insect resistance, being transferred to additional crop species as well as new approaches to enhance the nutritional quality of food crops or to produce pharmaceuticals in transgenic plants.

1 Methods to Establish Transgenic Plants

As an introduction to transgenic or genetically modified plants (GMP), we would first like to describe the transformation methods that are currently used, as well as the molecular requirements for the stable introduction and expression of the gene(s) of interest.

1.1 Transformation Methods

In general, two different approaches to transferring foreign DNA into plant cells can be distinguished: (1) a vector-mediated transformation method (via *Agrobacterium)*; and (2) direct gene transfer methods.

1.1.1 Agrobacterium Transformation

In the first case, the natural ability of the bacterial phytopathogen *Agrobacterium* to transfer DNA into plant cells is exploited. *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes* are soil microorganisms that incite crown gall tumors and hairy root disease, respectively, in a wide range of dicotyledonous as well as in some monocotyledonous plants. For plant transformation, mainly *A. tumefaciens* is used, and this bacterium contains a large tumor-inducing plasmid (Ti). During infection, a specific segment of the plasmid DNA referred to as T-DNA (transferred DNA) is inserted into the plant genome. The T-DNA contains genes for phytohormones, which are responsible for cell proliferation and the formation of the crown gall, as well as genes for the formation of special nutrients (opines) in the plant cell. In order to use *Agrobacterium* as a tool in genetic engineering, these genes have been deleted and replaced by the genes of interest. This was possible since only the 25-bp T-DNA border sequences on the right and left border are needed for DNA transfer. A wide

variety of such "disarmed" (nontumor-inducing) vectors have been developed. *Agrobacterium*-mediated transformation is the most commonly used method for most dicotyledonous plants. Advantages include the typical insertion of only one or a few copies of the transgene, and the transfer of relatively large segments of DNA with only minimal rearrangements.

1.1.2 Direct Gene Transfer

For many years, *Agrobacterium* could not be used to transform monocotyledonous and other recalcitrant species, so direct gene transfer methods have been developed as an alternative. These include DNA uptake into protoplasts (protoplast transformation) and the shooting of DNA-coated particles into tissues (particle bombardment).

1.1.2.1 **Transformation of protoplasts.** In contrast to animal cells, plant cells possess a solid cell wall, which is the first barrier to overcome when foreign genes are to be transferred into them. One way to circumvent this barrier is to use plant cells in which the cell wall has been digested enzymatically, resulting in the so-called *protoplasts*. DNA uptake into protoplasts can then be stimulated by the use of either polyethylene glycol (PEG-transformation) or electric pulses (electroporation). Where an appropriate protoplast-to-plant regeneration system is available, large numbers of transformed clones can be regenerated to fertile transgenic plants.

In cases where a barrier to gene transfer has not been detected, virtually every protoplast system has proven transformable, though with considerable differences in the efficiency. However, problems can arise from the regeneration of fertile plants from protoplasts as the regeneration is strongly species- and genotype-dependent, and undesired somaclonal variation can occur due to the relatively long tissue culture phase. One advantage of the protoplast transformation method is its great independence from other patented techniques.

1.1.2.2 **Particle bombardment.** The new method of using high-velocity microprojectiles to deliver DNA into plant tissue was developed by Sanford and colleagues in 1987. A particle gun is used to accelerate DNA-coated microprojectiles into cells, past the cell wall and the cell membrane. The microprojectile is small enough ($0.5 - 5\ \mu m$) to enter the plant cell without causing too much damage, yet it is of large enough mass to penetrate the cell wall and carry an appropriate amount of DNA on its surface.

The main advantage of particle bombardment is the absence of biological incompatibilities that are found when using biological vectors. Organelles such as chloroplasts have also been transformed using particle bombardment. Unfortunately, particle bombardment, as well as other direct DNA uptake methods often result in complex insertion loci, which can cause gene silencing.

1.2 Tissue Requirements

The ability to cultivate plant tissue *in vitro* is a prerequisite in almost all current transformation protocols. Transformation requires competent (i.e. transformable) cultured cells that are embryogenic or organogenic. Plant cells suitable for regeneration are either cocultivated with *Agrobacterium* or bombarded. For *Agrobacterium* transformation, mostly leaf discs or immature embryos are used for particle

bombardment in monocotyledonous plant scutellar tissues.

1.3 Molecular Requirements

A typical plant gene consists of a promoter, a coding sequence, a transcription terminator, and a polyadenylation signal. The expression level of the gene is mainly determined by these components, but can also be affected by the surrounding sequences.

1.3.1 Promoter

The promoter is the main determinant of the expression pattern in the transgenic plant. Constitutive promoters direct expression in all or almost all tissues, independently of developmental or environmental signals. The promoter directing the synthesis of the cauliflower mosaic virus 35S RNA is the most frequently used constitutive promoter; other constitutive promoters are derived from agrobacterial T-DNA, such as the nopaline synthase (nos) promoter or the octopine synthase (ocs) promoter. Both are mainly used in dicotyledonous plants.

In monocotyledonous plants, it is mostly promoters from the rice actin 1 gene (*act1*) or the maize ubiquitin 1 gene (*ubi1*) that are used. To enhance expression levels, the first intron of the respective genes has been added to the expression cassettes. In dicotyledonous plants, the addition of introns seems to have less pronounced effects on the expression level. Nevertheless, the insertion of an intron is always necessary when the expression of the gene product in bacteria must be completely avoided (e.g. for genes conferring bacterial resistance; see Sect. 3.1.2).

In cases where transgene expression should be directed to certain tissues or developmental stages, regulated promoters are required. Promoter elements responsible for expression in seeds, tubers, vegetative organs, and leaves have been isolated.

Environmental influences can induce gene expression after wounding, heat- or cold stress, or anaerobiosis. The use of natural inducible promoters has the disadvantage of causing pleiotropic effects, since endogenous genes will also be turned on. Chemically induced promoters are favored for the production of pharmaceuticals in transgenic plants, as production of the desired compound can be restricted to a certain time point.

1.3.2 Codon Usage

Expression of the gene of interest can also be influenced at the level of translation: the codon usage differs between plants and bacteria. Therefore, the removal of rare codons, e.g. codons not frequently used in plant cells, from genes of bacterial origin can enhance expression considerably. This was first described for a modified *B. thuringiensis* toxin gene where expression was enhanced up to 100-fold.

1.3.3 Selectable Marker and Reporter Genes

In plant transformation systems, the efficacy of stable gene transfer is low; hence, systems that allow the selection of the transformed cells are required. A selection system consists of a selective agent (which interferes with the plant metabolism) and a selectable marker gene (which codes for a protein, enabling inactivation or evasion of the selective agent).

The most commonly used selectable marker genes in transgenic plants code for antibiotic or herbicide resistance including the following:

1.3.3.1 **Antibiotic resistance genes.**

- *nptII* gene: The *neo* or *nptII* gene, isolated from transposon Tn5 from *Escherichia coli K*12 codes for the neomycin phosphotransferase (NPTII). This enzyme detoxifies a range of aminoglycoside antibiotics such as neomycin, kanamycin, and geneticin. These antibiotics are added to the culture medium after the transformation procedure. Only transformed cells or tissues will survive and will be regenerated to plants.
- *hpt* gene: The *hpt* gene has been isolated from *E. coli* and codes for hygromycin phosphotransferase, which detoxifies the antibiotic hygromycin B. Most plant tissues show a higher sensitivity to hygromycin B than to kanamycin or geneticin. In particular, cereals that are resistant to kanamycin and geneticin can be selected with hygromycin.
- *bla* gene: The $bla_{\text{TEM-1}}$ gene codes for the TEM-1 β-lactamase, the most encountered ampicillin resistance marker in molecular biology. TEM-1 β-lactamase attacks narrow-spectrum cephalosporins and all the anti-gram-negative-bacterium penicillins, except temocillin. Mostly, the ampicillin resistance in transgenic plants is not used as a selectable marker for plant transformation, but is a reminiscence of the transformation method. With direct gene transfer methods, it is used as a selectable marker during the cloning procedure prior to plant transformation. When present on the plasmid used for transformation, the ampicillin resistance gene is transferred and integrated together with the genes of interest, and remains under the control of its prokaryotic promoters.
- *aadA*: The *aadA* gene confers resistance to streptomycin and spectinomycin. The gene has been found in association with several transposons (Tn7, Tn21, ...) and is ubiquitous among gram-negative bacteria.

1.3.3.2 **Herbicide resistance genes.**

- *bar*/pat gene: the *bar* gene from *Streptomyces hygroscopicus* and the *pat* gene from *Streptomyces viridichromogenes* code for phosphinotricin acetyltransferase (PAT), which confers resistance to the herbicide compound, phosphinotricin. It is an inhibitor of glutamine synthetase, a plant enzyme involved in ammonia assimilation. Besides phosphinotricin, the commercially available nonselective herbicides Basta and bialaphos can be used as selective agents. Selection can be applied in the culture medium or by spraying of regenerated plantlets.
- *EPSPS* gene: The *5-enolpyruvylshikimate-5-phosphate synthase* (EPSPS) gene codes for the enzyme 5-enolpyruvylshikimate-5-phosphate synthase (EPSPS) that is involved in the skimatic acid pathway of the aromatic amino acid biosynthesis. EPSPS is present in all plants, bacteria, and fungi, but not in animals. The nonselective herbicide glyphosate binds to and inactivates EPSPS. By transferring a glyphosate-resistant EPSPS gene from the soil bacterium *Agrobacterium* strain CP4 to plants, glyphosate-resistant crop plants have been obtained. Glyphosate is the active ingredient of the commercially available herbicide Roundup.

1.3.3.3 **Reporter genes.** In contrast to selectable marker genes, reporter genes do not confer resistance to selective agents inhibitory to plant development. Reporter genes code for products that can be detected directly or that catalyze reactions whose products are detectable.

- *gusA* gene: The *gusA* (uidA) gene encodes β-glucuronidase (GUS), which hydrolyzes a wide range of β-glucuronides. Substrates for GUS are available for spectrometric, fluorometric, and histochemical assays. The GUS reporter gene system allows easy quantification with high sensitivity. One disadvantage of this system is its nonviability: examined cells and tissues cannot be regenerated into transgenic plants. Therefore, this system is mainly used to optimize a specific transformation protocol.
- *luc* gene: The *luciferase* reporter gene assay uses bioluminescence reactions with substrate/enzyme combinations, which lead to detectable light emission. The *luc* gene originates from the firefly *Photinus pyralis*, and codes for a luciferase, which decarboxylates beetle luciferin. In contrast to the GUS reporter system, the luciferase reporter gene is nonlethal to plant cells.

1.3.3.4 **Alternative marker and excision systems.** Owing to discussions on the possible risk of a gene transfer of antibiotic resistance genes from transgenic plants to clinically important microorganisms, alternative marker systems have been developed. Additionally, through the use of excision systems the presence of antibiotic or herbicide resistance genes in the final product can be avoided.

Alternative markers Several alternative marker systems have recently been developed. For example, Chua and colleagues successfully selected transgenic plants using an inducible *Agrobacterium* gene encoding for isopentenyltransferase (IPT). This enzyme catalyzes the first step in biosynthesis of cytokinins, a class of phytohormones. Only cells containing this gene are able to form shoots and differentiate into mature plants.

As another alternative selection system, phosphomannose isomerase (PMI) was successfully used in transgenic sugar beet and maize. PMI catalyzes the interconversion of mannose-6-phosphate and fructose-6-phosphate. Except for leguminous plants, PMI is absent from most plants, and therefore, only transgenic plants expressing the *E. coli manA* gene can survive on media containing mannose as a carbon source.

Excision systems Another approach is the use of recombination systems that enable the excision of marker genes in the successfully transformed plant. Four different site-specific recombination systems have been shown to function in plant cells. The best-characterized system in plants is the CRE/lox system of bacteriophage P1: the CRE recombinase enzyme recognizes specifically and catalyzes recombination between two lox sequences. Marker genes flanked by these lox sites can be precisely excised in the presence of the CRE recombinase. One disadvantage of the system was that the CRE protein had to be introduced in the transgenic plant via secondary transformation or sexual crossing with a CRE transformant. When inducible developmental stage promoters or tissue-specific promoters are used, both components of the CRE/lox system can be transferred to one transformant. After selection of the transformants, the promoter will be intentionally induced, resulting in excision of the marker gene. This system is also applicable to vegetatively propagated plants.

A recent publication describes the efficient excision of selectable marker genes that are framed by *attP*-sites without the

induction of a recombinase. The molecular basis of the reaction is still unknown, but presumably it is due to illegitimate recombination. Further experiments are necessary to reveal the underlying mechanism and to verify its general applicability.

2 GM Plants Already on the Market

In Europe, 10 genetically modified (crop) plants have been approved for placing on the market according to 90/220 EEC. These approvals covered different uses as applied for by the applicant: in two cases, the approvals included use as food and feed, namely, Roundup Ready soybean (Monsanto) and insect-resistant maize (Novartis, formerly Ciba-Geigy). Nevertheless, only few commercial plantings are carried out in Europe, and in Germany GMPs are grown only on an experimental scale.

Since 1997, GMPs intended for food or feed use in the EU must be notified or approved according to the "Novel Food Regulation." Since then, refined oil from herbicide-tolerant oil seed rape, refined oil from herbicide-tolerant and insect-resistant cotton, and processed products from herbicide-resistant and insect-resistant maize have been notified (Table 1).

Worldwide, GMPs have been grown on 67.7 million hectares in 2003 according to ISAAA. Since 1996, the global area of transgenic crops increased 40-fold, with an increasing proportion grown by developing countries. The main producers of GMPs are the US (42.8 million hectares), Argentina (13.9 million hectares), Canada (4.4 million hectares), Brazil (3 million hectares), China (2.8 million hectares), and South Africa (0.4 million hectares).

Genetically modified soybeans occupied 41.4 million hectares; GM maize was planted on 15.5 million hectares, transgenic cotton on 7.2 million hectares, and GM oilseed rape on 3.6 million hectares.

Some 73% of the GMPs contain a herbicide resistance gene, followed by 18% carrying insect resistance genes, and 8% both herbicide- and insect-resistance genes (stacked genes).

2.1 Herbicide Resistance in Soybean, Maize, Oilseed Rape, Sugar Beet, Rice, and Cotton

As mentioned above, herbicide resistance is the leading trait in commercialized GMP, with 25 lines having been approved for cultivation and/or food and feed use worldwide (Table 2).

2.2 Insect Resistance in Maize, Potatoes, Tomatoes, and Cotton

Insect resistance is the trait found second most often in approved GMPs. Insect-resistant lines, as well as lines that acquire an insect resistance gene in combination with a herbicide resistance or virus-resistance gene, are listed in Table 3.

2.3 Virus Resistance, Male Sterility, Delayed Fruit Ripening, and Fatty Acid Contents in GMPs

GMPs with virus resistance (potato, squash, and papaya; Table 4), male sterility (oilseed rape, maize, and chicory; Table 5), delayed fruit ripening (tomato; Table 6) as well as modified fatty acid content (soybean, oilseed rape; Table 7) have also been approved in some countries.

Tab. 1 Genetically modified crop plants and products approved in the EU.

Plant (line)	*Trait (gene)*	*Applicant*	*Approved use*	*Year*
Soybean (GTS-40-3-2)	HR (*CP4 EPSPS*)	Monsanto	Import and processing	1996
Maize (Bt 176)	IR (*cryIAb*), HR (*bar*), ABR (*bla*)	Ciba-Geigy	Cultivation; import and processing	1997
Maize (Bt 11)	IR (*cryIAb*), HR (*pat*)	Northrup King	Import and processing; food use	1998; 2004
Maize (MON 810)	IR (*cryIAb*)	Monsanto	Cultivation; flour, semolina, starch and starch products, oil	1998; 1997
Maize (MON 809)	IR (*cryIAb*), HR (*CP4 EPSPS, gox*)	Pioneer	Food additives	1998
Maize (T 25)	HR (*pat*), ABR (partial *bla*)	AgrEvo	Cultivation; refined oil, starch and starch products, fermented or heat-treated products from maize flour	1998
Oil seed rape (MS1 × RF1; MS1 × RF2)	MS/RF (*barnase/barstar*), HR (*bar*), ABR (*npt*II)	PGS	Cultivation; refined oil	1997
Oil seed rape (MS8 × RF3)	MS/RF (*barnase/barstar*), HR (*bar*)	PGS	Refined oil	1999
Oil seed rape (Topas 19/2)	HR (*pat*), ABR (*npt*II)	AgrEvo	Import and processing (feed); refined oil	1995; 1997
Oil seed rape (Falcon GS 40/90)	HR (*pat*)	AgrEvo	Refined oil	1999
Oil seed rape (Liberator L62)	HR (*pat*)	AgrEvo	Refined oil	1999
Oil seed rape (GT 73)	HR (*CP4 EPSPS, gox*)	Monsanto	Refined oil	1997
Cotton (MON 1445/1698)	HR (*CP4 EPSPS*), ABR (*npt*II, *aad*)	Monsanto	Refined oil	2002
Cotton (MON 531/757/1076)	IR (*cryIAc*), ABR (*npt*II, *aad*)	Monsanto	Refined oil	2002

Notes: HR: herbicide resistance (glyphosate resistance *CP4 EPSPS* gene, glyphosate resistance oxidoreductase gene (*gox*), glufosinate resistance *pat* or *bar* gene);
ABR: antibiotic resistance (ampicillin resistance (*bla*), kanamycin resistance (*npt*II)), streptomycin resistance gene (*aad*);
IR: insect resistance (*Bacillus thuringiensis cryIA(b)*);
MS: male sterility (*barnase*);
RF: restorer of fertility (*barstar*).

Tab. 2 Herbicide-resistant crop plants and products worldwide.

Plant (line)	*Trait (gene)*	*Applicant*	*Country*
Soybean (GTS-40-3-2)	HR (*CP4 EPSPS*)	Monsanto	Argentina, Australia, Brazil, Canada, China, Czech Republic, EU, Japan, Korea, Mexico, Philippines, Russia, Switzerland, South Africa, UK, Uruguay, USA
Soybean (A 2704-12, A 2704-21, A 5547-35)	HR (*pat*)	AgrEvo	Canada, Japan, USA
Soybean (A 5547–127)	HR (*pat*), ABR (partial *bla*)	AgrEvo	USA
Soybean (GU 262)	HR (*pat*), ABR (partial *bla*)	AgrEvo	USA
Soybean (W 62, W 98)	HR (*bar*), RP (*gus*)	AgrEvo	USA
Maize (GA 21)	HR (*maize EPSPS*)	Monsanto	Argentina, Australia, Canada, China, Japan, Korea, Philippines, USA
Maize (NK 603)	HR (*CP4 EPSPS*)	Monsanto	Argentina, Australia, Canada, Japan, Philippines, South Africa, USA
Maize (MON 832)	HR (*CP4 EPSPS, gox*), ABR (*npt*II)	Monsanto	Canada
Maize (B 16 = DLL 25)	HR (*bar*), ABR (*bla*)	DeKalb	Canada, Japan, Philippines, USA
Maize (T 14, T 25)	HR (*pat*), ABR (partial *bla*)	AgrEvo	Argentina, Australia, Canada, EU, Japan, Philippines, USA
Oilseed rape (GT 73)	HR (*CP4 EPSPS, gox*)	Monsanto	Australia, Canada, China, EU, Japan, Philippines, USA

Oilseed rape (GT 200)	HR (*CP4 EPSPS, gox*)	Monsanto	Canada, USA
Oilseed rape (HCN 92 = Topas 19/2)	HR (*pat*), ABR (*npt*II)	AgrEvo	Canada, EU, Japan, USA
Oilseed rape (HCN 10)	HR (*pat*)	AgrEvo	Canada, Japan, USA
Oilseed rape (HCN 28 = T 45)	HR (*pat*)	AgrEvo	Australia, Canada, Japan, USA
Oilseed rape (Oxy 235)	HR (*bxn*)	Rhone Poulenc	Australia, Canada, Japan, USA
Brassica rapa (ZSR 500, 501, 502)	HR (*CP4 EPSPS, gox*)	Monsanto	Canada
Cotton (MON 1445/1698)	HR (*CP4 EPSPS*), ABR (*npt*II, *aad*)	Monsanto	Argentina, Australia, Canada, China, Japan, Philippines, South Africa, USA
Cotton (BXN)	HR (*bxn*), ABR (*npt*II)	Calgene	Australia, Canada, Japan, USA
Cotton (LL Cotton 25)	HR (*bar*)	Bayer Crop Science	Canada, USA
Cotton (19-51A)	HR (*als*)	DuPont	USA
Sugar beet (T120-7)	HR (*pat*), ABR (*npt*II)	AgrEvo	Canada, Japan, USA
Sugar beet (GTSB 77)	HR (*CP4 EPSPS, gox*)	Novartis	Australia, USA
Wheat (MON 71800)	HR (*CP4 EPSPS*)	Monsanto	USA
Rice (LLRICE 06, 62)	HR (*bar*)	AgrEvo	USA
Flax (FP967)	HR (*als*) ABR (*npt*II)	University of Saskatoon	Canada, USA

Notes: HR: herbicide resistance (glyphosate resistance *CP4/maize EPSPS* gene, glyphosate resistance oxidoreductase gene (*gox*), glufosinate resistance *pat* or *bar* gene, bromoxynil herbicide resistance nitrilase gene (*bxn*), sulfonylurea resistance *als* gene);
ABR: antibiotic resistance (ampicillin resistance gene *bla*, kanamycin resistance gene *npt*II, streptomycin resistance gene (*aad*));
RP: reporter gene (*gus*).

Tab. 3 Insect-resistant crop plants and products worldwide.

Plant (line)	*Trait (gene)*	*Applicant*	*Country*
Maize (MON 810)	IR (*cryIAb*)	Monsanto	Argentina, Australia, Canada, EU, Japan, South Africa, Switzerland, USA
Maize (MON 802)	IR (*cryIAb*), HR (*CP4 EPSPS, gox*), ABR (*npt*II)	Monsanto	Canada, Japan, USA
Maize (MON 80100)	IR (*cryIAb*), HR (*CP4 EPSPS, gox*), ABR (*npt*II)	Monsanto	USA
Maize (MON 863)	IR (*cryIIIBb2*), ABR (*npt*II)	Monsanto	Australia, Canada, Japan, Philippines, USA
Maize (MON 809)	IR (*cryIA(b)*), HR (*CP4 EPSPS, gox*)	Pioneer	Canada, Japan, USA
Maize (Bt 176)	IR (*cryIAb*), HR (*bar*), ABR (*bla*)	Ciba-Geigy	Argentina, Australia, Canada, EU, Japan, Switzerland, USA
Maize (Bt 11)	IR (*cryIAb*), HR (*pat*)	Northrup King (Syngenta)	Argentina, Australia, Canada, EU, Japan, Philippines, South Africa, Switzerland, UK, USA
Maize (DBT 418)	IR (*cryIAc*, protease inhibitor II (*pin*II)); HR (*bar*), ABR (*bla*)	DeKalb	Argentina, Australia, Canada, Japan, Philippines, USA
Maize (TC1507)	IR (*cryIFa2*), HR (*pat*)	Mycogen/ Pioneer	Canada, Japan, Philippines, South Africa, USA

Potato (Bt 6, Russet Burbank New Leaf)	IR (*cryIIIA*), ABR (*npt*II)	Monsanto	Canada, Japan, Philippines, USA
Potato (ATBT04-6 etc. 4 lines Atlantic and Superior New Leaf)	IR (*cryIIIA*), ABR (*npt*II)	Monsanto	Australia, Canada, Japan, Philippines, USA
Potato (SEMT 15–15, New Leaf Y)	IR (*cryIIIA*), VR (*coat protein PVY*) ABR (*npt*II, *aad*)	Monsanto	Australia, Canada, Philippines, USA
Potato (RBTM 21–350, Russet Burbank New Leaf Plus)	IR (*cryIIIA*), VR (*replicase PLRV, helicase PLRV*) ABR (*npt*II)	Monsanto	Australia, Canada, Japan, USA
Cotton (MON 531, 757, 1076, Bollgard)	IR (*cryIAc*), ABR (*npt*II, *aad*)	Monsanto	Argentina, Australia, Canada, China, India, Japan, Mexico, Philippines, South Africa, USA
Cotton (MON 15985, Bollgard II)	IR (*cryIAc, cryIAb*), ABR (*npt*II, *aad*), RP (*uidA*)	Monsanto	Australia, Canada, Japan, Philippines, USA
Cotton (31807/31808)	IR (*cryIAc*), HR (*bxn*) ABR (*npt*II)	Calgene	Japan, USA
Tomato (5345)	IR (*cryIAc*), ABR (*npt*II, *aad*)	Monsanto	Canada, USA

Notes: IR: insect resistance (delta endotoxin genes of *Bacillus thuringiensis* (*cryIAb, cryIAc, cryIIIA, cryIFa2*; protease inhibitor II (*pin*II));
HR: herbicide resistance (glyphosate resistance *CP4 EPSPS* gene, glyphosate resistance oxidoreductase gene (*gox*), glufosinate resistance *pat* or *bar* gene, bromoxynil herbicide resistance nitrilase gene (*bxn*));
VR: virus resistance (potato virus Y (PVY) coat protein gene, replicase, and helicase gene of potato leafroll virus (PLRV), respectively);
ABR: antibiotic resistance (ampicillin resistance (*bla*), kanamycin resistance (*npt*II), streptomycin resistance (*aad*));
RP: reporter gene (β-D-glucuronidase (*gus*)).

Tab. 4 Virus-resistant crop plants and products worldwide.

Plant (line)	*Trait (gene)*	*Applicant*	*Country*
Potato (SEMT 15–15, New Leaf Y)	VR (*coat protein PVY*), IR (*cryIIIA*), ABR (*npt*II, *aad*)	Monsanto	Australia, Canada, Philippines, USA
Potato (RBTM 21–350, Russet Burbank New Leaf Plus)	VR (*replicase PLRV, helicase PLRV*), IR (*cryIIIA*), ABR (*npt*II)	Monsanto	Australia, Canada, Japan, USA
Squash (CZW-3)	VR (*coat protein CMV, coat protein ZYMV, coat protein WMV 2*), ABR (*npt*II)	Asgrow	Canada, USA
Squash (ZW20)	VR (*coat protein ZYMV, coat protein WMV 2*)	Upjohn	Canada, USA
Papaya (55-1/63-1)	VR (*coat protein PRSV*), RP (*gus*), ABR (*npt*II)	Cornell University	Canada, USA

Notes: VR: virus resistance (coat protein genes of potato virus Y (PVY), cucumber mosaic virus (CMV), zucchini yellows mosaic virus (ZYMV) and watermelon mosaic virus (WMV) 2, respectively; replicase and helicase genes of potato leafroll virus (PLRV));
IR: insect resistance (delta endotoxin gene of *Bacillus thuringiensis* (cryIIIA));
ABR: antibiotic resistance (kanamycin resistance (*npt*II), streptomycin resistance (*aad*));
RP: reporter gene (*gus*).

Tab. 5 Male sterility in crop plants worldwide.

Plant (line)	*Trait (gene)*	*Applicant*	*Country*
Oilseed rape (MS1, RF1 (PGS1))	MS (*barnase*), FR (*barstar*), HR (*bar*), ABR (*npt*II)	PGS	Australia, Canada, EU, Japan, USA
Oilseed rape (MS1, RF2 (PGS2))	MS (*barnase*), FR (*barstar*), HR (*bar*), ABR (*npt*II)	PGS	Australia, Canada, EU, Japan, USA
Oilseed rape (MS8× RF3)	MS (*barnase*), FR (*barstar*), HR (*bar*)	PGS	Australia, Canada, Japan, USA
Oilseed rape (PHY14, PHY35)	MS (*barnase*), FR (*barstar*), HR (*bar*)	PGS	Japan
Oilseed rape (PHY36)	MS (*barnase*), FR (*barstar*), HR (*bar*)	PGS	Japan
Maize (MS3)	MS (*barnase*), HR (*bar*), ABR (*bla*)	PGS	Canada, USA
Maize (MS6)	MS (*barnase*), HR (*bar*), ABR (*bla*)	PGS	USA
Maize (676, 678, 680)	MS (*dam*), HR (*pat*)	Pioneer	USA
Chicory (RM3-3, RM3-4, RM3-6)	MS (*barnase*), HR (*bar*), ABR (*npt*II)	Bejo Zaden	EU, USA

Notes: MS: male sterility (*barnase* from *Bacillus amyloliquefaciens, dam* (DNA adenine methylase from *E. coli*);
FR: fertility restoration (*barstar* from *Bacillus amyloliquefaciens*);
HR: herbicide resistance (glufosinate resistance *bar* gene);
ABR: antibiotic resistance (ampicillin resistance (*bla*), kanamycin resistance (*npt*II)).

Tab. 6 Delayed fruit ripening in tomatoes and tomato products worldwide.

Plant (line)	*Trait (gene)*	*Applicant*	*Country*
Tomato (FLAVR SAVR)	DR (*PG*, antisense) ABR (*npt*II)	Calgene	Canada, Japan, Mexico, USA
Tomato (B, Da, F)	DR (*PG*, sense or antisense) ABR (*npt*II)	Zeneca	Canada, USA
Tomato (8338)	DR (*ACCd*) ABR (*npt*II)	Monsanto	USA
Tomato (35 1 N)	DR (*sam*-K) ABR (*npt*II)	Agritope	USA
Tomato (1345-4)	DR (*ACC*, sense) ABR (*npt*II)	DNA Plant Technology Corporation	Canada, USA

Notes: DR: delayed ripening (polygalacturonase gene in sense or antisense direction (*PG*); 1-amino-cyclopropane-1-carboxylic acid deaminase gene (*ACCd*); *S*-adenosylmethionine hydrolase gene with Kozak consensus sequence (*sam*-K); aminocyclopropane cyclase synthase gene in sense orientation (*ACC*));
ABR: antibiotic resistance (kanamycin resistance (*npt*II)).

Tab. 7 Modified fatty acid content in crop plants and products worldwide.

Plant (line)	*Trait (gene)*	*Applicant*	*Country*
Soybean (G94-1, G94-19, G168, high-oleic soybean)	MFA (*GmFad2-1*), ABR (*bla*), RG (*gus*)	DuPont	Australia, Canada, Japan, USA
Oilseed rape (23-18-17, 23–198, high-laurate canola)	MFA (*BayTE*), ABR (*npt*II)	Calgene	Canada, USA

Notes: MFA: modified fatty acid content (delta 12 desaturase gene (*GmFad2-1*); thioesterase gene from *Umbellaria californica* (*Bay TE*);
ABR: antibiotic resistance (ampicillin resistance (*bla*), kanamycin resistance (*npt*II));
RP: reporter gene (*gus*).

3 GM Plants "In the Pipeline"

3.1 Input Traits

In the following chapters, selected examples will be presented of studies being conducted in the genetic engineering of crop plants. First, attempts to reduce the production costs of crop plants are summarized. These studies refer to crop plants with genetically engineered resistance against insects, diseases, abiotic stresses, and improved agronomic properties.

3.1.1 Insect Resistance

The best-characterized insecticidal proteins are the delta-endotoxins of *B. thuringiensis* (*B.t.*-toxins), which have been used as biopesticides in agriculture (also organic farming), forestry, and as a mosquito vector control for many years.

B.t. insecticidal activity is highly specific in that the endotoxins are nontoxic to nontarget insects, birds, and mammals.

Besides the commercially available *B.t.*-maize, -potato, -cotton and -tomato lines (see Table 3), insect-resistant rice (*cryIAb* and *cryIAc*), soybeans (*cryIAc*), oilseed rape (*cryIAc*), and eggplants (*cryIIIB*) have been developed. Field trials have also been performed in the United States with insect-resistant sunflower, lettuce, grapefruit, sugarcane, apple, walnut, grape and peanut.

In order to control corn rootworm infections, Mycogen, and Pioneer Hi-Bred are developing transgenic maize, which contains two novel proteins from *B. thuringiensis* strain PS149B1. These proteins belong to another class of insecticidal proteins with no homology to the delta-endotoxins (Bt PS149B1 toxins).

Another approach to achieve insect-resistant plants is to use plant defense proteins such as proteinase inhibitors, lectins or α-amylase inhibitor. Lectins are thought to confer resistance toward sap-sucking insects (e.g. the rice brown planthopper), which act as vector for virus transmission.

Morton and colleagues transferred the α-amylase inhibitor 1 gene from beans to peas and conducted field trials with the insect-resistant pea plants, and reported high insect mortality in transgenic plants. One advantage of using α-amylase inhibitor genes is the long history of human consumption of beans, and the fact that bean α-amylase has no effect on starch digestion in humans.

In an attempt to confer broad-spectrum resistance to storage pests, maize was transformed with the chicken avidin gene. The mode of action of avidin is to cause a deficiency of the vitamin biotin in insects; hence, a thorough safety testing for human consumption must be carried out before the commercial use of biotin maize.

3.1.2 Virus, Fungal, Bacterial, and Nematode Resistance

3.1.2.1 **Virus resistance.** One strategy to obtain virus-resistant plants is to transfer genes from the pathogen itself into the plant (pathogen-derived resistance). The most widely used approach is to express the virus coat protein in transgenic plants. In theory, the expression of viral genes disrupts viral infection or symptom development. All but one of the commercially available virus-resistant plants contain viral coat proteins (see Table 4), and this technique is extended to other plants such as rice, plum tree, tomato, pea, and peanut. Additionally, field trials have been performed in the United States with coat protein–mediated virus-resistant wheat, soybean, sugarcane, sugar beet, cucumber, sweet potato, grapefruit, pineapple, and papaya.

Another form of pathogen-derived resistance is the use of viral replicase genes (or RNA-dependent RNA polymerase genes), which presumably act via posttranscriptional gene silencing. This technique has been used to confer resistance to potato leafroll virus in potato (see Table 4), to barley yellow dwarf virus in oats, cucumber mosaic virus in tomato, rice tungro spherical virus in rice, and wheat streak mosaic virus in wheat.

Since varying degrees of virus resistance have been obtained with coat protein–mediated resistance, approaches have been formulated to ameliorate resistance against cucumber mosaic virus via satellite RNA, especially in tomato. However, this approach has raised controversy

as a single point mutation in the satellite RNA can transform it into a harmful necrogenic form.

In order to protect plants against more than one virus, ribosome-inactivating proteins (RIPs) have been expressed in transgenic plants. RIPs are strong inhibitors of protein synthesis and, depending on the plant species from which they originate, they show different levels of toxicity against different hosts. Pokeweed antiviral protein (PAP) confers resistance to PVX and PVY in transgenic potatoes, while PA-PII confers resistance to TMV, PVX, and fungal infections in tobacco, respectively.

On a more experimental scale are approaches to achieve virus resistance using antibodies against the virus coat protein. Such antibodies are able to neutralize virus infection, presumably by interacting with newly synthesized coat protein and disrupting viral particle formation. Like RIPs, broad-spectrum antibodies might be used to protect plants against a wider range of viruses, as has been demonstrated in the case of potyviruses.

3.1.2.2 **Fungal resistance.** Fungal resistance can be conferred via the activation of specific self-defense mechanisms in the plant. One of the mechanisms is the so-called *hypersensitive response* (HR), which enables plants to enclose the pathogen in the infected area by the formation of necrotic lesions. HR induces many defense-related signal molecules such as salicylic acid, ethylene, and phytoalexin. HR is also characterized by an accumulation of pathogenesis-related (PR) proteins that include fungal cell wall–degrading enzymes, antimicrobial peptides, thionins, lipid transfer proteins, and proteinase inhibitors.

In rice, the introduction of chitinase and thaumatin-like protein, respectively, led to increased resistance to sheath blight (*Rhizoctonia solani*). Enhanced resistance to the rice blast fungus *Magnaporthe grisea* was observed upon constitutive expression of chitinase and defense-related protein genes in transgenic rice.

Pathogenesis-related proteins from plants have been used to confer fungal resistance in alfalfa, cucumber, oilseed rape, tomatoes, wheat, grape vine, and oranges.

Other antifungal genes of plant origin comprise genes for RIPs, genes for phytoalexins, and anthocyanin genes. An example of an antifungal gene from nonplant sources that has been transferred to plants is the human lysozyme gene.

In the United States, field trials have been performed with fungal-resistant wheat, barley, cotton, rice, and bananas using different antifungal proteins.

Individual PR-proteins, however, have a narrow spectrum of antifungal activity, and need to function collectively in order to provide a modest but long-term resistance. Therefore, research is currently focusing on genes from mycoparasitic fungi as a source for improving resistance to fungal pathogens. An endochitinase of the mycoparasitic fungus *Trichoderma harzianum* has been transferred to tobacco and potato, and has been reported to confer a high level and a broad spectrum of resistance.

A similar approach has been taken by transferring an antifungal protein from a virus that persistently infects *Ustilago maydis*, to wheat. Transgenic wheat plants exhibit increased resistance against stinking mut (*Tilletia tritici*).

3.1.2.3 **Bacterial resistance.** Resistance to bacterial infections is not yet as

developed as virus- and fungal resistance, due partly to the fact that bacterial diseases are a major problem in only some crop plants such as potato, tomato, rice, and certain fruit trees.

The most efficient form of protection is genetic resistance, which is based on single dominant or semidominant genes. These R genes usually confer race-specific resistance, and their effectiveness is based on their interaction with complementary pathogen avirulence (AV) genes in the pathogen, the so-called *gene-for-gene interaction*. Resistance to bacterial blight caused by *Xanthomonas oryzae* pv. *oryzae* was achieved by transferring the resistance gene *Xa21* from a wild rice species to the elite indica rice (IR) variety 'IR72'. Accordingly, the resistance gene *Bs2* from pepper was transferred to tomato, which then exhibited resistance to bacterial spot disease.

The tomato disease resistance gene *Pto* specifies race-specific resistance to *Pseudomonas syringae* pv. *tomato* carrying the *avrPto* gene. By overexpressing *Pto,* a race-nonspecific resistance was observed in transgenic tomatoes.

The resistance based on single dominant gene expression always bears the danger of early evolution of counterresistance in the pathogen due to the emergence of strains that no longer express the specific avirulence gene product. Therefore, new resistance genes are being investigated for use in pyramiding strategies (combination of resistance genes against the same pathogen, but with different targets). One example is the *AP1* gene from sweet pepper, which delays the hypersensitive response when expressed in transgenic rice plants, and which can be used in combination with *Xa21* or other resistance genes.

In several plant species, bifunctional enzymes with lysozyme activity have been detected, which are thought to be involved in defense against bacteria. After transfer of the bacteriophage T4 lysozyme gene, transgenic potatoes showed a decreased susceptibility toward *Erwinia carotovora atroseptica* infections.

Insects produce antimicrobial peptides as major defense response to pathogen attack, and these include sarcotoxins, cecropins, and attacins. The latter has been transferred to apples and pears, which have improved resistance to *Erwinia amylovora*, causing fire blight. Similarly, the expression of synthetic antimicrobial peptide chimeras in transgenic tobacco led to broad-spectrum resistance against both bacterial and fungal pathogens. Bacterial resistant grapes, transformed with the antibiotic protein mangainin from toads are being field tested in the United States.

3.1.2.4 **Nematode resistance.** In the United States, field tests are being conducted with nematode-resistant carrots, tomatoes, potatoes, and pineapples. In the case of carrot, tomato, and potato, cysteine proteinase inhibitors from cowpea and rice, respectively, have been expressed in the transgenic plants. Proteinase inhibitors are an important element of natural plant defense strategy. The cowpea trypsin inhibitor (CpTI), a serine proteinase inhibitor, and oryzacystatin (Oc-I), an inhibitor of cysteine proteinase, have been shown to be effective against proteinases of a cyst nematode. In field trials, no detrimental effect on plant growth was observed in transgenic potatoes. When two partial nematode-resistant potato varieties were transformed with the cystatin, full resistance to potato cyst nematode *Globodera* was achieved. Cysteine proteinase inhibitors are of particular interest because they are the only class of proteinase that is not expressed in the digestive system of mammals.

3.1.3 Tolerance Against Abiotic Stress

The effects of weather, erosion, and depleted soils expose plants to a variety of stresses. Genetic engineering has been used to provide plants with additional stress response genes in order to counteract these environmental stresses. These genes can be grouped into two categories: (1) genes that respond directly against a particular stress; and (2) genes that regulate stress gene expression and signal transduction. Transfer or overexpression of both types of genes has been used in transgenic approaches to enhance tolerance against abiotic stresses. In the following section, salt, drought, and cold stress are considered together, as, on a cellular basis, all act as dehydration stress.

3.1.3.1 **Genes conferring dehydration-stress tolerance in transgenic plants.** Plants react to these stresses by displaying complex, quantitative traits that involve the function of many genes. Expression of these genes leads to the accumulation of low molecular weight components such as osmolytes, synthesis of late-embryogenesis-abundant (LEA) proteins, and activation of detoxifying enzymes. In transgenic approaches, genes for the enzymes that are responsible for the production of these compounds have been transferred to nontolerant plants. Most of these studies have been performed in model plants such as *Arabidopsis* or tobacco, but some investigations have been extended to food crops such as rice.

Bajaj and colleagues have summarized the results from these studies in a recent review. After transfer of genes encoding enzymes for osmoprotectants such as glycinebetaine, proline or putrescine, transgenic rice plants exhibited increased tolerance to salt and drought stress. Constitutive overexpression of the oat arginine decarboxylase gene in rice led to severely affected developmental patterns. When the arginine decarboxylase gene was under the control of an ABA-inducible promoter, the transgenic rice plants showed an increase in biomass under salinity-stress conditions.

Transgenic rice plants expressing the arginine decarboxylase gene (*adc*) from *Datura* produced high levels of the polyamine putrescine under stress, protecting the plants from drought.

The expression of a LEA protein in the transgenic rice plants led to increased tolerance to water deficit and salt stress in the transgenic plants. As it is believed that abiotic stress primarily affects plants through oxidative damage, genes for detoxifying enzymes have been transferred to sensitive plants. McKersie et al. showed that transgenic alfalfa overexpressing Mn superoxide dismutase to reduce free radicals was more tolerant of water deficit and freezing, and had better winter survival rates. It is envisaged that the use of stress-inducible promoters and the introduction of multiple genes will improve dehydration-stress tolerance.

3.1.3.2 **Regulatory genes encoding transcription factors.** The other promising approach to confer tolerance to dehydration stress is to transfer regulatory genes. As the products of these genes regulate gene expression and signal transduction under stress conditions, their overexpression can activate the expression of many stress-tolerance genes simultaneously. The overexpression of the transcription factor DREB1A in transgenic *Arabidopsis* led to increased tolerance to drought, salt, and freezing stresses. However, the constitutive overexpression of DREB1A also resulted in severe growth retardation under normal growth conditions.

In contrast, the stress-inducible expression of this gene had minimal effects on plant growth, and provided greater tolerance to stress conditions than genes driven by the 35S promoter. Drought tolerant wheat lines transformed with DREB1A are currently being field tested in Mexico.

Lee and colleagues cloned the functional homolog of the yeast Dbf2 kinase that enhances salt, drought, cold, and heat tolerance upon overexpression in yeast, as well as transgenic plant cells. However, the utilization of this gene to engineer transgenic crops with enhanced stress tolerance has still to be shown.

3.1.3.3 **Additional genes that confer tolerance to salt stress.** The detrimental effects of salt on plants are a consequence of both a water deficit resulting in osmotic stress, and the effects of excess sodium ions on key biochemical processes. To tolerate high levels of salts, plants should be able to use ions for osmotic adjustment and internally to distribute these ions to keep sodium away from the cytosol. The first transgenic approaches introduced genes that modulated cation transport systems. Hence, transgenic tomato plants overexpressing a vacuolar Na^+/H^+ antiport were able to grow, flower, and produce fruits in the presence of 200 mM sodium chloride.

3.1.3.4 **Genes conferring tolerance to low iron.** In arid and semiarid regions of the world, the soils are alkaline in nature and therefore, crop yields are limited by a lack of available iron. Under iron stress, some plants release specific Fe(III)-binding compounds, known as siderophores, which bind the otherwise insoluble Fe(III) and transport it to the root surface. To increase the quantity of siderophores released under conditions of low iron availability, Takahashi et al. transferred two barley genes coding for the enzyme nicotianamine aminotransferase (*naat-A* and *naat-B*) together with the endogenous promoters into rice plants. Transgenic rice plants excreted about 1.8 times more siderophore under iron stress than wild-type rice plants. This relatively small increase allowed transgenic rice plants to withstand iron deprivation remarkably better, resulting in a fourfold increase in grain yield as compared with wild-type plants.

Alternately, increasing the rate-limiting step of Fe(III) chelate reduction, which reduces iron to the more soluble Fe(II) form, might enhance iron uptake in alkaline soils.

3.1.3.5 **Aluminum tolerance.** Aluminum toxicity is one of the major factors limiting crop productivity in acidic soils, which comprise about 40% of the world's arable land. One of the most visible symptoms of aluminum toxicity is the inhibition of root elongation. One possible mechanism of aluminum tolerance is the chelation of aluminum by organic ions like citrate or malate in the rhizosphere or within root cells. Overproduction of citrate was shown to result in aluminum tolerance in transgenic tobacco, papaya, and oilseed rape. Expression of the *ALMT1* gene (aluminum-activated malate transporter (ALMT)) from wheat in transgenic barley plants resulted in high-level aluminum tolerance.

3.1.4 Improved Agronomic Properties

3.1.4.1 **Acceleration of sprouting time in potato.** The ability to control sprouting time in potato tubers is of considerable

economic importance to the potato industry. At present, the potato industry uses a range of chemical treatments in order to obtain the desired control. Using a biotechnological approach, the *pyrophosphatase* gene from *E. coli* under the control of the tuber-specific patatin promoter was transferred into potatoes. The *pyrophosphatase* gene was chosen because of the central role of inorganic phosphate in starch degradation and sucrose biosynthesis. It is believed that starch breakdown and subsequent formation of various metabolites is needed for growth of the sprout.

Transgenic potatoes displayed a significantly accelerated sprouting: the transgenic tubers sprouted on average six to seven weeks faster than did control tubers. In addition, after cold storage the majority of transgenic tubers sprouted within one week, whereas the wild-type tubers needed eight weeks or more.

3.1.4.2 **Reduction of generation time in citrus.** Citrus trees have a long juvenile phase that delays their reproductive development by between 6 and 20 years. With the aim of accelerating flower time, juvenile citrus seedlings were transformed with the *Arabidopsis LEAFY* (*LFY*) and *APETALA* (*AP1*) genes, which promote flower initiation in *Arabidopsis*. Both types of transgenic citrus produced fertile flowers as early as the first year. These traits are submitted to the offspring as dominant traits, generating trees with a generation time of one year from seed to seed. Constitutive expression of *LFY* also promoted flower initiation in transgenic rice.

3.2 Traits Affecting Food Quality

The following sections provide examples of how the nutritional value of food crops can be improved through genetic engineering, though none of these genetically engineered food plants has yet reached the marketplace. In addition to the safety of the GMPs for health and environment, the effectiveness of this approach in comparison to alternative methods must be shown for these traits.

3.2.1 Increased Vitamin Content

One of the most advanced projects to fortify food crops with vitamins is the so-called *golden rice*. To enable provitamin A biosynthesis in the rice endosperm, four additional enzymes are required. Immature rice embryos were cotransformed with two *Agrobacterium* constructs containing the *psy* and the *lcy* gene from daffodil, coding for the phytoene synthase and the lycopene β-cyclase, respectively, as well as the *crtI* gene from *Erwinia uredovora*, coding for a bacterial phytoene desaturase. Ten plants harboring all transferred genes were recovered, and all showed a normal vegetative phenotype and were fertile. In most cases, the transformed endosperms were yellow, indicating carotenoid formation. After transfer of the new trait into locally best-adapted varieties, either by traditional breeding or *de novo* transformation, it is hoped that vitamin-A deficiency in the developing countries can be prevented.

In contrast to rice, tomato plants already produce carotenoids, and tomato products are viewed as the principal dietary source of lycopene and one of the major sources of β-carotene. Since lycopene and β-carotene are considered beneficial to health (reducing chronic conditions such as coronary heart disease and certain cancers), an attempt was made to enhance the carotenoid content and profile of tomato fruits via genetic engineering. Therefore, the bacterial carotenoid gene

(*crtI*), which encodes for the phytoene desaturase, has been introduced into tomato plants. Transgenic plants showed an increased β-carotene content (up to 45% of the total carotenoid content), though the total carotenoid content was not elevated.

The regular uptake of vitamin C (ascorbic acid) is essential for humans since they lack the ability to synthesize it. In plants, ascorbic acid can be regenerated from its oxidized form in a reaction catalyzed by dehydroascorbate reductase (DHAR). The overexpression of a DHAR cDNA from wheat in maize and tobacco led to a two- to fourfold increase in ascorbic acid levels in foliar and kernels.

3.2.2 Production of Very-long-chain Polyunsaturated Fatty Acids

Very-long-chain polyunsaturated fatty acids (VLCPUFAs) like arachidonic, eicosapentaenoic, or docosahexaenoic acid have important roles in human health and nutrition. They are components of membrane phospholipids, precursors of prostaglandins, required for the development of the fetal neuronal system and reported to reduce the incidence of cardiovascular diseases. The diet of most modern societies is relatively low in ω 3-PUFAs with a concomitant increased level of ω 6-PUFA intake. An important source of VLCPUFAs is oily fish, whereas all plant oils do not contain VLCPUFAs. Now first results have been obtained to genetically engineer the capacity to synthesize these fatty acids in oilseed species. The seed-specific expression of cDNAs encoding fatty acid acyl-desaturases and elongases resulted in the accumulation of up to 5% of VLCPUFAs in linseed.

3.2.3 Increased Iron Level

As cereal grains are deficient in certain essential mineral nutrients, including iron, several approaches have been used to increase iron accumulation and alter iron metabolism. Since ferritin is a general iron storage protein in all living organisms, ferritin genes have been introduced into rice and wheat plants. Goto et al. generated transgenic rice plants expressing soybean ferritin under the control of the seed-specific rice Glu-B1 promoter. Transgenic rice seeds accumulated up to three times more iron as wild-type seeds. Likewise, increased iron levels were found in transgenic rice seeds expressing bean ferritin under the control of the related Gt-1 promoter.

In order to not only increase iron accumulation, but also improve its absorption in the human intestine, two approaches have been adopted. First, the level of the main inhibitor of iron absorption, phytic acid, was decreased by the introduction of a heat-tolerant phytase from *Aspergillus fumigatus*. Transgenic rice plants exhibited at least double phytase activity, whereas in one individual transgenic line, the phytase activity of the grains increased about 130-fold. Second, as cysteine peptides are considered a major enhancer of iron absorption, the endogenous cysteine-rich metallothionein-like protein gene (*rgMT*) was overexpressed, and the cysteic acid content increased significantly in transgenic seeds. The authors suggested that high-phytase rice, with increased iron content and rich in cysteine-peptide, has the potential to greatly improve iron supply in rice-eating populations.

3.2.4 Improved Amino Acid Composition

Potatoes contain limited amounts of the essential amino acids lysine, tryptophan, methionine, and cysteine. In order to improve the nutritional value of potatoes, a nonallergenic seed albumin gene (*AmA1*) from *Amaranthus hypochondriacus* was

transferred to potato plants. The seed-specific albumin was under the control of a tuber-specific and a constitutive promoter, respectively. In both transgenic lines, a 35 to 45% increase in total protein content was reported in transgenic tubers, which corresponded to an increase in most essential amino acids. Additionally, a twofold increase in tuber number and a 3.0- to 3.5-fold increase in tuber yield were observed.

3.2.5 Reduction in the Content of Antinutritive Factors

Cassava is one of the few plants in nature that contains toxic cyanogenic glycosides in the leaves and roots. Sufficient processing of the harvested roots normally renders the cassava safe, although the processing that renders toxic cassava varieties safe also removes certain nutritional value. Therefore, Moller et al. have isolated the genes responsible for cyanogenic glycoside production and are now transforming cassava with antisense constructs of the respective genes *CYP79D1* and *CYP79D2*.

Another strategy for reducing the cyanide toxicity of cassava roots is to introduce a gene that codes for the enzyme hydroxynitrile lyase (HNL). This enzyme breaks down the major cyanogen acetone cyanohydrin, and is expressed only in leaves. After transformation of the cDNA encoding HNL, the HNL activities in cassava roots were comparable with those in leaves. Field trials will determine whether the expression of HNL in roots in fact reduces the cyanide toxicity of cassava food products.

3.2.6 Production of "Low-calorie Sugar"

Koops and colleagues have developed a new sugar beet that produces fructan, a low-calorie sweetener, by inserting a single gene from Jerusalem artichoke that encodes an enzyme for converting sucrose to fructan (1-sucrose:sucrose fructosyl transferase [1-sst]). Short-chain fructans have the same sweetness as sucrose, but provide no calories as humans lack the fructan-degrading enzymes necessary to digest them. Longer-chain fructans form emulsions having a fatlike texture, while fructans also promote the growth of beneficial bacteria present in the gut. Transgenic sugar beet roots produce the same amount of total sugar, but expression of the *1-sst* gene resulted in conversion of more than 90% of the stored sucrose into fructans. Under greenhouse conditions, the "fructan beets" developed normally and had almost the same amount of root dry weight as normal sugar beets. The yield of 110 μmol g^{-1} fresh weight of fructan makes the extraction of these compounds economically interesting.

3.2.7 Seedless Fruits and Vegetables

In plants that are able to develop fruits without fertilization (parthenocarpic fruits), the seeds are absent – a feature that can increase fruit acceptance by consumers. To achieve parthenocarpic development, it is common practice to treat flower buds with synthetic auxinic hormones. To mimic the hormonal effects by genetic engineering, the expression of a gene able to increase auxin content and activity should be induced in the ovule. Rotino and colleagues used the *iaaM* gene from *P. syringae* pv. *savastanoi* under the control of an ovule-specific promoter (*DefH9*) from *Antirrhinum majus* to induce parthenocarpic development in transgenic tobaccos and eggplants. In transgenic eggplants, fruit setting took place even under environmental conditions prohibitive for the untransformed line. Fruit size and weight were similar to that obtained by pollination in transgenic

and control plants. It was envisaged that this approach might also be valuable in other horticultural species such as pepper and tomatoes, in which fruit quality is susceptible to uneven pollination and seed set during lower temperature fluctuations associated with "winter season" production. A similar approach might also be used to produce seedless watermelons rather than the cumbersome triploid seed production system now employed.

3.3 Traits that Affect Processing

3.3.1 Altered Gluten Level in Wheat to Change Baking Quality

The bread-making quality of wheat flour depends primarily on the presence of high molecular weight (HMW) glutenins. Wheat is unique among cereals in having this property. The glutenin proteins are encoded by six genes, and the total glutenin content of the grain is proportional to the expression of these genes. It has been shown that the quality of dough can be influenced both by the quantity and quality of the expressed genes. When specific glutenin genes were added back to wheat lines missing some of the glutenin genes, the dough-mixing characteristics were improved significantly. This same technique could be applied to wheat lines that have already been optimized for baking, and could result in flour with glutenin levels higher than the current maximum of 10% of total protein. Concordantly, the quality of durum wheat (*Triticum turgidum* L. var. *durum*) for bread and pasta making has been modified by insertion of HMW glutenin subunit genes.

3.3.2 Altered Grain Composition in Barley to Improve Malting Quality

β-Glucan is the major constituent of the cell wall of the starchy endosperm of barley. These cell wall molecules are very large, water-soluble, and produce viscous worts if not sufficiently reduced in size by hydrolysis. This causes slow filtration of wort and beer, as well as glucan precipitate in the finished product. β-Glucan hydrolysis is likely to be a function of the level of (1–3, 1–4)-β-glucanase produced by the aleurone, and how much of the enzyme survives high-temperature kilning and mashing. Doubling the amount of β-glucanase activity is likely to ensure sufficient β-glucan hydrolysis, and this could be achieved either by increasing the amount of enzyme synthesized or by changing the heat stability of β-glucanase. The latter has been achieved by transferring a fungal thermotolerant endo-1,4-β-glucanase to two barley cultivars. The amount of heterologous enzyme has been shown to be sufficient to reduce wort viscosity by decreasing the soluble β-glucan content.

In another approach, biochemically active wheat thioredoxin *h* has been overexpressed in the endosperm of transgenic barley grain. Such overexpression in germinated grain effected an up to fourfold increase in the activity of the debranching enzyme pullulanase, a rate-limiting enzyme in the breakdown of starch. The breaking down of starch is a key step in the malting process, and tests with transgenic varieties showed that the time required could be reduced by up to a day.

3.4 Traits of Pharmaceutical Interest

3.4.1 Production of Vaccines

Vaccines are designed to elicit an immune response without causing disease. While typical vaccines are composed of killed or attenuated disease-causing organisms, recombinant vaccines are desirable as an

alternative, as they generally cause fewer side effects than those that occur when the whole organism is delivered. Many candidate proteins have been identified that may function as effective subunit vaccines. Currently, the most common large-scale production of recombinant proteins are genetically engineered bacteria and yeast, due to the ease of manipulation and their rapid growth. However, recombinant proteins overexpressed in microorganisms must be extensively purified to remove host proteins and compounds. Transgenic plants provide an alternative system, with the great practical advantage of directly producing an edible plant tissue for oral immunization. In addition, edible plant-based recombinant vaccines are safe, provide nutrition, and are easy to administer.

At present, a great deal of research is focused on the understanding of transgene expression, stability, and processing in plants. Typical experiments investigating in-planta protein expression employ plant model systems such as potato, tomato, and maize. The ideal plant for human vaccination would be bananas, because they are readily eaten by babies, are consumed uncooked, and are indigenous to many developing countries. Since regeneration and growth to maturity can take up to three years in bananas, it is nevertheless necessary to optimize high-level expression of a vaccine antigen in model plants.

For viral infections, virus-like particles (VLPs), which form by self-assembly of viral surface proteins, are effective as vaccines. The hepatitis B surface antigen (HbsAg) has successfully been used as a vaccine against hepatitis B virus. The vaccine was first produced in yeast by Merck and SmithKline Beecham, and was the first recombinant subunit vaccine. Hence, it served as a model for a first attempt to produce a plant-based vaccine. HbsAg was expressed in tobacco and potato plants and shown to assemble into VLPs that are similar to the yeast-derived commercial vaccine.

In the United States, preliminary clinical trials have been performed, and two out of three volunteers who ate transgenic lettuce carrying a hepatitis B antigen displayed a good systemic response. Likewise, 19 of 20 people who ate a potato vaccine aimed at the Norwalk virus showed an immune response.

Currently, clinical trials with a plant-based vaccine against a bacterial pathogen are underway in the United States. Arntzen and colleagues inserted a gene for a part of an *E. coli* enterotoxin (LT-B subunit gene) that caused diarrhea in humans, into potato plants, and all but one of the 11 volunteers who ate the transgenic raw potatoes produced antibodies to the toxin.

A common problem of vaccine antigens expressed in plants has been the low level of expression. In LT-B-expressing plants, this has been overcome by using a "plant codon usage optimized" synthetic gene encoding LT-B and the use of the tuber-specific patatin promoter.

Future considerations must include containment of the transgenes as well as quality control for antigen content. These considerations are particularly important since inappropriate dosing of vaccines can impair their effectiveness, and constant dosing can lead (potentially) to immunological tolerance. In a more recent discussion, the feasibility of direct oral immunization by eating transgenic plants was doubted because of the problem of lot-to-lot consistency and administering pills prescribed by doctors was favored.

3.4.2 Production of Pharmaceuticals

Some recombinant proteins are already available commercially, including human erythropoietin or glucocerebrosidase. Until now, commercial production has used fermentation in *E. coli* and yeast or mammalian cell systems. However, these expression systems have significant limitations: bacteria cannot perform the complex posttranslational modifications required for bioactivity of many human proteins and production of proteins such as growth regulators or cell cycle inhibitors would negatively impact the transgenic animal cell culture.

Plants have several advantages compared with traditional systems for the molecular farming of pharmaceutical proteins. These include: (1) low production costs; (2) reduced time to market; (3) unlimited supply; (4) eukaryotic protein processing; and (5) safety from blood- or animal tissue-borne human pathogens. The production of a range of therapeutic proteins in transgenic plants has proven their capability for production of bioactive human proteins of pharmaceutical value: the proteins appear fully functional and structurally comparable with the analogous proteins produced in animal cell culture.

Examples of human therapeutic proteins include serum proteins such as hemoglobin as blood substitute, interferons as viral protection agents, lysosomal enzymes lacking in patients with Gaucher or Fabry disease, and other proteins such as hirudin, which is effective as an anticoagulant.

The majority of these pharmaceuticals are produced in model plants that are easy to transform, such as tobacco and potato, by using the constitutive 35S promoter from the cauliflower mosaic virus. However, as high levels of protein accumulation may negatively impact on yield and/or growth of the transgenic plants, inducible or tissue-specific promoters are preferable. Tobacco is an excellent biomass producer, but does produce toxic compounds. Crop-based production systems (e.g. wheat, rice, corn) lack these toxic substances and provide an existing infrastructure for their cultivation, harvest, distribution, and processing. This makes food crops highly attractive but leads to the potential for the unintended presence of therapeutic proteins in human food. Recently publicized incidents in which genetically modified crops have inadvertently mixed with those destined for human consumption have highlighted the need for mechanisms to ensure the segregation of plants that express pharmaceuticals.

Cereal crops (rice and wheat) were first used as production and storage system for a single-chain antibody against carcinoembryogenic antigen (CEA), a well-characterized tumor-associated marker antigen. In fact, in dried seeds, the antibodies could be stored for at least five months without significant loss of activity of the antibody. An alternate technique employs targeting of the recombinant protein to the oil bodies of *Brassica napus* seeds. A synthetic gene coding for a hirudin variant was fused to an *Arabidopsis* oleosin gene and transferred to oilseed rape. The recombinant hirudin was correctly targeted and accumulated on the oilbodies of transgenic seeds, which, because of their lower density, could easily be separated by flotation centrifugation. The functional biopharmaceutical was then released by protease treatment.

A recent publication showed the potential for producing recombinant proteins in the guttation fluid of tobacco. Guttation, which is the loss of water and dissolved materials from uninjured plant organs, leads

to the production of a guttation fluid. This can be collected throughout the plant's life, thereby providing a continuous and nondestructive system for recombinant protein production.

4 Outlook

It is likely that, in the future, a rapid development of molecular and technological methods will take place, and this will comprise the initial development of new transgenic approaches in model plants such as *Arabidopsis*, followed by the transfer of ready-established methods to most crop plants. The combination of genetic engineering and molecular marker technology with conventional crossings and subsequent selection will enable a rapid progress in plant breeding.

Despite these possibilities, limitations of widespread application can be foreseen in the patent situation, and this will lead to a partial unpredictability in production costs. A hindrance to the worldwide trade – and therefore also to the marketability of genetically modified foods – is founded in the different legal regulations, for example, in the United States and Europe. Additionally, the lack of acceptance by the public of genetically modified food, especially in central Europe, might delay the commercial potential of transgenic plants, even when there might be clear advantages for the consumer.

Acknowledgment

We thank Nina Mitra for her valuable technical assistance in preparing this manuscript.

See also Genetics, Molecular Basis of; Plant Gene Expression, Regulation of; Reproduction of Plants.

Bibliography

Books and Reviews

AGBIOS (Agriculture & Biotechnology Strategies (Canada) Inc.) (2002) GMO Database. http://www.agbios.com/default.asp (data from December 2004).

Bajaj, S., Targolli, J., Liu, L.-F., Ho, T.-H., Wu,, R. (1999) Transgenic approaches to increase dehydration-stress tolerance in plants, *Mol Breeding* **5**, 493–503.

Daniell, H., Khan, M.S., Allision, L. (2002) Milestones on chloroplast genetic engineering: an environmentally friendly era in biotechnology, *Trends in Plant Sci.* **7**, 84–91.

Dixon, R.A., Lamb, C.J., Masoud, S., Sewalt, V.J.H., Paiva, N.L. (1996) Metabolic engineering: prospects for crop improvement through the genetic manipulation of the phenylpropanoid biosynthesis and defense responses – a review, *Gene* **179**(1), 61–71.

Hammond, J., McGarvey, P., Yushibov, V. (Eds.) (1999) *Plant Biotechnology: New Products and Applications*, Springer-Verlag, Berlin, New York.

ISAAA (International Service for the Acquisition of Agri-biotech Applications) (2004) ISAAA Briefs No 30-2003. Global Status of Commercialized Transgenic Crops: 2003. http://www.isaaa.org/ (accessible December 2004).

Ma, J.K.-C., Drake, P.M.W., Christou, P. (2003) The production of recombinant pharmaceutical proteins in plants, *Nat. Rev. Genet.* **4**, 794–805.

Pew Initiative on Food and Biotechnology (2001) Harvest on the horizon: future uses of agricultural biotechnology. http://pewagbiotech.org/research/harvest/harvest.pdf. (accessible December 2004).

Potrykus, I., Spangenberg, G. (Eds.) (1993) *Gene Transfer to Plants*, Springer Lab Manual, Berlin.

Primary Literature

Abbadi, A., Domergue, F., Bauer, J., Napier, J.A., Welti, R., Zähringer, U., Cirpus, P., Heinz, E. (2004) Biosynthesis of Very-Long-Chain polyunsaturated fatty acids in transgenic oilseeds: Constraints on their accumulation, *Plant Cell* **16**, 2734–2748.

Arakawa, T., Chong, D.K.X., Langridge, W.H.R. (1998) Efficacy of a food plant-based oral cholera toxin B subunit vaccine, *Na. Biotechnol.* **16**, 292–297.

Arencibia, A.D., Carmona, E.R., Cornide, M.T., Castiglione, S., O'Reilly, J., Chinea, P., Oramas, P., Sala, F. (1999) Somaclonal variation in insect-resistant sugarcane (Saccharum hybrid) plants produced by electroporation, *Transgenic Res.* **8**, 349–360.

Atkinson, H.J., Urwin, P.E., Hansen, E., McPherson, M.J. (1995) Designs for engineered resistance to root-parasitic nematodes, *Trends Biotechnol.*. **13**, 369–374.

Barro, F., Rooke, L., Békés, F., Gras, P., Tatham, A.S., Fido, R., Lazzeri, P., Shewry, P.R., Barcelo, P. (1997) Transformation of wheat with high molecular weight subunit genes results in improved functional properties, *Nat. Biotechnol.* **15**, 1295–1299.

Capell, T., Bassie, L., Christou, P. (2004) Modulation of the polyamine biosynthetic pathway in transgenic rice confers tolerance to drought stress, *PNAS* **101**, 9909–9914.

Celis, C., Scurrah, M., Cowgill, S., Chumbiauca, S., Green, J., Franco, J., Main, G., Kiezebrink, D., Visser, R.G.F., Atkinson, H.J. (2004) Environmental biosafety and transgenic potato in a centre of diversity for this crop, *Nature* **432**, 222–225.

Chakraborty, S., Chakraborty, N., Datta, A. (2000) Increased nutritive value of transgenic potato by expressing a nonallergenic seed albumin gene from *Amaranthus hypochondriacus*, *Proc. Natl. Acad. Sci. U.S.A.* **97**(7), 3724–3729.

Cheng, X., Sardana, R., Kaplan, H., Altosaar, I. (1998) *Agrobacterium*-transformed rice plants expressing synthetic *cryIA(b)* and *cryIA(c)* genes are highly toxic to striped stem borer and yellow stem borer, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2767–2772.

Chen, Z., Young, T.E., Ling, J., Chang, S.-C., Gallie, D.R. (2003) Increasing vitamin C content of plants through enhanced ascorbate recycling, *PNAS* **100**, 3525–3530.

Cho, M.-J., Wong, J.H., Marx, C., Jiang, W., Lemaux, P.G., Buchanan, B.B. (1999) Overexpression of thioredoxin *h* leads to enhanced activity of starch debranching enzyme (pullulanase) in barley grain, *Proc. Natl. Acad. Sci. U.S.A.* **96**(25), 14641–14646.

Clausen, M., Kräuter, R., Schachermayr, G., Potrykus, I., Sautter, C. (2000) Antifungal activity of a virally encoded gene in transgenic wheat, *Nat. Biotechnol.* **18**, 446–449.

Corneille, S., Lutz, K., Svab, Z., Maliga, P. (2001) Efficient elimination of selectable marker genes from the plastid genome by the cre-lox site-specific recombination system, *Plant J.* **27**, 171–178.

De Block, M., Botterman, J., Vandewiele, M., Dockx, J., Thoen, C., Gossele, V., Movva, N.R., Thompson, C., Van Montague, M., Leemans, J. (1987) Engineering herbicide resistance in plants by expression of a detoxifying enzyme, *EMBO J.* **6**, 2513–2518.

De la Fuente, J.M., Ramirez-Rodriguez, V., Cabrera-Ponce, J.L., Herrera-Estrella, L. (1997) Aluminum tolerance intransgenic plants by alteration of citrate synthesis, *Science* **276**, 1566–1568.

Delhaize, E., Ryan, P.R., Hebb, D.M., Yamamoto, Y., Sasaki, T., Matsumoto, H. (2004) Engineering high-level aluminium tolerance in barley with the ALMT1 gene, *PNAS* **101**, 15249–15254.

Düring, K., Porsch, P., Fladung, M., Lörz, H. (1993) Transgenic potato plants resistant to the phytopathogenic bacterium *Erwinia carotovora*, *Plant J.* **3**(4), 587–598.

Farré, E.M., Bachmann, A., Willmitzer, L., Trethewey, R.N. (2001) Acceleration of potato tuber sprouting by the expression of a bacterial pyrophosphatase, *Nat. Biotechnol.* **19**, 268–272.

Fujimoto, H., Yamamoto, M., Kyozuka, J., Shimamoto, K. (1993) Insect-resistant rice generated by introduction of a modified δ-endotoxin gene from *Bacillus thuringiensis*, *Biotechnology* **11**, 1151–1155.14.

Gal-On, A., Wang, Y., Faure, J.E., Pilowsky, M., Zelcer, A. (1998) Transgenic resistance to cucumber mosaic virus in tomato: blocking of long-distance movement of the virus in lines harbouring a defective viral replicase, *Phytopathology* **88**(10), 1101–1107.

Gandikota, M., de Kocho, A., Chen, L., Ithal, N., Fauquet, C., Reddy, A.R. (2001) Development of transgenic rice plants expressing maize

anthocyanin genes and increased blast resistance, *Mol. Breeding* **7**(1), 73–83.

Gatehouse, A.M.R., Hilder, V.A., Powell, K.S., Wang, M., Davison, G.M., Gatehouse, L.N., Down, R.E., Edmonds, H.S., Boulter, D., Newell, C.A., Merryweather, A., Hamilton, W.D.O., Gatehouse, J.A. (1994) Insect-resistant transgenic plants: choosing the gene to do the 'job', *Biochem. Soc. Trans.* **22**, 944–949.

Goto, F., Yoshihara, T., Shigemoto, N., Toki, S., Takaiwa, F. (1999) Iron fortification of rice seeds by the soybean ferritin gene, *Nat. Biotechnol.* **17**, 282–286.

Hain, R., Reif, H.J., Krause, E., Langbartels, R., Kindl, H., Vornam, B., Wiese, W., Schmelzer, E., Schreier, P.H., Stocker, R.H., Stenzel, K. (1993) Disease resistance results from foreign phytoalexin expression in a novel plant, *Nature* **361**, 153–156.

Hayakawa, T., Zhu, Y., Itoh, K., Izawa, T., Shimamoto, K., Toriyama, S. (1992) Genetically engineered rice resistant to rice stripe virus, an insect-transmitted virus, *Proc. Natl. Acad. Sci. U.S.A.* **89**(20), 9865–9869.

He, X.Z., Dixon, R.A. (2000) Genetic manipulation of isoflavone 7-O-methyltransferase enhances biosynthesis of 4′-O-methylated isoflavonoid phytoalexins and disease resistance in alfalfa, *Plant Cell* **12**(9), 1689–1702.

He, Z., Fu, Y., Hu, G., Zhang, S., Yu, Y., Sun, Z. (2004) Phosphomannose-isomerase (*pmi*) gene as selectable marker for rice transformation via *Agrobacterium*, *Plant Sci.* **166**, 17–22.

He, Z., Zhu, Q., Dabi, T., Li, D., Weigel, D., Lamb, C. (2000) Transformation of rice with the *Arabidopsis* floral regulator *LEAFY* causes early heading, *Transgenic Res.* **9**, 223–227.

Jefferson, R.A., Kavanagh, T.A., Bevan, M.W. (1987) GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants, *EMBO J.* **6**, 39001–33908.

Joersbo, M., Donaldson, I., Petersen, S.G., Brunstedt, J., Okkels, F.T. (1998) Analysis of mannose selection used for transformation of sugar beet, *Mol. Breeding* **4**, 111–117.

Kasuga, M., Liu, Q., Yamaguchi-Shinozaki, K., Shinozaki, K. (1999) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor, *Nat. Biotechnol.* **17**, 287–291.

Komarnytsky, S., Borisjuk, N.V., Borisjuk, L.G., Alam, M.Z., Raskin, I. (2000) Production of recombinant proteins in tobacco guttation fluid, *Plant Physiol.* **124**, 927–933.

Kramer, K.J., Morgan, T.D., Throne, J.E., Dowell, F.E., Bailey, M., Howard, J.A. (2000) Transgenic avidin maize is resistant to storage insect pests, *Nat. Biotechnol.* **18**, 670–674.

Kunkel, T., Niu, Q.-W., Chan, Y.-S., Chua, N.-H. (1999) Inducible isopentenyl transferase as a high-efficiency marker for plant transformation, *Nat. Biotechnol.* **17**, 916–919.

Langridge, W.H.R. (2000) Edible vaccines, *Sci. Am.* **283**(3), 66–71.

Lin, W., Anuratha, C.S., Datta, K., Potrykus, K., Muthukrishnan, S., Datta, S.K. (1995) Genetic engineering of rice for resistance to sheath blight, *Biotechnology* **13**, 686–691.

Lodge, J.K., Kaniewski, W.K., Tumer, N.E. (1993) Broad-spectrum resistance in transgenic plants expressing pokeweed antiviral protein, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7089–7093.

Lorito, M., Woo, S.L., Fernandez, I.G., Colucci, G., Harman, G.E., Pintor-Toro, J.A., Filippone, E., Muccifora, S., Lawrence, C.B., Zoina, A., Tuzun, S., Scala, F. (1998) Genes from mycoparasitic fungi as source for improving plant resistance to fungal pathogens, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7860–7865.

Mason, H.S., Lam, D.M.K., Arntzen, C.J. (1992) Expression of hepatitis B surface antigen in transgenic plants, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 11745–11749.

McElroy, D., Brettell, R.I.S. (1994) Foreign gene expression in transgenic cereals, *Trends Biotechnol.* **12**, 62–68.

Moellenbeck, D.J., Peters, M.L., Bing, J.W., et al. (2001) Insecticidal proteins from *Bacillus thuringiensis* protect corn from corn root worms, *Nat. Biotechnol.* **19**, 668–672.

Morton, R.L., Schroeder, H.E., Bateman, K.S., Chrispeels, M.J., Armstrong, E., Higgins, T.J. (2000) Bean alpha-amylase inhibitor 1 in transgenic peas (*Pisum sativum*) provides complete protection from pea weevil (*Bruchus pisorum*) under field conditions, *Proc. Natl. Acad. Sci. U.S.A.* **97**(8), 3820–3835.

Nuutila, A.M., Ritala, A., Skadsen, R.W., Mannonen, L., Kaupinnen, V. (1999) Expression of fungal thermotolerant endo-1,4-beta-glucanase in transgenic barley seeds during germination, *Plant Mol. Biol.* **41**(6), 777–783.

Oldach, K., Becker, D., Lörz, H. (2001) Heterologous expression of genes mediating enhanced

fungal resistance in transgenic wheat, *Mol. Plant-Microbe Interact.* **14**(7), 832–838.

Osusky, M., Zhou, G., Osuska, L., Hancock, R.E., Kay, W.W., Misra, S. (2000) Transgenic plants expressing cationic peptide chimeras exhibit broad-spectrum resistance to phytopathogens, *Nat. Biotechnol.* **18**, 1162–1166.

Ow, D.W., Wood, K.V., DeLuca, M., de Wet, J.R., Helinski, D.R., Howell, S.H. (1986) Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants, *Science* **234**, 856–859.

Parmenter, D.L., Boothe, J.G., Van Rooijen, G.J.H., Yeung, E.C., Moloney, M.M. (1995) Production of biological active hirudin in plant seeds using oleosin partitioning, *Plant Mol. Biol.* **29**, 1167–1180.

Pena, L., Martin-Trillo, M., Juarez, J., Pina, J.A., Navarro, L., Martinez-Zapater, J.M. (2001) Constitutive expression of *Arabidopsis LEAFY* or *APETALA1* genes in citrus reduces their generation time, *Nat. Biotechnol.* **19**, 263–267.

Rao, K.V., Rathore, K.S., Hodges, T.K., Fu, X., Stoger, E., Sudhakar, D., Williams, S., Christou, P., Bharathi, M., Bown, D.P., Powell, K.S., Spence, J., Gatehouse, A.M.R., Gatehouse, J.A. (1998) Expression of snowdrop lectin (GNA) in transgenic rice plants confers resistance to rice brown planthopper, *Plant J.* **15**(4), 469–477.

Römer, S., Fraser, P.D., Kiano, J.W., Shipton, C.A., Misawa, N., Schuch, W., Bramley, P.M. (2000) Elevation of the provitamin A content of transgenic tomato plants, *Nat. Biotechnol.* **18**, 666–669.

Rotino, G.L., Perri, E., Zottini, M., Sommer, H., Spena, A. (1997) Genetic engineering of parthenocarpic plants, *Nat. Biotechnol.* **15**, 1398–1401.

Roy, M., Wu, R. (2001) Arginine decarboxylase transgene expression and analysis of environmental stress tolerance in transgenic rice, *Plant Sci.* **160**, 869–875.

Sanford, J.C., Klein, T.M., Wolf, E.D., Allen, N. (1987) Delivery of substances into cells and tissues using a particle bombardment process, *J. Part. Sci. Tech.* **5**, 27–37.

Sayre, R.T. (2000) Cyanogen reduction in transgenic cassava. Generation of safer food product for subsistence farmers, ISB News Report, August 2000. http//www.isb.vt.edu/articles/aug0003.htm (accessible December 2004).

Schaffrath, U., Mauch, F., Freydell, E., Schweizer, P., Dudler, R. (2000) Constitutive expression of the defense-related *Rir1b* gene in transgenic rice plants confers enhanced resistance to the rice blast fungus *Magnaporthe grisea*, *Plant Mol. Biol.* **43**, 59–66.

Sevenier, R., Hall, R.D., van der Meer, I., Hakkert, H.J.C., van Tunen, A.J., Koops, A.J. (1998) High-level fructan accumulation in a transgenic sugar beet, *Nat. Biotechnol.* **16**, 843–846.

Stöger, E., Vaquero, C., Torres, E., Sack, M., Nicholson, L., Drossard, J., Williams, S., Keen, D., Perrin, Y., Christou, P., Fischer, R. (2000) Cereal crops as viable production and storage systems for pharmaceutical scFv antibodies, *Plant Mol. Biol.* **42**, 583–590.

Stommel, J.R., Tousignant, M.E., Wai, T., Pasini, R., Kaper, J.M. (1998) Viral satellite RNA expression in transgenic tomato confers field tolerance to cucumber mosaic virus, *Plant Dis.* **82**(4), 391–396.

Tacket, C.O., Mason, H.S., Losonsk, G., Clements, J.D., Levine, M.M., Arntzen, C.J. (1998) Immunogenicity in humans of a recombinant bacterial antigen delivered in a transgenic potato, *Nat. Med.* **4**, 607–609.

Tai, T.H., Dahlbeck, D., Clark, E.T., Gajiwala, P., Pasion, R., Whalen, M.C., Stall, R.E., Staskawicz, B.J. (1999) Expression of the *Bs2* pepper gene confers resistance to bacterial spot disease in tomato, *Proc. Natl. Acad. Sci. U.S.A.* **96**(24), 14153–14158.

Takahashi, M., Nakanishi, H., Kawasaki, S., Nishizawa, N.K., Mori, S. (2001) Enhanced tolerance of rice to low iron availability in alkaline soils using barley nicotianamine aminotransferase genes, *Nat. Biotechnol.* **19**, 466–469.

Takaichi, M., Oeda, K. (2000) Transgenic carrots with enhanced resistance against two major pathogens, *Erysiphe heraclei* and *Alternaria dauci*, *Plant Sci.* **153**, 135–144.

Tang, X., Xie, M., Kim, Y.J., Zhou, J., Klessig, D.F., Martin, G.B. (1999) Overexpression of *Pto* activates responses and confers broad resistance, *Plant Cell* **11**(1), 15–29.

Tavladoraki, P., Benvenuto, E., Trinca, S., De Martinis, D., Cattaneo, A., Galeffi, P. (1993) Transgenic plants expressing a functional single-chain Fv antibody are specifically protected from virus attack, *Nature* **366**, 469–472.

Tepfer, M. (1993) Viral genes and transgenic plants, *Biotechnology* **11**, 1125–1132.

Tu, J., Ona, I., Zhang, Q., Mew, T.W., Khush, G.S., Datta, S.K. (1998) Transgenic rice variety 'IR 72' with Xa21 is resistant to bacterial blight, *Theoret. Appl. Genet.* **97**(1/2), 31–36.

Tu, J., Zhang, G., Datta, K., Xu, C., He, Y., Zhang, Q., Khush, G.S., Datta, S.K. (2000) Field performance of transgenic elite commercial hybrid rice expressing *Bacillus thuringiensis* δ-endotoxin, *Nat. Biotechnol.* **18**, 1101–1104.

Urwin, P.E., Green, J., Atkinson, H.J. (2003) Expression of a plant cystatin confers partial resistance to Globodera, full resistance is achieved by pyramiding a cystatin with natural resistance, *Mol Breeding.* **12**, 263–269.

Urwin, P.E., Lilley, C.J., McPherson, M.J., Atkinson, H.J. (1997) Resistance to both cyst and root-knot nematodes conferred by transgenic *Arabidopsis* expressing a modified plant cystatin, *Plant J.* **12**(2), 455–461.

Valanzuolo, S., Catello, S., Colombo, M., Dani, M., Monti, M.M., Uncini, L., Petrone, P., Spingo, P. (1994) Cucumber mosaic virus resistance in transgenic San marzano tomatoes, *Acta Hortic.* **376**, 377–386.

Ye, X., Al-Babili, S., Klöti, A., Zhang, J., Lucca, P., Beyer, P., Potrykus, I. (2000) Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm, *Science* **287**, 303–305.

Zhang, H.-X., Blumwald, E. (2001) Transgenic salt-tolerant tomato plants accumulate salt in the foliage but not in fruit, *Nat. Biotechnol.* **19**, 765–768.

Zubko, E., Scutt, C., Meyer, P. (2000) Intrachromosomal recombination between attP regions as a tool to remove selectable marker genes from tobacco transgenes, *Nat. Biotechnol.* **18**, 442–445.

Zuo, J., Niu, Q.-W., Møller, S.G., Chua, N.-H. (2001) Chemical-regulated, site-specific DNA excision in transgenic plants, *Nat. Biotechnol.* **19**, 157–161.

Translation of RNA to Protein

Robert Ashley Cox and Henry Rudolph Victor Arnstein
National Institute for Medical Research, London, England, UK

1 Introduction 582

2 mRNA Structure and the Genetic Code 584
2.1 Structure 584
2.2 Prokaryotic mRNA 585
2.3 Eukaryotic mRNA 585
2.4 The Genetic Code 586
2.4.1 Deviations from the Standard Genetic Code 586

3 Transfer RNA (tRNA) 587

4 Ribosome Structure and Function in Translation 590
4.1 Ribosome Structure 590
4.2 The Ribosome Cycle in Translation 594
4.2.1 Formation of Preinitiation Complexes 594
4.2.2 Initiation Complex Formation: Joining of the Large Ribosomal Subunit 598
4.2.3 Polypeptide Chain Synthesis: The Elongation–Translocation Cycle 599
4.3 High-resolution Structural Studies of the Ribosome 604

5 Translational Control of Gene Expression 608
5.1 mRNA Stability 609
5.2 Control by Interaction of Proteins with mRNA 611
5.3 Control by mRNA Structure 612
5.4 Control by Modification of Translation Factor Activity 612
5.4.1 Initiation Factors 612
5.4.2 Other Translation Factors 615
5.5 Effects of Antisense Polynucleotides 615
5.6 Availability of Amino Acids, tRNA Abundance, and Codon Usage 615

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

5.6.1 Amino Acids 615
5.6.2 Abundance of tRNAs and Codon Usage 615
5.7 Modulation of Ribosome Activity 615
5.8 Ribosome-inactivating Proteins 616

6 Concluding Remarks 616

Bibliography 617
Books and Reviews 617
Primary Literature 617

Keywords

Anticodon
Three consecutive bases in tRNA that bind to a specific mRNA codon by complementary antiparallel base pairing.

Antiparallel Base pairing
Pairing through specific hydrogen bonds between base residues of two polynucleotide chains or two segments of a single chain with phosphodiesterbonds running in the $5' \rightarrow 3'$ direction in one chain or segment and in the $3' \rightarrow 5'$ direction in the other. In DNA and RNA, the hydrogen bonds are usually formed between complementary base pairs (adenine (A) with either thymine (T) or uracil (U), and guanine (G) with cytosine (C)).

Codon
Three consecutive bases in mRNA or DNA, which code for an amino acid in protein synthesis.

Elongation
The stepwise addition of amino acids to the carboxyl terminus of a growing polypeptide chain.

Fidelity
The accuracy with which the RNA sequence is translated into the correct amino acid sequence.

Initiation
A multistep reaction between ribosomal subunits, charged initiator transfer RNA and messenger RNA that results in apposition of the ribosome-bound initiator Met-tRNA with an AUG initiator codon in mRNA. In this position, the ribosome is poised to form the first peptide bond.

Polarity
The asymmetry of a polynucleotide or polypeptide. In DNA, the two strands have opposite polarity; that is, they run in opposite directions (5′ → 3′ and 3′ → 5’). The polarity of a polypeptide is defined as running from the N-terminus to the C-terminus.

Reading Frame
One of three possible ways of translating groups of three nucleotides in mRNA. The appropriate reading frame is determined by the initiation codon.

Template Strand
The strand of the DNA double helix that is used as a template for transcription of RNA. It has a base sequence complementary to the RNA transcript.

Termination
The end of polypeptide synthesis, which is signaled by a codon for which there is no corresponding aminoacyl tRNA. When the ribosome reaches a termination codon in the mRNA, the polypeptide is released and the ribosome–mRNA–tRNA complex dissociates.

Translation
The stepwise synthesis of a polypeptide with an amino acid sequence determined by the nucleotide sequence of the mRNA coding region. The genetic code relates each amino acid to a group of three consecutive nucleotides termed a *codon*. Decoding of mRNA takes place in the 5′ → 3′ direction, and the polypeptide is synthesized from the amino to the carboxyl terminus.

Translocation
The stepwise advance of a ribosome along the mRNA, one codon at a time, with simultaneous transfer of peptidyl tRNA from the A site to the P site of the ribosome.

Proteins have a linear primary sequence of no more than 20 encoded amino acids and this underlying simplicity is used to express great subtlety in structure and versatility in function. For these reasons, proteins are essential to the structure and function of living cells. The assembly of polypeptide chains from amino acids and their subsequent modifications, leading to the final three-dimensional protein structure, are exceptionally complex processes; many components are involved and much of the cell’s energy is utilized. Each peptide bond requires the expenditure of four high-energy phosphate bonds (ATP is converted to AMP and two molecules of GTP are converted to GDP). This value excludes the energy used for initiation and release of the polypeptide chains and the cost of synthesizing and processing mRNA. The linear amino acid sequence of a protein is encoded within the gene as a linear deoxyribonucleotide sequence. Early steps in the biosynthesis of a protein include transcription of the gene and appropriate processing of the transcript leading to

the production of mature messenger RNA (mRNA). We describe the mechanisms involved in translating mRNA to produce a polypeptide chain which has the amino acid sequence specified by the gene. Translation takes place after mRNA is bound to small ribonucleoprotein particles called *ribosomes* through the mediation of aminoacyl tRNA which links the linear nucleotide sequence of mRNA with the linear amino acid sequence of the encoded polypeptide. The genetic message is read sequentially as mRNA moves relative to the ribosome. Thus, the polypeptide chain is formed by the sequential incorporation of amino acids into peptide linkage. The process of translation is achieved with high fidelity with less than one error per 10 000 amino acids incorporated into protein. The mechanisms of protein synthesis are essentially the same for all cells irrespective of whether they have a nucleus (eukaryotes) or not (prokaryotes).

1 Introduction

A gene or cistron is defined as the region of DNA that is transcribed into a functional RNA. The transcript functions either as such (e.g. tRNA, rRNA, snRNA) or as a messenger (mRNA), which after processing or editing as required, normally codes for one or more polypeptide chains in the translation process. A polynucleotide such as RNA is an asymmetrical polymer assembled from nucleoside triphosphates by a stepwise mechanism linking the 5′ position of one nucleotide by a phosphate bridge to the 3′ position of the adjacent nucleotide. In the finished polynucleotide chain, the first nucleotide residue has a 5′ position, which is not linked to another nucleotide, whereas the last nucleotide has an unlinked 3′ position. Thus, polynucleotide synthesis proceeds from the 5′ to the 3′ terminus and the polymer is said to have a 5′ to 3′ polarity. Usually, linear RNA sequences are written with the 5′ terminus on the left and the 3′ terminus on the right (Fig. 1a). Within the RNA chain, some bases may form antiparallel base pairs (Fig. 1b, as found in DNA).

The polypeptide chains of proteins are also asymmetrical polymers in which the amino acid residues are linked by peptide bonds between their α-amino and carboxyl groups (Fig. 2), leaving a free α-amino group at one end (the amino terminus) of the polymer and a free α-carboxyl group at the opposite end (the carboxyl terminus). The significance of the polarity of the RNA and proteins will become evident when the process of protein biosynthesis is explained below.

The genetic information stored in DNA is not usable directly for making proteins. Rather, it must be copied into a primary RNA transcript containing the coding sequences by an enzymatic transcription of segments of DNA containing the genes. In prokaryotes, the primary transcript is also the messenger (mRNA); that is, it can be used directly in polypeptide synthesis. In contrast, in eukaryotes the primary transcript (precursor-mRNA) is often much larger in size than the mature mRNA and requires extensive processing involving the excision of intervening and other noncoding sequences.

Messenger RNA serves as the template for protein synthesis; that is, the linear

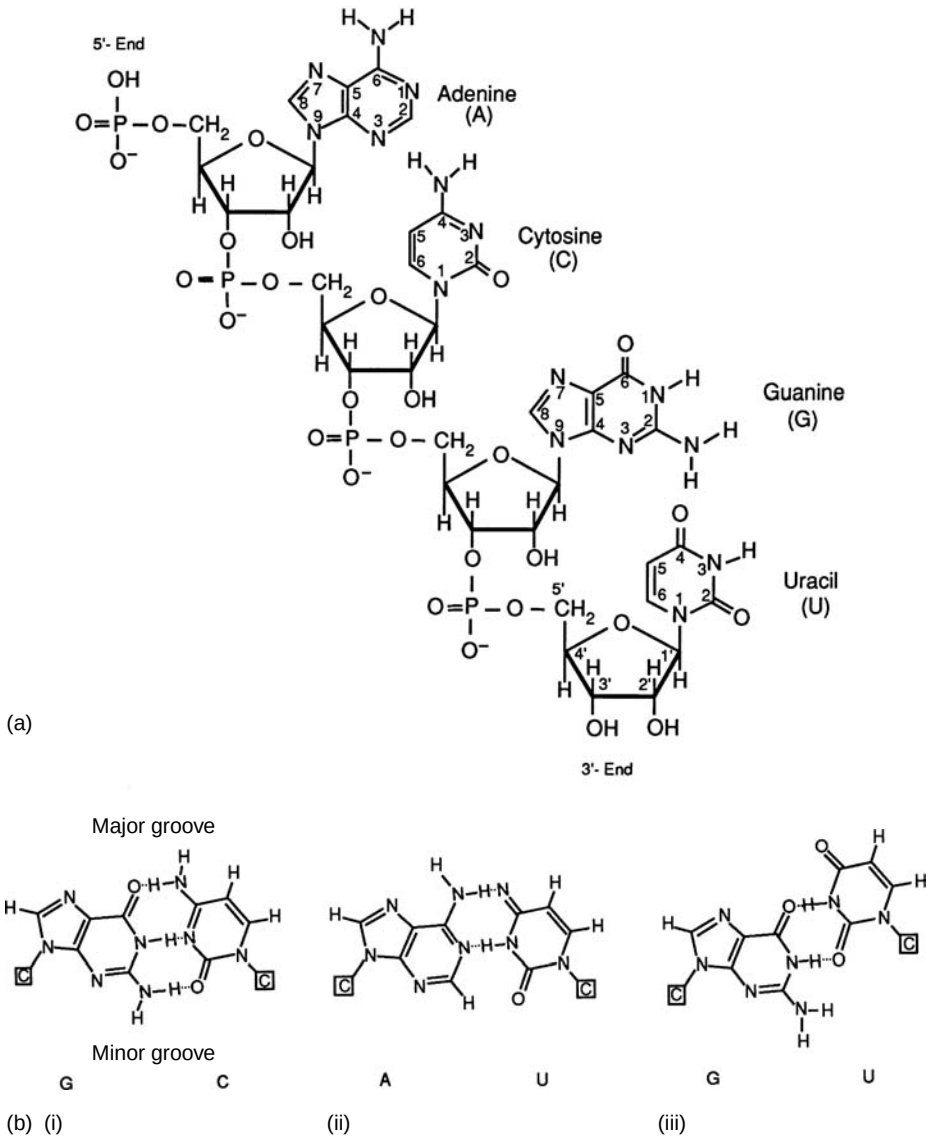

Fig. 1 Structural elements of ribonucleic acids. (a) Primary structure indicating the numbering system for purines, pyrimidines, and ribose. (b) Base-pairing interactions commonly found in RNA: (i) G + C base pair; (ii) A + U base pair; (iii) G + U base pair. Pairs (i) and (ii) are Watson and Crick base pairs; (iii) is a special type of base pairing found in intramolecular bihelical regions. The minor groove is on the side of the base pair with the glycosidic bond of which the carbon atom C 1′ of ribose is boxed. Note that DNA contains base pairs of types i and ii only but with thymine (5-methyluracil) in place of uracil. [From Arnstein, H.R.V., Cox, R.A. (1992) *Protein Biosynthesis*, Oxford University Press, Oxford. With permission.]

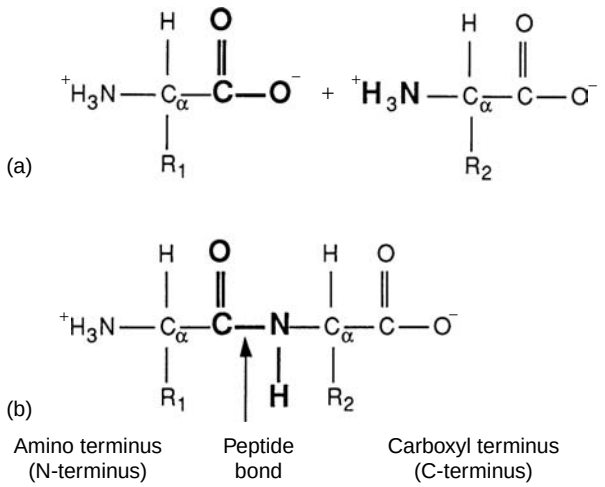

Fig. 2 Structure of the peptide bond. (a) Two L-amino acids with different side chains, R_1 and R_2; (b) a dipeptide formed from the two amino acids shown in (a).

nucleotide sequence of the mRNA dictates the amino acid sequence of the polynucleotide encoded originally by the gene. Conventionally, gene and mRNA nucleotide sequences are written in the 5′ to 3′ direction, which corresponds to the direction in which mRNA is decoded during polypeptide synthesis: The mRNA is read in the 5′ to 3′ direction, and the polypeptide is synthesized from the amino toward the carboxyl terminus.

The mechanism whereby RNA is translated into protein is complex, and the cell devotes considerable resources to the translational machinery. The components include 20 different amino acids, as well as transfer RNAs, aminoacyl tRNA synthetases, ribosomes, and a number of protein factors which cycle on and off the ribosomes and facilitate various steps in initiation of translations, elongation of the nascent polypeptide chain, and termination of synthesis with release of the completed polypeptide from the ribosome. The process depends on a supply of energy provided by ATP and GTP. The rate of protein synthesis is typically in the range of 6 (immature red blood cells of the rabbit) to 20 (*Escherichia coli* growing optimally) peptide bonds per second at 37 °C.

2 mRNA Structure and the Genetic Code

2.1 Structure

The sequence information of a gene that is copied (transcribed) into the nucleotide sequence of RNA from the complementary strand of DNA is called *the template strand*. The primary transcript is a single strand of RNA, which is a faithful copy of the nontemplate strand of DNA with substitution of U residues in place of T residues found in DNA. Sometimes, the primary transcript is altered, as described below, before it functions as mRNA; in these cases, the original unmodified transcript is the precursor or pre-mRNA. Usually, mRNAs have nontranslated sequences at the 5′ and

3′ ends in addition to the coding domain. These noncoding sequences sometimes affect the efficiency of translation and the stability of mRNA. The structures of typical mRNAs are shown in Fig. 3. The decoding process involves base pairing between three bases (designated a codon) in the mRNA and the three-base anticodon of a transfer RNA (tRNA). In a separate reaction, each tRNA is first linked to a particular amino acid and thus the pairing of mRNA with tRNA determines the sequence of amino acids in the resulting protein.

2.2 Prokaryotic mRNA

In prokaryotes, pre-mRNA usually undergoes little or no modification so that pre-mRNA and mRNA are very similar if not identical. Because pre-mRNA is colinear with DNA, DNA and proteins are usually colinear in these organisms. Gene expression in prokaryotes usually involves the cotranscription of several adjacent genes, and translation of mRNA sequences into polypeptides may begin at the 5′ end of mRNA while transcription is still in progress at the 3′ end.

2.3 Eukaryotic mRNA

In eukaryotes, the genetic information is stored mainly in the nucleus and to a minor degree in some organelles (mitochondria and chloroplasts). The description that follows pertains only to nuclear genes. Eukaryotic genes are more complicated than prokaryotic genes because the coding region is often discontinuous:

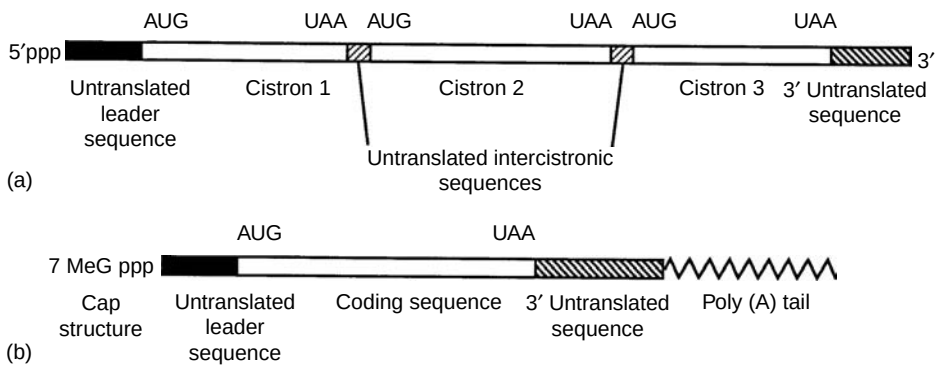

Fig. 3 Structure of typical mRNAs. (a) *Prokaryotic mRNA*. At the beginning of the transcript, there are two additional phosphate residues linked by pyrophosphate bonds to the 5′ phosphate group of the terminal nucleotide (see Fig. 1). The 5′ untranslated sequence often contains a ribosome-binding site (the Shine–Dalgarno sequence), which increases the efficiency of translation. AUG and UAA are representative initiation and termination codons, respectively. Cistrons (gene sequences) are separated by short (typically 12 to 24 nucleotides) noncoding intercistronic sequences. (b) *Eukaryotic mRNA*. Typically, the processed transcripts are monocistronic. The 5′ end is usually modified by the addition of 7-MeG, known as a cap structure, which is linked by two pyrophosphate groups to the terminal nucleotide of mRNA. The coding sequence is located between 5′ and 3′ untranslated sequences. In the majority of cases, the 3′ end is modified by the addition of 25 to 250 adenylate residues, termed the poly(A) tail. [From Arnstein, H.R.V., Cox, R.A. (1992) *Protein Biosynthesis*, Oxford University Press, Oxford. With permission.]

the coding sequences or exons are interrupted by intervening sequences (introns). Thus, genes and proteins are usually *not* colinear in eukaryotes. In the nucleus, a complicated set of splicing reactions removes all the introns from the pre-mRNA and fuses the exons into a continuous coding sequence. Other processing steps involve adding a "cap" to the 5′ end of the mRNA and a poly(A) "tail" to the 3′ end (see Fig. 3b). After completion of these nuclear maturation steps, the mRNA is transported to the cytoplasm where it is translated. As is the case with prokaryotic mRNA, the coding region is flanked by 5′ and 3′ nontranslated sequences.

2.4 The Genetic Code

The genetic code is triplet, comma-less, and nonoverlapping. As a consequence, a nucleotide sequence has three possible reading frames (Fig. 4). Because mRNA is normally translated into a unique polypeptide, an essential step in the translation process is the selection of the appropriate reading frame. This is achieved by starting translation at the initiation codon, usually AUG or less frequently GUG, which ensures that the following codons are read in phase within the required reading frame. Of the 64 theoretically possible triplets in the genetic code, 61 sense codons correspond to 20 genetically encoded amino acids found in all, or nearly all, proteins. When GUG is used as the initiation codon, it codes for methionine by interaction with the anticodon of the initiator Met-tRNA^{Met}, whereas elsewhere it codes for valine. All other codons specify only one amino acid but many amino acids are specified by two or more (up to six) codons; the code is unambiguous, but degenerate. The remaining three codons, termed *nonsense codons*, usually signify termination of synthesis and release of the finished polypeptide chain.

2.4.1 Deviations from the Standard Genetic Code

One of the nonsense codons, UGA, has an additional function in the synthesis of selenoproteins. The process involves the initial synthesis of selenocysteyl-tRNA from a novel seryl-tRNA and selenium. This tRNA contains an anticodon that is able to decode UGA and insert selenocysteine residues into the growing polypeptide chain, but only at UGA codons in a particular context of neighboring nucleotides. The insertion of selenocysteine residues into the polypeptide also requires a specific elongation factor T, which differs from the factor used for the

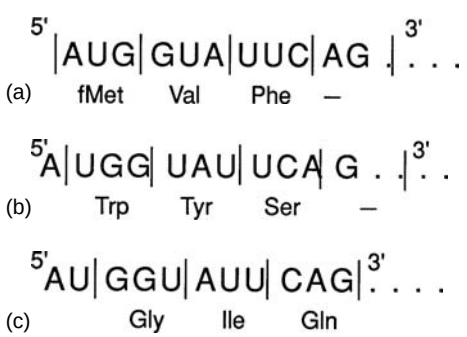

Fig. 4 Translation of a polynucleotide sequence into three alternative polypeptides using different reading frames.

incorporation of other aminoacyl tRNAs (see Sect. 4.2.3).

Mitochondria and chloroplasts, as well as certain organisms such as mycoplasma and ciliated protozoa, use a few non-standard genetic codons. For example, methionine is usually coded for by AUG (or occasionally GUG), but in human mitochondria this codon is replaced by AUA. Variations in the genetic code are thought to have arisen as a result of the loss of some tRNA genes and mutational pressure on DNA, giving rise to a predominance of either AT- or GC-rich codons.

3 Transfer RNA (tRNA)

Transfer RNA nomenclature: The amino acid linked to a charged tRNA is indicated by a prefix and the specificity of the tRNA in the aminoacylation reaction is shown as a superscript on the right; for example, Phe-tRNA^{Phe} indicates phenylalanine-specific tRNA charged with phenylalanine. The anticodon may be indicated as a right subscript or, alternatively, in the superscript after the amino acid; for example, tRNA$_{UGC}$ or tRNA$^{Ala/UGC}$. The right-hand subscript position is sometimes used to indicate the organism from which the tRNA is derived (e.g. tRNA$^{Val}_{yeast}$). The initiator tRNA, which is specific for methionine, is termed Met-tRNA$^{Met}_{f}$ or Met-tRNA$^{Met}_{i}$. Often the superscript ^{Met} is omitted. In prokaryotes and in the mitochondria of eukaryotes, the methionine residue of the charged initiator tRNA is formylated by transformylase using N^{10}-formyltetrahydrofolate as the donor, giving N-formylmet-tRNA$_f$. Commonly, this charged tRNA is termed fMet-tRNA$_f$. The methionine-specific elongator RNA, which inserted methionine into internal positions of the growing peptide chain is termed tRNA$^{Met}_{m}$ (or tRNA$_m$) when uncharged, and Met-tRNA$^{Met}_{m}$ (or Met-tRNA$_m$) when charged. By relating individual codons of mRNA to the cognate amino acids, tRNA functions as a key bilingual intermediate in the translation of the genetic code. All tRNAs are single-stranded molecules about 80 nucleotides long with a common 3′ terminal CCA sequence. Most of the bases are standard but some (e.g. pseudoU, dihydroU, and T) are derived by modification after transcription of the transfer RNA genes.

The secondary structure of tRNA is usually presented in two dimensions as a cloverleaf to highlight the regions of base pairing (Fig. 5a). X-ray crystallography reveals that additional hydrogen bonds give rise to an L-shaped tertiary structure (Fig. 5b). The CCA sequence carrying the amino acid is located distal to the anticodon.

The decoding process involves antiparallel base pairing between the three bases of mRNA codons and the complementary anticodons of tRNA during peptide bond formation (Fig. 6). The first and middle bases of the codon form conventional base pairs with the third and middle base of the anticodon, respectively, but the third base of the codon pairs with the first base of the anticodon by a less stringent interaction (e.g. base pairing of G with U as well as with C), giving rise to degeneracy. This so-called "wobble" considerably reduces the number of tRNA species required to decode the 61 sense codons. Thus, the protein synthesis system in the cytosol of eukaryotes contains only a few more than 40 different tRNAs, and in mitochondria, 22 to 24 species are sufficient.

The attachment of amino acids to tRNA involves the formation of an ester bond between the α-carboxyl group of the amino acid and the 3′ hydroxyl group of the terminal adenosine of tRNA. It requires specific enzymes, the aminoacyl tRNA synthetases. There are 20 different synthetases, each specific for one of the 20 amino acids, and each enzyme recognizes something unique in the structure of its cognate tRNA. The structural determinants that ensure accuracy of this

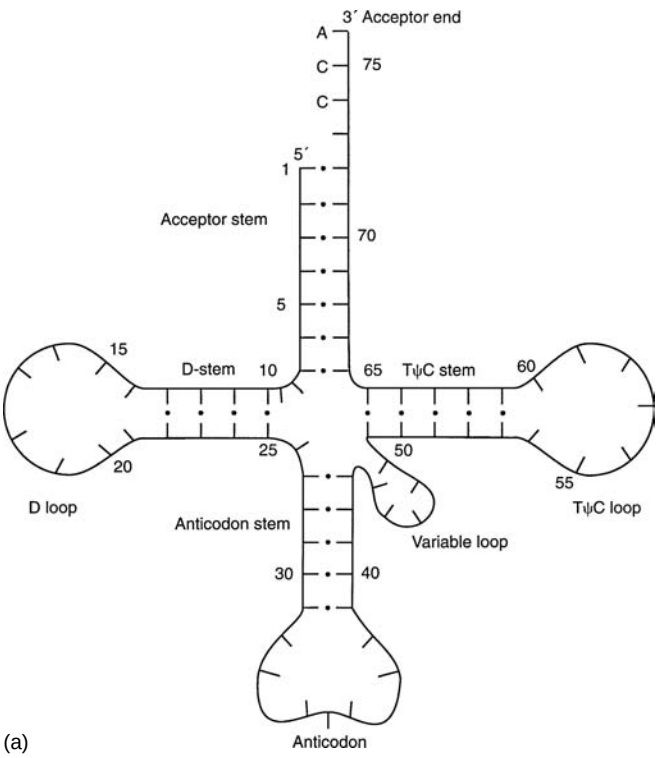

(a)

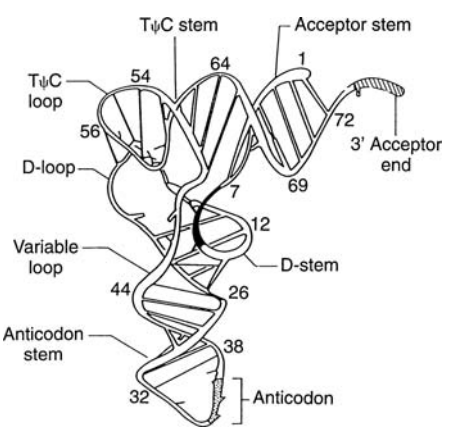

(b)

charging reaction vary for different tRNAs. The anticodon may play a part but sometimes even a single base elsewhere is sufficient to determine the specificity of the tRNA-synthetase interaction. The accuracy of the synthetase reaction in attaching an amino acid to its cognate tRNA is critically important to the fidelity of the

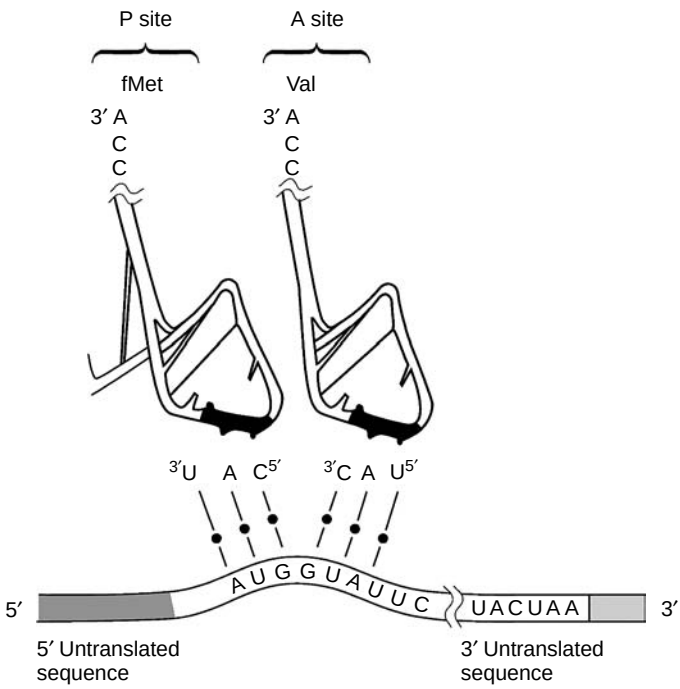

Fig. 6 Schematic illustration of base pairing between a codon and its anticodon. The diagram shows the interaction at the P site of the ribosome (see Fig. 9) between the initiation codon AUG of mRNA and the anticodon CAU of fMet-tRNA$_f^{Met}$ and the interaction at the A site between the codon GUA of mRNA and the anticodon UAC of Val-tRNA^{Val}. The polarity of an RNA species runs from the 5′ end to the 3′ end. The fragment of tRNA is representative of the general structure (see Fig. 5b) placed in the appropriate orientation. The interaction between the codon and the anticodon is antiparallel, and the three base pairs have a bihelical conformation. The shaded regions of mRNA denote untranslated sequences.

Fig. 5 Structure of phenylalanyl-transfer RNA. (a) Secondary cloverleaf structure. The 5′ and 3′ ends of the molecule are marked; the continuous line represents the sugar phosphate backbone. The short lines denote base residues and the dots denote base pairing through standard hydrogen bonds (see Fig. 1). The D loop contains dihydrouracil residues; the TψC loop contains thymine and pseudouridine. (b) Tertiary structure. The abbreviations for unusual bases are defined in (a). The sugar phosphate backbone is represented by double parallel lines. Standard base pairs are represented by short double parallel lines, and nonstandard base pairs by single lines. The anticodon sequence is stippled, and the acceptor end is shaded. [From Arnstein, H.R.V., Cox, R.A. (1992) *Protein Biosynthesis*, Oxford University Press, Oxford. With permission].

translation process. Once the aminoacyl tRNA has been formed, the subsequent incorporation of the amino acid residue into a polypeptide does not depend on the amino acid itself but only on the interaction between the anticodon of the aminoacyl tRNA with the codon of mRNA. Thus, an error in the synthetase reaction would lead to the incorporation of an inappropriate amino acid into the polypeptide.

Synthesis of aminoacyl tRNA (**III**) from amino acids (**I**) requires activation of the amino acid carboxyl group with formation of an intermediate enzyme-bound aminoacyladenylate (**II**) (Scheme 1).

The energy for the reaction (two high-energy phosphate bonds) is provided by ATP and stored in the ester bond of the aminoacyl tRNA to be used subsequently for peptide bond synthesis.

4 Ribosome Structure and Function in Translation

4.1 Ribosome Structure

Ribosomes are high molecular–weight complexes of RNA (rRNA) and proteins (Table 1), and the electron-dense particles are easily visualized within cells by electron microscopy (see Fig. 7). Usually, several

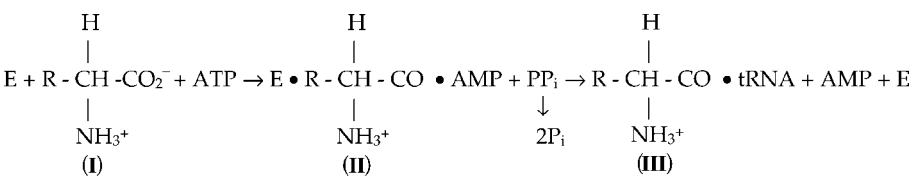

Where E = aminoacyl tRNA synthetase

Scheme 1

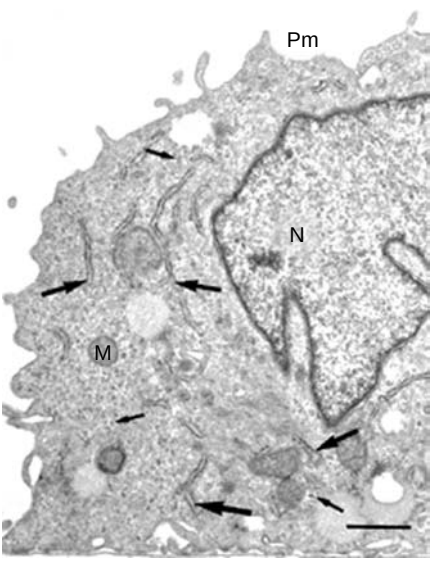

Fig. 7 Electron micrograph of a thin section through part of an epithelial cell showing plasma membrane (Pm), nucleus (N), mitochondria (m), and rough endoplasmic reticulum (large arrows) to which are attached numerous polysomes. Groups of free ribosomes (small arrows) occur in the cytoplasm. Bar represents 1000 nm. [Micrograph by I. D. J. Burdett, National Institute for Medical Research, London, UK.]

ribosomes become attached to one mRNA molecule giving rise to polyribosomes (also called *polysomes*; see Fig. 7). Ribosomes from various sources (prokaryotes, eukaryotic cytoplasm, mitochondria, chloroplasts, and kinetoplasts) vary in size from 20 to 30 nm in diameter, but all are composed of a large and a small subparticle or subunit and perform similar functions in protein synthesis. In general, cytoplasmic ribosomes of eukaryotes are larger than ribosomes of prokaryotes (Table 1). The principal functional domains of the ribosome and associated components are given in Fig. 8. More detailed resolution of the ribosome structure has allowed the

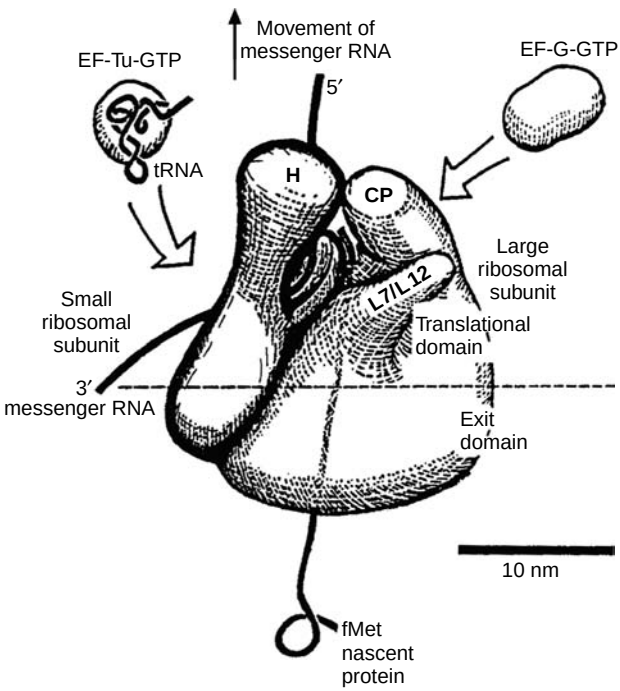

Fig. 8 Model of the *E. coli* ribosome, based on low resolution electron microscopy. The diagram shows the relative orientation of the large and small subunits and other functional components involved in polypeptide chain synthesis. H, head of the small subunit; CP, central protuberance and L7/L12 stalk of the large subunit; EF-Tu, elongation factor required for binding aminoacyl tRNA to the ribosomal A site; EF-G, elongation factor required for translocation of the peptidyl-tRNA from the A to the P site. The broken line indicates the boundary between the translational and exit domains that are involved in peptide bond formation and extrusion of the nascent polypeptide chain, respectively. Further details of ribosome structure and function are given in Fig. 11 for peptide bond formation, in Figs. 12 and 13 for the delivery of aminoacyl tRNA to the A site and Figs. 14 and 15 for translocation and Fig. 16 for the location of the nascent polypeptide chain. [From Arnstein, H.R.V., Cox, R.A. (1992) *Protein Biosynthesis*, Oxford University Press, Oxford. With permission.]

Tab. 1 Properties of ribosomes and ribosomal subunits.

Source	$S_{20,w}$	*Size [nm]*[a]	*Mass [Mda]*	*RNA:protein [w/w]*	*Axial ratio*	*rRNA mass [Mda]*	$S_{20,w}$ *(nominal)*
Ribosomes							
Prokaryotes							
E. coli	70	22.5 ± 2.5	2.6–2.9	2 : 1	– [f]		
Eukaryotes							
Cytoplasm	80	28.0 ± 2.8	3.4–4.5	1 : 1	–		
Chloroplast	70	22.5 ± 2.5	2.5–3.3	1 : 1	–		
Mitochondria[b]							
Protists	60–80		ca. 3.25	1 : 1.38	–		
Animals	55–60		2.7–3.2	1 : 1.7–1 : 4	–		
Fungi	67–80		4.2	1 : 1.13–1 : 1.6	–		
Higher plants	78		–	–	–		
Small ribosomal subunits							
Prokaryotes							
E. coli	30	22.0 ± 2.2	0.95	1 : 1	2 : 1	0.6	16
Eukaryotes							
Cytoplasm	36–41	25.0 ± 2.5	1.4	1 : 1	2 : 1	0.7	18
Chloroplast	28–35	22.0 ± 2.2	1.2	1 : 1	2 : 1	0.6	16
Mitochondria[b]							
Protists	32–45		–	–	–	0.2–0.35	11–16
Animals	32–40		1.1–1.6	–	–	0.31	12–14
Fungi	32–29		1.6–1.7	–	–	0.45–0.64	15–19
Higher plants	44		–	–	–	0.64	18

Large ribosomal subunits							
Prokaryotes							
E. coli	50	22.5 ± 2.2	1.75	2 : 1	1 : 1	1.17	23
						0.04	5
Eukaryotes							
Cytoplasm	60	28.0 ± 2.8	2.1–3.1	1 : 1	1 : 1	1.2–1.75	25–28
						0.05	5.8
						0.4	5
Chloroplast	46–54	22.5 ± 2.2	2.4	1 : 1	1 : 1	1.1	23
						0.04	5
						0.03[d]	4.55[d]
Mitochondria[b]							
Protists	45–60		–	–	–	0.4–0.74	12–24
Animals	25–35		1.65–2.0	–	–	0.5	16–21
Fungi	50		–	–	–	0.77–1.22	21–25
Higher plants	60		–	–	–	1.25	26
						0.04[e]	5[e]

[a]Largest dimension.
[b]Mitochondrial ribosomes are preferentially associated with the inner mitochondrial membrane and are more diverse than cytoplasmic or chloroplast ribosomes. Particular features include posttranscriptional oligoadenylation of the 3′ ends of both the smaller and larger rRNA components. Except for mitochondrial ribosomes of higher plants, the essential sequence motifs 5S rRNA and 5.8S rRNA are present in the large rRNA component and are not found as individual species.
[c]Mitochondria of protists such as *Trypanosoma brucei* are also known as kinetoplasts.
[d]4.5S rRNA corresponding to the 100 nucleotides at the 3′ end of the 23S rRNA in eubacteria occurs as a separate species in chloroplasts of higher plants.
[e]This species is present only in plant mitochondria and is absent from all other mitochondrial ribosomes.
[f]Data not available.

placement of mRNA, aminoacyl tRNA, peptidyl tRNA, and the nascent polypeptide chain (see Sect. 4.3).

The small subunit comprises a single rRNA of $0.3–0.7 \times 10^6$ Da and single copies of 20 to 30 unique proteins. It has a major function in binding initiator tRNA and mRNA in the initiation of protein synthesis and in decoding the genetic message.

The large subunit comprises a high molecular–weight rRNA ($0.6–1.7 \times 10^6$ Da) and often one or two smaller rRNAs ($0.03–0.05 \times 10^6$ Da) and 30 to 50 different proteins are present, with one exception, as single copies. The large subunit binds aminoacyl tRNA at the A site, peptidyl tRNA at the P site, and discharged tRNA at the E (exit) site. The large subunit contains the peptidyl transferase center and, unusually, this enzyme activity resides in the rRNA molecule itself rather than in the associated ribosomal proteins. This subunit is also involved in binding elongation factor G, which is required for translocation (Fig. 8).

4.2 The Ribosome Cycle in Translation

Polypeptide synthesis can be divided into three stages: initiation, elongation, and termination. Initiation involves the binding of a ribosome to mRNA with the initiation codon correctly placed in the P site. Elongation leads to the stepwise increase in the length of the polypeptide chain through the transfer of the growing chain to the amino group of aminoacyl tRNA. Termination of chain elongation and release of the completed polypeptide occurs when a termination codon reaches the A site. All stages require the participation of protein factors. Advances in establishing the structures of translational factors by X-ray crystallography have been rapid in the last decade.

4.2.1 Formation of Preinitiation Complexes

The ribosome cycle starts with the stepwise formation of an initiation complex from mRNA, charged initiator tRNA, and ribosomal subunits. A number of preinitiation complexes are formed as intermediates and the process is facilitated by initiation factors. In outline, prokaryotic and eukaryotic systems are similar, but there are a few differences, particularly as regards the complexity of the initiation factors and details of the mechanisms.

Prokaryotic systems. Three proteins, initiation factors IF-1, IF-2, and IF-3 (see Table 2), are required for the initiation of protein biosynthesis (see Fig. 9a). Ribosomal subunits are released by dissociation of ribosomes following translation of the mRNA. Dissociation is facilitated by the combined action of the initiation factors IF-1 and IF-3; IF-1 increases the rate of dissociation and IF-3 acts as an antiassociation factor when bound to the (30S) ribosomal subunit, thereby displacing the equilibrium in favor of subunit formation. Initiation factor IF-2 is also able to bind to the small 30S subunit and this association is stabilized by IF-1 and GTP, the latter acting as a steric effector without being hydrolyzed at this stage. IF-2 plays a central role in binding fMet-$tRNA_f$ to the 30S preinitiation complex by specific recognition of the *N*-formylmethionine residue attached to the initiator tRNA, thus restricting this interaction to charged initiator tRNA. All three factors bind to the 30S ribosomal subunit near the 3′ end of the 16S ribosomal RNA at adjacent sites that are

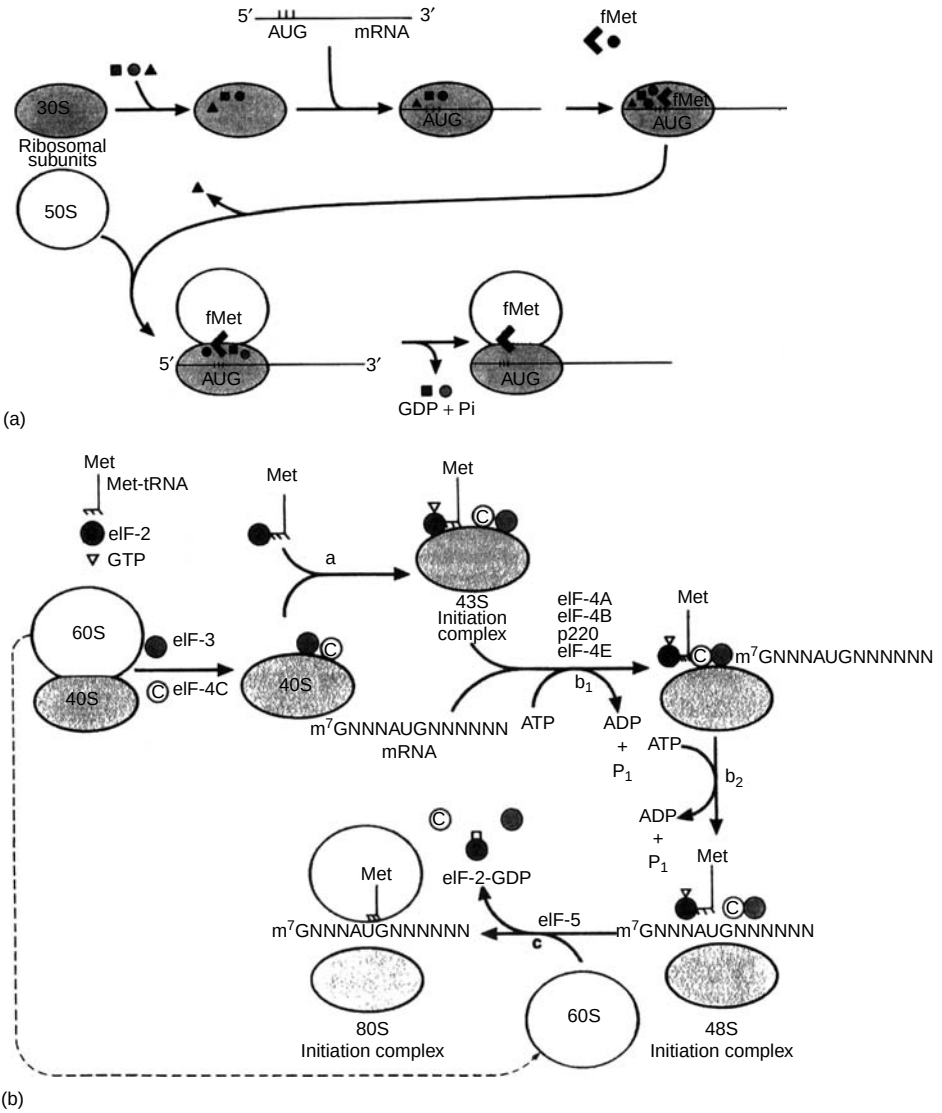

Fig. 9 Schematic diagram illustrating the formation of initiation complexes, (a) prokaryotic initiation. The following symbols are used for initiation factors and other components involved in the formation of the 70S initiation complex: ■, IF-1; ●, IF-2; ▲, IF-3; <, fMet-tRNA$_f$; ▽, GTP; other symbols are defined in the figure. (b) Eukaryotic initiation complexes in mammalian protein biosynthesis from components defined in the figure. The process may be divided into three stages: **a**, formation of a 43S preinitiation complex; **b**, binding of mRNA with formation of a 48S preinitiation complex; and **c**, synthesis of the 80S initiation complex containing the initiator tRNA in the correct position for peptide bond formation. [From Arnstein, H.R.V., Cox, R.A. (1992) *Protein Biosynthesis*, Oxford University Press, Oxford. With permission.]

Tab. 2 Prokaryotic initiation factors from *E. coli*.

Factor	M_r *[kDa]*	*Properties and function*
IF-1	9	Stimulates activity of IF-2; accelerates dissociation of unprogrammed ribosomes to subunits.
IF-2	100	Binds fMet-tRNA$_f$ to the ribosomal P site by a GTP-requiring reaction.
If-3	22	Binds natural mRNAs to the small ribosomal subunit probably by facilitating base pairing between the untranslated leader sequence and the 3′ end of 16S rRNA; prevents ribosomal subunit association when bound to the small subunit.

From Arnstein, H.R.V., Cox, R.A. (1992) *Protein Biosynthesis*, Oxford University Press, London. With permission.

located at the interface between the small and large ribosomal subunits.

In the next step, the initiator tRNA and mRNA associate with the 30S–IF-1–IF-2–IF-3 complex with release of IF-3. There is evidence from *in vitro* experiments that the binding of mRNA precedes that of the initiator tRNA.

Messenger RNA binds to the small ribosomal subunit immediately before formation of the final initiation complex with the initiation codon correctly positioned in the P site (see Fig. 9a). In the case of bacterial and bacteriophage messengers, the molecular recognition mechanism proposed by Shine and Dalgarno in 1974 involves base pairing between short nucleotide sequences, most often CUCC, near the 3′ end of the 16S ribosomal RNA and a complementary region, usually consisting of 3 to 9 bases on the 5′ side of the mRNA initiation codon, which has been found to be present in nearly all of more than 150 bacterial and bacteriophage messengers. Studies with mutants and mRNA fragments indicate that, in addition to the Shine–Dalgarno interaction, outlying upstream sequences in the leader region may also provide recognition signals between mRNAs and ribosomes, possibly by ensuring that the Shine–Dalgarno sequence is in an appropriate conformation.

The Shine–Dalgarno mechanism is also found in chloroplast protein synthesis as judged from sequence analysis of the 16S rRNA and mRNAs, but apparently not in mammalian mitochondria where the initiator codon occurs either directly at, or only a few nucleotides downstream from, the 5′ end of mRNA, which excludes the possibility of mRNA–rRNA base pairing in this region.

Eukaryotic systems. At least 12 proteins, the eukaryotic initiation factors (eIF) (see Table 3), are needed for initiation of protein biosynthesis (see Fig. 9b). The dissociation of cytosolic 80S ribosomes is facilitated by a complex initiation factor, eIF-3 (M_r approx. 5–700 000), consisting of 9 to 11 polypeptide chains, which binds to the small (40S) ribosomal subunit and prevents its reassociation to 80S ribosomes. Thus, this factor has antiassociation activity, but low molecular–weight proteins with similar activity have also been reported, and a protein, eIF-4C, of M_r 20 000, seems to function as an accessory factor to eIF-3 in the formation of a 43S ribosomal preinitiation complex. Also, another protein factor, eIF-6, of M_r 24 000, prevents

Tab. 3 Eukaryotic initiation factors.

Initiation factor	*Synonym*	*M_r (kDa) of factors and subunits*	*Properties and function*
eIF-1		15	Stabilizes initiation complexes
eIF-2		α, 38[a] β, 35[a] γ, 55	GTP-dependent binding of Met-tRNA$_f$ to the small ribosomal subunit
eIF-2B	GEF	27, 37, 52, 67, 85[a]	Conversion of eIF-2-GDP into eIF-3-GTPs
eIF-3		9–11 subunits 24–170[a]	Associates with 40S subunit to maintain dissociation; binds mRNA to 43S preinitiation complex
eIF-4A	50 kDa component of CBP-II	50	ATP-dependent unwinding of the secondary structure of the mRNA 5′ region; stimulates translation of exogenous mRNA in cell-free systems
eIF-4B		80[a]	mRNA binding; stimulates cell-free translation; ATPase activity of eIF-4A and eIF-4F; AUG recognition and recycling eIF-4F
eIF-4C		17	Ribosome dissociation; 60S subunit joining
eIF-4D		17	Formation of first peptide bond
eIF-4E	CBP-1; 24 kDa CBP	24–28[a]	Binds mRNA cap structure
eIF-4F	CBP-II; cap binding protein complex	24 (CBP-I)[a] 50 (eIF-4A), 200[a]	ATPase; unwinds mRNA secondary structure; stimulates cell-free translation
e-IF-4G		220	Stimulates protein synthesis by interacting with eIF-4E and poly(A) protein to circularize polysomes
eIF-5		60[a]	GTPase; release of eIF-2 and eIF-3 from preinitiation complex to allow joining of the 60S subunit
eIF-6		24	Antiassociation activity; binds o the 60S ribosomal subunit.

[a]Denotes subunit can be phosphorylated *in vivo*.
Note: GEF: guanine nucleotide exchange factor; CBP: cap binding protein; From Arnstein, H.R.V., Cox, R.A. (1992) *Protein Biosynthesis*, Oxford University Press, London. With permission.

reassociation by binding to the large (60S) ribosomal subunit.

Initiation factor eIF-2 gives a stable binary complex with GTP which binds the initiator tRNA, Met-tRNA$_f$, forming a ternary complex. Interaction of this ternary complex with the 40S ribosomal subunit containing bound initiation factors eIF-3 and eIF-4C gives rise to the 43S preinitiation complex, which is competent to bind messenger RNA in the presence of three further initiation factors, eIF-4A, eIF-4B, and eIF-4E, together with ATP.

The binding of cytosolic eukaryotic messenger RNAs to the small ribosomal subunit probably does not involve base-paring with the 18S rRNA, as no uninterrupted sequences of the Shine–Dalgarno type have been found. Instead, a "scanning model" has been proposed, in which the preinitiation complex, composed of the 40S ribosomal subunit, Met-tRNA$_f^{Met}$ and associated initiation factors binds at or near the 5′ cap of the mRNA and slides along the messenger until it encounters the first AUG triplet, at which point the 60S ribosomal subunit joins to give rise to the 80S initiation complex. Recognition of the cap is facilitated by cap binding proteins (CBP), which mediate an ATP-dependent melting of the mRNA secondary structure at the 5′ terminal region to allow the mRNA to thread through a channel in the neck of the 40S subunit. The cap structure is required for efficient binding and translation even in cases where the initiating AUG codon occurs hundreds of nucleotides downstream. As a rule, scanning by the 40S subunit stalls at the first AUG codon, which is recognized mainly by interaction with the anticodon of the Met-tRNA$_f^{Met}$. However, this recognition also depends in some way on eIF-2 and may be modulated as a result of deviation of the mRNA structure from the consensus sequence GCCGCCA/G CCAUGG. Sequence context may also account for the rare cases where initiation occurs downstream of the 5′ proximal AUG codon. Where this sequence context is unfavorable, initiation becomes inefficient, hence most 40S subunits will tend to initiate further along the mRNA at another AUG triplet in a more favorable context. This model also explains rare cases where initiation is not restricted to one particular AUG codon and translation of a single mRNA gives rise to two proteins.

4.2.2 Initiation Complex Formation: Joining of the Large Ribosomal Subunit

The last event in the initiation of protein synthesis involves the joining of the large ribosomal subunit to the preinitiation complex (Fig. 9). In the prokaryotic system, association of the 50S subunit with the 30S preinitiation complex takes place with hydrolysis of GTP by the GTPase activity of IF-2 and release of IF-1, IF-2, GDP, and P_i. GTP hydrolysis is essential for release of IF-2 from the initiation complex, which is prerequisite for allowing the fMet-tRNA$_f$ to engage in the formation of the first peptide bond. In eukaryotic protein synthesis, the 80S initiation complex is formed by joining the 60S ribosomal subunit to the 48S preinitiation complex consisting of the 40S ribosomal subunit, eIF-2, eIF-3, GTP, Met-tRNA$_f$, mRNA, and possibly eIF-4C. This coupling reaction requires an additional factor, eIF-5, which mediates the hydrolysis of GTP to GDP with release of eIF-2-GDP, P_i, and eIF-3 from the 48S preinitiation complex.

By this stage, all initiation factors have been released and are available for recycling, although the exact steps at which factors are released from intermediate complexes are not known in every case.

There is thus an initiation factor cycle within the ribosome cycle, and regulation of the activity of factors, particularly eIF-2, is an important control mechanism in translation (see Sect. 5.4.1).

4.2.3 Polypeptide Chain Synthesis: The Elongation–Translocation Cycle

In the initiation complex, location of the charged initiator tRNA in the P site of the ribosome (see Fig. 10a) allows transfer of the methionine residue to the amino group of another aminoacyl tRNA in the A site (Fig. 10b) by peptidyl transferase to form dipeptidyl tRNA (see Fig. 10c). Functional insertion of Met-$tRNA_f$ directly into the P site can be demonstrated using the trinucleotide AUG as a synthetic mRNA and another trinucleotide, for example, UUU, to bind an acceptor aminoacyl tRNA (in this case Phe-tRNA).

It is possible to measure the peptidyl transferase activity of the large subunit in the absence of mRNA by using the antibiotic puromycin, which resembles the 3′ terminal region of Phe-tRNA in structure; as an artificial acceptor it forms methionyl puromycin with Met-$tRNA_f$, which can be assayed.

Elongation. The first peptide bond is formed when the aminoacyl tRNA in the ribosomal A site is converted into the corresponding methionyl-aminoacyl-tRNA by transfer of the methionyl (or *N*-formylmethionyl) residue from the charged initiator tRNA in the P site (Fig. 10c). In artificial cell-free systems, any *N*-substituted aminoacyl tRNA, such as peptidyl tRNA or *N*-acetylaminoacyl-tRNA, can function in peptide bond synthesis as a donor in the P site in place of the charged initiator tRNA. The reaction is catalyzed by the peptidyltransferase activity of the large ribosomal subunit (see Fig. 11). No soluble cofactors appear to be involved, but monovalent cations (K^+) at a concentration of 100 mM or more and divalent cations (Mg^{2+}) below 2 mM are required.

Efficient entry of aminoacyl tRNA into the ribosomal A site requires the participation of an elongation factor, termed EF-Tu in prokaryotes (see Table 4) and eEF-1 (EF-1A) in eukaryotes (see Table 5), and GTP. The appropriate elongation factor forms a ternary complex with GTP and all aminoacyl tRNAs except initiator

Tab. 4 Properties of prokaryotic elongation and termination factors from *E. coli*.

	M_r [kDa]	Properties and function
Elongation factors		
EF-Tu	43	N-terminal acetyl-serine; heat labile; binds aminoacyl tRNA to the ribosomal A site
EF-Ts	30	Heat stable; regeneration of EF-Tu-GTP
EF-G	77	GTP-dependent translocation of peptidyl-tRNA and its mRNA codon from the A site to the P site.
Termination (release) factors		
RF1	36	Requires UAA or UAG codons for hydrolysis of peptidyl-tRNA
RF2	38	Requires UAA or UGA codons for hydrolysis of peptidyl-tRNA
RF3	46	Enhances RF1 and RF2 activity

From Arnstein, H.R.V., Cox, R.A. (1992) *Protein Biosynthesis*, Oxford University Press, London. With permission.

tRNA, but not with uncharged tRNA, thus ensuring that only appropriately charged tRNAs are efficiently bound in the A site. A special elongation factor showing extensive homology with both EF-Tu and IF-2 is involved in the synthesis of selenoproteins (see Sect. 2.4.1) from selenocysteyl-$tRNA^{UCA}$ in *E. coli*.

The above-mentioned aminoacyl tRNA binding reaction catalyzed by EF-Tu is the

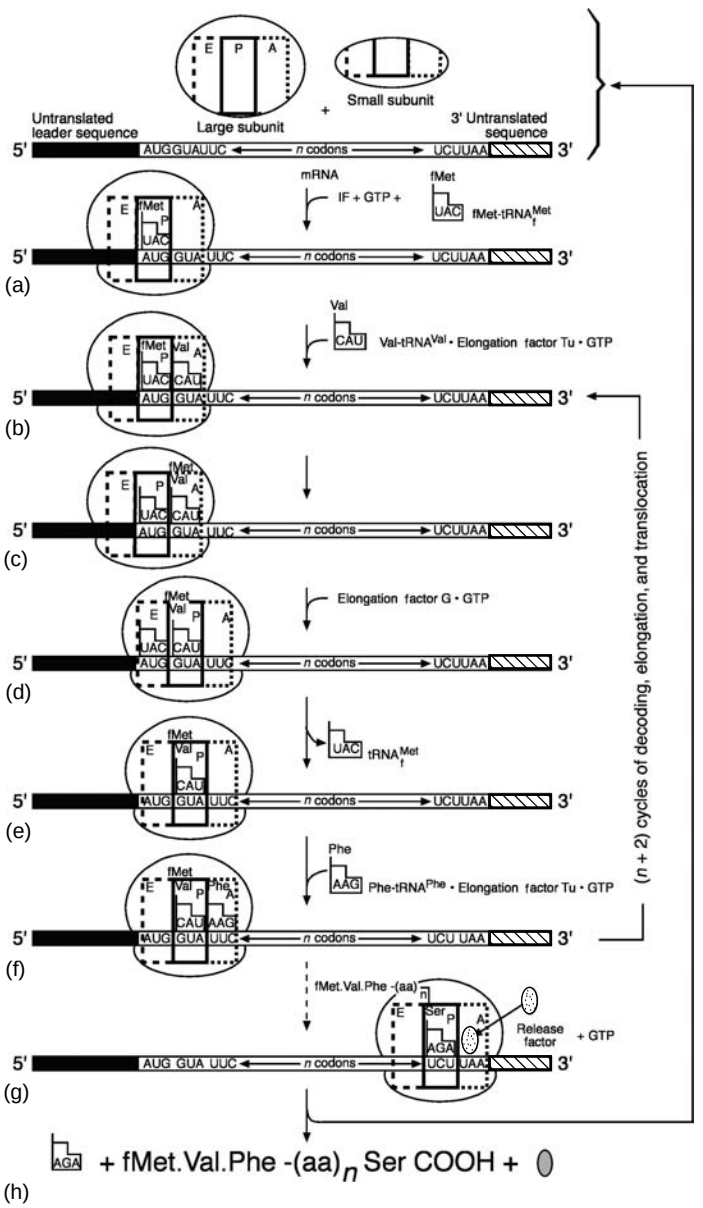

rate-limiting step in the elongation cycle; peptide bond formation and translocation are much faster. The initial binding of the ternary complex to the ribosome is readily reversed, but the interaction is stabilized by the subsequent codon recognition, which induces the GTPase conformation of EF-Tu leading immediately to the hydrolysis of the GTP component of the ternary complex to GDP. Hydrolysis of the GTP moiety causes a further change in the conformation of EF-Tu from the GTP-binding to the GDP-binding form. This conformational change leads to the release of aminoacyl tRNA, allowing its CCA end to align with the peptidyl transferase center of the ribosome and the instantaneous formation of the peptide bond. The elongation factor is later released from the ribosome as a complex with GDP. The operation of EF-Tu is thus similar to that of initiation factor 2 which binds charged initiator tRNA to the small ribosomal subunit. The kinetic changes and the changes in ribosome structure that accompany these events are enumerated below (see Sect. 4.3).

Following dissociation from the ribosome, the EF-Tu–GDP complex interacts with another elongation factor, EF-Ts,

Tab. 5 Eukaryotic elongation and termination factors.

	M_r [kDa]	Properties and function
Elongation factors from various yeast, animal, and plant cells		
eEF-1A (EF-1_L or eEF-Tu)	50–60	Analogous to EF-Tu
eEF-1B (eEF-Ts)	30	Analogous to EF-Ts
eEF-2	105	Contains essential SH groups and one residue of a posttranslationally modified histidine residue; GTP-dependent translocation analogous to EF-G
eEF-3	126	GTPase and ATPase activity; function not fully defined
Termination or release factors		
eRF-1	110	Two 55-kDa subunits; binds to the ribosome A site by a GTP and termination codon dependent reaction; hydrolyzes peptidyl-tRNA in the P site
eEF-2		GTPase; stimulates eRF-1 activity

From Arnstein, H.R.V., Cox, R.A. (1992) *Protein Biosynthesis*, Oxford University Press, London. With permission.

Fig. 10 Schematic view of the biosynthesis of a polypeptide by translation of prokaryotic mRNA. (a) Formation of the initiation complex with fMet-$tRNA_f^{Met}$ in the P site from mRNA, ribosomal subunits, initiation factors (IF), GTP, and fMet-$tRNA_f^{Met}$ (for details, see Fig. 8). The binding sites for aminoacyl tRNA, peptidyl-tRNA, and uncharged tRNA are designated A, P, and E, respectively. (b) Decoding of the codon GUA with Val-$tRNA^{Val}$. (c) Formation of the peptide bond by transfer of fMet to Val-$tRNA^{Val}$ forming fMet-Val-$tRNA^{Val}$. (d) Translocation of $tRNA_f^{Met}$ to the E site and of fMet-Val-$tRNA^{Val}$ to the P site with codon UUC aligned with the A site. (e) Ejection of $tRNA_f^{Met}$ from the ribosome. (f) Decoding of UUC with Phe-$tRNA^{Phe}$. (g) Decoding of UAA with release factor. Steps (b) to (f) constitute the elongation–translocation cycle. The position shown is reached by repeating the cycle $n + 2$ times after formation of fMet.Val-tRNA. (h) Release of the completed polypeptide, fMet.Val.Phe-$(aa)_n$Ser COOH, ribosomal subunits, release factor, mRNA and $tRNA^{Ser}$. The cycle may be repeated starting at (a).

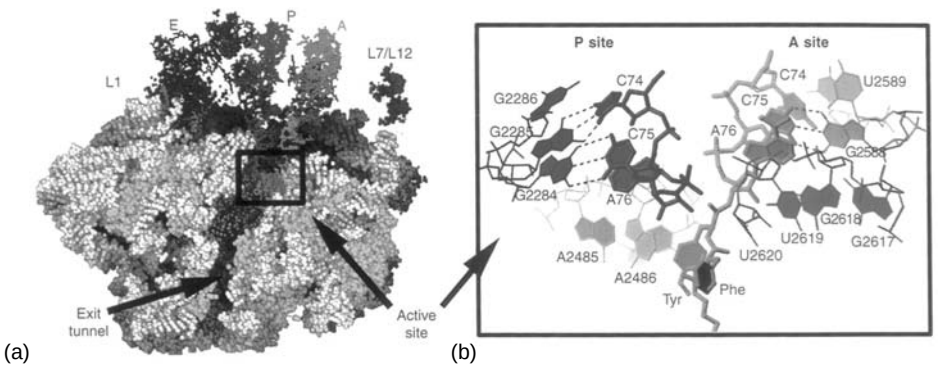

Fig. 11 Fine structure of the peptidyl transferase center interactions of the CCA ends (C74, C75, and A 76) of aminoacyl tRNA and peptidyl-tRNA with the rRNA moiety of the large subunit. Aminoacyl tRNA was represented by a C74 puromycin derivative of tRNA and peptidyl-tRNA was represented by CCA-phenylalanine-caproic acid-biotin. When these model compounds were bound to the A and P sites respectively, the large subunit catalyzed the formation of a peptide bond leading to newly extended peptidyl-tRNA (C74 puromycin–phenylalanine-caproic acid-biotin) in the A site and a deacylated tRNA in the P site. (a) a space filling representation of the 50S subunit (RNA in white and protein in yellow). The tRNAs in the A (green), P (purple), and E (brown) sites are placed according to the structure of the complete ribosome. The subunit has been split though the tunnel and the front section was removed to reveal the peptidyl transferase center (boxed). (b) Base-pairing interactions of CCA sequences of tRNA with 23S rRNA. The analog of deacylated tRNA (purple) is bound to rRNA through base pairing between C74 with G2285) and C75 with G2284. The analog of the newly formed peptidyl-tRNA is bound through base pairing between C75 and G1588. [From Schmeing, et al. (2002) A pretranslocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits, *Nat. Struct. Biol.* **9**, 225–230. With permission.] (see color plate p. xxxiii).

with formation of an EF-Tu–EF-Ts heterodimer and release of GDP. Reaction of the heterodimer with GTP regenerates the EF-Tu–GTP complex required for binding aminoacyl tRNA. The sequence of events is similar in eukaryotes with eEF-1A (M_r 50 000) corresponding to EF-Tu and eEF-1B (M_r 30 000) to EF-Ts.

Selection of the specific aminoacyl tRNA to be bound at the ribosomal A site is by base pairing between the relevant mRNA codon and the tRNA anticodon. Because this interaction involves only a triplet of bases and hence a maximum of nine hydrogen bonds (see Fig. 1B), it is intrinsically unstable at physiological temperatures and is stabilized by a network of interactions with 16S rRNA in the decoding site where bases of 16S rRNA interact with the codon–anticodon complex. It is inferred that these interactions allow sufficient time for peptide bond synthesis to occur. Also, the codon–anticodon pairing must be monitored for fidelity in order to minimize errors in translation. In *E. coli*, there is genetic and biochemical evidence that one of the proteins of the small ribosomal subunit, S12, is involved in ensuring the fidelity of normal translation and in causing the mistranslation that occurs in the presence of the antibiotic streptomycin due to incorrect codon–anticodon interactions.

Translocation. Translocation involves the movement of the ribosome along the

mRNA in the 5′ → 3′ direction. Immediately after synthesis of the first peptide bond, the ribosomal A site contains dipeptidyl tRNA while uncharged initiator tRNA remains in the P site. Thus, both these sites are occupied, and to allow the next aminoacyl tRNA to enter the A site it is necessary to eject the uncharged tRNA and shift the dipeptidyl tRNA from the A into the P site. This translocation (Fig. 10d) takes place as a concerted process involving movement of both messenger RNA and dipeptidyl RNA together into the P site, leaving the A site occupied by the next mRNA codon and free to accept the cognate aminoacyl tRNA (see Fig. 10e). At the same time, the deacylated tRNA moves first into an E (exit) site with subsequent ejection when the next aminoacyl tRNA enters the A site.

Translocation requires the participation of another elongation factor (EF-G in prokaryotes (Table 4) and EF-2 in eukaryotes (Table 5)) and GTP. It seems that when EF-G and GTP bind to the ribosome, translocation occurs but GTP hydrolysis is required only subsequently to release EF-G and GDP. The location of the EF-G binding site of the ribosome overlaps with that for EF-Tu; thus, EF-G must be released before EF-Tu–aminoacyl-tRNA–GTP complex can enter the A site. Analogous reactions occur in eukaryotic systems. The kinetic steps involved in translocation and the structural changes that take place are described below (Sect. 4.3).

After translocation, the ribosomal P site is occupied by dipeptidyl tRNA and the vacant A site contains the third mRNA codon. Entry of the next aminoacyl tRNA, selected as before by the codon–anticodon interaction into the A site (Fig. 10f) enables peptide bond synthesis to continue and repeated operation of the elongation–translocation cycle gives rise to a stepwise elongation of the nascent polypeptide chain, each complete cycle elongating the chain by one amino acid residue and moving the mRNA by one codon in the 5′ to 3′ direction. When the end of the coding sequence is reached, one of the termination (or stop) codons has entered the A site, translation stops, and the completed polypeptide chain is released.

Termination. (See Figs. 10g-h.) The presence of one of the three termination codons, UAA, UAG, or UGA, in the A site results in the binding of a release factor (Table 4) instead of an aminoacyl tRNA to the ribosome. In prokaryotes, two release factors have been identified, one (RF1) recognizing UAA and UAG, the other (RF2) functioning with UGA. Ribosomal binding and release of RF1 and RF2 are stimulated by a third factor, RF3, which interacts with GTP and GDP. In eukaryotic cells, such as reticulocytes, one release factor (eRF) has been found to function with all three termination codons, and the binding of this factor to ribosomes is stimulated by GTP but not GDP. Although the details are not entirely clear, GTP hydrolysis appears to be required for the release of the finished polypeptide chain by cleavage of the peptidyl tRNA bond and completion of the termination process leading to dissociation of the release factor from the ribosome.

Thus, at the end of the ribosome cycle, the coding sequence of messenger RNA has been translated to produce a particular polypeptide chain, and all the components involved become available for reuse in another round of the cycle (Fig. 10h). Usually, several ribosomes, each at a different stage in completing the process of translation, are attached to one mRNA molecule, giving rise to polyribosomes (also called *polysomes*; Fig. 7). In eukaryotic

cells, the efficiency of protein synthesis is stimulated by factor eIF4-G (Table 3), which interacts with both factor eIF-4E and a poly A-binding protein to link the 5′ and 3′ ends of mRNA. The resultant circularized polysomes show an enhanced ability to reinitiate after release of the ribosomal subunits from the messenger RNA at the end of a round of translation.

4.3 High-resolution Structural Studies of the Ribosome

Early structural studies based on electron microscopy provided sufficient information to allow the construction of static models showing the location of mRNA and the relative orientations of the ribosomal subunits (Fig. 8). The modern technique of high-resolution cryo-electron microscopy (cryo-EM) has provided views of the ribosome at a level of resolution (0.9 nm) approaching that achieved by X-ray diffraction studies. Although prokaryotic ribosomes are the principal subjects of structural studies because they are simpler than eukaryotic ribosomes, the insights into structure and function are thought to apply to all ribosomes.

During the last 10 years, X-ray diffraction studies have led to considerable advances in the elucidation of the structure of the ribosome in more detail, culminating in the determination of the *E. coli* ribosome at 0.78-nm resolution and of the small and large subunits at resolutions of 0.30 and 0.25 nm respectively. The structures have revealed the identity of each amino acid and each nucleotide. The findings provide insights at the atomic level into the reactions leading to the decoding of mRNA and the formation of the peptide bond (Fig. 11). Moreover, the importance of the role of rRNA in ribosome function has become evident from the structures; the functional regions of both small and large subunits are rich in RNA. The peptide bond is formed by nucleophilic attack on the ester carboxyl group of peptidyl tRNA bound to the site by the α-amino group of aminoacyl tRNA in the A site. Peptide bond formation is catalyzed by the peptidyl transferase center of the large subunit of the ribosome. Structural studies confirm that this catalytic site comprises 23S rRNA sequences only. The location of the peptidyl transferase center and the role of the CGA ends of tRNA are illustrated in Fig. 11. The three-dimensional structures also highlight the dynamic aspects of ribosome function leading to the view that the ribosome is a highly sophisticated motor driven by GTP with rRNA playing a leading role.

The first of the two GTP requiring steps is the delivery of aminoacyl tRNA to the A site of the ribosome. The distinct kinetic steps involved as the ternary complex of aminoacyl tRNA, EF-Tu and GTP binds to the ribosome are illustrated in Fig. 12 and the conformational changes that have been found to take place are depicted in Fig. 13. Please note that one molecule of GTP is converted to GDP irrespective of whether peptide bond formation ensues. The peptide bond is formed soon after the selection of the aminoacyl tRNA (the accommodation step in Fig. 12), leading to the newly formed and extended peptidyl tRNA residing in the A site. The second GTP requiring step then follows; namely, translocation of aminoacyl tRNA from the A site to the P site of the ribosome. The kinetically resolved stages of the translocation are shown in Fig. 14 and the conformational changes so far identified are illustrated in Fig. 15.

The ribosome also influences the secondary and tertiary structure (folding)

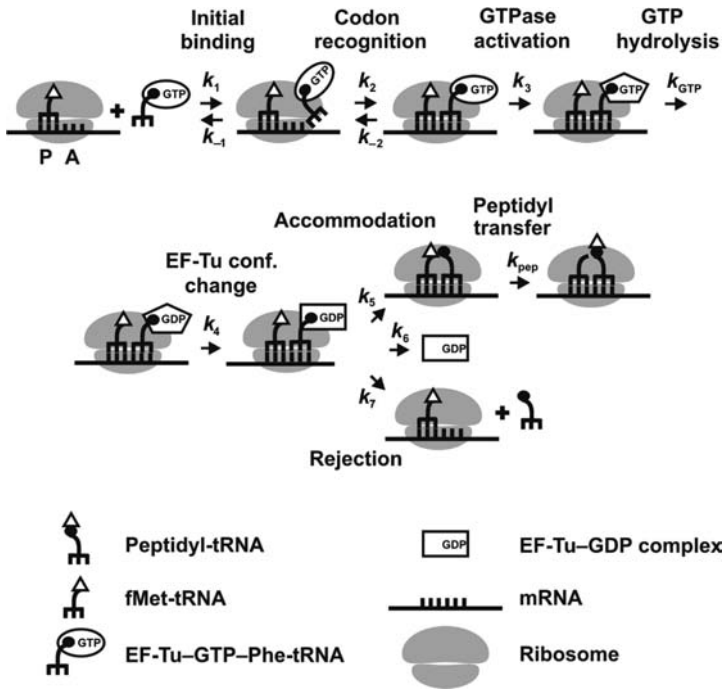

Fig. 12 Kinetically resolved stages in the interaction of ternary complex of aminoacyl tRNA•EF-Tu•GTP with the A site of the ribosome. Individual steps are shown by the numbered rate constants and by k_{GTP} and k_{pep}. The factor EF-Tu is shown in different conformations in GTP and GDP bound states and in the activated GTPase state. The ribosome is able to distinguish between cognate aminoacyl tRNA (for example, with anticodon 3′AAG 5′) and near cognate aminoacyl tRNA (for example, with anticodon 3′GAG 5′): this ability is called proof reading. [From Gromadski, K.B., Rodnina, M.V. (2004) Streptomycin interferes with conformational coupling between codon recognition and GTPase activation on the ribosome, *Nat. Struct. Mol. Biol.* **11**, 316–322. With permission.]

of the nascent polypeptide chain during its passage through the 10-nm long × 1 to 2 nm-diameter tunnel that is present in the large ribosomal subunit (Fig. 16). The wall of the tunnel comprises nucleotides of 23S rRNA and nonglobular parts of two ribosomal proteins, L4 and L2; proteins L23 and L29 flank the exit from the tunnel. Thus, the nascent polypeptide chain passes through an environment that influences the way in which it folds. As nascent polypeptide chains emerge from the tunnel, they are free to encounter molecular chaperones that prevent accumulation and aggregation of newly synthesized proteins. For example, the chaperone Trigger factor (TF) has a binding site involving protein L23 close to the exit of the tunnel. Signal recognition particles that convey secretory and transmembrane proteins to their appropriate destinations have bonding sites near to the exit of the tunnel in the regions of proteins L23 and L29.

The higher resolution studies of subunits have also revealed the mode

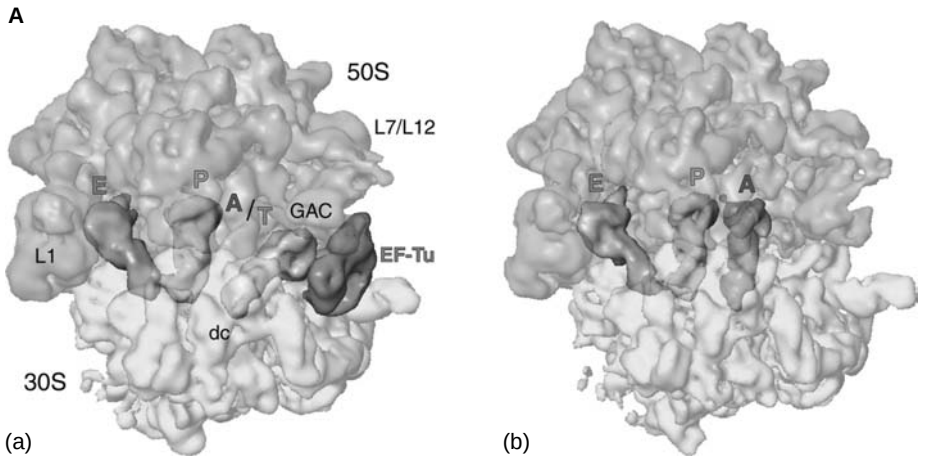

Fig. 13 Cryo-electron microscopy maps showing the incorporation of aminoacyl tRNA into the A site of a bacterial ribosome at a resolution of 0.9 nm and a conformational change in the GTPase-associated center. **A** The location of the binding site for the ternary complex of Phe-tRNA^{Phe}•EF-Tu•GDP-Kir. (a) The antibiotic Kirromycin (Kir) binds to the ternary complex and prevents dissociation of EF-Tu•GDP-Kir from the ribosome after GTP hydrolysis. The stalled complex is located below the L7/L12 stalk of the 50S subunit. EF-Tu is in red and A/T site tRNA is in pale pink. (b) Structure of the ribosome after peptide bond formation. The P site contains tRNA$_f^{Met}$ and the A site contains fMet•Phe-tRNA^{Phe} which is shown in magenta. In both panels, the ribosomal subunits are rendered semitransparent to illustrate the densities for tRNAs in the ribosomal intersubunit space. The sites for the tRNAs are labeled as follows: A/T, A/T site for the aa-tRNA bound to EF-Tu, reaching the decoding site in the A site of the 30S ribosomal subunit; A, aminoacyl site; P, peptidyl site; E, exit site. The P site tRNA is green, the E site tRNA is orange. **B** Conformational change in the GTPase-associated center of the 50S subunit. The cryo-electron microscopy maps show; (a) fMet-tRNA$_f^{Met}$ bound to the P site of the ribosome with a vacant A site; (b) fMet-tRNA$_f^{Met}$ in the P site and Phe-tRNA^{Phe}•EF-Tu•GDP-Kir complex with the A/T site between tRNA colored pink and EF-Tu colored red. The arrow indicates the site at the base of the L7/L12 stalk, which undergoes a conformational change during the binding of the ternary complex; (c) tRNA$_f^{Met}$ bound in the P site and fMet•Phe-tRNA^{Phe} bind in the A site. Insets (b) and (c) focus on the conformational changes of the GTPase-associated center. The view from the intersubunit space allows the movement of the GTPase-associated center to be observed. The E site is shown in orange; (d) Superposition of the 50S ribosomal subunits from the ribosome bound with the stalled ternary complex (semitransparent blue) and from the ribosome bearing a dipeptidyl-tRNA in the A site (solid red). Ribosomal subunits are blue (50S subunit) and yellow (30S subunit). Features of the large ribosomal subunit: L7/L12, stalk of proteins L7/L12 from the 50S ribosomal subunit; GAC, GTPase-associated center; SRC, sarcin-ricin loop; L16, ribosomal protein L16; L1, stalk of ribosomal protein L1. Features of the small ribosomal subunit: h, head of the 30S subunit; dc, decoding center. [From Valle, M., Zavialov, A., Li, W. et al. (2003) Incorporation of aminoacyl tRNA into the ribosome as seen by cryo-electron microscopy, *Nat. Struct. Biol.* **10**, 899–906. With permission] (see color plate p. xxxvii).

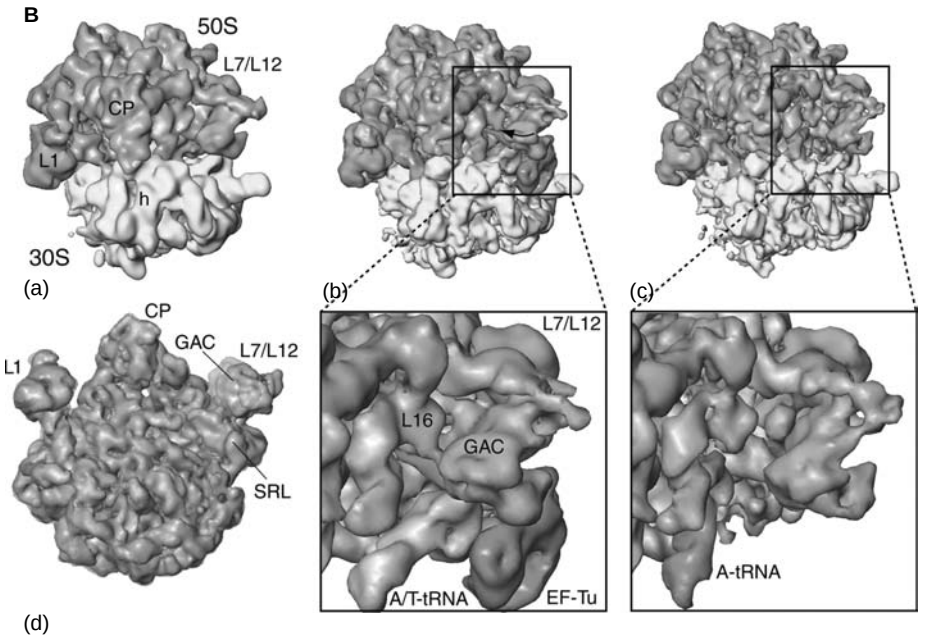

Fig. 13 (*Continued*)

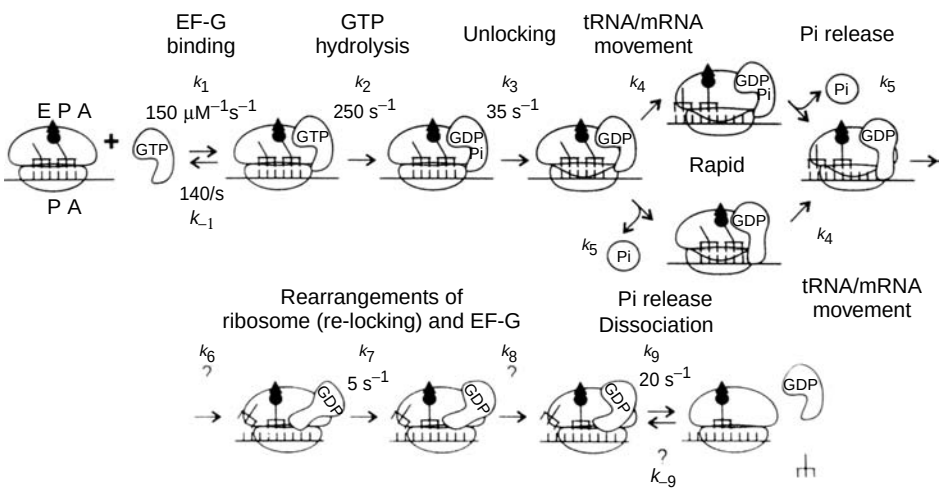

Fig. 14 A kinetic model of translocation. The diagram indicates several steps identified by several biochemical and rapid kinetic techniques. Ribosomes are shown in two conformations, closed and open. In the closed or locked conformation, bound tRNAs interact extensively with the ribosome. These interactions need to be disrupted to allow translocation (the open or unlocked conformation) and to be reestablished (relocked) after translocation. EF-G is shown in different conformations and orientations based on cryo-electron microscopy (see Fig. 5). [Reproduced from Peske, F., Savelsbergh, A., Katunin, V.I., Rodnina, M.V., Wintermeyer, W. (2004) Conformational changes of the small ribosomal subunit during elongation factor G-dependent tRNA-mRNA translocation, *J. Mol. Biol.* **343**, 1183–1194. With permission].

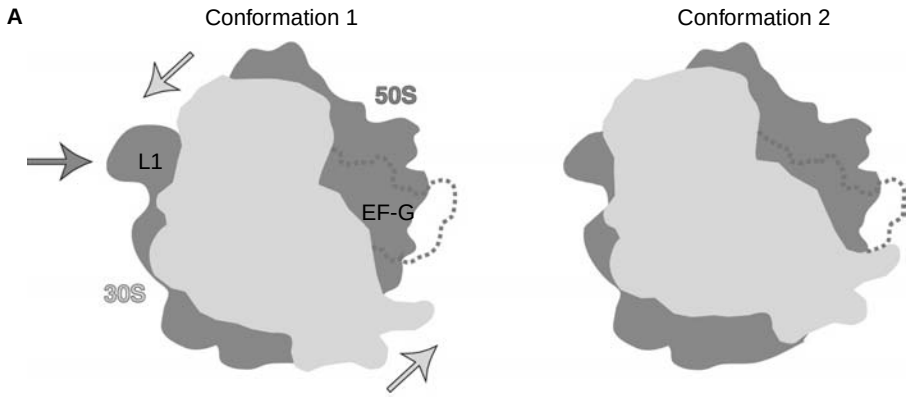

Fig. 15 Changes detected by cryo-electron microscopy (1.75–1.84 nm resolution) in the conformation of ribosomes during translocation. (A) After the peptide bond is formed, deacylated tRNA is located in the P site with peptidyl tRNA in the A site (conformation 1), the translocational state. A ratchet-like motion follows when EF-G•GTP is bound to the ribosome leading to conformation 2. The conformational change involves a counterclockwise rotation of the 30S subunit with respect to the 50S subunit with the L1 stalk moving in the direction opposite to the head of the 30S subunit. The outline of EF-G is shown by the dotted line. (B) Conformational changes in the L7/L12 stalk and in EF-G during translocation. The figures show that EF-G (shown in red) links the decoding site with the L7/L12 stalk. The cryo-electron microscopy maps shown in (c) and (d) correspond to conformation 2 in (a) with deacylated tRNA in the P site and peptidyl-tRNA in the A site and EF-G•GTP bound to the ribosome. The L7/L12 stalk (shown on the right of the dotted line) is bifurcated. Part of EF-G is shown close to the A site (marked A in (c)). A third conformation is observed after hydrolysis of GTP to GDP and before EF-G•GDP is released from the ribosome as shown in (a) and (b). In this conformation, the stalk is extended but not bifurcated and part of the EF-G is implicated in an arclike structure. The pretranslocational state is regained (see (a)) when EF-G•GDP is released from the ribosome. At the end of the translocation process, tRNAs in the A and P sites are transferred to the P and E sites, respectively, with the concomitant advance by one codon of mRNA. The insets show the appropriate orientations of the ribosome with 30S subunits and 50S subunits colored yellow and blue respectively. Features of the 30S subunit: b, body; h, head; sp, spur; ch, channel. Features of the 50S subunit: cp, central protuberance; L1, L1 protein; st, L7/L12 stalk. (A was reproduced from, Gas, A., Valle, M., Ehrenberg, M. and Frank, J. (2004). Dynamics of EF-G interaction with the ribosome explored by classification of a heterogenous cryo-EM dataset J. Structural Biol. **147**, 283–290. B was reproduced from Agrawal, R.K., Heagle, A.B., Penczek, P., Grassucci, R.A., Frank, J. (1999) EF-G dependent GTP hydrolysis induces translocation accompanied by large changes in the conformation of the 70S ribosome, *Nat. Struct. Biol.* **6**, 543–647. With permission) (see color plate p. xxxix).

of action of several antibiotics such as paromomycin, streptomycin, and spectinomycin, which modify the decoding function of the small subunit, and chloramphenicol, puromycin, and vernamycin, which affect peptide bond formation.

5 Translational Control of Gene Expression

Cells need to synthesize specific proteins in the required amounts at particular times and to deliver them to the correct locations.

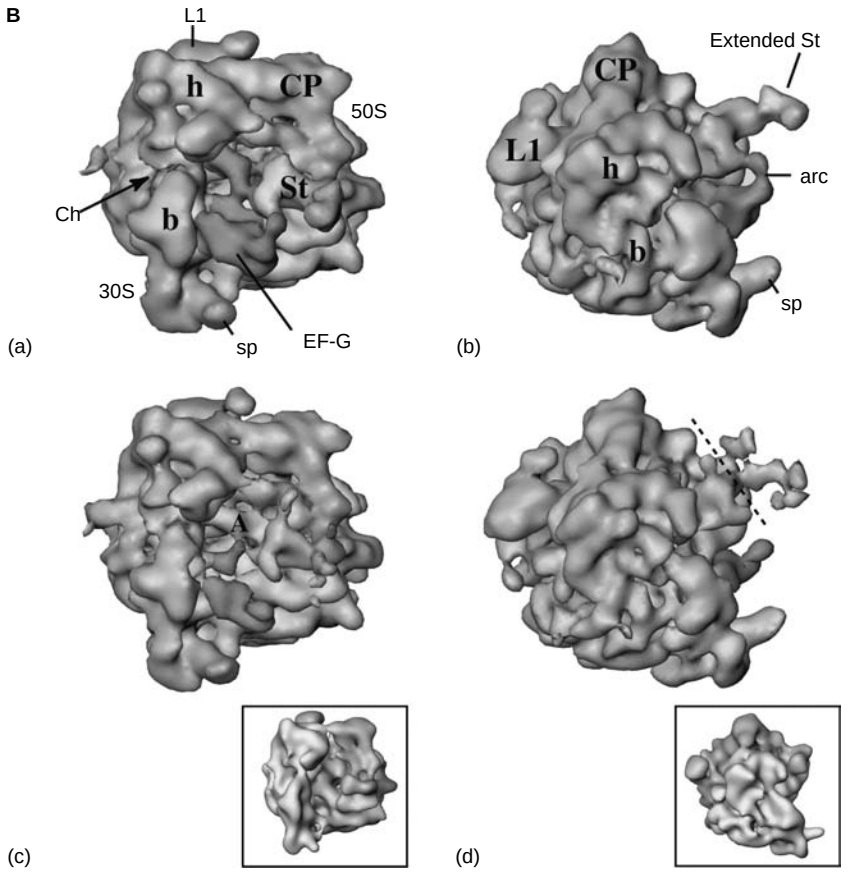

Fig. 15 (*Continued*)

These processes depend on a great many interactions and numerous mechanisms exist for the translational control of gene expression. Examples of the ways used by cells to control the translation of mRNA are presented in the sections that follow. This material is intended to be illustrative rather than comprehensive, and it is to be expected that novel control mechanisms will continue to be discovered.

5.1 mRNA Stability

Provided all other components of the translational systems are present in optimum amounts, control of translation may be achieved through the availability of the relevant messenger RNAs at the site of protein synthesis. The steady state level of mRNA is determined by the rate of its synthesis and degradation. Control of transcription is of major importance for synthesis in both prokaryotes and eukaryotes. The stability of mRNA depends on its primary and secondary structure as well as on the presence of factors such as stabilizing proteins and nucleases.

The secondary structure of mRNA is determined by its nucleotide sequence. For any mRNA, a number of coding sequences are possible because of the degeneracy

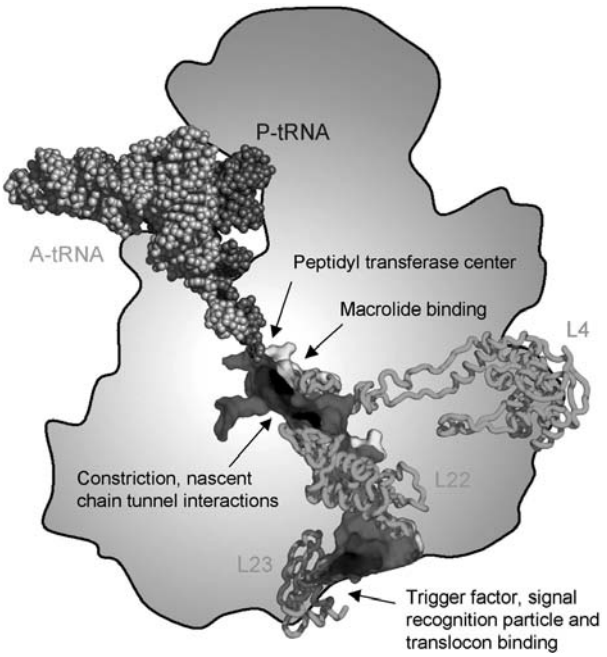

Fig. 16 Protein synthesis and the path of the nascent chain through the ribosome exit tunnel. The contour of the large ribosomal subunit is shown in gray. The aa-tRNA and pept-tRNA are represented by spheres, colored copper, and gray respectively. Ribosomal proteins L4, L22, and L23 are shown as green ribbons. The surface of the ribosomal exit tunnel corresponds to the empty space in the 50S subunit, which is sliced on the side. The surface is colored yellow outside and blue inside. The acceptor stems of aa-tRNA and pept-tRNA point toward the PTC, where peptide bond formation occurs. Nascent polypeptides escape the ribosome through the tunnel. They have to pass the narrow constriction where ribosomal proteins L4 and L22 contact the wall of the tunnel. This region is a target of antibiotics and the site of regulatory interactions between nascent chains and the tunnel. Ribosomal protein L22 mediates the interaction between nascent chains and the tunnel. Ribosomal protein L23 mediates the interaction between nascent chains emerging from the ribosome and cytosolic chaperones such as Trigger Factor (TF) and protein targeting factors such as Signal Recognition Factor (SRFF) and translocons. [From Jenni, S., Ban, N. (2003) The chemistry of protein synthesis and voyage through the ribosomal tunnel, *Curr. Opin. Struct. Biol.* **13**, 212–219. With permission] (see color plate p. xlii).

of the genetic code. Thus, degeneracy allows for particular features of secondary structure (often termed *cis* factors) to be favored, which may act either to stabilize or destabilize the mRNA, according to the needs of the cell, by determining its susceptibility to degradative enzymes (often termed *trans*-acting factors).

Whereas the structure of mRNA determines its susceptibility to degradative enzymes, the detailed mechanisms are complex. In prokaryotes, the enzymes involved include two endonucleases (RNase E and RNase III) and two exonucleases (polynucleotide phosphorylase and RNase II). Other nucleases may be active in particular cases such as phage infection. In eukaryotes, a major pathway involves the removal of the 3′ poly(A) tail (deadenylation), followed by removal of the 5′ cap, which renders the mRNA susceptible to rapid endonucleolytic degradation in the 5′ → 3′ direction.

5.2 Control by Interaction of Proteins with mRNA

Throughout the ribosome cycle, dynamic protein-mRNA interactions are functionally important in the initiation, elongation, and termination of polypeptide synthesis. In addition, more stable associations between proteins and mRNAs have been observed, particularly in eukaryotic cells. These messenger ribonucleoprotein complexes (mRNPs) occur both in polyribosomes and free in the cytosol, some of the latter being either temporarily or permanently unavailable for translation. Thus, protein-mRNA interactions contribute to the efficiency with which mRNAs are translated. Some proteins, such as the poly(A)-binding protein (p78), are present in most if not all mRNPs, whereas others appear to be cell specific and mRNA selective. In unfertilized sea urchin eggs and *Xenopus* oocytes, for example, untranslated messenger is sequestered by association with proteins that prevent translation until later stages of development. Duck reticulocytes contain globin mRNP, which cannot be translated *in vitro*, whereas the mRNA obtained by deproteinizing the complex can be translated, showing that in this case translation is prevented by the mRNP proteins.

Formation of a site-specific mRNA-protein complex is involved in the translational control of the biosynthesis of ferritin, an iron storage protein, which is stimulated in response to the presence of iron. In this instance, a cytoplasmic repressor protein of 85 kDa binds to a highly conserved 28-nucleotide stem-loop structure in the 5′ untranslated region of ferritin mRNAs in the absence of iron. In the presence of iron, the protein dissociates from the mRNA, which is then available for translation. A similar loop motif occurs in the 3′ untranslated region of transferrin receptor mRNA, which is also subject to translational control by an iron-responsive repressor.

During the cell cycle, histone mRNA is destabilized after completion of DNA replication, resulting in a 30- to 50-fold decrease. This change appears to be due to an increase in the level of free histones, which form a complex with histone mRNA by interaction with a stem-loop structure at the extreme 3′ terminus. Formation of this histone–histone mRNA complex is thought to activate a ribosome-associated 3′ → 5′ exonuclease, which degrades the histone mRNA. During the S phase, newly synthesized DNA binds free histones to form nucleosomes, thus preventing the degradation of histone mRNA at this stage of the cell cycle.

Specific regulation of gene expression at the level of translation also exists in prokaryotes. For example, the synthesis of *E. coli* threonyl-tRNA synthetase is negatively autoregulated by an interaction of the tRNA-like leader sequence

of its mRNA with the synthetase that inhibits translation by preventing the binding of ribosomes. The synthetase is displaced from the mRNA by $tRNA^{Thr}$, which thus acts as a translational antirepressor. This regulatory mechanism allows the cell to maintain a balance between the tRNA synthetase and its cognate tRNA.

Similarly, there is a mechanism used to control the synthesis of proteins encoded by a polycistronic mRNA. In this case, selective binding of the ribosomal protein to the region of the mRNA involved in the initiation of translation leads to the regulatory protein controlling both its own synthesis and that of other ribosomal proteins. A specific example is the role of ribosomal protein S4, which acts as a translational repressor of four ribosomal proteins (S4, S11, S13, and L17). Protein S4 appears to function as a repressor through an unusual "pseudoknot" linking a hairpin loop upstream of the ribosome-binding site with sequences 2 to 10 codons downstream of the initiation codon. (A pseudoknot structure contains intramolecular base pairs between base residues in the loop of a stem-loop structure and distal complementary regions of the RNA.) Stabilization of this structure by S4 would prevent the binding of ribosomes, and this control mechanism may contribute to the coordinated synthesis of the different ribosomal proteins required for ribosome assembly.

5.3 Control by mRNA Structure

The secondary structure of some eukaryotic mRNAs regulates translation by a mechanism involving a ribosomal frameshift which gives rise to a directed change of the translational reading frame to allow the synthesis of a single protein from two or more overlapping genes by suppression of an intervening termination codon. Several retroviruses use this mechanism to move from one reading frame to another in the expression of the viral RNA-dependent DNA polymerase. Other examples of the operation of such a frameshift include the synthesis of the reverse transcriptase enzymes of several retrotransposons, such as the yeast Ty1. The mechanism of changing the reading frame involves "slippery" sequences and a complex folding of the mRNA into a structure termed a *pseudoknot*.

In *E. coli* phages, translational control of the three cistrons of Q_β, f2, and related bacteriophage RNAs (Fig. 17) accounts for the synthesis of coat protein:replicase:A protein in the approximate ratio of 20 : 5 : 1. These quantitative differences are due to the differential and independent initiation of translation at each cistron as a result of differences in the secondary structure of the mRNA initiation sites. Furthermore, *in vivo* there is a delay in the synthesis of coat protein and this temporal control involves translational repression of the cistron by ribosomal protein S1. In addition, S1 functions as one of the subunits of the f2 RNA replicase; therefore, association of the newly synthesized translation product of the f2 replicase cistron with S1 will favor its dissociation from the phage RNA, thus allowing translation of the coat protein cistron to start.

5.4 Control by Modification of Translation Factor Activity

5.4.1 Initiation Factors

In eukaryotes, initiation of protein synthesis is inhibited by phosphorylation of the initiation factor eIF-2. In particular cases,

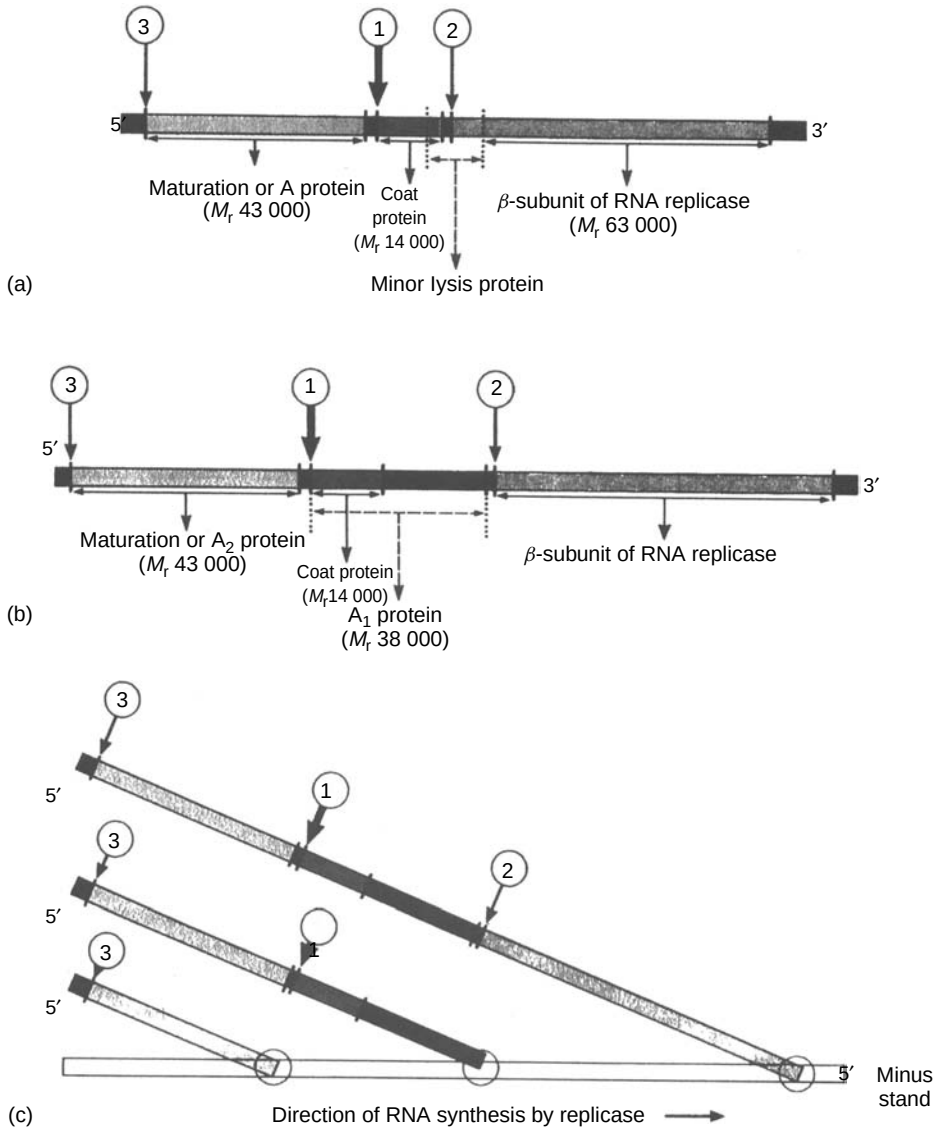

Fig. 17 Translational control of bacteriophage synthesis: (a) arrangements of MS2 and f2 bacteriophage cistrons, (b) arrangement of Q_β bacteriophage cistrons, and (c) replicative intermediate synthesizing new plus strands on the complementary minus-strand copy of the original bacteriophage RNA. The ribosome-binding sites (RBS) are indicated by the numbered arrows. The major RBS (1) binds ribosomes efficiently but can be blocked by ribosomal protein S1. The secondary RBS (2) becomes available only after translation of at least part of the coat protein cistron by ribosomes. RBS 3 is masked by the secondary structure of native bacteriophage RNA but is accessible in nascent RNA (c) or *in vitro* when the secondary structure is destroyed. Noncoding regions are shown in black. [Reproduced from Arnstein, H.R.V., Cox, R.A. (1992) *Protein Biosynthesis*, Oxford University Press, Oxford. With permission].

phosphorylation of eIF-2 is stimulated by a lack of heme or the presence of double-stranded RNA. Two different protein kinases capable of phosphorylating the α-subunit of eIF-2 have been characterized. One substrate, called the heme-controlled repressor (HCR) or heme-regulated inhibitor (HRI), is a cytoplasmic protein (95 000 Da) that is activated by phosphorylation. Double-stranded RNA activates phosphorylation of a 67 000-Da protein, which in turn phosphorylates eIF-2. Thus, a cascade of protein phosphorylation is involved. Phosphorylated eIF-2 is unable to exchange GDP for GTP, which prevents it from functioning in the binding of initiator tRNA to ribosomes.

Conditions other than lack of heme (e.g. heat shock, serum deprivation, or the presence of oxidized glutathione), which are known to inhibit protein synthesis, also stimulate the phosphorylation of eIF-2α. Conversely, the activity of eIF-4F is decreased by dephosphorylation of the 24-Da subunit (see Table 3) rather than by phosphorylation. Thus, a number of different kinases and phosphatases are involved in modulating the activities of different factors.

Small RNAs may also be involved in regulating the translation of mRNA in eukaryotic cells. Of the stimulatory RNAs, the best characterized is a small RNA of about 160 nucleotides, which accumulates in cells after infection with adenovirus. This virus-associated RNA, VA-RNA$_1$ sustains general protein synthesis by inhibiting the phosphorylation of the α-subunit of initiation factor eIF-2.

Two features of mRNA structure are implicated in several regulatory processes involving initiation factors, namely, the 5′ cap structure and a cytoplasmic polyadenylation element (*c*PE) located within the 3′ untranslated region. For example, embryonic cells have mRNAs that are stored until particular stages in development are reached. During storage, eIF-4E is thought to bind to the 5′ cap structure and a protein *c*PEB is bound to the *c*PE element; a third protein binds to both eIF-4E and *c*PEB. Similar nucleoprotein structures have been proposed to regulate the synthesis of key proteins during progression through the cell division cycle. Furthermore, eIF-2 and eIF-4E are targets for changing the patterns of protein synthesis associated with tumorigenesis. Phosphorylation of the α-subunit of eIF-2 (eIF-2α) regulates its activity. Phosphorylation of a small family of eIF-4E binding proteins (4E-BPs) regulates the binding of eIF-4E to eIF-4G. The activities of the appropriate kinases and phosphatases may, in turn, be controlled by cellular and viral oncogenes and tumor-suppressor genes.

A novel pathway for the initiation of translation has been found which permits initiation in a manner that is independent of the 5′ end of mRNA. Particular RNA sequences are involved, termed *internal RNA entry sites* (IRES). IRES elements have been found in both viral and cellular RNAs. The IRES elements intervene in the initiation of translation during cellular stress and during infection by viruses, for example, poliovirus or hepatitis C virus. The cricket paralysis virus RNA possesses an IRES element, which, in the absence of initiation factors and initiator tRNA, can bind to the P site and assemble 80S ribosomes. In this way, the reading frame is set so that the first codon is in the A site. The first translocation step takes place without a peptide bond being formed. However, aminoacyl tRNA is delivered to the P site so that chain elongation can

proceed normally. The cricket paralysis virus IRES bound to the small subunit and to ribosomes was visualized by cryo-electron microscopy.

5.4.2 Other Translation Factors

A Ca^{2+}/calmodulin-dependent protein kinase phosphorylates eEF-2. The phosphorylated factor appears to be inactive. Dephosphorylation of the factor by phosphatase restores its activity.

5.5 Effects of Antisense Polynucleotides

Antisense RNAs, which are polynucleotides with base sequences complementary to messenger RNAs, have been found in both prokaryotes and eukaryotes. Natural antisense RNAs are not common but synthetic RNAs directed at specific targets have been widely studied. It has been demonstrated that they can function as inhibitors of messenger RNA translation. In prokaryotes, the most effective inhibitors appear to have a base sequence complementary to the 5′ leader region, including the Shine–Dalgarno sequence, which is involved in the binding of mRNA to the small ribosomal subunit. In eukaryotes, translation of mRNA is inhibited by polyribonucleotides complementary to the 5′ untranslated region of mRNA, indicating a direct effect on initiation as in prokaryotes. Polynucleotides complementary to the 3′ untranslated region of mRNA also inhibit translation in some cells, and this effect may be due to destabilization of mRNA by ribonucleases specific for double-stranded RNA. The effect of antisense polynucleotides is not restricted to translation of mRNAs, but transcription as well as the processing of transcripts may also be inhibited.

5.6 Availability of Amino Acids, tRNA Abundance, and Codon Usage

5.6.1 Amino Acids

Polypeptide synthesis depends on an adequate supply of tRNAs charged with the 20 protein amino acids and appropriate interactions between their anticodons and the codons of mRNA. Peptide chain elongation is decreased or inhibited by lack of amino acids or other conditions giving rise to an imbalance or deficiency in aminoacyl tRNAs.

5.6.2 Abundance of tRNAs and Codon Usage

Different tRNAs are present in the cytosol in unequal amounts, and elongation rates are slower at codons corresponding to rare tRNA species. The existence of synonymous codons raises the question of preferential use of some codons and its possible significance in relation to translational efficiency and control. In some bacteria (e.g. *Pseudomonas aeruginosa*, which has a high content of G + C, 67.2%, in DNA), the most common codons are those with the strongest predicted codon–anticodon interaction – that is G + C base pairs – but this preference is not universal and, for example, does not apply to *E. coli*, which has a lower proportion of G + C (50%). Although codon usage may affect elongation rates, it is probably of less importance in translational control than the secondary structure of mRNA in relation to the rate of initiation of protein synthesis.

5.7 Modulation of Ribosome Activity

Specific ribosomal components have an important function in relation to the fidelity of protein synthesis. Thus, in *E. coli*

ribosomal protein, S12 determines the accuracy of codon–anticodon interactions and modulates the translational error frequency in the presence of the antibiotic streptomycin.

To what extent reversible modifications of ribosomal constituents are involved in translational control of protein synthesis is uncertain. Although phosphorylation of ribosomal protein S6 increases with cell proliferation, it is not known whether this change is directly related to the accompanying increase in protein synthesis by an effect on the translation rate.

5.8 Ribosome-inactivating Proteins

Many molds and plants produce toxins, which are protective agents, termed ribosome-inactivating proteins (RIPs), directed at particular cells and their ribosomes. These toxins are classified as either type I or type II RIPs according to the number of polypeptide chains.

Type I RIPs comprise a single polypeptide chain; for example, α-sarcin, an extracellular cytotoxin produced by *Aspergillus giganteus*, consists of a single chain of 150 amino acid residues.

Type II RIPs comprise two polypeptide chains, A and B. The A chain has the ability to inactivate ribosomes, and the B chain is a galactose-specific lectin responsible for the entry of the toxin into the target cell. Ricin, which is isolated from castor oil beans, is representative of type II RIPs.

Ribosomes are inactivated as a result of the RNA N-glycosidase activity of RIPs. These toxins have different specificities for particular cells and ribosomes. However, the target site for all RIPs is an adenylate residue (position 2660 in the *E. coli* 23S rRNA sequence) located within a highly conserved sequence of 12 nucleotides ($5'_{2654}$AGUACGAGAGGA$_{2665}3'$). Cleavage of the GpA$_{2660}$ internucleotide bond or depurination of A$_{2660}$ is sufficient to inactivate the ribosome. The target residue is located in the loop region of a stem–loop element of the secondary structure, which is termed the α-sarcin stem–loop. Thus, the target adenylate is either directly or indirectly essential for ribosome function. The α-sarcin stem–loop is known to be important for binding elongation factors EF-Tu and EF-G to the ribosome. RIPs have attracted interest as active components of reagents directed at particular targets such as cancer cells.

6 Concluding Remarks

The control of protein synthesis, either by regulation of the amount of mRNA available for translation or by the efficiency with which it is translated, is important in cell growth and development as a factor determining the level of cellular and extracellular proteins. Subversion of this control occurs in cells infected by viruses when the viral nucleic acid uses the protein-synthesizing machinery of the host cell and thereby changes normal cell metabolism in favor of the synthesis of viral proteins needed for the production of virus progeny.

The polypeptide chains of all proteins are synthesized by the process described above. This mechanism gives rise to primary polypeptide chains, which are often further modified; for example, by cleavage into smaller peptides, by structural modification of selected amino acid residues, by splicing of the polypeptide chain, or by the formation of covalent bonds between polypeptide chains. Some of these

secondary modifications are related to the correct folding of polypeptide chains and to the production of active enzymes or peptide hormones from inactive precursors (e.g. insulin from proinsulin). Also, the transport of proteins within the cell or the secretion of extracellular proteins is often linked to structural changes in polypeptide chains either during or after completion of synthesis.

See also Body Expression Map of Human Genome; DNA Replication and Transcription; Expression Systems for DNA Processes; Ribosome, High Resolution Structure and Function; RNA Secondary Structures; RNA Three-Dimensional Structures, Computer Modeling of.

Bibliography

Books and Reviews

Arnstein, H.R.V., Cox, R.A. (1992) *Protein Biosynthesis*, Oxford University Press, London.

Garrett, R.A., Douthwaite, S.R., Matheson, A.T., Moore, P.B., Noller, H.F. (Eds.) (2000) *The Ribosome: Structure, Function, Antibiotics and Cellular Interactions*, American Society for Microbiology, Washington, DC.

Hershey, J.W.B., Merrick, W.C. (2000) Pathway and Mechanism of Initiation of Protein Synthesis, in: Sonenberg, N., Hershey, J.W.B., Mathews, M.B. (Eds.) *Translational Control of Gene Expression*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 33–88.

Hershey, J.W.B., Mathews, M.B., Sonenberg, N. (Eds.) (1996) *The Pathway and Mechanism of Eukaryotic Protein Synthesis*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Hill, W.E., Dahlberg, A., Garrett, R.A., Moore, P.B., Schlessinger, D., Warner, J.R. (1990) *The Ribosome, Structure, Function and Evolution*, American Society for Microbiology, Washington, DC.

Matheson, A.T., Davies, J., Hill, W., Dennis, P. (1996) *Frontiers in Translation*, National Research Council of Canada, Ottawa, Ontario.

Nierhaus, K.H., Subramanian, A.R., Erdmann, V.A., Franceschi, F., Wittmann-Liebold, B. (Eds.) (1994) *Translational Apparatus: Structure, Function, Regulation and Evolution*, Plenum Press, New York.

Nierhaus, K.H., Wilson, D.N. (Eds.) (2004) *Protein Synthesis and Ribosome Structure: Translating the Genome*, Wiley-VCH Verlag, Weinheim.

Primary Literature

Agrawal, R.K., Heagle, A.B., Penczek, P., Grassucci, R.A., Frank, J. (1999) EF-G-dependent GTP hydrolysis induces translocation accompanied by large conformational changes in the 70S ribosome, *Nat. Struct. Biol.* **6**, 643–647.

Agrawal, R.K., Sharma, R.S., Kiel, M.C., Irokawa, G., Booth, T.M., Spahn, C.M.T., Grassucci, R.A., Kaji, A., Frank, J. (2004) Visualization of ribosome-recycling factor on the *Escherichia coli* 70S ribosome: Functional implications, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8900–8905.

Al-Karadaghi, S., Kristensen, O., Liljas, A. (2000) A decade of progress in understanding the structural basis of protein synthesis, *Prog. Biophys. Mol. Biol.* **73**, 167–193.

Auerbach, T., Bashan, A., Yonath, A. (2004) Ribosomal antibiotics: structural basis for resistance, synergism and selectivity, *Trends Biotechnol.* **22**, 570–576.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution, *Science* **289**, 905–920.

Bashan, A., Agmon, I., Zarivach, R., Schluenzen, F., Harms, J., Berisio, R., Bartels, H., Franceschi, F., Auerbach, T., Hansen, H.A.S., Kossoy, E., Kessler, M., Yonath, A. (2003) Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression, *Mol. Cell.* **11**, 91–102.

Blanchard, C.S., Gonzalez R.L. Jr., Kim, H.D., Chu, S., Puglisi, J.D. (2004) tRNA selection

and kinetic proofreading in translation, *Nat. Struct. Mol. Biol.*. **11**, 1008–1014.

Cao, Q., Richter, J.D. (2002) Dissolution of the maskin-eIF4E complex b cytoplasmic polyadenylation and poly(A)-binding protein controls cyclin B1 mRNA translation and oocyte maturation, *EMBO J.* **21**, 3852–3862.

Carter, A.R., Clemons, W.M., Brodersen, D.E., Morgan-Warren, R.J., Wimberly, B.T., Ramakrishnan, V. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics, *Nature* **407**, 340–349.

Castagnetti, S., Ephrussi, A. (2003) Orb and a long poly(A) tail are required for efficient oskar translation at the posterior pole of the Drosophila oocyte, *Development* **130**, 835–843.

Charlesworth, A., Cox, L.L., MacNicol, A.M. (2004) Cytoplasmic polyadenylation element (CPE)-and CPE binding protein (CPEB)-independent mechanisms regulate early class maternal mNA translational activation in Xenopus oocytes, *J. Biol. Chem.* **279**, 17650–17659.

Clemens, M.J. (2004) Targets and mechanisms for the regulation of translation in malignant transformation, *Oncogene* **23**, 3180–3188.

Cormier, P., Pyronnet, S., Salaun, P., Mulner-Lorillon, O., Sonenberg, N. (2003) Cap-dependent translation and control of the cell cycle, *Prog. Cell Cycle Res.* **5**, 469–475.

Cropp, T.A., Schultz, P.G. (2004) An expanding genetic code, *Trends Genet.* **20**, 625–630.

Dever, T. (2002) Gene-specific regulation by general translation factors, *Cell* **108**, 545–556.

Diedrich, G., Spahn, C.M.T., Stelzi, U., Schäfer, M.A., Wooten, T., Bochkariov, D.E., Coopeman, B.S., Traut, R.R., Nierhaus, K.H. (2000) Ribosomal protein L2 is involved in the association of the ribosomal subunits, tRNA binding A and P sites and peptidyl transfer, *EMBO J.* **19**, 5241–5250.

Dodson, R.E., Shapiro, D.J. (2002) Regulation of pathways of mRNA destabilization and stabilization, *Prog. Nucleic Acid Res. Mol. Biol.* **72**, 129–164.

Gao, H., Valle, M., Ehrenberg, M., Frank, J. (2004) Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset, *J. Struct. Biol.* **147**, 283–290.

Gebauer, F., Hentze, M.W. (2004) Molecular mechanisms of translational control, *Nat. Rev. Mol. Cell Biol.* **5**, 827–835.

Gilbert, R.J.C., Fucini, P., Connell, S., Fuller, S.D., Nierhaus, K.H., Robinson, C.V., Dobson, C.M., Stuart, D.I. (2004) Three-dimensional structures of translating ribosomes by cryo-EM, *Mol. Cell* **14**, 57–66.

Green, R., Lorsch, J.R. (2002) The path to perdition is paved with protons, *Cell* **110**, 665–668.

Green, R., Noller, H.F. (1997) Ribosomes and translation, *Annu. Rev. Biochem.* **66**, 674–716.

Groisman, I., Jung, M.Y., Sarkissian, M., Cao, Q., Richter, J.D. (2002) Translational control of the embryonic cell cycle, *Cell* **109**, 473–483.

Gromadski, K.B., Rodnina, M.V. (2004) Streptomycin interferes with conformational coupling betweeen codon recognition and GTPase activation on the ribosome, *Nat. Struct. Mol. Biol.* **11**, 316–322.

Grunberg-Manago, M. (1999) Messenger RNA stability and its role in control of gene expression in bacteria and phages, *Annu. Rev. Genet.* **33**, 193–227.

Halic, M., Becker, T., Pool, M.R., Spahn, C.M.T., Grassucci, R.A., Frank, J., Beckmann, R. (2004) Structure of the signal recognition particle interacting with the elongation-arrested ribosome, *Nature* **427**, 808–814.

Hansen, J.L., Schmeing, T.M., Moore, P.B., Steitz, T.A. (2002) Structural insights into peptide bond formation, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11670–11675.

Hardesty, B., Kramer, G. (2001) Folding of a nascent peptide on the ribosome, *Prog. Nucleic Acid Res. Mol. Biol.* **66**, 41–66.

Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., Yonath, A. (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium, *Cell* **107**, 679–688.

Hazelrigg, T. (2004) Lost in translation gets an oskar, *Dev. Cell* **6**, 611–613.

Hellen, C.U., Sarnow, P. (2001) Internal ribosome entry sites in eukaryotic mRNA molecules, *Genes Dev.* **15**, 1593–1612.

Herr, A.J., Atkins, J.F., Gesteland, R.F. (2000) Coupling of open reading frames by translational bypassing, *Annu. Rev. Biochem.* **69**, 343–372.

Hirokawa, G., Inokuchi, H., Kaji, H., Igarashi, K., Kaji, A. (2004) *In vivo* effect of inactivation of ribosome recycling factor - fate of ribosomes after unscheduled translation downstream

of open reading frame, *Mol. Microbiol.* **54**, 1011–1021.

Hoang, L., Frederick, K., Noller, H.F. (2004) Creating ribosomes with an all-RNA 30S subunit P site, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12439–12443.

Horton, L.E., Bushell, M., Barth-Baus, D., Tilleray, V.J., Clemens, M.J., Hensold, J.O. (2002) p53 activation results in rapid dephosphorylation of the eIF4E-binding protein 4E-BP1, inhibition of ribosomal protein S6 kinase and inhibition of translation initiation, *Oncogene* **21**, 5325–5334.

Ibba, M., Soll, D. (2000) Aminoacyl-tRN Asynthesis, *Annu. Rev. Biochem.* **69**, 617–650.

Jan, E., Hinzy, T.G., Sarnow, P. (2003) Divergent tRNA-like element supports initiation, elongation and termination of protein biosynthesis, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15410–15415.

Jenni, S., Ban, N. (2003) The chemistry of protein synthesis and voyage through the ribosomal tunnel, *Curr. Opin. Struct. Biol.* **13**, 212–219.

Katunin, V.I., Muth, G.W., Strobel, S.A., Wintermeyer, W., Rodnina, M.V. (2002) Important contribution to catalysis of peptide bond formation by a single ionizing group within the ribosome, *Mol. Cell* **10**, 339–346.

Krab, I.M., Parmeggiani, A. (2002) Mechanisms of EF-Tu, a pioneer GTPase, *Prog. Nucleic Acid Res. Mol. Biol.* **71**, 513–551.

Meyer, S., Temme, C., Wahle, W. (2004) Messenger RNA turnover in eukaryotes: pathways and enzymes, *Crit. Rev. Biochem. Mol. Biol.* **39**, 197–216.

Nakamura, A., Sato, K., Hanyu-Nakamura, K. (2004) Drosophila cup is an eIF4E binding protein that associates with Bruno and regulates oskar mRNA translation in oogenesis, *Dev. Cell* **6**, 69–78.

Nissen, P., Hansen, J., Ban, N., Moore, P.B., Steitz, T.A. (2000) The structural basis of ribosome activity in peptide bond synthesis, *Science* **289**, 920–930.

Noller, H.F. (2004) The driving force for molecular evolution of translation, *RNA* **10**, 1833–1837.

Peske, F., Savelsbergh, A., Katunin, V.I., Rodnina, M.V., Wintermeyer, W. (2004) Conformational changes of the small ribosomal subunit during elongation factor G-dependent tRNA-mRNA translocation, *J. Mol. Biol.* **343**, 1183–1194.

Prats, A.-C., Prats, H. (2002) Translational control of gene expression: role of IREs and consequences for cell transformation and angiogenesis, *Prog. Nucleic Acid Res. Mol. Biol.* **72**, 367–413.

Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation, *Cell* **108**, 557–572.

Rodnina, M.V., Wintermeyer, W. (2003) Peptide bond formation on the ribosome: structure and mechanism, *Curr. Opin. Struct. Biol.* **13**, 334–340.

Rodnina, M.V., Savelsbergh, A., Wintermeyer, W. (1999) Dynamics of translation on the ribosome: molecular mechanics of translocation, *FEMS Microbiol. Rev.* **23**, 317–333.

Ryabova, L.A., Pooggin, M.M., Hohn, T. (2002) Viral strategies of translation initiation: ribosomal shunt and reinitiation, *Prog. Nucleic Acid Res. Mol. Biol.* **72**, 1–39.

Sachs, A.B. (2000) Cell cycle-dependent translation initiation: IRES elements prevail, *Cell* **101**, 243–245.

Sachs, A.B., Varani, G. (2000) Eukaryotic translation initiation: there are (at least) two sides to every story, *Nat. Struct. Biol.* **7**, 356–361.

Schlüenzen, F., Pyetan, E., Fucini, P., Yonath, A., Harms, J.M. (2004) Inhibition of peptide bond formation by pleuromutilins: the structure of the 50S ribosomal subunit from *Deinococcus radiodurans* in complex with tiamulin, *Mol. Microbiol.* **54**, 1287–1294.

Schmeing, T.M., Seila, A.C., Hansen, J.L., Freeborn B., Soukup, J.K., Scaringe, S.A., Strobel, S.A., Moore, P.B., Steitz, T.A. (2002) A pre-translocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits, *Nat. Struct. Biol.* **9**, 225–230.

Schmidt, E.V. (2004) The role of c-myc in regulation of translation initiation, *Oncogene* **23**, 3217–3221.

Sengupta, J., Nilsson, J., Gursky, R., Spahn, C.M.T., Nissen, P., Frank, J. (2004) Identification of the versatile scaffold protein RACK1 on the eukaryotic ribosome by cryo-EM, *Nat. Struct. Mol. Biol.* **11**, 957–962.

Shine, J., Dalgarno, L. (1974) The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets

and ribosome binding sites, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1342–1346.
Shuman, S. (2001) Structure, mechanism and evolution of the mRNA capping apparatus, *Prog. Nucleic Acid Res. Mol. Biol.* **66**, 1–40.
Sonenberg, N., Dever, T.E. (2003) Eukaryotic translation factors and regulators, *Curr. Opin. Struct. Biol.* **13**, 56–63.
Spahn, C.M.T., Jan, E., Mulder, A., Grassucci, R.A., Sarnow, P., Frank, J. (2004) Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor, *Cell* **118**, 465–475.
Spremulli, L.L., Coursey, A., Navratil, T., Hunter, S.E. (2004) Initiation and elongation factors in mammalian mitochondrial protein biosynthesis, *Prog. Nucleic Acid Res. Mol. Biol.* **77**, 211–161.
Vagner, S., Galy, B., Pyronnet, S. (2001) Irresistible IRES. Attracting the translation machinery to internal ribosome entry sites, *EMBO Rep.* **2**, 893–898.
Valle, M., Zavialov, A., Li, W., Stagg, S.M., Sengupta, J., Nielsen, R.C., Nissen, P., Harvey, S.C., Ehrenberg, M., Frank, J. (2003) Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy, *Nat. Struct. Biol.* **10**, 899–906.
Vanzi, F., Vladimirov, S., Knudsen, C.R., Goldman, Y.E., Cooperman, B.S. (2003) Protein synthesis by single ribosomes, *RNA* **9**, 1174–1179.
Vila-Sanjurjo, A., Schuwirth, B.-S., Hau, C.W., Cate, J.H.D. (2004) Structural basis for the control of translation initiation during stress, *Nat. Struct. Mol. Biol.* **11**, 1054–1059.
Wild, K., Halic, M., Sinning, I., Beckmann, R. (2004) SRP meets the ribosome, *Nat. Struct. Mol. Biol.* **11**, 1049–1053.
Wilusz, C.J., Wilusz, J. (2004) Bringing the role of mRNA decay in the control of gene expression into focus, *Trends Genet.* **20**, 491–497.
Wimberly, B.T., Brodersen, D.E., Clemons, W.M. Jr., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit, *Nature* **407**, 327–339.
Yonath, A. (2002) The search and its outcome: high-resolution structures of ribosomal particles from mesophilic, thermophilic, and halophilic bacteria at various functional states, *Annu. Rev. Biophys. Biomed. Struct.* **31**, 257–273.
Yonath, A., Bashan, A. (2004) Ribosomal crystallography: initiation, peptide bond formation, and amino acid polymerization are hampered by antibiotics, *Annu. Rev. Microbiol.* **58**, 233–251.
Yusupov, M.M., Yusupova, C.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D., Noller, H.F. (2001) Crystal structure of the ribosome at 5.5 A resolution, *Science* **292**, 883–896.

Trans-Lesion DNA Synthesis: *see* Repair and Mutagenesis of DNA

Transplantation,organ and tissue: *see* Xenotransplantation in Pharmaceutical Biotechnology

Transport Across Cell Membranes: *see* Membrane Transport

Triacylglycerol Storage and Mobilization, Regulation of

Günter Müller, Stefan Petry, and Aventis Pharma Germany, a member of the Sanofi Aventis group
Therapeutic Department Metabolism, Frankfurt am Main, Germany

1 Triacylglycerol and Energy Storage 624
1.1 Storage and Mobilization of TAG 627
1.2 Lipolysis and Reesterification 627
1.3 TAG Storage/Mobilization and Disease 629
1.3.1 Diabetes Mellitus and Metabolic Syndrome 629
1.3.2 Lipotoxicity 630
1.3.3 Molecular Mechanisms of Lipotoxicity 634

2 Components for TAG Storage and Mobilization 635
2.1 TAG in Lipoproteins 635
2.2 TAG in Adipose Cells 635
2.2.1 Enzymes of TAG Synthesis 636
2.2.2 Lipid Droplets 641

3 Mechanism and Regulation of TAG Mobilization 656
3.1 HSL 656
3.2 cAMP 658
3.3 Phosphorylation of HSL 658
3.4 Dephosphorylation of HSL 661
3.5 Intrinsic HSL Activity 661
3.6 Translocation of HSL 662
3.6.1 Molecular Mechanism 662
3.6.2 Involvement of Perilipins 663
3.6.3 Involvement of ADRP 666
3.7 Feedback Inhibition of HSL 668
3.8 Expression of HSL 672
3.9 Non-HSL Lipases 672
3.10 Release of FFA 673

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 14
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30651-X

4 Physiological, Pharmacological, and Genetic Modulation of TAG Mobilization and Storage 675
4.1 Muscle Contraction 675
4.2 Glucose 676
4.3 Insulin 677
4.3.1 Molecular Mechanism 677
4.3.2 Desensitization 679
4.4 Leptin 680
4.5 Growth Hormone 680
4.6 Glucagon-like Peptide-1 681
4.7 TNF-α 681
4.8 Acylation-stimulating Protein 682
4.9 Acipimox, Nicotinic Acid, and Niacin 683
4.9.1 Physiological Mode of Action 685
4.9.2 Molecular Mechanism 686
4.9.3 Desensitization 687
4.10 Glimepiride and Phosphoinositolglycans 689
4.11 Differences in Regulation of TAG Storage and Mobilization between Visceral and Subcutaneous Adipocytes 690

5 Up-/Downregulation of Components of TAG Storage and Mobilization 691
5.1 HSL 691
5.2 ALBP 694
5.3 Perilipin 694
5.4 PKA 696
5.5 ASP 697
5.6 Caveolin 698

Bibliography 699
Books and Reviews 699
Primary Literature 699

Keywords

Diabetes Mellitus
Dysregulation of glucose and lipid metabolism in the course of reduced insulin secretion by pancreatic ß-cells and insulin action/sensitivity in muscle, adipose, and liver cells resulting in elevated blood glucose levels in the short term and diabetic late complications, atherosclerosis, and cardiovascular problems in the long term.

Hormone-sensitive Lipase
Hydrolase expressed in many cell types, which specifically cleaves mono/di/triacylglycerol in response to lipolytic stimuli, such as ß-adrenergic hormones, fasting, and muscle contraction.

Insulin
Major hormone regulating blood glucose and lipid homeostasis predominantly by inducing reesterification of postprandial glucose and free fatty acids to yield triacylglycerol for storage in lipid droplets predominantly in adipose tissue.

Lipid Droplets
Storage form of neutral lipids covered by a phospholipid monolayer and specific protein components involved in (regulation of) their synthesis and degradation.

Lipolysis
Enzymatic release of fatty acids and glycerol from mono/di/triacylglycerol stored in lipid droplets in many cell types.

Lipotoxicity
Impairment in responsiveness and sensitivity of insulin target cells (e.g. muscle, fat, and liver cells) toward insulin and of insulin-secreting β-cells toward glucose in the course of their prolonged exposure to elevated concentrations of fatty acids and intracellular accumulation of fatty acids in free and/or esterified form.

Translocation
Apparent movement of hormone-sensitive lipase from the cytoplasm to the surface of lipid droplets where cleavage of neutral lipids occurs in response to its stimulus-dependent phosphorylation.

Abbreviations

AA:	amino acid(s)
AC:	adenylate cyclase
AR:	adenosine receptor
α/β-AR:	α/β-adrenergic receptor
(M/D/T)AG:	(mono/di/tri)acylglycerol
DGAT:	acyl-CoA:diacylglycerol acyltransferase
DIGs:	detergent-insoluble glycolipid-enriched lipid raft microdomains
ER:	endoplasmic reticulum
(F)FA:	(free) fatty acids
H/LDL:	high/low density lipoprotein
HSL:	hormone-sensitive lipase

LD:	lipid droplets
LPL:	lipoprotein lipase
NA:	nicotinic acid
NEFA:	nonesterified fatty acids
NIDDM:	noninsulin-dependent diabetes mellitus
PDE:	phosphodiesterase
PI3K:	phosphatidylinositol-3-kinase
PKA/B/C:	protein kinase A/B/C
PP:	protein phosphatases
TNF-α:	tumor necrosis factor-α
VLDL:	very low-density lipoproteins
WAT:	white adipose tissue

A growing number of enzymes that are involved in lipid metabolic pathways are being identified and characterized. Here we summarize recent developments in the understanding of lipid storage and mobilization and assess the potential of lipid metabolic enzymes and mechanisms as targets for the pharmacological treatment of lipid disorders, noninsulin-dependent diabetes mellitus (NIDDM), obesity, and metabolic syndrome. We focus on enzymes and pathways that have been implicated in the biogenesis of lipid droplets and their degradation (lipolysis) on the basis of studies using (1) animal models that overexpress or harbor targeted deletions of the genes encoding the corresponding enzymes/proteins, (2) genetic and/or family studies of NIDDM and lipid disorders, and (3) modern biochemical/molecular biological methods for the identification and functional analysis of components involved in lipid metabolism.

1 Triacylglycerol and Energy Storage

Triacylglycerol (TAG) is the most concentrated form of energy available to biological tissues that emerged during biological evolution. For instance, the work covered by birds during their nonstop long-distance migrations is powered almost exclusively by fat reserves. The extra weight of carbohydrate required to produce the same calories would prevent the birds from ever taking off. In terms of human survival, the first unaided crossing of the polar ice cap was made possible by the very high butter-fat content of the 220-kg of food reserves transported by the sledges that were man-powered over the frozen wastes. The extra weight of a similar energy content as carbohydrate would have made this pioneering journey impossible in the short and medium terms. TAG also exhibits relatively low ''biological toxicity'' compared to FA (see below) and is well tolerated even at high concentrations in the blood plasma. TAG therefore provides a benign form of FA storage and transport. Before its use as an energy source, however, it must

undergo lipolytic mobilization to provide FA. The potent "toxicity" of FA in the long term requires a finely tuned regulatory mechanism that ensures that their release from TAG stores is correctly balanced by uptake and utilization. A number of clinical disorders of increasing incidence are associated with secondary hyperlipidemia affecting distinct lipoprotein subtypes via different mechanisms (Table 1).

Although the cells of most tissues synthesize TAG, only a few are adapted to store TAG in considerable quantity. Some, such as heart and skeletal muscle,

Tab. 1 Clinical disorders associated with secondary hyperlipidemia.

Disorder	***Lipoprotein type***	***Elevated plasma lipoprotein***	***Proposed mechanism***
Endocrine/Metabolic			
NIDDM	IV, V	VLDL, chylomicrons	VLDL production ↑ VLDL catabolism ↓
Hypothyroidism	IIa (III)	LDL (ß-VLDL)	LDL clearance ↓
Estrogen therapy	IV (V)	VLDL	VLDL production ↑
Glucocorticoid therapy	IIa, Iib	VLDL, LDL	VLDL production ↑ Conversion to LDL ↑
Hypopituitarism	Iib	VLDL, LDL	VLDL production ↑ Conversion to LDL ↑
Acromegaly	IV	VLDL	VLDL production ↑
Anorexia nervosa	IIa	LDL	Biliary excretion of bile acids/ cholesterol ↓
Lipodystrophy (congenital or acquired)	IV	VLDL	VLDL production ↑
Werner syndrome	IIa	LDL	Unknown
Acute intermittent porphyria	IIa	LDL	Unknown
Glycogen storage disease	IV (V)	VLDL	VLDL production ↑ VLDL catabolism ↓
NonEndocrine			
Alcohol	IV (V)	VLDL (chylomicrons)	VLDL production ↑
Nephrotic syndrome	IIa, Iib	VLDL, LDL	VLDL production ↑
Uremia	IV	VLDL	VLDL clearance ↓
Biliary obstruction/ cholestasis		LP-X	Diversion of biliary cholesterol and phospholipids in circulation
Hepatitis	IV	VLDL	LCAT activity/protein ↓
Systemic lupus erythematosis	I	Chylomicrons	Antibodies bind heparin and decrease LPL activity
Monoclonal gammapathy	IIa, III, IV	VLDL, IDL, LDL	Antibodies bind lipoproteins and interfere with catabolism

store small quantities of TAG for their own use (albeit the fact that the heart has been recently shown to have low capacity for lipoprotein export). In other tissues, for instance, in the pancreatic ß-cells, TAG synthesis, storage, and lipolysis may be involved in the regulation of the production of molecules, such as DAG and FA(-derivatives), which participate in various signal transduction processes or in providing a source of FA committed to membrane phospholipid synthesis. In mammals, of course, adipose tissue is the most important storage pool of whole-body TAG and, consequently, the body's largest energy reservoir; for example, an adult with 15 kg of body weight has >460 MJ of TAG fuel stores, which could provide 8.37 MJ daily for 2 months. The primary role of adipocytes is to store TAG during caloric excess and to mobilize this reserve when expenditure exceeds intake. Mature adipocytes are uniquely equipped to perform these functions. They possess the full complement of enzymes and regulatory proteins needed to carry out both TAG storage and mobilization (see below). In adipocytes, the regulation of these processes is exquisitely responsive to hormones, cytokines, and other factors that are involved in energy homeostasis.

Nevertheless, the liver also has an important TAG storage function. These two tissues export a proportion of their mobilized TAG. FA are released from adipose tissue, and TAG is exported as very low-density lipoproteins (VLDL) from the liver. Both organs are connected via a metabolic cycle in which FA released from adipose tissue are the major substrates for hepatic VLDL production and, in turn, FA from TAG contained in VLDL are returned to adipose tissue by the action of lipoprotein lipase (LPL). In addition, TAG in adipose tissue represents the major source of plasma glycerol, which accounts for about 90% of the substrates for hepatic gluconeogenesis at fasted condition in rodents. Central to the control of this intertissue cycle between white adipose tissue (WAT) and liver is the regulation of TAG mobilization and release by both organs. The different cellular and molecular mechanisms involved in the regulation of TAG mobilization from the storage pools in each of these tissues reflect the importance of the different roles of the lipids and FA, which are exported.

FA stored within the TAG of adipose cells constitutes the vast majority of energy reserves in animals. While the TAG pool in adipocytes provides a reservoir of energy for times of prolonged restricted caloric intake, there is also a normal "acute" flux of FA into and out of adipose tissue in response to meals and other factors, such as insulin. TAG hydrolysis in adipocytes, termed lipolysis, contributes to the increased levels of plasma nonesterified fatty acids (NEFA) by two mechanisms, one direct and the other indirect. In the postabsorptive state or during starvation FA are normally released from adipocytes directly into the plasma and transported to various tissues where they serve as a source of energy. In the insulin-resistant state, this direct release of FA is exacerbated, because the reduction in the antilipolytic action of insulin (see below) leads to increased lipolysis and a consequent increase in the release of FA from adipocytes into plasma. In contrast to the postabsorptive direct FA release, there is an indirect effect of adipocyte lipolysis on FA release in the postprandial state. There is a normal postprandial influx into adipocytes of FA produced by LPL action on VLDL and chylomicrons in the adipose endothelium, and this influx depends on a favorable downhill FA concentration gradient across the

adipocyte plasma membrane. Again, this indirect effect is exacerbated in the insulin-resistant state, since the increased lipolysis and, in consequence, the increased intracellular FA concentration in adipocytes abrogate the FA concentration gradient required for influx of FA. Thus, in an indirect fashion, adipocyte lipolysis can increase plasma FA by reducing the postprandial influx of FA into adipocytes. Given the apparent interrelated contribution of the cellular and LPL-mediated lipolytic reactions in adipose tissue to the plasma NEFA pool, it is important to understand the molecular basis of TAG hydrolysis by these mechanisms.

1.1 Storage and Mobilization of TAG

The two key enzymes involved in TAG mobilization within adipose tissue are LPL and (hormone-sensitive lipase) HSL. The function of LPL in mobilizing FA from TAG contained in VLDL and chylomicrons in the capillary endothelium of adipose tissue is not so clear as originally thought. It is now becoming evident that a variable (and regulated) proportion of NEFA released by LPL from TAG-rich lipoproteins does not enter adipose tissue but "leaks" into the plasma compartment and that this proportion may be under metabolic control dependent upon the nutritional state. For instance, in the postabsorptive state and during starvation, very little, if any, NEFA derived from TAG of VLDL enter the adipose tissue. The major portion enters the blood plasma and supplements the HSL-mediated provision of NEFA to the periphery during this period. Any excess NEFA produced, which override the immediate oxidative capacity of energy-requiring tissues, are reesterified by the liver and stored as TAG or VLDL. The potential hazards of hyperlipidemia for the cardiovascular system in the postprandial state are now well established and accepted (Table 1). The liver may, thus, play a major protective role during this period by removing excessive NEFA from the circulation, temporarily storing them as a "benign" derivative, TAG, and secreting them later as VLDL when the period of maximum danger has passed. This protective function may prove to be yet another facet of the liver's more general role as a frontline defense to protect body tissues from the potentially dangerous but inevitable consequences of major and abrupt physiological transitions. In this way, the liver may act as an essential (storage) buffer by which the supply of and demand for adipose tissue FA are matched in a way that does not result in or neutralize the potential FA-induced "lipotoxicity" to peripheral tissues.

1.2 Lipolysis and Reesterification

Adipocytes are highly differentiated cells specialized in handling large quantities of long-chain FA (Fig. 1). During lipid storage in the fed state, FA are released by LPL bound to the endothelial capillary wall into the blood from chylomicrons and lipoproteins assembled and secreted by intestinal and liver cells, respectively (step 1), or from albumin, move through the endothelium via passage through tight junctions or endothelial cells (step 2), bind to the outer leaflet of the plasma membrane of the target cells, and cross the membrane bilayer by simple diffusion or protein-mediated transport (step 3). While the direct involvement of FA not bound to albumin in FA uptake by adipocytes is generally accepted, it is still debated whether,

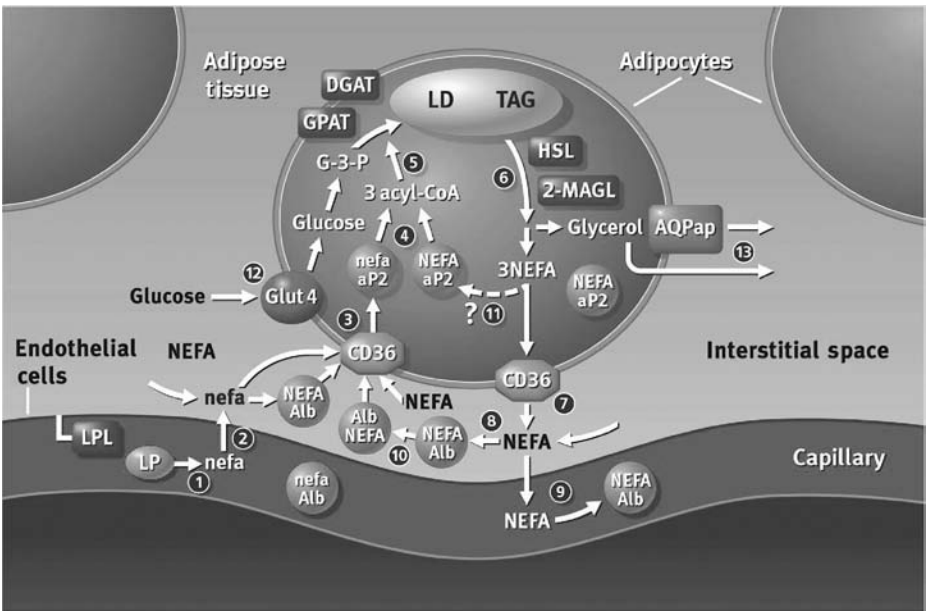

Fig. 1 Working model of FA reesterification and its control by the efficacy of removal of newly released NEFA from the interstitial space into the circulation. Elevation/reduction of the blood flow would decrease/increase the accumulation of NEFA in the interstitial space in the vicinity of adipocytes where the endogenously released NEFA intermingle with and are diluted by the nefa generated by LPL action from lipoproteins (LP) in the endothelial capillaries, and thereby decrease/increase the probability of reuptake and reesterification of NEFA released from adipocytes by HSL and 2-monoacylglycerol lipase (2-MAGL) action. This allows rapid adaptation of the supply of NEFA from adipose tissue according to the acute requirements of the organism (e.g. during starvation), which, to a certain degree, operates independent of the actual (insulin-controlled) lipolytic/antilipolytic state of the adipocytes. The major portion of NEFA/nefa in the interstitial space and capillaries is bound to serum albumin (Alb). The transport of glucose (as precursor for glycerol-3-phosphate, G-3-P) and the bidirectional flux of NEFA/nefa across the adipocyte plasma membrane is facilitated by GLUT4 and CD36/FAT, respectively, whereas simple diffusion and/or specific transport via AQPap is involved in the release of glycerol from adipocytes. See text for details.

and if so to what extent, the albumin-bound FA pool is directly engaged in this process. Using the hindlimb perfusion technique, it was demonstrated recently that FA incorporation into intramyocellular TAG depends not only on the unbound to albumin FA concentration but also, to some extent, on the total FA concentration. This supports the previous speculation that albumin-bound FA could be directly involved in the uptake process mediated by an interaction of the FA-albumin complex with an albumin receptor located at the cell surface, putative candidates for which are the albumin-binding proteins (ABP). They could either promote a direct transfer of FA from the complex into the plasma membrane or facilitate dissociation of FA molecules from albumin, and thus locally increase the number of FA molecules available for uptake. Subsequently, FA transported across the plasma

membrane are trapped in the cytoplasm by conversion into acyl-CoA (step 4), reesterified, and stored primarily as TAG in LD (lipid droplets) (step 5). During lipolysis of stored TAG in the fasting state, FA are released from intracellular LD primarily by action of HSL and 2-monoacylglycerol lipase (step 6), move to and cross the plasma membrane (step 7), and are released into the interstitial space, where they bind to albumin (step 8). Subsequently, the FA pass through the endothelial cells via transcytosis or diffuse through the spaces between them (tight junctions) to reach the blood (step 9).

The temporal and spatial relationship between breakdown (lipolysis) and resynthesis (reesterification) of TAG (TAG-FA cycling) has to be tightly controlled to either provoke or avoid generation of heat and energy expenditure through the utilization of ATP during substrate cycling. A substrate cycle exists when opposing, nonequilibrium reactions catalyzed by different enzymes are operating simultaneously. At least one of the reactions must involve the hydrolysis of ATP. Thus, a substrate cycle both liberates heat and increases energy expenditure in the absence of net conversion of substrate into product. Some of the NEFA formed during lipolysis can be reesterified to TAG, whereas little or no glycerol is reutilized by fat cells. This pathway of lipolysis and NEFA reesterification forms an important cycle for energy turnover, allowing the fat cell to respond rapidly to changes in peripheral requirements for NEFA (Fig. 1). This is reflected in the FA concentration gradient between plasma and adipose tissue (mainly fat cell cytosol), which is determined by both the bloodstream and uptake of NEFA by peripheral tissues. Two types of NEFA reesterification in fat cells can be recognized. Primary reesterification is the total amount of NEFA that is reesterified during a given situation and reflects the TAG synthesis capacity of the fat cells. Fractional reesterification is the proportion of NEFA reesterified in relation to the amount of NEFA formed by lipolysis in fat cells. The latter constitutes a futile cycle, energy-rich NEFA first being formed by lipolysis of TAG and then resynthesized to TAG. There is experimental evidence that NEFA have to be released from adipocytes into the interstitial space, where they form a common pool with NEFA generated by LPL action, and subsequently be taken up by the adipocytes for efficient reesterification rather than that NEFA accumulating in the cytoplasm can be used for a "short-circuit" reesterification. Fat cell reesterification can be determined *in vitro* by simultaneous measurement of the release of glycerol and NEFA using fluorescence, luminescence, or dual radioisotope techniques or a combination thereof.

1.3 TAG Storage/Mobilization and Disease

1.3.1 Diabetes Mellitus and Metabolic Syndrome

During the last decade, the traditional glucose-insulin axis with respect to examining glucose tolerance and as a basic pathophysiological mechanism operating in diabetes, although remaining diagnostically useful, has been criticized. The glucose-insulin axis has been overemphasized, while alterations in the insulin to NEFA ratio and "NEFA" metabolism received much less attention. The major change in the appreciation of the nature of molecules supplying body energy came with the discoveries that plasma NEFA are the major "lipid vehicles" in the mammalian organism. Furthermore, NEFA are

the only form of TAG mobilization from adipose tissue, subject to hormonal regulation. The very rapid turnover of plasma NEFA indicates that they are an important fuel even in nonfasting conditions and represent an energy source to body tissues equivalent to or greater than glucose. The importance of NEFA in the regulation of gluconeogenesis was extensively dealt with, but no clear pathogenic conclusion with relation to major diabetes perturbations emerged. A breakthrough was the demonstration in 1964, that infusion of a very small amount of insulin through the brachial artery in human forearm leads to a significant decrease in plasma NEFA concentrations without any concomitant effect on glucose uptake. Consequently, the restraint of NEFA release from adipose tissue is the most sensitive action of insulin compared with other insulin effects. The lack of restraint of NEFA mobilization in hypoinsulinemic states during pathogenesis of noninsulin-dependent diabetes mellitus (NIDDM) leads to a marked plasma NEFA increase, ectopic deposition of TAG and insulin resistance with regard to nonoxidative as well as oxidative glucose metabolism. The latter fact stresses the negative effect of excessive availability of NEFA in the short term on insulin-mediated glucose utilization in muscle. It involves a reduction of muscle glycolysis by inhibiting phosphofructokinase in the course of elevated cytosolic concentrations of citrate and consequent accumulation of glucose-6-phosphate. Pyruvate dehydrogenase is also inhibited by rising mitochondrial concentrations of acetyl-CoA and NADH. In the liver, the same changes in redox ratio and activation of pyruvate carboxylase by acetyl-CoA result in stimulation of gluconeogenesis. This mechanism links acute NEFA oversupply to impaired glucose tolerance or hyperglycemia in the developing or frank NIDDM, respectively.

Moreover, the increased plasma NEFA concentration in NIDDM patients provides substrate for VLDL production by the liver (see above), which is manifested as hypertriglyceridemia. Thus, drugs with the potential of lowering NEFA levels may reduce hepatic VLDL production, improve glucose tolerance and counteract lipotoxicity and therefore be useful for antidiabetic therapy. The generation and careful analysis of transgenic and knockout animals has aided insight into the recognized but partly misunderstood participation of cellular lipid metabolism in energy homeostasis. Numerous mouse models with disruptions in pathways relevant to FA as well as TAG storage and mobilization have been generated with dysregulated circulation of FA in the bloodstream as nonesterified molecules, bound to albumin, or in the core of lipoproteins, such as VLDL (see below).

1.3.2 Lipotoxicity

In the last decade, the deleterious effects of the deposition of TAG in nonadipose tissues during long-term elevations of plasma NEFA have been widely acknowledged. Whereas adipocytes have a unique capacity to store excess FA in the form of TAG in LD, nonadipose tissues, such as cardiac and skeletal muscle myocytes and pancreatic ß-cells, have a limited capacity for storage of TAG. In hyperlipidemic states, accumulation of excess TAG in nonadipose tissues leads to cell dysfunction and/or cell death, a phenomenon known as lipotoxicity and lipoapoptosis, respectively. Excess lipid accumulation in skeletal muscle is associated with the development of insulin resistance.

The "lipotoxic" consequences of excessive FA release from adipose tissue for pancreatic ß-cells have been described in a provocative report by Unger and colleagues. Abnormal ß-cell function and insulin resistance are the major characteristics in the pathogenesis of NIDDM. Lipotoxicity was proposed as an attractive concept to explain parallel developments in the impairment of insulin signaling and ß-cell function. Before the onset of overt NIDDM, insulin resistance was closely associated with ß-cell dysfunction characterized by ß-cell hyperplasia and insulin hypersecretion. During further progression of the disease, however, the ß-cell could no longer compensate for peripheral insulin resistance and declined in its function. From animal models of obesity-linked NIDDM, such as the obese and insulin-resistant Zucker rats, as well as *Psammomys obesus*, both impairment of insulin secretion linked to intracellular fat deposition and loss of cell mass by FA-induced apoptotic cell death contribute to the development of relative insulinopenia. Exposure of NEFA has cytostatic and proapoptotic effects on human pancreatic ß-cells. Lipid overload in pancreatic ß-cells also leads to dysregulated insulin secretion followed by apoptotic cell death. Furthermore, human autopsy studies show a relative reduction of ß-cell mass in patients with NIDDM compared with weight-matched nondiabetic subjects. These findings support the idea that, in genetically predisposed human subjects, prolonged exposure to elevated NEFA may contribute to ß-cell death and development/progression of NIDDM. The intracellular signaling pathways for FA-induced ß-cell apoptosis are incompletely understood. Different mechanisms for lipoapoptosis, such as increased ceramide formation, mitochondrial cytochrome *c* release, and nitric oxide generation, have been suggested. Furthermore, in different cell types, the FA-dependent signaling process might involve activation of different protein kinases C isoforms. However, the exact role of PKC isoforms in FA-induced apoptosis has not been defined (see below).

TAG storage and lipolysis in the pancreatic ß-cell may play a dual role. First, by way of storage, as a protective mechanism against the toxic effects of excess NEFA, and second by way of lipolysis to provide a signal (FA, DAG) that participates in the stimulus/secretion-coupling that regulates acute insulin release. With regard to the latter function, islets depleted of TAG show an abrogated stimulation of insulin release by glucose. HSL may balance TAG storage and TAG mobilization in pancreatic ß-cells, and, thus, via PKA, provide a critical link in the stimulus/secretion-coupling response. HSL is predominantly expressed in pancreatic islets as a 3.1-kb mRNA encoding an 89-kDa isoform containing an additional 43 aa N-terminal to the adipocyte form of the protein, as well as the 2.8-kb mRNA encoding the 84-kDa adipocyte form. HSL immunoreactivity in islets is detected primarily in ß-cells, but some HSL is also observed in α-cells. Long-term incubation of ß-cells or islets with high concentrations of glucose increases HSL mRNA and protein expression, as well as HSL activity. This induction of HSL expression appears to be regulated transcriptionally and also appears to depend on the metabolism of glucose. Treatment of ob/ob mice with leptin increased HSL expression in islets along with a decrease in islet TAG content and an improvement in insulin secretion. HSL is definitely functional in islets, since islet TAG content is increased 2

to 2.5-fold in HSL null mice compared with wild-type mice. These observations together with studies on insulin release in HSL null mice upon glucose challenge suggest that HSL might be important in glucose-induced insulin secretion and are consistent with the current view that FA metabolites are critical for insulin secretion, but that excessive NEFA can lead to lipotoxicity and dysfunction of ß-cells. The exact function of HSL in insulin secretion from ß-cells awaits further study.

The FA supply to ß-cells for lipid-signaling processes can be from exogenous and endogenous sources. Studies with HSL-deficient mice show roles for HSL in the maintenance of both exogenous and endogenous FFA supplies and, in at least some circumstances, the preservation of normal glucose-stimulated insulin secretion (GSIS). Thus, with respect to the exogenous FFA supply and consistent with previous reports, HSL deletion in male mice was associated with markedly reduced plasma FFA and TAG levels during fasting. For the endogenous supply, there is evidence of reduced access to islet TAG stores, as the TAG stores were increased in association with reduced islet TAG lipase activity and lipolysis, particularly in 7-month-old mice. Given that the TAG lipase activity was most affected in older mice, and plasma lipid levels were most reduced in the fasting state in male mice, ß-cell islet FA deprivation was almost certainly greatest in the older fasted male mice. Therefore, the finding of the greatest loss of GSIS in the fasted 7-month-old male $HSL^{-/-}$ mice, together with the reversibility of this loss with FFA replacement, are completely consistent with the essential role of FA supply in normal insulin secretion. Furthermore, triacsin C-experiments showed that activation of FFA to long-chain acyl-CoA is necessary for this FA effect, consistent with the malonyl-CoA/long-chain acyl-CoA signaling pathway of insulin secretion. These findings are also consistent with triacsin C diminishing FFA augmentation of insulin secretion stimulated by glucose or nonfuel secretagogues. The distal mechanism by which long-chain acyl-CoA promote insulin secretion, however, is not known. Possibilities include a direct long-chain acyl-CoA effect on exocytotic processes or indirect actions *via* acylation of proteins or esterification into lipid-signaling molecules, such as phosphatidate and DAG. Importantly, available data underscore a role for ß-cell HSL in the provision of an endogenous FFA supply when exogenous FA supplies are low. The importance of lipolysis for the formation of a lipid-derived coupling signal was confirmed by the fact that inhibition of lipolysis in rat islets by orlistat reduced insulin secretion.

The islet neutral cholesterylester hydrolase and TAG lipase activity data are consistent with lipase activity distributions in other tissues, which show the responsibility of HSL for most or all of the cholesterylester hydrolase and part of the TAG lipase activity. However, it remains unknown if DAG hydrolase activity is reduced in $HSL^{-/-}$ islets. Although HSL is the main source of DAG hydrolase activity in adipocytes, this activity in liver and skeletal muscle may occur *via* another enzyme(s)(see below). Interestingly, the relative importance of HSL-TAG lipase activity to overall TAG lipase activity increased as the $HSL^{-/-}$ mice aged. This provides an explanation for the detection of reduced basal lipolysis in the older mice only. The higher islet TAG content in the islets

from 4-month-old $HSL^{-/-}$ mice, however, strongly suggests abnormal lipolysis was also present in these islets, although hardly measurable as glycerol release in the *in vitro* assay. Clearly, the lipase and lipolysis results are indicative of at least one other lipase with TAG lipase activity in islets. The degree to which this other enzyme(s) is upregulated in the ß-cell of the $HSL^{-/-}$ mice to compensate for the total deficiency of HSL remains to be elucidated.

Of considerable interest is the demonstration of glucose-stimulated lipolysis in islets from $HSL^{+/+}$ and $HSL^{-/-}$ mice. Thus, glucose-stimulated lipolysis has now been shown to occur in mouse and rat islets. This process may prove to be important for signal transduction in GSIS. The finding that glucose increased lipolysis in $HSL^{-/-}$ islets is indicative of an alternate lipase being responsive to glucose. Why this does not prevent loss of GSIS, however, is not known. It may have been due to differences in subcellular localization of the lipases. Interestingly, islet HSL has recently been shown to be associated with insulin secretory granules.

Several groups have now generated $HSL^{-/-}$ mice, and much heterogeneity is reported in the metabolic phenotype, particularly in relation to the measurement of insulin sensitivity and ß-cell function. Insulin sensitivity assessed in one study by the intraperitoneal insulin tolerance test was reduced, whereas in another study, insulin sensitivity, calculated from the intraperitoneal glucose tolerance test, was enhanced in the $HSL^{-/-}$ mice, as glucose levels during the tolerance test remained unaltered in the face of reduced circulating insulin in $HSL^{-/-}$ mice. Furthermore, GSIS in isolated islets from fasted $HSL^{-/-}$ mice was impaired in male mice but enhanced in female mice. Although at first sight the data and literature seem confusing, some patterns relating to nutritional state and sex are now evolving. With respect to nutritional state, serum FFA and TAG levels are usually unchanged in the fed state, but markedly reduced in the fasted state in $HSL^{-/-}$ mice compared to $HSL^{+/+}$ mice. Similarly, hepatic TAG content has been reported as unchanged or increased in the fed state and decreased in the fasted state in $HSL^{-/-}$ mice. With the well-known inverse correlations between plasma FFA and hepatic TAG content and insulin sensitivity, it is conceivable that these abnormally wide fluctuations in whole animal lipid partitioning with nutritional state cause swings between relatively reduced (fed state) and enhanced (fasted state) insulin sensitivity in the $HSL^{-/-}$ mice compared to wild-type mice. This favors the hypothesis on the differences in the nutritional state (fed vs. fasted) at the time of assessment. The degree to which insulin sensitivity changes with the nutritional state, however, may also relate to the amount of fat in the diet and whether the mice are male or female.

It is of interest to discuss these findings in relation to results obtained with mice transgenic for HSL specifically in ß-cells. These mice, having been fed a high-fat diet, have impaired glucose tolerance with poor GSIS. It was concluded that excess HSL activity in the islet prevented detoxification of the increased fat supply, as the potentially protective diversion of FA to TAG stores was blocked. Both these HSL-transgenic mice and the $HSL^{-/-}$ mice highlight the importance of normal lipolysis pathways with normal regulation of lipid stores for normal insulin secretion.

1.3.3 Molecular Mechanisms of Lipotoxicity

In various experimental systems, saturated and unsaturated FA differ in their contributions to lipotoxicity. Previous studies in (Chinese hamster ovary) CHO cells, cardiac myocytes, pancreatic ß-cells, breast cancer cell lines, and hematopoietic precursor cell lines suggest that lipotoxicity from the accumulation of long-chain FA is specific for saturated FA. This selectivity has been attributed to the generation of specific proapoptotic lipid species or signaling molecules in response to saturated but not unsaturated FA. The nature of such signals may differ across cell types, but includes reactive oxygen species generation, *de novo* ceramide synthesis, nitric oxide generation, decreases in (phosphatidylinositol-3-kinase) PI3K, primary effects on mitochondrial structure and function and suppression of antiapoptotic factors, such as *Bcl*II.

A recent study provided evidence that the differential toxicity of these FA is directly related to their ability to promote TAG accumulation. It was demonstrated that exogenous or endogenously generated unsaturated FA rescue palmitate-induced apoptosis by promoting palmitate incorporation into TAG. Moreover, unsaturated as well as saturated long-chain FA were toxic in cells with impaired TAG synthetic capacity (e.g. lack of DGAT activity). *In vivo*, TAG accumulation and storage in nonadipose tissues may occur in the setting of mismatch between cellular FA influx and FA utilization. TAG accumulation in nonadipose cells provoked by either enhanced synthesis or impaired mobilization in response to acute FA overload may represent an initial cellular defense mechanism against lipotoxicity. Consequently, accumulation of TAG in nonadipose tissues probably serves as a barometer of the FA overload state in human disorders such as hyperlipidemia and lipodystrophies, and in animal models such as the Zucker (diabetic) fatty rat or the ob/ob mouse. Accumulation of cellular TAG *per se* does not seem to be toxic, at least initially. Rather, storage of excess FA in TAG pools of "inert" LD probably diverts these molecules from pathways that lead to cellular dysfunction and/or cytotoxicity and may thus serve as a buffer against lipotoxicity. In pathologic states, lipotoxicity may occur over time, despite TAG accumulation and storage, when either the cellular capacity for TAG storage is exceeded or when TAG pools are hydrolyzed (during HSL action), resulting in increased cellular NEFA levels. Thus, the duration and extent of TAG overload in conjunction with (regulation of) TAG storage and mobilization may determine whether a cell is protected or damaged.

This mechanism may also underlie the FA-induced apoptosis in rat and human pancreatic ß-cells. Chronic treatment with high physiological levels of saturated FA, but not with mono- or polyunsaturated FA, triggers apoptosis in about 20% of cultured ß-cells. Apoptosis restricted to saturated FA was also observed in primary cultured human ß-cells, suggesting that this mechanism is potentially relevant *in vivo* in humans. Apoptosis was accompanied by a rapid nuclear translocation of protein kinase C-delta (PKC-d) and subsequent lamin B1 disassembly. This translocation was impaired by inhibition of phospholipase C, which also considerably reduced apoptosis. Furthermore, lamin B1 disassembly and apoptosis were decreased by cell transfection with a dominant-negative mutant form of PKC-d. These data suggest that nuclear translocation and kinase activity of PKC-d are both necessary for saturated FA-induced

lipoapoptosis. In analogy to recent findings with CHO cells, it is conceivable that in ß-cells accumulation of DAG generated by phospholipase C action and the resulting activation of PKC-d can be prevented by conversion of DAG into and storage as TAG via an ill-defined regulatory process of DAG acylation that is stimulated by unsaturated FA.

2 Components for TAG Storage and Mobilization

2.1 TAG in Lipoproteins

Homozygous VLDL receptor knockout mice are resistant to both genetic and diet-induced obesity caused by a decreased peripheral and whole-body uptake of NEFA with no alterations in either food intake or fat absorption. The reduction in adipocyte TAG storage as shown by a decreased average fat cell size in VLDL receptor-deficient rodents implies an impaired FA delivery to adipose tissue in the absence of this lipoprotein receptor. Analogously, hepatic overexpression of human apolipoprotein C-I leads to hyperlipidemia accompanied by decreased visceral fat depots and lack of subcutaneous WAT in mice. Conversely, overexpression of apoA-II, the second most abundant HDL component, increases adiposity and insulin resistance in relation to decreased skeletal muscle glucose utilization. In the capillaries of skeletal muscle and adipose tissue at the luminal face of endothelial cells, LPL catalyzes the rate-limiting step in the hydrolysis of TAG from circulating VLDL and chylomicrons. Thus, LPL plays an important role in directing fat partitioning. In fact, complete LPL deficiency in mice results in minimal amounts of tissue lipids, leading to neonatal death due to marked hypoglycemia and hypertriglyceridemia. In heterozygotes, only mild hypertriglyceridemia with impaired LDL clearance and mild hyperinsulinemia accompanied by an approximately 20% decrease in fasting glucose concentrations were observed. Exclusive LPL deficiency in adipose tissue on a standard genetic background leads to normal growth and body composition. This implies that, although LPL controls FA entry into adipose tissue, fat mass is preserved by endogenous FA synthesis. When the same lack of LPL in adipose tissue is generated on an ob/ob background, a diminished weight gain is attained owing to an impaired lipid accumulation in adipocytes. Conversely, targeted overexpression of LPL in skeletal muscle or liver has no effect on body weight, but produces an increase in FA uptake into the respective tissue that adversely affects glucose metabolism.

2.2 TAG in Adipose Cells

Adipocytes, which vary enormously in size (20–200 μm in diameter), are embedded in a connective tissue matrix and are uniquely adapted to store and mobilize energy as TAG and NEFA/glycerol, respectively. Surplus energy is assimilated by fat cells and stored as TAG in LD. To accommodate the TAG, adipocytes can change their diameter up to 20-fold and their volumes by several 100-fold. Because ~90% of the cell volume is constituted by LD(s), the nucleus and the thin cytoplasm rim are pushed to the periphery of the adipocytes. WAT is actively involved in cell function regulation through a complex network of endocrine, paracrine, and autocrine signals that influence the response of many tissues,

including hypothalamus, pancreas, liver, skeletal muscle, and others. Until recently, adipocytes were seen as passive tissues for the storage of excess energy as TAG. However, they clearly act as endocrine secretory cells, releasing in constitutive or regulated fashion, several hormones (e.g. leptin), growth factors (TGFß), cytokines (e.g. TNF-α), and other factors (e.g. ASP) that act as feedback signals for the adipose tissue itself or for other tissues. Adipose tissue signaling pathways, arranged in a hierarchical fashion, constitute one of the voices of the body that enable the organism to adapt to a range of metabolic challenges, such as starvation, stress, and infection, as well as periods of energy excess.

2.2.1 Enzymes of TAG Synthesis

Accumulating evidence for important functions of enzymes involved in lipid synthesis and its control has been obtained through targeted disruption in rodent models of the corresponding genes (Fig. 2). However, direct links between abnormal expression or genetic variants and human disorders, such as obesity, hyperlipidemia, insulin resistance, and NIDDM await further clarification.

2.2.1.1 **Glycerol-3-Phosphate Dehydrogenase.** Glycerol-3-phosphate dehydrogenase (GPDH) is a ubiquitously expressed enzyme that participates in TAG synthesis and in shuttling NADH into mitochondria for oxidative metabolism. Transgenic mice overexpressing GPDH show a normal body weight but with an increased interscapular brown fat depot and virtually no WAT. Conversely, animals lacking mitochondrial GPDH show a decreased body weight compared with wild-type mice with no reported effect on adipose tissue depots but a marked insulin release defect when the malate–aspartate shuttle is blocked. This indicates an important role of the NADH shuttle or glycerol phosphate system in glucose-induced

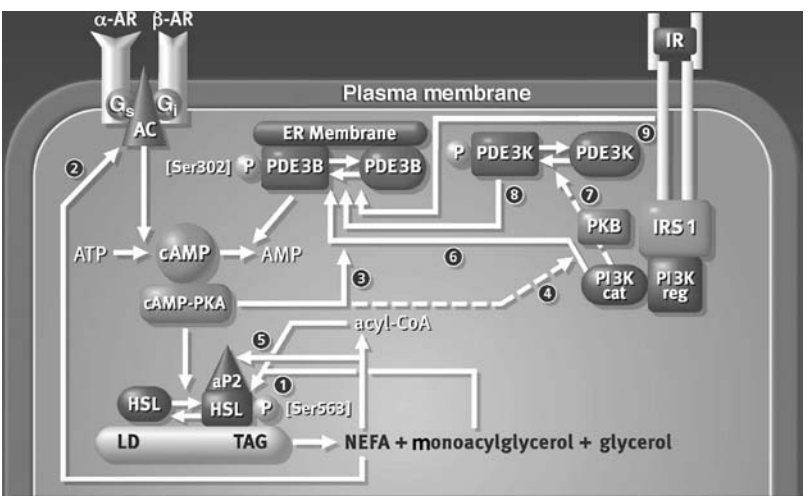

Fig. 2 Schematic model of the main signaling and metabolic pathways coupling hormonal stimuli and NEFA to (regulated) TAG synthesis, storage, and mobilization in adipocytes. PKA-dependent phosphorylation is depicted as filled circles. See text for details.

activation of mitochondrial metabolism and insulin secretion. Therefore, the two mouse models with altered expression of GPDH provide evidence for an involvement of this enzyme in lipogenesis and WAT development as well as in the regulation of the glycolytic pathway.

2.2.1.2 **Acetyl-CoA Carboxylases.**

Malonyl-CoA is a, key metabolite generated by acetyl-CoA carboxylases (ACC), which plays a pivotal role in linking FA and carbohydrate metabolism through regulation of FA synthesis and oxidation, respectively, in response to hormonal and dietary influences (Fig. 2). In mammals, ACC exists in two isoforms, ACC1 and ACC2, which are encoded by two different genes. ACC1 is the principal isoform in lipogenic tissues, whereas ACC2 is predominantly expressed in oxidative tissues, such as the heart and skeletal muscle. Both isoforms are expressed in the liver, where FA synthesis and oxidation coexist. ACC1 is a cytosolic enzyme, whereas ACC2 is associated with the mitochondrial membrane through its extended amino terminus. This difference in compartmentalization might contribute to the diverse roles of the two isoforms in fat metabolism; ACC1 is believed to regulate fat synthesis in lipogenic tissues, whereas ACC2 controls the rate of lipid oxidation in oxidative tissues. However, the recent identification of a mitochondrial form of FAS indicates that ACC2 might also have a role in FA synthesis in mitochondria.

ACC activity is mainly regulated by citrate and AMP kinase (AMPK). Citrate is believed to be an allosteric activator of ACC, as evidenced by a strong positive correlation between levels of malonyl-CoA and citrate concentrations. For example, elevated ACC activity occurs concomitantly with a threefold increase in cellular citrate concentration following the acetoacetate treatment of acini cells isolated from lactating rat mammary gland. Phosphorylation of ACC by AMPK inhibits its activity, which is accompanied by a reduction in malonyl-CoA levels. This is seen both in reperfused rat hearts following ischemia and in skeletal muscle following treadmill exercise or electrical stimulation. Furthermore, the incubation of isolated rat soleus muscles with the AMPK activator AICAR decreases ACC2 activity and simultaneously increases the rate of FA oxidation.

The effect of ACC2 on FA oxidation is also observed in knockout mice that are deficient in this enzyme. *Acc2*-null mice have a elevated rate of FA oxidation concomitant with decreased malonyl-CoA levels in the skeletal muscle and heart. These knockout mice are resistant to diet-induced obesity and related diabetes, with improved insulin sensitivity and decreased FA levels. The mRNA levels of uncoupling proteins (UCPs) are markedly higher in the adipose, heart (UCP2), and muscle (UCP3) tissues of mutant mice, which together with increased ß-oxidation might account for an upregulation of energy expenditure. Further analysis has also confirmed an important role of ACC2 in the malonyl-CoA/CPT1 axis, as illustrated by an increase in the CPT1 activity of hepatocytes that are isolated from *Acc2*-null mice. Surprisingly, CPT1 activity in skeletal muscle was not notably changed by ACC2 deficiency. The knockout mice have a normal lifespan with no apparent pathophysiological conditions caused by ACC2 deficiency. Several mammalian ACC inhibitors have been developed and tested for physiological effects on TAG synthesis, although ACC2-specific inhibitors have not been reported. One of the mammalian ACC inhibitors, CP-610431, shows dual inhibition of both

ACC1 and ACC2. When tested *in vivo*, several of the dual inhibitors reduced TAG synthesis and increased FA oxidation in rats, with a concurrent reduction in malonyl-CoA levels in the liver, muscle, and heart. These results further support a role for malonyl-CoA and validate the concept that the inhibition of ACC enzymes with small molecules might provide a viable treatment option for obesity. In addition, the development of dual inhibitors can prevent the feedback responses that are caused by the inhibition of a single ACC isoform, as observed in *Acc2*-knockout mice, therefore providing better overall efficacy.

2.2.1.3 **Fatty Acid Synthase.** Mammalian fatty acid synthase (FAS) catalyzes the *de novo* synthesis of saturated FA, such as myristate, palmitate, and stearate, using acetyl- and malonyl-CoA. It functions as a homodimer of a multifunctional protein that contains seven catalytic domains and a site for the prosthetic group 4′-phosphopantetheine. The enzyme is abundantly expressed in lipogenic tissues, such as liver, adipose, and lactating breast. FAS is believed to be important in coordination with CPT and ACC in maintaining energy homeostasis by converting excess food intake into lipids for storage and providing energy by upregulating the rate of mitochondrial oxidation. FAS might also have a role in carcinogenesis and increased FAS levels in cancer tissues indicate a poor prognosis. Much of our understanding of the metabolic role of FAS comes from investigations using two FAS inhibitors: cerulenin and C75. C75 was initially developed for its antitumor activity. Administration of C75 caused a dose-dependent decrease in food intake in BALB/c mice, and blocked the fasting-induced upregulation of orexigenic neuropeptides and the downregulation of anorexigenic neuropeptides in the hypothalamus. Similar results were observed with cerulenin, although with much less efficacy. Intracerebroventricular (ICV) injection of C75 resulted in the dose-dependent inhibition of feeding, which could be circumvented by ICV injection of neuropeptide Y.

One of the most popular models put forward to explain the weight-loss effects of FAS inhibitors posits that the level of malonyl-CoA in a subset of neurons regulates feeding, and inhibition of FAS causes a buildup of malonyl-CoA in the central nervous system, thereby suppressing the appetite of animals. This model was supported by the observations that administration of C75 to fasted mice rapidly increased hypothalamic malonyl-CoA and blocked feeding when the mice were presented with food. However, this model was challenged by the recent finding that C75 lacked target specificity for proopiomelanocortin (POMC) neurons and acted as a nonspecific neuronal activator. In support of the argument, C75 was reported to act as a possible alkylating agent and cerulenin was shown to inhibit protein acylation. In addition, C75 was reported to activate CPT1. Treatment of mice with C75 led to increases in both energy consumption and FA oxidation in diet-induced obesity that could be reversed by etomoxir, which is an inhibitor of CPT1.

2.2.1.4 **Glycerol-3-Phosphate Acyltransferase.** The synthesis of TAG and glycerophospholipids starts with the acylation of glycerol-3-phosphate by glycerol-3-phosphate acyltransferase (GPAT) to form lysophosphatidic acid (Fig. 2). GPAT activity in most mammalian tissues involves at least two isoforms of GPAT enzymes, the mitochondrial (GPAT1) and endoplasmic reticulum (erGPAT) isoforms, which are

encoded by two different genes. erGPAT accounts for 90% of the total enzyme activity in most tissues, with the exception of the liver, in which GPAT1 contributes about 50%. GPAT1 expression and enzyme activity are regulated by nutritional and hormonal factors in a manner that is consistent with a crucial role in TAG synthesis. Overexpression of GPAT1 in mammalian cell lines results in a notable increase in TAG synthesis. By contrast, inactivation of GPAT1 in mice causes a reduction in weight gain accompanied by significant decreases in hepatic TAG content, plasma TAG, and low-density-lipoprotein TAG secretion. Ablation of GPAT1 also results in mice that show a marked shift in the incorporation of unsaturated FA into TAG and phospholipids.

How GPAT1 enzyme deficiency affects TAG synthesis remains elusive, as mitochondria lack a complete set of TAG synthesis enzymes, such as DGAT, which catalyzes the final step in TAG synthesis. One proposed mechanism for the fat loss that is observed in *Gpat1*-knockout mice predicts a role for GPAT1 as a mediator of lipid oxidation. Accordingly, as GPAT1 and carnitine palmitoyltransferase-1 (CPT1) reside on the outer mitochondrial membrane, and both seem to be reciprocally regulated by AMP-activated kinase, elimination of GPAT1 activity would prevent the direction of fatty acyl-CoAs toward TAG synthesis and avoid competition with CPT1 for acyl-CoAs. Further investigation is needed to verify this hypothesis; for example, comparing the rate of mitochondrial oxidation between *Gpat1*-knockout mice and controls. Although no obvious pathophysiological conditions are associated with GPAT1 deletion, a real concern with targeting GPAT1 for obesity is its effect on phospholipid synthesis. For example, the inhibition of GAPT1 could potentially impair the synthesis of cardiolipin, which is a glycerophospholipid that is important in maintaining the normal function of mitochondria. Despite the considerable progress that has been made in deciphering individual residues that are important for the catalytic activity of GPAT1, information is still lacking on the three-dimensional structure of the enzyme and on any reported efforts to develop GPAT1 inhibitors.

2.2.1.5 **Monoacylglycerol Acyltransferase.** Monoacylglycerol acyltransferase (MGAT) catalyzes the first step in TAG resynthesis, which is involved in dietary absorption in enterocytes and represents one of the two main biochemical pathways that are involved in TAG synthesis. The activity of each pathway is determined by the relative abundance of 2-monoacylglycerol and FFA relative to glycerol-3-phosphate. In intestinal mucosa under normal lipid-absorption conditions, the MAG pathway contributes 80% of the TAG that is incorporated into chylomicrons and is believed to inhibit the glycerol-3-phosphate pathway. Consistent with a role in dietary fat absorption, all three of the MGAT genes (MGAT1-3) are predominantly expressed in the gastrointestinal tract. Although high MGAT2 expression at mRNA levels was detected in the human liver, little MGAT activity was identified in this organ. MGAT expression and activity in the liver is inducible by diabetes, obesity, and lactation in rodents. MGAT expression at the mRNA level is also detectable in adipogenic tissues, although at much lower levels than in the small intestine, which indicates a possible role of the enzyme in *de novo* lipid synthesis. However, the MGAT2 protein is below detection limits in all nongastrointestinal tissues. MGAT2 expression and activity in the small intestine are

induced by a high-fat diet, which correlates well with the pattern of dietary fat absorption; this enzyme is therefore an attractive drug target for dietary fat absorption. Its primary site of expression is limited to the gastrointestinal tract, which would minimize any potential side effects of an MGAT inhibitor on other tissues. Furthermore, discharge of MAG and FA (in contrast to TAG) as an unabsorbed energy source in the feces might minimize the gastrointestinal side effects that are caused by orlistat. MGAT is a member of the acyltransferase superfamily of enzymes, which have recently been shown to be druggable through the successful development of small-molecule inhibitors of acyl-CoA:cholesterol acyltransferase and 1-acyl-glycerol-3-phosphate acyltransferase as potential treatments for hypercholesterolemia and cancer, respectively.

2.2.1.6 **Acylglycerol-3-Phosphate Acyltransferase.** 1-Acyl-glycerol-3-phosphate acyltransferase (AGPAT) catalyzes the acylation of lysophosphatidic acid at the *sn*-2 position to form phosphatidic acid, which is an important intermediate in the *de novo* biosynthesis of TAG and glycerophospholipids. Dephosphorylation of phosphatidic acid results in the formation of *sn*-1,2-DAG, which joins the MGAT pathway for TAG synthesis that is catalyzed by diacylglycerol acyltransferase (DGAT). *AGPAT2* is one of only two identified human genes that, when mutated, cause congenital generalized lipodystrophy, which is a rare autosomal recessive disorder that is also known as Berardinelli–Seip syndrome. Mutation of *AGPAT2* also causes voracious appetite and increased energy expenditure as a result of leptin deficiency. AGPAT activity is markedly induced during 3T3-L1 adipocyte differentiation, and overexpression of AGPAT1 in 3T3-L1 adipocytes results in increased TAG synthesis. One obvious concern for intervention in AGPAT2 activity as a treatment for obesity is the potential side effect of lipodystrophy, which is associated with diabetes and extreme insulin resistance. However, this might not occur with small-molecule inhibitors, which are unlikely to impose total inhibition of the enzyme activity *in vivo*. AGPAT2 has proved to be druggable, as small-molecule inhibitors of AGPAT2 that were developed as a potential treatment for cancer have been shown to prevent cellular growth and induce apoptosis of tumor cells.

2.2.1.7 **Diacylglycerol Acyltransferase.** The final step in TAG synthesis occurs by the action of diacylglycerol acyltransferase (DGAT) with DAG and fatty acyl-CoA as substrates (Fig. 2) that merges the MGAT and glycerol-3-phosphate pathways. In contrast to MGAT enzymes, which are predominantly expressed in tissues that are involved in dietary fat absorption, both DGAT1 and DGAT2 are widely expressed in a range of tissues. DGAT1 polymorphism is linked with certain types of human obesity. Levels of DGAT1 protein increase sharply following the differentiation of 3T3-L1 into mature adipocytes, and the overexpression of DGAT1 and DGAT2 in cell lines results in a marked increase in TAG synthesis in mature adipocytes and stably transfected McA-RH7777 cells. In contrast to the minor role of DGAT in dietary fat absorption, mice that are deficient in DGAT1 show a reduction in adiposity and resistance to diet-induced obesity, which is accompanied by an increase in energy expenditure and hyperlocomotive activity. The phenotype can be recapitulated by adipose-tissue transplantation from *Dgat1*-knockout mice to normal controls, which

points to the adipocyte as the principal regulatory site for DGAT1 function.

By contrast, targeted inactivation of DGAT2 resulted in lethal neonatal lipopenia and skin-barrier abnormalities. DGAT2 is crucial to TAG synthesis in various tissues, as evidenced by the near absence of TAG in liver and white adipose tissue, and a greater than 70% reduction in plasma levels of TAG, FA, and glucose in *Dgat2*-knockout mice. The lethal phenotype is not compensated for by DGAT1; this possibly is a result of the different roles of the proteins in regulating endogenous TAG synthesis, as evidenced by the different subcellular localizations of TAG catalyzed by these two enzymes. Furthermore, the content of linoleic acid (C18:2) in FA from liver and plasma is greatly reduced in *Dgat2*-null mice, which indicates an important role for DGAT2 in the selective retention of essential FA that are vital for the maintenance of normal skin function. Although the heterozygous *Dgat2*-knockout mice seem to be indistinguishable from wild-type controls, further analysis is needed to investigate whether they are more resistant to diet-induced obesity. The difference in phenotype between *Dgat1*- and *Dgat2*-knockout mice has confirmed the theory that two distinct DGAT enzymes are responsible for storage and mobilization, and that separate pools of TAG exist for storage and oxidation. Several natural products from various microorganisms have been reported to inhibit DGAT activity, although the specificity of these compounds has not yet been confirmed against cloned DGAT enzymes.

2.2.2 Lipid Droplets

2.2.2.1 Morphology and Lipid Composition. In different types of mammalian cells, intracellular lipids are stored in discrete LD. White adipocytes have only one or a few large TAG-rich LD (10–100 μm diameter) per cell. Brown adipocytes, which occur in some hibernating animals and the fetuses/neonates of many mammals, contain many smaller storage LD (2–10 μm diameter). Small LD (<1 μm) are also present in various other cells such as liver, muscle, kidney, intestine, and mammary gland, cultured lines such as 3T3-L1 fibroblasts, CHO cells, melanocytes, cultured pulmonary epithelial cells, various animal tissues and macrophages. Other cells, such as cultured HepG2 hepatoma cells and primary lung lipofibroblasts may contain LD up to 2 to 3 μm in diameter. Cholesterylester-rich LD are contained in steroidogenic cells of adrenal cortex, testes, and ovary.

First reports on the composition of LD in mammalian cells in the 1960s revealed a high content of TAG and a small amount of protein in LD from bovine heart. Subsequently, membrane-bound LD of rat liver, which consist mainly of neutral lipids, and a lipid-rich fraction of pig liver, which mainly contains TAG, were isolated and characterized. Further investigations revealed that the composition of cellular LD varies depending on the cell type. LD from bovine heart have similarities to the fat globule of bovine milk in structure and, to a certain extent, in chemical composition. Under electron microscopy, bovine heart LD appear as irregular bodies 0.5 to 2 μm in diameter. Occasionally, a few larger particles (5–8 μm in diameter) were detected as distinct structures without indication of fusion. They consist of a core of TAG surrounded by a phospholipid monolayer with some proteins attached to the surface.

Approximately 95% of total lipids of LD from bovine heart are constituted by

TAG. The amount of protein present in these LD is about 5% of total mass, and the amount of phospholipids varies from 3 to 7% of total lipid. The major phospholipids forming the monolayer of these LD are phosphatidylcholine (ca. 50%) and phosphatidylethanolamine (30–40%). The NEFA content is very low. The chemical composition of LD in beef heart with their high TAG content and the rather small amount of phospholipids resembles the composition of chylomicrons. A striking difference, however, is the lack of cholesterol and cholesterylesters in LD from beef heart, whereas these lipid species occur at approximately 1 to 2% of total lipid in chylomicrons. Furthermore, the protein content of LD is two to three times higher than that of chylomicrons. In contrast to bovine heart LD, those of stellate cells from the rat liver consist of retinylesters, TAG, free cholesterol and a small amount of phospholipids. A general characteristic of LD regardless of the cell type appears to be the content of approximately 5% phospholipids of the total mass.

However, electron microscopy of conventional resin-embedded ultrathin sections cannot visualize any membranous structure around the LD. In the ultrathin section of LD specimens fixed by aldehydes and then by osmium tetroxide there appears to be no phospholipid, and its periphery is usually seen as a thin intermittent line. In many diagrams, the LD surface has been depicted as a phospholipid monolayer with the hydrophilic head group facing the cytoplasm and the hydrophobic acyl chains extending into the LD content. It has also been assumed that LD form by accumulation of TAG between the two leaflets of the ER (endoplasmic reticulum) membrane because hydrophobic TAG molecules may exist stably in the aqueous cytoplasm only when covered with amphiphilic molecules with the hydrophilic moiety facing outward. However, experimental evidence to support the above assumptions is scarce. Only freeze-fracture electron microscopy showed that the fracture plane along the LD surface is occasionally continuous with the cytoplasmic leaflet of the ER membrane. Because it is difficult to retain lipid-rich structures by conventional morphological methods, cryoelectron microscopy is of advantage since it observes biological specimens at atomic resolution without fixation or staining. The microscopy showed that the LD surface is indeed a hemi-membrane, or a phospholipid monolayer.

A phospholipid monolayer in the surface is consistent with the current model that LD are formed by TAG deposition between the two leaflets of the ER membrane and may remain connected to it. Distribution of acyl-CoA:cholesterol acyltransferase-1, a major enzyme that synthesizes cholesterylester, in the entire ER seems to indicate that LD may bud anywhere along the membrane. However, Cap-LC/ESI mass spectrometry showed that FA moieties of phosphatidylcholine and lyso-phosphatidylcholine in LD are distinct from those in the rough ER. The results do not rule out the generation of the LD surface generated from the ER membrane but indicate that the former is a highly differentiated domain. Mature LD might be independent of the ER. Alternatively, the LD may be connected to the ER, but some molecular mechanism may demarcate the LD surface from the bulk ER membrane as postulated for other ER domains. Whatever is true, TAG synthesized in wide areas of the ER do not deposit indiscriminately but are concentrated to loci specialized to make

LD. ADRP or other LD-associated proteins may be involved (see below).

Cap-LC/ESI mass spectrometry has also shown that the LD surface differs from the detergent-insoluble glycolipid- and sphingolipid/cholesterol-rich microdomains (DIGs) in FA composition. Interestingly, DIGs contain a higher proportion of long saturated FA than rough ER, whereas LD do not. This indicates that not only interaction between sphingolipids and cholesterol but interaction between saturated phospholipids and cholesterol are also important in forming DIGs, as predicted from model membrane studies. The discovery of caveolin sequestration (see below) led to speculation that the LD surface might be similar to DIGs. However, the present results do not support this in terms of the FA composition. Furthermore, thin-layer chromatography showed that the relative amount of sphingomyelin and free cholesterol in comparison with phosphatidylcholine is far less in LD than in DIGs. In addition to the ER, the LD of mammalian cells seem to be closely associated with other organelles. A close association of LD of beef heart with mitochondria has been reported, suggesting a role of the LD as a source for FA utilized for energy production by mitochondrial ß-oxidation. Cytoskeletal intermediate filament proteins, particularly vimentin, were detected forming a network in the vicinity of LD. The interaction of intermediate filaments and LD was confirmed by transmission electron microscopy and field-emission scanning electron microscopy.

2.2.2.2 **Protein Composition.** Little is known about proteins on the surface of mammalian intracellular LD. Two groups of proteins whose association with LD in animal cells is well documented are the perilipins and adipocyte differentiation related protein (ADRP). The human ortholog of ADRP has been termed adipophilin. Other proteins at or near the surface of the LD of adipocytes include the "capsular" proteins and vimentin. Furthermore, certain enzymes, for example, eicosanoid-forming enzymes, associate with LD in various types of cells.

Perilipins and ADRP Perilipins belong to a family of polypeptides that are expressed in adipocytes and steroidogenic cells of adrenal cortex, testes, and ovary. Two related classes of full-length cDNAs have been isolated that encode perilipins A and B from rat adipocyte cDNA expression library. Perilipin A and B, 56 kDa (517 aa) and 46 kDa, respectively, share a common amino-terminal sequence of 406 aa. Similar to adipocytes, steroidogenic cells express perilipin A, but these cells also contain relatively high amounts of perilipin C, a protein of approximately 42 kDa that is not detectable in adipocytes. Perilipins are encoded by a single copy gene that gives rise to the three protein isoforms by alternative mRNA splicing with perilipin A being the most abundant. In murine adipocytes, a forth perilipin, named *perilipin D*, occurs at the mRNA level. The perilipin D polypeptide, however, has not yet been reported. Perilipins exhibit sequence similarity to a few other proteins, one of which is ADRP. The amino termini of perilipin and ADRP are highly homologous at a region of approximately 100 aa. ADRP was first identified during a search for genes being expressed in an early stage of differentiation of murine adipocytes in culture. The mRNA of ADRP is present in various tissues and cultured cell lines. Immunocytochemical examination revealed that ADRP localizes to LD in murine 3T3-L1 adipocytes, murine MA-10 Leydig cells,

CHO fibroblasts, and human HepG2 hepatoma cells. In MA-10 Leydig cells, ADRP and perilipins colocalize to LD. In 3T3-L1 adipocytes, the two proteins were found on small LD although at different stages of differentiation. ADRP is present on these LD in early differentiated adipocytes and is then replaced by perilipins, which can only be found on the surface of the LD at a later growth stage. It was suggested that ADRP is involved in the management of TAG stores in nonadipose tissues. Interestingly, recently, ADRP was found to colocalize with insulin and at the LD surface in human islet cells. Incubation with glucose or FA alone and a combination of glucose and FA increased ADRP expression by 2- and 10-fold, respectively. Oleate induced ADRP expression more potently than palmitate, which correlated with higher levels of neutral lipid accumulation by islets. Consistent with increased ADRP expression, the initial uptake rate of FA in islet cells was elevated in islet cells treated with FA *plus* glucose compared with control cells. In addition, ADRP expression is considerably increased in ob/ob mice islets and its upregulation is tightly correlated with TAG accumulation in islets, demonstrating an inverse relationship with impairment of ß-cell function. Thus, ADRP may play a role in regulating TAG metabolism also in ß-cells and, in particular, be involved in the loss of ß-cell function during the pathogenesis of NIDDM.

Perilipins remain tightly bound to LD through alkaline carbonate or urea wash conditions, and require detergents to remove them. Stable expression of perlipin A in 3T3-L1 fibroblasts and subsequent immunofluorescence microscopy and immunoblotting revealed that neither the amino- nor the carboxyl-terminus is required to target perilipin A to LD. Full-length perilipin A is associated with LD via hydrophobic interactions, as shown by the persistence of perilipins on LD after centrifugation through alkaline carbonate solution. Mutagenesis studies indicate that the signals responsible for tight anchoring of perilipin A to LD are, most probably, multiple, partially redundant targeting sequences of moderately hydrophobic amino acids located within the 25% of the central domain. These signals seem to be, at least in part, constituted by the three regions of moderate hydrophobic character and the five domains with characteristics of amphipathic ß-pleated sheets (see below). Cytochemical analysis by immunogold labeling suggests that the perilipins are located on or within the limiting surface of the LD.

TIP47 A cDNA of another protein, named TIP47 or pp17, which has similarity to ADRP has also been described. TIP47/pp17 is a cargo protein involved in the trafficking of the mannose-6-phosphate receptor between endosomes and the Golgi network. Perilipin A, ADRP, and TIP47/pp17 exhibit strong sequence homology at their amino-terminal regions. Starting from the amino terminus, perilipin A contains a sequence of 105 aa similar to adipophilin and TIP47 (aa 17–121), five 10-aa domains with amphipathic ß-pleated sheet character (between aa 11 and 182), three sequences of moderate hydrophobicity (aa 243–260, 320–342, 349–364), a highly acidic region (aa 291–318), and six consensus sites for PKA phosphorylation. Two such hydrophobic regions and three areas of amphipathic ß-pleated sheet structure were detected in ADRP. A protein named S3-12 is also highly homologous to ADRP in a 33-aa repeated sequence. The mRNAs of perilipins, ADRP, and S3-12 are expressed much higher in adipose than in other tissues.

It was suggested that lipids regulate transcription of perilipin mRNA, because the amount of perilipin mRNA and the corresponding protein increase in the presence of exogenous lipids. In analogy, formation of the mRNA of ADRP and protein expression in preadipocytes was also shown to be stimulated by FA in a time- and dose-dependent way. Stimulation could be only achieved by long-chain FA, whereas the degree of saturation had no influence. Treatment of such stimulated cells with actinomycin D inhibited the expression of ADRP, suggesting that FA stimulate the gene expression at the transcriptional level. Recently, the protein expression of perilipin and ADRP in human muscle was studied in greater detail. Perilipin was detectable in biopsies of vastus lateralis muscle (VLM) only at levels consistent with adipocyte infiltration. ADRP was highly expressed in VLM, irrespective of whether derived from obese nondiabetic or diabetic subjects. Weight loss of the latter led to increased whole-body insulin sensitivity and ADRP content in VLM. Treatment of the obese nondiabetic subjects with either troglitazone or metformin in combination with glibenclamide considerably improved the glycemic control, as reflected in lowering of HbA1c and plasma fasting TAG levels, which was well correlated to post-treatment increases in ADRP expression in VLM. These results demonstrate that ADRP is the predominant LD-associated protein in VLM and that its expression is upregulated under conditions of improved glucose tolerance. It is tempting to speculate that increased expression of ADRP may act to sequester FA in TAG of LD contained in skeletal muscle and thereby protects the muscle from the inhibitory (''lipotoxic,'' see above) effects of FA (or derivatives thereof) on insulin signaling and action. In consequence, this study provided additional evidence that TAG in LD represents the ''benign'' storage form of FA and accumulates during overfeeding of myocytes and β-cells with FA, which is compatible rather than interferes with insulin action and secretion, at least up to a certain extent of LD deposition.

CGI-58 Recent studies have expanded the list of LD-associated proteins to include several with known enzymatic functions in sterol synthesis and lipid metabolism, as well as numerous proteins with unknown functions. Recently, the identification of CGI-58 as a component of LD isolated from cultured 3T3-L1 adipocytes was reported. The name CGI-58 derives from a Comparative Gene Identification initiative that annotated 150 novel transcripts conserved between *Caenorhabditis elegans* and humans. CGI-58 is a member of the esterase, thioesterase, and lipase subfamily of α/β-hydrolase fold enzymes. The protein sequence includes a highly conserved active site that, in most α/β-hydrolases, includes a nucleophilic serine residue known to be important in catalysis. Interestingly, the CGI-58 sequence encodes an asparagine in place of this catalytic serine. Significantly, eight different mutations in CGI-58 were recently identified as causative factors in Chanarin–Dorfman Syndrome, a neutral lipid storage disorder (NLSD) characterized by ichthyosis, hepatic steatosis, and hepatomegaly, developmental defects, and the accumulation of TAG-containing LD in leukocytes, basal keratinocytes, hepatocytes, myocytes, and other cells. The eight mutations known to lead to NLSD include truncations, null alleles, and three distinct point mutations. This suggests that the lipid storage phenotype of NLSD most likely arises from loss

of function of CGI-58, leading to excessive storage of TAG in LD.

CGI-58 mRNA is highly expressed in adipose tissue and testes, tissues that also express perilipins, and at lower levels in liver, skin, kidney, and heart. Both endogenous CGI-58 and an ectopic CGI-58-GFP chimera show diffuse cytoplasmic localization in 3T3-L1 preadipocytes, but localize almost exclusively to the surfaces of LD in differentiated 3T3-L1 adipocytes. The localization of endogenous CGI-58 was investigated in 3T3-L1 cells stably expressing mutated forms of perilipin using microscopy. CGI-58 binds to LD coated with perilipin A or mutated forms of perilipin with an intact C-terminal sequence from amino acid 382 to 429, but not to LD coated with perilipin B or mutated perilipin A lacking this sequence. Immunoprecipitation studies confirmed these findings, but also showed coprecipitation of perilipin B and CGI-58. Remarkably, activation of PKA by the incubation of 3T3-L1 adipocytes with isoproterenol and isobutylmethylxanthine disperses CGI-58 from the surfaces of LD to a cytoplasmic distribution. This shift in subcellular localization can be reversed by the addition of propranolol to the culture medium. Thus, CGI-58 binds to perilipin A-coated LD in a manner that is dependent upon the metabolic status of the adipocyte and the activity of PKA.

The CGI-58 binding site was mapped to a sequence of 48 aa between positions 382 and 429 near the C terminus of perilipin A; this sequence spans the junction at aa405 where the sequences of perilipin isoforms A and B diverge. Consistent with these results, CGI-58 failed to localize to LD coated with perilipin B, suggesting that a portion of the unique C terminus of perilipin A is required for binding. The deletion of short sequences between aa382 and 429 failed to identify a smaller binding site, suggesting that either CGI-58 binds to multiple sequences spread throughout the region, or the intact 48aa sequence is required to fold into a conformation that exposes a smaller binding interface. The data obtained from coimmunoprecipitation experiments were mostly consistent with cell-based experiments; however, a notable difference was the coprecipitation of CGI-58 with perilipin B. Since perilipin B contains the sequence from aa382 to 405, the results suggest that the position or conformation of a partial docking site for CGI-58 may be altered by the unique 17-aa sequence of perilipin B (406–422) when perilipin is associated with LD, but exposed when cells are lysed, and the proteins are solubilized away from LD.

Interestingly, CGI-58 translocates off of adipocyte LD and disperses throughout the cytosol when lipolysis is stimulated, suggesting that CGI-58 may play a role in lipid metabolism either when it is bound to LD under basal conditions, or during its movement away from LD following activation of PKA. While the phosphorylation of PKA sites 1 and 2 of perilipin A is unnecessary for the displacement of CGI-58 from LD, mutations altering PKA sites 4, 5, and 6 attenuate dispersion. Since CGI-58 contains at least one PKA consensus site, it is likely that the phosphorylation of serines in PKA consensus sites of both perilipin A and CGI-58 are required for the full dispersion of CGI-58 into the cytoplasm. Thus, the perilipin-dependent binding of CGI-58 to LD is modulated by the activation of PKA. It was hypothesized that perilipins form a scaffold at the surfaces of adipocyte LD that serves as an organizing center for lipid metabolic enzymes and transporters that is altered by metabolic signals. While proteins including CGI-58

bind to perilipin-coated LD under basal conditions, other proteins, such as HSL, are recruited to LD when perilipins have been phosphorylated by PKA. Under the latter conditions, protein components of basal LD may be displaced to accommodate association of the lipolytic machinery.

The role of CGI-58 in lipid metabolism remains to be elucidated. CGI-58 is a member of the α/β-hydrolase fold subfamily of lipases, esterases, and thioesterases. Mutations of human CGI-58 are responsible for excessive TAG storage in many types of cells. Studies of TAG metabolism in fibroblasts from NLSD patients show that CGI-58 is required for normal rates of TAG turnover. However, TAG lipase activity in NLSD cell lysates was comparable to that of control cells. Additionally, the rate of TAG clearance increased in NLSD cells incubated with triacsin C, an inhibitor of acyl-CoA synthetases, although not quite to the higher rates observed in control cells. These observations suggest that lipolysis occurs in NLSD cells, but that TAG clearance is blunted by rapid reesterification. Thus, CGI-58 is a factor that is required to assist TAG clearance, but may not be a lipase.

CGI-58 shows a subcellular localization pattern that is paradoxically opposite to that of HSL. CGI-58 disperses off LD when HSL translocates onto the LD, and the rate of lipolysis becomes elevated. Interestingly, basal rates of lipolysis are relatively normal in adipocytes isolated from HSL null mice, suggesting that as yet uncharacterized mechanisms catalyze TAG turnover under basal conditions. Thus, CGI-58 may interact with perilipin-coated LD to play a role in basal TAG hydrolysis. Alternatively, CGI-58 may function to transport products of lipolysis away from LD following the activation of PKA and the dispersion of CGI-58 into the cytosol. Since adipose TAG metabolism appears not to be severely compromised in NLSD patients, it is likely that adipocytes have redundant mechanisms to avoid the detrimental effects of mutations in CGI-58. Thus, despite much lower levels of expression, CGI-58 may play a more critical role in TAG catabolism in nonadipose tissues.

Other reports have described CGI-58 as a component of LD isolated from oleic acid-treated Chinese hamster ovary cells and cultured human epithelial cells, cells that lack perilipins but instead, coat LD with adipophilin. Furthermore, CGI-58 has been identified as a perilipin-binding protein by yeast two-hybrid assay using the N-terminal sequence of perilipin from aa 1 to 250 as bait. Direct binding between recombinant CGI-58 and recombinant truncated perilipin A that included aa 1–416 was demonstrated. In contrast, CGI-58 binding to LD in intact cells requires sequences in the C terminus of perilipin A, but not the N terminus. Interaction between CGI-58 and adipophilin by yeast two-hybrid and GST pull-down assays using recombinant proteins was also reported. In contrast, significant localization of CGI-58 to lipid droplets in 3T3-L1 preadipocytes was not found. Furthermore, the PAT family of proteins (perilipin, ADRP, TIP47) is highly conserved within N-terminal sequences, but structurally divergent in the C termini, where a binding site for CGI-58 on perilipin A has been identified.

Adipophilin Adipophilin (ADPH) is a prominent LD-associated protein found in many mammalian cell types, including hepatocytes, adipocytes, muscle cells, and mammary epithelial cells, either during development or in the mature functioning cell. ADPH has proposed functions in LD formation, FA transport, and milk lipid

secretion. Although it can be detected at the plasma membrane in mammary epithelial cells as a result of milk globule secretion, the weight of evidence suggests that ADPH specifically targets LD. However, little is known about the structural and biochemical properties involved in its association with LD or how it accomplishes its biological functions. Genetic analysis has demonstrated that ADPH is related to the family of LD-associated proteins that include perilipin (see above) and the mannose-6-phosphate receptor targeting protein TIP47, through a conserved region in their amino-termini, called the "PAT" domain (perilipin, ADPH, TIP47). The lack of significant sequence conservation outside this region indicated that it is an important determinant in the association of these proteins with LD. However, a recent study revealed that neither the PAT domain nor the carboxy-terminal half of ADPH is required for efficient targeting to or biogenesis of LD, but that, nevertheless, specific regions located elsewhere within ADPH are responsible for LD targeting. Previous work indicates that secretion of LD from mammary epithelial cells during lactation involves the formation of a tripartite complex between ADPH, the mammary-specific transmembrane protein, butyrophilin, and xanthine oxidoreductase. As formation of such a complex probably involves regions of ADPH distinct from those involved in LD targeting, the PAT domain or the carboxy-terminal half of ADPH may represent candidate regions for interaction with butyrophilin and/or xanthine oxidoreductase to trigger LD secretion.

Caveolins In primary and cultured adipose cells, caveolin-1 and caveolin-2 constitute a framework of caveolae that represent bulb-shaped plasma membrane invaginations (50–100 nm in size) enriched in cholesterol, (glyco)sphingolipids (sphingomyelin) as well as glycolipids (gangliosides, GPI lipids) and harboring a special set of (dually acylated) signaling proteins, glycosyl-phosphatidylinositol-anchored (GPI) proteins and the caveolae structural and marker proteins, caveolin 1–3. Based purely on ultrastructural comparisons and tissue expression profiles, the adipocyte seems to have the highest concentrations of caveolae and the highest levels of caveolin-1 and -2, that is, more than any other cell type. Indeed, electron micrographs of adipocytes dating back to 1963 show that caveolae account for up to 30% of the surface area of the adipocyte plasma membrane. Furthermore, in 3T3-L1 adipocytes, a widely used model system for studying adipogenesis, the number of caveolae increases 9-fold, and caveolin-1 and -2 expression increases 20-fold during differentiation from fibroblasts to the adipocyte state.

Caveolins are supposed to anchor to the plasma, Golgi, and ER membranes by the central hydrophobic domain with both amino- and carboxy-termini exposed to the cytoplasm. Caveolae fulfil important functions in protein and lipid/cholesterol transport (endocytosis, potocytosis, transcytosis, polarized transport) as well as in signal transduction. Caveolin-2, especially its ß-isoform, has been shown being targeted to the surface of LD, by immunofluorescence and immunoelectron microscopy, and by subcellular fractionation. Brefeldin A-treatment induced further accumulation of caveolin-2 along with caveolin-1 in LD. Analysis of mouse caveolin-2 deletion mutants revealed that the central hydrophobic domain (aa 87–119) and the amino-terminal (aa 70–86) and carboxy-terminal (aa 120–150) hydrophilic domains are all

necessary for the localization in LD. The amino- and carboxy-terminal domains appeared to be related to membrane binding to and exit from ER, respectively. This implies that caveolin-2, which is synthesized on cytosolic ribosomes and then incorporated into the ER membrane in posttranslational fashion, is transported to LD as a membrane protein. Thus, caveolin-2 seems to differ from the hitherto reported LD proteins, such as the perilipins, in that it exists as an integral membrane protein in the ER and plasma membrane. As anticaveolin-2 antibody, recognizing the N-terminal segment, decorated only the P face of LD in a freeze-fracture replica, caveolin-2 may take the same orientation on the LD surface as in the membrane. This is consistent with the hypothesis that the LD surface is a half-membrane, or a phospholipid monolayer, with amphiphilic lipid molecules with their hydrophilic and hydrophobic portions oriented toward the cytoplasm and the LD content, respectively. No labeling for caveolin-2 was seen in the LD surface, but was found in small clusters both in cryosections and in freeze-fracture replicas. The labeling for adipophilin shows the same pattern at the LD surface. These results suggest that the whole LD surface may not be homogeneous and that some segregation may occur in the half membrane. Interestingly, a recent study revealed unesterified cholesterol in LD. Furthermore, LD of eosinophils contain proteins typical for DIGs related to intracellular signaling, such as MAPK (mitogen-activated protein kinases) and Lyn. Together, these results suggest, intriguingly, that LD might be a novel membrane domain where caveolins may functionally regulate signaling proteins.

The recruitment of caveolin-1 to LD in BFA-treated cells may also be due to the hydrophobic domain shared by all caveolins. However, caveolin-1 does not appear to have the same degree of affinity for LD as caveolin-2. First, in cells expressing caveolin-1 alone, its redistribution to LD after BFA-treatment occurred slowly and only in some cells. Second, without the BFA-treatment, caveolin-1 was not found in LD even when overexpressed. The LD localization signal of caveolin-1 is likely to be less effective than that of caveolin-2. Endogenous caveolin-2 was not detected in LD under normal culture conditions. This may be because caveolin-2α, expressed predominantly in most cells, is less efficient than caveolin-2ß in LD targeting. But the accumulation of caveolin-2 in LD is transient, so that its distribution in LD can only be detected when the protein is expressed abundantly or when its trafficking is perturbed. The apparent morphological continuity of ER and LD and the persistence of LD labeling in BFA-treated cells indicate that caveolin-2 goes directly from ER to LD. Whether LD exists independently from the ER-Golgi pathway, or LD is an intermediate compartment between ER and Golgi, is not known. The results of BFA experiments may simply indicate that an excess of caveolin-2 in ER overflowed to LD. However, BFA might also block other pathways (e.g. from LD to the Golgi) and have caused LD retention. The two pathways are not mutually exclusive and need further studies for definition. Trafficking of caveolins to LD is important because it may be linked to that of TAG molecules. A testable question is whether the TAG composition of LD is changed by the presence of caveolins and/or by BFA-treatment. Also, if and how is LD related to the reported trafficking routes of caveolin-1? Interestingly, overexpression of a mutant caveolin-3 in BHK cells leads to its association with the

surface of large LD containing TAG and endogenous caveolins 1 to 3. Fluorescence, electron, and video microscopy observations are consistent with formation of the LD by maturation of subdomains of the ER. The mutant caveolin caused the intracellular accumulation of free cholesterol in late endosomes, a decrease in surface cholesterol, and a decrease in cholesterol efflux and synthesis. Incubation of the cells with oleic acid induced a significant accumulation of wild-type caveolins in the enlarged LD. Taken together, the available data hint at a role of caveolins in LD biogenesis.

2.2.2.3 Biogenesis of LD.

Role of Caveolins LD biogenesis is incompletely understood. The observations presented above suggest that although the caveolins are not normally associated with internal lipid compartments and particles, such as LD, they can certainly traffic to these locations under certain conditions. As such the caveolins represent the first known integral membrane protein components of LD. This characteristic may also explain the initially surprising finding that caveolin-1 can behave as a soluble protein. Immunogold labeling first detected soluble caveolin-1 in the lumen of the ER after cells were exposed to cholesterol oxidase for conversion of cholesterol at the outer leaflet of the plasma membrane into cholesterone. However, it remained a mystery how caveolin-1 as an integral membrane protein firmly embedded in the cytoplasmic leaflet of the (ER) membrane can move to the luminal leaflet for subsequent release into the lumen. Subsequently, a small pool of soluble caveolin-1 was found in the cytosol of fibroblasts in a complex with chaperones. Finally, the analysis of caveolin-1 distribution in different cell types revealed targeting of caveolin-1 to the cytosol (skeletal muscle cells and keratinocytes), and to the lumen of secretory vesicles (serous cells of the pancreas, fundic stomach, and salivary gland). Both the secreted and the cytosolic caveolin-1 exhibited characteristics for incorporation in lipoprotein-like particles, which may explain their apparent solubility. Several groups provided additional evidence that the cytosol of many cells contains a pool of apparently "soluble" and "lipoprotein-like" caveolin-1. This caveolin-1 is associated with cholesterol and behaves like a protein that is embedded in a particle with the size and buoyant density of HDL. Interestingly, caveolin-1 is secreted by exocrine secretory cells in HDL-like particles that contain apoA-1, raising the possibility that caveolin-rich lipid particles in the cytoplasm are involved in the assembly of secreted lipoproteins.

On the basis of the structure of caveolins, with their long stretch of hydrophobic amino acids acting as membrane-anchoring domain, it is unlikely that they can switch the "membraneous" into a true "soluble" conformation compatible with the aqueous milieu of the cytosol. Rather, the involvement of caveolins in the biogenesis of various lipid particles by budding from either the cytoplasmic or the luminal leaflets of the ER membrane may explain their cell type–specific localization either in the cytosol (as LD or as precursor particles for milk lipid granules destined for budding from the plasma membrane) or in the lumen of the ER and secretory vesicles (as nascent lipoproteins destined for movement along and release by the secretory pathway), respectively. This unifying concept for the biogenesis of apparently "soluble" and "membraneous" caveolin-containing structures requires a "flip-flop" mechanism for translocation of caveolins

from the cytoplasmic ER membrane leaflet (with amino- and carboxy termini exposed to the cytoplasm) to the luminal ER membrane leaflet (both termini exposed to the lumen), which may be induced during cholesterol depletion (e.g. by cholesterol oxidase). In conclusion, caveolin-1 can be integrated in (one leaflet of) various cellular membranes (plasma membrane, Golgi, ER, caveolae) or in (the phospholipid monolayer of) various cytosolic and secreted lipoprotein-like particles (LD, milk lipid granules, lipoproteins) and in each case may regulate the accessibility of these structures toward degradation by (phospho)lipases in a way similar to that now commonly accepted for perilipins.

Taken together, the prevailing model for LD biogenesis rationalizes that caveolins (but not most other membrane proteins) might be able to enter the LD (Fig. 3). Since LD are surrounded by a phospholipid monolayer and are thought to be derived from the ER, where TAG and cholesterylesters are synthesized (see above), it is tempting to speculate that TAG accumulate in the hydrophobic core of the bilayer, forming a bulge that eventually buds from the ER membrane to form free LD (Fig. 3a). Accumulation of TAG in the bilayer core would initially force opposite leaflets of the bilayer apart, increasing bilayer thickness. Such a thickened bilayer could not accommodate transmembrane proteins that have hydrophilic domains on both sides of the membrane and these proteins would be excluded from the forming LD (Fig. 3b). Caveolins, by contrast, lack luminal hydrophilic domains, and could diffuse freely between the ER membrane and the monolayer surrounding the nascent LD. Thus, caveolins would not need to dissociate from the ER membrane and expose their hydrophobic domains to the cytosol to enter the LD (Fig. 4, steps 1 and 2). This model suggests that caveolins can enter LD only while they are forming and are still in contact with the ER. Consequently, accumulation of caveolins in the ER would trigger their targeting to LD and concomitantly LD biogenesis. Alternatively, transport of caveolins from the plasma membrane and/or Golgi apparatus to the ER via a retrograde pathway might somehow lead to LD targeting. If caveolins normally pass through LD during biosynthetic transport or intracellular cycling, they must exit the LD rapidly as they are not normally (i.e. absence of BFA, wild-type caveolin) detected there.

Although it is unclear how ER accumulation of caveolins might lead to an apparent concentration of the proteins in LD, several possible explanations can be imagined. First, caveolins might have a high affinity for LD, either through specific binding to a protein or (phospho)lipid of the LD, or through preference of caveolin for the physical state of the LD phospholipid monolayer or TAG core leading to its spontaneous partitioning. Efficient packaging of caveolins into transport vesicles might normally prevent entry into the LD. Accumulation in the ER might saturate some component of the transport machinery, possibly a specific cargo receptor, giving the protein time to access the LD. Alternatively, accumulation of caveolins in the ER might change the properties of the ER membrane. The altered properties might increase the affinity of caveolins for LD, either by changing the physical properties of the membrane or by inducing a conformational change in caveolins.

Role of Perilipins In LD, accumulation of perilipins and TAG occurs roughly in parallel. When FA synthesis was inhibited, the concentration of TAG decreased by 90% and correspondingly the amount

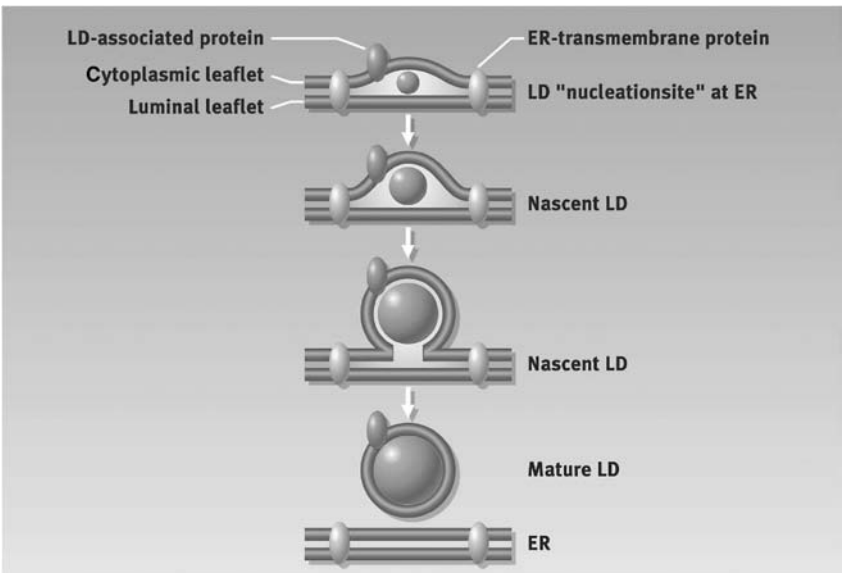

(a)

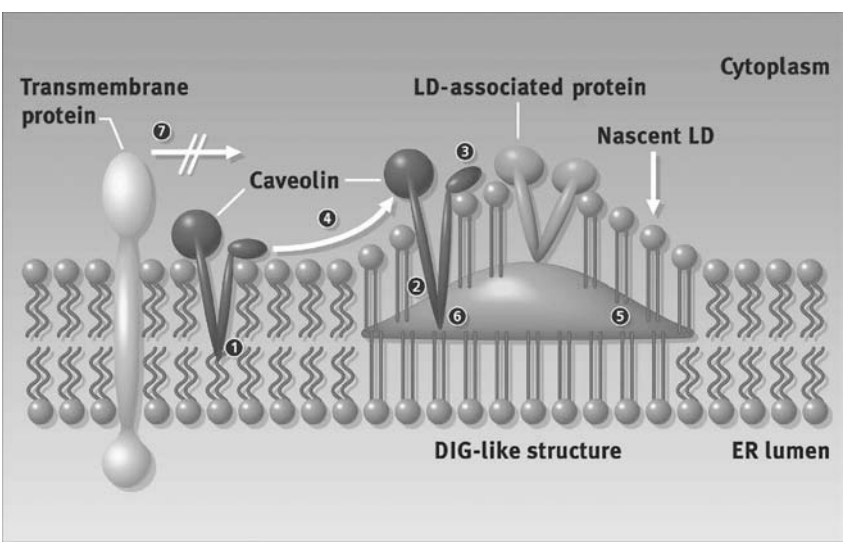

(b)

Fig. 3 ''Budding'' model for the biogenesis of intracellular LD. (a) LD formation is initiated by accumulation of TAG between the leaflets of the ER membrane. During budding of the forming LD, typical transmembrane proteins with hydrophilic domains protruding into the ER lumen and cytoplasm are segregated from membrane proteins, anchored in the outer leaflet, only, such as caveolin. After detachment of the matured LD from the ER membrane, the TAG core of the LD is surrounded by a phospholipid monolayer with embedded caveolin having both its amino- and carboxy-terminus exposed to the cytoplasm. (b) The topology and segregation of transmembrane proteins and caveolin are shown in greater detail. See text for detail.

of perilipin dropped. The availability of only a small amount of TAG may cause formation of small LD with insufficient surface space for the association of perilipin. Decreased synthesis of FA resulting in liberation of FA from TAG may lead to subsequent detaching of perilipin from the surface of the shrinking LD and subsequent degradation of the protein. Undifferentiated 3T3-L1 fibroblasts transfected with constructs that encode perilipin A accumulate numerous small LD that are associated with perilipin A. These results indicate that appropriate targeting of perilipins is not restricted to adipocytes or steroidogenic cells, and further suggests that perilipins may serve as nucleation sites for TAG deposition (Fig. 2). As nearly all cells can synthesize TAG, the formation of LD in the transfected fibroblasts may reflect trapping of TAG in perilipin-bound structures. This is compatible with the observed consequences of introducing perilipin A into other cells that do not normally express this protein. It remains to be demonstrated, however, whether perilipins are directed toward LD, or *vice versa* LD are transported to perilipin aggregates, or association of both components occurs in parallel (Fig. 4).

Growth Adipocyte differentiation in WAT involves the coalescence of small nascent LD to form the one or more large LD found in mature cells. By contrast, most other cell types contain numerous smaller LD (see above). In adipocytes, cytoskeletal intermediate filament proteins, particularly vimentin, form cages around small LD and thereby prevent the latter from fusing (Fig. 4, step 3). Indeed, a vimentin network is seen near LD in newly

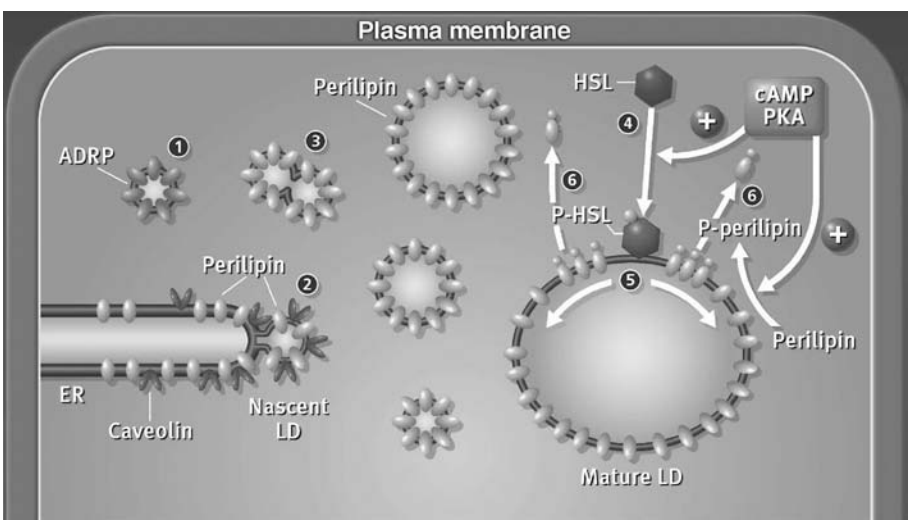

Fig. 4 Hypothetical model for integration of the major mechanisms of TAG storage and mobilization: (1) conversion of fibroblasts/preadipocytes into adipocytes, (2) biogenesis of forming LD at the ER membrane, (3) fusion of smaller LD to mature LD, (4) translocation of HSL from the cytosol to the surface of LD, (5) disassembly and reorganization at the LD surface of perilipin and (6) its translocation from LD to the cytosol. Both processes may contribute to relieve the barrier that prevents HSL from gaining access to the LD. PKA-dependent phosphorylation is depicted as filled circles. See text for details.

differentiating adipocytes. As the cells mature, a decrease in the number of vimentin fibers coincides with, and might play a role in, LD coalescence. In fact, in 3T3-L1 fibroblasts, expression of a dominant-negative vimentin reduces the formation of LD. However, the involvement of vimentin in LD biogenesis came under debate when knockout mice lacking vimentin were shown to exhibit normal biosynthesis of adipocyte LD, even though no compensatory formation of other intermediate-filament proteins could be demonstrated. Nevertheless, coalescence of adipocyte LD must be highly regulated – given the 1000-fold difference in the volume of LD in white and brown adipose tissue – but the mechanism of adipocyte LD maturation remains to be elucidated.

Cell-Free System Very recently, a microsome-based, cell-free system containing substrates for TAG synthesis and a phospholipase D-activating protein was introduced that assembles newly formed TAG into spherical LD. These LD were recovered in the low-density fraction by gradient ultracentrifugation and were similar in size and appearance to those isolated from rat adipocytes and 3T3-L1 fibroblasts. Caveolin-1 and -2, vimentin, ADRP, and the 78-kDa glucose–regulated protein, GRP-78, were associated with LD from this cell-free system. The caveolin was soluble in 1% TX-100, as was the caveolin on LD from 3T3-L1 cells. These data suggest that caveolin is released from the microsomes along with LD, indicating that caveolin may help sort newly formed TAG from the ER into a cytosolic form in a budding process of LD from the ER membrane. The ADRP on the LD was also derived from the microsome-associated pool, suggesting that it can be targeted to the ER membrane. ADRP can be acylated and these acyl moieties may participate in the interaction between ADRP and the membrane, since covalently linked FA are involved in the membrane targeting of other proteins (e.g. ADP ribosylation factor-1). Thus, ADRP binds to the microsomal membrane and participates in the budding of the small LD. Perilipins were not expressed in the 3T3-L1 fibroblasts used to prepare the microsomes and consequently were not present in the partially purified cytosolic phospholipase D activator preparation. Therefore, perilipins were not present on the LD assembled in the cell-free system. Consequently, small LD can be assembled in the absence of perilipins. GRP-78 is a chaperon protein that contains a KDEL sequence and is therefore restricted to the ER/Golgi region of the secretory pathway. GRP-78 appeared to be present on the cytosolic side of the surface layer of LD, suggesting that the surface layer is derived from the membrane of the ER. This agrees with the model proposed for the assembly of LD (see above), according to which the TAG will, during the biosynthesis, "oil-out" between the leaflets of the ER bilayers (Fig. 3). The LD formed by complete budding from the ER would be covered by a phospholipid monolayer and proteins originally bound to both the luminal and cytosolic surface of the ER membrane.

The LD from the cell-free system, like those from 3T3-L1 cells, contained TAG, DAG, phosphatidylcholine, phosphatidylethanolamine, and phosphatidylserine. The assembly of these LD-like structures depended on the rate of TAG biosynthesis and required an activator present in the high-speed supernatant from homogenized rat adipocytes. The activator-induced phospholipase D activity, and its effect on the release of the LD from the microsomes, was inhibited by butan-1-ol or 2,3-diphosphoglycerate.

The activator could be replaced by a constitutively active phospholipase D or its product, phosphatidic acid. In conclusion, LD seem to be formed from regions of the microsomal membrane that contain caveolin, vimentin, and ADRP in a process that depends on DGAT activity and is driven by the activity of a phospholipase D and the formation of phosphatidic acid. Phosphatidic acid acts as an intracellular messenger or as cone-shaped lipid that alters the curvature of the membrane. Its formation seems to be important for the fission of transport vesicles. Recent data suggest that it plays a role in (regulation of) the budding step within the biogenesis of LD, as has previously been proposed for the assembly of VLDL in the ER lumen.

Targeting Mechanism Three studies have addressed the question of how LD-associated proteins accumulate at the monolayer surface of LD during their biogenesis. Amino-terminal truncation mutants of caveolin-1, -2, -3, among them Cav-3DGV, are targeted to LD (and the early Golgi). Overexpressed caveolin-2 was found to be associated with LD (and the Golgi). Caveolin-1 with an ER retention signal fused to its carboxy terminus, such as Cav-1KKSL, as well as an internal-deletion mutant of caveolin-1, are highly concentrated in LD (and the ER and Golgi, respectively). Furthermore, blockade of the anterograde transport along the secretory pathway by Brefeldin A leads to increased expression of endogenous caveolin at the surface of LD. Consequently, (transient) accumulation of caveolins at the ER as a result of overexpression or impaired release into the secretory pathway during a folding defect caused by the mutation or by Brefeldin A-treatment may be sufficient for their efficient partitioning from the cytoplasmic membrane leaflet of the ER to the phospholipid monolayer of LD. According to this view, the association of caveolins with LD under physiological conditions, rather than the unusual experimental conditions of inefficient transport, has still to be demonstrated.

The LD-associated proteins described so far can be classified on the basis of their biosynthesis and topology: ADRP, perilipins, and HSL are translated on free ribosomes and recruited to the LD from the cytosol as a type of peripheral membrane protein, the latter two shuttling between the LD and the cytosol in reversible fashion in response to the nutritional state of the adipocyte (see below). Caveolins are cotranslationally inserted into the ER in a signal-recognition-particle-dependent fashion, with their central hydrophobic domain penetrating into the ER cytoplasmic membrane leaflet and the two flanking hydrophilic domains extending into the cytoplasm (Fig. 3b, step 1). Maintaining this topology the caveolins would diffuse laterally from the bulk ER to the surface phospholipid monolayer of nascent LD still attached to the ER without exposing their hydrophobic domains to the cytosol, whereas transmembrane proteins would be excluded from the LD (Fig. 3b, step 4). However, the apparent concentration of proteins at LD requires additional specific targeting information and deciphering mechanisms, albeit a consensus sequence for LD targeting has not been identified. The proteins could interact directly with phospholipids located at the LD surface (Fig. 3b, step 2) or with TAG deposited in the LD core (step 6) or with other LD proteins (step 3). Since LD probably arise (also) *de novo* (in contrast to the biogenesis of membranes) in the absence of preformed structures that could function as nucleation or docking sites for the incorporation of newly synthesized

LD proteins, some LD proteins have to come into direct contact with phospholipids/TAG of the LD and discriminate them from those of the ER. Caveolins represent candidates for LD proteins triggering the initial biogenetic steps of LD assembly on the basis of their high affinity for DIGs, which are characterized by tight packing and a high degree of order of their (saturated) lipid acyl chains (see above). The apparent affinity of caveolins for LD may well rely on their affinity for DIGs, which is supported by the finding that the LD-targeted Cav-3DGV mutant resists solubilization by cold nonionic detergent, as is true for typical DIG-associated proteins. Consequently, a physical state of the LD surface monolayer distinct from the ER cytoplasmic leaflet and more "DIG-like" may attract typical LD proteins, such as caveolins (Fig. 3b, step 7). However, putative targeting information encoded by the "DIG milieu" of nascent LD is not compatible with a recent more detailed analysis of the phospholipids contained in LD and DIGs, which did not reveal close similarity with regard to length and saturation of the constituent FA (see above). Alternatively, contact of the TAG molecules in the LD core with the phospholipids of the LD surface (Fig. 3b, step 5) might trigger a tight packing of the latter, creating a special lipidic environment that causes the selective recruitment of certain phospholipids and proteins to nascent LD.

3 Mechanism and Regulation of TAG Mobilization

3.1 HSL

HSL activity was first identified in 1964 as an epinephrine-sensitive lipolytic activity in adipose tissue. It was named according to its property of being activated by hormones such as catecholamines, ACTH, and glucagon. In 1981, the isolation of a relatively pure and biologically active HSL from rat adipocytes was reported. First cDNA clones were isolated for both rat and human HSL in 1988 and the structure of the gene for human HSL was elucidated in 1993. The rat and human cDNAs predict 768 and 775 amino acids, respectively, which is in good agreement with the molecular mass of 84 000. Despite 82% identity between the amino acid sequence of rat and human HSL and their similar molecular masses, human HSL migrates exhibit slightly lower electrophoretic mobility during SDS-PAGE compared to rat HSL, that is, 88 versus 84 kDa. HSL has a uniquely broad substrate specificity, hydrolyzing with varying efficacy tri-, di- and monoacylglycerols, as well as cholesteryl, steroid fatty acyl, retinyl and *p*-nitrophenyl esters. However, HSL lacks detectable phospholipase activity. HSL prefers cleavage of DAG (10- and 5-fold vs. TAG and MAG, respectively) and the 1- or 3-ester bond of its acylglycerol substrate (4-fold vs. 2-ester bond) esterified to polyunsaturated and shorter versus saturated and longer FA. Phosphorylation under standard *in vitro* assay conditions leads to moderate increases in activity against TAG and cholesterylesters, but not against DAG and MAG. It remains to be clarified whether phosphorylation of HSL provokes differential effects on its activity also *in vivo*. Rat adipocyte HSL was recognized to exist as a dimer in solution according to gel chromatography and sucrose gradient centrifugation. This finding was then confirmed using the human recombinant HSL and sedimentation equilibrium analysis. Kinetic studies finally revealed that the HSL functional dimer is characterized

by greater maximal activity with no difference in substrate affinity compared to its monomeric subunits.

Reflecting on its unique catalytic and regulatory properties, HSL exhibits only minimal sequence similarity with enzymes of the lipase gene family. On the basis of inhibition of HSL activity by diisopropylphosphofluoridate and the covalent linkage of this inhibitor to the HSL protein, HSL has been categorized as a serine esterase. The Gly-X-Ser-X-Gly motif contains the catalytic serine in most lipases, and HSL contains this consensus pentapeptide (Gly-Asp-Ser-Asp-Gly). Site-directed mutagenesis has confirmed that the serine of this pentapeptide is within the catalytic site of HSL. Human HSL also exhibits a region of similarity with lipase 2 of *Moraxella* TA144, an antarctic bacterium that hydrolyzes lipids at low temperature. Strikingly, HSL also manages to cleave TAG with higher efficacy at reduced temperatures compared to other mammalian lipases. The higher order domain structure of HSL could not be derived from the amino acid sequence. Two different functional domains – the catalytic domain and the regulatory domain – are encoded by different exons of the human gene suggesting the mosaic nature of the HSL protein. However, the overall lack of homology with other lipases precludes assignment of specific functions to different regions of the protein. Other lipases with little primary sequence homology display highly similar structural features based on crystal structure, but the crystal structure of HSL is urgently needed to gain further information. Detailed biochemical studies have provided evidence for the existence of a distinct lipid-binding domain that is separate from the catalytic region. Mild proteolysis of HSL with trypsin almost completely destroyed its activity against trioleoylglycerol with only marginal reductions against water-soluble *p*-nitrophenylbutyrate, indicating that HSL contains a lipid-binding domain. As trypsin treatment generates a stable 17.6-kDa fragment, it is reasonable to assume that HSL is composed of separately folded protease-resistant domains joined by short protease-sensitive linker regions. Studies on interfacial activation, that is, enhancement of lipolytic activity by the presence of a lipid-water interface, provided additional evidence for the existence of a separate lipid-binding domain. Intact HSL is activated by the presence of phospholipid vesicles or emulsified trioleoylglycerol, but these lipids have no effect on activity of the 17.6-kDa fragment on the water-soluble substrate, strongly suggesting that proteolysis has separated the lipid-binding domain from the catalytic domain. A more stable peptide comprising aa 1–323 can be generated by prolonged incubation of HSL with endoproteinase Lys-C suggesting that this fragment represents a separate structural domain. Circular dichroism and fluorescence spectroscopy analyzes during denaturation with guanidine hydrochloride and heat revealed two major steps of unfolding, which is in agreement with the existence of two distinct structural domains.

The amino-terminal 323-aa domain of possibly globular structure is encoded by exons 1–4 and has no primary or secondary structural similarity with known proteins, but interacts with the adipocyte lipid-binding protein, ALBP. The carboxy-terminal portion of HSL is similar to acetylcholinesterase, bile salt-stimulated lipase and several fungal lipases, and is composed of α/ß-hydrolase folds that constitute the catalytic site. The secondary

structure elements that constitute this scaffold are expected to be within regions formed by amino acids 332–535 and 688–766 (rat). Despite lack of significant primary sequence homology, the organization of the secondary structure predicted for the carboxy-terminal 450-aa of HSL is similar to the secondary structure of two fungal lipases from *Geotrichum candidum* and *Candida rugosa*. Secondary structure prediction in this region allowed the identification of the central elements of these α/ß-hydrolase folds according to which a three-dimensional model for this domain was proposed. Elucidation of the crystal structure for the Brefeldin A esterase from *Bacillus subtilis*, a protein with significant homology to HSL, confirmed the usefulness of this approach. On the basis of molecular modeling, it was proposed and later confirmed by site-directed mutagenesis, that Ser423, Asp703, and His733 (rat) constitute the catalytic triad for HSL and are found within this carboxy-terminal region. This portion is interrupted by the 150-aa regulatory module, which according to secondary structure prediction, does not contain α-helices and ß-sheets but harbors the known phosphorylation sites for PKA, AMPK, and ERKs and therefore has been assigned a regulatory function.

3.2
cAMP

TAG mobilization by lipolysis as a strictly regulated process has been known since the early 1960s, when it was established that fast-acting hormones such as ACTH and epinephrine increased lipolysis, and that insulin counteracted this activation. It was soon recognized that cAMP was involved in the regulation of the catecholamine-sensitive lipolytic activity in adipose tissue. Catecholamines can either stimulate lipolysis via three subtypes of ß-AR, which are positively coupled to cAMP-synthesizing AC (adenylate cyclase) by G_s proteins, or inhibit lipolysis via α_2-AR, negatively coupled to the AC by G_i proteins (Fig. 2). Concerning the functional significance of adrenoceptor-mediated increases in cAMP levels, two points should be emphasized. First, lipolytic agents generally increase the levels of cAMP far above the concentration required for maximal activation of PKA. Second, considerable species and tissue specificity exists with regard to the distribution of adrenoceptor subtypes. In humans, for instance, the interplay between α_2- and ß-AR is important in modulating cAMP levels in adipocytes. It is now generally accepted that lipolytic response of fat cells is controlled mainly by the sympathetic nervous system and by plasma insulin levels and depends on the balance between stimulatory and inhibitory pathways. cAMP activates PKA and this kinase then phosphorylates the two main targets involved in the control of lipolysis in the adipocyte, HSL and the perilipins (Fig. 2). Phosphorylation of these proteins dramatically increases lipolysis (see below). Insulin is the physiologically most important antilipolytic hormone. The ability of insulin to antagonize hormone-induced lipolysis can largely be explained by its ability to lower cAMP levels and thereby PKA activity, leading to a decrease in the phosphorylation state of HSL and perilipin.

3.3
Phosphorylation of HSL

Phosphorylation of partially purified HSL by PKA, leading to a moderate activation of its activity *in vitro*, was described in the early 1970s and later confirmed.

Subsequent phosphopeptide mapping and phosphoamino acid analysis suggested that HSL was phosphorylated on a single serine residue, named the regulatory site. Partial amino acid sequencing of phosphopeptides generated from bovine HSL, together with determination of the primary structure of rat HSL, allowed identification of Ser563 as the regulatory site. However, more recent data have dramatically challenged this view on the short-term regulation of HSL. The finding that mutation of Ser563 did not abolish PKA-induced activation of HSL led to the identification of two novel PKA sites, Ser659 and Ser660, that seem to be responsible in intact primary adipocytes for the activation of HSL in response to isoproterenol stimulation. The role of phosphorylation of Ser563 remains elusive (Fig. 5).

Apart from three PKA phosphorylation sites described above, HSL is phosphorylated *in vivo* in hormonally quiescent cells at a site named the *basal site*, corresponding to Ser565 in rat HSL, that is, two residues carboxy-terminal to Ser563. Glycogen synthase kinase-4, Ca^{2+}/calmodulin-dependent kinase II, and AMP-activated protein kinase (AMPK) phosphorylate Ser565 *in vitro* without any direct effect on enzyme activity (Fig. 5). AMPK has been proposed to be the physiologically relevant kinase, based on its involvement in other aspects of lipid metabolism and on its proposed role as fuel gauge. Furthermore, because phosphorylation by AMPK prevented subsequent phosphorylation of Ser563, and *vice versa*, it was proposed that phosphorylation of Ser565 exerts an antilipolytic role. This proposal is supported by experiments showing that preincubation with AICAR an activator of AMPK – causes a moderate reduction of the lipolytic response to catecholamines in primary adipocytes. The data indicating mutually exclusive

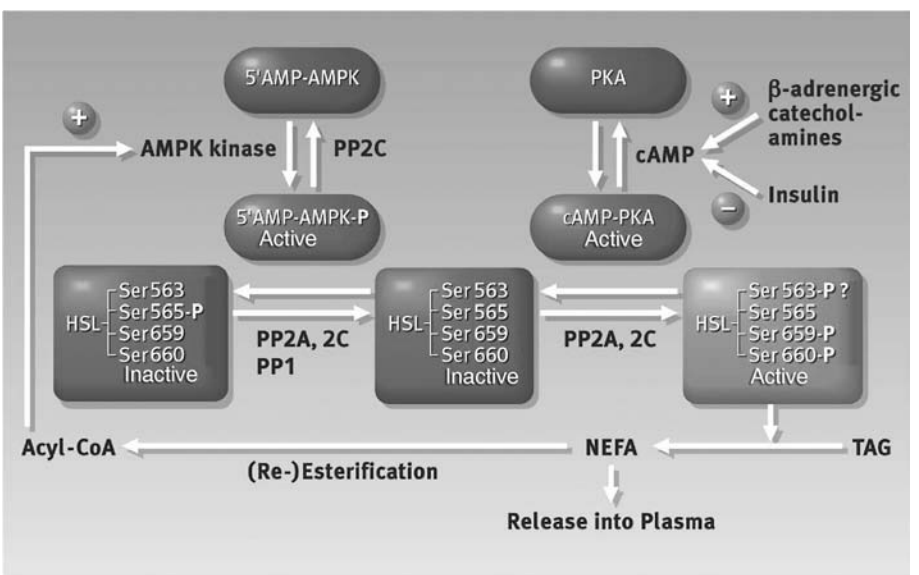

Fig. 5 Working model of the interplay between basal and regulatory phosphorylation sites at HSL, including the kinases and phosphatases involved. See text for details.

phosphorylation at the basal and regulatory sites imply that activation of HSL by PKA has to be preceded by dephosphorylation of the basal site. Whether this dephosphorylation is an important regulatory step or merely a result of a high constitutive protein phosphatase activity in the cell is not clear. Moreover, the recent finding that Ser563 is not essential for HSL activation raises some questions regarding this antilipolytic role of Ser565 phosphorylation. Also, the moderate antilipolytic effect of AICAR should be interpreted with caution because this substance has many cellular effects and has been questioned as a specific AMPK activator.

In fact, recent studies have elucidated completely novel aspects of AMPK physiology as it may be directly involved in the ß-adrenergic regulation of lipolysis. Much experimental evidence has promoted the idea that AMPK acts as an intracellular energy sensor, stimulated by the increased intracellular AMP/ATP ratio when cells are stressed by conditions such as hypoxia/ischemia in the heart and excessive contraction in skeletal muscle. Activated AMPK accelerates ATP-producing pathways, such as FA and glucose oxidation, while reducing ATP consumption, ultimately leading to the preservation or restoration of adequate high-energy phosphates. Whereas the importance of AMPK to control lipid metabolism in liver and muscle is well established, its role in regulating lipolysis in adipose tissue has remained controversial. Previous observations led to the hypothesis that AMPK antagonizes lipolysis in adipocytes (see above), presumably to prevent futile cycling and depletion of ATP during unrestrained simultaneous lipolysis and reesterification. However, subsequent mutational analysis of the PKA phosphorylation sites of HSL raised serious doubts on the negative effect of AMPK phosphorylation on PKA-induced HSL activation, which were then strengthened by the finding that isoproterenol stimulated AMPK phosphorylation and activity in isolated rat adipocytes, in apparent contradiction to an antilipolytic role for AMPK. Confirming and extending the latter study, it was demonstrated recently, that treatment of murine cultured 3T3-L1 adipocytes with isoproterenol or forskolin promoted the phosphorylation of AMPK at a critical activating Thr172 residue in a concentration- and time-dependent fashion. This correlated well with stimulation of AMPK activity as measured in the immune complex. Analogs of cAMP mimicked the effect of isoproterenol and forskolin on AMPK phosphorylation. Treatment of the adipocytes with insulin reduced both basal and forskolin-induced AMPK phosphorylation via a pathway dependent on PI-3K. Overexpression of a dominant-inhibitory mutant of AMPK blocked isoproterenol-induced lipolysis by about 50%. These data indicate that in adipocytes a novel pathway operates through which cAMP can lead to the activation of AMPK, and that this pathway is required for maximal stimulation of lipolysis. However, overexpression of a constitutively active AMPK version had a minimal effect on both basal and isoproterenol-induced lipolysis only, which may indicate that AMPK is required but not sufficient for lipolysis induction (in the absence of elevated cAMP). Moreover, it remains unclear from this study how agents that increase cAMP or the nucleotide itself activate AMPK in intact cells. Thr172, the phosphorylation site on AMPK responsible for most of the increase in AMPK activity, is not a consensus PKA phosphorylation site, suggesting that PKA regulates AMPK activity through

an indirect mechanism, such as phosphorylating and activating an unknown upstream AMPK kinase. Alternatively, the effect of cAMP could be independent of PKA. The other open question is how AMPK regulates lipolysis. As already mentioned, phosphorylation of HSL at Ser563, Ser659, and Ser660 by PKA alone does not account for the maximal activity of HSL as *in vitro* PKA phosphorylation causes a 1.5- to 2.2-fold increase in HSL lipase activity only, while isoproterenol stimulates FA release from adipocytes more than 50-fold. AMPK phosphorylates Ser565, which is located in the regulatory domain of HSL, as are the PKA phosphorylation sites, *in vitro*. Phosphorylation of Ser565 by AMPK might, speculatively, be involved in regulation of the translocation of HSL to the LD (see below). This would explain the recent finding that mutation of this residue abrogated the ability of HSL to translocate to LD. In addition, or alternatively, several other proteins, participation of which in lipolysis has been documented, such as perilipin (see below), may represent direct or indirect targets of AMPK. Taken together, the available experimental evidence hints at a positive modulatory role of AMPK in ß-adrenergic stimulation of lipolysis, during which it undergoes phosphorylation and activation and then possibly phosphorylates HSL at Ser565, leading to its translocation to LD. Physiologically, this putative mechanism nicely fits the commonly accepted role of AMPK as intracellular energy sensor since it guarantees that ATP depletion in adipocytes and in other tissues coordinated via the ß-adrenergic response leads to enhanced lipolysis, resulting in the release of FA that can be oxidized by target tissues to compensate for the increased demand for high energy phosphates.

3.4 Dephosphorylation of HSL

Two major studies devoted to clarifying the action of (protein phosphatases) PP on HSL have demonstrated a several-fold higher activity of PP2A and PP2C than PP1 toward the PKA site (Ser563), even though the source of both HSL and PP differed in each study. The same was true for the basal site, that is, Ser565. Because of their higher abundance in adipose tissue, PP2A and PP1 are the main PP acting on HSL *in vivo*. A role for PP1 in selective dephosphorylation of the basal site has been suggested. This, however, was not supported by another study, which showed a higher preference of PP1 for the PKA sites than for the basal site. The dephosphorylation of the novel PKA sites, Ser659 and Ser660, as well as the possible existence of specific HSL phosphatases, remains to be investigated (Fig. 5).

3.5 Intrinsic HSL Activity

Historically, the major consequence of HSL phosphorylation was thought to be an increase in the catalytic activity of the lipase toward TAG, which thereby triggers lipolysis, but more recent observations indicate that activation of lipolysis involves a more complex and elegant series of reactions. Phosphorylation of HSL *in vitro* leads to a moderate (twofold) increase in activity against emulsified TAG substrates, whereas ß-adrenergic stimulation of adipocytes leads to a >50-fold increase in the hydrolysis of TAG housed within LD. A reasonable explanation for this discrepancy is that (1) cytosolic HSL in unstimulated adipocytes is already highly active but has limited access to its substrate contained within intracellular LD

and; (2) the major consequence of lipolytic stimulation is the migration of HSL toward the LD and a change in the LD surface that provides access of the lipase to its TAG substrate in the LD core. That HSL is present in an active state in unstimulated adipocytes is supported by comparisons of lipolysis in intact control and stimulated cells with the lipolytic activities of homogenates or sonicates derived from these cells. Whereas lipolysis in intact stimulated cells is far greater than in intact control cells, there is little between the measurable lipolytic activities in homogenates or sonicates of these control and stimulated cells. Importantly, high HSL activity is found in sonicates or homogenates of control cells that exhibit only minimal lipolysis before disruption of the cells. The conditions of cell disruption in these studies are unlikely to foster phosphorylation of HSL because both PKA and ATP are diluted, and chelators in the homogenization medium reduce the effective concentration of Mg^{2+}, which is required for PKA activity. Hence, it appears that the non-PKA-phosphorylated HSL in homogenates or sonicates of control adipocytes can hydrolyze TAG at a rate nearly equal to that exhibited by PKA-phosphorylated HSL when acting on exogenous emulsified TAG substrates. The great increase in the lipolytic activity in sonicates of unstimulated adipocytes may result from the large increase in surface area of TAG arising from dispersion of the LD. It has been speculated that the increase was due to destruction of the integrity of the LD surface. Implicit in both explanations is the idea that, in the intact adipocyte, HSL is active but has restricted access to the TAG substrate. As reviewed below, this is compatible with data showing that lipolytic stimulation is accompanied by a migration of HSL to the LD *plus* a dramatic change in the structure of the LD surface.

3.6 Translocation of HSL

3.6.1 Molecular Mechanism

cAMP-stimulated lipolysis in cultured adipocytes is accompanied by redistribution of HSL from the aqueous supernatant of homogenates of unstimulated cells to the particulate fraction and the floating fat cake of homogenates of stimulated cells. These findings were later confirmed by the direct immunological demonstration of HSL movement from the cytosol to the LD, leading to the hypothesis that a major consequence of PKA activation is the translocation of the phosphorylated HSL from the cytosol to the LD. Subsequently, this hypothesis was confirmed by examining the location of HSL in 3T3-L1 adipocytes with the use of immunofluorescence microscopy. In unstimulated cells, the enzyme is disturbed diffusely in the cytosol, but migrates to the LD surface upon stimulation of lipolysis by the ß-AR agonist, isoproterenol. This translocation is reversible since upon addition of the ß-AR antagonist, propranolol, to stimulated cells, HSL resumes a diffuse distribution. Furthermore, HSL leaks from adipocytes into the medium when digitonin is used to permeabilize unstimulated, but not stimulated cells. These data suggest that HSL is freely soluble in the cytosol of unstimulated adipocytes and that its binding to the LD surface in stimulated cells is reasonably avid. Furthermore, HSL remains bound to LD following fractionation of stimulated cells. The temporal kinetics of HSL translocation to and away from the LD parallels the temporal kinetics of the onset and termination of lipolysis. Because the effects of isoproterenol are mimicked

by nonhydrolyzable cAMP analogs, the translocation reaction is downstream of PKA activation. There appears to be no cytoskeletal involvement in the translocation reaction, since translocation is unimpeded by cytoskeletal poisons such as nocodazole, cytochalasin B, and the combination of these two agents. Thus, an important consequence of lipolytic activation of adipocytes is the translocation of HSL from the cytosol to its substrate, that is, the TAG in the core of LD (Fig. 4, step 4).

Meanwhile, HSL translocation seems to be more complex and may be accompanied by and/or trigger a conformational change of HSL, leading to higher affinity for TAG and/or phospholipids of the LD. *In vitro*, even nonphosphorylated HSL binds readily to phospholipid vesicles and mixed phospholipid/DAG monolayers. However, in intact cells, the interaction with the LD only takes place upon lipolytic stimulation and phosphorylation of HSL. Remarkably, a mutant variant in which the newly described phosphorylation sites (Ser659, Ser660) have been replaced by alanines was unable to translocate. Of interest is whether the different sites for PKA phosphorylation on HSL have different roles in promoting cellular lipolysis, that is, in triggering HSL translocation to LD versus increasing intrinsic HSL activity toward TAG in the LD core. HSL translocation proceeds normally when either of these two sites is mutated singly, but is abolished on simultaneous mutation of both sites. By analogy with other lipases, HSL might exist in two conformational states, an active form – corresponding to the "open" form exposing a large hydrophobic lid structure as observed for other lipases – and an inactive "closed" form. In intact cells, phosphorylation of HSL would be required to trigger the transition from the "closed" to the "open" form, exposing the hydrophobic lid structure for interaction with TAG. *In vitro*, HSL may be permanently stabilized in the "open" form by the nonionic detergent required to keep the enzyme soluble and active in its purified form, thus making phosphorylation unnecessary. This mechanism of stabilization of the "open" form by a micelle of tensioactive molecules has been elegantly demonstrated for pancreatic lipase.

Recently, the involvement of multiple mechanisms for hormonal regulation of lipolysis by HSL has convincingly been demonstrated with transgenic mice expressing human HSL in WAT. HSL mRNA levels were increased in WAT threefold versus endogenous HSL. The lipolytic activity against TAG was elevated in WAT of transgenic compared with wild-type mice, which may explain the observed more pronounced reduction of body weight and fat mass in transgenic compared with wild-type mice after a four-day calorie restriction. However, and surprisingly, the cAMP-inducible lipolytic response was lower in adipocytes prepared from transgenic mice. Thus, these data suggest that mechanisms other than phosphorylation, translocation, and intrinsic activity of HSL may become rate limiting for lipolytic degradation of TAG in LD of intact adipocytes, at least under conditions of HSL overexpression, some of which and the components putatively involved are discussed below.

3.6.2 Involvement of Perilipins

Initially it was proposed that the function of perilipin could be to anchor HSL to its TAG substrate as a "docking protein" at the surface of LD when phosphorylated. However, attempts to coimmunoprecipitate (PKA-phosphorylated) HSL and perilipin offer no evidence of any direct interaction between these two proteins, nor

does a yeast two-hybrid screen. A salient feature of perilipins A and B is their phosphorylation by PKA. The A isoform has six predicted sites for PKA phosphorylation, three of which are in perilipin B. Most, if not all, of these sites in perilipin A appear to be phosphorylated upon lipolytic activation of adipocytes. The PKA-mediated phosphorylations occur on serine residues. By contrast, in unstimulated cells, one or more threonine residues, as yet not identified, are phosphorylated. Because perilipin A is polyphosphorylated by PKA during lipolytic activation of adipocytes, a role for perilipin in lipolysis has been hypothesized. As cytosolic HSL seems to be in an activated state in the absence of PKA phosphorylation, one must ask why the lipase does not bind to and hydrolyze TAG in unstimulated adipocytes. There are two possible explanations. First, despite the evidence that in intact cells non-phosphorylated HSL behaves like a freely soluble cytosolic protein, this HSL form might be constrained from reaching the LD due to its association with another protein (see below). Secondly, perilipins could serve to block HSL action in unstimulated cells. During the last decade, the common hypothesis has been that non-PKA-phosphorylated perilipins suppress lipolysis by denying HSL access to TAG in the core of the LD (Fig. 4, step 5). Several lines of evidence support this:

Mutational analysis of the three consensus PKA sites in perilipin A that are common to both perilipin A and B (Ser81, Ser222, Ser276) was performed to analyze the role of PKA-mediated hyperphosphorylation of perilipin in regulating lipolysis. When the mutant protein (delta3) was expressed in cells overexpressing acyl-CoA synthase-1 (ACS1) and fatty acid-binding protein-1 (FATP1), like wild-type perilipin A, the mutant one targeted itself to the surface of LD, increased TAG accumulation, and decreased basal lipolysis. However, unlike perilipin A, perilipin-delta3 blocked PKA-stimulated lipolysis in ACS1/FATP1 cells overexpressing HSL. Thus, PKA-stimulated phosphorylation of perilipin A is necessary to abrogate perilipin's A inhibitory action on lipolysis. Because perilipin A and B share the same PKA phosphorylation sites that were mutated, it can be speculated that perilipin B may also modulate PKA-stimulated lipolysis. If perilipins A and B assume similar extended conformations at the LD surface, the greater protection against hydrolysis of the longer splice variant, perilipin A, may be attributed to its greater length and thus greater coverage of the LD surface, that is, it is the extended unique carboxy-terminal tail (12 kDa) of perilipin A that is important in protecting against lipolysis in the unstimulated state. Since perilipins are major substrates of PKA, which polyphosphorylates perilipins on lipolytic stimulation, that is, in parallel to lipid hydrolysis, it follows that phosphorylation of PKA sites in the amino-terminal region induces a conformational change in the carboxy-terminal region that exposes portions of the LD surface to HSL action. Furthermore, expression of native and mutated forms of perilipin A and B in CHO fibroblasts revealed that perilipin A inhibits TAG hydrolysis by 87% when PKA is quiescent but that activation of PKA and phosphorylation of perilipin A provokes a sevenfold lipolytic activation. No HSL was detected by immunoblotting in these cells nor could any PKA-stimulated lipolytic activity in CHO cell homogenates be measured. Thus, the lipase responsible for TAG hydrolysis in CHO cells appears unresponsive to PKA, and PKA

regulation of lipolysis in CHO cells expressing perilipin A is concluded to result solely from the PKA phosphorylation of perilipin, which is supported by the finding that mutation of selected PKA sites eliminated this regulated response. Unlike perilipins, ADRP has no consensus sequences for PKA phosphorylation and thus is unlikely to be phosphorylated in lipolytically stimulated cells.

In agreement with the above findings, perilipin A acts cooperatively with HSL in lipolysis when both proteins are introduced into CHO cells. Thus, the ability of perilipin to regulate lipolysis of TAG in the core of LD appears to be manifested with any lipase that has access to the LD surface. The enhancement of lipolytic activity by perilipin A and the endogenous lipase of CHO cells is clearly time-dependent, requiring at least 30 min after stimulation to become manifest. In contrast, when both HSL and perilipin are introduced into CHO cells, the lipolytic activation occurs with no detectable lag. Phosphorylation of HSL is assumed to foster this immediate reaction, in contrast to the endogenous lipase, which is apparently not a PKA substrate.

Additional support for the barrier hypothesis for perilipin is based on analyzes of the stimulation of adipose cell lipolysis with the cytokine, tumor necrosis factor-α (TNF-α). Several cytokines, including TNF-α, increase lipolysis in adipocytes. Unlike the fast-acting lipolytic hormones, which activate PKA and increase lipolysis within minutes, TNF-α requires at least 6 to 8 h to initiate a lipolytic response. Moreover, TNF-α does not increase PKA activity. However, the cytokine does terminate expression of the perilipin gene, and within 24 h of exposure of the cells to TNF-α the perilipin content of cells is reduced by up to approximately 50%. This reduction in perilipin accounts for the increased lipolysis since adenovirus-mediated expression of perilipins A or B prevents TNF-α-stimulated lipolysis. In contrast to the endogenous perilipin, that produced by the adenovirus construct is not attenuated by the cytokine. Thus, the data indicate that perilipins not phosphorylated by PKA inhibit lipolysis and that phosphorylation of perilipins abrogates this barrier (Fig. 4, step 5).

Most interestingly, increased PKA activity in 3T3-L1 adipocytes is correlated with dramatic and reversible changes at the LD surface. Immunofluorescence staining for perilipin decreased at the periphery of large LD, suggesting that PKA activation lowers the density of the perilipin coat at the LD surface, either by an increase of the total LD surface area leading to "dilution" of perilipin or, more likely, loss of perilipin from the LD due to its translocation into the cytosol. The apparent reciprocal translocation of HSL (Fig. 4, step 4) and perilipin (Fig. 4, step 6) between LD and cytosol seems to depend on the "age" of the adipocytes. In young male rats (160–200 g), the translocation of HSL in response to isoproterenol closely paralleled the stimulation of lipolysis, but there was no translocation of perilipin from LD to the cytosol. By contrast, upon lipolytic stimulation and perilipin phosphorylation in adipocytes from older rats, there was no translocation of HSL but a significant translocation of perilipin away from the LD. Furthermore, the maximum rate of lipolysis induced by isoproterenol was markedly lower in cells from older rats than in those from younger rats. A similar redistribution of perilipin has also been reported for 3T3-L1 adipocytes. This "reverse" perilipin translocation is unlikely to involve a conformational change that allows perilipin to become freely soluble

in the cytosol as it is extremely hydrophobic. Instead, phosphorylation of perilipin and/or some other mechanism may trigger an alteration in the LD surface, rendering LD more susceptible to disruption by fractionation procedures, resulting in the release of apparently "free" perilipin in the cytosolic fraction. In conclusion, the results with older rats emphasize that lipolytic stimulation is not due to HSL translocation alone. The current speculation is that this diminution of the perilipin coating may represent one mechanism whereby perilipin contributes to lipolysis in adipocytes. However, it appears that the translocation of perilipin is not essential for the stimulation of lipolysis in all cells, or may not be the only or predominant mechanism as it is not observed in adipocytes from young male rats in which translocation of HSL was apparent upon lipolytic stimulation.

3.6.3 Involvement of ADRP

As already mentioned, perilipin can shield lipid esters in LD from lipase activity. When phosphorylated by PKA upon catecholamine stimulation, perilipin not only downregulates its barrier function, but also promotes translocation of lipases and enables efficient hydrolysis of TAG. On the other hand, ADRP is not likely to function as a shield against lipases, but it binds to NBD-FA and NBD-cholesterol, stimulates uptake of long-chain FA, and increases the cellular content of TAG. These properties suggest the importance of ADRP in the intracellular lipid trafficking, but the whole picture of its physiological function has not been delineated. The function of ADRP was examined by searching for proteins that interact with ADRP. By the yeast two-hybrid screening, ADP-ribosylation factor 1 (ARF1), a GTP-binding protein, was obtained as a candidate, and confirmed that it binds to ADRP both *in vivo* and *in vitro*. Interestingly, the GDP-bound ARF1 showed a higher affinity to ADRP than its GTP-bound form. Consistent with the property, when cells were treated with BFA, a reagent that inhibits ARF guanine nucleotide exchange factors, ADRP was dissociated from LD. The result indicates that an ARF1-dependent mechanism could affect association of ADRP to LD. It also showed that LD is a target of BFA action.

In addition, ARF1 is known to interact with a number of effector proteins, including phospholipase D, phosphatidylinositol 4-phosphate 5-kinase, and clathrin adaptor protein. These effectors interact with ARF1-GTP. Preferential binding to ARF1-GDP was also reported previously for p23, a transmembrane protein, which mediates recruitment of ARF1-GDP to the Golgi membrane. ARF-guanine nucleotide exchange factors also bind to ARF1-GDP. But ADRP does not have apparent sequence similarity to these proteins, and appears to be dissociated from LD upon binding to ARF1-GDP. It is not clear why dissociation of ADRP does not occur in the normal untreated cell, in which ARF1-GDP should exist in the cytosol, but several possibilities can be speculated. First, even though ARF1-GDP continuously dissociates ADRP from LD at a certain rate, it may be at equilibrium with ADRP that is newly synthesized and recruited to LD. Second, although a molecule corresponding to Rab-guanine nucleotide dissociation inhibitor has not been reported for ARF, it might exist, and ARF1 bound with the putative ARF1-guanine nucleotide dissociation inhibitor may not be able to interact with ADRP. Third, most ARF1 may be readily converted to ARF1-GTP by ARF1-guanine nucleotide exchange factors in the normal state, and thus dissociation of ADRP from LD may occur only when

ARF1-GDP increases by some stimulation. A number of ARF-GTPase activating proteins reported to exist in mammalian cells may be involved.

How does binding of ARF1-GDP cause dissociation of ADRP from LD? Previous studies showed that two independent segments, located in the amino and carboxyl termini of ADRP, can target the molecule to LD, whereas the central portion interposed between them is dispensable. It is now clear, that the central segment, 174–282, is involved in binding to ARF1. The relatively hydrophilic profile of the central segment suggests that it does not adhere to LD through hydrophobic interactions and may be able to interact with cytosolic molecules. In this context, it is noteworthy that the corresponding portion of TIP47 constitutes a unique α/β-domain and the first helix of the four-helix bundle, and makes a hydrophobic cleft in between where a lipid molecule may bind. Sequence alignment suggests that the central domain of ADRP takes a similar structure. Binding of ARF1 to critical domains may influence ADRP functionally and/or structurally, and eventually cause its dissociation from LD.

ADRP and perilipin are thought to be translated on free ribosomes, and they resemble each other in that several independent segments are involved in binding to LD. These properties may suggest that ADRP and perilipin adhere to LD by a similar mechanism. However, they are divergent in two important aspects. First, despite the sequence similarity in the amino-terminal PAT-1 domain, binding to LD is mediated by relatively dissimilar domains. Second, phosphorylation/dephosphorylation of perilipin regulates its binding to LD, whereas ADRP is not likely to be phosphorylated *in vivo*. Instead, FA may be involved in binding of ADRP to LD. A recent study showed apparent degradation of ADRP after the BFA-induced dissociation as well as the lack of promoting effect of ARF1-GTP on new LD formation. These findings suggest that ADRP dissociates from LD in an irreversible manner, and that the mechanism of LD binding may be different between ADRP and perilipin. BFA has been used as a reagent to inhibit vesicular transports. In addition, it causes redistribution of caveolins to LD (see above). In untreated cells, caveolins are present in the plasma membrane, endosome, and the Golgi, but after BFA treatment, they become concentrated in LD. This phenomenon has been thought to occur by translational diffusion of caveolins from the ER membrane to the LD surface, that is, due to inhibition of the ER-to-Golgi transport, an excessive amount of caveolins accumulate in the ER membrane and overflow to the LD surface. But in view of the effect of BFA upon ADRP and LD, the mechanism causing the redistribution of caveolins to LD may need to be reconsidered. One possible model is that ADRP normally coats the LD surface and blocks translational influx of ER membrane molecules by diffusion, but that dissociation of ADRP caused by BFA enables influx of ER membrane molecules into LD. As a modified version of this model, it is possible that dissociation of ADRP from LD may not be completed by binding to ARF1-GDP, but may occur only when the incoming ER proteins physically dislocate ADRP. In this case, the effect of BFA on ADRP is both direct and indirect. The latter model could explain the length of time required for the BFA effect on ADRP to become apparent.

It is noteworthy that LD of some cell types contains abundant free cholesterol. Free cholesterol in the LD surface may be important in sequestering a selected set

of proteins including caveolins. Despite the fact that inhibition of the ER-to-Golgi transport by BFA could cause accumulation of caveolins in the ER membrane and the subsequent diffusion to LD, the prominent concentration of caveolins in LD, and their retention even after BFA wash out cannot be explained simply by the overflow model. Dissociation of ADRP, a cholesterol-binding protein, and the resultant availability of free cholesterol may be important for the retention of caveolins in LD. Besides the direct action on ADRP, other ARF1-related pathways may work to cause the observed effect on LD. For example, phospholipase D was shown to be involved in new LD formation *in vitro* (see above). If turnover of LD is rapid in living cells and continuous formation is required to maintain constant LD volume, suppression of phospholipase D by BFA may contribute to reduce the total LD volume in a cell.

3.7 Feedback Inhibition of HSL

HSL is inhibited in a noncompetitive manner by oleoyl-CoA, oleic acid, and 2-monopalmitoylglycerol, with 50% inhibition at 0.1, 0.5, and 500 μM, respectively. Feedback inhibition of HSL by oleoyl-CoA and oleic acid may therefore prevent accumulation of free FA and cholesterol in the cell, whereas 2-monoacylglycerol may act as a feedback inhibitor if the capacity of 2-monoacylglycerol lipase is exceeded. The inhibition of HSL by oleoyl-CoA and oleic acid at concentrations well below their critical micellar concentrations (reported for fatty acyl-CoAs, 30–60 μM; for FA 0.8–1 mM) rules out any detergent effects, implying that inhibition is via binding to a specific site on the enzyme protein [Fig. 6]. FA inhibit AC and therefore cAMP accumulation in the adipocyte. Furthermore, PKA can phosphorylate and thereby stimulate phosphodiesterase, PDE3, directly or possibly indirectly by phosphorylating and activating the postulated PDE3 kinase (PDE3K), which finally leads to cAMP degradation. These mechanisms would reduce HSL activation by phosphorylation. The four inhibitory feedback loops may therefore act together to prevent further accumulation of FA released from lipolysis and therefore possible cell damage via detergent properties inherent in high concentrations of FA (Fig. 6).

The observed inhibition of HSL by 2-monopalmitoylglycerol may also be of significance as a feedback control of lipolysis. As a result of preferential cleavage of 1(3)-ester bonds by HSL, the hydrolysis of TAG and subsequently of DAG yields 2-monoacylglycerol. This is a poor substrate for HSL and therefore a distinct 2-monoacylglycerol lipase is required for the final step of lipolysis to yield free NEFA and glycerol. If the rate of production of 2-monoacylglycerol by HSL exceeds its rate of hydrolysis by the 2-monoacylglycerol lipase, the resultant accumulation of this lipid may prevent its further generation. However, the physiological conditions under which this might occur are unknown, as 2-monoacylglycerol lipase is reported to be present in adipocytes in sufficient quantity and is not rate limiting, even during hormonally stimulated lipolysis.

The two key enzymes in FA synthesis and cholesterol synthesis, ACC1, and HMG-CoA reductase, are also inhibited by oleic acid and oleoyl-CoA in a noncompetitive manner and over similar concentration ranges as for HSL. All three enzymes are phosphorylated by the AMPK. Phosphorylation of ACC1 and HMG-CoA reductase results in direct inhibition of

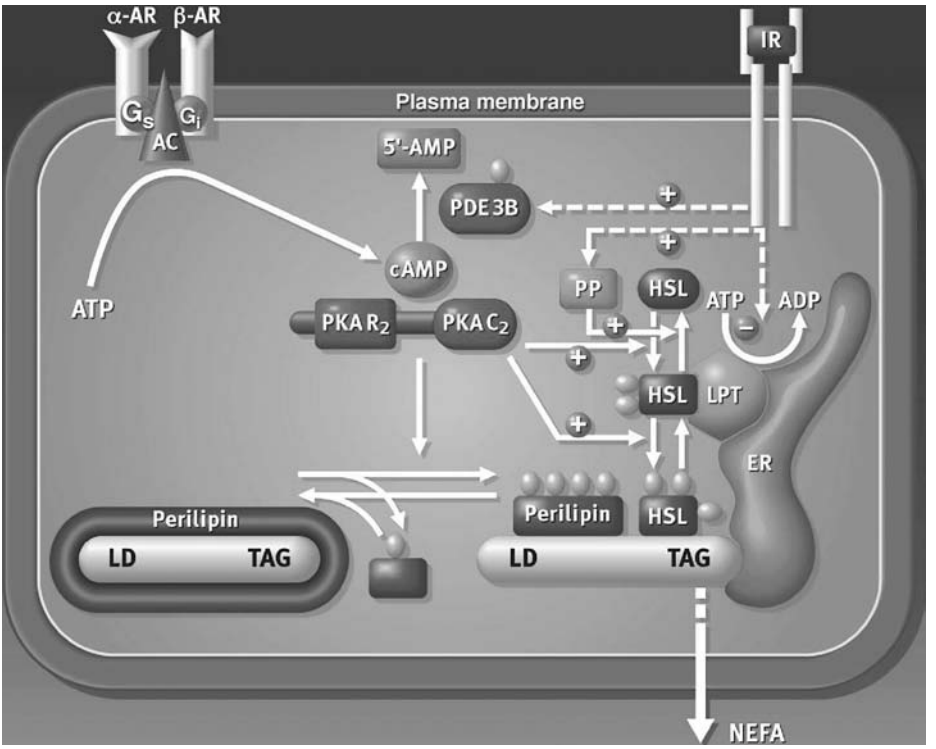

Fig. 6 (Hormonal) regulation of TAG mobilization in adipocytes by translocation of HSL and perilipin as well as the molecular mechanisms involved (phosphorylation by PKA, dephosphorylation by protein phosphatases PP and interaction with lipotransin, LPT, near the ER and LD; the role of lipotransin in HSL translocation reported by one group so far, only, remains to be confirmed and analyzed in greater detail). Phosphorylation is depicted as filled circles. See text for details.

these enzymes. AMPK is itself activated by nanomolar concentrations of fatty acyl-CoA. Thus, a coordinated feedback mechanism may exist whereby elevated levels of FA or fatty acyl-CoAs regulate the levels of free cholesterol and FA in the cell via inhibition of their *de novo* synthesis and mobilization from their stores through lipolysis.

Utilizing an yeast two-hybrid screen of a rat adipose tissue library, it was demonstrated that HSL specifically interacts with adipocyte lipid-binding protein (ALBP or aP2) and the amino-terminal 300 aa of HSL has been identified as the region responsible for this interaction. This adipose-specific fatty acid-binding protein is expressed during adipocyte differentiation and makes up to 6% of cytosolic protein in mature fat cells. Fatty acid-binding proteins (FABP) are abundant low-molecular-weight cytoplasmic proteins that are involved in intracellular trafficking and targeting as well as metabolism of FA. ALBP binds FA, retinoids, and other hydrophobic ligands and is postulated to shuttle FA within the aqueous cytosol toward the membranes of

the relevant intracellular organelles that are involved in TAG synthesis and oxidation. This transfer is believed to occur via a collision mechanism that depends on the interaction of organelle membrane phospholipids with basic residues of ALBP, resulting in partitioning of FA to organelles for metabolic utilization. It has been proposed that intracellular FABP function to sequester FA, thus serving as an intracellular buffer. Because HSL and ALBP interact, it was also hypothesized that ALBP might prevent direct feedback inhibition of HSL by high local concentrations of NEFA released at the site of hydrolysis (see above). Consistent with this view, adipocytes from ALBP-null mice exhibit markedly reduced basal and stimulated lipolysis both *in situ* and *in vivo*, but show few phenotypic changes when fed a standard low-fat diet, with the exception of slightly lower plasma glucose and TAG concentrations in the fasted state. This may be based on increased expression of keratinocyte fatty acid-binding protein in ALBP-null mice, normally expressed at low levels in adipose tissue. However, in contrast to wild-type controls, high-fat feeding did not lead to hyperinsulinemia and insulin resistance in ALBP-null mice, despite similarly elevated body weights. No major abnormalities in adipocyte TAG synthesis were observed in ALBP-null mice with compensation by keratinocyte FABP. Therefore, deficiency of ALBP was associated with partial protection from mild insulin resistance resulting from short-term high-fat feeding, possibly by relieving HSL from feedback inhibition by NEFA. In the opposite concept for the function of ALBP, it delivers FA to HSL, thereby mediating the inhibition of its catalytic activity by FA. Evidence has accumulated that the interaction of ALBP and HSL constitutes an additional mechanism whereby the hydrolytic activity of HSL is regulated (Fig. 7). Thus, incubation of ALBP with purified recombinant HSL *in vitro* led to an increase in TAG hydrolysis. This ALBP-induced increase in hydrolytic activity was due to at least two components. First, a small nonspecific effect of added ALBP, perhaps altering the surface tension of the TAG substrate and thereby leading to interfacial activation of HSL, and, second, a specific effect. The specific effect of the ALPB-induced increase in HSL hydrolytic activity appeared to be based primarily on the ability of ALBP to interact with HSL. The ability of ALBP to bind FA did not seem to be required for this effect since an FA-binding mutant of ALBP, which binds normally to HSL, displayed a similar capacity to increase activity as wild-type ALBP. There was no evidence that the ALBP-induced increase in HSL activity was due to changes in the dimerization of HSL, which has been reported previously. Furthermore, the ability of ALBP to increase HSL hydrolytic activity was also demonstrated *in situ* in cells cotransfected with HSL and ALBP as compared with HSL and vector alone. Importantly, the ability of FA to inhibit HSL hydrolytic activity was attenuated by coincubation with ALBP. Here, the specific effect of ALBP to protect HSL from FA-induced inhibition appeared to depend in part on the ability of ALBP both to interact with HSL and to bind FA since the ALBP FA-binding mutant did not preserve HSL activity as effectively as native ALBP when exposed to higher concentrations of FA. The level of ALBP *in vivo* is, notably, extremely high (estimated to be 250–400 μM), suggesting that, within the cellular context, ALBP can relieve product inhibition by sequestration of FA to a much greater degree than measured *in vitro* or *in situ*. On the basis of the amount of FA released from isolated adipocytes and the

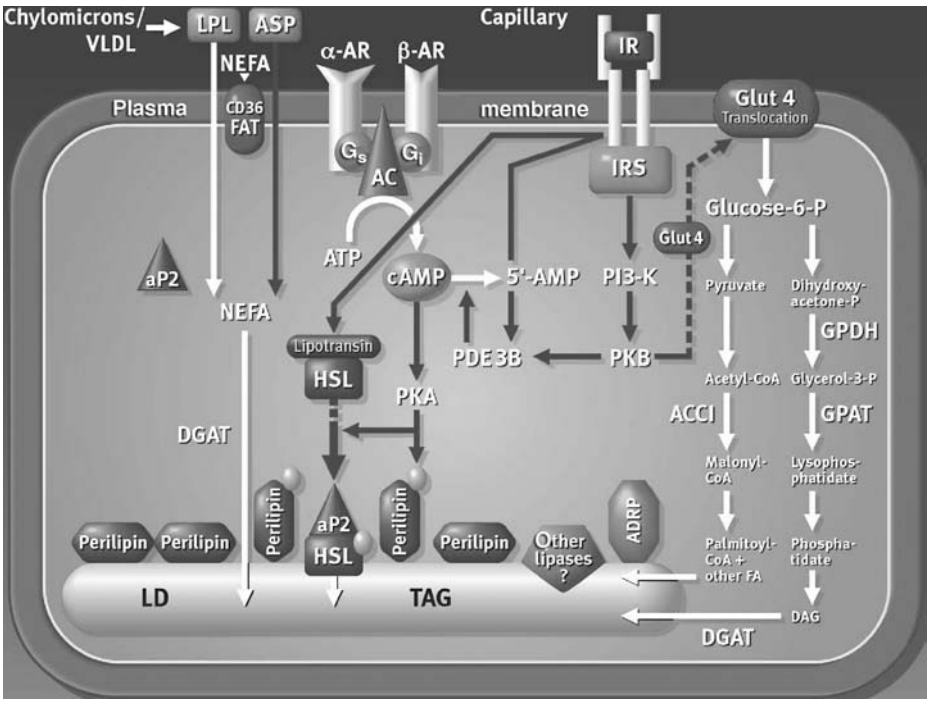

Fig. 7 Working model for the short-term inhibition of HSL activity/phosphorylation by insulin, including mechanisms of feedback inhibition by lipolytic end products and positive antilipolytic cross-talk to the insulin signaling cascade in adipocytes. Operation of a putative PDE3K downstream of PKB as well as direct phosphorylation/activation of PDE3B by the insulin receptor (IR) and/or PI3K is indicated. See text for details.

estimated water volume of an adipocyte, the intracellular concentration of FA is estimated to be $600 \pm 74\ \mu M$. The near 1 : 1 stoichiometry of the total FA pool with ALBP and the high affinity of FA for this protein, measured by a combination of titration calorimetry and fluorescence displacement assays, suggest that FA in the adipose cell are largely, if not exclusively, protein-bound.

By analysis of a series of deletional mutants of HSL in GST pull-down experiments, the region of HSL interacting with ALBP was localized to aa 192–200. In particular, site-directed mutagenesis of His194 or Glu199 eliminated the ability of HSL to interact with ALBP. Thus, these two residues, either directly or indirectly, are critical for mediating the interaction of HSL with ALBP. Interestingly, HSL mutants H194L and E199A, each of which retained normal basal hydrolytic activity but were unable to bind ALBP, failed to display an increase in hydrolytic activity when cotransfected with wild-type ALBP. The retention of normal basal hydrolytic activity by these mutants argues against mutations at these sites affecting HSL dimerization, since monomeric HSL displays only 1/40 the activity of the dimer. Moreover, this is consistent with the finding that no apparent increase in HSL dimerization occurred upon incubation with ALBP. Consequently, the capacity

of ALBP to increase the hydrolytic activity of HSL is, apparently, not due solely to the ability of ALBP to bind and sequester FA, but also appears to depend on the physical interaction of ALBP with HSL. This suggests that the binding of ALBP to HSL might either alter its conformation, allowing TAG to access the catalytic site more efficiently or preventing FA from inhibiting HSL through a conformational change or steric inhibition. Experiments with ALBP-null mice indicate that these functional consequences of the interaction of ALBP with HSL occur *in vivo*. Thus, the current observations are consistent with the proposal that ALBP (aP2) and HSL constitute a lipolytic complex (Fig. 7). This complex functionally results in an increase in the hydrolytic activity of HSL brought about by the physical interaction of HSL with ALBP and by the ability of ALBP to bind and sequester FA. Thereby, lipolysis and intracellular trafficking of FA are controlled in an organized fashion.

3.8 Expression of HSL

Existing adipocyte-like cell lines, such as 3T3-L1, 3T3-F442A, and BFC-1, together with primary adipocyte cultures, have been used to investigate the effects of different nutritional and hormonal factors on HSL expression. cAMP and phorbol esters decrease HSL expression in mature 3T3-F442A and BFC-1 adipocytes via two apparently distinct mechanisms. Insulin, growth hormone, and retinoic acid and oleic acid reportedly had no effect on HSL expression in mature 3T3-F442A adipocytes. In a study on isolated rat adipocytes in culture, dexamethasone increased HSL mRNA fourfold, whereas adrenaline and growth hormone had no effect.

The expression of HSL protein and mRNA levels are lower in subcutaneous fat stores compared with internal (visceral, omental) fat depots in the rat, suggesting a possible basis for the differences in the rate of lipolysis among various fat depots. In contrast, subcutaneous fat in humans has a higher HSL mRNA expression and HSL activity than omental fat. Human subcutaneous fat cells are larger and there is a positive correlation between fat cell size and HSL expression. When controlled for adipocyte cell size, the amount of HSL protein and HSL mRNA levels in subcutaneous adipocytes show a strong correlation with maximum lipolytic activity.

3.9 Non-HSL Lipases

Until recently, HSL was considered to be the principle or the only mediator of lipolysis in WAT. However, gene-targeted HSL-deficient mice with no detectable HSL show active HSL-independent lipolysis. Although $HSL^{-/-}$ mice have a severely blunted response to adrenergic stimulation, the basal lipolytic rate of isolated adipocytes was somewhat greater in $HSL^{-/-}$ cells than in normal cells. In $HSL^{-/-}$ mouse embryonic fibroblasts, the basal lipolytic rate is equal in $HSL^{-/-}$ cells and controls. In homogenates of WAT and of embryonic fibroblasts from $HSL^{-/-}$ mice, total TAG hydrolase (lipase) activity is 40 to 45% of that of normal mice. Therefore, HSL appears to be responsible for hormone-stimulated lipolysis, but the enzyme(s) mediating basal lipolysis remained unknown until the recent use of functional proteomics to detect non-HSL lipase(s) in mouse WAT.

After cell fractionation of intraabdominal WAT, most non-HSL neutral lipase

activity is localized in the 100 000 × *g* infranatant and fat cake fractions. By oleic acid-linked agarose chromatography of infranatant followed by elution in a CHAPS gradient, two peaks of esterase activity were identified using *p*-nitrophenyl butyrate as a substrate. One of the peaks contained most of the lipase activity. In the corresponding fractions, gel permeation chromatography and SDS-PAGE, followed by tandem mass spectrometric analysis of excised Coomassie Blue–stained peptides, revealed carboxylesterase 3 (TAG hydrolase [TGH]). TGH is also the principle lipase of WAT fat cake extracts. Partially purified WAT TGH had lipase activity as well as lesser but detectable neutral cholesterylester hydrolase activity. Western blotting of subcellular fractions of WAT and confocal microscopy of fibroblasts following *in vitro* adipocytic differentiation are consistent with a distribution of TGH to ER, cytosol, and the LD. TGH is responsible for a major part of non-HSL lipase activity in WAT *in vitro* and may mediate some or all HSL-independent lipolysis in adipocytes.

Both subcellular fractionation and confocal immunofluorescence microscopy suggest that TGH is present in the cytosol, in the ER, and also on the LD surface. TGH has previously been isolated from pig liver microsomes and from rat liver cytosol. In WAT, the distribution of TGH with that of the ER luminal enzyme, α-glucosidase II was compared. A greater fraction of TGH than of α-glucosidase II was present in the cytosol, suggesting that the presence of TGH in this compartment may not be due simply to spillage. Other possibilities cannot be eliminated, such as nonhomogeneous distribution of enzymes within adipocyte ER and selective disruption of TGH-enriched ER during preparation. Of note, the microsomal proteins TGH, calnexin, and α-glucosidase II were all present in fat cake. Part of this may be artifactual, due to nonspecific hydrophobic affinity of these proteins for the LD during subcellular fractionation. On confocal microscopy, some TGH immunofluorescence is in immediate proximity to the LD surface, and some merge with that of perilipin. These observations are consistent with a true *in vivo* relationship between the LD surface and ER. This has previously been suggested by electron microscopic data showing proximity of some ER cisternae with the LD surface and by reports that LD may arise from the ER and of the presence of other ER proteins on the LD surface (see above). Given that TGH, a microsomal protein, accounts for most non-HSL intracellular lipase activity in WAT at neutral pH *in vitro*, the possibility that ER may be directly involved in adipocyte lipolysis merits further experimentation.

3.10 Release of FFA

While liberated FA from adipocytes was believed to be transported by simple diffusion, recent studies have identified several membrane proteins that transport NEFA across the adipocyte plasma membrane. These include the FA transport protein, FATP, plasma membrane fatty acid-binding protein and fatty acid translocase, which are expressed in tissues with a high metabolic capacity for FA, such as WAT, skeletal muscle, and heart. Protein-mediated FA transport is assumed to occur via facilitated diffusion along the concentration gradient formed between the intracellular and plasma NEFA concentrations. Thus, during starvation the release of NEFA from adipose tissue into

the plasma is determined by the adipocyte lipolytic state, the coronary flow and the peripheral removal of NEFA. CD36 is a class B scavenger receptor that binds multiple ligands, including FA, and has been proposed to function as the predominant transporter of long-chain NEFA, acting as a gatekeeper. Adipose cells of CD36 null mice show a decreased capacity to incorporate long-chain NEFA into TAG. Although neither the effects of CD36 deficiency on adipose mass, adipocyte size, or insulinemia have been reported, an approximately 30% reduction in plasma glucose concentrations has been observed. Transgenic mice overexpressing CD36 in skeletal muscle showed increased glucose and insulin concentrations together with decreased plasma concentrations of NEFA and TAG at the same time as having increased FA oxidation in muscle. FATP expression is augmented by fasting and damped by insulin.

A number of studies in recent years have implicated changes in circulating FA and FA metabolism with impaired insulin sensitivity in skeletal muscle. A recent study demonstrates for the first time that in muscle from obese subjects and NIDDM patients, rates of FA transport are increased. This was associated with the increased plasmalemmal content of CD36. This greater influx of FA contributes to the excess accumulation of intramyocellular TAG, since the rates of long-chain FA esterification are increased in skeletal muscle from obese subjects. Although there has been some debate as to whether long-chain FA traverses the plasma membrane via unregulated diffusion or protein-mediated processes, molecular evidence indicates that both processes occur. Using animal models, translocation of CD36 from an intracellular depot to the plasma membrane by muscle contraction and by insulin has been demonstrated, and this increase in plasmalemmal CD36 increases FA uptake into muscle tissue. Interestingly, in human muscle of obese and NIDDM patients, the skeletal muscle FA transport rates and plasmalemmal CD36 were significantly increased, while the total available pool of CD36 (plasmalemmal + intracellular) was not altered, which parallels previous observations in muscle from obese Zucker rats. Changes in plasmalemmal FABP were not observed in either insulin-resistant animal muscle or human muscle from obese and NIDDM patients. The impairment of CD36 cycling between subcellular compartments is reminiscent of a similar disturbance in the insulin-responsive glucose transport system. It is well known that GLUT4 cycling between the cell surface and intracellular depots is impaired in skeletal muscle in obesity and NIDDM despite similar levels of total GLUT4. Thus, the impairment in CD36 and GLUT4 cycling are juxtaposed. In obesity and NIDDM, CD36 is retained at the cell surface while GLUT4 is retained within the intracellular depots. Given the strong association between FA transport and intracellular TAG, a marker of insulin resistance (see above), it is tempting to speculate about a cross talk between CD36 and GLUT4 signaling pathways.

The increase in rates of long-chain FA esterification and the positive association between FA transport and TAG concentration in skeletal muscle strongly suggests that the increased long-chain FA transported into muscle contributes to the accumulation of intramyocellular TAG. It has been proposed that the TAG concentrations in obesity and NIDDM are increased owing to an increase in plasmalemmal CD36. This increase then serves to increase the rate of FA transport into the muscle cell, where the excess long-chain

FA that are taken up are primarily stored as TAG. Excess intramyocellular TAG accumulation has been associated with impairments in insulin signaling in human obesity and NIDDM, although TAG depots are most likely to be a surrogate measure of other FA metabolites that interfere with insulin signaling in muscle (see above). Taken together, the increase in the FA transport rate is another disturbance in lipid handling that can occur in skeletal muscle in humans. The key mechanism accounting for this increase is not altered expression of FA transporters, but rather subcellular relocation of CD36, indicating that CD36 cycling between the plasma membrane and its intracellular compartment has been altered in obesity and NIDDM. The increased influx of FA, along with their increased rate of esterification, contributes to the increased intramuscular accumulation of TAG. Thus, the increased rate of FA transport may be a critical step in the long-chain FA-mediated development of insulin resistance.

4 Physiological, Pharmacological, and Genetic Modulation of TAG Mobilization and Storage

4.1 Muscle Contraction

Acute muscular contraction increases HSL activity in mouse soleus. In agreement with this finding, muscle contraction, independent of hormonal influence, increases intramyocellular TAG utilization. The mechanism underlying the contraction-induced increase in HSL activity is unknown. Furthermore, whether intramyocellular HSL activity can be hormonally regulated during muscular contraction is also unclear. The regulation of HSL in skeletal muscle by external hormonal signals during exercise is thought to be unlikely, inasmuch as calcium and metabolites related to the energy status of the cell are expected to be the dominant regulators. However, this may not be the case, since, surprisingly, insulin has been found to play a pivotal role in the regulation of FA oxidation and TAG utilization during intense muscular contraction, when the role of calcium and metabolites would be expected to exert a strong influence. This suggests that the well-known early decline in insulin during exercise may be a major factor in permitting increased TAG mobilization and utilization by muscle. The role of epinephrine in regulating intramyocellular lipolysis during contraction is also relatively unexamined. Infusion of various physiological concentrations of epinephrine during low-intensity cycling (25% peak oxygen uptake) enhances whole-body lipolysis but actually reduces whole-body lipid oxidation. Furthermore, increasing the exercise intensity to 45% peak oxygen uptake blunts the lipolytic action of epinephrine. Thus, the most important hormonal signal early in exercise that permits enhanced lipid utilization by muscle may be the decline in insulin, rather than the increase in epinephrine. The contraction-induced activation of HSL suggests the involvement of calcium in lipolysis control, possibly via Ca^{2+}/calmodulin-dependent kinase(s). Although early observations pointed to a role of calcium at different levels of the adipocyte lipolytic cascade, this has been contradicted by recent reports suggesting that an increase in intracellular calcium concentrations inhibits lipolysis. However, tissue-specific mechanisms may well exist for lipolysis control.

HSL of 84-kDa size has been reported to be expressed in cardiac muscle and skeletal muscle. In skeletal muscle, the expression of HSL is higher in oxidative than glycolytic muscle, and HSL expression is reduced in the muscle of 24-month-old rats, perhaps contributing to the increase in muscle TAG content observed with aging. HSL activity in muscle is stimulated by catecholamines acting via β-AR and cAMP and by contraction acting independently of sympathetic tone or catecholamines. Exercise training does not affect the expression of HSL protein in muscle, but decreases the sensitivity of stimulation of muscle HSL activity by epinephrine. Heart-specific transgenic overexpression of HSL prevents the accumulation of cardiac TAG normally seen in fasted rodents. In addition, heart-specific overexpression of HSL alters the expression of cardiac genes for FA oxidation, transcription factors, signaling molecules, cytoskeletal proteins, and histocompatibility antigens. Thus, HSL in cardiac and skeletal muscle plays a role in controlling the accumulation of LD as well as the mobilization of TAG dependent on the working load. At present there is no experimental evidence for a direct inhibition of HSL by insulin in skeletal muscle.

4.2 Glucose

Glucose deprivation of primary cultured adipocytes results in a slight decline in HSL expression, whereas incubation with high glucose and insulin maintains HSL expression and leads to an increase in basal and stimulated lipolysis. Recently, a glucose-responsive region was mapped within the proximal promoter of human HSL and the involvement of upstream stimulatory factor-1 and -2 binding to a consensus E-box within this region was shown to be responsible for transcriptional regulation in response to glucose metabolism between glucose-6-phosphate and triose phosphates. Glucose deprivation also reduces HSL mRNA levels by a factor of 2.5 to 3 in 3T3-F442A cells. The physiological significance of the glucose/insulin effects on HSL expression remains to be determined, but it may reflect an adaptation to nutritional changes.

Glucose oversupply has been associated with an increase in muscle TAG and malonyl-CoA. Muscular TAG content is strongly related to insulin resistance (see above), and malonyl-CoA is known to inhibit carnitine palmitoyltransferase 1 (CPT-1), which is responsible for the transport of long-chain acyl-CoA into mitochondria. Intramuscular levels of both long-chain acyl-CoA and malonyl-CoA are increased after chronic glucose infusion in rats. These changes correlate with the chronic activation of PKC-ε. However, *de novo* lipogenesis is not assumed to take place in human skeletal muscles to a significant extent. Interestingly, even if glucose is not involved in lipid accumulation, it has been reported to have a negative effect on insulin signaling mediated by PKC isoforms in several cell systems.

Chronic exposure of cultured skeletal muscle cells to high glucose concentrations caused the accumulation of TAG and increased DGAT-1 activity. These changes were accompanied by reduced basal and insulin-stimulated glucose uptake and glycogen synthesis. The conversion of labeled glucose into lipids indicates that *de novo* lipogenesis can take place in skeletal muscle cells, and suggests that the lipid accumulation caused by hyperglycemia is a result of increased lipogenesis. Surprisingly, the elevated lipid synthesis stimulated by glucose was accompanied by reduced palmitate uptake, whereas glycogen

content and expression of GLUT1 and GLUT4 were unchanged. This suggests that chronic hyperglycemia stimulates lipogenesis in myotubes rather than glycogen accumulation. Previous studies have shown that in rats chronically infused with glucose, intramyocellular TAG content is doubled and insulin sensitivity in muscle is decreased. However, hyperglycemia and hyperinsulinemia occur in parallel with metabolic and hormonal changes *in vivo*, and confounding effects of liver and adipose tissue cannot be excluded. Chronic incubation with 20-mM glucose increased intramyocellular TAG content in the absence of exogenously supplied FA and increased the incorporation of glucose into TAG. The increased conversion of [^{14}C]glucose into FFA and complex lipids also implies an effect on lipogenesis. *De novo* lipogenesis is low in skeletal muscle (about 5% of the [^{14}C]glucose incubated was recovered as cellular lipids), but apparently lipogenesis is stimulated under certain conditions (e.g. when glucose is the only source of energy supplied). Interestingly, hyperglycemia increases the activity of DGAT-1. Studies have recently focused on the role of DGAT-1 in obesity and insulin resistance. Mice lacking DGAT-1 have reduced levels of tissue TAG and increased sensitivity to insulin and leptin. Although TAG formation is increased by DGAT-1 after hyperglycemia in human myotubes, DGAT-1 mRNA expression is not elevated. The mechanism by which DGAT-1 activity is increased remains to be elucidated.

The metabolic changes induced by hyperglycemia most likely occur upstream of the Krebs cycle, since glucose oxidation to CO_2 was not altered. One possible explanation could be that glucose oversupply increases the production of the substrates acetyl-CoA and malonyl-CoA for FA biosynthesis. However, an increase in malonyl-CoA would be expected to inhibit CPT-1 and FA oxidation. Hyperglycemia does not alter the oxidation of oleic and palmitic acid. The increase in FFA from labeled glucose suggests that malonyl-CoA and long-chain acyl-CoA may be increased, although possibly not to levels sufficient to inhibit CPT-1. Alternatively, there could be multiple pools of malonyl-CoA in muscle: a cytosolic pool involved in FA synthesis and a pool localized to mitochondria that is involved in the regulation of FA oxidation. The long-chain acyl-CoA or other related lipid products formed might activate one or more PKC isozymes. In line with this fact is the activation of several PKC isoforms in the skeletal muscle of glucose-infused rats.

FA uptake and oxidation are often regulated in concert. Cellular uptake of palmitic acid was decreased after glucose oversupply, while uptake of oleic acid was not and FA oxidation was unaffected. These findings suggest that FA uptake and oxidation can be regulated separately. In agreement, high carbohydrate availability was found to be associated with an increase in palmitate uptake in perfused rat muscle. However, this may be due to the fact that a more acute exposure to a high glucose concentration has been used, where rat muscles were perfused for up to 40 min. The mechanism of reduced uptake of palmitic acid in myotubes is currently unknown, but is apparently specific. There may be a link between increased lipogenesis and the uptake and metabolism of exogenously added palmitic acid.

4.3 Insulin

4.3.1 Molecular Mechanism

Insulin, the most important physiological inhibitor of catecholamine-induced

lipolysis, induces phosphorylation and activation of a phosphodiesterase (PDE), leading to a decrease in cAMP levels and concomitant decrease of PKA activity (Fig. 7). The decrease in cAMP is mainly the result of an insulin-mediated activation of PDE3. By catalyzing the hydrolysis of cAMP and cGMP, PDEs regulate intracellular concentrations and biological responses of these second messengers. PDEs constitute a group of structurally related enzymes that belong to at least nine related gene families (PDE1-9), which differ in their primary structures, affinities for cAMP and cGMP, responses to specific effectors, and sensitivity for inhibitors and regulatory mechanisms. Two PDEs (PDE3 and PDE4) have high affinity for cAMP and are present in fat cells. PDE3 can be distinguished from PDE4 by its high affinity for both cAMP (low K_m) and cGMP (which downregulates PDE3 activity) and is associated with the ER membrane. Activation of PDE3 but not PDE4 plays a role in the antilipolytic effect of insulin *in vivo* in human fat tissue. Two distinct PDE3 subfamilies, PDE3A and PDE3B, products of distinct but related genes, have been identified, of which PDE3B is expressed in adipocytes. The importance of PDE3 as an upstream regulator of HSL activity in insulin's antilipolytic pathway has been shown with the use of specific and cell-permeable PDE3 inhibitors. Specific inhibition of PDE3 in intact cells completely blocks the antilipolytic action of insulin. Activation of PDE3 is associated with phosphorylation of serine-302 in intact cells. This site is also phosphorylated in response to cAMP-increasing hormones, leading to feedback regulation of the cAMP level. In the presence of both hormones, that is, the physiological condition during which insulin exerts its antilipolytic action, there is more than additive phosphorylation of serine-302 associated with a more than additive activation of PDE3, suggesting positive cross talk between the pathways upstream of PDE3. Thus, the regulation of cAMP content, PKA activity, and thereby the activity of HSL is very complex and includes feedback loops as well as cross talk between different pathways.

Upstream regulation of PDE3B can be summarized as follows (Fig. 7): The insulin-stimulated PDE3 kinase (PDE3K), apparently a serine kinase activity, may be identical to that intrinsically associated with the insulin receptor. PDE3B may thus represent the first substrate of the insulin receptor kinase, which is phosphorylated at serine leading to its activation (Fig. 7). However, other experimental evidence argues for the dual specificity protein/lipid kinase, PI3K, consisting of a regulatory and catalytic subunit, as the major upstream regulator of PDE3B and, consequently HSL, since treatment of the cells with the PI3K inhibitor wortmannin completely blocks activation of PDE3K, phosphorylation/activation of PDE3B, and finally the antilipolytic action of insulin (Fig. 7). Upon insulin stimulation, PI3K binds to specific tyrosine-phosphorylated motifs on the insulin receptor substrate protein-1 (IRS-1), the major substrate for the insulin receptor tyrosine kinase (IR). The signal transduction between PI3K and PDE3K is not yet known, but probably involves protein kinase B (PKB; Fig. 7). The well-known insulin-stimulated serine kinases, MAPK or p70S6 kinases, apparently are not involved in the antilipolytic pathway of insulin.

In addition to the short-term insulin regulation of cAMP degradation, long-term effects of insulin on HSL have been reported. HSL activity, immunoreactive protein, and mRNA levels in adipose tissue were increased in response to

streptozotocin-induced insulin deficiency in the rat, whereas acute treatment with insulin returned HSL activity to normal without altering the increased amounts of HSL immunoreactive protein and mRNA. Increased HSL activity after prolonged exposure of isolated adipocytes to insulin has been reported in one study, whereas others did not reveal an effect of insulin at the mRNA expression level of HSL. Thus, long-term insulin regulation of HSL and the involvement of pretranslational mechanisms remain under debate, while short-term treatment with insulin controls HSL by posttranslational mechanisms.

4.3.2 Desensitization

Long-term exposure of 3T3-L1 adipocytes to insulin at or near physiological concentrations results in increased basal glycerol release in the medium without a concomitant rise in NEFA release. These conditions also altered β-adrenergic-stimulated lipolysis, exaggerating glycerol but preventing NEFA release. Stimulation of lipolysis by long-term insulin treatment, as measured by glycerol release is an unexpected finding given the short-term antilipolytic action of insulin. This finding, however, agrees with previous studies that showed increased glycerol release after prolonged insulin treatment of isolated rat adipocytes. One of these studies showed that the effect of long-term insulin depends on the presence of high glucose concentrations in the medium. Yet, the role of the hexosamine pathway in mediating this effect of long-term insulin treatment seems unlikely as increased glycerol release occurred irrespective of the presence of glutamine in the medium. In this study, long-term exposure to insulin was done in culture medium, which contained 25-mM glucose. Yet, at the end of the incubation period, glycerol was collected for 1 h in a buffer without glucose. Thus, it seems that extracellular glucose could be necessary as a priming factor, particularly under high glucose conditions, but is not required for the actual metabolic flux responsible for glycerol generation. Collectively, these studies demonstrate the ability of insulin, even at subnanomolar concentrations, to stimulate glycerol release upon prolonged exposure as opposed to its short-term inhibitory effect on lipolysis.

In contrast to the effect on glycerol release, long-term exposure to insulin was not associated with increased NEFA concentrations in the medium. Despite being the two end products of TAG breakdown, glycerol and NEFA released from adipocytes rarely reach the expected 1 : 3 molar ratio. Short-term insulin stimulation rapidly inhibits NEFA release from adipose tissue, through a combined inhibitory action on HSL and a stimulatory effect on NEFA reesterification. Recent findings support the possibility that although long-term insulin stimulates glycerol release, the short-term insulin stimulatory effect on NEFA reesterification is maintained under prolonged treatment. Because the antilipolytic effect of insulin and the stimulation of NEFA reesterification have been shown to involve PI3K, the two pathways seem to diverge downstream of PI3K. These seemingly independent pathways to affect glycerol and NEFA release from adipocytes also respond to leptin, which was shown to stimulate glycerol but not NEFA release from isolated adipocytes (see below). A role for leptin in mediating the effects of long-term insulin on glycerol and NEFA release from adipose tissue is an intriguing possibility. In conclusion, long-term treatment with insulin apparently leads to desensitization of the signaling pathway downregulating HSL activity without

concomitant negative interference with stimulation of the reesterification pathway. The physiological role of long-term insulin regulation of HSL activity remains unclear.

4.4 Leptin

Studies with rodents have demonstrated that leptin induces lipolysis both *in vitro* and *in vivo*. The effect was absent in animals lacking a functional leptin receptor, that is, db/db mice and in Zucker rats. Intriguingly, FA release does not accompany glycerol release in leptin-induced lipolysis, which suggests simultaneous induction of FA oxidation. In these studies, effects on HSL expression were not investigated. Thus, it is not known whether the lipolytic effect of leptin is mediated via phosphorylation-induced activation of HSL, an induction of HSL expression, a combination of both, or some other mechanism. In another study, however, long-term treatment of mice with leptin increased HSL mRNA expression by 30% in WAT, whereas no effect was seen in brown adipose tissue. The leptin signal is presumably transduced via the JAK/STAT pathway. Leptin was also shown to directly activate the TAG/FA substrate cycle, lipolysis, and FA oxidation, shifting fuel preference from carbohydrate to FA oxidation. In particular, studies with isolated rat adipocytes in primary culture demonstrated that leptin impairs insulin inhibition of isoproterenol-induced lipolysis and PKA activation. These insulin effects were reduced by leptin (2 nM) with a half-life of 8 h. Leptin concentrations below 1 nM led to a rightward shift of the corresponding insulin concentration–response curves. At leptin concentrations above 30 nM, the responsiveness was diminished, and also resulted in nearly complete relief of lipolysis from insulin control. The IC_{50} for leptin was about 3 nM after 15 h of preincubation of the primary adipocytes. The natural splice variant des-Gln49-leptin exhibited a significantly lower potency. Adipocytes regained full sensitivity for insulin's antilipolytic action within a few hours after leptin removal. Consequently, in ob/ob mice, the loss of leptin stimulation of uncoupling and TAG/FA substrate cycling triggered, at least in part, by leptin blockade of the antilipolytic activity of insulin will simultaneously contribute to the decrease in TAG mobilization and utilization as well as to the increase in TAG synthesis and storage. In conclusion, accelerated TAG/FA substrate cycling in adipose tissue provides a new mechanism by which leptin triggers increased metabolic rate and energy expenditure.

4.5 Growth Hormone

Growth hormone is another hormone signaling *via* the JAK/STAT pathway and with documented lipolytic effects, albeit with variable efficiency, depending on species and experimental system. Using primary rat adipocytes, growth hormone was found to increase HSL activity, that is, lipolysis, without causing an increase in HSL mRNA levels, supporting a role for posttranslational mechanisms in controlling lipolysis. In agreement, cGMP levels appear to be elevated moderately upon challenge with growth hormone. cGMP has been shown to mimic cAMP as an activator of lipolysis. This is supported by *in vitro* studies showing that cGMP-dependent protein kinase phosphorylates

and activates HSL. cGMP-dependent protein kinase is present at very low levels in adipose tissue, and the physiological role of this kinase and cGMP for activation of lipolysis, if any, is not known. Recently, nitric oxide was shown to exert modulatory actions on lipolysis in adipose tissue via cGMP-independent mechanisms. The effects were complex, with different redox forms acting either stimulatory or inhibitory.

4.6 Glucagon-like Peptide-1

Inhibition of lipolysis in rat islets by orlistat has been shown to inhibit the incretin action of glucagon-like peptide-1 (GLP-1). It was therefore a surprise that GLP-1 was able to rescue the loss of GSIS in isolated islets from fasted male $HSL^{-/-}$ mice. Hence, HSL is not essential for the incretin action of GLP-1. Although it is possible that GLP-1 was able to overcome the effect of FA deprivation via signaling pathways independent of FA metabolism, the findings could also be consistent with the action of GLP-1 to stimulate an alternative lipase (see above). The failure of GLP-1 to significantly alter glycerol release does not rule out this possibility. Given that FA deprivation has previously been shown to block insulin secretion to both fuel and nonfuel secretagogues like GLP-1, the latter possibility seems to be more likely. Clearly, the characteristics of other islet ß-cell lipase activities, including their responsiveness to PKA activators, such as GLP-1, warrant further investigation. Of relevance, isoproterenol, a ß-agonist believed to promote lipolysis through HSL, has been shown to stimulate lipolysis in adipocytes from $HSL^{-/-}$ mice. This supports the notion of another cAMP-mediated lipase in adipose tissue and perhaps in the ß-cell as well.

4.7 TNF-α

Of the various factors suggested to play a part in impaired adipocyte function in insulin resistance and NIDDM, increased TNF-α expression and production have attracted interest. TNF-α is produced and secreted from adipose tissue in obesity and thus acts in an autocrine fashion to alter adipocyte function during obesity-linked insulin resistance. Long-term exposure (>6 h) to TNF-α has been shown to stimulate lipolysis in adipocytes, despite inducing reduced HSL expression. An early study showed a downregulation of HSL gene expression upon TNF-α treatment of 3T3-L1 cells, as measured by Northern blot analysis and enzyme activity measurements. A similar effect, although much more moderate, was seen more recently at the protein level. In a study with primary rat adipocytes, however, no alteration of the levels of HSL protein occurred upon treatment with TNF-α. Studies using human adipose tissue have now shown that TNF-α inhibits LPL activity by downregulating its protein expression. Increased TNF-α mRNA levels are correlated with decreased LPL activity in human subcutaneous adipose tissue. In addition, TNF-α reduces the expression of FA transporters in adipose tissue of the Syrian hamster, and decreases the expression of enzymes involved in lipogenesis, such as acetyl-CoA carboxylase, FA synthase, and acyl-CoA synthase. TNF-α could thus impair the synthesis and entry of FA into the adipocyte, curtailing an increase in the intracellular TAG pool size.

TNF-α has also been found to promote lipolysis. Indirectly, the downregulation

of HSL observed in 3T3-L1 cells may reflect the known dedifferentiating effect of TNF-α in these cells, whereas, directly, the lipolytic effect of TNF-α could be explained by the decrease in LPL activity and intracellular reesterification and/or downregulation of G_i and perilipin expression. The latter effect of TNF-α, as well as the activation of lipolysis, was prevented by thiazolidinediones, a class of insulin-sensitizing antidiabetic agents. The mechanisms involved in regulation of lipolysis by TNF-α remain unclear. A study with human subcutaneous adipocytes demonstrated that, concomitant with an increase in lipolysis induced by TNF-α, there is activation of ERK1/2, leading to a decrease in the expression of PDE3B, together with an elevation in intracellular cAMP and hence activation of PKA. Since these two pathways are classically not coupled, the significance of changes in both is unclear in relation to an increase in lipolysis. Interestingly, TNF-α stimulated the three mammalian MAPK, p42/44 (ERK2/1), JNK, and p38 in time- and concentration-dependent manners in parallel to induction of lipolysis (3-fold at maximal dose). Lipolysis was completely abrogated by inhibitors specific for p42/44 and JNK but was not affected by a p38 inhibitor. These data suggest involvement of p42/44 and JNK in the TNF-α upregulation of lipolysis in human preadipocytes. The close association between the expression of enzymes involved in lipolysis and lipogenesis with adipogenesis has made it difficult to tease out process-specific actions of TNF-α. TNF-α can thus act on the adipocyte to shift lipid metabolism away from lipid accumulation and toward a reduction in the TAG pool, which could represent one mechanism by which TNF-α can reduce total adipose tissue weight.

Severe weight loss or cachexia is the detrimental end point of several diseases, including cancer, infection, and congestive heart failure. Unlike starvation, cachexia results from a severe loss of lean body mass (mostly skeletal muscle) and the depletion of the fat depots. The pathogenesis of this weight loss is multifactorial, but evidence suggests that cytokines are key players. The predominant reason for the loss of adipose tissue is the fall in the activity of LPL and an increase in the activity of HSL. Since TNF-α promotes lipolysis and inhibits lipogenesis, it represents an ideal player in the depletion of adipose tissue mass seen with cachexia. An elevation in plasma levels of TNF-α, as opposed to adipose tissue-derived TNF-α, is suggested to be responsible for the metabolic alterations in adipose tissue with cachexia.

4.8 Acylation-stimulating Protein

Acylation-stimulating protein (ASP) is identical to C3adesArg, a cleavage product of complement C3. Cleavage of complement C3 is mediated through the alternate complement pathway via the interaction of C3, factor B, and adipsin that generates C3a. Rapid cleavage of the carboxy-terminal arginine of C3a by carboxypeptidase N generates ASP. Adipocytes are one of the few cells capable of producing all three factors (factor B, adipsin, and C3) that are required for the production of ASP. ASP production increases consequent to adipocyte differentiation, and plasma ASP levels are elevated in obesity. Chylomicrons *in vitro* stimulate ASP production by adipocytes. *In vivo* arterial–venous gradients across a subcutaneous adipose tissue bed in humans demonstrate direct postprandial production of ASP. The postprandial increase

in ASP is adipose tissue–specific and is not observed in the general circulation. Altogether, these data suggest that ASP and TAG storage are metabolically intertwined. ASP acts as an adipocyte autocrine factor and has been proposed to play a central role in the metabolism of adipose tissue by increasing the efficiency of TAG storage in adipocytes, an action that results in more rapid postprandial TAG clearance. This is based on stimulation of both glucose uptake and NEFA storage in human adipose tissue via translocation of glucose transporters (GLUT1, GLUT3, and GLUT4) from intracellular sites to the cell surface and an increase in DGAT activity. In addition, ASP has been shown to inhibit HSL in adipocytes, independently and additively to insulin.

The pathways distal to PI3K that are involved in the effect of insulin on FFA reesterification and in the action of ASP on FFA reesterification and lipolysis were unknown until now. PKC blockers could not counteract the effects of ASP or insulin on fractional NEFA reesterification or lipolysis, suggesting that the actions of the two hormones on NEFA release are not mediated by PKC. This is in contrast to earlier results on the action of ASP on TAG synthesis and glucose transport in human fibroblasts, which was counteracted by PKC blockers. The PI3K blocker wortmannin counteracted the effect of insulin on NEFA release by reversing the effect of insulin on both fractional NEFA reesterification and lipolysis. The effects of ASP, however, were not affected by wortmannin. Thus, the pathway mediating the effect of ASP on NEFA release does not involve PI3K. This strongly supports the hypothesis that ASP and insulin follow different pathways in their action on NEFA release. Because the ASP receptor protein(s) has not been identified, further research in this area is difficult at present. The question whether a more distal pathway could be common for the actions of insulin and ASP on NEFA metabolism was addressed. Interest was focused on PDE, which previously was found to mediate not only the antilipolytic effect of insulin (see above) but also contributes to the stimulatory effect of the hormone on glucose transport in human fat cells. A nonselective PDE inhibitor that does not interact with the antilipolytic effects of adenosine, counteracted the actions of both ASP and insulin on NEFA release by reversing their effects on fractional NEFA reesterification and lipolysis. The selective PDE3 blocker OPC3911 could also counteract these effects of ASP and insulin. The selective PDE4 blocker rolipram could not counteract the effect of insulin on fractional NEFA reesterification and lipolysis. Rolipram did not alter the effect of ASP on lipolysis either, but it counteracted slightly the effect of ASP on fractional NEFA reesterification. These results agree with those obtained with direct measurements of PDE activity. PDE3 activity in fat cells was stimulated with both ASP and insulin in a concentration-dependent manner, whereas PDE4 activity was only slightly stimulated with ASP. Thus, the effects of both ASP and insulin on NEFA release appear to be mediated predominantly by PDE3. PDE4 seems to be involved in the action of ASP as well, although to a lesser extent than PDE3.

4.9 Acipimox, Nicotinic Acid, and Niacin

Acipimox (5-methylpyrazine carboxylic acid-4-oxide) is a newer, well-tolerated derivative of nicotinic acid (NA) which shows lipid-lowering properties in patients

with types II and IV hyperlipidemia. The fall in plasma TAG is associated with a decrease in total and LDL cholesterol and an increase in HDL cholesterol. NA has long been used in the treatment of dyslipidemia, producing a very desirable modification of multiple cardiovascular risk factors, increasing HDL, TAG, and lipoprotein(a), which results in a reduction in mortality. NA is an effective therapeutic agent. However, it has to be administered at high doses and has a characteristic side effect profile defined by intense, but transient, prostaglandin-mediated cutaneous vasodilation ("flushing") that affects patient compliance. Despite their long history of clinical use, the precise mechanism of action of acipimox and NA was unknown until recently, although inhibition of adipocyte lipolysis via the activation of a G_i-coupled receptor is believed to contribute. It has been postulated that a reduction in NEFA liberated from adipose tissue results in reduced influx of NEFA into the liver and hepatic reesterification of NEFA and thereby diminishes the production rate of hepatic TAG available for VLDL and LDL synthesis, which in part explains the hypolipidemic effects observed during NA and acipimox therapy.

Whereas most studies have focused on the effects of NA on lipid metabolism, the action of NA on carbohydrate metabolism is less well understood. After acute NA administration, glucose concentrations have been reported to decrease, rise or remain unaltered in rats and humans. Results of glucose tolerance tests after acute NA intake have also been inconsistent. Chronic administration of NA has consistently resulted in deterioration of glucose tolerance and elevation of fasting blood glucose concentrations in normal humans and impairment of glycemic control in NIDDM patients. These effects are contrary to expectations based on the glucose-FA cycle. If reduction of lipolysis and NEFA availability reduces oxidation of fat in tissues and availability of gluconeogenic precursors in the liver, insulin sensitivity should improve and glucose concentrations fall, as observed with inhibitors of mitochondrial FA oxidation. It is unclear whether this failure of NA to act as an antidiabetic drug is caused by its unfavorable pharmacokinetic profile (short half-life time) and/or its mode of action as inhibitor of adipocyte lipolysis.

Niacin is an effective, unique lipid-regulating agent that beneficially reduces plasma TAG, cholesterol, and atherogenic apolipoprotein B (apoB)-containing lipoproteins (VLDL, LDL, and lipoprotein [a]) and increases antiatherogenic apoA-I-containing HDL levels. Several clinical trials have demonstrated that treatment with niacin significantly reduces total mortality and coronary events and retards the progression or induces the regression of coronary atherosclerosis. Despite its wide usage as a broad-spectrum lipid-regulating agent, the cellular and molecular mechanisms by which niacin modulates the hepatic lipid metabolism and the production of VLDL and LDL particles are incompletely understood. Using HepG2 cells as an *in vitro* model system, it has previously been shown that niacin inhibits the incorporation of radiolabeled oleic acid or glycerol into TAG, suggesting decreased *de novo* synthesis of TAG by niacin. Additionally, niacin may also increase intracellular apoB degradation in HepG2 cells. Because the synthesis and availability of TAG play a critical role in intracellular apoB processing and secretion of apoB-containing lipoproteins; previous studies suggested that niacin, by inhibiting TAG synthesis, increased intracellular apoB degradation,

resulting in reduced secretion of apoB-containing particles. In support of these *in vitro* studies, earlier turnover studies in humans suggested that niacin decreased the production (transport) rate of TAG from radiolabeled FA, thus decreasing TAG-rich lipoproteins (e.g. VLDL) and their product LDL.

4.9.1 Physiological Mode of Action

The effect of acipimox on adipose tissue lipolysis has been examined under three experimental conditions. Firstly, in the presence of high levels of the nonselective ß-AR agonist isoproterenol, secondly in the presence of low levels of adenosine deaminase (ADA), which when used alone has a minimal effect on basal lipolysis, yet when combined with submaximal concentrations of isoproterenol, substantially facilitate the lipolytic response. Thirdly, acipimox was examined in the presence of high levels of ADA. Fat cells produce adenosine from cAMP via consecutive action of PDE and 5′-nucleotidase, which, upon release into the medium, act at specific AR to inhibit AC, this inhibitory effect being mediated by the GTP-dependent, pertussis toxin-sensitive regulatory protein, G_i. High levels of ADA, by converting the endogenous adenosine into inosine, relieve the inhibition of AC, which results in the observed increase in basal lipolysis. The data show that acipimox has only a minor, if any, antilipolytic effect on lipolysis stimulated by isoproterenol alone. In contrast, when the lipolytic rate is increased by a lower concentration of isoproterenol in combination with ADA, acipimox becomes a potent antilipolytic agent. Wieser and Fain reported similar results for NA, with the addition of ADA increasing the sensitivity to catecholamine-induced lipolysis toward inhibition by NA. The dependence of acipimox upon removal of adenosine suggests that the antilipolytic mechanisms downstream of the AR and the putative acipimox receptor (see below) and upstream of AC are identical for adenosine and acipimox, respectively.

These results show that acipimox acts to inhibit cAMP-stimulated lipolysis through suppression of intracellular cAMP levels, with subsequent decrease in PKA activity. The antilipolytic activity of acipimox is associated with a quantitative shift of HSL from the LD to the cytosol in rat adipocytes, suggesting that, by decreasing intracellular cAMP levels, acipimox can reduce the association of HSL with TAG substrate of the LD. Pharmacokinetic studies show that the plasma concentration of acipimox in man after a single 250 mg oral dose reaches 33 μM, and remains above 7 μM for 6 to 8 h. In comparison, acipimox is a potent antilipolytic agent at a concentration of 10 μM and above, in human adipose tissue stimulated maximally and submaximally by isoproterenol. Thus, plasma acipimox levels correlate well with its *in vitro* effect on lipolysis and cAMP synthesis in adipocytes.

Work on NA showed that it also decreased intracellular cAMP levels through inhibition of AC. The conditions were identical to those required to inhibit the enzyme by the antilipolytic hormonal factors, prostaglandin E_2, adrenaline, and the adenosine analog, PIA. These findings indicated that NA inhibits AC via a receptor-mediated process in a hormone-like manner, which indicates the involvement of the regulatory protein, G_i. This was confirmed by the discovery that the antilipolytic effects of prostaglandin E_2, PIA, and NA were abolished by the pretreatment of adipocytes with pertussis toxin, to inactivate G_i. Use of the AR antagonist, 3-isobutyl-1-methylxanthine demonstrated that NA and adenosine act through

separate receptors to inhibit adipocyte AC activity. Since NA shares its ability to inhibit lipolysis via a G_i-dependent mechanism with acipimox, and since the drugs are structurally related, it seemed possible that acipimox suppresses intracellular cAMP levels through binding to the same G_i-linked receptor (see below).

4.9.2 Molecular Mechanism

G-protein activation by NA and derivatives was recently assessed as stimulation of guanosine 5′-(γ-[^{35}S]-thio)triphosphate binding, and [^{3}H]NA was used to specifically label binding sites. NA (EC_{50} 1 μM) stimulated [^{35}S]GTPγS binding in membranes from rat adipocytes and spleen, but not from other tissues. G-protein activation in adipocyte membranes in the presence of maximally activating concentrations of a selective A_1 AR agonist, 2-chloro-N^6-cyclopentyl-adenosine, and NA was almost additive, indicating that G-proteins of mostly distinct pools were activated by these agonists. G-protein activation by NA and related substances in spleen and adipocytes revealed identical pharmacological profiles. [^{3}H]NA specifically detected guanine nucleotide-sensitive binding sites of identical pharmacology in adipocyte and spleen membranes. These data indicate that NA most probably acts on a specific G-protein-coupled receptor (GPCR).

To identify the G_i-linked GPCR for NA, orphan receptors were selected on the basis of their tissue expression profiles for a rational screening. Subsequently, a subset of 10 orphan GPCR, which by mRNA distribution analysis exhibited significant expression levels in both adipose tissue and spleen, were expressed in an appropriate mammalian cell line to allow measurement of functional GTP$_\gamma$S binding following NA treatment. Thereby, HM74 was identified as a low-affinity receptor for NA, which represents an orphan GPCR previously cloned from a cDNA library derived from human monocytes. The half-maximal effector concentration for NA with 1 mM at HM74 was 1000-fold higher than that previously reported in membranes produced from rat adipose tissue or spleen. Subsequent identification of a novel paralog of HM74, termed HM74A, by a molecular biology approach and its expression in various test systems revealed that HM74A acts as a high-affinity receptor for NA. The activity and affinity of NA was in good agreement with that previously reported. In the [^{3}H]NA displacement assay, both the absolute potency and the rank order of potency of various HM74A ligands was the same, whether tested against the stable CHO cell line expressing recombinant human HM74A or human adipocytes. These data strongly suggest that HM74A is the G_i-linked GPCR for NA on human adipocytes. Compatible with this conclusion, nicotinamide, which unlike NA produces no alteration in lipoprotein profiles, acted only as a very weak agonist at HM74A with 1000-fold lower potency than NA. In contrast, acipimox has also been identified as a full high-affinity agonist for HM74A. Thus, activation of HM74A seems to account for the inhibition of lipolysis observed with NA and acipimox. Therefore, of the NA derivatives tested in man, it appears that potency at HM74A is linked with their efficacy in normalizing lipoprotein profiles.

Using HepG2 cells, the effect of niacin on the mRNA expression and microsomal activity of DGAT1 and DGAT2, the last committed but distinctly different enzymes for TAG synthesis was examined. Addition of niacin to the DGAT assay reaction mixture dose-dependently (0–3 mM) inhibited DGAT activity by 35 to 50%,

and the IC_{50} was found to be 0.1 mM. Enzyme kinetic studies showed apparent K_m values of 8.3 and 100 μM using [^{14}C]oleoyl-CoA and *sn*-1,2-dioleoylglycerol as substrates, respectively. A decrease in apparent V_{max} was observed with niacin, whereas the apparent K_m remained constant. A Lineweaver–Burk plot of DGAT inhibition by niacin showed a noncompetitive type of inhibition. Niacin selectively inhibited DGAT2 but not DGAT1 activity. Niacin inhibited overt DGAT activity. Niacin had no effect on the expression of DGAT1 and DGAT2 mRNA. These data suggest that niacin directly and noncompetitively inhibits DGAT2 but not DGAT1, resulting in decreased TAG synthesis and hepatic atherogenic lipoprotein secretion, thus indicating a major target site for its mechanism of action.

The inhibitory effects of niacin on FA synthesis and inhibition of peripheral lipolysis (e.g. adipose tissue) and decreased FA mobilization may be additional mechanisms of action to explain TAG reduction by niacin. In the absence of other regulatory events, niacin-mediated inhibition in adipocyte FA mobilization over prolonged steady state conditions would result in excessive synthesis of TAG and obesity. However, this mechanism is unlikely because niacin is not known to affect body weight clinically. On the basis of observations in hepatocytes, it is likely that niacin may also inhibit DGAT2 activity in adipocytes. This may provide a regulatory mechanism to limit excessive adipocyte TAG synthesis and related obesity events in the face of the effect of niacin on adipocyte FA mobilization. Although the biological actions of niacin are not clearly known to be mediated by receptor-mediated events, recent reports indicate that the orphan G-protein-coupled receptor mouse PUMA-G and human HM74 act as cellular receptors for niacin (see above). PUMA-G and its human ortholog HM74 are highly expressed in white and brown adipose tissue and are also detected in various other tissues, including lung, adrenal gland, and spleen. However, PUMA-G and HM74 were not detected in liver tissue. Using PUMA-G-deficient mice, it has been demonstrated that PUMA-G mediates the niacin-induced inhibition of FFA release from isolated adipocytes. Because of the lack of PUMA-G and/or HM74 in liver tissue, it is very unlikely that PUMA-G/HM74 mediates the effect of niacin on DGAT activity and TAG synthesis in hepatocytes.

In conclusion, inhibition of DGAT2 activity in liver by niacin may result in a decreased rate of TAG synthesis and its reduced availability for intracellular apoB lipidation and translocation across the ER membrane, resulting in observed effects of increased apoB degradation. Thus, these results suggest DGAT2 as a major site of action of niacin to inhibit TAG synthesis, resulting in decreased secretion of apoB-containing atherogenic lipoproteins. Niacin may represent a new therapeutic class of drugs that regulate atherogenic apoB-containing lipoprotein (e.g. VLDL and LDL) secretion by specifically targeting DGAT2.

4.9.3 Desensitization

The established phenomenon that exposure of a cell to an agonist can cause desensitization, so that a second exposure to the agonist is less effective than the first, has been studied extensively and various mechanisms have been described. First, short-term exposure to a ligand (minutes) can induce a change in the receptor. For example, ß-AR can become phosphorylated by a specific kinase (ß-AR kinase) after exposure to an agonist. Second, after

more prolonged exposure to an agonist, the number of receptors on the cell surface can decrease by a process termed *downregulation*. Downregulation is generally considered to occur in two steps. Initial rapid sequestration of receptors from the cell surface through endocytosis is followed by a slower loss of total cellular receptors involving intracellular degradation, although at least a portion is often recycled back to the cell surface, as has been demonstrated for the insulin receptor. In addition to increased degradation of receptors, recent studies for ß-AR have suggested that alterations in receptor mRNA turnover, that is, decreased transcription and/or increased mRNA degradation, may also play a role. These phenomena can readily explain homologous desensitization, where exposure of a cell to a hormone results in subsequent insensitivity to that hormone. Another commonly reported phenomenon is known as heterologous desensitization, in which treatment of a cell with a hormone subsequently can make the cell less sensitive to another hormone that operates through a different receptor. Heterologous desensitization has been described for many signaling systems, including GPCRs coupled to AC and to phospholipase C. Mechanisms of heterologous desensitization, however, are relatively poorly understood.

A_1-AR can, reportedly, be downregulated in rat adipocytes following prolonged incubation with the nonmetabolizable adenosine analog and agonist, PIA. Although there are two major classes of AR, known as A_1 and A_2 receptors, adipocytes primarily possess receptors of the A_1 class. These receptors are inhibitory toward AC, and their action is thought to be mediated via a pertussis toxin-sensitive G-protein known as G_i, of which three subtypes 1–3 have been identified in rat adipocytes. Prolonged incubation of adipocytes with PIA leads to a marked downregulation of A_1-AR. In addition, an approximately 50% decrease in G_i, as measured by pertussis-catalyzed ADP-ribosylation, has been found. This observation suggested that G_i was downregulated along with the AR. Subsequently, a series of specific antisera was used to demonstrate that PIA causes a marked (ca. 90%) loss of G_i1 and G_i3 with a more modest (50%) loss of G_i2. When PIA is washed away, G_i downregulation was reversed, suggesting that it is a real regulatory phenomenon rather than a simple toxic effect. Furthermore, Stiles and coworkers reported similar findings in adipocytes derived from rats after chronic infusion of PIA. More recent studies suggest that G-protein downregulation is a common phenomenon resulting from chronic exposure to agonists. Since several agonists can presumably couple through a single G-protein, it is clear that this phenomenon of G-protein downregulation could account for heterologous desensitization.

In addition to impairment of signal transduction activity at the level of G-proteins, heterologous desensitization of GPCR regulation of lipolysis can be caused by changes at the level of cell surface expression of GPCRs or signaling activity of signaling components, such as AC, PDE, and PKA (catalytic activity), as well as function of the lipolytic end effectors (e.g. HSL/perilipin translocation apparatus). Interestingly, prolonged incubation of rat adipocytes in primary culture with NA at concentrations up to 1 mM had no effect on the levels of G_i1-3 as well as AC. However, pretreatment of the cells with 1 mM NA for four days decreased their sensitivity for subsequent acute inhibition of β-adrenergic stimulation of lipolysis and

AC activity by NA by 1 order of magnitude (but had no effect on the sensitivity of the cells toward PIA). The molecular mechanism of the homologous desensitization of lipolysis by NA, which may form the molecular basis for this observed loss of efficacy of NA and acipimox treatment during chronic antihyperlipidemic therapy, remains to be elucidated.

4.10 Glimepiride and Phosphoinositolglycans

Recent studies have identified two completely different exogenous stimuli, which mimic the antilipolytic activity of insulin in insulin target cells. They bypass activation of the insulin receptor, and rather act by triggering redistribution of signaling components between DIGs of different lipid and protein composition at the target cell plasma membrane. (1) Phosphoinositolglycan-(-peptides) (PIG[-P]) prepared by chemical synthesis or isolation from natural sources reduce isoproterenol-stimulated lipolysis in concentration-dependent fashion to almost basal levels in primary and cultured adipocytes. PIG(-P) are derived from the polar core glycan head group of glycosylphosphatidylinositol (GPI) anchored proteins consisting of 2′,3′-cyclic phospho-*myo*-inositol coupled to five sugar moieties in typical glycosidic linkages (PIG portion) and, optionally, contain an additional terminal ethanolamine-linked tripeptide residue (P portion); (2) The hypoglycemic sulfonylurea, glimepiride, which in NIDDM patients exerts an insulin-independent blood glucose-decreasing activity in addition to its insulin-releasing potency blocks isoproterenol-stimulated lipolysis in adipocytes from insulin-resistant Zucker fatty rats or in adipocytes made insulin-resistant *in vitro* by incubation in primary culture in the presence of high concentrations of insulin and glucose.

Interestingly, hydrophilic PIG-P and lipophilic glimepiride, albeit differing completely in structure, have in common part of the signaling pathway that cross talks to the insulin signal transduction cascade and thereby negatively regulates HSL activity and lipolysis in adipocytes. The available data support the following model for their molecular mode of action. PIG(-P) bind as agonists/antagonists to the receptor, p115, a transmembrane protein that is anchored at the cell surface within DIGs. Upon binding, PIG(-P) displace endogenous GPI proteins from binding to p115 by competition with their GPI head groups and thereby reduce their concentration at DIGs. In contrast, glimepiride is not recognized by a proteinaceous receptor of the adipocyte plasma membrane but spontaneously inserts into DIGs via direct hydrophobic interactions in nonsaturatable fashion with GPI lipids thereby altering the structural organization of DIGs. Both events induce the redistribution of lipid-modified signaling protein, which are known to reside within DIGs on the basis of their acylation/glypiation and/or direct binding to caveolin. In particular, GPI proteins and palmitoylated nonreceptor tyrosine kinases of the src class, such as pp^{59}Lyn, translocate from DIGs of lower buoyant density (higher cholesterol content) to those of higher buoyant density (lower cholesterol content) accompanied by dissociation of pp^{59}Lyn from the DIG resident protein, caveolin. This relieves pp^{59}Lyn from inhibition by caveolin, interaction of which with various signaling proteins downregulates their signaling activity. In turn, activated pp^{59}Lyn phosphorylates IRS-1 at tyrosine residues, presumably at the same sites used by

the insulin receptor. In consequence, activation of the PI3K pathway leads to upregulation of PDE3B and downregulation of HSL activity via the same series of events engaged by insulin (see above). Very recent experiments revealed a possible explanation for the molecular link between the interaction of glimepiride with DIGs and the resulting reorganization and redistribution of structural (GPI proteins) and signaling (nonreceptor tyrosine kinases) components within the DIGs. The binding of glimepiride to GPI lipids triggers the activation of a DIG-associated GPI-specific phospholipase that cleaves GPI proteins and GPI lipids. PIG(-P) thereby generated may bind to the p115 at the DIGs and subsequently induce redistribution of signaling proteins from DIGs and the same downstream series of events, as recently demonstrated for exogenously added synthetic or natural PIG(-P). Alternatively, GPI cleavage may directly affect the structure of DIGs and thereby cause the redistribution of DIG components at both the outer (GPI proteins) and inner (nonreceptor tyrosine kinases) leaflets of the DIG membrane, accumulation of which critically depends on the presence of glycolipids. The latter possibility seems more likely since p115 is apparently not involved in mediating the antilipolytic activity of glimepiride. It seems plausible that the loss of long-chain FA from the outer leaflet of DIGs caused by redistribution of GPI proteins or cleavage of GPI lipids within DIGs simultaneously triggers the alteration of the FA composition of the DIG inner leaflet. This ''adjustment to a common FA milieu'' across the leaflets of DIGs may facilitate the redistribution of pp^{59}Lyn. Interestingly, the GPI-specific phospholipase stimulated by glimepiride in cultured and primary adipocytes is reportedly activated also by insulin in the presence of glucose.

4.11 Differences in Regulation of TAG Storage and Mobilization between Visceral and Subcutaneous Adipocytes

Individuals with peripheral obesity possess fat distribution subcutaneously in gluteofemoral areas and the lower part of the abdomen, and are at little risk of metabolic complications, such as NIDDM. Conversely, individuals with upper-body obesity accumulate fat in subcutaneous and visceral deposits and are more susceptible to metabolic problems, in particular when visceral fat deposits are abundant. Visceral fat deposits, located in the body cavity, are composed of the omental and mesenteric fat, and comprise the minor component of total body fat, representing 20% and 5–8% of total body fat in men and women, respectively. In upper-body obesity, fat excess is present not only in the visceral abdominal regions but also in the subcutaneous abdominal regions. There are several explanations but no clear proof why upper-body obesity is more at risk of developing metabolic disease than lower-body obesity. Since the visceral deposit is in direct contact with the liver through the portal circulation and considering the alterations of hormonal control of lipolysis of its adipocytes (insulin vs. catecholamines, see below), the ''portal paradigm'' was postulated on the basis that visceral adipocytes, through enhancement of lipolysis due to both reduced insulin-induced antilipolysis and enhanced catecholamine-induced lipolysis, will release portal NEFA that disturb liver metabolism. Chronic NEFA excess will in turn lead to glucose intolerance, hyperinsulinemia, insulin resistance, and dyslipidemia. However, experimental evidence is accumulating that nonvisceral upper-body fat rather than visceral fat

deposits is the predominant source of excess postprandial systemic NEFA availability in upper-body obese women and NIDDM patients. This raises the question as to whether the extent of visceral fat just represents another marker of ectopic fat distribution (i.e. fat present at the wrong place in tissues such as muscle, heart, liver, and pancreatic ß-cells). Alternatively, visceral adipocytes may possess properties that could be the origin of metabolic disturbances. In fact, striking adipose location–related differences have been reported in adipocyte responsiveness to insulin and catecholamines. Moreover, differences have been described in the characteristics of various biochemical pathways regulating TAG storage and mobilization between subcutaneous and visceral fat deposits.

In detail, adipocytes from visceral adipose tissue are more resistant to insulin-induced antilipolysis and reesterification of NEFA than those from leg and nonvisceral body fat both *in vitro* and *in vivo*. Various functional differences in these cells have been identified at the level of the insulin receptor and the postreceptor insulin-signaling cascade. PDE3B involved in lipolysis regulation by insulin (see above, Fig. 7) and protein tyrosine phosphatases dephosphorylating the insulin receptor, such as PTP1b, could be affected in differential fashion in visceral versus subcutaneous adipocytes. This concurs with the endogenous PTP1b activity found to be elevated in visceral adipose tissue and might contribute to the relative insulin resistance of this fat deposit. Exacerbating the impairment of insulin inhibition of lipolysis, the lipolytic response toward catecholamines is more pronounced in isolated adipocytes from visceral adipose tissue than from subcutaneous gluteal, femoral, and abdominal adipose tissue. This higher lipolytic activity can be explained by altered expression or function of HSL (increase) and/or proteins interacting with either HSL, such as ALBP (increase), or the LD, such as perilipin (decrease) (see above, Fig. 7). An enhanced α_2-adrenoceptor responsiveness associated with a concomitant decrease in ß-adrenoceptor responsiveness explains the lower lipolytic effect of catecholamines in gluteal and femoral adipocytes of normal and obese women and abdominal adipocytes of obese men compared with visceral adipocytes. Conversely, visceral adipocytes exhibit the highest $ß_{1,2}$-adrenoceptor-mediated and the weakest α_2-adrenoceptor-mediated lipolytic and antilipolytic responsiveness, respectively, to catecholamines, which seems to correlate with decreased expression of α_2-adrenoceptors and concomitantly increased expression of ß-adrenoceptors. In obese subjects, unrestrained lipolysis leads to excessive NEFA release from visceral hypertrophied adipocytes and may prevent their further enlargement in contrast to subcutaneous adipocytes, which actually represent the largest ones. The search for adipose tissue-specific local differences in the expression of genes that regulate the differentiation and expansion of adipose tissue as well as the regulation of TAG storage and mobilization is under way using genomic and proteomic approaches.

5 Up-/Downregulation of Components of TAG Storage and Mobilization

5.1 HSL

Recently, the functional significance of HSL in adipose tissue metabolism has

begun to be clarified in studies using HSL null mice. These mice showed normal growth rates and body weights. While the epididymal retroperitoneal and femoral WAT depots of HSL null mice remained unchanged, they displayed a 65% increase in brown adipose tissue mass compared with wild-type control mice. Inactivation of HSL resulted in the complete absence of neutral cholesterylester hydrolase activity in adipose tissue (both white and brown). However, TAG lipase activity in WAT was reduced by only 40% and TAG lipase activity in brown adipose tissue was similar to wild-type mice. Basal lipolysis, that is, glycerol release, was reduced in isolated adipose cells from HSL null mice in one study, but was unaffected and seemingly increased in another. Nonetheless, there was a marked defect or complete absence of catecholamine-stimulated glycerol release in adipose cells from HSL null mice, whereas catecholamine-stimulated NEFA release was still observed, but attenuated. This apparent discrepancy in the release of glycerol and NEFA from adipose cells of HSL null mice has been clarified by the observation that the basal and catecholamine-induced DAG content increased markedly in white and brown adipose tissue of HSL null mice. Therefore, the studies with HSL null mice appear to substantiate that HSL is the rate-limiting enzyme for DAG hydrolysis in adipose tissue and is essential for hormone-stimulated lipolysis. Remarkably, the absence of HSL was not associated with the development of obesity. However, adipose cells from HSL null mice, while displaying size heterogeneity, tended to be hypertrophic. Moreover, due to defective lipolysis during fasting, there was a reduction in circulating NEFA and a decreased hepatic production of VLDL secondary to the diminished release of NEFA from adipose tissue. This was associated with an induction of LPL in WAT, as well as in skeletal and cardiac muscle, but a decrease of LPL in brown adipose tissue. It remains to be determined whether operation of the TAG lipase in HSL null mice is a consequence of its compensatory upregulation in HSL-deficient adipocytes, only, or is indicative for expression of a second TAG-specific HSL-like lipase in wild-type adipocytes, too. Taken together, the findings on the lack of obesity and mild adipocyte hypertrophy observed in HSL null mice suggest that other lipases could also play a role in TAG mobilization.

Interestingly, in HSL null mice after prolonged fasting, plasma NEFA and TAG levels as well as hepatic TAG stores were reported to be significantly lower than with wild-type mice. This low hepatic TAG content was associated with improved hepatic insulin sensitivity since insulin caused a greater reduction in endogenous hepatic glucose production in HSL null mice (by 71%) than in wild-type mice (by 31%). This increase in hepatic insulin sensitivity was associated with elevated insulin receptor protein levels and activation of components of the insulin-signaling cascade, such as phosphorylation of PKB and activity of PI3K. The low hepatic TAG content in HSL null mice can be explained by their low plasma NEFA levels, since liver-specific NEFA uptake is commonly regarded to be a concentration-driven process facilitated by specific membrane transporters (see above). The inverse relationship between insulin sensitivity and hepatic TAG content may be explained by alterations in gene expression during activation of nuclear transcription factors such as peroxisome proliferator-activated receptors by intracellular TAG and/or fatty acyl derivatives. Furthermore, the diminished plasma NEFA levels *per se* cannot be excluded from being, at least partly,

responsible for the increased hepatic insulin sensitivity, since plasma NEFA concentrations are inversely correlated with insulin sensitivity (see above). Strikingly, in HSL null mice, no differences were observed in insulin-mediated whole-body glucose uptake compared with wild-type mice as revealed during hyperinsulinemic euglycemic clamp studies. The lack of improvement of whole-body insulin sensitivity may be explained by the reported absence of significant differences in muscle TAG content in combination with unchanged insulin-induced PKB phosphorylation in HSL null compared with wild-type mice despite lower plasma NEFA levels.

Recently, HSL null mice have been shown to have impaired insulin secretion, a 2- to 2.5-fold increase in islet TAG content, and elevated basal insulin secretion from isolated islets that fails to rise further upon challenge with glucose. Comparison of the gene expression profile between islets isolated from wild-type and HSL null mice using microarray chip technology and Taqman quantitative PCR for selected genes revealed changes in genes that are involved in lipid metabolism (methyl sterol oxidase, short chain acyl-CoA dehydrogenase), insulin response (insulin-induced growth response protein) and cytoskeleton (profilin, cofilin). In particular, the mRNAs for UCP-2, SREBP-1c, and PPAR-γ were upregulated in islets from HSL null mice as compared with wild-type mice. The transcriptional upregulation of UCP-2, which is regulated by SREBP-1c and PPAR-γ, was associated with elevated NEFA and correlated with impaired glucose-stimulated insulin secretion. Finally, C/EBPα, which is involved in cytokine-regulated apoptosis of pancreatic ß-cells, was also upregulated. Collectively, these findings suggest that HSL is important in maintaining lipid homeostasis in ß-cells by directly and indirectly controlling gene expressing for multiple metabolic and signaling pathways and hence in the regulation of glucose-stimulated insulin secretion.

A recent study using transgenic mice overexpressing HSL specifically in ß-cells provided novel and important insights with regard to the role of HSL in the development of lipotoxicity. These transgenic mice were characterized by impaired glucose and severely blunted glucose-stimulated insulin secretion upon challenge with a high-fat diet. Their islets displayed both elevated HSL activity and forskolin-induced lipolysis compared with wild-type islets, which resulted in significantly reduced TAG levels in transgenic compared with wild-type islets provided the mice have been fed a high-fat diet. Thus, the rate of influx of NEFA into TAG of the islet LD and the capacity to mobilize this TAG pool seem to determine the emergence of islet lipotoxicity. This is in accord with the recently reported inverse correlation between apoptosis and TAG accumulation in cultured ß-cells, suggesting a cytoprotective function of TAG in cytoplasmic LD formed by normal ß-cells against NEFA-induced islet dysfunction. Moreover, prolonged high-fat feeding of mice is accompanied by downregulation of the expression of islet HSL. Consequently, physiological downregulation of HSL or pharmacological inhibition of HSL and possibly of other as yet unknown TAG lipases in ß-cells could be interpreted in terms of the operation of a defense mechanism against the emergence of NEFA-induced islet cell dysfunction.

In conclusion, HSL degrading TAG of LD in islets seems to exert at least three different functions in the regulation of insulin secretion and its coupling to external stimuli. (1) In the short-term,

upon challenge of ß-cells with an insulin secretagogue, HSL is phosphorylated and activated via (glucagon-like peptide 1 and gastric inhibitory polypeptide) receptor/cAMP-dependent mechanisms lipolytically releasing NEFA and/or NEFA derivatives from islet TAG. These may act as acute stimulus (glucose)-secretion coupling factors for the exocytosis of insulin-containing secretory vesicles. (2) In the long-term diet-induced NIDDM and the increased influx of NEFA derived from plasma lipoproteins through the action of LPL and plasma NEFA into ß-cells lead to the accumulation of TAG in LD of ß-cells, which apparently exceeds the degradation capacity of the basal HSL. Thus, the low activity state of HSL under these conditions guarantees the storage of NEFA (derivatives), which potentially compromise the insulin secretory mechanism and induce ß-cell apoptosis. (3) Long-lasting aberrations in the regulation of lipolytic activity, such as overexpression of HSL, failure to downregulate HSL during prolonged high-fat feeding or the mass effect of TAG overstorage *per se* leading to incremental lipolytic release of NEFA even at low constitutive HSL activity, may result in exceeding a certain threshold level of NEFA flux. This triggers the lipotoxic pathway, presumably involving binding of NEFA (derivatives) as ligands to transcription factors, such as PPAR, and other mechanisms. The resulting alterations in the expression of genes regulating insulin secretory vesicle biogenesis and exocytosis finally lead to dampening of glucose-induced insulin secretion.

5.2 ALBP

In the absence of ALBP, intracellular concentrations of NEFA unbound to protein were increased in adipose cells, and basal as well as isoproterenol-stimulated lipolysis is decreased by 40%. These ALBP-deficient rodents showed a normal growth rate, body weight, and body composition compared with wild-type littermates due to functional compensation by the keratinocyte-type FABP. However, when exposed to a high-fat diet, ALBP-null mice developed diet-induced obesity, reaching a greater total weight gain than the control mice as a consequence of an increased fat pad weight. Both lean and obese ALBP-null mice displayed normoglycemia and normoinsulinemia, providing evidence for an uncoupling of obesity from insulin resistance through ALBP deficiency.

5.3 Perilipin

Recently, the amount of perilipin protein and lipolysis rates was analyzed in human subcutaneous adipocytes of nonobese and obese women. A polymorphism in intron 6 (rs891460 A/G) in the perilipin gene seems to be associated with a 50% decrease in perilipin expression and increased basal lipolysis (AA vs. GG carriers) in adipocytes of the obese women, which in subjects matched for body mass index and adipocyte volume was inversely correlated with the plasma concentrations of NEFA and glycerol and thus may contribute to the development of insulin resistance. Furthermore, they found two- to fourfold increases in noradrenaline-induced lipolysis. These data also reinforce the role of perilipin for the regulation of lipolysis in human fat cells.

The participation of perilipin in the regulation of TAG storage and mobilization has been addressed by knockout approaches from two different experimental groups.

Perilipin-deficient mice showed an increased metabolic rate together with an increased basal lipolytic rate in adipocytes, which confer resistance to diet-induced obesity. Nonetheless, they were severely impaired in their ability to respond to ß-adrenergic stimulation by an appropriate elevation in circulating NEFA. Although absence of perilipin resulted in leanness and reversal of obesity with an approximately 50% decrease in adipose mass and adipocyte size, which was not reflected in the lean body mass (that even increased slightly), it did not affect glucose tolerance, even with perilipin ablation being able to worsen peripheral glucose disposal. Perhaps most surprising is that mice lacking perilipin are more drastically affected in regard to regulation of FA mobilization than those deficient in HSL (see above). These observations could be explained if perilipin is not only required to maintain adipose cells in the quiescent state but is also involved in the functional lipolytic activation in adipose cells. To test this hypothesis experimentally, HSL translocation has been investigated in adipocytes differentiated *in vitro* from embryonic fibroblasts of wild-type and perilipin-null mice. Strikingly, they found that HSL fails to translocate to LD in cells lacking perilipin. Furthermore, introduction of normal perilipin A but not of an unstimulatable mutant variant (lacking the amino-terminal PKA phosphorylation sites) into CHO fibroblasts significantly stimulated the PKA-induced translocation of HSL. CHO cells expressing both perilipin and HSL were able to elicit a greater and more rapid response to a lipolytic stimulus than when either of these proteins was expressed separately. Thus, the perilipin-null mice emphasize the two major actions of perilipin: (1) to protect TAG against hydrolysis in the basal state and thus permit TAG storage; and (2) to facilitate HSL translocation to LD, leading to hydrolysis of TAG and release of FA in the stimulated state. Interestingly, the perilipin-mediated translocation process does not seem to depend on additional factors unique to adipocytes since it can be reconstituted by expressing both perilipin A and HSL in CHO fibroblasts. Moreover, phosphorylation of perilipin A at its three most amino-terminal PKA sites is apparently required to promote HSL translocation to LD. Recently, it was demonstrated that expression of wild-type perilipin alone, but not of a mutant lacking these PKA sites, in CHO cells increases the sensitivity of lipolysis toward ß-adrenergic stimulation. In contrast, LD coated with ADRP from cells that do not express perilipin (i.e. all with the exception of adipocytes and steroidogenic cells) neither facilitate HSL binding to the LD surface nor respond to elevated PKA activity nor are capable of restraining lipolysis to a minimum. The accelerated lipolytic activity exhibited by perilipin A-coated LD may result from increased accessibility of HSL (or some unknown neutral lipase) to TAG within LD when perilipin is phosphorylated at its three most amino-terminal PKA sites. Importantly, the binding of HSL to LD seems to be based on its direct interaction with the TAG substrate since translocation of HSL is blocked upon mutating the serine residue in its catalytic triad.

Two hypotheses have been proposed that reconcile the present data. Phosphorylated perilipin could directly recruit HSL to the LD surface by direct binding to the enzyme. However, pull-down and two-hybrid experiments have not provided experimental evidence of direct interaction between perilipin and HSL so far. Alternatively, phosphorylated perilipin may modify the

LD surface to indirectly facilitate interaction of HSL with the TAG core within the LD. HSL translocation and initiation of lipolysis occur within minutes. Initially, HSL and perilipin colocalize on LD but then, within a much longer period, perilipin slowly departs from large LD, and is subsequently recovered from dispersed, much smaller, LD (see above). Recent studies indicate that even after the departure of perilipin, HSL remains bound to the larger LD, which now lack perilipin, where it continues to cleave TAG. Taken together, the present data favor the indirect mechanism, although the two hypotheses are not mutually exclusive. In any case, the mechanism underlying the action of phosphorylated perilipin seems to modify the TAG of LD such that it is more effectively presented as a substrate, for example, by triggering dispersion of LD upon contact with phosphorylated perilipin leading to a larger total LD surface area and thereby facilitated access of HSL. It is now generally accepted that protein–phospholipid/cholesterol interactions can modify the shape of the plasma membrane and this mechanism seems to form the molecular basis for the biogenesis of caveolae (see above), for instance. The operation of similar mechanisms for neutral lipids seems to be more speculative.

5.4 PKA

PKA is a heterotrimer consisting of two regulatory cAMP-binding and two catalytic subunits. Four regulatory isoforms (RIα, RIβ, RIIα, and RIIβ) and two catalytic isoforms (Cα and Cβ) are expressed in the mouse, and each is encoded by a separate gene. The RIIβ subunit is expressed principally in three tissues known to regulate energy homeostasis: brown adipose tissue, WAT, and brain. It has been suggested that the induction of PKA in certain tissues may decrease obesity. For example, activation of the adipose-specific ß-AR, which signals via PKA, decreased obesity in both genetically obese (ob/ob) and diet-induced obese mice, suggesting that signaling mechanisms through this pathway are important in preventing obesity.

Studies in mice lacking RIIβ have revealed an unexpected role for this protein in regulating energy balance. These animals remained remarkably lean even when challenged with a high-fat diet and had increased metabolic activity, manifested by increases in body temperature, uncoupling protein-1 concentration, and lipolysis. Biochemical studies have shown that loss of RIIβ was compensated by increased expression of RIα, which is more sensitive to cAMP activation, resulting in a net increase in basal PKA activity, but markedly impaired cAMP-induced activation of HSL in mutant WAT, yet left cAMP-regulated gene expression relatively unperturbed. The disruption of lipolytic stimulation was seen both *in vitro* and *in vivo* and occurred equally for signaling from the ß$_3$-receptor-specific agonist CL316.243 and a nonspecific ß-agonist, isoproterenol. It is unlikely that the signaling defect in RIIβ mutant WAT was caused by decreased expression of either ß-AR or HSL, given that mRNA levels for these genes were unaffected. Conceivably, impairment of lipolytic hormonal response could be a consequence of the chronic stimulation of basal lipolysis seen in RIIβ null mice due to increased basal PKA activity. However, adipocytes from knockout mice deficient in the G-protein subunit, $G_i\alpha_2$, showed a threefold increase in basal cAMP levels and an elevated basal

rate of lipolysis but retained normal maximal response to ß-AR agonists. These studies suggest that increasing basal PKA activity in adipose tissue and brain ameliorates obesity.

In a recent study, wild-type and RIIβ null mice maintained on the C57BL/6 genetic background strain were fed a high-fat, high-carbohydrate diet, which is known to induce obesity and diabetes in C57BL/6 mice. RIIβ null mice were resistant to weight gain and hyperinsulinemia. *In vivo* insulin sensitivity and glucose disposal were dramatically improved in the RIIβ null mice, as were plasma lipid profiles. When mice were corrected for differences in body weight, improved insulin-mediated glucose disposal was still observed in the RIIβ null mice, suggesting an obesity-independent effect of RIIβ on promoting insulin resistance. This suggests that PKA activity in both adipose tissue and brain is important for determining body composition, food intake, and diabetogenic parameters.

It has been speculated that at least some of the elements mediating lipolytic stimulation (e.g. ß-AR, AC, PKA, HSL) be colocalized within adipocytes to facilitate efficient signal transduction, and that lipolytic stimulation is impaired in RIIβ-deficient rodents because RIIβ participates specifically in the formation of this complex. Previous studies suggest a compartmentalized apparatus that mediates lipolytic stimulation as follows: (1) At any given intracellular cAMP concentration, the lipolytic response toward catecholamines is greater than toward forskolin, a nonspecific AC activator. Hence, low concentrations of isoproterenol ($\sim$10 nM) can stimulate lipolysis without measurably altering overall cAMP levels, whereas low concentrations of forskolin (0.1–1.0 μM) increase intracellular cAMP without affecting lipolysis. (2) The concentration of isoproterenol or ß$_3$-specific agonists required for half-maximal activation of AC activity in adipocyte membranes is 80-fold greater than the concentration required to activate lipolysis in intact adipocytes. (3) A signaling complex including ß$_2$-AR, PKA, and phosphatases has been isolated that appears to be assembled by the scaffold protein, gravin. In summary, catecholamines stimulate lipolysis more potently than they increase overall intracellular cAMP, suggesting a preferential association of ß-AR, PKA, and perhaps its substrate, HSL, in a ''lipolytic'' complex.

5.5 ASP

Since mice lacking complement C3 cannot produce ASP, C3 gene knockout mice are consequently ASP-null animals. ASP null mice show a reduced WAT weight, distributed evenly through all depots in females while primarily affecting gonadal and perirenal localizations in males, as well as reduced body fat. In addition, male mice have delayed TAG clearance. The reduced adipose mass was accompanied by hypoleptinemia with a modest increase in food intake. When fed on either a low-fat or high-fat diet, male ASP null mice showed no differences in body weight compared with wild-type animals. Female ASP null mice, however, had a decreased body weight compared with their wild-type counterparts. The relatively mild phenotype of these mice suggests a minor role of ASP in TAG storage and mobilization control or the existence of compensatory mechanisms rescuing part of the physiological function of ASP.

Very recently, the effect of ASP deficiency has been examined in ob/ob mice, which are leptin-deficient. The ob/ob mice have been used to test for protection from obesity in several double-knockout models. In most cases, the decrease in body weight, which was evident in the single knockout, was enhanced when examined on the background of the ob/ob obese mouse model. Compared with age-matched ob/ob mice, ob/ob C3 null mice had delayed postprandial TAG and FA clearance, associated with decreased body weight and according to the homeostasis model assessment index increased insulin sensitivity. By contrast, food intake in the double-knockout mice was slightly elevated compared with ob/ob mice. The hyperphagia/leanness was balanced by a significant increase in energy expenditure and oxygen consumption. These results suggest that the ASP regulation of energy storage may influence energy expenditure and dynamic metabolic balance.

Since ASP increases TAG synthesis and reesterification and decreases intracellular TAG mobilization, such as norepinephrine-stimulated NEFA release from fat cells as well, the effect of ASP, at least in adipose tissue, is to effectively decrease substrate cycling and increase TAG storage. ASP deficiency could possibly release the brake on this cycling process, allowing for increased substrate cycling and augmenting energy expenditure. This pattern is different to that observed for insulin, which had marked effects on both lipolysis and fractional NEFA reesterification. When fat cells were incubated with maximally effective concentrations of each protein, ASP inhibited NEFA release during lipolysis to a lesser degree than insulin, whereas it stimulated fractional NEFA reesterification to the same, or even greater, extent as insulin.

5.6 Caveolin

Caveolin null mice shed additional light onto the proposed new roles of caveolin and caveolae in TAG storage and mobilization (see above). Caveolin-1 null mice were found to store less TAG in WAT than control mice. This defect led to resistance to diet-induced obesity and to increased blood concentrations of NEFA and TAG. Because caveolin can be recovered with LD, at least under certain experimental conditions (see above), it is tempting to speculate that the lack of caveolin interferes with normal LD biogenesis or enables HSL to gain access to TAG in the LD core. Alternatively, in the absence of caveolin-1 adipocytes lack functional caveolae and this could lead to diminished FA uptake. As the flux of FA into primary adipocytes and 3T3-L1 adipocytes follows saturatable kinetics, facilitated membrane transport has been proposed as the uptake mechanism. Interestingly, labeling of membrane proteins with photoreactive long-chain FA identified caveolin-1 as the major FA-binding protein in adipocytes. Together with the demonstrated association of the caveolins with LD (see above), it is possible that caveolae and the caveolins act as portals for the uptake and transport of FA to LD. However, it cannot be excluded that the negative effect of missing caveolin-1 on TAG storage is secondary to impairment of insulin signaling to the glucose transport system and the resulting reduction of glucose uptake and FA esterification. Caveolae and lipid rafts (DIGs) act as scaffolding platform for the so-called Cbl-CAP pathway that is required for insulin stimulation of glucose transport in addition to the IRS-PI3K pathway.

See also Intracellular Fatty Acid Binding Proteins and Fatty Acid Transport; Intracellular Fatty Acid Binding Proteins in Metabolic Regulation; Lipid and Lipoprotein Metabolism; Metabolic Basis of Cellular Energy.

Bibliography

Books and Reviews

Cianflone, K., Maslowska, M., Sniderman, A.D. (1999) Acylation stimulating protein (ASP), an adipocyte autocrine: new directions, *Semin. Cell. Dev. Biol.* **10**, 31–41.

Frühbeck, G., Gomez-Ambrosi, J., Muruzabal, F.J., Burrell, M.A. (2001) The adipocyte: a model for integration of endocrine and metabolic signaling in energy metabolism regulation, *Am. J. Physiol. Endocrinol. Metab.* **280**, 827–847.

Gibbons, G.F., Islam, K., Pease, R.J. (2000) Mobilisation of triacylglycerol stores, *Biochim. Biophys. Acta* **1483**, 37–57.

Holm, C., Osterlund, T., Laurell, H., Contreras, J.A. (2000) Molecular mechanisms regulating hormone-sensitive lipase and lipolysis, *Annu. Rev. Nutr.* **20**, 365–393.

Jensen, M.D. (1997) Lipolysis: contribution from regional fat, *Annu. Rev. Nutr.* **17**, 127–139.

Lehner, R., Kuksis, A. (1996) Biosynthesis of triacylglycerols, *Prog. Lipid Res.* **35**, 169–201.

McGarry, J.D. (2002) Banting Lecture 2001: dysregulation of fatty acid metabolism in the etiology of type 2 diabetes, *Diabetes* **51**, 7–18.

Murphy, D.J., Vance, J. (1999) Mechanism of lipid-body formation, *Trends Biochem. Sci.* **24**, 109–115.

Shafrir, E., Raz, I. (2002) Diabetes: mellitus or lipidus, *Diabetologia* **46**, 433–440.

Su, H.S., Wang, D. (1998) Nutritional and hormonal regulation of enzymes in fat synthesis: studies on fatty acid synthase and mitochondrial glycerol-3-phosphate acyltransferase gene transcription, *Annu. Rev. Nutr.* **18**, 331–351.

Unger, R.H. (2002) Lipotoxic diseases, *Annu. Rev. Med.* **53**, 319–336.

Zilversmit, D.B. (1979) Atherogenesis: a postprandial phenomenon, *Circulation* **60**, 473–485.

Primary Literature

Abu-Elheiga, L., Matzuk, M.M., Abo-Hashema, K.A., Wakil, S.J. (2001) Continuous fatty acid oxidation and reduced fat storage in mice lacking acetyl-CoA carboxylase 2, *Science* **291**, 2613–2616.

Arner, P., Goldfinger, M., Marcus, C., Engfeldt, P. (1991) Effects of lipolytic and antilipolytic agents on glucose transport in human fat cells, *Int. J. Obes.* **15**, 327–331.

Arthur, M.J., Atshaves, B.P., Frolov, A., Foxworth, W.D., Kier, A.B., Schroeder, F. (1999) Cellular uptake and intracellular trafficking of long-chain fatty acids, *J. Lipid Res.* **40**, 1371–1383.

Berk, P.D., Zhou, S.L., Kiang, C.L., Stump, D., Bradbury, M., Isola, L.M. (1997) Uptake of long-chain free fatty acids is selectively upregulated in adipocytes of Zucker rats with genetic obesity and noninsulin-dependent diabetes mellitus, *J. Biol. Chem.* **272**, 8830–8835.

Botion, L.M., Green, A. (1999) Long-term regulation of lipolysis and hormone-sensitive lipase by insulin and glucose, *Diabetes* **48**, 1691–1697.

Cases, S., Smith, S.J., Zheng, Y.W. (1998) Identification of a gene encoding an acyl-CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13018–13023.

Christie, A.W., McCormick, D.K.T., Emmison, N., Kraemer, F.B., Alberti, K.G.M.M., Yeaman, S.J. (1996) Mechanism of antilipolytic action of acipimox in isolated rat adipocytes, *Diabetologia* **39**, 45–53.

Cianflone, K., Maslowska, M. (1995) Differentiation-induced production of ASP in human adipocytes, *Eur. J. Clin. Invest.* **25**, 817–825.

Cianflone, K., Roncari, D.A.K., Maslowska, M., Baldo, A., Forden, J., Sniderman, A.D. (1994) Adipsin/acylation-stimulating protein system in human adipocytes: regulation of triacylglycerol synthesis, *Biochemistry* **33**, 9489–9495.

Clifford, G.M., Londos, C., Kraemer, F.B., Vernon, R.G., Yeaman, S.J. (2000) Translocation

of hormone-sensitive lipase and perilipin upon lipolytic stimulation of rat adipocytes, *J. Biol. Chem.* **275**, 5011–5015.

Cnop, M., Hannaert, J.C., Hoorens, A., Eizirik, D.L., Pipeleers, D.G. (2001) Inverse relation between cytotoxicity of free fatty acids in pancreatic islet cells and cellular triglyceride accumulation, *Diabetes* **50**, 1771–1777.

Coe, N.R., Simpson, M.A., Bernlohr, D.A. (1999) Targeted disruption of the adipocyte lipid-binding protein (aP2 protein) gene impairs fat cell lipolysis and increases cellular fatty acid levels, *J. Lipid Res.* **40**, 967–972.

Crepaldi, G., Avogaro, P., Descovich, G.C. (1988) Plasma lipid-lowering activity of acipimox in patients with type II and type IV hyperlipoproteinaemia: a double blind trial, *Atherosclerosis* **70**, 115–121.

Cummings, D.E., Brandon, E.P., Planas, J.V., Motamed, K., Idzerda, R.L., McKnight, G.S. (1996) Genetically lean mice result from targeted disruption of the RII beta subunit of protein kinase A, *Nature* **382**, 622–626.

Degerman, E., Smith, C.J., Tornqvist, H., Vasta, V., Manganiello, V., Belfrage, P. (1990) Evidence that insulin and isoprenaline activate the cGMP-inhibited low-Km cAMP phosphodiesterase in rat fat cells by phosphorylation, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 533–537.

Dugail, I., Le Lay, S., Varret, M., Le Liepvre, X., Dagher, G., Ferr, P. (2003) New insights into how adipocytes sense their triglyceride stores. Is cholesterol a signal, *Horm. MeTab. Res.* **35**, 204–210.

Edens, N.K., Hirsch, J., Hirsch, J.J. (1990) Mechanism of free fatty acid reesterification in human adipocytes in vitro, *Lipid Res.* **31**, 1423–1431.

Efendic, S., Alm, B., Löw, H. (1970) Effects of Ca^{2+} on lipolysis in human omental adipose tissue in vitro, *Horm. MeTab. Res.* **2**, 287–291.

Egan, J.J., Greenberg, A.S., Chang, M.K., Wek, S.A., Moos, M.C., Londos, C. (1992) Mechanism of hormone-stimulated lipolysis in adipocytes: translocation of hormone-sensitive lipase to the lipid storage droplet, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8537–8541.

Elks, M.L., Manganiello, V.C. (1985) Antilipolytic action of insulin: role of cAMP phosphodiesterase activation, *Endocrinology* **116**, 2119–2221.

Eriksson, H., Tornqvist, H. (1997) Specific inhibition of the cGMP-inhibited cAMP phosphodiesterase blocks the insulin-like antilipolytic effect of growth hormone in rat adipocytes, *Mol. Cell. Biochem.* **169**, 37–42.

Eriksson, H., Ridderstrale, M., Degerman, E., Ekholm, D., Smith, C.J., Manganiello, V., Belfrage, P., Tornqvist, H. (1995) Evidence for the key role of the adipocyte cGMP-inhibited cAMP phosphodiesterase in the antilipolytic action of insulin, *Biochim. Biophys. Acta* **1266**, 101–104.

Fain, J.N., Shepherd, R.E. (1975) Free fatty acids as feedback regulators of adenylate cyclase and cyclic 3 : 5-AMP accumulation in rat fat cells, *J. Biol. Chem.* **250**, 6586–6592.

Febbraio, M., Abumrad, N.A., Hajjar, D.P. (1999) A null mutation in CD36 reveals an important role in fatty acid and lipoprotein metabolism, *J. Biol. Chem.* **274**, 19055–19062.

Frayn, K.N. (1998) Regulation of fatty acid delivery in vivo, *Adv. Exp. Med. Biol.* **441**, 171–179.

Fredrikson, G., Tornqvist, H., Belfrage, P. (1986) Hormone-sensitive lipase and monoacylglycerol lipase are both required for complete degradation of adipocyte triacylglycerol, *Biochim. Biophys. Acta* **876**, 288–293.

Fulcher, G.R., Walker, M., Catalano, C., Agius, I., Alberti, K.G.M.M. (1992) Metabolic effects of suppression of non-esterified fatty acid levels with acipimox in obese NIDDM subjects, *Diabetes* **41**, 1400–1408.

Garg, A., Grundy, S.M. (1990) Nicotinic acid therapy for dyslipidemia in noninsulin-dependent diabetes mellitus, *JAMA* **264**, 723–726.

Gasic, S., Tian, B., Green, A. (1999) Tumor necrosis factor alpha stimulates lipolysis in adipocytes by decreasing Gi protein concentrations, *J. Biol. Chem.* **274**, 6770–6775.

Germinario, R., Sniderman, A.D., Manuel, S., Lefebre, S.P., Baldo, A., Cianflone, K. (1993) Coordinate regulation of triacylglycerol synthesis and glucose transport by acylation-stimulating protein, *Metabolism* **42**, 574–580.

Green, A. (1987) Adenosine receptor downregulation and insulin resistance following prolonged incubation of adipocytes with an A1 adenosine receptor agonist, *J. Biol. Chem.* **262**, 15702–15707.

Green, A., Johnson, J.L., Milligan, G. (1990) Downregulation of Gi subtypes by prolonged incubation of adipocytes with an A1

adenosine receptor agonist, *J. Biol. Chem.* **265**, 5206–5210.

Green, A., Dobias, S.B., Walters, D.J., Brassier, A.R. (1994) Tumor necrosis factor increases the rate of lipolysis in primary cultures of adipocytes without altering levels of hormone-sensitive lipase, *Endocrinology* **134**, 2581–2588.

Green, A., Milligan, G., Dobias, S.B. (1992) Gi downregulation as a mechanism for heterologous desensitization in adipocytes, *J. Biol. Chem.* **267**, 3223–3229.

Guo, Z. (1999) Regional postprandial fatty acid metabolism in different obesity phenotypes, *Diabetes* **48**, 1586–1592.

Haemmerle, G., Zimmermann, R., Hayn, M., Theussl, C., Waeg, G., Wagner, E., Sattler, W., Magin, T.M., Wagner, E.F., Zechner, R. (2002) Hormone-sensitive lipase deficiency in mice causes diglyceride accumulation in adipose tissue, muscle, and testis, *J. Biol. Chem.* **277**, 4806–4815.

Hagström-Toft, E., Bolinder, J., Eriksson, S., Arner, P. (1995) Role of phosphodiesterase III in the antilipolytic effect of insulin in vivo, *Diabetes* **44**, 1170–1175.

Hauner, H., Petruschka, T., Russ, M., Röhrig, K., Eckel, J. (1995) Effects of tumour necrosis factor alpha (TNF alpha) on glucose transport and lipid metabolism of newly-differentiated human fat cells in cell culture, *Diabetologia* **38**, 764–771.

Hermoso, J., Pignol, D., Penel, S. (1997) Neutron crystallography evidence of lipase-colipase complex activation by a micelle, *EMBO J.* **16**, 5531–5536.

Hirsch, H.A., Rosen, O.M. (1984) Lipolytic stimulation modulates the subcellular distribution of hormone-sensitive lipase in 3T3-L1 cells, *J. Lipid Res.* **25**, 665–677.

Hopp, J.F., Palmer, W.K. (1990) Effect of electrical stimulation on intracellular triacylglycerol in isolated skeletal muscle, *J. Appl. Physiol.* **68**, 348–354.

Hotamisligil, G., Shargill, N., Spiegelman, B. (1993) Adipose expression of tumor necrosis factor alpha: direct role in obesity-linked insulin resistance, *Science* **259**, 87–91.

Hotamisligil, G., Arner, P., Caro, J.F., Atkinson, R.L., Spiegelman, B.M. (1995) Increased adipose tissue expression of tumor necrosis factor alpha in human obesity and insulin resistance, *J. Clin. Invest.* **95**, 2409–2415.

Ibrahimi, A., Bonen, A., Blinn, W.D. (1999) Muscle-specific overexpression of FAT/CD36 enhances fatty acid oxidation by contracting muscle, reduces plasma triglycerides and fatty acids, and increases plasma glucose and insulin, *J. Biol. Chem.* **274**, 26761–26766.

Jepson, C.A., Yeaman, S.J. (1992) Inhibition of hormone-sensitive lipase by intermediary lipid metabolites, *FEBS Lett.* **310**, 197–200.

Kang, E.S., Betts, D., Fain, J.N., Bahouth, S.W., Myers, L.K. (1993) Chronic exposure of rat fat cells to insulin enhances lipolysis and activation of partially purified hormone-sensitive lipase, *Diabetes* **42**, 1415–1424.

Kishida, K., Kuriyama, H., Funahashi, T. (2000) Aquaporin adipose, a putative glycerol channel in adipocytes, *J. Biol. Chem.* **275**, 20896–20902.

Kraemer, F.B., Shen, W.-J. (2002) Hormone-sensitive lipase: control of intracellular tri-(di)acylglycerol and cholesteryl ester hydrolysis, *J. Lipid Res.* **43**, 1585–1594.

Kraemer, F.B., Patel, S., Saedi, M.S., Sztalryd, C. (1993) Detection of hormone-sensitive lipase in various tissues. I. Expression of an HSL/bacterial fusion protein and generation of anti-HSL antibodies, *J. Lipid Res.* **34**, 663–671.

Landau, C., Chen, Y.D., Skowronski, R., Hollenbeck, C.B., Japan, J.B., Reaven, G.M. (1992) Effect of nicotinic acid on plasma glucose concentration in normal individuals, *Horm. MeTab. Res.* **24**, 424–428.

Langfort, J., Ploug, T., Ihlemann, J., Holm, C., Galbo, H. (2000) Stimulation of hormone-sensitive lipase activity by contractions in rat skeletal muscle, *Biochem. J.* **351**, 207–214.

Langfort, J., Ploug, T., Ihlemann, J., Saldo, M., Holm, C., Galbo, H. (1999) Expression of hormone-sensitive lipase and its regulation by adrenaline in skeletal muscle, *Biochem. J.* **340**, 459–465.

Langfort, J., Ploug, T., Ihlemann, J., Enevoldsen, J.H., Stallknecht, B., Saldo, M., Kjaer, M., Holm, C., Galbo, H. (1998) Hormone-sensitive lipase (HSL) expression and regulation in skeletal muscle, *Adv. Exp. Med. Biol.* **441**, 219–228.

Large, V., Arner, P., Reynisdottir, S., Grober, J., Harmelen van, V., Holm, C., Langin, D. (1998) Comparison of hormone-sensitive lipase expression and activity in relation to lipolysis in human fat cells, *J. Lipid Res.* **39**, 1688–1695.

Leibel, R.L., Hirsch, J. (1985) A radioisotopic technique for analysis of free fatty acid

reesterification in human adipose tissue, *Am. J. Physiol.* **248**, 140–147.

Listenberger, L.L., Han, X., Lewis, S.E., Cases, S., Farese, R.V., Ory, D.S., Schaffer, J.E. (2003) Triglyceride accumulation protects against fatty acid–induced lipotoxicity, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3077–3082.

Lorenzen, A., Stannek, C., Lang, H., Andrianov, V., Kalvinsh, I., Schwabe, U. (2001) Characterization of a G-protein-coupled receptor for nicotinic acid, *Mol. Pharmacol.* **59**, 349–357.

Man, M.Z., Hui, T.Y., Schaffer, J.E., Lodish, H.F., Bernlohr, D.A. (1996) Regulation of the murine adipocyte fatty acid transporter gene by insulin, *Mol. Endocrinol.* **10**, 1021–1028.

Maslowska, M., Vu, H., Phelis, S., Sniderman, A.D., Rhodes, B., Blank, D., Cianflone, K. (1999) Plasma acylation-stimulating protein, adipsin and lipids in non-obese and obese populations, *Eur. J. Clin. Invest.* **29**, 679–686.

McKnight, G.S., Cummings, D.E., Amieux, P.S., Sikorski, M.A., Brandon, E.P., Planas, J.V., Motamed, K., Idzerda, R.L. (1998) Cyclic AMP, PKA, and the physiological regulation of adiposity, *Rec. Prog. Horm. Res.* **53**, 139–161.

Mottagui-Tabar, S., Ryden, M., Löfgren, P., Faulds, G., Hoffstedt, J., Brookes, A.J., Andersson, I., Arner, P. (2003) Evidence for an important role of perilipin in the regulation of human adipocyte lipolysis, *Diabetologia* **46**, 789–797.

Müller, G., Dearey, E.A., Pünter, J. (1993) The sulphonylurea drug, glimepiride, stimulates release of glycosyl-phosphatidylinositol-anchored plasma membrane proteins from 3T3 adipocytes, *Biochem. J.* **289**, 509–521.

Müller, G., Ertl, J., Gerl, M., Preibisch, G. (1997) Leptin impairs metabolic actions of insulin in isolated rat adipocytes, *J. Biol. Chem.* **272**, 10585–10593.

Müller, G., Hanekop, N., Wied, S., Frick, W. (2002) Cholesterol depletion blocks redistribution of lipid raft components and insulin-mimetic signaling by glimepiride and phosphoinositolglycans in rat adipocytes, *Mol. Med.* **8**, 120–136.

Müller, G., Jung, C., Frick, W., Bandlow, W., Kramer, W. (2002) Interaction of phosphoinositolglycan(-peptides) with plasma membrane lipid rafts triggers insulin-mimetic signaling in rat adipocytes, *Arch. Biochem. Biophys.* **408**, 7–16.

Müller, G., Wied, S., Crecelius, A., Kessler, A., Eckel, J. (1997) Phosphoinositolglycan-peptides from yeast potently induce metabolic insulin actions in isolated rat adipocytes, cardiomyocytes and diaphragm, *Endocrinology* **138**, 3459–3475.

Müller, G., Wied, S., Wetekam, E.-M., Crecelius, A., Unkelbach, A., Pünter, J. (1994) Stimulation of glucose utilization in 3T3 adipocytes and rat diaphragm in vitro by the sulfonylurea, glimepiride and glibenclamide, is correlated with modulations of the cAMP regulatory cascade, *Biochem. Pharmacol.* **48**, 985–996.

Murray, I., Havel, P.J., Sniderman, A.D., Cianflone, K. (2000) Reduced body weight, adipose tissue, and leptin levels despite increased energy intake in female mice lacking acylation-stimulating protein, *Endocrinology* **141**, 1041–1049.

Nurjhan, N., Consoli, A., Gerich, J. (1992) Increased lipolysis and its consequences on gluconeogenesis in noninsulin-dependent diabetes mellitus, *J. Clin. Invest.* **89**, 169–175.

Okada, T., Kawano, Y., Sakakibara, T., Hazeki, O., Ui, M. (1994) Essential role of PI 3-kinase in insulin-induced glucose transport and antilipolysis in rat adipocytes, *J. Biol. Chem.* **269**, 3568–3573.

Osuga, J.-I., Ishibashi, S., Oka, T., Yagyu, H., Tozawa, R., Fujimoto, A., Shionoira, F., Yahagi, N., Kraemer, F.B., Tsutsumi, O., Yamada, N. (2000) Targeted disruption of hormone-sensitive lipase results in male sterility and adipocyte hypertrophy, but not in obesity, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 787–792.

Pan, D.A., Lillioja, S., Kriketos, A.D., Milner, M.R., Baur, L.A., Bogardus, C., Jenkins, A.B., Storlien, L.H. (1997) Skeletal muscle triglyceride levels are inversely related to insulin action, *Diabetes* **46**, 983–988.

Prentki, M., Vischer, S., Glennon, M.C., Regazzi, R., Deeney, J.T., Corkey, B.E. (1992) Malonyl-CoA and long-chain acyl-CoA esters as metabolic coupling factors in nutrient-induced insulin secretion, *J. Biol. Chem.* **267**, 5802–5810.

Puhakainen, I., Koivisto, V.A., Yki-Jarvinen, H. (1992) Lipolysis and gluconeogenesis from glycerol are increased in patients with noninsulin-dependent diabetes mellitus, *J. Clin. Endocrinol. MeTab.* **75**, 789–794.

Rahn, T., Ridderstrale, M., Tornqvist, H., Manganiello, V., Fredrikson, G., Belfrage, P., Degerman, E. (1994) Essential role of phosphatidylinositol 3-kinase in insulin-induced activation and phosphorylation of the cGMP-inhibited cAMP phosphodiesterase in rat adipocytes studies using the selective inhibitor wortmannin, *FEBS Lett.* **350**, 314–317.

Rahn-Landstrom, T., Mei, J., Karlsson, M., Manganiello, V., Degerman, E. (2000) Downregulation of cyclic-nucleotide phosphodiesterase 3B in 3T3-L1 adipocytes induced by tumour necrosis factor a and cAMP, *Biochem. J.* **346**, 337–343.

Randle, P.J., Garland, P.B., Hales, C.N., Newsholme, E.A. (1963) The glucose fatty acid cycle: its role in insulin sensitivity and the metabolic disturbances of diabetes mellitus, *Lancet* **1**, 785–789.

Ray, H., Beylot, M., Arner, P., Larrouy, D., Langin, D., Holm, C., Large, V. (2003) The presence of a catalytically inactive form of hormone-sensitive lipase is associated with decreased lipolysis in abdominal subcutaneous adipose tissue of obese subjects, *Diabetes* **52**, 1417–1422.

Razani, B. (2002) Caveolin-1 deficient mice are lean, resistant to diet-induced obesity, and show hypertriglyceridemia with adipocyte abnormalities, *J. Biol. Chem.* **277**, 8635–8647.

Rebrin, K., Steil, G.M., Mittelman, S.D., Bergman, R.N. (1996) Causal linkage between insulin suppression of lipolysis and suppression of liver glucose output in dogs, *J. Clin. Invest.* **98**, 741–749.

Reue, K., Doolittle, M.H. (1996) Naturally occurring mutations in mice affecting lipid transport and metabolism, *J. Lipid Res.* **37**, 1387–1405.

Roduit, R., Masiello, P., Wang, S.P., Li, H., Mitchell, G.A., Prentki, M. (2001) A role for hormone-sensitive lipase in glucose-stimulated insulin secretion: a study in hormone-sensitive lipase deficient mice, *Diabetes* **50**, 1970–1975.

Ryden, M., Dicker, A., Harmelen van, V. (2002) Mapping of early signaling events in tumor necrosis factor-a-mediated lipolysis in human fat cells, *J. Biol. Chem.* **277**, 1085–1091.

Schreyer, S.A., Cummings, D.E., McKnight, G.S., LeBoeuf, R.C. (2001) Mutations of the RIIbeta subunit of protein kinase A prevents diet-induced insulin resistance and dyslipidemia in mice, *Diabetes* **50**, 2555–2562.

Shen, W.J., Sridhar, K., Bernlohr, D.A., Kraemer, F.B. (1999) Interaction of rat hormone-sensitive lipase with adipocyte lipid-binding protein, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5528–5532.

Shen, W.-J., Liang, Y., Hong, R., Patel, R., Natu, V.K., Sridhar, K., Jenkins, A., Bernlohr, D.A., Kraemer, F.B. (2001) Characterization of the functional interaction of adipocyte lipid-binding protein with hormone-sensitive lipase, *J. Biol. Chem.* **276**, 49443–49448.

Shimabukuro, M., Zhou, Y.T., Levi, M., Unger, R.H. (1998) Fatty acid-induced beta cell apoptosis: a link between obesity and diabetes, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2498–2502.

Smith, S.J., Cases, S., Jensen, D.R. (2000) Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking Dgat, *Nat. Genet.* **25**, 87–90.

Smith, C.J., Vasta, V., Degerman, E., Belfrage, P., Manganiello, V. (1991) Hormone-sensitive cyclic GMP-inhibited cyclic AMP phosphodiesterase in rat adipocytes. Regulation of insulin- and cAMP-dependent activation by phosphorylation, *J. Biol. Chem.* **266**, 13385–13390.

Smith, F., Rouet, P., Lucas, S., Mairal, A., Sengenes, C., Lafontan, M., Vaulont, S., Casado, M., Langin, D. (2002) Transcriptional regulation of adipocyte hormone-sensitive lipase by glucose, *Diabetes* **51**, 293–300.

Souza, S.C., Muliro, K.V., Liscum, L. (2002) Modulation of hormone-sensitive lipase and protein kinase A-mediated lipolysis by perilipin A in an adenoviral reconstituted system, *J. Biol. Chem.* **277**, 8267–8272.

Souza, S.C., Yamamoto, M.T., Franciosa, M.D., Lien, P., Greenberg, A.S. (1998) BRL 49653 blocks the lipolytic actions of tumor necrosis factor alpha: a potential new insulin-sensitizing mechanism for thiazolidinediones, *Diabetes* **47**, 691–695.

Souza, S.C., Vargas De, L.M., Yamamoto, M.T., Lien, P., Franciosa, M.D., Moss, L.G., Greenberg, A.S. (1998) Overexpression of perilipin A and B blocks the ability of tumor necrosis factor alpha to increase lipolysis in 3T3-L1 adipocytes, *J. Biol. Chem.* **273**, 24665–24669.

Stremmel, W., Strohmeyer, G., Borchard, F., Kochwa, S., Berk, P.D. (1985) Isolation and partial characterization of a fatty acid-binding protein in rat liver plasma membranes, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4–8.

Su, C.L., Sztalryd, C., Contreras, J.A., Holm, C., Kimmel, A.R., Londos, C. (2003) Mutational analysis of the hormone-sensitive lipase translocation reaction in adipocytes, *J. Biol. Chem.* **278**, 43615–43619.

Synak, M., Gorecka, M., Langfort, J., Smol, E., Zernicka, E. (2003) Palmitic acid incorporation into intramyocellular acylglycerols depends on both total and unbound to albumin palmitic acid concentration, *Biochem. Cell. Biol.* **81**, 35–41.

Tansey, J.T., Huml, A.M., Vogt, R. (2003) Functional studies on native and mutated forms of perilipins, *J. Biol. Chem.* **278**, 8401–8408.

Tansey, J.T., Sztalryd, C., Gruia Gray, J. (2001) Perilipin ablation results in a lean mouse with aberrant adipocyte lipolysis, enhanced leptin production, and resistance to diet-induced obesity, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6494–6499.

Tao, Y.Z., Cianflone, K., Sniderman, A.D., Colby-Germinario, S.P., Germinario, R.J. (1997) Acylation-stimulating protein (ASP) regulates glucose transport in the rat L6 muscle cell line, *Biochim. Biophys. Acta* **1344**, 221–229.

Tavintharan, S., Kashyap, M.L. (2001) The benefits of niacin in atherosclerosis, *Curr. Atheroscler. Rep.* **3**, 74–82.

Tornvall, P., Walldius, G.W. (1991) A comparison between nicotinic acid and acipimox in hypertriglyceridaemia–effects on serum lipids, lipoproteins, glucose tolerance and tolerability, *J. Int. Med.* **230**, 415–421.

Uysal, K.T., Scheja, L., Wiesbrock, S.M., Bonner-Weir, S., Hotamisligil, G.S. (2000) Improved glucose and lipid metabolism in genetically obese mice lacking aP2, *Endocrinology* **141**, 3388–3396.

Wang, S.P., Laurin, N., Himms-Hagen, J., Rudnicki, M.A., Levy, E., Robert, M.-F., Pan, L., Oligny, L., Mitchell, G.A. (2001) The adipose tissue phenotype of hormone-sensitive lipase deficiency in mice, *Obes. Res.* **9**, 119–128.

Wang, M.Y., Lee, Y., Unger, R.H. (1999) Novel form of lipolysis induced by leptin, *J. Biol. Chem.* **274**, 17541–17544.

Winzell, M.S., Svensson, H., Enerbäck, S., Ravnskjaer, K., Mandrup, S., Esser, V., Arner, P., Alves-Guerra, M.-C., Miroux, B., Sundler, F., Ahren, B., Holm, C. (2003) Pancreatic b-cell lipotoxicity induced by overexpression of hormone-sensitive lipase, *Diabetes* **52**, 2057–2065.

Wise, A., Foord, S.M., Fraser, N.J., Barnes, A.A., Elshourbagy, N. (2003) Molecular identification of high and low-affinity receptors for nicotinic acid, *J. Biol. Chem.* **278**, 9869–9874.

Wood, S.L., Emmison, N., Borthwick, A.C., Yeaman, S.J. (1993) The protein phosphorylation responsible for dephosphorylation of hormone-sensitive lipase in isolated rat adipocytes, *Biochem. J.* **295**, 531–523.

Wootan, M.G., Bernlohr, D.A., Storch, J. (1993) Mechanism of fluorescent fatty acid transfer from adipocyte fatty acid-binding to membranes, *Biochemistry* **32**, 8622–8627.

Yasruel, Z., Cianflone, K., Sniderman, A.D., Rosenbloom, M.M., Walsh, M., Rodriguez, M.A. (1991) Effect of acylation-stimulating protein on the triacylglycerol synthetic pathway of human adipose tissue, *Lipids* **26**, 495–499.

Yin, W., Mu, J., Birnbaum, M.J. (2003) Role of AMP-activated protein kinase in cyclic-AMP-dependent lipolysis in 3T3-L1 adipocytes, *J. Biol. Chem.* **278**, 43074–43080.

Zierath, J.R., Livingston, J.N., Thorne, A. (1998) Regional difference in insulin inhibition of non-esterified fatty acid release from human adipocytes: relation to insulin receptor phosphorylation and intracellular signalling through the insulin receptor substrate-1 pathway, *Diabetologia* **41**, 1343–1354.

Triplet Repeat Diseases

Stephan J. Guyenet and Albert R. La Spada
University of Washington, Seattle, WA 98195

1 A Novel Mechanism of Genetic Mutation Emerges 3
1.1 Repeat Sequences of All Types and Sizes 3
1.2 Trinucleotide Repeat Expansion as a Cause of Disease: Unique Features Explain Unusual Genetics 4

2 Repeat Diseases and Their Classification 5
2.1 Summary of Repeat Diseases 5
2.2 Differences in Repeat Sequence Composition and Location within Gene 6
2.3 Classification Based upon Mechanism of Pathogenesis and Nature of Mutation 6

3 Type 1: The CAG/Polyglutamine Repeat Diseases 9
3.1 Spinal and Bulbar Muscular Atrophy 9
3.2 Huntington's Disease 12
3.3 Dentatorubral Pallidoluysian Atrophy 15
3.4 Spinocerebellar Ataxia Type 1 16
3.5 Spinocerebellar Ataxia Type 2 18
3.6 Spinocerebellar Ataxia Type 3/Machado–Joseph Disease 19
3.7 Spinocerebellar Ataxia Type 6 21
3.8 Spinocerebellar Ataxia Type 7 22
3.9 Spinocerebellar Ataxia Type 17 25
3.10 Role of Aggregation in Polyglutamine Disease Pathogenesis 25
3.11 Protein Context 29
3.12 Transcriptional Dysregulation 30
3.13 Proteolytic Cleavage 30

4 Type 2: the Loss-of-function Repeat Diseases 33
4.1 Fragile X Syndrome 33

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

4.2 Fragile XE Mental Retardation 34
4.3 Friedreich's Ataxia 36
4.4 Progressive Myoclonus Epilepsy Type 1 39

5 Type 3: the RNA Gain-of-function Repeat Diseases 40
5.1 Myotonic Dystrophy Type 1 40
5.2 Myotonic Dystrophy Type 2 42
5.3 Spinocerebellar Ataxia Type 8 43
5.4 The Fragile X Tremor – Ataxia Syndrome (FXTAS) 44

6 Type 4: The Polyalanine Diseases 46
6.1 Overview 46
6.2 Oculopharyngeal Muscular dystrophy 46
6.3 Synpolydactyly (Syndactyly Type II) 48
6.4 Cleidocranial Dysplasia 48
6.5 Holoprosencephaly 48
6.6 Hand-foot-genital Syndrome 49
6.7 Blepharophimosis-ptosis-epicanthus Inversus Syndrome 49
6.8 Syndromic and Nonsyndromic X-linked Mental Retardation 49
6.9 Congenital Central Hypoventilation Syndrome 50

7 Unclassified Repeat Diseases Lacking Mechanistic Explanations 50
7.1 Spinocerebellar Ataxia Type 10 50
7.2 Spinocerebellar Ataxia Type 12 51
7.3 Huntington's Disease Like 2 (HDL2) 52

Bibliography 54
Books and Reviews 54
Primary Literature 54

Keywords

Aggregate
Accumulation of proteinaceous and/or ribonucleic acid material into a structure that is visible in a cell at the light microscope level.

Anticipation
Worsening severity of a disease phenotype as the causal (typically dominant) genetic mutation is transmitted from one generation to the next in a family segregating the disease of interest.

Gain of function
Refers to a type of mutation that imparts a novel activity or action to the gene product containing the mutation.

Loss of function
Refers to a type of mutation that eliminates the action of the gene product encoded by the gene within which the mutation resides.

Repeat Expansion
An elongation of a repeat to a larger size that no longer falls within the size distribution range typically seen in the normal population; this process is now recognized as a mechanism of human genetic mutation.

Trinucleotide
Three DNA base pairs of specific sequence composition (e.g. cytosine-adenine-guanine).

The repeat expansion disorders are a group of human diseases that are caused by the elongation of a DNA repeat sequence. In this chapter, we provide an overview of the discovery of repeat expansion as an important cause of human disease, and we summarize the molecular genetics and mechanistic basis of 27 microsatellite repeat disorders. Comparison of the many repeat expansion disorders reveals distinct categories of repeat diseases, allowing us to propose a classification of the repeat expansion disorders based upon mutation sequence and pathogenic mechanism. The four types of repeat expansion disorders defined by this approach are the CAG/polyglutamine repeat diseases; the loss-of-function repeat diseases; the RNA gain-of-function repeat diseases; and the polyalanine diseases. Although the genetic basis for most of these diseases was determined less than a decade or so ago, considerable advances have been made in our understanding of how ''dynamic mutations'' produce molecular pathology and human disease.

1
A Novel Mechanism of Genetic Mutation Emerges

1.1
Repeat Sequences of All Types and Sizes

Long before the sequencing of the human genome was undertaken, the discovery of bacterial enzymes that recognize specific DNA sequences (''recognition sites'') and cleaved them, yielded a new methodology for differentiating individuals (''molecular fingerprinting''). This methodology also found application in the mapping of inherited genetic diseases, taking advantage of human variation in the form of so-called restriction fragment length polymorphisms (RFLP's). In the search for even more informative genetic markers, investigators uncovered a variety of very

short repeat sequences ("short tandem repeats" – STR's; "simple sequence length polymorphisms" – SSLP's) that were highly dispersed throughout the human genome. These latter repeat sequences came to be known as "microsatellites" to differentiate them from the "minisatellites" that were being used for molecular fingerprinting. Minisatellites were originally defined as tandem arrays of 14–100 bp repeating sequences. In the case of minisatellites, the repeat sequence was in essence a "consensus" as deviations from the exact repeat sequence were common. Microsatellites, however, were typically pure perfect repeats of less than 13 bp, the most common microsatellites being dinucleotide repeats, trinucleotide repeats, or tetranucleotide repeats. The discovery of minisatellites and microsatellites yielded a virtual bonanza of reagents for molecular fingerprinting and genetic linkage mapping, while also providing evolutionary biologists and population geneticists with intriguing material to attempt to reconstruct evolutionary relationships and inter- or intraspecies relationships. Although human geneticists and evolutionary biologists were applying microsatellites in different ways for their own studies, they shared the commonly held belief that such repeats were neutral and therefore unlikely to be of much functional consequence, let alone play a role in causing human disease. Of course, all that would soon change in the last decade of the twentieth century.

1.2 Trinucleotide Repeat Expansion as a Cause of Disease: Unique Features Explain Unusual Genetics

In 1991, two groups working on seemingly unrelated inherited genetic diseases independently made paradigm-shifting discoveries. In one case, a CAG trinucleotide repeat expansion within the first coding exon of the androgen receptor (*AR*) gene was found to be the cause of an X-linked neuromuscular disorder known as spinal and bulbar muscular atrophy (SBMA or Kennedy's disease). CAG encodes the amino acid glutamine; thus, elongation of a polyglutamine tract within the AR protein was hypothesized to be the molecular basis for the motor neuron degeneration in SBMA. In the other case, a disorder known as the fragile X syndrome of mental retardation (FRAXA), also X-linked but much more common, was reported to result from expansion of a CGG repeat. In the latter case, although originally envisioned to encode a polyarginine tract, the CGG repeat turned out to be in the 5′ untranslated region of a novel gene, the so-called *FMR-1* gene (for "fragile X mental retardation-1").

The identification of triplet repeat expansions as the cause of two inherited diseases was an exciting turning point in the field of human molecular genetics not only because of the novel nature of these findings, but also because of the unusual genetic characteristics of this new type of mutation. Analysis of families segregating FRAXA revealed the existence of three distinct allele categories defined by the length of the pure CGG repeat: a normal size range; a disease size range; and an intermediate, "premutation" size range. As discussed below, individuals carrying premutation-sized CGG repeats never develop the mental retardation phenotype, but instead are at risk for passing on even larger CGG repeats to their children or grandchildren who then display the mental retardation phenotype. An important tenet that emerged from these early studies was that expanded CGG repeats (larger

than the normal size range) displayed an exceptionally high rate of further mutation. This observation reversed a commonly held view of the genetic material – the notion that any single nucleotide in the human genome displayed a mutation rate of $\sim 1 \times 10^{-5}$. In the case of expanded CGG repeats in premutation carriers, the rate approached unity (10^0)! Besides displaying this high mutation rate, there were other unusual features: (1) the CGG repeats showed a marked tendency to further expansion, suggesting that repeat mutation was a polar process; and (2) the sex of the individual transmitting the premutation-sized CGG allele determined whether a large expansion into the disease range would be possible. For FRAXA, expansion into the disease range could only occur if the premutation allele was transmitted by a female carrier. All of these unusual aspects of FRAXA CGG repeat genetics thoroughly accounted for the bizarre non-Mendelian inheritance patterns described in FRAXA families as the "Sherman paradox."

The recognition of repeat expansion as the cause of the neurodegenerative disorder SBMA and as an explanation for the puzzling genetics of FRAXA set the stage for further discoveries in the field of neurogenetics (the study of inherited neurological disorders). One disorder known as *myotonic dystrophy* (DM), the most common of all muscular dystrophies, had been the center of a genetic controversy that had gone on for nearly a century. The controversy involved the debate over whether the clinical phenomenon of anticipation truly existed or was simply an artifact of clinical study (Anticipation may be defined as a progressively earlier age of disease onset with increasing disease severity in successive generations of a family segregating an inherited disorder). Although anticipation was initially proposed as a defining feature of DM in 1918, a number of leading geneticists, among them Penrose, L.S., dismissed its authenticity, claiming that it was a product of ascertainment bias due to better clinically defining the profound variable expressivity in this disorder. This view persisted from its promulgation in the 1950s, reinforced by the concept that the genetic material is seldom subject to alteration or modification that could be heritable. However, with the discovery that expanded trinucleotide repeats could further expand, and that indeed the expansion process was a prerequisite for the FRAXA disease phenotype – accounting for maternal inheritance and greater percentages of affected individuals in more recent generations – a role for repeats in diseases displaying anticipation was entertained. Consequently, the third repeat mutation disorder to be identified – just one year after the SBMA and FRAXA discoveries – was DM. Studies of the causal CTG repeat expansion in DM demonstrated a strong correlation between disease severity (i.e. age of onset and rate of progression) and the length of the CTG repeat. In this way, it became clear that anticipation was a genuine phenomenon and that expanded repeat mutational instability was its long awaited molecular explanation. The mutational instability characteristics of expanded repeats have led to their designation as so-called "dynamic mutations."

2 Repeat Diseases and Their Classification

2.1 Summary of Repeat Diseases

The list of diseases caused by microsatellite repeat expansions (involving repeats of

3–12 bp) now includes more than 25 disorders (Table 1). Certain aspects of the list of repeat diseases deserve emphasis. Twenty of the repeat diseases either principally or exclusively affect the neuraxis – that is, anywhere from the brain to the cerebellum/brainstem, to the spinal cord, to the peripheral nerve or muscle – and are degenerative disorders. Almost all of these inherited neurological disorders are caused by large expanded repeats that display the property of pronounced genetic instability (dynamic mutation). On the other hand, the developmental malformation syndromes for which repeat expansion mutations have been implicated all involve modestly sized disease repeats by comparison, and these repeats do not exhibit such dynamic mutation genetic instability.

2.2 Differences in Repeat Sequence Composition and Location within Gene

When faced with the task of categorizing the various repeat expansion diseases into different classes, a number of approaches are possible. We have found that consideration of the sequence of the repeat and its location within the gene are the most useful characteristics to apply for grouping the different repeat diseases. As shown in Table 1, there are many different types of repeats varying in length and sequence composition. However, among the recurrent sequence types are CAG trinucleotide repeats, CTG trinucleotide repeats, and GCG trinucleotide repeats. The rest of the repeat sequences are unique in composition and differ widely in size, ranging from 3 to 12 bp as noted above. Comparison of the location of a repeat within the gene it is affecting also yields different types of repeats. The largest single-repeat location category is within the coding region of a gene, which applies to both CAG (glutamine) and GCG (alanine) repeats, and has been proposed but not yet demonstrated for CTG (leucine) repeats. The rest of the locations defined for repeats vary widely, ranging from the gene's promoter to its 5′ untranslated region to an intron to the 3′ untranslated region, and no more than two to three repeat expansions can be placed in each of these categories at this time.

2.3 Classification Based upon Mechanism of Pathogenesis and Nature of Mutation

To allow us to reconstruct how the different repeat expansion mutations cause molecular pathology in the various disorders that they cause, we have chosen to categorize the 25 repeat disorders into four classes (Table 2). The first class of disorders, the Type 1 repeat diseases, are the "CAG-polyglutamine disorders." This class of repeat diseases includes nine inherited neurodegenerative disorders (SBMA, Huntington's disease, dentatorubral pallidoluysian atrophy, and six forms of spinocerebellar ataxia) that all share the common feature of being caused by a CAG repeat located within the coding region of a gene. Upon CAG repeat expansion, a mutant protein with an extended polyglutamine tract is produced, making the protein then adopt an abnormal conformation and misfold to initiate the pathogenic cascade. The resultant pathology is believed to primarily stem from a gain of function of the mutant protein imparted by the expanded polyglutamine tract. As discussed below, much effort has gone into trying to define what the gain-of-function effect is for each polyglutamine disease protein and into determining if

Tab. 1 Compilation of repeat expansion diseases in humans.

Disease	*Gene*	*Repeat*	*Normal*	*Premutation*	*Disease*
Blepharophimosis-ptosis-epicanthus inversus	*FOXL2*	GCG	14	–	24
Cleidocranial dysplasia	*RUNX2*	GCG	11–17	–	27
Congenital central hypoventilation	*PHOX-2B*	GCG	20	–	25–29
Dentatorubralpallidoluysian atrophy (Haw River syndrome)	*Atrophin-1*	CAG	7–35	–	49–88
Fragile X syndrome of mental retardation	*Fmr-1*	CGG	6–53	45–200[a]	200–>2000
FRAXE mental retardation	*Fmr-2*	GCC	6–40	61–200	200–1000
Friedreich's ataxia	*Frataxin*	GAA	7–38	33–60	66–1700
Hand-foot-genital	*HOX-A13*	GCG	18	–	24–27
Huntington's disease	*Htt/IT15*	CAG	6–35	27–36	39–250
Huntington's disease like 2	*JPH3*	CAG	7–27	–	50–60
Holoprosencephaly	*ZIC2*	GCG	15	–	25
Myoclonus epilepsy type 1	*Cystatin B*	CCCCGCCCCGCG	2–3	12–17	30–150
Myotonic dystrophy type 1	*DMPK*	CTG	5–37	50–80	50–4000
Myotonic dystrophy type 2	*ZNF9*	CCTG	104–176	?	175–11 000
Oculopharyngeal muscular dystrophy	*PABP2*	GCG	6–7	–	8–17
Spinocerebellar ataxia type 1	*Ataxin-1*	CAG	6–44	–	39–>10
Spinocerebellar ataxia type 2	*Ataxin-2*	CAG	13–33	–	31–>10
Spinocerebellar ataxia type 3 (Machado-Joseph disease)	*Ataxin-3*	CAG	12–40	–	55–84
Spinocerebellar ataxia type 6	*CACNA1A*	CAG	4–18	–	19–33
Spinocerebellar ataxia type 7	*Ataxin-7*	CAG	4–35	–	37–306
Spinocerebellar ataxia type 8	?	CTG	16–>800	?	71–>1000
Spinocerebellar ataxia type 10	*Ataxin-10*	ATTCT	10–22	?	800–4500
Spinocerebellar ataxia type 12	*PPP2R2B*	CAG	7–32	–	55–78
Spinocerebellar ataxia type 17	*TBP*	CAG	25–48	–	43–66
Syndromic and nonsyndromic X-linked mental retardation	*ARX*	GCG	10–16	–	17–23
Synpolydactyly	*HOX-D13*	GCG	9–15	–	16–29
X-linked spinal & bulbar muscular atrophy	*AR*	CAG	5–34	–	37–70

[a] Male premutation carriers of the FMR-1 CGG repeat develop the Fragile X tremor-ataxia syndrome.
–, no premutation alleles identified, so they may not exist.
?, existence of premutation alleles is unknown.

Tab. 2 Classification of the repeat expansion diseases.

CAG/polyglutamine disorders	*Loss-of-function disorders*	*RNA gain-of-function disorders*	*Polyalanine disorders*
Dentatorubral pallidoluysian atrophy	Fragile X syndrome of MR	Myotonic dystrophy type 1	Blepharophimosis-ptosis-epicanthus inversus
Huntington's disease	FRAXE MR	Myotonic dystrophy type 2	
Spinal and bulbar muscular atrophy	Friedreich's ataxia	Spinocerebellar ataxia type 8	Cleidocranial dysplasia
Spinocerebellar ataxia type 1	Myoclonus epilepsy type 1	Fragile X tremor-ataxia syndrome	Congenital central hypoventilation
Spinocerebellar ataxia type 2			Hand-foot-genital syndrome
Spinocerebellar ataxia type 3			Holoprosencephaly
Spinocerebellar ataxia type 6			Oculopharyngeal muscular dystrophy
Spinocerebellar ataxia type 7			Synpolydactyly
Spinocerebellar ataxia type 17			X-linked MR (syndromic & nonsyndromic)

MR = mental retardation.

shared pathways of toxicity are initiated in the different diseases. At least one, and perhaps a greater number of these disorders may principally involve a simultaneous dominant-negative partial loss of the normal function of the disease protein. The next class of disorders, the Type 2 repeat diseases, are a much more disparate group of repeat disorders. The Type 2 repeat diseases are the "Loss-of-function repeat disorders." These disorders include different repeats that vary in sequence composition and gene location, but share a final common pathway of disease pathogenesis – a loss of function of the disease gene within which they occur. This group includes various classic trinucleotide repeat disorders such as the two fragile X syndromes of mental retardation (FRAXA and FRAXE) and Friedreich's ataxia – but also encompasses the dodecamer repeat expansion in progressive myoclonic epilepsy type 1, and possibly the CAG repeat expansion in Huntington's disease like-2 (*HDL2*) gene. Strong evidence for a loss-of-function pathway in the form of nonrepeat loss-of-function mutations supports many of these classifications. The third group of repeat diseases, the Type 3 disorders, comprise a shared class because all of them have been proposed to involve the production of a toxic RNA species. This category of repeat diseases is thus called the *RNA gain-of-function disorders*. Included among these disorders are two closely related forms of DM, the common and classic myotonic dystrophy type 1 (DM1) and its uncommon phenocopy, myotonic dystrophy type 2 (DM2). Another member of this group is the recently described fragile X tremor-ataxia syndrome (FXTAS) in male premutation carriers – a fascinating example of two different disease pathways operating upon the same expanded repeat mutation based upon size range differences. One form of spinocerebellar ataxia (SCA8) with an unclear mechanism of pathogenesis has also been provisionally placed into this category, based upon current working models of how its repeat causes disease. The last class of repeat disease, the Type 4 disorders, are the "GCG-polyalanine disorders" that are grouped together because all involve short GCG repeat tracts falling within the coding regions of unrelated genes that become expanded to moderately sized GCG repeats. With the exception of oculopharyngeal muscular dystrophy, all are developmental malformation syndromes, and while gain-of-function polyalanine toxicity has been proposed for a number of these disorders, loss of function due to the polyalanine expansion seems more likely for others. Finally, a number of repeat disorders, spinocerebellar ataxia type 10 (SCA10), spinocerebellar ataxia type 12 (SCA12), and Huntington's disease like 2 (HDL2), currently defy classification because very little is known about their molecular basis. These diseases will be considered in the final section of this chapter.

3 Type 1: The CAG/Polyglutamine Repeat Diseases

3.1 Spinal and Bulbar Muscular Atrophy

Spinal and bulbar muscular atrophy (SBMA; Kennedy's disease) is a late-onset neurodegenerative disease with an inheritance pattern resembling X-linked recessive. It has a prevalence of about 1 in 50 000 males. Patients suffer a late-onset, progressive degeneration of primarily lower motor neurons in the anterior horn of the spinal cord and in the bulbar region of the

brainstem; however, variable involvement of sensory neurons in the dorsal root ganglia also occurs. SBMA typically presents with cramps, followed by proximal muscle weakness and atrophy. Patients often exhibit dysarthria, dysphagia, and fasciculations of the tongue and lips. Affected individuals have symptoms of mild androgen insensitivity, such as gynecomastia, reduced fertility, and testicular atrophy.

SBMA is caused by a polymorphic (CAG)*n* repeat in the first exon of the *AR* gene, which is expressed as a glutamine tract. Unaffected individuals carry 5 to 34 triplet repeats, while affected individuals carry 37 to 70 repeats. SBMA exhibits a paternal expansion bias. The disease does not appear to involve a simple loss-of-function mechanism, as complete loss of AR does not result in motor neuron degeneration.

AR is widely expressed in males and females, and is a member of the steroid receptor–thyroid receptor superfamily with a highly conserved DNA binding domain, ligand binding domain and transactivation domain (Fig. 1). In its inactive state, it forms an apo-receptor complex with heat-shock proteins (HSPs) 70 and 90, and resides in the cytoplasm. Upon binding androgen, it dissociates from these HSP chaperone proteins and translocates to the nucleus. Once in the nucleus, AR dimers transactivate certain genes, many of which are responsible for generating and maintaining male characteristics. Although the glutamine expansion does not affect the binding of its ligand, androgen (testosterone), the glutamine tract is in the major transactivation domain, and may affect transactivation competence. However, the effect of the polyglutamine expansion upon AR transactivation competence remains controversial.

Many lines of evidence suggest that AR must translocate to the nucleus to exert its toxicity. Nuclear inclusions (NIs) are present in motor neurons of the spinal cord and brainstem in SBMA patients. In a Drosophila model of SBMA, retinal expression of mutant AR only yielded a degenerative phenotype in the presence of ligand. As in humans, androgen binding causes the nuclear translocation of AR in mice. Transgenic male SBMA mice produce testosterone and will only

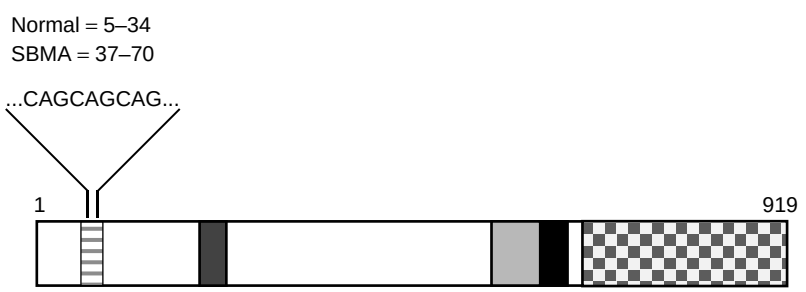

Fig. 1 Diagram of the androgen receptor. The androgen receptor is a member of the steroid receptor-thyroid hormone receptor superfamily, and consequently displays a stereotypical architecture. The CAG repeat – polyglutamine tract (striped box) resides within the amino-terminal domain, which mediates transcription activation through an "activation function" domain (charcoal gray box). Additional conserved domains include the DNA binding domain (light gray box), nuclear localization signal (black box), and the ligand binding domain (checkered box). Expansion of the CAG repeat to alleles of $\geq$37 triplets is the cause of spinal and bulbar muscular atrophy (SBMA).

develop motor neuron disease if expressing polyglutamine-expanded full-length AR protein, yet when castrated, such SBMA transgenic mice do not develop a phenotype. In the same study, female transgenic mice hemizygous for expanded AR develop motor neuron degeneration when exposed to exogenous androgen. Importantly, even rare human females who are homozygous for an AR CAG repeat expansion mutation do not develop SBMA, despite widespread expression of mutant AR throughout the CNS (central nervous system). Thus, SBMA is not a true X-linked recessive disorder, but rather is classified as a sex-limited disorder, since expression of the disease phenotype is dependent upon male levels of androgen.

A number of studies on SBMA patients and mouse models have characterized the NIs seen in this disease. While NIs are widely distributed in lower motor neurons of the spinal cord and brainstem, NIs also occur in a variety of nonneuronal tissues that appear to function normally. NIs colocalize with components of the proteasome and with molecular chaperones. How this contributes to the phenotype is unknown, although HSP70 overexpression attenuates toxicity in cell culture and in transgenic mouse. The presence of proteasome components and HSPs in NIs may thus be revealing a protective intervention by the cell, or alternatively may be deleterious due to depletion of these important cellular proteins. Interestingly, only antibodies raised to amino-terminal fragments of AR detect NIs, indicating that proteolysis may play a role in the disorder. Caspase-3 has been shown to cleave AR in a polyglutamine tract length-dependent manner *in vitro*, and this may have pathogenic significance, as truncated AR is more toxic than full-length protein in cell culture studies and transgenic mouse models. The phosphorylation of AR is modulated by androgen, and this posttranslational modification appears to enhance caspase-3 cleavage.

One possible mechanism of expanded AR toxicity is through transcription interference. Polyglutamine-expanded AR interacts with the transcription coactivator CREB-binding protein (CBP) in a polyglutamine tract length-dependent manner, colocalizes with CBP in spinal cord NIs from patients, and can interfere with CBP-dependent transcription. CBP is a transcription cofactor that regulates the expression of vascular endothelial growth factor (VEGF), among other genes. VEGF is important in motor neuron health, as deletion of a portion of its promoter called the *hypoxia response element* (HRE) causes motor neuron degeneration even in normoxic mice. Pathologically expanded AR reduces VEGF transcript expression in males, with VEGF165 isoform expression reduced at both the RNA and protein levels. Adding VEGF165 to a motor neuron-like cell line (MN-1) expressing polyglutamine-expanded AR significantly rescues its cell death, again supporting the role of transcription interference in SBMA and suggesting that VEGF165 may serve as a neurotrophic factor for motor neurons.

While no effective treatment for SBMA has yet been validated in human patients, the role of testosterone in SBMA disease progression has received considerable attention. Interestingly, testosterone supplementation was initially used as a treatment for SBMA. Rather than aggravating the disease as one might fear, it was reported to attenuate the phenotype slightly, but did not significantly retard disease progression. Such a beneficial effect of testosterone may be due

to anabolic effects on muscle strength or to a downregulation of AR in response to elevated androgen levels. Still another possibility is that AR-mediated transcription in motor neurons somehow performs a trophic function in the face of damage or injury. Very recent work, however, strongly indicates that elimination of ligand by surgical or pharmacological castration is a very effective treatment in SBMA transgenic mouse models, even showing efficacy when SBMA mice display an advanced phenotype. As ligand binding is associated with nuclear translocation of mutant AR and aberrant effects upon nuclear transcription appear crucial for SBMA disease progression, abrogation of nuclear localization may account for the success of castration. Consistent with this hypothesis are results of studies with flutamide, a drug that appears to block AR-dependent transactivation without preventing its nuclear translocation. While pharmacological castration with leuprolide is highly effective therapeutically, flutamide does not ameliorate symptoms or disease progression in an SBMA mouse model. Attempted translation of these preclinical trial results to human SBMA patients is currently underway.

3.2 Huntington's Disease

Huntington's disease (HD) is an autosomal dominant disorder with a prevalence of 1 per 15 000 persons worldwide. It is a debilitating disease that often presents clinically in the fourth or fifth decade of life with chorea (i.e. spontaneous, involuntary dancelike movements). Personality change and cognitive impairment may precede the clinical onset by years, and ultimately culminate in dementia after onset of the movement disorder. Chorea gives way to bradykinesia and rigidity late in the disease. CNS atrophy occurs most prominently in the striatum, which is reduced to a fraction of its original size (Fig. 2). However, significant neurodegeneration and neuron loss in the cortex is also typical, while cerebellum, brainstem, and spinal cord are relatively spared – except in juvenile-onset cases.

HD is caused by a (CAG)*n* repeat expansion in the *huntingtin* gene (*htt*), which encodes a 350-kDa protein containing 67 exons. Unaffected individuals carry 6 to 35 repeats, while affected individuals carry 39 to 250 repeats. The largest repeats cause juvenile-onset HD and display a paternal transmission bias. The htt protein is ubiquitously expressed in brain tissue,

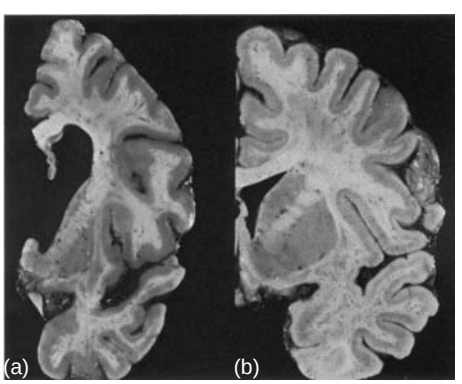

Fig. 2 Huntington's disease (HD) neuropathology. Hemi-coronal sections of postmortem brains from (a) a classic, adult-onset HD patient and (b) a normal control reveal marked degeneration of the striatum (midmedial region) and considerable atrophy of cortical regions. (From Robataille et al. (1997) *Brain pathol.* **7**, 901. Used with permission).

with highest levels in striatal interneurons and cortical pyramidal cells. Two htt mRNA transcripts have been detected: one 10 kb and the other 13 kb in length, expressed most highly in CNS neurons. Ultrastructurally, the wild-type full-length protein is predominantly cytoplasmic and is located in the pre- and postsynaptic regions of dendrites and axons. Htt protein associates with microtubules, vesicles, and organelles.

HD appears to have a predominantly dominant mechanism due to its inheritance pattern, evidence that homozygotes are no more severely affected than heterozygotes, and the observation that heterozygous deletion of huntingtin does not cause HD. The fact that no HD-causing loss-of-function mutations have been documented in the huge *htt* gene further supports a gain-of-function mechanism. However, postnatal elimination of htt expression in regions of the cortex can cause striatal degeneration in mice, so the notion that gain of function fully accounts for the HD phenotype is being reexamined. As htt is a regulator of transcription activation and/or mediator of vesicular brain-derived neurotrophic factor (BDNF) transport up and down axons, many investigators now envision the effects of expanded htt as twofold: simultaneously causing protein misfolding leading to gain-of-function toxicity, and inducing partial loss of an ill-defined normal function.

In 1997, it was first reported that mutant htt forms dense amyloid-type aggregates in the nucleus, perikarya, and neuropil of neurons from a transgenic mouse model and in human patients. The role of aggregation in the polyglutamine diseases and in a wide range of neurodegenerative disorders including Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, and the prion diseases, thus emerged as an important theme at the end of the last decade. While initially it was thought that the formation of large aggregates is the basis of polyglutamine neurotoxicity, the weight of evidence now suggests little correlation between visible aggregate formation and disease pathology. However, the occurrence of aggregates along with neuropathology suggests that aggregation is inextricably linked to the pathogenicity of polyglutamine disease proteins. We will address the role of polyglutamine aggregation in neurotoxicity in detail in a separate section.

Many theories have been proposed to account for how polyglutamine-expanded htt causes neurotoxicity. A thorough discussion of this literature goes beyond the scope of this chapter, so the reader is referred to the relevant books and reviews in the Bibliography for a more intensive treatment of this topic. Major theories of htt neurotoxicity that will be considered herein include transcription dysregulation, proteasome inhibition, mitochondrial dysfunction/excitotoxicity, and proteolytic cleavage. As multiple independent toxicity events may be occurring concomitantly, these disease pathways are not mutually exclusive.

Expanded htt protein may interfere with transcriptional processes. Many transcription factors, such as CBP, contain glutamine tracts that mediate important protein–protein interactions. CBP interacts directly with htt, and mutant htt toxicity is ameliorated in striatal neurons *in vitro* when CBP lacking the htt interaction domain is overexpressed. CBP mediates the transcription of a number of neuronal survival factors, such as BDNF, and functions by acetylating histones, one aspect of chromatin remodeling that permits transcription to occur. Studies of HD fruit fly and mouse models have highlighted the

potential importance of histone acetylation by demonstrating that histone deacetylase inhibitors (HDAC Is) can successfully rescue degenerative phenotypes in these model organisms. Htt can affect transcription mediated by p53, Sp1, and REST interaction, thereby potentially altering the expression of a large number of genes. Differential expression of survival factors, neurotransmitter receptors, and a number of other genes may thus contribute to HD pathogenesis.

The ubiquitin-proteasome protein degradation pathway has also been implicated in HD. Nuclear inclusions of htt colocalize with ubiquitin, which indicates that the expanded protein has been identified as misfolded and thus targeted for proteasomal degradation. However, polyQ sequences longer than 9 glutamines are impossible for eukaryotic proteasomes to cleave. The proteasomes of htt-transfected cells are consequently less capable of degrading proteins other than htt, as demonstrated by the reduced degradation of GFP-tagged proteins in culture. If proteasome components are clogged with mutant protein and/or sequestered into inclusions, they may be unable to degrade other misfolded or damaged proteins that carry out important functions. Alternatively, accumulation of improperly degraded proteins may interfere with other normal cellular processes, such as autophagy or mitochondrial oxidative phosphorylation.

Perhaps the longest and most thoroughly studied htt toxicity pathway involves metabolic disturbances and excitotoxicity. Glucose metabolism and oxygen consumption are reduced in HD brains as measured by PET. Severe deficits in the activity of complexes II/III and IV of the mitochondrial electron transport chain (ETC) are evident in HD brains. Inhibitors of complex II of the mitochondrial electron transport chain, such as 3-nitropropionic acid (3-NPA), can cause a selective degeneration of the striatum in rat and primate models when injected systemically, since the striatum has among the highest energy demands of all neuronal regions. 3-NPA reduces levels of ATP, resulting in mitochondrial and cellular depolarization with activation of voltage-dependent NMDA receptors. Excitotoxic damage may act in concert with increased production of free radicals to cause selective striatal degeneration in HD. Some of the earliest animal models of HD were thus generated by exposing the striatum or entire brain of rats or primates to metabolic or excitotoxic insults. In 1976, the first such model of HD was created by injection of the glutamate analog and excitotoxin kainic acid into the striatum of rats. Such lesioned rats exhibited a selective degeneration of neurons in the striatum reminiscent of HD. Recent studies have suggested that impaired Ca++ flux due to aberrant interaction of htt with the inositol phosphate-3 receptor (IP3-R) may underlie excitotoxic pathology. Studies of mitochondria from HD patients reveal abnormal mitochondrial Ca++ handling and decreased mitochondrial depolarization thresholds, in support of this hypothesis. One very recent study has found that deletion of peroxisome proliferator-activated receptor (PPAR) gamma coactivator 1 alpha (PGC-1α), a key mediator of mitochondrial biogenesis, yields HD-like striatal degeneration in mice. Thus, metabolic insults and excitotoxicity may be key steps in HD disease pathogenesis.

Another important theory of htt neurotoxicity posits that proteolytic cleavage of htt is a required step in neuronal dysfunction and the degeneration process. As noted above, the htt protein is enormous, with full-length product consisting of >3100 amino acids. Analysis of human

HD material has indicated an absence of midprotein and C-terminal epitopes in htt aggregates. Careful biochemical studies of htt have characterized a variety of putative caspase and calpain cleavage sites. Various studies suggest that the more truncated the polyglutamine-expanded htt protein, the more toxic it is in cell culture and in animal models, and the more likely it is to enter the nucleus and produce toxicity there. In the case of htt, a series of proteolytic cleavage steps culminating with cleavage to an ~100 amino acid peptide fragment by an aspartyl protease has been proposed to yield a final "toxic fragment." As it turns out, the "toxic fragment hypothesis" (as it has also been called) may be applicable to a number of polyQ diseases, which will be reviewed later in this section.

3.3 Dentatorubral Pallidoluysian Atrophy

Dentatorubral pallidoluysian atrophy (DRPLA) is a rare, autosomal dominant neurodegenerative disorder most prevalent in Japan. Adult-onset DRPLA typically involves progressive cerebellar ataxia, choreoatheosis, epilepsy, and dementia, while juvenile-onset cases also display myoclonus, epilepsy, and mental retardation. Neuropathological abnormalities include degeneration of the dentate nucleus of the cerebellum, rubral nucleus, and globus pallidus, as well as a more generalized degeneration and gliosis involving the brainstem, cerebellum, cortex, and pons. DRPLA is caused by a polymorphic (CAG)*n* repeat in the carboxy-terminal coding region of the *atrophin-1* gene on chromosome 12. Normal individuals carry 6–35 repeats, while affected individuals carry 49–88 repeats. Anticipation is prominent in DRPLA, as very large repeat expansions can occur in a single generation, typically via paternal transmission.

The DRPLA disease protein, atrophin-1, is widely expressed and appears to be predominantly cytoplasmic. Atrophin-1 contains both a putative nuclear localization signal (NLS) and nuclear export signal (NES), and nuclear localization of normal atrophin-1 is observed. Nuclear localization of polyglutamine-expanded atrophin-1 has been linked to increased toxicity in cell culture models. Ubiquitinated NIs are present in neurons and glia from patient brains, and are also immunoreactive for small ubiquitinlike modifier (SUMO) protein. Atrophin-1 is a substrate for c-Jun N-terminal kinase (JNK), with JNK's affinity for atrophin-1 inversely proportional to the size of the polyglutamine tract expansion.

Although the relevance of JNK phosphorylation to atrophin-1 function is unknown, other insights into the function of atrophin-1 have been reported. Using *Drosophila melanogaster* as a model system, Zhang et al. took advantage of the existence of a fly ortholog of atrophin-1 (*Atro*) and created lines of flies carrying mutations in the *Atro* gene. These flies demonstrated severe developmental abnormalities due to complex patterning defects, and subsequent experiments revealed that *Atro* is a transcription corepressor whose activity diminishes with increasing polyglutamine tract length. Independent studies of human atrophin-1 in cell culture studies found evidence for transcription interference of polyglutamine-expanded atrophin-1 with CREB-mediated gene activation. Transcription dysregulation may thus play a prominent role in DRPLA. Another study suggests that disturbed carbohydrate metabolism may contribute to the DRPLA neurodegenerative phenotype.

3.4 Spinocerebellar Ataxia Type 1

Spinocerebellar ataxia type 1 (SCA1) is an autosomal dominant disorder characterized primarily by a progressive cerebellar ataxia. It accounts for about 6% of all autosomal dominant cerebellar ataxias (ADCAs) worldwide. SCA1 patients suffer from coordination difficulties including dysarthria, dysphagia, and ophthalmoplegia. The cerebellum undergoes atrophy, gliosis, and severe loss of Purkinje cells, accompanied by degeneration of the dentate nucleus, inferior olive, and some brainstem nuclei. Disease onset typically occurs in the third or fourth decade of life, but presentation in childhood or adolescence to late life may be seen. SCA1 is caused by the expansion of a coding (CAG)*n* repeat in the amino-terminal coding region of the *ataxin-1* gene. The ataxin-1 CAG repeat is highly polymorphic, ranging in size from 6 to 44 triplets in unaffected individuals. The repeat length associated with disease ranges from 39 to more than 100 CAGs. Unaffected individuals with more than 20 repeats have CAT triplet interruptions within their (CAG)*n* repeat tracts. Such interruptions stabilize the repeat expansion, while absence of the CAT repeat interspersion is noted in SCA1 patient alleles. As in many other polyglutamine disorders, there is strong evidence supporting a gain-of-function mechanism in SCA1.

Ataxin-1 is widely expressed in the CNS and throughout the periphery, although expression levels are several-fold higher in nervous system tissues. The protein is predominantly nuclear in the CNS, however, some cytoplasmic staining is apparent in Purkinje cells of the cerebellum and in brainstem nuclei. Ataxin-1 knockout mice do not develop SCA1, although they do exhibit impairments in motor and spatial learning. These mice also have decreased paired-pulse facilitation in the CA1 region of the hippocampus, suggesting that ataxin-1 may normally function in synaptic plasticity and learning.

Large ataxin-1 containing NI's occur in the brainstem of affected individuals, and are immunoreactive for ubiquitin, the 20 S proteasome subunit, and HSPs HDJ2 and Hsc70. Work done on SCA1 in transgenic mice by the Orr and Zoghbi laboratories has been crucial in formulating models of not only SCA1 disease pathogenesis but also for influencing views of the molecular basis of all polyglutamine diseases. Indeed, the first mouse model for a polyglutamine disease was generated by transgenic overexpression of polyglutamine-expanded ataxin-1 in Purkinje cells (Fig. 3). This SCA1 transgenic mouse model has laid the foundation for

Fig. 3 The original spinocerebellar ataxia type 1 (SCA1) mouse model. (a) Diagram of the *Pcp2-SCA1* transgene construct. A Purkinje cell-specific expression cassette based upon inclusion of the promoter (straight line), first two noncoding exons (black boxes), and first intron (bent line) of the Purkinje cell protein 2 (*Pcp2*) gene and a SV40 polyadenylation sequence (open box) was the basis for this landmark work. Ataxin-1 cDNAs (gray box) containing either 30 CAGs (control) or 82 CAGs (expanded) were inserted into the *Pcp2* expression cassette. Sites of various PCR primer sets are also shown. (b) *Pcp2-SCA1* CAG-82 mice display ataxia. Still photographs of a 30-week-old *Pcp2-SCA1* CAG-82 mouse from a transgenic line that greatly overexpresses the mutant ataxin-1 transgene illustrates the inability of this mouse to maintain its balance when ambulating. Loss of balance when walking is consistent with the gait ataxia seen in human SCA1 patients. (From Burright et al. (1995) *Cell* **82**, 937, used with permission).

numerous follow-up studies of polyglutamine disease pathogenesis. For example, expression of mutant ataxin-1 lacking an intact self-association domain precluded aggregate formation, but permitted neurotoxicity, demonstrating that NIs are not required for SCA1 in mice. In a later study, crossing of SCA1 transgenic mice with mice lacking a ubiquitin ligase enzyme yielded SCA1 mice incapable of aggregate formation. These mice were more severely affected than their transgenic counterparts

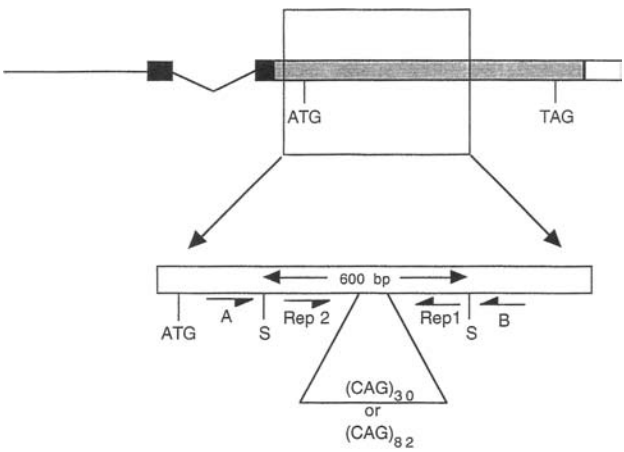

(a)

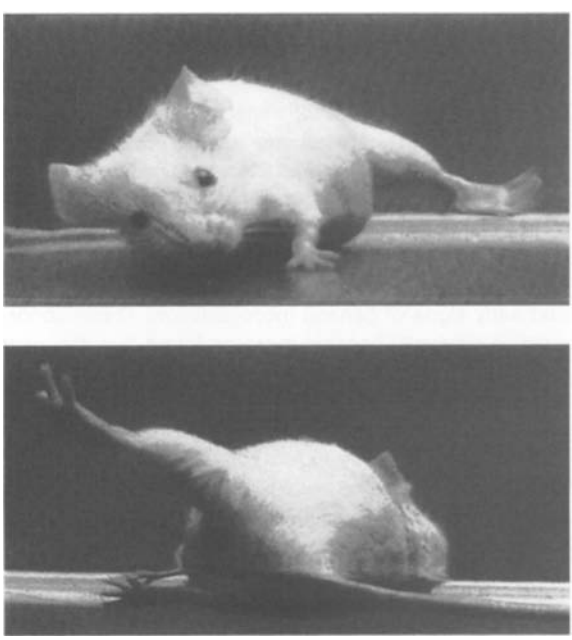

(b)

due to absence of the ubiquitin ligase. This work supported the view that visible aggregate formation may represent a protective cellular response for neutralizing misfolded polyglutamine-containing peptides. In another study, mice expressing ataxin-1 with a mutated NLS were found not to develop SCA1, suggesting that nuclear localization is absolutely required for SCA1 molecular pathology. Although targeting ataxin-1 to Purkinje cells appears sufficient to recapitulate a dramatic SCA1-like disease in mice, a subsequent knockin mouse model of SCA1 indicated that expression in other regions of the CNS yields a more representative disease phenotype.

Over the past few years, numerous leads have emerged in the search for the pathogenic basis of SCA1. In one line of investigation, phosphorylation of serine 776 of the ataxin-1 protein was shown to affect pathogenesis, highlighting the importance of this posttranslational modification. SCA1 transgenic mice in which serine 776 had been mutated to an alanine, exhibited a dramatically attenuated phenotype and a complete lack of NI's. In an accompanying study, interaction of polyglutamine-expanded ataxin-1, but not wild-type ataxin-1, with several isoforms of the phosphoserine/threonine binding protein 14-3-3 was reported. This extremely abundant peptide is thought to serve a regulatory function by binding proteins and determining their subcellular localization, among other things. The interaction between ataxin-1 and 14-3-3 was shown to be dependent upon the Akt-mediated phosphorylation of serine 776. A novel mechanism for ataxin-1 toxicity was thus proposed: upon Akt phosphorylation of polyglutamine-expanded ataxin-1, ataxin-1 binds 14-3-3, is stabilized, and ultimately accumulates in the nucleus. The downstream effects of the nuclear accumulation of mutant ataxin-1 on neuronal function, however, remain undefined.

Transcriptional dysregulation is also a reasonable hypothesis for SCA1 pathogenesis. One study has demonstrated that ataxin-1 interacts with polyglutamine binding protein 1 (PQBP-1), and that this interaction results in interference with RNA polymerase-dependent transcription. Independent studies have supported a role for ataxin-1 as a transcription corepressor. In pull-down assays and in Drosophila, ataxin-1 interacts with the proteins SMRT (silencing mediator of retinoid and thyroid hormone receptors) and HDAC3 (histone deacetylase 3), both transcriptional repressors. Aggregates of polyglutamine-expanded ataxin-1 sequester SMRTER, the Drosophila ortholog of SMRT. Transcription repressors and histone deacetylases also appeared in an earlier screen for modulators of the Ataxin-1 phenotype in the fly. In addition, two separate studies of gene expression alterations in presymptomatic SCA1 transgenic mice have uncovered changes in the levels of transcripts encoding proteins involved in Ca++ flux and metabolism. It thus appears that transcription dysregulation may be a key step in SCA1 disease pathogenesis.

3.5 Spinocerebellar Ataxia Type 2

Spinocerebellar ataxia type 2 (SCA2) is an autosomal dominant, progressive cerebellar ataxia that accounts for about 13% of all ADCA cases. Although patients display problems with voluntary coordinated movements as in the rest of the SCAs, its main distinguishing clinical feature is extremely slow saccadic eye movements. Other symptoms may include hyporeflexia, myoclonus, and action tremor. Patients suffer a gradual degeneration of

the cerebellum, inferior olive, pons, and spinal cord. Both Purkinje and granule cells degenerate in the cerebellum, and CNS atrophy in some patients can be widespread. Interestingly, in certain cases, there is involvement of the substantia nigra, with such patients displaying a prominent degree of parkinsonism. The disorder is caused by a coding (CAG)*n* expansion in a novel gene of unknown function, named *ataxin-2*. The SCA2 CAG repeat displays little polymorphism in the normal population, with 95% of the population possessing 22 or 23 repeats in each allele. The remaining 5% of alleles in unaffected individuals do nonetheless range from 15–31 CAG repeats. Such unaffected individuals typically have two CAA interruptions in their CAG repeat tract. Affected SCA2 patients typically display CAG repeats numbering from 32–63 triplets, and such disease alleles are always uninterrupted CAG tracts. There is a nonlinear inverse correlation between expansion size and age of onset in SCA2, with some studies documenting a paternal transmission bias for larger expansions.

Both wild-type and expanded ataxin-2 mRNA is widely expressed, with highest levels in the substantia nigra and Purkinje cells of the cerebellum. The 140-kDa protein is cytoplasmic and its function remains uncertain. Since it contains Sm1 and Sm2 motifs common in proteins involved in RNA splicing and protein–protein interactions, involvement in RNA processing has been proposed. A yeast two-hybrid screen indicated that ataxin-2 has a binding partner, ataxin-2 binding protein-1 (A2BP1). A2BP1 is highly conserved throughout the animal kingdom and its expression pattern corresponds well with that of ataxin-2. A2BP1 contains a domain that is also conserved in RNA-binding proteins, the RNP motif. Thus, a complex including ataxin-2 and A2BP1 may be involved in RNA processing or metabolism.

The *ataxin-2* gene is evolutionarily conserved. The murine ortholog of ataxin-2 does not contain a polyglutamine repeat tract, however, but instead possesses a single glutamine residue at the analogous location. (Absence of a substantial glutamine repeat tract is typical for mouse orthologs of polyglutamine disease proteins.) In the case of ataxin-2, study of the Drosophila ortholog (*Datx2*), which does show two regions of marked amino acid similarity and does contain polyglutamine repeat regions, has yielded some potentially important insights into ataxin-2s normal function. Modulation of *Datx2* dosage resulted in mutant phenotypes whose cause could be traced to aberrant actin filament formation (Fig. 4). This work suggests that alteration of ataxin-2-mediated regulation of cytoskeletal structure could affect dendrite formation or other aspects of neuronal function in SCA2. As SCA2 is one of the few polyglutamine disorders to display prominent cytosolic aggregates, such a model of SCA2 pathogenesis seems plausible. In other experiments, eliminating the *C. elegans* ortholog of ataxin-2 (ATX-2) yielded a lethal phenotype, indicating that ATX-2 is required for early embryonic development of this nematode worm. It remains to be seen how these observations in worms and flies will apply to ataxin-2 function in mammals, especially since no simple or conditional knockout of the mouse *ataxin-2* gene has been performed.

3.6 Spinocerebellar Ataxia Type 3/Machado–Joseph Disease

Spinocerebellar ataxia type 3 (SCA3) is the most common inherited dominant

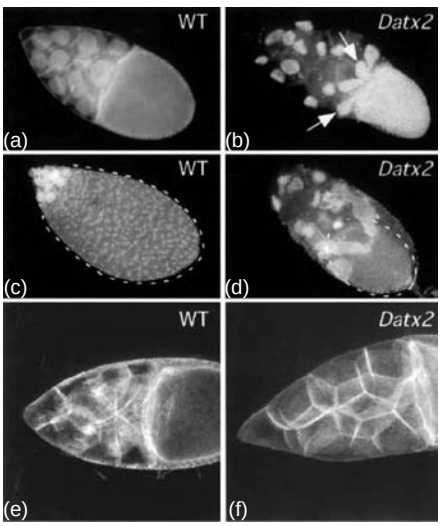

Fig. 4 Studies of the ataxin-2 ortholog in *Drosophila melanogaster* reveals a role for Drosophila ataxin-2 (*Datx2*) in actin filament formation during oogenesis. Egg chambers from normal (Wild-Type (WT)) and mutant flies with reduced expression (*Datx2*) of Drosophila ataxin-2 were stained with DAPI (blue) and phalloidin (red) to indicate nuclei and filamentous actin respectively (a–d). The egg chambers of WT flies prior to cytoplasmic transport (a) display well demarcated, separated but interconnected cells (blue) as expected, while the egg chambers of *Datx2* flies (b) contain irregularly arranged cells. After cytoplasmic transport, the egg chambers of WT flies (c) show one greatly enlarged oocyte (dashed line) with only a small section of compressed cells, while the egg chambers of *Datx2* flies (d) have failed to yield an enlarged oocyte, instead retaining dispersed and large adjacent cells. Confocal images of egg chambers prior to cytoplasmic transport stage reveal a prominent actin filament network in WT flies (e), but a remarkably transparent actin filament network in *Datx2* flies (f). The decreased density of the actin filament network underlies the cytoplasmic "dumping" defect in the *Datx2* flies. (From Satterfield, T.F., Jackson, S.M., Pallanck, L.J. (2002) A Drosophila homolog of the polyglutamine disease gene *SCA2* is a dosage-sensitive regulator of actin filament formation, *Genetics* **162**, 1687–1702, used with permission of *Genetics*.) (See color plate p. xxiii).

spinocerebellar ataxia worldwide. SCA3 is also known as Machado–Joseph disease (MJD) because of its initial description in a group of Portuguese residents of the Azores islands. It is a progressive, autosomal dominant cerebellar ataxia whose clinical features include ophthalmoplegia, dystonia, dysarthria, and signs of lower motor neuron disease, such as tongue and facial fasciculations. Degeneration occurs in the spinocerebellar tracts, dentate nuclei, red nuclei, substantia nigra, and spinal cord. This disease is unique among the SCAs because the cerebellar cortex and inferior olive are largely spared. SCA3 is caused by a (CAG)*n* repeat near the 3′ end of the coding region of a novel gene (*ataxin-3*). Normal alleles range from 12–40 CAG repeats while affected individuals carry 55–84 repeats. Unlike many other triplet repeat disorders, there is a substantial gap between the largest normal repeat allele and the smallest disease-causing repeat allele. Some researchers have proposed that this could be due to a SCA3 founder effect. The presence of SCA3 in every major racial population worldwide, however, would require the founder to be truly ancient. There is the typical inverse correlation between repeat number and age of onset in SCA3, and in this disease, paternal expansion bias is characteristic, as documented by repeat sizing of sperm.

Ataxin-3 is a ubiquitously expressed 42-kDa protein, making it the smallest of the polyglutamine proteins. Ataxin-3 is highly conserved in eukaryotes, with homology to ENTH and VHS domain proteins involved in regulatory adaptor functions and membrane trafficking. Ataxin-3 has several splice variants which reside in the nucleus and the cytoplasm, and it appears developmentally regulated. The existence of NIs was first documented in the brains of patients with SCA3, and this feature of polyglutamine-expanded

ataxin-3 remains evident in SCA3 cell culture and mouse models. Ataxin-3 NIs are ubiquitinated and contain numerous transcription factors.

A number of recent discoveries regarding the domain structure and normal function of ataxin-3 suggest pathways by which polyglutamine repeat expansion could result in disease pathogenesis. Comparison of ataxin-3 amino acid sequences across a wide range of eukaryotic species revealed the presence of an extremely highly conserved amino-terminal sequence that was named the *josephin domain*. Study of this region indicates that it may play a role in aggregate formation in concert with the expanded polyglutamine tract. A rather intriguing feature of ataxin-3 that was discovered independently is the presence of multiple ubiquitin interaction motifs (UIMs), and the demonstration that ataxin-3 is a polyubiquitin binding protein. This suggests a role for ataxin-3 in mediating protein refolding and degradation.

In addition to a possible normal role in protein surveillance, other studies have found ataxin-3 directly interacts with the histone acetyltransferases CBP and p300, and can block histone acetyltransferase activity by inhibiting access of such coactivators to their histone substrates. This appears to be mediated by the interaction of ataxin-3 with histones. *In vitro* and *in vivo* studies of ataxin-3 further revealed an interaction with histones and the chromatin remodeling machinery that led to a repression of transcription activation. Thus, polyglutamine-expanded ataxin-3 may also affect transcriptional processes once it begins to accumulate in the nuclear compartments of the cell types where it is expressed.

3.7 Spinocerebellar Ataxia Type 6

Spinocerebellar ataxia type 6 (SCA6) is an autosomal dominant, slowly progressing cerebellar ataxia that accounts for ~10–20% of ADCA worldwide. It is characterized predominantly by cerebellar dysfunction that may have an episodic component. Other common features may include dysarthria, nystagmus, loss of vibration sense and proprioception, and imbalance. Histopathological changes include loss of Purkinje cells, cerebellar granule neurons, and neurons in the dentate nucleus and inferior olive. SCA6 is caused by a coding (CAG)*n* expansion in exon 47 of the gene *CACNA1A*, which encodes the α_{1A} subunit of the P/Q-type voltage-gated calcium channel. This gene is located on chromosome 19 at band p13. The SCA6 CAG repeat is small compared to other polyglutamine disorders, with a pathogenic range of only 19–33 triplets. Unaffected individuals carry alleles of 4–18 repeats. There is an inverse correlation between repeat length and age of disease onset, and minimal intergenerational and somatic instability has been reported for the SCA6 repeat expansion in affected patients.

The 9.8 kb *CACNA1A* transcript implicated in SCA6 is expressed throughout the CNS, and most highly in cerebellar Purkinje cells and granule neurons. The α_{1A} subunit is the pore-forming component of the channel, which is important in neurotransmitter release at the synapse. Polyglutamine expansions in this subunit cause variable changes in calcium transmission rates, depending on the system and the β subunit coexpressed. Expansions do not cause a reduction in membrane channel density in HEK293 cells, suggesting that aggregation does not occur

at pathogenic repeat lengths. Ubiquitin-negative neuronal inclusions are visible in the cytoplasm of SCA6 patient Purkinje cells, however.

In addition to displaying a disease allele range that does not overlap with the other polyQ diseases, there are several other reasons to suspect that SCA6 is different from the rest of the polyglutamine repeat group. The first difference is the existence of two disorders, episodic ataxia type 2 (EA2) and familial hemiplegic migraine (FMH), that are both allelic to SCA6, as both are caused by point mutations in the *CACNA1A* gene. EA2 resembles SCA6 as affected EA2 individuals suffer from a slowly progressive form of episodic ataxia, accompanied by nystagmus, dysarthria, loss of balance and sometimes cerebellar atrophy. EA2 is usually caused by truncation mutations in *CACNA1A*, resulting in loss-of-function of the calcium channel. CACNA1A knockout mice similarly develop ataxia and late-onset cerebellar atrophy, characteristic of SCA6 and EA2 patient phenotypes. Another reason SCA6 is different from other polyQ disorders is that the CACNA1A protein probably does not undergo a structural change that converts it into an aggregate-prone, beta-sheet adopting, amyloid-like conformer. Indeed, even the largest SCA6 polyglutamine tract is below the threshold required for stable β-pleated sheet formation in other polyglutamine disorders. Consistent with this prediction, channel localization of mutant polyglutamine-expanded CACNA1A protein is not affected in cell culture models, and its channel function is not completely abolished. Cytoplasmic aggregates in patient material are ubiquitin-negative, suggesting that the protein may not be grossly misfolded. This evidence supports a model whereby a dominant-negative loss of function of the P/Q-type voltage-gated calcium channel due to association of the mutant α_{1A} subunit with other subunits causes SCA6. The coincidence that this disorder is caused by a polyQ tract expansion and causes a progressive cerebellar ataxia cannot be ignored, however, more data will be required to soundly refute a possible concomitant toxic gain-of-function effect.

3.8 Spinocerebellar Ataxia Type 7

Spinocerebellar ataxia type 7 (SCA7) is an autosomal dominant, progressive cerebellar ataxia. It is unique among autosomal dominant SCAs, as patients typically develop visual impairment in addition to their cerebellar ataxia. The visual impairment is due to a cone-rod dystrophy that results in retinal degeneration. Patients first develop problems distinguishing colors, but ultimately go blind in this type of retinal degeneration. SCA7 patients may present either with cerebellar ataxia or visual impairment. The likelihood of their presentation is dictated by the size of their CAG repeat disease allele, with larger repeats typically favoring presentation with visual impairment. Affected individuals display prominent dysarthria, and can develop increased reflexes, decreased vibration sense, and oculomotor disturbances. Neuronal degeneration and reactive gliosis occur in the cerebellar cortex, dentate nucleus, inferior olive, pontine nuclei, and occasionally in the basal ganglia. NIs are widespread. Infantile-onset SCA7 has been documented, and in this fatal form of the disease, nonneuronal tissues such as the heart and the kidney are severely affected. SCA7 is caused by a highly polymorphic (CAG)*n* repeat in the 5′ coding region of the *ataxin-7* gene. Unaffected individuals carry 4–35 repeats,

while affected individuals carry 37–306 repeats. The SCA7 trinucleotide repeat is one of the most unstable of all polyglutamine disease genes, sometimes expanding by as much as 250 repeats in a single generation. There is a pronounced paternal expansion bias, with large expansions occurring in male germ cells, frequently causing embryonic lethality that results in reduced transmission. Marked repeat instability can occur in the brain, and produce large somatic expansions on occasion.

Ataxin-7 is a ubiquitously expressed protein of 892 amino acids. It is expressed most highly in heart, skeletal muscle, and pancreas. Expression levels within the CNS are highest in the cerebellum and brainstem. One splice variant, ataxin-7b, is expressed predominantly in the CNS (32). Ataxin-7 contains a functional arrestin domain, a protein interaction domain that is highly selective for phosphorylated forms of its interacting protein(s). This suggests that it interacts with specific phospho-proteins, although none have been discovered to date. It also contains two SH3 domains, which are protein interaction domains that bind proline-rich sequences and mediate a number of cell signaling processes. Ataxin-7 also contains three putative NLSs and one NES.

Ataxin-7 is conserved throughout eukaryotes, and its yeast ortholog SGF73 is part of a multisubunit histone acetyltransferase complex called *SAGA* (*Spt/Ada/Gcn5 acetyltransferase*). The human orthologs of SAGA comprise the so-called STAGA (SPT3/TAF9/ADA2/GCN5 acetyltransferase) complex, and are essential transcription coactivators required for the transcription of certain genes. STAGA components immunoprecipitate with ataxin-7. Although pathogenic expansion of ataxin-7 does not alter its ability to be integrated into the STAGA complex, the presence of the polyglutamine-expanded ataxin-7 has a dominant-negative effect upon the GCN5 histone acetyltransferase activity of the STAGA complex, resulting in transcription dysregulation. The transcription dysregulation caused by polyglutamine-expanded ataxin-7 likely causes a disease phenotype by altering the ability of certain transcription factors to activate expression of their target genes.

The best characterized example of transcription dysregulation by polyglutamine-expanded ataxin-7 is its interference with the cone-rod homeobox protein (CRX), a glutamine domain containing transcription factor expressed only in the retina and the pineal gland. Ataxin-7 interacts directly and functionally with CRX, according to studies performed *in vitro* and in a mouse model of SCA7. Importantly, the interaction between ataxin-7 and CRX appears to involve the glutamine tract regions found in both proteins. Autosomal dominant mutations in CRX can cause a cone-rod dystrophy in humans, further supporting a model in which CRX's diminished transactivation competence is central to the SCA7 retinal degeneration phenotype. Several transcription factors, including CBP, can be found in SCA7. CRX may be but one of a number of transcription factors whose function is diminished by polyQ-expanded ataxin-7 interaction and dysregulation of STAGA complex-mediated gene expression.

One intriguing feature of SCA7 was discovered upon generation of transgenic mice expressing ataxin-7 with the mouse prion protein promoter. This promoter drives expression in every tissue, with the occasional notable exception of the Purkinje cells of the cerebellum. Despite lack of Purkinje cell expression of ataxin-7, mice

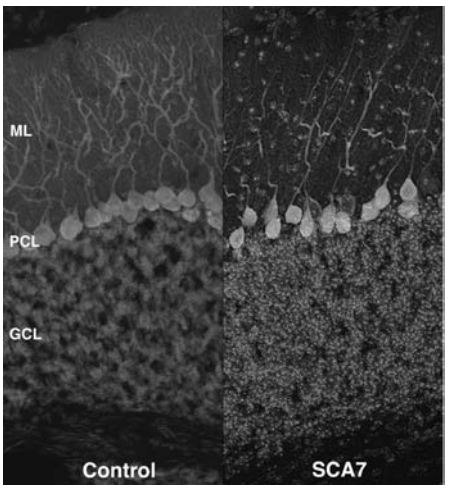

Fig. 5 Noncell autonomous Purkinje cell degeneration in a mouse model of spinocerebellar ataxia type 7 (SCA7). Confocal microscopy analysis of cerebellar sections from a SCA7 transgenic mouse (SCA7) created with an ataxin-7 CAG-92 containing murine prion protein expression vector and from an age- and sex-matched nontransgenic littermate (Control). Staining with an anti-ataxin-7 antibody (magenta), a calbindin antibody (green), and DAPI (blue) reveals a healthy, normal-appearing cerebellum characterized by properly oriented Purkinje cells with extensive dendritic arborization in the "Control" mice. However, SCA7 transgenic mice display pronounced Purkinje cell degeneration as evidenced by decreased dendritic arborization and displacement of Purkinje cell bodies. Interestingly, although numerous neurons in the granule cell layer (GCL) and the molecular layer (ML) display aggregates of ataxin-7, there is no accumulation of mutant ataxin-7 in the degenerating Purkinje cells due to lack of appreciable expression there. As the Purkinje cells degenerate without expressing the mutant protein, the degeneration is described as noncell autonomous. (Adapted from Garden, G.A., Libby, R.T., Fu, Y.H., Kinoshita, Y., Huang, J., Possin, D.E., Smith, A.C., Martinez, R.A., Fine, G.C., Grote, S.K., et al. (2002) Polyglutamine-expanded ataxin-7 promotes noncell-autonomous Purkinje cell degeneration and displays proteolytic cleavage in ataxic transgenic mice, *J. Neurosci.* **22**, 4897–4905, used with permission of the *Journal of Neuroscience*.) (See color plate p. xxiv).

developed a cerebellar ataxia accompanied by degeneration of the Purkinje cells (Fig. 5). This noncell-autonomous degeneration may point to a disease mechanism involving withdrawal of trophic support by communicating neurons (olivary, deep cerebellar, brainstem, or granule neurons) or dysfunction of glutamate transporters expressed by surrounding glia. Damage to inferior olivary neurons or Bergmann glia can indeed cause the degeneration of Purkinje cells, lending credence to this theory.

Another interesting feature of SCA7 that is common to other neurodegenerative disorders is the prominence of morphological and functional degeneration without pronounced apoptosis. Some neurons in SCA7 humans and mouse models exhibit indentations in the nuclear envelope, reduced arborization, ectopy, and increased autophagy. (Autophagy is the bulk degradation of cellular components by a membrane-bound autophagosome and is accelerated by a number of cellular insults.) Caspase-3, a proteolytic enzyme classically associated with apoptosis, is activated at abnormally high levels in SCA7 patient brains. This may suggest that caspase activation, rather than causing apoptosis, is contributing to a nonapoptotic degenerative process.

In normal human neurons, ataxin-7 is variably located in the nucleus or the cytoplasm. In SCA7 patients, ataxin-7 gradually undergoes a shift in localization into NIs. This process has been replicated in SCA7 transgenic mice, in which ataxin-7 immunoreactivity shifts from the cytoplasm to the nucleus, and ultimately forms foci there. In patient's brains, these NIs colocalize with promyelocytic leukemia (PML) protein, which is an integral part of nuclear bodies (NBs). NBs are associated with transcription regulation and the

ubiquitin-proteasome pathway; thus, accumulation of ataxin-7 in NBs may represent an attempt by the cell to degrade misfolded protein.

3.9 Spinocerebellar Ataxia Type 17

Spinocerebellar ataxia type 17 (SCA17) is the most recently discovered polyglutamine repeat disease. It is an autosomal dominant, progressive disorder characterized by dementia as well as cerebellar ataxia and involuntary movement abnormalities. Its symptoms are diverse and heterogeneous, but typically begin with behavioral disturbances and cognitive impairment. This is followed by ataxia, rigidity, dystonia, hyperreflexia, and rarely parkinsonism. Neuropathologically, patients may suffer degeneration of the cortex, cerebellum (including Purkinje cells), inferior olive, caudate nucleus, and medial thalamic nuclei. SCA17 is caused by the expansion of a coding (CAG)n repeat in the TATA-binding protein (TBP) gene on chromosome 6q27. Unaffected individuals carry 25–48 repeats, while affected individuals carry 43–66 repeats. The area of overlap between affected and unaffected alleles indicates incomplete penetrance at intermediate repeat lengths. The pathogenic threshold is higher than for most other polyglutamine disorders, yet there is an inverse correlation between repeat length and age of onset.

TBP is a ubiquitously expressed transcription initiation factor that is a core component of the RNA polymerase II transcription factor D (TFIID) complex. TBP possesses DNA binding activity in the TFIID complex, and is therefore required for the transcription of numerous genes. SCA17 patients have NIs immunoreactive for TBP, polyglutamine, and ubiquitin. Given TBP's central role in transcription regulation, the basis of SCA17 disease pathogenesis is proposed to involve transcription dysregulation, but this is yet to be proven.

3.10 Role of Aggregation in Polyglutamine Disease Pathogenesis

In most polyglutamine disorders and in many neurodegenerative diseases in general, protein aggregation is a prominent feature. It occurs in almost all polyglutamine diseases, despite the lack of domain or structural similarity between the different disease proteins. Aggregates have long been considered reliable indicators of disease, although their pattern and onset often do not correspond well with the cell-type specificity of disease pathology. As the role of aggregation in pathogenesis has been one of the most hotly debated issues in the field of neurodegeneration, it may have implications for Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis. Why? There are several characteristics of polyglutamine-mediated aggregation that are presumably common to all of the disorders that show aggregation. First of all, aggregates are rich in β-pleated sheets and have many of the properties of amyloid. This is supported by Congo red birefringence and immunoreactivity with antiamyloid antibodies. Various studies have shown that the conformational change is associated with the production of visible aggregates rather than in the soluble nonaggregated phase, where polyglutamine tracts are thought to remain in a random coil conformation.

In vitro, the kinetics of aggregation are consistent with nucleated-growth polymerization, in which the rate-limiting step is

the formation of misfolded peptide assemblies, often referred to as *oligomers*. A mechanism involving nucleated-growth polymerization would predict that only one (or a few) visible aggregates would appear in each affected cell, and this appears to be the case for most polyglutamine disorders. In the cellular milieu however, where membranes and intermolecular interactions compartmentalize proteins, this prediction may not be valid. The discovery of polyglutamine microaggregates confirms this suspicion. The threshold for aggregation of polyglutamine tracts is similar to the disease threshold for most polyglutamine diseases. *In vitro*, longer polyglutamine tracts have a lower concentration threshold for aggregation and nucleate more quickly than shorter tracts. This is one possible explanation for the correlation between tract length and disease onset and progression.

Some data support a role for aggregation in polyglutamine disease pathogenesis. Some proponents of this theory invoke the fact that aggregates sequester many proteins besides the disease protein. Intermolecular interactions between transcription factors often involve glutamine tracts or glutamine-rich regions. Some studies have shown that normal proteins with glutamine tracts or glutamine-rich regions are enriched in polyglutamine aggregates in cell culture and animal models. Several transcription factors colocalize with such aggregates, including CBP, TBP, and numerous TBP-associated factors (TAFs). CBP is responsible for the prosurvival effects of BDNF, and its soluble concentration is lowered in HD patient's brains. Postnatal mice lacking CREB and its homolog CREM develop a progressive degeneration of the hippocampus and striatum.

Titration of enzymes and factors required for protein refolding and degradation away from the soluble phase and into aggregates has been proposed as a potential cause of cell toxicity in neurons with aggregates. Caspase activation is another way in which aggregation could be linked to pathology. Caspase recruitment into aggregates can lead to their activation, which could result in dysfunction or cell death in neurons. Further supporting the role of aggregates in polyglutamine protein toxicity, injection of preformed polyglutamine aggregates into the nuclei of cultured cells causes cell death, while the injection of nonpolyglutamine aggregates does not. At the same time, there is strong evidence that soluble polyglutamine protein, rather than aggregates, is the primary source of toxicity. Importantly, the pattern of aggregates observed in human patients often does not correlate with the pattern of neuronal dysfunction. For example, in the striatum of HD patients, large interneurons contain aggregates more frequently than medium spiny neurons, yet the former neurons are largely spared, while the latter are most vulnerable. HD transgenic mice expressing full-length mutant htt will develop inclusions in many brain regions many of which are neuropathologically unaffected, while few inclusions are detected in the striatum, the region of the brain displaying the most prominent pathology.

Aggregation and toxicity have been directly dissociated in other polyglutamine disease model systems. Studies of the SCA1 knockin mouse model revealed that neurons lacking aggregates were more susceptible to dysfunction and demise, while those neurons displaying prominent aggregates were protected. Similarly, SCA7 transgenic mice exhibit retinal pathology before the occurrence of visible aggregates. Finally, in a very provocative study, SCA1

transgenic mice lacking the ubiquitin ligase E6-AP were significantly less capable of forming large and numerous aggregates in neurons in comparison to SCA1 transgenic mice on a wild-type background, but displayed earlier onset of an ataxic phenotype and accelerated neurodegeneration. This suggests that the aggregates may be protective, although the toxicity of compromising the proteasome may have contributed to the accelerated phenotype. Thus ensued a contentious debate over the role of aggregate formation in polyglutamine disease – with some workers espousing the view that aggregates were responsible for disease pathology, others suggesting that the aggregation process was a protective coping mechanism of the cell and thereby beneficial, and still others insisting that aggregates were incidental and irrelevant. This debate was complicated by the existence of "microaggregates," small clumps of aggregated protein visible only at the electron microscope level.

Over the last few years, studies deconstructing polyglutamine tract aggregation into a multistep process suggest a parsimonious explanation for the divergent views. Using a variety of biophysical approaches, including transmission electron microscopy (TEM), Fourier transform infrared spectroscopy (FTIR), and atomic force microscopy (AFM), one study dissected the process of huntingtin (htt) exon 1 peptide aggregation and found evidence for a number of sequential morphological and structural intermediates (Fig. 6). By showing that misfolded htt exon 1–44Q adopted intermediate structures, such work opened up the possibility that intermediates (not visible at the light microscope level) are the toxic species and that the ultimate visible aggregated forms of htt exon 1–44Q are neutralized versions of mutant protein. To examine the role of aggregate formation in polyglutamine toxicity, another group then tracked survival time versus diffuse, soluble htt protein expression levels in htt-transfected striatal neurons undergoing aggregate formation. Comparison of neurons that developed aggregates over time versus those that did not confirmed that the level of soluble nonaggregated mutant polyglutamine protein was the more reliable predictor of cellular toxicity. Such studies have shifted our attention to the role of the intermediates (oligomers, protofibrils, etc.) as the toxic species instead of the final, visible aggregates.

A reasonable model for polyglutamine toxicity predicts that the process starts with a protein that misfolds because of the presence of an expanded polyglutamine tract. The misfolded protein is initially detectable in the soluble phase due to the cell's ability to maintain the protein in a properly folded state and direct it to the degradation machinery. However, the refolding capacity of the cell is ultimately exceeded, and since the degradation machinery cannot turnover the misfolded mutant polyglutamine protein, accumulation occurs. Misfolded polyglutamine proteins can spontaneously change their structural properties and adopt abnormal conformations. These abnormally folded proteins can then nucleate, forming oligomers. Oligomers then form protofibrils that grow into fibrils. The transition to the fibril stage is characterized by the attainment of a β-sheet structure, so at this point the structures are amyloid-like. Fibrils then grow into fibers (also known as microaggregates), which then assemble into aggregates visible under a light microscope. According to such a model, blocking an intermediate step could be therapeutically effective. Consistent with this, one

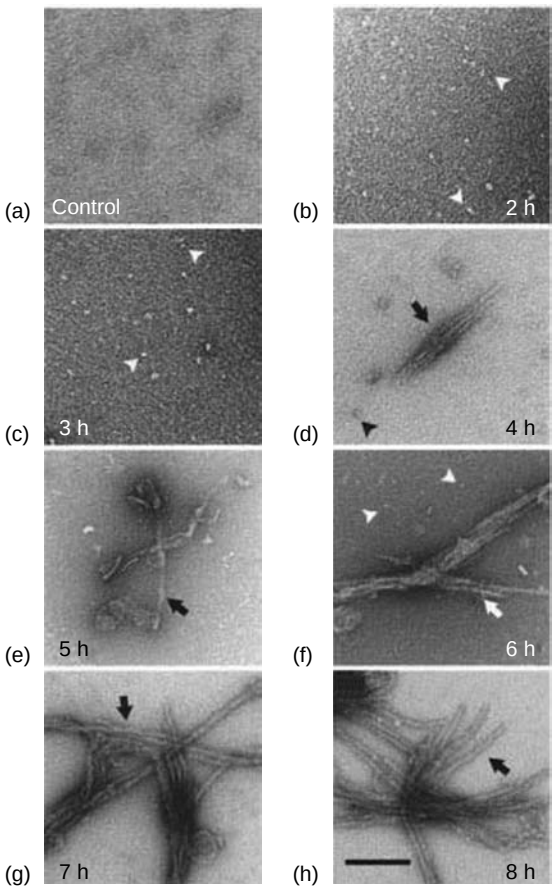

Fig. 6 Polyglutamine-expanded huntingtin protein undergoes a variety of structural alterations on its way to becoming a visible aggregate. In this experiment, transmission electron microscopy tracked the structure of huntingtin exon 1 peptide with 44 glutamines after release from a linked affinity tag. (a) Prior to affinity tag release, nothing is observed. (b-c) After affinity tag release, globular structures become apparent. (d) This is followed by formation of short fibers (4–5 nm in diameter). (e-f) Fiber formation becomes more prominent. However, during this time, large globular assemblies remain (arrowheads). (g-h) Eventually, only fibers are present, and the fibers begin to adopt a uniform appearance, suggesting consolidation into protofibrils. (From Poirier, M.A., Li, H., Macosko, J., Cai, S., Amzel, M., Ross, C.A. (2002) Huntingtin spheroids and protofibrils as precursors in polyglutamine fibrilization, *J. Biol. Chem.* **277**, 41032–41037, used with permission of the *Journal of Biological Chemistry*.).

group has nicely shown that Congo red can bind polyglutamine peptides once they are amyloid-like and prevent their conversion into fibers, while an independent group has shown that Congo red delivery to an HD mouse model is a highly effective treatment intervention.

One satisfying aspect of this model is that it allows us to simultaneously view aggregates as harmful, protective, and innocuous. How? First, aggregates clearly must be toxic as their creation is predicated upon the production of earlier intermediate toxic forms. At the same time, the final visible aggregates are less toxic than the earlier intermediates, so anything that enhances their sequestration into visible aggregates is beneficial. Finally, since oligomers are difficult to detect, it is often not possible to know whether cells are successfully sequestering the toxic precursors into aggregates or if high levels of toxic intermediates are building up. So, aggregates are thus also incidental, since their presence does not provide us with insight into the steady state levels of the toxic precursors. In conclusion, advances in our understanding of the role of aggregate formation in polyglutamine disease processes suggest that aggregates may be simultaneously viewed as harmful, beneficial, and incidental.

3.11 Protein Context

The importance of protein context in the pathogenesis of polyglutamine disorders is a subject of great interest to many researchers, because it is directly related to the mechanisms of toxicity in each disease. It is evident in the polyglutamine disorders that protein context plays a role in cell-type specificity. For example, a polyglutamine expansion in the Htt protein causes a severe degeneration of the striatum and cortex, while the same size glutamine tract in ataxin-1 causes a degeneration of structures in the cerebellum and brainstem, while sparing the striatum and cortex. This cannot be attributed to gross differences in expression patterns, since both proteins are pan-neural (Fig. 7).

One effect that protein context may have on pathology is to affect subcellular localization. For example, the presence of a functional NLS or NES dictates preferential subcellular localization in the nucleus or cytoplasm. Mice expressing expanded ataxin-1 with a mutated NLS do not develop the SCA1 phenotype, while those without mutated NLSs do. Interaction domains also affect polyglutamine protein toxicity. The ability of expanded polyglutamine proteins to interact with other molecules is a promising avenue in the search for mechanisms of cell-type specificity. Yeast two-hybrid screens and other techniques have yielded interaction partners for a number of polyglutamine proteins. For example, Htt's interactions with transcription factors may prove to be central to its pathological effects. At the same time, it is also evident that polyglutamine tracts are innately toxic. Pure polyglutamine tracts cause toxicity in cell culture. Mice that express a glutamine tract of 150 amino acids in the Hprt protein, which does not naturally contain any such tracts, exhibit progressive neurological deficits and NIs. Thus, it appears that pathology in each polyglutamine disease is due to a gain of function of the glutamine tract that is then modulated by the protein context in which it resides.

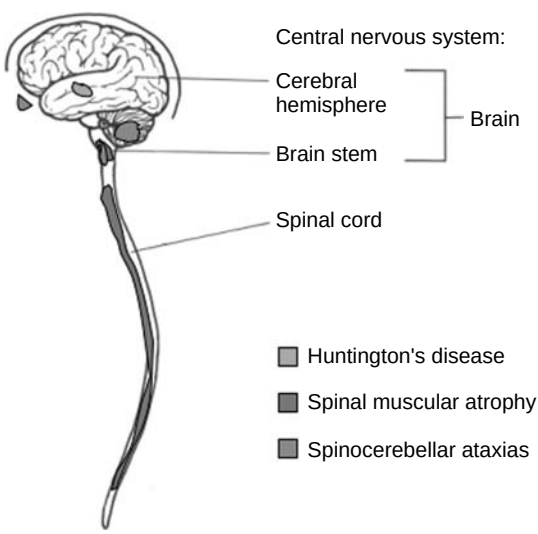

Fig. 7 The conundrum of cell-type specificity in the polyglutamine diseases. Although the different polyglutamine disease proteins are expressed throughout the central nervous system, only select populations of neurons degenerate in the different disorders. The principal regions of selective vulnerability are shown for certain of the polyglutamine diseases.

3.12 Transcriptional Dysregulation

Transcriptional dysregulation is one theory of polyglutamine toxicity that has gained support from many lines of evidence in several disorders, and it appears increasingly likely that it is one of the fundamental causes of pathogenesis. Microarray technology and other genomic techniques are facilitating rapid advances in this area of the polyglutamine field. The majority of polyglutamine disorders are caused by proteins that are either transcription factors/cofactors or interact closely with transcription factors. TBP and AR are transcription factors, atrophin-1 is a transcriptional corepressor, ataxin-7 is part of a transcriptional coactivator complex, and ataxin-1, htt, and ataxin-3 all interact with various transcription factors. It is therefore reasonable to suspect that an abnormally long polyglutamine tract could interfere with the transcriptional activity of these proteins and/or their interactors.

CBP is one of the most studied transcription factors in the polyglutamine field. It is a transcription activator that mediates part of the cellular response to cAMP, and it can interact with numerous polyglutamine proteins in the soluble or insoluble phase, including huntingtin, ataxin-3, ataxin-7, and AR. Expanded htt represses CBP-regulated genes, and overexpression of CBP causes a considerable rescue of the cell death phenotype in HD and SBMA cell culture models. Postnatal mice lacking CBP's upstream activator CREB and its homolog CREM display a profound degeneration of the striatum and hippocampus. Thus, interference with CBP appears to be a common theme in polyglutamine pathology.

Another common theme in the transcription dysregulation equation is the tendency for the polyglutamine disease proteins to alter transcription factor/coactivator-mediated histone acetyltransferase (HAT) activity. The ability of a gene to be transcribed depends upon its chromatin structure, and thus upon the degree of histone acetylation in its vicinity. This is because acetylated histones cause chromatin to be in an "open," transcription-friendly conformation. Acetylation status depends upon the balance between the activity of HATs and their countervailing counterparts, the HDACs. CBP, p300, and p300/CBP-associated factor (PCAF) are all HATs that are inhibited by polyglutamine-expanded htt exon 1 peptide and ataxin-3. In cell culture, expression of a mutant htt fragment reduces global histone acetylation.

Finally, according to a very recent study, ataxin-7 directly interacts with GCN5 as part of the STAGA complex, and upon polyglutamine expansion, mutant ataxin-7 causes a dominant-negative effect upon GCN5 HAT activity. This results in CRX transcription interference that may contribute to the SCA7 retinal degeneration phenotype. Inhibiting HDACs with drugs known as *HDAC inhibitors* (HDAC Is) attenuates the phenotype of HD and SBMA in mouse and fly models, presumably by shifting the cells' acetylation status. HDAC Is such as sodium butyrate and especially suberoylanilide hydroxamic acid (SAHA) have thus emerged as possible candidates for therapeutic trials in human polyglutamine disease patients.

3.13 Proteolytic Cleavage

The occurrence of proteolytic cleavage in polyglutamine diseases first became apparent when it was shown in HD that htt can be cleaved by extracts derived

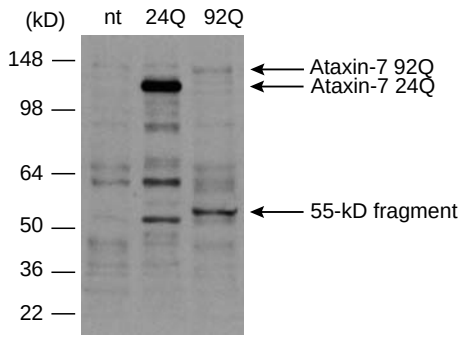

Fig. 8 Proteolytic cleavage in the polyglutamine diseases. Production of amino-terminal truncated proteins is observed in many of the polyglutamine diseases, as is shown here for SCA7. In this experiment, nuclear fractions of retinal lysates from age- and sex-matched nontransgenic (nt), ataxin-7 CAG-24 (24Q), and ataxin-7 CAG-92 (92Q) mice were prepared and immunoblotted with an antibody directed against the amino-terminal region of the protein ataxin-7. As shown here, in addition to soluble full-length ataxin-7 (which is typically reduced in cells where it is aggregating into insoluble inclusions), an ~50–60 kD fragment is detected. (Adapted from Garden, G.A., Libby, R.T., Fu, Y.H., Kinoshita, Y., Huang, J., Possin, D.E., Smith, A.C., Martinez, R.A., Fine, G.C., Grote, S.K., et al. (2002) Polyglutamine-expanded ataxin-7 promotes noncell-autonomous Purkinje cell degeneration and displays proteolytic cleavage in ataxic transgenic mice, *J. Neurosci.* **22**, 4897–4905. Used with permission of the *Journal of Neuroscience*.).

from apoptotic cells. Subsequent publications showed that only amino-terminal epitopes of htt protein are detectable in aggregates. SBMA, DRPLA, and SCA7 were then added to the list of diseases in which aggregates are only immunoreactive for a glutamine-containing fragment of the disease protein. Since then, evidence of proteolytic cleavage has been published for nearly all known polyglutamine disorders. Experimental studies revealed that inhibition of caspase cleavage reduces aggregate formation and toxicity in a cell culture model of HD, underscoring the relevance of proteolytic cleavage in polyglutamine disease pathogenesis. Cleavage promotes aggregation and/or toxicity in SCA3 and SCA7 (Fig. 8). Polyglutamine tract containing htt fragments are more toxic in cell culture than full-length htt, and the most widely used HD mouse model expresses only exon 1 of the *htt* gene.

There are several mechanisms by which cleavage may modulate polyglutamine neurotoxicity. Abnormal proteolysis of polyglutamine proteins may lead to a toxic species that can cause damage in the soluble or insoluble phase. Alternatively, proteolysis may be a normal event in protein turnover and the inability to clear cleaved protein may be the problem. Another possibility is that soluble polyglutamine proteins may cause the activation of proteases such as caspases, which then cleave the disease protein and send the cell on a path to degeneration and ultimately apoptosis. As we will see, each theory has support and not all are mutually exclusive.

Abnormal proteolysis of polyglutamine disease proteins is one possible mechanism of toxicity. Some studies have shown that polyglutamine repeat length modulates susceptibility to proteolytic cleavage, but results have been inconclusive. Few studies have been performed using human brain tissue, however. Considerable evidence exists to show that polyglutamine tracts themselves are resistant to degradation by mammalian proteasomes. The 20S and 26S eukaryotic proteasomes are incapable of cutting within polyglutamine tracts longer than 9 amino acids *in vitro*. The inability of the proteasome to digest glutamine tracts may contribute to the toxicity and aggregation of expanded polyglutamine peptide fragments.

Proteolysis may be a normal part of the turnover of wild-type and mutant htt. One group published evidence that caspase-3 cleaves both mutant and wild-type htt in the cytoplasm of HD brains, suggesting that this particular cleavage is a normal event in its turnover. Since this fragment was located exclusively in the cytoplasm, it also implies that further cleavage occurs before the nuclear translocation of htt. Another study suggested that pepstatin-sensitive aspartyl endopeptidases such as the presenilins and cathepsins D and E may be involved in the normal turnover of htt. This process, which may normally aid in maintaining the balance between htt synthesis and degradation, could generate a toxic, highly aggregation-prone (and hence intermediate oligomer/protofibril-prone) species when it cleaves mutant htt.

Another possible source of toxic cleavage products is the ubiquitin-proteasome system, which is responsible for labeling and digesting misfolded proteins, among other things. Polyglutamine tracts beyond a threshold of about 35 glutamines adopt an abnormal β-pleated sheet conformation, which may cause the host protein to be ubiquitinated and thus targeted for degradation. In support of this theory, a common feature of polyglutamine diseases is the presence of ubiquitin-positive inclusions. Arguing against this idea is an experiment showing that in a cell culture model of SCA1, increased proteasome degradation due to the introduction of a degradation signal reduces aggregates and toxicity. Another potential source of toxic fragments is the autophagy pathway, which is responsible for the bulk degradation of cellular components. According to one study, degradation of htt by autophagy and the autophagosome-associated cathepsin D creates toxic htt fragments.

There are many proteases that are suspected to play a role in the proteolytic cleavage of polyglutamine proteins, the most studied of which are the caspases. Caspases are cysteine- dependent aspartyl proteases that play a crucial role in apoptosis, but are increasingly being investigated for their nonapoptotic functions. In healthy cells, caspases are present predominantly as inactive proenzymes, which must be cleaved for full activity. Caspase-3 is the only protease that has been shown to cleave htt in HD patient's material, although caspases 1 and 6 have also been implicated in cell culture and *in vivo*. Inhibiting the cleavage of htt by any of these three proteases is protective, and an HD mouse model expressing dominant-negative caspase-1 showed significantly delayed disease onset. Caspase-3 has also been implicated in the proteolytic cleavage of atrophin-1, ataxin-3 and AR. Caspases are activated in response to a number of cellular insults, and in postmitotic cells such as neurons where inhibitors of apoptosis proteins (IAPs) are highly expressed, this may cause cellular damage that does not result in classical apoptosis.

Calpains are another family of proteases implicated in the cleavage of polyglutamine proteins. They are calcium-activated cysteine proteases that exist predominantly as proenzymes. As mouse models of HD and material from HD patients exhibit abnormal mitochondrial calcium regulation, some investigators have proposed that altered calcium flux is the mechanism of calpain activation. Indeed, several cell culture studies support the role of calpain proteases I, II, and "m" in htt cleavage. As mentioned previously, other aspartyl proteases such as the Alzheimer's disease-associated presenilins and autophagy-associated cathepsins have also been implicated in htt cleavage.

4
Type 2: the Loss-of-function Repeat Diseases

4.1
Fragile X Syndrome

Fragile X syndrome (FRAXA) is an X-linked disorder with a prevalence of 1 in 4000 in males and 1 in 8000 in females, making it the most common form of inherited mental retardation. Neurological presentation frequently includes mild to severe mental retardation, hyperactivity, poor eye contact, high-pitched speech, and flapping or biting hand movements. Physical signs in males include long, prominent ears, and jaws, macrocephaly, postpubescent macroorchidism, and occasionally, connective tissue abnormalities. The mutation responsible for FRAXA is typically an expansion of a polymorphic (CGG)*n* repeat found in the 5′-untranslated region of the fragile X mental retardation-1 (FMR1) gene. Expanded chromosomes have a folate-sensitive fragile site at Xq27.3 that can be viewed under a light microscope under special cell culture conditions. Normal alleles contain 6–53 triplets punctuated by one or more AGGs, which are considered to have a stabilizing influence on the repeat. Disease alleles contain expansions beyond 200 and up to 2000 repeats, with no AGG interruptions. Pathogenic expansions result exclusively from maternal transmission. FRAXA appears to be a loss-of-function disorder, since deletion of FMR1 and loss-of-function point mutations can also cause FRAXA. This view is further supported by the FMR1 knockout mouse, which reproduces certain aspects of the human disorder.

FRAXA patients have reduced levels of the *FMR1* gene product, FMRP, and there is a linear correlation between reduced protein levels and IQ test scores. Expansion of the disease allele results in hypermethylation of the (CGG)*n* tract, which spreads to a nearby CpG island in the FMR1 promoter region. Some of the proposed secondary structures formed by the FRAXA repeat contain C–C mispairs, which are good targets for human DNA methyltransferase. Another theory invokes the RNAi protein Dicer's ability to cleave CGG repeat RNA, postulating that the resulting siRNAs recruit DNA methyltransferases to the FMR1 locus. Affected alleles also display condensed chromatin, loss of histone acetylation and increased histone methylation. These data support a process whereby DNA methylation recruits transcription silencing machinery, which subsequently suppresses FMR1 transcription in the nearby promoter region. Interestingly, while premutation carriers had been viewed as perfectly normal for decades, recent work indicates that some premutation carriers (60–200 uninterrupted CGG repeats) display a phenotype distinct from FRAXA. Females have a predisposition to premature ovarian failure (POF), and males may develop late-onset ataxia and tremor with the presence of neuronal intranuclear inclusions (NIs) that consist of RNA. (This new FXTAS is discussed in a separate section.) Expression studies have shown that premutation carriers may express up to seven times more FMR1 mRNA than normal individuals. This upregulation is probably not due to compensation for loss of function, because individuals with point mutations resulting in loss of function of FMRP do not have higher levels of the transcript. Abnormal transcript levels due to premutations or full mutations are thought to be as a result of the expansion's effect

on transcription initiation rather than on mRNA stability.

The FMR1 promoter lacks a functional TATA box and initiator sequence. Four functional transcription factor binding sites have been identified in normal individuals: two sites binding USF1/USF2 and Nrf-1, and two GC boxes that bind members of the "Sp" family of transcription factors. Sp1 and Nrf-1 binding is disrupted in cells from FRAXA patients. The *FMR1* gene encodes the widely expressed FMRP, which is an RNA-binding protein most highly expressed in the brain and testes. It is 60% homologous to two other proteins, FXR1P and FXR2P, with which it interacts. The protein is thought to bind approximately 4% of all brain mRNA transcripts, through its RGG box domain. It selectively recognizes RNA-containing hairpin or tetraplex secondary structures ("G-quartet"), and shuttles into and out of the nucleus and associates with polyribosomes in messenger ribonucleoprotein complexes (mRNPs). The FMRP-containing mRNP complexes also contain Pur α and mStaufen, proteins involved in the transport of neuronal granules. These granules, which contain RNA and associated proteins, are transported to dendritic spines in a metabotropic glutamate receptor 5 (mGluR5)-dependent manner. FMRP can suppress the translation of certain transcripts *in vitro* and *in vivo*. FMRP may thus regulate the transport, localization, and translation of certain mRNAs in an activity-dependent manner (Fig. 9). In Drosophila, FMRP interacts with components of the RNAi machinery, which is involved in gene silencing and thus translational control. Although FMRP does not seem to affect the siRNA pathway, it may regulate microRNAs (miRNAs), which are noncoding RNAs thought to control the translation of mRNAs by binding to their 3′-untranslated region. FMRP associates with miRNAs, and proteins in miRNA-containing complexes in mammals, interactions that may be relevant to FMRPs regulation of translation.

FMRP is important for the development of dendritic spines and synaptic plasticity. It regulates the expression of MAP1B, an important regulator of microtubule stability. FMRP knockout mice have abnormally high levels of MAP1B, resulting in increased microtubule stability. This may affect the development of dendrites and/or dendritic spines. The absence of FMRP in hippocampal neurons results in immature dendritic spine morphology and delayed synaptic connections, perhaps contributing to the neurological phenotype observed in FRAXA. FMRP may also affect mGluR5-dependent long term depression (LTD). LTD is a process by which neuronal activity can cause a lasting desensitization of neurons to depolarization. LTD-associated protein synthesis at synapses may be enhanced by mGluR5 activation, while it appears to be suppressed by FMRP. This is supported by evidence of enhanced LTD in the FMR-1 knockout mouse. FMRP suppression and mGluR5 activation of LTD-dependent local protein synthesis may be opposing forces that are out of balance in FRAXA, perhaps accounting for part of the cognitive abnormalities.

4.2 Fragile XE Mental Retardation

Fragile XE mental retardation (FRAXE) is an X-linked disorder with a prevalence of about 1–4% that of FRAXA. It accounts for approximately two-thirds of families with nonspecific X-linked mental retardation (MRX), a classification in which mental retardation is the only consistent

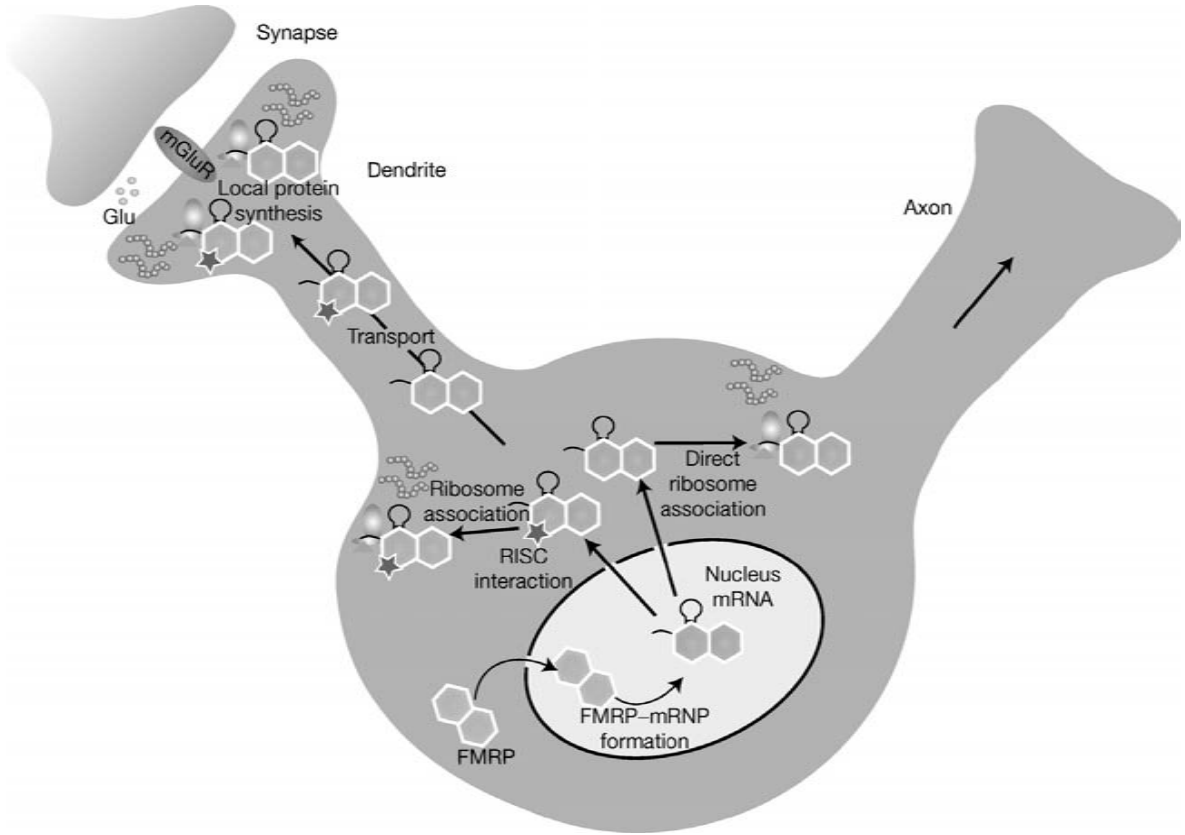

Fig. 9 Advances in the understanding of FMRP's normal function. FMRP binds to recently transcribed messenger RNAs in the nucleus to form a ribonucleoprotein complex (mRNP). The FMRP–mRNP complex moves out of the nucleus and then either directly associates with ribosomes *or* interacts with the RNA-induced silencing complex (RISC). The FMRP–mRNP complex is also transported to dendrites. Whether in the perinuclear cytosol or in dendrites, FMRP is believed to regulate protein translation. (From Jin, P., Alisch, R.S., Warren, S.T. (2004) RNA and microRNAs in fragile X mental retardation, *Nat. Cell Biol.* **6**, 1048–1053. Used with permission of *Nature Cell Biology*.).

clinical feature. Other characteristics are variable and may include hyperactivity, mild facial hypoplasia and nasal abnormalities. FRAXE is caused by the expansion of a (GCC)*n* repeat in the promoter region of *FMR2*, a gene 600 kb downstream of *FMR-1*. It also overlaps with the putative promoter of a gene called *FMR3*, transcribed in the opposite direction. Both maternal and paternal transmission can result in expansion. Normal individuals bear 6–35 GCC repeats, premutation carriers bear 61–200 repeats, and full mutation carriers bear >200 repeats. Pathogenic expansion results in a folate-sensitive fragile site at the disease locus.

In mice, FMR2 mRNA is expressed in adult and fetal brain, kidney, lung, and placenta, with highest brain levels in the hippocampus and amygdala. Transcripts of both FMR2 and FMR3 are reduced beyond detection in FRAXE patients. This may be due to a methylation-dependent silencing process similar to FRAXA. The presence of the repeat within the preinitiation region of the FMR2 promoter may also suggest a more direct disruption of transcription. FMR2 mRNA is highly expressed in the hippocampus, an area critical for learning and memory. Its paralogs AF4 and LAF4 are both transcription transactivator proteins, and the FMR2 protein seems to be a potent transcription activator itself. Furthermore, the FMR2 protein is nuclear, consistent with its proposed role in transcription regulation. The Drosophila ortholog of FMR2/AR4, Lilliputian, is essential for proper organ development, and its loss of function is lethal. FMR2 knockout mice display impairment in the conditioned fear test and enhanced LTP (long-term potentiation) in the hippocampus. The role of FMR3 in FRAXE, if any, is unknown. The mechanism of FRAXE pathogenesis may therefore involve silencing of the human *FMR2* gene resulting in altered transcription in many parts of the developing nervous system and mature brain. The neurological deficits characteristic of FRAXE may involve effects of altered transcription on the hippocampus, amygdala, and possibly other brain structures.

4.3 Friedreich's Ataxia

With a prevalence of about one per 50 000 individuals in the Caucasian population, Friedreich's ataxia (FRDA) is the most common inherited ataxia in this ethnic group. It is a multisystem degenerative disease that is unusual among the triplet repeat disorders due to its autosomal recessive inheritance. Neurological symptoms include gait, limb and truncal ataxia, loss of position and vibration senses, diminished tendon reflexes, and dysarthria. Neuropathology changes include degeneration of the posterior columns of the spinal cord, loss of large primary sensory neurons in the dorsal root ganglia (DRG), and mild, late-onset degeneration of the cerebellar cortex. Other common clinical features are cardiomyopathy, diabetes mellitus, scoliosis, and other skeletal abnormalities. Patients often present with symptoms in childhood, become wheelchair-bound by their late teens or early twenties, and have reduced lifespans due to cardiac failure. Adult presenting patients, however, can have nearly normal lifespans with more protracted progression and less severe nonneuronal involvement. Indeed, until the identification of the causal mutation, many of these adult-onset cases went undiagnosed.

FRDA results from the expansion of a polymorphic (GAA)*n* repeat in the first intron of the gene *X*25, now known as frataxin. FRDA patients have at least one

expanded allele. For the disease to occur, the second allele must contain either an expanded repeat, or rarely, a copy of frataxin containing a loss-of-function mutation. No patients have been described with loss-of-function allele mutations in both *FRDA* genes. This is probably due to prenatal lethality, as frataxin-null mice die in utero. Disease severity and onset age are determined by the size of the (GAA)*n* repeat of the smaller expansion allele. Unaffected individuals carry at least one allele between 7–38 repeats, while affected individuals carry two alleles of 66–1700 repeats. Interruptions of the (GAA)*n* repeat result in later onset and attenuated presentations. GAA expansion at the FRDA locus can result in enormous repeats in a single generation. Maternal transmission can result in expansion or contraction, while paternal transmission results primarily in contraction.

FRDA patients exhibit reduced frataxin RNA and protein levels, and evidence suggests a defect in transcription or RNA maturation. Current models propose that frataxin RNA elongation may be disturbed by triplex structure formation between expanded DNA strands during transcription. DNA triplexes are formed when one strand of a double-stranded DNA molecule folds back upon itself, and interacts with two previously annealed strands. This creates a local structure comprising three DNA strands held together by hydrogen bonds. Certain sequences favor this process, among them extended (GAA)*n* repeats. Triplex formation in the first intron would presumably affect transcript elongation but not initiation. In support of this model are *in vitro* transcription experiments demonstrating no effect of repeat length on transcript initiation.

Frataxin is a 210 amino acid protein that is well conserved from prokaryotes to mammals. It contains mitochondrial targeting signal sequences and localizes to the mitochondrial matrix. Frataxin expression occurs at the primary sites of pathology: dorsal root ganglia, spinal cord, sensory nerves, heart, and pancreas. These are tissues that rely upon high levels of oxidative metabolism and consequently are rich in mitochondria. In addition, such cell types are often postmitotic, meaning that most dividing cell types are spared in FRDA patients.

Studies with FRDA patient material have demonstrated an increased heart iron content and a deficiency in iron–sulfur cluster-containing proteins, including aconitase, a protein involved in iron homeostasis. In addition, fibroblasts derived from FRDA patients are abnormally sensitive to iron and hydrogen peroxide induced stress. Ablation of the yeast frataxin homolog results in respiratory dysfunction, abnormal accumulation of mitochondrial iron, impaired biogenesis of iron–sulfur proteins, and increased sensitivity to oxidative stress. Conditional knockout of frataxin in striated muscle results in a heart-specific phenotype resembling the cardiac abnormalities in human FRDA (Fig. 10). This supports a genetic mechanism involving loss of function of the disease protein. Frataxin may be part of a complex that delivers iron to Iron–sulfur clusters (ISCs), which are cofactors essential to the activity of many important cellular proteins. The accumulation of iron outside the mitochondria, although probably not central to FRDA pathogenesis, supports a deficiency in iron delivery to ISCs. Among ISC-dependent cofactors are proteins involved in mitochondrial electron transport and thus respiration. Disturbances in oxidative metabolism are often associated with increases in the production of

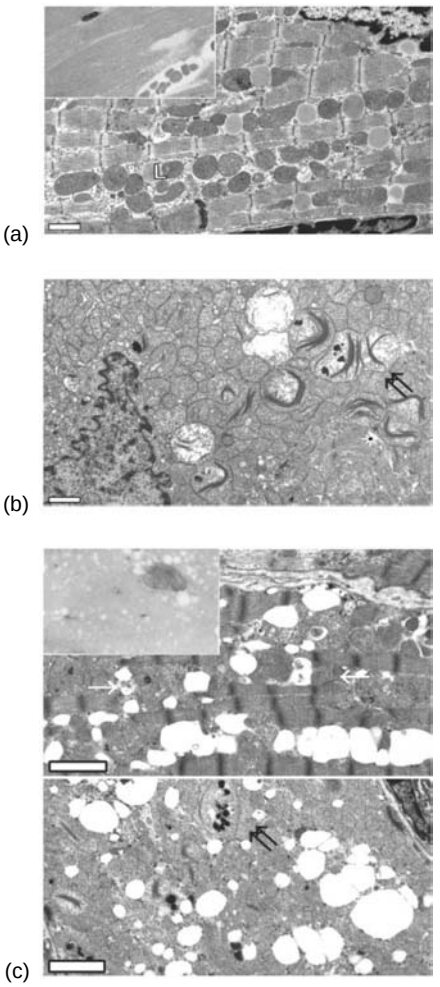

Fig. 10 Friedreich's ataxia mice display cardiac muscle pathology.
(a) Transmission electron microscopy of mice lacking expression of frataxin in only their muscle reveals the abnormal accumulation of lipid droplets (L) in cardiac muscle at 4 weeks of age. (b) By 7 weeks of age, mitochondria appear abnormal and iron deposits are visible (arrows). (c) With further progression, large vacuoles emerge (inset-top) as mitochondria undergo continued prominent degeneration. Ultimately, the mitochondria become engorged with electron-dense material consistent with iron deposits. (From Seznec, H., Simon, D., Monassier, L., Criqui-Filipe, P., Gansmuller, A., Rustin, P., Koenig, M., Puccio, H. (2004) Idebenone delays the onset of cardiac functional alteration without correction of Fe-S enzymes deficit in a mouse model for Friedreich ataxia, *Hum. Mol. Genet.* **13**, 1017–1024, used with permission of *Human Molecular Genetics*).

toxic reactive oxygen species (ROS). FRDA therefore is thought to result from reduced frataxin levels, leading to abnormal iron–sulfur metabolism, mitochondrial dysfunction, oxidative stress, and tissue degeneration. FRDA is thus caused by the dynamic mutation of a nuclear-encoded mitochondrial protein. The disease shares features with classic mitochondrial disorders such as MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis, and strokelike episodes syndrome) and MERRF (myoclonus epilepsy with ragged red fibers), which are caused by stable mutations to mitochondrial-encoded proteins. FRDA appears to be a classic mitochondrial disorder with an unusual genetic basis – dynamic mutation.

Given the data suggesting a role for oxidative metabolism and iron transport in FRDA pathogenesis, antioxidants and iron transport molecules are being evaluated as treatments for this disease. Several clinical trials have been conducted in FRDA

patients using idebenone, a synthetic analog of coenzyme Q10 and a potent antioxidant. Some studies have shown substantial improvements in the heart function of patients, although the lack of randomized placebo controls and the variability of the disorder render the results controversial. A recent conditional knockout mouse model of FRDA also supports the use of idebenone as a treatment for the disorder, making treatment of FRDA patients with this antioxidant a distinct possibility.

4.4 Progressive Myoclonus Epilepsy Type 1

Progressive myoclonus epilepsy type 1 (EPM1) is a rare, autosomal recessive neurological disorder most prevalent in Finland and parts of North Africa. Its most prominent symptoms are progressive photosensitive myoclonus and tonic-clonic epilepsy. Some patients also experience a progressive cerebellar ataxia and cognitive decline. Neurodegeneration occurs in the thalamus, spinal cord, and the Purkinje and granular neurons of the cerebellum. Typical age of onset is 6 to 16 years of age. EPM1 is sometimes caused by missense mutations in the *cystatin B* gene (CSTB); however, analysis of affected patients lacking such mutations revealed a dodecamer repeat expansion upstream of CSTB. The dodecamer repeat is located between 66 and 77 bp 5′ of two putative transcription start sites, and has the sequence CCCCGCCCCGCG. Unaffected individuals have a repeat number of 2–3, premutation carriers have 12–17 repeats, and affected individuals carry 30–150 repeat alleles.

CSTB is a highly conserved gene in the cystatin family of cysteine protease inhibitors. CSTB mRNA is ubiquitous, with high transcript levels in the hippocampus. The protein binds to and inhibits lysosomal proteases such as cathepsins B, H, L, and S. Pathogenic expansion of the repeat causes a reduction in CSTB transcription in some cell types. Cstb knockout mice develop symptoms similar to those seen in EPM1 patients, and missense mutations causing the disease disrupt the ability of CSTB to bind cysteine proteases. Together, these data suggest a loss-of-function mechanism for EPM1. Lowered inhibition of cysteine proteases may result in neuronal damage, causing the phenotype observed in EPM1.

There are several ways in which the repeat might reduce CSTB expression. The first is that reduced gene expression may be due to an increase in the distance between promoter elements and the transcription start site. AP-1 binding sites have been located upstream of the repeat. One study demonstrated a two- to fourfold reduction in promoter activity when the repeat is expanded. An equivalent reduction also occurred when similarly sized heterologous DNA was used in place of the expanded repeat. Independent studies have confirmed a large reduction in promoter activity due to repeat expansion. Another way the repeat could affect CSTB expression is through alterations in chromatin structure. The repeat region is G-C rich and might exclude nucleosomes or promote the formation of secondary structures. Abnormal DNA secondary structures have been observed in the EPM1 repeat region. These alterations could affect the expression of *CSTB* or other genes in the vicinity. A third possible mechanism for expression changes lies in the sequence of the repeat region. Each dodecamer repeat contains a GC box, which could act as an Sp1 binding site *in vivo*. In cases of abnormally large repeat numbers,

increased Sp1 binding could contribute to the suppression of CSTB transcription.

5
Type 3: the RNA Gain-of-function Repeat Diseases

5.1
Myotonic Dystrophy Type 1

Myotonic dystrophy or DM is an autosomal dominant, multisystem disorder with a prevalence of 1 in 100 000 worldwide, and 1 in 8000 in European and North American Caucasian populations. Patients typically present with proximal or distal muscle dysfunction including weakness, pain, and myotonia (failure of muscle relaxation). DM exhibits a combination of characteristic symptoms: cardiac conduction abnormalities, subcapsular iridescent cataracts, and unusual endocrine changes. Other features include testicular atrophy, type II diabetes, and late-onset cognitive impairment. In its most severe form, congenital DM, mental retardation, craniofacial deformities, and other developmental abnormalities are present. It exhibits both paternal and maternal transmission, although there is an almost exclusive maternal transmission in congenital DM. Myotonic dystrophy type 1 (DM1) is caused by a (CTG)n expansion in the 3′-untranslated region of the gene, dystrophica myotonica protein kinase (DMPK), on chromosome 19. Unaffected individuals carry 5–37 repeats, while affected individuals carry 50–4000 repeats. DM exhibits pronounced anticipation and dramatic somatic instability. Interestingly, the clinical phenomenon of anticipation (worsening severity as a disease gene is transmitted from one generation to the next) was first described nearly a century ago in a family segregating DM1. Although the anticipation phenomenon was dismissed as an artifact of ascertainment by geneticists of the mid twentieth century, anticipation is now known to be a genuine feature of dynamic mutation diseases. Anticipation results from the tendency of disease repeats to expand and the inverse correlation between repeat length and age of disease onset.

Several theories have been advanced to explain the molecular basis of DM1, including haploinsufficiency of DMPK, local chromatin effects on neighboring genes, and gain of function exerted by expanded RNAs. DMPK is proposed to have many functions, some of which could relate to the disease, such as modulation of skeletal muscle sodium channels, RNA metabolism, calcium homeostasis, and the cell stress response. Initial expression studies reported a decrease in DMPK RNA and protein levels, but a DMPK knockout mouse designed to test the haploinsufficiency theory exhibited only mild myopathy that was inconsistent with the DM phenotype. The (CTG)n tract is a strong nucleosome assembly site, and it was therefore hypothesized to have trans-effects on the expression of neighboring genes. According to this model, the myriad symptoms of DM1 are caused by disturbances in the expression of multiple nearby genes due to the expansion – making DM1 a "contiguous gene syndrome." There are several genes in close proximity to the *DMPK* gene (i.e. <5 kb) whose roles in DM1 were considered; the most studied of these genes has been SIX5. The SIX5 homolog in Drosophila is required for normal eye development and its mouse homolog is involved in regulating distal limb muscle development. Expression data on SIX5 in

DM1 patients has been inconsistent, however. A Six5 knockout mouse developed cataracts, although they were not of the type observed in DM1. Another gene implicated by this model is *DMWD*, which is expressed in the testis and suspected to be involved in male infertility. Studies indicated that the expression of DMWD is not altered.

The third theory for the molecular basis of DM proposes that an RNA gain-of-function mechanism is responsible. RNA foci, which are accumulations of expanded transcripts, accumulate in the nuclei of patient cells (Fig. 11). The RNA gain-of-function theory was buttressed by the discovery that DM2, a disorder with symptoms almost identical to DM1, is caused by a (CCTG)*n* expansion in intron 1 of the zinc finger 9 protein (ZNF9). The two genes responsible for DM1 and DM2 are unrelated and reside on different chromosomes. Genes near the two loci bear no obvious resemblances. The only striking parallels between the two expansions are: (1) they both contain CTG triplets; and (2) they both occur in transcribed but noncoding regions of the genome. More support for the RNA gain-of-function model came from a mouse model generated by Mankodi et al. in 2000, which contained a (CTG)*n* expansion in the 3′ untranslated region of the skeletal actin (*HSA*) gene. This mouse exhibited myopathy typical of DM1, although the skeletal muscle-restricted expression pattern of HSA precludes broader conclusions. Thus, there are several lines of evidence suggesting that CUG expansion-containing RNAs are capable of causing DM.

RNA-binding proteins and transcription factors colocalize with the RNA foci found in DM, potentially altering nuclear processes. Elevated levels of CUG-containing RNA has been shown to alter gene splicing in specific transcripts which could be relevant to DM: cardiac troponin T (cTNT), involved in cardiomyopathy; Insulin Receptor (IR), involved in diabetes; and Clc-1, the main chloride channel in muscle. Abnormal splicing of cTNT, IR, and ClC-1 are proposed to account for the cardiac abnormalities, insulin insensitivity and myotonia observed in DM.

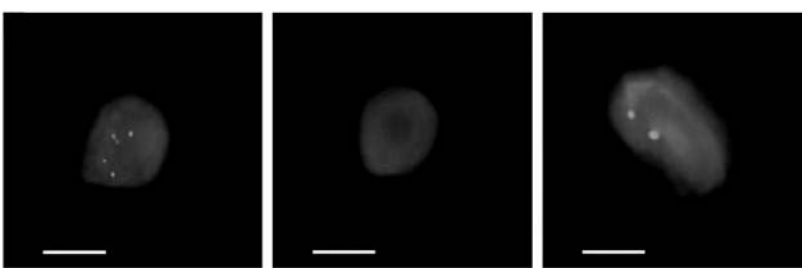

(a) (b) (c)

Fig. 11 *In situ* hybridization of muscle sections with fluorescently labeled antisense oligonucleotide probes reveals accumulation of mutant RNA in DM1 and DM2. (a) Probing of DM2 muscle with a CAGG probe indicates that multiple RNA foci are present. (b) Probing of normal muscle with a CAGG probe demonstrates absence of RNA foci. (c) Probing of DM1 muscle with a CAG probe yields prominent RNA foci. (From Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W., Ranum, L.P. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9, *Science* **293**, 864–867, used with permission of *Science*).

Splicing abnormalities cause a reduction in the membrane concentration of ClC-1 and reduce chloride conductance to levels consistent with myotonia. These spicing alterations are thought to be due to the repeat-expanded RNA's effects on two families of RNA-binding proteins: "CUG-BP1 and ETR-3-like factors" (CELF) and "muscleblind-like proteins" (MBNL). CELF proteins regulate pre-mRNA splicing in cTNT, IR, and ClC-1. In patient tissue and cell culture, CUG-BP1 levels and activity are increased in response to elevated levels of CUG-containing RNA. This may be due to a lengthening of the protein's half-life. MBNL proteins were named after their Drosophila ortholog *muscleblind*, which is required for photoreceptor and muscle differentiation in flies. The three known proteins in this family (MBNL1 (MBNL), MBNL2 (MBLL), and MBNL3 (MBXL)) are splicing regulators that are thought to act antagonistically to CELF family proteins. They colocalize with RNA foci *in vivo*, and a mouse model lacking specific isoforms of MBNL1 recapitulates the myotonia, cataracts, and splicing dysregulation observed in DM.

The splicing alterations seen in DM patients are consistent with loss of function of MBNL proteins or an increase in CELF protein activity in muscle and brain. Because of the colocalization of MBNL proteins with RNA foci, it has been proposed that sequestration and subsequent depletion of these proteins from the cellular milieu is responsible for the symptoms of DM. Although this process probably plays a critical role in the disease, recent evidence suggests it is not solely responsible. The most plausible current theory explaining DM pathogenesis proposes an imbalance between the antagonistic MBNL and CELF proteins, resulting in specific splicing abnormalities that cause DM's diverse range of symptoms.

5.2 Myotonic Dystrophy Type 2

Myotonic dystrophy type 2 (DM2) is an autosomal dominant, multisystem disorder very similar to DM1. The majority of its symptoms resemble DM1: progressive weakness, myotonia, cardiac disturbances, iridescent cataracts, and insulin insensitivity. There are some notable differences, however. DM2, unlike DM1, predominantly affects proximal muscles at its onset, which is why many cases were originally classified as proximal myotonic myopathy (PROMM). Other interesting differences are that mental retardation is not observed in DM2, DM2 patients show increased sweating, and DM2 congenital forms have not been observed. DM2 is caused by a (CCTG)*n* tetranucleotide expansion in the first intron of the zinc finger protein 9 (*ZNF9*) gene on chromosome 3q21.3. The tetranucleotide repeat can expand to stunning lengths, with the longest cases comprising 44 kb of DNA, making them the longest tracts observed in the repeat expansion disorders. The DM2 locus exhibits marked somatic instability. Over the course of a patient's lifetime, the average repeat length increases substantially, as judged by blood drawings.

The similarities between DM1 and DM2 are not restricted to the clinical presentation. The expansions are both large, CTG-containing tracts that are transcribed, but not translated. Both disorders cause nuclear RNA foci that sequester specific RNA-binding proteins (Fig. 11), including CELF and MBNL family members. Similar splicing abnormalities are also observed in DM2. Since the genes associated with the repeat expansions in DM1 and DM2 are

unrelated, it is likely that the cause of both disorders is a toxic gain of function of long tracts of CUG-containing RNA. A mouse model generated by Mankodi et al. in 2000, which contained a (CTG)*n* expansion in the 3′ untranslated region of the skeletal actin (*HSA*) gene, exhibited myopathy typical of DM. Thus, even outside the context of the DM1 and DM2 loci, CTG-containing expansions can cause DM-like symptoms. This supports the RNA gain-of-function theory, as transcribed CTG-containing expansions are inherently capable of causing disease. Differences between DM1 and DM2 may be due to regional or temporal expression patterns or differences in the affinity of RNA-binding proteins for CTG or CCTG tracts.

5.3 Spinocerebellar Ataxia Type 8

Spinocerebellar ataxia type 8 (SCA8) is a dominantly inherited cerebellar ataxia. Affected individuals suffer from late-onset, slowly progressing gait ataxia as well as dysarthria, oculomotor incoordination, spasticity, and decreased vibration sense. The cerebellar cortices and vermis undergo a slowly progressing but dramatic atrophy, while the brainstem exhibits little evidence of degeneration. Patients may become wheelchair-bound as early as their fourth decade of life. Using a direct method for cloning expanded triplet repeats (i.e. RAPID cloning), a gene containing an expanded CTG tract was identified in a large family (MN-A) segregating the SCA8 phenotype. Interestingly, this CTG is contained within a gene on the long arm of chromosome 13 that is transcribed, but apparently not translated into a protein product. Thus, it was proposed that the production of an RNA transcript containing this expanded CTG repeat tract is the cause of SCA8. Numerous subsequent studies, however, have indicated that possession of the expanded CTG repeat tract appears necessary, but not sufficient, for the production of the SCA8 phenotype. Thus, reduced penetrance is viewed as a key feature of SCA8 at this time. On the basis of the available genetic data, normal individuals always carry fewer than 70 CTG repeats, while affected SCA8 patients can carry anywhere from 71 to >1000 CTG repeats. As extremely large CTG repeat alleles (>800 triplets) were shown to not cause disease in early reports, this was initially attributed to lack of stability of the mutant RNA product. However, further work has revealed considerable overlap between disease-causing CTG repeat expansions and nonpenetrant CTG repeat alleles, indicating that a secondary factor – either a trans-acting genetic factor or an environmental factor – must be present to yield the SCA8 disease phenotype. In the case of the original MN-A family, evidence for a *cis* modifier appears responsible for the extremely high penetrance in this large pedigree. Thus, while the causality of the CTG repeat expansion in SCA8 has been somewhat controversial, review of the current literature suggests that the CTG repeat expansion is directly involved in SCA8, but may not alone be sufficient to produce the disorder.

Another unique feature of the SCA8 CTG repeat is its extreme and unusual genetic instability. Paternal transmissions generally result in contractions, while maternal transmissions generally result in expansions. Expansions of up to 600 repeats have occurred in one generation through maternal transmission. Large deletions in expanded alleles often occur in sperm cells, offering an explanation for the paternal contraction bias. Pathogenic expansions

often have 5′ triplet interruptions, the role of which is unclear.

The molecular basis of how the expanded SCA8 RNA causes disease remains unclear. The transcript has been detected exclusively in the brain, and appears to be transcribed in the CTG orientation. Translation of a polyglutamine tract from the CAG-containing transcript in the opposite direction is precluded by the existence of stop codons flanking the repeat. The longest transcript identified to date contains 6 exons, and many alternatively spliced forms of the gene have been detected, although all these isoforms are expressed at very low levels. None contain significant open reading frames. In 1999, Koob et al. reported that a gene partially overlaps the 5′ end of the *SCA8* gene locus on the antisense strand. This gene, called *Kelch-like 1 (KLHL1)*, for its homology to the Drosophila *KELCH* gene, is highly conserved and predicted to encode an actin-binding protein of 748 amino acids. Its expression overlaps with that of the SCA8 transcript, suggesting that the *SCA8* gene may produce an antisense RNA whose function is to regulate KLHL1 expression. Whether or how this is occurring remains uncertain at this time.

With the discovery of the DM RNA toxic gain-of-function pathway, an emerging theory for SCA8 pathology holds that the production of a CUG-expanded transcript results in RNA gain-of-function toxicity within the restricted neuronal populations where the *SCA8* gene is expressed. In support of this hypothesis, transgenic mice derived with a human bacterial artificial chromosome containing the entire *SCA8* gene with a 118 CTG repeat expansion develop neurological disease. The severity of their phenotype depends on the expression level of the transgene. In the more moderate expressing lines, the SCA8 mice display a slowly progressive gait ataxia reminiscent of the human disease. While no protein product has been detected, 1C2 and ubiquitin antibody staining reveal intranuclear inclusions in cerebellar neurons, consistent with the RNA-containing inclusions observed in FXTAS brains. An alternative interpretation is that translation of CTG into a polyleucine tract is occurring. Further work will be needed to distinguish these two possibilities.

Studies of the SCA8 CTG repeat expansion in *D. melanogaster* have also been informative. Whether directing expression of the *SCA8* gene to fly retina with a normal CTG repeat tract or an expanded CTG repeat tract, a neurodegenerative eye phenotype results. Modifier screens using fly stocks carrying mutations in genes encoding RNA-binding protein yielded a number of genes that could either enhance or suppress this retinal degeneration phenotype. Interestingly, Drosophila *muscleblind*, whose mammalian counterpart has been implicated in the DM RNA toxicity pathway, modified the retinal degeneration caused by expanded SCA8 CTG repeat expression more so than the retinal degeneration caused by normal SCA8 CTG repeat expression. Such data supports the hypothesis that the SCA8 CTG repeat expansion is producing neurotoxicity by altering the function of RNA-binding proteins as in DM. Although it would be premature to definitively conclude that SCA8 is an RNA gain-of-function repeat disease, provisional classification in this category is appropriate.

5.4 The Fragile X Tremor – Ataxia Syndrome (FXTAS)

Fragile X tremor-ataxia syndrome (FXTAS) is a rare and unusual disorder

that is associated with premutation of the FRAXA locus in males. Patients develop a progressive ataxia and intention tremor. This is sometimes accompanied by dementia, parkinsonism, and autonomic dysfunction. Neuropathological changes include degeneration of the cerebellum and ubiquitin-positive intranuclear inclusions in neurons and glia. FXTAS is associated with the expansion of a (CGG)*n* tract in the 5′-untranslated region of the *FMR1* gene, found on chromosome Xq27.3. Expansion of this tract beyond 200 repeats causes FRAXA, a disorder resulting from a reduction in the expression of FMR1. Males with tracts between 55 and 200 CGG repeats are considered to be in the "premutation range," and are now considered to be at risk for FXTAS beyond middle age.

In contrast to FRAXA, FXTAS patients show increased expression of the FMR1 transcript, with longer repeats corresponding to higher transcript levels. Premutation carriers with more than 100 repeats have an average of five times more FMR1 mRNA than individuals in the normal repeat range. Despite the increase in transcript level, the protein product of the *FMR1* gene, FMRP, is slightly reduced in FXTAS patients. Since elevated transcript

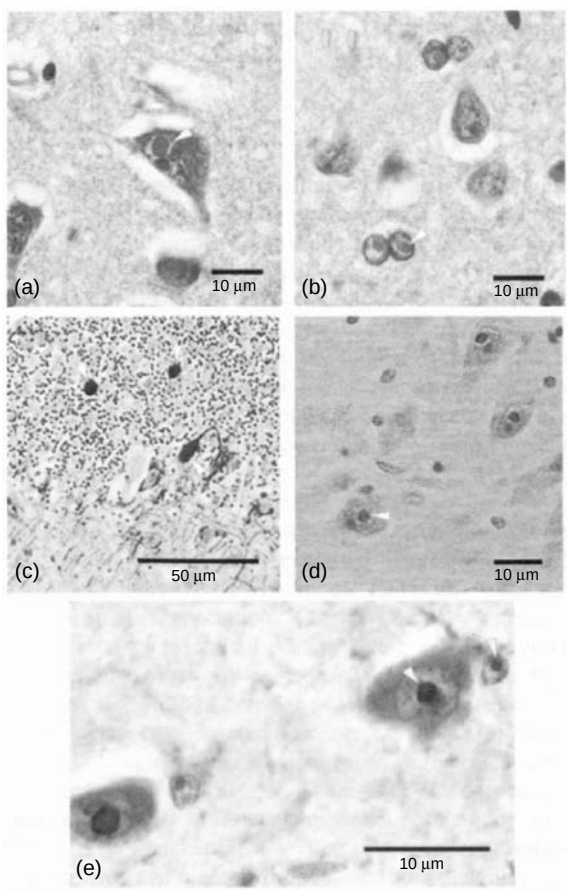

Fig. 12 Intranuclear inclusions are present in the brains of FXTAS patients. (a) Hematoxylin & eosin staining of cerebral neurons reveals a refractile, eosiniphilic nuclear inclusion of ~5 μM in diameter (white arrowhead). (b) Hematoxylin & eosin staining of cerebral astrocytes reveals refractile, eosiniphilic nuclear inclusions of ~2 μM in diameter (white arrowhead). (c) Antineurofilament antibody staining of cerebellum demonstrates presence of dystrophic neurites, consistent with ongoing Purkinje cell degeneration. (d) Antiubiquitin antibody staining of cerebral neurons labels intranuclear inclusions. (e) Antiubiquitin positive intranuclear inclusions form in both neurons (white arrowhead; larger cell) and astrocytes (white arrowhead; smaller cell). (From Greco et al. (2002) *Brain* **125**, 1760, used with permission).

levels are present without great alteration in protein levels, this disorder may be due to an RNA gain of function of the FMR1 transcript. Indeed, immunostaining of FXTAS patients' brain sections reveal intranuclear inclusions in neurons and glia, believed to be comprised of accumulated Fmr1 RNA transcripts and various proteins (Fig. 12). To determine the molecular basis of FXTAS, a knockin mouse model of FXTAS was generated using a human 98 CGG repeat, and was noted to produce elevated levels of the Fmr1 transcript and display intranuclear ribonucleoprotein inclusions, containing ubiquitin, Hsp40 and the 20S proteasome subunit. The inclusions may be a response to RNA toxicity or a result of the aggregation of CGG-binding proteins. While the mechanism of FXTAS may be due to a dominant gain of function of triplet-expanded RNA on RNA-binding proteins as in DM, there is currently no direct evidence supporting this theory.

6 Type 4: The Polyalanine Diseases

6.1 Overview

Polyalanine disorders are characterized by small expansions in trinucleotide repeats encoding alanine tracts. Most are rare, autosomal dominant developmental malformations. Although the normal function of the alanine tract is unknown, they tend to be found in transcription factors. The majority of alanine tract expansions associated with disease loci are also located in transcription factors, most of which play a role in development. In contrast to polyglutamine expansions, polyalanine expansions are small and stable, not exceeding 30 triplets. Most are composed of "imperfect" alanine tracts, including combinations of GCG, GCA, GCC, and GCT codons. This composition argues against an expansion mechanism involving single-strand, hairpin-like secondary structures as in other repeat disorders. A more plausible mechanism for polyalanine expansion is unequal crossing-over, in which alleles mispair during meiotic crossing-over, resulting in one expanded and one contracted tract.

6.2 Oculopharyngeal Muscular dystrophy

Oculopharyngeal muscular dystrophy (OPMD) is a predominantly autosomal dominant, late-onset disorder characterized by progressive drooping eyelids, dysphagia, and proximal limb weakness. Certain skeletal muscles in affected patients degenerate and contain nuclear inclusions and rimmed vacuoles. This effect is particularly striking in the levator palpebra and pharyngeal muscles, responsible for lifting the eyelids and swallowing. OPMD is caused by an alanine expansion in the gene polyadenine-binding protein 2 (PABP2). PABP2 normally contains a 10-alanine tract encoded by $(GCG)_6(GCA)_3GCG$. Affected individuals carry 12–17 alanines, probably resulting from unequal crossing-over of the two alleles. Repeat length appears to correlate with disease severity.

PABP2 is an abundant, ubiquitously expressed pre-mRNA-binding protein that plays a role in controlling the formation and length of mRNA polyA tails. NIs immunoreactive for PABP2 are present in the skeletal muscle cells of OPMD patients and sequester polyA-containing transcripts. The aggregates are filamentous and contain PABP2, ubiquitin and

proteasome subunits, leading investigators to conclude that expansion of the alanine tract probably causes PABP2 to misfold and aggregate. Aggregation in OPMD has been linked to toxicity in several experimental systems. Polyalanine-expanded PABP2 aggregates and causes cell death in cultured cells and transgenic mice. Indeed, widespread expression of the human *PABP2* gene in transgenic mice yielded polyalanine length-dependent muscle pathology, including rimmed vacuoles, central nuclei, and numerous dystrophic changes (Fig. 13). Deleting the C-terminal oligomerization domain, overexpressing chaperones, or exposing cells to aggregation inhibitors such as Congo red and doxycycline, reduces aggregate formation and toxicity in cell culture. A polyalanine-expanded peptide, similar to the amino-terminal region of PABP2 was shown to adopt a β-sheet conformation, whereas the same peptide with 7 alanines adopted an α-helical conformation. This is reminiscent of polyglutamine proteins, which also adopt a β-sheet conformation when the polyglutamine tract exceeds a threshold of about 35 glutamines. Expanded polyalanine proteins also activate caspase-3 and -8 in cultured cells, another feature reminiscent of polyglutamine toxicity.

In 2004, Wirtschafter et al. proposed a model for the selective vulnerability of

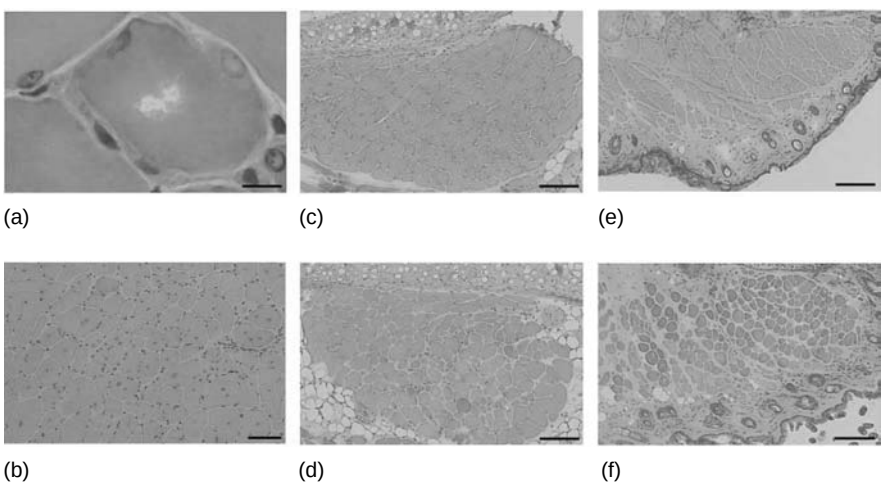

Fig. 13 Mice expressing polyalanine-expanded PABP2 protein display prominent muscle pathology consistent with the OPMD phenotype. Transgenic mice were generated using the pCAGGS expression vector that consists of the chicken beta-actin promoter and a CMV enhancer, and thus ubiquitously express either human PABP2 protein with six alanines (c, e) or nine alanines (a, b, d, f) at roughly comparable levels. Hematoxylin & eosin staining of sections from the soleus muscle (a, b), the pharynx muscle (c, d), and the eyelid muscle (e, f) reveal normal muscle histology in the six-alanine PABP2 expressing mice (c, e). However, mice expressing PABP2 with nine alanines display nonuniform muscle fiber size and prominent connective tissue in pharynx and eyelid musculature (d, f). At high power, cytoplasmic vacuoles, reminiscent of OPMD "rimmed vacuoles" are apparent in soleus muscle from the nine-alanine PABP2 expressing mice (a). Low power examination further indicates frequent central nuclei in the soleus muscle from such mice (b). (From Hino et al. (2004) *Hum. Mol. Genet.* **13**, 181, used with permission).

extraocular muscles in OPMD. Unlike other skeletal muscles, extraocular muscles are not postmitotic and they continually undergo remodeling. This requires the frequent upregulation of genes involved in cell cycling and protein synthesis. Failure of correct mRNA polyadenylation or transport in these cells may have a cumulative toxic effect resulting in the progressive degeneration of these muscles. The mechanism of OPMD could be due to an interference with polyadenylation, disturbances in intracellular trafficking of mRNA, or toxicity due to the aggregation of misfolded and/or aggregated PABP2 species. Further research will be required to distinguish between these possibilities.

6.3 Synpolydactyly (Syndactyly Type II)

Synpolydactyly (SPD) is a rare, autosomal dominant developmental disorder characterized by fused and extra digits (syndactyly and polydactyly, respectively). It is caused by the expansion of a polyalanine tract in the amino-terminal region of the transcription factor HOX-D13. *HOX* genes act in concert to coordinate axial patterning in animals. Tracts of 7–14 alanines in HOX-D13 have been linked to SPD, and disease severity is proportional to repeat length. Multiple studies support a "dominant-negative" role for expanded HOX-D13 protein in SPD. Mice null for Hox-d13 have a phenotype less severe than SPD, while mice with alanine tract expansions have a form that more closely resembles the disorder. Mice lacking Hox11, Hox12, and Hox13 have a phenotype similar to SPD, suggesting that the alanine expansion in SPD antagonizes the function of other *HOX* genes. This is supported by a genetic complementation study. In further support of this theory, humans with suspected loss-of-function mutations in the *HOX-D13* gene do not have a phenotype consistent with SPD.

6.4 Cleidocranial Dysplasia

Cleidocranial dysplasia (CCD) is a rare, autosomal dominant developmental disorder characterized by holes in the skull, dental malformations, absent or hypoplastic clavicles and maxillae, and other skeletal malformations. The primary cause of CCD is thought to be loss-of-function mutations in the gene *RUNX2*. This gene affects osteoblast differentiation and is a member of the Runt family of transcription factors. In one family, phenotypically distinct from classic CCD, an expansion from 17 to 27 alanines in RUNX2 has been detected. This family exhibited brachydactyly and a mild CCD phenotype. The difference in phenotype between the polyalanine expansion mutation family and typical presumed haploinsufficient, loss-of-function CCD patients supports a gain-of-function effect of the expanded alanine tract in this atypical family. Contraction of the tract in RUNX2 is common and does not cause a detectable phenotype.

6.5 Holoprosencephaly

Holoprosencephaly (HPE) is a common developmental malformation resulting in partial or full cyclopia, failure to develop midline structures in the ventral forebrain, and prenatal lethality. In rare cases, HPE is caused by the heterozygous expansion of a 15-amino acid alanine tract in the protein ZIC2. This protein is one member of a family of zinc finger proteins believed to regulate neurulation,

left–right axis formation and other developmental processes. Other individuals with heterozygous loss-of-function mutations have a phenotype indistinguishable from those with alanine tract expansions. Partial loss of function of Zic2 in mice similarly causes developmental abnormalities similar to HPE. Therefore, expansion of the alanine tract likely causes loss of function of ZIC2 in the case of HPE.

6.6 Hand-foot-genital Syndrome

Hand-foot-genital syndrome (HFGS) is a rare, dominantly inherited developmental abnormality characterized by malformation of the distal limbs and lower urogenital tract. Short thumbs, short great toes, and abnormal carpals and tarsals are some of its salient features. Some HFGS patients have alanine tract expansions in the protein HOX-A13, which is in the same family as the SPD-associated protein HOX-D13. HOX-A13 contains three alanine tracts, the most C-terminal of which is the most commonly mutated. Expansions enlarge the second or third tract by 6 to 9 alanines. In humans, deletion of HOX-A13 causes a phenotype that is mild in comparison to HFGS caused by alanine tract mutations. Hoxa13-null mice also have a milder phenotype than mice carrying a frameshift deletion suspected to confer gain of function. This evidence suggests a dominant-negative mechanism for HFGS caused by alanine tract expansions.

6.7 Blepharophimosis-ptosis-epicanthus Inversus Syndrome

Blepharophimosis-ptosis-epicanthus inversus syndrome (BPES) is a rare, autosomal dominant developmental disorder resulting in malformation of the upper eyelids and forehead, and occasionally premature ovarian failure in women. It is caused by a number of different mutations in the gene *Forkhead L2* (*Foxl2*), the most common of which is expansion of its carboxy-terminal 14-amino acid alanine tract. FOXL2 is a highly conserved transcription factor whose role in the ovary has been studied most thoroughly. It is expressed in the ovaries and eyelids during development and adulthood. Studies in which Foxl2 have been ablated in mice show that it is important for the differentiation of ovarian granulosa cells, and its absence causes accelerated follicle cell depletion leading to POF. Knockout mice also display craniofacial abnormalities consistent with BPES, suggesting that a dominant-negative mechanism may be responsible in humans.

6.8 Syndromic and Nonsyndromic X-linked Mental Retardation

Several loosely related disorders are associated with alanine tract expansions in the Aristaless related homeobox (ARX) protein on chromosome Xp22.13. One is nonsyndromic X-linked mental retardation (XLMR), a heterogeneous condition in which mental retardation is the main consistent feature. Several syndromic XLMR disorders linked to alanine expansion in ARX include West syndrome (WS) and Partington syndrome (PRTS). WS causes progressive mental retardation with abnormal EEG and infantile seizures, while PRTS causes mental retardation, dysarthria, and dystonic movements of the hands. Brain anatomy appears normal in these disorders.

ARX is a paired-class homeodomain protein that is expressed in the ventricular

and marginal zones of the developing mouse brain and continues to be expressed in adult cortex. It is suspected to play a role in neuroepithelial cell differentiation and maintenance of neuronal subtypes in the adult cortex. ARX-null mice have small brains and neuronal migration deficits. Expansions of two different alanine tracts in ARX cause XLMR: a 12-alanine tract at amino acids 144–155, and a second tract at amino acids 100–115. Both sites cause highly variable forms of XLMR, with expansions in the former alanine tract resulting in nonsyndromic XLMR, WS or PRTS. Alanine tract expansions in ARX probably result in partial loss of function of the protein. Heterozygous female carriers of expanded ARX do not exhibit XLMR, suggesting that one normal copy of the protein is sufficient for normal cognitive ability. Also, humans with null mutations of ARX suffer from X-linked lissencephaly with abnormal genitalia (XLAG), a much more severe condition causing major developmental abnormalities in the brain and genitalia. Thus, XLMR due to alanine expansions in ARX may be because of partial loss of function leading to subtle developmental defects and/or failure to maintain specific neuronal populations.

6.9
Congenital Central Hypoventilation Syndrome

Congenital central hypoventilation syndrome (CCHS) is a rare, autosomal dominant disorder causing a failure of autonomic control of breathing. It attenuates or abolishes responses to hypercarbia and hypoxemia. In the majority of cases, it is caused by the expansion of one of two alanine tracts in the protein PHOX-2B. Mutations expand the 20-residue tract to 25–29 alanines. PHOX-2B is a paired-class transcription factor containing a homeodomain. A loss-of-function mutation in murine Phox2b is homozygous lethal, and specifically prevents the development of parasympathetic ganglia. Heterozygous mice show chronic pupil dilation but no parasympathetic or respiratory disturbances. Also, a patient hemizygous for a 5-Mb deletion including Phox2b does not have CCHS. No cases of CCHS have been reported in which PHOX2B is truncated before the homeobox domain. This evidence suggests that the expansion of alanine tracts in PHOX2B may cause a subtle, dominant-negative effect on the development of respiratory control pathways.

7
Unclassified Repeat Diseases Lacking Mechanistic Explanations

7.1
Spinocerebellar Ataxia Type 10

Spinocerebellar ataxia type 10 (SCA10) is an autosomal dominant, progressive ataxia that exhibits nearly pure cerebellar signs. It appears restricted to individuals of Mexican ethnicity. SCA10 patients typically present with gait ataxia, followed by dysarthria, dysphagia, and ocular dysmetria, and most patients also experience recurrent motor seizures. Cerebellar atrophy is the most prominent neuropathological change. SCA10 is caused by the expansion of a highly polymorphic pentanucleotide repeat, (ATTCT)*n*, on chromosome 22. Unaffected individuals carry 10–22 repeats. While expanded alleles have not been successfully PCR amplified, transcript sizes on Northern blots and fragment sizes on Southern blots indicate that disease alleles can range from ~800 to 4500 repeats. Anticipation occurs in SCA10, independently

supporting the causal nature of the repeat expansion. The ATTCT repeat is located in intron 9 of the *SCA10* gene, a previously unrecognized 66-kb gene that encodes a novel putative 475 amino acid protein of unknown function with few recognizable motifs or domains. The 2-kb SCA10 transcript is ubiquitously expressed though the highest levels of expression occur in the brain, testis, and adrenal glands. Within the brain, it is most highly expressed in the cerebellum and associated structures. The carboxy-terminal portion of the protein appears to contain "armadillo repeats," which are responsible for the membrane association of β-catenins. Ataxin-10 does not seem to associate with membranes, however. The entire protein is highly conserved between humans and rodents, and potential orthologs exist in Arabidopsis and Drosophila.

There are several proposed mechanisms for SCA10 pathogenesis. The most obvious hypothesis is that the large intronic expansion affects ataxin-10 expression, perhaps by altering local chromatin structure. siRNA knockdown of ataxin-10 in cell culture experiments yields higher rates of cell death in cerebellar neurons than in cortical neurons. However, the expression levels of the SCA10 transcript are not reduced in patient's lymphoblast cells, arguing against simple haploinsufficiency. Consequently, RNA gain-of-function toxicity has been proposed as the potential mechanism; however, there are no further data at this time to support such a hypothesis. Additional work will need to be done to distinguish between these and other possibilities, and will need to account for the cell-type specific pattern of neurodegeneration that occurs in the face of apparently widespread expression of the *SCA10* gene mutation.

7.2 Spinocerebellar Ataxia Type 12

Among the more recent additions to the unstable repeat disease group is spinocerebellar ataxia type 12 (SCA12), a rare, autosomal dominant disorder that may be most prevalent in Indian populations. Its symptoms are distinct from the other SCAs, typically beginning with an action tremor of the upper extremities and progressing to include hyperreflexia, mild cerebellar dysfunction, bradykinesia, increased muscle tone, psychiatric symptoms, and dementia. The brains of SCA12 patients likely undergo a slow, generalized atrophy that is most prominent in the cortex, but also results in loss of Purkinje cells in the cerebellum. Disease onset typically occurs in the third or fourth decade, and a gradually progressive disease course is typical.

SCA12 is caused by a (CAG)*n* expansion in chromosome 5q31 – q33. The expansion is 5′ to the *PPP2R2B* gene, encoding a regulatory subunit of the protein phosphatase 2A enzyme (PP2A). This gene has many transcription start sites, some of which include the repeat (but many of which do not). GENSCAN predicts an exon including the expansion that would encode a polyserine tract, but this prediction is of low probability and considered unlikely. Other evidence suggests that the expansion is located in the promoter region, and this is currently the most widely accepted view. Unaffected individuals carry 7–32 CAG repeats while affected individuals carry 55–78 CAG triplets. The most common repeat size in unaffected individuals is 10 CAGs. The repeat is fairly stable, with only modest expansions and contractions resulting equally from maternal and paternal transmission. A significant correlation between repeat length and age of onset has not been documented.

The protein PP2A is an essential serine/threonine phosphatase expressed in all known eukaryotic cells. It is involved in diverse cellular functions, including cell growth, differentiation, DNA replication, neurotransmitter release, and apoptosis. PPP2R2B is a brain-specific regulatory subunit of PP2A. The class of regulatory subunits including PPP2R2B may affect PP2A's phosphatase activity for certain substrates, including histone-1, vimentin, and tau. It may also affect PP2A's subcellular localization.

There are several possible explanations for SCA12 pathogenesis. The first is that expanded PPP2R2B may generate a polyamino acid tract-containing protein, resulting in toxicity. Northern blots probing for sequence flanking the CAG repeat did not detect a PPP2R2B transcript, indicating that if a repeat-containing exon exists and is transcribed, it is not present at appreciable levels. Nevertheless, the possibility of polyglutamine, polyserine, or polyalanine toxicity, though unlikely, cannot be completely ruled out at this time. A second possible cause of SCA12 pathogenesis is RNA gain of function, as occurs in DM and FXTAS. The repeat expansions are smaller in SCA12 than in DM, however, and more importantly, CAG tract-containing transcripts appear to be rather scarce. A third possibility is altered splicing of the PPP2R2B transcript. Several different amino-termini are possible, the ratio of which may affect PP2A's subcellular localization. Yet another theory of SCA12 pathogenesis is that the expansion affects PPP2R2B transcript levels. Repeat expansion causes a substantial increase in PPP2R2B expression as measured in reporter assays using a neuroblastoma cell line. Altered levels of the protein could affect PP2A's specificity or subcellular localization. This has the potential to disturb a multitude of processes in the CNS. At this time, all of the above theories of SCA12 CAG repeat expansion neurotoxicity remain plausible.

7.3 Huntington's Disease Like 2 (HDL2)

Perhaps the most exciting and enigmatic recent discovery in the repeat expansion field is that of the mutational basis of a disorder known as Huntington's disease like 2 or HDL2. HDL2 is so named because it is in essence a genocopy of classical HD, as the original HDL2 pedigree was labeled with a diagnosis of HD until HD CAG repeat testing indicated that this family's HD-like disease did not result from a CAG repeat expansion in the *htt* gene. HDL2 patients present with weight loss and diminished coordination, and then develop tremors, dysarthria, hyperreflexia, and rigidity. Patients display psychiatric involvement, chorea, and dystonia, and ultimately become demented. Death occurs 15–25 years after onset, when patients become bedridden as in typical HD cases. MRI findings reveal marked atrophy of the caudate and of the cerebral cortex, making the neuropathology indistinguishable from classic HD. HDL2 patients do not show cerebellar signs or neuropathology.

As soon as it was found that HDL2 patients do not have the HD CAG repeat expansion, direct cloning methods for triplet repeat expansions of the CAG/CTG type were applied to patient samples and an expanded CAG/CTG repeat ($n = 55$) was isolated. Sequence flanking this repeat indicated that HDL2 is caused by CTG repeat expansions in the *junctophilin-3* (*JPH3*) gene, one of a family of structural proteins whose function is to link

Ca++ channels on the ER with voltage sensors on the plasma membrane. Analysis of JPH3 CTG repeat indicates that normal individuals typically carry 7–27 CTG repeats while affected HDL2 patients usually have expansions of 50–60 CTG repeats. Widespread screening of HD-like patients from around the globe suggests that the HDL2 JPH3 CTG repeat expansions are most common in individuals of African ethnicity.

The question of how the JPH3 CTG repeat expansion causes HDL2 remains unknown, but there are at least four possible explanations. These alternative (but not mutually exclusive) theories stem from the documented alternative processing of the *JPH3* gene, which permits the prediction of the CTG repeat tract as: (1) part of intron 1; (2) as part of the 3′ untranslated region; or (3) as encoding either a polyleucine or polyalanine tract. One theory is that haploinsufficiency of the Ca++ regulating brain- and testes-specific junctophilin-3 protein is responsible for HDL2. While JPH3 knockout mice display motor incoordination, no histological abnormalities are found in their brains. Further work with these mice is ongoing to evaluate this hypothesis. Evidence against simple haploinsufficiency has come from study of HDL2 patient's brain material, however. 1C2 and ubiquitin antibody immunostaining reveal intranuclear inclusions in neurons throughout the brains of these patients, with dramatic similarity in distribution to patients with classic HD (Fig. 14). At this time, the molecular basis of HDL2 is unknown; however, the incredible overlap between HD and HDL2 in terms of clinical phenotype and neuropathology strongly suggests that solving HDL2 should have profound implications for our mechanistic understanding of HD.

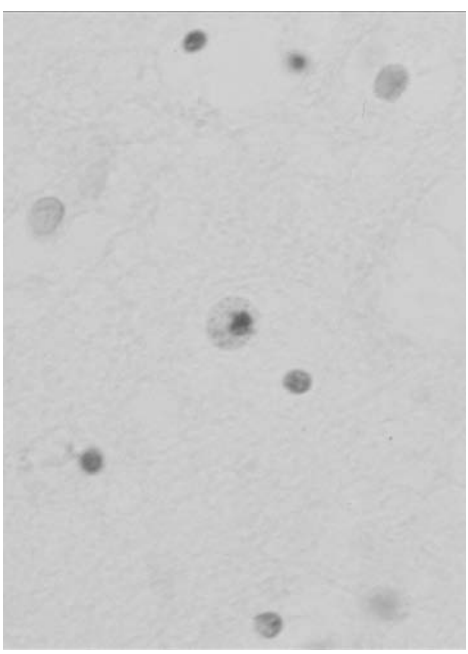

Fig. 14 HDL2 patients have 1C2-positive intranuclear inclusions. 1C2 staining of cerebral cortex (frontal lobe) from an HDL2 patient reveals a prominent intranuclear inclusion that resembles the intranuclear inclusions seen in classic HD patients. As the 1C2 antibody is directed against expanded, misfolded polyglutamine tracts, which are not predicted to be expressed from the causal *HDL2* gene, the explanation for the presence of such nuclear inclusions in HDL2 patients remains unknown. (From Margolis, R.L., O'Hearn, E., Rosenblatt, A., Willour, V., Holmes, S.E., Franz, M.L., Callahan, C., Hwang, H.S., Troncoso, J.C., Ross, C.A. (2001) A disorder similar to Huntington's disease is associated with a novel CAG repeat expansion, *Ann. Neurol.* **50**, 373–380, used with permission of *Annals of Neurology*, and John Wiley & Sons, publisher).

See also Genetics, Molecular Basis of; Motor Neuron Diseases: Cellular and Animal Models; Motor Neuron Diseases: Molecular Mechanism, Pathophysiology, and Treatments; Noncoding Tandemly Repeated DNA Sequences.

Bibliography

Books and Reviews

Brown, L.Y., Brown, S.A. (2004) Alanine tracts: the expanding story of human illness and trinucleotide repeats, *Trends Genet.* **20**, 51–58.

Harper, P.S. (2001) *Myotonic Dystrophy*, Saunders, London, UK.

Harper, P.S., Perutz, M. (2001) *Glutamine Repeats and Neurodegenerative Diseases: Molecular Aspects*, Oxford University Press, New York.

Jin, P., Alisch, R.S., Warren, S.T. (2004) RNA and microRNAs in fragile X mental retardation, *Nat. Cell Biol.* **6**, 1048–1053.

Ranum, L.P., Day, J.W. (2004) Pathogenic RNA repeats: an expanding role in genetic disease, *Trends Genet.* **20**, 506–512.

Wells, R.D., Warren, S.T. (1998) *Genetic Instabilities and Hereditary Neurological Diseases*, Academic Press, San Diego, CA.

Primary Literature

Albrecht, M., Golatta, M., Wullner, U., Lengauer, T. (2004) Structural and functional analysis of ataxin-2 and ataxin-3. *Eur. J. Biochem.* **271**, 3155–3170.

Amiel, J., Laudier, B., Attie-Bitach, T., Trang, H., de Pontual, L., Gener, B., Trochet, D., Etchevers, H., Ray, P., Simonneau, M. et al. (2003) Polyalanine expansion and frameshift mutations of the paired-like homeobox gene *PHOX2B* in congenital central hypoventilation syndrome, *Nat. Genet.* **33**, 459–461.

Arrasate, M., Mitra, S., Schweitzer, E.S., Segal, M.R., Finkbeiner, S. (2004) Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death, *Nature.* **431**, 805–810.

Banfi, S., Servadio, A., Chung, M., Capozzoli, F., Duvick, L.A., Elde, R., Zoghbi, H.Y., Orr, H.T. (1996) Cloning and developmental expression analysis of the murine homolog of the spinocerebellar ataxia type 1 gene (*Sca1*), *Hum. Mol. Genet.* **5**, 33–40.

Brais, B., Bouchard, J.P., Xie, Y.G., Rochefort, D.L., Chretien, N., Tome, F.M., Lafreniere, R.G., Rommens, J.M., Uyama, E., Nohira, O. et al. (1998) Short GCG expansions in the *PABP2* gene cause oculopharyngeal muscular dystrophy, *Nat. Genet.* **18**, 164–167.

Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.P., Hudson, T. et al. (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member, *Cell.* **69**, 385.

Brown, V., Jin, P., Ceman, S., Darnell, J.C., O'Donnell, W.T., Tenenbaum, S.A., Jin, X., Feng, Y., Wilkinson, K.D., Keene, J.D. et al. (2001) Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome, *Cell.* **107**, 477–487.

Brown, L.Y., Odent, S., David, V., Blayau, M., Dubourg, C., Apacik, C., Delgado, M.A., Hall, B.D., Reynolds, J.F., Sommer, A. et al. (2001) Holoprosencephaly due to mutations in ZIC2: alanine tract expansion mutations may be caused by parental somatic recombination, *Hum. Mol. Genet.* **10**, 791–796.

Burnett, B., Li, F., Pittman, R.N. (2003) The polyglutamine neurodegenerative protein ataxin-3 binds polyubiquitynated proteins and has ubiquitin protease activity, *Hum. Mol. Genet.* **12**, 3195–3205.

Calado, A., Tome, F.M., Brais, B., Rouleau, G.A., Kuhn, U., Wahle, E., Carmo-Fonseca, M. (2000) Nuclear inclusions in oculopharyngeal muscular dystrophy consist of poly(A) binding protein 2 aggregates which sequester poly(A) RNA, *Hum. Mol. Genet.* **9**, 2321–2328.

Chen, S., Peng, G.H., Wang, X., Smith, A.C., Grote, S.K., Sopher, B.L., La Spada, A.R. (2004) Interference of Crx-dependent transcription by ataxin-7 involves interaction between the glutamine regions and requires the ataxin-7 carboxy-terminal region for

nuclear localization, *Hum. Mol. Genet.* **13**, 53–67.

Chen, H.K., Fernandez-Funez, P., Acevedo, S.F., Lam, Y.C., Kaytor, M.D., Fernandez, M.H., Aitken, A., Skoulakis, E.M., Orr, H.T., Botas, J. et al. (2003) Interaction of Akt-phosphorylated ataxin-1 with 14-3-3 mediates neurodegeneration in spinocerebellar ataxia type 1, *Cell.* **113**, 457–468.

Cummings, C.J., Reinstein, E., Sun, Y., Antalffy, B., Jiang, Y., Ciechanover, A., Orr, H.T., Beaudet, A.L., Zoghbi, H.Y. (1999) Mutation of the E6-AP ubiquitin ligase reduces nuclear inclusion frequency while accelerating polyglutamine-induced pathology in SCA1 mice, *Neuron.* **24**, 879–892.

David, G., Abbas, N., Stevanin, G., Durr, A., Yvert, G., Cancel, G., Weber, C., Imbert, G., Saudou, F., Antoniou, E. et al. (1997) Cloning of the *SCA7* gene reveals a highly unstable CAG repeat expansion, *Nat. Genet.* **17**, 65–70.

Davies, S.W., Turmaine, M., Cozens, B.A., DiFiglia, M., Sharp, A.H., Ross, C.A., Scherzinger, E., Wanker, E.E., Mangiarini, L., Bates, G.P. (1997) Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation, *Cell.* **90**, 537–548.

Day, J.W., Ricker, K., Jacobsen, J.F., Rasmussen, L.J., Dick, K.A., Kress, W., Schneider, C., Koch, M.C., Beilman, G.J., Harrison, A.R. et al. (2003) Myotonic dystrophy type 2: molecular, diagnostic and clinical spectrum, *Neurology.* **60**, 657–664.

De Baere, E., Beysen, D., Oley, C., Lorenz, B., Cocquet, J., De Sutter, P., Devriendt, K., Dixon, M., Fellous, M., Fryns, J.P. et al. (2003) FOXL2 and BPES: mutational hotspots, phenotypic variability, and revision of the genotype-phenotype correlation, *Am. J. Hum. Genet.* **72**, 478–487.

DiFiglia, M., Sapp, E., Chase, K.O., Davies, S.W., Bates, G.P., Vonsattel, J.P., Aronin, N. (1997) Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain, *Science.* **277**, 1990–1993.

Dunah, A.W., Jeong, H., Griffin, A., Kim, Y.M., Standaert, D.G., Hersch, S.M., Mouradian, M.M., Young, A.B., Tanese, N., Krainc, D. (2002) Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease, *Science.* **296**, 2238–2243.

Ellerby, L.M., Hackam, A.S., Propp, S.S., Ellerby, H.M., Rabizadeh, S., Cashman, N.R., Trifiro, M.A., Pinsky, L., Wellington, C.L., Salvesen, G.S. et al. (1999) Kennedy's disease: caspase cleavage of the androgen receptor is a crucial event in cytotoxicity, *J. Neurochem.* **72**, 185–195.

Fu, Y.H., Kuhl, D.P., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J., Holden, J.J., Fenwick, R.G., Jr., Warren, S.T., et al. (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox, *Cell.* **67**, 1047–1058.

Garden, G.A., Libby, R.T., Fu, Y.H., Kinoshita, Y., Huang, J., Possin, D.E., Smith, A.C., Martinez, R.A., Fine, G.C., Grote, S.K., et al. (2002) Polyglutamine-expanded ataxin-7 promotes noncell-autonomous Purkinje cell degeneration and displays proteolytic cleavage in ataxic transgenic mice, *J. Neurosci.* **22**, 4897–4905.

Gecz, J. (2000) The *FMR2* gene, FRAXE and non-specific X-linked mental retardation: clinical and molecular aspects, *Ann. Hum. Genet.* **64**, 95–106.

Goodman, F.R., Scambler, P.J. (2001) Human *HOX* gene mutations, *Clin. Genet.* **59**, 1–11.

Grabczyk, E., Usdin, K. (2000) The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner, *Nucleic Acids Res.* **28**, 2815–2822.

Greene, E., Handa, V., Kumari, D., Usdin, K. (2003) Transcription defects induced by repeat expansion: fragile X syndrome, FRAXE mental retardation, progressive myoclonus epilepsy type 1, and Friedreich ataxia, *Cytogenet. Genome Res.* **100**, 65–76.

Hagerman, R.J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., Grigsby, J., Gage, B., Hagerman, P.J. (2001) Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X, *Neurology.* **57**, 127–130.

Helmlinger, D., Hardy, S., Sasorith, S., Klein, F., Robert, F., Weber, C., Miguet, L., Potier, N., Van-Dorsselaer, A., Wurtz, J.M., et al. (2004) Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes, *Hum. Mol. Genet.* **13**, 1257–1265.

Hillman, M.A., Gecz, J. (2001) Fragile XE-associated familial mental retardation protein 2 (FMR2) acts as a potent transcription activator, *J. Hum. Genet.* **46**, 251–259.

Hodgson, J.G., Agopyan, N., Gutekunst, C.A., Leavitt, B.R., LePiane, F., Singaraja, R., Smith, D.J., Bissada, N., McCutcheon, K., Nasir, J., et al. (1999) A YAC mouse model for Huntington's disease with full-length mutant huntingtin, cytoplasmic toxicity, and selective striatal neurodegeneration, *Neuron*. **23**, 181–192.

Holmes, S.E., O'Hearn, E., McInnis, M.G., Gorelick-Feldman, D.A., Kleiderlein, J.J., Callahan, C.A., Ingersoll-Ashworth, R.G., Sherr, M., Sumner, A.J. (1999) Expansion of a novel CAG repeat in the 5′ region of a gene encoding a subunit of protein phosphatase 2A is associated with spinocerebellar ataxia type 12, *Am. J. Hum. Genet*. **65**, A4.

Holmes, S.E., O'Hearn, E., Rosenblatt, A., Callahan, C., Hwang, H.S., Ingersoll-Ashworth, R.G., Fleisher, A., Stevanin, G., Brice, A., Potter, N.T., et al. (2001) A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2, *Nat. Genet*. **29**, 377–378.

Huntington's Disease, Collaborative, Research and Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes, *Cell* **72**, 971–983.

Huynh, D.P., Figueroa, K., Hoang, N., Pulst, S.M. (2000) Nuclear localization or inclusion body formation of ataxin-2 are not necessary for SCA2 pathogenesis in mouse or human, *Nat. Genet*. **26**, 44–50.

Ishikawa, K., Fujigasaki, H., Saegusa, H., Ohwada, K., Fujita, T., Iwamoto, H., Komatsuzaki, Y., Toru, S., Toriyama, H., Watanabe, M., et al. (1999) Abundant expression and cytoplasmic aggregations of [alpha]1A voltage-dependent calcium channel protein associated with neurodegeneration in spinocerebellar ataxia type 6, *Hum. Mol. Genet*. **8**, 1185–1193.

Jiang, H., Mankodi, A., Swanson, M.S., Moxley, R.T., Thornton, C.A. (2004) Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons, *Hum. Mol. Genet*. **13**, 3079–3088.

Katsuno, M., Adachi, H., Kume, A., Li, M., Nakagomi, Y., Niwa, H., Sang, C., Kobayashi, Y., Doyu, M., Sobue, G. (2002) Testosterone reduction prevents phenotypic expression in a transgenic mouse model of spinal and bulbar muscular atrophy, *Neuron*. **35**, 843–854.

Kitamura, K., Yanazawa, M., Sugiyama, N., Miura, H., Iizuka-Kogo, A., Kusaka, M., Omichi, K., Suzuki, R., Kato-Fukui, Y., Kamiirisa, K., et al. (2002) Mutation of ARX causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans, *Nat. Genet*. **32**, 359–369.

Klement, I.A., Skinner, P.J., Kaytor, M.D., Yi, H., Hersch, S.M., Clark, H.B., Zoghbi, H.Y., Orr, H.T. (1998) Ataxin-1 nuclear localization and aggregation: role in polyglutamine-induced disease in SCA1 transgenic mice, *Cell*. **95**, 41–53.

Koob, M.D., Moseley, M.L., Schut, L.J., Benzow, K.A., Bird, T.D., Day, J.W., Ranum, L.P. (1999) An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8), *Nat. Genet*. **21**, 379–384.

Kremer, E.J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S.T., Schlessinger, D., Sutherland, G.R., Richards, R.I. (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n, *Science*. **252**, 1711–1714.

La Spada, A.R., Wilson, E.M., Lubahn, D.B., Harding, A.E., Fischbeck, K.H. (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy, *Nature*. **352**, 77–79.

La Spada, A.R., Roling, D.B., Harding, A.E., Warner, C.L., Spiegel, R., Hausmanowa Petrusewicz, I., Yee, W.C., Fischbeck, K.H. (1992) Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy, *Nat. Genet*. **2**, 301–304.

La Spada, A.R., Fu, Y., Sopher, B.L., Libby, R.T., Wang, X., Li, L.Y., Einum, D.D., Huang, J., Possin, D.E., Smith, A.C., et al. (2001) Polyglutamine-expanded ataxin-7 antagonizes CRX function and induces cone-rod dystrophy in a mouse model of SCA7, *Neuron*. **31**, 913–927.

Lafreniere, R.G., Rochefort, D.L., Chretien, N., Rommens, J.M., Cochius, J.I., Kalviainen, R., Nousiainen, U., Patry, G., Farrell, K., Soderfeldt, B. et al. (1997) Unstable insertion in the 5′ flanking region of the *cystatin B* gene is the most common mutation in progressive myoclonus epilepsy type 1, EPM1, *Nat. Genet*. **15**, 298–302.

Lalioti, M.D., Antonarakis, S.E., Scott, H.S. (2003) The epilepsy, the protease inhibitor

and the dodecamer: progressive myoclonus epilepsy, cystatin b and a 12-mer repeat expansion, *Cytogenet. Genome Res.* **100**, 213–223.

Lalioti, M.D., Scott, H.S., Antonarakis, S.E. (1999) Altered spacing of promoter elements due to the dodecamer repeat expansion contributes to reduced expression of the *cystatin B* gene in EPM1, *Hum. Mol. Genet.* **8**, 1791–1798.

Lin, X., Antalffy, B., Kang, D., Orr, H.T., Zoghbi, H.Y. (2000) Polyglutamine expansion down-regulates specific neuronal genes before pathologic changes in SCA1, *Nat. Neurosci.* **3**, 157–163.

Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W., Ranum, L.P. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9, *Science.* **293**, 864–867.

Loesch, D.Z., Huggins, R.M., Bui, Q.M., Epstein, J.L., Taylor, A.K., Hagerman, R.J. (2002) Effect of the deficits of fragile X mental retardation protein on cognitive status of fragile x males and females assessed by robust pedigree analysis, *J. Dev. Behav. Pediatr.* **23**, 416–423.

Lunkes, A., Lindenberg, K.S., Ben-Haiem, L., Weber, C., Devys, D., Landwehrmeyer, G.B., Mandel, J.L., Trottier, Y. (2002) Proteases acting on mutant huntingtin generate cleaved products that differentially build up cytoplasmic and nuclear inclusions, *Mol. Cell.* **10**, 259–269.

Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., Lawton, M., Trottier, Y., Lehrach, H., Davies, S.W., et al. (1996) Exon 1 of the *HD* gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice, *Cell.* **87**, 493–506.

Mankodi, A., Logigian, E., Callahan, L., McClain, C., White, R., Henderson, D., Krym, M., Thornton, C.A. (2000) Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat, *Science.* **289**, 1769–1773.

Margolis, R.L., O'Hearn, E., Rosenblatt, A., Willour, V., Holmes, S.E., Franz, M.L., Callahan, C., Hwang, H.S., Troncoso, J.C., Ross, C.A. (2001) A disorder similar to Huntington's disease is associated with a novel CAG repeat expansion, *Ann. Neurol.* **50**, 373–380.

Matsuura, T., Fang, P., Lin, X., Khajavi, M., Tsuji, K., Rasmussen, A., Grewal, R.P., Achari, M., Alonso, M.E., Pulst, S.M., et al. (2004) Somatic and germline instability of the ATTCT repeat in spinocerebellar ataxia type 10, *Am. J. Hum. Genet.* **74**, 1216–1224.

Matsuura, T., Yamagata, T., Burgess, D.L., Rasmussen, A., Grewal, R.P., Watase, K., Khajavi, M., McCall, A.E., Davis, C.F., Zu, L., et al. (2000) Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10, *Nat. Genet.* **26**, 191–194.

McCampbell, A., Taylor, J.P., Taye, A.A., Robitschek, J., Li, M., Walcott, J., Merry, D., Chai, Y., Paulson, H., Sobue, G., et al. (2000) CREB-binding protein sequestration by expanded polyglutamine, *Hum. Mol. Genet.* **9**, 2197–2202.

Monckton, D.G., Cayuela, M.L., Gould, F.K., Brock, G.J., Silva, R., Ashizawa, T. (1999) Very large (CAG)(n) DNA repeat expansions in the sperm of two spinocerebellar ataxia type 7 males, *Hum. Mol. Genet.* **8**, 2473–2478.

Mortlock, D.P., Innis, J.W. (1997) Mutation of HOXA13 in hand-foot-genital syndrome, *Nat. Genet.* **15**, 179–180.

Mosemiller, A.K., Dalton, J.C., Day, J.W., Ranum, L.P. (2003) Molecular genetics of spinocerebellar ataxia type 8 (SCA8), *Cytogenet. Genome Res.* **100**, 175–183.

Mundlos, S., Otto, F., Mundlos, C., Mulliken, J.B., Aylsworth, A.S., Albright, S., Lindhout, D., Cole, W.G., Henn, W., Knoll, J.H., et al. (1997) Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia, *Cell.* **89**, 773–779.

Muragaki, Y., Mundlos, S., Upton, J., Olsen, B.R. (1996) Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13, *Science.* **272**, 548–551.

Mutsuddi, M., Marshall, C.M., Benzow, K.A., Koob, M.D., Rebay, I. (2004) The spinocerebellar ataxia 8 noncoding RNA causes neurodegeneration and associates with staufen in Drosophila, *Curr. Biol.* **14**, 302–308.

Nagafuchi, S., Yanagisawa, H., Sato, K., Shirayama, T., Ohsaki, E., Bundo, M., Takeda, T., Tadokoro, K., Kondo, I., Murayama, N., et al. (1994) Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p, *Nat. Genet.* **6**, 14–18.

Nakamura, K., Jeong, S.Y., Uchihara, T., Anno, M., Nagashima, K., Nagashima, T., Ikeda, S., Tsuji, S., Kanazawa, I. (2001) SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine

in TATA-binding protein, *Hum. Mol. Genet.* **10**, 1441–1448.

Okazawa, H., Rich, T., Chang, A., Lin, X., Waragai, M., Kajikawa, M., Enokido, Y., Komuro, A., Kato, S., Shibata, M., et al. (2002) Interaction between mutant ataxin-1 and PQBP-1 affects transcription and cell death, *Neuron.* **34**, 701–713.

Oostra, B.A., Willemsen, R. (2003) A fragile balance: FMR1 expression levels, *Hum. Mol. Genet.* **12**, **Spec No 2**, R249–R257.

Orr, H.T., Chung, M.Y., Banfi, S., Kwiatkowski, Jr., T.J., Servadio, A., Beaudet, A.L., McCall, A.E., Duvick, L.A., Ranum, L.P., Zoghbi, H.Y. (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1, *Nat. Genet.* **4**, 221–226.

Paulson, H.L., Perez, M.K., Trottier, Y., Trojanowski, J.Q., Subramony, S.H., Das, S.S., Vig, P., Mandel, J.L., Fischbeck, K.H., Pittman, R.N. (1997) Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3, *Neuron.* **19**, 333–344.

Penrose, L.S. (1948) The problem of anticipation in pedigrees of dystrophica myotonica, *Ann. Eugen.* **14**, 125–132.

Perutz, M.F., Johnson, T., Suzuki, M., Finch, J.T. (1994) Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases, *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5355–5358.

Poirier, M.A., Li, H., Macosko, J., Cai, S., Amzel, M., Ross, C.A. (2002) Huntingtin spheroids and protofibrils as precursors in polyglutamine fibrilization, *J. Biol. Chem.* **277**, 41032–41037.

Pulst, S.M., Nechiporuk, A., Nechiporuk, T., Gispert, S., Chen, X.N., Lopes-Cendes, I., Pearlman, S., Starkman, S., Orozco-Diaz, G., Lunkes, A., et al. (1996) Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2, *Nat. Genet.* **14**, 269–276.

Rasmussen, A., Matsuura, T., Ruano, L., Yescas, P., Ochoa, A., Ashizawa, T., Alonso, E. (2001) Clinical and genetic analysis of four Mexican families with spinocerebellar ataxia type 10, *Ann. Neurol.* **50**, 234–239.

Sanchez, I., Mahlke, C., Yuan, J. (2003) Pivotal role of oligomerization in expanded polyglutamine neurodegenerative disorders, *Nature.* **421**, 373–379.

Sanpei, K., Takano, H., Igarashi, S., Sato, T., Oyake, M., Sasaki, H., Wakisaka, A., Tashiro, K., Ishida, Y., Ikeuchi, T., et al. (1996) Identification of the *spinocerebellar ataxia type 2* gene using a direct identification of repeat expansion and cloning technique, DIRECT, *Nat. Genet.* **14**, 277–284.

Satterfield, T.F., Jackson, S.M., Pallanck, L.J. (2002) A Drosophila homolog of the polyglutamine disease gene *SCA2* is a dosage-sensitive regulator of actin filament formation, *Genetics.* **162**, 1687–1702.

Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G.P., Davies, S.W., Lehrach, H., Wanker, E.E. (1997) Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo, *Cell.* **90**, 549–558.

Schilling, G., Wood, J.D., Duan, K., Slunt, H.H., Gonzales, V., Yamada, M., Cooper, J.K., Margolis, R.L., Jenkins, N.A., Copeland, N.G., et al. (1999) Nuclear accumulation of truncated atrophin-1 fragments in a transgenic mouse model of DRPLA, *Neuron.* **24**, 275–286.

Seznec, H., Simon, D., Monassier, L., Criqui-Filipe, P., Gansmuller, A., Rustin, P., Koenig, M., Puccio, H. (2004) Idebenone delays the onset of cardiac functional alteration without correction of Fe-S enzymes deficit in a mouse model for Friedreich ataxia, *Hum. Mol. Genet.* **13**, 1017–1024.

Sopher, B.L., Thomas, P.S., Jr., LaFevre-Bernt, M.A., Holm, I.E., Wilke, S.A., Ware, C.B., Jin, L.W., Libby, R.T., Ellerby, L.M., La Spada, A.R. (2004) Androgen receptor YAC transgenic mice recapitulate SBMA motor neuronopathy and implicate VEGF164 in the motor neuron degeneration, *Neuron.* **41**, 687–699.

Stevanin, G., Camuzat, A., Holmes, S.E., Julien, C., Sahloul, R., Dode, C., Hahn-Barma, V., Ross, C.A., Margolis, R.L., Durr, A., et al. (2002) CAG/CTG repeat expansions at the Huntington's disease-like 2 locus are rare in Huntington's disease patients, *Neurology.* **58**, 965–967.

Stromme, P., Mangelsdorf, M.E., Shaw, M.A., Lower, K.M., Lewis, S.M., Bruyere, H., Lutcherath, V., Gedeon, A.K., Wallace, R.H., Scheffer, I.E., et al. (2002) Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy, *Nat. Genet.* **30**, 441–445.

Sutherland, G.R., Baker, E. (1992) Characterisation of a new rare fragile site easily confused

with the fragile X, *Hum. Mol. Genet.* **1**, 111–113.

Uda, M., Ottolenghi, C., Crisponi, L., Garcia, J.E., Deiana, M., Kimber, W., Forabosco, A., Cao, A., Schlessinger, D., Pilia, G. (2004) Foxl2 disruption causes mouse ovarian failure by pervasive blockage of follicle development, *Hum. Mol. Genet.* **13**, 1171–1181.

Ueno, S., Kondoh, K., Kotani, Y., Komure, O., Kuno, S., Kawai, J., Hazama, F., Sano, A. (1995) Somatic mosaicism of CAG repeat in dentatorubral-pallidoluysian atrophy (DRPLA), *Hum. Mol. Genet.* **4**, 663–666.

Wallis, J., Williamson, R., Chamberlain, S. (1990) Identification of a hypervariable microsatellite polymorphism within D9S15 tightly linked to Friedreich's ataxia, *Hum. Genet.* **85**, 98–100.

Warren, S.T. (1997) Polyalanine expansion in synpolydactyly might result from unequal crossing-over of HOXD13, *Science.* **275**, 408–409.

Watase, K., Weeber, E.J., Xu, B., Antalffy, B., Yuva-Paylor, L., Hashimoto, K., Kano, M., Atkinson, R., Sun, Y., Armstrong, D.L., et al. (2002) A long CAG repeat in the mouse Sca1 locus replicates SCA1 features and reveals the impact of protein solubility on selective neurodegeneration, *Neuron.* **34**, 905–919.

Wellington, C.L., Ellerby, L.M., Gutekunst, C.A., Rogers, D., Warby, S., Graham, R.K., Loubser, O., van Raamsdonk, J., Singaraja, R., Yang, Y.Z., et al. (2002) Caspase cleavage of mutant huntingtin precedes neurodegeneration in Huntington's disease, *J. Neurosci.* **22**, 7862–7872.

Wellington, C.L., Ellerby, L.M., Hackam, A.S., Margolis, R.L., Trifiro, M.A., Singaraja, R., McCutcheon, K., Salvesen, G.S., Propp, S.S., Bromm, M., et al. (1998) Caspase cleavage of gene products associated with triplet expansion disorders generates truncated fragments containing the polyglutamine tract, *J. Biol. Chem.* **273**, 9158–9167.

Willemsen, R., Hoogeveen-Westerveld, M., Reis, S., Holstege, J., Severijnen, L.A., Nieuwenhuizen, I.M., Schrier, M., van Unen, L., Tassone, F., Hoogeveen, A.T., et al. (2003) The FMR1 CGG repeat mouse displays ubiquitin-positive intranuclear neuronal inclusions; implications for the cerebellar tremor/ataxia syndrome, *Hum. Mol. Genet.* **12**, 949–959.

Wirtschafter, J.D., Ferrington, D.A., McLoon, L.K. (2004) Continuous remodeling of adult extraocular muscles as an explanation for selective craniofacial vulnerability in oculopharyngeal muscular dystrophy, *J. Neuroophthalmol.* **24**, 62–67.

Yoo, S.Y., Pennesi, M.E., Weeber, E.J., Xu, B., Atkinson, R., Chen, S., Armstrong, D.L., Wu, S.M., Sweatt, J.D., Zoghbi, H.Y. (2003) SCA7 knockin mice model human SCA7 and reveal gradual accumulation of mutant ataxin-7 in neurons and abnormalities in short-term plasticity, *Neuron.* **37**, 383–401.

Yvert, G., Lindenberg, K.S., Picaud, S., Landwehrmeyer, G.B., Sahel, J.A., Mandel, J.L. (2000) Expanded polyglutamines induce neurodegeneration and trans-neuronal alterations in cerebellum and retina of SCA7 transgenic mice, *Hum. Mol. Genet.* **9**, 2491–2506.

Zakany, J., Duboule, D. (1996) Synpolydactyly in mice with a targeted deficiency in the HoxD complex, *Nature.* **384**, 69–71.

Zhang, S., Xu, L., Lee, J., Xu, T. (2002) Drosophila atrophin homolog functions as a transcriptional corepressor in multiple developmental processes, *Cell* **108**, 45–46.

Zhuchenko, O., Bailey, J., Bonnen, P., Ashizawa, T., Stockton, D.W., Amos, C., Dobyns, W.B., Subramony, S.H., Zoghbi, H.Y., Lee, C.C. (1997) Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel, *Nat. Genet.* **15**, 62–69.

Zuccato, C., Tartari, M., Crotti, A., Goffredo, D., Valenza, M., Conti, L., Cataudella, T., Leavitt, B.R., Hayden, M.R., Timmusk, T., et al. (2003) Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes, *Nat. Genet.* **35**, 76–83.

Tumor Immunology: *see* Immunology of Cancer

Tumorigenesis, Epigenetic Mechanisms: *see* Epigenetic Mechanisms in Tumorigenesis

Two-photon Microscopy and Imaging

Patrick Theer, Bernd Kuhn, Dorine Keusters, and Winfried Denk
Max Planck Institute for Medical Research, Heidelberg, Germany

1 Introduction 63
1.1 Optical Sectioning 64
1.2 Increased Penetration Depth 66
1.3 Reduced Photodamage and Photobleaching 67

2 The Multiphoton Microscope 67
2.1 Light Sources 68
2.2 Excitation Pathway 69
2.3 Final Focusing and Resolution 69
2.4 Detection 70

3 Limitations 71
3.1 Temporal Resolution 71
3.2 Spatial Resolution 72
3.3 Penetration Depth 73
3.4 Photodamage and Photobleaching 74

4 Fluorophores 74
4.1 Comparison of 1PA and 2PA 74
4.2 Intrinsic Chromophores 75
4.3 Synthetic Dyes 76
4.3.1 Ion Indicators 76
4.3.2 Voltage-sensitive Dyes 76
4.4 Genetically Encoded Fluorophores and Indicators 77

5 Applications 77
5.1 Neurobiology 77
5.2 Calcium Imaging 79
5.3 Imaging of Metabolic Activity 80

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

5.4 Photoactivation (Uncaging) 80
5.5 Human Diseases and Therapy 81

6 Future Directions 82

Bibliography 83
Books and Reviews 83
Primary Literature 83

Keywords

Confocal Microscopy (CM)
Microscopy technique where optical sectioning is achieved by using a detector pinhole that is confocal to the excitation focus to reject the out-of-focus signal.

Fluorescence
Light-emitting process of special molecules, called *fluorophores*, after excitation with photons of higher energy.

Laser-scanning Microscopy
Microscopy technique, in which the light source is focused to a diffraction-limited spot and the image is formed by scanning this spot across the sample while sequentially recording the signal generated at each position.

Multiphoton Absorption (MPA)
Process where two or more photons are absorbed simultaneously to excite a transition with an energy equal to the sum of the energy of the photons.

Multiphoton Microscopy (MPM)
Microscopy technique where the contrast mechanism is based on a multiphoton absorption process. Usually, but not always, the detected signal is fluorescence.

Two-photon Microscopy (2PM)
Most widely used variant of multiphoton microscopy.

Whole-field Detection
Detection scheme for laser-scanning microscopy where the light from the whole field of view is detected.

Wide-field Imaging
Imaging technique where the whole field of view is illuminated simultaneously and imaged onto a spatially resolving detector, for example, a charge-coupled device (CCD) camera.

Multiphoton microscopy (MPM) is an imaging technique that employs signals, such as fluorescence, generated by a multiphoton absorption (MPA) process as its contrast mechanism. Compared to other fluorescence microscopy techniques, MPM often has a higher sensitivity, superior penetration into scattering tissue, and causes less photodamage to the sample and less photobleaching of the fluorophore. These advantages make MPM the method of choice for optophysiological experiments and also make MPM the only technique available for high-resolution *in vivo* imaging deep within scattering tissue, such as the brain. The two-photon microscope (2PM) is the kind of MPM used in the overwhelming majority of applications since it has virtually all of the specific advantages of MPM. Recent advances in staining techniques including the booming development of genetically encoded fluorophores keep increasing the field of actual and possible MPM applications.

1 Introduction

Over the past 15 years, multiphoton microscopy (MPM) has developed into an indispensable imaging tool in the life sciences. MPM has higher sensitivity, superior penetration into scattering tissue, and often causes less photodamage in samples than alternative imaging techniques such as confocal microscopy (CM). This means that MPM – combined, in particular, with one of many new genetically encoded and functional dyes – has become the method of choice for *in vivo* imaging in strongly scattering tissues such as skin, muscle, and tumors, and even deep within intact organs such as kidney, pancreas, lymph nodes, and brain.

This article provides an overview of functional principles, instrumentation, and some of the applications of MPM. Further details can be found in the books and reviews and primary literature listed at the end of this article.

In laser-scanning microscopy, a three-dimensional image is generated by scanning the position of a tight laser focus around in the sample while recording, as a function of the focus position, a signal – such as fluorescence – that is generated by absorption of light from the beam.

Molecules that can easily absorb light generally have a structure where several atoms of the molecule are connected by conjugated single–double bonds. As a consequence, the most weakly bound electrons are delocalized and thus interact easily with the electromagnetic field. Near-ultraviolet or visible light is usually sufficient to promote the outermost electron to an excited level (Fig. 1). In general, the molecule is excited electronically and vibrationally, with the latter excitation decaying very fast ($\sim 10^{-12}$ s). From the state reached after vibrational relaxation, which is still electronically excited, the molecule then relaxes within nanoseconds (10^{-9} s) to a state that is only vibrationally excited. The energy difference can be carried away by fluorescence emission of a photon.

The difference between one photon (1P)-excited and multiphoton (MP)-excited fluorescence microscopy is that molecular absorption is of a single photon in the first case and a photon multiplet in the second (Fig. 1). In a 1P process, the photon energy equals the energy difference between the ground state and an excited state. Light

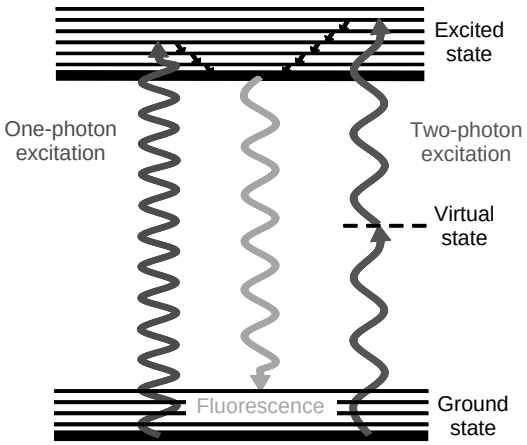

Fig. 1 Principle of one-photon and two-photon excited fluorescence. In 1PE fluorescence, a molecule is brought to the excited state by absorbing a single (blue) photon. Vibrational relaxation brings the molecule to the lowest excited state, from which it returns to the ground state by emitting a longer-wavelength (green) photon. In 2PE fluorescence, the molecule is brought to the excited state by simultaneously absorbing two NIR (red) photons. The molecule again vibrationally relaxes, and emits a fluorescence photon from the same level as in 1PE fluorescence (see color plate p. xxv).

with a quantum energy less than the gap between the ground state and the lowest lying electronically excited state cannot excite the molecule by a 1P process, but if the light intensity is sufficiently high, multiple photons of lower energy can cooperate and combine their energies to promote the molecule into the excited state. For example, two 800 nm photons can excite a transition that normally requires 400 nm light. One may picture this process as the first photon exciting the molecule to a virtual state from where the second photon then brings the molecule into its (real) excited state. For this process to occur, the two photons have to arrive at the fluorophore within the lifetime of the virtual state, which is less than 1 fs ($1\ \text{fs} = 10^{-15}$ s) as defined by the Heisenberg uncertainty principle ($\tau \leq \hbar/E$). The likelihood that two photons arrive at a fluorophore almost simultaneously is proportional to the intensity squared and is all but negligible for light sources other than a focused laser. Therefore, two-photon absorption (2PA) was observed only after the invention of the laser. It is the strongly superlinear dependence on the light intensity that gives 2PM (and MPM in general) its main advantages.

1.1 Optical Sectioning

In 1P microscopy, the signal generated is proportional to the light intensity (the power divided by the beam area: $I = P/A$). As a result, the signal generated in each slice of a sample depends on the slice thickness, but not the distance from the focus, since variations in the beam area are, as long as the power stays constant, exactly balanced by inverse variations in the excitation rate per molecule. When imaging thick samples, the light from the focus is thus always contaminated, in fact, dominated by fluorescence generated above and below the focus. This is the bane of wide-field fluorescence imaging of 3-dimensional (3D) samples. In CM, this problem is solved by placing a pinhole in front of the detector. The pinhole is confocal with the excitation focus and rejects light that did not originate from that focus (Fig. 2). This property of the confocal microscope, called *optical sectioning*, vastly improves the ability to obtain 3-D fluorescence images, particularly of thick biological samples.

The first nonlinear optical microscope used the second-harmonic signal and

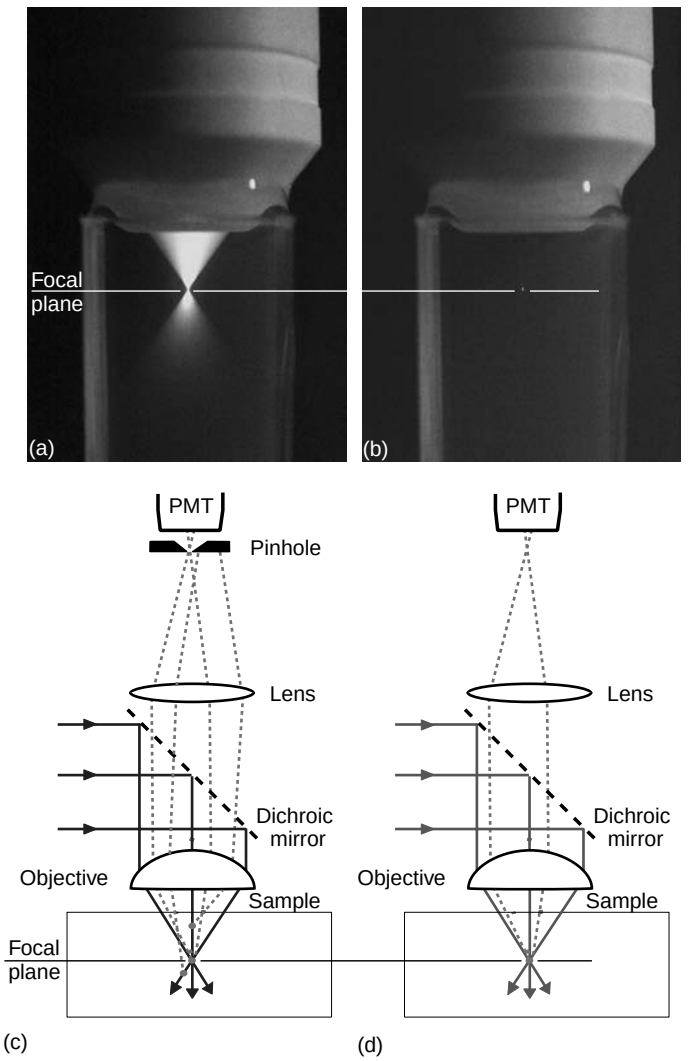

Fig. 2 Optical sectioning in one-photon confocal and in two-photon microscopy. (a) With 1PE, fluorescence is generated throughout the illumination cone. For 2PE (b), because of the nonlinear intensity dependence, fluorescence generation is limited to a small volume around the focus. In CM, optical sectioning is achieved by placing a confocal pinhole in front of the detector. This pinhole rejects (out-of-focus) fluorescence that is generated above and below the focus (c), fluorescence generated by scattered excitation light (e), and in-focus fluorescence that is scattered on its way to the detector (e). Because no out-of-focus fluorescence is generated in a two-photon microscope, no pinhole is necessary (d) and even fluorescence that is scattered on its way to the surface can be detected (f).

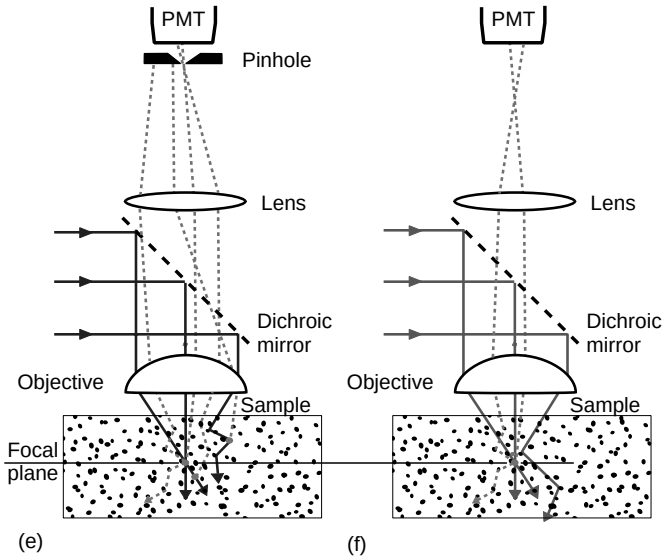

Fig. 2 (*Continued*)

wide-field imaging. Scanning the focus, which was introduced later, strongly enhances the focal but not the out-of-focus signal, and thus provides optical sectioning. The hopes placed on harmonic imaging for biological specimens were, however, not fulfilled, at least initially.

Two photon–excited fluorescence microscopy proved to be much more immediately applicable to biological problems, due, first, to the central importance of fluorescent labeling in biological microscopy (see Sect. 4) and second, to the differences in the physical characters of fluorescence and harmonic generation. The main difference being that harmonic generation is a coherent process, where field strengths rather than intensities add, and the generated signal depends on the square of the chromophore concentration, and is liable to interferences effects; in fluorescence, which is incoherent, light from different fluorophores does not interfere with each other and the signal increases linearly with the fluorophore concentration.

Common to scanned-focus nonlinear microscopies is that they provide excitation-based optical sectioning. This is because, different from the linear (1P) case, the total excitation in a slice intersecting the beam strongly drops with distance to the focus (Fig. 2). In fluorescence microscopy, excitation-based rather than confocal optical sectioning is particularly helpful since photodamage and photobleaching are enormously reduced (see below).

1.2 Increased Penetration Depth

MPM can generate high-resolution images from focal planes deep inside the tissue, much deeper than CM, because signal is generated only at the focus allowing scattered fluorescence to be collected without loss of resolution or optical sectioning. In CM, only light passing through the confocal pinhole is detected. However, due to scattering, the fraction of the signal

originating from the focus that does pass the pinhole in the CM decreases exponentially with imaging depth. Compensating for the loss in signal by increasing the excitation power in CM in most cases would lead to an unacceptable increase in photodamage (see below). A further reason for the deeper penetration of MPM is that, for the same fluorescent label it uses excitation light with a wavelength that is longer (up to two times for 2PM) and hence scattered much less. Occasionally helpful is the larger separation between the excitation and emission wavelengths in MP excited fluorescence, which allows a more spectrally complete detection of the fluorescence.

1.3 Reduced Photodamage and Photobleaching

Protracted and high-resolution CM imaging is often limited by excitation light-induced damage to the sample (photodamage) or the fluorophore (photobleaching), which in 1PM occur throughout the sample even though information is only gained from the focal slice. In MPM, damage is confined to those fluorophores (and their immediate vicinity) that do provide information. A reduced relative MP absorbance of intrinsic fluorophores may further contribute to the vastly improved tissue viability sometimes seen even in tissue that scatters only little.

2 The Multiphoton Microscope

The layout (Fig. 3) of a generic MPM is very similar to that of a laser-scanning CM, from which an MPM can often be constructed by modification. In this section, the components of an MPM will be discussed individually.

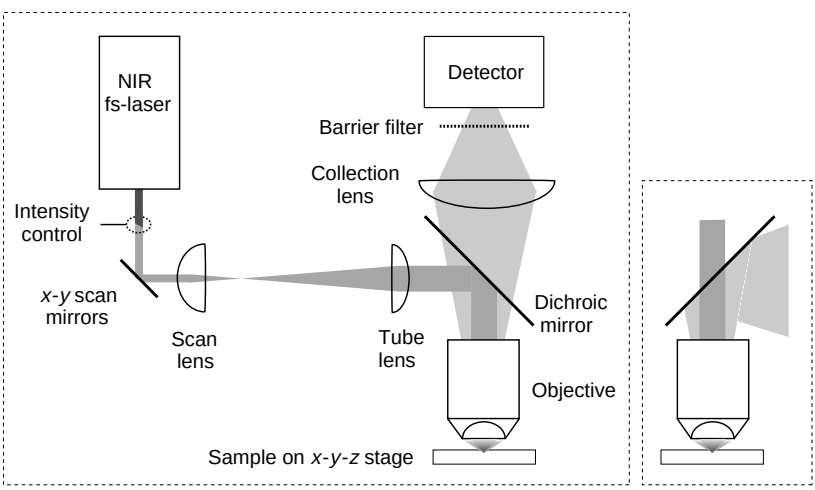

Fig. 3 Setup of a generic two-photon microscope. Light from an ultrafast laser or laser amplifier is focused to a diffraction-limited spot by an objective. The beam is raster-scanned across the sample and the fluorescence generated at each position is collected by the objective, separated from the excitation light by a dichroic mirror and additional filters either in a transmission or reflectance configuration (see inset), and detected by a whole-field detector.

2.1 Light Sources

The development of reliable ultrafast laser sources operating in the near infrared (NIR) coincided with the invention of MPM and has been an important factor in the broad adoption of MPM, since with continuous wave (cw) lasers, even 2PM is only marginally possible for brightly labeled samples. Today the vast majority of MPMs use mode-locked titanium sapphire lasers (Ti:sapphire, Ti:Al_2O_3), which can be tuned from about 680 to 1100 nm, allowing 2PE of almost all relevant labels, and 3PE into the deeper UV. The pulse length of this laser type is around ~100 fs with a repetition rate of around 100 MHz, i.e. every 10 ns a short burst of light arrives at the sample while for the remainder of this period no light is incident upon the sample. Alternative laser sources that produce similar pulse lengths are based on Cr:LiSrAlFl, Cr:Forsterite, or Nd:YLiF crystals, or on Yb-doped optical fibers, all of which have much smaller tuning ranges, but with the exception of the Cr:LiSrAlFl laser, operate at longer wavelengths than can be easily provided by the Ti:Sapphire laser. A very large tuning range but less total power is provided by optical parametric amplifiers.

An ultrafast laser increases the signal levels in the 2PM by 5 orders of magnitude as compared to a cw-laser of the same average power. This is because the average rate at which fluorescence photons are generated by 2PA is proportional to the average square of the instantaneous light intensity I:

$$\langle n_F(\mathbf{r})\rangle \propto \delta C\langle I(\mathbf{r},t)^2\rangle = \delta C\langle I(\mathbf{r},t)\rangle^2 \frac{\langle I(\mathbf{r},t)^2\rangle}{\langle I(\mathbf{r},t)\rangle^2}, \quad (1)$$

where $\langle\,\rangle$ indicates an average over time, δ is the two-photon absorption cross section (often expressed in units of Göppert-Mayer (GM); 1 GM $= 10^{-50}$ cm^{4}s/photon), and C is the fluorophore concentration. For a spot with area A, power and intensity are related by:

$$I(\mathbf{r},t) = \frac{P(\mathbf{r})}{A}. \quad (2)$$

For a square pulse with pulse length τ, the "two-photon advantage" is

$$\frac{\langle I(\mathbf{r},t)^2\rangle}{\langle I(\mathbf{r},t)\rangle^2} = \frac{\frac{1}{T}\int_0^T I(\mathbf{r},t)^2 \mathrm{d}t}{\left(\frac{1}{T}\int_0^T I(\mathbf{r},t)\mathrm{d}t\right)^2} = \frac{1}{f\tau}. \quad (3)$$

Equation 1 then becomes:

$$\langle n_F(\mathbf{r})\rangle \propto \delta C\frac{P(\mathbf{r})^2}{A^2}\frac{1}{f\tau}. \quad (4)$$

Thus, for example, to obtain the same signal with 1 ps pulses rather than with 100 fs pulses, the average power must be higher by a factor of $\sqrt{10}$.

Implicit in this derivation is the assumption that the 2P absorption cross section δ is constant across the spectrum of the laser pulse. For very short pulses this is not true because – due to the uncertainty principle – the shorter the pulse duration ($\Delta\tau$), the broader its spectral width ($\Delta\nu$). For a transform-limited (unchirped) Gaussian-shaped pulse $\Delta\tau\Delta\nu = 0.44$. A 100-fs laser pulse centered around 800 nm then ranges (full width at half maximum; FWHM) over 10 nm in wavelength and a 20-fs pulse over 50 nm, which is already as wide as the main features in a typical fluorophore absorption spectrum. However, in practice the main limitation for using shorter pulses is group delay dispersion (GDD), which occurs because

different wavelength components of the laser pulse travel at different speeds inside all materials, including of course, optical glass. After passing the microscope optics, the long-wavelength ("red") components of the pulse arrive at the focus ahead of the short-wavelength ("blue") components, leading to a so-called *chirped* pulse. Because the wider spectrum of a shorter pulse accentuates this effect, an initially shorter pulse can actually become longer once it reaches the focus than an initially longer pulse. Owing to the GDD in a typical microscope, a 20-fs pulse, for example, is stretched to about 700 fs, while a 100-fs pulse is stretched to only 200 fs. Without compensation (prechirping, see below), the shortest pulses that can be attained at the sample are about 150 fs. Therefore, lasers producing 100 to 200 fs pulses are typically used in 2PM. It is possible (but rarely necessary) to compensate for the microscope's GDD by the use of an optical arrangement (consisting of prisms, gratings, or special dielectric coatings) that has negative GDD. Negative GDD prechirps the pulse, giving the blue components a head start over the red components. This head start compensates all or part of the speed disadvantage that the blue components have when traveling through the microscope optics. Prechirping allows the arrival of transform-limited pulses as short as 15 fs at the focus. For such short pulses, compensation of higher-order dispersion and chromatic aberrations needs to be considered as well. For higher than 2PA, the effects of using shorter pulses are even more dramatic (the 3PA rate, for example, increases by 10 orders of magnitude compared with cw-excitation when using the typical Ti:sapphire laser).

2.2 Excitation Pathway

The pathway for the excitation light in a MPM is essentially that of a laser-scanning CM: shutter, beam conditioning and intensity control, scan mirrors, scan lens, tube lens, objective, and, finally, the sample (see Fig. 3). As in every laser-scanning microscope, the scan mirrors (or a point in between for proximity coupled scanners) is conjugate to (i.e. imaged into) the objective exit pupil. Coatings on lenses and mirrors need, of course, to be IR transmissive or reflective, respectively. The intensity control needs to be compatible with ultrashort pulses and should allow the automatic compensation of depth dependent losses (see Sect. 3.3) and should be fast enough to shut off the beam during retrace, i.e. while the focus is moved from the end of a scan line to the beginning of the next.

2.3 Final Focusing and Resolution

Many standard microscope objectives can be used for MPM, albeit with compromises in correction and transmissivity. While this was an issue in the early times of MPM, most objectives now are either corrected and transparent throughout the IR or come as special "IR" versions. To minimize tissue damage (see below), detection efficiency has to be maximized by keeping the transmission at the fluorescence wavelength(s) as high as possible. Optical correction at the fluorescence wavelength is not necessary. The numerical aperture (NA) of the objective determines the resolution (laterally $\Delta\rho \propto 1/NA$, axially $\Delta z \propto 1/(NA)^2$) and the peak intensity ($\propto (NA)^2$), and thus the peak excitation rate per fluorophore ($\propto (NA)^4$), but (and

this is a special property of 2PA, that neither 1P nor higher-order multiphoton absorption, MPA possess) when the concentration of fluorophores is uniform, the *total* amount of fluorescence generated throughout the focal volume $V = \Delta\rho^2\Delta z$ is independent of the NA:

$$\langle n_{\mathrm{F}}\rangle_{\mathrm{tot}} \propto \delta CP^2 \frac{NA^4}{f\tau}\Delta\rho^2\Delta z \propto \delta CP^2 \frac{1}{f\tau}. \tag{5}$$

A more extensive treatment that takes into account the exact spatial and temporal intensity distribution of the light in the sample yields:

$$\langle n_{\mathrm{F}}\rangle_{\mathrm{tot}} = \frac{4\eta g n\lambda}{\pi h^2 c^2}\delta CP^2 \frac{1}{f\tau}, \tag{6}$$

where η is the quantum-efficiency of the fluorophore (i.e. the percentage of excited molecules that emit their excess energy in the form of a photon), $g = f\tau\langle I(\mathbf{r},t)^2\rangle/\langle I(\mathbf{r},t)\rangle^2$ and depends on the exact temporal intensity distribution of the pulse. For the square pulse used in equation (3), $g = 1$. For more realistic shapes such as a Gaussian or a hyperbolic secant intensity distribution, $g = 0.66$ and 0.59, respectively; n, λ, h, and c are the index of refraction, wavelength of the excitation light, Planck's constant, and the vacuum speed of light, respectively.

While scattering *per se* need not reduce the resolution even if only a minor part of the beam energy reaches the focus unscattered (ballistically), eventually the stronger attenuation of the lateral rays due to their longer path length inside the tissue does lead to a reduction in the effective NA, and thus a larger focus size. Further, practical, considerations in selecting an objective are the detection efficiency (see below), working distance, which tends to decrease with increasing NA, and the immersion medium, which will usually be water or a physiological salt solution. Refractive-index mismatch between immersion medium and sample that was not taken into account during the lens design results in a loss of resolution and signal strength.

2.4 Detection

To preserve the main strength of MPM, i.e. the reduction of photodamage, the detection efficiency must not be compromised. This is because molecular excitation inevitably causes photodamage to the sample (see Sect. 3.4) and thus the number of molecular excitations necessary to generate a sufficient number of detected photons need to be minimized. In nonscattering samples, the main determining factor for detection efficiency is the solid angle of collection (assuming, of course, the avoidance of absorption and reflection losses as much as possible). Thus, a high NA objective is crucial if epi detection (through the objective lens, the usual configuration) is used. Better still is a combination of epi detection (where, for working distance reasons, a compromise with respect to the NA may be necessary) with transdetection through a high NA condenser. The separation of the fluorescence from the excitation light by a dichroic mirror is standard practice in fluorescence microscopy, as is the further separation into multiple spectral channels. Suppression of the excitation light often is easier in MPM because the wavelength separation is larger and the sensitivity of some detectors declines sharply toward longer wavelengths. In contrast to a laser-scanning CM, there is no need in the MPM to "de-scan" the fluorescence in order to thread it through the confocal pinhole. In fact, the detection optics in

a MPM can be rather crude but need to accept large opening angles and have a large field of view for deep imaging in scattering samples.

Deep imaging in scattering samples is where no other known fluorescence microscopy technique can compete with MPM because MPM can image deep in scattering samples virtually without loss of resolution and sensitivity as long as the detection pathway collects a large fraction of the diffuse fluorescence light emerging from the sample. Unlike in other techniques, the diffuse fluorescence contains high-resolution information about the focus because the excitation is highly localized to a volume of less than one femtoliter. The improvement possible by using whole-field detection has been directly demonstrated by switching quickly between whole-field and descanned detection. The main drawback of whole-field detection is its incompatibility with detectors that only accept a small phase space volume (the product of angular and area acceptances). This precludes the use of photon-counting avalanche photodiodes (APDs), which are the highest quantum-efficiency photon detectors available, and of spectrometer detectors. Fortunately, the characteristics of photo multiplier tubes (PMTs), which have large sensitive areas and acceptance angles and are the standard detectors for CM, have also improved, closing the quantum-efficiency gap.

3 Limitations

In this section, the limitation of spatial and temporal resolution, penetration depth and photodamage will be discussed.

3.1 Temporal Resolution

The temporal resolution of optical techniques can be very high. In fact, some of the fastest measurements are based on pulsed lasers. When fluorescence is used, the excited-state lifetime (in the nanosecond range) is the ultimate limit. In practice, temporal resolution is often limited by the time required to obtain a sufficient signal-to-noise ratio. When a whole image is to be acquired, this time has to be multiplied with the number of pixels in a frame. Another limitation of all laser-scanning microscopes is that the focus has to be raster-scanned across the field of view. In most CMs and MPMs, this is achieved by mechanically scanning the beam using galvanometer mounted mirrors, which are limited in frequency response to the low kilohertz regime, setting the rate at which scan lines can be acquired to about 1 kHz or so, depending somewhat on the extent of the scan. A high-resolution (1000 lines) image thus takes at least one second. By sacrificing one spatial dimension and scanning a single line only, a time resolution of 1 ms is common, and when measuring only from one spot, a time resolution in the microsecond range can easily be achieved. On the other hand, to acquire a high-resolution volume image (a stack of optical sections) can take many minutes. When trying to measure dynamic physiological signals from many locations simultaneously, as it is desirable, for example, in neuroscience, image acquisition time is a serious limitation. Attempts have, therefore, been made to speed up the scanning process by either using resonant scanners, or rotating polygonal mirrors, or by scanning the beam nonmechanically using an acousto-optical deflector (AOD). Owing to the spread wavelength spectrum

of the excitation pulses, the AOD, which is based on diffraction by a sound wave, is plagued by angular dispersion, which can only be sufficiently compensated for a limited field of view. The main attraction of AOD scanning is the capability of random-access scanning: the focus can be moved between distant location within microseconds without illuminating (and potentially damaging) the sample area in between.

Acquisition speed is limited ultimately, as mentioned above, by the fluorescence decay time, which has to be shorter than the pixel dwell time. But the limited rate at which a fluorophore can produce photons, even if the excitation light is arbitrarily intense, usually enforces a longer pixel dwell time just to collect enough photons. Beyond those limits, an increase is only possible by scanning multiple points, arranged either in a 2D pattern or in a line. This approach is rather successful in CM, but it requires confocal or at least spatially resolved detection, since at each point in time, fluorescence can come from any of the foci. Thus, using it in MPM means giving up one of the main advantages of MPM, namely the ability to detect fluorescence irrespective of how it reaches the detector (see above) and accepting either image degradation or signal loss. Other drawbacks are anisotropic lateral and reduced axial resolution (for line illumination) and the increased laser-power requirement (for any multipoint method).

3.2 Spatial Resolution

In the generic configuration (with whole-field detection), there is no spatial selectivity in the detection and hence the resolution of MPM (i.e. the size of the point-spread function) is completely determined by the intensity distribution (in fact, the point-spread function equals the intensity squared) at the focus. Thus, the resolution depends on the wavelength of the excitation light and on the NA of the objective. With 700-nm light and an NA of 1.3, a resolution (FWHM) of ~200 nm in the lateral and 600 nm in the axial direction can be achieved.

Even though in CM the resolution is determined by the combined (multiplied) excitation and detection point-spread functions, the mathematical expression for the resolution of CM and 2PM are identical. This holds, however, only under the somewhat unrealistic assumption that the pinhole has zero diameter and the 1PE and 2PE wavelengths and the emission wavelength are all the same. Realistically, the pinhole has to be opened up to gain sufficient signal, the emission wavelength is longer (both reducing the resolution of CM) and the excitation wavelength is close to twice as long for 2P as for1P (reducing the resolution in 2PM).

As a result, compared with CM, the resolution in 2PM (and MPM, in general) is typically somewhat worse, but can be improved by reintroducing a confocal pinhole, albeit at the expense (particularly in scattering specimens) of detection efficiency. 2PM can also be combined with the 4Pi microscope, which greatly improves the resolution in the axial direction, and where using 2PE rather than 1PE considerably reduces side lobes in the point-spread function and thus fills in the holes in the modulation transfer function.

In general, at a given wavelength, resolution improvements beyond the Abbe limit require optical nonlinearity, which is exploited in the Stimulated Emission Depletion (STED) microscope. This microscope is based on the saturation of

excited-state depletion, which has potentially unlimited resolution, because this saturation contains arbitrarily high orders of nonlinearity. 2PE, by contrast, contains only a second order (quadratic) nonlinearity, and consequently, in 2PM the resolution improvement beyond the Abbe limit is moderate.

3.3 Penetration Depth

In scattering samples where absorption of the light is negligible, the intensity of the ballistic light, i.e. light that has not been scattered, decreases with depth in the sample as:

$$I_{\text{ball}} = I_0 \exp\left(-\frac{d}{l_s}\right), \qquad (7)$$

where l_s is the scattering length, i.e. the average distance a photon travels before it is scattered. For near-infrared wavelengths, l_s in brain tissue is around 200 μm. To estimate the imaging depth, one can assume that only the ballistic light that reaches the focus contributes to the generation of two-photon fluorescence, which seems to hold for up to at least several scattering mean-free-path lengths. Then the number of fluorescence photons generated when the focus is at a depth d below the surface can be written as:

$$\langle n_F \rangle_{\text{tot}} = \frac{1}{2}\eta\delta C P^2 \frac{g}{f\tau}\frac{8n\lambda}{\pi h^2 c^2} e^{(-2d/l_s)}. \qquad (8)$$

As long as the detection system is designed to capture most of the fluorescence, which consists almost entirely of scattered light, the maximum imaging depth depends only on the laser power that is available. With the power available from standard mode-locked Ti:sapphire lasers, imaging depths of 2 to 3 scattering lengths (400–600 μm) are achieved routinely.

To increase the imaging depth further, the efficiency in generating 2P fluorescence needs to be increased, which is possible by reducing the laser pulse–repetition rate while increasing the pulse energy. The use of a so-called *regenerative amplifier*, which generates pulses that are about 250 times as intense as those directly from the laser "oscillator" but are also by the same factor less frequent, resulting in an unchanged average power, permits imaging to about 1000 μm. A somewhat technical but important issue is that each pixel needs to contain at least one pulse, limiting the pixel rate to several hundred kilohertz (the repetition rate of the amplifier). At such pulse rates, the average 2PA rate is increased by a factor of about 250.

Now the imaging depth is, however, no longer limited by the available laser power but instead by the fluorescence background generated near the surface of the sample. The reason for this is that in order to keep the fluorescence generation at the focus constant, the amount of ballistic light that reaches the focus must also be kept the same, and thus the laser power that needs to enter the sample must increase exponentially (equation 7). At large depths, this exponential increase overcomes the quadratic decrease in light intensity, which results from the increased beam cross section with distance from the focus. Eventually, the intensity at the surface becomes comparable to the intensity at the focus and surface fluorescence starts to dominate. For samples with staining throughout the sample, this limit may well be impossible to overcome, thus defining the ultimate depth limit of 2PM (a similar logic applies to higher-order MPM). This limit increases somewhat with the NA, but that may not help much since at higher NAs, focus blurring due to wavefront aberrations becomes a more serious issue, and

may, in turn, require the use of adaptive optics as a remedy.

3.4 Photodamage and Photobleaching

Cell viability is obviously important for imaging living tissue. Excitation of intrinsic or introduced chromophores often leads to photochemical effects, such as destruction of the chromophore (bleaching) or damage and subsequent cell-death, which are mostly mediated by reactive oxygen species. This problem is exacerbated in laser-scanning CM because of the rather poor utilization of the generated fluorescence light and the appearance of superlinear photochemical effects at high intensities, probably due to excited-state absorption. Utilization of the generated fluorescence is generally much better in MPM, with the improvement increasing as the specimen becomes more scattering and the imaging depth increases.

However, in clear (i.e. nonscattering) specimens, where the detection efficiency in the CM for focal photons can be rather high, bleaching and photodamage in the 2PM can actually be exacerbated, but is also clearly reduced in some cases compared with CM. Strategies for avoiding or reducing damage depend on the underlying photophysics and photochemistry. If damage is due to single-photon absorption, increasing the 2PE efficiency by reducing pulse length or repetition rate is helpful. Damage is, however, rarely dominated by 1PA for typical tissue, but can become so in pigmented cells or if much longer wavelengths are used and water absorption becomes an issue. If higher-order instantaneous nonlinearities, such as 3P absorption or optical breakdown are the culprits, increasing the pulse length can help. Chemical nonlinearities such as locally overwhelming cellular repair mechanisms should be avoidable by faster scanning. Often, as mostly anecdotal evidence suggests, damage decreases with increasing wavelength, which may be due to reduced 2PA by endogenous chromophores. While photobleaching can be quantitatively measured, this is more difficult with damage to cells, since photostress may be present but may not be evident below a certain instantaneous or cumulative threshold. This is consistent with the observation in several studies that at low intensities (up to $10^{14} - 10^{15}$ W m^{-2} which corresponds to 1–5 mW at the focus) no obvious damage occurs. As the intensity increases, severe damage, optical breakdown, and strong luminescence eventually do occur.

4 Fluorophores

4.1 Comparison of 1PA and 2PA

While the fluorescence process itself is largely independent of the mode of excitation (and hence the fluorescence emission spectra are the same for 1PE and 2PE), the absorption spectra of 1PE and 2PE are expected and found to be quite different in some cases. The reasons for this are: (1) To excite a molecule from its ground state to the excited state, quantum-mechanical selection rules have to be fulfilled in addition to energy conservation. For example, the angular momentum (usually only the photon spin is relevant since the absorption of a photon with nonzero orbital angular momentum is unlikely) that is carried by a photon must be taken up by the molecule upon absorption. The selection rules for

2PA are different from those for 1PA, since in the first case, two photons, each carrying a spin of $1\hbar$, are absorbed. This means that the angular momentum of the molecule has to be unchanged (if two photons of opposite spin are absorbed) or changed by $2\hbar$. A change by $1\hbar$ (also corresponding to a change in parity) is not allowed. However, for complex asymmetric molecules, these selection rules are not necessarily as strict, due to interaction with molecular vibrations and rotations. (2) Since the transition to the excited state occurs in a sense via all molecular states, which serve as a combined "virtual" intermediate state, excitation is possible even if there is no direct orbital overlap between ground and target state.

For the use of fluorescent dyes in MPM, it is very important to know their MP-absorption spectra in order to choose the best dye and then excite it optimally. While it is not easy to measure MP-absorption spectra, at least the 2P-absorption spectra of many commonly used chromophores are now available (http://www.drbio.cornell.edu/) and it turns out that all dyes used with 1PE can also be used with 2PE, albeit with a larger spread in their cross sections because, in a way, transition matrix elements enter twice in the 2PA process.

In some cases, the 2PA spectra plotted on a $\lambda/2$ scale are rather similar to the 1PA spectra plotted on a λ scale. In most cases, however, the excitation maxima are shifted to the blue and the spectra are broader. Additional spectral features at wavelengths longer than the "red" 1P absorption edge are not seen, nor would they be expected. The reason for this is that fluorescence emission only occurs for molecules where the transition to the lowest lying excited state 1P is allowed. There have been a number of attempts to predict 2P cross sections and to develop, on the basis of theoretical calculations, chromophores with large 2P cross sections.

In the following, we will give a short overview of chromophores that are widely used with 2PM, with particular focus on dyes with biological relevance. Under this premise, the chromophores can be categorized into intrinsic, synthetic, and genetically encoded dyes, each group containing dyes with a variety of different indicator features.

4.2 Intrinsic Chromophores

Biological tissue is naturally fluorescent to a varying degree even without adding chemically synthesized molecules or introducing the gene of a fluorescent protein (see below). The molecular origin of this intrinsic "autofluorescence" depends on the excitation wavelength. For UV excitation, nucleotides and the aromatic amino acids are dominant, for excitation with visible wavelengths, other compounds such as nicotinamide adenine dinucleotide (phosphate) NAD(P)H and flavin adenine dinucleotide FAD are the main contributors.

Tryptophane fluorescence is rarely used for 2PM, but 2PE fluorescence correlation spectroscopy (FCS) is possible for proteins containing a large number of tryptophane residues. NAD(P)H is of special interest as its fluorescence intensity depends on its oxidation state. With excitation at 800 nm, NAD(P)H and flavoprotein fluorescence can be used together for a quantitative ratiometric measurement of cellular metabolism.

One of the advantages of MPA for exciting cellular autofluorescence is that it permits access to transitions that

require for 1PE, ultraviolet light of wavelengths that do not pass most microscope objectives.

On one hand, endogenous fluorophores can be useful because tissues can be imaged without dyes having to be applied to the tissue. On the other hand, endogenous fluorophores contribute most of the unwanted background when exogenous dyes, which provide enormous specificity and tailored indicator properties, are to be imaged. A special case is the green fluorescent protein (GFP, see below), which occurs naturally in the jellyfish and has a biological function there, but in other organisms GFP is an exogenous stain in the sense that it needs to be introduced by molecular genetic means.

4.3 Synthetic Dyes

Because of the specific advantage of MPM for the observation of living tissue, the most frequently used synthetic dyes are those that provide functional signals, called *fluorescent indicators*, in particular, ion-sensitive, and (more recently) voltage-sensitive probes.

4.3.1 Ion Indicators

In many cellular responses, changes in ion concentrations are involved. Most important is Ca^{2+}, with its key position in cellular signal transduction; also important are Na^+, K^+, Mg^{2+}, H^+, and Cl^-. Ion indicators go back to Arnaldus de Villanova, a doctor and alchemist (~1300 AD), who invented litmus, an indicator that changes its color (absorption spectrum) depending on the concentration of H^+ ions. The first 2PM calcium measurements where done using Indo-1, a BAPTA-based Ca^{2+} indicator, which is, however, excited by rather short wavelength and has a rather moderate 2P cross section. Much larger cross sections are found in xanthene-derived indicators, which also use the BAPTA Ca^{2+}-binding group and are now available with a wide range of Ca^{2+} affinities, covering the entire physiological range. A selection of different peak emission wavelengths (from 530 to 670 nm) ensures discrimination ability against GFP or other cellular labels if needed.

In tissue imaging, labeling without damage to cells is crucial and several methods have been used in connection with MPM. (1) Loading individual cells by diffusion or iontophoresis from dye-filled micropipettes in brain slices or *in vivo*; (2) applying cell-permeable acetoxymethoxyl (AM) esters variants, which become trapped in cells by intracellular esterases. AM loading is straightforward in isolated cells, but is also possible in brain slices and even *in vivo*; and (3) particle ("gene")-gun delivery.

For an overview of fluorescent indicators see http://www.probes.com/. All of these indicators can be efficiently 2P excited somewhere between 700 and 1000 nm, i.e. they are accessible when using a Ti:sapphire laser.

4.3.2 Voltage-sensitive Dyes

Voltage-sensitive dyes (VSDs) measure electrical activity directly rather than indirectly via Ca^{2+} influx. Most VSD-based measurements have, until recently, not used laser-scanning microscopy, since the fractional fluorescence changes that occur with physiological voltage swings are small and thus need for detection and quantification of a large number of fluorescent photons. Recently, it was shown that by using band-edge excitation, fractional changes are substantially increased, and are for equivalent excitation wavelengths, larger with 2PE than with 1PE.

4.4 Genetically Encoded Fluorophores and Indicators

Labeling with synthetic dyes is in most cases not cell-type specific. For example, when injecting AM ester-Ca^{2+} indicator into brain tissue, many cell types take up the dye, and it is impossible to stain, for example, only neurons. Genetically encoded protein "dyes," often called XFPs by generalizing from GFP, the green fluorescent protein, with "X" standing for the color such as for yellow in YFP and for cyan in CFP, overcome this problem because their expression can be put under the control of specific promoters that are only active in certain cell types. XFPs can also be fused to other proteins to then map their dynamic distribution in the cell. Fusion protein with XFPs have also been designed to act as indicators for Ca^{2+}, Cl^{-}, pH, and other cellular signals and physiological parameters such as cAMP, phosphorylation, redox potential, protein kinase B and C, as well as membrane voltage. The molecular genetic methods used to label intact animals are either the creation of transgenic animals, infection with properly engineered viral vectors, or *in vivo* electroporation.

For imaging in intact animals XFPs and MPM are rather synergistic, since together they solve the deep labeling and the deep imaging problem, with 2PM even allowing imaging through the thinned but intact skull.

5 Applications

On the basis of the unique imaging properties described earlier, MPM has become the preferred and virtually the only method that allows morphological and functional imaging in thick living tissue and in live animals. While many of the essential uses of MPM are in the field of neuroscience, the range of biological structures and processes investigated using MPM have continuously grown. Applications of MPM now include photodynamic therapy, noninvasive optical biopsy of human skin, the study of cellular metabolism, embryogenesis, immunology, tumor pathophysiology, and neurodegenerative disease. In the following, we will discuss some of these applications in more detail.

5.1 Neurobiology

After it became clear that 2PM is particularly well suited to high-resolution imaging in scattering tissue, it was applied to neural tissue, which is strongly scattering and requires the use of intact pieces of tissue for functional studies owing to the high degree of cellular interconnectivity. Early studies focused on Ca^{2+} dynamics in dendritic spines. Then it became clear that it is possible to image Ca^{2+} dynamics with high spatial and temporal resolution in the wholly intact brain as well, where 2PM can now be used to guide recording pipettes to selected cells. It has also been possible to study the structural stability and plasticity of neuronal morphology in explanted tissue and, over periods of months, in intact animals.

2PM has helped to reveal neural structure and function at various levels ranging from individual synapses to entire neural networks (see also Fig. 4). The retina, a spatially separate but functionally integral part of the brain, has been a particularly rewarding specimen for 2PM. Because of the

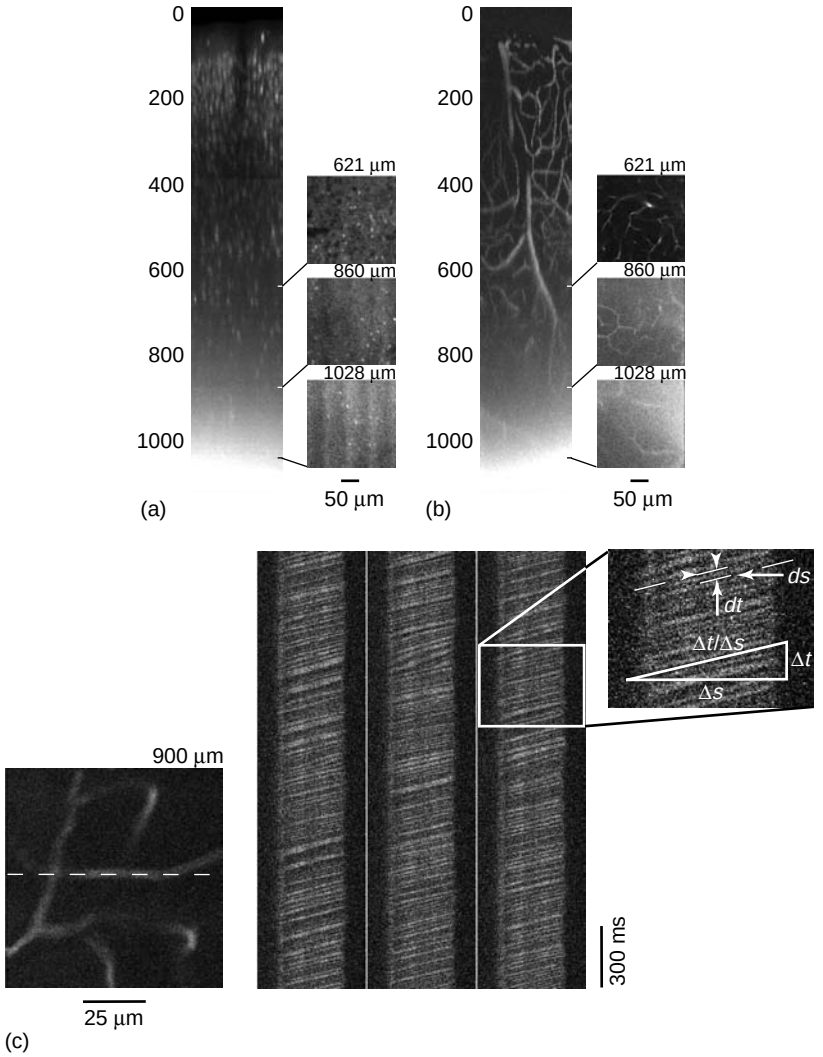

Fig. 4 (a) *x-z* projection and single planar scans of GFP labeled neurons and (b) stained vasculature obtained throughout almost the entire gray matter of the mouse neocortex. (c) Blood flow measurement in the mouse neocortex at 900 µm below the brain surface. Blood cells appear as shadows in surrounding blood plasma. Their motion traces out shaded bands in an image consisting of line scans repeatedly taken along the capillaries (dashed line in the planar scan). Blood flow parameters as velocity, linear density, average spacing, and flux of red blood cells can be inferred from the slope (dt/ds) of, and the distance (ds) and time (dt) between shaded bands.

retina's high sensitivity for UV and visible light, 1PE of common fluorescent probes inevitably perturbs the observed specimen; in fact, intensities used commonly for fluorescence microscopy completely bleach the photopigments within seconds. In

contrast, photoreceptors are blind to light in the near-infrared wavelength region, allowing the use of IR-excited 2PM to measure light stimulus-evoked calcium signal in the functionally intact retina, which has been used to study the mechanism of motion detection in retinal cells (see Fig. 5).

5.2 Calcium Imaging

"Calcium imaging" constitutes the bulk of ion imaging, both with 1PM and with MPM, because of the important role of Ca^{2+} in coupling electrical processes to biochemical processes, for example, in muscular contraction and neurotransmitter secretion, and because $[Ca^{2+}]$ transients can often be used as a proxy for electrical excitation. The particular advantages of MPM are, again, excitation-based optical sectioning and highly efficient fluorescence collection in scattering tissue. A disadvantage of MPM is that ratiometric indicators that require dual-wavelength excitation are difficult to use because a second laser is required for the second excitation wavelength (at this point in time, changing the wavelength even in a fully automated Ti:Sapphire lasers still takes seconds). Therefore, ratiometric calcium imaging with 2PM has to resort

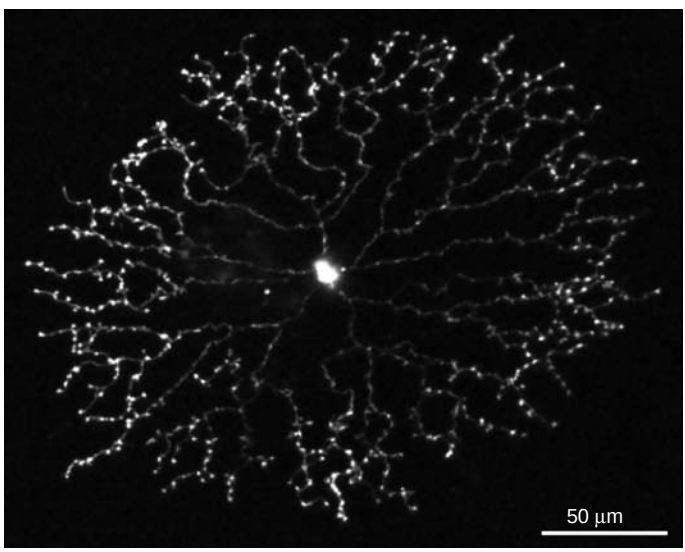

(a)

Fig. 5 Two-photon optophysiology in the retina. Dye-filled "starburst" amacrine cell (a) in flat-mounted rabbit retina. Like many amacrine cells, this neuron bears no axon; it receives inputs and makes output synapses with its dendrites. Starburst cells are involved in the detection of image motion. Using (b) two-photon microscopy, light stimulus-evoked Ca^{2+} signals (green trace) were recorded in the dendritic tips of a starburst amacrine cell. Simultaneously, membrane voltage (black trace) was measured at the soma using a patch electrode (see schematic drawing). The light stimulus, a concentric sinusoidal wave, induced stronger responses when it was expanding (left) than when it was contracting (right) (see color plate p. xxv).

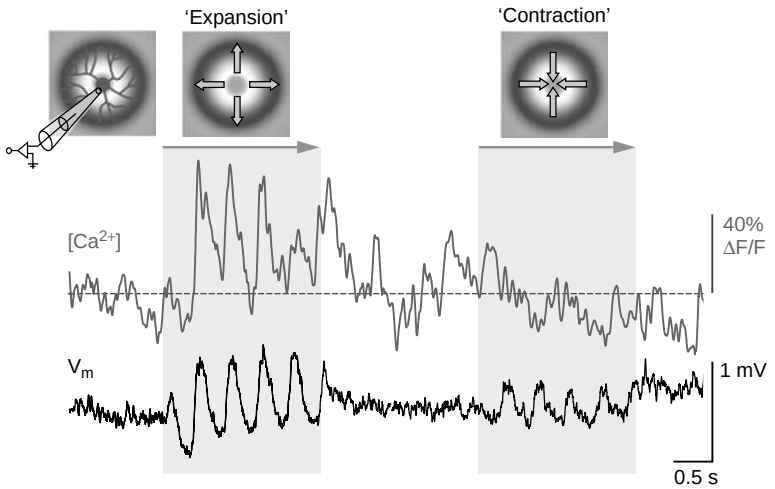

(b)

Fig. 5 (*Continued*)

to Indo-1, which has drawbacks, such as a short absorption wavelength and fast bleaching, but has been used to measure calcium oscillations in tumor mast cells and calcium transients in cardiac myocytes. More common nowadays is the use of dye mixtures that provide virtually all the advantages of proper ratiometric indicators. Furthermore, single-wavelength indicators are often sufficient, particularly if one is interested mainly in the temporal $[Ca^{2+}]$ dynamics (see Figs. 5 and 6). Satisfactory calibration can often be achieved by using the intensity prior to stimulation as a reference, such as was done in the early studies of $[Ca^{2+}]$ dynamics in dendrites, and dendritic spines in brain slices, and in whole animals.

5.3 Imaging of Metabolic Activity

The fluorescence from the reduced pyridine nucleotides NADH and NADPH, which are metabolic intermediates, can be used to localize and characterize the activity of respiratory enzymes and thus is an intrinsic probe of cellular metabolism. 1P-CM of two-dimensional NADH/NADPH-fluorescence maps is possible, as has been shown in rabbit cornea, but is hampered by photobleaching and photodamage due to the UV illumination (~360 nm) needed for the excitation of NADH/NADPH. Those problems are substantially reduced with 2PE so that now nearly continuous NADH/NADPH imaging is possible throughout the entire thickness (400 μm) of the rabbit cornea. Because metabolic events can be fast but may need to be monitored for extended time periods, low-damage imaging is of paramount importance. With MPM, it has become possible to follow the response to glucose application in β-cells inside intact pancreatic islets and in muscle cells.

5.4 Photoactivation (Uncaging)

The excellent localization of MPE can be used to confine photochemical generation

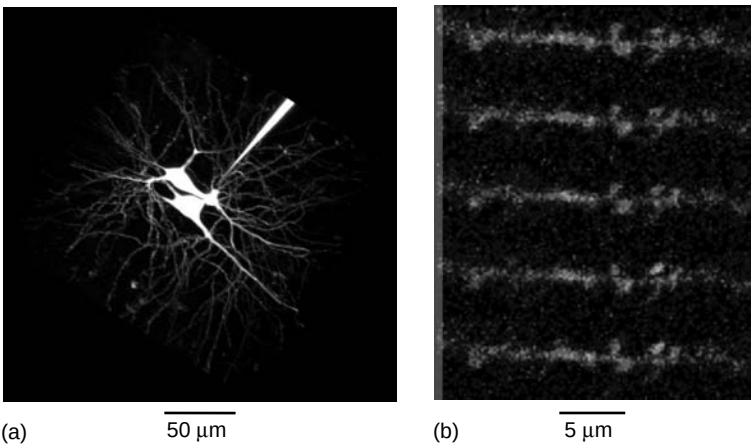

Fig. 6 Measurement of calcium dynamics in dendritic spines of a hippocampal neuron (a) using simultaneous excitation, at 910 nm, of two different fluorophores – a red fluorescent Ca^{2+}-insensitive dye (Alexa-594) helping in visualization of small structures and a green fluorescent Ca^{2+} indicator (Fluo-5F), which is very dim at rest, but shows large fluorescence changes in activated spines. (b) Following extracellular electric stimulation, presynaptic fibers release transmitter in the synaptic cleft whose binding to corresponding receptors on the postsynaptic membrane causes a Ca^{2+} rise in a single dendritic spine leading to an increase in green fluorescence observed by successive planar scans of a dendritic branchlet (see color plate p. xxvi).

of chemical compounds, such as fluorescent tracers or biological signaling molecules, to volumes smaller than a femtoliter. Typically, this is achieved by the illumination of a photolabile precursor that then absorbs two or more photons and decays via a photochemical pathway into the desired product. Scanning 2P photochemical microscopy, first demonstrated by mapping the distribution of neurotransmitter receptors on cultured muscle-like cells, has, however, been hampered by the very low 2PE cross sections of the available "caged" compounds (an example of the rule that weak 1P absorbers tend to be very weak 2P absorbers) even though there has been recent progress in the synthesis of better caged compounds, having allowed the study of calcium "sparks" in cardiac muscle cells and the mapping of glutamate sensitivities on dendritic spines. A combination of 2P uncaging and bleaching of fluorescence has been used to measure diffusional resistance of spine necks.

5.5 Human Diseases and Therapy

In conjunction with animal models, MPM has provided some insights into human diseases such as Alzheimer's dementia. There, 2PM *in vivo* imaging allows the observation over time of how senile plaques (amyloid-beta peptide aggregates) develop. This is extremely helpful when evaluating antiplaque therapies, since the fate of individual plaques can be followed over weeks and months by repeated imaging through a thinned skull. These experiments have shown that, surprisingly, plaques do not change size after formation.

2PM has also been used to measure blood flow at the level of single capillaries and to observe leucocytes-endothelial interactions in tumors *in vivo* at tissue depths that are not accessible otherwise. Such studies have revealed, for example, the depth dependent dynamics of tumor angiogenesis. In the area of immunology, 2PM allows the observation of lymphocytes in their native environment such as monitoring the motility of T cells that invade the brain during neuroinflammation.

Potential clinical applications on MPM are optical skin biopsy and photodynamic therapy. Traditional biopsy is an invasive procedure, which requires the removal and fixation of tissue before imaging. During this process, biochemical information is often poorly preserved. Optical biopsy does not require the removal of any tissue sample, but it remains to be seen whether the quality of 2PM images will become good enough to allow a pathological analysis with accuracy comparable to that of traditional biopsies.

Photodynamic therapy is based on the preferential accumulation of certain photosensitizing agents in tumorous or otherwise abnormal tissues and the subsequent selective tissue destruction by illumination. 2PE photodynamic therapy has the potential of providing more spatially specific destruction of, for example, cancer tissue than 1PE, which can be confined not as easily or not at all.

6 Future Directions

Among the "exponentially" growing number of papers on MPM, about half are still on methods development. With these ongoing improvements and extensions of MPM, the range of applications that benefit from MPM continues to expand. Much of the current improvement efforts are directed at excitation and detection efficiencies, which, as was discussed at length earlier, is the main factor determining imaging depth and specific (i.e. per information gained) photodamage. Novel chemical fluorophores with large two photon–absorption cross sections still await significant biological applications. Closer to being really useful appear to be semiconductor quantum dots (QDs), which have the largest 2PA cross sections measured, are very photostable, and are available with widely varying but narrow emission spectra. Recent advances have all but solved early problems with solubility, quenching, and toxicity. Different coats allow covalent linking to biorecognition molecules, such as antibodies or biotin/avidin, so that QDs can be used as specific fluorescent probes. In conjunction with their broad excitation spectra, QDs are very well suited for multilabel imaging.

An entirely different route to improving excitation efficiency is the use of coherent control, which involves tailoring the phase and amplitude of a laser pulse to optimize the optical response of a molecule, to achieve selective excitation, or to reduce photobleaching.

Inhomogeneities in the index of refraction across the tissue lead to degradation of the focus and thus a reduction in 2PE efficiency. In particular at large depths, therefore, adaptive aberration correction can significantly improve excitation.

In order to image structures that are beyond the penetration depth of the MPM, overlying tissue has to be removed. In order to keep the lateral extent of such removal, limited "endoscopic" approaches are being developed that use a small diameter gradient-index (GRIN) lens as

the objective. This limits the NA and the field of view, but may be the only way to reach deeper brain structures such as the hippocampus *in vivo*.

Regular MP microscopes are bulky and thus most *in vivo* imaging is done in anesthetized and immobilized animals. To achieve the goal of high-resolution imaging in freely moving animals, scanners and optics have to be miniaturized. This requires a different scanning mechanism because the weight of a galvo scanner rules out its use in a headpiece that can weigh at most a few tens of grams without encumbering an animal such as a rat too much. An alternative is the "2P-fiberscope" in which scanning is achieved by deflecting the free end of the fiber that is used to deliver the excitation light to the headpiece. The deflection can, for example, be achieved by using resonant oscillations. One challenge peculiar to nonlinear fiber microscopy is that special measures, including the use of large core or microstructured fibers, have to be taken to limit the degradation of the pulse shape by nonlinear optical effects in the beam-delivery fiber.

See also Alzheimer's Disease; Calcium Biochemistry; Metabolic Basis of Cellular Energy; Molecular Neurobiology, Single-Cell; Neuron Chemistry.

Bibliography

Books and Reviews

Denk, W., Svoboda, K. (1997) Photon upmanship: why multiphoton imaging is more than a gimmick, *Neuron* **18**(3), 351–357.

Denk, W., Piston, D.W., Webb, W.W. (1995) Two-photon molecular excitation in laser scanning microscopy, in: Pawley, J. (Ed.) *The Handbook of Confocal Microscopy*, Plenum, New York, pp. 445–458.

Diaspro, A., 2002 *Confocal and Two-photon Microscopy: Foundations, Applications, and Advances*, Wiley-Liss, New York, pp. xix, 567.

Helmchen, F., Denk, W. (2002) New developments in multiphoton microscopy, *Curr. Opin. Neurobiol.* **12**(5), 593–601.

König, K. (2000) Multiphoton microscopy in life sciences, *J. Microsc. (Oxford)* **200**, 83–104.

Masters, B.R., Thompson, B.D. (Eds.) (2003) *Selected Papers on Multi-photon Excitation Microscopy*, SPIE Milestone Series, Vol. MS175. SPIE Press, Bellingham, D.C.

Mertz, J. (2004) Nonlinear microscopy: new techniques and applications, *Curr. Opin. Neurobiol.* **14**(5), 610–616.

So, P.T.C., et al. (2000) Two-photon excitation fluorescence microscopy, *Annu. Rev. Biomed. Eng.* **2**, 399–429.

Zipfel, W.R., Williams, R.M., Webb, W.W. (2003) Nonlinear magic: multiphoton microscopy in the biosciences, *Nat. Biotechnol.* **21**(11), 1368–1376.

Primary Literature

Abbe, E. (1873) Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung, *Schultzes Arch. Mikrosk. Anat.* **9**, 413–468.

Albota, M., et al. (1998) Design of organic molecules with large two-photon absorption cross sections, *Science* **281**(5383), 1653–1656.

Barad, Y., et al. (1997) Nonlinear scanning laser microscopy by third harmonic generation, *Appl. Phys. Lett.* **70**(8), 922–924.

Bardeen, C.J., et al. (1999) Effect of pulse shape on the efficiency of multiphoton processes: implications for biological microscopy, *J. Biomed. Opt.* **4**(3), 362–367.

Beaurepaire, E., Mertz, J. (2002) Epifluorescence collection in two-photon microscopy, *Appl. Opt.* **41**(25), 5376–5382.

Beaurepaire, E., Oheim, M., Mertz, J. (2001) Ultra-deep two-photon fluorescence excitation in turbid media, *Opt. Commun.* **188**(1–4), 25–29.

Bennett, B.D., et al. (1996) Quantitative subcellular imaging of glucose metabolism within intact pancreatic islets, *J. Biol. Chem.* **271**(7), 3647–3651.

Bewersdorf, J., Pick, R., Hell, S.W. (1998) Multifocal multiphoton microscopy, *Opt. Lett.* **23**(9), 655–657.

Bhawalkar, J.D., et al. (1997) Two-photon photodynamic therapy, *J. Clin. Laser Med. Surg.* **15**, 201–204.

Bird, D., Gu, M. (2003) Two-photon fluorescence endoscopy with a micro-optic scanning head, *Opt. Lett.* **28**(17), 1552–1554.

Birge, R.R. (1986) 2-photon spectroscopy of protein-bound chromophores, *Acc. Chem. Res.* **19**(5), 138–146.

Brakenhoff, G.J., et al. (1996) Real-time two-photon confocal microscopy using a femtosecond, amplified Ti:sapphire system, *J. Microsc. (Oxford)* **181**, 253–259.

Brown, E.B., et al. (2001) In vivo measurement of gene expression, angiogenesis and physiological function in tumors using multiphoton laser scanning microscopy, *Nat. Med.* **7**(7), 866–870.

Centonze, V.E., White, J.G. (1998) Multiphoton excitation provides optical sections from deeper within scattering specimens than confocal imaging, *Biophys. J.* **75**(4), 2015–2024.

Chalfie, M., et al. (1994) Green fluorescent protein as a marker for gene expression, *Science* **263**(5148), 802–805.

Chan, W.C.W., Nie, S.M. (1998) Quantum dot bioconjugates for ultrasensitive nonisotopic detection, *Science* **281**(5385), 2016–2018.

Chance, B., Thorell, B. (1959) Localization and kinetics of reduced pyridine nucleotide in living cells by microfluorometry, *J. Biol. Chem.* **234**(11), 3044–3050.

Christie, R.H., et al. (2001) Growth arrest of individual senile plaques in a model of Alzheimer's disease observed by in vivo multiphoton microscopy, *J. Neurosci.* **21**(3), 858–864.

Chung, M.A., Lee, K.S., Jung, S.D. Two-photon absorption cross sections of dithienothiophene-based molecules, *ETRI J.* 2002. **24**(3), 221–225.

Cohen, L.B., et al. (1974) Changes in axon fluorescence during activity: molecular probes of membrane potential, *J. Membr. Biol.* **19**(1), 1–36.

de Grauw, G.J., et al. (1999) Imaging properties in two-photon excitation microscopy and effects of refractive-index mismatch in thick specimens, *Appl. Opt.* **38**(28), 5995–6003.

Denk, W. (1994) 2-photon scanning photochemical microscopy-mapping ligand-gated ion-channel distributions, *Proc. Natl. Acad. Sci. U.S.A.* **91**(14), 6629–6633.

Denk, W., et al. (1994) Anatomical and functional imaging of neurons using 2-photon laser-scanning microscopy, *J. Neurosci. Methods* **54**(2), 151–162.

Denk, W., Detwiler, P.B. (1999) Optical recording of light-evoked calcium signals in the functionally intact retina, *Proc. Natl. Acad. Sci. U.S.A.* **96**(12), 7035–7040.

Denk, W., Strickler, J.H., Webb, W.W. (1990) Two-photon laser scanning fluorescence microscopy, *Science* **248**, 73–76.

Denk, W., Sugimori, M., Llinas, R. (1995) Two types of calcium response limited to single spines in cerebellar Purkinje cells, *Proc. Natl. Acad. Sci. U.S.A.* **92**(18), 8279–8282.

Eilers, J., et al. (2001) GABA-mediated Ca^{2+} signaling in developing rat cerebellar Purkinje neurones, *J. Physiol. (London)* **536**(2), 429–437.

Engert, F., Bonhoeffer, T. (1999) Dendritic spine changes associated with hippocampal long-term synaptic plasticity, *Nature* **399**(6731), 66–70.

Euler, T., Detwiler, P.B., Denk, W. (2002) Directionally selective calcium signals in dendrites of starburst amacrine cells, *Nature* **418**(6900), 845–852.

Fan, G.Y., et al. (1999) Video-rate scanning two-photon excitation fluorescence microscopy and ratio imaging with cameleons, *Biophys. J.* **76**(5), 2412–2420.

Feierabend, M., Ruckel, M., Denk, W. (2004) Coherence-gated wave-front sensing in strongly scattering samples, *Opt. Lett.* **29**(19), 2255–2257.

Fork, R.L., Martinez, O.E., Gordon, J.P. (1984) Negative dispersion using pairs of prisms, *Opt. Lett.* **9**(5), 150–152.

Friedrich, D.M. (1982) 2-photon molecular-spectroscopy, *J. Chem. Educ.* **59**(6), 472–481.

Gannaway, J.N., Sheppard, C.J.R. (1978) Second-harmonic imaging in the scanning optical microscope, *Opt. Quantum Electron.* **10**(5), 435–439.

Goeppert-Mayer, M. (1931) Ueber Elementarakte mit zwei Quantenspruengen, *Ann. Phys.* **9**, 273.

Griesbeck, O. (2004) Fluorescent proteins as sensors for cellular functions, *Curr. Opin. Neurobiol.* **14**(5), 636–641.

Grutzendler, J., Kasthuri, N., Gan, W.B. (2002) Long-term dendritic spine stability in the adult cortex, *Nature* **420**(6917), 812–816.

Gu, M. (1996) Resolution in 3-photon fluorescence scanning microscopy, *Opt. Lett.* **21**(13), 988–990.

Gu, M., Sheppard, C.J.R. (1995) Comparison of 3-dimensional imaging properties between 2-photon and single-photon fluorescence microscopy, *J. Microsc. (Oxford)* **177**, 128–137.

Hanninen, P., Soini, E., Hell, S. (1994) Continuous wave excitation two-photon fluorescence microscopy, *J. Microsc. (Oxford)* **176**, 222–225.

Hell, S., Stelzer, E.H.K. (1992) Fundamental improvement of resolution with a 4pi-confocal fluorescence microscope using 2-photon excitation, *Opt. Commun.* **93**(5–6), 277–282.

Hell, S., Wichmann, J. (1994) Breaking the diffraction resolution limit by stimulated emission: stimulated emission depletion fluorescence microscopy, *Opt. Lett.* **19**(11), 780–782.

Hell, S.W., Dyba, M., Jakobs, S. (2004) Concepts for nanoscale resolution in fluorescence microscopy, *Curr. Opin. Neurobiol.* **14**(5), 599–609.

Hellwarth, R., Christiansen, P. (1974) Nonlinear optical microscopy examination of structure in polycrystalline ZnSe, *Opt. Commun.* **12**(3), 318–322.

Helmchen, F., et al. (1999) In vivo dendritic calcium dynamics in deep-layer cortical pyramidal neurons, *Nat. Neurosci.* **2**(11), 989–996.

Helmchen, F., et al. (2001) A miniature head-mounted two-photon microscope high-resolution brain imaging in freely moving animals, *Neuron* **31**(6), 903–912.

Helmchen, F., Tank, D.W., Denk, W. (2002) Enhanced two-photon excitation through optical fiber by single-mode propagation in a large core, *Appl. Opt.* **41**(15), 2930–2934.

Hentschel, M., et al. (2001) Attosecond metrology, *Nature* **414**(6863), 509–513.

Hockberger, P.E., et al. (1999) Activation of flavin-containing oxidases underlies light-induced production of H2O2 in mammalian cells, *Proc. Natl. Acad. Sci. U.S.A.* **96**(11), 6255–6260.

Jovin, T.M. 2003 Quantum dots finally come of age, *Nat. Biotechnol.* **21**(1), 32–33.

Jung, J.C., Schnitzer, M.J. (2003) Multiphoton endoscopy, *Opt. Lett.* **28**(11), 902–904.

Kaiser, W., Garrett, C.B.G. (1961) Two-photon excitation in CaF_2:Eu^{2+}, *Phys. Rev. Lett.* **7**(6), 229–231.

Kawano, H., et al. (2003) Attenuation of photobleaching in two-photon excitation fluorescence from green fluorescent protein with shaped excitation pulses, *Biochem. Biophys. Res. Commun.* **311**(3), 592–596.

Kettunen, P., et al. (2002) Imaging calcium dynamics in the nervous system by means of ballistic delivery of indicators, *J. Neurosci. Methods* **119**(1), 37–43.

Kim, K.H., Buehler, C., So, P.T.C. (1999) High-speed, two-photon scanning microscope, *Appl. Opt.* **38**(28), 6004–6009.

Kleinfeld, D., et al. (1998) Fluctuations and stimulus-induced changes in blood flow observed in individual capillaries in layers 2 through 4 of rat neocortex, *Proc. Natl. Acad. Sci. U.S.A.* **95**(26), 15741–15746.

Koester, H.J., et al. (1999) Ca2+ fluorescence imaging with pico- and femtosecond two-photon excitation: signal and photodamage, *Biophys. J.* **77**(4), 2226–2236.

König, K., Tirlapur, U.K. (2002) Cellular and Subcellular Perturbations during Multiphoton Microscopy, in: Diaspro, A. (Ed.) *Confocal and Two-photon Microscopy. Foundations, Applications and Advances*, Wiley-Liss, New York, pp. 191–205.

Kuhn, B., Fromherz, P., Denk, W. (2004) High sensitivity of stark-shift voltage-sensing dyes by one- or two-photon excitation near the red spectral edge, *Biophys. J.* **87**(1), 631–639.

Lakowicz, J.R., et al. (1999) Advances in fluorescence spectroscopy: multi-photon excitation, engineered proteins, modulation sensing and microsecond rhenium metal-ligand complexes, *Acta. Phys. Pol. A* **95**(1), 179–196.

Larson, D.R., et al. (2003) Water-soluble quantum dots for multiphoton fluorescence imaging in vivo, *Science* **300**(5624), 1434–1436.

Lechleiter, J.D., Lin, D.T., Sieneart, I. (2002) Multi-photon laser scanning microscopy using an acoustic optical deflector, *Biophys. J.* **83**(4), 2292–2299.

Lipp, P., Niggli, E. (1998) Fundamental calcium release events revealed by two-photon excitation photolysis of caged calcium in guinea-pig cardiac myocytes, *J. Physiol. (London)* **508**(3), 801–809.

Lippitz, M., et al. (2002) Two-photon excitation microscopy of tryptophan-containing proteins, *Proc. Natl. Acad. Sci. U.S.A.* **99**(5), 2772–2777.

Lombardo, J.A., et al. (2003) Amyloid-beta antibody treatment leads to rapid normalization of plaque-induced neuritic alterations, *J. Neurosci.* **23**(34), 10879–10883.

Maiti, S., et al. (1997) Measuring serotonin distribution in live cells with three-photon excitation, *Science* **275**(5299), 530–532.

Maletic-Savatic, M., Malinow, R., Svoboda, K. (1999) Rapid dendritic morphogenesis in CA1 hippocampal dendrites induced by synaptic activity, *Science* **283**(5409), 1923–1927.

Margrie, T.W., et al. (2003) Targeted whole-cell recordings in the mammalian brain in vivo, *Neuron* **39**(6), 911–918.

Masters, B.R., Kriete, A., Kukulies, J. (1993) Ultraviolet confocalfluoresence microscopy in the in vitro cornea: redox metabolic imaging, *Appl. Opt.* **32**(4), 592–596.

Masters, B.R., So, P.T.C., Gratton, E. (1997) Multiphoton excitation fluorescence microscopy and spectroscopy of in vivo human skin, *Biophys. J.* **72**(6), 2405–2412.

Matsuzaki, M., et al. (2001) Dendritic spine geometry is critical for AMPA receptor expression in hippocampal CA1 pyramidal neurons, *Nat. Neurosci.* **4**(11), 1086–1092.

McClain, W.M. Excited state symmetry assignment through polarized two-photon absorption studies of fluids, *J. Chem. Phys.* 1971. **55**(6), 2789–2796.

Mempel, T.R., et al. (2004) In vivo imaging of leukocyte trafficking in blood vessels and tissues, *Curr. Opin. Immunol.* **16**(4), 406–417.

Miller, M.J., et al. (2002) Two-photon imaging of lymphocyte motility and antigen response in intact lymph node, *Science* **296**(5574), 1869–1873.

Minsky, M. (1961) Microscopy Apparatus, in US Patent, 3013467.

Minta, A., Kao, J.P., Tsien, R.Y. (1989) Fluorescent indicators for cytosolic calcium based on rhodamine and fluorescein chromophores, *J. Biol. Chem.* **264**(14), 8171–8178.

Miyawaki, A., et al. (1997) Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin, *Nature* **388**(6645), 882–887.

Muller, M., et al. (1998) Dispersion precompensation of 15 femtosecond optical pulses for high-numerical-aperture objectives, *J. Microsc. (Oxford)* **191**, 141–150.

Neil, M.A.A., et al. (2000) Adaptive aberration correction in a two-photon microscope, *J. Microsc. (Oxford)* **200**, 105–108.

Nielsen, T., et al. 2001 High efficiency beam splitter for multifocal multiphoton microscopy, *J. Microsc.* **201**, 368–376.

Nitsch, R., et al. (2004) Direct impact of T cells on neurons revealed by two-photon microscopy in living brain tissue, *J. Neurosci.* **24**(10), 2458–2464.

Norris, T.B. (1992) Femtosecond pulse amplification at 250 kHz with a Ti-sapphire regenerative amplifier and application to continuum generation, *Opt. Lett.* **17**(14), 1009–1011.

Ouzounov, D.G., et al. (2002) Delivery of nanojoule femtosecond pulses through large-core microstructured fibers, *Opt. Lett.* **27**(17), 1513–1515.

Pastirk, I., et al. (2003) Selective two-photon microscopy with shaped femtosecond pulses, *Opt. Express* **11**(14), 1695–1701.

Patterson, G.H., Piston, D.W. (2000) Photobleaching in two-photon excitation microscopy, *Biophys. J.* **78**(4), 2159–2162.

Pawley, J., (Ed.) 1995 *Handbook of Biological Confocal Microscopy*, 2nd edition, Kluwer Academic Publishers, Norwell, MA.

Pelliccioli, A.P., Wirz, J 2002 Photoremovable protecting groups: reaction mechanisms and applications, *Photochem. Photobiol. Sci.* **1**(7), 441–458.

Piston, D.W., et al. (1994) 2-photon-excitation fluorescence imaging of 3-dimensional calcium-ion activity, *Appl. Opt.* **33**(4), 662–669.

Piston, D.W., Knobel, S.M. (1999) Real-time analysis of glucose metabolism by microscopy, *Trends Endocrinol. Metab.* **10**(10), 413–417.

Piston, D.W., Masters, B.R., Webb, W.W. (1995) 3-dimensionally resolved Nad(P)H cellular metabolic redox imaging of the in-situ cornea with 2-photon excitation laser-scanning microscopy, *J. Microsc. (Oxford)* **178**, 20–27.

Potter, S.M., Pine, J., Fraser, S.E. (1996) Neural transplant staining with DiI and vital imaging by 2-photon laser-scanning microscopy, *Scanning Microsc. Suppl.* **10**, 189–199.

Prasher, D.C., et al. (1992) Primary structure of the Aequorea victoria green-fluorescent protein, *Gene* **111**(2), 229–233.

Shen, Y.R. (1984) *The Principles of Nonlinear Optics*, Wiley, New York.

Sheppard, C.J.R. (1980) Scanning optical microscope, *Electron. Power* **26**(2), 166–172.

Shirakawa, A., Sakane, I., Kobayashi, T. (1998) Pulse-front-matched optical parametric amplification for sub-10-fs pulse generation tunable in the visible and near infrared, *Opt. Lett.* **23**(16), 1292–1294.

Spence, D.E., Kean, P.N., Sibbett, W. (1991) 60-fsec pulse generation from a self-mode-locked Ti:sapphire laser, *Opt. Lett.* **16**(1), 42–44.

Squier, J., Muller, M. (2001) High resolution nonlinear microscopy: a review of sources and methods for achieving optimal imaging, *Rev. Sci. Instrum.* **72**(7), 2855–2867.

Squirrell, J.M., et al. (1999) Long-term two-photon fluorescence imaging of mammalian embryos without compromising viability, *Nat. Biotechnol.* **17**(8), 763–767.

Stelzer, E.H.K., et al. (1994) Nonlinear absorption extends confocal fluorescence microscopy into the ultra-violet regime and confines the illumination volume, *Opt. Commun.* **104**(4–6), 223–228.

Stosiek, C., et al. (2003) In vivo two-photon calcium imaging of neuronal networks, *Proc. Natl. Acad. Sci. U.S.A.* **100**(12), 7319–7324.

Svaasand, L.O., Ellingsen, R. (1983) Optical properties of human brain, *Photochem. Photobiol.* **38**(3), 293–299.

Svoboda, K., et al. (1997) In vivo dendritic calcium dynamics in neocortical pyramidal neurons, *Nature* **385**(6612), 161–165.

Svoboda, K., Tank, D.W., Denk, W. (1996) Direct measurement of coupling between dendritic spines and shafts, *Science* **272**(5262), 716–719.

Szmacinski, H., Gryczynski, I., Lakowicz, J.R. (1998) Spatially localized ballistic two-photon excitation in scattering media, *Biospectroscopy* **4**(5), 303–310.

Tan, Y.P., et al. (1999) Fast scanning and efficient photodetection in a simple two-photon microscope, *J. Neurosci. Methods* **92**(1–2), 123–135.

Theer, P., Hasan, M.T., Denk, W. (2003) Two-photon imaging to a depth of 1000 mu m in living brains by use of a Ti:Al2O3 regenerative amplifier, *Opt. Lett.* **28**(12), 1022–1024.

Trachtenberg, J.T., et al. (2002) Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex, *Nature* **420**(6917), 788–794.

Treacy, E. (1969) Optical Pulse Compression With Diffraction Gratings, *IEEE J. Quantum Electron.* **QE-5**(9), 454–458.

Tsien, R.Y. (1981) A non-disruptive technique for loading calcium buffers and indicators into cells, *Nature* **290**(5806), 527–528.

Tsien, R.Y. (1989) Fluorescent-Probes of Cell Signaling, *Annu. Rev. Neurosci.* **12**, 227–253.

White, J., Amos, W., Fordham, M. (1987) An evaluation of confocal versus conventional imaging of biological structures by fluorescence light microscopy, *J. Cell Biol.* **105**, 41–48.

Wilson, T., Sheppard, C. (1984) *Theory and Practice of Scanning Optical Microscopy*, Academic Press, London.

Xu, C., Webb, W.W. (1996) Measurement of two-photon excitation cross sections of molecular fluorophores with data from 690 to 1050 nm, *J. Opt. Soc. Am. B* **13**(3), 481–491.

Yaroslavsky, A.N., et al. (2002) Optical properties of selected native and coagulated human brain tissues in vitro in the visible and near infrared spectral range, *Phys. Med. Biol.* **47**(12), 2059–2073.

Yasuda, R., et al. (2004) Imaging calcium concentration dynamics in small neuronal compartments, *Sci. STKE* **2004**(219), pl5.

Yuste, R., Denk, W. (1995) Dendritic spines as basic functional units of neuronal integration, *Nature* **375**(6533), 682–684.

Zhang, J., et al. (2002) Creating new fluorescent probes for cell biology, *Nat. Rev. Mol. Cell Biol.* **3**(12), 906–918.

Zumbusch, A., Holtom, G.R., Xie, X.S. (1999) Three-dimensional vibrational imaging by coherent anti-stokes Raman scattering, *Phys. Rev. Lett.* **82**(20), 4142–4145.

Ubiquitin Mediation Complex: *see* Proteasomes

Ubiquitin-Proteasome System for Controlling Cellular Protein Levels

Michael H. Glickman[1] *and Aaron Ciechanover*[2]
[1] *Department of Biology, Israel Institute of Technology, Haifa, Israel*
[2] *Faculty of Medicine, Israel Institute of Technology, Haifa, Israel*

1 Mechanisms of Ubiquitination and Protein Degradation 91
1.1 Ubiquitination 91
1.2 Selection of Proteins for Degradation 93
1.3 Degradation 94
1.4 Regulation of the Ubiquitin System 95

2 Biological Processes Regulated by the Ubiquitin System 96

3 Modification by Other Ubiquitin-like Proteins 98

4 Ubiquitination in Health and Disease 100
4.1 Cancer 100
4.2 Neurological Disorders 102
4.3 Cystic Fibrosis 103
4.4 Immune Response and Inflammatory Disorders 104

5 Summary and Outlook 105

Bibliography 105
Books and Reviews 105
Primary Literature 106

Keywords

Deubiquitination
It is the process of removing a ubiquitin tag from a target protein. Deubiquitination is carried out by ubiquitin-specific proteases (also known as deubiquitinating enzymes or

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

DUBs) that hydrolyze the peptide or isopeptide bond following the last residue of ubiquitin.

ER-associated Degradation (ERAD)
It removes proteins from the ER membrane or ER lumen. ERAD involves retrotranslocation of the target protein from the ER into the cytoplasm, ubiquitination, and proteolysis by the proteasome.

Proteasome (a.k.a. 26S proteasome)
It is a large, multisubunit ATP-dependent protease found in all eukaryotic cells responsible for most regulatory proteolysis. The proteasome is composed of a cylindrical *20S core particle* encompassing the proteolytically active subunits, and a *19S regulatory particle* that recognizes polyubiquitinated proteins and prepares them for proteolysis in the 20S core particle.

Proteolysis or Protein Degradation
It is the hydrolysis of peptide bonds in proteins yielding short peptides or amino acids.

Ubiquitin
(Ub) is a highly conserved, compact, and stable 76-residue protein found in all eukaryotes that can be covalently attached at its free carboxyl-terminus to amino groups on other proteins. This posttranslational modification of proteins by ubiquitin is called *ubiquitination* and is carried out by ubiquitinating enzymes.

Ubiquitinating Enzymes
They are a modular assembly of enzymes that are responsible for tagging a designated protein with *ubiquitin*. Ubiquitinating enzymes include an ATP-utilizing E1 that activates the free carboxyl-terminus of ubiquitin, a number of E2s that transfer the activated ubiquitin to the target protein, and numerous E3s that select unique substrates and mediate the action of the proper E2s.

Ubiquitination
It is the posttranslational modification of proteins with ubiquitin or a chain of multiple ubiquitin units. Usually, the outcome of ubiquitination is recognition and proteolysis by the *26S proteasome*. In some cases, ubiquitination does not lead to degradation but serves to alter intrinsic properties of the target protein, or direct its subcellular localization.

Ubiquitin-like Proteins
(Ubl) are small proteins that share homology with ubiquitin either structurally or in sequence. Similar to ubiquitin, most can be attached to the amino group of a lysine side chain on a target protein.

Cells contain many different kinds of proteins, each fulfilling structural, functional, or regulatory roles. Monitoring the state of all these proteins, as well as continuously adjusting their levels to suit demands is paramount to survival. The presence of damaged or mutated proteins, as well as altered levels of normal proteins could cause pathological conditions and even cell death. To exercise such quality control, cells are continuously spending energy both to synthesize new proteins, and to simultaneously degrade them, even though many may still be functional. An important characteristic of regulatory degradation is that it is specific; only the correct proteins are removed in a time-coordinated manner. Such extraordinary specificity is achieved by a modular system that identifies the proteins to be degraded, marks them by covalently attaching ubiquitin to an amino residue, and finally proteolyses the substrate by the 26S proteasome. Recognition of target proteins is carried out by a specific ubiquitin-protein ligase, called an E3. This protein recognizes the substrate and usually directs a ubiquitin-conjugating enzyme, an E2, to attach ubiquitin, a small 76 amino acid protein, onto the substrate. Ubiquitin molecules are often added to one another as well as to the substrate, resulting in chains of ubiquitin extending from the protein targeted for degradation. These polyubiquitin conjugates are then shuttled to the 26S proteasome, a large proteolytic complex, where they are degraded. Interestingly, ubiquitination is a reversible process, with deubiquitinating enzymes able to remove ubiquitin from the target before it can be recognized by the proteasome. Hence, transfer of the polyubiquitinated conjugate to the proteasome must happen swiftly or be shielded from these enzymes. The balance of these processes allows the ubiquitin-proteasome system to control the cellular levels and half lives of thousands of proteins making it a key player in basic biological pathways such as cell division, differentiation, signal transduction, trafficking, and quality control. Not surprisingly, aberrations in the system have been implicated in the pathogenesis of many diseases, certain malignancies, neurodegenerative disorders, inflammation, and immune response. Understanding the underlying mechanisms involved is important for the development of novel, mechanism-based drugs.

1 Mechanisms of Ubiquitination and Protein Degradation

1.1 Ubiquitination

The cellular levels of many proteins are kept in check by regulated degradation via the ubiquitin-proteasome pathway. Substrates are first tagged by covalent attachment of multiple ubiquitin molecules. These tagged proteins are then proteolysed by the 26S proteasome complex simultaneous with release of the ubiquitin tag. This last process is mediated by deubiquitinating enzymes (DUBs), a number of which are attached to the proteasome itself and work together to define the efficiency of the overall process.

Ubiquitin is a 76-residue protein and one of the most (if not the most) evolutionarily conserved proteins; its sequence and structure being almost identical in all

eukaryotes. Conjugation of ubiquitin to the protein substrate requires the participation of a slew of ubiquitinating enzymes that are usually broken down into three distinct classes. Initially, the ubiquitin-activating enzyme (UBA; also known as E1), activates ubiquitin in an ATP-dependent reaction to generate a high-energy thiol ester intermediate between the carboxyl-terminus of ubiquitin and the active site cysteine of E1 (E1-S∼ubiquitin). One of several E2 enzymes (ubiquitin-conjugating enzymes; UBCs) transfers the activated ubiquitin from E1, via an additional high-energy thiol ester intermediate, E2-S∼ubiquitin, to the substrate that is specifically bound to a member of the ubiquitin-protein ligase family, E3 (Fig.1). E3s catalyze the last step in the conjugation process: covalent attachment of ubiquitin to the substrate. The protein to be degraded is recognized by a specific E3. This protein directs the addition of ubiquitin from the E2 onto the substrate. There are several classes of E3 enzymes. Members of the RING finger-containing E3s, the largest family of ubiquitin ligases, mediate the transfer of the activated ubiquitin directly from the E2 to the E3-bound substrate. In the case of another class of E3, the so-called HECT (*Homologous to the E6-AP C-Terminus*) domain E3s, the ubiquitin is transferred from the E2 enzyme to an active site cysteine residue on the E3 to generate yet a third high-energy thiol ester intermediate, ubiquitin-S∼E3. Ubiquitin is then transferred from the E3 to the substrate protein that is bound to the ligase.

At the end of the ubiquitination cascade, the ubiquitin molecule is attached by

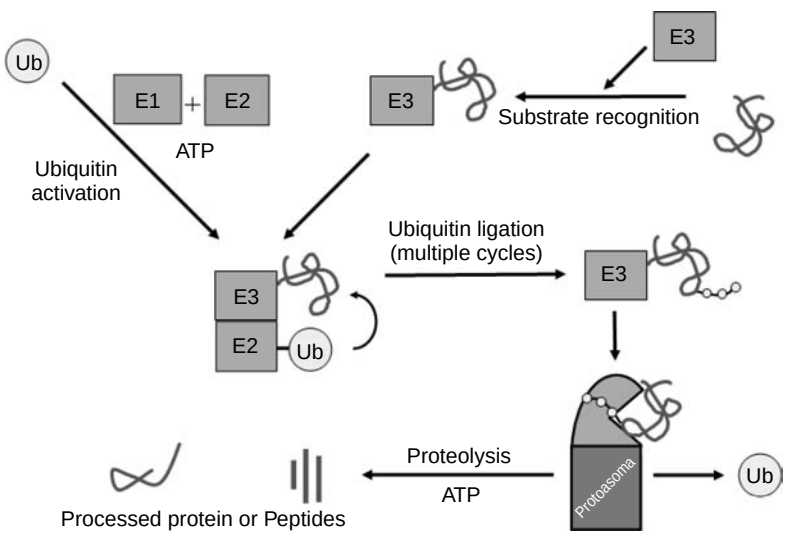

Fig. 1 General scheme of the ubiquitin system. Ubiquitin is activated by the ubiquitin-activating enzyme, E1, followed by its transfer to a ubiquitin-conjugating enzyme, E2. E2 transfers the activated ubiquitin to the protein substrate that is bound to a ubiquitin ligase, E3, which is specific for the substrate. Successive conjugation of ubiquitin moieties to one another generates a polyubiquitin chain that serves as the recognition signal for the 26S proteasome. The substrate is degraded by the 26S proteasome into short peptides, while ubiquitin is released by deubiquitinating enzymes (DUBs) for reuse.

its carboxyl-terminus to an amino group in the substrate protein. Generally, this is the amino group of an internal lysine side chain (an ε-NH_2 group), resulting in a covalent amide bond also known as an *isopeptide bond*. In some cases, ubiquitin is conjugated to the primary amino-terminus forming a linear amide bond; essentially a peptide bond. The ubiquitination reaction continues successively, adding an activated ubiquitin molecule to an internal lysine residue on the previously conjugated ubiquitin in each round, creating a polyubiquitin chain. In some instances, chain extension may require an additional auxiliary ligase called an *E4*. In this scenario, the first ubiquitin moiety is conjugated to the substrate by one E3, while chain elongation is catalyzed by a different ligase, the E4. It is thought that chains containing more than four ubiquitin moieties can bind to the proteasome with particularly high affinity, thus serving to target the protein to which they are conjugated for proteasomeal degradation.

1.2 Selection of Proteins for Degradation

A major task for the ubiquitin system is how to achieve high specificity and selectivity toward its numerous substrates. How to guarantee that the correct substrate will be removed whereas another, often similar protein will remain unscathed? In other words, why are certain proteins stable in the cell, while others are extremely short-lived? Complicating matters, some proteins are inherently short-lived and constitutively degraded while others are stable under most other conditions and degraded only upon a signal such as extracellular stimuli or a particular point in the cell cycle. Within the ubiquitin system, substrate selection is carried out predominantly by the E3 ligases. As such, E3s are the most diverse component of the pathway; to date, over 1000 different E3s or subunits of E3 ligase complexes have been identified in the human genome based on specific, commonly shared structural motifs. Each E3 is specific for a substrate or class of substrates.

Substrates must be recognized by a specific E3 and bind to it as a prerequisite to their ubiquitination. In most cases, however, the substrates are not recognized in a constitutive manner but must undergo posttranslational modification such as specific phosphorylation, dephosphorylation, or oxidation that renders them susceptible for recognition by an E3. In some cases, recognition of the target proteins by the E3 depends on association with an auxiliary protein (such as molecular chaperones) that serves to link the substrate with the appropriate ligase. In fact, many E3 ligases are themselves multisubunit complexes with each subunit participating in a unique network of protein–protein interactions. Certain subunits detach from, or dock to, the core E3 platform tweaking the ligase properties. Such modular behavior allows a limited set of proteins to target a much larger pool of substrates, each at the correct time and cellular context. Well-studied examples are the cullin-based E3 ligases, which regulate many steps of the cell cycle. The cullin subunit and the ring-finger component Rbx1 (plus a few additional subunits) serve as a stable platform while the substrate-recognition domain is a modular component that can be replaced as needed thus directing the cullin-based E3 to ubiquitinate a wholly new subset of substrates. In still other targeting mechanisms, substrates have to dissociate from a complexed form in order to be susceptible for E3 recognition and ubiquitination. This is a common feature of transcription

factors that usually have to dissociate from their DNA segment before they can bind to an E3 ligase to be ubiquitinated. Stability of yet other proteins depends on oligomerization. Thus, modifying enzymes (such as kinases), ancillary proteins, or DNA sequences to which substrates bind, also play an important role in the recognition process, in addition obviously to the E3s. In some instances, it is the E3 that must "be switched on" by undergoing posttranslational modification in order to yield an active form that recognizes the target substrate more efficiently.

1.3 Degradation

Degradation of polyubiquitinated substrates is carried out by the 26S proteasome. This multicatalytic protease degrades polyubiquitinated proteins to short peptides and releases ubiquitin for reuse by the ubiquitination machinery. There are a few reports indicating that the 26S proteasome – *in vivo* as well as *in vitro* – can also proteolyse certain (usually unstable or misfolded) proteins that are not covalently modified with the ubiquitin tag. However, the proteasome is the only enzyme that can proteolyse polyubiquitinated proteins. It is composed of two subcomplexes: a 20S core particle (CP) that contains the catalytic sites and proteolytic activity, and a 19S regulatory particle (RP) that is essential for recognition and treatment of ubiquitinated substrates.

The 20S CP is a barrel-shaped structure composed of four stacked rings, two identical outer α-rings and two identical inner β-rings. The eukaryotic α- and β-rings are each composed of seven distinct subunits, giving the 20S complex the general structure of $\alpha_{1-7}\beta_{1-7}\beta_{1-7}\alpha_{1-7}$. Some (though not all) of the β-subunits are proteolytically active, with the remainder serving structural roles in forming the barrel structure. The proteolytically active β-subunits are threonine proteases, members of a unique N-terminal nucleophile hydrolase family. The active site nucleophile is the N-terminal threonine (the protein undergoes activation by posttranslational cleavage *after* it properly assembles into the 20S structure), while the proton acceptor is the N-terminus of the protein itself. This unique active site structure has drawn efforts to design proteasome-specific inhibitors that would not affect other cellular proteases. A number of proteasome inhibitors that target the 20S CP β-subunits are indeed available, with one, Valcade^{tm}, recently receiving FDA approval for treatment of multiple myeloma in humans.

Substrates must enter the lumen of the 20S CP where proteolysis occurs by transverseing through the α-ring at either of the two ends of the 20S barrel. It appears that they can enter from both ends simultaneously enhancing the proteolytic capacity of the proteasome. The α-ring at either of the two ends of the 20S barrel can be capped by a 19S RP. One important function of the 19S RP is to recognize ubiquitinated proteins and other potential substrates of the proteasome. Indeed, a number of polyubiquitin-binding subunits have been identified in the 19S RP, and are thought to recognize, transfer, bind, and anchor the tagged protein to the 19S RP. However, the interplay between these ubiquitin-binding subunits and how each contributes to proteasome function has yet to be deciphered. A second function of the 19S RP is to open an orifice in the α-ring of the 20S CP that will allow entry of the substrate into the internal proteolytic chamber. Even in its opened-state, the passage through the α-ring is too narrow to allow folded proteins through. It is assumed that the

19S regulatory particle unfolds substrates and inserts the unstructured polypeptide chain into the 20S CP. Both the channel opening function and the unfolding of the substrate require metabolic energy, and indeed, the 19S RP contains six different ATPase subunits. Biochemical characterization and mutational analysis have shown these ATPases to be involved in the unfolding and channel gating properties.

Following degradation of the substrate, short peptides derived from the substrate are released from the proteasome. The bulk of these peptides are further degraded in the cytoplasm into free amino acids by cytosolic amino- and carboxypeptidases. A small fraction of product peptides is transported by the TAP (transporter associated with antigen processing) across the ER membrane, further trimmed by specific proteases within the ER, bind to the MHC class I complex, and are carried to the cell surface where they are presented to cytotoxic T cells. In this manner, the proteasome is the major producer of antigenic peptides. Since the proteasome degrades both natural proteins as well as foreign (viral or mutated) proteins, both "self" and "nonself" antigens are produced. When these peptides are derived from a "nonself" protein, the T cell lyses the presenting cell. In rare instances, the proteasome may even splice nonconsecutive segments together, creating a new peptide with a novel sequence not found "as is" in the protein it was derived from. If substantiated as a general phenomenon, peptide splicing will have far reaching implications to the study of antigen presentation.

It should be emphasized that proteasomal degradation does not always continue to completion. In some cases, the proteasome can carve the ubiquitinated substrate in a limited manner, releasing a stable and functional truncated product. The best-studied example is the case of the NF-κB transcriptional regulator: an active subunit (p50 or p52) is generated by the proteasome from a longer inactive precursor (p105 or p100).

In addition to proteolysing the substrate (partially or in full) in the 20S CP, a totally unrelated proteolytic activity is present in the 19S RP and is intimately linked to proteasome function. Proteasomes hydrolyze the peptide or isopeptide bond immediately following the final residue of ubiquitin, thus severing the link between ubiquitin and the substrate or one ubiquitin unit and another in a polyubiquitin chain. This ubiquitin-specific protease activity has been mapped to a subunit of the 19S RP, Rpn11/POH1, as well as to a number of tightly associating components such as UCH37 and ubp6/USP14. The combined action of these DUBs releases ubiquitin, sparing it from being degraded along with the target protein and allowing it to be reused by the ubiquitin system.

1.4 Regulation of the Ubiquitin System

The ubiquitin-proteasome pathway can be regulated at the level of ubiquitination or at the level of proteasome activity. Since conjugation and proteasomal degradation are required for numerous cellular functions, regulation must be delicately executed. In a few cases, general regulation of the entire pathway is observed in response to physiological signals that call for increased or decreased proteolysis. For example, in fungi and plants, a dramatic increase in levels of ubiquitin, ubiquitinating enzymes, and proteasome subunits is observed under many stress conditions such as heat shock or oxidative damage. In animals, upregulation of the entire pathway

is observed during massive degradation of skeletal muscle proteins that occurs under normal fasting, and also under pathological conditions such as malignancy-induced cachexia, severe sepsis, metabolic acidosis, or following denervation. In general, these conditions are correlated with accumulation of damaged or misfolded proteins, which should be efficiently removed for cell survival; hence, the need for increased rates of proteolysis. Conversely, it has recently been shown that some conditions require less proteolysis for survival of the cell. During long-term starvation or prolonged and severe heat exposure, proteasome activity is inhibited and general proteolysis rates are slowed down in yeast cells.

In most cases, however, regulation is specific targeting defined substrates that are recognized by specific ligases. By far, the greatest degree of molecular complexity in the ubiquitin system is among the E3 ligases. There are hundreds of ubiquitin ligases encoded in the genomes of multicellular eukaryotes, in addition to hundreds of substrate-recognition subunits that function together with the E3 ligases to target specific substrates. The modular nature of these ligase complexes ensures that slight changes in ratios of the components will direct the pathway to degradation of specific pathways. Thus, many substrate-recognition components of E3 ligases are either differentially expressed or stabilized only upon certain physiological signals.

E3 activity or specificity can also be regulated by posttranslational modifications of defined motifs within the ligase or the substrate. The targeting motif can be a single amino acid residue (e.g. the N-terminal residue), a specific sequence (the destruction box in cyclins), or a domain (such as a hydrophobic patch) that is not normally exposed. In other cases, the motif is generated by a posttranslational modification such as phosphorylation (two neighboring serine residues in the case of IκBα, or a single serine residue in the case of p27), or oxidation (hydroxyproline in the case of the hypoxia-inducible factor-1α—HIF1α). Interestingly, phosphorylation is required also to activate certain ligases, such as c-Cbl. Many proteasome subunits, both in the 20S CP and in the 19S RP can be modified by phosphorylation or glycosylation. These modifications alter proteasome stability or activity, though it is still unclear whether the effect is general toward all substrates, or more interestingly, influence the efficiency of degradation of a subset of substrates.

Eukaryotic genomes also usually encode several dozen different deubiquitinating enzymes. Such a large number suggests that some of them may act only on particular substrates, and raises the possibility that the rate of degradation of some proteins is determined by how fast ubiquitin is removed from them rather than by how fast it is added. Furthermore, substrate delivery to the 26S proteasome probably requires accessory or adaptor molecules, which may be differentially expressed to promote degradation or some ubiquitinated substrates over others.

2 Biological Processes Regulated by the Ubiquitin System

The list of cellular proteins that are targeted by ubiquitin is growing rapidly. By removing these cellular proteins, ubiquitin-mediated proteolysis plays an important role in regulating many basic cellular processes. Among these processes are cell cycle, cell division, differentiation, organ development, cellular response to stress

and extracellular effectors, morphogenesis of neuronal networks, modulation of cell surface receptors ion channels and secretory pathways, DNA repair, transcriptional regulation, transcriptional silencing, long-term memory, circadian rhythms, regulation of the immune and inflammatory responses, and biogenesis of organelles.

The paradigm of regulated degradation, where involvement of the ubiquitin-proteasome pathway is well understood, is the cell cycle. Numerous cell cycle regulators such as cyclins, cyclin-dependent kinase inhibitors, proteins involved in sister chromatid separation, tumor suppressors, transcriptional activators and their inhibitors are all short-lived proteins that are rapidly proteolysed upon the correct cues. For example, mitosis must follow duplication of the chromosomes, so certain factors required for DNA synthesis are degraded at the end of S phase and are stabilized only once mitosis has been completed. Progression of mitosis requires that certain inhibitors of mitosis are quickly degraded only once the chromosomes are aligned and are ready to be separated to opposite poles, allowing the cell cycle to advance. The SCF family of ubiquitin-protein ligases is largely responsible for protein ubiquitination in the G1/S phase and the related APC/cyclosome complex performs similar functions in G2/M.

Signal transduction at numerous levels is kept in check by specific ubiquitination and degradation of key components from cell surface receptors, through kinases or phosphatases and down to the ultimate transcription factors. Receptor downregulation (such as growth-hormone receptor) by ubiquitin-dependent degradation is an important aspect of signal transduction. Downstream adaptor proteins such as c-Cbl have been shown to be ubiquitinated, as well. Recently, a number of kinases from the MapK, or other signaling pathways have been shown to be substrates of ubiquitinating enzymes. The synthesis of proteins and transcriptional regulation is also regulated in part by the ubiquitin-proteasome system. Numerous transcription factors are rapidly degraded, usually when dissociated from their DNA sequence. An incomplete list of short-lived transcription factors that are substrates of specific E3 ligases includes members of the AP-1, Id, HIF, P53, NF-κB families. Finally, metabolic pathways are under strict regulation as well. For example, the rate-limiting enzymes in both the sterol and unsaturated fatty acid biosynthetic pathways are located in the membrane of the ER, and their levels are regulated by ubiquitin-dependent degradation. Ubiquitin-dependent degradation of these enzymes is crucial for reducing their activity when levels of sterols or unsaturated fatty acids, respectively, become too high.

Proteasome-dependent proteolysis is a vital component of combating viral infections and aiding the immune response. Upon viral infection, Interferon-γ causes mammalian cells to synthesize new catalytic subunits of the 20S proteasomes with a preference for cleaving peptide bonds between hydrophobic amino acids. These βi subunits are incorporated into newly assembled proteasomes – the so-called *immuno-proteasomes* – with an increased ability to generate protein fragments appropriate for MHC class I presentation. Interferon-γ also induces expression of the 11S regulator, which facilitates the production of antigenic peptides for optimal binding to MHC class I molecules and accelerating the cytotoxic lymphocyte response.

The ubiquitin-proteasome system plays a critical role in heat shock and oxidative damage response by removing mutated,

denatured, or misfolded proteins. It appears that in some cases, chaperones (such as Hsp70) cooperate with the ubiquitinating machinery to recognize and ubiquitinate unfolded proteins. By selectively removing these proteins, the system plays a key role in protection against aggregate buildup and in the quality control of the proteome.

All secretory proteins and most transmembrane proteins are inserted into or pass through the membrane of the ER. These proteins must fold, assemble into complexes with other proteins, and undergo various posttranslational modifications. If these proteins fail to fold or assemble properly, or lack appropriate glycosylation, they are prevented from proceeding to the Golgi and instead they are sent back to the cytosol for proteasomal degradation in a process known as *ER-associated degradation* (ERAD). Elaborate machinery exists on the cytoplasmic side of the ER membrane to recognize target proteins, to polyubiquitinate, to extract from the membrane in an ATP-dependent manner and to shuttle to the proteasome for proteolysis. Extraction from the membrane into the cytosol is probably carried out mainly by a hexameric ATPase ring called *Cdc48* or *p*97. This complex is located on the cytoplasmic surface of the membrane, can bind to polyubiquitinated substrates and undergo ATP-dependent conformational changes that ratchet the substrate through the ER channel.

Finally, it is important to note that ubiquitination also plays nondegradative roles in regulating certain biological processes. For example, histone H2A is monoubiquitinated in many places, though not proteolysed. Histone ubiquitination is thought to influence chromatin structure and accessibility. Ubiquitination can also serve to target proteins to certain localizations in the cell. In one case, monoamine oxidase A is ubiquitinated and inserted into the outer membrane of the mitochondria though there is no evidence so far that it is targeted for proteasome proteolysis on the way. Modification by polyubiquitin chains that polymerase via lysine 63 on each ubiquitin molecule (rather than the more common polymerization via Lys48) plays a role in a variety of processes, including endocytosis, postreplicative DNA, stress response, mitochondrial DNA inheritance, ribosomal function. Even though such chains probably do bind the proteasome, experimental evidence so far suggests that this type of modification does not involve proteolysis of the target substrate, but rather the modification plays a role in the activation/inactivation of the target protein.

3 Modification by Other Ubiquitin-like Proteins

In parallel to the ubiquitin-conjugating system, there are additional protein conjugation systems. The modifiers are always small globular proteins structurally related to ubiquitin and are known as *ubiquitin-like modifiers* (Ubls). These include SUMO, Nedd8/Rub1, ISG15, Apg8, and Apg12. In all the cases so far, the conjugation process has been carried out in an analogous manner to ubiquitination, utilizing conjugating enzymes that are similar to (yet distinct from) ubiquitinating enzymes. Modification is always on a lysine side chain of the modified proteins to which the carboxyl-terminus of the Ubl is covalently attached. Thus,

SUMO (small ubiquitin modifier), for example, is conjugated by a heterodimeric E1 – Aos1/Uba2, and the E2-conjugating enzyme Ubc9. Though Ubc9 can recognize the SUMOylation motif and transfer SUMO directly to certain substrates, for some proteins, specific E3 enzymes have been described that mediate these processes. A notable distinction between ubiquitination and modification by Ubls, however, is that the latter modification is monomeric, with only one Ubl molecule attached to the target. Recently, evidence has been provided that the ubiquitin-like modifiers SUMO-2 and SUMO-3 may form short chains. Similarly to ubiquitination, modification by Ubls is reversible on account of ubiquitin-like proteases (Ulps) that can shave the modifier off its target. A rapidly increasing number of Ulps have been identified, usually with clear homology to deubiquitinating enzymes.

Both enzymes and substrates of the ubiquitin system as well as other proteins have been found to be modified by ubiquitin-like proteins. The process serves many functions, such as activation of enzymes and transcriptional regulators, subcellular localization, or protecting proteins from ubiquitination. One role of modification by Ubls is to modulate the ubiquitin system. Modification of E3 ubiquitin ligases by ubiquitin-like molecules affects their activity. For example, conjugation of NEDD8/Rub1 to the cullin component of the SCF ligase complex increases its affinity to bind the E2 component of the conjugation machinery, thus increasing overall efficiency of conjugation. In the case of substrates, since modification occurs on lysines – the same residue to which ubiquitin is conjugated – modification can block ubiquitination and consequently enhance cellular stability. For example, in the case of IκBα, the inhibitor of the transcriptional regulator NF-κB, modification by SUMO-1 was shown to protect the substrate from ubiquitination. The activity of yet other transcription factors such as C-Jun, p53, c-Myb to name a few is also altered upon SUMOylation. In a completely different case, SUMOylation of RanGAP1 targets the protein to its final subcellular destination in the nuclear pore complex, NPC. Many SUMOylated proteins are found in promyelocytic leukemia (PML) nuclear bodies suggesting that SUMO may serve as a targeting signal to these or other locals. Complicating matters is that p53 can also be modified by NEDD8, though the exact relationship between ubiquitination, SUMOylation, and NEDDylation of this protein is not yet understood. Modification by the Ubl proteins Apg12 and Apg8 participates in autophagy, while modification by ISG15 functions in interferon-γ signaling.

A completely different group of ubiquitin-like proteins are those that are not involved in protein modification. Thus, certain proteins contain an intrinsic domain that folds into a ubiquitin-like domain. Many are proteins that are part of the ubiquitin system: among them are some E3 ligases, deubiquitinating enzymes, and polyubiquitin-binding proteins. For example, Parkin is an E3 ligase that contains a ubiquitin-like domain that probably facilitates its interaction with other components of the system. hHR23/Rad23, hPLIC/Dsk2, and Ddi1 are ubiquitin-binding proteins that all have an N-terminal Ubl. These proteins have all been implicated in the recognition and turnover of substrates targeted for degradation. They probably shuttle polyubiquitinated substrate to the

proteasome to which they bind via their Ubl domain.

4
Ubiquitination in Health and Disease

Inactivation of the key junctions in the ubiquitin–proteasome pathway, such as the enzyme E1 or the proteasome is lethal. Yet, mutations or acquired changes in enzymes or recognition motifs in substrates that do not affect vital pathways or that only partially affect proteolysis may result in a broad array of diseases. Pathological cases include those that result from loss of function – stabilization of a substrate that is normally targeted for degradation by evading the ubiquitin system or due to a mutation in a ubiquitin system enzyme – and those that result from gain of function such as abnormal or accelerated degradation of the protein target. Following are a number of well-studied cases that tie the ubiquitin system to pathogenesis of malignancies, neurodegenerative disorders, genetic diseases, and disorders of the immune and inflammatory responses.

4.1
Cancer

Alterations in ubiquitination and deubiquitination reactions have been directly implicated in the etiology of many malignancies. In general, cancers can result from *stabilization* of oncoproteins, or *destabilization* of tumor suppressor gene products. Some of the natural substrates of the system are oncoproteins that if not properly removed from the cell, can promote malignant transformation. For instance, ubiquitin targets N-Myc, c-Myc, c-Fos, c-Jun, Src, and the adenovirus E1A proteins. Similarly, inadvertent destabilization of tumor suppressors such as p53 and p27 has also been implicated in the pathogenesis of malignancies.

In the case of uterine cervical carcinoma, the level of the tumor suppressor protein p53 is extremely low. Most of these malignancies are caused by high-risk strains of the human papillomavirus (HPV). Detailed studies have shown that the suppressor is targeted for ubiquitin-mediated degradation by the virally encoded oncoprotein E6. Degradation is mediated by the normal and naturally occurring HECT domain E3 enzyme E6-associated protein – E6-AP, where E6 serves as an ancillary protein that allows recognition of p53. E6-AP will not recognize p53 in the absence of E6. E6 associates with both the ubiquitin ligase and the target substrate and brings them to the necessary proximity that is assumed to allow catalysis of conjugation to occur. Removal of the suppressor by the oncoprotein is probably an important mechanism used by the virus to transform cells. While the nature of the native substrates of E6-AP is still elusive, a mutation in the enzyme has been implicated directly as the cause of Angelman Syndrome characterized by mental retardation, seizures, out of context frequent smiling and laughter, and abnormal gait. It is possible that the protein, which accumulates as a result of the mutation, is selectively toxic to the developing neuronal cells.

Similar to the case of p53, low levels of the cyclin-dependent kinase inhibitor $p27^{Kip1}$ have been demonstrated in colorectal, prostate, and breast cancers. p27 acts as a negative growth regulator/tumor suppressor that binds and negatively regulates CDK2/cyclin E and CDK2/cyclin A complexes and thus does not allow cell cycle progression from G1 and entrance into the S phase. As noted, its level is markedly

reduced in several cancers, and in many of these cases there was a strong correlation between the low level of p27 and the aggressiveness of the disease – tumor grading, clinical staging, and poor prognosis of the patients. Dissection of the mechanism that underlies the decrease in p27 revealed that the protein is of the WT species and it is probably abnormal activation of the ubiquitin system, and, in particular, dysregulated overexpression of its cognate E3, the F-box protein Skp2, that leads to rapid removal of the suppressor.

β-catenin plays an important role in signal transduction and differentiation of the colorectal epithelium, and aberrations in ubiquitin-mediated regulation of its levels may play an important role in the multistep development of colorectal tumors. In the absence of signaling, casein kinase I (CKI) and glycogen synthase kinase 3β (GSK3β) are active and promote phosphorylation, recruitment of the β-TrCP ubiquitin ligase, ubiquitination, and subsequent degradation of β-catenin. Signaling promotes dephosphorylation, stabilization, and subsequent activation of β-catenin via complex formation with otherwise inactive subunits of transcription factor complexes. β-Catenin interacts with the 300 kDa tumor suppressor APC (adenomatous polyposis coli) and Axin to generate a complex that appears to regulate, in a yet unknown manner, its intracellular level. Aberrations in degradation of β-catenin that lead to its stabilization, accumulation, and subsequent oncogenic activation, can result from two distinct mechanisms: (1) mutations in the phosphorylation motif of the protein by CKI and GSK3β, and (2) mutations in the targeting APC/Axin machinery.

Mutations in components of the ubiquitination machinery can also cause malignancies. Mutations in one germline copy of *VHL* predisposes individuals to a wide range of malignancies, including more then 80% of sporadic cases of renal cell carcinoma, pheochromocytoma, cerebellar hemangioblastomas, and retinal angiomas. A hallmark of $VHL^{-/-}$ tumors is a high degree of vascularization that arises from constitutive expression of the α-subunit of the master switch transcription factor hypoxia-inducible factor-1 (HIF-1), which results in overexpression of hypoxia-inducible gene products including the crucial vascular endothelial growth factor (VEGF). It has been recently shown that pVHL is a subunit in a ubiquitin ligase that is involved in targeting of HIF-1α for ubiquitin- and proteasome-mediated degradation. HIF-1α heterodimerizes with the constitutively expressed HIF-1β to generate the active transcription factors. Under normoxic, and obviously hyperoxic, conditions, HIF-1α is hydroxylated specifically on Pro residue 564 to generate a hydroxyproline derivative. This hydroxylated proline residue is recognized by the pVHL E3 complex that targets the molecule for ubiquitination and subsequent degradation. Under hypoxic conditions, HIF-1α is stable, as the efficiency of the hydroxylation reaction under these conditions is extremely low. Loss of VHL function stabilizes HIF-1α, which can explain the stimulation of vascular growth in tumors in which VHL is mutated or lacking. Since overexpression of VEGF alone or many of the other known target proteins of HIF does not lead to malignant transformation, and since WT VHL can restore normal growth control in these malignant cells, researchers assume that pVHL and/or HIF must have additional, yet unknown substrates.

Defects in deubiquitinating enzymes can also lead to aberrations in growth control and to tumorigenesis. Overexpression of the ubiquitin proteases (UBP) Unp has

been found in lung carcinoma in humans and in mouse tumors. Unp may have a role in regulating the degradation of specific, yet unidentified, substrates in these tissues, though it does not seem to be crucial for general proteolysis. An interesting deubiquitinating enzyme is CYLD that is involved in deubiquitinating K63-Ub from NEMO (NF-kB essential modifier), TRAF2 [tumor necrosis factor receptor (TNFR)-associated factor 2] and TRAF6. Ubiquitination of NEMO [a regulator of the IKK (IκB kinase) signaling complex] and the TRAF proteins (which are ubiquitin ligases), and generation of polyubiquitin chains linked via Lys63 of the ubiquitin moiety, does not target these proteins for degradation, but results in their activation. Thus, inhibition of deubiquitination of NEMO, by mutation in CYLD, for example, may lead to uncontrolled activation of the IKK complex with increased activity of NF-κB. Indeed, CYLD was found mutated in familial cylindromatosis, a rare pathology characterized by predisposition to multiple tumors of the skin appendages and the salivary gland.

4.2 Neurological Disorders

Accumulation of ubiquitin conjugates and/or inclusion bodies associated with ubiquitin, proteasome, and certain disease-characteristic proteins, have been reported in a broad array of chronic neurodegenerative diseases, such as the neurofibrillary tangles of Alzheimer's disease (AD), brainstem Lewy bodies (LBs) – the neuropathological hallmark in Parkinson's disease (PD), and nuclear inclusions in CAG repeat expansion (polyglutamine extension) disorders such as occurring in Huntington's disease. However, in all these cases, a direct pathogenetic linkage to aberrations in the ubiquitin system has not been established. One factor that complicates the establishment of such linkage is the realization that many of these diseases, such as Alzheimer's and Parkinson's, are not defined clinical entities, but rather syndromes with different etiologies. Accumulation of ubiquitin conjugates in Lewy inclusion bodies in many of these cases may be secondary, and reflects unsuccessful attempts by the ubiquitin and proteasomal machineries to remove damaged/abnormal proteins. While the initial hypothesis was that inclusion bodies are generated because of the inherent tendency of the abnormal proteins to associate with one another and aggregate, it is now thought that the process maybe more complex and involves active cellular machineries, including inhibition of the ubiquitin system by the aggregated proteins. This aggregation of brain proteins into defined lesions is emerging as a common but poorly understood mechanistic theme in sporadic and hereditary neurodegenerative disorders.

The case of Parkinson's disease highlights the complexity of the involvement of the ubiquitin system in the pathogenesis of neurodegeneration. Mutations in several proteins such as α-synuclein that plays a role in synaptic vesicles dynamics and that leads to its stabilization and accumulation, or in the deubiquitinating enzyme UCH-L1, have been described that link the ubiquitin system to the pathogenesis of the disease. One important player in the pathogenesis of Parkinson's disease is Parkin, which is a RING-finger E3. Mutations in the gene appear to be responsible for the pathogenesis of autosomal recessive juvenile Parkinsonism (AR-JP), one of the most common familial

forms of Parkinson's disease. Parkin ubiquitinates and promotes the degradation of several substrates, such as CDCrel-1 (cell division control related protein), a synaptic vesicle-enriched septin GTPase or the Pael receptor, a putative G protein-coupled ER transmembrane polypeptide. It is possible that aberration in the degradation of one of these substrates that leads to its accumulation is neurotoxic and underlies the pathogenesis of AR-JP. However, it should be noted that such a protein has not been identified with any certainty. Complicating the situation is the recent finding that inactivation of the Parkin gene in mice does not lead to any symptoms that are even reminiscent of Parkinson's disease.

Alzheimer's disease is characterized by accumulation/association of ubiquitin with the phosphorylated form of Tau in neurofibrillary tangles and senile plaques, two lesions that are characteristic of the neuronal abnormalities associated with the disease. However, the role of Tau and other putative target proteins in the pathogenesis of the disease is still not clear. A more direct, relationship between the ubiquitin system and pathogenesis of Alzheimer's disease was established with the discovery of a frameshift mutation in the ubiquitin transcript, which leads to extension of the molecule with 20 amino acid residues [Ub(+1)], and, which has been selectively observed in the brains of Alzheimer's disease patients, including those with late onset, nonfamilial disease. Ub(+1) is an efficient acceptor for polyubiquitination, though it cannot be activated by E1 (as it lacks the essential G76 residue) and be transferred to a substrate or to another ubiquitin moiety. The resulting polyubiquitin chains are refractory to disassembly by deubiquitinating enzyme, in particular, isopeptidase T that requires for its activity an exposed G76 residue at the proximal ubiquitin moiety. The accumulated polyubiquitin chains block proteasomal degradation, which results in neuronal apoptosis. Thus, expression of Ub(+1) in the brain, that increases apparently with aging, can potentially result in dominant inhibition of the ubiquitin-proteasome system, leading to accumulation of toxic proteins with neuropathologic consequences. Since (Ub(+1)) was described also in other disorders such as Down's syndrome or supranuclear palsy, it is clear that it is not entirely specific to Alzheimer's disease and a major problem of how the mutation leads to distinct pathologies in different patients remains unsolved.

4.3 Cystic Fibrosis

The cystic fibrosis gene encodes the CF transmembrane regulator (CFTR) that is a chloride channel. Only a small fraction of the protein matures to the cell surface, whereas most of it is degraded from the endoplasmic reticulum (ER) prior to its maturation by the ubiquitin system. One frequent mutation in the channel is ΔF508. The mutation leads to an autosomal recessive inherited multisystem disorder characterized by chronic obstruction of airways and severe maldigestion due to exocrine pancreatic dysfunction. Despite normal ion channel function, $CFTR^{\Delta F508}$ does not reach the cell surface at all, and is retained in the ER from which it is degraded. It is possible that the rapid and efficient degradation results in complete lack of cell surface expression of the F508 protein, and therefore contributes to the pathogenesis of the disease.

4.4 Immune Response and Inflammatory Disorders

A wide array of immune and inflammatory disorders can be caused by untoward activation of the immune system central transcription factor NF-κB. Activation of the factor stimulates transcription of many cytokines, adhesion molecules, inflammatory response, and stress proteins, and immune system receptors. The factor is activated by the ubiquitin system via a two-step proteolytic mechanism: (1) limited processing of the precursor protein p105 to yield the active subunit p50, and (2) signal-induced phosphorylation and subsequent degradation of the inhibitor IκBα that enables translocation of the factor into the nucleus where it initiates specific transcriptional activity. An interesting case in that respect involves mutations in NEMO. Mutations in the protein lead to a series of diseases that affect the skin, among them incontinentia pigmenti (IP), hypohidrotic/anhidrotic ectodermal dysplasia, but also, as expected, immune deficiency. The immunological and infectious features observed in patients result from impaired NF-κB signaling, including cellular response to LPS, and a variety of cytokines.

As described above, the HPV evolved a mechanism for proteolytic removal of p53 that enables continuous replication and propagation of the virus under conditions of DNA damage (insertion of the viral genome into the cellular DNA) that normally would have ended with p53-induced apoptosis. Two other viruses evolved mechanisms that also utilize the ubiquitin system, here to escape immune surveillance. In one case, the Epstein-Barr Virus (EBV) EBV nuclear antigen 1 (EBNA1) persists in healthy virus carriers for life, and is the only viral protein regularly detected in all EBV-associated malignancies, such as Burkitt's lymphoma. Unlike EBNAs 2–4 that are strong immunogens, EBNA1 is not processed and cannot elicit a cytoxic T-cell response. The persistence of EBNA1 contributes, most probably, to some of the pathologies caused by the virus. An interesting structural feature common to all EBNA1 protein is a relatively long and unusual glycine–alanine repeat at the N-terminal domain of the molecule. These Gly–Ala repeats constitute a cis-acting element that inhibits processing of the protein by the 26S proteasome and subsequent presentation of potential antigenic epitopes. Thus, the evolution of the Gly–Ala repeat enabled the virus to evade proteolysis and subsequent presentation to the immune system. An additional interesting observation involves the pathobiology of the human cytomegalovirus (CMV). The virus genome encodes two ER proteins, US2 and US11, that downregulate the expression of MHC class I heavy chain molecules. The MHC molecules are synthesized, transported to the ER where they are glycosylated, but shortly thereafter, in cells expressing US2 or US11, are transported back to the cytosol, deglycosylated, and degraded by the proteasome following ubiquitination. It appears that the viral products bind to the MHC molecules and escort\dislocate them to the translocation machinery where they are transported back into the cytosol.

Thus, it appears that evolution of many viruses has involved intimate recognition with a variety of proteolytic processes. This enabled the evolution of viral mechanisms that enhance the function, via subversion of the normal proteolytic machinery, of the viral replication and propagation machinery.

5
Summary and Outlook

Ubiquitination is involved in the majority of regulated protein degradation in the cell. For proteolytic purposes, usually a polyubiquitin chain is assembled on the protein allowing for tight binding to the 26S proteasome and efficient degradation. However, it should be emphasized that the process is not a unidirectional irreversible process and that ubiquitinated proteins are in a dynamic state balancing further rounds of ubiquitination, deubiquitination, or degradation by the proteasome. Ubiquitin-like proteins have also been found to be regulators of ubiquitination enzymes, or of intracellular targeting and localization. Many questions are still unanswered. Is ubiquitin a strict requirement for proteasome-dependent intracellular degradation, or can some substrates be degraded in an ubiquitin-independent manner? Furthermore, does the ubiquitin-proteasome pathway play additional nonproteolytic roles, such as protein remodeling, transcriptional activation or altering protein–protein interactions or with nucleic acids?

Because of the central role the ubiquitin system plays in such a broad array of basic cellular processes, development of drugs that modulate the activity of the system may be difficult. Inhibition of enzymes common to the entire pathway, such as E1 or the proteasome, may inhibit many processes nonspecifically, although at the right dosage they could affect some targets more than others. One example in which such inhibitors have shown to be beneficial is multiple myeloma, a malignancy that affects the bone marrow. Alternatively, development of small molecules that interfere with the recognition step between specific E3s and their substrates, or site-specific inhibitors of deubiquitinating enzymes may be productive approaches as well.

See also Alzheimer's Disease; Motor Neuron Diseases: Cellular and Animal Models; Mucoviscidosis (Cystic Fibrosis), Molecular Cell Biology of; Oncology, Molecular; Proteasomes; Protein and Nucleic Acid Enzymes; Transcription Factor NF-κB: Function, Structure, Regulation, Pathways, and Applications; Triplet Repeat Diseases.

Bibliography

Books and Reviews

Amerik, A.Y., Hochstrasser, M. (2004) Mechanism and function of deubiquitinating enzymes, *Biochim. Biophys. Acta* **1695**, 189–207.

Glickman, M.H., Ciechanover, A. (2002) The Ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction, *Physiol. Rev.* **82**, 373–428.

Goldberg, A.L., Elledge, S.J. Wade, J. (2001) *The Cellular Chamber of Doom*, Scientific American, January, 68–73.

Hicke, L., Dunn, R. (2003) Regulation of membrane protein transport by ubiquitin and ubiquitin-binding proteins, *Annu. Rev. Cell Dev. Biol.* **19**, 141–172.

Hilt, W., Wolf, D.H. (Eds.) (2000) *Proteasomes: The World of Regulatory Proteolysis*, Eurekah.com/LANDES BIOSCIENCE Publishing Company, Georgetown, TX.

Pickart, C.M. (2001) Mechanisms of ubiquitination, *Annu. Rev. Biochem.* **70**, 503–533.

Pickart, C.M., Cohen, R.E. (2004) Proteasomes and their kin: proteases in the machine age, *Nat. Rev. Mol. Cell Biol.* **5**, 177–187.

Schwartz, D.C., Hochstrasser, M. (2003) A superfamily of protein tags: ubiquitin, SUMO

and related modifiers, *Trends Biochem. Sci.* **28**, 321–328.

Weissman, A.M. (2001) Themes and variations on ubiquitylation, *Nat. Rev. Mol. Cell Biol.* **2**, 169–179.

Wolf, D.H., Hilt, W. (2004) The proteasome: a proteolytic nanomachine of cell regulation and waste disposal, *Biochim. Biophys. Acta (BBA) – Mol. Cell Res.* **1695**, 19.

Primary Literature

Bloom, J., Pagano, M. (2003) Deregulated degradation of the cdk inhibitor p27 and malignant transformation, *Semin. Cancer Biol.* **13**, 41–47.

Boyd, S.D., Tsai, K.Y., Jacks, T. (2000) An intact HDM2 RING-finger domain is required for nuclear exclusion of p53, *Nat. Cell Biol.* **2**, 563–568.

Buschmann, T., Fuchs, S.Y., Lee, C.G., Pan, Z.Q., Ronai, Z. (2000) SUMO-1 modification of Mdm2 prevents its self-ubiquitination and increases Mdm2 ability to ubiquitinate p53, *Cell* **101**, 753–762.

Chen, Z., Hagler, J., Palombella, V.J., Melandri, F., Scherer, D., Ballard, D., Maniatis, T. (1995) Signal-induced site-specific phosphorylation targets I κ B α to the ubiquitin-proteasome pathway. *Genes Dev.* **9**, 1586–1597.

Ciechanover, A., Hod, Y., Hershko, A. (1978) A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes, *Biochem. Biophys. Res. Commun.* **81**, 1100–1105.

Ciechanover, A., Heller, H., Katz-Etzion, R., Hershko, A. (1981) Activation of the heat-stable polypeptide of the ATP-dependent proteolytic system, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 761–765.

Ciechanover, A., Elias, S., Heller, H., Hershko, A. (1982) "Covalent affinity" purification of ubiquitin-activating enzyme, *J. Biol. Chem.* **257**, 2537–2542.

Ciechanover, A., Heller, H., Elias, S., Haas, A.L., Hershko, A. (1980) ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1365–1368.

Elsasser, S., Chandler-Militello, D., Muller, B., Hanna, J., Finley, D. (2004) Rad23 and Rpn10 serve as alternative ubiquitin receptors for the proteasome, *J. Biol. Chem.* **279**, 26817–26822.

Etlinger, J.D., Goldberg, A.L. (1977) A soluble ATP-dependent proteolytic system responsible for the degradation of abnormal proteins in reticulocytes, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 54–58.

Finley, D., Ciechanover, A., Varshavsky, A. (1984) Thermolability of ubiquitin-activating enzyme from the mammalian cell cycle mutant ts85, *Cell* **37**, 43–55.

Finley, D., Ozkaynak, E., Varshavsky, A. (1987) The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation, and other stresses, *Cell* **48**, 1035–1046.

Finley, D., Bartel, B., Varshavsky, A. (1989) The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis, *Nature* **338**, 394–401.

Fraser, J., Luu, H.A., Neculcea, J., Thomas, D.Y., Storms, R.K. (1991) Ubiquitin gene expression: response to enviormental changes, *Curr. Genet.* **1–2**, 17–23.

Galan, J., Haguenauer-Tsapis, R. (1997) Ubiquitin Lys63 is involved in ubiquitination of a yeast plasma membrane protein, *EMBO J.* **16**, 5847–5854.

Geyer, R.K., Yu, Z.K., Maki, C.G. (2000) The MDM2 RING-finger domain is required to promote p53 nuclear export, *Nat. Cell Biol.* **2**, 569–573.

Giles, R.H., van Es, J.H., Clevers, H. (2003) Caught up in a Wnt storm: Wnt signaling in cancer, *Biochim. Biophys. Acta* **1653**, 1–24.

Glotzer, M., Murray, A.W., Kirschner, M.W. (1991) Cyclin is degraded by the ubiquitin pathway, *Nature* **349**, 132–138.

Goebl, M.G., Yochem, J., Jentsch, S., McGrath, J.P., Varshavsky, A., Byers, B. (1988) The yeast cell cycle gene CDC34 encodes a ubiquitin-conjugating enzyme, *Science* **241**, 1331–1335.

Goldknopf, I.L., Busch, H. (1977) Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugate-protein A24, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 864–868.

Goldstein, G., Scheid, M.S., Hammerling, V., Boyse, E.A., Schlesinger, D.H., Niall, H.D. (1975) Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 11–15.

Gong, L., Yeh, E.T. (1999) Identification of the activating and conjugating enzymes of the NEDD8 conjugation pathway, *J. Biol. Chem.* **274**, 12036–12042.

Gray, D.A., Inazawa, J., Gupta, K., Wong, A., Ueda, R., Takahashi, T. (1995) Elevated expression of Unph, a proto-oncogene at 3p21.3, in human lung tumors, *Oncogene* **10**, 2179–2183.

Gross-Mesilaty, S., Hargrove, J.L., Chiechanover, A. (1997) Degradation of tyrosine aminotransferase (TAT)via the ubiquitin-proteasome pathway, *FEBS Lett.* **405**(2), 175–180.

Haas, A.L., Rose, I.A. (1982) The mechanism of ubiquitin activating enzyme, *J. Biol. Chem.* **257**, 10329–10337.

Haas, A.L., Rose, I.A., Hershko, A. (1981) Purification of the ubiquitin activating enzyme required for ATP-dependent protein breakdown, *Fed. Proc.* **40**, 1691.

Heessen, S., Masucci, M.G., Dantuma, N.P. (2005) The UBA2 domain functions as an intrinsic stabilization signal that protects Rad23 from proteasomal degradation, *Mol. Cell* **18**, 225–235.

Hershko, A., Tomkins, G.M. (1971) Studies on the degradation of tyrosine aminotransferase in hepatoma cells in culture, *J. Biol. Chem.* **246**, 710–714.

Hershko, A., Ciechanover, A., Rose, I.A. (1979) Resolution of the ATP-dependent proteolytic system from reticulocytes: a component that interacts with ATP, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3107–3110.

Hershko, A., Ciechanover, A., Rose, I.A. (1981) Identification of the active amino acid residue of the polypeptide of ATP-dependent protein breakdown, *J. Biol. Chem.* **256**, 1525–1528.

Hershko, A., Eytan, E., Ciechanover, A., Haas, A.L. (1982) Immunochemical analysis of the turnover of ubiquitin-protein conjugates in intact cells, *J. Biol. Chem.* **257**, 13964–13970.

Hershko, A., Heller, H., Elias, S., Chiechanover, A. (1983) Components of ubiquitin-protein ligase system, *J. Biol. Chem.* **258**, 8206–8214.

Hershko, A., Ciechanover, A., Heller, H., Haas, A.L., Rose, I.A. (1980) Proposed role of ATP in protein breakdown: conjugation of proteins with multiple chains of the polypeptide of ATP-dependent proteolysis, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1783–1786.

Hicke, L., Riezman, H. (1996) Ubiquitination of a yeast plasma membrane receptor signals its ligand-stimulated endocytosis, *Cell* **84**, 277–287.

Honda, R., Tanaka, H., Yasuda, H. (1997) Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53, *FEBS Lett.* **420**, 25–27.

Huang, D.T., Walden, H., Duda, D., Schulman, B.A. (2004) Ubiquitin-like protein activation, *Oncogene* **23**, 1958–1971.

Huang, L., Kinnucan, E., Wang, G., Beaudenon, S., Howley, P.M., Huibregtse, J.M., Pavletich, N.P. (1999) Structure of an E6AP-UbcH7 complex: Insights into ubiquitination by the E2-E3 enzyme cascade, *Science* **286**, 1321–1326.

Hunt, L.T., Dayhoff, M.O. (1977) Amino-terminal sequence identity of ubiquitin and the nonhistone component of nuclear protein A24, *Biochem. Biophys. Res. Commun.* **74**, 650–655.

Ivan, M., Kaelin, W.G. (2001) The von Hippel-Lindau tumor suppressor protein, *Curr. Opin. Genet. Dev.* **11**, 27–34.

Jensen, D.E., et al. (1998) BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression, *Oncogene* **16**, 1097–1112.

Kishino, T., Lalande, M., Wagstaff, J. (1997) UBE3A/E6-AP mutations cause Angelman syndrome, *Nat. Genet.* **15**, 70–73.

Kitada, T., et al. (1998) Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism, *Nature* **392**, 605–608.

Klotzbucher, A., Stewart, E., Harrison, D., Hunt, T. (1996) The 'destruction box' of cyclin A allows B-type cyclins to be ubiquitinated, but not efficiently destroyed, *EMBO J.* **15**, 3053–3064.

Kovalenko, A., Chable-Bessia, C., Cantarella, G., Israël, A., Wallach, D., Courtois, G. (2003) The tumour suppressor CYLD negatively regulates NF-kB signalling by deubiquitination, *Nature* **424**, 801–805.

Lorick, K.L., Jensen, J.P., Fang, S., Ong, A.M., Hatakayama, S., Weissman, A.M. (1999) RING finger mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11364–11369.

Lueders, J., Demand, J., Hoehfeld, J. (2000) The ubiquitin-related BAG-1 provides a link between the molecular chaperones Hsc70/Hsp70 and the proteasome, *J. Biol. Chem.* **275**, 4613–4617.

Lyapina, S., et al. (2001) Promotion of NEDD-CUL1 conjugate cleavage by COP9 signalosome, *Science* **292**, 1382–1385.

Ma, C.P., Slaughter, C.A., DeMartino, G.N. (1992) Identification, purification, and characterization of a protein activator (PA28)

of the 20S proteasome, *J. Biol. Chem.* **267**, 10515–10523.

Ma, C.P., Vu, J.H., Proske, R.J., Slaughter, C.A., DeMartino, G.N. (1994) Identification, purification, and characterization of a high molecular weight ATP-dependent activator (PA700) of the 26S proteasome, *J. Biol. Chem.* **269**, 3539–3547.

Mahaffey, D., Rechsteiner, M. (1999) Discrimination between ubiquitin-dependent and ubiquitin-independent proteolytic pathways by the 26S proteasome subunit 5a, *FEBS Lett.* **450**, 123–125.

Maniatis, T. (1999) A ubiquitin ligase complex essential for the NF-kB, Wnt/Wingless, and Hedgehog signaling pathways, *Genes Dev.* **13**, 505–514.

Margottin, F., et al. (1998) A novel human WD protein, human b-TrCp, that interacts with HIV-1 Vpu, connects CD4 to the ER degradation pathway through an F-box motif, *Mol. Cell* **1**, 565–574.

Marunouchi, T., Yasuda, H., Matsumoto, Y., Yamada, M. (1980) Disappearance of a basic chromosomal protein from cells of a mouse temperature-sensitive mutant defective in histone phosphorylation, *Biochem. Biophys. Res. Commun.* **95**, 126–131.

Matsumoto, Y., Yasuda, H., Marunouchi, T., Yamada, M. (1983) Decrease in uH2A (protein A24) of a mouse temperature-sensitive mutant, *FEBS Lett.* **151**, 139–142.

Matsumoto, Y., Yasuda, H., Mita, S., Marunouchi, T., Yamada, M. (1980) Evidence for the involvement of H1 histone phosphorylation in chromosome condensation, *Nature* **284**, 181–183.

Maxwell, P.H., et al. (1999) The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis, *Nature* **399**, 271–275.

Mitch, W.E., Goldberg, A.L. (1996) Mechanisms of muscle wasting: The role of the ubiquitin-proteasome pathway, *N. Engl. J. Med.* **335**, 1897–1905.

Mizushima, N., et al. (1998) A protein conjugation system essential for autophagy, *Nature* **395**, 395–398.

Moazed, D., Johnson, D. (1996) A deubiquitinating enzyme interacts with SIR4 and regulates silencing in S. cerevisiae, *Cell* **86**, 667–677.

Molinari, M., Milner, J. (1995) p53 in complex with DNA is resistant to ubiquitin-dependent proteolysis in the presence of HPV-16 E6, *Oncogene* **10**, 1849–1854.

Orian, A., et al. (2000) SCF-b-TrCP ubiquitin ligase-mediated processing of NF-kB p105 requires phosphorylation of its C-terminus by IkB kinase, *EMBO J.* **19**, 2580–2591.

Orlowski, M., Cardozo, C., Michaud, C. (1993) Evidence for the presence of five distinct proteolytic components in the pituitary multicatalytic proteinase complex. Properties of two components cleaving bond on the carboxyl side of branched chain and small neutral amino acids, *Biochemistry* **32**, 1563–1572.

Ortega, J., Bernard Heymann, J., Kajava, A.V., Ustrell, V., Rechsteiner, M., Steven, A.C. (2005) The axial channel of the 20 S proteasome opens upon binding of the PA200 activator, *J. Mol. Biol.* **346**, 1221.

Osaka, F., et al. (2000) Covalent modifier NEDD8 is essential for SCF ubiquitin-ligase in fission yeast, *EMBO J.* **19**, 3475–3484.

Panaro, F.J., Na, E., Dharmarajan, K., Pan, Z.Q., Valenzuela, D.M., DeChiara, T.M., Stitt, T.N., Yancopoulos, G.D., Glass, D.J. (2001) Identification of ubiquitin ligases required for skeletal muscle atrophy, *Science* **294**, 1704–1708.

Pickart, C.M., Rose, I.A. (1985) Functional heterogeneity of ubiquitin carrier proteins, *J. Biol. Chem.* **260**, 1573–1581.

Pugh, C.W., Ratcliffe, P.J. (2003) The von Hippel-Lindau tumor suppressor, hypoxia-inducible factor-1 (HIF-1) degradation, and cancer pathogenesis, *Semin. Cancer Biol.* **13**, 83–89.

Rabinovich, E., Kerem, A., Frohlich, K.U., Diamant, N., Bar-Nun, S. (2002) AAA-ATPase p97/Cdc48p, a cytosolic chaperone required for endoplasmic reticulum-associated protein degradation, *Mol. Cell. Biol.* **22**, 626–634.

Saeki, Y., Saitoh, A., Toh-e, A., Yokosawa, H. (2002) Ubiquitin-like proteins and Rpn10 play cooperative roles in ubiquitin-dependent proteolysis, *Biochem. Biophys. Res. Commun.* **293**, 986–992.

Saeki, Y., Tayama, Y., Toh-e, A., Yokosawa, H. (2004) Definitive evidence for Ufd2-catalyzed elongation of the ubiquitin chain through Lys48 linkage, *Biochem. Biophys. Res. Commun.* **320**, 840–845.

Scheffner, M., Huibregtse, J.M., Vierstra, R.D., Howley, P.M. (1993) The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein

ligase in the ubiquitination of p53, *Cell* **75**, 495–505.

Scheffner, M., Werness, B.A., Huibregtse, J.M., Levine, A.L., Howley, P.M. (1990) The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53, *Cell* **63**, 1129–1136.

Shamu, C.E., Story, C.M., Rapoport, T.A., Ploegh, H.L. (1999) The pathway of US11-dependent degradation of MHC class I heavy chains involves a ubiquitin- conjugated intermediate, *J. Cell Biol.* **147**, 45–58.

Shimura, H., et al. (2000) Familial Parkinson disease gene product, Parkin, is a ubiquitin-protein ligase, *Nat. Genet.* **25**, 302–305.

Tsvetkov, L.M., Yeh, K.H., Lee, S.J., Sun, H., Zhang, H. (1999) p27Kip1 ubiquitination and degradation is regulated by the SCFSkp2 complex through phosphorylated Thr187 in p27, *Curr. Biol.* **9**, 661–664.

Turner, C.T., Du, F., Varshavsky, A. (2000) Peptides accelerate their uptake by activating ubiquitin-dependent proteolytic pathway, *Nature* **405**, 579–583.

Vassilev, L.T., Vu, B.T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., Fotouhi, N., Liu, E.A. (2004) *In vivo* activation of the p53 pathway by small molecule antagonists of MDM2, *Science* **303**, 844–848.

Ward, C.L., Omura, S., Kopito, R.R. (1995) Degradation of CFTR by the ubiquitin-proteasome pathway, *Cell* **83**, 121–127.

Wigley, W.C., Fabunmi, R.P., Lee, M.G., Marino, C.R., Muallem, S., DeMartino, G.N., Thomas, P.J. (1999) Dynamic association of proteasomal machinery with the centrosome, *J. Cell Biol.* **145**, 481–490.

Wilkinson, K.D., Urban, M.K., Haas, A.L. (1980) Ubiquitin is the ATP-dependent proteolysis factor I of rabbit reticulocytes, *J. Biol. Chem.* **255**, 7529–7532.

Xie, Y., Varshavsky, A. (2001) RPN4 is a ligand, substrate, and transcriptional regulator of the 26S proteasome: a negative feedback circuit, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3056–3061.

Xirodimas, D., Saville, M., Bourdon, J., Hay, R., Lane, D. (2004) Mdm2-mediated NEDD8 conjugation of p53 inhibits its transcriptional activity, *Cell* **118**, 83–97.

Yaron, A., et al. (1997) Inhibition of NF-kB cellular function via specific targeting of the IkB-ubiquitin ligase, *EMBO J.* **16**, 6486–6494.

Ye, Y., Meyer, H.H., Rapoport, T.A. (2001) The AAA ATPase Cdc48/p97 and its partners transport proteins from the ER into the cytosol, *Nature* **414**, 652–656.

Zhang, Y., Chang, C., Gehling, D.J., Hemmati-Brivanlou, A., Derynck, R. (2001) Regulation of Smad degradation and activity by Smurf2, an E3 ubiquitin ligase, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 974–979.

Zhang, Y., Gao, J., Chung, K.K., Huang, H., Dawson, V.L., Dawson, T.M. (2000) Parkin functions as an E2-dependent ubiquitin-protein ligase and promotes the degradation of the synaptic vesicle-associated protein, CDCrel-1, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13354–13359.

Ultraviolet Radiation Damage to DNA: *see* DNA Damage Induced by Ultraviolet Radiation

Unnatural Amino Acid Incorporation: *see* Protein Expression by Expansion of the Genetic Code

Vaccination, Genetic

Joerg Reimann[1], *Martin Schleef*[2], *and Reinhold Schirmbeck*[1]
[1] *University of Ulm, Ulm, Germany*
[2] *Plasmid Factory, Bielefeld, Germany*

1 Principles 113
1.1 The Emerging Technique of DNA Vaccination 113
1.2 The Novel Aspect of DNA Vaccination 113

2 Techniques 115
2.1 Basic Structure of a Eukaryotic Expression Vector 115
2.2 Increasing the Antigenic Information Delivered by DNA Vaccines 116
2.3 Strategies to Enhance the Immunogenicity of DNA Vaccines 117
2.3.1 Synthetic Genes 117
2.3.2 Promoter/Enhancer Sequences 117
2.3.3 Enhancing Transfection Efficiency of DNA Vaccines 117
2.3.4 Subcellular Targeting of Antigen Expression 118
2.3.5 Targeting Antigens to APCs 118
2.3.6 Codelivery of Antigen with Immune-stimulating/modulating Cytokines and Chemokines 118
2.3.7 Codelivery of Transcription Factors 119
2.3.8 CpG-containing Oligodeoxynucleotides(ODN) 119
2.3.9 Stress Proteins 119
2.3.10 Copriming Specific "Help" 120
2.4 Production and Quality Control of DNA Vaccines 120
2.4.1 Quality Control and Safety in Plasmid DNA Vaccine Production 120
2.4.2 Plasmid Identity 120
2.4.3 Animal-free DNA Production 121
2.4.4 Genomic DNA 121
2.4.5 Host Cell Impurities 122
2.4.6 Importance of Supercoiled Forms 122
2.5 Stability of DNA Vaccines 124
2.5.1 Storage Stability of Plasmid DNA 124

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

2.5.2 Stability in Plasmid Vector Applications 124
2.6 Delivery of DNA Vaccines 124
2.6.1 Combining DNA Vaccines and Protein Vaccines 125
2.6.2 Routes of DNA Vaccine Delivery 125

3 Applications 125
3.1 Priming Immune Responses by DNA Vaccination 125
3.1.1 Priming T-cell Responses 125
3.1.2 Priming B-cell (Antibody) Responses 126
3.2 DNA Vaccination as an Experimental Tool 126
3.3 Potential Advantages of DNA Vaccination 126

4 Perspectives 127
4.1 Achievement and Unresolved Issues of the Technology 127
4.2 Preclinical and Clinical Trials that Test Potential Applications of DNA Vaccination in Clinical Practice 127
4.3 Risks and Safety Issues Raised by DNA Vaccination 128

Acknowledgment 129

Bibliography 129
Books and Reviews 129
Primary Literature 129

Keywords

DNA (genetic or nucleic acid) vaccination
Stimulating an immune response to antigen(s) produced *in vivo* by cells transfected with injected expression plasmid DNA.

In genetic or nucleic acid (DNA or RNA) vaccination, antigenic information is delivered by *in vivo* expression of antigen from inoculated vector DNA (or RNA). Combining different antigen-encoding sequences, modifying their sequence and/or codelivering them with sequences encoding immune-stimulating factors (readily achieved through recombinant DNA technologies) strikingly enhances and modulates the immunogenicity of genetic vaccines and extends the repertoire of epitopes to which different compartments of the adaptive immune system can be specifically primed. Large-scale production of DNA vaccines, their formulation and alternative modes of delivery are currently optimized. Experimental DNA vaccines have demonstrated efficacy in priming antibody and/or cellular immune responses that protect against a variety of pathogens or transplantable tumors in animal models.

Genetic vaccination has therefore become an unrivaled tool in preclinical research. Although highly attractive in principle, the practical development of DNA vaccination for clinical applications is at an early and experimental stage, and will have to resolve different issues before entering widespread use.

1 Principles

1.1 The Emerging Technique of DNA Vaccination

Active immune prophylaxis (vaccination) has been undeniably one of the greatest achievements in medicine. Vaccine development started centuries ago with the identification and large-scale production of attenuated pathogens. As these induced only limited clinically apparent disease after infection but expressed many of the immunogenic determinants of the pathogenic strain, they could elicit specific, protective immunity with a low burden of clinical disease after deliberate infection. This vaccination strategy evolved into the use of defined products of pathogens feasible after their molecular identification, inactivation (if they displayed a pathogenic potential like toxins, for e.g.), and formulation (to enhance their often weak intrinsic immunogenicity).

Genetic (nucleic acid) vaccination is a recently developed immunization technique in which antigen is expressed in immunogenic form (i.e. able to specifically prime an immune response) from inoculated nucleic acid fragments containing antigen-encoding sequences. The nucleic acid fragments are engineered to efficiently support expression of the respective antigen(s) *in situ*. Nucleic acid fragments used as genetic vaccines can be either plasmid DNA (DNA vaccines), or messenger (m) RNA (RNA vaccines). If cloned into suitable expression vectors, and delivered in an appropriate formulation at a suitable dose by a suitable route, these vaccines efficiently prime humoral and cellular immune responses to a large spectrum of protein antigens. The different steps in developing and testing a DNA vaccine are outlined in Table 1.

1.2 The Novel Aspect of DNA Vaccination

Genetic vaccination introduced new prospects into the field of vaccinology, which remain to be fully explored. Antigens produced *in situ* from DNA or RNA vaccines undergo all posttranslational modifications, such as folding, glycosylation, and oligomerizations (with host cell-derived or pathogen-derived molecules). This is a key issue in eliciting protective, neutralizing antibody (B cell) responses to many pathogens, as these usually bind only conformational epitopes. This is difficult to achieve by recombinant protein vaccines (expressed in a heterologous source and subjected to extensive purification procedures), but comparatively easy to achieve by genetic vaccines. Furthermore, genetic immunization has been shown to be exceptionally potent in stimulating cellular (T cell) responses. Antigenic peptides are generated from proteins produced from the DNA or RNA vaccine within a cell of the host. These can be introduced into

Tab. 1 The objectives of consecutive steps in DNA vaccination and the issues that have to be addressed.

Step	*Objective*	*Issues to be addressed*
1.	Construction of DNA vaccine	• Type of vector • Selection of antigen, antigenic domains, epitopes • Choice of costimulator (cytokine) to be codelivered
2.	Evaluate antigen expression from DNA vaccine	• Test amount, integrity, longevity of antigen expressed from the DNA vaccine in transient transfection assays
3.	Large-scale DNA vaccine production	• Production and isolation of supercoiled plasmid DNA on a large scale • Quality control (quantitative assessment of contaminants) • Define storage and shipment conditions
4.	Formulation of DNA vaccine	• Nonpackaged (''naked'') plasmid DNA • Plasmid DNA packaged into liposomes, lipoplexes, polymers, cationic peptides, aluminum phosphate • Particle-coated plasmid DNA • Transformation of bacteria
5.	Delivery of DNA vaccine	• Intramuscular, subcutaneous, intradermal injection • Directly ''shooting'' DNA-coated gold microprojectiles into cells of the skin or the mucosa with the ''gene gun'' • Electroporation • Oral infection with DNA vaccine-carrying bacteria
6.	*In vivo* expression of antigen from a DNA vaccine	• *In situ* production of secreted antigen for priming antibody responses (humoral immunity) • Production and processing of antigen to (cross) prime T-cell responses (cellular immunity)
7.	Evaluate the vaccine-induced, specific immunity	• Follow over weeks postvaccination the specific serum antibody titers elicited • Measure numbers and functional properties of specifically primed CD4 and CD8 T-cell responses • Test response to boost injections • Test longevity of specific immunological memory • Test protective efficacy of the specific immunity
8.	Evaluate long-term safety issues	• Persistence of integrated plasmid DNA in the vaccinated individual • Autoimmune reactions

different processing pathways to generate presentation-competent major histocompatibility complex (MHC) (class I or class II) molecules, an essential prerequisite to prime cellular (T cell) responses. Modifications introduced into the antigen expressed by a DNA vaccine can facilitate its processing, thereby enhancing its antigenicity for T cells. Hence, both compartments of the specific immune system can be specifically stimulated by genetic vaccines.

Large numbers of vaccine candidates have been identified in different pathogens

through sequencing of their genomes. DNA vaccination is an attractive tool to evaluate this expanding repertoire of potentially interesting antigens to experimentally identify vaccine candidates, and to evaluate and modulate their immunogenicity. This can be achieved through either expression library immunization, or genomic vaccination. DNA vaccines opened new ways to produce (native, variant, or chimeric) antigens in heterologous expression systems including complex, chimeric antigens that prime multivalent responses to different proteins of the same or different pathogens. Genetic vaccination has been shown to efficiently prime immune responses to viral, bacterial, parasitic, and tumor antigens in informative, preclinical animal models (including e.g. mice, rats, woodchuck, pig, rabbit, sheep, cattle, dogs, monkeys, chimps, fish, and chicken). Some of these vaccines demonstrated protective value, that is, either induced specific resistance to pathogens, or primed immunity that specifically rejected transplantable tumors. DNA vaccination contributed to the rational design of adjuvants by defining factors that modulate the immunogenicity and/or the polarization of an immune response (i.e. the repertoire of immune effector functions specifically activated).

Though similar to vaccines based on recombinant viruses, genetic vaccination has distinct advantages. The antigen load delivered by a genetic vaccine is restricted and readily controlled. DNA or RNA vaccines do not replicate, thus avoiding the complications of uncontrolled infections in immunodeficiency conditions. Many viral proteins selectively and efficiently interfere with distinct steps of antigen-processing pathways that generate and/or present antigenic peptides. They thereby inhibit immune responses. This does not operate in genetic vaccines, which can focus the response to selected determinants and exclude suppressive interfering determinants or molecules. This stresses the value of genetic vaccination as an attractive candidate for prophylactic or therapeutic immunization against intracellular and extracellular (viral, bacterial, or parasitic) pathogens and cancer.

2 Techniques

2.1 Basic Structure of a Eukaryotic Expression Vector

A "DNA vaccine" is a nonreplicative, closed circular, double-stranded plasmid DNA often 2 to 10 kb in size. Plasmids >10 kb in size are often ineffective as DNA vaccines. The vector contains 2 units. In the *transcription unit*, the target antigen-encoding sequence is cloned downstream from appropriate promoter/enhancer sequences that control transcription in human cells. Additional sequences regulating transcription (e.g. intron sequences, termination, and polyadenylation signals) are usually inserted into this unit. The *bacterial backbone* is required to produce the recombinant DNA in bacteria. It contains a bacterial origin of replication and a prokaryotic selection marker. These sequences are not essential constituents of DNA vaccines as minimal expression constructs with a transcription unit (as a covalently closed, linear DNA fragment) but no bacterial backbone have been successfully used as DNA vaccines. Plasmids used for DNA vaccination should be designed to minimize chances for integration into the host cell genome although it is difficult to exclude integration events.

Most genetic vaccines are DNA vectors because work with DNA is easier than work with RNA. However, successful vaccination with antigen-encoding mRNA has been reported. The half-life of antigen expression, the persistence of the vaccine (and the related safety issues), and the intrinsic adjuvant effect differ between RNA- and DNA-based vaccines.

2.2 Increasing the Antigenic Information Delivered by DNA Vaccines

A key objective in vaccinology is the development of multivalent vaccines. This facilitates acceptance and logistics of the vaccine, and decreases its cost. It is furthermore desirable to include different antigens of the same pathogen to increase the chance to prime protective immunity in an outbred population. DNA vaccine technology offers different options to achieve this goal. The technology comprises expression strategies that range from the expression of "minigenes" encoding only a single epitope (at the one extreme), to the coexpression of different, large antigens at stoichiometrically defined ratios (at the other extreme).

Minigenes inserted into DNA vaccines focus the immune response to only a single epitope. This allows selective priming of a monospecific response, excluding interferences from coprimed and potentially suppressive (regulatory or immunodominant) responses. Although an interesting tool to study the requirements for the induction and regulation of monospecific immune responses, the approach has limited practical value. Combining different minigenes into a single sequence leads to polytope vaccines. Different short sequences encoding only a single epitope inserted into a DNA vaccine like "pearls on a string" are coexpressed by these vaccines that have been shown to successfully prime multispecific responses in inbred animal species. By fusing sequences encoding large antigens (or antigenic domains), DNA vaccines that produce chimeric antigens can be generated. Such constructs specifically stimulate multispecific T-cell responses. Coexpression of different antigens by either polycistronic expression constructs or bidirectional promoters, or a combination of both, generates complex expression vectors in which up to four complete antigens can be coexpressed at stoichiometrically defined ratios in immunogenic form. The large size of such vectors limits their usefulness as DNA vaccines. Expression of different antigens from alternative reading frames of the same nucleotide sequence is another way to increase the antigenic information delivered by a DNA vaccine of limited size. Coinjection of mixtures of DNA plasmid vectors encoding different antigens can prime multispecific antibody and T-cell responses. Extreme cases of mixing different expression plasmids are the immunization with mixtures of plasmids containing either large numbers of transcripts generated in expression libraries, or genomic fragments representing the complete genome of a pathogen. These studies showed that as few as one plasmid in 10^5 can give rise to an immune response. Rapidly evolving strategies for expressing antigens in DNA vaccines thus cover a broad range, from a strictly monospecific T-cell response, to a single epitope (present in a minigene), to multispecific T- and B-cell responses, to epitopes on independent antigens coexpressed by complex expression constructs.

2.3 Strategies to Enhance the Immunogenicity of DNA Vaccines

Immune responses to DNA vaccines alone have been weak, especially in humans. To be immunogenic (i.e. to prime a specific immune response), an antigen has to be produced by a DNA vaccine *in situ* in sufficient quantity with "costimulatory signals" mostly provided by antigen-presenting cells (APC). DNA vaccines can be designed to optimize antigen expression (through the choice of promoter sequences and optimization of codon usage) in the relevant tissues (secondary lymphoid tissues, i.e. lymph nodes) by appropriate APCs (macrophages, B cells, or dendritic cells). "Costimulatory signals" are ligands that activate receptors expressed by cells of the innate immune system or APC. Some of the molecularly defined "costimulatory factors" are of microbial origin while others are produced early in infection as "danger signals" by the host. Codelivery of such factors is often essential to prime a specific response, but can sometimes be bypassed by copriming responses in different compartments of the specific immune system (e.g. eliciting $CD4^+$ T-cell and $CD8^+$ T-cell responses to different epitopes of the same antigen). Here, the specific interaction of APCs with two different T-cell subsets allows initiation of both T-cell responses through mutual "help."

2.3.1 Synthetic Genes

Recently, synthetic genes have been designed with optimized codon usage that can strikingly enhance translation efficiency of antigens from some vector constructs. This technology is particularly interesting for expressing bacterial antigens in eukaryotic cells.

2.3.2 Promoter/Enhancer Sequences

To support optimal translation of the protein antigen from a DNA vaccine, the antigen-encoding sequence in the transcription unit has to be inserted downstream from strong promoter/enhancer sequences. Different promoter sequences can be used to construct effective DNA vaccines. Viral sequences (e.g. from cytomegalo, papova, or retroviruses), especially those from the immediate early region of the human cytomegalovirus (HCMV), have been favored as they drive high-level production of proteins in many cell types. Control sequences from other viruses (e.g. retrovirus or papova virus) or bacteria (e.g. borrelia) have been demonstrated to support high-level antigen expression from DNA vaccines that efficiently prime humoral and cellular immune responses. Alternatives to this viral control element have been developed because interferons (released early in an immune response) suppress many viral promoters. Mammalian cell–derived promoter sequences (controlling e.g. desmin, elongation factor-1α or metallothionein expression) have therefore been used to drive antigen production in DNA vaccine. Attempts to selectively express antigen from DNA vaccines in APCs (e.g. dendritic cells) through tissue-specific promoter/enhancer sequences have been less successful, presumably because tissue specificity of expression requires large, discontinuous sequence elements that are not yet correctly defined and are difficult to accommodate into a small DNA vaccine plasmid.

2.3.3 Enhancing Transfection Efficiency of DNA Vaccines

The DNA vaccine has to be delivered to the cytosol of the cell. The most critical process in transfection is the transport

of DNA from the cytosol to the nucleus. Tagging DNA vector constructs with viral sequences, or coupling DNA to nuclear localization signal peptides can facilitate the delivery of vector DNA to nuclei where transcription can take place.

2.3.4 Subcellular Targeting of Antigen Expression

To facilitate B-cell priming (antibody responses), antigen is usually expressed as a membrane-anchored or secreted protein because antibodies recognize cell surface or extracellular antigens in their native conformation. Antigens (or its fragments) should be expressed as intracellular proteins to facilitate T-cell priming because the T-cell receptor for antigen recognizes peptides generated by partial proteolytic degradation of antigens within cells. Antigen produced in the cytosol is degraded by proteasomes into peptides that are translocated into the endoplasmic reticulum by specialized transported molecules, where they are loaded to MHC class I molecules before trafficking to the cell surface. This is the prerequisite for priming $CD8^+$ T-cell responses. Priming $CD4^+$ T-cell responses is facilitated if antigens (or their antigenic fragments or epitopes) are targeted to a vesicular pathway. This allows their degradation in a specialized endolysosomal compartment. The generated peptides are "loaded" to MHC class II molecules that are transported to the cell surface. This is a prerequisite for priming $CD4^+$ T-cell responses. Subcellular targeting of antigen traffic by appropriate engineering of the DNA vaccine can thus be used to prime selected compartments of the specific immune system.

2.3.5 Targeting Antigens to APCs

Uptake of secreted antigen produced from a DNA vaccine can be targeted to professional APCs to enhance its immunogenicity. This can be achieved by fusing the antigen-encoding sequence to a ligand of a receptor expressed (preferentially or exclusively) by APCs. Ligands binding CD58, CD40, or CD80/CD86 have been used for this purpose. Such strategies also facilitate uptake of antigen by APC through receptor-mediated endocytosis, an efficient way to deliver antigen directly to some antigen-processing compartments. Similarly, antigen released at the site of DNA vaccine inoculation can be targeted to lymph nodes (the site of priming of immune responses) by fusion with CD62L (L-selectin) that binds to a receptor selectively expressed by high endothelia of venules of these tissues.

2.3.6 Codelivery of Antigen with Immune-stimulating/modulating Cytokines and Chemokines

Short-range humoral mediators regulate most steps of the immune response, that is, specific activation (priming), clonal expansion, effector cell differentiation, and memory cell generation. The factors exert their function *in situ* in a paracrine or autocrine manner at the right site (often in synapse-like contact areas between cells) at the right time (often as a short pulse at a critical stage of the process) in the right dose (often at very low concentrations). It is difficult to develop a DNA vaccine that codelivers a cytokine or chemokine signal together with the antigen and can fulfill all these requirements. Nevertheless, many cytokines have been reported to enhance priming of humoral and/or cellular immune responses when codelivered with DNA vaccines. In addition to enhancing the immunogenicity of an antigen, codelivered cytokines modulate the spectrum of immune effector functions that are specifically stimulated.

Cytokines reported to enhance and modulate DNA vaccine-induced immune responses include GM-CSF, IL-2, IL-15, IL-6, interferon-α, IL-10, or IL-12/IL-23. Expression plasmids encoding chemokines act as strong adjuvant for eliciting Th1-derived immunity when inoculated together with DNA vaccines. In these experimental settings, cytokines were codelivered by DNA vaccines as fusion constructs with the antigen, as plasmid mixtures, or as polycistronic constructs. Delivery of cytokines or chemokines (that are autoantigens) by DNA vaccination can prime autoreactive immune reactivities that neutralize the bioactivity of these factors. Repeated administration of DNA vaccines containing chemokine-encoding genes inhibited the development and progression of arthritis through the development of chemokine-neutralizing autoantibodies, even when administered after the onset of disease. However, the appearance of autoantibodies binding cytokines or chemokines is also a regular phenomenon in many chronic infections. It is unexpected that such a wide variety of cytokines with very different modes of action enhance priming of specific immunity by DNA vaccines. It is uncertain whether the reported findings have general validity for vaccination protocols in species other than the mouse. However, they certainly provided information on the molecular basis of "adjuvant effects" that until recently were largely empirically defined.

2.3.7 Codelivery of Transcription Factors

Proteins produced from DNA vaccines by transfected cells can include transcription factors or antiapoptotic molecules. An example is the codelivery of interferon regulatory factors (IRFs). Using an experimental DNA vaccination model, it was shown that IRF-3 had good activity for T cells, IRF-7 had good activity for both antibody and T cells, and IRF-1 had good activity for antibody. This opens a new field of manipulating intracellular signaling pathways to induce, enhance, or modulate inflammatory reactions to create a supportive milieu for the initiation of an adaptive immune response. A further option is to extend the life span of antigen-bearing APCs by codelivering antiapoptotic signals that prolong the period they can specifically stimulate T-cell responses. These strategies target the differentiation and the functional life span of APC, the key regulators at the interface between the innate and adaptive immune system that regulate priming, magnitude, and polarization of an immune response.

2.3.8 CpG-containing Oligodeoxynucleotides(ODN)

Motifs in DNA sequences containing cytosine/guanine (CpG) nucleotides flanked by 2 to 3 other nucleotides exert potent immune-stimulating signals through binding to the toll-like receptor (TLR) 9. Such motifs are present in many plasmid DNAs used for genetic immunization. It has been proposed that they provide a critical contribution to the immunogenicity of these constructs. This is not supported by the observation that TLR9-deficient mice are perfectly able to respond to DNA vaccines. The role of CpG-containing motifs in the immunogenicity of DNA vaccines is thus unresolved.

2.3.9 Stress Proteins

Heat shock or stress proteins (hsp) have been incorporated into DNA-, protein-, or peptide-based vaccines as either antigens, chaperones, or adjuvant. T- cell reactivities to many bacterial and protozoan

pathogens (e.g. mycobacteria, chlamydia, yersinia, malaria plasmodia, or leishmania) are directed against hsp epitopes. An immunogenicity-potentiating role of hsp is apparent in DNA vaccines in which antigen and hsp molecules are linked, either by fusing sequence encoding antigens to hsp-encoding sequences, or by fusing hsp-capturing domains to antigen-encoding sequences. This strikingly enhances the potency of the vaccine to elicit, for example, tumor-rejecting T-cell responses indicating a potent adjuvant effect of hsp molecules.

2.3.10 Copriming Specific "Help"

DNA vaccines can deliver large protein antigens that either contain naturally or are engineered to contain epitopes that stimulate $CD4^+$ T-cell "help." This specific T-cell reactivity can "help" prime either B cells (enhancing antibody responses), or $CD8^+$ T cells (cell-mediated cytolytic and/or cytokine response). This has been used to increase the potency of DNA vaccines for priming antibody or cytotoxic T-cell responses.

2.4 Production and Quality Control of DNA Vaccines

Traditional production methods used sophisticated methodology to separate DNA from contaminating organic components. In contrast, new processes for manufacturing plasmid DNA pharmaceuticals have been significantly improved and are shorter but also fulfill established GMP (*good manufacturing practices*) guidelines. The core elements used in manufacturing and quality control are nearly identical (generic) for plasmids <10 kb in size, but require modifications for the manufacturing and formulation process of larger plasmids. The first step in plasmid production is the stable transformation of a defined *Escherichia coli* host strain with the plasmid DNA vector, followed by cultivation in a bioreactor, alkaline lysis of the bacterial biomass, followed by several filtration and chromatography steps to obtain DNA pharmaceuticals of the required quality. The most important requirements for the plasmid purification process are:

1. separation of plasmid DNA from substances derived from the DNA isolation process or the bacterial host cell;
2. purification of the "supercoiled" (ccc) form from other plasmid DNA forms to obtain a homogeneous product; and
3. protection of the ccc form throughout the manufacturing, formulation, storage, and delivery process.

2.4.1 Quality Control and Safety in Plasmid DNA Vaccine Production

Plasmid DNA quality depends on manufacturing, storage, and application. Its safety depends on vector construction, characterization, and testing by toxicology before clinical trials. These parameters are well defined but subjected to ongoing improvements regarding analytical techniques. Table 2 shows relevant quality control tests for in-process-control (IPC) and product release. A safe, state-of-the-art manufactured, well-controlled product is usually requested, the relevant guidelines for the specification for clinical material are currently subject to continuous improvements.

2.4.2 Plasmid Identity

Plasmid DNA is tested for identity by digestion with a panel of restriction enzymes followed by agarose gel electrophoresis (AGE) to determine the length of the digestion fragments obtained. These have to

Tab. 2 Important physical and microbial specifications for quality assurance and quality control of plasmid DNA vaccines.

	Test	*Analytical method*
1.	DNA concentration	UV Absorption (260 nm)
2.	General purity	UV Scan (220–320 nm)
3.	Homogeneity (ccc content)	CGE (*capillary gel electrophoresis*)
4.	Purity (visible)	Visual inspection
5.	Purity (genomic DNA)	Agarose gel (visual); Southern blot, Quantitative PCR
6.	Purity (RNA)	Agarose gel (visual); fluorescence assay, quantitative PCR
7.	Purity (protein)	BCA test
8.	Purity (LPS)	LAL test
9.	Purity (microorganisms)	Bioburden test; sterility test
10.	Identity (vector structure)	Restriction fragment lengths conforms to reference in AGE (3–4 enzymes)
11.	Identity (sequence)	Sequencing (double strand)

Notes: AGE: agarose gel electrophoresis; BCA: Bicinchoninic acid; ccc: covalently closed circular or supercoiled; CGE: capillary gel electrophoresis; LAL: *Limulus* amebocyte lysate; LPS: lipopolysaccharide.
Source: Schleef, M., Schmidt, T. (2004) Animal-free production of ccc-supercoiled plasmids for research and clinical applications, *J. Gene Med.* **6**, S45–S53.

conform to the calculated lengths of the fragments. The nucleotide sequence of the plasmid is determined by sequencing the plasmid DNA.

2.4.3 Animal-free DNA Production

Because RNA from *E. coli* is the major contaminant of the plasmid DNA produced, a typical and critical process in its purification is usually the treatment with ribonucleases (e.g. RNase A from bovine pancreas) to destroy RNA molecules. Current pharmaceutical DNA manufacturing technology avoids bovine RNase treatment to increase the quality of the product, even in large-scale production of GMP-grade DNA. To ensure product purity and safety, no enzymes should be used in the purification and cultivation processes that thus should be carried out without animal-derived substances.

2.4.4 Genomic DNA

A major impurity in plasmid DNA preparations is chromosomal and genomic DNA (gDNA) from plasmid-producing bacterial cells. The most critical step is alkaline lysis, in which gDNA as well as plasmid DNA are denatured by alkaline pH shift. As chromosomal DNA is extremely sensitive to shearing, this can result in its fragmentation. Some gDNA fragments can migrate as distinct bands in AGE while smaller fragments can be detected as undefined smears in overloaded AGE analyses. Sensitive assays, for example, Southern blot hybridization or quantitative PCR, are often used to quantify gDNA contamination. While contaminations of gDNA in plasmid DNA preparations were often in the range of 5 to 10% in the past, novel purification technologies allowed a significant reduction.

DNA species must be completely removed from the manufacturing environments at product change. This aspect of changeover is important if different DNAs are produced within the same facility.

2.4.5 Host Cell Impurities

Host cell–derived contaminants (e.g. proteins, RNA, lipopolysaccharides) should be reduced to minimal concentrations in the plasmid DNA purification process. Proteins can be detected by colorimetric assays. Quantification of residual RNA is performed by fluorescence assays after digestion of plasmid DNA with DNase. Lipopolysaccharides can be detected in the *Limulus polyphemus* lysate assay. A striking reduction of these impurities has to be achieved before the manufactured DNA can be used in research or clinics.

2.4.6 Importance of Supercoiled Forms

Structural homogeneity of plasmid DNA is usually determined by AGE. Different forms of plasmid DNA with an identical nucleotide sequence are present within a given DNA preparation from *E. coli* cells and can be assigned to different AGE bands. However, the assignment of bands to different plasmid topologies is complicated because the electrophoretic mobility of plasmids of different shapes changes with the electrophoretic operating conditions. In addition, the quantification of DNA plasmid forms based on signal intensity of stained bands in AGE may not be reliable due to nonlinear responses. It is well known that typically only one band (the ccc form) is observed when only a small amount of plasmid DNA is applied to AGE while two prominent bands (ccc and a slower migrating, nicked, open (oc) form), are usually revealed by standard AGE. Capillary gel electrophoresis (CGE) allows the identification and more reliable quantification of all forms of plasmid DNA. CGE is performed using thin (diameter 100 μm), coated capillaries with a length of 40 to 60 cm, filled with a liquid polymer, for example, a solution of hydroxypropylmethylcellulose. Electrophoretic separation takes place by applying a high 5 to 30 kV voltage followed by high-resolution detection using special intercalating dyes. This allows quantitative detection of all plasmid forms by laser-induced fluorescence. This automated system offers high reproducibility, reliable quantification, and short analysis time. In contrast to AGE, quantification of plasmid forms by CGE is possible in a wide range of linearity and needs only 50-ng plasmid DNA.

The ccc plasmid topology is the most compact structure and is expected to be the most active topology. Other plasmid topologies also appear often and have to be eliminated from the product. If one strand is broken (nicked), the oc form results. This is caused either naturally by processes within the plasmid-producing bacteria or during processing (enzymatic or mechanic degradation) of the biomass. Linear forms are generated if both strands are cleaved once at approximately the same position. In addition, plasmids may appear as oligomeric forms (e.g. concatemers) detectable in different topologies (Fig. 1a,b). The only intact, nondamaged form is the ccc supercoil DNA because linear forms and oc forms are randomly damaged at different gene locations and may be inactive if promoter- or antigen-encoding regions are destroyed. Any process development must therefore be subjected to IPC systems to assess the characteristics of the produced plasmid molecules. In particular, methods are required for obtaining supercoiled ccc plasmid DNA in pure form.

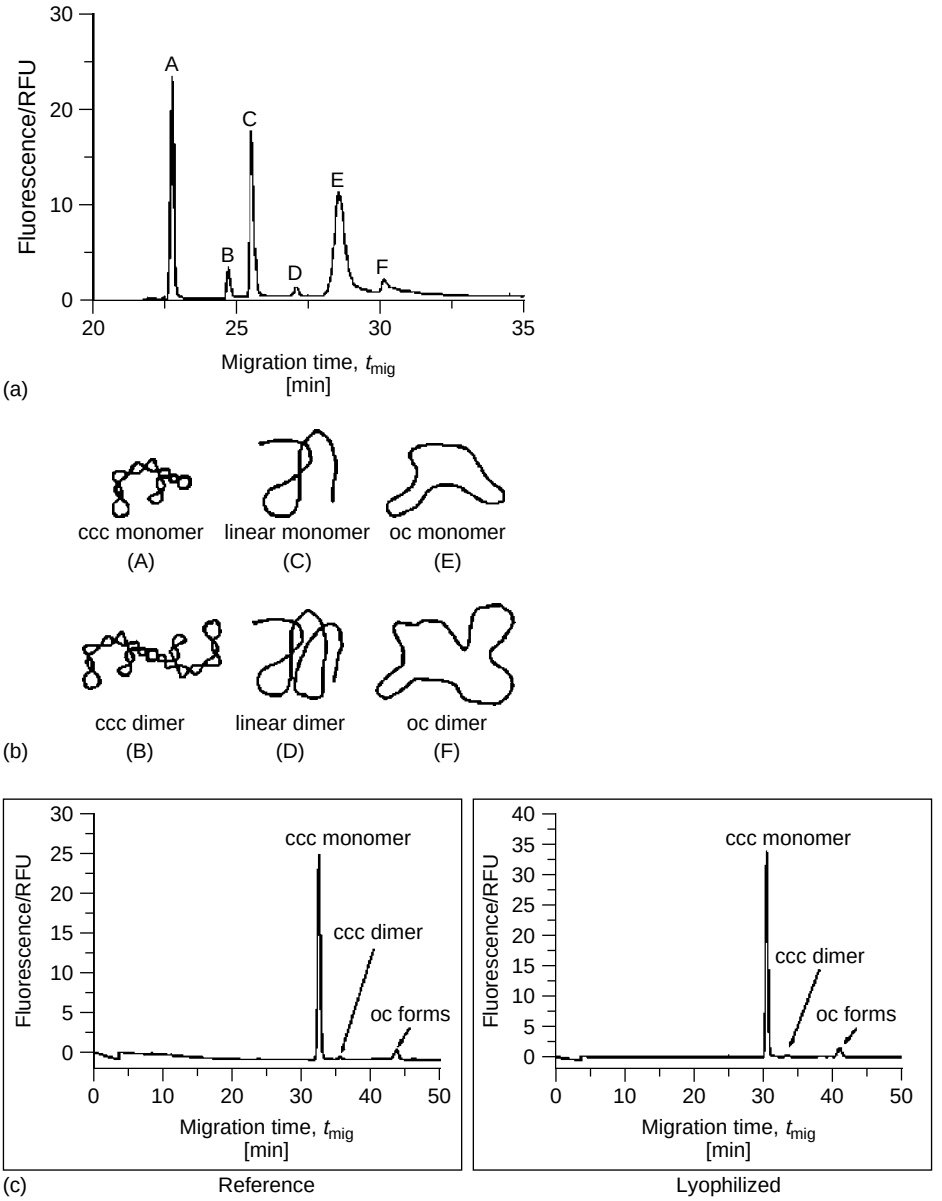

Fig. 1 (a) Electropherogram indicating individual peaks of (b) various plasmid topologies, (c) visualized in the corresponding graph (a and b from Schleef, M., Schmidt, T. (2004) Animal-free production of ccc-supercoiled plasmids for research and clinical applications, *J. Gene Med.* **6**, S45–S53). The comparison of the plasmid topology pattern in DNA before (reference) and after lyophilization indicated no difference ((c) from Schmidt, T., Voß, C., Schleef, M. (2005) From Bulk to Delivery: Plasmid Manufacturing and Storage, in: Schleef, M. (Ed.) *Plasmid Pharmaceuticals*, Wiley-VCH, Weinheim).

2.5 Stability of DNA Vaccines

Manufacturing, formulation, storage, and application of DNA vaccines should be easy and cost efficient. Physical and chemical stability of plasmid DNA is required for DNA-based pharmaceuticals that can be stored, shipped, and applied even under critical environmental conditions. DNA delivery sometimes requires protection of the active pharmaceutical ingredient, which is an issue of DNA formulation.

2.5.1 Storage Stability of Plasmid DNA

Recent studies have shown alterations of plasmid DNA stored at different temperatures (e.g. −80, −20, or +4 °C) over 13 to 30 months. Plasmid DNA stored at −20 or −80 °C shows no degradation of the intact ccc into the damaged oc form. In contrast, ccc plasmid DNA stored at +4 °C is degraded into damaged oc and linear forms within months. Hence, storage conditions are critical to maintain the bioactivity of plasmid DNA pharmaceuticals. An attractive option in plasmid DNA storage and shipment is lyophilization, a technology often applied to other vaccines or pharmaceuticals. Although lyophilization of plasmid DNA has been accompanied by damage to its ccc supercoil topology and insufficient reconstitution, recent improvements in lyophilization could be shown not to damage ccc supercoil plasmid DNA (Fig. 1c).

2.5.2 Stability in Plasmid Vector Applications

Very little is known about the *in vivo* fate of a formulated DNA vaccine after its injection. Complexing DNA with cationic agents provides some physical protection *in vivo* but raises the question whether plasmid topology is maintained during formulation. This can be determined by CGE analysis of small samples reextracted from the complexes.

2.6 Delivery of DNA Vaccines

DNA vaccines can be delivered either as nonpackaged (“naked”) plasmid DNA, or as formulated or “packaged” plasmid DNA. A single (intramuscular, subcutaneous, or intradermal) injection of a fairly high dose (50–100 μg/mouse) of “naked” plasmid DNA into mice usually primes an immune response. Unfortunately, this is not reproducible in other species, including man. Many formulations have therefore been used to formulate plasmid DNA in an effort to facilitate *in vivo* its half-life, its transfection efficacy, and the expression of an immunogenic product from the vector. These included incorporation into cationic liposomes, virosomes or lipoplexes; binding to cationic polymers, peptides or proteins; mixing with aluminum phosphate; *in vivo* electroporation; or treatment with chloroquine (to reduce plasmid DNA degradation in an acid endolysosomal compartment after uptake). Plasmid DNA vaccines can be introduced into invasive bacterial cells (from strains that are used for vaccination). Following infection after oral vaccination, the DNA vaccine is directly delivered to the cell. Another promising delivery technique is a biolistic system in which DNA-coated gold microprojectiles are directly propelled into cells in the living animal (“gene gun”). This technique has been successful in genetic vaccination of many animal species, and has been employed in clinical trials. The plethora of alternatives in this rapidly evolving field of delivery techniques is an indicator that the issue is not yet resolved

in a practical way. This is a major drawback in introducing this technique into clinical practice.

2.6.1 Combining DNA Vaccines and Protein Vaccines

Traditional protein antigen vaccines and DNA vaccines can be combined. Protein antigens and DNA vaccines emulsified together in aluminum phosphate can prime multivalent immune responses to both components of the vaccine. DNA vaccines, delivered as a complex with antigenic, cationic peptides, or proteins can prime immune responses to epitopes in the DNA vaccines and in the cationic antigens. These preliminary data suggest that DNA vaccines could be a part of multicomponent vaccines.

2.6.2 Routes of DNA Vaccine Delivery

Similar to the many formulation protocols, many different routes of plasmid DNA delivery have been described. These include intramuscular, subcutaneous, or intradermal injection of "naked" or formulated plasmid DNA, or intradermal and mucosal delivery of plasmid DNA by the gene gun. Finding a reliable, easy, safe, and accepted route for delivering DNA vaccines in clinical practice is a high priority in the field.

3 Applications

3.1 Priming Immune Responses by DNA Vaccination

DNA vaccines effectively prime and boost immune responses, which is evident by the generation of specific (B and/or T cell) reactivity and/or the specific induction of protective memory against pathogens. An immune response primed by a conventional (e.g. protein-based) vaccine can be boosted by DNA vaccination. A response primed by DNA vaccination can be boosted by alternative vaccine formulations of the relevant antigen (e.g. a protein antigen or a recombinant virus), a protocol apparently eliciting superior protective immunity to some clinically important pathogens. Most experimental studies have been performed on the mouse, in which immune responses can be readily primed by different DNA vaccination protocols and in which many specific readout assays are available. In contrast, DNA vaccination had mixed successes in species other than the mouse, including man.

3.1.1 Priming T-cell Responses

In contrast to most alternative vaccines, DNA vaccines are particularly potent in priming MHC class I-restricted CD8 T-cell responses and MHC class II-restricted CD4 T-cell responses. Priming CD8 T-cell responses by genetic vaccination can be indirect or direct. Antigens produced from a DNA vaccine by transiently transfected cells (myocytes, keratinocytes) seem to be transferred to professional APC that present their epitopes in an immunogenic way. This indirect way of antigen expression, processing, and presentation is called *cross-priming*. In contrast, resident dendritic cells (DC) in the targeted tissue may be directly transfected by a DNA vaccine and process the antigen it produces for immunogenic presentation. Direct priming may operate after intradermal delivery of low amounts of plasmid DNA with the gene gun. It is complicated by the finding that resident DC usually are immature and tend to elicit tolerogenic or regulatory rather than effector/memory T-cell responses. Engineering antigenic

constructs in DNA vaccines to target APC in lymph nodes, codelivery of factors that enhance their immunogenicity, and an increase of antigenic information they carry offer attractive prospects for optimizing potent T cell–stimulating vaccines. DNA vaccination could prime T-cell responses to nonstructural viral antigens that support cross-strain protection, and could override nonresponder status. The exceptional potency of DNA vaccines to elicit specific T-cell responses has generated the hope that it might contribute to solve some major current health issues, such as the development of vaccines to be used against, for example, HIV, mycobacteria, papillomaviruses, or cancer.

3.1.2 Priming B-cell (Antibody) Responses

DNA vaccines expressing secreted antigen readily prime antibody responses. In mice, injection of a high dose of a DNA vaccine induces high titers of specific IgG2a antibodies (indicating Th1 priming), while injection of a low dose of the same DNA vaccine induces preferentially specific IgG1 antibodies (indicating Th2 priming). Antibody responses have been elicited after DNA vaccination to subdominant (cryptic) epitopes of antigen fragments that were difficult to detect by conventional vaccination strategies. *In situ* production of native antigen with all post-translational modifications allows priming of neutralizing antibodies to conformational epitopes that are usually CD4 T-cell "help"-dependent.

3.2 DNA Vaccination as an Experimental Tool

Expression library immunization is a DNA vaccine–based tool to identify vaccine candidates out of a large number of gene products expressed by a pathogen. Although its usefulness has only been demonstrated in a few systems, it may gain in interest as more genomes of pathogens are sequenced. Owing to the ease of deleting or recombining fragments of open reading frames, DNA vaccines are furthermore useful in identifying antigenic domains or epitopes of complex antigens. Once identified, these antigenic determinants can be combined into polyvalent, chimeric vaccines that contain immunodominant domains from different antigens and/or different pathogens. The selective incorporation of an antigenic domain or epitope into a DNA vaccine has been used to generate monospecific immunological probes, either monoclonal antibodies, or T-cell clones.

3.3 Potential Advantages of DNA Vaccination

Genetic vaccination can provide long-lived immunity, and multivalent vaccines can be given by a single inoculation. As DNA vaccines to different pathogens are produced using similar or identical generic production techniques, the development of formulation, delivery, and safety issues are greatly simplified. A particular feature of DNA vaccination is its exceptional potency to prime T-cell responses. DNA vaccines have been shown to override low responder status in preclinical animal models to subdominant T-cell epitopes. It can prime T cells to nonstructural virus proteins, thereby extending the repertoire of vaccine candidates that can be used to control virus infections. Priming T-cell immunity to (nonvariant) internal viral antigens can establish cross-strain protection against pathogenic viruses, a goal difficult to obtain with conventional vaccines that rely on the induction of

neutralizing antibody responses against variant envelope proteins. Newborn mice immunized with DNA vaccines develop balanced Th1/Th2 immune responses. DNA vaccination is thus an attractive option to prime antiviral Th1 immunity in neonates or young children, overcoming the bias for specific tolerance induction of the neonatal immune system. Hence, this type of vaccine may have advantages for protection early in life, an objective difficult to achieve with most conventional vaccines currently available.

4 Perspectives

4.1 Achievement and Unresolved Issues of the Technology

The technology introduced only a decade ago has made striking progress in vector design, large-scale DNA production, and optimizing the antigenic information and the immunogenicity of antigens, their domains or selected epitopes. Major unresolved issues are the efficient delivery of the vaccine in man and the critical assessment of safety issues. While the former needs an easy, safe, and reliable technique soon, the latter will only be resolved with clinical experience accumulating over extended time spans.

4.2 Preclinical and Clinical Trials that Test Potential Applications of DNA Vaccination in Clinical Practice

In informative preclinical animal models of *infectious diseases,* DNA vaccination proved effective as a *prophylactic* vaccine against a broad spectrum of pathogens including, for example; influenza, malaria plasmodia, leishmania, mycobacteria, simian immunodeficiency virus, papilloma virus, and herpes viruses. Even more challenging is the development of therapeutic vaccines against HIV, hepatitis B and C virus, and malaria plasmodia or mycobacteria. DNA vaccine-based specific immunotherapy protocols against these and other pathogens that cause chronic infection are under investigation in various informative animal species.

The specific *immunotherapy of cancer* relies on induction of efficient cellular immunity to tumor-associated antigens by therapeutic vaccination. Because it efficiently primes T cells, DNA vaccination has played a prominent role in designing treatment schemes for the treatment of growing tumors in preclinical models.

DNA vaccination primes potent Th1 immunity. An established Th2 bias of an immune response can be specifically converted into a Th1 phenotype by boosting with the relevant antigen in conjunction with a strong Th1-driving stimulus. DNA vaccination has been successfully used in a preclinical model to specifically suppress established *allergy* by converting the pathogenic Th2 response into a harmless Th1 response.

These are some of the many potential areas of interest where DNA vaccination may find clinical applications. Clinical trials using either DNA vaccination alone, or DNA vaccine priming and boosting with recombinant viruses have been initiated. These include vaccination of uninfected volunteers to assess safety and immunogenicity, and therapeutic vaccination of chronically infected or cancer-bearing patients to assess safety and therapeutic efficacy. An infectious disease targeted with high priority is AIDS caused by the human

immunodeficiency virus (HIV). DNA vaccines containing >20 CD8 T cell–defined epitopes (polytope vaccines) of HIV, its envelope (gp160, gp40, gp120) protein or its antigenic fragments, its nucleocapsid gag (p17, p24), a fusion protein containing gag and polymerase, or its small regulator proteins (tat, nef) from different HIV clades have been tested in Phase I clinical trials using different formulations and different constructs, sometimes with selected cytokine adjuvants (e.g. IL-2). In addition, some clinical trials have been initiated to assess the efficacy of DNA-based vaccination for the specific control of herpes viruses; human papilloma virus, Ebola virus; *Plasmodium falciparum* (causing malaria), or mycobacteria. The therapeutic value of DNA vaccines is currently tested in melanomas, renal and prostate cancers, hepatocellular carcinoma, chronic lymphatic leukemia, and colon cancer. It is expected that these studies will generate a wealth of information on the safety, immunogenicity, and therapeutic efficacy of DNA vaccination strategies.

4.3 Risks and Safety Issues Raised by DNA Vaccination

Only limited experience on the long-term safety of DNA vaccination has been derived from preclinical animal models and their validity for the critical assessment of the risks under clinical conditions is uncertain. The following potential risks of DNA vaccination have been discussed.

Integration of plasmid DNA into coding or regulatory sequences of cellular DNA could lead to *insertional mutagenesis*. It is conceivable that insertion into the genome of the germline could occur as a rare event. The integrated "foreign" DNA could contribute to tumor development by activating oncogenes, inactivating tumor suppressor genes, or inducing chromosomal instability through the induction of chromosomal breaks or rearrangements. DNA vaccines usually produce only low amounts of antigen that may induce low zone tolerance rather than immunity. This possibility has not been verified experimentally up to now. In contrast, it has been shown that DNA vaccine-primed responses that were undetectable post-priming could be readily boosted. Some DNA vaccines were exceptionally potent raising the concern that they could induce autoimmunity specific for either codelivered, autologous costimulator molecules (cytokines, chemokines), or for the transfected cell. As autoantibody responses against cytokines regularly occur during tissue damage in the course of chronic infections, it is uncertain whether this is a complication typical for DNA vaccination. Anti-DNA antibodies have been only rarely induced by DNA vaccines in experimental animal models. As the *in vivo* transfection of cells by injected plasmid DNA of the vaccine is not controlled, and destructive Th1 immune responses are elicited, immune-mediated tissue lesions could occur. Whether this is an advantageous (deleting genetically altered cells) or deleterious event is uncertain. Antibiotic resistance genes are usually transferred with the DNA vaccine, but the stable induction of this resistance in vaccinated individuals is an unlikely event because resistance genes are expressed under prokaryotic promoter control in the bacterial backbone, and not expected to be expressed at high levels for extended periods of time by human cells. Furthermore, most of the antibiotics used for selection are not of clinical importance.

Acknowledgment

We greatly appreciate the helpful comments of Drs. M.L. Michel (Institut Pasteur, Paris, France) and D.M. Klinman (FDA, Bethesda, USA).

See also Cytomegalovirus and Varicella–zoster Virus Vaccines; Genetics, Molecular Basis of; Medicinal Chemistry; Nucleic Acid Packaging of DNA Viruses; Nucleic Acid Packaging of RNA Viruses; Oncology, Molecular; Retrovirus Integration, Biological and Biochemical Machineries for; Somatic Gene Therapy; Vaccine Approaches for Protection Against HIV; Vaccine Research and Development Methodology; Vector System: Plasmid DNA.

Bibliography

Books and Reviews

Barry, M.A., Johnston, S.A. (1997) Biological features of genetic immunization, *Vaccine* **15**, 788–791.

Donnelly, J.J., Ulmer, J.B., Liu, M.A. (1994) Immunization with DNA, *J. Immunol. Methods* **176**, 145–152.

Donnelly, J.J., Ulmer, J.B., Shiver, J.W., Liu, M.A. (1997) DNA vaccines, *Annu. Rev. Immunol.* **15**, 617–648.

Ertl, H.C., Xiang, Z. (1996) Novel vaccine approaches, *J. Immunol.* **156**, 3579–3582.

Gurunathan, S., Klinman, D.M., Seder, R.A. (2000) DNA vaccines: immunology, application and optimization, *Annu. Rev. Immunol.* **18**, 927–974.

Johnston, S.A., Tang, D.C. (1994) Gene gun transfection of animal cells and genetic immunization, *Methods Cell Biol.* **43**, 353–365.

Krieg, A.M. (2002) CpG motifs in bacterial DNA and their immune effects, *Annu. Rev. Immunol.* **20**, 709–760.

Mor, G., Singla, M., Steinberg, A.D., Hoffman, S.L., Okuda, K., Klinman, D.M. (1997) Do DNA vaccines induce autoimmune disease? *Hum. Gene Ther.* **8**, 293–300.

Roman, M., Spiegelberg, H.L., Broide, D., Raz, E. (1997) Gene immunization for allergic disorders, *Springer Semin. Immunopathol.* **19**, 223–232.

Schleef, M. (1999) Issues of Large Scale Plasmid DNA Manufacturing, in: Rehm, H.-J., Reed, G., Puehler, A., Stadler, P. (Eds.) *Recombinant Proteins, Monoclonal Antibodies and Therapeutic Genes*, Wiley-VCH, Weinheim, pp. 443–469.

Schleef, M., Schmidt, T. (2004) Animal-free production of ccc-supercoiled plasmids for research and clinical applications, *J. Gene Med.* **6**, S45–S53.

Schmidt, T., Friehs, K., Flaschel, E. (2001) Structures of Plasmid DNA, in: Schleef, M. (Ed.) *Plasmids for Therapy and Vaccination*, Wiley-VCH, Weinheim, pp. 29–42.

Schmidt, T., Voß, C., Schleef, M. (2005) From Bulk to Delivery: Plasmid Manufacturing and Storage, in: Schleef, M. (Ed.) *Plasmid Pharmaceuticals*, Wiley-VCH, Weinheim.

Siegrist, C.-A. (1997) Potential advantages and risks of nucleic acid vaccines for infant immunization, *Vaccine* **15**, 798–800.

Stevenson, F.K. (2004) DNA vaccines and adjuvants, *Immunol. Rev.* **199**, 5–278.

Primary Literature

Amara, R.R., Villinger, F., Altman, J.D., Lydy, S.L., O'Neil, S.P., Staprans, S.I., Montefiori, D.C., Xu, Y., Herndon, J.G., Wyatt, L.S., Candido, M.A., Kozyr, N.L., Earl, P.L., Smith, J.M., Ma, H.L., Grimm, B.D., Hulsey, M.L., Miller, J., McClure, H.M., McNicholl, J.M., Moss, B., Robinson, H.L. (2001) Control of a mucosal challenge and prevention of AIDS by a multiprotein DNA/MVA vaccine, *Science* **292**, 69–74.

Arichi, T., Saito, T., Major, M.E., Belyakov, I.M., Shirai, M., Engelhard, V.H., Feinstone, S.M., Berzofsky, J.A. (2000) Prophylactic DNA vaccine for hepatitis C virus (HCV)

infection: HCV-specific cytotoxic T lymphocyte induction and protection from HCV-recombinant vaccinia infection in an HLA-A2.1 transgenic mouse model, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 297–302.

Barouch, D.H., Craiu, A., Kuroda, M.J., Schmitz, J.E., Zheng, X.X., Santra, S., Frost, J.D., Krivulka, G.R., Lifton, M.A., Crabbs, C.L., Heidecker, G., Perry, H.C., Davies, M.E., Xie, H., Nickerson, C.E., Steenbeke, T.D., Lord, C.I., Montefiori, D.C., Strom, T.B., Shiver, J.W., Lewis, M.G., Letvin, N.L. (2000) Augmentation of immune responses to HIV-1 and simian immunodeficiency virus DNA vaccines by IL-2/Ig plasmid administration in rhesus monkeys, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4192–4197.

Boyer, J.D., Ugen, K.E., Chattergoon, M., Wang, B., Shah, A., Agadjanyan, M.G., Bagarazzi, M.L., Javadian, A., Carrano, R., Coney, L., Williams, W.V., Weiner, D.B. (1997) DNA vaccination as anti-human immunodeficiency virus immunotherapy in infected chimpanzees, *J. Infect. Dis.* **176**, 1501–1509.

Boyer, J.D., Cohen, A.D., Vogt, S., Schumann, K., Nath, B., Ahn, L., Lacy, K., Bagarazzi, M.L., Higgins, T.J., Baine, Y., Ciccarelli, R.B., Ginsberg, R.S., MacGregor, R.R., Weiner, D.B. (2000) Vaccination of seronegative volunteers with a human immunodeficiency virus type 1 env/rev DNA vaccine induces antigen-specific proliferation and lymphocyte production of β-chemokines, *J. Infect. Dis.* **181**, 476–483.

Boyle, J.S., Brady, J.L., Lew, A.M. (1998) Enhanced responses to a DNA vaccine encoding a fusion antigen that is directed to sites of immune induction, *Nature* **392**, 408–411.

Broide, D., Schwarze, J., Tighe, H., Gifford, T., Nguyen, M.D., Malek, S., van Uden, J.H., Martin, O.E., Gelfand, E.W., Raz, E. (1998) Immunostimulatory DNA sequences inhibit IL-5, eosinophilic inflammation, and airway hyperresponsiveness in mice, *J. Immunol.* **161**, 7054–7062.

Calarota, S.A., Leandersson, A.C., Bratt, G., Hinkula, J., Klinman, D.M., Weinhold, K.J., Sandstrom, E., Wahren, B. (1999) Immune responses in asymptomatic HIV-1-infected patients after HIV-DNA immunization followed by highly active antiretroviral treatment, *J. Immunol.* **163**, 2330–2338.

Casares, S., Inaba, K., Brumeanu, T.D., Steinman, R.M., Bona, C.A. (1997) Antigen presentation by dendritic cells after immunization with DNA encoding a major histocompatibility complex class II-restricted viral epitope, *J. Exp. Med.* **186**, 1481–1486.

Charo, J., Ciupitu, A.M., Le-Chevalier-De-Preville, A., Trivedi, P., Klein, G., Hinkula, J., Kiessling, R. (1999) A long-term memory obtained by genetic immunization results in full protection from a mammary adenocarcinoma expressing an EBV gene, *J. Immunol.* **163**, 5913–5919.

Coon, B., An, L.L., Whitton, J.L., von-Herrath, M.G. (1999) DNA immunization to prevent autoimmune diabetes, *J. Clin. Invest.* **104**, 189–194.

Cornell, K.A., Bouwer, H.G., Hinrichs, D.J., Barry, R.A. (1999) Genetic immunization of mice against Listeria monocytogenes using plasmid DNA encoding listeriolysin O, *J. Immunol.* **163**, 322–329.

Darji, A., Guzman, C.A., Gerstel, B., Wachholz, P., Timmis, K.N., Wehland, J., Chakraborty, T., Weiss, S. (1997) Oral somatic transgene vaccination using attenuated *S. typhimurium*, *Cell* **91**, 765–775.

Dupuis, M., Denis, M.K., Woo, C., Goldbeck, C., Selby, M.J., Chen, M., Otten, G.R., Ulmer, J.B., Donnelly, J.J., Ott, G., McDonald, D.M. (2000) Distribution of DNA vaccines determines their immunogenicity after intramuscular injection in mice, *J. Immunol.* **165**, 2850–2858.

Feltquate, D.M., Heaney, S., Webster, R.G., Robinson, H.L. (1997) Different T helper cell types and antibody isotypes generated by saline and gene gun DNA immunization, *J. Immunol.* **158**, 2278–2284.

Fennelly, G.J., Khan, S.A., Abadi, M.A., Wild, T.F., Bloom, B.R. (1999) Mucosal DNA vaccine immunization against measles with a highly attenuated *Shigella flexneri* vector, *J. Immunol.* **162**, 1603–1610.

Fissolo, N., Riedl, P., Reimann, J., Schirmbeck, R. (2004) DNA vaccines prime $CD8^+$ T cell responses to epitopes of viral antigens produced from overlapping reading frames of a single coding sequence, *Eur. J. Immunol.* **35**, 117–127.

Fournillier, A., Nakano, I., Vitvitski, L., Depla, E., Vidalin, O., Maertens, G., Trepo, C., Inchauspe, G. (1998) Modulation of immune responses to hepatitis C virus envelope E2 protein following injection of plasmid DNA

using single or combined delivery routes, *Hepatology* **28**, 237–244.

Fu, T.M., Friedman, A., Ulmer, J.B., Liu, M.A., Donnelly, J.J. (1997) Protective cellular immunity: cytotoxic T-lymphocyte responses against dominant and recessive epitopes of influenza virus nucleoprotein induced by DNA immunization, *J. Virol.* **71**, 2715–2721.

Gurunathan, S., Sacks, D.L., Brown, D.R., Reiner, S.L., CHarest, H., Glaichenhaus, N., Seder, R.A. (1997) Vaccination with DNA encoding the immunodominant LACK parasite antigen confers protective immunity to mice infected with *Leishmania major*, *J. Exp. Med.* **186**, 1137–1147.

Hassett, D.E., Zhang, J., Whitton, J.L. (1997) Neonatal DNA immunization with a plasmid encoding an internal viral protein is effective in the presence of maternal antibodies and protects against subsequent viral challenge, *J. Virol.* **71**, 7881–7888.

Iwasaki, A., Torres, C.A., Ohashi, P.S., Robinson, H.L., Barber, B.H. (1997) The dominant role of bone marrow-derived cells in CTL induction following plasmid DNA immunization at different sites, *J. Immunol.* **159**, 11–14.

Klavinskis, L.S., Barnfield, C., Gao, L., Parker, S. (1999) Intranasal immunization with plasmid DNA-lipid complexes elicits mucosal immunity in the female genital and rectal tracts, *J. Immunol.* **162**, 254–262.

Kodihalli, S., Haynes, J.R., Robinson, H.L., Webster, R.G. (1997) Cross-protection among lethal H5N2 influenza viruses induced by DNA vaccine to the hemagglutinin, *J. Virol.* **71**, 3391–3396.

Kuhröber, A., Wild, J., Pudollek, H.P., Chisari, F.V., Reimann, J. (1997) DNA vaccination with plasmids encoding the intracellular (HBcAg) or secreted (HBeAg) form of the core protein of hepatitis B virus primes T cell responses to two overlapping K^b- and K^d-restricted epitopes, *Int. Immunol.* **9**, 1203–1212.

Kuklin, N., Daheshia, M., Karem, K., Manickan, E., Rouse, B.T. (1997) Induction of mucosal immunity against herpes simplex virus by plasmid DNA immunization, *J. Virol.* **71**, 3138–3145.

Kwissa, M., Lindblad, E.B., Schirmbeck, R., Reimann, J. (2003) Codelivery of a DNA vaccine and a protein vaccine with aluminum phosphate stimulates a potent and multivalent immune response, *J. Mol. Med.* **81**, 502–510.

Kwissa, M., Unsinger, J., Schirmbeck, R., Hauser, H., Reimann, J. (2000) Polyvalent DNA vaccines with bidirectional promoters, *J. Mol. Med.* **78**, 495–506.

Lai, W.C., Pakes, S.P., Ren, K., Lu, Y.S., Bennett, M. (1997) Therapeutic effect of DNA immunization of genetically susceptible mice infected with virulent mycoplasma pulmonis, *J. Immunol.* **158**, 2513–2516.

Lee, A.Y., Manning, W.C., Arian, C.L., Polakos, N.K., Barajas, J.L., Ulmer, J.B., Houghton, M., Paliard, X. (2000) Priming of hepatitis C virus-specific cytotoxic T lymphocytes in mice following portal vein injection of a liver-specific plasmid DNA, *Hepatology* **31**, 1327–1333.

Letvin, N.L., Montefiori, D.C., Yasutomi, Y., Perry, H.C., Davies, M.E., Lekutis, C., Alroy, M., Freed, D.C., Lord, C.I., Handt, L.K., Liu, M.A., Shiver, J.W. (1997) Potent, protective anti-HIV immune responses generated by bimodal HIV envelope DNA plus protein vaccination, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9378–9383.

Li, X., Sambhara, S., Li, C.X., Ewasyshyn, M., Parrington, M., Caterini, J., James, O., Cates, G., Du, R.P., Klein, M. (1998) Protection against respiratory syncytial virus infection by DNA immunization, *J. Exp. Med.* **188**, 681–688.

Lin, Y.L., Chen, L.K., Liao, C.L., Yeh, C.T., Ma, S.H., Chen, J.L., Huang, Y.L., Chen, S.S., Chiang, H.Y. (1998) DNA immunization with Japanese encephalitis virus nonstructural protein NS1 elicits protective immunity in mice, *J. Virol.* **72**, 191–200.

Livingston, B., Crimi, C., Newman, M., Higashimoto, Y., Appella, E., Sidney, J., Sette, A. (2002) A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes, *J. Immunol.* **168**, 5499–5506.

Lodmell, D.L., Ray, N.B., Parnell, M.J., Ewalt, L.C., Hanlon, C.A., Shaddock, J.H., Sanderlin, D.S., Rupprecht, C.E. (1998) DNA immunization protects nonhuman primates against rabies virus, *Nat. Med.* **4**, 949–952.

Lowrie, D.B., Tascon, R.E., Bonato, V.L., Lima, V.M., Faccioli, L.H., Stavropoulos, E., Colston, M.J., Hewinson, R.G., Moelling, K., Silva, C.L. (1999) Therapy of tuberculosis in mice by DNA vaccination, *Nature* **400**, 269–271.

MacGregor, R.R., Boyer, J.D., Ugen, K.E., Lacy, K.E., Gluckman, S.J., Bagarazzi, M.L.,

Chattergoon, M.A., Baine, Y., Higgins, T.J., Ciccarelli, R.B., Coney, L.R., Ginsberg, R.S., Weiner, D.B. (1998) First human trial of a DNA-based vaccine for treatment of human immunodeficiency virus type 1 infection: safety and host response, *J. Infect. Dis.* **178**, 92–100.

Macklin, M.D., McCabe, D., McGregor, M.W., Neumann, V., Meyer, T., Callan, R., Hinshaw, V.S., Swain, W.F. (1998) Immunization of pigs with a particle-mediated DNA vaccine to influenza a virus protects against challenge with homologous virus, *J. Virol.* **72**, 1491–1496.

Maloy, K.J., Erdmann, I., Basch, V., Sierro, S., Kramps, T.A., Zinkernagel, R.M., Oehen, S., Kundig, T.M. (2001) Intralymphatic immunization enhances DNA vaccination, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3299–3303.

Mancini, M., Hadchouel, M., Tiollais, P., Michel, M.L. (1998) Regulation of hepatitis B virus mRNA expression in a hepatitis B surface antigen transgenic mouse model by IFN-g-secreting T cells after DNA-based immunization, *J. Immunol.* **161**, 5564–5570.

Mancini-Bourgine, M., Fontaine, H., Scott-Algara, D., Pol, S., Brechot, C., Michel, M.L. (2004) Induction or expansion of T-cell responses by a hepatitis B DNA vaccine administered to chronic HBV carriers, *Hepatology* **40**, 874–882.

Manickan, E., Yu, Z., Rouse, B.T. (1997) DNA immunization of neonates induces immunity despite the presence of maternal antibody, *J. Clin. Invest.* **100**, 2371–2375.

Martinez, X., Brandt, C., Saddallah, F., Tougne, C., Barrios, C., Wild, F., Dougan, G., Lambert, P.H., Siegrist, C.-A. (1997) DNA immunization circumvents deficient induction of T helper type 1 and cytotoxic T lymphocyte responses in neonates and during early life, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 8726–8731.

Murphey, C.M., Wilson, L.A., Trichel, A.M., Roberts, D.E., Xu, K., Ohkawa, S., Woodson, B., Bohm, R., Blanchard, J. (1999) Selective induction of protective MHC class I-restricted CTL in the intestinal lamina propria of rhesus monkeys by transient SIV infection of the colonic mucosa, *J. Immunol.* **162**, 540–549.

Piedrafita, D., Xu, D., Hunter, D., Harrison, R.A., Liew, F.Y. (1999) Protective immune responses induced by vaccination with an expression genomic library of *Leishmania major*, *J. Immunol.* **163**, 1467–1472.

Pilon, S.A., Piechocki, M.P., Wei, W.Z. (2001) Vaccination with cytoplasmic erbb-2 DNA protects mice from mammary tumor growth without anti-erbb-2 antibody, *J. Immunol.* **167**, 3201–3206.

Robinson, H.L., Montefiori, D.C., Johnson, R.P., Manson, K.H., Kalish, M.L., Lifson, J.D., Rizvi, T.A., Lu, S., Hu, S.L., Mazzara, G.P., Panicali, D.L., Herndon, J.G., Glickman, R., Candido, M.A., Lydy, S.L., Wyand, M.S., McClure, H.M. (1999) Neutralizing antibody-independent containment of immunodeficiency virus challenges by DNA priming and recombinant pox virus booster immunizations, *Nat. Med.* **5**, 526–534.

Rosenecker, J., Naundorf, S., Gersting, S.W., Hauck, R.W., Gessner, A., Nicklaus, P., Muller, R.H., Rudolph, C. (2003) Interaction of bronchoalveolar lavage fluid with polyplexes and lipoplexes: analysing the role of proteins and glycoproteins, *J. Gene Med.* **5**, 49–60.

Schirmbeck, R., Fissolo, N., Chaplin, P., Reimann, J. (2003) Enhanced priming of multispecific, murine $CD8^+$ T cell responses by DNA vaccines expressing stress protein-binding polytope peptides, *J. Immunol.* **171**, 1240–1246.

Schirmbeck, R., Kwissa, M., Fissolo, N., Elkholy, S., Riedl, P., Reimann, J. (2002) Priming polyvalent immunity by DNA vaccines expressing chimeric antigens with a stress protein-capturing, viral J-domain, *FASEB J.* **16**, 1108–1110.

Schirmbeck, R., Zheng, X., Roggendorf, M., Geissler, M., Chisari, F.V., Reimann, J., Lu, M. (2001) Targeting murine immune responses to selected T cell- or antibody-defined determinants of the hepatitis B surface antigen by plasmid DNA vaccines encoding chimeric antigen, *J. Immunol.* **166**, 1405–1413.

Schmidt, T., Friehs, K., Flaschel, E. (1996) Rapid determination of plasmid copy number, *J. Biotechnol.* **49**, 219–229.

Schmidt, T., Friehs, K., Schleef, M., Voss, C., Flaschel, E. (2002) Quantitative analysis of plasmid forms by agarose and capillary gel electrophoresis, *Anal. Biochem.* **274**, 235–240.

Sedegah, M., Jones, T.R., Kaur, M., Hedstrom, R., Hobart, P., Tine, J.A., Hoffman, S.L. (1998) Boosting with recombinant vaccinia increases immunogenicity and protective efficacy of

malaria DNA vaccine, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7648–7653.

Smith, G.J. III, Helf, M., Nesbet, C., Betita, H.A., Meek, J., Ferre, F. (1999) Fast and accurate method for quantitating *E. coli* host-cell DNA contamination in plasmid DNA preparations, *BioTechniques* **26**, 518–522, 524–526.

Spellerberg, M.B., Zhu, D., Thompsett, A., King, C.A., Hamblin, T.J., Stevenson, F.K. (1997) DNA vaccines against lymphoma: promotion of anti-idiotypic antibody responses induced by single chain Fv genes by fusion to tetanus toxin fragment C, *J. Immunol.* **159**, 1885–1892.

Syrengelas, A.D., Levy, R. (1999) DNA vaccination against the idiotype of a murine B cell lymphoma: mechanism of tumor protection, *J. Immunol.* **162**, 4790–4795.

Tedeschi, V., Akatsuka, T., Shih, J.W., Battegay, M., Feinstone, S.M. (1997) A specific antibody response to HCV E2 elicited in mice by intramuscular inoculation of plasmid DNA containing coding sequences for E2, *Hepatology* **25**, 459–462.

Walther, W., Stein, U., Voss, C., Schmidt, T., Schleef, M., Schlag, P.M. (2003) Stability analysis for long-term storage of naked DNA: impact on nonviral in vivo gene transfer, *Anal. Biochem.* **318**, 230–235.

Walther, W., Stein, U., Fichtner, I., Voss, C., Schmidt, T., Schleef, M., Nellessen, T., Schlag, P.M. (2002) Intratumoral low-volume jet-injection for efficient nonviral gene transfer, *Mol. Biotechnol.* **21**, 105–115.

Wang, R., Doolan, D.L., Le, T.P., Hedstrom, R.C., Coonan, K.M., Charoenvit, Y., Jones, T.R., Hobart, P., Margalith, M., Ng, J., Weiss, W.R., Sedegah, M., de Taisne, C., Norman, J.A., Hoffman, S.L. (1998) Induction of antigen-specific cytotoxic T lymphocytes in humans by a malaria DNA vaccine, *Science* **282**, 476–480.

Wild, J., Grüner, B., Metzger, K., Kuhröber, A., Pudollek, H.-P., Hauser, H., Schirmbeck, R., Reimann, J. (1998) Polyvalent vaccination against hepatitis B surface and core antigen using dicistronic expression plasmids, *Vaccine* **16**, 353–360.

Yasutomi, Y., Robinson, H.L., Lu, S., Mustafa, F., Lekutis, C., Arthos, J., Mullins, J.I., Voss, G., Manson, K., Wyand, M., Letvin, N.L. (1996) Simian immunodeficiency virus-specific cytotoxic T-lymphocyte induction through DNA vaccination of rhesus monkeys, *J. Virol.* **70**, 678–681.

Youssef, S., Maor, G., Wildbaum, G., Grabie, N., Gour, L.A., Karin, N. (2000) C-C chemokine-encoding DNA vaccines enhance breakdown of tolerance to their gene products and treat ongoing adjuvant arthritis, *J. Clin. Invest.* **106**, 361–371.

Zhang, J., Silvestri, N., Whitton, J.L., Hassett, D.E. (2002) Neonates mount robust and protective adult-like $CD8^+$ T-cell responses to DNA vaccines, *J. Virol.* **76**, 11911–11919.

Vaccine Approaches for Protection Against HIV

Stephen M. Smith[1], *Mahender Singh*[1], *and Kuan-Teh Jeang*[2]
[1] *Saint Michael's Medical Center & New Jersey Medical School, Newark, NJ, USA*
[2] *Laboratory of Molecular Microbiology, NIAID, NIH, Bethesda, MD, USA*

1 Introduction 136

2 Immune Responses Elicited by Viruses and Vaccines 137

3 Challenges to HIV-1 Vaccine Development 139

4 Vaccine Development Stages 140
4.1 Preclinical Phase 140
4.2 Human Trials 142

5 Animal Models 142

6 Vaccine Approaches 144
6.1 Inactivated Virus 144
6.2 Virus Subunit 145
6.3 Live-attenuated Virus 146
6.4 Viral and Bacterial Vectors 146
6.5 DNA Vectors 147

7 Conclusion 148

Bibliography 148
Books and Reviews 148
Primary Literature 149

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

Keywords

AIDS
Acquired immunodeficiency syndrome.

CTL
Cytotoxic T lymphocytes; a type of white blood cells that can kill cells infected with viruses.

DNA
Deoxyribonucleic acid.

HIV
Human immunodeficiency virus.

SIV
Simian immunodeficiency virus.

Vaccine
An agent that stimulates an immune response in the body in order to prevent future infection.

Human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) has reached pandemic proportions globally. Currently, there are approximately 40 million individuals infected with HIV-1. The most likely solution for this disease is a preventive vaccine that can be distributed worldwide. Despite intense efforts, the development of a vaccine against HIV has not been realized. Here, we review the challenges that have impeded progress and the approaches that are currently under consideration. Vaccination results from human trials and animal models are also discussed.

1 Introduction

The first cases of Acquired Immunodeficiency Syndrome (AIDS) were described in homosexual men in the United States in 1981. Within three years after the initial reports, French and American scientists confirmed that the causative agent for AIDS was a retrovirus, the human immunodeficiency virus (HIV). Today, more than 20 years later, over 20 million

individuals have died of AIDS; and in 2004, approximately 40 million people globally are infected and living with HIV-1.

The worldwide trend for HIV/AIDS is not encouraging. In 2003 alone, 3 million individuals became newly infected with HIV-1 in sub-Saharan Africa; an additional 1 million new cases of HIV-1 infection emerged in Asia. Even in developed nations, HIV-1 infection remains on the rise. Today, more than 950 000 and 580 000 individuals in the United States and Western Europe, respectively, are seropositive for HIV-1. New advances in the treatment and prevention of HIV-1 also appear to be insufficient to halt the spread of HIV-1. The United Nations estimate that only 7% of people in developing countries have access to antiretroviral treatment, and that only 20% of people worldwide are reached by HIV-1 prevention programs. Indeed, data from countries with ongoing HIV/AIDS treatment/prevention initiatives have shown that such approaches have failed to effectively blunt the spread of AIDS. In this light, past vaccination successes in controlling viral epidemics such as smallpox and polio suggest that a cost-effective preventive HIV/AIDS vaccine is reasonably the best public health solution.

The development of an effective, prophylactic HIV-1 vaccine has proven to be extremely difficult. Despite an extraordinary devotion of time, effort, and money, the incremental advances in our attempts to generate immunogens, which offer protective immunity to HIV-1, have been discouraging. Nevertheless, this slow progress has to be viewed in the context that to date no effective vaccine against any retrovirus has been developed. Scientifically, we are just commencing our understanding of the *in vivo* immunological parameters that dictate durable control over retroviral infections. In this chapter, we summarize the approaches, the progress, and our current state of immunological knowledge in developing an HIV/AIDS vaccine.

2 Immune Responses Elicited by Viruses and Vaccines

Viruses come in different flavors. Smallpox, polio, and measles viruses cause acute infections. These viruses, after an episodic presence, are normally cleared by the host with no lasting sequelae. By contrast, other viruses like the herpesviruses, present and maintain chronic and/or latent infections. The latter pathogens establish a life-long parasitic relationship with their hosts, and likely survive in this relationship because they have evolved mechanisms for evading immune-mediated elimination.

The adaptive humoral and cell-mediated immunities are the prototypic responses mounted by human host to all viral infections (Fig. 1). In response to replication-competent viruses, humans are, in principle, capable of generating neutralizing antibodies (humoral immunity) and cytotoxic T lymphocytes (CTL) (cell-mediated immunity). Our current understanding of virology and vaccinology suggests that the key immune response, which clears an acute infection, preventing chronicity, correlates with virus-specific neutralizing antibody. On the other hand, cell-mediated immunity is generally deemed necessary to control a chronic viral infection, once established. The latter view does not, however, exclude a protective role for antibodies in chronic infection.

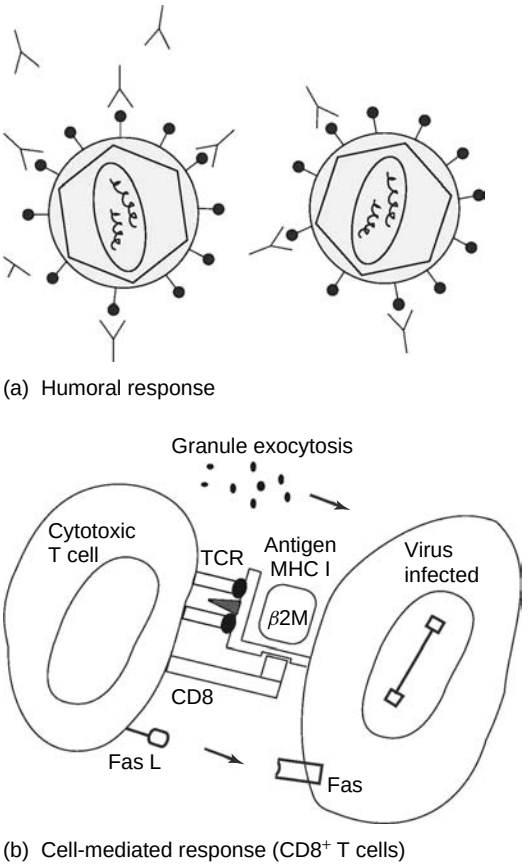

Fig. 1 The two major types of adaptive immunity triggered by replication-competent viruses in mammalian hosts. (a) Humoral or antibody responses are elicited. Antibodies represented by the Y symbol can target epitopes on the virus, "neutralizing" the virus and preventing infection of cells. (b) MHC class I restricted cytotoxic T-lymphocyte response can also be provoked. The diagram shows a cytotoxic ($CD8^+$) T-lymphocyte, which recognizes through its T-cell receptor (TCR) antigen (brown triangle) presented by the MHC class I molecule on the surface of a virus-infected cell. The cytotoxic T cell can kill the virus-infected cell via two mechanisms, exocytosis of toxic granules that perforate target cell membranes and induction of apoptosis using Fas ligand (Fas L)–Fas interaction. $\beta 2M = \beta$-2 microglobulin. (See color plate p. xxvii.)

A requisite step in the lifecycle of all retroviruses is the integration of the viral genome into the host's cellular chromosome. Hence, infection by a retrovirus, if initial entry into permissive cells is not prevented by neutralizing antibody, has the potential to persist and spread chronically within the host so long as infected cells are not eliminated. Early infection studies of HIV-1 in the chimpanzee model showed that passive infusion of neutralizing antibodies can prevent the establishment of a chronic infection. However, should the body's initial opportunity to generate neutralizing antibody to prevent HIV-1 infection of cells fail, a chronic and progressive viral infection then takes hold. The natural viral history suggests that humans do not effectively respond with protective and sterilizing antibody to initial HIV-1 infection, and immunity against reinfection is never established. Indeed, HIV-1 does establish a chronic progressive infection and can rapidly mutate its genome to escape the restriction enforced by initial neutralizing antibodies. Because natural humoral correlates of protection are not observed, it has been reasoned that cell-mediated CTL response acts to control chronic *in vivo* HIV-1 replication and dictates the rapidity of progression to AIDS.

3
Challenges to HIV-1 Vaccine Development

How does one then go about developing a vaccine against HIV-1? A major obstacle to immunological success is the plasticity of the viral genome. Currently, HIV-1 exists in many different subtypes (clades), each differing from the other by about 10% in their respective genetic content. In this context, one also notes that HIV-1 turnover *in vivo* is astonishingly high with up to 10^9 virions being produced each day. Superimpose onto this number the fact that the reverse transcriptase enzyme of HIV-1 is highly error prone (mutation rate of approximately 1 per 10^5 bases), then one expects for a genome size of 10^4 bases, if all things were equal, that every possible mutation would occur at every position of the genome everyday. This mathematics argues that for each epitope to which a specific immune response is raised by the host, the virus could and would change the selected epitope and thus, rapidly affect immunological evasion. If this argument holds true, then it is very unlikely that a useful HIV vaccine can be developed. Accordingly, the foremost challenge for a successful vaccine is to identify critical viral epitopes, which the virus for reasons of functional constraint, cannot mutate for purposes of evading immune restriction. If such immutable epitopes are also broadly conserved across different HIV-1 clades, then success could be contemplated.

A recent report lends tangible credence to the concept for the existence of immutable epitopes to which neutralizing antibodies could be elicited against HIV-1. Hence, researchers recently showed that a caveolin-binding motif (CBD1) in gp41 is highly conserved across all HIV-1 clades, and that a broadly neutralizing antibody against CBD-1 halts virus infection in tissue culture. Building upon this new finding, logical next steps in vaccine development would be to uncover additional epitopes in Env and Gag, which are either linearly or conformationally conserved across clades and to test these for ability to elicit durable neutralizing antibody.

If a neutralizing antibody vaccine cannot be developed to prevent HIV infection, it still may be possible to develop a CTL-eliciting immunogen, which would control virus replication *in vivo* and delay or prevent the onset of AIDS. In general, HIV-1 infection in humans is not itself symptomatic, but rather it causes gradual immune system deterioration. Over time, $CD4^+$ T cells are depleted. The definition of acquired immunodeficiency syndrome or AIDS is when a person's $CD4^+$ T cells fall below 200 cells μL^{-1} of blood. A person with AIDS is susceptible to many opportunistic infections and malignancies, which do not occur in people with intact immune systems. These opportunistic illnesses are the causes of death. On average, HIV infection leads to AIDS within 8 to 10 years and has a plasma viral load of 10^4–10^5 copies mL^{-1} (Fig. 2). The level of viremia closely correlates with the rate of $CD4^+$ T-cell depletion. The higher a person's viral load is, the quicker that person loses $CD4^+$ T cells and advances to AIDS. The goal of a vaccine, which induces anti-HIV CTL, is to significantly reduce viremia by eliminating productively infected cells. A reduction in viremia will slow the rate of immune suppression and will decrease the rate of transmission to others. This approach contrasts with that of all other viral vaccines, which prevent or limit acute infection. In the same way, it eludes antibodies; HIV is facilely capable of generating CTL-escape mutants. The solution here rests on coming up with CTLs, which target immutable epitopes. There

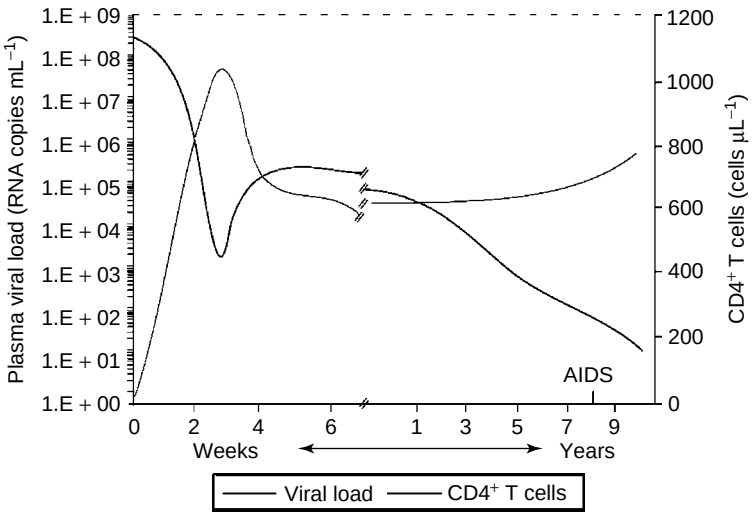

Fig. 2 Typical HIV-1 infection is characterized by acute-phase viremia followed by chronic intermediate levels of plasma viremia, a steady loss of peripheral $CD4^+$ T cells and the development of AIDS after 7 to 10 years. (Adapted from Feinberg, M.B., Moore, J.P. (2002) AIDS vaccine models: challenging challenge viruses, *Nat. Med.* **8**(3), 207–210.)

is emerging evidence that such immutable CTL targets exist in the HIV-1 genome. For example, we recently showed that a CTL epitope in the second exon of the viral Tat protein appears to fit that definition. Going forward, the scientific challenges are to induce high titers of neutralizing antibodies and potent cellular immune responses to conserved viral epitopes, which cannot be changed by HIV-1 in response to immune restriction without drastic loss of function for the virus.

4 Vaccine Development Stages

4.1 Preclinical Phase

Preclinical testing, which occurs in laboratories or animal models, of vaccine candidates takes years to decades (Fig. 3). Initial basic experiments defined HIV's proteins and its basic life cycle. HIV-1 is a retrovirus of the lentivirus genus. The HIV-1 RNA genome is ~9 kb and encodes 9 polypeptides. Posttranslation proteolytic digestion results in the formation of 15 discrete proteins. The major structural proteins are encoded by three genes, *gag* (group-specific antigen), *pol* (polymerase), and *env* (envelope) (Fig. 4), while the accessory proteins, Vif, Vpu, Vpr, and Nef, and the regulatory proteins, Tat and Rev, are the primary translation products of multiply spliced mRNAs. In vaccine development, each viral protein (15 separate proteins and many subtypes of each) represents a possible vaccine target. After selecting a viral target or targets, researchers next choose a delivery system or approach, such as recombinant adenovirus. Then, animal models are used to test the immunogenicity and the safety of the vaccine strategy. In many cases,

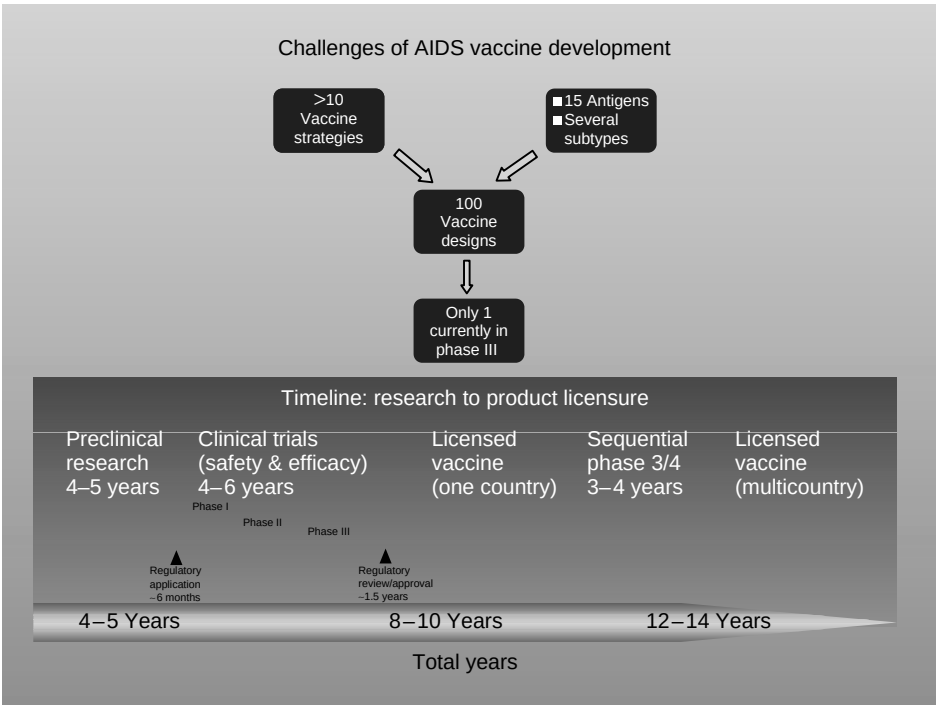

Fig. 3 Challenges to HIV vaccine development. Each stage of vaccine development takes years to complete. In the absence of an immunologic goal or target, researchers have attempted to raise immune responses against many viral antigens through several different approaches. Despite these efforts, only one vaccine type has been tested in Phase III trials in the 23 years of the epidemic.

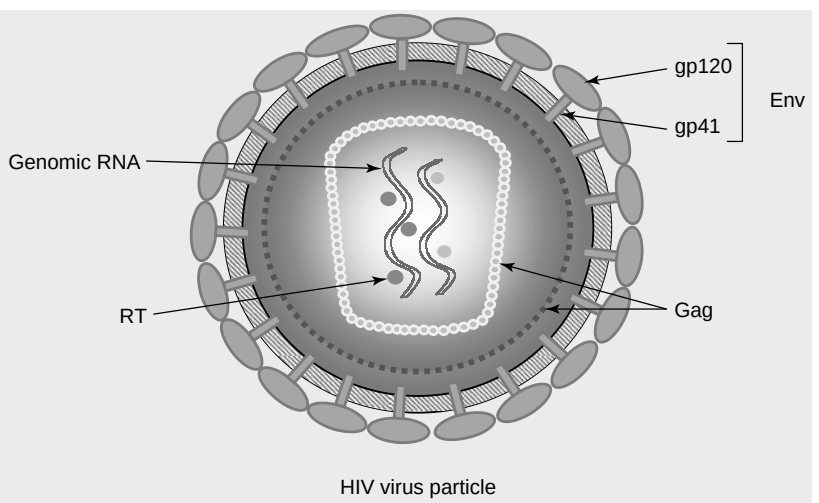

Fig. 4 Schematic of HIV virus particle.

small efficacy studies are performed in the macaque model of AIDS (see the following). Once a vaccine candidate successfully passes these tests, it can then be considered for human testing or clinical trials. Of the multitude of HIV vaccine candidates tested in preclinical studies, only one has been completely tested in an efficacy trial (Phase III), which, as discussed, failed.

4.2 Human Trials

Vaccines, like new therapeutics, are evaluated in three successive stages of clinical testing, referred to as Phase I, II, and III trials. In Phase I testing, the side effects and safety of a given vaccine are measured in a small number of healthy volunteers who are at low risk of infection. Typically, 10 to 30 individuals are inoculated with the vaccine at different doses and immunization schedules. Participants are closely monitored for the development of adverse reactions, and immune responses to the vaccine are evaluated secondarily. Phase I trials are usually completed in approximately 1 year. In Phase II trials, several hundred people are given the vaccine to collect additional short-term safety data and, more so, to measure the immunogenicity of the vaccine. Participants in Phase II trials are individuals who have varied rates of risk for acquiring the infection, and these trials typically take 1.5 to 2 years to complete. Once a vaccine has been determined to be safe and sufficiently immunogenic, it is tested in an efficacy, or Phase III, trial.

Several thousand volunteers must be enrolled in a Phase III vaccine trial. The study must be placebo controlled and double-blinded such that neither the researcher nor the volunteer knows whether he/she is receiving placebo or vaccine. Further, Phase III studies must be conducted in a population with substantial risk of HIV-1 infection and can take three years to complete. Combined, Phase I, II, and III trials for a candidate HIV vaccine may require up to 6 years to yield the necessary safety, immunogenicity, and efficacy data.

The National Institutes of Health in the United States has formed the HIV Vaccine Trials Network (HVTN) to help organize and sponsor vaccine research. The self-stated mandate of HVTN is to conduct "all phases of clinical trials, from evaluating candidate vaccines for safety and the ability to stimulate immune responses to testing vaccine efficacy. The HVTN mission is to develop and test preventive HIV vaccines. This research is done through multicenter clinical trials in a global network of domestic and international sites." Other countries, including Great Britain, Thailand, Kenya, Uganda, South Africa, Australia, and Italy, actively support HIV vaccine trials. Several pharmaceutical companies have developed significant HIV vaccine projects, most notably Merck, Aventis, and Glaxo Smith Kline (GSK). For several years, the International AIDS Vaccine Initiative (IAVI), a nonprofit global organization, has been an advocacy group for HIV vaccine research and testing. Now, IAVI is directly supporting clinical testing. The HIV vaccine effort is not lacking for governmental, nongovernmental, or industrial support. Rather, the many institutions and foundations lack a clear immunologic target.

5 Animal Models

HIV-1 does not productively infect any other animal, with the possible exception

of chimpanzees. Small laboratory animals such as mice, guinea pigs, and rabbits have been used to evaluate the safety and immunogenicity of various vaccine preparations. These animals have also been used for testing the potency of adjuvants and immunostimulatory molecules (e.g. TNF-α, IL-2, IL-12). However, such animals cannot be used for evaluating the efficacy of candidate vaccines, since these animals are not infected by HIV-1. The lack of a good small animal model for HIV-1 infection and pathogenicity has been one of the major difficulties in AIDS vaccine development.

Macaques, which originate from Asia, are the animals, most often used in HIV vaccine development. Monkeys and apes from Africa harbor more than 30 lentiviruses, called *simian immunodeficiency viruses* (SIVs). The SIVs are related to HIV in their genomic organization and sequence homology. None of these viruses causes disease in its natural host; however, SIV causes AIDS when inoculated into Asian macaques. SIV does not infect macaques in the wild. This model was developed after researchers serendipitously discovered that macaques, which had been housed with African monkeys, developed AIDS via cross-species transmission of SIV. Infection in captive macaques is used as a disease model to study simian AIDS.

Laboratory-generated chimeras between HIV-1 and SIV, called *simian–human immunodeficiency viruses* (SHIV), have also been created to be used in pathogenesis and vaccine studies in macaques. SHIVs typically have gag, pol, vif, vpx, vpr, and the long terminal repeat (LTR) of SIV, while tat, rev, and env are derived from HIV-1. While not all SHIVs are pathogenic in macaques, a subset of pathogenic SHIVs have been identified and shown to cause rapid immunodeficiency and disease progression. As a result, several pathogenic SHIVs have been utilized as challenge viruses in vaccine studies.

Among the different species of macaques, rhesus macaques (Macaca mulatta), cynomolgus macaques (Macaca fascicularis), and pigtailed macaques (Macaca nemestrina) are most commonly used for pathogenesis and vaccine studies. Early studies focused on chimpanzees as the only nonhuman animal susceptible to infection with HIV-1; however, these animals are infrequently used today because of their failure to develop significant viremia and disease following HIV-1 inoculation. Moreover, legal and ethical issues have created mounting pressure for laboratories in the United States to discontinue research using chimpanzees. Consequently, SIV infection of macaques has become the most frequently used model to evaluate candidate vaccines. Despite the ability to establish SIV infection, this model is not without drawbacks. Owing to the high cost of macaques and limited supply, vaccine studies are often constrained to the use of a small number of animals (<20). Often, in the interest of economy, this initial group of animals is further divided into subgroups of 4 to 5 macaques to simultaneously evaluate several related strategies. An unfortunate consequence of such small group sizes is the loss of statistical power and the potential for misinterpretation of results due to inapparent biases. However, in the absence of an alternative, research in the SIV-macaque model has provided valuable information on the induction of immune responses and the efficacy against pathogenic virus challenge for many candidate vaccines. In the following sections, we will review data generated in this model encompassing several different

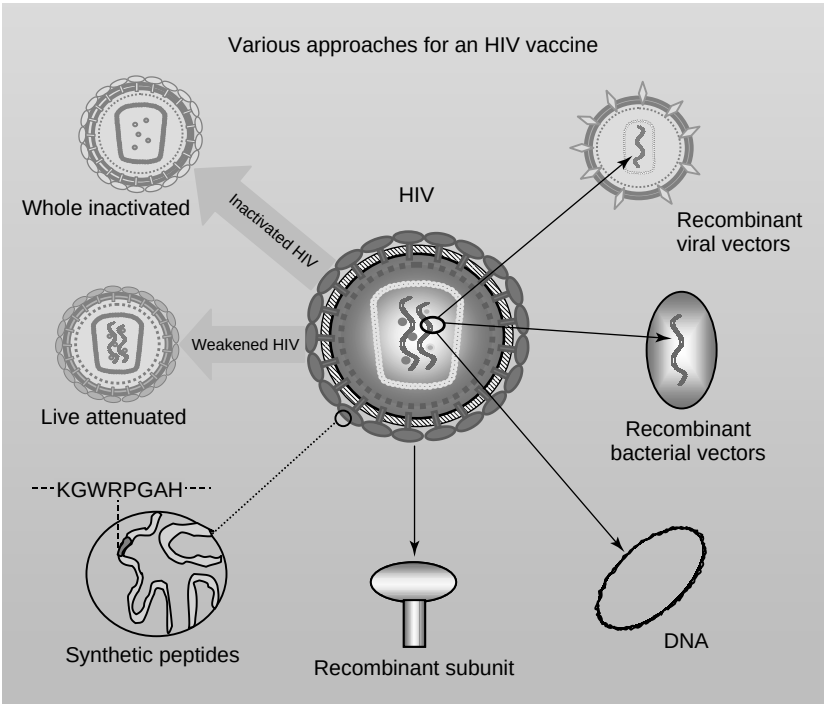

Fig. 5 Various approaches for an HIV vaccine are illustrated. (Please see text for elaboration.)

vaccine strategies (Fig. 5) and the limited work conducted in human clinical trials.

6 Vaccine Approaches

6.1 Inactivated Virus

Inactivated virus particles have been used successfully as vaccines against several human pathogens, including poliovirus, influenza, and hepatitis A virus. In this simple strategy, virus is grown in tissue culture or other vessel, such as chicken eggs. Next, virus is purified and chemically inactivated, while retaining the overall particle structure. The dead virus particles are then injected into the host. In response to such vaccines, humans develop a primary antibody or humoral response. In some viral diseases, these antibodies confer significant protection against infection and/or severe disease.

In the late 1980s and early 1990s, several researchers used this approach against SIV in macaques. This approach worked well in initial experiments. In the following years, whole-inactivated vaccines were shown to protect macaques from pathogenic SIV infection by intravenous, intrarectal, and intravaginal routes. The initial optimism of these results turned out to be based on false hope. SIV used for vaccine preparations was grown in human cells and the human cellular proteins were incorporated into the vaccine virus particles. Such vaccine preparations induced antibody

responses in vaccinated animals against the human cellular proteins and afforded protection against the challenge viruses, which were also propagated in human cells. High levels of antibodies against human cellular proteins were detected in the protected animals and the degree of protection correlated with the level of these antibody titers. This observation of xeno antigen–induced protection was further confirmed by experiments where macaques immunized with human cells alone, HLA-DR, or HLA class I proteins were protected against SIV grown in human cells, but not against SIV grown in macaque cells. Several researchers tried to exploit the finding of xeno-antigen responses by designing vaccine strategies based on alloimmunization, but concerns over inducing autoimmune responses rendered this approach unattractive.

6.2 Virus Subunit

With the advent of recombinant molecular biology, it is simple to express viral proteins individually and separately. The strategy has been extremely successful against many viruses, including hepatitis B virus. In this approach, the gene for a viral protein, usually the viral envelope, is expressed *in vitro*. The viral protein is purified and then injected into the host with an adjuvant. These protein-based vaccines illicit a strong humoral response, but a very limited cellular immune response. In some cases, the viral subunit approach induces a similar humoral response as the inactivated viral vaccine (i.e. influenza), but causes fewer side effects. Many attempts using the subunit approach were made in the macaque model to induce protective antibodies against pathogenic SIV. However, they were not successful. Most approaches used some form of viral envelope, typically gp120, since this is the only viral protein exposed to antibodies.

Despite the many failures in the macaque model with the subunit approach, the National Institutes of Health supported two large Phase III vaccine trials, which used this simplistic strategy. This vaccine, AIDSVAX, is a monomeric version of the normally trimeric HIV-1 envelope glycoprotein, gp120. The vaccine was found to induce a strong antibody response against linear epitopes on gp120 and a $CD4^+$ T-cell proliferative response. However, in two independent trials, which tested slight variations of the same approach, the vaccine failed to confer protection against HIV-1 infection and did not significantly alter virus replication, as measured by plasma viral load.

The first of two Phase III trials of AIDSVAX was conducted in the United States, Puerto Rico, Canada, and the Netherlands and enrolled 5109 men and 309 women at high risk for acquiring HIV-1 infection. The volunteers were randomized to either the vaccine (AIDSVAX B/B) or placebo arm of the study at a ratio of 2 : 1. Those in the vaccine arm received AIDSVAX B/B, which contains two forms of gp120 derived from subtype B virus; the most common subtype found throughout North America and Europe. Of those enrolled, 5009 participants completed the three-year study. While adverse reactions were minimal, the infection rates for the vaccine and placebo groups were 5.7 and 5.8%, respectively, and were not statistically different. The second efficacy trial was conducted in Thailand and enrolled 2546 injection drug users. For this study, the vaccine was changed to AIDSVAX B/E to reflect the subtypes of HIV more commonly found in Asia (subtypes B and E). This trial again

showed no protection against HIV-1 infection providing convincing evidence that immunization with monomeric gp120 is ineffective as a vaccine strategy. Secondary to its lack of efficacy, researchers currently use this approach in conjunction with another strategy and never alone.

6.3 Live-attenuated Virus

Inactivated virus particles are capable of inducing immune responses, which for some viral diseases can be protective. Live-attenuated versions, however, often produce stronger immune responses in a larger percentage of hosts than a comparable inactivated or dead version. The classic comparison of these two approaches is that between the Salk (inactivated virion) polio vaccine and the Sabin (attenuated viruses) polio vaccine. The Salk vaccine, while effective in many, does not stimulate the same degree of immune response as the Sabin vaccine. By replicating within the host, an attenuated vector more effectively mimics a limited natural infection and elicits humoral and cell-mediated immune responses. In general, attenuated virus vaccines are more efficacious than their inactivated counterparts. The major disadvantage of attenuated viral vaccines is the possibility or propensity to cause disease. The Sabin vaccine, for instance, causes 1 case of vaccine-associated poliomyelitis per 1 000 000 recipients, while the Salk vaccine causes none.

In HIV vaccine development, no attenuated virus candidate has been evaluated in clinical trials. Since HIV is a retrovirus, even an attenuated version would not be completely cleared from the host. While scientists can easily modify or delete each gene from HIV, no one knows how to attenuate HIV in such a way that it makes the virus safe and yet effective. In the macaque model, certain forms of attenuated SIV protect animals from infection with wild-type virus. Several researchers have confirmed this finding. Unfortunately, in neonatal macaques and even in some adults, this attenuated viral vaccine replicates efficiently enough to cause gradual depletion of the immune system. Forms of attenuated virus, which replicate less efficiently, fail to induce protective immunity in any animals. In conclusion, at least in the SIV-macaque model, attenuated vaccines, as currently conceived through gene deletion or interruption, are not capable of safely protecting a large population of animals. To improve the safety profile, some groups, like ours, continue to work on strategies that allow conditional control of the attenuated strain. In these models, the attenuated HIV's replication would be dependent upon the coadministration of a drug. Once the drug was discontinued, the attenuated viral vaccine would cease replicating. This approach has not yet been tested in macaques.

6.4 Viral and Bacterial Vectors

As mentioned earlier, attenuated viral vaccines have the benefit of replicating within the host and thus induce more potent immune responses. Investigators have used molecular techniques to produce viral and bacterial vectors, which express HIV proteins. Several recombinant viruses have been tested in macaques. These include adenovirus, different alphaviruses, various pox viridae, and even rhabdoviruses. As cellular parasites, each viral vector replicates within the cell and uses the cell's protein producing machinery. Intracellular expression of foreign or viral proteins leads to enhanced cell-mediated immune responses. In the macaque model, many different recombinant viruses have shown

promise and are able to induce an immune response that controls the replication of SIV. Unlike the Sabin vaccine against poliovirus, these recombinant virus vaccines do not protect the animals against infection, but rather against disease.

Recombinant viruses, which encode HIV proteins, are now being tested in human trials. A canarypox vector called ALVAC is currently in Phase III testing. ALVAC has been studied extensively in Phase I and II trials, and has been shown to have an excellent safety profile. This is because of the fact that canarypox infects mammalian cells, in which it completes a single cycle of replication but does not release infectious virions. In the current clinical trial (designated RV144), ALVAC is produced by Aventis and engineered to express HIV-1 subtype E env (gp120), subtype B env (gp41), gag, and protease (designated ALVAC-HIV-vCP1521).

Since protection against HIV has been impossible to achieve with any one vector or vaccine, researchers are now using two or more vaccines in what is termed *prime-boost strategy*. The first vaccine is given to prime the immune system's response against HIV antigens and then the second is administered to boost this response. In fact, most planned Phase II and III trials employ this approach. One prime-boost strategy is DNA vaccination (see the following) followed by recombinant poxvirus. Both vectors induce primarily a cell-mediated immune response. This approach has worked well in the macaque model against pathogenic SHIVs, but not against pathogenic SIV. In other prime-boost regimens, the first vector is given to raise cell-mediated responses, while the second, usually soluble protein, is given to induce a humoral response. This latter approach is currently in the Phase III study, RV144, in which ALVAC is used to prime the cellular immune response and soluble gp120, AIDSVAX, is administered as the boost to raise antibodies.

Recently, Phase I and II trials with another recombinant poxvirus, modified vaccinia Ankara (MVA) was completed. The immune response targets were not reached with this vector. Plans for a Phase III trial were scrapped. Recombinant adenovirus has been very effective in protecting macaques against disease and is being studied in humans in Phase II testing. Several additional HIV-1 vaccine strategies are now being tested in Phase I trials. A comprehensive list can be found on the International AIDS Vaccine Initiative or IAVI Web site (http://www.iavireport.org/trialsdb/).

Bacteria such as Salmonella and Listeria can survive well intracellularly. Researchers have used recombinant techniques to manufacture attenuated versions of these bacteria, which express HIV proteins, to deliver intracellular expression. In small animal models, these bacteria do elicit anti-HIV immune responses. These approaches have been extensively studied in the macaque model or in human trials yet.

6.5 DNA Vectors

Direct inoculation of purified preparations of DNA as a means of immunization has become an exciting area of research, stemming from an unexpected observation in mice. DNA, encoding antigens, or even adjuvant cytokines, can be delivered either intradermally, intramuscularly or mucosally, from where it is taken up by host cells. Host cell production of the encoded proteins may enhance presentation of antigenic peptides on major histocompatibility complex (MHC)

Class I molecules, and facilitate the development of antigen-specific CTL responses. Cytokine delivery by DNA allows for localized delivery to the same tissue in which antigen is being expressed and processed. DNA vaccination is currently being tested in early clinical trials. Early data suggest that human immune responses to DNA vectors are less than that seen in mice.

7 Conclusion

In the history of medicine, no vaccination effort has been as large or as extensive as that against HIV-1. Many different approaches and multiple attempts at each have not yielded a clear avenue to follow. From these efforts, we have learned that prevention of infection through vaccination (i.e. attenuated SIV) is possible. However, we also know that it is difficult to induce neutralizing antibodies by vaccination, despite the recognition that this approach offers the most promise for a truly preventive vaccine. CTL-based vaccines can protect against disease in some macaque models of HIV, but not in others, and escape from CTL can lead to vaccine failure. HIV-1 vaccine trials are expensive and the current scientific landscape suggests that a successful HIV-1 vaccine may still be years away. Until that time, widespread use of antiretroviral drugs offers a tangible hope to slow the epidemic.

See also AIDS/HIV, Molecular and Cell Biology; Cytomegalovirus and Varicella–zoster Virus Vaccines; HIV Therapeutics, Biochemistry of; Liposome Gene Transfection; Medicinal Chemistry; Retrovirus Integration, Biological and Biochemical Machineries for; Vaccination, Genetic; Vaccine Research and Development Methodology.

Bibliography

Books and Reviews

Apetrei, C., Marx, P.A., et al. (2004) The evolution of HIV and its consequences, *Infect. Dis. Clin. North Am.* **18**(2), 369–394.

Burton, D.R., Desrosiers, R.C., et al. (2004) Public health. A sound rationale needed for phase III HIV-1 vaccine trials, *Science* **303**(5656), 316.

Cohen, J. (2003a) Clinical research. A setback and an advance on the AIDS vaccine front, *Science* **300**(5616), 28–29.

Cohen, J. (2003b) HIV/AIDS. Vaccine results lose significance under scrutiny, *Science* **299**(5612), 1495.

Cohen, J. (2003c) Public health. AIDS vaccine still alive as booster after second failure in Thailand, *Science* **302**(5649), 1309–1310.

Cohen, J. (2003d) Public health. AIDS vaccine trial produces disappointment and confusion, *Science* **299**(5611), 1290–1291.

Feinberg, M.B., Moore, J.P. (2002) AIDS vaccine models: challenging challenge viruses, *Nat. Med.* **8**(3), 207–210.

Guatelli, J.C., Siliciano, R.F., et al. (2002) Human Immunodeficiency Virus, in: Richman, D.D., Whitley, R.J., Hayden, F.G. (Eds.) *Clinical Virology*, American Society for Microbiology Press, Washington, DC, pp. 685–729.

Hirsch, V.M. (2004) What can natural infection of African monkeys with simian immunodeficiency virus tell us about the pathogenesis of AIDS? *AIDS Rev.* **6**(1), 40–53.

Hirsch, V.M., Lifson, J.D. (2000) Simian immunodeficiency virus infection of monkeys as a model system for the study of AIDS pathogenesis, treatment, and prevention, *Adv. Pharmacol.* **49**, 437–477.

Joag, S.V. (2000) Primate models of AIDS, *Microbes Infect.* **2**(2), 223–229.

Johnson, R.P. (2002) Mechanisms of protection against simian immunodeficiency virus infection, *Vaccine* **20**(15), 1985–1987.

Marx, P.A. (1995) Attenuated retrovirus vaccines and AIDS, *Science* **270**(5239), 1219–1220; author reply 1220–1222.

McCarthy, M. (2003) HIV vaccine fails in phase 3 trial, *Lancet* **361**(9359), 755–756.

Shearer, G.M., Pinto, L.A., et al. (1999) Alloimmunization for immune-based therapy and vaccine design against HIV/AIDS, *Immunol. Today* **20**(2), 66–71.

Smith, S.M. (2002) HIV vaccine development in the nonhuman primate model of AIDS, *J. Biomed. Sci.* **9**(2), 100–111.

Smith, S.M. (2004) HIV CTL escape: at what cost? *Retrovirology* **1**(1), 8.

Primary Literature

Abel, K., Compton, L., et al. (2003) Simian-human immunodeficiency virus SHIV89.6-induced protection against intravaginal challenge with pathogenic SIVmac239 is independent of the route of immunization and is associated with a combination of cytotoxic T-lymphocyte and alpha interferon responses, *J. Virol.* **77**(5), 3099–3118.

Ahmed, R.K., Makitalo, B., et al. (2002) Spontaneous production of RANTES and antigen-specific IFN-gamma production in macaques vaccinated with SHIV-4 correlates with protection against SIVsm challenge, *Clin. Exp. Immunol.* **129**(1), 11–18.

Akahata, W., Ido, E., et al. (2000) DNA vaccination of macaques by a full genome HIV-1 plasmid which produces noninfectious virus particles, *Virology* **275**(1), 116–124.

Allen, T.M., Mortara, L., et al. (2002) Tat-vaccinated macaques do not control simian immunodeficiency virus SIVmac239 replication, *J. Virol.* **76**(8), 4108–4112.

Almond, N., Kent, K., et al. (1995) Protection by attenuated simian immunodeficiency virus in macaques against challenge with virus-infected cells, *Lancet* **345**(8961), 1342–1344.

Almond, N., Rose, J., et al. (1997) Mechanisms of protection induced by attenuated simian immunodeficiency virus. I. Protection cannot be transferred with immune serum, *J. Gen. Virol.* **78**(Pt 8), 1919–1922.

Amara, R.R., Smith, J.M., et al. (2002) Critical role for Env as well as Gag-Pol in control of a simian-human immunodeficiency virus 89.6P challenge by a DNA prime/recombinant modified vaccinia virus Ankara vaccine, *J. Virol.* **76**(12), 6138–6146.

Amara, R.R., Villinger, F., et al. (2002) Different patterns of immune responses but similar control of a simian-human immunodeficiency virus 89.6P mucosal challenge by modified vaccinia virus Ankara (MVA) and DNA/MVA vaccines, *J. Virol.* **76**(15), 7625–7631.

Arthur, L.O., Bess, J.W. Jr., et al (1995) Macaques immunized with HLA-DR are protected from challenge with simian immunodeficiency virus, *J. Virol.* **69**(5), 3117–3124.

Baba, T.W., Jeong, Y.S., et al. (1995) Pathogenicity of live, attenuated SIV after mucosal infection of neonatal macaques, *Science* **267**(5205), 1820–1825.

Baba, T.W., Liska, V., et al. (1999) Live attenuated, multiply deleted simian immunodeficiency virus causes AIDS in infant and adult macaques, *Nat. Med.* **5**(2), 194–203.

Baba, T.W., Liska, V., et al. (2000) Human neutralizing monoclonal antibodies of the IgG1 subtype protect against mucosal simian-human immunodeficiency virus infection, *Nat. Med.* **6**(2), 200–206.

Barouch, D.H., Santra, S., et al. (2000) Control of viremia and prevention of clinical AIDS in rhesus monkeys by cytokine-augmented DNA vaccination, *Science* **290**(5491), 486–492.

Bertley, F.M., Kozlowski, P.A., et al. (2004) Control of simian/human immunodeficiency virus viremia and disease progression after IL-2-augmented DNA-modified vaccinia virus Ankara nasal vaccination in nonhuman primates, *J. Immunol.* **172**(6), 3745–3757.

Bogers, W.M., Bergmeier, L.A., et al. (2004) A novel HIV-CCR5 receptor vaccine strategy in the control of mucosal SIV/HIV infection, *AIDS* **18**(1), 25–36.

Buckner, C., Gines, L.G., et al. (2004) Priming B cell-mediated anti-HIV envelope responses by vaccination allows for the long-term control of infection in macaques exposed to a R5-tropic SHIV, *Virology* **320**(1), 167–180.

Buge, S.L., Ma, H.L., et al. (2003) Gp120-alum boosting of a Gag-Pol-Env DNA/MVA AIDS vaccine: poorer control of a pathogenic viral challenge, *AIDS Res. Hum. Retroviruses* **19**(10), 891–900.

Cafaro, A., Caputo, A., et al. (1999) Control of SHIV-89.6P-infection of cynomolgus

monkeys by HIV-1 tat protein vaccine, *Nat. Med.* **5**(6), 643–650.

Cafaro, A., Caputo, A., et al. (2000) SHIV89.6P pathogenicity in cynomolgus monkeys and control of viral replication and disease onset by human immunodeficiency virus type 1 tat vaccine, *J. Med. Primatol.* **29**(3–4), 193–208.

Chackerian, B., Briglio, L., et al. (2004) Induction of autoantibodies to CCR5 in macaques and subsequent effects upon challenge with an R5-tropic simian/human immunodeficiency virus, *J. Virol.* **78**(8), 4037–4047.

Chan, W.L., Rodgers, A., et al. (1995) Immunization with class I human histocompatibility leukocyte antigen can protect macaques against challenge infection with SIVmac-32H, *AIDS* **9**(3), 223–228.

Clements, J.E., Montelaro, R.C., et al. (1995) Cross-protective immune responses induced in rhesus macaques by immunization with attenuated macrophage-tropic simian immunodeficiency virus, *J. Virol.* **69**(5), 2737–2744.

Connor, R.I., Montefiori, D.C., et al. (1998) Temporal analyses of virus replication, immune responses, and efficacy in rhesus macaques immunized with a live, attenuated simian immunodeficiency virus vaccine, *J. Virol.* **72**(9), 7501–7509.

Cranage, M.P., Baskerville, A., et al. (1992) Intrarectal challenge of macaques vaccinated with formalin-inactivated simian immunodeficiency virus, *Lancet* **339**(8788), 273–274.

Cranage, M.P., Whatmore, A.M., et al. (1997) Macaques infected with live attenuated SIVmac are protected against superinfection via the rectal mucosa, *Virology* **229**(1), 143–154.

Dale, C.J., Liu, X.S., et al. (2002) Chimeric human papilloma virus-simian/human immunodeficiency virus virus-like-particle vaccines: immunogenicity and protective efficacy in macaques, *Virology* **301**(1), 176–187.

Daniel, M.D., Kirchhoff, F., et al. (1992) Protective effects of a live attenuated SIV vaccine with a deletion in the nef gene, *Science* **258**(5090), 1938–1941.

Das, A.T., Zhou, X., et al. (2004) Viral evolution as a tool to improve the tetracycline-regulated gene expression system, *J. Biol. Chem.* **279**(18), 18776–18782.

Deacon, N.J., Tsykin, A., et al. (1995) Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients, *Science* **270**(5238), 988–991.

Desrosiers, R.C., Wyand, M.S., et al. (1989) Vaccine protection against simian immunodeficiency virus infection, *Proc. Natl. Acad. Sci. U.S.A.* **86**(16), 6353–6357.

Doria-Rose, N.A., Ohlen, C., et al. (2003) Multigene DNA priming-boosting vaccines protect macaques from acute CD4+-T-cell depletion after simian-human immunodeficiency virus SHIV89.6P mucosal challenge, *J. Virol.* **77**(21), 11563–11577.

Evans, D.T., Chen, L.M., et al. (2003) Mucosal priming of simian immunodeficiency virus-specific cytotoxic T-lymphocyte responses in rhesus macaques by the Salmonella type III secretion antigen delivery system, *J. Virol.* **77**(4), 2400–2409.

Ferrantelli, F., Hofmann-Lehmann, R., et al. (2003) Post-exposure prophylaxis with human monoclonal antibodies prevented SHIV89.6P infection or disease in neonatal macaques, *AIDS* **17**(3), 301–309.

Ferrantelli, F., Rasmussen, R.A., et al. (2004) Complete protection of neonatal rhesus macaques against oral exposure to pathogenic simian-human immunodeficiency virus by human anti-HIV monoclonal antibodies, *J. Infect. Dis.* **189**(12), 2167–2173.

Fuller, D.H., Rajakumar, P.A., et al. (2002) Induction of mucosal protection against primary, heterologous simian immunodeficiency virus by a DNA vaccine, *J. Virol.* **76**(7), 3309–3317.

Gardner, M., Rosenthal, A., et al. (1995) Passive immunization of rhesus macaques against SIV infection and disease, *AIDS Res. Hum. Retroviruses* **11**(7), 843–854.

Gundlach, B.R., Lewis, M.G., et al. (2000) Evidence for recombination of live, attenuated immunodeficiency virus vaccine with challenge virus to a more virulent strain, *J. Virol.* **74**(8), 3537–3542.

Hel, Z., Nacsa, J., et al. (2002) Containment of simian immunodeficiency virus infection in vaccinated macaques: correlation with the magnitude of virus-specific pre- and postchallenge CD4+ and CD8+ T cell responses, *J. Immunol.* **169**(9), 4778–4787.

Hirsch, V.M., Fuerst, T.R., et al. (1996) Patterns of viral replication correlate with outcome in simian immunodeficiency virus (SIV)-infected macaques: effect of prior immunization with a trivalent SIV vaccine in modified vaccinia virus Ankara, *J. Virol.* **70**(6), 3741–3752.

Hofmann-Lehmann, R., Rasmussen, R.A., et al. (2001) Passive immunization against oral AIDS virus transmission: an approach to prevent mother-to-infant HIV-1 transmission? *J. Med. Primatol.* **30**(4), 190–196.

Hofmann-Lehmann, R., Vlasak, J., et al. (2001) Postnatal passive immunization of neonatal macaques with a triple combination of human monoclonal antibodies against oral simian-human immunodeficiency virus challenge, *J. Virol.* **75**(16), 7470–7480.

Horton, H., Vogel, T.U., et al. (2002) Immunization of rhesus macaques with a DNA prime/modified vaccinia virus Ankara boost regimen induces broad simian immunodeficiency virus (SIV)-specific T-cell responses and reduces initial viral replication but does not prevent disease progression following challenge with pathogenic SIVmac239, *J. Virol.* **76**(14), 7187–7202.

Hovanessian, A.G., Briand, J.P., et al. (2004) The caveolin-1 binding domain of HIV-1 glycoprotein gp41 is an efficient B cell epitope vaccine candidate against virus infection., *Immunity* **21**(5), 617–627.

Izumi, Y., Ami, Y., et al. (2003) Intravenous inoculation of replication-deficient recombinant vaccinia virus DIs expressing simian immunodeficiency virus gag controls highly pathogenic simian-human immunodeficiency virus in monkeys, *J. Virol.* **77**(24), 13248–13256.

Johnson, R.P., Lifson, J.D., et al. (1999) Highly attenuated vaccine strains of simian immunodeficiency virus protect against vaginal challenge: inverse relationship of degree of protection with level of attenuation, *J. Virol.* **73**(6), 4952–4961.

Kantakamalakul, W., Cox, J., et al. (2001) Cytotoxic T lymphocyte responses to vaccinia virus antigens but not HIV-1 subtype E envelope protein seen in HIV-1 seronegative Thais, *Asian Pac. J. Allergy Immunol.* **19**(1), 17–22.

Kestler, H.W. III, Ringler, D.J., et al. (1991) Importance of the nef gene for maintenance of high virus loads and for development of AIDS, *Cell* **65**(4), 651–662.

Koopman, G., Mortier, D., et al. (2004) Vaccine protection from CD4+ T-cell loss caused by simian immunodeficiency virus (SIV) mac251 is afforded by sequential immunization with three unrelated vaccine vectors encoding multiple SIV antigens, *J. Gen. Virol.* **85**(Pt 10), 2915–2924.

Kuate, S., Stahl-Hennig, C., et al. (2003) Single-cycle immunodeficiency viruses provide strategies for uncoupling in vivo expression levels from viral replicative capacity and for mimicking live-attenuated SIV vaccines, *Virology* **313**(2), 653–662.

Langlois, A.J., Weinhold, K.J., et al. (1992) Detection of anti-human cell antibodies in sera from macaques immunized with whole inactivated virus, *AIDS Res. Hum. Retroviruses* **8**(9), 1641–1652.

Learmont, J.C., Geczy, A.F., et al. (1999) Immunologic and virologic status after 14 to 18 years of infection with an attenuated strain of HIV-1. A report from the Sydney Blood Bank Cohort, *N. Engl. J. Med.* **340**(22), 1715–1722.

Lee, D., Graham, B.S., et al. (2004) Breakthrough infections during phase 1 and 2 prime-boost HIV-1 vaccine trials with canarypox vectors (ALVAC) and booster dose of recombinant gp120 or gp160, *J. Infect. Dis.* **190**(5), 903–907.

Lehner, T., Shearer, G.M., et al. (2000) Alloimmunization as a strategy for vaccine design against HIV/AIDS, *AIDS Res. Hum. Retroviruses* **16**(4), 309–313.

Lena, P., Villinger, F., et al. (2002) Co-immunization of rhesus macaques with plasmid vectors expressing IFN-gamma, GM-CSF, and SIV antigens enhances anti-viral humoral immunity but does not affect viremia after challenge with highly pathogenic virus, *Vaccine* **20**(Suppl. 4), A69–A79.

Letvin, N.L., Huang, Y., et al. (2004) Heterologous envelope immunogens contribute to AIDS vaccine protection in rhesus monkeys, *J. Virol.* **78**(14), 7490–7497.

Lifson, J.D., Rossio, J.L., et al. (2004) Evaluation of the safety, immunogenicity, and protective efficacy of whole inactivated simian immunodeficiency virus (SIV) vaccines with conformationally and functionally intact envelope glycoproteins, *AIDS Res. Hum. Retroviruses* **20**(7), 772–787.

Ling, B., Veazey, R.S., et al. (2002) SIV(mac) pathogenesis in rhesus macaques of Chinese and Indian origin compared with primary HIV infections in humans, *AIDS* **16**(11), 1489–1496.

Lohman, B.L., McChesney, M.B., et al. (1994) A partially attenuated simian immunodeficiency virus induces host immunity that correlates

with resistance to pathogenic virus challenge, *J. Virol.* **68**(11), 7021–7029.

Makitalo, B., Lundholm, P., et al. (2004) Enhanced cellular immunity and systemic control of SHIV infection by combined parenteral and mucosal administration of a DNA prime MVA boost vaccine regimen, *J. Gen. Virol.* **85**(Pt 8), 2407–2419.

Malkevitch, N., Patterson, L.J., et al. (2003) A replication competent adenovirus 5 host range mutant-simian immunodeficiency virus (SIV) recombinant priming/subunit protein boosting vaccine regimen induces broad, persistent SIV-specific cellular immunity to dominant and subdominant epitopes in Mamu-A*01 rhesus macaques, *J. Immunol.* **170**(8), 4281–4289.

Marx, P.A., Compans, R.W., et al. (1993) Protection against vaginal SIV transmission with microencapsulated vaccine, *Science* **260**(5112), 1323–1327.

Marzio, G., Verhoef, K., et al. (2001) In vitro evolution of a highly replicating, doxycycline-dependent HIV for applications in vaccine studies, *Proc. Natl. Acad. Sci. U.S.A.* **98**(11), 6342–6347.

Mascola, J.R., Lewis, M.G., et al. (1999) Protection of macaques against pathogenic simian/human immunodeficiency virus 89.6PD by passive transfer of neutralizing antibodies, *J. Virol.* **73**(5), 4009–4018.

Mascola, J.R., Stiegler, G., et al. (2000) Protection of macaques against vaginal transmission of a pathogenic HIV-1/SIV chimeric virus by passive infusion of neutralizing antibodies, *Nat. Med.* **6**(2), 207–210.

Mascola, J.R., Lewis, M.G., et al. (2003) Cellular immunity elicited by human immunodeficiency virus type 1/simian immunodeficiency virus DNA vaccination does not augment the sterile protection afforded by passive infusion of neutralizing antibodies, *J. Virol.* **77**(19), 10348–10356.

Matano, T., Kobayashi, M., et al. (2004) Cytotoxic T lymphocyte-based control of simian immunodeficiency virus replication in a preclinical AIDS vaccine trial, *J. Exp. Med.* **199**(12), 1709–1718.

Mills, K.H., Page, M., et al. (1992) Protection against SIV infection in macaques by immunization with inactivated virus from the BK28 molecular clone, but not with BK28-derived recombinant env and gag proteins, *J. Med. Primatol.* **21**(2–3), 50–58.

Mooij, P., Nieuwenhuis, I.G., et al. (2004) Qualitative T-helper responses to multiple viral antigens correlate with vaccine-induced immunity to simian/human immunodeficiency virus infection, *J. Virol.* **78**(7), 3333–3342.

Mossman, S.P., Pierce, C.C., et al. (2004) Protective immunity to SIV challenge elicited by vaccination of macaques with multigenic DNA vaccines producing virus-like particles, *AIDS Res. Hum. Retroviruses* **20**(4), 425–434.

Mothe, B.R., Weinfurter, J., et al. (2003) Expression of the major histocompatibility complex class I molecule Mamu-A*01 is associated with control of simian immunodeficiency virus SIVmac239 replication, *J. Virol.* **77**(4), 2736–2740.

Murphey-Corb, M., Martin, L.N., et al. (1989) A formalin-inactivated whole SIV vaccine confers protection in macaques, *Science* **246**(4935), 1293–1297.

Muthumani, K., Bagarazzi, M., et al. (2003) A Gag-Pol/Env-Rev SIV239 DNA vaccine improves CD4 counts, and reduce viral loads after pathogenic intrarectal SIV(mac)251 challenge in rhesus macaques, *Vaccine* **21**(7–8), 629–637.

Mwau, M., Cebere, I., et al. (2004) A human immunodeficiency virus 1 (HIV-1) clade a vaccine in clinical trials: stimulation of HIV-specific T-cell responses by DNA and recombinant modified vaccinia virus Ankara (MVA) vaccines in humans, *J. Gen. Virol.* **85**(Pt 4), 911–919.

Negri, D.R., Baroncelli, S., et al. (2004) Protective efficacy of a multicomponent vector vaccine in cynomolgus monkeys after intrarectal simian immunodeficiency virus challenge, *J. Gen. Virol.* **85**(Pt 5), 1191–1201.

Nilsson, C., Makitalo, B., et al. (2001) Enhanced simian immunodeficiency virus-specific immune responses in macaques induced by priming with recombinant semliki forest virus and boosting with modified vaccinia virus Ankara, *Vaccine* **19**(25–26), 3526–3536.

Nilsson, C., Sutter, G., et al. (2002) Immunization with recombinant modified vaccinia virus Ankara can modify mucosal simian immunodeficiency virus infection and delay disease progression in macaques, *J. Gen. Virol.* **83**(Pt 4), 807–818.

Nitayaphan, S., Pitisuttithum, P., et al. (2004) Safety and immunogenicity of an HIV subtype B and E prime-boost vaccine combination in

HIV-negative thai adults, *J. Infect. Dis.* **190**(4), 702–706.

Otsyula, M.G., Miller, C.J., et al. (1996) Fetal or neonatal infection with attenuated simian immunodeficiency virus results in protective immunity against oral challenge with pathogenic SIVmac251, *Virology* **222**(1), 275–278.

Pal, R., Venzon, D., et al. (2002) ALVAC-SIV-gag-pol-env-based vaccination and macaque major histocompatibility complex class I (A*01) delay simian immunodeficiency virus SIVmac-induced immunodeficiency, *J. Virol.* **76**(1), 292–302.

Parren, P.W., Marx, P.A., et al. (2001) Antibody protects macaques against vaginal challenge with a pathogenic R5 simian/human immunodeficiency virus at serum levels giving complete neutralization in vitro, *J. Virol.* **75**(17), 8340–8347.

Patterson, L.J., Malkevitch, N., et al. (2003) Potent, persistent induction and modulation of cellular immune responses in rhesus macaques primed with Ad5hr-simian immunodeficiency virus (SIV) env/rev, gag, and/or nef vaccines and boosted with SIV gp120, *J. Virol.* **77**(16), 8607–8620.

Patterson, L.J., Malkevitch, N., et al. (2004) Protection against mucosal simian immunodeficiency virus SIV(mac251) challenge by using replicating adenovirus-SIV multigene vaccine priming and subunit boosting, *J. Virol.* **78**(5), 2212–2221.

Pauza, C.D., Trivedi, P., et al. (2000) Vaccination with tat toxoid attenuates disease in simian/HIV-challenged macaques, *Proc. Natl. Acad. Sci. U.S.A.* **97**(7), 3515–3519.

Putkonen, P., Thorstensson, R., et al. (1991) Prevention of HIV-2 and SIVsm infection by passive immunization in cynomolgus monkeys, *Nature* **352**(6334), 436–438.

Quinto, I., Puca, A., et al. (2004) High attenuation and immunogenicity of a simian immunodeficiency virus expressing a proteolysis-resistant inhibitor of NF-kappaB, *J. Biol. Chem.* **279**(3), 1720–1728.

Ramsburg, E., Rose, N.F., et al. (2004) Highly effective control of an AIDS virus challenge in macaques by using vesicular stomatitis virus and modified vaccinia virus Ankara vaccine vectors in a single-boost protocol, *J. Virol.* **78**(8), 3930–3940.

Santra, S., Barouch, D.H., et al. (2004) Recombinant poxvirus boosting of DNA-primed rhesus monkeys augments peak but not memory T lymphocyte responses, *Proc. Natl. Acad. Sci. U.S.A.* **101**(30), 11088–11093.

Schacker, T.W., Hughes, J.P., et al. (1998) Biological and virologic characteristics of primary HIV infection, *Ann. Intern. Med.* **128**(8), 613–620.

Sharpe, S.A., Cope, A., et al. (2004) Macaques infected long-term with attenuated simian immunodeficiency virus (SIVmac) remain resistant to wild-type challenge, despite declining cytotoxic T lymphocyte responses to an immunodominant epitope, *J. Gen. Virol.* **85**(Pt 9), 2591–2602.

Shiver, J.W., Fu, T.M., et al. (2002) Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency-virus immunity, *Nature* **415**(6869), 331–335.

Shiver, J.W., Casimiro, D.R., et al. (2003) Replication-Defective Adenovirus Vector Vaccines Attenuate SIVmac239 and SHIV89.6P Challenge Infections: Effects of Challenge Virus, MHC Class I Expression and Multiple Vaccine Antigen Expression. *10th Conference on Retroviruses and Opportunistic Infect: Abstract No. 85lb*, Boston, MA, http://www.aegis.com/.

Silvera, P., Wade-Evans, A., et al. (2001) Mechanisms of protection induced by live attenuated simian immunodeficiency virus: III. Viral interference and the role of CD8+ T-cells and beta-chemokines in the inhibition of virus infection of PBMCs in vitro, *J. Med. Primatol.* **30**(1), 1–13.

Silvera, P., Richardson, M.W., et al. (2002) Outcome of simian-human immunodeficiency virus strain 89.6p challenge following vaccination of rhesus macaques with human immunodeficiency virus tat protein, *J. Virol.* **76**(8), 3800–3809.

Smith, S.M. (2004) HIV CTL escape: at what cost? *Retrovirology* **1**(1), 8.

Smith, S.M., Markham, R.B., et al. (1996) Conditional reduction of human immunodeficiency virus type 1 replication by a gain-of-herpes simplex virus 1 thymidine kinase function, *Proc. Natl. Acad. Sci. U.S.A.* **93**(15), 7955–7960.

Smith, S.M., Holland, B., et al. (1999) Retrospective analysis of viral load and SIV antibody responses in rhesus macaques infected with pathogenic SIV: predictive value for disease progression, *AIDS Res. Hum. Retroviruses* **15**(18), 1691–1701.

Smith, S.M., Khoroshev, M., et al. (2001) Constitutively dead, conditionally live HIV-1

genomes. Ex vivo implications for a live virus vaccine, *J. Biol. Chem.* **276**(34), 32184–32190.

Smith, S.M., Pentlicky, S., et al. (2003) An in vivo replication-important function in the second coding exon of tat is constrained against mutation despite cytotoxic T lymphocyte selection, *J. Biol. Chem.* **278**(45), 44816–44825.

Stahl-Hennig, C., Dittmer, U., et al. (1996a) Attenuated SIV imparts immunity to challenge with pathogenic spleen-derived SIV but cannot prevent repair of the nef deletion, *Immunol. Lett.* **51**(1–2), 129–135.

Stahl-Hennig, C., Dittmer, U., et al. (1996b) Rapid development of vaccine protection in macaques by live-attenuated simian immunodeficiency virus, *J. Gen. Virol.* **77**(Pt 12), 2969–2981.

Stebbings, R.J., Almond, N.M., et al. (2002) Mechanisms of protection induced by attenuated simian immunodeficiency virus, *Virology* **296**(2), 338–353.

Stott, E.J. (1991) Anti-cell antibody in macaques, *Nature* **353**(6343), 393.

Takeda, A., Igarashi, H., et al. (2003) Protective efficacy of an AIDS vaccine, a single DNA priming followed by a single booster with a recombinant replication-defective Sendai virus vector, in a macaque AIDS model, *J. Virol.* **77**(17), 9710–9715.

ten Haaft, P., Almond, N., et al. (2001) Comparison of early plasma RNA loads in different macaque species and the impact of different routes of exposure on SIV/SHIV infection, *J. Med. Primatol.* **30**(4), 207–214.

Veazey, R.S., Shattock, R.J., et al. (2003) Prevention of virus transmission to macaque monkeys by a vaginally applied monoclonal antibody to HIV-1 gp120, *Nat. Med.* **9**(3), 343–346.

Verhoef, K., Marzio, G., et al. (2001) Strict control of human immunodeficiency virus type 1 replication by a genetic switch: tet for tat, *J. Virol.* **75**(2), 979–987.

Vogel, T.U., Beer, B.E., et al. (2002) Induction of anti-simian immunodeficiency virus cellular and humoral immune responses in rhesus macaques by peptide immunogens: correlation of CTL activity and reduction of cell-associated but not plasma virus load following challenge, *J. Gen. Virol.* **83**(Pt 1), 81–91.

Vogel, T.U., Reynolds, M.R., et al. (2003) Multispecific vaccine-induced mucosal cytotoxic T lymphocytes reduce acute-phase viral replication but fail in long-term control of simian immunodeficiency virus SIVmac239, *J. Virol.* **77**(24), 13348–13360.

Voss, G., Manson, K., et al. (2003) Prevention of disease induced by a partially heterologous AIDS virus in rhesus monkeys by using an adjuvanted multicomponent protein vaccine, *J. Virol.* **77**(2), 1049–1058.

Wang, S.W., Bertley, F.M., et al. (2004) An SHIV DNA/MVA rectal vaccination in macaques provides systemic and mucosal virus-specific responses and protection against AIDS, *AIDS Res. Hum. Retroviruses* **20**(8), 846–859.

Warren, J.T., Dolatshahi, M. (1993) First updated and revised survey of worldwide HIV and SIV vaccine challenge studies in nonhuman primates: progress in first and second order studies, *J. Med. Primatol.* **22**(2–3), 203–235.

Whatmore, A.M., Cook, N., et al. (1995) Repair and evolution of nef in vivo modulates simian immunodeficiency virus virulence, *J. Virol.* **69**(8), 5117–5123.

Willey, R.L., Byrum, R., et al. (2003) Control of viremia and prevention of simian-human immunodeficiency virus-induced disease in rhesus macaques immunized with recombinant vaccinia viruses plus inactivated simian immunodeficiency virus and human immunodeficiency virus type 1 particles, *J. Virol.* **77**(2), 1163–1174.

Wolff, J.A., Malone, R.W., et al. (1990) Direct gene transfer into mouse muscle in vivo, *Science* **247**(4949 Pt 1), 1465–1468.

Wyand, M.S., Manson, K.H., et al. (1996) Vaccine protection by a triple deletion mutant of simian immunodeficiency virus, *J. Virol.* **70**(6), 3724–3733.

Wyatt, R., Kwong, P.D., et al. (1998) The antigenic structure of the HIV gp120 envelope glycoprotein, *Nature* **393**(6686), 705–711.

Zhao, J., Pinczewski, J., et al. (2003) Improved protection of rhesus macaques against intrarectal simian immunodeficiency virus SIV(mac251) challenge by a replication-competent Ad5hr-SIVenv/rev and Ad5hr-SIVgag recombinant priming/gp120 boosting regimen, *J. Virol.* **77**(15), 8354–8365.

Vaccine Biotechnology: *see* Genetic Engineering of Vaccines

Vaccine Research and Development Methodology

Jens-Peter Gregersen
Chiron Vaccines, Chiron-Behring GmbH & Co KG, Marburg, Germany

1 From the Research Concept to a Development Candidate 159

2 Vaccine Research Projects 160

3 Scientific Challenges of Vaccine R&D 161

4 Technical Aspects of Vaccine Development 165
4.1 Preclinical Development 166
4.2 Production Facilities 169
4.3 Clinical Development 169
4.4 Licensing and Registration of Vaccine Products 170

5 Economic Aspects of Vaccine Development 171
5.1 Vaccine Development Cost 171
5.2 Risks and Chances 172

6 Perspectives 173

Bibliography 175
Books and Reviews 175
Primary Literature 175

Keywords

Clinical Research and Development
Testing and evaluation of a medicinal product in clinical trials in human beings or in the veterinary target species. Clinical studies for medicinal product development are

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

subdivided into three phases, wherein Phase I denotes the earliest basic safety studies, Phase II means extended efficacy and safety studies, and Phase III encompasses large-scale trials under practical conditions. Clinical trials conducted after product licensing are often referred to as *Phase IV clinical studies*.

Preclinical Research and Development
Activities preceding the clinical testing of a medicinal product. The term is usually applied to summarize all technical work required to develop a medicinal product, including process and analytical methods development, formulation development, characterization, and safety studies.

Vaccine
A medicinal product, which is applied in order to confer protection against diseases, preferably against infectious diseases. Different types of vaccines are separated by the nature of their active ingredients or antigens as *inactivated vaccines*, consisting of inactivated (usually by chemically agents), whole microorganisms, *subunit vaccines*, consisting of one or few components of the whole microorganism, *live attenuated vaccines*, consisting of weakened or ''attenuated'' microorganisms, *conjugate vaccines*, wherein the desired antigen is linked or ''conjugated'' to a carrier that stimulates a more effective immune response to the antigen, and *toxoid vaccines*, consisting of inactivated bacterial toxins. *DNA vaccines* are using the genetic material encoding for the desired antigen rather than the antigen itself. Vaccines may contain mineral salts or more complex molecules and compositions as *adjuvants* to enhance the immune response of the vaccinated individual.

Vaccine research and vaccine development are commonly combined by the term R&D because, in practice, these two different disciplines cannot be easily separated. However, there are fundamental differences. Research is mainly motivated by, and aiming at, scientific publications, which are best achieved by new methods, inventions, and discoveries. Development is aiming at innovations, that is, new products that will have an impact on our daily life. In research, a vaccine could be an antigen or a preparation that has the potential of eliminating or inhibiting microorganisms. In order to convert this into a useful vaccine, developers must then add several other dimensions to the research product, namely, quality, safety, a specifically defined clinical efficacy, and practical utility. For a successful development not only scientific but also technological, commercial, and regulatory hurdles must be taken. Designing practical utility into a product encompasses and combines almost any aspect of the product, including a good tolerability, acceptable application schemes with only few vaccinations, a proven and perceived effectiveness, comfortable presentation forms, formulations that guarantee good stability and shelf lives, and, of course, adequate product cost.

1
From the Research Concept to a Development Candidate

Concepts for new vaccines arise from research and are based on combined scientific findings collected over many years and by various scientific institutions and disciplines. New vaccine concepts are regularly presented and proposed in large numbers by scientific publications or patent applications, but these concepts rarely result in new vaccines. After being tested in mice, most concepts slowly fade, since the original results cannot be reproduced under more practical conditions or turn out to be insufficiently effective to justify additional work. On the other hand, there are also organizational and financial aspects that represent serious hurdles. Most academic institutions and scientists simply do not have adequate resources to perform vaccine studies in specific models or even in monkeys or primates. Whereas vaccine antigen candidates can be designed and made by only one or a few individuals, studying these more intensively would normally require other specialists, specific facilities, and, of course, much more money. The initial research project now competes for scarce resources and needs very convincing data to make it to the next stage.

Scientific collaborations across institutional walls are an almost absolute prerequisite for continuing projects beyond testing in small laboratory animals and in order to proceed into a more intensive and application-oriented research phase. During this secondary research phase, promising concepts are taken up, reproduced, and improved until finally – and in only very few cases – a viable product and development concept can be put together. Almost invariably the efficacy of the candidate vaccine needs to be improved and made more reliable. For many new indications, even the tools and models must be established first, by which immunological effects or protection can be adequately measured.

For those few candidates that remain attractive after being studied in a more reliable way or in better models, it will then be important to assess the technical and economic aspects of the vaccine candidate very carefully. As these strongly depend on available facilities, general expertise, and specific experience with certain techniques needed, these aspects are normally evaluated by the developing organization during a project evaluation or "predevelopment" phase. At the end of this phase, a development concept should be available, which at least fulfills the following three criteria:

1. There should be sufficient evidence that the vaccine candidate is effective and protective in humans or in the target animal species. This normally presupposes that meaningful animal models have been established and that the vaccine has been tested successfully.
2. There should be a defined technical base or verified options by which the vaccine can be reliably and safely produced on a large scale. This includes, for example, cell culture or expression systems, purification schemes, and formulations that are qualified for the production of pharmaceutical products and do not contain hazardous components that cannot be removed during later process steps.
3. The expected product cost and the resulting sales prices must be in balance with the envisaged benefit of using the vaccine and expected revenues should be able to recoup the development cost in a reasonable period of time.

Thus, there should be rather clear ideas as to how the vaccine is to be manufactured and how it is characterized in its main qualities. If this base is not yet known or is based only on assumptions, a targeted product development in its strict sense is not possible, as neither the way to go nor the target or end result are known. In this case, the project should still be considered to be a research project. However, particularly in the case of vaccines, development projects are frequently started with many uncertainties, assumptions, and compromises, as vaccines are highly complex compositions, which cannot be characterized entirely and completely by analytical methods. Vaccines are products that are defined to a great deal by the process by which they are made, by the analytical tools by which they are tested, and even by the facilities in which they are manufactured. As a consequence, most vaccine development projects have no clear starting point and research and process development activities run in parallel. Although partly impossible, this should be avoided as far as possible, as development activities need many more people and are considerably more expensive than research. No developing organization has sufficient resources to run numerous complex development projects in parallel or to change the direction of a development again and again. By defining adequate criteria and by a proper project organization, critical aspects of a development project can be identified early, so that these are evaluated during the applied research phase prior to the onset of product development.

2 Vaccine Research Projects

An excellent overview of ongoing research activities for vaccines is provided by the Jordan Reports issued by the US National Institutes of Health. According to the latest issues of these reports, the number of vaccine R&D projects in the United States in the year 2000 and 2002 amounted to more than 500 projects. Almost one-third of these were various efforts to develop vaccines against AIDS. A list of the main target indications pursued by recent vaccine research and development efforts is given in Table 1.

The top positions of the vaccine research "hit list" have not changed very much over the past few decades. Well-known viral and

Tab. 1 Main infectious agents or targets for new vaccines in advanced R&D.

Viruses	*Bacteria*	*Parasites*	*Tumors*
HIV/AIDS	*Streptococcus*	Malaria	B-cell lymphomas
Hepatitis C virus	*Helicobacter pylori*	*Leishmania*	Melanomas
Herpes simplex viruses	*Borrelia*/Lyme disease	*Schistosoma*	Prostata carcinoma
Cytomegalovirus	*Salmonella*	*Toxoplasma*	CEA-tumors
HRSV	Enterotoxigenic *E. coli*	*Trypanosoma*	
Parainfluenza	*Shigella*		
Rotavirus			

Notes: HIV/AIDS: human immunodeficiency virus/acquired immunodeficiency syndrome; HRSV: human respiratory syncytialvirus; CEA: carcino-embryonal antigen: an antigen that is frequently found on colorectal, bronchial, and breast cancer cells.

bacterial infections continue to occupy the most prominent positions. However, the number of individual projects for many of these vaccine indications has increased considerably. The simple reason for this is that formerly complete microorganisms or subfractions thereof, but rarely purified single antigens, had to be used as vaccine candidates. Modern molecular biology and recombinant techniques result in individual antigens or even smaller immune response stimulating units, such as peptide epitopes, which may be varied or combined by almost endless options. Of course, this increases the number of candidates significantly and offers many new chances and possibilities, but it does not necessarily increase the chances of success for each individual approach. Molecular biology has opened up various new possibilities to approach antiparasite vaccines and tumor vaccines but also, in these particularly complex fields, the number of projects dealing with conventional "whole" organisms or cells is quite remarkable. Antitumor vaccine projects indicate that vaccines should no longer be regarded only as infection prophylaxis. Immunizations can and will in future also be used as therapeutic measures. Vaccine research even covers approaches that attempt to induce temporal infertility by the induction of antihormonal antibodies.

In comparison to current vaccine R&D projects, the number of newly licensed vaccines is extremely small. Most newly licensed products are improvements or combinations of existing vaccines; real vaccine novelties are very rare. Thus, the chances that a vaccine project in advanced research finally ends up as a vaccine product is minimal and is certainly far below 1%. These low success rates in research inevitably lead to long research phases. Short time intervals of around five years between the first publication or patent application of a new vaccine concept and the start of full development are an extremely short, applied research phase for vaccines. These may be applicable to some veterinary vaccines, for which vaccine protection of a candidate vaccine can be measured directly in the target species. For vaccines against human diseases, 10 or more years appear to be a more realistic average estimate for this phase. If one adds those further 10 to 12 years that it takes on average to develop a vaccine product, one must assume that after the basic concept has been published or patented for the first time, about 20 years are needed to successfully develop a new vaccine product. This is not different from the average time interval of two decades between concept and first appearance on the market for various innovative technical products developed during the past 100 years (including e.g. not only complex products, such as antibiotics, the pacemaker, and radar but also presumably simple products such as the zipper, dry soup mixes, powdered coffee, ballpoint pens, and liquid shampoo). At their times, all innovations had to overcome serious hurdles, such as scientific challenges, technical difficulties, and usually financial limitations also.

3
Scientific Challenges of Vaccine R&D

Science and technologies are the driving forces that enable us to develop new vaccines. Regarding the basic technologies, there are few discoveries to be named that had a significant positive influence and resulted in new vaccines. Cultivation of pure bacterial cultures is still the

fundamental base for most bacterial vaccines. A remarkable breakthrough came with the invention and development of cell culture techniques in the 1950s, which led to several new or significantly improved antiviral vaccines, including the current state-of-the-art vaccines against poliomyelitis, mumps, measles, rubella, and cell culture rabies vaccines. Current influenza vaccines are still all based on embryonated egg-derived virus antigen, but cell culture technologies are now also applied to develope new influenza vaccines. Compared to these fundamental bacterial and viral propagation technologies, molecular biology and recombinant techniques up to now had a rather limited success with essentially only one licensed recombinant human vaccine for hepatitis B. DNA vaccines may be regarded as yet another new and basic technology for new vaccines, but only a decade after their discovery, they certainly did not yet have enough time to mature to practical applications. Monoclonal antibodies or anti-idiotype antibodies, however, did not lead to new vaccines as expected, although these basic techniques were initially quoted as a major breakthrough in vaccine research.

Apart from a few essential technologies, continuous research in virology, microbiology, parasitology, and immunology are the foundations for vaccine research. However, even the most detailed knowledge about cytokines and their regulation of immune responses, or of fundamental genetic mechanisms controlling the growth and replication of microorganisms cannot be expected to bring any direct or immediate success. For the past and for the foreseeable future, it seems that it is more the pragmatic, application-oriented research that primarily fosters vaccine development. Complex immunological hurdles must be overcome in order to arrive at a new vaccine target, and that is mainly done by establishing suitable animal models and by testing all sorts of vaccine candidate antigens in these models in a very pragmatic way for their protective effects.

Recent efforts to develop a vaccine against AIDS serve as a good example of illustrating the importance of suitable models for vaccine development. In the United States in the year 2000, 135 different AIDS vaccine projects were known. Only 10% of these were considered to be basic research and development (R&D) projects, that is, they are mainly in a phase of selecting, constructing, and making the desired antigen. The rest of all these projects were allocated to preclinical testing phases in animals or to clinical testing in humans (compare Fig. 1). Less than one-third of these projects seemed to have passed small animal testing successfully and appeared to be worth testing in monkeys. Only 4.4% of the antigen candidates proceeded to trials in chimpanzees. A substantial proportion of 44% of vaccine candidates was tested in humans for safety and efficacy; however, only 1.5% were already in Phase III clinical trials, indicating that these two different vaccine candidates appear interesting enough to go into widespread field testing for efficacy. The low number of projects in basic R&D shows that after two decades of AIDS research, there are not too many new antigens or entirely new approaches to be discovered. In the absence of reliable animal models, the relatively high number of projects in early human clinical trials and the low number in later stage clinical trials very clearly demonstrate that in this case research is essentially performed in human clinical trials – with all the inherent limitations. Consequently, the

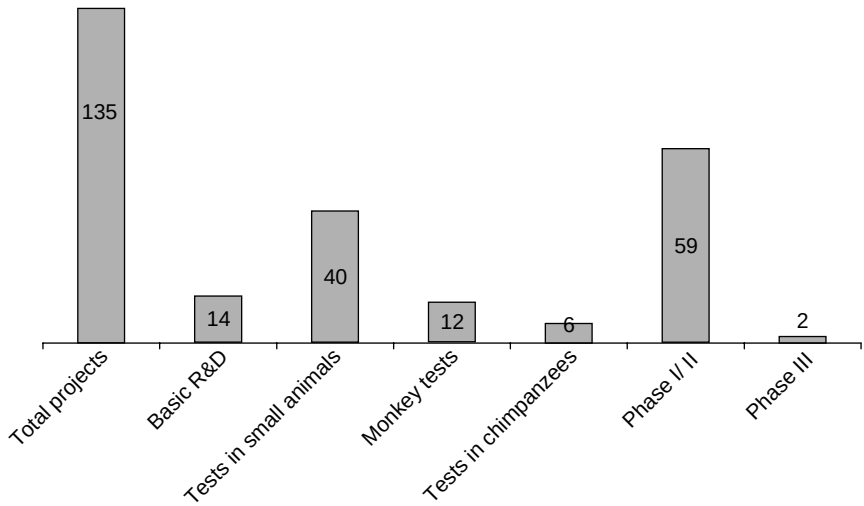

Fig. 1 AIDS vaccine candidates in research and development. The numbers apply to projects identified in the United States in year 2000.

chances of success are low, while at the same time the cost of such research is extremely high.

What are the scientific challenges and difficulties to be overcome on the way toward an effective AIDS vaccine? As summarized in Table 2, infectious microorganisms and parasites have developed various mechanisms by which they effectively prevent their elimination by the host's immune response. All of these negative attributes have been found to be associated with HIV infections. HIV does not only evade the immune responses by presenting itself by different subtypes, by varying its main immunogenic antigens during the protracted course of infection in an infected individual, or by hiding itself in a nonaccessible form by integrating its genome into host cell's genes, it even interferes actively with several important immune functions and modifies these for its own benefit and support. Of particular relevance is the selective preference of HIV for CD4 immune cells, as disturbance of their function can result in numerous deleterious effects. The ability of HIV to persist and replicate in macrophages enables HIV to convert the migrating immune cells into an efficient vehicle across normal barriers. HIV is not only insufficiently neutralized by antibodies, it even uses bound antibodies to get access to immune cells, such as macrophages, which carry receptors for the Fc fragment of antibodies.

Whereas AIDS and HIV was only chosen as examples that contributes any imaginable difficulty to vaccine development, Table 2 also lists many other current vaccine projects and their specific difficulties. A limited number of different serotypes may still be overcome by making and combining several similar vaccines, once a successful vaccine against one of these has been accomplished. Thus, vaccines against parainfluenza infections appear reasonably feasible. Other indications, such as malaria, herpesvirus infections or Lyme disease/Borreliosis, however, represent

Tab. 2 Immunological challenges on the way toward new vaccines.

Attribute	***Examples***	
Different serotypes or subtypes to be covered by the vaccine	Parainfluenza:	3 major pathogenic subtypes
	Dengue:	4 subtypes
	Malaria:	4 major pathogenic plasmodium species
	Borreliosis:	4 genetic and immunological types
	Hepatitis C:	6 major genotypes and >100 subtypes known
	HIV/AIDS:	>10 subtypes known
Antigenic variation of major immunogens	Malaria:	High variance of major antigens within the parasite
	HIV:	Antigens vary during the course of infection even within the same patient
	Trypanosoma:	Periodic switching of major surface glycoproteins
Genetic restrictions of immune recognition and immune responses	Malaria:	Multiple HLA restrictions for recognition of Plasmodium falciparum CTL epitopes even within the same individual
	HIV:	HLA-restricted CTL escape mutations associated with viral load and disease progression
Microorganism not accessible to immune responses	HIV:	Virus genetically integrated in host cell genomes
	Herpesviruses:	Virus persists in a latent state in neuronal cells
Microorganism persists in immune cells and may spread with these into tissues or across blood-brain barrier	HIV:	Persistence and active replication in, for example, macrophages
	Herpesviruses:	Can infect endothelial cells and macrophages
	Borreliosis:	Borrelia survive in macrophages. Complement membrane complexes and macrophages in the endoneutrium of Lyme neuroborreliosis
	Hepatitis C:	Macrophages and T cells found to be infected by HCV
Immune-enhancement and immune-mediated disease	HIV:	Antibody and Fc receptor–mediated enhancement of infection and disease
	Dengue:	Antibody-mediated enhancement of infection
	Borreliosis:	Immune-mediated neuropathology and arthritis
	Respiratory syncytial virus:	Inactivated vaccine induced high serum antibodies and aggravated disease upon infection

Notes: HIV/AIDS: human immunodeficiency virus/acquired immunodeficiency syndrome; HLA: human leukocyte antigens; CTL: cytotoxic T-lymphocytes; HCV: hepatitis C virus.

quite significant scientific immunological challenges, because the responsible microorganisms combine many unfavorable immunological characteristics. Finally, the example of a respiratory syncytial virus (RSV) vaccine developed and tested in the late 1960s may serve as an example to illustrate the difficulties and practical effects that some of these imponderable aspects can have. This RSV vaccine turned out to enhance a later disease, rather than preventing it. More than 30 years after those results were published, there is still no real explanation for the underlying mechanism and almost all further efforts to develop a new vaccine were stuck in a preclinical phase.

Another important aspect, which seems to be underrated in many current vaccine research projects, is the fact that most vaccines are not sufficiently effective if these are based on only single antigens. Controlled vaccine studies performed under ideal conditions in genetically homogeneous or inbred animals quite often lead to the false assumption that a fully protective vaccine antigen has been identified.

Tab. 3 Protectivity of different forms of vaccine antigens.

Type of antigen	*Amount of antigen [μg]*	*Vaccinations*
Purified whole virus particles	1	1
Virion subunits (12 S)	10	2
Virus protein 1	200	3
Oligopeptide (N-142-160-C)	200	1

Notes: Type of antigen, amount of antigen, and frequency of immunization required to achieve protection against foot-and-mouth-disease virus infection in the guinea pig model.

However, when the same vaccine is then studied under more practical conditions by fewer numbers of immunizations, in the presence of acceptable and better-tolerated adjuvants, it becomes evident that the selected antigen candidate alone is simply not effective enough. Table 3 summarizes the experiences made with different foot-and-mouth-disease (FMD) experimental vaccines. Results from model studies with this type of vaccine can be correlated reasonably well with protective response in the target species. The FMD virus consists only of three structural proteins and the most relevant virus-neutralizing antigenic epitopes are known to be located on virus protein 1. Thus, FMD vaccines appeared as an excellent target for new vaccines based on recombinant technologies. The standard vaccine, made of inactivated whole virus particles, required a relatively low amount of antigen and only one immunization in order to confer protection. Efforts to make smaller subunit or single-protein vaccines resulted only in a similar protection if several immunizations and/or massively increased antigen doses were given. These results had been established rather early in the molecular biology vaccine era, but, despite intensive further research, a commercially viable recombinant FMD vaccine has never been achieved. Meanwhile, conventional whole virus vaccines were successful enough to allow for measures to eradicate the disease completely in those countries where the vaccine has been used intensively.

4 Technical Aspects of Vaccine Development

In an ideal situation, vaccine development commences with a proven, protective, and well-defined antigenic composition.

Successful vaccine development then normally takes 10 or more years, but only a small proportion of all development candidates finally end up as a licensed product; the vast majority remains stuck in early development phases or is abandoned. Figure 2 summarizes the essential tasks of a vaccine development project and may give a rough impression of what is to be expected. For the sake of clarity, several time dependencies and overlaps have been neglected in this graphic overview.

An extensive range of national and international rules and guidelines exist, covering almost any aspect of pharmaceutical and vaccine development and registration. These guidelines describe standards that are not binding in a legal sense, but adherence to these is strongly recommended, as during later registration and licensing, the product will be judged by the same rules. Deviations from guideline recommendations may be inevitable for certain aspects and particularly for vaccines, but these should only be considered if convincing reasons for doing so can be presented. A summary of relevant guideline's requirements along with specific interpretations and applications for biopharmaceuticals and vaccines is quoted under Books and Reviews and may be consulted by prospective developers in order to get a reasonable understanding of the guiding principles.

4.1 Preclinical Development

Preclinical development comprises the technical and scientific elaboration of a process to manufacture the desired product on a large scale. Firstly, cell cultures and microorganisms to be used must be established as Master Cell Banks and Working Cell Banks, or as Master and Working Seeds, respectively. These ensure a constant supply of well-characterized, uniform, biological starting materials. Numerous tests *in vitro* and *in vivo* are required to guarantee the absence of undesired adventitious agents and to confirm the identity of these cell banks and microbial seeds.

Starting with a single aliquot of the Working Cell Bank and/or the Working Seed material, a process is then established and brought up to a final scale. The term "upstream process" typically means a cell culture or fermentation process up to some 100 L, but, for very common vaccines, larger scales may be chosen. Downstream processing summarizes activities during the purification process and typically includes recovery and concentration steps, followed by a secondary purification or "polishing" to remove specific impurities and process-related impurities introduced during earlier steps. Inactivation of bacteria or viruses or detoxifying steps for toxoid vaccines is usually included after an initial concentration step.

Formulation development includes the design of adequately buffered and well-tolerated, stable formulations, adjuvantation, the development of specific application forms, combination of vaccines into compatible vaccines, and, particularly for live attenuated vaccines, the development of a lyophilization process. Formulation development also extends to the selection or design of final syringes or other presentation forms and to filling and packing processes. Stability monitoring programs for intermediate and final products are of adamant importance for any development work and should be started as early as possible to avoid difficulties at a late development stage.

Analytical development encompasses all activities to design and use adequate methods to control and specify all parts of

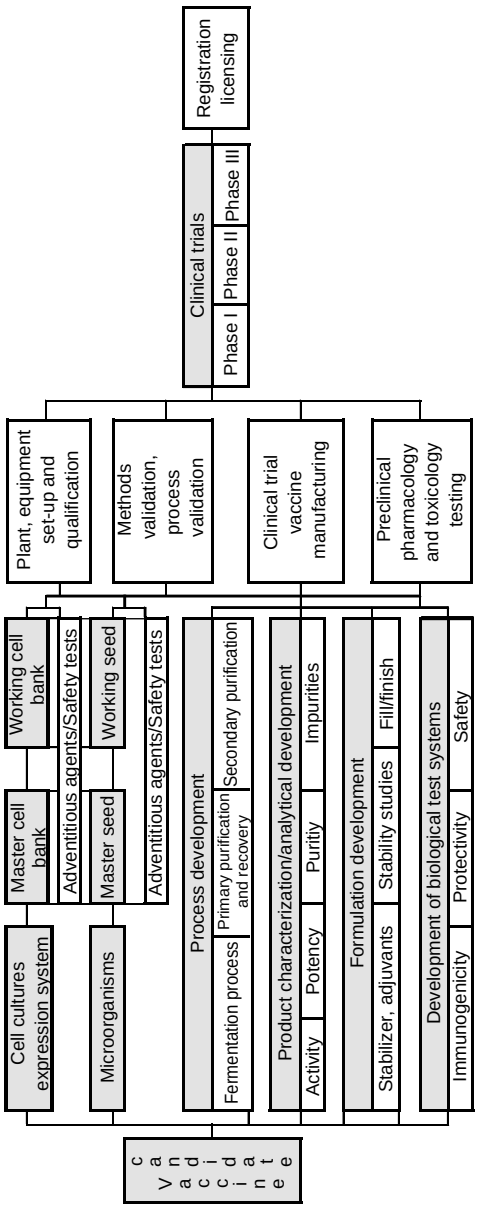

Fig. 2 Schematic diagram of a vaccine development process.

the process and the product. This includes testing of starting materials, intermediate products, and the final product, for example, for identity, specific activity, conformation, purity, and impurities. For a vaccine developed according to today's standards, a range of about 100 different tests and methods will be required. Most of these tests need to be validated for their specific purpose in order to assess the methods specificity, sensitivity, statistical exactness, and its limitations.

In parallel with process development, biological tests, and model systems must be available to monitor the vaccine's potency, immunogenicity, or protectivity at any stage, as vaccines are particularly labile products and minor modifications of the process can have a significant – mainly negative – influence on the vaccine antigen. Likewise, biological models must be at hand to study the vaccine's basic pharmacological, immunological, toxicological, and potential immunotoxicological characteristics. As far as this can be adequately studied, these include dose responses, characterization of induced humoral and cellular immune responses or of their major contributing protective mechanisms, longevity of immune responses, and potential immunological side effects. Although vaccines rarely present severe tolerability or toxicological risks, abbreviated classic toxicological testing is mandatory before the onset of studies in humans. Most vaccines need to be tested only in local and systemic tolerance studies and in repeated dose studies in standard toxicology models, but for new adjuvants and certain new excipients, representing a significant part of the vaccine composition, even a complete toxicology program, including two years of carcinogenicity studies, may be needed. Further toxicology and safety studies addressing specific risks, such as embryonal, fetal, or peri- and neonatal toxicity may be required for certain vaccines and applications or if risks are expected or known. The recent withdrawal of a newly licensed Rotavirus vaccine that was suspected to cause intussusceptions and fatal bowl obstructions in vaccinated children may serve as an example that such studies may be required not on entirely hypothetical grounds alone. In this particular case, however, the true reason for the fatalities could also be a mere coincidence and the higher medicinal attention and reporting of fatality cases in vaccinated individuals.

Owing to the biological origin of many starting materials, risks associated with prions and potentially contaminating viruses must be addressed. Organizational measures are to be put in place to avoid risk materials in addition to testing for adventitious agents. Potential risks by starting materials or process contaminants can further be evaluated and assessed by model studies with various viral and microbial agents. If specific risks are identified and if safety margins appear low, specific countermeasures are to be included into the process. As far as possible, within the technical limitations of these safety studies, a residual risk of less than 1 in 1 000 000 cases should be aimed for. In practical terms this means, for example, that an unnoticed contaminating virus is inactivated or eliminated by the process to a degree that no active virus would be found in a vaccine volume equivalent to 1 000 000 doses. For live attenuated vaccines, viral safety must also be assured by assessing the genetic and phenotypic stability of the vaccine virus and by evaluating the chances and consequences of transmissions of the virus to unvaccinated individuals.

4.2 Production Facilities

Facilities and equipment for the manufacturing of a vaccine are an immanent part of the registration dossier for the product. Any major change would have to be approved by the regulating authorities. Thus, at least for the later clinical phases, the product should be made in a specific plant and with dedicated equipment. For a development project, this means that after the process has been defined, large investments into buildings, facilities, and equipment are to be expected. Owing to the inherent risks of these investments, pilot plants should be available to produce initial clinical trial vaccine lots on an intermediate scale. A developing organization may even choose to go into Phase III clinical trials with a vaccine that has been produced in a pilot plant and to seek registration for this "preliminary" product. This approach delays the investment decision to a later point of time when all development risks have been abolished, but inevitably requires new clinical trials for the vaccine that is later on made in the final plant and extends the time to the market by several years. The sum to be invested greatly depends upon the scale of operation and dosage volume of the vaccine. For a complete vaccine plant including all auxiliary functions, the total investment may well accumulate to far above € or US$100 million. Vaccine producers who can use their existing infrastructure, such as buildings, filling and packaging facilities, raw material and media production areas, quality control laboratories, and so on, would have to invest significantly less. For small or start-up companies, outsourcing and outlicensing may be chosen to reduce risky capital investments, as only vaccines with high market expectations justify establishing a complete, own manufacturing operation.

4.3 Clinical Development

The clinical development of a new vaccine is done in three phases and normally lasts three to seven years. The duration mainly depends not only on the novelty and complexity of the vaccine indication to be explored but also on the availability of measurable immunological surrogate markers of protection. If the vaccine's efficacy must be evaluated by comparing randomly occurring cases of the disease in test groups and in alternatively treated control groups, clinical studies can be extremely long lasting, demanding, and risky.

Prerequisites of all clinical trials are adequate preclinical pharmacological and toxicological safety assessments, including animal studies, to justify tests in humans. On the basis of the available safety data and documentation, approval for clinical trials must be obtained by the relevant ethics committees and health authorities. Trials will only be admitted if these are conducted according to preestablished, systematic, and written procedures for the organization and conduct of the trials, for data collection, documentation, and statistical verification of the trial results. The "informed consent" of all participating trial subjects and medical personnel is essential. For trials involving children or mentally handicapped persons, the informed consent must be given by parents or by the person responsible. Clinical trials are to be planned and conducted according to "good clinical practice" standards that require controlled and randomized trials where possible. Control groups are to be treated by established products or treatments. Placebo treatments are

only admitted where no alternative treatment exists.

During the initial Phase I, the basic safety features of the vaccine candidate are intensively studied in a limited number (<100) of patients or healthy volunteers. The main purpose of these studies is to confirm the vaccine's local and general tolerance before it is applied in further clinical trial subjects, but Phase I vaccine studies can partly be used for a first dose finding, and immunological evaluations for adequate immune responses. During Phase I trials, vaccines rarely fail due to safety concerns, but quite frequently due to insufficient or inconsistent immune responses below expected levels.

Phase II clinical studies usually comprise no more than several hundred subjects and are normally done as controlled studies comparing the test vaccine along with an alternative prophylactic or therapeutic treatment. Clinical evaluations are mainly addressing the vaccine's effectiveness and safety, doses, application schemes, and possibly also different target groups selected by age, specific risks, countries, or by epidemiological criteria.

Phase III clinical studies are expanded, controlled, or uncontrolled trials on efficacy and safety in various clinical settings and under practical conditions. Altogether several hundred to several thousand trial subjects are enrolled at various trial sites, which are often distributed over several countries in order to study different epidemiological situations, ethnic populations, and deviating local medical practices. Phase III studies can also be evaluated for risk–benefit relationships and address practicability aspects as well as interactions by other products or concomitantly applied medical treatments. Postmarketing clinical trials of the licensed product, often referred to as *Phase IV clinical trials*, are, nowadays, rather often, also requested as part of a conditioned licensing of pharmaceutical products, mainly in order to specifically investigate those aspects that can only be assessed by large statistical cohorts.

For live attenuated vaccines, specific safety aspects must also be studied clinically. As live viruses or bacteria replicate in the vaccine and may be shed into the environment, the potential transmission of vaccine microorganisms to unvaccinated subjects must be studied. If transmission is possible or likely, the vaccine's genotypic and phenotypic stability must be carefully studied and confirmed.

4.4 Licensing and Registration of Vaccine Products

On the basis of previous experience and evaluations, the process of getting a vaccine through the evaluation at different national licensing authorities on average takes about two years, which includes time periods for working off and answering questions not adequately covered by the registration dossier. As vaccines and other biological pharmaceuticals are particularly complex compositions that cannot be adequately characterized by specific quality control methods, the entire process, manufacturing facilities, analytical methods used to specify the product and its starting materials, and all ingredients are considered as being an inherent characterizing part of the product. Any change to these affects the product's license and requires approval by the licensing authority. Changing essential elements, such as production cell substrates or microbial strains, critical test methods such as potency assays, purification methods, or

formulations would almost inevitably be seen as a change to the product that needs to be verified by new clinical trials. Furthermore, each individual batch produced must be approved and released by the authorities.

5 Economic Aspects of Vaccine Development

Without any doubt, the development of vaccines is a very costly and long-lasting process that bears a significant risk of failure. The following paragraphs intend to provide some deeper insights into the specific risks and chances, cost, and time requirements to develop a new vaccine, as the knowledge of those basics drawn from experience may by helpful in decision making. After all, successful vaccine development depends not only on good science and technical methods but also to a great extent on adequate management decisions.

5.1 Vaccine Development Cost

The number of successful vaccine projects is fairly low and retrospective evaluations of the specific cost incurred by these development projects over a time period of 10 or more years are difficult. However, cost evaluations covering developments from the late 1960s to the early 1990s exist, which summarize the development cost of various pharmaceutical developments. Although chemical drug products dominated these figures, several vaccine projects were also assessed. With all the inherent variability, we can reasonably assume that these figures also give adequate estimates for vaccine products. These evaluations show that pharmaceutical development cost during those years tended to increase by a factor of about 10 within a decade. As demonstrated by a simple graph (Fig. 3), the rising cost is only in part due to the normal inflation rate, but clearly correlates with the increasing regulatory demands,

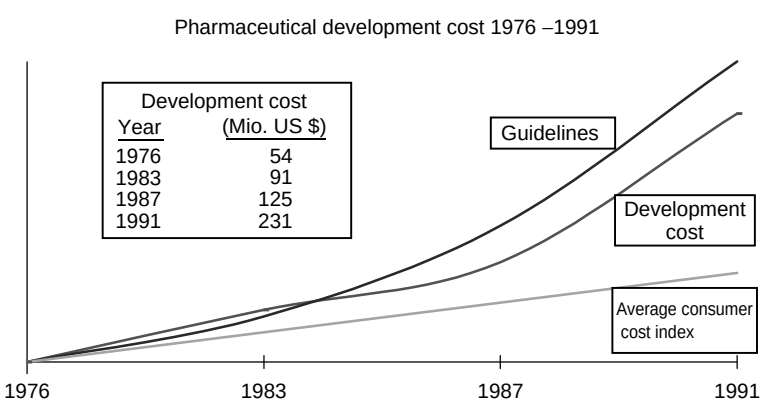

Fig. 3 Pharmaceutical development cost. Development cost denoted for various mainly chemical pharmaceuticals, including vaccine projects. Correlation with rising regulatory requirements as indicated by the counts of existing guidelines at the indicated point of time. Average consumer cost index for a household of four persons taken from figures released by the Feral German Statistics Office (Statistisches Bundesamt). Graphs show relative figures adjusted to a uniform scale.

as exemplified by the number of applicable guidelines.

The latest well-founded figure of US$231 million, published in 1991, was based on evaluations of 93 successful product developments. This sum has since been quoted on many occasions and has partly been willfully and generously projected to later dates. Thus, quoted sums of US$500 to 800 million may be encountered to describe the cost and risks of pharmaceutical development projects. However, these figures are misleading if several important details about the original calculations are not mentioned: those 231 millions include to a great extent, opportunity cost (calculated by an interest rate of 9% of the invested capital over a period of 12 years) and the cost of many unsuccessful or abandoned projects (assuming a success rate of 23%), furthermore, tax credits were not accounted for. All in all, the underlying ex-pocket expenses must be assumed to be about one-fourth of the total sum, and, according to today's standards, direct cost between € or US$ 60 and 100 million may be assumed as a realistic estimate for the development of an average new vaccine. If, however, there is no suitable infrastructure and if investments into completely new production facilities are to be made, this could easily double the cost.

Apart from capital investments, personnel is the most relevant cost factor to be considered. Owing to the high number of persons involved in the preclinical and technical development and to the long duration of these activities (on average about four years until the start of clinical trials and several years beyond until registration), preclinical and technical activities account for about one half or more of the development cost. Clinical development normally causes about one-fourth of the development cost; the remaining quarter is evenly spread throughout the developing organization and covers overheads, technical support functions, quality control, and quality assurance, as well as various other specialists, for example, for quality assurance, regulatory affairs, patenting, and market research.

Taken together, the average vaccine development project requires about 170 man-years of work with average total expenditures per person and workplace in the pharmaceutical industry being around €180 000 to 200 000 or for the United States around US$220 000 to 240 000. This results in roughly € or US$30 to 40 million for personnel and workplace expenses. External cost of around 20% and highly variable capital investments into plant and facilities are then to be added.

5.2 Risks and Chances

The success of a project during and for the entire development process can be estimated by the numbers of projects that make it until the next development phase and finally end up as a commercial product. On the basis of data for products developed during the preceding decades until 1994, one must assume that only 50% of the preclinical vaccine development projects enter the clinical phase and another 50% is abandoned during the clinical trials. Having passed all preceding hurdles, product registration seemed to be uncritical, as only a loss of 1 out of 100 vaccine projects was noted. For pharmaceutical drug products, overall success rates of 11% were found, that is, 100 product candidates entering the preclinical development resulted in only 11 licensed products. Vaccine projects appeared to be more successful with an average of 22% licensed products per 100 projects (compare

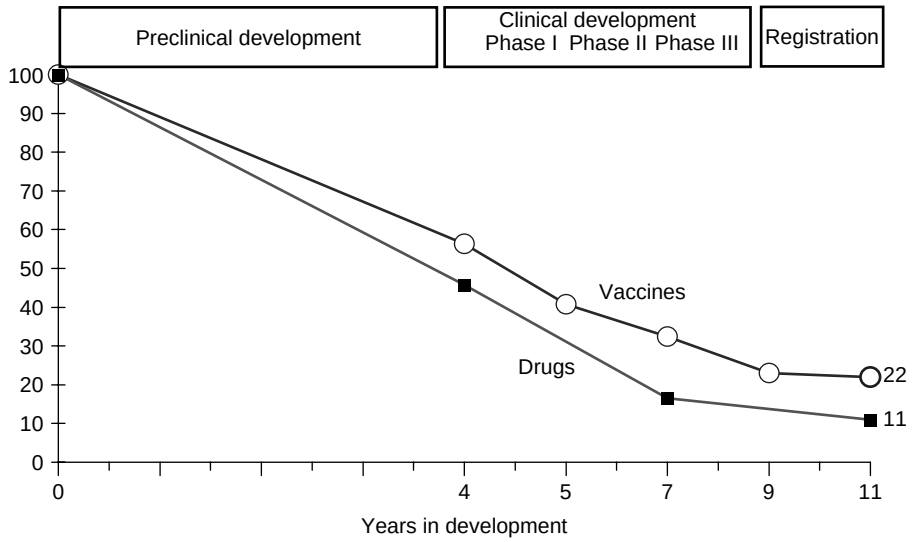

Fig. 4 Success rates of vaccine development projects in comparison with drug development. Data represent the percentage of projects, which successfully completed the respective development phase. Data summarized from assessments that mainly refer to projects of the 1980s. As discussed in the text, the success rates for vaccines developed today are most likely lower than shown here.

Fig. 4). However, the figures presented in Fig. 1 for AIDS vaccine projects of the recent past show that average figures can also be grossly misleading.

Most current vaccine candidates are dealing with quite "difficult" infectious diseases, which, under natural conditions, do not induce a lasting protective or sterile immunity, thus doubts about the applicability of those earlier risk evaluations to current vaccine projects are justified. Even if AIDS vaccine projects are not considered, a snapshot view on more recent vaccine developments supports the suspicion that success rates for today's projects and particularly for new vaccine indications are much lower. As shown in Table 4, the success rates of preclinical development in more recent times appeared to be below 50% and only 14 to 15% of all projects were found in Phase III clinical trials. Whereas preclinical projects represented a very wide spectrum of entirely new vaccines, the majority of Phase III clinical trials were covering alternatives to already existing vaccines, such as competitor's developments, combinations, or improved formulations. Less than 5% of these advanced projects were approaches to develop entirely new vaccines. These figures represent only a static view upon the vaccine development for a certain time period, but they clearly indicate that nowadays – and particularly for really new vaccines – development success rates clearly below 5% appear more realistic than earlier estimates that were above 20%.

6 Perspectives

Judged by the number of scientific publications in microbiological and biotechnology

Tab. 4 Human vaccine development projects in the years 2000 and 2002.

Project phase	*No. of vaccine projects*		
	Year 2000	*Year 2002*	*Percentage*
Preclinical development	349	331	100
Advanced to Phase I clinical development	158	154	45–47
Advanced to Phase II clinical development	102	106	29–32
Advanced to Phase III clinical development	51	47	14–15
Thereof alternatives to existing vaccines in Phase III	37	32	10–11
Thereof new vaccine indications in Phase III	14	15	4–5

Notes: Data extracted from listed vaccine projects in the year 2000 and 2002 in the United States without consideration of AIDS vaccine projects. Vaccine projects in Phase III clinical trials for new indications include vaccines against *Coccidioides immitis*, group B *streptococcus*/*Streptococcus pneumoniae*, *Plasmodium falciparum*, *Trypanosoma cruzi*, *Leishmania major*, *Mycobacterium leprae*, Meningococcus B and C, Rotavirus, *Vibrio cholerae*, and *Pytium insidiosum*.

journals, vaccine R&D appears to have a great attraction for scientists from all pertinent scientific disciplines. Whenever new methods and technologies became available, these have always and immediately triggered a huge number of new vaccine research projects and stimulated research into formerly hopeless vaccines. Along with the good reputation that vaccines enjoy, this scientific background is an excellent base for new vaccines. It also provides a favorable environment to attract the required capital that is needed to develop vaccines into marketable products. However, considering the high risks and the long duration of vaccine R&D, there must also be other reasons why investors and pharmaceutical companies invest in this field.

Vaccines represent only a small proportion of the pharmaceutical market, but, nevertheless, vaccines are extremely successful products. Firstly, vaccines effectively prevent diseases, rather than only curing these. Owing to these advantages, vaccines have often created their own markets and have even defended their market shares against competition by very effective therapeutics or antibiotics. Secondly, most vaccines are recommended by public health authorities and thus enjoy a rather safe position on the market. Furthermore, there are usually only rather limited numbers of competitive products because vaccines are far too complex to become an easy target for producers of generic imitations. And, finally, vaccines usually have a very long life span. As long as vaccine products are not neglected and become outdated, but are constantly adapted to a better state of the art, vaccines do not lose their market position, unless they are too successful and by and by eliminate the need to use the vaccine.

Thus, vaccine R&D can be very rewarding for both scientists and investors. Regarding the risks, however, the investor has a quite different perspective than the scientist. The investor may contain risks by putting capital into many different projects and enterprises, thus participating in the statistically very successful "average vaccine." To a limited extent, large companies who develop vaccines can also apply the same strategy. However, small enterprises and individual scientists working for only

one or a few R&D projects have only few options to manage and reduce risks. They often choose a high-risk approach by aiming only for "block buster" products. In this case, they must be aware that competition in this field will be also very strong, which increases the risks even more. However, within the given financial limitations, risks could also be spread over a certain number of projects in early R&D phases, preferably by approaching different indications and concentrating on an attractive new or improved technology.

Scientifically and commercially, vaccine R&D is certainly not a playing ground for those who expect fast success and revenues. Any organization that intends to invest into vaccine R&D should be prepared – both mentally and financially – to endure for at least 10 to 20 years.

See also Dendritic Cells; Human and Veterinary Classical Vaccines against Bacterial Diseases; Immunologic Memory; Immunology; Liposome Gene Transfection; Medicinal Chemistry; Mucosal Vaccination; Transcription Factor NF-κB: Function, Structure, Regulation, Pathways, and Applications.

Bibliography

Books and Reviews

Gregersen, J.P. (1994) *Research and Development of Vaccines and Pharmaceuticals from Biotechnology. A Guide to Effective Project Management, Patenting and Product Registration*, VCH Publishers, Weinheim, New York, Basel, Cambridge, Tokyo.

International Conference on Harmonization (ICH). International regulatory guidelines, harmonized between the EU, USA, and Japan, and covering various aspects are available as ICH documents under: http://www.ich.org/ich5.html

The Rules Governing Medicinal Products in the European Community, Volume I-IV (1989), Vol. III plus Addendum 1, 2, 3 (1990–1994), as well as numerous newer guidelines, which are not contained in the above mentioned volumes can be obtained from: Office des Publications Officielles des Communantés Européenes, Luxemburg. These guidelines are also available via http://www.eudra.org/Document.htm

U.S. Department of Health and Human Services: National Institutes of Health, National Institute of Allergy and Infectious Diseases (latest issue from 2002) *The Jordan Report. Accelerated Development of Vaccines 2002.* http://www.niaid.nih.gov/dmid/vaccines/jordan20/default.htm

US-FDA guidelines and Points to Consider (PCT) documents, as well as lists of all available documents about human biological medicines are available from: Congressional, Consumer and International Affairs Staff, Metro Park North, Building 3, 5600 Fishers Lane, Rockville, MD 20857 and via internet under http://www.fda.gov/cber/guidelines.htm

Primary Literature

Batelle Memorial Laboratories. (1973) *Science, Technology and Innovation*. Report to the National Science Foundation, USA.

Bienz-Tadmore, B., Dicerbo, P.A., Tadmore, G., Lasagna, L. (1992) Biopharmaceuticals and conventional drugs: Clinical success rates, *BioTechnology* **10**, 521–525.

Di Masi, J.A., Hansen, R.W., Grabowski, H.G., Lasagna, L. (1991) Cost of innovation in the pharmaceutical industry, *J. Health Economics* **10**, 107–142.

Donelson, J.E., Hill, K.L., El-Sayed, N.N. (1998) Multiple mechanisms of immune evasion by African trypanosomes, *Mol. Biochem. Parasitol.* **91**, 51–66.

Du Chateau, B.K., Munson, E.L., England, D.M., Lovrich, S.D., Callister, S.M., Jensen, J.R., Schell, R.F. (1999) Macrophages interact with enriched populations of distinct T lymphocyte

subsets for the induction of severe destructive Lyme arthritis, *J. Leukoc. Biol.* **65**, 162–170.

Ellis, D.J., Pekar, P.P. (1983) *Planning Basics for Managers*, AMACON, New York.

Goulder, P.J., Brander, C., Tang, Y., Tremblay, C., Colbert, R.A., Addo, M.M., Rosenberg, E.S., Nguyen, T., Allen, R., Trocha, A., Altfeld, M., He, S., Bunce, M., Funkhouser, R., Pelton, S.I., Burchett, S.K., McIntosh, K., Korber, B.T., Walker, B.D. (2001) Evolution and transmission of stable CTL escape mutations in HIV infection, *Nature* **412**, 334–338.

Gregersen, J.P. (1997) Vaccine Development: The Long Road from Initial Idea to Product Licensure, in: Levine, M.M., Woodrow, G.W., Kaper, J.B., Cobon, G.S. (Eds.) *New Generation Vaccines*, Chap. 7, 2nd edition, Marcel Dekker Inc., New York, Basel, Hong Kong, pp. 1165–1177.

Haase, A.T. (1986) Pathogenesis of lentivirus infections, *Nature* **322**, 130–136.

Halstead, S.B. (1979) In vivo enhancement of dengue virus infection in rhesus monkeys by passively transferred antibody, *J. Infect. Dis.* **140**, 527–533.

Hansen, R.W. (1979) The Pharmaceutical Development Process: Estimates of Development and the Effects of Proposed Regulatory Changes, in: Chien, R.J. (Ed.) *Issues in Pharmaceutical Economics*, Lexington Books, Cambridge, MA, pp. 151–186.

Homsy, J., Meyer, M., Levy, J.A. (1990) Serum enhancement of human immunodeficiency virus (HIV) infection correlates with disease in HIV-1 infected individuals, *J. Virol.* **64**, 1437–1440.

Horzinek, M.C. (1984) *Kompendium der allgemeinen Virologie*, Enke Verlag, Stuttgart, Germany.

Kim, H.W., Canchola, J.G., Brandt, C.D., Pyles, G., Chanock, R.M., Jensen, K., Parrot, R.H. (1969) Respiratory syncytialvirus disease in infants despite prior administration of antigenic inactivated vaccine, *Am. J. Epidemiol.* **89**, 422–434.

Lane, H.C., Depper, J.M., Greene, W.C., Whalen, G., Waldmann, T.A., Fauci, A.S. (1985) Quantitative analysis of immune functions in patients with acquired immunodeficiency syndrome: Evidence for a selective defect in soluble antigen recognition, *New Engl. J. Med.* **313**, 79–84.

Langle, L., Occelli, R. (1983) The cost of a new drug, *J. d'Economie Medicale* **1**, 77–106.

Le, T.P., Coonan, K.M., Hedstrom, R.C., Charoenvit, Y., Sedegah, M., Epstein, J.E., Kumar, S., Wang, R., Doolan, D.L., Maguire, J.D., Parker, S.E., Hobart, P., Norman, J., Hoffman, S.L. (2000) Safety, tolerability and humoral immune responses after intramuscular administration of a malaria DNA vaccine to healthy adult volunteers, *Vaccine* **18**, 1893–1901.

Luciw, P. (1996) Human Immunodeficiency Viruses and their Replication, in: Field, B.N., Knipe, D.M., Howley, P.M. (Eds.) *Field's Virology*, Chap. 60, Vol. 2, 3rd edition, Lippincott-Raven, Philadelphia, pp. 1881–1952.

Maimone, D., Villanova, M., Stanta, G., Bonin, S., Malandrini, A., Guazzi, G.C., Annunziata, P. (1997) Detection of Borrelia burgdorferi DNA and complement membrane attack complex deposits in the sural nerve of a patient with chronic polyneuropathy and tertiary Lyme disease, *Muscle Nerve* **20**, 969–975.

Montgomery, R.R., Nathanson, M.H., Malawista, S.E. (1993) The fate of Borrelia burgdorferi, the agent for Lyme disease, in mouse macrophages. Destruction, survival, recovery, *J. Immunol.* **150**, 909–915.

Moore, C.B., John, M., James, I.R., Christiansen, F.T., Witt, C.S., Mallal, S.A. (2002) Evidence of HIV-adaptation to HLA-restricted immune responses at a population level, *Science* **296**, 1439–1443.

Odeberg, J., Cerboni, C., Browne, H. Karre, K., Moller, E., Carbone, E., Soderberg-Naucler, C. (2002) Human cytomegalovirus (HCMV)-infected endothelial cells and macrophages are less susceptible to natural killer lysis independent of the downregulation of classical HLA class I molecules or expression of the HCMV class I homologue, UL 18, *Scand. J. Immunol.* **55**, 149–161.

Roizman, B., Sears, A.E. (1996) Herpes Simplex Viruses and their Replication, in: Field, B.N., Knipe, D.M., Howley, P.M. (Eds.) *Field's Virology*, Chap. 72, Vol. 2, 3rd edition, Lippincott-Raven, Philadelphia, pp. 2231–2296.

Rosen, S. (1976) *Future Facts*, Simon & Schuster, New York.

Smith, J.D., Chitnis, C.E., Craig, A.G., Roberts, D.J., Hudson-Taylor, D.E., Peterson, D.S., Pinches, R., Newbold, C.I., Miller, L.H. (1995)

Switches in expression of Plasmodium falciparum VAR genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes, *Cell* **82**, 101–110.

Struck, M.M. (1994) Biopharmaceutical R&D success rates and development times, *BioTechnology* **12**, 674–677.

Struck, M.M. (1996) Vaccine R&D success rates and development times, *Nat. Biotechnol.* **14**, 591–593.

Takeda, A., Tuazon, C., Ennis, F.A. (1988) Antibody-enhanced infection by HIV-1, *Science* **242**, 550–583.

Wang, R., Doolan, D.L., Le, T.P., Hedstrom, R.C., Coonan, K.M., Charoenvit, Y., Jones, T.R., Hobart, P., Margalith, M., Ng, J., Weiss, W.R., Sedegah, M., de Taisne, C., Norman, J.A., Hoffman, S.L. (1998) Induction of antigen-specific cytotoxic T lymphocytes in humans by a malaria DNA vaccine, *Science* **282**, 476–480.

Wiggins, S.N. (1987) *The Cost of Developing a New Drug*, Pharmaceutical Manufacturers' Association, Washington, DC.

Zignego, A.L., De Carli, M., Monti, M., Careccia, G., La Villa, G., Giannini, C., D'Elios, M.M., Del Prete, G., Gentilini, P. (1995) Hepatitis C virus infection of mononuclear cells from peripheral blood and liver infiltrates in chronically infected patients, *J. Med. Virol.* **47**, 58–64.

Vaccines (Classical) against bacterial disease: *see* Human and Veterinary Classical Vaccines against Bacterial Diseases

Variation and Conservation of the Development of Animal Bodies: Evolutionary Developmental Biology

Francoise Mazet[1] *and Sebastian M. Shimeld*[2]
[1] *University of Reading, Whiteknights, Reading, UK*
[2] *University of Oxford, Oxford, UK*

1 Historical Perspective and Recent Developments 181
1.1 Relationship between Ontogeny and Phylogeny in Late Nineteenth/Early Twentieth Century 181
1.2 Impact of Molecular Biological, and Other Techniques 182
1.3 Important Questions 183
1.3.1 The Microevolution of Development 183
1.3.2 Canalization and Constraint in the Evolution of Development 183
1.3.3 The Basis for Major Transitions in the Development of Body Form 183

2 Animal Phylogeny and Gene Ontology 184
2.1 The Critical Nature of Phylogenetics in Defining Taxonomic Relationships and Gene Classification 184
2.2 Identification of a Conserved Developmental Genetic Toolkit in Animals 187
2.3 Hox Genes and AP Axis: A Detailed Example from the Toolkit 188
2.4 The Evolution of this Toolkit via Duplication 189

3 Microevolution of Developmental Processes 189
3.1 Evolution via Changes in Coding or Noncoding Sequences 190
3.2 Parallelism in Morphological Evolution 191

4 Developmental Basis of Innovations 192
4.1 The Diploblast/Triploblast Divide: Evolutionary Hypotheses and Key Data 192
4.2 Gene Duplication and Vertebrate Origins 193

5 Loss of Complexity in Animal Development 193
5.1 Phylogeny Suggests Loss of Genetic Complexity is Widespread 193

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

5.2 Examples of Developmental Regulatory Changes Associated with Loss: Snakes and Sticklebacks 194

6 Future Prospects 196
6.1 The Impact of Genomics and Postgenomics 196
6.2 Understanding Regulatory Evolution: Computational and Empirical Methods United 196

Bibliography 197
Books and Reviews 197
Primary Literature 197

Keywords

Bilaterians
The bilaterally symmetrical animals, including all triploblasts, excluding the diploblasts and sponges.

Clade
A group of species linked by a common ancestral species. Also called *monophyletic group*.

Diploblasts
Animals possessing only two cell layers (the ectoderm and the endoderm). Includes the cnidarians (jellyfish) and the ctenophores (comb jellies).

Genetic toolkit
Set of genes necessary to construct an embryo. Contains mostly transcription factors and signaling molecules.

Homeobox
Conserved region of 180 nucleotides, translated in a 60 amino acid domain (the homeodomain). Homeobox genes code for transcription factors called *homeoproteins*. Examples: *Otx*, *Hox*.

Homologs
Genes sharing sequence similarity and inferred to have originally derived from an ancestral gene at an indeterminate point. Example: otx genes.

Orthologs
Genes in different species that are related to one another by speciation events. Example: *otd* in *D. melanogaster* is orthologous to *Otx1* and *Otx2* in mammals.

Paralogs
Paralogous genes are those that in different, or the same, species are related by a gene duplication event. Example: *Otx1* and *Otx2* in mice.

Promoter
Noncoding region of the DNA that lies near the start site of transcription. Contains binding sites for the transcriptional machinery.

Pseudogenes
Nonfunctional genes. A pseudogene can be transcribed.

Triploblasts
Animals possessing three cell layers: the ectoderm, the mesoderm, and the endoderm.

The body (or phenotype) of a multicellular organism has two histories. One is ontogenetic, which is the development of the adult body from a fertilized egg. The other is phylogenetic, which is the evolutionary history of the organism. These two histories are fundamentally connected, as evolutionary change to the shape and size of an organism's body occurs via change in the processes regulating embryonic development, while selection on the adult phenotype determines whether such changes persist or are eradicated. Evolutionary developmental biology seeks to understand how evolutionary change in developmental processes has occurred, and how it has led over evolutionary time to the wide diversity of body forms that exist today or have existed in the past. At the core of the subject is the study of the genes that regulate development, and how change in such genes results in evolutionary change of the phenotype.

1 Historical Perspective and Recent Developments

1.1 Relationship between Ontogeny and Phylogeny in Late Nineteenth/Early Twentieth Century

Throughout the later part of the nineteenth century and early twentieth century, the study of embryology and evolutionary biology were deeply intertwined. The comparative study of embryonic development provided authors such as Haeckel with the basis for evolutionary hypotheses, and helped to refine an emerging view of the phylogenetic relationship of living animal groups. This cozy relationship was essentially split in the 1920s and 1930s by the rise of a new biological discipline, genetics. Evolutionary biology embraced genetics, becoming increasingly quantitative in its approaches, and resulting in the birth of evolutionary and population genetics. Embryology became increasingly experimental, adopting both genetic

and manipulatory methods for perturbing embryos, and laying the foundations for the study of developmental processes that we now refer to as developmental biology.

With some notable exceptions, the two disciplines remained largely separate until the 1980s, when key technical advances opened the door to a reunited field; evolutionary developmental biology. Essentially, the technical advances relate to our ability to examine and manipulate DNA, the common bedrock on which both disciplines could be mounted (see Sect. 1.2). Developmental biologists were now able to examine the genes responsible for regulating embryonic development, and to determine the precise genetic lesions responsible for specific mutant phenotypes. A milestone was passed in 1984, with the discovery of Homeobox genes and the subsequent demonstration of the conserved role of Hox type Homeobox genes in the development of animals with widely divergent body plans. Homeobox gene conservation was soon discovered to be but the tip of an iceberg of conserved genes and pathways that regulate animal development.

The new techniques that allowed us to read gene sequences and understand the relations between genotype and phenotype were a giant step that transformed the study of evolutionary developmental biology, opening the door to detailed, rigorous examination of hypotheses often first posed over a hundred years ago.

1.2 Impact of Molecular Biological, and Other Techniques

Molecular biology techniques such as polymerase chain reaction (PCR), Southern blotting, and DNA library screening have facilitated the rapid isolation of genes. With the application of such techniques to multiple model species, it rapidly became clear that different animals had genes that share a high level of similarity in their sequences. Degenerate PCR could then be used to isolate related genes from different species on the basis of such similarities. This technique drastically accelerated the number of genes characterized in both model and nonmodel systems, freeing gene isolation from traditional genetic studies.

New techniques also allowed the visualization of gene products, enabling researchers to examine their distribution in embryos. The characterization of a gene expression pattern is performed using two main techniques to visualize the products of the gene: *in situ* hybridization for the visualization of mRNA transcripts and immunolocalization for the visualization of proteins. Both techniques can be performed either in whole fixed embryos, in dissected parts of embryos, or on sections. Expression domains can also be localized in living embryos and adults using Green Fluorescent Protein (GFP) technology.

All these techniques, however, do not give any insight into the function of a gene or the protein it encodes. This can be assessed in several ways. Historically, the most informative method has been by creating and analyzing mutants, though this is extremely time consuming and labor intensive. Of great interest is the recent emergence of antisense technologies, such as RNAi and morpholino oligonucleotides, which allow gene function to be assessed without the need for time-consuming mutant screening programs. Additional to these loss-of-function approaches, gain-of-function experiments can also be informative, and are usually undertaken by injection or electroporation of RNA or DNA constructs into embryos. Finally, genomic and postgenomic techniques for

examining genes and proteins on a large scale are playing an increasingly important role in evolutionary developmental biology. In this case, researchers can follow the impact of one change on the expression of thousands of other molecules at once, and thus better understand the real impact of each genomic modification. In the following, we address some of the key evolutionary developmental questions currently under investigation, highlighting recent advances in our understanding and suggesting how future developments in biology will impact on this field.

1.3 Important Questions

1.3.1 The Microevolution of Development

One of the central questions of evolutionary biology has been the relationship between microevolution and macroevolution, that is, how population and speciation level processes contributed to the evolution of diverse animal body plans. As discussed earlier, it has been demonstrated that genes act through development to create morphological structures, and from this, it follows that evolutionary change in morphology is linked to changes in developmental processes. To understand this, we need to know how variation in genes arises, the nature of variation in developmental processes in extant populations, and we need to be able to examine the effect of selection on this variation. These issues are addressed in Sect. 3.

1.3.2 Canalization and Constraint in the Evolution of Development

Animal evolution is constrained by physical laws, such as gravity, rates of diffusion, and rates of chemical interactions, which act to limit morphology and physiology in certain ways. An important question is whether developmental processes also provide constraints on the evolution of animal morphology. For example, all vertebrates share fundamental similarities in their body plans, despite hundreds of millions of years of separation. Does this imply that they are unable to "undo" their basic design (i.e. that certain phenotypic end points cannot be reached), or is it simply that radical changes are possible, but selection continually removes them? An answer to this question would help us determine whether canalization of developmental processes constrains the evolution of animal form. Similarly, are all processes and periods of embryonic development equally subject to evolutionary change, or is phenotypic variation more likely to arise and be subject to selection at specific times and places in development? 'Answers to this too would help us understand why body plans often appear to be highly conserved in the face of enormous finer-scale variation, as seen in highly speciose taxa such as the insects for example'? The consensus view is that development does constrain the evolution of morphology, but this is not proven and many researchers hold alternative views.

1.3.3 The Basis for Major Transitions in the Development of Body Form

Taxonomists divide the animal kingdom into some 30 phyla, each with a different body plan. Notably, the majority of such body plans appear to have evolved in a surprisingly short window in the Cambrian, the so-called *Cambrian explosion*. How did such body plan diversity evolve, and how do other major changes in development and morphology evolve? In particular, can such differences be explained by the cumulative effect of microevolutionary processes,

with the extinction of intermediates giving the illusion of major transitions, or do other factors, such as those discussed earlier in Sect. 1.3.2, affect body plan evolution? The discovery that homeotic mutations, in which one part of an animal's body can be completely replaced by a duplication of another part, such as an antenna by a leg, can be caused by changes in single genes of the Hox complex (see Sect. 2.3), was taken by many to indicate that major changes in body form could be produced by relatively small mutations. However most evolutionary biologists now dispute that such mutations and the "hopeful monsters" they produced could have contributed to the evolution of animal morphology in such a simple way. These issues are addressed here from two different standpoints, the evolution of new complex characters that define major groups of animals (innovations) in Sect. 4, and the loss of complexity in Sect. 5.

Related to this issue is the degree of underlying similarity that appears to be present in the basic developmental programs of different phyla. One example of this is the role of the *Pax6* gene in eye development in numerous phyla with quite different eye structures, for example, insects, nemertene worms, and vertebrates. At its basic level, this suggests that the common ancestor of these taxa (and by extrapolation all bilaterally symmetrical animals: see Sect. 2) had a photoreceptor whose development was regulated by *Pax6*. Similarly, while vertebrates and insects have nervous systems that are superficially very different, comparisons of neural patterning in *Drosophila* and mouse have suggested further conservation of more intricate mechanisms between widely divergent taxa (Fig. 1). These studies have demonstrated that the positioning of cell types along both the dorsoventral and anteroposterior axes of the nervous systems of these species reflects a similar underlying profile of gene expression. This suggests a degree of molecular similarity that underlies the different nervous systems of insects and vertebrates, and therefore that the common ancestor of these two animal lineages also used such mechanisms to pattern its nervous system. The evolution of the vertebrate and insect nervous systems, therefore, can be seen as elaborations built upon a complex, preexisting, molecular ground plan, indicating how such major differences might have evolved from a common ancestor.

2 Animal Phylogeny and Gene Ontology

2.1 The Critical Nature of Phylogenetics in Defining Taxonomic Relationships and Gene Classification

The "tree of life" is a pictorial representation of the relationships between living organisms. Early trees, such as that of Ernst Haeckel, tended to depict an apparent progression from what was regarded as simple organisms to those considered more complex, and ultimately, man. Formalization of the tree-building process began with the science of cladistics, founded by Willy Hennig. Numerous attempts have been made to reconstruct animal phylogeny using adult as well as embryological morphological characters, with the results differing in many details. While most agree on major relationships, such as the division of bilaterians into protostomes and

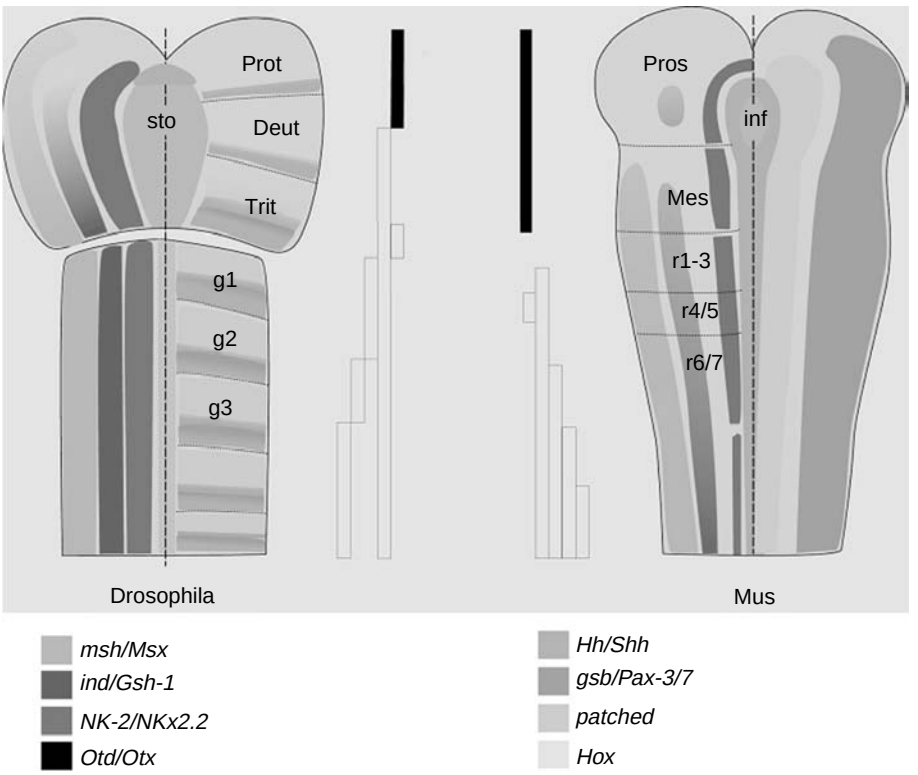

Fig. 1 Comparisons of gene expression patterns in the anterior central nervous system (CNS) of *Drosophila* and Mouse embryos. The expression of *msh/Msx* (blue), *ind/Gsh-1* (brown), *NK-2/Nkx-2.2* (red), *Otd/Otx* (black bars) and *Hox* orthologs (white bars) has a similar topological arrangement in the two species, despite the considerable differences in their nervous systems. The expression patterns of *Hh/Shh* (violet), *gsb/Pax-3/7* (green) and *patched* (yellow) differ between the two species. The similarities in gene expression are often considered to show that their central nervous systems derive from a single ancestral nervous system of considerable complexity (Reprinted with permission from Arendt, D., Nubler-Jung, K. (1999) Comparison of early nerve cord development in insects and vertebrates, *Development* **126**, 2309–2325. Copyright 1999 The Company of Biologists Ltd.) (See color plate p. xxviii).

deuterostomes, and the basal placement of diploblastic phyla such as the cnidarians, ctenophores, and sponges, such variability in proposed phylogenetic relationships between taxa is a real problem for those seeking to reconstruct the evolution of animal development, as interpretation of comparisons between taxa is entirely dependent on how they relate to each other.

Over the past 20 years, molecular phylogenetics, primarily based on shared derived similarities in the sequences of genes, has helped resolve some of these ambiguities. In theory, most molecular changes should not result in phenotypic change, and hence shared derived molecular characters should give a view of animal phylogeny that is independent of

morphology. In practice, debate continues over whether molecular or morphological phylogenetics gives the most accurate view of the tree of life. Despite this ambiguity, the use of molecular phylogenetics has proven to be a powerful tool that has been used to redraw relationships between animal phyla. A good example of this is the phylogeny of the animal kingdom based on 18S ribosomal RNA gene sequences and some *Hox* gene sequences. This agreed with the morphological consensus in dividing bilaterians into protostomes and deuterostomes, but further divided the protostomes into the ecdysozoa (including the arthropods and the nematodes) and the lophotrochozoa (including the annelids, molluscs, and brachiopods). This rearrangement of the phylogenetic relationships of animal phyla has had profound consequences for our interpretation of animal evolution, suggesting that some monophyletic groups previously created on the basis of shared morphological

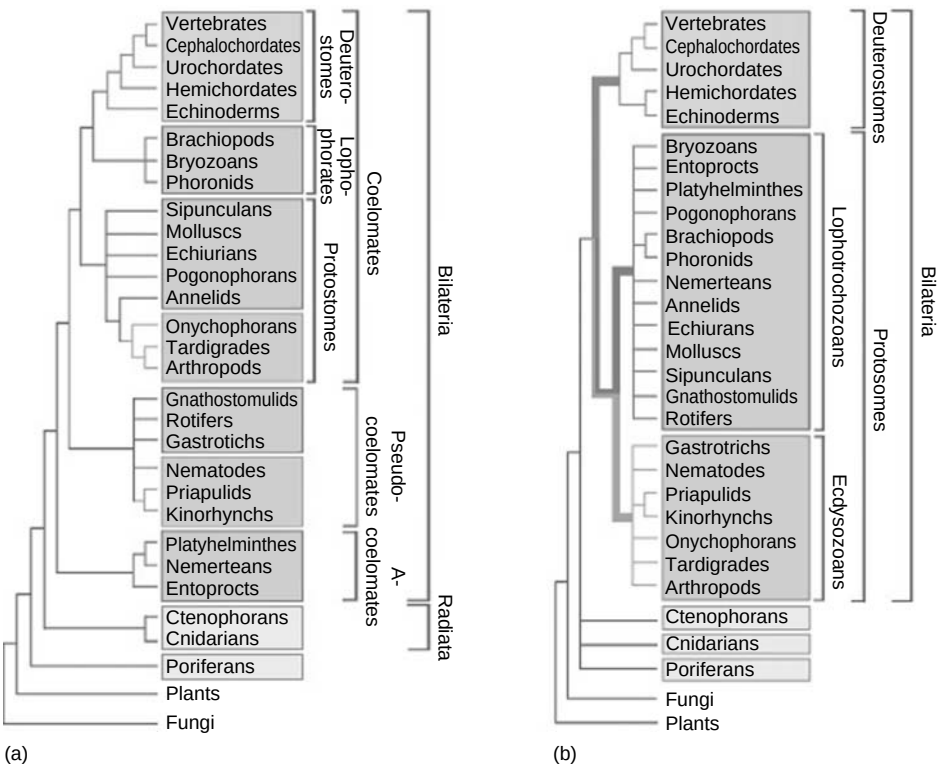

Fig. 2 A metazoan phylogenetic tree, based on morphology and embryology. The bilaterians are subdivided into three major groups: the coelomates, pseudocoelomates, and acoelomates. The tree in (b) is based on more recent DNA-based analyses. Under such an analysis, pseudocoelomates and acoelomates are included into the protostome clade, which is itself subdivided into lophotrocozoans and ecdysozoans (Reprinted with permission from Adoutte, A., Balavoine, G., Lartillot, N., Lespinet, O., Prud'homme, B., de Rosa, R. (2000) *Proc. Natl. Acad. Sci.* **97**(9), 4453–4456. Copyright 2000 National Academy of Sciences, USA).

resemblance were artifacts (Fig. 2) In turn, some researchers have taken this to suggest that the presumed shared morphologies were in fact convergently evolved.

2.2 Identification of a Conserved Developmental Genetic Toolkit in Animals

The developmental genetic toolkit can be defined as the minimum set of genes necessary to construct the embryo, and as such, excludes housekeeping genes that are necessary for the metabolic processes of the adult body. This is a loose definition and more of a working concept than a strict description of what makes a gene "developmental." At a practical level, these genes mostly encode transcription factors and components of intercellular signaling pathways. Transcription factors interact with DNA to regulate the expression of other genes. Differential expression of transcription factors is a fundamental part of cell differentiation. The most famous part of the genetic toolkit is the large family of transcription factors called *homeobox genes* that appear to have some functions conserved in most, if not all, phyla (see Table 1). Other major families of transcription factors in animals include the zinc finger, leucine zipper, basic helix-loop-helix, Tbx, and Fox families. Signaling pathway molecules include members of the TGFβ, Wnt, Hedgehog, Notch, and fibroblast growth factor (FGF) pathways. While transcription factors work inside a cell's nucleus and are essentially cell-autonomous, signaling pathways mediate communication between cells and are the key components in organizing the spatial distribution of differentiating cells. This is a critical step in building a multicellular organism. Many such genes were originally characterized in *Drosophila* and research over the past 20 years has demonstrated that most such genes are shared by many different animal phyla, and are therefore likely to have been evolved at an early point in animal

Tab. 1 Some examples from the developmental genetic toolkit.

Homeobox transcription factors	*Conserved roles*
Hox	Anteroposterior axis
Sox	Neural differentiation
Vnd/Gsx	CNS lateral patterning
Cad/Cdx	Posterior patterning
Otd/Otx	Anterior patterning
So/Six	Neural differentiation
Msh/Msx	Neural differentiation
Other transcription factors	*Conserved roles*
Fd1/FoxC	Dorsal mesoderm differentiation
Bin/FoxF	Visceral mesoderm differentiation
Achaete-Scute/Ath	Proneural differentiation
Tbx genes	Various
Secreted signaling molecules	*Developmental roles*
Wingless/Wnt	Various
Hedgehog	Various
BMP/chordin	Dorsoventral patterning
Notch	Various

evolution, before the radiation of the lineages that have given rise to most living animals. Key examples in this context are the *Hox* genes.

2.3 Hox Genes and AP Axis: A Detailed Example from the Toolkit

The *Hox* genes form a subset of the homeobox gene family. Originally isolated in *Drosophila melanogaster*, Hox orthologs have since been found in all animals. Of special interest is the conservation of the clustered genomic organization of *Hox* genes in several bilaterian lineages (Fig. 3), suggesting first that the common ancestor of both protostomes and deuterostomes had a complement of at least eight genes organized in a single cluster, and second that the clustered organization is important for the proper function of these genes. *Hox* genes are deployed in the developing embryos of several bilaterian lineages in

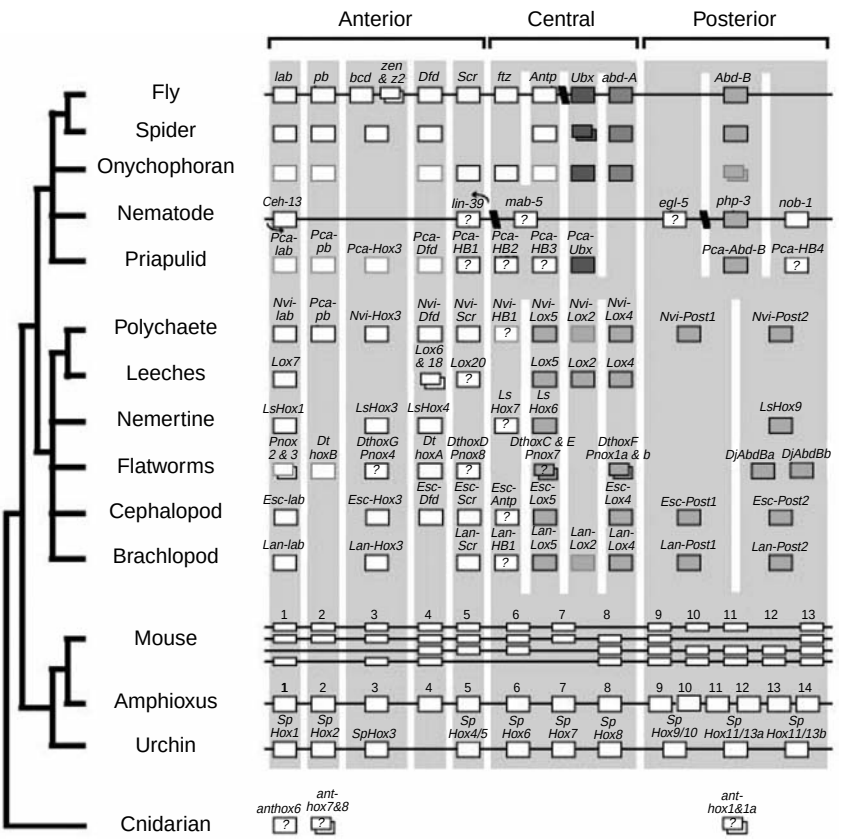

Fig. 3 *Hox* gene clusters in bilaterians. Each shaded column corresponds to an orthologous group. Genes are marked by a question mark when orthology is uncertain. Posterior *Hox* genes have been duplicated in the deuterostome lineage (lower three species) and are all related to the fruit fly gene *Abd-B*. Note that all mouse clusters lack several Hox orthologs (Reprinted with permission from Balavoine, G., deRosa, R., Adoutte, A. (2002) *Mol. Phylogenet. Evol.* **24**, 366–373.)

a strikingly similar way. Specifically, their role in the development of the anteroposterior axis of the early embryo correlates with their relative location within the cluster. Thus, genes located at one end of the cluster are expressed anteriorly in the embryo, while those toward the other end have more posterior expression. This conserved feature is named *spatial colinearity*.

In many taxa, *Hox* genes are also involved in other developmental processes, for example: the fine sculpting of bristle patterns on the legs of drosophilid flies, and the differentiation of the hematopoietic cell lineage in vertebrates. These probably reflect the recent co-option of the *Hox* genes to these new roles, reflecting a fundamental property of the evolution of genes and pathways involved in development: their frequent co-option in the evolution of new morphology.

2.4 The Evolution of this Toolkit via Duplication

When compared, the genomes of *Drosophila* and vertebrates show that, although they possess members of many of the same gene families, the number of members of each gene family is often different between the two lineages. Typically, the excess of copies is in favor of the vertebrates. Extra copies of genes arise by gene duplication, a phenomenon that is common in the evolution of genomes. A paradox is that, directly after the duplication event, the copies of the gene should be functionally redundant, suggesting a lack of selective advantage and rapid elimination of additional copies. Genome comparisons show, however, that many gene duplicates have been conserved for long enough to acquire differences in their patterns of expression. Such acquisition of different spatial or temporal expression patterns is one of the accepted explanations as to why duplicates have been preserved. A model, called the *Duplication-Degeneration-Complementation* (DDC) model, provides the most thorough and satisfying explanation of this process, and takes into account not only variations within the coding sequence but also modifications in regulatory sequences. The DDC model proposes three possible fates for a duplicated gene (Fig. 4): degeneration, when a copy is eliminated or made nonfunctional (pseudogene); neofunctionalization, in which the duplicates diverge by one acquiring a new function (which was not present with the ancestral gene); subfunctionalization, in which the duplicates diverge by reciprocal loss of some of the parental functions. Evidence so far suggests that all three fates happen depending on the gene. In summary, duplication is clearly an important force in evolution of the gene families that regulate development, and will be discussed in combination with gene loss in Sect. 5.

3 Microevolution of Developmental Processes

Loosely defined, the microevolution of development includes the study of within-population variation in developmental processes, and the contribution of development to speciation level processes. There are several key questions in this field that remain to be answered, including, for example: How much genetically determined variation is there in natural populations which gives rise to developmental changes? What is the nature of this

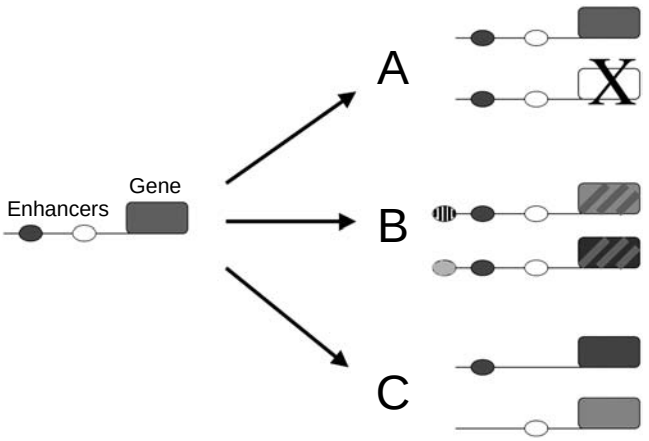

Fig. 4 Fates of duplicated genes as proposed by the DDC model. (A) Degeneration. One copy is eliminated after duplication. (B) Neofunctionalization. Each copy acquires a new expression that can replace or complement the ancestral one by acquisition of new regulatory sequences in front of each gene. (C) Subfunctionalization. Each duplicate becomes essential to the development of the embryo because of reciprocal loss of the original regulatory sequences. Note that these models are not mutually exclusive.

variation; does it primarily involve coding sequences or noncoding sequences? How does this yield phenotypic variation on which selection can operate? Do gene networks or other processes insulate individuals from phenotypic change by buffering developmental processes? Does convergent evolution of morphology tend to involve parallel genetic changes in different lineages? This section illustrates how answers to some of these questions are emerging from recent studies.

3.1 Evolution via Changes in Coding or Noncoding Sequences

While mutation rate can vary substantially across the genome, there is no evidence to indicate that different regions of genes, such as introns, exons, promoters, or enhancers, have generically different mutation rates. Thus, at a basic level, genetic variation can arise in all parts of a gene. When comparing orthologous genes from different species, however, translated exons usually appear more highly conserved than noncoding sequences. This implies that variation in coding regions, specifically of a sort that affects the amino acid sequence of the encoded protein, may be under tighter restraint than other regions of genes. With respect to developmental genes, several studies have demonstrated functional conservation of protein sequences over immense evolutionary distances. For example, the *Otx* genes encode homeobox genes expressed in the anterior of the developing embryos of several phyla. Nagao and colleagues were able to rescue the defects of fruit flies mutated for the gene *Otd*, by inserting the human orthologs, *Otx2* and *Otx1*, demonstrating that the human proteins could replace the *Drosophila* protein despite over 500 hundred million years

of evolutionary separation. Indeed, such studies form the basis of the concept of a conserved developmental toolkit, outlined in Sect. 2. This is not to imply that protein-encoding sequences do not evolve. They do, and this can have functional consequences, for example, some insect *Hox* genes have acquired new roles linked to the evolution of new protein motifs. However, it seems that variation between species is concentrated in noncoding sequences. While much of this may reflect genetic drift of nonfunctional sequence, it also includes the regions responsible for regulating the spatial and temporal deployment of gene transcription, the promoters and enhancers.

A small number of studies have addressed the evolution of the regulation of developmental genes at a functional level. One of the best-studied models in this regard is the fruit fly *Drosophila*. A key example of this is the *Ultrabithorax* (*Ubx*) gene of *Drosophila* and its role in the evolution of leg trichome pattern. Trichomes are small protrusions of the cuticle that cover adults and larvae in patterns that are typically species specific. Stern demonstrated that regulatory changes controlling transcription of the *Ubx* gene resulted in subtle differences in the distribution of the Ubx protein, and were the cause of subsequent subtle differences in leg trichome patterns between species of *Drosophila*. Such subtle changes in gene regulation are presumed to be responsible for the differences in the expression patterns of orthologous genes between species, and are presumed to underlie much of the evolution of developmental processes.

There are also other factors that affect the evolution of enhancers. The *Drosophila even-skipped* (*eve*) stripe 2 enhancer is one of the best-studied regulatory elements. *eve* is expressed in stripes in a pair rule pattern in the *Drosophila* blastoderm embryo, with expression in different stripes controlling a distinct enhancer. Ludwig and colleagues addressed the question of the relationship between genotypic and phenotypic variation in the stripe 2 enhancer. Comparison of the sequence of this enhancer between different *Drosophila* species showed considerable sequence variation, despite the ability of each sequence to direct correct spatiotemporal gene expression in its endogenous setting. By moving the enhancer between species, they then showed that a heterologous enhancer could essentially work in another species, despite the differences in primary nucleotide sequence.

This shows that function can be preserved despite high levels of sequence change in the enhancer region. This raised the possibility that the way enhancers work permits substantial genetic variation without this impacting on the phenotype. At a practical level, this makes understanding enhancer evolution from sequence information extremely problematic, and it is yet to be definitively established whether similar forces operate in species other than *Drosophila*, although, as gene regulation seems to be fundamentally similar between taxa, there is no *a priori* reason to think they should not.

3.2 Parallelism in Morphological Evolution

Similar selective forces, provided by similar ecological conditions, tend to drive the evolution of similar adaptive morphologies in different lineages. Recent studies have raised the surprising possibility that such changes in morphology may often reflect independent evolution of the same developmental pathway. A good example of this is the role of the *shavenbaby* (*svb*) gene in the evolution of larval trichome patterns in

Drosophilid flies. Most *Drosophila* species have been found to have a distinct pattern of trichomes on the larval cuticle. In addition, some species such as *Drosophila sechellia*, have lost trichomes, adopting what is known as a "naked" morphology. These naked species are scattered through *Drosophila* phylogeny, clearly showing that the morphology has evolved, convergently, on several occasions. Shavenbaby performs a key regulatory function in determining where trichomes will develop, and recent studies show that selection for naked morphology has targeted this locus in the evolution of different naked species. This is a clear case of genetic parallelism in the evolution of a convergent phenotype, and raises the possibility that other developmental traits that vary between species may in fact reflect parallel genetic changes.

4 Developmental Basis of Innovations

Innovations are characterized by novelty, with the suggestion that innovations provide some adaptive advantage to a taxon. In many respects their definition is somewhat subjective. Characters described as innovations are often inferred to have evolved at the base of a lineage comprising the taxa in which they are found. In reality, a clear distinction between species that possess a particular innovation and species that do not is probably only apparent due to the extinction of intermediate forms. Examples of such innovations include, in the bilateria, the presence of the mesodermal layer and bilateral symmetry of the body, and, in vertebrates, the neural crest cells and the internal skeleton. Each have been proposed to be the defining features of major animal lineages whose evolution endowed descendants with considerable adaptive advantage.

4.1 The Diploblast/Triploblast Divide: Evolutionary Hypotheses and Key Data

The origin of the mesoderm is often considered as key to the evolution of the complex organs and body plans of the triploblasts. All bilaterians possess this cell layer, and as such are sometimes called the *triploblasts*. In contrast, the members of the cnidaria and ctenophora (jellyfish, anemones and allies, and comb jellies respectively) are usually referred to as *diploblasts*, in that they lack a mesodermal cell layer. They also do not have obvious bilateral symmetry, instead possessing overt radial symmetry organized around an oral–aboral axis. Both morphological and molecular phylogenies agree in placing the diploblasts as basal to the bilateria. Hence, the cnidarians and ctenophores are considered good systems to investigate how the mesoderm evolved. Three basic possibilities exist in this regard. First, mesoderm may be a genuine innovation, the origin of which cannot be traced in diploblasts. Second, there may be a tissue in diploblasts that can be considered as a predecessor to the mesoderm. Third, the diploblasts may be secondarily simplified, and have lost the mesoderm.

The characterization of several cnidarian genes, orthologous to genes known as *mesodermal* markers in triploblasts, sheds light on this issue. Many such cnidarian genes are expressed primarily in the gastrodermis, an endodermal cell layer, in agreement with the hypothesis that the mesoderm is of an endodermal origin and can be traced to a diploblastic tissue. One scenario is that among the genes specifying the endodermal cell layer in

the common ancestor of diploblasts and triploblasts, some were recruited for the specification of the new mesodermal cell layer. The data do not, however, exclude the possibility that diploblasts are degenerate.

4.2 Gene Duplication and Vertebrate Origins

Vertebrates are members of the chordate phylum, characterized by a dorsal neural tube, a notochord and an endostyle. In addition, all vertebrates possess several distinct characters usually considered to be novelties, including migrating neural crest cells, a large regionalized brain, and the endoskeleton (cartilage and bone). The evolution of the vertebrates from an invertebrate chordate ancestor remains a mystery for a key reason: there is no evidence of a living or fossilized animal that shows evidence of a morphological transition state. However, it is possible to trace the origins of some of these new structures by comparing the morphology, developmental processes, and genetics of the vertebrates with their closest living relatives, such as the amphioxus and the ascidians. These organisms are classified with the vertebrates in the chordate phylum, and their embryonic and larval stages share many similarities with vertebrate embryos.

Molecular comparisons have proven to be a popular way to investigate the invertebrate vertebrate transition. A fascinating finding from these studies is that vertebrates possess more members of most transcription factor and signaling molecular gene families than does any invertebrate studied to date. Ohno in 1970 first raised the idea that large-scale gene duplication, via duplication of the whole genome may have played an important role in early vertebrate evolution. Currently, while the mechanism of duplication is still debated, it is widely accepted that some major duplication events happened at the dawn of the vertebrate evolution. As discussed earlier, the direct consequences of gene duplication for a specific gene can vary. What is the consequence of simultaneous duplication of numerous or even all of an organisms genes? This is not known; however, many researchers consider the co-occurrence of such gene duplications with the emergence of vertebrates as indicative that the extra genetic material may have played an important role in the evolution of vertebrates.

5 Loss of Complexity in Animal Development

Much of the earlier discussion focused on the evolution of new characters in animal lineages. However, loss of complex structures and genetic mechanisms is also a powerful force in the evolution of developmental processes and morphology. The two sides of such loss, the genes and the structures that they regulate, are discussed here.

5.1 Phylogeny Suggests Loss of Genetic Complexity is Widespread

When comparing the gene complements of two species, the presence of a specific gene in one species but not in the other can result from two processes. First, it may have evolved specifically in the lineage leading to one species. Second, it may have evolved earlier but been lost by the other species. Mapping the gene complements of species onto a phylogeny reveals some surprising results: it

is apparent that loss of genes that are highly conserved in other taxa is extremely common. Critical examples include the apparent loss of the hedgehog intercellular signaling pathway by the nematode *Caenorhabditis elegans*, and of Hox cluster cohesion in several lineages including *C. elegans* and urochordates. Furthermore, a recent large-scale analysis of the expressed genes in the cnidarian *Acropora millepora* (a coral) identified many genes shared with vertebrates but missing in both *Drosophila* and *C. elegans*. This indicates the loss of these genes in the lineage leading to these species, as their presence in a basally diverging lineage such as a cnidarian clearly demonstrates their early origin in animal evolution. Similarly, the urochordate *Ciona intestinalis* lacks several genes found in both protostomes and vertebrates. Because of the position of urochordates close to the vertebrates in the metazoan tree, we can infer that *Ciona* has lost these genes. In conclusion, there are several examples of lineages having lost genes. In most cases, the direct consequences of such losses to development is unknown, however, many animals that show this feature, including *Drosophila*, *Ciona*, and *C. elegans*, also show a drastic reorganization of early developmental programs, suggesting a fundamental association between gene loss and such evolutionary change.

5.2 Examples of Developmental Regulatory Changes Associated with Loss: Snakes and Sticklebacks

Considerable diversity in animal evolution is associated with the loss or reduction of structures. However, these modifications are not usually due to the loss of a gene, nor are they a restoration of ancestral, simpler morphology. Instead, a loss of body complexity can often be related to the modification of expression of key developmental genes. In vertebrates, for example, the invasion of new ecological niches and the origin of new locomotor adaptations like burrowing or swimming have repeatedly involved the loss or partial reduction of limbs. Obvious examples include whales (which have lost their hindlimbs) and snakes (which have essentially lost both sets of limbs).

Snakes evolved from tetrapod lizards. Their axial skeleton consists of up to hundreds of morphologically similar thoracic vertebrae, while the forelimbs are absent and the hindlimbs are severely reduced or absent. Because *Hox* genes are involved in the patterning of the vertebrate limbs as well as their morphological antecedents, the fins, a logical step was to analyze the expression of these transcription factors in snakes. The genes *Hoxb5*, *Hoxc6*, and *Hoxc8*, expressed in the region where thoracic vertebrate form in other tetrapods, have been examined in a species of python that has no forelimbs, and has hindlimbs reduced to rudimentary outgrowths (Fig. 5). The expression of the genes was found to be altered in a way that prefigured the expanded thoracic zone, thus eliminating the axial territory in which forelimbs usually develop. Hindlimb buds were initiated, but the signaling pathways that are normally required for limb development were not properly activated, stalling further outgrowth and development.

In sticklebacks, a large spine in the pelvic region serves to limit predation in marine populations; however, in some freshwater stickleback populations, this spine is reduced or lost entirely. Shapiro and colleagues performed genome-wide

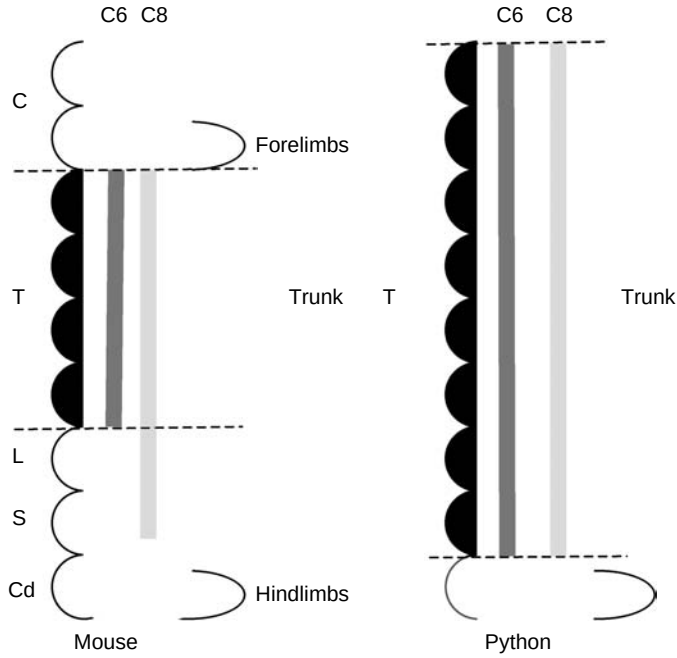

Fig. 5 In limbed vertebrates, the coexpression of *Hoxc6* (in dark gray) and *Hoxc8* (in light gray) in the mesodermal derivatives correlates with the thoracic-type vertebrae (T) and with the forelimb/brachial plexus region. In python, both domains are extended anteriorly and posteriorly changing the identity of the vertebrae and most likely inhibiting the induction of the forelimb development. Rudiments of the hindlimbs are still present in the python and the morphology of the vertebrae at the same level is of an intermediate type. C: cervical vertebrae, T: thoracic vertebrae, L: lumbar vertebrae, S: sacral vertebrae, Cd: caudal vertebrae. See Cohn, M.J., Tickle, C. (1999) Developmental basis of limblessness and axial patterning in snakes, *Nature* **399**, 474–479 for more details.

mapping of several natural populations of stickleback fishes and deduced that the modification was related to the alteration of the expression of the gene *Pitx1*. The experimental manipulation of *Pitx1* does indeed affect pelvic appendage development; however, *Pitx1* is also known to be involved in the development of numerous morphological features, including the proper development of some craniofacial structures. The loss of this gene thus affects multiple parts of the developing embryo, and indeed, no silencing mutation of the gene itself was detected in the mutant sticklebacks. This lead to the hypothesis that the mutation instead affects a regulatory region specifically involved in driving *Pitx1* expression during the late development of the pelvic area, with other regulatory regions remaining intact. This highlights the probability that such subtle changes in gene regulatory elements have the potential to bring about startlingly large changes in morphology, while preserving other essential gene functions.

6 Future Prospects

6.1 The Impact of Genomics and Postgenomics

Clearly, genomics will play an increasingly important role in evolutionary developmental biology, as in many other areas of biology. The current bottleneck in applying genome-scale approaches to nonmodel organisms is financial rather than technical, and this hurdle is beginning to be jumped, with genome sequences of several nonmodel taxa (including the mosquito, honey bee, and sea squirt) already completed, and others (including a sea-anemone, flatworm, mollusk, and annelid) following close behind. How can this wealth of sequence data be utilized to understand animal evolution?

First, the generation of whole genome sequences and expressed sequence tag (EST) sets allow the entire gene complement of an organism to be assessed, giving a full picture of gene family diversity. When plotted onto a phylogeny, this can indicate global patterns of gene loss and gene family expansion over evolutionary timescales. Interesting examples in this context are the lineage specific expansions of nuclear hormone receptor genes in the Caenorhabditid nematodes, and the Krab-like C2H2 zinc finger transcription factors in mammals. Such lineage specific differences will be a prime source of information, detailing how the diversity of genomes and resultant morphologies have evolved.

Second, genome resources allow the rapid isolation of clones for experimental studies, either via clone libraries or via PCR. Many studies of nonmodel taxa published over the past 15 years have concerned single genes in single species, reflecting the time-consuming nature of generating such data. Genome resources greatly shorten the time-consuming gene isolation and characterization stages, allowing more ambitious projects focusing on multiple genes, pathways, and networks.

Third, and perhaps most exciting, a genome sequence provides instant access to potential regulatory elements, at least *in silico*. Modeling of regulatory networks within organisms is a rapidly moving area of research, most advanced in model taxa such as yeast and the nematode *C. elegans*. As the principles governing network function are deduced from these model data sets, it will render possible their application to nonmodel systems. Ultimately, the changes in developmental programs that give rise to different morphologies are translated via such networks from mutations in an organism's DNA.

6.2 Understanding Regulatory Evolution: Computational and Empirical Methods United

The development of computational models raises the possibility that we may eventually be able to fully understand the evolution of animal from genotype to phenotype. The wealth of data means that statistical analyses become possible and mathematical models simulating evolutionary processes are now under construction.

Although the experimentation *in vivo* will still be necessary, developmental biologists and clinical biologists can now compare data concerning the decoded genomic sequences and apply them to their own research. For instance, it is possible to access large data sets of regulatory sequences and thus to infer the

possible regulatory pathways for any given gene in any given genome.

See also Brain Development; Genetic Intelligence, Evolution of; Genetic Variation and Molecular Evolution; Molecular Systematics and Evolution; Preimplantation Mammalian Development, Regulation of Gene Expression in; Principles and Applications of Embryogenomics; Protein Repertoire, Evolution of; Regulome; Spatial and Temporal Expression Patterns in Animals.

Bibliography

Books and Reviews

Arendt, D., Nubler-Jung, K. (1999) Comparison of early nerve cord development in insects and vertebrates, *Development* **126**, 2309–2325.

Bennett, K.D. (2004) Continuing the debate on the role of quaternary environmental change for macroevolution, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **359**, 295–303.

Ferrier, D.E., Holland, P.W. (2001) Ancient origin of the Hox gene cluster, *Nat. Rev. Genet.* **2**, 33–38.

Gellon, G., McGinnis, W. (1998) Shaping animal body plans in development and evolution by modulation of Hox expression patterns, *Bioessays* **20**, 116–125.

Gilbert, S. (1998) Conceptual breakthroughs in developmental biology, *J. Biosci.* **23**, 169–176.

Gilbert, S., Opitz, J., Raff, R. (1996) Resynthesizing evolutionary and developmental biology, *Dev. Biol.* **173**, 357–372.

Gould, S.J., Lewontin, R.C. (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme, *Proc. R. Soc. Lond. B. Biol. Sci.* **205**, 581–598.

Hinchliffe, J.R. (2002) Developmental basis of limb evolution, *Int. J. Dev. Biol.* **46**, 835–845.

Holland, P.W. (2001) Beyond the Hox: how widespread is homeobox gene clustering?, *J. Anat.* **199**, 13–23.

Primary Literature

Adoutte, A., Balavoine, G., Lartillot, N., de Rosa, R. (1999) Animal evolution. The end of the intermediate taxa?, *Trends Genet.* **15**, 104–108.

Adoutte, A., Balavoine, G., Lartillot, N., Lespinet, O., Prud'homme, B., de Rosa, R. (2000) The new animal phylogeny: reliability and implications, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4453–4456.

Aguinaldo, A.M., Turbeville, J.M., Linford, L.S., Rivera, M.C., Garey, J.R., Raff, R.A., Lake, J.A. (1997) Evidence for a clade of nematodes, arthropods and other moulting animals, *Nature* **387**, 489–493.

Amores, A., Suzuki, T., Yan, Y.L., Pomeroy, J., Singer, A., Amemiya, C., Postlethwait, J.H. (2004) Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish, *Genome Res.* **14**, 1–10.

Andrioli, L.P., Vasisht, V., Theodosopoulou, E., Oberstein, A., Small, S. (2002) Anterior repression of a *Drosophila* stripe enhancer requires three position-specific mechanisms, *Development* **129**, 4931–4940.

Balavoine, G., de Rosa, R., Adoutte, A. (2002) Hox clusters and bilaterian phylogeny, *Mol. Phylogenet. Evol.* **24**, 366–373.

Buckley, P.G., Mantripragada, K.K., Benetkiewicz, M., et al. (2002) A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications, *Hum. Mol. Genet.* **11**, 3221–3229.

Cohn, M.J., Tickle, C. (1999) Developmental basis of limblessness and axial patterning in snakes, *Nature* **399**, 474–479.

de Rosa, R., Grenier, J.K., Andreeva, T., Cook, C.E., Adoutte, A., Akam, M., Carroll, S.B., Balavoine, G. (1999) Hox genes in brachiopods and priapulids and protostome evolution, *Nature* **399**, 772–776.

Dearolf, C.R., Topol, J., Parker, C.S. (1989) Transcriptional control of *Drosophila* fushi tarazu zebra stripe expression, *Genes. Dev.* **3**, 384–398.

Erwin, D.H. (2000) Macroevolution is more than repeated rounds of microevolution, *Evol. Dev.* **2**, 78–84.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J. (1999) Preservation of duplicate genes by complementary, degenerative mutations, *Genetics* **151**, 1531–1545.

Freeman, G., Martindale, M.Q. (2002) The origin of mesoderm in phoronids, *Dev. Biol.* **252**, 301–311.

Fujiwara, S., Kawamura, K. (2003) Acquisition of retinoic acid signaling pathway and innovation of the chordate body plan, *Zoolog. Sci.* **20**, 809–818.

Gabaldon, T., Huynen, M.A. (2004) Prediction of protein function and pathways in the genome era, *Cell Mol. Life. Sci.* **61**, 930–944.

Galant, R., Carroll, S.B. (2002) Evolution of a transcriptional repression domain in an insect Hox protein, *Nature* **415**, 910–913.

Gauchat, D., Mazet, F., Berney, C., Schummer, M., Kreger, S., Pawlowski, J., Galliot, B. (2000) Evolution of Antp-class genes and differential expression of Hydra Hox/paraHox genes in anterior patterning, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4493–4498.

Ghazi, A., Paul, L., VijayRaghavan, K. (2003) Prepattern genes and signaling molecules regulate stripe expression to specify *Drosophila* flight muscle attachment sites, *Mech. Dev.* **120**, 519–528.

Hartmann, C., Jackle, H. (1995) Spatiotemporal relationships between a novel *Drosophila* stripe expressing gene and known segmentation genes by simultaneous visualisation of transcript patterns, *Chromosoma* **104**, 84–91.

Hartmann, C., Taubert, H., Jackle, H., Pankratz, M.J. (1994) A two-step mode of stripe formation in the *Drosophila* blastoderm requires interactions among primary pair rule genes, *Mech Dev.* **45**, 3–13.

Hobmayer, B., Rentzsch, F., Kuhn, K., Happel, C.M., von Laue, C.C., Snyder, P., Rothbacher, U., Holstein, T.W. (2000) WNT signalling molecules act in axis formation in the diploblastic metazoan Hydra, *Nature* **407**, 186–189.

Jakob, W., Sagasser, S., Dellaporta, S., Holland, P., Kuhn, K., Schierwater, B. (2004) The Trox-2 Hox/ParaHox gene of Trichoplax (Placozoa) marks an epithelial boundary, *Dev. Genes. Evol.* **214**, 170–175.

Knight, R.D., Shimeld, S.M. (2001) Identification of conserved C2H2 zinc-finger gene families in the Bilateria, *Genome Biol.* **2**, 0016.1–0016.8.

Kortschak, R.D., Samuel, G., Saint, R., Miller, D.J. (2003) EST analysis of the cnidarian Acropora millepora reveals extensive gene loss and rapid sequence divergence in the model invertebrates, *Curr. Biol.* **13**, 2190–2195.

Kwon, C., Hays, R., Fetting, J., Orenic, T.V. (2004) Opposing inputs by Hedgehog and Brinker define a stripe of hairy expression in the *Drosophila* leg imaginal disc, *Development* **131**, 2681–2692.

La Rosee, A., Hader, T., Taubert, H., Rivera-Pomar, R., Jackle, H. (1997) Mechanism and Bicoid-dependent control of hairy stripe 7 expression in the posterior region of the *Drosophila* embryo, *Embo. J.* **16**, 4403–4411.

La Rosee-Borggreve, A., Hader, T., Wainwright, D., Sauer, F., Jackle, H. (1999) hairy stripe 7 element mediates activation and repression in response to different domains and levels of Kruppel in the *Drosophila* embryo, *Mech. Dev.* **89**, 133–140.

Lacalli, T.C., Wilkinson, D.A., Harrison, L.G. (1988) Theoretical aspects of stripe formation in relation to *Drosophila* segmentation, *Development* **104**, 105–113.

Lander, E.S., Linton, L.M., Birren, B., et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Lardelli, M., Ish-Horowicz, D. (1993) *Drosophila* hairy pair-rule gene regulates embryonic patterning outside its apparent stripe domains, *Development* **118**, 255–266.

Ludwig, M.Z., Patel, N.H., Kreitman, M. (1998) Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change, *Development* **125**, 949–958.

Ludwig, M.Z., Bergman, C., Patel, N.H., Kreitman, M. (2000) Evidence for stabilizing selection in a eukaryotic enhancer element, *Nature* **403**, 564–567.

Lynch, M., Conery, J.S. (2000a) The evolutionary fate and consequences of duplicate genes, *Science* **290**, 1151–1155.

Lynch, M., Force, A. (2000b) The probability of duplicate gene preservation by subfunctionalization, *Genetics* **154**, 459–473.

Lynch, M., O'Hely, M., Walsh, B., Force, A. (2001) The probability of preservation of a newly arisen gene duplicate, *Genetics* **159**, 1789–1804.

Lynch, M., Mosher, C., Huff, J., Nettikadan, S., Johnson, J., Henderson, E. (2004)

Functional protein nanoarrays for biomarker profiling, *Proteomics* **4**, 1695–1702.

Malicki, J., Cianetti, L.C., Peschle, C., McGinnis, W. (1992) A human HOX4B regulatory element provides head-specific expression in *Drosophila* embryos, *Nature* **358**, 345–347.

Mantripragada, K.K., Buckley, P.G., de Stahl, T.D., Dumanski, J.P. (2004) Genomic microarrays in the spotlight, *Trends Genet.* **20**, 87–94.

Martindale, M.Q., Henry, J.Q. (1999) Intracellular fate mapping in a basal metazoan, the ctenophore Mnemiopsis leidyi, reveals the origins of mesoderm and the existence of indeterminate cell lineages, *Dev. Biol.* **214**, 243–257.

Martindale, M.Q., Pang, K., Finnerty, J.R. (2004) Investigating the origins of triploblasty: 'mesodermal' gene expression in a diploblastic animal, the sea anemone Nematostella vectensis (phylum, Cnidaria; class, Anthozoa), *Development* **131**, 2463–2474.

Master, V.A., Kourakis, M.J., Martindale, M.Q. (1996) Isolation, characterization, and expression of Le-msx, a maternally expressed member of the msx gene family from the glossiphoniid leech, Helobdella, *Dev. Dyn.* **207**, 404–419.

Mazet, F., Shimeld, S.M. (2002b) Gene duplication and divergence in the early evolution of vertebrates, *Curr. Opin. Genet. Dev.* **12**, 393–396.

Mazet, F., Yu, J.K., Liberles, D.A., Holland, L.Z., Shimeld, S.M. (2003) Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria, *Gene* **316**, 79–89.

McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A., Gehring, W.J. (1984) A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans, *Cell* **37**, 403–408.

Minguillon, C., Jimenez-Delgado, S., Panopoulou, G., Garcia-Fernandez, J. (2003) The amphioxus Hairy family: differential fate after duplication, *Development* **130**, 5903–5914.

Monteiro, A.S., Okamura, B., Holland, P.W. (2002) Orphan worm finds a home: Buddenbrockia is a myxozoan, *Mol. Biol. Evol.* **19**, 968–971.

Nagao, T., Leuzinger, S., Acampora, D., Simeone, A., Finkelstein, R., Reichert, H., Furukubo-Tokunaga, K. (1998) Developmental rescue of *Drosophila* cephalic defects by the human Otx genes, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 3737–3742.

Ohno, S. (1970) *Evolution by Gene Duplication*, Springler-Verlag, New York.

Orihara, M., Hosono, C., Kojima, T., Saigo, K. (1999) Identification of engrailed promoter elements essential for interactions with a stripe enhancer in *Drosophila* embryos, *Genes Cells* **4**, 205–218.

Pankratz, M.J., Seifert, E., Gerwin, N., Billi, B., Nauber, U., Jackle, H. (1990) Gradients of Kruppel and knirps gene products direct pair-rule gene stripe patterning in the posterior region of the *Drosophila* embryo, *Cell* **61**, 309–317.

Peterson, K.J. (2004) Isolation of Hox and Parahox genes in the hemichordate Ptychodera flava and the evolution of deuterostome Hox genes, *Mol. Phylogenet. Evol.* **31**, 1208–1215.

Raible, F., Arendt, D. (2004) Metazoan evolution: some animals are more equal than others, *Curr. Biol.* **14**, R106–R108.

Riddihough, G., Ish-Horowicz, D. (1991) Individual stripe regulatory elements in the *Drosophila* hairy promoter respond to maternal, gap, and pair-rule genes, *Genes Dev.* **5**, 840–854.

Rodriguez-Esteban, C., Tsukui, T., Yonei, S., Magallon, J., Tamura, K., Izpisua Belmonte, J.C. (1999) The T-box genes Tbx4 and Tbx5 regulate limb outgrowth and identity, *Nature* **398**, 814–818.

Ronshaugen, M., McGinnis, N., McGinnis, W. (2002) Hox protein mutation and macroevolution of the insect body plan, *Nature* **415**, 914–917.

Ruvkin, G., Hobert, O. (1998) The taxonomy of developmental control genes in *Caenorhabditis elegans*, *Science* **282**, 2033–2041.

Sanicola, M., Sekelsky, J., Elson, S., Gelbart, W.M. (1995) Drawing a stripe in *Drosophila* imaginal disks: negative regulation of decapentaplegic and patched expression by engrailed, *Genetics* **139**, 745–756.

Scholz, C.B., Technau, U. (2003) The ancestral role of Brachyury: expression of NemBra1 in the basal cnidarian Nematostella vectensis (Anthozoa), *Dev. Genes Evol.* **212**, 563–570.

Scott, M.P., Weiner, A.J. (1984) Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu

loci of *Drosophila*, *Proc. Natl. Acad. Sci. U. S. A.* **81**, 4115–4119.

Seaver, E.C., Paulson, D.A., Irvine, S.Q., Martindale, M.Q. (2001) The spatial and temporal expression of Ch-en, the engrailed gene in the polychaete Chaetopterus, does not support a role in body axis segmentation, *Dev. Biol.* **236**, 195–209.

Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jonsson, B., Schluter, D., Kingsley, D.M. (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks, *Nature* **428**, 717–723.

Shimeld, S.M., Holland, P.W. (2000) Vertebrate innovations, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4449–4452.

Small, S., Kraut, R., Hoey, T., Warrior, R., Levine, M. (1991) Transcriptional regulation of a pair-rule stripe in *Drosophila*, *Genes Dev.* **5**, 827–839.

Small, S., Blair, A., Levine, M. (1992) Regulation of even-skipped stripe 2 in the *Drosophila* embryo, *Embo. J.* **11**, 4047–4057.

Stanojevic, D., Small, S., Levine, M. (1991) Regulation of a segmentation stripe by overlapping activators and repressors in the *Drosophila* embryo, *Science* **254**, 1385–1387.

Stern, D.L. (1998) A role of Ultrabithorax in morphological differences between *Drosophila* species, *Nature* **396**, 463–466.

Stollewerk, A. (2002) Recruitment of cell groups through Delta/Notch signalling during spider neurogenesis, *Development* **129**, 5339–5348.

Stollewerk, A., Schoppmeier, M., Damen, W.G. (2003) Involvement of Notch and Delta genes in spider segmentation, *Nature* **423**, 863–865.

Sucena, E., Stern, D.L. (2000) Divergence of larval morphology between *Drosophila* sechellia and its sibling species caused by cis-regulatory evolution of ovo/shaven-baby, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4530–4534.

Sucena, E., Delon, I., Jones, I., Payre, F., Stern, D.L. (2003) Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism, *Nature* **424**, 935–938.

Takio, Y., Pasqualetti, M., Kuraku, S., Hirano, S., Rijli, F.M., Kuratani, S. (2004) Evolutionary biology: lamprey Hox genes and the evolution of jaws, *Nature* **429**, 1–262.

Technau, U., Scholz, C.B. (2003) Origin and evolution of endoderm and mesoderm, *Int. J. Dev. Biol.* **47**, 531–539.

Venter, J.C., Adams, M.D., Myers, E.W., et al. (2001) The sequence of the human genome, *Science* **291**, 1304–1351.

Williams, N.A., Holland, P.W. (1998) Molecular evolution of the brain of chordates, *Brain. Behav. Evol.* **52**, 177–185.

Zhu, H., Snyder, M. (2001) Protein arrays and microarrays, *Curr. Opin. Chem. Biol.* **5**, 40–45.

Zhu, H., Bilgin, M., Bangham, R., et al. (2001) Global analysis of protein activities using proteome chips, *Science* **293**, 2101–2105.

Vascular Development and Angiogenesis

M. Luisa Iruela-Arispe
University of California, Los Angeles, CA, USA

1 General Overview 203

2 A Stem Cell for the Vascular System: The Hemangioblast 204

3 Development of the Vascular System 205
3.1 Early Vascular Morphogenesis 205
3.2 Vascular Patterning and Remodeling 205
3.2.1 Vascular Patterning 205
3.2.2 Vascular Remodeling 206
3.3 Arteries and Veins 207

4 Lymphangiogenesis 208

5 Key Signaling Pathways in Vascular Morphogenesis 210
5.1 The VEGF Signaling Pathway 211
5.1.1 Soluble VEGFR1, an Endogenous VEGF Antagonist 212
5.2 The Notch Signaling Pathway 214
5.3 Angiopoietins – Tie Signaling Axis 215
5.4 Ephrins and Eph Receptors 216
5.5 TGF-β Signaling 216
5.6 Platelet-derived Growth Factor 217

6 Multiple Forms of Vascular Expansion 218
6.1 Circulating Endothelial Cell Progenitors 218
6.2 Vascular Intussusception 220

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

7 Mechanical Forces and Angiogenesis 220

8 Angiogenesis Inhibitors 221

Acknowledgment 222

Bibliography 222
Books and Reviews 222
Primary Literature 223

Keywords

Angiogenesis
The process by which new blood vessels are originated through the sprouting of endothelial cells from preexistent vessels.

Vasculogenesis
Differentiation of endothelial and precursor cells from the mesoderm and their coalescence into tubes to form a primary vascular plexus.

The development of the vascular system is initiated by the local differentiation of mesenchymal cells into hemangioblasts. These are pluripotent progenitors for both endothelial and hemtopoetic cells. Subsequently, endothelial cells form vascular channels that coalesce and circulation begins thereafter. Recruitment of smooth muscle cells and pericytes contributes to the remodeling stage characterized by the emergence of a hierarchical tree of vessels. Concomitantly, a second array of vessels emerges from the cardinal vein, the lymphatic vasculature. The use of molecular biology and genetics has allowed us to identify the key signaling pathways involved in the specification, differentiation and homeostasis of both the vascular and lymphatic systems. Here we have discussed the details associated with the formation of the vasculature in concert with the key signaling pathways responsible for these events.

1 General Overview

The vascular system develops shortly after gastrulation. The formation of the embryonic vasculature is initiated by the appearance of blood islands from progenitor cells (hemangioblasts) in the visceral yolk sac. Within the blood islands, hemangioblasts differentiate into either hematopioetic or endothelial cells, with the former located within the channel space and the endothelial cells lying on the edges. This first phase of blood vessel formation is known as *vasculogenesis*. A second phase then begins, which is referred to as *angiogenesis*, involving active sprouting of vessels from preexisting blood vessels. Interconnection of these primitive vessels results in the formation of a primary vascular plexus, which undergoes a complex process of remodeling. Eventually, the fetal vasculature emerges from angiogenic growth, selective fusion, and regression of primitive vessels. Figure 1 illustrates the major processes in blood vessel formation and highlights the key signaling pathways required for each step.

The morphogenesis of the heart is concomitant with the differentiation of blood islands. Cardiac progenitor cells migrate to the ventral midline, forming a linear heart tube. As the inflow and outflow tracks are formed, the heart is joined to the yolk sac and flow begins. All subsequent vascular development occurs in the presence of blood flow.

The circulatory system is the first functioning organ system to develop in vertebrate embryos and it is essential for viability and survival. This dependency has been advantageous to investigators. The unsuspected contribution of several regulatory molecules has been revealed by hemorrhage and embryonic lethality. During the last two decades, the use of homologous recombination to ablate gene function has provided major breakthroughs in our understanding of vascular development. Key signaling pathways to vascular morphogenesis include vascular endothelial growth factor (VEGF), Notch, angiopoietins, ephrins, transforming growth factor (TGF)-β, and platelet-derived growth factor (PDGF).

The identification of genes involved in embryonic vascular morphogenesis is important to gain a concrete understanding of how vessels are formed. Many of the events that take place during development are recapitulated in situations of neoangiogenesis in the adult. Thus,

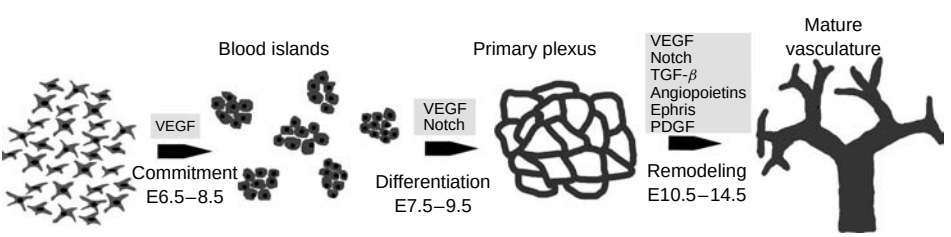

Fig. 1 Development of blood vessels: initially, undifferentiated mesenchymal cells under the instructive signals of VEGF become committed to hematopoietic and endothelial cell fate originating small groups of cells, known as *blood islands*. These clumps of cells gain lumen, coalesce, and form a primary plexus. Subsequently, a series of remodeling events takes place with selective fusion and regression, as well as active expansion by sprouting of new vessels. The end result is a hierarchic mature vascular plexus.

this information is central to the generation of novel and more effective therapies for the manipulation of vascular growth during pathological conditions in the adult.

2
A Stem Cell for the Vascular System: The Hemangioblast

The notion of a common progenitor for both endothelial and hematopoietic cells was first suggested in the early part of the twentieth century by Florence Sabin, based on her studies of vessel formation in the chick blastodisc. Those initial observations were morphological and based on the proximity between endothelial cells and hematopoietic cells in the embryos. Today, we know that endothelial and hematopoietic progenitors share expression of many genes, including VEGFR1, VEGFR2, Tie1 and Tie2, Tal1/SCL, and Runx1. We also know that genetic ablation of VEGFR1 and R2 affects both hematopoietic and endothelial development. Although these findings did not provide a concrete lineage link, they were consistent with at least an interdependency of both cell types.

Subsequently, the utilization of "ES cell *in vitro* differentiation assays" offered an opportunity to further explore the lineage association between endothelial and hematopoietic cells. Using a subpopulation of embryonic stem (ES) cells that expressed VEGFR2, Choi and colleagues demonstrated that under appropriate culture conditions those cells could give rise to both blast-colony forming cells (BL-CFC) and endothelial cells. Subsequent analysis demonstrated that not all VEGFR2 cells could generate BL-CFC, but cells coexpressing both VEGFR2 with Tal1/SCL are enriched for BL-CFC. Tal1/SCL is a transcription factor required for the initiation of primitive and definitive hematopoietic development. Overexpression of Tal1/SCL in Flk1-expressing ES cells significantly increased the number of BL-CFCs in culture. Consistent with those observations, ES cells lacking Tal1/SCL do not differentiate into endothelial cells or hematopoietic cells. However, some degree of endothelial cell development has been noted in Tal1-null mice.

Additional evidence for the existence of a common link between endothelial and hematopoietic cells comes from the finding that during development, certain endothelial cells give rise to hematopoietic progenitors. Functional evaluations *in vivo* have demonstrated that definitive hematopoietic stem cells (HSC) can arise from yolk sac, and in the intraembryonic aorta-gonad-mesonephros (AGM) region. A large morphological body has also noted budding of hematopoietic progenitors from the ventral region of the dorsal aorta, umbilical, and vitelline arteries in chicken, mouse, and human embryos. This potential of endothelial cells to give rise to hematopoietic progenitors is temporally restricted to early developmental stages. Currently, there is no evidence that mature endothelial cells can give rise to hematopoietic progenitors.

Primordia of blood vessels form by the interconnection of hemangioblast clumps that coalesce into blood lakes, and subsequently give rise to a primitive network of endothelial tubes. The process, known as *vasculogenesis*, marks the initiation of vascular development. While, initially confined to an extraembryonic location (yolk sac), shortly thereafter (8–12 h) vasculogenesis is also noted within the embryo proper.

3 Development of the Vascular System

3.1 Early Vascular Morphogenesis

The primitive vascular plexus is composed of differentiated endothelial cells that initiate their function as barrier between the bloodstream and tissues. The main characteristic of this plexus is its uniformity. Channels of the vascular plexus are of the same width, and are separated by similar distances, with no evidence of vascular hierarchy. The vascular plexus is transient and must be altered to accommodate the physical forces associated with the pressures of blood flow. Failure to differentiate this primitive vascular plexus results in embryonic death, generally by E9.5 to E12.5, a feature that has enabled the identification of several genes critical to vascular remodeling.

The vascular plexus expands by vascular sprouting (angiogenesis). During angiogenesis, endothelial cells migrate and proliferate in response to various stimuli, assembling into tubules containing cell–cell tight junctions characteristic of the endothelium. Recently, studies in the developing retina have indicated that the column of cells forming a sprout is heterogeneous. The leading cell, referred to as the *tip cell* within a vascular sprout, does not proliferate in response to growth factors. The tip cell responds to VEGF signaling by sending cellular processes (filopodia) and sensing its environment. The presumption is that, much like the neurons, these cells act to provide directionality to the sprout and the location of the future vascular bed. Recruitment and differentiation of supporting cells associated with mature vessels, such as smooth muscle cells, fibroblasts, and pericytes is also initiated at this time, but is not as prevalent in this stage. Recruitment events are highly regulated and require the coordinated action of a large number of molecules. However, the main group of signaling pathways that is essential to the early morphogenetic events include: VEGF and Notch. Genetic inactivation of molecules within these two signaling pathways results in lethality at E10.5. In addition, haploinsufficiency has been noted in VEGF and delta4 (ligand for Notch) embryos, suggesting that dosage of these ligands is critical to vascular morphogenesis.

3.2 Vascular Patterning and Remodeling

As it develops, the vascular system forms stereotypical blueprints reproducible within the same species and in a few cases even conserved across ontogeny. Patterning results from the combined effect provided by directionality of sprouting, selective vascular fusion, and apoptosis. Thus, some vessels are eliminated, while others increase in size providing the typical treelike structure of blood vessels. Although in both events, patterning and remodeling occur concurrently, for didactic they will be addressed sequentially.

3.2.1 Vascular Patterning

The remarkable conservation of vascular patterning suggests that specific genetic programs coordinate its formation. Interestingly, there is a growing body of experimental evidence to indicate that many of the signaling pathways involved in neural pathfinding also guide endothelial cells during development. The interconnection between neural and vascular patterning has been demonstrated by the finding that

sensory nerves instruct arterial differentiation and blood vessel patterning in the skin. There is an expanding list of neural guidance gene families that regulate migration and proliferation of endothelial cells, suggesting that developmental similarities coupled with physical associations between the vascular and nervous system rely on common molecular mechanisms.

Recently, the sprouting and selective fusion of growing capillaries has been documented by recording zebra fish embryos labeled with fli-EGFP, a construct that enables visualization of endothelial cells. This method has shown that tip cells in sprouts exhibit numerous active filopodia, which extend and retract in an intermittent fashion. Their directionality is guided by the combined set of attractive and repulsive cues, which, in turn, determine the selection of the branching site, the direction, and the fusion. Elucidation of the signaling pathways involved in this process is still in its infancy; however, a subset of these has been identified. Clearly, VEGF plays an important role in the guidance of sprouts, but more specifically the VEGF isoforms that bind to matrix proteins (VEGF188) are responsible for stimulation of sprouts and guidance cues.

Neuropilin, a receptor initially identified for its ability to promote axonal guidance, was later found to bind specific splice isoforms of VEGF. Genetic experiments later confirmed that ablation of neuropilin 1 resulted in abnormal vascular development. In particular, these mice exhibit poor branching in the subventricular zone of the brain indicating that neuropilin 1 is necessary for tip cell guidance, at least in the brain.

Another important group of signaling molecules in patterning are the plexins. In particular, loss of plexin D1 in both zebra fish and in mouse result in abnormalities in pathfinding. Plexin D1 null mice showed abundant branching in the intersomitic vessels with concomitant loss of the normal vascular pattern. The results indicate that disruption of plexin D1 signaling removes the repulsion signals that are necessary for correct guidance during vascular sprouting events. The ligand for plexin D1, Semaphorin 3E has also been inactivated in mouse and reveals a very similar phenotype.

3.2.2 Vascular Remodeling

Remodeling is best understood in the retina. The relatively planar distribution of the capillaries combined with postnatal development make this an excellent system to study vascular morphogenesis. In this system, it is easy to note how specific vessels fuse to form larger vascular channels and others regress giving the vasculature its typical hierarchical structure.

Several pathways have been implicated in vascular remodeling. Essentially, an interruption in the evolution of the primary plexus into the hierarchical vascular tree is considered as a remodeling defect. Signaling molecules, which contribute to the recruitment of mural cells, such as Tie2 and Angiopoietin-1 result in remodeling defects. Also, pathways that affect the lumen and the stability of vessels have also been implicated in remodeling, such as Notch 1. These signaling pathways and their specific contributions will be described later in this chapter.

The contribution of extracellular matrix molecules appears to be extremely important in vascular stabilization and, consequently it can play a role in remodeling events. There is experimental evidence, for example, that type XVIII collagen, a component of basement membranes, is

required for the regression of hyaloid vessels in the developing eye. The hyaloid vessels are formed to provide nutrients to the lens capsule. After this structure differentiates, these vessels regress as part of the normal development of the eye. Absence of type XVIII collagen prevents this regression. The specific mechanisms that drive the regression of these vessels by type XVIII collagen are yet to be determined. It is likely that the signals from matrix molecules are conveyed through integrins. Several integrin proteins have been inactivated and some have revealed important roles in vascular stability, including α v, $\alpha 5$, and $\beta 3$ among others. Thus, it appears that extracellular matrix proteins, through integrins, are important contributors to vascular remodeling and stability.

3.3 Arteries and Veins

A functional vascular circuit requires the separation of arterial and venous networks that only interconnect within the capillary beds of distal target organs. Whereas the endothelial tubes that makeup primitive vascular plexus are morphologically indistinguishable, mature arteries and veins differ in their arrangement of extracellular matrix, smooth muscle cells, and other supporting cells that provide specialized mechanical and physiological properties. Thus, it is conceivable that the formation of arterial and venous networks might differ from one another. Indeed, it appears that during differentiation, endothelial cells become committed to either arterial or venous identities. However, their specific timing of differentiation and segregation during vascular morphogenesis is poorly understood.

Acquisition of arterial or venous features was previously attributed to local environmental factors. Specifically, hemodynamic forces such as blood pressure and blood flow imprinted arterial or venous character to vessels, depending upon the pressure load imposed. During development, however, this was not to be the case. Genetic predetermination drives fate of arterial and venular endothelial cells before any physical regulatory event takes place. Evidence of a genetic program that directs the development of arterial–venous identity comes largely from the characterization of gene-targeted mice and mutational analysis in zebra fish. The first compelling data arose from the observation that the transmembrane ligand, ephrin B2, and its cognate tyrosine kinase receptor, EphB4 are differentially expressed within the arterial and venous endothelium, respectively, of embryonic and extraembryonic vessels prior to the onset of circulation. Genetic inactivation of either ephrin B2 or EphB4 disrupts angiogenic remodeling of arteries and veins that is likely due to inappropriate interactions between arterial and venous angiogenic sprouts. As a result, animals die at midgestation from vascular anomalies and lack of remodeling.

More recently, the Notch signaling pathway has been implicated as an important mediator of arterial differentiation in zebra fish. In fact, inactivation of Notch in zebra fish blocks arterial differentiation, as determined by the expression of arterial-specific markers. Supporting these findings, expression of Notch ligands and receptors in mice is restricted to the arterial vasculature. Similarly, expression of neuropilin 1 and 2 is also segregated in arteries and veins, respectively. Combined, the findings support a critical role for genetic programs regulating arterial and venous fates. However, these genetic programs do not explain how arterial and venous endothelial cells are guided to common

distal targets along parallel but distinct paths.

Failure to establish or maintain proper arterial–venous boundaries in mutants may partially relate to a disruption of normal sprouting mechanisms during angiogenesis. Normally, arteries and veins sprout from the primitive vasculogenic vessels to form parallel but distinct vascular networks that only interconnect at the capillary beds of the target organs. Patients afflicted by hereditary hemorrhagic telangiectasia (HHT) suffer from multiple arteriovenous malformations, direct, and abnormal connections between arterial and venous vascular beds. As will be discussed in more detail, HHT is caused by haploinsufficiency in either activin receptor-like kinase 1 (ALK-1) or endoglin, both are members of the TGF-β superfamily of receptor and ligands. Mice lacking Akt1 or endoglin lose the morphological, molecular, and functional distinctions between arteries and veins. These observations led to the hypothesis that genes responsible for guiding endothelial tubes to their appropriate destinations might be differentially expressed in mutant mice lacking Akt or endoglin. Differential evaluation of these mice resulted in the identification of Robo-4, an endothelial-specific member of the Robo family of proteins. Combined, it appears that establishment of arterial and venous identity has to occur in parallel to patterning and selective fusion of new sprouts.

Finally, it is important to stress that arterial and venous molecular signatures, although established early during development, may not be sufficient to maintain the arterial–venous identity. It has been experimentally demonstrated that after transplanting the arterial endothelium into veins and vice versa, the transplanted endothelium adopts the specific molecular programs of the host vessel, suggesting that local cues modulate vessel specificity. The specific identity of these cues is unknown; however, it is also likely that mechanical forces, including flow and pressure might play important roles.

4 Lymphangiogenesis

In addition to blood vessels, the circulatory system also includes a group of lymphatic vessels. Unlike the circulatory loop formed by the heart and blood vessels, the lymphatic vasculature consists of blind-ended capillaries that permeate most tissues to drain interstitial fluid and proteins, and return them to the blood circulation by large lymphatic conduits (thoracic duct) that drain into the left subclavian vein. Lymphatics are also responsible for fat uptake from the gut and play an essential role in the immune system by directing leukocytes and antigen presenting cells to the lymph nodes. The blood and lymphatic systems complement each other to maintain tissue homeostasis; an interdependence that is also reflected in their coordinated development during embryogenesis.

As first described by Florence Sabin in 1901, the lymphatic vessels arise in midgestation (E10.5 in mouse) after the development of the blood vascular network from progenitor cells in the cardinal vein. Subsequent to the formation of the lymph sacs, progressive expansion of lymphatic vessels occurs by active sprouting, in a manner similar to blood vessel sprouting, but under regulatory control of a different cast of molecules. Prox-1, a homeobox gene related to the Drosophila Prospero, is one of such molecules. Mice lacking this transcription factor fail to develop lymphatics and die

at midgestation from generalized edema. It appears that Prox-1 is essential for the commitment and differentiation of endothelial cells from the cardinal vein to the lymphatic lineage. Figure 2 illustrates the process of lymphangiogenesis from the budding of endothelial cells from the cardinal vein to the organization of lymphatic capillaries. In the absence of Prox expression, cells that bud from the cardinal vein continue to display a cell-surface marker profile of the blood vascular system suggesting that expression of Prox-1 is sufficient for the specification of lymphatic endothelial cells. In fact, overexpression of Prox-1 is sufficient to reprogram cultured blood vascular cells into lymphatic endothelial cells.

Maturation of the lymphatic vascular system, as well as homeostatic function of lymphatics vessels requires expression of VEGF-C and VEGFR3. Activation of VEGFR3 by either VEGF-C or D leads to enhanced proliferation, migration, and survival of cultured human adult lymphatic endothelial cells. In addition, activation of VEGFR3 by overexpression of VEGF-C in adult transgenic mice results in lymphangiogenesis, and blocking of VEGFR3 signaling *in vivo* by expression of a soluble form of the receptor leads to lymphedema in adult. Finally, the association of missense mutation of the human VEGFR3 gene in patients with chronic lymphedema (Milroy's disease) supports a requirement for VEGFR3 signaling in the adult. It should be stressed that expression of VEGFR3 is initially not restricted to lymphatic vessels. In fact, VEGFR3 is highly expressed in vascular endothelium during development and mice die from disruptions in the morphogenesis of the vascular system. After E13.5 in the mouse, VEGFR3 is downregulated in blood vessels and becomes highly expressed in lymphatic endothelium.

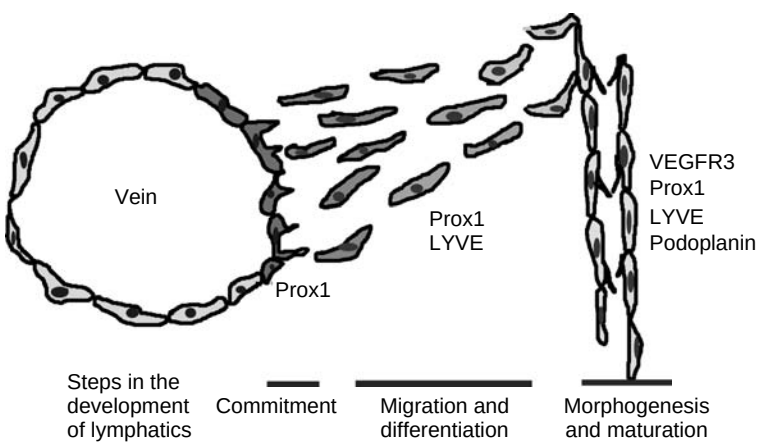

Fig. 2 Lymphangiogenesis: the initiation of lymphatic vessels occurs by the specification of subdomains within veins. These showed commitment to the lymphatic fate by expressing Prox1, a transcription factor that regulates expression of certain lymphatic genes. Subsequently these cells depart from the parental veins and migrate to organize a parallel vascular system that only drains its contents into major veins, but it is otherwise completely independent from blood vessels. Markers for lymphatic endothelial cells include: VEGFR3, Prox1, LYVE, and podoplanin.

VEGFR2 is expressed in lymphatic endothelium, as well as in blood vascular endothelium. In lymphatic vessels, VEGF-C is able to activate both VEGFR2 and VEGFR3. However, mice deficient for VEGF-C display a defect in the formation of the initial lymphatic vessels sprouts from veins, but specification of the lymphatic lineage is not impaired. Evaluation of heterozygous mice for VEGF-C revealed that dosage of this ligand is critical for normal lymphatic function in the adult. In fact, VEGF-C +/− mice display lymphedema in the lower limbs as well as chylous ascites from impaired lymphatic function in the gut.

The angiopoietin-Tie2 signaling pathway is also important in lymphangiogenesis. The contribution of Ang-Tie, however, appears to be more important for the function rather than to the development of lymphatics. Mice lacking Ang2 develop chylous ascites, an accumulation of interstitial fluid in the gut that is associated with a loss of gut lymphatic function. This requirement for Ang2 in the lymphatic system can be substituted by Ang1, as demonstrated by gene replacement. In addition, a subset of embryos deficient for Tie1 displays severe tissue edema that is independent of placental or cardiac dysfunction.

Abnormal development of peripheral lymphatics results in severe edema and impairment of lymphatic development in mouse models can be lethal. Equally importantly, lymphatic vessels are created by tumors and these vessels are used by tumors to metastasize from their site of origin. Growth of lymphatics in some tumors even appears to be rate limiting for metastasis. Thus, identification of molecular pathways regulating lymphatic growth may provide new strategies to treat vascular and neoplastic diseases.

5 Key Signaling Pathways in Vascular Morphogenesis

The molecular nature of the signaling pathways involved in blood vessel development has become an intense area of investigation, and great strides have been made toward elucidating the functions of the required genes. The mouse has proven to be a robust model system for studying vascular morphogenesis, not only because its developmental processes are similar to humans but also because of the wealth of available methods to manipulate the genome and create mutant animals. Many mutant mouse lines with a multiplicity of early and late vascular defects are currently available and allow for detailed analysis of the role of single genes or combination of genes during development, homeostatic functions, and under pathological conditions.

There are four most extensively studied ligand/receptor families that are essential for vascular morphogenesis: VEGF-VEGFRc; Notch-DSL; angiopoietin(Ang)-Tie2, and ephrin B2-ephB4. Their critical roles during development have been established by diverse ranges of approaches including the use of genetically ablated mouse lines for each gene. In addition, the TGF-β and PDGF pathways are also essential and will be discussed. Figure 1 provides an overall assessment of the impact of each of these pathways during vascular maturation. While we will constrain our comments to these major signaling pathways, it is important to stress that a large number of extracellular matrix proteins and their receptors, adhesion molecules, and other growth factors contribute extensively to the overall process of vascular maturation.

5.1 The VEGF Signaling Pathway

Initially identified for its ability to mediate vascular permeability, VEGF was later found to be a mitogenic agent for endothelial cells. Today, VEGF is recognized as an essential cytokine for the development and homeostasis of the cardiovascular system. Inactivation of a single VEGF allele results in embryonic lethality at midgestation due to severe cardiovascular defects. Interestingly, increase of VEGF in the heart, also leads to embryonic lethality. In the adult, overexpression of VEGF by basal epidermal cells (expressing keratin-14 promoter) results in increased microvascular density, leukocyte adhesion, and hypersensitivity to inflammation. Together these results indicate that unlike most genes, alterations in VEGF levels can result in significant pathological outcomes.

Although the major focus of interest has been on VEGF-A (or simply VEGF), this molecule is only one member of a family of proteins that also comprises VEGF-B, VEGF-C, VEGF-D, VEGF-E, and PlGF (placenta growth factor). These proteins interact with three major tyrosine kinases: VEGFR1, VEGFR2, and VEGFR3 and two nonreceptor tyrosine kinases; neuropilin 1 and neuropilin 2, which also bind to other ligands. Figure 3 provides a schematic view of these receptors and their respective specificity for each ligand. While this chapter will focus mostly on VEGF-A, referred herein as VEGF, much research has been done in the biological functions of the other family members.

The biologic activities of VEGF are mediated principally by two related

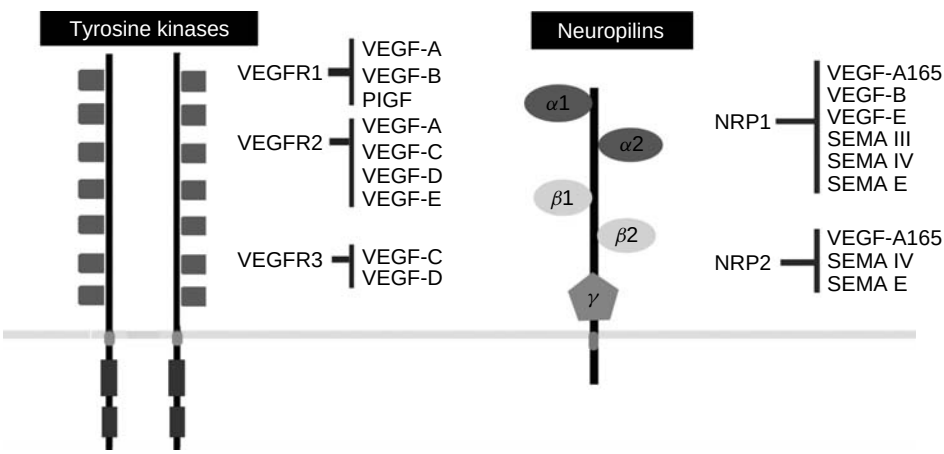

Fig. 3 VEGF receptors and ligands: VEGF signaling is essential for blood vessel development and homeostasis. This pathway includes a series of tyrosine kinases: VEGFR1, VEGFR2, and VEGFR3 that belong to the class III receptor tyrosine kinases of the platelet-derived growth factor receptor superfamily. All three receptors are characterized by the presence of seven immunoglobulin homology domains (in the extracellular region) and two intracellular kinase domains. The relative binding affinities of each receptor for the ligands is indicated. Neuropilins and accessory receptors to the VEGF pathway. The structure of neuropilin (NRP) 1 and 2 comprise two α, two β, and one γ subdomains, in addition to a short intracellular tail that does not show significant homology to any known proteins. The interaction of NRP 1 and 2 with their respective ligands is shown on the right.

cell-surface receptors, Flt1/VEGFR1 and KDR/Flk-1/VEGFR2 with extracellular ligand-binding domains and intracellular protein-tyrosine kinase domains. It has been determined that VEGFR1 is the high affinity receptor, while VEGFR2 is the low affinity receptor. In addition, some VEGF isoforms can bind to neuropilin receptors; these receptors will be discussed separately. VEGFR2 is the main signaling receptor for VEGF. Activation of this receptor has been shown to promote differentiation of progenitors, migration, mitogenesis, survival, nitric oxide release, and vascular permeability. In contrast, the biological role of VEGFR1 remains enigmatic. Treatment of VEGFR1-expressing endothelial cells with VEGF typically does not stimulate migration or DNA synthesis. Other reports using chimeric receptors or neutralizing anti-VEGFR1 antibodies to distinguish VEGFR1 and VEGFR2-mediated pathways suggest that VEGFR1 may indeed transmit signals that down-modulate endothelial cell responses to VEGF via VEGFR2. VEGFR1 appears to suppress both VEGF-stimulated chemotaxis and proliferation in endothelial cells.

Other evidence indicates that VEGFR1 performs important functions during vascular development. These might be mediated by VEGF, but could also be directed through activation by placenta-derived growth factor, another ligand for VEGFR1. Furthermore, in the adult, VEGFR1 appears to regulate monocyte function.

In situ hybridization reveals that patterns of VEGFR1 expression partially parallel those of VEGFR2 and VEGF in the developing mouse embryo. Moreover, homozygous knockout of VEGFR1 in mice results in embryonic lethality between days E9.5 and E11.5 characterized by a superabundance of endothelial cells and their hemangioblast precursors and the failure to form vascular channels. It was postulated that VEGFR1, as a buffer for embryonic VEGF, serves to keep endothelial cell expansion within limits appropriate for efficient vascular channel assembly. The VEGFR1-null phenotype is distinct from that seen in VEGFR2 homozygous knockout mice, which died *in utero* between days E8.5 and E9.5 with marked deficits in hematopoietic and endothelial cell numbers. Intriguingly, transgenic mice homozygous for tyrosine kinase–null VEGFR1 containing intact extracellular and transmembrane domains developed a normal, functional vasculature, but showed defects in monocyte chemotaxis. These findings reinforce the concept that, in endothelial cells, VEGFR1 may serve a modulatory function by virtue of its ligand-binding activity and potential to interfere with VEGFR2 or neuropilin 1, in addition to a signaling function.

While in neonatal mice expression of VEGF is required for survival, the function of VEGF signaling in adult vessels is less clear. There is evidence that activation of this signaling pathway is required for the maintenance of fenestrations in certain vascular beds. Blockade of VEGF signaling in adult mice has resulted in reduced capillary density and proteinuria.

5.1.1 Soluble VEGFR1, an Endogenous VEGF Antagonist

In addition to the major VEGFR1 mRNA species of 6.5 to 8.5 kb (presumed to encode the full-length receptor protein), Northern blotting typically detects other VEGFR1 hybridizing mRNA, originally detected in human placenta and endothelial cells, which arises by alternative RNA splicing. This mRNA encodes a protein, termed *soluble* or *secreted* VEGFR1 (or

sFlt1), in which a unique intron 13-encoded C-terminal peptide replaces the membrane-anchoring and tyrosine kinase segments, resulting in release of VEGF-binding receptor fragment from cells. Mouse VEGFR1 gene structure reveals that sFlt-1 mRNA is generated by the failure to excise intron 13 and the use of a presumed intronic polyadenylation signal. Recombinant sFlt1 binds VEGF with high affinity (Kd 20 pM) and acts as a potent inhibitor of VEGF biological activities *in vitro*. sFlt1 binds heparin, which may favor its retention close to sites of release in tissues containing heparan sulfate proteoglycans. Additional soluble forms of VEGFR1 have been described although their origin and properties remain unclear.

sFlt1 and chimeric analogs have been used experimentally as potential therapeutic agents and to test the role of VEGF in a variety of physiologic and pathological processes *in vivo*. For example, sFlt1 has been investigated as an experimental therapy for neoplasms that constitutively produce VEGF. Increased expression of sFlt1, either via transfected or infected tumor cells or administration of adenoviral vectors can dramatically inhibit vascularization and/or growth of tumor xenografts in mice. Similarly, virus-directed expression of sFlt1 reduced subretinal neovascularization in rats and ischemia-induced retinal neovascularization in mice. In a small-scale study, sFlt1 protein in the systemic circulation was marginally increased in patients with proliferative retinopathy compared to healthy controls.

Although it is clear that sFlt1, when overexpressed, can effectively inhibit biological actions of VEGF *in vivo*, little is known about the physiological function(s) of sFlt1 or the mechanisms controlling its biosynthesis. Since sFlt1 expression can occur only where the parental VEGFR1 pre-mRNA is produced, the fundamental level of control is transcriptional, governed by the endothelial-selective promoter elements for VEGFR1. In cell cultures or tissue explants, expression of mRNAs for VEGFR1, but not for VEGFR2, can be induced by hypoxia. This pattern of VEGFR1 inducibility can be recapitulated in animals exposed to systemic hypoxia.

Neuropilins and semaphorins. Neuropilins are transmembrane receptors best understood for their interaction with semaphorins. However, recently it has become clear that in addition to semaphorins, neuropilins also bind to VEGF isoforms 164 and 188. Eight classes of semaphorins have been identified and most of these appear to be specific to the neural system. They provide negative neuronal guidance cues: repel axons and collapse growth cones. Class-3 semaphorins, however, bind to neuropilin receptors also expressed by endothelial cells, in addition to neurons. There are six members of class-3 semaphorins (sema A–F) and they exhibit some relative degree of specificity. Thus, sema-3A activates neuropilin 1 and sema 3F activates neuropilin 2.

Clear demonstration of the impact of semaphorins in the cardiovascular system has been obtained from genetic knockout studies. Sema-3A–null mouse in 129/SV background die from cardiac failure approximately 3 days after birth. In the CD-1 background, they also exhibit a number of vascular defects in the head and abnormalities in trunk vessels. At E9.5, cranial blood vessels of the mutant embryos fail to remodel and the formation of the anterior cardinal vein is disrupted. Neuropilin 1–null mice also showed vascular remodeling defects with onset of lethality at E12.5 to E13.5. They also display agenesis of some branchial arch-derived vessels and

transposition of aortic arches. Overexpression of neuropilin 1 in mice also results in vascular defects. These include: vascular dilation, excess of vascular growth, and cardiac malformation.

Sema-3C–null mice die just after birth due to aortic arch malformations and septation defects in the heart. Interestingly, these were not noted in neuropilin 2–null mice. Instead, neuropilin 2 deficiency showed reduction of small lymphatic vessels and capillaries.

Interestingly, removal of both neuropilins showed a more severe vascular phenotype than either receptor alone. Embryonic lethality occurs at E8.5 with defects in yolk sac and embryonic vasculature. Overall, there seems to be a decreased number of endothelial cells in a manner similar to the one observed in VEGFR2–null mouse. Together these findings indicate that activation of neuropilins by semaphorins is an essential component of vascular morphogenesis.

5.2 The Notch Signaling Pathway

The Notch signaling pathway impacts a large variety of cell types including hematopoietic, neuronal, muscle, epithelial, and more recently, endothelial. It has been shown to participate in cell fate decisions either by initiating differentiation processes or by maintaining the undifferentiated state of cells. The contribution of Notch to the vasculature was initially recognized by the association of ligands and receptors with hereditary vascular anomalies. CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy), is a disease of adult onset manifested by strokes, migraines, and progressive dementia. CADASIL has been linked to mutations in notch 3 resulting in a progressive degeneration of the smooth muscle layer surrounding cerebral and skin arterioles. Alagille syndrome has been attributed to mutations in jagged 1. Patients suffering from this syndrome exhibit abnormally formed blood vessels, arterial stenosis, and heart disease, in addition to hepatic lesions and skeletal defects.

In mammals, there are four distinct Notch receptors (notch 1–4) and five ligands (jagged 1–2 and δ1, 3, and 4). Unlike the VEGF signaling pathway, Notch receptors and ligands are transmembrane, thus interaction requires cell–cell contacts. Most frequently, associations occur between cells (homotypic or heterotypic) resulting in trans-signaling events. However, activation of receptors in cis can also occur. There does not appear to be ligand specificity between the ligands and receptors; however, recent experimental evidence suggests that not all receptor/ligand pair results in signaling. Activation of Notch requires a series of proteolytic events that are trigged by binding to receptors. The enzymes implicated in processing include members of the ADAM family (ADAM10/Kuzbanian and gamma secretase). Interestingly, genetic inactivation of both these enzymes has resulted in embryonic lethality with vascular anomalies similar to those found in Notch mutants. Once released, the intracellular domain of Notch (ICD) translocates to the nucleus where it binds to the transcription factor RBPJ-k (also known as *CBF-1/RJBk* in mammals, and Su(H) in Drosophila). The complex then activates a cohort of specific downstream effectors. The best characterized direct targets of Notch activation include the Hairy genes in Drosophila and the related bHLH transcription factor genes, Hes/Hey in vertebrates.

Targeted inactivation of notch 1 has confirmed its relevance to the development of the cardiovascular system. Embryos homozygous for a null allele of notch 1 die by day 9.5 with clear defects in somitogenesis and severe cardiovascular anomalies. Targeted mutation of another Notch receptor, notch 4, showed no defects in homozygous mutant mice, but compound homozygous for notch 1 and 4 had a more severe vascular defect than embryos homozygous for the notch 1 receptor alone. Interestingly, gain of function of notch 4 in the endothelium has an abnormal vascular development and early lethality. Like VEGF, it seems that vascular morphogenesis is extremely susceptible to levels of Notch signaling.

Several Notch ligands have been inactivated by homologous recombination. Embryos lacking a functional jagged-1 gene die at embryonic day 10.5, with vascular defects including lack of vascular remodeling and absent vitelline vessels. More recently, inactivation in delta4 results in haploinsufficiency with lethality at E 9.5. Mice die from lack of remodeling of the primary vascular plexus, a finding that was evident in both the yolk sac and in the embryo proper.

While all Notch receptors and ligand mutants do not link this pathway in hemangioblast or endothelial cell specification, a possible contribution in Notch signaling in cell fate choice might occur in the lymphatic system. Activated Notch appears to directly transactivate VEGFR3, and thus, it might contribute to the specification of lymphatic endothelial cells from veins.

5.3 Angiopoietins – Tie Signaling Axis

Initially isolated through a secretion-trap expression strategy aimed at identifying the ligand for Tie2, Angiopoietin-1 belongs to the family of ligands currently comprising four members (Ang 1–4). From these, Ang 1 and 2 are more understood with respect to their roles in developmental and pathological angiogenesis. Although these two ligands bind to the receptor tyrosine kinase Tie2 with similar affinities, Ang2 can inhibit Ang1 under some specific physiological contexts. Some studies *in vitro*, however, have shown that Ang2 can induce Tie2 phosphorylation in endothelial cells depending on dose, duration of exposure, and whether the source is autocrine or paracrine. Interestingly, Ang2 is expressed at angiogenic and vascular remodeling sites and it contributes to the detachment of smooth muscle cells. This enables the exit, migration, and proliferation of endothelial cells during the process of angiogenesis. These findings explain why Ang2 can either participate in the regression of the formation of new blood vessels, since destabilization of vessels is also a requirement for angiogenesis.

Inactivation of Tie2 by homologous recombination results in embryonic lethality by midgestation (E12.5) from cardiovascular abnormalities. Mice exhibit lack of remodeling of the primary plexus and sprouting deficiency. Genetic ablation of Ang-1 results in a similar phenotype, indicating that at least during development, these two molecules are codependent. Gain of function of Ang1 results in increased vascularization and abnormal permeability.

Although ligands for the related kinase, Tie1, are yet to be identified, inactivation of this receptor is also embryonically lethal (from E13.5 to neonates). Deficiency in Tie1 results in edema and microvascular fragility with hemorrhage. Together the data indicates that Tie1 is not required for the early differentiation of endothelial cells

and, unlike Tie2, is not necessary for initial patterning.

In adult settings, it is clear that Ang1 is a natural antipermeability factor protecting against excessive plasma leakage and counteracting the pathological permeability induced by VEGF.

5.4 Ephrins and Eph Receptors

The family of Eph receptors includes a large group of receptor tyrosine kinases (14 in total) that regulate an impressive array of biological effects, including: embryonic patterning, neuronal targeting, epithelial differentiation, and vascular development, among others.

Activation of Eph receptors by their ephrin ligands requires cell–cell contact, as ephrins are also membrane bound. Signaling between ligands and receptors is bidirectional and dependent on the dimerization state of the ligand. The effects of ephrins have been well studied in neurons. In this system, ephrins provide guidance by repulsion of contact of the growing axon. There have been eight ephrin ligands identified thus far; however, only a subset of these appear to impact the cardiovascular system.

The first indication that Eph signaling was involved in vascular morphogenesis came from genetic inactivation analysis. In fact, removal of either ephrin B2 or EphB4 results in early embryonic lethality (E10.5) with alterations in vascular remodeling and lack of arterial and venular identities. Interestingly, expression analysis of LacZ reporter mice under the regulatory control of ephrin B2 and EphB4 revealed complementary expression patterns, with ephrin B2 in arteries and EphB4 in veins. Expression of ephrin B2 and EphB2 is also found in mesenchymal cells, suggesting a possible role for this signaling system in the recruitment of mural cells during remodeling events.

The signaling consequences initiated by Eph-ephrins have been also explored *in vitro*. Thus, signaling of EphA2 by ephrin A1 in aortic endothelial cells attenuates proliferative signals mediated by VEGF, through what appears to be the Erk pathway. In addition, this system regulates cell–cell recognition, as was anticipated by the defects noted in the null mouse.

5.5 TGF-β Signaling

TGF-β is a member of a large superfamily that includes: bone morphogenetic proteins, activins, inhibits, and Mullerian inhibitory substance all relevant to developmental processes. Three members of the TGF-β family (TGF-β1–3) have been identified all with partially overlapping expression, but distinct functions. The growth factors are secreted as latent forms and its activation is dependent on either proteolytic processing or binding to thrombospondin-1. Signal transduction by TGF-β requires a series of serine/threonine receptors, accessory receptors, Smad proteins, and Smad transcription factors that convey these signals to specific genes.

Because of its early expression and broad distribution, it was extremely surprising to find that targeted disruption of the TGF-β1 gene does not necessarily lead to embryonic lethality or congenital anomalies and mice tend to die 3 weeks after birth from immunological deficiencies. Initially, the potential issue of redundancy was a plausible explanation. Subsequently, it was shown that TGF-β could be transferred from mother to fetuses by the placenta

and to pups through the milk. This transfer is likely to account for the lack of a more severe phenotype, as the transfer would rescue the lack of endogenous production of the growth factor. While the maternal source of TGF-β could rescue some animals, still only about 50% of homozygous mice were recovered. The lethality was associated with defects in both hematopoiesis and vasculogenesis resulting in lethality around E9.5 to E11.5.

Genetic inactivation studies have clearly shown that TGF-β signaling is important for differentiation of endothelial cells. However, a further understanding of its contribution has been difficult to ascertain, particularly since the effects of TGF-β appear to be multiple and altered by other factors. In two-dimensional cultures, TGF-β has been shown to inhibit endothelial cell proliferation. Interestingly, in three-dimensional cultures, TGF-β does not affect proliferation, but appears to modulate differentiation. More recently, these opposing results have been conciliated by a more concrete understanding of TGF-β signaling pathways.

TGF-β signaling is mediated by its type I and II serine/threonine kinase receptors. There are two type I receptors: ALK-1 and 5 and one type II receptor, also known as TGF-β receptor II (TGFβRII). By activation of Smad1/5, it has been shown that ALK1 is responsible for the stimulation of endothelial cell proliferation and migration; while ALK5 through activation of Smad2/3 results in inhibition of cell proliferation and migration. Consequently, the net expression of these two receptors and their respective target Smads dictates the end result of TGF-β response in endothelial cells.

Mutations in ALK1 or on its accessory receptor, endoglin, have been linked to HHT. Consistent with these findings, ALK1-null mice die at E10.5 to E11.5 with defects in smooth muscle cell development and angiogenesis. ALK1 knockout animals also display vascular branching defects, dilations, and abnormal vascular connections (arterial–venular fusions). Mice lacking either TGF-βRII or ALK5 die at E10.5 due to defects in vascular development clearly seen in the yolk sac. Mice lacking endoglin also display embryonic lethality, with defects in vascular remodeling and abnormalities in arterial–venous boundaries. The contribution of this signaling pathway to vascular development has also been reinforced by the experimental inactivation of several Smad proteins in mice.

5.6 Platelet-derived Growth Factor

The PDGF family of dimeric growth factors shares a significant degree of sequence similarity to VEGF, yet its expression patterns and functional properties are clearly distinct. PDGFs and their tyrosine kinase receptors are expressed and impact a large number of tissues including fibroblasts, smooth muscle cells, neurons, and endothelium. This expression pattern explains why deregulation of this pathway has been associated with a myriad of human diseases, including atherosclerosis, fibrosis, and cancers.

The classical PDGFs include homo- and heterodimeric associations of the A and B chains, forming PDGF AA, BB, and AB respectively. The A and B polypeptides share similar protein domains, and are secreted as fully active factors. Like VEGF, these classical PDGFs also contain "retention matrix" motifs that allow them to interact with extracellular matrix proteins and regulate their biological availability to specific sites. Signaling occurs by

three receptors resulting from the association of two transmembrane polypeptides: PDFGR-α and β that also function as homo and heterodimers.

The contribution of PDGFs to the vasculature has been revealed by gene inactivation in mice. Phenotypic analysis of null mice implicated PDGF-B in the maturation stages of vascular development. PDGF-BB is expressed by endothelial cells and its receptor, PDGF-β, is located in vascular smooth muscle cells and pericytes. Absence of PDGF-B results in reduced pericyte association with capillaries, vascular fragility, and hemorrhage with subsequent lethality. Both PDGF-B and PDGFR-β-null mice had decreased numbers and proliferation of smooth muscle cells/pericytes progenitors. This was more noticeable in the brain, heart, and brown adipose tissue.

Recently, two additional growth factors have been identified in this family: PDGF-C and D. In contrast to A and B, the novel PDGFs are secreted as zymogens and require activation by proteolytic cleavage of their N-terminal CUB domains. Following proteolytic processing, PDGF-C binds as a homodimer to PDGFR-α, whereas active PDGF-D interacts with PDGFR-β. Both are also able to activate PDGFR-α and β heterodimers in cells that expressed these two chains. Genetic inactivation of PDGF-C results in defects in secondary palate formation and the dermis. The contribution, if any, of PDGF-C and D in the vascular system is yet to be elucidated.

6 Multiple Forms of Vascular Expansion

Angiogenesis is also a vital and dynamic process in adult vertebrate organisms. Vascular expansion is essential for normal tissue growth, wound healing, endometrial cycling, and pregnancy. Moreover, angiogenesis can be a beneficial adaptation to myocardial ischemia and peripheral vascular insufficiency.

Postnatal neovascularization has been thought to occur by a sprouting mechanism, that is, angiogenesis. While this process has been well studied and characterized during development, additional modes of vascular expansion have been recently proposed. In particular, the contribution of circulating endothelial cell progenitors (ECPs) has been clearly documented in a wide array of models and the existence of "resident-cell progenitors" that contribute to the growth of vessels upon injury has also been recognized. In addition, it appears that a mechanism of intraluminal bridging of vessels, known as *intussusception*, has also been noted as a mode for vascular expansion. These processes are discussed in greater length in the following sections.

6.1 Circulating Endothelial Cell Progenitors

The growth of blood vessels in adult animals has been traditionally considered to occur only as the result of angiogenesis, sprouting from preexisting vessels, rather than *de novo* endothelial recruitment from progenitor cells. Recent studies, however, have documented the existence of bone marrow–derived endothelial cells that participate in adult vascular growth, particularly in the context of rapidly generated tumor vasculature. Characterization of these circulating endothelial precursors has demonstrated that they are likely hematopoietic in origin and are recruited by VEGF signals to home in sites of vascular growth.

Similar to embryonic progenitors, adult bone marrow–derived endothelial precursors, including Sca1, appear to share antigenic determinants with hematopoietic progenitors. They also respond to similar factors, prominent among them is VEGF, which mobilizes Sca1+ cells, and stimulates their proliferation and migration, and probably their differentiation. ECPs have been detected in the peripheral blood and have the capacity to differentiate *in vivo* and *in vitro*. Their incorporation into sites of neovascularization has also been documented. Specifically, bone marrow–derived endothelial cell precursors (BMD-ECP) have been shown to participate in neoangiogenesis after postmyocardial ischemia, limb ischemia, wound healing, atherosclerosis, endothelialization of vascular grafts, retinal neovascularization, and cardiac impaired neovascularization during aging. Furthermore, the growth of at least certain tumors appears to be dependent on the recruitment of ECPs from bone marrow. Transplantation of wild-type marrow into *Id3/4* knockout mouse was reported to rescue the otherwise impaired tumor neovascularization of *Id3/4* deficient animals. At least in this model system, tumor vascularization was dependent on bone marrow–derived endothelial cells.

Contributions of bone marrow cells to the vascular smooth muscle cell compartment have been examined as well. Smooth muscle cells from arterioles in the infarcted heart were found to have bone marrow origin. Furthermore, smooth muscle–like cells can be derived from cultures flk-positive ES cells. These studies indicate that bone marrow cells can contribute to smooth muscle cell formation in pathological conditions, but do not address their potential role in normal vascular formation.

While investigators have agreed on their presence, the relative contribution of vascular progenitors to sites of neovascularization in the adult has been a point of heated debate, mostly because the assessments are not equivalent. Engraftment of ECPs, either from blood circulation or adult tissues have been reported to range from 3.5 to 10%. The information contrasts other studies that indicate progenitor contribution of up to 95%. Again, the source of bone marrow (or purified subpopulation) and the experimental models have been variable, therefore, this question remains to be answered.

If, indeed, the contribution of bone marrow–derived progenitors to the growth of the tumor vasculature is significant, manipulation of the ECP population might result in suppression of angiogenesis and tumor expansion. Consequently, several groups have now initiated significant efforts to ascertain the potential value of ECPs for therapeutic intervention. However, progress in this arena has been hampered by lack of a solid means to identify these cells.

The cell-surface makeup and full characterization of ECPs is still unclear and the subject of active investigation. Experiments using CD34+ sorted precursors have shown that ECPs display this marker, both while in the bone marrow and also in the circulation. In contrast, other studies have reported the presence of tissue-resident CD34 negative cells that can contribute to endothelial cells. These latter studies were performed with local progenitors, so it is possible that there are indeed multiple precursors, some bone marrow–derived; and some resident. More consistently, CD34+, AC133 (CD133 or prominin A in mouse) and VEGFR2 are the markers most agreed upon in the literature. Why has it been difficult to

fully characterize this population of bone marrow–derived cells? Models that can identify these cells are currently missing, and the isolation of ECPs using CD34 is challenging, with low yield. Other investigators have used a combination of markers, yet the proof of principle for their identity relies on vascular integration. This adds another layer of difficulty, as once they become incorporated, the differentiation pathways are initiated and the makeup of the precursor population is lost. Thus, the transient nature of the phenotype has hindered progress in this area. Clearly, genetic markers that could trace this population of cells can add a strong advantage to this research.

6.2 Vascular Intussusception

The term *intussusception* is used to describe the process of intravascular growth by the formation of transendothelial cell bridges. These bridges are subsequently reinforced by tissue columns that grow into the lumen of these vessels. Subsequently the bridges gain mural cells, that is, pericytes and/or smooth muscle components, and undergo remodeling. Through this process a large vessel can branch, or be subdivided into smaller daughter vessels. Little information is available in the literature about this process. At this point, it is difficult to ascertain the frequency of intussusception *in vivo* and its relevance to vascular expansion.

7 Mechanical Forces and Angiogenesis

Mechanical factors, such as shear stress, wall tension, and stretch, have long been implicated in vascular growth. Ligation of vessels and subsequent interruption in flow results in severe disruption in normal vascular patterning and morphogenesis. However, caveats in the interpretation of these experiments include the possible contribution of oxygenation.

Technological advances have allowed the application of laminar, pulsatile, and circular shear stress on cultured endothelial cells to study the effect of these forces on gene expression. There is increasing experimental evidence that shear stress, and other consequences of vascular activity, modulate gene expression and that different types of stress activate distinct gene programs. Stretch of cardiomyocytes activates several second messenger pathways including Ras/MAP kinase, tyrosine kinases, and protein kinase C. Stretch has been shown to activate all three MAP kinase family members. Furthermore, it has been shown that stretching of skeletal muscle results in transcriptional activation of VEGF.

In vivo evidence also indicates a role for stretching in angiogenesis. Stretching the left ventricle for 30 min via an intraventricular balloon resulted in nearly a sixfold increase in VEGF message level. Stretching of endothelial cells also precipitates changes that activate and facilitate angiogenesis. They include rearrangement of the cytoskeleton and the release of plasminogen activators and metalloproteinases, which disrupt the basement membrane thus enabling cell migration, proliferation, and activation of ion channels with consequent mobilization of calcium. These factors modulate the relative generation of cellular force by the cytoskeleton and integrins, which combined with the relative rigidity offered by the matrix results in either thinner or wider capillary tubes *in vitro*. Clearly, a concrete understanding of how mechanical

forces are interpreted by cells during angiogenesis is precarious; however, this is an active area of investigation that is likely to reveal important functions of hemodynamics forces in the vasculature.

8 Angiogenesis Inhibitors

The concept that tumor progression could be regulated by pharmacological and/or genetic suppression of blood vessel growth has engendered a long-standing interest in the identification of molecules or synthetic compounds that block angiogenesis. Among recognized angiogenesis inhibitors are: platelet factor 4 (PF4), thrombospondin-1 (TSP1) and 2 (TSP2), angiostatin, endostatin, pigment epithelial-derived factor, and more recently proteolytic fragments of type IV collagen. Interference in the signaling pathways of specific stimulators, such as VEGF/VPF (vascular permeability factor) has proven effective in blocking blood vessel growth. In addition, blockage of integrin binding, in particular, $\alpha_v ß_3$, has been shown to suppress angiogenesis in several tumor types.

Antiangiogenesis has been appealing for cancer therapy for three main reasons: (1) it is likely that most tumors are dependent on angiogenesis, thereby providing a common target in the treatment of widely heterogeneous disease; (2) endothelial cells are considered to be less likely to develop adaptations to bypass drug effects (i.e. drug-resistant phenotype, as seen in some tumors); (3) it is anticipated that tumor vessels are proliferative providing a differential target than the quiescent vessels present in normal tissues.

Results from angiogenesis research during the last decade have questioned some of those earlier assumptions. Clinical trials designed to test antiangiogenic therapy have met with variable and rather disappointing success. In particular, not all tumor vessels appear equally susceptible to a single modality of antiangiogenic therapy. The reasons for this outcome are likely multiple. For example, the participation of tumor cells to the vascular wall (''vascular mimicry'') has been demonstrated in uveal carcinoma, and more recently in melanoma. How responsive are these cells to antiangiogenic therapy? The molecular nature of tumor microvessels appears to be more variable than anticipated and differences in the tumor microenvironment will likely influence therapeutic outcome. It has been well acknowledged that the ''context'' poses a direct (and reciprocal) effect in the tumor cells. Additional reasons for the ''disconnect'' between bench and clinical results have been attributed to poor design of trials and efficacy of some of the tested drugs. More recently, the evaluation of ''combination therapy'' (antiangiogenic and ''metronomic'' chemotherapy) has received experimental attention with a higher degree of success than either therapy alone. Regardless, the relative success of therapeutic approaches using VEGF has engendered hope for this modality of therapy both in cancer (to block vessels) and in cardiac ischemia (to enhance angiogenesis).

VEGF has shown experimental success in animal models to treat vascular insufficiency in the heart and in peripheral blood vessels. Accordingly, clinical studies have been undertaken in which VEGF has been administered as recombinant protein or expressed from plasmids or viral vectors for stimulatory reasons. The situation is particularly complex in diabetic patients, where the potential benefits of

proangiogenic therapy for peripheral vascular disease must be weighed against the risks of exacerbating the progression of "diabetic retinopahy," a condition characterized by excessive vascular proliferation. Inhibition of VEGF signaling has been used for suppression of vascular growth during cancer progression. In addition, the role of VEGF in other processes, such as atherosclerosis, arthritis, and retinopathies has engendered a lot of interest in attempting to block this signaling pathway in pathologies other than tumors. The efficacy of these treatments is currently under evaluation. It is clear that understanding the control of cellular events associated with VEGF-dependent angiogenesis will facilitate the rational design of therapeutic interventions.

Acknowledgment

I would like to acknowledge members of my laboratory for stimulating discussions and thank Dr. Tim Lane for important inputs and suggestions.

Bibliography

Books and Reviews

Alitalo, K., Carmeliet, P. (2002) Molecular mechanisms of lymphangiogenesis in health and disease, *Cancer Cell* **1**, 219–227.

Alva, A.J., Iruela-Arispe, M.L. (2004) Notch signaling in vascular morphogenesis, *Curr. Opin. Hematol.* **11**, 278–283.

Asahara, T., Isner, J.M. (2002) Endothelial progenitor cells for vascular regeneration, *J. Hematother. Stem. Cell Res.* **2**, 171–178.

Asahara, T., Kalka, C., Isner, J.M. (2000) Stem cell therapy and gene transfer for regeneration, *Gene Ther.* **6**, 451–457.

Asahara, T., Murohara, T., Sullivan, A., Silver, M., van der Zee, R., Li, T., Witzenbichler, B., Schatteman, G., Isner, J.M. (1997) Isolation of putative progenitor endothelial cells for angiogenesis, *Science* **275**, 964–967.

Betsholtz, C., Karlsson, L., Lindahl, P. (2001) Developmental roles of platelet-derived growth factors, *BioEssays* **23**, 494–507.

Bissell, M.J., Radisky, D. (2001) Putting tumours in context, *Nat. Rev. Cancer* **1**, 46–54.

Carmeliet, P. (2000) Mechanisms of angiogenesis and arteriogenesis, *Nat. Med.* **6**, 389–395.

Chen, Y.G., Hata, A., Lo, R.S., Wotton, D., Shi, Y., Pavletich, N., Massague, J. (1998) Determinants of specificity in TGF-beta signal transduction, *Genes Dev.* **12**, 2144–2152.

Cheng, N., Brantley, D.M., Chen, J. (2002) The ephrins and Eph receptors in angiogenesis, *Cytokine Growth Factor Rev.* **13**, 75–85.

Conway, E.M., Collen, D., Carmeliet, P. (2001) Molecular mechanisms of blood vessel growth, *Cardiovasc. Res.* **49**, 507–521.

Eichmann, A., Le, Noble, F., Autiero, M., Carmeliet, P. (2005) Guidance of vascular and neural network formation, *Curr. Opin. Neurobiol.* **15**, 108–115.

Ema, M., Rossant, J. (2003) Cell fate decisions in early blood vessel formation, *Trends Cardiovasc. Med.* **13**, 254–259.

Ferrara, N. (2002) VEGF and the quest for tumour angiogenesis factors, *Nat. Rev. Cancer* **10**, 795–703.

Ferrara, N., Davis-Smyth, T. (1997) The biology of vascular endothelial growth factor, *Endocr. Rev.* **18**, 42–45.

Ferrara, N., Alitalo, K. (1999) Clinical applications of angiogenic growth factors and their inhibitors, *Nat. Med.* **5**, 1359–1364.

Gale, N.W., Yancopoulos, G.D. (1997) Ephrins and their receptors, a repulsive topic? *Cell Tissue Res.* **290**, 227–241.

Glagov, S., Zarins, C., Giddens, D.P., Ku, D.N. (1988) Hemodynamics and atherosclerosis. Insights and perspectives gained from studies of human arteries, *Arch. Pathol. Lab. Med.* **112**, 1018–1031.

Godin, I., Cumano, A. (2002) The hare and the tortoise: an embryonic haematopoietic race, *Nat. Rev. Immunol.* **2**, 593–604.

Hanahan, D., Folkman, J. (1996) Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis, *Cell* **86**, 353–364.

Hendrix, M.J., Seftor, R.E., Seftor, E.A., Gruman, L.M., Lee, L.M., Nickoloff, B.J.,

Miele, L., Sheriff, D.D., Schatteman, G.C. (2002). Transendothelial function of human metastatic melanoma cells: role of the microenvironment in cell-fate determination, *Cancer Res.* **62**(3), 665–668.

Holder, N., Klein, R. (1999) Eph receptors and ephrins, effectors of morphogenesis, *Development* **126**, 2033–2044.

Hood, J.D., Cheresh, D.A. (2002) Role of integrins in cell invasion and migration, *Nat. Rev. Cancer* **2**, 91–100.

Iruela-Arispe, M.L., Dvorak, H.F. (1997) Angiogenesis: a dynamic balance of stimulators and inhibitors, *Thromb Haemost.* **1**, 672–677.

Jones, N., Dumont, D.J. (2000) Tek/Tie2 signaling, new and old partners, *Cancer Metastasis Rev.* **19**, 13–17.

Kageyama, R., Ohtsuka, T. (1999) The Notch-Hes pathway in mammalian neural development, *Cell Res.* **9**, 179–188.

Kerbel, R., Folkman, J. (2002) Clinical translation of angiogenesis inhibitors, *Nat. Rev. Cancer* **10**, 727–739.

Lebrin, F., Deckers, M., Bertolino, P., Ten Dijke, P. (2005) TGF-beta receptor function in the endothelium, *Cardiovasc. Res.* **65**, 599–608.

Oliver, G., Detmar, M. (2002) The rediscovery of the lymphatic system: old and new insights into the development and biological function of the lymphatic vasculature, *Genes Dev.* **16**, 773–783.

Orioli, D., Klein, R. (1997) The Eph receptor family, axonal guidance by contact repulsion, *Trends Genet.* **13**, 354–359.

Petrova, T.V., Makinen, T., Alitalo, K. (1999) Signaling via vascular endothelial growth factor receptors, *Exp. Cell Res.* **253**, 117–130.

Rafii, S., Lyden, D., Benezra, R., Hattori, K., Heissig, B. (2002) Vascular and haematopoietic stem cells: novel targets for anti-angiogenesis therapy? *Nat. Rev. Cancer* **11**, 826–835.

Risau, W., Flamme, I. (1995) Vasculogenesis, *Annu. Rev. Cell Dev. Biol.* **11**, 73–91.

Robinson, C.J., Stringer, S.E. (2001) The splice variants of vascular endothelial growth factor (VEGF) and their receptors, *J. Cell Sci.* **114**, 853–865.

Saharinen, P., Tammela, T., Karkkainen, M.J., Alitalo, K. (2004) Lymphatic vasculature: development, molecular regulation and role in tumor metastasis and inflammation, *Trends Immunol.* **25**, 387–395.

Shima, D.T., Mailhos, C. (2000) Vascular developmental biology, getting nervous, *Curr. Opin. Genet. Dev.* **10**, 536–542.

Struhl, G., Adachi, A. (1998) Nuclear access and action of notch in vivo, *Cell* **93**, 649–660.

Stupack, D.G., Cheresh, D.A. (2002) Get a ligand, get a life: integrins, signaling and cell survival, *J. Cell Sci.* **19**, 729–738.

Weinstein, B.M. (2005) Vessels and nerves: marching to the same tune, *Cell* **120**, 299–302.

Yu, T.W., Bargmann, C.I. (2001) Dynamic regulation of axon guidance, *Nat. Neurosci.* **4**(Suppl. 1) 1169–1176.

Primary Literature

Adams, R.H., Wilkinson, G.A., Weiss, C., Diella, F., Gale, N.W., Deutsch, U., Risau, W., Klein, R. (1999) Roles of ephrinB ligands and EphB receptors in cardiovascular development: demarcation of arterial/venous domains, vascular morphogenesis, and sprouting angiogenesis, *Genes Dev.* **13**, 295–306.

Artavanis-Tsakonas, S., Rand, M.D., Lake, R.J. (1999) Notch signaling: cell fate control and signal integration in development, *Science* **284**, 770–776.

Arthur, H.M., Ure, J., Smith, A.J., Renforth, G., Wilson, D.I., Torsney, E., Charlton, R., Parums, D.V., Jowett, T., Marchuk, D.A., Burn, J., Diamond, A.G. (2000) Endoglin, an ancillary TGF-beta receptor, is required for extraembryonic angiogenesis and plays a key role in heart development, *Dev. Biol.* **217**, 42–53.

Asahara, T., Murohara, A., Sullivan Silver, M., van der Zee, R., Li, T., Witzenbichler, B., Schatteman, G., Isner, J.M. (1997) Isolation of putative progenitor endothelial cells for angiogenesis, *Science* **275**, 964–967.

Basson, C.T., Kocher, O., Basson, M.D., Asis, A., Madri, J.A. (1992) Differential modulation of vascular cell integrin and extracellular matrix expression in vitro by TGF-beta 1 correlates with reciprocal effects on cell migration, *J. Cell. Physiol.* **153**, 118–128.

Behar, O., Golden, J.A., Mashimo, H., Schoen, F.J., Fishman, M.C. (1996) Semaphorin III is needed for normal patterning and growth of nerves, bones and heart, *Nature* **383**, 525–528.

Bertrand, J.Y., Giroux, S., Golub, R., Klaine, M., Jalil, A., Boucontet, L., Godin, I., Cumano, A. (2005) Characterization of

purified intraembryonic hematopoietic stem cells as a tool to define their site of origin, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 134–139.

Bhattacharya, V., McSweeney, P.A., Shi, Q., Bruno, B., Ishida, A., Nash, R., Storb, R.F., Sauvage, L.R., Hammond, W.P., Wu, M.H. (2000) Enhanced endothelialization and microvessel formation in polyester grafts seeded with CD34(+) bone marrow cells, *Blood* **95**, 581–585.

Brekken, R.A., Overholser, J.P., Stastny, V.A., Waltenberger, J., Minna, J.D., Thorpe, P.E. (2000) Selective inhibition of vascular endothelial growth factor (VEGF) receptor 2 (KDR/Flk-1) activity by a monoclonal anti-VEGF antibody blocks tumor growth in mice, *Cancer Res.* **60**, 5117–5124.

Carmeliet, P., Ferreira, V., Breier, G., Pollefeyt, S., Kieckens, L., Gertsenstein, M., Fahrig, M., Vandenhoeck, A., Harpal, K., Eberhardt, C., Declercq, C., Pawling, J., Moons, L., Collen, D., Risau, W., Nagy, A. (1996) Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele, *Nature* **380**, 435–439.

Chen, H., Ikeda, U., Shimpo, M., Maeda, Y., Shibuya, M., Ozawa, K., Shimada, K. (2000) Inhibition of vascular endothelial growth factor activity by transfection with the soluble FLT-1 gene, *J. Cardiovasc. Pharmacol.* **36**, 498–502.

Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J.C., Keller, G. (1998) A common precursor for hematopoietic and endothelial cells, *Development* **125**, 725–732.

Chung, Y.S., Zhang, W.J., Arentson, E., Kingsley, P.D., Palis, J., Choi, K. (2002) Lineage analysis of the hemangioblast as defined by FLK1 and SCL expression, *Development* **129**, 5511–5520.

Clauss, M., Weich, H., Breier, G., Knies, U., Rockl, W., Waltenberger, J., Risau, W. (1996) The vascular endothelial growth factor receptor Flt-1 mediates biological activities. Implications for a functional role of placenta growth factor in monocyte activation and chemotaxis, *J. Biol. Chem.* **271**, 17629–17634.

Conlon, R.A., Reaume, A.G., Rossant, J. (1995) Notch1 is required for the coordinate segmentation of somites, *Development* **121**, 1533–1545.

Crosby, J.R., Kaminski, W.E., Schatteman, G., Martin, P.J., Raines, E.W., Seifert, R.A., Bowen-Pope, D.F. (2000) Endothelial cells of hematopoietic origin make a significant contribution to adult blood vessel formation, *Circ. Res.* **87**, 728–730.

Davis, R.L., Turner, D.L. (2001) Vertebrate hairy and Enhancer of split related proteins: transcriptional repressors regulating cellular differentiation and embryonic patterning, *Oncogene* **20**, 8342–8357.

Davis, S., Aldrich, T.H., Jones, P.F., Acheson, A., Compton, D.L., Jain, V., Ryan, T.E., Bruno, J., Radziejewski, C., Maisonpierre, P.C., Yancopoulos, G.D. (1996) Isolation of angiopoietin-1, a ligand for the TIE2 receptor, by secretion-trap expression cloning, *Cell* **87**, 1161–1169.

Dawson, D.W., Volpert, O.V., Gillis, P., Crawford, S.E., Xu, H., Benedict, W., Bouck, N.P. (1999) Pigment epithelium-derived factor: a potent inhibitor of angiogenesis, *Science* **5425**, 245–248.

Detmar, M., Brown, L.F., Schon, M.P., Elicker, B.M., Velasco, P., Richard, L., Fukumura, D., Monsky, W., Claffey, K.P., Jain, R.K. (1998) Increased microvascular density and enhanced leukocyte rolling and adhesion in the skin of VEGF transgenic mice, *J. Invest. Dermatol.* **111**, 1–6.

Dickson, M.C., Martin, J.S., Cousins, F.M., Kulkarni, A.B., Karlsson, S., Akhurst, R.J. (1995) Defective haematopoiesis and vasculogenesis in transforming growth factor-beta 1 knock out mice, *Development* **121**, 1845–1854.

Djonov, V., Schmid, M., Tschanz, S.A., Burri, P.H. (2000) Intussusceptive angiogenesis: its role in embryonic vascular network formation, *Circ. Res.* **86**, 286–292.

Duarte, A., Hirashima, M., Benedito, R., Trindade, A., Diniz, P., Bekman, E., Costa, L., Henrique, D., Rossant, J. (2004) Dosage-sensitive requirement for mouse Dll4 in artery development, *Genes Dev.* **18**, 2474–2478.

Dumont, D.J., Yamaguchi, T.P., Conlon, R.A., Rossant, J., Breitman, M.L. (1992) Tek, a novel tyrosine kinase gene located on mouse chromosome 4, is expressed in endothelial cells and their presumptive precursors, *Oncogene* **7**, 1471–1480.

Dumont, D.J., Fong, G.H., Puri, M.C., Gradwohl, G., Alitalo, K., Breitman, M.L. (1995) Vascularization of the mouse embryo: a study of flk-1, tek, tie, and vascular endothelial growth factor expression during development, *Dev. Dyn.* **203**, 80–92.

Dumont, D.J., Gradwohl, G., Fong, G.H., Puri, M.C., Gertsenstein, M., Auerbach, A., Breitman, M.L. (1994) Dominant-negative and targeted null mutations in the endothelial receptor tyrosine kinase, tek, reveal a critical role in vasculogenesis of the embryo, *Genes Dev.* **8**, 1897–1909.

Dumont, D.J., Jussila, L., Taipale, J., Lymboussaki, A., Mustonen, T., Pajusola, K., Breitman, M., Alitalo, K. (1998) Cardiovascular failure in mouse embryos deficient in VEGF receptor-3, *Science* **282**, 946–949.

Dunker, N., Krieglstein, K. (2000) Targeted mutations of transforming growth factor-beta genes reveal important roles in mouse development and adult homeostasis, *Eur. J. Biochem.* **267**, 6982–6988.

Edelberg, J.M., Tang, L., Hattori, K., Lyden, D., Rafii, S. (2002) Young adult bone marrow-derived endothelial precursor cells restore aging-impaired cardiac angiogenic function, *Circ. Res.* **10**, 89–93.

Elefanty, A.G., Begley, C.G., Metcalf, D., Barnett, L., Kontgen, F., Robb, L. (1998) Characterization of hematopoietic progenitor cells that express the transcription factor SCL, using a lacZ "knock-in" strategy, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11,897–11,902.

Ema, M., Faloon, P., Zhang, W.J., Hirashima, M., Reid, T., Stanford, W.L., Orkin, S., Choi, K., Rossant, J. (2003) Combinatorial effects of Flk1 and Tal1 on vascular and hematopoietic development in the mouse, *Genes Dev.* **17**, 380–393.

Esser, S., Wolburg, K., Wolburg, H., Breier, G., Kurzchalia, T., Risau, W. (1998) Vascular endothelial growth factor induces endothelial fenestrations in vitro, *J. Cell Biol.* **140**, 947–959.

Feiner, L., Webber, A.L., Brown, C.B., Lu, M.M., Jia, L., Feinstein, P., Mombaerts, P., Epstein, J.A., Raper, J.A. (2001) Targeted disruption of semaphorin 3C leads to persistent truncus arteriosus and aortic arch interruption, *Development* **128**, 3061–3070.

Ferrara, N., Carver-Moore, K., Chen, H., Dowd, M., Lu, L., O'Shea, K.S., Powell-Braxton, L., Hillan, K.J., Moore, M.W. (1996) Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene, *Nature* **380**, 439–442.

Fong, G.H., Rossant, J., Gertsenstein, M., Breitman, M.L. (1995) Role of the Flt-1 receptor tyrosine kinase in regulating the assembly of vascular endothelium, *Nature* **376**, 66–70.

Fong, G.H., Zhang, L., Bryce, D.M., Peng, J. (1999) Increased hemangioblast commitment, not vascular disorganization, is the primary defect in flt-1 knock-out mice, *Development* **126**, 3015–3025.

Fortini, M.E., Artavanis-Tsakonas, S. (1994) The suppressor of hairless protein participates in notch receptor signaling, *Cell* **79**, 273–282.

Francis, S.E., Goh, K.L., Hodivala-Dilkes, K., Bader, B.L., Stark, M., Davidson, D., Hynes, R.O. (2002) Central roles of alpha5beta1 integrin and fibronectin in vascular development in mouse embryos and embryoid bodies, *Arterioscler. Thromb. Vasc. Biol.* **22**, 927–933.

Fukai, N., Eklund, L., Marneros, A.G., Oh, S.P., Keene, D.R., Tamarkin, L., Niemela, M., Ilves, M., Li, E., Pihlajaniemi, T., Olsen, B.R. (2002) Lack of collagen XVIII/endostatin results in eye abnormalities, *EMBO J.* **21**, 1535–1544.

Fuller, T., Korff, T., Kilian, A., Dandekar, G., Augustin, H.G. (2003) Forward EphB4 signaling in endothelial cells controls cellular repulsion and segregation from ephrinB2 positive cells, *J. Cell Sci.* **116**, 2461–2470.

Gale, N.W., Yancopoulos, G.D. (1999) Growth factors acting via endothelial cell-specific receptor tyrosine kinases, VEGFs, angiopoietins, and ephrins in vascular development, *Genes Dev.* **13**, 1055–1066.

Gale, N.W., Dominguez, M.G., Noguera, I., Pan, L., Hughes, V., Valenzuela, D.M., Murphy, A.J., Adams, N.C., Lin, H.C., Holash, J., Thurston, G., Yancopoulos, G.D. (2004) Haploinsufficiency of delta-like 4 ligand results in embryonic lethality due to major defects in arterial and vascular development, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15949–15954.

Gale, N.W., Thurston, G., Hackett, S.F., Renard, R., Wang, Q., McClain, J., Martin, C., Witte, C., Witte, M.H., Jackson, D., Suri, C., Campochiaro, P.A., Wiegand, S.J., Yancopoulos, G.D. (2002) Angiopoietin-2 is required for postnatal angiogenesis and lymphatic patterning, and only the latter role is rescued by Angiopoietin-1, *Dev. Cell* **3**, 411–423.

Garcia-Porrero, J.A., Godin, I.E., Dieterlen-Lievre, F. (1995) Potential intraembryonic hemogenic sites at pre-liver stages in the mouse, *Anat. Embryol. (Berl.)* **192**, 425–435.

Gehling, U.M., Ergun, S., Schumacher, U., Wagener, C., Pantel, K., Otte, M., Schuch, G., Schafhausen, P., Mende, T., Kilic, N., Kluge,

K., Schafer, B., Hossfeld, D.K., Fiedler, W. (2000) In vitro differentiation of endothelial cells from AC133-positive progenitor cells, *Blood* **95**, 106–112.

Gerber, H.P., Condorelli, F., Park, J., Ferrara, N. (1997) Differential transcriptional regulation of the two vascular endothelial growth factor receptor genes. Flt-1, but not Flk-1/KDR, is up-regulated by hypoxia, *J. Biol. Chem.* **272**, 23659–23667.

Gerber, H.P., Hillan, K.J., Ryan, A.M., Kowalski, J., Keller, G.A., Rangell, L., Wright, B.D., Radtke, F., Aguet, M., Ferrara, N. (1999) VEGF is required for growth and survival in neonatal mice, *Development* **126**, 1149–1159.

Gerhardt, H., Golding, M., Fruttiger, M., Ruhrberg, C., Lundkvist, A., Abramsson, A., Jeltsch, M., Mitchell, C., Alitalo, K., Shima, D., Betsholtz, C. (2003) VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia, *J. Cell Biol.* **161**, 1163–1177.

Gerety, S.S., Anderson, D.J. (2002) Cardiovascular ephrinB2 function is essential for embryonic angiogenesis, *Development* **129**, 1397–1410.

Gerety, S.S., Wang, H.U., Chen, Z.F., Anderson, D.J. (1999) Symmetrical mutant phenotypes of the receptor EphB4 and its specific transmembrane ligand ephrin-B2 in cardiovascular development, *Mol. Cell* **4**, 403–414.

Gitler, A.D., Lu, M.M., Epstein, J.A. (2004) PlexinD1 and semaphorin signaling are required in endothelial cells for cardiovascular development, *Dev. Cell* **7**, 107–116.

Gluzman-Poltorak, Z., Cohen, T., Herzog, Y., Neufeld, G. (2000) Neuropilin-2 is a receptor for the vascular endothelial growth factor (VEGF) forms VEGF-145 and VEGF-165, *J. Biol. Chem.* **275**, 18040–18045; Erratum (2000) *J. Biol. Chem.* **275**, 29922.

Gluzman-Poltorak, Z., Cohen, T., Shibuya, M., Neufeld, G. (2001) Vascular endothelial growth factor receptor-1 and neuropilin-2 form complexes, *J. Biol. Chem.* **276**, 18688–18694.

Goumans, M.J., Valdimarsdottir, G., Itoh, S., Lebrin, F., Larsson, J., Mummery, C., Karlsson, S., ten Dijke, P. (2003) Activin receptor-like kinase (ALK)1 is an antagonistic mediator of lateral TGFβ/ALK5 signaling, *Mol. Cell* **12**, 817–828.

Gu, C., Yoshida, Y., Livet, J., Reimert, D.V., Mann, F., Merte, J., Henderson, C.E., Jessell, T.M., Kolodkin, A.L., Ginty, D.D. (2005) Semaphorin 3E and plexin-D1 control vascular pattern independently of neuropilins, *Science* **307**, 265–268.

Hamaguchi, I., Huang, X.L., Takakura, N., Tada, J., Yamaguchi, Y., Kodama, H., Suda, T. (1999) In vitro hematopoietic and endothelial cell development from cells expressing TEK receptor in murine aorta-gonad-mesonephros region, *Blood* **93**, 1549–1556.

Hamano, Y., Zeisberg, M., Sugimoto, H., Lively, J.C., Maeshima, Y., Yang, C., Hynes, R.O., Werb, Z., Sudhakar, A., Kalluri, R. (2003) Physiological levels of tumstatin, a fragment of collagen IV alpha3 chain, are generated by MMP-9 proteolysis and suppress angiogenesis via alphaV beta3 integrin, *Cancer Cell* **3**, 589–601.

Hattori, K., Dias, S., Heissig, B., Hackett, N.R., Lyden, D., Tateno, M., Hicklin, D.J., Zhu, Z., Witte, L., Crystal, R.G., Moore, M.A., Rafii, S. (2001) Vascular endothelial growth factor and angiopoietin-1 stimulate postnatal hematopoiesis by recruitment of vasculogenic and hematopoietic stem cells, *J. Exp. Med.* **9**, 1005–1014.

He, Z., Tessier-Lavigne, M. (1997) Neuropilin is a receptor for the axonal chemorepellent semaphorin III, *Cell* **90**, 739–751.

Heimark, R.L., Twardzik, D.R., Schwartz, S.M. (1986) Inhibition of endothelial regeneration by type-beta transforming growth factor from platelets, *Science* **233**, 1078–1080.

Hellstrom, M., Kalen, M., Lindahl, P., Abramsson, A., Betsholtz, C. (1999) Role of PDGF-B and PDGFR-beta in recruitment of vascular smooth muscle cells and pericytes during embryonic blood vessel formation in the mouse, *Development* **126**, 3047–3055.

Herzog, Y., Kalcheim, C., Kahane, N., Reshef, R., Neufeld, G. (2001) Differential expression of neuropilin-1 and neuropilin-2 in arteries and veins, *Mech. Dev.* **109**, 115–119.

Hiratsuka, S., Minowa, O., Kuno, J., Noda, T., Shibuya, M. (1998) Flt-1 lacking the tyrosine kinase domain is sufficient for normal development and angiogenesis in mice, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 934954.

Irrthum, A., Karkkainen, M.J., Devriendt, K., Alitalo, K., Vikkula, M. (2000) Congenital hereditary lymphedema caused by a mutation that inactivates VEGFR3 tyrosine kinase, *Am. J. Hum. Genet.* **67**, 295–301.

Iruela-Arispe, M.L., Sage, E.H. (1993) Endothelial cells exhibiting angiogenesis in

vitro proliferate in response to TGF-beta 1, *J. Cell. Biochem.* **52**, 414–430.

Isogai, S., Lawson, N.D., Torrealday, S., Horiguchi, M., Weinstein, B.M. (2003) Angiogenic network formation in the developing vertebrate trunk, *Development* **130**, 5281–5290.

Jackson, K.A., Majka, S.M., Wang, H., Pocius, J., Hartley, C.J., Majesky, M.W., Entman, M.L., Michael, L.H., Hirschi, K.K., Goodell, M.A. (2001) Regeneration of ischemic cardiac muscle and vascular endothelium by adult stem cells, *J. Clin. Invest.* **11**, 302–395.

Johnson, D.W., Berg, J.N., Baldwin, M.A., Gallione, C.J., Marondel, I., Yoon, S.J., Stenzel, T.T., Speer, M., Pericak-Vance, M.A., Diamond, A., Guttmacher, A.E., Jackson, C.E., Attisano, L., Kucherlapati, R., Porteous, M.E., Marchuk, D.A. (1996) Mutations in the activin receptor-like kinase 1 gene in hereditary haemorrhagic telangiectasia type 2, *Nat. Genet.* **13**, 189–195.

Jones, N., Dumont, D.J. (2000) Tek/Tie2 signaling, new and old partners, *Cancer Metastasis Rev.* **19**, 13–17.

Jones, N., Master, Z., Jones, J., Bouchard, D., Gunji, Y., Sasaki, H., Daly, R., Alitalo, K., Dumont, D.J. (1999) Identification of Tek/Tie2 binding partners. Binding to a multifunctional docking site mediates cell survival and migration, *J. Biol. Chem.* **274**, 30896–30905.

Joutel, A., Corpechot, C., Ducros, A., Vahedi, K., Chabriat, H., Mouton, P., Alamowitch, S., Domenga, V., Cecillion, M., Marechal, E., Maciazek, J., Vayssiere, C., Cruaud, C., Cabanis, E.A., Ruchoux, M.M., Weissenbach, J., Bach, J.F., Bousser, M.G., Tournier-Lasserve, E. (1996) Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia, *Nature* **383**, 707–710.

Kabrun, N., Buhring, H.J., Choi, K., Ullrich, A., Risau, W., Keller, G. (1997) Flk-1 expression defines a population of early embryonic hematopoietic precursors, *Development* **124**, 2039–2048.

Kanno, S., Oda, N., Abe, M., Terai, Y., Ito, M., Shitara, K., Tabayashi, K., Shibuya, M., Sato, Y. (2000) Roles of two VEGF receptors, Flt-1 and KDR, in the signal transduction of VEGF effects in human vascular endothelial cells, *Oncogene* **19**, 2138–2146.

Karkkainen, M.J., Ferrell, R.E., Lawrence, E.C., Kimak, M.A., Levinson, K.L., McTigue, M.A., Alitalo, K., Finegold, D.N. (2000) Missense mutations interfere with VEGFR-3 signalling in primary lymphoedema, *Nat. Genet.* **25**, 153–159.

Karkkainen, M.J., Haiko, P., Sainio, K., Partanen, J., Taipale, J., Petrova, T.V., Jeltsch, M., Jackson, D.G., Talikka, M., Rauvala, H., Betsholtz, C., Alitalo, K. (2004) Vascular endothelial growth factor C is required for sprouting of the first lymphatic vessels from embryonic veins, *Nat. Immunol.* **5**, 74–80.

Kawasaki, T., Kitsukawa, T., Bekku, Y., Matsuda, Y., Sanbo, M., Yagi, T., Fujisawa, H. (1999) A requirement for neuropilin-1 in embryonic vessel formation, *Development* **126**, 4895–4902.

Kendall, R.L., Wang, G., Thomas, K.A. (1996) Identification of a natural soluble form of the vascular endothelial growth factor receptor, FLT-1, and its heterodimerization with KDR, *Biochem. Biophys. Res. Commun.* **226**, 324–328.

Kim, N.H., Oh, J.H., Seo, J.A., Lee, K.W., Kim, S.G., Choi, K.M., Baik, S.H., Choi, D.S., Kang, Y.S., Han, S.Y., Han, K.H., Ji, Y.H., Cha, D.R. (2005) Vascular endothelial growth factor (VEGF) and soluble VEGF receptor FLT-1 in diabetic nephropathy, *Kidney Int.* **67**, 167–177.

Kitsukawa, T., Shimono, A., Kawakami, A., Kondoh, H., Fujisawa, H. (1995) Overexpression of a membrane protein, neuropilin, in chimeric mice causes anomalies in the cardiovascular system, nervous system and limbs, *Development* **121**, 4309–4318.

Kolodkin, A.L., Levengood, D.V., Rowe, E.G., Tai, Y.T., Giger, R.J., Ginty, D.D. (1997) Neuropilin is a semaphorin III receptor, *Cell* **90**, 753–762.

Krebs, L.T., Xue, Y., Norton, C.R., Shutter, J.R., Maguire, M., Sundberg, J.P., Gallahan, D., Closson, V., Kitajewski, J., Callahan, R., Smith, G.H., Stark, K.L., Gridley, T. (2000) Notch signaling is essential for vascular morphogenesis in mice, *Genes Dev.* **14**, 1343–1352.

Kulkarni, A.B., Huh, C.G., Becker, D., Geiser, A., Lyght, M., Flanders, K.C., Roberts, A.B., Sporn, M.B., Ward, J.M., Karlsson, S. (1993) Transforming growth factor beta 1 null mutation in mice causes excessive inflammatory response and early death, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 770–774.

Lacaud, G., Gore, L., Kennedy, M., Kouskoff, V., Kingsley, P., Hogan, C., Carlsson, L.,

Speck, N., Palis, J., Keller, G. (2002) Runx1 is essential for hematopoietic commitment at the hemangioblast stage of development in vitro, *Blood* **100**, 458–466.

Lai, C.M., Brankov, M., Zaknich, T., Lai, Y.K., Shen, W.Y., Constable, I.J., Kovesdi, I., Rakoczy, P.E. (2001) Inhibition of angiogenesis by adenovirus-mediated sFlt-1 expression in a rat model of corneal neovascularization, *Hum. Gene Ther.* **12**, 1299–1310.

Lamouille, S., Mallet, C., Feige, J.J., Bailly, S. (2002) Activin receptor-like kinase 1 is implicated in the maturation phase of angiogenesis, *Blood* **100**, 4495–4501.

Larsson, J., Goumans, M.J., Sjostrand, L.J., van Rooijen, M.A., Ward, D., Leveen, P., Xu, X., ten Dijke, P., Mummery, C.L., Karlsson, S. (2001) Abnormal angiogenesis but intact hematopoietic potential in TGF-beta type I receptor-deficient mice, *EMBO J.* **20**, 1663–1673.

Lawson, N.D., Weinstein, B.M. (2002) Arteries and veins: making a difference with zebrafish, *Nat. Rev. Genet.* **3**, 674–682.

Lawson, N.D., Scheer, N., Pham, V.N., Kim, C.H., Chitnis, A.B., Campos-Ortega, J.A., Weinstein, B.M. (2001) Notch signaling is required for arterial-venous differentiation during embryonic vascular development, *Development* **128**, 3675–3683.

Le Noble, F., Moyon, D., Pardanaud, L., Yuan, L., Djonov, V., Matthijsen, R., Breant, C., Fleury, V., Eichmann, A. (2004) Flow regulates arterial-venous differentiation in the chick embryo yolk sac, *Development* **131**, 361–375.

Letterio, J.J., Geiser, A.G., Kulkarni, A.B., Roche, N.S., Sporn, M.B., Roberts, A.B. (1994) Maternal rescue of transforming growth factor-beta 1 null mice, *Science* **264**, 1936–1938.

Leung, D.W., Cachianes, G., Kuang, W.J., Goeddel, D.V., Ferrara, N. (1989) Vascular endothelial growth factor is a secreted angiogenic mitogen, *Science* **246**, 1306–1309.

Leveen, P., Pekny, M., Gebre-Medhin, S., Swolin, B., Larsson, E., Betsholtz, C. (1994) Mice deficient for PDGF B show renal, cardiovascular, and hematological abnormalities, *Genes Dev.* **8**, 1875–1887.

Li, H., Gu, B., Zhang, Y., Lewis, D.F., Wang, Y. (2005) Hypoxia-induced increase in soluble Flt-1 production correlates with enhanced oxidative stress in trophoblast cells from the human placenta, *Placenta* **26**, 210–217.

Li, D.Y., Sorensen, L.K., Brooke, B.S., Urness, L.D., Davis, E.C., Taylor, D.G., Boak, B.B., Wendel, D.P. (1999) Defective angiogenesis in mice lacking endoglin, *Science* **284**, 1534–1537.

Li, L., Krantz, I.D., Deng, Y., Genin, A., Banta, A.B., Collins, C.C., Qi, M., Trask, B.J., Kuo, W.L., Cochran, J., Costa, T., Pierpont, M.E., Rand, E.B., Piccoli, D.A., Hood, L., Spinner, N.B. (1997) Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1, *Nat. Genet.* **16**, 243–251.

Lyden, D., Hattori, K., Dias, S., Costa, C., Blaikie, P., Butros, L., Chadburn, A., Heissig, B., Marks, W., Witte, L., Wu, Y., Hicklin, D., Zhu, Z., Hackett, N.R., Crystal, R.G., Moore, M.A., Hajjar, K.A., Manova, K., Benezra, R., Rafii, S. (2001) Impaired recruitment of bone-marrow-derived endothelial and hematopoietic precursor cells blocks tumor angiogenesis and growth, *Nat. Med.* **7**, 194–101.

Madri, J.A., Pratt, B.M., Tucker, A.M. (1988) Phenotypic modulation of endothelial cells by transforming growth factor-beta depends upon the composition and organization of the extracellular matrix, *J. Cell Biol.* **106**, 1375–1384.

Mailhos, C., Modlich, U., Lewis, J., Harris, A., Bicknell, R., Ish-Horowicz, D. (2001) Delta4, an endothelial specific notch ligand expressed at sites of physiological and tumor angiogenesis, *Differentiation* **69**, 135–144.

Maisonpierre, P.C., Suri, C., Jones, P.F., Bartunkova, S., Wiegand, S.J., Radziejewski, C., Compton, D., McClain, J., Aldrich, T.H., Papadopoulos, N., Daly, T.J., Davis, S., Sato, T.N., Yancopoulos, G.D. (1997) Angiopoietin-2, a natural antagonist for Tie2 that disrupts in vivo angiogenesis, *Science* **277**, 55–60.

Majka, S.M., Jackson, K.A., Kienstra, K.A., Majesky, M.W., Goodell, M.A., Hirschi, K. (2003) Distinct populations of vascular progenitors in skeletal muscle are derived from bone marrow and exhibit different cell fates during vascular regeneration, *J. Clin. Invest.* **111**, 71–79.

Man, S., Bocci, G., Francia, G., Green, S.K., Jothy, S., Hanahan, D., Bohlen, P., Hicklin, D.J., Bergers, G., Kerbel, R.S. (2002) Antitumor effects in mice of low-dose (metronomic) cyclophosphamide administered continuously

through the drinking water, *Cancer Res.* **10**, 731–735.

McAllister, K.A., Grogg, K.M., Johnson, D.W., Gallione, C.J., Baldwin, M.A., Jackson, C.E., Helmbold, E.A., Markel, D.S., McKinnon, W.C., Murrel, J., McCormick, M.K. (1994) Endoglin, a TGF-â binding protein of endothelial cells, is the gene for hereditary haemorrhagic telangiectasia type 1, *Nat. Genet.* **8**, 345–351.

McCarty, J.H., Monahan-Earley, R.A., Brown, L.F., Keller, M., Gerhardt, H., Rubin, K., Shani, M., Dvorak, H.F., Wolburg, H., Bader, B.L., Dvorak, A.M., Hynes, R.O. (2002) Defective associations between blood vessels and brain parenchyma lead to cerebral hemorrhage in mice lacking alphav integrins, *Mol. Cell. Biol.* **22**, 7667–7677.

Miao, H.Q., Soker, S., Feiner, L., Alonso, J.L., Raper, J.A., Klagsbrun, M. (1999) Neuropilin-1 mediates collapsin-1/semaphorin III inhibition of endothelial cell motility: functional competition of collapsin-1 and vascular endothelial growth factor-165, *J. Cell Biol.* **146**, 233–242.

Miao, H., Wei, B., Peehl, D.M., Li, Q., Alexandrou, T., Scheling, J.R., Rhim, J.S., Sedor, J.R., Burnett, E., Wang, B. (2001) Activation of EphA receptor tyrosine kinase inhibits the Ras/MAPK pathway, *Nat. Cell Biol.* **3**, 527–530.

Millauer, B., Wizigmann-Voos, S., Schnurch, H., Martinez, R., Moller, N.P., Risau, W., Ullrich, A. (1993) High affinity VEGF binding and developmental expression suggest Flk-1 as a major regulator of vasculogenesis and angiogenesis, *Cell* **72**, 835–846.

Miquerol, L., Langille, B.L., Nagy, A. (2000) Embryonic development is disrupted by modest increases in vascular endothelial growth factor gene expression, *Development* **127**, 3941–3946.

Mochizuki, Y., Nakamura, T., Kanetake, H., Kanda, S. (2002) Angiopoietin 2 stimulates migration and tube-like structure formation of murine brain capillary endothelial cells through c-Fes and c-Fyn, *J. Cell Sci.* **115**, 175–183.

Motoike, T., Markham, D.W., Rossant, J., Sato, T.N. (2003) Evidence for novel fate of Flk1+ progenitor: contribution to muscle lineage, *Genesis* **35**, 153–159.

Mukouyama, Y.S., Shin, D., Britsch, S., Tanigucbi, M., Anderson, D.J. (2002) Sensory nerves determine the pattern of arterial differentiation and blood vessel branching in the skin, *Cell* **109**, 693–705.

Neufeld, G., Cohen, T., Shraga, N., Lange, T., Kessler, O., Herzog, Y. (2002) The neuropilins, multifunctional semaphorin and VEGF receptors that modulate axon guidance and angiogenesis, *Trends Cardiovasc. Med.* **12**, 13–19.

North, T., Gu, T.L., Stacy, T., Wang, Q., Howard, L., Binder, M., Marin-Padilla, M., Speck, N.A. (1999) Cbfa2 is required for the formation of intra-aortic hematopoietic clusters, *Development* **126**, 2563–2575.

Oda, T., Elkahloun, A.G., Pike, B.L., Okajima, K., Krantz, I.D., Genin, A., Piccoli, D.A., Meltzer, P.S., Spinner, N.B., Collins, F.S., Chandrasekharappa, S.C. (1997) Mutations in the human Jagged1 gene are responsible for Alagille syndrome, *Nat. Genet.* **16**, 235–242.

Oh, S.P., Seki, T., Goss, K.A., Imamura, T., Yi, Y., Donahoe, P.K., Li, L., Miyazono, K., ten Dijke, P., Kim, S., Li, E. (2000) Activin receptor-like kinase 1 modulates transforming growth factor-beta 1 signaling in the regulation of angiogenesis, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2626–2631.

Orlic, D., Kajstura, J., Chimenti, S., Jakoniuk, I., Anderson, S.M., Li, B., Pickel, J., McKay, R., Nadal-Ginard, B., Bodine, D.M., Leri, A., Anversa, P. (2001) Bone marrow cells regenerate infarcted myocardium, *Nature* **6829**, 701–705.

O'Reilly, M.S., Boehm, T., Shing, Y., Fukai, N., Vasios, G., Lane, W.S., Flynn, E., Birkhead, J.R., Olsen, B.R., Folkman, J. (1997) Endostatin: an endogenous inhibitor of angiogenesis and tumor growth, *Cell* **2**, 277–285.

O'Reilly, M.S., Holmgren, L., Shing, Y., Chen, C., Rosenthal, R.A., Moses, M., Lane, W.S., Cao, Y., Sage, E.H., Folkman, J. (1994) Angiostatin: a novel angiogenesis inhibitor that mediates the suppression of metastases by a Lewis lung carcinoma, *Cell* **2**, 315–318.

Oshima, M., Oshima, H., Taketo, M.M. (1996) TGF-beta receptor type II deficiency results in defects of yolk sac hematopoiesis and vasculogenesis, *Dev. Biol.* **179**, 297–302.

Othman-Hassan, K., Pate, K., Papoutsi, M., Rodriguez-Niedenfuhr, M., Christ, B., Wilting, J. (2001) Arterial identity of endothelial cells is controlled by local cues, *Dev. Biol.* **237**, 398–409.

Palis, J., Yoder, M.C. (2001) Yolk-sac hematopoiesis: the first blood cells of mouse and man, *Exp. Hematol.* **29**, 927–936.

Park, M., Lee, S.T. (1999) The fourth immunoglobulin-like loop in the extracellular domain of FLT-1, a VEGF receptor, includes a major heparin-binding site, *Biochem. Biophys. Res. Commun.* **264**, 730–734.

Partanen, J., Puri, M.C., Schwartz, L., Fischer, K.D., Bernstein, A., Rossant, J. (1996) Cell autonomous functions of the receptor tyrosine kinase TIE in a late phase of angiogenic capillary growth and endothelial cell survival during murine development, *Development* **122**, 3013–3021.

Patan, S., Munn, L.L., Jain, R.K. (1996) Intussusceptive microvascular growth in a human colon adenocarcinoma xenograft: a novel mechanism of tumor angiogenesis, *Microvasc. Res.* **51**, 260–272.

Peichev, M., Naiyer, A.J., Pereira, D., Zhu, Z., Lane, W.J., Williams, M., Oz, M.C., Hicklin, D.J., Witte, L., Moore, M.A., Rafii, S. (2000) Expression of VEGFR-2 and AC133 by circulating human CD34(+) cells identifies a population of functional endothelial precursors, *Blood* **3**, 952–958.

Piccoli, D.A., Spinner, N.B. (2001) Alagille syndrome and the Jagged1 gene, *Semin. Liver Dis.* **21**, 525–534.

Puri, M.C., Rossant, J., Alitalo, K., Bernstein, A., Partanen, J. (1995) The receptor tyrosine kinase TIE is required for integrity and survival of vascular endothelial cells, *EMBO J.* **14**, 5884–5891.

Rabbany, S.Y., Heissig, B., Hattori, K., Rafii, S. (2003) Molecular pathways regulating mobilization of marrow-derived stem cells for tissue revascularization, *Trends Mol. Med.* **3**, 109–117.

Rastinejad, F., Polverini, P.J., Bouck, N.P. (1989) Regulation of the activity of a new inhibitor of angiogenesis by a cancer suppressor gene, *Cell* **3**, 345–355.

Reynolds, L.E., Wyder, L., Lively, J.C., Taverna, D., Robinson, S.D., Huang, X., Sheppard, D., Hynes, R.O., Hodivala-Dilke, K.M. (2002) Enhanced pathological angiogenesis in mice lacking beta3 integrin or beta3 and beta5 integrins, *Nat. Med.* **8**, 27–34.

Reyes, M., Dudek, A., Jahagirdar, B., Koodie, L., Marker, P.H., Verfaillie, C.M. (2002) Origin of endothelial progenitors in human postnatal bone marrow, *J. Clin. Invest.* **109**, 337–346.

Rodriguez-Manzaneque, J.C., Lane, T.F., Ortega, M.A., Hynes, R.O., Lawler, J., Iruela-Arispe, M.L. (2001) Thrombospondin-1 suppresses tumor growth by a novel mechanism that includes blockade of matrix metalloproteinase-9 activation, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12485–12490.

Ruhrberg, C., Gerhardt, H., Golding, M., Watson, R., Ioannidou, S., Fujisawa, H., Betsholtz, C., Shima, D.T. (2002) Spatially restricted patterning cues provided by heparin-binding VEGF-A control blood vessel branching morphogenesis, *Genes Dev.* **16**, 2684–2698.

Ruchoux, M.M., Maurage, C.A. (1997) CADASIL, Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy, *J. Neuropathol. Exp. Neurol.* **56**, 947–964.

Sabin, F.R. (1920) Studies on the origin of blood vessels and of red corpuscles as seen in the living blastoderm of the chick during the second day of incubation, *Contrib. Embryol.* **9**, 213–262.

Salven, P., Mustjoki, S., Alitalo, R., Alitalo, K., Rafii, S. (2003) VEGFR-3 and CD133 identify a population of CD34+ lymphatic/vascular endothelial precursor cells, *Blood* **101**, 168–172.

Sato, T.N., Qin, Y., Kozak, C.A., Audus, K.L. (1993) Tie-1 and tie-2 define another class of putative receptor tyrosine kinase genes expressed in early embryonic vascular system, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9355–9358; Erratum. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1205–1206.

Sato, Y., Kanno, S., Oda, N., Abe, M., Ito, M., Shitara, K., Shibuya, M. (2000) Properties of two VEGF receptors, Flt-1 and KDR, in signal transduction, *Ann. N. Y. Acad. Sci.* **902**, 201–205.

Sato, T.N., Tozawa, Y., Deutsch, U., Wolburg-Buchholz, K., Fujiwara, Y., Gendron-Maguire, M., Gridley, T., Wolburg, H., Risau, W., Qin, Y. (1995) Distinct roles of the receptor tyrosine kinases Tie-1 and Tie-2 in blood vessel formation, *Nature* **376**, 70–74.

Sawano, A., Takahashi, T., Yamaguchi, S., Aonuma, M., Shibuya, M. (1996) Flt-1 but not KDR/Flk-1 tyrosine kinase is a receptor for placenta growth factor, which is related to vascular endothelial growth factor, *Cell Growth Differ.* **7**, 213–221.

Serini, G., Valdembri, D., Zanivan, S., Morterra, G., Burkhardt, C., Caccavari, F., Zammataro, L., Primo, L., Tamagnone, L., Logan, M., Tessier-Lavigne, M., Taniguchi, M., Puschel, A.W., Bussolino, F. (2003) Class 3 semaphorins control vascular morphogenesis by inhibiting integrin function, *Nature* **424**, 391–397.

Senger, D.R., Galli, S.J., Dvorak, A.M., Perruzzi, C.A., Harvey, V.S., Dvorak, H.F. (1983) Tumor cells secrete a vascular permeability factor that promotes accumulation of ascites fluid, *Science* **219**, 983–985.

Shalaby, F., Rossant, J., Yamaguchi, T.P., Gertsenstein, M., Wu, X.F., Breitman, M.L., Schuh, A.C. (1995) Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice, *Nature* **376**, 62–66.

Shalaby, F., Ho, J., Stanford, W.L., Fischer, K.D., Schuh, A.C., Schwartz, L., Bernstein, A., Rossant, J. (1997) A requirement for Flk1 in primitive and definitive hematopoiesis and vasculogenesis, *Cell* **89**, 981–990.

Shivdasani, R.A., Mayer, E.L., Orkin, S.H. (1995) Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL, *Nature* **373**, 432–434.

Shull, M.M., Ormsby, I., Kier, A.B., Pawlowski, S., Diebold, R.J., Yin, M., Allen, R., Sidman, C., Proetzel, G., Calvin, D. (1992) Targeted disruption of the mouse transforming growth factor-beta 1 gene results in multifocal inflammatory disease, *Nature* **359**, 693–699.

Shutter, J.R., Scully, S., Fan, W., Richards, W.G., Kitajewski, J., Deblandre, G.A., Kintner, C.R., Stark, K.L. (2000) Dll4, a novel Notch ligand expressed in arterial endothelium, *Genes Dev.* **14**, 1313–1318.

Soker, S., Takashima, S., Miao, H.Q., Neufeld, G., Klagsbrun, M. (1998) Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor, *Cell* **92**, 735–745.

Sorensen, L.K., Brooke, B.S., Li, D.Y., Urness, L.D. (2003) Loss of distinct arterial and venous boundaries in mice lacking endoglin, a vascular-specific TGFbeta coreceptor, *Dev. Biol.* **261**, 235–250.

Soriano, P. (1994) Abnormal kidney development and hematological disorders in PDGF beta-receptor mutant mice, *Genes Dev.* **8**, 1888–1896.

Stalmans, I., Ng, Y.S., Rohan, R., Fruttiger, M., Bouche, A., Yuce, A., Fujisawa, H., Hermans, B., Shani, M., Jansen, S., Hicklin, D., Anderson, D.J., Gardiner, T., Hammes, H.P., Moons, L., Dewerchin, M., Collen, D., Carmeliet, P., D'Amore, P.A. (2002) Arteriolar and venular patterning in retinas of mice selectively expressing VEGF isoforms, *J. Clin. Invest.* **109**, 327–336.

Stehbens, W.E. (1996) Structural and architectural changes during arterial development and the role of hemodynamics, *Acta Anat. (Basel)* **157**, 261–274.

Sugimoto, H., Hamano, Y., Charytan, D., Cosgrove, D., Kieran, M., Sudhakar, A., Kalluri, R. (2003) Neutralization of circulating vascular endothelial growth factor (VEGF) by anti-VEGF antibodies and soluble VEGF receptor 1 (sFlt-1) induces proteinuria, *J. Biol. Chem.* **278**, 12605–12608.

Suri, C., Jones, P.F., Patan, S., Bartunkova, S., Maisonpierre, P.C., Davis, S., Sato, T.N., Yancopoulos, G.D. (1996) Requisite role of angiopoietin-1, a ligand for the TIE2 receptor, during embryonic angiogenesis, *Cell* **87**, 1171–1180.

Suri, C., McClain, J., Thurston, G., McDonald, D.M., Zhou, H., Oldmixon, E.H., Sato, T.N., Yancopoulos, G.D. (1998) Increased vascularization in mice overexpressing angiopoietin-1, *Science* **282**, 468–471.

Takashima, S., Kitakaze, M., Asakura, M., Asanuma, H., Sanada, S., Tashiro, F., Niwa, H., Miyazaki, Ji.J., Hirota, S., Kitamura, Y., Kitsukawa, T., Fujisawa, H., Klagsbrun, M., Hori, M. (2002) Targeting of both mouse neuropilin-1 and neuropilin-2 genes severely impairs developmental yolk sac and embryonic angiogenesis, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3657–3662.

Teichert-Kuliszewska, K., Maisonpierre, P.C., Jones, N., Campbell, A.I., Master, Z., Bendeck, M.P., Alitalo, K., Dumont, D.J., Yancopoulos, G.D., Stewart, D.J. (2001) Biological action of angiopoietin-2 in a fibrin matrix model of angiogenesis is associated with activation of Tie2, *Cardiovasc. Res.* **49**, 659–670.

Thurston, G., Suri, C., Smith, K., McClain, J., Sato, T.N., Yancopoulos, G.D., McDonald, D.M. (1999) Leakage-resistant blood vessels in mice transgenically overexpressing angiopoietin-1, *Science* **286**, 2511–2514.

Thurston, G., Rudge, J.S., Ioffe, E., Zhou, H., Ross, L., Croll, S.D., Glazer, N., Holash, J., McDonald, D.M., Yancopoulos, G.D.

(2000) Angiopoietin-1 protects the adult vasculature against plasma leakage, *Nat. Med.* **6**, 460–463.

Torres-Vazquez, J., Gitler, A.D., Fraser, S.D., Berk, J.D., Pham, V.N., Fishman, M.C., Childs, S., Epstein, J.A., Weinstein, B.M. (2004) Semaphorin-plexin signaling guides patterning of the developing vasculature, *Dev. Cell* **7**, 117–123.

Urness, L.D., Sorensen, L.K., Li, D.Y. (2000) Arteriovenous malformations in mice lacking activin receptor-like kinase-1, *Nat. Genet.* **26**, 328–331.

Uyttendaele, H., Ho, J., Rossant, J., Kitajewski, J. (2001) Vascular patterning defects associated with expression of activated Notch4 in embryonic endothelium, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5643–5648.

Veikkola, T., Jussila, L., Makinen, T., Karpanen, T., Jeltsch, M., Petrova, T.V., Kubo, H., Thurston, G., McDonald, D.M., Achen, M.G., Stacker, S.A., Alitalo, K. (2001) Signalling via vascular endothelial growth factor receptor-3 is sufficient for lymphangiogenesis in transgenic mice, *EMBO J.* **20**, 1223–1231.

Villa, N., Walker, L., Lindsell, C.E., Gasson, J., Iruela-Arispe, M.L., Weinmaster, G. (2001) Vascular expression of Notch pathway receptors and ligands is restricted to arterial vessels, *Mech. Dev.* **108**, 161–164.

Wang, H.U., Chen, Z., Anderson, D. (1998) Molecular distinction and angiogenic interaction between embryonic arteries and veins revealed by ephrin-B2 and its receptor Eph-B4, *Cell* **93**, 741–753.

Wigle, J.T., Oliver, G. (1999) Prox1 function is required for the development of the murine lymphatic system, *Cell* **98**, 769–778.

Xue, Y., Gao, X., Lindsell, C.E., Norton, C.R., Chang, B., Hicks, C., Gendron-Maguire, M., Rand, E.B., Weinmaster, G., Gridley, T. (1999) Embryonic lethality and vascular defects in mice lacking the Notch ligand Jagged1, *Hum. Mol. Genet.* **8**, 723–730.

Yamaguchi, T.P., Dumont, D.J., Conlon, R.A., Breitman, M.L., Rossant, J. (1993) flk-1, an flt-related receptor tyrosine kinase is an early marker for endothelial cell precursors, *Development* **118**, 489–498.

Yamashita, J., Itch, H., Hirashima, M., Minetaro, O., Nishikawa, S., Yurugi, T., Naito, M., Kazuwa, N., Nishikawa, S.A. (2000) FIk1-positive cells derived from embryonic stem cells serve as vascular progenitors, *Nature* **408**, 92–96.

Yamashita, J., Itoh, H., Hirashima, M., Ogawa, M., Nishikawa, S., Yurugi, T., Naito, M., Nakao, K., Nishikawa, S. (2000) Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors, *Nature* **408**, 92–96.

Yuan, L., Moyon, D., Pardanaud, L., Breant, C., Karkkainen, M.J., Alitalo, K., Eichmann, A. (2002) Abnormal lymphatic vessel development in neuropilin 2 mutant mice, *Development* **129**, 4797–4806.

Ziegler, B.L., Valtieri, M., Porada, G.A., De Maria, R., Muller, R., Masella, B., Gabbianelli, M., Casella, I., Pelosi, E., Bock, T., Zanjani, E.D., Peschle, C. (1999) KDR receptor: a key marker defining hematopoietic stem cells, *Science* **285**, 1553–1558.

Zhang, Z., Zou, W., Wang, J., Gu, J., Dang, Y., Li, B., Zhao, L., Qian, C., Qian, Q., Liu, X. (2005) Suppression of tumor growth by oncolytic adenovirus-mediated delivery of an antiangiogenic gene, soluble Flt-1, *Mol. Ther.* **11**, 553–562.

Vector System: Plasmid DNA

Rajkumar Banerjee[1] *and Leaf Huang*[2]
[1] *Indian Institute of Chemical Technology, Hyderabad, India*
[2] *University of Pittsburgh, Pittsburgh, PA 15213*

1 Overview of Plasmid-mediated Gene Therapy 290

2 Direct Delivery with Naked Plasmid DNA 292
2.1 History and Recent Efforts Toward Attaining Long-term Expression 292
2.2 Recent Therapeutic Uses: Treating Cancer 293
2.3 Understanding Physiological Role of Protein-transduction Pathways 294
2.4 Gene Transfer to Skeletal Muscle 294
2.5 Gene Delivery for Myocardial Diseases 295
2.6 Gene Therapy for Angiogenesis 296
2.7 Gene Therapy for Autoimmune Diseases 297

3 Improving Plasmid DNA-mediated Gene Transfer by Electroporation 297

4 Improving Plasmid DNA Transfer Mediated by Gene Gun 298

5 Lipid-based Vectors 299
5.1 Idea Behind Lipid-based Vectors: Recent and Classical Uses 299
5.1.1 Structures of First Generation Cationic Lipids 300
5.2 Cellular Barriers for Transfection 302
5.2.1 Structure of Lipid/DNA Complex 302
5.2.2 Entry of DNA into Cells 302
5.2.3 Fate of Complex Inside the Cell: Escape into Cytosol 304
5.2.4 Entry of DNA in Nucleus 305

6 Second Generation of Lipidic Delivery System (Lipopolyplex) 305
6.1 Polycation-condensed DNA Entrapped in Cationic Liposome: LPD-I Formulation 305
6.2 Polycation-condensed DNA in Anionic Liposomes: LPD-II Formulation 307

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

7 Emulsion-mediated Gene Transfer 308
7.1 General Development 308
7.2 Reconstituted Chylomicron Remnants for Gene Transfer 308

8 Use of Cationic Liposome/DNA Complex (Lipoplex) 309
8.1 *In Vivo* Biodistribution Through Intravenous Injection: Therapeutic Implications 309
8.2 Directly Injected Lipoplexes Used Against Cancer and Other Diseases 310
8.3 Intraperitoneal Injection of Lipoplex for Treating Cancer 312

9 Gene Delivery by Polymeric Systems (Polyplex) 312
9.1 General Development and Uses 312
9.2 Targeted Gene Delivery by Antibodies Conjugated with Polycations 314

10 Nonviral Vector Related Cytotoxicity 314

11 Conclusion 315

Acknowledgment 316

Bibliography 316
Books and Reviews 316
Primary Literature 317

Keywords

Gene Therapy
A therapeutic technique to eradicate genetic malfunctions by introducing new gene or repairing defective gene in the host cell. The therapy, initially aimed toward treating inheritable disorders, is also used in treating acquired disorders such as AIDS, cancer and so on.

Lipoplex
A charged complex made of DNA and cationic lipid. The complex is formed by electrostatic combination of negatively charged DNA with positively charged cationic lipids. Preformed cationic liposomes made of cationic lipids and/or colipids interact and condense DNA by cooperative interaction of positively charged head group of aggregated surface of lipids with negatively charged DNA.

Lipopolyplex
An entity formed by mixing the preformed liposomes made of cationic, anionic or neutral lipids, and/or colipids with cationic polymer induced precondensed DNA. Electron microscopic studies revealed a lipid-based envelope, carrying the polyplex-containing core.

Nonviral Gene Delivery
A special gene delivery technique, which does not use viral mechanism to deliver genes to cells. It involves various physical techniques, chemical methods, and biomimetic synthetic agents to introduce genes to cells.

Polyplex
Another charged complex as lipoplex, but here the cationic entities are polymers. Cationic polymers associated with net positive charges also condense negatively charged DNA by electrostatic combination to form polyplex.

Abbreviations

CHEMS:	cholesteryl hemisuccinate
DC-Chol:	3$\tilde{\beta}$-[*N*-(*N*′,*N*′-dimethylaminoethane)-carbamoyl] cholesterol
DDAB:	dimethyldioctadecylammonium bromide
14Dea2:	*O*,*O*′-ditetradecanolyl-*N*-(trimethylammonio acetyl) diethanolamine chloride
DHDEAB:	(*N*,N-di-*n*-hexadecyl-*N*,*N*-dihydroxyethylammonium bromide)
DMDHP:	*N*,*N*-(2-hydroxyethyl)-*N*-methyl-2,3-bis(myristoyloxy)-1-propanaminium iodide
DMRIE:	*N*-(2-hydroxyethyl)-*N*,*N*-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium chloride
DOPE:	1,2-dioeoyl-sn-glycero-3-phosphatidylethanolamine
DORIE:	*N*-[1-(2,3-dioleyloxy)propyl]-*N*-hydroxyethyl-*N*,*N*-dimethylammonium chloride
DOSPA:	2,3-dioleyloxy-*N*-[2(sperminecar-boxamido)ethyl]-*N*,*N*-dimethyl-1-propanaminium trifluoroacetate
DOTAP:	*N*-[1-(2,3-dioleyloxy)propyl]-*N*′,*N*′,*N*′-trimethylammonium chloride
DOTIM:	1-[2-[9-(*Z*)-octadecenoyloxy]ethyl]]-2-[8](*Z*)-heptadecenyl]-3-[hydroxyethyl]imidazolinium chloride
DOTMA:	*N*-[1-(2,3-dioleyloxy)propyl]-*N*,*N*,*N*-trimethylammonium chloride
GAP-DLRIE:	*N*-(3-aminopropyl)-*N*,*N*-dimethyl-2,3-bis(dodecyloxy)-1-propaniminium bromide
LPLL:	lipopoly(L-lysine)
MP:	monomethylphosphonate
PAMAM:	polyamidoamine
PEG:	polyethyleneglycol
PEI:	polyethyleneimines
PLL:	poly-L-lysine
PO:	phosphodiester
PS:	phosphorothioate
TC-Chol:	3β-[*N*-(*N*′,*N*′,*N*′-trimethylaminoethane)carbamoyl] cholesterol

In gene therapy, nonviral gene delivery has earned a special position on its own right. This delivery system enjoys several favorable properties, such as, the delivery efficiency is not limited by the size of the genetic cargo, and it is less immunogenic and less toxic than their viral counterpart. One of the methods of delivery is direct injection of naked DNA to the organs of interest. Plasmid DNAs (pDNA), which carry recombinant genes of interest, are used for introducing genes to cells and organs. Various physical methods, for example, gene gun and electroporation, increase the efficiency of the naked DNA incorporation. Lipid-based vectors have the advantage of protecting the DNA against degradation by endogenous DNase *in vivo*, and can be targeted to a specific cellular site in some special cases. Varieties of lipid-based systems have been designed, mostly containing cationic lipids of different structures. Second-generation vectors containing polymers are more stable and more targetable than the first generation vectors. Although significant progress has been made, nonviral vectors are still limited in their efficiency and by some cytotoxicity inherited in the bacterial pDNA, which hosts the gene of interest. Understanding mechanisms and means to overcome the cellular barriers for the DNA delivery will undoubtedly promote further development of pDNA mediated gene delivery.

1 Overview of Plasmid-mediated Gene Therapy

Postgenomic era has thrust new responsibilities on humankind. Many defective genes responsible for various disease phenotypes have already been discovered. It now becomes a daunting challenge for multidisciplinary field experts to think of strategies to introduce normal genes efficiently into the target cells and/or tissues to complement the defective genes. Hence, gene therapy faces new tasks. This methodology originally designed to correct inheritable disorders, is now being considered for treating acquired disorders, such as cancer and AIDS.

Gene therapy is largely classified into two modes of gene delivery, it either utilizes viruses or it does not. Replication defective viruses such as retrovirus, adenovirus, adeno-associated virus, herpes simplex virus, papilloma virus, Sendai virus, and so on with part of the genes replaced by a therapeutic gene are used to transduce cells with very high efficiency. They are in use for various gene therapy approaches either clinically or experimentally. There are several drawbacks, which have partially eclipsed or limited the widespread use of viral vectors albeit their excellent transducing efficacy in cells and tissues. For example, induction of strong host-immune response followed by toxicity is common in adenoviral vector–based treatment, which limits the duration of the transgene expression. Additionally, the imposing immunogenicity makes repeated dosing impossible. Retroviral vector requires dividing cells for gene expression, but it poses the risk of random integration into the host genome, leading to mutation and neoplastic transformation. A large, industrial scale production of viral vectors is difficult. There is limitation for the size of the gene cassette to be included in viral vectors. These concerns have prompted

the development of nonviral vectors as a less toxic and more scalable alternative for gene therapy. However, nonviral vectors still need improvement, both in efficiency as well as in the duration of the transgene expression. These aspects are the central concerns of the current nonviral vector development.

Nonviral vectors often use recombinant plasmid DNA (pDNA) as the carrier of the therapeutic gene. Recombinant DNA is an artificial, synthetic DNA-based recombinant molecule, made by addition of several DNA fragments obtained from different sources, whereas pDNA is an extrachromosomal, circular, predominantly supercoiled bacterial DNA, which can be produced on a large scale by fermentation in bacterial cultures. Through simple recombination techniques, any therapeutic or marker gene of any size can be introduced into pDNA. Nowadays with the help of existing advanced technologies, a single lot can produce multigram quantities of pDNA, which can be finally purified using simple physical and chemical methods. Generally, standard pDNA is 4–10 kbp in size (molecular weight: 1×10^6 to 3×10^6). Other than the usual bacterial sequences, pDNA for gene therapy contains an expressible sequence of genomic DNA, popularly known as *complementary DNA* (cDNA) that codes for a specific protein. This genetic sequence, also called *transgene,* remains under the controlling influence of another gene sequence called *promoter*, located upstream of the coding sequence of transgene. This promoter sequence controls transcription, that is, transgene to messenger RNA conversion. A transcription termination gene is often inserted downstream of coding sequence to ensure proper termination of transgene transcription. A polyadenylation sequence and sometimes an intron, that is, a nonexpressible gene sequence, is also added to ensure proper processing of the m-RNA product of the transgene (see Fig. 1). The major advantage of plasmid expression vectors is that the cells upon transfection produce homogeneous, correctly folded, posttranslationally modified proteins within itself. This thwarts the attempt to purify and isolate proteins, which often leads to partial degradation or inactivation of the protein.

This chapter is, however, intended to summarily emphasize the role of pDNA as a gene-carrying cask, which finds its role in gene therapy using nonviral-based delivery techniques toward deciphering various biological functions, and use

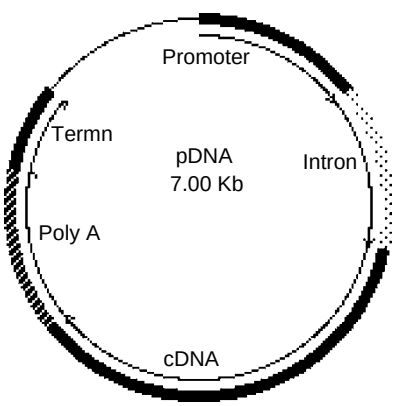

Fig. 1 Schematic representation of a typical plasmid DNA. Promoter: promoter sequence; Intron: intron sequence; cDNA: complementary DNA encoding a gene; Poly A: polyadenylation sequence; Termn: termination sequence.

those facts in finding gene therapy–based therapeutically useful tools against various disease pathologies.

2
Direct Delivery with Naked Plasmid DNA

2.1
History and Recent Efforts Toward Attaining Long-term Expression

The simplest of all DNA delivery systems is the injection of naked pDNA to the organ or tissue of interest. Skeletal muscle was first transfected by intramuscular (IM) injection of naked pDNA. Since then, a plethora of tissues such as, the interstitial space of liver, thyroid, heart muscles, brain, urological organs, and so on have been transfected with direct gene transfer. The delivery by direct injection of naked DNAs containing one or more anticancer genes was also tested against various tumors with mixed efficacies.

This technique of direct injection of pDNA was thoroughly used in several other experimentations as well. This technique of direct injection has become a useful tool to analyze the gene expression and promoter function in the respective organ. Utilizing viral promoter, such as that from cytomegalovirus (CMV), a high level of initial expression was attained, which, however, declined sharply to near background within two to three weeks. Although within two days of pDNA injection the maximum decrease in the expression level was observed, but the pDNA continued to be scantily detected till the end of 12 weeks (0.2 copies per genome). The major cause of this early decline was believed to be because of viral promoter activated immune responses, and this fact was later corroborated by experimentations in immunosuppressed mice that showed extended levels of transgene expressions.

Sustained therapeutic benefit needs assurance for long-lasting gene expression of pDNA associated therapeutic genes without eliciting any expression related cytotoxicity. Classically, several experimentations were performed, with fewer cases showing chances of extended expression time. For example, high-expressing human factor IX (hFIX) pDNA upon transfection in mouse liver yielded therapeutic-level gene expression over $1\frac{1}{2}$ years, eliciting no transgene expression–related cytotoxicity or long-term therapy related toxicity. However, with liver regeneration, there was a decline in transgene expression, which suggested the possibility of decline in maintenance of plasmid, rather than the possibility of transgene integration into the host genome. Hence, sustained gene expressions required a transcriptionally active vector DNA that could persist for a long time in the organ. Let us also not forget that a random genomic integration in host genome is contrary to health benefit. Therefore, with the help of this transcriptionally active vector, pDNA nonviral gene transfer could lead to extended life of transgene-expression level without leaving scope for genomic integration.

For extended transgene expressions of conventional plasmid vectors, a new technique was involved. The plasmid vector was modified to contain Eppstein–Barr virus (EBV) episomal vectors that consisted of *EBNA1* (EBV nuclear antigen-1) gene, and oriP element. Interestingly, nuclear localization and transcriptional upregulation of the oriP-bearing plasmid was assisted with the help of EBNA-1. This enabled efficacious transgene expression and long-term maintenance of the expression in cells transfected with the EBV/lipoplex. EBV plasmid vector played an important

role in various disorders such as congenital, malignant, chronic, and infectious diseases. This plasmid was used in gene therapy against tumor malignancies, such as in reducing tumor size by intratumoral injection of EBV plasmid vector encoding *HSV1-TK* anticancer gene. *EBV* plasmid-based interleukin genes, *IL-12* and *IL-18* gene transfer to tumors elicited natural killer (NK) cell activities and cytotoxic T-cell (CTL) responses, which lead to tumor retardation. Cardiomyopathic animals experienced extended cardiac output upon intracardiomuscular transfer of EBV-based β-adrenergic receptor gene. EBV-based plasmid gene transfer was also used for genetic vaccination against acute viral infections.

Another school of thought sprouted while developing and fine-tuning the conditions for long-term transgene expression. We know that virus-mediated gene transfer lead to random integration into the host chromosome. This often leads to abnormality in gene expression. However, site-specific genetic integration into inconspicuous genomic sequence might be one of the most-sought-after preconditions for stable, long-lasting, and powerful transgene expressions mediated through recombinant plasmids. This became possible with phage integrases. These enzymes mediate unidirectional site-specific recombination between two specific DNA sequences, one containing phage attachment site (attP) and the other bacterial attachment site (attB). These DNA sequences are short, yet specific on a genetic scale. A plasmid containing the bacterial attachment site flanking the gene of interest in both the ends was transfected into cells. The integrases of the serine family showed efficient recombination in mammalian cells. They mediated efficient integration at introduced att sites or native sequences with partial att sites (pseudo-attP). Human α_1-antitrypsin and hFIX gene associated with attB cassette, coadministered with integrase plasmid, upon liver delivery to mice showed elevated long-term expression (therapeutic level of protein obtained for about 8 months). Upon partial hepatectomy, the expression did not decline, indicating that the DNA was integrated to the mouse genome. Similar results were obtained during the functional restoration of human type VII collagen protein in skin cells obtained from recessive dystrophic epidermolysis bullosa (RDEB) patients. Additionally, this technique could be useful in developing big transgenic animals dedicated to the preparation of therapeutic proteins on large scales.

2.2 Recent Therapeutic Uses: Treating Cancer

Nonviral gene therapy had become a useful tool in generating various therapeutic effects in diverse animal models. One of these effects is to induce anticancer response in tumor models by introducing anticancer genes. NO synthase II (*NOS II*) gene that catalyzes nitric oxide (NO) production was incorporated in plasmid under CMV promoter control. This plasmid, upon injecting in medullary thyroid cancer (MTC), elicited anticancer effect in tumorigenic cells. The gene triggered a suicidal effect (apoptosis) on the tumor cells by the production of NO, which specifically activated the macrophages against the fast dividing tumor cells, leaving normal cells untouched. NO was also observed to mediate bystander antitumor effect: The activated *NOS* gene in a tumor cell induced the cytotoxicity in the neighboring tumor cells. The fact that NO mediated tumorogenic effect does not require transfection of all neoplastic cells promised a

capable suicide gene therapy approach to human cancer.

Fas ligand is another apoptotic inducing molecule that showed similar bystander effect. The membrane protein receptor Fas and its ligand FasL mutual interaction initiate an apoptotic signal in Fas-bearing cells. pDNA could be used to deliver these apoptosis-inducing genes to initiate killing of transfected and nontransfected surrounding cells. On direct injection of FasL encoding pDNA vector into the inflamed thyroid, pathogenic lymphocytes were inhibited to enter into thyroid, leaving the already infiltrated T cells dead. Thus, FasL expressing in thyroicytes might lead to potential remedial therapy for the experimental autoimmune thyroiditis (EAT).

2.3 Understanding Physiological Role of Protein-transduction Pathways

Direct pDNA transfer technique could find its use in examining the role of a physiologically related protein-transduction signal pathway toward certain endogenous disease phenotype, such as investigating the functional role of the tissue kallikrein-kinin system (KKS) in the central control of blood pressure homeostasis. Kallikrein is a proteinase enzyme that converts kininogen to vasodialative kinin peptides. To understand the role of vasodialative KKS on the pathogenesis of hypertension, the human tissue kallikrein gene *CMV-cHK*, was intracerebroventricularly injected into hypertensive rats. Cortex, cerebellum, brain stem, hippocampus, and hypothalamus identifiably expressed human tissue kallikrein protein. The expression level and its effect could lead to verify the role of the KKS system.

Similarly, the method of Direct injection of pDNA to rat fibroblast had been used to study the consequences of overexpressing dominant negative forms of the ubiquitous mitogen-activated protein kinase (MAP-kinase) toward inhibition of cell proliferation. Direct pDNA microinjection technique was also used to demonstrate the absolute requirement of the protein Cyclin A in DNA synthesis, as opposed to the popular belief that the Cyclin A was uniquely involved in the mitotic activation only. An SV40 expression vector encoding the full-length, antisense Cyclin A RNA, upon direct microinjection into synchronized mammalian cells, completely suppressed Cyclin A expression at S phase, resulting in inhibition of DNA synthesis.

2.4 Gene Transfer to Skeletal Muscle

A variety of diseases such as muscular dystrophies, chronic ischemic limb syndromes and so on, were treated by the technique of direct gene transfer to skeletal muscle. Specifically engineered pDNAs and the vector systems were developed that enabled regulated and tissue-specific transgene expression in skeletal muscle *in vivo*. Naked pDNA–based gene transfer had been used in different animal models, for example, in correcting Duchenne Muscular Dystrophy, to supply sources of therapeutic protein systemically, for genetic vaccination against pathogens and tumor cells. However, larger animals showed limited clinical utility owing to inefficient gene expression in them.

On quantifying the gene-uptake in the muscles, the IM injection showed less than 1% uptake of injected dose and was limited to cells adjacent to the needle track. Using hypertonic sucrose, or muscle revitalizers, the efficiency and reproducibility

of gene expression could be increased. Intraarterial injection of naked pDNA into the femoral arteries of rats showed 2 orders of magnitude higher transgene-expression level throughout the muscles of hind limb compared to that obtained in direct muscular injection. Myofibers were 10% more transfected through intravascular delivery than with the direct IM injection. This intraarterial injection technique was extended to the nonhuman primates also. Hypothetically, the pDNA upon intravascular injection extravasates out by the intravascular pressure following the convective flow across the endothelium, and was soon picked up by the *in situ* muscle cells with the help of membrane receptors. This was evidenced by the fact that the naked pDNA was taken up by hepatocytes *in vivo* by a receptor-mediated process. Several DNA receptors had been discovered in human leukocytes, peritoneal macrophages, and in wide variety of tissues and tumor cells. For the enhancement of IM gene expression, a fresh look is needed toward the poorly understood mechanism of molecular recognition and characterization of the cell surface receptor(s) involved in the binding and internalization of DNA. It is reported that rapid injection of pDNA in a large volume (e.g. 5 μg of DNA/20 g mouse in 1.5 to 2.0 mL of saline solution) through the tail vein left the injected DNA in the inferior vena cava. The DNA flowed back to the tissues linked to this vascular system, primarily the liver. The hydrodynamic pressure forced DNA into the liver cells before it was mixed with blood. By this process, the liver showed the highest expression of gene; internal organs such as lung, spleen, heart, and kidney were also efficiently transfected. It was also shown that briefly clamping the vena cava following tail vein injection of pDNA in a small volume efficiently transfected both liver and diaphragm. The result is potentially important because diaphragm is barely transfected by hydrodynamics-based method. The full-length dystrophin cDNA can be delivered to the diaphragm for the treatment of Duchenne Muscular Dystrophy. It is well known that patients with Duchenne Muscular Dystrophy often suffer from fatal respiratory failure due to the dystrophic diaphragm muscle collapse.

2.5 Gene Delivery for Myocardial Diseases

Myocardial gene delivery via direct injection or through coronary vasculature with naked pDNA usually showed low levels of gene transfection. However, the results so obtained had invaluable impact in characterizing the role of promoters in cardiac tissue and further helped in examining the effect on transferred foreign gene expression by the influences of naturally occurring mechanical and hormonal stimuli of the myocardium.

It was demonstrated that the tissue-specific promoter chimeras injected into the heart could respond accurately to shift in thyroid hormone levels *in vivo*. Injection of pDNA with gene constructs driven by cellular promoters resulted in detectable levels of reporter gene activities. The cellular promoter was derived from the rat α-myosin heavy chain (α-MHC) gene. The expression of this gene *in vivo* is restricted to cardiac muscle and is positively regulated by thyroid hormone. The regulatory portion of genes expressing specifically in cardiac muscles could be identified using this method. Direct DNA injection had been extended to evaluate and characterize the activation properties of a cardiac-specific promoter/enhancer of the slow/cardiac troponin C (cTnC) gene that express in cardiac striated muscles.

Myocardial direct DNA injection was also utilized to analyze the transcriptional regulation of brain creatine kinase (BCK) gene in the developing heart. pDNA constructs containing BCK promoter and *CAT* or luciferase reporter gene was delivered into the left lateral wall and apex of the ventricle on the heart. The study might provide insight into the embryonic gene expressing mechanism during cardiogenesis. Because the *BCK* gene, the major gene for cytoplasmic creatine kinase expressed in the embryonic heart, is downregulated during cardiogenesis, it is reinstated in response to stimuli such as ischemia, hypertrophy, or heart failure in the adult.

The method of direct pDNA injection was used to explore the effect of specific pathophysiological state on cardiac gene expression, such as ischemia, myocardial infarction, reperfusion injury, hypertension and so on. Ischemia is a disease state that is formed when tissues are starved of blood supply and nutrients because of deficient supply of blood through possibly narrowed or blocked arteries. Sporadic myocardial ischemia is prevalently associated with coronary arterial diseases. Ischemia related disease phenotype needs a therapeutic gene, which could be selectively upregulated by the elevated signals for ischemic activity and consequently downregulated when the activity represses. In this context, Prentice et al. introduced expression plasmids containing muscle-specific α-MHC promoters and hypoxia-responsive enhancer (HRE) elements linked to a reporter gene in cultured cells or into the rabbit myocardium, and measured the regulation of these constructs by hypoxia or experimental ischemia. It was shown that the expression of reporter gene was induced by both hypoxia *in vitro* and by a short interval of ischemia *in vivo*.

There were different reports concerning the stability of plasmid-based transgenes in both skeletal and cardiac muscles. It was shown that the rat cardiac myocytes could express *β-galactosidase* gene under the control of the Rous sarcoma virus promoter, by the injection of pDNA encoding the reporter gene directly into the left ventricular wall. β-galactosidase expressed in cardiac myocytes was detected in rat hearts for at least four weeks after injection of the *β-galactosidase* gene. In postmitotic cardiac and skeletal muscle cells, the transgene expression of the pDNA declines with time, probably due to the episomal localization of the DNA. The reason that the striated muscles showed higher capacity for uptake and expressing of pDNA following direct injection was not clear; the efficient gene transfer might be induced by cellular membrane rupturing and destabilization, followed by inflammation caused by the injection needle.

2.6 Gene Therapy for Angiogenesis

Because ischemia is formed as a result of deficient supply of blood to tissues, formation of new blood vessels is a remedy against this disease. One of the mechanisms that involves the formation of new blood vessels is angiogenesis. Angiogenesis is the process of new blood vessel development for the vascularization of various organs, for wound healing, and to allow cancer development and proliferation.

Therapeutic angiogenesis involves replenishing angiogenic growth factors by administering recombinant proteins or endothelial growth-factor gene. The recombinant proteins have severe limitation on its usage, as they are expensive and difficult to produce on a large scale.

On the other hand, gene therapy provides a systemic and long-term effect with modification in the effective dosage of the therapeutic agent. To evade potential problems of pathological angiogenesis, transient gene expression is usually preferred for this kind of treatment. Tsurumi et al. introduced naked pDNA encoding vascular endothelial growth factor (VEGF) by (IM injection into ischemic hind limb muscles of a rabbit model and observed that the vessels and blood-capillaries were increased in rabbit muscles injected with VEGF compared to controls. An enhanced vascularity-induced perfusion followed by increased blood flow in the ischemic limbs was also observed. In clinical trial, Simovic et al. introduced naked pDNA encoding human *VEGF* gene by direct IM injection to chronic ischemic limbs of patients, to treat peripheral neuropathy caused by critical limb ischemia. The patients showed decreased neuropathic disability in the treated limbs, indicating that long-term therapy might improve the integrity in tissues of ischemic limb and consequent retrieval of limb.

2.7 Gene Therapy for Autoimmune Diseases

Autoimmune disease is a pathogenic condition in which ones immune system erroneously targets and aggresses a person's own cells, tissues, and organs. The most prevalent symptom of this disease is inflammation, which is caused by the abundant presence of a large number of immune cells and molecules in the target site of the body. Autoimmune diseases or other inflammatory conditions were primarily and successfully confronted by gene therapy through the delivery of cytokines or cytokine inhibitors. Interferon γ. (IFN-γ), interleukin-1 (IL-1, α or β), IL-12, and tumor necrosis factor α (TNFα) are recurrently addressed inflammatory cytokines in illnesses related to autoimmune/inflammatory diseases. Other than these cytokines, transforming growth factor β (TGF-β) is also a key regulatory cytokine, because TGF-β inhibits T and B cell responses, irregularity of which leads to elevation of autoimmunity invoked disease conditions.

In animal models, pDNA constructs with the encoding anti-inflammatory cytokine genes for IL-10, IL-4, and TGF-β1 were injected into either tibialis anterior or rectus femoris muscles in nonobese diabetic (NOD) mouse against autoimmune diabetic disease. Although there was no marked decrease in severity of insulitis, the diabetes was reduced in NOD mice injected with IL10 as compared to untreated NOD mice. In another experiment, treatment of autoimmunity prone NOD mice with pCMV-TGF-β1 resulted in considerable elevation of TGF-β1 level in the plasma. The increased levels of TGF-β1 exerted various immunosuppressive effects such as suppression of delayed-type hypersensitivity (DTH), and prevention of insulitic and diabetic incidence in this kind of mice. TGF-β1, IL-4, and IFN-γ gene coding plasmid vectors were also injected intramuscularly into rodent models for treating experimental allergic encephalomyelitis (EAE), systemic lupus erythematosus (SLE), colitis, and Streptococcal cell wall-induced arthritis (SCW arthritis).

3 Improving Plasmid DNA-mediated Gene Transfer by Electroporation

Electroporation is a process of exposing cells to controlled electric field for the

purpose of cellular membrane permeability. The electric pulses, intensely localized to destabilize the membrane, allow molecules to enter cells. A variety of genetic materials were inserted into the cells *in vitro* by electroporation. Electrochemotherapy facilitates the cellular entry of hydrophilic anticancer agents such as bleomycin to obtain drastically improved antitumor effect in malignant melanoma.

Recently, *in vivo* electroporation has emerged as a leading technology for developing nonviral gene therapies and nucleic acid vaccines (NAV). Naked pDNA injections accompanied by electroporation showed 10 to 1000-fold increases in gene expressions as compared to the nonelectroporated direct injections. However, the duration of a gene expression following *in vivo* electroporation was dependent on target tissue and plasmid constructs. Despite this fact, a broader variety of cells showed reporter gene expressions.

The major advantage of electroporation is that a treatment-area specific *in vivo* gene expression is possible wherein specific tissue could be directed by using specifically designed electrodes. A high level of controlled expression of specific gene was obtained by using tissue-specific promoter directed against specific tissue with a different pulsing parameter.

Various disease pathologies underwent electroporation-mediated *in vivo* gene transfer. The successful utilization of this technique was done in targeting hepatic parenchyma, hepatocellular carcinoma (HCC), skin, skeletal muscle, mouse testes, melanoma, human primary myoblast, glomeruli, brain, human primary hematopoietic stem cells, human esophageal tumor, and rat skeletal muscle for correcting anemia of renal failure. Furthermore, electroporation of TGF-β1 gene to the wound models in diabetic mice resulted in higher cell proliferation, denser yet organized extracellular matrix formation, collagen synthesis, and angiogenesis in the treated wound site.

Electroporation utilizing high electric field strength induces tissue damage. This restricts its usage. Efficient and safe electrotransfer for delivering exogenous material into tissues must be developed before the clinical potential of gene therapy can be realized. Recently, a novel syringe electrode was developed. With this electrode, electric field was directly delivered to the cells in which pDNA had been injected, the injected DNA could be confined to the high field region without affecting the adjacent areas. High transfection efficiency was observed with comparatively lower field strength that caused minimal tissue damage.

A technique called *microelectroporation* was utilized locally to introduce a transgene into chick embryo and for expressing the gene in a spatial and temporal manner. By the microelectroporation technique, DNA molecules were efficiently introduced into the optic vesicle, sensory placodes, surface ectoderm, neuroepithelium of the central nervous system (CNS), and into the somites and limb mesenchyme.

4 Improving Plasmid DNA Transfer Mediated by Gene Gun

Gene gun is a physical way of administering gene *in vitro* or *in vivo*. pDNA makes electrostatic complex with gold or tungsten microparticles. The DNA-metal particle upon targeting to specific tissue is forcibly expelled by a flow of regulated, highly compressed helium gas. The delivery is highly localized to the tissue part. The delivery

to different tissue depths and areas is regulated by adjusting the speed and hence impact-pressure of the projectile. The process involves easy and speedy preparation of the delivery vehicle while keeping DNA intact. This technique sometimes allows DNA to gain direct access to nucleus bypassing the endosome/lysosome. These cellular compartments may induce possible enzymatic degradation. Because of the benefit of accessibility, this technique is especially suitable for gene transfer to skin and for superficial wounds. There are several examples in various animal models, which have shown a high level of transgene expression in the epidermis and dermis of the skin.

By the gene gun technique, mouse skin was transfected with IL-6 and hemagglutinin encoded DNA to elicit protective immune responses against equine influenza virus. Several different large animals such as rhesus monkeys, pigs, and horses were also immunized against virus by transfecting gene encoding the viral antigen.

Detectable transgene activities were also noticed in various nonsuperficial organs, such as lungs, pancreas, kidney, muscle, and cornea. Gene gun was also used to treat against tumor growth. Intradermal tumor upon *IL-12* gene delivery by gene gun gave detectable levels of the gene product at the treatment site. This eventually led to complete tumor regression within seven days. TGF-β1 encoded plasmid bombardment to rat tissue enhanced the tensile strength of the tissue by almost twofold, as compared to the control constituting nongene gun mediated direct injection. Porcine partial-thickness wounds, when transfected with a vector expressing epidermal growth factor (EGF) showed an increased rate of epithelial cell formation. This increased epithelialization shortened the wound healing time by 20%.

5 Lipid-based Vectors

5.1 Idea Behind Lipid-based Vectors: Recent and Classical Uses

The lipid-based DNA delivery system was first utilized on the premise that naked DNA upon injection *in vivo* might be degraded by endogenous DNase. The DNA would remain protected inside the lipid-based delivery vehicle during its run through the circulating blood stream. With necessary chemical modification, this vehicle would target the DNA toward specific cellular sites.

These lipids are a class of molecule that self-assemble in aqueous or organic media. Depending on its molecular architecture and in the presence or absence of cosolute, they assume structures such as micelle, emulsion, or liposome in aqueous media, and reverse micelle in organic media. However, among all the systems, liposomal systems found most biological usage, probably owing to its close resemblance with the cellular membrane. They constitute a bilayered structure, which can encapsulate water-soluble molecules in its aqueous hydrophilic core or water-insoluble molecules in its hydrophobic bilayer. Liposome-based drug delivery systems have shown promise in clinical use and several products have already been approved by the FDA. Fraley et al. had shown for the first time that encapsulating DNA in liposomal aqueous core could also deliver DNA to cellular targets. They used a composite liposomal system containing anionic and nonionic lipids. The DNA encapsulation efficiency for this system was low. Toward enhancing gene delivery and targeting ability of the liposome to specific tissue site, anionic

liposomes encapsulating DNA were coated with target-specific antibody to make pH-sensitive immunoliposomes. These liposomes upon intraperitoneal (IP) injection, showed specific transfection to tumor cells in an ascites tumor model in nude mice. Very recently, the immunoliposomes were revisited for targeting brain. An exogenous gene was targeted to brain through the blood brain barrier by intravenous (i.v.) injection of pegylated immunoliposomes conjugated with antibody to rat transferrin receptor. The whole entity was targeted to brain via the transcytosis of transferrin receptor.

Felgner et al. for the first time used cationic liposomes for gene transfection. The primary idea was to electrostatically condense DNA with more than one equivalent of cationic lipid and the negatively charged cells would take up the net positively charged entity when fed to the cell. Overall gene delivery to the cells was very high compared to any contemporary nonviral gene delivery techniques. This opened up a new vista in the field of gene therapy. Soon a plethora of literature developed, which chronicled the painstaking fine-tuning of the chemical structure of lipids by systematic structure-activity studies. These studies were done to dig out the structure of the potentially most potent cationic lipid for gene transfection.

pDNA played a significant role in this context. pDNA reconstructed with recombinant marker/reporter gene and suitable promoter is primarily used as a tool to screen the efficient transfection lipid from a library of cationic lipidic molecules. The electrostatic lipid–DNA complex that is formed by the recombination of these first generation cationic lipids with DNA is termed *lipoplex*. Cationic polymers also condense DNA and, with or without the presence of lipid, these polymers can deliver DNA to cells. This makes the basis of the second-generation gene delivery systems. Cationic polymers, cationic dendrimers, and so on condense DNA to make a complex called "polyplex". If liposomes, cationic or anionic are used to encapsulate the precondensed cationic polymer–DNA complex, and are used to deliver DNA to cells, then the whole entity is termed *lipopolyplex*. We will talk about these second-generation gene delivery systems later in this chapter.

5.1.1 Structures of First Generation Cationic Lipids

In the process of finding out the best cationic transfection lipid, several groups throughout the world, over the last 15 to 17 years, explored different chemistries to evolve different chemical structures, which made quite a big library of molecules. It is very interesting to know that several architectural differences in the molecular infrastructure of cationic lipids evolved, which lead to the formation of cationic lipids of varied transfection ability. Here, we will briefly yet broadly classify the categories of cationic lipids with varied degrees of chemical moieties (Fig. 2).

1. *Lipids with glycerol backbone containing aliphatic carbon chain(s)*: DOTMA, DMRIE, DORIE, DOSPA, GAP-DLRIE, DOTAP, DMDHP and so on. These lipids are linked to glycerol moiety through ether or ester bond and have either a single cationic charge with none, single, or multiple hydroxyl groups, or have multiple cationic charges by possessing spermidine derivatives.
2. *Lipids with long-chain alkylamine/alkylamide moiety*: DDAB, DOGS, LPLL, GS2888, 14DEA2, diC14amidine,

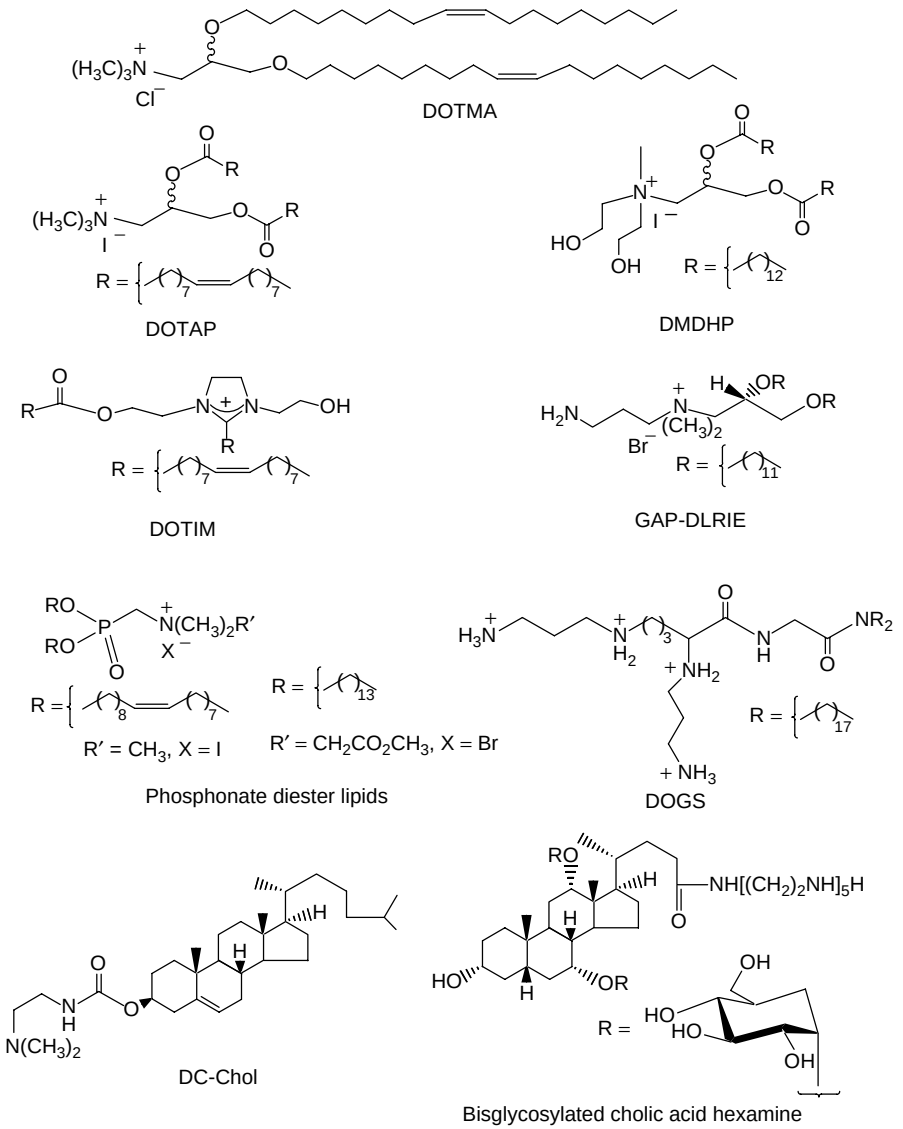

Fig. 2 Structures of some representative cationic lipids.

DHDEAB, TFX™ and so on. These are basically nonglycerol-based lipids with single or multiple charges. The charges arise because of the presence of ammonium or multiple amines or spermidines. Some of these lipids also possess multiple hydroxyl groups.

3. *Lipids with cholesterol moiety*: DC-Chol and other cholesterol derivatives, biglycosylated cholic acid derivative and cholesterol with polyamines, spermidines and so on.
4. *Lipids with L-α-phosphatidylethanolamine derivatives with spermidines*:

1,2-dipalmitoylphosphatidylethanolamidospermine (DPPES) or L-α-phosphatidylcholine with phosphonate diester.
5. *Lipids with imidazole derivative*: DOTIM or nonglycerol-based lipids with long-chain alkyl acyl carnitine ester.

There were many other lipids that were synthesized and tested for gene transfection. Only a few of these lipids had shown consistent and relatively high level of transfection in various cell lines and in animal models *in vivo*. Most of these first generation formulations were not long circulating in the blood, owing to the fact that the DNA/cationic lipid complex disintegrates in the presence of negatively charged blood proteins. Additionally, they showed no specific targeting of cells. Finally, in spite of such a big library of molecules, no consensus thumb-rule could be predicted or ascertained regarding the magic structure of lipid, which might show best transfection in all kinds of cells consistently. This is because there were so many parameters involved in the whole process of transfection. In order to develop the potentially best nonviral gene delivery system, we have to understand various barriers confronting transfection. Here, we will briefly discuss these barriers.

5.2 Cellular Barriers for Transfection

To develop an efficient gene delivery system, it seems necessary to understand the extra- and intracellular processes involved in the overall transfection mechanism. This will lead to comprehending the challenge before the task of developing efficient lipid-based nonviral vectors for gene therapy applications. For this purpose, cationic liposomes and pDNA are used widely to understand the cellular mechanism involved in the transfection (see Fig. 3).

5.2.1 Structure of Lipid/DNA Complex

Positively charged cationic liposomes electrostatically interact with negatively charged nucleic acid sequences to form fused, aggregated complexes, which are capable of entering a cell. These aggregates have a very heterogeneous size distribution with respect to particle size and net charge. The lipid-to-DNA mass and charge ratio, which defines the size and net surface charge of the complex, is an important factor to determine transfection efficiency. Hence, the structural features associated with this kind of complex need complete interpretation. Structural features revealed by electron microscopy include lipid-coated DNA strands, aggregates of liposome intercalating DNA, DNA entrapped between the lamellae of aggregated multilamellar structure, and tubular structure consisting of fused liposome around DNA. Polycations such as polylysine or multivalent cationic lipid condense DNA into small compact particles. Cells can take these compacted DNA much more efficiently than the lipid/DNA condensed particles obtained by the interaction of DNA by monovalent cationic lipids.

5.2.2 Entry of DNA into Cells

The presence of glycophospholipids and negatively charged proteins such as proteoglycans, glycosaminoglycans and so on imparts negative charges on the cell surfaces. It was hypothesized that membrane fusion between cationic liposome and negatively charged cell membrane is the primary means of cell entry. The hypothesis was based on the supposition that cationic and anionic liposomes readily

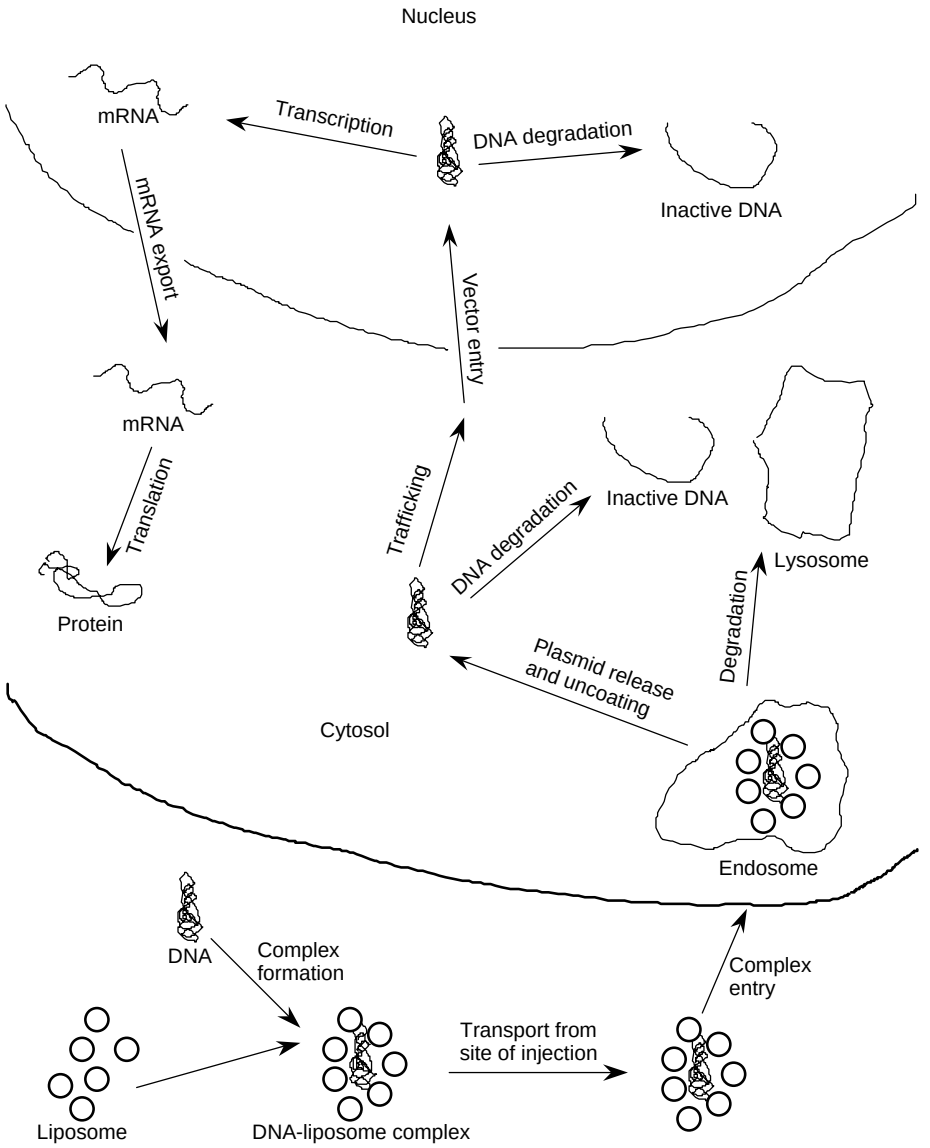

Fig. 3 Schematic illustration of the processes involved in gene delivery and expression.

fuse. It was conceptualized that the interaction of the cationic complex with the cell is primarily electrostatic, and does not involve any specific receptor for the cationic moiety. Other experiments also suggested that the mechanism of DNA transfer to animal cells from cationic liposomes might not entail a simple fusion of liposomal and cellular membrane. However, it was shown that heparin sulfates, the highly anionic polysaccharides, serve as an important cell surface receptor for cationic

lipid–DNA complexes. The nondifferentiated edged airway epithelial cells endocytose liposome–DNA complexes consequent to their relatively high negative charge and phagocytic activity compared to relatively more differentiated cells. Evidences now strongly suggest that slow endocytosis of intact lipid/DNA complex could be the primary method for cellular entry.

5.2.3 Fate of Complex Inside the Cell: Escape into Cytosol

After cellular entry, the lipoplex is engulfed into the lower pH compartment called *early endosomes* in the perimembranous region. The intracellular trafficking was ascertained by an experiment wherein a liposome composed of lipopolylysine (LPLL) and DOPE (1,2-dioleoyl-sn-glycero-3-phosphatidylethylamine) was first used to condense DNA to form electron-dense particles recognized by thin-section transmission EM and upon treatment to the cells a majority of them were found to reside in vesicular compartments. The endosomal contents usually pass into lysosome, the cell organelle in the perinuclear region that houses a host of degredative enzymes. The hydrolytic enzymes degrade most if not all of the lysosomal contents. So, it becomes necessary that the endosomal content must free itself at the early endosome stage to keep the DNA intact. Disruption of endosomal vesicle is visualized in the liposomal system containing DOPE as a helper colipid. DOPE, with its tendency to promote significant polymorphic changes in the lipid phase, stimulates membrane fusion or destabilization. This is followed by leakage of endosomal content into the cytoplasm. DOPE, in aqueous media, assumes inverted hexagonal phase II structure, which is frequently obtained in regions of membrane where it fuses with another membrane. Thus, one may assume that DOPE, or liposome formulation containing DOPE, might fuse with endosomal membrane and destabilize it to leak out the content from the endosome.

If the early endosomal release is not possible, then another way to keep DNA intact in lysosome is to protect the DNA from lysosomal degradation. Cationic liposomes formulated with cholesterol are believed to offer a useful role in keeping DNA intact. It was demonstrated that the lysosomal enzymes work at lower pH, that is, pH < 6. It was also shown that cholesterol-containing liposomes, which possess greater stability and lower ion-permeability compared to DOPE-containing liposomes, provide an improved stability to the lipid–DNA complex in the cytosol. It is easily conceivable that if the endosomal content passes onto lysosome before being released from endosomes, the lipoplex could remain secured in the lysosome.

Cationic lipids can destabilize a cellular membrane because of its intrinsic detergent property. Therefore, destabilization of endosomal and/or lysosomal membrane may be a contribution from the cationic lipids itself. In the same context, it was shown that the cationic lipid/DOPE or cationic lipid/cholesterol liposome formulation, exhibit surface anisotropies in terms of increased liposomal surface pH. The surface pH of the liposomal formulations exhibits at least two pH units higher than the pH of the solution at which they are made. Therefore, a liposomal solution made at physiological pH may in reality exhibit a surface pH > 9, which is detrimental for both the stability of endosome and activity of lysosomal enzymes. Endosomal disruptions were also done with fusogenic peptides, which promote pH-dependent fusion of small liposomes when

associated with lipid bilayer. When these peptides were codelivered with lipid/DNA complex, they imparted formidable endosomal disruption by changing its usual random coil conformation into amphipathic α-helix conformation at lower pH, resulting in consequent cytoplasmic delivery of DNA.

5.2.4 Entry of DNA in Nucleus

The success of lipid-mediated gene transfer is severely limited by the inefficient transport of transfected DNA from the cytoplasm into the nucleus. In general, macromolecules enter the nucleus through nuclear pores. Molecular aggregates with size more than 55 Å in diameter or molecular weight greater than 40 KD utilize the nuclear pore complex (NPC) to access the nucleus. NPC bind with molecular aggregates associated with nuclear localization signal (NLS) peptide and help it to translocate across the nuclear membrane. One of the typical signals is the NLS from SV40 large T antigen (PKRRRKV), which had been conjugated to many different molecules to gain nuclear access. NLS peptide-conjugated pDNA was delivered efficiently inside the nucleus with an enhanced gene expression. Similarly, SV40 T antigen NLS codelivered with DOTAP cationic liposome mediated efficient gene transfer and expression in the cell. The genes delivered inside the nucleus require uncoating from the lipidic shell before the transcription starts. It is generally assumed that pDNA is displaced from the complex by anionic macromolecules in the nucleus. In this regard, an alternative hypothesis for cytosolic release of DNA from lipid/DNA complex had been proposed. Demonstration through model studies showed that the pDNA was released from cationic liposome/DNA complexes by anionic liposomes exhibiting compositions mimicking the cytoplasmic face of the lipid monolayer of the cellular membrane. Membrane destabilization followed by flip-flop of the liposome/DNA complex by the anionic lipids by electrostatic interaction resulted in charge neutralized lipidic ion pairs followed by release of pDNA.

To evade the nuclear transport, an alternative approach was developed, which utilizes T7-based cytoplasmic expression vectors. Here, the inefficiently nuclear-transported DNA could be expressed in the cytoplasm itself.

6 Second Generation of Lipidic Delivery System (Lipopolyplex)

6.1 Polycation-condensed DNA Entrapped in Cationic Liposome: LPD-I Formulation

One of the prerequisites for a bioentity entering the cell and the nucleus is to possess a small size. The delivery of DNA was efficient through cationic lipidic formulation mediated gene delivery, but the nuclear transport from cytoplasm was quite an inefficient process. Generally, lipoplexes formed by multiple charged cationic lipids made small (~20 nm), highly condensed DNA/lipidic complex. These complexes were found to be more transfecting than the lipoplex formed by monocationic lipids and DNA. Synthetic cationic polymers like poly-L-lysine (PLL) could also condense DNA into very small compacted particles. This compact complex was allowed to freely mix with cationic liposomal solution. The overall complex so formed was termed lipid-protamine-DNA (LPD-I). A sucrose gradient ultracentrifugation of the heterogeneous mixture showed the existence of various populations associated

with varied amounts of lipid. Negative-stain EM studies of purified complexes showed electron-dense structures ranging from elongated rod-shaped to ball-shaped particles. The purified fractions were several fold more transfection efficient than unpurified ones. The fractions that contained more lipids were more efficient in gene transfection than those containing fewer lipids. The transfection efficiency of LPD-I was many folds higher than the corresponding cationic liposome–DNA complex. It is hypothesized that (1) the small compacted structure of DNA resulted in high cellular and nuclear uptake, (2) the DNA is protected efficiently against enzymatic degradation, and (3) PLL may have mimicked the NLS for nuclear delivery of DNA. An alternative cationic polymer, protamine sulfate, from salmon sperm was used as a substitution for PLL. Protamine sulfate is a small (MW ~4000 vs ~18 500 for PLL), highly positively charged peptide and very basic due to the presence of 21 arginine residues, which contains an NLS. It is naturally occurring (whereas PLL is synthetic), a USP grade compound, FDA approved and finds use as an antidote to heparin-induced anticoagulation. With a routine history of human administration, the issues of toxicity and immunogenicity were minimal for protamine sulfate.

Both protamine sulfate and PLL were compared for their transfection efficiencies. With the same amount of DNA and cationic lipid, PLL reached its efficiency plateau at an amount that was half the amount of protamine sulfate, but the overall efficiency of protamine sulfate was 2 to 7-fold higher than PLL in different cell lines. The extent of transfection efficiency for LPD-I was 7 to 45-fold higher than the levels shown by DNA/cationic lipid complex. Phosphate and free-base forms of protamine showed lower levels of transfection and the activity was comparable to lipoplex. It is noteworthy that arginine, which exists as a salt of sulfate, and the lysine, which exists as a free base or as a phosphated salt, showed a variance in the relative transfection efficiencies even though the percentage of basic (arginine + lysine) residues remained relatively constant.

Upon i.v. injection, LPD-I made with DOTAP had shown very high gene expression in heart, lung, liver, spleen, and kidney, with the highest expression found in the lung. The *in vivo* gene expression had steadily increased with increase in cationic lipid and/or protamine, but the protamine–DNA charge ratio was kept below 2 : 1 to avoid the formation of large aggregates. After LPD injection, transgene expression was detected as early as 1 h, peaked at 6 h, and declined thereafter. Lung had consistently shown the highest gene expression, which lasted for 2 days. Even after 4 days of LPD injection, gene expression could still be detected in lung, spleen, and liver.

In a biodistribution study with labeled DNA and liposome, it was revealed that DNA was rapidly removed from circulation by the liver after injection of protamine–DNA complexes, and uptake by the lung was always less than 10% of injected dose at all time points. But DNA when formulated in LPD was trapped in lung to an extent of 40% within 5 min of LPD injection. In Southern blot analysis, it was revealed that the pDNA injected with LPD was detectable in the lung in a much higher quantity than the control, even after 6 h of i.v. injection. Intraportal injection gave significantly lower gene expression than with i.v. injection.

6.2 Polycation-condensed DNA in Anionic Liposomes: LPD-II Formulation

Some disadvantages always accompany the cationic lipid-mediated gene transfection. Cytotoxic effect in cells/tissues is the primary concern that requires early attention. The cationic lipids exhibit relatively large sizes, while complexation with DNA provide suboptimal DNA condensation and have limited efficiency and lack of tissue specificity, whereas anionic liposomes used in gene therapy have shown poor encapsulation efficiencies due to the large size and excessive negative charge of uncondensed pDNA. The utilization of anionic liposome was revisited during a recent approach, which also utilized the concept of condensing DNA by cationic polymers. A delivery vector was developed wherein polylysine-condensed pDNA was entrapped into folate-targeted anionic liposomes via charge interaction. It had structural similarity as LPD-I and it was named *LPD-II*. It differed from LPD-I in that anionic lipids instead of cationic lipids were used. This novel vector was more efficient and less cytotoxic compared to conventional cationic liposomal vectors.

Folate-targeted LPD-II particles were generated by mixing anionic liposomes composed of DOPE/CHEMS/folate-PEG-DOPE and the cationic DNA–polylysine (1 : 0.75, w/w) complexes. Structural analysis of LPD-II by negative-stain EM showed that the DNA–polylysine (which appears as rod shaped individually) and lipid complex appeared as a highly electron dense, spherical core with a low-density coating. The mean diameter of these particles was 74 ± 14 nm, that is, smaller than the empty liposomes.

KB cells expressing folate receptors were transfected with LPD-II particles containing luciferase reporter gene and high transfection efficiency was observed. The activity could be inhibited by the presence of excess free folate. Control LPD-II particles generated with nontargeted liposomes was only active at low lipid/DNA ratios, suggesting that the transfection by LPD-II particles was only receptor dependent when the overall charge was negative. Compared with DC-Chol/DOPE/DNA-liposome complexes, LPD-II showed ~20 to 30-fold more transfection activity. On replacing DOPE with DOPC in the original formulation, the transfection was severely reduced. This indicated that the fusogenic activity of DOPE was essential for the transfection activity of LPD-II particles.

Their low physiological stability, inefficient cellular uptake, and the lack of tissue specificity currently limit the therapeutic applications of antisense oligonucleotides (ODN). The use of various vectors renders phosphodiester (PO) ODN resistant to enzymatic digestion. KB cells, which overexpress folate binding protein, were also transfected with LPD-II containing targeting ligand folate to deliver ODN against epidermal growth-factor receptor (EGFR). This resulted in downregulation of EGFR and growth inhibition of KB cells. The modified backbone-ODN, namely, phosphorothioate (PS) and monomethylphosphonate (MP) are more stable to enzymatic degradation compared to PO ODN, but they suffer from increased toxicity and decreased specificity. In a study, PO ODN against EGFR had shown growth inhibitory effect to KB cells compared to that of PS/PO ODN when delivered with LPD-II, indicating that LPD-II could also protect PO ODN from attack by enzymes inside cells.

7
Emulsion-mediated Gene Transfer

7.1
General Development

Emulsions are one of the most widely studied colloidal dispersion systems for the delivery of drugs. The oil-in-water emulsion is made of oil dispersed in an aqueous phase with a suitable emulsifier such as phospholipids, nonionic or ionic surfactants. Castor oil or soybean oil is predominantly used as the core oil phase. Nonionic surfactants such as Tween, Span, Brij, and pluronic copolymers are used as coemulsifiers. The ionic coemulsifiers are phospholipids or cationic lipids. A number of structure/activity studies had been done with different emulsion formulations, which were subsequently used for gene delivery *in vitro* and *in vivo*. The nonionic surfactants such as Tween, Span, Brij, and pluronic copolymers were found to be excellent coemulsifiers when used along with castor oil, DC-Chol, and DOPE. The *in vitro* transfection study on BL-6 cells showed that the Tween surfactant containing formulations had more serum resistivity and exhibited higher transfection in serum-containing media than in the absence of serum. One of the Brij containing formulations, that is, one with 2-oxyethylene chains, showed the highest transfection efficiency in the presence of serum. The toxicity of each formulation is minimal. In a DOTAP, soybean oil, and pDNA emulsion complex, it was observed that in spite of the change of zeta potential with varying amounts of DNA, the structure and the size of the emulsion complex remained mostly unchanged. The stability of this emulsion complex was shown to be high, and inhibited DNase-I digestion of pDNA. In serum-containing media, the emulsion showed much higher transfection efficiency compared to lipofectamine/DNA transfection complex. On inclusion of polyethyleneglycol-PE in the emulsion complex, a high level transfection was observed even in the media containing 90% of serum. This result suggested that *in vivo* transfection could be done with this emulsion complex. Another soybean oil-DOTAP emulsion was used to transfer genes to the epithelial cells of the mouse nasal cavity via intranasal instillation. The emulsion showed enhanced stability against heparin exchange and exhibited higher level of transfection compared to commercially available liposome/DNA complex in nasal cavity mucosa.

7.2
Reconstituted Chylomicron Remnants for Gene Transfer

Chylomicrons are triglyceride rich lipoproteins that are slowly modified during circulation in blood. The core glyceride structure is hydrolyzed by the lipoprotein lipase. It is by the apolipoprotein-specific receptors that the hepatocytes in the liver consume these remnants. It had been shown that even reconstituted chylomicron remnants (RCR) formed by purified lipids could also be taken up by the hepatocytes following i.v. injections. In our lab, we modified the system and used it for delivery of DNA into liver cells. DNA being predominantly hydrophilic in nature cannot be included in the hydrophobic interior of the RCR. TC-Chol, a quaternary ammonium analog of DC-Chol, was used in various amounts to make complexes with DNA. The resultant hydrophobic complexes upon extraction from aqueous solution were emulsified with commercially available

lipids (olive oil: L-α-phosphatidyl choline: L-α-lysophosphatidyl choline: cholesteryl oleate) by homogenization. DNA/TC-Chol complex was incorporated into the internal oil space of RCR and remained protected against Dnase-I digestion. The RCR containing DNA/TC-Chol complex, injected intraportally, gave the transgene luciferase expression in the liver 100-fold higher than with naked DNA injection. The expression was also obtained in the spleen, lung, and heart, but was 25 to 800-fold lower than that in the liver. The gene expressions obtained through tail vein injection was 100-fold lower than that of the mice injected intraportally. This was likely because the RCR did not contain any apolipoproteins on its surface, which could have otherwise facilitated the receptor-mediated uptake. Moreover, the colloidal stability might not have been as good as the ones containing apolipoproteins. RCR did not employ any protein or peptide that acts as an antigen. So, it could be injected repetitively by using a catheter method that had been established for multiple portal vein infusion. We have recently improved the formulation by adding pegylated lipid to the surface of RCR and showed that the circulation time of the particle was prolonged significantly.

8 Use of Cationic Liposome/DNA Complex (Lipoplex)

Lipoplexes are widely used for transferring genes to various *in vivo* targets with comparative ease and least toxic effect to the host. Use of lipoplexes in gene therapy clinical trials comprises almost one-fourth of the total gene therapy clinical trials. pDNA containing various therapeutic genes were used to treat a variety of disease pathologies. A vast amount of literature is available on the *in vitro*, preclinical, and clinical evaluations. Lipoplexes enjoy the property of easy administration in body without eliciting any extradiscomfort to the patient or the animal. Intravenous, intradermal, intraperitoneal, and direct injections are the common ways through which lipoplexes are injected in to the body. In the targeting front, several ligand associated cationic liposomes were developed to target genes to specific cell targets such as cancer. Recently, a cationic liposomal stealth gene delivery system was developed utilizing a sigma ligand associated polyethylene glycol–based lipid. The system could target and deliver genes specifically to sigma receptor overexpressing breast adenocarcinoma cells without showing any targeting ability to sigma receptor non-overexpressing cancer cells. Additionally, the stealth lipoplex was serum stable. The PEG lipid was chemically modified with a neuroleptic drug, haloperidol (used for over three decades to treat psychotic patients), to target cancer cells that overexpress sigma receptors. No conclusive links were ascertained between cancer pathogenicity and overexpressing sigma receptors in certain cancer cells. But, this versatile sigma receptor targeting stealth cationic liposomal system could be potentially used to target therapeutic anticancer genes to breast cancer or other cancer models that were implicated with overexpressing sigma receptors.

8.1 *In Vivo* Biodistribution Through Intravenous Injection: Therapeutic Implications

Intravenous injection is a widely used mode of gene delivery in animals. On i.v. lipoplex injection *in vivo*, the residence

of DNA primarily in heart and lungs even after nine days with minimal toxicity was observed. Intravenous injected lipoplex expressed transgene in almost all organs including lung, kidney, heart, spleen, liver, brain and so on, and the expression stayed for nine weeks with apparently no treatment related toxicity. Toxicity and antitumor response was evaluated on mice and pigs with high doses of lipoplex containing *MHC* gene incorporated pDNA.

Recently, more work has been done to increase the overall transfection efficiency with much higher targeting capability and reproducibility of the liposomal delivery system. Intravenously administered lipoplex avidly reacts with blood components. So, it is necessary to keep the complex intact in the blood till it reaches the organ of interest. It was shown that in the presence of erythrocytes, cationic lipid/cholesterol formulation did not induce fusion between erythrocytes, whereas the cationic lipid/DOPE formulation possessing high fluidity in its structure induced fusion between the erythrocytes after a short incubation period. This offered an explanation as to why cholesterol makes a superior formulation with cationic lipids for *in vivo* purpose.

A repeated systemic i.v. injection of cytokine gene (*IFNβ*1) by lipoplexes gave a systemic expression of human interferon-β in mice, thereby increasing the possibility of cytokines used for therapeutic purposes in a systemic manner. An enhanced and highly selective liver targeting by i.v. injection of cationic lipoplex containing β-sitosterol beta-D-glucoside (*Sit-G*) gene was also observed. Intravenous injections showed gene expression in all major organs including heart, lung, liver, spleen, and kidney, with lung being most efficiently transfected. For efficient targeting and gene expression in the lung, i.v. injection was favored over intratracheal instillation. Uyechi et al. had shown by injecting fluorescently tagged lipoplexes through the vein that the entire lung lobe was homogeneously fluorescent, whereas intratracheal administration resulted in regional distribution of lipoplex, concentrated around bronchioles and distal airways.

LPD-I lipopolyplex particles upon i.v. injection were also distributed to spleen along with the major organs, where they were endocytosed and released to the cytoplasm. Administrations of these particles containing unmethylated CpG motif in pDNA elicit immunostimulation that subsequently showed tumorostatic effect. The effect of CpG will be discussed in detail in the later part of this chapter. However, the spleen delivery idea was utilized to deliver antigenic peptides entrapped in LPD particles to the cytoplasm of splenic antigen presenting cells. It is known that *E6* and *E7* are two oncogenes responsible for the maintenance of the malignant state of HPV-positive tumors. Therefore, an arginine containing small peptide epitope derived from tumor antigen, HPV-E7 protein was electrostatically attached to pDNA. The preattached peptide-DNA was condensed by protamine and finally the whole entity was enclosed in LPD using cationic lipid. The particle upon i.v. injection was delivered to the splenic APCs and thereby induced E7 mediated immune response against syngeneic HPV-induced tumor in mice.

8.2 Directly Injected Lipoplexes Used Against Cancer and Other Diseases

Direct injection to tissue is a common approach for cationic lipid-mediated gene

therapy. Intratumoral injection of DNA-liposome complexes containing either E6 or E7 antisense plasmid resulted in significant growth inhibition of C3 tumors grown in a syngeneic mouse model. A direct injection of recombinant pDNA containing murine class I major histocompatibility complex (*MHC*) gene into localized arterial segment of various major organs showed that the direct gene transfer by liposomal transfection did not lead to treatment related toxicity, autoimmunity, or gonadal localization of the transgene in mice. The toxicity of gene delivery by DNA liposomes was also analyzed in pigs and rabbits *in vivo*. There were no clinically significant immunopathology in major organs such as, brain, heart, lung, liver, kidney, spleen, and skeletal muscles. To stimulate local tumor immunity in patients with stage IV melanoma, Nabel et al. injected pDNA encoding MHC class I protein complexed with cationic lipid directly into the cutaneous tumor nodules. Treated lesions exhibited the presence of T cell, followed by an enhanced reactivity of tumor infiltrating lymphocytes. As a result, local inhibition of tumor followed by complete diminution of tumor was observed in some of the treated patients. Mohr et al. had shown that direct lipolex injection to intrahepatic hepatocellular carcinoma produced by human HCC cells appeared far superior to systemic administration for gene therapy for localized intrahepatic tumors, because the direct administration to tumors left the surrounding normal hepatic cells untouched. In another typical example, *in vivo* direct intratumor injection of plasmid containing the coding sequence for the human *IL-2* gene complexed with cationic lipid formulation resulted in retention of intact pDNA in the tumor tissue and IL-2 secretion by cell cultures derived from the injected tumors. Formulation of this lipid with the cationic lipid inhibited DNA degradation and enhanced *in vivo* transfection efficiency over pDNA alone.

Airway administration of liposome complexes was used for the treatment of pulmonary diseases including cystic fibrosis (CF). Cationic liposome/DNA complex showed no adverse effect toward airway epithelial integrity; hence, the cationic lipid-based delivery system proved to be appropriate for use in human trials for CF. A series of preclinical trials were done in CF patients with intranasal instillation to evaluate the risk factors associated with the treatment. Since there was no apparent toxicity associated with lipoplexes as was seen from these trials, progress has been made in delivering the complexes to the entire lung by aerosol in CF patients. By nebulization, lipoplex was delivered into the airways of mutant mice to obtain human cystic fibrosis transmembrane conductance regulator (CFTR) cDNA expression in the respiratory tract. A study conducted on CF patients revealed that pCMV-CFTR/cationic liposome complex on administration to the nasal epithelium gave no evidence of excess nasal inflammation, or any adverse events related to active treatment. Gene transfer and expression assayed by PCR revealed the presence of transgene DNA in seven of the eight treated patients up to 28 days after treatment. Intranasal instillation technique was also used in the mouse model, to incorporate cationic lipid/DNA complexes. The CFTR dysfunctional gene gives rise to several multiple defects in airway epithelia, one of them being altered Cl^- and Na^+ permeability. Zhang et al. had shown that the goblet cells were more efficiently targeted with lipoplexes than any other cells in the entire spectrum of lung airway epithelia. This was ascertained by

the fact that an efficiently reduced mucous sulfation to levels seen in non-CF airways was observed with lipoplex/DNA in spite of low levels of *CFTR* gene expression in lung epithelial cells in human bronchial xenograft model of mice as compared to non-CF airways of control mice. These kinds of apparent complexities in CFTR function presented challenges in the design of different lipoplex formulations that were capable of generating the endogenous patterns of *CFTR* gene expression in specific lung epithelial cells.

8.3 Intraperitoneal Injection of Lipoplex for Treating Cancer

IP injection of lipoplex was done to transfect cells in the peritoneum region. Nude mice bearing disseminated human ovarian tumors derived from the p185-overexpressing SKOV-3 ovarian cancer cells were injected with *E1A* gene/lipoplex intraperitoneally. These tumors resemble stage III of human ovarian cancer. The expression of E1A protein decreased the expression of p185 oncoprotein, and hence increased the survival rate of mice. Seventy percent of the treated group survived for 1.5 years from the last injection, but the untreated group barely survived more than 16 weeks. The treatment of complex containing 1/13 of the original lipid dose also worked as efficiently as the normal dose. There was no apparent toxicity or major organ pathologic change. There was no trace of E1A DNA in the liver, lung, heart, spleen, brain, uterus, and ovaries of the treated mice even after $1\frac{1}{2}$ years.

9 Gene Delivery by Polymeric Systems (Polyplex)

9.1 General Development and Uses

An alternative nonviral gene delivery system has been developed, which utilizes polymers, cationic or neutral. Cationic polymers are predominantly used since they efficiently condense DNA to very small particles. These complexes are called *polyplexes*. Wu et al. for the first time utilized polylysine-asialo-orosomucoid conjugate to condense pDNA and targeted the pDNA to liver. pDNA condensed with protamine/polylysine conjugate of iron transport protein, transferrin, was efficiently delivered to eukaryotic cells. pDNA was delivered to primary peripheral blood lymphocytes utilizing CD3 T-cell receptors. pDNA condensed by CD3 antibody–polylysine conjugate showed receptor-mediated endocytosis to efficiently internalize pDNA into T lymphocytes. Depending on the cell-binding ligand, specific targeting was obtained in different cell lines.

Several of the most effective polymeric delivery systems were polyamidoamine (PAMAM) dendrimers and polyethyleneimines (PEI). Being nonbiodegradable, these synthetic polymers posed a potential toxicity to cell, and hence biodegradable polypeptides like PLL and protamine were used for condensation and delivery of gene but with limited efficacy in transfection. They were usually used with cationic lipids to obtain enhanced transfection activity. Among the biodegradable polymers, chitosan and β-cyclodextrin–based polymers were also used for gene transfection.

Although the cationic polymers shared the same mechanism of DNA condensation, the transfection efficiency greatly

varied between polymers. Even different molecular weight and isomeric forms of the same polymer showed different physicochemical characteristics, transfection efficiency, and toxicity.

Ligand conjugated polymers were used for *in vivo* targeting and expression of gene. Kircheis et al. reported an enhanced level of transfection in subcutaneous Neuro2a tumors on intratumoral injection of transferrin-PEI/pDNA compared with naked pDNA injection. On pegylation, Tf-PEI/pDNA complex became serum resistant without losing its targetability, and the complex could be efficiently targeted to neuro2a tumors in mouse tumor model after i.v. injection. Neurotensin cross-linked PLL polyplexed with pDNA upon injection into the substantia nigra of rat, showed a high level of transfection in nigrostriatal dopamine neurons that was detectable up to 15 days.

EBV-based plasmid vector containing thymidine kinase (*TK*) gene coupled with PAMAM dendrimer (EBV/polyplex) was used in suicide gene therapy of cancer. Intratumoral injection of *TK* gene containing EBV/dendrimer polyplex into Huh7 HCC tumors in severly compromised immuno-deficient (SCID) mice showed remarkable suppression of tumor growth leading to prolongation of survival time. Gene transfer to the lung was obtained by i.v. injection of G9 PAMAM dendrimer-complexed pDNA into mice. This resulted in high levels of transgene expression in the alveoli at 12 and 24 h, followed by a second peak of expression three to five days after administration. But bronchial epithelium was primarily targeted with the direct endobronchial administration of this polyplex. Topical *in vivo* delivery of pDNA to the hairless mice skin was done with PAMAM dendrimer polyplex. The polyplex was incorporated in or coated on the surface of poly (DL-lactide-co-glycolide) (PLGA) or collagen-based biodegradable membranes.

In an effort to transfer gene to rabbit carotid artery, Turunen et al. used DNA/fractured dendrimer (generation 6) polyplex to obtain a high level of gene transfer (4.4%) compared with what was obtained by lipoplex. The arterial gene transfer was particularly useful because it could be used as a tool for treating various vascular diseases. *In vivo* gene transfer method, in this study, employed a gene delivery reservoir (collar) around the carotid artery, which served as a reservoir for the gene delivery solution. This type of local gene transfer with cationic polyplex provided a technically efficient way of treating arterial diseases during vascular surgeries, such as by-pass and other surgeries. These dendrimer-based polyplexes showed a clear advantage over polylysine in that polyplexes largely buffer the drop of endosomal pH inside the endosome leading to increase in *in vitro* transfection efficiencies.

Qin et al. used starburst PAMAM dendrimers to transfer genes into a murine cardiac transplantation model. These starburst dendrimers are a special class of its kind, which are highly branched spherical polymers with large numbers of amino groups on the surface. At the time of transplantation of whole heart in the recipient mice, the dendrimer/β-gal pDNA was directly injected into the graft tissue. X-gal staining revealed a highly efficient and widespread transgene expression in both myocytes and the graft infiltrating cells with the peak lasting up to 14 days. For organ transplantation, severe tissue-rejection is a common immune response. Viral IL-10, a cytokine synthesis inhibitory factor is able to regulate a variety of negative immune responses by suppressing the

synthesis of IFN-γ, or inhibiting IL-1, IL-6, IL-8, IL-12, and TNF-α. Direct injection of this dendrimer/viral *IL-10* gene with α-MHC promoter polyplex to the cardiac tissue showed an increased survival of the cardiac allograft. As little as 0.31 μg of the injected pDNA led to an increased mean survival from 13.9 to 38.6 days.

9.2 Targeted Gene Delivery by Antibodies Conjugated with Polycations

Polyamines had shown efficient transfection to lung endothelium. An efficient targeted transfection vector to lungs could be achieved by conjugating a targeting ligand against platelet endothelial cell adhesion molecule-1 (PECAM-1) to polyamines. This ligand–polyamine complex was targeted efficiently to the pulmonary endothelial cells. A chemical vector was synthesized by covalent conjugation of polyethylenimine and anti-PECAM antibody (Ab). The cationic complex was shown to deliver DNA specifically to mouse lung endothelial cells. The highest gene expression was obtained at relatively low plus-to-minus charge ratios. The PEI, conjugated with a control IgG, did not enhance transfection of mouse lung endothelial cells. Intravenous injection of this anti-PECAM Ab-PEI /DNA showed an increased lung expression in mice compared to other modes of injection.

Immunostimulatory cytokine, TNF-α in blood was found to be about fivefold less in the mice injected with PEI-anti-PECAM Ab/DNA than in the mice injected with PEI/DNA. The decrease in TNF-α was partially reversible by pretreating mice with Ab to PECAM. Immunosuppressant dexamethasone, when injected along with anti-PECAM Ab/DNA, not only improved the persistence and level of gene expression in the lung but also shortened the refractory period for repeated dosing. This supported a potential therapeutic role of dexamethasone in lung gene transfer with Ab-polymer conjugates.

Similarly, Ferkol et al. targeted the polymeric immunoglobulin receptor (pIgR), which are expressed in lung and liver tissues and transferred pDNA to these tissues. The targeting ligand, anti-secretory component (SC) Fab antibody was covalently linked to PLL. The polycation upon condensing pDNA was delivered successfully to airway epithelium *in vivo*. Tissues that do not express the pIgR, spleen, and heart, were not transfected. In addition, conjugate prepared with irrelevant Fab fragments did not produce detectable transgene activity. This complex specifically targeted pIgR-bearing tissues, but after repeated dosing increased humoral immune response against anti-SC Fab antibody was observed.

10 Nonviral Vector Related Cytotoxicity

It is known from the 1980s that the bacterial DNA stimulates the formation of cytotoxic IFN-α, β, and IL-12 when the DNA is taken up by macrophages. It in turn leads to NK cell activation and production of proinflammatory cytokine IFN-γ. This is accompanied by the proliferation of B cell and hence reduction of apoptosis and release of IL-6 and IL-12. These proinflammatory effects were found to be due to some immunostimulatory sequences in prokaryotic DNA that contained unmethylated CpG dinucleotide motif flanked by two 5′ purines and two 3′ pyrimidines. pDNA, which is derived from bacterial DNA, induces these immune

responses. The unmethylated CpG motif–containing sequence occurs four times more frequently in prokaryote DNA than in eukaryotic DNA. Moreover, the CpG motifs are usually 75% more methylated in mammalian DNA than in prokaryotic DNA. On methylation of the cytosine bases in plasmid, the immunostimulatory effect is decreased considerably. Immature dendritic cells tend to produce proinflammatory IFN-α, β, IL-6, IL-12, and TNF-α on exposure to CpG containing oligonucleotide or bacterial plasmid.

These immune-stimulatory effects leading to inflammatory cytokine productions had negative impact on the systemic gene delivery of cationic lipid/DNA complexes. On recognition of pDNA by splenic macrophages during circulation, there was every chance that the pDNA would elicit an immune response, which might lead to phagocytosis of the complex and hence to decreased transgene expression. A high level of proinflammatory cytokine also led to inactivation of several promoters, resulting in a decrease in transgene expression. The death of animals by high dose lipoplex injection for obtaining high transgene expression could be attributed both to the high concentration of lipid and to pDNA mediated toxicity. In the case of local lipoplex administration in animals, minimal toxicity was observed. However, on i.v. injection or intratracheal instillation, high levels of IFN-γ and TNF-α were observed. Pretreating cytokine-neutralizing antibodies during i.v. injection of lipoplex, CpG triggered inflammation and immune responses were minimized and prolonged gene expression was obtained. Repeated dosing without any antibody treatment led to silenced transgene expression after one or two weeks.

It is not clear though how cytokine production decreases the transgene expression, but various ongoing efforts to minimize CpG related immune responses and toxicity, and to enhance transgene expression is worth mentioning. Hofmann et al. had used PCR-amplified fragments containing encoded therapeutic gene and regulatory elements for preparing LPD. On delivering, a similar level of gene expression comparable to pDNA lipoplex was obtained. However, a much lower level of cytokine response was observed, which sustained the gene expression for a longer period than pDNA. PCR fragment contains fewer CpG motifs than the full-length pDNA, which led to reduced CpG triggered adverse effects. Yew et al. had shown that on mutating CpG or its flanking motifs in the plasmid, a decreased level of cytokine and increased transgene expression could be obtained. A limited interaction of plasmid with immune cells could also lead to decreased cytokine response. It could be achieved by sequentially injecting cationic liposomes and free pDNA. Song et al. used this process for the purpose of efficiently transfecting the lung by prolonging the residency time and interaction of DNA with pulmonary endothelium. Tan et al. used the same concept to show that sequential injection led to the formation of a lower level of cytokines as compared to lipoplex injection. It is evident from the above efforts that understanding the detailed mechanism of CpG induced immune response is required for increasing the efficacy of pDNA mediated gene delivery.

11 Conclusion

Since the first attempt of utilizing pDNA for gene delivery to cells, the two decades have seen many attempts to enhance the

efficacy of genetic vectors, to test various therapeutically important genes, to understand the mechanism of gene delivery, associated toxicity, and factors inhibiting gene expression. A number of gene therapy clinical trials were also performed. Although no single vector is superior to other vectors, each *in vivo* gene transfer application will find its vector system for optimal performance. Understanding cellular barriers and possible means to overcome them will undoubtedly further improve the performance of a nonviral vector.

Acknowledgment

The original work from Leaf Huang's lab described in this review has been supported by NIH grants CA 74918, AR 45925, DK 44935, AI 48851. ''Fast Track Research Grant for Young Scientists'' from Department of Science & Technology, Govt. of India supported the work done in Rajkumar Banerjee's lab as described in this chapter.

See also Autoantibodies and Autoimmunity; Vascular Development and Angiogenesis; Oncology, Molecular; Plasmids; Somatic Gene Therapy; Therapeutic Angiogenesis; Vectors and Gene Therapy.

Bibliography

Books and Reviews

Bird, A.P. (1986) CpG-rich islands and the function of DNA methylation, *Nature* **321**, 209–213.

Budker, V., Budker, T., Zhang, G., Subbotin, V., Loomis, A., Wolff, J.A. (2000) Hypothesis: naked plasmid DNA is taken up by cells in vivo by a receptor-mediated process, *J. Gene Med.* **2**, 76–88.

Chonn, A., Cullis, P.R. (1995) Recent advances in liposomal drug-delivery systems, *Curr. Opin. Biotechnol.* **6**, 698–708.

Davidson, J.M., Krieg, T., Eming, S.A. (2000) Particle-mediated gene therapy of wounds, *Wound Repair. Regen.* **8**, 452–459.

Davis, L.I. (1995) The nuclear pore complex, *Annu. Rev. Biochem.* **64**, 865–896.

Demeneix, B., Behr, J.-P., Boussif, O., Zanta, M.A., Abdallah, P., Remy, J.-S. (1998) Gene transfer with lipospermines and polyethylenimines, *Adv. Drug Delivery Rev.* **30**, 85–95.

Groth, A.C., Calos, M.P. (2004) Phage integrases: biology and applications, *J. Mol. Biol.* **335**, 667–668.

Huang, L., Li, S. (1997) Liposomal gene delivery: a complex package, *Nat. Biotechnol.* **15**, 620–621.

Jaroszeski, M.J., Gilbert, R., Nicolau, C., Heller, R. (1999) In vivo gene delivery by electroporation, *Adv. Drug. Delivery Rev.* **35**, 131–137.

Kamps, J.A., Scherphof, G.L. (1998) Receptor versus non-receptor mediated clearance of liposomes, *Adv. Drug Delivery Rev.* **32**, 81–97.

Korzarsky, K.F., Wilson, J.M. (1993) Gene therapy: adenovirus vectors, *Curr. Opin. Genet. Dev.* **3**, 499–503.

Ledley, F.D. (1995) Nonviral gene therapy: the promise of genes as pharmaceutical products, *Hum. Gene Ther.* **6**, 1129–1144.

Marshall, D.J., Leiden, J.M. (1998) Recent advances in skeletal-muscle-based gene therapy, *Curr. Opin. Genet. Dev.* **8**, 360–365.

Mazda, O. (2002) Improvement of nonviral gene therapy by Epstein-Barr virus (EBV)-based plasmid vectors, *Curr. Gene Ther.* **2**, 379–392.

Mir, L.M., Bureau, M.F., Rangara, R., Schwartz, B., Scherman, D. (1998) Long-term, high level in vivo gene expression after electric pulse-mediated gene transfer into skeletal muscle, *Crit. Rev. Acad. Sci. III* **321**, 893–899.

Pante, N., Aebi, U. (1996) Molecular dissection of the nuclear pore complex, *Crit. Rev. Biochem. Mol. Biol.* **31**, 153–199.

Singh, M., Ravin, L.J. (1986) Parenteral emulsions as drug carrier systems, *J. Parenter. Sci. Technol.* **40**, 34–41.

Somiari, S., Malone, J.G., Drabick, J.J., Gilbert, R., Heller, R., Jaroszeski, M.J., Malone, R.W. (2000) Theory and in vivo application of electroporative gene delivery, *Mol. Ther.* **2**, 178–187.

Spooner, R.A., Deonarain, M.P., Epenetos, A.A. (1995) DNA vaccination for cancer treatment, *Gene Ther.* **2**, 173–180.

Webster, K.A., Prentice, H., Discher, D.J., Hicks, M.C., Bisphoric, N.H. (1998) in: Wheelan, W.J. (Ed.) *Molecular Biology in the Conquest of Disease*, Oxford University Press, Oxford, UK, p. 37.

Wolff, J.A. (1999) in: Friedmann, T. (Ed.) *Development of Human Gene Therapy*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, p. 279.

Yang, N.S., Hogge, G.H., MacEwen, E.G. (1999) in: Huang, L. Hung, M.C. Wagner, E. (Eds.) *Nonviral Vectors for Gene Therapy*, Academic Press, San Diego, CA, p. 171.

Primary Literature

Acsadi, G., Dickson, D.R., Love, A., Jani, A., Walsh, F.S., Gurusinghe, A., Wolff, J.A., Davies, K.E. 1991. Human dystrophin expression in mdx mice after intramuscular injection of DNA constructs. *Nature* **352**, 815–818.

Aihara, H., Miyazaki, J. (1998) Gene transfer into muscle by electroporation in vivo, *Nat. Biotechnol.* **16**, 867–870.

Akao, T., Nakayama, T., Takeshia, K., Ito, A. (1994) Design of a new cationic amphiphile with efficient DNA-transfection ability, *Biochem. Mol. Biol. Int.* **34**, 915–920.

Alton, E.W., Middleton, P.G., Caplen, N.J., Smith, S.N., Steel, D.M., Munkonge, F.M., Jeffery, P.K., Geddes, D.M., Hart, S.L., Williamson, R., et al. (1993) Non-invasive liposome-mediated gene delivery can correct the ion transport defect in cystic fibrosis mutant mice, *Nat. Genet.* **5**, 135–142.

Alvarez-Maya, I., Navarro-Quiroga, I., Meraz-Rios, M.A., Aceves, J., Martinez-Fong, D. (2001) In vivo gene transfer to dopamine neurons of rat substantia nigra via the high-affinity neurotensin receptor, *Mol. Med.* **7**, 186–192.

Andree, C., Swain, W.F., Page, C.P., Macklin, M.D., Slama, J., Hatzis, D., Eriksson, E. (1994) In vivo transfer and expression of a human epidermal growth factor gene accelerates wound repair, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12188–12192.

Anwer, K., Shi, M., French, M.F., Muller, S.R., Chen, W., Liu, Q., Proctor, B.L., Wang, J., Mumper, R.J., Singhal, A., Rolland, A.P., Alila, H.W. (1998) Systemic effect of human growth hormone after intramuscular injection of a single dose of a muscle-specific gene medicine, *Hum. Gene Ther.* **9**, 659–670.

Ardehali, A., Fyfe, A., Laks, H., Drinkwater, D.C., Jr., Qiao, J.H., Lusis, A.J. (1995) Direct gene transfer into donor hearts at the time of harvest, *J. Thorac. Cardiovasc. Surg.* **109**, 716–719.

Aronsohn, A.I., Hughes, J.A. (1998) Nuclear localization signal peptides enhance cationic liposome-mediated gene therapy, *J. Drug Target.* **5**, 163–169.

Bailey, A.L., Cullis, P.R. (1997) Membrane fusion with cationic liposomes: effects of target membrane lipid composition, *Biochemistry* **36**, 1628–1634.

Banerjee, R., Das, P.K., Chaudhuri, A. (1998) Interfacial indazolization: novel chemical evidence for remarkably high exo-surface pH of cationic liposomes used in gene transfection, *Biochim. Biophys. Acta* **1373**, 299–308.

Banerjee, R., Das, P.K., Srilakshmi, G.V., Chaudhuri, A., Rao, N.M. (1999) Novel series of non-glycerol-based cationic transfection lipids for use in liposomal gene delivery, *J. Med. Chem.* **42**, 4292–4299.

Batteux, F., Tourneur, L., Trebeden, H., Charreire, J., Chiocchia, G. (1999) Gene therapy of experimental autoimmune thyroiditis by in vivo administration of plasmid DNA coding for Fas ligand, *J. Immunol.* **162**, 603–608.

Behr, J.P.-., Demeneix, B., Loeffler, J.-P., Perez-Mutul, J. (1989) Efficient gene transfer into mammalian primary endocrine cells with lipopolyamine-coated DNA, *Proc. Natl. Acad. Sci.* **86**, 6982–6986.

Belperron, A.A., Feltquate, D., Fox, B.A., Horii, T., Bzik, D.J. (1999) Immune responses induced by gene gun or intramuscular injection of DNA vaccines that express immunogenic regions of the serine repeat antigen from Plasmodium falciparum, *Infect. Immun.* **67**, 5163–5169.

Benita, S., Levy, M.Y. (1993) Submicron emulsions as colloidal drug carriers for intravenous administration: comprehensive

physicochemical characterization, *J. Pharm. Sci.* **82**, 1069–1079.

Benn, S.I., Whitsitt, J.S., Broadley, K.N., Nanney, L.B., Perkins, D., He, L., Patel, M., Morgan, J.R., Swain, W.F., Davidson, J.M. (1996) Particle-mediated gene transfer with transforming growth factor-beta1 cDNAs enhances wound repair in rat skin, *J. Clin. Invest.* **98**, 2894–2902.

Bennett, R.M., Gabor, G.T., Merritt, M.M. (1985) DNA binding to human leukocytes. Evidence for a receptor-mediated association, internalization, and degradation of DNA, *J. Clin. Invest.* **76**, 2182–2190.

Bennett, M.J., Aberle, A.M., Balasubramaniam, R.P., Malone, J.G., Malone, R.W., Nantz, M.H. (1997) Cationic lipid-mediated gene delivery to murine lung: correlation of lipid hydration with in vivo transfection activity, *J. Med. Chem.* **269**, 4069–4078.

Bettan, M., Emmanuel, F., Darteil, R., Caillaud, J.M., Soubrier, F., Delaere, P., Macfoudi, D.B, Duverger, N., Scherman, D. (2000) High-level protein secretion into blood circulation after electric pulse-mediated gene transfer into skeletal muscle, *Mol. Ther.* **2**, 204–210.

Bielinska, A.U., Yen, A., Wu, H.L., Zahos, K.M., Sun, R., Weiner, N.D., Baker, J.R., Jr., Roessler, B.J. (2000) Application of membrane-based dendrimer/DNA complexes for solid phase transfection in vitro and in vivo, *Biomaterials* **21**, 877–887.

Bottega, R., Epand, R.M. (1992) Inhibition of protein kinase C by cationic amphiphiles, *Biochemistry* **31**, 9025–9030.

Boulton, T.G., Nye, S.H., Robbins, D.J., Ip, N.Y., Radziejewska, S.D., Morgenbesser, R.A., DePinho, R.A., Panayotatos, N., Cobbs, M.H., Yancopoulos, G.D. (1991) ERKs: a family of protein-serine/threonine kinases that are activated and tyrosine phosphorylated in response to insulin and NGF, *Cell* **65**, 663–675.

Brisson, M., He, Y., Li, S., Yang, J.-P., Huang, L. (1999) A novel T7 RNA polymerase autogene for efficient cytoplasmic expression of target genes, *Gene Ther.* **6**, 263–270.

Bryant, M., Drew, G.M., Houston, P., Hissey, P., Campbell, C.J., Braddock, M. (2000) Tissue repair with a therapeutic transcription factor, *Hum Gene Ther* **11**, 2143–2158.

Buschle, M., Cotten, M., Kirlappos, H., Mechtler, K., Schaffner, G., Zauner, W., Birnsteil, M.L., Wagner, E. (1995) Receptor-mediated gene transfer into human T lymphocytes via binding of DNA/CD3 antibody particles to the CD3 T cell receptor complex, *Hum. Gene Ther.* **6**, 753–761.

Caplen, N.J., Alton, E.W., Middleton, P.G., Dorin, J.R., Stevenson, B.J., Gao, X., Durham, S.R., Jeffery, P.K., Hodson, M.E., Coutelle, C., Huang, L., Porteous, D.J., Williamson, R., Geddes, D.M. (1995) Liposome-mediated CFTR gene transfer to the nasal epithelium of patients with cystic fibrosis, *Nat. Med.* **1**, 39–46.

Chadwick, S.L., Kingston, H.D., Stern, M., Cook, R.M., O'Connor, B.J., Lukasson, M., Balfour, R.P., Rosenberg, M., Cheng, S.H., Smith, A.E., Meeker, D.P., Geddes, D.M., Alton, E.W. (1998) Safety of a single aerosol administration of escalating doses of the cationic lipid GL-67/DOPE/DMPE-PEG5000 formulation to the lungs of normal volunteers, *Gene Ther.* **5**, 937–942.

Chang, Y., Prud'homme, G.J. (1999) Intramuscular administration of expression plasmids encoding interferon-gamma receptor/IgG1 or IL-4/IgG1 chimeric proteins protects from autoimmunity, *J. Gene Med.* **1**, 415–423.

Chen, C.H., Ji, H., Suh, K.W., Choti, M.A., Pardoll, D.M., Wu, T.C. (1999) Gene gun-mediated DNA vaccination induces antitumor immunity against human papillomavirus type 16 E7-expressing murine tumor metastases in the liver and lungs, *Gene Ther.* **6**, 1972–1981.

Chesnoy, S., Lee, P.Y., Huang, L. (2003) Intradermal injection of transforming growth factor-beta1 gene enhances wound healing in genetically diabetic mice, *Pharm. Res.* **20**, 345–350.

Chesnoy, S., Durand, D., Doucet, J., Stolz, D.B., Huang, L. (2001) Improved DNA/emulsion complex stabilized by poly(ethylene glycol) conjugated phospholipid, *Pharm. Res.* **18**, 1480–1484.

Conry, R.M., LoBuglio, A.F., Loechel, F., Moore, S.E., Sumerel, L.A., Barlow, D.L., Curiel, D.T. (1995) A carcinoembryonic antigen polynucleotide vaccine has in vivo antitumor activity, *Gene Ther.* **2**, 59–65.

Coster, H.G. (1965) A quantitative analysis of the voltage-current relationships of fixed charge membranes and the associated property of "punch-through", *Biophys. J.* **5**, 669–686.

Cowdery, J.S., Chace, J.H., Yi, A.K., Krieg, A.M. (1996) Bacterial DNA induces NK cells to produce IFN-gamma in vivo and increases

the toxicity of lipopolysaccharides, *J. Immunol.* **156**, 4570–4575.

Daleke, D.L., Hong, K.L., Papahadjopoulos, D. (1992) Endocytosis of liposomes by macrophages: binding, acidification and leakage of liposomes monitored by a new fluorescence assay, *Biochim. Biophys. Acta* **1024**, 352–366.

Danko, I., Fritz, J.D., Jiao, S., Hogan, K., Latendresse, J.S., Wolff, J.A. (1994) Pharmacological enhancement of in vivo foreign gene expression in muscle, *Gene Ther.* **1**, 114–121.

Davis, N.L., Brown, K.W., Johnston, R.E. (1996) A viral vaccine vector that expresses foreign genes in lymph nodes and protects against mucosal challenge, *J. Virol.* **70**, 3781–3787.

Davis, H.L., Whalen, R.G., Demeneix, B.A. (1993) Direct gene transfer into skeletal muscle in vivo: factors affecting efficiency of transfer and stability of expression, *Hum. Gene Ther.* **4**, 151–159.

Dileo, J., Banerjee, R., Whitmore, M., Nayak, J.V., Falo, L.D., Jr., Huang, L. (2003) Lipid-protamine-DNA-mediated antigen delivery to antigen-presenting cells results in enhanced anti-tumor immune responses, *Mol Ther.* **7**, 640–648.

Dowty, M.E., Williams, P., Zhang, G., Hagstrom, J.E., Wolff, J.A. (1995) Plasmid DNA entry into postmitotic nuclei of primary rat myotubes, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4572–4576.

Dubendorff, J.W., Studier, F.W. (1991) Creation of a T7 autogene. Cloning and expression of the gene for bacteriophage T7 RNA polymerase under control of its cognate promoter, *J. Mol. Biol.* **219**, 61–68.

Dujardin, N., Van Der Smissen, P., Preat, V. (2001) Topical gene transfer into rat skin using electroporation, *Pharm. Res.* **18**, 61–66.

Duzgunes, N., Goldstein, J.A., Friend, D.S., Felgner, P.L. (1989) Fusion of liposomes containing a novel cationic lipid, N-[2,3-(dioleyloxy)propyl]-N,N,N-trimethylammonium: induction by multivalent anions and asymmetric fusion with acidic phospholipid vesicles, *Biochemistry* **28**, 9179–9184.

Dworetzky, S.I., Lanford, R.E., Feldherr, C.M. (1988) The effects of variations in the number and sequence of targeting signals on nuclear uptake, *J. Cell Biol.* **107**, 1279–1287.

Espinos, E., Liu, J.H., Bader, C.R., Bernheim, L. (2001) Efficient non-viral DNA-mediated gene transfer to human primary myoblasts using electroporation, *Neuromuscul. Disord.* **11**, 341–349.

Farhood, H., Serbina, N., Huang, L. (1995) The role of dioleoyl phosphatidylethanolamine in cationic liposome mediated gene transfer, *Biochim. Biophys. Acta* **1235**, 289–295.

Fazio, V.M., Fazio, S., Rinaldi, M., Catani, M.V., Zotti, S., Ciafre, S.A., Seripa, D., Ricci, R., Farace, M.G. (1994) Accumulation of human apolipoprotein-E in rat plasma after in vivo intramuscular injection of naked DNA, *Biochem. Biophys. Res. Commun.* **200**, 298–305.

Feldherr, C.M., Akin, D. (1990) The permeability of the nuclear envelope in dividing and nondividing cell cultures, *J. Cell Biol.* **111**, 1–8.

Felgner, P.L., Gadek, T.R., Holm, M., Roman, R., Chan, H.W., Wenz, M., Northrop, J.P., Ringold, G.M., Danielsen, M. (1987) Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7413–7417.

Felgner, J.H., Kumar, R., Sridhar, C.N., Wheeler, C.J., Tsai, Y.J., Border, R., Ramsey, P., Martin, M., Felgner, P.L. (1994) Enhanced gene delivery and mechanism studies with a novel series of cationic lipid formulations, *J. Biol. Chem.* **269**(:) 2550–2561.

Ferkol, T., Perales, J.C., Eckman, E., Kaetzel, C.S., Hanson, R.W., Davis, P.B. (1995) Gene transfer into the airway epithelium of animals by targeting the polymeric immunoglobulin receptor, *J. Clin. Invest.* **95**, 493–502.

Ferkol, T., Palle, A.P., Eckman, E., Perales, J.C., Trzaska, T., Tosi, M., Redline, R., Davis, P.B. (1996) Immunologic responses to gene transfer into mice via the polymeric immunoglobulin receptor, *Gene Ther.* **3**, 669–678.

Fischer, D., Bieber, T., Li, Y., Elsasser, H., Kissel, T. (1999) A novel non-viral vector for DNA delivery based on low molecular weight, branched polyethylenimine: effect of molecular weight on transfection efficiency and cytotoxicity, *Pharm. Res.* **16**, 1273–1279.

Floch, V., Loisel, S., Guenin, E., Herve, A.C., Clement, J.C., Yaouanc, J.J., des Abbayes, H., Ferec, C. (2000) Cation substitution in cationic phosphonolipids: a new concept to improve transfection activity and decrease cellular toxicity, *J. Med. Chem.* **43**, 4617–4628.

Folkman, J., Merler, E., Abernathy, C., Williams, G. (1971) Isolation of a tumor factor

responsible for angiogenesis, *J. Exp. Med.* **133**, 275–288.

Fraley, R., Subramani, S., Berg, P., Papahadjopoulos, D. (1980) Introduction of liposome-encapsulated SV40 DNA into cells, *J. Biol. Chem.* **255**, 10431–10435.

Friend, D.S., Papahadjopoulos, D., Debs, R.J. (1996) Endocytosis and intracellular processing accompanying transfection mediated by cationic liposomes, *Biochim. Biophys. Acta* **1278**, 41–50.

Fuller, D.H., Murphey-Corb, M., Clements, J., Barnett, S., Haynes, J.R. (1996) Induction of immunodeficiency virus-specific immune responses in rhesus monkeys following gene gun-mediated DNA vaccination, *J. Med. Primatol.* **25**, 236–241.

Gainer, A.L., Suarez-Pinzon, W.L., Min, W.P., Swiston, J.R., Hancock-Friesen, C., Korbutt, G.S., Rajotte, R.V., Warnock, G.L., Elliott, J.F. (1998) Improved survival of biolistically transfected mouse islet allografts expressing CTLA4-Ig or soluble Fas ligand, *Transplantation* **66**, 194–199.

Gao, X., Huang, L. (1991) A novel cationic liposome reagent for efficient transfection of mammalian cells, *Biochem. Biophys. Res. Commun.* **179**, 280–285.

Gao, X., Huang, L. (1993) Cytoplasmic expression of a reporter gene by co-delivery of T7 RNA polymerase and T7 promoter sequence with cationic liposomes, *Nucleic Acids Res.* **21**, 2867–2872.

Gao, X., Huang, L. (1996) Potentiation of cationic liposome-mediated gene delivery by polycations, *Biochemistry* **35**, 1027–1036.

Gao, X., Jaffurs, D., Robbins, P.D., Huang, L. (1994) A sustained, cytoplasmic transgene expression system delivered by cationic liposomes, *Biochem. Biophys. Res. Commun.* **200**, 1201–1206.

Gershon, H., Ghirlando, R., Guttman, S.B., Minsky, A. (1993) Mode of formation and structural features of DNA-cationic liposome complexes used for transfection, *Biochemistry* **32**, 7143–7151.

Giladi, E., Raz, E., Karmeli, F., Okon, E., Rachmilewitz, D. (1995) Transforming growth factor-beta gene therapy ameliorates experimental colitis in rats, *Eur. J. Gastroenterol. Hepatol.* **7**, 341–347.

Gill, D.R., Southern, K.W., Mofford, K.A., Seddon, T., Huang, L., Sorgi, F., Thomson, A., MacVinish, L.J., Ratcliff, R., Bilton, D., Lane, D.J., Littlewood, J.M., Webb, A.K., Middleton, P.G., Colledge, W.H., Cuthbert, A.W., Evans, M.J., Higgins, C.F., Hyde, S.C. (1997) A placebo-controlled study of liposome-mediated gene transfer to the nasal epithelium of patients with cystic fibrosis, *Gene Ther.* **4**, 199–209.

Girard, F., Strausfeld, U., Fernandez, A., Lamb, N.J.C. (1991) Cyclin A is required for the onset of DNA replication in mammalian fibroblasts, *Cell* **67**, 1169–1179.

Godbey, W., Wu, K., Mikos, A. (1999) Size matters: molecular weight affects the efficiency of poly(ethylenimine) as a gene delivery vehicle, *J. Biomed. Mater. Res.* **45**, 268–275.

Gustafsson, J., Arvindson, G., Karlsson, G., Almgren, M.B. (1995) Complexes between cationic liposomes and DNA visualized by cryo-TEM, *Biochim. Biophys. Acta* **1235**, 305–312.

Hara, T., Tan, Y., Huang, L. (1997) In vivo gene delivery to the liver using reconstituted chylomicron remnants as a novel nonviral vector, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 14547–14552.

He, Y., Huang, L. (1997) Growth inhibition of human papillomavirus 16 DNA-positive mouse tumor by antisense RNA transcribed from U6 promoter, *Cancer Res.* **57**, 3993–3999.

Hefeneider, S.H., Cornell, K.A., Brown, L.E., Bakke, A.C., McCoy, S.L., Bennett, R.M. (1992) Nucleosomes and DNA bind to specific cell-surface molecules on murine cells and induce cytokine production, *Clin. Immunol. Immunopathol.* **63**, 245–251.

Heller, L., Jaroszeski, M.J., Coppola, D., Pottinger, C., Gilbert, R., Heller, R. (2000) Electrically mediated plasmid DNA delivery to hepatocellular carcinomas in vivo, *Gene Ther.* **7**, 826–829.

Heller, R., Jaroszeski, M., Atkin, A., Moradpour, D., Gilbert, R., Wands, J., Nicolau, C. (1996) In vivo gene electroinjection and expression in rat liver, *FEBS Lett.* **389**, 225–228.

Herweijer, H., Zhang, G., Subbotin, V.M., Budker, V., Williams, P., Wolff, J.A. (2001) Time course of gene expression after plasmid DNA gene transfer to the liver, *J. Gene Med.* **3**, 280–291.

Hickman, M.A., Malone, R.W., Lehmann-Bruinsma, K., Sih, T.R., Knoell, D., Szoka, F.C., Walzem, R., Carlson, D.M., Powell, J.S. (1994) Gene expression following direct

injection of DNA into liver, *Hum. Gene Ther.* **5**, 1477–1483.

Hofman, C.R., Dileo, J.P., Li, Z., Li, S., Huang, L. (2001) Efficient in vivo gene transfer by PCR amplified fragment with reduced inflammatory activity, *Gene Ther.* **8**, 71–74.

Hong, K., Zheng, W., Baker, A., Papahadjopoulos, D. (1997) Stabilization of cationic liposome-plasmid DNA complexes by polyamines and poly(ethylene glycol)-phospholipid conjugates for efficient in vivo gene delivery, *FEBS Lett.* **400**, 233–237.

Hsu, D.H., de Waal Malefyt, R., Fiorentino, D.F., Dang, M.N., Vieira, P., de Vries, J., Spits, H., Mosmann, T.R., Moore, K.W. (1990) Expression of interleukin-10 activity by Epstein-Barr virus protein BCRF1, *Science* **250**, 830–832.

Hudde, T., Rayner, S.A., Comer, R.M., Weber, M., Isaacs, J.D., Waldmann, H., Larkin, D.F., George, A.J. (1999) Activated polyamidoamine dendrimers, a non-viral vector for gene transfer to the corneal endothelium, *Gene Ther.* **6**, 939–943.

Huettinger, M., Retzek, H., Eder, M., Goldenberg, H. (1988) Characteristics of chylomicron remnant uptake into rat liver, *Clin. Biochem.* **21**, 87–92.

Hussain, M.M., Maxfield, F.R., Mas-Oliva, J., Tabas, I., Ji, Z.S., Innerarity, T.L., Mahley, R.W. (1991) Clearance of chylomicron remnants by the low density lipoprotein receptor-related protein/alpha 2-macroglobulin receptor, *J. Biol. Chem.* **266**, 13936–13940.

Hwang, S.J., Bellocq, N.C., Davis, M.E. (2001) Effects of structure of beta-cyclodextrin-containing polymers on gene delivery, *Bioconjugate Chem.* **12**, 280–290.

Hwang, S.H., Hayashi, K., Takayama, K., Maitani, Y. (2001) Liver-targeted gene transfer into a human hepatoblastoma cell line and in vivo by sterylglucoside-containing cationic liposomes, *Gene Ther.* **8**, 1276–1280.

Jakob, T., Walker, P.S., Krieg, A.M., Udey, M.C., Vogel, J.C. (1998) Activation of cutaneous dendritic cells by CpG-containing oligodeoxynucleotides: a role for dendritic cells in the augmentation of Th1 responses by immunostimulatory DNA, *J. Immunol.* **161**, 3042–3049.

Jarnagin, W.R., Debs, R.J., Wang, S.S., Bisell, D.M. (1992) Cationic lipid-mediated transfection of liver cells in primary culture, *Nucleic Acids Res.* **20**, 4205–4211.

Kabakov, A.E., Saenko, V.A., Poverenny, A.M. (1991) LDL-mediated interaction of DNA and DNA-anti-DNA immune complexes with cell surface, *Clin. Exp. Immunol.* **83**, 359–363.

Kalderon, D., Roberts, B.L., Richardson, W.D., Smith, A.E. (1994) A short amino acid sequence able to specify nuclear location, *Cell* **39**, 499–509.

Kim, T.W., Chung, H., Kwon, I.C., Sung, H.C., Jeong, S.Y. (2000) In vivo gene transfer to the mouse nasal cavity mucosa using a stable cationic lipid emulsion, *Mol. Cells* **10**, 142–147.

Kircheis, R., Schüller, S., Brunner, S., Ogris, M., Heider, K.H., Zauner, W., Wagner, E. (1999) Polycation-based DNA complexes for tumor-targeted gene delivery in vivo, *J. Gene Med.* **1**, 111–120.

Kircheis, R., Kichler, A., Wallner, G., kursa, M., Ogris, M., Felzmann, T., Buchberger, M., Wagner, E. (1997) Coupling of cell-binding ligands to polyethylenimine for targeted gene delivery, *Gene Ther.* **4**, 409–418.

Kitsis, R.N., Buttrick, P.M., McNally, E.M., Kaplan, M.L., Leinwand, L.A. (1991) Hormonal modulation of a gene injected into rat heart in vivo, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4138–4142.

Klinman, D.M., Yamshchikov, G., Ishigatsubo, Y. (1997) Contribution of CpG motifs to the immunogenicity of DNA vaccines, *J. Immunol.* **158**, 3635–3639.

Krieg, A.M., Yi, A.K., Matson, S., Waldschmidt, T.J., Bishop, G.A., Teasdale, R., Koretzky, G.A., Klinman, D.M. (1995) CpG motifs in bacterial DNA trigger direct B-cell activation, *Nature* **374**, 546–549.

Kukowska-Latallo, J.F., Bielinska, A.U., Johnson, J., Spindler, R., Tomalia, D.A., Baker, J.R., Jr. (1996) Efficient transfer of genetic material into mammalian cells using Starburst polyamidoamine dendrimers, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4897–4902.

Kukowska-Latallo, J.F., Raczka, E., Quintana, A., Chen, C., Rymaszewski, M., Baker, J.R., Jr. (2000) Intravascular and endobronchial DNA delivery to murine lung tissue using a novel, nonviral vector, *Hum. Gene Ther.* **11**, 1385–1395.

Kuriyama, S., Mitoro, A., Tsujinoue, H., Nakatani, T., Yoshiji, H., Tsujimoto, T., Yamazaki, M., Fukui, H. (2000) Particle-mediated gene transfer into murine livers using a newly developed gene gun, *Gene Ther.* **7**, 1132–1136.

Labat-Moleur, F., Steffan, A.M., Brisson, C., Perron, H., Feugeas, O., Furstenberger, P.,

Oberling, F., Brambilla, E., Behr, J.P. (1996) An electron microscopy study into the mechanism of gene transfer with lipopolyamines, *Gene Ther.* **3**, 1010–1017.

Larsen, D.L., Dybdahl-Sissoko, N., McGregor, M.W., Drape, R., Neumann, V., Swain, W.F., Lunn, D.P., Olsen, C.W. (1998) Coadministration of DNA encoding interleukin-6 and hemagglutinin confers protection from influenza virus challenge in mice, *J. Virol.* **72**, 1704–17048.

Lee, R.J., Huang, L. (1996) Folate-targeted, anionic liposome-entrapped polylysine-condensed DNA for tumor cell-specific gene transfer, *J. Biol. Chem.* **271**, 8481–8487.

Lee, P.Y., Chesnoy, S., Huang, L. (2004) Electroporatic delivery of TGF-beta1 gene works synergistically with electric therapy to enhance diabetic wound healing in db/db mice, *J. Invest. Dermatol.* **123**, 791–798.

Lee, K.D., Hong, K., Papahadjopoulos, D. (1992) Recognition of liposomes by cells: in vitro binding and endocytosis mediated by specific lipid headgroups and surface charge density, *Biochim. Biophys. Acta* **1103**, 185–197.

Lee, E.R., Marshall, J., Siegel, C.S., Jiang, C., Yew, N.S., Nichols, M.R., Nietupski, J.B., Ziegler, R.J., Lane, M.B., Wang, k.X., Scheule, R.K., Harris, D.J., Smith, A.E., Cheng, S.H. (1996) Detailed analysis of structures and formulations of cationic lipids for efficient gene transfer to the lung, *Hum. Gene Ther.* **7**, 1701–1717.

Leong, K.W., Mao, H.-Q., Truong-Le, V.L., Roy, K., Walsh, S.M., August, J.T. (1998) DNA-polycation nanospheres as non-viral gene delivery vehicles, *J. Control Release* **53**, 183–193.

Leventis, R., Silvius, J.R. (1990) Interactions of mammalian cells with lipid dispersions containing novel metabolizable cationic amphiphiles, *Biochim. Biophys. Acta* **1023**, 124–132.

Li, S., Huang, L. (1997) In vivo gene transfer via intravenous administration of cationic lipid-protamine-DNA (LPD) complexes, *Gene Ther.* **4**, 891–900.

Li, L.H., McCarthy, P., Hui, S.W. (2001) High-efficiency electrotransfection of human primary hematopoietic stem cells, *FASEB J.* **15**, 586–588.

Li, S., Tan, Y., Viroonchatapan, E., Pitt, B.R., Huang, L. (2000) Targeted gene delivery to pulmonary endothelium by anti-PECAM antibody, *Am. J. Physiol. Lung Cell Mol. Physiol.* **278**, L504–L511.

Li, S., Tseng, W.C., Stolz, D.B., Wu, S.P., Watkins, S.C., Huang, L. (1999) Dynamic changes in the characteristics of cationic lipidic vectors after exposure to mouse serum: implications for intravenous lipofection, *Gene Ther.* **6**, 585–594.

Li, S., Wu, S.P., Whitmore, M., Loeffert, E.J., Wang, L., Watkins, S.C., Pitt, B.R., Huang, L. (1999) Effect of immune response on gene transfer to the lung via systemic administration of cationic lipidic vectors, *Am. J. Physiol.* **276**, L796–L804.

Lin, H., Parmacek, M.S., Morle, G., Bolling, S., Leiden, J.M. (1990) Expression of recombinant genes in myocardium in vivo after direct injection of DNA, *Circulation* **82**, 2217–2221.

Liu, F., Huang, L. (2001) Improving plasmid DNA-mediated liver gene transfer by prolonging its retention in the hepatic vasculature, *J Gene Med.* **3**, 569–576.

Liu, F., Huang, L. (2002) A syringe electrode device for simultaneous injection of DNA and electrotransfer, *Mol. Ther.* **5**, 323–328.

Liu, F., Song, Y., Liu, D. (1999) Hydrodynamics-based transfection in animals by systemic administration of plasmid DNA, *Gene Ther.* **6**, 1258–1266.

Liu, F., Nishikawa, M., Clemens, P.R., Huang, L. (2001) Transfer of full-length DMD to the diaphragm muscle of Dmd(mdx/mdx) mice through systemic administration of plasmid DNA, *Mol. Ther.* **4**, 45–51.

Liu, F., Qi, H., Huang, L., Liu, D. (1997) Factors controlling the efficiency of cationic lipid-mediated transfection in vivo via intravenous administration, *Gene Ther.* **4**, 517–523.

Liu, F., Yang, J., Huang, L., Liu, D. (1996) Effect of non-ionic surfactants on the formation of DNA/emulsion complexes and emulsion-mediated gene transfer, *Pharm. Res.* **13**, 1642–1646.

Lunn, D.P., Soboll, G., Schram, B.R., Quass, J., McGregor, M.W., Drape, R.J., Macklin, M.D., McCabe, D.E., Swain, W.F., Olsen, C.W. (1999) Antibody responses to DNA vaccination of horses using the influenza virus hemagglutinin gene, *Vaccine* **17**, 2245–2258.

Macklin, M.D., McCabe, D., McGregor, M.W., Neumann, V., Meyer, T., Callan, R., Hinshaw, V.S., Swain, W.F. (1998) Immunization of pigs with a particle-mediated DNA vaccine

to influenza A virus protects against challenge with homologous virus, *J. Virol.* **72**, 1491–1496.

Mahvi, D.M., Burkholder, J.K., Turner, J., Culp, J., Malter, J.S., Sondel, P.M., Yang, N.S. (1996) Particle-mediated gene transfer of granulocyte-macrophage colony-stimulating factor cDNA to tumor cells: implications for a clinically relevant tumor vaccine, *Hum. Gene Ther.* **7**, 1535–1543.

Maruyama-Tabata, H., Harada, Y., Matsumura, T., Satoh, E., Cui, F., Iwai, M., Kita, M., Hibi, S., Imanishi, J., Sawada, T., Mazda, O. (2000) Effective suicide gene therapy in vivo by EBV-based plasmid vector coupled with polyamidoamine dendrimer, *Gene Ther.* **7**, 53–60.

Matsubara, H., Maeda, T., Gunji, Y., Koide, Y., Asano, T., Ochiai, T., Sakiyama, S., Tagawa, M. (2001) Combinatory antitumor effects of electroporation-mediated chemotherapy and wild-type p53 gene transfer to human esophageal cancer cells, *Int. J. Oncol.* **18**(:) 825–829.

Matsui, H., Johnson, L.G., Randell, S.H., Boucher, R.C. (1997) Loss of binding and entry of liposome-DNA complexes decreases transfection efficiency in differentiated airway epithelial cells, *J. Biol. Chem.* **272**, 1117–1126.

McDonald, P., Hicks, M.N., Cobbe, S.M., Prentice, H. (1995) Gene transfer in models of myocardial ischemia, *Ann. N.Y. Acad. Sci.* **752**, 455–459.

McNeil, P.L., Khakee, R. (1992) Disruptions of muscle fiber plasma membranes. Role in exercise-induced damage, *Am. J. Pathol.* **140**, 1097–1109.

Meyer, O., Schughart, K., Pavirani, A., Kolbe, H.V. (2000) Multiple systemic expression of human interferon-beta in mice can be achieved upon repeated administration of optimized pcTG90-lipoplex, *Gene Ther.* **7**, 1606–1611.

Miao, C.H., Thompson, A.R., Loeb, K., Ye, X. (2001) Long-term and therapeutic-level hepatic gene expression of human factor IX after naked plasmid transfer in vivo, *Mol. Ther.* **3**, 947–957.

Middleton, P., Caplen, G.N.J., Gao, X., Huang, L., Gaya, H., Geddes, D.M., Alton, E.W. (1994) Nasal application of the cationic liposome DC-Chol:DOPE does not alter ion transport, lung function or bacterial growth, *Eur. Respir. J.* **7**, 442–445.

Mislick, K.A., Baldeschweiler, J.D. (1996) Evidence for the role of proteoglycans in cation-mediated gene transfer, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12349–12354.

Mohr, L., Yoon, S.K., Eastman, S.J., Chu, Q., Scheule, R.K., Scaglioni, P.P., Geissler, M., Heintges, T., Blum, H.E., Wands, J.R. (2001) Cationic liposome-mediated gene delivery to the liver and to hepatocellular carcinomas in mice, *Hum. Gene Ther.* **12**, 799–809.

Momose, T., Tonegawa, A., Takeuchi, J., Ogawa, H., Umesono, K., Yasuda, K. (1999) Efficient targeting of gene expression in chick embryos by microelectroporation, *Dev. Growth Differ.* **41**, 335–344.

Monroe, W.T., McQuain, M.M., Chang, M.S., Alexander, J.S., Haselton, F.R. (1999) Targeting expression with light using caged DNA, *J. Biol. Chem.* **274**, 20895–20900.

Moore, K.W., Vieira, P., Fiorentino, D.F., Trounstine, M.L., Khan, T.A., Mosmann, T.R. (1990) Homology of cytokine synthesis inhibitory factor (IL-10) to the Epstein-Barr virus gene BCRFI, *Science* **248**, 1230–1234.

Moradpour, D., Schauer, J.I., Zurawski, V.R., Wands, J.R., Boutin, R.H. (1996) Efficient gene transfer into mammalian cells with cholesteryl-spermidine, *Biochem. Biophys. Res. Commun.* **221**, 82–88.

Mounkes, L.C., Zhong, W., Palacin, G.C., Heath, T.D., Debs, R.J. (1998) Proteoglycans mediate cationic liposome-DNA complex-based gene delivery in vitro and in vivo, *J. Biol. Chem.* **273**, 26164–26170.

Mukherjee, A., Prasad, T.K., Rao, N.M., Banerjee, R. (2005) Haloperidol associated stealth liposomes: a potent carrier for delivering genes to human breast cancer cells, *J. Biol. Chem.* (Article in Press).

Muramatsu, T., Shibata, O., Ryoki, S., Ohmori, Y., Okumura, J. (1997) Foreign gene expression in the mouse testis by localized in vivo gene transfer, *Biochem. Biophys. Res. Commun.* **233**, 45–49.

Nabel, G.J., Gordon, D., Bishop, D.K., Nickoloff, B.J., Yang, Z., Aruga, A., Cameron, M.J., Nabel, E.G., Chang, A.E. (1996) Immune response in human melanoma after transfer of an allogeneic class I major histocompatibility complex gene with DNA-liposome complexes, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 15388–15393.

Nabel, E.G., Gordon, D., Yang, Z.Y., Xu, L., San, H., Plautz, G.E., Wu, B.Y., Gao, X., Huang, L., Nabel, G.J. (1992) Gene transfer

in vivo with DNA-liposome complexes: lack of autoimmunity and gonadal localization, *Hum. Gene Ther.* **3**, 649–656.

Nakamura, H., Watanabe, Y., Funahashi, J.I. (2000) Misexpression of genes in brain vesicles by in ovo electroporation, *Dev. Growth Differ.* **42**, 199–201.

Nanney, L.B., Paulsen, S., Davidson, M.K., Cardwell, N.L., Whitsitt, J.S., Davidson, J.M. (2000) Boosting epidermal growth factor receptor expression by gene gun transfection stimulates epidermal growth in vivo, *Wound Rep. Reg.* **8**, 117–127.

Nitta, Y., Tashiro, F., Tokui, M., Shimada, A., Takei, I., Tabayashi, K., Miyazaki, J. (1998) Systemic delivery of interleukin 10 by intramuscular injection of expression plasmid DNA prevents autoimmune diabetes in nonobese diabetic mice, *Hum. Gene Ther.* **9**, 1701–1707.

Ogino, H., Yasuda, K. (1998) Induction of lens differentiation by activation of a bZIP transcription factor, L-Maf, *Science* **280**, 115–118.

Ogris, M., Brunner, S., Schuller, S., Kircheis, R., Wagner, E. (1999) PEGylated DNA/transferrin-PEI complexes: reduced interaction with blood components, extended circulation in blood and potential for systemic gene delivery, *Gene Ther.* **6**, 595–605.

Olivares, E.C., Hollis, R.P., Chalberg, T.W., Meuse, L., Kay, M.A., Calos, M.P. (2002)Site-specific genomic integration produces therapeutic Factor IXlevels in mice, *Nat. Biotechnol.* **20**, 1124–1128.

Ortiz-Urda, S., Thyagarajan, B., Keene, D.R., Lin, Q., Fang, M., Calos, M.P., Khavari, P.A. (2002) Stable nonviral genetic correction of inherited human skin disease, *Nat. Med.* **8**, 1166–1170.

Parker, S.E., Khatibi, S., Margalith, M., Anderson, D., Yankauckas, M., Gromkowski, S.H., Latimer, T., Lew, D., Marquet, M., Manthorpe, M., Hobart, P., Hersh, E., Stopeck, A.T., Norman, J. (1996) Plasmid DNA gene therapy: studies with the human interleukin-2 gene in tumor cells in vitro and in the murine B16 melanoma model in vivo, *Cancer Gene Ther.* **3**, 175–185.

Parmacek, M.S., Ip, H.S., Jung, F., Shen, T., Martin, J.F., Vora, A.J., Olson, E.N., Leiden, J.M. (1994) A novel myogenic regulatory circuit controls slow/cardiac troponin C gene transcription in skeletal muscle, *Mol. Cell. Biol.* **14**, 1870–1875.

Piccirillo, C.A., Prud'homme, G.J. (1999) Prevention of experimental allergic encephalomyelitis by intramuscular gene transfer with cytokine-encoding plasmid vectors, *Hum. Gene Ther.* **10**, 1915–1922.

Piccirillo, C.A., Chang, Y., Prud'homme, G.J. (1998) TGF-beta1 somatic gene therapy prevents autoimmune disease in nonobese diabetic mice, *J. Immunol.* **161**, 3950–3956.

Pollard, H., Remy, J.S., Loussouarn, G., Demolombe, S., Behr, J.-P., Escande, D. (1998) Polyethylenimine but not cationic lipids promotes transgene delivery to the nucleus in mammalian cells, *J. Biol. Chem.* **273**, 7507–7511.

Porteous, D.J., Dorin, J.R., McLachlan, G., Davidson-Smith, H., Davidson, H., Stevenson, B.J., Carothers, A.D., Wallace, W.A., Moralee, S., Hoenes, C., Kallmeyer, G., Michaelis, U., Naujoks, K., Ho, L.P., Samways, J.M., Imrie, M., Greening, A.P., Innes, J.A. (1997) Evidence for safety and efficacy of DOTAP cationic liposome mediated CFTR gene transfer to the nasal epithelium of patients with cystic fibrosis, *Gene Ther.* **4**, 210–218.

Prabhakar, B.S., Allaway, G.P., Srinivasappa, J., Notkins, A.L. (1990) Cell surface expression of the 70-kD component of Ku, a DNA-binding nuclear autoantigen, *J. Clin. Invest.* **86**, 1301–1305.

Prentice, H., Kloner, R.A., Li, Y., Newman, L., Kedes, L. (1996) Ischemic/reperfused myocardium can express recombinant protein following direct DNA or retroviral injection, *J. Mol. Cell Cardiol.* **28**, 133–140.

Prentice, H., Bishopric, N.H., Hicks, M.N., Discher, D.J., Wu, X., Wylie, A.A., Webster, K.A. (1997) Regulated expression of a foreign gene targeted to the ischaemic myocardium, *Cardiovasc. Res.* **35**, 567–574.

Prud'homme, G.J. (2000) Gene therapy of autoimmune diseases with vectors encoding regulatory cytokines or inflammatory cytokine inhibitors, *J. Gene Med.* **2**, 222–232.

Qin, L., Ding, Y., Bromberg, J.S. (1996) Gene transfer of transforming growth factor-beta 1 prolongs murine cardiac allograft survival by inhibiting cell-mediated immunity, *Hum. Gene Ther.* **7**, 1981–1988.

Qin, L., Ding, Y., Pahud, D.R., Chang, E., Imperiale, M.J., Bromberg, J.S. (1997) Promoter

attenuation in gene therapy: interferon-gamma and tumor necrosis factor-alpha inhibit transgene expression, *Hum. Gene Ther.* **8**, 2019–2029.

Qin, L., Pahud, D.R., Ding, Y., Bielinska, A.U., Kukowska-Latallo, J.F., Baker, J.R., Jr., Bromberg, J.S. (1998) Efficient transfer of genes into murine cardiac grafts by Starburst polyamidoamine dendrimers, *Hum. Gene Ther.* **9**, 553–560.

Qin, L., Chavin, K.D., Ding, Y., Tahara, H., Favaro, J.P., Woodward, J.E., Suzuki, T., Robbins, P.D., Lotze, M.T., Bromberg, J.S. (1996) Retrovirus-mediated transfer of viral IL-10 gene prolongs murine cardiac allograft survival, *J. Immunol.* **156**, 2316–2323.

Radler, J.O., Koltover, I., Salditt, T., Safinya, C.R. (1997) Structure of DNA-cationic liposome complexes: DNA intercalation in multilamellar membranes in distinct interhelical packing regimes, *Science* **275**, 810–814.

Rakhmilevich, A.L., Turner, J., Ford, M.J., McCabe, D., Sun, W.H., Sondel, P.M., Grota, K., Yang, N.S. (1996) Gene gun-mediated skin transfection with interleukin 12 gene results in regression of established primary and metastatic murine tumors, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6291–6296.

Raz, E., Duddler, J., Lotz, M., Bairs, S.M., Berry, C.C., Eisenberg, R.A., Carson, D. (1995) Modulation of disease activity in murine systemic lupus erythematosus by cytokine gene delivery, *Lupus* **4**, 286–292.

Regnier, V., De Morre, N., Jadoul, A., Preat, V. (1999) Mechanisms of a phosphorothioate oligonucleotide delivery by skin electroporation, *Int. J. Pharm.* **184**, 147–156.

Rensen, P.C., van Dijk, M.C., Havenaar, E.C., Bijsterbosch, M.K., Kruijt, J.K., van Berkel, T.J. (1995) Selective liver targeting of antivirals by recombinant chylomicrons–a new therapeutic approach to hepatitis B, *Nat. Med.* **1**, 221–225.

Ritchie, M.E. (1996) Characterization of human B creatine kinase gene regulation in the heart in vitro and in vivo, *J. Biol. Chem.* **271**, 25485–25491.

Rizzuto, G., Cappelletti, M., Mennuni, C., Wiznerowicz, M., DeMartis, A., Maione, D., Ciliberto, G., La Monica, N., Fattori, E. (2000) Gene electrotransfer results in a high-level transduction of rat skeletal muscle and corrects anemia of renal failure, *Hum. Gene Ther.* **11**, 1891–1900.

Rizzuto, G., Cappelletti, M., Maione, D., Savino, R., Lazzaro, D., Costa, P., Mathiesen, I., Cortese, R., Ciliberto, G., Laufer, R., La Monica, N., Fattori, E. (1999) Efficient and regulated erythropoietin production by naked DNA injection and muscle electroporation, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6417–6422.

Rols, M.P., Delteil, C., Golzio, M., Dumond, P. (1998) In vivo electrically mediated protein and gene transfer in murine melanoma, *Nat. Biotechnol.* **16**, 168–171.

Rols, M.P., Bachaud, J.M., Giraud, P., Chevreau, C., Roche, H., Teissie, J. (2000) Electrochemotherapy of cutaneous metastases in malignant melanoma, *Melanoma Res.* **10**, 468–474.

Ruysschaert, J.M., El Quahabi, A., Willeaume, V., Huez, G., Fuks, R., Vandenbremden, M., Di Stefano, P. (1994) A novel cationic amphiphile for transfection of mammalian cells, *Biochem. Biophys. Res. Commun.* **103**, 1622–1628.

Sakurai, F., Nishioka, T., Saito, H., Baba, T., Okuda, A., Matsumoto, O., Taga, T., Yamashita, F., Takakura, Y., Hashida, M. (2001) Interaction between DNA-cationic liposome complexes and erythrocytes is an important factor in systemic gene transfer via the intravenous route in mice: the role of the neutral helper lipid, *Gene Ther.* **8**, 677–686.

San, H., Yang, Z.Y., Pompili, V.J., Jaffe, M.L., Plautz, G.E., Xu, L., Felgner, J.H., Wheeler, C.J., Felgner, P.L., Gao, X., et al. (1993) Safety and short-term toxicity of a novel cationic lipid formulation for human gene therapy, *Hum. Gene Ther.* **4**, 781–788.

Schwartz, B., Benoist, C., Abdallah, B., Rangara, R., Hassan, A., Scherman, D., Demeneix, B.A. (1996) Gene transfer by naked DNA into adult mouse brain, *Gene Ther.* **3**, 405–411.

Sebestyen, M.G., Ludtke, J.J., Bassik, M.C., Zhang, G., Budker, V., Lukhtanov, E.A., Hagstrom, J.E., Wolff, J.A. (1998) DNA vector chemistry: the covalent attachment of signal peptides to plasmid DNA, *Nat. Biotechnol.* **16**, 80–85.

Shi, N., Pardridge, W.M. (2000) Noninvasive gene targeting to the brain, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7567–7572.

Siess, D.C., Vedder, C.T., Merkens, L.S., Tanaka, T., Freed, A.C., McCoy, S.L., Heinrich, M.C., Deffebach, M.E., Bennett, R.M., Hefeneider, S.H. (2000) A human gene coding for a membrane-associated nucleic

acid-binding protein, *J. Biol. Chem.* **275**, 33655–33662.

Sikes, M.L., O'Malley, B.W., Jr., Finegold, M.J., Ledley, F.D. (1994) In vivo gene transfer into rabbit thyroid follicular cells by direct DNA injection, *Hum. Gene Ther.* **5**, 837–844.

Simovic, D., Isner, J.M., Ropper, A.H., Pieczek, A., Weinberg, D.H. (2001) Improvement in chronic ischemic neuropathy after intramuscular phVEGF165 gene transfer in patients with critical limb ischemia, *Arch. Neurol.* **58**, 761–768.

Smith, J.G., Walzem, R.L., German, J.B. (1993) Liposomes as agents of DNA transfer, *Biochim. Biophys. Acta.* **1154**, 327–340.

Soler, M.N., Bobe, P., Benihoud, K., Lemaire, G., Roos, B.A., Lausson, S. (2000) Gene therapy of rat medullary thyroid cancer by naked nitric oxide synthase II DNA injection, *J. Gene Med.* **2**, 344–352.

Solodin, I., Brown, C.S., Bruno, M.S., Chow, C.-Y., Jang, E.-H., Debs, R.J., Heath, T.D. (1995) A novel series of amphiphilic imidazolinium compounds for in vitro and in vivo gene delivery, *Biochemistry* **34**, 13537–13544.

Song, Y.K., Liu, F., Liu, D. (1998) Enhanced gene expression in mouse lung by prolonging the retention time of intravenously injected plasmid DNA, *Gene Ther.* **5**, 1531–1537.

Song, X.Y., Gu, M., Jin, W.W., Klinman, D.M., Wahl, S.M. (1998) Plasmid DNA encoding transforming growth factor-beta1 suppresses chronic disease in a streptococcal cell wall-induced arthritis model, *J. Clin. Invest.* **101**, 2615–26121.

Sorgi, F.L., Bhattacharya, S., Huang, L. (1997) Protamine sulfate enhances lipid-mediated gene transfer, *Gene Ther.* **4**, 961–968.

Sparwasser, T., Koch, E.S., Vabulas, R.M., Heeg, K., Lipford, G.B., Ellwart, J.W., Wagner, H. (1998) Bacterial DNA and immunostimulatory CpG oligonucleotides trigger maturation and activation of murine dendritic cells, *Eur. J. Immunol.* **28**, 2045–2054.

Stamatos, L., Leventis, R., Zuckermann, M.J., Silvius, J.R. (1988) Interactions of cationic lipid vesicles with negatively charged phospholipid vesicles and biological membranes, *Biochemistry* **27**, 3917–3925.

Stein, C.A., Tonkinson, J.L., Zhang, L.M., Yakubov, L., Gervasoni, J., Taub, R., Rotenberg, S.A. (1993) Dynamics of the internalization of phosphodiester oligodeoxynucleotides in HL60 cells, *Biochemistry* **32**, 4855–4861.

Sternberg, B., Sorgi, F.L., Huang, L. (1994) New structures in complex formation between DNA and cationic liposomes visualized by freeze-fracture electron microscopy, *FEBS Lett.* **356**, 361–366.

Stewart, M.J., Plautz, G.E., Del Buono, L., Yang, Z.Y., Xu, L., Gao, X., Huang, L., Nabel, E.G., Nabel, G.J. (1992) Gene transfer in vivo with DNA-liposome complexes: safety and acute toxicity in mice, *Hum. Gene Ther.* **3**, 267–275.

Straubinger, R.M., Papahadjopoulos, D., Hong, K.L. (1990) Endocytosis and intracellular fate of liposomes using pyranine as a probe, *Biochemistry* **29**, 4929–4939.

Straubinger, R.M., Hong, K., Friend, D.S., Papahadjopoulos, D. (1983) Endocytosis of liposomes and intracellular fate of encapsulated molecules: encounter with a low pH compartment after internalization in coated vesicles, *Cell* **32**, 1069–1079.

Sukharev, S.I., Klenchin, V.A., Serov, S.M., Chernomordik, L.V., Chizmadzhev, Y.A. (1992) Electroporation and electrophoretic DNA transfer into cells. The effect of DNA interaction with electropores, *Biophys. J.* **63**, 1320–1327.

Suzuki, T., Shin, B.C., Fujikura, K., Matsuzaki, T. (1998) Direct gene transfer into rat liver cells by in vivo electroporation, *FEBS Lett.* **425**, 436–440.

Tabata, H., Nakajima, K. (2001) Efficient in utero gene transfer system to the developing mouse brain using electroporation: visualization of neuronal migration in the developing cortex, *Neuroscience* **103**, 865–872.

Takagi, T., Hashiguchi, M., Mahato, R.I., Tokuda, H., Takakura, Y., Hashida, M. (1998) Involvement of specific mechanism in plasmid DNA uptake by mouse peritoneal macrophages, *Biochem. Biophys. Res. Commun.* **245**, 729–733.

Takeuchi, J., Koshiba-Takeuchi, K., Matsumoto, K., Vogel-Hopker, A., Naitoh-Matsuo, M., Ogura, K., Takahashi, N., Yasuda, K., Ogura, T. (1999) Tbx5 and Tbx4 genes determine the wing/leg identity of limb buds, *Nature* **398**, 810–814.

Tan, Y., Li, S., Pitt, B.R., Huang, L. (1999) The inhibitory role of CpG immunostimulatory motifs in cationic lipid vector-mediated transgene expression in vivo, *Hum. Gene Ther.* **10**, 2153–2161.

Tang, M.X., Szoka, F.C., Jr. (1997) The influence of polymer structure on the interactions of cationic polymers with DNA and morphology of the resulting complexes, *Gene Ther.* **4**, 823–832.

Tang, M.X., Redeman, C.T., Szoka, F.C., Jr. (1996) In vitro gene delivery by degraded polyamidoamine dendrimers, *Bioconjugate Chem.* **7**, 703–714.

Templeton, N.S., Lasic, D.D., Frederik, P.M., Strey, H.H., Roberts, D.D., Pavlakis, G.N. (1997) Improved DNA: liposome complexes for increased systemic delivery and gene expression, *Nat. Biotechnol.* **15**, 647–652.

Tousignant, J.D., Gates, A.L., Ingram, L.A., Johnson, C.L., Nietupski, J.B., Cheng, S.H., Eastman, S.J., Scheule, R.K. (2000) Comprehensive analysis of the acute toxicities induced by systemic administration of cationic lipid:plasmid DNA complexes in mice, *Hum. Gene Ther.* **11**, 2493–2513.

Tsujie, M., Isaka, Y., Nakamura, H., Imai, E., Hori, M. (2001) Electroporation-mediated gene transfer that targets glomeruli, *J. Am. Soc. Nephrol.* **12**, 949–954.

Tsurumi, Y., Takeshita, S., Chen, D., Kearney, M., Rossow, S.T., Passeri, J., Horowitz, J.R., Symes, J.F., Isner, J.M. (1996) Direct intramuscular gene transfer of naked DNA encoding vascular endothelial growth factor augments collateral development and tissue perfusion, *Circulation* **94**, 3281–3290.

Turunen, M.P., Hiltunen, M.O., Ruponen, M., Virkamaki, L., Szoka, F.C., Jr., Urtti, A., Yla-Herttuala, S. (1999) Efficient adventitial gene delivery to rabbit carotid artery with cationic polymer-plasmid complexes, *Gene Ther.* **6**, 6–11.

Uyechi, L.S., Gagne, L., Thurston, G., Szoka, F.C., Jr. (2001) Mechanism of lipoplex gene delivery in mouse lung: binding and internalization of fluorescent lipid and DNA components, *Gene Ther.* **8**, 828–836.

Vigneron, J.P., Oudrhiri, N., Fauquet, M., Vergely, L., Bradley, J.C., Basseville, M., Lehn, P., Lehn, J.-M. (1996) Guanidinium-cholesterol cationic lipids: efficient vectors for the transfection of eukaryotic cells, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9682–9686.

Vrancken Peeters, M.J., Perkins, A.L., Kay, M.A. (1996) Method for multiple portal vein infusions in mice: quantitation of adenovirus-mediated hepatic gene transfer, *Biotechniques* **20**, 278–285.

Wagner, E., Zenke, M., Cotten, M., Beug, H., Birnsteil, M.L. (1990) Transferrin-polycation conjugates as carriers for DNA uptake into cells, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3410–3414.

Walker, S., Sofia, M.J., Kokarla, R., Kogan, N.A., Wierichs, L., Longley, C.B., Bruku, K., Axelrod, H.R., Midha, S., Babu, S., Kahne, D. (1996) Cationic facial amphiphiles: a promising class of transfection agents, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1585–1590.

Wang, C.Y., Huang, L. (1987) pH-sensitive immunoliposomes mediate target-cell-specific delivery and controlled expression of a foreign gene in mouse, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7851–7855.

Wang, C., Chao, C., Madeddu, P., Chao, L., Chao, J. (1998) Central delivery of human tissue kallikrein gene reduces blood pressure in hypertensive rats, *Biochem. Biophys. Res. Commun.* **244**, 449–454.

Wang, J., Guo, X., Xu, Y., Barron, L., Szoka, F.C., Jr. (1998) Synthesis and characterization of long chain alkyl acyl carnitine esters. Potentially biodegradable cationic lipids for use in gene delivery, *J. Med. Chem.* **41**, 2207–2215.

Warrant, R.W., Kim, S.H. (1978) alpha-Helix-double helix interaction shown in the structure of a protamine-transfer RNA complex and a nucleoprotamine model, *Nature* **271**, 130–135.

Wells, D.J. (1993) Improved gene transfer by direct plasmid injection associated with regeneration in mouse skeletal muscle, *FEBS Lett.* **332**, 179–182.

Wheeler, C.J., Felgner, P.L., Tsai, Y.J., Marshall, J., Sukhu, L., Doh, S.J., Hartikka, J., Nieptupski, J., Manthorpe, M., Nichols, M., Plewe, M., Liang, X., Norman, J., Smith, A., Cheng, S.J. (1996) A novel cationic lipid greatly enhances plasmid DNA delivery and expression in mouse lung, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 11454–11459.

Whitmore, M., Li, S., Huang, L. (1999) LPD lipopolyplex initiates a potent cytokine response and inhibits tumor growth, *Gene Ther.* **6**, 1867–1875.

Wilson, R.W., Bloomfield, V.A. (1979) Counterion-induced condensation of deoxyribonucleic acid. a light-scattering study, *Biochemistry* **18**, 2192–2196.

Wolff, J.A., Ludtke, J.J., Acsadi, G., Williams, P., Jani, A. (1992) Long-term persistence of

plasmid DNA and foreign gene expression in mouse muscle, *Hum. Mol. Genet.* **1**, 363–369.

Wolff, J.A., Williams, P., Acsadi, G., Jiao, S., Jani, A., Chong, W. (1991) Conditions affecting direct gene transfer into rodent muscle in vivo, *Biotechniques* **11**, 474–485.

Wolff, J.A., Malone, R.W., Williams, P., Chong, W., Acsadi, G., Jani, A., Felgner, P.L. (1990) Direct gene transfer into mouse muscle in vivo, *Science* **247**, 1465–1468.

Wu, G.Y., Wu, C.H. (1988) Receptor-mediated gene delivery and expression in vivo, *J. Biol. Chem.* **263**, 14621–14624.

Xie, K., Huang, S., Dong, Z., Juang, S.H., Wang, Y., Fidler, I.J. (1997) Destruction of bystander cells by tumor cells transfected with inducible nitric oxide (NO) synthase gene, *J. Natl. cancer Inst.* **89**, 421–427.

Xie, K., Huang, S., Dong, Z., Juang, S.H., Gutman, M., Xie, Q.W., Nathan, C., Fidler, I.J. (1995) Transfection with the inducible nitric oxide synthase gene suppresses tumorigenicity and abrogates metastasis by K-1735 murine melanoma cells, *J. Exp. Med.* **181**, 1333–1343.

Xu, F., Szoka, F.C. Jr (1996) Mechanism of DNA release from cationic liposome/DNA complexes used in cell transfection, *Biochemistry* **35**, 5616–5623.

Yamamoto, S., Kuramoto, E., Shimada, S., Tokunaga, T. (1988) In vitro augmentation of natural killer cell activity and production of interferon-alpha/beta and -gamma with deoxyribonucleic acid fraction from Mycobacterium bovis BCG, *Jpn. J. Cancer Res.* **79**, 866–873.

Yanai, K., Hirota, K., Taniguchi-Yanai, K., Shigematsu, Y., Shimamoto, Y., Saito, T., Chowdhury, S., Takiguchi, M., Arakawa, M., Nibu, Y., Sugiyama, F., Yagami, K., Fukamizu, A. (1999) Regulated expression of human angiotensinogen gene by hepatocyte nuclear factor 4 and chicken ovalbumin upstream promoter-transcription factor, *J. Biol. Chem.* **274**, 34605–34612.

Yew, N.S., Zhao, H., Wu, I.H., Song, A., Tousignant, J.D., Przybylska, M., Cheng, S.H. (2000) Reduced inflammatory response to plasmid DNA vectors by elimination and inhibition of immunostimulatory CpG motifs, *Mol. Ther.* **1**, 255–262.

Yew, N.S., Wang, K.X., Przybylska, M., Bagley, R.G., Stedman, M., Marshall, J., Scheule, R.K., Cheng, S.H. (1999) Contribution of plasmid DNA to inflammation in the lung after administration of cationic lipid:pDNA complexes, *Hum. Gene Ther.* **10**, 223–234.

Yi, A.K., Chace, J.H., Cowdery, J.S., Krieg, A.M. (1996) IFN-gamma promotes IL-6 and IgM secretion in response to CpG motifs in bacterial DNA and oligodeoxynucleotides, *J. Immunol.* **156**, 558–564.

Yi, S.W., Yune, T.Y., Kim, T.W., Chung, H., Choi, Y.W., Kwon, I.C., Lee, E.B., Jeong, S.Y. (2000) A cationic lipid emulsion/DNA complex as a physically stable and serum-resistant gene delivery system, *Pharm. Res.* **17**, 314–320.

Yoo, J.J., Soker, S., Lin, L.F., Mehegan, K., Guthrie, P.D., Atala, A. (1999) Direct in vivo gene transfer to urological organs, *J. Urol.* **162**, 1115–1118.

Yoshida, S., Kashiwamura, S.I., Hosoya, Y., Luo, E., Matsuoka, H., Ishii, A., Fujimura, A., Kobayashi, E. (2000) Direct immunization of malaria DNA vaccine into the liver by gene gun protects against lethal challenge of Plasmodium berghei sporozoite, *Biochem. Biophys. Res. Commun.* **271**, 107–115.

Yoshimura, T., Shono, M., Imai, K., Hong, K. (1995) Kinetic analysis of endocytosis and intracellular fate of liposomes in single macrophages, *J. Biochem.* **117**, 34–41.

Yoshimura, K., Rosenfeld, M.A., Nakamura, H., Scherer, E.M., Pavirani, A., Lecocq, J.P., Crystal, R.G. (1992) Expression of the human cystic fibrosis transmembrane conductance regulator gene in the mouse lung after in vivo intratracheal plasmid-mediated gene transfer, *Nucleic Acids Res.* **20**, 3233–3240.

Yu, D., Matin, A., Xia, W., Sorgi, F., Huang, L., Hung, M.C. (1995) Liposome-mediated in vivo E1A gene transfer suppressed dissemination of ovarian cancer cells that overexpress HER-2/neu, *Oncogene* **11**, 1383–1388.

Zabner, J., Fasbender, A.J., Moninger, T., Poellinger, K.A., Welsh, M.J. (1995) Cellular and molecular barriers to gene transfer by a cationic lipid, *J. Biol. Chem.* **270**, 18997–19007.

Zabner, J., Cheng, S.H., Meeker, D., Launspach, J., Balfour, R., Perricone, M.A., Morris, J.E., Marshall, J., Fasbender, A., Smith, A.E., Welsh, M.J. (1997) Comparison of DNA-lipid complexes and DNA alone for gene transfer to cystic fibrosis airway epithelia in vivo, *J. Clin. Invest.* **100**, 1529–1537.

Zelphati, O., Szoka, F.C., Jr. (1996) Intracellular distribution and mechanism of delivery of oligonucleotides mediated by cationic lipids, *Pharm. Res.* **13**, 1367–1372.

Zhang, G., Budker, V., Wolff, J.A. (1999) High levels of foreign gene expression in hepatocytes after tail vein injections of naked plasmid DNA, *Hum. Gene Ther.* **10**, 1735–1737.

Zhang, G., Budker, V., Williams, P., Subbotin, V., Wolff, J.A. (2001) Efficient expression of naked DNA delivered intraarterially to limb muscles of nonhuman primates, *Hum. Gene Ther.* **12**, 427–438.

Zhang, Y., Jiang, Q., Dudus, L., Yankaskas, J.R., Engelhardt, J.F. (1998) Vector-specific complementation profiles of two independent primary defects in cystic fibrosis airways, *Hum. Gene Ther.* **9**, 635–648.

Zheng, Q.A., Chang, D.C. (1991) High-efficiency gene transfection by in situ electroporation of cultured cells, *Biochim. Biophys. Acta* **1088**, 104–110.

Zhou, X., Huang, L. (1994) DNA transfection mediated by cationic liposomes containing lipopolylysine: characterization and mechanism of action, *Biochim. Biophys. Acta* **1189**, 195–203.

Zhu, N., Liggitt, D., Liu, Y., Debs, R. (1993) Systemic gene expression after intravenous DNA delivery into adult mice, *Science* **261**, 209–211.

Ziady, A., Ferkol, T., Dawson, D., Perlmutter, D., Davis, P. (1999) Chain length of the polylysine in receptor-targeted gene transfer complexes affects duration of reporter gene expression both in vitro and in vivo, *J. Biol. Chem.* **274**, 4908–4916.

Zuidam, N.J., Barenholz, Y. (1998) Electrostatic and structural properties of complexes involving plasmid DNA and cationic lipids commonly used for gene delivery, *Biochim. Biophys. Acta* **1368**, 115–128.

Vector Targeting in Gene Therapy

Yosuke Kawakami and David T. Curiel
University of Alabama at Birmingham, Birmingham

1 Introduction 333

2 Transcriptional Targeting 334
2.1 Gastrointestinal Cancers 335
2.2 Hepatocellular Carcinoma 336
2.3 Pancreas/Pancreatic Carcinoma 336
2.4 Lung Cancers/Mesothelioma 337
2.5 Breast Cancer 338
2.6 Urological Cancers 339
2.7 Gynecological Cancers 340
2.8 Brain Tumors 340
2.9 Thyroid Tumors 341
2.10 Malignant Melanoma 341
2.11 Other Cancers 342
2.12 Normal Tissue/Organs 342
2.13 Transcriptional Targeting of Replication-competent Ad 342
2.14 Summary 343

3 Transductional Targeting 345
3.1 Genetic Capsid Modification 346
3.1.1 Chimeric Ad Vector with Different Serotype or Species Fiber 346
3.1.2 Ad Vector with Modified Fiber Proteins 347
3.2 Adaptor Molecules 347
3.3 Genetic Capsid Modification Combined with Adaptor Molecules 348
3.4 Genetic Capsid Modification with Incorporation of Larger-sized Targeting Molecules 350

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

4 Transductional and Transcriptional Dual Targeting 350
4.1 Combination of Targeting with Adaptor Molecule and Transcriptional Targeting 351
4.2 Genetic Capsid Modification Coupled with Transcriptional Targeting 351

5 Conclusions 352

Acknowledgments 353

Bibliography 353
Books and Reviews 353
Primary Literature 353

Keywords

Fiber (of Viruses)
Extracellular structure protruding from each vertex of the virus capsid, which plays an important role in virus infection by binding to its primary receptor.

Promoter
A DNA sequence recognized by RNA polymerase, which initiates transcription and contributes to specific transgene expression by regulation of viral DNA transcription process.

Transcription
The process of copying DNA to RNA by RNA polymerase, in which protein expression level can be controlled for specificity.

Transcriptional Targeting
To achieve selective transgene expression by placing the therapeutic gene under the control of specific promoters that are activated in target cells but not in untarget cells.

Transductional Targeting
To achieve selective transgene expression by redirecting binding of the fiber capsid protein away from a native receptor to an alternative cell-selective receptor.

Tropism
Movement in an organism in response to some environmental stimulus, which is utilized as affinity of adenovirus vectors to the cells representing infectivity in vector targeting.

In gene therapy, current vector systems demonstrate insufficient expression and lack of specificity, resulting in poor curative effects in target cells and/or high toxicity to nontarget cells. Thus, vector optimization is critical for the derivation of efficient clinical gene therapeutics. Recently, adenoviral vectors have been targeted at the level of transduction and transcription to overcome these issues. Using genetic capsid modification techniques, adapter molecules with tissue-specific antibodies and tissue-specific promoters (TSPs), these vectors can be efficiently targeted. These strategies are applicable for many types of target cells and are compatible with a variety of targeting molecules. These combinations are aimed at improving therapeutic index of diseases such as cancers, as well as nonmalignant diseases such as pulmonary hypertension.

1 Introduction

Genetic-based therapies, such as gene therapy, represent a novel approach under investigation for the diagnosis and treatment of a wide variety of diseases. As with most conventional therapies, the balance between target cell and nontarget cell toxicity determines the therapeutic index of gene therapy. All gene therapy strategies are fundamentally based on the expression of transgenes and require efficient delivery of the therapeutic gene. However, correlative laboratory studies have demonstrated the limited ability of current generation vector systems to efficiently transduce targeted cells. In addition, limitations in vector specificity can lead to transduction of normal cells and untoward toxicity even in the setting of compartmental dosing. Thus, vector optimization is critical for the derivation of efficient clinical gene therapeutics. In this regard, adenovirus (Ad) serotype 5 (Ad5) is the predominant adenovirus vector for gene therapy, since it can transduce both dividing and nondividing cells and infect a wide variety of cell types. However, an increasing number of tumor types have been reported to be relatively refractory to Ad5 infection due to the paucity of the native Ad5 receptor, coxsackievirus-adenovirus receptor (CAR). On the other hand, its wide tropism, which is mediated by Ad5 fiber protein and CAR, allows infection of many types of cells expressing CAR in normal human tissues such as hepatocytes, precluding *in vivo* targeting and provoking immune responses that can prevent repeated administration and even shorten the duration of therapeutic gene expression.

In the last two decades, Ad5 has been well characterized in its life cycle, capsid proteins function and, more importantly, the biology of infection in cells. Thus, there is much data on which to base Ad vector targeting strategies. Briefly, three distinct sequential steps are required for Ad infection and transgene expression: (1) binding of the Ad to specific receptors on the surface of the target cell; (2) internalization of the virus; and (3) transfer of the viral genome to the nucleus in which the transgene is expressed through transcription and translation (Fig. 1). On the basis of these steps, three opportunities for targeting intervention are implied: (1) Ad delivery to targeted organs/sites; (2) Ad binding

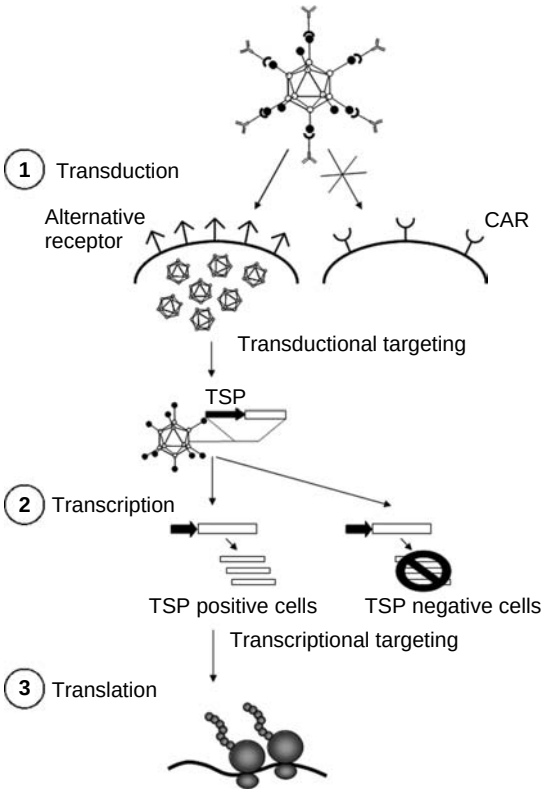

Fig. 1 Strategies for retargeting of Ad vectors.

and entry in targeted cells through primary and/or secondary Ad receptor; (3) protein expression of the transgene in the targeted cells through transcription and translation. This review outlines a series of Ad vectors that have addressed the issue of the vector targeting of gene therapy.

2 Transcriptional Targeting

To date, vectors employed for cancer gene therapy have included constitutively active promoters such as the cytomegalovirus (CMV) promoter, simian virus 40 (SV40) promoter and cytomegalovirus immediate-early enhancer, chicken β-actin promoter and the rabbit β-globin polyadenylation signal (CAG) promoter. The CMV promoter is a strongly positive regulator but lacks expression specificity. Owing to this and other reasons, mammalian cellular promoters are being explored for cancer gene therapy; the goal is specific and it is the persistent expression of therapeutic genes in tumors or target tissues.

One goal of targeting is to minimize the ectopic transgene expression in order to prevent toxicity when the therapeutic gene is cytocidal or an enzyme that activates a prodrug. Adverse effects due to gene expression in nontarget cells, most abundantly liver toxicity and toxicity to the bone marrow, have been reported in preclinical studies. Furthermore, ectopic gene expression in immune cells may be responsible for an immune response to the

transgene, thus limiting therapeutic efficacy. Target cell–specific gene expression may be accomplished through restriction of gene expression by using tissue-specific promoters (TSPs). To achieve this, efficient gene therapy regimens require transgene expression in the tumor, which has been denoted as the "on" status, and absence of expression in the relevant normal tissues, particularly in liver, which has been denoted as the "off" status.

Most TSPs reflect the natural property of the normal cells in target tissue/organs. Pathological changes in target tissues/organs, such as degeneration or malignant transformation, can alter expression of TSPs (i.e. activation of TSPs may be caused by these pathological changes). Thus, utilizing TSPs can enable a successful transcriptional targeting by maintaining higher transductional activity and specificity. To identify the best TSP for targeting cancer, one approach is to use "organ specific" promoters. Some of the targeted organs such as thyroid, breast, and pancreas secrete organ-specific proteins, such as thyroglobulin (TG), lactalbumin, and insulin. Promoters of these proteins are candidates of TSP and promoter activities can be predicted via monitoring these proteins. On the other hand, ectopic expression of these TSPs in normal cells in the organ must also be considered, except in particular cancers in which the function of the organ does not require preservation, such as uterus and prostate. However, TSPs often become overexpressed during transformation and may have a high therapeutic index without damaging normal cells/tissues. Another approach is to use "cancer specific" promoters, the genes of which are linked to malignant transformation and are expressed in a variety of cancers. In this case, promoter activity should also be analyzed in normal cells, such as stem cells and germ cells.

2.1 Gastrointestinal Cancers

Advanced gastrointestinal cancers, especially those associated with distant metastatic disease, are excellent candidates for gene therapy. A number of colon cancer targeting strategies have been reported. Carcinoembryonic antigen (CEA) promoter is the most frequently used TSP for colon cancer targeting. CEA is an oncofetal protein and is often detected at high levels in the serum of patients with colon cancers. Cytosine deaminase (CD), herpes simplex virus-thymidine kinase (HSV-tk), and diphtheria toxin are common therapeutic genes, and suicide gene therapy strategies using Ad vector constructed with *HSV-tk* or *CD* gene driven by the CEA promoter have been reported. One approach utilized the Cre/*loxP* recombination system, consisting of a Cre-producing Ad driven by the CEA promoter (Ad.CEA-Cre) and another Ad designed for inducible expression of the *HSV-TK* gene by Cre (Ad.lox-TK), to overcome the low activity of CEA promoter.

Human telomerase reverse transcriptase (hTERT) promoter, which is the catalytic subunit of the telomerase, is another promoter widely used for colon cancer and other cancers. It is expressed in most cancer cells but is quiescent in normal cells. Ad vectors with hTERT promoter driving expression of *CD* or *Bax* gene have been currently used in colon cancer models. Cyclooxygenase-2 (cox-2), an inducible isoform of the cyclooxygenase family, is closely linked to carcinogenesis and progression of colon cancers and about half of benign colon polyps. An Ad vector constructed with a cox-2 driven

HSV-tk gene demonstrated specific tumor cell killing with reduced liver toxicity using *in vitro* and *in vivo* models. Other colon cancer–specific promoters such as, the CTP1 promoter (beta-catenin-dependent promoter), the EGP-2 promoter (human epithelial glycoprotein-2 promoter), and the p21/WAF-1 promoter can also drive CD, HSV-tk or iNOS; these are useful promoters for carcinoma directed treatment modalities. For example, the EGP-2 promoter is epithelium specific, but there is highly abundant expression in the human epithelium, while the p21/WAF-1 promoter is an X-ray inducible promoter.

Gastric cancer is one of the leading causes of cancer-related death, and it remains endemic, especially in Asia, Africa, and parts of Europe. Targeted gene therapy could be one of the candidates for an innovative therapeutic approach in the treatment of advanced gastric cancer. The majority of gastric as well as colon cancers have been reported to show increased expression of cox-2 and CEA. Ad vectors with cox-2 or CEA-driven *HSV-tk* or *CD* genes showed specific cell killing for the treatment of gastric cancer. Nakaya et al. reported that suicide gene therapy with α-fetoprotein (AFP) promoter-driven HSV-tk was useful for the treatment of AFP-producing gastric cancer.

2.2 Hepatocellular Carcinoma

Hepatocellular carcinoma (HCC) is one of the most common malignancies worldwide. Since most conventional therapies have shown no survival benefit, gene therapy may provide an alternative therapeutic approach for the treatment of HCC. Many trials for targeting HCC have been reported and most of these were targeted using the AFP promoter. AFP is a glycoprotein that is similar to albumin and is produced by the fetal liver. The AFP gene is normally expressed in fetal liver and is transcriptionally silent in adult liver but is overexpressed in HCC. The level of AFP elevation in HCC has been shown to correlate with poor tumor differentiation, increased tumor burden, early recurrence following tumor resection, and unfavorable prognosis. A number of strategies using AFP promoter driving the suicide genes, CD or HSV-tk, have been reported, some of which were constructed in combination with the Cre/*loxP* system or with the phosphoglycerate kinase (PKG) promoter to overcome the low transcriptional activity of the AFP promoter. Other TSPs for HCC have not been reported yet in the context of Ad-based gene therapy. However, Onishi et al. demonstrated hTERT transcripts in surgically resected HCC by *in situ* hybridization and suggested that the hTERT promoter is a good candidate as a target for HCC.

2.3 Pancreas/Pancreatic Carcinoma

The pancreas is both an exocrine and an endocrine organ. Amylase and insulin are representative proteins that are secreted from this organ. Therefore, targeting pancreatic carcinoma using these promoters is a rational strategy. The homeodomain transcription factor activating the insulin promoter, the pancreatic duodenal homeobox-1 (PDX-1), is an essential regulator of pancreatic endocrine cell development and adult islet β-cell function. Recently, expression of PDX-1 was found in several human pancreatic cancer cell lines. Wang et al. showed that the rat insulin promoter (RIP), used to drive the

HSV-tk and *CD* genes, resulted in cytotoxicity of PDX-1–positive insulinoma cells. The murine pancreatic amylase promoter was also tested for LacZ reporter gene expression in the pancreas for the treatment of metabolic disease therapy. In approximately one-third of pancreatic cancers, the proto-oncogene, *erbB2*, is overexpressed due to transcriptional upregulation of the *erbB2* gene with or without amplification. Harris et al. constructed an Ad vector with the *CD* gene driven by the erbB2 promoter, which demonstrated pancreatic cell–specific killing. Several other cancer-specific promoters such as the midkine, cox-2, and hTERT promoters, driving either reporter or suicide genes have shown specific transduction and therapeutic effect in pancreatic carcinoma.

2.4 Lung Cancers/Mesothelioma

Lung cancer has the highest mortality rate of all malignancies in developed countries; thus, novel therapeutic approaches such as gene therapy is required in this field. Small-cell lung cancer (SCLC), which comprises about 25% of lung cancers, characteristically secretes a variety of neuropeptides, such as adrenocorticotrophin hormone (ACTH), gastrin-releasing peptide (GRP), gastrin, neuron-specific enolase (NSE), cholecystokinin, and arginine vasopressin promoter (AVP). Many of these peptides can act as autocrine growth factors to the tumor. Yoshizawa et al. constructed vectors with HSV-tk driven by the NSE or GRP promoter and demonstrated a therapeutic effect in SCLC in an *in vitro* or *in vivo* model. Coulson et al. showed specific reporter gene transduction in SCLC cells using an AVP driven vector, indicating that AVP is also a candidate TSP for SCLC targeting.

Carcinogenesis related promoters, which are also candidate TSPs for other cancers, demonstrate lung cancer–specific transduction. For example, vascular endothelial growth factor (VEGF), which is one of the most effective mitogens specific for endothelial cells (EC), has been implicated in the neovascularization of a variety of tumors including lung cancers, especially in hypoxic conditions such as that found inside solid tumors. The mouse lung cancer cell line, A11, which was transfected with a vector harboring the VEGF promoter fused to the *HSV-tk* gene were demonstrated to give specific sensitivity to ganciclovir (GCV). This effect was enhanced by exposure to hypoxia followed by reoxygenation. The cytotoxic effect of the proapoptotic gene, *Bax*, was evaluated using an Ad vector expressing the human *Bax* gene under the control of the human VEGF promoter, suggesting a possible therapeutic application using this cancer-specific proapoptotic gene. Hexokinase type II (HKII), one of the hexokinases (HKs; ATP: D-hexose-6-phosphotransferase) that catalyzes the first committed step of glycolysis, is found in insulin-responsive tissues such as skeletal muscle, adipose tissue, and heart. In addition, tumors at various stages of malignancies show an increase in HK II activity compared with normal tissues. Katabi et al. reported that HKII promoter is another candidate TSP for non–small cell lung cancer by showing specific reporter gene expression and therapeutic effect of suicide gene therapy in an *in vivo* model.

Other TSPs, such as EGP-2/Ep-CAM and hTERT, which were previously utilized in gastrointestinal cancers and HCCs, were also successful for targeting lung cancer. Ad vectors with HSV-tk, CD and Bax, driven by these promoters demonstrated

specific gene transfer and therapeutic effects using both *in vitro* and *in vivo* models. Fukazawa et al. reported that a dual promoter system using the human surfactant protein A1 (hSPA1) promoter and the hTERT promoter were useful for targeting lung cancer as demonstrated by reporter gene analysis.

2.5 Breast Cancer

Breast cancer remains one of the most common malignancies in women today and many gene therapy strategies are studied in this field. In some women, breast cancer is a local disease without distant spread. However, most women with primary breast cancer have subclinical metastases. Therefore, systemic administration of targeted Ad vector is a reasonable approach in the treatment of breast cancer patients with metastasis.

Human α-lactalbumin (ALA) is an enzyme involved in lactose production and the breast cancer antigen ALA, which is normally expressed in the fully differentiated lactating mammary gland at the onset of parturition. ALA was shown to be expressed in more than 60% of breast cancer tissues in the clinical cases tested. The ovine β-lactoglobulin (BLG) protein is the most abundant whey protein in the milk of ruminants and it is regulated in a hormone-dependent manner similar to ALA, although it is not normally expressed in mice or humans. Leonard et al. constructed a series of Ad vectors that express *CD* gene under the transcriptional control of either the ALA or the BLG promoter, and demonstrated a therapeutic effect in a breast cancer xenograft model. CEA, previously mentioned as a TSP for GI cancers, is also a tumor marker of breast cancers, and increased CEA level in the serum has been observed in approximately 10% of the total breast cancer patients and 40 to 70% of recurrent ones. Qiao et al. reported that an Ad vector with *HSV-tk* gene driven by the CEA promoter demonstrated CEA-specific breast cancer cell killing using both *in vitro* and *in vivo* models.

Highly malignant cancers, such as breast cancer, are usually thermoresistant because they produce heat shock protein 70 (hsp70). Braiden et al. examined the effects of hsp promoter-mediated suicide gene therapy using the HSV-tk/GCV system for breast cancer in combination with hyperthermia. This report demonstrated that combination strategies were synergistically effective in mediating Fas-dependent apoptosis for breast cancer–specific gene therapy. Another stress-inducible chaperone protein with antiapoptotic properties is GRP78. GRP78 is overexpressed in transformed cells and during glucose starvation, acidosis, and hypoxic conditions, characteristic of poorly vascularized tumors. Dong et al. demonstrated that the GRP78 promoter is able to eradicate tumors using murine cells in immunocompetent models by driving expression of the *HSV-tk* suicide gene. Further, this system offers the advantage of positron-emission tomography (PET) imaging in combination with the suicide gene therapy in a breast cancer model.

Breast cancers usually arise from epithelial cells, and epithelial-based cancers have a variety of target specific TSPs. L-plastin (*LP*), which gene codes for an actin-binding protein, has been shown to be expressed at high levels in human epithelial cancer cells, but not in normal cells. Thus, LP is also a candidate TSP for targeting breast cancer. An Ad vector with an LP-driven *CD* or *E1* gene demonstrated a therapeutic effect in breast

cancer xenograft model. Additionally, the chemokine receptor, CXCR4, was recently reported to be markedly upregulated in breast cancer cells, but undetectable in normal mammary epithelial cells. An Ad vector with a CXCR4 promoter-driven reporter gene showed breast cancer–specific transduction, indicating that it is also a candidate TSP for targeting breast cancer. In breast cancer, overexpression of *erb*B-2 has been observed in approximately 20% and has been associated with reduced relapse-free patients and overall poor patient survival. Pandha et al. have shown that the tumor-specific *erb*B-2 promoter driving the *CD* gene resulted in levels of CD expression in a phase I clinical trial of breast cancer patients. Finally, two promoters of note are the hTERT and hexokinase promoters. The analysis of reporter gene expression driven by hTERT promoter or hexokinase promoter has demonstrated breast cancer–specific transduction, indicating that they are candidate TSPs for targeting breast cancer.

2.6 Urological Cancers

Gene therapy studies in urological diseases have been performed in prostate, bladder, and renal cell cancers. A variety of clinical gene therapy trials for prostate cancer have been reported, because of the large numbers of patients that are available and regional localization allowing a variety of direct intratumoral approaches. Prostate-specific antigen (PSA), a 34-kDa chymotrypsin-like serine protease and its promoter are known to be highly specific to prostate tissue and multiple PSA-targeted gene therapy strategies have been utilized for prostate cancer using reporter genes and suicide genes. However, due to low activity, some PSA promoter approaches have used a Cre/*loxP* system to enhance promoter activity. Prostate-specific membrane antigen (PSMA) is a type-2 membrane protein expressed in the prostate, and it is highly expressed in metastatic or poorly differentiated adenocarcinomas. Ad vectors with *HSV-tk* or *CD* gene driven by PSMA promoter, demonstrated a therapeutic effect using *in vitro* and *in vivo* models. These results illustrate that PSMA is also a candidate TSP for targeting prostate cancer.

Dihydrotestosterone (DHT)-inducible third-generation probasin-derived promoter, ARR(2)PB, which has been modified to contain two androgen response elements, has also been one of the prostate cancer–specific promoters used. Therapeutic viruses constructed with the Bax or Bad apoptotic protein gene driven by ARR(2)PB induced specific overexpression of these gene and apoptosis in prostate cancer cells. The T-cell receptor γ-chain alternate reading frame protein (TARP), is a protein that is uniquely expressed in prostate epithelial cells and prostate cancer cells. Cheng et al. showed that the transcriptional activity of an Ad vector with a chimeric sequence comprising of the TARP promoter and the PSA enhancer is highly active in testosterone-deprived prostate cancer cells. Caveolin-1 (a structural component of caveolae), human glandular kallikrein 2 (hK2) and osteocalcin (a major noncollagenous bone matrix protein) are upregulated in higher-grade stage, androgen resistant or metastatic prostate cancer. Ad vectors with these promoters have been demonstrated to possess prostate cancer cell–specific transduction or killing with reporter gene expression or suicide gene therapy.

Compared to prostate cancer, very few gene therapy strategies have been studied in bladder cancer or renal cell carcinoma

(RCC). In bladder cancer gene therapy, Ad vectors with either an *LacZ* or a *CD* gene driven by the L-plastin promoter demonstrated specific reporter gene expression and therapeutic effect, respectively. An Ad vector with the luciferase reporter gene driven by CXCR4 (one of the chemokine receptors), showed specific transduction of RCC cells with liver off status *in vivo*, indicating that the CXCR4 promoter is a novel candidate for transcriptional targeting of RCC.

2.7 Gynecological Cancers

In the gynecological field, ovarian and cervical cancers are good candidates for gene therapy. A number of targeting strategies for ovarian cancer have been reported, with targeting using hTERT promoter being utilized most frequently. Ad vectors in combination with luciferase, *HSV-tk* and *Bax* gene demonstrated ovarian cancer–specific transduction and cell killing. KDR/flk-1 and flt-1 are receptors of VEGF and are known to display dysregulated expression in both tumor vasculature and their progressions. These promoters were used for Ad construction with reporter and *CD* genes, showing usefulness for targeting certain ovarian cancers, such as nonendothelial originated ovarian cancer and teratocarcinomas.

The whey-acidic protein human epididymis protein 4 (HE4) is frequently overexpressed in ovarian cancer, suggesting that the HE4 promoter is highly transcriptionally active in this disease. Luciferase expression Ad vector driven by the HE4 promoter exhibited significant expression in ovarian cancer cells, but not in normal cells. Secretory leukoprotease inhibitor (SLPI) is a 12-kDa serine protease inhibitor expressed in some human carcinomas, including breast, lung, endometrium, and ovary. Selective Ad-mediated transgene expression could be achieved through the use of the SLPI promoter in the context of ovarian cancer. Mesothelin (MSLN), a cell surface glycoprotein, is overexpressed in ovarian cancer, but not in normal tissues with the exception of mesothelial cells. MSLN-driven luciferase expression Ad vector achieved transcriptional targeting in the context of ovarian cancer cells. Cox-2, midkine, hK2 and LP promoters have also been used for targeting ovarian cancer cells in the context of Ad vectors with reporter genes and suicide genes, demonstrating their utility as candidate TSPs for ovarian cancer.

For targeting cervical cancer cells, several TSPs, such as SLPI, cox-2, midkine, flt-1, VEGF, Survivin, and CXCR4 promoters have recently been screened using luciferase reporters in Ad vectors, with results demonstrating that VEGF and midkine are the most promising TSPs. Ad vectors with hTERT promoter-driven Bax and monocyte chemoattractant protein-1 (*MCP-1*) gene have also demonstrated tumor-specific cell killing in cervical cancer cells.

2.8 Brain Tumors

Brain tumors are some of the most fatal malignancies, and do not respond to conventional therapies, such as chemotherapy and radiotherapy. A number of gene therapy strategies, such as suicide gene therapy, have been performed with respect to these types of cancers. However, few systemic transcriptional targeting strategies have been studied in this field, because brain tumors exist in a compartment protected by a blood-brain barrier. This fact

also means that little amounts of virus flowing out from the compartment would occur in the context of local administration of therapeutic Ad vectors. Glial fibrillary acidic protein, GFA2, is an intermediate filament protein expressed primarily in astrocytes. Vandier et al. constructed a recombinant Ad vector, Adgfa2TK, in which the HSV-tk was gene driven by GFA2 promoter, demonstrating inhibition of glioma cells growing *in vitro* and *in vivo*. Ad vectors with the hTERT promoter driving the caspase-8 or the Fas associated protein with death domain (*FADD*) genes showed specific cell killing, which indicated that hTERT is also a candidate TSP for targeting gliomas. Human glandular kallikrein 2 has also been reported to be useful for gliomas as well as cervical cancer mentioned earlier.

Pituitary tumors producing prolactin (PRL) were successfully treated using Ad vectors containing the *HSV-tk* gene driven by the PRL promoter both *in vitro* and *in vivo*. Proopiomelanocortin (*POMC*) gene encodes ACTH, which causes Cusing's syndrome, when overexpressed. The POMC promoter provided useful treatment of ACTH-producing pituitary tumors via Ad transduction.

2.9 Thyroid Tumors

The thyroid gland is a relatively common site of malignant neoplasms, giving rise to 90% of all endocrine cancers, and targeted therapy is especially important for this disease in regard to preserving normal thyroid function. Thyroid carcinomas are pathologically classified into papillary, follicular, anaplastic, and medullary cancers. The two most frequent entities, papillary and follicular thyroid cancers, are usually referred to as *differentiated thyroid carcinoma* and are always positive for TG immunohistochemical staining. The Ad vectors with *HSV-tk* gene driven by the TG promoter–mediated thyroid cancer–specific cell-killing ability, with less toxicity to normal cells. On the other hand, some poorly differentiated, metastatic or most anaplastic thyroid carcinomas show diminished TG expression, accompanied by loss of some thyroid-specific transcription factors. The promoter from the calcitonin/calcitonin gene-related peptide (*CT/CGRP*) gene was tested for targeting medullary thyroid carcinoma, and showed higher, specific transduction by reporter gene analysis. Undifferentiated thyroid carcinoma was successfully targeted using an hTERT promoter-driven luciferase Ad vector, and an hTERT-targeted HSV-tk Ad vectors demonstrated specific cell killing.

2.10 Malignant Melanoma

Melanoma is a relatively rare disease, but its biological characteristics are well investigated because of its accessibility. Melanoma inhibitory activity (MIA) protein, which is highly expressed in melanoma, plays an important role in melanoma metastasis and invasion, and MIA has been identified as a highly specific and sensitive marker for malignant melanoma. Tyrosinase, the enzyme catalyzing the rate-limiting step in melanin production, is exclusively expressed in melanocytes, pigment cells of the retina and melanoma cells. Rothfels et al. reported that MIA and tyrosinase promoters mediated suicide gene therapy of malignant melanoma, resulting in melanoma-specific cell killing. Survivin, a novel member of the inhibitor of apoptosis

(IAP) protein family, is expressed in human cancers, but is quiescent in normal differentiated tissues. A survivin-driven luciferase expression Ad vector showed high activity in melanoma cells with liver off status. Cox-2 is also reported to be expressed in melanoma cells and an Ad vector containing cox-2 promoter driving the luciferase gene, demonstrating activity of the cox-2 promoter in melanoma cell lines as well as in primary melanoma cells but not in normal cells.

2.11 Other Cancers

Several transcriptional targeting strategies have been reported in squamous cell cancer and pediatric tumors. Two inducible promoters, early growth response (Egr-1), which is radiation sensitive, and Hsp70 were demonstrated to induce squamous cell cancer–specific transduction. In pediatric tumors, tyrosine hydroxylase promoter and midkine promoter were tested using neuroblastoma and Wilms' tumor cells, demonstrating their usefulness for targeting these cancers. Osteocalcin (OS) is a major noncollagenous protein of bone regulated by 1,25dihydroxyvitamin DJ [1,25-(OH)pDa] and it is believed to be expressed only by differentiated osteoblasts. The OS promoter for osteosarcoma and the hTERT promoter for fibrosarcoma were reported as useful TSPs for specific cell killing or therapeutic effects.

2.12 Normal Tissue/Organs

Fewer numbers of gene therapy strategies have been studied for nonmalignant diseases for the following reasons. First, nonmalignant targeted cells, such as endothelial cells, muscle cells, neurons and hematopoietic cells, are relatively resistant to Ad infection. Further reduction of transgene expression through transcriptional targeting would result in insufficient therapeutic effects, since most transcriptional targeting involves the loss of transgene activity in exchange for specificity. Second, most therapeutic genes for nonmalignant diseases are nontoxic ones, such as enzymes, hormones, and other important factors of cell function, which do not cause nontargeting cell death that would have resulted in severe side effects. Some gene therapy strategies for nonmalignant diseases, which would be inappropriate for targeting by constitutively active promoters, utilize targeting by TSPs (see Table 1).

2.13 Transcriptional Targeting of Replication-competent Ad

Conditionally replicative adenovirus (CRAd) agents have recently been applied for a variety of neoplastic diseases as a novel cancer gene therapy (see Fig. 2). In this approach, the virus replicates conditionally in transduced tumor cells, which lead to infection of neighboring cancer cells with progeny virus, resulting in virus-mediated oncolysis. Therefore, it is important to target CRAd agents more strictly than replication-incompetent Ad since even marginal levels of expression of TSP in untargeted normal cells could result in severe toxicity. Some TSPs that have been proved to be useful for targeting in the context of replication deficient Ad vectors have been studied in replication-competent CRAds agents (Table 2).

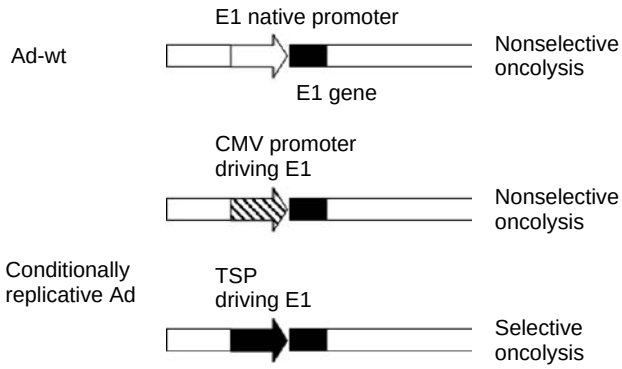

Fig. 2 Schematic representation of types of CRAd agents.

Tab. 1 Transcriptional targeting of nonmalignant cells.

Promoter	*Target cells*	*Transgene*	*Investigator*
Flt-1 (VEGFR-2)	Endothelial cells	Reporter gene	Nicklin
ICAM-1			
Von Willebrand factor			
KDR (VEGFR-1)	HUVEC	CD	Huang
Skeletal a-actin	Skeletal muscle	Reporter gene	Frauli
b-Enolase			
Creatine kinase			
Synapsin 1	Neuron	Reporter gene	Kugler
		BclX	Kilic
ICAM-2	Endothelial cells	Endoglin	Velasco
K18 (keratin 18)	Airway cells	Reporter gene	Toietta
Platelet-specific glycopron Iib	Platelets	Factor VIII	Shi
	Megakaryocytes		
Fascin	Dendritic cells (DC)	Reporter gene	Ross
Dectin-2	Langerhans cells (LC)	Reporter gene	Morita
Albumin	Hepatocytes	Reporter gene	Walther
α1-Antitrypsin		Reporter gene	Walther
Liver-type pyruvate kinase (LPK)		Reporter gene	Park

Notes: DC: dendritic cells; LC: Langerhans cells; LPK: liver-type pyruvate kinase.

2.14 Summary

A number of promoters have already been and will be reported as TSPs for transcriptional targeting, allowing us to choose a variety of targeting strategies that would be the best for the patients or diseases. In addition, transcriptional targeting is more advantageous compared to transductional targeting in respect to a "liver off" phenotype. However, the use of many TSPs results in reduced transgene expression in exchange for excellent specificity, which often causes insufficient therapeutic effect. Transductional augmentation, such

Tab. 2 Transcriptional targeting of replication-competent Ad.

Promoter	*Driving gene*	*Target disease/cells*	*Investigator*
Calcitonin	ICP4	Leiomyosarcoma	Yamamura
Cyclooxygenase-2	E1A	Pancreatic cancer	Yamamoto
	E1A	Ovarian cancer	Kanerva
	E1A	Esophageal cancer	Davydova
	E1A	Bladder cancer	Shirakawa
Flk-1	E1A	Endothelial cells	Savontaus
Human telomerase reverse transcriptase (hTERT)	E4	Breast cancer	Hernandez-Alcoceba
	E1A	Gastric cancer Cervical cancer Hepatocellular carcinoma Renal cell carcinoma	Wirth
	E1A	Hepatocellular carcinoma	Huang
	E1A	Lung cancer Hepatocellular carcinoma Cervical cancer Brain tumor	Kim
	E1A, E1B	Lung cancer	Kawashima
	E1A	Breast cancer Ovarian cancer Colon cancer	Huang
	E1A	Lung cancer Hepatocellular carcinoma Prostate cancer Colon cancer Pancreatic cancer Breast cancer	Irving
IAI.3B	E1A	Ovarian cancer	Hamada
L-plastin	E1A	Breast cancer Melanoma	Zhang
Midkine	E1A	Neuroblastoma Ewing's sarcoma	Adachi
	E1A	Glioma	Kohno
	E1A	Hepatocellular carcinoma	Yu
MN/CA9	E1A	Cervical cancer	Lim
Secretory laukoprotease inhibitor (SLPI)	E1A	Lung cancer	Maemondo
Surfactant protein B (SPB)	E4	Hepatocellular carcinoma Cervical cancer Colon cancer Prostate cancer	Doronin
Tyrosinase	E1A	Melanoma	Nettelbeck
	E1A	Melanoma	Peter
	E1A	Melanoma	Liu
	E1A, E4	Melanoma	Banerjee
Vascular endothelium growth factor (VEGF)	E1A	Ovarian cancer	Lam

Notes: SPB: Surfactant protein B.

as high dosage of the vector in combination with transductional targeting may result in a good therapeutic index.

3 Transductional Targeting

The extensively characterized capsid structure, genome, and replication cycle of Ad, particularly of the most commonly employed serotype 5, have allowed the molecular modifications required for their utilization of Ads as gene transfer vectors. Adenoviral infection is initiated by the recognition of the primary cellular receptor the CAR on target cells by the C-terminal portion of the fiber protein, termed the *knob*. After binding of the fiber knob domain, entry of the virus into the cell occurs via interaction of the Arg-Gly-Asp (RGD) sequence located in the viral penton base protein with cellular integrins (see Fig. 3). The development of genetically modified adenovirus (Ad) vectors with specificity for a single-cell type will require both the introduction of novel tropism determinants and the ablation of endogenous tropism. Consequently, it will not be possible to exploit the native cellular entry pathway in the propagation of these targeted Ad vectors. On this basis, Ad vector can be targeted at the level of transduction through CAR-independent infection pathways by both ablation of its wide native tropism and introduction of novel infectivity preference for target cells, a concept that is otherwise known as *retargeting*.

As most transductional targeting strategies have shown, the advantage of transductional targeting is that the retargeted Ad vector would function not only with respect to enhanced specificity but also with respect to enhanced activity of transgene expression due to improved infectivity via the receptor of interest that would typically be sacrificed in exchange for specificity in most cases of transcriptional targeting. Retargeting of Ad vectors was originally initiated with two different approaches, genetic Ad capsid modification and adaptor targeting molecules, both of which are designed to attain CAR-independent transduction pathway while forgoing the native CAR pathway. Recently, the combination

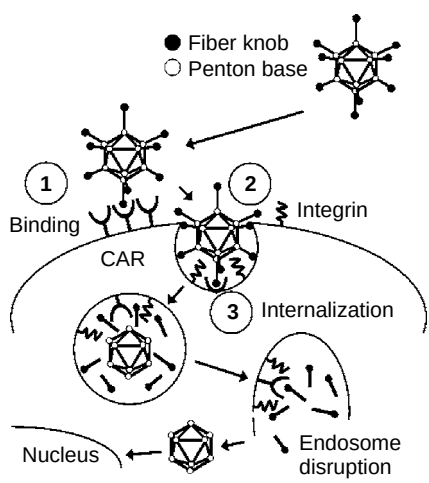

Fig. 3 The pathway of Ad entry.

of these two distinct methods was successfully achieved in an Ad vector by using the phage T4 fibritin motif.

3.1 Genetic Capsid Modification

3.1.1 Chimeric Ad Vector with Different Serotype or Species Fiber

The adenovirus fiber protein is responsible for attachment of the virion to cell surface receptors. The identity of the cellular receptor with which subgroup C adenoviruses, including Ad5 and Ad2, interact is CAR. Other subgroups including A–F except B also share this common CAR receptor for primary attachment. There is also evidence that suggests that Ad5 can also associate with other receptors including MHC class I proteins, heparin sulfate. Subgroup B adenoviruses, such as Ad3, Ad7, Ad11, Ad16, and Ad35 do not utilize CAR for cell entry. Several putative receptors for these viruses have been identified including CD46 for Ad35 and CD80/86 for Ad3. Stevenson et al. carried out a series of competition binding experiments using recombinant native fiber polypeptides from Ad5 and Ad3 and chimeric fiber proteins in which the head domains of Ad5 and Ad3 were exchanged. The results demonstrated that the determinants of Ad receptor binding are located in the knob (head) domain of the fiber, suggesting the possibility of altering the receptor specificity of the fiber protein by manipulation of sequences contained in the head domain. This very concept was initially tested by Krasnykh et al. and Stevenson et al., whereby an Ad vector containing chimeric fibers composed of the Ad3 fiber knob domain fused to the Ad5 fiber tail and shaft (Ad5/3). Ad5/3, which targets the Ad3 receptor, displayed high efficiency in circumventing CAR deficiency and enhancing gene delivery into neuroblastoma, B-lymphocytes, ovarian cancer, RCC, squamous cell carcinoma of the head and neck and melanoma cells.

Shayakhmetov et al. screened different Ad serotypes, namely, Ad3, Ad4, Ad5, Ad9, Ad35, and Ad41, for interaction with noncycling human CD34(+) cells and K562 cells at the level of virus attachment, internalization, and replication. Ad serotype 35 emerged as the variant with the highest tropism for CD34(+) cells, resulting in the completion of the chimeric Ad vector Ad5/F35, the fiber of which consists of the tail and shaft domains of Ad5 and the knob domain of Ad35. Ad5/F35 demonstrated enhanced transduction in many nonmalignant cells, such as hematopoietic cells, bone marrow cells, endothelial cells, fibroblasts, DCs, and retinal cells as well as glioma, leukemia, osteosarcoma, melanoma, colon cancer, and cervical cancer cells.

Various other chimeric Ad vectors have been reported, including Ad2 with Ad17 fiber (Ad2(17f)), Ad5 with Ad16 fiber (Ad5.Fib16) and Ad5 with Ad11 fiber (Ad5/11), demonstrating efficient transduction in airway epithelia, cardiovascular tissue, and hematopoietic cells, respectively. Interestingly, Schoggins et al. demonstrated that a chimeric Ad5 vector with the Ad41 fiber decreases transduction to hepatocytes and is useful for liver nontargeting. The Toronto strain of canine adenovirus type 2 (CAV2) exhibits native tropism and it has been evaluated as a gene therapy vector. An Ad5 vector containing the nonhuman CAV2 knob can efficiently bind to and enter CAR-deficient cells, with up to 30-fold augmentation in gene delivery in comparison to the Ad5 control vector.

3.1.2 Ad Vector with Modified Fiber Proteins

Retargeting approaches based on genetic alternation of the virion capsid via fiber knob modification with small peptides have achieved cell-specific gene delivery. Certain targeting motifs may be incorporated in this manner to route the virus toward nonnative cellular entry pathways. Wickham et al. constructed two Ad vectors that contained insertions in the Ad fiber coat protein to redirect virus binding to either α(v) integrin or heparan sulfate cellular receptors by inserting Arg-Gly-Asp (RGD) or polylysin (pK7) at the C terminus of the fiber knob region respectively. These vectors demonstrated increased transduction 5- to 500-fold in numerous cell types lacking sufficient levels of the primary Ad receptor including macrophage, endothelial, smooth muscle, fibroblast, and T cells. Krasnykh and Dmitriev et al. have shown that peptides incorporated into the HI loop of the fiber knob remains available for binding in the context of mature virions containing modified fibers, suggesting that heterologous ligands, such as RGD, could be incorporated into the HI loop of the fiber knob and that this locale possesses properties consistent with its employment in adenovirus retargeting strategies. These integrin- and/or heparan sulfate-targeting Ad vector demonstrated enhanced transgene expression in dendritic, leukemic, myeloma, head and neck cancer, ovarian cancer, and renal cell cancer cells.

Douglas et al. demonstrated retargeting of an Ad vector via nonadenoviral, artificial receptor with six histidine (6His) incorporated Ad in the knob and anti-6His single-chain antibody (scFv) expressing cells. Nicklin et al. generated genetically modified Ad fiber proteins with selective tropism to EC by engineering human endothelial cell-binding peptides SIGYPLP into the HI loop of the Ad fiber, demonstrating efficient and selective tropism for EC compared with control Ad vectors.

In summary, genetic capsid modification allows progeny virions to maintain the same physical and genome structure as their parent viruses. This property is essential when applying targeting methods to replication. Genetic capsid modification has proven to be a useful method for enhancing the infectivity of CRAds as described below. Despite the gains that have been achieved with transductional targeting via genetic methods, challenges still remain. In designing serotype or xenotype chimeras, it is difficult to predict the targeting effect in advance since cellular receptors and therefore the viral tropism of Ads other than those conventionally applied as vectors are unknown. Genetic fiber modification is complicated by the limited size restriction of peptides that can be incorporated due to the structural incompatibility between the fiber and the targeting ligand. Moreover, RGD and pK7, which are recognized by integrins and heparan sulfate that are broadly expressed in many types of cells, may contribute to enhance transductional activity but not its specificity. Combination with the other control mechanisms, such as transcriptional targeting, may be recommended to achieve more selective results.

3.2 Adaptor Molecules

Several retargeting strategies to abrogate native tropism and redirect Ad uptake through defined receptors have been attempted using bifunctional adaptor molecules. In this strategy, one element of the bispecific adaptor binds to the

Ad knob, blocking its interaction with CAR and hence its native tropism. Such attachment may be accomplished with anti-adenovirus knob Fab, single-chain Fv fragments (scFv), and the extracellular domain of CAR. The second component of the bispecific adaptor is chemically or genetically conjugated to the knob-binding portion, introducing specificity for the target cells. Some of these molecules have been antibodies (Fab, scFv) or ligands that bind to specific receptors expressed on target cells.

Particularly, these candidate molecules have included fibroblast growth factor (FGF) for Kaposi'e sarcoma, melanoma, and ovarian cancer cells; epidermal growth factor (EGF) for glioma, head and neck, skin, ovarian, and lung cancer cells; CD40 ligand for DC and ovarian cancer cells; epithelial cell adhesion molecule (Ep-CAM) for head and neck, ovarian, hepatocellular, and prostate cancer cells; angiotensin converting enzyme (ACE) for pulmonary ECs; CEA for breast cancer cells; PSMA for prostate cancer cells; MSLN for ovarian cancer cells; and high molecular weight melanoma-associated antigen (HMWMAA) for melanoma cells and the pancarcinoma antigen tumor-associated glycoprotein 72 (TAG-72), for ovarian cancer. (Table 3)

Transductional targeting with adaptor molecules was originally developed using a chemical conjugate comprising the Fab fragment of a monoclonal antibody directed against the Ad5 fiber and a full monoclonal antibody against the target molecules. However, chemical conjugates consisting of two individual portions may be difficult to manufacture as they have homogeneity property, which is undesirable in a product intended for clinical use. Recombinant fusion proteins offer a number of technological advantages including simplified production and purification when compared with chemical conjugates. Therefore, using bispecific adaptor-based Ad targeting with a recombinant fusion protein consisting of an extracellular portion of the Ad receptor CAR linked to an scFv or ligand to the targeting molecule may be a more rational approach (Table 3).

The use of bifunctional adaptors allows a myriad of molecules to be considered as candidates for rational vector targeting design. This method may be accomplished without concerns regarding size restrictions or perturbation of virus function that are commonly encountered in genetic modification techniques. With specific, defined targets incorporated into the design, cell receptor levels may be examined beforehand with flow cytometry or PCR analyses. Also, a great advantage is the fact that bispecific conjugates may be applied to the multitude of available Ad5-based vectors that have already been constructed for specific disease contexts. These vectors contain the unmodified Ad5 knob with which the conjugates can interact. On the other hand, retargeting in this case is successful only during the initial transduction, but not in replication-dependent situations involving subsequent infection by progeny virions, unless the adaptor molecules are continuously and appropriately provided *in situ*. Therefore, it is unclear how feasible this method would be in the context of replication-competent Ad vectors.

3.3 Genetic Capsid Modification Combined with Adaptor Molecules

A potential barrier to the development of Ad targeting with the aforementioned adaptor molecules for cell-specific delivery lies in the fact that several types of targeting

Tab. 3 Transductional targeting with adaptor molecules.

Targeting molecule	*Target cell recognition site*	*Ad recognition site*	*Type of molecule*	*Target cell*	*Investigator*
Anti-6 Histidine Fab	Six histidine	Anti-Ad5 fiber scFv	Fusion protein	Glioma expressing anti-His tag sFv	Douglas
Fibroblast growth factor (FGF)	Anti-FGF Fab	Anti-Ad5 knob Fab	Chemical conjugate	Kaposi's sarcoma	Goldman
				Melanoma	Gu
				Ovarian cancer	Rancourt
Epidermal growth factor receptor (EGFR)	Anti-EGFR Fab	Anti-Ad5 knob Fab	Chemical conjugate	Glioma	Miller
				Head and neck cancer	Blackwell
	Anti-EGFR scFv	Anti-Ad5 fiber scFv	Fusion protein	Glioma	Haisma
				Ovarian cancer	
				Skin cancer	
				Lung cancer	
	Anti-EGFR scFv	Soluble CAR	Fusion protein	Head and neck cancer	Dmitriev
				Ovarian cancer	
				Skin cancer	
				Skin cancer	Hemminki
c-erbB-2	Anti-c-erbB-2 scFv	Soluble CAR	Fusion protein	Breast cancer	Kashentseva
				Ovarian cancer	
CD40	Anti-CD40 Fab	Anti-Ad5 knob Fab	Chemical conjugate	Dendritic cells	de Gruijl
	Anti-CD40 scFv	Soluble CAR	Fusion protein	Dendritic cells	Pereboev
				Ovarian cancer	Hakkarainen
	CD40 ligand	Soluble CAR	Fusion protein	Dendritic cells	Pereboev
Angiotensin converting enzyme (ACE)	Anti-ACE Fab	Anti-Ad5 knob Fab	Chemical conjugate	Pulmonary epithelial cells	Reynolds
Epithelial cell adhesion molecule (Ep-CAM)	Anti-EpCAM Fab	Anti-Ad5 knob Fab	Chemical conjugate	Head and neck cancer	Haisma
				Ovarian cancer	
				Hepatocellular carcinoma	
Prostate-specific membrane antigen (PSMA)	Anti-PSMA Fab	Anti-Ad5 knob Fab	Chemical conjugate	Prostate cancer	Kraaji
Carcinoembryonic antigen (CEA)	Anti-CEA scFv	Anti-Ad5 fiber scFv	Fusion protein	Breast cancer	Korn
High molecular weight melanoma-associated antigen (HMWMAA)	Anti-HMWMAA	Anti-Ad5 fiber scFv	Fusion protein	Melanoma	Nettelbeck

protein ligands require posttranslational modifications, such as the formation of disulfide bonds. This fact poses a problem especially for Ad capsid proteins that do not undergo such posttranslational modifications and are mostly localized in the nucleus during assembly of the virions. In addition, recombinant or chemically engineered adaptor molecules have to be specifically tailored for each particular targeting situation. To overcome these problems, a new general targeting strategy, which combines genetic modifications of the Ad capsid with a protein bridge approach, was recently developed, allowing the direct formation of vector–ligand complexes with monoclonal antibodies against cell surface antigens for targeting of Ad vectors. This capsid modification method entails the genetic incorporation of the Fc-binding domain derived from staphylococcal protein A into the Ad fiber protein. Significant augmentation of transduction efficiency using EGFR or CD40 specific monoclonal antibodies was achieved by applying this scheme.

Another approach of capsid modification with a protein bridge is avidin-based targeting with metabolically biotinylated Ad vector. In this strategy, a biotin acceptor peptide (BAP) was genetically fused to the Ad capsid fiber protein, which, during propagation would be biotinylated by the endogenous biotin ligase in 293 cells to produce covalently biotinylated virions. The resulting biotinylated vector could be retargeted to new receptors by conjugation to biotinylated antibodies via tetrameric avidin, resulting in increased transduction 10- to 30-fold over untargeted Ad-Fiber-BAP-TR in the case of biotinylated anti-CD59 or anti-CD71 antibodies in HeLa cells. In addition to the many advantages of targeting with adaptor molecules, the combination of genetic capsid modification and adaptor proteins is useful especially for the screening of targeting molecules and would enable dual targeting with plural antibodies.

3.4 Genetic Capsid Modification with Incorporation of Larger-sized Targeting Molecules

Transductional targeting using adaptor molecules requires a more complicated procedure, especially for clinical usage. However, as noted above, attempts to alter Ad tropism by genetic modification of the Ad fiber have had limited success due to structural conflicts between the fiber and the targeting ligand. Additionally, adaptor molecule systems may only work effectively in situations not requiring virus replication. Therefore, novel approaches compatible with adenovirus replication are required to apply targeting to oncolytic Ad vectors. Krasnykh and Belousova et al. reported the use of the phage T4 fibritin trimerization motif to facilitate the genetic incorporation of larger targeting ligands into the Ad fiber. They validated this strategy with the human CD40 ligand (CD40L), demonstrating the incorporation of a large sized targeting molecule in the modified Ad5 fiber that had not been accomplished previously. This complex chimera was fully functional, conferring specific transgene expression in CD40 expressing cells.

4 Transductional and Transcriptional Dual Targeting

As stated above, various transductional and transcriptional approaches have been

devised to improve targeting of Ad vectors. Transductional targeting alters the natural infection pathway of the Ad vector to enhance gene delivery to the target tissue and to reduce transgene expression in the liver. However, genetic modifications of Ad to ablate CAR recognition have not reduced hepatic transgene expression. Secondary interactions between the RGD motif in the Ad penton base and cell surface integrins, which normally mediate internalization of the virion after primary attachment to CAR, may account for some of the residual hepatocyte transduction. These findings suggest the need for complementary approaches to achieve improved targeting.

Transcriptional targeting has achieved exceptional specificity to the target cells while maintaining a "liver off" profile. However, transcriptional targeting alone is of little merit if the target cells are poorly transduced. Also, no promoter is entirely specific without having to sacrifice promoter strength. On the basis of these considerations, combining the complementary approaches of transductional and transcriptional targeting, neither of which is perfect or "non-leaky" by itself may be a rational approach.

4.1 Combination of Targeting with Adaptor Molecule and Transcriptional Targeting

The first report of dual targeting by combining both transductional and transcriptional targeting was accomplished in the pulmonary endothelium. The transductional targeting utilized was on the basis of a bispecific antibody consisting of anti-ACE (pulmonary endothelium marker) and anti-Ad5 knob Fabs while transcriptional targeting was achieved with the endothelial-specific promoter flt-1. Transcriptionally targeted Ad vectors were complexed with transductional targeting proteins to redirect the transduction of cells in a specific manner. This approach resulted in a synergistic, 300 000-fold improvement in the selectivity of transgene expression for the lungs in comparison to the usual site of vector sequestration, the liver. For targeting brain tumors and osteosarcoma cells, the bispecific antibody consisting of anti-EGFR and anti-Ad5 knob Fabs was utilized for transductional and OS promoter, which is TSP for osteosarcoma, was utilized for transcriptional targeting, respectively, showing enhanced transduction of reporter genes while retaining specificity. The aforementioned fusion protein consisting of EGF, one of the EGFR ligands, and sCAR was also utilized for dual targeting. Ad vectors constructed with a reporter or the *HSV-tk* gene driven by the SLPI promoter, which is one of the TSPs for ovarian cancer, demonstrated enhanced transduction and therapeutic effect in ovarian cancer both *in vitro* and *in vivo*.

4.2 Genetic Capsid Modification Coupled with Transcriptional Targeting

Genetic capsid modification has demonstrated the ability of the cyclic RGD-4C peptide to retarget Ad tropism toward αv integrins, leading to enhanced gene delivery to a repertoire of cell types. However, vectors modified in this manner are not selective in that they can transduce a wide variety of cells. In fact, RGD-modified vectors were originally developed to broaden the tropism of Ad-based vectors and achieve enhanced infectivity. Strategies to maintain the increased

transduction afforded by RGD modification, but on a more cell-selective basis, have been exploited by combination with transcriptional targeting using TSPs. Work et al. combined transcriptional targeting using the endothelial-specific promoter, flt-1, with genetic fiber modification using the RGD motif for targeting endothelial cells, showing that double modification substantially shifted transduction profiles toward vascular endothelial cells in rat hepatocytes and endothelial cells. Nicklin et al. also combined transcriptional targeting on the basis of the flt-1 promoter with genetic fiber modification using the human endothelial cell-binding peptide SIGYPLP, demonstrating efficient and selective transduction of some cancer cell lines.

The combination of genetic capsid modification and transcriptional targeting with TSP gives full play to its abilities in CRAd, since the enhanced infectivity achieved by transductional targeting with capsid modification would lead to robust replication in target cancer cells, producing progeny virus that can go on to infect the surrounding cells. These replication and lateralization functions of CRAds greatly influence their efficacy. As noted in the previous section, fiber-modified Ad vectors, such as Ad5RGD and Ad5/3, show enhanced infectivity in various cancer cells by exploiting CAR-independent infection pathways. Cox-2 promoter-based CRAds with RGD fibers or Ad5/3 chimeric fibers have been constructed, demonstrating dramatic improvements in oncolysis of pancreatic, esophageal, and ovarian cancer cells both *in vitro* and *in vivo*. For melanoma treatment, a tyrosinase promoter–based CRAd was generated with an Ad5/3 chimeric fiber, also resulting in enhanced oncolysis.

5 Conclusions

In any treatment modality, a low therapeutic index resulting from poor curative effects in target cells and/or high toxicity to nontarget cells presents great obstacles hindering successful treatment of the disease. In this regard, Ad vectors can be targeted at the level of transduction and transcription to overcome these issues, contributing enhanced specificity and efficiency of therapeutic transgene expression. Transcriptional targeting is practical only for specificity, while transductional targeting can achieve both increased efficiency and specificity. Each targeting strategy possesses both advantages and disadvantages, perhaps making neither one alone adequate for gene therapy goals.

On the basis of the above considerations, combined transductional and transcriptional targeting represents a viable targeting method for the following two reasons. First, combined targeting enables both efficient and specific transduction that is essential for effective therapy. Additionally, increased transgene expression resulting from transductional targeting approaches that provide enhanced gene delivery would make it possible to apply TSPs that show high specificity, but have been disregarded due to their low promoter activity. Second, transgenes were screened stringently for specificity through both transductional and transcriptional targeting.

The number of gene therapy strategies carried out to date, however, has fallen short of expectations. It is obvious that seeking better efficiency and specificity are major considerations that need to be addressed. Current methodology for dual targeting may be too complicated at present for clinical usage. As introduced above, radically modified Ad vectors with

large-sized targeting molecules have been generated and they mainly contribute to simplifying clinical protocols through the combination of transcriptional targeting portion inside the capsid. Further investigation of Ad biology and identification of novel targeting molecules will lead to an improvement of current gene therapy approaches.

Acknowledgments

The authors thank Dr. J. Michael Mathis, Director, Gene Therapy Program, LSU Health Sciences Center, Shreveport, for critical reading of the manuscript. This work was supported by the following grants: National Institutes of Health R01 CA86881 (D.T.C.), R01 HL67962 (D.T.C.), Muscular Dystrophy Association MDA359 (D.T.C.) and The Department of Defense W81XWH-05-1-0035 (D.T.C.).

See also Cytomegalovirus and Varicella–zoster Virus Vaccines; Liposome Gene Transfection; Medicinal Chemistry; Oncology, Molecular; Retrovirus Integration, Biological and Biochemical Machineries for; Somatic Gene Therapy.

Bibliography

Books and Reviews

Alemany, R., Balague, C. (2000) Replicative adenoviruses for cancer therapy, *Nat. Biotechnol.* **18**, 723–727.

Culver, K.W., Blaese, R.M. (1994) Gene therapy for cancer, *Trends Genet.* **10**, 174–178.

DeGroot, L.J., Jameson, J.L. (2001) *Endocrinology*, 4th edition, Saunders, Philadelphia, PA.

Fengzhi, L. (2003) Survivin study: what is the next wave? *J. Cell. Physiol.* **197**, 8–29.

Freeman, S.M., Zwiebel, J.A. (1993) Gene therapy of cancer, *Cancer Invest.* **11**, 676–688.

Haviv, Y.S., Curiel, D.T. (2003) Engineering regulatory elements for conditionally-replicative adeno-viruses, *Curr. Gene Ther.* **3**, 357–385.

Herrmann, F. (1995) Cancer gene therapy: principles, problems, and perspectives, *J. Mol. Med.* **73**, 157–163.

Kirn, D., Martuza, R.L., Zwiebel, J. (2001) Replication-selective virotherapy for cancer: biological principles, risk management and future directions, *Nat. Med.* **7**, 781–787.

Kruyt, F.A., Curiel, D.T. (2002) Toward a new generation of conditionally replicating adenoviruses: pairing tumor selectivity with maximal oncolysis, *Hum. Gene Ther.* **13**, 485–495.

Nettelbeck, D.M. (2003) Virotherapeutics: conditionally replicative adenoviruses for viral oncolysis, *Anticancer Drugs* **14**, 577–584.

Oosterhoff, D., van Beusechem, V.W. (2004) Conditionally replicating adenoviruses as anticancer agents and ways to improve their efficacy, *J. Exp. Ther. Oncol.* **4**, 37–57.

Schmidt-Wolf, G., Schmidt-Wolf, I.G. (1994) Human cancer and gene therapy, *Ann. Hematol.* **69**, 273–279.

Shively, J.E., Beatty, J.D. (1985) CEA-related antigens: molecular biology and clinical significance, *Crit. Rev. Oncol. Hematol.* **2**, 355–399.

Thean, E.T., Toh, B. (1990) Serum human alpha-lactalbumin as a marker for breast cancer, *Br. J. Cancer* **61**, 773–775.

Verma, I.M., Somia, N. (1997) Gene therapy – promises, problems and prospects, *Nature* **389**, 239–242.

Williams, C., Shattuck-Brandt, R.L., DuBois, R.N. (1999) The role of COX-2 in intestinal cancer, *Ann. N. Y. Acad. Sci.* **899**, 72–83.

Wilson, J.E. (1995) Hexokinases, *Rev. Physiol. Biochem. Pharmacol.* **126**, 65–198.

Woll, P.J. (1996) *Growth Factors and Lung Cancer*, Lippincott-Raven, Philadelphia, PA.

Primary Literature

Adachi, Y., Matsubara, S., Muramatsu, T., Curiel, D.T., Reynolds, P.N. (2002) Midkine promoter-based adenoviral suicide gene

therapy to midkine-positive pediatric tumor, *J. Pediatr. Surg.* **37**, 588–592.

Adachi, Y., Reynolds, P.N., Yamamoto, M., Wang, M., Takayama, K., Matsubara, S., Muramatsu, T., Curiel, D.T. (2001) A midkine promoter-based conditionally replicative adenovirus for treatment of pediatric solid tumors and bone marrow tumor purging, *Cancer Res.* **61**, 7882–7888.

Adams, V., Kempf, W., Hassam, S., Briner, J. (1995) Determination of hexokinase isoenzyme I and II composition by RTPCR: increased hexokinase isoenzyme II in human renal cell carcinoma, *Biochem. Mol. Med.* **54**, 53–58.

Akbulut, H., Zhang, L., Tang, Y., Deisseroth, A. (2003) Cytotoxic effect of replication-competent adenoviral vectors carrying L-plastin promoter regulated E1A and cytosine deaminase genes in cancers of the breast, ovary and colon, *Cancer Gene Ther.* **10**, 388–395.

Alemany, R., Curiel, D.T. (2001) CAR-binding ablation does not change biodistribution and toxicity of adenoviral vectors, *Gene Ther.* **8**, 1347–1353.

Ali, S., McClenaghan, M., Simons, P.J., Clark, J.A. (1990) Characterization of the alleles encoding ovine beta-lactoglobulins A and B, *J. Mol. Biol.* **91**, 201–207.

Anderson, L.M., Krotz, S., Weitzman, S.A., Thimmapaya, B. (2000) Breast cancer-specific expression of the Candida albicans cytosine deaminase gene using a transcriptional targeting approach, *Cancer Gene Ther.* **7**, 845–852.

Banerjee, N.S., Rivera, A.A., Wang, M., Chow, L.T., Broker, T.R., Curiel, D.T., Nettelbeck, D.M. (2004) Analyses of melanoma-targeted oncolytic adenoviruses with tyrosinase enhancer/promoter-driven E1A, E4, or both in submerged cells and organotypic cultures, *Mol. Cancer Ther.* **3**, 437–449.

Barbee, R.W., Stapleton, D.D., Perry, B.D., Re, R.N., Murgo, J.P., Valentino, V.A., Cook, J.L. (1993) Prior arterial injury enhances luciferase expression following in vivo gene transfer, *Biochem. Biophys. Res. Commun.* **190**, 70–78.

Barker, S.D., Dmitriev, I.P., Nettelbeck, D.M., Liu, B., Rivera, A.A., Alvarez, R.D., Curiel, D.T., Hemminki, A. (2003) Combined transcriptional and transductional targeting improves the specificity and efficacy of adenoviral gene delivery to ovarian carcinoma, *Gene Ther.* **10**, 1198–1204.

Barker, S.D., Coolidge, C.J., Kanerva, A., Hakkarainen, T., Yamamoto, M., Liu, B., Rivera, A.A., Bhoola, S.M., Barnes, M.N., Alvarez, R.D., Curiel, D.T., Hemminki, A. (2003) The secretory leukoprotease inhibitor (SLPI) promoter for ovarian cancer gene therapy, *J. Gene Med.* **5**, 300–310.

Barnett, B.G., Tillman, B.W., Curiel, D.T., Douglas, J.T. (2002) Dual targeting of adenoviral vectors at the levels of transduction and transcription enhances the specificity of gene expression in cancer cells, *Mol. Ther.* **6**, 377–385.

Bauerschmitz, G.J., Nettelbeck, D.M., Kanerva, A., Baker, A.H., Hemminki, A., Reynolds, P.N., Curiel, D.T. (2002) The flt-1 promoter for transcriptional targeting of teratocarcinoma, *Cancer Res.* **62**, 1271–1274.

Belousova, N., Korokhov, N., Krendelshchikova, V., Simonenko, V., Mikheeva, G., Triozzi, P.L., Aldrich, W.A., Banerjee, P.T., Gillies, S.D., Curiel, D.T., Krasnykh, V. (2003) Genetically targeted adenovirus vector directed to CD40-expressing cells, *J. Virol.* **77**, 11367–11377.

Berry, N.B., Cho, Y.M., Harrington, M.A., Williams, S.D., Foley, J., Nephew, K.P. (2004) Transcriptional targeting in ovarian cancer cells using the human epididymis protein 4 promoter, *Gynecol. Oncol.* **92**, 896–904.

Bilbao, R., Gerolami, R., Bralet, M.P., Qian, C., Tran, P.L., Tennant, B., Prieto, J., Brechot, C. (2000) Transduction efficacy, antitumoral effect, and toxicity of adenovirus-mediated herpes simplex virus thymidine kinase/ganciclovir therapy of hepatocellular carcinoma: the woodchuck animal model, *Cancer Gene Ther.* **7**, 657–662.

Blackwell, J.L., Li, H., Gomez-Navarro, J., Dmitriev, I., Krasnykh, V., Richter, C.A., Shaw, D.R., Alvarez, R.D., Curiel, D.T., Strong, T.V. (2000) Using a tropism-modified adenoviral vector to circumvent inhibitory factors in ascites fluid, *Hum. Gene Ther.* **11**, 1657–1669.

Braiden, V., Ohtsuru, A., Kawashita, Y., Miki, F., Sawada, T., Ito, M., Cao, Y., Kaneda, Y., Koji, T., Yamashita, S. (2000) Eradication of breast cancer xenografts by hyperthermic suicide gene therapy under the control of the heat shock protein promoter, *Hum. Gene Ther.* **11**, 2453–2463.

Breidenbach, M., Rein, D.T., Everts, M., Glasgow, J.N., Wang, M., Passineau, M.J., Alvarez, R.D., Korokhov, N., Curiel, D.T. (2005) Mesothelin-mediated targeting of adenoviral vectors for ovarian cancer gene therapy, *Gene Ther.*, **12**, 187–193.

Cao, G.W., Qi, Z.T., Pan, X., Zhang, X.Q., Miao, X.H., Feng, Y., Lu, X.H., Kuriyama, S., Du, P. (1998) Gene therapy for human colorectal carcinoma using human CEA promoter contro led bacterial ADP-ribosylating toxin genes human CEA: PEA & DTA gene transfer, *World J. Gastroenterol.* **4**, 388–391.

Cao, G., Kuriyama, S., Gao, J., Nakatani, T., Chen, Q., Yoshiji, H., Zhao, L., Kojima, H., Dong, Y., Fukui, H., Hou, J. (2001) Gene therapy for hepatocellular carcinoma based on tumour-selective suicide gene expression using the alpha-fetoprotein (AFP) enhancer and a housekeeping gene promoter, *Eur. J. Cancer* **37**, 140–147.

Casado, E., Gomez-Navarro, J., Yamamoto, M., Adachi, Y., Coolidge, C.J., Arafat, W.O., Barker, S.D., Wang, M.H., Mahasreshti, P.J., Hemminki, A., Gonzalez-Baron, M., Barnes, M.N., Pustilnik, T.B., Siegal, G.P., Alvarez, R.D., Curiel, D.T. (2001) Strategies to accomplish targeted expression of transgenes in ovarian cancer for molecular therapeutic applications, *Clin. Cancer Res.* **7**, 2496–2504.

Cheng, W.S., Kraaij, R., Nilsson, B., van der Weel, L., de Ridder, C.M., Totterman, T.H., Essand, M. (2004) A novel TARP-promoter-based adenovirus against hormone-dependent and hormone-refractory prostate cancer, *Mol. Ther.* **10**, 355–364.

Coulson, J.M., Stanley, J., Woll, P.J. (1999) Tumour-specific arginine vasopressin promoter activation in small-cell lung cancer, *Br. J. Cancer* **80**, 1935–1944.

Daniele, B., Bencivenga, A., Megna, A.S., Tinessa, V. (2004) Alpha-fetoprotein and ultrasonography screening for hepatocellular carcinoma, *Gastroenterology* **127**, S108–S112.

Davidoff, A.M., Stevenson, S.C., McClelland, A., Shochat, S.J., Vanin, E.F. (1999) Enhanced neuroblastoma transduction for an improved antitumor vaccine, *J. Surg. Res.* **83**, 95–99.

Davydova, J., Le, L.P., Gavrikova, T., Wang, M., Krasnykh, V., Yamamoto, M. (2004) Infectivity-enhanced cyclooxygenase-2-based conditionally replicative adenoviruses for esophageal adenocarcinoma treatment, *Cancer Res.* **64**, 4319–4327.

de Gruijl, T.D., Luykx-de Bakker, S.A., Tillman, B.W., van den Eertwegh, A.J., Buter, J., Lougheed, S.M., van der Bij, G.J., Safer, A.M., Haisma, H.J., Curiel, D.T., Scheper, R.J., Pinedo, H.M., Gerritsen, W.R. (2002) Prolonged maturation and enhanced transduction of dendritic cells migrated from human skin explants after in situ delivery of CD40-targeted adenoviral vectors, *J. Immunol.* **169**, 5322–5331.

Dematteo, R.P., McClane, S.J., Fisher, K., Yeh, H., Chu, G., Burke, C., Raper, S.E. (1997) Engineering tissue-specific expression of a recombinant adenovirus: selective transgene transcription in the pancreas using the amylase promoter, *J. Surg. Res.* **72**, 155–161.

Dmitriev, I., Kashentseva, E., Rogers, B.E., Krasnykh, V., Curiel, D.T. (2000) Ectodomain of coxsackievirus and adenovirus receptor genetically fused to epidermal growth factor mediates adenovirus targeting to epidermal growth factor receptor-positive cells, *J. Virol.* **74**, 6875–6884.

Dmitriev, I., Krasnykh, V., Miller, C.R., Wang, M., Kashentseva, E., Mikheeva, G., Belousova, N., Curiel, D.T. (1998) An adenovirus vector with genetically modified fibers demonstrates expanded tropism via utilization of a coxsackievirus and adenovirus receptor-independent cell entry mechanism, *J. Virol.* **72**, 9706–9713.

Dong, D., Dubeau, L., Bading, J., Nguyen, K., Luna, M., Yu, H., Gazit-Bornstein, G., Gordon, E.M., Gomer, C., Hall, F.L., Gambhir, S.S., Lee, A.S. (2004) Spontaneous and controllable activation of suicide gene expression driven by the stress-inducible grp78 promoter resulting in eradication of sizable human tumors, *Hum. Gene Ther.* **15**, 553–561.

Doronin, K., Kuppuswamy, M., Toth, K., Tollefson, A.E., Krajcsi, P., Krougliak, V., Wold, W.S. (2001) Tissue-specific, tumor-selective, replication-competent adenovirus vector for cancer gene therapy, *J. Virol.* **75**, 3314–3324.

Douglas, J.T., Rogers, B.E., Rosenfeld, M.E., Michael, S.I., Feng, M., Curiel, D.T. (1996) Targeted gene delivery by tropism-modified adenoviral vectors, *Nat. Biotechnol.* **14**, 1574–1578.

Douglas, J.T., Miller, C.R., Kim, M., Dmitriev, I., Mikheeva, G., Krasnykh, V., Curiel, D.T. (1999) A system for the propagation of

adenoviral vectors with genetically modified receptor specificities, *Nat. Biotechnol.* **17**, 470–475.

Dvorak, H.F., Brown, L.F., Detmer, M., Dvrak, A.M. (1995) Vascular permeability factor/vascular endothelial growth factor, microvascular hyperpermeability, and angiogenesis, *Am. J. Pathol.* **146**, 1029–1039.

Eberhart, C.E., Coffey, R.J., Radhika, A., Giardiello, F.M., Ferrenbach, S., DuBois, R.N. (1994) Up-regulation of cyclooxygenase 2 gene expression in human colorectal adenomas and adenocarcinomas, *Gastroenterology* **107**, 1183–1188.

Findlay, J., Brew, K. (1972) The complete amino acid sequence of human alpha-lactalbumin, *Eur. J. Biochem.* **27**, 65–86.

Frauli, M., Ribault, S., Neuville, P., Auge, F., Calenda, V. (2003) Adenoviral-mediated skeletal muscle transcriptional targeting using chimeric tissue-specific promoters, *Med. Sci. Monit.* **9**, BR–78–84.

Fukazawa, T., Maeda, Y., Sladek, F.M., Owen-Schaub, L.B. (2004) Development of a cancer-targeted tissue-specific promoter system, *Cancer Res.* **64**, 363–369.

Gerrish, K., Gannon, M., Shih, D., Henderson, E., Stoffel, M., Wright, C.V., Stein, R. (2000) Pancreatic beta cell-specific transcription of the pdx-1 gene. The role of conserved upstream control regions and their hepatic nuclear factor 3beta sites, *J. Biol. Chem.* **275**, 3485–3492.

Glasgow, J.N., Kremer, E.J., Hemminki, A., Siegal, G.P., Douglas, J.T., Curiel, D.T. (2004) An adenovirus vector with a chimeric fiber derived from canine adenovirus type 2 displays novel tropism, *Virology* **324**, 103–116.

Goldberg, M.A., Schneider, T.J. (1994) Similarities between the oxygen-sensing mechanisms regulating the expression of vascular endothelial growth factor and erythropoietin, *J. Biol. Chem.* **269**, 4355–4359.

Goldman, C.K., Rogers, B.E., Douglas, J.T., Sosnowski, B.A., Ying, W., Siegal, G.P., Baird, A., Campain, J.A., Curiel, D.T. (1997) Targeted gene delivery to Kaposi's sarcoma cells via the fibroblast growth factor receptor, *Cancer Res.* **57**, 1447–1451.

Gonzalez, R., Vereecque, R., Wickham, T.J., Vanrumbeke, M., Kovesdi, I., Bauters, F., Fenaux, P., Quesnel, B. (1999) Increased gene transfer in acute myeloid leukemic cells by an adenovirus vector containing a modified fiber protein, *Gene Ther.* **6**, 314–320.

Gonzalez, R., Vereecque, R., Wickham, T.J., Facon, T., Hetuin, D., Kovesdi, I., Bauters, F., Fenaux, P., Quesnel, B. (1999) Transduction of bone marrow cells by the AdZ.F(pK7) modified adenovirus demonstrates preferential gene transfer in myeloma cells, *Hum. Gene Ther.* **10**, 2709–2717.

Gu, J., Andreeff, M., Roth, J.A., Fang, B. (2002) hTERT promoter induces tumor-specific Bax gene expression and cell killing in syngenic mouse tumor model and prevents systemic toxicity, *Gene Ther.* **9**, 30–37.

Gu, J., Kagawa, S., Takakura, M., Kyo, S., Inoue, M., Roth, J.A., Fang, B. (2000) Tumor-specific transgene expression from the human telomerase reverse transcriptase promoter enables targeting of the therapeutic effects of the Bax gene to cancers, *Cancer Res.* **60**, 5359–5364.

Gu, D.L., Gonzalez, A.M., Printz, M.A., Doukas, J., Ying, W., D'Andrea, M., Hoganson, D.K., Curiel, D.T., Douglas, J.T., Sosnowski, B.A., Baird, A., Aukerman, S.L., Pierce, G.F. (1999) Fibroblast growth factor 2 retargeted adenovirus has redirected cellular tropism: evidence for reduced toxicity and enhanced antitumor activity in mice, *Cancer Res.* **59**, 2608–2614.

Gugala, Z., Olmsted-Davis, E.A., Gannon, F.H., Lindsey, R.W., Davis, A.R. (2003) Osteoinduction by ex vivo adenovirus-mediated BMP2 delivery is independent of cell type, *Gene Ther.* **10**, 1289–1296.

Haisma, H.J., Grill, J., Curiel, D.T., Hoogeland, S., van Beusechem, V.W., Pinedo, H.M., Gerritsen, W.R. (2000) Targeting of adenoviral vectors through a bispecific single-chain antibody, *Cancer Gene Ther.* **7**, 901–904.

Haisma, H.J., Pinedo, H.M., Rijswijk, A., der Meulen-Muileman, I., Sosnowski, B.A., Ying, W., Beusechem, V.W., Tillman, B.W., Gerritsen, W.R., Curiel, D.T. (1999) Tumor-specific gene transfer via an adenoviral vector targeted to the pan-carcinoma antigen EpCAM, *Gene Ther.* **6**, 1469–1474.

Hakkarainen, T., Hemminki, A., Pereboev, A.V., Barker, S.D., Asiedu, C.K., Strong, T.V., Kanerva, A., Wahlfors, J., Curiel, D.T. (2003) CD40 is expressed on ovarian cancer cells and can be utilized for targeting adenoviruses, *Clin. Cancer Res.* **9**, 619–624.

Hamada, K., Kohno, S., Iwamoto, M., Yokota, H., Okada, M., Tagawa, M., Hirose, S., Yamasaki, K., Shirakata, Y., Hashimoto, K., Ito, M. (2003) Identification of the human IAI.3B promoter element and its use in the construction of a replication-selective adenovirus for ovarian cancer therapy, *Cancer Res.* **63**, 2506–2512.

Harris, J.D., Gutierrez, A.A., Hurst, H.C., Sikora, K., Lemoine, N.R. (1994) Gene therapy for cancer using tumour-specific prodrug activation, *Gene Ther.* **1**, 170–175.

Havenga, M.J., Lemckert, A.A., Grimbergen, J.M., Vogels, R., Huisman, L.G., Valerio, D., Bout, A., Quax, P.H. (2001) Improved adenovirus vectors for infection of cardiovascular tissues, *J. Virol.* **75**, 3335–3342.

Haviv, Y.S., van Houdt, W.J., Lu, B., Curiel, D.T., Zhu, Z.B. (2004) Transcriptional targeting in renal cancer cell lines via the human CXCR4 promoter, *Mol. Cancer Ther.* **3**, 687–691.

Haviv, Y.S., Blackwell, J.L., Kanerva, A., Nagi, P., Krasnykh, V., Dmitriev, I., Wang, M., Naito, S., Lei, X., Hemminki, A., Carey, D., Curiel, D.T. (2002) Adenoviral gene therapy for renal cancer requires retargeting to alternative cellular receptors, *Cancer Res.* **62**, 4213–4281.

Hemminki, A., Dmitriev, I., Liu, B., Desmond, R.A., Alemany, R., Curiel, D.T. (2001) Targeting oncolytic adenoviral agents to the epidermal growth factor pathway with a secretory fusion molecule, *Cancer Res.* **61**, 6377–6381.

Hernandez-Alcoceba, R., Pihalja, M., Qian, D., Clarke, M.F. (2002) New oncolytic adenoviruses with hypoxia- and estrogen receptor-regulated replication, *Hum. Gene Ther.* **13**, 1737–1750.

Huang, X., Zhang, W., Wakimoto, H., Hamada, H., Cao, X. (2002) Adenovirus-mediated tissue-specific cytosine deaminase gene therapy for human hepatocellular carcinoma with different AFP expression levels, *J. Exp. Ther. Oncol.* **2**, 100–106.

Huang, T.G., Savontaus, M.J., Shinozaki, K., Sauter, B.V., Woo, S.L. (2003) Telomerase-dependent oncolytic adenovirus for cancer treatment, *Gene Ther.* **10**, 1241–1247.

Huang, Z.H., Yang, W.Y., Gong, X.W., Qian, Y., Che, X.Y. (2004) Killing effect of suicide gene system under control by KDR promotor on human umbilical vein endothelial cells, *Di Yi Jun Yi Da Xue Xue Bao* **24**, 139–143.

Huang, Q., Zhang, X., Wang, H., Yan, B., Kirkpatrick, J., Dewhrist, M.W., Li, C.Y. (2004) A novel conditionally replicative adenovirus vector targeting telomerase-positive tumor cells, *Clin. Cancer Res.* **10**, 1203–1205.

Huard, J., Lochmuller, H., Acsadi, G., Jani, A., Massie, B., Karpati, G. (1995) The route of administration is a major determinant of the transduction efficiency of rat tissues by adenoviral recombinants, *Gene Ther.* **2**, 107–115.

Ido, A., Uto, H., Moriuchi, A., Nagata, K., Onaga, Y., Onaga, M., Hori, T., Hirono, S., Hayashi, K., Tamaoki, T., Tsubouchi, H. (2001) Gene therapy targeting for hepatocellular carcinoma: selective and enhanced suicide gene expression regulated by a hypoxia-inducible enhancer linked to a human alpha-fetoprotein promoter, *Cancer Res.* **61**, 3016–3021.

Ikegami, S., Tadakuma, T., Ono, T., Suzuki, S., Yoshimura, I., Asano, T., Hayakawa, M. (2004) Treatment efficiency of a suicide gene therapy using prostate-specific membrane antigen promoter/enhancer in a castrated mouse model of prostate cancer, *Cancer Sci.* **95**, 367–370.

Imamura, H., Matsuyama, Y., Tanaka, E., Ohkubo, T., Hasegawa, K., Miyagawa, S., Sugawara, Y., Minagawa, M., Takayama, T., Kawasaki, S., Makuuchi, M. (2003) Risk factors contributing to early and late phase intrahepatic recurrence of hepatocellular carcinoma after hepatectomy, *J. Hepatol.* **38**, 200–207.

Irving, J., Wang, Z., Powell, S., O'Sullivan, C., Mok, M., Murphy, B., Cardoza, L., Lebkowski, J.S., Majumdar, A.S. (2004) Conditionally replicative adenovirus driven by the human telomerase promoter provides broad-spectrum antitumor activity without liver toxicity, *Cancer Gene Ther.* **11**, 174–185.

Jiang, Q., Ge, K., Xu, D.H., Sun, L.Y., Zheng, Z.C., Liu, X.Y. (1998) Use of carcinoembryonic antigen gene promoter in colorectal carcinoma-specific suicidal gene therapy, *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)*. **30**, 1–8.

Kaliberov, S.A., Buchsbaum, D.J., Gillespie, G.Y., Curiel, D.T., Arafat, W.O., Carpenter, M., Stackhouse, M.A. (2002) Adenovirus-mediated transfer of BAX driven by the vascular endothelial growth factor promoter

induces apoptosis in lung cancer cells, *Mol. Ther.* **6**, 190–198.

Kanerva, A., Bauerschmitz, G.J., Yamamoto, M., Lam, J.T., Alvarez, R.D., Siegal, G.P., Curiel, D.T., Hemminki, A. (2004) A cyclooxygenase-2 promoter-based conditionally replicating adenovirus with enhanced infectivity for treatment of ovarian adenocarcinoma, *Gene Ther.* **11**, 552–559.

Kanerva, A., Mikheeva, G.V., Krasnykh, V., Coolidge, C.J., Lam, J.T., Mahasreshti, P.J., Barker, S.D., Straughn, M., Barnes, M.N., Alvarez, R.D., Hemminki, A., Curiel, D.T. (2002) Targeting adenovirus to the serotype 3 receptor increases gene transfer efficiency to ovarian cancer cells, *Clin. Cancer Res.* **8**, 275–280.

Kashentseva, E.A., Seki, T., Curiel, D.T., Dmitriev, I.P. (2002) Adenovirus targeting to c-erbB-2 oncoprotein by single-chain antibody fused to trimeric form of adenovirus receptor ectodomain, *Cancer Res.* **62**, 609–616.

Kasono, K., Blackwell, J.L., Douglas, J.T., Dmitriev, I., Strong, T.V., Reynolds, P., Kropf, D.A., Carroll, W.R., Peters, G.E., Bucy, R.P., Curiel, D.T., Krasnykh, V. (1999) Selective gene delivery to head and neck cancer cells via an integrin targeted adenoviral vector, *Clin. Cancer Res.* **5**, 2571–2579.

Katabi, M.M., Chan, H.L., Karp, S.E., Batist, G. (1999) Hexokinase type II: a novel tumor-specific promoter for gene-targeted therapy differentially expressed and regulated in human cancer cells, *Hum. Gene Ther.* **10**, 155–164.

Katz, M.H., Spivack, D.E., Takimoto, S., Fang, B., Burton, D.W., Moossa, A.R., Hoffman, R.M., Bouvet, M. (2003) Gene therapy of pancreatic cancer with green fluorescent protein and tumor necrosis factor-related apoptosis-inducing ligand fusion gene expression driven by a human telomerase reverse transcriptase promoter, *Ann. Surg. Oncol.* **10**, 762–772.

Kawakami, Y., Li, H., Lam, J.T., Krasnykh, V., Curiel, D.T., Blackwell, J.L. (2003) Substitution of the adenovirus serotype 5 knob with a serotype 3 knob enhances multiple steps in virus replication, *Cancer Res.* **63**, 1262–1269.

Kawashima, T., Kagawa, S., Kobayashi, N., Shirakiya, Y., Umeoka, T., Teraishi, F., Taki, M., Kyo, S., Tanaka, N., Fujiwara, T. (2004) Telomerase-specific replication-selective virotherapy for human cancer, *Clin. Cancer Res.* **10**, 285–292.

Kelly, F.J., Miller, C.R., Buchsbaum, D.J., Gomez-Navarro, J., Barnes, M.N., Alvarez, R.D., Curiel, D.T. (2000) Selectivity of TAG-72-targeted adenovirus gene transfer to primary ovarian carcinoma cells versus autologous mesothelial cells in vitro, *Clin. Cancer Res.* **6**, 4323–4333.

Kijima, T., Osaki, T., Nishino, K., Kumagai, T., Funakoshi, T., Goto, H., Tachibana, I., Tanio, Y., Kishimoto, T. (1999) Application of the Cre recombinase/loxP system further enhances antitumor effects in cell type-specific gene therapy against carcinoembryonic antigen-producing cancer, *Cancer Res.* **59**, 4906–4911.

Kilic, E., Hermann, D.M., Kugler, S., Kilic, U., Holzmuller, H., Schmeer, C., Bahr, M. (2002) Adenovirus-mediated Bcl-X(L) expression using a neuron-specific synapsin-1 promoter protects against disseminated neuronal injury and brain infarction following focal cerebral ischemia in mice, *Neurobiol. Dis.* **11**, 275–284.

Kim, E., Kim, J.H., Shin, H.Y., Lee, H., Yang, J.M., Kim, J., Sohn, J.H., Kim, H., Yun, C.O. (2003) Ad-mTERT-delta19, a conditional replication-competent adenovirus driven by the human telomerase promoter, selectively replicates in and elicits cytopathic effect in a cancer cell-specific manner, *Hum. Gene Ther.* **14**, 1415–1428.

Kiwaki, K., Kanegae, Y., Saito, I., Komaki, S., Nakamura, K., Miyazaki, J.I., Endo, F., Matsuda, I. (1996) Correction of ornithine transcarbamylase deficiency in adult spf(ash) mice and in OTC-deficient human hepatocytes with recombinant adenoviruses bearing the CAG promoter, *Hum. Gene Ther.* **7**, 821–830.

Kohno, S., Nakagawa, K., Hamada, K., Harada, H., Yamasaki, K., Hashimoto, K., Tagawa, M., Nagato, S., Furukawa, K., Ohnishi, T. (2004) Midkine promoter-based conditionally replicative adenovirus for malignant glioma therapy, *Oncol. Rep.* **12**, 73–78.

Komata, T., Koga, S., Hirohata, S., Takakura, M., Germano, I.M., Inoue, M., Kyo, S., Kondo, S., Kondo, Y. (2001) A novel treatment of human malignant gliomas in vitro and in vivo: FADD gene transfer under the control of the human telomerase reverse transcriptase gene promoter, *Int. J. Oncol.* **19**, 1015–1020.

Komata, T., Kondo, Y., Kanzawa, T., Ito, H., Hirohata, S., Koga, S., Sumiyoshi, H., Takakura, M., Inoue, M., Barna, B.P., Germano, I.M., Kyo, S., Kondo, S. (2002) Caspase-8 gene therapy using the human telomerase reverse transcriptase promoter for malignant glioma cells, *Hum. Gene Ther.* **13**, 1015–1025.

Korn, T., Nettelbeck, D.M., Volkel, T., Muller, R., Kontermann, R.E. (2004) Recombinant bispecific antibodies for the targeting of adenoviruses to CEA-expressing tumour cells: a comparative analysis of bacterially expressed single-chain diabody and tandem scFv, *J. Gene Med.* **6**, 642–651.

Korokhov, N., Mikheeva, G., Krendelshchikov, A., Belousova, N., Simonenko, V., Krendelshchikova, V., Pereboev, A., Kotov, A., Kotova, O., Triozzi, P.L., Aldrich, W.A., Douglas, J.T., Lo, K.M., Banerjee, P.T., Gillies, S.D., Curiel, D.T., Krasnykh, V. (2003) Targeting of adenovirus via genetic modification of the viral capsid combined with a protein bridge, *J. Virol.* **77**, 12931–12940.

Koshikawa, N., Takenaga, K., Tagawa, M., Sakiyama, S. (2000) Therapeutic efficacy of the suicide gene driven by the promoter of vascular endothelial growth factor gene against hypoxic tumor cells, *Cancer Res.* **60**, 2936–2941.

Kraaij, R., Van Rijswijk, A.L., Oomen, M.H., Haisma, H.J., Bangma, C.H. (2005) Prostate specific membrane antigen (PSMA) is a tissue-specific target for adenoviral transduction of prostate cancer in vitro, *Prostate*, **62**, 253–259.

Krasnykh, V.N., Mikheeva, G.V., Douglas, J.T., Curiel, D.T. (1996) Generation of recombinant adenovirus vectors with modified fibers for altering viral tropism, *J. Virol.* **70**, 6839–6846.

Krasnykh, V., Belousova, N., Korokhov, N., Mikheeva, G., Curiel, D.T. (2001) Genetic targeting of an adenovirus vector via replacement of the fiber protein with the phage T4 fibritin, *J. Virol.* **75**, 4176–4183.

Krasnykh, V., Dmitriev, I., Mikheeva, G., Miller, C.R., Belousova, N., Curiel, D.T. (1998) Characterization of an adenovirus vector containing a heterologous peptide epitope in the HI loop of the fiber knob, *J. Virol.* **72**, 1844–1852.

Kremer, E.J., Boutin, S., Chillon, M., Danos, O. (2000) Canine adenovirus vectors: an alternative for adenovirus-mediated gene transfer, *J. Virol.* **74**, 505–512.

Kubo, H., Gardner, T.A., Wada, Y., Koeneman, K.S., Gotoh, A., Yang, L., Kao, C., Lim, S.D., Amin, M.B., Yang, H., Black, M.E., Matsubara, S., Nakagawa, M., Gillenwater, J.Y., Zhau, H.E., Chung, L.W. (2003) Phase I dose escalation clinical trial of adenovirus vector carrying osteocalcin promoter-driven herpes simplex virus thymidine kinase in localized and metastatic hormone-refractory prostate cancer, *Hum. Gene Ther.* **14**, 227–241.

Kugler, S., Kilic, E., Bahr, M. (2003) Human synapsin 1 gene promoter confers highly neuron-specific long-term transgene expression from an adenoviral vector in the adult rat brain depending on the transduced area, *Gene Ther.* **10**, 337–347.

Kutchera, W., Jones, D.A., Matsunami, N., Groden, J., McIntyre, T.M., Zimmerman, G.A., White, R.L., Prescott, S.M. (1996) Prostaglandin H synthase 2 is expressed abnormally in human colon cancer: Evidence for a transcriptional effect, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4816–4820.

Lam, J.T., Kanerva, A., Bauerschmitz, G.J., Takayama, K., Suzuki, K., Yamamoto, M., Bhoola, S.M., Liu, B., Wang, M., Barnes, M.N., Alvarez, R.D., Siegal, G.P., Curiel, D.T., Hemminki, A. (2004) Inter-patient variation in efficacy of five oncolytic adenovirus candidates for ovarian cancer therapy, *J. Gene Med.* **6**, 1333–1342.

Lee, E.J., Martinson, F., Kotlar, T., Thimmapaya, B., Jameson, J.L. (2001) Adenovirus-mediated targeted expression of toxic genes to adrenocorticotropin-producing pituitary tumors using the proopiomelanocortin promoter, *J. Clin. Endocrinol. Metab.* **86**, 3400–3409.

Lee, A.K., Delellis, R.A., Rosen, P.P., Herbert-Stanton, T., Tallberg, K., Garcia, C., Wolfe, H. (1984) Alpha-lactalbumin as an immunohistochemical marker for metastatic breast carcinomas, *Am. J. Surg. Pathol.* **8**, 93–100.

Lim, H.Y., Ahn, M., Chung, H.C., Gardner, T.A., Kao, C., Lee, S.J., Kim, S.J. (2004) Tumor-specific gene therapy for uterine cervical cancer using MN/CA9-directed replication-competent adenovirus, *Cancer Gene Ther.* **11**, 532–538.

Lin, T., Huang, X., Gu, J., Zhang, L., Roth, J.A., Xiong, M., Curley, S.A., Yu, Y., Hunt, K.K., Fang, B. (2002) Long-term tumor-free survival from treatment with the GFP-TRAIL fusion gene expressed from the hTERT promoter in breast cancer cells, *Oncogene* **21**, 8020–8028.

Liu, Y., Ye, T., Sun, D., Maynard, J., Deisseroth, A. (2004) Conditionally replication-competent adenoviral vectors with enhanced infectivity for use in gene therapy of melanoma, *Hum. Gene Ther.* **15**, 637–647.

Liu, J., Zou, W.G., Lang, M.F., Luo, J., Sun, L.Y., Wang, X.N., Qian, Q.J., Liu, X.Y. (2002) Cancer-specific killing by the CD suicide gene using the human telomerase reverse transcriptase promoter, *Int. J. Oncol.* **21**, 661–666.

Liu, F.F., Miller, N., Levin, W., Zanke, B., Cooper, B., Henry, M., Sherar, M.D., Pintilie, M., Hunt, J.W., Hill, R.P. (1996) The potential role of HSP70 as an indicator of response to radiation and hyperthermia treatments for recurrent breast cancer, *Int. J. Hyperthermia* **12**, 197–208.

Lovekin, C., Ellis, I.O., Locker, A., Robertson, J.F., Bell, J., Nicholson, R., Gullick, W.J., Elston, C.W., Blamey, R.W. (1991) c-erbB-2 oncoprotein expression in primary and advanced breast cancer, *Br. J. Cancer* **63**, 439–443.

Maemondo, M., Saijo, Y., Narumi, K., Kikuchi, T., Usui, K., Tazawa, R., Matsumoto, K., Nakamura, T., Sasaki, K., Takahashi, M., Niitsu, Y., Nukiwa, T. (2004) Gene therapy with secretory leukoprotease inhibitor promoter-controlled replication-competent adenovirus for non-small cell lung cancer, *Cancer Res.* **64**, 4611–4620.

Mallam, J.N., Hurwitz, M.Y., Mahoney, T., Chevez-Barrios, P., Hurwitz, R.L. (2004) Efficient gene transfer into retinal cells using adenoviral vectors: dependence on receptor expression, *Invest. Ophthalmol. Vis. Sci.* **45**, 1680–1687.

Matsumoto, K., Yoshimoto, J., Sugo, H., Kojima, K., Futagawa, S., Matsumoto, T. (2002) Relationship between the histological degrees of hepatitis and the postoperative recurrence of hepatocellular carcinoma in patients with hepatitis C, *Hepatol. Res.* **23**, 196–201.

McLaughlin, P.M., Trzpis, M., Kroesen, B.J., Helfrich, W., Terpstra, P., Dokter, W.H., Ruiters, M.H., de Leij, L.F., Harmsen, M.C. (2004) Use of the EGP-2/Ep-CAM promoter for targeted expression of heterologous genes in carcinoma derived cell lines, *Cancer Gene Ther.* **11**, 603–612.

Messina, M., Yu, D.M., Learoyd, D.L., Both, G.W., Molloy, P.L., Robinson, B.G. (2003) High level, tissue-specific expression of a modified calcitonin/calcitonin gene-related peptide promoter in a human medullary thyroid carcinoma cell line, *Mol. Cell. Endocrinol.* **164**, 219–224.

Miller, C.R., Buchsbaum, D.J., Reynolds, P.N., Douglas, J.T., Gillespie, G.Y., Mayo, M.S., Raben, D., Curiel, D.T. (1998) Differential susceptibility of primary and established human glioma cells to adenovirus infection: targeting via the epidermal growth factor receptor achieves fiber receptor-independent gene transfer, *Cancer Res.* **58**, 5738–5748.

Mink, S., Hartig, E., Jennewein, P., Doppler, W., Cato, A.C. (1992) A mammary cell-specific enhancer in mouse mammary tumor virus DNA is composed of multiple regulatory elements including binding sites for CTF/NFI and a novel transcription factor, mammary cell-activating factor, *Mol. Cell. Biol.* **12**, 4906–4918.

Mizuguchi, H., Hayakawa, T. (2002) Adenovirus vectors containing chimeric type 5 and type 35 fiber proteins exhibit altered and expanded tropism and increase the size limit of foreign genes, *Gene* **285**, 69–77.

Molina, R., Filella, X., Zanon, G., Pahisa, J., Alicarte, J., Munoz, M., Farrus, B., Ballesta, A.M. (2003) Prospective evaluation of tumor markers (c-erbB-2 oncoprotein, CEA and CA 15.3) in patients with locoregional breast cancer, *Anticancer Res.* **23**, 1043–1050.

Morimoto, E., Inase, N., Miyake, S., Yoshizawa, Y. (2001) Adenovirus-mediated suicide gene transfer to small cell lung carcinoma using a tumor-specific promoter, *Anticancer Res.* **21**, 329–331.

Morita, A., Ariizumi, K., Ritter, R. III, Jester, J.V., Kumamoto, T., Johnston, S.A., Takashima, A., (2001) Development of a Langerhans cell-targeted gene therapy format using a dendritic cell-specific promoter, *Gene Ther.* **8**, 1729–1737.

Muller, A., Homey, B., Soto, H., Ge, N., Catron, D., Buchanan, M.E., McClanahan, T., Murphy, E., Yuan, W., Wagner, S.N., Barrera, J.L., Mohar, A., Verastegui, E., Zlotnik, A. (2001) Involvement of chemokine receptors in breast cancer metastasis, *Nature* **410**, 50–56.

Nakamura, M., Kyo, S., Kanaya, T., Yatabe, N., Maida, Y., Tanaka, M., Ishida, Y., Fujii, C., Kondo, T., Inoue, M., Mukaida, N. (2004) hTERT-promoter-based tumor-specific expression of MCP-1 effectively sensitizes cervical

cancer cells to a low dose of cisplatin, *Cancer Gene Ther.* **11**, 1–7.

Nakaya, H., Ishizu, A., Ikeda, H., Tahara, M., Shindo, J., Itoh, R., Takahashi, T., Asaka, M., Ishikura, H., Yoshiki, T. (2003) In vitro model of suicide gene therapy for alpha-fetoprotein-producing gastric cancer, *Anticancer Res.* **23**, 3795–3800.

Nettelbeck, D.M., Rivera, A.A., Balague, C., Alemany, R., Curiel, D.T. (2002) Novel oncolytic adenoviruses targeted to melanoma: specific viral replication and cytolysis by expression of E1A mutants from the tyrosinase enhancer/promoter, *Cancer Res.* **62**, 4663–4670.

Nettelbeck, D.M., Rivera, A.A., Davydova, J., Dieckmann, D., Yamamoto, M., Curiel, D.T. (2003) Cyclooxygenase-2 promoter for tumour-specific targeting of adenoviral vectors to melanoma, *Melanoma Res.* **13**, 287–292.

Nettelbeck, D.M., Rivera, A.A., Kupsch, J., Dieckmann, D., Douglas, J.T., Kontermann, R.E., Alemany, R., Curiel, D.T. (2004) Retargeting of adenoviral infection to melanoma: combining genetic ablation of native tropism with a recombinant bispecific single-chain diabody (scDb) adapter that binds to fiber knob and HMWMAA, *Int. J. Cancer* **108**, 136–145.

Nicklin, S.A., Reynolds, P.N., Brosnan, M.J., White, S.J., Curiel, D.T., Dominiczak, A.F., Baker, A.H. (2001) Analysis of cell-specific promoters for viral gene therapy targeted at the vascular endothelium, *Hypertension* **38**, 65–70.

Nicklin, S.A., Von Seggern, D.J., Work, L.M., Pek, D.C., Dominiczak, A.F., Nemerow, G.R., Baker, A.H. (2001) Ablating adenovirus type 5 fiber-CAR binding and HI loop insertion of the SIGYPLP peptide generate an endothelial cell-selective adenovirus, *Mol. Ther.* **4**, 534–542.

Nicklin, S.A., Dishart, K.L., Buening, H., Reynolds, P.N., Hallek, M., Nemerow, G.R., Von Seggern, D.J., Baker, A.H. (2003) Transductional and transcriptional targeting of cancer cells using genetically engineered viral vectors, *Cancer Lett.* **201**, 165–173.

Nishimura, R., Nagao, K., Miyayama, H., Matsuda, M., Baba, K., Matsuoka, Y., Yamashita, H. (2003) Elevated serum CA15-3 levels correlate with positive estrogen receptor and initial favorable outcome in patients who died from recurrent breast cancer, *Breast Cancer* **10**, 220–227.

Nyati, M.K., Sreekumar, A., Li, S., Zhang, M., Rynkiewicz, S.D., Chinnaiyan, A.M., Rehemtulla, A., Lawrence, T.S. (2002) High and selective expression of yeast cytosine deaminase under a carcinoembryonic antigen promoter-enhancer, *Cancer Res.* **62**, 2337–2342.

Okada, N., Mori, N., Koretomo, R., Okada, Y., Nakayama, T., Yoshie, O., Mizuguchi, H., Hayakawa, T., Nakagawa, S., Mayumi, T., Fujita, T., Yamamoto, A. (2005) Augmentation of the migratory ability of DC-based vaccine into regional lymph nodes by efficient CCR7 gene transduction, *Gene Ther.* **12**, 129–139.

Olmsted-Davis, E.A., Gugala, Z., Gannon, F.H., Yotnda, P., McAlhany, R.E., Lindsey, R.W., Davis, A.R. (2002) Use of a chimeric adenovirus vector enhances BMP2 production and bone formation, *Hum. Gene Ther.* **13**, 1337–1347.

Onishi, T., Nouso, K., Higashi, T., Toshikuni, N., Nakatsukasa, H., Kobayashi, Y., Uemura, M., Yumoto, E., Fujiwara, K., Sato, S., Nakamura, S., Yokoyama, J., Hanafusa, T., Shiratori, Y. (2003) Cellular distribution of telomerase reverse transcriptase in human hepatocellular carcinoma, *J. Gastroenterol. Hepatol.* **18**, 1168–1174.

Pandha, H.S., Martin, L.A., Rigg, A., Hurst, H.C., Stamp, G.W., Sikora, K., Lemoine, N.R. (1999) Genetic prodrug activation therapy for breast cancer: a phase I clinical trial of erbB-2-directed suicide gene expression, *J. Clin. Oncol.* **17**, 2180–2189.

Park, H.S., Cheon, J., Cho, H.Y., Ko, Y.H., Bae, J.H., Moon, D.G., Kim, J.J. (2003) In vivo characterization of a prostate-specific antigen promoter-based suicide gene therapy for the treatment of benign prostatic hyperplasia, *Gene Ther.* **10**, 1129–1134.

Park, C.W., Park, Y.M., Lee, G.T., Lee, Y., Woo, S., Cha, J.Y., Ahn, C.W., Cha, B.S., Kim, K.S., Ahn, Y.H., Lee, H.C. (2004) Targeting of therapeutic gene expression to the liver by using liver-type pyruvate kinase proximal promoter and the SV40 viral enhancer active in multiple cell types, *Biochem. Biophys. Res. Commun.* **314**, 131–137.

Parrott, M.B., Adams, K.E., Mercier, G.T., Mok, H., Campos, S.K., Barry, M.A. (2003) Metabolically biotinylated adenovirus for cell targeting, ligand screening, and vector purification, *Mol. Ther.* **8**, 688–700.

Peng, S.Y., Lai, P.L., Chu, J.S., Lee, P.H., Tsung, P.T., Chen, D.S., Hsu, H.C. (1993)

Expression and hypomethylation of alpha-fetoprotein gene in unicentric and multicentric human hepatocellular carcinomas, *Hepatology* **17**, 35–41.

Peng, X.Y., Won, J.H., Rutherford, T., Fujii, T., Zelterman, D., Pizzorno, G., Sapi, E., Leavitt, J., Kacinski, B., Crystal, R., Schwartz, P., Deisseroth, A. (2001) The use of the L-plastin promoter for adenoviral-mediated, tumor-specific gene expression in ovarian and bladder cancer cell lines, *Cancer Res.* **61**, 4405–4413.

Pereboev, A.V., Nagle, J.M., Shakhmatov, M.A., Triozzi, P.L., Matthews, Q.L., Kawakami, Y., Curiel, D.T., Blackwell, J.L. (2004) Enhanced gene transfer to mouse dendritic cells using adenoviral vectors coated with a novel adapter molecule, *Mol. Ther.* **9**, 712–720.

Pereboev, A.V., Asiedu, C.K., Kawakami, Y., Dong, S.S., Blackwell, J.L., Kashentseva, E.A., Triozzi, P.L., Aldrich, W.A., Curiel, D.T., Thomas, J.M., Dmitriev, I.P. (2002) Coxsackie-virus-adenovirus receptor genetically fused to anti-human CD40 scFv enhances adenoviral transduction of dendritic cells, *Gene Ther.* **9**, 1189–1193.

Peshavaria, M., Henderson, E., Sharma, A., Wright, C.V., Stein, R. (1997) Functional characterization of the transactivation properties of the PDX-1 homeodomain protein, *Mol. Cell. Biol.* **17**, 3987–3996.

Peter, I., Graf, C., Dummer, R., Schaffner, W., Greber, U.F., Hemmi, S. (2003) A novel attenuated replication-competent adenovirus for melanoma therapy, *Gene Ther.* **10**, 530–539.

Plate, K.H., Breier, G., Weich, H.A., Risau, W. (1992) Vascular endothelial growth factor is a potential tumour angiogenesis factor in human gliomas in vivo, *Nature* **359**, 276–280.

Pollmann, A., Kabisch, H., Block, A., Muller, J., Hellwinkel, O.J. (2004) Limited specificity of promoter constructs for gene therapy in osteosarcoma, *Int. J. Mol. Med.* **14**, 737–742.

Pramudji, C., Shimura, S., Ebara, S., Yang, G., Wang, J., Ren, C., Yuan, Y., Tahir, S.A., Timme, T.L., Thompson, T.C. (2001) In situ prostate cancer gene therapy using a novel adenoviral vector regulated by the caveolin-1 promoter, *Clin. Cancer Res.* **7**, 4272–4279.

Qiao, J., Doubrovin, M., Sauter, B.V., Huang, Y., Guo, Z.S., Balatoni, J., Akhurst, T., Blasberg, R.G., Tjuvajev, J.G., Chen, S.H., Woo, S.L. (2002) Tumor-specific transcriptional targeting of suicide gene therapy, *Gene Ther.* **9**, 168–175.

Quirk, S.J., Gannell, J.E., Funder, J.W. (1988) Alpha-lactalbumin production by rat mammary gland: prepartum emergence of resistance to glucocorticoids and progestins, *Mol. Cell. Biol.* **58**, 183–189.

Rancourt, C., Rogers, B.E., Sosnowski, B.A., Wang, M., Piche, A., Pierce, G.F., Alvarez, R.D., Siegal, G.P., Douglas, J.T., Curiel, D.T. (1998) Basic fibroblast growth factor enhancement of adenovirus-mediated delivery of the herpes simplex virus thymidine kinase gene results in augmented therapeutic benefit in a murine model of ovarian cancer, *Clin. Cancer Res.* **4**, 2455–2461.

Rasmussen, H., Rasmussen, C., Lempicki, M., Durham, R., Brough, D., King, C.R., Weichselbaum, R. (2002) TNFerade Biologic: preclinical toxicology of a novel adenovector with a radiation-inducible promoter, carrying the human tumor necrosis factor alpha gene, *Cancer Gene Ther.* **9**, 951–957.

Rea, D., Havenga, M.J., van Den Assem, M., Sutmuller, R.P., Lemckert, A., Hoeben, R.C., Bout, A., Melief, C.J., Offringa, R. (2001) Highly efficient transduction of human monocyte-derived dendritic cells with subgroup B fiber-modified adenovirus vectors enhances transgene-encoded antigen presentation to cytotoxic T cells, *J. Immunol.* **166**, 5236–5244.

Rein, D.T., Breidenbach, M., Nettelbeck, D.M., Kawakami, Y., Siegal, G.P., Huh, W.K., Wang, M., Hemminki, A., Bauerschmitz, G.J., Yamamoto, M., Adachi, Y., Takayama, K., Dall, P., Curiel, D.T. (2004) Evaluation of tissue-specific promoters in carcinomas of the cervix uteri, *J. Gene Med.* **6**, 1281–1289.

Rempel, A., Bannasch, P., Mayer, D. (1994) Differences in expression and intracellular distribution of hexokinase isoenzymes in rat liver cells of different transformation stages, *Biochim. Biophys. Acta* **1219**, 660–668.

Reynolds, P.N., Dmitriev, I., Curiel, D.T. (1999) Insertion of an RGD motif into the HI loop of adenovirus fiber protein alters the distribution of transgene expression of the systemically administered vector, *Gene Ther.* **6**, 1336–1339.

Reynolds, P.N., Nicklin, S.A., Kaliberova, L., Boatman, B.G., Grizzle, W.E., Balyasnikova, I.V., Baker, A.H., Danilov, S.M., Curiel, D.T. (2001) Combined transductional and

transcriptional targeting improves the specificity of transgene expression in vivo, *Nat. Biotechnol.* **19**, 838–842.

Reynolds, P.N., Zinn, K.R., Gavrilyuk, V.D., Balyasnikova, I.V., Rogers, B.E., Buchsbaum, D.J., Wang, M.H., Miletich, D.J., Grizzle, W.E., Douglas, J.T., Danilov, S.M., Curiel, D.T. (2000) A targetable, injectable adenoviral vector for selective gene delivery to pulmonary endothelium in vivo, *Mol. Ther.* **2**, 562–578.

Ristimaki, A., Honkanen, N., Jankala, H., Sipponen, P., Harkonen, M. (1997) Expression of cyclooxygenase-2 in human gastric carcinoma, *Cancer Res.* **57**, 1276–1280.

Rivera, A.A., Davydova, J., Schierer, S., Wang, M., Krasnykh, V., Yamamoto, M., Curiel, D.T., Nettelbeck, D.M. (2004) Combining high selectivity of replication with fiber chimerism for effective adenoviral oncolysis of CAR-negative melanoma cells, *Gene Ther.* **11**, 1694–1702.

Ross, R., Sudowe, S., Beisner, J., Ross, X.L., Ludwig-Portugall, I., Steitz, J., Tuting, T., Knop, J., Reske-Kunz, A.B. (2003) Transcriptional targeting of dendritic cells for gene therapy using the promoter of the cytoskeletal protein fascin, *Gene Ther.* **10**, 1035–1040.

Rothfels, H., Paschen, A., Schadendorf, D. (2003) Evaluation of combined gene regulatory elements for transcriptional targeting of suicide gene expression to malignant melanoma, *Exp. Dermatol.* **12**, 799–810.

Sakai, Y., Kaneko, S., Sato, Y., Kanegae, Y., Tamaoki, T., Saito, I., Kobayashi, K. (2001) Gene therapy for hepatocellular carcinoma using two recombinant adenovirus vectors with alpha-fetoprotein promoter and Cre/lox P system, *J. Virol. Methods* **92**, 5–17.

Savontaus, M.J., Sauter, B.V., Huang, T.G., Woo, S.L. (2002) Transcriptional targeting of conditionally replicating adenovirus to dividing endothelial cells, *Gene Ther.* **9**, 972–979.

Schmidt, M., Heimberger, T., Gruensfelder, P., Schler, G., Hoppe, F. (2004) Inducible promoters for gene therapy of head and neck cancer: an in vitro study, *Eur. Arch. Otorhinolaryngol.* **261**, 208–215.

Schoggins, J.W., Gall, J.G., Falck-Pedersen, E. (2003) Subgroup B and F fiber chimeras eliminate normal adenovirus type 5 vector transduction in vitro and in vivo, *J. Virol.* **77**, 1039–1048.

Shayakhmetov, D.M., Papayannopoulou, T., Stamatoyannopoulos, G., Lieber, A. (2000) Efficient gene transfer into human CD34(+) cells by a retargeted adenovirus vector, *J. Virol.* **74**, 2567–2583.

Shayakhmetov, D.M., Li, Z.Y., Ternovoi, V., Gaggar, A., Gharwan, H., Lieber, A. (2003) The interaction between the fiber knob domain and the cellular attachment receptor determines the intracellular trafficking route of adenoviruses, *J. Virol.* **77**, 3712–3723.

Shen, L.Z., Wu, W.X., Xu, D.H., Zheng, Z.C., Liu, X.Y., Ding, Q., Hua, Y.B., Yao, K. (2002) Specific CEA-producing colorectal carcinoma cell killing with recombinant adenoviral vector containing cytosine deaminase gene, *World J. Gastroenterol.* **8**, 270–275.

Shi, Q., Wilcox, D.A., Fahs, S.A., Kroner, P.A., Montgomery, R.R. (2003) Expression of human factor VIII under control of the platelet-specific alphaIIb promoter in megakaryocytic cell line as well as storage together with VWF, *Mol. Genet. Metab.* **79**, 25–33.

Shirakawa, T., Hamada, K., Zhang, Z., Okada, H., Tagawa, M., Kamidono, S., Kawabata, M., Gotoh, A. (2004) A cox-2 promoter-based replication-selective adenoviral vector to target the cox-2-expressing human bladder cancer cells, *Clin. Cancer Res.* **10**, 4342–4348.

Singh, J., Hamid, R., Reddy, B.S. (1997) Dietary fat and colon cancer: modulation of cyclooxygenase-2 by types and amount of dietary fat during the postinitiation stage of colon carcinogenesis, *Cancer Res.* **57**, 3465–3470.

Slager, E.H., van der Minne, C.E., Goudsmit, J., van Oers, J.M., Kostense, S., Havenga, M.J., Osanto, S., Griffioen, M. (2004) Induction of CAMEL/NY-ESO-ORF2-specific CD8+ T cells upon stimulation with dendritic cells infected with a modified Ad5 vector expressing a chimeric Ad5/35 fiber, *Cancer Gene Ther.* **11**, 227–236.

Song, J.S., Kim, H.P., Yoon, W.S., Lee, K.W., Kim, M.H., Kim, K.T., Kim, H.S., Kim, Y.T. (2003) Adenovirus-mediated suicide gene therapy using the human telomerase catalytic subunit (hTERT) gene promoter induced apoptosis of ovarian cancer cell line, *Biosci. Biotechnol. Biochem.* **67**, 2344–2350.

Soudais, C., Laplace-Builhe, C., Kissa, K., Kremer, E.J. (2001) Preferential transduction of neurons by canine adenovirus vectors and

their efficient retrograde transport in vivo, *FASEB J.* **15**, 2283–2285.

Southgate, T.D., Windeatt, S., Smith-Arica, J., Gerdes, C.A., Perone, M.J., Morris, I., Davis, J.R., Klatzmann, D., Lowenstein, P.R., Castro, M.G. (2000) Transcriptional targeting to anterior pituitary lactotrophic cells using recombinant adenovirus vectors in vitro and in vivo in normal and estrogen/sulpiride-induced hyperplastic anterior pituitaries, *Endocrinology* **141**, 3493–3505.

Stahlecker, J., Gauger, A., Bosserhoff, A., Buttner, R., Ring, J., Hein, R. (2000) MIA as a reliable tumor marker in the serum of patients with malignant melanoma, *Anticancer Res.* **20**, 5041–5044.

Stecher, H., Shayakhmetov, D.M., Stamatoyannopoulos, G., Lieber, A. (2001) A capsid-modified adenovirus vector devoid of all viral genes: assessment of transduction and toxicity in human hematopoietic cells, *Mol. Ther.* **4**, 36–44.

Steffens, S., Sandquist, A., Frank, S., Fischer, U., Lex, C., Rainov, N.G., Kramm, C.M. (2004) A neuroblastoma-selective suicide gene therapy approach using the tyrosine hydroxylase promoter, *Pediatr. Res.* **56**, 268–277.

Stevenson, S.C., Rollence, M., Marshall-Neff, J., McClelland, A. (1997) Selective targeting of human cells by a chimeric adenovirus vector containing a modified fiber protein, *J. Virol.* **71**, 4782–4790.

Stevenson, S.C., Rollence, M., White, B., Weaver, L., McClelland, A. (1995) Human adenovirus serotypes 3 and 5 bind to two different cellular receptors via the fiber head domain, *J. Virol.* **69**, 2850–2857.

Stoffers, D.A., Zinkin, N.T., Stanojevic, V., Clarke, W.L., Habener, J.F. (1997) Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence, *Nat. Genet.* **15**, 106–110.

Szary, J., Kalita, K., Przybyszewska, M., Dus, D., Kieda, C., Janik, P., Szala, S. (2001) KDR promoter can transcriptionally target cytosine deaminase suicide gene to cancer cells of nonendothelial origin, *Anticancer Res.* **21**, 3471–3475.

Takeda, T., Inaba, H., Yamazaki, M., Kyo, S., Miyamoto, T., Suzuki, S., Ehara, T., Kakizawa, T., Hara, M., DeGroot, L.J., Hashizume, K. (2003) Tumor-specific gene therapy for undifferentiated thyroid carcinoma utilizing the telomerase reverse transcriptase promoter, *J. Clin. Endocrinol. Metab.* **88**, 3531–3538.

Takeda, T., Yamazaki, M., Minemura, K., Imai, Y., Inaba, H., Suzuki, S., Miyamoto, T., Ichikawa, K., Kakizawa, T., Mori, J., DeGroot, L.J., Hashizume, K. (2002) A tandemly repeated thyroglobulin core promoter has potential to enhance efficacy for tissue-specific gene therapy for thyroid carcinomas, *Cancer Gene Ther.* **9**, 864–874.

Tanaka, M., Inase, N., Miyake, S., Yoshizawa, Y. (2001) Neuron specific enolase promoter for suicide gene therapy in small cell lung carcinoma, *Anticancer Res.* **21**, 291–294.

Tanyi, J.L., Lapushin, R., Eder, A., Auersperg, N., Tabassam, F.H., Roth, J.A., Gu, J., Fang, B., Mills, G.B., Wolf, J. (2002) Identification of tissue- and cancer-selective promoters for the introduction of genes into human ovarian cancer cells, *Gynecol. Oncol.* **85**, 451–458.

Tillman, B.W., de Gruijl, T.D., Luykx-de Bakker, S.A., Scheper, R.J., Pinedo, H.M., Curiel, T.J., Gerritsen, W.R., Curiel, D.T. (1999) Maturation of dendritic cells accompanies high-efficiency gene transfer by a CD40-targeted adenoviral vector, *J. Immunol.* **162**, 6378–6383.

Tirone, T., Fagan, S.P., Templeton, N.S., Wang, X., Brunicardi, F.C. (2001) Insulinoma-induced hypoglycemic death in mice is prevented with beta cell specific gene therapy, *Ann. Surg. Oncol.* **233**, 603–611.

Toietta, G., Koehler, D.R., Finegold, M.J., Lee, B., Hu, J., Beaudet, A.L. (2003) Reduced inflammation and improved airway expression using helper-dependent adenoviral vectors with a K18 promoter, *Mol. Ther.* **7**, 649–658.

Torbenson, M., Kannangai, R., Abraham, S., Sahin, F., Choti, M., Wang, J. (2004) Concurrent evaluation of p53, beta-catenin, and alpha-fetoprotein expression in human hepatocellular carcinoma, *Am. J. Clin. Pathol.* **122**, 377–382.

Tsujii, M., DuBois, R.N. (1995) Alterations in cellular adhesion and apoptosis in epithelial cells overexpressing prostaglandin endoperoxide synthase 2, *Cell* **83**, 493–501.

Uchida, A., O'Keefe, D.S., Bacich, D.J., Molloy, P.L., Heston, W.D. (2001) In vivo suicide gene therapy model using a newly discovered prostate-specific membrane antigen promoter/enhancer: a potential alternative

approach to androgen deprivation therapy, *Urology* **58**, 132–139.

Ueda, K., Iwahashi, M., Nakamori, M., Nakamura, M., Matsuura, I., Yamaue, H., Tanimura, H. (2001) Carcinoembryonic antigen-specific suicide gene therapy of cytosine deaminase/5-fluorocytosine enhanced by the cre/loxP system in the orthotopic gastric carcinoma model, *Cancer Res.* **61**, 6158–6162.

Ueda, K., Iwahashi, M., Nakamori, M., Nakamura, M., Matsuura, I., Ojima, T., Yamaue, H. (2003) Improvement of carcinoembryonic antigen-specific prodrug gene therapy for experimental colon cancer, *Surgery* **133**, 309–317.

van der Eb, M.M., Cramer, S.J., Vergouwe, Y., Schagen, F.H., van Krieken, J.H., van der Eb, A.J., Rinkes, I.H., van de Velde, C.J., Hoeben, R.C. (1998) Severe hepatic dysfunction after adenovirus-mediated transfer of the herpes simplex virus thymidine kinase gene and ganciclovir administration, *Gene Ther.* **5**, 451–458.

Vandier, D., Rixe, O., Besnard, F., Kim, M., Rikiyama, T., Goldsmith, M., Brenner, M., Gouyette, A., Cowan, K.H. (2000) Inhibition of glioma cells in vitro and in vivo using a recombinant adenoviral vector containing an astrocyte-specific promoter, *Cancer Gene Ther.* **7**, 1120–1126.

Varga, M.J., Weibull, C., Everitt, E. (1991) Infectious entry pathway of adenovirus type 2, *J. Virol.* **65**, 6061–6070.

Varnavski, A.N., Calcedo, R., Bove, M., Gao, G., Wilson, J.M. (2005) Evaluation of toxicity from high-dose systemic administration of recombinant adenovirus vector in vector-naive and pre-immunized mice, *Gene Ther.* **12**, 427–436.

Velasco, B., Ramirez, J.R., Relloso, M., Li, C., Kumar, S., Lopez-Bote, J.P., Perez-Barriocanal, F., Lopez-Novoa, J.M., Cowan, P.J., d'Apice, A.J., Bernabeu, C. (2001) Vascular gene transfer driven by endoglin and ICAM-2 endothelial-specific promoters, *Gene Ther.* **8**, 897–904.

Volk, A.L., Rivera, A.A., Kanerva, A., Bauerschmitz, G., Dmitriev, I., Nettelbeck, D.M., Curiel, D.T. (2003) Enhanced adenovirus infection of melanoma cells by fiber-modification: incorporation of RGD peptide or Ad5/3 chimerism, *Cancer Biol. Ther.* **2**, 511–515.

Volpers, C., Thirion, C., Biermann, V., Hussmann, S., Kewes, H., Dunant, P., von der Mark, H., Herrmann, A., Kochanek, S., Lochmuller, H. (2003) Antibody-mediated targeting of an adenovirus vector modified to contain a synthetic immunoglobulin g-binding domain in the capsid, *J. Virol.* **77**, 2093–2104.

Von Seggern, D.J., Huang, S., Fleck, S.K., Stevenson, S.C., Nemerow, G.R. (2000) Adenovirus vector pseudotyping in fiber-expressing cell lines: improved transduction of epstein-barr virus-transformed B cells, *J. Virol.* **74**, 354–362.

Walther, W., Stein, U. (1996) Cell type specific and inducible promoters for vectors in gene therapy as an approach for cell targeting, *J. Mol. Med.* **74**, 379–392.

Wang, X.P., Yazawa, K., Yang, J., Kohn, D., Fisher, W.E., Brunicardi, F.C. (2004) Specific gene expression and therapy for pancreatic cancer using the cytosine deaminase gene directed by the rat insulin promoter, *J. Gastrointest. Surg.* **8**, 98–108.

Wesseling, J.G., Yamamoto, M., Adachi, Y., Bosma, P.J., van Wijland, M., Blackwell, J.L., Li, H., Reynolds, P.N., Dmitriev, I., Vickers, S.M., Huibregtse, K., Curiel, D.T. (2001) Midkine and cyclooxygenase-2 promoters are promising for adenoviral vector gene delivery of pancreatic carcinoma, *Cancer Gene Ther.* **8**, 990–996.

Wickham, T.J., Mathias, P., Cheresh, D.A., Nemerow, G.R. (1993) Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment, *Cell* **73**, 309–319.

Wickham, T.J., Tzeng, E., Shears, L.L. II, Roelvink, P.W., Li, Y., Lee, G.M., Brough, D.E., Lizonova, A., Kovesdi, I. (1997) Increased in vitro and in vivo gene transfer by adenovirus vectors containing chimeric fiber proteins, *J. Virol.* **71**, 8221–8229.

Wirth, T., Zender, L., Schulte, B., Mundt, B., Plentz, R., Rudolph, K.L., Manns, M., Kubicka, S., Kuhnel, F. (2003) A telomerase-dependent conditionally replicating adenovirus for selective treatment of cancer, *Cancer Res.* **63**, 3181–3188.

Work, L.M., Ritchie, N., Nicklin, S.A., Reynolds, P.N., Baker, A.H. (2004) Dual targeting of gene delivery by genetic modification of adenovirus serotype 5 fibers and cell-selective transcriptional control, *Gene Ther.* **11**, 1296–1300.

Worthington, J., McCarthy, H.O., Barrett, E., Adams, C., Robson, T., Hirst, D.G. (2004) Use of the radiation-inducible WAF1 promoter to drive iNOS gene therapy as a novel anti-cancer treatment, *J. Gene Med.* **6**, 673–680.

Wu, H., Han, T., Lam, J.T., Leath, C.A., Dmitriev, I., Kashentseva, E., Barnes, M.N., Alvarez, R.D., Curiel, D.T. (2004) Preclinical evaluation of a class of infectivity-enhanced adenoviral vectors in ovarian cancer gene therapy, *Gene Ther.* **11**, 874–878.

Xie, X., Zhao, X., Liu, Y., Young, C.Y., Tindall, D.J., Slawin, K.M., Spencer, D.M. (2001) Robust prostate-specific expression for targeted gene therapy based on the human kallikrein 2 promoter, *Hum. Gene Ther.* **12**, 549–561.

Yamamoto, M., Alemany, R., Adachi, Y., Grizzle, W.E., Curiel, D.T. (2001) Characterization of the cyclooxygenase-2 promoter in an adenoviral vector and its application for the mitigation of toxicity in suicide gene therapy of gastrointestinal cancers, *Mol. Ther.* **3**, 385–394.

Yamamoto, M., Davydova, J., Wang, M., Siegal, G.P., Krasnykh, V., Vickers, S.M., Curiel, D.T. (2003) Infectivity enhanced, cyclooxygenase-2 promoter-based conditionally replicative adenovirus for pancreatic cancer, *Gastroenterology* **125**, 1203–1218.

Yamamura, H., Hashio, M., Noguchi, M., Sugenoya, Y., Osakada, M., Hirano, N., Sasaki, Y., Yoden, T., Awata, N., Araki, N., Tatsuta, M., Miyatake, S.I., Takahashi, K. (2001) Identification of the transcriptional regulatory sequences of human calponin promoter and their use in targeting a conditionally replicating herpes vector to malignant human soft tissue and bone tumors, *Cancer Res.* **61**, 3969–3977.

Yoshimura, I., Ikegami, S., Suzuki, S., Tadakuma, T., Hayakawa, M. (2002) Adenovirus mediated prostate specific enzyme prodrug gene therapy using prostate specific antigen promoter enhanced by the Cre-loxP system, *J. Urol.* **168**, 2659–2664.

Yotnda, P., Onishi, H., Heslop, H.E., Shayakhmetov, D., Lieber, A., Brenner, M., Davis, A. (2001) Efficient infection of primitive hematopoietic stem cells by modified adenovirus, *Gene Ther.* **8**, 930–937.

Yu, L., Hamada, K., Namba, M., Kadomatsu, K., Muramatsu, T., Matsubara, S., Tagawa, M. (2004) Midkine promoter-driven suicide gene expression and -mediated adenovirus replication produced cytotoxic effects to immortalized and tumour cells, *Eur. J. Cancer* **40**, 1787–1794.

Zabner, J., Chillon, M., Grunst, T., Moninger, T.O., Davidson, B.L., Gregory, R., Armentano, D. (1999) A chimeric type 2 adenovirus vector with a type 17 fiber enhances gene transfer to human airway epithelia, *J. Virol.* **73**, 8689–8695.

Zhang, R., Straus, F.H., DeGroot, L.J. (2001) Adenoviral-mediated gene therapy for thyroid carcinoma using thymidine kinase controlled by thyroglobulin promoter demonstrates high specificity and low toxicity, *Thyroid* **11**, 115–123.

Zhang, L., Akbulut, H., Tang, Y., Peng, X., Pizzorno, G., Sapi, E., Manegold, S., Deisseroth, A. (2002) Adenoviral vectors with E1A regulated by tumor-specific promoters are selectively cytolytic for breast cancer and melanoma, *Mol. Ther.* **6**, 386–393.

Zhang, Y., Yu, J., Unni, E., Shao, T.C., Nan, B., Snabboon, T., Kasper, S., Andriani, F., Denner, L., Marcelli, M. (2002) Monogene and polygene therapy for the treatment of experimental prostate cancers by use of apoptotic genes bax and bad driven by the prostate-specific promoter ARR(2)PB, *Hum. Gene Ther.* **13**, 2051–2064.

Zhu, Z.B., Makhija, S.K., Lu, B., Wang, M., Kaliberova, L., Liu, B., Rivera, A.A., Nettelbeck, D.M., Mahasreshti, P.J., Leath, C.A., III, Yamaoto, M., Alvarez, R.D., Curiel, D.T. (2004) Transcriptional targeting of adenoviral vector through the CXCR4 tumor-specific promoter, *Gene Ther.* **11**, 645–658.

Zhu, Z.B., Makhija, S.K., Lu, B., Wang, M., Kaliberova, L., Liu, B., Rivera, A.A., Nettelbeck, D.M., Mahasreshti, P.J., Leath, C.A., Barker, S., Yamaoto, M., Li, F., Alvarez, R.D., Curiel, D.T. (2004) Transcriptional targeting of tumors with a novel tumor-specific survivin promoter, *Cancer Gene Ther.* **11**, 256–262.

Vectors and Gene Therapy

Edward A. Burton[1], *David J. Fink*[2], *and Joseph C. Glorioso*[2]
[1] *University of Pittsburgh, Pittsburgh, PA, USA*
[2] *University of Michigan, Ann Arbor, MI, USA*

1 Introduction 236

2 The Science of Gene Transfer 236

3 Principles of Vector Engineering 239
3.1 Retroviruses 239
3.2 Lentiviruses 240
3.3 Adenoviruses 241
3.4 Adeno-associated Viruses 246
3.5 Herpes Simplex Virus 248
3.6 Vector Targeting 252
3.7 Regulation of Gene Expression 254
3.8 Nonviral Vectors 255

4 Clinical Applications and Genes 257
4.1 Single Gene Disorders 257
4.1.1 Hematological Disease 257
4.1.2 Restoration of Circulating Proteins 258
4.1.3 Respiratory Disease 260
4.1.4 Neurological Disease 261
4.1.5 Muscle 262
4.2 Delivery of Proteins for a Pharmacological Action 263
4.2.1 Transplantation 263
4.2.2 Ischemic Heart Disease 264
4.2.3 Neurological Disease 265
4.2.4 Cancer Gene Therapy 266

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

5 Adverse Events 267
5.1 Multiple Organ Failure Secondary to Cytokine Storm 267
5.2 Leukemia Secondary to Insertional Mutagenesis 268

6 From Bench to Clinic – HSV-mediated Gene Therapy of Pain as an Example 270

7 Future Prospects – What will Gene Therapy Likely be About in 10 Years? 272

Bibliography 273
Books and reviews 273
Primary Literature 274

Keywords

Adeno-associated Virus
A nonenveloped single-stranded DNA virus that may be modified to produce a gene transfer vector.

Adenovirus
A nonenveloped double-stranded DNA virus that may be modified to produce a gene transfer vector.

***Ex vivo* Gene Transfer**
Transfer of genetic material to cells cultured outside the body of a living organism.

Gene Therapy
The introduction of genetic sequences into cells of a living organism, resulting in an alteration of cell or tissue function in a manner that corrects or treats a pathological process.

Gene Transfer Vector
Reagent used to deliver genetic material to the interior of cells.

Herpes Simplex Virus
An enveloped neurotropic double-stranded DNA virus that may be modified to produce a gene transfer vector.

***In vivo* Gene Transfer**
Transfer of genetic material to cells by introduction of a vector into the body of a living organism.

Lentivirus
An enveloped single-stranded RNA virus that may be modified to produce a gene transfer vector.

Nonviral Vector
Gene transfer vector that does not contain viral components.

Replication defective
A virus that has been modified to prevent its replication in target cells following gene transfer.

Tropism
The natural host cell range of a vector.

Viral Vector
Gene transfer vector made by genetic engineering of a virus.

Gene therapy may be defined as *the introduction of genetic sequences into cells of a living organism, resulting in an alteration of cell or tissue function in a manner that corrects or treats a pathological process.*

This might be accomplished in one of two broad ways: (1) The therapeutic genetic material is transported *ex vivo* into cells taken from a patient – the cells can be subsequently reintroduced, with or without expansion, as an homologous cell transplant. The reintroduced cells engraft and produce a therapeutic product that can alter the biology of the transduced cell to correct a genetic abnormality or deficiency, or may serve as a factory for production of a protein product used by other tissues exposed to the circulation. (2) The therapeutic genetic material is introduced into the cells of the body by direct *in vivo* gene transfer. Direct *in vivo* gene transfer may be used if procurement of appropriate cell explants requires unacceptably invasive procedures, or for gene transfer to essentially nonregenerating tissues (e.g. brain or heart).

The genetic material transferred (''transgene'') may encode a complete therapeutic gene to replace a defective function, or coding sequences whose products interfere with an acquired or inherited pathologic cellular function. Gene transfer may also be used to enhance normal function to achieve a therapeutic effect, such as the induction of an immune response or the reversal of autoimmune activity. Finally, a large effort in gene therapy has been directed toward the use of gene transfer to destroy unwanted tissue, such as malignant cells, or fight infectious agents, such as human immune deficiency virus (HIV).

The diverse potential gene therapy applications suggest that genes as molecular medicines hold great promise for the future. In part this promise is beginning to be realized, but much more research (and testing in patients) is required before gene therapy becomes standard medical practice.

1 Introduction

In this article, we discuss the science and the current status of gene therapy. This includes a review of methods used for gene transfer, current attempts to apply gene therapy to treat disease, and a discussion of serious adverse reactions to gene therapy. One example of gene therapy will be described in more detail in order to better acquaint the reader with the steps needed to develop this type of therapy. Finally, an attempt will be made to make some projections regarding the future of gene therapy research and practice by describing where the field may be headed in the next 10 years.

2 The Science of Gene Transfer

A major effort in gene therapy research has concerned the development of efficient, safe methods to introduce genetic information into cells either outside or within the body. These gene delivery tools are referred to as *gene vectors* and can be constructed from modified viruses (''viral vectors'') or directly from chemical constructs (''nonviral vectors'').

The advent of gene cloning and the ability to make recombinant viruses in the early 1980s portended an era in which viral gene vectors would become the working tools for gene therapists (Table 1). All that was needed was the removal of a critical gene from the virus that allowed it to grow or cause disease (Table 2) and replace that gene with one that is useful to the cell. Gene therapy researchers soon found that reporter genes like the *Escherichia coli lac*Z gene encoding β-galactosidase could be introduced into almost any tissue by various recombinant viruses, and β-gal activity detected using a colorigenic agent. Depending on the target cell, a nonintegrating or integrating vector could be used to introduce the reporter gene. Curing a disease using this technology should have been a simple matter of replacing the reporter transgene with a therapeutic gene. So why did this clear path to success fail to rapidly yield new therapeutic products? To answer this question, it is important to understand the technology of virus-mediated gene transfer and the nature of clinical studies targeted by this approach to treatment.

Why choose viruses? It has been known for many years that viruses can deliver their genes to cells by the process of virus infection. Depending on the class of virus, a variety of highly evolved mechanisms for redirecting cellular processes for making new virus particles come into play. Viruses are essentially inert particles outside the body and are vulnerable to the outside world, making their expectant viability short-lived, in the absence of available hosts. The life cycle of viruses, in the simplest terms, involve virus attachment to a susceptible cell, entry into the cell, capsid removal in a process referred to as *uncoating* to expose the virus genome in an active form, either in the cytoplasm or nucleus. The virus genes are transcribed and expressed as proteins or active RNA molecules, requiring at least partial use of the cellular transcriptional and translational machinery. The most sophisticated viruses have hundreds of genes (e.g. vaccinia), whereas the most simple have as few as four to six genes (e.g. polyoma virus and adeno-associated virus). The viral genome, single- or double-stranded DNA or RNA, is replicated and encapsidated into new virus particles that are released by cell lysis or by a

Tab. 1 Major viruses used to construct viral vectors.

	Genome	*Genomic integration?*	*Envelope?*	*Latent infection?*	*Helper virus?*	*Cellular receptor*	*Route of cellular entry*	*Disease caused by wild-type virus*
Lentivirus	ssRNA; 9.5 kb	Yes	Yes	Yes	No	CD4, CXCR5 ([a]phosphatidyl serine)	Plasmalemma ([a]endosome)	AIDS
Adenovirus	dsDNA; 35 kb	No	No	No	No	CAR, integrin $\alpha v\beta 5$	Endosome	Respiratory infection; conjunctivitis; gastroenteritis
Adeno-associated virus	ssDNA; 5 kb	Yes	No	Yes	Yes	Integrin $\alpha v\beta 5$, FGF-1R	Endosome	None known
Herpes simplex virus	dsDNA; 150 kb	No	Yes	Yes	No	Glycosaminoglycan, nectin1α and HVEM	Plasmalemma	Cold sore; rarely encephalitis

[a]Although wild-type HIV enters cells through the plasmalemma, vectors pseudotyped with the vesicular stomatitis virus G-spike glycoprotein enter through the endocytotic pathway.

Tab. 2 Properties of some viral vectors.

	Vector	*Transgene insert capacity [kb]*	*Genes removed from vector genome to prevent replication/toxicity*	*Complementation of essential functions*	*Easy prepara-tion of high titers and pure stock?*	*Easy insertion of transgene sequences?*	*Toxicity*
Lentivirus	HIV-based	6	All except 5′ portion of gag	Transient plasmid transfection	No	Yes	Insertional mutagenesis
Adenovirus	E1, E3 deleted Ad5	8	E1, E3	Complementing cell line	Yes	No	Inflammation
	'Gutless'	30	All	Complementing cell line and helper virus	No	No	Inflammation
Adeno-associated virus	AAV2-based	4.5	All	Helper virus or transient plasmid transfection	No	Yes	Insertional mutagenesis
Herpes simplex virus	Genomic HSV-1 vectors	30–40	ICP4, ICP27, ICP22, ICP47, ICP0, U_L41	Complementing cell line	Yes	No	Minimal
	Amplicon	120	All	Transient BAC transfection	No	No	Minimal

process called *budding* in which the virus capsid is enclosed in a membrane or envelope. These nascent particles are then competent to invade other cells. For details of viral replication and life cycles, see excellent and exhaustive chapters in Field's Virology. Summary figures of the life cycles of major viruses used as vectors are shown below.

Viruses have evolved sophisticated mechanisms for either rapid spread or the ability to remain relatively quiescent in cells of the body until an opportunity presents itself for transmission to a naïve host. Many viruses (e.g. adenovirus) do not persist in the host but rather rapidly replicate and produce large numbers of infectious particles that can be spread to other susceptible hosts. Vectors based on such viruses are less useful for long-term gene delivery but can be used for applications that require transient transgene expression, such as the induction of immunity or destruction of tumors. The life cycle of other viruses (e.g. retroviruses and lentiviruses) involves a persistent phase in which the virus genome may become integrated into a host chromosome and can thus be passed on to daughter cells following cell division. Vectors created from these viruses can be useful for applications in which persistent transgene expression is required. Herpes viruses persist in the host as nonintegrated forms that either remain associated with nondividing cells or can replicate their genetic information in synchrony with cell division. Most viruses that persist in the host have evolved sophisticated means for transiently avoiding host detection by the immune system, often reaching a stalemate between their elimination by the immune system and their disease-causing activity. Most human infecting viruses that persist do so by a system referred to as *latency* where few, if any, viral genes are expressed in cells where latency is established and virus persistence generally lasts for the life of the host.

Early studies of viral vector-mediated gene transfer for therapeutic purposes have been focused on a few viruses (Tables 1 and 2). The choice of viral vector for a specific application depends on the biology of the wild-type virus and the degree to which the virus life cycle and genes have been characterized. As more insight is gained into virus biology, it is likely that additional viral gene transfer vectors will be developed.

3 Principles of Vector Engineering

3.1 Retroviruses

It has been known for some time that retroviruses can integrate their genomes into the host cell by a process of reverse transcription, using reverse transcriptase to allow the RNA tumor virus genome to become integrated into the DNA genome of a cell. Virus entry into the nucleus is largely dependent on cell division and thus these viruses most efficiently infect dividing cells.

One of the best-studied retroviruses is the mouse Moloney leukemia virus. It was discovered in the 1980s that "packaging" cell lines could be engineered to express the virus coat protein and envelope protein constitutively. On transfection of these packaging cells with a DNA copy of the virus, infectious packaged particles were produced, even though the vector genome contained only terminal repeat elements and the packaging signal from the original virus – the internal sequences

having been replaced by nonviral genes. These infectious packaged particles could be used to transduce other cells, and by virtue of their terminal repeat elements, the nonviral genes were incorporated into cellular DNA. Following successful transfer of the adenosine deaminase (*ADA*) gene in preclinical studies, this first virus gene transfer vector was used in the first gene transfer clinical trials in the early 1990s in an attempt to treat the ADA-deficiency form of severe combined immunodeficiency (SCID). Gene transfer with these first-generation vectors was not efficient enough to produce a therapeutic outcome, and indeed the study design almost ensured failure. An initially unrecognized shortcoming of retroviruses and lentiviruses is the propensity for the vector genome to integrate into the cell genome, preferentially near sites of active transcription. This process can lead to rare insertional mutagenesis events and activation of cellular genes that have oncogenic potential. In a more recent clinical trial using a retroviral vector to treat X-lined SCID, two of the patients cured of the disease by gene transfer developed leukemic transformation as a result of insertional mutagenesis (see Sect. 5.2). Investigators are now attempting to engineer retroviral vectors, which are either directed in their integration activity or lack the ability to activate cellular genes.

3.2 Lentiviruses

Vectors constructed from modified lentiviruses, a group of RNA viruses closely related to the oncogenic retroviruses (Figs. 1 and 2), have come to the forefront in the last five years because of their efficiency of gene transduction, their larger capacity for foreign DNA, and their ability to infect nondividing cells. Engineering human lentiviruses to create a vector requires retention of the natural integrase function part of the nucleocapsid and the viral polymerase (Fig. 3). Methods have been developed in which elements of lentiviruses, the Moloney virus, and the envelope components of vesicular stomatitis virus (VSV-G) have been incorporated into three expression plasmids, which can be transfected into cells for vector production (Fig. 4). Transfection methods are not as efficient as using packaging cell lines for retroviral vectors, but lentiviral vectors can be concentrated by centrifugation since the VSV-G envelope component is stable. Another advantage of VSV-G is that it utilizes a common cell surface proteoglycan, which is ubiquitously expressed

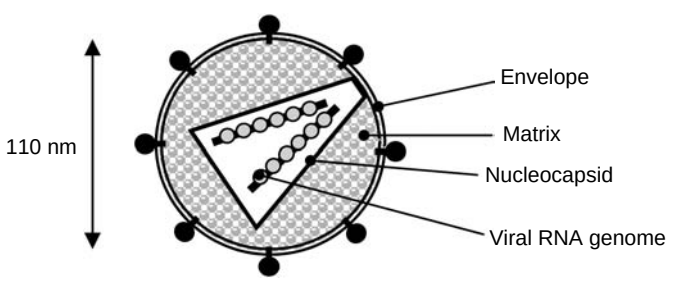

Fig. 1 The structure of Human Immunodeficiency Virus – A schematic depiction of a mature HIV-1 virion is shown, to illustrate key structural components of this prototypical lentivirus.

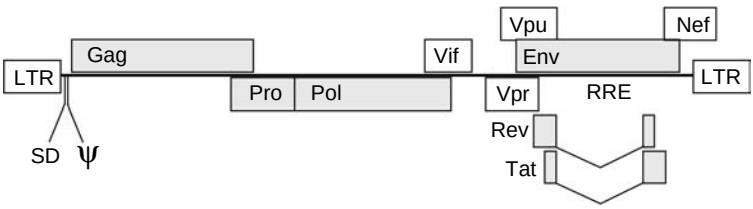

Fig. 2 The HIV-1 genome – The HIV-1 RNA genome is illustrated diagrammatically (not to scale) to demonstrate overlapping genomic locations and splicing of viral genes. Abbreviations: LTR, long terminal repeat (containing promoter for viral transcripts); SD, major splice donor; Ψ, packaging signal; RRE, rev-response element. The names of viral genes are shown within the boxes. *gag* encodes the core antigens, *pol/pro* encodes the reverse transcriptase and protease enzymes necessary for genomic integration and posttranslational processing of viral proteins respectively, and *env* encodes the viral envelope proteins necessary for cell entry. *tat* encodes a transactivator of viral transcription and *rev* encodes a splicing/export regulator – see Fig. 3.

on many cells as a receptor, and thus the host range for the lentiviral vector is greatly broadened. It should be noted that despite the use of human immune deficiency virus (HIV) components for many applications, this vector does not cause AIDS or other diseases. Other lentiviruses may also be suitable for the construction of lentiviral-based vectors. Nevertheless, lentiviral vectors have the same propensity for random integration that has been observed with retroviral vectors, although chromosomal breaks are less targeted by lentivirus than by oncogenic retrovirus since the lentivirus has the ability to cut and religate double-stranded DNA as part of the enzyme systems that are contained within the particle. Lentiviral vectors have proven especially useful in transduction of brain and bone marrow stem cells. They have yet to be tested clinically, but this is likely to occur within the next few years.

3.3 Adenoviruses

After retroviruses, adenoviruses have been used most often for gene transfer (Fig. 5). Replication-defective viruses lacking the tumor-causing gene of adenovirus, *E1a*, had been made in the early 1980s and propagated in cell lines that contained the *E1a* gene integrated into the cell chromosome (see Figs. 6, 7). Such cell lines were designated as complementing cells since the missing viral function was supplied *in trans* on infection. The deletion of E1a prevented the virus from replicating in most wild-type cells (Fig. 7), although there was leaky expression of the other viral functions, including structural proteins and the viral polymerase. Transduction of a number of target tissues in experimental models was demonstrated using early adenoviral (AdV) vectors. Deletion and complementation of further adenoviral genes enabled generation of more efficient and less toxic vectors. One attractive property of AdV vectors is their remarkable ability to express transgenes at very high levels. Because the wild-type virus causes a respiratory infection, it was initially thought that Ad vectors would prove to be ideal for gene transfer to lung, for example, in the common inherited respiratory disease, cystic fibrosis. Unfortunately,

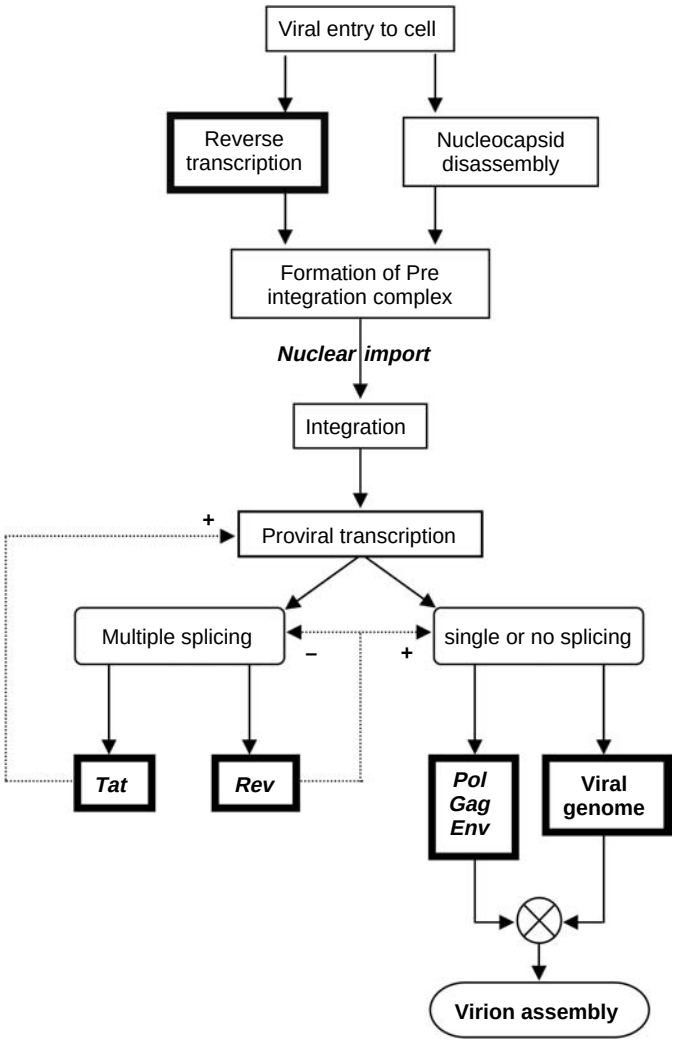

Fig. 3 Major events in the life cycle of lentiviruses – The early part of the life cycle is depicted in the upper portion of the diagram; entry, reverse transcription, formation of the preintegration complex (PIC) and nuclear import of the PIC culminate in the integration of proviral DNA into the host cell genome. The late part of the cycle is characterized by transcription of viral genes. All of the viral genes are encoded by a single primary transcript, and the expression of different proteins is controlled by regulation of RNA splicing and posttranslational protein processing. Initially, multiply spliced RNA isoforms encode regulatory proteins (*tat* and *rev*) that *trans*activate proviral transcription, and drive a shift toward the production of unspliced or singly spliced RNA. The latter encodes the virion structural proteins and forms the viral genome, enabling assembly of virions. Vectors are constructed by deleting all viral genes from the genome contained in the particle, thereby disrupting replication at multiple levels – see Fig. 4.

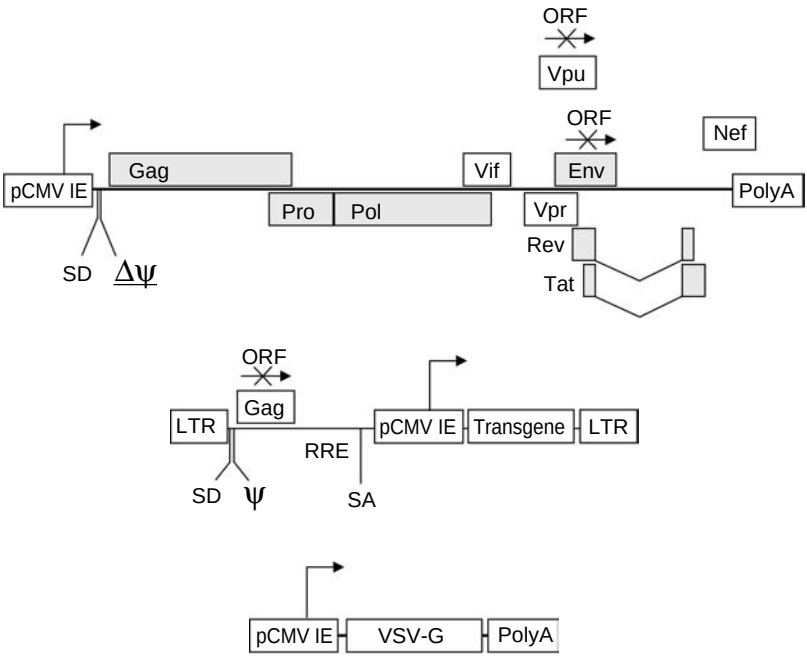

Fig. 4 Construction of replication-defective lentivirus vectors – The system described by Naldini et al. is illustrated. Replication-defective lentiviral vectors may be generated by cotransfection of mammalian cells with the three plasmids shown. A. The ***packaging construct*** provides all of the genes that encode viral structural proteins. Thus, gag-pol, rev, and tat are expressed from this construct. The genes encoding these essential proteins are not incorporated into the virion, however, as the viral LTRs and packaging signal (ψ) are deleted from this vector. In addition, the accessory gene *Vpu* and envelope gene contain frame-shift mutations, abolishing expression of the relevant proteins. B. The ***transfer vector*** contains an expression cassette for the therapeutic or experimental transgene, in addition to the viral LTRs and packaging signals. This vector sequence is transcribed, and the resulting RNA is packaged into lentiviral particles generated from the structural proteins encoded by plasmid A. As the transfer vector contains no sequences encoding functional viral genes (the *gag* gene is truncated and frame shifted), the resulting vectors are replication defective. C. The ***envelope plasmid*** contains an expression cassette encoding a viral envelope protein to replace that encoded by the lentiviral *env* gene, which is deleted from the packaging construct. The G-spike envelope glycoprotein from Vesicular Stomatitis Virus is commonly used to pseudotype lentiviral vectors.

AdV vectors have not proven useful for this application for two major reasons. First, adenoviral receptors are not present on the apical (airway) surface of airway cells, and so intrabronchial or aerosol application of the vector does not give rise to efficient infection of the appropriate target cell required for cystic fibrosis transmembrane conductance regulator (CFTR) expression in the lung. Second, innate and acquired immune response to vector enhanced by the leaky gene expression in the first-generation Ad vector constructs made them susceptible to immune elimination. The immune response to Ad vectors has proven to be a difficult problem for many

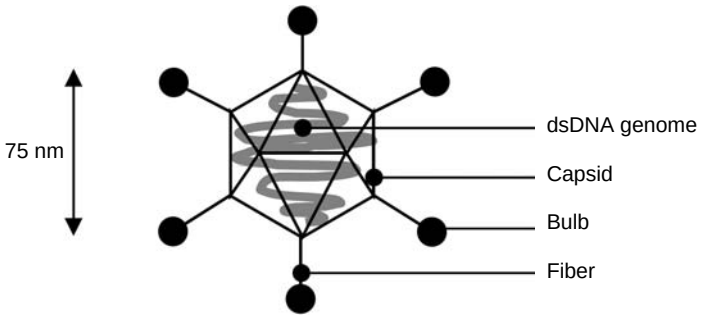

Fig. 5 The structure of Adenovirus – A schematic depiction of a mature adenovirus virion is shown to illustrate key structural components of the virus.

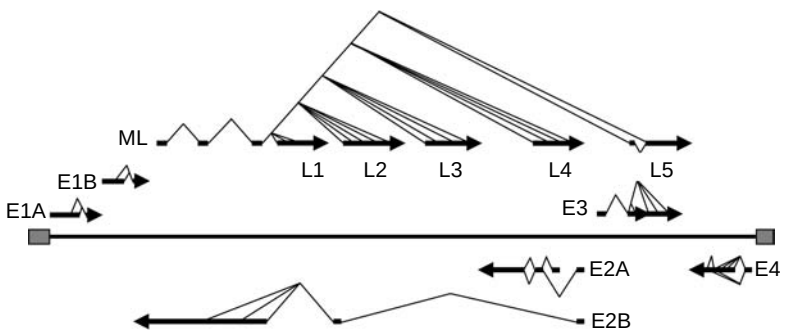

Fig. 6 The adenovirus genome – The adenoviral genome consists of 35 kb of dsDNA, in which the viral gene-encoding sequence is enclosed by two inverted terminal repeats (grey boxes). Although there are only eight basic transcriptional units, approximately 40 transcripts are produced by a bewildering array of alternative splicing events. In particular, the major late promoter gives rise to transcripts that share a tripartite leader sequence, culminating in a splice boundary with one of a possible 18 coding sequences. These are grouped, according to the polyA signal at which the transcripts terminate, into L1, L2, and so on. Viral transcription depends on the presence of the viral gene E1A (see Fig. 7); first-generation vectors were deleted for E1A and E1B and the genes supplied *in trans* using special cell lines. Later vectors are deleted for other genes, for example E3, in addition to E1. The so-called "gutless" vectors contain only the packaging signals and transgene DNA, the viral genome being supplied *in trans* by use of a helper virus to express structural proteins, and thus allow the generation of virion-like particles. These vectors can accommodate up to 30 kb of transgene sequence, but are difficult to prepare in high titer and purity.

proposed applications. Alterations to the vector, which eliminate all of the viral components from the vector genome, improve vector stability *in vivo*, but this is not sufficient to eliminate immune rejection of AdV-infected cells. This was most dramatically demonstrated by the finding that AdV inoculation of liver in immune-competent mice results in high levels of hepatocyte transduction, but expression is lost.

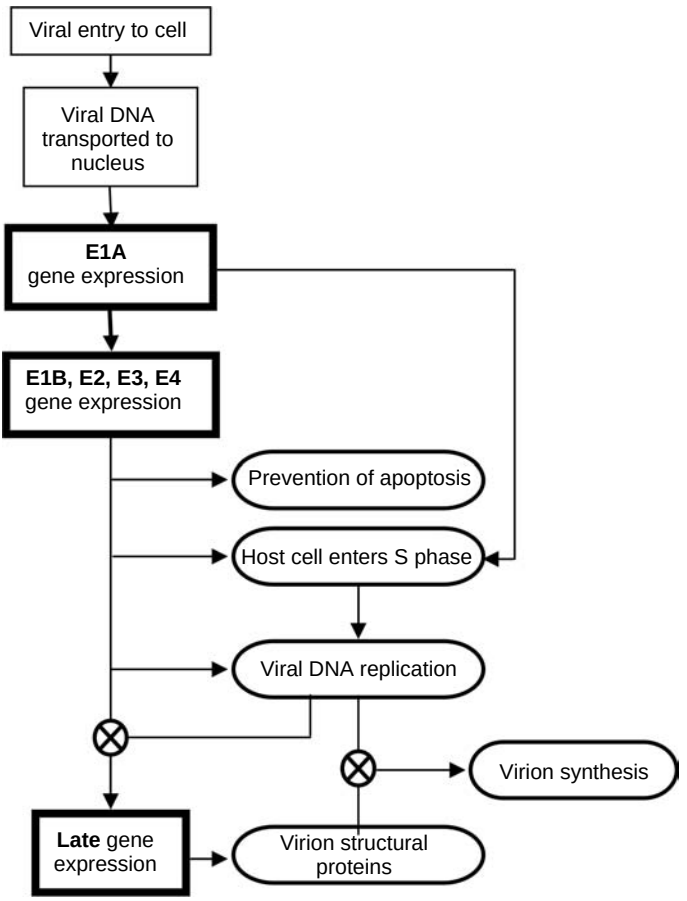

Fig. 7 Major events in the life cycle of adenovirus – The major events in regulation of viral gene expression are shown. Following entry to the nucleus, expression of the viral gene *E1A* is essential to allow expression of the remaining components of the viral genome. The gene expression program of adenovirus results in the subversion of a number of cellular processes, facilitating the production of progeny virions. There is no latent infection and the result of infection is cell lysis and release of adenoviral particles. Deletion of the E1 region halts the viral transcriptional program, and the resulting particles can be used as gene transfer vectors. Later versions of these vectors have multiple genes deleted to enhance the cloning capacity of the vectors and reduce cytotoxicity resulting from leaky expression of the remaining viral genes.

In contrast, the use of immunodeficient mice in the same experiments resulted in long-term gene expression. In addition to immune rejection, AdV vectors cannot be administered to patients with high levels of circulating antibody, and the vector cannot be readily repeat dosed in patients, even those without preexisting immunity, since vector inoculation results in high-titer antibody production. There have been

attempts to protect AdV from immune recognition by coating the particle and AdVs of alternative serotypes or from other species or have been isolated with different coat protein structures in the hope that these particles will avoid immune recognition. The AdV particle itself, however, has the ability to induce cytokines (for example interleukin-6) that can result in elimination of the virus genome. A liver gene transfer clinical trial, using high-titer AdV, led to the death of one patient as a result of a vector-induced cytokine storm, leading to organ failure (see Sect. 5.1). There are two important uses of AdV vectors that are likely to be effective in the future. The first is the use of AdVs carrying immune-activating genes in the treatment of solid tumors, and the second is the use of recombinant AdVs for vaccination against newly arising infectious agents. Both applications take advantage of high-level transgene expression, efficient gene delivery, and immunogenicity of the vector substrate.

3.4 Adeno-associated Viruses

Adeno-associated virus (AAV) was initially discovered as a contaminant of adenoviral preparations (Figs. 8, 9 and 10). AAV-based vectors have enjoyed a broad appeal following their initial development. The reasons for this include the following: (1) AAV causes no known disease; (2) the AAV genome can be manipulated as a simple plasmid; (3) the small size of AAV allows the particle to infiltrate many different tissues; (4) certain tissues (e.g. muscle) are transduced very efficiently. Although its small size (22 nm) does not allow packaging of more than 5 kb of foreign sequence, this is sufficient to accommodate most gene coding sequences as cDNAs (i.e. the coding sequence of a gene with the introns and flanking regions deleted). *In vitro*, wild-type AAV integrates into a specific location on chromosome 19q, but this specific integration event is dependent on *replicase*, a viral gene that must be eliminated from vectors because of its considerable toxicity to cells. AAV replication requires several helper functions that can be provided by either adenovirus or herpes simplex virus. Like adenovirus, AAV is highly susceptible to neutralization and cannot be easily repeat dosed, but AAV vector-mediated transgene expression in some tissues provides for long-term expression; the lack of induction of immunity to the transgene product may reflect low infectivity in antigen-producing cells, such as the dendritic cells located in the skin. The optimal method for producing AAV is still

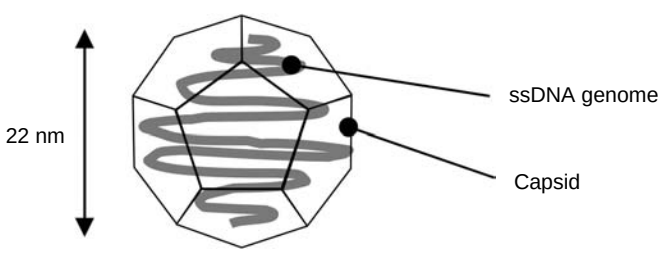

Fig. 8 The structure of adeno-associated virus – A schematic depiction of a mature adeno-associated virus virion is shown, to illustrate key structural components of the virus.

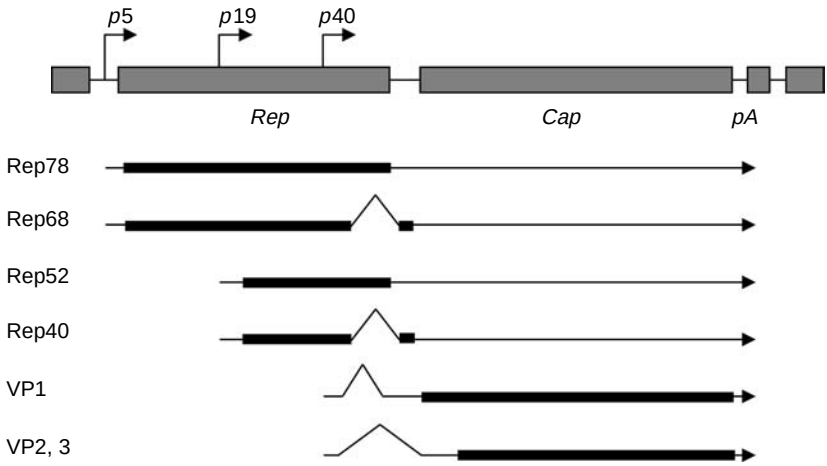

Fig. 9 The adeno-associated virus genome – The AAV genome consists of 5 kb of ssDNA. The coding sequences are flanked by inverted terminal repeats (grey boxes). There are two major genes, *rep* (encoding proteins necessary for gene expression, genomic integration, and replication) and *cap* (encoding the structural components of the virion). There are three viral promoters giving rise to six transcripts by alternative splicing. All of the transcripts terminate at a single polyA signal. Seven proteins are encoded by the six transcripts because the VP2,3 transcript can be translated using one of two different translational start codons. Gene transfer vectors are generated by deletion of *rep* and *cap*, supplying these genes *in trans* by cotransfection of a plasmid, to allow generation of particles. Unfortunately, the products of the *rep* gene are toxic to host cells, and so packaging cell lines have not been possible to generate.

under development, but the most common approach is a triple-plasmid transfection method in which one plasmid is the vector, another expresses the replicase and capsid functions, and the final plasmid provides the adenovirus helper functions. This method produces AAV vector particles without contamination from helper virus since the adenovirus packaging signals have been removed, but is not very efficient. The scale-up manufacture of AAV is also complicated by the remarkably low transduction efficiency of AAV, compared with other vectors. Transduction with AAV *in vivo* thus requires very high doses of virus per cell. Although AAV may integrate into the host genome during its normal life cycle, this is inefficient and cell division is usually required. In contrast to other vectors, AAV appears to require concatemerization in the nucleus for expression, a process that can take weeks for maximum transgene expression. Unless the efficiency of AAV transduction can be improved, its utility as a vector for the treatment of patients will be limited; manufacturing considerations will preclude treatment of more than a few patients at the required doses of $>10^{12}$ per patient. High vector dosing also risks provoking toxic immune responses. Despite these concerns, AAV remains promising. For certain applications, such as gene transfer to the heart, it may prove to be the vector of choice. In addition, new serotypes are proving to be more efficient in some tissues such as liver where previous vectors based on AAV type 2, for

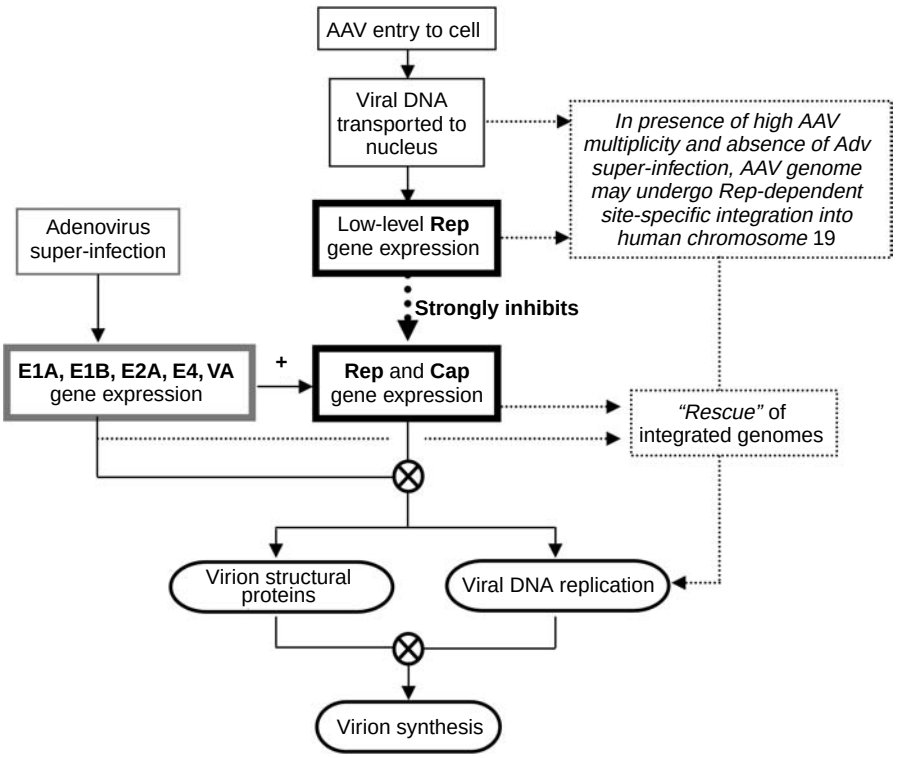

Fig. 10 Major events in the life cycle of adeno-associated virus – Following entry of AAV into the cell, one of two chains of events may ensue the following: 1. In the absence of super infection with a helper virus (AdV is shown, but HSV can also provide helper functions), concatemerized AAV genomes may undergo *rep*-dependent site-specific integration into human chromosome 19, giving rise to a latent infection. Alternatively, the viral genome may persist in the cell nucleus. Low-level expression of rep proteins is strongly inhibitory to expression of viral transcripts. 2. Infection of an AAV-harboring cell with adenovirus, or transfection with a plasmid containing the minimal essential AdV helper functions (E1A, E1B, E2A, E4, and VA) allows high-level expression of *rep* and *cap*, and also forces the cell to enter the S-phase of the cell cycle. This allows replication of AAV viral DNA, synthesis of virion components and assembly of AAV virions. High-level expression of *rep* and *cap*, in association with AdV helper functions can rescue integrated genomes from the chromosome of a latently infected cell. Replication-defective vectors are generated by deletion of *rep* and *cap* from the AAV genome. The vector particles are prepared by cotransfection of packaging cells with a *rep/cap* plasmid, an AdV helper function plasmid, and a gene transfer plasmid containing the transgene of interest and the AAV packaging signals. As the resulting particles are deficient in the *rep* gene, integration into the host chromosome is very inefficient and occurs randomly, but with a predilection for transcriptionally active genes.

example, were considerably less effective. Perhaps, when increased manufacturing and transduction efficiency is achieved, opportunities for engineering better vectors may arise.

3.5 Herpes Simplex Virus

The last vector under serious consideration for human studies is Herpes simplex virus

(HSV) type 1 (Figs. 11, 12, 13 and 14). The wild-type virus causes the common cold sore, and is transmitted through the skin by direct contact with a lesion from an infected individual. The wild-type virus can replicate in skin and mucous membranes but does not persist in the epithelium, instead entering sensory nerve terminals in the skin from which they are transported in a retrograde manner to the nerve cell body in sensory ganglia (Fig. 13). In the sensory ganglia, the lytic viral life cycle is curtailed within days, after which the wild-type virus enters a latent state, with concomitant silencing of lytic genes. During latency, only the latency locus is transcriptionally active, and latently infected nuclei accumulate a stable intron called the *latency-associated transcript* or LAT RNA. The ability to persist in neurons for life suggests that HSV may be engineered for gene transfer to nerves and that vector delivery can result from simple skin inoculation. The general strategy for engineering HSV vectors from the wild-type virus is to eliminate genes that contribute to virus replication and reactivation from latency while taking advantage

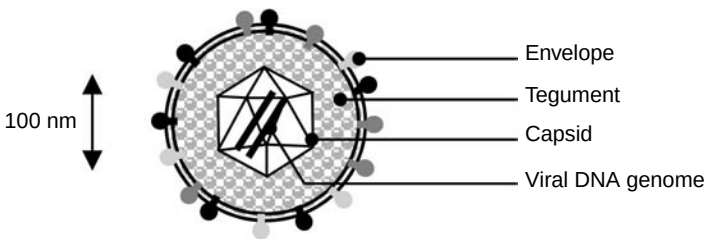

Fig. 11 The structure of herpes simplex virus – A schematic depiction of a mature HSV-1 virion is shown to illustrate the key structural components of the viral particle.

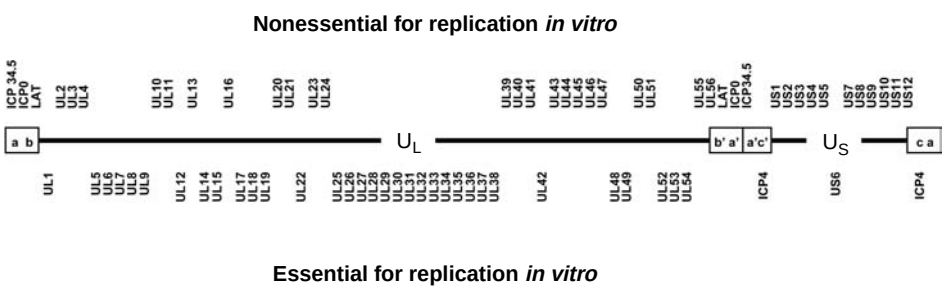

Fig. 12 The HSV-1 genome – The HSV-1 genome is illustrated diagrammatically. The genome is 154 kb and encodes 84 genes. The genome is divided into unique long and short segments flanked by repeat sequences. Genes that are essential and nonessential for viral replication *in vitro* are indicated. Nonessential genes (which encode a variety of functions that optimize the interaction of HSV with the host organism) can be removed from the virus without compromising replication *in vitro*, allowing their replacement with transgenes of interest. The capacity for the insertion of foreign DNA sequences into the HSV genome is large; genomic vectors can accommodate around 40 kb of transgenic material and amplicons (particles that have all viral genes removed and supplied in trans) can accommodate 120 kb of foreign sequence.

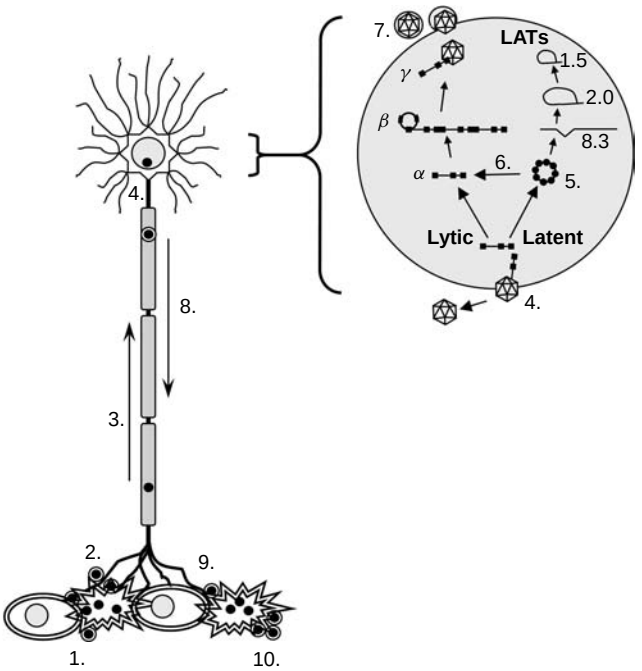

Fig. 13 The life cycle of HSV-1 *in vivo* – The key events occurring during infection of a human host are depicted schematically as follows: 1 – Lytic cycle of replication at epithelial port of entry; 2 – Virions released from epithelia enter sensory nerve terminals; 3 – Nucleocapsid and tegument undergoes retrograde axonal transport to soma; 4 – Viral DNA enters neuronal nucleus and either initiates lytic cascade of gene expression or becomes latent; 5 – During latency, viral genome remains episomal and nuclear. Only the LAT genes are expressed; 6 – Immunosuppression, intercurrent illness, or other stimulus ''reactivates'' lytic infection; 7 – Virions formed by budding from nuclear membrane; 8 – Nucleocapsid and glycoproteins transported separately by anterograde axonal transport; 9 – Virion assembly and egress from nerve terminal; 10 – Recurrent epithelial infection at or near the site of primary lesion.

of the latent genome as a platform for transgene expression (Fig. 14). The virus depends on expression of four genes, activated upon entry into the nucleus and referred to as *immediate early functions* for replication to occur. In the absence of these genes, other viral functions fail to be expressed or are expressed at exceedingly low levels. Thus, despite the fact that the virus contains approximately 84 genes (Fig. 12), engineering HSV vectors is not as complex as it may seem. Two of the immediate early genes are essential functions, but complementation of all four genes is required for high virus vector yield on production in complementing cells. The latency promoter elements have been partially characterized and can be used to drive the transgene expression for considerable time periods in sensory nerves. While

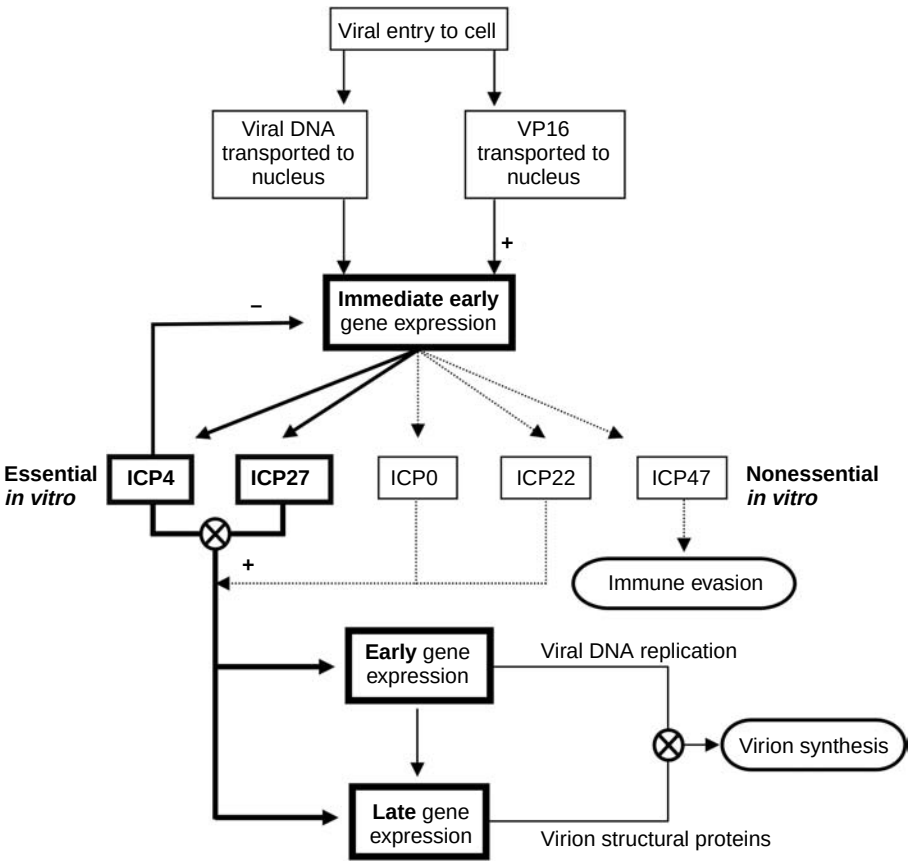

Fig. 14 Regulation of HSV-1 gene expression during lytic infection – This flow chart illustrates the important regulatory events in gene expression occurring during lytic HSV-1 infection. In order to proceed to the later stages of infection, during which the viral genome is replicated and new virions assembled, both of the essential immediate-early genes *ICP4* and *ICP27* must be expressed. Deletion of one or other essential *IE* gene results in a replication-defective virus by eliminating early and late gene expression; these vectors are easily propagated and prepared to high titer and purity in cell lines expressing ICP4 and ICP27. Later vectors have multiple *IE* genes deleted to reduce cytotoxicity. In addition, many of the large number of genes that are nonessential for replication *in vitro* can be deleted to allow insertion of multiple or large transgene sequences. Herpes amplicon genomes contain the HSV origin of replication and packaging signals, in addition too the transgene of interest, but do not contain any viral coding sequences. A BAC containing the HSV genome expresses all viral structural genes in a cotransfection preparation, to allow generation of particles. Although the amplicon system may be thought advantageous owing to the absence of viral genes and large capacity for insertion of transgenic material, (120 kb), amplicons are very difficult to prepare in useful amounts and current technology strongly favors the use of replication-defective genomic vectors.

multiple essential viral genes have been complemented in cell lines engineered for this purpose, a packaging cell line has not yet been produced that takes full advantage of the large genome for delivering foreign DNA. This may be difficult since a number of the virus functions are toxic to cells and their expression requires coordinated regulation, a feat difficult to accomplish using virus genes embedded in cellular chromosomes. Nevertheless, efforts are ongoing to eliminate more virus functions, creating additional room for foreign sequences. The large genome of HSV of 152 kb in length suggests its potential utility for transferring multiple genes and perhaps even genomic sequences. Current vectors can accommodate up to 50 kb.

Another class of HSV-based vectors is referred to as *amplicons*. These are plasmids that mimic defective interfering particles that arise spontaneously during high multiplicity passage of the virus in cell culture. Amplicons can be generated with wild-type or conditional mutant virus as a helper, provided the amplicon contains a packaging signal and an origin of replication. As the viral genes are supplied *in trans*, the capacity of insertion of cloned DNA is large. Unfortunately, the *Ori* signal increases the recombination frequency, and it is therefore difficult to avoid recombination between the amplicon and helper virus, so that completely pure vector stocks are difficult to obtain. This problem can be ameliorated to a substantial degree by using helper plasmids for amplicon packaging, but virus reconstitution may occur. Manufacture of amplicons is a difficult task for use in patients, since consistent preparations are difficult to achieve. Thus, vectors that are produced by complementation are more likely to be developed as gene therapy products, at least with current technology. As is the case with AdV and AAV-based vectors, a substantial proportion of the population has been infected with HSV and has antibodies against the virus. However, antiviral immunity is not effective in preventing HSV infection, rather it is the extent of virus replication and frequency of reactivation that is controlled by the immune response. Immune competent individuals can have HSV reactivation events, but this frequency is considerably greater in immune-compromised individuals. Reactivation, of course, does not occur with replication-defective vectors.

3.6 Vector Targeting

Over the past decade, substantial emphasis has been placed on manipulating the host range of viral vectors. Targeting could be used (1) to transduce cells that are not naturally infected by the wild-type virus or (2) to increase the uptake of vector into a limited subset of cells. In the best-case scenario, a vector could be engineered to be defective in its natural targeting function, but redirected to recognize an epitope on a desired target cell, allowing systemic vector inoculation, to achieve distributed, yet targeted, transduction. Enveloped viruses have been retargeted by using different envelope glycoproteins from other enveloped viruses. In addition, novel ligands have been inserted into the coding sequence of specific virus envelope genes. However, virus binding to target cells has not always led to infection, since a second fusion step is required either at the cell surface or within the endosome compartment (Figs. 15, 16). Recent studies with AAV have discovered sequences that can be modified without disrupting virus uncoating. Although efficient retargeting has

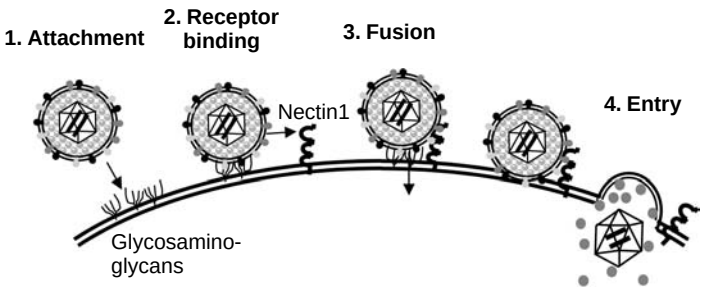

Fig. 15 Cell entry of an enveloped virus – HSV-1 is shown for illustration. Following an initial attachment event (involving interaction of HSV glycoproteins gC and gB with cell surface glycosaminoglycans), a specific receptor interaction occurs between viral glycoprotein gD and cell surface nectin-1 (also called HveC). This results in fusion of viral envelope with the cell membrane, a process dependent on viral glycoproteins gB and gH/L. The nucleocapsid and tegument proteins then enter the cytoplasm, leaving the viral glycoproteins on the cell surface. Subsequently, the nucleocapsid and some of the tegument proteins are transported to the cell nucleus. The viral DNA is uncoated and enters the nucleus with some of the tegument proteins, initiating viral gene transcription (see Fig. 14).

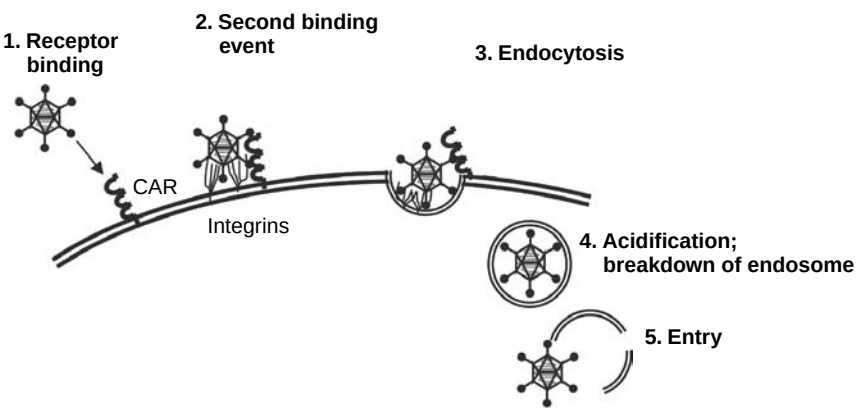

Fig. 16 Adenovirus is shown for illustration. Following an initial receptor binding event, involving interaction of the terminal bulb of the virus fibre protein with the coxsackie-adenovirus receptor (CAR), a second receptor interaction occurs between the AdV penton base protein and cell surface integrins. This results in the endocytosis of the receptor and virus. As the endosome matures, and its environment acidifies, viral components cause the endosome to break down, releasing the partially degraded viral particle into the cell. The nucleocapsid is then transported to the nucleus, where uncoating and nuclear entry of the viral genome are followed by viral gene expression.

not yet been achieved with AAV vectors, enhancement of entry into cells that express low levels of AAV receptor has been achieved by modification of the capsid to display a cell-specific receptor on the AAV virion. The most effective retargeting of viral vectors has involved the modification of coat protein components of adenovirus. Adenovirus infects cells by binding to the coxsackie/adenovirus receptor (CAR) using the extended knob structure of the hexon capsid protein at the particle surface. The virus penton base binds to a specific integrin, $\alpha V\beta 5$, using a specific motif arginine-glutamate-aspartate (RGD). This dual receptor binding recognition is typical of many viruses and ensures specificity. The adenovirus is endocytosed and is released from endosome as it acidifies. A number of strategies have been used to retarget adenoviruses. These include extension of the hexon, using an antibody ligand or a receptor ligand, or engineering the knob function directly by genetic engineering. In addition, the RGD amino acid motif has been altered, so that different integrins can be recognized. This combination has specifically altered receptor interactions that lead to adenovirus infection, in a quantitatively significant manner. Such retargeting holds promise for safe, directed *in vivo* use, especially for targeting cancer cells involved in metastasis.

3.7 Regulation of Gene Expression

If gene delivery cannot be restricted to the cell of interest, targeting of gene expression can be attempted using promoters with tissue specificity. It should be noted that promoter function in the background of a virus genome may be unpredictable. In addition, some tissue-specific promoters are complex, requiring large amounts of sequence (including elements located within introns) to be active, and most vector systems cannot accommodate such large sequences. Considerable effort has gone into determining the critical motifs for promoter specificity and a variety of interesting sequences have been characterized, including locus control regions, matrix attachment functions, transcription factor recognition sequences, and insulator elements. Transcriptional control has proven to be a complex problem to unravel; native promoters in their minimum state have not proven satisfactory. Moreover, as indicated earlier, the context of the gene is important, such that some promoters are quite active until the virus genome becomes integrated or is subject to chromatin formation. Engineering vectors that respond to chromatin remodeling for gene expression has not been satisfactorily worked out, so that gene control in vectors often relies on empirical rather than rational methods. There are exceptions; "housekeeping" gene promoters tend to be functional most often (if not most vigorously), although targeted gene expression involves another level of complexity in design. For some vectors, the virus contains naturally occurring strong constitutive promoters, such as the retrovirus LTRs, or tissue-specific promoters such as the nerve-specific latency promoter of HSV. While this is a promising area of investigation for further gene vector development, the use of enhancer elements out of context can lead to an expression that is not appropriately controlled.

A related issue concerns control of when and how long the gene product should be expressed. Some vector-expressed hormones, for example, are highly active in small doses – long-term expression could be harmful or could lead to malignant transformation. It would be highly useful

if a gene switch were available for use in vectors, in which a safe and bioavailable drug could be used to turn gene expression on and off. Several systems have been developed (Table 3). The common theme shared by these systems is that a chimeric transcription factor both (1) recognizes a specific target promoter driving the therapeutic transgene, and (2) is activated by a drug. These systems require the constitutive expression of one or more transcriptional activator gene products. There are several problems associated with all of the available systems: the promoters are usually "leaky" (low-level transcription persists even in the absence of the activating drug or presence of the repressor); the constitutively expressed chimeric transcription factors are frequently immunogenic; maintenance of activator gene transcription is as difficult as effecting long-term expression of any transgene; and the targeted promoter may become unresponsive due to methylation, chromatin structure, and so on. Despite these difficulties, evidence for gene switch function in animal models appears promising; perhaps these systems will be improved in various ways, including humanization of the activator gene products to reduce immunogenicity. Another strategy for achieving the same end result might be engineered gene products that are constitutively expressed, but functionally inert in the absence of an "activator" drug. Such a system might be simpler to apply, but more difficult to design or apply broadly and may only work in special cases. This remains a fertile area of investigation for the field.

3.8 Nonviral Vectors

A compelling argument for the gene therapy field has been that if nonviral vectors could be developed with similar efficiency and maintenance of transgene expression could be done as viral vectors, several difficult problems in the field would be addressed. These include scalable and more controlled manufacturing processes, since DNA can be readily produced in bioreactors to very high concentrations. The products would be more consistent and better characterized, and delivery could be simple, involving direct injection into tissue. Theoretically, the immune response to nonviral vectors should be attenuated because there are no protein components in the particles – if true, repeat dosing might also be possible. These arguments have been especially attractive to the pharmaceutical industry and thus a considerable amount of resources has been invested in the field of nonviral gene delivery.

Despite these efforts, several problems have arisen that have been very difficult to address. The first is that DNA alone has proven to be highly immunogenic, particularly when CpG repeat sequences are present, and when used with various delivery agents such as cationic liposomes. The second is that targeting is generally not possible, even when specific ligands are attached to the DNA or the delivery vehicle, and gene delivery by nonviral vectors is relatively inefficient compared to viral vector-mediated gene transfer. Third, transgene expression is highly variable and generally very short lived. There have been some exceptions to these general observations. DNA delivery to some tissues such as skeletal muscle is reasonably efficient, and nonviral vector-mediated DNA delivery has proven to be useful for immune priming as part of vaccine strategies.

Tab. 3 Examples of inducible promoter systems.

Type	*Inducer/ repressor*	*DNA response element*	*Chimeric Transcription factor(s)*				*Mechanism of regulation*
			Name	*DNA binding domain*	*Ligand Binding domain*	*Transactivation domain*	
Tet-off	Tetracyclines	Bacterial tet operon sequences (tetO)	tTA (tet transactivator)	Bacterial tet repressor		HSV VP-16	tTA binds to tetO in absence of tetracycline
Tet-on	Tetracyclines	Bacterial tet operon sequences (tetO)	rtTA (reverse tet transactivator)	Mutated bacterial tet repressor		HSV-VP-16	rtTA binds to tetO in presence of tetracycline
Progesterone regulated	Mifepristone (RU486)	Yeast Gal4 upstream activating sequence (UASG)	GL-VP	Yeast Gal4 transcription factor	Mutated human progesterone receptor	HSV VP-16 or human NFκB p65	GL-VP binds to UASG in presence of RU486
Ecdysone regulated	Ecdysone, muristerone	Drosophila ecdysone response element	1.VgEcR	Drosophila ecdysone receptor		HSV VP-16	VgEcR forms active complex with hRXRα or drosophila USP in presence of ecdysone
			2. Human retinoic acid receptor X or drosophila ultraspiracle (USP)				
Rapamycin regulated	Rapamycin	ZFHD1 response element	1. ZFHD1-FKBP 2. FRAP-p65	Human ZFHD1 none	Human FKBP Truncated human FRAP	None Truncated human NFκ B p65	Rapamycin binding brings together functional units of bipartite transcription factor

4 Clinical Applications and Genes

4.1 Single Gene Disorders

Gene delivery to treat Mendelian single gene disorders would allow the fundamental genetic abnormalities to be addressed, rather than the current best practice of intermittently delivering gene products or devising some other compensatory strategy for the genetic lesion. The list of potential gene therapy targets in this category is very large – theoretically, any genetic disease could be treated this way. There are, however, a few examples that have been instructive in terms of understanding the factors likely to determine a successful outcome, and some disease targets are attractive, because the factors governing successful development of therapy are likely to be less stringent than in other diseases. The following paragraphs present a selection of applications, including some where clinical efficacy has already been demonstrated and others where practical gene therapy seems a remote prospect at present.

4.1.1 Hematological Disease

These disorders provide an attractive target for the gene therapist. The diseases are often very severe or fatal and have limited effective therapy at present. Many of the diseases are well characterized at a molecular level, so that the genes that should be delivered are well known. Finally, many of these conditions result from defects in genes necessary for the function of cells during or after differentiation from bone marrow multipotent precursors. Consequently, the transduction event may be carried out *ex vivo* in autologous harvested bone marrow stem cells, so that the vector only enters the target cells, and a degree of selection or expansion is possible before the cells are returned to the host. A list of possible disease targets in this category is shown in Table 4.

Gene therapy for SCID has been ongoing for more than a decade, beginning in 1990. While there are some 60 genetic defects that lead to the SCID phenotype, the availability of patients and a clear understanding of the disease process has focused attention on two genes, ADA deficiency and the gamma c chain of the cytokine receptor (γc). ADA deficiency leads to toxicity to developing immune cells (T and B lymphocytes) and γc deficiency results in a defective growth factor receptor on T and B cells that is needed for responsiveness to cytokine induction of proliferation and maturation of immune lymphocytes. Current management of ADA deficiency is carried out by infusing the ADA protein into the blood, which is taken up by immune cells primarily macrophage. This results in partial restoration of the immune system. The enzyme is PEGalated to enhance stability in the serum and to increase uptake. The first landmark SCID-ADA clinical trial was carried out by French Anderson and Michael Blease, who were then members of the National Institutes of Health (NIH). The patients were given autologous bone marrow cells derived from the circulation that had been infected with a Moloney-based retroviral vector carrying the ADA gene. Infusion of these cells did prove to be safe, but patients continued to receive PEG-ADA for safety reasons. Other trials have continued along these lines until a recent study in Italy carried out by Claudio Bordignon has attempted this treatment in the absence of PEG-ADA therapy. Patients appear to be responding without toxicity and it is likely that the potential success of this trial

Tab. 4 Possible disease targets for gene transfer to hematopoietic stem cells.

Disease	*Gene*
Red Cell Defects	
β-Thalassaemia	β-Globin
Sickle-cell disease	β-like globin
G6PD deficiency	G6PD
Porphyria	Ferrochelatase, others
Cytoskeletal defects	Spectrin
Leukocyte defects	
Chronic granulomatous disease	gp91phox, p47phox
Leukocyte adhesion defect	CD18
Stem-cell defects	
Fanconi anemia	FANC A–E
Immune deficiency syndromes	
Severe combined immunodeficiency (SCID)	γc, ADA, JAK3 kinase
Wiskott–Aldrich syndrome	WASP
CD40 ligand deficiency	CD154
Agammaglobulinemia	BTK
Lysosomal storage disease	
Gaucher disease	Glucocerebrosidase
Hurler's syndrome	α-Iduronidase
Leukodystrophies	X-adrenoleukodystrophy ALDP
Metachromatic leukodystrophy	Arylsulphatase A

can be attributed to the removal of PEG-ADA therapy. ADA expression should lead to selection of the gene-repaired cells *in vivo*, whereas continued use of PEG-ADA prevents this selection. This brings up an important point for gene therapy, which is that the most efficient outcomes may occur in circumstances where transduction occurs in a tissue where there is cellular turnover, and the transduced cells acquire a selective growth advantage. This seems to have contributed to the unprecedented clinical success of the SCIDX gene therapy trial described in the section below.

A number of other targets for HSC gene therapy are contemplated, including hemoglobinopathies, leukodystrophies, lysosomal storage diseases, porphyrias, and leukocyte defects.

4.1.2 Restoration of Circulating Proteins

Some hereditary diseases are caused by genetic defects that abolish the expression of circulating proteins. Hemophilias are the commonest of these disorders. Absence of circulating clotting factor VIII (hemophilia A) or IX (hemophilia B) results in defective blood coagulation, with a tendency to bleed spontaneously into joints (hemarthrosis), following dental extractions or surgical procedures, or following even minor trauma causing life-threatening hemorrhage. Both diseases are transmitted as Mendelian X-linked traits and are therefore usually manifested only in boys. Current treatment involves infusion of purified protein factors, providing temporary resolution of the clotting defect at large expense, inconvenience, and at a finite risk of transmitting infection,

where factors are obtained from pooled blood donations.

These diseases are attractive targets for gene therapists for several reasons. First, they are relatively common and disabling. Second, the blood levels of the clotting factors necessary to restore physiological clotting function are low; levels of >1% normal are rarely associated with spontaneous hemorrhage; levels above 5% normal are associated with posttraumatic bleeding only; and levels above 20% normal are associated with a normal phenotype. This means that complete restoration of the blood levels produced physiologically (by the liver and endothelium) is not necessary to have a clinical impact, and also that the regulation of expression levels might not be as critical as in some diseases, where, for example, stoichiometry with other components of a multiprotein complex is important, or the levels of a potently bioactive protein controlled closely to prevent toxicity. Finally, the requirements for restoration of function are that the protein should be present in the bloodstream, and so the site or tissue of transduction might not be critical, as long as the "depot" tissue has access to the circulation. This means that transduction of more accessible sites than the site of physiological production (the liver) might be possible, or that engineered cells created by *ex vivo* transduction might be able to correct the underlying defect.

Using animal models of hemophilia (there are both mouse and dog models), transient restoration of hemostatic levels of factors VIII and IX was demonstrated using first- and second-generation adenoviruses. However, the inflammatory response that occurs in response to adenovirus was responsible for limiting the time span of gene expression, an effect that was partially mitigated using a gutless AdV vector. Similar studies using lentiviruses in mouse models have shown a sustained increase in levels of factor IX, but not of factor VIII owing to an immune response to the human protein used in the mouse model. More recently, adeno-associated viral vectors have been shown to allow sustained expression of clotting factors at therapeutic levels, following transduction of either liver or muscle in animal models. In one clinical trial, hemophilia B patients have been treated with hepatic arterial infusions of an AAV2 vector carrying the factor IX gene. At low dosage (2×10^{11} particles), no toxicity or elevation of factor IX above background was detectable. One patient received an intermediate dosage (1×10^{12} particles). Circulating factor IX levels of 10 to 12% were detectable within 2 to 3 weeks after transduction. Unfortunately, the levels of factor IX then dropped back to the baseline as the levels of circulating transaminases increased, indicative of hepatocellular injury. A similar dose of vector produced a therapeutic level of factor IX lasting 1 to 2 years in the canine model. It is unclear at present whether the pattern seen in the single reported patient will be reiterated, or whether there will clinical benefit, in other patients. Another study using AAV2 injected into muscle reported a transient increase in factor IX levels at a lower vector dose, but no measurable increase at a higher vector dose. Finally, a phase I trial used endogenous fibroblasts transduced with the factor VIII gene *ex vivo* to investigate cell-based delivery systems. Following expansion, $1-4 \times 10^{8}$ cells were reimplanted into the omentum of six patients. Four of these had detectable elevations in serum factor VIII levels, but these were transient and all had returned to baseline by one year. It seems that better understanding of the processes regulating antibody responses to factors VIII and

IX, the immune response to viral vector components, and the regulatory events involved in long-term gene expression and implanted cell senescence, may be necessary to take this work forward.

α1-antitrypsin (α1AT) deficiency is another example of a genetic disease associated with loss of a circulating protein. This serine protease inhibitor is normally produced in the liver and secreted into the circulation. Patients with loss of the circulating enzyme inhibitor develop abnormal lung proteolysis and a severe form of emphysema. As one primary site of pathology is the lung, it has been suggested that therapy directed at local extracellular expression in the lung might be easier to accomplish than systemic delivery, and equally therapeutically efficacious for the lung disease. However, to prevent the liver disease that accompanies α1AT deficiency in the circulation, it will be necessary to express normal α1AT in the liver. This is because the liver pathology is caused by misfolding of a mutant protein rather than loss of its function, and so local synthesis of wild-type protein will be necessary to suppress production of the mutant. Animal studies have been carried out using adeno-associated virus and adenovirus vectors expressing α1AT. By hepatic or muscular delivery of the vectors, it is possible to produce detectable levels of circulating protein for sustained periods. One reported clinical trial used a liposome complex to deliver the AAT gene to the nasal epithelium of subjects with α1AT deficiency. Other clinical trials have yet to commence, including one aiming to express circulating α1-AT in skeletal muscle using an AAV vector.

4.1.3 Respiratory Disease

Cystic fibrosis (CF) is the commonest autosomal recessive genetic disorder in the western world, with an estimated gene carrier frequency of 1 : 20, resulting in 1 : 1600 livebirths being affected by the condition. CF is a multiorgan disease affecting the respiratory system, pancreas, digestive system, skin, and reproductive organs. The product of the *CF* gene, the CFTR is an integral membrane protein involved in ionic flux in exocrine glands. Its loss results in extremely thickened secretions causing bacterial colonization, repeated infection, and bronchiectasis in the lungs, and exocrine dysfunction in the pancreas. The latter is amenable to enzyme replacement therapy; however, despite optimal antibiotic therapy, CF patients generally die in their teens or in their 20s from the respiratory complications of the disease.

Gene therapy has focused on the introduction of the normal *CFTR* gene into bronchial airway epithelial cells to restore normal CFTR function and thereby alleviate the respiratory problems that cause fatality. In principle, the airway epithelium should be easily accessible to viral and nonviral vectors applied topically, through aerosols or bronchoscopic instillation. As the nasal epithelium is histologically similar to that found in the bronchioles and exhibits similar functional deficits in cystic fibrosis, some studies have focused on gene transfer to nasal epithelium to show proof of principle. Early studies showed proof of principle that a number of different vector platforms could introduce the *CFTR* gene into airway epithelium, resulting in expression of CFTR. Despite early optimism, however, clinical gene transfer to correct the abnormal pulmonary phenotype of CF has proved very difficult and therapeutic results have yet to be achieved. Latest clinical gene transfer trials show that adenovirus and adeno-associated virus vectors are extremely inefficient in transducing the bronchial

epithelium, and there are major problems with repeat dosing each of these vectors (as would be necessary because of short-lived gene expression) owing to immunity. Although nonviral techniques, including liposomes and cationic lipid formulations, can be repeatedly administered, they appear even less efficient than the viral techniques, and are not entirely devoid of toxicity – one study reporting fever, muscle and joint pain in half of the subjects after aerosol delivery. Recent advances have included identification of factors such as mucus and glycocalyx that contribute to the poor efficiency of transduction *in vivo*. Although measures designed to mitigate these effects may enhance gene transfer, unfortunately, the location of receptors for adenovirus and AAV on the basal (nonairway) surface of the bronchiolar epithelium may prove an insurmountable obstacle to the use of these vectors for topical gene transfer in CF, unless their tropism can be redirected to apical receptors. Recently, other viral vectors have been developed that may transduce airway cells via the apical (airway) surface, including lentivirus pseudotyped with envelope glycoproteins from viruses that naturally infect the airway epithelia via the apical surface. Finally, inefficient transduction of epithelia has been demonstrated using nonviral vectors and oligonucleotides via the bronchial circulation, raising the possibility that gene delivery to the basal surface of the cells may be achievable. It remains to be seen whether these modified vectors or other approaches can solve the problems of efficiency, gene expression longevity, and toxicity inherent in techniques used hitherto.

4.1.4 Neurological Disease

Gene transfer to the brain presents a number of formidable challenges. Brain tissue is inaccessible to vectors from the circulation, enclosed in dense bone, extremely sensitive to functional disruption by mechanical manipulation, nonregenerating and physiologically vital. This means that *ex vivo* transduction cannot be carried out, except in rare circumstances where there is good reason to think that bone marrow stem cell transduction will allow correction of a neurological phenotype. For example, in the severe childhood cerebral inflammatory type of adrenoleukodystrophy, bone marrow transplantation early in the course of the illness appears to arrest the progressive cerebral disease, and it is reasonable to suppose that genetic modification of endogenous bone marrow stem cells might also be efficacious. For most neurogenetic diseases, however, *in vivo* gene transfer will be necessary to correct the genetic abnormality present in the brain or spine. This might be achieved through direct inoculation of vector into the area of the brain in which the pathological process is manifest. For example, replacement of the *Parkin* gene in juvenile onset Parkinson's disease might be targeted to the substantia nigra as the bulk of the initial pathology appears here. Alternatively, some viruses are transported into the CNS from peripheral sites, or within the CNS from one location to another, by axonal transport. This may allow vector inoculation at a remote and more accessible site, for example, muscle inoculation in gene therapy targeted at diseases of the motor neuron.

The pathology in many types of hereditary neurological diseases is widely distributed, and there is no good method at present for widespread *in vivo* gene delivery in the brain, as would be needed to correct the phenotype of recessive diseases like Friedreich's ataxia, fragile X

mental retardation, or Hallervorden-Spatz disease. Consequently, clinical optimism at present is limited to diseases in which the pathology is localized, or rare conditions in which the missing gene product is normally located extracellularly or can be taken up by cells adjacent and remote to those transduced. Examples of the latter groups of diseases include the rare mucopolysaccharidoses with CNS involvement, where successful phenotypic outcomes have been seen following gene transfer in animal models.

Two further barriers present obstacles for gene therapy of single-gene diseases of the CNS. First, most of the adult-onset and many of the childhood-onset conditions are inherited in an autosomal dominant pattern and are caused by pathological gain of function or dominant negative effects. Examples of these diseases include the spinocerebellar ataxias, Huntington's disease, inherited prion diseases, and familial Alzheimer's diseases, the commonest forms of dystonia and some developmental diseases. Gene therapy in these cases would depend on gene targeting, possibly with gene replacement, rather than gene delivery on its own. There has been progress in this field, including development of RNA interference technology, but there have been no clinical gene therapy trials in these diseases. Second, the majority of these disease run their course over years and decades, and, as the brain is not an accessible site for repeated vector inoculation, long-term expression of transgenes would be highly desirable before gene therapy for these disease could be contemplated. It is likely that, for these type of applications, success will be achieved by exploiting the properties of viruses that persist long-term in the nervous system as part of their natural life cycle.

4.1.5 Muscle

Duchenne muscular dystrophy (DMD) is a common (1 : 3500 male births) genetic disease of muscle that results in progressive muscle wasting and death by the age of 15 to 20 from respiratory impairment or cardiac disease. It is caused by the loss of a muscle protein, dystrophin, which forms part of a large complex of proteins that localize to the sarcolemma of muscle fibers and function to maintain the integrity of the muscle cell membrane during contraction. The dystrophin gene is very large (the cDNA is >10 kb) and therefore can only be accommodated within gutless adenovirus or herpes simplex virus vectors – both of which are inefficient vehicles for transduction of muscle. Much work has been carried out in trying to develop truncated functional versions of the dystrophin gene that may be accommodated in adeno-associated virus vectors, which appear more promising as vectors in muscle tissue. A group of conditions, the sarcoglycanopathies, are closely related to DMD. They are caused by mutations in genes encoding other components of the protein complex of which dystrophin is part. These genes are smaller and can be readily accommodated within adeno-associated virus vectors. Successful correction of either the dystrophin- or sarcoglycan-deficient phenotypes presents a number of significant challenges, however, even if a useful vector system can be developed. First, the tissue compartment requiring transduction by the vector is vast and is distributed throughout the body, necessitating many multiple points of inoculation or systemic delivery. Second, functional recovery of an individual muscle is likely to depend on the transduction of a threshold proportion of muscle fibers, by introduction of a transgene at a number of sites along the syncitial muscle

fiber, so inefficient or incomplete transduction would likely be ineffective even if distributed over the large tissue compartment. Although a number of preclinical and clinical studies have shown proof of concept that dystrophin and other missing proteins can be expressed transiently in dystrophic muscle, no clinical studies have yet demonstrated any phenotypic benefit.

4.2 Delivery of Proteins for a Pharmacological Action

Gene therapy has much broader potential application than the delivery of genes that are missing in inherited Mendelian recessive diseases. There are many examples of potently active biological proteins whose delivery to specific sites might be useful to correct the abnormalities in certain diseases. Delivery of these proteins might be best accomplished by gene transfer, either because of convenience or costs issues (it is very expensive to manufacture purified recombinant proteins, and difficult to continuously administer them to sites such as the brain or joints), or because of biological necessity (for example, proteins that act intracellularly, or proteins intended to act at the cell surface of targeted cells only, would have to be expressed within the target cell unless some form of highly efficient protein transduction could be developed). The following paragraphs describe some potential applications of gene therapy to allow protein pharmacology rather than gene function replacement.

4.2.1 Transplantation

Rejection is a serious and common consequence of organ transplantation, necessitating the use of potent and toxic immunosuppressant drugs, often with incomplete suppression of rejection. Rejection of transplanted tissue depends on host CD4 and CD8 T-cells becoming activated in response to the transplantation of foreign antigens. There has been much work focusing on molecular definition of the events that follow stimulation of the T-cell receptor by antigen bound to the surface of antigen presenting cells (APC). It appears that interaction of T-cell CD28 with antigens CD80 and CD86 present on the antigen presenting cells favors immune activation, whereas signaling through another T-cell receptor, CTLA-4, results in downregulation of the immune response. Systemic blockade of the CD28 – CD80/86 interaction in animal models, using a chimeric antibody, results in systemic immune suppression. Recent work has examined the possibility of expressing the blocking antibody locally in the allograft by gene transfer; although this improved graft survival, it also resulted in systemic immune suppression. A similar approach to blocking the interaction between T-cell ligand CD154 with the APC receptor CD40 resulted in suppression of acute rejection, but not chronic rejection. Following costimulation, T helper cells secrete cytokines, with particular profiles that correlate with either rejection (Th1 response) or lack of rejection (Th2 response). Consequently, it has been suggested that creating a Th2 response in the transplanted tissue may prevent rejection. As most cytokines are associated with severe adverse side effects on systemic administration, the local production of Th2 cytokines, for example IL-10, in the graft has been suggested, using gene transfer. This approach enabled modest prolongation of graft survival in animal models. Other parallel approaches have focused on genetic modification of antigen presenting cells, causing them to synthesize immune

tolerance–inducing factors like TGF-β, or factors that specifically cause elimination of activated T cells, for example, *Fas* ligand. Finally, another approach to the induction of immune tolerance using gene transfer relies on the observation that hosts carrying allogeneic bone marrow transplants exist in a state of mixed host-donor hematopoietic cell chimerism, in which specific tolerance to donor antigens develops. On the basis of this observation, gene transfer strategies have been devised in which donor MHC molecules are expressed in host hematopoietic cells using *ex vivo* gene transfer. In several animal models, this seems to induce hyporesponsiveness to the transplanted antigens, allowing prolonged graft survival without compromising systemic immunity. None of these approaches has yet proven sufficiently successful in preclinical studies to warrant testing in a clinical trial, although the field is moving rapidly and the potential applications for gene transfer expanding in parallel with understanding the processes involved in transplant rejection.

4.2.2 Ischemic Heart Disease

Ischemic heart disease is the leading cause of death in the western world, and is most commonly caused by atheromatous occlusion of the coronary vessels. The blood supply to the myocardium becomes compromised and there may be a supply/demand mismatch resulting in cardiac ischemia or infarction. Conventional therapy includes addressing factors that may contribute to the development of atheroma, drugs that improve cardiac perfusion or reduce cardiac oxygen and energy demand, or various re-vascularization procedures. The success rate of these procedures is less than perfect, and restenosis is common after revascularization procedures like angioplasty or bypass grafting. Consequently, there has been considerable interest in gene therapy approaches that might help to encourage endogenous neovascularization in areas of ischemic myocardium through the local expression of potent peptide factors that encourage growth of new blood vessels (for example vascular endothelial growth factors, VEGF1 and VEGF2, and fibroblast growth factors FGF1 and FGF4), or the introduction of genes encoding transcription factors that are instrumental in executing the coordinated neovascularization response (for example hypoxia-inducible factor-1α, HIF1α). Transfer of these genes to ischemic myocardium or limb tissues has been demonstrated in preclinical studies using adenovirus and adeno-associated virus vectors in addition to nonviral vectors. First- and second-generation AdV gene transfer, as elsewhere in the body, gives rise to nonsustained gene expression and a local inflammatory response. In many of the preclinical studies, expression of the growth factor or transcription factor resulted in significant angiogenesis. Clinical studies have demonstrated the feasibility of intracardiac or intracoronary administration of the gene transfer reagents encoding angiogenic growth factors. Most of the studies have been small and primarily aimed at assessing safety rather than efficacy. Even so, markers of cardiac ischemia that would not be expected to be susceptible to placebo artifact, such as SPECT scanning or angiography, were improved in two of the trials examining VEGF2 expression in the myocardium or endocardium. A single double-blind placebo-controlled trial examined intracoronary gene transfer using an adenovirus vector encoding FGF-4. This trial did not demonstrate any statistically significant benefit from the treatment. Understanding the factors, for example,

the longevity of gene expression, that underlie successful revascularization is still incomplete and it is expected that further advances in gene delivery technology and vascular biology will pave the way for success in this field.

4.2.3 Neurological Disease

The common sporadic form of Parkinson's disease does not have a known or defined genetic basis, yet many potential protein-based interventions have been identified through work in tissue culture and animal models. These include intracellular proteins (for example antiapoptotic factors to prevent cell loss or enzymes to enable functional rescue of the abnormal neurochemical phenotype following cell loss) or extracellular proteins (for example growth factors that might allow restoration of the damaged basal ganglia circuitry). One recent study described the intraputaminal infusion of recombinant glial cell-line derived neurotrophic factor (GDNF) in PD patients, with beneficial effects on functional imaging surrogates of cell integrity. The infusion of GDNF may prove to be clinically efficacious and was associated with few minor adverse effects. The route of administration is, however, cumbersome and expensive; therefore, *GDNF* gene delivery has many possible advantages over the local infusion of recombinant factor. These advantages might include the capacity for prolonged local synthesis and delivery of the factor *in situ*, in addition to convenience and cost issues. However, it will be essential to develop vectors that allow regulated gene expression and sustained transgene expression before *GDNF* gene delivery might be contemplated.

Using a different approach for *PD* gene therapy, a clinical trial is about to commence using an AAV vector to transfer a gene encoding the enzyme glutamic acid decarboxylase (GAD) to the brain. This strategy is based on a number of observations concerning the functional abnormalities present in the basal ganglia of patients with PD. Neurons of the subthalamic nucleus (STN) are generally overactive in PD and secrete an excitatory neurotransmitter, glutamate, which acts to increase the firing rate of postsynaptic neurons in the internal *globus pallidus* and *substantia nigra pars reticulata*. It is known that electrically evoked reduction in STN overactivity in PD can be therapeutically efficacious. GAD is involved in the synthesis of an inhibitory transmitter, γ-amino butyric acid (GABA). STN neurons expressing GAD produce mixed excitatory/inhibitory postsynaptic responses. This may functionally alleviate the effects of STN overactivity by partially antagonizing the postsynaptic responses to STN overactivity. The clinical trial has recently begun to recruit patients and will address toxicity and dosing issues rather than therapeutic effectiveness in the first instance.

Other examples of gene transfer–mediated protein pharmacology in neurology might include the targeting of neuroimmunological regulatory molecules in diseases with a prominent inflammatory or immune component, such as multiple sclerosis (MS). For example, it is known that the antiinflammatory cytokine β-interferon is weakly efficacious when administered systemically in relapsing-remitting MS. Various animal models of autoimmune myelin destruction have shown that pathological effects can be mitigated by expression of a range of antiinflammatory cytokines in the cerebrospinal fluid and extracellular fluid of the CNS parenchyma, following transduction of the meninges as depot sites for production. It is possible that this may be

a viable way to deliver immunomodulatory molecules to the CNS.

4.2.4 Cancer Gene Therapy

Despite recent advances in clinical oncology and radiotherapy, many types of cancer are still associated with a poor prognosis, and many of the current best treatments are highly toxic. Gene therapy thus provides an attractive experimental approach to developing novel therapeutic reagents to tackle malignancy. Of human clinical gene therapy trials carried out so far, the majority have been in patients with various cancers, reflecting both the desperate need for better treatments and the perceived potential for this technology to yield favorable results.

The aims of cancer gene therapy are rather different to those of other gene therapy applications, in that the gene therapist seeks to selectively destroy target cells. This has been accomplished in a number of ways, including the following:

(a) *Suicide gene therapy.* The transgene product is toxic to expressing cells. An example is the HSV-thymidine kinase-ganciclovir (TK-GCV) system in which expression of viral TK allows activation of a prodrug (GCV) that interferes with DNA replication in rapidly dividing cells.
(b) *Radiosensitization.* The transgene product enhances the toxicity of gamma irradiation in the tumor. TNF-α is an example.
(c) *Immunotherapy.* The transgene product stimulates an immune response directed against tumor. Examples include CD80, TNFα, GM-CSF, IL-2, interferon-γ, and IL-12.
(d) *Restoration of cellular functions altered during oncogenesis.* Many tumors acquire means of avoiding apoptosis in response to the mutations in cellular DNA that often occur during oncogenesis. A common example is the loss of p53 signaling that occurs in many types of solid tumors. Replacement of p53 function in this instance can cause the malignant cells to undergo apoptosis. Other examples might include delivery of other tumor suppressor genes that are mutated in the tumor, or targeting oncogene products or their related cellular pathways to suppress tumor growth.
(e) *Disruption of tumor blood supply.* Most tumors excite an angiogenic response that allows continuing, and adequate, metabolic support for the expanding the tumor mass. Gene therapy can be designed to interfere with this process, causing the tumor to outgrow its blood supply and undergo necrosis.

The mode of action of these reagents is to effect selective destruction of malignant cells. It follows that some means of targeting is essential to prevent destruction of normal tissue. This might be accomplished in several ways. Firstly, strategies might be designed that rely on limiting the expression of a universally toxic gene product to cancer cells. For example, targeted viral delivery, by engineering the tropism of the vector, or targeted transgene transcription in malignant cells, by using specific regulatory elements in the transgene expression cassette, might accomplish the limited spatial expression of a toxic suicide gene product. Secondly, strategies might be deigned that rely on the differential toxicity of a gene product expressed in both normal and malignant cells. Gene products that are directly and selectively toxic to cancer cells could be expressed by both normal and malignant cells, with a preferential effect on the latter. An example might include the HSV-TK

system, where the rapid cell division characteristic of malignant cells renders them especially sensitive to disruption in cellular DNA synthesis. Alternatively, gene products could be expressed within a tissue, in order to induce an immune response, which is preferentially directed against the tumor, by virtue of its abnormal antigenic signature. The first of these general approaches is still in early stages of development, as the problems inherent in effecting targeted vector transduction and cell-specific vector-mediated transcription have yet to be fully overcome. In addition, these targeting strategies will depend upon better definition of tumor-specific features, such as tumor-restricted cell surface receptors and cis-acting regulatory elements that are preferentially active in cancer cells. This type of strategy is extremely attractive, however, because there is a theoretical possibility that avid and highly specific targeting would allow systemic administration of the vector, with resulting selective destruction of metastatic tumor deposits. This is rarely achievable in current practice, with few notable examples, such as the ablation of metastatic well-differentiated thyroid tumors through targeting a defined cellular phenotype, the uptake of iodine, allowing selective sequestration of radioactive iodine. It is hoped that similar molecular targeting through gene therapy will allow parallel approaches to be taken in targeting defined cellular processes unique to tumor tissue.

Various gene transfer vectors have been used in preclinical and clinical trials of cancer gene therapy, including replication-deficient lentiviruses, adenoviruses and herpes simplex viruses, and conditionally replicating HSV and adenoviruses. It is not yet known which vector is optimal, and it is likely that different vectors will show utility for specific applications. For example, the high-level transient expression and immunogenicity of AdV may have application for suicide gene therapy and immunotherapy applications in rapidly dividing tumors, whereas the long-term gene expression and capacity for multiple transgenes seen with HSV vectors may make them good candidates for use in slower growing tumors where multiple simultaneous pathways must be targeted to achieve tumor eradication, for example, glioblastoma.

5 Adverse Events

Considering the number of patients treated with gene therapy in clinical trials, there have been relatively few serious adverse events that have threatened the lives of participants. Nevertheless, those that have occurred have been rather dramatic.

5.1 Multiple Organ Failure Secondary to Cytokine Storm

In the late 1990s, a dose-escalation phase I trial was carried out by the Human Gene Therapy Institute at the University of Pennsylvania. This trial involved the use of replication-defective adenovirus, carrying the gene encoding ornithine transcarbamylase (OTC), to treat inherited OTC deficiency. OTC normally functions as part of the urea cycle, to detoxify nitrogen produced during hepatic amino acid metabolism. The disease phenotype varies in severity; the most significant disease is associated with a completely defective gene. Failure to detoxify hepatic nitrogen leads to hyperammonemia and metabolic coma, which may be fatal if left untreated. It was hoped that successful

safety testing of the AdV-OTC vector would eventually lead to the treatment of children suffering from this fatal disease, following successful gene transfer in preclinical studies. However, following review of the protocol, the study was redesigned to enroll adult patients with mild disease. One such patient, Jesse Gelsinger, received the highest dose of vector (3.8×10^{23} infectious particles) at the top end of the dose-escalation schedule. Within hours, he developed a fever, accompanied by biochemical evidence of hepatic damage. Inflammatory cytokines appeared in the blood, followed by signs of disseminated intravascular coagulation, and respiratory and multiorgan failure. He died within four days of receiving the vector. A second patient had received a similar dose of vector, but did not experience severe side effects.

It was concluded from this study that adenovirus introduced through the hepatic portal vein was not confined to the liver, and that the virus had a greater propensity to infect Kupffer cells than hepatocytes in humans. Kupffer cells are reticuloendothelial cells, which were probably responsible for the release of high levels of circulating cytokines following transduction with recombinant adenovirus. This result differed enormously from the experience in mouse models of OTC deficiency, in which the AdV vector largely infected hepatocytes by the same route of administration. Studies in primates showed toxicity similar to that observed in the patient who died.

This outcome demonstrated that the initial immune response to very high doses of vector can be potentially life threatening. Indeed, high doses of any virus will probably result in toxicity, due to the significant load of virus proteins, rather than the transgene. A fire storm of criticism directed at the field of gene therapy followed this unfortunate adverse event. The development of stricter control over gene therapy clinical research was an inevitable result. This outcome at Penn essentially eliminated the use of adenovirus-based vector for liver gene transfer and reduced the interest in using adenovirus *in vivo* for applications other than cancer gene therapy or vaccines, where the immunogenicity of the vector would be advantageous. This has, in turn, focused attention on other vector systems and accelerated progress in other areas.

5.2 Leukemia Secondary to Insertional Mutagenesis

The one true cure of an inherited disease by gene therapy was achieved in the treatment of severe combined immunodeficiency (SCID). There are 60 genetic defects that can lead to the SCID phenotype. Deficiency of the common γc chain of the cytokine receptor results in failure of maturing B and T cells to respond to cytokine induction for both proliferation and maturation. The γc gene is located on the X chromosome, and correspondingly, the disease is inherited as a Mendelian X-linked recessive trait. Death at an early age, secondary to recurrent infections is the usual outcome, unless the children can be treated by donor-matched bone marrow transplantation, which has approximately an 80% success rate. Children without matched donors are treated by transplant with a relatively low success rate and at considerable risk.

Replacement of the missing γc chain gene by *ex vivo* gene transfer to autologous CD34+ hematopoietic stem cells was carried out in patients after successful outcomes in preclinical studies, using a retroviral vector encoding the γc cDNA under transcriptional control of the retroviral

LTR promoter. The transduced cells were reintroduced into the patients, and initial results showed repopulation of bone marrow with differentiating cell lineages and appearance of the missing differentiated cell types in peripheral blood. Correspondingly, there was recovery of immune functions, with many of the 10 treated patients no longer experiencing recurrent infections and dependence on antibiotic therapy (see above).

Unfortunately, after an interval of three years, it became apparent that two of the patients had developed clonal expansion of a T-lymphocyte subpopulation, with peripheral lymphocytosis, bone marrow infiltration and splenomegaly – clinical signs of leukemia. In each case, the clone, which in retrospect was detectable in samples from the patients at earlier time points in the follow-up process, contained a single copy of the retrovirus integrated into the host cell chromosome 11, within or near the *LMO2* gene. The leukemia cells contained additional dyskaryotic abnormalities, including a t(6 : 13) translocation in one patient and clonal populations with various abnormalities including trisomy 11 in the other. Both patients are undergoing chemotherapy at the time of writing this review.

It is acknowledged that there was no reason to suspect that this outcome would occur – a decade of clinical trials using retroviruses to transduce hematopoietic stem cells preceded the SCIDX trial, without a single case of leukemia occurring. In all reports of mice receiving retrovirally transduced HSCs, there is one description of a single animal that developed myeloid leukemia secondary to integration of the vector within a protooncogene that was previously implicated in leukemogenesis. LMO2 is a leukemogenic oncoprotein, as determined by murine transgenic overexpression experiments, which is activated in human childhood acute leukemia. It is transcriptionally active in progenitor cells, which may explain the propensity of the vector to integrate within this gene. It is unclear how integration into this gene contributed to the formation of leukemia in the SCIDX trial patients. The prolonged latency between retroviral integration and malignant clonal emergence argues that additional factors must have been involved, other than insertional mutagenesis following the transduction event. For example, unregulated expression of the γc chain from the viral LTR promoter may have contributed to malignant transformation; there are reported precedents of leukemias that are dependent on γc cytokine signaling. Furthermore, the disease may have provided unique circumstances resulting in oncogenic interaction between multiple factors. The combination of LMO2 upregulation, unregulated γc expression, and the disease-related immune deficiency preventing detection and eradication of malignant clones may have allowed subclones with additional oncogenic mutations to emerge. These possibilities are currently being investigated further.

The implications for the development of gene therapy mediated by integrating vectors are significant. First, there will be more interest in the integration of vectors into protooncogenes, and preclinical trials and toxicity studies will have to address site of integration issues. Second, vector design may be modified to incorporate tissue-specific regulatory elements, self-inactivating viral LTR sequences and insulators to prevent the inadvertent activation of neighboring genes following integration. Third, it may be desirable to incorporate a suicide gene into the vectors so that transduced clones that become

malignant can be eradicated pharmacologically. Finally, research in understanding the processes that determine the site of vector integration may be expanded, with the aim of learning how to manipulate the mechanisms involved and direct insertion into an inert part of the genome.

While there is still great skepticism regarding the future of gene therapy, the majority of workers remain optimistic that many of the problems encountered are not substantially more difficult than those seen in other fields where new technology is under development, including transplantation and cancer chemotherapy.

6
From Bench to Clinic – HSV-mediated Gene Therapy of Pain as an Example

One example of therapeutic gene transfer is the use of HSV vectors in the treatment of chronic pain. Chronic pain is defined as pain that persists beyond the course of the acute insult, or pain that accompanies a chronic primary process that cannot be cured. Although the majority of chronic pain can be effectively treated with existing remedies, a significant minority of patients cannot, resulting in substantial morbidity and cost. The neuroanatomic pathways involved in pain are identical to those involved in the perception of acute painful stimuli, but alterations in gene expression in neuronal and nonneuronal elements at many levels of the neuraxis produce an altered substrate that result in symptoms that are refractory to medical management. The traditional pharmacologic approach to chronic pain is epitomized by the search for small molecule agonists that would be selective for elements in the pain pathway, but most of those components are not unique to nociceptive neurotransmission. Opioid receptors, for example, are present at all levels of the neuraxis, including the central and peripheral terminals of primary nociceptors, second-order neurons in dorsal horn of spinal cord, and nuclei in the brain and brain stem and are also found in nonneural tissues including bladder, gut, and inflammatory cells. Even though opiate drugs are highly effective, their continuous use for treatment of chronic pain is limited by side effects that occur because activation of receptors in nonnociceptive pathways results in sedation, constipation, urinary retention, and respiratory suppression. Continued use results in tolerance so that increasing doses of the drug are required to achieve maintain the therapeutic effect and abuse is a potential problem. The gene therapy approach to the treatment of chronic pain exploits gene transfer to achieve local release of neuroactive peptides to reduce pain. HSV-based vectors are uniquely suited to this purpose because subcutaneous inoculation of the vector can be used to transduce dorsal root ganglia (DRG) in an anatomically predictable manner to effect the release of analgesic peptides from axonal terminals of the DRG neurons in the dorsal horn of spinal cord.

Animal models have been developed that recapitulate the essential features of inflammatory pain, neuropathic pain, and pain caused by metastatic cancer, which can be used to predict the response of human patients to analgesic drugs. We tested HSV vectors in each of these models. We constructed an HSV vector containing the coding sequence for human proenkephalin under the control of a human cytomegalovirus immediate-early promoter (HCMV IEp) in a vector background deleted for a single essential immediate-early HSV gene (*ICP4*), termed *vector SHPE*. We demonstrated *in vitro* that neurons transduced with the vector

produce human proenkephalin, and release enkephalin into the medium, and demonstrated *in vivo* that subcutaneous inoculation of the vector resulted in transduction of DRG neurons and production of proenkephalin *in vivo*. In the formalin test, a model of inflammatory pain created by injection of a dilute solution of formalin, inoculation with SHPE produced a substantial and statistically significant reduction pain that was reversed by intrathecal naltrexone, a relatively selective delta opiate receptor antagonist. The effect of vector inoculation was maximal in animals tested one week after vector administration, waned over the subsequent three weeks, but could be reestablished by reinoculation of the vector.

Pain resulting from nerve damage (neuropathic pain) can be modeled by selective ligation of a single spinal nerve. Subcutaneous inoculation of SHPE into the plantar surface of the foot one week after selective L5 spinal nerve ligation provided an antiallodynic effect that began one week after inoculation and persisted for five weeks. Like the analgesic effect of SHPE in the formalin model of inflammatory pain, the antiallodynic effect of SHPE inoculation in neuropathic pain was reversed by intraperitoneal naloxone, and reinoculation of the vector after the initial effect had waned promptly reestablished the allodynic effect. The antiallodynic effect of SHPE-mediated enkephalin expression was continuous throughout the day, was additive with morphine, shifting the ED_{50} for morphine from 1.8 mg kg^{-1} to 0.15 mg kg^{-1}, and the effect of SHPE-mediated enkephalin expression persisted in the face of tolerance to morphine. Animals treated with the maximum dose of morphine (10 mg kg^{-1} twice a day) developed tolerance by one week, so that morphine no longer provided any antiallodynic effect. In contrast, animals that had been inoculated with SHPE continued to demonstrate an antiallodynic effect even after three weeks of morphine treatment.

An important consideration in moving from preclinical animal studies to human trials is the selection of an appropriate condition so that the potential risks of the novel therapy are justified, and even though chronic pain is a cause of significant suffering, the condition itself is not life threatening. Therefore, we examined the analgesic effect of the subcutaneous inoculation of SHPE in a murine model of pain resulting from cancer metastatic to bone. Tumor-bearing mice develop a spontaneous pain behavior that can be rated. Animals inoculated with a control vector manifested spontaneous pain-related behavior that increased over the subsequent two weeks, while mice inoculated with SHPE one week after establishment of the tumor showed a substantial and statistically significant reduction in pain-related behavior, and an effect that was synergistic with morphine and could be reversed by intrathecal naltrexone.

On the basis of the preclinical data in the several different rodent models of pain, we are moving forward with plans for a human trial. The plans for this phase I/II safety/dose-escalation trial of the proenkephalin-expressing vector was reviewed by the recombinant DNA Advisory Committee (RAC) of the NIH in June 2002, and the request for an investigational new drug (IND) waiver to the Food and Drug Administration has been submitted. The HSV vector for the human trial will be deleted for two IE HSV genes (*ICP4* and *ICP27*) and contains deletions in the promoters for two other IE genes (*ICP22* and *ICP47*). The proenkephalin transgene has been

placed in both copies of ICP4, so that in the unlikely event of recombination with a latent wild-type virus, the recombinants would be replication defective. In the trial, we will enroll patients with cancer metastatic to a vertebral body, resulting in pain refractory to maximal medical management. The primary outcome of this phase I/II trial will involve standard measures of safety assessed by common criteria. Secondary measures will include an evaluation of the focal pain using an analog pain "thermometer", global measures of total pain using a similar tool, as well as assessment of pain-related phenomena such as sleep and depression and the concurrent use of analgesic medication, including opiate drugs.

In animal models, HSV-mediated gene transfer to DRG neurons achieved by subcutaneous inoculation is effective in expressing peptides to be released locally from sensory neurons to achieve focal therapeutic effects. The proposed human trial will allow us to determine whether the same approach may be used in the treatment of patients, and will be the first step in the use of this therapy for pain and polyneuropathy. Proenkephalin is not the only transgene that may be used in gene therapy for chronic pain. In other studies, we have found that HSV-mediated expression of glutamic acid decarboxylase, antiinflammatory peptides (interleukin 4 or the tumor necrosis factor alpha soluble receptor), or the glial cell derived neurotrophic factor, all have analgesic effects in different models of chronic pain. HSV-based vectors may ultimately serve as the platform to deliver a number of different genes whose products have analgesic properties in the treatment of specific types of chronic pain.

7
Future Prospects – What will Gene Therapy Likely be About in 10 Years?

Gene transfer as a method to alter the biology of cells has almost unlimited potential providing safe and effective methods for gene delivery and expression can be developed. It appears likely that this can be achieved and indeed certain human viruses have largely accomplished this task. Vectors that integrate have great potential for gene delivery to dividing cells where long-term expression is required. The retroviruses are currently the vector of choice and lentiviruses in particular may become the dominant vector type since cell division is not required for vector entry into the nucleus and integration can be mediated by the vector integrase. However, integration is essentially random with a propensity for integrating at sites of open chromatin and active genes. The problems associated with gene inactivation by insertional mutagenesis or the activation of potential oncogenes is not just theoretical, posing safety concerns with the use of these vectors. Future work on the development of these viruses will endeavor to target integration, remove promiscuous enhancers in the vector LTRs, and attempt to take advantage of tissue-specific enhancers that may limit expression to certain stages of development or cell type. Also, the possibility exists that certain phage integrases may be useful if expressed from the vector along with movable elements that might achieve targeting to specific loci or chromosomal sites. Another important development may result from a better understanding of mechanisms of vector uncoating and gene expression. For example, AAV transduces many cell types with poor efficiency, requiring large multiplicities to

achieve transgene expression. Uncoating may be enhanced through use of different serotypes or through manipulation of the capsid. AAV may also be delivered in recombinant form within the genome of larger vectors such as HSV, where transient expression of rep, for example, may induce excision and subsequent integration of the AAV genome into a safe site within chromosome 19. HSV has a very large packaging potential, which can be fully realized if there were a packaging cell line available. The engineering of such cell lines is possible and may be achieved in the coming years. Other possibilities for improvement involve the establishment of precise methods of vector targeting and transgene expression. There are likely to be numerous viral or cellular genomic sequences that act in cis to influence transgene expression and these must be discovered, characterized, and applied in the context of the virus genome. The rules that govern gene expression in the virus genome are still unclear. Engineering the virus coat or envelope will continue until suitable methods are developed. Secondary targeting at the nuclear membrane may also be possible. The future will also include multigene therapies requiring the interaction of multiple gene functions to achieve a therapeutic outcome including genes, which trigger the differentiation and survival of stem cells that can rebuild damaged tissues. In general, gene transfer may become a very important tool for directed tissue development.

Diseases that are the most likely targets for gene therapy within the next 5 to 8 years will include the modification of responses in the host that either lead to improved immunity to infectious diseases or cancer, alter the host response to painful stimuli, reduce autoimmune reactions, and improve transplant tissue survival or improve the outcome for damaged tissue such as ischemia in the heart. Also, in the short term, it is likely that certain metabolic diseases such as hemophilia or inborn errors of metabolism that result in immunodeficiency or toxic metabolic states will show signs of success. Longer term, the use of gene transfer for tissue engineering, and stem cell applications for tissue repair will began to unfold. While these comments are only speculative, there is reason to believe that many of these applications are nearing the era of real benefit to patients and even if only a few of these applications are successful in the near term, gene therapy as a practical modality of medical therapy will have reached its early stages of promise.

See also Cytomegalovirus and Varicella–zoster Virus Vaccines; Liposome Gene Transfection; Medicinal Chemistry; Nucleic Acid Packaging of DNA Viruses; Nucleic Acid Packaging of RNA Viruses; Plasmids; Regulation of Stem Cell; Retrovirus Integration, Biological and Biochemical Machineries for; Somatic Gene Therapy.

Bibliography

Books and reviews

Bagley, J., Iacomini, J. (2003) Gene therapy progress and prospects: gene therapy in organ transplantation, *Gene Ther.* **10**, 605–611.

Burton, E.A., Glorioso, J.C., Fink, D.J. (2003) Gene therapy progress and prospects: Parkinson's disease, *Gene Ther.* **10**, 1721–1727.

Burton, E.A., Fink, D.J., Glorioso, J.C. (2002) Gene delivery using herpes simplex virus vectors, *DNA Cell Biol.* **21**, 915–936.

Burton, E.A., Wechuck, J.B., Wendell, S.K., Goins, W.F., Fink, D.J., Glorioso, J.C. (2001) Multiple applications for replication-defective Herpes Simplex virus vectors, *Stem Cells* **19**, 358–377.

Cheng, S.H., Smith, A.E. (2003) Gene therapy progress and prospects: gene therapy of lysosomal storage disorders, *Gene Ther.* **10**, 1275–1281.

Fields, B.N., Knipe, D.M., Howley, P.M. (1996) *Fields Virology*, 3rd Edition Lippincott- Raven, Philadelphia, PA.

Griesenbach, U., Ferrari, S., Geddes, D.M., Alton, E.W. (2002) Gene therapy progress and prospects: cystic fibrosis, *Gene Ther.* **9**, 1344–1350.

Khan, T.A., Sellke, F.W., Laham, R.J. (2003) Gene therapy progress and prospects: therapeutic angiogenesis for limb and myocardial ischemia, *Gene Ther.* **10**, 285–291.

Luciw, P.A. (1996) Chapter 60: Human Immunodeficiency Viruses and Their Replication, in: Field's, B.N., Knipe, D.M., Howley, P.M. (Eds.) *Fields Virology*, Lippincott-Raven, Philadelphia, PA pp. 1881–1952.

Niidome, T., Huang, L. (2002) Gene therapy progress and prospects: nonviral vectors, *Gene Ther.* **9**, 1647–1652.

Roizman, B., Sears, A.E. (1996) Chapter 72: Herpes Simplex Viruses and Their Replication, in: Fields, B.N., Knipe, D.M., Howley, P.M. (Eds.) *Fields Virology*, Lippincott-Raven, Philadelphia, PA pp. 2231–2295.

St George, J.A. (2003) Gene therapy progress and prospects: adenoviral vectors, *Gene Ther.* **10**, 1135–1141.

Stecenko, A.A., Brigham, K.L. (2003) Gene therapy progress and prospects: alpha-1 antitrypsin, *Gene Ther.* **10**, 95–99.

Walsh, C.E. (2003) Gene therapy progress and prospects: gene therapy for the hemophilias, *Gene Ther.* **10**, 999–1003.

Primary Literature

Aitken, M.L., Moss, R.B., Waltz, D.A., Dovey, M.E., Tonelli, M.R., McNamara, S.C., Gibson, R.L., Ramsey, B.W., Carter, B.J., Reynolds, T.C. (2001) A phase I study of aerosolized administration of tgAAVCF to cystic fibrosis subjects with mild lung disease, *Hum. Gene Ther.* **12**, 1907–1916.

Aiuti, A., Vai, S., Mortellaro, A., Casorati, G., Ficara, F., Andolfi, G., Ferrari, G., Tabucchi, A., Carlucci, F., Ochs, H.D., Notarangelo, L.D., Roncarolo, M.G., Bordignon, C. (2002) Immune reconstitution in ADA-SCID after PBL gene therapy and discontinuation of enzyme replacement, *Nat. Med.* **8**, 423–425.

Aiuti, A., Slavin, S., Aker, M., Ficara, F., Deola, S., Mortellaro, A., Morecki, S., Andolfi, G., Tabucchi, A., Carlucci, F., Marinello, E., Cattaneo, F., Vai, S., Servida, P., Miniero, R., Roncarolo, M.G., Bordignon, C. (2002) Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning, *Science* **296**, 2410–2413.

Akkina, R.K., Walton, R.M., Chen, M.L., Li, Q.X., Planelles, V., Chen, I.S. (1996) High-efficiency gene transfer into CD34+ cells with a human immunodeficiency virus type 1-based retroviral vector pseudotyped with vesicular stomatitis virus envelope glycoprotein G, *J. Virol.* **70**, 2581–2585.

Amalfitano, A., Hauser, M.A., Hu, H., Serra, D., Begy, C.R., Chamberlain, J.S. (1998) Production and characterization of improved adenovirus vectors with the E1, E2b, and E3 genes deleted, *J. Virol.* **72**, 926–933.

Anderson, D.B., Laquerre, S., Goins, W.F., Cohen, J.B., Glorioso, J.C. (2000) Pseudotyping of glycoprotein D-deficient herpes simplex virus type 1 with vesicular stomatitis virus glycoprotein G enables mutant virus attachment and entry., *J. Virol.* **74**, 2481–2487.

Armentano, D., Thompson, A.R., Darlington, G., Woo, S.L. (1990) Expression of human factor IX in rabbit hepatocytes by retrovirus-mediated gene transfer: potential for gene therapy of hemophilia B, *Proc. Natl. Acad. Sci. U. S. A.* **87**, 6141–6145.

Babiss, L.E., Young, C.S., Fisher, P.B., Ginsberg, H.S. (1983) Expression of adenovirus E1a and E1b gene products and the Escherichia coli XGPRT gene in KB cells, *J. Virol.* **46**, 454–465.

Bagley, J., Iacomini, J. (2003) Gene therapy progress and prospects: gene therapy in organ transplantation, *Gene Ther.* **10**, 605–611.

Bai, Q., Burton, E.A., Goins, W.F., Glorioso, J.C. (2002) Modifying the tropism of HSV vectors for gene therapy applications, in: *Vector Targeting for Therapeutic Gene Delivery*,

Douglas, J., Curiel, D. (Eds.) John Wiley & Sons, New York.

Barr, D., Tubb, J., Ferguson, D., Scaria, A., Lieber, A., Wilson, C., Perkins, J., Kay, M.A. (1995) Strain related variations in adenovirally mediated transgene expression from mouse hepatocytes in vivo: comparisons between immunocompetent and immunodeficient inbred strains, *Gene Ther.* **2**, 151–155.

Batshaw, M.L., Roan, Y., Jung, A.L., Rosenberg, L.A., Brusilow, S.W. (1980) Cerebral dysfunction in asymptomatic carriers of ornithine transcarbamylase deficiency, *N. Engl. J. Med.* **302**, 482–485.

Bellon, G., Michel-Calemard, L., Thouvenot, D., Jagneaux, V., Poitevin, F., Malcus, C., Accart, N., Layani, M.P., Aymard, M., Bernon, H., Bienvenu, J., Courtney, M., Doring, G., Gilly, B., Gilly, R., Lamy, D., Levrey, H., Morel, Y., Paulin, C., Perraud, F., Rodillon, L., Sene, C., So, S., Touraine-Moulin, F., Pavirani, A., et al. (1997) Aerosol administration of a recombinant adenovirus expressing CFTR to cystic fibrosis patients: a phase I clinical trial, *Hum. Gene Ther.* **8**, 15–25.

Ben-Gary, H., McKinney, R.L., Rosengart, T., Lesser, M.L., Crystal, R.G. (2002) Systemic interleukin-6 responses following administration of adenovirus gene transfer vectors to humans by different routes, *Mol. Ther.* **6**, 287–297.

Bensadoun, J.C., Pereira de Almeida, L., Fine, E.G., Tseng, J.L., Deglon, N., Aebischer, P. (2003) Comparative study of GDNF delivery systems for the CNS: polymer rods, encapsulated cells, and lentiviral vectors, *J. Controlled Release* **87**, 107–115.

Blaese, R.M., Culver, K.W., Miller, A.D., Carter, C.S., Fleisher, T., Clerici, M., Shearer, G., Chang, L., Chiang, Y., Tolstoshev, P., et al. (1995) T lymphocyte-directed gene therapy for ADA-SCID: initial trial results after 4 years, *Science* **270**, 475–480.

Bloom, D.C., Lokensgard, J.R., Maidment, N.T., Feldman, L.T., Stevens, J.G. (1994) Long-term expression of genes in vivo using non-replicating HSV vectors, *Gene Ther.* **1**, S36–S38.

Brigham, K.L., Lane, K.B., Meyrick, B., Stecenko, A.A., Strack, S., Cannon, D.R., Caudill, M., Canonico, A.E. (2000) Transfection of nasal mucosa with a normal alpha1-antitrypsin gene in alpha1-antitrypsin-deficient subjects: comparison with protein therapy, *Hum. Gene Ther.* **11**, 1023–1032.

Buning, H., Ried, M.U., Perabo, L., Gerner, F.M., Huttner, N.A., Enssle, J., Hallek, M. (2003) Receptor targeting of adeno-associated virus vectors, *Gene Ther.* **10**, 1142–1151.

Burton, E.A., Glorioso, J.C., Fink, D.J. (2003) Gene therapy progress and prospects: Parkinson's disease, *Gene Ther.* **10**, 1721–1727.

Burton, E.A., Bai, Q., Goins, W.F., Glorioso, J.C. (2001) Targeting gene expression using HSV vectors, *Adv. Drug Delivery Rev.* **53**, 155–170.

Burton, E.A., Wechuck, J.B., Wendell, S.K., Goins, W.F., Fink, D.J., Glorioso, J.C. (2001) Multiple Applications For Replication-Defective Herpes Simplex Virus Vectors, *Stem Cells* **19**, 358–377.

Caplen, N.J., Alton, E.W., Middleton, P.G., Dorin, J.R., Stevenson, B.J., Gao, X., Durham, S.R., Jeffery, P.K., Hodson, M.E., Coutelle, C., et al. (1995) Liposome-mediated CFTR gene transfer to the nasal epithelium of patients with cystic fibrosis, *Nat. Med.* **1**, 39–46.

Cavazzana-Calvo, M., Hacein-Bey, S., de Saint Basile, G., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J.L., Bousso, P., Deist, F.L., Fischer, A. (2000) Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease, *Science* **288**, 669–672.

Chao, H., Monahan, P.E., Liu, Y., Samulski, R.J., Walsh, C.E. (2001) Sustained and complete phenotype correction of hemophilia B mice following intramuscular injection of AAV1 serotype vectors, *Mol. Ther.* **4**, 217–222.

Chao, H., Liu, Y., Rabinowitz, J., Li, C., Samulski, R.J., Walsh, C.E. (2000) Several log increase in therapeutic transgene delivery by distinct adeno-associated viral serotype vectors, *Mol. Ther.* **2**, 619–623.

Chen, X., Schmidt, M.C., Goins, W.F., Glorioso, J.C. (1995) Two herpes simplex virus type 1 latency-active promoters differ in their contributions to latency-associated transcript expression during lytic and latent infections, *J. Virol.* **69**, 7899–7908.

Chow, C.M., Athanassiadou, A., Raguz, S., Psiouri, L., Harland, L., Malik, M., Aitken, M.A., Grosveld, F., Antoniou, M. (2002) LCR-mediated, long-term tissue-specific gene expression within replicating episomal

plasmid and cosmid vectors, *Gene Ther.* **9**, 327–336.

Colombo, F., Zanusso, M., Casentini, L., Cavaggioni, A., Franchin, E., Calvi, P., Palu, G. (1997) Gene stereotactic neurosurgery for recurrent malignant gliomas, *Stereotact. Funct. Neurosurg.* **68**, 245–251.

Corbeau, P., Kraus, G., Wong-Staal, F. (1996) Efficient gene transfer by a human immunodeficiency virus type 1 (HIV-1)-derived vector utilizing a stable HIV packaging cell line, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 14070–14075.

Croyle, M.A., Chirmule, N., Zhang, Y., Wilson, J.M. (2002) PEGylation of E1-deleted adenovirus vectors allows significant gene expression on readministration to liver, *Hum. Gene Ther.* **13**, 1887–1900.

Crystal, R.G., McElvaney, N.G., Rosenfeld, M.A., Chu, C.S., Mastrangeli, A., Hay, J.G., Brody, S.L., Jaffe, H.A., Eissa, N.T., Danel, C. (1994) Administration of an adenovirus containing the human CFTR cDNA to the respiratory tract of individuals with cystic fibrosis, *Nat. Genet.* **8**, 42–51.

Curiel, D.T. (1999) Strategies to adapt adenoviral vectors for targeted delivery, *Ann. N. Y. Acad. Sci.* **886**, 158–171.

David, A., Chetritt, J., Guillot, C., Tesson, L., Heslan, J.M., Cuturi, M.C., Soulillou, J.P., Anegon, I. (2000) Interleukin-10 produced by recombinant adenovirus prolongs survival of cardiac allografts in rats, *Gene Ther.* **7**, 505–510.

Davidson, B.L., Allen, E.D., Kozarsky, K.F., Wilson, J.M., Roessler, B.J. (1993) A model system for in vivo gene transfer into the central nervous system using an adenoviral vector, *Nat. Genet.* **3**, 219–223.

DeLuca, N.A., McCarthy, A.M., Schaffer, P.A. (1985) Isolation and characterization of deletion mutants of herpes simplex virus type 1 in the gene encoding immediate-early regulatory protein ICP4, *J. Virol.* **56**, 558–570.

DeMatteo, R.P., Chu, G., Ahn, M., Chang, E., Burke, C., Raper, S.E., Barker, C.F., Markmann, J.F. (1997) Immunologic barriers to hepatic adenoviral gene therapy for transplantation, *Transplantation* **63**, 315–319.

Desmaris, N., Bosch, A., Salaun, C., Petit, C., Prevost, M.C., Tordo, N., Perrin, P., Schwartz, O., de Rocquigny, H., Heard, J.M. (2001) Production and neurotropism of lentivirus vectors pseudotyped with lyssavirus envelope glycoproteins, *Mol. Ther.* **4**, 149–156.

Dmitriev, I.P., Kashentseva, E.A., Curiel, D.T. (2002) Engineering of adenovirus vectors containing heterologous peptide sequences in the C terminus of capsid protein IX, *J. Virol.* **76**, 6893–6899.

Dobson, A.T., Margolis, T.P., Gomes, W.A., Feldman, L.T. (1995) In vivo deletion analysis of the herpes simplex virus type 1 latency-associated transcript promoter, *J. Virol.* **69**, 2264–2270.

Dobson, A.T., Margolis, T.P., Sedarati, F., Stevens, J.G., Feldman, L.T. (1990) A latent, nonpathogenic HSV-1-derived vector stably expresses beta-galactosidase in mouse neurons, *Neuron* **5**, 353–360.

Dobson, A.T., Sederati, F., Devi Rao, G., Flanagan, W.M., Farrell, M.J., Stevens, J.G., Wagner, E.K., Feldman, L.T. (1989) Identification of the latency-associated transcript promoter by expression of rabbit beta-globin mRNA in mouse sensory nerve ganglia latently infected with a recombinant herpes simplex virus, *J. Virol.* **63**, 3844–3851.

Dolman, C.L., Clasen, R.A., Dorovini-Zis, K. (1988) Severe cerebral damage in ornithine transcarbamylase deficiency, *Clin. Neuropathol.* **7**, 10–15.

Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D., Naldini, L. (1998) A third-generation lentivirus vector with a conditional packaging system, *J. Virol.* **72**, 8463–8471.

During, M.J., Kaplitt, M.G., Stern, M.B., Eidelberg, D. (2001) Subthalamic GAD gene transfer in Parkinson disease patients who are candidates for deep brain stimulation, *Hum. Gene Ther.* **12**, 1589–1591.

Einfeld, D.A., Schroeder, R., Roelvink, P.W., Lizonova, A., King, C.R., Kovesdi, I., Wickham, T.J. (2001) Reducing the native tropism of adenovirus vectors requires removal of both CAR and integrin interactions, *J. Virol.* **75**, 11284–11291.

Emery, D.W., Yannaki, E., Tubb, J., Stamatoyannopoulos, G. (2000) A chromatin insulator protects retrovirus vectors from chromosomal position effects, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9150–9155.

Emi, N., Friedmann, T., Yee, J.K. (1991) Pseudotype formation of murine leukemia virus with the G protein of vesicular stomatitis virus, *J. Virol.* **65**, 1202–1207.

Ferrari, G., Rossini, S., Giavazzi, R., Maggioni, D., Nobili, N., Soldati, M., Ungers, G., Mavilio, F., Gilboa, E., Bordignon, C. (1991) An in vivo model of somatic cell gene therapy for human severe combined immunodeficiency, *Science* **251**, 1363–1366.

Fields, B.N., Knipe, D.M., Howley, P.M. (1996) *Fields Virology*, 3rd Edition, Lippincott- Raven, Philadelphia, PA.

Fink, J.K., Correll, P.H., Perry, L.K., Brady, R.O., Karlsson, S. (1990) Correction of glucocerebrosidase deficiency after retroviral-mediated gene transfer into hematopoietic progenitor cells from patients with Gaucher disease, *Proc. Natl. Acad. Sci. U. S. A.* **87**, 2334–2338.

Fink, D.J., Sternberg, L.R., Weber, P.C., Mata, M., Goins, W.F., Glorioso, J.C. (1992) In vivo expression of beta-galactosidase in hippocampal neurons by HSV-mediated gene transfer, *Hum. Gene Ther.* **3**, 11–19.

Frenkel, N., Singer, O., Kwong, A.D. (1994) Minireview: the herpes simplex virus amplicon–a versatile defective virus vector, *Gene. Ther.* **1**(Suppl 1), S40–S46.

Friedman, R.L. (1985) Expression of human adenosine deaminase using a transmissible murine retrovirus vector system, *Proc. Natl. Acad. Sci. U. S. A.* **82**, 703–707.

Furlan, R., Poliani, P.L., Marconi, P.C., Bergami, A., Ruffini, F., Adorini, L., Glorioso, J.C., Comi, G., Martino, G. (2001) Central nervous system gene therapy with interleukin-4 inhibits progression of ongoing relapsing-remitting autoimmune encephalomyelitis in Biozzi AB/H mice, *Gene Ther.* **8**, 13–19.

Gansbacher, B. (2003) Report of a second serious adverse event in a clinical trial of gene therapy for X-linked severe combined immune deficiency (X-SCID). Position of the European Society of Gene Therapy (ESGT), *J. Gene. Med.* **5**, 261–262.

Gill, S.S., Patel, N.K., Hotton, G.R., O'Sullivan, K., McCarter, R., Bunnage, M., Brooks, D.J., Svendsen, C.N., Heywood, P. (2003) Direct brain infusion of glial cell line-derived neurotrophic factor in Parkinson disease, *Nat. Med.* **9**, 589–595.

Giraud, C., Winocour, E., Berns, K.I. (1994) Site-specific integration by adeno-associated virus is directed by a cellular DNA sequence, *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10039–10043.

Girod, A., Ried, M., Wobus, C., Lahm, H., Leike, K., Kleinschmidt, J., Deleage, G., Hallek, M. (1999) Genetic capsid modifications allow efficient re-targeting of adeno-associated virus type 2, *Nat. Med.* **5**, 1438.

Goins, W.F., Lee, K.A., Cavalcoli, J.D., O'Malley, M.E., DeKosky, S.T., Fink, D.J., Glorioso, J.C. (1999) Herpes simplex virus type 1 vector-mediated expression of nerve growth factor protects dorsal root ganglion neurons from peroxide toxicity, *J. Virol.* **73**, 519–532.

Goins, W.F., Sternberg, L.R., Croen, K.D., Krause, P.R., Hendricks, R.L., Fink, D.J., Straus, S.E., Levine, M., Glorioso, J.C. (1994) A novel latency-active promoter is contained within the herpes simplex virus type 1 UL flanking repeats, *J. Virol.* **68**, 2239–2252.

Goins, W.F., Yoshimura, N., Phelan, M.W., Yokoyama, T., Fraser, M.O., Ozawa, H., Bennett, N.J., de Groat, W.C., Glorioso, J.C., Chancellor, M.B. (2001) Herpes simplex virus mediated nerve growth factor expression in bladder and afferent neurons: potential treatment for diabetic bladder dysfunction, *J. Urol.* **165**, 1748–1754.

Gordon, Y.J., Johnson, B., Romanowski, E., Araullo Cruz, T. (1988) RNA complementary to herpes simplex virus type 1 ICP0 gene demonstrated in neurons of human trigeminal ganglia, *J. Virol.* **62**, 1832–1835.

Goss, J.R., Goins, W.F., Lacomis, D., Mata, M., Glorioso, J.C., Fink, D.J. (2002) Herpes simplex-mediated gene transfer of nerve growth factor protects against peripheral neuropathy in streptozotocin-induced diabetes in the mouse, *Diabetes* **51**, 2227–2232.

Goss, J.R., Mata, M., Goins, W.F., Wu, H.H., Glorioso, J.C., Fink, D.J. (2001) Antinociceptive effect of a genomic herpes simplex virus-based vector expressing human proenkephalin in rat dorsal root ganglion, *Gene Ther.* **8**, 551–556.

Grines, C.L., Watkins, M.W., Helmer, G., Penny, W., Brinker, J., Marmur, J.D., West, A., Rade, J.J., Marrott, P., Hammond, H.K., Engler, R.L. (2002) Angiogenic Gene Therapy (AGENT) trial in patients with stable angina pectoris, *Circulation* **105**, 1291–1297.

Guillot, C., Guillonneau, C., Mathieu, P., Gerdes, C.A., Menoret, S., Braudeau, C., Tesson, L., Renaudin, K., Castro, M.G., Lowenstein, P.R., Anegon, I. (2002) Prolonged blockade of CD40-CD40 ligand interactions by gene transfer of CD40Ig results in long-term heart allograft survival and donor-specific

hyporesponsiveness, but does not prevent chronic rejection, *J. Immunol.* **168**, 1600–1609.

Guillot, C., Mathieu, P., Coathalem, H., Le Mauff, B., Castro, M.G., Tesson, L., Usal, C., Laumonier, T., Brouard, S., Soulillou, J.P., Lowenstein, P.R., Cuturi, M.C., Anegon, I. (2000) Tolerance to cardiac allografts via local and systemic mechanisms after adenovirus-mediated CTLA4Ig expression, *J. Immunol.* **164**, 5258–5268.

Hacein-Bey, S., Basile, G.D., Lemerle, J., Fischer, A., Cavazzana-Calvo, M. (1998) gammac gene transfer in the presence of stem cell factor, FLT-3L, interleukin-7 (IL-7), IL-1, and IL-15 cytokines restores T-cell differentiation from gammac(-) X-linked severe combined immunodeficiency hematopoietic progenitor cells in murine fetal thymic organ cultures, *Blood* **92**, 4090–4097.

Hacein-Bey, H., Cavazzana-Calvo, M., Le Deist, F., Dautry-Varsat, A., Hivroz, C., Riviere, I., Danos, O., Heard, J.M., Sugamura, K., Fischer, A., De Saint Basile, G. (1996) Gammac gene transfer into SCID X1 patients' B-cell lines restores normal high-affinity interleukin-2 receptor expression and function, *Blood* **87**, 3108–3116.

Hacein-Bey-Abina, S., von Kalle, C., Schmidt, M., Le Deist, F., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J.L., Fraser, C.C., Cavazzana-Calvo, M., Fischer, A. (2003) A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency, *N. Engl. J. Med.* **348**, 255–256.

Hacein-Bey-Abina, S., Le Deist, F., Carlier, F., Bouneaud, C., Hue, C., De Villartay, J.P., Thrasher, A.J., Wulffraat, N., Sorensen, R., Dupuis-Girod, S., Fischer, A., Davies, E.G., Kuis, W., Leiva, L., Cavazzana-Calvo, M. (2002) Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy, *N. Engl. J. Med.* **346**, 1185–1193.

Haj-Ahmad, Y., Graham, F.L. (1986) Development of a helper-independent human adenovirus vector and its use in the transfer of the herpes simplex virus thymidine kinase gene, *J. Virol.* **57**, 267–274.

Hao, S., Mata, M., Goins, W., Glorioso, J.C., Fink, D.J. (2003) Transgene-mediated enkephalin release enhances the effect of morphine and evades tolerance to produce a sustained antiallodynic effect in neuropathic pain, *Pain* **102**, 135–142.

Harding, B.N., Leonard, J.V., Erdohazi, M. (1984) Ornithine carbamoyl transferase deficiency: a neuropathological study, *Eur. J. Pediatr.* **141**, 215–220.

Heim, D.A., Hanazono, Y., Giri, N., Wu, T., Childs, R., Sellers, S.E., Muul, L., Agricola, B.A., Metzger, M.E., Donahue, R.E., Tisdale, J.F., Dunbar, C.E. (2000) Introduction of a xenogeneic gene via hematopoietic stem cells leads to specific tolerance in a rhesus monkey model, *Mol. Ther.* **1**, 533–544.

Hermonat, P.L., Muzyczka, N. (1984) Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells, *Proc. Natl. Acad. Sci. U. S. A.* **81**, 6466–6470.

Hershfield, M.S. (1995) PEG-ADA: an alternative to haploidentical bone marrow transplantation and an adjunct to gene therapy for adenosine deaminase deficiency, *Hum. Mutat.* **5**, 107–112.

Hershfield, M.S., Chaffee, S., Sorensen, R.U. (1993) Enzyme replacement therapy with polyethylene glycol-adenosine deaminase in adenosine deaminase deficiency: overview and case reports of three patients, including two now receiving gene therapy, *Pediatr. Res.* **33**, S42–S47; discussion S47-8.

Hofmann, A., Nolan, G.P., Blau, H.M. (1996) Rapid retroviral delivery of tetracycline-inducible genes in a single autoregulatory cassette, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 5185–5190.

Hoppe, U.C., Marban, E., Johns, D.C. (2000) Adenovirus-mediated inducible gene expression in vivo by a hybrid ecdysone receptor, *Mol. Ther.* **1**, 159–164.

Hyde, S.C., Southern, K.W., Gileadi, U., Fitzjohn, E.M., Mofford, K.A., Waddell, B.E., Gooi, H.C., Goddard, C.A., Hannavy, K., Smyth, S.E., Egan, J.J., Sorgi, F.L., Huang, L., Cuthbert, A.W., Evans, M.J., Colledge, W.H., Higgins, C.F., Webb, A.K., Gill, D.R. (2000) Repeat administration of DNA/liposomes to the nasal epithelium of patients with cystic fibrosis, *Gene Ther.* **7**, 1156–1165.

Iida, A., Chen, S.T., Friedmann, T., Yee, J.K. (1996) Inducible gene expression by retrovirus-mediated transfer of a modified tetracycline-regulated system, *J. Virol.* **70**, 6054–6059.

Im, S.A., Gomez Manzano, C., Fueyo, J., Liu, T.J., Ke, L.D., Kim, J.S., Lee, H.Y., Steck, P.A., Kyritsis, A.P., Yung, W.K. (1999)

Antiangiogenesis treatment for gliomas: transfer of antisense-vascular endothelial growth factor inhibits tumor growth in vivo, *Cancer Res.* **59**, 895–900.

Kaiser, J. (2003) Gene therapy. Seeking the cause of induced leukemias in X-SCID trial, *Science* **299**, 495.

Kanno, H., Hattori, S., Sato, H., Murata, H., Huang, F.H., Hayashi, A., Suzuki, N., Yamamoto, I., Kawamoto, S., Minami, M., Miyatake, S., Shuin, T., Kaplitt, M.G. (1999) Experimental gene therapy against subcutaneously implanted glioma with a herpes simplex virus-defective vector expressing interferon-gamma, *Cancer Gene Ther.* **6**, 147–154.

Kantoff, P.W., Kohn, D.B., Mitsuya, H., Armentano, D., Sieberg, M., Zwiebel, J.A., Eglitis, M.A., McLachlin, J.R., Wiginton, D.A., Hutton, J.J., et al. (1986) Correction of adenosine deaminase deficiency in cultured human T and B cells by retrovirus-mediated gene transfer, *Proc. Natl. Acad. Sci. U. S. A.* **83**, 6563–6567.

Kashentseva, E.A., Seki, T., Curiel, D.T., Dmitriev, I.P. (2002) Adenovirus targeting to c-erbB-2 oncoprotein by single-chain antibody fused to trimeric form of adenovirus receptor ectodomain, *Cancer Res.* **62**, 609–616.

Kawada, T., Nakazawa, M., Nakauchi, S., Yamazaki, K., Shimamoto, R., Urabe, M., Nakata, J., Hemmi, C., Masui, F., Nakajima, T., Suzuki, J., Monahan, J., Sato, H., Masaki, T., Ozawa, K., Toyo-Oka, T. (2002) Rescue of hereditary form of dilated cardiomyopathy by rAAV-mediated somatic gene therapy: amelioration of morphological findings, sarcolemmal permeability, cardiac performances, and the prognosis of TO-2 hamsters, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 901–906.

Kay, M.A., Manno, C.S., Ragni, M.V., Larson, P.J., Couto, L.B., McClelland, A., Glader, B., Chew, A.J., Tai, S.J., Herzog, R.W., Arruda, V., Johnson, F., Scallan, C., Skarsgard, E., Flake, A.W., High, K.A. (2000) Evidence for gene transfer and expression of factor IX in haemophilia B patients treated with an AAV vector, *Nat. Genet.* **24**, 257–261.

Keir, S.D., Mitchell, W.J., Feldman, L.T., Martin, J.R. (1995) Targeting and gene expression in spinal cord motor neurons following intramuscular inoculation of an HSV-1 vector, *J. Neurovirol.* **1**, 259–267.

Khan, T.A., Sellke, F.W., Laham, R.J. (2003) Gene therapy progress and prospects: therapeutic angiogenesis for limb and myocardial ischemia, *Gene Ther.* **10**, 285–291.

Kim, S.H., Chung, J.M. (1992) An experimental model for peripheral neuropathy produced by segmental spinal nerve ligation in the rat, *Pain* **50**, 355–363.

Kim, V.N., Mitrophanous, K., Kingsman, S.M., Kingsman, A.J. (1998) Minimal requirement for a lentivirus vector based on human immunodeficiency virus type 1, *J. Virol.* **72**, 811–816.

Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y., Shimizu, N. (1998) Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism, *Nature* **392**, 605–608.

Knowles, M.R., Hohneker, K.W., Zhou, Z., Olsen, J.C., Noah, T.L., Hu, P.C., Leigh, M.W., Engelhardt, J.F., Edwards, L.J., Jones, K.R., et al. (1995) A controlled study of adenoviral-vector-mediated gene transfer in the nasal epithelium of patients with cystic fibrosis, *N. Engl. J. Med.* **333**, 823–831.

Kochanek, S., Schiedner, G., Volpers, C. (2001) High-capacity 'gutless' adenoviral vectors, *Curr. Opin. Mol. Ther.* **3**, 454–463.

Koeberl, D.D., Alexander, I.E., Halbert, C.L., Russell, D.W., Miller, A.D. (1997) Persistent expression of human clotting factor IX from mouse liver after intravenous injection of adeno-associated virus vectors, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1426–1431.

Kohn, D.B., Sadelain, M., Glorioso, J.C. (2003) Occurrence of leukaemia following gene therapy of X-linked SCID, *Nat. Rev. Cancer* **3**, 477–488.

Koponen, J.K., Kankkonen, H., Kannasto, J., Wirth, T., Hillen, W., Bujard, H., Yla-Herttuala, S. (2003) Doxycycline-regulated lentiviral vector system with a novel reverse transactivator rtTA2S-M2 shows a tight control of gene expression in vitro and in vivo, *Gene Ther.* **10**, 459–466.

Kordower, J.H., Emborg, M.E., Bloch, J., Ma, S.Y., Chu, Y., Leventhal, L., McBride, J., Chen, E.Y., Palfi, S., Roitberg, B.Z., Brown, W.D., Holden, J.E., Pyzalski, R., Taylor, M.D., Carvey, P., Ling, Z., Trono, D., Hantraye, P., Deglon, N., Aebischer, P. (2000) Neurodegeneration prevented by lentiviral vector delivery of GDNF in primate models of Parkinson's disease, *Science* **290**, 767–773.

Krisky, D.M., Marconi, P.C., Oligino, T., Rouse, R.J., Fink, D.J., Glorioso, J.C. (1997) Rapid method for construction of recombinant HSV gene transfer vectors, *Gene Ther.* **4**, 1120–1125.

Krisky, D.M., Marconi, P.C., Oligino, T.J., Rouse, R.J., Fink, D.J., Cohen, J.B., Watkins, S.C., Glorioso, J.C. (1998) Development of herpes simplex virus replication-defective multigene vectors for combination gene therapy applications, *Gene Ther.* **5**, 1517–1530.

Krisky, D.M., Wolfe, D., Goins, W.F., Marconi, P.C., Ramakrishnan, R., Mata, M., Rouse, R.J., Fink, D.J., Glorioso, J.C. (1998) Deletion of multiple immediate-early genes from herpes simplex virus reduces cytotoxicity and permits long-term gene expression in neurons, *Gene Ther.* **5**, 1593–1603.

Lachmann, R.H., Efstathiou, S. (1997) Utilization of the herpes simplex virus type 1 latency-associated regulatory region to drive stable reporter gene expression in the nervous system, *J. Virol.* **71**, 3197–3207.

Lamartina, S., Silvi, L., Roscilli, G., Casimiro, D., Simon, A.J., Davies, M.E., Shiver, J.W., Rinaudo, C.D., Zampaglione, I., Fattori, E., Colloca, S., Gonzalez Paz, O., Laufer, R., Bujard, H., Cortese, R., Ciliberto, G., Toniatti, C. (2003) Construction of an rtTA2(s)-m2/tts(kid)-based transcription regulatory switch that displays no basal activity, good inducibility, and high responsiveness to doxycycline in mice and non-human primates, *Mol. Ther.* **7**, 271–280.

Lang, F.F., Yung, W.K., Sawaya, R., Tofilon, P.J. (1999) Adenovirus-mediated p53 gene therapy for human gliomas, *Neurosurgery* **45**, 1093–1104.

Le Gal La Salle, G., Robert, J.J., Berrard, S., Ridoux, V., Stratford-Perricaudet, L.D., Perricaudet, M., Mallet, J. (1993) An adenovirus vector for gene transfer into neurons and glia in the brain, *Science* **259**, 988–990.

Ledley, F.D., Grenett, H.E., McGinnis-Shelnutt, M., Woo, S.L. (1986) Retroviral-mediated gene transfer of human phenylalanine hydroxylase into NIH 3T3 and hepatoma cells, *Proc. Natl. Acad. Sci. U. S. A.* **83**, 409–413.

Lee, L.Y., Patel, S.R., Hackett, N.R., Mack, C.A., Polce, D.R., El-Sawy, T., Hachamovitch, R., Zanzonico, P., Sanborn, T.A., Parikh, M., Isom, O.W., Crystal, R.G., Rosengart, T.K. (2000) Focal angiogen therapy using intramyocardial delivery of an adenovirus vector coding for vascular endothelial growth factor 121, *Ann. Thorac. Surg.* **69**, 14–23; discussion 23-4.

Li, X., Mukai, T., Young, D., Frankel, S., Law, P., Wong Staal, F. (1998) Transduction of CD34+ cells by a vesicular stomach virus protein G (VSV-G) pseudotyped HIV-1 vector. Stable gene expression in progeny cells, including dendritic cells, *J. Hum. Virol.* **1**, 346–352.

Lim, B., Apperley, J.F., Orkin, S.H., Williams, D.A. (1989) Long-term expression of human adenosine deaminase in mice transplanted with retrovirus-infected hematopoietic stem cells, *Proc. Natl. Acad. Sci. U. S. A.* **86**, 8892–8896.

Linden, R.M., Ward, P., Giraud, C., Winocour, E., Berns, K.I. (1996) Site-specific integration by adeno-associated virus, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 11288–11294.

Loeser, J., Butler, S., Chapman, C., Turk, D. (2001) *Bonica's Management of Pain*, 3rd Edition, Lippincott Williams & Wilkins, Philadelphia, PA.

Loiler, S.A., Conlon, T.J., Song, S., Tang, Q., Warrington, K.H., Agarwal, A., Kapturczak, M., Li, C., Ricordi, C., Atkinson, M.A., Muzyczka, N., Flotte, T.R. (2003) Targeting recombinant adeno-associated virus vectors to enhance gene transfer to pancreatic islets and liver, *Gene Ther.* **10**, 1551–1558.

Losordo, D.W., Vale, P.R., Hendel, R.C., Milliken, C.E., Fortuin, F.D., Cummings, N., Schatz, R.A., Asahara, T., Isner, J.M., Kuntz, R.E. (2002) Phase 1/2 placebo-controlled, double-blind, dose-escalating trial of myocardial vascular endothelial growth factor 2 gene transfer by catheter delivery in patients with chronic myocardial ischemia, *Circulation* **105**, 2012–2018.

Luo, J., Kaplitt, M.G., Fitzsimons, H.L., Zuzga, D.S., Liu, Y., Oshinsky, M.L., During, M.J. (2002) Subthalamic GAD gene therapy in a Parkinson's disease rat model, *Science* **298**, 425–429.

Machein, M.R., Kullmer, J., Fiebich, B.L., Plate, K.H., Warnke, P.C. (1999) Vascular endothelial growth factor expression, vascular volume, and, capillary permeability in human brain tumors, *Neurosurgery* **44**, 732–740.

Mack, C.A., Song, W.R., Carpenter, H., Wickham, T.J., Kovesdi, I., Harvey, B.G., Magovern, C.J., Isom, O.W., Rosengart, T., Falck-Pedersen, E., Hackett, N.R., Crystal, R.G.,

Mastrangeli, A. (1997) Circumvention of anti-adenovirus neutralizing immunity by administration of an adenoviral vector of an alternate serotype, *Hum. Gene Ther.* **8**, 99–109.

Maestri, N.E., Clissold, D., Brusilow, S.W. (1999) Neonatal onset ornithine transcarbamylase deficiency: A retrospective analysis, *J. Pediatr.* **134**, 268–272.

Maestri, N.E., Brusilow, S.W., Clissold, D.B., Bassett, S.S. (1996) Long-term treatment of girls with ornithine transcarbamylase deficiency, *N. Engl. J. Med.* **335**, 855–859.

Manno, C.S., Chew, A.J., Hutchison, S., Larson, P.J., Herzog, R.W., Arruda, V.R., Tai, S.J., Ragni, M.V., Thompson, A., Ozelo, M., Couto, L.B., Leonard, D.G., Johnson, F.A., McClelland, A., Scallan, C., Skarsgard, E., Flake, A.W., Kay, M.A., High, K.A., Glader, B. (2003) AAV-mediated factor IX gene transfer to skeletal muscle in patients with severe hemophilia B, *Blood* **101**, 2963–2972.

Marconi, P., Tamura, M., Moriuchi, S., Krisky, D.M., Niranjan, A., Goins, W.F., Cohen, J.B., Glorioso, J.C. (2000) Connexin 43-enhanced suicide gene therapy using herpesviral vectors, *Mol. Ther.* **1**, 71–81.

Markowitz, D., Goff, S., Bank, A. (1988) Construction and use of a safe and efficient amphotropic packaging cell line, *Virology* **167**, 400–406.

Mastrangeli, A., Harvey, B.G., Yao, J., Wolff, G., Kovesdi, I., Crystal, R.G., Falck-Pedersen, E. (1996) ''Sero-switch'' adenovirus-mediated in vivo gene transfer: circumvention of anti-adenovirus humoral immune defenses against repeat adenovirus vector administration by changing the adenovirus serotype, *Hum. Gene Ther.* **7**, 79–87.

Mazarakis, N.D., Azzouz, M., Rohll, J.B., Ellard, F.M., Wilkes, F.J., Olsen, A.L., Carter, E.E., Barber, R.D., Baban, D.F., Kingsman, S.M., Kingsman, A.J., O'Malley, K., Mitrophanous, K.A. (2001) Rabies virus glycoprotein pseudotyping of lentiviral vectors enables retrograde axonal transport and access to the nervous system after peripheral delivery, *Hum. Mol. Genet.* **10**, 2109–2121.

McLaughlin, S.K., Collis, P., Hermonat, P.L., Muzyczka, N. (1988) Adeno-associated virus general transduction vectors: analysis of proviral structures, *J. Virol.* **62**, 1963–1973.

Mercier, S., Gahery-Segard, H., Monteil, M., Lengagne, R., Guillet, J.G., Eloit, M., Denesvre, C. (2002) Distinct roles of adenovirus vector-transduced dendritic cells, myoblasts, and endothelial cells in mediating an immune response against a transgene product, *J. Virol.* **76**, 2899–2911.

Miao, C.H., Ohashi, K., Patijn, G.A., Meuse, L., Ye, X., Thompson, A.R., Kay, M.A. (2000) Inclusion of the hepatic locus control region, an intron, and untranslated region increases and stabilizes hepatic factor IX gene expression in vivo but not in vitro, *Mol. Ther.* **1**, 522–532.

Min, W.P., Gorczynski, R., Huang, X.Y., Kushida, M., Kim, P., Obataki, M., Lei, J., Suri, R.M., Cattral, M.S. (2000) Dendritic cells genetically engineered to express Fas ligand induce donor-specific hyporesponsiveness and prolong allograft survival, *J. Immunol.* **164**, 161–167.

Miyanohara, A., Sharkey, M.F., Witztum, J.L., Steinberg, D., Friedmann, T. (1988) Efficient expression of retroviral vector-transduced human low density lipoprotein (LDL) receptor in LDL receptor-deficient rabbit fibroblasts in vitro, *Proc. Natl. Acad. Sci. U. S. A.* **85**, 6538–6542.

Miyoshi, H., Takahashi, M., Gage, F.H., Verma, I.M. (1997) Stable and efficient gene transfer into the retina using an HIV-based lentiviral vector, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 10319–10323.

Miyoshi, H., Blomer, U., Takahashi, M., Gage, F.H., Verma, I.M. (1998) Development of a self-inactivating lentivirus vector, *J. Virol.* **72**, 8150–8157.

Miyoshi, H., Smith, K.A., Mosier, D.E., Verma, I.M., Torbett, B.E. (1999) Transduction of human CD34+ cells that mediate long-term engraftment of NOD/SCID mice by HIV vectors, *Science* **283**, 682–686.

Mizuguchi, H., Xu, Z.L., Sakurai, F., Mayumi, T., Hayakawa, T. (2003) Tight positive regulation of transgene expression by a single adenovirus vector containing the rtTA and tTS expression cassettes in separate genome regions, *Hum. Gene Ther.* **14**, 1265–1277.

Mizuguchi, H., Koizumi, N., Hosono, T., Utoguchi, N., Watanabe, Y., Kay, M.A., Hayakawa, T. (2001) A simplified system for constructing recombinant adenoviral vectors containing heterologous peptides in the HI loop of their fiber knob, *Gene Ther.* **8**, 730–735.

Molinier-Frenkel, V., Gahery-Segard, H., Mehtali, M., Le Boulaire, C., Ribault, S., Boulanger, P., Tursz, T., Guillet, J.G., Farace, F. (2000) Immune response to recombinant

adenovirus in humans: capsid components from viral input are targets for vector-specific cytotoxic T lymphocytes, *J. Virol.* **74**, 7678–7682.

Moriuchi, S., Oligino, T., Krisky, D., Marconi, P., Fink, D., Cohen, J., Glorioso, J.C. (1998) Enhanced tumor cell killing in the presence of ganciclovir by herpes simplex virus type 1 vector-directed coexpression of human tumor necrosis factor-alpha and herpes simplex virus thymidine kinase, *Cancer Res.* **58**, 5731–5737.

Morral, N., O'Neal, W., Rice, K., Leland, M., Kaplan, J., Piedra, P.A., Zhou, H., Parks, R.J., Velji, R., Aguilar-Cordova, E., Wadsworth, S., Graham, F.L., Kochanek, S., Carey, K.D., Beaudet, A.L. (1999) Administration of helper-dependent adenoviral vectors and sequential delivery of different vector serotype for long-term liver-directed gene transfer in baboons, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 12816–12821.

Mount, J.D., Herzog, R.W., Tillson, D.M., Goodman, S.A., Robinson, N., McCleland, M.L., Bellinger, D., Nichols, T.C., Arruda, V.R., Lothrop, C.D., Jr., High, K.A. (2002) Sustained phenotypic correction of hemophilia B dogs with a factor IX null mutation by liver-directed gene therapy, *Blood* **99**, 2670–2676.

Muller, O.J., Kaul, F., Weitzman, M.D., Pasqualini, R., Arap, W., Kleinschmidt, J.A., Trepel, M. (2003) Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors, *Nat. Biotechnol.* **21**, 1040–1046.

Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M., Trono, D. (1996) In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector, *Science* **272**, 263–267.

Nathwani, A.C., Davidoff, A.M., Hanawa, H., Hu, Y., Hoffer, F.A., Nikanorov, A., Slaughter, C., Ng, C.Y., Zhou, J., Lozier, J.N., Mandrell, T.D., Vanin, E.F., Nienhuis, A.W. (2002) Sustained high-level expression of human factor IX (hFIX) after liver-targeted delivery of recombinant adeno-associated virus encoding the hFIX gene in rhesus macaques, *Blood* **100**, 1662–1669.

Niranjan, A., Moriuchi, S., Lunsford, L.D., Kondziolka, D., Flickinger, J.C., Fellows, W., Rajendiran, S., Tamura, M., Cohen, J.B., Glorioso, J.C. (2000) Effective treatment of experimental glioblastoma by HSV vector-mediated TNFalpha and HSV-tk gene transfer in combination with radiosurgery and ganciclovir administration, *Mol. Ther.* **2**, 114–120.

No, D., Yao, T.P., Evans, R.M. (1996) Ecdysone-inducible gene expression in mammalian cells and transgenic mice, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 3346–3351.

Noone, P.G., Hohneker, K.W., Zhou, Z., Johnson, L.G., Foy, C., Gipson, C., Jones, K., Noah, T.L., Leigh, M.W., Schwartzbach, C., Efthimiou, J., Pearlman, R., Boucher, R.C., Knowles, M.R. (2000) Safety and biological efficacy of a lipid-CFTR complex for gene transfer in the nasal epithelium of adult patients with cystic fibrosis, *Mol. Ther.* **1**, 105–114.

Oligino, T., Poliani, P.L., Wang, Y., Tsai, S.Y., O'Malley, B.W., Fink, D.J., Glorioso, J.C. (1998) Drug inducible transgene expression in brain using a herpes simplex virus vector, *Gene Ther.* **5**, 491–496.

Palmer, T.D., Hock, R.A., Osborne, W.R., Miller, A.D. (1987) Efficient retrovirus-mediated transfer and expression of a human adenosine deaminase gene in diploid skin fibroblasts from an adenosine deaminase-deficient human, *Proc. Natl. Acad. Sci. U. S. A.* **84**, 1055–1059.

Palmer, J.A., Branston, R.H., Lilley, C.E., Robinson, M.J., Groutsi, F., Smith, J., Latchman, D.S., Coffin, R.S. (2000) Development and optimization of herpes simplex virus vectors for multiple long-term gene delivery to the peripheral nervous system, *J. Virol.* **74**, 5604–5618.

Park, F., Ohashi, K., Kay, M.A. (2000) Therapeutic levels of human factor VIII and IX using HIV-1-based lentiviral vectors in mouse liver, *Blood* **96**, 1173–1176.

Parker, J.N., Gillespie, G.Y., Love, C.E., Randall, S., Whitley, R.J., Markert, J.M. (2000) From the cover: engineered herpes simplex virus expressing IL-12 in the treatment of experimental murine brain tumors, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 2208–2213.

Paulus, W., Baur, I., Boyce, F.M., Breakefield, X.O., Reeves, S.A. (1996) Self-contained, tetracycline-regulated retroviral vector system for gene delivery to mammalian cells, *J. Virol.* **70**, 62–67.

Pawliuk, R., Westerman, K.A., Fabry, M.E., Payen, E., Tighe, R., Bouhassira, E.E., Acharya, S.A., Ellis, J., London, I.M., Eaves, C.J., Humphries, R.K., Beuzard, Y., Nagel, R.L., Leboulch, P. (2001) Correction of sickle cell

disease in transgenic mouse models by gene therapy, *Science* **294**, 2368–2371.

Perricone, M.A., Morris, J.E., Pavelka, K., Plog, M.S., O'Sullivan, B.P., Joseph, P.M., Dorkin, H., Lapey, A., Balfour, R., Meeker, D.P., Smith, A.E., Wadsworth, S.C., St George, J.A. (2001) Aerosol and lobar administration of a recombinant adenovirus to individuals with cystic fibrosis. II. Transfection efficiency in airway epithelium, *Hum. Gene Ther.* **12**, 1383–1394.

Podsakoff, G.M. (2001) Lentiviral vectors approach the clinic but fall back: National Institutes of Health Recombinant DNA Advisory Committee review of a first clinical protocol for use of a lentiviral vector, *Mol. Ther.* **4**, 282–283.

Poliani, P.L., Brok, H., Furlan, R., Ruffini, F., Bergami, A., Desina, G., Marconi, P.C., Rovaris, M., Uccelli, A., Glorioso, J.C., Penna, G., Adorini, L., Comi, G., t Hart, B., Martino, G. (2001) Delivery to the central nervous system of a nonreplicative herpes simplex type 1 vector engineered with the interleukin 4 gene protects rhesus monkeys from hyperacute autoimmune encephalomyelitis, *Hum. Gene Ther.* **12**, 905–920.

Ponnazhagan, S., Erikson, D., Kearns, W.G., Zhou, S.Z., Nahreini, P., Wang, X.S., Srivastava, A. (1997) Lack of site-specific integration of the recombinant adeno-associated virus 2 genomes in human cells, *Hum. Gene Ther.* **8**, 275–284.

Puck, J.M., Deschenes, S.M., Porter, J.C., Dutra, A.S., Brown, C.J., Willard, H.F., Henthorn, P.S. (1993) The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated in X-linked severe combined immunodeficiency, SCIDX1, *Hum. Mol. Genet.* **2**, 1099–1104.

Qian, H.S., Channon, K., Nepliouevа, V., Wang, Q., Finer, M., Tsui, L., George, S.E., McArthur, J. (2001) Improved adenoviral vector for vascular gene therapy: beneficial effects on vascular function and inflammation, *Circ. Res.* **88**, 911–917.

Ragot, T., Vincent, N., Chafey, P., Vigne, E., Gilgenkrantz, H., Couton, D., Cartaud, J., Briand, P., Kaplan, J.C., Perricaudet, M., et al. (1993) Efficient adenovirus-mediated transfer of a human minidystrophin gene to skeletal muscle of mdx mice, *Nature* **361**, 647–650.

Raper, S.E., Yudkoff, M., Chirmule, N., Gao, G.P., Nunes, F., Haskal, Z.J., Furth, E.E., Propert, K.J., Robinson, M.B., Magosin, S., Simoes, H., Speicher, L., Hughes, J., Tazelaar, J., Wivel, N.A., Wilson, J.M., Batshaw, M.L. (2002) A pilot study of in vivo liver-directed gene transfer with an adenoviral vector in partial ornithine transcarbamylase deficiency, *Hum. Gene Ther.* **13**, 163–175.

Reddy, P.S., Sakhuja, K., Ganesh, S., Yang, L., Kayda, D., Brann, T., Pattison, S., Golightly, D., Idamakanti, N., Pinkstaff, A., Kaloss, M., Barjot, C., Chamberlain, J.S., Kaleko, M., Connelly, S. (2002) Sustained human factor VIII expression in hemophilia A mice following systemic delivery of a gutless adenoviral vector, *Mol. Ther.* **5**, 63–73.

Reiser, J., Harmison, G., Kluepfel-Stahl, S., Brady, R.O., Karlsson, S., Schubert, M. (1996) Transduction of nondividing cells using pseudotyped defective high-titer HIV type 1 particles, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 15266–15271.

Rivera, V.M., Clackson, T., Natesan, S., Pollock, R., Amara, J.F., Keenan, T., Magari, S.R., Phillips, T., Courage, N.L., Cerasoli, F., Jr., Holt, D.A., Gilman, M. (1996) A humanized system for pharmacologic control of gene expression, *Nat. Med.* **2**, 1028–1032.

Rock, D.L., Nesburn, A.B., Ghiasi, H., Ong, J., Lewis, T.L., Lokensgard, J.R., Wechsler, S.L. (1987) Detection of latency-related viral RNAs in trigeminal ganglia of rabbits latently infected with herpes simplex virus type 1, *J. Virol.* **61**, 3820–3826.

Rosen, C.A., Sodroski, J.G., Haseltine, W.A. (1985) The location of cis-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat, *Cell* **41**, 813–823.

Rosenfeld, M.A., Siegfried, W., Yoshimura, K., Yoneyama, K., Fukayama, M., Stier, L.E., Paakko, P.K., Gilardi, P., Stratford-Perricaudet, L.D., Perricaudet, M., et al. (1991) Adenovirus-mediated transfer of a recombinant alpha 1-antitrypsin gene to the lung epithelium in vivo, *Science* **252**, 431–434.

Rosenfeld, M.A., Yoshimura, K., Trapnell, B.C., Yoneyama, K., Rosenthal, E.R., Dalemans, W., Fukayama, M., Bargon, J., Stier, L.E., Stratford-Perricaudet, L., et al. (1992) In vivo transfer of the human cystic fibrosis transmembrane conductance regulator gene to the airway epithelium, *Cell* **68**, 143–155.

Rosengart, T.K., Lee, L.Y., Patel, S.R., Sanborn, T.A., Parikh, M., Bergman, G.W., Hachamovitch, R., Szulc, M., Kligfield, P.D., Okin, P.M., Hahn, R.T., Devereux, R.B., Post, M.R., Hackett, N.R., Foster, T., Grasso, T.M., Lesser, M.L., Isom, O.W., Crystal, R.G. (1999) Angiogenesis gene therapy: phase I assessment of direct intramyocardial administration of an adenovirus vector expressing VEGF121 cDNA to individuals with clinically significant severe coronary artery disease, *Circulation* **100**, 468–474.

Roth, D.A., Tawa, N.E., Jr., O'Brien, J.M., Treco, D.A., Selden, R.F. (2001) Nonviral transfer of the gene encoding coagulation factor VIII in patients with severe hemophilia A, *N. Engl. J. Med.* **344**, 1735–1742.

Ruffini, F., Furlan, R., Poliani, P.L., Brambilla, E., Marconi, P.C., Bergami, A., Desina, G., Glorioso, J.C., Comi, G., Martino, G. (2001) Fibroblast growth factor-II gene therapy reverts the clinical course and the pathological signs of chronic experimental autoimmune encephalomyelitis in C57BL/6 mice, *Gene Ther.* **8**, 1207–1213.

Ruiz, F.E., Clancy, J.P., Perricone, M.A., Bebok, Z., Hong, J.S., Cheng, S.H., Meeker, D.P., Young, K.R., Schoumacher, R.A., Weatherly, M.R., Wing, L., Morris, J.E., Sindel, L., Rosenberg, M., van Ginkel, F.W., McGhee, J.R., Kelly, D., Lyrene, R.K., Sorscher, E.J. (2001) A clinical inflammatory syndrome attributable to aerosolized lipid-DNA administration in cystic fibrosis, *Hum. Gene Ther.* **12**, 751–761.

Sacks, W.R., Greene, C.C., Aschman, D.P., Schaffer, P.A. (1985) Herpes simplex virus type 1 ICP27 is an essential regulatory protein, *J. Virol.* **55**, 796–805.

Sailaja, G., HogenEsch, H., North, A., Hays, J., Mittal, S.K. (2002) Encapsulation of recombinant adenovirus into alginate microspheres circumvents vector-specific immune response, *Gene Ther.* **9**, 1722–1729.

Sakhuja, K., Reddy, P.S., Ganesh, S., Cantaniag, F., Pattison, S., Limbach, P., Kayda, D.B., Kadan, M.J., Kaleko, M., Connelly, S. (2003) Optimization of the generation and propagation of gutless adenoviral vectors, *Hum. Gene Ther.* **14**, 243–254.

Samulski, R.J., Zhu, X., Xiao, X., Brook, J.D., Housman, D.E., Epstein, N., Hunter, L.A. (1991) Targeted integration of adeno-associated virus (AAV) into human chromosome 19, *EMBO J.* **10**, 3941–3950.

Sarkar, N., Ruck, A., Kallner, G.S.Y.H., Blomberg, P., Islam, K.B., van der Linden, J., Lindblom, D., Nygren, A.T., Lind, B., Brodin, L.A., Drvota, V., Sylven, C. (2001) Effects of intramyocardial injection of phVEGF-A165 as sole therapy in patients with refractory coronary artery disease–12-month follow-up: angiogenic gene therapy, *J. Intern. Med.* **250**, 373–381.

Scherr, M., Battmer, K., Blomer, U., Schiedlmeier, B., Ganser, A., Grez, M., Eder, M. (2002) Lentiviral gene transfer into peripheral blood-derived CD34+ NOD/SCID-repopulating cells, *Blood* **99**, 709–712.

Schlegel, R., Tralka, T.S., Willingham, M.C., Pastan, I. (1983) Inhibition of VSV binding and infectivity by phosphatidylserine: is phosphatidylserine a VSV-binding site? *Cell* **32**, 639–646.

Schmeisser, F., Donohue, M., Weir, J.P. (2002) Tetracycline-regulated gene expression in replication-incompetent herpes simplex virus vectors, *Hum. Gene Ther.* **13**, 2113–2124.

Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., Bushman, F. (2002) HIV-1 integration in the human genome favors active genes and local hotspots, *Cell* **110**, 521–529.

Schwei, M.J., Honore, P., Rogers, S.D., Salak-Johnson, J.L., Finke, M.P., Ramnaraine, M.L., Clohisy, D.R., Mantyh, P.W. (1999) Neurochemical and cellular reorganization of the spinal cord in a murine model of bone cancer pain, *J. Neurosci.* **19**, 10886–10897.

Seshidhar Reddy, P., Ganesh, S., Limbach, M.P., Brann, T., Pinkstaff, A., Kaloss, M., Kaleko, M., Connelly, S. (2003) Development of adenovirus serotype 35 as a gene transfer vector, *Virology* **311**, 384–393.

Shapiro, E., Krivit, W., Lockman, L., Jambaque, I., Peters, C., Cowan, M., Harris, R., Blanche, S., Bordigoni, P., Loes, D., Ziegler, R., Crittenden, M., Ris, D., Berg, B., Cox, C., Moser, H., Fischer, A., Aubourg, P. (2000) Long-term effect of bone-marrow transplantation for childhood-onset cerebral X-linked adrenoleukodystrophy, *Lancet* **356**, 713–718.

Shi, W., Bartlett, J.S. (2003) RGD inclusion in VP3 provides adeno-associated virus type 2

(AAV2)-based vectors with a heparan sulfate-independent cell entry mechanism, *Mol. Ther.* **7**, 515–525.

Smith, C., Lachmann, R.H., Efstathiou, S. (2000) Expression from the herpes simplex virus type 1 latency-associated promoter in the murine central nervous system, *J. Gen. Virol.* **3**, 649–662.

Snyder, R.O., Flotte, T.R. (2002) Production of clinical-grade recombinant adeno-associated virus vectors, *Curr. Opin. Biotechnol.* **13**, 418–423.

Snyder, R.O., Miao, C.H., Patijn, G.A., Spratt, S.K., Danos, O., Nagy, D., Gown, A.M., Winther, B., Meuse, L., Cohen, L.K., Thompson, A.R., Kay, M.A. (1997) Persistent and therapeutic concentrations of human factor IX in mice after hepatic gene transfer of recombinant AAV vectors, *Nat. Genet.* **16**, 270–276.

Sonntag, K.C., Emery, D.W., Yasumoto, A., Haller, G., Germana, S., Sablinski, T., Shimizu, A., Yamada, K., Shimada, H., Arn, S., Sachs, D.H., LeGuern, C. (2001) Tolerance to solid organ transplants through transfer of MHC class II genes, *J. Clin. Invest.* **107**, 65–71.

Spivack, J.G., Fraser, N.W. (1987) Detection of herpes simplex virus type 1 transcripts during latent infection in mice, *J. Virol.* **61**, 3841–3847.

Starcich, B., Ratner, L., Josephs, S.F., Okamoto, T., Gallo, R.C., Wong-Staal, F. (1985) Characterization of long terminal repeat sequences of HTLV-III, *Science* **227**, 538–540.

Stecenko, A.A., Brigham, K.L. (2003) Gene therapy progress and prospects: alpha-1 antitrypsin, *Gene Ther.* **10**, 95–99.

Stevens, J.G., Wagner, E.K., Devi Rao, G.B., Cook, M.L., Feldman, L.T. (1987) RNA complementary to a herpesvirus alpha gene mRNA is prominent in latently infected neurons, *Science* **235**, 1056–1059.

Su, H., Arakawa-Hoyt, J., Kan, Y.W. (2002) Adeno-associated viral vector-mediated hypoxia response element-regulated gene expression in mouse ischemic heart model, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 9480–9485.

Su, H., Lu, R., Kan, Y.W. (2000) Adeno-associated viral vector-mediated vascular endothelial growth factor gene transfer induces neovascular formation in ischemic heart, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 13801–13806.

Suzuki, M., Singh, R.N., Crystal, R.G. (1996) Regulatable promoters for use in gene therapy applications: modification of the 5′-flanking region of the CFTR gene with multiple cAMP response elements to support basal, low-level gene expression that can be upregulated by exogenous agents that raise intracellular levels of cAMP, *Hum. Gene Ther.* **7**, 1883–1893.

Suzuki, Y., Isogai, K., Teramoto, T., Tashita, H., Shimozawa, N., Nishimura, M., Asano, T., Oda, M., Kamei, A., Ishiguro, H., Kato, S., Ohashi, T., Kobayashi, H., Eto, Y., Kondo, N. (2000) Bone marrow transplantation for the treatment of X-linked adrenoleukodystrophy, *J. Inherit. Metab. Dis.* **23**, 453–458.

Sykes, M. (2001) Mixed chimerism and transplant tolerance, *Immunity* **14**, 417–424.

Takanashi, J., Kurihara, A., Tomita, M., Kanazawa, M., Yamamoto, S., Morita, F., Ikehira, H., Tanada, S., Kohno, Y. (2002) Distinctly abnormal brain metabolism in late-onset ornithine transcarbamylase deficiency, *Neurology* **59**, 210–214.

Takayama, T., Kaneko, K., Morelli, A.E., Li, W., Tahara, H., Thomson, A.W. (2002) Retroviral delivery of transforming growth factor-beta1 to myeloid dendritic cells: inhibition of T-cell priming ability and influence on allograft survival, *Transplantation* **74**, 112–119.

Tao, N., Gao, G.P., Parr, M., Johnston, J., Baradet, T., Wilson, J.M., Barsoum, J., Fawell, S.E. (2001) Sequestration of adenoviral vector by Kupffer cells leads to a nonlinear dose response of transduction in liver, *Mol. Ther.* **3**, 28–35.

Tratschin, J.D., West, M.H., Sandbank, T., Carter, B.J. (1984) A human parvovirus, adeno-associated virus, as a eucaryotic vector: transient expression and encapsidation of the procaryotic gene for chloramphenicol acetyltransferase, *Mol. Cell. Biol.* **4**, 2072–2081.

Tuchman, M., McCullough, B.A., Yudkoff, M. (2000) The molecular basis of ornithine transcarbamylase deficiency, *Eur. J. Pediatr.* **159**(Suppl 3), S196–S198.

Vale, P.R., Losordo, D.W., Milliken, C.E., McDonald, M.C., Gravelin, L.M., Curry, C.M., Esakof, D.D., Maysky, M., Symes, J.F., Isner, J.M. (2001) Randomized, single-blind, placebo-controlled pilot study of catheter-based myocardial gene transfer for therapeutic angiogenesis using left ventricular electromechanical mapping in patients with chronic myocardial ischemia, *Circulation* **103**, 2138–2143.

Wade-Martins, R., Smith, E.R., Tyminski, E., Chiocca, E.A., Saeki, Y. (2001) An infectious

transfer and expression system for genomic DNA loci in human and mouse cells, *Nat. Biotechnol.* **19**, 1067–1070.

Walsh, C.E. (2003) Gene therapy progress and prospects: gene therapy for the hemophilias, *Gene Ther.* **10**, 999–1003.

Wang, L., Muramatsu, S., Lu, Y., Ikeguchi, K., Fujimoto, K., Okada, T., Mizukami, H., Hanazono, Y., Kume, A., Urano, F., Ichinose, H., Nagatsu, T., Nakano, I., Ozawa, K. (2002) Delayed delivery of AAV-GDNF prevents nigral neurodegeneration and promotes functional recovery in a rat model of Parkinson's disease, *Gene Ther.* **9**, 381–389.

Wesseling, J.G., Bosma, P.J., Krasnykh, V., Kashentseva, E.A., Blackwell, J.L., Reynolds, P.N., Li, H., Parameshwar, M., Vickers, S.M., Jaffee, E.M., Huibregtse, K., Curiel, D.T., Dmitriev, I. (2001) Improved gene transfer efficiency to primary and established human pancreatic carcinoma target cells via epidermal growth factor receptor and integrin-targeted adenoviral vectors, *Gene Ther.* **8**, 969–976.

Woolf, C.J., Salter, M.W. (2000) Neuronal plasticity: increasing the gain in pain, *Science* **288**, 1765–1769.

Wu, H., Seki, T., Dmitriev, I., Uil, T., Kashentseva, E., Han, T., Curiel, D.T. (2002) Double modification of adenovirus fiber with RGD and polylysine motifs improves coxsackievirus-adenovirus receptor-independent gene transfer efficiency, *Hum. Gene Ther.* **13**, 1647–1653.

Xiao, X., Li, J., Samulski, R.J. (1998) Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus, *J. Virol.* **72**, 2224–2232.

Xie, Q., Bu, W., Bhatia, S., Hare, J., Somasundaram, T., Azzi, A., Chapman, M.S. (2002) The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10405–10410.

Yaksh, T.L. (1999) Spinal systems and pain processing: development of novel analgesic drugs with mechanistically defined models, *Trends Pharmacol. Sci.* **20**, 329–337.

Ye, X., Rivera, V.M., Zoltick, P., Cerasoli, F., Jr., Schnell, M.A., Gao, G., Hughes, J.V., Gilman, M., Wilson, J.M. (1999) Regulated delivery of therapeutic proteins after in vivo somatic cell gene transfer, *Science* **283**, 88–91.

Zhang, X.Y., La Russa, V.F., Bao, L., Kolls, J., Schwarzenberger, P., Reiser, J. (2002) Lentiviral vectors for sustained transgene expression in human bone marrow-derived stromal cells, *Mol. Ther.* **5**, 555–565.

Zielske, S.P., Gerson, S.L. (2002) Lentiviral Transduction of P140K MGMT into Human CD34(+) Hematopoietic Progenitors at Low Multiplicity of Infection Confers Significant Resistance to BG/BCNU and Allows Selection in Vitro, *Mol. Ther.* **5**, 381–387.

Zou, L., Zhou, H., Pastore, L., Yang, K. (2000) Prolonged transgene expression mediated by a helper-dependent adenoviral vector (hdAd) in the central nervous system, *Mol. Ther.* **2**, 105–113.

Zufferey, R., Nagy, D., Mandel, R.J., Naldini, L., Trono, D. (1997) Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo, *Nat. Biotechnol.* **15**, 871–875.

Zwaagstra, J., Ghiasi, H., Nesburn, A.B., Wechsler, S.L. (1989) In vitro promoter activity associated with the latency-associated transcript gene of herpes simplex virus type 1, *J. Gen. Virol.* **70**, 2163–2169.

Zwaagstra, J.C., Ghiasi, H., Nesburn, A.B., Wechsler, S.L. (1991) Identification of a major regulatory sequence in the latency associated transcript (LAT) promoter of herpes simplex virus type 1 (HSV-1), *Virology* **182**, 287–297.

Zwaagstra, J.C., Ghiasi, H., Slanina, S.M., Nesburn, A.B., Wheatley, S.C., Lillycrop, K., Wood, J., Latchman, D.S., Patel, K., Wechsler, S.L. (1990) Activity of herpes simplex virus type 1 latency-associated transcript (LAT) promoter in neuron-derived cells: evidence for neuron specificity and for a large LAT transcript, *J. Virol.* **64**, 5019–5028.

Zwiebel, J.A., Freeman, S.M., Kantoff, P.W., Cornetta, K., Ryan, U.S., Anderson, W.F. (1989) High-level recombinant gene expression in rabbit endothelial cells transduced by retroviral vectors, *Science* **243**, 220–222.

Vertebrate Mitochondrial DNA

Max Ingman and Ulf Gyllensten
Uppsala University, Uppsala, Sweden

1 Structure and Function 368
1.1 Structure of the Mitochondrial Genome 368
1.2 Gene Content and Function 369
1.3 Gene Organization 370
1.4 Genetic Code 371
1.5 Transcription and Replication 372

2 Origin of Mitochondria 373

3 Genetics of Mitochondrial DNA 374
3.1 Inheritance 374
3.2 Heteroplasmy and Segregation 374
3.3 Recombination 375
3.4 Substitution Rate 376

4 Mitochondrial Dysfunction and Human Disease 376
4.1 Point Mutations 377
4.2 Deletion-insertion Mutations 377
4.3 Mutations and Neurodegenerative Diseases 378
4.4 Contribution of Mitochondrial Mutations to Aging 378

5 Tracing Human History Using Mitochondrial DNA 379
5.1 Pros and Cons of Using Mitochondrial DNA for Evolutionary Studies 379
5.2 Mitochondrial Genomics 381
5.3 Human Migration Patterns 382
5.4 Natural Selection in Mitochondrial DNA 383

Bibliography 383
Books and Reviews 383
Primary Literature 384

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

Keywords

Chloroplast
A cytoplasmic organelle that is central to the photosynthetic process in plants.

Heteroplasmy
The occurrence of multiple forms of mitochondrial DNA (mtDNA) in the same individual.

Mitochondria
Cellular organelles containing the protein machinery for oxidative phosphorylation.

Oxidative Phosphorylation
A cell metabolism process by which cellular energy (ATP) is produced from ADP.

The mitochondria of eukaryotes contain a genome of their own, the mitochondrial DNA (mtDNA). The mitochondrial genome of metazoans is a covalently closed circular molecule with a size of about 16 kilonucleotide pairs (kntp). Considerable length differences exist between organism groups (e.g. ciliata >40 kntp, fungi 17–176 kntp, higher plants 200–800 kntp). In vertebrates, the genome encodes 2 rRNAs, 22 tRNAs, and 13 proteins involved in oxidative phosphorylation. This is a multicopy genome with a predominantly maternal inheritance, and in vertebrates, a high substitution rate. A number of mtDNA mutations in humans have been associated with muscular or neurological disease, and the accumulation of mutations in this genome may result in premature aging and contribute to the natural aging process. MtDNA has been one of the main tools used in studies of human evolution and has provided some of the strongest evidence for an African origin of modern humans. This review is biased toward the vertebrate genome due to the wealth of information available from this group. A number of unique characteristics set vertebrate mitochondrial DNA apart from the mtDNA in other organisms, justifying this focus.

1 Structure and Function

1.1 Structure of the Mitochondrial Genome

The mitochondria of multicellular animals (Metazoa) as well as of other eukaryotes contain a genome of their own: mitochondrial DNA. Most mtDNAs are organized as covalently closed circular molecules (or appears as such when a gene map is constructed). Linear molecules have been observed in the coelenterates (hydra), fungi (hansenula, candida), ciliates

(paramecium, tetrahymena), chlorophyte algae (Chlamydomonas), and chrysophyte algae (Ochromonas). Metazoan mtDNA genome range from about 14 kilonucleotide pairs (kntp) in the nematode *Caenorhabditis elegans* to about 42 kntp in scallop, *Placeopecten magellanicus*. All metazoan mtDNA genomes are organized as a single molecule, except that of the cnidarian *Hydra attenuata*, which appears to consist of two 8-kntp circles. By contrast, the mtDNA of other organismal groups are highly variable in size. For instance, the genome of *Paramecium aurelia* exceeds 40 kntps; fungi are sized between 17–176 kntps and higher plants have a genome size of 200 and 800 kntp. Particularly in plants, the genome is divided into a number of circles, with the subgenomic circles ranging from 60 to 240 kntp. In certain protozoans (order *Kinetoplastida*), the mtDNA is found in certain compartments of the mitochondria (kinetoplasts) and consists of maxicircles (14–39 kntp) and minicircles (0.6–2.5 kntp).

1.2 Gene Content and Function

In contrast to the enormous size variation among the mtDNAs of different species, the gene content is to a large extent conserved. Metazoan mtDNA contains the genes for two rRNAs homologous to the 16 S and 23 S ribosomal RNAs of *Escherichia coli*, 22 tRNAs, and 13 open reading frames (ORFs). Six of these ORFs encode enzymes, or components thereof, that are involved in oxidative phosphorylation (cytochrome *b*, CO subunits I–III, ATPase6 and ATPase8). The other seven ORFs encode subunits of the respiratory chain NADH dehydrogenases (commonly referred to as ND1 and ND4L). In addition, a single noncoding region termed the *displacement loop* or D loop is found between the coding sequences (See Fig. 1). The term "D loop" refers to the short three-stranded structure formed by displacement of the heavy (H) strand by a short, newly synthesized, H-strand segment. Although noncoding, the D loop contains a number of regulatory signal sequences necessary for the initiation of H-strand replication and for both H-strand and L-strand (light strand) transcription. Extensive length variation has been observed between the D loops of different metazoans and appears to be due to different types of directly repeated segments. While the D-loop region spans about 1 to 2 kntp in vertebrates, it is only about 100 ntp in some species of sea urchins. The intraspecific variability in the length of the D loop is also extensive in some species and is similarly caused by variation in the number of directly repeated segments.

In addition to the homologs for the protein-coding genes in the mtDNA of all metazoan species, plants carry the genes for several other subunits of the respiratory chain (e.g. ATP synthetase F1 complex, and ATP synthetase subunit 9), as well as genes not previously found in mitochondrial genomes, such as the gene for reverse transcriptase. Also, all plant mtDNA carry the gene for 5 S rRNA, not present in animals and only occasionally in fungal mtDNA.

In protozoans, the genome (maxicircle) contains the same protein-coding genes as in metazoan mtDNA and in addition, components of complex I (ND7 and ND8) that are found in the nuclear DNA of metazoan species or the chloroplast DNA of plants. Other components that have been found in the protozoan mtDNA genome are genes for apocytochrome *b* and CO subunits I, II, and III. In some fungi, a number of additional ORFs have

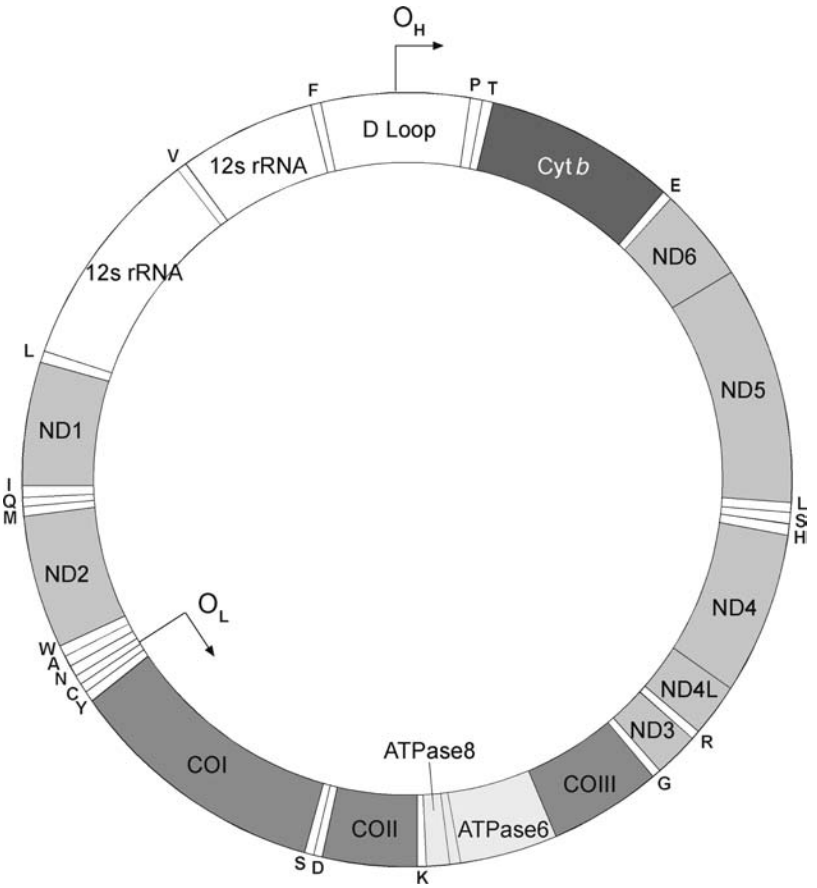

Fig. 1 Outline of the mammalian mitochondrial genome. The location of genes encoding the 2 rRNAs and the 13 proteins are indicated. The locations of the tRNA genes are indicated as white bars in between the protein-coding genes and labeled using standard nomenclature. The major noncoding region (D loop) is located between the tRNA phenylalanine and tRNA proline genes. The origins and direction of replication for the heavy (O_H) and light (O_L) strands are also shown.

been identified, but except for a few (e.g. the variant 1 protein (VAR1) thought to encode a mitochondrial ribosomal protein), no function has been assigned to the products of these ORFs. The genes involved in the translation process also vary in distribution among taxa. While the tRNA genes are present in the mtDNA of most taxa, the number of rRNA genes located in the mtDNA varies between 28 for aspergillus to zero for trypanosomes. The genes for the ribosomal proteins are located in the mtDNA of bryophytes and to some extent in angiosperms and some protists, but are lacking entirely from the mtDNA of other organisms.

1.3 Gene Organization

Most vertebrates and invertebrates have a conserved gene content, but the gene order

may vary considerably between phyla. For instance, among five vertebrate species: human, mouse, bovine, rat, and Xenopus (phylum Chordata), the gene order is identical. The gene order in species from the phyla Echinodermata and Anthropoda differ from the vertebrate gene order by a minimum of two translocations and three inversions. By contrast, species from the phylum Nematoda have such deviating gene orders that no simple scheme has been constructed to account for the differences relative to the gene order of vertebrates.

The organization of animal mtDNA is extremely compact and in contrast to that of nuclear genes, the genome lacks introns (excluding parts of the D loop for which no functional role has been defined). As the only exception to date, the sea anemone (Metridium) apparently contains intron sequences. Also, the coding sequences of many species display variable degrees of overlap, either between the 3′-ends of two genes that are encoded from opposite strands of the molecule or between genes encoded by the same strand. For example, ATPase8 overlaps with ATPase6 by 2 to 46 ntps in vertebrates and in higher invertebrates and ND4L overlaps with ND4 by 7 ntp in vertebrates.

The genome size of fungal mtDNA spans a wide size range (17.6–176 kntp) while being relatively conserved with respect to gene content. This size variation is due to expansion of AT-rich intergenic spacer regions and the existence of large introns. For instance, in species of Podosora, the *CO1* gene covers a total of 24.5 kntp (26% of the genome) and includes 16 introns. The fungi mtDNA genome also appears to encode two proteins associated with introns: intron-encoded maturases, which function in intron excision during posttranscriptional processing of precursor RNAs, and intron-encoded site-specific endonucleases, necessary for intron transposition. In the light of the enormous diversity in genome sizes among fungi, it is not surprising to find that gene order or gene organization differ considerably among species. It is believed that the fungal mtDNA genomes, as a result of the high frequency of intergenic noncoding iterated spacer sequences, can change in organization to exhibit a higher frequency of recombination.

Our understanding of genome organization in higher plants may be biased, since information is to a large extent based on flowering plants. As is the case in fungi, the large and diverse genome size (200–2400 kntp) in flowering plants is due to the occurrence of intron sequences. To an even larger extent than in fungi, plant mtDNA appear to have a propensity for rapid structural change coupled with a low rate of sequence divergence. The plasticity of the genome has been related to the existence of repeated blocks, several kntps long, that occur in both orientations and function as recombination hotspots. Recombination between such regions will generate "minicircles," which can join other "minicircles" carrying the same repeat motif. Possibly as a consequence of this organization, the gene order is not strongly conserved among plant species, and genetic exchange between the chloroplast and mitochondrial genomes appears to occur frequently. In contrast to fungi, most of the noncoding mtDNA in plants show no distinctive nucleotide bias (e.g. no AT-rich spacer as in fungi).

1.4 Genetic Code

On the basis of comparisons between the predicted amino acid sequences (using

nucleotide sequence data) and protein sequencing, mitochondrial DNA has a unique genetic code. For example, rather than functioning as a stop codon as it does in nuclear DNA, the triplet TGA specifies tryptophan in all mtDNAs except those of plants. This alteration is likely to have occurred before the divergence of Fungi, Protista, and Metazoa. With the exception of cnidarians, the AGA and AGG triplets of mtDNA specify serine instead of arginine as in the nuclear genome. This alteration appears to have occurred after the cnidarian ancestral lineage diverged from that leading to all other invertebrates. In echinoderms, the AUA codon specifies isoleucine and AAA specifies asparagine instead of lysine. The AAA codon is also thought to specify asparagine in platyhelminths, indicating that the specificity of AAA has changed from lysine to asparagine on two independent evolutionary lineages.

The metazoan transcription machinery also employs unorthodox translation initiation codons. All four ATN triplets are being used as initiation codons in mtDNA in different species (although the subset of codons used varies considerably between species). Finally, the translation terminations of some mammalian mitochondrial proteins end either in T alone or TA, instead of the complete termination codon TAA or TAG. The primary transcript of the mtDNA is a multicistronic RNA, and precise RNA cleavage and posttranscriptional polyadenylation will generate the complete termination codons.

1.5 Transcription and Replication

Transcription of vertebrate mtDNA is initiated at two promoters, heavy-strand promoter (HSP) and light-strand promoter (LSP) located in the major noncoding region of the mtDNA. These promoters mainly direct the synthesis of transcripts from one strand, although they have been shown to be bidirectional under certain circumstances. The evolution of separate promoters for the different strands may be a characteristic that has evolved in the mammalian lineage, since in amphibians (Xenopus) the promoters have a much more pronounced bidirectionality and in birds (chicken) the single available promoter is completely bidirectional. At least two proteins are necessary for transcription initiation: mitochondrial transcription factor 1 (mtTF1) and mtDNA polymerase. The mtTF1 binds −10 to −40 ntps upstream from the promoter and initiates transcription together with the mtRNA polymerase. The function of the mtTF1 is highly conserved among mammals since the mouse and the human mtTF1 can be experimentally exchanged with each other. MtTF1 shows sequence similarity with nuclear high mobility group (HMG) proteins and a nuclear rDNA transcription factor (pol1). There is no known difference between the HSP and LSP promoters with regard to the proteins required for proper initiation of transcription, supporting the hypothesis that they originated by duplication. Termination of transcription of mammalian mtDNA appears to occur at a single site located at the 3′ terminus of the rRNA region.

Mammalian mtDNA has two origins of replication – one for each strand – and a similar organization appears to exist in all vertebrate mtDNA. The origin for heavy-strand (O_H) synthesis is located within the D loop, while that for the light strand (O_L) is surrounded by a cluster of tRNA genes (Fig. 1). Replication always begins at the O_H and synthesis of the light strand is not initiated until

the O_L has been displaced as a single-stranded molecule. The replication event begins with the synthesis of an RNA strand and later shifts into the synthesis of a DNA strand. In the case of the mouse, RNA synthesis appears to begin at the transcription promoter, and therefore replication priming at the O_H depends on transcription promoter function. One transcription initiation event could thus serve for both transcription and replication priming. The transition from RNA to DNA is over a set of conserved sequence blocks (CSB) in the D loop and mediated by an RNAase MRP (mitochondrial RNA processing) enzyme. Priming from the O_L is through a DNA primase that recognizes the target site (O_L). With regard to O_H priming, a stretch of about 15 ribonucleotides is synthesized before the transition to DNA is made. In this case, the transition is located in a tRNA gene: this is the second occasion during which a tRNA gene sequence forms part of the regulatory sequences in mtDNA. After priming and transition to DNA, the mitochondrial γ-polymerase extends each of the strands. This enzyme has the 3′ to 5′ exonuclease activity associated with proofreading activity.

2 Origin of Mitochondria

The difference in the genetic code, circular genome organization and the localization of the mtDNA relative to the nuclear genome suggest that mtDNA has a separate evolutionary origin from the nuclear genome. Two alternative hypotheses have been presented: (1) the mitochondrion and its genome arose through a process of intracellular compartmentalization and functional speciation, occurring within a single kind of cell (autogenous origin), or (2) the mitochondria are evolutionary remnants of endosymbionts, probably of bacterial origin (xenogenous origin). Phylogenetic analyses of mitochondrial sequence data have been performed to address these hypotheses. The most informative analyses have been based on comparisons of ribosomal RNA sequences because of the presence of these genes in all the genome types compared, their functional constancy and equivalence, and the highly conserved primary and secondary structure of the rRNA allowing for precise and relatively unambiguous alignment of sequences. Analyses of rRNA sequences have suggested a new primary subdivision of living organisms. Instead of the five kingdoms (Animalia, Plantae, Fungi, Protista, and Monera) previously recognized, it has been suggested that organisms should be divided into three primary lineages: the archaebacteria, the eubacteria, and the eukaryotes. Archaebacteria contain two major divisions of organisms: (1) sulfur-dependent, extremely thermophilic bacteria, and (2) methane producers and extreme halophiles. Phylogenetic analyses as well as a wealth of other information clearly indicate that archaebacteria have a different origin than eubacteria (containing plastid genomes such as mitochondria and bacteria), although there is still debate over whether they are monophyletic (having a single origin) or paraphyletic (having multiple genetic origins).

In the phylogenetic tree, the mitochondrial tRNA sequences are found in the eubacterial group, showing that they must have a different evolutionary origin than the eukaryotic cells in which they are presently found. In particular, the closest contemporary relatives of mitochondrial rRNA are members of a subgroup of

purple bacteria, providing unambiguous evidence in favor of the endosymbiotic origin of mitochondria of animals, fungi, plants, and ciliates from within aerobic eubacteria. Extant relatives of the aerobic eubacteria family that form intracellular associations with eukaryotes include rhizobacteria, agrobacteria, and rickettsiae.

Although the general origin of mitochondria and mtDNA has been elucidated, a number of issues remain to be determined. Was there a single event of symbiosis or multiple events? Are mitochondrial and plastids monophyletic or polyphyletic? Did mitochondria and plastids (chloroplasts) arise simultaneously or serially? How has lateral gene flow among chloroplasts, plastids, and the nuclear genome affected the present gene content of the different genomes?

It is well known that genes can be transferred between the mitochondrion and the chloroplast, as well as between the chloroplast and the nucleus. Among animals, a number of copies of mtDNA fragments have been inserted into the nuclear genome, but no examples of expressed sequences have been reported. In the mungbean and the cowpea, a copy of the mitochondrial *COII* gene has been found in the nucleus and is transcribed. This transfer is estimated to have occurred 60–200 million years ago, while the loss of mitochondrial COII expression appears to be much more recent. The probable pathway of gene transfer in this case involved reverse transcription of an RNA intermediate. The frequency of transfers that are followed by successful expression appears to be relatively low in higher organisms. This may be partly because, to accomplish expression of the gene, the transcription control element must be present and the gene must contain the sequence of the signal peptide necessary for transport into the mitochondria.

3 Genetics of Mitochondrial DNA

3.1 Inheritance

The mitochondrial DNA of most metazoan species is predominately maternally inherited. In interspecific crosses of mice, the mtDNA of 5–10 mitochondria in the midpiece of the sperm is able to enter the egg and become established in the zygote. The contribution in mice is on the order of 10 paternal mitochondria in 10 000 maternal mitochondria per generation. In humans, one case has been observed where paternal contribution has been documented. Given the very large number of transmissions that have been examined in different human studies, all consistent with an exclusively maternal mode of inheritance, the circumstances under which paternal contribution can occur in humans must be very unique. When paternal contribution occurs, it may generate individuals with multiple types of mtDNA. Low levels of paternal contribution of mtDNA have also been reported in Drosophila. In bivalves (mussels), the contribution appears to be almost equal among sexes, and in forest trees, the mitochondria are inherited through the paternal parent. Hence, there is wide variation between different organisms with respect to the pattern of inheritance of the organelle.

3.2 Heteroplasmy and Segregation

In metazoan species, single individuals frequently contain a population of

molecules differing either by point mutations or in length: a state known as *heteroplasmy*. A heteroplasmic state could arise within individuals through mutational mechanisms, or the mutant mtDNA might be acquired through paternal contribution. The high frequency of some of the variant molecules (50–90%) implies that forces other than the mutation rate are involved in determining the frequency of variant molecules. Mutated mtDNA may increase in proportion as a consequence of selection or strong segregation (genetic drift) during oogenesis or embryogenesis. Different tissues may show different frequencies of variant molecules due to the independence of cellular division and mitochondrial replication.

A number of reports have described the stability of the heteroplasmic state between generations in metazoans, suggesting that segregation during oogenesis is limited. In contrast to these studies, extremely rapid and dramatic shifts in the frequency of different mtDNA types have been reported over a single generation in some mammalian species (Holstein cows). Also, in studies of mtDNA mutations associated with human disease, dramatic changes in the frequency of different mutations have sometimes been seen from one generation to the next. However, the rules governing the transmission and segregation of mitochondria between sexual generations are at present only partly known. Analyses of single cells have shown different levels of heteroplasmy. The mitochondrial vesicles within each of these cells continuously fuse and divide and the mtDNA content of individual vesicles may vary from zero to approximately 15 copies. However, it is not known whether mtDNA distribute among vesicles during the process of fusion.

3.3 Recombination

Recombination in mitochondrial DNA is quite common in plants, where it gives rise to the large number of different sized molecules present. Particular recombination sites (repeats) in plant mtDNA appear to function as hot spots for sequence exchange. Also, in fungi, mtDNA appears to participate in recombination. In mammals, mitochondria are usually inherited from the mother through the egg cytoplasm. In humans, the egg cytoplasm contains a large number of mitochondrial particles whereas the sperm midpiece only carries five to ten. During fertilization, these paternal mitochondria are labeled for degradation through the ubiquitin pathway, resulting in an exclusively maternal inheritance pattern. This uniparental mode of inheritance means that recombination between maternal and paternal mitochondria should not occur.

As early as 1980, Giles and colleagues demonstrated maternal inheritance in human mitochondrial DNA. This was confirmed in intraspecific mouse crosses, while in interspecific crosses, paternal contribution was observed. On the basis of phylogenetic and linkage disequilibrium (LD) analyses, evidence of recombination has been claimed between paternal and maternal human mitochondrial DNA. The results of these studies have been controversial, with criticisms focusing mainly on the statistical method used and a biased dataset. Reanalyses of the dataset using the standard LD measure *D*' show that there is no association of linkage disequilibrium with physical distance.

Paternal mtDNA has been detected in the muscle tissue of a single patient suffering from myopathy while the blood, hair

roots, and cultured fibroblasts of this individual contained only maternal mtDNA. This individual was later shown to harbor recombinant mtDNA, demonstrating that it is possible for paternal mitochondria from the sperm to escape the degradative ubiquitin pathway and potentially recombine with maternal mtDNA. While this may happen in isolated cases, the absence of evidence of recombination in large datasets of human mtDNA sequences suggests that it is rare and does not contribute significantly to the evolution of human mitochondrial DNA.

3.4 Substitution Rate

The substitution rate in the human mitochondrial genome is typically 5 to 10 times higher than that of nuclear DNA, a figure that is consistent across vertebrates. The reason for this higher rate is thought to be due to either a high rate of nucleotide misincorporation or low DNA polymerase repair efficiency. A high concentration of oxidative radicals (superoxide anions H_2O_2) by-products of oxidative phosphorylation have been implicated as a contributing factor to this high substitution rate. These compounds react with lipids, proteins, and nucleic acids and form thymine glycols and 8-hydroxyguanosine in DNA, which may inhibit replication and transcription, leading to longer exposure of triple-stranded DNA to mutagenic compounds. In addition, studies on the reliability of mitochondrial DNA replication have shown that the net average fidelity of the human mitochondrial DNA polymerase (pol γ) is about 1 error per 1–20 million base pairs, a rate that is 1000-fold higher than the error frequency expected for the replication of nuclear DNA. Also, the high turnover rate of mitochondria in tissues would provide more rounds of replication during which errors could be generated. Compounding these factors is an enhanced chance of fixation of new mutations due to low functional constraints on mitochondrial products.

However, an elevated rate of substitution has not been detected for invertebrate, or plant mtDNA, or for chloroplast DNA, where the rate appears to be similar or lower than the substitution rate of nuclear genes. For instance, in plants, the mtDNA substitution rate has been found to be about sixfold lower than that of nuclear genes and threefold lower than that of chloroplast DNA.

4 Mitochondrial Dysfunction and Human Disease

The mitochondrial genome supplies parts of the protein machinery that is necessary for efficient oxidative phosphorylation, using a series of five multiple-subunit enzymes located within the mitochondrial inner membrane. The components needed for oxidative phosphorylation are encoded by both nuclear and mitochondrial genes. The nuclear genome provides a total of 65 subunits, compared to the 13 mitochondrially encoded subunits. A genetic defect in oxidative phosphorylation could be due to mutations in genes of either the nuclear or the mitochondrial genome. On the basis of the number of different genes involved, it may be assumed that nuclear mutations causing mitochondrial malfunction should predominate by a factor of 5. However, differences in the frequency by which new mutations are introduced in the two genomes (see Sect. 3.4), as well

as the clonal inheritance pattern of mitochondrial DNA, may compensate for this difference and indeed result in a higher proportion of mitochondrial dysfunction being due to mtDNA mutations.

4.1 Point Mutations

Point mutations resulting in amino acid replacements have been found to be associated with Leber's hereditary optic neuropathy (LHON) and neurogenetic muscle weakness, ataxia, and retinitis pigmentosa (NARP). A number of different point mutations have been associated with these symptoms, some (like the one at ntp1 1778) account for more than 50% of the LHON cases, while a number of others are private mutations that either alone, or through synergistic effects, cause a similar phenotype. The segregation of the main LHON mutation may occur very rapidly: the proportion of mutated genomes has been found to rise from less than 50% to almost fixation in a single sexual generation.

NARP patients have a very diverse set of symptoms ranging from ataxia and dementia to proximal neurogenic muscle weakness. They have in common a missense mutation at ntp 8993, resulting in a change of amino acid 156 of the ATPase6 from a leucine to an arginine.

Since the mitochondria have their own set of tRNAs, mutation in any of these may cause systemic effects by partial or complete inhibition of mitochondrial mRNAs. A number of different point mutations have been found to be associated with diseases such as myoclonic epilepsy with red-ragged fibers (MERRF) and mitochondrial encephalomyopathy, lactic acidosis, and stroke-like symptoms (MELAS). Patients suffering from MERRF (point mutation at ntp 8344) have uncontrolled myoclonic epilepsy and mitochondrial myopathy that may translate into symptoms such as hearing loss, dementia, myoclonus, and respiratory failure. The severity of the symptoms appears to be related to the proportion of mutated genomes, and the disease has a progressive development. Therefore, an individual with a moderate proportion of mutated mtDNA (20%) and no symptoms at age 20 may have developed severe complications by the age of 65. This progression is believed to be due to the age-related decline of oxidative phosphorylation. Similarly, most MELAS patients have point mutations in the tRNA for leucine (ntp 3243), which also result in an inactivation of an overlapping transcription termination sequence.

4.2 Deletion-insertion Mutations

Insertions/deletions of the mtDNA have been found in patients with Kearns–Sayre syndrome (KSS), chronic external ophthalmoplegia (CEOP) and Pearson's syndrome. The majority of cases of KSS and CEOP involve the deletion of the region in between, but not including, the origins of replication (O_H and O_L). Conceivably, deletions that encompass the two origins of replication, inhibit replication and such mutations are lost during mitochondrial regeneration. More than 95% of cases carry deletions between O_H and O_L: the most frequent deletion (between ntp 8468 and 13 446) occurs in 30 to 50% of patients. In most of the CEOP, KSS, and Pearson's syndrome cases, only a single type of deletion is found, suggesting that these mutations have been generated by a single event, either through slip-replication, homologous recombination, or topoisomerase cleavage. Since most cases are single spontaneous mutations with no family history, they are

likely to have been generated *de novo* in a specific cell lineage during embryogenesis. This may also explain the large differences in the proportion of genomes carrying deletions between tissue types. In many KSS and CEOP patients, there is a progression of the severity of the disease with age, concomitant with an increase in the proportion of deleted mtDNA in the muscle fibers. The deleted genomes in these diseases may be postulated to accumulate over time simply as a consequence of replicative advantage (i.e. their shorter length). In the case of point mutations, such as those described in Sect. 4.1, additional factors (e.g. selection) must be invoked to explain the rapid shift observed between generations or tissue types.

Duplications extending from the *COII* gene to the cytochrome *b* gene, and spanning both origins of replication, have been found in the mtDNA of patients with Pearson's syndrome. Because of the added number of replication origins, these mutations may have a replicative advantage that contributes to their rapid accumulation.

Cases of respiratory failure, lactic acidosis, and failure of muscle or kidneys have also been found to be associated with extremely low amounts of mtDNA. In a case of lethal mitochondrial myopathy, only 2 to 12% of the normal amount of mtDNA was detected. Similar symptoms have been observed in AIDS patients as a result of zidovudine treatment, which is intended to block viral DNA replication, but, as a side effect also has a similar effect on mtDNA.

4.3 Mutations and Neurodegenerative Diseases

The diseases discussed thus far are but a few of those in which mitochondrial mutations have been implicated. Some families with diabetes mellitus have been shown to carry a deficiency of oxidative phosphorylation coupled with an mtDNA deletion, which, in the heteroplasmic state, removed O_L and large parts of the mtDNA genome. Also, mtDNA mutations have been implicated in neurodegenerative diseases (e.g. Parkinson's, Alzheimer's, Huntington's). The brain tissue of Parkinson's disease patients has been shown to contain mtDNA deletions and a deficiency of complex I has been reported in both the brain and muscle tissue. Most Parkinson's cases are sporadic, and only occasionally will family members be affected, as would be expected if the decline in oxidative phosphorylation were genetically based. Similarly, some Alzheimer's patients have complex IV defects and most of the cases appear to be spontaneous (with a fraction being autosomal dominant). Attempts have been made to correlate the frequency of mtDNA deletions with Alzheimer's disease, but at present there is no strong evidence that these patients carry aberrant levels of mutations compared with age-matched asymptomatic controls. In addition, decreased level of CO activity was recently reported in certain brain regions of a patient with schizophrenia. This finding indicates that mitochondrial dysfunction may also be involved in the etiology of psychiatric disorders.

4.4 Contribution of Mitochondrial Mutations to Aging

A confounding factor in the analysis of mitochondrial diseases is the natural accumulation of mutations with age. A number of studies have shown that insertion–deletion mutations in rat skeletal

muscle and deletions in human hearts tend to accumulate with age. Similar studies indicate that certain point mutations may also accumulate with age. Taken together, these studies suggest that there may be a progressive accumulation of mtDNA mutations, which may in turn contribute to the decline in oxidative phosphorylation that is characteristic of aging individuals. Therefore, it has been postulated that biological aging is related to and partially caused by the accumulation of mitochondrial mutations. One concern is that the frequency of mutations in asymptomatic aging people appears to be too low to be able to exert a significant effect on oxidative phosphorylation. Also, experiments from cell hybrids indicate that the main age-related decline in oxidative phosphorylation is under the control of nuclear, rather than mitochondrial, genes. However, the age-dependent accumulation of mtDNA mutations is still important in the prognosis for carriers of mtDNA disease mutations. Recently, a mouse model in which the proofreading activity of the mitochondrial polymerase had been partly altered showed the accumulation of a high number of mutations and also symptoms of premature aging. This provides evidence that high levels of mtDNA mutations may result in symptoms similar to that observed during biological aging.

5 Tracing Human History Using Mitochondrial DNA

5.1 Pros and Cons of Using Mitochondrial DNA for Evolutionary Studies

MtDNA has been heavily utilized for studies of human evolution. This is largely due to a handful of appealing features that set mtDNA apart from nuclear DNA. Possibly the most important factor is that, like the Y chromosome, mitochondria are uniparentally inherited. While the lack of apparent recombination means that only the history of women in a population can be investigated, it allows the tracing of a direct genetic line where all polymorphism is due to mutation (however, see Sect. 3.3). The high substitution rate in the mitochondrial genome compounds this advantage and allows for phylogenetic resolution of more closely related individuals than is possible with nuclear DNA sequences of similar length. If mtDNA is evolving 5 to 10 times faster than nuclear DNA, then 5 to 10 times the amount of nuclear sequence data would be required to give the same amount of phylogenetic data. Also, as the mitochondrial genome is of an immediately finite length (~16 500 ntp), a wealth of data from the same locus is available for comparison. In fact, over 10 000 human mitochondrial hypervariable region sequences (from the mitochondrial D loop) have been produced to date, many for the purpose of forensic analyses. The D loop is an important source of information for forensic scientists for many of the same reasons that it appeals to population geneticists. It can be possible to extract mtDNA from sources when the nuclear DNA is degraded beyond use. This is mainly because many more copies of the mitochondrial genome exist in each cell. In fact, mtDNA has been extracted, amplified and sequenced from the fossilized remains of Neanderthal bones at least 30 000 years old.

However, estimation of the number of nucleotide substitutions between a pair of sequences is not a straightforward task. If the genetic distance between two mtDNA sequences is very high, there

is the likelihood that the same site may have undergone substitution more than once (multiple substitution). Another confounding factor coupled to a high genetic distance is parallel substitution, in which the same site undergoes the same nucleotide substitution in multiple lineages independently. In either case, two substitutions would have occurred, but neither will be observed in a pairwise comparison. Both these situations will lead to underestimation of the genetic distance between the sequences and will affect the calculation of the time at which they diverged. In a sample of 53 mitochondrial D-loop sequences, it was found that some sites had undergone parallel substitution in up to 10 separate lineages. This means that at some loci, a large percentage of sequences share polymorphisms by chance rather than by descent.

The problem of multiple substitutions is illustrated in Fig. 2. While the substitution rate has remained relatively constant, the accumulation of substitutions between these species is not linear over time. In fact, the rate of accumulation of pairwise synonymous substitutions ranges from 17 per million years in species that diverged 2 million years ago to 4.5 per million years in species that diverged 20 million years ago. This is due to a saturation of mutations, making multiple substitutions more common. However, in agreement with the molecular clock hypothesis, the rate of accumulation of nonsynonymous substitutions (and therefore amino acid replacements) remains fairly constant.

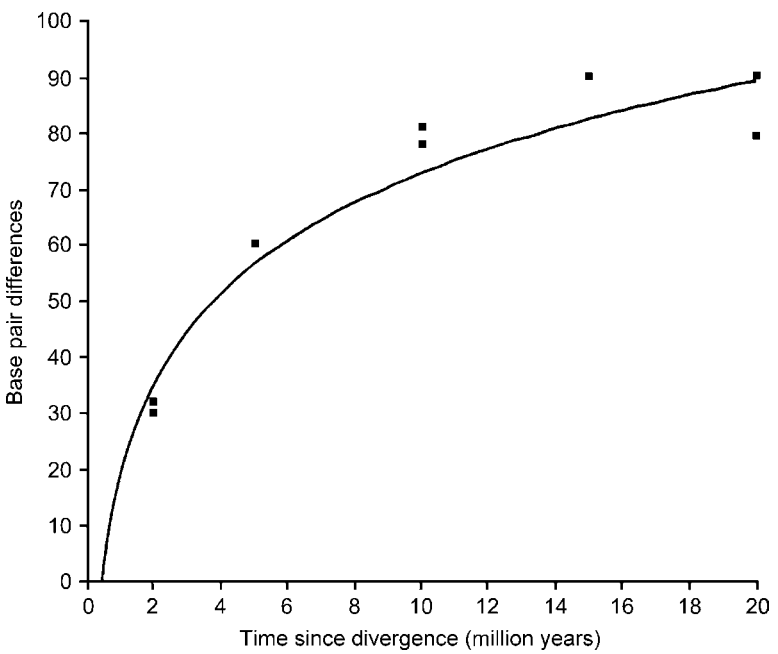

Fig. 2 Total nucleotide substitutions between pairs of mitochondrial cytochrome *c* oxidase subunit II gene (684 bp) from various bovid taxa *versus* time since divergence (from Janecek, L.L., Honeycutt, R.L., Adkins, R.M., Davis, S.K. (1996) Mitochondrial gene sequences and the molecular systematics of the Artiodactyl subfamily Bovinae, *Mol. Phylogenet. Evol.* **6**, 107–119).

Of course, this is an extreme example and, because there is considerably less time depth, the problem is not so pronounced for analysis of human mitochondrial DNAs. Nevertheless, due to these complications, it is necessary to use a mathematical model to estimate the number of nucleotide substitutions between pairs of sequences.

5.2 Mitochondrial Genomics

Prior to the year 2000, few studies presented phylogenetic analyses based on the complete mitochondrial genome, but rather focused on particular sites that are thought to be variable, or the rapidly evolving mitochondrial control region (D loop). One of the human population genetics studies that stands out above all others employed mtDNA genetic variation to study the evolutionary relationships between individuals of diverse origins. In 1987, a group of researchers at the University of California campus at Berkeley presented data that indicated a recent African origin of modern humans. This landmark study was based on restriction fragment length polymorphism (RFLP) variation among the mtDNA of 147 humans from five populations. In this study, 12 restriction enzymes were used to cut the mtDNA of each of these individuals at an average of 370 positions. Using this technique, Cann and colleagues assayed a total of about 9% of the mtDNA of each individual and identified 195 polymorphic sites. The 147 individuals included in the study yielded 133 distinct mitochondrial haplotypes, which were assembled into a phylogenetic tree using the maximum parsimony method. The oldest group (clade) on the tree was comprised of mtDNA types derived from African people and, given the number of differences between the two most divergent mtDNA types, the common ancestor of all the haplotypes was estimated to have existed (in Africa) between 140 000 and 280 000 years ago.

The mitochondrial D loop comprises about 1100 ntp and makes up less than 7% of the complete mitochondrial genome. Sequencing the D loop for use in molecular population genetics has been popular due to the high degree of polymorphism that can be found in this relatively short sequence. The advantage of copious amounts of published data from the human D loop continues to make it an attractive source of phylogenetic information. The main problem of the D loop stems from extensive heterogeneity in substitution rate that results from an extreme rate of substitution at some sites. The consequence of this is a high frequency of multiple substitutions and, as a result, specific algorithms for estimating genetic distance have appeared in an attempt to take this into account. Nevertheless, information gleaned from D-loop sequences has added significantly to our level of understanding of human evolution and the genetic relationships among populations.

Since 2000, when the first large-scale study utilized complete mitochondrial genomes, more than 2000 complete or near complete human mitochondrial genomes have been published. Complete mitochondrial genomes provide increased phylogenetic resolution of closely related lineages, as well as facilitating more accurate estimations of genetic dates for demographic events since all substitutions are available, including singleton changes (present in just one lineage). While RFLP data can define some major clades, the complete genome data provide more topological support and also distinguish a

number of additional clades. In addition, the availability of complete mitochondrial genomes has allowed a more detailed analysis of the pattern of selection across the mitochondrial coding region as well as comprehensive surveys for evidence of mitochondrial recombination.

5.3 Human Migration Patterns

Our understanding of the global pattern of the movements of modern human peoples is based on a synthesis of physical and cultural anthropology with molecular data from multiple genetic systems and loci (see Fig. 3). Also taken into account are geological factors such as the degree of glaciation, which can affect migration paths and raise or lower sea levels.

Analyses of mtDNA have provided important information about human evolution and migration for nearly 20 years. The earliest and some of the strongest evidence for the ''recent African origin'' hypothesis have been based on mtDNA. This hypothesis is now a general (but incomplete) consensus and states that anatomically modern humans evolved in Africa 100 000 to 200 000 years ago, and all extant humans are descended from a small number of females that lived at that time. MtDNA analyses in concert with studies of nuclear DNA and physical evidence have also shown that modern humans exited Africa in the last 100 000 years to the Levant, the Mediterranean region between Turkey and Egypt. From there, they radiated toward East Asia, Europe, Australia, and eventually to the Americas. As modern humans

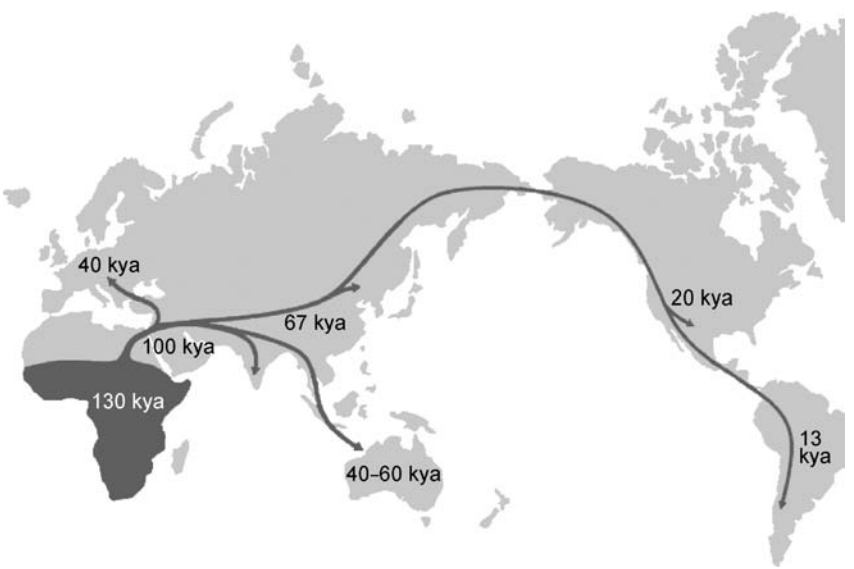

Fig. 3 The origin and dispersal of modern humans. Evidence from mitochondrial genomes bolsters the hypothesis that the place of origin was sub-Saharan Africa between 100 000 and 200 000 years ago and that the dispersal from Africa occurred within the past 100 000 years. The earliest known fossil and archaeological evidence on each continent, shown on the map, is consistent with this view (redrawn after Hedges, S.B. (2000) A start for population genomics, *Nature* **408**, 652–653).

advanced across the world, they must have encountered archaic humans that were resident in their new range. A key question is whether there was total replacement of these previous occupants or if admixture between modern and archaic forms has contributed to the contemporary human gene pool. While barriers to fertility are slow to develop, making admixture biologically feasible, the absence of evidence of this from clonal loci such as mtDNA and the Y chromosome suggests that archaic forms such as *Homo erectus* have not contributed significantly to current human genetic variation.

5.4 Natural Selection in Mitochondrial DNA

There are several examples of nonneutrality of mtDNA mutations, such as cytoplasmic male sterility in plants and in humans, and in a muscle disease known as *mitochondrial myopathy*. Effects of natural selection in mitochondrial coding regions are also evident in the discordance between the numbers of synonymous and nonsynonymous substitutions. Also, when mitochondrial lineages are subdivided into groupings that are thought to reflect their historical habitat, a larger proportion of nonsynonymous substitutions are seen among the "arctic" lineages relative to the "tropical" and "temperate" lineages in the *ATPase6* gene. This could be due to changing environmental pressures on mitochondrial lineages as humans moved from the African climate to exploit new habitats.

This may be especially important in cold environments as mitochondrial oxidative phosphorylation (OXPHOS) employs dietary calories to produce ATP, which can be utilized to generate heat for body temperature maintenance. Mitochondrial lineages may therefore vary functionally depending on their long-term environment and the rate of evolution of lineages from different habitats could be quite different.

See also Metabolic Basis of Cellular Energy; Mitochondrial Porins, Eukaryotic.

Bibliography

Books and Reviews

Chinnery, P.F., Turnbull, D.M. (2001) Epidemiology and treatment of mitochondrial disorders, *Am. J. Med. Genet.* **106**, 94–101.

Clayton, D.A. (1991) Replication and transcription of vertebrate mitochondrial DNA, *Annu. Rev. Cell. Biol.* **7**, 453–478.

DiMauro, S., Hirano, M., Schon, E.A. (2000) Mitochondrial encephalomyopathies: therapeutic approaches, *Neurol. Sci.* **21**, 901–908.

Hedges, S.B. (2000) A start for population genomics, *Nature* **408**, 652–653.

Larsson, N.G., Rustin, P. (2001) Animal models for respiratory chain disease, *Trends Mol. Med.* **12**, 578–581.

Lightowlers, R.N., Chinnery, P.F., Turnbull, D.M. (1997) Mammalian mitochondrial genetics: heredity, heteroplasmy and disease, *Trends Genet.* **13**, 450–455.

MITOMAP. (2005) A Human Mitochondrial Genome Database (http://www.mitomap.org).

MtDB. (2005) Human Mitochondrial Genome Database (http://www.genpat.uu.se/mtDB).

Oldfors, A., Tulinius, M. (2003) Mitochondrial encephalomyopathies, *J. Neuropathol. Exp. Neurol.* **62**, 217–227.

Rabinowitz, M., Swift, H. (1970) Mitochondrial nucleic acids and their relation to the biogenesis of mitochondria, *Physiol. Rev.* **50**, 376–427.

Shadel, G.S., Clayton, D.A. (1997) Mitochondrial DNA maintenance in vertebrates, *Annu. Rev. Biochem.* **66**, 409–435.

Wolstenholme, D.R., Jeon, K.W. (1992) *Mitochondrial genomes*, Academic press, London.

Primary Literature

Anderson, S., Bankier, A.T., Barrell, B.G. (1981) Sequence and organization of the human mitochondrial genome, *Nature* **290**, 457–465.

Andrews, R.M., Kubacka, I., Chinnery, P.F. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, *Nat. Genet.* **23**, 147.

Awadalla, P., Eyre-Walker, A., Smith, J.M. (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA, *Science* **286**, 2524–2525.

Brown, W.M., George, M. Jr., Wilson, A.C. (1979) Rapid evolution of animal mitochondrial DNA, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1967–1971.

Brown, W.M., Prager, E.M., Wang, A., Wilson, A.C. (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution, *J. Mol. Evol.* **18**, 225–239.

Budar, F., Touzet, P., De Paepe, R. (2003) The nucleo-mitochondrial conflict in cytoplasmic male sterilities revisited, *Genetica* **117**, 3–16.

Cann, R.L., Stoneking, M., Wilson, A.C. (1987) Mitochondrial DNA and human evolution, *Nature* **325**, 31–36.

Cavelier, L., Johannisson, A., Gyllensten, U. (2000) Analysis of mtDNA copy number and composition of single mitochondrial particles using flow cytometry and PCR, *Exp. Cell. Res.* **25**, 79–85.

Cavelier, L., Jazin, E.E., Eriksson, I., Prince, J., Bave, U., Oreland, L., Gyllensten, U. (1995) Decreased cytochrome-c oxidase activity and lack of age-related accumulation of mitochondrial DNA deletions in the brains of schizophrenics, *Genomics* **29**, 217–224.

Cavelier, L., Erikson, I., Tammi, M., Jalonen, P., Lindholm, E., Jazin, E., Smith, P., Luthman, H., Gyllensten, U. (2001) MtDNA mutations in maternally inherited diabetes: presence of the 3397 ND1 mutation previously associated with Alzheimer's and Parkinson's disease, *Hereditas* **135**, 65–70.

Eyre-Walker, A., Smith, N.H., Smith, J.M. (1999) How clonal are human mitochondria? *Proc. R. Soc. Lond. B. Biol. Sci.* **266**, 477–483.

Giles, R.E., Blanc, H., Cann, H.M., Wallace, D.C. (1980) Maternal inheritance of human mitochondrial DNA, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6715–6719.

Gyllensten, U., Wharton, D., Wilson, A.C. (1985) Maternal inheritance of mitochondrial DNA during backcrossing of two species of mice, *J. Hered.* **76**, 321–324.

Gyllensten, U., Wharton, D., Josefsson, A., Wilson, A.C. (1991) Paternal inheritance of mitochondrial DNA in mice, *Nature* **352**, 255–257.

Harding, A.E., Sweeney, M.G., Govan, G.G. (1995) Pedigree analysis in Leber hereditary optic neuropathy families with a pathogenic mtDNA mutation, *Am. J. Hum. Genet.* **57**, 77–86.

Hauswirth, W.W., Laipis, P.J. (1982) Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4686–4690.

Hayashi, J., Ohta, S., Kikuchi, A. (1991) Introduction of disease-related mitochondrial DNA deletions into HeLa cells lacking mitochondrial DNA results in mitochondrial dysfunction, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10614–10618.

Howell, N., Halvorson, S., Kubacka, I. (1992) Mitochondrial gene segregation in mammals: is the bottleneck always narrow? *Hum. Genet.* **90**, 117–120.

Ingman, M., Gyllensten, U. (2001) Analysis of the complete human mtDNA genome: methodology and inferences for human evolution, *J. Hered.* **92**, 454–461.

Ingman, M., Kaessmann, H., Pääbo, S., Gyllensten, U. (2000) Mitochondrial genome variation and the origin of modern humans, *Nature* **408**, 708–713.

Janecek, L.L., Honeycutt, R.L., Adkins, R.M., Davis, S.K. (1996) Mitochondrial gene sequences and the molecular systematics of the Artiodactyl subfamily Bovinae, *Mol. Phylogenet. Evol.* **6**, 107–119.

Jenuth, J.P., Peterson, A.C., Fu, K. (1996) Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA, *Nat. Genet.* **14**, 146–151.

Johnson, A.A., Johnson, K.A. (2001a) Exonuclease proofreading by human mitochondrial DNA polymerase, *J. Biol. Chem.* **276**, 38097–38107.

Johnson, A.A., Johnson, K.A. (2001b) Fidelity of nucleotide incorporation by human mitochondrial DNA polymerase, *J. Biol. Chem.* **276**, 38090–38096.

Koehler, C.M., Lindberg, G.L., Brown, D.R. (1991) Replacement of bovine mitochondrial DNA by a sequence variant within one generation, *Genetics* **129**, 247–255.

Kraytsberg, Y., Schwartz, M., Brown, T.A., Ebralidse, K., Kunz, W.S., Clayton, D.A., Vissing, J., Khrapko, K. (2004) Recombination of human mitochondrial DNA, *Science* **304**, 981.

Krings, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M., Paabo, S. (1997) Neandertal DNA sequences and the origin of modern humans, *Cell* **90**, 19–30.

Lansman, R.A., Clayton, D.A. (1975) Selective nicking of mammalian mitochondrial DNA in vivo: photosensitization by incorporation of 5-bromodeoxyuridine, *J. Mol. Biol.* **99**, 761–776.

Larsson, N.G., Tulinius, M.H., Holme, E., Oldfors, A., Andersen, O., Wahlström, J., Aasly, J. (1992) Segregation and manifestations of the mtDNA tRNALys A→G(8344) mutation of myoclonus epilepsy and ragged-red fibers (MERRF) syndrome, *Am. J. Hum. Genet.* **51**, 1201–1212.

Larsson, N.G., Wang, J., Wilhelmsson, H., Oldfors, A., Rustin, P., Lewandoski, M., Barsh, G.S., Clayton, D.A. (1998) Mitochondrial transcription factor A is necessary for mtDNA maintenance and embryogenesis in mice, *Nat. Genet.* **18**, 231–236.

Lindholm, E., Cavelier, L., Howell, W.M., Eriksson, I., Jalonen, P., Adolfsson, R., Blackwood, D.H., Muir, W.J., Brookes, A.J., Gyllensten, U., Jazin, E.E. (1997) Mitochondrial sequence variants in patients with schizophrenia, *Eur. J. Hum. Genet.* **5**, 406–412.

Maddison, D., Ruvolo, M., Swofford, D. (1992) Geographic origins of human mitochondrial DNA: phylogenetic evidence from control region sequences, *Syst. Biol.* **41**, 111–124.

McShane, M.A., Hammans, S.R., Sweeney, M. (1991) Pearson syndrome and mitochondrial encephalomyopathy in a patient with a deletion of mtDNA, *Am. J. Hum. Genet.* **48**, 39–42.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., Sukernik, R.I., Olckers, A., Wallace, D.C. (2003) Natural selection shaped regional mtDNA variation in humans, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 171–176.

Rantanen, A., Jansson, M., Oldfors, A., Larsson, N.G. (2001) Downregulation of Tfam and mtDNA copy number during mammalian spermatogenesis, *Mamm. Genome* **12**, 787–792.

Rotig, A., Bessis, J.L., Romero, N. (1992) Maternally inherited duplication of the mitochondrial genome in a syndrome of proximal tubulopathy, diabetes mellitus, and cerebellar ataxia, *Am. J. Hum. Genet.* **50**, 364–370.

Saada, A., Shaag, A., Mandel, H. (2001) Mutant mitochondrial thymidine kinase in mitochondrial DNA depletion myopathy, *Nat. Genet.* **29**, 342–344.

Schwartz, M., Vissing, J. (2002) Paternal inheritance of mitochondrial DNA, *N. Engl. J. Med.* **347**, 576–580.

Shoubridge, E.A. (2001) Nuclear genetic defects of oxidative phosphorylation, *Hum. Mol. Genet.* **10**, 2277–2284.

Silva, J.P., Kohler, M., Graff, C., Oldfors, A., Magnuson, M.A., Berggren, P.O., Larsson, N.G. (2000) Impaired insulin secretion and beta-cell loss in tissue-specific knockout mice with mitochondrial diabetes, *Nat. Genet.* **26**, 336–340.

Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., Dominko, T., Simerly, C., Schatten, G. (1999) Ubiquitin tag for sperm mitochondria, *Nature* **402**, 371–372.

Tamas, I., Klasson, L., Canbäck, B., Näslund, A.K., Eriksson, A.-S., Wernegreen, J.J., Sandström, J.P., Moran, N.A., Andersson, S.G.E. (2002) 50 million years of genomic stasis in endosymbiotic bacteria, *Science* **296**, 2376–2379.

Tamura, K., Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees, *Mol. Biol. Evol.* **10**, 512–526.

Trifunovic, A., Wredenberg, A., Falkenberg, M., Spelbrink, J.N., Rovio, A.T., Bruder, C.E., Bohlooly, Y.M., Gidlof, S., Oldfors, A., Wibom, R., Tornell, J., Jacobs, H.T., Larsson, N.G. (2004) Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature* **429**, 417–423.

Wakeley, J. (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA, *J. Mol. Evol.* **37**, 613–623.

Wolfe, K.H., Li, W.-H., Sharp, P.M. (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 9054–9058.

Vesicle Budding and Fusion: *see* Membrane Traffic: Vesicle Budding and Fusion

Viral Inhibitors and Immune Response Mediators: The Interferons

Anthony Meager
Division of Immunology and Endocrinology, National Institute for Biological Standards and Control, South Mimms, Herts, UK

1 Historical Perspective 390

2 Interferon Genes and Proteins 391
2.1 Identification and Nomenclature 391
2.2 Interferon Genes 392
2.3 Interferon Proteins 393
2.3.1 IFN-α 393
2.3.2 IFN-ß 393
2.3.3 IFN-κ 396
2.3.4 IFN-ω 396
2.3.5 IFN-τ 396
2.3.6 IFN-γ (type II IFN) 396
2.3.7 IFN-λ (type III IFNs) 397

3 Regulation of IFN Synthesis 397
3.1 Type I IFNs 397
3.1.1 IFN-α 397
3.1.2 IFN-ß 399
3.1.3 IFN-κ 400
3.1.4 IFN-ω 400
3.1.5 IFN-τ 400
3.2 Type II IFN 400
3.2.1 IFN-γ 400
3.3 Type III IFNs 401
3.3.1 IFN-λs 401

4 Regulation of Cellular Gene Expression by IFNs 401
4.1 IFN Receptors 401

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

4.1.1 Molecular Characteristics 401
4.1.2 Receptor Signaling 402
4.2 IFN-response Gene Sequences 404

5 Proteins Induced by IFNs 404

6 Biological Activities Associated with IFNs 405
6.1 Antiviral Activity 405
6.1.1 Molecular Mechanisms 405
6.1.2 Defense Mechanisms of Viruses 407
6.2 Antiproliferative Activity 408
6.3 Antimicrobial Activity 408
6.4 Immunoregulatory Activity 409

7 Potential Physiological and Pathophysiological Roles of IFNs 410
7.1 Do IFNs have Physiological Roles? 410
7.2 Ways of Studying IFN "Physiology" 411
7.3 IFNs and Pathophysiological Phenomena 412
7.4 IFNs and Pathogenesis 412

8 Clinical Uses of IFNs 414

9 Final Remarks 418

Bibliography 419
Books and Reviews 419
Primary Literature 419

Keywords

Cytokine
One of a class of inducible biologically active proteins that exercise specific, receptor-mediated effects in target cells or in the cytokine-producing cells themselves.

Homology
Degree of structural similarity.

Interferon
One of a group of cytokines with antiviral activity, also displaying antimicrobial, antiproliferative, and immunomodulatory activities.

Interferon Receptors
Transmembrane glycoproteins specifically binding IFN and capable of signal transduction across the cell surface membrane.

Interferon/Type I
Acid-stable species of interferon, including, IFN-α, -ß, -ω, -κ, and -τ.

Interferon/Type II
Acid-labile species of interferon, IFN-γ.

Interferon/Type III
A small family of interferon-like molecules known as IFN-λs. Alternatively designated interleukin-28A and B and interleukin-29.

Subtype
One of two or more molecularly and antigenically related IFNs.

Tumor Cell
Cell that is capable of producing a tumor and that is unresponsive to mediators of normal cell growth control.

The interferons (IFNs), a category of biologically active proteins or cytokines with common antiviral activity, are induced by a variety of pathological stimuli, including viruses, bacteria, protozoa, and foreign antigens, as well as by endogenous cellular interactions. IFNs are secreted locally from induced cells to stimulate, via specific cell surface receptors, host defense mechanisms in the surrounding tissues. Such mechanisms are manifested as a wide spectrum of antiviral, antimicrobial, antitumor, and immunomodulatory actions, and have led to the belief that IFNs could have therapeutic potential in the treatment of many infections and invasive diseases.

Clinical investigations, begun in earnest in the 1980s with the abundant availability of "recombinant" IFNs, have demonstrated that IFNs, used as single therapeutic agents, induce beneficial responses in a select number of virus-mediated and malignant diseases, but not in the major human cancers, such as breast, lung, and colon. IFNs also induce significant, undesirable, but reversible, side effects. Nevertheless, IFNs continue to be widely used, for example, in the cancer area, and increasingly so for disease indications such as chronic hepatitis, caused by hepatitis-B and -C viruses, and multiple sclerosis, a chronic degenerative disease of unknown etiology.

1
Historical Perspective

The phenomenon of "interference" in virus infections was studied from the end of the 1940s. In 1957, Isaacs and Lindemann discovered a biological substance produced in virus-infected cells that in turn produced an antiviral response in cell cultures infected by the same virus. The word *interferon* (IFN) was coined for this substance. Initially, there were difficulties in characterizing IFN because under laboratory conditions it was made in only minute quantities. Gradually however, with the introduction of large-scale production facilities and the development of efficient purification procedures, the nature of IFN began to emerge. First, it was established that IFN was protein in nature. Second, it was discovered that there were different molecular species (types) of IFN. Third, it became evident that IFNs were associated with a growing number of biological activities beside their antiviral activity. One of these activities, the antitumor action, was the spur to increased efforts during the 1970s to develop IFN as an anticancer agent. Progress was made particularly in the area of purification and in the sequencing of the N-terminal regions of pure IFN proteins. However, the development of rDNA technology in the late 1970s was to revolutionize both the production of IFNs and their characterization. This technology was also to reveal that one type of human IFN, designated IFN-α, was composed of many different, but structurally related, IFN proteins, now known as subtypes. Using nucleotide-sequencing techniques, the amino acid sequences of IFN molecules could be deduced once complementary DNA (cDNA) copies of individual IFN messenger RNA (mRNA) were cloned. Genes for individual IFN proteins were also isolated, characterized, and their chromosomal location established. By comparing the IFN genes of other animal species to those of humans, evolutionary relationships and the ancestry of IFN genes were deduced.

The early 1980s were times of intensive research investigations into all aspects of IFN biochemistry and biology and the period that saw the instigation of many clinical trials to evaluate IFN as a potential therapeutic agent in cancer patients. Reports of initial successes in the clinic were followed by more sober analyses of IFN's worth as an anticancer agent. IFN clearly had potential, but tumor responses were seldom large enough to cure cancer patients. In addition, IFN produced a number of undesirable side effects in patients. The cloning in the mid-1980s of the second major type of IFN, now designated IFN-γ, led to a further expansion of biochemical and biological investigations and more clinical trials.

The period from around 1985 to date has witnessed many new discoveries about IFNs and their biological actions and featured the cloning of their cell surface receptors, responsible for triggering cellular responses following IFN binding. There has been a rationalization of the clinical results leading to the now-accepted view that IFN administrated as a single anticancer agent is beneficial in only a handful of malignancies, for example, hairy cell leukemia (which is extremely rare), renal cell carcinoma, AIDS-related Kaposi's sarcoma and chronic myelogenous leukemia (CML). Nevertheless, further attempts to exploit IFN's potential have and are being made by combining it with other anticancer drugs. IFN-α is now having some success as an antiviral agent in the treatment of hepatitis-B and -C virus infections,

while IFN-β has emerged as a therapeutic agent for multiple sclerosis.

2 Interferon Genes and Proteins

2.1 Identification and Nomenclature

Originally two types of human IFN were defined on the basis of the capacity of the biological activity of IFN preparation to withstand acidification to pH 2. These were designated type I (acid-stable) IFN and type II (acid-labile) IFN. On the basis of the cell source, type I IFN was subdivided into two subclasses: *leukocyte IFN* and *fibroblast IFN*. These were subsequently shown to be antigenically distinct. Type II IFN was often referred to as *immune IFN* to denote its production by stimulated T lymphocytes.

In 1978, an international IFN nomenclature committee recommended that *leukocyte IFN* should become IFN-α, *fibroblast IFN*, IFN-ß, and *immune IFN*, IFN-γ. These designations were widely accepted, but nowadays with increased knowledge of IFN genes, protein structures, and receptors, IFN-α and IFN-β are often classified as type I IFNs and IFN-γ as type II IFN. However, complications arose when it was discovered that IFN-α was heterogeneous and contained many different, but structurally and antigenically related species, now commonly referred to as subtypes. Some groups labeled these subtypes αA, αB, αC, αD, and so on, while others adopted the numerical system of α_1, α_2, α_4, and so on. The latter nomenclature now prevails. However, there is no general correspondence between letters and numbers, for example, $\alpha A = \alpha 2$; $\alpha B = \alpha 8$; $\alpha D = \alpha 1$. When an IFN preparation is a mixture of IFN-α subtypes, these are often designated IFN-αn. The discovery in the mid-1980s of additional IFN-α-like proteins antigenically distinct from the main family of IFN-α subtypes then created another problem. At first, the new IFN-α-like proteins were classified as IFN-α subclass II, with members of the original IFN-α family being referred to as IFNα subclass I. However, a radical departure from the sequence of letters in the Greek alphabet resolved this complex nomenclature. IFN-α subclass II has become IFN omega (ω)! Recently, another human type I IFN distantly related to either IFN-α or IFN-ω and expressed only in epidermal keratinocytes has been characterized; this is designated *keratinocyte-derived IFN*, or IFN-κ. The most recent addition to the human type I IFN family is IFN-ε, which appears distantly related to IFN-β but whose function remains uncertain.

In mice, similar IFN-α subtypes, IFN-β, and IFN-ε are present, but IFN-κ and IFN-ω are absent. However, a type I IFN-like molecule called limitin, which is 30% related to human IFN-ω, has recently been identified in mice. In other animal species (e.g. bovines), besides multiple IFN-α subtypes similar to those in humans, there are frequently several different IFN-ß and/or IFN-ω subtypes, although besides mice IFN-ω is also absent in canines. Further, other classes of IFN-α-like proteins have been identified in certain animal species, but not in humans; the *trophoblast* (τ) and *delta* (δ) IFNs are produced by the trophectoderm and are involved in the maternal recognition of pregnancy in ruminants (e.g. bovines) and pigs, respectively.

All members of the type I (α, β, δ, ε, κ, τ, and ω) and II (γ) IFNs induce antiviral activity. A further small family of human cytokines with limited homology to type I

IFNs have recently been characterized as also having antiviral activity. They have been designated IFN-λ1 (alternatively, interleukin-29 (IL-29)), IFN-λ2 (IL-28A), and IFN-λ3 (IL-28B) and collectively as type III IFNs. Their counterparts probably exist in other animal species, but IFN-λ1 (IL-29) is absent in mice.

The nomenclature system for designating IFNs according to their animal species of origin is reasonably straightforward. The first two letters of the animal name are most often used: human, Hu; murine, Mu; bovine, Bo, and so on. However, Mo is sometimes used for mouse IFN and Ra for rat IFN to distinguish between these murine species. In other cases, no abbreviations are used, as in, for example, chick IFN and rabbit IFN.

2.2 Interferon Genes

The number of IFNs varies among different animal species, but those of the IFN-α family are usually the most numerous. In humans, there are 13 nonallelic IFN-α genes (Note that these genes are designated as *IFNA1*, *IFNA2*, and so on, so that they correspond exactly to the IFN-α proteins they encode), plus four nonallelic *IFNA-like* pseudogenes (i.e. genes that cannot give rise to functional IFN-α molecules). In addition, allelic variants of some *IFNA* subtype genes (e.g. for *IFNA2*) have been found. There is one HuIFN-ß gene (designated *IFNB1*), one IFN-ε gene (designated *IFNE1*) one HuIFN-κ gene (designated *IFNK*), one HuIFN-ω gene (designated *IFNW1*) and several *IFNW*-like pseudogenes. All of these genes and pseudogenes are located on the short arm of human chromosome 9 where they are arranged in tandem. The equivalent site for mouse *IFNA* genes is in mouse chromosome 4. Each of the genes specifying IFN-α, -ß, -ε and -ω proteins has no introns, but the *IFNK* gene contains one intron in the 3′ noncoding region. Lack of introns is an unusual feature among mammalian cytokine genes, in which there are usually three or more introns (IFNs are considered to belong to the cytokine superfamily of multifunctional, intercellular biological mediators). The lack of introns suggests a very ancient origin of the common ancestral gene. It has been predicted that the primordial type I IFN gene arose some 500 million years ago. A split about 400 million years ago appears to have given rise to the first *IFNA* and *IFNB* genes. Since then, the *IFNA* gene has evolved and duplicated many times to give the multiple *IFNA* genes found in present day mammals. Around 100 million years ago, an *IFNA* gene apparently diverged enough to give rise to the *IFNW* gene family, which is found in most mammals. In ruminants, the *IFNW* gene family appears to have split around 60 to 80 million years ago to give the *IFNT* gene family, which is now only found in the Ruminantia suborder Artiodactyls (sheep, cattle). The relationship of the pig IFN-δ gene and the mouse limitin gene to the *IFNA*, *W*, *T* gene family remains uncertain. However, *IFNE1* and *IFNK* genes are more likely related to the *IFNB* gene than to the *IFNA*, *W*, *T* genes.

In contrast, throughout evolution there appears to have been no duplication of the IFN-γ gene (designated *IFNG*). Only a single *IFNG* gene has been found in all animal species tested. This gene has four exons and three introns, a common organizational feature of many cytokine genes, and is located on chromosome 12 in humans and chromosome 10 in the mouse. The *IFNG* gene also shares other structural features in common with cytokine

genes such as an inducible transcriptional enhancer in the 5′ flanking region found, for example, in the interleukin-2 gene.

The genes for IFN-λ1-3 (type III IFNs) are located on human chromosome 19 and contain introns. These genes appear to have emerged about the same time as type I IFN genes and are widespread in mammals, although the mouse lacks a functional IFN-λ1 gene.

All the active IFN genes have regulatory regions in their 5′ noncoding promoter sequences, which control transcription. In all coding regions, the part of the gene that encodes the mature IFN protein is preceded by a nucleotide sequence specifying an N-terminal signal peptide.

2.3 Interferon Proteins

2.3.1 IFN-α

All IFN-α's are secreted proteins and, as mentioned earlier, are transcribed from mRNA as precursor proteins, pre-IFN-α, containing N-terminal signal polypeptides of about 23 amino acids. The signal polypeptide is cleaved off before the mature IFN-α molecule is secreted from the cell. The mature form of most HuIFN-α subtypes contains 166 amino acids (except HuIFN-α_2, 165 amino acids: (Fig. 1). IFN-α subtypes from the mouse are of a similar size with 166 or 167 amino acids (except mouse IFN-α_4, 162 amino acids). The calculated molecular weight of HuIFN-α subtypes is approximately 18.5 kDa, although apparent molecular weights in sodium dodecyl sulfate (SDS)-polyacrylamide gels vary between 17 and 26 kDa. The amino acid sequences of HuIFN-α subtypes are highly related, with complete identity at 85 of the 166 (amino acid) positions. Many of the positions where amino acids differ from subtype to subtype are conservative substitutions. Interestingly, HuIFN-α subtypes contain four cysteine residues whose positions (1, 29, 99, and 139) are highly conserved (Fig. 1). These four cysteines form two disulphide bridges (1–99, 29–139), which induce folding of the IFN-α molecule and whose integrity is essential for biological activity. IFN-α subtypes are predicted to have a high proportion (ca 60%) of α-helical regions and are folded to form globular proteins. The three-dimensional structure of HuIFN-α_2 revealed by X-ray crystallographic analysis comprises 5α-helices (A–E) connected by loops in the "classic" up-up-down-down configuration. This α-helical "bundle" structure IFN-αs share with many other cytokines, particularly those of the hematopoietin family. Only one of the HuIFN-α subtype sequences contain sites for N-linked glycosylation; however, limited O-linked glycosylation may occur in several. Natural HuIFN-α_2, for example, appears to contain some O-linked oligosaccharides (N.B. Recombinant IFN-α subtypes produced by *Escherichia coli* are nonglycosylated since bacteria lack the biosynthetic machinery to add sugar residues to polypeptides).

Although the mouse IFN-α subtypes, of which 10 have been cloned and expressed, are only approximately 40% related in amino acid sequence to their HuIFN-α subtype counterparts, they are likely to be structurally similar. However, in contrast to HuIFN-α subtypes, most mouse IFN-α subtypes contain one N-linked glycosylation site and are probably glycoproteins when produced from mouse somatic cells.

2.3.2 IFN-ß

Pre-HuIFN-ß contains 187 amino acids, of which 21 comprise the N-terminal

```
α1        -23M A S P F A L L M V L V V L S C K S S C S L G1C D L P
α2           M A L T F A L L V A L L V L S C K S S C S V G1C D L P
αII(w)       M V L L L P L L V A L P L C H C G P C G S L S1C D L P
β         -21M T N K C L L Q I A L L L C F S T T A L S1M S Y N L L
γ         -23M K Y T S Y I L A F Q L C I V L - G S L G C Y C Q1D P

α1        E T H S L D N R R T L M L L A Q M S R I S P S S C L M D R
α2        Q T H S L G S R R T L M L L A Q M R K I S L F S C L K D R
αII(w)    Q N H G L L S R N T L V L L H Q M R R I S P F L C L K D R
β         G F L Q R S S N F Q C Q K L L W Q L N G R L E Y C L K D R
γ         Y V K E A E N L K K Y F N A G H S D V A D N G T L F L G I

α1        H D F G F P Q   E E F D G N Q F Q K A P A I S V L H   E L
α2        H D F G F P Q   E E F - G N Q F Q K A E T I P V L H   E M
αII(w)    R D F R F P Q   E M V K G S Q L Q K A H V M S V L H   E M
β         M N F D I P E   E I K Q L Q Q F Q K E D A A L T I Y   E M
γ         L K N W K E E   S D R K I M Q S Q I V S F Y F K L -   - -

α1        I Q Q I F N L F T T K D S S A A W D E D L L D K F C T E L
α2        I Q Q I F N L F S T K D S S A A W D E T L L D K F Y T E L
αII(w)    L Q Q I F S L F H T E R S S A A W N M T L L D Q L H T E L
β         L Q N I F A I F R Q D S S S T G W N E T I V E N L L A N V
γ         F K N F K D D Q S I Q K S V E T I K E D M N V K F F N S N

α1        Y Q   Q L N D L E A C V M Q E E R V G E T P L M N A D S I
α2        Y Q   Q L N D L E A C V I Q G V G V T E T P L M K E D S I
αII(w)    H Q   Q L Q H L E T C L L Q V V G E G E S A G A I S S P A
β         Y H   Q I N H L K T V L E E K L E K E D F T R G K L M S S
γ         K K   K R D D F E K L T N Y S V T D L N V Q R K A I H E L

α1        L A V K K Y F R R I T L Y L T E K K Y S P C A W E V V R A
α2        L A V R K Y F Q R I T L Y L K E K K Y S P C A W E V V R A
αII(w)    L T L R R Y F Q G I A V Y L K E K K Y S D C A W E V V R M
β         L H L K R Y Y G R I L H Y L K A K E Y S H C A W T I V R V
γ         I Q V M A E L - - - - - - - - - - - - S P A A K T G K R K

α1           E I M R S L S L S T N L Q E R L R R K E166
α2           E I M R S F S L S T N L Q E S L R S K E165
αII(w)       E I M K S L F L S T N M Q E R L R S K D A D L G SS172
β            E I L R N F Y F I N R L T G Y L R N166
γ            R - - - - - - - S Q M L - F R G R R A S Q143
```

Fig. 1 Amino acid sequences of human IFNs, α_1, α_2, αII (ω), ß and γ. Sequences are aligned to show maximum structural relatedness. Sites of N-linked glycosylation (Asn = N) are underlined.

signal polypeptide. The signal polypeptide is cleaved on secretion to yield the mature HuIFN-ß protein of 166 amino acids, the same size as most HuIFN-α subtypes (Fig. 1). However, HuIFN-ß shows only approximately 30% amino acid

sequence relatedness with HuIFN-α and is antigenically distinct from it. The mature HuIFN-ß molecule lacks the N-terminal cysteine residue at position 1 found in HuIFN-α subtypes, but contains three other cysteines at position 17, 31, and 141. The cysteines at position 31 and 141 are equivalent to those found at positions 29 and 139 in HuIFN-α and are joined to form a single disulphide bond. Replacement of the cysteine residue at position 17 by serine does not result in any loss of biological activity, whereas serine substitution of Cys^{141} does. Unlike most HuIFN-α subtypes, HuIFN-ß's amino acid sequence contains one site, asparagine at position 80, for N-linked glycosylation and it is known that HuIFN-ß produced by human fibroblasts is a glycoprotein of approximately 23 kDa. Like HuIFN-α subtypes, HuIFN-ß is predicted to have a high α-helical content, and folds up as an α-helical bundle of 5 domains.

Mouse IFN-ß contains 161 amino acids and shows some 60% relatedness to HuIFN-ß: in contrast, however, mouse IFN-ß contains only one cysteine residue and thus intramolecular disulphide bonds are impossible. There are three N-glycosylation sites present in mouse IFN-ß, and the rather high apparent molecular weight of 35 kDa probably reflects extensive glycosylation of this molecule. The three-dimensional crystal structure of recombinant mouse IFN-ß has been solved, revealing that the structure consists of five α-helices folded into a compact α-helical bundle (Fig. 2). There are many structural similarities between mouse IFN-β and HuIFN-α_2. Subsequently, the three-dimensional structure of HuIFN-β has also been shown to be similar, but dimers may form through a zinc-binding site present at the interface between two protomers. On the basis of comparative sequence analyses, it is predicted that in all mammalian IFN-α, IFN-ß, and IFN-ω proteins, the five α-helical domains are conserved. Pig IFN-δ, mouse limitin, IFN-ε, and ruminant IFN-τ are

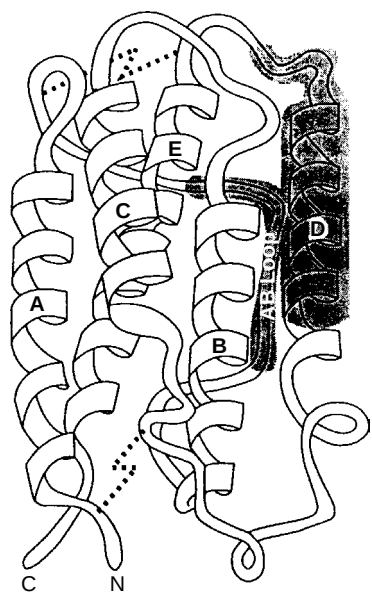

Fig. 2 Schematic drawing of the "side" view of a recombinant MuIFN-ß molecule. Helices A-E and an intervening loop AB are labeled N- and C-termini are marked. The three shaded areas along loop AB, helix D, and loop DE appear to comprise a spatially continuous binding site for the IFN receptor(s). (Reprinted with permission of Senda, et al. (1992) *EMBO J.* **11**, 3193; Copyright EMBO/IRL Press Ltd., UK).

also predicted to have the same basic structure.

2.3.3 IFN-κ

HuIFN-κ, made principally by keratinocytes, is a recently identified novel member of the type I IFN family displaying around 30% homology to HuIFN-α, HuIFN-β, or HuIFN-ω. In distinction to the latter, HuIFN-κ contains an additional cysteine at its C-terminus following the four conserved cysteines, and an insertion of 12 amino acids between the C and D helices. Its full amino acid sequence is therefore predicted to comprise 180 amino acids. However, sequence alignment with other type I IFNs predicts that HuIFN-κ has an overall structure typical of this family. Molecular modeling of HuIFN-κ has indicated a structure consisting of five α-helices as in other type I IFNs.

2.3.4 IFN-ω

The structure of IFN-ω is similar to that of IFN-α subtypes, but there have been sufficient amino acid substitutions to make it only about 60% related and thus antigenically distinct from them (HuIFNω is only 30% related to HuIFN-ß). Like pre-IFN-α, pre-IFN-ω has a signal sequence of 23 amino acids. However, the mature form of HuIFN-ω contains 172 amino acids, six more at the C-terminal end than HuIFN-α subtypes (Fig. 1). The four cysteine residues occur in precisely the same positions, 1, 29, 99, and 139, as they do in HuIFN-α subtypes. With regard to glycosylation, HuIFN-ω resembles HuIFN-ß more than HuIFN-α in having a single N-glycosylation site at Asn78.

IFN-ω is not found in mice or dogs, but three functional IFN-ω subtypes have been identified in sheep, cattle, and horses.

2.3.5 IFN-τ

Over a decade ago, it was discovered that ovine trophoblastic protein-1 (OTP-1), a hormonelike molecule having antiluteolytic activity produced by the preimplantation ruminant conceptus during the second and third weeks of pregnancy, has antiviral activity. This led to the finding that OTP-1 was itself structurally related to type 1 IFN, in particular to IFN-ω. As a consequence, OTP-1 and related proteins have been designated by one group of researchers as trophoblast IFNs or IFN-τ. Preovine IFN-τ contains 195 amino acids; a signal sequence of 23 amino acids followed by the mature polypeptide of 172 amino acids. IFN-τ molecules all contain the precisely conserved cysteines at positions 1, 29, 99, and 139 found both in IFN-α and IFN-ω molecules. Like IFN-ω, some IFN-τ subtypes possess a potential site for N-glycosylation on Asn78. It is predicted that IFN-τ will have a similar three-dimensional structure to IFN-α and IFN-ß (i.e. a bundle configuration of five α-helices).

IFN-τ subtypes are only found in the Ruminantia suborder Artiodactyla (e.g. sheep, goats, cattle), where they appear to have evolved to perform a special antiluteolytic function in early pregnancy. In man and other mammals, the fertilized egg, the conceptus, immediately attaches to the uterine wall and therefore there is no special need to maintain the corpus luteum as in ruminants, where the conceptus undergoes extensive preimplantation development.

2.3.6 IFN-γ (type II IFN)

The nucleotide sequence of HuIFN-γ mRNA contains an open-reading frame of 166 codons. At first, it was believed that the mature HuIFN-γ protein contained 146 amino acids, the 20 N-terminal amino

acids of pre-HuIFN-γ forming the signal peptide. However, this proved incorrect, inasmuch as cleavage of the signal peptide and the mature HuIFN-γ occurred between residues 23 and 24 of the pre-HuIFN-γ (Fig. 1). It was also shown that the N-terminal amino acid of HuIFN-γ was pyroglutamic acid and that several of the C-terminal amino acids of the natural 143 amino acid polypeptide could be cleaved off without loss of biological activity. The predicted molecular weight of the unmodified polypeptide is approximately 17 kDa, but there are two N-glycosylation sites (Asn25 and Asn97), both of which are used to yield fully glycosylated HuIFN-γ molecules of 25 kDa (Note that HuIFN-γ preparations derived from T lymphocytes often contain minor amounts of a 20-kDa molecule representing glycosylation at one site only.). In its native state, HuIFN-γ is not only heavily glycosylated but is also a homodimer of two 25-kDa molecules (subunits). The role of the oligosaccharide side chains is not known. Since unglycosylated HuIFN-γ produced from *E. coli* containing IFN-γ cDNA is biologically active, they appear to be dispensable.

The amino acid sequence of HuIFN-γ shows very little homology with those of type I IFNs (α, ß, κ, ω, or τ). Despite this, rather interestingly the three-dimensional structure of HuIFN-γ (illustrated in Fig. 3), which was reported in 1991, contains some features in common with IFN-ß. Like the type I IFNs, IFN-γ molecules are essentially α-helical in content, with no ß-sheet. IFN-γ subunits have six α-helical regions as opposed to the five present in the IFN-ß structure. On dimerization of IFN-γ subunits, five of the 12 helices are folded to form a structural domain that resembles the folded arrangement of the five helices in IFN-ß. Other similarities in folding topology between IFN-γ and IFN-ß are apparent, suggesting that despite their present low degree of homology, an ancient precursor to both types might have existed.

Mouse IFN-γ containing 133 amino acids is smaller than HuIFN-γ and is only 40% related in sequence. It also is heavily glycosylated and the molecular weight of the natural dimeric molecule is around 38 kDa. Its structure, and those of IFN-γ from other species, are expected to be similar to that worked out for the three-dimensional structure of HuIFN-γ.

2.3.7 IFN-λ (type III IFNs)

The IFN-λs contain 200 amino acids: in sequence IFN-λ1 is 81% related to IFN-λ2; IFN-λ2 is 96% related to IFN-λ3. IFN-λ1 contains an N-glycosylation site that is not present in IFN-λ2 or IFN-λ3. The IFN-λs are distantly related to type I IFNs (15–19% homology) and to the interleukin-10 (IL-10; 11–13% homology) family of cytokines. The 3-D structures of IFN-λs are not yet known, but since cysteine patterns are conserved on alignment of amino acid sequences with those of IFN-α and IL-10, it is predicted that IFN-λs are helical cytokine family members.

3 Regulation of IFN Synthesis

3.1 Type I IFNs

3.1.1 IFN-α

IFN-α can be induced by a whole variety of stimuli, including viruses, bacteria, xenogeneic or allogeneic tumor cells, virus-infected cells, B-cell mitogens, and so on. IFN-α subtypes are produced by stimulated or activated null lymphocytes,

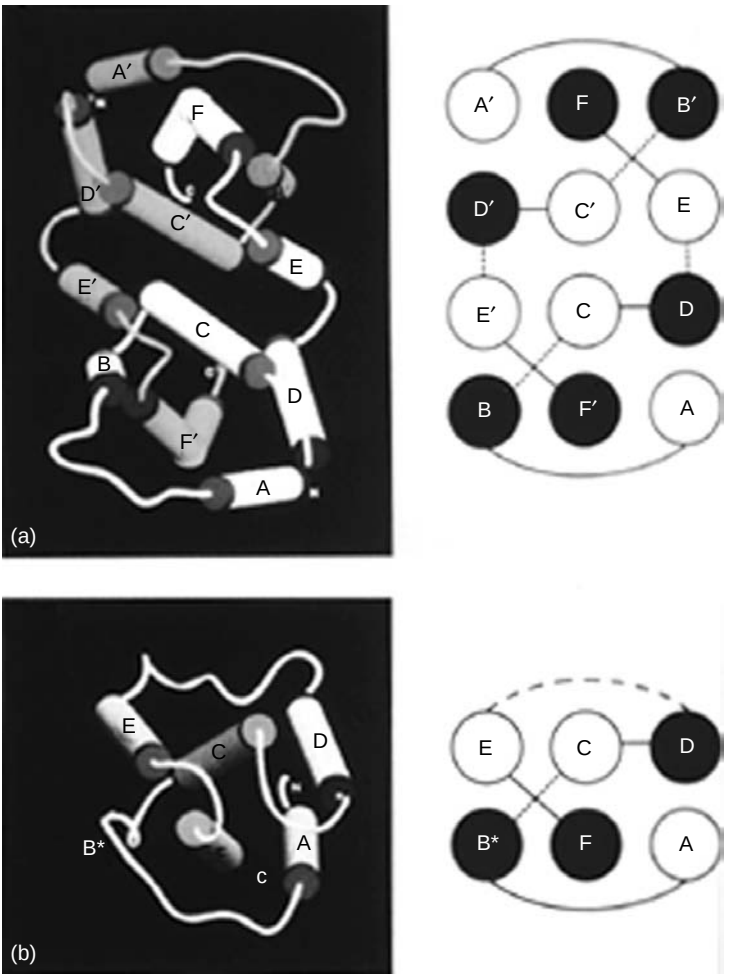

Fig. 3 Schematic drawings of (a) the recombinant human IFN-γ D′ dimer and (b) IFN-ß. On the left-hand side, the α helices are represented as cylinders, and the nonhelical regions are shown as tubes. The N- and C-terminal ends are marked. On the right-hand side, the α-helices are shown as circles, and the nonhelical regions are shown as solid or dashed lines, depending on whether they are above or below the plane of the figure. The regions labeled B* represent a kinked region of IFNß that is not helical, but shows some helical features. (Reprinted with permission of Ealick, et al. (1991) *Science*, **252**, 698. Copyright 1991 by American Association for the Advancement of Science).

monocytes, and lymphoblastoid cell lines of B-cell origin. More recently, plasmacytoid dendritic cells, a subset found mainly in lymph nodes and thymus, have been characterized as IFN producer cells (IPC) that produce large amounts of IFN-α following virus stimulation. The proportion of each IFN-α subtype in the mixtures

of subtypes produced as a consequence of stimulation/activation varies according to the nature of the stimulus and the producer cell type. It is still unclear why so many IFN-α subtypes are required and how their proportions are regulated at the level of their genes (but see below).

Normally, *IFNA* genes are tightly repressed, but when cells are infected by a virus, for example, "virus-responsive elements" in the *IFNA* genes are activated leading to derepression. *IFNA* genes contain several IFN gene regulatory elements (IREs), termed positive-response domains (PRDs) and negative-response domains (NRDs), in their 5′ noncoding promoter region whose sequences appear to govern inducibility by various inducers. Such IREs are acted upon by host cell interferon regulatory factors (IRF), a family of 9 known structurally related proteins, which act as transcriptional enhancers (e.g. IRF-1, −3, −5, −7) or repressors (e.g. IRF-2). It is now believed that IRF-3, which is constitutively present and phosphorylated by a serine kinase early after virus infection, plays a key role in IFN-α induction. Phosphorylated IRF-3 dimerizes, translocates to the nucleus, and activates the transcription of an "early" *IFNA* (most likely *IFNA4*) gene together with *IFNB*. The newly made "early" IFNs act in an autocrine/paracrine manner to activate type I IFN signaling pathways (see under Sect. 4.1.2) leading to transcription of IFN responsive genes, including IRF-7, an inducible transcription factor, which, following phosphorylation in virus-infected cells, activates several other *IFNA* genes plus many host defense genes (e.g. cytokines, chemokines). Thus, IFN-α production appears to be amplified in a controlled manner subject to the virus-dependent, sequential activation of specific IRFs. However, *IFNA* gene transcription leading to synthesis of IFN-α mRNAs continues only for a short interval before the genes are shut down again, even if the presence of the inducer is maintained, though IFN-α mRNA expression can be extended by certain drugs (e.g. cycloheximide, which blocks protein synthesis and hence the formation of labile repressors such as IRF-2).

3.1.2 IFN-ß

The regulation of transcription of the *IFNB* gene has been more closely studied than that of *IFNA* genes. Here, it is known that within the 120 bp 5′ to the transcription initiation site, there are at least six IRE, four PRD (I–IV), and two NRD (I, II). PRDI and PRDII are involved in both positive control and in postinduction repression. Several protein factors have been identified, which bind to PRDI including IRF-1 and IRF-3, transcriptional enhancers, and IRF-2, a repressor. PRDII resembles an NFκB transcription factor binding site and agents that activate NFκB (e.g. viruses, double-stranded RNA, phorbol esters), therefore promote IFN-ß induction. In contrast to *IFNA* gene regulation, active NFκB is probably a requirement for efficient *IFNB* gene transcription.

Although the major regulatory elements within the *IFNB* gene promoter have been identified, the mechanisms by which the gene is differentially induced and shut off are still not completely understood. However, research into the signaling pathways involved in following virus, double-stranded RNA (dsRNA), or cytokine stimulation has increased our knowledge of potential mechanisms for induction. For example, it is apparent that dsRNA, often an intermediate in viral genome-directed

nucleic acid synthesis, can act as a "proximal" inducer (e.g. via Toll-like receptor-3) of IRF-3 and NFκB activation, which are then able to activate transcription of the *IFNB* gene. The production of IFN-β appears to be the first stage in the cooperative activation of other IRFs (e.g. IRF-5, IRF-7), which appear to be crucial for transcriptional induction of *IFNA* and *IFNW* genes. In the absence of IRF-2, viral induction of IFN-ß is increased.

3.1.3 IFN-κ

Unlike other type I IFN genes, the *IFNK* gene appears to be constitutively active to some extent, particularly in IFN-κ-producer cells such as keratinocytes. However, IFN-κ mRNA expression is reported to be upregulated by viral infection, by other type I IFNs, and characteristically, by IFN-γ (type II IFN) in monocytes and dendritic cells. Currently, there is no information about which transcription factors are involved in *IFNK* gene activation.

3.1.4 IFN-ω

The *IFNW* gene is virus-inducible, like *IFNA* and *IFNB* genes, and, since there is much sequence identity with many of the *IFNA* subtype genes, it is assumed that its regulation will also be similar to that applying to *IFNA* genes. As a proportion of the antiviral activity present in the IFN mixture secreted by virally induced leukocytes, IFN-ω constitutes approximately 15% of the total.

3.1.5 IFN-τ

The *IFNT* genes are not virus-inducible and are therefore regulated in a different manner to the other type I IFN genes. IFN-τ is produced by trophoblast cells and the temporal pattern of expression depends on the time frame of maternal recognition of pregnancy, which varies from one to three weeks in ruminants. For instance, ovine (Ov) IFN-τ is only produced in large amounts between the 13th and 21st days of pregnancy, although OvIFN-τ mRNA can be detected as early as 8 to 10 days. It is believed that a factor(s) in the luteal phase environment of the uterus is required to trigger and maintain expression of IFN-τ genes. The promoter regions of *IFNT* genes appear to be under different regulatory control than other type I IFN genes, despite a high degree of superficial similarity in nucleotide sequences across all members of this gene family.

3.2 Type II IFN

3.2.1 IFN-γ

IFN-γ is produced mainly by antigen- or mitogen-stimulated mature, T lymphocytes and natural killer (NK) cells. Some T-cell lymphoblastoid lines, especially those chronically infected with HLTV-I, also can be stimulated to produce IFNγ. Like IFN-τ, IFN-γ does not appear to be virus-inducible. The gene structure of *IFNG* is quite apart from the intron-less type I IFN genes and resembles more the four exon–three intron arrangement found in cytokine genes (e.g. IL-2, IL-4, and GM-CSF). There are also similarities in the 5′ flanking region, upstream from the initiation site, of the *IFNG* gene and the 5′ flanking regions of such cytokine genes. *IFNG* gene transcription is probably controlled by binding of nuclear transcription factors to regulatory sequences in the 5′ flanking region. Since phorbol esters can act as (co)-stimulators of IFN-γ induction, it is probable that activation of NFκB is involved in gene derepression.

3.3 Type III IFNs

3.3.1 IFN-λs

Transcription of IFN-λ genes is stimulated by the same inducers of IFN-α, -β, and -ω genes, that is, several viruses, dsRNA (e.g. poly I.C.), in a variety of cell lines and dendritic cells. Virus-infected cells produce IFN-λ activity. The regulation of mRNA expression of IFN-λ genes is likely to involve a positive feedback mechanism, similar to that operating for IFN-α/β genes, but details are not yet available.

4 Regulation of Cellular Gene Expression by IFNs

4.1 IFN Receptors

4.1.1 Molecular Characteristics

IFNs exercise their actions in cells via IFN-specific cell surface receptors. These receptors bind IFNs with high affinity and transduce the signal occasioned by ligand (IFN)-binding across the cell membrane into the cytoplasm. All type I IFNs appear to share a common receptor, whereas the sole type II IFN, IFN-γ, has a distinct receptor. This division of receptors has been corroborated by the cloning of both type I and type II IFN cell surface binding proteins, which revealed little apparent relatedness of their respective amino acid sequences. However, these IFN binding proteins, two for type I IFN and one for type II IFN, share structural features that are frequently found in cytokine receptors; each has a large N-terminal extracellular domain, which is glycosylated, and which forms the binding site for IFN, a comparatively small transmembrane region and an intracellular domain of variable size (Fig. 4). The genes for the type I IFN binding proteins are located on human chromosome 21 and that for the type II IFN binding protein on human chromosome 6. It seems probable that the type I IFN binding proteins are intimately associated and together form high-affinity receptors. The type II IFN binding protein also appears to require a second component or accessory factor (AF), coded for by a gene located on chromosome 21, to form a functional receptor. One group has deduced that possibly the two types of receptor are evolutionarily connected and in their extracellular domains resemble the structures of cytokine receptors of the hematopoietin family (Fig. 4). It is now widely accepted that IFN receptors together with the individual receptors for IL-10 and its paralogs constitute a distinct subclass of hematopoietin receptors, currently referred to as *type II cytokine receptor family* (CRF2).

Comparison of protein sequences of the IFN receptors of different species has indicated that these have evolved with their cognate IFNs in a concerted fashion, and this has resulted in some degree of species specificity for most IFNs. For example, most HuIFN-α subtypes, HuIFN-ß and HuIFN-γ are poorly bound by mouse IFN receptors and vice versa. However, in some instances some IFNs are able to act across species; for example, HuIFN-α subtypes bind as well or better to bovine type I IFN receptors and can trigger antiviral activity in bovine cells.

It has recently been shown that the receptors for IFN-λs are contained within the CRF2 family. All IFN-λs bind to a unique receptor chain known as IL-28Rα (CRF2-12), which forms a functional heterodimeric receptor by combining with the accessory receptor chain of IL-10,

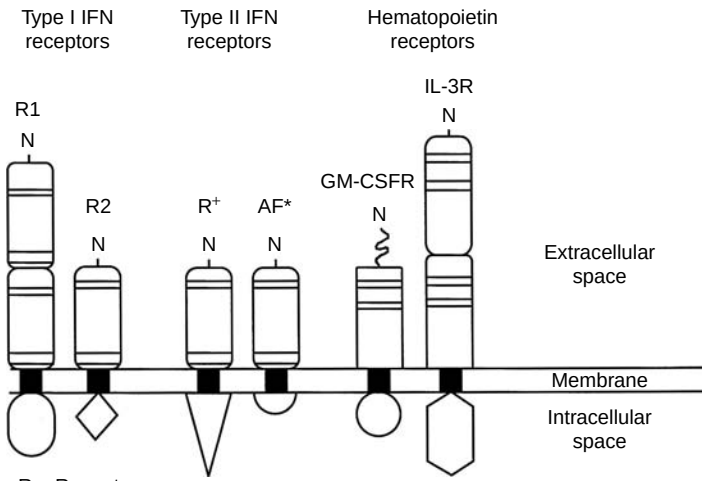

+R – Receptor
*AF – Accessory factor

Fig. 4 Schematic drawing of the domain organization of the human type I IFN and type II IFNγ receptors and representative, analogous hematopoietin receptors. The extracellular segments of the granulocyte-macrophage colony-stimulating factor receptor (GM-CSFR) and interleukin-3 receptor (IL-3R) are drawn to linear scale, and conserved cysteines are marked by black bands. Adapted from Bazan, Shared architecture of hormone binding domains in type I and type II interferon receptors, *Cell* (1990) **61**, 753. Copyright Cell Press.

designated IL-10R2. It is not unusual for different cytokine ligands to share receptor components; besides IL-10 and IFN-λs, IL-10R2 is also an accessory receptor chain of the IL-22 and IL-26 receptors.

4.1.2 Receptor Signaling

The intracellular domains of the type I and type II IFN receptors are not predicted to have kinase activity of any sort: many growth factor receptors (e.g. epidermal growth factor receptor) contain integral tyrosine kinase, which is essential for phosphorylation of downstream, intracellular mediators of the signaling pathway. Nevertheless, IFN receptors do associate with intracellular, nonreceptor kinases. The type I IFN-receptor chains are linked with nonreceptor tyrosine kinases, designated TYK2 (R1) and JAK1 (R2), whereas the type II IFN-receptor chains are associated with JAK1 (R1) and JAK2 (R2 or AF). In type I IFN signaling, TYK2 and JAK1 tyrosine kinases further link to what is known as interferon-stimulated gene factor-3 (ISGF-3), a complex of three/four distinct proteins that contains the primary transcriptional activator for IFN-induced genes. ISGF-3 is normally present in an inactive form being composed of inactive ISGF-3α (three related proteins: p84/91, two alternatively spliced transcripts of the same gene, and p113) and a constitutive and IFN-inducible functional component ISGF-3γ (now synonymous with IRF-9) that is associated with p113. The ISGF-3α components have been redesignated as "Signal Transducers and Activators of Transcription" (STATs); p84/91 and p113 are thus STAT1 and STAT2, respectively. STAT1 contains a DNA-binding domain

and a nuclear location signal (NLS), which is absent in STAT2; IRF-9 does have an NLS and thus by binding to STAT2 confers ability to be translocated to the nucleus. Following phosphorylation by TYK2 and JAK1, STAT1 and STAT2 dimerize and the ISGF-3 complex (STAT1/STAT2/IRF-9) translocates to the nucleus, where it acts as a transcription factor for the activation of type I IFN-inducible genes (Fig. 5). In contrast, type II IFN (IFN-γ) stimulation leads to phosphorylation of only p91 STAT1: homodimers of STAT1 act as transcription factors for the activation of IFN-γ inducible genes.

For type III IFNs, the IFN-λs, it appears that the same signal transduction pathway operating for type I IFNs is activated, despite the distinct receptors for IFN-λs. This similarity in signaling is probably due to JAK1 and TYK2 association with IL-28Rα and IL-10R2, respectively, the two receptor chains of the IFN-λ (IL-28/-29) receptor complex, and the contingent activation of ISGF3 (see above).

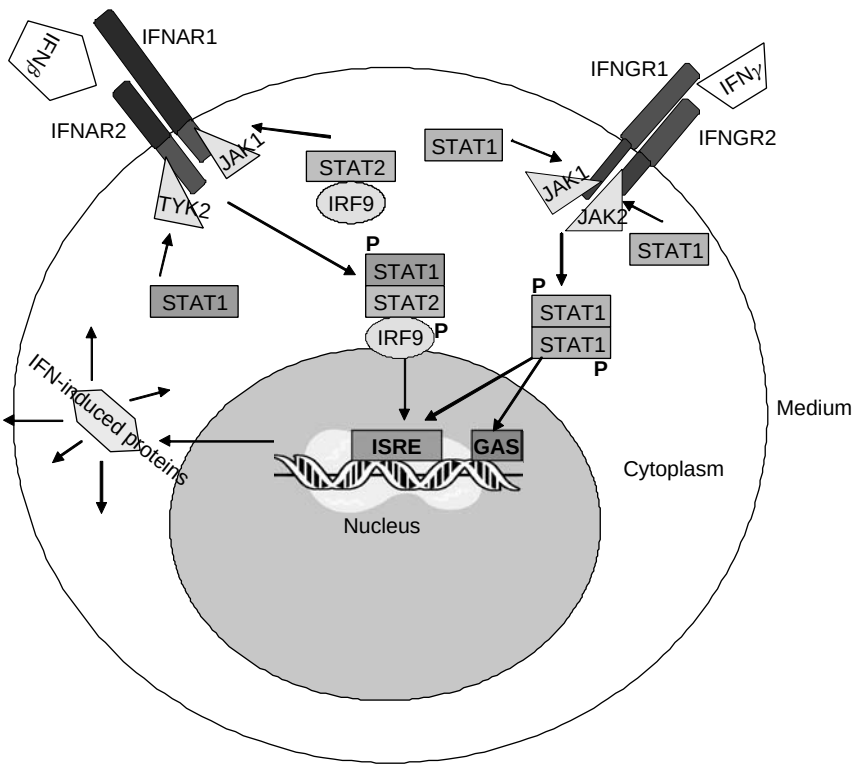

Fig. 5 Schematic drawing depicting signal transduction pathways from type I and type II IFN receptors at the cell membrane to the cell nucleus. ISGF3, interferon-stimulated gene factor-3. The ISGF3α complex contains three structurally related proteins (p84), p91[STAT1] and p113 [STAT2], and the ISGF3γ subunit is composed of a single protein, p48/IRF-9. ISRE, interferon-stimulated response element, present in IFN-inducible genes. GAS, γ-activated sequences, present in IFN-γ-inducible genes. JAK1, JAK2 and TYK2, nonreceptor protein tyrosine kinases involved in the phosphorylation of ISGF3α proteins, p91 [STAT1] and p113 [STAT2].

While the "JAK-STAT pathway" is the major intracellular signaling mechanism activated in IFN-treated cells, it may not be the only one. There has been some, albeit somewhat controversial, evidence to implicate protein kinase C (PKC) pathways as well. In particular, the type II IFN receptor appears to trigger PKC translocation from membrane to cytoplasm upon IFN-γ binding. However, possibly other non-PKC serine/threonine kinases are activated and are responsible for further signal amplification and activation of transcription factors. IFN-γ is a known inducer of IRF-9, which is necessary for the translocation of type I IFN-activatable ISGF-3 from cytoplasm to nucleus.

4.2 IFN-response Gene Sequences

Soon after the binding of type I IFNs to their cognate receptors, a number of nuclear genes are activated and transcripts (mRNA) of a variety of proteins are synthesized. The type I IFN-inducible genes have in common a regulatory nucleotide sequence in their 5′ flanking promoter region, and this type of sequence, which resembles the IRE present in IFN genes, has been designated IFN-stimulated response element (ISRE). The resemblance between ISRE and IRE could account for the fact that many IFN-inducible genes are transcriptionally activated by virus infection or dsRNA, which also activate the transcription of IFN genes (but not *IFNT* genes). The ISRE for type II IFN (IFN-γ), are similar, but generally known as γ-activation sequences (GAS). In contrast to the almost consensus ISRE sequences in type I IFN-inducible genes, it appears that GAS are likely to be heterogeneous. It is evident that signaling pathways of both type I and type II IFN systems share common signal transduction components (e.g. STAT1) leading to "convergent" transcriptional activation of a common set of IFN-inducible genes (Fig. 5). However, IFN-γ also induces a unique set of protein genes not induced by type I IFNs suggesting the involvement of other transcription factors. For example, IRF-1 (formerly known as ISGF-2) forms complexes with ISRE in both type I and type II IFN-stimulated cells, but is most particularly involved in transcription of type II IFN-inducible genes, for example, class II major histocompatibility (MHC) antigen genes. Current experimental evidence strongly indicates that type III IFNs, the IFN-λs, induce the same genes as type I IFNs, although probably more weakly.

5 Proteins Induced by IFNs

It is well established that the biological activity of IFNs depends on selective protein synthesis. Thus, it is clear that at least some of the set of IFN-inducible proteins must mediate the biological activity of IFN. IFN-inducible proteins, whose number probably runs into the 100s, include both those proteins induced early after IFN stimulation and those proteins that are produced at later times, often in response to the actions of early IFN-inducible proteins. Many IFN-inducible proteins are induced by both type I IFN and type II IFN, although some may be preferentially induced by either type of IFN. Yet other proteins are exclusively induced by either type I or type II IFN. In some cases, IFN-inducible proteins may be completely absent from the cell before IFN stimulation, but in other cases the effect of IFN causes an overall

increase in production of proteins that are already being expressed. Table 1 shows an incomplete list of IFN-inducible proteins together with their likely functions and their relative inducibility by type I and type II IFNs. Note that since IFN-inducible proteins tend also to be cell type specific, not all of the proteins listed will be expressed in all cell types.

6 Biological Activities Associated with IFNs

6.1 Antiviral Activity

6.1.1 Molecular Mechanisms

As mentioned earlier, viruses often induce IFN responsive genes, including type I IFNs, *IFNA, IFNB*, and *IFNW* and type

Tab. 1 IFN-inducible proteins.

Protein	*Function*	*Induction by*	
		Type I IFN	*Type II IFN*
2,5 A-synthetase	dsRNA-dependent synthesis of ppp-(A2p)n-A (2-5A); activator of RNase L	+	+
Protein (p68) kinase (PKR), dsRNA-activatable protein kinase	Phosphorylation of peptide initiation factor eIF-2α	+	+
MHC class I (HLA-A, B, C)	Antigen presentation to CTL	+	+
MHC class II (HLA-DR)	Antigen presentation to Th-lymphocytes	±	+
Indoleamine 2, 3-dioxygenase	Tryptophan catabolism	−	+
Guanylate-binding proteins (GBP; γ67)	GTP, GDP binding	±	+
Mx	Specific inhibition of influenza virus	+	−
IRF1/ISGF2	Transcription factor	±	+
IRF2	Transcription factor	+	−
IP-10	Related to chemotactic IL-8-like cytokines	±	+
Metallothionein	Metal detoxification	+	+
(TNF) receptors	Mediate TNF action	±	+
IL-2 receptors	Mediate IL-2 action	−	+
Intercellular adherence molecule-1 (ICAM-1)	Endothelial cell adhesion protein	−	+
Immunoglobulin Fc-receptor (FcR)	Ig binding by macrophages/neutrophils	±	+
Thymosin ß4	Induction of terminal transferase in lymphocytes	+	?
ß2-microglobulin	Immune function	+	+
Nitric oxide synthetase	Production of nitric oxide from arginine. Increased microbicidal activity in macrophages	−	+
200 family	From cluster of 6 genes; p204 is located in nucleolus	+	−
6–16	Unknown; extracellular	+	±
1–8/9–27	Cell surface proteins	+	+

III IFNs. The corresponding IFNs probably act in a paracrine manner on adjacent cells to induce antiviral mechanisms. The vast number of viruses with different replication strategies notwithstanding, it appears that these agents can be countered by relatively few IFN-inducible "antiviral" proteins. One of the best characterized of these proteins is an enzyme known as 2-5A synthetase which, in the presence of dsRNA (often an intermediate of viral RNA synthesis), catalyzes the formation of an unusual oligonucleotide, ppp (A2′p)nA (2-5A), which in turn activates a latent endonuclease, RNase L. This endonuclease degrades viral (and cellular) mRNA and therefore inhibits viral protein synthesis. Small RNA viruses (*Picornaviridae*), such as Mengo virus or murine encephalomyocarditis virus (EMCV), whose replication is cytoplasmic are most affected by the induction of 2-5A synthetase. A second important antiviral protein is a dsRNA-activatable protein kinase, now designated PKR, which in the active form phosphorylates the peptide initiation factor, eIF2, involved in polyribosomal translation of mRNA. Phosphorylated eIF2 is inactive and thus protein synthesis is inhibited. This inhibition has been associated with the loss of replicating capacity of rhabdoviruses such as vesicular stomatitis virus (VSV). Both type I and type II IFNs have the potential to induce 2-5A synthetase and PKR, but induction occurs more slowly in type II IFN-treated cells. A third "antiviral" protein, Mx, is however primarily induced by type I IFNs; type III IFNs probably weakly induce Mx too. The Mx protein inhibits the replication of influenza viruses in rodent cells, specifically by inhibiting the nuclear stage of influenza viral nucleic acid synthesis.

Probably there are other antiviral proteins and mechanisms by which virus replication is inhibited. For instance, some viruses (e.g. herpesvirus and certain retroviruses) appear to be inhibited at the relatively late stage of virus particle maturation and budding. Table 2 summarizes the antiviral mechanisms applying to different virus families.

Tab. 2 Antiviral mechanisms for different virus groups.

Virus family	***Principal antiviral mechanism***
Small RNA viruses (picornaviruses), including rhinoviruses, poliovirus, encephalomyocarditis virus, Coxsackie viruses	2-5A synthetase system, inhibition of viral polypeptide synthesis by RNase L degradation of viral mRNA.
Rhabdoviruses, e.g. vesicular stomatitis virus (VSV)	dsRNA-activatable protein kinase, phosphorylation of initiation factor eIF-2 to inhibit viral polypeptide synthesis
Influenza viruses (orthomyxoviruses)	Mx protein, inhibition of early nuclear phase transcription of viral mRNAs (type I IFN only)
Retroviruses, e.g. Rous sarcoma virus, human immunodeficiency viruses (HIV-1, 2)	Acute infections: early inhibition of proviral DNA synthesis or integration. Chronic infections: late inhibition of virion (virus particle) assembly or maturation
Herpesviruses, e.g. herpes simplex viruses 1 and 2, varicella virus	Possibly early inhibition of viral protein synthesis by 2-5A synthetase system. Late stage inhibition of viral glycoprotein assembly

6.1.2 Defense Mechanisms of Viruses

In the course of evolution, many viruses have developed countermechanisms by which they can disrupt the antiviral mechanisms induced by IFNs. The IFN-inducible, dsRNA-dependent protein kinase PKR is one of the main targets, and different viruses overcome this by various means. For example, adenoviruses and Epstein–Barr virus (a member of the herpesvirus family) produce small RNA molecules, VAI-, and EBER-RNAs, respectively, which bind to PKR and block its activation by dsRNA. Reoviruses and vaccinia virus (a member of the poxvirus family) produce proteins, sigma 3 and SKI respectively, that bind to dsRNA and thus impair the activation of PKR. Interestingly, if IFN-treated VSV-infected cells are coinfected by vaccinia virus, VSV-replication is rescued, presumably by the inhibitory effect of SKI on PKR. Vaccinia also produces a nonfunctional protein analog of eIF2, which competes with eIF2 for phosphorylation by PKR and thus dilutes out the antiviral effect of activated PKR. Other viruses (Table 3) may "defeat" the kinase by yet other means.

The 2-5A synthetase-RNase L system can also be attacked. For example, encephalomyocarditis virus (EMCV), a picornavirus, can inactivate RNase L in several cell lines, but this inactivation is usually blocked by IFN treatment. Herpesviruses in contrast appear to impair RNase L activation by producing competing analogs of 2-5A.

Some viruses even have the capability to block the transcription of (normally) IFN-inducible genes. The E1A protein of adenoviruses, for example, which is produced early in viral replication, prevents activation of ISGF3 by IFNα (type I IFN). Hepatitis C virus uses a viral protease to block IRF-3 activation and IFN synthesis, while in hepatitis-B virus-infected cells, the so-called virus-specified "terminal protein" inhibits both type I and type II IFN-inducible gene expression.

Tab. 3 Viral defense mechanisms against IFN antiviral effects.

Virus family	*Defense mechanism*
Small RNA viruses, e.g. encephalomyocarditis virus	Inactivation of RNase L of 2-5A synthetase system
Herpesvirus	Viral analogs of 2-5A block 2-5A synthetase
Epstein–Barr virus (a herpesvirus)	Production of small RNA molecules, EBER-DNA, which block activity of dsRNA-dependent protein kinase (PKR)
Adenoviruses	Production of small RNA molecules, VAI-RNA, which block activity of dsRNA-dependent protein kinase (PKR). Early adenoviral E1a protein blocks activation of ISGF-3 by type I IFN
Reoviruses and poxviruses, in particular vaccinia virus	Production of viral proteins that bind dsRNA and therefore inhibit activation of PKR. Vaccinia also produces a nonfunctional analog of the initiation factor eIF-2, which competes for PKR-mediated phosphorylation with the cellular PKR
Influenza viruses	Activate preexisting inhibitor of PKR
Human immunodeficiency virus-1 (HIV-1)	HIV-1 TAT gene mediated inhibition of PKR synthesis. HIV-1 TAR RNA may also inhibit PKR activity

Finally, members of the poxvirus family, including vaccinia, encode IFN binding proteins, which are soluble and thus able to act as decoy receptors for either type I or type II IFNs. These viral decoy IFN receptors competitively bind IFNs and therefore reduce their antiviral activity.

6.2 Antiproliferative Activity

The antiviral actions wrought by IFNs are mediated by enzymes whose mechanisms have broad implications for cell growth and proliferation. Virus replication may be regarded as a form of pathological growth of a foreign, cell-like, entity at the cell's expense. In the presence of IFN, enzymes that generally curtail protein synthesis are activated, but because viral protein synthesis is normally rapid, the inhibitory effect on virus replication appears more dramatic than on the slower and more complex cellular growth. Possibly, the type I IFN system was evolved as part of a complex, interactive network of intercellular mediators that regulate cell growth and proliferation.

The dsRNA-dependent protein kinase, PKR, is constitutively produced in low amounts in a variety of cell types and is made in larger amounts following IFN induction. If a mutant form of this enzyme, which is unable to phosphorylate eIF2, is introduced into cells they undergo neoplastic transformation. This result suggests that normal PKR acts as a tumor-suppressor gene product whose activity is abrogated in the presence of dominant-negative (catalytically inactive) PKR mutants. The 2-5A synthetase-RNase L system could also have tumor-suppressor activity. For example, the levels of these two enzymes are high in growth-arrested cells. Introduction of 2-5A-like oligoadenylates into proliferating cells causes growth impairment, probably through activation of RNase L.

The foregoing observations indicate that enzymes such as PKR, 2-5A synthetase, and RNase L are involved in negative growth control mechanisms. That these enzymes are increased and activated in IFN-treated cells correlates well with IFN's antiproliferative activity. The increase in synthesis of these enzymes is however dependent on IFN-inducible transcription factors. Two of these transcription factors, IRF-1 and IRF-2, behave antagonistically; IRF-1 is a transcriptional activator while IRF-2 is a transcriptional inhibitor of the same set of IFN-inducible genes. IRF-1 has a short half-life and thus transcription of IFN-inducible genes is normally inhibited by the longer lasting IRF-2. After IFN induction, IRF-1 is transiently increased and leads to derepression of IFN-inducible genes. If IRF-2 is overexpressed, cells become transformed, probably because even low-level constitutive production of IFN-inducible gene products, such as PKR and 2-5A synthetase, is blocked. This transformation can be reversed by overexpressing IRF-1.

6.3 Antimicrobial Activity

A variety of microorganisms, including bacteria, mycoplasma, rickettsiae, chlamydia, protozoa, and fungi, have been cited as IFN inducers. Since, however, they lack IFN receptors, they are not directly susceptible to IFN actions. Nevertheless, IFNs tend to inhibit the growth of many of these microorganisms, particularly those that are intracellular parasites. It is unlikely, though, that the actions of the enzymes (e.g. 2-5A synthetase, PKR) that

inhibit viral replication will affect the replication of intracellular parasites, since their biosynthetic machinery is protected by their cell membrane. Other indirect mechanisms are therefore implicated. For example, type II IFN induces the enzyme indoleamine 2,3-dioxygenase (IDO), which catalyzes the conversion of tryptophan to N-formyl-kynurenine, resulting in tryptophan starvation, an activity that could also contribute to IFN-mediated antiproliferative effect. The growth of certain intracellular parasites (e.g. the protozoan *Toxoplasma gondii*) is inhibited by type II IFN, and tryptophan deprivation appears to be one of the inhibitory mechanisms involved. In phagocytic cells such as macrophages, which can be invaded by protozoans (e.g. *Leishmania*), IFNs, particularly type II IFN, inhibit parasite growth, probably by the induction of synthesis of reactive oxygen intermediates (ROI) and reactive nitrogen intermediates (RNI, e.g. $NO^{\bullet}$), which are toxic. The production of ROI following type II IFN induction may also contribute to the killing of intraerythrocytic malarial parasites.

There is rather less evidence to support a protective role of IFNs in bacterial and mycobacterial infections. Few bacteria appear to be growth inhibited by IFNs, but one exception is *Legionella pneumophilia*, whose replication inside macrophages is inhibited by type II IFN. Again, the production of ROI appears to be involved. It is noteworthy that patients with chronic granulomatous disease (CGD), who are inherently deficient in the respiratory burst oxidase system required for ROI generation, suffer from repeated severe bacterial infections of the deep tissues. At least some CGD patients benefit from type II IFN therapy.

6.4 Immunoregulatory Activity

Emerging evidence strongly suggests the involvement of IFNs not only in innate defense mechanisms but also in shaping adaptive immune responses to microbial attack. For example, type I IFN production by PDC following viral stimulation probably mediates T and NK cell maturation. However, the most crucial action is the increased expression of MHC antigens in leukocytes and other cell types triggered by IFNs. (The MHC antigens are cell surface glycoproteins essential for recognition of foreign, nonself, antigens). Both type I and type II IFNs induce expression of class I MHC antigens (i.e. HLA-A, B, C), but only type II IFN can induce *de novo* synthesis of class II MHC antigens (i.e. HLA-DR, DP, DQ). Class II MHC antigen expression is critically required for presentation of foreign antigen by macrophages and dendritic cells to T-helper (T_h) ($CD4^+$) lymphocytes leading to their activation. T_h-lymphocytes then secrete a range of cytokines that regulate B-lymphocyte proliferation and immunoglobulin (antibody) synthesis. Class I MHC antigen expression occurs more widely and is required for recognition of foreign antigen by cytotoxic T lymphocytes (CTL, $CD8^+$). For example, recognition of virus-infected cells by CTL depends on class I MHC antigen presentation of viral antigens at the cell membrane. Another category of leukocytes composed of large granular lymphocytes and known as NK cells are also activated by IFN, particularly type I IFN. It is not presently known how type I IFN augments the cytotoxic function of NK cells whose recognition of virus-infected cells or tumor cells appears to be independent of MHC antigen expression.

In the presence of antigen-specific antibodies, macrophages can effect cell-mediated cytotoxicity. Such antibody-dependent cell-mediated cytotoxicity (ADCC) can be enhanced by IFN through augmentation of immunoglobulin G-Fc receptor (IgG-FcR) expression.

Other cell surface proteins whose expression is increased by IFNs include cell adhesion molecules. Intercellular adhesion molecule 1 (ICAM-1) in particular is highly induced by type II IFN, but not by type I IFN, in vascular endothelial cells. The increased expression of ICAM-1 and other cell adhesion molecules leads to lymphocytes and monocytes bearing counterreceptors, for example, those expressing the integrin LFA-1, sticking to the endothelial cell lining of blood vessels from where they can migrate into injured or infected tissues.

Apart from the upregulation of cell surface molecules, IFNs also appear to be involved in the proliferation and functions of B lymphocytes, particularly immunoglobulin (antibody) synthesis. For example, type I IFNs inhibit antibody production when present at high concentrations but may stimulate antibody production at low concentrations. In comparison, type II IFN may act as a costimulant for B-cell proliferation as well as an immunoglobulin isotype switch in antibody-producing cells. Type II IFN (IFN-γ) enhances IgG_{2a} synthesis and inhibits IgG_1 and IgE production, an action that is opposed to that of interleukin-4 (IL-4). Type II IFN may also act by an unknown mechanism as a maturation factor for certain classes of cytotoxic T lymphocytes (CTL). IFNs, in general, are inhibitory for the proliferation of hematopoietic progenitor cells and therefore can depress hematopoiesis and the production of leukocytes.

7 Potential Physiological and Pathophysiological Roles of IFNs

7.1 Do IFNs have Physiological Roles?

All of the biological activities ascribed to IFNs were discovered through *in vitro* experimentation. Here it is possible to pick and choose conditions that favor particular outcomes (e.g. the antiviral response) by adjusting doses of IFN, times of incubation, levels of virus challenge, and so on. Experiments are frequently performed in cell systems where IFNs are the only exogenous, biologically active reagents present. The whole range of IFN-inducible activities is not demonstrable in a single-cell type; thus *in vitro* experimentation, often involving multiple cell "targets," including tumor cell lines, tends to illustrate, and perhaps exaggerate, the potential of IFNs as multifunctional regulators rather than define their "true" physiological roles. *In vivo*, IFNs together with other cytokines are often coinduced; therefore, it can be difficult to deduce whether IFN, or rather a particular type of IFN, has a major physiological role or one or more subordinate or accessory roles. Furthermore, although IFNs can be shown to have actions that regulate cell growth and differentiation *in vitro*, it is much harder to show unequivocally that IFNs are involved in physiological processes such as embryogenesis, organ and tissue development, and homeostatic mechanisms. That IFNs have the potential for inducing antiviral, antimicrobial, and antitumor activity suggests that their true physiological worth is as regulators of host defense mechanisms, to prevent pathophysiological events.

7.2 Ways of Studying IFN "Physiology"

Probably the best approach to determine how IFNs work *in vivo* is to investigate animals in which a deficiency of the IFN system exists or can be induced. For the IFN system, however, natural genetic deficiencies, either in the IFN genes or in the IFN-receptor genes, are probably very rare. Nevertheless, it is now possible to knockout the function of particular genes by specific site-directed disruption of the gene, for example, in mice. This approach involves targeting a nonfunctional gene by means of a vector into the mouse embryonic stem (ES) cell genome resulting in disruption of gene expression. Subsequently, ES cell lines are selected for the disrupted gene and injected into blastocysts. A program of breeding of the resulting chimeric animals and selection of offspring carrying the disrupted gene in the germline eventually leads, as long as this is not prenatally lethal, to the generation of viable null or knockout mice, which are homozygous for the disrupted gene.

To date, knockout mice with disrupted IFN, IFN-receptor, IRF, JAK, and other IFN-stimulated genes have been generated. For example, knockout mice with either a disrupted type II IFN gene or a disrupted type II IFN-receptor gene have been produced. Both kinds of mice developed normally and were healthy, but they exhibited a defect in natural resistance when challenged with pathogens (e.g. vaccinia virus or *Listeria monocytogenes* or mycobacteria). Macrophages in these mice showed impaired ROI and nitric oxide production and expression of class II MHC antigen expression; NK cell activity, but not CTL activity, was reduced while splenocytes proliferated in an uncontrolled manner in response to mitogen or alloantigen. Knockout mice deficient in the IFN receptor lacked an antigen-specific IgG_{2a} response. It thus has been clearly demonstrated that type II IFN (IFN-γ) is essential for immune responses to pathogens by several cell types.

Production of macrophage nitric oxide is crucially dependent on type II IFN-inducible nitric oxide synthetase (iNOS), whose gene transcription is regulated by IRF-1, a transcriptional enhancer. Thus, mice deficient in IRF-1 are also deficient in iNOS and susceptible to mycobacterial (e.g. *M. tuberculosis*) infections, which they are unable to clear. IRF-1 knockout mice also share other phenotypic similarities, such as reduced NK activity, with IFN-γR knockout mice.

Although IFN-γR knockout mice were susceptible to vaccinia virus, they were no more susceptible to other viruses (e.g. Semliki Forest virus (SFV), VSV) than their normal littermates. In contrast, knockout mice defective in type I IFN-receptor R1 gene, while displaying no overt phenotypic abnormalities, were unable to counter infection with several different viruses, including SFV and VSV. These findings indicate that type I and type II IFNs exert their antiviral effects through different, partially nonredundant pathways.

Certain mouse strains are more susceptible to infection by viruses or microorganisms, and in some cases this can be attributed to low type I IFN production. In addition, injection of mice with antitype I IFN antibody has been shown to increase their susceptibility to a range of virus infections. Children who suffer repeated rhinovirus (common cold) infections also have been shown to secrete lower than average levels of type I IFNs.

There are many molecularly different species of type I IFN (e.g. IFN-α_1,

$\alpha_2, \ldots \alpha_n$, IFN-ß, IFN-$\kappa$, IFN-$\omega$, IFN-$\tau$). Except for IFN-$\tau$, which has a special role as an antiluteolytic agent in ruminant pregnancy, there is no good evidence why such a large diversity of type I IFN molecules is necessary. Qualitatively, there are no differences in the *in vitro* biological activities of IFN-α, IFN-ß, and IFN-ω. Quantitative differences among these IFNs can be demonstrated, but these probably reflect their differing affinities for the type I IFN receptor. Possibly, differences in their hydrophobicity causes individual type I IFNs to locate more efficiently to particular tissues or anatomical sites but, as yet, there is little to support this hypothesis.

7.3 IFNs and Pathophysiological Phenomena

IFN genes are normally switched off and only expressed transiently during intervals of acute illness or injury. IFNs are thus induced to contribute to the host's specific and nonspecific defense mechanisms to combat infectious diseases and resolve tissue injuries. They are part of a network of mediators, including many other cytokines, responsible for regulatory immune responses and inflammatory reactions. However, if immunity and inflammation are excessive, pathophysiological phenomena often arise. IFNs are usually secreted locally, but can, if overexpressed, find their way into the circulation and produce symptoms such as fever, malaise, and headache, often associated with such acute viral infections as influenza. Such symptoms can be reproduced in patients receiving therapeutic doses of exogenous IFN-α or IFN-ß for malignancies and so it is clear that it is the IFN molecule itself that has the capacity to induce undesirable "side effects."

IFNs, and particularly type II IFN (IFN-γ), are proinflammatory. By increasing recognition of infected cells through augmentation of MHC antigen expression and enhancing cell adhesion molecules, leukocytes (neutrophils, lymphocytes, and macrophages) are recruited to sites of infected or injured tissues, are activated by antigen and cytokines to become cytotoxic or phagocytic, and synthesize a range of inflammatory and vasoactive mediators. In damaged skin, for instance, this results in local reddening and swelling before healing and resolution of the damaged area occurs. Infection or injury occurring a second time at the same site sometimes leads to a more severe inflammatory reaction as leukocytes are "primed" to infiltrate the site. This phenomenon is known as delayed hypersensitivity (DH). The severe complications (e.g. hemorrhagic shock) observed in children suffering from a second dengue virus infection (normally of a different serotype to the one that caused the primary infection) probably result from DH reactions. In a murine model, DH has been shown to be inhibited by neutralizing antibody against IFN-γ, thus strongly implicating IFN-γ as a proximal mediator of DH.

7.4 IFNs and Pathogenesis

On some occasions, the presence of IFNs correlates well with disease progression. Again, in a murine model, *Plasmodium berghei* induces cerebral malaria, often a severe complication of human infection with *Plasmodium falciparum*. Neutralizing antibodies to IFN-γ injected prior to infection with *P. berghei*, however, prevented the development of cerebral malaria. IFN-γ and other cytokines (e.g. TNF-α) are also partly responsible for the formation

of granulomas, nodular-like tissue masses characteristic of infections caused by some microorganisms, particularly mycobacteria (e.g. *Mycobacterium leprae*, the microbe that causes leprosy). Granulomas prevent spread of the mycobacteria, and since they contain large numbers of macrophages and neutrophils, mycobacteria can be killed by IFN-γ induction of ROI and nitric oxide synthesis. If, for example, ROI production fails to clear the mycobacteria and persists over an extended period, however, this cell-mediated response may lead to chronic inflammation and tissue damage. In such cases, chronic disease might result not from excessive IFN-γ levels, but from continuous, low-level production, which is insufficient to induce resolution of the disease. Other types of acute illnesses (e.g. viral hepatitis) may also become chronic due to a failure of the IFN system to clear the virus or microorganism. In both hepatitis-B and -C virus infections, viral mechanisms actively inhibit type I IFN production and action, and probably contribute to the establishment of chronic virus infection.

IFNs have also been associated with autoimmune diseases, but while levels of certain IFN types may be raised in such diseases, it is difficult to distinguish whether the IFNs are contributing to pathogenesis or whether they are by-products of underlying pathological processes. Autoimmune diseases usually result from a breakdown in immunological tolerance to self-antigens, but the causative agents are mostly unknown. It is, however, widely believed that viruses, bacteria, or mycoplasma, often infecting individuals years prior to the onset of disease, are responsible for lymphocyte sensitization to self-antigens. There are some genetic factors also (e.g. MHC antigen haplotypes) that increase susceptibility to particular autoimmune diseases. The inbred New Zealand Black (NZB) mice, for example, "spontaneously" develop autoantibodies and glomerulonephritis; this autoimmune disease progresses more rapidly if the mice are injected with type I or type II murine IFN and less rapidly if the mice are given neutralizing antitype II murine IFN antibody. Type I IFN-receptor deficiency in homozygous IFN-α/βR gene deleted NZB mice also resulted in reduced lupus-like disease. These findings suggest that IFNs have a pathogenic role in this autoimmune disease of NZB mice.

A small number of people suffer from systemic lupus erythematosus (SLE), an autoimmune disease characterized by high titers of autoantibodies of a range of specificities, including nuclear components, DNA, and nucleosomes leading to immune complex induced vasculitis and kidney failure, a feature in common with the NZB mice's autoimmune disease. Circulating IFN-α can be found in some SLE patients, with levels correlating to disease activity. Its role remains enigmatic, but there is increasing evidence of IFN-α being involved on both the initiation and maintenance of the central autoimmune processes of SLE. It is hypothesized that a specific class of dendritic cells is stimulated by viruses to secrete IFN-α in the first phase and by immune complexes in the second phase, thus creating a vicious cycle mechanism driving autoimmune processes. The unabated IFN-α production might contribute to autoimmune disease in a number of ways (e.g. by stimulating B-cell production of autoantibodies, by increasing MHC antigen expression and thus stimulating autoreactive T and B cells. In support of this hypothesis, an IFN-α "signature" of many IFN-induced genes has been demonstrated in white blood cells from patients with active SLE.

Tab. 4 IFNs and disease correlations.

Disease	*IFN type*	*Putative pathogenic mechanism(s)*
Systemic lupus erythematosus (SLE)	IFN-α	Stimulation of autoantibodies, autoreactive T and B cells
Insulin-dependent diabetes mellitus (IDDM)	IFN-α/IFN-γ	Upregulation of MHC class I and class II antigens in ß-islet cells, generation of autoreactive cytotoxic T lymphocytes
Behçets disease (eye lesions)	"Acid labile" – IFN-α/IFN-γ	Increased MHC class I and class II antigen expression. Increased autoantibody production
Grave's ophthalmopathy (immunological cytotoxicity against eye muscle cells)	IFN-γ	Increased MHC class II antigen expression. Increased ADCC of eye muscle cells
Multiple sclerosis	IFN-γ	Generation of autoreactive CTL in central nervous system
Sjögren's syndrome (chronic inflammatory disease of exocrine glands)	IFN-γ	Downregulation of type I IFN receptors in NK cells leading to loss of NK cell activity
Aplastic anemia (primary disorder of hematopoietic system)	IFN-α/IFN-γ	Inhibition of hematopoietic precursor proliferation

In tissue-restricted autoimmune diseases, such as insulin-dependent diabetes mellitus (IDDM), localized IFN secretion in a particular tissue or tissues may also be involved. For example, the ß-islet cells in the pancreas of IDDM patients hyperexpress class I and class II MHC antigens suggesting the presence of types I and II IFN, respectively. This has been confirmed by immunohistochemical staining of sections of pancreases. There are several other autoimmune or autoimmune-like diseases for which IFNs have been associated and these are summarized in Table 4.

8 Clinical Uses of IFNs

The potent antiviral activity of IFNs together with their potential antitumor actions provided the spur to the large-scale manufacture of IFNs for the purpose of clinical evaluation in a variety of viral and malignant diseases. In the early 1970s, large-scale production was limited by the quantities of human buffy coats (leukocytes) which could be used to make type I IFN. Later in that decade, human lymphoblastoid cells (e.g. Namalwa cell line), which could be grown to large culture volumes, became available for type I IFN production. By the 1980s, following the cloning of IFN-α subtype genes, some IFN species were mass-produced in *E. coli* or, in some instances, in mammalian cells, leading to abundant availability of certain subtypes (e.g. IFN-α_2). Production of IFN-ß, IFN-ω and IFN-γ by this means followed. Synthesis or construction of consensus sequence IFNA genes and various hybrid IFNA genes (e.g. A1/A2, A1/A8) has led to the production of nonnatural IFN-α molecules with enhanced activities or stability.

As a result of the anecdotal or often uncontrolled clinical studies of the 1970s, there was great enthusiasm in the 1980s, both from manufacturers of IFNs and clinicians, to test the therapeutic potential of interferons. It was only after many controlled or randomized studies had been conducted that it became obvious that IFN administered as a single agent was not beneficial for the treatment of the majority of malignant diseases, including the major cancers, lung, breast, colon. The initial optimism all but vanished and was replaced in the mid-1980s by a more sober and realistic appreciation of the potential therapeutic value of IFNs.

One major problem remains, that is: there is little real understanding of how IFNs were having their antitumor effects. Although IFNs had been shown to have antiproliferative activity in many tumor cell lines *in vitro*, it was also known that other tumor cell lines that were resistant to IFNs *in vitro* were nevertheless destroyed by IFN treatment when grown as transplantable tumors in mice. This indicated that IFNs probably acted on tumors *in vivo* by indirect mechanisms (e.g. augmentation of NK cell activity or enhancement of MHC antigen expression and antitumor immunity). In other words, their antitumor effects had more to do with modifying biological responses than direct antiproliferative actions. IFNs and cytokines in general are sometimes referred to as biological response modifiers (BRMs). However, it is still not clear what biological and pharmacological responses are necessary for IFN-induced tumor regression to occur.

In patients IFNs are administered in high doses, often close to the maximum tolerated dose (MTD): IFN, a ''biological drug'' is thus being used as if it were a chemical drug, a practice that sometimes leads to severe side effects (Table 5) and toxicity. Therefore, it is difficult to establish whether IFN is acting on tumors in the manner of a chemotherapeutic agent (i.e. direct action) or as a BRM (i.e. indirect action) or as both. It has become apparent nonetheless that IFNs administered at low doses are generally less effective in causing tumor regression. In addition, high-dose IFN therapy over a long period causes leukopenia (i.e. reduced numbers of leukocytes that respond to IFN and mediate indirect IFN antitumor effects), relative immunosuppression, and autoimmune phenomena. It is therefore difficult, against a background of generalized toxic effects, to ''harness'' whatever antitumor activity IFNs may have for particular tumor types.

Tab. 5 Side effects of type I IFN therapy.

Observed clinical responses (frequent)
Fever
Chills
Malaise/Fatigue
Myalgia
Headache
Anorexia/weight loss
Pharmacological toxicity (infrequent)
Liver and kidney toxicity
Inhibition of hematopoiesis, leukopenia
Central nervous system toxicity, e.g. confusion, altered mental states
Heart disturbances, e.g. arrhythmia, ischemia
Immunological responses
Depressed natural killer cell activity and ADCC on prolonged therapy schedules
Induction of autoantibodies

So far IFNs have only been partially successful in the treatment of cancers, and only for a very limited number of types of malignancy. Clinical trials have shown that the efficacy of individual IFN-α subtypes, in particular IFN-α_2, is more or

less the same as heterogeneous, multicomponent, type I IFN preparations such as lymphoblastoid and leukocyte IFN, both of which contain mixtures of IFN-α subtypes. IFN-ß is probably as efficacious, but it has not been used as widely as IFN-α preparations in antitumor therapies. IFN-γ, despite its apparent greater immunoregulatory activity, has been found to be rather less potent than IFN-α. All IFNs in high doses cause significant, but readily reversible, side effects (Table 5). The most responsive cancer to IFN-α therapy is a very rare form of leukemia known as hairy cell leukemia (HCL), where up to an 80% response rate has been reported. In HCL patients, the "hairy cells" invade the spleen and bone marrow. It has been convincingly shown that IFN-α therapy continued over several months leads to clearance of "hairy cells" and in some patients, a long-term remission is achieved. Some patients develop neutralizing antibodies against IFN, particularly IFN-α_2, which may cause resistance to further treatment. In these cases, beneficial responses maybe rescued by switching to a multicomponent IFN-α preparation. On the whole, IFN-α therapy of HCL appears at least as effective and durable as chemotherapy with the drug pentostatin (2-deoxycoformycin).

IFN-α therapy has also been shown to slow down the progression of CML. In this cancer, leukemic cells grow slowly in the initial chronic phase and persist for 2 to 4 years, but there follows a dramatic "blast crisis," producing rapidly proliferating myeloid leukemia cells and a fatal outcome. CML patients can be treated with IFN-α in the chronic phase, and some achieve complete hematological remission (i.e. disappearance of leukemic cells, and often increased survival). Recently, IFN-α therapy of CML has been replaced by the administration of tyrosine kinase inhibitors such as Gleevec®, which block c-abl activity in leukemic cells, and thus their proliferation.

Other malignancies in which IFN-α therapy seems to work, although with a lower percentage of patients responding than in HCL and CML, include low-grade non-Hodgkin's lymphoma, cutaneous T-cell lymphoma, AIDS-related Kaposi's sarcoma, carcinoid tumors, renal cell carcinoma, squamous epithelial tumors of the head and neck, multiple myeloma, and malignant melanoma (Table 6). In most of these cancers, complete responses are low compared to partial responses, but IFN-α may help with maintenance therapy of disease in some cases. In preclinical systems, the combination of IFN therapy and conventional chemotherapy has appeared to offer greater chances of producing effective treatment of many cancers, but in clinical trials this strategy has produced mostly disappointing results, probably because optimal combinations of IFNs with cytotoxic drugs cannot be deduced owing to a lack of understanding of the physiological interaction of these agents.

As mentioned earlier, IFN therapy induces leukopenia and generalized immunosuppression. This effect may be useful for the treatment of diseases in which there is uncontrolled leukocytosis (e.g. thrombocytosis: markedly elevated platelet numbers) associated with various myeloproliferative diseases. Parenteral (subcutaneous or intramuscular) injection of IFN-ß appears to benefit multiple sclerosis (MS) patients in the early stages of this disease by leading to a reduced rate of exacerbations. This, however, could be a result of the relative immunosuppression caused by IFN-ß (e.g. by inhibiting IFN-γ actions and thereby

Tab. 6 Responses seen in human malignancies following IFNα therapy.

		Approximate response rate[a] [%]
Hematological malignancies:	Hairy cell leukemia	80
	Chronic myelogenous leukemia (preblast crisis)	79
	Cutaneous T-cell lymphoma	52
	Non-Hodgkin's lymphoma	46
	Multiple myeloma	20
Solid tumors:	Squamous tumors of head and neck	91%
	Carcinoid tumors	55
	Bladder cancer	65
	Cervical cancer	43
	Kaposi's sarcoma (AIDS-related)	30
	Ovarian cancer	19
	Renal cell carcinoma	17

Notes: Intermediate and high-grade lymphomas, Hodgkin's disease, and chronic lymphocytic leukemia, all less than 20%. Malignant melanoma, breast cancer, colorectal carcinoma, lung cancer, osteogenic sarcoma, all around 10% or less.
[a]Response rate includes complete, partial, and minor responses.

suppressing the growth and activity of autoreactive T lymphocytes in the central nervous system.

Type II IFN, IFN-γ, has been rather less successful in the treatment of cancers, but uniquely it has been found to benefit some chronic granulomatous disease (CGD) patients, probably by boosting and thus correcting the impaired ROI producing capacity of macrophages and neutrophils. IFN-γ may also be useful in the adjunctive treatment of leishmaniasis and mycobacterial diseases.

In the treatment of viral diseases, IFNs, despite having proven antiviral activity *in vitro*, have not proved to be the panacea for most common viral infections of man. IFNs prevent the replication of rhinoviruses in cell cultures when administered to volunteers in the form of a nasal spray, but IFNs cannot "cure" colds once established. IFNs are only partially effective in preventing influenza infection. Topical applications of IFN in the form of creams or ointments to herpesvirus lesions (e.g. in herpes zoster) and genital warts caused by papilloma viruses have been tried but have given limited beneficial effects. It has been found that parenteral administration (i.e. intramuscular or intravenous injection) has greater beneficial effects of IFN-α on virus-caused lesions and warts, although not to the extent that IFN therapy has become the treatment of choice. Another wartlike disease, juvenile laryngeal papilloma (JLP), which can severely obstruct the airways of young children, also has been found to respond beneficially to IFN-α therapy. Nevertheless, it appears neither curative nor of substantial value as an adjunctive agent in the long-term management of JLP.

Probably, the most successful application of IFN-α therapy in virus-mediated

diseases is in the treatment of chronic active hepatitis, caused by either hepatitis-B or -C viruses. A certain proportion of chronic active hepatitis-B patients respond positively to IFN-α treatment. For instance, viral infectivity markers disappear and seroconversion and cure follows. It is interesting that in the case of hepatitis-B virus, viral activity is probably responsible for inhibiting the endogenous IFN system, and therefore the administration of exogenous IFN-α constitutes a replacement therapy. In hepatitis C virus infection, the most common cause of transfusion-linked hepatitis, IFN-α is the treatment of choice. Recent improvements in combination therapies using "pegylated" IFN-α_2 (the pegylation markedly increases the half-life of IFN in the circulation) and the antiviral drug ribovarin have resulted in cure rates approaching 70%. However, some strains of the virus, in particular genotype 1, which is the most prevalent worldwide, are strongly resistant to IFN therapy. This resistance has been recently demonstrated to be associated with the capacity of the virus to block IRF-3 activation, which is essential for antiviral activity.

Type I IFN, both IFN-α and IFN-ß, has been shown to inhibit HIV-1 replication *in vitro*. However, *in vivo* there is little evidence showing IFN-α has any long-term beneficial effect in HIV-1 positive individuals. IFN-α can cause tumor regression in some AIDS patients with Kaposi's sarcoma, but this possibly has more to do with IFNs antitumor or antiangiogenic activity than with its antiviral one. Combination therapies for HIV-1 infections involving IFN-α, antiviral drugs such as 3′-azido-2′,3′-dideoxythymidine (AZT) and cytokines, such as granulocyte-macrophage colony-stimulating factor (GM-CSF), appeared promising, but have not been followed up.

9
Final Remarks

Since their discovery over four decades ago, IFNs have ceased to be poorly characterized antiviral factors; both at the molecular and cell biological levels; they are now among the best defined and better understood biologically active intercellular mediators. However, despite the large body of information concerning IFNs and their actions, there are still many issues to be resolved. Why, for example, is the type I IFN family so heterogeneous? There is a need to find more ways to translate the potent antiviral, antimicrobial, and antitumor activities of IFNs into the means for combating human disease. Future progress will depend on further understanding not only of IFN-mediated actions at the molecular and cellular level but also of the numerous ways IFNs can interact with other biologically active mediators, including cytokines, hormones, and chemical drugs.

See also Cytokines: Interleukins; Medicinal Chemistry; Molecular Mediators: Cytokines; Nucleic Acid Packaging of DNA Viruses; Signaling Through JAKs and STATs: Interferons Lead the Way; Transcription Factor NF-κB: Function, Structure, Regulation, Pathways, and Applications; Virus-Cell Interactions.

Bibliography

Books and Reviews

Barnes, B., Lubyova, B., Pitha, P.M. (2002) On the role of IRF in host defense, *J. Interferon Cytokine Res.* **22**, 59–71.

Baron, S., Coppenhaven, D.H., Dianzani, F., et al. (Eds.) (1992) *Interferon: Principles and Medical Applications*. International Society for Interferon Research, University of Texas Medical Branch, Galveston, TX.

Billiau, A. (1996) Interferon-γ: biology and role in pathogenesis, *Adv. Immunol.* **62**, 61–90.

Brierley, M.M., Fish, E.N. (2002) IFN-α/β receptor interactions to biologic outcomes: understanding the circuitry, *J. Interferon Cytokine Res.* **22**, 835–845.

Byrne, G.I., Turco, J. (Eds.) (1988) *Interferon and Non-viral Pathogens*, Dekker, New York.

Corssmit, E.P.M., De Metz, J., Sauerwein, H.P., Romijn, J.A. (2000) Biologic responses to IFN-α administration in humans, *J. Interferon Cytokine Res.* **20**, 1039–1047.

De Maeyer, E., De Maeyer-Guignard, J. (1988) *Interferons and other Regulatory Cytokines*, Wiley, New York.

De Maeyer, E., DeMaeyer-Guignard, J. (1998) Interferon gamma, in: Mire-Sluis, A., Thorpe, R. (Eds.) *Cytokines*, Academic Press, San Diego, CA, p. 391.

Gutterman, J.U. (1994) Cytokine therapeutics: lessons from interferon α, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1198.

Ihle, J., Kerr, I. (1995) JAKs and STATs in signaling by the cytokine receptor superfamily, *Trends Genet.* **11**, 69–74.

Langer, J.A., Cutrone, E.C., Kotenko, S. (2004) The class II cytokine receptor (CRF2) family: overview and patterns of receptor-ligand interactions, *Cytokine Growth Factor Rev.* **15**, 33–48.

Meager, A. (1995) Interferons in clinical use: an appraisal, *J. Biotechnol. Health Care* **2**, 281–317.

Meager, A. (1998) *The Molecular Biology of Cytokines*, John Wiley, Chichester, UK.

Meager, A. (2002) Biological assays for interferons, *J. Immunol. Methods* **261**, 21–36.

Mogensen, K.E., Lewernz, M., Reboul, J., et al. (1999) The type I interferon receptor: structure, function, and evaluation of a family business, *J. Interferon Cytokine Res.* **19**, 1069–1098.

Pestka, S., Krause, D.K., Walter, M.R. (2004) Interferons, interferon-like cytokines, and their receptors, *Immunol. Rev.* **202**, 8–32.

Roberts, R.M., Cross, J.C., Leaman, D.W. (1992) Interferons as hormones of pregnancy, *Endocr. Rev.* **13**, 432–452.

Ronnblom, L., Alm, G.V. (2002) The natural interferon-α producing cells in systemic lupus erythematosus, *Hum. Immunol.* **63**, 1181–1193.

Sen, G.C., Lengyel, P. (1992) The interferon system: a bird's eye view of its biochemistry, *J. Biol. Chem.* **267**, 5017–5020.

Stark, G.R., Kerr, I.M., Williams, B.R., et al. (1998) How cells respond to interferons, *Annu. Rev. Biochem.* **67**, 227.

Stewart, W.E. II, (1979) *The Interferon System*, Springer-Verlag, Vienna.

Wadler, S., Schwartz, E.L. (1990) Antineoplastic activity of the combination of interferon and cytotoxic agents against experimental and human malignancies: a review, *Cancer Res.* **50**, 3473–3486.

Primary Literature

Aboud, M., Hassan, Y. (1983) Accumulation and breakdown of RNA-deficient intracellular virus particles in interferon-treated NIH 3T3 cells chronically producing Moloney murine leukaemia virus, *J. Virol.* **45**, 489–495.

Adolf, G.R. (1990) Monoclonal antibodies and enzyme immunoassays specific for human interferon (IFN) ω1: evidence that IFN-ω1 is a component of human leukocyte IFN, *Virology* **175**, 410–417.

Aguet, M., Dembic, Z., Merlin, G. (1988) Molecular cloning and expression of the human interferon-γ receptor, *Cell* **55**, 273–280.

Allen, G., Diaz, M. (1996) Nomenclature of the human interferon proteins, *J. Interferon Cytokine Res.* **16**, 181–185.

Barber, G.N., Jagus, R., Meurs, E.F., et al. (1995) Molecular mechanisms responsible for malignant transformation by regulatory and catalytic domain variants of the interferon-induced enzyme RNA-dependent protein kinase, *J. Biol. Chem.* **270**, 17423–17428.

Borden, E.C., Hogan, T.F., Voelkel, J.G. (1982) Comparative antiproliferative activity in vitro

of natural interferons α and β for diploid and transformed human cells, *Cancer Res.* **42**, 4948–4953.

Bouillon, M., Audette, M. (1993) Transduction of retinoic acid and γ-interferon signal for intercellular adhesion molecule-1 expression on human tumour cell lines: evidence for the late-acting involvement of protein kinase C inactivation, *Cancer Res.* **53**, 826–832.

Capon, D.J., Shepard, H.M., Goeddel, D.V. (1985) Two distinct families of human and bovine interferon-α genes are co-ordinately expressed and encode functional polypeptides, *Mol. Cell. Biol.* **5**, 768–779.

Coccia, E.M., Severa, M., Giacomini, D., et al. (2004) Viral infection and Toll-like receptor agonists induce a differential expression of type I and λ interferons in human plasmacytoid and moncyte-derived dendritic cells, *Eur. J. Immunol.* **34**, 796–805.

Da Silva, A.J., Brickelmaier, M., Majeau, G.R., et al. (2002) Comparison of the gene expression patterns induced by treatment of human umbilical vein endothelial cells with IFN-α2b vs. IFN-β1a: understanding the functional relationship between distinct type I interferons that act through a common receptor, *J. Interferon Cytokine Res.* **22**, 173–188.

Deblandre, G.A., Marinx, O.P., Evans, S.S., et al. (1995) Expression cloning of an interferon-inducible 17-kDa membrane protein implicated in the control of cell growth, *J. Biol. Chem.* **270**, 23860–23866.

Diaz, M.O., Ziemin, S., Le Beau, M.M., et al. (1988) Homozygous deletion of the α- and β-genes in human leukaemia and derived cell lines, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5259–5263.

Dumoutier, L., Tounsi, A., Michaels, T., et al. (2004) Role of the interleukin-28 receptor tyrosine residues for the antiproliferative activity of IL-29/IFN-λ1: similarities with type I interferon signaling, *J. Biol. Chem.* **279**, 32269–32274.

Ealick, S.E., Cook, W.J., Vijay-Kumar, S., et al. (1991) Three-dimensional structure of recombinant interferon-γ, *Science* **252**, 698–702.

Fellous, M., Kamoun, M., Gresser, I., Bono, R. (1979) Enhanced expression of HLA antigens and β2-microglobulin on interferon-treated human lymphoid cells, *Eur. J. Immunol.* **9**, 446–449.

Gray, P.W., Leung, D.W., Pennica, D., et al. (1982) Expression of human immune interferon cDNA in *E. coli* and monkey cells, *Nature (Lond.)* **295**, 503–508.

Gupta, S.L., Holmes, S.L., Mehra, L.L. (1982) Interferon action against reovirus: activation of interferon-induced protein kinase in mouse L929 cells upon reovirus infection, *Virology* **120**, 495–499.

Hauptmann, R., Swetly, P. (1985) A novel class of human type I interferons, *Nucleic Acids Res.* **13**, 4739–6219.

Heron, I., Hokland, M., Berg, K. (1978) Enhanced expression of β2-microglobulin and HLA antigens on lymphoid cells by type I interferons, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 6215–6219.

Hokland, M., Berg, K. (1981) Interferon enhances the antibody-dependent cellular cytotoxicity (ADCC) of human polymorphonuclear leukocytes, *J. Immunol.* **127**, 1585–1588.

Horisberger, M.A. (1992) Interferon-induced human protein MxA is a GTPase which binds transiently to cellular proteins, *J. Virol.* **66**, 4705–4709.

Hovanessian, A.G., Wood, J.N. (1980) Anticellular and antiviral effects of ppp(2′5′A)$_n$, *Virology* **101**, 81–90.

Karpusus, M., Nolte, M., Benton, C.B., et al. (1997) The crystal structure of human interferon β at 2.2-A resolution, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11813–11818.

Kotenko, S.V., Gallagher, G., Baurin, V.V., et al. (2003) IFN-λs mediate antiviral protection through a distinct class II cytokine receptor complex, *Nat. Immunol.* **4**, 69–77.

Kroger, A., Koster, M., Schroeder, K., et al. (2002) Activities of IRF-1, *J. Interferon Cytokine Res.* **22**, 5–14.

Kumar, R., Choubey, D., Lengyel, P., Sen, G.C. (1988) Studies on the role of the 2′-5′-oligoadenylate synthetase-RNase L pathway in β-interferon-mediated inhibition of encephalomyocarditis virus replication, *J. Virol.* **62**, 3175–3181.

LaFleur, D.W., Nardelli, B., Tsareva, T., et al. (2001) Interferon-kappa, a novel type I interferon expressed in human keratinocytes, *J. Biol. Chem.* **276**, 39765.

McNurlan, M., Clemens, M. (1986) Inhibition of cell proliferation by interferons: relative contributions of changes in protein synthesis and breakdown to growth control of

human lymphoblastoid cells, *Biochem. J.* **237**, 871–876.

Meurs, E.F., Galabru, J., Barber, G.N., et al. (1993) Tumor suppressor function of the interferon-induced double-stranded RNA-activated protein kinase, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 232–236.

Müller, U., Steinhoff, U., Reis, L., et al. (1994) Functional role of type I and type II interferons in antiviral defense, *Science* **264**, 1918–1921.

Nagata, S., Mantei, N., Weissmann, C. (1980) The structure of one of the eight or more distinct chromosomal genes for human interferon-α, *Nature (Lond.)* **287**, 401–408.

Pellegrini, S., John, J., Shearer, M., et al. (1989) Use of a selectable marker regulated by alpha interferon to obtain mutations in the signalling pathway, *Mol. Cell. Biol.* **9**, 4605–4612.

Pestka, S., Meager, A. (1995) Interferon standardization and designations, *J. Interferon Cytokine Res.* **17**(Suppl. 1), S9.

Qin, X.-Q., Runkel, L., Deck, C., et al. (1997) Interferon-β induces S phase accumulation selectively in human transformed cells, *J. Interferon Cytokine Res.* **17**, 355–367.

Radhakrishnan, R., Walter, L.J., Hruza, A., et al. (1996) Zinc mediated dimer of human interferon-α_{2b} revealed by X-ray crystallography, *Structure* **4**, 1453–1463.

Resnitzky, D., Tiefenbrun, N., Berissi, H., Kimchi, A. (1992) Interferons and interleukin-6 suppress phosphorylation of the retinoblastoma protein in growth-sensitive hematopoietic cells, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 402–406.

Rice, A.P., Duncan, R., Hershey, J.W.B., Kerr, I.M. (1985) Double-stranded RNA-dependent protein kinase and 2-5A system are both activated in interferon-treated, encephalomyocarditis virus-infected HeLa cells, *J. Virol.* **54**, 894–898.

Rigby, W.F., Ball, E.D., Guyre, P.M., Fanger, M.W. (1985) The effect of recombinant-DNA-derived interferons on the growth of myeloid progenitor cells, *Blood* **65**, 858–861.

Ronni, T., Melén, K., Malygin, A., Julkunen, J. (1993) Control of IFN-inducible MxA gene expression in human cells, *J. Immunol.* **141**, 1715–1726.

Sheppard, P., Kindsvogel, W., Xu, W., et al. (2003) IL-28, IL-29 and their class II cytokine receptor IL-28R, *Nat. Immunol.* **4**, 63–68.

Stojdl, D.F., Lichty, B., Knowles, S., et al. (2000) Exploiting tumor-specific defects in the interferon pathway with a previously known oncolytic virus, *Nat. Med.* **6**, 821–825.

Tan, S.-L., Katze, M.G. (1999) The emerging role of the interferon-induced PKR protein kinase as an apoptotic effector: a new face of death? *J. Interferon Cytokine Res.* **19**, 543–554.

Taniguchi, T., Mantei, N., Schwarzstein, M., et al. (1980) Human leukocyte and fibroblast interferons are structurally related, *Nature (Lond.)* **285**, 547–549.

Wang, E., Pfeffer, L.M., Tamm, I. (1981) Interferon α induces a protein kinase C-ε (PKC-ε) gene expression and a 4.7 kb PKC-ε related transcript, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6281–6285.

Wong, L.H., Kramer, K.G., Hatzinisiriou, I., et al. (1997) Interferon-resistant human melanoma cells are deficient in ISGF3 components, STAT1, STAT2, and p48-ISGF3 gamma, *J. Biol. Chem.* **272**, 28779–28784.

Zhang, K., Kumar, R. (1994) Interferon-α inhibits cyclin E- and cyclin D1-dependent CDK-2 kinase activity associated with RB and E2F in Daudi cells, *Biochem. Biophys. Res. Comm.* **200**, 522–528.

Zhou, A., Hassel, B.A., Silverman, R.H. (1993) Expression cloning of 2-5A-dependent RNAase: a uniquely regulated mediator of interferon action, *Cell* **72**, 753–765.

Virus-Cell Interactions

Andrew J. Mouland[1], *Nikolaus Heveker*[2], *and Anne Gatignol*[1]
[1] *Lady Davis Institute for Medical Research and McGill University, Montréal, Canada*
[2] *Centre de Recherche, Hôpital Sainte-Justine and Université de Montréal, Montréal, Canada*

1 Virus Entry 430
1.1 Virus Attachment to Target Cells 430
1.2 Viral Receptors 431
1.3 Virus Fusion 433

2 Innate Cellular Response to Virus Infection 434
2.1 Immediate Cell Response to Virus Infection 434
2.1.1 Cytokine and Chemokine Induction upon Virus Infection 434
2.1.2 Interferon α/β Production 434
2.1.3 IFN-γ Production 435
2.2 The Interferon Pathway 435
2.3 Activation of Host Genes in Response to Interferon 436
2.3.1 $2'-5'$ A Synthetase and RNAse L 436
2.3.2 Mx GTPase 437
2.3.3 RNA-Specific Adenosine Deaminase, ADAR1 437
2.3.4 Protein Kinase PKR 437
2.3.5 ISG56 437
2.3.6 Major Histocompatibility Complex Proteins 437
2.3.7 Other Genes Induced by Virus Infection 438
2.4 Viral Antagonists to the IFN Response 438
2.4.1 Inhibitors of IFN Synthesis 438
2.4.2 Inhibitors of IFN Signaling 439
2.4.3 Inhibitors of ISGs 440

3 Early Events in Virus Replication 442
3.1 Uncoating 442

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

3.2 Decapsidation and the TRIMs 442
3.3 Reverse Transcription and the APOBECs 444
3.4 Intracellular Replication and Trafficking 445
3.5 Trafficking to the Nucleus 446
3.6 Cellular Proteins and Nuclear Import 446
3.7 Cellular Proteins and Integration 447

4 Viral Gene Expression 448
4.1 Synthesis of Viral mRNA 448
4.1.1 Transcription of DNA Virus Genomes 448
4.1.2 Transcription of Viruses Integrated into the Host Chromosome 450
4.1.3 Synthesis of Coding mRNA by RNA Viruses 451
4.1.4 Virus-induced Cell Transcription Shut-off 452
4.2 Nuclear Export 452
4.2.1 Export of Viral RNA 452
4.2.2 Virus-induced Cellular Nuclear Export Shut-off 454
4.3 Translation 455
4.3.1 Initiation of Translation of Viral mRNA 455
4.3.2 Translational Regulation by Cellular Factors 455
4.3.3 Virus-induced Cell Translation Shut-off 456

5 Virus–Host Interactions Involved in Viral Egress 457
5.1 Ribonucleoprotein Transport 457
5.2 Viral Protein Egress Mediated Through Vesicular Trafficking Mechanisms 459
5.3 Virion Assembly: Morphogenesis and Budding 459
5.4 Gag–Host Interactions: The Gag Interactome 460
5.5 Cellular Proteins Incorporated in the Virion 461

6 Cell-to-Cell Viral Transmission 462
6.1 The Virological Synapse 463
6.2 Dendritic Cell to T-cell Transmission 464
6.3 Nanotubes and Viral Transmission 464

7 Pathogenesis Induced by Virus–Cell Interactions 465
7.1 Virus–Cell Interactions Leading to Oncogenicity 465
7.1.1 Cellular Genes Targeted by Viral Oncoproteins 465
7.1.2 Viral Oncoproteins 465
7.2 Viral Proteins Contributing to Cell Death 467
7.2.1 Anti-apoptotic Viral Genes 467
7.2.2 Proapoptotic Viral Genes 468
7.3 Viral Proteins that Affect Host Immunity 468
7.3.1 Virus-encoded Cytokine and Cytokine Receptor Homologs 468
7.3.2 Inhibitors of MHC/HLA Molecules 469
7.3.3 HIV-1 Example: Downregulation of CD4 by Several Mechanisms 469

7.4 Viral-induced Neurotoxicity 470
7.4.1 Viruses Inducing Neurotoxicity 470
7.4.2 Example of HIV-associated Dementia 470

8 Perspectives on Therapies Based on Virus–Cell Interactions 471
8.1 Virus Entry as an Antiviral Target 471
8.2 Antiviral Action of Interferons 471
8.3 Inhibition of Viral Gene Expression 472
8.4 Inhibition of Assembly 473
8.5 Viruses as Therapeutic Agents 474

9 Conclusion 474

Acknowledgments 474

Bibliography 475
Books and Reviews 475
Primary Literature 478

Keywords

Attachment Factor
Cell surface molecule that permits virus binding without triggering conformational rearrangements in the viral envelope, and without permitting viral entry.

Innate Immunity
Immediate cellular response without previous exposure of the host cell to an infectious agent.

Interferons
Large family of multifunctional secreted proteins involved in antiviral defense, cell growth regulation, and immune activation.

Intracellular Trafficking
Translocation of cellular and viral macromolecules within a cell. A phenomenon that occurs at both early and late stages of viral replication.

Receptor
Cellular surface molecule that binds virus and permits viral entry.

Ribonucleoprotein
A complex that is composed of both RNA and protein.

Viral Restriction
A block to virus replication due to the presence of constitutively-expressed genes, or to the absence of a required component in the infected cell.

Abbreviations

AAV:	adeno-associated virus
ACE2:	angiotensin converting enzyme 2
ADAR:	adenosine deaminase
Ag:	antigen
AIDS:	acquired immunodeficiency syndrome
ALV:	avian leukosis virus
APOBEC:	apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like
ASLV:	avian sarcoma and leukosis virus
ASV:	avian sarcoma virus
ATF:	activating transcription factor
BAF:	barrier-to-autointegration factor
BDV:	Borna disease virus
CAT:	cationic amino acid transporter
CBP:	cAMP Binding Protein
ccc:	covalently closed circle
CD4:	cluster differentiation 4
CDK:	cyclin dependent kinase
CIITA:	MHC class II transcription factor
CK2:	calcium kinase 2
CMV:	cytomegalovirus
c-onc:	cellular proto-oncogenes
c-PKA:	cAMP-dependent protein kinase
CPSF:	cleavage and polyadenylation specificity factor
CREB:	cyclic-AMP response element binding
CRM-1:	chromosome maintenance region 1
CTD:	carboxy terminal domain
CTE:	constitutive transport RNA elements
CTL:	cytotoxic T lymphocytes
DC:	dendritic cells
DC-SIGN:	dendritic cell-specific ICAM-3 grabbing nonintegrin
dsRNA:	double-stranded RNA
EBV:	Epstein–Barr virus
eEF1α:	eukaryotic elongation factor-1 alpha
EGF:	epidermal growth factor
EIAV:	equine infectious anaemia virus

eIF:	eukaryotic initiation factor
EMCV:	encephalomyocarditis virus
eRF1:	eukaryotic release factor 1
ESCRT:	endosomal sorting complex required for transport
FADD:	Fas associated death domain
FIV:	feline immunodeficiency virus
FMDV:	foot-and-mouth disease virus
FV:	foamy virus
Fv-1:	Friend virus susceptibility factor-1
Gag:	group specific antigen
gD:	glycoprotein D
GLUT:	glucose transporter
GM-CSF:	granulocyte-macrophage colony-stimulating factor
HAT:	histone acetyl transferase
HBV:	hepatitis B virus
HCMV:	human cytomegalovirus
HCV:	hepatitis C virus
HDAC:	histone deacetylases
HDM2:	human double minute 2 protein
HDV:	hepatitis delta virus
HHV8/KSHV:	human herpes virus 8/Kaposi's sarcoma-associated herpesvirus
HIV:	human immunodeficiency virus
HIVD:	HIV-associated dementia
HLA:	human leukocyte antigen
HMG:	high-mobility group protein
HNF:	hepatocyte nuclear factors
hnRNP:	heterogeneous nuclear ribonucleoprotein
hnRNP A2:	heterogeneous nuclear ribonucleoprotein A2
HPV:	human papillomavirus
HSPG:	heparan sulfate proteoglycans
HSV:	herpes simplex virus
HTLV-I:	human T-cell leukemia virus
IFN:	interferon
IFNAR:	IFN-α Receptor
IFNGR:	IFN-γ Receptor
IKK:	IκB kinase
IL:	interleukin
IN:	integrase
INI1 or hSNF5:	integrase interactor 1
iNOS:	inducible nitric oxide synthase
IRES:	internal ribosome entry site
IRF:	IFN regulatory factor
IRS-1:	insulin receptor substrate 1
IS:	immunological synapse

ISG:	IFN-stimulated gene
ISGF3:	IFN-stimulated gene factor 3
ISRE:	IFN-stimulated response elements
IκB:	inhibitory factor κB
Jak:	Janus kinase
kDa:	kiloDaltons
LEDGF/p75:	lens epithelium-derived growth factor
LT:	large T antigen
LTR:	long terminal repeat
MA:	matrix
MAP:	mitogen-activated protein
MCP-1:	monocyte chemoattractant protein 1
MCV:	molluscum contagiosum virus
MHC:	major histocompatibility complex
MIP-1:	macrophage inflammatory protein 1
MKKK:	MAP 3 kinase
ML:	major late
MLV:	murine leukemia virus
MPSV:	Mason-Pfizer simian virus
MTOC:	microtubule organizing center
MVB:	multivesicular body
NC:	nucleocapsid
NES:	nuclear export signal
NF-κB:	nuclear factor-κB
NK:	natural killer
NLS:	nuclear localization signal
NP:	nucleoprotein
OAS:	2′–5′ oligoadenylate synthetases
P/CAF:	P300/CBP-associated factor
PA:	polymerase subunit A
PABP:	polyA binding protein
PACT:	PKR activator
PDI:	protein disulfide isomerase
PIC:	preintegration complex
PKC:	protein kinase C
PKR:	dsRNA-activated protein kinase
PML:	promyelocytic leukemia protein
pRb:	retinoblastoma susceptibility protein
PyV:	polyomavirus
RANTES:	regulated upon activation, normal T-cell expressed and secreted
RdRp:	RNA-dependent RNA polymerase
Ref1:	restriction factor 1
RISC:	RNA-induced silencing complex
RLI:	RNAse L Inhibitor

RNAi:	RNA interference
RNAP:	RNA polymerase
RNAse L:	endoribonuclease L
RNP:	ribonucleoprotein
RRE:	Rev responsive element
RsSV:	Rous sarcoma virus
RSV:	respiratory syncytial virus
RT:	reverse transcriptase
RVFV:	Rift Valley fever virus
RxRE:	Rex responsive element
Sam68:	Src-associated during mitosis 68 kDa protein
SARS-CoV:	coronavirus that causes severe acute respiratory syndrome
SFV:	Semliki Forest virus
siRNA:	small-interfering RNA
SIV:	simian immunodeficiency virus
SOCS3:	suppressor of cytokine signaling-3
ST:	small T antigen
Stat:	signal transducer and activator of transcription
SU:	envelope surface unit
SV40:	simian virus 40
TAP:	transporter associated with antigen processing
TAR:	trans-activation response element
TBEV:	tick-borne encephalitis virus
TBP:	TATA–binding protein
TGFβ:	transforming growth factor β
TLR:	toll-like receptor
TM:	envelope transmembrane unit
TNFR:	tumor necrosis factor receptor
TNFα:	tumor necrosis factor α
TORC3:	transducer of regulated CREB protein 3
TRAIL:	TNF-related apoptosis-inducing ligand
TRBP:	TAR RNA binding protein
TRIM:	tripartite repeat motif protein
Tsg101:	tumor susceptibility gene 101 protein
UNG2:	uracil-DNA glycosylase 2
vhs:	virion host shut-off
Vif:	viral infectivity factor
v-onc:	viral oncogenes
Vpr:	viral protein R
VS:	virological synapse
VV:	vaccinia virus
VZV:	varicella zoster virus
WNV:	West Nile virus
X-PBP:	X-promoter binding protein

Viruses are obligatory parasites of their host cells. During the viral replication cycle, many host-cell components are co-opted or hijacked to favor viral replication and virus production. The interactions between the virus and the host induce multiple cellular responses that modify normal cell function and can lead either to clearing of the virus or to enhanced virus replication and pathogenesis. Most, if not all steps of the viral replication cycle are characterized by critical virus–host interactions. These begin with the earliest steps of the replication cycle (e.g. viral entry into the target cell) to later steps including mRNA synthesis, genome replication, and protein and viral genome trafficking within the nucleus and cytoplasm to sites of viral assembly. The understanding of the molecular details of these virus-cell interactions will ultimately help in the design of new and specific therapies against viral infection. The purpose of this review is to highlight various virus–cell interactions. Because of the extensive nature of this topic, we will be presenting general themes and specific examples of virus–cell interactions where appropriate rather than a comprehensive catalog of every virus–cell interaction that has been uncovered.

1 Virus Entry

Virus entry designates the process that ultimately permits delivery of the viral content into the target cell cytoplasm, to be replicated by viral and cellular proteins. This process can be grossly divided into two distinct major processes: recognition of the surface of a suitable target cell by the virus and the translocation process itself, that is, for enveloped viruses, fusion between the viral and cellular lipid membranes. These functions, receptor binding and virus-cell fusion, are mediated by the viral envelope proteins. The first step, the recognition of suitable target cells itself comprises several distinct, though partly overlapping, functions that are interactions with defined structures on the cell surface.

Nonenveloped viruses, such as small DNA-containing viruses and the RNA-containing viruses of the Picornaviridae and Reoviridae families, are devoid of a viral lipid bilayer and thus have to break the host-cell membrane instead of merging with it. Their entry mechanism into target cells is therefore fundamentally different from that of enveloped viruses, on which we will concentrate in this review. However, the principles guiding entry remain the same: nonenveloped viruses use cellular receptors to recognize and attach to target cells and receptor binding is followed by the activation of a viral fusion protein.

1.1 Virus Attachment to Target Cells

Attachment is defined as a low affinity, but high avidity interaction that permits velcro-like virus–cell association. These interactions facilitate the following entry steps, especially the interaction of the viral envelope with specific protein receptors, but alone do not lead to major conformational rearrangements in the viral proteins and the cellular attachment factor. Most attachment factors are repetitive carbohydrate structures, such as heparan sulfate

proteoglycans (HSPG) and glycolipids, but some lipids may also be involved in attachment. The nonexhaustive list of viruses binding to cell surface HSPG comprises viruses as different as papillomaviruses, parvoviruses, dengue virus, hepatitis C virus (HCV), herpesviruses, and human immunodeficiency virus (HIV). Herpes simplex viruses (HSV) are somewhat more selective in that they require the activity of specific sulphotransferases to generate HSV glycoprotein D (gD) binding sites. The importance of the contribution of the virus–HSPG interaction to viral entry is illustrated by the finding that treatment of target cells with heparitinase drastically reduces infectivity. Similarly, a role for phosphatidylserine in vesicular stomatitis virus (VSV) entry has been described, but the interaction of the viral envelope protein with the lipid does not actually fulfill the criteria for a viral receptor (see the following), but rather those of an attachment factor.

Another virus–carbohydrate interaction occurs between the glycosylated viral envelope proteins and C-type lectins such as dendritic cell-specific ICAM-3 grabbing nonintegrin (DC-SIGN), L-SIGN or hMGL. Virus binding to lectins seems, as interaction with HSPG, neither to induce major conformational changes in the envelope, nor to trigger membrane fusion and thus entry. Virus binding to DC-SIGN, which was first identified with HIV, nevertheless has dramatic consequences, as it increases the half-life of infectious virions and facilitates transmission to susceptible target cells. The lectin-bound virus may be internalized in intracellular vesicles, travels with the dendritic cell, and is presented to susceptible target cells at distant sites such as the lymph node. Virus–lectin interactions therefore enhance viral spread, and are far from being unique to HIV. As yet, DC-SIGN has been shown to interact with HCV, Ebola virus, Marburg virus, and coronavirus that causes severe acute respiratory syndrome (SARS-CoV).

A third mechanism by which viruses can enhance the efficiency of virus–cell contacts is the incorporation of cellular receptors in the viral envelop during egress. For example, incorporation of ICAM-1 and other cellular adhesion molecules into HIV-1 virions clearly affects the viral replication cycle and enhances the efficiency of infection (Table 1). Although not yet systematically studied in other viruses, it is possible that this mechanism has evolved as a general mechanism for enveloped viruses to optimize viral entry.

1.2 Viral Receptors

Viral receptors in the strict sense of the term are transmembrane proteins that are sufficient to permit viral fusion with any given cell type. Unlike attachment factors, they lead to conformational rearrangements in the envelope protein(s), and unfolding cascades of events that ultimately permit viral entry. Therefore, *in vivo* cell tropism directly relates to the capacity of viruses to infect targets via the expressed viral receptor. Of note, for many viruses there are multiple receptors that form either a receptor complex (e.g. receptor and coreceptor), or are alternatively used to gain entry into different types of target cells. A good example for this interplay between obligatory and/or alternative viral receptors are those permitting HIV-1 entry into target cells. All known HIV-1 strains depend on the single transmembrane domain receptor cluster differentiation 4 (CD4). Interaction between the viral envelope surface unit (SU) and CD4 is followed by envelope-binding

Tab. 1 Examples of virion-incorporated cellular proteins.

Gene	*Virus*	*Putative function*	*Incorporation*[a]
AIP1/Vps28	Retroviruses	Assembly/budding	Via Gag
APOBEC3G	Retroviruses	Deamination; infectivity	Via viral RNP and Gag
β-actin	HHV8/KHSV, HIV-1	Structural	Associated with tegument-capsid
c-PKA	HIV-1	Infectivity	Via Gag
CK2	VSV	Kinase activity on L, P and M proteins	Via viral RNP
CK2	Influenza virus	Kinase activity on NP, PA	Inside virion
Cyclophilin A	HIV-1	Uncoating; infectivity	Via Gag
DC-SIGN	Retroviruses	cell-to-cell transmission	In envelope
ERK2/MAPK	HIV-1, HTLV-I	Kinase activity on CA, MA	Via MA and CA
HBV-associated kinase	HBV	Infectivity	Via core
HP68	Retroviruses	Capsid/Gag assembly	Via Gag
HSV-associated kinase	HSV	Infectivity	Via tegument proteins
ICAMs	Retroviruses	Infectivity; cell-to-cell transmission; host-cell attachment	Via Gag
INI1	Retroviruses	Assembly, PIC import and integration	Via Gag/Pol
Myosin	HHV8/KHSV	Unknown	Associated with tegument-capsid
Nucleolin	MLV	Virion assembly	Via NC
PKC	Sendai virus	Kinase	Inside virion
Ribonucleotide reductase	HSV	Serine/threonine kinase	Via NC
Ribosome-associated protein	HBV	Infectivity	Via core
Staufen	Retroviruses, reovirus, other RNA viruses	Gene expression, RNA encapsidation, infectivity	Via retroviral RNP and genomic RNA
Tal1	Retroviruses	Virion budding	Via Tsg101
tRNA synthetases	Retroviruses	Virion assembly, reverse transcription	Via Gag/Pol
Tsg101	HIV-1	Assembly/budding	Via Gag
Uracil-DNA glycosylase (UNG2)	Retroviruses	PIC import and integration	Via Gag/Pol and/or Vpr
Virion-associated matrix-interacting protein, VAN	HIV-1	RNA localization; acts with HIV-1 MAp17	Via MA p17 domain of Gag
Vps4b	Retroviruses	Assembly/budding	Via Gag

[a] Location or mediator of incorporation is indicated.

to one of two alternative chemokine receptors: CCR5, which is used by all viral strains in the initial phase of infection, or CXCR4, which is mainly used by strains isolated in advanced stages of HIV disease. These chemokine receptors contain seven transmembrane domains and their extracellular regions are far closer to the cellular membrane than the envelope-binding domains of CD4. Consequently, a two-step mechanism for this virus–receptor interaction has been proposed: the first interaction permitting "capture" of the virion and induction of conformational changes in the envelope that expose a structure that will interact with the chemokine coreceptor. This second interaction triggers, by a still not well-understood mechanism, the activation of the fusion peptide in the envelope transmembrane unit (TM), which will lead eventually to membrane fusion (see the following). In the case of HIV, both receptor and coreceptor restrict viral tropism, and the switch from CCR5-usage to CXCR4-usage permits the virus to gain access to immature and naïve $CD4^+$-T cells. However, anecdotal reports also describe the occasional usage of CD8 as a primary receptor of some HIV strains. The identification of CD134, but not CD4, as a primary receptor for the closely related feline immunodeficiency virus (FIV), which also depends on CXCR4, further illustrates that there is, in principle, some evolutionary plasticity for receptor usage by viruses, as long as the new target cells permit efficient viral replication downstream of virus entry.

For many viruses, only one receptor has been identified as yet. These include one transmembrane domain receptor with several globular extracellular domains, such as the epidermal growth factor (EGF) receptor, which permits entry of the human cytomegalovirus (HCMV) into target cells, or the angiotensin converting enzyme 2 (ACE2) that is a receptor for the SARS-CoV. Receptors containing multiple membrane spanning domains can also function alone as viral receptors. One example is the human serotonin receptor 5HT2AR that mediates entry of the human polyoma virus JCV. Moreover, most retroviruses depend on multiple transmembrane domain containing transporter proteins, such as the glucose transporter (GLUT)-1 for the human T-cell leukemia virus (HTLV-I), the cationic amino acid transporter (CAT)1 for ecotropic murine leukemia virus (MLV), and many other similar transporters for the gamma (γ)-retroviruses.

There is an increasing number of viruses for which a complex interplay between several cell surface receptors, coreceptors, and attachment factors is reported, such as HSV, HCV, HIV, and adeno-associated virus 2 (AAV-2), and it is not always easy to determine which role is played by which molecule. Additional complexity comes from entry factors that are not mandatory but alternative, and determine entry in different target cells, that is viral tropism. Therefore, viral entry may be expected to remain a dynamic research field with many surprises in store.

1.3 Virus Fusion

Fusion between the viral and cellular lipid bilayers is ultimately achieved by the action of the fusion peptide. The action of this peptide is irreversible as it inserts into the host's cell membrane. This is triggered by events that differ between viruses, and is in many cases not well understood. Interestingly, there is increasing evidence that, at least in the case of HIV-1, the envelope-coreceptor interaction alone is insufficient to trigger efficient fusion

peptide activation, but requires the activity of extracellular enzymes of the host, such as protein disulfide isomerase (PDI).

There are some differences in the fusogenic structures of different viruses, and these can be classified as N-terminal (class I) fusion peptides or internal (class II) fusion peptides. Some viruses, such as influenza virus, are efficiently endocytosed by the host cell after receptor binding, and the decrease of the pH in endosomes triggers activation of the fusion peptide. For HIV-1, fusion is pH-independent and can occur at the plasma membrane, although endocytic HIV-1 uptake cannot categorically be ruled out. The event that triggers the activation of the HIV-1 TM fusion peptide remains elusive, but depends on SU interaction with chemokine receptors. Intriguingly, the fusion events downstream from fusion peptide activation seem similar between HIV-1 and influenza virus, despite their distance in evolutionary terms. Indeed, in both cases, intramolecular interactions between proximal and distal helices within the triple-stranded helical fusion protein bundles lead to rapprochement of the viral and cellular bilayers.

2 Innate Cellular Response to Virus Infection

2.1 Immediate Cell Response to Virus Infection

2.1.1 Cytokine and Chemokine Induction upon Virus Infection

As an immediate response to virus attachment or entry, the host cell will produce a large number of cytokines and chemokines. This induction has been well documented for herpesviruses, poxviruses, paramyxoviruses, influenza virus, adenovirus, VSV, hepatitis B virus (HBV), HIV-1, and HTLV-I, but probably occurs for most viruses. The triggering event is either the virus–receptor interaction, the early release of viral glycoproteins, nucleocapsid, or double-stranded (ds)RNA. The cytokines produced vary with the inducer and include tumor necrosis factor (TNF) α, various interleukins (IL)1–18, granulocyte-macrophage colony-stimulating factor (GM-CSF), macrophage inflammatory protein (MIP)-1α/β, transforming growth factor (TGF)β, monocyte chemoattractant protein (MCP)-1, regulated upon activation normal T-cell expressed and secreted (RANTES), and interferon (IFN)-α/β, and γ. Interestingly, infection with Lassa, Ebola, and Marburg viruses, three viruses that cause hemorrhagic fevers with high mortality rates, do not induce cytokine production when they cause death, emphasizing the protective role of the innate cell response.

2.1.2 Interferon α/β Production

One of the most important cell responses is the production of IFNα/β proteins (Fig. 1). Virus attachment to toll-like receptor (TLR) family proteins, viral nucleocapsid (NC), viral glycoproteins, and double-stranded RNA (dsRNA) will trigger the activation of multiple signal transduction cascades. One such cascade is the activation of the IκB kinase (IKK)-α, -β, and -γ complex that will lead to the activation of the inhibitory factor κB (IκBα) and the liberation of nuclear factor-κB (NF-κB). Another cascade is the IKK-related complex composed of TANK/I-TRAF, IKK-ε, and TBK1 that will lead to the activation of the IFN regulatory factor (IRF)3. Activating transcription factor (ATF) -2/cJun dimers will also be produced from the mitogen-activated protein (MAP) 3 kinase (MKKK) pathway.

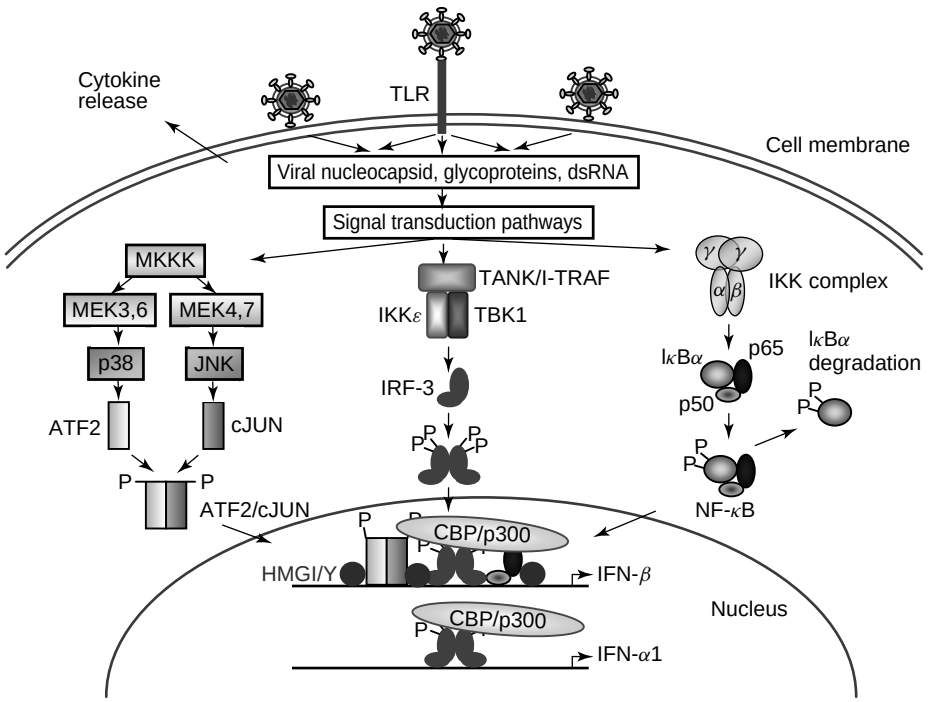

Fig. 1 Early events of the innate immune response after virus infection leading to IFN production and cytokine release.

These factors will be activated by phosphorylation and will translocate to the nucleus and bind to the IFN-α and -β promoters. IRF3 homodimers will transcriptionally induce the production of IFNα1, whereas ATF2/cJun, IRF3/IRF3, and NF-κB with the chromatin associated high-mobility group protein (HMG)I/Y will recruit the histone acetyl transferase (HAT) cAMP binding protein (CBP)/p300 and the chromatin remodeling protein SWI/SNF leading to the activation of IFN-β promoter. IRF1, IRF5, and IRF7 are also involved in the early production of IFN-α/β as shown for HIV-1 and paramyxoviridae.

2.1.3 IFN-γ Production

IFN-α/β are produced in most cells in response to a large number of viruses, whereas IFN-γ is produced mostly in T cells, macrophages, natural killer (NK) cells, and polymorphonuclear neutrophils. Although IFN-γ is mainly induced by antigens (Ags), mitogens, and cytokines, human papillomavirus (HPV), HSV, encephalomyocarditis virus (EMCV), influenza, respiratory syncytial virus (RSV), cytomegalovirus (CMV), and HBV induce its production. Its induction pathways are not fully elucidated, but likely involve IL-12, IL-18, and MIP-1α-mediated signaling.

2.2 The Interferon Pathway

The IFN-α/β produced in response to viral infection are secreted and bind to their receptor composed of IFN-α Receptor (AR)-1

and IFNAR-2 chains, which are associated with the Janus kinases Tyk2 and Jak1, respectively (Fig. 2). IFN-γ binds to the receptor IFN-γ Receptor (GR)1 and IFNGR2 chains associated with Jak1 and Jak2. Both will activate the Jak/Stat signal transduction pathway. The Jaks will phosphorylate Stat1 and Stat2 leading to the formation of Stat1-Stat2 heterodimers and Stat1 homodimers. Stat1-Stat2 heterodimers, mainly induced by IFN-α/β, will associate with IRF9 (p48) and the trimer, called ISGF3, will translocate to the nucleus and bind to the IFN stimulation response element (ISRE), a DNA element in the promoter of IRF7 and IFN-stimulated genes (ISGs). IRF7 will become phosphorylated, will associate with IRF-3, and will further contribute to an increased activation of IFN-α and -β promoters during the delayed events. Following activation by IFN-γ, Jak1 and Jak2 will induce Stat1 phosphorylation and dimerization. The Stat1 dimer will translocate to the nucleus and will bind to the IFN-γ-activated site (GAS) element. Stat1 bound to the GAS motifs will activate or inhibit IFN-γ regulated genes.

2.3 Activation of Host Genes in Response to Interferon

IFNs induce the synthesis of more than 300 ISGs. Some of them have been studied extensively for their antiviral activities, and the activity of many others are likely to be elucidated in the near future.

2.3.1 2′–5′ A Synthetase and RNAse L

2′–5′ Oligoadenylate synthetases (OAS) are a family of enzymes activated by dsRNA that catalyze the synthesis from

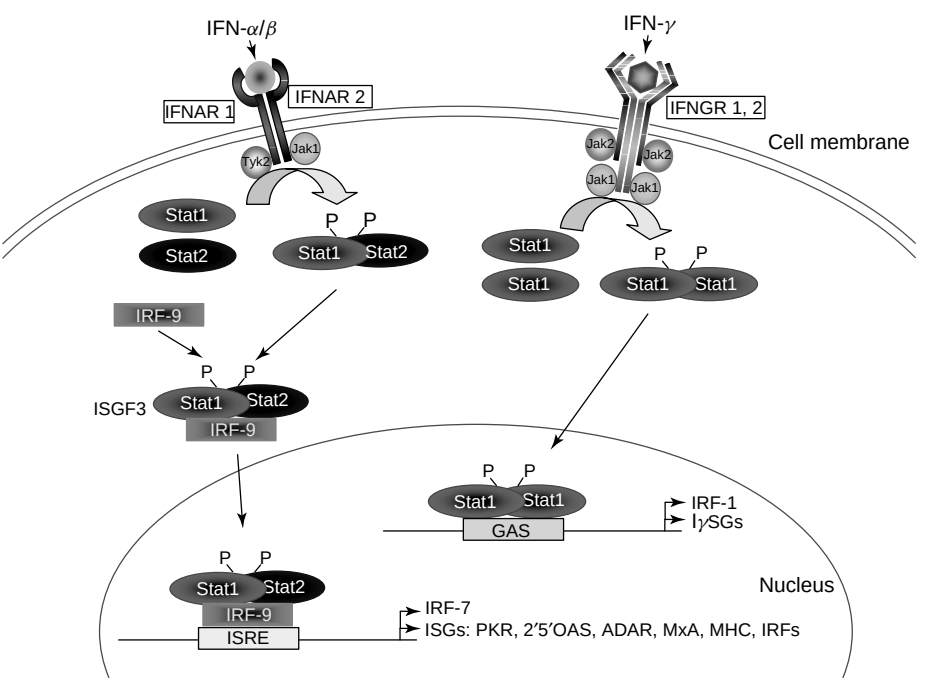

Fig. 2 Interferon pathways mediated by the binding of IFN-α/β and IFN-γ to their receptors.

ATP of oligomers of adenosine linked by 2′ 5′ phosphodiester bonds. The 2′–5′A oligoadenylates bind to endoribonuclease L (RNAse L) with high affinity. They induce its dimerization and endonuclease activity on viral and cellular RNA. In the context of virus infection, HIV-1 transactivation response (TAR), HTLV-I Rex responsive element (RxRE), Epstein–Barr virus (EBV), EBER-1, and adenovirus VAI RNAs are very good activators. In Picornaviridae, the activation of the OAS pathway correlates very well with the inhibition of viral replication. This system also plays a role in the antiviral effects of IFN-α/β against vaccinia virus (VV), EMCV, and reovirus.

2.3.2 Mx GTPase

The IFN-α/β-induced Mx proteins were first identified in the infection by influenza and Thogoto viruses, two members of the Orthomyxoviridae family. They belong to the dynamin-like GTPase family and exert their activity either by blocking the transport of viral nucleocapsid to the cytoplasm or by inhibiting viral RNA synthesis in the nucleus. MxA shows antiviral activities on several RNA virus members of the Orthomyxo-, Bunya-, Paramyxo-, Rhabdo-, and Togaviridae, but not Picornaviridae.

2.3.3 RNA-Specific Adenosine Deaminase, ADAR1

IFN-α and IFN-γ induce adenosine deaminase (ADAR1), that converts adenosine to inosine in viral RNA and cellular mRNA. A-to-I editing modifies RNA polymerase and ribosome recognition of the RNA and alters the sequence of replicated RNA and its translation. Extensive RNA editing leading to hypermutation occurs in measles virus, parainfluenza virus 3, VSV, Borna disease virus (BDV), avian leukosis virus (ALV), and polyomavirus (PyV), whereas limited selective editing is essential for hepatitis delta virus (HDV) protein synthesis.

2.3.4 Protein Kinase PKR

The IFN-inducible serine/threonine protein kinase PKR is activated by autophosphorylation, dsRNA, and homodimerization. Its best-characterized substrates are the α-subunit of the translation initiation factor 2 (eIF2α) and IκBα. The phosphorylation of eIF2α leads to inhibition of translation of viral and cellular mRNA, which strongly contributes to an antiviral state. All dsRNAs more than 30 nt in length can potentially activate PKR. Specifically, highly structured RNA such as HIV TAR, reovirus S1, and HDV RNAs are very good PKR activators. The antiviral activity of PKR has been demonstrated in cell culture for EMCV, influenza, HIV-1, VV, and VSV. This activity is certainly very broad and its intensity varies with the presence or absence of PKR antagonists in the viral genome.

2.3.5 ISG56

ISG56 is very strongly induced by virus infections including VSV, EMCV, Sendai, and influenza viruses as well as IFN, IRF3, and dsRNA. The gene encodes the P56 protein that interacts with the p48 subunit of the translation initiation factor eIF3 and blocks initiation of protein synthesis. Overexpression of the protein does not block VSV or EMCV replication in cell culture, but it contributes, with PKR, to the inhibition of HCV translation by IFN treatment.

2.3.6 Major Histocompatibility Complex Proteins

The major histocompatibility complex (MHC)/human leukocyte antigen (HLA)

class I and II molecules present the antigenic peptides derived from viral protein antigens to cytotoxic T cells. MHC-I and MHC-II are both activated during viral infection. MHC-I is induced by both IFN-α/β and IFN-γ, whereas only IFN-γ induces MHC-II proteins. Stat1 and IRF1 are involved in the IFN-induced expression of MHC class I antigens. The MHC-II transcription factor (CIITA) activates MHC-II protein expression and is the key mediator of IFN-γ activity. MHC-restricted recognition and killing of infected cells by lymphocytic T cells is a key component of the host response to viral infection.

2.3.7 Other Genes Induced by Virus Infection

Recent studies have used DNA microarrays to analyze the RNA profile of cells either treated by IFN or overexpressing IRF3. Similar studies have used virus-infected cells including CMV, papillomavirus, HCV, HSV, enterovirus, HBV, HDV, varicella zoster virus (VZV), adenovirus, and HIV-1. They all show the induction or suppression of known genes but also reveal unknown genes that may have a functional importance in the cell response to viral infection. Among them, the induction of ISG15, a ubiquitin-like protein, the inducible nitric oxide synthase (iNOS), and the P200 family involved in inhibition of gene transcription, cell growth, and differentiation show an increasing functional importance in viral replication.

2.4 Viral Antagonists to the IFN Response

A functional innate immune response to viral infection can lead to the suppression of viral replication. To complete their life cycle, viruses have evolved several mechanisms that involve virus–cell interactions to counter the cellular responses to viral infection. These include the inhibition of IFN synthesis, blocking IFN signaling and ISG function (Table 2). An additional mechanism to counter these cellular responses is to interfere with normal immune response.

2.4.1 Inhibitors of IFN Synthesis

The first mechanism to inhibit the IFN response is to prevent its synthesis either during the early events or during the late events of viral replication. To overcome the cellular IFN response, several viruses encode inhibitors of IRF3. HCMV pp65 suppresses the activation of IRF3. Ebola VP35, influenza NS1, HPV16 E6 bind to IRF3 and inhibit its activation and transduction pathways. Influenza NS1 and the V proteins of some paramyxoviruses block the activation of IRF3, IRF7, NFκB, and ATF2/cJun transduction pathways by dsRNA or viral products. Poxvirus protein N1L inhibits NF-κB and IRF3 signaling by TLR by association with the components of the IKK complexes. HCV NS3/4A protease blocks IRF3 activation by blocking more than one step upstream of the IKKε/TBK1 activity. In addition, E7 of HPV16 binds IRF1 and human herpes virus 8/Kaposi's sarcoma-associated herpesvirus (HHV8/KSHV) encodes a viral (v)IRF1 protein and a vIRF3 that prevent IRF1 and IRF3 activities by binding to CBP/p300.

ORF-C of HBV, the ML protein of Thogotovirus, the matrix protein of VSV, and NSs proteins of bunyaviruses inhibit IFN α/β synthesis, but the mechanism is not well characterized. A stronger inhibition of IFN synthesis by wild type measles virus compared to the vaccine strain may correlate with the expression of C and V virulence factors.

Tab. 2 Viral antagonists of the IFN response.

Virus	*Viral protein*	*Mechanism of inhibition*
1. Inhibition of IFN synthesis		
HCMV	pp65	Inhibits IRF3
Ebola	VP35	Binds to and inhibits IRF3
HCV	NS3/4A	Blocks IRF3 phosphorylation
HPV16	E6	Binds to and inhibits IRF3
	E7	Binds IRF1
Influenza	NS1	Blocks the activation of IRF3, IRF7, NFκB and ATF2/cJun
Paramyxoviruses	V	Blocks the activation of IRF3, IRF7, NFκB and ATF2/cJun
HHV8	vIRF1 and vIRF3	Binds CBP/p300, prevents IRF1, IRF3 activity
Adenovirus	E1A	Binds CBP/p300, prevents IRF3 activity
HBV	ORF-C	Unknown
Thogotovirus	ML	Unknown
VSV	Matrix	Unknown
Bunyaviruses	NSs	Unknown
2. Inhibition of IFN signaling		
Myxoma virus	MT-7	Decoy receptor, binds IFN-γ
Vaccinia virus	B8R	Decoy receptor, binds IFN-γ
	B18R	Decoy receptor, binds IFN-α/β
	VH1	Dephosphorylates Stat1
Adenovirus	Unknown	Downregulates IFN-γ-R2 receptor chain
	E1A	Decreases Stat1 and IRF9
		Interferes with Stat1-CBP and Stat1-IRF1
HCMV	Unknown	Decreases Jak1 and IRF9
Murine polyomavirus	LT	Binds to Jak1
Sendai virus	C	Interacts with Stat1, inhibits Tyk2, Stat1 phosphorylation
hPIV-2, SV5, MUV	V	Degrades or inhibits Stat1 or 2
HPV18	E6	Binds to Tyk2
HPV 16 and 31	E6	Downregulates Stat1
	E7	Binds to IRF9
VSV	Matrix	Blocks Stat3 phosphorylation
HSV-1	ICP0	Inhibits ISGs transcription
	UL41, UL13	Induces SOCS, a host Jak–Stat inhibitor
HCV	Core	Induces SOCS, a host Jak–Stat inhibitor

2.4.2 Inhibitors of IFN Signaling

IFN signaling is initiated by association of IFN with its cognate receptor. This step is inhibited in poxviruses that synthesize viral mimics of both IFN-α/β and IFN-γ receptors. Myxoma virus MT-7 and VV B8R are decoy receptors that bind IFN-γ, whereas B18R from VV and B18R-like from other orthopoxviruses bind IFN-α/β. They prevent IFNs binding to their native receptors and the subsequent signaling pathways. Interference with signal initiation also occurs with adenoviruses that downregulate the expression of IFNγ-R2 receptor chain.

Another major target of viral interference is the Jak–Stat signaling pathway. Although some proteins like HBx of HBV, or HTLV-I Tax activate the Jak–Stat pathways, many viruses encode inhibitors to shut down the IFN response. HCMV decreases the levels of Jak1 and IRF9,

whereas the large T (LT) antigen of murine PyV binds to Jak1, inhibiting IFN-α/β and -γ pathways. The C proteins of Sendai virus physically interact with Stat1 and mediate Tyk2, Stat1 phosphorylation inhibition and Stat1 degradation. Other members of the paramyxoviridae have a V protein that interacts with Stats and either degrade or inhibit Stat1 or 2. HPV18 E6 protein binds to Tyk2, prevents its binding to IFNAR1 and inhibits the Jak–Stat pathway of IFN-α/β. HPV16 and HPV31 E6 downregulate Stat1, whereas the E7 protein blocks ISGF3 formation by competitively binding to IRF9. Similarly, adenovirus E1A protein decreases the levels of Stat1 and IRF9, which blocks ISGF3 formation. E1A also interferes with Stat-CBP/P300 and Stat1-IRF1 interactions, resulting in a block in transcriptional activation of Stat1. VSV matrix protein blocks Stat phosphorylation via the IL-6 signal transducer gp130, and VV virion phosphatase VH1 dephosphorylates Stat1. HSV-1 ICP0 blocks the transcription of ISGs by a process that involves the host proteasome–mediated degradation.

HSV-1 and HCV infection lead to a marked suppression of IFN-induced phosphorylation of Jak1, Tyk2, Stat1, and Stat2 indicating an inhibition that precedes the Jak phosphorylation step. Indeed, these two viruses transcriptionally induce the suppressor of cytokine signaling-3 (SOCS3), a host Jak/Stat inhibitor, at an early step of infection. This activity is mediated by the UL41 RNAse, and UL13 protein kinase of HSV1, and by the core protein of HCV.

2.4.3 Inhibitors of ISGs

A very common way for viruses to counteract the IFN response is to inhibit the function of ISGs that inhibit virus replication. One of the best-known mechanisms is the inhibition of PKR by viruses. The OAS/RNAse L pathway and ADAR activity are also targeted by viral components. Interestingly, viruses have evolved mechanisms to counteract the activity of ISGs by several mechanisms (Table 3).

The best-studied poxvirus, VV, encodes for two proteins, E3L and K3L, which inhibit the activity of PKR and RNAse L. E3L is a dsRNA binding protein (dsRBP) that directly binds to PKR and sequesters dsRNA to prevent PKR activity. E3L also inhibits OAS activity by dsRNA competition and ADAR1 deaminase function. In contrast, K3L is an eIF2α homolog that acts as a pseudosubstrate of PKR to competitively inhibit eIF2α phosphorylation and restore viral translation. Several other poxviruses express or are predicted to encode homologs of E3L and/or K3L, but not molluscum contagiosum virus (MCV). In addition, Myxoma virus encodes a protein, M156R, which is a structural mimic of eIF2α.

HSV1 is the best-studied virus of the Herpes family. The virus encodes a protein, Us11, which is sufficient to prevent PKR activation. Us11 binds dsRNA; it has no known RNA binding domain but an RXP motif that mediates the binding. It could act either by sequestering dsRNA or by direct binding to PKR. Us11 also inhibits PKR activation by the cellular PKR activator (PACT), by protein–protein interactions. HSV-1 has a second mechanism to increase viral translation. The virus encodes $\gamma_1$34.5 protein, which associates with phosphatase 1α to dephosphorylate eIF2α. Us11 homologs can be found in several other α-herpesviruses like HSV-2 and simian B virus, whereas the γ-herpes virus EBV

Tab. 3 Viral inhibition of IFN-induced enzymes.

Virus	*Viral factor*	*Mechanism of inhibition*
1. PKR inhibitors		
Vaccinia virus	E3L	Binds dsRNA and PKR
	K3L	eIF2α homolog, pseudosubstrate for PKR
HSV1	Us11	Binds dsRNA, PKR and PACT
	$\gamma_1$34.5	Dephosphorylates eIF2α
EBV	SM	Unknown
	EBER RNA	Binds to PKR with no activation
Adenovirus	VA RNA	Binds to PKR with no activation
HIV-1	Short TAR RNA	Binds to PKR with no activation
		Recruits TRBP to inhibit PKR
	Tat	eIF2α homolog, pseudosubstrate for PKR
HCV	NS5A, E2	Binds to PKR
Influenza	NS1	Binds dsRNA and PKR
		Induces p58^{IPK} that inhibits PKR
Reovirus	σ3	Binds dsRNA and PKR
Rotavirus	NSP3	Binds dsRNA and PKR
SV40	LT	Unknown but after PKR phosphorylation
2. OAS/RNAse L inhibitors		
Vaccinia virus	E3L	Sequesters dsRNA
Influenza	NS1	Sequesters dsRNA
HSV1	Us11	Sequesters dsRNA
		Synthesizes 2′5′A antagonists
Reovirus	σ3	Sequesters dsRNA
Rotavirus	NSP3	Sequesters dsRNA
EMCV		Induces RNAse L inhibitor, RLI
HIV-1		Induces RNAse L inhibitor, RLI

encodes for the SM protein with RXP motifs that mediate PKR inhibition. HSV1 and HSV2 also induce the synthesis of 2′5′A derivatives that antagonize the OAS system.

Other mechanisms to inhibit PKR include the synthesis of large amounts of viral RNA that prevent PKR activation. This has been shown to be the case with the accumulation of EBV EBER, adenovirus VAI, and HIV-1 TAR RNAs. Viral dsRBPs like influenza NS1, rotavirus NSP3, and reovirus σ3 bind to PKR and prevent its activation, whereas HCV NS5A and E2 also bind to PKR and may participate in the IFN responsiveness in patients. HIV-1 Tat acts as a pseudosubstrate for PKR by mimicking eIF2α. Finally, simian virus 40 (SV40) LT acts after the phosphorylation step, all contributing to the inhibition of PKR activity.

Some viruses use the cellular machinery to inhibit PKR. Influenza NS1 induces the cellular inhibitor of PKR p58^{IPK} by promoting its dissociation with its negative regulator hsp40, whereas HIV-1 TAR RNA recruits the TAR RNA binding protein (TRBP), a potent cellular PKR inhibitor in close proximity to PKR. Finally, while proteins that sequester dsRNA can inhibit the OAS/RNAse L system, EMCV and HIV-1 downregulate RNase L activity through the increased expression of RNAse L Inhibitor (RLI).

3 Early Events in Virus Replication

3.1 Uncoating

Several types of enveloped viruses enter cells via a fusogenic route such that the lipid bilayer that envelopes the virus fuses with the plasma membrane. This is followed by the penetration of the genetic material and auxiliary factors of the capsid into the cytosol. The viral capsid must first pass through the dense cortical actin network located immediately beneath the plasma membrane. The cytoplasm represents a tightly packed, and somewhat organized environment, indicating that this early event of the replication cycle must be tightly regulated. While viral proteins such as HIV-1 Nef have been proposed to aid in this first assault on the cell, few details are known about this. Subsequent to this fusion, a capsid uncoating event (decapsidation) occurs. The movement of these large macromolecular complexes has been shown to be along cytoskeletal elements such as microtubules (Fig. 3). For HSV-1, adenovirus and HIV-1 for example, directed vectorial transport has been visualized microscopically (see next section) to direct the capsids containing the genetic material toward the nucleus. Few other details about this process are known because this subject of entry inhibition has been difficult to address experimentally. However, recent evidence and the creation of *in vitro* virus fusion and decapsidation system have allowed further insight into this process. This system allows for reconstitution of proteins and cellular environments to study this for MLV, equine infectious anaemia virus (EIAV), Rous sarcoma virus (RsSV), and avian sarcoma and leukosis virus (ASLV). The substrates used in this system are viral particles that are trapped before completing membrane fusion. These virions are induced to fuse out of endosomes and the viral cores are released into solution where they are amenable to biochemical manipulation. This system revealed that membrane fusion is not sufficient to stimulate the formation of a reverse transcription complex. Instead, ATP hydrolysis and cellular factors that are >5 kiloDaltons (kDa) in size are required. One salient observation is the necessity for an energy source, ATP, reciprocating this requirement *in vivo*. These *in vitro* systems however will require improvement and greater flexibility to recapitulate the *in vivo* setting. The identification of host-cell proteins in this process is essential and it will also be important to study other types of viruses in order to gain insight into and to fully appreciate the process of decapsidation.

3.2 Decapsidation and the TRIMs

The interest in this event has come to the forefront of HIV-1 research through the discovery of restriction factors that prevent decapsidation of retroviral capsids following viral entry. This viral restriction was noted in very early studies that showed restricted replication of MLV in mouse cells of different genetic backgrounds. Expression of distinct alleles of the mouse Friend virus susceptibility factor-1 (Fv-1) locus dictated susceptibility to infection by B- or N-tropic MLV. Fv-1, it turns out, bears a striking similarity to MLV coat proteins. Restriction factor 1 (Ref1) was also characterized to be a restriction factor for MLV infection of human cells, but this restriction can be overcome by infecting the cells with MLV, suggesting that

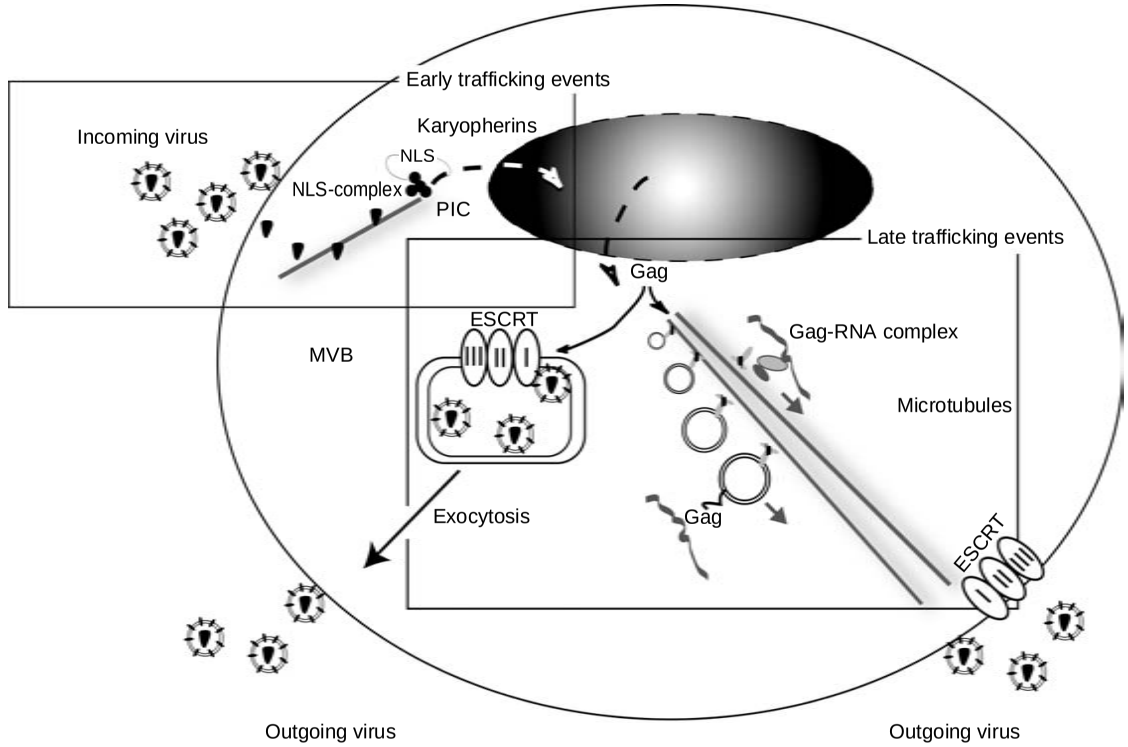

Fig. 3 The early trafficking events following entry of viruses into the cell, enveloped viruses are uncoated and the RNA–protein complex is trafficked along microtubules toward the nucleus (in HIV-1, this RNA–protein complex is translocated to the pericentriolar region or MTOC). Viral proteins that contain NLS interact with karyopherin proteins in the nuclear pore and form NLS-complexes that are imported into the nucleus. For viruses that require mitosis for entry (e.g. MLV), the nuclear membrane must be broken down during cell division before the NLS-complex can be exposed to the nuclear milieu. Following nuclear export, Gag is synthesized and associates to membranes via an amino-terminal myristoylation signal. Gag can be brought to MVB by direct interaction with class E type proteins for viral budding. This is reportedly the principal means of Gag trafficking in HIV-1 host cells like macrophages. The other scenario pertains to $CD4^+$ T cells budding mechanisms. Gag can traffic along microtubules attached to vesicular membranes to reach the cell plasma membrane, where it can bud using the same budding machinery involving class E proteins and the engagement of the ESCRT complexes. Viral RNA (blue squiggle) can be trafficked accordingly as a naked Gag-RNA complex or associated to vesicles (See color plate p. xxxii).

Ref1 can be saturated to allow infection by noncognate viruses like HIV-1. Infection of monkey cells by HIV-1 was not possible for reasons that have now become increasingly clear. New and Old World monkey cells express a protein that restricted HIV-1 infection and the block was shown to be at the level of decapsidation. While the restriction was linked to the HIV-1 capsid (CA) protein domain Lv-1, the tripartite repeat motif protein (TRIM) 5α was identified as the dominant cellular restriction factor that restricts HIV-1 infection of monkey cells. TRIM5α inhibits decapsidation of incoming HIV-1 capsids, thereby preventing subsequent steps in infection. Interestingly, a single amino acid substitution found in the TRIM5α protein from monkeys blocks HIV-1 decapsidation and reversion can restore decapsidation to salvage infection. Additional details of this process will undoubtedly reveal mechanistic information as well as identify other host factors that are involved in this process. Attention to this early step of infection could be the basis of future anti-retroviral compounds.

3.3 Reverse Transcription and the APOBECs

The reverse transcription process has been the focus of recent efforts to understand the reasons behind host factor-induced restriction to HIV-1 infection. Recent work points to restriction factors that are incorporated in virions from virus-producing cells that restrict early steps of virus replication. In the case that will be presented below, the acquisition of host-cell proteins is important for the generation of infectious virus particles. In the so-called "viral infectivity factor (Vif) permissive cells," HIV-1 infectious particles are generated regardless of whether Vif is expressed in producer cells. However, in cell lines that were found to be nonpermissive, infectious virus can only be produced when the HIV-1 auxiliary factor, Vif, is expressed in virus-producing cells.

Apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like (APOBEC) proteins have been shown to be important for restricting retroviral replication by triggering massive deamination of deoxycytidine within the retroviral minus (first)-strand complementary (c)DNA during a postentry step of the retroviral replication cycle. The impact on viral replication is evidenced by high frequencies of retroviral genomic strand G-to-A transition mutations. Recent work also points to the activity of ribonucleotide deamination, capable of mediating ribocytidine-to-deoxythymidine changes in genomes of RNA viruses, indicating that this may actually occur prior to reverse transcription. This added level of complexity may be a larger part of the innate defense mechanism to viral infection, as described above for TRIM5α. In HIV-1, the auxiliary protein Vif was shown to be key to overcoming this antiviral (or innate) response to infection by binding to and interfering with APOBEC's incorporation when it was expressed. This is achieved, at least in part, by impairing APOBEC3G translational efficiency and by targeting the APOBEC3G for proteasomal destruction prior to virion incorporation. Vif associates with APOBEC3G and directs it to the Cul5-elonginB-elonginC-RBx1 ubiquitin ligase complex and, as a consequence, APOBEC3G becomes polyubiquitinated and degraded by proteasomes. However, expression of Vif-defective HIV-1 was followed by massive deamination as shown by A->G mutations in sequence analyses to result in a catastrophic defect in infectivity. APOBEC proteins have now been shown to interact with a few viral proteins

to mediate incorporation and subsequent editing activity: APOBEC3G can also associate to the Gag NC domain. It will be interesting to determine whether Vif and NC coordinate to promote APOBEC interaction.

Interesting new data have revealed the negative replicative activity of APOBEC proteins in several viruses in addition to HIV-1, including HIV-2, simian immunodeficiency virus (SIV), MLV, EIAV, and HCV. The biochemical consequence of deamination appears to be identical even if most of these viruses (such as in the latter three viruses) do not naturally express Vif. This suggests that a Vif-like activity may exist in the cognate hosts of these viruses, or alternatively, these viruses have evolved to replicate in APOBEC-negative tissues. This perhaps reflects the narrow and restricted expression pattern for several of the APOBEC genes.

The relevance for APOBEC function lies in APOBECs' ability to restrict several viruses and this appears to occur across species. MLV replication for instance is inhibited by APOBEC action through cytidine deamidation of the *de novo* synthesized DNA. In SIVs, however, Vif's interaction with either APOBEC3G or APOBEC3F does not restrict strains of SIV in human cells. Thus, we do not have the entire picture of what is turning out to be a more complex picture of APOBEC function in retroviruses. APOBECs may interact with each other suggesting additive antiviral activities of multiple APOBEC proteins for a single virus. There are at least 11 APOBEC proteins expressed in vertebrates that can be coordinately expressed, and this is evidence for this added level of complexity. A multimodal mechanism of action for APOBEC proteins is suggested in viral replication, but this model must await identification of functional relationships between these proteins. Implicating multiple APOBEC proteins with temporal and activated functions timed by the viral replication cycle could reveal new details about the mechanisms of viral pathogenesis. The determination of the cellular and viral composition of the APOBEC complexes will likely be the subject of intense research in the upcoming year.

3.4 Intracellular Replication and Trafficking

Once inside the cell, virus capsids must make their way to the nucleus or other intracellular sites for reverse transcription, RNA or DNA replication. A formidable task must be undertaken by the virus to transport its genetic material to the nuclear membrane. Active transport along a microtubule-based cytoskeletal network appears to be the case for many viruses including adenoviruses, HSV, VV, and retroviruses following decapsidation, and for endocytosed viruses from the parvovirus family. This hijacking of vesicular transport mechanisms might be easier to visualize; nevertheless, viruses that unload their capsids into the host cells will likely require additional virus–cell interactions to traffic. The co-opting of the microtubule network and molecular motor proteins appears to be the mechanism of choice. Adenovirus-activated host cell–signaling cascades based on protein kinase A and MAP kinases require a *bona fide* extracellular virus–cell interaction to mediate entry, or this virus remains inert in the cytosol when entering the cytoplasm. HIV-1, and likely other retroviral capsids, are unloaded and track along microtubules to the microtubule organizing center (MTOC) region before entering the nucleus (Fig. 3). Some evidence demonstrates that viruses

can mediate the restructuring of the microtubule network early and late in their replication cycle by activated signaling cascades as mentioned above or by the specific action of regulatory proteins, for example, Rev in the case of HIV-1. A recent example of this is the significant restructuring of microtubules during HSV1 infection, in order to facilitate the nuclear localization of the major virion tegument protein, VP22.

3.5 Trafficking to the Nucleus

Viral genomes of viruses that replicate via a nuclear step or a DNA intermediate, such as retroviruses, hepadnaviruses, HSV1, influenza virus, and adenovirus, must bypass the nuclear envelope and enter the nucleus. Most viruses enter the nucleus via nuclear pores. However, these cellular structures have a pore size estimated to be 25 to 39 nm in diameter, which is too small for most viral capsids to penetrate, except hepadnavirus capsids. This suggests therefore that there is a requirement for an active transport mechanism to achieve this critical step of the viral life cycle. This is supported by the identification of MLV mutants that cannot bring their PICs into the nucleus in dividing cells and suggests an active transport process to bypass the nuclear barrier. Active transport is now known to be regulated by both viral and host proteins in both mitotic and nondividing cells.

To achieve nuclear import of the viral genome in HIV-1 for example, a preintegration complex or PIC is formed. This is composed of double-stranded linear DNA, viral proteins Matrix (MA), RT, integrase (IN), and the auxiliary protein, viral protein R (VpR). This PIC has an estimated Stokes diameter of 56 nm. Except for lentiviruses such as HIV-1, PICs from most retroviruses are unable to enter intact nuclei and must therefore "wait" for the breakdown of the nuclear membrane occurring during mitosis. Consequently, these retroviruses, such as MLV, are dependent in part on the cell cycle and cannot replicate in nondividing cells. In contrast, lentiviruses are able to productively infect nondividing cells such as macrophages or quiescent T lymphocytes, indicating that PICs are able to actively cross the nuclear membrane even if the cells are not cycling. Some other retroviruses seem to have an intermediate capacity to enter the nucleus, since the PICs of RsSV and foamy viruses (FV) are able to penetrate intact nuclei with a low efficiency, but their replication is dramatically increased in dividing cells.

3.6 Cellular Proteins and Nuclear Import

The karyopherin receptor complex is a heterodimer of an α-subunit, which recognizes nuclear localization signal (NLS) found within karyophilic proteins (both viral and cellular) and a β-subunit which docks the complex at nucleoporins that scaffold the nuclear pore. Several nucleophilic viral and host-cell factors that harbor NLSs have been shown to play key roles in this process. For HIV-1, multiple redundant functions for viral proteins might serve as the signal to route the PIC to the nucleus and through the nuclear pore, and host cell–protein factors associated with PICs might supply NLS activity. Indeed, several viral and host factors have been identified as components of the HIV-1 PIC. These include, for example Vpr, MA, IN, HMGI(Y), Ku, and barrier-to-autointegration factor (BAF). Several reports indicate a possible role for direct

interactions between karyopherin α and β with HIV-1 IN in the nuclear import of HIV-1 PIC. HIV-1 MA also possesses multiple NLSs and this is believed to be a primary determinant for nuclear import of the HIV-1 PICs via karyopherin α and β. The HIV-1 auxiliary protein, Vpr, colocalizes with karyopherin β and regulates karyopherin β-mediated docking of the HIV-1 PIC at the nuclear envelope, but this evidence remains to be substantiated.

Before migrating to the nucleus, HIV-1 PICs, containing IN, induce the cytoplasmic recruitment of integrase interactor 1 (INI1 or hSNF5) and promyelocytic leukemia protein (PML, or TRIM19, a member of the TRIM family of proteins), a protein found in nuclear bodies. Other viruses encode NLS-containing proteins that direct nuclear import such as nucleoprotein of influenza A virus. HPV16 L2 protein interacts with several karyopherins by its NLSs to promote nuclear entry via these import pathways (Fig. 3). Thus, to be determined are the fine molecular details on how particular host proteins associated with the PICs might supply an NLS activity to PIC. A recent example of this is the newly discovered involvement of importin-7 in HIV-1 PIC transport. Importin-7 is mildly hydrophobic, belongs to the importin-β superfamily, and is involved in nuclear import and chaperoning of basic nucleic acid–binding proteins, such as ribosomal proteins and histone H1. Via direct interaction with IN, importin-7 contributes to nuclear import of PIC in HIV-1. Importin-7 has also been demonstrated to have a principal role in nuclear import of adenovirus type 2 DNA, suggesting that importin-7 could have a more general role in mediating nuclear import of viral nucleic acids and complexes.

3.7 Cellular Proteins and Integration

As a suitable target for antiviral therapeutics, the process of proviral DNA integration by retroviruses and in particular HIV-1 has been a subject of intense research. This process involves the introduction of proviral DNA into the host chromosome by a series of biochemical reactions involving primarily the IN protein. The discovery of the roles of viral and cellular proteins in this process has led to a better understanding of this complex process. We will briefly mention some of the virus–cell interactions involved in this process. At least two of these proteins (i.e. INI1 and Uracil-DNA glycosylase 2 or UNG2) are acquired by the virus during assembly for later function in integration (Table 1).

Viruses ensure their survival by packaging some of the required cellular and viral components during assembly for postentry replicative events. This can have either a positive (in this case) or negative impact on viral replication (see Sect. 3.3 and Table 1). For integration, several cellular proteins have been found to interact with structural and polymerase proteins to achieve virion incorporation. INI1 was found to play a role in both virus assembly and earlier events in PIC nuclear import and integration. INI1 is incorporated into HIV-1 virions, presumably through its interaction with HIV-1 integrase (Table 1). Interaction between these proteins has been mapped to amino acids 183–243 of INI1 and the central domain of HIV-1 IN (amino acids 48–188). INI1 is part of a complex that relieves repressive chromatin structures, allowing the transcriptional machinery to access its targets more effectively. The encoded nuclear protein may also bind to and enhance the

DNA joining activity of HIV-1 IN. UNG2 is a glycosylase that removes uracil from both double- and single-stranded DNA in nuclear chromatin, thus contributing to base excision repair in the integration reaction. It is believed to be incorporated into virions via IN in the context of the precursor Gag/Pol protein since a virus that harbors two single point mutations in IN (L172A/K173A) is deficient for UNG2. Alternatively, it may be brought into the virus by a specific interaction with the auxiliary protein Vpr. In this case, this virus–host interaction was shown to influence HIV-1 mutation rate during reverse transcription. Other, not so well studied host proteins are involved in this process through characterized interactions with IN such as lens epithelium-derived growth factor (LEDGF/p75) considered to be the main lentiviral IN-to-chromatin tethering factor in several lentiviruses, but not in oncoretroviruses like MLV and HTLV-1. The IN/LEDGF/p75 interaction can also target the PIC to chromosomes and stimulate IN strand transfer activity at the same time. Finally, the HMG I(Y) protein mentioned above as a component of the HIV-1 PIC has been identified in a variety of PICs from several retroviruses and are active in integration assays for HIV-1, avian sarcoma virus (ASV), and MLV.

4 Viral Gene Expression

4.1 Synthesis of Viral mRNA

4.1.1 Transcription of DNA Virus Genomes

DNA viruses synthesize their mRNA from their DNA genome. The family with the largest genomes, the poxviruses, replicate entirely in the cytoplasm of the host cell. They show a high level of independence from the host cell, but borrow several host proteins to synthesize their mRNA. Poxviruses have similar transcription mechanisms, and VV is the model for this family. VV transcription can be separated into three phases: early, intermediate, and late gene transcription. The RNA polymerase and transcription factors are included in the virion, which allows the virus to synthesize all early class genes without any cellular factors. These include proteins participating in DNA replication, nucleotide synthesis, and intermediate gene transcription. Whereas early transcription occurs within the virus core particle, intermediate transcription starts after DNA replication in a more open environment. Key functions for both initiation and termination of intermediate and late transcription appear to be borrowed from the cell. Several host factors, including a nuclear DNA binding protein, have been characterized within the transcription complex formed around the viral RNA polymerase.

Herpesviruses have a nuclear step in their replication cycle. The prototype of the α-herpes viruses is HSV-1. The transcriptional control of its immediate early promoter by viral and cellular factors determines a latency stage or a lytic cycle. The virion protein VP16 is a key regulator of this mechanism through its association with two host-cell factors, Oct1 and HCF-1. All three proteins form a multiprotein complex that binds to the immediate early promoter through the Oct1 DNA binding domain. The precise interaction with Oct1 and HCF-1 mediates the activation and triggers the lytic cycle, whereas the binding of Oct2 or the association of VP16 with HCF-2 leads to promoter inactivation and

a latent mode of infection. Zta, a bZIP protein of the γ-herpesvirus EBV, is a trans-activator that acts by interactions with DNA, TFIIA-D transcription factors and other transcriptional activators. Like VP16 on HSV1, it triggers the lytic cycle of EBV.

Small nuclear DNA viruses including polyoma-, parvo-, papilloma-, and adenoviruses do not code for DNA or RNA polymerases and therefore rely entirely on the cellular enzymes to complete their replication and transcription. The ability of the cellular DNA polymerases to replicate the viral DNA, and the presence or absence of various transcription factors will therefore restrict the replication and the expression of these viruses to specific species and cell types. The early promoter of HPVs is recognized by cellular transcription factors. The first viral proteins to be expressed are the replication factors E1 and E2. They form a complex at the viral origin to recruit cellular polymerases and accessory proteins to mediate replication. Binding of low concentrations of E2 at sites adjacent to cellular transcription factors activates the early promoter, whereas at high concentrations, it induces a repression by blocking the binding of cellular factors.

The transcription of adenovirus requires complex interactions between cellular and viral factors. It starts with the recognition of the E1A transcription unit by cellular proteins alone. As the two forms of E1A proteins lack a DNA binding activity, they associate with transcription factors, proteins of the general transcription machinery, coactivators and chromatin-modifying enzymes to activate transcription of viral early genes. They also release cellular repression of the viral E2 promoter. The E2 late promoter is then activated by a joint coordination between the viral E1B and the host-cell factor YB-1 to produce the E2 viral replication proteins. The expression of E2 genes triggers DNA replication. The IVa2 promoter is bound by a cellular repressor and this promoter inhibition is released when viral DNA synthesis increases and binds the repressor. The IV2a protein, a DNA binding protein, is then produced and stimulates the major late (ML) promoter. The complete interplay between viral and cellular factors leads to the formation of viral particles.

HBV has a partially double-stranded DNA, which is converted to covalently closed circle (ccc) in the nucleus by the host cell. This ccc DNA is the major transcriptional template for the virus, and its efficient regulation is the result of a complex interplay between trans-acting factors and their corresponding cis motifs. HBV transcription is mediated by four different promoters and two enhancer elements. The precore promoter is regulated by a number of cellular transcription factors including nuclear receptors. These are ubiquitous nuclear receptors or hepatocyte nuclear factors (HNF) 3 and HNF4. They have activating or repressing activity leading to the synthesis of either the precore or the core RNA with different biological activities. Whereas the basal preS1 promoter is regulated by the TATA-binding protein (TBP) and HNF-1 binding sites, its activity is enhanced by SP1, NF1, and HNF-3 to produce the large surface antigen. The TATA-less promoter S2 is regulated by a CCAAT motif bound by NF-Y/CBF transcription factors. It directs the production of the middle and major surface antigen transcript. The S2 promoter is also indirectly activated by the accumulation of the large surface antigen to the endoplasmic reticulum, which likely releases transcription factors that bind to the S2 promoter. The X promoter regulates the synthesis of

the mRNA coding for HBx, the viral trans-activator. The X promoter is regulated by ubiquitous and liver-specific transcription factors such as the X-promoter binding protein (X-PBP). HBx does not bind DNA directly but binds to and modifies the function of transcription factors to increase the activity of all the other HBV promoters.

4.1.2 Transcription of Viruses Integrated into the Host Chromosome

Retroviruses are RNA viruses that convert their RNA into DNA by reverse transcription. This DNA is then transported to the nucleus and integrated into the host genome. Basal transcription is regulated by cellular transcription factors and activators. Some retroviruses also encode viral trans-activators whose function requires a tight interaction with host factors. Two typical examples, the HTLV-I Tax and the HIV-1 Tat proteins, have a similar function but important mechanistic differences (Fig. 4). The HTLV-I long terminal repeat (LTR) contains three 21-bp repeats required for Tax activation. Each of these Tax-responsive sequences (TxRE) contains cyclic-AMP response element binding (CREB)/ATF binding sites flanked by GC-rich sequences. In the nucleus of the host cell, Tax dimerizes and interacts with CREB, CBP, and P/CAF cofactors. The formation of a ternary complex, Tax/CREB/TxRE, mediates a specific HTLV-I Tax trans-activation, as Tax is not able to activate transcription at other CREB/ATF sites on cellular promoters. Three Tax-interacting factors are involved in this specificity. P/CAF binds to the C-terminal activation domain of Tax, and the binding correlates with the trans-activation ability of Tax. The SR-related protein TAXREB803/SRL300 is a DNA binding protein that binds Tax and the 3′ GC-rich sequence of TxRE. It functions as a Tax coactivator in the expression of HTLV-I genes. The transducer of regulated CREB protein 3 (TORC3), a coactivator of CREB, also binds Tax and increases transcriptional activation at the HTLV-I LTR. Although the exact coordination of all these host and viral factors remains to be fully determined, their requirement in Tax trans-activation suggests the formation of a large multiprotein complex on the TxRE site that will contact factors of the general transcriptional machinery.

The HIV-1 LTR is controlled by a large number of transcription factors with SP1 and NFκ-B as the most proximal from the transcription start site. Cellular components control the basal level of HIV-1 LTR expression, whereas the activated level is regulated by the viral trans-activator Tat. In contrast to all other known trans-activators, Tat has an RNA target, the TAR element, located at the 5′ end of all HIV-1 mRNAs. *In vivo* trans-activation by Tat occurs by a multistep mechanism that requires a coordinated activity of cellular and

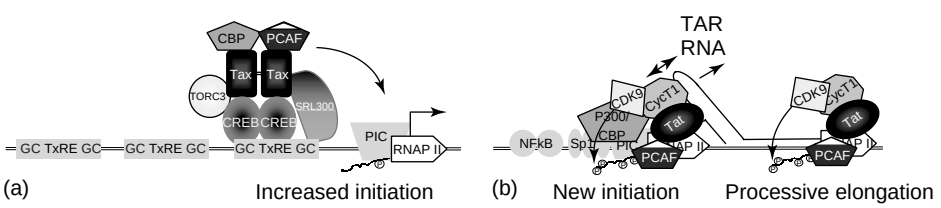

Fig. 4 Comparison of the trans-activation mechanisms mediated by HTLV-I Tax and HIV-1 Tat with their associated host factors.

viral factors. As a first step, Tat is acetylated by P/CAF at K28 and the Ac28-Tat has a very high affinity for the positive transcriptional elongation factor (P-TEF)b, composed of Cyclin T1 (CycT) and the cyclin dependent kinase (CDK) 9. Indeed, Tat binds directly to CycT, which itself binds to CDK9. The Tat-CycT-CDK9 complex has a very high affinity for TAR RNA and forms a ribonucleoprotein complex. This complex recruits p300/CBP, which acetylates Tat at K50. As a result, Ac50Tat-CycT-CDK9 has a decreased affinity for TAR and will be transferred to the next preinitiation complex paused after the carboxy terminal domain (CTD) of the RNA polymerase (RNAP) II has been hypophosphorylated by CDK7. The association of p300/CBP-Ac50Tat-CycT-CDK9 with the preinitiation complex forms an efficient transcription elongation complex in which CDK9 hyperphosphorylates the RNAP II CTD and maintains this phosphorylation during the elongation of HIV-1 mRNAs. These complex interactions between cellular and viral factors provide a very efficient trans-activation mechanism of HIV-1.

4.1.3 Synthesis of Coding mRNA by RNA Viruses

RNA viruses that do not use a DNA intermediate for their replication synthesize their mRNA and replicate using a viral RNA-dependent RNA polymerase (RdRp). This viral enzyme or complex of enzymes renders the synthesis of mRNAs largely independent from the cell and limits virus–cell interactions to a few necessary exchanges. Plus-strand RNA viruses, including Picorna-, Flavi-, Corona-, and Togaviridae, copy the genomic RNA to form a complementary minus-strand RNA using a virus-encoded RdRp. This new RNA in turn serves as a template for the synthesis of new genomic RNA. The sites of viral RNA synthesis are exclusively in the cytoplasm and have been localized to structures derived from diverse intracellular membranes, including the lysosomes, endoplasmic reticulum (ER), Golgi apparatus, and endosomes. Although most RdRps function *in vitro*, their activity remains low, and template specificity is a property of the viral polymerase complexes found in infected cells. Membrane-associated replication complexes, composed of viral proteins, replicating RNA, and altered host-cell membranes, represent a common feature of all plus-strand RNA viruses investigated thus far.

Most negative-strand RNA viruses replicate in the cytoplasm of infected cells. Their RdRp carries out two types of RNA synthesis using the same RNA template associated with the nucleocapsid N protein: (1) mRNA synthesis, called *transcription*, in which the RdRp synthesizes each mRNA successively, and stalls to allow polyadenylation and new start; (2) *replication*, corresponding to RNA synthesis of full-length antigenomic RNA. In the first case the RdRp pauses frequently, whereas it is highly processive in the second, likely due to the coupling of replication and viral assembly. Cellular proteins participate in the formation of either the transcriptase or the replicase complex and likely govern RdRp processivity.

All the dsRNA viruses also encode their own enzymes necessary for transcription and replication. The best-studied ones are the rotaviruses in the Reoviridae family, in which a core particle is released in the cell cytoplasm after the removal of the outer membrane. The transcription from the dsRNA occurs within the core particle and the exchange with the cell is limited to pores that allow the entry of nucleotides, *S*-adenosyl methionine, and ions, as well as the exit of mRNA molecules to the

cytoplasm. Translation takes place in the cytoplasm, whereas genome replication occurs in viral inclusions called *viroplasms*.

4.1.4 Virus-induced Cell Transcription Shut-off

One dramatic result of a virus–cell interaction is the inhibition of nuclear transcription, a phenomenon referred to as *transcription shut-off*. This mechanism has been described with cytopathic RNA viruses, which replicate exclusively in the cytoplasm and encode their own RdRp. The best-studied example of this phenomenon occurs following poliovirus infection. Once the viral RNA is in the cell cytoplasm, it is translated by the host-cell translational machinery into a polyprotein that is then cleaved into three precursor proteins (P1-P3) by virus-encoded proteases. The P3 precursor will then give rise to 3CD, a precursor protease that contains an NLS in the additional region. The self-cleavage of this precursor will generate the viral protease $3C^{pro}$ in the nucleus. $3C^{pro}$ contributes to the shut-off of all three transcription systems by the degradation of transcription factors including TBP. Among the same family, the foot-and-mouth disease virus (FMDV) also induces transcription shut-off by the degradation of histone 3 by its $3C^{pro}$.

Among the negative-strand RNA viruses, the M protein of VSV also induces transcription shut-off by targeting TBP, but the mechanism is not degradation. The Rift Valley fever virus (RVFV) has evolved a different mechanism by targeting TFIIH. The NSs protein of RVFV interacts with the p44 component of TFIIH to form nuclear filamentous structures that also contain the XPB subunit of TFIIH. Therefore, NSs sequesters p44 and XPB subunits and prevents TFIIH formation. TFIIH has a function in RNAP II, and RNAP I transcription, as well as in nucleotide excision repair (Table 4). Viruses take advantage of transcription shut-off to favor their own synthesis, and this activity is certainly correlated to pathogenesis, although not always proven. For some viruses like HDV, the target factor remains unknown and it is likely that other viruses and other mechanisms will be discovered in the future.

4.2 Nuclear Export

4.2.1 Export of Viral RNA

Viruses that synthesize their mRNA in the nucleus need to export it to the cytoplasm using the cellular export machinery. Two well-described examples are HIV-1 Rev and HTLV-I/II Rex, two viral proteins with similar function and mechanism (Fig. 5). Their function is to export unspliced and singly spliced viral RNAs through the binding to a specific RNA sequence, the Rev/Rex responsive element RRE or RxRE. Following HIV genome integration in the host chromosome, transcription is initiated. mRNA from the entire genome is generated and then fully spliced by the host splicing machinery. In the early phase, only the multiply spliced mRNAs are exported to the cytoplasm to translate Tat and Rev and Nef proteins (HIV), or Tax and Rex (HTLV), whereas longer, intron-containing mRNAs are retained in the nucleus. Tat and Rev, Tax and Rex are translocated to the nucleus where Tat and Tax increase transcription, and Rev and Rex bind to RRE and RxRE respectively, to mediate the export of unspliced and singly spliced mRNAs. This starts the late phase of viral expression. RRE is not present on multiply spliced HIV RNAs but all HTLV transcripts contain RxRE. Rev and Rex bind to their RNA

Tab. 4 Mechanisms of virus-induced cell expression shut-off.

Virus	*Viral protein*	*Mechanism of shut-offion*
1. Transcription shut-off		
Poliovirus	3C protease	Degrades TBP, CREB, Oct-1, P53, TFIIICα, SL1
FMDV	3C protease	Degrades Histone 3
VSV	M	Inhibits TBP
HSV	ICP22 and UL13	Alters RNAPII hyperphosphorylation
	ICP4, ICP27 and ICP8	Displaces TFIIE from RNAP II
RVFV	NSs	Prevents TFIIH formation
2. Nuclear export shut-off		
Polio and rhinovirus	Protease	Degrades Nup153 and p62 in the NPC
VSV	M	Interacts with and inhibits Nup98
Parvovirus	NS2	Sequesters CRM-1
HBV	X	Sequesters CRM-1
Influenza	NS1	Interacts with and inhibits CPSF and PABP
HSV	ICP27	Binds to splicing factors
3. Translation shut-off		
HSV	UL41 (*vhs*)	Degrades cellular and viral mRNA
Picornaviruses	2A-protease	Cleaves the N-terminus of eIF4G
		Impairs all cap-dependent mRNA translation
Adenovirus	100 kDa	Dephosphorylates eIF4E
Rotavirus	NSP3	Binds to eIF4G and excludes PABP

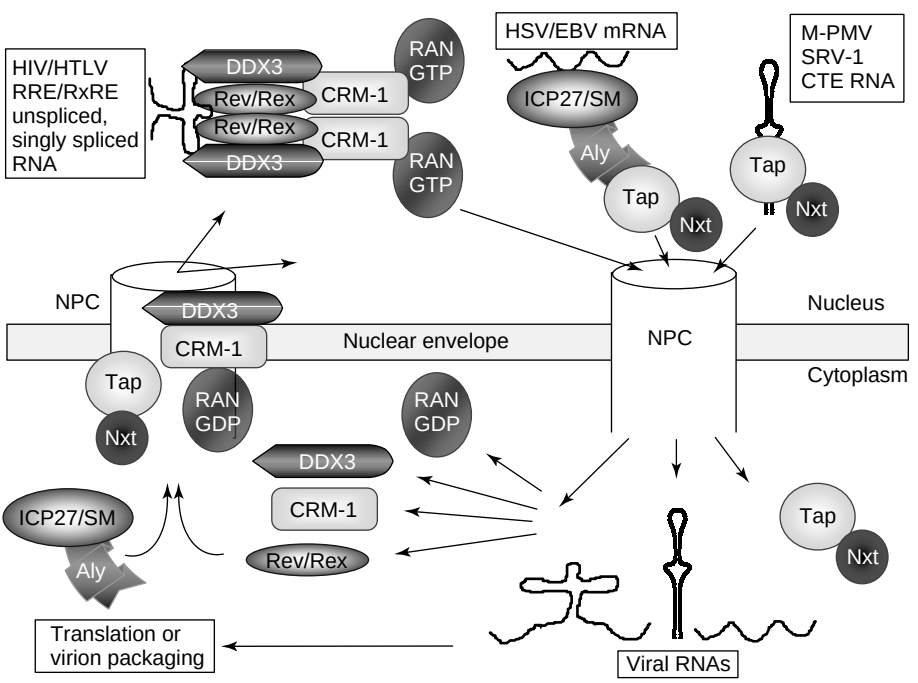

Fig. 5 Mechanisms of viral nuclear mRNA export and their cellular cofactors.

targets through an arginine-rich motif and to the chromosome maintenance region 1 (CRM-1) through their nuclear export signal (NES). Rev (and maybe Rex) also binds to an RNA helicase, DDX3, required for its function and to CRM-1. CRM-1 binds to RAN-GTP, and this ribonucleoprotein complex binds to nucleoporins localized in the nuclear pore complex (NPC) to deliver the RNA in the cytoplasm for translation and virion packaging. Singly spliced and unspliced mRNAs encode Gag, Gag-pol, Env, and other accessory proteins. Many other cellular proteins help the nuclear mRNA export by different mechanisms. For example, Src-associated during mitosis 68 kDa protein (Sam68) and human Rev-interacting protein (hRIP) are cofactors for Rev function, whereas heterogeneous nuclear ribonucleoprotein A2 (hnRNPA2) promotes HIV RNA export through a different pathway. In contrast, hnRNPA1 competes with Rex for RxRE binding.

Other RNA export mechanisms have been described. Retroviruses like the Mason–Pfizer simian virus (MPSV) and simian retrovirus type-1 have constitutive transport RNA elements (CTE) that are recognized by the cellular protein Tap to mediate constitutive transport of their RNA. Tap forms a heterodimer with a small cofactor Nxt, and the CTE-Tap-Nxt complex mediates a high affinity association with components of the NPC to mediate mRNA nuclear export. Herpesviruses also use the Tap-Nxt pathway for nuclear export. The HSV protein ICP27 and its EBV homolog SM, selectively bind to their viral mRNAs and to a cellular factor Aly that itself binds to the Tap-Nxt complex, to direct the mRNA to the NPC. The mechanisms used by other DNA viruses are not so well known, but some studies suggest that adenovirus use several pathways, whereas HBV has a posttranscriptional regulatory element (PRE) with some functional analogies with the CTE mechanism.

Most RNA viruses replicate in the cytoplasm, but influenza, Thogoto, and BDV negative-strand viruses replicate their RNA in the nucleus and then need to export it to the cytoplasm. The influenza M1 protein promotes termination of RNA replication and binds to vRNPs. This complex then recruits the NS2/NEP protein with an NES that interacts with CRM-1. Therefore, influenza and also probably BDV use the CRM-1/RAN-GTP pathway, but no such evidence has been found for Thogoto virus.

4.2.2 Virus-induced Cellular Nuclear Export Shut-off

Whereas viruses that synthesize their RNA in the nucleus subvert the functional cellular import–export machineries for their benefit, some viruses affect this pathway and mediate a general inhibition of nuclear-cytoplasmic trafficking. In poliovirus and rhinovirus infections, the analysis of the NPC revealed that at least two proteins, Nup153 and p62, were proteolyzed shortly after infection, likely due to the activity of a viral protease. As a consequence, several shuttling cellular proteins were found mainly in the cytoplasm. VSV M protein interacts with nucleoporin Nup98 to mediate inhibition of nuclear mRNA transport. IFN treatment increases the level of Nup98 and Nup96, reversing the effect of M, suggesting that the host has evolved mechanisms to counteract the IFN-antagonist properties of VSV M. The parvovirus NS2 and the HBV X proteins use a different mechanism, which is the cytoplasmic sequestration of CRM-1, therefore affecting important cellular processes. The influenza virus NS1 protein inhibits pre-mRNA splicing and

polyadenylation by interaction with the cleavage and polyadenylation specificity factor (CPSF) and the polyA binding protein (PABP), leading to the accumulation of pre-mRNA in the nucleus (Table 4). The investigation of more mechanisms will certainly determine more export pathways affected during viral infection.

4.3 Translation

4.3.1 Initiation of Translation of Viral mRNA

The early steps of virus replication result from an equilibrium between viral and cellular factors. Some viruses, however, have evolved toward a large autonomy from the cell. In contrast, no virus encodes the components necessary for autonomous translation of viral RNA. Therefore, all viruses have to compete with cellular protein synthesis to ensure that their RNAs are translated efficiently. As a consequence, viruses use the full variety of mechanisms available for cellular mRNA translation. The main mechanism is the classical scanning mechanism, whereby the 40S ribosome binds at the 5′ cap structure on the mRNA, and scans toward the translation initiation codon. Variations of this mechanism include ribosome reinitiation, leaky scanning, non-AUG start codons, frameshifting, termination suppression, and ribosome shunting; all are used to different extents by viruses. Several viruses, mainly from the Picornaviridae and Flaviviridae families, have uncapped mRNAs and use an internal ribosome entry site (IRES) to initiate their translation.

The scanning mechanism of translation starts with the recognition of the m7G cap nucleotide structure by the eukaryotic initiation factor (eIF) 4E (cap binding protein) and eIF4G associated with the RNA helicase eIF4A. EIF4G binds to eIF3, which is linked to the 40S ribosome, and to the PABP, which is bound to the 3′ polyA tail. PABP links the 3′ and 5′ ends of the mRNA to ensure efficient translation. Translation initiation requires the loading of the 40S ribosomal subunit linked to eIF2-GTP and the initiator Met-tRNA. Some viruses encode factors that can substitute for cellular factors in the translation scanning process. Rotavirus mRNAs released in the cytoplasm are capped but not polyadenylated. To replace the PABP and its link to eIF4G, rotaviruses encode NSP3, a dsRBP that binds the 3′ end of the mRNA and eIF4G with greater affinity than PABP. Thus, NSP3 subverts the host translation machinery to enhance the translation of virus-encoded mRNAs. Influenza virus may use a similar mechanism because its NS1 protein binds to viral mRNA and eIF4G, but the mechanism is not so well understood (Fig. 6).

The IRES mechanism has been described for several viruses but seems to be the sole translation mechanism used by Picornaviridae, of which poliovirus has been the best studied. Its mRNA is uncapped and its structured 5′ end directly recruits the 40S ribosomal subunit without the requirement for eIF4E, PABP, or the N-terminus of eIF4G. Its 3′ mRNA end is polyadenylated and binds PABP, but it is not clear if this protein is involved in mediating 5′->3′ mRNA association at some step. The flavivirus HCV IRES is also well studied. It does not require eIF4G and its 3′ mRNA end is not polyadenylated. Its translation is initiated by direct binding of the 40S ribosomal subunit and its associated eIF1A, 2 and 3 factors to the RNA.

4.3.2 Translational Regulation by Cellular Factors

In addition to the activity of factors from the general translation machinery, other

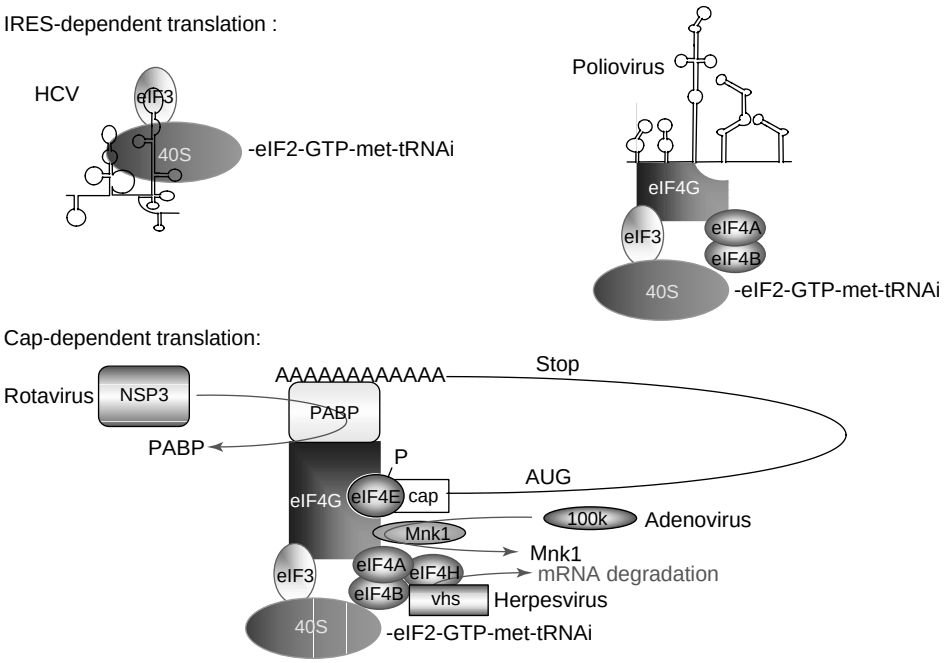

Fig. 6 Subversion of the host cell–protein synthesis by viral proteins.

cellular factors influence translation efficiency of either cap-dependent or cap-independent mechanisms. One general translation inhibition factor is PKR (as described earlier), which is induced by IFN and activated by dsRNA, heparin, or PACT. Its activity results in a general viral and cellular translation shut-off, but cellular and viral factors counteract its activity. One factor that acts both on PKR and on a PKR-independent pathway is the cellular protein TRBP that increases translation from transcripts containing HIV-1, TAR, and other structured RNAs. Similarly, but probably by a different mechanism, the lupus autoantigen, La, increases translation from HIV-1 TAR, but also binds to IRESs. Indeed, La and the polypyrimidine tract binding protein (PTB) seem to be required for efficient translation from poliovirus and HCV IRESs. Retroviruses provide another example of exploiting host translation factors. The production of Gag-Pol protein of MLV occurs by misreading a stop codon in 5% of the translation. This codon is normally recognized by the eukaryotic release factor 1 (eRF1) bound to eRF3 to produce the Gag protein, but the binding of the MLV RT with eRF1 inhibits eRF1 action and enhances the efficiency of readthrough. Other examples of alternate mechanisms can be found in a large variety of viruses as the result of cell-adapted virus mechanisms.

4.3.3 Virus-induced Cell Translation Shut-off

As a consequence of using the cell's translation machinery, all viruses limit the

availability of cellular translation factors for cellular protein synthesis. In addition, some viruses have evolved specific pathways to induce a rapid shut-off of host cell–protein synthesis to favor their own translation. One well-described example is HSV. Upon viral entry, a rapid shut-off of host cell–protein synthesis can be observed by the disruption of preexisting polyribosomes and degradation of host mRNA in the absence of *de novo* viral gene expression. This mechanism is triggered by the virion host shut-off (*vhs*) protein, an mRNA-specific Rnase encoded by HSV UL41. This protein, located in the virion tegument, degrades directly or indirectly cellular mRNA in the cytoplasm through interactions with one or more cellular translation initiation factor (Table 4).

Other mechanisms involve the subversion of translation factors. Picornaviruses encode a protease (2A-pro) that cleaves the N-terminus of eIF4G, therefore preventing eIF4E and PABP binding and impairing all cap-dependent mRNA translation. Among this family, EMCV does not cleave eIF4G but activates an eIF4E repressor. During adenovirus translation shut-off, eIF4E is strongly dephosphorylated, which inhibits cap-dependent translation. Adenoviruses encode a 100-kDa protein that replaces the Mnk1 kinase (which normally phosphorylates the eIF4E cap binding protein) to mediate the dephosphorylation of eIF4E. To selectively continue to translate viral cap-dependent mRNAs, adenovirus uses ribosome shunting, which involves loading of the 40S ribosome onto the mRNA in a cap-dependent manner, followed by direct translocation to the initiation codon without scanning the mRNA. Several other viruses clearly induce a translation shut-off, and more information about these mechanisms remains to be discovered.

5 Virus–Host Interactions Involved in Viral Egress

5.1 Ribonucleoprotein Transport

Interesting new data reveal a role for cellular proteins and major cellular machineries in late viral trafficking events. The assembly of viral particles requires the coordinated movement of all of the viral components to sites of assembly. It has been known for some time that the cytoskeletal network is used as a scaffold for the movement of viral proteins and capsids (or assembly complexes) toward the periphery of cell for release during viral egress (Fig. 3). This is the case for several retroviruses and HSV1. MPSV, a type B retrovirus, not only accumulates at a pericentriolar region but its spherical capsids are also formed there and are transported to the plasma membrane on microtubules to acquire an envelope during the budding process. Many studies have employed host cell–microtubule-destabilizing reagents such as nocodazole to disassemble the microtubule cytoskeleton. This has provided evidence for a directed movement along microtubules and the activity of host-cell kinesin and adaptor proteins to propel the viral components and assembly complexes. HSV1 capsids also require directed transport to the periphery for budding and utilize the cytoskeletal machinery to achieve this. HSV1, however engages dynein, and not one of the conventional kinesin molecules, as the motor protein to propel it essentially out of the cell. HIV-1 trafficking is achieved by microtubule-dependent machinery and it can be stopped by nocodazole treatment. HIV-1 and MLV Gag's association to the molecular motor protein, kinesin KIF4A,

likely aids in propelling HIV-1 and MLV Gag assembly complexes through the cytosol toward the plasma membrane to sites of virus budding. To be propelled, VV mimics signaling and phosphorylation events to promote polymerization of actin filaments likely to provide a scaffold for viral exit from the cell. This has been visualized in cells by identifying actin-rich cytosolic extensions that trail VV particles in cells and in tubular cellular protrusions (or pseudopods). VV (especially the enveloped form of this virus) also utilizes the microtubule network for virus exit from the cell. Similar types of pseudopod structures have also been identified during HIV-1 production in CD4+ T lymphocytes. The actin-based cytoskeleton has also been shown to be important for retroviral morphogenesis, budding, and production, indicating that the two cytoskeletal machineries based on tubulin and actin are both important for viral translocation in the cell.

Nuclear proteins that are involved in transcription and RNA posttranscriptional processing, for example, play downstream roles in replication, and could potentially influence how the virus interacts with host factors in subsequent steps. For instance, nuclear RNA binding proteins initially interact with genomes of retroviruses, mediate its transport, but also potentially influence how the genomes are translated and metabolized. This could be due to multiple functions for a protein, its spatial distribution, or it could serve as a signal for the engagement of some proteins, and a corresponding disengagement of others. The impact of these interactions can have dramatic effects on viral gene expression and as a consequence on pathogenesis. This has been shown in studies examining RNA export, and host proteins such as several hnRNP proteins (A1, A2), and RNA helicases (DDX3, RNA helicase A), proteins that have both nuclear and cytosolic functions during viral replication. The pathway by which retroviral RNA exits the nucleus, via different transportin/exportin proteins, or by CRM-1, influences viral gene expression patterns likely to impact on the spatial segregation of critical virus–host interactions with proteins and limiting membranes required for viral morphogenesis. RNA viruses have the tendency to attract many types of RNA binding proteins and this can impact on host protein acquisition, for instance, of proteins that are involved later in viral assembly or proteins that have function in the encapsidation of viral genomes into virions. Cellular proteins such as nucleolin, hRIP, eukaryotic elongation factor-1α (eEF1α), hnRNP proteins, and Staufen interact with retroviral genomic RNA during some of the major gene expression events (transcription, splicing, translation). Many of these proteins are also found in highly purified viral particles (Table 1). In many ways this would be expected because viral RNAs are trafficked intracellularly like most cellular RNAs. Interestingly, Staufen is found in a complex containing the RNA helicase proteins RHA, DDX1, and DDX3, proteins that are involved in nucleocytoplasmic RNA trafficking. It is part of the cytosolic Gag–ribonucleoprotein (RNP) complex and it is also found in purified virus (Table 3). Staufen's association with influenza virus NS1 in the nucleus, indicates that this RNA trafficking protein plays roles in and out of the nucleus during viral replication in several viruses. Furthermore, this work may indicate a more widespread function of these RNA trafficking proteins during the viral gene expression stages of a variety

of viruses from HIV-1, HSV, adenovirus, and influenza virus in their respective host cells. Live cell imaging of MLV RNA trafficking reveals tight apposition of the viral RNP complex that contains RNA and Gag, to recycling endosomal membranes and a dependence on the microtubule cytoskeleton. It is clear from this recent work that viral RNA translocation requires critical associations between host-cell proteins and the cell's vesicular sorting machinery during viral assembly.

5.2 Viral Protein Egress Mediated Through Vesicular Trafficking Mechanisms

While the process of viral egress is becoming the focus of intense investigation, nowhere have we investigated this process more intensely than in retroviruses. As one of the least understood frontiers in viral replication, research in this area has underscored the importance of the vesicular trafficking machinery in viral protein and RNA egress. A multitude of proteins involved in vesicular sorting mechanisms have been closely associated with viral proteins in several cell types. We will discuss some of these and the cellular compartments to which they belong and the implications for viral replication.

Virion assembly, a phenomenon that was once thought to have been mediated merely by diffusion of components to the plasma membrane to initiate assembly and aggregation, is shaping up to be a highly regulated process requiring the activity of a multitude of cellular proteins and cellular machineries. HIV-1 infection in macrophages plays a determinant role in the establishment, persistence, and pathogenesis of infection. In macrophages, however, it has been appreciated for many years that virus accumulates in intracellular, vacuole-like compartments. Ultrastructural analyses have shown that this intracellular compartment is the late endosome/multivesicular body (MVB). Virus-like structures and several organellar markers of the MVB compartment such as CD63, Lamp-1, and Lamp-2 have been identified ultrastructurally. Many of these proteins are also found in purified virus preparations (Table 1). Gag targeting and assembly in this compartment can occur in macrophages and $CD4^+$ T cells and other epithelial cells (HeLa).

5.3 Virion Assembly: Morphogenesis and Budding

The Gag proteins of several retroviruses harbor late domains that drive viral budding at the MVB. They were so named because virus assembly was arrested at late stages when they were mutated. Several late domain motifs (P(T/S)AP, PPXY, and YP(X)nL) have been identified in enveloped retroviruses such as in EIAV, MLV, MPSV, RsSV, and HTLV-I and they have also been found in ortho- and paramyxoviruses and in Ebola virus. In HIV-1 and SIV, viral budding is mediated by a conserved stretch of proline-rich amino acids with the motif P-T/S-A-P with their p6 domain. These data reveal that these different families of enveloped viruses likely bud using a common mechanism in their respective host cells.

Deletion or mutagenesis of the late domain resulted in a dramatic decrease in viral release and morphologically detectable budding defects. Because viral budding is a phenomenon that occurs on cellular membranes, these morphological data suggested that late domains might interact with the machinery involved in

membrane and vesicle formation. Molecular biology analyses revealed that late domains interacted with cellular class E proteins. The HIV-1 p6 domain for instance is bound by the tumor susceptibility gene 101 protein (Tsg101). This protein is classified as "class E" because its deletion gives rise to abnormally enlarged endosomal membrane compartments that fail to mature to MVB. Tsg101 protein's normal function is to help sort proteins into the MVBs. The HIV-1 Gag-Tsg101 interaction essentially hijacks the host's MVB vesicle formation machinery. This in turn recruits proteins of the 350 kDa endosomal sorting complex required for transport (ESCRT) -1, which regulate MVB biogenesis and cellular vacuolar sorting pathways to membranes of the late endosome (Fig. 3). ESCRT-1 is transiently recruited to limiting membranes of endosomes where it recruits two other complexes (ESCRT-2 and -3) to initiate MVB formation. Other cellular components have been identified to play a role in HIV-1 budding, for instance AIP1, another class E protein, but this host-cell protein binds adjacent LRSLF and YPDL motifs in HIV-1 and EIAV, respectively. It is likely that the over 22 candidate human class E proteins play a role in virus budding but most might impact MVB biogenesis in general rather than directly interacting with viral components.

There are fundamental differences in the site of budding of retroviruses depending on cell type, but these processes are topologically identical in that the membranes must curve and bud away from the cytoplasm. Both $CD4^+$ T cell (plasma membrane) and macrophage (MVB) (Fig. 3) virion budding membranes should contain all the machinery needed for this process. Emphasis is now focused on identifying class E and other cellular factors that may induce curvature in these membranes to influence budding (AIP1 could be one of them). The influence of viral budding could impact on pathogenesis since MVB budding pathway used principally by macrophages could promote directional release of virions. The impact on pathogenesis could also be substantial especially since the tropism of HIV-1 changes with time (from macrophage to T-tropic) and the budding pathway changes as a consequence. Recent work has elucidated the nature of viral protein egress in different cell types.

5.4 Gag–Host Interactions: The Gag Interactome

Group specific antigen or Gag is the major viral structural protein and plays a central role in many aspects of the viral replication cycle. It is generally presented as a polyprotein that is cleaved upon maturation of virions during assembly and viral morphogenesis to give rise to multiple smaller proteins. In retroviruses, these proteins have multiple functions throughout the life cycle. As listed above, these proteins interact with cellular proteins for function. Recent work has focused on late events of retroviral replication, a less well-characterized mesh of virus–cell interactions (see the following). Documenting these interactions is crucial to understanding the process of HIV-1 replication and pathogenesis, and it provides the foundation for the development of new therapeutic and prevention strategies to combat the acquired immunodeficiency syndrome (AIDS). The importance of Gag and the central role of this protein in viral replication were the initiative to create the *HIV-1, Human Protein Interaction Database* (www.ncbi.nlm.nih.gov/RefSeq/HIVInteractions/) by the National Institute

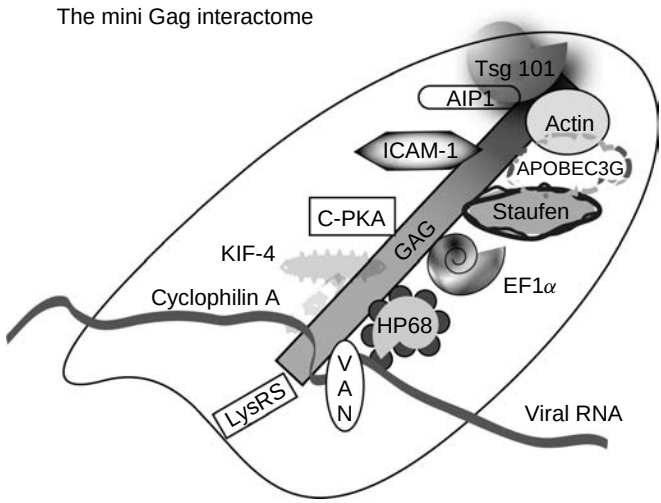

Fig. 7 Some of the host-cell proteins that interact with retroviral Gag protein. This cartoon depicts that Gag interacts with multiple host proteins in addition to the genomic RNA (blue squiggle) throughout the viral life cycle. These interactions can impact on viral egress and assembly (motor protein KIF4A, Actin; HP68, lysyl-tRNA synthetase (LysRS), cyclophilin A), gene expression (EF1α), viral budding (Tsg101, AIP1), early events to influence cell-to-cell transmission and infectivity (ICAM-1), cell restriction to viral infection (APOBEC3G, Cyclophilin A) and signaling, and posttranslational modifications of Gag (catalytic subunit of cAMP-dependent protein kinase, c-PKA). Location of host-cell protein does not indicate region of Gag to which it binds (See color plate p. xxix).

of Allergy and Infectious Diseases of the National Institutes of Health (USA), in which we find an up-to-date and curated list of documented HIV-1-host cell protein interactions. Some of the most recent interactions with Gag have been highlighted earlier and Fig. 7 shows some of these in a cartoon. Table 1 lists a few of these interactions and their possible role in the replication cycle.

5.5 Cellular Proteins Incorporated in the Virion

A partial list of virion-associated host-cell proteins (studied extensively in HIV-1) is provided in Table 1. A few examples of virion-associated host-cell proteins have already been described in the sections earlier (ICAMs, Sect. 1.1), APOBEC3G (Sect. 3.3) INI1 and UNG2 (Sect. 3.7), eEF1α and Staufen (Sect. 5.1), and class E proteins (Sect. 5.3) and their impact on viral replication steps described. These proteins are acquired at various steps of the viral assembly pathway, between the nucleus and the budding membranes. Function can be ascribed to many of these on the basis of biochemical data and genetic data (using RNA interference, RNAi or antisense RNA approaches), and in most cases, the virion association can be explained by specific interactions that occur between viral protein and RNA (or the

viral RNP) components and these host proteins. Earlier reports demonstrate that virion incorporation of host proteins found in the virus envelope may merely reflect the levels of expression at the cell surface in a particular cell type. The expression of several adhesion molecules on the virus envelope (LFA-3, ICAM-1, and MHC-I and II) was studied, and the adhesion molecule profile present on the envelope of primary isolates of HIV-1 was shown to be cell-type dependent. Nevertheless, very recent information on ICAM-1 in HIV-1 demonstrates that its virion incorporation is mediated by a specific interaction with Gag. One of the first proteins to be identified in retroviruses was Cyclophilin A, a peptidyl-prolyl *cis*-isomerase that interacts directly with CA protein of HIV-1 Gag. Its role in viral replication appeared significant since blocking its incorporation by mutagenesis or with the use of cyclosporine A resulted in compromised infectivity. This virus–cell interaction continues to gain attention. It has been shown to impact on viral replication in virus-producing and target cells. These roles include regulation of assembly, decapsidation, reverse transcription and most recently, postentry viral restriction to infection.

While it would seem evident that there would be a limit to the number of proteins that can be packaged into a virion, recent work demonstrates that host protein packaging can be regulated in the virus-producing cells, especially when the host protein is detrimental to virus replication. In cells that are nonpermissive to Vif for example, Vif expression can sequester or bind APOBEC3G, to prevent incorporation into virions and the catastrophic effect on replication as described above (see Sect. 3.3). This example demonstrates the fact that during virus assembly there are additional levels of regulation to influence host-cell protein incorporation that may rely on either viral or host-cell proteins. Both APOBEC3G and Staufen for example are packaged in a Gag-independent manner, but may require the inclusion of viral RNA and a functional RNP assembly complex to take them into the virus. Other documented examples of virion-associated proteins include HO3, a putative histidyl-tRNA synthetase in HIV-1, but not MLV, lysyl-tRNA synthetase in HIV-1, and the nucleolar protein, nucleolin, in MLV. Future work will uncover the impact of host-cell protein packaging in HIV-1 and in other viruses, but it is clear that selective virion incorporation within the viral envelope or inside the virus has a significant impact on viral infectivity and pathogenesis.

6 Cell-to-Cell Viral Transmission

The propagation of viruses within a host organism is known to occur via three main mechanisms. The first and classical mode of entry is by cell-free virions that infect cells via a receptor-mediated mechanism followed by entry into the cytoplasm (see Sect. 1). This might be the least efficient mechanism in that free-floating virions must seek out permissive host cells. The second is via effector cells that engulf or trap virions usually via a receptor-mediated mechanism on C-type lectins and other host cell surface receptors such as DC-SIGN. These cells do not become productively infected but have this capacity to entrap virus particles and mediate the transfer of virus at other sites in the body. This mechanism is believed to contribute to virus spread and to be important for transferring virus to cellular

compartments other than the primary site of contact. The third mechanism of virus propagation is characterized by cell-to-cell transmission of virions. This transmission route likely represents the most efficient means of infection, but is probably the least well characterized. This mechanism has been recently studied in good detail for HTLV-I and HIV-1 and there is some evidence to suggest that the SARS-CoV can also be transmitted via this route of transmission. There are several advantages for the virus to be transmitted this way and these pertain to immunological escape, creation of viral reservoirs, and dissemination of virus throughout the body, to areas, for example, that might be impervious to antiviral therapies.

6.1 The Virological Synapse

Cell-to-cell transmission occurs via the so-called virological synapse (VS) that is formed between virus-harboring or infected cells and an uninfected cell. It is important first to draw on its similarities with the immunological synapse (IS), a well-characterized structure that is formed between T cells, cytotoxic T lymphocytes (CTL), NK T cells, and target cells. The main similarity is the interaction between adhesion molecules (such as ICAM-1) and Talin/LFA-1 between cells. Noteworthy is that immune cells will continually sample each other, infected or not, or in the case of the IS, antigen presenting or not. So the bombardment might appear to be haphazard, suggesting that additional membrane signals derived from a virus-expressing cell (i.e. its envelope) will be important to engage a cell and to initiate cell-to-cell transfer of virus components. Evidence for the involvement of Env comes from very early work, which demonstrated that the addition of recombinant HIV-1 envelope protein to uninfected cells in culture caused a dramatic reorganization of its cognate receptor and coreceptor molecules and the cytoskeleton (actin). Other similarities are the dramatic reorganization of the cytoskeletal machinery, as well as the nature of the cell-to-cell interactions that appear to be transient in nature, as shown by live cell microscopy studies (although this has not been accurately timed in the case for the VS).

Immunofluorescence imaging studies reveal a tight membrane apposition between infected and uninfected T cells. Electron microscopy also reveals the presence of mature virions in this extracellular compartment, suggesting that virions bud out and are concentrated in the VS to favor immediate uptake by the target cell. Recruitment to the site of cell-to-cell contact of many receptors in the uninfected cells such as CD4, and CXCR4/CCR5 coreceptors ensues. Once engaged, a series of cellular events occur. Imaging data suggest that a reorganization of cytoskeletal networks based on actin and tubulin occurs in both the infected and uninfected or target cell. The actin coactivator, Talin, is observed at the site of cell-to-cell contact. The remarkable observation was also made showing that the microtubule network also redistributes to this area. The principal cellular organizing center for RNA, protein, and vesicular trafficking, the MTOC or pericentriole, migrates to the vicinity of the VS. This surprising response, perhaps mediated either by the cell–cell apposition, viral expression, or cellular signaling cascades in the cell, might provide efficient targeting of Gag, viral RNA, and other components to these sites of virion budding. This may be in part an effective strategy to propagate viral

genomes, but also represents a viral transmission route that masks viral antigens from an immune response that can clearly be detrimental to viral replication. While this mode of transmission appears to be the only mode of transmission for HTLV-1, other types of viruses such as HIV-1, HSV1, and maybe SARS-CoV use this transmission route. Cell-to-cell contact of an HSV1-infected cell induces a significant restructuring of cell surface receptors and the engagement of immunoglobulin-like receptors, nectin1 (PRR1/HveC/HIgR), nectin2 (PRR2/HveB), and the mannose-6-phosphate receptors. Because HTLV-I succumbs to a block in cell-free virus infection, it will be interesting to determine the virus–cell interactions that are overcome when HTLV-I cell-to-cell transmission occurs. Cell-to-cell contact and what ensues clearly subvert the normal cellular processes to allow efficient spread of viral particles to many cell types and cellular compartments.

6.2
Dendritic Cell to T-cell Transmission

As the primary site of infection, the epithelial cells that line the vagina are usually the sites of initial HIV-1 virus exposure. This milieu is rich in immature dendritic cells (DC) and upon encountering viral antigens, as they would do following a sexual encounter with a virus-infected individual, virus is trapped within DC. This is mediated via the same DC-SIGN molecules on the surface of the DC and likely, its receptor that is expressed on the surface of virus particles (ICAM). Virions are engulfed in the DCs and these cells migrate to secondary lymphoid tissues, where they have the potential to transmit virus via cell-to-cell transmission to other immune cells. The DC cells have been shown to mediate cell-to-cell transmission to T cells (likely in these secondary tissues), without becoming infected themselves and this would necessarily result in the first steps in infection. Immunofluorescence imaging reveals the formation of a VS at the site of contact between DC and T cells, a similar redistribution of cytoskeletal proteins and receptor molecules described above. Importantly, DC-SIGN represents the means for other pathogens including bacteria, parasites, and emerging viruses such as SARS-CoV to usurp the primary immune defense mechanisms and to disseminate in the body.

6.3
Nanotubes and Viral Transmission

Long protrusions have been demonstrated to form between immune cells following cell-to-cell contact. These have been termed *nanotubes*, and have been visualized by both light and electron microscopy. These are very narrow membrane channels. Complexes have been shown to move by saltatory and directed transport between cells via this tunnel (via channeled microtubules and cellular kinesin motors). Complexes and even cellular organelles up to several microns in diameter are capable of trafficking through these channels. While only shown to occur following IS formation to date, these types of structures can form between immune cells and any other cell type in coculture. This structure could potentially be formed between uninfected and infected cells and could provide an efficient means to transfer viral genetic material and proteins to many uninfected cells at the same time. This could be a means once again for viruses to avoid the immune system. Whether viral transmission occurs via nanotubes as

a consequence of specific virus–cell interactions will be the subject of further experimentation.

7 Pathogenesis Induced by Virus–Cell Interactions

7.1 Virus–Cell Interactions Leading to Oncogenicity

Many viruses induce cancers in animals including humans. The first group of viruses is the DNA tumor viruses of the Adeno-, Polyoma-, and Papillomaviridae families. HPV infection represents a major risk for the development of human cervical carcinomas. Adenoviruses have not been associated with human tumors, but some serotypes transform cells in tissue culture and induce tumors in rodents. PyV have a host restriction. In permissive cells, they induce a lytic cycle, whereas in nonpermissive cells, viral entry results in abortive infection or cell transformation. Although the exact origin of SV40 in some human tumors is still under investigation, JC and BK polyomaviruses are human pathogens with an oncogenic potential, but their association with human cancers may be the result of a combination with other factors. Members of the Herpesviridae, EBV and HHV8 can also induce tumors in specific conditions. Retroviruses constitute a large group of RNA tumor viruses, although not all retroviruses cause cancers. Several retroviruses induce different types of cancers in animals, and HTLV is the first discovered human retrovirus that induces cancer. The mechanisms used by these viruses are quite complex, involving specific genes that contribute to cell-cycle deregulation by different mechanisms.

7.1.1 Cellular Genes Targeted by Viral Oncoproteins

The cellular retinoblastoma susceptibility protein (pRb) and the human tumor suppressor p53 play major roles in the cell cycle and are the target of viral oncoproteins. The Rb family "pocket proteins" negatively regulate progression from G0 to G1 and progression to S phase, thus controlling cell-cycle progression. pRb sequesters the E2F family of transcription factors, therefore preventing E2F-dependent genes like c-fos and c-myc to be activated. In addition, pRb recruits chromatin remodeling factors, histone deacetylases (HDAC) and Grg1/Brm1 to further repress E2F-dependent transcription. p53 protein is a tumor suppressor encoded by a gene whose disruption causes 50 to 55% of all human cancers. p53 has a key role in cell-cycle arrest and apoptosis induction when DNA becomes extensively damaged. In normal cells, p53 is a short-lived protein maintained at very low levels bound to its inhibitor human double minute 2 protein (HDM2). p53 production is activated in response to various stimuli including DNA damage, oncogene expression, or viral expression. It translocates to the nucleus upon activation and stimulates the expression of genes implicated in cell-cycle arrest, DNA repair, or apoptosis. These two mechanisms overlap when p53 production increases and stimulates transcription of CDK inhibitors, resulting in decreased pRb phosphorylation and inactive E2F that remains bound to pRb. pRb and p53 constitute major targets of oncoproteins of DNA viruses, but also contribute to oncogenesis induced by retroviruses.

7.1.2 Viral Oncoproteins

SV40, JC, and BK polyomaviruses all encode two proteins that contribute to oncogenicity, the LT and the small T (ST)

Ag; LT is sufficient to transform cells. LT stimulates the cell cycle by binding to several proteins that control cell-cycle progression and apoptosis. LT binds to pRb and p53, and these interactions are major factors of the oncogenic transformation. LT disrupts pRb-E2F interactions, allowing E2F activation of c-fos and c-myc that promote G1/S phase transition. LT binds to and inactivates p53, thus preventing inhibition of cell cycle and apoptosis. Another interaction of LT is with the insulin receptor substrate (IRS-1), which causes its translocation to the nucleus and may affect DNA repair and chromosomal stability. LT also binds to β-catenin and enhances expression of c-myc and Cyclin D1. ST cooperates with LT, but also binds to protein phosphatase 2A and inhibits its activity. This inhibition results in an increased transcription of Cyclins D and A, and in the downregulation of the cyclin kinase inhibitor, p27, to contribute to cell transformation.

Adenovirus E1A, E1B, and E4 gene products contribute to cell-cycle deregulation. The two forms of E1A protein inactivate the pRb checkpoint, allowing S phase progression. E1A also interacts with chromatin remodeling factors, including the HAT CBP/p300 and HDACs. These interactions affect transcription of several cellular and viral genes that contribute to cell-cycle regulation, and also increase p53 production and stabilization. To counteract its activity, the E1B and E4 gene products inactivate p53 to further prevent cell-cycle inhibition and apoptosis.

HPV E6 and E7 are oncoproteins that both contribute to the induction, the progression, and the severity of human cancers. E6 binds to p53, and this interaction results in the degradation of p53 via the ubiquitin proteolytic pathway. This activity results in a strong inhibition of apoptosis. E6 also activates the expression of the telomerase enzyme, thereby decreasing the stability of chromosome ends. It also associates with CBP/p300, which reduces the transcriptional trans-activation of p53. E7 associates with Rb and HDAC and thereby releases E2F that increases cell-cycle progression.

Herpes viruses can progress to a lytic or to a latent phase and some members induce cancers. HCMV, HHV-6, EBV, and HHV-8 all encode proteins that inhibit the p53 transcription pathway, promoting oncogenesis. The clear implication of γ-herpesviruses EBV and HHV-8 in human cancers is linked to genetic abnormalities but has additional molecular pathways to induce cell-cycle deregulation. HHV-8, responsible for Kaposi's sarcoma, has at least three latent genes encoding proteins that deregulate cell cycle. The viral cyclin (vCyc) inactivates pRb. LANA-1 interacts with and represses p53 and pRb; it also interacts with H-ras. vFLIP inhibits Fas-induced apoptosis and activates NF-κB. EBV genes involved in oncogenesis are the EBNA1 nuclear antigen, the latent membrane proteins (LMPs 1, 2A, and 2B) and the BamHIA transcripts.

Many retroviruses are oncogenic, inducing mainly sarcomas and hematopoietic tumors. The mechanism of tumor induction by retroviruses involves viral oncogenes (v-onc), which are derived from cellular genes called *proto-oncogenes (c-onc)*. Retroviruses that carry v-onc genes transform cells *in vitro*, but are generally replication-defective because some or all the genes required for viral replication were deleted when the oncogene was captured. Therefore, virus production can only be obtained with a helper virus

that provides the gene products for replication and virion packaging. More than 30 v-onc genes and their cellular proto-oncogenes (c-onc) homologs have been discovered including src, ras, raf, jun, fos, and myc among the best-studied ones. Many tumors are also induced by retroviruses that lack oncogenes. In this case, the oncogenic process is generally long and requires provirus integration in the vicinity of a proto-oncogene, a mechanism called *proviral insertional mutagenesis*. The overexpression of the c-onc promotes cell growth deregulation. v-onc genes have a large variety of activities and are classified in families on the basis of the function of the protein encoded by the corresponding c-onc gene. They include growth factors, growth factor receptors, protein tyrosine kinases, adaptor signaling proteins, G proteins, serine/threonine kinases, and transcription factors. v-onc transcription factors are often modified compared to their cellular counterpart and induce activation or repression of several genes leading to cellular dysfunction. Many of the v-onc genes cooperate, and insertional mutagenesis may target several loci. In some retrovirus infections, tumor suppressor genes may also be involved, indicating that retrovirus-induced tumors are multifactor processes. HTLV-I is different from other retroviruses because its trans-activator Tax is its major oncoprotein, and it lacks a cellular homolog. Tax has numerous effects leading to cell-cycle deregulation. It activates the expression of Cyclin D, cdk2, 4, and 6 and represses most cdk inhibitors, which promotes cell-cycle progression by allowing constitutive activation of cyclin–cdk complexes. Tax also mediates repression of DNA repair by inhibition of DNA polymerase β and p53 and trans-activation of PCNA.

7.2 Viral Proteins Contributing to Cell Death

Viruses and cells use a wide variety of strategies to affect or maintain the balance between growth and death of infected cells. Apoptosis is a programmed cell death for the elimination of unwanted cells, including virus-infected cells. Two main signaling pathways are used by the cell to reach apoptosis. The death receptor pathway, which includes CD95/Fas and the tumor necrosis factor receptor (TNFR)1 and uses the adaptor molecule Fas associated death domain (FADD), leads to the activation of caspase 8 and 10. The mitochondrial pathway leads to the activation of caspase 9, which forms the apoptosome with cytochrome *c* and Apaf-1. Bcl-2, an integral mitochondrial membrane protein, prevents apoptosis, whereas other members of the Bcl-2 family such as Bax have proapoptotic activities. As described in the previous section, p53 induces growth arrest and has apoptotic properties.

7.2.1 Anti-apoptotic Viral Genes

Virus infection induces the expression of IFN and in turn IFN activates genes that play a role in apoptosis. These include TNF-related apoptosis-inducing ligand (TRAIL), PKR, IRF-1, and PML. To maintain host-cell survival and to ensure viral replication, viruses have evolved strategies to prevent cell death. As examples, the herpesvirus vFLIP, the poxvirus CrmA, and the insect baculovirus p35 encode proteins that inhibit caspase 8 and 3 activation. The γ-herpesviruses encode BHRF1 and KSBcl-2 that inhibit the mitochondrial apoptotic process. HSV-1 LANA, adenovirus E1B, HPV E6, and polyomavirus LT inhibit p53-mediated apoptosis. EBV, LMP1, and HTLV-I Tax may also stimulate the expression of

cell survival factors by activating the NF-κB pathway.

7.2.2 Proapoptotic Viral Genes

Some viruses have an advantage to promote cell-cycle arrest for a specific step in their replication cycle or to induce a lytic phase. For these purposes, they encode proteins with proapoptotic properties. γ-herpesviruses are able to arrest or modulate cell-cycle progression during the lytic phase. EBV Zta and the HHV-8 K-bZIP mediate these changes through increased c/EBPα protein production and interaction with this protein production. c/EBPα upregulates the expression of $p21^{WAF1}$, also involved in cell-cycle regulation. Zta and K-bZIP use another route to halt the cell cycle; they both interact with p53. Zta was shown to upregulate p53 and $p27^{KIP1}$, which contributes to cell-cycle arrest and apoptosis, but the mechanism is likely more complex, involving the modulation of several other genes. HIV-1 has evolved several ways to mediate apoptosis. Its Vpr protein induces G2 cell-cycle arrest in proliferating T lymphocytes. HIV-1 protease directly cleaves and activates procaspase 8. HIV-1 envelope gp120 induces apoptotic cell death of lymphocytes, neurons, and myocardiocytes by interaction with surface receptors. The HPV E2 protein not only induces cell-cycle arrest by blocking the transcription of E6 and E7 genes but also induces apoptosis through the caspase 8 pathway. In addition, E2 is cleaved by caspases during apoptosis, which enhances its apoptotic activity.

7.3 Viral Proteins that Affect Host Immunity

To achieve a successful invasion of their host, viruses have developed mechanisms to counteract the cell immune response. In addition to interfering with the innate immune pathways described in Sect. 2, viruses interfere with adaptive immunity by different mechanisms involving B and T lymphocytes. Examples that have been studied at the molecular level include chemokine mimicry and decreasing MHC-I/HLA expression, but viruses subvert many other cellular pathways.

7.3.1 Virus-encoded Cytokine and Cytokine Receptor Homologs

Many viruses, especially those in the herpes- and poxvirus families encode cytokine and chemokine variants, as well as soluble or membrane-bound cytokine and chemokine receptors. Viral (v)IL-6 encoded by HHV8 and vIL-10 encoded by HCMV and EBV modify the host immune response and lead to the activation and growth of infected cells, thereby significantly contributing to pathogenesis. Similarly, soluble cytokine receptor homologs or original viral chemokine binding proteins serve as scavengers that interfere with inflammation, immune activation, and the migration of antiviral immune effector cells. The biological role of virus-encoded chemokine receptors, such as the HHV8-encoded ORF74, remains somewhat enigmatic. Although ORF74 is constitutively active and sufficient for the development of Kaposi's sarcoma-like lesions, other virus-encoded chemokine receptors, such as the syntenic *H. saimiri*-encoded ECRF3, are not, suggesting different functions for different receptors. Moreover, the finding that HHV-8 ORF74 is expressed in the lytic and not the latent phase of HHV-8 replication makes it difficult to conceive that the receptor plays a major role in the transformation of latently HHV-8 infected cells. Like ORF74, the HCMV-encoded US28 is constitutively active, but not transforming, and has been proposed to serve as a

chemokine scavenger. However, the receptor also induces chemokinesis in vascular smooth muscle cells, which is consistent with the epidemiological association of HCMV infection with atherosclerosis. Hence, although the biological function of virus-encoded immune modulators remains still partly elusive, it is clear that they are major determinants for virus pathogenesis.

7.3.2 Inhibitors of MHC/HLA Molecules

The MHC-I/HLA antigen presentation pathway plays a crucial role in immunity to viruses. MHC-I molecules are expressed on the cell surface and present peptides derived from intracellular proteins to CTLs, which then eliminate the infected cells. The MHC-I antigen presentation pathway involves: (1) protein processing in the cytosol by the proteasome; (2) peptide translocation into the ER lumen by the transporter associated with antigen processing (TAP); (3) MHC-I folding and assembling in the ER lumen with ER proteins; (4) association between the MHC-I complex with TAP and peptide binding; and (5) transport of peptide loaded MHC-I through the secretory pathway to the plasma membrane. Many viruses have evolved mechanisms to inhibit the MHC-I pathway, thus enabling virus-infected cells to escape CTL lysis. One first mechanism is to decrease the amount of MHC-I molecules either by inhibiting their synthesis, by sequestration, or by promoting their degradation. Herpes viruses represent a family in which most, if not all, members encode proteins that inhibit MHC-I Ag presentation. In HSV, ICP27 mediated cell shut-off and *vhs*-induced mRNA decay contribute to the loss of MHC-I mRNA and protein. HCMV UL18 and UL40 suppress MHC-I expression. Adenovirus E1A inhibits MHC-I expression, whereas E1B inhibits transport and expression. Bovine papillomavirus and HPV E5 inhibit expression of MHC-I/HLA by retention in the Golgi apparatus, whereas HIV-1 Nef promotes their sequestration in the trans-Golgi network. HCMV US2 and US11 promote ubiquitination of the MHC-I heavy chain and its degradation through the proteasome pathway. HHV7 U21 binds to MHC-I molecules in the ER and targets them to the lysosome for degradation. HHV8 K3 and K5 proteins cause a rapid downregulation of MHC-I molecules from the plasma membrane by clathrin-dependent endocytosis. Another mechanism is to turn off the mechanism of peptide presentation. HSV ICP47 acts as a competitive inhibitor of peptide binding to TAP, whereas HCMV US6 interacts with TAP and inhibits peptide translocation. The adenovirus gene product E3/19K (E19) binds to TAP and prevents TAP/MHC-I association. HCMV US3 and US10 inhibit the export of peptides loaded MHC-I from the ER. These viral mechanisms contribute to viral evasion from the cell immune system, ensure prolonged survival in the infected host and facilitate virus transmission and pathogenesis.

7.3.3 HIV-1 Example: Downregulation of CD4 by Several Mechanisms

HIV induces a severe immunodeficiency due to several mechanisms all of which contribute to this pathogenesis. In addition to decreasing MHC-I expression at the cell surface by its Nef protein, HIV induces the degradation of its receptor CD4. A decline of CD4+ T lymphocytes, which have a helper T-cell function and a cytotoxic capacity, contribute to decreased cellular and humoral immunity. HIV-1 encodes three proteins that contribute to CD4 downregulation. The Nef protein decreases the number of CD4 molecules

at the cell surface by accelerating their clathrin-mediated endocytosis that target them to the lysosome for degradation. The Env gp160 precursor has a high affinity for newly synthesized CD4 molecules and retains them in the ER. The Vpu protein binds to the cytoplasmic tail of CD4 and to β-TRCP/Skp1 that targets the complex to the proteasome degradation pathway. The combination of these mechanisms contributes to increased virus spread and increased viral pathogenicity.

7.4 Viral-induced Neurotoxicity

7.4.1 Viruses Inducing Neurotoxicity

Several viruses induce neurotoxicity and several brain disorders different from brain cancers. The neurotropic polyomavirus JCV is the etiologic agent of the fatal demyelinating disease, progressive multifocal leukoencephalopathy. Whereas its LT induces brain tumors in nonpermissive cells, JVC-infected oligodendrocytes are permissive for viral lytic infection and are destroyed, leading to demyelination. Virus-induced demyelination can also be induced by neurotropic strains of mouse hepatitis virus, Theiler's virus and Semliki Forest virus (SFV). Several flaviviruses, including West Nile virus (WNV), tick-borne encephalitis virus (TBEV) and the Japanese encephalitis also cause neuropathies. Herpes viruses and rabies that infect neurons, poliovirus that infects anterior horn cells have also been implicated in neuropathogenesis. For most viruses, the mechanism of neurotropism and neuropathy is not fully elucidated. HIV-1-associated dementia has been well studied and has the particularity to induce neuronal cell death with no or rare neuronal infection.

7.4.2 Example of HIV-associated Dementia

HIV-1 infects the brain via infected macrophage/microglial cells that cross the blood brain barrier. Although this infection occurs early after HIV infection, HIV-associated dementia (HIVD) occurs late in the disease in about 20% of infected individuals and is characterized by several cellular disorders including macrophage activation and increased death of astrocytes and neurons. Productive infection of microglia leads to the release of HIV virions and soluble Tat protein. When the virion enters a new cell, the envelope may be shed outside and promotes independent toxicity. Viral Tat and envelope have been implicated in the neurotoxic process of HIV. Astrocytes that are also infected by HIV, produce a very limited amount of virus, but release cytokines that participate to the HIVD. Tat stimulates microglia to synthesize neurotoxic molecules including proinflammatory cytokines and free radicals. Envelope proteins retain the ability to interact with the cognate HIV receptors. The interaction of soluble HIV-1 gp120 does not require the presence of CD4, at least on neuronal cells. This interaction leads to CXCR4 signaling, and may result in neuronal apoptosis, in the same was as the natural CXCR4 ligand, the chemokine SDF-1. The gp120 interaction with CXCR4 on other cells of the nervous system such as astrocytes induces signaling that leads to activation of these targets, which, in turn, secrete neurotoxic products. Other lentiviruses like FIV share some functional features in envelope mediated neurological disorders. HIV-1 and FIV env and HIV-1 Tat induce metalloproteinase (MMP)-2 in macrophages, which also participates in neuronal death. Overall, the productive infection of microglial cells, the nonproductive infection

of astrocytes, the secretion of neurotoxic molecules, the viral Tat and Env proteins, all lead to neuronal and astrocytic cell death and contribute to HIVD.

8
Perspectives on Therapies Based on Virus–Cell Interactions

8.1
Virus Entry as an Antiviral Target

Virus entry as an antiviral target holds several specific advantages as well as drawbacks as compared to viral enzyme drug targets. Blocking entry by viral receptor ligands that successfully compete with the envelope for the available binding sites means that a protein of the host, and not the virus, is targeted. This strategy is currently pursued with the development of inhibitors of HIV-1 entry into target cells that block the chemokine entry coreceptors. Being encoded by the host genome, this kind of target is invariable and will therefore not evolve to give rise to drug resistant variants, a mechanism that is the major cause of failure of therapies using HIV-1 RT and PR inhibitors. Other viral strategies to achieve resistance can however not be prevented and may constitute an obstacle for the success of coreceptor-targeting drugs: the virus may bypass the requirement for a given coreceptor by using another, alternative receptor. For example, it will be important to evaluate whether the use of CCR5 antagonists will induce a coreceptor switch in the virus, provoking the emergence of CXCR4-using strains. Another conceptual problem of targeting host molecules rather than viral proteins is that this kind of therapy often interferes with the natural functions of this host protein, thereby inducing intolerable side effects. This point clearly depends on the given target, as illustrated by the HIV-1 chemokine coreceptors. CCR5 is a safe target, since a homozygous functional deletion of the CCR5 gene exists with relatively high frequencies in some human populations without readily detectable phenotype. However, CXCR4 is a housekeeping receptor with fundamental importance. Genetic ablation of either of the genes coding for CXCR4 or its ligand, stromal cell-derived factor 1 (SDF-1) in mice is perinatally lethal. Although the importance of CXCR4 in human adults may be somewhat lower, long-term treatment with CXCR4 ligands seems difficult due to severe side effects. Another approach to interfere with viral entry is an HIV-1 fusion inhibitor that targets the conformational rearrangements of the viral fusion peptide. The first of these compounds, Enfuvirtide, is already used in the clinic. Thus far, no other example of successful pharmacologic inhibition of the entry process of viruses other than HIV is known. Whether entry inhibition may prove an antiviral principle of general interest that is useful for the therapy of different viral infections, remains to be investigated.

8.2
Antiviral Action of Interferons

Various forms of IFNs α and β are currently being used in therapeutic treatment of viral diseases including Hepatitis B and C. HPV diseases are also treated with IFNs α, β, and γ with varied success: from no response to complete effectiveness. IFN treatment has been attempted against HIV infection but multiple clinical trials have shown it to be completely ineffective. Considering the multiple mechanisms and

counter mechanisms of the virus and the host described above, it is clear that the outcome of IFN treatment will be dependent on both of these responses. Cells infected by different strains from the same virus or family of viruses can respond differently to IFN treatment. As an example, in the Poxviruses, a treatment against MCV infection is currently being tested with an immunomodulator that induces IFN production, because MCV does not encode for E3L and K3L homologs and cannot inhibit PKR by this means. The effectiveness of IFNs against hepatitis C infection is about 40%, and correlations have been made with the activity of NS5A against PKR in some studies but not in others. A more complete knowledge of virus–host interactions will help the design of combination therapies using compounds that will prevent the inhibition of the IFN response and thereby will lead to a more effective IFN therapy.

8.3 Inhibition of Viral Gene Expression

The targeting of viral gene expression steps from RNA synthesis to translation has been attempted with various small molecules, some of which are currently being used in therapeutics. Recent improvement in our understanding of virus–cell interactions have allowed rational design of strategies to disrupt vital interactions by competitive peptides, transdominant molecules, or the use of RNA-based technologies to either degrade mRNAs or prevent their translation. Finding molecules to disrupt virus–cell interactions can be first achieved by the search of inactive mimics of cellular factors. These can be small peptides that can bind competitively to a viral target without mediating the cellular function. They can also be transdominant mutant proteins that bind to the viral protein or RNA but have no function. Examples have been developed for HIV-1 with peptides that can block the TAR RNA to prevent trans-activation, and with the use of Rev mutants that have a transdominant activity by a mutation in the NES site. Large-scale screening for inactivating peptides can be achieved by using phage display libraries in assays that target a specific function. Recent examples of viruses that can be inhibited by protein fragments using this technology include hepatitis A virus, HCV, HIV, and rotavirus.

RNA-based technologies include antisense, molecular decoys, ribozymes, and small-interfering (si)RNAs. They all use RNA molecules that target the viral RNA or mRNAs coding for cellular proteins required for viral replication. RNA processing of the RNA duplexes involves cellular functions that are used to cleave the RNA or block its translation. Antisense technology uses exogenous ssRNA (or DNA) that will hybridize to the viral RNA to form duplexes that will then be cleaved by the cellular RNAse H and degraded. Alternative and more potent technology uses the coupling of these antisense RNA molecules to a molecule that sterically inhibits a function. In this case, the RNA targets the 5′ end of the mRNA and the steric block prevents its translation. Antisense nucleic acids, also called *aptamers*, have been used to block HIV, HCV, HBV, RSV, and HPV, but the future of their therapeutic use remains to be evaluated. RNA molecular decoys use structured RNAs that mimic a viral RNA target to sequester viral and cellular factors. They have been developed for the HIV-1 TAR RNA to prevent the Tat-CycT1-CDK9 complex formation and the trans-activation mechanism. Ribozymes (Rz) are catalytic RNA molecules

that combine an antisense RNA to an enzymatic activity that specifically cleaves the targeted RNA. Various forms include hairpin, hammerhead, RNAse P, and HDV (δ) Rz. Their *in vitro* activity indicates that they act independently of any cellular factor, but some variations between *in vitro* and *in vivo* functions, and the very good adaptation of the δRz to the cell environment suggest that cell factors may improve their efficacy in some cases. They have been successfully used to inhibit HBV, HCV, hepatitis E virus, herpesviruses, HIV, HPV, influenza, and rhinovirus.

RNAi allows the targeting of specific genes to explore their function during viral replication. 21–23-bp siRNAs are transfected into mammalian cells to induce the degradation of the homologous mRNA by RNAi. The activity of siRNAs depends on a cellular enzyme, Dicer that associates in an RNA-induced silencing complex (RISC) leading to dsRNA cleavage at the site of hybridization. The list of viruses that have been targeted is growing very rapidly and include HBV, HPV, several herpesviruses, and the polyomavirus JC among the DNA viruses. The targeted RNA viruses include the picornaviruses (poliovirus, rhinovirus, FMDV, HAV, enterovirus), the flaviviruses (HCV, Dengue, WNV), the togavirus, SFV, the SARS-CoV, rotaviruses, the paramyxoviruses RSV, and the orthomyxovirus influenza. Retroviruses HIV-1, HTLV-I, and RsSV have also been successfully targeted. Cellular factors required for viral replication have also been inhibited by RNAi. Decreased HIV replication can be achieved by inhibiting CD4, CCR5, and CXCR4 coreceptors, NFκB, CycT, CDK9, Staufen, Cyclophilin A, Lys-tRNA synthetase, PARP, SPT5, and TRBP. Inhibition of HCV can be obtained by decreasing RNA helicase p68, La, PTB, eIF2Bγ, and hVAP33 levels. A decrease of La can also inhibit HBV and poliovirus replication.

Some of these RNA-based methods, such as siRNAs against influenza and FMDV have been tested successfully in animals, whereas targets against HIV, HCV, and HPV are expressed on viral vectors for gene therapy approaches to treat or prevent infection. *In vivo* assays using RNAi-based strategies, only three years after the discovery of function in mammalian cells, show promise of this therapeutic intervention.

8.4 Inhibition of Assembly

At the forefront of basic research are the efforts placed on the development of new antiviral therapies. Therapies targeting viral enzymes and viral protein function have taken a front seat in the armamentarium against several viruses. Effectiveness is a concern, while escape mutants or viral resistance to drug therapies represent the biggest problem especially for HIV-1. Strategies that include state-of-the-art technologies such as RNAi have met with good success, but again viruses including HIV and poliovirus have managed to elude these site-specific targets by mutagenesis. Targeting of later events such as viral assembly is really an area of intense research with the basic cellular machinery only recently uncovered. One example relied on small molecules to target and block Gag function (i.e. CAP-1, CAP-2 compounds for HIV-1) but have not been touted as inhibitors of specific virus–cell interactions *per se*.

Another emerging and fast-moving field is the cellular mechanisms that HIV and other viruses co-opt to get themselves out of the cell. This includes multiple proteins and interdependent complexes that

bifurcate with other regulatory networks such as limiting membrane compartments and ubiquitination machineries. It would seem that if we were to block the virus from getting out, we would have the key to a successful therapy. Targeting vesicular trafficking machinery, including the cytoskeletal elements on which vesicles track, may create a hostile environment for the cell, but if it's to block viral spread it's worth it!

8.5 Viruses as Therapeutic Agents

Another aspect of virus–host interactions in therapeutics is the use of viruses as delivery vectors or therapeutic agents. Viruses with no pathogenic potential or viruses in which the pathogenic determinants have been removed can be used to deliver therapeutic DNA or RNA into cells by gene therapy. For this purpose, retroviruses, adenoviruses, adeno-associated virus (AAV), vaccinia and herpes-derived vectors have been used or are currently in clinical trials. To improve safety, most modifications of these vectors include reducing the risk of insertional mutagenesis and directing the expression of therapeutic genes to specific tissues. Some of these vectors and other viruses have been developed to replicate in and to specifically kill cancer cells. Because of this property, they have been called *oncolytic viruses*. Adenoviruses in which E1B is deleted replicate only in cells with an inactivated p53 gene and cause cell lysis, but do not replicate in normal cells. HSV deleted for the ICP34.5 gene replicate better in cancer cells than in the adjacent tissue. Reovirus is nonpathogenic in humans but has a distinctive ability to replicate in cancer cells with an activated Ras pathway. Sindbis virus has a natural tropism for tumor cells and modified SFV can stimulate immunity to tumor-associated antigens and induce apoptosis to destroy tumor cells. An attenuated strain of Newcastle disease virus and VSV mutants selectively lyses tumor cells versus normal cells based on tumor-specific defects in the interferon-mediated antiviral response. Several oncolytic viruses have been tested in animal models and are currently assayed in clinical trial against various cancers and hold strong promises as new therapeutic agents.

9 Conclusion

The very low mutation rate of the host genome and the *in vivo* effectiveness demonstrated in some strategies provide strong justification to pursue the study of virus–cell interactions with the principal aim to uncover effective therapeutic targets to cure or treat infectious diseases. It is definite that should a particular virus–cell interaction be interrupted by targeting cellular protein function, rather than viral protein function, the question of resistance will not be a major limitation to the application of this knowledge to a therapy. The use of oncolytic viruses to treat cancers and other disorders has already demonstrated that the study of virus–host interactions will be beneficial to human health.

Acknowledgments

The authors apologize to readers for all the original research articles that could not be cited because of the space limitation and invite them to refer to appropriate reviews. Special thanks go to Drs. Nicholas Acheson (McGill University, Montréal, Canada) and

David Ott (National Cancer Institute, NIH, USA) for their extensive critical analysis and corrections. We also wish to thank the members of our laboratories, Drs. Levon Abrahamyan and Sébastien Lainé for their critical reading. The work done in the authors' laboratories is supported by the Canadian Institutes of Health Research (CIHR), the Fonds de la Recherche en Santé du Québec (FRSQ), the Canadian Foundation for Innovation (CFI) and the Canadian Foundation for AIDS Research (CANFAR). Andrew J Mouland is a CIHR new investigator; Nikolaus Heveker and Anne Gatignol are research scientists from the FRSQ.

See also Cytokines: Interleukins; Innate Immunity; Molecular Mediators: Cytokines; Nucleic Acid Packaging of RNA Viruses; Pathogens: Innate Immune Reponses; Retrovirus Integration, Biological and Biochemical Machineries for.

Bibliography

Books and Reviews

Akkina, R., Banerjea, A., Bai, J., Anderson, J., Li, M.J., Rossi, J. (2003) siRNAs, ribozymes and RNA decoys in modeling stem cell-based gene therapy for HIV/AIDS, *Anticancer Res.* **23**, 1997–2005.

Amara, A., Littman, D.R. (2003) After hrs with HIV, *J. Cell Biol.* **162**, 371–375.

Bannwarth, S., Gatignol, A. (2005) HIV-1 TAR RNA: the target of molecular interactions between the virus and its host, *Curr. HIV Res.* **3**, 61–71.

Barber, G.N. (2001) Host defense, viruses and apoptosis, *Cell Death Differ.* **8**, 113–126.

Bedard, K.M., Semler, B.L. (2004) Regulation of picornavirus gene expression, *Microbes Infect.* **6**, 702–713.

Berger, E.A., Murphy, P.M., Farber, J.M. (1999) Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease, *Annu. Rev. Immunol.* **17**, 657–700.

Bieniasz, P.D. (2004) Intrinsic immunity: a frontline defense against viral attack, *Nat. Immunol.* **5**, 1109–1115.

Bose, S., Banerjee, A.K. (2003) Innate immune response against nonsegmented negative strand RNA viruses, *J. Interferon Cytokine Res.* **23**, 401–412.

Brinton, M.A. (2002) The molecular biology of West Nile Virus: a new invader of the western hemisphere, *Annu. Rev. Microbiol.* **56**, 371–402.

Broyles, S.S. (2003) Vaccinia virus transcription, *J. Gen. Virol.* **84**, 2293–2303.

Bukrinsky, M.I., Haffar, O.K. (1999) HIV-1 nuclear import: in search of a leader, *Front. Biosci.* **4**, D772–D781.

Castedo, M., Perfettini, J.L., Andreau, K., Roumier, T., Piacentini, M., Kroemer, G. (2003) Mitochondrial apoptosis induced by the HIV-1 envelope, *Ann. N. Y. Acad. Sci.* **1010**, 19–28.

Cathomas, G. (2003) Kaposi's sarcoma-associated herpesvirus (KSHV)/human herpesvirus 8 (HHV-8) as a tumour virus, *Herpes* **10**, 72–77.

Chan, D.C., Kim, P.S. (1998) HIV entry and its inhibition, *Cell* **93**, 681–684.

Collot-Teixeira, S., Bass, J., Denis, F., Ranger-Rogez, S. (2004) Human tumor suppressor p53 and DNA viruses, *Rev. Med. Virol.* **14**, 301–319.

Compton, T. (2004) Receptors and immune sensors: the complex entry path of human cytomegalovirus, *Trends Cell Biol.* **14**, 5–8.

Cullen, B.R. (2003) Nuclear mRNA export: insights from virology, *Trends Biochem. Sci.* **28**, 419–424.

Derse, D., Heidecker, G. (2003) Virology. Forced entry–or does HTLV-I have the key?, *Science* **299**, 1670–1671.

Dlamini, Z., Mbita, Z., Zungu, M. (2004) Genealogy, expression, and molecular mechanisms in apoptosis, *Pharmacol. Ther.* **101**, 1–15.

Garcia-Sastre, A. (2004) Identification and characterization of viral antagonists of type I interferon in negative-strand RNA viruses, *Curr. Top. Microbiol. Immunol.* **283**, 249–280.

Gatza, M.L., Watt, J.C., Marriott, S.J. (2003) Cellular transformation by the HTLV-I tax

protein, a jack-of-all-trades, *Oncogene* **22**, 5141–5149.

Goff, S.P. (2004) Genetic reprogramming by retroviruses: enhanced suppression of translational termination, *Cell Cycle* **3**, 123–125.

Goodbourn, S., Didcock, L., Randall, R.E. (2000) Interferons: cell signalling, immune modulation, antiviral response and virus countermeasures, *J. Gen. Virol.* **81**, 2341–2364.

Gorry, P.R., Ong, C., Thorpe, J., Bannwarth, S., Thompson, K.A., Gatignol, A., Wesselingh, S.L., Purcell, D.F.J. (2003) Astrocyte infection by HIV-1: mechanisms of restricted virus replication, and role in the pathogenesis of HIV-1-associated dementia, *Curr. HIV Res.* **1**, 463–473.

Grandvaux, N., tenOever, B.R., Servant, M.J., Hiscott, J. (2002) The interferon antiviral response: from viral invasion to evasion, *Curr. Opin. Infect. Dis.* **15**, 259–267.

Greene, W.C., Peterlin, B.M. (2002) Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy, *Nat. Med.* **8**, 673–680.

Haasnoot, J., Cupac, D., Berkhout, B. (2003) Inhibition of virus replication by RNA interference, *J. Biomed. Sci.* **10**, 607–616.

Harris, R.S., Liddament, M.T. (2004) Retroviral restriction by APOBEC proteins, *Nat. Rev. Immunol.* **4**, 868–877.

Helt, A.M., Galloway, D.A. (2003) Mechanisms by which DNA tumor virus oncoproteins target the Rb family of pocket proteins, *Carcinogenesis* **24**, 159–169.

Hewitt, E.W. (2003) The MHC class I antigen presentation pathway: strategies for viral immune evasion, *Immunology* **110**, 163–169.

Hiscott, J., Grandvaux, N., Sharma, S., Tenoever, B.R., Servant, M.J., Lin, R. (2003) Convergence of the NF-kappaB and interferon signaling pathways in the regulation of antiviral defense and apoptosis, *Ann. N. Y. Acad. Sci.* **1010**, 237–248.

Hui, E.K. (2002) Virion-associated protein kinases, *Cell. Mol. Life Sci.* **59**, 920–931.

Jayaram, H., Estes, M.K., Prasad, B.V. (2004) Emerging themes in rotavirus cell entry, genome organization, transcription and replication, *Virus Res.* **101**, 67–81.

Jeang, K.T. (2001) Functional activities of the human T-cell leukemia virus type I Tax oncoprotein: cellular signaling through NF-kappa B, *Cytokine Growth Factor Rev.* **12**, 207–217.

Johnston, J.B., McFadden, G. (2003) Poxvirus immunomodulatory strategies: current perspectives, *J. Virol.* **77**, 6093–6100.

Katze, M.G., He, Y., Gale, M. (2002) Viruses and interferon: a fight for supremacy, *Nat. Rev. Immunol.* **2**, 675–687.

Kean, K.M. (2003) The role of mRNA 5′-noncoding and 3′-end sequences on 40S ribosomal subunit recruitment, and how RNA viruses successfully compete with cellular mRNAs to ensure their own protein synthesis, *Biol. Cell* **95**, 129–139.

Khan, A.U., Lal, S.K. (2003) Ribozymes: a modern tool in medicine, *J. Biomed. Sci.* **10**, 457–467.

Kolakofsky, D., Le Mercier, P., Iseni, F., Garcin, D. (2004) Viral DNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis, *Virology* **318**, 463–473.

Koromilas, A.E., Li, S., Matlashewski, G. (2001) Control of interferon signaling in human papillomavirus infection, *Cytokine Growth Factor Rev.* **12**, 157–170.

Kozak, M. (2002) Pushing the limits of the scanning mechanism for initiation of translation, *Gene* **299**, 1–34.

Lee, K., KewalRamani, V.N. (2004) In defense of the cell: TRIM5alpha interception of mammalian retroviruses, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10496–10497.

Levesque, K., Finzi, A., Binette, J., Cohen, E.A. (2004) Role of CD4 receptor down-regulation during HIV-1 infection, *Curr. HIV Res.* **2**, 51–59.

Longworth, M.S., Laimins, L.A. (2004) Pathogenesis of human papillomaviruses in differentiating epithelia, *Microbiol. Mol. Biol. Rev.* **68**, 362–372.

Lopez, S., Arias, C.F. (2004) Multistep entry of rotavirus into cells: a versaillesque dance, *Trends Microbiol.* **12**, 271–278.

Maggon, K., Barik, S. (2004) New drugs and treatment for respiratory syncytial virus, *Rev. Med. Virol.* **14**, 149–168.

McFadden, G., Lalani, A., Everett, H., Nash, P., Xu, X. (1998) Virus-encoded receptors for cytokines and chemokines, *Semin. Cel. Dev. Biol.* **9**, 359–368.

Minghetti, L., Visentin, S., Patrizio, M., Franchini, L., Ajmone-Cat, M.A., Levi, G. (2004)

Multiple actions of the human immunodeficiency virus type-1 Tat protein on microglial cell functions, *Neurochem. Res.* **29**, 965–978.

Mogensen, T.H., Paludan, S.R. (2001) Molecular pathways in virus-induced cytokine production, *Microbiol. Mol. Biol. Rev.* **65**, 131–150.

Mohr, I. (2004) Neutralizing innate host defenses to control viral translation in HSV-1 infected cells, *Int. Rev. Immunol.* **23**, 199–220.

Moolla, N., Kew, M., Arbuthnot, P. (2002) Regulatory elements of hepatitis B virus transcription, *J. Viral Hepat.* **9**, 323–331.

Moore, J.P., Trkola, A., Dragic, T. (1997) Co-receptors for HIV-1 entry, *Curr. Opin. Immunol.* **9**, 551–562.

Morita, E., Sundquist, W.I. (2004) Retrovirus budding, *Annu. Rev. Cell Dev. Biol.* **20**, 395–425.

Murphy, P.M. (2001) Viral exploitation and subversion of the immune system through chemokine mimicry, *Nat. Immunol.* **2**, 116–122.

Nath, A. (2002) Human immunodeficiency virus (HIV) proteins in neuropathogenesis of HIV dementia, *J. Infect. Dis.* **186**, S193–S198.

Ott, D.E. (2002) Potential roles of cellular proteins in HIV-1, *Rev. Med. Virol.* **12**, 359–374.

Penin, F., Dubuisson, J., Rey, F.A., Moradpour, D., Pawlotsky, J.M. (2004) Structural biology of hepatitis C virus, *Hepatology* **39**, 5–19.

Pereira, C.F., Paridaen, J.T. (2004) Anti-HIV drug development–an overview, *Curr Pharm Des.* **10**, 4005–4037.

Piguet, V., Sattentau, Q. (2004) Dangerous liaisons at the virological synapse, *J. Clin. Invest.* **114**, 605–610.

Piller, S.C., Caly, L., Jans, D.A. (2003) Nuclear import of the pre-integration complex (PIC): the Achilles heel of HIV?, *Curr. Drug Targets* **4**, 409–429.

Relph, K., Harrington, K., Pandha, H. (2004) Recent developments and current status of gene therapy using viral vectors in the united kingdom, *BMJ* **329**, 839–842.

Rosenberg, N., Jolicoeur, P. (1997) Retroviral Pathogenesis, in: Coffin, J.M., Hughes, S.H., Varmus, H.E. (Eds.) *Retroviruses*. New York, Cold Spring Harbor Laboratory Press.

Samuel, C.E. (2001) Antiviral actions of interferons, *Clin. Microbiol. Rev.* **14**, 778–809.

Schneider, R.J., Mohr, I. (2003) Translation initiation and viral tricks, *Trends Biochem. Sci.* **28**, 130–136.

Schroder, K., Hertzog, P.J., Ravasi, T., Hume, D.A. (2004) Interferon-gamma: an overview of signals, mechanisms and functions, *J. Leukoc. Biol.* **75**, 163–189.

Seet, B.T., McFadden, G. (2002) Viral chemokine-binding proteins, *J. Leukoc. Biol.* **72**, 24–34.

Sen, G.C. (2001) Viruses and interferons, *Annu. Rev. Microbiol.* **55**, 255–281.

Sherman, M.P., Greene, W.C. (2002) Slipping through the door: HIV entry into the nucleus, *Microbes Infect.* **4**, 67–73.

Sinclair, A.J. (2003) bZIP proteins of human gammaherpesviruses, *J. Gen. Virol.* **84**, 1941–1949.

Skehel, J.J., Wiley, D.C. (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin, *Annu. Rev. Biochem.* **69**, 531–569.

Skoczylas, C., Fahrbach, K.M., Rundell, K. (2004) Cellular targets of the SV40 small-t antigen in human cell transformation, *Cell Cycle* **3**, 606–610.

Smiley, J.R. (2004) Herpes simplex virus virion host shutoff protein: immune evasion mediated by a viral RNase?, *J. Virol.* **78**, 1063–1068.

Stevenson, M. (2003) Dissecting HIV-1 through RNA interference, *Nat. Rev. Immunol.* **3**, 851–858.

Strebel, K. (2003) Virus-host interactions: role of HIV proteins Vif, Tat, and rev, *AIDS* **4**, S25–S34.

Tailor, C.S., Lavillette, D., Marin, M., Kabat, D. (2003) Cell surface receptors for gammaretroviruses, *Curr. Top. Microbiol. Immunol.* **281**, 29–106.

Toulme, J.J., Darfeuille, F., Kolb, G., Chabas, S., Staedel, C. (2003) Modulating viral gene expression by aptamers to RNA structures, *Biol. Cell* **95**, 229–238.

Ugolini, S., Mondor, I., Sattentau, Q.J. (1999) HIV-1 attachment: another look, *Trends Microbiol.* **7**, 144–149.

Verani, A., Gras, G., Pancino, G. (2005) Macrophages and HIV-1: dangerous liaisons, *Mol. Immunol.* **42**, 195–212.

Weidman, M.K., Sharma, R., Raychaudhuri, S., Kundu, P., Tsai, W., Dasgupta, A. (2003) The interaction of cytoplasmic RNA viruses with the nucleus, *Virus Res.* **95**, 75–85.

White, M.K., Khalili, K. (2004) Polyomaviruses and human cancer: molecular mechanisms underlying patterns of tumorigenesis, *Virology* **324**, 1–16.

Witvrouw, M., Van Maele, B., Vercammen, J., Hantson, A., Engelborghs, Y., De Clercq, E., Pannecouque, C., Debyser, Z. (2004) Novel inhibitors of HIV-1 integration, *Curr. Drug Metab.* **5**, 291–304.

Wysocka, J., Herr, W. (2003) The herpes simplex virus VP16-induced complex: the makings of a regulatory switch, *Trends Biochem. Sci.* **28**, 294–304.

Young, L.S., Rickinson, A.B. (2004) Epstein-Barr virus: 40 years on, *Nat. Rev. Cancer* **4**, 757–768.

Younis, I., Green, P.L. (2005) The human T-cell leukemia virus Rex protein, *Front. Biosci.* **10**, 431–445.

Primary Literature

Ahmed, M., McKenzie, M.O., Puckett, S., Hojnacki, M., Poliquin, L., Lyles, D.S. (2003) Ability of the matrix protein of vesicular stomatitis virus to suppress beta interferon gene expression is genetically correlated with the inhibition of host RNA and protein synthesis, *J. Virol.* **77**, 4646–4657.

Alvarez, C.P., Lasala, F., Carrillo, J., Muniz, O., Corbi, A.L., Delgado, R. (2002) C-type lectins DC-SIGN and L-SIGN mediate cellular entry by Ebola virus in cis and in trans, *J. Virol.* **76**, 6841–6844.

Artsaenko, O., Tessmann, K., Sack, M., Haussinger, D., Heintges, T. (2003) Abrogation of hepatitis C virus NS3 helicase enzymatic activity by recombinant human antibodies, *J. Gen. Virol.* **84**, 2323–2332.

Ashrafi, G.H., Haghshenas, M.R., Marchetti, B., O'Brien, P.M., Campo, M.S. (2005) E5 protein of human papillomavirus type 16 selectively downregulates surface HLA class I, *Int. J. Cancer* **113**, 276–283.

Bai, J., Sui, J., Zhu, R.Y., Tallarico, A.S., Gennari, F., Zhang, D., Marasco, W.A. (2003) Inhibition of Tat-mediated transactivation and HIV-1 replication by human anti-hCyclinT1 intrabodies, *J. Biol. Chem.* **278**, 1433–1442.

Baize, S., Kaplon, J., Faure, C., Pannetier, D., Georges-Courbot, M.C., Deubel, V. (2004) Lassa virus infection of human dendritic cells and macrophages is productive but fails to activate cells, *J. Immunol.* **172**, 2861–2869.

Barbouche, R., Miquelis, R., Jones, I.M., Fenouillet, E. (2003) Protein-disulfide isomerase-mediated reduction of two disulfide bonds of HIV envelope glycoprotein 120 occurs post-CXCR4 binding and is required for fusion, *J. Biol. Chem.* **278**, 3131–3136.

Barnard, R.J., Narayan, S., Dornadula, G., Miller, M.D., Young, J.A. (2004) Low pH is required for avian sarcoma and leukosis virus Env-dependent viral penetration into the cytosol and not for viral uncoating, *J. Virol.* **78**, 10433–10441.

Barth, H., Schafer, C., Adah, M.I., Zhang, F., Linhardt, R.J., Toyoda, H., Kinoshita-Toyoda, A., Toida, T., Van Kuppevelt, T.H., Depla, E., Von Weizsacker, F., Blum, H.E., Baumert, T.F. (2003) Cellular binding of hepatitis C virus envelope glycoprotein E2 requires cell surface heparan sulfate, *J. Biol. Chem.* **278**, 41003–41012.

Basler, C.F., Mikulasova, A., Martinez-Sobrido, L., Paragas, J., Muhlberger, E., Bray, M., Klenk, H.D., Palese, P., Garcia-Sastre, A. (2003) The Ebola virus VP35 protein inhibits activation of interferon regulatory factor 3, *J. Virol.* **77**, 7945–7956.

Basyuk, E., Galli, T., Mougel, M., Blanchard, J.M., Sitbon, M., Bertrand, E. (2003) Retroviral genomic RNAs are transported to the plasma membrane by endosomal vesicles, *Dev. Cell* **5**, 161–174.

Beausejour, Y., Tremblay, M.J. (2004) Interaction between the cytoplasmic domain of ICAM-1 and Pr55Gag leads to acquisition of host ICAM-1 by human immunodeficiency virus type 1, *J. Virol.* **78**, 11916–11925.

Bériault, V., Clément, J.F., Levesque, K., Lebel, C., Yong, X., Chabot, B., Cohen, E.A., Cochrane, A.W., Rigby, W.F., Mouland, A.J. (2004) A late role for the association of hnRNP A2 with the HIV-1 hnRNP A2 response elements in genomic RNA, Gag, and Vpr localization, *J. Biol. Chem.* **279**, 44141–44153.

Bode, J.G., Ludwig, S., Ehrhardt, C., Albrecht, U., Erhardt, A., Schaper, F., Heinrich, P.C., Haussinger, D. (2003) IFN-alpha antagonistic activity of HCV core protein involves induction of suppressor of cytokine signaling-3, *FASEB J.* **17**, 488–490.

Bosio, C.M., Aman, M.J., Grogan, C., Hogan, R., Ruthel, G., Negley, D., Mohamadzadeh, M., Bavari, S., Schmaljohn, A. (2003) Ebola and Marburg viruses replicate in monocyte-derived dendritic cells without inducing the

production of cytokines and full maturation, *J. Infect. Dis.* **188**, 1630–1638.

Bounou, S., Leclerc, J.E., Tremblay, M.J. (2002) Presence of host ICAM-1 in laboratory and clinical strains of human immunodeficiency virus type 1 increases virus infectivity and CD4(+)-T-Cell depletion in human lymphoid tissue, a major site of replication in vivo, *J. Virol.* **76**, 1004–1014.

Breiman, A., Grandvaux, N., Lin, R., Ottone, C., Akira, S., Yoneyama, M., Fujita, T., Hiscott, J., Meurs, E.F. (2005) Inhibition of RIG-I-dependent signaling to the interferon pathway during hepatitis C virus expression and restoration of signaling by IKKε, *J. Virol.* **79**, 3969–3978.

Cen, S., Guo, F., Niu, M., Saadatmand, J., Deflassieux, J., Kleiman, L. (2004) The interaction between HIV-1 Gag and APOBEC3G, *J. Biol. Chem.* **279**, 33177–33184.

Chatel-Chaix, L., Clément, J.F., Martel, C., Bériault, V., Gatignol, A., DesGroseillers, L., Mouland, A.J. (2004) Identification of Staufen in the human immunodeficiency virus type 1 Gag ribonucleoprotein complex and a role in generating infectious viral particles, *Mol. Cell. Biol.* **24**, 2637–2648.

Chen, W., Yan, W., Du, Q., Fei, L., Liu, M., Ni, Z., Sheng, Z., Zheng, Z. (2004) RNA interference targeting VP1 inhibits foot-and-mouth disease virus replication in BHK-21 cells and suckling mice, *J. Virol.* **78**, 6900–6907.

Chiu, Y.L., Cao, H., Jacque, J.M., Stevenson, M., Rana, T.M. (2004) Inhibition of human immunodeficiency virus type 1 replication by RNA interference directed against human transcription elongation factor P-TEFb (CDK9/CyclinT1), *J. Virol.* **78**, 2517–2529.

Clawson, G.A., Miranda, G.Q., Sivarajah, A., Xin, P., Pan, W., Thiboutot, D., Christensen, N.D. (2004) Inhibition of papilloma progression by antisense oligonucleotides targeted to HPV11 E6/E7 RNA, *Gene Ther.* **11**, 1331–1341.

Coil, D.A., Miller, A.D. (2004) Phosphatidylserine is not the cell surface receptor for vesicular stomatitis virus, *J. Virol.* **78**, 10920–10926.

Costa-Mattioli, M., Svitkin, Y., Sonenberg, N. (2004) La autoantigen is necessary for optimal function of the poliovirus and hepatitis C virus internal ribosome entry site in vivo and in vitro, *Mol. Cell. Biol.* **24**, 6861–6870.

Cros, J.F., Palese, P. (2003) Trafficking of viral genomic RNA into and out of the nucleus: influenza, Thogoto and Borna disease viruses, *Virus Res.* **95**, 3–12.

Daher, A., Longuet, M., Dorin, D., Bois, F., Segeral, E., Bannwarth, S., Battisti, P.-L., Purcell, D., Benarous, R., Vaquero, C., Meurs, E.F., Gatignol, A. (2001) Two dimerization domains in the trans-activation response RNA-binding protein (TRBP) individually reverse the protein kinase R inhibition of HIV-1 long terminal repeat expression, *J. Biol. Chem.* **276**, 33899–33905.

DiPerna, G., Stack, J., Bowie, A.G., Boyd, A., Kotwal, G., Zhang, Z., Arvikar, S., Latz, E., Fitzgerald, K.A., Marshall, W.L. (2004) Poxvirus protein N1L targets the I-kappaB kinase complex, inhibits signaling to NF-kappaB by the tumor necrosis factor superfamily of receptors, and inhibits NF-kappaB and IRF3 signaling by toll-like receptors, *J. Biol. Chem.* **279**, 36570–36578.

Doepker, R.C., Hsu, W.L., Saffran, H.A., Smiley, J.R. (2004) Herpes simplex virus virion host shutoff protein is stimulated by translation initiation factors eIF4B and eIF4H, *J. Virol.* **78**, 4684–4699.

Donelan, N.R., Dauber, B., Wang, X., Basler, C.F., Wolff, T., Garcia-Sastre, A. (2004) The N- and C-terminal domains of the NS1 protein of influenza B virus can independently inhibit IRF-3 and beta interferon promoter activation, *J. Virol.* **78**, 11574–11582.

Dorin, D., Bonnet, M.C., Bannwarth, S., Gatignol, A., Meurs, E.F., Vaquero, C. (2003) The TAR RNA-binding protein, TRBP, stimulates the expression of TAR-containing RNAs *in vitro* and *in vivo* independently of its ability to inhibit the dsRNA-dependent kinase PKR., *J. Biol. Chem.* **278**, 4440–4448.

Dormitzer, P.R., Nason, E.B., Prasad, B.V., Harrison, S.C. (2004) Structural rearrangements in the membrane penetration protein of a non-enveloped virus, *Nature* **430**, 1053–1058.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells, *Nature* **411**, 494–498.

Elphick, G.F., Querbes, W., Jordan, J.A., Gee, G.V., Eash, S., Manley, K., Dugan, A., Stanifer, M., Bhatnagar, A., Kroeze, W.K., Roth, B.L., Atwood, W.J. (2004) The human

polyomavirus, JCV, uses serotonin receptors to infect cells, *Science* **306**, 1380–1383.

Enninga, J., Levy, D.E., Blobel, G., Fontoura, B.M. (2002) Role of nucleoporin induction in releasing an mRNA nuclear export block, *Science* **295**, 1523–1525.

Forgues, M., Difilippantonio, M.J., Linke, S.P., Ried, T., Nagashima, K., Feden, J., Valerie, K., Fukasawa, K., Wang, X.W. (2003) Involvement of Crm1 in hepatitis B virus X protein-induced aberrant centriole replication and abnormal mitotic spindles, *Mol. Cell. Biol.* **23**, 5282–5292.

Fortin, J.F., Cantin, R., Lamontagne, G., Tremblay, M. (1997) Host-derived ICAM-1 glycoproteins incorporated on human immunodeficiency virus type 1 are biologically active and enhance viral infectivity, *J. Virol.* **71**, 3588–3596.

Ge, Q., Filip, L., Bai, A., Nguyen, T., Eisen, H.N., Chen, J. (2004) Inhibition of influenza virus production in virus-infected mice by RNA interference, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8676–8681.

Geijtenbeek, T.B., Torensma, R., van Vliet, S.J., van Duijnhoven, G.C., Adema, G.J., van Kooyk, Y., Figdor, C.G. (2000) Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses, *Cell* **100**, 575–585.

Guo, F., Cen, S., Niu, M., Javanbakht, H., Kleiman, L. (2003) Specific inhibition of the synthesis of human lysyl-tRNA synthetase results in decreases in tRNA(Lys) incorporation, tRNA(3)(Lys) annealing to viral RNA, and viral infectivity in human immunodeficiency virus type 1, *J. Virol.* **77**, 9817–9822.

Hagmaier, K., Jennings, S., Buse, J., Weber, F., Kochs, G. (2003) Novel gene product of Thogoto virus segment 6 codes for an interferon antagonist, *J. Virol.* **77**, 2747–2752.

Hesselgesser, J., Taub, D., Baskar, P., Greenberg, M., Hoxie, J., Kolson, D.L., Horuk, R. (1998) Neuronal apoptosis induced by HIV-1 gp120 and the chemokine SDF-1 alpha is mediated by the chemokine receptor CXCR4, *Curr. Biol.* **8**, 595–598.

Hilgard, P., Stockert, R. (2000) Heparan sulfate proteoglycans initiate dengue virus infection of hepatocytes, *Hepatology* **32**, 1069–1077.

Hoeprich, S., Zhou, Q., Guo, S., Shu, D., Qi, G., Wang, Y., Guo, P. (2003) Bacterial virus phi29 pRNA as a hammerhead ribozyme escort to destroy hepatitis B virus, *Gene Ther.* **10**, 1258–1267.

Holm, P.S., Bergmann, S., Jurchott, K., Lage, H., Brand, K., Ladhoff, A., Mantwill, K., Curiel, D.T., Dobbelstein, M., Dietel, M., Gansbacher, B., Royer, H.D. (2002) YB-1 relocates to the nucleus in adenovirus-infected cells and facilitates viral replication by inducing E2 gene expression through the E2 late promoter, *J. Biol. Chem.* **277**, 10427–10434.

Huang, W., Kiefer, J., Whalen, D., Flint, S.J. (2003) DNA synthesis-dependent relief of repression of transcription from the adenovirus type 2 IVa(2) promoter by a cellular protein, *Virology* **314**, 394–402.

Jolly, C., Kashefi, K., Hollinshead, M., Sattentau, Q.J. (2004) HIV-1 cell to cell transfer across an Env-induced, actin-dependent synapse, *J. Exp. Med.* **199**, 283–293.

Kashiwakura, Y., Tamayose, K., Iwabuchi, K., Hirai, Y., Shimada, T., Matsumoto, K., Nakamura, T., Watanabe, M., Oshimi, K., Daida, H. (2005) Hepatocyte growth factor receptor is a coreceptor for adeno-associated virus type 2 infection, *J. Virol.* **79**, 609–614.

Kim, K., Trang, P., Umamoto, S., Hai, R., Liu, F. (2004) RNase P ribozyme inhibits cytomegalovirus replication by blocking the expression of viral capsid proteins, *Nucleic Acids Res.* **32**, 3427–3434.

Kim, S.J., Jang, M.H., Stapleton, J.T., Yoon, S.O., Kim, K.S., Jeon, E.S., Hong, H.J. (2004) Neutralizing human monoclonal antibodies to hepatitis A virus recovered by phage display, *Virology* **318**, 598–607.

Koga, H., Ohshima, T., Shimotohno, K. (2004) Enhanced activation of Tax-dependent transcription of human T-cell Leukemia Virus type I (HTLV-I) long terminal repeat by TORC3, *J. Biol. Chem.* **279**, 52978–52983.

Kunke, D., Grimm, D., Denger, S., Kreuzer, J., Delius, H., Komitowski, D., Kleinschmidt, J.A. (2000) Preclinical study on gene therapy of cervical carcinoma using adeno-associated virus vectors, *Cancer Gene Ther.* **7**, 766–777.

Le May, N., Dubaele, S., Proietti De Santis, L., Billecocq, A., Bouloy, M., Egly, J.M. (2004) TFIIH transcription factor, a target for the Rift Valley hemorrhagic fever virus, *Cell* **116**, 541–550.

Li, M.J., Bauer, G., Michienzi, A., Yee, J.K., Lee, N.S., Kim, J., Li, S., Castanotto, D., Zaia, J., Rossi, J.J. (2003) Inhibition of HIV-1 infection by lentiviral vectors expressing Pol

III-promoted anti-HIV RNAs, *Mol. Ther.* **8**, 196–206.

Lopez-Bueno, A., Valle, N., Gallego, J.M., Perez, J., Almendral, J.M. (2004) Enhanced cytoplasmic sequestration of the nuclear export receptor CRM1 by NS2 mutations developed in the host regulates parvovirus fitness, *J. Virol.* **78**, 10674–10684.

Ma, Q., Jones, D., Borghesani, P.R., Segal, R.A., Nagasawa, T., Kishimoto, T., Bronson, R.T., Springer, T.A. (1998) Impaired B-lymphopoiesis, myelopoiesis, and derailed cerebellar neuron migration in CXCR4- and SDF-1-deficient mice, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9448–9453.

Manel, N., Kim, F.J., Kinet, S., Taylor, N., Sitbon, M., Battini, J.L. (2003) The ubiquitous glucose transporter GLUT-1 is a receptor for HTLV, *Cell* **115**, 449–459.

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., Trono, D. (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts, *Nature* **424**, 99–103.

Margottin, F., Bour, S.P., Durand, H., Selig, L., Benichou, S., Richard, V., Thomas, D., Strebel, K., Benarous, R. (1998) A novel human WD protein, h-ßTrCP, that interacts with HIV-1 Vpu, connects CD4 to the ER degradation pathway through an F-box motif, *Mol. Cell* **1**, 565–574.

Marques, J.T., Trindade, G.D., Da Fonseca, F.G., Dos Santos, J.R., Bonjardim, C.A., Ferreira, P.C., Kroon, E.G. (2001) Characterization of ATI, TK and IFN-alpha/betaR genes in the genome of the BeAn 58058 virus, a naturally attenuated wild Orthopoxvirus, *Virus Genes* **23**, 291–301.

Marzi, A., Gramberg, T., Simmons, G., Moller, P., Rennekamp, A.J., Krumbiegel, M., Geier, M., Eisemann, J., Turza, N., Saunier, B., Steinkasserer, A., Becker, S., Bates, P., Hofmann, H., Pohlmann, S. (2004) DC-SIGN and DC-SIGNR interact with the glycoprotein of Marburg virus and the S protein of severe acute respiratory syndrome coronavirus, *J. Virol.* **78**, 12090–12095.

Matthews, T., Salgo, M., Greenberg, M., Chung, J., DeMasi, R., Bolognesi, D. (2004) Enfuvirtide: the first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes, *Nat. Rev Drug Discov* **3**, 215–225.

McCaffrey, A.P., Nakai, H., Pandey, K., Huang, Z., Salazar, F.H., Xu, H., Wieland, S.F., Marion, P.L., Kay, M.A. (2003) Inhibition of hepatitis B virus in mice by RNA interference, *Nat. Biotechnol.* **21**, 639–644.

McDonald, D., Wu, L., Bohks, S.M., KewalRamani, V.N., Unutmaz, D., Hope, T.J. (2003) Recruitment of HIV and its receptors to dendritic cell-T cell junctions, *Science* **300**, 1295–1297.

McDonald, D., Vodicka, M.A., Lucero, G., Svitkina, T.M., Borisy, G.G., Emerman, M., Hope, T.J. (2002) Visualization of the intracellular behavior of HIV in living cells, *J. Cell. Biol.* **159**, 441–452.

McLaren, M., Asai, K., Cochrane, A. (2004) A novel function for Sam68: enhancement of HIV-1 RNA 3′ end processing, *RNA* **10**, 1119–1129.

Moore, M.J., Dorfman, T., Li, W., Wong, S.K., Li, Y., Kuhn, J.H., Coderre, J., Vasilieva, N., Han, Z., Greenough, T.C., Farzan, M., Choe, H. (2004) Retroviruses pseudotyped with the severe acute respiratory syndrome coronavirus spike protein efficiently infect cells expressing angiotensin-converting enzyme 2, *J. Virol.* **78**, 10628–10635.

Mossman, K.L., Smiley, J.R. (2002) Herpes simplex virus ICP0 and ICP34.5 counteract distinct interferon-induced barriers to virus replication, *J. Virol.* **76**, 1995–1998.

Mouland, A.J., Xu, H., Cui, H., Krueger, W., Munro, T.P., Prasol, M., Mercier, J., Rekosh, D., Smith, R., Barbarese, E., Cohen, É.A., Carson, J.H. (2001) RNA trafficking signals in human immunodeficiency virus type 1, *Mol. Cell. Biol.* **21**, 2133–2143.

Norman, K.L., Hirasawa, K., Yang, A.D., Shields, M.A., Lee, P.W. (2004) Reovirus oncolysis: the Ras/RalGEF/p38 pathway dictates host cell permissiveness to reovirus infection, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11099–11104.

Nydegger, S., Foti, M., Derdowski, A., Spearman, P., Thali, M. (2003) HIV-1 egress is gated through late endosomal membranes, *Traffic* **4**, 902–910.

Otto, G.A., Puglisi, J.D. (2004) The pathway of HCV IRES-mediated translation initiation, *Cell* **119**, 369–380.

Ping, Y.H., Chu, C.Y., Cao, H., Jacque, J.M., Stevenson, M., Rana, T.M. (2004) Modulating HIV-1 replication by RNA interference directed against human transcription elongation factor SPT5, *Retrovirology* **1**, 46.

Power, C., Zhang, K., van Marle, G. (2004) Comparative neurovirulence in lentiviral

infections: The roles of viral molecular diversity and select proteases, *J. Neurovirol.* **10**, 113–117.

Pozharskaya, V.P., Weakland, L.L., Zimring, J.C., Krug, L.T., Unger, E.R., Neisch, A., Joshi, H., Inoue, N., Offermann, M.K. (2004) Short duration of elevated vIRF-1 expression during lytic replication of human herpesvirus 8 limits its ability to block antiviral responses induced by alpha interferon in BCBL-1 cells, *J. Virol.* **78**, 6621–6635.

Saha, K., Zhang, J., Gupta, A., Dave, R., Yimen, M., Zerhouni, B. (2001) Isolation of primary HIV-1 that target CD8+ T lymphocytes using CD8 as a receptor, *Nat. Med.* **7**, 65–72.

Sanchez-Velar, N., Udofia, E.B., Yu, Z., Zapp, M.L. (2004) hRIP, a cellular cofactor for Rev function, promotes release of HIV RNAs from the perinuclear region, *Genes Dev.* **18**, 23–34.

Schlegel, R., Tralka, T.S., Willingham, M.C., Pastan, I. (1983) Inhibition of VSV binding and infectivity by phosphatidylserine: is phosphatidylserine a VSV-binding site?, *Cell* **32**, 639–646.

Selinka, H.C., Giroglou, T., Sapp, M. (2002) Analysis of the infectious entry pathway of human papillomavirus type 33 pseudovirions, *Virology* **299**, 279–287.

Shimojima, M., Miyazawa, T., Ikeda, Y., McMonagle, E.L., Haining, H., Akashi, H., Takeuchi, Y., Hosie, M.J., Willett, B.J. (2004) Use of CD134 as a primary receptor by the feline immunodeficiency virus, *Science* **303**, 1192–1195.

Shukla, D., Liu, J., Blaiklock, P., Shworak, N.W., Bai, X., Esko, J.D., Cohen, G.H., Eisenberg, R.J., Rosenberg, R.D., Spear, P.G. (1999) A novel role for 3-O-sulfated heparan sulfate in herpes simplex virus 1 entry, *Cell* **99**, 13–22.

Stopak, K., de Noronha, C., Yonemoto, W., Greene, W.C. (2003) HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability, *Mol. Cell* **12**, 591–601.

Streblow, D.N., Soderberg-Naucler, C., Vieira, J., Smith, P., Wakabayashi, E., Ruchti, F., Mattison, K., Altschuler, Y., Nelson, J.A. (1999) The human cytomegalovirus chemokine receptor US28 mediates vascular smooth muscle cell migration, *Cell* **99**, 511–520.

Summerford, C., Samulski, R.J. (1998) Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions, *J. Virol.* **72**, 1438–1445.

Swanson, C.M., Puffer, B.A., Ahmad, K.M., Doms, R.W., Malim, M.H. (2004) Retroviral mRNA nuclear export elements regulate protein function and virion assembly, *EMBO J.* **23**, 2632–2640.

Tachibana, K., Hirota, S., Iizasa, H., Yoshida, H., Kawabata, K., Kataoka, Y., Kitamura, Y., Matsushima, K., Yoshida, N., Nishikawa, S., Kishimoto, T., Nagasawa, T. (1998) The chemokine receptor CXCR4 is essential for vascularization of the gastrointestinal tract, *Nature* **393**, 591–594.

Tang, C., Loeliger, E., Kinde, I., Kyere, S., Mayo, K., Barklis, E., Sun, Y., Huang, M., Summers, M.F. (2003) Antiviral inhibition of the HIV-1 capsid protein, *J. Mol. Biol.* **327**, 1013–1020.

Terstegen, L., Gatsios, P., Ludwig, S., Pleschka, S., Jahnen-Dechent, W., Heinrich, P.C., Graeve, L. (2001) The vesicular stomatitis virus matrix protein inhibits glycoprotein 130-dependent STAT activation, *J. Immunol.* **167**, 5209–5216.

Toggas, S.M., Masliah, E., Rockenstein, E.M., Rall, G.F., Abraham, C.R., Mucke, L. (1994) Central nervous system damage produced by expression of the HIV-1 coat protein gp120 in transgenic mice, *Nature* **367**, 188–193.

Towers, G.J., Hatziioannou, T., Cowan, S., Goff, S.P., Luban, J., Bieniasz, P.D. (2003) Cyclophilin a modulates the sensitivity of HIV-1 to host restriction factors, *Nat. Med.* **9**, 1138–1143.

Turelli, P., Doucas, V., Craig, E., Mangeat, B., Klages, N., Evans, R., Kalpana, G., Trono, D. (2001) Cytoplasmic recruitment of INI1 and PML on incoming HIV preintegration complexes: interference with early steps of viral replication, *Mol. Cell* **7**, 1245–1254.

Vanherberghen, B., Andersson, K., Carlin, L.M., Nolte-'t Hoen, E.N., Williams, G.S., Hoglund, P., Davis, D.M. (2004) Human and murine inhibitory natural killer cell receptors transfer from natural killer cells to target cells, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16873–16878.

von Schwedler, U.K., Stuchell, M., Muller, B., Ward, D.M., Chung, H.Y., Morita, E., Wang, H.E., Davis, T., He, G.P., Cimbora, D.M., Scott, A., Krausslich, H.G., Kaplan, J., Morham, S.G., Sundquist, W.I. (2003) The

protein network of HIV budding, *Cell* **114**, 701–713.

Wang, X., Huong, S.M., Chiu, M.L., Raab-Traub, N., Huang, E.S. (2003) Epidermal growth factor receptor is a cellular receptor for human cytomegalovirus, *Nature* **424**, 456–461.

Wang, S.Z., Bao, Y.X., Rosenberger, C.L., Tesfaigzi, Y., Stark, J.M., Harrod, K.S. (2004) IL-12p40 and IL-18 modulate inflammatory and immune responses to respiratory syncytial virus infection, *J. Immunol.* **173**, 4040–4049.

Wang, C., Pflugheber, J., Sumpter, R., Sodora, D.L., Hui, D., Sen, G.C., Gale, M. (2003) Alpha interferon induces distinct translational control programs to suppress hepatitis C virus RNA replication, *J. Virol.* **77**, 3898–3912.

Wang, Z., Ren, L., Zhao, X., Hung, T., Meng, A., Wang, J., Chen, Y.G. (2004) Inhibition of severe acute respiratory syndrome virus replication by small interfering RNAs in mammalian cells, *J. Virol.* **78**, 7523–7527.

Xiong, H., Zeng, Y.C., Lewis, T., Zheng, J., Persidsky, Y., Gendelman, H.E. (2000) HIV-1 infected mononuclear phagocyte secretory products affect neuronal physiology leading to cellular demise: relevance for HIV-1-associated dementia, *J. Neurovirol.* **6**(Suppl.1), S14–S23.

Yedavalli, V.S., Neuveut, C., Chi, Y.H., Kleiman, L., Jeang, K.T. (2004) Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function, *Cell* **119**, 381–392.

Yokota, S., Yokosawa, N., Okabayashi, T., Suzutani, T., Miura, S., Jimbow, K., Fujii, N. (2004) Induction of suppressor of cytokine signaling-3 by herpes simplex virus type 1 contributes to inhibition of the interferon signaling pathway, *J. Virol.* **78**, 6282–6286.

Youn, H.G., Matsumoto, J., Tanaka, Y., Shimotohno, K. (2003) SR-related protein TAXREB803/SRL300 is an important cellular factor for the transactivational function of human T-cell lymphotropic virus type 1 Tax, *J. Virol.* **77**, 10015–10027.

Zennou, V., Perez-Caballero, D., Gottlinger, H., Bieniasz, P.D. (2004) APOBEC3G incorporation into human immunodeficiency virus type 1 particles, *J. Virol.* **78**, 12058–12061.

Zou, Y.R., Kottmann, A.H., Kuroda, M., Taniuchi, I., Littman, D.R. (1998) Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development, *Nature* **393**, 595–599.

Zurcher, T., Marion, R.M., Ortin, J. (2000) Protein synthesis shut-off induced by influenza virus infection is independent of PKR activity, *J. Virol.* **74**, 8781–8784.

Virus-free Gene Transfer Systems in Somatic Gene Therapy

Oliver Kayser[1] *and Albrecht F. Kiderlen*[2]
[1] *Pharmaceutical Biology, Rijksuniversiteit Groningen, Groningen, The Netherlands*
[2] *Robert Koch-Institut, Berlin, Germany*

1 What is Gene Therapy? 486

2 Strategies in Gene Therapy 487

3 Gene-transfection Systems 489
3.1 Physical Gene-transfection Systems 490
3.1.1 Electroporation 490
3.1.2 Bioballistics 490
3.2 Chemical Gene-transfection Systems 490
3.2.1 Cationic Liposomes (Lipoplex) 491
3.2.2 Polymer Particles (Polyplex) 494
3.2.3 Poly-L-Lysine (PLL) and Poly-L-Arginine (PLA) 495
3.2.4 Polyethyleneimine (PEI) 496
3.2.5 Dendrimers 496
3.2.6 Chitosan 497
3.2.7 Poly(2-Dimethylamino)Ethylmethacrylate 497

4 Perspectives 498

Bibliography 499
Books and Reviews 499
Primary Literature 499

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

Keywords

Liposomes
Microscopic spherical vesicles that form spontaneously when phospholipids are hydrated. The enclosed aqueous core may be used to transport hydrophilic, the uni- or multilamellar shell to transport lipophilic molecules.

Somatic
Referring to cells of the body, as opposed to sex cells or gametes of the germ line.

Transfection
Insertion of small pieces of DNA into a cell. If the DNA is expressed without integration into the host cell DNA, the transfection is transient. If the DNA integrates into the host cell DNA, it is replicated whenever the host cell divides, and the transfection is stable.

Vectors
Sensu strictu: DNA elements used to amplify and express foreign DNA in a host cell.
Sensu latu: Also used synonymous with ''gene-transfection systems,'' that is, biological, physical, and chemical carriers that augment the entry of pieces of DNA into a cell.

In somatic gene therapy, nucleic acid is transferred into and subsequently expressed in somatic cells (i.e. not in stem cells) in order to alter or substitute gene functions as a means of prophylactic or therapeutic treatment of patients. The nucleic acid itself is the pharmaceutical drug. For transfer of genes, viral (retro, adeno, and adenovirus-associated viruses), physical (electroporation and bioballistics), and chemical systems are utilized. Chemical and physical systems are united as nonviral systems.

1 What is Gene Therapy?

Gene therapy may be defined as the expression of a gene that has been introduced into a target cell or a target tissue in order to alter an existing function or to introduce a new function with the aim of curing or protecting a patient from a specific disease. In many countries, this is restricted by law to somatic cells. In Germany, for example, genetic manipulation of germ cells is forbidden according to §5 Embryoschutz Gesetz (EschG) and is under penalty of up to 5-years imprisonment. German legislation on gene transfection as a form of therapy or medicine is still incomplete. An amendment specifically covering somatic gene therapy has not yet been passed. German Drug Regulation (§2 and §3 Deutsches Arzneimittelgesetz, AMG) defines DNA introduced for therapeutic purposes in a very general manner as a pharmaceutical product. German Gene Technology Regulation (§2 (3) Deutsches Gentechnikgesetz, GenTG) explicitly excludes the regulation

of gene therapy. Interestingly, gene therapy might be indirectly affected: theoretically (*sensu strictu*), a patient undergoing gene therapy becomes a genetically modified organism. Consequently, his release from hospital should be subject to authorization. When discussing DNA as a pharmaceutical, a look at European legislation, for example, at the guidelines of the European Agency for the evaluation of medicinal products (EMEA) is helpful. According to Council Regulations (EEC) No. 2309/93–ANNEX (List A), DNA is considered a pharmaceutical product for gene therapy. The necessary vector is simply defined as an additive.

Most likely, gene therapy will only work for a limited number of "suitable" diseases in a restricted commercial environment. It is very costly and affords a highly complex technology as well as an individually tailored strategy for gene delivery, depending on the relevant circumstances and aims. Monogenetic diseases such as hemophilia, sickle cell anemia, adenosine-deaminase (ADA)-deficit, cystic fibrosis, or Duchenne muscular dystrophy are likely first candidates, as they require the replacement or substitution of only a single gene. Trisomy 21 (Down syndrome), on the other hand, which is also frequently mentioned in this context, is a poor candidate, as it is much more difficult to silence the additional chromosome 21 than to replace nonfunctional genes. Furthermore, the monogenetic origin of Down syndrome is under debate. In different forms of cancer and heart or circulatory failures or of acquired genetic dysfunctions such as hepatitis or AIDS, two or more genes must be substituted or otherwise manipulated. Here, successful gene therapy seems unlikely at the moment due to technical limitations in DNA transport into target cells and in tissue-specific forms of application.

Among the known innate diseases, gene therapy has been most intensively investigated in ADA deficiency and cystic fibrosis. Both reduce mean life expectancy to below 20 years, accompanied by severe symptoms that strongly affect the quality of life.

This review covers nonviral gene-transfection strategies of current interest with special reference to experimental results found *in vivo* and to clinical trials.

2 Strategies in Gene Therapy

Independent of the respective method of gene transfer, two basic strategies for gene therapy may be discriminated (Fig. 1). Following the *ex vivo* strategy, cells or tissue are first removed from the patient, then exposed *in vitro* to the therapeutic genetic construct. If the transfection has been successful, the material is reimplanted in the patient. In the alternative *in vivo* methods, the therapeutic gene or DNA-sequence is integrated into a vector and this construct is injected locally or systemically into the patient. Among other deficits, the latter method is characterized by poor tissue selectivity, rapid extracellular DNA degradation and the danger of inducing oncogenes when using viral vectors.

Further differentiation is possible at the molecular level (Fig. 2): An intact (therapeutic) gene may simply be added to the defect one (Fig. 2A), a missing gene may be substituted (Fig. 2B), or a malfunctioning gene product may be inhibited, for example, during gene translation by giving mRNA-complementary antisense oligonucleotides (Fig. 2C). The first antisense-oligonucleotide drug, Vitravene™ (with

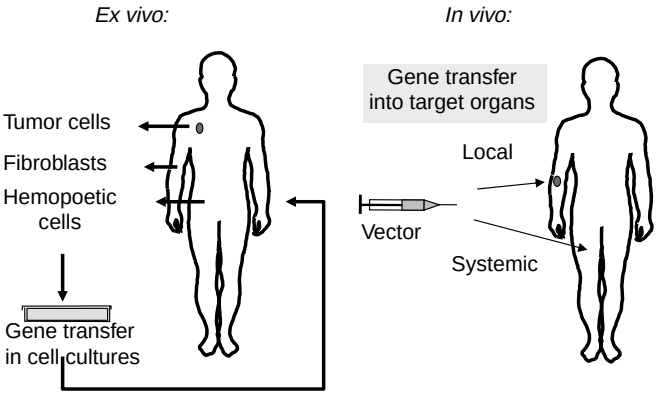

Fig. 1 *Ex vivo* and *in vivo* gene therapy strategies.

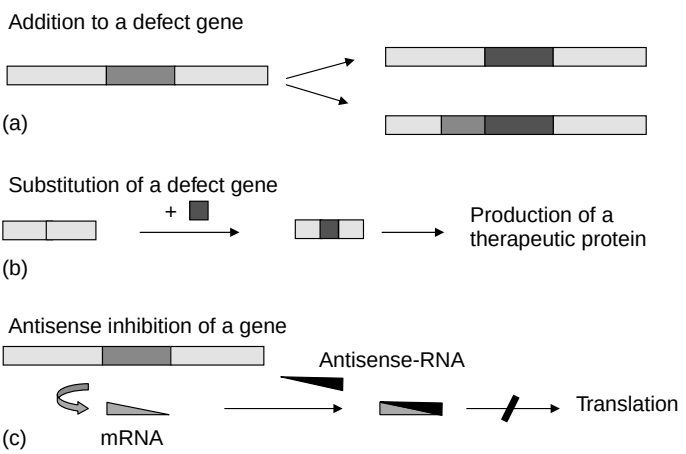

Fig. 2 Therapeutic gene functions.

formivirsen as the active ingredient), has been licensed for treating cytomegalovirus (CMV)-retinitis in AIDS patients.

The description of plasmid vectors, the biotechnology necessary for their production as pharmaceuticals and licensing specifications are not major subjects of this article. These can be studied in comprehensive overviews by Hutchins or Ferreira.

Intensive work is ongoing, both in designing new synthetic vectors and in improving DNA as a therapeutic agent. The transfection efficiency of "naked" DNA is low and it is rapidly degraded in the cytosol. The expression of the therapeutic gene may be enhanced by eukaryotic promoters of viral origin such as CMV or simian virus. The influence of the 5′UTR-site on the translation efficiency of mRNA is also being studied. The insertion of at least one intron into the cDNA may lead to 100-fold enhancement of mRNA. The addition of a suitable terminal poly(A) signal (bovine growth factor) has similar effects. Special attention is given to the development of tissue-specific promoters, restricting transfected gene expression to the necessary therapeutic

sites. A very elegant strategy for controlling the expression of the transfected gene is by turning it on or off with common oral drugs such as tetracycline or progesterone-antagonist. These interact with a mutated receptor acting as a transgene, thus initiating a signal transduction cascade, which finally induces the transcription (or its inhibition) of the respective gene for just as long as the drug is kept at a sufficient level.

3 Gene-transfection Systems

Owing to their rapid degradation in the cytosol, genes or DNA sequences are only rarely transfected as "naked" molecules. The commonly used gene transfer methods may be segregated into biological, physical, and chemical systems (Table 1). The biological systems involving viral vectors (retro, adeno, or pox viruses) make up 77% of all gene-transfection studies published to date. However, physical and chemical methods have experienced a relative increase in recent years (12% from 1998 to 2003). This tendency reflects the often highly serious and hardly controllable side effects of viral systems. Despite recent modifications of wild-type vectors, the future of such systems for human application appears limited. One main problem of viral vectors is their strong immunogenicity. Already the first (high-dose) application may initiate an immune reaction against the given proteins, which may lead to all sorts of allergic reactions, even lethal anaphylactic shock, when repeated. Possible reversion of the virus to wild type is another danger. The potential induction of oncogenes by retroviruses must also be mentioned.

With this background, the development of chemical or physical gene transfer systems appears especially interesting, combining simple usage with maximum safety. Further requirements for an ideal gene transfer system are minimal infectivity and immunogenicity, defined

Tab. 1 Gene transfer methods in somatic gene therapy.

	Method	*Application*	*Tissue selectivity*	*Transfection efficiency*	*Expression duration*
Chemical	Liposomes, Ca-phosphate-precipitation	*Ex vivo*/(*in vivo*)	No	Low	Transient/ stabile
Physical	Microinjection, electroporation, "particle-bombardment"	*Ex vivo*/(*in vivo*)	No	Moderate	Transient
Biological	Nonviral:ligand/ receptor	(*Ex vivo*)/*in vivo*	Yes	Low–moderate	Transient
	Viral: (e.g. retro-, adeno-, AAV)	*Ex vivo*/*in vivo*	Some	Moderate–high	Transient/ stabile

From Gottschalk, U., Chan, S. (1998) Somatic gene therapy. Present situation and future perspective, *Arzneimittelforschung* **48**, 1111–1120.

chemical and physical characteristics, and the possibility of multiple dosing.

When looking for suitable alternatives to viral vectors, the pharmaceutical industry can offer a broad spectrum of thoroughly investigated and readily available medicinal carrier systems. As far as molecular transport and organ or cell specificity are concerned, the demands on modern drug delivery and gene transfer systems exhibit so much similarity that the latter may profit significantly in the fields of biotechnology and pharmacology. Among the chemical vectors, liposomes, polymer nanoparticles, and polylysine particles deserve special discussion. Physical transfection systems such as electroporation and bioballistics are further examples for an efficient transfer of existing know how in pharmaceutical technology to somatic gene therapy. It might be mentioned that gene transfer in a nonviral system must correctly be addressed as *transfection*, whereas, in biological systems, it should be referred to as *transduction*. Table 1 suggests a systematic arrangement of methods used in gene transfer technology. In the following sections, the most important contemporary nonviral gene transfer systems are described and assessed.

3.1 Physical Gene-transfection Systems

3.1.1 Electroporation

During electroporation, cells or tissues are exposed to an electric field of high voltage (up to 1 kV). Short, rapid pulses cause transient membrane instability and the formation of pores with a mean lifetime of up to minutes. Soluble DNA constructs that have been added to the culture medium or injected into the tissue may thereby enter the cell and ultimately reach the nucleus. The basic principle is also known to pharmacists as iontophoresis and is used for transdermal drug application. This gene-transfection method is so far well accepted by patients and is safe, especially due to the low risk of infection. To date, mainly liver, muscle, and skin cells have been transformed this way, mostly via the transdermal route.

3.1.2 Bioballistics

Bioballistical methods are already widely used in biotechnology, for example as "gene guns" for injecting DNA vaccines (Fig. 3). For this, linear or circular (plasmid-) DNA) is adsorbed to nanometer-sized gold or tungsten particles. These are shot from a cylinder with compressed nitrogen or helium as propellant, thereby reaching speeds of up to 900 m s^{-1}. When applied to skin, the DNA/metal particles pass through dead tissue such as the *Stratum corneum* reaching living cells. Statistically, only one of 10 000 particles reaches the interior of viable cells, thereby loosing its DNA. For the DNA to then enter the nucleus, active transport mechanisms are probably necessary. Bioballistical gene transfer methods are obviously not suitable for systemic application. Further drawbacks are the need for very stable DNA and the high developing costs. Their greatest advantage, on the other hand, is the fact that related technologies are already approved and commercially available such as the Accell® (Powderject) (Fig. 3) or the Helios® (BioRad) systems, which are used for vaccination.

3.2 Chemical Gene-transfection Systems

A multitude of synthetic chemical vectors are being developed. Of these, cationic lipids and cationic polymers are probably

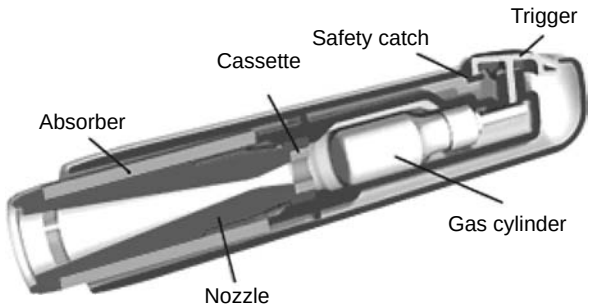

Fig. 3 Accela® gene gun (powderject). From: Burkoth, T.L., Bellhouse, B.J., Hewson, G., Longridge, D.J., Muddle, A.G., Sarphie, D.F. (1999) Transdermal and transmucosal powdered drug delivery, *Crit. Rev. Ther. Drug Carrier Syst.* **16**, 331–384.

the most thoroughly investigated. Already, in 1987, Felgner et al. could demonstrate *in vitro* gene transfection using cationic lipids. Transfer efficiency was systematically improved, leading to the first clinical studies on cystic fibrosis patients.

Irrespective of their highly variant molecular composition and 3D structure, chemical vectors such as liposomes and polymer particles have many biological and physical features in common. Positively charged amine functions are especially important, as these can be loaded with the negatively charged DNA molecules. Optimal interaction between plasmid DNA and cationic additives lead to the development of colloidal, positively charged particles, which may adhere to and be taken up by negatively charged cells a fact that is a cornerstone of chemical vector technology. Type and structure of the amine functions determine the stability of the complex, its cellular uptake in a phagosome, its release from the phagosome into the cytosol and dissociation of the DNA molecule, and even DNA transport to the nucleus.

Nucleic acids as drugs are still rather unusual in pharmaceutics. They are highly negatively charged and range in size from 10^3 kDa (oligonucleotides) to 10^6 kDa (genes). Their targets are invariably intracellular. In contrast to most conventional pharmaceuticals, nucleic acids are too large and too strongly charged to pass cell membranes by simple passive means. Furthermore, free, that is, "naked" nucleic acids are rapidly degraded by cellular nucleases. Chemical vectors therefore have the additional job to protect the DNA they carry from enzymatic degradation. In the context of gene transfer, liposomes are also referred to as lipoplex, polymers as polyplex, and combinations as lipopolyplex particles.

3.2.1 Cationic Liposomes (Lipoplex)

Cationic liposomes were developed and already complexed with DNA over 20 years ago. However, it was in the Human Genome Project that their potential as gene transfer vehicles really became apparent to geneticists and pharmacists. Cationic liposomes consist of cationic phospholipids, which may be divided according to their number of tertiary amine functions into monovalent and multivalent cationic lipids. An overview of the most commonly used cationic lipids is given in Fig. 4.

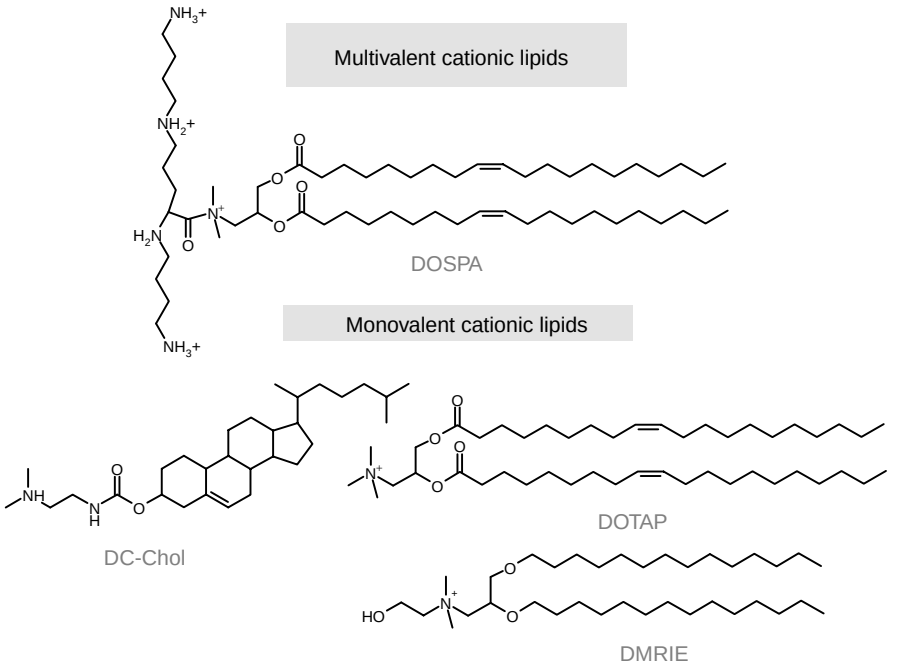

Fig. 4 Cationic lipids: DOSPA: 2,3-Dioleoyloxy-*N*-[2-(spermincarboxyamido)ethyl]-*N*,*N*-dimethyl-1-propanamminiumchloride; DOTAP: 1,2-Dioleoyloxypropyl-3-*N*,*N*,*N*-trimethyl-ammoniumchloride; DC-chol: 3β-[*N*-(*N*′,*N*′-dimethylaminoethane) carbamoyl]-cholesterol; DMRIE: 1,2-Dimyristyloxypropyl-3-*N*,*N*-dimethylhydroxy-ammoniumbromide. From Li, Z., Ma, Z. (2001) Nonviral gene therapy, *Curr. Gene Ther.* **1**, 201–226 and Simoes, S., Pires, P., Düzgünes, N., Pedroso de Lima, M.C. (1999) Cationic liposomes as gene transfer vectors: barriers to successful application in gene therapy, *Curr. Opin. Mol. Ther.* **1**, 147–157.

DOTAP (1,2-Dioleoyloxypropyl-3-*N*,*N*,*N*-trimethyl-ammoniumchloride) and DC-chol (3β-[*N*′,*N*′-dimethylaminoethane)-carbamoyl]-cholesterol are possibly the most important. However, owing to high toxicity, they must be mixed with helper lipids such as DOPE before processing to liposomes. A well-known transfection agent is Lipofectin®, which consists of equal parts DOTMA and DOPE.

The high general toxicity of the first cationic lipids initiated the synthesis of a variety of derivatives. The published *in vitro* data allows a first structure/activity analysis. The presence of a tertiary amine function and of 1 or 2 hydrophobic aliphatic chains are most important for producing stable liposomes and for binding negatively charged DNA. The tertiary amine functions can be combined with the hydrophobic lipid chains by either ester or ether bonds. Ester bonds are split more readily under physiological conditions than ether bonds and the lipoplex particles are consequentially metabolized more rapidly. These different metabolic characteristics are exemplified by DOTMA and DOTAP; the former is synthesized as ester, the latter as ether.

The aliphatic lipid chain may be saturated (DMRI) or contain double bonds (DOTAP, DOSPA). *In vivo*, dialkyl chains, with a length of 12 to 18 carbon atoms each, seem to promote the best gene transfection. A substitution of the alkyl chains with cholesterol, as in DC-chol, shows further advantages in the treatment of cystic fibrosis. Cholesterol structures have strong affinities to bronchoepithelial cells and interact less with plasma proteins.

Positively charged liposomes loaded with negatively charged DNA must maintain an overall positive charge in order to approach and contact target cells that are normally negatively charged. Liposomes that have been overloaded with DNA and have thus received a negative total charge exhibit low endocytosis rates and enhanced metabolic degradation. Proper selection of the cationic lipids and the lipid-to-DNA ratio are decisive for the rate of complexation and the colloidal structure of the product. Finally, conformations may change in time, a process also referred to as "aging" of the DNA/lipid complex.

Recent studies show that intracellular stability of the lipid/DNA complex and the necessary dissociation of the DNA molecule from its vector depend both on the pKs of the chosen lipids and on the pH of the product. Safinya et al. describe how not only the main cationic lipid but the helper lipids as well are influenced by the physical characteristics of the resulting lipid phase. For instance, when formulating DOTAP liposomes with DOPC, lamellar structures are achieved. However, when exchanging DOPC for DOPE as helper lipid, inverse hexagonal micelles are produced. One may speculate whether different lipid phase characteristics lead to defined biopharmaceutical variations.

Cationic liposomes have also been intensively investigated *in vivo*. In combination with neutral helper lipids, they appear to be well tolerated even at high concentrations; they are neither immunogenic nor do they induce toxic side effects. This applies to different routes of application such as intravenous, pulmonary, or nasal, none of which have been described as being significantly unpleasant by the recipient patients. Nevertheless, cationic liposomes also pose a number of problems such as high serum protein binding and extremely rapid metabolic degradation in the liver.

Liposomes are generally the most efficient at a diameter ranging from 400 to 500 nm. Larger particles are almost completely removed from the circulation by cells of the mononuclear phagocytes system of liver and lung and therefore do not reach other target areas. Pegylation is one means of improving pharmacokinetics. However, the useful extent of pegylation is restricted as further reduction of the zeta potential at the surface of the liposomes correlates with their reduced cellular uptake.

Low transfection rates are a further problem. Reasons are – among others – insufficient lysosomal release of the therapeutic DNA from its carrier and/or its rapid enzymatic degradation within the lysosome or in the cytosol. The DNA that was integrated in the liposomes must be released to become effective (see Fig. 5). Studies on microinjections of lipid/DNA complexes directly into the nucleus clearly show reduced transfection rates. As the nuclear pores have mean diameters of only 25 to 50 nm and passage for macromolecules is further controlled by the nuclear pore complex, passive diffusion into the karyosol is limited to particle sizes below 45 kDa. Conventional recombinant plasmids normally have a

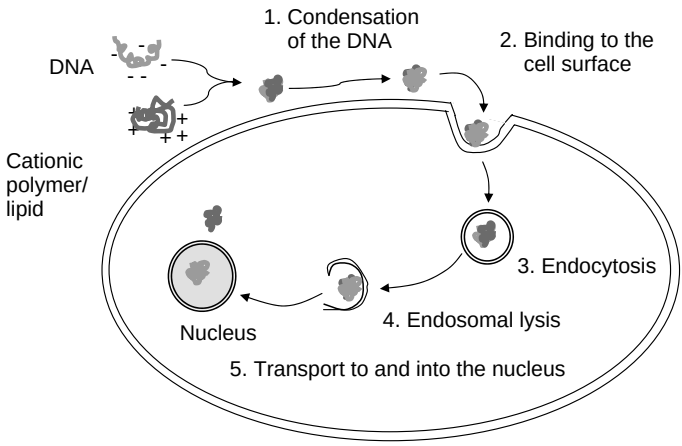

Fig. 5 Uptake, transport, and release of DNA in particulate gene transfer systems.

molecular weight ranging from 50 to 100 kDa and therefore require active transport into the nucleus.

Admission of plasmid DNA into the karyosol is steered by nuclear localization sequences, which are assisted by importin-beta, guanin-nucleotide-binding protein (Ran) and nuclear transport factor (NTF). In the karyosol, the therapeutic DNA is transcribed by RNA polymerases into mRNA, which is transported back into the cytosol where it attaches to ribosomes for translation into protein. In order to improve its transport into the nucleus, the therapeutic DNA may be coupled to such nucleus localizing sequences. The latter can be found naturally in certain viruses, which have developed this strategy for a most efficient nuclear invasion. For example, Rudolph et al. coupled DNA with short TAT sequences taken from the arginine-rich motif of the HIV-1-TAT protein and achieved significant enhancement of transfection. Of the 101 amino acid–long HIV-TAT sequence, they synthesized a 12 amino acid–long oligopeptide, bound it to polyarginine and thus achieved 390-fold enhanced transfection rates.

Unsatisfactory transfection rates and low cell or organ specificity also initiated the development of simple liposomes to virosomes or immunoliposomes. Virosomes are liposomes that contain viral proteins or fusiogenic peptides for enhanced DNA release from the endosome and for generally improved gene-transfection rates. Immunoliposomes are characterized by target-specific monoclonal antibodies bound to their surface. These were first developed in the 1980s for tumor-specific delivery of liposome-entrapped drugs and have since been modified for gene therapy.

3.2.2 Polymer Particles (Polyplex)

A further class of synthetic gene vectors that has received attention in past years is cationic polymers, which condense and package DNA with high efficiency. Polymerized or oligomerized branched or nonbranched amino acid chains composed of lysine or arginine are common. Polyethyleneimine, however, developed in 1995 by Boussif and already used for gene-transfection experiments *in vitro*

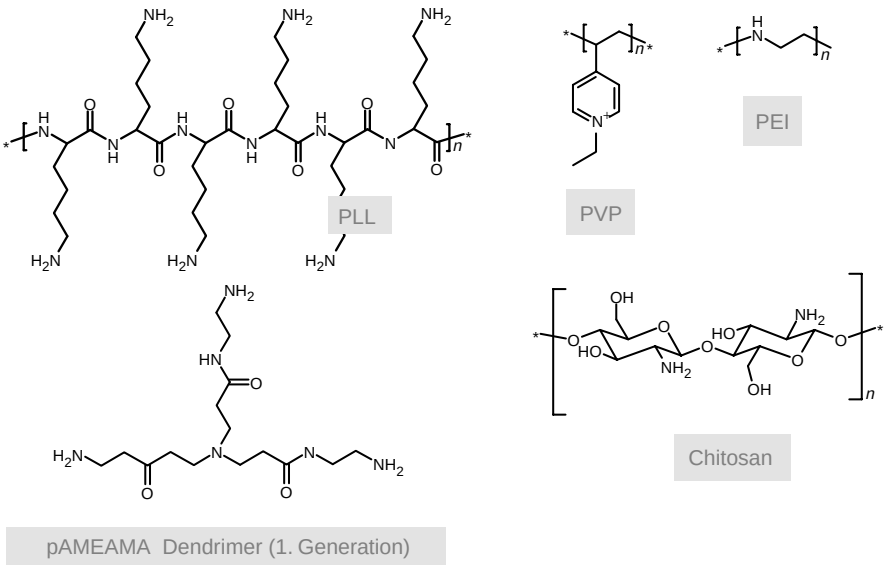

Fig. 6 Chemical structures of cationic polymers.

as well as *in vivo*, appears to be the most promising cationic polymer at the moment (see Fig. 6). Basically, cationic polymer gene transfer systems reveal the same pharmaceutical problems and intracellular barriers as described above for cationic liposomes.

3.2.3 Poly-L-Lysine (PLL) and Poly-L-Arginine (PLA)

Poly-L-lysine (PLL) has already been intensively used as polymeric gene transfer system. It is synthesized by polymerization of the *N*-carboxyanhydrid of lysine. Arginine is polymerized in a similar manner. PLL/DNA complexes are produced by dissolving both components in aqueous media and precipitating the particulate complexes. These particles, which normally range from 400 to 500 nm, are capable of transporting nucleic acids ranging from short molecules to large artificial yeast chromosomes. The first *in vivo* studies, however, revealed substantial toxicity combined with low DNA-transfection efficiency. Different chemical modifications and variations in particle size were then tested. Toxicity is substantially reduced by coating the particles with PEG derivatives, and the transfection rate is enhanced by attaching ligands such as transferrin, folate, or target-specific monoclonal antibodies. Interestingly, pegylation also achieved significant reduction in unwanted hepatic metabolization of the particles.

Apart from lysine, polymers of other cationic amino acids such as arginine and histidine were also investigated. Conjugation of histidine to ε-L-histidine enabled the development of highly interesting PLH/DNA complexes characterized by high transfection rates. One explanation for the elevated transfection efficiency may be that the highly protonated histidine structure that develops in the generally acidic (pH 6) endosomal environment may cause rapid destruction of the endosomal membrane and thus an enhanced

release of the therapeutic DNA into the cytosol.

3.2.4 Polyethyleneimine (PEI)

Linear or branched polyethyleneimines generally range in size from 1.8 to 800 kDa. They are synthesized by cationic polymerization. Starting from 2-substituted-2-oxazoline-monomers, linear polyethyleneimine (PEI) with a mean molecular weight of 22 kDa A, also known as ExGEn 500, are produced by hydrolysis. PEI/DNA complexes have already been used in many *in vivo* studies in animals following iv injection; clinical studies, however, have not been reported to date. PEI/DNA complexes have repeatedly revealed very high transfection efficacy. One advantage of PEI/DNA complexes over lipoplex or PLL particles is their intrinsic buffer capacity at lysosomal pH, leading, as described above for PLL particles, to rapid destruction of the lysosomal membrane and DNA release. This effect, also termed *proton sponge*, is brought about by the chemical structure of PEI. Polymerization produces particles with primary amines at the surface and secondary as well as tertiary amines in the interior. This causes a shift in pKa from 6.9 to 3.9. The strong protonation at a pH below 6 induces an osmotic gradient across the endosomal membrane, resulting in an influx of water, swelling, and, finally, disruption of the endosome with release of the PEI/DNA particles into the cytosol.

Though the literature reveals some discrepancies, there seems to exist an inverse relationship between the molecular weight of the PEI particles and their transfection efficiency. The interesting physicochemical characteristics of PEI-based gene transfer systems encouraged their further development, for example, to dendrimers (see below). In direct comparison, linear PEI complexes (e.g. PEI 22) seem to possess better transfection characteristics than branched ones (e.g. PEI 25). One explanation may be a premature dissociation and subsequent degradation of the DNA molecule.

Clinical trials with PEI complexes as gene transfer systems could so far not be undertaken because of their frequently intolerable general toxicity. Depending on the chemical structure, the lethal dose for mice ranges from 40 to 100 mg kg^{-1} body weight. The main problem lies in the strong interaction between PEI complexes and erythrocytes, leading to their aggregation and the danger of emboli. Pegylation of the PEI complexes may only partially help solve the problem, as a high degree of pegylation generally reduces particle uptake and thus transfection efficacy [77].

3.2.5 Dendrimers

The name "dendrimer" refers to the star- or tree-shaped, branched structures of this relatively new class of cationic gene transfer systems. They are frequently synthesized from polyamidoamines with special chemical or physical features. Probably best known are the "starburst" – dendrimers with particle sizes ranging from 5 to 100 nm. These particles reveal a highly regular branches "dendritic" symmetry. Starburst dendrimers are three-dimensional oligomeric or polymeric compounds, which, initiated from small molecules as nuclei, are built layer by layer ("generations") by repeated chemical reaction cycles. This allows an exquisite steering of the final size, three-dimensional form, and surface chemistry of a starburst polymer by the individual selection of

components and binding procedures for each generation.

The physical and chemical characteristics of dendrimers are mainly the result of number and type or amine functions on the particle surface, but the secondary and tertiary amines in the inside also affect their biological features. Despite their high molecular weights, dendrimers are soluble in water. They complex DNA with great efficiency, thus giving excellent vehicles for gene transfer. Their high transfer efficiency, however, is probably less due to high DNA adsorption rates and more to protonation of the amine functions after endosomal uptake. As described above for the PEI complexes, this induces an osmotic gradient, leading to osmotic lysis of the organelle and enhanced release of the complex into the cytosol. Dendrimer/DNA complexes have proven their gene transfer efficacy in *in vivo* studies. However, as already exhibited by the PEI complexes, they show strong, undesired interaction with erythrocytes, causing hemolysis. Again, the free primary amine functions on the particle surface are held responsible. Depending on the type of dendrimer and the target cell, cationic dendrimers also reveal general cytotoxicity at concentrations ranging from 50 to 300 $\mu g\ mL^{-1}$. On the other hand, no dendrimer has to date been reported to induce tumors or to substantially affect the immune system.

3.2.6 Chitosan

Chitosan is a fiber produced by hydrolyzing chitin, mostly from crustaceans. Owing to its free amine functions, chitosan may also be protonated (pKa = 5.6). In *in vitro* studies with Hela cells, chitosan/DNA complexes showed gene-transfection potency similar to that reported for PEI/DNA complexes. Plain chitosan, however, is almost insoluble in water at neutral pH (but soluble at acidic pH). For this reason, tri-methylated, quaternary chitosan derivatives have been produced that are sufficiently soluble under physiological conditions and easily complex DNA molecules. In *in vitro* experiments with COS-1 and CaCo-2 cells, these innovative chitosan derivatives proved superior to nontreated chitosan polymer, particularly as they showed no unspecific cytotoxicity.

3.2.7 Poly(2-Dimethylamino) Ethylmethacrylate

Methacrylate polymers are used in pharmaceutical technology for microencapsulation. They are synthesized by polymerization of monomeric dimethylaminoacrylic acid to poly(2-dimethylamino)ethylmethacrylate (pDMAEMA), which is both simple and cheap. Their low general toxicity makes these polymers interesting also as gene transfer vehicles. In *in vitro* experiments with OVCAR-3 and COS-7 cells, some pDMAEMA/DNA particles showed high transfection rates. This proved to be highly dependent on their size and charge. In HEPES-buffer (pH 7.4), pDMAEMA particles exhibit a positive zeta potential of around 25 mV and an average size of 100 to 200 nm. Following endosomal uptake, again the outer primary amine functions are protonated, leading to osmotic lysis of the endosome and release of the pDMAEMA/DNA particles into the cytosol.

Studies on DNA absorption onto pDMAEMA particles show that linear DNA (e.g. antisense oligonucleotides) is adsorbed more strongly than circular plasmid DNA. However, this stronger adsorption has a negative influence on the gene transfer efficiency, as the DNA molecule is less likely to dissociate from

the pDMAEMA particles in the cytosol. For this reason, circular DNA is preferred. As expected, DNA that is adsorbed to pDMAEMA is protected from degradation by DNAse I.

4 Perspectives

The nonviral gene-transfection systems introduced in this review bear significant advantages over the viral systems, but have serious drawbacks as well. One fact in favor is that many such systems are already well established in classical areas of pharmaceutical technology. Their production methods have already been optimized and safety aspect investigated in detail. Simple transfer of knowledge may substantially reduce developing costs. Nonviral systems are noninfectious. They allow significantly higher DNA-loading rates than viral systems, which reach their limits around 30 kb (Herpes virus). Nonviral systems are only weakly immunogenic and therefore allow – in stark contrast to viral systems – multiple application.

The main drawback of nonviral systems is that they normally only lead to transient expression of the therapeutic gene as it is not permanently integrated into the host genome. In consequence, the therapeutic gene transfer must be regularly repeated, possibly over a long period of time. Further disadvantages are insufficient cell or tissue specificity and low DNA transfer rates from the cytosol to the nucleus. Taken together, the gene-transfection performances of nonviral systems are even weaker than those of viral systems.

To date, it is still not possible to make a clear decision between viral and nonviral gene transfer systems. Mixed systems, hybrid vectors, which can be envisaged as "de-nucleated" viruses, are in the pipeline. A combination of viral surface proteins and liposomes or the integration of therapeutic DNA into artificial cells or viruses are further innovative ideas for improving somatic gene therapy. Most important is rapid progress on three fields: In cell biology, unspecific and specific intracellular trafficking of macromolecules still raises questions. In biochemistry, further DNA carriers must be brought forward for testing, and pharmaceutical technology must supply improved and cell-/tissue-specific drug delivery systems. Brought together in a rational form, such progress should make somatic gene therapy possible for selected disease forms already in the near future.

Somatic gene therapy might develop into one of the most important therapeutic strategies of the near future. Innate or acquired genetic defects are held responsible for a number of diseases such as hemophilia, cystic fibrosis (mucoviscidosis), adenosine-deaminase deficit and AIDS. Substituting or supplementing malfunctioning or missing genetic information by transiently or permanently inserting the appropriate gene appears to be a plausible therapeutic strategy especially from the patients' point of view. Attempts in gene therapy began in the early 1990s with great expectations that have only partially been met. No disease with a defined genetic background has so far been causally cured by gene therapy. Viral gene transfer systems have caused severe problems that could not be brought under control to date. The death of the 18-year-old Jesse Gelsinger is a tragic evidence for the basic deficits of viral transfection systems. In consequence, attention is now focusing on chemical and physical gene transfer systems.

See also Liposome Gene Transfection; Somatic Gene Therapy; Vaccination, Genetic; Vector System: Plasmid DNA.

Bibliography

Books and Reviews

Brooks, G. (2002) *Gene Therapy: The Use of DNA as a Drug*, Pharmaceutical Press, London.

EMEA Council Regulation 2309/93 as proclaimed 22nd July 1993 (OJ EC L 214).

Parker, A.L., Newman, C., Briggs, S., Seymour, L., Sheridan, P.J. (2003) Nonviral gene delivery: techniques and implications for molecular medicine, *Expert Rev. Mol. Med.* **22**, 113–142.

Pfeifer, A., Verma, I.M. (2001) Gene therapy: promises and problems, *Annu. Rev. Genomics Hum. Genet.* **2**, 177–211.

Rubanyi, G.M. (2001) The future of human gene therapy, *Mol. Aspects Med.* **22**, 113–142.

Primary Literature

Anwer, K., Shi, M., French, M.F., et al. (1998) Systemic effect of human growth hormone after intramuscular injection of a single dose of a muscle-specific gene medicine, *Hum. Gene Ther.* **9**, 659–670.

Arigita, C., Zuidam, N.J., Crommelin, D.J., Hennik, W.E. (1999) Association and dissociation characteristics of polymer/DNA complexes used for gene delivery, *Pharm. Res.* **16**, 1534–1541.

Barrett, A.D. (2002) Arilvax (Powderject), *Curr. Opin. Investig. Drugs* **3**, 992–995.

Behr, J. (1997) The proton sponge: a trick to enter cells the viruses did not exploit, *Chimia* **51**, 34–36.

Biggar, W.D., Klamut, H.J., Demacio, P.C., Stevens, D.J., Ray, P.N. (2002) Duchenne muscular dystrophy: current knowledge, treatment, and future prospects, *Clin. Orthop.* **401**, 88–106.

Blaese, R.M., Culver, K.W., Miller, A.D., et al. (1995) T lymphocyte-directed gene therapy for ADA-SCID: initial trial results after 4 years, *Science* **270**, 475–480.

Bordignon, C., Notarangelo, L.D., Nobili, N., et al. (1995) Gene therapy in peripheral blood lymphocytes and bone marrow for ADA-immunodeficient patients, *Science* **270**, 470–475.

Boussif, O., Lezoualch, F., Zanta, M.A., et al. (1995) A versatile vector for gene and oligonucleotide transfer into cells in culture and *in vivo*: polyethylenimine, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7297–7301.

Bragonzi, A., Dina, G., Villa, A., et al. (2000) Biodistribution and transgene expression with nonviral cationic vector/DNA complexes in the lungs, *Gene Ther.* **7**, 1753–1760.

Brissault, B., Kichler, A., Guis, C., Leborgne, C., Danos, O., Cheradame, H. (2003) Synthesis of linear polyethylenimine derivatives for DNA transfection, *Bioconjug. Chem.* **14**, 581–587.

Burkoth, T.L., Bellhouse, B.J., Hewson, G., Longridge, D.J., Muddle, A.G., Sarphie, D.F. (1999) Transdermal and transmucosal powdered drug delivery, *Crit. Rev. Ther. Drug Carrier Syst.* **16**, 331–384.

Chan, C.K., Jans, D.A. (2002) Using nuclear targeting signals to enhance non-viral gene transfer, *Immunol. Cell Biol.* **80**, 19–130.

Chemin, I., Moradpour, D., Wieland, S., et al. (1998) Liver-directed gene transfer: a linear polyethylenimine derivative mediates highly efficient DNA delivery to primary hepatocytes *in vitro* and *in vivo*, *J. Viral. Hepat.* **5**, 369–375.

Cloninger, M.J. (2002) Biological applications of dendrimers, *Curr. Opin. Chem. Biol.* **6**, 742–748.

Coleman, M.E., DeMayo, F., Yin, K.C., et al. (1995) Myogenic vector expression of insulin-like growth factor I stimulates muscle cell differentiation and myofiber hypertrophy in transgenic mice, *J. Biol. Chem.* **270**, 12109–12116.

De Jong, G., Telenius, A., Vanderbyl, S., Meitz, A., Drayer, J. (2001) Efficient *in vitro* transfer of a 60-Mb mammalian artificial chromosome into murine and hamster cells using cationic lipids and dendrimers, *Chromosome Res.* **9**, 475–485.

Dykes, G.M. (2001) Dendrimers: a review of their appeal and applications, *J. Chem. Technol. Biotechnol.* **76**, 903–918.

Felgner, P.L., Gadek, T.R., Holm, M., et al. (1987) Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7413–7417.

Felgner, P.L., Tsai, Y.J., Sukhu], L., et al. (1995) Improved cationic lipid formulations for *in vivo* gene therapy, *Ann. N. Y. Acad. Sci.* **772**, 126–139.

Feltquate, D.M. (1998) DNA vaccines: vector design, delivery, and antigen presentation, *J. Cell. Biochem.* **30–31**(Suppl.), 304–311.

Ferreira, G.N.M., Monteiro, G.A., Duarte, M.F., Cabral, P., Cabral, J.M.S. (2000) Downstream processing of plasmid DNA for gene therapy and DNA vaccine applications, *Trends Biotechnol.* **18**, 380–938.

Fischer, A. (2001) Gene therapy: some results, many problems to solve, *Cell Mol. Biol.* **47**, 1269–1275.

Flotte, T.R. (1999) Gene therapy for cystic fibrosis, *Curr. Opin. Mol. Ther.* **1**, 510–516.

Gebhart, C.L., Kabanov, A.V. (2001) Evaluation of polyplexes as gene transfer agents, *J. Control. Release* **73**, 401–416.

Gehl, J. (2003) Electroporation: theory and methods, perspectives for drug delivery, gene therapy and research, *Acta Physiol. Scand.* **177**, 437–447.

German Regulation for Gene Technology (Gentechnik Gesetz, GenTG) as proclaimed 16th December 1993. *BGBl* part I: 2066. Last altered Art. 1 G, 21st December 2004; 2005 I 186.

German Regulation for the Protection of Human Embryos (Embryoschutz Gesetz, EschG) as proclaimed 13th December 1990, *BGBl* part I (1990) pp. 2747.

Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., Bujard, H. (1995) Transcriptional activation by tetracyclines in mammalian cells, *Science* **268**, 1766–1769.

Gottschalk, U., Chan, S. (1998) Somatic gene therapy. Present situation and future perspective, *Arzneimittelforschung* **48**, 1111–1120.

Huang, M.T., Gorman, C.M. (1990) Intervening sequences increase efficiency of RNA 3′ processing and accumulation of cytoplasmic RNA, *Nucleic Acids Res.* **18**, 937–947.

Hutchins, B. (2000) Characterization of plasmids and formulations for non-viral gene therapy, *Curr. Opin. Mol. Ther.* **2**, 131–135.

Johnson-Saliba, M., Jans, D.A. (2001) Gene therapy: optimizing DNA delivery to the nucleus, *Curr. Drug Targets* **2**, 371–399.

Kanikkannan, N. (2002) Iontophoresis-based transdermal delivery systems, *BioDrugs* **16**, 339–347.

Kingdon, H.S., Lundblad, R.L. (2002) An adventure in biotechnology: the development of haemophilia A therapeutics – from whole-blood transfusion to recombinant DNA to gene therapy, *Biotechnol. Appl. Biochem.* **35**, 141–148.

Koltover, I., Wagner, K., Safinya, C.R. (2000) DNA condensation in two dimensions, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14046–14051.

Koltover, I., Salditt, T., Radler, J.O., Safinya, C.R. (1998) An inverted hexagonal phase of cationic liposome-DNA complexes related to DNA release and delivery, *Science* **281**, 78–81.

Kozak, M. (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation, *J. Biol. Chem.* **266**, 19867–19870.

Kozak, M. (1992) Regulation of translation in eukaryotic systems, *Ann. Rev. Cell. Biol.* **8**, 197–225.

Kukowska-Latallo, J.F., Bielinska, A.U., Johnson, J., Spindler, R., Tomalia, D.A., Baker, J.R. Jr. (1996) Efficient transfer of genetic material into mammalian cells using Starburst polyamidoamine dendrimers, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4897–4902.

Lee, E.R., Marshall, J., Siegel, C.S., et al. (1996) Detailed analysis of structures and formulations of cationic lipids for efficient gene transfer to the lung, *Hum. Gene Ther.* **7**, 1701–1717.

Li, Z., Ma, Z. (2001) Nonviral gene therapy, *Curr. Gene Ther.* **1**, 201–226.

Lin, A.J., Slack, N.L., Ahmad, A., et al. (2000) Structure and structure-function studies of lipid/plasmid DNA complexes, *J. Drug Target.* **8**, 13–27.

Lollo, C.P., Banaszczyk, M.G., Mullen, P.M., et al. (2002) Poly-L-lysine-based gene delivery systems. Synthesis, purification, and application, *Methods Mol. Med.* **69**, 1–13.

Ma, H., Diamond, S.L. (2001) Nonviral gene therapy and its delivery systems, *Curr. Pharm. Biotechnol.* **2**, 1–17.

Mahato, R.I., Smith, L.C., Rolland, A. (1999) Pharmaceutical perspectives of nonviral gene delivery, *Adv. Genet.* **41**, 95–156.

Malik, N., Wiwattanapatepee, R., Klopsch, R., et al. (2000) Dendrimers: relationship between structure and biocompatibility *in vitro* and preliminary studies on the biodistribution of 125J-labelled polyamidoamine dendrimers *in vivo*, *J. Control. Release* **65**, 133–148.

Mannisto, M., Vanderkerken, S., Toncheva, V., et al. (2002) Structure-activity relationships of poly(L-lysines): effects of pegylation and molecular shape on physicochemical and biological properties in gene delivery, *J. Control. Release* **83**, 169–182.

Marschall, P., Malik, N., Larin, Z. (1999) Transfer of YACs up to 2.3 Mb intact into human cells with polyethylenimine, *Gene Ther.* **6**, 1634–1637.

Merlin, J.L., N'Doye, A., Bouriez, T., Dolivet, G. (2002) Polyethylenimine derivatives as potent nonviral vectors for gene transfer, *Drug News Perspect.* **15**, 445–451.

Midoux, P., LeCam, E., Coulaud, D., Delain, E., Pichon, C. (2002) Histidine containing peptides and polypeptides as nucleic acid vectors, *Somat. Cell. Mol. Genet.* **27**, 27–47.

Newkome, G.R., Moorefield, C.N., Vogtle, F. (2001) *Dendrimers and Dendrons: Concepts, Syntheses, Applications*, Wiley-VCH, New York.

Niculescu-Duvaz, D., Heyes, J., Springer, C.J. (2003) Structure-activity relationship in cationic lipid mediated gene transfection, *Curr. Med. Chem.* **10**, 1233–1261.

Orr, R.M. (2001) Technology evaluation: fomivirsen, ISIS pharmaceuticals Inc/CIBA vision, *Curr. Opin. Mol. Ther.* **3**, 288–294.

Osaka, G., Carey, K., Cuthbertson, A., et al. (1996) Pharmacokinetics, tissue distribution, and expression efficiency of plasmid [33P]DNA following intravenous administration of DNA/cationic lipid complexes in mice: use of a novel radionuclide approach, *J. Pharm. Sci.* **6**, 612–618.

Petersen, H., Fechner, P.M., Marti, A.L., et al. (2002) Polyethylenimine-graft-poly(ethylene glycol) copolymers: influence of copolymer block structure on DNA complexation and biological activities as gene delivery system, *Bioconjug. Chem.* **13**, 845–854.

Plank, C., Tang, M.X., Wolfe, A.R., Szoka, F.C. Jr. (1999) Branched cationic peptides for gene delivery: role of type and number of cationic residues in formation and *in vitro* activity of DNA polyplexes, *Hum. Gene Ther.* **10**, 319–332.

Pouton, C.W., Lucas, P., Thomas, B.J., Uduehi, A.N., Milroy, D.A., Moss, S.H. (1998) Polycation-DNA complexes for gene delivery: a comparison of the biopharmaceutical properties of cationic polypeptides and cationic lipids, *J. Control. Release* **53**, 289–299.

Qin, L., Ding, Y., Pahud, D.R., Chang, E., Imperiale, M.J., Bromberg, J.S. (1997) Promoter attenuation in gene therapy: interferon-gamma and tumor necrosis factor-alpha inhibit transgene expression, *Hum. Gene Ther.* **8**, 2019–2029.

Rädler, J.O., Koltover, I., Salditt, T., Safinya, C.R. (1997) Structure of DNA-cationic liposome complexes: DNA intercalation in multilamellar membranes in distinct interhelical packing regimes, *Science* **275**, 810–814.

Romano, G., Claudio, P.P., Kaiser, H.E., Giordano, A. (1998) Recent advances, prospects and problems in designing new strategies for oligonucleotide and gene delivery in therapy, *In Vivo* **12**, 59–67.

Römpp, L. (1999) in Decjwer, W.D., Pühler, A., Schmid, E. (Eds.) *Biotechnologie und Gentechnik*, 2nd edition, Georg Thieme Verlag, Stuttgart, New York.

Rudolph, C., Plank, C., Lausier, J., Schillinger, U., Müller, R.H., Rosenecker, J. (2003) Oligomers of the arginine-rich motif of the HIV-1 TAT protein are capable of transferring plasmid DNA into cells, *J. Biol. Chem.* **278**, 11411–11418.

Safinya, C.R. (2001) Structures of lipid-DNA complexes: supramolecular assembly and gene delivery, *Curr. Opin. Struct. Biol.* **11**, 440–448.

Satkauskas, S., Bureau, M.F., Puc, M., et al. (2002) Mechanisms of *in vivo* DNA electrotransfer: respective contributions of cell electropermeabilization and DNA electrophoresis, *Mol. Ther.* **5**, 133–140.

Schaffer, D.V., Fidelman, N.A., Dan, N., Lauffenburger, D.A. (2000) Vector unpacking as a potential barrier for receptor-mediated polyplex gene delivery, *Biotechnol. Bioeng.* **67**, 598–606.

Schatzlein, A.G. (2001) Non-viral vectors in cancer gene therapy: principles and progress, *Anti-Cancer Drugs* **12**, 275–304.

Shi, N., Boado, R.J., Pardridge, W.M. (2001) Receptor-mediated gene targeting to tissues *in vivo* following intravenous administration of pegylated immunoliposomes, *Pharm. Res.* **18**, 1091–1095.

Simoes, S., Pires, P., Düzgünes, N., Pedroso de Lima, M.C. (1999) Cationic liposomes as gene transfer vectors: barriers to successful application in gene therapy, *Curr. Opin. Mol. Ther.* **1**, 147–157.

Stevenson, B.J., Carothers, A.D., Wallace, W.A., et al. (1997) Evidence for safety and efficacy of DOTAP cationic liposome-mediated CFTR gene transfer to the nasal epithelium of patients with cystic fibrosis, *Gene Ther.* **4**, 210–218.

Stiriba, S.E., Frey, H., Haag, R. (2002) Dendritic polymers in biomedical applications: from potential to clinical use in diagnostics and therapy, *Angew. Chem., Int. Ed. Engl.* **41**, 1329–1334.

Templeton, N.S. (2003) Myths concerning the use of cationic liposomes *in vivo*, *Expert Opin. Biol. Ther.* **3**, 57–69.

Thanou, M., Florea, B.I., Geldof, M., Junginger, H.E., Borchard, G. (2002) Quaternized chitosan oligomers as novel gene delivery vectors in epithelial cell lines, *Biomaterials* **23**, 153–159.

Tomalia, D.A., Killat, G.R. (1985) in: Kroschwitz, J.I. (Ed.) *Encyclopedia of Polymer Science and Engineering 1*, 2nd edition, Wiley, New York, pp. 680–739.

Tomlinson, E., Rolland, A.P. (1996) Controllable gene therapy: pharmaceutics of non viral gene delivery systems, *J. Control. Release* **39**, 357–372.

URL: http://www.wiley.co.uk/genmed/clinical/ (1st June 2003)

Vacik, J., Dean, B.S., Zimmer, W.E., Dean, D.A. (1998) Cell-specific nuclear import of plasmid DANN, *Gene Ther.* **6**, 1006–1014.

van Steenis, J.H., van Maarseveen, E.M., Verbaan, F.J. et al. (2003) Preparation and characterization of folate-targeted pEG-coated pDMAEMA-based polyplexes, *J. Control. Release* **87**, 167–176.

Verbaan, F. (2002) *Colloidal Gene Delivery Systems – In Vivo Fate of Poly(2-(Dimethyl Amino) Ethyl Methacrylate)-Based Transfection Complexes*, Ph.D.-thesis, Utrecht University, Holland, pp. 14–15.

Wagner, E., Ogris, M., Zauner, W. (1998) Polylysine-based transfection systems utilizing receptor-mediated delivery, *Adv. Drug Deliv. Rev.* **30**, 97–113.

Wang, Y., O'Malley, B.W., Tsai, S.Y. (1997) Inducible system designed for future gene therapy, *Methods Mol. Biol.* **63**, 401–413.

Wattiaux, R., Laurent, N., Wattiaux-De Coninck, S., Jadot, M. (2000) Endosomes, lysosomes: their implication in gene transfer, *Adv. Drug Deliv. Rev.* **41**, 201–208.

Wheeler, C., Felgner, P.L., Tsai, Y.J., et al. (1996) A novel cationic lipid greatly enhances plasmid DNA delivery and expression in mouse lung, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 11454–11459.

Wightman, L., Kircheis, R., Rossler, V., et al. (2001) Different behavior of branched and linear polyethylenimine for gene delivery *in vitro* and *in vivo*, *J. Gene Med.* **3**, 362–372.

Wolff, J.A., Malone, R.W., Williams, P., et al. (1990) Direct gene transfer into mouse muscle *in vivo*, *Science* **247**, 1465–1468.

Yew, N.S., Wysokenski, D.M., Wang, K.X., et al. (1997) Optimization of plasmid vectors for high-level expression in lung epithelial cells, *Hum. Gene Ther.* **8**, 575–584.

Yu, R.Z., Geary, R.S., Leeds, J.M., et al. (1999) Pharmacokinetics and tissue disposition in monkeys of an antisense oligonucleotide inhibitor of Ha-ras encapsulated in stealth liposomes, *Pharm. Res.* **16**, 1309–1315.

Zauner, W., Ogris, M., Wagner, E. (1998) Polylysine-based transfection systems utilizing receptor mediated delivery, *Adv. Drug Deliv. Rev.* **30**, 115–131.

Zou, S.M., Erbacher, P., Remy, J.S., Behr, J.P. (2000) Systemic linear polyethylenimine (L-PEI)-mediated gene delivery in the mouse, *J. Gene Med.* **2**, 128–134.

Zuidam, N.J., Posthuma, G., de Vries, E.T.J., Crommelin, D.J.A., Hennink, W.E., Storm, G. (2000) Effects of physicochemical characteristics of poly(2-(dimethylamino) ethyl methacrylate)-based polyplexes on cellular association and internalization, *J. Drug Target.* **8**, 51–66.

Viruses, DNA Packaging of: ***see*** **Nucleic Acid Packaging of DNA Viruses**

Viruses, RNA Packaging of: ***see*** **Nucleic Acid Packaging of RNA Viruses**

Vitamin Receptors

Krishnamurti Dakshinamurti
University of Manitoba Winnipeg, MB, Canada

1 Introduction 507

2 Vitamin A 508
2.1 Absorption and Transport 508
2.2 Retinoid Receptors: Cellular Binding Proteins 509
2.3 Nuclear Retinoic Acid Receptors 509

3 Vitamin D 511
3.1 General 511
3.2 Vitamin D Receptors 511
3.3 Vitamin D Responsive Genes 512
3.4 Vitamin D Target Cells 512

4 Vitamin E 513
4.1 General 513
4.2 Transport of *d*-Tocopherol 514
4.3 Binding of α-Tocopherol to Erythrocyte Membranes 514
4.4 α-Tocopherol and Cell Signaling 514
4.5 α-Tocopherol and Atherogenesis 514

5 Ascorbic Acid 515
5.1 General 515
5.2 Ascorbic Acid Transport and Transporters 516
5.3 Ascorbic Acid and Extracellular Matrix Production 518
5.4 Ascorbic Acid and Noncollagen Gene Expression 519
5.5 Ascorbic Acid and Differentiation 520

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

6 Biotin 521
6.1 Noncovalent Biotin-binding Proteins 521
6.1.1 Avidins 521
6.1.2 Nonavidin Biotin-binding Proteins 522
6.2 Role of Biotinidase as a Biotin-carrier Protein 524
6.3 Nuclear Biotin-binding Protein 525
6.4 Requirement of Biotin by Cells in Culture 525
6.5 Biotin and Embryological Development 526
6.6 Biotin Requirement for Specific Functions 527
6.7 Effect of Biotin on the Regulation and Synthesis of Specific Proteins 528
6.8 Biotinylation of Histones 529

7 Vitamin B_6 530
7.1 General 530
7.2 Pyridoxal Phosphate–binding Proteins 530
7.3 Nuclear Pyridoxal Phosphate Binding 531
7.4 Function of Pyridoxal Phosphate in the Nucleus 531

Bibliography 532
Books and Reviews 532
Primary Literature 532

Keywords

Receptor
Proteins with a recognition function for a vitamin or its derivative, that is, proteins that bind to the vitamin ligand.

Response Element
The characteristic DNA-binding sites for hormones, vitamins, or their active forms.

Transcription
Synthesis of mRNA or mRNA precursor.

Transcription Factor
Direct-acting factors that control transcription by interacting with DNA sites near the target gene.

Vitamer
Derivatives of a vitamin that function as a vitamin.

Vitamins
Accessory factors of diverse chemical structure required for the maintenance and growth of protozoans and metazoans.

Interlocking ligand–receptor systems constitute an intra- and intercellular communication system in which exogenous or endogenous molecules such as vitamins, hormones, cytokines, and foreign macromolecules convey signals through association with membrane structures (i.e. receptors). Sequential interactions between ligands and receptors on membranes of cells or organelles coordinate metabolism within the same and between different cells and organs. Cell communication is regulated by activation events and by processes that limit such activation.

This article presents current information on the ligand (vitamin)-receptor (binding protein) interaction as it applies to vitamins, from the site of enteric absorption to the site of vitamin function. For the absorption and transport of most vitamins, specific proteins, which have a recognition factor incorporated into them, are involved. At the plasma membrane of the cell, a second recognition factor has to function before the vitamin is internalized. Intracellular vitamin-binding proteins and their interaction with cellular systems resulting in the biological action of the vitamins provide the last step in the successive stages of communication. Thus, we have the concept of a ligand (vitamin), an extraorganismic molecule, reacting with a specific protein entity, the receptor, facilitating absorption and transport of the ligand to its site of action with the resultant sequence of physiological events to which we refer as *vitamin function*. The term ''receptor,'' in the context of vitamins, refers to proteins with a recognition function, not to receptors in the traditional pharmacological sense. It includes proteins that bind to the vitamin ligand and function as extracellular or intracellular transporters as well as the smaller number of vitamin-binding nuclear proteins. In the latter case, vitamins might act as small molecular modulators of protein-DNA interaction, a situation analogous to that of some hormones. The designation ''receptors'' for these categories of vitamin-binding proteins appears to be appropriate inasmuch as it connotes a function more than ligand binding.

The transcriptional patterns of selected genes are often altered by external signals that the cell receives. Thus, changes in nutritional supplies can either induce or repress the transcription of specific genes or gene sets. In many cases, the regulation of transcription (up or down) is mediated by alterations in the availability of essential transcription factors or by change in the factors' ability to bind to specific regulatory sequence motifs.

1 Introduction

Vitamins are required by all protozoans and metazoans that have been studied. They do not conform to any single chemical classification. Some vitamins, such as retinol (vitamin A), cholecalciferol (vitamin D), α-tocopherol (vitamin E), and phylloquinone (vitamin K), are lipid soluble; others, such as thiamine (vitamin B_1), riboflavin (vitamin B_2), niacin, pyridoxine (vitamin B_6), panthothenic acid, lipoic acid, biotin, folic acid, cyanocobalamin (vitamin B_{12}), and ascorbic acid (vitamin C), are watersoluble.

While the unlimited availability of vitamins may have permitted the evolutionary loss of the ability of animals to synthesize vitamins, the ability to absorb vitamins from the environment or from digested food is essential. The modes of uptake and intra- and intercellular transport of these two broad categories of vitamins are different based on their solubility characteristics. Lipoprotein binders are implicated in the transport of the lipid-soluble vitamins. Investigators recognized early, among the water-soluble vitamins, the bindings of cyanocobalamin and folate, respectively, to specific proteins. lncreasing evidence has accumulated to indicate that the absorption of many water-soluble vitamins in the gastrointestinal tract is through saturable mechanisms. In addition, the concentrations of many water-soluble vitamins are much higher in the brain than in plasma. This is true of placental transport as well, implicating specific transporters for individual vitamins across the blood-brain and placental barriers, respectively.

Absorption of vitamins by most cells occurs through membrane-associated transport systems, which are present in low concentrations. These receptorlike proteins may consist of an extracellular vitamin-binding domain linked to a transmembrane anchor. The fertilized egg and embryo obtain vitamins from the external environment using cell surface vitamin receptor/transporter proteins. The vitamins are used directly without storage. In contrast, the eggs of birds and reptiles contain both the vitamins and their respective vitamin-binding proteins deposited in the yolk for transport to the oocyte.

The physiological effects of vitamins follow their entry into the cell, and subsequent metabolic transformation and association with specific apoenzymes, resulting in an accentuation of a metabolic pathway. In this respect, vitamins are metabolic initiators. Where the vitamin function is other than as a coenzyme or prosthetic group of an enzyme, its effect is through regulation of cellular protein synthesis. The mechanisms of action of lipid-soluble vitamins are similar to those of the steroid hormones. Water-soluble vitamins such as biotin, ascorbic acid, and vitamin B_6 also seem to have a modulatory influence on cellular protein synthesis acting through similar mechanisms.

2 Vitamin A

2.1 Absorption and Transport

Vitamin A, an essential nutrient in the diet, plays a significant role in diverse physiological processes such as development, differentiation, and vision. Its role in embryogenesis is evident from certain known effects of a deficiency or excess of vitamin A, namely, malformation or teratogenicity. Vitamin A deficiency has profound effects on the immune system, and retinoids seem to affect differentiation, the stimulation of phagocytosis, and the modulation of cytokines and eicosanoids. Several clinical studies have demonstrated beneficial effects of retinoids on skin diseases such as acne, psoriasis, keratodermas, and skin cancers. Retinoids play a role in the development of the central body axis and limbs of vertebrates. High levels of endogenous retinoids are found in proximity to the developing axis in a variety of species. They exert their wide spectrum of biological functions through their ability to regulate gene expression at specific targets. The

term retinoid includes the natural forms of vitamin A as well as synthetic analogues, which may or may not have the biological activity of vitamin A. The natural forms of vitamin A include retinol, retinal, and retinoic acid (RA).

Ingested retinyl esters are hydrolyzed to retinol by the lipases in the intestinal lumen. Retinoid production from carotenoids occurs by oxidative or eccentric cleavage of the central double bond of carotenoids. The transport of retinol across the brush border of enterocyte is the first step in the absorption of vitamin A into the organism. A carrier-mediated mechanism specific for *all-trans*- and 3-dehydroretinol with passive diffusion of the 13-*cis* compound and retinal has been suggested.

In the intestinal mucosal cell, free retinol is esterified and incorporated into chylomicrons for transport. Chylomicron remnants are primarily cleared by liver, although extrahepatic uptake occurs in other tissues such as bone marrow and spleen. In the liver, retinyl esters from chylomicron remnants are taken up by the parenchymal cells, hydrolyzed to retinol, and bound to retinol-binding protein (RBP). This series of events is followed by paracrine transfer of retinol from hepatocytes to the perisinusoidal stellate cells for storage. Retinol is mobilized from the stellate cells bound to RBP and released into circulation. Retinol-RBP complex binds to transthyretin in the plasma compartment and thus, thyroid hormone and RBP-retinol complex are transported together. Retinol reaches target cells bound to the RBP transport protein. Transport of retinol across cell membrane follows receptor-mediated endocytosis, internalization, degradation of RBP, and liberation of retinol within the cells.

2.2 Retinoid Receptors: Cellular Binding Proteins

Retinol is oxidized to RA and other compounds such as 9-*cis* RA. Within the cell, retinol and RA, respectively, bind to ligand-specific cellular binding proteins: cellular retinol-binding protein (CRBP) and cellular retinoic acid–binding protein (CRABP). These proteins belong to the superfamily of fatty acid–binding proteins. CRABP-I, CRABP-II, and CRBP have been cloned from human and rat sources. CRABP-I and CRABP-II differ in their tissue distribution and affinity for RA. CRABP-II, the major form expressed in human epidermis, is upregulated by RA during differentiation. The promoter of the CRABP-II gene has been shown to have an RA-response element. Retinoid translocation into target cells is regulated by binding to specific high-affinity cellular receptors.

2.3 Nuclear Retinoic Acid Receptors

Almost all the biological activity of vitamin A is due to the ability of its metabolite RA (all-*trans* RA, or atRA) to regulate gene expression, a receptor-mediated event. By analogy with lipid-soluble hormones, the RA receptor is a nuclear-acting transcription factor. These receptors belong to a highly conserved superfamily of proteins with similar structural and functional domains.

Following the isolation of the human RA receptor (hRAR_α), it was shown that its DNA-binding and ligand-binding domain (LBD) was highly conserved with other members of the nuclear receptor family. There are three subtypes of RAR (RAR_α, RAP_β, RAR_γ) in various species, and their LBDs are highly conserved. A second

class of retinoid-responsive transcription factor – the human retinoid X receptor ($hRXR_{\alpha}$), isolated later – had a sequence dissimilar to that of RAR, although both had the ability to respond to RA. RAR and RXR significantly differ in terms of their binding to tritiated all-*trans* RA, leading to the suggestion that a metabolite of RA might be the ligand for RXR. An isomer of RA, 9-*cis*-retinoic acid (9cRA) was identified as this RA metabolite, which was 40 times more potent than RA in transacting RXR. It is also active with RAR. RXRs bind and respond to 9cRA, and they are able to activate target organs in the absence of RAR gene products.

The superfamily of RARs and RXRs share a common structure consisting of functionally distinct domains. The N-terminal A/B domain, which is the least homogenous between the receptors, participates in ligand-independent transactivation function (AF-1). AF-1 synergizes with the AF-2 function of the E domain. Domain C is very highly conserved and is involved in DNA binding. This domain also contains a pair of zinc-binding motifs (zinc fingers) that make contact with the major groove of the DNA. Domain D is the hinge region and is involved in determining the receptor's response as well as binding with corepressors. Domain E, which is also highly conserved, is the LBD and also has the dimerization motifs and ligand-dependent transactivation activity. Domain F is the C-terminal end of RARs and is absent in RXRs.

The presence of two receptor systems suggests that target genes – responsive to each receptor system – are distinct. RA seems to modulate positively the expressions of CRBP and CRABP. CRBP binds retinol with high affinity and is involved in its esterification to retinyl esters. Similarly, CRABP binds to RA and seems to be involved in the oxidative inactivation of RA. An additional regulatory step is the isomerization of RA to 9cRA.

Gene expression is modulated by members of the nuclear receptor superfamily by binding as homo- or heterodimers to specific DNA sequences, the hormone response elements (HREs). Sequence analysis of the response elements activated by RARs and RXRs indicates that similar to the HREs for other nonsteroid members of the nuclear receptor family, the retinoid response elements consist of a minimal half-site consensus sequence (AGGT-CA) configured into a variety of structured motifs, including direct repeats and palindromes. Unlike the heterogeneity of RAREs, RXREs are exclusively of the DR-1 type (a direct repeat separated by one nucleotide). Most of the RXREs discovered seem to occur in promoters of genes involved in vitamin A and lipid metabolism, suggesting a role for RXRs in several distinct but coregulated metabolic pathways. RXR has a role both as a ligand-dependent transcription factor (as homodimer) and as a partner (as heterodimer) to several nuclear receptors. It is interesting that 9cRA, which promotes RXR homodimer formation, is derived from RA, which activates only the RAR heterodimer. RAR-RXR heterodimer can repress RXR homodimer function. Thus, under limiting concentration of RXR, the target genes for RAR will be preferentially activated. When the concentration of 9cRA is high, the RXR homodimer is stabilized. Also, the RAR-RXR heterodimer has a higher affinity for DNA than the RXR homodimer, and might compete more favorably for binding to RXREs. A combinatorial regulation of receptor expression and ligand availability seems to increase the diversity of the signaling pathways. Experiments using knockout mice and cultured cells indicate

that each subtype of RAR and RXR has a distinct effect on target gene expression. For example, RXR_α seems to be involved in determining the life span of hepatocytes. In skin, RXR_α plays an important role in postnatal skin maturation and in hair follicle growth. It is also important in adipogenesis. The cellular retinoid-binding proteins also play a significant role in the retinoid-signaling pathway by regulating the intracellular transport and metabolism of retinoids.

3 Vitamin D

3.1 General

Vitamin D is grouped along with vitamins A, E, and K as one of the fat-soluble vitamins; it is formed in the skin of mammals by ultraviolet irradiation, upon conversion of the precursor 7-dehydrocholesterol to cholecalciferol (vitamin D_3). Its physiological function is related to whole-body calcium homeostasis, inasmuch as a deficiency of this vitamin leads to rickets in children and osteomalacia in adults. Ultraviolet irradiation prevents or heals bone disorders induced by a deficiency.

In the vitamin-deficient state, the absorptive capacity of small intestine for calcium is decreased. One of the physiological functions of vitamin D is the maintenance of a constant extracellular concentration of calcium acting in concert with parathyroid hormone and possibly, calcitonin, thus preventing hypo- or hypercalcemia. Cholecalciferol is converted to 1,25-dihydroxycholecalciferol, the biologically active form of the vitamin. This is a two-step process. The product of the initial hydroxylation, which takes place in the liver, is 25-hydroxycholecalciferol ($25[OH]D_3$). This is secreted into blood, where it circulates bound to a specific vitamin D–binding globulin. The final step takes place in the kidney through a 1-hydroxylation reaction, and the product 1,25-dihydroxycholecalciferol ($1,25[OH]_2D_3$) is the active form of vitamin D. In addition to the role in calcium homeostasis, vitamin D is now recognized to be involved in regulating growth and differentiation of a variety of cell types.

3.2 Vitamin D Receptors

Most of the biological actions of vitamin D occur through a $1,25(OH)_2D_3$-mediated regulation of the transcription of specific genes in target tissues. This active form of vitamin D binds to a specific nuclear receptor (VDR), which belongs to the superfamily of steroid, thyroid hormone (TR), and vitamin A (RAR) receptors. Ligand binding is necessary for the high-affinity binding of VDR to specific vitamin D DNA response elements (VDREs) located in the promoter region of the target genes. This is similar to the binding of TR and RAR to their cognate receptors, since TR, RAR, and VDR belong to the same subfamily of receptors. Interaction between the ligand-receptor complex with VDREs modulates the transcription of genes specific for the biological action of vitamin D.

Osteoporosis is an age-related loss of bone mineral. Bone mineral mass is affected by both genetic and nongenetic factors. A correlation between alleles of VDR and bone density has been identified. In bone, $1,25(OH)_2D_3$ recruits bone-resorbing osteoclasts from hematopoietic precursor cells. It also modulates, through VDR, the bone-forming osteoblasts. Restriction fragment length polymorphism

analysis indicates the presence of two alleles of VDR. In a study of 250 Caucasian twins and a second study of 311 unrelated healthy women, it was shown that common allelic variants of the VDR gene can be used to predict differences in bone mineral density, hence susceptibility to osteoporosis. The magnitude of the VDR allele difference between the two genotypes could mean a 10-year difference for postmenopausal women in the onset of vertebrate fractures.

3.3 Vitamin D Responsive Genes

Vitamin D acting through $1,25(OH)_2D_3$ enhances intestinal calcium absorption by stimulating transcellular calcium transport. The stimulation is achieved by both nongenomic and genomic actions of vitamin D. One such genomic action is the induction of the calcium-binding protein calbindin-D, which could facilitate the diffusion of calcium from the brush border to the basolateral membrane of the enterocyte. Other vitamin D–responsive calcium-binding proteins such as Ca^{2+}-ATPase and the integral membrane calcium-binding protein have been characterized. A list of 51 gene products that are reported to be $1,2(OH)D_3$ sensitive in terms of altered mRNA levels has been assembled. This list includes genes associated in the mineral homeostasis, vitamin D metabolism, cell differentiation and proliferation, bone matrix protein, extracellular matrix proteins, oncogenes, chromosomal proteins, growth factors, signal transduction proteins, peptide hormones, and energy metabolism. In clonal rat pituitary tumor (GH) cells in culture, $1,25(OH)_2D_3$ selectively increased prolactin production, prolactin mRNA, and agonist-induced prolactin release. In primary cultures of rat anterior pituitary cells, $1,25(OH)_2D_3$ selectively enhanced TRH-induced thyrotropin (TSH) release. In studies using primary dispersed and clonal pituitary cells, $1,25(OH)_2D_3$ has been shown to stimulate pituitary hormone production and agonist-induced hormone release. There was an upregulation of thyrotropin-releasing hormone (TRH)-binding sites as shown by an increase in TRH receptor–specific mRNA. Incubation of pituitary GH4C1 cells with vitamin D markedly reduces the response of rat growth hormone (GH) mRNA to thyroid hormone (T3) and RA. Rat GH gene expression is stimulated by T3. The 5′-flanking region of rat GE gene contains a TRE that mediates T3 stimulation. RA also increases the transcription of the GH gene. The RA response was mapped to the same sequence that mediated regulation by T3, indicating that TRE also functioned as RARE. Vitamin D inhibited the T3 and RA transactivation of the rat GH gene by interference with the common response element present in the 5′-flanking region of the gene. Vitamin D also inhibited the activation of other T3- and RA-responsive promoters. The formation of TR-VDR and RAR-VDR heterodimers, which have low affinities for these response elements, can inhibit binding of more active heterodimers with RXR. These results highlight the repressive role of VDR and indicate that the biological response to T3 and RA could be selectively inhibited in cells that contain receptors for vitamin D.

3.4 Vitamin D Target Cells

It has been established that many tissues other than the kidney convert $25(OH)D_3$, to $1,25(OH)_2D_3$ although it is not released into the blood stream. They seem to act in

an autocrine or paracrine fashion by binding locally to VDR and affecting cell proliferation, differentiation or apoptosis locally. Such target tissues include the central nervous system, skin and hair follicles, and immune and endocrine systems. Vitamin D modulates this growth regulatory pathways of many normal and transformed cells. Treatment of cells with $1,25(OH)_2D_3$ induces antiproliferative and/or proapoptotic effects or differentiation.

In many transformed cells including breast, prostrate, and leukemic cells, $1,25(OH)_2D_3$ induces cell cycle arrest in Go/Gi. This arrest is associated with the inhibition of cyclin-dependent kinases. In colon cancer cells, the VDR ligand complex interacts with protein phosphatases, which inactivate p70S6 kinase that is essential for Gi/S phase transition.

$1,25(OH)_2D_3$ induces apoptosis in cell lines derived from breast cancer, prostrate cancer, myeloma, retinoblastoma, neuroblastoma, and glioma. Its actions are also synergistic with other triggers of apoptosis. The sensitivity to $1,25(OH)_2D_3$-mediated apoptosis is related to the expression as well as the subcellular localization of the BCl-2 family of pro- and antiapoptotic proteins.

Vitamin D is a regulator of proliferation, differentiation, and apoptosis of cells of different lineage, and growth inhibition of tumors in response to vitamin D or its analogs has much therapeutic potential.

4 Vitamin E

4.1 General

"Vitamin E" is the generic term for two groups of naturally occurring fat-soluble compounds, tocopherols and tocotrienols. The members of each group are designated α, β, γ, or δ, depending on the numbers and position of the methyl groups attached to the chroman nucleus. The most biologically active of these compounds is α-tocopherol, which functions as an antioxidant in membranes to prevent lipid peroxidation. The intake of γ-tocopherol is two to four times that of α-tocopherol because of its prevalence in many diets. Both α- and γ-tocopherol are absorbed equally by the intestine and transferred to the bloodstream by chylomicrons, which are taken up by the liver. However, α-tocopherol is present in human plasma at a concentration many times higher than that of γ-tocopherol, suggesting a mechanism of discrimination in the liver. This accounts for the generally recognized lower biopotency of γ-tocopherol.

It has been suggested that specific membrane tocopherol-binding proteins are involved in peripheral cellular uptake of tocopherol. A 46-KDa tocopherol associated protein (TAP) with substantial sequence homology to α-TPP is expressed in most tissues with high levels in liver, brain, and prostate.

α-TTP has been purified and characterized and shown to consist of two closely similar isoforms of molecular mass 30.5 KDa with high specificity for binding to the α isomer of tocopherol. This protein is present in rat liver and absent in other rat tissues such as heart, spleen, testes, and lung. The α-TTP gene was sequenced and shown to have homology with bovine and human 11-*cis*-retinaldehyde-binding protein of the retina. It has been suggested that the liver α-TTP, in view of its specificity for the α isomer, discriminates against the γ-tocopherol, which is excreted into the bile. The α-tocopherol is transferred to

very low density lipoprotein (VLDL). A familial vitamin E deficiency, which results in spinocerebellar lesions with progressive development of ataxia and areflexia, has been described. Vitamin E is almost absent in the plasma of these patients, although lipid absorption, pancreatic, and intestinal functions, as well as lipoprotein transport, were normal. The only abnormality was in the transfer of α-tocopherol into VLDL, with the suggestion of an abnormality in the α-TTP of these patients.

4.2
Transport of *d*-Tocopherol

There does not seem to be a special carrier for Vitamin E in circulation. It is incorporated into lipoprotein particles in a nonspecific manner. In liver cells, tocopherol is transferred from chylomicron remnants to cytosolic α-tocopherol transfer protein (α-TTP), which mediates its transfer to rough endoplasmic reticulum and the golgi, the site of VLDL synthesis. α-TTP is stereo specific so that only α-tocopherol is incorporated into the nascent VLDL particles.

4.3
Binding of α-Tocopherol to Erythrocyte Membranes

α-Tocopherol has been shown to decrease the hemolytic damage to erythrocytes by oxidative stress. Since the disruption of erythrocyte membrane is the likely locus of oxidative stress, and since α-tocopherol has been proposed to have a role in membrane structure and function, it is possible that α-tocopherol-binding proteins or receptors are present on these membranes. Red blood cell membranes have specific binding sites for α-tocopherol. These binding sites have the properties expected of membrane receptors (i.e. specificity, saturability, moderate affinity, time-, protein-, and temperature-dependence, reversibility). The binding of α-tocopherol to the membrane is highly specific, inasmuch as only γ-tocopherol competes with α-tocopherol for binding to the membrane.

4.4
α-Tocopherol and Cell Signaling

In terms of function, the antioxidant properties of α-tocopherol have received most attention. The nonantioxidant cell-signaling roles for α-tocopherol have now been recognized. Smooth muscle cell proliferation and protein kinase C (PKC) activity are inhibited by the α isoform, whereas other isomers and related chemicals with similar antioxidant properties are inactive. The inhibition of PKC by α-tocopherol is seen in other cell types as well. The inhibition of PKC by α-tocopherol is related to its activation of a protein phosphatase 2A. This phosphatase dephosphorylates PKC-α and inhibits its activity.

4.5
α-Tocopherol and Atherogenesis

Atherosclerosis, a major cause of cardiovascular disease, is an inflammatory process. Proinflammatory mediators such as prostaglandin E (PGE_2) and thromboxane A_2 (TXA_2) are important in the development of atherosclerosis. The age-associated increase in PGE_2 is the result of increased cyclooxygenase activity (COX), which converts arachidonic acid into PGE_2, the precursor of thromboxane and eicosanoids. The activity of the COX-2 isoform is regulated by growth factors, tumor promoters, cytokines, glucocorticoids, and lipopolysaccharides. Increased arachidonic acid metabolism by COX stimulates

cell proliferation through activation of PKC. α-tocopherol has been shown to decrease PGE_2 production through a decrease in COX activity, a posttranslational effect of vitamin E. Other homologs of vitamin E are also effective to different extents in inhibiting COX activity, although this difference is not correlated to their antioxidant potency. A direct inhibition of COX-2 by the tocopherols is indicated.

The expression of a spectrum of genes involved in the inflammatory and proliferative response of cells is controlled by the nuclear factor kappa B (NF-kB). This is present as a dimer in an inactive form while attached to an inhibitory protein. Activation by inflammatory or proliferative cytokines, reactive oxygen species, or bacterial wall components is followed by phosphorylation and proteolytic removal of the inhibitory component. The active NF-kB transfers to the nucleus where it interacts with the regulatory kappa B elements in the promoter and enhancer regions, thus regulating transcription of inducible genes. Incubation of human Jurkat T-cells with α-tocopheryl acetate or succinate resulted in a dose-dependent inhibition of NF-kB activation. Macrophage cell line THP-1 pretreated with α-tocopherol and then activated with lipopolysaccharide also showed a similar inhibition of NF-kB activation, thus indicating that α-tocopherol by itself inhibits NF-kB activation and/or translocation to the nucleus. Activated NF-kB has been identified in atherosclerotic plaques. It has been shown to regulate genes involved in development of atherosclerosis. NF-kB is also regulated by cellular redox status. Vitamin E, primarily the α-tocopherol isoform, thus regulates NF-kB activity either directly by modulating the key steps in the pathway of activation of NF-kB or indirectly by altering the intracellular redox status.

α-tocopherol has been shown to be involved in the expression of α-tropomyosin, cell adhesion proteins, and scavenger receptors. Soon after balloon injury, smooth muscle cells that migrate to the intima have decreased amounts of tropomyosin, which returns to normal levels later. α-tocopherol, but not the β-isomer, has been shown to induce tropomyosin in rat smooth muscle cells and thus, decrease the contractility of smooth muscle cells. Activation of endothelial cells results in the release of vascular cytokines, which induce the expression of cell surface adhesion molecules such as VCAM-1 and ICAM-1, which are necessary for the recruitment of neutrophils. The expression of the adhesion molecules is dependent on the activation of NF-kB. Treatment of endothelial cells with oxidized LDL transcriptionally increases the levels of VCAM1. Pretreatment of endothelial cells with α-tocopherol inhibited, dose-dependently, the oxidized LDL-induced expression of adhesion molecules, indicating that the antiatherogenic action of vitamin E might be related to its downregulation of cell adhesion molecules. CD36 is a scavenger receptor that transports oxidized LDL into the cytosol. α-tocopherol has been shown to downregulate the expression of CD36 and thus to inhibit the transport of oxidized LDL with resultant reduction in the formation of foam cells.

5 Ascorbic Acid

5.1 General

The antioxidant activity of ascorbic acid in all living organisms is the best-recognized function of this vitamin. Vitamin C and

vitamin E eliminate cytotoxic free radicals by redox cycling. Ascorbate and vitamin E function at physiological pH as donors of single hydrogen atoms, enabling these vitamins to react efficiently with free radicals and not with molecular oxygen. Ascorbic acid is a cofactor in the synthesis of collagen and proteoglycans, the two major constituents of the extracellular matrix. The synthesis of collagen by fibroblasts involves enzymes that hydroxylate lysine and proline in the procollagen molecule. Ascorbate has also been shown to be a cofactor of the enzyme peptidylglycine α–aminating monooxygenase, which amidates neuroendocrine peptides. The concentration of ascorbate in brain and endocrine glands is high. In addition, the presence of a complex system to regulate its intracellular and intravesicular availability indicates a critical role for ascorbate in these structures. Ascorbate is released from brain and endocrine tissues and can affect cellular functions by acting on plasma membrane receptors.

5.2 Ascorbic Acid Transport and Transporters

Most animals, exception primates and guinea pig, synthesize ascorbate from glucose in the liver. Hence, intestinal transport of ascorbate has special significance in primates and the guinea pig as the source of ascorbic acid. Vitamin C (L-dihydroascorbic acid) is a stereospecific compound. The L stereoisomer is the biologically active compound: the D form has only one-twentieth its biological activity. Once inside the cell, the D epimer (D-isoascorbic acid) is as efficient as the L form, at least in one aspect of its biological action, that of collagen synthesis. Thus, the cell membrane is the primary site of stereoselectivity. The carrier proteins in the cell membrane are stereospecific in terms of binding and transporting the vitamin across the cell membrane to its intracellular site of action.

Dietary vitamin C is a mixture of ascorbate and dehydroascorbic acid. Both compounds are antiscorbutic. Dehydroascorbic acid, the oxidized form, is not an acid, does not ionize, and is more hydrophobic than ascorbic acid. A net transepithelial absorption and transport of this form was shown in everted and noneverted loops of guinea pig intestine. The vitamer reaching the serosal side is in the reduced state. Such simultaneous absorption and reduction is characteristic of species that do not synthesize ascorbic acid. On the contrary, there is no net absorption of dehydroascorbic acid in the intestine of the rat, a species capable of synthesizing ascorbic acid in the liver.

In contrast to the intestinal transport of ascorbate by various mammalian species, renal handling of this substance seems to be very similar. Ascorbate is filtered in the glomerulus and is reabsorbed in the proximal convoluted tubule. The uptake was saturable, inhibited by D-iso-ascorbate and dehydroascorbic acid, temperature dependent, and susceptible to metabolic inhibitors. Ascorbate is reabsorbed by the kidney in a sodium-dependent active process not shared by other acid anions.

When incubated in a medium containing physiological amounts of ascorbic acid, human neutrophils accumulate millimolar amounts using two transporters. The high-affinity transporter with an apparent Michaelis constant K_m of 2 to 5 μM is saturated at normal extracellular ascorbic acid concentrations. The low-affinity transporter has an apparent K_m of 6 to 7 mM, and is in the linear phase of uptake at normal extracellular concentrations of ascorbic acid. The intracellular concentration of ascorbic acid in neutrophils, which

is dependent on the extracellular concentration, functions to preserve neutrophil integrity. It also protects host tissue by inactivating free radicals and oxidants produced by the phagolysosomes.

Ascorbic acid is essential for the formation of bone by osteoblasts. *In vivo* studies indicate that administered ^{14}C-labeled ascorbic acid accumulates preferentially at sites of active osteogenesis. *In vitro* studies have shown that ascorbic acid is essential for the formation of mineralized extracellular matrix by rabbit long bone cells and rat calvaria cells. Two cell systems, a clonal osteoblastlike cell line ROS 17/2.8 and a primary culture of rat calvaria cells, have been used as *in vitro* osteoblast models. Both cell cultures accumulate ascorbate. Consistent with a carrier-mediated mechanism, transport was saturable, temperature dependent, stereo selective, and dependent on external Na^+ concentration. Ascorbate taken up is rapidly compartmentalized in the lumen of the rough endoplasmic reticulum and is bound or metabolized to a nonpermeant species, in keeping with its putative role as a cofactor in the posttranslational modification of procollagen. Using the clonal osteosarcoma cell line UMR-106, it was shown that transforming growth factor β (TGFB), an important modulator of cell proliferation and differentiation, stimulates carrier-mediated ascorbate transport. Stimulation of ascorbate uptake was seen at TGFB concentrations similar to that required for stimulating the production of collagenous extracellular matrix. The enhancement by TGFB of ascorbate uptake was associated with an increase in the apparent V_{max} of the rate of transport without any change in the affinity of the transporter for ascorbic acid, indicating that either the number of functional transporters or the rate at which individual transporters translocate ascorbate is increased.

The effect of TGFB is completely blocked by cycloheximide. This would suggest that TGFB increases the rate of synthesis of the Na^+ ascorbate cotransporter or a regulatory protein that interacts with the existing transporters to increase their turnover number.

Ascorbic acid is transported into 3T6 fibroblasts, which synthesize collagen, by a carrier-mediated saturable active Na^+ cotransporter (SVCT) process that appears to be the mechanism of transport in most cell types. Ascorbic acid, the antiscorbutic form, stimulates chondrocyte expression of SVCT, which transports the L form more efficiently than the D-isoform. It was observed that serum from various animals contained a heat-labile factor that interacted with bacterial endotoxin and inhibited ascorbate transport. The inhibitor may have been generated during endotoxin-induced activation of complement in unheated serum. The inhibitor has been purified from insulin-activated human serum. This heat-stable serum protein factor, which is produced when serum complement is activated, has a molecular mass of 9000 Da. The reactivity of the purified inhibitor with antiserum to C3a suggests that the inhibitor is C3a or a very similar molecule. The inhibitor of ascorbate transport increases the apparent K_m for ascorbate without affecting the V_{max}. In view of the similarity of ascorbate transport in many cell types, it is possible that this protein can inhibit ascorbate transport in other cell types as well, resulting in depletion of tissue ascorbate during infection or autoimmune process when serum complement is activated.

Following intracardiac injection of [^{14}C]ascorbic acid into guinea pigs, the neurohypophysis takes up the label to

a high concentration (exceeded only by the adrenals). The uptake of [^{14}C]ascorbic acid by isolated nerve endings from ox neurohypophysis was demonstrated to be a saturable process, suggesting the presence of both glucose-dependent and sodium-dependent uptakes. Cortisol and triiodothyronine inhibited the uptake. Posttranslational processing of vasopressin and oxytocin precursors involves a trypsinlike enzyme, a carboxypeptidase, and an amidating enzyme (peptidylglycyl-α-amino monooxygenase (PAM)). PAM, which requires copper, molecular oxygen, and ascorbic acid, may be limiting in the processing of amidated peptides. In a study of pancreatic islet cells, in which several peptides (e.g. pancreastatin, TRH, pancreatic polypeptide, amylin) are C-terminally amidated, the presence of PAM activity has been reported. Research has demonstrated the presence of PAM, a high content of ascorbic acid, as well as a carrier-mediated process for its uptake in the islet cells. As in neurohypophyseal nerve terminals, ascorbate uptake was inhibited by triiodothyronine and glucocorticoids.

5.3 Ascorbic Acid and Extracellular Matrix Production

Collagen is a structural protein found in tissues such as bone, cartilage, tendon, and skin. The collagens are characterized by a triple-helical structure. For the three peptides to form a stable helix, a significant number of proline residues in the peptide chain have to be hydroxylated by the enzyme prolyl hydroxylase, which requires ascorbic acid. Elastin, another ECM (extracellular matrix) protein, is found in elastic tissues and also contains hydroxylated proline residues. In addition to its role in the hydroxylation reaction, vitamin C stimulates the synthesis of collagen in a number of cell types. Elastin synthesis is also affected by Vitamin C.

Vitamin C increases the amount of procollagen mRNA by stimulating the transcription of the procollagen gene and by stabilizing the mRNA. Ascorbic acid treatment of human skin fibroblasts increased procollagen mRNA twofold. In primary avian tendon (PAT) cells, there was a similar sixfold increase of procollagen mRNA. In PAT cells treated with α,α-dipyridyl, an inhibitor of hydroxylase, there was a significant decrease in the rate of procollagen synthesis without any inhibitory effect on procollagen mRNA. These results suggest that ascorbic acid might directly increase the transcription of the procollagen gene through interaction with an ascorbic acid–specific cis-regulatory element on the procollagen genes.

Cartilage is replaced by bone-forming cells during endochondral longitudinal bone growth at the epiphysis, and fracture healing. During the developmental process, cartilage tissue undergoes orderly morphological changes involving proliferation, cellular hypertrophy, production of ECM, and mineralization. Immature cells produce type I and type II collagen, and fibronectin. Proliferating chondrocytes synthesize and secrete large amounts of type II and type IX collagen, as well as cartilage-specific proteoglycans. Hypertrophic, nondividing cells produce type X collagen. Mineralizing tissue expresses alkaline phosphatase and type X collagen. Those changes in collagen types are accompanied by changes in their mRNA level. Proliferation of prehypertorphic cells requires vitamin C. The increase in alkaline phosphatase activity followed the expression of collagen type X. Mineralization in chondrocytes was also ascorbate dependent. Although vitamin C is not necessary

for general cell growth, it is essential for the production of ECM.

5.4 Ascorbic Acid and Noncollagen Gene Expression

Related to the role of ascorbic acid in normal ECM production is the report of the action of ascorbic acid on type IV collagenase (matrix metalloproteinase-2, MMP-2). A low intake of vitamin C during pregnancy is associated with a high risk of premature rupture of the membrane (PROM). A new effect of ascorbic acid as a modulator of MMP-2 gene expression at the level of transcription has been reported, and it suggests that collagen degradation might be a significant factor in PROM.

Skin fibroblasts from patients with Ehlers–Danlos syndrome type VI (EDS VI), an inherited connective tissue disorder, have decreased activity of lysyl hydroxylase (LH). LH hydroxylates lysine residues in peptide linkage, a key step in posttranslational modification of collagen. EDS patients with LH deficiency have extreme joint hypermobility, skin hyperextensibility, bruisability, and delayed wound healing. Treatment of skin fibroblasts from EDS patients with ascorbic acid increased LH levels, which paralleled LH mRNA levels. Ascorbate increased total collagen production twofold.

Ascorbic acid deficiency was studied in ODS rats (genotype od/od) with an inherited defect in ascorbic acid synthesis. Ascorbic acid deficiency decreased the serum apolipoprotein A-1 concentration, mediated by a decrease in its mRNA levels. Ascorbic acid deficiency also decreased the renal level of fatty acid–binding protein and the serum level of α_{2M}-globulin. The hepatic level of α_{2M}-globulin mRNA of the ascorbic acid–deficient rat was significantly reduced. The kidney fatty acid–binding protein is a proteolytic fragment of α_{2M}-globulin, which is synthesized in the liver. In further work, the effects of ascorbic acid deficiency in the ODS rats on the hepatic gene expression of the positive acute phase proteins, haptoglobin and α_1-acid glycoprotein, as well as negative acute phase proteins, apolipoprotein A-1 and albumin, were studied. Ascorbic acid deficiency significantly elevated the serum concentration of haptoglobin and lowered the concentrations of apolipoprotein A-1 and albumin. The increases in haptoglobin and α_1-acid glycoprotein corresponded to increases in their mRNA levels. Correspondingly, the decreases in apolipoprotein A-1 and albumin matched decreases in hepatic mRNA levels of these proteins in ascorbic acid–deficient ODS rats.

Ascorbic acid has a significant role in catecholamine synthesis in the sympathetic nervous system as a cofactor of dopamine-β-hydroxylase. Incubation of neuroblastoma cell line SK-N-SH with ascorbic acid for 2 h results in enhanced synthesis of 3,4-dihydroxyphenylalanine (DOPA) and dopamine. After five days of incubation of the cells with high levels of ascorbic acid, there was a threefold increase of tyrosine hydroxylase gene expression, whereas dopamine β-hydroxylase gene expression was not altered. The results suggest that ascorbic acid might be a useful adjunct in the treatment of early Parkinson's disease.

Ascorbic acid deficiency in guinea pigs caused decreases in the amount of various forms of cytochrome P450 in liver microsomes. Ascorbic acid deficiency decreases the expression of cytochrome P450 mRNA, specifically for the 1A1 and 1A2 subtypes, indicating that transcription of P450 is regulated by ascorbic acid in guinea pigs.

Vitamin C inhibits NF-κB activation by tumour necrosis factor (TNF) through activation of p38 mitogen-activated protein kinase (MAPK). NF-κB is a eukaryotic transcription factor, which regulates the expression of a number of genes involved in immune and inflammatory responses. It is activated by pathogenic signals including the proinflammatory cytokines interleukin-1 (IL-1) and TNF. In the cell cytosol, NF-κB exists in a latent form consisting of a transcriptionally active dimer (p50-p65) bound to an inhibitor protein IκB. Upon stimulation by cytokine inducers, IκB is phosphorylated by IκB kinase (IKK) and is then ubiquitinated and degraded by 26S proteosomes. The released NF-κB dimer then translocates to the nucleus, binds to κB elements in the promoters, and activates target gene expression. Vitamin C inhibits cytokine stimulation of NF-κB. This is not due to an antioxidant effect of vitamin C as other antioxidants do not mimic or augment the action of vitamin C. The effect of vitamin C is mediated through activation of stress-activated protein kinase p38 MAPK. It is not known yet how vitamin C activates p38 MAPK. Activated p38 MAPK inhibits TNF activation of NF-κB as well as the phosphorylation of IκB and activation of IKK. These actions of vitamin C correspond closely to those of vitamin E on NF-κB activation.

A role for vitamin C in augmenting the chemotherapeutic response of cervical carcinoma HeLa cells has been proposed. Human papilloma virus (HPV) is involved in the pathogenesis of cervical cancer. The oncopotential of high-risk GPV types is due to E6 and E7 oncoproteins, which transform the infected cell. The E6 oncoprotein targets p38 protein, which is ubiquitinated and degraded. Cell cycle regulation is affected. These oncoproteins also activate telomerase, contributing to the immortalization of transformed cells. HPV-positive HeLa cells were incubated with millimolar concentrations of vitamin C to study its effect on the transcriptional regulation of HPV E6/7 oncogene expression. Vitmain C treatment resulted in a downregulation of the viral oncoprotein E6, which paralleled a decrease of activator protein (AP-1) members c-jun and c-fos in a time- and dose-dependent manner. The downregulation of E6 was associated with an upregulation of the proapoptotic p53 and Bax protein, and a downregulation of apoptosis inhibitor Bcl-2.

Vitamin C mediates cellular responses to counter ultraviolet (UV) radiation-mediated cell damage and cell death. Ascorbic acid rescues keratinocytes from UV-mediated cytotoxicity directly by scavenging free radical species. In addition, ascorbic acid induces transcriptional downregulation of IL-1α mRNA. Oxidative stress due to UV radiation activates redox-sensitive transcription factors including NF-κB and members of the AP-1 complex such as c-jun and c-fos. Ascorbate inhibits the AP-1-dependent transactivation of specific promoters. Ascorbic acid is a negative regulator of AP-1 and AP-1-dependent transcription.

5.5 Ascorbic Acid and Differentiation

Embryonic stem cells can differentiate into specialized cells including cardiac myocytes. In a study of the effect of inducers of cardiac differentiation, additional treatment of these cells with ascorbic acid markedly enhanced differentiation with the induced cells displaying spontaneous and rhythmic contractile activity and also staining positively for sarcomeric myosin and α-actin. Ascorbic acid also induced the

expression of cardiac genes in a developmentally controlled manner. This effect of L-ascorbic acid was not mimicked by other antioxidants including vitamin E.

Ascorbic acid has been shown to enhance the expression of several genes in human keratinocytes. The upregulation of mut L homologue-1 (MLH1) and its downstream target p73, were seen as early as 2 h after treatment with ascorbic acid. Modulation of MLH1 and p73 gene expression improves cellular susceptibility to apoptosis triggered by DNA-damaging agents, and offers a potential mechanism of the anticarcinogenic action of vitamin C. Combinations of vitamin C and vitamin E modulate cellular antioxidant network and block survivin gene expression in androgen-responsive as well as in nonresponsive human prostate cancer cell lines through inducing apoptosis via suppression of survivin, an antiapoptotic factor.

6 Biotin

6.1 Noncovalent Biotin-binding Proteins

Biotin in most biological materials is covalently bound to a lysine residue of proteins (carboxylases). The noncovalent binding of biotin to various proteins seems to be crucial for the uptake, transport, and intracellular functions other than the prosthetic group functions of the vitamin.

A group of proteins bind to biotin noncovalently. Both avidin (the biotin-binding protein of egg white) and strepavidin (the extracellular biotin-binding protein of *Streptomyces avidinii*) have exceedingly high affinities for biotin, with a dissociation constant K_3 of 10^{-15} M. These are the strongest noncovalent bindings known between a protein and a small molecular weight ligand. Other biotin-binding proteins (e.g. the egg yolk biotin-binding proteins, biotinidase, biotin-holocarboxylase synthetase) have progressively lower affinities for biotin. Biotin-binding proteins vary considerably along the evolutionary scale. In *Escherichia coli*, the biotin-binding protein biotin-holocarboxylase synthetase has been shown to regulate the synthesis of the enzymes of biotin biosynthesis pathway.

In the laying hen, there are three biotin-binding proteins. In addition to avidin, which is present in albumen of egg, there are two biotin-binding proteins in egg yolk. The presence of distinct biotin-binding proteins in the yolk suggests a specialized role for biotin in the regulation of the development of the embryo. In mammals, only two proteins have so far been shown to bind noncovalently to biotin.

6.1.1 Avidins

Avidin has been isolated from hen egg white and the oviduct of the avian species. It is a basic glycoprotein with a molecular mass of 67 kDa and an isoelectric pH of 10.5. It is a tetramer of four identical subunits, each with 128 amino acid residues and one biotin-binding site. The composition and structural arrangement of the carbohydrate moiety of avidin has been shown to be heterogeneous. Strepavidin is a nonglycosylated neutral protein secreted by *Streptomyces avidinii*. Its binding to biotin is remarkably similar to that of avidin, with a dissociation constant of 4×10^{-14} M. Both avidin and strepavidin are tetrameric, with each subunit containing one biotin-binding site. However, they lack immunological cross-reactivity and evolutionary relatedness. Avidin is a secretory protein produced by the oviduct and concentrated in the egg white. It is

induced by progesterone. In addition to being an oviductal secretory protein, avidin is synthesized in areas of tissue damage and inflammation caused by diverse agents such as toxic doses of actinomycin D, septic infection, or thermal injury in chicken. This progesterone-independent synthesis of avidin is inhibited by antiinflammatory steroids.

The avidin-biotin system has been used in a wide range of biological studies including isolation, localization, cytochemistry, immunochemistry, and diagnostics, as well as for probing genes. Since there is some nonspecific adsorption of avidin, strepavidin is generally used instead. However, strepavidin has been shown to bind with high affinity to cell surfaces. Strepavidin itself can displace fibronectin from cell surfaces. The RYD (Arg-Tyr-Asp) sequence of strepavidin mimics the RGD (Arg-Gly-Asp) sequences, which is a universal recognition domain for adhesion receptors, as well as for the receptors for fibronectin, fibrinogen, and other adhesive proteins.

6.1.2 Nonavidin Biotin-binding Proteins

In addition to avidin and strepavidin, the two examples of proteins with exceptionally strong affinity for biotin, other biotin-binding proteins have lower affinities for biotin. These include the egg yolk biotin-binding proteins: Biotin-holocarboxylase synthetase and biotinidase.

Biotin-Holocarboxylase Synthetase. Biotin carboxylases are synthesized in the form of the apoproteins that undergo posttranslational covalent modification by the addition of biotin, the prosthetic group, to the ε-amino group of lysine of the apoprotein. This covalent attachment of biotin to a specific lysine in the apoenzyme is catalyzed by biotin-holocarboxylase synthetase in a two-step reaction. The synthetase catalyzes the biotination of apocarboxylase from heterologous sources.

The biotin locus in *E. coli* was shown to be located in the same region as *gal* and *att*, and the order was determined to be *gal, att, bio*. The biotin locus is composed of five consecutive genes in the order *A B F C D* (representing the genes coding for various enzymes of biotin biosynthesis) and three unlinked genes *bioH, bioR* (a regulatory gene coding for the biotin repressor protein), and *bioP* (a permeability gene). Transcription of this cluster of genes is from two "face-to-face" partially overlapping promoters. One transcript includes the *Bio F C D* genes, whereas the second transcript includes *bioA* gene and an open reading frame of unknown function. Transcription from both promoters was blocked *in vivo* upon addition of high levels of biotin to the medium of *E. coli* cultures. Starvation of *E. coli* biotin auxotrophs resulted in increased transcription. In the biotin operon, the repressor is the biotin-protein ligase and the corepressor is biotinyl AMP. Maximal rates of biotin operon transcription are associated with conditions of biotin limitation.

Biotinidase. Acid hydrolysis of biotin proteins releases free biotin, whereas proteolytic hydrolysis yields biotin peptides. The smallest among these is biocytin, ε-*N*-biotinyl-L-lysine. The name biotinidase (biotin-amide amidohydrolase, EC 3.5.1.12) was proposed for this enzyme, which also hydrolyzes biocytin. The rat liver enzyme is associated with membranes, although not an integral part of the membrane structure. The enzymes extracted from diverse sources such as rat liver and human serum are similar in many of their properties (e.g. optimal pH

and K_m for substrate, acidic nature, inhibition by chaotropic agents and sulfhydryl agents). The major difference between these two proteins is in the heat stability of the rat liver enzyme. It appears that the liver is the site of synthesis of the serum enzyme. Human serum biotinidase has been purified to homogeneity. The enzyme is a glycoprotein monomer with a molecular mass of 68 kDa. It contains sialic acid residues. However, these are not required for enzyme activity. Although the predicted amino acid sequence from the isolated biotinidase clone did not share extensive homology with avidin and strepavidin, there is a conservation of sequence around tryptophan residues that has been shown to be critical for biotin binding. It is possible that the hydrophobic binding of biotin to biotinidase and egg yolk biotin-binding proteins require tryptophan residues in the active center.

Egg Yolk Biotin-binding Proteins. The proteins of egg yolk differ from those of egg white in regard to time and place of synthesis. The entire egg white is secreted by the magnum of the oviduct within 4 to 6 h of ovulation. The yolk proteins, in contrast, are serum proteins synthesized by the liver, and contain high-affinity binding proteins for riboflavin, thiamine, cholecalciferol, retinol, vitamin B_{12}, and iron. Vitamins and minerals are transported protein-bound to the oocyte. The transport proteins such as the riboflavin and thiamine-binding proteins and transferrin present in both egg yolk and egg white are products of the respective single genes.

The exception seems to be the biotin-binding proteins. A biotin-binding protein (BBP-I) isolated from hen egg yolk was shown to be distinct from avidin. Avidin binds to biotin very tightly and does not exchange, whereas the yolk biotin-binding protein exchanges biotin at the body temperature of the bird. Thus, biotin could be made available to the developing embryo from its reserve in the yolk. Only about 15% of the biotin of the whole egg is associated with avidin, under normal conditions of biotin nutrition of the bird. The rest is associated with the yolk BBP. The plasma of laying hens and egg yolk contain another biotin-binding protein (BBP-II), which is stable up to 45 °C, whereas BBP-I was stable up to 65 °C. Both are tetrameric isoproteins, normally saturated with biotin, and together they account for most of the biotin of hen plasma and yolk. Yolk BBPs have a nutritional role, that of transporting biotin to the developing embryo. BBP-II is preferentially deposited in the yolk and is specifically involved in the transport of biotin to the oocyte. Both BBP-I and BBP-II cross-react with anti-BBP-II serum. The maximal production of BBP-I is attained at a lower level of dietary biotin ($\approx$50 μg kg^{-1}) than for BBP-II ($\approx$250 μg kg^{-1}). At this higher level of dietary biotin, the production of BBP-II is severalfold higher than that of BBP-I. The observation that the amount of these proteins is limited by dietary biotin within the normal dietary range would suggest that biotin is required for their synthesis, stability, or secretion. The synthesis of both BBP-I and BBP-II is stimulated during egg laying, indicating that the gene(s) for these two proteins are induced by estrogen. The difference in their response to dietary biotin content as well as in their distribution between hen and chick plasma would suggest that they are either products of different genes or differentially modified products of the same gene. Both estrogen and the specific ligand for BBP-I and BBP-II are involved in

the regulation of their synthesis, although the mechanism of action of biotin in this sequence is not known yet.

A biotin-binding protein analogous to BBP-II has been characterized in the serum of the pregnant or estrogenized female rat. This protein is not present in the sera of normal male rats. The molecular mass of this protein is 66 kDa and the isoelectric pH is 4.1. It exhibits immunological cross-reactivity with the purified egg yolk biotin-binding protein, BBP-II. The biological role of this protein is not known. The physiological relevance of the biotin-binding protein during rodent embryonic development was demonstrated by selective immunoneutralization of the maternal vitamin carrier, an intervention that led to a drastic decrease in the transport of biotin to the embryo, which in turn resulted in early embryonic mortality – without affecting maternal biotin nutrition.

6.2 Role of Biotinidase as a Biotin-carrier Protein

On the basis of investigations with HeLa cells and human fibroblasts, we suggested that biotin transport in mammalian cells might involve pinocytosis through the functioning of specific circulating proteins. Both biotinidase and biotin-binding activities coincided during serum fractionation, suggesting that biotinidase is the only protein that binds biotin in human serum and might be the biotin-carrier protein in human serum. We determined the binding of biotin to pure human serum biotinidase. The Scatchard plot as well as displacement curve analysis showed two linear parts, indicating that two classes of sites with high and low affinities for biotin (0.5 and 50 nM, respectively), were present. The high-capacity, low-affinity binding corresponds to that of other vitamin-binding proteins.

The decreased level of plasma biotin in biotinidase-deficient patients was tentatively attributed to decreased levels of biotinidase or a specific biotin-binding protein in plasma. Other indirect evidence for the existence of a biotin-binding protein comes from the study of epileptic patients on long-term therapy with anticonvulsants. Significantly low levels of serum biotin were observed in these patients, suggesting that these anticonvulsants probably compete with biotin for binding to a circulatory protein. These anticonvulsants share with biotin the common structural feature of a cyclic carbamide group. Protein binding involved this group of biotin and the anticonvulsants. We have shown that all these anticonvulsants compete with biotin for the binding to biotinidase, indicating that biotin in human serum is essentially noncovalently bound to protein(s) and that biotinidase is the only protein that exchanges with [^{3}H]*d*-biotin.

The uptake of biotin by various human cell lines was reported by us to be saturable. On the basis of the binding of avidin and avidin-biotin complex to rat liver plasma membrane, we suggested that avidin in these systems was mimicking a natural biotin-binding carrier involved in biotin transport. At concentrations of biotin less than 5 μM, absorption of biotin by rat jejunal segments proceeded largely by a saturable process. The biphasic transport of biotin and biocytin in the rat small intestine suggests that at biotin concentrations in the gut below 50 nM, the saturable uptake mechanism would operate. Such a system would have a tremendous advantage in the context of fluctuating amounts of biotin ingested in the diet. This is born out by the rarity

of primary biotin deficiency in humans. The late onset type of multiple carboxylase deficiency (MCD) was shown to be due to the deficiency of biotinidase, the patients lacking the system for absorbing biotin in the nanomolar range. Biotinidase is the only protein in brush border membrane of rat intestine that binds to biotin, indicating that biotinidase *in vivo* may have a role in the transport of biotin.

6.3 Nuclear Biotin-binding Protein

Intracellular fractionation of biotin in various tissues of the rat and the chicken indicated that a significant amount of biotin was associated with the nuclear fraction. The biotin content of biotin-deficient rat liver is about one-tenth that of normal rat liver, and a significant 20% of this is present in the nuclear fraction. Biotin is present in cell nuclei prepared from various tissues of cultured cells even though such nuclei do not possess any carbon dioxide–fixing capability, indicating that biotin in nuclei does not function as the prosthetic group of carboxylases. During biotin deficiency, biotin in the nuclear fraction seems to be conserved, whereas it is lost preferentially from other cellular organelles. Biotin in nuclei has been shown to be bound noncovalently to a protein. This protein binds reversibly to biotin *in vitro*, with a maximal binding of 3.54 pmol μg of protein and with a dissociation constant for biotin of 2.2×10^{-7} M. Polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulphate indicates an apparent subunit molecular mass of 69 kDa for this protein.

There have been other reports of the presence of high concentrations of biotin in optically clear nuclei (OCN). Solid cell nests resembling mulberries (morules) have been reported in colorectal adenomas and carcinomas, endometrial tissue and in pulmonary endodermal tumors resembling fetal lung. The nuclei in morules in several organs are referred to as OCN, and have been found to contain significant amounts of biotin. OCN with significant accumulations of biotin unrelated to the development of a morule appear in endometrium during pregnancy and in ovarian endometroid carcinoma.

6.4 Requirement of Biotin by Cells in Culture

A requirement for biotin by cells in culture would be expected in view of the obligatory involvement of biotin in the metabolism of carbohydrates, lipids, and deaminated residues of certain amino acids. However, various earlier studies indicated that cells in culture do not require biotin. Using biotin-depleted fetal bovine serum (FBS) and Eagle's minimum essential medium, we demonstrated a requirement for biotin by HeLa cells, human fibroblasts and rous sarcoma virus-transformed (RST) baby hamster kidney (BHK) cells. This was based on the viability, biotin content, and activities of biotin-dependent and -independent enzymes. We also demonstrated that there was a significant decrease in the incorporation of leucine into protein of the homogenate or cytosol of biotin-deficient HeLa cells compared to cells grown in a biotin-supplemented medium. When biotin was added to the biotin-deficient medium, there was a twofold increase in the incorporation of leucine into proteins. Experiments using puromycin and cordycepin indicated that the appearance of new RNAs in the cytoplasm occurs when biotin-deficient cells are supplemented with biotin. Mammalian cultured cells, when they

do not receive specific signals, which include specific growth factors, come to a halt in a quiescent, nongrowing variant of the G_1 state, referred to as G_0. The synthesis of components of the cell cycle control system is switched off.

Normal cells in G_1 arrest due to serine starvation start incorporating [^{3}H] thymidine into DNA as soon as serine is restored to the medium. Biotin-deficient HeLa cells under similar conditions do not incorporate [^{3}H] thymidine into DNA even when serine is restored to the medium. However, within 4 h of supplementation of biotin to the biotin-deficient medium, the incorporation of [^{3}H] thymidine into DNA reaches a maximum. By this time, there is stimulation of protein synthesis. The two phenomena are related, and the growth-promoting effect of biotin might be achieved through the synthesis of certain proteins.

A requirement for high-density lipoprotein (HDL) has been shown for the growth of Madin–Darby canine kidney cells grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with transferrin when the cells were exposed to a mixture (1 : 1) of DMEM and F-12 medium. The components of the F-12 medium responsible for support of growth in the absence of HDL include biotin, which is absent from DMEM, and choline, which is present in insufficient concentrations in DMEM. It is significant that the HDL fraction of plasma has a considerable amount of biotin associated with it. This biotin is nondialysable and is noncovalently attached to protein.

Various transformed cell lines produce transforming growth factors that enable them to grow in a serum-free medium supplemented with transferrin and insulin. The L1subline derived from the 3T3 mouse fibroblast cell line has the capacity to differentiate in the resting state into a cell type having the characteristics of adipocytes. When they reach confluence and start to differentiate, they greatly increase their rate of triglyceride synthesis. The increase in lipogenic rate parallels a coordinated increase in the activities of key enzymes of the fatty acid biosynthetic pathway. This increase correlates with a marked rise in nuclear run-off transcription rates for these mRNAs during differentiation. The process of differentiation can be accelerated in a number of ways, such as by increasing the amount of serum in the culture medium or by adding insulin or biotin. When serum in the cultures was dialyzed extensively, the morphological and enzyme changes characteristic of adipocyte differentiation were induced without fat deposition. It was presumed that dialysis removed biotin from serum, and the role of biotin was linked to the increase in acetyl-CoA carboxylase. However, it took 24 to 48 h after the addition of biotin to the medium before deposition of triglyceride. The delay suggests that some factor required for the induction of the whole set of lipogenic enzymes (which, with the exception of acetyl-CoA carboxylase, are not biotin enzymes) might be formed under the influence of biotin.

6.5 Biotin and Embryological Development

Congenital malformations have been reported in embryos of domestic fowl maintained on a biotin-deficient diet. Maternal biotin deficiency was strongly teratogenic in mice even when dams did not exhibit any typical signs of biotin deficiency. At mid-gestation, biotin-deficient embryos weighed less than normal embryos and had external malformations such as micognathia and micromelia.

There was a marked reduction in the size of the palatal process on day 15.5 of gestation, which may be due to altered proliferation of mesenchyme. The same may be said for the proliferation of mesenchyme in the limb bud.

We investigated the development of the palatal process in culture. After 72 h of organ culture, >90% of the explants from normal mouse embryos (biotin replete) were at stage 6 development. The corresponding figure for explants from biotin-deficient embryos cultured in a biotin-deficient medium was 6.5%. If the deficient explants were cultured in a medium containing biotin (10^{-8} M), the percentage at stage 6 of development rose to >30. Administration of biotin (20 mg) to biotin-deficient dams 24 h prior to removal of the embryos resulted in 33% of the explants at stage 6 development when cultured in a biotin-deficient medium, and >50% at stage 6 if cultured in a medium containing biotin (10^{-7} M).

In view of the role of biotin carboxylases in the major metabolic pathways, it was possible that the teratogenicity of biotin deficiency might be due to the accumulation of intermediary or secondary metabolites. However, there was no detrimental effect of any of the organic acid intermediates or secondary metabolites on palatal closure of the explants when these compounds were added to the organ culture medium at a concentration of 10^{-4} M. These results highlight the continuous requirement for biotin during the proliferation of the mesenchyme, perhaps for the synthesis of growth factors during organogenesis.

6.6 Biotin Requirement for Specific Functions

Evidence of delayed spermatogenesis and decreased numbers of spermatozoa due to biotin deficiency in the rat has been reported in early literature. In mammals, spermatogenesis is dependent primarily upon testosterone, which is produced by Leydig cells and acts on the Sertoli and peritubular cells of the seminiferous tubules to drive spermatogenesis. We have shown that testicular and serum levels of testosterone are decreased in the biotin-deficient rat. Biotin deficiency was accompanied by a significant degree of sloughing of the seminiferous tubule epithelium in these rats. Treatment of biotin-deficient rats with gonadotrophins or biotin increases serum levels of testosterone. However, even when testosterone levels are maintained at high levels in biotin-deficient rats by testosterone implants, the increase in serum testosterone does not result in normal spermatogenesis. The administration of biotin alone or biotin in addition to testosterone, to biotin-deficient rats, leads to normal spermatogenesis, thus suggesting that biotin might be involved in the formation of local testicular factor(s), which are required in addition to testosterone and follicle-stimulating hormone for the normal interaction between Leydig, Sertoli, and peritubular cells. The identity of these factors is not known, although we have shown that biotin is required for the synthesis of testicular proteins.

Brain-derived neurotrophic factor (BDNF) is a member of a family of cell-signaling molecules (neurotrophins), which have an important role in neuronal development and plasticity. BDNF has been implicated in developmental and adult plasticity within telencephalic brain regions that control learned vocal behavior in songbirds. BDNF expression seems to correlate with specific stages of songbird vocal learning. High levels of biotin have been reported in specific telencephalic nuclei (RA and HVC) among

juvenile males. It is possible that there may be a specific upregulation of biotinylated proteins within the HVA and RA nuclei in juvenile males. The developmental expression of this is correlated with vocal learning. High levels of biotin in the hippocampus, a brain region important for learning and memory, has also been reported, emphasizing the important role that biotin-regulated mechanisms might play in neuronal plasticity.

6.7 Effect of Biotin on the Regulation and Synthesis of Specific Proteins

Holocarboxylase synthetase (HCS) catalyzes the covalent attachment of biotin to the biotin-dependent carboxylases. MCD in humans is an inborn error of metabolism characterized by lack of carboxylase activities due to a lack of HCS activity. The obligatory participation of HCS in the biotin-dependent stimulation of the levels of HCS mRNA and those of acetyl-CoA carboxylase and the α subunit of propionyl CoA carboxylase in human fibroblasts has been reported. A regulatory role for biotin in the control of HCS and carboxylase mRNA levels via signaling cascade involving HCS, guanylate cyclase, and cGMP-dependent protein kinase has been proposed.

The asialoglycoprotein receptor is characteristic of fully differentiated hepatocytes. The human hepatoblastoma line HepG2 expresses maximum receptor activity only in confluent cultures. HepG2 cells grown to confluency in minimal essential medium made 10% with respect to FBS demonstrate an asialoglycoprotein receptor with ligand-binding characteristics and molecular mass comparable to those of the receptors purified from human liver. However, dialysis or ultrafiltration, with removal of the low molecular weight fraction of FBS, dramatically reduced expression of asialoglycoprotein receptor. The addition of biotin or biocytin to medium supplemented with dialyzed FBS during logarithmic growth supported normal expression of asialoglycoprotein receptor by confluent HepG2 cells. Biotin was most effective at a concentration of 10^{-8} M. Isolation of mRNA from HepG2 revealed no difference in asialoglycoprotein receptor transcripts when cells were grown in a dialyzed or undialyzed medium. Thus, a biotin-dependent posttranscriptional event seems to permit the ultimate expression of asialoglycoprotein receptor by HepG2 cells.

Glucokinase is specifically induced by biotin. Biotin is not a part of this protein. Earlier work from our laboratory demonstrated that liver glucokinase activity was altered in response to the biotin status of rats. Biotin also plays a role in the precocious development of glucokinase in young rats. In all these experiments, increase in enzyme activity was associated with protein synthesis. The synthesis of glucokinase is under developmental, nutritional, and hormonal control. The *in vivo* regulation of glucokinase mRNA of biotin-injected starved rats had increased 3.9-fold in comparison to normal-fed rats. Following this increase, there was a rapid decay of glucokinase mRNA, which was hardly detectable 4 h after biotin administration. The induction of glucokinase mRNA at 1 h following biotin administration to starved rats was 19-fold in comparison to starved rats not receiving the biotin injection. The induction of glucokinase mRNA by biotin is quite marked, relatively rapid, and correlated with the increase in glucokinase activity. In "run-on" transcription experiments with isolated liver nuclei,

biotin administration increased glucokinase gene transcription by about 6.7-fold. The increased transcription of glucokinase gene was not simply due to an increase in overall transcriptional efficiency, inasmuch as the transcription of the β-actin gene was unaffected by biotin. The difference between induction of mRNA and rate of transcription could also be due to the stabilization of glucokinase mRNA by biotin.

In both fasted and diabetic rats, hepatic phophoenol pyruvate carboxykinase (PEPCK) activities are markedly increased. Refeeding a high carbohydrate diet to fasted rats decreased PEPCK mRNA, which, in turn, is due to the repression of transcription of PEPCK gene by insulin. Since we had shown that the administration of biotin causes a rapid and marked induction of glucokinase mRNA in starved rats, we investigated the effect of biotin on the regulation of PEPCK in the liver and kidney of diabetic rats. After administration of biotin, hepatic PEPCK mRNA decreased to 15% of the non-biotin-injected control level at 3 h after injection. The level of hepatic mRNA had recovered significantly by 5 h after biotin injection. In parallel studies on insulin administration to diabetic rats, the inhibition was nearly fully reversed by 7 h after insulin injection. The relative rates of liver PEPCK gene transcription at various time intervals following biotin administration to diabetic rats was estimated in "run-on" transcription experiments. The transcription of the actin gene, included as internal control, was not influenced by biotin administration. Biotin suppressed the transcription rates of hepatic PEPCK gene by 55% at 30 min, and gradually, the rate of transcription increased back to the original level.

It is to be noted that the inhibition of transcription by biotin is dominant over other stimulatory effects. The elevated plasma glucagons characteristic of the fasting or diabetic condition induces PEPCK synthesis through enhanced transcription of the PEPCK gene. There are many similarities between biotin and insulin in their action on the enzymes of glucose metabolism. Both induce the mRNA that encodes glucokinase, a key glycolytic enzyme, and repress the mRNA that encodes PEPCK, a key gluconeogenic enzyme. This was the first demonstration that a water-soluble vitamin, biotin, exerted both positive and negative transcriptional control over key enzymes of glucose metabolism.

6.8 Biotinylation of Histones

Preteolytic hydrolysis of biotin proteins yields biotin peptides, the smallest of them being biocytin, ε-*N*-biotinyl-L-lysine. Biocytin and synthetic biotin peptides are hydrolyzed by biotinidase. Our studies indicated that this might not be its major function because the optimum pH for biotinidase hydrolysis of biocytin is 4.5 to 5, whereas the pH of serum is 7.4. The concentration of biocytin in serum is in the nanomolar range, whereas the affinity constant of biotinidase for biocytin is in the micromolar range. We proposed that biotinidase functions as a biotin-carrier protein. Later work indicated that biotinidase itself is biotinylated by biocytin. It was further shown that histone is a specific acceptor of biotin transferred from biotinylated biotinidase. This raised the possibility of histones being an endogeneous substrate of biotinyl transfer. Biotinylation of histones occurs *in vivo* in human cells, and there is increased biotinylation in response to cell proliferation. Histones *in vivo* are modified by

acytylation, methylation, phosphorylation, ubiquitination, and poly (ADP-ribsylation). Biotinylation of histones, by analogy with other modifications of histones, could lead to increased transcription of DNA.

7 Vitamin B_6

7.1 General

The term "vitamin B_6" refers to a group of naturally occurring pyridine derivatives that include pyridoxine, pyridoxal, pyridoxamine, and their phosphorylated derivates. Vitamin B_6 is essential for normal growth, development, and biological function. Pyridoxal phosphate, generally, is the active form of this vitamin and is a cofactor for many enzymes involved with the metabolism of amino acids and lipids. In addition to a role in general metabolism, its functions in the nervous and immune systems are well documented.

7.2 Pyridoxal Phosphate–binding Proteins

Pyridoxal phosphate in plasma is bound to albumin with high affinity and exists as a Schiff base. Virtually none of the pyridoxal phosphate in plasma is dialyzable. This binding protects against the renal glomerular filtration of pyridoxal phosphate. Pyridoxal is less tightly bound to albumin than pyridoxal phosphate, hence is susceptible to glomerular filtration. However, pyridoxal appearing in the glomerular filtrate is reabsorbed. The binding of pyridoxal phosphate to albumin also protects it against hydrolysis by alkaline phosphatase. Liver is the primary source of pyridoxal phosphate in plasma, and possesses a unique transport mechanism that possibly releases pyridoxal phosphate as an albumin complex. The complex contains pyridoxal phosphate and albumin in a molar ratio of 2 : 1. The binding of pyridoxal to hemoglobin results in the accumulation of the vitamer in the erythrocyte. This makes it possible for pyridoxal as well to be a transport form of vitamin B_6. The role of protein binding in the regulation of the liver cellular pyridoxal phosphate has also been indicated. When hepatic cytosol is dialyzed exhaustively, about 50% of the pyridoxal phosphate is still bound to proteins. This binding protects pyridoxal phosphate from hydrolysis by the liver plasma membrane–associated phosphatase activity. Thus, protein binding within cells as well as in plasma is a significant modifier of the hydrolysis of pyridoxal phosphate by phosphatase. It has been proposed that the distribution of pyridoxal phosphate between protein-bound and free forms within the cell, together with the enzymatic hydrolysis of pyridoxal phosphate, maintains the concentration of this vitamer form. In liver cytosol, pyridoxal phosphate is found as five major protein-bound forms.

Skeletal muscle is the major repository of vitamin B_6 in the body. About 60 to 90% of the total vitamin B_6 present in muscle is associated with glycogen phosphorylase, an enzyme that comprises about 5% of the total soluble protein of muscle. Pyridoxal phosphate is bound principally to glycogen phosphorylase. Phosphorylase requires pyridoxal phosphate for catalytic activity. Pyridoxal phosphate binds to Lys-679 in rabbit muscle glycogen phosphorylase with the 5′ phosphate adjacent to the substrate-binding site. The function of muscle phosphorylase as a storage depot for pyridoxal phosphate has been demonstrated by the effect of

vitamin B_6 deficiency on muscle phosphorylase activity.

7.3 Nuclear Pyridoxal Phosphate Binding

In rats fed on a diet adequate in vitamin B_6, the fraction of total pyridoxal phosphate found in the nuclei of liver cells was 21%, and this increased to 39% in rats fed on a vitamin B_6-deficient diet, indicating a conservation of the vitamin in the nuclear compartment. This is analogous to the distribution of biotin in biotin-replete and biotin-deficient animals. Pyridoxal phosphate is the only significant form of vitamin B_6 found in the nucleus, and there is one particularly prominent pyridoxal phosphate–binding protein in the nucleoplasmic fraction of nuclear preparations obtained from HTC cells. Results of analysis by sodium dodecyl sulfatelpolyacrylamide gel electrophoresis (SDS-PAGE) indicate that this protein has an apparent molecular mass of 50 to 55 kDa. The presence of pyridoxal phosphate in the nucleus could be of physiological significance. It is not known how the nucleus acquires pyridoxal phosphate or whether the phosphorylation of pyridoxal takes place in the nucleus, given the general assumption that phosphorylated vitamin B_6 vitamers do not cross membranes of mammalian cells.

7.4 Function of Pyridoxal Phosphate in the Nucleus

The function of the protein-bound pyridoxal phosphate in cell nucleus is of much interest. Various reports indicate that pyridoxal phosphate affects steroid hormone activity by altering the interaction of steroid receptor complexes with DNA, chromatin, and nuclei. Rat hepatoma cells when grown in presence of 5-mM pyridoxine have a significantly reduced glucocorticoid-dependent induction of tyrosine aminotransferase, and the enzyme activity is increased when the cells are grown in a medium deficient in pyridoxine. The addition of pyridoxal phosphate to the medium containing mouse mammary gland explants, however, results in a significant decrease of both dexamethasone binding to nuclear steroid receptor and dexamethasone-stimulated casein mRNA synthesis. The possibility that pyridoxal phosphate might act as a modulator of protein-DNA interaction was suggested.

The possibility that vitamin B_6 might regulate transcriptional activation by glucocorticoid receptors as well as by other members of the steroid hormone receptor family has been investigated. Elevated vitamin B_6 concentrations suppressed transcription mediated through glucocorticoid receptor. Conversely, such transcription was enhanced in mild vitamin B_6 deficiency. Under these conditions, the levels of hormone-independent constitutive gene expression were not affected. It was also shown that transcriptional activities mediated by androgen, progesterone, and estrogen were reduced by elevated vitamin B_6 levels and enhanced under conditions of vitamin deficiency. The glucocorticoid-induced expressions of cytosolic aspartate aminotransferase and cystathionase genes are modulated by pyridoxal phosphate. Gene expression is inhibited by pyridoxal phosphate, whereas it is enhanced in livers from pyridoxine-deficient rats.

The level of steroid-induced gene expression from simple promoters containing only hormone response elements and a TATA sequence was not affected by changes in the vitamin B_6 status. However, the modulatory influence of vitamin B_6 status was restored when a binding site

for a transcription factor, nuclear factor (NFl), was included within the hormone responsive promoter. These results suggest that vitamin B_6 modulates steroid hormone–mediated gene expression through its influence on a functional interaction between the steroid hormone receptors and the transcription factor NFI. Thus, the vitamin B_6 status of the cells introduces an additional level of cell-specific control over the regulation of gene expression by steroid hormones.

See also Medicinal Chemistry; Receptor Biochemistry; Regulation of Stem Cell; Vitamins, Structure and Function of.

Bibliography

Books and Reviews

Christakose, S., Dhawah, P., Liu, Y., Peng, X., Porta, A. (2003) New insights into the mechanism of vitamin D action, *J. Cell. Biochem.* **88**, 695–705.

Dakshinamurti, K. (Ed.) (1994) *Vitamin Receptors*, Cambridge University Press, Cambridge, MA.

Dakshinamurti, K. (2003) Regulation of Gene Expression by Biotin, Vitamin B6 and Vitamin C, in: Zempleni, J., Daniel, H. (Eds.) *Molecular Nutrition*, CABI International, Oxford, UK.

Holick, M.F. (2003) Vitamin D: a millennium perspective, *J. Cell. Biochem.* **88**, 296–307.

Noy, N. (2000) Retinoid-binding proteins: mediators of retinoid action, *Biochem. J.* **348**, 481–495.

Porter, T.D. (2003) Supernatant protein factor and tocopherol-associated protein: an unexpected link between cholesterol synthesis and vitamin E, *J. Nutr. Biochem.* **14**, 1–8.

Rodriguez-Melandez, R., Zempleni, J. (2003) Regulation of gene expression by biotin, *J. Nutr. Biochem.* **14**, 680–690.

Primary Literature

Alcain, F.J., Buron, M.I., Rodriguez-Aguilera, J.C., Villaba, J.M.Z., Navas, P. (1990) Ascorbate free radical stimulates the growth of a human promyelocyte leukemia cell line, *Cancer Res.* **50**, 5887–5891.

Allgood, V.E., Cidlowski, J.A. (1992) Vitamin B6 modulates transcriptional activation by multiple members of the steroid hormone receptor superfamily, *J. Biol. Chem.* **267**, 3819–3824.

Allgood, V.E., Powell-Oliver, F.E., Cidlowski, J.A. (1990) Vitamin B6 influences glucocorticoid receptor-dependent gene expression, *J. Biol. Chem.* **265**, 12424–12433.

Baez-Saldana, A., Diaz, G., Espinoza, B., Ortega, E. (1998) Biotin deficiency induces changes in sub-population of spleen lymphocytes in mice, *Am. J. Clin. Nutr.* **67**, 431–437.

Bhullar, R.P., Dakshinamurti, K. (1985) The effects of biotin on cellular function of hela cells, *J. Cell. Physiol.* **122**, 425–430.

Boeckx, R.L., Dakshinamurti, K. (1974) Biotin-mediated protein synthesis, *Biochem. J.* **140**, 549–556.

Borboni, P., Magnaterra, R., Rabini, R.A., Staffolanni, R., Porzio, O., Sesti, G., Fusco, A., Mazzanti, L., Lauro, R., Marlier, L.N.J.L. (1996) Effect of biotin on glucokinase activity, mRNA expression and insulin release in cultured B cells, *Acta Diabetol.* **33**, 154–158.

Bowie, A.G., O'Neill, L.A.J. (2000) Vitamin C inhibits NF-kBactivation by TNF via the activation of p38 mitogen-activated protein kinase, *J. Immunol.* **165**, 7180–7188.

Brigelius_Flohe, R., Flohe, I. (1996) Ascorbic acid, cell proliferation and cell differentiation in culture, *Subcell. Biochem.* **25**, 83–107.

Catani, M.V., Ross, A., Costanzo, A., Sabatini, S., Levrero, M. (2001) Induction of gene expression by activator protein-1 in the ascorbate protection against 4v-induced damage, *Biochem. J.* **356**, 77–85.

Chalifour, L.E., Dakshinamurti, K. (1982) The requirement of human fibroblasts in culture, *Biochem. Biophys. Res. commun.* **104**, 1047–1053.

Chalifour, L.E., Dakshinamurti, K. (1982) The characterization of the uptake of avidin-biotin complex by hela cells, *Biochim. Biophys. Acta* **721**, 64–69.

Chauhan, J., Dakshinamurti, K. (1991) Transcriptional regulation of the glucokinase gene, *J. Biol. Chem.* **266**, 10035–10038.

Collins, J.C., Paeitta, E., Green, R., Morell, A.G., Stockert, R.J. (1988) Biotin-dependent expression of the asialoglycoprotein receptor in HepG2, *J. Biol. Chem.* **263**, 11280–11283.

Dakshinamurti, K., Chalifour, L.E. (1981) the biotin requirement of hela cells, *J. Cell. Physiol.* **107**, 427–438.

Dakshinamurti, K., Chauhan, J. (1989) Biotin, *Vitam. Horm.* **45**, 337–384.

Dakshinamurti, K., Cheah-Tan, C. (1968) Liver glucokinase of the biotin-deficient rat, *Can. J. Biochem.* **46**, 75–80.

Dakshinamurti, K., Cheah-Tan, C. (1968) Biotin-mediated synthesis of hepatic glucokinase in the rat, *Arch. Biochem. Biophys.* **127**, 17–21.

Dakshinamurti, K., Li, W. (1994) Transcriptional regulation of liver phosphoenolpyruvate carboxykinase in the rat, *Mol. Cell. Biochem.* **132**, 127–132.

Dakshinamurti, K., Litwack, S. Biotin and protein synthesis in the rat liver, *J. Biol. Chem.* **245**, 5600–5605.

Dakshinamurti, K., Mistry, S.P. (1963) Tissue and intracellular distribution of biotin-C14OOH in rats and chickens, *J. Biol. chem.* **238**, 294–296.

Dakshinamurti, K., Mistry, S.P. (1963) Amino acid incorporation in biotin deficiency, *J. Biol. Chem.* **238**, 297–301.

Dakshinamurti, K., Chalifour, L., Bhullar, R.P. (1985) Requirement of biotin and the function of biotin in cells in culture, *Ann. N.Y. Acad. Sci.* **447**, 38–55.

Dakshinamurti, K., Modi, V.V., Mistry, S.P. (1968) Biotin deficiency produces hyperglycemia, *Proc. Soc. Exp. Biol. Med.* **127**, 396–400.

Dakshinamurti, K., Sharma, S.K., Geiger, J.D. (2003) Neuroprotective actions of pyridoxine, *Biochim. Biophys. Acta* **1647**, 225–230.

Dakshinamurti, K., Terrago-Litwak, L., Hong, H.C. (1970) Biotin and glucose metabolism, *Can. J. Biochem.* **48**, 493–500.

De La Vega, L.A., Stockert, R.J. (2000) Regulation of the insulin and asialoglycoprotein receptors via cGMP-dependent protein kinase, *Am. J. Physiol.* **279**, C2037–C2042.

Franceschi, R.T., Iyer, B.S., Cui, Y. (1994) Effects of ascorbic acid on collagen matrix formation and osteoblast differentiation in murine MC3T3 – F1 cells, *J. Bone Miner. Res.* **9**, 843–854.

Hung, C.F., Lu, K.H. (2001) Vitamin C inhibited DNA adduct formation and arylamine N-acetyltransferase activity and gene expression in rat glial tumor cells, *Neurochem. Res.* **26**, 1107–1112.

Ikeda, S., Takasu, M., Satsuda, T., Kakinuma, A. (1997) Ascorbic acid deficiency decreases the renal level of kidney fatty acid-binding protein by lowering the alpha 2u-globulin gene expression in liver in scurvy-prone ODS rats, *J. Nutr.* **127**, 2173–2178.

Ikeda, S., Horio, F., Kakinuma, A. (1998) Ascorbic acid deficiency changes hepatic gene expression of acute phase proteins in scurvy-prone ODS rats, *J. Nutr.* **128**, 832–838.

Johnson, R., Norstrom, E., Soderstrom, K. (2000) Increased expression of endogenous biotin, but not BNDF, telencephalic song regions during zebra finch vocal learning, *Dev. Brain. Res.* **120**, 113–123.

Komatsu, S.I., Watanabe, H., Oka, T., Tsuge, H., Nii, H., Kato, N. (2001) Vitamin B6-supplemented diets compared with a low vitamin b6 diet suppresses azomethane-induced colon tumorigenesis in mice by reducing cell proliferation, *J. Nutr.* **131**, 2204–2207.

Kurata, S., Senoo, H., Hata, R. (1993) Transcriptional activation of type I collagen genes by ascorbic acid-2-phosphates in human skin fibroblasts and its failure in cells from a patient with alpha 2(I)-chain-defective Ehles-Danlos syndrome, *Exp. Cell. Res.* **206**, 63–71.

Mori, T., Kitamura, R., Imaoka, S., Funae, Y., Kitada, M., Kamataki, T. (1992) Examination of lipid peroxidation in liver microsomes of guinea pigs as a causal factor in the decrease in the content of P450 due to ascorbic acid deficiency, *Res. Commun. Chem. Pathol. Pharmacol.* **75**, 209–219.

Oka, T., Komori, N., Kuwahata, M., Hiroi, Y., Shimoda, T., Okada, M., Natori, Y. (1995) Pyridoxal 5-phosphate modulates expression of cytosolic aspartate amino transferase gene by inactivation of glucocorticoid receptor, *J. Nutr. Sci. Vitaminol.* **41**, 363–375.

Oka, T., Komori, N., Kuwahata, M., Sasa, T., Suzuki, I., Okada, M., Natori, Y. (1993) Vitamin B6 deficiency causes activation of

RNA polymerase and general enhancement of gene expression in rat liver, *FEBS Lett.* **330**, 409–413.

Paulose, C.S., Thliveris, J., Viswanathan, M., Dakshinamurti, K. (1989) Testiculur function in biotin-deficient adult rats, *Horm. Metab. Res.* **21**, 661–665.

Reddy, V.G., Khanna, N., Singh, N. (2001) Vitamin C augmrnts chemotherapeutic response of cervical carcinoma hela cells by stabilizing p53, *Biochem. Biophys. Res. Commun.* **282**, 409–415.

Rodriguez-Melendez, R., Perez-Andrade, M.E., Diaz, A., Delorate, A., Camacho-Arroyo, I., Ciceron, I., Ibarra, I., Velazquez, A. (1999) Different effects of biotin deficiency and replenishment on rat liver pyruvate and propionyl-CoA carboxylases and on their mRNAs, *Mol. Genet. Metab.* **66**, 16–23.

Rodriguez-Melendez, R., Cano, S., Mendez, S.T., Velazquez, A. (2001) Biotin regulates the genetic expression of holocarboxylase synthetase and mitochondrial carboxylases in rats, *J. Nutr.* **131**, 1909–1913.

Sato, A., Nishioka, M., Awata, S., Nakayama, K., Okada, M., Horiuchi, S., Okabe, N., Sasa, T., Oka, T., Natori, Y. (1996) Vitamin B6 deficiency accelerates metabolic turnover of cystathionase in rat liver, *Arch. Biochem. Biophys.* **330**, 409–413.

Seitz, G., Gerbhardt, S., Beck, J.F., Bohm, W., Lode, H.N., Niethammer, D., Bruchelt, G. (1998) Ascorbic acid stimulates DOPA synthesis and tyrosine hydroxylase gene expression in the human neuroblastoma cell line SK-N-SH, *Neurosci. Lett.* **244**, 33–36.

Singh, I.N., Dakshinamurti, K. (1988) Stimulation of guanylate cyclase and RNA polymerase II activities in hela cells and fibroblasts by biotin, *Mol. Cell. Biochem.* **79**, 47–55.

Sone, H., Ito, M., Suguyama, K., Ohneda, M., Maebashi, M., Furukawa, Y. (1999) Biotin enhances glucose stimulated insulin secretion in the isolated perfused pancreas of the rat, *J. Nutr. Biochem.* **10**, 237–243.

Spence, J.T., Koudelka, A.P. (1984) Effects of biotin upon intracellular level of cGMP and the activity of glucokinase in cultured rat hepatocytes, *J. Biol. Chem.* **259**, 6393–6396.

Stanley, J.S., Griffin, J.B., Zempleni, J. (2001) Biotinylation of histones in human cells. Effects of cell proliferation, *Eur. J. Biochem.* **268**, 5424–5429.

Vesely, D.L. (1982) Biotin enhances guanylate kinase, *Science* **216**, 1329–1330.

Wang, H., Prevsner, J. (1999) Detection of endogenous biotin in various tissues: novel function in the hippocampus and implications for its use in avidin-biotin technology, *Cell Tissue Res.* **296**, 511–516.

Watanabe, T., Dakshinamurti, K., Persaud, T.V.N. (1995) Effect of biotin on palatal development of mouse embryos in organ culture, *J. Nutr.* **125**, 2114–2121.

Zempleni, J., Mock, D.M. (1999) Mitogen-induced proliferation increases biotin uptake into human peripheral blood mononuclear cells, *Am. J. Physiol.* **276**, C1079–C1084.

Zempleni, J., Mock, D.M. (2000) Utilization of biotin in proliferating human lymphocytes, *J. Nutr.* **130**, 335S–337S.

Zempleni, J., Mock, D.M. (2001) Biotin homeostasis during the cell cycle, *Nutr. Res. Rev.* **14**, 45–63.

Vitamins, Structure and Function of

Donald B. McCormick
Department of Biochemistry, Emory University, Atlanta, GA, USA

1 Classification 537
1.1 Discovery 537
1.2 Nomenclature 537

2 Individual Vitamins 537
2.1 Fat-soluble Cases 537
2.1.1 Vitamin A 537
2.1.2 Vitamin D 539
2.1.3 Vitamin E 541
2.1.4 Vitamin K 541
2.2 Water-soluble Cases 542
2.2.1 Thiamine 542
2.2.2 Riboflavin 545
2.2.3 Vitamin B_6 546
2.2.4 Niacin and Niacinamide 547
2.2.5 Folacin 548
2.2.6 Vitamin B_{12} 549
2.2.7 Biotin 550
2.2.8 Pantothenic Acid 550
2.2.9 Vitamin C 551

3 Perspectives 552
3.1 Quasi-vitamins 552
3.2 Pharmacologic Uses 552

Bibliography 553

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

Keywords

Apoenzyme
A protein moiety of an enzyme that requires a coenzyme.

Avitaminosis
A disease condition, described as a deficiency syndrome, resulting from the lack of a vitamin.

Coenzyme
An organic molecule, generally derived from a vitamin that functions catalytically in an enzyme system.

Cofactor
A natural reactant, usually either a metal ion or coenzyme, required in an enzyme-catalyzed reaction.

Holoenzyme
A catalytically active enzyme constituted by coenzyme bound to apoenzyme.

Hypervitaminosis
An unhealthy condition resulting from excess of a vitamin.

Hypovitaminosis
An unhealthy condition resulting from too little of a vitamin; interchangeable with avitaminosis.

A vitamin is an organic compound that occurs as a natural component of foods and must be supplied exogenously in small amounts to maintain growth, health, and reproduction of an organism. Hence, a vitamin is essential because it cannot be made in the organism that requires it; yet it functions, usually after metabolic alteration, in indispensable ways. Since only small quantities, usually microgram to milligram amounts per human adult per day, are required to avoid a deficiency disease, vitamins are classified as micronutrients along with the trace elements.

1 Classification

1.1 Discovery

Symptoms of diseases that were likely the result of insufficient dietary intake of one or more vitamins are described in early writings. For example, signs of beri-beri (lack of thiamine) are described in the ancient Chinese herbals (ca. 2600 B.C.), and the Eber papyrus (ca. 1150 B.C.) gives the signs of scurvy (lack of vitamin C). That diet could influence such diseases was also suspected in early writings from ancient Greek, Roman, and Arab physicians, who recognized that ingestion of liver would both prevent and cure night blindness, which we know today to be an early symptom of hypovitaminosis A. Such empirical relationships set the stage for the experimental phase of vitamin discovery that became scientifically based around the beginning of this century.

1.2 Nomenclature

The historical manner in which vitamins have been named was first based on capital Arabic letters (A, B, C, etc.). When it was realized that some vitamins (e.g. B) were mixtures more complex than initially perceived, subscript numbers were added to the individually isolated compounds (e.g. vitamins B_1, B_2). In some instances, subscripts were used as it became evident that members of a group are functionally similar but differ in one or more particulars of structure [e.g. vitamin A, which is required for vision, exists as A_1 (retinol) and A_2 (3-dehydroretinol)]. Common chemical names, which are receiving greater usage, give a better indication of the types of compound involved. In some cases, these have replaced the earlier alphanumeric designations. Thus vitamin H is now called *biotin*, and vitamin B_3 is known as *niacin*. The common names often reflect the presence of a specific atom (e.g. *thia*mine denotes sulfur). Chemically functional groups are also indicated (e.g. pyrido*xamine* with the amine function). Even larger portions of the molecular structure can be reflected (e.g. *ribo*flavin with a ribityl chain). Parts of some names reflect a biofunctional property as with chole*calciferol,* which is an alcohol that *carries* calcium, or reflect a source/location, as in *phyllo*quinone from plants.

2 Individual Vitamins

2.1 Fat-soluble Cases

The relative solubilities of vitamins continue to provide a generally useful division. Members of the *fat-soluble* group (i.e. A, D, E, K) are more soluble in organic solvents, which also relates generally to their fatlike absorption, transport, and storage. Hence, they are rather well retained in the body and in some cases become toxic at high dosage. Table 1 lists the fat-soluble vitamins required by humans and many animals. Aspects of chemistry, metabolism, and function of individual groups are considered in Sects. 2.1.1 through 2.1.4.

2.1.1 Vitamin A

The structure of retinol (A_1) as one form of vitamin A is shown in Fig. 1. 3-Dehydroretinol (A_2) has an additional double bond between carbons 3 and 4. These alcohol compounds are yellowish

Tab. 1 Fat-soluble vitamins.

Group/vitamers	*Adult RDA*[a]	*Dietary sources*	*Physiologic roles*	*Deficiency*	*Excess*
A Retinol Retinal	800–1000 μg as retinal equivalents	Provitamins are carotenoids with a β-ionone ring in green and yellow/orange vegetables; retinol in liver, whole milk products	Visual pigments, cellular differentiation and proliferation	Night blindness, xeroph-thalmia, keratomalacia	Headache, vomiting, diplopia, alopecia, desquamation, bone abnormalities
D Cholecalciferol Ergocalciferol	5–10 μg as cholecalciferol (= 400 IU of D)	Provitamins are 7-dehydrocholes-terol in skin and ergosterol from ergot; fortified milk and dairy products, eggs, margarine	Adsorption of calcium and mineralization of bones and teeth	Rickets (children) and osteomalacia (adults)	Nausea and vomiting, anorexia, diarrhea and loss of weight; calcification of soft tissue, especially kidney and heart
E Tocopherols Tocotrienols	8–10 mg as α-toco-pherol equivalents	Seeds/vegetable oils, margarine, and shortening; green leafy vegetables	Antioxidant for cell membrane unsaturated lipids and low-density lipoprotein	Anemia (premature infants), nerve and muscle degeneration	Relatively nontoxic, but may antagonize vitamin K in those on anticoagulant therapy
K Phylloquinones Menaquinones	55–80 μg	Green leafy vegetables; lesser amounts in milk and dairy products, meat, eggs, cereals, fruits	Formation of blood clotting proteins that bind calcium ions	Hemorrhagic disease	Relatively nontoxic, but synthetic menadione can lead to jaundice

[a]The lower range is for women, but needs are modestly higher in pregnancy or lactation.

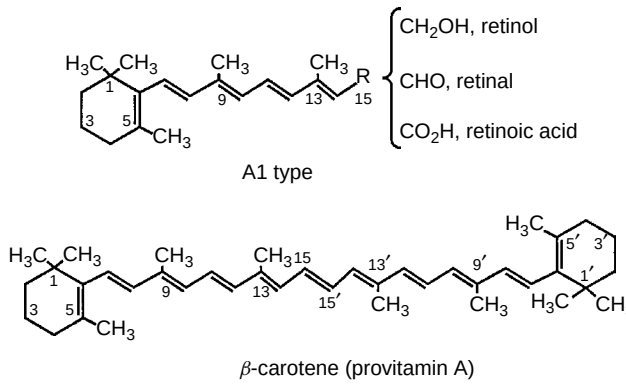

Fig. 1 Vitaminic forms of A_1, A_2, and carotene.

oils or low-melting solids that are practically insoluble in water but soluble in organic solvents and mineral oil. Vitamin A is sensitive to oxygen and ultraviolet light. Vitamin A_1 predominates, especially as long-chain fatty acid esters, in the liver of mammals and saltwater fish (e.g. cod), whereas the biologically less active vitamin A_2 is found in freshwater fish oils. Higher animals are unable to synthesize the β-ionone-type ring structure, but can derive the aldehyde retinal from provitamin carotenes and cryptoxanthin of plants. Retinal is then reversibly reduced to retinol. Retinoic acid is the irreversible oxidation product from retinal.

The structure for the most common and effective provitamin A, β-carotene, is also given in Fig. 1. The other carotenes yield less vitamin A_1 activity. Such carotenoid compounds, which constitute the yellow-to-orange pigments of most vegetables and fruits, vary considerably in their availability to the human; they often constitute the main dietary source of what ultimately becomes vitamin A.

The emulsification to micelles of forms of vitamin A and provitamin A by bile salts enhances their uptake by mucosal cells of the small intestine. Absorbed free retinol is reesterified with long-chain fatty acids within the cell. Retinyl esters in association with chylomicrons then pass via the lymphatic system to the liver and are stored in a lipoglycoprotein complex. Retinol associated with plasma retinol–building protein is released from hepatocytes to form, with circulating prealbumin, a molecular aggregate of sufficient size to avoid loss through glomerular filtration.

The best-understood physiological function of vitamin A is its participation as retinal in vision. Retinol and retinal are involved in a reversible cycle with opsin and rhodopsin, resulting in optic nerve transmission. Other broader functions of vitamin A occur in reproduction and growth. Systemic effects of an optimal level of vitamin A are the stabilization of cellular and intracellular membranes, maintenance of the integrity of epithelial tissue, and synthesis of glycoproteins. Retinoic acid also functions in the cell development and differentiation.

2.1.2 Vitamin D

The two forms of vitamin D, ergocalciferol (D_2) and cholecalciferol (D_3), are steroid-derived compounds. D_2 does not occur naturally but is produced by ultraviolet

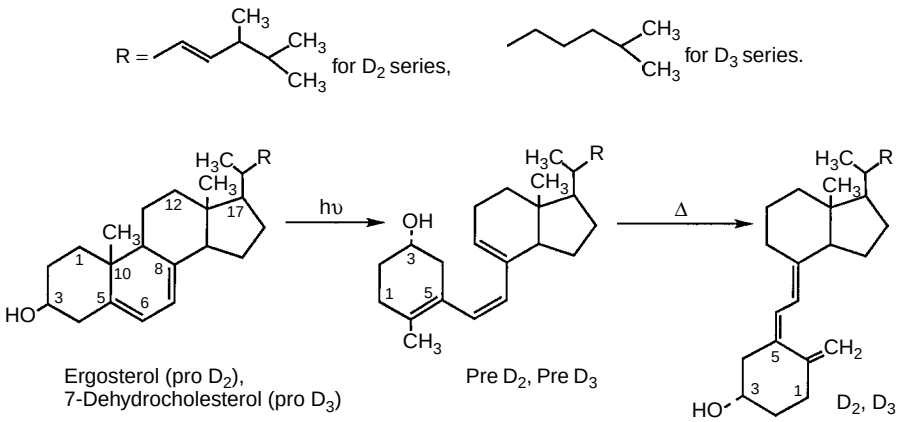

Fig. 2 Provitamins, previtamins, and vitamins D_2, and D_3.

irradiation of ergosterol, which occurs in molds, yeast, and higher plants. Vitamin D_3 in our bodies derives from irradiation of provitamin 7-dehydrocholesterol in skin. As shown by structures in Fig. 2, the only chemical difference between the two vitamin and provitamin forms is in the side chain. There are at least 10 compounds known to yield vitamin D–active compounds upon irradiation. Vitamin D forms white, odorless crystals that are soluble in fats and organic solvents. It is generally stable to heat, alkalies, acids, and oxidation.

The extent to which ultraviolet irradiation of 7-dehydrocholesterol in skin provides vitamin D_3 varies with seasonal changes in sunlight, amount of clothes worn, degree of skin pigmentation, and other factors. The availability and consumption of such vitamin D_3–containing foods as fatty fish, eggs, liver, and butter also varies. Hence, widespread fortification of food has been adopted to ensure an adequate supply. Because dairy foods in the United States supply about 75% of dietary calcium in a calcium/phosphorus ratio near-optimal for bone growth, milk, and infant formulas are desirable vehicles for vitamin D fortification. Though a poor source of vitamin D, milk that is now marketed contains 10 g (400 IU) of added vitamin D per quart.

Vitamin D is efficiently absorbed from the gastrointestinal tract unless there is fat malabsorption due to lack of bile salts, pancreatic insufficiency, or a defect of intestinal mucosal function. When absorbed, the vitamin is bound directly to chylomicrons. Much of the vitamin initially supplied to the liver is hydroxylated at the terminal side chain position to yield 25-hydroxy-D. This 25-hydroxycholecalciferol, which represents the major metabolite of D_3 in plasma, circulates by being bound to a vitamin D–binding α-globulin having a molecular weight (MW) of approximately 52 000. In kidneys, a second hydroxylation occurs at the 1α-position. This change is regulated by vitamin D status and by calcium/phosphate levels, which are controlled by parathormone (PTH) and calcitonin.

The 1α, 25-dihydroxycholecalciferol formed in the kidney complexes with the vitamin D–binding protein for transfer via blood to target tissues (e.g. small

intestine, bone). In the intestinal mucosal cell, the hormonally active dihydroxy-D mediates synthesis of a specific calcium-binding protein.

2.1.3 Vitamin E

The form of vitamin E that is biologically most active, and on which units are based, is D-α-tocopherol. There are eight related natural compounds in this vitaminic group, which is biosynthesized in plants and is especially abundant in vegetable oils. Structures are shown in Fig. 3. These compounds are soluble in fat solvents and insoluble in water, stable in acid and heat in the absence of oxygen, but labile to oxygen in alkaline solutions and to ultraviolet light.

Vitamin E is absorbed from the small intestine in the presence of bile. Most tocopherol enters the bloodstream via lymph, where the tocopherol is associated with chylomicrons and very low-density lipoproteins. The vitamin is stored in most tissues, with the largest amount stored in adipose tissue. Rapid exchange of tocopherol occurs between the erythrocyte membranes and plasma lipoproteins. When physiological amounts are administered, only a small fraction of the dose appears in urine.

The best-defined role of vitamin E is as an antioxidant for unsaturated fatty acyl moieties of lipids within membranes. Oxidative damage to polyunsaturated fatty acyl parts of phospholipids membrane can occur as a result of hydrogen peroxide production by flavoprotein oxidases. Concurrent free radical damage can ensue.

2.1.4 Vitamin K

Compounds in the vitamin K series, shown in Fig. 4, are the phylloquinones (K_1 type) synthesized in plants and the menaquinones (K_2 type) of bacterial origin. Several synthetic analogs and derivatives have been used; most relate to or derive from menadione (K_3), which lacks a side-chain substituent at position 3. The K vitamins are insoluble in water but dissolve in organic fat solvents. They are destroyed by alkaline solutions and reducing agents and are also sensitive to ultraviolet light.

The absorption of natural vitamin K from the small intestine is facilitated by bile. Vitamins K_1 and K_2 are bound to chylomicrons for transport from mucosal cells to the liver. Menadione (K_3) is more rapidly and completely absorbed. Vitamin K in the bloodstream is associated with circulating β-lipoproteins for transport. Whereas only traces of urinary metabolites of vitamins K_1 and K_2 appear in urine, a considerable

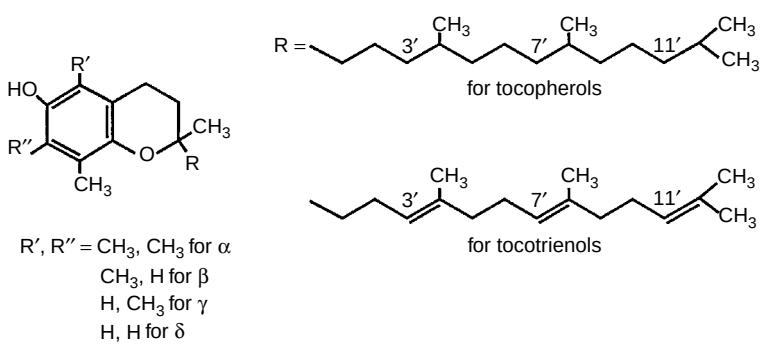

Fig. 3 Vitaminic forms of E.

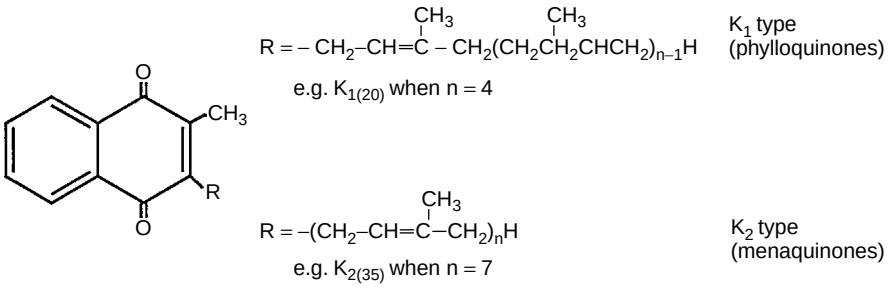

Fig. 4 Vitaminic forms of K.

portion of vitamin K_3 (menadione hydroquinone) is conjugated to β-glucuronide and sulfate esters, which are excreted.

Initially, vitamin K was recognized as a dietary antihemorrhagic factor. It has been found to be necessary for liver syntheses of plasma-clotting factors II (prothrombin), VII (proconvertin), IX (plasma thromboplastin component), and X (Stuart factor). It is now known that reduced K (hydroquinone) participates in the oxygen-dependent incorporation of CO_2 into the specific L-glutamyl residues of prothrombin and other plasma proteins. The formation of bone Gla protein (γ-carboxylglutamyl proteins) or osteocalcin, the possible regulators of mineralization, also depends on vitamin K.

Bis-4-hydroxycoumarin (dicumarol), the anticlotting compound from spoiled sweet clover, and synthetic 4-hydroxycoumarins such as warfarin, act as anticoagulants by interfering with the conversion of provitamin K to the functional hydroquinone. Hence, 4-hydroxy-coumarins are able to suppress the formation of prothrombin and other vitamin K–dependent Ca^2-binding proteins.

2.2 Water-soluble Cases

Vitamin C and the vitamins in the B-complex group are *water soluble* and as such share the fate of other solutes that are more compatible with an aqueous, physiological medium. Characteristics include a lesser tendency to be retained for long time in the body and a greater loss by way of urinary excretion. Table 2 lists the water-soluble vitamins required by the human and many plants and animals.

The eight vitamin groups that comprise the B-complex function after metabolic conversions to coenzymes, which are involved in holoenzymic systems, are also responsible for catalyzing numerous reactions. In the case of vitamin C, a cosubstrate-like function is seen with its oxidation coupled to certain hydroxylations. Aspects of chemistry, metabolism, and function of individual groups are considered in Sects. 2.2.1 through 2.2.9.

2.2.1 Thiamine

Thiamine (vitamin B_1) is a pyrimidyl-substituted thiazole (Fig. 5). The principal, if not the sole, coenzyme form is the pyrophosphate ester (TPP). Monophosphate and triphosphate esters occur naturally. Small amounts of thiamine and its phosphates are present in most plant and animal tissues, but more abundant sources are unrefined cereal grains, liver, heart, kidney, and lean cuts of pork. Thiamine is somewhat heat labile, particularly in alkaline solutions.

Tab. 2 Water-soluble vitamins.

Group/Vitamers	*Adult RDA*[a]	*Dietary sources*	*Physiologic roles*	*Deficiency*
B_1: Thiamine	1.1–1.2 mg	Pork, organ meat, whole grains, legumes	Coenzyme functions in metabolism of carbohydrates and branched-chain amino acids	Beri-beri, polyneuritis, Wernicke–Korsakoff syndrome
B_2: Riboflavin	1.1–1.3 mg	Milk and dairy products, meat, green vegetables	Coenzyme functions in numerous redox reactions	Cheilosis, angular stomatitis, dermatitis
Niacin Nicotinic acid Nicotinamide	14–16 mg as niacin equivalents	Liver, lean meat, grains, legumes; can be formed from tryptophan	Cosubstrates/coenzymes for H-transfer with numerous dehydrogenases	Pellagra with diarrhea, dermatitis, and dementia
B_6 Pyridoxine Pyridoxamine Pyridoxal	1.3 mg	Meat, vegetables, whole-grain cereals	Coenzyme functions in metabolism of amino acids, glycogen, and sphingoid bases	Nasolateral seborrhea, glossitis, peripheral neuropathy (epileptiform convulsions in infants)
Folacin Folic acid Folyl polyglutamates	400 µg	Liver, yeast, leafy vegetables, legumes, whole-wheat products	Coenzyme functions in single- carbon transfers in metabolism of nucleic and amino acids	Megaloblastic anemia, mental disorder, gastrointestinal disturbances, glossitis, pallor
B_{12}: Cyanocobalamin	2.4 µg	Meat, eggs, dairy products (not in plant foods)	Coenzyme functions in metabolism of odd-number fatty acids and methyl transfers	Pernicious anemia, neurological disorders, pallor
Biotin	60–100 µg (ranges recommended)	Liver, yeast, egg, yolk, soy flour, cereals	Coenzyme functions in bicarbonate-dependent carboxylations	Fatigue, depression, nausea, dermatitis, muscular pains

(*continued overleaf*)

Tab. 2 (*Continued*)

Group/vitamers	*Adult RDA*[a]	*Dietary sources*	*Physiologic roles*	*Deficiency*
Pantothenic acid	30 mg (adequate intake)	Animal tissues, whole-grain cereals, legumes; widely distributed	Constituent of CoA and phosphopantetheine involved in fatty acid metabolism	Fatigue, sleep disturbances, impaired coordination, nausea
C L-ascorbic acid Dehydroascorbic acid	5 mg (adequate intake)	Vegetables and fruits, especially broccoli and citrus	Formation of collagen and carnitine; important as antioxidant	Scurvy with petechiae, ecchy moses, peri-follicular hemorrhage, spongy and bleeding gums

[a]The lower range is for women, but needs are modestly higher in pregnancy or lactation.

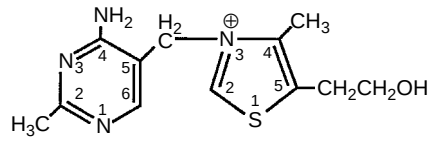

Fig. 5 Structure and ring numbering for thiamine.

Thiamine is readily absorbed in the small intestine by an active transport process when intake is less than 5 mg/day; at higher levels of intake, passive diffusion increasingly contributes to absorption. Phosphorylation takes place in the jejunal mucosa to yield thiamine pyrophosphate. Thiamine is carried by the portal blood to the liver. The free vitamin occurs in the plasma, but the coenzyme, TPP, predominates in the cellular components. About half the body stores are found in skeletal muscles, with much of the remainder in heart, liver, kidneys, and nervous tissue, including brain. Thiamine, as well as several of its catabolites, is excreted into the urine by the renal tubules.

As thiamine deficiency develops, there is a rather rapid loss of the vitamin from all tissues except the brain. The decrease of TPP in the erythrocytes roughly parallels the decrease of this coenzyme in other tissues. During this time, the thiamine in urine falls to near zero; the urinary metabolites remain high for some time before decreasing.

TPP functioning as the Mg(II)-coordinated coenzyme for the so-called active aldehyde transfers in the oxidative decarboxylation of α-keto acids, and the formation of α-ketols (ketoses) catalyzed by transketolase.

2.2.2 **Riboflavin**

Riboflavin (vitamin B_2) is a ribityl-substituted isoalloxazine (Fig. 6). This vitamin and its natural derivatives are yellow fluorescent compounds that are widely distributed throughout the plant and animal kingdoms. Principal among these flavins are the coenzymes, flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). Such coenzyme forms are rich sources of the vitamin in liver, kidney, and heart. Many vegetables are also good sources, but cereals are rather low in flavin content. Raw milk is a good source of the vitamin, but considerable loss can occur from exposure to light during pasteurization and bottling, or as a result of irradiation to increase the vitamin D content. Flavins are stable to heat but are decomposed by light.

The process by which riboflavin and lesser amounts of natural derivatives are released by digestion of complexes with food proteins involves hydrolysis of coenzyme forms of the vitamin (mainly

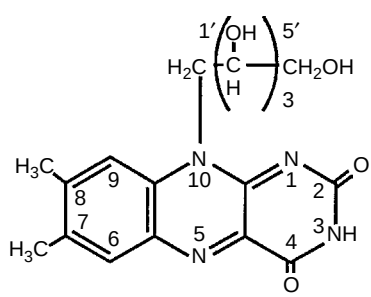

Fig. 6 Structure and numbering for riboflavin.

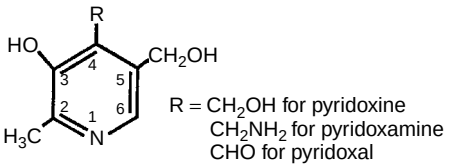

Fig. 7 Natural forms of vitamin B_6.

FAD; less FMN) as a consequence of gastric acidification. The vitamin is primarily absorbed in the proximal small intestine by a saturable transport system that is rapid and proportional to intake before leveling off at higher doses. Bile salts appear to facilitate the uptake. Much of the riboflavin circulating in blood plasma is loosely associated with albumin; lesser amounts complex with other proteins. One subfraction of immunoglobulin G (IgG) avidly binds a small portion of the total flavin in blood. Conversion of riboflavin to the coenzymes FMN and FAD occurs within the cell cytoplasm of most tissues, but particularly in the small intestine, liver, heart, and kidneys. Thyroxine and triiodothyronine stimulate FMN and FAD synthesis in mammalian systems. FAD is the predominant flavocoenzyme present in tissues, where it is mainly complexed with numerous flavoprotein dehydrogenases and oxidases. Since there is little storage of riboflavin as such, the urinary excretion reflects dietary intake.

In bound coenzymatic form, riboflavin participates in oxidation–reduction reactions in numerous metabolic pathways and in energy production via the respiratory chain. Flavins serve as redox carriers upon differential binding to proteins, participating, as well, in both one- and two-electron transfers. In reduced (1,5-dihydro) form, flavins react rapidly with oxygen.

2.2.3 Vitamin B_6

The three natural members of the vitamin B_6 group – pyridoxine (pyridoxol), pyridoxamine, and pyridoxal – are all 2-methyl-3-hydroxy-5-hydroxymethyl pyridines, as shown in Fig. 7. Both pyridoxamine-5′-phosphate and pyridoxal-5′-phosphate (PLP) interconvert as coenzyme forms during reactions catalyzed by aminotransferases (transaminases). Vitamin B_6 is widely distributed in animal and plant tissues, where the phosphorylated forms, and particularly PLP, predominate. Meats, poultry, and fish are good sources, as are yeast, certain seeds, and bran; somewhat more limited sources are milk, eggs, and green leafy vegetables. The common commercial form of the vitamin is pyridoxine hydrochloride, which is a water soluble, white, and crystalline solid. Solutions of the B_6 vitamins are decomposed by light, especially in the UV region, at neutral to alkaline pH. There is significant loss of PLP during the thermal processing of foods.

The three B_6 vitamers are readily absorbed by the mucosal cells. PLP binds for catalytic function with numerous specific apoenzymes throughout the cell. The erythrocytes, in addition, trap PLP as a conjugate Schiff base with hemoglobin. Glycogen phosphorylase contains most of the PLP in skeletal muscle.

Free vitamin is released when the phosphates are hydrolyzed by the nonspecific alkaline phosphatase located on the plasma membrane of cells or when pyridoxine-5′-β-D-Glucoside is hydrolyzed by the nonspecific glucosidase within the cells. The vitamin released from animal-derived foods is mainly pyridoxal; from plants, pyridoxine is preponderant. Some PLP is also released into the circulation by the

liver. Although PLP is the principal tissue form of vitamin B_6 and pyridoxal constitutes much of the circulating vitamin, the main catabolite excreted in urine is 4-pyridoxic acid.

As coenzyme PLP, vitamin B_6 functions in the metabolism of proteins, carbohydrates, lipids, and amino acids. PLP coenzyme-dependent reactions include formation of amines such as epinephrine and norepinephrine, formation of aminolevulinate in heme synthesis, and phosphorolysis of the α-1,4-linkages of glycogen.

2.2.4 Niacin and Niacinamide

Though the term *niacin* is chemically synonymous with the nicotinic (pyridine-3-carboxylic) acid, it is now used as the generic name for the specific compound as well as for the derivatives exhibiting qualitatively the biological activity of niacinamide (nicotinamide, nicotinic acid amide). Thus, *niacin activity* and *niacin deficiency* carry this broader meaning in nutritional literature. Figure 8 gives the structures of both vitamers.

The coenzymes nicotinamide adenine dinucleotide (NAD) (NAD; diphosphopyridine nucleotide (DPN)) and nicotinamide adenine dinucleotide phosphate (NADP) (NADP; triphosphopyridine nucleotide TPN) represent most of the niacin activity found in food sources that include yeast, lean meats, liver, and poultry. Milk, canned salmon, and several leafy green vegetables have lesser amounts. Protein provides a considerable portion of niacin equivalent because of the tryptophan content. As much as two-thirds of niacin required by adults can be derived from tryptophan metabolism via nicotinic acid ribonucleotide to NAD and NADP. Free forms of the vitamin are white, stable solids that are quite soluble in water. The oxidized coenzymes are labile to alkali, whereas the reduced (dihydro) coenzymes are labile to acid. Reduction of the oxidized coenzymes commonly occurs by the addition of a hydride ion to the para (4) position of the nicotinamide ring, with the simultaneous formation of a solvated proton. NADH and NADPH (but not NAD and NADP) absorb light in the near-ultraviolet region (339 nm).

The coenzymes are hydrolyzed in the intestinal tract, and both the acid and amide forms of the vitamin are readily absorbed. Nicotinic acid and nicotinamide are both present in the blood and move between blood and cerebrospinal fluid. Both compounds are converted to the coenzyme forms in blood cells, kidneys, brain, and liver. In the tissues, most of the vitamin is present as nicotinamide in NAD and NADP, although liver may contain a significant fraction of the free vitamin. There is little storage of niacin as such. Although nicotinamide can be converted to nicotinic acid by a rather widespread microsomal deamidase, there is no direct reamidation of nicotinic acid. Humans excrete 1-methylnicotinamide and 1-methyl-3-carboxamido-6-pyridone, which are primary urinary metabolites.

There are hundreds of enzymes that require the nicotinamide moiety within

Nicotinic acid (niacin) Nicotinamide (niacinamide)

Fig. 8 Structures for niacin and niacinamide.

either NAD or NADP. Most of these oxidoreductases function as dehydrogenases and catalyze such diverse reactions as the conversion of alcohols (often sugars and polyols) to aldehydes or ketones, hemiacetals to lactones, aldehydes to acids, and certain amino acids to keto acids. Generally, most NAD-dependent enzymes are involved in catabolic reactions, whereas NADP systems are more common to biosynthetic reactions.

2.2.5 **Folacin**

Folacin is the term covering both folic acid (pteroyl monoglutamate) and its natural polyglutamyl forms. The structure of the unconjugated acid, given in Fig. 9, comprises a pterin linked through a *p*-aminobenzoyl residue to the glutamate terminus. Reduction proceeds through the 7,8-dihydro to the 5,6,7,8-tetrahydro level before coenzymic activity is generated. Green leafy vegetables (e.g. spinach) are excellent sources of the vitamin.

Folacin is digested in the upper small intestine by cleavage of excess glutamyl residues by *conjugases*: Substances that hydrolyze the predominant coenzymic forms present in foods to liberate the free folate, which is then absorbed. Folate-binding proteins occur in plasma. In tissues, folate is reduced by the NADPH-dependent dihydrofolate reductase to tetrahydrofolate, which is then polyglutamylated by an ATP-dependent synthetase.

Coenzymic cycling is generally at the tetrahydro level for one-carbon-transfer reactions, though an oxidation–reduction shift from the tetrahydro to dihydro occurs with thymidylate synthetase. In this case, the transfer of a methylene function from 5,10-methylene tetrahydrofolyl coenzyme to deoxyuridylate involves its reduction to a methyl on the thymidylate product concomitant to oxidation of the pterin portion to generate the dihydrofolyl coenzyme that must be recycled. Other important roles of tetrahydrofolyl (THF) coenzymes relating to biosynthesis of nucleic acid precursors include the transformylase-catalyzed conversions of glycinamide ribonucleotide to the *N*-formyl compound with the participation of 5,10-methenyl-THF, and 5-aminoimidazole-4-carboxamide ribonucleotide to the 5-formamido compound with the participation of 10-formyl-THF.

With respect to amino acid metabolism, enzyme systems include the hydroxymethyltransferase-catalyzed conversion of serine with THF to glycine and

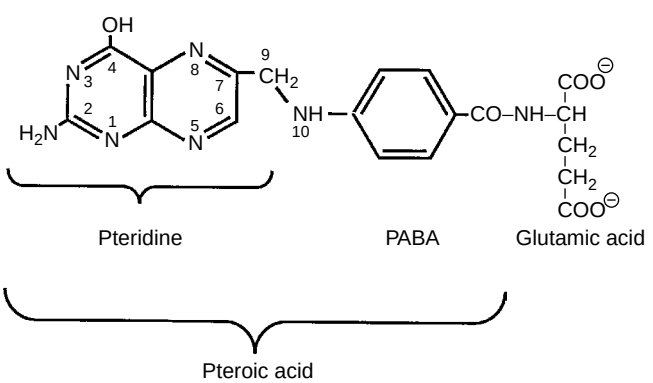

Fig. 9 Structure with numbering for folic acid.

5,10-methylene-THF, the formiminotransferase reaction with formiminoglutamate and THF to yield glutamate plus 5-formimino-THF, and the biosynthesis of methionine from homocysteine plus 5-methyl-THF to regenerate THF.

2.2.6 Vitamin B_{12}

The commercial form of vitamin B_{12} is cyanocobalamin, shown in Fig. 10, but the tissue-predominate B_{12} has water or the hydroxyl ion in place of the coordinate cyanide. Primary structural features include a central cobinamide with a corrin ring system and an affixed dimethylbenzimidazole nucleotide with coordination of this nitrogen base to the corrin cobalt. The cobalamins are photolabile and also unstable at extremes of pH. Cyanocobalamin exhibits a distinctive absorption spectrum with a long wavelength maximum at 550 nm for the reddish aqueous solutions. Best food sources are meats and dairy products; higher plants, which cannot synthesize this vitamin, provide little or none. Hence, true vegetarians are at risk for deficiency.

Upon ingestion, vitamin B_{12} (extrinsic factor) forms a complex with a glycoprotein called *intrinsic factor* that is absorbed in the distal ileum. Additional B_{12}-binding proteins (transcobalamins, R proteins) help vector the vitamin through plasma and into cells. For conversion to functional coenzymes, the aquocobalamin (B_{12a}) is reduced through the radical (B_{12r}) to the Co(I) reactive B_{12s} that can react with ATP to add the 5′-deoxyadenosyl moiety of coenzyme B_{12} in a transferase-catalyzed step or with methyl-THF to form

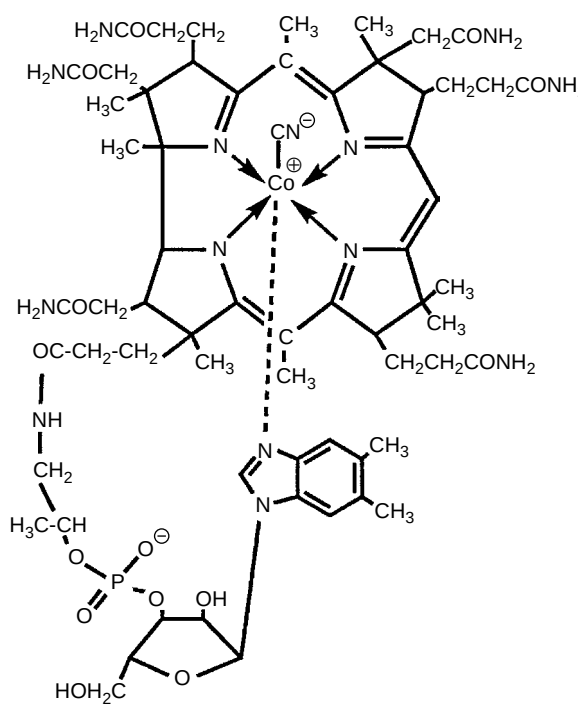

Fig. 10 Structure of cyanocobalamin (B_{12}).

methyl B_{12} that participates in methionine biosynthesis.

Besides the essential function of B_{12} in formation of methionine in mammals including the human, the coenzyme form is required for the methylmalonyl–coenzyme A mutase that routes L-methylmalonyl–CoA to succinyl–CoA and the Krebs cycle. More broadly in other organisms, B_{12} coenzymes function in ribonucleotide reductase, diol dehydratase, ethanolamine ammonia lyase, L-β-lysine mutase, glutamate mutase, and so on. In some cases (e.g. the adenylcobamide coenzyme used in a bacterial glutamate mutase), the N base that is a dimethyl-benzimidazole in the Co-B_{12} in our bodies is replaced by a different N base.

2.2.7 Biotin

Biotin (vitamin H) is a fused biocyclic system. In most organisms, this vitamin, shown in Fig. 11, occurs mainly bound to protein. Good sources of biotin include liver, kidney, pancreas, eggs, yeast, and milk.

Digestion of dietary proteins containing bound biotin yields considerable biocytin (ε-N-biotinyl lysine). Biocytin and biotin are readily absorbed. Biotinidase releases biotin from biocytin. Biotin is cleared from the circulating blood and is taken up by such tissues as liver, muscle, and kidney. Careful balance studies show that in humans, the urinary excretion of biotin often exceeds dietary intake and fecal excretion is always more than dietary intake because of microfloral biosynthesis. For the average adult, total body content of biotin is perhaps only 1 mg.

Four biotin-dependent enzymes have been found in human tissues. They are carboxylases for acetyl–CoA, propionyl–CoA, methylcrotonyl–CoA, and pyruvate.

2.2.8 Pantothenic Acid

Pantothenic acid is the amide formed from pantoic acid and alanine, as shown in Fig. 12. This vitamin is ubiquitous in nature and is synthesized by most microorganisms and plants. The vitamin is an integral part of a covalently attached prosthetic group of acyl carrier proteins and is within the structure of coenzyme A. The vitamin is widely distributed in foods and is particularly abundant in animal sources, legumes, and whole-grain cereals. Excellent food sources include egg yolk, kidney, liver, and yeast. Fair sources include broccoli, lean beef, skimmed milk, sweet potatoes, and molasses. More than half the pantothenate in wheat may be lost during the manufacture of flour, and up to a third is lost during the cooking of meat. The most common commercial synthetic form is the calcium salt.

Coenzyme A, the form in which much of the pantothenic acid is ingested, is hydrolyzed by intestinal pyrophosphatase

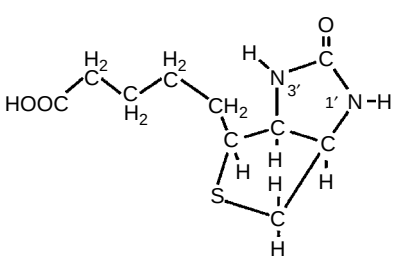

Fig. 11 Structure of biotin.

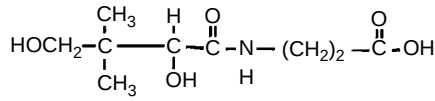

Fig. 12 Structure of pantothenic acid.

and phosphatase to pantetheine, which together with pantothenate is absorbed into the portal circulation. Within cells, resynthesis of coenzyme compounds occurs. About 80% of the vitamin in animal tissues is in CoA form. Cleavage enzymes operate during turnover and release of the vitamin that is excreted in the urine. Only a small fraction of pantothenate is secreted in milk, and even less into colostrum.

The myriad acyl thiol esters of CoA, of which pantothenic acid is a constituent, are central to the metabolism of numerous compounds (especially lipids) and the ultimate catabolic disposition of carbohydrates and ketogenic amino acids. For example, acetyl–CoA, which derives from the metabolism of carbohydrates, fats, and amino acids, can acetylate compounds such as choline and hexosamines to produce essential biochemicals; it can also condense with other metabolites such as oxaloacetate or with itself to supply citrate and cholesterol, respectively. Another essential role of pantothenic acid is its participation in the 4′-phosphopantetheine moiety of acyl carrier protein (ACP), where it facilitates chain elongation during fatty acid biosynthesis. Although the reactive thiol function of CoA and ACP is not an integral part of pantothenate, the steric and chemical properties conferred by the vitamin structure are important for enzymic recognition.

2.2.9 Vitamin C

Vitamin C (L-ascorbic acid) is a white, crystalline solid that is readily soluble in water. Ascorbic acid is a relatively strong reductant with an E_0 (pH 7) of 0.58 V. Ascorbic acid is reversibly oxidized to dehydroascorbic acid (ascorbone); both compounds are shown in Fig. 13. The dehydro form is more labile than the reduced form. Some mammals, including the human, lack the enzyme L-gulonolactonase that catalyzes the formation of 2-keto-L-gulonolactone, which spontaneously tautomerizes to L-ascorbic acid. The best sources of the vitamin are citrus fruits, berries, melons, tomatoes, green peppers, raw cabbage, and leafy green vegetables. Losses during processing, especially with heat and aerobic conditions, can be considerable.

Absorption of vitamin C occurs readily, mostly from the stomach, where some of the ascorbic acid is converted to the dehydro form. At physiological pH, the uncharged dehydroascorbic acid passes across cell membranes faster than the

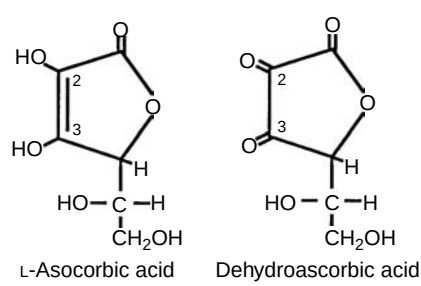

Fig. 13 Structure for L-ascorbic and dehydroascorbic acids.

anionic L-ascorbate. Passive diffusion of vitamin C accounts significantly for entry into some cells, such as leukocytes and erythrocytes, but an active transport mechanism also operates, especially for platelets, liver, adrenals, and retina. Vitamin C is found in most tissues, but glandular tissues such as the pituitary, adrenal cortex, corpus luteum, and thymus have the highest amounts, and the retina has 20 to 30 times the plasma concentration. The half-life for vitamin C in humans is only about 16 days. Ascorbate, dehydroascorbic acid, and lesser amounts of a number of catabolites are present in urine.

The most clearly established and critical functional role for ascorbic acid is a cofactor for protocollagen hydroxylase, the enzyme responsible for hydroxylation of prolyl and lysyl residues within nascent peptides in the connective tissue proteins. Among these are collagen and related proteins that comprise the intercellular material of cartilage, dentin, and bone. Vitamin C may also be involved in carnitine biosynthesis, tyrosine metabolism, microsomal drug metabolism, synthesis of epinephrine and antiinflammatory steroids by the adrenals, folic acid metabolism, and leukocyte functions. Absorption of Fe(II) is enhanced by simultaneous ingestion of the vitamin.

3 Perspectives

3.1 Quasi-vitamins

Some compounds not listed in Tables 1 and 2 are legitimate vitamins for certain species, though not the human. For example, lipoic and *P*-aminobenzoic acids for certain microorganisms or carnitine for the common mealworm represent essential organic micronutrients that cannot be synthesized adequately by the organisms that depend on dietary supply of these compounds. More than micro amounts of choline and myoinositol are required at early growth stages for rodents. There are, furthermore, several compounds that at one time or another have been falsely claimed to be vitamins with reputed benefit to humans. These include such ineffective factors as B_{15} or pangamic acid and B_{17} or laetrile. Only the 13 vitaminic groups listed in Tables 1 and 2 presently account for human needs.

3.2 Pharmacologic Uses

As mentioned in connection with the fat-soluble vitamins, particularly A and D, toxicity can result from the ingestion of too much, especially chronically. Excessive amounts of at least a couple of water-soluble vitamins can also pose a health hazard. Hepatoxicity can result from high amounts of nicotinic acid, and neurologic dysfunction has been associated with chronic excess of pyridoxine.

There is a growing certainty, however, that amounts of L-ascorbate (C) and α-tocopherol (E) that are modestly above the present recommended daily allowance (RDA) levels, have beneficial effects as natural antioxidants. The decrease in some specific cancers has been associated epidemiologically with higher intakes of foods rich in C, E, and carotenoids, including the provitamin A β-carotene. Extra folic acid also seems to protect the fetuses in a subset of women who have a history of neural tube defects in their newborns. More specific and direct experimentation will be needed to clarify such pharmacological roles of vitamins.

See also Medicinal Chemistry; Micronutrients, Trace Elements; Retinoids and the Chemical Biology of the Visual Cycle; Vitamin Receptors.

Bibliography

Bender, D.A. (1992) *Nutritional Biochemistry of the Vitamins*, Cambridge University Press, Cambridge.

Bowman, B.A., Russell, R.M., Eds. (2001) *Present Knowledge in Nutrition*, 8th ed. ILSI Press, Washington, D.C.

Combs, G.E., Jr. (1992) *The Vitamins, Functional Aspects in Nutrition and Health*, 2nd ed. Academic Press, San Diego, CA.

DRI (Dietary Reference Intakes) for Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline (1998) Institute of Medicine (IOM), National Academy Press, Washington, D.C.

McCormick, D.B., Greene, H.L. (1999) Vitamins. In *Tietz Textbook of Clinical Chemistry*, 3rd ed., C. Burtis and E. Ashwood, Eds. Saunders, Philadelphia.

Rucker, R.B., Suttie, J.W., McCormick, D.B., Machlin, L.J., Eds. (2001) *Handbook of Vitamins*, 3rd ed. Marcel Dekker, New York.

Sauberlich, H.E., Machlin, L.J. (1992) Beyond deficiency. New views on the function and health effects of vitamins. *Ann. N.Y. Acad. Sci.* 669.

Shils, M.E., Olson, J.A., Shike, M., Ross, A.C., Eds. (1999) *Modern Nutrition in Health and Disease*, 9th ed. Williams & Wilkins, Baltimore.

Whole Genome Human Chromosome Physical Mapping

Cassandra L. Smith, Giang H. Nguyen, Denan Wang, and Nickolaev Bukanov
Boston University, MA, USA

1 **Introduction 559**

2 **Genomic Mapping Terms and Concepts 560**

3 **Evolving Approaches for Physical Mapping 561**

4 **Genomic Restriction Maps 561**

5 **Genomic Clone Libraries 565**

6 **Cloneless Genomic Libraries 566**

7 **Gap Closure 568**

8 **Prospectus 568**

Bibliography 569
Books and Reviews 569
Primary Literature 569

Keywords

***Alu* Repeat**
A 300-bp DNA interspersed repeat sequence that occurs about once every 5000 bp. The most common repeat element in the human genome, it is CG rich and appears to occur preferentially in light Giemsa bands on condensed chromosomes.

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

Bacterial Artificial Chromosome (BAC)
An artificial chromosome cloning system in *E. coli* based on elements of the native circular chromosome.

Cloneless Library
A library of genomic fragments consisting of gel slices of electrophoretically fractionated DNA.

Comparative Genomic Hybridization (CGH)
A molecular cytogenetic method capable of detecting and locating relative genomic sequence copy number differences between pairs of DNA samples. Equal concentrations of differentially labeled DNAs are hybridized simultaneously to metaphase chromosome spreads. Regions with deleted or amplified DNA sequences are seen as changes in the ratio of the intensities of these two labels along the target chromosome.

Expressed Sequence Tag (EST)
A segment of a sequence from cDNA clone that corresponds to an mRNA. Mapping ESTs helps make maps and points directly to expressed genes.

Fluorescence *In Situ* Hybridization (FISH)
Hybridization of nucleic acid probes to chromosomes immobilized on microscope slides or filters.

Inter-*Alu* PCR
A polymerase chain reaction method that uses primers contained in *Alu* repeat element to amplify single-copy sequences between adjacent *Alu* elements.

Jumping Libraries
A library of clones containing DNA sequences from the ends of the same restriction fragment.

***Kpn* Repeat**
The second most commonly occurring repeat sequence in the human genome, occurring on average once every 50 000 bp. It is a relatively AT-rich, LINE repeat and occurs preferentially in dark Giemsa bands on condensed chromosomes.

Linking Libraries
Libraries of clones containing DNA sequences that span a selected restriction enzyme recognition site.

Locus
A region on a chromosome linked to a functional unit such as a gene, telomere, centromere, and replication origin.

Long Interspersed Repeat Elements (LINEs)
Long interspersed repeating sequences that appear to be similar to retroposons.

Multiplex Analysis
An analytical approach to increase throughput by collecting multiple data simultaneously.

P1-derived Artificial Chromosomes (PAC)
An *E. coli* bacteriophage cloning system used for cloning of fragments up to 150 kb in size.

Polonies
Isolation of DNA sequences on beads rather than in clones.

Polymerase Chain Reaction (PCR)
A method for amplifying DNA by alternatively denaturing double-stranded DNA, annealing pairs of primers located near each other on complementary strands, and synthesizing the DNA between the primers using DNA polymerase.

Polymorphism Link-up
A mapping approach that establishes continuity between restriction fragments by taking advantage of the naturally occurring polymorphism in different DNAs. In some cell lines, hybridized probes appear to identify different fragments, whereas in others they appear to identify the same fragment. The pattern of occurrence of these fragments can be used to assess whether the probes identify the same or adjacent fragments.

Pulsed Field Gel (PFG) Electrophoresis
A method of electrophoresis that exposes nucleic acids to alternating electrical fields. Fractionation is based on the speed at which the molecules can change directions.

Radiation Hybrid Maps
Radiation hybrid maps are created by analyzing randomly broken DNA cloned into hybrid cell lines. The breakage is done using X rays and the centiRay distance is a function of how often two loci remain together in a library of radiation hybrid cell lines.

RARE (RecA-assisted Restriction Endonuclease) Cleavage
A site-specific DNA cleavage method based on the ability of RecA protein from *E. coli* to pair an oligonucleotide to its homologous sequence in duplex DNA. This three-stranded DNA complex is protected from methylase. After methylation and removal of RecA, restriction endonuclease cleavage is limited to the site previously protected from methylation. If pairs of oligonucleotides are used, a specific fragment can be cleaved out of genomes.

Retroposons
A group of DNA sequence elements that appear to transpose through an RNA intermediate. Retroposon elements do not code for reverse transcriptase, do not have terminally redundant sequences, and do have a 39-poly$(A)_n$ stretch. A variable-sized target duplication occurs at the site of integration.

Sequence-tagged Restriction Site (STAR)
A short DNA sequence used to identify the DNA surrounding a restriction nuclease cleavage site.

Sequence-tagged Site (STS)
A short DNA sequence used to identify a DNA segment.

Short Interspersed Repeat Elements (SINEs)
Repeated sequences less than 500-bp long and present in a high copy number (10^5 per human genome). The *Alu* repeat element consists of SINEs.

Yeast Artificial Chromosome (YAC)
An artificial yeast chromosome constructed by cloning genomic fragments into vectors that can replicate in yeast. YACs have the following characteristics: a yeast centromere, two telomeric sequences, and a selectable marker.

This entry reviews the types of physical map that can be constructed, the methods used to construct them and their most likely current use. In particular, the emphasis is on the construction of genomic restriction maps and ordering overlapping libraries using top-down mapping approaches to enable *de novo* genome sequencing and for gap filling in sequencing projects. Also explored is the use of these approaches in functional studies.

1 Introduction

In the past, most genome studies were limited to organisms having well-developed genetic systems. Thus, a few model systems were studied intensely (e.g. bacteriophage lambda, *Eschericia coli*, Drosophila, mice) in a large number of different ways. The accumulation of diverse knowledge on these model organisms further promoted research on these model systems. Now, a number of molecular approaches have been developed that allow the construction of physical maps for virtually any chromosome. This means that analysis of an uncharacterized organism can begin with physical dissection including sequencing. Knowledge gained from this initial foray provides the foundation for functional studies. Hence, a new synergism in biology is provided by bottom-up genome studies.

Bacterial genomes are small and range from ~1 to 15 Mb in size, reflecting in part, the relative ability of these organisms to be free living. The entire genomic sequence of an increasing number of bacterial genomes is becoming available. These genomes are small enough so that sequencing of random clones is used to obtain the entire genome sequence. Hence, physical genomic maps for these organisms are used to study global gene expression or other global genomic activities.

In the past, the study of some lower eukaryotic genomes such as protozoa was hampered by the lack of well-developed genetic systems. Also, these chromosomes did not condense during cell division and could not be visualized microscopically. The application of pulsed field gel electrophoresis (PFG) to examine the genetic make up of these organisms was quite useful. PFG separates DNA chromosomal DNAs or large fragments that range up to ~10 Mb in size. Protozoan genomes range in size from the length that overlaps large bacterial genomes, up to about 100 Mb. Surprisingly, one parasitic protozoan genome, the *Giardia lamblia* genome, was found to be only about 12 Mb in size (i.e. smaller than the *Saccharomyces* genome).

Until recently, large, complex genomes, like the human genome (~3000 Mb in size) have been particularly recalcitrant to molecular dissection. Human and other higher eukaryotic chromosomes are 50 to 300 Mb in size, condense and are visible microscopically. Differential staining of the condensed chromosomes allows a finer division into regions ("bands") estimated to be 5 to 30 Mb in size that appear to reflect regional differences in GC content. The banding patterns serve as anchors to a large amount of genomic data. Also, the division of the genome into chromosomes and chromosomal bands provides convenient pieces for top-down mapping approaches.

In top-down mapping approaches, the genome is divided into units to facilitate study. For instance, chromosomes naturally divide the genome, and chromosome-banding patterns represent another division. Further division will depend on the method of analysis. Conventional recombinant DNA approaches to genome analysis allowed the characterization of molecules up to only about 0.05 Mb. The "resolution gap," 0.05 to 10 Mb was exactly the size range that was most amenable to study by PFG techniques and the entire size range of DNA molecules could be analyzed. PFG also promoted the extension of recombinant DNA methods to large cloning system such as yeast artificial chromosomes (YACs), P1 artificial

chromosomes (PACs), and bacterial artificial chromosomes (BACs).

2
Genomic Mapping Terms and Concepts

Many molecular techniques with varying degrees of resolution can be used to characterize chromosomes. Hence, there is no unifying concept of what constitutes a chromosomal physical map. Further, there are ambiguities in the literature with respect to the distinction of physical versus genetic maps.

Any map will consist of markers or objects. The order, and perhaps the distance, between pairs of objects will be known (some might argue that a map must consist of both order and distance). Object ordering along a chromosome should be maintained irrespective of the method used to construct the map, whereas map distances are method dependent. For instance, the amount of recombination along a chromosome is not constant. Hence, it is not surprising that a comparison of the physical and genetic distances along the long arm of chromosome 21 revealed at least a sixfold variation in the distances. It is quite clear that the ultimate map is the entire sequence of a chromosomal DNA. All maps and objects will be anchored to the DNA sequence once it is available. Some imprecision will always exist because map objects may be imprecisely defined, or have imprecisely defined locations by their very nature.

Classically, a genetic map was composed of chromosomal loci and the amount of recombination between the loci was the genetic distance. For eukaryotic organisms, genetic distance, expressed in centimorgans (cM), is a measure of the coinheritance of genetic markers. This type of analysis requires the examination of two loci in multiple generations. In bacteria like *E. coli*, recombination is measured as the time of transfer and integration of DNA from one cell into the chromosome of another cell.

In some instances, the placement of genes on various physical maps has been referred to as a *genetic map* (i.e. "a gene map"), although only physical distances and locations were known. Cytogenetic maps lead to further confusion in terminology. In these maps, the presence or absence of a gene or map object is correlated with an observable genomic location (i.e. a chromosome band).

Another concept whose use results in some confusion is "locus." Classically, this term defined the location genes (i.e. "genetic locus"). More generally, however, a chromosomal locus, representing a location on a chromosome, can consist of objects other than genes. Examples include DNA sequences, sequenced tagged sites (STS), expressed sequenced tags (ESTs), probe sites, restriction enzyme sites, clone sites, centromeres, telomeres, and chromosomal breakpoints (such as those that occur naturally or are induced by ultraviolet breakage). To add to the confusion, new genetic markers based on anonymous DNA sequences now define genetic loci because they are used in genetic mapping experiments. In this entry, a *chromosomal locus* is any chromosomal location that has been identified in a distinctive manner.

Here, a physical map is considered to be any map consisting of objects that have been located by physical rather than by genetic methods. Thus, a physical map can consist of objects located along the chromosome, such as a chromosomal band, a breakpoint, a genomic restriction fragment, and the location of a clone

on a chromosome. Here, the ordering of restriction fragments and clones will be used. These maps are constructed using molecular methods.

3
Evolving Approaches for Physical Mapping

The term "physical maps" has been used to describe both genomic restriction maps and genomic clone libraries. Low-resolution whole-genome restriction maps are created by ordering fragments that have been fractionated by size electrophoretically. These maps have been created by hybridization experiments using cloned sequences as probes or by PCR (polymerase chain reaction) testing for the presence of specific STSs/ESTs.

Another physical type of map is a radiation hybrid map. In this case, randomly fragmented DNA is cloned to create a hybrid cell line library. Different fragments are present in each hybrid cell line that is tested for the presence of a library of loci. The distance loci are reflected in the frequency of co-occurrence of two loci in the same cell line.

Also, genome restriction maps may be created by analysis of restriction sites contained on overlapping clones. The resolution of such restriction maps depends on the restriction enzyme used, and the frequency at which its recognition site occurs within the genomic DNA sample. Furthermore, an overlapping library may in itself represent a map consisting of ordered objects whose size can be approximated but not stated with certainty.

More efficient "cloneless" library approaches use PCR to analyze gel slices of lanes containing specific fragments. Here, the gel slices represent the cloneless library units that can be used in place of cloned DNA. In the future, cloneless approaches will likely replace time-consuming clone library construction and ordering.

4
Genomic Restriction Maps

The first step in creating a genomic restriction map is choosing the DNA source. For many organisms or chromosomes, this is obvious, since there is a well-characterized isolate or a cell line that may be useful. For most eukaryotic genomes, if available, some complications associated with the analysis of polymorphic diploid DNA can be avoided by using DNA from hybrid cell lines containing a single chromosome.

Usually, the genomic DNA is extracted and purified intact, in agarose to prevent shear damage. Small, 10-Mb chromosomal DNAs may be sized directly by PFG. Other chromosomal DNAs must be cleaved with a restriction enzyme before they are subjected to PFG analysis. The largest size standard for PFG electrophoresis is 6 Mb. Thus, analytical results use this limit, although enhanced versions of the technique allow the fractionation of molecules greater than 6 Mb.

Usually, maps are constructed using restriction enzymes that have large recognition sequences or a recognition site that occurs infrequently in the genome of interest. In some cases, it has been useful to test a battery of enzymes. The usefulness of a particular enzyme may be estimated roughly from the size of the site or the GC content of the test organism. Even so, these predictions are somewhat inaccurate because the genomic DNA sequences do not occur at random, whereas calculated occurrences assume randomness. Furthermore, the frequency of occurrence

in known sequences may not be representative of the entire genome because in the past many molecular studies have focused on gene sequencing.

Besides restriction enzymes, a number of other enzymes or methods have been adapted for cutting genomic DNA into specific, large pieces. These methods usually depend on the formation of an unusual structure (e.g. a triplex or D-loop structure) having an associated single-stranded region. D-loops are formed with the aid of RecA protein at targeted sites. In one strategy, termed the *Achilles' heel* strategy, or RARE (RecA-assisted restriction endonuclease) cleavage, genomic DNA containing a D-loop is treated with DNA methylase. The methylase recognizes and modifies specific sequence except when the sites occur within the D-loop. The methylase is removed, and the DNA is then subjected to digestion with the cognate restriction enzyme. Only the restriction enzyme recognition site in the previously formed D-loop is now susceptible to cleavage by the restriction enzyme.

This type of approach allows for site-directed cleavage of genomic DNA at specific locations and has been used to map the end of human chromosomes, as only one RARE cleavage is required. The combination of two RARE reactions from nearby sites would produce an interstitial fragment whose length would be equivalent to the distance between the two sites. Furthermore, fragments smaller than 10 Mb in size could be purified from the remaining genomic DNA using PFG electrophoresis.

Genomic map construction is similar to putting a puzzle together. It is much easier if all the pieces are identified in advance. For small genomes, this is accomplished with a restriction enzyme that produces reasonable number of fragments resolvable by PFG analysis. In these experiments, all genomic DNA is visualized by simple ethidium bromide staining.

A method was developed for visualizing all the megabase restriction fragments for individual chromosomes from large genomes. This approach analyzes PFG-fractionated restriction fragments from genomic DNA that is obtained from monosomic hybrid cell lines. Here, a species-specific interspersed repetitive hybridization probe is used to identify the megabase restriction fragments. For example, the megabase restriction fragments from chromosome 20 are visualized with a human-specific *Alu* probe in Fig. 1. This approach works as long as the restriction enzyme digestion goes to completion. Otherwise, partial digestion products can obscure the complete and especially larger digestion fragments. For the human genome, the restriction enzymes *Not* I, *Sgr*A I, *Fse* I, and *Asc* I have this desired characteristic and cut the genome into fragments that appear to average 1 Mb in size. In higher eukaryotic organisms, partial cleavage by a restriction enzyme is due to partial CpG methylation of their recognition sites inhibiting cleavage.

Most organisms will have several interspersed repeats that are of different lengths. The most commonly occurring interspersed repeat in the human genome is the *Alu* repeat. This repeat is a short interspersed repeating element (SINE) that is estimated to occur about every 5000 bp. The second most commonly occurring interspersed human repeat is the *Kpn* repeat. The *Kpn* element, a LINE (long interspersed) repeat, is estimated to occur at a frequency tenfold less than that of the *Alu* repeat. Theoretically, these two repeats used as hybridization probes to hybrid cell line DNA cleaved with the appropriate enzyme should reveal all the

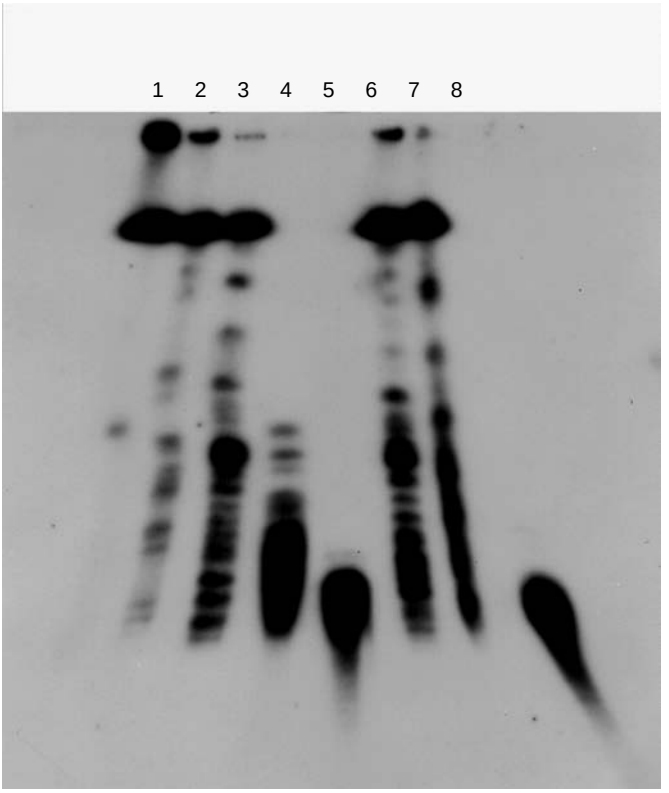

Fig. 1 Detection of PFG-fractionated restriction fragments of human chromosomes. DNA from a monosomic hybrid cell line containing chromosome 20 was digested to completion with different enzymes, PFG fractionated and hybridized to the human-specific *Alu* probe. The recognition sites for the enzymes used are: (1) *Sgf* I (GCGATCGC); (2) *Asc* I (GGCGCGCC); (3) *Fse* I (GGCCGG/CC); (4) *Pme* I (GTTTAAAC); (5) *Pac* I (TTAATTAA); (6) *Not* I (GCGGCCGC); (7) *Sgr*AI (CRCCGGYG); and (8) *Swa* I (ATTT/AAAT).

human megabase fragments. However, this approach may not distinguish two fragments of the same or very similar sizes even with high-resolution PFG fractionations.

The use of repeat sequences as hybridization probes can reveal information about the size and distribution of restriction fragments, but do not reveal order. Regional specific repetitive sequences, like those that occur at telomeric and centromeric regions may be used to identify the ends of physical map, as well as to provide an important anchor for regions of condensed chromosomes visualized microscopically. In human-hybrid cell lines, use of such sequences identifies ends of individual chromosomes.

In conventional mapping experiments, interstitial megabase fragments are linked to single-copy sequences by hybridization experiment or by the more efficient

PCR approach described in the following. Single-copy sequences, also called *STSs*, that have been located on genetic maps can be used for the regional assignment of restriction fragments and can serve as anchors between genetic and physical maps. The accuracy of fragment location will be a function of the resolution of the genetic map. When this approach is applied to a sample like that in Fig. 1, the cataloged megabase restriction fragments containing interspersed repeats are linked to specific single-copy sequences and, if known, genomic locations. ESTs are STS linked to mRNAs and their use would make a map of expressed genes. Note that any PCR primer pair can be used for mapping even if the genetic location of the specific sequence is unknown.

Complete genomic restriction maps do not identify neighboring fragments. At least one neighbor can be identified unambiguously using an STS to analyze partially digested DNA. The difficulty of interpreting the partial digestion data increases dramatically with the number of partial digestion products. For instance, it is important to consider that both the neighboring fragments may be of the same size, that is, the possibility that a single partial product band could represent two different products of the same size must always be borne in mind. This means that neighboring fragments must be confirmed in partial digestion experiments using STSs from the two complete digest fragments. Many times, the confusion associated with partial digest data can be sorted out by obtaining partial digest data on nearly adjacent fragments. In this approach, different sets of partial digest fragments are identified by sequences on different and complete digest fragments. The correct map interpretation is one that is consistent with all the data.

The analysis of partial products is simplified if the STS is from a telomere region or is situated on a small fragment located next to a very large fragment. In both cases, the partial digest information reflects the order of fragments in only one direction. Otherwise, the partial digest products results will represent bidirectional information. This approach also eliminates the problem of comigration of same-size products associated with bidirectional partial mapping data.

Megabase restriction fragments themselves may also be fingerprinted by cleavage with a second enzyme. The products are best analyzed using probes located at the end of the original megabase fragment. The partial digest product data will be different for the two probes located at different positions. The interpretation of the distribution of the cleavage sites of the second enzyme must be consistent with data obtained using the two interstitial (or terminal) probes. This approach is particularly useful for analyzing megabase fragments that are at the limit of the PFG technique, because the products of the second enzyme will be within the PFG limit.

There are several other approaches to proving that two restriction fragments are adjacent. By far the most powerful is the use of linking libraries – small-insert libraries that contain DNA segments from adjacent megabase restriction fragments. Hence, a *Not* I linking clone used as a hybridization probe to genomic DNA digested with *Not* I will identify adjacent *Not* I fragments. A complete linking library would suffice to construct, in the most efficient manner, a complete genomic restriction map. Without a complete linking library, partial digest strategies combined with polymorphism link-up and double restriction enzyme strategies can provide information on neighboring fragments. In

both approaches, parallel analysis of a region provides a regional signature.

The same polymorphism link-up refers to the results obtained when one probe is used to analyze restriction enzyme digested DNA from different cells. For instance, a restriction site may be missing in one cell line, either because there is a mutation at the restriction enzyme site or because methylation is interfering with restriction enzyme cleavage. When a number of DNA sources are examined, it is often possible to fingerprint the polymorphism of a restriction site (i.e. assuming that the site is polymorphic). The probes that are usually located on separate, but adjacent, fragments can, in some DNAs, be found on the same fragment. Many times this latter fragment will be equal to the total size of the two smaller fragments seen in other cell lines. In such a case, the pattern of occurrence of the two distinct smaller fragments, detected by the different probes, is self-consistent. This is very similar to the more familiar approach of fingerprinting a DNA using many different restriction enzymes.

5 Genomic Clone Libraries

The construction and ordering of an overlapping library begins with the selection of the genomic DNA sample as discussed earlier. The next decision entails choice of the type of vector that will be used. The most distinguishing feature of vector possibilities is the size of the DNA that can be cloned into them. It is helpful to begin whole-genome mapping projects using the largest DNA fragments possible, minimizing the number of clones to order. In the past, the largest cloning vectors were cosmids that contained 40-kb cloned sequences. Today, several systems are available for the cloning of larger segments. For instance, megabase fragments can be cloned into YACs (yeast artificial chromosomes) in yeast cells. These libraries have been replaced by the easier to use and more stable PAC (P1 artificial chromosomes) and BAC (bacterial artificial chromosomes) libraries that clone fragments of about 100 kb in size. Today, the construction and ordering of PAC/BAC libraries still remains a major effort.

Usually, today's top-down sequencing approaches begin by creating large clone BAC libraries. Then, random, small overlapping fragments of each BAC are subcloned into sequencing vectors. The large clone libraries are ordered using a variety of techniques (see the following), whereas sequencing is done on an unordered library. Ultimately, the small-insert library is ordered when overlaps are identified while the sequence is put together.

Clone Ordering. In the past, chromosome walking (or "crawling") experiments, involved hybridizing individual clones to an entire genomic library to detect overlapping clones. Although this approach is guaranteed to detect overlapping clones, it is slow and laborious.

Today, many efficient genomic approaches are used to order all the clones in a library. The genomic approaches analyze many samples at the same time in a "multiplex" approach and generally use methods that are easy to automate. For instance, clones are analyzed in pools to minimize the number of experiments that are needed to order the entire library. In some cases, a tiered pooling strategy is used, where the first testing is done on a small group of superpools, each composed of different pools. Only the pools that make

up a positive superpool needs to be further analyzed. Some designed pooling strategies have each clone present in multiple but distinct pools. The pattern of positive signals to such a set of pools will reveal the identity of the overlapping clones.

Most bottom-up, fingerprinting methods require that libraries be constructed using partial restriction enzyme digests to ensure that overlaps exist between different clones and that all or most of the genome is cloned. There are a number of ways to fingerprint clones to identify overlaps. Bottom-up strategies for library ordering usually involve testing individual clones to search for an overlapping restriction fragment or restriction site pattern. This type of fingerprinting is easy to automate. Clones may also be fingerprinted using interspersed repetitive sequences. Here, clones, or restriction fragments of the clones, are hybridized to a set of short oligonucleotides or longer interspersed repetitive sequences, and the patterns of hybridization are used to link clones.

A variation in this approach that has proven to be quite efficient is the use of genomic DNA restriction fragments as hybridization probes to order clone libraries. Megabase restriction fragment probes allow the assignment of clones to particular fragments. If the fragments are mapped, the clones can then be regionally assigned in the genome. Clones located in different regions are ordered simultaneously by further hybridization experiments using probes generated from a pool of restriction fragments. Here, the pools are created from gel slices containing small restriction fragments. Each pool consists of genomic restriction fragments of a particular size range from different regions of the genome. Overlapping clones will be located on the same megabase fragment and hybridize to the same pool of small fragments. This approach increases the efficiency of library ordering by eightfold.

Another ordering approach involves detecting overlaps by DNA sequencing. Here, the sequences may be chosen at random sites (e.g. STSs) or the sequences can be collected at specific locations originally called *sequence-tagged restriction sites* (STARs). The STAR approach may be used with both partial and complete digest libraries. DNA sequence information may be collected at specific restriction sites, including the ends of the cloned sequences. Clone linking is done using a linking clone library that spans the restriction enzyme site when the first library was created from completely digested DNA. Clones that generated from partial restriction enzyme digestions have ends that overlap. The later method eliminates the need for up-front clone ordering, but a highly representative library is required, and sequence read lengths must be long enough to span most small repeats.

6
Cloneless Genomic Libraries

Cloneless libraries have several advantages over clone libraries. These libraries can be made rapidly and inexpensively from any genomic source and all genomic DNA is present in the library. These libraries are made from agarose or acrylamide gels containing size fractionation DNA of low or high complexity. The gel slice containing the cloneless fraction is treated in the same manner as a cloned DNA sample.

In most of the discussion below, only complete digest cloneless libraries are discussed. However, partially digested or random broken DNAs could be used to

generate a cloneless library. In essence, the analysis and use of a cloneless library is very similar to that of a clone library.

The gel lane is carefully sliced into thin pieces using a clean coverslip for each cut to insure minimum cross-contamination. The DNA in each slice is used as a source for testing, for instance, with PCR. For small and medium size genome, the number of unique fragments within a slice will be small. For large genomes, there are many fragments in a single slice. In the case of samples from the gels, like those shown in Fig. 1, the only visible fragment is from the human DNA but other rodent fragments are present. The cloneless approach is a hybrid between genomic mapping and clone ordering.

One consideration with the cloneless approach is that DNA in agarose is believed to be difficult to manipulate enzymatically. However, others and we have developed robust and reliable simple methods for using DNA in agarose that avoids purification. Most methods were developed some time ago for PFG applications. Additional methods were developed more recently for PCR that involve the addition of preservatives to the stored fraction and the PCR.

Another potential problem with cloneless fractions is the low amount of DNA in each slice. There are a number of PCR methods that solve this problem by randomly amplifying the DNA in a fraction (or clone) to generate a large amount of additional material. Furthermore, these methods have been developed to minimize the introduction of mutations. Hence, the amplified DNA appears to be representative of the starting template.

Cloneless fractions can be used for a variety of experiments in addition to that of genomic sequencing. For instance, a recently developed method termed *comparative genome hybridization* (CGH) uses differentially labeled cDNA, or genomic DNA, from two samples. The labeled samples are mixed together in equal amounts and used as a hybridization probe. The target samples are metaphase chromosomes or clones. Quantitative analysis of the hybridization signal provides information on the relative amounts of DNA from different chromosomal regions. In summary, CGH experiments provide positional information of differences between samples.

An alternative to conventional CGH would utilize genome DNA arrays that have been generated from size-fractionated genomic restriction fragments. An array consisting of 2-mm slices from a 15-cm gel lane would have ~40-Mb resolution. Each fraction could be cleaved with a second enzyme and subjected to a second electrophoretic size fractionation. The ~1000 slices generated from the entire second fractionation would provide DNA at ~3-Mb resolution.

The cloneless, like a clone library approach does not provide positional genomic information directly. For instance, clones are linked to genomic position by FISH (fluorescence *in situ* hybridization) experiments to metaphase chromosomes. Similarly, the DNA in each slice can be used in a FISH experiment to metaphase chromosome and locate it to a specific genomic region. Genomic restriction maps, or ordered clone, or cloneless libraries can be used in place of metaphase chromosomes especially in those organisms, like protozoan, whose chromosomes do not condense during cell division.

The approaches used to order clone libraries may be applied to cloneless libraries. For instance, fraction can be linked to STSs/ESTs. Alternatively, hybridization experiments or other methods that were

described earlier can be used to detect overlaps with clone libraries or a second cloneless library generated with a second enzyme.

Cloneless libraries can be used in place of X-ray hybrid maps. X-ray hybrid maps are created by determining the frequency that two loci are separated by random breaks in DNA generated by X rays. In this approach, randomly broken DNA is cloned into a hybrid cell line and a library of different cell lines are created with different pieces of DNA. Then, each cell line is tested for the presence of a marker and a centiRay distance is calculated that reflects the frequency at which two loci occur in the same cell line. Like genetic mapping, the closer two loci are, the more likely they will be on the same randomly generate fragments. Here, instead of cell lines, randomly broken DNA fractionated by size could be used to create a cloneless library. Then each fraction could be tested for each loci and the distance between loci calculated as a function of the number of fractions that contain both loci.

7 Gap Closure

It is easy to start maps but difficult to finish them. Each gap in each map presents a unique problem. Gaps may be true gaps or pseudogaps. Pseudogaps contain repetitive sequences and require different strategies for closure than true gaps.

True gaps contain unique sequences and need to be treated individually. The best strategy will depend on the putative size of the gap, the amount of polymorphism in the region, the number of unassigned megabase fragments or clones. Gaps that arise from unclonable sequences may be filled using a cloneless library. For instance, STSs could be developed from cloneless fraction that overlap two regions as we have done for chromosome 21. For instance, single-copy human sequences located between *Alu* elements can be generated from inter-*Alu* PCR amplifications and used as probes to identify fragments in gap regions. Alternatively, the template contained within a fraction can be labeled and used as a hybridization probe to an array of samples at the borders of gaps. Ultimately, the DNA within a gap is identified in a cloneless library and directly sequenced using a direct genomic sequences approach.

8 Prospectus

What is the role of physical maps today and in the future? An increasing sequencing efficiency is coupled to a decreasing need to create and order expensive and time-consuming clone libraries. For small genomes, whole-genome sequencing projects use a bottom-up sequencing approach with unordered clone libraries. Soon a cloneless sequencing approach using "polonies" may replace the need for cloning of these small genomes. The polonies approach sequences pools of single DNA molecules that have been immobilized on beads. Sequencing is done on a library of immobilized DNAs simultaneously. Today, this method allows sequencing read of up to ~14 bases per fragment end, or ~28 bases per polony, and up to 1×10^9 beads/slide. Alternatively, imaging methods that directly view the sequence of individual DNA molecules may be used. Meanwhile, adaption of the cloneless approach for larger genome sequencing projects should considerably reduce *de novo* sequencing by eliminating

the need for creating clone libraries, as well as enable gap closure.

See also Chromosome Organization within the Nucleus; DNA Libraries; Gene Distribution in the Human Genome; Genome Scanning Method; Restriction Landmark Genomic Scanning (RLGS); Genomic DNA Libraries, Construction and Applications; Genomic Sequencing (Core Article); Nucleic Acid and Protein Sequence Analysis and Bioinformatics; Nucleic Acids (DNA) Sequencing, Transcriptional; Shotgun Sequencing (SGS).

Bibliography

Books and Reviews

Cantor, C.R., Smith, C.L., Matthew, M.K. (1988) Pulsed field gel electrophoresis of very large DNA molecules, *Annu. Rev. Biophys. Biophys. Chem.* **17**, 287–304.

Ledbetter, S.A., Nelson, D.L. (1991) Genome Amplification Using Primers Directed to Interspersed Repetitive Sequences (Irs-Pcr), in: McPherson, M.J., Quirke, P., Taylor, G.R. (Eds.) *Polymerase Chain Reaction I: A Practical Approach*, IRL Press, Oxford.

Murphy, M.I., Hammond, H.A., Caskey, C.T. (1996) Solid-phase automated sequencing of Pcr-amplified genomic DNA, *Methods Mol. Biol.* **65**, 163–176.

Schalkwyk, L.C., Francis, F., Lehrach, H. (1995) Techniques in mammalian genome mapping, *Curr. Opin. Biotechnol.* **6**(1), 37–43.

Smith, C.L., Condemine, G. (1990) New approaches for physical mapping of small genomes, *J. Bacteriol.* **172**(3), 1167–1172.

Smith, C.L., Lawrance, S.K., Gillespie, G.A., Cantor, C.R., Weissman, S.M., Collins, F.S. (1987) Strategies for mapping and cloning macroregions of mammalian genomes, *Methods Enzymol.* **151**, 461–489.

Smith, C.L., Klco, S., Zhang, T.Y., Fang, H., Oliva, R., Fan, J.B., Bremer, M., Lawrence, S. (1993) Analysis of Megabase DNA Using Pulsed Field Gel Electrophoresis, in: Adolph, K.W. (Ed.) *Methods in Molecular Genetics*, Academic Press, San Diego, CA.

Van der ploeg, L.H.T., Gottesdiener, K.M., Korman, S.H., Weiden, M., Le Blancq, S. (1992) Protozoan Genomes, in: Ulanovsky, L., Burmeister, M. (Ed.) *Methods in Molecular Biology, Pulsed-Field Gel Electrophoresis. Protocols, Methods, and Theories*, Humana Press, Totowa, NJ.

Welsh, J., McClelland, M. (1994) Fingerprinting Using Arbitrarily Primed Pcr: Application to Genetic Mapping, Population Biology, Epidemiology, and Detection of Differentially Expressed Rnas, in: Mullis, K.B., Ferre, F., Gibbs, R.A. (Eds.) *The Polymerase Chain Reaction*, Birkhauser, Boston, MA..

Primary Literature

Albarghouthi, M.N., Barron, A.E. (2000) Polymeric matrices for DNA sequencing by capillary electrophoresis, *Electrophoresis* **21**(18), 4096–4111.

Barnes, W.M. (1994) Pcr amplification of up to 35-Kb DNA with high fidelity and high yield from lambda bacteriophage templates, *Proc. Natl. Acad. Sci. U.S.A.* **91**(6), 2216–2220.

Cai, W., Jing, J., Irvin, B., Ohler, L., Rose, E., Shizuya, H., Kim, U.J., Simon, M., Anantharaman, T., Mishra, B., Schwartz, D.C. (1998) High-resolution restriction maps of bacterial artificial chromosomes constructed by optical mapping, *Proc. Natl. Acad. Sci. U.S.A.* **95**(7), 3390–3395.

Dederich, D.A., Okwuonu, G., Garner, T., Denn, A., Sutton, A., Escotto, M., Martindale, A., Delgado, O., Muzny, D.M., Gibbs, R.A., Metzker, M.L. (2002) Glass bead purification of plasmid template DNA for high throughput sequencing of mammalian genomes, *Nucleic Acids Res.* **30**(7), e32.

Ferrin, L.J., Camerini-Otero, R.D. (1994) Long-range mapping of gaps and telomeres with Reca-assisted restriction endonuclease (Rare) cleavage, *Nat. Genet.* **6**(4), 379–383.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty,

B.A., Merrick, J.M. et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* rd, *Science* **269**(5223), 496–512.

Florijn, R.J., Bonden, L.A., Vrolijk, H., Wiegant, J., Vaandrager, J.W., Baas, F., den Dunnen, J.T., Tanke, H.J., van Ommen, G.J., Raap, A.K. (1995) High-resolution DNA fiber-fish for genomic DNA mapping and colour barcoding of large genes, *Hum. Mol. Genet.* **4**(5), 831–836.

Fonstein, M., Haselkorn, R. (1995) Physical mapping of bacterial genomes, *J. Bacteriol.* **177**(12), 3361–3369.

Goldenberger, D., Perschil, I., Ritzler, M., Altwegg, M. (1995) A simple ''Universal'' DNA extraction procedure using Sds and proteinase K is compatible with direct Pcr amplification, *PCR Methods Appl.* **4**(6), 368–370.

Green, E.D. (2001) Strategies for the systematic sequencing of complex genomes, *Nat. Rev. Genet.* **2**(8), 573–583.

Hatada, I., Hayashizaki, Y., Hirotsune, S., Komatsubara, H., Mukai, T. (1991) A genomic scanning method for higher organisms using restriction sites as landmarks, *Proc. Natl. Acad. Sci. U.S.A.* **88**(21), 9523–9527.

Heiskanen, M., Kallioniemi, O., Palotie, A. (1996) Fiber-fish: experiences and a refined protocol, *Genet. Anal.* **12**(5–6), 179–184.

Hozier, J., Graham, R., Westfall, T., Siebert, P., Davis, L. (1994) Preparative in situ hybridization: selection of chromosome region-specific libraries on mitotic chromosomes, *Genomics* **19**(3), 441–447.

Kallioniemi, A., Kallioniemi, O.P., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F., Pinkel, D. (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors, *Science* **258**(5083), 818–821.

Kim, U.J., Birren, B.W., Slepak, T., Mancino, V., Boysen, C., Kang, H.L., Simon, M.I., Shizuya, H. (1996) Construction and characterization of a human bacterial artificial chromosome library, *Genomics* **34**(2), 213–218.

Kovacic, R.T., Comai, L., Bendich, A.J. (1995) Protection of megabase DNA from shearing, *Nucleic Acids Res.* **23**(19), 3999–4000.

Lawrence, J.B., Carter, K.C., Gerdes, M.J. (1992) Extending the capabilities of interphase chromatin mapping, *Nat. Genet.* **2**(3), 171–172.

Marra, M.A., Kucaba, T.A., Dietrich, N.L., Green, E.D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D., Waterston, R.H. (1997) High throughput fingerprint analysis of large-insert clones, *Genome Res.* **7**(11), 1072–1084.

Michalet, X., Ekong, R., Fougerousse, F., Rousseaux, S., Schurra, C., Hornigold, N., van Slegtenhorst, M., Wolfe, J., Povey, S., Beckmann, J.S., Bensimon, A. (1997) Dynamic molecular combing: stretching the whole human genome for high-resolution studies, *Science* **277**(5331), 1518–1523.

Newell, W., Beck, S., Lehrach, H., Lyall, A. (1998) Estimation of distances and map construction using radiation hybrids, *Genome Res.* **8**(5), 493–508.

Paez, J.G., Lin, M., Beroukhim, R., Lee, J.C., Zhao, X., Richter, D.J., Gabriel, S., Herman, P., Sasaki, H., Altshuler, D., Li, C., Meyerson, M., Sellers, W.R. (2004) Genome coverage and sequence fidelity of Phi29 polymerase-based multiple strand displacement whole genome amplification, *Nucleic Acids Res.* **32**(9), e71.

Ricke, D.O., Liu, Q., Gostout, B., Sommer, S.S. (1995) Nonrandom patterns of simple and cryptic triplet repeats in coding and noncoding sequences, *Genomics* **26**(3), 510–520.

Smith, C.L., Warburton, P., Gaal, A., Cantor, C.R. (1986) Analysis of genome organization and rearrangements by pulsed gradient gel electrophoresis, *Genet. Eng.* **8**, 45–70.

Wu, C.C., Nimmakayala, P., Santos, F.A., Springman, R., Scheuring, C., Meksem, K., Lightfoot, D.A., Zhang, H.B. (2004) Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping, *Theor. Appl. Genet.* **109**(5), 1041–1050.

Zabarovska, V.I., Gizatullin, R.Z., Al-Amin, A.N., Podowski, R., Protopopov, A.I., Lofdahl, S., Wahlestedt, C., Winberg, G., Kashuba, V.I., Ernberg, I., Zabarovsky, E.R. (2002) A new approach to genome mapping and sequencing: slalom libraries, *Nucleic Acids Res.* **30**(2), E6.

Whole Genome Shotgun Sequencing: *see* Shotgun Sequencing (SGS)

Xenotransplantation in Pharmaceutical Biotechnology

Gregory J. Brunn[1,2] *and Jeffrey L. Platt*[1,3]
[1] *Transplantation Biology Program, Mayo Clinic, Rochester, MN*
[2] *Department of Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN*
[3] *Departments of Surgery, Immunology and Pediatrics, Mayo Clinic, Rochester, MN*

1 Introduction 575
1.1 Alternative Sources of Transplantable Tissues 575
1.2 The Pig as a Source of Tissues and Organs for Clinical Xenotransplantation 576

2 The Biologic and Immunologic Responses to Xenotransplantation 577
2.1 Graft Vascularization 577
2.2 Hyperacute Rejection 577
2.3 Complement Activation 579
2.4 Acute Vascular Rejection 581
2.5 Accommodation 583
2.6 Cellular Mediated Immune Responses 583

3 Physiologic and Infectious Hurdles to Xenotransplantation 584
3.1 Interspecies Challenges 584
3.2 Zoonosis 585

4 A Scenario for the Clinical Application of Xenotransplantation 586

Bibliography 587
Books and Reviews 587
Primary Literature 587

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

Keywords

Accommodation
Establishment of an organ transplant in a recipient from whom antidonor antibodies have been depleted. Rejection does not ensue when antidonor antibodies return to the circulation because of changes in the antibody, in the antigen, or in susceptibility of the organ to rejection.

Acute Vascular Rejection
A rejection that may begin within 24 hours of implantation of a vascular graft and leading to graft destruction over the ensuing days to weeks. Mediated by graft reactive antidonor antibodies, activation of the complement cascade, and possibly NK cells and macrophages.

Genetic Engineering
A general term that covers the use of various experimental techniques to produce molecules of DNA-containing new genes or novel combinations of genes, usually for insertion into a host cell for cloning.

Humoral Immunity
The component of the immune system involving B-lymphocytes, antibodies, and related cytokines. B-lymphocytes respond to the presence of antigen and differentiate to antibody-secreting plasma cells. The antibodies in turn neutralize antigen, activate phagocytic cells, recruit lytic enzymes and activate the complement cascade to destroy antigen-bearing cells.

Hyperacute Rejection
A rejection that usually develops in less than one hour from the implantation of a vascular graft, a form of antibody-mediated, usually irreversible, damage to a transplanted organ manifested predominantly by diffuse thrombotic lesions, usually confined to the organ itself and only rarely disseminated.

Xenotransplantation
The surgical removal of an organ or tissue from one species and transplanting it into a member of a different species, for example: the use of a baboon heart in a human being.

Zoonosis
The transmission of a disease from an animal or nonhuman species to humans. The natural reservoir is a nonhuman animal.

For nearly a century, xenotransplantation has been seen as a potential curative approach to replacing the function of diseased or damaged organs. Until recently, however, the application of xenotransplantation has seemed only a remote possibility. Genetic engineering of large animals has allowed the ability to add genes and remove genes from lines of animals, which could provide an enduring source of tissues and organs for clinical application. Genetic engineering could address the immunologic, physiologic, and infectious barriers to xenotransplantation, and could expand the utility of xenotransplantation to provide a source of cells with defined expression of exogenous genes for treatment of human disease.

1 Introduction

1.1 Alternative Sources of Transplantable Tissues

No area of medicine has generated more excitement or controversy than the field of transplantation. Organ allotransplantation allows "curative" treatments for failure of the heart, kidney, liver, and lungs by replacing these diseased organs with physiologically normal ones. Replacement of β cells via pancreatic islet or whole pancreas transplantation offers curative treatment to patients with diabetes. The main limitation to applying transplantation for the treatment of disease is a shortage of human donors. This shortage limits the clinical application of organ transplantation to approximately 5% of the number of transplants that would be performed where the supply of organs is unlimited. Possible solutions to this limitation have garnered considerable interest and include use of artificial organs, "engineered tissues," stem cell transplants and xenotransplants. Although some newer technologies have excited interest, xenotransplants of the heart, lung, kidney, and pancreatic islets are known to function well enough to sustain life. Enthusiasm for xenotransplantation also stems from the possibility that animal tissues and organs might be less susceptible to disease recurrence compared to allotransplants. Advances in cellular and molecular biology and in genetics open possibilities for use of cells, tissues, and organs to address the complications of disease, not only by replacement of abnormal cells and tissues but also by the use of transplanted tissues to impart novel physiological functions. In this regard and for some purposes, xenografts may be an ideal vehicle for introducing a novel gene or biochemical process that could be of value to the transplant recipient.

If interest in xenotransplantation is substantial, the hurdles to its application are equally so. For the past three decades, the first and preeminent obstacle to transplanting organs and tissues between species has been the immune reaction of the host against the graft. A second, and still theoretical, hurdle is the possibility that beyond the immune barrier, there might be physiologic limitations to the survival or function of a xenograft and the possibility that a xenotransplant might engender medical complications for the xenogeneic host. A third hurdle is the possibility that a xenograft might transfer

infectious agents from the donor to the host and that from the host such agents might spread to other members of society. This communication will consider the current state of efforts to overcome the various hurdles to xenotransplantation and will evaluate how genetic engineering might be applied to this end.

1.2 The Pig as a Source of Tissues and Organs for Clinical Xenotransplantation

Although it might be intuitive that the best source of xenogeneic tissues for clinical transplantation is nonhuman primates, it is the pig that is the focus of most efforts in this field. The reasons for favoring the pig as a xenotransplant source include the availability of pigs in large numbers, the ease with which the pig can be bred, the limited risk of zoonotic disease engendered by the use of pigs, and the possibility of introducing new genes into the germline of the pig.

Genetic engineering of pigs using transgenic techniques and nuclear transfer has certain advantages over conventional gene therapy (Table 1). Introducing genetic material directly into the porcine germ cell obviates the need for a vehicle, which may vary in reliability of gene delivery and may introduce secondary unintended consequences due to the vector itself. Second, the genetic material introduced into the germline can be expressed constitutively in

Tab. 1 Genetic engineering in xenotransplantation: conventional gene therapy versus transgenic therapy versus cloning.

	Conventional gene therapy	*Conventional transgenic techniques*	*Cloning*
Delivery	Vector or vehicle required	Injection of genetic material directly into pro-nuclei of fertilized egg	Transfection of cultured somatic cells
Expression	Dependent on ability of each cell to take up genetic material	Genetic material introduced into the germline, leading to expression in a line of animals	Genetic material introduced into the germline, leading to expression in a line of animals
	Requires treatment for every transplant or recipient		
	May require repeated treatment	One manipulation	One manipulation
Immunogenicity	Delivery vehicle or transgene may be immunogenic	The transgene may elicit immune response	The transgene may elicit immune response
Target of genetic manipulation	The recipient and the graft may be transduced	Genetic manipulation of the donor only	Genetic manipulation of the donor only
Genetic manipulation	Gene addition Dominant negative	Gene addition Dominant negative	Gene addition Dominant negative Gene knock out

all cells, especially in stem cells, and passed on to subsequent generations. Third, with the use of transgenic techniques, only the donor is manipulated; in conventional gene therapy, both the donor and the recipient may be affected.

Recent advances in cloning pigs through nuclear transfer also allow "knocking out" genes. Besides the advantage of gene knockouts, nuclear transfer can be done with cultured somatic cells obviating the need for embryonic stem cells.

2 The Biologic and Immunologic Responses to Xenotransplantation

2.1 Graft Vascularization

All xenografts elicit an immune response, including antibodies, cell-mediated immunity, natural killer cells and inflammation. However, the fate of xenografts confronted with these responses is dictated in part by the way in which the graft receives its vascular supply (Fig. 1). Isolated cells, such as hepatocytes, and "free" tissues, such as pancreatic islets and skin, derive their vascular supply through the ingrowth of host blood vessels. The process of neovascularization, as such, might be impaired in a xenograft by incompatibility of donor growth factors with the host microvasculature. To the extent that neovascularization or graft function depends upon hormones and cytokines of host origin, the function of the xenograft might also be impaired. As the host microcirculation is established, however, a xenogeneic tissue may be relatively protected from attack by host immune elements. Whole-organ grafts provide their own microcirculation and growth factors, and, as a result, incompatibility between the donor and the recipient is less likely to have an impact on cellular function. On the other hand, because the circulation is of donor origin, the immune, inflammatory, and coagulation systems of the recipient can act directly on donor cells, sometimes with dramatic and devastating consequences.

2.2 Hyperacute Rejection

An organ transplanted from a pig into a primate such as a human is subject to hyperacute rejection. Hyperacute rejection begins immediately upon reperfusion of the graft and destroys the graft within minutes to a few hours. Hyperacute rejection is characterized histologically by interstitial hemorrhage and thrombosis, the thrombi consisting mainly of platelets.

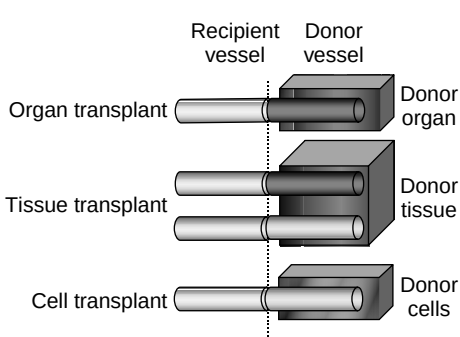

Fig. 1 Mechanisms of xenograft vascularization. Organ xenografts receive recipient blood exclusively through the donor blood vessels (top). Free tissue xenografts (e.g. Pancreatic islets and skin) are vascularized partly by the ingrowth of recipient blood vessels and partly by spontaneous anastomosis of donor and recipient capillaries (middle). Cellular xenografts (e.g. hepatocytes and bone-marrow cells) are vascularized by the ingrowth of recipient blood vessels (bottom).

Research over the past decade has clarified the molecular basis for the hyperacute rejection of pig organs by primates, and this knowledge has led to the development of new and incisive therapeutic approaches to averting this problem. Hyperacute rejection was once considered the most daunting hurdle to clinical application of xenotransplantation; however, hyperacute rejection can now be prevented in nearly every case.

Hyperacute rejection of porcine organ xenografts by primates is initiated by the binding of xenoreactive natural antibodies to the graft. Xenoreactive natural antibodies are present in the circulation without a known history of sensitization. Contrary to expectations, xenoreactive antibodies are predominantly directed against only one antigen, a saccharide consisting of terminal Galα1,3Gal. The importance of Galα1,3Gal as the primary antigenic barrier to xenotransplantation was demonstrated recently by experiments in which anti-Galα1,3Gal antibodies were specifically depleted from baboons using immunoaffinity columns before transplantation of pig organs. Antibody-binding to the newly transplanted organs was largely curtailed, and hyperacute rejection did not occur.

Although the identification of the relevant antigen for pig-to-primate xenotransplantation allows specific depletion of the offending antibodies, more enduring and less intrusive forms of therapy would be preferred. One approach to overcoming the antibody–antigen reaction is to develop lines of pigs with low levels of antigen expression. Various genetic approaches aimed at "remodeling" the antigenicity of donor tissues by reducing Galα1,3Gal expression are under investigation. These approaches can be separated into three categories: (1) interference with the function of α1,3-galactosyltransferase (α1,3GT), the enzyme that catalyzes the synthesis of the Galα1,3Gal moiety; (2) expression of α-galactosidase, which cleaves α-galactosyl residues; and (3) deletion of the gene encoding α1,3GT from the pig genome to prevent synthesis of the saccharide. The utility of genetic modification of pig tissues to reduce Galα1,3Gal expression has recently been demonstrated by Sharma and colleagues, who generated transgenic pigs expressing the H-transferase. Transgenic pigs expressing H-transferase express H antigen at the terminus of some sugar chains instead of Galα1,3Gal. Another genetic approach to modify expression of Galα1-3Gal was proposed by Osman et al. Expression of α-galactosidase, which cleaves α-galactosyl residues, in conjunction with other galactosyltransferases, significantly reduces expression of Galα1-3Gal. A third strategy to modify Galα1,3Gal was demonstrated by Miyagawa and coworkers who generated transgenic pigs expressing the human β-1,4 – *N*-acetylglucosaminyltransferase III gene (GnT-III). This enzyme catalyzes transfer of *N*-acetylglucosamine to maturing mannose-modified proteins as they pass through the Golgi apparatus and leads to diminished Galα1,3Gal expression both by competing with α1,3GT and by preventing subsequent modifications by α1,3GT by insertion of an *N*-acetylglucosamine onto the growing mannose chain. The net result was diminished natural antibody reactivity on transgenic pig tissue when transplanted into primates. Although these advances illustrate the utility of genetic modification of pig donor tissue, the main limitation of these genetic approaches is that residual Galα1,3Gal may be sufficient to allow rejection reactions to occur.

The most obvious approach to developing xenograft donors with diminished reactivity with host antibodies would be to genetically target or "knock out" the enzyme α1,3-galactosyltransferase. Embryonic stem cells were used to knock this gene out in mice, demonstrating that removal of this enzyme is not lethal. The intense interest in generating pigs that lack the α1,3GT gene has fueled a race to delete this gene from the pig genome, and recently several groups have succeeded in this endeavor. Prather and colleagues and Ayares and coworkers used similar strategies to disrupt one allele of the α1,3GT gene (GGTA1) in pigs by first targeting the gene for disruption in fetal porcine fibroblasts. Selected clones were used as nuclear donors for enucleated pig oocytes, the resulting embryos implanted into surrogate gilts. Both approaches yielded live, healthy piglet clones in which one copy of GGTA1 had been disrupted. These achievements demonstrated that nuclear transfer technology could be applied to pig embryos, which are notoriously fragile and difficult to manipulate. Recently, the generation of cloned pigs harboring a functional knockout of both alleles of α1,3GT was reported. This first success required some serendipity in that the α1,3GT-deficient pigs were found not to be homozygous knockouts, but rather a functional knockout arose from a process in which a knockout of one allele paired with a spontaneous single base change in the remaining GGTA1 gene resulted in an inactivating amino acid substitution in α1,3GT. Pigs with both alleles targeted were described by Dor and colleagues. These pigs were miniature swine with homozygous deletion of the α1,3GT gene (α1,3GT$^{-/-}$), do not express Galα1,3Gal epitopes on any tissue and, as might be expected, make anti-Galα1,3Gal reactive natural antibodies that are cytotoxic to cells from α1,3GT$^{+/+}$ pigs.

The availability of pigs with inactivation or deletion of α1,3GT should allow confirmation of the importance of Galα1,3Gal for hyperacute rejection, but may reveal other hurdles to xenotransplantation. For example, a cautionary note has recently come to light, indicating that pig cells that lack both copies of the GGTA1 gene may still synthesize the Galα1,3Gal antigen, albeit at very low levels. A further caution was suggested by Miyata et al., who found that murine lungs lacking Galα1,3Gal still bound natural antibodies and activate complement when perfused by human serum, and these events are associated with diminished lung perfusion. Thus, removal of Galα1,3Gal may not avert other potent xenogeneic immune responses, as will be discussed below.

2.3 Complement Activation

A second and essential step in the development of hyperacute rejection is activation of the complement system of the recipient on donor blood vessels. Complement activation is triggered by the binding of complement-fixing xenoreactive antibodies such as natural antibodies directed against Galα1,3Gal to graft endothelium and, to a smaller extent perhaps, by reperfusion injury. Regardless of the mechanism leading to complement activation, a xenograft is extraordinarily sensitive to complement-mediated injury because of multiple defects in the regulation of complement (Fig. 2). Under normal circumstances, the complement cascade is regulated or inhibited by various proteins in the plasma and on the surface of cells. These proteins protect normal

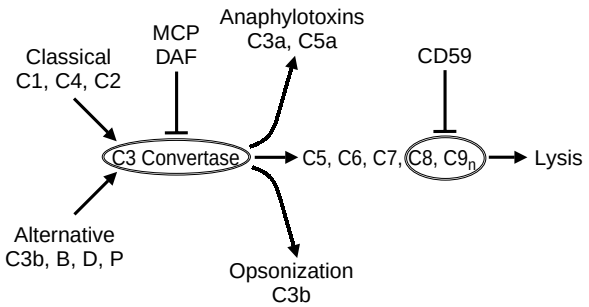

Fig. 2 Regulation of the complement system. The complement cascade, which can be activated via the classical or alternative pathway, is regulated under normal circumstances by various proteins in the plasma and on the cell surface. Three of the cell-surface complement regulatory proteins are shown here. Decay-accelerating factor (DAF) and membrane cofactor protein (MCP) regulates complement activation by dissociating or promoting the degradation of C3 convertase. CD59, also known as protectin, prevents the functions of terminal complement complexes by inhibiting C8 and C9. An organ graft transplanted into a xenogeneic recipient is especially sensitive to complement-mediated injury because decay-accelerating factor, membrane cofactor protein, and CD59 expressed on the xenograft endothelium cannot effectively regulate the complement system of the recipient.

cells from suffering inadvertent injury during the activation of complement. The proteins that regulate the complement cascade function in a species-restricted fashion; that is, complement regulatory proteins inhibit homologous complement far more effectively than heterologous complement. Accordingly, the complement regulatory proteins expressed in a xenograft are ineffective at controlling the complement cascade of the recipient, and the graft is subject to severe complement-mediated injury.

To address this problem, lines of animals have been developed that lack the Galα1,3Gal epitope recognized by natural antibodies (see Sect. 2.2) or that are transgenic for human complement regulatory proteins and that are able to control activation of complement in the xenograft (Fig. 2). Animals transgenic for human decay-accelerating factor (hDAF), which regulates complement at the level of C3 together with CD59, which regulates complement at the level of C8 and C9 or CD46, which controls complement activation at the level of C3 and C4, have demonstrated that the expression of even low levels of human decay-accelerating factor and CD59 or CD46 in porcine-to-primate xenografts is sufficient to allow a xenograft to avoid hyperacute rejection. These results, and the dramatic prolongation of xenograft survival achieved by expressing higher levels of human decay-accelerating factor in the pig, underscore the importance of complement regulation as a determinant of xenograft outcomes.

One of the major obstacles to testing the effects of transgenes in pig organs has

been the difficulty in generating transgenic pigs. Recent work by Lavitrano, et al. may accelerate the rate at which transgenic pigs may be generated and tested in transplant models. These investigators used sperm-mediated gene transfer to incorporate the human decay-accelerating factor gene into pigs and obtained a high efficiency of transgenesis (80% of pigs incorporated hDAF into the genome) and hDAF expression (43% of the transgenic pigs). The transgenic hDAF was functional *in vitro*, and transmitted to progeny as expected. This method in theory could be used to introduce multiple genes at once, or a tailor-made set of human genes that may be useful for transplant-mediated genetic therapies, as mentioned earlier.

2.4 Acute Vascular Rejection

If hyperacute rejection of a xenograft is averted, a xenograft is subject to the development of acute vascular rejection, so named because of its resemblance to acute vascular rejection of allografts. Acute vascular rejection (sometimes called *delayed xenograft rejection*) may begin within 24 hours of reperfusion and leads to graft destruction over the following days and weeks. Although the factors important in the pathogenesis of acute vascular rejection are incompletely understood, there is growing evidence that acute vascular rejection is triggered at least in part by the binding of xenoreactive antibodies to the graft. The importance of xenoreactive antibodies in triggering acute vascular rejection is suggested by three lines of evidence: (1) antidonor antibodies are present in the circulation of recipients whose grafts are subject to acute vascular rejection; (2) depletion of antidonor antibodies delays or prevents acute vascular rejection; and (3) administration of antidonor antibodies leads to the development of acute vascular rejection. Recent studies suggest that among the antibodies that provoke acute vascular rejection are those directed against Galα1-3Gal. This problem thus constitutes further impetus for the continued development of α1,3Gal-deficient pigs. Regardless of which elements of the immune system trigger acute vascular rejection, it is commonly thought that this type of rejection, and especially the intravascular coagulation characteristically associated with it, are caused by the activation of endothelial cells in the transplant. Activated endothelial cells express procoagulant molecules, such as tissue factor, and proinflammatory molecules, such as E-selectin and cytokines. The pathogenesis of acute vascular rejection is summarized in Fig. 3.

Although various therapeutic manipulations have proven successful in preventing hyperacute rejection, acute vascular rejection poses a more difficult problem, in part, because therapies are needed on an ongoing basis. For this reason, genetic modification of the donor may prove more important for dealing with acute vascular rejection than with hyperacute rejection. The various possible approaches for combating acute vascular rejection are listed in Table 2. Among these approaches, the reduction of Galα1,3Gal in xenotransplant donors may be an important part of the overall strategy, to the extent that Galα1,3Gal proves to be an important antigenic target in acute vascular xenograft rejection. Preliminary studies suggest that the level of antibody-binding needed to initiate acute vascular rejection is considerably lower than the level needed to initiate hyperacute rejection. Accordingly, the antigen expression

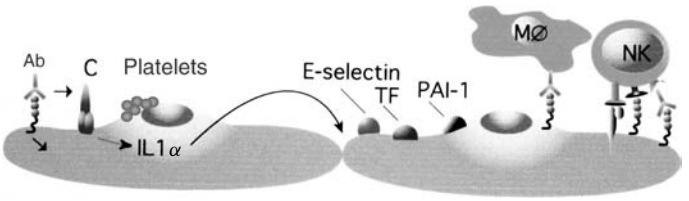

Fig. 3 Pathogenesis of acute vascular rejection. Activation of endothelium by xenoreactive antibodies (Ab), complement (C), platelets and perhaps by inflammatory cells (natural killer [NK] cells and macrophages [MØ]) leads to the expression of new pathophysiologic properties. These new properties, such as the synthesis of tissue factor (TF) and plasminogen activator inhibitor type 1 (PAI-1), promote coagulation; the synthesis of E-selectin and cytokines such as IL1α promote inflammation. These changes in turn cause thrombosis, ischemia, and endothelial injury, the hallmarks of acute vascular rejection. (Adapted from Platt, J. L. (1998) New directions for organ transplantation, *Nature* **392**(Suppl. 6679), 11–17, with permission).

Tab. 2 Therapeutic strategies for acute vascular xenograft rejection.

Possible mechanism targeted	***Manipulation of:***	
	Recipient	***Donor***
Antibody–antigen interaction	Specific depletion of xenoreactive antibodies Prevention of xenoreactive antibody synthesis (e.g. cyclophosphamide, leflunomide)	Generating transgenic pigs with low levels of antigen Generation of pig clones lacking antigen
Complement activation	Systemic anticomplement therapy (e.g. CVF, sCR1, gamma globulin)	Generation of donor pigs transgenic for human complement regulatory proteins
Endothelial-cell activation	Administration of anti-inflammatory agents	Inhibition of NFκB function Introduction of protective genes
Molecular incompatibilities	Administration of inhibitors (e.g. inhibitors of complement or coagulation)	Introduction of compatible molecules

would have to be reduced very significantly to achieve therapeutic benefit for acute vascular rejection. The availability of α1,3galactosyltransferase-deficient pigs should prove to be an ideal model to test this concept. In addition to lowering antigen expression, it is likely that expression of human complement regulatory proteins will be helpful in preventing acute vascular rejection. Preliminary studies suggest that interfering with the antigen–antibody reaction and controlling the complement cascade may be sufficient to prevent acute vascular rejection for at least

some period of time. These goals were accomplished by using animals transgenic for human decay-accelerating factor and CD59 as a source of organs and baboons depleted of immunoglobulin as recipients. Cozzi and associates achieved prolonged survival of xenografts, presumably preventing acute vascular rejection, by using transgenic pigs expressing high levels of decay-accelerating factor and cynomolgus monkeys treated with very high doses of cyclophosphamide. The immunosuppression perhaps prevented the synthesis of antidonor antibodies.

Work in rodents points to the potential involvement of natural killer (NK) cells and macrophages in mediating acute vascular rejection. However, the ability of immunoglobulin manipulation to prevent acute vascular rejection suggests that the involvement of NK cells and macrophages might be less important than *in vitro* studies and studies in rodents have suggested. On the other hand, NK cells might exacerbate the injury triggered by xenoreactive antibodies, as human NK cells have been shown to activate porcine endothelial cells *in vitro*.

2.5 Accommodation

Fortunately, the presence of antidonor antibodies in the circulation of a graft recipient does not inevitably trigger acute vascular rejection. If antidonor antibodies are temporarily depleted from a recipient, an organ transplant can be established so that rejection does not ensue when the antidonor antibodies are returned to the circulation. This phenomenon is referred to as *accommodation*. Accommodation may reflect a change in the antibodies, in the antigen, or in the susceptibility of the organ to rejection. If accommodation can be established, it may be especially important in xenotransplantation because it would obviate the need for ongoing interventions to inhibit antibody-binding to the graft. One potential approach to accommodation may be the use of genetic engineering to reduce the susceptibility of an organ transplant to acute vascular rejection and the endothelial-cell activation associated with it. Unfortunately, successful intervention at the level of such effector mechanisms has yet to be achieved. However, disruption of antibody–antigen interaction has brought about accommodation in human subjects.

2.6 Cellular Mediated Immune Responses

Organ transplants and cellular and free tissue transplants are subject to cellular rejection. In allotransplantation, cellular rejection is controlled by conventional immunosuppressive therapy, but there is concern that, for several reasons, cellular rejection may be especially severe in xenotransplants. First, the great variety of antigenic proteins in a xenograft may lead to recruitment of a diverse set of "xenoreactive" T cells. Second, the binding of xenoreactive antibodies and activation of the complement system may lead to amplification of elicited immune responses. For example, deposition of complement in a graft may cause activation of antigen-presenting cells, in turn stimulating T-cell responses. Still another factor that might amplify the elicited immune response to a xenotransplant involves "immunoregulation," which ordinarily would circumscribe cellular immune responses, but may fail or be deficient across species. Such failure could reflect limitations in

the recognition of xenogeneic cells or incompatibility of relevant growth factors, as but two examples.

Induction of immunologic tolerance has been an erstwhile goal of transplant surgeons and physicians. Especially in the case of xenotransplantation, if the current immunosuppressive regimens are not sufficient, induction of immunologic tolerance may be required. At least three approaches are being pursued: (1) the generation of mixed hematopoietic chimerism, (2) the establishment of microchimerism by various means, and (3) thymic transplantation. The development of mixed hematopoietic chimerism through the introduction of donor bone marrow has worked very well across rodent species, although success may be limited by xenoreactive antibodies and the engraftment impaired by incompatibility of host growth factors or microenvironment. Fortunately, there is evidence that these problems can be overcome. Various approaches to peripheral tolerance, such as the blockade of costimulation by administration of a fusion protein consisting of a soluble form of the CTLA-4 molecule and immunoglobulin (CTLA-4-Ig), are being pursued.

Still another factor in the cellular response to a xenotransplant involves the action of NK cells. Natural killer cell functions can be amplified by cell-surface receptors that recognize Galα1-3Gal. Natural killer cell functions are downregulated by receptors that recognize homologous major histocompatibility complex (MHC) class I. Human NK cells may be especially active against xenogeneic cells because of stimulation on the one hand and failure of downregulation on the other. The possible involvement of NK cells in xenograft rejection might be addressed by generation of transgenic pigs expressing on the cell-surface MHC-like molecules that will more effectively recognize corresponding receptors on NK cells and that will downregulate the function of NK cells.

How a xenogeneic donor could be modified genetically to enhance the development of tolerance or to limit elicited immune responses is still uncertain. Clearly, efforts to control the natural immune barriers to xenotransplantation may contribute to limiting the elicited immune response. To the extent that recipient T cells recognize donor cells directly, that is, the T-cell receptors of the recipient recognizing native MHC antigens on donor cells, a xenogeneic donor might be engineered in such a way to reduce corecognition (through CD4 and CD8) or costimulation (through CD28 or other T-cell surface molecules) or to express inhibitory molecules such as CD59 or Fas ligand. These approaches and the expression of inhibitory molecules, which are being considered as approaches to gene therapy in allotransplantation, may well prove more effective in xenotransplantation because inhibitory genes can be introduced as transgenes and thereby expressed in all relevant cells in the graft. Another useful and perhaps necessary approach will involve genetic modifications to allow the survival and function of donor bone-marrow cells.

3 Physiologic and Infectious Hurdles to Xenotransplantation

3.1 Interspecies Challenges

Progress in addressing some of the immunological obstacles to xenotransplantation has brought into focus the question of

the extent to which a xenotransplant would function optimally in a foreign host. A recent demonstration that the porcine kidney and the porcine lung can replace the most important functions of the primate kidney and primate lung is encouraging. Subtle defects in physiology across species may nevertheless exist. Organs such as the liver, which secrete a variety of proteins and which depend on complex enzymatic cascades, may prove incompatible with a primate host. Accordingly, one important application of genetic engineering in xenotransplantation may be the amplification or modulation of xenograft function to allow for more complete establishment of physiologic function or to overcome critical defects; for example, recent studies by Akhter and associates and Kypson and coauthors aimed at improving the function of cardiac allografts by manipulation of β-adrenergic signaling, and this technique might be adapted to the xenotransplant to improve cardiac function. On the other hand, most cellular processes and biochemical cascades are intrinsically regulated to meet the overall physiologic needs of the whole individual. The key question, then, is which of the many potential defects actually need to be repaired.

Another potential hurdle to the clinical application of xenotransplantation is the possibility that the xenograft may disturb normal metabolic and physiologic functions in the recipient. For example, Lawson and coworkers have shown that porcine thrombomodulin fails to interact adequately with human thrombin and protein C to generate activated protein C. This defect could lead to a prothrombotic diathesis because of failure of generation of activated protein C. Of even greater concern is the possibility that the transplantation of an organ, such as the liver, could add prothrombotic or proinflammatory products into the blood of the recipient. Although perhaps a great many physiologic defects can be detected at the molecular level, the critical question will be which of these defects is important at the whole-organ level or with respect to the well-being of the recipient, and which must be repaired by pharmaceutical or genetic means.

3.2 Zoonosis

The increasing success of experimental xenotransplants and therapeutic trials bring to the fore the question of zoonosis, that is, infectious disease introduced from the graft into the recipient. The transfer of infectious agents from the graft to the recipient is a well-known complication of allotransplantation. To the extent that infection of the recipient in this way increases the risks of transplantation, that risk can generally be estimated and a decision made on the basis of the risk versus the potential benefits conferred by the transplant. The concern about zoonosis in xenotransplantation is not so much the risk to the recipient of the transplant, but the risk that an infectious agent will be transferred from the recipient to the population at large. Fortunately, all of the microbial agents known to infect the pig can be detected by screening and potentially eliminated from a population of xenotransplant donors. There is concern, however, that the pig may harbor endogenous retroviruses that are inherited with genomic DNA and that might become activated and transferred to the cells of the recipient.

The possibility that an endogenous retrovirus from swine might pose a hurdle to xenotransplantation was first raised by Patience et al. Reports from several laboratories have raised concern about transmission of porcine retroviruses. Patience and coauthors found that the porcine endogenous retrovirus (PERV) can be activated in pig cells, leading to the release of particles that can infect human cell lines. Whether this virus or other endogenous viruses can actually infect across species and whether such infection would lead to disease are unknown, but remain a subject of current epidemiologic investigation. Ogle and colleagues found that human hematopoietic stem cells introduced into swine acquire PERV by fusing with host cells to form human–swine hybrid synkaryons. The hybrid cells appear to have a survival advantage over unfused cells and transmit the virus to uninfected human cells. If cross-species infection does prove to be an important issue, genetic therapies might also be used to address this problem. The simplest genetic therapy would involve breeding out the organism, but this approach might fail if the organism were widespread or integrated at multiple loci. Some genetic therapies have been developed to potentially control human immunodeficiency viruses. Although these therapies have generally failed because it has been difficult or impossible to gain expression of the transferred genes in stem cells and at levels sufficient to deal with high viral loads, the application of such therapies might be much easier in xenotransplantation because the therapeutic genes could be delivered through the germline. Ultimately, if elimination of endogenous retroviruses were necessary, it could potentially be accomplished by gene targeting and cloning, as discussed above.

4
A Scenario for the Clinical Application of Xenotransplantation

Successful application of xenotransplantation in the clinical arena requires insights into not only immunology but also into physiology and infectious disease, all of which have been discussed briefly here in the context of genetic therapy. In recent years, important advances have been made in elucidating the immunologic hurdles of pig-to-primate transplantation. Although this scientific progress is important and exciting, xenotransplantation will likely enter the clinical arena through a step-by-step process. A first step, free tissue xenografting, is in limited clinical trials already and preliminary evidence is encouraging as porcine free tissue xenografts appear to endure in a human recipient. One immediate application of free tissue xenografting would be treatment of cirrhosis caused by hepatitis virus, using targeted infusion of porcine hepatocytes. The promise of this approach is enhanced because (1) pig hepatocytes are resistant to viral reinfection, (2) rat models of cirrhotic liver failure indicate that porcine hepatocyte xenotransplants may endure and function well, and (3) predicted demand for hepatic transplantation due to hepatitis C-induced cirrhotic liver disease is likely to worsen the already acute shortage of livers available for transplant. Another potential extension of free tissue xenografting is transplantation of xenogeneic islets of Langerhans for the treatment of type one diabetes. Xenogeneic islet transplants may prove to be less liable to destruction by the autoimmune processes that underlies this disease. Temporary or ''bridge'' organ transplantation will probably follow free tissue xenografting. Bridge transplants will not address the problem of the shortage of

human organs, but incisive analysis of the outcomes of these transplants will provide important information about the remaining immunologic hurdles and the potential physiologic and infectious considerations. With this information, further therapies including genetic engineering may allow the use of porcine organs as permanent replacements. Even then, one can envision ongoing efforts to apply genetic therapies that will optimize graft function and limit the complications of transplantation.

While it may be that the use of pigs as a source of organs and tissues for transplantation is not far off, exciting advances in tissue engineering, stem cell technology, and *in vitro* organogenesis may broaden the use of animals in human medicine. Adult and embryonic stem cell culture has given rise to organ-specific tissues with functional characteristics of the corresponding organs. Although these cultures are unlikely to yield fully developed functional organs for transplantation into humans, pigs or other animals could be used as recipients of these culture-initiated tissues and allow completion of development. These organs, grown and maintained in animals, may then be available for transplantation on an "as needed" basis. Pigs may thus serve as xenograft "recipients" prior to becoming organ "donors." The recent success in cloning of animals, including pigs, raises the possibility of transferring nuclei from a human patient's cells into enucleated stem cells of an animal, and then growing the cells in animals to generate differentiated human tissue that is autologous with the patient. The lessons learned from genetic manipulation of animals in the quest to make animal organs suitable for transplantation into humans may find their best application in generating animals suitable for use as biological reactors to grow human organs suitable for transplantation into humans.

See also Animal Biotechnology and Modeling; Immune Defence, Cell Mediated; Immunology; Medicinal Chemistry; Nuclear Transfer for Cloning Animals.

Bibliography

Books and Reviews

Cascalho, M., Platt, J.L. (2001) Xenotransplantation and other means of organ replacement, *Nat. Rev. Immunol.* **1**, 154–160.

Ogle, B.M., Platt, J.L. (2002) Approaches to the replacement of the function of failing organs, *Curr. Opin. Organ Transplant.* **7**(1), 28–34.

Platt, J.L. (2004) Future approaches to replacement of organs, *Am. J. Transplant.* **4**(Suppl. 6), 5–6.

Platt, J.L. (2001) Immunology of Xenotransplantation, in: Austen, K.F., Frank, M.M., Atkinson, J.P., Cantor, H. (Eds.) *Samter's Immunologic Diseases*, Lippincott Williams & Wilkins, Philadelphia, PA, pp. 1132–1146.

Primary Literature

2001 Annual Report of the U.S. Organ Procurement and Transplantation Network and the Scientific Registry for Transplant Recipients: Transplant Data 1991–2000. Department of Health and Human Services, Health Resources and Services Administration, Office of Special Programs, Division of Transplantation, Rockville, MD; United Network for Organ Sharing, Richmond, VA; University Renal Research and Education Association, Ann Arbor, MI. Retrieved December 9, 2002 from the World Wide Web: http://www.optn.org//data/annualReport.asp.

Akhter, S.A., Skaer, C.A., Kypson, A.P., McDonald, P.H., Peppel, K.C., Glower, D.D., Lefkowitz, R.J., Koch, W.J. (1997) Restoration of β-adrenergic signaling in failing cardiac ventricular myocytes via adenoviral-mediated

gene transfer, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 12100–12105.

Aksentijevich, I., Sachs, D.H., Sykes, M. (1992) Humoral tolerance in xenogeneic BMT recipients conditioned by a nonmyeloablative regimen, *Transplantation* **53**, 1108–1114.

Alvarado, C.G., Cotterell, A.H., McCurry, K.R., Collins, B.H., Magee, J.C., Berthold, J., Logan, J.S., Platt, J.L. (1995) Variation in the level of xenoantigen expression in porcine organs, *Transplantation* **59**, 1589–1596.

Bach, F.H., Winkler, H., Ferran, C., Hancock, W.W., Robson, S.C. (1996) Delayed xenograft rejection, *Immunol. Today* **17**, 379–384.

Betthauser, J., Forsberg, E., Augenstein, M., Childs, L., Eilertsen, K., Enos, J., Forsythe, T., Golueke, P., Jurgella, G., Koppang, R., Lesmeister, T., Mallon, K., Mell, G., Misica, P., Pace, M., Pfister-Genskow, M., Strelchenko, N., Voelker, G., Watt, S., Thompson, S., Bishop, M. (2000) Production of cloned pigs from in vitro systems, *Nat. Biotechnol.* **18**, 1055–1059.

Byrne, G.W., McCurry, K.R., Martin, M.J., McClellan, S.M., Platt, J.L., Logan, J.S. (1997) Transgenic pigs expressing human CD59 and decay-accelerating factor produce an intrinsic barrier to complement-mediated damage, *Transplantation* **63**, 149–155.

Calne, R.Y. (1970) Organ transplantation between widely disparate species, *Transplant. Proc.* **2**, 550–556.

Cascalho, M., Platt, J.L. (2001) The immunological barrier to xenotransplantation, *Immunity* **14**, 437–446.

Chari, R.S., Collins, B.H., Magee, J.C., Kirk, A.D., Harland, R.C., McCann, R.L., Platt, J.L., Meyers, W.C. (1994) Treatment of hepatic failure with ex vivo pig-liver perfusion followed by liver transplantation, *N. Engl. J. Med.* **331**(4), 234–237.

Chopek, M.W., Simmons, R.L., Platt, J.L. (1987) ABO-incompatible renal transplantation: initial immunopathologic evaluation, *Transplant. Proc.* **19**, 4553–4557.

Collins, B.H., Parker, W., Platt, J.L. (1994) Characterization of porcine endothelial cell determinants recognized by human natural antibodies, *Xenotransplantation* **1**, 36–46.

Cooper, D.K.C., Human, P.A., Lexer, G., Rose, A.G., Rees, J., Keraan, M., Du Toit, E. (1988) Effects of cyclosporine and antibody adsorption on pig cardiac xenograft survival in the baboon, *J. Heart Transplant.* **7**, 238–246.

Cotterell, A.H., Collins, B.H., Parker, W., Harland, R.C., Platt, J.L. (1995) The humoral immune response in humans following cross-perfusion of porcine organs, *Transplantation* **60**, 861–868.

Cozzi, E., Yannoutsos, N., Langford, G.A., Pino-Chavez, G., Wallwork, J., White, D.J.G. (1997) Effect of Transgenic Expression of Human Decay-Accelerating Factor on the Inhibition of Hyperacute Rejection of Pig Organs, in: Cooper, D.K.C., Kemp, E., Platt, J.L., White, D.J.G. (Eds.) *Xenotransplantation: the Transplantation of Organs and Tissues Between Species*, Springer, Berlin, Germany, pp. 665–682.

Daggett, C.W., Yeatman, M., Lodge, A.J., Chen, E.P., Gullotto, C., Frank, M.M., Platt, J.L., Davis, R.D. (1998) Total respiratory support from swine lungs in primate recipients, *J. Thorac. Cardiovasc. Surg.* **115**, 19–27.

Dai, Y., Vaught, T.D., Boone, J., Chen, S., Phelps, C.J., Ball, S., Monahan, J.A., Jobst, P.M., McCreath, K.J., Lamborn, A.E., Cowell-Lucero, J.L., Wells, K.D., Colman, A., Polejaeva, I.A., Ayares, D.L. (2002) Targeted disruption of the alpha1,3-galactosyltransferase gene in cloned pigs, *Nat. Biotechnol.* **20**(3), 251–255.

Dalmasso, A.P. (1992) The complement system in xenotransplantation, *Immunopharmacology* **24**, 149–160.

Dalmasso, A.P., Vercellotti, G.M., Platt, J.L., Bach, F.H. (1991) Inhibition of complement-mediated endothelial cell cytotoxicity by decay-accelerating factor: potential for prevention of xenograft hyperacute rejection, *Transplantation* **52**, 530–533.

Deacon, T., Schumacher, J., Dinsmore, J., Thomas, C., Palmer, P., Kott, S., Edge, A., Penney, D., Kassissieh, S., Dempsey, P., Isacson, O. (1997) Histological evidence of fetal pig neural cell survival after transplantation into a patient with Parkinson's disease, *Nat. Med.* **3**, 350–353.

Diamond, L.E., Quinn, C.M., Martin, M.J., Lawson, J., Platt, J.L., Logan, J.S. (2001) A human CD46 transgenic pig model system for the study of discordant xenotransplantation, *Transplantation* **71**, 132–142.

Dor, F.J., Tseng, Y.L., Cheng, J., Moran, K., Sanderson, T.M., Lancos, C.J., Shimizu, A.,

Yamada, K., Awwad, M., Sachs, D.H., Hawley, R.J., Schuurman, H.J., Cooper, D.K. (2004) alpha1,3-Galactosyltransferase gene-knockout miniature swine produce natural cytotoxic anti-Gal antibodies, *Transplantation* **78**(1), 15–20.

Evans, R.W. (2001) Coming to Terms with Reality: Why Xenotransplantation is a Necessity, in: Platt, J.L. (Ed.) *Xenotransplantation*, ASM Press, Washington, DC., pp. 29–51.

Good, A.H., Cooper, D.K.C., Malcolm, A.J., Ippolito, R.M., Koren, E., Neethling, F.A., Ye, Y., Zuhdi, N., Lamontagne, L.R. (1992) Identification of carbohydrate structures that bind human antiporcine antibodies: implications for discordant xenografting in humans, *Transplant. Proc.* **24**, 559–562.

Goodman, D.J., von Albertini, M., Willson, A., Millan, M.T., Bach, F.H. (1996) Direct activation of porcine endothelial cells by human natural killer cells, *Transplantation* **61**, 763–771.

Gritsch, H.A., Glaser, R.M., Emery, D.W., Lee, L.A., Smith, C.V., Sablinski, T., Arn, J.S., Sachs, D.H., Sykes, M. (1994) The importance of nonimmune factors in reconstitution by discordant xenogeneic hematopoietic cells, *Transplantation* **57**, 906–917.

Groth, C.G., Korsgren, O., Tibell, A., Tollemar, J., Moller, E., Bolinder, J., Ostman, J., Reinholt, F.P., Hellerstrom, C., Andersson, A. (1994) Transplantation of porcine fetal pancreas to diabetic patients, *Lancet* **344**(8934), 1402–1404.

Ihrcke, N.S., Wrenshall, L.E., Lindman, B.J., Platt, J.L. (1993) Role of heparan sulfate in immune system-blood vessel interactions, *Immunol. Today* **14**, 500–505.

Inverardi, L., Clissi, B., Stolzer, A.L., Bender, J.R., Pardi, R. (1996) Overlapping recognition of xenogeneic carbohydrate ligands by human natural killer lymphocytes and natural antibodies, *Transplant. Proc.* **28**, 552.

Kypson, A.P., Peppel, K., Akhter, S.A., Lilly, R.E., Glower, D.D., Lefkowitz, R.J., Koch, W.J. (1998) *Ex vivo* adenoviral-mediated gene transfer to the adult rat heart, *J. Thorac. Cardiovasc. Surg.* **115**(3), 623–630.

Lai, L., Kolber-Simonds, D., Park, K.W., Cheong, H.T., Greenstein, J.L., Im, G.S., Samuel, M., Bonk, A., Rieke, A., Day, B.N., Murphy, C.N., Carter, D.B., Hawley, R.J., Prather, R.S. (2002) Production of α-1,3-galactosyltransferase knockout pigs by nuclear transfer cloning, *Science* **295**(5557), 1089–1092.

Lanier, L.L., Phillips, J.H. (1996) Inhibitory MHC class I receptors on NK cells and T cells, *Immunol. Today* **17**, 86–91.

Lavitrano, M., Bacci, M.L., Forni, M., Lazzereschi, D., Di Stefano, C., Fioretti, D., Giancotti, P., Marfe, G., Pucci, L., Renzi, L., Wang, H., Stoppacciaro, A., Stassi, G., Sargiacomo, M., Sinibaldi, P., Turchi, V., Giovannoni, R., Della Casa, G., Seren, E., Rossi, G. (2002) Efficient production by sperm-mediated gene transfer of human decay-accelerating factor (hDAF) transgenic pigs for xenotransplantation, *Proc. Natl. Acad. Sci. U. S. A.* **99**(22), 14230–14235.

Lawson, J.H., Platt, J.L. (1996) Molecular barriers to xenotransplantation, *Transplantation* **62**, 303–310.

Lawson, J.H., Platt, J.L. (2000) Xenotransplantation: Prospects for Clinical Application, in: Owen, W., Pereira, B., Sayegh, M. (Eds.) *Dialysis and Transplantation: A Companion to Brenner and Rector's The Kidney*, W. B. Saunders Co., Philadelphia, PA, 653–660.

Lawson, J.H., Sorrell, R.D., Platt, J.L. (1997) Thrombomodulin activity in porcine-to-human xenotransplantation (abstract), *Circulation* **96**, 565.

Leventhal, J.R., Matas, A.J., Sun, L.H., Reif, S., Bolman, R.M., Dalmasso, A.P., Platt, J.L. (1993) The immunopathology of cardiac xenograft rejection in the guinea pig-to-rat model, *Transplantation* **56**, 1–8.

Li, H., Ricordi, C., Demetris, A.J., Kaufman, C.L., Korbanic, C., Hronakes, M.L., Ildstad, S.T. (1994) Mixed xenogeneic chimerism (mouse+rat→mouse) to induce donor-specific tolerance to sequential or simultaneous islet xenografts, *Transplantation* **57**, 592–598.

Lin, S.S., Platt, J.L. (1996) Immunologic barriers to xenotransplantation, *J. Heart Lung. Transplant.* **15**, 547–555.

Lin, S.S., Kooyman, D.L., Daniels, L.J., Daggett, C.W., Parker, W., Lawson, J.H., Hoopes, C.W., Gullotto, C., Li, L., Birch, P., Davis, R.D., Diamond, L.E., Logan, J.S., Platt, J.L. (1997) The role of natural anti-Galα1-3Gal antibodies in hyperacute rejection of pig-to-baboon cardiac xenotransplants, *Transpl. Immunol.* **5**, 212–218.

Lin, S.S., Weidner, B.C., Byrne, G.W., Diamond, L.E., Lawson, J.H., Hoopes, C.W.,

Daniels, L.J., Daggett, C.W., Parker, W., Harland, R.C., Davis, R.D., Bollinger, R.R., Logan, J.S., Platt, J.L. (1998) The role of antibodies in acute vascular rejection of pig-to-baboon cardiac transplants, *J. Clin. Invest.* **101**, 1745–1756.

Ljunggren, H-G., Karre, K. (1990) In search of the 'missing self': MHC molecules and NK cell recognition, *Immunol. Today* **11**, 237–244.

Magee, J.C., Collins, B.H., Harland, R.C., Lindman, B.J., Bollinger, R.R., Frank, M.M., Platt, J.L. (1995) Immunoglobulin prevents complement mediated hyperacute rejection in swine-to-primate xenotransplantation, *J. Clin. Invest.* **96**, 2404–2412.

Malyguine, A.M., Saadi, S., Platt, J.L., Dawson, J.R. (1996) Human natural killer cells induce morphologic changes in porcine endothelial cell monolayers, *Transplantation* **61**, 161–164.

Malyguine, A.M., Saadi, S., Holzknecht, R.A., Patte, C.R., Sud, N., Platt, J.L., Dawson, J.R. (1997) Induction of procoagulant function in porcine endothelial cells by human NK cells, *J. Immunol.* **159**, 4659–4664.

McCurry, K.R., Kooyman, D.L., Alvarado, C.G., Cotterell, A.H., Martin, M.J., Logan, J.S., Platt, J.L. (1995) Human complement regulatory proteins protect swine-to-primate cardiac xenografts from humoral injury, *Nat. Med.* **1**, 423–427.

McCurry, K.R., Parker, W., Cotterell, A.H., Weidner, B.C., Lin, S.S., Daniels, L.J., Holzknecht, Z.E., Byrne, G.W., Diamond, L.E., Logan, J.S., Platt, J.L. (1997) Humoral responses in pig-to-baboon cardiac transplantation: implications for the pathogenesis and treatment of acute vascular rejection and for accommodation, *Hum. Immunol.* **58**, 91–105.

McPaul, J.J., Stastny, P., Freeman, R.B. (1981) Specificities of antibodies eluted from human cadaveric renal allografts, *J. Clin. Invest.* **67**, 1405–1414.

Miyagawa, S., Hirose, H., Shirakura, R., Naka, Y., Nakata, S., Kawashima, Y., Seya, T., Matsumoto, M., Uenaka, A., Kitamura, H. (1988) The mechanism of discordant xenograft rejection, *Transplantation* **46**, 825–830.

Miyagawa, S., Murakami, H., Takahagi, Y., Nakai, R., Yamada, M., Murase, A., Koyota, S., Koma, M., Matsunami, K., Fukuta, D., Shirakura, R., Taniguchi, N. (2001) Remodeling of the major pig xenoantigen by *N*-acetylglucosaminyltransferase III in transgenic pig, *J. Biol. Chem.* **276**, 39310–39319.

Miyata, Y., Platt, J. (2003) Xeno: still stuck without α-Gal, *Nat. Biotechnol.* **21**(4), 359–360.

Nagata, H., Ito, M., Cai, J., Edge, A., Platt, J.L., Fox, I.J. (2003) Treatment of cirrhosis and liver failure in rats by hepatocyte xenotransplantation, *Gastroenterology* **124**(2), 422–431.

Ogle, B.M., Butters, K.B., Plummer, T.B., Ring, K.R., Knudsen, B., Litzow, M.R., Cascalho, M., Platt, J.L. (2004) Spontaneous fusion of cells between species yields transdifferentiation and retroviral in vivo, *FASEB J.* **18**, 548–550.

Onishi, A., Iwamoto, M., Akita, T., Mikawa, S., Takeda, K., Awata, T., Hanada, H., Perry, A.C.F. (2000) Pig cloning by microinjection of fetal fibroblast nuclei, *Science* **289**, 1188–1190.

Osman, N., McKenzie, I.F., Ostenried, K., Ioannou, Y.A., Desnick, R.J., Sandrin, M.S. (1997) Combined transgenic expression of alpha-galactosidase and alpha1,2-fucosyltransferase leads to optimal reduction in the major xenoepitope Galalpha(1,3)Gal, *Proc. Natl. Acad. Sci. U. S. A.* **94**(26), 14677–14682.

Parker, W., Lin, S.S., Platt, J.L. (2001) Antigen expression in xenotransplantation: how low must it go? *Transplantation* **71**, 313–319.

Parker, W., Bruno, D., Holzknecht, Z.E., Platt, J.L. (1994) Characterization and affinity isolation of xenoreactive human natural antibodies, *J. Immunol.* **153**, 3791–3803.

Parker, W., Saadi, S., Lin, S.S., Holzknecht, Z.E., Bustos, M., Platt, J.L. (1996) Transplantation of discordant xenografts: a challenge revisited, *Immunol. Today* **17**, 373–378.

Patience, C., Takeuchi, Y., Weiss, R.A. (1997) Infection of human cells by an endogenous retrovirus of pigs, *Nat. Med.* **3**, 282–286.

Paul, L.C., Claas, F.H.J., van Es, L.A., Kalff, M.W., de Graeff, J. (1979) Accelerated rejection of a renal allograft associated with pretransplantation antibodies directed against donor antigens on endothelium and monocytes, *N. Engl. J. Med.* **300**, 1258–1260.

Perper, R.J., Najarian, J.S. (1967) Experimental renal heterotransplantation. III. Passive transfer of transplantation immunity, *Transplantation* **5**, 514–533.

Phelps, C.J., Koike, C., Vaught, T.D., Boone, J., Wells, K.D., Chen, S.H., Ball, S., Specht, S.M.,

Polejaeva, I.A., Monahan, J.A., Jobst, P.M., Sharma, S.B., Lamborn, A.E., Garst, A.S., Moore, M., Demetris, A.J., Rudert, W.A., Bottino, R., Bertera, S., Trucco, M., Starzl, T.E., Dai, Y., Ayares, D.L. (2003) Production of alpha 1,3-galactosyltransferase-deficient pigs, *Science* **299**(5605), 411–414.

Platt, J.L. (2000) Hyperacute rejection: fact or fancy, *Transplantation* **69**(6), 1034–1035.

Platt, J.L., Logan, J.S. (1996) Use of transgenic animals in xenotransplantation, *Transplant. Rev.* **10**, 69–77.

Platt, J.L., Fischel, R.J., Matas, A.J., Reif, S.A., Bolman, R.M., Bach, F.H. (1991) Immunopathology of hyperacute xenograft rejection in a swine-to-primate model, *Transplantation* **52**, 214–220.

Platt, J.L., Turman, M.A., Noreen, H.J., Fischel, R.J., Bolman, R.M., Bach, F.H. (1990) An ELISA assay for xenoreactive natural antibodies, *Transplantation* **49**, 1000–1001.

Platt, J.L., Vercellotti, G.M., Dalmasso, A.P., Matas, A.J., Bolman, R.M., Najarian, J.S., Bach, F.H. (1990) Transplantation of discordant xenografts: a review of progress, *Immunol. Today* **11**, 450–456.

Polejaeva, I.A., Chen, S., Vaught, T.D., Page, R.L., Mullins, J., Ball, S., Dai, Y., Boone, J., Walker, S., Ayares, D.L., Colman, A., Campbell, K.H.S. (2000) Cloned piglets produced by nuclear transfer from adult somatic cells, *Nature* **407**, 86–90.

Porter, K.A. (1992) Renal Transplantation, in: Heptinstall, R.H. (Ed.) *Pathology of the Kidney*, Vol. 3, Little Brown and Company, Boston, MA, pp. 1799–1933.

Sachs, D.H., Sablinski, T. (1995) Tolerance across discordant xenogeneic barriers, *Xenotransplantation* **2**, 234–239.

Sandrin, M.S., Vaughan, H.A., Dabkowski, P.L., McKenzie, I.F.C. (1993) Anti-pig IgM antibodies in human serum react predominantly with Galα(1,3)Gal epitopes, *Proc. Natl. Acad. Sci. U. S. A.* **90**, 11391–11395.

Sharma, A., Okabe, J.F., Birch, P., Platt, J.L., Logan, J.S. (1996) Reduction in the level of Gal (α1,3) Gal in transgenic mice and pigs by the expression of an α(1,2) fucosyltransferase, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 7190–7195.

Sharma, A., Naziruddin, B., Cui, C., Martin, M.J., Xu, H., Wan, H., Lei, Y., Harrison, C., Yin, J., Okabe, J., Mathews, C., Stark, A., Adams, C.S., Houtz, J., Wiseman, B.S., Byrne, G.W., Logan, J.S. (2003) Pig cells that lack the gene for alpha1-3 galactosyltransferase express low levels of the gal antigen, *Transplantation* **75**(4), 430–436.

Starzl, T.E., Demetris, A.J., Murase, N., Thomson, A.W., Trucco, M., Ricordi, C. (1993) Donor cell chimerism permitted by immunosuppressive drugs: a new view of organ transplantation, *Immunol. Today* **14**, 326–332.

Sullenger, B.A., Gallardo, H.F., Ungers, G.E., Gilboa, E. (1990) Overexpression of TAR sequences renders cells resistant to human immunodeficiency virus replication, *Cell* **63**, 601–608.

Sykes, M., Lee, L.A., Sachs, D.H. (1994) Xenograft tolerance, *Immunol. Rev.* **141**, 245–276.

Thall, A.D., Malý, P., Lowe, J.B. (1995) Oocyte Galα1,3Gal epitopes implicated in sperm adhesion to the zona pellucida glycoprotein ZP3 are not required for fertilization in the mouse, *J. Biol. Chem.* **270**(37), 21437–21440.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., Jones, J.M. (1998) Embryonic stem cell lines derived from human blastocysts, *Science* **282**(5391), 1145–1147.

White, D., Wallwork, J. (1993) Xenografting: probability, possibility, or pipe dream? *Lancet* **342**, 879–880.

Zhao, Y., Swenson, K., Sergio, J.J., Arn, J.S., Sachs, D.H., Sykes, M. (1996) Skin graft tolerance across a discordant xenogeneic barrier, *Nat. Med.* **2**(11), 1211–1216.

X-Ray Diffraction of Biomolecules: *see* Macromolecules, X-Ray Diffraction of Biological

Yeast, DNA Repair in: *see* DNA Repair in Yeast

Zebrafish (*Danio rerio*) Genome and Genetics

Ralf Dahm, Robert Geisler, and Christiane Nüsslein-Volhard
Max Planck Institute for Developmental Biology, Tübingen, Germany

1 The Zebrafish as a Model Organism for Developmental Biology 596
1.1 General Remarks 596
1.2 Advantages of the Zebrafish as a Model Organism 596
1.3 Zebrafish Development 596
1.4 Phylogeny 599
1.5 Laboratory Strains 599

2 Genetic Tools and Methods Available to Study Development in the Zebrafish 602
2.1 General Remarks 602
2.2 Genetic Screens (Forward Genetics) 602
2.2.1 Chemical Mutagenesis 602
2.2.2 Insertional Mutagenesis 604
2.2.3 Radiation Mutagenesis 607
2.3 Reverse Genetics 607
2.3.1 Morpholino Knockdown 608
2.3.2 TILLING 610
2.3.3 RNA Interference 612
2.4 Parthenogenesis 612
2.5 Transgenesis 616
2.6 Microarrays 617

3 Mapping of the Zebrafish Genome 617
3.1 Genetic Mapping 617
3.2 Radiation Hybrid Mapping 619
3.3 Physical Mapping 620
3.4 Genome Sequencing 621

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Volume 15
Edited by Robert A. Meyers.
Copyright © 2005 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30652-8

4 Resources on the Internet 622
4.1 General 622
4.2 Mapping 622
4.3 Sequencing 622
4.4 Resource Centers 622

Bibliography 622
Books and Reviews 622
Primary Literature 623

Keywords

Bacterial Artificial Chromosomes (BACs)
Genomic clones with 100 to 200 kb insert size used for physical mapping and genome sequencing.

Expressed Sequence Tags (ESTs)
Single-pass sequences of cDNA clones that provide rapid access to genes as well as information on expression and polymorphism.

Forward Genetics
Methods that identify an unknown gene by its mutant phenotype, usually by mutagenesis screening with a chemical mutagen or retrovirus, and cloning the affected locus.

Knockdown
Abolishment of activity of a known gene by injection of morpholino oligonucleotides complementary to its transcript into the fertilized egg, mainly feasible for genes required early in development.

Microsatellite Markers
Genetic markers consisting of a sequence of short repeats of variable length (in particular $(CA)_n$) flanked by a pair of primer sequences, useful for genetic mapping because they are widespread, codominant, and easily scored by PCR amplification and gel electrophoresis.

Positional Cloning
Molecular cloning of mutant loci for which no obvious candidate gene exists, by genetically mapping the mutation to a specific contig or genomic clone, and sequencing the transcripts of this critical region.

Radiation Hybrid Mapping
Mapping method for sequence-tagged site (STS) markers, based on fusion cell lines generated from irradiated cells of the species of interest and rodent cells. If two markers are found by PCR to be present or absent in the same radiation hybrid cell lines, they are likely to be close to each other.

Restriction Fingerprinting
Physical mapping method in which genomic clones are digested with a restriction enzyme, overlaps between the clones are identified by their similar restriction patterns, and contigs are constructed from the overlapping clones.

Reverse Genetics
Methods that identify the mutant genotype of a known gene for further investigation of its function, for example by knockdown or TILLING.

Single Nucleotide Polymorphisms (SNPs)
Changes of a single nucleotide between strains, useful for genetic mapping, for example by sequencing.

Targeting Induced Local Lesions In Genomes (TILLING)
Identification of mutations in a known gene by screening of a library of mutagenized animals, either by sequencing or with CEL-I heteronuclease; also referred to as targeted knockout.

Teleost Genome Duplication
Hypothesized genome duplication (tetraploidization) event in the evolution of ray-finned fish, followed by functional specialization or loss of duplicate genes; as a result, some mammalian genes have two zebrafish orthologs with distinct functions and expression domains.

Owing to its transparent, easily accessible embryos, simple breeding, and short generation time, the zebrafish has become one of the most important model organisms to study the genetic control of embryonic development. Zebrafish research initially focused on forward genetics (mutagenesis screening), but reverse genetic methods, transgenesis, and microarrays are now equally available to characterize known genes. The zebrafish genome consists of 25 chromosome pairs and has an estimated haploid size of 1.5 Gb. Several genome maps exist for the zebrafish. The meiotic MGH map is generally used for placement of mutant loci and, in combination with the T51 radiation hybrid map and BAC fingerprinting, forms the basis of genome sequencing. A genome sequence is being generated by a combination of whole-genome shotgun and BAC-based sequencing, and is expected to be finished in 2005.

1 The Zebrafish as a Model Organism for Developmental Biology

1.1 General Remarks

In recent years, the zebrafish has become a favorite model organism for biologists studying developmental as well as cell biological processes. It is uniquely suited to identifying novel regulatory genes in vertebrate development and elucidating their function in the developing animal. No other vertebrate model organism offers the same combination of transparent and easily accessible embryos, efficient mutagenesis screening that has led to several hundred characterized mutations, and more recently, a sequenced genome as well as microarray and knockout technologies.

In addition, the zebrafish is of interest to the pharmaceutical industry because the embryos can be used for screening small molecule libraries in high-throughput pipelines. This approach can be applied to toxicology and, with mutants that are models for human diseases, to drug discovery and development.

The potential of the zebrafish system has been only partially realized so far. A main reason is that prior to genome sequencing, cloning of zebrafish mutations relied mainly on candidate genes and was therefore biased to finding mutations in genes already known from other organisms. However, this is changing rapidly, and novel regulatory genes are increasingly being discovered and characterized in the zebrafish.

1.2 Advantages of the Zebrafish as a Model Organism

The zebrafish has many properties that make it ideally suited as a model organism for experimental biologists:

- Zebrafish are easy to maintain and breed, minimizing the cost and effort to keep large populations of animals.
- They develop externally (outside of the female), allowing easy access to the embryos at all stages of development.
- The embryos are sturdy and large enough for experimental manipulations, such as microinjections or transplantation of cells. However, the embryos are small enough to fit into standard 96-well plates, facilitating automated analyses.
- Sperm samples can easily be obtained (even from live fish), stored, and later used for *in vitro* fertilizations.
- Embryonic development of the zebrafish is synchronous, that is, at any given time, all the embryos in a clutch will be at the same developmental stage.
- Compared with other vertebrate models for development, such as the mouse or *Xenopus*, the organs in zebrafish larvae are composed of fewer cells, making them easier to study. However, the organs still function in much the same way as those in larger animals.
- Zebrafish have large numbers of offspring.

1.3 Zebrafish Development

In addition to these advantages, there are two main embryological advantages that set the zebrafish apart from other vertebrate model organisms: its very rapid early development and the fact that its development can easily be observed in great detail.

For a vertebrate, embryonic development in the zebrafish is extremely rapid (see Fig. 1). Embryogenesis takes approximately 2 days, the larvae hatch on the

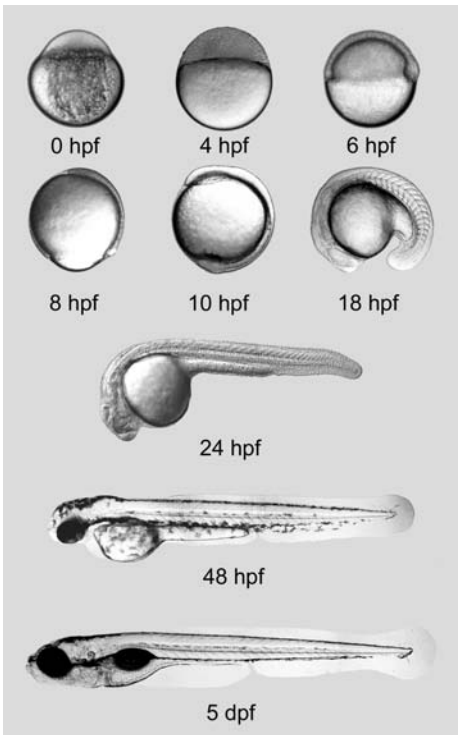

Fig. 1 A photographic overview of key aspects of early zebrafish development. hpf: hours postfertilization; dpf: days postfertilization. 0 hpf: At the one-cell stage immediately after fertilization, the zygote sits on top of a large ball of yolk. 4 hpf: Four hours following fertilization, the cells have undergone several rounds of cleavages to form a mass of cells located on top of the yolk. Please note that in contrast to species such as *Xenopus*, the yolk cell does not undergo cleavage divisions. 6 hpf: Gastrulation begins with epiboly of the cells. At this stage, the antero-posterior and dorso-ventral axes can be determined. 8 hpf: Onset of thickening in the region of the future head. 10 hpf: First somite forms (right) and the optic lobes evaginate from the diencephalon. 18 hpf: Body plan becomes recognizable, and the first voluntary muscle contractions are observable as movements of the tail. 24 hpf: Most vertebrate-specific body features and organs are already recognizable, including a compartmentalized brain, eyes, ears, and all internal organs. However, the majority of the organs are not yet functional. Notably though, the heart starts beating just before the end of the first day of development. Over the next 24 hpf, numerous cell types differentiate, and the organs gradually begin to function. 48 hpf: Zebrafish start to hatch and start swimming. 5 dpf: Larvae swim well and search for food independently. The speed of zebrafish development is temperature dependent. Zebrafish develop at temperatures ranging from approximately 18 to 32 °C. The higher the temperature, the faster the development. Normally, zebrafish are kept at temperatures ranging between 25 and 28.5 °C. In this text, all statements relating to the development of any given structure at any time point always refer to development having occurred at 28.5 °C. (Adapted from Haffter, P., Granato, M., Brand, M., Mullins, M.C., Hammerschmidt, M., Kane, D.A., Odenthal, J., van Eeden, F.J., Jiang, Y.J., Heisenberg, C.P., Kelsh, R.N., Furutani-Seiki, M., Vogelsang, E., Beuchle, D., Schach, U., Fabian, C., Nusslein-Volhard, C. (1996) The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio, Development* **123**, 1–36.)

third day of development and as of day five, the vast majority of cell types have differentiated and the organs have taken up their function. This rapid development allows the observation of developmental processes and the completion of experiments generally within a few hours to days.

Despite the rapidity of embryonic and early larval development in zebrafish, the time needed to reach sexual maturity (generation time) is between two and four months (depending on the conditions under which they are raised; see Fig. 2).

The development of the zebrafish is similar to the embryogenesis in higher-ordered vertebrates, including humans and other mammals. But unlike in mammals, the embryonic development of zebrafish occurs outside the female. This allows easy access to the embryos without the need to sacrifice the female. Moreover, unlike in birds and reptiles, the shell of the zebrafish eggs (the chorion) is transparent, allowing the continuous observation of the developing embryo in its "natural" environment without the need to open the

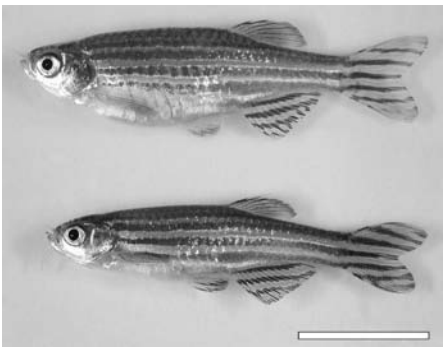

Fig. 2 Photographs of age-matched adult female (top) and male (bottom) zebrafish. The characteristic horizontal stripes running along the body and fins gave this species its name. Adult zebrafish measure about 3 to 5 cm in length. Females can be recognized by their larger belly accommodating the eggs. Males tend to be more reddish in color than female zebrafish. The scale bar is 1 cm. (See color plate p. xxix).

shell. This property is shared with, for example, the African claw-toed frog *Xenopus*, another model for early development.

The most important advantage of the zebrafish, however, is that the embryos themselves are completely transparent during early development. This is due both to their relatively small size and the absence of pigment throughout the first 36 h of development (the time when pigmentation in the melanocytes and the cells of the retinal pigment epithelium begins). Yet, even after the onset of pigmentation, most parts of the body can still be observed between the relatively few melanophores. Also, in contrast to *Xenopus*, in zebrafish, the yolk does not participate in the cleavage divisions of the early embryo but remains distinct from the embryo throughout development.

This transparency of the zebrafish embryos and larvae allows the visualization of many processes that occur deep inside the developing animal. These include, the development of internal organs, such as the brain, the cardiovascular system or the digestive tract, the migration of cells as well as the path finding of axons. In the zebrafish, these developmental processes can be visualized in real time without the need for invasive intervention (e.g. dissection or histological sectioning) or sophisticated imaging techniques (e.g. MRI).

The external development and the transparency of the zebrafish embryos facilitate numerous experimental approaches. These include, for example, laser manipulations, such as cell ablation or uncaging experiments as well as the generation of extensive fate maps. Moreover, in techniques, such as mRNA *in situ* hybridization and antibody staining to test for the expression patterns of genes and proteins, the translucency of the embryos in many cases obliterates the need for sectioning as the staining can be performed and imaged on whole embyros.

In contrast to mammals, zebrafish can generate large numbers of offspring. When kept under optimal conditions, a single female can lay up to 200 eggs per week. This greatly facilitates high-throughput approaches and statistical analyses, including the reliable identification of mutant phenotypes, for example in a genetic screen. This fact, together with the small size and easy maintenance of zebrafish embryos as well as their rapid embryonic development, allows for forward genetic approaches to be taken that are often too labor intensive and too costly to be carried out with other vertebrate models, such as the mouse. The zebrafish thus combines being a vertebrate with many of the advantages of invertebrate model organisms.

For a summary of key properties of commonly used animal model organisms, please refer to Table 1.

1.4 Phylogeny

The zebrafish (*Danio rerio*, formerly also known as *Brachydanio rerio*) is a small tropical freshwater fish. Its natural habitats are rivers of South Asia, mainly northern India as well as northern Pakistan, Nepal, and Bhutan. The zebrafish belongs to the family of the cyprinids (*Cyprinidae*) in the class of ray-finned fishes (*Actinopterygii*) and within this class to the bony fishes (teleosts or *Teleostei*) to which most extant ray-finned fishes belong (Fig. 3). The teleost fishes comprise other model organisms:

- Medaka (*Oryzias latipes*), a species, which is mainly used to study developmental biology.
- The three-spined stickleback (*Gasterosteus aculeatus*) and the guppy (*Poecilia reticulata*), the two species with a long tradition in behavioral studies and more recently, in population genetics also.
- The two pufferfish species *Takifugu rubripes* (also referred to as *Fugu rubripes*) and *Tetraodon nigroviridis* (recently also referred to as *Monotreta nigroviridis*), which with their relatively small and simple genomes are important organisms to understand genome architecture, organization, and function. *Takifugu rubripes* was the first fish and second vertebrate species whose genome was fully sequenced (in 2002). The genome of *T. nigroviridis* is currently also being sequenced.

These five species belong to the teleost crown group, a large radiation whose last common ancestor with the zebrafish lived about 100 million years ago (a similar evolutionary time span as that between humans and mice). The lineages leading to the cyprinids and mammals split about 400 to 450 million years ago.

It is assumed that during the evolution of the ray-finned fishes, there has been a genome duplication event, which lead to a tetraploidization of the genome. This duplication was followed by a functional specialization of some of the duplicated genes on the one hand and the loss of other genes on the other hand. As a result, an estimated 20% of mammalian genes have two zebrafish orthologs with distinct functions and expression domains. This subfunction partitioning can help to understand different aspects of the mammalian gene function as they can be studied in isolation.

1.5 Laboratory Strains

Several wild-type laboratory strains of the zebrafish have been established by the major research laboratories. While these strains are inbred to some extent, none of them is completely isogenic, and all of them have somewhat overlapping distributions of marker alleles. The choice of a mapping strain is thus often a matter of personal preference. Only the most common strains will be mentioned here.

The primary strains used for mutagenesis experiments are AB and Tübingen (Tü). The AB strain was generated by G. Streisinger in Eugene, Oregon, by crossing the strains obtained at a pet shop. It has been used for a major mutagenesis screen by the laboratory of W. Driever, and for several smaller

Tab. 1 Comparison of key properties of commonly used model organisms.

Organism	*Size egg (mm)*	*Length adult*	*Number of offspring*	*Duration of embryogenesis*	*Generation time*	*Genome size [bp]*	*Number of genes*
Caenorhabditis elegans	0.05	1 mm	300–350/life time	20 h	50 h	0.097×10^9	20 000
Drosophila melanogaster	0.5	3 mm	50/day	24 h	9 days	0.18×10^9	13 600
Zebrafish (*Danio rerio*)	0.7	5 cm	200/week	2 days	2–3 months	1.5×10^9	est. 30 000
Xenopus laevis	1–1.3	10–14 cm	300–1000/2–3 months	2–2.5 days	1–2 years	3.1×10^9	Unknown
Xenopus tropicalis	0.6–0.8	4–5 cm	1000–3000/2–3 months	2 days	3–4 months	1.7×10^9	Unknown
Chicken (*Gallus gallus*)	60–80	30 cm	1/1–2 days	6 days	9 months	1.1×10^9	Unknown
Mouse (*Mus musculus*)	0.1	6–10 cm	4–8/4–6 weeks	14 days	2 months	2.6×10^9	30 000

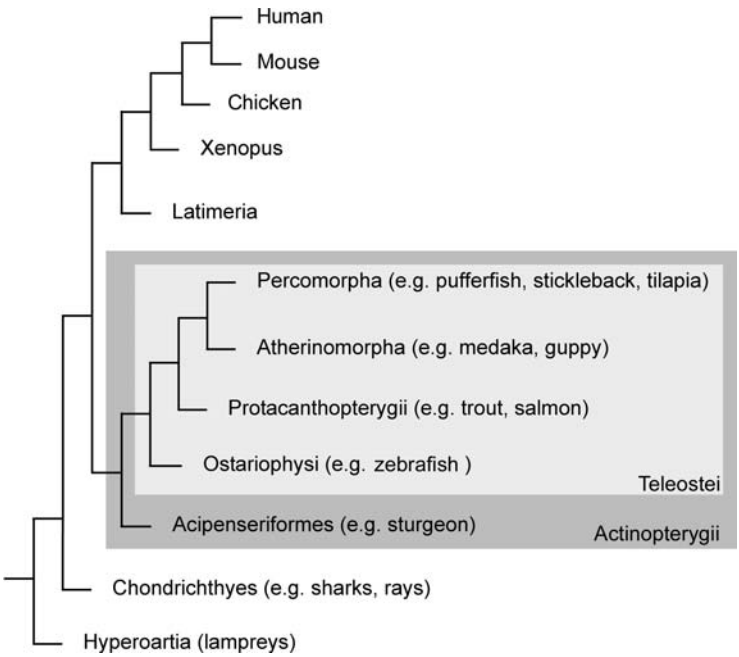

Fig. 3 Schematic representation of the major vertebrate lineages.

screens. The Tübingen strain, likewise originating from a pet shop, has been used as the genetic background of most mutagenesis screens including the two largest ones, performed in the laboratory of C. Nüsslein-Volhard. For this reason, it was also chosen as the strain to be sequenced in the zebrafish genome project.

Wild-type India Calcutta (WIK) was generated from a wild catch in India as a highly polymorphic reference strain for mapcrosses with mutations in the Tü background. In crosses of Tü x WIK, 68% of microsatellite markers give a scorable polymorphism. India (IN) was used for the MGH meiotic mapping panel in combination with AB; however, it is more difficult to breed than WIK and may contain mutations that affect viability. The strain SJD was isogenized over several generations by the laboratory of S. L. Johnson, and it is more polymorphic in crosses to Tü (75% scorable polymorphism). It would thus be better suited as a reference strain, but it is more difficult to breed because of a skewed sex ratio (a common problem with inbred zebrafish strains).

In the zebrafish, visible markers are not as useful for the maintenance of mutations as, for example, in *Drosophila* because it has 25 chromosomes and lacks balancer chromosomes (which would prevent recombination between the marker and the mutant locus). A common strain with visible markers is Tübingen long fin (TL, formerly named Tup long fin), which carries a recessive mutation (leo^{t1}) that causes spotting, and a dominant mutation (lof^{dt2}) causing long fins.

Wild-type strains can be obtained either from the Zebrafish International Resource Center (ZIRC) or from the laboratories in which they originated.

2 Genetic Tools and Methods Available to Study Development in the Zebrafish

2.1 General Remarks

When compared to other model organisms, such as the nematode *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster*, or the mouse *Mus musculus*, the number of tools and methods that are available when working with the zebrafish is still relatively limited. This is not surprising when considering that the zebrafish is a comparatively recent addition to the list of model organisms. It was established largely by George Streisinger in the late 1970s and early 1980s. Since then, the zebrafish community has grown steadily, the tools and methods available to zebrafish researchers have increased considerably over the past two decades, and there is now a substantial number of protocols, reagents, as well as transgenic and mutant zebrafish lines.

2.2 Genetic Screens (Forward Genetics)

To date, genetic screens are the most successful approach for identifying genes with an essential function in the development of an organism. Genetic screens rely on the random induction of mutations in the genome and the subsequent search for individuals with a mutant phenotype. They have the advantage that they are relatively unbiased with respect to prior knowledge, and thus also allow the identification of previously unknown genes. However, being able to identify mutants for a given process, morphological structure or behavior relies on the capability to develop an assay for the detection of the corresponding phenotypes. Furthermore, as genetic screening relies on the identification of a phenotype, it selects genes with unique or at least partially nonredundant functions. Genes with redundant functions, however, cannot be detected in genetic screens.

The possibility to perform large-scale genetic screens in the zebrafish is a powerful feature of this model organism. When compared to other vertebrate model organisms (mouse, chicken, frog), the generation time of the zebrafish is similar (2–4 months). But the large number of offspring greatly facilitates phenotypic and genetic analyses, such as the collecting of sufficient numbers of mutant individuals for different assays and phenotypic analyses as well as of recombinant individuals for the subsequent mapping of mutations (see below). Moreover, the moderate space required to breed large numbers of zebrafish significantly reduces the cost of such undertakings. Finally, the speed of zebrafish development and the ease with which it can be observed greatly facilitate the identification of mutant phenotypes for virtually any organ, tissue, or cell type.

Several mutagenesis protocols have been established for the zebrafish. The mutagens used include ENU, radiation (mainly γ-rays), as well as insertional mutagenesis with retroviruses or transposable elements.

2.2.1 Chemical Mutagenesis

The generation of mutants by chemical mutagenesis has been one of the most successful approaches in elucidating the function of genes and the genetic control of biological processes, particularly those of embryonic development. One of the most widely used substances to induce mutations is the chemical *N*-ethyl-*N*-nitrosourea (ENU). ENU is a potent mutagen, which has been used to induce mutations in various organisms. For

example, in the mouse, it has the highest mutation rate of all germ cell mutagens examined, and also for the zebrafish, ENU has been the mutagen of choice in the majority of genetic screens carried out to date.

ENU is a synthetic alkylating agent, which induces mutations by transferring its ethyl group to oxygen or nitrogen radicals in individual bases of DNA, including the N1, N3, and N7 groups of adenine, the O2 and N3 groups of cytosine, the O6, N3, and N7 groups of guanine, and the O2, O4, and N3 groups of thymine. Generally, ENU induces point mutations: The ethylation of a base can lead to the modified base being misread during DNA replication, resulting in the integration of a noncomplementary base in the second DNA strand (see Fig. 4). Following a second round of replication, one of the four double helices derived from the initially modified double helix carries a complete base-pair substitution. ENU-induced mutations chiefly involve the base thymidine (AT to TA transversions and AT to GC transitions as shown in Fig. 3 are the most common mutations induced by ENU). However, chromosomal rearrangements (translocations, inversions, deletions) have also been described for post-meiotically mutagenized male germ cells.

When animals are exposed to ENU (in the case of zebrafish by bathing them in a solution containing ENU), it induces mutations in all cells of the body, including in the precursors of the germ cells. It has been shown for mice that the rate of mutagenesis by ENU in testes is highest in spermatogonial stem cells

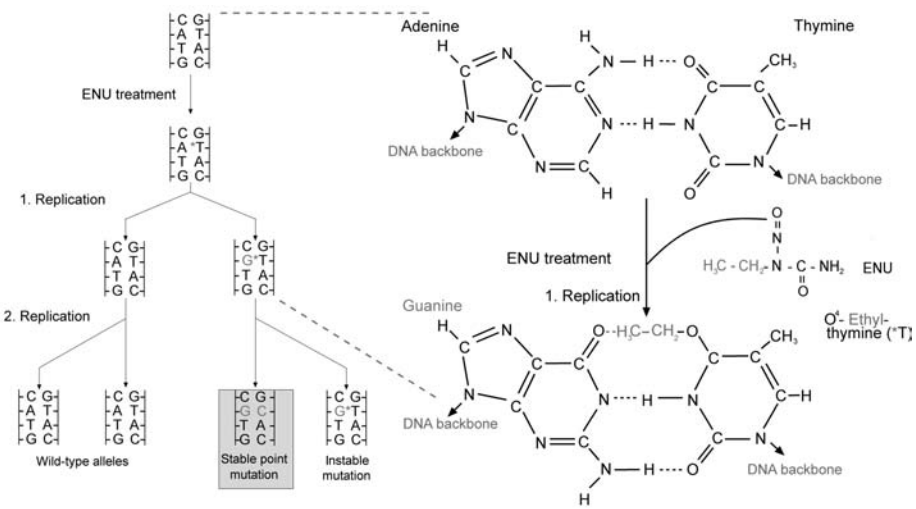

Fig. 4 Mechanism of mutagenesis by ENU leading to an AT → GC transition. The ethyl group in ENU (shown in red or indicated as a red asterisk) is transferred to the O^4-atom of thymine, resulting in an O^4-ethylthymine adduct. Failing an excision of the O^4-ethylthymine adduct by the DNA repair machinery, the O^4-ethylthymine mispairs with a guanine in a subsequent replication of the DNA. In this mispairing, the ENU-derived ethyl group forms a third hydrogen bond with the O^6 of the guanine base in the complementary DNA strand. A successive second replication leads to a complete AT → GC transition in one of the four double helices produced, thus fixing the mutation. Substitutions of bases are indicated in blue. Figure adapted after Noveroske, J.K., Weber, J.S., Justice, M.J. (2000) The mutagenic action of *N*-ethyl-*N*-nitrosourea in the mouse, *Mamm. Genome* **11**, 478–483. (See color plate p. xxx).

where optimal doses of ENU result in one mutation per gene in every 175 to 655 gametes (corresponding to a specific locus rate of $1.5–5.7 \times 10^{-3}$). In the zebrafish, the specific locus rate ranges from 0.6 to 4.8×10^{-3}, indicating that either the mutagenesis rates that can be achieved are slightly lower than those for the mouse or that the optimum conditions for mutagenesis in the zebrafish have not yet been found.

As the mutations induced by ENU primarily occur only in one of the DNA strands (see Fig. 3), there is a difference in the offspring resulting from mutagenized germ cells, depending on whether the exposure to ENU occurred before or after meiosis: mutations in premeiotic spermatogonia lead to sperm, which carry the mutation in both strands (due to DNA replication during meiosis), whereas mutations in postmeiotic germ cells result in sperm with only one of the DNA strands mutagenized. As a consequence, animals arising from spermatogonia carry the mutation in every cell. However, the progeny from germ cells that were mutagenized postmeiotically are mosaic with respect to the mutation.

Figure 5 depicts a typical breeding scheme for a two generation ENU mutagenesis screen in the zebrafish. Genetic screens using the zebrafish have resulted in the isolation of nearly 2000 mutant lines up until today. The vast majority of these mutations are recessive and affect embryonic or larval development (most are embryonic or larval lethal). However, a few dominant mutations as well as some mutations with adult phenotypes have been isolated. For a list of published mutants with brief phenotypic and – where available – molecular characterizations, refer to the searchable database at http://www.zfin.org.

2.2.2 Insertional Mutagenesis

The most significant disadvantage of working with mutants generated by chemical mutagenesis is the difficulty in identifying the mutation. The localization of chemically mutated genes in the genome requires a laborious genetic mapping of the mutated locus followed by positional cloning and/or a candidate gene approach (see below). This drawback can be circumvented if the mutagenesis is achieved by the insertion of a known sequence (e.g. a transposon or a retroviral vector), which can be used as a starting point for the sequencing of the regions flanking the insert. This strategy has been successfully applied in mutagenesis screens in other model organisms. Famously, the P-element transposon has been used to generate numerous mutants in *Drosophila*, and retroviral vectors were employed to mutate, for example, mice.

In the zebrafish, insertional mutants have been generated using one of three methods: injection of plasmid DNA, transposons, and pseudotyped retroviral vectors. However, the only efficient protocols for insertional mutagenesis in the zebrafish have been established using retroviruses.

Briefly, retroviral vectors are injected into zebrafish embryos at the blastula stage, resulting in mosaic founder fish carrying insertions also in the precursors of the germ cells (up to 25–30 insertions in the germ line per individual with a given insertion occurring in 3–20% of the gametes). Incrossing of these founders generates a nonmosaic F_1 generation. Typically, approximately 20% of these fish have 5 to 10 insertions per individual. The F_1 fish carrying multiple insertions are incrossed to raise F_2 families with at least 10 independently segregating inserts,

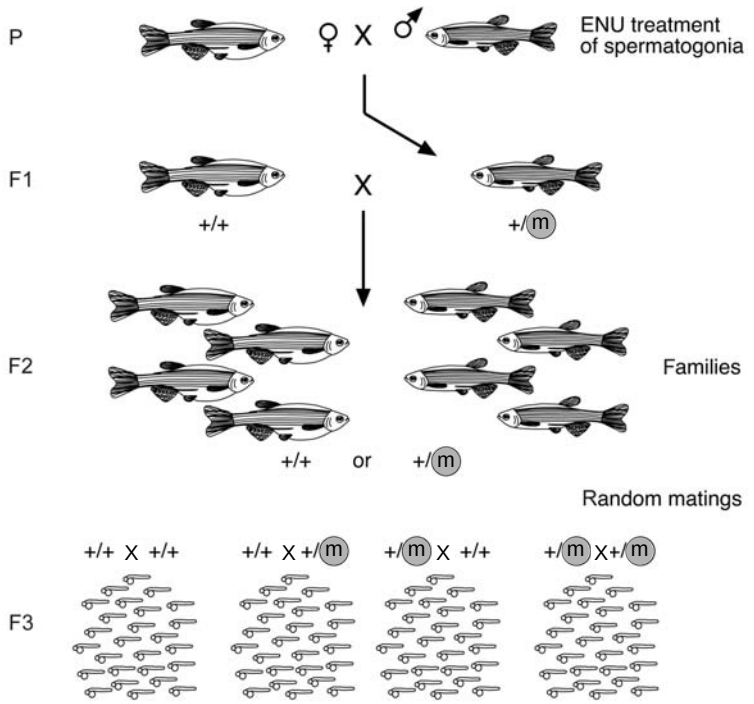

Fig. 5 Depiction of a mutagenesis and crossing scheme. Male fish are mutagenized with ENU and subsequently mated with wild-type female fish. The fish of the resulting F_1 generation are heterozygous for individual mutations (depicted as "m" in this example). These F_1 fish are again mated with wild-type fish in single-pair matings. The offspring from each of these matings are commonly referred to as "F_2 families." In these F_2 families, approx. 50% of the individuals are heterozygous for a particular mutation (+/m); the other 50% carry two wild-type alleles of this locus (+/+). Random single-pair matings between the siblings of an F_2 family give rise to the F_3 generation. This is the first generation in which a particular mutation can occur in a homozygous state and which hence first show a mutant phenotype for recessive mutations (the vast majority of ENU-induced mutations are recessive; rarely dominant phenotypes can already be observed in the F_1 and F_2 generations). On the level of the mutated locus, four different cases can be distinguished: Two homozygous F_2 individuals are mated (+/+ X +/+), a genotypically wild-type F_2 female is mated with a heterozygous F_2 male (+/+ X +/m), a heterozygous F_2 female is mated with a genotypically wild-type F_2 male (+/m X +/+), and two heterozygous F_2 individuals are mated (+/m X +/m). In the case of a recessive mutation, the latter cross results in phenotypically mutant F_3 offspring (approximately 25% of the individuals). (Figure adapted from Haffter, P., Granato, M., Brand, M., Mullins, M.C., Hammerschmidt, M., Kane, D.A., Odenthal, J., van Eeden, F.J., Jiang, Y.J., Heisenberg, C.P., Kelsh, R.N., Furutani-Seiki, M., Vogelsang, E., Beuchle, D., Schach, U., Fabian, C., Nusslein-Volhard, C. (1996) The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio, Development* **123**, 1–36.)

which in turn are incrossed and their offspring (F_3) screened for phenotypes.

In a zebrafish mutagenesis screen using a pseudotyped retroviral vector, approximately one in every 80 insertions results in an embryonic/early larval phenotype that can be identified on the basis of morphologically observable criteria. With an average number of 12 insertions that can be screened per F_2 family, the mutagenesis rate in zebrafish insertional mutagenesis screens using retroviruses is an estimated seven- to tenfold lower than in equivalent genetic screens using ENU. This means that significantly larger numbers of insertion lines have to be screened in order to obtain the same number of mutant phenotypes as when using an ENU protocol. However, once an interesting phenotype resulting from an insertion has been identified, the affected gene can be cloned much faster than is normally the case for ENU-induced mutations (approximately 2 weeks for insertions compared to several months or more for ENU mutations).

To date, over 500 zebrafish mutants with retroviral insertions have been identified. The range of phenotypes, which can be obtained with either chemical or insertional mutagenesis, is comparable. As observed in genetic screens based on chemical mutagenesis, roughly one-third of these mutants show defects in the development of one or very few parts of the body, organs, or cell types, while two-thirds display general or pleiotropic phenotypes, such as necrosis, heart edema, or small heads and eyes.

A major advantage of mutagenesis approaches relying on the insertion of retroviral or transposon sequences is that they allow a very easy cloning of the disrupted genes via the inserted sequences. The experience with cloning the mutated genes in the zebrafish insertional mutants has shown that in 80% of the cases, it was sufficient to sequence up to a few kilobases of the genomic DNA flanking the inserted sequence to identify the gene. This has led to the cloning of over 280 disrupted genes in the zebrafish to date – including some whose function *in vivo* had not previously been described. In comparison, approximately 100 chemically mutated zebrafish genes were cloned to date.

The differences in the efficiencies of mutagenesis and identification of the mutated genes lead to a difference in the kinds of genes, which tend to be identified using the two methods. Owing to the very large numbers of mutants generated in chemical mutagenesis screens, generally only those displaying specific developmental phenotypes are kept. The comparatively lower numbers generated in insertional mutagenesis screens leave more capacity to also keep mutant lines with less specific phenotypes. The need to focus on specific phenotypes in chemical mutagenesis screens is increased by the fact that the effort required to clone a chemically mutated gene is considerably bigger when compared to the identification of a gene disrupted by an insertion. This results in mutations in genes with more pleiotropic effects during development being underrepresented in collections of mutants resulting from chemical mutagenesis screens. Additionally, the cloning of the majority of ENU-mutated zebrafish genes at least partially relied on a candidate gene approach (favoring the identification of known genes), while the cloning of insertion mutations is less biased. As a consequence, the majority of ENU mutations identified to date are in transcription factors as well as signaling molecules and their receptors. The insertion mutants, on the other hand, also identified a number of genes involved in more general

cellular processes, such as the control of the cell cycle, vesicle trafficking, protein sorting between organelles or protein glycosylation.

2.2.3 Radiation Mutagenesis

In the past, ionising radiation has been a commonly used means for inducing mutations in a wide range of organisms. Ionising radiation carries enough energy to cause double-strand breaks in DNA helices and thus leads to chromosomal rearrangements, such as deletions, inversions, and translocations. As the majority of mutations induced by ionising radiation tend to be large chromosomal deficiencies generally comprising more than one gene, the usefulness of such mutants in the study of gene function is limited. Moreover, many radiation-induced mutations tend to be dominant lethal, and therefore mutants carrying such mutations cannot be maintained. For these reasons, ionising radiation has only been used infrequently and on a comparatively small scale to generate zebrafish mutants, and, compared to chemical and insertional mutagenesis, radiation plays a marginal role in the zebrafish as a means to generate mutants.

There are however studies in which γ-radiation has been used to induce mutations in the zebrafish, and several mutants with developmental defects have been isolated. γ-rays were shown to efficiently induce mutations, for example, in mature sperm cells, fertilized eggs, and blastula embryos of the zebrafish with specific locus mutation rates of up to 10^{-2}. With the current protocols, the survival rate to adulthood is approximately 25% for zebrafish raised using irradiated sperm in *in vitro* fertilizations and between 25 and 50% for irradiated zygotes (http://zfin.org/zf_info/zfbook/chapt7/7.9.html).

2.3 Reverse Genetics

In recent years, genome sequencing projects have yielded an enormous wealth of information, including numerous predicted genes with, as of yet, unidentified functions. The elucidation of the functions of these genes is one of the major challenges in the years to come. Reverse genetics approaches offer efficient ways to address this challenge. They allow to specifically knock out or knock down genes of interest either by generating mutants in these genes or by preventing the translation of the corresponding mRNA(s) into proteins respectively. The function of the targeted genes is subsequently elucidated via analyses of the resulting phenotypes.

For two decades, the mouse was the only vertebrate model organism for which reverse genetics methods could be established. These methods are based on homologous recombination in embryonic stem (ES) cells: an endogenous locus is replaced by a homologous extra-genomic construct engineered to carry a desired mutation. Following a selection step, ES cells carrying the mutation are introduced into mouse blastocysts where they participate in the normal development and contribute progeny to all organs including the germ line. These chimeric mice (always males) are mated with wild-type females to generate heterozygous offspring, which is generally outcrossed to generate a larger number of heterozygous F_2 individuals. The F_2 is then incrossed to produce the first homozygous embryos.

Despite efforts to establish zebrafish ES cell lines, there is currently no protocol for ES cell–based knockout or targeted gene expression by homologous recombination in the zebrafish. However, the advent of reverse genetics technology in the form

of morpholino knockdown and, more recently, the TILLING (Targeting Induced Local Lesions In Genomes) approach had a significant impact on zebrafish research. Particularly, in combination with the zebrafish genome-sequencing project, which is nearing its completion, these two techniques will increasingly contribute to elucidate the function of novel genes. Taken together, the morpholino and TILLING techniques will allow reverse genetics approaches to be performed routinely and on a large scale in the zebrafish.

2.3.1 Morpholino Knockdown

Morpholinos provide a very rapid means to obtain a phenotype, and thus assay the *in vivo* function of a specific gene. The technique is based on injecting modified oligonucleotides that prevent the expression of the targeted gene by blocking the translation of the corresponding mRNA(s). As zebrafish produce large numbers of offspring, which develop externally, they are ideally suited to large-scale knockdown approaches. Importantly, their eggs are large enough for sufficient quantities of morpholinos to be injected to block the expression of target genes over the full duration of embryonic development. Since the establishment of protocols for morpholino knockdown in the zebrafish, the technique has rapidly spread through the zebrafish community and is now widely used to rapidly confirm or exclude candidate genes for mutants obtained in mutagenesis screens as well as to obtain phenotypes for novel genes. Particularly for the latter, the zebrafish will play an increasingly important role in the future.

When performing a targeted mutagenesis of a locus by homologous recombination in mice, it typically takes one year to obtain the first homozygous offspring and another six months to generate the first 100 homozygous embryos to work with. Importantly, one heterozygous female has to be sacrificed each time to obtain (on an average) two homozygous embryos. This makes targeted mutagenesis in the mouse a time-consuming, laborious, and costly process. In comparison, the knockdown of genes by morpholino injection in zebrafish is faster, as there is no need for time-consuming manipulations of ES cells or for the breeding and crossing of several generations of animals to obtain homozygous individuals. Instead, the desired phenotype can be observed in the "first" generation from within a few hours to days. Also, in a single experiment, hundreds of embryos can be injected in a couple of hours.

Morpholinos cannot be used in mice, mainly because the morpholinos are rapidly diluted during mouse development, and their effect is thus lost before a number of interesting processes occur. This is due to two features characteristic for mammalian development: Firstly, the eggs are small such that only relatively small quantities of morpholino can be injected. Secondly, the embryos rapidly grow following implantation into the uterus.

Morpholinos are short oligonucleotides (18- to 25-bases long). In contrast to naturally occurring nucleic acids, the backbone of morpholinos is chemically modified: Each of the four bases A, C, G, and T is linked to a six-membered morpholine ring (hence the name) instead of the sugar moieties normally occurring in nucleic acids. These rings are connected with each other via nonionic phosphorodiamidate units (see Fig. 6). This chemical modification renders morpholinos resistant to degradation by nucleases. Morpholinos are generally designed to be complementary (antisense) to the region including the START

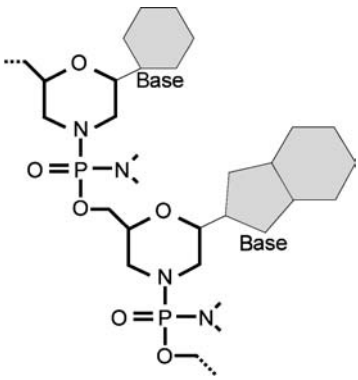

Fig. 6 Schematic representation of the backbone of a morpholino oligonucleotide (morpholino or MO). Instead of being linked to a ribose molecule (as is the case in RNA) or deoxyribose molecule (as in DNA), the bases in morpholinos are linked to a six-membered morpholine ring. These rings are connected via nonionic phosphorodiamidate groups (instead of an ionic phosphate group as in RNA and DNA). The bases linked to the morpholine ring are the same as in DNA: adenine, cytosine, guanine, and thymine.

codon of translation (ATG) or a crucial splice site in the mRNA of a targeted gene.

When injected into a cell, the morpholinos bind to their target sequence in the endogenous mRNA transcripts and block their translation into protein. The resulting effect of a morpholino injection is a phenotype that closely resembles or is identical to a loss-of-function mutant. In principle, any gene can be knocked down using morpholinos, including those that act early during development as maternally deposited mRNAs (a class for which it is relatively difficult to obtain homozygous mutants).

In zebrafish, morpholinos are generally injected into the egg yolk shortly after fertilization. They exert their effect almost immediately and effectively block the translation of the targeted mRNA for between three and four days of development (which in the zebrafish comprises the entire period of embryonic development). After that time, the effect of the morpholinos gradually diminishes as they are progressively diluted (and degraded) in the growing animal. As a consequence, only phenotypes that manifest themselves during this period can be observed, and genes acting later than three to four days after fertilization can generally not be targeted by morpholino injection.

Unlike nucleic acids used to knock down genes in other model organisms (“interfering RNA,” RNAi; see below), morpholinos are not charged and not toxic even at high concentrations, allowing large amounts to be injected. Varying the amount of morpholino injected makes it possible to attain different dose-dependent strengths of a phenotype – not unlike the situation observable in allelic series.

Morpholinos are also very well suited to overcome the problem of redundancy of gene functions – a weakness of mutagenesis screens and homologous recombination experiments. By simultaneously injecting two (or more) morpholinos targeted to related genes, these genes can be inactivated simultaneously. Similarly, morpholinos can be used to knock down additional genes in a mutant background. This way, the knockdown of gene functions through morpholinos is likely to contribute considerably to elucidating the genetic networks acting in development.

Despite their advantages, the knockdown with morpholinos has significant limitations. Morpholino injections can only lead to lack-of-function phenotypes and never serve to generate the variety

of sophisticated mutations achievable with homologous recombination in mice (e.g. conditional knockouts and knockins). Moreover, the dilution of the morpholino as the embryo grows restricts the time when phenotypes can be reliably identified and characterized to the first, approximately, three days of development. Also, a morpholino knockdown phenotype is not as reliable and reproducible as a genetic mutant. In many instances, the knockdown phenotypes will faithfully phenocopy the mutant phenotype, but the injection of morpholinos can result in nonspecific phenotypes, such as widespread cell death and neuronal degeneration. These have to be separated from genuine phenotypes by including a control morpholino in each experiment (e.g. a morpholino with the same sequence except for three mismatched bases). Furthermore, a morpholino targeted at a particular gene may in some instances "cross react" and block the production of other gene products. Finally, the knockdown by morpholino injection is transient and not passed on to subsequent generations. This necessitates new injections for each experiment that is being undertaken with knockdown fish. However, many of these drawbacks are overcome by the TILLING technique to generate knockout mutants.

2.3.2 TILLING

TILLING is an effective method to isolate mutants with defined mutations, including the knockout of genes. It constitutes a random mutagenesis and the subsequent screening for mutations in target genes. The method was initially developed for *Arabidopsis thaliana* but has since been used in other species, including the zebrafish.

In TILLING, mutations are randomly induced in the genome via chemical mutagenesis (e.g. with ENU as described above). The mutagenized fish are mated with wild-type animals to produce F_1 fish, which carry individual mutations in a heterozygous state (Fig. 7). However, instead of using these F_1 fish to raise F_2 families and subsequently screening their offspring for phenotypes as would be done in a genetic screen, the F_1 fish are directly examined for the presence of mutations. This is done by taking biopsies (fin clips) from which genomic DNA is isolated. This DNA is then screened for mutations in (known) genes of interest. Screening can be done either by sequencing parts of the genes with specific primer pairs flanking a region of interest or with an enzyme-mediated assay based on the property of the plant endonuclease CEL-I to recognize and cleave heteroduplexes. The possibility to automate certain steps of the TILLING procedure allows the screening of large numbers of mutagenized genomes for mutations in genes of interest. Once a mutation of interest has been identified in the genome of a particular fish, outcrossing this individual to a wild-type fish generates more offspring carrying the mutation. Incrossing of this F_2 generation results in homozygous F_3 individuals, which can be phenotypically analyzed.

The TILLING procedure has recently been successfully tested in the zebrafish and is now used on a large scale to generate zebrafish knockout mutants in a broad range of genes (see www.zf-models.org). In a test study, a library of 4608 male and female F_1 zebrafish derived from ENU-mutagenized founder fish was screened for mutations in 16 genes. The overall mutation frequency in this experiment was one mutation in approximately 235 000 bp. For the majority of target genes, only one PCR amplification product (amplicon) was screened for

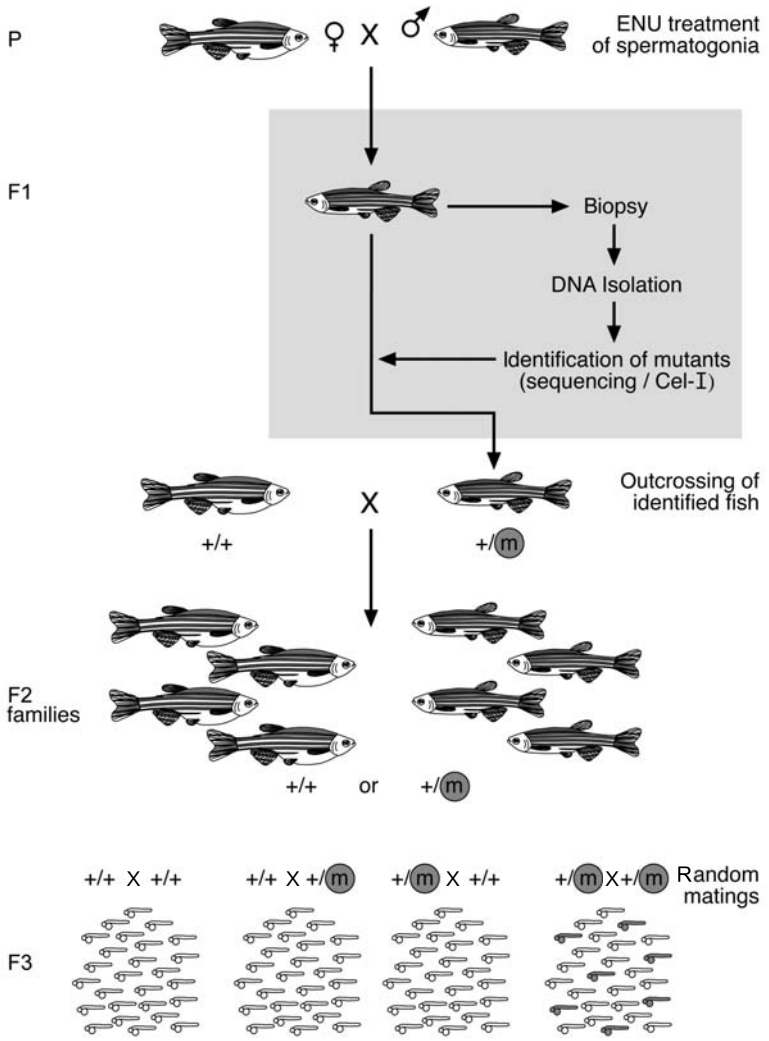

Fig. 7 Schematic illustration of the procedure used to generate, identify, and recover TILLING mutants. For details on the procedure see text.

mutations. The size of the amplicon was such that, on average, one-quarter of the coding sequence was screened. In addition to coding sequence, these amplicons comprised an average of 26% noncoding sequences, such as untranslated regions and intronic sequences.

This approach resulted in the identification of 255 mutations. The majority (80%) of all mutations found were located in coding sequences, whereas 20% were found in noncoding regions. Of the mutations identified in coding regions, most were missense mutations resulting in the exchange of an amino acid, roughly one-third were silent, and approximately one in ten were nonsense, resulting in a premature STOP codon (14 in total). Of the mutations

identified in noncoding regions, approximately 10% (seven in total) altered splice donor or acceptor consensus sequences and are presumed to cause a change in the splicing of the affected genes. A detailed characterization of the mutants arising from this TILLING project has not yet been done. However, it is presumed that the 14 nonsense mutations and the seven mutations affecting splice donor/acceptor sites will most likely give rise to complete loss-of-function or knockout phenotypes for these genes. This would signify that, in total, 21 null mutations (roughly 10%) could be identified in 13 of the 16 target genes in a period of a few months.

These pilot studies suggest that TILLING can be performed in a high-throughput setup in the zebrafish. Importantly, the study has shown that the TILLING technique is sufficiently sensitive to detect mutations against a background of naturally occurring single nucleotide polymorphisms (SNPs) – a finding of particular importance for species such as the zebrafish for which no extensively inbred lines exist. It must, however, be noted that in contrast to methods relying on homologous recombination in mice, TILLING cannot be used to introduce designed mutations. Instead, it relies on the spectrum of mutations inducible randomly by chemical mutagenesis.

2.3.3 RNA Interference

Similar to the knockdown with morpholinos, RNA interference (RNAi) is a method to transiently knockdown specific gene products. RNAi relies on a conserved cellular response to double-stranded RNA (dsRNA). When dsRNA is introduced into a cell, it leads to a series of events resulting in the sequence-specific degradation and translational inhibition of the homologous endogenous mRNA and thus the posttranscriptional silencing of the targeted gene. Since its discovery in *C. elegans* in 1998, RNAi as a mechanism to silence genes has been employed in a variety of fungi, plant and animal species, including model organisms such as *Neurospora, Arabidopsis, C. elegans, Drosophila*, and mice.

There have also been attempts to establish RNAi-mediated knockdown protocols for the zebrafish. However, the ratio of specific phenotypes was generally low and variable, and there have been studies reporting significant nonspecific effects of RNAi on the development of zebrafish. Although dsRNA-mediated posttranscriptional silencing appears to be an almost universal cellular mechanism and it should be possible to develop protocols to overcome these problems in principle, this has not yet been achieved, and morpholino injection is currently the method of choice to generate knockdown phenotypes in the zebrafish.

2.4 Parthenogenesis

As the fertilization of the eggs in zebrafish occurs outside the body of the female, the process can be manipulated experimentally, for instance, to induce the development of gynogenetic and androgenetic haploid or diploid embryos. The generation of individuals with a genome derived from only one parent (uniparental inheritance) can be useful for F_1 mutant screens as such screens require significantly fewer progeny to be raised in order to identify recessive mutations of interest than conventional diploid screens, which necessitate the production of an F_3 generation (see above). Moreover, gynogenetic and androgenetic individuals

can be used to determine the frequencies of meiotic recombination events during gametogenesis in either females or males, as well as to study genomic imprinting and sex determination.

For an overview of the procedures employed to generate gyno- and androgenetic haploid and diploid embryos, see Fig. 8.

Generation and development of gynogenetic haploid embryos. Haploid embryos are generated by fertilizing eggs with sperm that has previously been inactivated by irradiation with UV light. UV irradiation destroys the sperms' DNA but does not affect their capability to activate the eggs. The developing embryos accordingly

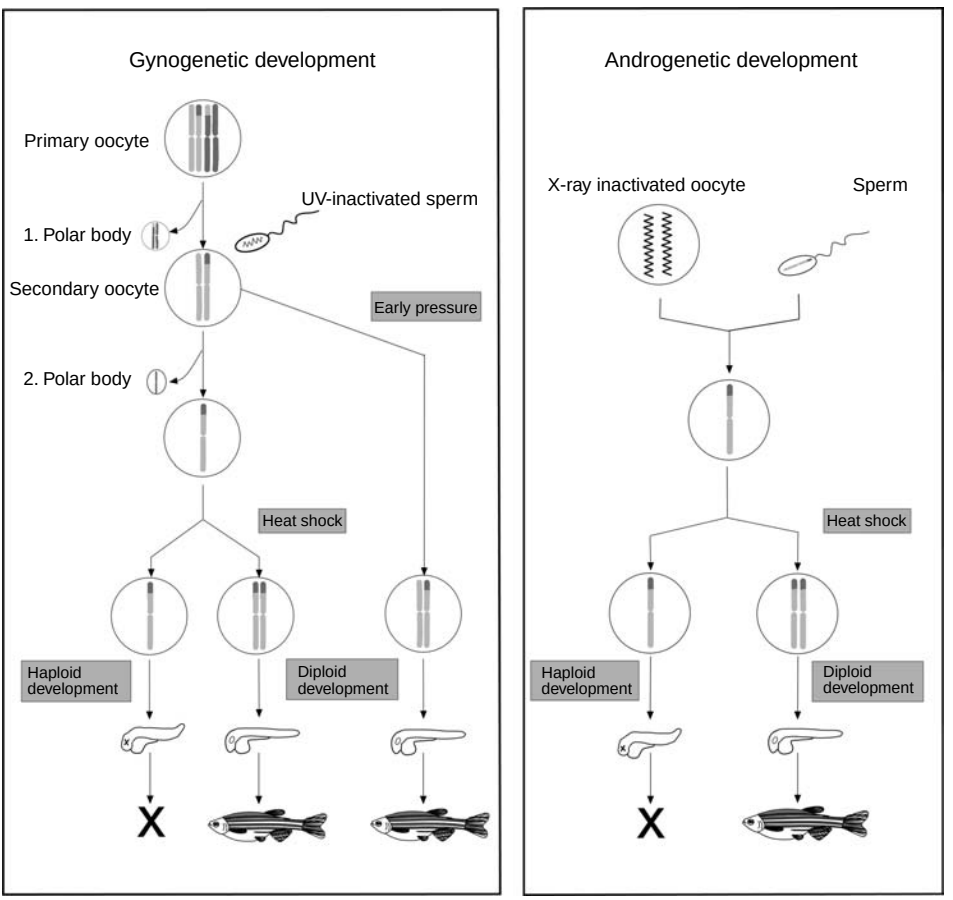

Fig. 8 Schematic representation of the methods used to generate gyno- and androgenetic haploid and diploid zebrafish embryos. For simplicity's sake, only one chromosome is shown. The colors (red and blue) denote the parental origin (maternal and paternal, respectively) of a chromosome/part of a chromosome in the oocyte or the sperm. The presence of two colors in a single chromosome arm indicates a recombination event(s) between the maternal and the paternal chromosome that occurred in the first meiosis leading to the germ cells. The X in the lower row signifies that the respective fish are not viable. For details on the procedures see text. (See color plate p. xxxi).

are haploid, lacking the paternal genetic complement. In consequence, any mutation(s) present in the maternal genome will manifest themselves in the developing haploid progeny. This can be exploited, for example, to directly screen the offspring of mutagenized females or of heterozygous F_1 females arising from the mating of a mutagenized male. If a mutation is present in both DNA strands of a locus (e.g. all insertions and all ENU mutations in the F_1), 50% of the progeny of these females will show the mutant phenotype. The phenotypes resulting from the induced mutations in the maternal genome are frequently similar to the phenotype in diploid individual homozygous for the mutation, although the haploid phenotypes tend to be stronger.

However, these advantages of haploid screens are counterbalanced by significant drawbacks. The viability of haploid embryos is significantly reduced when compared to diploid individuals. They complete embryogenesis but generally arrest quite early during larval development. Haploid embryos usually die by day four or five of development (none live to maturity). As a consequence, haploid screens are only really practical to uncover mutations that produce a phenotype within, approximately, the first three days of development. Moreover, haploid embryos typically display a set of background phenotypes, such as smaller cells, a shorter, thicker, and occasionally bent body axis, as well as defects in organogenesis (e.g. less distinct brain morphology, variable number of ears, defects in the cardiovascular system: swollen pericardial cavity, reduced circulation of blood) against which mutant phenotypes have to be identified. However, as these background phenotypes are relatively consistent, an experienced screener can often distinguish them from genuine phenotypes resulting from a mutation. Additionally, as with all genetic screens, the ratio of individuals showing a particular phenotype to those not showing the phenotype is a useful measure to decide. Nonetheless, the haploid background phenotypes may impede the observation of some phenotypes. Owing to these disadvantages, haploid screens are of only minor significance. However, they may prove valuable to identify mutations that affect early development with comparatively less effort. Once a mutation has been identified, the carrier fish can be bred to produce individuals diploid for the mutation.

Generation and development of gynogenetic diploid embryos. In contrast to haploid embryos, parthenogenetic diploid embryos display less severe developmental abnormalities, and can even survive to become fertile adults. This considerably facilitates the identification of mutant phenotypes and removes the restriction to day three of development up until which phenotypes can be identified in a haploid background. The generation and screening of gynogenetic diploid fish is particularly useful for the identification of maternal-effect mutations as well as of recessive mutations affecting adult traits.

Gynogenetic diploid embryos are generated by diploidizing haploid embryos produced through the fertilization of eggs with UV-inactivated sperm. Following the activation of an oocyte with UV-inactivated sperm, there are two phases during which there are two complements of the maternal genome in the zygote: prior to the completion of the second meiotic division (i.e. before the second polar body is extruded from the zygote) and later during the first mitosis (after the DNA has been replicated, but before zygote undergoes the first cleavage division). If either the extrusion of the

second polar body or the first cleavage is prevented, the embryo will possess a diploid maternal complement (see Fig. 8).

Exposure of eggs in the metaphase of the second meiotic division to a high hydrostatic pressure causes the spindle to disintegrate and the two complements of chromosomes are not separated. As a consequence, there is no extrusion of the polar body, and the egg carries a diploid female complement, which is replicated and distributed to the daughter cells as a normal diploid zygotic complement would. This method to prevent the formation of the second polar body is known as *early pressure* (*EP*). The survival rate following a diploidization with EP typically ranges between 50 and 80%. The gynogenetic diploid fish are viable, develop normally, and become fertile but are not as robust generally as diploid fish with a paternal and a maternal complement.

Despite the fact that the genome of EP-induced gynogenetic diploid fish derives from a single complement, not all loci are identical (i.e. homozygous). The heterozygosity is due to the fact that the sister chromatids present in the oocyte in the second meiosis have already undergone recombination during the first meiosis. The probability of a given locus to be homozygous in an embryo depends on the distance of this locus from the centromere. Loci closely linked to the centromere will be homozygous in nearly all embryos (i.e. 50% of embryos are homozygous for one allele and the other 50% are homozygous for the other allele). The greater the distance of a locus from the centromere, however, the fewer embryos will be homozygous for this locus. As a consequence, EP-based mutagenesis screens show a bias toward the identification of loci nearer the centromeres. Moreover, the fact that the ratio of phenotypically wild-type to mutant embryos is not constant owing to the unpredictable occurrence of recombination events, which hampers the reliable distinction of mutant phenotypes from background phenotypes.

An alternative method for generating gynogenetic diploid embryos is through a heat shock. The separation of the chromosomes and the ensuing cytokinesis of the first mitotic division can be prevented by exposing the egg to a heat shock (generally for 2 min at a temperature of 41 °C starting at 13 min following fertilization of the eggs with UV-inactivated sperm). The efficiency of this procedure to produce diploid embryos is not as high as that usually obtained with EP. Generally, only between 10 and 20% of the individuals survive the procedure and grow up to become fertile adults. Following the heat shock, development continues normally with cleavage divisions. What would have been the second such cleavage is now the first and produces the first two daughter cells.

The outcome of a diploidization induced by heat shock is different from that resulting from exposure of the egg to EP. In embryos created through heat shock, the meioses of the oocyte have been completed, and one set of sister chromatids of the remaining female complement of chromosomes extruded in a polar body. As the diploid state of these embryos is achieved through the replication of a single set of chromatids (i.e. one double helix), the embryos are homozygous for all loci, and 50% of the progeny will carry any given locus in a homozygous state. This advantage of (heat shock) HS-induced gynogenetic diploidy is, however, counterbalanced by the lower survival rates when compared to EP-based methods.

Generation and development of androgenetic haploid and diploid embryos. Similar to the production of gynogenetic diploid embryos carrying only a maternal genetic contribution, androgenetic haploid and diploid embryos can be produced. In these cases, the DNA of the oocytes is destroyed by exposing them to X-rays or UV light, and they are subsequently fertilized with normal sperm. The efficiency of androgenetic haploid production ranges from 10 to 50%. These embryos can either be left to develop as haploid androgenetic embryos or they can be diploidized by preventing the first mitotic cleavage division by heat shock as described above.

The androgenetic haploid and the diploid embryos also show distinct phenotypes. Like gynogenetic haploid embryos, androgenetic haploid embryos display a characteristic set of phenotypes, including a shorter, broader body as well as smaller melanocytes, and their development arrests around day four. The fact that both androgenetic and gynogenetic haploid embryos display very similar phenotypes indicates that these are due to the haploid state of these fish and independent of the parental origin of the genome. Androgenetic diploid embryos exhibit a retarded embryonic and larval development compared to normal diploid embryos, but later catch up in size and develop into fertile adult fish. However, the irradiation of the eggs necessary for androgenesis frequently causes sufficient cellular damage to make the phenotypic screening of the resulting embryos very difficult.

2.5 Transgenesis

The expression of exogenous genes or endogenous genes under the control of modified enhancer/promoter elements is a commonly used way of studying the functions of a gene *in vivo*. There are techniques to achieve both transient and stable expressions of transgenes in the zebrafish. For transient expression studies, the DNA constructs are generally injected into fertilized eggs or early embryos. However, these transient expressions generally result in only half of the injected embryos showing an expression in the expected cell types (e.g. as compared to *in situ* hybridization patterns obtained for the gene). Moreover, the expression is frequently not found in all the cells of the expected cell type. In order to achieve an expression of a transgene that faithfully mimics the expression of an endogenous promoter in the zebrafish, a stable transgenic line has to be created.

To achieve a stable transgenic zebrafish line, several techniques have been developed. These include the infection with retroviral vectors, penetration with DNA-coated microprojectiles, and electroporation. However, the introduction of transgenes is commonly achieved by microinjection of plasmid DNA into the embryo at the one-cell stage. This technique is fast (over 100 eggs can be injected in an hour) and comparatively gentle on the embryos. Furthermore, it results in high rates of integration of up to 25% for transient transfections and between 5 and 30% for stable transgenic lines (with germ lines mosaic for the transgene; between 1 and 50% of the F_1 offspring of these transgenic fish are hemizygous for the transgene).

In recent years, several genetic tools have become available that allow a sophisticated control and analysis of the expression of transgenes in zebrafish. These include:

- Reporter genes, such as the green fluorescent protein (GFP) as well as its derivative cyan (CFP) and yellow (YFP) versions under the control of

tissue- or cell type–specific promoters and enhancers for promoter/enhancer studies or cell lineage experiments.
- Promoters (e.g. the hsp70 promoter) that can be activated by heat shock for analyses of the effect of a transgene at specific times during development.
- The Gal4-UAS system to efficiently drive the expression of transgenes.

Owing to the transparency of its embryos, the spatiotemporal expression of fluorescent transgenes, such as GFP, in the various organs and tissues can be particularly well observed and imaged. This possibility has been exploited in numerous studies and is currently being used in a GFP-based enhancer trap screen performed by the laboratory of T. Becker at the University of Bergen in Norway.

2.6 Microarrays

Microarray hybridization allows to study the expression of thousands of genes in parallel. Probes representing individual transcripts are spotted or synthesized on a glass slide or silicon chip, and hybridized with fluorescently labeled target cDNA from a mutant, knockdown, or other conditions of interest. Microarrays are either based on cDNA clones or oligonucleotides. For a species like zebrafish where a large amount of genome sequence has already been generated, oligonucleotide arrays are advantageous as cross-hybridization can be ruled out by selecting unique oligonucleotide sequences for the probes. Two sets of 16 399 65-mer and 14 240 50-mer zebrafish oligonucleotides, as well as a photolithographically synthesized array representing 14 900 transcripts, are commercially available for spotting. Microarrays have only recently been introduced in the zebrafish field but hold great promise to unravel regulatory pathways in development and to discover genes of interest that are not accessible by reverse genetics methods because of redundant functions.

3 Mapping of the Zebrafish Genome

3.1 Genetic Mapping

Genetic mapping of mutant loci in the zebrafish can be carried out either by half-tetrad analysis or by genome scanning, usually with microsatellite markers. The latter is generally the method of choice because it is technically less demanding and more accurate.

Four genetic maps have been generated for the zebrafish (HS, MGH, MOP, and GAT; cf. Table 2). The established reference map for the zebrafish is the meiotic MGH map produced by the laboratory of M. Fishman at the Massachusetts General Hospital, Boston. It is based on microsatellite markers (CA repeats) scored on a panel of 44 diploid F_2 fish from a cross between the strains AB and India. When the MGH map was published in 1998, it contained 705 markers covering 2350 centimorgans (cM) and was already complete, that is it had one linkage group for each of the 25 zebrafish chromosomes. Production of the map continued until 2001 when a total of 3845 markers was scored.

In order to place a zebrafish mutation on the MGH map, mutant carriers are crossed with a polymorphic reference line. F_1 brothers and sisters are mated, and phenotypically mutant F_2 fish and their wild-type siblings are collected. A genome scan is then carried out by amplifying a subset of the MGH markers both from

Tab. 2 Mapping panels for the zebrafish.

Name	*Type*	*Markers placed as of February 2004*	*Laboratory*
An integrated map of the zebrafish genome (ZMAP)	Computed	30 566 (1)	J. Postlethwait
Goodfellow T51 (T51)	Radiation Hybrid	14 989 (2)	R. Geisler, Y. Zhou
Loeb/NIH/5000/4000 (LN54)	Radiation Hybrid	4192 (1)	I. Dawid
Heat shock (HS)	Meiotic	3888 (1)	J. Postlethwait, W. Talbot
Boston MGH Cross (MGH)	Meiotic	3881 (1)	M. Fishman
Mother of Pearl (MOP)	Meiotic	676 (1)	J. Postlethwait
Gates et al(GAT)	Meiotic	383 (1)	W. Talbot

Sources: (1) http://zfin.org
(2) http://wwmap.tuebingen.mpg.de

a mutant and from a sibling pool and comparing the band intensities of the marker alleles between the pools (batched segregant analysis, Fig. 9). If a significant difference is detected, the marker in question is likely to be linked to the mutation. This potential linkage can be confirmed by repeating the PCR with the individual mutant and wild-type F_2 fish that were used for the pooling, and counting the recombinations. An unambiguous map position can be obtained by identifying linkage to additional markers. To date, over 400 zebrafish mutations have been mapped at a resolution of approximately 1 cM (600 kb) by genome scanning with a set of 192 microsatellite markers, mainly by the laboratory of R. Geisler. An alternative to this approach is the typing of single nucleotide polymorphism (SNP) markers by microarray hybridization, which has been demonstrated by W. Talbot. Microarrays offer the prospect of performing batched segregant analysis for a very large number of markers in parallel once sufficiently accurate detection methods are available.

Results of such rough mapping can be sufficient for cloning a mutation of interest if a suitable candidate gene can be identified, and if evidence such as a matching expression pattern or a morpholino-induced knockdown phenotype can be obtained for the gene. In the absence of good candidate genes or supporting evidence, however, the map position needs to be refined. This involves identifying the closest polymorphic MGH markers and collecting recombinants for these markers from a large number of individuals (typically 1000–2000 embryos, corresponding to 2000–4000 meioses). The recombinants are then used for mostly SNP-based mapping of genomic clones or expressed sequence tags (ESTs) from the region of the mutation. SNP are typically identified and typed by sequencing. Genome sequencing showed that SNPs within the Tü strain occur at an average frequency of 1/200 bp. In the vicinity of the mutation, recombinations should be lost. The region in which no recombinations are detected on either side is referred to as the critical region; when starting out with a sufficient number of recombinants, it can be reduced to a single contig of the zebrafish genome sequence, or even to a single genomic clone. Transcripts of the critical region are

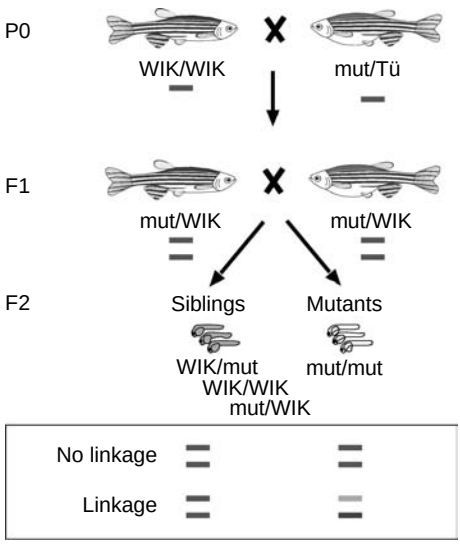

Fig. 9 Map cross and bulked segregant analysis of a recessive zebrafish mutation. WIK/WIK, reference fish; mut/Tü, fish carrying the mutation in Tü background. Band sizes and intensities of a representative SSLP marker are indicated schematically. In case of no linkage between the mutation and the marker, the intensities in the mutant and sibling pool are the same. In case of a linkage, the Tü band is stronger in the mutant pool and the WIK band in the sibling pool (Figure from Geisler, R. (2002) Mapping and Cloning, in: Nüsslein-Volhard, C., Dahm, R. (Eds.) *Zebrafish, A Practical Approach*, Oxford University Press, Oxford, UK, pp. 175–212.) (See color plate p. xxx).

identified and, if necessary, cloned based on annotation available at the ENSEMBL website, and then sequenced in order to find the mutation.

3.2 Radiation Hybrid Mapping

Mapping panels for radiation hybrid (RH) mapping are generated by fusing lethally irradiated cells of the species of interest with mouse or hamster cells. Hybrid cell lines are then grown, each of which retains a large number of random fragments of the species of interest, either inserted into the host chromosomes or as separate minichromosomes. Markers to be mapped are tested for their presence or absence in DNA of the cell lines by PCR. If two markers are present in many of the same RH lines (i.e. they have a similar RH score vector), they are likely to be close to each other in the genome. By ordering markers such that the ones with the most similar score vectors are next to each other, maps can be generated. Radiation hybrid mapping is thus similar to genetic mapping, with radiation-induced breaks in place of genetic recombination sites. Distances are given in centirays (1 cR indicates 1% breakage). The main advantages over genetic mapping are a higher resolution and the ability to map almost any sequence-tagged site (STS) marker without the need to identify a polymorphism (since RH lines are interspecies hybrids). For these reasons, RH mapping has become the method of choice for mapping of genes and genomic clones in the zebrafish.

The T51 radiation hybrid panel for the zebrafish was produced by the laboratory of P. Goodfellow. A map of the zebrafish genome based on the T51 panel containing 1451 STSs was generated by the laboratory of R. Geisler in 1999, demonstrating the feasibility of radiation hybrid mapping in a nonmammalian vertebrate. Since then the laboratories of R. Geisler and Y. Zhou have added several thousand STS markers to the map, mainly ESTs and BAC ends. As of January 2004, 14 989 markers could be placed on the T51 map, corresponding to an average marker spacing of 100 kb. The map is anchored on the meiotic MGH

map by microsatellite markers scored on both panels.

The LN54 radiation hybrid panel for the zebrafish was produced by a consortium led by I. Dawid. It has a lower resolution than the T51 panel and was used for the mapping of 4192 markers. Mapping services are available for the RH maps; markers are either mapped on request or RH DNA is provided to interested researchers by the mapping labs, and the resulting data can be submitted online.

Data from these mapping efforts have been used for studies of the conserved synteny between zebrafish and human, which revealed that 80% of the genes belong to conserved synteny groups. Although a genome duplication is believed to have taken place in the teleost lineage, most human genes have only one ortholog in zebrafish, perhaps indicating an extensive loss of duplicated genes (and those paralog pairs that do exist in zebrafish are sufficiently diverged so that they do not pose a problem for mapping).

3.3 Physical Mapping

A physical map of even higher resolution than a radiation hybrid map can be obtained by fingerprinting of genomic clones with large inserts such as Bacterial Artificial Chromosomes (BACs, 100–200 kb insert size). This method identifies contigs made up of overlapping genomic clones. Overlaps are detected either by hybridization with oligonucleotide probes (oligonucleotide fingerprinting) or by restriction digest (restriction fingerprinting). In the former case, clones are deemed to overlap if they hybridize with the same probes, in the latter, if they share a sufficient number of restriction fragments of similar size. In the zebrafish, the oligonucleotide fingerprinting method was used only on chromosome 20, while restriction fingerprinting was carried out on the entire genome by the laboratories of R. Geisler, S. Humphray, and R. Plasterk. Four BAC libraries were constructed from the Tü line for the purposes of fingerprinting and sequencing. A total of 213 000 BACs from these four libraries were fingerprinted by restriction digest with the enzyme HindIII, resulting in 3775 contigs with a combined 21-fold genome coverage. Contig data are available online for use in the positional cloning of zebrafish mutations, and are also used by the Sanger Institute to select BACs for the clone-based sequencing of the zebrafish genome. Furthermore, the contigs are being anchored on the T51 radiation hybrid map by mapping of BAC ends in a collaboration of R. Geisler and Y. Zhou. Most of the contigs are expected to be anchored by the end of 2004. The resulting physical map of the zebrafish genome is used by the ENSEMBL project for assembling of the genome sequence.

Cytogenetics in the zebrafish is lagging behind since the 25 chromosome pairs are relatively small and similar in size and are therefore difficult to identify visually. No sex chromosomes are known. An alternative is offered by fluorescent *in situ* hybridization (FISH), but because of the relative ease of genetic mapping, this is not commonly carried out in zebrafish, and FISH data are available only for some chromosomes. As a consequence of this, the zebrafish community agreed in 2003 to number zebrafish chromosomes following the numbering of the genetic linkage groups (LGs), in contrast to other organisms where chromosomes are numbered by size.

3.4 Genome Sequencing

The sequencing of zebrafish transcripts is largely based on the EST approach. ESTs are obtained by single-pass sequencing of cDNA clones and offer rapid access to transcribed sequences as well as information on genetic polymorphism. In the zebrafish, EST sequencing has been carried out mostly by the WashU-Zebrafish Genome Resources Project led by S. L. Johnson, which has generated over 200 000 ESTs, while smaller numbers have been produced by the laboratories of C. C. Liew and J. Peng. The WashU ESTs have been and assembled in over 25 000 "WZ" clusters, which represent the majority of all zebrafish genes, while those of J. Peng have been assembled in "IMBC" clusters. In a separate effort, zebrafish "UniGene" clusters are being assembled by the NCBI. Complementing these efforts, the Zebrafish Gene Collection is a recent NIH-sponsored effort to generate full-length cDNA libraries and sequences. Curated information on zebrafish genes is available from NCBI's LocusLink service as well as from the ZFIN and ENSEMBL websites.

In February 2001, the Sanger Institute began a zebrafish genome project, which follows a combination of two strategies: sequencing of BAC clones selected from the physical map of the genome and whole-genome shotgun sequencing. The BACs are selected to represent a minimum tiling path across the fingerprint contigs. The genome size is estimated to be 1.5 to 1.6 Gb. As of February 2004, 1.0 Gb of BAC clone sequences (0.6-fold coverage) and 9.1 Gb of whole-genome shotgun sequences (5.7-fold coverage) have been generated. In addition, sequencing of BAC ends has been carried out both by the Sanger Institute and by the laboratory of S. Schuster. A finished genome sequence (with approximately 10-fold coverage) is expected to be available by the end of 2005.

Assembly of the zebrafish genome sequence has proved more difficult than in other organisms because of two factors:

1. There is a high degree of polymorphism that is visible in genome sequencing because both the BACs and the whole-genome shotgun library used for sequencing were made from sperm or embryos derived from several adult fish (due to constraints of the DNA isolation method employed).
2. Additionally, the zebrafish genome contains a high number of repeats. A repeat library was compiled by R. Waterman that contains repeats found more than 50 times and which covers approximately 40% of the genome. Masking of repeats is therefore essential when performing similarity searches or designing PCR primers for zebrafish sequences.

As of February 2004, the sequence was assembled into 62 895 supercontigs (which still contain short stretches of unknown sequence). These supercontigs are mostly based on FPC contigs, which are in turn anchored on the T51 RH map by BAC ends, thus providing rough chromosomal positions. The genome sequence is subsequently annotated by the ENSEMBL project of the Sanger Institute and the European Molecular Biology Laboratory and is accessible with annotations through a genome browser on the ENSEMBL website.

4 Resources on the Internet

4.1 General

The Zebrafish Information Network ⟨http://zfin.org/⟩

Community database and repository of genome data, which provides information on genes, maps, stocks, expression patterns, publications, researchers, and suppliers.

4.2 Mapping

MGH Zebrafish WWW Server ⟨http://zebrafish.mgh.harvard.edu/⟩

Homepage for the MGH genetic map.

Tübingen Map of the Zebrafish Genome ⟨http://wwwmap.tuebingen.mpg.de/⟩

Homepage for the T51 RH map (calculated with RH TSP Map software), mutant mapping, data submission to the T51 map.

The Children's Hospital Zebrafish Genome Project Initiatives

⟨http://zfrhmaps.tch.harvard.edu/ZonRHmapper/⟩

Alternate homepage for the T51 RH map (calculated with SAMapper), instant mapping service.

4.3 Sequencing

ENSEMBL Zebrafish Genome Browser ⟨http://www.ensembl.org/Danio_rerio/⟩

Assembled and annotated zebrafish genome sequence.

The Sanger Institute: Danio rerio Sequencing Project ⟨http://www.sanger.ac.uk/Projects/D_rerio/⟩

Information and links on genome sequencing, database of fingerprinting contigs, online RepeatMasker.

WashU-Zebrafish Genome Resources Project ⟨http://zfish.wustl.edu/⟩

Database of ESTs and EST clusters, BLAST server for EST clusters.

NCBI Zebrafish Genome Resources ⟨http://www.ncbi.nlm.nih.gov/genome/guide/zebrafish/index.html⟩

Homepage for NCBI resources and services including sequence databases, BLAST, and a map viewer.

Zebrafish Gene Collection ⟨http://zgc.nci.nih.gov/⟩

Full-length cDNA libraries and sequences.

4.4 Resource Centers

Zebrafish International Resource Center ⟨http://zfin.org/zf_info/stckctr/stckctr.html⟩

Wild-type and mutant zebrafish stocks, miscellaneous clones, antibodies and materials, zebrafish pathology, and health services.

RZPD German Resource Center for Genome Research ⟨http://www.rzpd.de/⟩

cDNA and genomic libraries, library pools and filters, and individual clones.

See also Genetics, Molecular Basis of; Genomic Sequencing (Core Article); RNA Interference; Shotgun Sequencing (SGS); Transgenic Fish; Whole Genome Human Chromosome Physical Mapping.

Bibliography

Books and Reviews

Amemiya, C.T., Zhong, T.P., Silverman, G.A., Fishman, M.C., Zon, L.I. (1999) Zebrafish

YAC, BAC, and PAC genomic libraries, *Methods Cell Biol.* **60**, 235–258.

Amsterdam, A. (2003) Insertional mutagenesis in zebrafish, *Dev. Dyn.* **228**, 523–534.

Clack, J.A. (2002) *Gaining Ground – The Origin and Evolution of Tetrapods*, Indiana University Press, Indiana, IN, p. 369

Detrich, H.W., Westerfield, M., Zon, L.I. (Eds.) (1999a) *The Zebrafish: Biology*, Methods in Cell Biology Series, Vol. 59, Academic Press, San Diego, CA.

Detrich, H.W., Westerfield, M., Zon, L.I. (Eds.) (1999b) *The Zebrafish: Genetics and Genomics*, Methods in Cell Biology Series, Vol. 60, Academic Press, San Diego, CA.

Frohnhöfer, H.G. (2002) Table of Zebrafish Mutations, in: Nüsslein-Volhard, C., Dahm, R. (Eds.) *Zebrafish, A Practical Approach*, Oxford University Press, Oxford, UK, pp. 237–292.

Geisler, R. (2002) Mapping and Cloning, in: Nüsslein-Volhard, C., Dahm, R. (Eds.) *Zebrafish, A Practical Approach*, Oxford University Press, Oxford, UK, pp. 175–212.

Gilmour, D.T., Jessen, J.R., Lin, S. (2002) Manipulating Gene Expression in the Zebrafish, in: Nüsslein-Volhard, C., Dahm, R. (Eds.) *Zebrafish, A Practical Approach*, Oxford University Press, Oxford, UK, pp. 121–143.

Gong, Z. (1999) Zebrafish expressed sequence tags and their applications, *Methods Cell Biol.* **60**, 213–233.

Hannon, G.J. (2002) RNA interference, *Nature* **418**, 244–251.

Johnson, S.L., Zon, L.I. (1999) Genetic backgrounds and some standard stocks and strains used in zebrafish developmental biology and genetics, *Methods Cell Biol.* **60**, 357–359.

Kimmel, C.B. (1989) Genetics and early development of zebrafish, *Trends Genet.* **5**, 283–288.

Kwok, C., Critcher, R., Schmitt, K. (1999) Construction and characterization of zebrafish whole genome radiation hybrids, *Methods Cell Biol.* **60**, 287–302.

Nüsslein-Volhard, C., Dahm, R. (Eds.) (2002) *Zebrafish, A Practical Approach*, Practical Approach Series, Oxford University Press, Oxford, UK, p. 303.

Pelegri, F. (2002) Mutagenesis, in: Nüsslein-Volhard, C., Dahm, R. (Eds.) *Zebrafish, A Practical Approach*, Oxford University Press, Oxford, UK, pp. 145–174.

Pelegri, F., Schulte-Merker, S. (1999) A Gynogenesis-Based Screen for Maternal-Effect Genes in the Zebrafish, *Danio rerio*, in: Detrich, H.W., Westerfield, M., Zon, L.I. (Eds.) *The Zebrafish: Genetics and Genomics*, Methods in Cell Biology Series, Vol. 60, Academic Press, San Diego, CA, p. 1–20.

Sola, L., Gornung, E. (2001) Classical and molecular cytogenetics of the zebrafish, Danio rerio (Cyprinidae, Cypriniformes): an overview, *Genetica* **111**, 397–412.

Talbot, W.S., Schier, A.F. (1999) Positional cloning of mutated zebrafish genes, *Methods Cell Biol.* **60**, 259–286.

Till, B.J., Colbert, T., Tompa, R., Enns, L.C., Codomo, C.A., Johnson, J.E., Reynolds, S.H., Henikoff, J.G., Greene, E.A., Steine, M.N., Comai, L., Henikoff, S. (2003) High-throughput TILLING for functional genomics, *Methods Mol. Biol.* **236**, 205–220.

Walker, C. (1999) Haploid Screens and Gamma-Ray Mutagenesis, in: Detrich, H.W., Westerfield, M., Zon, L.I. (Eds.) *The Zebrafish: Genetics and Genomics*, Methods in Cell Biology Series, Vol. 60, Academic Press, San Diego, CA, pp. 44–70.

Primary Literature

Amsterdam, A., Burgess, S., Golling, G., Chen, W., Sun, Z., Townsend, K., Farrington, S., Haldi, M., Hopkins, N. (1999) A large-scale insertional mutagenesis screen in zebrafish, *Genes Dev.* **13**, 2713–2724.

Barbazuk, W.B., Korf, I., Kadavi, C., Heyen, J., Tate, S., Wun, E., Bedell, J.A., McPherson, J.D., Johnson, S.L. (2000) The syntenic relationship of the zebrafish and human genomes, *Genome Res.* **10**, 1351–1358.

Chen, W., Burgess, S., Golling, G., Amsterdam, A., Hopkins, N. (2002) High-throughput selection of retrovirus producer cell lines leads to markedly improved efficiency of germ line-transmissible insertions in zebra fish, *J. Virol.* **76**, 2192–2198.

Colbert, T., Till, B.J., Tompa, R., Reynolds, S., Steine, M.N., Yeung, A.T., McCallum, C.M., Comai, L., Henikoff, S. (2001) High-throughput screening for induced point mutations, *Plant Physiol.* **126**, 480–484.

Corley-Smith, G.E., Lim, C.J., Brandhorst, B.P. (1996) Production of androgenetic zebrafish (*Danio rerio*), *Genetics* **142**, 1265–1276.

Cresko, W.A., Yan, Y.L., Baltrus, D.A., Amores, A., Singer, A., Rodriguez-Mari, A., Postlethwait, J.H. (2003) Genome duplication, subfunction partitioning, and lineage divergence: Sox9 in stickleback and zebrafish, *Dev. Dyn.* **228**, 480–489.

Davidson, A.E., Balciunas, D., Mohn, D., Shaffer, J., Hermanson, S., Sivasubbu, S., Cliff, M.P., Hackett, P.B., Ekker, S.C. (2003) Efficient gene delivery and gene expression in zebrafish using the sleeping beauty transposon, *Dev. Biol.* **263**, 191–202.

Driever, W., Solnica-Krezel, L., Schier, A.F., Neuhauss, S.C., Malicki, J., Stemple, D.L., Stainier, D.Y., Zwartkruis, F., Abdelilah, S., Rangini, Z., Belak, J., Boggs, C. (1999) A genetic screen for mutations affecting embryogenesis in zebrafish, *Development* **123**, 37–46.

Ekker, M., Ye, F., Joly, L., Tellis, P., Chevrette, M. (1999) Zebrafish/mouse somatic cell hybrids for the characterization of the zebrafish genome, *Methods Cell Biol.* **60**, 303–321.

Felsenfeld, A.L., Walker, C., Westerfield, M., Kimmel, C., Streisinger, G. (1990) Mutations affecting skeletal muscle myofibril structure in the zebrafish, *Development* **108**, 443–459.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*, *Nature* **391**, 806–811.

Fürnzler, D., Her, H., Knapik, E.W., Clark, M., Lehrach, H., Postlethwait, J.H., Zon, L.I., Beier, D.R. (1998) Gene mapping in zebrafish using single-strand conformation polymorphism analysis, *Genomics* **51**, 216–222.

Gaiano, N., Amsterdam, A., Kawakami, K., Allende, M., Becker, T., Hopkins, N. (1996) Insertional mutagenesis and rapid cloning of essential genes in zebrafish, *Nature* **383**, 829–832.

Gates, M.A., Kim, L., Egan, E.S., Cardozo, T., Sirotkin, H.I., Dougan, S.T., Lashkari, D., Abagyan, R., Schier, A.F., Talbot, W.S. (1999) A genetic linkage map for zebrafish: comparative analysis and localization of genes and expressed sequences, *Genome Res.* **9**, 334–347.

Geisler, R., Rauch, G.-J., Baier, H., van Bebber, F., Broß, L., Davis, R.W., Dekens, M., Finger, K., Fricke, C., Gates, M.A., Geiger, H., Geiger-Rudolph, S., Gilmour, D., Glaser, S., Gnügge, L., Habeck, H., Hingst, K., Holley, S., Keenan, J., Kirn, A., Knaut, H., Lashkari, D., Maderspacher, F., Martyn, U., Neuhauss, S., Neumann, C., Nicolson, T., Pelegri, F., Ray, R., Rick, J., Roehl, H., Roeser, T., Schauerte, H.E., Schier, A.F., Schönberger, U., Schönthaler, H.-B., Schulte-Merker, S., Seydler, C., Talbot, W.S., Weiler, C., Nüsslein-Volhard, C., Haffter, P. (1999) A radiation hybrid map of the zebrafish genome, *Nat. Genet.* **23**, 86–89.

Goff, D.J., Galvin, K., Katz, H., Westerfield, M., Lander, E.S., Tabin, C.J. (1992) Identification of polymorphic simple sequence repeats in the genome of the zebrafish, *Genomics* **14**, 200–202.

Golling, G., Amsterdam, A., Sun, Z., Antonelli, M., Maldonado, E., Chen, W., Burgess, S., Haldi, M., Artzt, K., Farrington, S., Lin, S., Nissen, R., Hopkins, N. (2002) Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development, *Nat. Genet.* **31**, 135–140.

Gong, Z., Yan, T., Liao, J., Lee, S.E., He, J., Hew, C.L. (1997) Rapid identification and isolation of zebrafish cDNA clones, *Gene* **201**, 87–98.

Haffter, P., Granato, M., Brand, M., Mullins, M.C., Hammerschmidt, M., Kane, D.A., Odenthal, J., van Eeden, F.J., Jiang, Y.J., Heisenberg, C.P., Kelsh, R.N., Furutani-Seiki, M., Vogelsang, E., Beuchle, D., Schach, U., Fabian, C., Nüsslein-Volhard, C. (1996) The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*, *Development* **123**, 1–36.

Halpern, M.E., Ho, R.K., Walker, C., Kimmel, C.B. (1993) Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation, *Cell* **75**, 99–111.

Hatta, K., Kimmel, C.B., Ho, R.K., Walker, C. (1991) The cyclops mutation blocks specification of the floor plate of the zebrafish central nervous system, *Nature* **350**, 339–341.

Henion, P.D., Raible, D.W., Beattie, C.E., Stoesser, K.L., Weston, J.A., Eisen, J.S. (1996) Screen for mutations affecting development of zebrafish neural crest, *Dev. Genet.* **18**, 11–17.

Johnson, S.L., Africa, D., Horne, S., Postlethwait, J.H. (1995) Half-tetrad analysis in zebrafish: mapping the *ros* mutation and the centromere of linkage group I, *Genetics* **139**, 1727–1735.

Johnson, S.L., Gates, M.A., Johnson, M., Talbot, W.S., Horne, S., Baik, K., Rude, S., Wong, J.R.,

Postlethwait, J.H. (1996) Centromere-linkage analysis and consolidation of the zebrafish genetic map, *Genetics* **142**, 1277–1288.

Kauffman, E.J., Gestl, E.E., Kim, D.J., Walker, C., Hite, J.M., Yan, G., Rogan, P.K., Johnson, S.L., Cheng, K.C. (1995) Microsatellite-centromere mapping in the zebrafish (Danio rerio), *Genomics* **30**, 337–341.

Kawakami, K., Amsterdam, A., Shimoda, N., Becker, T., Mugg, J., Shima, A., Hopkins, N. (2000) Proviral insertions in the zebrafish hagoromo gene, encoding an F-box/WD40-repeat protein, cause stripe pattern anomalies, *Curr. Biol.* **10**, 463–466.

Kelly, P.D., Chu, F., Woods, I.G., Ngo-Hazelett, P., Cardozo, T., Huang, H., Kimm, F., Liao, L., Yan, Y.L., Zhou, Y., Johnson, S.L., Abagyan, R., Schier, A.F., Koster, R.W., Fraser, S.E. (2001) Tracing transgene expression in living zebrafish embryos, *Dev. Biol.* **233**, 329–346.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., Schilling, T.F. (1995) Stages of embryonic development of the zebrafish, *Dev. Dyn.* **203**, 253–310.

Knapik, E.W., Goodman, A., Ekker, M., Chevrette, M., Delgado, J., Neuhauss, S., Shimoda, N., Driever, W., Fishman, M.C., Jacob, H.J. (1998) A microsatellite genetic linkage map for zebrafish (Danio rerio), *Nat. Genet.* **18**, 338–343.

Knapik, E.W., Goodman, A., Atkinson, O.S., Roberts, C.T., Shiozawa, M., Sim, C.U., Weksler-Zangen, S., Trolliet, M.R., Futrell, C., Innes, B.A., Koike, G., McLaughlin, M.G., Pierre, L., Simon, J.S., Vilallonga, E., Roy, M., Chiang, P.W., Fishman, M.C., Driever, W., Jacob, H.J. (1996) A reference cross DNA panel for zebrafish (Danio rerio) anchored with simple sequence length polymorphisms, *Development* **123**, 451–460.

Kurita, K., Burgess, S.M., Sakai, N. (2004) Transgenic zebrafish produced by retroviral infection of in vitro-cultured sperm, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1263–1267.

Kwok, C., Korn, R.M., Davis, M.E., Burt, D.W., Critcher, R., McCarthy, L., Paw, B.H., Zon, L.I., Goodfellow, P.N., Schmitt, K. (1998) Characterization of whole genome radiation hybrid mapping resources for non-mammalian vertebrates, *Nucleic Acids Res.* **26**, 3562–3566.

Li, Y.X., Farrell, M.J., Liu, R., Mohanty, N., Kirby, M.L. (2000) Double-stranded RNA injection produces null phenotypes in zebrafish, *Dev. Biol.* **217**, 394–405.

Lin, S., Gaiano, N., Culp, P., Burns, J.C., Friedmann, T., Yee, J.K., Hopkins, N. (1994) Integration and germ-line transmission of a pseudotyped retroviral vector in zebrafish, *Science* **265**, 666–669.

Lo, J., Lee, S., Xu, M., Liu, F., Ruan, H., Eun, A., He, Y., Ma, W., Wang, W., Wen, Z., Peng, J. (2003) 15,000 unique zebrafish EST clusters and their use in microarray for profiling gene expression patterns during embryogenesis, *Genome Res.* **13**, 455–466.

McPherson, J.D., Chevrette, M., Dawid, I.B., Johnson, S.L., Ekker, M. (1999) Radiation hybrid mapping of the zebrafish genome, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9745–9750.

Moens, C.B., Yan, Y.L., Appel, B., Force, A.G., Kimmel, C.B. (1996) valentino: a zebrafish gene required for normal hindbrain segmentation, *Development* **122**, 3981–3990.

Mohideen, M.A., Moore, J.L., Cheng, K.C. (2000) Centromere-linked microsatellite markers for linkage groups 3, 4, 6, 7, 13, and 20 of zebrafish (*Danio rerio*), *Genomics* **67**, 102–106.

Molina, A., Di Martino, E., Martial, J.A., Muller, M. (2001) Heat shock stimulation of a tilapia heat shock protein 70 promoter is mediated by a distal element, *Biochem. J.* **356**, 353–359.

Nasevicius, A., Ekker, S.C. (2000) Effective targeted gene 'knockdown' in zebrafish, *Nat. Genet.* **26**, 216–220.

Nechiporuk, A., Finney, J.E., Keating, M.T., Johnson, S.L. (1999) Assessment of polymorphism in zebrafish mapping strains, *Genome Res.* **9**, 1231–1238.

Oates, A.C., Bruce, A.E., Ho, R.K. (2000) Too much interference: injection of double-stranded RNA has nonspecific effects in the zebrafish embryo, *Dev. Biol.* **224**, 20–28.

Postlethwait, J.H., Talbot, W.S. (2000) Genetic linkage mapping of zebrafish genes and ESTs, *Genome Res.* **10**, 558–567.

Postlethwait, J.H., Woods, I.G., Ngo-Hazelett, P., Yan, Y.-L., Kelly, P.D., Chu, F., Huang, H., Hill-Force, A., William, S.T. (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes, *Genome Res.* **10**, 1890–1902.

Postlethwait, J.H., Yan, Y.L., Gates, M.A., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E.S., Force, A., Gong, Z., Goutel, C., Fritz, A., Kelsh, R., Knapik, E., Liao, E., Paw, B., Ransom, D., Singer, A., Thomson, M., Abduljabbar, T.S., Yelick, P., Beier, D., Joly, J.S., Larhammar, D., Rosa, F., Westerfield, M., Zon, L., Johnson, S., Talbot, W.S. (1998) Vertebrate genome evolution and the zebrafish gene map, *Nat. Genet.* **18**, 345–349.

Postlethwait, J., Johnson, S., Midson, C.N., Talbot, W.S., Gates, M., Ballinger, E.W., Africa, D., Andrews, R., Carl, T., Eisen, J.S., Horne, S., Kimmel, C.B., Hutchinson, M., Johnson, M., Rodriguez, A. (1994) A genetic linkage map for the zebrafish, *Science* **264**, 699–703.

Rawls, J.F., Frieda, M.R., McAdow, A.R., Gross, J.P., Clayton, C.M., Heyen, C.K., Johnson, S.L. (2003) Coupled mutagenesis screens and genetic mapping in zebrafish, *Genetics* **163**, 997–1009.

Raz, E., van Luenen, H.G., Schaerringer, B., Plasterk, R.H., Driever, W. (1998) Transposition of the nematode *Caenorhabditis elegans* Tc3 element in the zebrafish *Danio rerio*, *Curr. Biol.* **8**, 82–88.

Scheer, N., Camnos-Ortega, J.A. (1999) Use of the Gal4-UAS technique for targeted gene expression in the zebrafish, *Mech. Dev.* **80**, 153–158.

Stickney, H.L., Schmutz, J., Woods, I.G., Holtzer, C.C., Dickson, M.C., Kelly, P.D., Myers, R.M., Talbot, W.S. (2002) Rapid mapping of zebrafish mutations with SNPs and oligonucleotide microarrays, *Genome Res.* **12**, 1929–1934.

Streisinger, G., Walker, C., Dower, N., Knauber, D., Singer, F. (1981) Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*), *Nature* **291**, 293–296.

Stuart, G.W., McMurray, J.V., Westerfield, M. (1988) Replication, integration and stable germ-line transmission of foreign sequences injected into early zebrafish embryos, *Development* **103**, 403–412.

Tawk, M., Tuil, D., Torrente, Y., Vriz, S., Paulin, D. (2002) High-efficiency gene transfer into adult fish: a new tool to study fin regeneration, *Genesis* **32**, 27–31.

Ton, C., Hwang, D.M., Dempsey, A.A., Tang, H.C., Yoon, J., Lim, M., Mably, J.D., Fishman, M.C., Liew, C.C. (2000) Identification, characterization, and mapping of expressed sequence tags from an embryonic zebrafish heart cDNA library, *Genome Res.* **10**, 1915–1927.

Ungar, A.R., Helde, K.A., Moon, R.T. (1998) Production of androgenetic haploids in zebrafish with ultraviolet light, *Mol. Mar. Biol. Biotech.* **7**, 320–326.

Wargelius, A., Ellingsen, S., Fjose, A. (1999) Double-stranded RNA induces specific developmental defects in zebrafish embryos, *Biochem. Biophys. Res. Commun.* **263**, 156–161.

Westerfield, M., Liu, D.W., Kimmel, C.B., Walker, C. (1990) Pathfinding and synapse formation in a zebrafish mutant lacking functional acetylcholine receptors, *Neuron* **4**, 867–874.

Wienholds, E., Schulte-Merker, S., Walderich, B., Plasterk, R.H. (2002) Target-selected inactivation of the zebrafish rag1 gene, *Science* **297**, 99–102.

Wienholds, E., van Eeden, F., Kosters, M., Mudde, J., Plasterk, R.H., Cuppen, E. (2003) Efficient target-selected mutagenesis in zebrafish, *Genome Res.* **13**, 2700–2707.

Woods, I.G., Kelly, P.D., Chu, F., Ngo-Hazelett, P., Yan, Y.L., Huang, H., Postlethwait, J.H., Talbot, W.S. (2000) A comparative map of the zebrafish genome, *Genome Res.* **10**, 1903–1914.

Xiao, T., Shoji, W., Zhou, W., Su, F., Kuwada, J.Y. (2003) Transmembrane sema4E guides branchiomotor axons to their targets in zebrafish, *J. Neurosci.* **23**, 4190–4198.

Yan, Y.L., Talbot, W.S., Egan, E.S., Postlethwait, J.H. (1998) Mutant rescue by BAC clone injection in zebrafish, *Genomics* **50**, 287–289.

Zardoya, R., Meyer, A. (1996) Evolutionary relationships of the coelacanth, lungfishes, and tetrapods based on the 28S ribosomal RNA gene, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5449–5454.

Zhao, Z., Cao, Y., Li, M., Meng, A. (2001) Double-stranded RNA injection produces nonspecific defects in zebrafish, *Dev. Biol.* **229**, 215–223.